UNIVERSITAT POLITÈCNICA DE VALÈNCIA

DEPARTAMENTO DE MATEMÁTICA APLICADA

MODEL-BASED ANALYSIS AND METABOLIC DESIGN OF A CYANOBACTERIUM FOR BIO-PRODUCTS SYNTHESIS

PhD Thesis by:

Julián Triana Dopico

Advisors:

Pedro José Fernández de Córdoba Castellá
Arnau Montagud Aquino
Javier Fermín Urchueguía Schölzel

Valencia, 2014
Insitut Universitari de Matemàtica Pura i Aplicada

Pedro José Fernández de Córdoba Castellá, Catedrático de Universidad del Departamento de Matemática Aplicada de la Universitat Politècnica de València, Arnau Montagud Aquino, Investigador del Institut Curie, París, y Javier Fermín Urchueguía Schölzel, Catedrático de Universidad del Departamento de Física Aplicada de la Universitat Politècnica de València,

INFORMAN

Que Julián Triana Dopico ha realizado bajo nuestra dirección la Tesis Doctoral titulada "Model-based analysis and metabolic design of a cyanobacterium for bio-products synthesis", con la que se presenta para optar al Grado de Doctor en Matemática y que, a nuestro juicio, reúne las condiciones de calidad exigibles para una Tesis Doctoral.

Y para que conste a los efectos oportunos, firmamos la presente en Valencia, a 24 de julio de 2014.

Pedro J. Fernández de Córdoba Castellá Arnau Montagud Aquino Javier F. Urchueguía Schölzel

Present *philosophiae doctor* thesis is framed within the scientific exchange and collaboration between the *University of Pinar del Rio* (UPR, Cuba), and the Interdisciplinary Modelling Group, InterTech (www.intertech.upv.es), linked to the *Institut Universitari de Matemàtica Pura i Aplicada* at *Universitat Politècnica de València* (UPV, Spain).

The Eagle soars in the summit of Heaven,

The Hunter with his dogs pursues his circuit.

O perpetual revolution of configured stars,

O perpetual recurrence of determined seasons,

O world of spring and autumn, birth and dying!

The endless cycle of idea and action,

Endless invention, endless experiment,

Brings knowledge of motion, but not of stillness;

Knowledge of speech, but not of silence;

Knowledge of words, and ignorance of the Word.

All our knowledge brings us nearer to our ignorance,

All our ignorance brings us nearer to death,

But nearness to death no nearer to God.

Where is the Life we have lost in living?

Where is the wisdom we have lost in knowledge?

Where is the knowledge we have lost in information? ...

Thomas Stearns Eliot, "Choruses from the Rock" ***(1940)***

Foreword

Otra cumbre que se alcanza en la compleja red de la vida. Un nodo diferente e imperecedero... Primero, porque la ruta hacia este supone muchos nodos intermedios que bifurcan el flujo de nuestros esfuerzos. Una vía, cuya regulación, implica que nuestra voluntad se conecte con nuestro objetivo. Segundo, porque representa un "hub" que involucrará encrucijadas hacia destinos ambivalentes: metas personales y profesionales. Así de ininteligible, suele ser.

No tendría sentido hablar de esta "red" sino menciono a todas las personas que forman parte de ella. Lo poco o lo mucho que he logrado, sin dudas, es gracias a que han estado presente en cada camino. Mi hogar ha sido el punto de partida. Atesoro una gran familia que simboliza mi estandarte y mi brújula. Todo lo que soy es consecuencia directa de la mímesis perfecta de mis padres. Todo el decoro es para ellos: mis padres, mi hermano menor y su esposa, y mi hermano mayor en la distancia. Mi familia grande: tíos y primos, constituyen parte indisoluble de cada conexión. La tranquilidad de contar con su apoyo incondicional, es reconfortante.

La sabiduría, el optimismo y la filantropía de mis directores, han hecho posible que este proyecto haya llegado a término. Mi infinito agradecimiento por aceptarme en su grupo y guiarme muchas veces desde la distancia. Está de más mencionar que en Cuba siempre tendrán su otra casa. De igual forma, quisiera resaltar que la culminación de esta tesis fue fruto del estrecho vínculo entre la ONG "InterTech-Cooperación" y la Universidad de Pinar del Río (UPR). Gracias a todos sus miembros, los que están y los que un día estuvieron, su altruismo y amparo me hicieron sentir como en casa.

El papel preponderante de todos mis amigos y compañeros de trabajo ha sido esencial para apuntalar esta tesis. Contar con su ayuda, ha sido estimulante. Amigos de dentro y fuera de Cuba, de estudio, de carrera, de trabajo y de la cotidianidad. Gracias por haber compartido el espacio de un aula, de una residencia estudiantil, de un departamento, de un piso, de la vida. Gracias por la conexión.

La gratitud, como ciertas flores, no se da en la altura y mejor reverdece en la tierra buena de los humildes

José Martí *(1853-1895)*

Abstract

The current investigation is aimed at the reconstruction and analysis of genome-scale metabolic models. Specifically, it is focused on the use of mathematical-computational simulations to predict the cellular metabolism behavior towards bio-products production. The photosynthetic cyanobacterium *Synechococcus elongatus* PCC7942 was studied as biological system.

This prokaryotic has been used in several studies as a biological platform for the synthesis of several substances for industrial interest. These studies are based on the advantage of autotrophic systems, which basically requires light and CO_2 for growth. The main objective of this thesis is the integration of different types of biological information, whose interaction can be extract applicable knowledge for economic interests. To this end, our study was addressed to the use of methods for modeling, analyzing and predicting the behavior of metabolic phenotypes of cyanobacterium.

The work has been divided into chapters organized sequentially, where the starting point was the *in silico* metabolic network reconstruction.

This process intent to join in a metabolic model of all chemical reactions codified in genome. The stoichiometric coefficients of each reactions, can be arranged into a sparse matrix (stoichiometric matrix), where the columns corresponds to reactions and rows to metabolites. As a result of this process the first model was obtained (*i*Syf646) than later was updated to another (*i*Syf715). Both were generated from data *-omics* published in databases, scientific reviews as well as textbooks. To validate them, each one of the stoichiometric matrix together with relevant constraints were used by simulation techniques based on linear programming. These reconstructions have to be flexible enough to allow autotrophic growth under which the organism grows in nature.

Once the reconstructions were validated, environmental variations can be simulated and we were able to study its effects through changes in outline system parameters. Subsequently, synthetic capabilities were evaluated from the *in silico* models in order to design metabolic engineering strategies. To do this a genetic variation was simulated in reactions network, where the disturbed stoichiometric matrix was the object of the quadratic optimization methods. As a result sets of

optimal solutions were generated to enhanced production of various metabolites of energetic and industrial interest such as: ethanol, higher chain alcohols, lipids and hydrogen.

Qualitatively distinct patterns of metabolic pathway utilization were identified by generation of phenotypic phase planes for biomass growth and synthesis of the bio-products as objective functions. We analyzed the variations of CO_2 and light uptakes rates over the genome-scale metabolic network.

Finally, genome-scale metabolic models allow us to establish criteria to integrate different types of data to help of find important points of regulation that may be subject to genetic modification. These regulatory centers have been investigated under drastic changes of CO_2 concentration on ambient and have been inferred operational principles of cyanobacterium metabolism.

In general, this thesis presents the metabolic capabilities of photosynthetic cyanobacterium *Synechococcus elongatus* PCC7942 to produce substances of interest, being a potential biological platform for clean and sustainable production.

Resumen

La presente investigación se orienta a la reconstrucción y análisis de los modelos metabólicos a escala genómica. Específicamente, se centra en el uso de las simulaciones matemático-computacionales para predecir el comportamiento del metabolismo celular hacia la producción de bio-productos. Como sistema biológico fue estudiado la cianobacteria fotosintética *Synechococcus elongatus* PCC7942.

Este procariota ha sido utilizado en diversos estudios como plataforma biológica para la síntesis de varias sustancias de interés industrial. Estos trabajos parten de la ventaja de este sistema autótrofo, el cual solo requiere de luz y CO_2 para su crecimiento. El principal objetivo de esta tesis es la integración de diferentes tipos de información biológica, de cuya interacción se pueda extraer conocimiento aplicable a intereses económicos. Para ello, nuestro estudio se dirigió al uso de métodos para modelar, analizar y predecir el comportamiento de los fenotipos metabólicos de la cianobacteria.

El trabajo ha sido estructurado en capítulos organizados secuencialmente, donde el punto de partida fue la reconstrucción *in silico* de la red metabólica de este microorganismo. Este proceso intenta agrupar en un modelo todas las reacciones químicas propias del metabolismo celular codificadas en el genoma. Los coeficientes estequiométricos de cada una de las reacciones del conjunto, pueden ser ordenados en una matriz dispersa (matriz estequiométricas), donde las columnas corresponden a las reacciones y las filas a los metabolitos. Como resultado de este proceso se obtuvo un primer modelo (*i*Syf646) que posteriormente fue actualizado a otro (*i*Syf715). Ambos fueron generados a partir de datos -*ómicos* publicados en bases de datos, revistas científicas así como en libros de texto. Para validarlos, las matrices estequiométricas de cada uno, junto a restricciones pertinentes, fueron utilizadas por técnicas de simulación basadas en programación lineal. Los modelos tenían que ser lo suficientemente flexible como para simular el crecimiento autotrófico bajo el cual este organismo crece en la naturaleza.

Una vez validadas las reconstrucciones, se pudieron simular variaciones ambientales y estudiar sus efectos mediante cambios en los parámetros de contorno del sistema. Seguidamente fueron evaluadas las capacidades de síntesis de los

modelos *in silico* con la finalidad de diseñar estrategias de ingeniería metabólica. Para ello fueron simuladas variaciones genéticas en la red de reacciones, donde las matrices estequiométricas perturbadas fueron objeto de métodos de optimización cuadrática. Como resultados se generaron conjuntos de soluciones óptimos hacia la producción mejorada de varios metabolitos de interés energético e industrial como son: etanol, alcoholes de cadena larga, lípidos e hidrógeno.

Fueron identificados cualitativamente distintos patrones de utilización de las vías metabólicas, mediante la generación de planos de fases fenotípicas para el crecimiento de la biomasa y la síntesis de bio-productos como funciones objetivos. Analizamos las variaciones de las velocidades de entrada de CO_2 y luz sobre el modelo metabólico a escala genómica.

Finalmente, los modelos metabólicos a escala genómica fueron utilizados para encontrar puntos importantes de regulación que pueden ser objeto de modificación genética. Estos centros reguladores han sido investigados bajo cambios drásticos de la concentración de CO_2 en el ambiente y se han inferido principios operacionales del metabolismo de esta cianobacteria.

En general, el estudio realizado en esta tesis presenta las capacidades metabólicas de la cianobacteria fotosintética *Synechococcus elongatus* PCC7942 para producir sustancias de interés, siendo una plataforma biológica potencial de producción limpia y sostenible.

Resum

La present investigació s'orienta a la reconstrucció i l'anàlisi dels models metabòlics a escala genòmica. Específicament, se centra en l'ús de les simulacions matemàtiques-computacionals per a predir el comportament del metabolisme cel·lular en la producció de bio-productes. Com a sistema biològic s'ha estudiat el cianobacteri fotosintètic *Synechococcus elongatus* PCC7942.

Aquest procariota ha sigut utilitzt en diversos estudis com a plataforma biològica per a la síntesi de diverses substàncies d'interès industrial. Aquests treballs parteixen de l'avantatge d'aquest sistema autòtrof, el qual sols requereix llum i CO_2 per al seu creixement. El principal objectiu d'aquesta tesi és la integració de diferents tipus d'informació biològica, de la interacció de la qual es pot extreure coneixement aplicable a interessos econòmics. Per a això, el nostre estudi s'ha dirigit a l'ús de mètodes per a modelar, analitzar i predir el comportament dels fenotips metabòlics del cianobacteri.

El treball ha estat estructurat en capítols organitzats seqüencialment, on el punt de partida fou la reconstrucció *in silico* de la xarxa metabòlica d'aquest microorganisme. Aquest procés intenta agrupar en un model totes les reaccions químiques pròpies del metabolisme cel·lular codificades al genoma. Els coeficients estequiomètrics de cadascuna de les reaccions del conjunt poden ser ordenats en una matriu dispersa (matriu estequiomètrica), on les columnes corresponen a les reaccions i les files als metabòlits. Com a resultat d'aquest procés s'obtingué un primer model (*i*Syf646) que posteriorment fou actualitzat a un altre (*i*Syf715). Ambdós foren generats a partir de dades òmiques publicades en bases de dades, revistes científiques i llibres de text. Per a validar-los, les matrius estequiomètriques de cadascun, així com les reaccions pertinents, foren utilitzades per tècniques de simulació basades en programació lineal. Els models havien de ser suficientment flexibles com per a simular el creixement autotròfic sota el qual aquest organisme creix a la natura.

Una vegada validades les reconstruccions, es pogueren simular variacions ambientals i estudiar els seus efectes mitjançant canvis en els paràmetres de contorn del sistema. Seguidament s´avaluaren les capacitats de síntesi dels models *in silico* amb la finalitat de dissenyar estratègies d'enginyeria metabòlica. Amb aquesta finalitat es

simularen variacions genètiques en la xarxa de reaccions, on les matrius estequiomètriques pertorbades foren objecte de mètodes d'optimització quadràtica. Com a resultats es generaren conjunts de solucions òptims cap a la producció millorada de diversos metabòlits d'interés energètic i industrial com ara: etanol, alcohols de cadena llarga, lípids i hidrogen.

S'identificaren qualitativament diferents patrons d'utilització de les vies metabòliques, mitjançant la generació de mapes de fases fenotípiques per al creixement de la biomassa i la síntesi de bio-productes com a funcions objectiu. Analitzàrem les variacions de les velocitats d'entrada de CO_2 i llum sobre el model metabòlic a escala genòmica.

Finalment, els models metabòlics a escala genòmica foren utilitzats per a trobar punts importants de regulació que poden ser objecte de modificació genètica. Aquests centres reguladors han sigut investigats sota canvis dràstics de la concentració de CO_2 en l'ambient i s'han inferit principis operacionals del metabolisme d'aquest cianobacteri.

En general, l'estudi realitzat en aquesta tesi presenta les capacitats metabòliques del cioanobacteri fotosintètic *Synechococcus elongatus* PCC7942 per a produir substàncies d'interès, com a plataforma biològica potencial de producció neta i sostenible.

MODEL-BASED ANALYSIS AND METABOLIC DESIGN OF A CYANOBACTERIUM FOR BIO-PRODUCTS SYNTHESIS

Index

Aims, Objectives and Thesis approach

Systems Biology represents a new approach to decoding life. The ability to generate detailed lists of biological components, determine their interactions and generate genome-wide datasets has led to the emergence of this discipline. These actions form the basis for computer modeling and simulation which are the main study objects in Systems Biology.

Complex biological processes can be simplified to mathematical models and analyze their functions by computer simulations. The process of building mathematical models and running computer simulations is iterative. The mathematical, *"in silico"*, models will have some analytical, interpretative, and predictive capabilities because the functional states of reconstructed networks are directly related to cellular phenotypes. The main difference between *in silico* and *in vivo* organism is that the *in silico* version is incomplete and missing some features. This means, that some features of the organisms have been preferred as research goals over others. Therefore, we must formulate experimentally testable hypotheses based on the *in silico* analysis, perform the experiments, and update the models.

On the other hand, organisms have to abide by a series of constraints, including those arising from basic natural laws, spatial constraints, and also from the environment in which they live. Many possible biological functions are achievable under these constraints, and organisms modify their behavior by imposing constraints through various regulatory mechanisms to select useful functional states from the allowable states. A constraints-based approach emerges from these considerations that enable the simultaneous analysis of physiochemical factors and biological properties.

Metabolism of an organism can be modeled into a network of metabolites and enzymes. This reconstructed network is called constraint-based stoichiometric models. This should integrate all genomic, genetic and biochemical information known for a given organism. Metabolic 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.

Constraint-based stoichiometric models can be used to study optimal behaviors, to assess possible genetic and environmental perturbations on the system, to integrate transcriptomic data into the metabolic network, and so on. The following work will study topics at this crossroad: the use of a biological system, in this case a cyanobacteria, in order to obtain bio-products, and understand their metabolism as a whole using mathematical models.

This thesis is devoted to the reconstruction and use of such model aimed at improving bio-products producing strategies in cyanobacterium *Synechococcus elongatus* PCC7942.

Objectives

The principal objectives of this dissertation are the following:

a) Reconstruct a genome-scale metabolic model for Synechococcus elongatus PCC7942.

Cyanobacterium *Synechococcus elongatus* PCC7942 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) Validate reconstructed metabolic models for Synechococcus elongatus PCC7942.

Model validation usually focuses on testing whether the growth capabilities, or any particular objective flux, correspond to a given set of experimental data. The validation of metabolic models is the starting point for the assessment of metabolic capabilities.

c) Analyze environmental and genetic variations imposed on the metabolic network under a systemic perspective.

Cyanobacterium *Synechococcus elongatus* PCC7942 will not be a desirable production platform if researchers do not know its behavior under perturbations. Genome-scale metabolic model allows the integrative study of the entire metabolism under such variations or mutations performed on its genome. This may allow detecting which variations are critical to the well-being of this organism.

d) Define strategies for the production of substances of socioeconomic importance through metabolic engineering designs.

Various metabolites have been identified as desirable products that can be produced from this organism: ethanol, higher chain alcohols, lipids and hydrogen. Their theoretical production limits need to be assessed and these enhanced-production mutants need to be studied and discussed under a system-wide perspective.

e) Analyze the discrete metabolic phenotypes

A finite number of qualitatively distinct patterns of metabolic pathway utilization can be identified in the optimal solution space (metabolic phenotypes) by a sensitivity analysis over the genome-scale metabolic network.

f) Integrate transcriptomics data into metabolic model

Finally, strategies need to be performed in order to efficiently integrate different levels of biological information. In our case, we have started this by focusing on integrating transcriptomics data into our model. Genome-scale metabolic models allow establishing integrative approaches to such include different data and infer novel conclusions for the preceding step. This hypothesis-driven method could be important when knowledge useful for metabolic engineering design can be retrieved from it.

Thesis approach

This thesis tries to bridge the Sciences of biology and computational sciences arriving to System Biology. In the beginning of the manuscript (**Chapter 1**) we have briefly outlined some of the basic concepts of System Biology and its importance. In addition, we have looked at cyanobacteria biotechnological applications as bio-products production platform. The following chapters encompass different consecutive aspects of this project.

In **Chapter 2** we have described the reconstruction process of a genome-scale metabolic model of *Synechococcus elongatus* PCC7942. The modeling process is explained in detail, two versions of our model with well-described biomass composition are presented and connectivity analyses are done.

Chapter 3 is devoted to the studies of the fluxomic data of *Synechococcus elongatus* PCC7942 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. Alternative optima solutions are calculated from flux variability analysis. Finally, metabolic network stability under certain perturbations is evaluated with robustness analysis.

Genetic perturbations are studied in **Chapter 4**, where essential genes are evaluated as well as mutations that lead *Synechococcus elongatus* PCC7942 to be a good production platform of value-added metabolites, such as ethanol, higher chain alcohols, lipids and hydrogen. Single, double and triple knockout strategies are studied and theoretical production limits are assessed in the light of these overproducing strains.

With the goal of studying the resulting optimal metabolic phenotypes, in **Chapter 5** we have analyzed the phenotypic phase planes generated for the growth and productions of the industrially-relevant bio-products mentioned above, varying carbon and energy sources such as CO_2 and light.

Integration of transcriptomics data on the metabolic network is the scope of **Chapter 6**. Metabolic-reactions connectivity is analyzed under CO_2 acclimation. Regulatory hubs upon CO_2 acclimation regime are identified and explained in a system-wide integrative manner.

Finally, **Chapter 7** gathers conclusions among all chapters and comments the planned workflow to analyze the potentialities of *Synechococcus elongatus* PCC7942 as production platform for bio-products.

Scientific contributions

• *Synechococcus elongatus* PCC7942 metabolic models

Triana J, Montagud A, Siurana M, Gamermann D, Torres J, Tena J, Fernández de Córdoba P, Urchueguía JF. **Generation and evaluation of a genome-scale metabolic network model of *Synechococcus elongatus* PCC7942**. *Submitted at Metabolites.*

Triana J, Montagud A, Gamermann D, Fernández de Córdoba P, Urchueguía JF. ***In silico* analysis for bio-products synthesis through genome-scale reconstruction of the *Synechococcus elongatus* PCC7942 metabolic network**. *Manuscript in preparation.*

• Systems Biology works

Reyes R, Gamermann D, Montagud A, Fuente D, **Triana J**, Urchueguía JF, Fernández de Córdoba P. (2012) **Automation on the generation of genome scale metabolic models**. *Journal of Computational Biology*, **19(12)**: 1295-1306.

Jaime-Infante RA, Hernández-Martínez Z, **Triana-Dopico J**, Fosado-Tellez O, Montagud-Aquino A, Gamermann D, Fernández de Córdoba-Castellá P, Urchueguía-Schölzel JF. (2014) **Herramienta para la optmización de flujos metabólicos en sistemas biológicos.** *Investigación Operacional*, **35(2)**: 96-103.

Gamermann D, Montagud A, Jaime Infante RA, **Triana J**, Fernández de Córdoba P, Urchieguía JF. **PyNetMet: Python tools for efficient work with networks and metabolic models**. *Computational and Mathematical Biology*, 3(5): 1-11.

Pacheco Y, Reyes R, **Triana J**, Gamermann D, Montagud A, Fernández de Córdoba P, Urchueguía JF. **Integrated database for metabolic models reconstruction using COPABI.** *Manuscript in preparation.*

• Synthetic Biology works

Gamermann D, Montagud A, Aparicio P, Navarro E, **Triana J**, Villatoro FR, Urchueguía JF, Fernández de Córdoba P. (2012) **A modular synthetic device to calibrate promoters**. *Journal of Biological Systems*, **20(1)**: 1-24.

• Book

Pacheco-Suárez Y, Reyes-Chirino R, **Triana-Dopico J**. (2012) **Servicio Web cliente orientado a la obtención de la información biológica disponible en la base de datos KEGG**. *Ed. Acad Mica spa Ola*. ISBN-10: 3848471051. ISBN-13: 9783848471058.

• Bachelor Thesis supervised

Pacheco-Suárez, Yarlenis. **Web service cliente orientado a la extracción de la información biológica disponible en la base de datos KEGG**. Universidad de Pinar del Río "Hermanos Saíz Montes de Oca", 2011.

Armenteros-Pérez, Javier. **Plataforma computacional para el acceso a la información biológica**. Universidad de Pinar del Río "Hermanos Saíz Montes de Oca", 2011.

Hernández-Martínez, Zenén. **Herramientas para el análisis *in silico* de la distribución de flujos metabólicos en sistemas biológicos**. Universidad de Pinar del Río "Hermanos Saíz Montes de Oca", 2012.

Morejón Guerra, Leslie Miguel. **Módulo para la creación de modelos metabólicos de sistemas biológicos (COPABI)**. Universidad de Pinar del Río "Hermanos Saíz Montes de Oca", 2012.

Suárez-Ordaz, Dariel. **Herramienta para el completamiento de Bases de Datos Biológicas**. Universidad de Pinar del Río "Hermanos Saíz Montes de Oca", 2012.

Soto-González, Jean Carlos. **Sistema de Escritorio para Simulaciones Dinámicas (dFBA) en Redes Metabólicas**. Universidad de Pinar del Río "Hermanos Saíz Montes de Oca", 2012.

Rodríguez-Romeu, Raidel. **Herramienta para la predicción del comportamiento metabólico en sistemas biológicos bajo perturbaciones**. Universidad de Pinar del Río "Hermanos Saíz Montes de Oca", 2013.

• Other biotechnological works

Amador-Cañizares Y, Alvarez-Lajonchere L, Guerra I, Rodríguez-Alonso I, Martínez-Donato G, **Triana J**, González-Horta EE, Pérez A, Dueñas-Carrera S. (2008) **Induction of IgA and sustained deficiency of cell proliferative response in chronic hepatitis C**. *World Journal of Gastroenterology*, **14(44)**: 6844-6852. doi:10.3748/wjg.14.6844.

No podrás nadar hacia nuevos horizontes sino tienes el valor de perder de vista la costa

William Faulkner, *(1897-1962)*

Chapter 1. Introduction

Systems Biology is the research area that studies life processes from a systemic approach. This field requires of the collaboration of researchers from diverse backgrounds, including biology, biochemistry, computer science, mathematics, statistics, physics and chemistry. These collaborations are indispensable because the large-scale knowledge integration required for understand a certain biological phenomenon. Here, the boundaries between the different disciplines disappear, creating a new science, a new universe of knowledge. This introduction and overview of system modeling in biology aims to build the ground that supports the core of this thesis.

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

Triana J, Montagud A, Siurana M, Gamermann D, Torres J, Tena J, Fernández de Córdoba P, Urchueguía JF. **Generation and evaluation of a genome-scale metabolic network model of *Synechococcus elongatus* PCC7942**. *Submitted at Metabolites*.

Triana J, Montagud A, Gamermann D, Fernández de Córdoba P, Urchueguía JF**. *In silico* analysis for bio-products synthesis through genome-scale reconstruction of the *Synechococcus elongatus* PCC7942 metabolic network**. In preparation.

Chapter 1. Introduction

1.1 Systems Biology approach

During the middle of the past century the reductionist approaches have influenced the biological sciences. These approaches focused on the generation of information about individual cellular components, their chemical composition, and often their biological functions (Palsson, 2006). With the spreading out of novel technologies, increasing biological complex datasets have been generated. Sequencing and genetic synthesis technologies in biological research have been exponentially enhanced (1.5-fold/year since 1960s, 6-fold per year since 2005) and have reached breakthroughs starting from the first sequenced genomes to the first genome-scale synthesis of several organisms (Carr and Church, 2009; Church, 2013). Thus, assorted *-omics* data are now available that enable us to leave the reductionist approaches and use integrative paradigms (see figure 1.1) (Palsson, 2006; Church, 2013). One of the new research fields that emerge from this panorama is Systems Biology (SB).

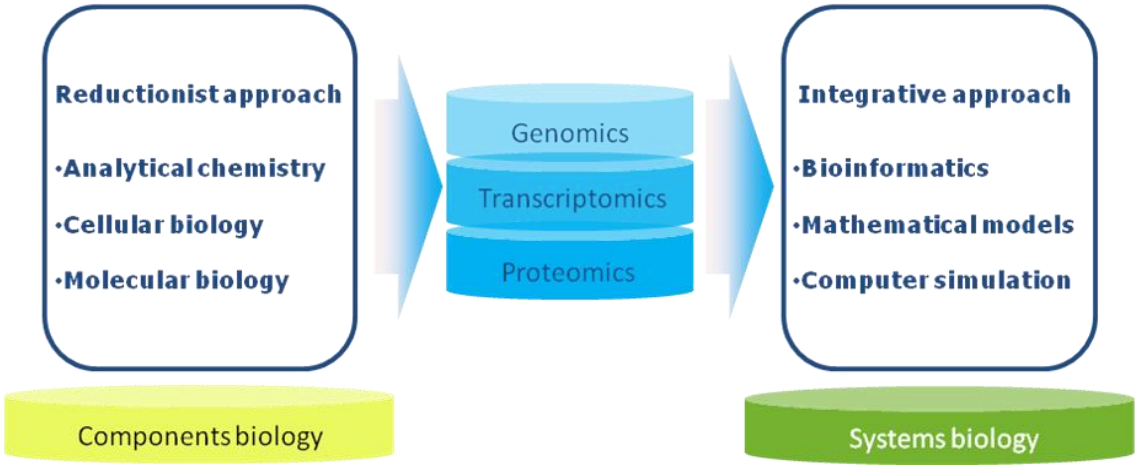


Figure 1.1. Biology approach evolution. From reductionist to integrative approach.

Several questions arise from the paradigm shift in cell and molecular biology due to the change of cell parts analysis to system analysis. Many questions protrude from the list, such as: What is SB?, What is attempted to achieve with the integrative approaches? or What are the basis of these types of analyzes?

SB is the bottom-up approach to quantitatively explain the properties of biological systems from the modeling and simulation of the interactions and characteristics of its macromolecular components (Snoep *et al.*, 2006). It is the interdisciplinary research of

biological processes in which the interactions of internal and external elements that influence the cell functionality, is represented by a mathematical system. Systems biology has two general aims: *"a narrow one, which is to discover how complex networks of proteins work, and a broader one, which is to integrate the molecular and network data with the generation and function of organism phenotypes"* (Bard, 2013).

The typical workflow implies a plurality of tasks and levels of information. This thorough process integrates the biological components that participate in the process, the study of its interactions and its reconstruction in a model, the analysis and mathematical depiction of the network model and its use as a basis for analyze, interpret and predict experimental outcomes (see figure 1.2). Prediction here means generating specific hypotheses that can then be experimentally tested in order to gain higher insight into the biological entities. These *in silico* models of reconstructed networks are then improved in an iterative fashion (Palsson, 2006).

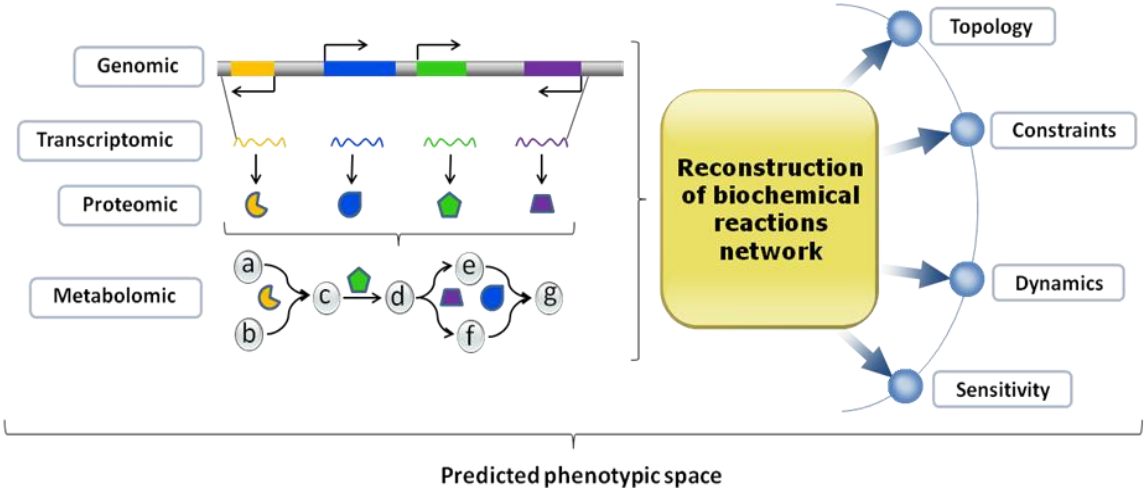


Figure 1.2. The plurality of information (high-throughput data) needed for model reconstruction and *in silico* modeling methods implemented in Systems Biology.

1.2 The genome-scale metabolic network model

As we mentioned above, reductionist viewpoints cannot, by definition, provide a coherent understanding of whole cell functions. That´s why the modeling of whole biological systems, as a top-down approach, has received increasing attention. Genome-scale metabolic models are examples of modeling approaches that have been developed to predict systems-level phenotypes, and which have had success in recent years (O'Brien *et al.*, 2013). Many of them have been extended to include different

kinds of information, for example gene product expression levels (Montagud *et al.*, 2010; O'Brien *et al.*, 2013), as well as to integrate many more cellular processes and can be used to simulate dynamic cellular states, such as the whole-cell model of the human pathogen *Mycoplasma genitalium* (Karr *et al.*, 2012). Thus, metabolic models have been useful for generating new hypotheses and targeting promising areas for biotechnology (Esvelt and Wang, 2013).

Mathematical metabolic models accounting for genome-scale information have been reconstructed for nearly twenty organisms (and numbers and rising each month), mostly through thorough curation efforts. Reconstructed *in silico* organisms are computer representations of their *in vivo* counterparts (Palsson, 2006). However, because biological systems are inherently nonlinear and seldom show emergent properties, "the whole is more than the sum of the components" (Szallasi *et al*., 2006), the construction of the metabolic networks is not only a compilation of chemical reactions but also gathering of constraints, evidences and such that will make up the basis for *in silico* analyses of the organism´s behavior. The reconstruction of these models is based on genetic information available in the genomes of organisms. Furthermore, in order to build a meaningful model, researcher needs experimental data together with established knowledge, such as physiological and biochemical information that is accessible from literature, journal articles, experiments and databases. At present, the majority of the organisms lack enough data to support this process. Therefore, the *in silico* analyses of these reconstructions could lead to failed predictions and model updates will take place successively. Thus, the process of building mathematical models and running computer simulations of complex biological processes is iterative (see figure 1.3). On the other hand, the quality of the metabolic models depends on the accuracy of the information. In certain instances, lack of clearness and quality turns out to be a problem that undermines the faithfulness of the reconstructed models, especially for erroneous entries and false negatives and false positives (Weise *et al*., 2006). The relationships between complex metabolic processes usually falls to properly determine the processing of substrates into products and their stoichiometry, if this transformation is spontaneous or catalyzed by enzymes or if cofactors are involved. Additionally, the sub-cellular localization of the reaction and

some thermodynamic aspects such as irreversibility must be known (Förster *et al*. 2003).

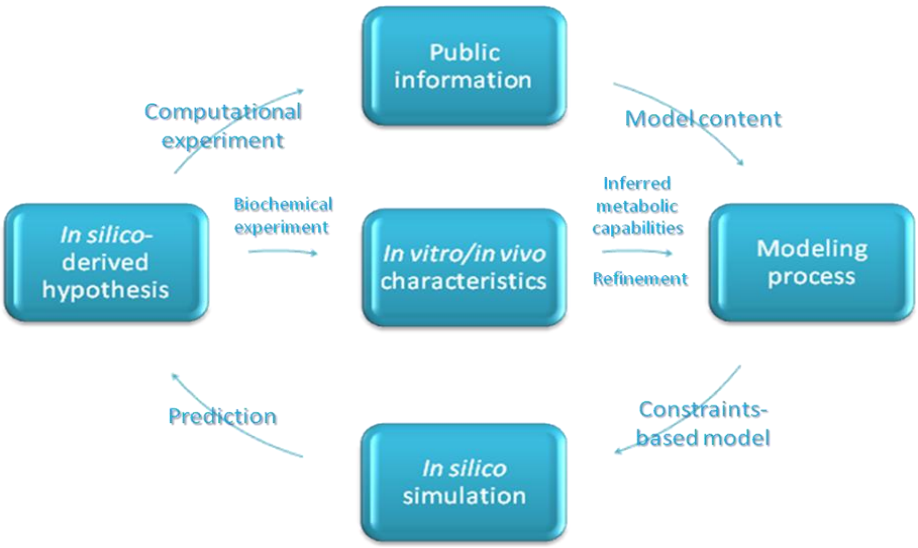


Figure 1.3. The iterative workflows for *in silico* metabolic model reconstruction. Adapted from (Palsson, 2000).

Several protocols have been published to define in detail each one of the steps that should lead to a proper reconstruction, as well as the software packages and databases that can assist in this labor (Förster *et al*. 2003; Notebaart *et al*., 2006; Feist *et al*., 2009; Thiele and Palsson, 2010). Presently, the process of reconstruction is long and arduous mainly due to its manual construction and for quality control checks (Thiele and Palsson, 2010). Some reports assert that a genome-scale metabolic network reconstruction can easily take from several months up to 2 years (Förster *et al*. 2003; Duarte *et al*., 2007). Furthermore, some works have attempted to automate the metabolic reconstruction, or at least some parts of it, in order to cut down the time needed for such a project, such as the COPABI project (Reyes *et al.*, 2012). However, these efforts have been hampered with the current problems in databases information and genome annotations (Feist *et al*., 2009). Thereby, resulting algorithms still need the supervision of experts in order to be able to generate quality metabolic networks models as a basis for predictive analysis (Thiele and Palsson, 2010).

Some authors declare that the genome-scale constraint-based metabolic models are a natural continuation of genome annotation. Others suggest that does not systematic metabolic model reconstruction pipeline exists yet.

1.3 Metabolic network analysis

1.3.1 Metabolic regulation and control

Work from Voet and Voet in 2004 is a very nice way of seeing metabolism from a system-wide perspective. They brought together thermodynamics, biology and system biology in order to understand metabolic behavior. They explained that living organisms are thermodynamically open systems that tend to maintain a steady-state rather that reaching equilibrium. Equilibrium that would only happens in the case of death for living things. Thus, the flux (rate of flow) of intermediates through a metabolic pathway is constant; that is, the rates of synthesis and breakdown of each pathway intermediate maintain a constant concentration. Such a state will most probably be one of maximum thermodynamic efficiency. Therefore, these authors state that, regulation of the steady-state (its homeostasis) must be maintained when the flux changes through the pathway in response to changes in demand (Voet and Voet, 2004).

Thus, we have here a steady-state and the regulation of its elements in order to be able to adapt this steady state to whichever perturbation that may happen. In this sense, the concepts of metabolic control and metabolic regulation are usually confused. Usually, metabolic regulation is defined as a process by which the steady-state flow of metabolites through a pathway is maintained, whereas metabolic control is the influence exerted on the enzymes of a pathway in response to an external signal in order to alter the flux of metabolites (Crabtree and Newsholme, 1987; Kacser and Burns, 1995).

As Voet and Voet continue, there are two principal reasons why metabolic flux must be controlled: one is to provide products at the rate they are needed, that is, to balance supply with demand; and the second is to maintain the steady-state consentrations of the intermediates in a pathway within a narrow flux range (homeostasis) (Voet and Voet, 2004).

According to these authors, organisms maintain homeostasis for several reasons:

1. In an open system, such as metabolism, the steady-state is the state of maximum thermodynamic efficiency.

2. Many intermediates participate in more than one pathway, so that changing their concentrations may disturb a delicate balance.
3. The rate at which a pathway can respond to a control signal slows if large changes in intermediate concentrations are involved.
4. Large changes in intermediate concentrations may have deleterious effects on cellular osmotic properties. (Voet and Voet, 2004)

The level of metabolic flux, and hence, the concentrations of intermediates at which a pathway is maintained, vary with the necessities of the organism through a highly responsive system of precise controls. Such pathway are analogous to rivers that have been dammed to provide a means of generating electricity, in the words of Voet and Voet. Although water is continually flowing in and out of the lake formed by the dam, a relatively constant water level is maintained. The rate of water outflow from the lake is precisely controlled at the dam and is varied in response to the need for electrical power (Voet and Voet, 2004).

1.3.2 Metabolic flux

A metabolic pathway is a sequence of enzyme-catalyzed reactions. To define the flux through the pathway researchers have to consider each one of its reaction steps. The flux of metabolites, *J*, through each reaction step is the rate of the forward reaction, v_f, less that of the reverse reaction, v_r (Voet and Voet, 2004):

$$J = v_f - v_r$$

At equilibrium, by definition, there is no flux ($J = 0$), although v_f and v_r may be quite large. At the other extreme lie reactions that are far from equilibrium, $v_f >> v_r$, so that the flux is essentially equal to the rate of the forward reaction, $J \approx v_f$. The flux through a steady-state pathway is constant and is set (generated) by the pathway's rate-determining step (or steps). Consequently, control of flux through a metabolic pathway requires: (1) that the flux through this flux-generating step varies in response to the organism's metabolic requirements and (2) that this change in flux be communicated throughout the pathway to maintain a steady-state (Kacser and Burns, 1995; Fell, 1997; Voet and Voet, 2004).

According to the points made by the authors Voet and Voet, it has been seen that the fractional change in flux through a metabolic pathaway´s rate-determining step(s)

to the fractional change in subtrate concentration necessary to communicate that change to the following reaction steps is governed by (Voet and Voet, 2004):

$$\frac{\Delta J}{J} = \frac{\Delta[S]}{[S]} \quad \frac{v_f}{v_f - v_r}$$

where $\Delta J/J$ is the fractional change in flux through the rate-determinig step(s), S is the product of the rate-determinig step(s) and $\Delta[S]/S$ is the fractional change in v_f ($\Delta v_f/v_f$), assuming the simplest and most common situation of $[S] << K_M$ in Michaelis-Menten equation.

Following these authors' reasoning, this quantity is a measure of the sensitivity of a reaction´s fractional change in flux to its fractional change in substrate concentration. This quantity is also a measure of reactions reversibility, that is, how close it is to equilibrium:

a) In an irreversible reaction, v_r approaches 0 (relative to v_f) and therefore $v_f/(v_f - v_r)$ approaches 1. The reaction therefore requires a nearly equal fractional increase in its substrate concentration in order to respond to a fractional increase in fluxes.
b) As a reaction approaches equilibrium, v_r approaches v_f and hence $v_f/(v_f - v_r)$ approaches infinity. The reaction´s response toa fractional increase in flux therefore requires a much smaller fractonal increase in its substrate concentration.

Consequently, the ability of a reaction to communicate a change in flux increases as the reaction approaches equilibrium. A series of sequential reactions that are all near equilibirum therefore have the same flux and maintain concentrations of intermediates in a steady-state (homeostasis) (Voet and Voet, 2004).

A typical feature of metabolic control exerted by the cell is the control of the rate of the enzyme that is responsible of the rate-limiting step of every metabolic pathway.These enzymes are the so-called regulatory enzymes and most of them are allosterics. Thus, they are subject to feedback inhibition and are often also controlled by covalent modification (Fell, 1997; Voet and Voet, 2004).

Voet and Voet rise several questions regarding this control: Are these regulatory enzymes really rate limiting for the pathway? Is there really only one step in the pathway that is rate limiting, or migth there be a number of enzymes contributing to

the regulation of the pathway? Does controlling these enzymes really control the flux of metabolites through the pathway or is the function of feedback inhibition really to maintain a steady-state? These are complicated questions with complicated answers, and these answers will not be the main scope of present thesis (Voet and Voet, 2004).

1.3.3 Constraint-based computer simulation

1.3.3.1 Constraints on cellular functions

The evolution process is intrinsically associated to biological sciences. Biological systems have adapted to diverse environments over the evolutive history, and as they replicate, they produce descendants that are not genetically identical to the parent, consequently generating a population with slight differences between individuals. As Palsson wrote it in 2006, "To survive in a given environment, organisms must satisfy myriad constraints, which limit the range of available phenotypes. All expressed phenotypes resulting from the selection process must satisfy the governing constraints. Therefore, clear identification and statement of constraints to define ranges of allowable phenotypic states provides a fundamental approach to understanding biological systems that is consistent with our understanding of the way in which organism operate and evolve" (Palsson, 2006).

Different types of constraints limit cellular functions and will have to have a presence in our models. There are both nonadjustable and adjustable constraints. The former can be used to restrict the range of possible behaviors, for example, the physic-chemical factors, the connectivity or the capacity characterized by maximum and minimum flux range; and the later can be used to further limit allowable behaviors, such as: the reaction rates or adjustable capacity factor by transcriptional regulation. One has to bear in mind that these constraints can be adjusted through an evolutionary process or through changing environmental conditions. In addition, the adjustable constraints may vary in a bio-population from one individual to another. There are diverse forms to classify these restrictions, and many authors have discussed them from different points of view. One of these authors, Price *et al.* mentioned four constraints categories (Price *et al.*, 2004b), which are: fundamental physico-chemical

constraints, spatial or topological constraints, condition-dependent environmental constraints and regulatory or self-imposed constraints.

a) Physico-chemical constraints: Many of these constraints govern cellular processes, and these constraints are inviolable and provide "hard" constraints on cell functions. Conservation of mass, energy and momentum represent hard constraints. The interior of a cell are densely packed, forming an environment where the viscosity may be on the order of 100-1000 times that of water. Diffusion rates inside a cell may be slow, especially for macromolecules (Elowitz *et al.*, 1999). The confinement of a large number of molecules within a semi-permeable membrane causes high osmolarity. Therefore, cells require mechanisms for dealing with osmotic pressure generated, such as sodium-potassium pumps or a cell wall (Werner and Heinrich, 1985; Lew and Bookchin, 1986). Intracellular reaction rates are determined by local concentrations inside cells and might be limited by mass-transport. Reactions have maximal reaction rates (denoted with v_{max}) estimated to be about a million molecules per μm^3 per second. Furthermore, biochemical reactions must result in a negative free-energy change to proceed in the forward direction.

b) Topological constraints: The crowding of molecules inside cells leads to topological, or three-dimensional, constraints. The linear dimension of the bacterial genome is on the order of 1000 times that of the length of the cell. DNA must therefore be tightly packed in the nuclear region in an accessible and functional configuration since DNA is only functional if it is accessible. Thus, at least two competing needs (to be tightly packed but easily accessible) constraint the physical arrangement of genome.

c) Environmental constraints: These constraints on cells, such as nutrient availability, pH, temperature, osmolarity, the availability of electron acceptors, etc., are typically time and condition-dependent. Environmental constraints are of fundamental importance for the quantitative analysis of microorganisms functions; however, natural environment may be hard to define precisely. Defined media and well-documented environmental conditions are needed to integrate data from various laboratories into quantitative models that are both

accurately descriptive and predictive. Laboratory experiments with undefined media composition are often of limited use for quantitative *in silico* modeling.

d) Regulatory constraints: These constraints are fundamentally different than the three categories discussed above. They are self-imposed and are subject to evolutionary change, and can thus vary in time. For this reason, these constraints may be referred to as regulatory restraints, in contrast to physico-chemical constraints, the topological constraints, and time-dependent environmental constraints. Based on environmental conditions, regulatory constraints provide a mechanism to eliminate suboptimal phenotypic states and confine cellular functions to behaviors of increased fitness. Regulatory constraints are implemented by the cell in a variety of ways, including the amount of gene products made (transcriptional and translational regulation) and their activity (enzyme regulation).

Constraints can be applied to the analysis of reconstructed networks to narrow achievable behaviors, and can be applied in a successive manner. The imposition of a constraint leads to solutions spaces rather than the computation of a single solution, the hallmark of theory-based models. Cellular behaviors (*i.e.*, functional states) can be attained within this solution space. Each allowable behavior basically represents a different phenotype based on the component list, the biochemical properties of the components, their interconnectivity, and the imposed constraints. The constraint-based approach leads to *in silico* analysis procedures that are helpful in analyzing, interpreting, and occasionally predicting the genotype-phenotype relationship.

1.3.3.2 Methods for analyzing metabolic network states

There are three types of metabolic models (Baart and Martens, 2012):

a) Steady-state models, which only take into account the flux of metabolites through the system in steady-state.

b) Steady-state kinetic models, which take into account the flux of metabolites through the system in steady-state and contain at least one kinetic equation that relates the concentration of a metabolite to a reaction rate.

c) Dynamic kinetic models, which take into account the kinetics of all different enzymes involved in the reaction network.

In an ideal situation, the model contains different levels of information, including reaction stoichiometry, reaction kinetics and regulatory information. However, the required kinetic and regulatory information is very scarce, despite there are several studies in this direction (Tomita, 2001; Covert *et al.*, 2004; Bruggeman and Westerhoff, 2006; Bruggeman *et al.*, 2008; De Mey *et al.*, 2010). For this reason researchers generally use a (pseudo) steady-state approximation, using the reaction stoichiometry in combination with mass balancing of the fluxes (Taymaz-Nikerel *et al.*, 2010). The metabolic capabilities of the constructed network may be calculated using constraint-based computer simulation methods. Towards this, a variety of tools/algorithms is available and can be applied on metabolic models. Among them, we can find *Flux Balance Analysis* (FBA) (Varma and Palsson, 1993; Edwards *et al.*, 1999), *Metabolic Flux Analysis* (MFA) (Schilling *et al.*, 1999; Varma and Palsson, 1994c), *Elementary Flux Modes* (EFM) (Schuster *et al.*, 1999), *Extreme Pathways* (ExPa) (Schilling *et al.*, 1999), *Robustness Analysis* (RNA) (Edwards and Palsson, 2000), *Phenotypic Phase Plane analysis* (PhPP) (Edwards *et al*., 2002), *Minimization of Metabolic Adjustment* (MOMA) (Segrè *et al.*, 2002), *Flux Variability Analysis* (FVA) (Mahadevan and Schilling, 2003) and *Regulatory On-Off Minimization* (ROOM) (Shlomi *et al.*, 2005).

These methods, among others, can be largely classified into categories according to its purposes (see figure 1.4), such as: finding best or optimal states in the allowable range; investigating flux dependencies; studying state all allowable states; altering possible phenotypes as a consequence of genetic variations; and defining and imposing further constraints (Price *et al.*, 2004b).

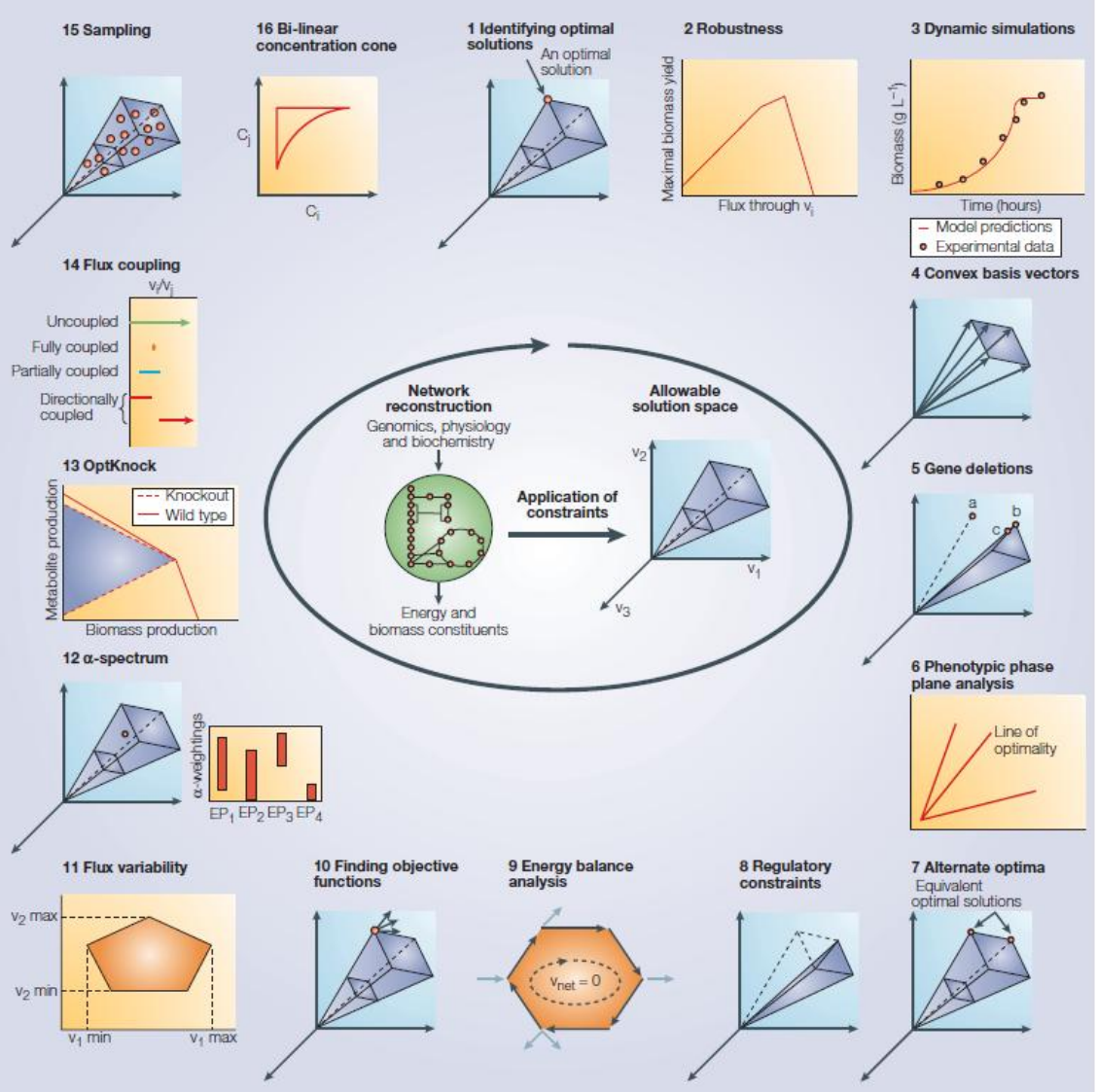


Figure 1.4. A summary of system biology´s tools for constraint-based analysis. As indicated, several methods are being developed at various laboratories to analyze the solution space. Taken from (Price *et al.*, 2004b).

1.3.3.3 Finding optimal states

Each one of the metabolic reactions that take place into the cell can be represented as chemical equations with the stoichiometry associated to each species. All the stoichiometric information can be represented in a matrix array: the stoichiometric matrix. Indirectly, additional information could be associated with this matrix such as: enzyme complex formation, transcriptomics data, open reading frames, etc. "Therefore, once reconstructed, the stoichiometric matrix represents a biochemically, genetically, and genomically structured database. This matrix is formed from the stoichiometric coefficients of the chemical equations derived from reaction network. It

is organized such that every column corresponds to a reaction and every row corresponds to a metabolite; and the entries in the matrix are stoichiometric coefficients. Each column that describes a reaction is constrained by the rules of chemistry, such as elemental balancing. Every row thus describes the reactions in which that compound participates and therefore how the reactions are interconnected" (Varma and Palsson, 1994a; Palsson, 2006).

Mathematically, the stoichiometric matrix is a linear transformation of the metabolic flux vector to a vector of time derivates of the concentration vector. The vector produced by a linear transformation is in four fundamental subspaces. The resulting vector is in two orthogonal subspaces: the column and left null space; and the vector being mapped is also in two orthogonal subspaces: the row and null spaces (Palsson, 2006).

For example, these vector subspaces are distinguished each other by (Palsson, 2006):

a) *Null space*. The null space of *S* contains all the steady-state flux distributions allowable in the network. The steady state is of much interest since most homeostatic states are close to being steady states.
b) *Row space*. The row space of *S* contains all the dynamic flux distributions of a network and thus the thermodynamic driving forces that change the rate of reaction activity.
c) *Left null space*. The left null space of *S* contains all the conservation relationships, or *time invariants*, that a network contains. The sums of conserved metabolites or conserved metabolic pools do not change with time and are combinations of concentration variables.
d) *Column space*. The column space of *S* contains all the possible time derivatives of the concentration vector and thus shows how the thermodynamic driving forces move the concentration state of the network.

Since there are typically only two, three, or four compounds that participate in a reaction out of hundreds of compounds participating in a network, the stoichiometric matrix is sparse, and thus, mostly comprises zero elements (Varma and Palsson, 1994a; Palsson, 2006). In biological networks the number of reactions is typically greater than

the number of metabolites. This results in a plurality of feasible steady-state flux distribution. Although infinite in number, the steady-state solutions lie in a restricted region, the so-called null space of the stoichiometric matrix. The null space can be used to define the range of all allowable phenotypes of a given network, since it specifies all the steady-state flux distributions that it can embed (Varma and Palsson, 1994a; Palsson, 2006). Conversely, due to the fact that only a particular set of phenotypes are expressed under particular conditions, the optimization techniques can be used to find particular solutions that maximize or minimize a particular biological objective. The desired objective is described mathematically, which takes the form of an objective function (that we will call Z).

The mathematical representation of the objective functions enables the formulation of a range of functionalities and network states of interest. The definition of these objectives can be used to assess the metabolic potentialities of the genome-scale metabolic network (Price *et al.*, 2002; Papin *et al.*, 2002), to represent exploration of physiologically meaningful objectives (Edwards *et al.*, 2001; Ibarra *et al.*, 2002), or design metabolic engineering strategies towards an *in silico* strains creation to improve production of a desired bio-product (Varma and Palsson, 1994b; Liao *et al.*, 1996; Burgard *et al.*, 2003). Some of them are displayed in figure 1.5.

The objective function can be either a linear or nonlinear function. Generally, these functions together with the imposed constraints are linear and thus, linear optimization or linear programming (LP) can be used to solve the optimization problem (Bonarius *et al.*, 1997; Kauffman *et al.*, 2003). In 1992, Savinell and Palsson developed and published an approach that has been adapted to LP: the Flux Balance Analysis (FBA) (Savinell and Palsson, 1992a, 1992b). Since then, FBA is a widely used tool for studying biochemical networks, in particular genome-scale metabolic network reconstructions (Orth *et al.*, 2010).

In detail, FBA assumes that metabolic networks will reach a steady-state constrained by the stoichiometry. The stoichiometric constraints lead to an underdetermined system; however, a bounded solution space of all feasible fluxes can be identified (Schilling *et al.*, 2000a).

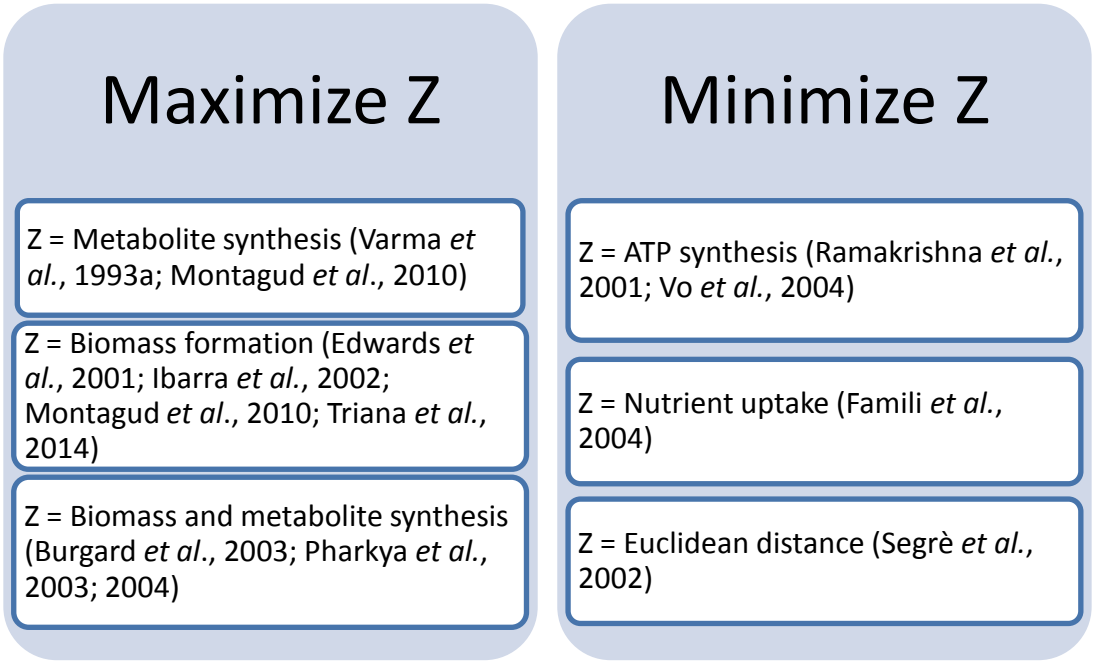


Figure 1.5. Some examples of biological objective functions that are widely used to perform *in silico* simulations.

This solution space can be further restricted by specifying maximum and minimum fluxes through any particular reaction and by specifying other constraints (see figure 1.6). Thus, obtaining these constraints gives us the performance capability of the metabolic network, and the constraints can be refined by adding experimental data (Förster *et al.*, 2003).

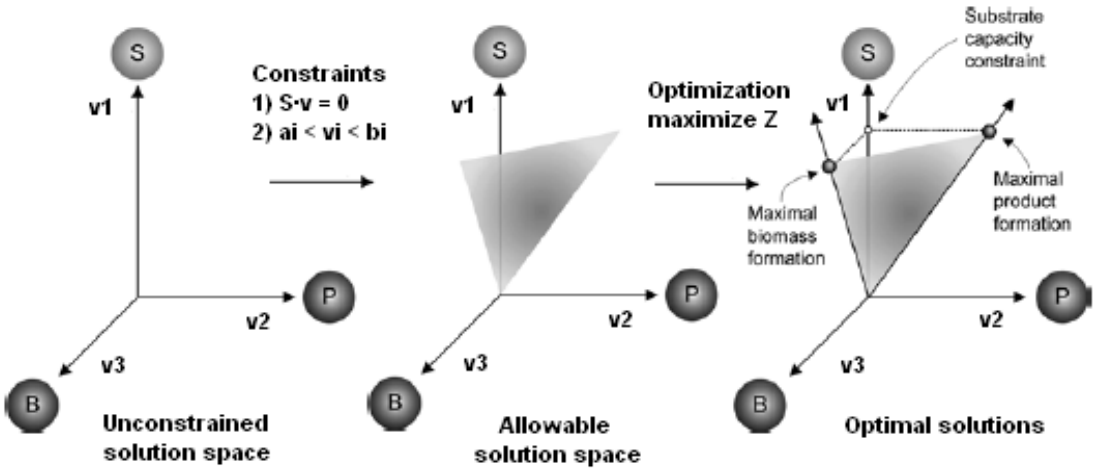


Figure 1.6. Graphical scheme of constraint-based modeling. With no constraints, the flux distribution of a biological network may lie at any point in a solution space. When mass balance constraints imposed by the stoichiometric matrix (S) and some of substrate conversion rates are known from measurements, a solution that optimizes a particular objective, like maximize biomass formation or maximize product formation can be calculated using linear programming. Adapted from (Orth *et al.*, 2010) and (Baart and Martens, 2012).

Once the solution space describing the capability of the organism is defined, the network's behavior can be studied by optimizing the steady-state behavior with respect to some objective function. The simulation results can then be experimentally verified and used to further strengthen the model. Ultimately, the iterative model refinement procedure can result in predictive models of cellular metabolism.

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) (http://www.sysbio.se/BioMet). Other software are COBRA Toolbox (Becker *et al.*, 2007), the latest version of Pathway Tools (Latendresse *et al.*, 2012) and some freely packages available for academic research such as PyNetMet (Gamermann *et al*., 2014), CellNetAnalyzer (Klamt *et al.*, 2007) and FBA-SimVis (Grafahrend-Belau *et al.*, 2009), that allows the solving of linear programming problems such as Flux Balance Analysis.

Several reports show that FBA predictions, with maximum growth rate as the objective function, are consistent with experimental flux data, thus, the assumption of optimality for a wild-type bacterium is justifiable. However, the same argument may not be valid for genetically engineered knockouts or other bacterial strains that were not exposed to long-term evolutionary pressure. Segrè *et al.* addressed this point by introducing the method of minimization of metabolic adjustment (MOMA) improving the prediction efficiency of FBA for studying biological system mutants (Segrè *et al.*, 2002). These authors tested the hypothesis that by knocking out metabolic fluxes cells undergo a minimal redistribution with respect to the flux configuration of the wild type.

MOMA and FBA are constraint-based analysis methods based on the same stoichiometric constraints. But, in MOMA the optimal growth flux for mutants is relaxed. "Instead, for perturbations such as gene deletions, MOMA approximates metabolic phenotype by performing distance minimization in flux space" (Segrè *et al.*, 2002). Thus, we can go further into the study of the genetic variations that can perturb the metabolic network.

Unlike FBA, the mathematical formulation in MOMA yields a quadratic programming problem. In MOMA we search for a point in the feasible space of the mutant strain that has minimal distance from a given flux vector of the feasible space of the wild-type strain (see figure 1.7) (Segrè *et al.*, 2002). The goal is to find the vector that belonging to the feasible space for the mutant strain such that the Euclidean distance:

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

is minimized.

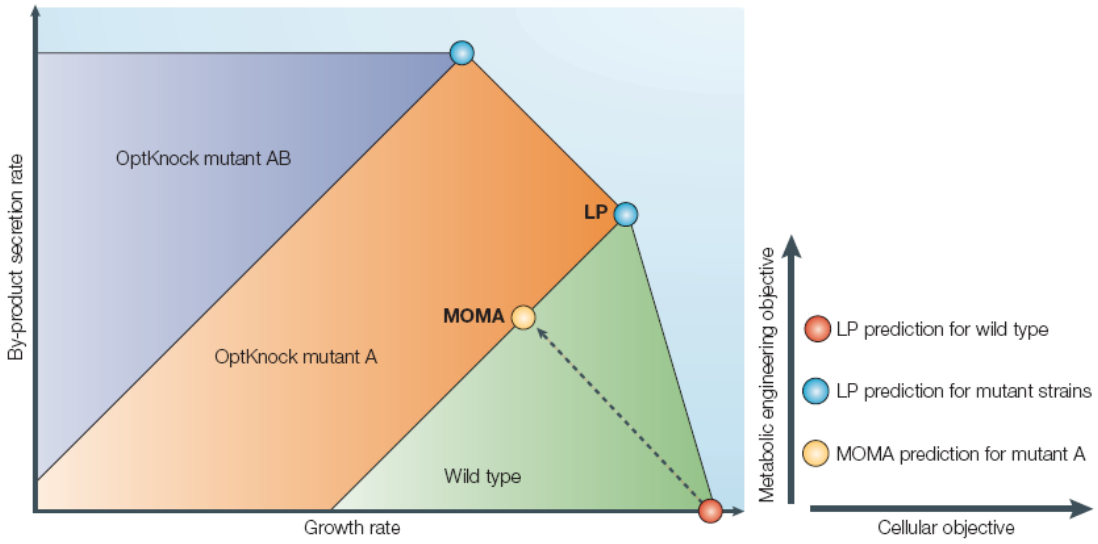


Figure 1.7. Altered solution spaces**.** Solution spaces are altered by perturbations in the underlying biochemical network, such as occur with gene deletions. The alternative MOMA knockout solution, calculated through quadratic programming, can be thought of as a projection of the FBA optimum onto the feasible space of the mutant. The mutant FBA optimum and the corresponding MOMA solution are, in general, distinct. Taken from (Price *et al.*, 2004b).

1.4 *In silico* guided metabolic engineering for bio-products synthesis from CO_2 and photons

Molecular biology methods have allowed the rational perturbation of cells. "Historically, these methods have been limited to gene insertions or mutations at random or at a few pre-defined locations across the genome. The handful of methods capable of targeted gene editing suffered from low efficiencies, significant labor costs, or both. Recent advances have dramatically expanded our ability to engineer cells in a directed and combinatorial manner" (Esvelt and Wang, 2013).

Ducat *et al.* in a special issue of Applied Microbiology journal commented: "Laboratory-scale systems for bioconversion of CO_2 and light to bio-products have been investigated since the biotechnology generation of valuable metabolites has been widely used. Most bioindustrial processes rely on microorganisms metabolizing carbon compounds to generate a diverse array of valuable chemicals, such as amino acids, vitamins and organic acids. In such production schemes, the cost of this compounds feedstock, such as carbohydrate, is a significant fraction of the total production cost. The use of photosynthetic organisms offers an alternative production approach, in which carbohydrate feedstock costs would be eliminated. Moreover, with the advent of global warming, there is growing interest in processes that couple CO_2 capture to chemical synthesis through the use of photosynthetic microorganisms" (Ducat *et al.,* 2011). The need is bringing forth great creativity in uncovering new candidate bio-molecules that can be made via metabolic engineering (Wackett, 2011).

Even though several microorganisms have been employed for the biotechnological synthesis of value-added compounds, the productive potential of cyanobacterial group has remained largely unexplored. Cyanobacteria are the only known prokaryotes with the capability to perform oxygenic photosynthesis and are attracting increasing attention as suitable host organisms for the production of bio-products. Cyanobacteria species have many advantages as hosts for biotechnological practices, such as: rapid genetics, simple input requirements, carbon sequestration, and tolerance of marginal agricultural environments, among others. Ducat *et al.,* in 2011 reviewed the recent research involving the engineering of cyanobacterial species for the production of valuable bioindustrial compounds, including natural cyanobacterial products (e.g. sugars and isoprene), biofuels (e.g. alcohols, alkanes and hydrogen), and other commodity chemicals. These authors highlighted biological and economic obstacles to scale cyanobacterial production and discussed the methods for increasing cyanobacterial production efficiencies (Ducat *et al.,* 2011). While some strategies for improving photochemical efficiencies within the carbon fixation machinery might be possible, many are impacted by insufficient knowledge of cyanobacterial biology. Cyanobacteria are under-represented in the scientific literature relative to other microbes, although information regarding these species has been steadily increasing in

recent years (Burja *et al.*, 2003). "It is particularly important to recognize that cyanobacteria are not simply single-celled plants; their metabolism also differs in fundamental ways. High-throughput analysis and systems biology approaches, such as metabolomics, might help close the gap in the understanding of cyanobacterial biochemistry" (Ducat *et al.,* 2011).

One of the attractive bio-products that have been targeted in cyanobacteria is biofuels. This is because the necessity to develop and improve sustainable energy resources due to the finite nature of fossil fuels (Quintana *et al*., 2011). Nowadays, there are many factors that suggest that the biofuels landscape is disorganized. It is controlled by the rules imposed by economic forces and driven by the necessity of finding new sources of energy, particularly for motor fuels. "These next generation fuels include long-chain alcohols, terpenoid hydrocarbons, and diesel-length alkanes. Renewable fuels contain carbon derived from carbon dioxide. The carbon dioxide is derived directly by photosynthetic fuel-producing organisms or via intermediary biomass polymers that were previously derived from carbon dioxide" (Wackett, 2011). Computational modeling improvements and modular construction approaches could be the key that support the analysis and design of novel biofuels. As well as, when metabolic engineering is an alternative to take into account, the efficiency associated to this alternative is an important key for make it relatively inexpensive. "For example, novel metabolic networks have been constructed to make long-chain alcohols and hydrocarbons that have superior fuel properties and are cheaper over the current alcohols. A particularly exciting approach is to implement a direct utilization of solar energy to make a usable fuel. A number of approaches use the components of current biological systems, but re-engineer them for more direct, efficient production of fuels" (Wackett, 2011).

1.4.1 Cyanobacterium model as a potential platform for metabolic engineering

Cyanobacteria are a large group of photosynthetic oxygen-evolving prokaryotes of a great breadth of morphologies and ecologies. They originated during the Precambrian era (2.8×10^9 years ago) and since then play key roles in global carbon and nitrogen cycles (Shih *et al.,* 2013). These microorganisms can be found in almost every

terrestrial and aquatic habitat: in oceans, freshwater and even in bare rock and soil. They can occur as planktonic cells or form phototrophic biofilms in fresh water and marine environments, they occur in damp soil, or even on temporarily moistened rocks in deserts. Cyanobacteria account for 20-30% of Earth's photosynthetic productivity and convert solar energy into biomass-stored chemical energy at the rate of ~450 TW (Pisciotta *et al.,* 2010). Furthermore, some of these organisms are a significant factor in the oxygen cycle and then in global ecology.

The genus *Synechococcus* encompasses cyanobacteria that have a broader distribution in freshwater, marine environments and are less abundant in oligotrophic (low nutrient) regions. This genus is among the most important photosynthetic bacteria in the marine environment, estimated to account for about 25% of the primary production that occurs in typical marine habitats, making them one of the most significant photosynthetic bacteria (Scanlan and Nyree, 2002). *Synechococcus elongatus* (previously known as *Anacystis nidulans*) is specie into this genus that has a rod-shaped appearance and is oligotrophic, having the ability to survive in freshwater environments with low nutrients. Living habitats include freshwater hot springs and other freshwater habitats preferably with a mesophilic or moderate temperature range (Waterbury *et al.*, 1986). Two genomic strains of this cyanobacterium have been sequenced and are closely related to each other: *Synechococcus elongatus* PCC6301 (Sugita *et al*., 2007) and *Synechococcus elongatus* PCC7942 (van den Hondel *et al*., 1980; Van der Plas *et al*., 1992; Chen *et al*., 2008).

Synechococcus elongatus PCC7942 represent a model organism for studying several behaviors, such as circadian rhythms because it has an “endogenous timing mechanism” by which it can create and maintain a 24 hour clock period (Andersson *et al*., 2000; Kondo *et al*., 1993). Because, it was the first cyanobacterium to be cloned by DNA added exogenously (Shestakov and Khyen, 1970), and due to their highly versatile metabolism and their ability to directly convert solar energy into hydrocarbons, this microorganism has been investigated for biotechnological applications. Basically, it uses the sun’s energy, H_2O and CO_2 (together with some minerals) to synthesize its storage components, *i.e.* carbohydrates, lipids and proteins. These storage components form a potential feedstock which can be converted into bio-products.

Thus, these microorganisms possess special features which make them a promising model to transform all these carbon sources into valuable substances (Quintana *et al*., 2011). In spite of this, its domestication is at the forefront of current global challenges, such as the efficient supply of carbon skeletons from non-fossil resources and the efficient sequestration of atmospheric CO_2.

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 some products studied in chapter 4, and is one of the motivations of this thesis. Mathematical modeling of *Synechococcus elongatus* PCC7942 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.

Chapter 2. Reconstruction of cyanobacterium genome-scale metabolic network

The reconstruction of genome-scale metabolic models and their validation represent a great challenge in systems biology. This challenge allows researchers to use these as stoichiometric simulation models, phenotypic functions of interest can be predicted. Here, we present a comprehensive integration of different types of information into a cyanobacterium metabolic network model, as well as the study of its topological properties.

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

Triana J, Montagud A, Siurana M, Gamermann D, Torres J, Tena J, Fernández de Córdoba P, Urchueguía JF. **Generation and evaluation of a genome-scale metabolic network model of *Synechococcus elongatus* PCC7942**. *Submitted at Metabolites*.

Triana J, Montagud A, Gamermann D, Fernández de Córdoba P, Urchueguía JF. ***In silico* analysis for bio-products synthesis through genome-scale reconstruction of the *Synechococcus elongatus* PCC7942 metabolic network**. *In preparation.*

Reyes R, Gamermann D, Montagud A, Fuente D, Triana J, Urchueguía JF, Fernández de Córdoba P. (2012) **Automation on the generation of genome scale metabolic models**. *Journal of Computational Biology*, **19(12)**: 1295-1306.

Chapter 2. Reconstruction of cyanobacterium genome-scale metabolic network

2.1 Introduction

The development of genomic sequencing and genetic mapping together with omics-science opened up the way towards the quantitative study of biological systems. In this panorama, systems biology has emerged as a promising predictive science on a large and quantitatively deep scale (Snoep *et al*. 2006). In this field, the metabolic pathways and its capabilities constitute the central insight (Papin *et al*., 2003). The reconstruction of the metabolic genotype into an *in silico* model is increasingly considered an important technology for industrial applications. Optimal characteristics within the metabolic network capabilities under physiological and genetic conditions, such as different growth media and gene knockout candidates can be identified (Edwards *et al*., 2001). These analyses have been used by researchers to design metabolic engineering strategies in a variety of problems (Montagud *et al*., 2010, 2011; Park *et al*., 2011; Milne *et al*. 2011). One of such addressed problems has been the upcoming energy crisis shortage, causing the funding of many projects that focus on the biomass-derived fuels production. The production of biofuels in microorganism has been explored in attractive hosts like cyanobacteria. These organisms have simple growth requirements; grow to high densities, and use light, carbon dioxide, and other inorganic nutrients efficiently (Deng and Coleman, 1999).

The oligotrophic cyanobacterium *Synechococcus elongatus* PCC7942 has become an interesting platform for bioconversion of light and CO_2, into value-added substances. This freshwater and unicellular microorganism (formerly known as *Anacystis nidulans* R2), stands as a prokaryotic model for exploring the circadian rhythms (Andersson *et al*., 2000; Kondo *et al*., 1993). Engineering techniques for genetic manipulation of cyanobacterial including this strain have been reported by, for example, Clerico *et al*., 2007, highlighting the usefulness of this system as a genetic host. Additionally, it was the first cyanobacterium to be cloned with exogenous DNA (Shestakov and Khyen, 1970).

This chapter presents a manually curated genome-scale reconstruction for *Synechococcus elongatus* PCC7942, which represents the first metabolic model for this biological system.

2.2 *Synechococcus elongatus* PCC7942 genome

Several cyanobacterial sequencing project have been completed and others, not less important, are in progress. Nowadays, more than 50 genomes are available in different databases like GenBank of the National Center for Biotechnology Information (NCBI) (Benson *et al*., 2013), Cyanobase (http://genome.kazusa.or.jp/cyanobase), and others. Most of them were performed, not only to elucidate ecological aspects or some distinctive feature of the genus, but for their biotechnological potentialities.

In 2004, the DOE Joint Genome Institute (JGI) concluded the genome sequencing project of the *Synechococcus elongatus* PCC7942, for which professor Susan S. Golden´s group, of the Texas A&M University, was the leader. The sequence assembly process included shotgun sequencing, as a result of JGI project, and transposon-mediated mutagenesis and sequencing strategy to determine the sequences surrounding transposon insertions in essentially every gene, as proposed by Golden´s lab (Holtman *et al*., 2005).

The annotation of the genome for protein and RNA-encoding regions was carried out by two computer predictions. As part of this project, the annotation was performed by Computational Biology Department at ORNL (DOE Oak Ridge National Laboratory) for JGI and the same time by to the Annotation Engine of TIGR (The Institute for Genomic Research) that provide manual curation support (Holtman *et al*., 2005, You Chen´s doctoral dissertation, 2007). Along this, the complete genome of *Synechococcus elongatus* PCC6301, a strain of this specie provided the possibility in determining the genetic loci that account for this major difference (Sugita *et al*., 2007).

A circular chromosome of approximately 2.7 Mb, an essential large plasmid (pANL, ~46.3 kb), and a non-essential small plasmid (pANS, ~8.6 kb) are the genome components of *S. elongatus* PCC7942 (van den Hondel *et al*., 1980; Van der Plas *et al*., 1992; Chen *et al*., 2008). Genomic details can be studied in Table II.A.

Table II.A. Summary of genome features of *Synechococcus elongatus* PCC7942.

	Chromosome	Plasmid pANL	Plasmid pANS
Length of DNA (base pairs)	2695903	46366	7835
G+C (%)	~ 55.47	52.9	~59
RNA genes	54	-	-
rRNA genes	6	-	-
tRNA genes	45	-	-
Other RNA genes	3	-	-
Protein genes	2856	50	8
With predicted function	1682	17	-
Without predicted function	1174	33	-
Total genes	2906	50	8

2.3 *Synechococcus elongatus* PCC7942 metabolic model

2.3.1 Reconstruction procedure

The reconstruction of genome-scale metabolic models is a process which aims to achieve the simulation of the cell metabolism. The closeness of the simulations to the *in vivo* behavior of a given organism will depend on the quality of the data and on the caution of the researcher performing each step of this procedure. Several referential works can instruct and lead researchers in this task (Förster *et al*., 2003; Notebaart *et al*., 2006; Feist *et al*., 2009; Thiele and Palsson, 2010).

The start of this process consisted on the download of the current annotation and genomic sequence files for the *Synechococcus elongatus* PCC7942. A NCBI Entrez Gene repository (NCBI, 2013), a specialized database of genomes and genetic information, provided the fundamental starting point for this endeavor. Subsequently, PathoLogic, a component of Pathway Tools software (Karp *et al*., 2002) was used to construct a database from the genomic sequence and annotations files. This resulted in an object-oriented database that includes all genes, proteins, metabolites and reactions present in the organism. Alternatively, we used COPABI computational platform to build a similar database and then to automatically generate the metabolic model following probabilistic criteria of uniqueness and completeness (Reyes *et al*., 2012). This strategy

was conceived in order to double check the results of both tools and to take into account the probabilistic criteria in order to fill gaps in metabolic pathways or to exclude duplicated reactions suggested by COPABI. The figure 2.1 summarizes the process.

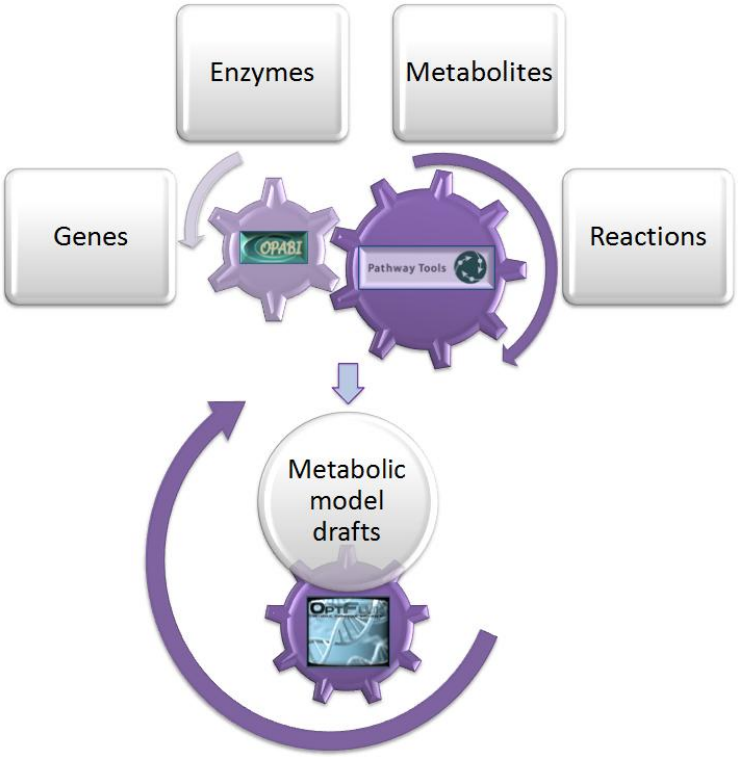


Figure 2.1. Genome-scale metabolic models reconstruction process.

At this point a depuration process of all genes and metabolic pathways was started; biochemistry books and journal articles were needed to complete the task. In addition, public repositories which useful for the reconstruction such as the specific databases for metabolic pathways, enzymes and metabolic compounds are displayed in Table II.B.

Table II.B. Examples of databases that assisted in the *Synechococcus elongatus* PCC7942 reconstruction process.

Databases of genome and genetic information	
NCBI Entrez Gene (National Center of Biotechnology Information, provides, among others, genomic information)	www.ncbi.nlm.nih.gov/sites/entrez
DDBJ(DNA Data Bank of Japan)	www.ddbj.nig.ac.jp
EMBL-Bank (Europe's primary nucleotide sequence database)	www.ebi.ac.uk/embl
KEEG (Kyoto Encyclopedia of Genes and Genome)	www.kegg.com
BioCyc (collection of genomes and metabolic pathways)	www.biocyc.org
JGI Genomes (DOE-Joint Genome Institute, is a joint repository of various microorganisms and eukaryotic	genome.jgi.doe.gov

genomes)	
CyanoBase (Database of cyanobacterias genomes)	genome.kazusa.or.jp/cyanobase
Databases of metabolic pathways	
KEEG (Kyoto Encyclopedia of Genes and Genome)	www.kegg.com
BioCyc (collection of genomes and metabolic pathways)	www.biocyc.org
BioCarta (Biological Pathways databases)	www.biocarta.com
Databases of enzymes	
ExPASy-Enzyme (Enzyme nomenclature database)	www.expasy.org/enzyme
BRENDA (The Comprehensive Enzyme Information System)	www.brenda-enzymes.org
IntEnz (Integrated relational Enzyme database)	www.ebi.ac.uk/intenz
SABIO (Biochemical Reaction Kinetics Database)	sabiork.h-its.org
CAZy (Carbohydrate Active enZyme database)	www.cazy.org
Databases of metabolic compounds	
EBI-ChEBI (Chemical Entities of Biological Interest)	www.ebi.ac.uk/chebi
PDB (The Chemical Component Dictionary)	remediation.wwpdb.org/ccd.html
LIPIDMAPS (LIPID Metabolites and Pathways)	www.lipidmaps.org
LipidBank (The official database of Japanese Conference on the Biochemistry of Lipids (JCBL))	lipidbank.jp
KNApSAcK (A Comprehensive Species-Metabolite Relationship Database)	kanaya.naist.jp/KNApSAcK

We thoroughly reviewed the experimental data that support the presence of metabolic genes. Reactions such as the ones involved in genetic replication, gene expression and cell division, which are not commonly included within the metabolic models, also have been assessed and removed from final model (Thiele and Palsson, 2010). We had in consideration the presence of isoenzymes and protein subunits that form multimeric enzymes. Thus, the reactions catalyzed by multimeric enzymes or enzymatic complexes were described as a single reaction (Förster *et al*., 2003).

The enzymes nomenclature, the stoichiometry, the name of the metabolites and the reversibility were checked for each biochemical reaction and verified with the help of the Enzyme nomenclature database (Bairoch, 2000) and KEGG pathway database (Kanehisa *et al.*, 2008). Some enzymes, metabolic pathways and compounds databases (see Table II) provided assistance regarding synonyms use, stoichiometric coefficients of metabolites and directionality of chemical transformation. In some cases a single enzyme may catalyze different reactions; in our model we identified each of these reactions with different identifiers as part of the name. In order to follow the law of mass conservation we balanced all elements in every reaction, except protons, so that chemical conversions were coherent. Moreover, some metabolites are frequently reported in reactions' databases in a non-specific form (e.g. 'an alcohol', 'an electron acceptor', etc) due to lack of experiments that corroborate the existence of certain

chemical transformations. This non-specificity was insufficient for our modeling goals and, thus, specific nomenclature was sought. Certain published metabolic models, such as *Synechocystis* sp. PCC6803 (Montagud *et al*., 2010, 2011), that preserve phylogenetic relationship with *Synechococcus elongatus* PCC7942 (belonging to the same Phylum), have served as a reference to solve these issues. Through this analysis, we assessed the possibility of including missing cofactors (e.g. water molecule or hydrogen ion, among others) in some type of reactions, like the ones catalyzed by hydrolases, oxidoreductases or transferases. Else, if state of the art was unable to specify a single cofactor requirement, like NADH or NADPH, two reactions were included in the reconstructed metabolic network. If no conclusive evidence about reversibility was found, reactions were set to be reversible.

No lumped reactions were left in the model. Thereby, some traditional pathways like photosynthesis or oxidative phosphorylation were described as a set of reactions, thus enabling the tracing of the corresponding fluxes. On the other hand, non-enzymatic reactions were included in the metabolic model. The inclusion of these filled the synthesis or degradation of some compounds that appear at the beginning or at the end of a metabolic pathway, known as dead-end metabolites (Thiele and Palsson, 2010).

Upon completion of the metabolic model draft, we looked for possible reactions gaps in the metabolic network. Many gaps result from the presence of blocked reactions (that may be related with dead-end metabolites disconnection), as well as, from incorrect gene annotation but for which one can find proper biochemical evidence (Thiele and Palsson, 2010). In that debugging step we filled out the gaps by an exhaustive search in databases and journal articles looking for experimental data that supports the addition of these reactions to the model. Examples of these are two reactions whose presence is necessary to complete the glyoxylate shunt, specifically the genes that encode for malate synthase (EC number: "*2.3.3.9"*) and isocitrate lyase (EC number: *"4.1.3.1"*), such enzymatic activities have been measured (Pearce and Carr, 1967) but presently do not have a cognate Open Reading Frame (ORF) associated to them. Another reaction that was included in our models even though no cognate gene has been indentified is the aquaporins reaction that was suggested in a study of

the osmotic stress effects on the photosynthetic machinery (Allakhverdiev *et al.*, 2000).

Nonetheless, some dead-end metabolites were kept in the curated model; meaning that they are required for metabolic model´s functionality, like the biomass formation (buildings blocks) and there is no evidence at all of their connection to another part of the metabolism or to a possible transport out of the cell. In fact, many reactions are necessary for the synthesis of these or others buildings blocks, but that have no corresponding enzyme-coding gene assigned. Hence, many gene and enzymes were needed to be included in the model to allow the formation of biomass. Another essential point in the debugging process was the removing of internal loops that are thermodynamically unfeasible, e.g. futile cycles (like substrate cycles (Voet and Voet, 2004)) and Type III-extreme pathway (Schilling *et al.*, 2000b). That´s crucial since several constraint-based approaches, such as FBA, does not account for regulation, thus futile cycles cannot be shut down otherwise and could retrieve unnatural flux behaviors.

2.3.2 Versions

2.3.2.1 *i*Syf646 metabolic model

After exhaustive searches the first resulting metabolic model for *S. elongatus* PCC7942 (*i*Syf646) consisted in 835 reactions and 803 metabolites. These reactions are encoded by 646 genes, to which 472 enzymes were assigned. The presences of protein complexes and multimeric enzymes, explain the differences between the number of enzymes and genes.

A set of reactions with no cognate genes are present in *i*Syf646: 13 non-enzymatic (spontaneous) conversions, 16 passive transport reactions and 11 unassigned reaction (the majority according to KEGG report), have been included. In the debugging process we also included 80 reactions whose ORFs were not annotated in the genome on the basis of biochemical evidence or physiological considerations. Additionally, 319 reactions of *i*Syf646 were found to be reversible, 54 external metabolites and 40 exchange reactions have been taken into account. Table II.C shows the model overview.

The stoichiometric model comprises 62 metabolic pathways that include the synthesis and degradation reactions of primary and secondary metabolites. Thus, some processes like photosynthesis, oxidative phosphorylation and glycogen biosynthesis are described in detail.

*i*Syf646 genome-scale metabolic model is in OptGene (Patil *et al*., 2005) format.

2.3.2.2 *i*Syf715 metabolic model

Our second version of *S. elongatus* PCC7942 (*i*Syf715) metabolic model was generated as an upgrade of the first release. We assessed the incorporation of several reactions for which, at that time, we did not have enough information. Secondly, we included and corrected many isoenzymes and complexes that were incomplete in *i*Syf646. More specifically, this version of the model includes 851 reactions and 838 metabolites, 326 reactions were found to be reversible. The bulk of reactions are catalyzed by 530 enzymes encoded by 715 genes. We have previously explained the reason for the difference between the number of genes and reactions, which we note here again.

Reactions with no cognate genes are still present in *i*Syf715. Here, the only difference relates to the reactions whose ORFs were not annotated in the genome, in this case 76 reactions, that were included due to experimental evidence of their presence. *i*Syf715 genome-scale metabolic model is available in Appendix 1.1 (in OptGene (Patil *et al*., 2005) format). Table II.C displays the general characteristics of both models.

Both final models include central metabolic pathways such as glycolysis/gluconeogenesis pathway, the Calvin-Benson cycle, the pentose phosphate pathway, incomplete reactions within the tricarboxylic acid cycle (TCA), as well as the complete set of anabolic pathways involved in biosynthesis of chlorophyll, glycogen, amino acids, lipids, nucleotides, vitamins, cofactors, etc. Pathways for glyoxylate synthesis (*via* ribulose-1,5-bisphosphate carboxylase/oxygenase and the shunt across the TCA), and aminosugars metabolism are also included.

Table II.C. Distribution of the model reactions as per cognate genes.

GENERAL OVERVIEW	***i*Syf646**	***i*Syf715**
Number of genes	646	715
Number of metabolic reactions	835	851
Number of metabolites	803	838
Enzymes	472	530
Multimeric enzymes and enzymatic complexes	69	79
REACTION OVERVIEW		
Reversible reactions	319	326
Irreversible reactions	516	525
Reactions with assigned genes	**715**	**735**
Enzymatic conversion	703	710
Protein-mediated transport (active and passive-mediated transports)	12	25
Reactions with no cognate genes	**120**	**116**
Non-enzymatic conversion (spontaneous)	13	13
Passive transport reactions (simple diffusion)	16	16
EC reactions not annotated	80	76
Unassigned reactions	11	11

Photosynthetic electron transfer associated with the thylakoid membrane is represented as a set of 10 separate reactions, including light capture by photosystem II (PSII) and photosystem I (PSI), electron transfer between the two photosystems, and cyclic electron transfer which involves PSI ferredoxin.

2.3.3 Formulation of biomass equation

Once the cyanobacteria metabolism has been reconstructed, we focused on the stoichiometric network analysis. As we explained in the previous chapter, to calculate the quantity of the optimal behavior through some constraint-based approach, like FBA, it is essential to define an objective function. The biomass growth is the most commonly used objective function to simulate the metabolic flux distribution, and it

has become a standard to assess the best metabolic engineering strategies (Edwards *et al*., 2001; Ibarra *et al*., 2002).

From a biochemical point of view, one might think that describing a comprehensive biomass synthesis equation would be a difficult task. Especially if one considers molecular plurality and the quasi-static and dynamic intermolecular relationships among elements. However, in this case, the formulation of biomass composition lies on the stoichiometric coefficients of all of the substances that have a modular construction (building blocks). They consist of linked monomeric units that make up the lowest level of structural hierarchy of the cell. In particular, biomass growth is expressed by transforming the building blocks, such as: amino acids, deoxyribonucleotides, ribonucleotides, lipids, carbohydrates, antenna chromophores, some cofactors, etc, into one mole of biomass. Thus, the growth flux is defined as a metabolic flux utilizing the biosynthetic precursors, X_m, in the appropriate ratios to produce biomass:

$$\sum_{\text{all m}} d_m \cdot X_m \rightarrow \text{biomass}$$

where d_m is stoichiometric coefficients (biomass fraction) of the metabolite X_m (Edwards *et al.*, 2002).

Seeking information about weight fractions of macromolecules and monomers to reflect the composition of any organism is critical. Frequently, data related to the relative amounts of these metabolites are not available in literature, or the published information is shown in a particular physiological condition not usable for our goals. The quantities measured in other phylogenetically related biological system, could be a close approximation to the metabolic reality of the concerned organism.

As a part of the reconstruction process we detailed a biomass equation for *S. elongatus* PCC7942. Little is known about the specific molecular quantities of this cyanobacterium. However, the previous study of Rosales-Loaiza *et al*. in *Synechococcus* sp., isolated from a hypersaline waterhole, served as a reference in the composition of total protein, chlorophyll *a,* β-carotene and zeaxanthin (Rosales-Loaiza *et al*., 2008). We regard that total protein quantities per gram of dry weight (gDW) is not enough to describe the composition of this macromolecule in the cell. Especially for being one of

the polymers more diverse and valuable in terms of the monomeric units composition. Hence, we adapted the amino acid quantities by selecting the well-studied biomass composition of *Synechococcus sp.* PCC 7002 metabolic model as a template (Hamilton and Reed, 2012). Bearing in mind that the optimum growth rates (calculated by sensitivity analysis) do not change drastically by varying the monomeric composition of the major macromolecules (Varma and Palsson, 1994a). Because the photosynthetic carbon assimilation in cyanobacteria results in the accumulation of polysaccharides, mostly glycogen (Nakamura *et al*., 2005), we agree to define the composition of total carbohydrates as the amount of this polymer. Here, the carbohydrate composition measured in *Synechococcus* sp. strain PCC 7002 (Xu *et al.*, 2012) was assumed in our models.

We opted to include values of carotenoid pigment, in this case trans-lycopene, as biomass precursors. This substance was taken from data reported for *Synechosystis sp.* PCC6803 (Miao *et al*., 2003). Moreover, we estimate ratios between the concentrations of chlorophyll a and phycocyanobiline measures in *S. elongatus* (González-Barreiro *et al.*, 2004). Thus, phycocyanobiline´s amounts were incorporated into biomass equation according to the chlorophylls quantities fixed.

In addition, the lipids amounts were based on the data displayed for *S. elongatus* PCC7942 used as a limiting resource to feeding an herbivore (Martin-Creuzburg *et al*., 2009). Additionally, the molar quantities for the deoxyribonucleotides and ribonucleotides were defined from the information available in the works of Herdman *et al*. and Allen and Smith, respectively (Herdman *et al*., 1979; Allen and Smith, 1969).

Furthermore, with the goal to form a demand objective function (*viz.*: biomass growth) the information on the maintenance energy requirements must be included. The energy is used for both growth-associated and nongrowth-associated maintenance functions (Stouthamer, 1979; Förster *et al*., 2003). Some of them are the cells active transports, membrane potentials, turn-over of macromolecules, maintenance of concentration gradients (pH or osmotic pressure), mobility and the ATP cost that is required for the polymerization of amino acids and nucleotides and for the synthesis of building blocks. Because the lack of data, we used the same maintenance energy requirements for *Synechocystis* sp. PCC6803 growth, in this instance 59.28 mmol ATP

gDW^{-1} were added to biomass equation (Montagud *et al*., 2010). This equation is reaction "Biomass" in the whole metabolic model showed in Appendix 1.1 and Table II.D illustrates its composition.

Table II.D. Biomass composition of both *S. elongatus* PCC7942 metabolic models.

Amino acid counts of the proteome		**Metabolites**	**mmol/gDW**
Alanine	897	**Proteins**	0.000459
Arginine	526	**Carbohydrates**	
Aspartate	518	Glycogen	0.53439
Asparagine	374	**Antenna chromophores**	
Cysteine	102	Zeaxanthin	0.00079
Glutamine	576	Beta-carotene	0.000875
Glutamate	614	Trans-lycopene	0.00820225
Glycine	702	Chlorophyll a	0.0057
Histidine	197	Phycocyanobiline	0.0285
Isoleucine	628	**Deoxyribonucleotides**	
Leucine	128	dATP	0.0201156
Lysine	417	dTTP	0.0201156
Methionine	194	dGTP	0.02538445
Phenylalanine	406	dCTP	0.02538445
Proline	512	**Ribonucleotides**	
Serine	548	AMP	0.140389293
Threonine	580	UMP	0.140389293
Tryptophan	149	GMP	0.123745851
Tyrosine	294	CMP	0.123745851
Valine	638	**Lipids**	
		14C-lipid	0.028
		16C-lipid	0.0042
		18C-lipid	0.00448
		(9Z)16C-lipid	0.0066

	(9Z)18C-lipid	0.00625

At this time, both models were designed that could be applied to linear optimization to simulate cellular behavior (see next Chapter). However, this step is not conclusive, since the computed results could not be in agreement with experimental ones. This incongruity allows the iterative adjustment of the model, as part of debugging process, until that gene-reactions list be closer to natural state.

2.3.4 Network topology. Connectivity analysis

The topological analysis of *Synechococcus elongatus* PCC7942 metabolic network can lead us to understand how the metabolites and the interactions between them (biochemical reactions) determine the metabolic function into the cell. In representative manner, compounds involved in a specific number of reactions, will form a node with certain connections on the map of metabolic reactions. This approach has been widely used in systems biology, with many works from many researchers from different fields, like Mathematics and Statistics (Barabási and Oltvai, 2004; Tanaka, 2005; Mahadevan and Palsson, 2005).

Many studies have shown that biological systems share an important property with others complex system: most of these networks are scale-free and thereby the nodes connection can be approximated by a power law distribution (Barabási and Albert, 1999; Albert *et al.*, 2000; Barabási and Bonabeau, 2003; Barabási and Oltvai, 2004). That probability distribution implies that small events are very common, while oversized occurrences are less so. Most nodes have just a few connections and some have a tremendous number of links (known as metabolic hubs). That means, the probability of a given node has exactly "*k*" links follows a distribution $P(k) \sim k^{-\gamma}$, where the exponent "γ" is the degree exponent. Typically the exponent values fall in the range $2<\gamma<3$ (Barabási and Albert, 1999).

We gave the task of finding what metabolites are more connected in the network. Also, all reported hubs were compared with their relative presence in other existing genome-scale metabolic models (Table II.E).

Table II.E. in *S. elongatus* PCC7942 and others metabolic models. "*N*" are the number of connections.

Metabolic hubs	*N* in *i*Syf646	*N* in *i*Syf715	*N* in *i*Syp611	*N* in *i*Syn811	*N* in *i*Cce806	*N* in i*AF1260*	*N* in *i*FF708
H2O	238	243	202	219	245	697	-
phosphate	174	169	154	112	174	81	113
ADP	165	159	125	111	143	253	131
ATP	150	148	129	136	153	338	166
H^+	139	149	291	153	350	923	188
diphosphate	107	110	103	84	108	28	-
CO2	72	69	58	72	64	53	66
AMP	68	74	63	21	66	86	48
NADPH	61	74	69	68	78	66	57
$NADP^+$	59	72	64	68	71	39	61
L-glutamate	49	52	44	44	50	52	56
NAD^+	48	46	50	52	66	79	58
NADH	43	45	52	48	67	75	52
oxygen O2	43	45	29	40	33	40	31
S-adenosyl-L-methionine	37	37	13	28	25	18	19
ammonia	34	44	26	28	33	22	-
coenzyme A	32	29	26	23	28	71	39
pyruvate	31	32	30	20	32	61	20
L-glutamine	31	30	25	21	28	18	23
glutathione	31	32	9	26	11	17	10
S-adenosyl-L-homocysteine	26	25	7	24	15	12	14

*i*Syp611: *Synechococcus sp.* PCC 7002 metabolic model (Hamilton and Reed, 2012).
*i*Syn811: *Synechocystis* sp. PCC6803 metabolic model (Montagud *et al*., 2011).
*i*Cce806: *Cyanothece sp.* ATCC 51142 metabolic model (Vu et al., 2012).
*i*AF1260: *E. coli* metabolic model (Feist *et al*., 2007).
*i*FF708: *Saccharomyces cerevisiae* metabolic model (Förster *et al*., 2003).

For more details we decided to assess metabolites connectivity to reactions. In this case, the links between metabolites to enzymes as well as metabolites more connected as substrates or as products (see Table II.F). Similarly, we studied connectivity among them, highlighting those with larger connections according the edges directionality (in-connected and out-connected) (see Table II.G).

Table II.F. Most connected metabolites to enzymes, as a substrate and as a product in *S. elongatus* PCC7942.

Most connected metabolites to enzymes (Metabolite/#connection)		**Most connected metabolites as substrates (Metabolite/#connection)**		**Most connected metabolites as products (Metabolite/#connection)**	
***i*Syf646**	***i*Syf715**	***i*Syf646**	***i*Syf715**	***i*Syf646**	***i*Syf715**
H2O (232)	H2O (234)	H2O (193)	H2O (195)	H2O (127)	H2O (130)
ATP (159)	ATP (156)	ATP (158)	ATP (154)	ADP (113)	H^+ (116)
H^+ (143)	H^+ (155)	H^+ (109)	H^+ (119)	Phosphate (109)	ADP (110)
Phosphate (119)	Phosphate (117)	NADPH (55)	NADPH (63)	H^+ (108)	phosphate (105)
ADP (115)	ADP (113)	Phosphate (45)	$NADP^+$ (50)	Diphosphate (93)	Diphosphate (95)
Diphosphate (95)	Diphosphate (97)	$NADP^+$ (43)	ADP (43)	CO2 (69)	CO2 (67)
CO2 (73)	CO2 (72)	ADP (42)	phosphate (43)	$NADP^+$ (55)	$NADP^+$ (63)
$NADP^+$ (60)	$NADP^+$ (70)	NAD^+ (37)	NAD^+ (38)	ATP (44)	NADPH (50)
NADPH (60)	NADPH (70)	NADH (30)	NADH (30)	NADPH (43)	ATP (45)
NAD^+ (48)	NAD^+ (47)	Oxygen O2 (29)	Oxygen O2 (27)	AMP (39)	AMP (42)
NADH (43)	L-glutamate (44)	S-adenosyl-L-methionine (28)	L-glutamate (27)	NADH (36)	NADH (37)
L-glutamate (43)	AMP (43)	L-glutamate (25)	S-adenosyl-L-methionine (27)	NAD^+ (34)	NAD^+ (33)
AMP (40)	NADH (43)	Glutathione (23)	Glutathione (23)	L-glutamate (33)	L-glutamate (33)
Oxygen O2 (34)	Ammonia (33)	Diphosphate (22)	Diphosphate (22)	S-adenosyl-L-homocysteine (23)	Ammonia (27)
S-adenosyl-L-methionine (29)	Oxygen O2 (32)	Malonyl-ACP (22)	Malonyl-ACP (22)	Ammonia (22)	S-adenosyl-L-homocysteine (22)
Ammonia (27)	S-adenosyl-L-methionine (28)	CO2 (18)	CO2 (18)	Coenzyme A (21)	Pyruvate (21)
Glutathione (27)	Glutathione (28)	L-glutamine (18)	Pyruvate (17)	5-oxoproline (20)	Coenzyme A (20)
Pyruvate (23)	Pyruvate (26)	Pyruvate (15)	L-glutamine (17)	Cysteinylglycine (19)	5-oxoproline (20)
Coenzyme A (23)	Malonyl-ACP (23)	Acetyl-CoA (14)	Acetyl-CoA (16)	Pyruvate (18)	Cysteinylglycine (19)
S-adenosyl-L-homocysteine (23)	Coenzyme A (22)	2-ketoglutarate (13)	Ammonia (16)	Oxygen O2 (16)	Oxygen O2 (14)
Malonyl-ACP (23)	S-adenosyl-L-homocysteine (22)	Coenzyme A (12)	2-ketoglutarate (13)	Tetrahydrofolate (13)	2-ketoglutarate (13)
L-glutamine (21)	L-glutamine (21)	Ammonia (12)	AMP (12)	2-ketoglutarate (11)	Tetrahydrofolate (13)

Table II.G. Most out-connected and in-connected metabolites in *S. elongatus* PCC7942.

Most out-connected metabolites (Metabolite/#connection)		**Most in-connected metabolites (Metabolite/#connection)**	
***i*Syf646**	***i*Syf715**	***i*Syf646**	***i*Syf715**
H2O (203)	H2O (206)	Phosphate (166)	Phosphate (158)
ATP (150)	ATP (146)	ADP (163)	ADP (155)
H^+ (110)	H^+ (119)	H2O (144)	H2O (152)
NADPH (57)	NADPH (68)	H^+ (110)	H^+ (116)
Phosphate (56)	Phosphate (55)	Diphosphate (105)	Diphosphate (108)
ADP (48)	$NADP^+$ (55)	AMP (67)	AMP (73)
$NADP^+$ (46)	ADP (49)	CO2 (66)	$NADP^+$ (64)
oxygen O2 (39)	oxygen O2 (41)	$NADP^+$ (53)	CO2 (62)
NAD^+ (38)	NAD^+ (39)	NADPH (48)	NADPH (56)
S-adenosyl-L-methionine (35)	L-glutamate (37)	ATP (47)	ATP (48)
L-glutamate (33)	S-adenosyl-L-methionine (35)	L-glutamate (40)	L-glutamate (41)
NADH (32)	NADH (33)	NADH (38)	NADH (40)
Diphosphate (31)	Diphosphate (31)	NAD+ (37)	Ammonia (36)
L-glutamine (27)	L-glutamine (25)	Coenzyme A (31)	NAD+ (35)
CO2 (25)	Glutathione (25)	Ammonia (28)	Coenzyme A (29)
Glutathione (25)	CO2 (24)	S-adenosyl-L-homocysteine (26)	Pyruvate (26)
Pyruvate (21)	Ammonia (24)	Pyruvate (25)	S-adenosyl-L-homocysteine (25)
L-aspartate (21)	Pyruvate (23)	oxygen O2 (21)	oxygen O2 (21)
2-ketoglutarate (19)	L-alanine (23)	5-oxoproline (20)	Acetate (20)
L-alanine (19)	AMP (21)	Acetate (19)	5-oxoproline (20)
Acetyl-CoA (17)	L-aspartate (21)	Tetrahydrofolate (19)	tetrahydrofolate (19)
Coenzyme A (17)	Acetyl-CoA (19)	Cysteinylglycine (19)	UDP (19)

From comparisons between the most connected metabolites in different microorganisms, we appreciate a similarity regarding the usage of these compounds by these diverse metabolic machineries.

Unsurprisingly, in all the analyses we could see that water is the more connected compound. Water has direct involvement in various kinds of reactions, both as a substrate or product, most notably in hydrolysis, condensation and reduction-oxidation reactions.

Other highly connected nodes are carrier molecules such as: ADP, ATP, $NADP^+$, NAD^+, phosphate and oxygen, most of them are involved in the majority of energy

releasing and/or consuming pathways. Most of the metabolic pathways are regulated, at least in part, by levels of nucleotides such as ATP and ADP. The ATP molecule carries energy between different routes as it is a reactive intermediate capable of coupling endergonic reactions to exergonic ones. Furthermore, free energy released by catabolic processes is conserved through the reduction of the coenzyme $NADP^+$ to NADPH. In turn, the NAD^+/NADH redox couple functions as the electron acceptor in many exergonic metabolites oxidations, while also serves as electron sink along a series of redox complex reactions of increasing reduction potential until a final acceptor, such as the oxygen in Electron Transport Chain (Voet and Voet, 2004).

A few amino acids and peptides were shown to have high values of connectivity, such as: L-glutamate, L-glutamine, glutathione and cysteinylglycine. The first one, among other functions, plays a central role in processes related to transamination as well as the synthesis of other amino acids. While L-glutamine, by its ability to donate the amino group, has an essential function as intermediate in the vitamin B_{12} pathway, in the purine, folate and glutathione metabolism, among others (Voet and Voet, 2004). In this latter process, wherein also is involved the peptide cysteinylglycine, is produced the largest endogenous antioxidant in the cells, the glutathione per se (Pompella *et al.*, 2003).

The S-adenosyl-L-methionine/S-adenosyl-L-homocysteine pair is another example of metabolites that behave as typical hubs. Several methylases use the first as substrate, making it a major biological methylating agent, and yield the second as product (Finkelstein and Martin, 2000). The S-adenosyl-L-methionine is an essential metabolite in certain metabolic pathways like the porphyrin and chlorophyll metabolism.

Additional highly connected metabolites, and no less important, are the ammonia, coenzyme A and pyruvate, constituting either the substrates or products of many metabolic pathways, for example: glycolysis, tricarboxylic acid (TCA) cycle, glyoxylate shunt and amino acids metabolism.

The network topology analysis of both *S. elongatus* PCC7942 metabolic models (*i*Syf646 and *i*Syf715) can help in the understanding of how metabolites and their interactions determine their metabolic function in the cell. As many studies have

shown, most of these networks are scale-free and thereby the nodes connection can be estimated by a power law distribution (Barabási and Albert, 1999; Albert *et al.*, 2000). For the connectivity distribution analysis of *i*Syf715 we used a systematic mathematical approach: the Pareto´s law in terms of the cumulative distribution function ($P\ (K>k) \sim k^{-\gamma}$) to get a proper fit (Newman, 2005; Hardy, 2010). We used the cumulative distribution rather than obtaining a classic log-log scale plot of the distribution of connections number among number of nodes, because the tail smoothes out in the cumulative distribution and no data is "obscured" as in the logarithmic binning procedure (Adamic and Huberman, 2002). In particular, fitting this continuous distribution, we find an equation y=-1.189*x+7.23 for *i*Syf646 whose slope is equivalent to an exponent of γ=2.189, and equation y=-1.203*x+7.19 for *i*Syf715 with γ=2.203. See figure 2.2 for more details.

Our analysis points towards the fact that both genome-scale metabolic networks are characterized by a power-law distribution with high non-uniformity as we can see at the cumulative distribution toward the right; most of the nodes have only a few associations. This corresponds with the partial results for metabolites connections shown in Tables II.E, II.F and II.G. It follows from this that exist, evidently, a reduced number of chemical transformation that the majority of metabolites can undergo (data not shown). The biological significance of this hierarchical connectivity should be related with evolutionary process, where the hubs were the first compounds that were present in the earliest cells predecessor´s metabolism (Wagner and Fell, 2001). Likewise, the "attack tolerance" of a network is such that the removal of the most highly connected nodes has the broadest impact on network functions (Jeong *et al.*, 2000; 2001). It should be noted that highly connected nodes may represent effective targets for metabolic engineering; or at least should be considered in the design of strategies for the production of other metabolites. However, topological properties of networks must be interpreted in the more biologically relevant functional network states and their properties. One such consideration, for instance, is that a metabolic network must make all the biomass components of the cell in order for it to grow. Thus, even eliminating a step in a linear low-flux pathway leading to the synthesis of

cofactors, vitamins, or amino acids will prevents a genome-scale metabolic network from supporting growth.

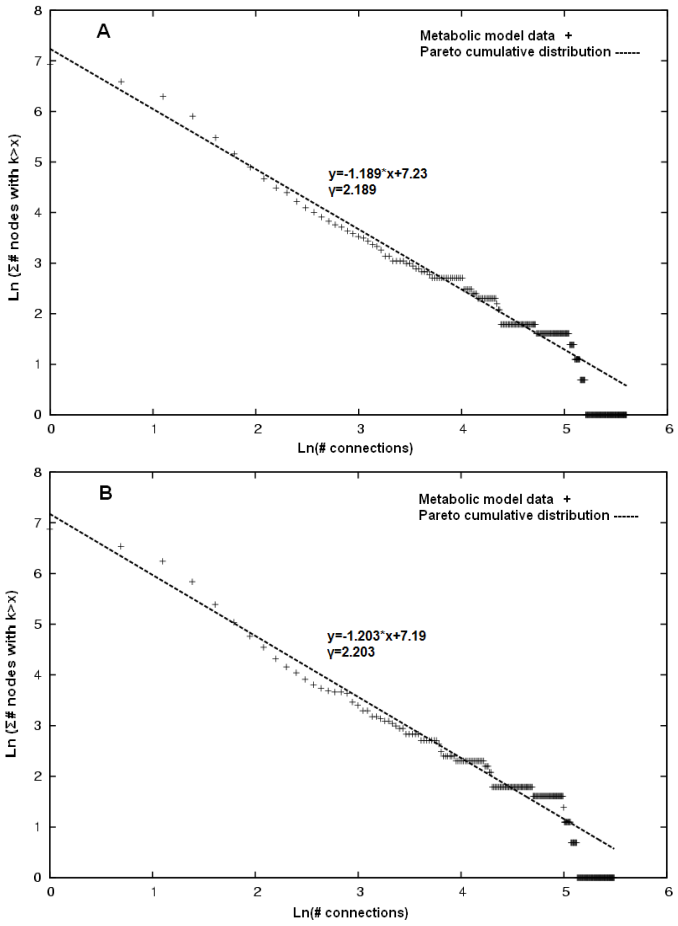


Figure 2.2. Connectivity distribution of two *S. elongatus* PCC7942 metabolic models. **A**. Cumulative distribution toward the right via Pareto function plot of *i*Syf646. **B**. Cumulative distribution toward the right via Pareto function plot of *i*Syf715.

In order to compare the compounds participation in networks, we computed the connectivity for all metabolites in others genome-scale metabolic network reconstructions. The distributions of node connectivities were found for *Synechocystis* sp. PCC6803 (*i*Syn811) (Montagud *et al.*, 2011), *Saccharomyces cerevisiae* (*i*FF708) (Förster *et al*., 2003), *Spirulina platensis* C1 (*i*AK692) (Klanchui *et al.*, 2012) and *Clostridium beijerinckii* NCIMB 8052 (*i*CM925) (Milne *et al.*, 2011). As seen in the figure 2.3, all the reconstructed networks follow a power-law distribution. There is a high probability of low connectivity as well as a low probability of high connectivity. Once again, the most highly connected nodes are carrier molecules.

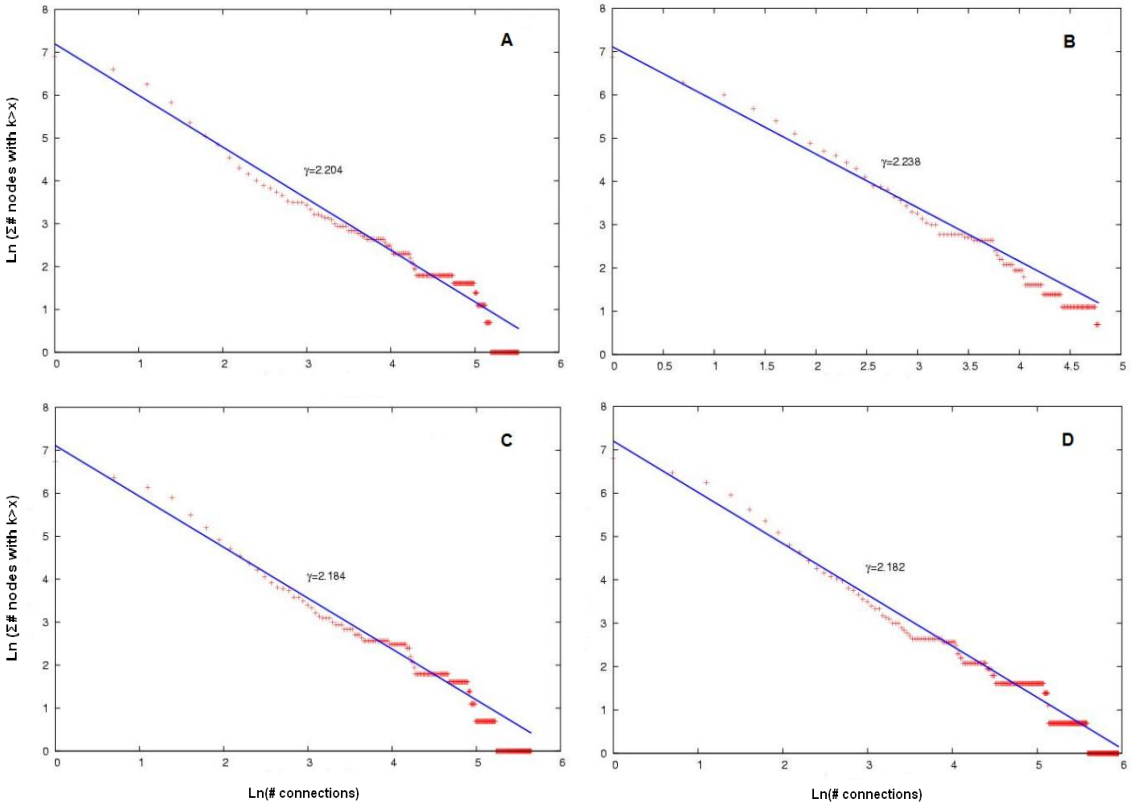


Figure 2.3. Connectivity distribution of four genome-scale metabolic network model reconstructions. Red crosses and blue lines indicate the metabolic model data and Pareto cumulative distribution, respectively. **A**. Cumulative distribution toward the right via Pareto function plot of *i*Syn811. **B**. Cumulative distribution toward the right via Pareto function plot of *i*FF708. **C.** Cumulative distribution toward the right via Pareto function plot of *i*AK692. **D.** Cumulative distribution toward the right via Pareto function plot of *i*CM925.

2.4 Conclusions

We have successfully reconstructed the first genome-scale metabolic network for *Synechococcus elongatus* PCC7942. This chapter has detailed the reconstruction process. By this proceeding we obtained a first version of the metabolic model, *i*Syf646, which was subsequently updated in a second version *i*Syf715. The curated models represent an up-to-date database that encompasses all knowledge available in public databases, scientific publications and textbooks on the metabolism of this cyanobacterium. The model has been compiled in OptGene and SBML to enable its use with different software.

This laborious task was initially performed with the support of two semi-automatic reconstruction tools: Pathway Tools and COPABI. These softwares, along with some manual curation steps, yielded the *i*Syf646 model consisting in 646 genes, 835

reactions and 803 metabolites. While the upgraded version, *i*Syf715, comprises 715 genes, 851 reactions and 838 metabolites. Moreover, in the last version we identified 76 enzymatic reactions needed for the correct function of the metabolism, but with no annotated cognate gene. These genes are interesting targets for experimental studies as we have seen that their presence is required in order to build up the basic cellular components.

From the topological perspective the characteristics of the model are very similar to other published organisms' providing support for an evolutionary study of the structure and organizational properties of metabolic networks, in the line of recent works (Gamesman *et al.*, 2014). The connectivity analysis of the model using the Pareto cumulative distribution shows a scale-free behaviour with a highly non-uniformity and a hierarchical connectivity of the metabolites, which is typical of biological networks and points towards functional properties discussed in other works (Csete and Doyle, 2004). Thus, *i*Syf646 and *i*Syf715 represent a common blueprint for the large-scale organization of interactions among all cellular constituents.

These models join the growing list of genome-scale metabolic models and specifically to those belonging to this phylum, which has its potential as the photosynthetic model organisms.

Synechococcus elongatus PCC7942 metabolic model is a useful tool in the fields of systems biology and metabolic engineering. Applicability of *i*Syf646 and *i*Syf715 is demonstrated by using a variety of computational analyses in following chapters.

Chapter 3. *In silico* fluxomic behavior through constraints-based approach

A significant milestone in a modeling program is the successful representation of the behavior of the biological system by a model. One of the main reasons of model usage is hypothesis building and testing, which allows to rapidly analyze the effects of manipulating experimental conditions in the model without having to perform complex and costly experiments, and ideally decreasing the number of performed experiments. In this chapter, we predicted the repertoires of metabolic fluxes that optimize a given biological function, as well as we validated the results with reported experimental data.

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

Triana J, Montagud A, Siurana M, Gamermann D, Torres J, Tena J, Fernández de Córdoba P, Urchueguía JF. **Generation and evaluation of a genome-scale metabolic network model of *Synechococcus elongatus* PCC7942**. *Submitted at Metabolites*.

Triana J, Montagud A, Gamermann D, Fernández de Córdoba P, Urchueguía JF. ***In silico* analysis for bio-products synthesis through genome-scale reconstruction of the *Synechococcus elongatus* PCC7942 metabolic network**. *In preparation.*

Chapter 3. *In silico* fluxomic behavior through constraints-based approach

3.1 Introduction

The reconstructed metabolic models allow direct correlation between the genomic information and metabolic activity at the flux level. Through the use of a mathematical framework researchers can simulate and quantify the optimal interactive behavior of hundred of reactions (and its cognate genes) with hundred metabolites, under different conditions (Förster *et al*., 2003; Edwards *et al*., 2001).

With such metabolic models, researchers can quantitatively predict maximum theoretical production yields. Examples of these yields are biomass growth (Ibarra *et al*., 2002; Edwards *et al*., 2001), synthesis of metabolite with economical value (Varma *et al*., 1993a) or both simultaneously (Burgard *et al*., 2003, Pharkya *et al*., 2004), as well as the minimization of ATP production (Vo *et al*., 2004) or the uptake rate of nutrients (Famili *et al*., 2003). These are just some examples that have already been shown to be effective as a means of improving production yields.

The applicability of genome-scale metabolic models has been demonstrated through the use of several computational analyses. Constraint-based approaches, such as FBA, are one of the most common used to simulate phenotypic behavior under imposed physiological and/or genetic conditions (Stephanopoulos *et al*., 1998; Price *et al*., 2003; Durot *et al*., 2009). FBA aims to obtain, through the optimization of a cellular objective (usually growth), the space of allowable flux distributions of a biological system under steady-state conditions. The optimization problem is subject to a set of constraints associated with minimum (lower) and maximum (upper) bounds in every reaction, which are defined by thermodynamical and experimental data. Finally, the resulting flux distribution can be contrasted with *in vivo* information and, thus, the metabolic model can be used for further analyses (Price *et al*., 2004b; Orth *et al*., 2010).

Cyanobacteria, and within them *Synechococcus elongatus* PCC7942, are oxygenic photosynthetic prokaryotes. This genus represents important in vivo biomass photoautotrophic producers that are widespread in diverse environments (Whitton and Potts, 2000). Also, they possess unique biochemical properties which make them a

promising model to transform carbon sources into valuable substances (Quintana *et al*., 2011).

The focus of this study is on a mathematical analysis of the reconstructed *Synechococcus elongatus* PCC7942 metabolic network. The constraint-based optimization methods, introduced in the first chapter, are widely used and can be repeatedly applied to compute the metabolic flux distribution under different conditions.

3.2 Finding optimal states

There are a few number of genome-scale metabolic models for cyanobacteria, and therefore, that completely accounts for photosynthesis. The reconstruction and analysis of the *Synechocystis* sp. PCC6803 metabolic model provides an indisputable reference point, when photosynthetic description for these prokaryotes it comes. There have been published several studies that have increased our understanding of the metabolic behavior of this organism and have targeted its use as a biological production platform (Yang *et al*., 2002a, 2002b; Shastri and Morgan, 2005; Kun *et al*., 2008; Fu, 2008; Montagud *et al*., 2010; Knoop *et al*., 2010; Montagud *et al*., 2011; reviewed in Montagud *et al.*, 2013). Other interesting works, such as the metabolic reconstruction of *Cyanothece sp.* ATCC 51142, have provided a detailed overview of interactions between components of photosynthetic and respiratory systems (Vu *et al*., 2012). By means of a bi-level mixed-integer programming approach, the tool CONGA, was useful to develop a stoichiometric model for another cyanobacterium within the *Synechococcus* genus. Thus, the *Synechococcus sp.* PCC 7002 metabolic model was described thanks to the comparison with *Cyanothece sp.* ATCC 51142 reconstruction (Hamilton and Reed, 2012).

It´s well known that some experimental parameters imposed on growth conditions, such as: unrestricted light, carbon, or nitrogen-limited cultures, may lead to significant differences in cell´s macromolecular composition. The measurements of biomass molecular quantities would be essential to fulfill the biomass formulation in this circumstance. Despite this, we have performed simulation studies with the same biomass equation and varying these parameters. We have done this having in mind that simulation results are very little altered by changes in the biomass monomeric

composition as was described by some works (Varma and Palsson, 1994a; Shastri and Morgan, 2005).

Our developed *i*Syf646 and *i*Syf715 models constitute the theoretical baseline of *S. elongatus* PCC7942´s metabolism at the genomic level. The models were simulated by constraint-based approaches methods, as those mentioned above, through setting objectives functions and constraints for the quantitative calculation of optimal states. The validation of our models was the starting point for the assessment of metabolic capabilities which will be discussed in the next chapter.

3.2.1 Constraints settings for system simulation

Synechococcus elongatus PCC7942 is said to be an obligate photoautotroph organism, thus we defined a set of constraints for this growth conditions. This means, that energy comes from light and carbon source from CO_2. In order to be able to simulate this biological system, we need assume a quasi steady-state and the imposition of thermodynamical and biological constraints (Covert *et al.*, 2003; Price *et al.*, 2004b). This allows reduce the undetermined possible solution space to a biologically-feasible solution space that contains the metabolic flux distribution (null vector space of the stoichiometric matrix) (see first Chapter).

The autotrophic growth for *i*Syf646 and *i*Syf715 was simulated using a two-step optimization. The first step was the maximization of biomass growth while the light intake was unconstrained. Next, the maximum growth value was incorporated as a constraint to minimize the light uptake rate (the second step). This was designed with the aim to estimate physiologically meaningful photon uptake values that tallied experimental measurements. Here, an important point for accurate calculation is that the capacity constraints must be close as possible to natural fluxes.

Firstly, in order to estimate a theoretical maximum illumination (radiant flux density), the surface area per weight of biomass had to be calculated. Considering the geometry of the cell as a prolate spheroid (see methods) in which length and width of cell are reported to be 3.57 ± 0.12 µm and 1.47 ± 0.09 µm, respectively, and the dry weight is 3.87 ± 0.03 ng, approximately (Rosales-Loaiza *et al*., 2005). Then, we decided to find an appropriate irradiance value because the lack of our own experimental data for growth versus irradiance. Some studies have tested the growth of various

Synechococcus sp. strains, fixing irradiance values between 0.03 mE m^{-2} s^{-1} and 0.234 mE m^{-2} s^{-1} (Bertilsson *et al.*, 2003; Rosales-Loaiza *et al*., 2005; Fu *et al*., 2007). However, Rosales-Loaiza *et al*. have asserted that in 12:12 hours photoperiod with values around to 0.156 mE m^{-2} s^{-1} this cyanobacterium had the highest growth (Rosales-Loaiza *et al*., 2008). Thus, assuming this irradiance value the theoretical maximum illumination that would reach the cell membrane, in units, used in our model can be calculated as 1.96 mE gDW^{-1} h^{-1}. This light flux value is assumed to be the fraction of the total radiant flux density, which is actually converted into chemical energy by the system.

The dissolved inorganic carbon (Ci) exists, on aquatic ecosystems, mainly as two slowly interconvertible species, CO_2 and carbonic acid ($H_2CO_3 \rightarrow HCO_3^- + H^+$). Active transport for CO_2 and HCO_3^- are present in *Synechococcus elongatus* PCC7942 in at least four Ci uptake systems with different affinities (Badger and Price, 2003; Price *et al*., 2004a). Although the CO_2 and HCO_3^- are interconvertible, it has been shown that both are simultaneously and continuously transported by the cells (Espie *et al*., 1988). These uptake systems constitute an efficient CO_2-concentrating mechanism in picocyanobacteria together with the carboxysome that contains the enzyme that catalyzes carbon fixation (ribulose-1,5-bisphosphate carboxylase/ oxygenase (RuBisCO)) (Price *et al*., 2004a). Kajiwara and co-workers reported that this cyanobacterium achieved a maximum CO_2 uptake rate of 0.025 g L^{-1} h^{-1} at a cell mass concentration of 0.286 g L^{-1} (Kajiwara *et al*., 1997). Thus, the first optimization was carried out by constraining the CO_2 and HCO_3^- uptake rates at 1.99 mmol gDW^{-1} h^{-1}.

Additionally, some transport systems across the membrane such as: phosphate, water, sulphate, nitrate, ammonia as well as carbon monoxide and hydrogen peroxide transport were included in the model and properly bounded. Some of the reversible reactions involving NADH and NADPH were constrained to be irreversible so that spurious transhydrogenation was controlled.

The second simulation aims at a minimization of photon uptake rate, constraining the biomass growth rate at the value obtained from the first optimization. In biological terms, we assume that the cells will optimally growth using the amount of measured CO_2 and only needing the minimized photon usage.

Main constraints across the autotrophic growth condition can be seen in Table III.A.

Table III.A. Principal constraints across the autotrophic growth condition. Units in mmol gDW^{-1} h^{-1}, except for light input which is in mE gDW^{-1} h^{-1}.

Constraints	Values in first Optimization*	Values in second Optimization*
Light input in PSI	0; 1.96	0; 0.1
Light input in PSII	0; 1.96	0; 0.1
CO_2 uptake rates	0; 1.99	0; 1.99
HCO_3^- uptake rates	0; 1.99	0; 1.99
Nitrate uptake rates	0; 160	0; 160
CO uptake rates	-10; 10	-10; 10
Sulphate uptake rates	-104; 104	-104; 104

* Values indicate, consecutively, minimum and maximum boundaries.

3.2.2 Fluxes' vector space of optimal metabolic growth. Metabolic models validation

With the purpose to evaluate and validate the predictive accuracy of *i*Syf646 and *i*Syf715, we used FBA to simulate a given physiological behavior. Model validation usually focuses on testing whether the growth capabilities, or any particular objective flux, correspond to a given set of experimental data. In fact, the known inability of the organism can also be evaluated to further confirm the reconstruction accuracy (Thiele and Palsson, 2010). Due to the biphasic nature of cyanobacterium growth, we look for reported data for exponential growth phase with which to compare the *in silico* simulation of the models. Maximum specific growth rate of S*ynechococcus elongatus* PCC7942 has been reported between 0.0519 h^{-1} to 0.0551 h^{-1} (Kuan, 2013; Kajiwara *et al.*, 1997), despite scarce information about studies on optimization the specific growth rate of this prokaryote. Model was simulated under the aforementioned conditions while optimizing biomass functions to simulate photoautotrophic growth. Interestingly, our simulations resulted in a maximum specific growth of 0.05987 h^{-1}. The slight difference with experimental data could be the result of several factors, including regulation, stress and feedback inhibition, which cannot be captured in

constraints-based stoichiometric models. Moreover, the growth of many laboratory strains are not consistent with the computed optimal by FBA because they are not necessarily evolved for growth maximization (Lewis *et al.*, 2010a; Imam *et al.*, 2011). These results are, therefore, an overall acceptable validation of the genome-scale metabolic model.

As part of the second optimization problem we constrained the *in silico* growth rate to this value and sought for the minimum of light uptake as the objective function. This minimization resulted in photon uptake for photosystem I (reaction "*_lightI"*, in *i*Syf715) and photosystem II (reaction "*_lightII"*, in *i*Syf715) being 0.1 mE gDW^{-1} h^{-1}.

After verifying that our models could reproduce experimental growth rate, we investigated the metabolic fluxes' vector that corresponds to that growth. As expected, during photoautotrophic growth we could see a flux distribution directed towards the CO_2 fixation at the Calvin-Benson cycle. Likewise the solution space reveals that the autotrophic growth flows in the gluconeogenic direction (see figure 3.1). High fluxes through RuBisCO (reaction "*4.1.1.39b"*, in *i*Syf715) and carbonic anhydrase (reaction "*4.2.1.1b"*, in *i*Syf715) are due to the need of carbon dioxide and carbonic acid as the carbon sources. These are in correspondence with the natural conversion of light energy and carbon sources into complex carbohydrate molecules in cyanobacterium cells. The demand of ATP and NADPH (reducing power) is covered by its synthesis associated with electronic transfers of light-excited photosystems. The presence of non-cyclic electron transport influences their synthesis. Similarly, the included cyclic electron transport contributes to produce ATP through generation of a proton motive force via the pumping of H^+ across the membrane. Here, a significant flows are observed over the ferredoxin-$NADP^+$ reductase (reaction "*_1.18.1.2"*, in *i*Syf715) and across the ATPase (reaction "*_3.6.3.14"*, in *i*Syf715), the governing complex in photophosphorylation process. Thus, the ATP/NADPH ratio is in correspondence with the hypothesis that the existence of both electron transports must be essential for efficient photosynthesis (Munekaga *et al.*, 2004).

As example, the photophosphorylation supplied the energy required for the active regeneration of both D-glyceraldehyde-3-phosphate from 3-phosphoglycerate as D-ribulose-5-phosphate from D-ribulose-1,5-bisphosphate. On the other hand, the

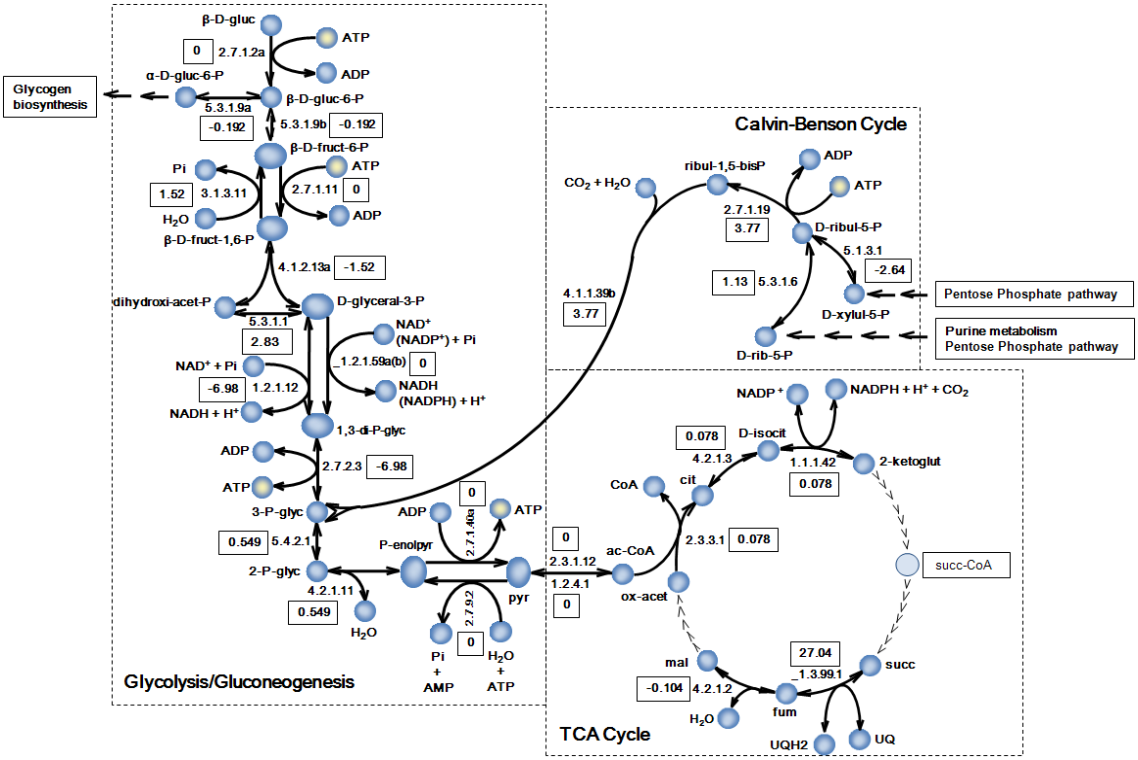
reducing power provided by the oxidation of water via non-cyclic electron transport is essential for many synthesis pathways, such as: L-threonine, L-lysine, L-proline, L-phenylalanine, L-valine, folate, coenzyme A, chlorophyll biosynthesis among other amino acid and cofactor metabolic pathways. Additionally, we have noted the activation of glycogen production through ADP-D-glucose synthesis, which serves as a form of energy storage. Glyoxylate shunt is basically inactive. Interestingly, we have also observed fluxes of acetyl-CoA consumption towards the biosynthesis of some biomass formation precursors, e.g. amino acids and fatty acids.

Results of the FBA simulations of some split-flux reactions in the cyanobacterium's central metabolism are shown in figure 3.1. Appendixe 1.2 contains the entire metabolic flux vector generated by the described optimization techniques.

It is noteworthy that reactions such as those catalyzed by glucokinase (reaction *"2.7.1.2a"*, in *i*Syf715), phosphofructokinase (reaction "*2.7.1.11"*, in *i*Syf715) and pyruvate kinase (reactions *"2.7.1.40a"*, in *i*Syf715), associated with catabolism of carbohydrates, exhibit no metabolic flux, which is in correspondence with photoautotrophic growth conditions (Buchanan *et al.*, 2000; Voet and Voet, 2004). Whereas others like those catalyzed by glucose-6-phosphate isomerase (reactions *"5.3.1.9a,b"*, in *i*Syf715), fructose 1,6-bisphosphate phosphatase (reactions *"3.1.3.11"*, in *i*Syf715) and phosphoribulokinase (reaction *"2.7.1.19"*, in *i*Syf715) as well as those that produce D-ribulose-1,5-bisphosphate and catalyzed by ribose-5-phosphate isomerase (reaction *"5.3.1.6"*, in *i*Syf715) and ribulose-phosphate 3-epimerase (reaction *"5.1.3.1"*, in *i*Syf715), operate at relatively high reaction rates.

Finally, we would like to focus on the high flux value of the reversible reaction "*_1.3.99.1*" catalyzed by succinate dehydrogenase. In this case the direction implies the succinate oxidation to fumarate reducing ubiquinone. The synthesis of fumarate is an essential reaction since it represents an intermediate node in many metabolic pathways that yield building blocks for biomass formation; such as: pyrimidine (specifically, uridine monophosphate (UMP)) and aspartate biosynthesis. It is known that this cyanobacterium does not have abundance of complex morphological characteristics (Robertson *et al.*, 2001); therefore, the reduced ubiquinone could be oxidized by other processes, such as photosynthesis electronic transfer, contributing to

the formation of NADPH and ATP without triggering their classical synthesis reactions' fluxes.



Abbreviations: β-D-gluc = β-D-glucose; β-D-gluc-6-P = β-D-glucose-6-phosphate; α-D-gluc-6-P = α-D-glucose-6-phosphate; β-D-fruct-6-P = beta-D-fructose-6-phosphate; β-D-fruct-1,6-P = β-D-fructose-1,6-bisphosphate; dihydroxi-acet-P = dihydroxy-acetone phosphate; D-glyceral-3-P = D-glyceraldehyde-3-phosphate; Pi = phosphate; 1,3-di-P-glyc = 1,3 diphosphateglycerate; 3-P-glyc = 3-phosphoglycerate; 2-P-glyc = 2-phosphoglycerate; P-enolpyr = phosphoenolpyruvate; pyr = pyruvate; ac-CoA = acetyl-CoA; ox-acet = oxaloacetate; mal = malate; fum = fumarate; UQH2 /UQ = reduced ubiquinone/oxidized ubiquinone; succ = succinate; succ-CoA = succinyl-CoA; 2-ketoglut = 2-ketoglutarate; D-isocit = D-isocitrate; cit = citrate; CoA = coenzyme A; ribul-1,5-bisP = D-ribulose-1,5-bisphosphate; D-ribul-5-P = D-ribulose-5-phosphate; D-rib-5-P = D-ribose-5-phosphate; D-xylul-5-P = D-xylulose-5-phosphate

Figure 3.1. Metabolic flux map of some reactions when optimizing biomass function to simulate growth. Here are shown the glycolysis and gluconeogenesis pathways, the incomplete TCA cycle and a partial depiction of the Calvin-Benson cycle present in cyanobacterium metabolism. Note that the empty arrows, in the TCA cycle, indicate the absence of these reactions in the pathway. The values beside each EC number, and locked in box, represent the predicted fluxes via FBA with v_{uptake} (CO_2; HCO_3^-) = 1.99 mmol gDW^{-1} h^{-1}. All of these are expressed in mmol gDW^{-1} h^{-1} and correspond to the maximum of the optimization problem. Fluxes are not necessarily consistent from one reaction to the next because other, smaller flux pathways have interplay with the reactions here.

3.2.3 Flux variability analysis

Although the FBA leads to an optimal objective function solution, it is true that there are alternate flux distributions that could yield equivalent optimal solutions. Thus, a given reaction can be having different flux while still resulting in an optimal solution of the optimization problem.

With the goal to find the allowable flux range for a particular reaction in these alternate optima, we used the Flux Variability Analysis (FVA) algorithm. This approach has been used to find ranges of values for all fluxes in the set of alternative optima without affecting the simulated objective function (Mahadevan and Schilling, 2003).

We looked for the range of minimum and maximum flux values of each reaction in *i*Syf646 and *i*Syf715. All reactions that have the same minimum and maximum flux value, *viz.:* non-zero and zero values, as well as those that have different non-zero minimum and maximum flux value were computed. Additionally, we decided to count all reactions that achieved non-zero minimum values whereas maximum flux is zero, and vice versa and those which varied their flow between -∞ and +∞, non-zero and +∞, -∞ and non-zero, zero and +∞ and between -∞ and zero. The figure 3.2 shows in details these results.

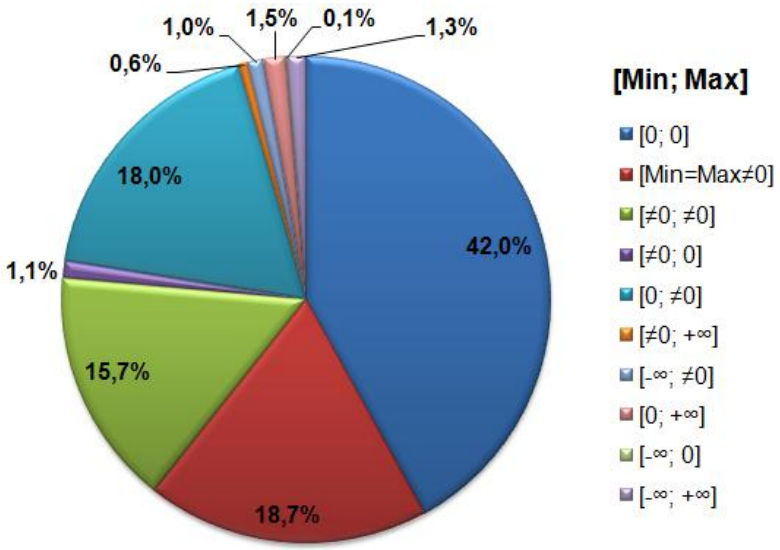


Figure 3.2. Overview of all minimum and maximum fluxes of each reaction given the biomass growth optimal value in *i*Syf646 and *i*Syf715.

As an example, the fluxes distributions of some diagrammed reactions in figure 3.1 were investigated by this approach. The range in which reactions can change without altering the maximum biomass growth (0.05987 h^{-1}) is shown in figure 3.3.

The variation observed on reactions *"5.4.2.1"* and *"4.2.1.11"* gives us an idea of the strong coupling between them. In this case, the maximum flow values indicates that both reactions occur in the sense of the isomerization of 3-phosphoglycerate to 2-phosphoglycerate, and dehydration of the latter to become phosphoenolpyruvate, respectively.

However, alternative solutions demonstrate that higher fluxes can take place in the meaning of gluoneogenesis. This is biologically possible since the 3-phosphoglycerate represents a hub within this pathway. In others words, it is a direct product of CO_2 fixation as well as the reduction of pyruvate through pyruvate-water dikinase (reaction *"2.7.9.2"*). So, its transformation until α-D-glucose-1-phosphate enables glycogen synthesis; a polysaccharide that accumulates and also forms an important component of *Synechococcus elongatus* PCC7942 biomass (Nakamura *et al.*, 2005).

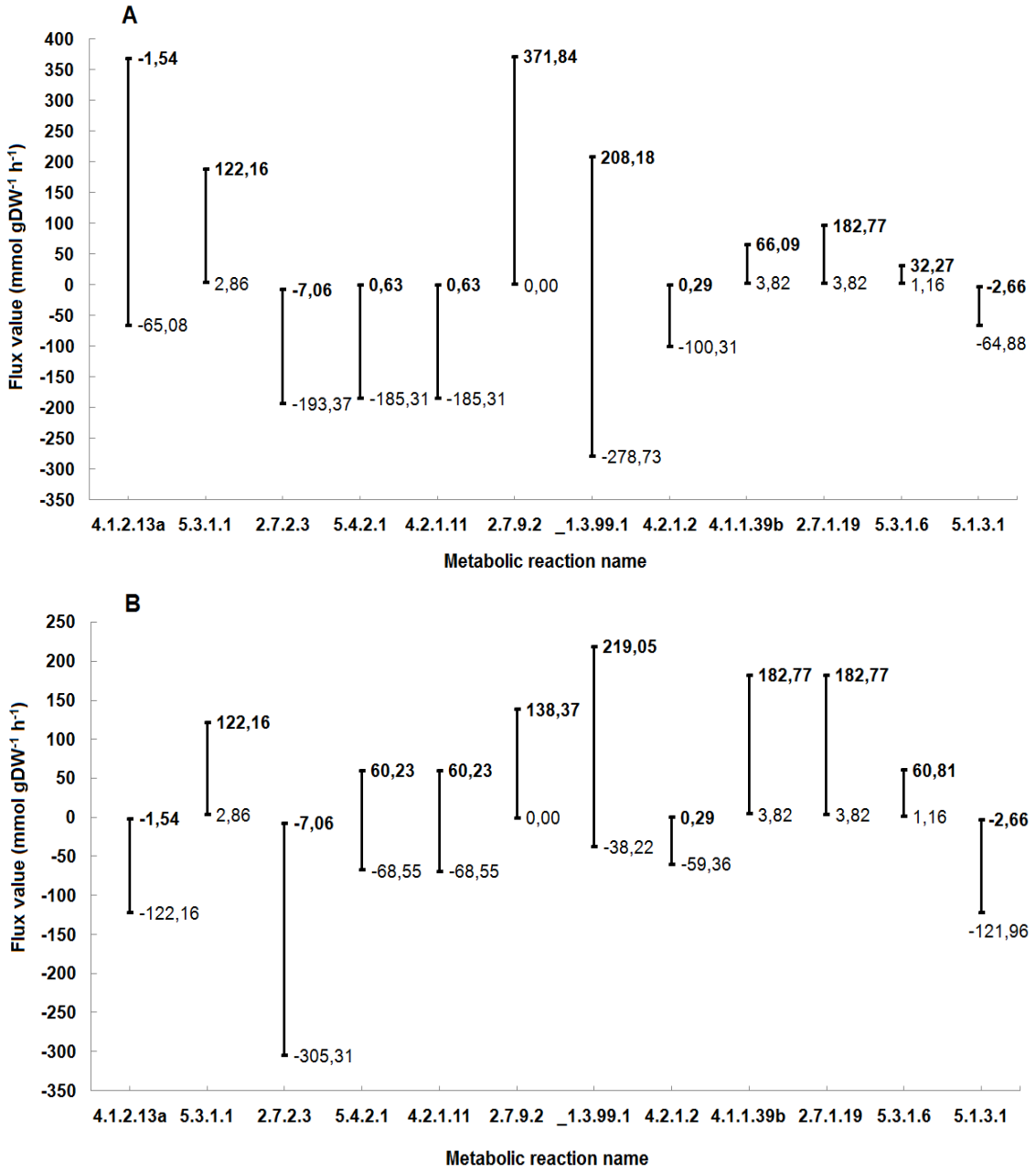


Figure 3.3. Flux distribution, calculated by FVA, of some central metabolism reactions in *i*Syf646 (A) and *i*Syf715 (B). The y-axis indicates the flux values of each variability range point

bounded by the maximum and minimum values. Bold values correspond to the predicted maximum values for each reaction.

Similarly, the variability seen in reaction *"_1.3.99.1"* indicates that this reaction may proceed, thus, in the direction of the oxidation to fumarate as in reducing to succinate, both with relatively high values. Further corroborating the TCA cycle amphibole nature. Besides the importance of flow direction towards the emergence of fumarate (mentioned above), it is also essential the succinate presence, as an intermediate metabolite, in many metabolic pathways; such as: glyoxylate shunt and pyrimidine biosynthesis. Although, the flux variability of reaction catalyzed by fumarate hidratase (reaction *"4.2.1.2"*) takes both positive and negative values, it appears that higher levels of metabolic flux are directed toward the malate production; a metabolite involved in the only active reaction (*"·2.3.3.9"*), and in reverse sense, in the glyoxylate shunt (see Appendix 1.2).

3.3 Robustness analysis of metabolic model networks

When we evaluate the optimal states by FVA, to some extent we are assessing the parametric sensitivity of the metabolic network. But many times it is appropriate to perform a Robustness Analysis (RNA) (Edwards and Palsson, 2000) to estimate the metabolic network stability under certain perturbations.

The notion of robustness has recently received considerable interest in diverse fields related to complex networks. Examples include Internet, social networks, and biology (Strogatz, 2001; Stelling *et al*., 2004; Kitano *et al*., 2004). In general, robustness means the persistence of a system´s characteristic behavior under perturbation or condition of uncertainty. The concept is closely related to stability in dynamical system theory, but usually employed with respect to a broader class of phenomena (Kitano, 2002; Carlson and Doyle, 2002).

We assessed the sensitivity of the metabolic network by iterative calculation of optimal states, varying an environmental parameter over a given range of values. We examined the effect of variations in CO_2 and HCO_3^- uptake rates as constraints that support the optimal biomass growth. The range of these parameters was established from zero to 80 mmol gDW^{-1} h^{-1}. In every calculation, these constraints varied in an increment within this range.

As the Ci flux constraint was increased from the zero value, it was determined that the capability of the metabolic network to support growth was sensitive to all Ci flux values between the range. Figure 3.4 displays the biomass growth plotted as a function of these variables, where we can observe a nearly linear behavior in the biomass production with increasing values of CO_2 and HCO_3^- uptake rates. *Synechococcus* sp. growth rates in high CO_2 level were observed by Fu *et al*. (Fu *et al*., 2007), as well as in other works with several autotrophic organisms (Burkhardt and Riebesell 1997, Burkhardt *et al*. 1999, Yang and Gao 2003, Kim *et al*. 2006).Theoretically, our results can be explained because the increased steady-state uptake flux of this substrate leads to amplify its incorporation to the greater part of building blocks´ carbon skeleton. In natural circumstances, the *in silico* results can be explained because there are constitutive and inducible transporters in response of CO_2 and HCO_3^- availability in the Ci pool (Shibata *et al.*, 2002; Raven, 2003). Even so, this sensitivity analysis shows us the capacity of these input variables to disrupt the growth of the cyanobacterium, and their importance in the models.

The CO_2 and HCO_3^- uptake rates required for fully carbon fixation by photosynthetic apparatus correspond to the peaks of the both curves. These optimal points correlate with perfect conversion of these substrates into precursors that build the biomass, perhaps for the high Ci concentration inside the carboxysome that causing the saturation of RuBisCO. The fact that there is a little difference between the growths values in the peaks is due to that HCO_3^- is a metabolite involved in a plurality of functions. Stands out as an intermediary in some biomass precursor's production, like fatty acids and pyrimidines. Also, this ion, which dominates in the alkaline oceans, is accumulated in the carboxysome and converted into CO_2 by carbonic anhydrase, a enzyme closely related with RuBisCO. Thus, the metabolic flux is distributed across all nodes associated with its consumption. Whereas CO_2, besides being the substrate of the CO_2-concentrating mechanism, is also a product of many reactions of primary metabolism. So, theoretically can be reused when build up in the carboxysome.

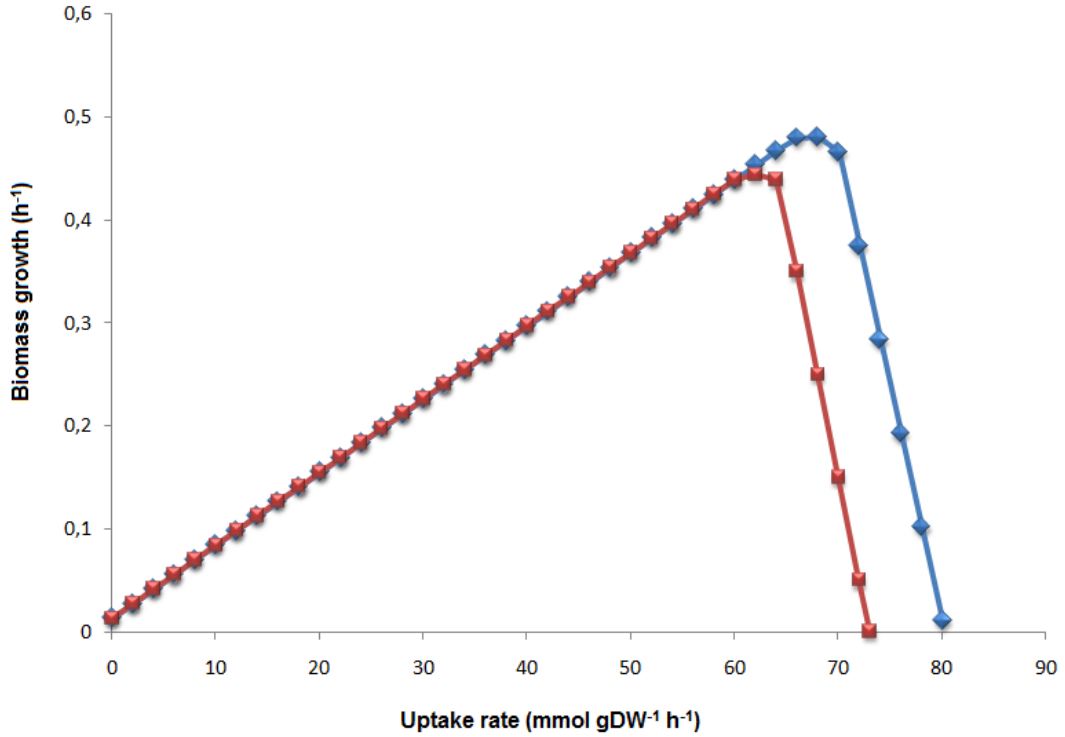


Figure 3.4. Effects on *Synechococcus elongatus* PCC7942 maximal growth of varying CO_2 and HCO_3^- uptake rates. Red squares and blue diamonds mark biomass production values in correspondence with changes in HCO_3^- and CO_2 uptake rates, respectively.

The Ci flux could be increased to about 3116% for HCO_3^- and 3417% for CO_2 of the *in silico* wildtype before severe limitations in the growth flux were encountered. Thus, the post-peak segment shows the decline in the cyanobacterium growth as the carbon input increases. In this region, too much Ci is taken up relative to photons and the growth rate drops due to dissipation of the excess Ci. It seems that the energy and electrons requirements for biomass production are channeled in reactions related with this dissipation. Beyond this, the segment represents an unrealistic physiological situation similar to those found by Edwards and Palsson (Edwards and Palsson, 2000) in the sensitivity analysis of *E. coli* metabolic network altering the flux of some reactions in the central metabolic pathways. Furthermore, this behavior is common under selective pressure, in which the biological system could adapt to new change by altering its phenotypic state (Rosales-Loaiza *et al*., 2005; Rosales-Loaiza *et al*., 2008).

The defined regions in the phenotypic phase plane (PhPP) analysis (see Chapter 5) when two parameters are varied simultaneously (Edwards *et al*., 2002), can explain this unrealistic physiological situation.

3.4 Conclusions

Through reconstruction and validation of *Synechococcus elongatus* PCC7942 metabolic model, we have simulated autotrophic flux landscapes. The quantifications made by different computational analyses reveal the applicability of these genome-scale metabolic network models, proving, once more, the usefulness of these *in silico* reconstructions.

The cyanobacterium's autotrophic growth was simulated by Flux Balance Analysis with OptGene software. To do this, the objective function taken into account was the biomass synthesis, a proxy of cell growth.

Additionally, we evaluated the flux variability across all reactions that support the optimal growth of this prokaryote. Moreover, the CO_2 and HCO_3^- uptake rates were subjected to a robustness analysis to investigate the optimal systemic effect of flux alteration in biomass growth. Reaffirming the important role of these main carbon sources in metabolic functionality.

In the next chapter we will continue exploring the metabolic capabilities of *Synechococcus elongatus* PCC7942, presenting a mutants study targeting at the improved production of biofuels.

3.5 Methods

3.5.1 Cell surface area calculation

With the goal to compute theoretical maximum illumination per weight of biomass we calculated cyanobacterium cell surface area as a prolate spheroid. This is a type of spheroid in which the polar axis is greater than the equatorial diameter (figure 3.5) (Hilbert and Cohn-Vossen, 1999).

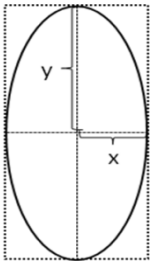


Figure 3.5. Ellipsoid from whose rotation around its major axis is generates a prolate spheroid. In where "y" indicates the polar axis and "x" indicates equatorial diameter.

Then, the surface area can be calculated as:

$$A = 2\pi x^2 (1 + \frac{y}{xe} \sin^{-1} e) \quad \text{(Beyer, 1987)}$$

where "e" may be identified as the eccentricity and can be calculated as:

$$e^2 = 1 - \frac{x^2}{y^2}$$

3.5.2 Flux Balance Analysis

We have used the FBA in all simulations in this chapter. In this optimization method the objective function is a linear equation which can be solved by linear programming (LP). This approach can be implemented in a computing environment, even for large-scale systems such as biological ones (Orth *et al*., 2010). The details of this constraints-based analysis are outlined in the first Chapter.

In general terms, the whole metabolic network can be represented in a matrix (S), known as stoichiometric matrix. This arrangement is composed of the stoichiometric coefficients of each internal metabolite in biochemical reactions. Mathematically, this matrix is a linear transformation of flux vector (v) into a vector of time derivatives of the metabolite's concentration vector (x):

$$\frac{dx}{dt} = S \cdot v$$

This equation represents the fundamental equation of dynamic mass balance that characterizes all functional states in the network.

The method assumes that the system operates under the assumption of steady-state, meaning that the amount of a metabolite (x_i) in all reactions does not vary in time (t). Thus,

$$\frac{dxi}{dt} = \sum_j Sij \cdot vj = 0$$

where the individual equation in the set represents the summation of all fluxes that form x_i and those that degrade it (Edwards *et al*., 1999). Taking this in vectors v and x, we have $S \cdot v = 0$, the main FBA constraint. This physicochemical limitation represents itself, a set of linear equations where the number of equations (one for each metabolite) is much smaller than the number of unknown variables (fluxes of

each reaction). Consequently, this set is indeterminate. Other physico-chemical constraints such as the reactions reversibility and the energy required for cell maintenance, as well as experimental constraints, are imposed on mathematical model associated to linear optimization method (Orth *et al*., 2010).

This problem can be formulated as follows:

$$\text{Max } (v_j)$$

$$\text{Subject to } S \cdot v_j = 0 \quad \forall j \in N$$

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

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

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

where is the rate of the j^{th} reaction. The elements of the flux vector v were constrained for the definition of reversible and irreversible reactions, , and ,r, respectively. Additionally, a set of equations was established, , that contain constrained metabolic reactions desirable for simulation, as well as uptake reactions which were bound by experimentally determined values from the literature.

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

3.5.3 Flux Variability Analysis

The range of maximum and minimum flux for a given reaction, that supports a particular optimal state, can be computed by this approach (Palsson, 2006). For a given flux (v_j), we can determine this range by solving two LP problems:

$$\text{Max or Min } (v_j)$$

$$\text{Subject to } S \cdot v_j = 0 \quad \forall j \in N$$

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

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

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

$$v_j = v_{opt}$$

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

3.5.4 Robustness Analysis

Environmental and genetic parameters can be varied over a range of values, in order to achieve this sensitivity analysis. This variation can be done in a staggered manner and the optimization problem can be solved by linear programming like FBA. In this case, every parameter value in the range is set as a constraint on the mathematical model (Edwards and Palsson, 2000). Thus, the LP problem can be solved for every incremental value along this range.

The method can be formulated as follows:

$$\text{Max } (v_{jn})$$

$$\text{Subject to } S \cdot v_j = 0 \quad \forall j \in N$$

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

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

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

$$v_j = c_n$$

where c_n is varied in an increment (L) between the minimum and maximum values of the range (x and y, respectively), for example, $c_1 = x$ and $c_2 = y$ with $c_{n+1} = c_n + (y-x)/(L-1)$. The results will generate a series of value for $v\,(v_{jn,}\, n \in [1, L])$ (Palsson, 2006).

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

Chapter 4. Assessment of metabolic capabilities

Model-based analyses have become a particularly important approach for understanding the functions and capabilities of metabolic networks. Many interesting ideas involve the assessment of the consequences of the loss or gain of gene function. Several of these could have clear relevance for biotechnological applications. Here, we present an investigation in which we designed new strains of cyanobacterium that can achieve certain functional state of biotechnological interest.

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

Triana J, Montagud A, Siurana M, Gamermann D, Torres J, Tena J, Fernández de Córdoba P, Urchueguía JF. **Generation and evaluation of a genome-scale metabolic network model of *Synechococcus elongatus* PCC7942**. *Submitted at Metabolites*.

Triana J, Montagud A, Gamermann D, Fernández de Córdoba P, Urchueguía JF. ***In silico* analysis for bio-products synthesis through genome-scale reconstruction of the *Synechococcus elongatus* PCC7942 metabolic network**. *In preparation.*

Chapter 4. Assessment of metabolic capabilities

4.1 Introduction

Since the reconstruction of the metabolic network has become a powerful tool for analysis in systems biology, many researchers have guided their interest in this field targeting economically-relevant goals. Analysis of the genetic modifications in genome-scale metabolic models is one approach to target these projects. In this context, constraint-based methods have been used to indicate mutations that will divert flux towards a preferred designed pathway even in the absence of detailed mechanistic or kinetic data (Bar-Even *et al.*, 2010).

These studies have allowed developing strategies for genetically engineered microbial hosts for enhanced production of various bio-products (Kim *et al*., 2012). There are many examples of successful investigations where microorganisms are used as biological synthesis platform. Some of them have emerged and conducted through *in silico* driven metabolic engineering (Zha *et al.*, 2009; Xu *et al.*, 2011; Ranganathan *et al.*, 2012; Lin *et al.*, 2013).

Despite the fact that the applications of the genome-scale metabolic models were initially found in the implementation of metabolic engineering designs, it is nonetheless true that their profits in other fields have been particularly important. Among these we note understanding of microbial pathogens characteristics and mode of action, many biomedical and pharmaceutical applications, to name a few (Kim *et al.*, 2010, 2011; Duarte *et al.*, 2007; Lewis *et al.*, 2010b).

Genetically modified cyanobacteria are potential biocatalysts for substances production. They possess higher photosynthetic levels and growth rates compared to other algae and higher plants and can be easily manipulated through biotechnological techniques. Their nutritional requirements are very basic, namely, light and CO_2 as the energy and carbon source, respectively, water and mineral salts (especially phosphorous-containing salts) (Quintana *et al.*, 2011). These advantages have suggested the use of these prokaryotes as potential candidates, as an example, for biofuels productions (Rittmann, 2008; Montagud *et al.*, 2013).

After the reconstruction of the metabolic network of *Synechococcus elongatus* PCC7942, we intend to achieve one of the main objectives of this thesis: the *in silico*

design of bio-products cell factory. *Synechococcus elongatus* PCC7942 has become an attractive model to lead CO_2 bioconversion toward the heterologous production of value-added compounds such as: ethanol (Deng and Coleman, 1999), isobutyraldehyde and isobutanol (Atsumi *et al*., 2009) extracellular non-crystalline cellulose (Nobles and Brown, 2008) and bio-hydrogen (Dutta *et al.*, 2005). Although other strategies for improving photochemical efficiencies within the carbon fixation machinery might be possible, many are limited by inadequate knowledge of cyanobacterial biology and metabolic network properties.

The matter of this chapter is the evaluation of the metabolic capabilities of *Synechococcus elongatus* PCC7942. This CO_2 removal system will be the chassis for genetic variations analysis. By which we aim to create a light-based bio-substances production platform that include the third generation renewable energy resources.

4.2 Building and enhancing chemical assignments in metabolic network

Microbial cell as have been used for human purposes since the dawn of time. Biological systems are able to perform many reactions that result in industrially-interesting metabolites. These compounds, whose syntheses go both by simple or multiple chemical transformations, could be the subject for design new metabolic pathways in a particular organism.

The gathering of information related to genes, proteins and reactions, as well as their associations, described for *Synechococcus elongatus* PCC7942 metabolism enables simulations of the metabolic potentialities of this microorganism. This information also allows us to properly evaluate the gain or loss of specific genes in production strategies.

New metabolic phenotypes can be predicted after *in silico* genetic modifications. Nevertheless, hypothetical knowledge to which of these modifications are essential or not for the metabolic network functioning is required.

4.2.1 Gene essentiality analysis

We have used the metabolic reconstruction as a framework to make a comprehensive gene essentiality analysis for every ORF present in the genome annotation. This simulation was carried out, using FBA algorithm, on all single-gene knockouts in models under photoautotrophic condition. To determine the effect of a gene deletion, the reaction(s) associated with each gene in *i*Syf715 were individually deleted from stoichiometric matrix. The simulations were performed using CO_2 uptake rate of 1.99 mmol gDW^{-1} h^{-1} and and irradiance of 1.96 mE gDW^{-1} h^{-1}. Mutant phenotypes were classified as lethal (where there was no growth of the biomass), reduced growth (where growth was lower than that observed in the wild-type) and wild-type growth. In figure 4.1 we have shown the results of this analysis. For comparison purposes, we have also shown the published results for various genome-scale metabolic models.

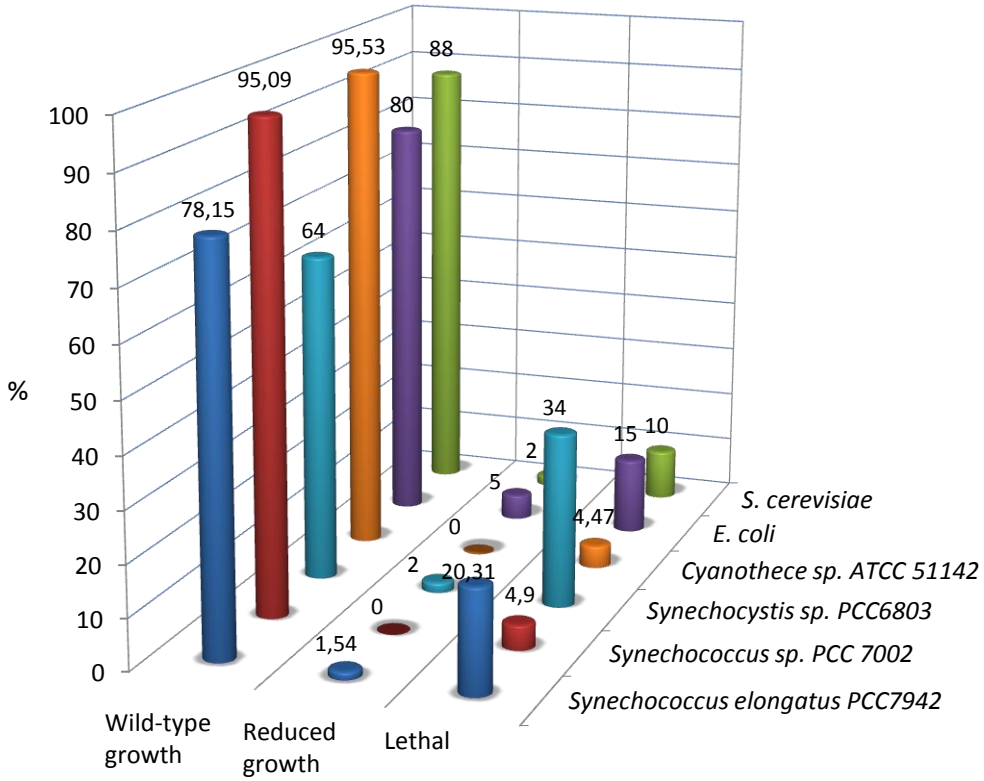


Figure 4.1. Distribution of gene knockout results for six model organisms simulated using FBA algorithm, classified as wild-type growth, reduced growth and lethal (references for each organism in text body).

Single gene knockouts analysis for *Synechococcus elongatus* PCC7942 showed that a core set of 162 reactions were predicted to be essential for growth, necessary for the formation of biomass. Enzymes associated with these reactions are encoded in 145

genes (20.31% of the total) in *i*Syf715. Furthermore, the simulation spots 11 genes that cause a reduced growth rate (1.54% of the total) under the same growth conditions.

As shown in figure 4.1, there are notable differences in the proportion of the essential genes among different organisms. Lethal genes counts among these organisms are as follows: in the metabolic network of *Synechococcus* sp. PCC 7002, 30 genes, 4.9 % of the total (Hamilton and Reed, 2012); *Cyanothece* sp. ATCC 51142, 36 genes, 4.47 % of the total (Vu *et al*., 2012); *Saccharomyces cerevisiae*, 148 genes, 10% of the total (Förster *et al.*, 2003) and *E. coli,* 187 genes, 15% of the total (Feist *et al.*, 2007). Comparatively, our *i*Syf715 model has a significantly larger fraction of essential metabolic genes. This goes in line with published results of *i*Syn669, a metabolic model of *Synechocystis* sp. PCC6803 with counts of 304 genes, 34 % of the total (Montagud *et al.*, 2010). An extensive experimental setup of *S. elongatus* PCC7942 mutants does not currently exist so to validate gene knockout simulations. However, gene essentiality analysis still allows us to generate hypotheses about genes and reactions that are potentially essential under one or more growth conditions. This approach has an intrinsic relationship with genome annotation as well as the growth medium and errors can drive to false positives and false negatives conclusions. These could be one of the main causes of the differences in these microorganisms.

4.2.2 Converting photons and CO_2 into photanol

The long-term availability of fuel fossils has raised great concern in the recent past. This, together with the environmental problems derived from their exploitation, has encouraged an effort focused on sustainable production methods (Milne *et al*., 2011). Presently, one of the more attractive alternatives of energy sources has been the production, by fermentation, of low molecular weight alcohols. There have been many investigations to improve this process. One of the most interesting approaches is a combination of the fermentative and photosynthetic processes on a single biological chassis.

This is based on the ability of photosynthetic microorganisms to produce glyceraldehyde-3-phosphate from CO_2 fixation, as well as the skill of fermentative microorganisms to use C3 sugars to obtain energy and its subsequent transformation to alcohols. This strategy, known as photanol approach, was initially developed for

Synechocystis sp. PCC6803 (Hellingwerf and Mattos, 2009). However, this technology could be theoretically feasible in other photosynthetic biological systems. Basically in those where the techniques of genetic engineering and biotechnology have been extensively tested, like *Synechococcus elongatus* PCC7942. Having this in mind, *i*Syf715 metabolic model was used to simulate its potentialities not only to produce a metabolite of socio-economic interest, but also to identify hot spots where we can improve its productivity.

4.2.2.1 Ethanol

One of the most desired and studied ideas have been the use of photosynthetic mechanisms for bioconversion of CO_2 to ethanol. Nowadays, established technologies for ethanol production use the fermentation of sugars from plant sources, C4 plants mainly, such as corn (Hill *et al.*, 2006), and sugarcane (Goldemberg, 2007). However, this technology is inefficient *per se* mostly because the energy produced is fewer to that the one stored in feedstock (Huber *et al.,* 2006). On the other hand, it is one of the most controversial alternatives due to its competition with food production (Rittmann, 2008).

In this regard, cyanobacteria could be one of the most promising non-food-related biomass sources for energy generation and, in particular, for ethanol synthesis. The first of these microorganisms to be transformed to produce ethanol was *Synechococcus elongatus* PCC7942 (Deng and Coleman, 1999). Genes encoding pyruvate decarboxylase and alcohol dehydrogenase II from *Zymomonas mobilis* were cloned into a vector and used to transform this cyanobacterium.

As an example of the usefulness of the present metabolic models we have designed *in silico* metabolic engineering strategies to improve the production of ethanol. The main purpose was to design a metabolic network (through reaction knockout simulations) which could potentially overproduce ethanol while still evolving some biomass. To this end, we used Minimization of Metabolic Adjustment (MOMA) algorithm (see methods) implemented in OptGene software (Patil *et al*., 2005). MOMA algorithm was developed to calculate the changes in fluxes within the solution space as a result of a gene deletion process (Segrè *et al.*, 2002). It has been reported that this

algorithm provides better description of flux distributions in mutants or under non-natural growth conditions than FBA algorithm.

We simulated single, double and triple knockout strategies in order to find possible genetic targets to increase ethanol production, but without drastically diminishing biomass production. We used *i*Syf715 metabolic model for which wild type optimal growth rate, under photoautotrophic condition, was 0.05987 h^{-1}. In the following figure it can be seen the simulations results for single knockouts.

A

Reaction in model	Gene
2.7.4.3a	adk
1.17.4.1c	nrdb
2.7.7.7b	dnaN
2.7.4.6d	ndkR
2.1.3.2	pyrB
3.5.2.3	pyrC
2.4.2.10	pyrE
4.1.1.23	pyrF
2.7.4.6a	ndkR
Ammonia H3N TRANS-RXN59G-178	amt1

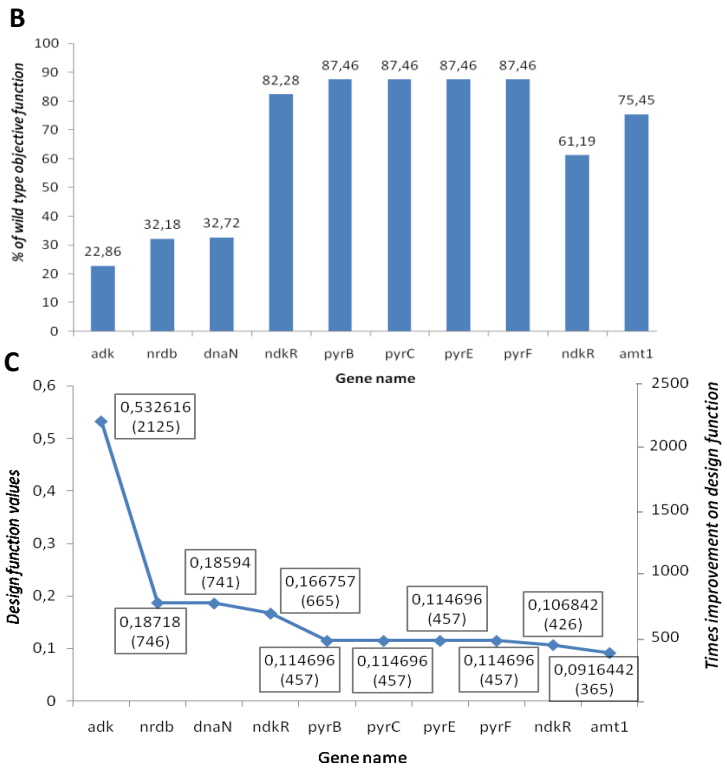


Figure 4.2. Proposed single knockouts for an improved ethanol production. **A**. List of top 10 single knockouts by which they reached a maximum on the *design function* (ethanol production). **B**. Percent depicting the *mutant objective function* (biomass production) relative to the *wild type objective function*. **C**. Values of the design function (first value in the text box) relative to the deleted genes as well as the times improvement on this design function (value enclosed in brackets). Units for *objective function* and *design function* in h^{-1} and mmol gDW^{-1} h^{-1}, respectively.

By deleting the *adk* gene that encodes the enzyme adenylate kinase (reaction *"2.7.4.3a"* in *i*Syf715 and ORF ID: *Synpcc7942_2213*), the biological system is capable of achieving a greater flow towards the production of ethanol, here, 0.532616 mmol gDW^{-1} h^{-1}, with a growth rate of 0.00644061 h^{-1} (Figure 4.2).

By interrupting the flux through this reaction (*2.7.4.3a : AMP + ATP <-> 2 ADP*), the consumption as much the substrates as the products is reduced. This could favor the

acetate synthesis, and therefore the acetaldehyde one, by the reaction catalyzed by acetyl-CoA synthetase from AMP and acetyl-CoA (reaction *"6.2.1.1a"* in *i*Syf715) or by the reaction catalyzed by acetate kinase from ADP and acetylphosphate (reaction *"2.7.2.1a"* in *i*Syf715). On the other hand, by reducing the consumption of ATP the NADH synthesis is favored and therefore the ethanol production.

Second best strategy would be to knockout the gene *nrdb* that codes for a class II ribonucleoside-diphosphate reductase (reaction "*1.17.4.1c"* in *i*Syf715 and ORF ID: *Synpcc7942_1609*). Similarly, it would avoid the dGDP formation and therefore diminish the phosphates groups transfer, catalyzed by nucleoside diphosphate kinase, that uses ATP (reaction *"2.7.4.6d"* in *i*Syf715).

Through double knockout simulation we observed an improved ethanol evolution despite some reduction in the biomass growth (see Figure 4.3). Interestingly, best double deletion strategy involves *adk* gene removal (reaction *"2.7.4.3a"* in *i*Syf715) continuing the results obtained in the single knockout analysis.

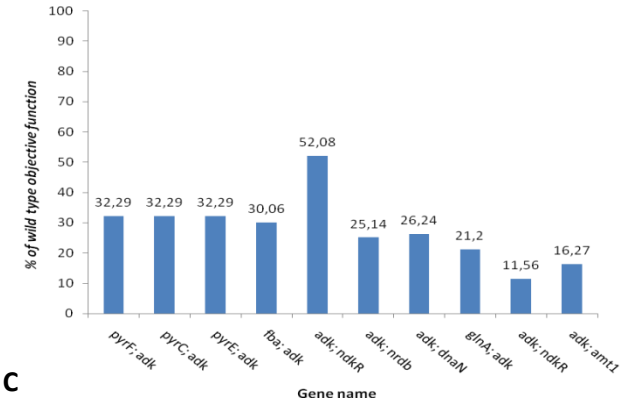
The best double knockout mutant strain for ethanol production would target the gene coding for adenylate kinase and the gene *pyrF* coding for orotidine-5-phosphate decarboxylase (reaction *"4.1.1.23"* in *i*Syf715 and ORF ID: *Synpcc7942_2569*). This mutant would yield an ethanol synthesis rate of 0.633361 mmol gDW^{-1} h^{-1}, with a growth rate of 0.0090926 h^{-1}. The combination of these two deletions involves the reduction of direct and indirect uses of ATP by these two enzymes, respectively. In this way, it would increase the NADH pool and the flux of its oxidation by ethanol synthesis machinery.

Best *in silico* triple knockout strain corresponds to deletions of gene *adk* encoding for adenylate kinase enzyme (reaction *"2.7.4.3a"* in *i*Syf715 and ORF ID: *Synpcc7942_2213*), gene *ndkR* for nucleoside diphosphate kinase enzyme (reaction *"2.7.4.6a"* in *i*Syf715 and ORF ID: *Synpcc7942_2497*) and gene *pyrC* for dihydroorotase enzyme (reaction *"3.5.2.3"* in *i*Syf715 and ORF ID: *Synpcc7942_0486*). This simulated strain has an ethanol production of 0.724116 mmol gDW^{-1} h^{-1} with a growth rate of 0.00724567 h^{-1} (see figure 4.3). Again, these knockouts involve the reduction of the ATP use, that as explained above, the flow towards ethanol synthesis is increased.

A

Pair of reactions	Pair of genes
4.1.1.23; 2.7.4.3a	pyrF; adk
3.5.2.3; 2.7.4.3a	pyrC; adk
2.4.2.10; 2.7.4.3a	pyrE; adk
4.1.2.13a; 2.7.4.3a	fba; adk
2.7.4.3a; 2.7.4.6d	adk; ndkR
2.7.4.3a; 1.17.4.1c	adk; nrdb
2.7.4.3a; 2.7.7.7b	adk; dnaN
6.3.1.2; 2.7.4.3a	glnA; adk
2.7.4.3a; 2.7.4.6a	adk; ndkR
2.7.4.3a; ammonia H3N TRANS-RXN59G-178	adk; amt1

B



C

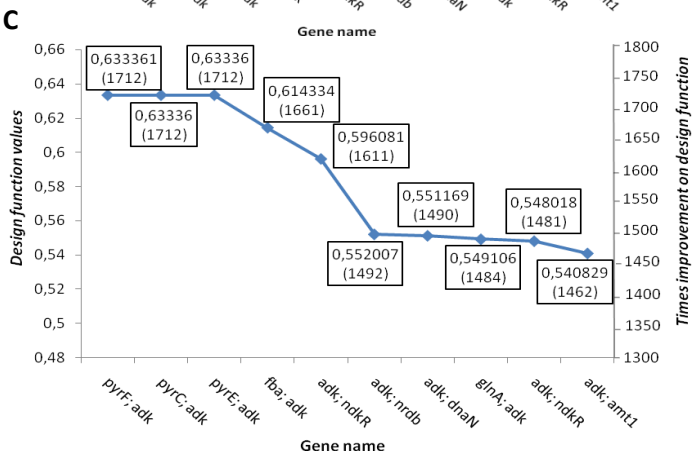


Figure 4.3. Proposed double knockouts for an improved ethanol production. **A**. List of top 10 single knockouts by which they reached a maximum on the *design function* (ethanol production). **B**. Percent depicting the *mutant objective function* (biomass production) relative to the *wild type objective function*. **C**. Values of the design function (first value in the text box) relative to the deleted genes as well as the times improvement on this design function (value enclosed in brackets). Units for *objective function* and *design function* in h^{-1} and mmol gDW^{-1} h^{-1}, respectively.

The graph in figure 4.4 shows the most frequent genetic deletions through which, theoretically, mutant strains would reach a maximum of the objective and design functions. It can be appreciated that the *adk* gene mutations are the most frequent in the phenotypes analyzed, followed by mutations in *ndkR* and *pyrC* genes.

From the engineering point of view, the synthesis of a desired compound has been boosted by manipulating one of the factors that affect the rate of reactions. Many studies have been achieved that through overexpressing the catalytic mechanisms related to its precursors (Duan *et al.*, 2010; Thykaer *et al.*, 2010).

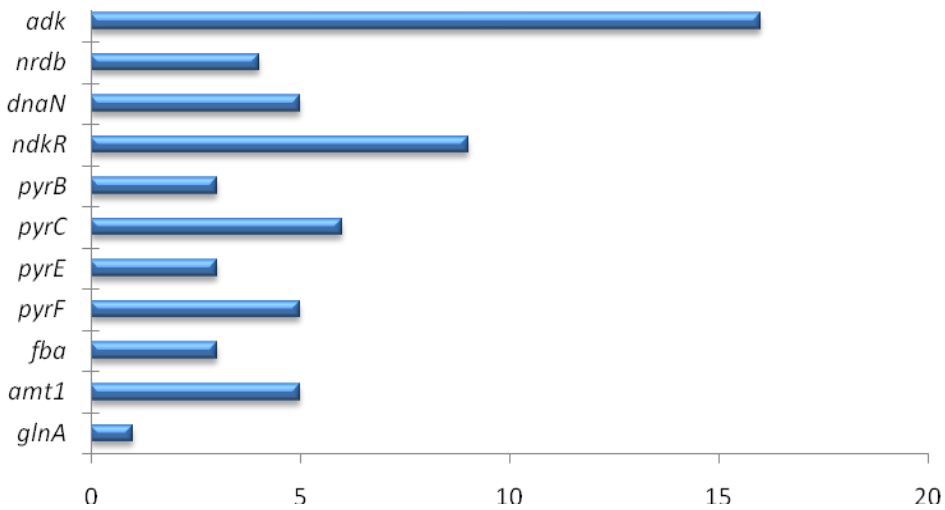


Figure 4.4. Most frequent mutations, proposed by MOMA analysis, which include the top 10 candidates of the single, double and triple knockouts for maximize ethanol production.

In this case, and taking the example of a Michaelis-Menten reaction, when the enzymes concentration becomes high enough to entirely transform the substrate to the enzyme-substrate complex, the second step of the reaction becomes rate limiting, and overall reaction rate becomes sensitive to further increases in enzymes concentration. This was applied in our steady-state system modeling. Thus, with the aim of focusing on other metabolic designs, we made sure that the rate of alcohol dehydrogenase (reaction *"1.1.1.1"* in *i*Syf715) was only restricted by its thermodynamical capabilities. Then, we added the missing gene encoding for the pyruvate decarboxylase enzyme (reaction *"·4.1.1.1"* in *i*Syf715), and again, we made sure that the rate was only restricted by its thermodynamical capabilities. This protein catalyze the directly conversion of pyruvate, considered a hubs within metabolic network, into acetaldehyde.

The estimation of theoretical maximum yield of ethanol production was performed by FBA and compared with results reported by Deng and Coleman in a recombinant *Synechococcus elongatus* PCC7942 (Deng and Coleman, 1999). The transformed cells by these authors were grown in 500 mL batch cultures and the specific ethanol synthesis rate was 0.2 mg L^{-1} d^{-1} (≈ $5.8 \cdot 10^{-9}$ mmol gDW^{-1} h^{-1}). We found that for CO_2 uptake rate of 1.99 mmol gDW^{-1} h^{-1} and an incident irradiance of 1.96 mE gDW^{-1} h^{-1}, our prediction for theoretical productivity of ethanol behaves as shown in the graph in Figure 4.2. For the maximum biomass growth (0.05987 h^{-1}), the

theoretical synthesis of ethanol was $3.81 \cdot 10^{-8}$ mmol gDW^{-1} h^{-1} which is not far from the experimental report. On the other hand, the *in silico* productivity gives us an idea of the metabolic capabilities of this cyanobacterium when the carbon flux is directed towards ethanol production (see figure 4.5).

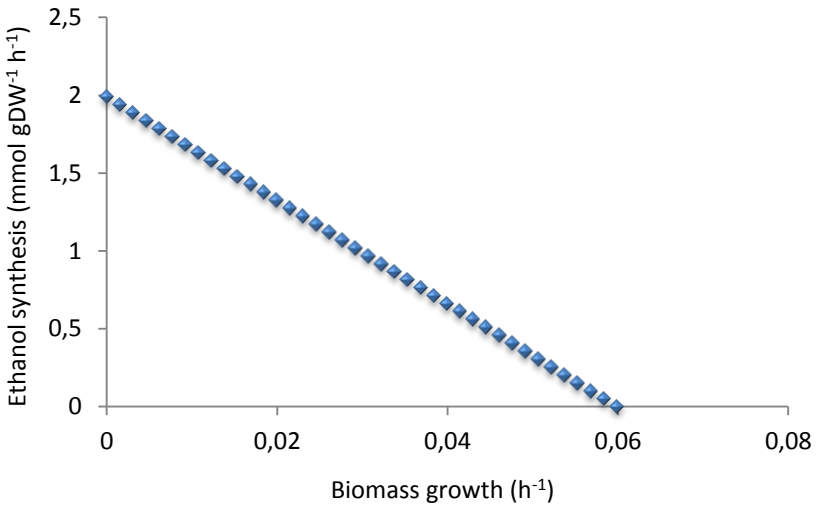


Figure 4.5. Theoretical productivity of ethanol as predicted by using *i*Syf715. Maximum ethanol production is shown as a function of minimal demand on biomass formation under autotrophic growth.

As we can see the maximal value for ethanol synthesis rate derived by *i*Syf715 was predicted to be larger when the biomass growth is minimal. The left part of the graph shows the behavior of a strain whose energy and reduced power are only intended to synthesize ethanol. While the right part represent the wild strain, which does not produce ethanol, since it is not described in the biomass equation. The graph can be used to estimate the theoretical maximum production, and therefore to organize the mutants previously simulated. It is not always interesting designing a strategy whose objectives are the production of ethanol with a low biomass growth. Usually, a compromise solution in the middle of that line will be the desired objective.

4.2.2.2 Higher chain alcohols

Despite that ethanol can be mixed with existing fuels and current engines can use it without any modification (Kaygusuz, 2009), this biofuel has several drawbacks: low energy density, high vapor pressure, and high hygroscopicity (Smith *et al*., 2010). These problems have led us to focus on other alcohols that present high energy content, low

volatility and low corrosivity, as well as being less hydroscopic. In this regard, higher alcohols, such as the isomers of n-butanol seem to be promising candidates (Cheng *et al*., 2012; Baez *et al*., 2011; Li *et al.*, 2011; Borden and Papoutsakis 2007; Mariano *et al*., 2009; Atsumi *et al*. 2008, 2009a, b; Shen and Liao 2008; Cann and Liao 2008; Connor and Liao 2008; Smith *et al.*, 2010). Some of them, like isobutanol and even the n-butanol, constitute the best targets as potential additives or substitutes of gasoline because of their multiple advantages. Nowadays current butanol demand is fulfilled by chemical synthesis, which requires high production cost and expensive catalyst (Carlini *et al.*, 2003). In fact, higher alcohols such as isobutanol, 2-methyl-1-butanol, 3-methyl-1-butanol and 1-propanol are rarely synthesized by biological systems even though small amounts have been detected as microbial by-products (Sentheshanuganathan, 1960; Dickinson *et al.*, 1997; 1998; 2000; 2003).

For successful *in situ* synthesis of these alcohols, and so as to prevent accumulation of heterologous metabolites and hence cytotoxicity (Pitera *et al.*, 2007), we looked for pathways that shared common elements with our organism. Non-fermentative pathways such as the Ehrlich pathway (Sentheshanmuganathan, 1958) has been one of the targets for this strategies. The idea is to alter the metabolic fluxes of the transformation of 2-keto acids, which are intermediates in amino acid biosynthesis pathways, to enhance the synthesis of higher alcohols. These types of carboxylic acids can be dehydrogenated to alcohols and reduced to aldehydes, by the action of alcohol dehydrogenases and 2-keto acid decarboxylases, which are not distinctive of prokaryotes (Konig, 1998). Thus, we added this new gene ("*kivd"* and reaction *"·4.1.1.74"* in *i*Syf715) as a heterologous pathway to produce biofuels.

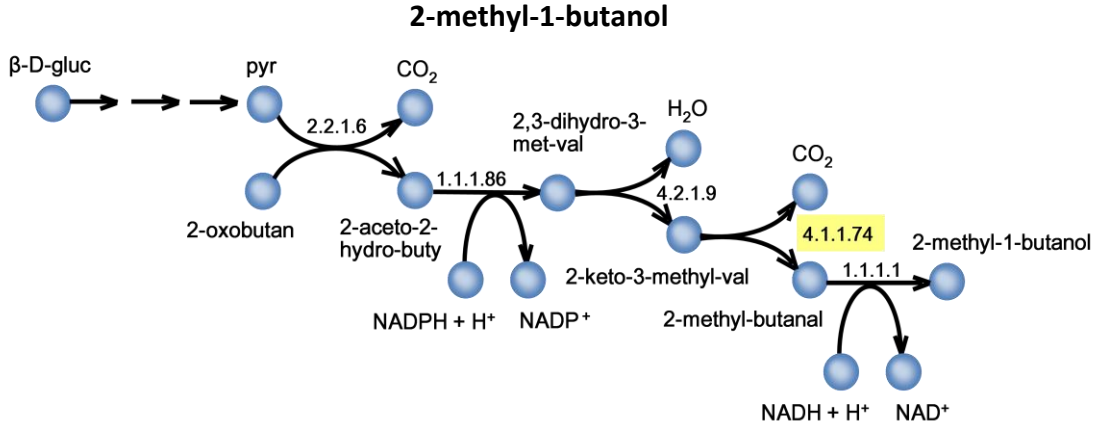
We decided to perform simulations of different alcohols productions with some 2-keto acids as precursors. For instance, 2-keto-3-methyl-valerate and 2-ketoisocaproate, which are intermediates of isoleucine and leucine biosynthesis pathway, respectively, can be converted to 2-methyl-1-butanol and 3-methyl-1-butanol, respectively. The valine biosynthesis pathway produces 2-keto-isovalerate, which is the precursor for isobutanol. Finally, we decided to simulate 1-propanol synthesis whose substrate is 2-ketobutyrate in isoleucine biosynthesis pathway. Although some of the 2-keto-acid decarboxylases (KDCs) have substrate-specific

activity (Atsumi *et al.*, 2008), all of these enzymes were modeled as one, having a wide substrate range.

We simulated single, double and triple knockout strategies, using MOMA algorithm implemented in OptGene software (Patil *et al.*, 2005), in order to find possible genetic targets to increase alcohols production. More specifically, we evaluated mutant phenotypes that maximize the production of alcohol. We used *i*Syf715 metabolic model whose wild type growth rate is 0.05987 h^{-1}.

The following table shows single knockout strains result of MOMA analyses, maximizing the synthesis of higher alcohols.

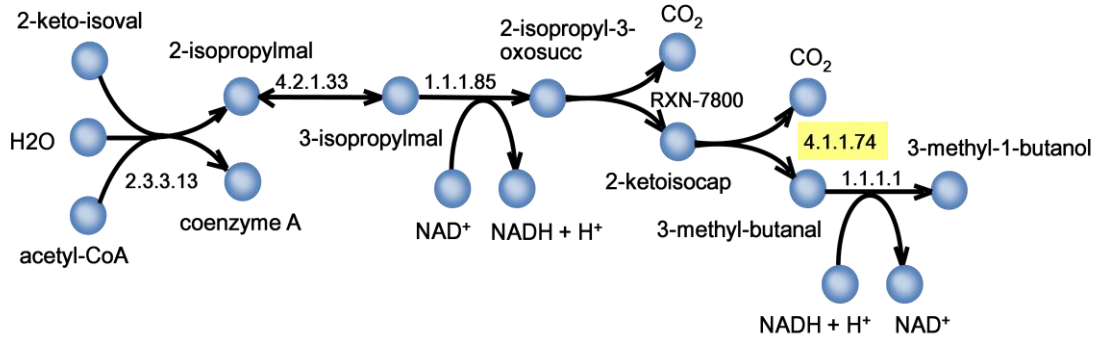
Table IV.A. Proposed single knockouts for an improved higher alcohols production. Units for *objective function* (biomass production) in h^{-1} and for *design function* (higher alcohols) in mmol gDW^{-1} h^{-1}. In all pathways, the knock-in is highlighted in yellow.



Abbreviation: β-D-gluc = β-D-glucose; pyr = pyruvate; 2-oxobutan = 2-oxobutanoate; 2-aceto-2-hydro-buty = 2-aceto-2-hydroxy-butyrate; 2,3-dihydro-3-met- val = 2,3-dihydroxy-3-methylvalerate; 2-keto-3-methyl-val = 2-keto-3-methyl-valerate.

Reaction in model	***Gene***	***% of wild type objective function***	***Design function***	***Times improvement on design function***
4.1.2.13a	*fba*	38,56	0,0783	2260
2.1.3.2	*pyrB*	80,74	0,0623	1798
3.5.2.3	*pyrC*	80,74	0,0623	1798
2.4.2.10	*pyrE*	80,74	0,0623	1798
4.1.1.23	*pyrF*	80,74	0,0623	1798
4.2.1.1b	*icfA/ecaA*	26,21	0,0551	1591
ammonia H3N TRANS-RXN59G-178	*amt1*	69,16	0,0447	1290
2.7.4.6d	*ndkR*	80,35	0,0273	788
1.4.1.1	*ald*	74,21	0,0220	635
2.7.7.6c	*rpoB*	91,59	0,0158	456

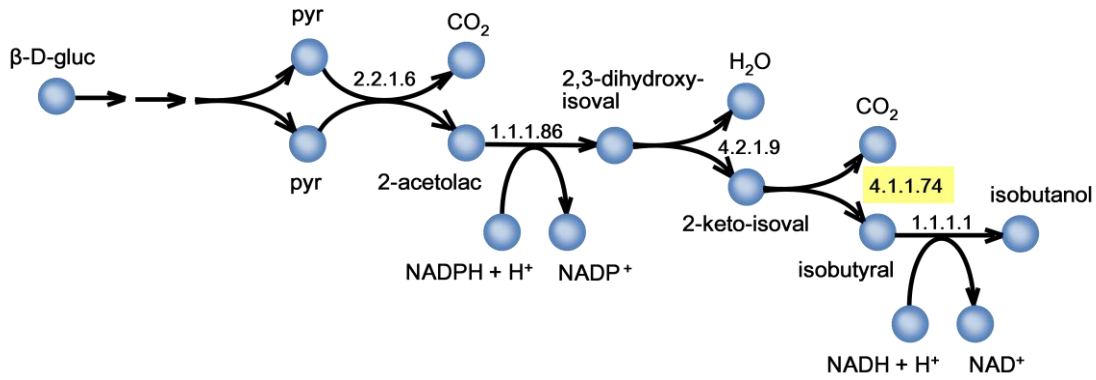
3-methyl-1-butanol



Abbreviation: 2-keto-isoval = 2-keto-isovalerate; 2-isopropylmal = 2-isopropylmalate; 3-isopropylmal = 3-isopropylmalate; 2-isopropyl-3-oxosucc = 2-isopropyl-3-oxosuccinate; 2-ketoisocap = 2-ketoisocaproate.

Reaction in model	*Gene*	*% of wild type objective function*	*Design function*	*Times improvement on design function*
4.2.1.1b	*icfA/ecaA*	16,18	0,0914	1589
2.7.4.6d	*ndkR*	67,92	0,0816	1418
1.17.4.1c	*nrdb*	20,69	0,0634	1102
2.7.7.7b	*dnaN*	21,44	0,0623	1083
ammonia H3N TRANS RXN59G-178	*amt1*	67,09	0,0613	1065
4.1.2.13a	*fba*	17,52	0,052	904
3.5.2.3	*pyrC*	82,17	0,0518	900
2.4.2.10	*pyrE*	82,17	0,0518	900
2.1.3.2	*pyrB*	82,17	0,0518	900
4.1.1.23	*pyrF*	82,17	0,0518	900

Isobutanol

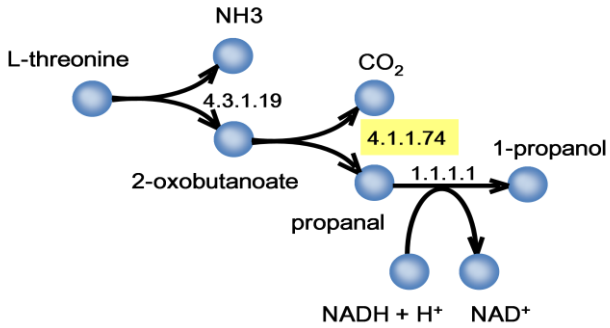


Abbreviation: β-D-gluc = β-D-glucose; pyr = pyruvate; 2-acetolac = 2-acetolactate; 2,3-dihydroxy-isoval = 2,3-dihydroxy-isovalerate; 2-keto-isoval = 2-keto-isovalerate; isobutyral = isobutyraldehyde.

Reaction in model	*Gene*	*% of wild type objective function*	*Design function*	*Times improvement on design function*
2.7.4.6d	*ndkR*	67,42	0,129	1304
4.1.2.13a	*fba*	31,21	0,125	1264
4.2.1.1b	*icfA/ecaA*	18,95	0,112	1132

ammonia H3N TRANS-RXN59G-178	*amt1*	67,65	0,0845	854
3.5.2.3	*pyrC*	80,73	0,0776	784
2.4.2.10	*pyrE*	80,73	0,0776	784
2.1.3.2	*pyrB*	80,73	0,0776	784
4.1.1.23	*pyrF*	80,73	0,0776	784
4.1.1.31	*ppc*	45,3	0,0756	764
1.17.4.1c	*nrdb*	26,12	0,0440	444

1-propanol



Reaction in model	***Gene***	***% of wild type objective function***	***Design function***	***Times improvement on design function***
2.1.3.2	*pyrB*	83,75	0,0532	1532
4.1.1.23	*pyrF*	83,75	0,0532	1532
3.5.2.3	*pyrC*	83,75	0,0532	1532
2.4.2.10	*pyrE*	83,75	0,0532	1532
4.1.2.13a	*fba*	23,75	0,0374	1077
1.4.1.1	*ald*	74,77	0,0344	991
4.2.1.1b	*icfA/ecaA*	32,63	0,0335	965
ammonia H3N TRANS-RXN59G-178	*amt1*	72,25	0,0319	919
4.2.1.2	*fum*	86,83	0,03	864
4.3.2.2a	*purB*	90,82	0,0205	590

In the case of 2-methyl-1-butanol, the simulation results indicate that by deleting *fba* gene, which encodes for an aldolase (reaction *"4.1.2.13a"* in *i*Syf715 and ORF ID: *Synpcc7942_1443*) has maximal synthesis of this alcohol. Our model achieves an alcohol flux production of 0.0783 mmol gDW^{-1} h^{-1}, with a growth rate of 0.0108696 h^{-1}. This is a key enzyme in the glycolytic pathway since it cleaves an aldol, through a retro-aldol condensation, that initiates the second phase of this path, in which triose

phosphates complete its oxidation. Additionally, this reaction is common to the gluconeogenesis pathway, an active route in the *Synechococcus elongatus* PCC7942 photosynthetic metabolism. By blocking this step the metabolic flux in gluconeogenic direction is prevented and the 3-phosphoglycerate from the Calvin-Benson cycle will be directed towards the pyruvate formation, one of the precursors of 2-methyl-1-butanol. The second best mutation targets to the mutation of *pyrB* gene that encodes for aspartate transcarbamylase (ATCase) (reaction *"2.1.3.2"* in *i*Syf715 and ORF ID: *Synpcc7942_0670*). Preventing the flux through this reaction, the L-aspartate remains available as substrate for other reactions, such as the catalyzed by aspartate kinase (reaction *"2.7.2.4"* in *i*Syf715) which is part of a set of reactions that transform this amino acid to L-threonine. In turn, the L-threonine is deaminated to 2-oxobutanoate (reaction *"4.3.1.19a"* in *i*Syf715) which is a precursor for the 2-methyl-1-butanol synthesis.

Analyzing the singles knockout for 3-methyl-1-butanol, we can see that the best single strategy corresponds to the deletion of the *icfA/ecaA* genes encoding for carbonic anhydrase enzyme (reactions *"4.2.1.1b"* in *i*Syf715 and ORF IDs: *Synpcc7942_1447, Synpcc7942_1388 and Synpcc7942_B2619*). As shown in Table IV.A, the model is capable of producing a flux of 0.0914 mmol gDW^{-1} h^{-1} of alcohol while diminishing growth to 0.00455735 h^{-1}. These deletions prevent the dehydration of the bicarbonate ion (HCO_3^-) and its conversion into CO_2. Thus, the CO_2 fixation by Calvin - Benson cycle is decreased and the carbon flux is redirected towards the production of this biofuel. This carbon diversion can explain the high rate of its synthesis as well as the low biomass growth rate. Furthermore, the second best strategy consists on *ndkR* gene mutation encoding for nucleoside diphosphate kinase (reaction *"2.7.4.6d"* in *i*Syf715 and ORF ID: *Synpcc7942_2497*). This mutation prevents the formation of dGTP and avoids its reaction with the pyruvate (reaction *"2.7.1.40d"* in *i*Syf715). This way, pyruvate remains available to be transformed into 2-keto-isovalerate, one of precursors of 3-methyl-1-butanol.

As in the previous case, the mutant phenotype that has the highest flux for isobutanol synthesis is the one without *ndkR, a* gene coding for nucleoside diphosphate kinase (reaction *"2.7.4.6d"* in *i*Syf715 and ORF ID: *Synpcc7942_2497*). As

we explained further above, this mutation could increase the 2-keto-isovalerate formation rate. In this case, the model is capable of producing 0.129 mmol gDW^{-1} h^{-1} of alcohol with a growth rate of 0.0189957 h^{-1}. Besides, the second best strategy involves the *fba* gene deletion, a gene coding for aldolase enzyme (reaction *"4.1.2.13a"* in *i*Syf715 and ORF ID: *Synpcc7942_1443*). This strategy redirects the flow to the end of the glycolytic pathway. This mutation could increase the level of pyruvate, substrate of the 2-keto-isovalerate formation and therefore of isobutanol.

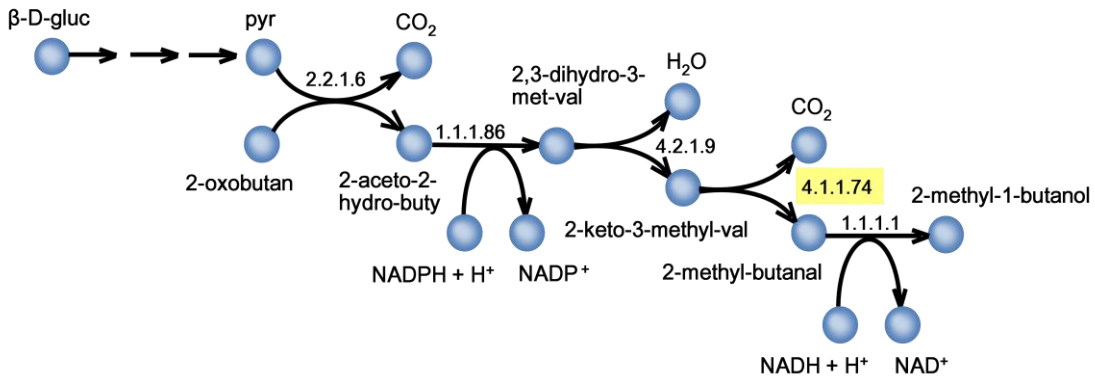
We have seen that higher flux in 1-propanol synthesis is reached when blocking the flux through the reactions catalyzed by either ATCase, orotidine-5-phosphate decarboxylase, dihydroorotase or orotate phosphoribosyltransferase and encoded by *pyrB, pyrF, pyrC* and *pyrE,* respectively (reactions "*2.1.3.2", "4.1.1.23", "3.5.2.3" and "2.4.2.10",* respectively*,* in *i*Syf715 and ORF ID: *Synpcc7942_0670, Synpcc7942_2569, Synpcc7942_0486* and *Synpcc7942_2592*). In particular the prokaryote is capable of achieving a flux of 0.0532 mmol gDW^{-1} h^{-1}, with a growth rate of 0.0235939 h^{-1}. As we explained, the *pyrB* deletion increases the conversion flux of L-aspartate to L-threonine and this into 2-oxobutanoate that is a precursor of 1-propanol. Additionally, by eliminating the activity of these enzymes, which are strongly coupled to pyrimidine biosynthesis pathway, the decarboxylation of orotidine-5'-phosphate and its CO_2 release are prevented. As a consequence, the gluconeogenic flux would be diminished and flux towards the production of this biofuel would be favored.

We also looked into combination of double knockouts. The following table shows the best double mutant strain result of MOMA analysis, in order to maximize the production of higher alcohols.

As in the case of ethanol, we observed improved alcohols synthesis with these double knockouts than with the single ones (see Table). Interestingly, the 10 best double deletions strategies involve *amt1* genes removal (reaction *"ammonia H3N TRANS-RXN59G-178"* in *i*Syf715 and ORF ID: *Synpcc7942_0442* and *Synpcc7942_2279*). This is in line with the results presented in above single knockout section.

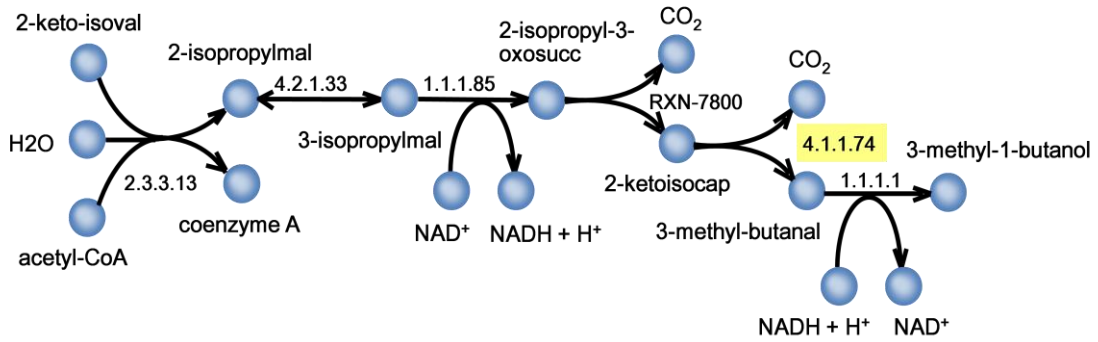
Table IV.B. Proposed double knockouts for an improved higher alcohols production. Units for *objective function* (biomass production) in h^{-1} and for *design function* (higher alcohols) in mmol gDW^{-1} h^{-1}. In all pathways, the knock-in is highlighted in yellow.

2-methyl-1-butanol



Abbreviation: β-D-gluc = β-D-glucose; pyr = pyruvate; 2-oxobutan = 2-oxobutanoate; 2-aceto-2-hydro-buty = 2-aceto-2-hydroxy-butyrate; 2,3-dihydro-3-met- val = 2,3-dihydroxy-3-methylvalerate; 2-keto-3-methyl-val = 2-keto-3-methyl-valerate.

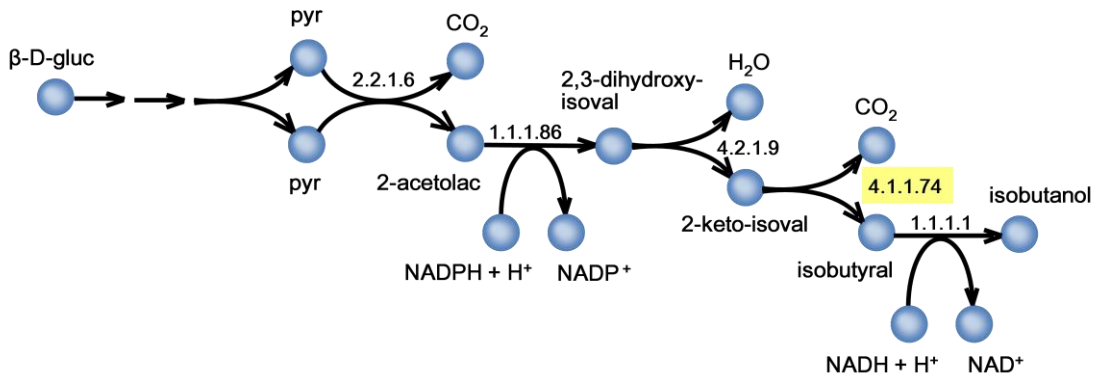
Pair of reactions	***Pair of genes***	***% of wild type objective function***	***Design function***	***Times improvement on design function***
3.5.2.3; ammonia H3N TRANS-RXN59G-178	*pyrC; amt1*	65,66	0,122	4846
2.4.2.10; ammonia H3N TRANS-RXN59G-178	*pyrE; amt1*	65,66	0,122	4846
2.1.3.2; ammonia H3N TRANS-RXN59G-178	*pyrB; amt1*	65,66	0,122	4846
4.1.1.23; ammonia H3N TRANS-RXN59G-178	*pyrF; amt1*	65,66	0,122	4846
3.5.2.3; _1.4.4.2	*pyrC;* gcvP	74,66	0,107	4250
3.5.2.3; _2.1.2.10a	*pyrC;* gcvT	74,66	0,107	4250
3.5.2.3; _1.8.1.4c	*pyrC; phdD*	74,66	0,107	4250
2.4.2.10; _1.4.4.2	*pyrE; gcvP*	74,66	0,107	4250
2.4.2.10; _2.1.2.10a	*pyrE; gcvT*	74,66	0,107	4250
2.4.2.10; _1.8.1.4c	*pyrE; phdD*	74,66	0,107	4250
2.1.3.2; _1.4.4.2	*pyrB; gcvP*	74,66	0,107	4250
2.1.3.2; _2.1.2.10a	*pyrB; gcvT*	74,66	0,107	4250
2.1.3.2; _1.8.1.4c	*pyrB; phdD*	74,66	0,107	4250
4.1.1.23; _1.4.4.2	*pyrF; gcvP*	74,66	0,107	4250
4.1.1.23; _2.1.2.10a	*pyrF; gcvT*	74,66	0,107	4250
4.1.1.23; _1.8.1.4c	*pyrF; phdD*	74,66	0,107	4250
3-methyl-1-butanol				



Abbreviation: 2-keto-isoval = 2-keto-isovalerate; 2-isopropylmal = 2-isopropylmalate; 3-isopropylmal = 3-isopropylmalate; 2-isopropyl-3-oxosucc = 2-isopropyl-3-oxosuccinate; 2-ketoisocap = 2-ketoisocaproate.

Pair of reactions	*Pair of genes*	*% of wild type objective function*	*Design function*	*Times improvement on design function*
1.17.4.1c; ammonia H3N TRANS-RXN59G-178	*nrdb; amt1*	11,22	0,138	1056
2.7.7.7b; ammonia H3N TRANS-RXN59G-178	*dnaN; amt1*	12,75	0,1369	1048
2.1.3.2; ammonia H3N TRANS-RXN59G-178	*pyrB; amt1*	65,06	0,1337	1023
4.1.1.23; ammonia H3N TRANS-RXN59G-178	*pyrF; amt1*	65,06	0,1337	1023
3.5.2.3; ammonia H3N TRANS-RXN59G-178	*pyrC; amt1*	65,06	0,1337	1023
2.4.2.10; ammonia H3N TRANS-RXN59G-178	*pyrE; amt1*	65,06	0,1337	1023
2.2.1.2; 2.7.4.6d	*tal; ndkR*	41,05	0,1251	957
2.7.4.6d; ammonia H3N TRANS-RXN59G-178	*ndkR; amt1*	57,07	0,125	957
2.7.9.2; 2.7.4.6d	*ppsA; ndkR*	65,83	0,1205	922
4.2.1.1b; ammonia H3N TRANS-RXN59G-178	*icfA/ecaA; amt1*	10,25	0,1178	901

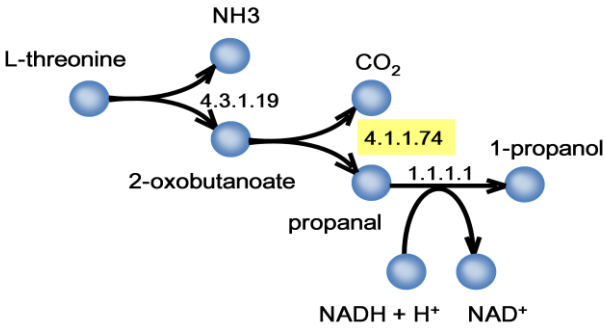
Isobutanol



Abbreviation: β-D-gluc = β-D-glucose; pyr = pyruvate; 2-acetolac = 2-acetolactate; 2,3-dihydroxy-isoval = 2,3-dihydroxy-isovalerate; 2-keto-isoval = 2-keto-isovalerate; isobutyral = isobutyraldehyde.

Pair of reactions	Pair of genes	% of wild type objective function	Design function	Times improvement on design function
3.6.1.1; 1.2.1.3c	ppa; puuC	2,39	0,36233	3418
4.1.2.13a; 4.1.1.31	fba; ppc	18,09	0,25385	2394
4.1.2.13a; ammonia H3N TRANS-RXN59G-178	fba; amt1	16,46	0,232886	2196
2.7.1.146; 2.7.4.6d	fba; ndkR	18,39	0,214276	2021
4.1.2.13a; 1.2.1.3c	fba; puuC	18,39	0,212528	2004
2.7.4.6d; 4.2.1.1b	ndkR; icfA/ecaA	5,0	0,205815	1941
2.7.1.146; 1.2.1.3c	pfk; puuC	28,46	0,202854	1613
4.1.2.13a; 1.2.4.1	fba; pdh	9,92	0,201563	1901
4.1.2.13a; 2.3.1.12	fba; aceF	9,92	0,201563	1901
4.1.2.13a; 1.8.1.4a	fba; phdD	9,924	0,20155	1901
2.7.4.6d; 1.2.4.1	ndkR; pdh	58,95	0,19966	1883
2.7.4.6d; 2.3.1.12	ndkR; aceF	58,95	0,19966	1883
2.7.4.6d; 1.8.1.4a	ndkR; phdD	58,95	0,19966	1883
2.7.9.2; 2.7.4.6d	ppsA; ndkR	61,9	0,195749	1846

1-propanol



Pair of reactions	Pair of genes	% of wild type objective function	Design function	Times improvement on design function
2.2.1.2; 4.2.1.1b	tal; icfA/ecaA	7,18	0,074099	1262
4.2.1.1b; ammonia H3N TRANS-RXN59G-178	icfA/ecaA; amt1	23,96	0,071594	1220
2.1.3.2; 1.2.1.3c	pyrB; puuC	86,64	0,071279	1214
4.1.1.23; 1.2.1.3c	pyrF; puuC	86,64	0,071279	1214
3.5.2.3; 1.2.1.3c	pyrC; puuC	86,64	0,071278	1214
2.4.2.10; 1.2.1.3c	pyrE; puuC	86,64	0,071278	1214

2.7.7.8d; ammonia H3N TRANS-RXN59G-178	*pnp; amt1*	48,27	0,069210	1179
4.1.2.13b; 1.2.1.3c	*fba; puuC*	85,72	0,068318	1164
3.1.3.37; 1.2.1.3c	*glpX; puuC*	85,72	0,068318	1164
3.5.2.3; 2.7.7.6c	*pyrC; rpoB*	80,98	0,068025	1159
2.4.2.10; 2.7.7.6c	*pyrE; rpoB*	80,98	0,068025	1159

In the case of 2-methyl-1-butanol, there are four candidates that are good double knockouts strategies that evolve same alcohol flux and same biomass production. The four involve the elimination of *amt1* genes combined with the deletions in the *pyrC, pyrE, pyrB and pyrF* genes *(*reaction "*3.5.2.3", "2.4.2.10"*, "*2.1.3.2*" and "*4.1.1.23*" in *i*Syf715, respectively, and ORF ID: *Synpcc7942_0486, Synpcc7942_0488, Synpcc7942_2592, Synpcc7942_0670* and *Synpcc7942_2569*). In all cases, we predicted 2-methyl-1-butanol synthesis rate of 0.122 mmol gDW^{-1} h^{-1}, with a growth rate of 0.0185017 h^{-1}.

As we explained before the mutation of these coupled genes have the same effect, over the orotidine-5'-phosphate decarboxylation. The removal of *amt1* genes would affect the ammonium input and thus also affects the back flow of the reaction *"1.4.1.1"*, which involves the use of pyruvate to form L-alanine, and the flow of the reaction *"3.5.1.1"* which includes the use of L-aspartate, among other reactions where ammonium is used. These changes would mean higher availability of both L-aspartate and pyruvate, precursor of higher alcohol.

The best double knockout for the 3-methyl-1-butanol production, involves the combined deletions of *nrdb* gene, which encodes ribonucleoside diphosphate reductase-class II (reaction *"1.17.4.1c"* in *i*Syf715 and ORF ID: *Synpcc7942_1609*) and *amt1* genes. With these deletions, the system is capable of reaching a maximum of 3-methyl-1-butanol synthesis of 0.138 mmol gDW^{-1} h^{-1}, with a growth rate of 0.0031591 h^{-1}. As was previously mentioned, the *amt1* knockouts would increase pyruvate levels as this metabolite is also a precursor of 2-keto-isovalerate. Whereas if *nrdb* gene is mutated the dGDP is not produced (in simulation, the flux of this reaction is in reverse direction). That's why this mutation may impede the dGDP´s phosphorilation into dGTP

(reaction *"2.7.4.6d"* in *i*Syf715) and in turn react with the pyruvate (reaction *"2.7.1.40d"* in *i*Syf715).

For isobutanol, the best strategy would be to knockout *ppa* gene that codifies for inorganic diphosphatase (reaction *"3.6.1.1"* in *i*Syf715 and ORF ID: *Synpcc7942_1383*) and *puuC* gene that codifies for aldehyde dehydrogenase (NAD^+) (reaction *"1.2.1.3c"* in *i*Syf715 and ORF ID: *Synpcc7942_0489*). Through this strategy the model is capable of reaching a maximum isobutanol production of 0.36233 mmol gDW^{-1} h^{-1}, with a growth rate of 0.000673601 h^{-1}. The first mutation would avoid the hydrolysis of diphosphate and therefore diminish the phosphates concentration required in reactions such as the one catalyzed by fructose-6-phosphate phosphoketolase (reaction *"4.1.2.22"* in *i*Syf715). It would also affect the carbon fixation machinery through the Calvin-Benson cycle redirecting more flux to the alcohol synthesis. This explains the low growth rate if we compare it to the other proposed strategies. Moreover, by removing the aldehyde dehydrogenase, NADH consumption is avoided, which is an indispensable cofactor for reducing isobutyraldehyde to isobutanol.

In the case of 1-propanol, the mutations retrieved by the analysis are *tal* and *icfA*/*ecaA* genes that code for transaldolase and carbonic anhydrase, respectively (reactions *"2.2.1.2"* and *"4.2.1.1b"* in *i*Syf715, respectively, and ORF ID: *Synpcc7942_2297*, *Synpcc7942_1447*, *Synpcc7942_1388* and *Synpcc7942_B2619*). With this mutant phenotype the cyanobacterium is capable of reaching an alcohol production of 0.0740992 gDW $mmol^{-1}$ h^{-1}, with a growth rate of 0.0020217 h^{-1}. The transaldolase deletion prevents the synthesis of D-glyceraldehyde-3-phosphate and D-sedoheptulose-7-phosphate in pentose phosphate pathway, and that these in turn react and form D-ribose-5-phosphate (reaction *"2.2.1.1a"* in *i*Syf715). This metabolite is the precursor of D-ribulose-5-phosphate, which is a key compound in the Calvin-Benson cycle. Thereby, this strategy would cut the flux to those reactions and metabolic pathways associated with cell growth. Similarly, and as we previously mentioned, if the genes for carbonic anhydrase are removed the CO_2 formation rate decreases as well as its fixation by the Calvin-Benson cycle. Thus, the carbon flux is redistributed to form 1-propanol. Once more, these mutations explain the predicted low growth for this organism.

We looked for possible triple knockouts that improved the production of each of these higher alcohols. In particular, we saw that 2-methyl-1-butanol production could be optimized by eliminating *pyrC, pyrB* and *amt1* genes giving production titters of 0.28 mmol gDW^{-1} h^{-1}, with a growth rate of 0.01996 h^{-1}. The impact of the elimination of each of these genes has been previously explained.

In the case of 3-methyl-1-butanol, the best triple knockout involves genes *ndkR, amt1 and icfA*/*ecaA*. The predicted synthesis of this alcohol is 0.189 mmol gDW^{-1} h^{-1}, with a growth rate of 0.010991 h^{-1}. Once again, the effect of these deletions on metabolic phenotype has been explained further above.

In the case of isobutanol, the best strategy includes the elimination of *fba, puuC* and *ppsA* genes, the latter encoding for pyruvate-water dikinase (reaction *"2.7.9.2"* in *i*Syf715 and ORF ID: *Synpcc7942_0781*). We predicted that the microorganism is capable of producing 0.4096 mmol gDW^{-1} h^{-1}, with a growth rate of 0.00938 h^{-1}. The involvement of the mutations in *fba* and *puuC* genes in *Synechococcus elongatus* PCC7942 metabolism has been previously explained, while for *ppsA* gene this effect relates to the possibility of increasing the pyruvate levels as precursor of this alcohol.

Finally, when we analyzed the triple mutant to enhance the 1-propanol synthesis, we identified genes *pyrC, icfA/ecaA* and *amt1*. Through this mutant phenotype, the cyanobacterium is capable of producing 0.0901 mmol gDW^{-1} h^{-1}, with a growth rate of 0.00962 h^{-1}. The effects of these knockouts have been described above.

Figure 4.6 shows the most frequent genetic deletions through which mutant strains would reach a maximum of alcohol production and cell growth. As it can be appreciated, the mutations *fba, pyrB, pyrC, pyrE*, *pyrF*, *icfA/ecaA, amt1* and *ndkR* genes are common in all the analyzed phenotypes and the two last genes stand out as the most frequent.

As in the ethanol case, we performed FBA analysis for estimate a theoretical maximum yield of all alcohols production. We considered a CO_2 uptake rate of 1.99 mmol gDW^{-1} h^{-1} and an incident irradiance of 1.96 mE gDW^{-1} h^{-1}. Our prediction for theoretical productivity of these alcohols behaves as shown in figure 4.7.

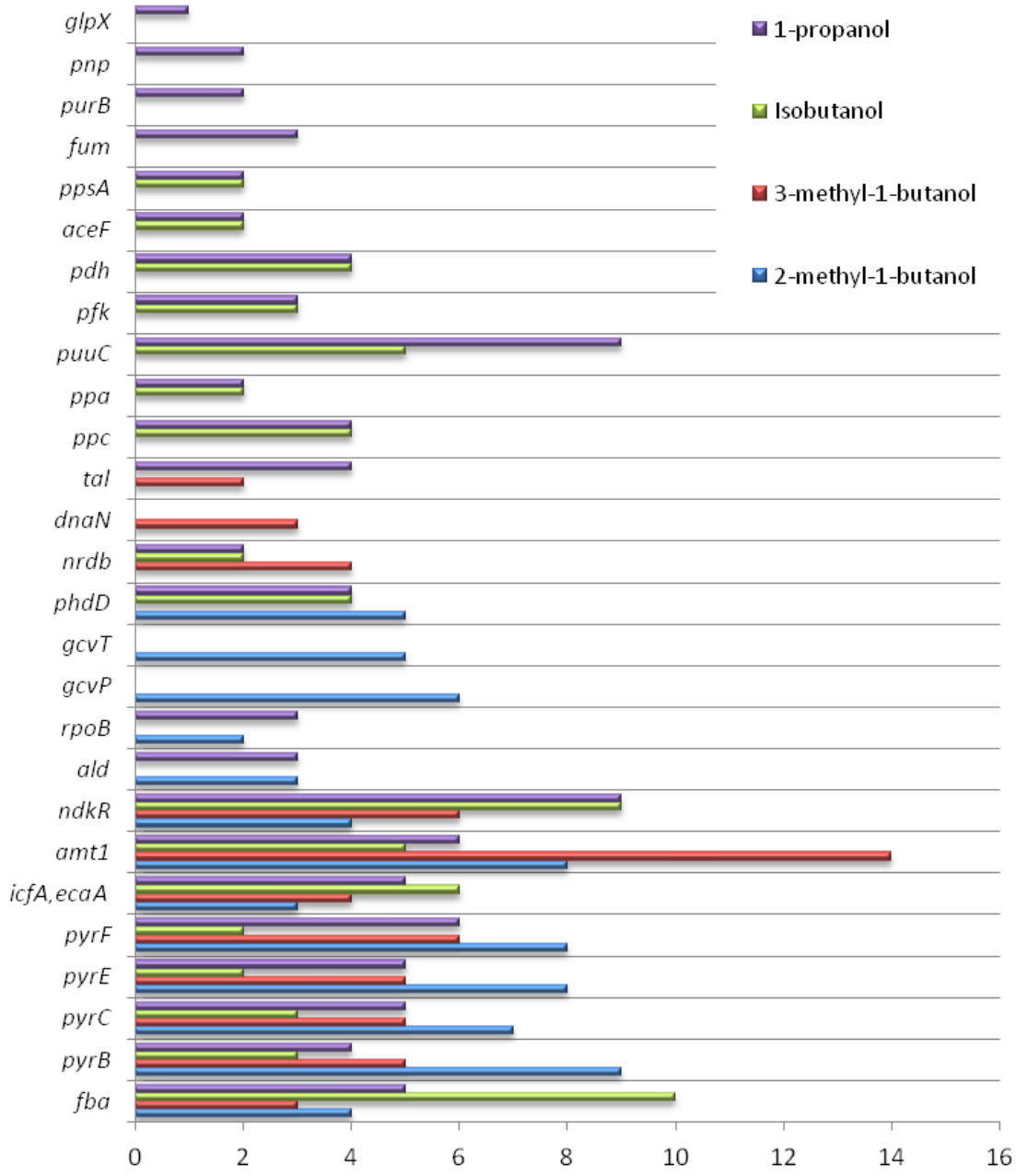


Figure 4.6. Most frequent mutations, proposed by MOMA analysis, which include the best candidates of the single, double and triple knockouts for maximize higher chain alcohols production.

In all cases, the simulation reveals the capability of the metabolic network when the flux is rerouted to the synthesis of these bio-products. As seen in the figure 4.7, the predicted productivity of both the 2-methyl-1-butanol as the 3-methyl-1-butanol behaves equally. Thus, for the maximum biomass growth (0.05987 h^{-1}), the theoretical synthesis of these metabolites was 2.42 10^{-9} mmol gDW^{-1} h^{-1}. Whereas for the same value of biomass growth, the theoretical synthesis of 1-propanol and isobutanol was 4.03 10^{-9} mmol gDW^{-1} h^{-1} and 3.02·10^{-9} mmol gDW^{-1} h^{-1}, respectively.

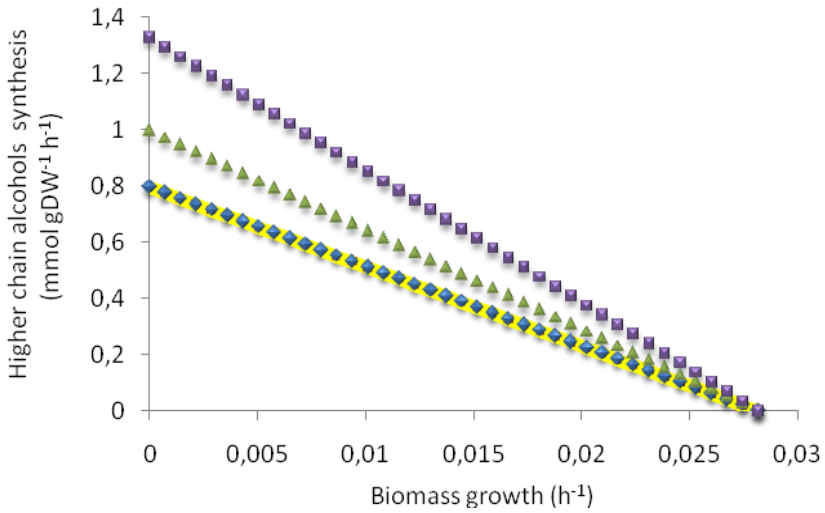


Figure 4.7. Theoretical productivity of higher chain alcohols as predicted by using *i*Syf715. Maximum ethanol production is shown as a function of minimal demand on biomass formation under autotrophic growth. The purple squares represent productivity of 1-propanol, green triangles represent isobutanol, yellow squares represent 3-methyl-1-butanol and blue diamonds correspond to 2-methyl-1-butanol.

In figure 4.7, the maximal value for higher chain alcohols synthesis rate derived by *i*Syf715 was predicted to be larger when the biomass growth is minimal. Similar to ethanol, the intercept to the y-axis represents an strain whose energy and reduced power are only intended to synthesize the higher alcohols. While the right part represent the wild strain, which does not produce the alcohols, since it is not described in the biomass equation. Note that the highest modular value of the slope corresponds to 1-propanol, followed by isobutanol and less and equal value for 3-methyl-1-butanol and 2-methyl-1-butanol. Once again, the graph can be used to estimate the theoretical maximum production, and therefore to organize the mutants previously simulated. It is not always interesting designing an strategy whose objectives are the production of ethanol with a low biomass growth. Usually, a compromise solution in the middle of that line will be the desired objective.

A metabolic engineering approach, using *Escherichia coli*, was performed in order to produce higher alcohols including those used in our analysis (Atsumi *et al*. 2008). Their production rates for 2-methyl-1-butanol, 3-methyl-1-butanol, isobutanol and 1-propanol were, approximately, 0.049 mmol gDW^{-1} h^{-1}, 0.095 mmol gDW^{-1} h^{-1}, 0.336 mmol gDW^{-1} h^{-1} and 0.0333 mmol gDW^{-1} h^{-1}, respectively. Although this strategy was developed in another biological host, it must be noted that in the present study these production values were obtained by *in silico* simulation with growth values close to the

optimal (see figure 4.8). This suggests how efficient the process could be if we wish to achieve a maximum in the production of these metabolites and at the same time reach a growth rate as close to maximum.

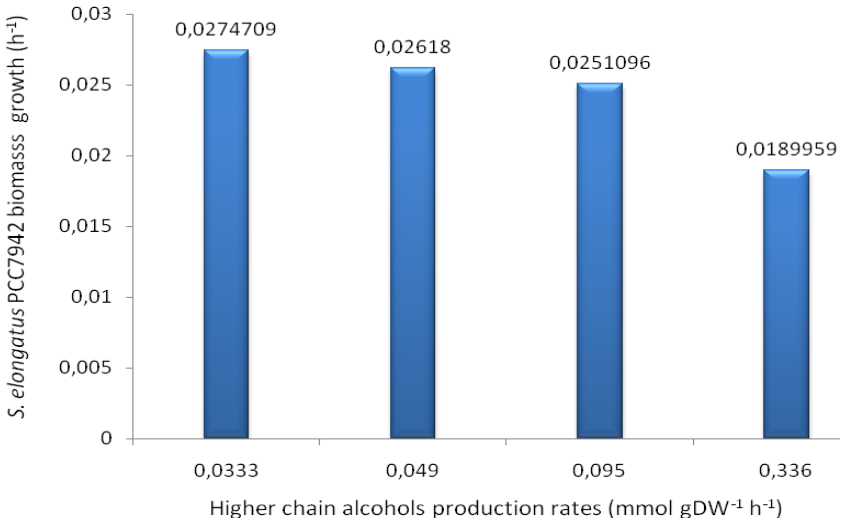


Figure 4.8. Relantionship between growth rate of *Synechococcus elongatus* PCC7942 (y-axis) with the synthesis rate values of higher chain alcohols reported for *E. coli* (x-axis) (Atsumi *et al*. 2008). These x-axis values correspond, from left to right, to 1-propanol, 2-methyl-1-butanol, 3-methyl-1-butanol and isobutanol.

An engineering approach to synthesize isobutyraldehyde and isobutanol, have been successfully developed in *Synechococcus elongatus* PCC7942 (Atsumi *et al*., 2009b). This paper reports a productivity of isobutanol production of 3.125 mg (isobutanol) L^{-1} h^{-1} at a biomass growth rate approximately 0.0169 h^{-1}. As a result of our theoretical calculations, the model can produce 0.408205 mmol gDW^{-1} h^{-1} of isobutanol at the same biomass growth, which is a 3779-fold increase over the results of Atsumi *et al.* This demonstrates the potentialities of this organism model as high-yield production platform if metabolic engineering strategies and by-product synthesis conditions were designed thoroughly.

4.2.3 Assessing lipids synthesis for biodiesel and industrial applications

Lipids-fuel, like fatty acid methyl esters (biodiesel), obtained from vegetable oils, animal fats and algae have emerged as a viable alternative to petroleum diesel. It has received attention as a renewable, biodegradable, and non-polluting fuel (Pearl, 2002; Kiss and Bildea, 2012) as well as, being a potential product from microbial biomass.

There is little information on whether the cyanobacteria produce and accumulate large amounts of lipids. Although these prokaryotes have low lipid content compared to other organisms such as microalgae (Francisco *et al.*, 2010), it has been found that lipid synthesis is related to high levels of photosynthesis and high biomass productivity, so the use of photosynthetic organisms for biodiesel production is becoming more and more attractive (Karatay and Donmez, 2011; Costa and Morais, 2011; Taher *et al.*, 2011; Da Rós *et al.*, 2013).

For simulation purposes, we designed an design objective function that groups all free fatty acids that are described in our model and which quantities have been reported by Da Rós and co-workers (Da Rós *et al.*, 2013). This objective function consists of octanoic acid (caprylic acid), decanoic acid (capric acid), dodecanoic acid (lauric acid), tetradecanoic acid (myristic acid), hexadecanoic acid (palmitic acid), octadecanoic acid (stearic acid), palmitoleic acid and oleic acid.

Additionally, we have chosen free saturated fatty acids (FFA), like palmitic acid (C16) and stearic acid (C18), to study the network capacity to synthesize them. These FFAs are considered value-added compounds of industrial interest. For example, sodium palmitate, which is commonly obtained by saponification of palmitic acid, is widely used as feedstock in the food industry, cosmetics and pharmaceutical industries (Elder, 1987; Burdock, 2005). Stearic acid is frequently used as lubricants, dispersing agent and softener in rubber compounds and in layer coatings. This acid is also employed in the manufacture of stearates and stearate dryers, and its salts are used for pharmaceuticals preparations and in soaps, detergents, cosmetics and dietary supplements production (Elder, 1987; Khalil *et al.*, 2000; Burdock, 2005).

We simulated single, double and triple knockout strategies in order to find possible genetic targets to increase this lipids production. Once again, we used MOMA algorithm implemented in OptGene software (Patil *et al*., 2005). As part of the input file format of this software, we defined a “design objective function" as palmitic and stearic acids syntheses or fatty acids synthesis and left biomass production as “biological objective function”. We used *i*Syf715 metabolic model for which wild type optimal growth rate was 0.05987 h^{-1}. In figure 4.9 we can see the results for single knockouts simulation of the fatty acids mixture´s genes.

A

Reaction in model	Gene
1.1.1.47a	gdh
1.1.1.47b	gdh
2.7.4.3a	adk
4.2.1.1b	icfA
_1.3.1.26a	dapB
_1.5.1.2a1	proC
_1.18.-.-	ftrC
2.3.1.35	argJ
1.1.1.86a	ilvC
6.3.4.18	purK

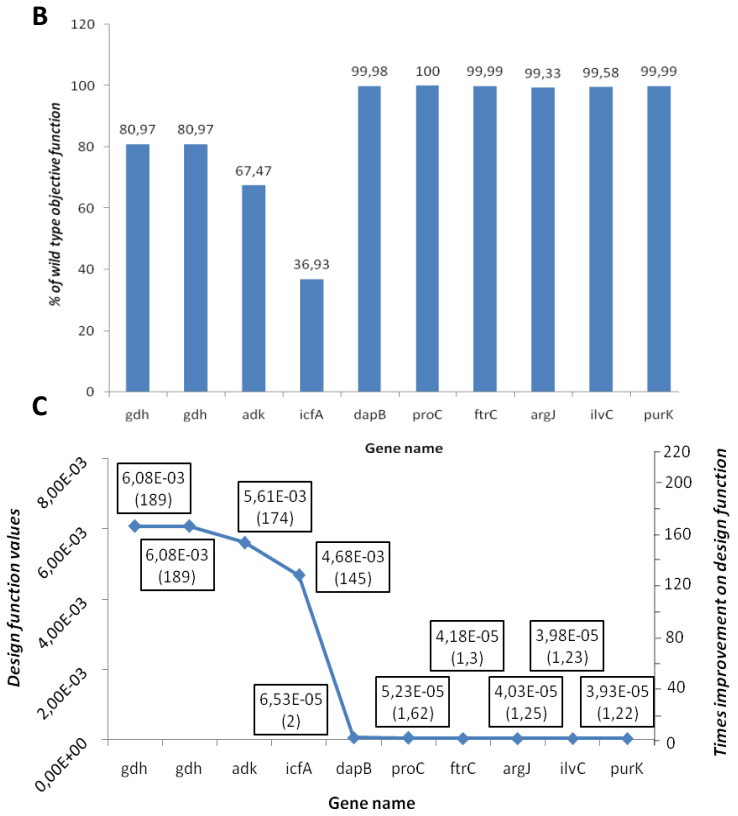


Figure 4.9. Proposed single knockouts for an improved fatty acids profile production. **A**. List of top 10 single knockouts by which they reached a maximum on the *design function* (fatty acids profile synthesis). **B**. Percent depicting the *mutant objective function* (biomass production) relative to the *wild type objective function*. **C**. Values of the design function (first value in the text box) relative to the deleted genes as well as the times improvement on this design function (value enclosed in brackets).The knockout in *icfA* gene implies knockout in *ecaA* gene. Units for *objective function* and *design function* in h^{-1} and mmol gDW^{-1} h^{-1}, respectively.

According to these results, the best single knockout strategy to increase the synthesis rate for the fatty acids involves the removal of one of the genes from the pentose phosphate pathway. In that case, the simulation predicts that the best strategy is deleting *gdh* gene coding for glucose-1-dehydrogenase (reaction *"1.1.1.47a"* and *"1.1.1.47b"* in *i*Syf715 and ORF ID: *Synpcc7942_1573*). With this deletion the resulting phenotype is capable of producing 0.00608 mmol gDW^{-1} h^{-1} fatty acids mix, as well as a growth rate of 0.0227973 h^{-1} (see Figure 4.9). With this knockout, both NADH and NADPH would be more available for the reactions catalyzed by enoyl-[acyl-carrier-protein] reductase (NADH) and for 3-oxoacyl-[acyl-carrier-protein] reductase, respectively (reactions *"1.3.1.9"* and *"1.1.1.100"*, respectively, in *i*Syf715).

The suppression of *adk* gene that codes for adenylate kinase (reaction *"2.7.4.3a"* in *i*Syf715 and ORF ID: *Synpcc7942_2213*), constitutes second best strategy to enhance

this lipid production. As in the ethanol case, this deletion could reduce the ATP consumption, which is essential in reactions, among many others to the ones of the fatty acid biosynthesis pathway, such as those catalyzed by acetyl-CoA carboxylase (reaction *"_6.4.1.2a"* in *i*Syf715) to yield malonyl-CoA, and the reaction catalyzed by biotin carboxylase (reaction *"_6.3.4.14"* in *i*Syf715) to form carboxybiotin-carboxyl-carrier protein which carboxylated the acetyl-CoA to malonyl-CoA.

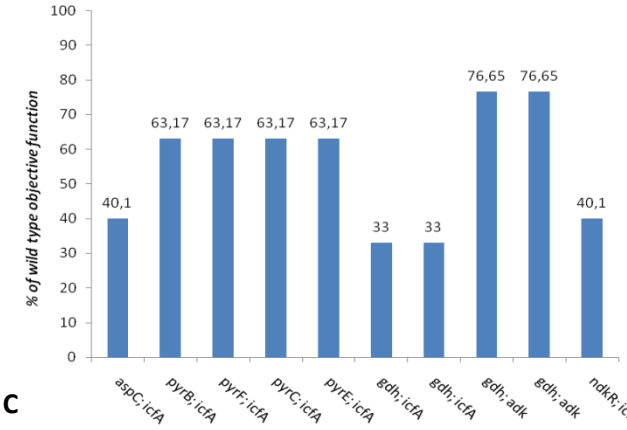
Furthermore, the best double knockouts candidates are shown in the figure 4.10. The double mutant that allows a theoretical maximum of fatty acid mix is one that combines the deletion of *aspC and icfA*/*ecaA* encoding for aspartate aminotransferase and carbonic anhydrase (reactions "*2.6.1.1a"* and *"4.2.1.1b"* in *i*Syf715 and ORF ID: *Synpcc7942_2545*, *Synpcc7942_1447*, *Synpcc7942_1388* and *Synpcc7942_B2619*). The resulting phenotype is capable of achieving a fatty acids yield of 0,0251 mmol gDW^{-1} h^{-1}, as well as a growth rate of 0.0113398 h^{-1} (Figure 4.10). This strategy includes the reverse of the reaction catalyzed by aspartate aminotransferase: the transformation of L-glutamate to L-aspartate. Suppressing metabolic flux through this transamination reaction prevents the use of this amino acid as substrate in reactions where the direct consumption of ATP is required, such as the reactions catalyzed by argininosuccinate synthase, aspartate kinase and phosphoribosylaminoimidazolesuccinocarboxamide synthase (reactions "*6.3.4.5a", "2.7.2.4"* and *"6.3.2.6"*, respectively, in *i*Syf715) or the indirect use of both this cofactor and HCO_3^-, by partially coupled reactions catalyzed by carbamoyl-phosphate synthetase and aspartate transcarbamylase (ATCase) (reactions *"6.3.5.5"* and *"2.1.3.2"* in *i*Syf715).

As we discussed previously, the availability of ATP would increase the total flow of the fatty acid biosynthesis pathway through the reactions catalyzed by acetyl-CoA carboxylase (reaction *"_6.4.1.2a"* in *i*Syf715) and biotin carboxylase (reaction *"_6.3.4.14"* in *i*Syf715).If the flux through the reaction catalyzed by carbonic anhydrase is blocked, it prevents the HCO_3^- conversion to CO_2. Through this strategy we can achieve that this carbon source remains available for the carboxylation of acetyl-CoA to malonyl-CoA catalyzed by acetyl-CoA carboxylase (reaction *"_6.4.1.2a"* in *i*Syf715), which represents the first step of fatty acid biosynthesis and one of its rate-controlling steps (Molenaar *et al.*, 2003).

A

Pair of reactions	Pair of genes
2.6.1.1a; 4.2.1.1b	aspC; icfA
2.1.3.2; 4.2.1.1b	pyrB; icfA
4.1.1.23; 4.2.1.1b	pyrF; icfA
3.5.2.3; 4.2.1.1b	pyrC; icfA
2.4.2.10; 4.2.1.1b	pyrE; icfA
1.1.1.47a; 4.2.1.1b	gdh; icfA
1.1.1.47b; 4.2.1.1b	gdh; icfA
1.1.1.47a ; 2.7.4.3a	gdh; adk
1.1.1.47b; 2.7.4.3a	gdh; adk
2.7.4.6d; 4.2.1.1b	ndkR; icfA

B



C

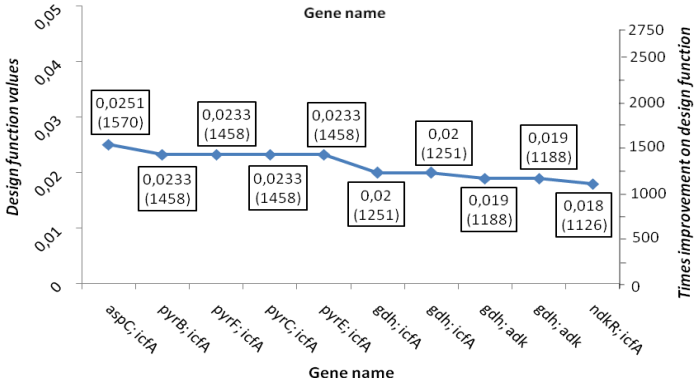


Figure 4.10. Proposed double knockouts for an improved fatty acids profile production. **A**. List of top 10 single knockouts by which they reached a maximum on the *design function* (ethanol evolution). **B**. Percent depicting the *mutant objective function* (biomass production) relative to the *wild type objective function*. **C**. Values of the design function (first value in the text box) relative to the deleted genes as well as the times improvement on this design function (value enclosed in brackets). The knockout in *icfA* gene implies knockout in *ecaA* gene. Units for *objective function* and *design function* in h^{-1} and mmol gDW^{-1} h^{-1}, respectively.

The study of triple knockout with MOMA algorithm pointed out that the best strategy is one that involved mutations in *icfA/ecaA, ndkR, and gdh* that encode for carbonic anhydrase, nucleoside diphosphate kinase and glucose-1-dehydrogenase, respectively (reactions "*4.2.1.1b", "2.7.4.6d"* and "*1.1.1.47a"*, respectively, in *i*Syf715 and ORF ID: *Synpcc7942_1447*, *Synpcc7942_1388* and *Synpcc7942_B2619, Synpcc7942_2497* and *Synpcc7942_1573*). As we have explained for each of these mutations, this strategy prevents to decrease the concentrations of HCO_3^-, ATP or NADH and NADPH, which are metabolites required in the synthesis of the fatty acids and the energetic pathways described in our model.

In figure 4.11 we list the most frequent genetic deletions with which our simulated mutant strains have a maximum fatty acids synthesis. Here, *icfA/ecaA* mutations are the most common ones followed by *gdh* and *adk* genes.

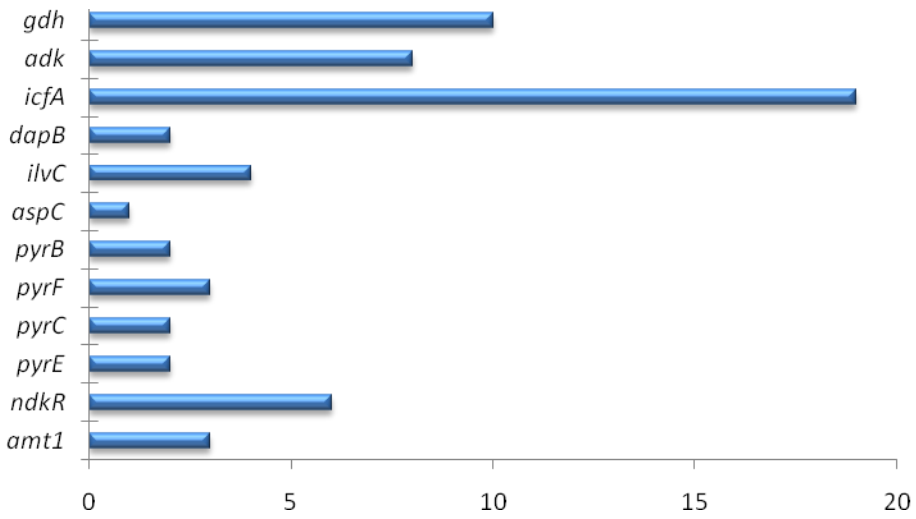


Figure 4.11. Most frequent mutations, proposed by MOMA analysis, within the top 10 candidates in single, double and triple knockouts for maximize fatty acids profile production. The knockout in *icfA* gene implies knockout in *ecaA* gene.

Figure 4.12 shows the results of the simulation for palmitic and stearic acids genes single knockouts. The best single knockouts in the direction of enhancing the synthesis of both palmitic and stearic acid coincides with the deletion of *icfA*/*ecaA* genes that encoding for carbonic anhydrase (reaction *"4.2.1.1b"* in *i*Syf715 and ORF ID: *Synpcc7942_1447*, *Synpcc7942_1388*, *Synpcc7942_B2619*). The system is capable of achieve maximum palmitic acid and stearic acid synthesis at 0.0457 and 0.0403 mmol gDW^{-1} h^{-1}, respectively, as well as a growth rates of 0.0104303 h^{-1} and 0.0105438 h^{-1} (Figure 4.12). As we explained before, this mutation contributes to avoid decrease of HCO_3^- concentration into the system.

In the case of palmitic acid, the second best single knockout involves the removal of one of the genes required in pentose phosphate pathway. Namely the *gdh* gene that codifies for glucose-1-dehydrogenase (reaction *"1.1.1.47a"* and *"1.1.1.47b"* in *i*Syf715 and ORF ID: *Synpcc7942_1573*). The biochemical effects of these mutations have been explained above. However, in the case of stearic acid, the second best candidate is related to blocking ammonium imports through one of its carriers (reaction *"ammonia H3N TRANS-RXN59G-178"* in *i*Syf715 and ORF ID: *Synpcc7942_0442* and *Synpcc7942_2279*). By means of the ammonia entrance blockage, the metabolic flux through the reactions catalyzed by *6.3.1.2*, *6.3.4.2* and *6.3.1.5*, is affected. All of these reactions relate the use of ATP as a phosphate groups donor or as energy currency,

thus, without these reactions, the flux through the fatty acid biosynthesis pathway is favored.

A

Reaction in model		Gene	
Palmitic acid	**Stearic acid**	**Palmitic acid**	**Stearic acid**
4.2.1.1b	4.2.1.1b	icfA	icfA
1.1.1.47a	ammonia H3N TRANS-RXN59G-178	gdh	amt1
1.1.1.47b	_1.10.9.1	gdh	pet
2.7.4.6d	1.1.1.47a	ndkR	gdh
ammonia H3N TRANS-RXN59G-178	1.1.1.47b	amt1	gdh
_1.10.9.1	2.7.4.6d	pet	ndkR
3.5.2.3	2.7.4.3a	pyrC	adk
2.4.2.10	3.5.2.3	pyrE	pyrC
4.1.1.23	2.4.2.10	pyrF	pyrE
2.1.3.2	4.1.1.23	pyrB	pyrF

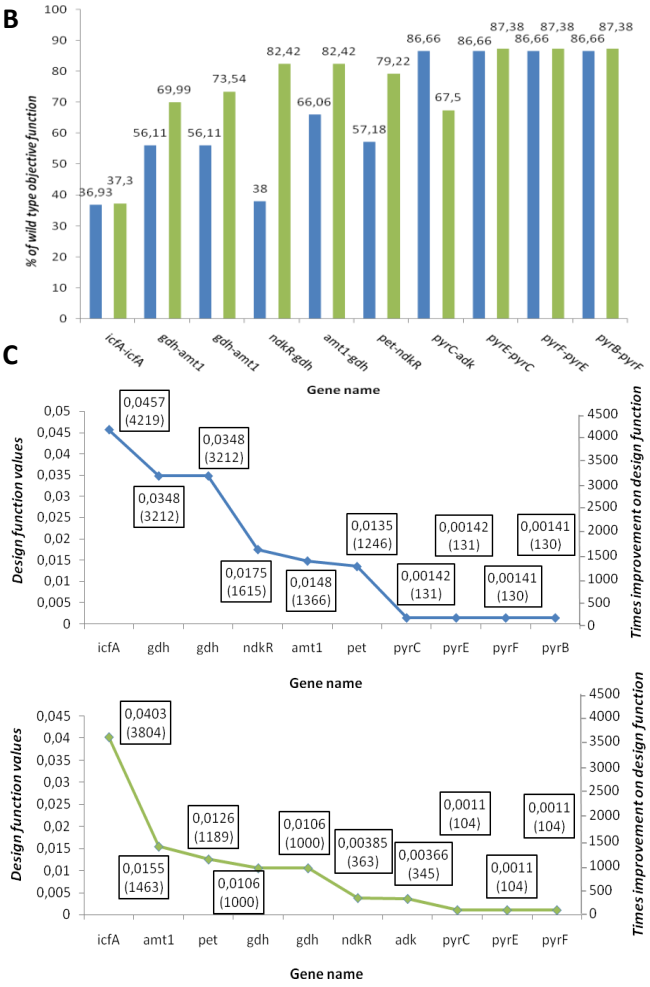


Figure 4.12. Proposed single knockouts for an improved palmitic and stearic acid production. **A**. List of top 10 single knockouts by which they reached a maximum on the *design function* (palmitic and stearic acid synthesis). **B**. Percent depicting the *mutant objective function* (biomass production) relative to the *wild type objective function*. Blue and green bars correspond to mutants whose design functions are the palmitic and stearic acid production, respectively. **C**. Values of the design function (first value in the text box) relative to the deleted genes as well as the times improvement on this design function (value enclosed in brackets). The blue and green lines belong to the palmitic and stearic acid production, respectively. The knockout in *icfA* gene implies knockout in *ecaA* gene. Units for *objective function* and *design function* in h^{-1} and mmol gDW^{-1} h^{-1}, respectively.

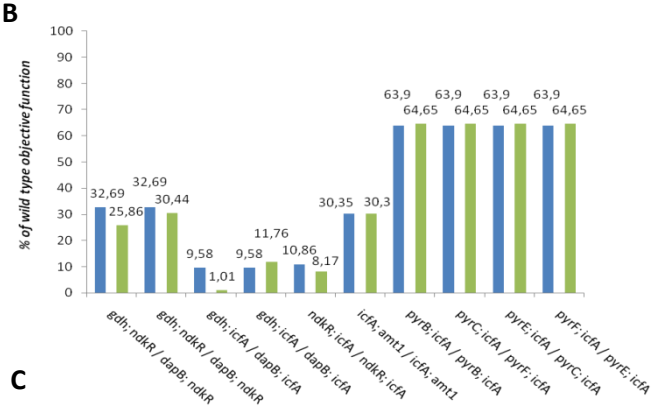
The simulation results for double-gene knockouts of palmitic and stearic acids are shown in figure 4.13. An increase in the lipids synthesis rate in both double mutants' simulations was observed. For palmitic acid synthesis, the best mutant strain would be one devoid of the *gdh* and *ndkR* genes coding for glucose-1-dehydrogenase and nucleoside diphosphate kinase, respetively (reactions *"1.1.1.47a"* and *"2.7.4.6d"*, respectively, in *i*Syf715 and ORF ID: *Synpcc7942_1573, Synpcc7942_2497*). The system

is capable of achieving a maximum in the palmitic acid synthesis of 0.0782 mmol gDW^{-1} h^{-1} and a growth rate of 0.00921336 h^{-1} (Figure 4.13). As can be seen, this design shown an increase, about twice, if compared to simulation for single mutants.

A

Pair of reactions		Pair of genes	
Palmitic acid	**Stearic acid**	**Palmitic acid**	**Stearic acid**
1.1.1.47a; 2.7.4.6d	_1.3.1.26a; 2.7.4.6d	gdh; ndkR	dapB; ndkR
1.1.1.47b; 2.7.4.6d	_1.3.1.26b; 2.7.4.6d	gdh; ndkR	dapB; ndkR
1.1.1.47a; 4.2.1.1b	_1.3.1.26a; 4.2.1.1b	gdh; icfA	dapB; icfA
1.1.1.47b; 4.2.1.1b	_1.3.1.26b; 4.2.1.1b	gdh; icfA	dapB; icfA
2.7.4.6d; 4.2.1.1b	2.7.4.6d; 4.2.1.1b	ndkR; icfA	ndkR; icfA
4.2.1.1b; ammonia H3N TRANS-RXN59G-178	4.2.1.1b; ammonia H3N TRANS-RXN59G-178	icfA; amt1	icfA; amt1
2.1.3.2; 4.2.1.1b	2.1.3.2; 4.2.1.1b	pyrB; icfA	pyrB; icfA
3.5.2.3; 4.2.1.1b	4.1.1.23; 4.2.1.1b	pyrC; icfA	pyrF; icfA
2.4.2.10; 4.2.1.1b	3.5.2.3; 4.2.1.1b	pyrE; icfA	pyrC; icfA
4.1.1.23; 4.2.1.1b	2.4.2.10; 4.2.1.1b	pyrF; icfA	pyrE; icfA

B



C

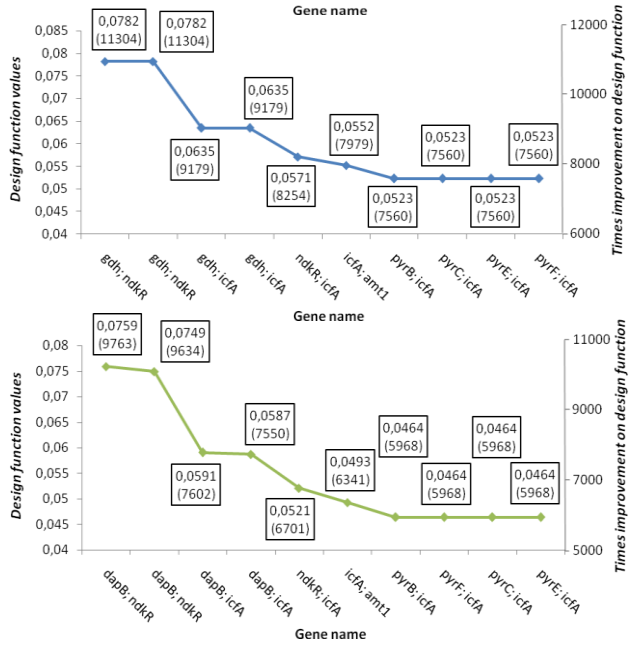


Figure 4.13. Proposed double knockouts for an improved palmitic acid and stearic acid production. **A**. List of top 10 single knockouts by which they reached a maximum on the *design function* (palmitic acid and stearic acid synthesis). **B**. Percent depicting the *mutant objective function* (biomass production) relative to the *wild type objective function*. Blue and green bars correspond to mutants whose design functions are the palmitic and stearic acid production, respectively. **C**. Values of the design function (first value in the text box) relative to the deleted genes as well as the times improvement on this design function (value enclosed in brackets). The blue and green lines belong to the palmitic and stearic acid production, respectively. The knockout in *icfA* gene implies knockout in *ecaA* gene. Units for *objective function* and *design function* in h^{-1} and mmol gDW^{-1} h^{-1}, respectively.

Like in result for palmitic acid, the best double mutant strain for stearic acid maintains the deletion in the gene *ndkR*. Nevertheless, this deletion combine the

knockout in the gene *dapB* that coding for the enzyme dihydrodipicolinate reductase (reaction *"2.7.4.6d"* and *"_1.3.1.26a-_1.3.1.26b"*, respectively, in *i*Syf715 and ORF ID: *Synpcc7942_2497, Synpcc7942_2136*). In this case, the system is capable of achieving a maximum in the stearic acid synthesis of 0.0759 mmol gDW^{-1} h^{-1} and a growth rate of 0.00727966 h^{-1} (Figure 4.13). The deletion of dihydrodipicolinate reductase entails the non-oxidation of tetrahydrodipicolinate and therefore, as explained above, the cofactors NADH and NADPH are kept accessible for the synthesis of fatty acids.

Through triple knockout simulation we observed an improved lipids production relative to the results of the single and double knockouts. The best *in silico* strain includes the knockouts of gene *ndkR* coding for nucleoside diphosphate kinase (reaction *"2.7.4.6d"* in *i*Syf715 and ORF ID: *Synpcc7942_2497*), gene *pyrF* coding for orotidine-5-phosphate decarboxylase (reaction *"4.1.1.23"* in *i*Syf715 and ORF ID: *Synpcc7942_2569*) and genes *icfA* and *ecaA* that codify for carbonic anhydrase (reaction *"4.2.1.1b"* in *i*Syf715 and ORF ID: *Synpcc7942_1447*, *Synpcc7942_1388*, *Synpcc7942_B2619*). This candidate includes the possibility of removing genes, as explained above, involved in the reactions using ATP directly or indirectly, such as the reactions *"2.7.4.6d"* and *"4.1.1.23"*; and at the elimination of the gene for carbonic anhydrase related the dehydration of bicarbonate ion.

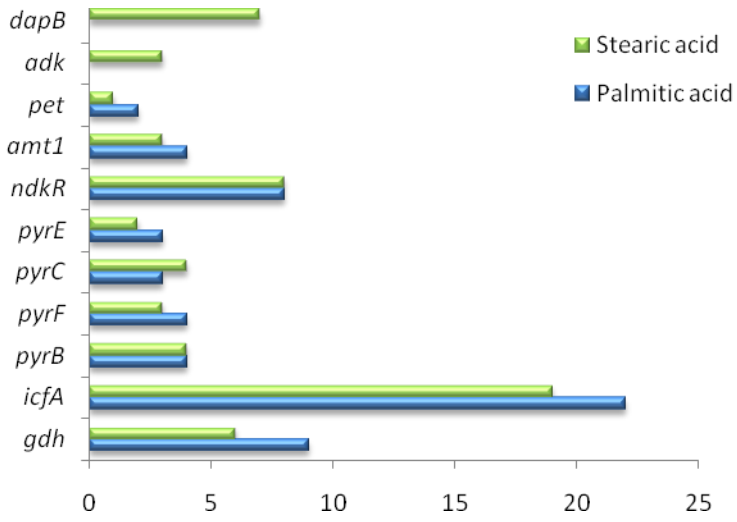


Figure 4.14. Most frequent mutations, proposed by MOMA analysis, within the top 10 candidates in single, double and triple knockouts for maximize palmitic acid and stearic acid production. The knockout in *icfA* gene implies knockout in *ecaA* gene.

Once again, we show the most frequent genetic deletions for which the mutant strains, theoretically, reach a maximum of the objective function (Figure 4.14). Mutations on *icfA/ecaA* genes, followed by the *ndkR* and *gdh* genes are most commonly predicted knockouts by MOMA simulation.

Through FBA simulation we estimated theoretical maximum yield of these fatty acids production. For comparative purposes we checked the results of the *in silico* simulation with lipid productivity published by Da Ros *et al.* for this specie (Da Rós *et al.*, 2013). Specifically, we compared theoretical productivity of the fatty acids profile, in which we took into account only 8 of the 21 reported free fatty acids.

We found that for CO_2 uptake rate of 1.99 mmol gDW^{-1} h^{-1} and an incident irradiance of 1.96 mE gDW^{-1} h^{-1}, our prediction for theoretical lipids productivity behaves as shown in the graph in Figure 4.2. For a biomass growth of 0.0246 h^{-1}, the theoretical synthesis of the fatty acids profile was 0.0501 mmol gDW^{-1} h^{-1}, a value that is very close to the published (≈0.0521 mmol gDW^{-1} h^{-1}) for a biomass productivity of 52.7 mg L^{-1} day^{-1} in relation to dry biomass. This is an important point within the validation process that has been mentioned from the previous chapters.

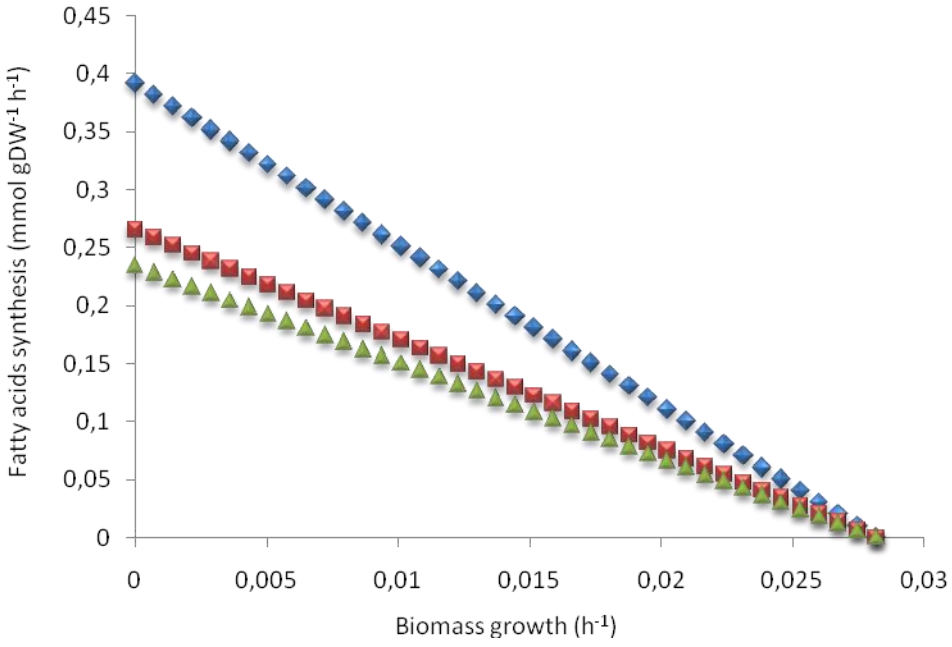


Figure 4.15. Theoretical productivity of fatty acids as predicted by using *i*Syf715. Maximum fatty acids production is shown as a function of minimal demand on biomass formation under autotrophic growth. The blue diamonds, red squares and green triangles correspond to productivities of fatty acids profile, palmitic acid and steraric acid, respectively.

As we expected, the function between lipids productivity and growth rate indicates the linear relationship between these variables. As our targeted lipids are considered building blocks, it is obvious that the carbon sources are drained towards the synthesis of biomass when this is maximized. However, the phenotypes where the maximization of these lipids is incorporated as a new feature, show high theoretical production rates when the carbon flux is channeled to towards the fatty acids biosynthesis. Unlike eukaryotic algae, cyanobacteria are not high natural lipid producers. Therefore, these *in silico* productivities give us an idea of the potentialities of possible genetic modifications into this biological system.

4.2.4 Assessing hydrogen evolution

Hydrogen gas (H_2) features huge potential as a clean energy resource, basically because reduction on carbon dioxide production. However, in spite of being considered a non-polluting fuel biotechnological development is needed to produce it more efficiently from renewable sources. Economically, advances on biological hydrogen production processes are still not as attractive as conventional H_2-production processes, so it is a challenge to make it more competitive in near future (Das, 2010). Biological systems possess a broad range of physicochemical mechanisms to produce hydrogen, like: photo-fermentations, dark-fermentation direct biophotolysis and indirect biophotolysis, (Kondratieva and Gogotov, 1983; Nandi and Sengupta, 1998; Das and Veziroglu, 2001; Hallenbeck and Benemann, 2002). Among this bioprosses, the photo-fermentation process is an exciting new area of technology development that should offer a potential production of usable hydrogen. This process has the strong advantages of being of using light as the energy source and is environmentally clean. It was found that biological hydrogen synthesis rates are not enough to power small hydrogen protons fuel cells on a continuous basis (Levin *et al.*, 2004). In this sense, many reports highlight that several cyanobacteria species have the capacity to synthesize H_2 in a natural way (Abed *et al.,* 2009; Das and Veziroglu 2001; Dutta *et al.,* 2005).

This biofuel is produced by metabolic pathways that include nitrogenase or hydrogenase depending on the type of cyanobacterium, which may be either nitrogen fixing or non-nitrogen fixing. *Synechococcus* sp. are non-nitrogen fixing organisms that

are suitable for higher hydrogen evolution (Lopes-Pinto *et al.*, 2002; Dutta *et al.,* 2005). One of the major drawbacks in the use of these microorganisms to produce hydrogen is that these enzymes are strongly inhibited by O_2 produced by photosynthesis. There are some engineering strategies to mitigate this problem, which basically try to redirect the metabolic flux towards the bio-hydrogen synthesis as well as to modify the hydrogenases to decrease the O_2´s negative allosteric effect (Angermayr *et al.*, 2009; Weyman 2010; McNeely *et al.*, 2010).

As part of efforts to find a solution to all these problems, we simulated single, double and triple knockout strategies in order to find possible genetic targets to increase hydrogen production in *S. elongatus* PCC7942. As before, we used MOMA algorithm implemented in OptGene software (Patil *et al*., 2005). As part of the input file format of this software, we defined the “design objective function" as hydrogen synthesis whereas the biomass synthesis was the “biological objective function”. We used *i*Syf715 metabolic model for which wild type optimal growth rate was 0.00789506 h^{-1}.

By deleting the genes *thiD, tmk, ubiG, Synpcc7249_2055, PduX and cobF* that code for the enzymes: phosphomethylpyrimidine kinase, thymidylate kinase, 3-demethylubiquinol 3-O-methyltransferase, 2-octaprenyl-6-methoxyphenol hydroxylase (enzyme not annotated in the genome), L-threonine kinase (enzyme not annotated in the genome) and precorrin-6A synthase (enzyme not annotated in the genome), respectively (reaction *"2.7.4.7"*, *“2.7.4.9b”*, *“·2.1.1.64”, “1.14.13.-a”*, *“·RXN-8626” and “·2.1.1.152”*, respectively in *i*Syf715 and ORF ID: *Synpcc7942_2379, Synpcc7942_0093 and Synpcc7942_2055*), the biological system is capable of achieving the greatest flux towards the production of hydrogen, here, $4.82 \cdot 10^{-5}$ mmol gDW^{-1} h^{-1} (Figure 4.16).

With the knockout in *thiD*, the system achieves the greatest flux towards the hydrogen evolution with a growth rate of 0.00783143 h^{-1}. When this reaction is silenced, hydroxymethylpyrimidine-phosphate phosphorylation can be prevented, diminishing the ATP consumption and increasing, indirectly, NADPH concentration as indispensable redox cofactor for hydrogen synthesis. Unconsumed ATP quantities would be available for NAD^+ phosphorylation through the reaction catalyzed by NAD^+ kinase (reaction *"2.7.1.23"* in *i*Syf715). As a result, $NADP^+$ pool would increase, so that

its reduction could be enhanced by the reaction catalyzed by ferredoxin-$NADP^+$ reductase (reaction *"_1.18.1.2"* in *i*Syf715) in photosynthesis.

A

Reaction in model	Gene
2.7.4.7	thiD
2.7.4.9a	tmk
·2.1.1.64a	ubiG
1.14.13.-b	Synpcc7942_2055
RXN-8626	PduX
·2.1.1.64b	ubiG
·2.1.1.152	cobF
6.3.5.11	CbiA
4.1.2.25	folB
4.2.1.9a	ilvD

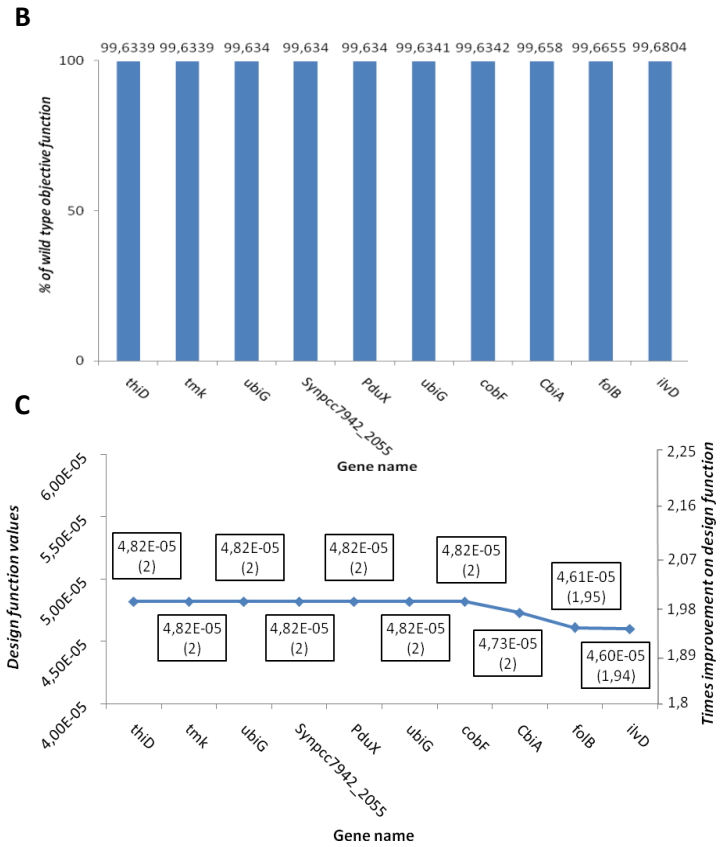


Figure 4.16. Proposed single knockouts for an improved hydrogen production. **A**. List of top 10 single knockouts by which they reached a maximum on the *design function* (hydrogen evolution). **B**. Percent depicting the *mutant objective function* (biomass production) relative to the *wild type objective function*. **C**. Values of the design function (first value in the text box) relative to the deleted genes as well as the times improvement on this design function (value enclosed in brackets). Units for *objective function* and *design function* in h^{-1} and mmol gDW^{-1} h^{-1}, respectively.

The best second mutation would involve blocking thymidylate kinase activity (reaction *"2.7.4.9a"* in *i*Syf715 and ORF ID: *Synpcc7942_0093*). This knockout from pyrimidine biosynthesis pathway would have similar effect to the one mentioned above: it would reduce the dTMP phosphorylation by ATP. It is noteworthy to state that the best strategies for singles mutants do not increase hydrogen evolution more than twice the wild type's titters.

There are little differences between double knockouts candidates and single knockouts in terms of improved hydrogen production (Figure 4.17). The best double deletions increased in 0.6 times the design function flux of the best single mutant. It

consists in the combination of deletions of the genes *pyrB* and *glgA* that encode for aspartate transcarbamylase (ATCase) and ADP-glucose transglycosylase, respectively (reactions *"2.1.3.2*" and *"_2.4.1.21b"* in *i*Syf715 and ORF ID: *Synpcc7942_0670*, *Synpcc7942_2518*). This *in silico* strain had a hydrogen evolution of $7.64 \cdot 10^{-5}$ mmol gDW^{-1} h^{-1} and with a growth rate of 0.00784371 h^{-1}.

The condensation of carbamoyl-phosphate with L-aspartate to form N-carbamoyl-L-aspartate is interrupted by removing the reaction catalyzed by the ATCase. This reaction, considered a flux-generating step of the pathway (Voet and Voet, 2004), occurs in the absence of ATP because carbamoyl-phosphate is intrinsically activated. However, interrupting the flux through this reaction could also affect the flux of the first reaction of pyrimidine biosynthesis, which is, precisely, the synthesis of carbamoyl-phosphate by enzyme carbamoyl phosphate synthetase, which involves the use of two molecules of ATP. Once again, via this deletion we could prevent the consumption of ATP and therefore it would have more reductant power for hydrogen production. In addition, by this strategy the transformation of HCO_3^- ion in carbamoyl-phosphate could be reduced, so the carbon incorporation through the carbonic anhydrase and RuBisCO could be optimized.

The mutation in *glgA* gene, which encodes for ADP-glucose transglycosylase, is aimed at eliminating the flux through glycogen synthesis, specifically, it aims at the reaction in which ADP-D-glucose units are added to the polymer growing chain. The glycogen biosynthesis requires an additional exergonic step since the direct conversion of alpha-D-glucose-1-phosphate to glycogen and phosphate is thermodynamically unfavorable. In prokaryotes, this process involves the reaction of alpha-D-glucose-1-phosphate with ATP catalyzed by ADP-glucose diphosphorylase (reaction *"2.7.7.27"* in *i*Syf715). With this strategy, ATP is once again enhanced.

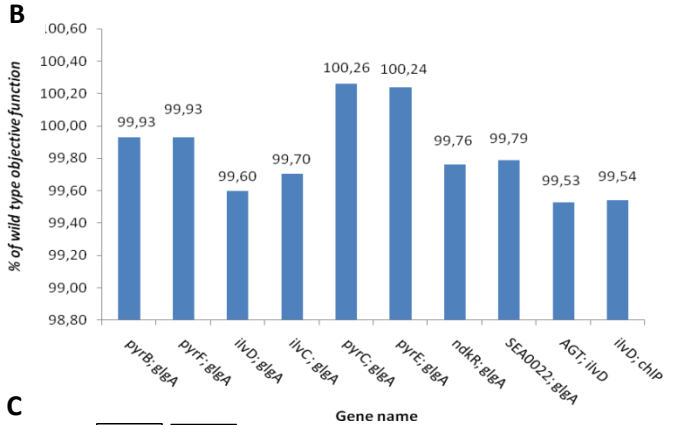
The eight best double knockouts candidates include the deletion of ADP-glucose transglycosylase, which represents an essential enzyme into glycogen metabolism, an active pathway of cyanobacterium carbon metabolism. The following seven combinations, in descending order, include simultaneous deletion of this gene with genes that encode for orotidine-5'-phosphate decarboxylase (reactions *"4.1.1.23*" in *i*Syf715), dihydroxy-acid dehydratase (reactions *"4.2.1.9a*" in *i*Syf715), acetohydroxy

acid isomeroreductase (reactions *"1.1.1.86a"* in *i*Syf715), dihydroorotase (reactions *"3.5.2.3"* in *i*Syf715), orotate phosphoribosyltransferase (reactions *"2.4.2.10"* in *i*Syf715), nucleoside diphosphate kinase (reactions *"2.7.4.6h"* in *i*Syf715) and protochlorophyllide reductase (reactions *"1.3.1.33b"* in *i*Syf715).

A

Pair of reactions	Pair of genes
2.1.3.2; _2.4.1.21b	pyrB; glgA
4.1.1.23; _2.4.1.21b	pyrF; glgA
4.2.1.9a; _2.4.1.21b	ilvD; glgA
1.1.1.86a; _2.4.1.21b	ilvC ; glgA
3.5.2.3; _2.4.1.21b	pyrC; glgA
2.4.2.10; _2.4.1.21b	pyrE; glgA
2.7.4.6h; _2.4.1.21b	ndkR; glgA
1.3.1.33b; _2.4.1.21b	SEA0022; glgA
2.6.1.44; 4.2.1.9a	AGT; ilvD
4.2.1.9a; 1.3.1.83	ilvD; chlP

B



C

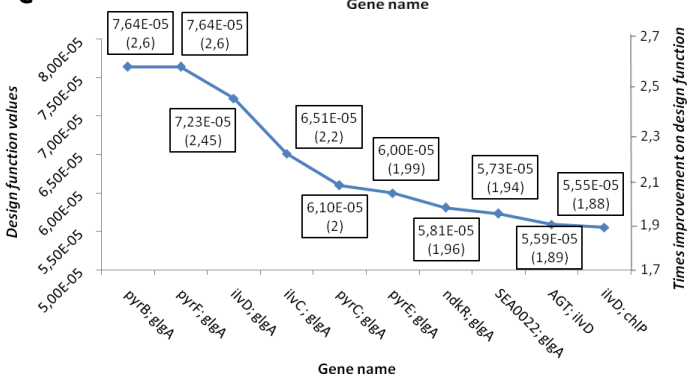


Figure 4.17. Proposed double knockouts for an improved hydrogen production. **A**. List of top 10 double knockouts by which they reached a maximum on the *design function* (hydrogen evolution). **B**. Percent depicting the *mutant objective function* (biomass production) relative to the *wild type objective function*. **C**. Values of the design function (first value in the text box) relative to the deleted genes as well as the times improvement on this design function (value enclosed in brackets). Units for *objective function* and *design function* in h^{-1} and mmol gDW^{-1} h^{-1}, respectively.

We looked to the other combination of knockouts that follows in order. It consists in the elimination of the genes for alanine-glyoxylate aminotransferase and dihydroxy-acid dehydratase (reactions *"2.6.1.44"* and *"4.2.1.9a"*, respectively, in *i*Syf715 and ORF IDs *Synpcc7942_2160*, *Synpcc7942_0626*). This strategy computes a hydrogen evolution of $5.55 \cdot 10^{-5}$ mmol gDW^{-1} h^{-1}, and with a growth rate of 0.00781212 h^{-1}. The first gene is associated with the transamination process, in which the amino group of L-alanine is transferred to glyoxylate to become pyruvate and L-glycine, respectively. This last metabolite is involved in some reactions in *i*Syf715, like *tRNA 6.1.1.14*,

6.3.4.13 and *6.3.2.3* where it involve the use of ATP to provide either a phosphate group or to energize the reactions. On the other hand, the other product of *2.6.1.44,* pyruvate, is involved in reaction *2.7.1.40c*, where this metabolite captures a phosphate group from GTP to become phosphoenolpyruvate. In our model, GTP comes from GDP and ATP by the reaction catalyzed by nucleoside diphosphate kinase (reaction *"2.7.4.6c"* in *i*Syf715). As noted before, our simulations suggest the optimization of the ATP usage as a cofactor in various pathways as well as the optimization of its use to phosphorylate NAD^+.

The second gene, which encodes for dihydroxy-acid dehydratase, involves dehydratation of 2,3-dihydroxy-isovalerate. This metabolite is the product of the reduction of 2-acetolactate by the acetohydroxy acid isomeroreductase enzyme (reaction *"1.1.1.86a"* in *i*Syf715) where NADPH is the reductant agent. This gene is a candidate to be removed from the genome of the cyanobacterium, because NADPH would remain available for hydrogen synthesis.

The results of the triple knockouts analysis shows that the best strategy is removing the *ndkR* gene encoding for nucleoside diphosphate kinase (reaction *"2.7.4.6h"* in *i*Syf715 and ORF ID: *Synpcc7942_2497*), the *ilvC* gene which codes for acetohydroxy acid isomeroreductase (reaction *"1.1.1.86a"* in *i*Syf715 and gene *Synpcc7942_1552*), and gene *glgA* for ADP-glucose transglucosylase (reaction *"_2.4.1.21b"* in *i*Syf715 and ORF ID: *Synpcc7942_2518*). With these knockouts, the hydrogen evolution was the highest, 0.00963842 mmol gDW^{-1} h^{-1} with a growth rate of 0.01602 h^{-1}. The mutations of *ndkR* and *glgA*, impact on the minimization of ATP consumption. *IlvC* gene mutation increases the NADPH accessibility for hydrogen production.

We have seen that through MOMA analysis, the most frequent genetic deletions that enhance hydrogen production and biomass evolution, figure 4.18, include *glgA* gene as the most frequent mutation, followed by mutations in *ilvD* and *ubiG* genes.

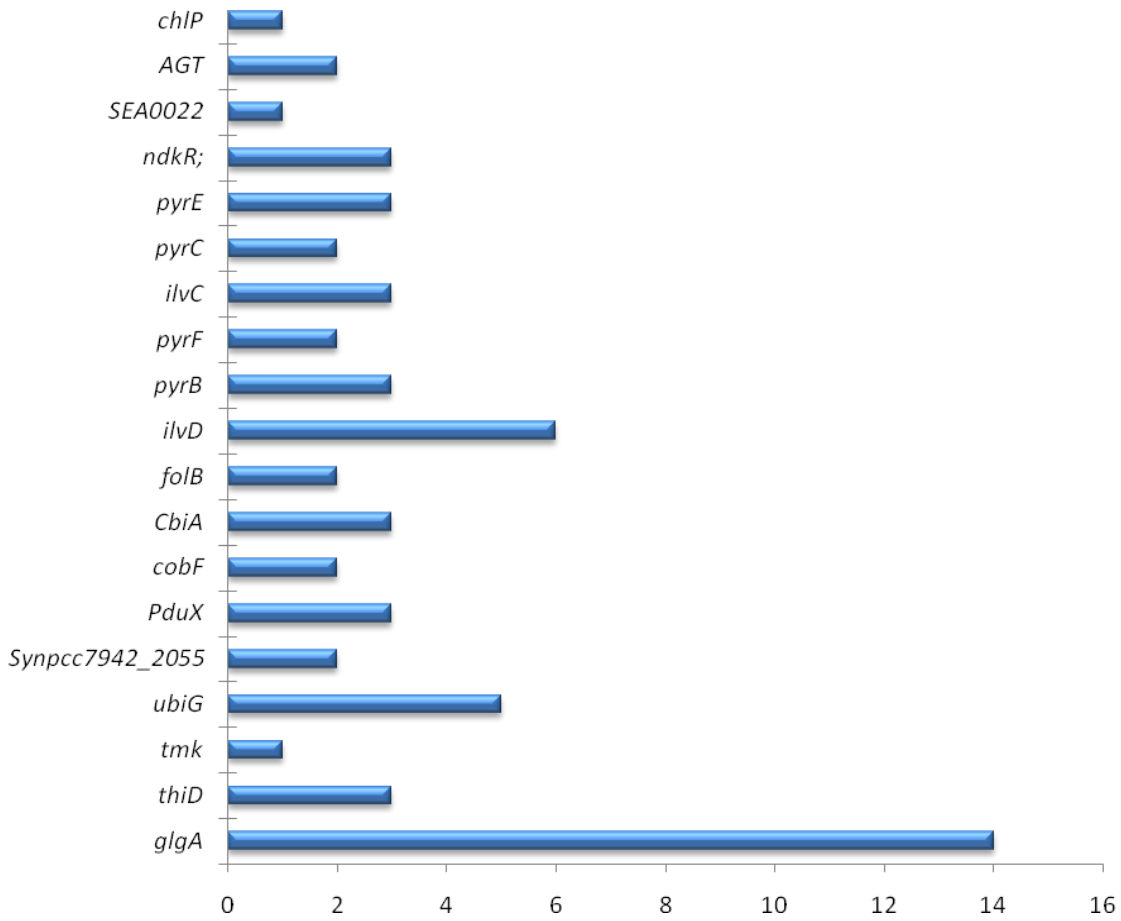


Figure 4.18. Most frequent mutations, proposed by MOMA analysis, which include the top 10 candidates of the single, double and triple knockouts for maximize hydrogen evolution.

As done previously, the estimation of theoretical maximum yield of hydrogen production was performed by FBA. Here, we compared our data with results reported by Asada *et al.* in a genetically modified *Synechococcus elongatus* PCC7942 strain (Asada *et al.*, 2000). This study demonstrated that transformed cells were capable of producing hydrogen by receiving the electrons from photosynthesis and catalyzed by clostridial hydrogenasa I protein. Although these authors did not report values of biomass growth rate, this research could serve, once again, as a reference to validate our metabolic models. We found that for CO_2 uptake rate of 1.99 mmol gDW^{-1} h^{-1} and an incident irradiance of 1.96 mE gDW^{-1} h^{-1}, our prediction for theoretical hydrogen productivity behaves as shown in figure 4.19.

For the maximum biomass growth (0.05987 h^{-1}), the theoretical synthesis of hydrogen was $6.04 \cdot 10^{-10}$ mmol gDW^{-1} h^{-1}. However, for a value of 0.0278 h^{-1} (very close to the maximum growth) the corresponding hydrogen evolution match perfectly with

the value reported by Asada *et al.* (≈0.002899 mmol gDW^{-1} h^{-1}) (Asada *et al.*, 2000) in cell-free extracts of *Synechococcus* wild-type cells (assuming that cell concentration is 10^8 cells ml^{-1} that correspond to an OD_{730} of 1.0 (Deng and Coleman, 1999)). Furthermore, H_2 production in cell-free extracts of transformant cells (0.0569 mmol gDW^{-1} h^{-1}, Asada *et al.*, 2000) is reached by *i*Syf715 with a growth rate of 0.02 h^{-1}, which is a coherent growth rate in this specie (Bertilsson *et al.*, 2003; Rosales-Loaiza *et al*., 2005; Fu *et al*., 2007).

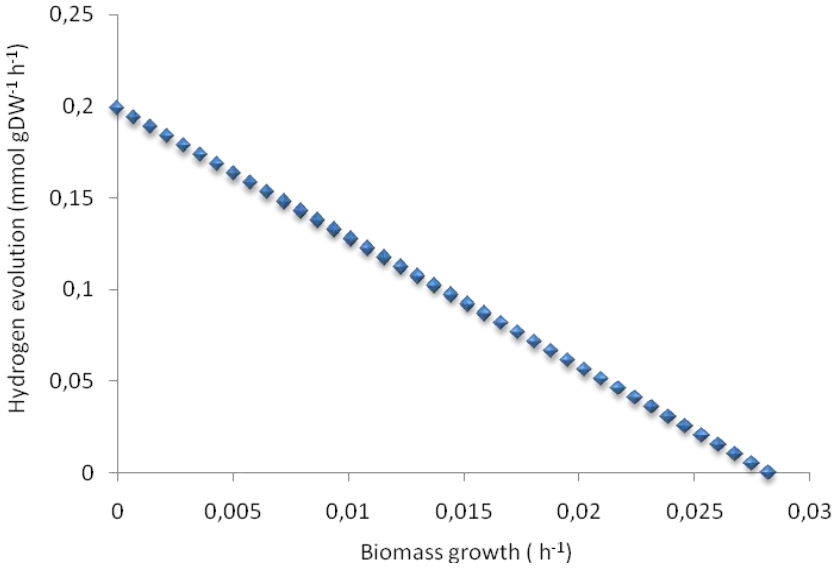


Figure 4.19. Theoretical productivity of hydrogen as predicted by using *i*Syf715. Maximum hydrogen production is shown as a function of minimal demand on biomass formation under autotrophic growth.

In figure 4.19, the hydrogen evolution maximal value for *i*Syf715 was predicted to be larger when the biomass growth is minimal. As can be seen, the maximum hydrogen evolution for growing cells was linearly related to the available electrons from NADPH, basically, generated by photosynthesis. As in the ethanol case, the intercept to the x-axis, where no electrons are available to support H_2 production, shows the reducing power that allows biomass growth alone. Also, the slope indicates the theoretical maximum that can be achieved if the electron flux is channeled towards the protons reduction. This graph allows us to evaluate what is the potential of this organism if all energy available on the cell was directed to hydrogen production. This value is obviously unreachable in naturally occurring cells, but its value is interesting to study what is the theoretical maximum of the hydrogen production.

To address if H_2 evolution is a direct function of the cyanobaterium carbon substrates, we analyzed the influence of both CO_2 and HCO_3^- uptakes rates into maximal hydrogen synthesis rate (figure 4.20). As we could imagine from the lineal dependence of carbon inputs and hydrogen evolution, higher values of hydrogen evolution correspond to higher carbon inputs. Interestingly, with CO_2 inputs higher of 15 mmol gDW^{-1} h^{-1}, hydrogen drains much more from CO_2 than from HCO_3^-.

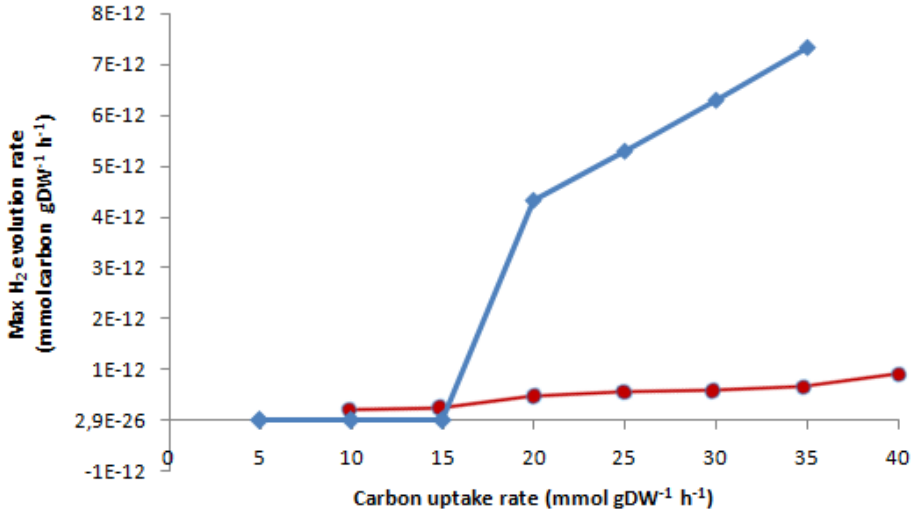


Figure 4.20. Relationship between carbon sources and maximal H_2 production rate. This last parameter was normalized per number of carbon atoms. The blue diamonds and red circles correspond to the influence of CO_2 and HCO_3^- uptake rates in maximal H_2 production rate, respectively.

4.3 Conclusions

Constraint-based methods, such as FBA and MOMA, are useful for aiding in metabolic engineering approaches. Furthermore, these methods have been used to assess possible genetic modifications that redirect the metabolic flux towards biological targets.

In the present study, we ventured into the metabolic capabilities of *Synechococcus elongatus* PCC7942, as a potential value added-compounds factory. We have demonstrated the use of this specie as a novel photon-fuelled production platform in order to generate metabolites of interest such as ethanol, higher chain alcohols, fatty acids and hydrogen. These novel studies can help discover strategies and understand experimental efforts that have been taken already in this organism for the synthesis of ethanol, isobutanol and hydrogen (Deng and Coleman, 1999; Atsumi *et al.*, 2009b; Asada *et al.*, 2000).

In our studies we have shown the enormous advantages of the introduction of non-intuitive perturbations in the metabolic network in order to improve the synthesis of these metabolites. Such as in the case of ethanol, lipids and higher chain alcohols synthesis.

It is important to note that while these predicted genotype-phenotype relationships demonstrate a compelling prospective use of the network, the majority of the proposed mutant phenotypes remain to be validated experimentally. Nevertheless, these predictions could be used to help define the scope and expected outcomes of such future studies.

4.4 Methods

4.4.1 Minimization of metabolic adjustment

In order to identify identity the gene knockout targets for the enhanced production of metabolites of interest, the method of minimization of metabolic adjustment (MOMA) (from Segrè *et al.*, 2002) using quadratic programming (QP) was employed. These calculations were performed with the OptGene software (Patil *et al.*, 2005), currently, available online at Biomet Toolbox (Cvijovic *et al.*, 2010) (http://www.sysbio.se/BioMet).

This tool has the advantage of simultaneously could help us to design higher production strains and couple metabolites production to growth rate, unlike other traditional computational tools. MOMA provides a mathematical approach based on the assumption that the mutant organism initially remains as close as possible to the organism in its native or wild state, in terms of values of metabolic fluxes (Segrè *et al.*, 2002).

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

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

is minimized.

Expressed as a standard QP problem, the objective is to minimize f (x) on a set of linear constraints:

$$f(x) = Lx - \frac{1}{2} x^T Q x$$

In where, the vector *L* of length *N* and matrix *Q* of *NxN* defined linear and quadratic part of the objective function, respectively, and x^T is the transpose of x. By observing that the minimization of the Euclidean distance is equivalent to minimizing its square and the constant terms can be omitted from the objective function, one can choose *Q* to be an unit matrix *NxN* and set *L =-W* and thus reduce of minimize D(w,x) to minimizing f(x). Here, we are interested in the case $w = v^{wt}$ (flux vector of the organism in the wild-type), *i.e.*, finding the point u^j in Φj that is closer to the point of wild-type (Segrè *et al.*, 2002).

Finally, the mathematical model is as follows:

Minimize $Z = (V_{ko} - V_{wt})^T (V_{ko} - V_{wt})$

Subject to $S \cdot v_j = 0 \quad \forall j \in N$

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

$v_{j,rev} \in R$

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

and $v_j = 0$ as a result of knockout

Where V_{ko} refers to the vector that reference the reactions of the organism in its mutant status. V_{wt} refers to the vector of the flux values of the reactions of the organism in its natural state. *S* represents the stoichiometric matrix of the system and *j* the indices of eliminated reactions results from mutations imposed on the biological system.

As part of input file format of this software, in all cases we defined the “design objective function" as bio-products synthesis whereas the biomass synthesis was the “biological objective function”.

4.4.2 Converting units of production rates to flux values

The production rate values of the higher alcohols in *E. coli* genetically engineered (Atsumi *et al*. 2008) were converted to flux values. For this purpose, we assume that

OD_{600} of 1.0 is equivalent to approximately 0.39 g L^{-1} of dry weight (Glazyrina *et al.*, 2010).

We calculated that the isobutanol synthesis rate in *Synechococcus elongatus* PCC7942 genetically engineered (Atsumi *et al*., 2009b) was ≈1.08 10^{-4} mmol gDW^{-1} h^{-1}. For that, we assume that OD_{730} of 1.0 is equivalent to approximately 10^{8} cells ml^{-1} (Deng and Coleman, 1999).

Chapter 5. Phenotypic phase plane analysis of *Synechococcus elongatus* PCC7942

Mathematical modeling of metabolic functions has been developed over the past several decades. One of the most commonly used methods is linear programming, implemented on powerful tools, such as FBA. This implementation has proven useful for the analysis of genome-scale reconstructed metabolic maps, an approach with considerable potential for the analysis of metabolic functions. However, its use is restricted to compute the flux distribution for a single growth condition with well defined constraints. We hereby present phenotypic phase plane analysis that gives the possibility to compute all biologically-feasible growth conditions defined by varying more than one parameter.

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

Triana J, Montagud A, Gamermann D, Fernández de Córdoba P, Urchueguía JF. ***In silico* analysis for bio-products synthesis through genome-scale reconstruction of the *Synechococcus elongatus* PCC7942 metabolic network**. *In preparation.*

Chapter 5. Phenotypic phase plane analysis of *Synechococcus elongatus* PCC7942

5.1 Introduction

Many authors believe the genome-scale metabolic maps and its analysis conducted by the FBA are the keys to studying systems biology (Thiele and Palsson, 2010; Orth *et al.*, 2010; Blazier and Papin, 2012). The genome-scale metabolic networks can be reconstructed from annotated genome sequence data, biochemical literature, bioinformatic analysis, and strain-specific information. Flux balance analysis has been successfully tested and validated on several metabolic networks (Orth *et al*., 2010), despite assuming the unrealistic situation of a "pure steady-state". As mentioned, this method has been widely used for describing and/or predicting metabolic functions and capabilities of a given organism. However, one of the drawbacks of using this tool is that it typically offers a limited view of the metabolic behavior of the system, a single snapshot. This is mainly due to large-scale metabolic models that are subject to a well-defined set of constraints under a single growth condition, in order to be able to have a single solution, not a space of solutions. Edwards *et al*. developed an interesting approach to generate a global view of the optimality properties of a network (Edwards *et al*., 2002). They suggest a different approach using FBA, named phenotypic phase planes (PhPPs) analysis, in which the optimal metabolic flux distribution is charted onto a single plane, which is defined by the availability of two external substrates. PhPPs is a linear optimization procedure which can be used to study the value of the objective function (a desired phenotype) as two constraints (external substrates) vary simultaneously. If the constraints vary, the shape of the cone that contains the solution space changes and the optimal flux vector may qualitatively change.

This method allows for identification, in this plane, of a finite number of qualitatively distinct patterns of metabolic pathway utilization (metabolic phenotypes), dividing it into discrete phases. In this parametric sensitivity analysis the shadow prices throughout the two-substrate space are computed and used, in the form of isoclines, to identify phase planes (Edwards *et al*., 2002). These regions can be used to classify the qualitative state of metabolic network in a range of two

constrained fluxes. Shadow prices define the incremental change in the objective function when an exchange flux is incrementally changed.

One important feature of the PhPP is the line of optimality (LO), which is defined as the line that represents the optimal relation between the two analyzed parameters (Edwards *et al*., 2002). When the cell is operating along the LO, calculated when the growth rate is used as the objective function, the cell is growing at its maximal biomass yield.

The main goal of this chapter is, first to generate *Synechococcus elongatus* PCC7942 PhPPs for the growth and production of the previously studied bio-products, *viz.*: ethanol, higher chain alcohols, lipids and hydrogen, with CO_2 as a carbon source at all levels of illumination; and finally, the study of the resulting optimal metabolic phenotypes.

5.2 CO_2 and light phenotype phase plane for biomass growth rate

In the previous chapter, parametric sensitivity of the optimal properties of cyanobaterium metabolic networks, such as robustness analysis, was assessed using only one parameter. However, PhPP analysis is a valuable method to characterize different functional states by projection of the steady-state solution space in two or three dimensions (Edwards *et al.*, 2002). Also, this analysis is developed to consider all possible outcome variations studying two constraining environmental variables. Thus, we decided to vary the CO_2 uptake rate and light input simultaneously then analyze its contributions to the objective function, in this case biomass growth.

The *i*Syf715 metabolic network was used to generate a PhPP. We calculated all points in a plane formed by using the light input on the x-axis and the CO_2 uptake rate on the y-axis. As shown in figure 5.1, we arbitrarily constrained our sensitivity analysis to a CO_2 uptake rate range from 0 to 95 mmol gDW^{-1} h^{-1} and light uptake rate from 0 to 52 mE gDW^{-1} h^{-1}. We varied the light input constraint among both photosystems: photosystem I (reaction *"_lightI"* in *i*Syf715) and photosystem II (reaction *"_lightII"* in *i*Syf715). This type of analysis allowed the steady-state flux distributions to be divided into a finite number of regions, each one with similar metabolic flux patterns and characterized by equivalent shadow prices.

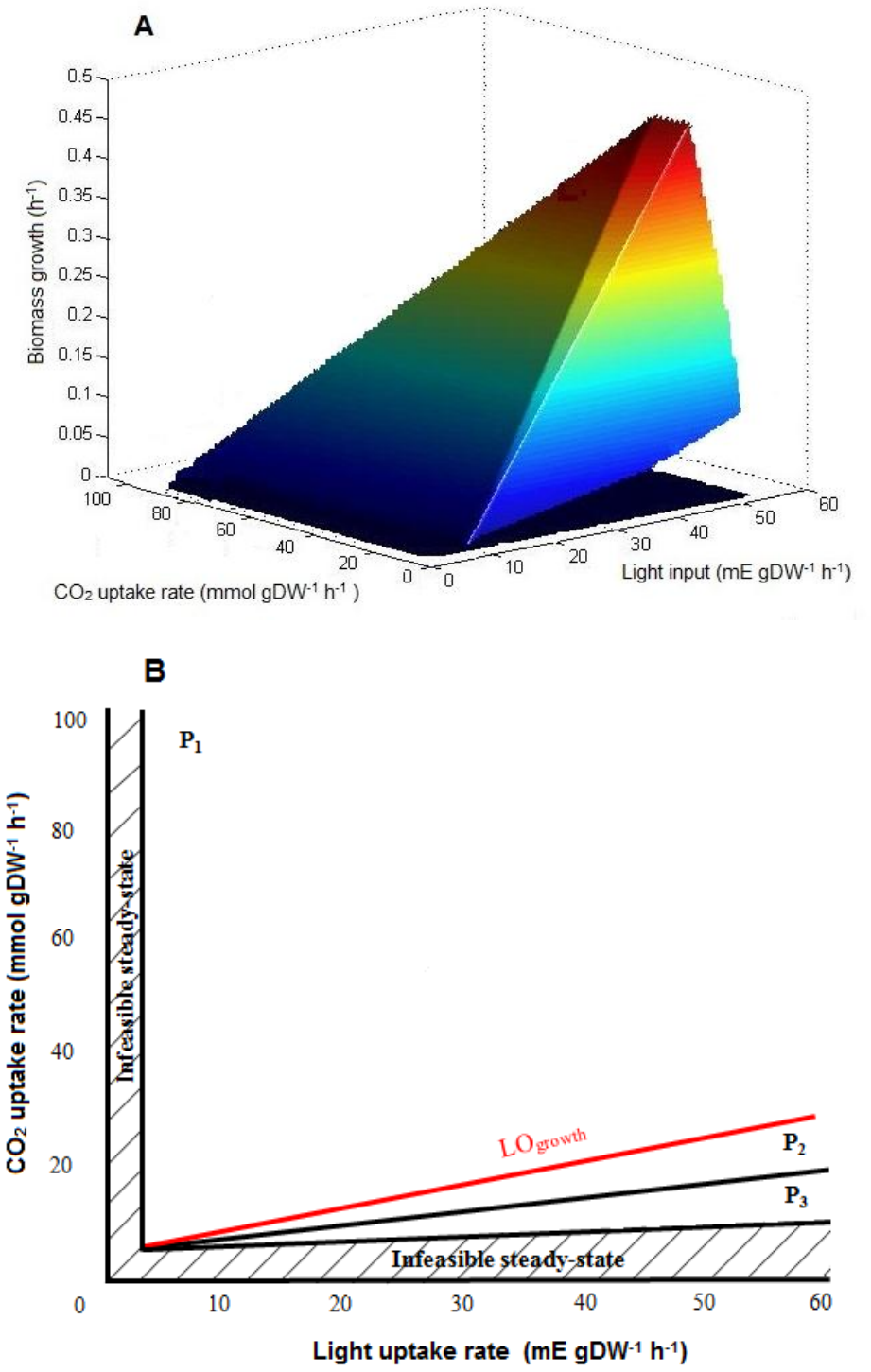


Figure 5.1. The *i*Syf715 CO_2-light phenotypic phase plane for biomass growth. **A.** The three-dimensional PhPP. **B**. A two-dimensional projection of the 3-D polytope in panel. The line of optimality is shown in red.

As can be seen in figure 5.1-A, the surface of the three-dimensional PhPP corresponds to the maximum growth rate allowable for each pair of CO_2 and light uptake rates in the x-y plane. The two-dimensional projection of the PhPP (Figure 5.1-B) has been divided into three different optimal phenotypes or phases ($P1_{CO2,light}$ -

$P3_{CO2,light}$), where optimal use of the metabolic network is fundamentally dissimilar. This projection allowed the metabolic fluxes distributions to be split, each one characterized by the effect that CO_2 uptake rate and light input had on the objective function.

Due to stoichiometric limitations, there also are two regions of the PhPP with infeasible steady-state flux distributions; the small areas along the x-axis and y-axis (Figure 5.1-B). Cell growth is infeasible in both regions since there are incompatibilities between the demands of energy and carbon sources, for biomass formation, and the low uptake rates of light (area along the y-axis) and CO_2 (area along the x-axis). If the substrates are taken up at the rates represented by these points, the metabolic network is not able to produce the mass, energy, and redox constraints while generating biomass. The metabolic network can only transiently operate in such a region (Palsson, 2006), not as a steady-state solution.

The phase $P1_{CO2,light}$, where the isoclines are vertical, we found the so-called phase of single substrate limitation (Edwards *et al*., 2002). In this case the slopes (α) of isoclines (see methods) are infinite. These phase´s arise because the shadow price for CO_2 goes to zero, and therefore has no value for the cell. In this phase, the CO_2 was provided beyond the optimal amount required by the metabolic network to support growth.

The LO represents the optimal light input for the complete carbon fixation in order to support the maximal biomass yield. Below the LO there is another optimal metabolic phenotype: $P2_{CO2,light}$. In this plane the absolute value of α is greater than unity, therefore, the photons that come into the system could be more valuable towards achieving the objective. This phenotype was characterized by CO_2 limitation, so that, the metabolic flux distribution might be useful to interpret the operational principles of this phenotype. Thus, the FBA simulation was carried out by fixing the light uptake rate at 30 mE gDW^{-1} h^{-1} and for all CO_2 uptake rates across all phases. Results show that the optimal fluxes of Calvin-Benson cycle and photosynthetic machinery were sensitive to the CO_2 uptake rate in this region. Additionally, we see that the O_2 evolution decreases as the CO_2 uptake decreases. This could be due a possible diversion of the optimal carbon flux in the synthesis of reduced by-products,

for instance, ethanol. The optimal flux in the acetyl-CoA formation and in the reduced branch of the pentose phosphate pathway was sensitive to CO_2 uptake, as was the gluconeogenesis increase with increasing CO_2 uptake rate.

The $P3_{CO2,light}$ region was also characterized by CO_2 limitation and was defined as a futile phase with excess photons. These "futile" phases are characterized by a negative effect of one of the substrates on the objective function (Edwards and Palsson, 2000). Interestingly, the shadow price for ATP is zero, suggesting a situation where the ATP production by ATPase complex (reaction *"_3.6.3.14"* in *i*Syf715), which coupled with photosynthesis, would exceed the biosynthetic/maintenance demands for biomass growth. The metabolic network would dissipate the photons excess by increasing the ATPase flux, which would generate ATP but also reduce the $NADP^+$ by ferredoxin-$NADP^+$ reductase (reaction *"_1.18.1.2"* in *i*Syf715). At the same time, the transhydrogenase reaction was used to convert the excess NADPH to NADH. When the CO_2 uptake rate and the light input force the lower limits of $P3_{CO2,light}$, the produced ATP and NADH formed reduced by-products such as hydrogen or ethanol, and no biomass was generated. The shadow prices for both by-products, indicates that they have no value to the cell in this region, and therefore, they were secreted to optimally balance the redox potential of the cell. Thus, the metabolic operation in this phase is wasteful and probably physiologically unstable; this might explain the observed post-peak segment in robustness analysis presented in chapter 3.

5.3 CO_2 and light phenotype phase plane for alcohols production

Use of *Synechococcus elongatus* PCC7942 metabolism was assessed for synthesis of ethanol, 2-methyl-1-butanol, 3-methyl-1-butanol, isobutanol and 1-propanol, with CO_2 and light as substrates. For each environmental variable, a PhPP was plotted to map the optimal metabolic characteristic for the alcohols productions.

Once again, we used the *i*Syf715 metabolic network to generate PhPPs and arbitrarily constrained our sensitivity analysis to CO_2 uptake rate ranging from 0 to 95 mmol gDW^{-1} h^{-1} and light uptake rate from 0 to 52 mE gDW^{-1} h^{-1}. We found two regions in the PhPPs, each one characterized by a qualitatively different optimal use of metabolic pathways (see figure 5.2).

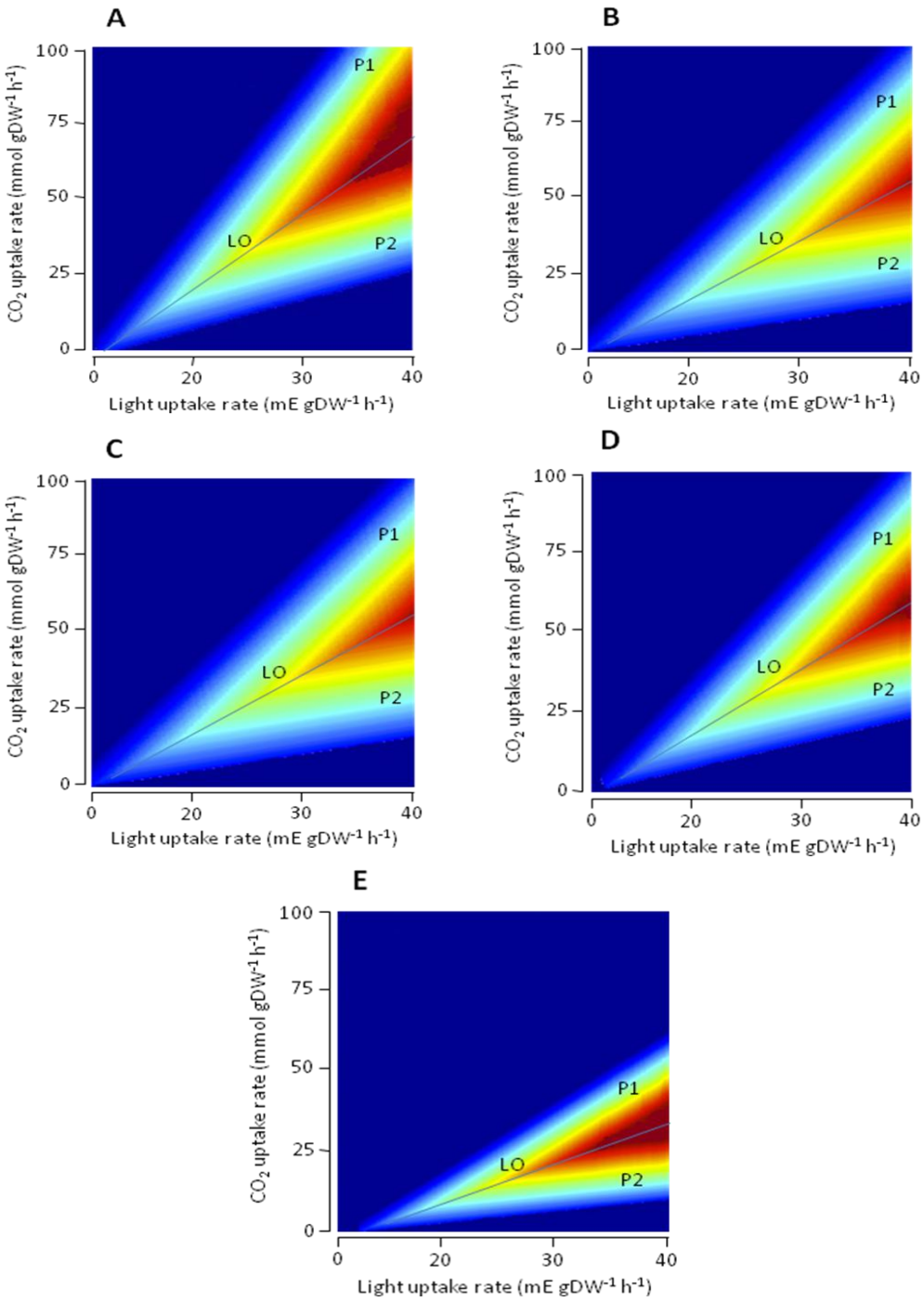


Figure 5.2. A two-dimensional projection of the 3-D polytope in panel of *i*Syf715 CO_2-light phenotypic phase plane for ethanol synthesis (A), 2-methyl-1-butanol (B), 3-methyl-1-butanol (C)**,** isobutanol (D) and 1-propanol (E). The line of optimality (LO) is shown in every projection.

In this case, all the CO_2-light PhPPs consist of two phases, each one corresponding to a futile region separated by the LO. Points on the line of optimality represent the optimal light input required for the complete carbon fixation to maximize the alcohols synthesis.

The first phase is a futile region ($P1_{CO2,light}$), that is defined by an excess of CO_2. A defining characteristic of this futile phase was that the shadow price of the 3-phosphoglycerate is zero. This would indicate a condition where the excess of CO_2 was dissipated by increasing the gluconeogenesis pathway flux. In the lower limits of the plane, the *in silico* analysis shows that an increase in the CO_2 uptake rate, at a fixed light input (30 mE gDW^{-1} h^{-1}), increases the objectives functions until a threshold is reached. At CO_2 uptake rates higher than this value, inhibitory effects on the alcohols synthesis are observed. Here, the available photons are not enough to synthesize the ATP and NADPH molecules needed for ethanol and higher chain alcohols formation and hence reduces its synthesis.

Similar to $P3_{CO2,light}$ region for biomass growth, the $P2_{CO2,light}$ region was characterized by CO_2 limitation with excess photons. This variable has a positive shadow price and it is inhibitory toward obtaining the biological objective. Here, too much photons are taken up relative to CO_2 and ethanol and higher alcohols synthesis rates drop due to forced dissipation of the excess light. We observed the effects of the zero value in the shadow prices for ATP and GTP. The metabolic network dissipated the excess photons by increasing the ATPase flux as well as the nucleoside-diphosphate kinase flux (reaction *"2.7.4.6c"* in *i*Syf715). When the CO_2 uptake rate and the light input defined the points on the lower limits of $P2_{CO2,light}$, the produced ATP and GTP were directed to the formation of a by-product like glycolaldehyde or to the export of organic carbon through the glycolysis, and, therefore, no ethanol was generated. That is consistent with the steady-state metabolomic and transcriptomic analyses of this cyanobacterium under low CO_2 acclimation (Schwarz *et al.*, 2011). There are distinguishable physiological differences between this phase and $P3_{CO2,light}$ region for biomass growth in terms of their secretion profiles.

There also are two small areas along the x-axis and y-axis of the PhPP with infeasible steady-state flux distributions (areas in dark blue). The low uptake rates of light (area along the y-axis) and CO_2 (area along the x-axis) do not supply enough source material to get the system working.

5.4 CO_2 and light phenotype phase plane for lipids synthesis

We analyzed the steady-state fatty acids profile productions of *i*Syf715 under the influence of variations in carbon (CO_2) and energy (light) uptake rates. Restricted parameters were bounded, one more time, from 0 to 95 mmol gDW^{-1} h^{-1} for CO_2 and 0 to 52 mE gDW^{-1} h^{-1} for light. The two-dimensional projection of the PhPPs result in three different optimal phenotypes (see figure 5.3).

The LO represents the optimal light input for the complete carbon fixation in order to support the maximal lipids yields. The phase $P1_{CO2,light}$ is characterized as a futile region. This phase has isoclines with positives slope and are thus phenotypically unstable, for example; if the light input is fixed, then lowering the CO_2 uptake rate toward the LO will increase the fatty acids synthesis rates. Thus, the maximum allowable CO_2 uptake rate would not be chosen to maximize the objective function.

The $P2_{CO2,light}$ metabolic phenotype, where α value is negative, is defined by dual limitation of the substrates. In this phase, the ratio of CO_2 uptake rate and light uptake rate is lower than on the LO. Again, the FBA simulation was performed by fixing the light uptake rate at 30 mE gDW^{-1} h^{-1} and for all CO_2 uptake rates across all phases. The cell is CO_2 limited and, NADH and NADPH are available in excess, meaning that the fatty acids synthesis would improve if its availabilities decreased. In order to maintain the cell's redox balance, the excess of these cofactors must be oxidized. In the case of NADH, this is done through the production of ethanol, which begins to be secreted in this phase. The NADPH oxidation is achieved by increasing Calvin-Benson cycle flux with the excess of photons.

As the ratio of CO_2 and light uptake rates is further decreased, cyclic electron transfer, which involves PSI ferredoxin, and ferredoxin-$NADP^+$ reductase become essential for biological objective in $P3_{CO2,light}$ region. The flow of carbon compounds decreases into the central metabolism, nonetheless, the intermediary anabolism, dependent on redox cofactor, becomes more effective.

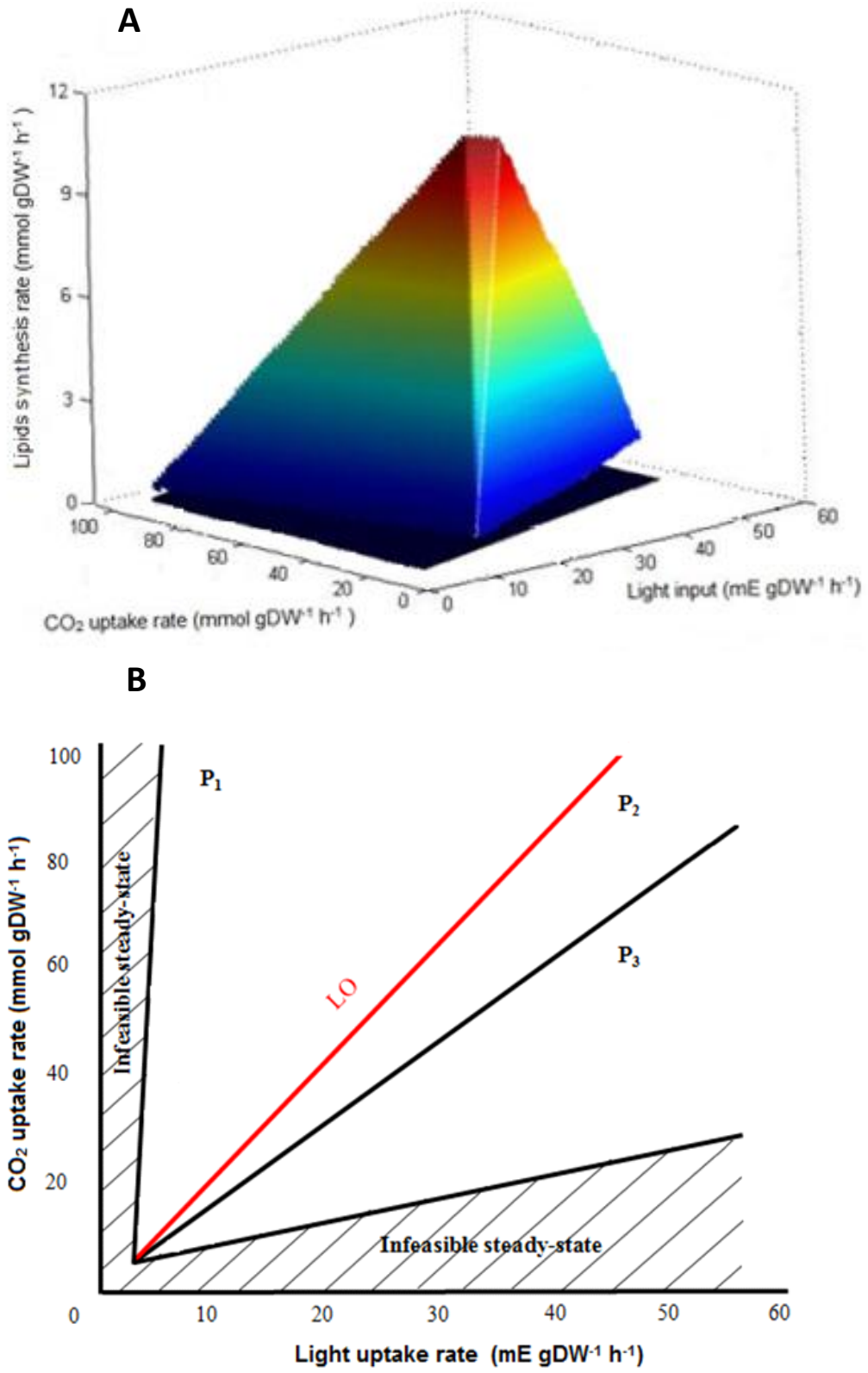


Figure 5.3. The *i*Syf715 CO_2-light phenotypic phase plane for lipids synthesis. **A.** The three-dimensional PhPP. **B**. A two-dimensional projection of the 3-D polytope in panel. The line of optimality is shown in red.

Once again, there are unfeasible steady-state flux distributions along the x-axis and y-axis of the PhPP. The low uptake rates of light (area along the y-axis) and CO_2

(area along the x-axis) do not supply enough source material to get the system working.

5.5 CO_2 and light phenotype phase plane for hydrogen evolution

Parametric sensitivity of the hydrogen optimal production in cyanobaterium metabolic network was assessed. We carried out a PhPP analysis to model metabolic functional states and analyze the contribution of CO_2 uptake rate and light input to this objective function. These parameters were constrained between 0 to 95 mmol gDW^{-1} h^{-1} and 0 to 52 mE gDW^{-1} h^{-1}, respectively.

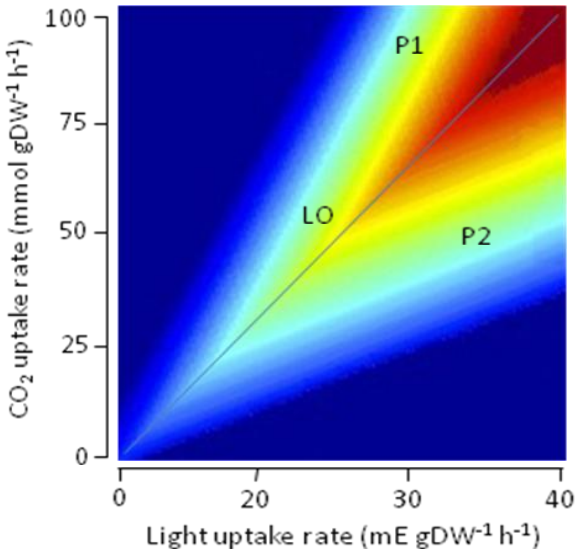


Figure 5.4. The *i*Syf715 CO_2-light phenotypic phase plane for hydrogen evolution. The line of optimality is shown between both regions.

With this approach, two futile phases were identified: $P1_{CO2,light}$ and $P2_{CO2,light}$, which are separated by the LO line (figure 5.4). As in previous analyses, the first phase was characterized by the excess of CO_2 and by pyruvate shadow price being zero. This metabolic condition suggests that the excess of CO_2 was dissipated by increasing the TCA cycle flux, specifically, the reaction catalyzed by isocitrate dehydrogenase ($NADP^+$) (reaction *"1.1.1.42a"* in *i*Syf715). In the lower limits of the plane, an increase in the CO_2 uptake rate, at a fixed light input (30 mE gDW^{-1} h^{-1}), augments the hydrogen synthesis rate until a threshold is reached. As in the alcohols PhPPs analyses, at CO_2 uptake rates higher than this value, inhibitory effects on the hydrogen synthesis are observed. Here, the available photons are not enough to synthesize the NADPH molecules needed for hydrogen formation and hence reduces its synthesis.

The line of optimality, between the phases, and two small areas along the x-axis and y-axis of the PhPP with unfeasible steady-state flux distributions, can be seen in the figure (areas in dark blue).

In $P2_{CO2,light}$, much more photons are being taken in relation to CO_2 and the hydrogen synthesis rates drops due to forced dissipation of the excess light. As in the alcohols PhPPs analyses, we observed the effects of the zero value in the shadow prices for ATP and GTP and the formation of by-product. We also observed a considerable decrease of the reducing branch of the pentose phosphate pathway fluxes and subsequent reduction of NADPH synthesis by this route.

5.6 Conclusions

The reconstruction of genome-scale metabolic models and their applications represent a great advantage of systems biology. Some methods used to analyze them, such as: phenotypic phase planes analysis, allow us for the prediction of a cell's phenotype based on its genotype and environmental conditions (Duarte *et al.*, 2004). We have proven the use of PhPP analyses to improve growth rates or useful metabolite production in biological systems. This approach consists on execution of several FBA on the model, and simultaneously the value of the objective function (or by-product fluxes) is tested when certain nutrient uptake constraints are co-varied.

With this approach we have further explored the capabilities of *Synechococcus elongatus* PCC7942 metabolism. We can generate new hypotheses on how this organism operates, and, more importantly, we have gained insights into what distinct patterns of pathway utilization the model uses when optimizing for different well-defined objective functions.

5.7 Methods

5.7.1 Computing the Phase Plane

To compute PhPPs we have followed the methodology described in (Edwards *et al.*, 2002). In general, as this analysis is based on the FBA is a linear optimization procedure and is based on computing shadow prices as defined in classical linear programming duality theory (Varma *et al.*, 1993b). The delimitation of the phases are defined by

calculating the shadow prices, which define the incremental change in the objective function if a constraining exchange flux is incrementally changed.

Mathematically, the shadow prices (π_i) are defined as the derivatives of the objective function (Z) at the boundary with respect to an exchange flux (b_i) (Edwards *et al*., 2002):

$$\pi_i = \frac{\partial Z}{\partial b_i}$$

The values of the shadow prices were used to compute the slope of isoclines within each region of the PhPP. The isoclines were also defined to interpret the metabolic phenotype and to represent the locus of points within the two-dimensional space that provide for the same value of the objective function (Edwards *et al*., 2002). A ratio of the shadow prices was used to define these slopes (α):

$$\alpha = -\frac{\gamma_x}{\gamma_y}\Big|_z = -((-dZ/db_x)/(-dZ/db_y))\Big|_z = -\frac{db_y}{db_x}\Big|_z$$

The negative sign in the equation was introduced in anticipation of its interpretation.

To generate all the regions in the PhPP we used MATLAB (The MathWorks Inc., Natwick, MA).

Chapter 6. Metabolome dynamic upon inorganic carbon acclimation

Sometimes, the high-dimensionality of biological networks has hampered the elucidation of the principles of certain cellular processes functioning. However, the topological studies of these interactions are interesting from diverse points of views. One of these is the integration of *-omics* data with the connectivity features of the network. Thus, the combination of the transcriptome data with metabolome information provides the opportunity to identify regulatory principles under certain physiological conditions.

Part of the contents of this chapter is based on parts of the following journal article:

Triana J., Montagud A., Gamermann D., Fernández de Córdoba P., Urchueguía J. F. ***In silico* analysis for bio-products synthesis through genome-scale reconstruction of the *Synechococcus elongatus* PCC7942 metabolic network**. In preparation.

Chapter 6. Metabolome dynamic upon inorganic carbon acclimation

6.1 Introduction

The high-throughput analytical chemistry´s emergence has forced changes in classical reductionist analyses towards new integrative approaches (Hood *et al.*, 2004). Nowadays, many biologists acknowledge the use of bioinformatics, mathematical models and computer simulation to make integrative analyses of the cell as a system. The analysis of multiple entities, such as genes, proteins or metabolites, has become a critical issue for the future development of biology (Palsson, 2006).

The relationship between genetics and cellular functions is hierarchical and involves many steps which are well described in the central dogma of molecular biology (Crick, 1958; Crick, 1970). However, genome-wide quantification of genes and the mapping of their functional relationship have been great challenges in systems biology. Several techniques have been used in order to analyze and extract knowledge from the vast amounts of -*omics* data, such as transcriptome and metabolome, yet the problem of integrating these layers of information and elucidate new system's principles is still unresolved.

Oliveira *et al.* proposed a new approach that considerably intents to reduce the dimensionality of the data analysis problem. These authors propose the use of genome-scale bio-molecular interaction information, such as physical or functional interactions between metabolites in reconstructed metabolic networks, constraining the solution space and revealing regulatory principles under certain physiological conditions (Oliveira *et al.*, 2008). The developed algorithm, named Reporter Features, allows for the identification of cellular regulatory focal points (*i.e. reporter features*). In the case of the metabolic network, the algorithm would unveil *reporter metabolites*, which can represent regulatory hubs of the metabolism.

Following the scope of studying the characteristics of *Synechococcus elongatus* PCC7942 as photo-biological production platform, in this chapter we have explored system-wide variations upon inorganic carbon acclimation. The transcriptional phenotyping of these inorganic carbon variations was reported by Schwarz *et al.* (Schwarz *et al.*, 2011).

6.2 *i*Syf715 as bio-molecular interaction network for integration

We applied the Reporter Feature algorithm to analyze the transcriptional changes in *S. elongatus* PCC7942 cells that had been shifted from high to low CO_2 levels of growth conditions.

In order to effectively perform modular analysis we used *i*Syf715 as a metabolites interaction network to integrate transcriptional information for inorganic carbon acclimation (Schwarz *et al.*, 2011). As we have seen in previous chapters, a perturbation in the CO_2 input flux of *i*Syf715 metabolic model drives big changes in the flux distribution through the network. This observation led us to perform analyses that help to identify around which metabolites the transcriptional changes are significantly concentrated. These metabolites, termed *reporter metabolites,* represent key regulatory nodes in the landscape of the metabolism. For this purpose, gene-reactions associations derived from the metabolic model reconstruction process were used (see method).

Synechococcus elongatus PCC7942 is strictly photoautotrophic and cannot use organic carbon such as glucose or nitrogen sources. This causes its metabolism to be less flexible to the availability of other nutrients, and much easier to elucidate. This advantage was taken into account by Schwarz *et al.* (2011), to analyze the acclimation of *S. elongatus* PCC7942 to low inorganic carbon (Ci). In this case, high CO_2 condition (HC)-acclimated wild type cells were grown for several days under 5% CO_2. Then, the exponentially growing cells under this regimen were shifted to low ambient CO_2 (LC) conditions (grown under 0.035% CO_2 by bubbling with ambient air), and the transcriptome was characterized after 6 and 24 h.

Our goals were to identify metabolites around which regulation is centered during CO_2 regime transitions and to find clusters of metabolic genes that were significantly co-regulated across these transitions (Patil and Nielsen, 2005). The analysis was performed with whole-genome *S. elongatus* PCC7942 microarrays data. For a study of the overall genome and its inorganic carbon acclimation regulation, as well as the fold change (Nodop *et al.*, 2008) gene expression levels in response to this shift refer to Schwarz *et al.*, (2011). In this chapter, we focused on the relationship between the

metabolism and its regulation, so genes with no direct relationship to a metabolic reaction were not considered.

The metabolic model file together with the p-values for differential expression under the two conditions (see methods) was used to rank the *reporter metabolites*. The top 20 ranking metabolites out of 52 significant hits ($p<0.05$) are shown in Table VI.A.

Drastic metabolic shifts are observed with the change from growth on HC to growth on LC for 6 and 24 h. In particular, *reporter metabolites* were found scattered across the whole metabolism (see Table VI.A), thus offering a global view of the transcriptional response in the metabolic network. During growth on high CO_2 condition, all building blocks for biomass synthesis can be derived from the coupling of mechanism for capturing CO_2 and HCO_3^- and steady-state photosynthetic carbon fixation. Although it is known that Ci uptake system is both constitutively expressed (Badger and Gallagher, 1987) as inducibly expressed, e.g. high affinity HCO_3^- uptake system, during growth at low Ci levels (Badger, 1987; Price and Badger, 1989). Thus, during growth on low ambient CO_2 condition, HCO_3^- transporters together with carbonic anhydrase form the central component of the CO_2 concentrating mechanism that support the production of the biomass precursors.

The output from *in silico* analysis accurately captures several reporter features under LC condition. For example, *H_2CO_3* and *H_2CO_3_extrac* are the reporters with the highest score and are involved in system uptake of carbon source. It is obvious that these metabolites represent regulatory hot spots in a Ci shift regime. *UDP-GlcNAc-enolpyruvate*, *N-acetylmuramic acid 6-phosphate* and *(R)-lactate* are important intermediates in aminosugars metabolism. The second and third *reporter metabolites* are substrate and product, respectively, of a reaction which is sequentially involved in *UDP-GlcNAc-enolpyruvate* synthesis. This later metabolite is implicated in redox reaction in where NADPH (cofactor required in Calvin-Benson cycle) is produced. It is valid to note that an increase in these metabolites concentration also causes an increase in the metabolic flux.

Table VI.A. The top 20 reporter metabolites from Reporter Features.

Metabolite	Number of neighbors
H_2CO_3_extrac	1
H_2CO_3	2
UDP-GlcNAc-enolpyruvate	2
cellobiose	1
phytofluene	1
tRNAtrp	1
L-tryptophanyl-tRNAtrp	1
N-acetylmuramic acid 6-phosphate	1
(R)-lactate	1
2-keto-isovalerate	4
Ca^{2+}_extrac	1
Ca^{2+}	1
ADP-ribose	1
R-4'-phosphopantothenoyl-L-cysteine	1
L-glutamine	15
2-keto-3-deoxy-6-phospho-gluconate	1
dGDP	4
4-amino-4-deoxychorismate	1
5-phospho-beta-D-ribosyl-amine	2
UDP-N-acetylmuramoyl-L-alanyl-D-glutamyl-meso-2,6-diaminoheptanedioate	2

Cellobiose stands as a reporter as the cell can hydrolyze this disaccharide to beta-D-glucose, as an alternative carbon source in response to low carbon fixation flux. *Phytofluene* participate in carotenoid biosynthesis, such as zeta-carotene. This type of pigment is the precursor of others which are intimately associated with the photosynthetic reaction centers and plays a vital role in protecting against potentially lethal photooxidative damage (Koyama, 1991). Also, these carotenoids serve as light-harvesting pigments (Siefermann-Harms, 1987) and are involved in the thermal

dissipation of excess light energy captured by the light-harvesting antenna (Demmig-Adams and Adams, 1992). These pigments are possibly reporter metabolites related to inorganic carbon acclimation since the biological system would protect from the excess of photon due to the low flux of carbon fixation. *tRNAtrp* and *L-tryptophanyl-tRNAtrp* are substrate and product, respectively, of the reaction in where the L-tryptophan is activated for the protein biosynthesis. *2-keto-isovalerate* is a 2-keto acid implicated in several pathways, such as: L-valine, L-leucine and coenzyme A biosynthesis. In the majority of these pathways, the participation of this 2-keto acid is indirectly related to decarboxylation processes of other 2-keto acids, and thus the release of CO_2. *Ca^{2+}_extrac* and *Ca^{2+}* correspond to the uptake of this cofactor. It is found that a transient increase in intracellular free Ca^{2+} is also triggered following nitrogen deprivation in *S. elongates* PCC7942, which has evidenced the role of this ion in the acclimation of this cyanobacterium to nitrogen starvation (Leganés *et al.*, 2009). This effect could be also related to inorganic carbon acclimation of this biological system and would make a nice hypothesis to test in the future. *ADP-ribose* is involved in purine metabolism and is a precursor of D-ribose-5-phosphate and thereby of D-ribulose-5-phosphate. The decrease in expression levels of the enzyme ADP-ribose diphosphatase (reaction *"3.6.1.13"* in *i*Syf715) appears to be a physiological response of the system to a reduction in carbon sequestration. Another reporter metabolite is *R-4'-phosphopantothenoyl-L-cysteine* which is related with coenzyme A biosynthesis. This metabolite is decarboxylated to pantetheine 4'-phosphate; consequently, the released CO_2 can be used by RuBisCO. The *L-glutamine* is an interesting hub that is intermediate in several pathways, like: amino acids metabolism, pyrimidine, purine and folate biosynthesis, cobalt, non-cobalt and coenzyme B12 pathway, glutathione and aminosugars metabolism. Conversely, it seems that only some reactions of these pathways, picture a scenario of global regulation. These reactions are those that involve the following enzymes: phosphoribosyldiphosphate 5-amidotransferase (reaction *"2.4.2.14"* in *i*Syf715), glutamine amidotransferase/cyclase (reaction "*4.1.3.-"* in *i*Syf715), glutaminyl-tRNA synthase (reaction "*tRNA _6.3.5.7"* in *i*Syf715), asparaginyl-tRNA synthase (reaction "*tRNA 6.3.5.6"* in *i*Syf715), aminodeoxychorismate synthase (reaction "*2.6.1.85"* in *i*Syf715), cobyrinate a,c-diamide synthase (reaction

"*6.3.5.11"* in *i*Syf715), adenosylcobyric acid synthase (reaction "*6.3.5.10"* in *i*Syf715) and gamma-glutamyltransferase (reaction "*_2.3.2.2k"* in *i*Syf715) picture a scenario of global regulation. *2-keto-3-deoxy-6-phospho-gluconate* is a reporter from pentose phosphate pathway. This indicates that the statistically significant differences in the expression of 2-dehydro-3-deoxy-phosphogluconate aldolase (reaction "*4.1.2.14"* in *i*Syf715), could be related to the successive pyruvate decarboxylations in some reactions or with the involvement of D-glyceraldehyde-3-phosphate in redox reactions which produces NADPH. On the other hand, it has recently been shown that this reaction has an increased carbon flux into the glycolytic pathway on HC/LC-shifted cyanobacteria cells (Huege *et al.*, 2011; Schwarz *et al.*, 2011). *dGDP* is an intermediate of purine metabolism. Interestingly, all the genes implicated in reactions in where *dGDP* is an intermediate, show differences in expression levels in both conditions. Other reporters are *4-amino-4-deoxychorismate* and *5-phospho-beta-D-ribosyl-amine*, which constitute intermediates in folate biosynthesis and purine metabolism, respectively. Here, chorismate and 5-phosphoribosyl 1-pyrophosphate are probably mostly used as a carbon skeletons for subsequent nitrogen assimilation via the glutamine synthetase/glutamate synthase cycle. The glutamate produced is subsequently used as an amino donator or building block for other biosyntheses such as amino acids, chlorophylls, antennas or heme/bilin. Finally, *UDP-N-acetylmuramoyl-L-alanyl-D-glutamyl-meso-2,6-diaminoheptanedioate* constitutes a metabolite belongings to peptidoglycan biosynthesis pathway. The low-growing *S. elongatus* PCC7942 show a decrease in the expression of the gene coding for UDP-N-acetylmuramoyl-L-alanyl-D-glutamate-2,6-diaminopimelate ligase, therefore the peptidoglycan synthesis could be affected. Thus, another hypothesis is that cell wall would be different under starvation of inorganic carbon than under high Ci regime.

To acquire a better overview of the hot spots within the metabolic network, we visualize each of 52 reporter metabolite in a whole metabolic map (see figure 6.1).

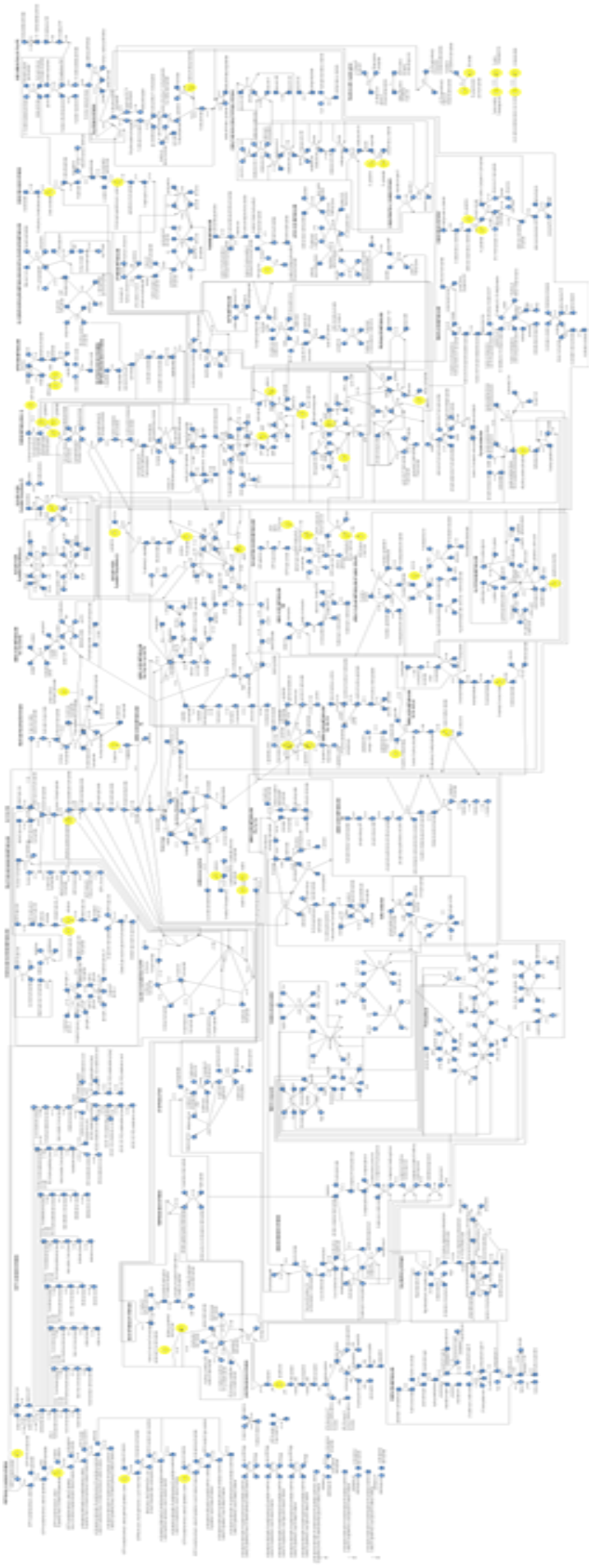


Figure 6.1. Reporter metabolites (yellow points) under HC/LC regime. Some of the metabolites are repeated in the graph.

By using the Reporter Subnetwork algorithm, we uncover sets of genes that significantly change their expression coordinately. We found a big set of 184 ORFs divided in two set of 88 and 96 ORFs. The first set consists of the genes from carbohydrate metabolism (89%) and amino acids metabolism (11%). The second set is representative of a variety of genes from different pathways such as nucleotide metabolism (42%), metabolism of cofactors (39%), nitrogen metabolism (18%) and carotenoids metabolism (1%).

It seems clear that more experimental data are necessary to clarify the regulatory principles of the inorganic carbon acclimation. Specially, the quantification of transients metabolites across the whole metabolic map.

6.3 Conclusions

Reporter Features can be applied to study a series of related perturbations where results can be used to construct an inferred regulatory map connecting physical or functional interactions between the perturbed elements (Oliveira *et al.*, 2008). In *i*Syf715, each perturbed element was linked to the Reporter Features calculated from the corresponding perturbation data, viz.: inorganic carbon acclimation. The resulting network is a representation of direct and/or indirect mechanisms of regulation that span the set of (related) perturbations used. As *reporter metabolites*, we have identified several metabolites that play a role as regulatory hubs when metabolism is being turned on from HC to LC for 6 and 24 h.

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.

The analyses presented in this chapter demonstrate the utility of this hypothesis-driven method when knowledge useful for metabolic design 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 alone is insufficient and, thus, algorithms such as Reporter Features can be successfully applied to a great extent to biological networks.

6.4 Methods

6.4.1 Transcriptome data analysis

Gene expression data from the comparison of two inorganic carbon acclimation conditions (Schwarz *et al.*, 2011) was integrated in the metabolic reconstruction. Reporter Features algorithm (Patil and Nielsen, 2005; Oliveira *et al.*, 2008), currently available online at Biomet Toolbox (Cvijovic *et al.*, 2010) (http://www.sysbio.se/BioMet), was used for this purpose. This algorithm works with three kinds of information:

(a) *p-values* for genes, resulting from, for example, Student's *t*-test or the Mann-Whitney *u*-test run on transcriptomic data,

(b) *interaction file,* where genes/reactions are connected to the corresponding features, in this case the corresponding substrates and products, and

(c) *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}).

$$z_{gene\ i} = \mathrm{cdf}^{-1}\left[1 - p_{gene\ i}\right]$$

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 neighbor *Nj* nodes (genes), *i.e.*:

$$z_{feature\ j} = \frac{1}{\mathrm{N_j}} \sum_{k=1}^{Nj} \mathrm{z_{gene\ k}}$$

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*:

$$z^{corrected}_{feature\ j} = \left(\frac{\mathrm{z_{feature\ j}} - m_N}{s_N} \right)$$

Chapter 7. Concluding remarks

Mathematical modeling has been practiced in various branches of science and engineering. Genome-scale models and their analyses can be useful for addressing several issues from the classical biology. Here, we draw a set of concluding remarks as well as the main workflow presented in this thesis.

Chapter 7. Concluding remarks

7.1 System biology is inherently mathematical

Biological systems can be characterized by the interactions between their components. Detailed catalogs of biological parts of cells have emerged and its chemical and causal interactions are being documented in those databases. The basic laws of chemistry governed the interactions of these molecular components. Multiple associations between biological components can be seen as a network, and can have different functional states. Several factors constrain these states such as: physicochemical, environmental, and biological. The number of possible functional states of networks typically grows much faster than the number of components in the network and also far exceeds the number of biologically useful states to an organism (Palsson, 2006). Metabolism is one of the best characterized cellular networks in terms of its biochemistry, kinetics and thermodynamics. These networks are based on all information of biochemical reactions, like stoichiometry, which can serve as base to form a matrix, named stoichiometric matrix. This network can be described at different levels of information enabling us to conceptualize their functionalities in a hierarchical fashion (Palsson, 2006).

Two-dimensional or systemic annotation of genomes is emerging and represents unity of effort in systems biology through metabolic network reconstruction. Network reconstruction is a detailed, laborious process that needs careful examination of all the components and links in the network. Procedures to perform this task have been developed. Numerous web resources and tools are available to aid in developing curated networks. Metabolic network interact with essentially all other cellular processes. The reconstruction of these processes and the integration of multiple networks will lead to the description of a comprehensive range of cellular functions. Such a multinetwork reconstruction represents a biochemically, genetically and genomically, structured database that provides the framework for analyzing *-omics* data types (Palsson, 2006).

Mathematically, the reconstruction process leads to the stoichiometric matrix which is becomes a key in systems biology. Structurally, or topologically, the stoichiometric matrix represents a reaction map, and mathematically, is a linear

mapping operation. Direct topological studies are interesting from a variety of standpoints. They focus on relatively easy to understand and intuitive properties of the structure of the network. Elementary topological properties relate to how connected a network is and how its components participate in forming the connectivity properties of the network. The number of reactions that compounds participate in network follows an approximate power law distribution in genome-scale matrices of metabolism (Palsson, 2006).

A wide range of constraint-based analysis methods have appeared and are being used to analyze various aspects of genome-scale models and the biological properties of the organisms that these models represents. Genome-scale reconstructions are mathematically represented and the governing constraints are imposed. This procedure leads to an *in silico* organism that contains all the known components of the real organism that it represents, and allows the simulation of allowable states given a set of governing constraints. Biological systems have to abide by a series of constraints, including those arising form basic natural laws, spatial constraints, and the environment in which they operate. Many possible biological functions are achievable under these constraints, and organisms willfully impose constraints through various regulatory mechanisms to select useful functional states from all allowable states. A constraint-based approach emerges from these considerations that enable the simultaneous analysis of physicochemical factors and biological properties.

Bounds and balance form linear constraints that comprise non-negativity constraints on the variables and by mass or flux balances. Specific points within these bounded solution spaces can be determined through optimization procedures, which are based on a stated objective function. These objectives can be used to probe network capabilities, to represent likely physiological objectives, and to represent candidate biological designs. If the objective function is linear, the linear programming can be used to find the optimal solution. In such cases, many different solutions (alternate optima) lead to the same optimal objective value. Thus, various analysis methods that look at optimal states and parameters variation have been programmed. As well as, several sophisticated *in silico* analysis methods, in which the objective function is non-linear, have been developed to carry out analyses of the consequence

of gene knockouts. Summarizing, these methods are principally based on the use of constraint-based optimization.

These considerations encompass the answers to many questions in systems biology. However, due to that systems and network analysis is at an early stage of development, there are some challenges that the field faces and constitute the horizon for future works.

7.2 Workflow

The scope of this thesis lied on two categories: biological content and *in silico* analyses methods applied to it.

The thesis work started from a model-driven systems biology workflow. We first gathered all kinds of biological information from *Synechococcus elongatus* PCC7942, including -*omics* elements, for example: genomics data, proteomics, transcriptomics, metabolomics, fluxomics, etc. Similarly, we deepened on the importance of the use of genome-scale metabolic models, with emphasis on its applicability to socioeconomic purposes. With retrieved information, we reconstructed the cyanobacterium metabolic model at genome-scale. Once reconstructed cellular metabolism, the accuracy and quality of the model was validated by comparison with experimental data for this cyanobacterium. This involved a series of computational simulations, by which the optimal states of the metabolic network were determined. Also, we calculated the variability of metabolic flux distributions without affecting the simulated objective function. We assessed the robustness of the network by studying the effects of varying CO_2 and HCO_3^- uptake rates on maximal growth.

The second phase was characterized by the analysis of the metabolic capabilities of this prokaryote. Using the reconstructed metabolic model we evaluated genetic perturbations which lead to the maximum in the synthesis of different bio-products. With this purpose, we studied economically important metabolites such as: ethanol, higher chain alcohols, lipids and hydrogen. Likewise, we proposed feasible metabolic designs by which the microorganism may be capable of overproduce these mentioned metabolites.

In the same way, the *in silico* metabolic model was used to predict cellular metabolic states. We characterized metabolic phenotypes and their respective

biomass, alcohols, lipids and hydrogen productions over a range of CO_2 and photons uptake rates.

Finally, the genome-scale metabolic model has served as a basis to integrate different biological information levels, like: transcriptome. In this way, we inferred new hot spots related to metabolic dynamic under inorganic carbon starvation, which can be useful in the design of metabolic engineering strategies.

Results of this PhD dissertation can be the foundation of future integrative approaches that will give continuity to the work performed in order to obtain an industrially-relevant production platform out of *Synechococcus elongatus* PCC7942.

Bibliography

Abed RMM, Dobretsov S, Sudesh K. (2009) **Applications of cyanobacteria in biotechnology**. *J Appl Microbiol*, **106**:1-12.

Adamic LA, Huberman BA. (2002) **Zipf's law and the Internet.** *Glottometrics*, **3**:143-150.

Albert R, Jeong H, Barabási AL. (2000) **Error and attack tolerance of complex networks**. *Nature*, **406**: 378-382.

Allakhverdiev SI, Sakamoto A, Nishiyama Y, Murata N. (2000) **Inactivation of photosystems I and II in response to osmotic stress in *Synechococcus*. Contribution of water channels**. *Plant Physiol*, **122**:1201-1208.

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

Andersson CR, Tsinoremas NF, Shelton J. *et al.* (2000) **Application of bioluminescence to the study of circadian rhythms in cyanobacteria**. *Methods Enzymol*, **305**:527-542.

Angermayr SA, Hellingwerf KJ, Lindblad P, Mattos MJT. (2009) **Energy biotechnology with cyanobacteria**. *Curr Opin Biotechnol*, **20**:257-263.

Asada Y, Koike Y, Schnackenberg J, et al. (2000) **Heterologous expression of clostridial hydrogenase in the cyanobacterium *Synechococcus* PCC7942**. *BBA-Gene Struct Expr*, **1490 (3)**:269-278.

Atsumi S, Hanai T, Liao JC. (2008) **Non-fermentative pathways for synthesis of branched-chain higher alcohols as biofuels**. *Nature* **451**:86-89.

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

Atsumi S, Li Z, Liao JC. (2009a) **Acetolactate synthase from *Bacillus subtilis* serves as a 2-ketoisovalerate decarboxylase for isobutanol biosynthesis in *Escherichia coli***. *Appl Environ Microb*, **75**:6306-6311.

Baart GJE, Martens DE. (2012) **Genome-scale metabolic models: reconstruction and analysis**. *Methods Mol Biol*, **799**:107-126.

Badger MR, Gallagher A. (1987) **Adaptation of photosynthetic CO_2 and HCO_3^- accumulation by the cyanobacterium *Synechococcus* PCC6301 to growth at different inorganic carbon concentrations**. *Aust J Plant Physiol*, **14**:189-210.

Badger MR, Price GD. (2003) **CO_2 concentrating mechanisms in cyanobacteria: molecular components, their diversity and evolution**. *J Exp Bot*, **54**:609-622.

Badger MR. (1987) **The CO_2 concentrating mechanism in aquatic phototrophs**. In: *The Biochemistry of Plants: a Comprehensive Treatise*, Vol. 10. Photosynthesis (Ed. by M.D. Hatch and N.K. Boardman), Academic Press, San Diego.

Baez A, Cho KM, Liao JC. (2011) **High-flux isobutanol production using engineered *Escherichia coli*: a bioreactor study with in situ product removal**. *Appl Microbiol Biot*, **90**:1681-1690.

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

Barabási AL, Albert R. (1999) **Emergence of scaling in random networks**. *Science*, **286**:509-512.

Barabási AL, Bonabeau E. (2003) **Scale-free networks**. *Sci Am*, **288**:50-59.

Barabási AL, Oltvai ZN. (2004) **Network biology understanding the cell's functional organization**. *Nat Rev Genet*, **5**:101-113.

Bard J. (2013) **Systems biology-the broader perspective**. *Cells*, **2**:414-431.

Bar-Even A, Noor E, Lewis NE, Milo R. (2010) **Design and analysis of synthetic carbon fixation pathways**. *PNAS,* **107**:8889-8894.

Becker SA, Feist AM, Mo ML, *et al.* (2007) **Quantitative prediction of cellular metabolism with constraint-based models: the COBRA Toolbox**. *Nat Protoc*, **2**:727-738.

Benson DA, Cavanaugh M, Clark K, *et al.* (2013) **GenBank**. *Nucleic Acids Res*, **41**:36-42.

Bertilsson S, Berglund O, Karl DM, Chisholm SW. (2003) **Elemental composition of marine *Prochlorococcus* and *Synechococcus*: Implications for the ecological stoichiometry of the sea**. *Limnol Oceanogr*, **48**:1721-1731.

Beyer WH. (1987) **CRC Standard Mathematical Tables**, 28th Ed. Boca Raton, FL: CRC Press.

Blazier AS, Papin JA. (2012) **Integration of expression data in genome-scale metabolic network reconstructions**. *Front Physiol*, **3**: 299.

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

Borden JR, Papoutsakis ET. (2007) **Dynamics of genomic-library enrichment and identification of solvent tolerance genes for *Clostridium acetobutylicum***. *Appl Environ Microb*, **73**:3061-3068.

Bruggeman FJ, Snoep JL, Westerhoff HV. (2008) **Control, responses and modularity of cellular regulatory networks: a control analysis perspective**. *IET Syst Biol*, **2**:397-410.

Bruggeman FJ, Westerhoff HV. (2006) **Approaches to biosimulation of cellular processes**. *J Biol Phys*, **32**:273-288.

Buchanan BB, Gruissem W, Jones RL (Eds.). (2000) **Biochemistry and molecular biology in plants**. *American Society of Plants Physiologists*, Rockville, MD.

Burdock GA (Ed.). (2005) **Fenaroli's handbook of flavor ingredients**. 5th Ed. Boca Raton, FL.

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

Burja AM, Dhamwichukorn S, Wright PC. (2003) **Cyanobacterial postgenomic research and systems biology**. *Trends Biotechnol*, **21**:504-511.

Burkhardt S, Riebesell U, Zondervan I. (1999) **Effects of growth rate, CO_2 concentration, and cell size on the stable carbon isotope fractions in marine phytoplankton**. *Geochim Cosmochom Acta*, **63**:3729-3741.

Burkhardt S, Riebesell U. (1997) **CO_2 availability affects elemental composition (C:N:P) of the marine diatom *Skeletonema costatum***. *Mar Ecol Prog Ser*, **155**:67-76.

Cann AF, Liao JC. (2008) **Production of 2-methyl-1-butanol inengineered *Escherichia coli***. *Appl Microbiol Biot*, **81**:89-98.

Carlini C, Macinai A, Marchionna M, Noviello M, Galletti AMR, Sbrana G (2003) **Selective synthesis of isobutanol by means of the Guerbet reaction: part 3:**

methanol/n-propanol condensation by using bifunctional catalytic systems based on nickel, rhodium and ruthenium species with basic components. *J Mol Catal A Chem*, **206**:409-418.

Carlson J, Doyle J. (2002) **Complexity and robustness**. *PNAS*, **99**:2538-2545.

Carr P, Church GM. (2009) **Genome engineering**. *Nat Biotechnol*, **27**:1151-1162.

Clerico EM, Ditty JL, Golden SS. (2007) **Specialized techniques for site-directed mutagenesis in cyanobacteria**. *Methods Mol Biol*, **362**:155-171.

Connor MR, Liao JC. (2008) **Engineering of an *Escherichia coli* strain for the production of 3-methyl-1-butanol**. *Appl Environ Microb*, **74**:5769-5775.

Costa JA, de Morais MG. (2011) **The role of biochemical engineering in the production of biofuels from microalgae**. *Bioresour Technol*, **102(1)**:2-9.

Covert MW, Famili I, Palsson BØ (2003). **Identifying constraints that govern cell behavior: a key to converting conceptual to computational models in biology?** *Biotechnol Bioeng*, **84**:763-772.

Covert MW, Knight EM, Reed JL, *et al.* (2004) **Integrating high-throughput and computational data elucidates bacterial networks**. *Nature*, **429**:92-96.

Crabtree B, Newsholme EA. (1987) **A systematic approach to describing and analyzing metabolic control systems**. *Trends Biochem Sci,* **12**:4-12.

Crick FH. (1958) **The biological replication of macromolecules**. *Symp Soc Exp Biol*, **12**:138-163.

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

Csete M, Doyle J. (2004) **Bow ties, metabolism and disease**. *Trends Biotechnol*, **22**:446-450.

Cvijovic M, Olivares-Hernández R, Agren R, *et al.* (2010a) **BioMet Toolbox -suppl**. *Nucleic Acids Res*, **38 Suppl**:W144-149.

Chen Y. (2007) **Functional genomics of the unicellular cyanobacterium *Synechococcus elongatus* PCC7942**. A Philosophiae Doctor thesis. (Available at: http://repository.tamu.edu/bitstream/handle/1969.1/ETD-TAMU-1552/CHEN-DISSERTATION.pdf?sequence=1).

Chen YC, Holtman KC, Magnuson RD, *et al.* (2008) **The complete sequence and functional analysis of pANL, the large plasmid of the unicellular freshwater cyanobacterium *Synechococcus elongatus* PCC7942.** *Plasmid*, **59**:176-192.

Cheng CL, Che PY, Chen BY, *et al.* (2012) **High yield bio-butanol production by solvent-producing bacterial microflora**. *Bioresource Technol*, **113**:58-64.

Church GM. (2013) **Reading and writing omes**. *Mol Syst Biol,* **9**:642-645.

Da Rós PCM, Silva CSP, Silva-Stenico ME, Fiore MF, De Castro HF. (2013) **Assessment of Chemical and Physico-Chemical Properties of Cyanobacterial Lipids for Biodiesel Production**. *Mar Drugs*, **11(7)**:2365-2381.

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

Das D. (2010) **Biohydrogen Production Technology**: the present scenario. *Akshay Urja*, **3**:18-21.

De Mey M, Taymaz-Nikerel H, Baart G *et al.* (2010) **Catching prompt metabolite dynamics in *Escherichia coli* with the BioScope at oxygen rich conditions**. *Metab Eng*, **12**:477-487.

Demmig-Adams B, Adams III W.W. (1992) **Photoprotection and other responses of plants to high light stress**. *Annu Rev Plant Physiol Plant MoI Biol*, **43**:599-626.

Deng MD, Coleman J. (1999) **Ethanol Synthesis by Genetic Engineering in Cyanobacteria.** *Appl Environ Microb*, **65**:523-528.

Dickinson JR, Harrison SJ, Dickinson JA, Hewlins MJ. (2000) **An investigation of the metabolism of isoleucine to active Amyl alcohol in *Saccharomyces cerevisiae***. *J Biol Chem*, **275**:10937-10942.

Dickinson JR, Harrison SJ, Hewlins MJ. (1998) **An investigation of the metabolism of valine to isobutyl alcohol in *Saccharomyces cerevisiae***. *J Biol Chem*, **273**:25751-25756.

Dickinson JR, Lanterman MM, Danner DJ, *et al.* (1997) **A ^{13}C nuclear magnetic resonance investigation of the metabolism of leucine to isoamyl alcohol in *Saccharomyces cerevisiae***. *J Biol Chem*, **272**:26871-26878.

Dickinson, J. R. *et al.* (2003)**The catabolism of amino acids to long chain and complex alcohols in Saccharomyces cerevisiae**. *J Biol Chem*, **278**:8028-8034.

Duan YX, Chen T, Chen X, Zhao XM. (2010) **Overexpression of glucose-6-phosphate dehydrogenase enhances riboflavin production in *Bacillus subtilis***. *Appl Microbiol Biot*, **85**:1907-1914.

Duarte NC, Becker SA, Jamshidi N, *et al.* (2007) **Global reconstruction of the human metabolic network based on genomic and bibliomic data**. *PNAS*, **104**:1777-1782.

Ducat DC, Way JC, Silver PA. (2011) **Engineering cyanobacteria to generate high-value products**. *Trends Biotechnol*, **29(2)**:95-103.

Durot M, Bourguignon PY, Schachter V. (2009) **Genome-scale models of bacterial metabolism: reconstruction and applications**. *FEMS Microbiol Rev,* **33**:164-190.

Dutta D, De D, Chaudhuri S, Bhattacharya SK. (2005) **Hydrogen production by Cyanobacteria**. *Microb Cell Fact*, **4**:1-11.

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

Edwards JS, Palsson BØ. (2000) **Robustness analysis of the *Escherichia coli* metabolic network**. *Biotechnol Prog,* **16**:927-939.

Edwards JS, Palsson BØ. (2000) **Metabolic flux balance analysis and the *in silico* analysis of *Escherichia coli* K-12 gene deletions**. *BMC Bioinformatics*, **1**:1-10.

Edwards JS, Ramakrishna R, Palsson BØ. (2002) **Characterizing the Metabolic Phenotype: a Phenotype Phase Plane Analysis**. *Biotechnol Bioeng*, **77**:27-36.

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.

Elder RL (ed). (1987) **Final report on the safety assessment of oleic acid, lauric acid, palmitic acid, myristic acid, and stearic acid**. *J Am Coll Toxicol*, **6(3)**: 321-401.

Elowitz MB, Surette MG, Wolf PE, *et al.* (1999) **Protein mobility in the cytoplasm of *Escherichia coli***. *J Bacteriol,* **181**:197-203.

Espie GS, Miller AG, Birch DG, Canvin DT. (1988) **Simultaneous transport of CO_2 and HCO_3^- by the cyanobacterium *Synechococcus* UTEX 6251**. *Plant Physiol*, **87**:551-554.

Esvelt KM, Wang HH. (2013) **Genome-scale engineering for systems and synthetic biology**. *Mol Syst Biol*, **9(641)**:1-17.

Famili I, Forster J, Nielsen J, *et al.* (2003) ***Saccharomyces cerevisiae*** **phenotypes can be predicted using constraint-based analysis of a genome-scale reconstructed metabolic network**. *PNAS,* **100**:13134-13139.

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**. *Mol Syst Biol*, **3**:121.

Feist AM, Herrgard MJ, Thiele I, *et al.* (2009) **Reconstruction of biochemical networks in microorganisms**. *Nat Rev Microbiology,* **7**:129-143.

Fell D. (1997) **Understanding the control of metabolism**, Portland Press, London.

Finkelstein J, Martin J. (2000) **Homocysteine**. *Int J Biochem Cell Biol,* **32**:385-389.

Förster J, Famili I, Fu P., *et al.* (2003) **Genome-scale reconstruction of the *Saccharomyces cerevisiae* metabolic network**. *Genome Res,* **13**:244-253.

Francisco EC, Neves DB, Jacob-Lopes E, Franco TT. (2010) **Microalgae as feedstock for biodiesel production: carbon dioxide sequestration, lipid production and biofuel quality**. *J Chem Technol Biotechnol*, **85**:395-403.

Fu FX, Warner ME, Zhang Y, *et al.* (2007) **Effects of increased temperature and CO_2 on photosynthesis, growth, and elemental ratios in marine *Synechococcus* and *Procholorococcus* (Cyanobacteria)**. *J Phycol*, **43**:485-496.

Fu P. (2008) **Genome-scale modeling of *Synechocystis* sp. PCC6803 and prediction of pathway insertion**. *J Chem Technol Biot*, **84**:473-483.

Gamermann D, Montagud A, Conejero JA, *et al*. (2014) **New approach for phylogenetic tree recovery based on genome-scale metabolic networks**. *J Comput Biol,* **21**:1-12.

Gamermann D, Montagud A, Jaime Infante RA, Triana J, Fernández de Córdoba P, Urchieguía JF. (2014) **PyNetMet: Python tools for efficient work with networks and metabolic models**. *Computational and Mathematical Biology*, **3(5)**:1-11.

Glazyrina J**,** Materne EM**,** Dreher T**,** *et al*. (2010) **High cell density cultivation and recombinant protein production with *Escherichia coli* in a rocking-motion-type bioreactor**. *Microb Cell Fact*, **9**:42.

Goldemberg J. (2007) **Ethanol for a sustainable energy future**. *Science*, **315**:808-810.

González-Barreiro O, Rioboo C, Cid A, Herrero C. (2004) **Atrazine-induced chlorosis in *Synechococcus elongatus* cells**. *Arch Environ Contam Toxicol*, **46(3)**:301-307.

Grafahrend-Belau E, Klukas C, Junker BH, Schreiber F. (2009) **FBA-SimVis: interactive visualization of constraint-based metabolic models**. *Bioinformatics,* **25(20)**:2755-2757.

Hamilton JJ, Reed JL. (2012) **Identification of Functional Differences in Metabolic Networks Using Comparative Genomics and Constraint-Based Models**. *PLoS ONE,* **7(4)**: e34670.

Hallenbeck PC, Benemann JR. (2002) **Biological hydrogen production: fundamentals and limiting processes**. *Int J Hydrogen Energ*, **27**:1185-1193.

Hardy M. (2010) **Pareto's Law**. *Math Intell,* **32(3)**:38-43.

Hellingwerf KJ, Mattos MJT. (2009) **Alternative routes to biofuels: light-driven biofuel formation from C_{O2} and water based on the 'photanol' approach**. *J Biotechnol*, **142**:87-90.

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

Hilbert D, Cohn-Vossen S. (1999) **Geometry and the Imagination**. New York: Chelsea Publishing Co.

Hill J, Nelson E, Tilman D, *et al.* (2006) **Environmental, economic, and energetic costs and benefits of biodiesel and ethanol biofuels**. *PNAS*, **103**:11206-11210.

Holtman CK, Chen Y, Sandoval P, *et al.* (2005) **High-throughput functional analysisof the *Synechococcus elongatus* PCC7942 genome**. *DNA Res,* **12**:103-115.

Hood L, Heath JR, Phelps ME, Lin B. (2004) **Systems biology and new technologies enable predictive and preventative medicine**. *Science*, **306**:640-643.

Huber GW, Iborra S, Corma A. (2006) **Synthesis of transportation fuels from biomass: chemistry, catalysts, and engineering**. *Chem Rev*, **106**:4044-4098.

Huege J, Goetze J, Scharz D, *et al.* (2011) **Modulation of the major paths of carbon in photorespiratory mutants of *Synechocystis***. *PLoS ONE,* **6**:e16278.

Ibarra RU, Edwards JS, Palsson BØ. (2002) ***Escherichia coli* K-12 undergoes adaptive evolution to achieve *in silico* predicted optimal growth**. *Nature*, **420**:186-189.

Imam S, Yilmaz S, Sohmen U, *et al.* (2011) ***i*Rsp1095: A genome-scale reconstruction of the *Rhodobacter sphaeroides* metabolic network**. *BMC Syst Biol*, **5**:116.

Jeong H, Mason SP, Barabási AL, Oltvai ZN. (2001) **Lethality and centrality in protein networks**. *Nature*, **411**:41-42.

Jeong H, Tombor B, Albert R, *et al*. (2000) **The large-scale organization of metabolic networks**. *Nature*, **407**:651-654.

Kacser H, Burns JA. (1995) (with additional comments by Kacser H. and Fell D. A.) **The control of flux**. *Biochem. Soc. Trans,* **23**:341-366.

Kajiwara S, Yamada H, Ohkuni N, *et al.* (1997) **Design of the bioreactor for carbon dioxide fixation by *Synechococcus* PCC7492**. *Energ Convers Manage,* **38**:529-532.

Kanehisa M, Araki M, Goto S, *et al.* (2008) **KEGG for linking genomes to life and the environment**. *Nucleic Acids Res*, **36**:480-484.

Karatay SE, Donmez G. (2011) **Microbial oil production from thermophile cyanobacteria for biodiesel production**. *Appl Energ*, **88**:3632-3635.

Karp PD, Paley S, Romero P. (2002) **The Pathway Tools software.** *Bioinformatics (Oxford, England),* **18(Suppl 1)**:225-232.

Karr JR, Sanghvi JC, Macklin DN, *et al.* (2012) **A whole-cell computational model predicts phenotype from genotype**. *Cell* **150**:389-401.

Kauffman KJ, Prakash P, Edwards JS. (2003). **Advances in flux balance analysis**. *Curr Opin Biotechnol,* **14**:491-496.

Kaygusuz K. (2009) **Bioenergy as a Clean and Sustainable Fuel. Energy sources, part A: recovery, utilization, and environmental effects**. *Energ Sourc*, **31**:1069-1080.

Khalil MH, Marcelletti JF, Katz LR, Katz DH, Pope LE. (2000) **Topical application of docosanol- or stearic acid-containing creams reduces severity of phenol burn wounds in mice**. *Contact Dermatitis*, 43(2):79-81.

Kim HU, Kim SY, Jeong H, *et al.* (2011) **Integrative genome-scale metabolic analysis of Vibrio vulnificus for drug targeting and discovery**. *Mol Syst Biol*, **7**:460.

Kim HU, Kim TY, Lee SY (2010) **Genome-scale metabolic network analysis and drug targeting of multi-drug resistant pathogen Acinetobacter baumannii AYE**. *Mol Biosyst*, **6**:339-348.

Kim JM, Lee K, Shin K, *et al.* (2006) **The effect of seawater CO_2 concentration on growth of a natural phytoplankton assemblage in a controlled mesocosm experiment.** *Limnol Oceanogr*, **51**:1629-1636.

Kim TY, Sohn SB, Kim YB, *et al.* (2012) **Recent advances in reconstruction and applications of genome-scale metabolic models**. *Curr Opin Biotechnol*, **23**:617-623.

Kiss AA, Bildea CS. (2012) **A review of biodiesel production by integrated reactive separation technologies**. *J Chem Technol Biot*, 87:861-879.

Kitano H. (2002) **System biology: A brief overview**. *Science*, **295**:1662-1664.

Kitano H. (2004) **Biological robustness**. *Nat Rev Genet*, **5**:826-836.

Klamt S, Saez-Rodriguez J, Gilles ED. (2007) **Structural and functional analysis of cellular networks with CellNetAnalyzer**. *BMC Syst Biol,* **1**:1-13.

Klanchui A, Chiraphan K, Atchara P, *et al.* (2012) ***i*AK692: A genome-scale metabolic model of *Spirulina platensis C1***. *BMC Syst Biol*, **6**:71.

Knoop H, Zilliges Y, Lockau W, Steuer R. (2010) **The metabolic network of *Synechocystis* sp. PCC 6803: Systemic properties of autotrophic growth**. *Plant Physiol*, **154**:410-422.

Kondatrieva EN, Gogotov IN. (1983) **Production of molecular hydrogen in microorganisms.** *Adv in Biochem Eng Biot*, **28**:139-161.

Kondo T, Strayer CA, Kulkarni RD. *et al.* (1993) **Circadian rhythms in prokaryotes: luciferase as a reporter of circadian gene expression in cyanobacteria**. *PNAS*, **90**:5672-5676.

Konig S. (1998) **Subunit structure, function and organisation of pyruvate decarboxylases from various organisms**. *Biochim. Biophys. Acta*, **1385**:271-286.

Koyama Y. (1991) **Structures and functions of carotenoids in photosynthetic systems**. *J. Photochem. Photobiol. B*, **9**:265-280.

Kuan, D. (2013) **Growth optimization of Synechococcus elongatus PCC7942 in lab flask and 2D photobioreactor**. A Master thesis. (Available at:

https://circle.ubc.ca/bitstream/handle/2429/45010/ubc_2013_fall_kuan_david.pdf?sequence=1).

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

Latendresse M, Krummenacker M, Trupp M, Karp PD. (2012) **Construction and completion of flux balance models from pathway databases**. *Bioinformatics*, **28**:388-396.

Leganés F, Forchhammer K, Fernández-Piñas F. (2009) **Role of calcium in acclimation of the cyanobacterium *Synechococcus elongatus* PCC 7942 to nitrogen starvation**. *Microbiology,* **155(1)**:25-34.

Levin DB, Pitt L, Love M. (2004) **Biohydrogen production: Prospects and limitations to practical application**. *Int J Hydrogen Energ*, **29(2)**:173-185.

Lew VL, Bookchin RM. (1986) **Volume, pH, and ion-content regulation in human red cells: analysis of transient behavior with an integrated model**. *J Membr Biol*, **92**:57-74.

Lewis NE, Hixson KK, Conrad TM, *et al*. (2010a) **Omic data from evolved *E. coli* are consistent with computed optimal growth from genome-scale models**. *Mol Syst Biol*, **6**:390.

Lewis NE, Schramm G, Bordbar A, *et al.* (2010b) **Large-scale *in silico* modeling of metabolic interactions between cell types in the human brain**. *Nat Biotechnol*, **28**:1279-1285.

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

Liao JC, Hou SY, Chao YP. (1996) **Pathway analysis, engineering and physiological considerations for redirecting central metabolism**. *Biotechnol Bioeng,* **52**:129-140.

Lin F, Chen Y, Levine R, *et al.* (2013) **Improving fatty acid availability for bio-hydrocarbon production in *Escherichia coli* by metabolic engineering**. *PLoS ONE* **8(10)**:e78595.

Lopes-Pinto FA, Troshina O, Lindblad P. (2002) **A brief look at three decades of research on cyanobacterial hydrogen evolution**. *Int J Hydrogen Energ*, **27:**1209-1215.

Mahadevan K, Palsson BØ. (2005) **Properties of metabolic networks: Structure versus function**. *Biophys J*, **88**:L7-L9.

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

Mariano AP, Borba Costa CB, De Angelis DF, *et al.* (2009) **Optimization strategies based on sequential quadratic programming applied for a fermentation process for butanol production**. *Appl Microbiol Biot*, **159**:366-381.

Martin-Creuzburg D, Sperfeld E, Wacker A. (2009) **Colimitation of a freshwater herbivore by sterols and polyunsaturated fatty acids**. *Proc R Soc B rspb*, **276(1663)**:1805-1814.

McNeely K, Xu Y, Bennette N, Bryant DA, Dismukes GC. (2010) **Redirecting reductant flux into hydrogen production via metabolic engineering of fermentative carbon metabolism in a cyanobacterium**. *Appl Environ Microb*, **76**:5032-5038.

Miao X, Wu Q, Wu G, Zhao N. (2003) **Changes in photosynthesis and pigmentation in an agp deletion mutant of the cyanobacterium *Synechocystis* sp**. *Biotechnol Lett*, **25**:391-396.

Milne CB, Eddy JA, Raju R. (2011) **Metabolic network reconstruction and genome-scale model of butanol-producing strain *Clostridium beijerincki i*NCIMB 8052**. *BMC Syst Biol,* **5**:130-145.

Montagud A, Fernández de Córdoba P, Urchueguía JF. (2013) ***Synechocystis* sp. PCC 6803 metabolic models study for the enhanced production of biofuels.** *Crit Rev Biotechnol*, **8551**:1-15.

Montagud A, Navarro E, Fernández de Córdoba P, *et al.* (2010) **Reconstruction and analysis of genome-scale metabolic model of a photosynthetic bacterium**. *BMC Syst Biol,* **4**:156-172.

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**. *Biotechnol J,* **6**:330-342.

Molenaar A, Mao J, Oden K, Seyfert HM. (2003) **All three promoters of the acetyl-coenzyme A-carboxylase ALPA-encoding gene are expressed in mammary epithelial cells of ruminants.** *J Histochem Cytochem*, **51**:1073-1081.

Munekaga Y, Hashimoto M, Miyaka C, *et al.* (2004) **Cyclic electron flow around photosystem I is essential for photosynthesis**. *Nature*, **429**:579-582.

Nakamura YJ, Takahashi A, Sakurai Y, *et al.* (2005). **Some cyanobacteria synthesize semi-amylopectin type α-polyglucans instead of glycogen**. *Plant Cell Physiol*, **46**:539-545.

Nandi R, Sengupta S. (1998) **Microbial production of hydrogen: an overview**. *Crit Rev in Microbiol*, **24**:61-84.

NCBI. (2013) **Entrez Genome for *Synechococcus elongatus* PCC7942** [//www.ncbi.nlm.nih.gov/genome/430].

Newman MEJ. (2005) **Power laws, Pareto distributions and Zipf's law**. *Contemp Phys*, **46(5)**: 323-351.

Nobles DR, Brown RM. (2008) **Transgenic expression of *Gluconacetobacter xylinus* strain ATCC 53582 cellulose synthase genes in the cyanobacterium *Synechococcus leopoliensis* strain UTCC 100**. *Cellulose*, **15**:691-701.

Nodop A, Pietsch D, Höcker R, *et al.* (2008) **Transcript profiling reveals new insights into the acclimation of themesophilic fresh-water cyanobacterium *Synechococcus elongatus* PCC 7942 to iron starvation**. *Plant Physiol*, **147**:747-763.

Notebaart RA, Van Enckevort FH, Francke C, *et al.* (2006) **Accelerating the reconstruction of genome-scale metabolic networks**. *BMC Bioinformatics*, **7**:296.

O'Brien EJ, Lerman JA, Chang RL, *et al.* (2013) **Genome-scale models of metabolism and gene expression extend and refine growth phenotype prediction**. *Mol Syst Biol*, **9**:693.

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

Orth JD, Thiele I, Palsson BØ. (2010) **¿What is flux balance analysis?** *Nature Biotechnology*, **28**:245-248.

Orth JD, Thiele I, Palsson BØ. (2010) **What is flux balance analysis?** *Nat Biotechnol,* **28**:245-248.

Palsson BØ. (2000) **The challenges of *in silico* biology**. *Nat Biotechnol*, **18**:1147-1150.

Palsson BØ. (2006) **Systems Biology - Properties of Reconstructed Networks**, Cambridge University Press: New York.

Papin JA, Price ND, Edwards JS, Palsson BØ. (2002) **The genome-scale metabolic extreme pathway structure in *Haemophilus influenzae* shows significant network redundancy**. *J Theor Biol,* **215**:67-82.

Papin JA, Price ND, Wiback SJ. (2003) **Metabolic pathways in the post-genome era**. *Trends Biochem Sci,* **28**:250-258.

Park JM, Kim TY, Lee SY. (2011) **Genome-scale reconstruction and *in silico*analysis of the *Ralstonia eutropha* H16 for polyhydroxyalkanoate synthesis, lithoautotrophic growth, and 2-methyl citric acid production**.*BMC Syst Biol,* **5**:101-112.

Patil KR, Nielsen J. (2005) **Uncovering transcriptional regulation of metabolism by using metabolic network topology.** *PNAS*, **102**:2685-2689.

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***. *J Gen Microbiology,* **49**:301-313.

Pearl GG. (2002) **Animal Fat Potential for Bioenergy use**. *Bioenergy 2002, The Biennial Bioenergy Conference, Boise, ID*.

Pharkya P, Burgard AP, Maranas CD. (2003) **Exploring the overproduction of amino acids using the bilevel optimization framework optKnock**. *Biotechnol Bioeng,* **84**:887-899.

Pharkya P, Burgard AP, Maranas CD. (2004) **OptStrain: A computational framework for redesign of microbial production systems**. *Genome Res*, **14**:2367-2376.

Pisciotta JM, Zou Y, Baskakov IV. (2010) **Light-Dependent Electrogenic Activity of Cyanobacteria**. *PLoS ONE* **5(5)**:e10821.

Pitera DJ, Paddon CJ, Newman JD, Keasling JD. (2007) **Balancing a heterologous mevalonate pathway for improved isoprenoid production in *Escherichia coli***. *Metab Eng*, **9**:193-207.

Pompella A, Visvikis A, Paolicchi A, *et al.* (2003) **The changing faces of glutathione, a cellular protagonist**. *Biochem Pharmacol,* **66**:1499-1503.

Price GD, Badger MR. (1989) **Ethoxyzolamide Inhibition of CO_2 Uptake in the Cyanobacterium *Synechococcus* PCC7942 without Apparent Inhibition of Internal Carbonic Anhydrase Activity**. *Plant Physiol*, **89**:37-43.

Price GD, Woodger FJ, Badger MR, *et al.* (2004a) **Identification of a SulP-type bicarbonate transporter in marine cyanobacteria**. *PNAS*, **101**:8228-18233.

Price ND, Papin JA, Palsson BØ. (2002) **Determination of redundancy and systems properties of the metabolic network of *Helicobacter pylori* using genome-scale extreme pathway analysis**. *Genome Res,* **12**:760-769.

Price ND, Papin JA, Schilling CH, *et al.* (2003) **Genome-scale microbial *in silico* models: the constraints-based approach**. *Trends Biotechnol,* **21**:162-169.

Price ND, Reed JL, Palsson BØ. (2004b) **Genome-scale models of microbial cells: evaluating the consequences of constraints**. *Nat Rev Microbiology,* **2**:886-897.

Quintana N, van der Kooy F, van de Rhee MD, *et al.* (2011) **Renewable energy from Cyanobacteria: energy production optimization by metabolic pathway engineering**. *Appl Microbiol Biot*, **91**:471-490.

Ramakrishna R, Ednards JS, McCulloch A, Palsson BØ. (2001) **Flux balance analysis of mitochondrial energy metabolism: Consequences of systemic stoichiometric constraints**. *Am J Physiol Regul Integr Comp Physiol,* **280**:R695-704.

Ranganathan S, Wei Tee T, Chowdhury A, *et al.* (2012) **An integrated computational and experimental study for overproducing fatty acids in *Escherichia coli***. *Metab Eng* **14**:687-704.

Raven JA. (2003) **Inorganic carbon concentrating mechanisms in relation to the biology of algae**. *Photosynth Res*, **77**:155-171.

Reyes R, Gamermann D, Montagud A, Fuente D, Triana J, Urchueguía JF y Fernández de Córdoba P. (2012) **Automation on the generation of genome scale metabolic models**. *J Comput Biol*, **19**:1295-1306.

Rittmann BE. (2008) **Opportunities for renewable bioenergy using microorganisms**. *Biotechnol Bioeng*, **100(2)**:203-212.

Rittmann BE. (2008) **Review Opportunities for renewable bioenergy using microorganisms**. *Biotechnol Bioeng*, **100**:203-212.

Robertson BR, Tezuka N, Watanabe MM. (2001) **Phylogenetic analyses of *Synechococcus* strains (cyanobacteria) using sequences of 16S rDNA and part of the phycocyanin operon reveal multiple evolutionary lines and reflect phycobilin content**. *Int J Syst Evol Microbiol,* **51**:861-871.

Rocha I, Maia P, Evangelista P, *et al.* (2010) **OptFlux: an open-source software platform for *in silico* metabolic engineering**. *BMC Syst Biol*, **4**:45.

Rosales-Loaiza N, Guevara M, Lodeiros C, Morales E. (2008) **Crecimiento y producción de metabolitos de la cianobacteria marina *Synechococcus* sp. (Chroococcales) en función de la irradiancia**. *Rev Biol Trop*, **56**:421-429.

Rosales-Loaiza N, Ortega J, Mora R, *et al.* (2005) **Influencia de la salinidad sobre crecimiento y composición bioquímica de la cianobacteria *Synechococcus* sp**. *Cienc Mar,* **31**:349-355.

Savinell JM, Palsson BØ. (1992a) **Optimal selection of metabolic fluxes for *in vivo* measurement. I. Development of mathematical methods**. *J Theor Biol*, **155**:201-214.

Savinell JM, Palsson BØ. (1992b) **Optimal selection of metabolic fluxes for in vivo measurement. II. Application to *Escherichia coli* and hybridoma cell metabolism**. *J Theor Biol*, **155**:215-242.

Scanlan DJ, Nyree JW. (2002) **Molecular ecology of the marine cyanobacterial genera *Prochlorococcus* and *Synechococcus***. *FEMS Microbiol Ecol*, **40**:1-12.

Schilling CH, Edwards JS, Letscher D, Palsson BØ. (2000a) **Combining pathway analysis with flux balance analysis for the comprehensive study of metabolic systems**. *Biotechnol Bioeng*, **71**:286-306.

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**. *J Theor Biol,* **203**:229-248.

Schilling CH, Schuster S, Palsson BØ *et al.* (1999) **Metabolic pathway analysis: basic concepts and scientific applications in the postgenomic era**. *Biotechnol Prog*, **15**:296-303.

Schuster S, Dandekar T, Fell DA. (1999) **Detection of elementary flux modes in biochemical networks: a promising tool for pathway analysis and metabolic engineering**. *Trends Biotechnol*, **17**:53-60.

Schwarz D, Nodop A, Hüge J, *et al.* (2011) **Metabolic and transcriptomic phenotyping of inorganic carbon acclimation in the cyanobacterium *Synechococcus elongatus* PCC 7942**. *Plant Physiol*, **155**:1640-1655.

Segrè D, Vitkup D, Church GM. (2002) **Analysis of optimality in natural and perturbed metabolic networks**. *PNAS*, **99**:15112-15117.

Sentheshanmuganathan S, Elsden SR. (1958) **The mechanism of the formation of tyrosol by *Saccharomyces cerevisiae***. *Biochem J*, **69**:210-218.

Sentheshanuganathan S. (1960) **The mechanism of the formation of higher alcohols from amino acids by *Saccharomyces cerevisiae***. *Biochem J*, **74**:568-576.

Shastri A, Morgan J. (2005) **Flux balance analysis of photoautotrophic metabolism**. *Biotechnol Progr*, **21**:1617-26.

Shen CR, Liao JC. (2008) **Metabolic engineering of *Escherichia coli* for 1-butanol and 1-propanol production via the keto-acid pathways**. *Metab Eng* **10**:312-320.

Shestakov SV, Khyen NT. (1970) **Evidence for genetic transformation in blue-green alga *Anacystis nidulans* R2**. *Molec Gen Genet,* **107**:372-375.

Shibata M, Ohkawa H, Katoh H, Shimoyama M, Ogawa T. (2002) **Two CO_2-uptake systems: four systems for inorganic carbon acquisition in *Synechocystis* sp. strain PCC6803**. *Funct Plant Biol*, **29**:123-129.

Shih PM, Dongying W, Amel L, *et al.* (2013) **Improving the coverage of thecyanobacterial phylum using diversity-driven genome sequencing**. *PNAS*, **110(3)**:1053-1058.

Shlomi T, Berkman O, Ruppin E. (2005) **Regulatory on/off minimization of metabolic flux changes after genetic perturbations**. *PNAS*, **102**:7695-7700.

Siefermann-Harms D. (1987) **The light-hawesting and protective functions of carotenoids in photosynthetic membranes**. *Plant Physiol*, **69**:561-568.

Smith KM, Cho KM, Liao JC. (2010) **Engineering *Corynebacterium glutamicum* for isobutanol production**. *Appl Microbiol Biot,* **87**:1045-1055.

Snoep JL, Bruggeman F, Olivier BG. (2006) **Towards building the silicon cell: A modular approach**. *BioSystems,* **83**:207-216.

Stelling J, Sauer U, Szallasi Z, *et al.* (2004) **Robustness of Cellular Functions**. *Cell*, **118**:675-685.

Stephanopoulos G, Aristidou AA, Nielsen JH. (1998) **Metabolic engineering: principles and methodologies**. Academic Press. San Diego.

Stouthamer AH. (1979) **The search for correlation between theoretical and experimental growth yields**. *Microb. Biochem,* **21**:1-48.

Strogatz S. (2001) **Exploring complex networks**. *Nature*, **410**:268-276.

Sugita C, Ogata K, Shikata M, *et al.* (2007) **Complete nucleotide sequence of the freshwater unicellular cyanobacterium *Synechococcus elongatus* PCC6301 chromosome: gene content and organization**. *Photosyn. Res*, **93**:55-67.

Szallasi Z, Periwal V, Stelling J. (2006) **Modeling in Cellular System Biology: From Concepts to Nuts and Bolts**. The MIT Press. Cambridge, Massachusetts.

Taher H, Al-Zuhair S, Al-Marzouqi AH, Haik Y, Farid MM. (2011) **A review of enzymatic transesterification of microalgal oil-based biodiesel using supercritical technology**. *Enzyme Res*, 468292.

Tanaka R. (2005) **Scale-rich metabolic networks**. *Phys Rev Lett*, **94**:168101-168104.

Taymaz-Nikerel H, Borujeni AE, Verheijen PJ *et al.* (2010) **Genome-derived minimal metabolic models for *Escherichia coli* MG1655 with estimated in vivo respiratory ATP stoichiometry**. *Biotechnol Bioeng*, **107**:369-381.

Thiele I, Palsson BØ. (2010) **A protocol for generating a high-quality genome-scale metabolic reconstruction**. *Nat Protoc,* **5**:93-121.

Thykaer J, Nielsen J, Wohlleben W, *et al.* (2010) **Increased glycopeptide production after overexpression of shikimate pathway genes being part of the balhimycin biosynthetic gene cluster**. *Metab Eng*, **12**:455-461.

Tomita M. (2001) **Whole-cell simulation: a grand challenge of the 21st century**. *Trends Biotechnol*, **19**:205-210.

Triana J, Montagud A, Siurana M, Gamermann D, Torres J, Tena J, Fernández de Córdoba P, Urchueguía JF. (2014) **Generation and evaluation of a genome-scale metabolic network model of *Synechococcus elongatus* PCC7942**. *Submitted at Metabolites*.

van den Hondel CA, Verbeek S, van der Ende A, *et al.* (1980) **Introduction of transposon Tn901 into a plasmid of *Anacystis nidulans*: preparation for cloning in cyanobacteria**. *PNAS*, **77**:1570-1574.

Van der Plas J. (1992) **Identification of replication and stability functions in the complete nucleotide sequence of plasmid pUH24 from the cyanobacterium *Synechococcus* sp. PCC 7942.** *Mol Microbiol*, **6**:653-664.

Varma A, Boesch BW, Palsson BØ. (1993a) **Biochemical production capabilities of *Escherichia coli***. *Biotechnol Bioeng,* **42**:59-73.

Varma A, Boesch BW, Palsson BØ. (1993b) **Stoichiometric interpretation of *Escherichia coli* glucose catabolism under various oxygenation rates**. *Appl Environ Microb,* **59**:2465-2473.

Varma A, Palsson BØ. (1993) **Metabolic capabilities of E coli II. Optimal growth patterns**. *J Theor Biol*, **165**:503-522.

Varma A, Palsson BØ. (1994a) **Metabolic Flux Balancing: Basic concepts, scientific and practical use**. *Nature*, **12**:994-998.

Varma A, Palsson BØ. (1994b) **Predictions for oxygen supply control to enhance population stability of engineered production strains**. *Biotechnol Bioeng,* **43**:275-285.

Varma A, Palsson BØ. (1994c) **Stoichiometric flux balance models quantitatively predict growth and metabolic by-product excretion in wild-type *Escherichia coli* W3110**. *Appl Environ Microb,* **60**:3724-3731.

Vo TD, Greenberg HJ, Palsson BØ. (2004) **Reconstruction and functional characterization of the human mitochondrial metabolic network based on proteomic and biochemical data**. *J Biol Chem,* **279**:39532-39540.

Voet D, Voet JG. (2004) **Biochemistry**. 3rd Ed. John Wiley & Sons, New York.

Vu TT, Stolyar SM, Pinchuk GE, *et al.* (2012) **Genome-scale modeling of light-driven reductant partitioning and carbon fluxes in diazotrophic unicellular cyanobacterium *Cyanothece* sp. ATCC 51142**. *PLoS Comput Biol,* **8(4)**:e1002460.

Wackett LP. (2011) **Engineering microbes to produce biofuels**. *Curr Opin Biotechnol*, **22**:1-6.

Wagner A, Fell DA. (2001) **The small world inside large metabolic networks**. *P Roy Soc Lond B Bio,* **268**:1803-1810.

Waterbury JB, Watson SW, Valois FW, Franks DG. (1986) **Biological and ecological characterization of the marine unicellular cyanobacterium *Synechococcus***. *Can Bull Fish Aquat Sci*, **214**:71-120.

Weise S, Grosse I, Klukas C *et al.* (2006) **Meta-All: a system for managing metabolic pathway information**. *BMC Bioinformatics*, **7**:465.

Werner A, Heinrich R. (1985). **A kinetic model for the interaction of energy metabolism and osmotic states of human erythrocytes. Analysis of the stationary ''*in vivo*'' state and of time dependent variations under blood preservation conditions**. *Biomed Biochim Acta*, **44**:185-212.

Weyman PD. (2010) **Expression of oxygen-tolerant hydrogenases in *Synechococcus elongatus***. *10th Cyanobacterial Molecular Biology Workshop*, Lake Arrowhead, USA.

Whitton BA, Potts M. (2000) **The ecology of cyanobacteria. Their diversity in time and space**. Kluwer Academic Publishers, Dordrecht, Netherlands.

Xu P, Ranganathan S, Fowler ZL, Maranas CD, Koffas MA. (2011) **Genome-scale metabolic network modeling results in minimal interventions that cooperatively force carbon flux towards malonyl-CoA**. *Metab Eng*, **13**:578-587.

Xu Y, Guerra LT, Li Z, Ludwig M, Dismukes GC, Bryant DA. (2012) **Altered carbohydrate metabolism in glycogen synthase mutants of *Synechococcus* sp. strain PCC** 7002: Cell factories for soluble sugars. *Metab Eng***, 16**:56-67.

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***. *Appl Microbiol Biot*, **58**:813-822.

Yang C, Hua Q, Shimizu K. (2002b) **Quantitative analysis of intracellular metabolic fluxes using GC-MS and two-dimensional NMR spectroscopy**. *J Biosci Bioeng*, **93**:78-87.

Yang Y, Gao K. (2003) **Effects of CO_2 concentrations on the freshwater microalgae, *Chlamydomonas reinhardtii*, *Chlorella pyrenoidosa* and *Scenedesmus obliquus* (Chlorophyta)**. *J Appl Phycol*, **15**:379-389.

Zha W, Rubin-Pitel SB, Shao Z, Zhao H. (2009) **Improving cellular malonyl-CoA level in *Escherichia coli* via metabolic engineering**. *Metab Eng*, **11**:192-198.

Appendixes

Appendix 1.1

*i*Syf715: *Synechococcus elongatus* PCC7942 genome-scale metabolic model

```
-REACTIONS

# GLYCOLYSIS
2.7.1.2a : beta-D-glucose + ATP -> beta-D-glucose-6-phosphate + ADP
2.7.1.2b : alpha-D-glucose + ATP -> alpha-D-glucose-6-phosphate + ADP
5.3.1.9a : alpha-D-glucose-6-phosphate <-> beta-D-glucose-6-phosphate
5.3.1.9b : beta-D-glucose-6-phosphate <-> beta-D-fructose-6-phosphate
2.7.1.11 : ATP + beta-D-fructose-6-phosphate -> ADP + beta-D-fructose-
1,6-bisphosphate
3.1.3.11 : beta-D-fructose-1,6-bisphosphate + H2O -> beta-D-fructose-
6-phosphate + phosphate O4P
4.1.2.13a : beta-D-fructose-1,6-bisphosphate <-> dihydroxy-acetone
phosphate + D-glyceraldehyde-3-phosphate
5.3.1.1 : D-glyceraldehyde-3-phosphate <-> dihydroxy-acetone phosphate
_1.2.1.59a : NAD+ + D-glyceraldehyde-3-phosphate + phosphate O4P ->
NADH + H+ + 1,3-diphosphateglycerate
_1.2.1.59b : NADP+ + D-glyceraldehyde-3-phosphate + phosphate O4P ->
NADPH + H+ + 1,3-diphosphateglycerate
1.2.1.12 : NAD+ + D-glyceraldehyde-3-phosphate + phosphate O4P <->
NADH + H+ + 1,3-diphosphateglycerate
2.7.2.3 : 1,3-diphosphateglycerate + ADP <-> 3-phosphoglycerate + ATP
5.4.2.1 : 3-phosphoglycerate <-> 2-phosphoglycerate
4.2.1.11 : 2-phosphoglycerate <-> phosphoenolpyruvate + H2O
2.7.1.40a : ADP + phosphoenolpyruvate -> ATP + pyruvate
2.7.9.2 : H2O + pyruvate + ATP -> phosphate O4P + phosphoenolpyruvate
+ AMP

# TCA CYCLE
2.3.3.1 : oxaloacetate + acetyl-CoA + H2O <-> citrate + coenzyme A
4.2.1.3a : citrate <-> cis-aconitate + H2O
4.2.1.3b : cis-aconitate + H2O <-> D-isocitrate
1.1.1.42a : D-isocitrate + NADP+ <-> oxalosuccinate + NADPH + H+
1.1.1.42b : oxalosuccinate <-> 2-ketoglutarate + CO2
_1.3.99.1 : succinate + UQ <-> fumarate + UQH2
4.2.1.2 : malate <-> fumarate + H2O

# GLYOXYLATE SHUNT
# not pres ! would made up glyoxylate shunt ~~IMPORTANT, following 2
reactions not in sequence!!~~
·4.1.3.1 : D-isocitrate -> glyoxylate + succinate
·2.3.3.9 : acetyl-CoA + H2O + glyoxylate <-> malate + coenzyme A
4.1.1.39a : D-ribulose-1,5-bisphosphate + O2 + H2O -> 3-
phosphoglycerate + 2-phosphoglycolate + 2 H+
_3.1.3.18 : 2-phosphoglycolate + H2O -> glycolate + phosphate O4P
_1.1.3.15 : glycolate + O2 <-> glyoxylate + H2O2
_4.1.1.2 : oxalate + H+ -> formate + CO2
```

```
# PENTOSE PHOSPHATE PATHWAY
1.1.1.49 : beta-D-glucose-6-phosphate + NADP+ <-> D-glucono-delta-
lactone-6-phosphate + NADPH + H+
3.1.1.31 : D-glucono-delta-lactone-6-phosphate + H2O -> 6-phospho-D-
gluconate
1.1.1.44 : 6-phospho-D-gluconate + NADP+ -> D-ribulose-5-phosphate +
CO2 + NADPH + H+
5.1.3.1 : D-ribulose-5-phosphate <-> D-xylulose-5-phosphate
5.3.1.6 : D-ribose-5-phosphate <-> D-ribulose-5-phosphate
5.3.1.9c : alfa-D-glucose-6-phosphate <-> beta-D-fructose-6-phosphate
2.2.1.1a : D-ribose-5-phosphate + D-xylulose-5-phosphate <-> D-
sedoheptulose-7-phosphate + D-glyceraldehyde-3-phosphate
2.2.1.1b : D-erythrose-4-phosphate + D-xylulose-5-phosphate <-> beta-
D-fructose-6-phosphate + D-glyceraldehyde-3-phosphate
_2.2.1.2 : beta-D-fructose-6-phosphate + D-erythrose-4-phosphate -> D-
glyceraldehyde-3-phosphate + D-sedoheptulose-7-phosphate
4.1.2.14 : 2-keto-3-deoxy-6-phospho-gluconate -> D-glyceraldehyde-3-
phosphate + pyruvate
4.1.2.4 : deoxyribose-5-phosphate <-> acetaldehyde + D-glyceraldehyde-
3-phosphate
1.1.1.47a : beta-D-glucose + NAD+ <-> D-glucono-1,5-lactone + NADH +
H+
1.1.1.47b : beta-D-glucose + NADP+ <-> D-glucono-1,5-lactone + NADPH +
H+

# CALVIN CYCLE ADDITION TO PPP
2.7.1.19 : D-ribulose-5-phosphate + ATP -> D-ribulose-1,5-bisphosphate
+ ADP
4.1.2.13b : dihydroxy-acetone phosphate + D-erythrose-4-phosphate ->
D-sedoheptulose-1,7-bisphosphate
3.1.3.37 : H2O + D-sedoheptulose-1,7-bisphosphate -> phosphate O4P +
D-sedoheptulose-7-phosphate
4.1.2.22 : beta-D-fructose-6-phosphate + phosphate O4P + 2 H+ ->
acetyl phosphate + D-erythrose 4-phosphate + H2O
# in "SPECIAL"
4.1.1.39b : D-ribulose-1,5-bisphosphate + CO2 + H2O -> 2 3-
phosphoglycerate + 2 H+

# AMINO ACIDS
# alanine - Ala
5.1.1.1 : L-alanine <-> D-alanine
1.4.1.1 : NAD+ + H2O + L-alanine <-> ammonia H3N + NADH + H+ +
pyruvate
tRNA 6.1.1.7 : tRNAala + L-alanine + ATP -> L-alanyl-tRNAala +
diphosphate + AMP
2.6.1.44 : L-alanine + glyoxylate -> pyruvate + glycine

# Asp, Arg, Asn, Lys and Thr
# aspartate - Asp
1.4.3.16a : O2 + H2O + L-aspartate -> ammonia H3N + H2O2 +
oxaloacetate
2.6.1.1a : L-aspartate + 2-ketoglutarate <-> L-glutamate +
oxaloacetate
# - from fumarate to Asp
4.3.2.2a : adenylo-succinate <-> fumarate + AMP
6.3.4.4 : L-aspartate + inosine-5'-phosphate + GTP <-> adenylo-
succinate + phosphate O4P + GDP
# - from Asp to Arg
2.1.3.3 : L-ornithine + carbamoyl-phosphate <-> citrulline + phosphate
O4P
```

```
6.3.4.5a : L-aspartate + citrulline + ATP <-> L-arginino-succinate +
diphosphate + AMP
4.3.2.1 : L-arginino-succinate <-> L-arginine + fumarate
# - from Asp to Asn
3.5.1.1 : L-asparagine + H2O <-> ammonia H3N + L-aspartate
# - from Asp to Thr
2.7.2.4 : L-aspartate + ATP -> L-aspartyl-4-phosphate + ADP
1.2.1.11 : NADPH + H+ + L-aspartyl-4-phosphate -> NADP+ + phosphate
O4P + L-aspartate-semialdehyde
_1.1.1.3a : L-aspartate-semialdehyde + NADPH + H+ <-> L-homoserine +
NADP+
_1.1.1.3b : L-homoserine + NAD+ <-> L-aspartate-semialdehyde + NADH +
H+
2.7.1.39 : L-homoserine + ATP -> O-phospho-L-homoserine + ADP
4.2.3.1a : O-phospho-L-homoserine + H2O -> phosphate O4P + L-threonine
# - from Asp to Lys
4.2.1.52 : pyruvate + L-aspartate-semialdehyde -> 2 H2O + L-2,3-
dihydrodipicolinate
_1.3.1.26a : tetrahydrodipicolinate + NADP+ <-> L-2,3-
dihydrodipicolinate + NADPH + H+
_1.3.1.26b : tetrahydrodipicolinate + NAD+ <-> L-2,3-
dihydrodipicolinate + NADH + H+
2.6.1.83 : L,L-diaminopimelate + 2-ketoglutarate <->
tetrahydrodipicolinate + L-glutamate + H2O
5.1.1.7 : L,L-diaminopimelate <-> meso-diaminopimelate
4.1.1.20 : meso-diaminopimelate -> CO2 + L-lysine
tRNA 6.1.1.12 : tRNAasp + L-aspartate + ATP -> L-aspartyl-tRNAasp +
diphosphate + AMP

# arginine - Arg
# - from Asp to Arg
tRNA 6.1.1.19 : tRNAarg + L-arginine + ATP -> L-arginyl-tRNAarg +
diphosphate + AMP
# - from Arg to putrescine
4.1.1.19 : L-arginine -> CO2 + agmatine
# not pres ! (would connect 4.1.1.19 to putrescine) ~~IMPORTANT,
following reaction not in sequence!!~~
·3.5.3.12 : agmatine + H2O -> N-carbamoyl putrescine + ammonia H3N
3.5.1.53 : N-carbamoyl putrescine + H2O -> putrescine + CO2 + ammonia
H3N

# asparagine - Asn
# - from Asp to Asn
tRNA 6.1.1.22 : tRNAasn + L-asparagine + ATP -> L-asparaginyl-tRNAasn
+ diphosphate + AMP
tRNA 6.3.5.6 : L-glutamine + L-aspartyl-tRNAAsn + ATP -> L-glutamate +
L-asparaginyl-tRNAAsn + phosphate O4P + ADP

# lysine - Lys
# - from Asp to Lys
tRNA 6.1.1.6 : tRNAlys + L-lysine + ATP -> L-lysyl-tRNAlys +
diphosphate + AMP
4.1.1.18 : L-lysine -> CO2 + cadaverine

# threonine - Thr
# - from Asp to Thr
THREOSPON-RXN : 2-amino-3-oxobutanoate <-> aminoacetone + CO2
# not pres ! (would connect 2-oxobutanoate whith Valine, Isoleucine
and Leucine biosynthesis) ~~IMPORTANT, following reaction not in
sequence!!~~
·4.3.1.19a : L-threonine -> 2-oxobutanoate + ammonia H3N
```

```
tRNA 6.1.1.3 : tRNAthr + L-threonine + ATP -> L-threonyl-tRNAthr +
diphosphate + AMP

# Ser, Cys and Gly
# serine - Ser
# - from 3PG to Ser
2.7.1.31 : glicerate + ATP <-> 3-phosphoglycerate + ADP
1.1.1.95 : 3-phosphoglycerate + NAD+ -> 3-phospho-hydroxypyruvate +
NADH + H+
# not pres ! (would connect pyruvate with this pathway) ~~IMPORTANT,
following reaction not in sequence!!~~
·_4.3.1.19b : L-serine <-> pyruvate + ammonia H3N
# - from Ser to Cys
2.3.1.30 : L-serine + acetyl-CoA <-> O-acetyl-L-serine + coenzyme A
2.5.1.47 : O-acetyl-L-serine + hydrogen sulfide <-> L-cysteine +
acetate
2.5.1.49a : O-acetyl-L-serine + thiosulfate + thioredoxin + H+ -> L-
cysteine + sulfite O3s + thioredoxin disulfide + acetate
# - from Ser to Gly (& viceversa)
2.1.2.1 : L-serine + tetrahydrofolate <-> 5,10-methylene-THF + glycine
+ H2O
2.6.1.45 : hydroxypyruvate + glycine <-> L-serine + glyoxylate
tRNA 6.1.1.11 : tRNAser + L-serine + ATP -> L-seryl-tRNAser +
diphosphate + AMP
2.6.1.51 : L-serine + pyruvate <-> hydroxypyruvate + L-alanine

# cysteine - Cys
# - from Ser to Cys
CYSPON-RXN : 3-sulfinyl-pyruvate + H2O -> sulfite O3s + pyruvate
2.6.1.1e : L-cysteine + 2-ketoglutarate <-> mercaptopyruvate + L-
glutamate
2.8.1.2 : mercaptopyruvate + sulfite O3s -> thiosulfate + pyruvate
tRNA 6.1.1.16 : tRNAcys + L-cysteine + ATP -> L-cysteinyl-tRNAcys +
diphosphate + AMP

# glycine - Gly
# - from Ser to Gly (& viceversa)
tRNA 6.1.1.14 : tRNAgly + glycine + ATP -> glycyl-tRNAgly +
diphosphate + AMP
3.4.11.1 : L-cysteine-glycine + H2O -> L-cysteine + glycine

# Glu, Gln, Pro and Trp
# glutamate - Glu
transport TRANS-RXN59G-639 : L-glutamate_extrac + H+_extrac -> L-
glutamate + H+
# - from Glu to Pro
2.7.2.11 : L-glutamate + ATP -> L-glutamate-5-phosphate + ADP
1.2.1.41 : L-glutamate gamma-semialdehyde + phosphate O4P + NADP+ <->
L-glutamate-5-phosphate + NADPH + H+
SPONTPRO-RXN : L-glutamate gamma-semialdehyde <-> H2O + pyrroline 5-
carboxylate
_1.5.1.2a1 : L-proline + NADP+ <-> pyrroline 5-carboxylate + NADPH +
H+
_1.5.1.2a2 : L-proline + NAD+ <-> pyrroline 5-carboxylate + NADH + H+
# - from Glu to Gln
6.3.1.2 : ammonia H3N + L-glutamate + ATP -> L-glutamine + ADP +
phosphate O4P
5.1.1.3 : L-glutamate <-> D-glutamate
1.4.7.1 : 2 L-glutamate + 2 Fd -> L-glutamine + 2-ketoglutarate + 2
Fd* + 2 H+
```

```
# not pres ! would interconvert D-Gln and L-Gln ~~IMPORTANT, following
reaction not in sequence!!~~
·5.1.1.9 : D-glutamine <-> L-glutamine
# - from Glu to Trp - chorismate comes from Phe biosynthesis
4.1.3.27 : chorismate + L-glutamine -> L-glutamate + anthranilate +
pyruvate
2.4.2.18 : N-(5'-phosphoribosyl)-anthranilate + diphosphate <->
anthranilate + 5-phosphoribosyl 1-pyrophosphate
5.3.1.24 : N-(5'-phosphoribosyl)-anthranilate <-> 1-(o-
carboxyphenylamino)-1'-deoxyribulose-5'-phosphate
4.1.1.48 : 1-(o-carboxyphenylamino)-1'-deoxyribulose-5'-phosphate ->
indole-3-glycerol-phosphate + CO2 + H2O
4.2.1.20a : indole-3-glycerol-phosphate + L-serine <-> L-tryptophan +
H2O + D-glyceraldehyde-3-phosphate
_4.2.1.20b : indole-3-glycerol-phosphate <-> indole + D-
glyceraldehyde-3-phosphate
4.2.1.20c : indole + L-serine -> L-tryptophan + H2O
2.6.1.1b : 3-sulfinoalanine + 2-ketoglutarate -> L-glutamate + 3-
sulfinyl-pyruvate
tRNA 6.1.1.17 : tRNAGlu + L-glutamate + ATP -> L-glutamyl-tRNAGlu +
diphosphate + AMP

# glutamine - Gln
transport TRANS-RXN59G-245 : L-glutamine_extrac + H+_extrac -> L-
glutamine + H+
# - from Glu to Gln
tRNA _6.3.5.7 : L-glutamine + L-glutamyl-tRNAGlu + ATP -> L-glutamate
+ L-glutaminyl-tRNAGln + phosphate O4P + ADP
# not pres ! would produce Gln's tRNA ~~IMPORTANT, following reaction
not in sequence!!~~
·tRNA 6.1.1.18 : tRNAgln + L-glutamine + ATP <-> L-glutaminyl-tRNAgln
+ diphosphate + AMP

# proline - Pro
# - from Glu to Pro
# - from Pro to pyruvate - not pres
_1.5.1.2b : L-1 pyrroline-3-hydroxy-5-carboxylate + NADH + H+ ->
trans-4-Hydroxy-L-proline + NAD+
_1.5.1.2c : L-1 pyrroline-3-hydroxy-5-carboxylate + NADPH + H+ ->
trans-4-Hydroxy-L-proline + NADP+
_2.6.1.1c : L-erythro-4-hydroxy-glutamate + 2-ketoglutarate -> D-4-
hydroxy-2-keto-glutarate + L-glutamate
4.1.3.16 : D-4-hydroxy-2-keto-glutarate <-> glyoxylate + pyruvate
_spont1 : L-1 pyrroline-3-hydroxy-5-carboxylate <-> L-4-hydroxy-
glutamate semialdehyde
tRNA 6.1.1.15 : L-proline + tRNApro + ATP -> L-prolyl-tRNApro +
diphosphate + AMP
5.2.1.8 : peptidylproline (omega = 180) -> peptidylproline (omega = 0)

# tryptophan - Trp
# - from Glu to Trp
tRNA 6.1.1.2 : ATP + tRNAtrp + L-tryptophan -> AMP + diphosphate + L-
tryptophanyl-tRNAtrp

# Phe and Tyr
# phenylalanine - Phe
# - chorismate production (common to Phe, Tyr and Trp)
2.5.1.54 : phosphoenolpyruvate + D-erythrose-4-phosphate + H2O -> 3-
deoxy-D-arabino-heptulosonate-7-phosphate + phosphate O4P
4.2.3.4 : 3-deoxy-D-arabino-heptulosonate-7-phosphate -> 3-
dehydroquinate + phosphate O4P
```

```
4.2.1.10 : 3-dehydroquinate -> H2O + 3-dehydro-shikimate
1.1.1.25 : NADP+ + shikimate <-> NADPH + H+ + 3-dehydro-shikimate
2.7.1.71 : shikimate + ATP -> shikimate-3-phosphate + ADP
2.5.1.19 : shikimate-3-phosphate + phosphoenolpyruvate <-> 5-
enolpyruvyl-shikimate-3-phosphate + phosphate O4P
4.2.3.5 : 5-enolpyruvyl-shikimate-3-phosphate -> phosphate O4P +
chorismate
# - from chorismate to prephenate (common to Phe and Tyr)
5.4.99.5 : chorismate <-> prephenate
# - from prephenate to Phe
4.2.1.51 : prephenate <-> phenylpyruvate + H2O + CO2
2.6.1.9b : phenylpyruvate + L-glutamate <-> L-phenylalanine + 2-
ketoglutarate
# - from Phe to Tyr (not pres)
4.2.1.96 : 4alpha-hydroxy-tetrahydrobiopterin -> dihydrobiopterin +
H2O
tRNA 6.1.1.20 : tRNAphe + L-phenylalanine + ATP -> L-phenylalanyl-
tRNAphe + diphosphate + AMP

# tyrosine - Tyr
# - chorismate production (common to Phe, Tyr and Trp)
# - from chorismate to prephenate (common to Phe and Tyr)
# - from prephenate to Tyr I
# not pres ! would interconnect prephenate with L-tyrosine
~~IMPORTANT, following reaction not in sequence!!~~
·1.3.1.12 : prephenate + NAD+ <-> 4-hydroxyphenylpyruvate + CO2 + NADH
+ H+
_2.6.1.1d : 4-hydroxyphenylpyruvate + L-glutamate <-> L-tyrosine + 2-
ketoglutarate
1.3.1.78 : L-arogenate + NADP+ -> L-tyrosine + CO2 + NADPH + H+
# - from prephenate to Tyr II - not pres
# - from Phe to Tyr - not pres
tRNA 6.1.1.1 : L-tyrosine + tRNAtyr + ATP -> L-tyrosyl-tRNAtyr +
diphosphate + AMP

# histidine - His
# - 5-phosphoribosyl 1-pyrophosphate comes from Pirimidine
biosynthesis
2.4.2.17 : phosphoribosyl-ATP + diphosphate <-> ATP + 5-phosphoribosyl
1-pyrophosphate
3.6.1.31 : phosphoribosyl-ATP + H2O -> phosphoribosyl-AMP +
diphosphate
3.5.4.19 : phosphoribosyl-AMP + H2O -> phosphoribosylformiminoAICAR-
phosphate
5.3.1.16 : phosphoribosylformiminoAICAR-phosphate ->
phosphoribulosylformimino-AICAR-P
4.1.3.- : phosphoribulosylformimino-AICAR-P + L-glutamine -> L-
glutamate + D-erythro-imidazole-glycerol-phosphate + aminoimidazole
carboxamide ribonucleotide
4.2.1.19 : D-erythro-imidazole-glycerol-phosphate -> imidazole acetol-
phosphate + H2O
2.6.1.9a : imidazole acetol-phosphate + L-glutamate <-> L-histidinol-
phosphate + 2-ketoglutarate
# not pres ! would close His production ~~IMPORTANT, following
reaction not in sequence!!~~
·3.1.3.15 : L-histidinol-phosphate + H2O <-> histidinol + phosphate
O4P
1.1.1.23a : histidinol + NAD+ -> histidinal + NADH + H+
1.1.1.23b : histidinal + NAD+ + H2O -> L-histidine + NADH + H+
```

```
tRNA 6.1.1.21 : tRNAhis + L-histidine + ATP -> L-histidyl-tRNAhis +
diphosphate + AMP

# Val, Ile and Leu
# valine - Val
_2.2.1.6d : 2 pyruvate -> 2-acetolactate + CO2
1.1.1.86a : 2,3-dihydroxy-isovalerate + NADP+ <-> 2-acetolactate +
NADPH + H+
4.2.1.9a : 2,3-dihydroxy-isovalerate -> 2-keto-isovalerate + H2O
2.6.1.42a : L-glutamate + 2-keto-isovalerate <-> L-valine + 2-
ketoglutarate
2.6.1.66 : L-valine + pyruvate <-> 3-methyl-2-oxobutanoic acid + L-
alanine
tRNA 6.1.1.9 : tRNAval + L-valine + ATP -> L-valyl-tRNAval
+diphosphate + AMP

# isoleucine - Ile
_2.2.1.6e : pyruvate + 2-oxobutanoate -> 2-aceto-2-hydroxy-butyrate +
CO2
_1.1.1.85b : D-erythro-3-methylmalate -> 2-oxobutanoate + CO2
1.1.1.86b : 2-aceto-2-hydroxy-butyrate + NADPH + H+ <-> 2,3-dihydroxy-
3-methylvalerate + NADP+
4.2.1.9b : 2,3-dihydroxy-3-methylvalerate -> 2-keto-3-methyl-valerate
+ H2O
2.6.1.42b : L-isoleucine + 2-ketoglutarate <-> L-glutamate + 2-keto-3-
methyl-valerate
tRNA 6.1.1.5 : tRNAile + L-isoleucine + ATP -> L-isoleucyl-tRNAile +
diphosphate + AMP

# leucine - Leu
2.3.3.13 : 2-keto-isovalerate + acetyl-CoA + H2O -> 2-isopropylmalate
+ coenzyme A
4.2.1.33a : 2-isopropylmalate <-> isopropylmaleate + H2O
4.2.1.33b : 3-isopropylmalate <-> isopropylmaleate + H2O
1.1.1.85a : 3-isopropylmalate + NAD+ -> 2-isopropyl-3-oxosuccinate +
NADH + H+
RXN-7800 : 2-isopropyl-3-oxosuccinate <-> 2-ketoisocaproate + CO2
2.6.1.42c : L-glutamate + 2-ketoisocaproate -> L-leucine + 2-
ketoglutarate
tRNA 6.1.1.4 : L-leucine + tRNAleu + ATP -> L-leucyl-tRNAleu +
diphosphate + AMP

# methionine - Met
# from L-homoserine to L-homocysteine
2.3.1.31 : acetyl-CoA + L-homoserine <-> coenzyme A + O-acetyl-L-
homoserine
2.3.1.46 : succinyl-CoA + L-homoserine <-> coenzyme A + O-succinyl-L-
homoserine
2.5.1.49b : O-acetyl-L-homoserine + hydrogen sulfide -> L-homocysteine
+ acetate
# from L-homocysteine to Met
3.3.1.1 : S-adenosyl-L-homocysteine + H2O <-> L-homocysteine +
adenosine
2.1.1.13 : L-homocysteine + 5-methyl-THF <-> L-methionine +
tetrahydrofolate
2.5.1.6 : ATP + L-methionine + H2O -> phosphate O4P + diphosphate + S-
adenosyl-L-methionine
2.1.1.37 : S-adenosyl-L-methionine + DNA <-> S-adenosyl-L-homocysteine
+ DNA 5-methylcytosine
4.1.1.50 : S-adenosyl-L-methionine + H+ -> S-adenosyl-metioninamine +
CO2
```

```
2.5.1.16a : S-adenosyl-metioninamine + putrescine -> spermidine + S-
methyl-5'-thioadenosine
2.4.2.28 : S-methyl-5'-thioadenosine + phosphate O4P -> adenine + 5-
methylthioribose-1-phosphate
5.3.1.23 : 5-methylthioribose-1-phosphate -> 5-methylthioribulose-1-
phosphate
4.2.1.109 : 5-methylthioribulose-1-phosphate -> 2,3-diketo-5-
methylthiopentyl-1-phosphate + H2O
3.1.3.77 : 2,3-diketo-5-methylthiopentyl-1-phosphate + H2O -> 1,2-
dihydroxy-3-keto-5-methylthiopentene + phosphate O4P
1.13.11.54 : 1,2-dihydroxy-3-keto-5-methylthiopentene + O2 -> 4-
methylthio-2-oxobutanoic acid + formate
1.13.11.53 : 1,2-dihydroxy-3-keto-5-methylthiopentene + O2 -> 3-
methylthio-propionato + formate + CO
tRNA 2.1.2.9b : 10-formyl-tetrahydrofolate + L-methionyl-tRNAmet + H2O
-> tetrahydrofolate + N-formyl-L-methionyl-tRNAfmet
tRNA 6.1.1.10 : tRNAmet + L-methionine + ATP -> L-methionyl-tRNAmet +
diphosphate + AMP

# additional aa reactions
3.5.2.10 : H2O + creatinine -> creatine
4.4.1.16 : selenocysteine + UQH2 <-> selenide + L-alanine + UQ

# NUCLEIC ACIDS
# Pyrimidine biosynthesis (C, T, U)
6.3.5.5 : 2 ATP + L-glutamine + HCO3- + H2O -> L-glutamate + 2 ADP +
phosphate O4P + carbamoyl-phosphate
2.1.3.2 : L-aspartate + carbamoyl-phosphate -> N-carbamoyl-L-aspartate
+ phosphate O4P
3.5.2.3 : dihydroorotate + H2O <-> N-carbamoyl-L-aspartate
1.3.98.1a : O2 + dihydroorotate <-> H2O2 + orotate
2.4.2.10 : orotidine-5'-phosphate + diphosphate <-> 5-phosphoribosyl
1-pyrophosphate + orotate
4.1.1.23 : orotidine-5'-phosphate -> CO2 + UMP
1.3.98.1b : dihydroorotate + fumarate <-> orotate + succinate

# - Uracil
3.1.3.5g : UMP + H2O <-> uridine + phosphate O4P
2.7.4.14a : ATP + UMP <-> ADP + UDP
_2.7.4.22 : ATP + UMP <-> ADP + UDP
2.4.2.9 : diphosphate + UMP <-> 5-phosphoribosyl 1-pyrophosphate +
uracil
3.6.1.19e : H2O + UTP -> diphosphate + UMP
2.7.4.6e : UDP + ATP <-> UTP + ADP
3.5.4.13a : H2O + CTP -> ammonia H3N + UTP
6.3.4.2 : ATP + UTP + ammonia H3N -> ADP + phosphate O4P + CTP
2.7.4.6h : dUDP + ATP <-> dUTP + ADP
3.6.1.19f : H2O + dUTP -> diphosphate + dUMP
2.7.4.9b : ATP + dUMP <-> ADP + dUDP
1.17.4.1e : dUDP + thioredoxin + H2O <-> UDP + thioredoxin disulfide
2.1.1.148 : 5,10-methylene-THF + dUMP + FADH2 -> dTMP +
tetrahydrofolate + FAD
2.7.7.8c : RNAn1 + phosphate O4P -> RNAn + UDP
2.7.7.6c : UTP + RNAn -> diphosphate + RNAn1
3.5.4.5a : cytidine + H2O <-> uridine + ammonia H3N
3.5.4.1a : H2O + cytosine -> ammonia H3N + uracil

# - Cytosine
2.7.4.6f : CDP + ATP <-> CTP + ADP
2.7.4.14b : ATP + CMP <-> ADP + CDP
```

```
1.17.4.1d : dCDP + thioredoxin + H2O <-> CDP + thioredoxin disulfide
3.1.3.5h : CMP + H2O <-> cytidine + phosphate O4P
2.7.4.6g : dCDP + ATP <-> dCTP + ADP
2.7.4.14c : ATP + dCMP <-> ADP + dCDP
3.1.3.5i : dCMP + H2O <-> deoxycytidine + phosphate O4P
3.5.4.5b : deoxycytidine + H2O <-> deoxyuridine + ammonia H3N
3.5.4.13b : H2O + dCTP -> ammonia H3N + dUTP
2.7.7.6d : CTP + RNAn -> diphosphate + RNAn1
2.7.7.8d : RNAn1 + phosphate O4P -> RNAn + CDP
2.7.7.7c : dCTP + DNAn <-> diphosphate + DNAn1
# - Thymine
3.1.3.5j : thymidine + H2O <-> dTMP + phosphate O4P
2.7.4.9a : ATP + dTMP <-> ADP + dTDP
2.7.4.6i : dTDP + ATP <-> dTTP + ADP
2.7.7.7d : dTTP + DNAn <-> diphosphate + DNAn1
_3.5.4.1b : H2O + thymine -> ammonia H3N + 5-methylcytosine

# Purine metabolism (A, G)
3.6.1.13 : ADP-ribose + H2O -> AMP + D-ribose-5-phosphate
2.7.6.1 : ATP + D-ribose-5-phosphate -> 5-phosphoribosyl 1-
pyrophosphate + AMP
2.4.2.14 : 5-phosphoribosyl 1-pyrophosphate + L-glutamine + H2O -> 5-
phospho-beta-D-ribosyl-amine + diphosphate + L-glutamate
6.3.4.13 : 5-phospho-beta-D-ribosyl-amine + ATP + glycine -> ADP +
phosphate O4P + 5-phospho-ribosyl-glycineamide
2.1.2.2a : 10-formyl-tetrahydrofolate + 5-phospho-ribosyl-glycineamide
<-> tetrahydrofolate + 5'-phosphoribosyl-N-formylglycineamide
6.3.5.3 : ATP + 5'-phosphoribosyl-N-formylglycineamide + L-glutamine +
H2O -> L-glutamate + ADP + phosphate O4P + 5-phosphoribosyl-N-
formylglycineamidine
6.3.3.1 : ATP + 5-phosphoribosyl-N-formylglycineamidine -> ADP +
phosphate O4P + 5-aminoimidazole ribonucleotide
6.3.4.18 : ATP + 5-aminoimidazole ribonucleotide + HCO3- -> ADP +
phosphate O4P + 5-carboxyamino-1-(5-phospho-D-ribosyl)imidazole
5.4.99.18 : 5-carboxyamino-1-(5-phospho-D-ribosyl)imidazole <-> 4-
carboxyaminoimidazole ribonucleotide
6.3.2.6 : ATP + 4-carboxyaminoimidazole ribonucleotide + L-aspartate
<-> ADP + phosphate O4P + 5'-phosphoribosyl-4-(N-succinocarboxamide)-
5-aminoimidazole
4.3.2.2b : 5'-phosphoribosyl-4-(N-succinocarboxamide)-5-aminoimidazole
<-> fumarate + aminoimidazole carboxamide ribonucleotide
2.1.2.3a : 10-formyl-tetrahydrofolate + aminoimidazole carboxamide
ribonucleotide <-> tetrahydrofolate + phosphoribosyl-formamido-
carboxamide
2.4.2.7c : aminoimidazole carboxamide ribonucleotide <-> 5-amino-4-
imidazolecarboxyamide

# - Inosine
3.5.4.10 : inosine-5'-phosphate + H2O <-> phosphoribosyl-formamido-
carboxamide
2.1.2.3b : inosine-5'-phosphate + H2O <-> phosphoribosyl-formamido-
carboxamide
3.1.3.5c : inosine-5'-phosphate + H2O <-> inosine + phosphate O4P
_2.7.4.3c : inosine-5'-phosphate + ITP <-> 2 IDP
3.6.1.19a : H2O + ITP -> diphosphate + inosine-5'-phosphate
3.6.1.19g : H2O + dITP -> diphosphate + dIMP
_1.17.4.1b : dIDP + thioredoxin + H2O <-> IDP + thioredoxin disulfide
2.7.4.6j : IDP + ATP <-> ITP + ADP
2.7.4.6b : dIDP + ATP <-> dITP + ADP
1.1.1.205 : H2O + NAD+ + inosine-5'-phosphate <-> xanthosine-5-
phosphate + NADH + H+
```

```
3.5.4.4a : adenosine + H2O -> inosine + ammonia H3N
# - Xanthosine
3.1.3.5d : xanthosine-5-phosphate + H2O <-> xanthosine + phosphate O4P
3.6.1.19b : H2O + XTP -> diphosphate + xanthosine
6.3.5.2 : xanthosine-5-phosphate + H2O + L-glutamine + ATP -> L-
glutamate + GMP + diphosphate + AMP
# - Adenine
3.1.3.5a : AMP + H2O <-> adenosine + phosphate O4P
2.4.2.7a : AMP + diphosphate <-> 5-phosphoribosyl 1-pyrophosphate +
adenine
2.7.4.3a : AMP + ATP <-> 2 ADP
2.7.4.6a : dADP + ATP <-> dATP + ADP
2.7.1.40b : dATP + pyruvate <-> dADP + phosphoenolpyruvate
1.17.4.1a : dADP + thioredoxin + H2O <-> ADP + thioredoxin disulfide
2.7.4.3b : ATP + dAMP <-> ADP + dADP
3.1.3.5b : dAMP + H2O <-> 5'-deoxyadenosine + phosphate O4P
2.7.7.7a : dATP + DNAn <-> diphosphate + DNAn1
2.7.7.6a : ATP + RNAn -> diphosphate + RNAn1
2.7.7.8a : RNAn1 + phosphate O4P -> RNAn + ADP
3.5.4.4b : 5'-deoxyadenosine + H2O <-> deoxyinosine + ammonia H3N
# - Guanine
3.1.3.5e : GMP + H2O <-> guanosine + phosphate O4P
2.4.2.7b : GMP + diphosphate <-> 5-phosphoribosyl 1-pyrophosphate +
guanine
2.7.4.8a : GMP + ATP <-> GDP + ADP
3.6.1.19c : H2O + GTP -> diphosphate + GMP
1.17.4.1c : dGDP + thioredoxin + H2O <-> GDP + thioredoxin disulfide
2.7.4.6c : GDP + ATP <-> GTP + ADP
2.7.1.40c : GTP + pyruvate <-> GDP + phosphoenolpyruvate
2.7.4.8b : dGMP + ATP <-> dGDP + ADP
2.7.4.6d : dGDP + ATP <-> dGTP + ADP
2.7.1.40d : dGTP + pyruvate <-> dGDP + phosphoenolpyruvate
3.1.3.5f : dGMP + H2O <-> deoxyguanosine + phosphate O4P
3.6.1.19d : H2O + dGTP -> diphosphate + dGMP
3.1.7.2 : guanosine 5'-diphosphate,3'-diphosphate + H2O <->
diphosphate + GDP
3.6.1.11 : H2O + guanosine 3'-diphosphate 5'-triphosphate <->
phosphate O4P + guanosine 5'-diphosphate,3'-diphosphate
2.7.7.6b : GTP + RNAn -> diphosphate + RNAn1
2.7.7.8b : RNAn1 + phosphate O4P -> RNAn + GDP
2.7.7.7b : dGTP + DNAn <-> diphosphate + DNAn1

# - ApppA production
2.7.7.4 : sulfate O4s + ATP -> adenosine 5'-phosphosulfate +
diphosphate
2.7.1.25 : adenosine 5'-phosphosulfate + ATP -> phosphoadenosine-5'-
phosphosulfate + ADP
_2.7.7.53a : 2 adenosine 5'-phosphosulfate + ATP -> sulfate O4s +
phosphate O4P + 5',5'''-diadenosine tetraphosphate + AMP
2.7.7.53b : ATP + ADP <-> phosphate O4P + 5',5'''-diadenosine
tetraphosphate
# - Purine cyclases
4.6.1.1a : ATP -> cyclic-AMP + diphosphate
4.6.1.1b : GTP -> cyclic-GMP + diphosphate

# Folate biosynthesis
3.5.4.16 : GTP + H2O -> formate + 7,8-dihydroneopterin triphosphate
4.1.2.25 : 7,8-dihydro-D-neopterin <-> glycolaldehyde + 6-
hydroxymethyl-7,8-dihydropterin
```

```
2.7.6.3 : 6-hydroxymethyl-7,8-dihydropterin + ATP -> (2-amino-4-
hydroxy-7,8-dihydropteridin-6-yl)methyl diphosphate + AMP
2.5.1.15a : p-aminobenzoate + (2-amino-4-hydroxy-7,8-dihydropteridin-
6-yl)methyl diphosphate -> 7,8-dihydropteroate + diphosphate
2.5.1.15b : p-aminobenzoate + 6-hydroxymethyl-7,8-dihydropterin ->
7,8-dihydropteroate + H2O
2.6.1.85 : L-glutamine + chorismate <-> L-glutamate + 4-amino-4-
deoxychorismate
# not pres ! (would connect formate to folate biosynthesis)
~~IMPORTANT, following reaction not in sequence!!~~
·6.3.4.3 : tetrahydrofolate + formate + ATP <-> ADP + phosphate O4P +
10-formyl-tetrahydrofolate
# not pres ! (would connect chorismate to folate biosynthesis and p-
aminobenzoate to aa production) ~~IMPORTANT, following reaction not in
sequence!!~~
·4.1.3.38 : 4-amino-4-deoxychorismate <-> p-aminobenzoate + pyruvate
6.3.2.12 : L-glutamate + 7,8-dihydropteroate + ATP -> 7,8-
dihydrofolate + phosphate O4P + ADP
# not pres ! (would complete this pathway) ~~IMPORTANT, following
reaction not in sequence!!~~
·1.5.1.3 : NADP+ + tetrahydrofolate <-> NADPH + H+ + 7,8-dihydrofolate
# - from tetrahydrofolate to a tetrahydrofolate polyglutamate
6.3.2.17a : tetrahydrofolate + L-glutamate + ATP <-> tetrahydrofolate-
L-glutamate + ADP + phosphate O4P
6.3.2.17b : tetrahydrofolate-L-glutamate + L-glutamate + ATP <-> a
tetrahydrofolate polyglutamate + ADP + phosphate O4P
# - from 7,8-dihydropteroate to dihydrobiopterin
6.3.2.17c : L-glutamate + 7,8-dihydropteroate + ATP -> 7,8-
dihydrofolate + phosphate O4P + ADP
_4.2.3.12 : 7,8-dihydroneopterin triphosphate -> 6-pyruvoyl-5,6,7,8-
tetrahydropterin + diphosphate
# - one carbon pool by folate
2.1.2.2b : H2O + 5,10-methenyltetrahydrofolate + 5-phospho-ribosyl-
glycineamide -> tetrahydrofolate + 5'-phosphoribosyl-N-
formylglycineamide
_1.4.4.2 : H-protein-lipoyllysine + glycine -> H-Protein-S-
aminomethyldihydrolipoyllysine + CO2
_2.1.2.10a : H-Protein-S-aminomethyldihydrolipoyllysine +
tetrahydrofolate -> [H protein]-dihydrolipoyllysine + 5,10-methylene-
THF + ammonia H3N
_1.8.1.4c : [H protein]-dihydrolipoyllysine + NAD+ -> H-protein-
lipoyllysine + NADH + H+
1.5.1.5 : 5,10-methylene-THF + NADP+ <-> NADPH + H+ + 5,10-
methenyltetrahydrofolate
3.5.1.10 : H2O + 10-formyl-tetrahydrofolate -> tetrahydrofolate +
formate
3.5.4.9a : H2O + 5,10-methenyltetrahydrofolate <-> 10-formyl-
tetrahydrofolate
6.3.3.2 : 5-formyl-tetrahydrofolate + ATP -> 5,10-
methenyltetrahydrofolate + ADP + phosphate O4P
2.1.2.10b : 5,10-methenyltetrahydrofolate + H2O <-> 5-formyl-
tetrahydrofolate
# not pres ! (would complete this pathway, reaction 2.1.1.13 HAS TO
CONSUME 5-methyl-THF) ~~IMPORTANT, following 2 reactions not in
sequence!!~~
·1.5.1.20a : 5-methyl-THF + NADP+ <-> 5,10-methylene-THF + NADPH + H+
·1.5.1.20b : 5-methyl-THF + NAD+ <-> 5,10-methylene-THF + NADH + H+
2.1.2.9a : 10-formyl-tetrahydrofolate + L-methionyl-tRNAfmet + H2O ->
tetrahydrofolate + N-formyl-L-methionyl-tRNAfmet

# Photosynthesis
```

```
# in "SPECIAL" _lightII : photonII_extrac -> photonII
# in "SPECIAL" _lightI : photonI_extrac -> photonI
# the following two reactions are catalyzed by 1.10.3.9 (according to
ExPaSy and BRENDA databases)
_PSII : 2 photonII + H2O + PSII -> 0.5 O2 + 2 H+_thylac + PSII*
_UQ : PSII* + UQ + 2 H+ -> PSII + UQH2
_citb6 : UQH2 + citb6f + 2 H+ -> UQ + citb6f* + 4 H+_thylac
_1.10.9.1 : citb6f* + 2 PC -> citb6f + 2 PC*
# the following two reactions are catalyzed by 1.97.1.1 (according to
ExPaSy and BRENDA databases)
_PSI : photonI + PC* + PSI -> PC + PSI*
_Fd : PSI* + Fd -> PSI + Fd*
_FNR : 2 Fd* + FNR -> 2 Fd + FNR*
_1.18.1.2 : FNR* + NADP+ -> FNR + NADPH + H+
_3.6.3.14 : 4 H+_thylac + phosphate O4P + ADP <-> 4 H+ + H2O + ATP
3.6.1.1 : H2O + diphosphate -> 2 phosphate O4P
2.7.4.1 : ATP + diphosphate <-> ADP + PPPi
#_NADPH + H+2 : NADPH + H+ -> NADPH + H+2 + NADP+
#_BidHyd : NADPH + H+2 + 2 H+ <-> H2
_NADH2 : NADH + H+ -> NADH2 + NAD+
_BidHyd : NADH2 + 2 H+ <-> H2

# - cyclic part
_cyclic : 2 Fd* + UQ + 2 H+ -> 2 Fd + UQH2
_NDH1 : NADPH + H+ + 2 PC -> NADP+ + 2 PC*

# Oxidative phosphorylation
_1.6.5.3 : NADH + UQ + 5 H+ -> NAD+ + UQH2 + 4 H+_thylac
_1.6.99.3 : NADH + UQ + 5 H+ -> NAD+ + UQH2 + 4 H+_thylac
_quinol oxidase : 2 UQH2 + O2 + 4 H+ -> 2 UQ + 2 H2O + 4 H+_thylac
_cit c : citb6f* + 2 cit c -> citb6f + 2 cit c*
_1.9.3.1 : 4 cit c* + O2 + 8 H+ -> 4 cit c + 4 H+_thylac + 2 H2O
_PSIaltern : photonI + cit c* + PSI -> cit c + PSI*

# Urea Cycle and metabolism of amino groups
2.3.1.35 : L-glutamate + N-acetyl-L-ornithine <-> N-acetyl-L-glutamate
+ L-ornithine
2.6.1.11 : N-acetyl-L-ornithine + 2-ketoglutarate <-> L-glutamate + N-
acetyl-L-glutamate 5-semialdehyde
1.2.1.38 : N-acetyl-L-glutamate 5-semialdehyde + NADP+ + phosphate O4P
<-> N-acetylglutamyl-phosphate + NADPH + H+
2.7.2.8 : N-acetyl-L-glutamate + ATP -> N-acetylglutamyl-phosphate +
ADP
2.3.1.1 : L-glutamate + acetyl-CoA -> N-acetyl-L-glutamate + coenzyme
A
4.2.1.1a : H2CO3 <-> H2O + CO2
_1.4.3.4 : O2 + H2O + N-acetylputrescine -> ammonia H3N + H2O2 + N4-
acetylaminobutanal
_1.2.1.3a : N4-acetylaminobutanal + NAD+ + H2O <-> N4-
acetylaminobutanoate + NADH + H+
_1.2.1.3b : 4-aminobutanal + NAD+ + H2O <-> 4-aminobutanoate + NADH +
H+

# Nitrogen metabolism
1.7.7.2 : nitrite + 2 Fd + H2O <-> 2 Fd* + nitrate NO3 + 2 H+
1.7.7.1 : ammonia H3N + 2 H2O + 6 Fd <-> nitrite + 6 Fd* + 7 H+
4.2.1.104 : cyanate + HCO3- + H+ <-> carbamate + CO2
3.5.5.1 : nitrile + 2 H2O <-> carboxylate + ammonia H3N

# Nicotinate and nicotinamide metabolism (NAD biosynthesis)
```

```
1.4.3.16b : O2 + L-aspartate -> H2O2 + iminoaspartate
_2.5.1.72 : iminoaspartate + dihydroxy-acetone phosphate ->
quinolinate + 2 H2O + phosphate O4P
2.4.2.19 : 5-phosphoribosyl 1-pyrophosphate + quinolinate -> CO2 +
diphosphate + nicotinate nucleotide
3.1.3.5k : nicotinate nucleotide + H2O -> nicotinate D-ribonucleoside
+ phosphate O4P
2.7.7.1a : nicotinate nucleotide + ATP -> deamido-NAD + diphosphate
2.7.7.18a : ATP + nicotinate nucleotide -> diphosphate + deamido-NAD
6.3.1.5 : ammonia H3N + ATP + deamido-NAD -> AMP + diphosphate + NAD+
2.7.7.1b : nicotinamide mononucleotide + ATP -> NAD+ + diphosphate
2.7.7.18b : ATP + nicotinamide mononucleotide -> diphosphate + NAD+
3.1.3.5l : nicotinamide mononucleotide + H2O -> N-ribosylnicotinamide
+ phosphate O4P
2.7.1.23 : NAD+ + ATP -> NADP+ + ADP
1.6.1.2 : NAD+ + NADPH <-> NADH + NADP+

# Ubiquinone biosynthesis (not complete, should be! -> forced to be
complete)
# - Ubiquinone branch
# not pres ! would begin UQ branch ~~IMPORTANT, following 2 reactions
not in sequence!!~~
·4.1.3.40 : chorismate <-> 4-hydroxybenzoate + pyruvate
·octaprenylsyn : all-trans-geranyl-geranyl diphosphate -> all-trans-
octaprenyl diphosphate
2.5.1.-a : all-trans-octaprenyl diphosphate + 4-hydroxybenzoate -> 3-
octaprenyl-4-hydroxybenzoate + diphosphate
4.1.1.-a : 3-octaprenyl-4-hydroxybenzoate -> 2-octaprenylphenol + CO2
# not pres ! would allow UQ biosynthesis ~~IMPORTANT, following
reaction not in sequence!!~~
·OCTAPRENYLPHENOL-HYDROX-RXN : 2 2-octaprenylphenol + O2 <-> 2 2-
octaprenyl-6-hydroxyphenol
·2.1.1.64b : S-adenosyl-L-methionine + 2-octaprenyl-6-hydroxyphenol <-
> S-adenosyl-L-homocysteine + 2-octaprenyl-6-methoxyphenol
1.14.13.-a : 2-octaprenyl-6-methoxyphenol + O2 + NADPH + H+ -> 2-
octaprenyl-6-methoxy-1,4-benzoquinone + NADP+ + H2O
2.1.1.201 : 2-octaprenyl-6-methoxy-1,4-benzoquinone + S-adenosyl-L-
methionine -> S-adenosyl-L-homocysteine + 2-octaprenyl-3-methyl-6-
methoxy-1,4-benzoquinone
1.14.13.-b : 2-octaprenyl-3-methyl-6-methoxy-1,4-benzoquinone + O2 +
NADPH + H+ <-> 3-demethylubiquinone-8 + H2O + NADP+
# not pres ! would allow UQ biosynthesis ~~IMPORTANT, following
reaction not in sequence!!~~
·2.1.1.64a : S-adenosyl-L-methionine + 2-octaprenyl-3-methyl-5-
hydroxy-6-methoxy-1,4-benzoquinone <-> S-adenosyl-L-homocysteine +
ubiquinone-8
_UQsyf : ubiquinone-8 -> UQ
# - Phylloquinone/Menaquinone branch
5.4.4.2 : chorismate -> isochorismate
2.2.1.9 : isochorismate + 2-ketoglutarate <-> 5-enolpyruvoyl-6-
hydroxy-2-succinyl-cyclohex-3-ene-1-carboxylate + CO2
4.2.99.20 : 5-enolpyruvoyl-6-hydroxy-2-succinyl-cyclohex-3-ene-1-
carboxylate <-> 2-succinyl-6-hydroxy-2,4-cyclohexadiene-1-carboxylate
+ pyruvate
4.2.1.113 : 2-succinyl-6-hydroxy-2,4-cyclohexadiene-1-carboxylate ->
o-succinylbenzoate + H2O
6.2.1.26 : ATP + o-succinylbenzoate + coenzyme A -> AMP + diphosphate
+ O-succinylbenzoyl-CoA
4.1.3.36 : O-succinylbenzoyl-CoA <-> coenzyme A + 1,4-dihydroxy-2-
naphthoate
```

```
2.5.1.74a : all-trans-octaprenyl diphosphate + 1,4-dihydroxy-2-
naphthoate -> demethylmenaquinone-8 + diphosphate + CO2
2.5.1.74b : 1,4-dihydroxy-2-naphthoate + phytyl diphosphate ->
demethylphylloquinone + CO2 + diphosphate
2.1.1.163a : demethylmenaquinone-8 + S-adenosyl-L-methionine -> S-
adenosyl-L-homocysteine + menaquinone-8
2.1.1.163b : demethylphylloquinone + S-adenosyl-L-methionine ->
phylloquinone + S-adenosyl-L-homocysteine
3.1.2.28 : 1,4-dihydroxy-2-naphthoyl-CoA + H2O -> 1,4-dihydroxy-2-
naphthoate + CoA

# Sulfur metabolism
1.8.4.8 : phosphoadenosine-5'-phosphosulfate + thioredoxin <-> sulfite
O3s + adenosine-3',5'-bisphosphate + thioredoxin disulfide
# not pres ! ~~IMPORTANT, following reaction not in sequence!!~~
·1.13.11.8 : sulfite O3s -> SO + O2 + H2O
1.8.7.1 : sulfite O3s + 6 Fd* + 6 H+ <-> hydrogen sulfide + 6 Fd + 3
H2O
3.1.3.7 : phosphoadenosine-5'-phosphosulfate + H2O -> adenosine 5'-
phosphosulfate + phosphate O4P

# Ethanol metabolism (part of Glycolysis)
6.2.1.1a : coenzyme A + acetate + ATP <-> acetyl-CoA + diphosphate +
AMP
1.2.1.3c : acetaldehyde + NAD+ + H2O <-> NADH + H+ + acetate
1.1.1.1 : acetaldehyde + NADH + H+ -> ethanol + NAD+
2.7.2.1a : ATP + acetate <-> ADP + acetylphosphate

# coenzyme A biosynthesis
_1.1.1.86c : 2-acetolactate + NADPH + H+ -> 2,3-dihydroxy-3-
methylbutanoate + NADP+
_4.2.1.9c : 2,3-dihydroxy-3-methylbutanoate -> 2-keto-isovalerate +
H2O
2.1.2.11 : 5,10-methylene-THF + 2-keto-isovalerate + H2O ->
tetrahydrofolate + 2-dehydropantoate
6.3.2.1 : beta-alanine + R-pantoate + ATP -> pantothenate +
diphosphate + AMP
_1.2.1.3d : beta-aminopropion aldehyde + NAD+ + H2O <-> NADH + H+ +
beta-alanine
2.7.1.33 : pantothenate + ATP -> D-4'-phosphopantothenate + ADP
6.3.2.5 : D-4'-phosphopantothenate + L-cysteine + CTP -> R-4'-
phosphopantothenoyl-L-cysteine + diphosphate + CMP
4.1.1.36 : R-4'-phosphopantothenoyl-L-cysteine -> pantetheine 4'-
phosphate + CO2
_2.7.7.3 : ATP + pantetheine 4'-phosphate -> diphosphate + 3'-
dephospho-CoA
_2.7.1.24 : ATP + 3'-dephospho-CoA -> ADP + coenzyme A


# Fatty acid biosynthesis
_6.4.1.2a : ATP + acetyl-CoA + HCO3- + H+ -> ADP + phosphate O4P +
malonyl-CoA
_6.4.1.2b : carboxybiotin-carboxyl-carrier protein + acetyl-CoA ->
biotin-BCCP (monomer) + malonyl-CoA
_6.3.4.14 : ATP + biotin-BCCP (monomer) + CO2 -> ADP + phosphate O4P +
carboxybiotin-carboxyl-carrier protein
2.3.1.39 : malonyl-CoA + an [acyl-carrier protein] -> coenzyme A + a
malonyl-ACP
2.3.1.41a : acetyl-CoA + an [acyl-carrier protein] -> an acetyl-ACP +
coenzyme A + CO2
```

```
2.3.1.41b : an acetyl-ACP + a malonyl-ACP -> acetoacetyl-[acp]
C4H5O2SR + CO2
1.1.1.100a : acetoacetyl-[acp] C4H5O2SR + NADPH + H+ <-> (3R)-3-
Hydroxybutanoyl-[acyl-carrier protein] C4H7O2SR + NADP+
4.2.1.-a : (3R)-3-Hydroxybutanoyl-[acyl-carrier protein] C4H7O2SR <->
but-2-enoyl-[acyl-carrier protein] C4H5OSR + H2O
1.3.1.9a : but-2-enoyl-[acyl-carrier protein] C4H5OSR + NADH + H+ <->
butyryl-[acyl-carrier protein] C4H7OSR + NAD+
2.3.1.41c : butyryl-[acyl-carrier protein] C4H7OSR + a malonyl-ACP ->
3-oxohexanoyl-[acyl-carrier protein] C6H9O2SR + CO2
1.1.1.100b : 3-oxohexanoyl-[acyl-carrier protein] C6H9O2SR + NADPH +
H+ <-> (R)-3-Hydroxyhexanoyl-[acyl-carrier protein] C6H11O2SR + NADP+
4.2.1.-b : (R)-3-Hydroxyhexanoyl-[acyl-carrier protein] C6H11O2SR <->
trans-hex-2-enoyl-[acyl-carrier protein] C6H9OSR + H2O
1.3.1.9b : trans-hex-2-enoyl-[acyl-carrier protein] C6H9OSR + NADH +
H+ <-> hexanoyl-[acyl-carrier protein] C6H11OSR + NAD+
2.3.1.41d : hexanoyl-[acyl-carrier protein] C6H11OSR + a malonyl-ACP -
> 3-oxooctanoyl-[acyl-carrier protein] C8H13O2SR + CO2 + an [acyl-
carrier protein]
1.1.1.100c : 3-oxooctanoyl-[acyl-carrier protein] C8H13O2SR + NADPH +
H+ <-> (R)-3-Hydroxyoctanoyl-[acyl-carrier protein] C8H15O2SR + NADP+
4.2.1.-c : (R)-3-Hydroxyoctanoyl-[acyl-carrier protein] C8H15O2SR <->
trans-oct-2-enoyl-[acyl-carrier protein] C8H13OSR + H2O
1.3.1.9c : trans-oct-2-enoyl-[acyl-carrier protein] C8H13OSR + NADH +
H+ <-> octanoyl-[acyl-carrier protein] C8H15OSR + NAD+
2.3.1.41e : octanoyl-[acyl-carrier protein] C8H15OSR + a malonyl-ACP -
> 3-oxodecanoyl-[acyl-carrier protein] C10H17O2SR + CO2 + an [acyl-
carrier protein]
1.1.1.100d : 3-oxodecanoyl-[acyl-carrier protein] C10H17O2SR + NADPH +
H+ <-> (R)-3-Hydroxydecanoyl-[acyl-carrier protein] C10H19O2SR + NADP+
1.3.1.9d : trans-dec-2-enoyl-[acyl-carrier protein] C10H17OSR + NADH +
H+ <-> decanoyl-[acyl-carrier protein] C10H19OSR + NAD+
4.2.1.-d : (R)-3-Hydroxydecanoyl-[acyl-carrier protein] C10H19O2SR <->
trans-dec-2-enoyl-[acyl-carrier protein] C10H17OSR + H2O
2.3.1.41f : decanoyl-[acyl-carrier protein] C10H19OSR + a malonyl-ACP
-> 3-oxododecanoyl-[acyl-carrier protein] C12H21O2SR + CO2 + an [acyl-
carrier protein]
1.1.1.100e : 3-oxododecanoyl-[acyl-carrier protein] C12H21O2SR + NADPH
+ H+ <-> (R)-3-Hydroxydodecanoyl-[acyl-carrier protein] C12H23O2SR +
NADP+
4.2.1.-e : (R)-3-Hydroxydodecanoyl-[acyl-carrier protein] C12H23O2SR
<-> trans-dodec-2-enoyl-[acyl-carrier protein] C12H21OSR + H2O
1.3.1.9e : trans-dodec-2-enoyl-[acyl-carrier protein] C12H21OSR + NADH
+ H+ <-> dodecanoyl-[acyl-carrier protein] C12H23OSR + NAD+
2.3.1.41g : dodecanoyl-[acyl-carrier protein] C12H23OSR + a malonyl-
ACP -> 3-oxotetradecanoyl-[acyl-carrier protein] C14H25O2SR + CO2 + an
[acyl-carrier protein]
1.1.1.100f : 3-oxotetradecanoyl-[acyl-carrier protein] C14H25O2SR +
NADPH + H+ <-> (R)-3-Hydroxytetradecanoyl-[acyl-carrier protein]
C14H27O2SR + NADP+
4.2.1.-f : (R)-3-Hydroxytetradecanoyl-[acyl-carrier protein]
C14H27O2SR <-> trans-tetradec-2-enoyl-[acyl-carrier protein] C14H25OSR
+ H2O
1.3.1.9f : trans-tetradec-2-enoyl-[acyl-carrier protein] C14H25OSR +
NADH + H+ <-> tetradecanoyl-[acyl-carrier protein] C14H27OSR + NAD+
2.3.1.41h : tetradecanoyl-[acyl-carrier protein] C14H27OSR + a
malonyl-ACP -> 3-oxohexadecanoyl-[acyl-carrier protein] C16H29O2SR +
CO2 + an [acyl-carrier protein]
1.1.1.100g : 3-oxohexadecanoyl-[acyl-carrier protein] C16H29O2SR +
NADPH + H+ <-> (R)-3-Hydroxypalmitoyl-[acyl-carrier protein]
C16H31O2SR + NADP+
```

```
4.2.1.-g : (R)-3-Hydroxypalmitoyl-[acyl-carrier protein] C16H31O2SR <-
> trans-hexadec-2-enoyl-[acyl-carrier protein] C16H29OSR + H2O
1.3.1.9g : trans-hexadec-2-enoyl-[acyl-carrier protein] C16H29OSR +
NADH + H+ <-> hexadecanoyl-[acyl-carrier protein] C16H31OSR + NAD+
2.3.1.179i : hexadecanoyl-[acyl-carrier protein] C16H31OSR + a
malonyl-ACP -> 3-oxooctadecanoyl-[acp] C18H33O2SR + CO2 + an [acyl-
carrier protein]
1.1.1.100h : 3-oxooctadecanoyl-[acp] C18H33O2SR + NADPH + H+ <-> 3-
Hydroxyoctadecanoyl-[acp] C18H35O2SR + NADP+
4.2.1.-h : 3-Hydroxyoctadecanoyl-[acp] C18H35O2SR <-> trans-octadec-2-
enoyl-[acp] C18H33OSR + H2O
1.3.1.9h : trans-octadec-2-enoyl-[acp] C18H33OSR + NADH + H+ <->
octadecanoyl-[acyl-carrier protein] C18H35OSR + NAD+
_1.14.19.1a : hexadecanoyl-[acyl-carrier protein] C16H31OSR + O2 + 2
H+ <-> (9Z)-hexadecanoyl-[acyl-carrier protein] C16H31OSR + H2O
_1.14.19.1b : octadecanoyl-[acyl-carrier protein] C18H35OSR + O2 + 2
H+ <-> (9Z)-octadecanoyl-[acyl-carrier protein] C18H33OSR + H2O
# not pres ! would end ACP release and fatty acid formation
~~IMPORTANT, following 13 reactions not in sequence!!~~
·3.1.2.14a : octanoyl-[acyl-carrier protein] C8H15OSR + H2O ->
octanoic acid + an [acyl-carrier protein]
·3.1.2.14b : decanoyl-[acyl-carrier protein] C10H19OSR + H2O ->
decanoic acid + an [acyl-carrier protein]
·3.1.2.14c : dodecanoyl-[acyl-carrier protein] C12H23OSR + H2O ->
dodecanoic acid + an [acyl-carrier protein]
·3.1.2.14d : tetradecanoyl-[acyl-carrier protein] C14H27OSR + H2O ->
tetradecanoic acid + an [acyl-carrier protein]
·3.1.2.14e : hexadecanoyl-[acyl-carrier protein] C16H31OSR + H2O ->
hexadecanoic acid + an [acyl-carrier protein]
·3.1.2.14f : (9Z)-hexadecanoyl-[acyl-carrier protein] C16H31OSR + H2O
-> (9Z)-hexadecanoic acid + an [acyl-carrier protein]
·3.1.2.14g : octadecanoyl-[acyl-carrier protein] C18H35OSR + H2O ->
octadecanoic acid + an [acyl-carrier protein]
·3.1.2.14h : (9Z)-octadecanoyl-[acyl-carrier protein] C18H33OSR + H2O
-> (9Z)-octadecanoic acid + an [acyl-carrier protein]
2.3.1.179a : acetyl-CoA + acyl-carrier protein HSR -> an acetyl-ACP +
CO2
2.3.1.179b : an acetyl-ACP + a malonyl-ACP -> acetoacetyl-[acp]
C4H5O2SR + CO2 + an [acyl-carrier protein]
2.3.1.179c : butyryl-[acyl-carrier protein] C4H7OSR + a malonyl-ACP ->
3-oxohexanoyl-[acyl-carrier protein] C6H9O2SR + CO2 + an [acyl-carrier
protein]
2.3.1.179d : hexanoyl-[acyl-carrier protein] C6H11OSR + a malonyl-ACP
-> 3-oxooctanoyl-[acyl-carrier protein] C8H13O2SR + CO2 + an [acyl-
carrier protein]
2.3.1.179e : octanoyl-[acyl-carrier protein] C8H15OSR + a malonyl-ACP
-> 3-oxodecanoyl-[acyl-carrier protein] C10H17O2SR + CO2 + an [acyl-
carrier protein]
2.3.1.179f : decanoyl-[acyl-carrier protein] C10H19OSR + a malonyl-ACP
-> 3-oxododecanoyl-[acyl-carrier protein] C12H21O2SR + CO2 + an [acyl-
carrier protein]
2.3.1.179g : dodecanoyl-[acyl-carrier protein] C12H23OSR + a malonyl-
ACP -> 3-oxotetradecanoyl-[acyl-carrier protein] C14H25O2SR + CO2 + an
[acyl-carrier protein]
2.3.1.179h : tetradecanoyl-[acyl-carrier protein] C14H27OSR + a
malonyl-ACP -> 3-oxohexadecanoyl-[acyl-carrier protein] C16H29O2SR +
CO2 + an [acyl-carrier protein]
2.3.1.180a : acetyl-CoA + acyl-carrier protein HSR -> an acetyl-ACP +
CO2
```

```
2.3.1.180b : an acetyl-ACP + a malonyl-ACP -> acetoacetyl-[acp]
C4H5O2SR + CO2 + an [acyl-carrier protein]
2.3.1.180c : butyryl-[acyl-carrier protein] C4H7OSR + a malonyl-ACP ->
3-oxohexanoyl-[acyl-carrier protein] C6H9O2SR + CO2 + an [acyl-carrier
protein]
2.3.1.180d : hexanoyl-[acyl-carrier protein] C6H11OSR + a malonyl-ACP
-> 3-oxooctanoyl-[acyl-carrier protein] C8H13O2SR + CO2 + an [acyl-
carrier protein]
2.3.1.180e : octanoyl-[acyl-carrier protein] C8H15OSR + a malonyl-ACP
-> 3-oxodecanoyl-[acyl-carrier protein] C10H17O2SR + CO2 + an [acyl-
carrier protein]
2.3.1.180f : decanoyl-[acyl-carrier protein] C10H19OSR + a malonyl-ACP
-> 3-oxododecanoyl-[acyl-carrier protein] C12H21O2SR + CO2 + an [acyl-
carrier protein]
2.3.1.180g : dodecanoyl-[acyl-carrier protein] C12H23OSR + a malonyl-
ACP -> 3-oxotetradecanoyl-[acyl-carrier protein] C14H25O2SR + CO2 + an
[acyl-carrier protein]
2.3.1.180h : tetradecanoyl-[acyl-carrier protein] C14H27OSR + a
malonyl-ACP -> 3-oxohexadecanoyl-[acyl-carrier protein] C16H29O2SR +
CO2 + an [acyl-carrier protein]
6.2.1.3 : coenzyme A + a fatty acid + ATP -> acyl-CoA + diphosphate +
AMP

# Porphyrin and chlorophyll metabolism
# - from glutamate-1-semialdehyde to uroporphyrinogen-III
tRNA 1.2.1.70 : glutamate-1-semialdehyde + NADP+ + tRNAGlu <-> L-
glutamyl-tRNAGlu + NADPH + H+
5.4.3.8 : glutamate-1-semialdehyde -> 5-amino-levulinate
4.2.1.24 : 2 5-amino-levulinate -> 2 H2O + porphobilinogen
2.5.1.61 : H2O + 4 porphobilinogen -> 4 ammonia H3N +
hydroxymethylbilane
4.2.1.75 : hydroxymethylbilane -> uroporphyrinogen-III + H2O
_uroporphyrinogen-I_spont : hydroxymethylbilane -> uroporphyrinogen-I
+ H2O
_uroporphyrin-I_spont : uroporphyrinogen-I -> uroporphyrin-I + 6 H+
4.1.1.37b : uroporphyrinogen-III -> 4 CO2 + coproporphyrinogen I
_coproporphyrin-I_spont : coproporphyrinogen I -> coproporphyrin I + 6
H+
_uroporphyrin-III_spont : uroporphyrinogen-III -> uroporphyrin-III + 6
H+
# - from uroporphyrinogen-III to protoporphyrinogen-IX
4.1.1.37a : uroporphyrinogen-III -> 4 CO2 + coproporphyrinogen III
_1.3.3.3 : coproporphyrinogen III + O2 + 2 H+ -> protoporphyrinogen-IX
+ 2 CO2 + 2 H2O
_1.3.99.22 : coproporphyrinogen III + 2 S-adenosyl-L-methionine ->
protoporphyrinogen-IX + 2 CO2 + 2 L-methionine + 2 5'-deoxyadenosine
# not pres ! (would connect protoporphyrinogen-IX to protoporphyrin IX
production) ~~IMPORTANT, following reaction not in sequence!!~~
·1.3.3.4 : protoporphyrinogen-IX + O2 -> protoporphyrin IX + 3 H2O
# - Heme production
4.99.1.1 : Fe2+ + protoporphyrin IX -> 2 H+ + protoheme IX
2.5.1.-c : protoheme IX + (E,E)-farnesyl diphosphate + H2O -> heme o +
diphosphate
_COX15 : heme o -> heme a
_1.14.99.3 : 3 O2 + 3 thioredoxin + protoheme IX -> biliverdin + 3 H2O
+ 3 thioredoxin disulfide + CO + Fe2+
1.3.7.5 : biliverdin + 2 Fd* -> (3Z)-phycocyanobilin + 2 Fd
# - Chlorophyll pathway
6.6.1.1 : ATP + protoporphyrin IX + Mg2+ + H2O -> Mg-protoporphyrin +
phosphate O4P + ADP + 2 H+
```

```
2.1.1.11 : Mg-protoporphyrin + S-adenosyl-L-methionine -> S-adenosyl-
L-homocysteine + Mg-protoporphyrin monomethyl ester
Zn-Bacterio-chlorophyll-a : Mg-protoporphyrin monomethyl ester + Zn2+
-> Zn-Bacterio-chlorophyll a + Mg2+
1.14.13.81a : Mg-protoporphyrin monomethyl ester + NADPH + H+ + O2 ->
131-hydroxy-magnesium-protoporphyrin IX 13-monomethyl ester + NADP+ +
H2O
1.14.13.81b : 131-hydroxy-magnesium-protoporphyrin IX 13-monomethyl
ester + NADPH + H+ + O2 -> 131-oxo-magnesium-protoporphyrin IX 13-
monomethyl ester + NADP+ + 2 H2O
1.14.13.81c : 131-oxo-magnesium-protoporphyrin IX 13-monomethyl ester
+ NADPH + H+ + O2 -> divinyl protochlorophyllide a + NADP+ + 2 H2O
1.3.1.33a : divinyl protochlorophyllide a + NADPH + H+ -> NADP+ +
divinylchlorophyllide a
# not pres ! (would connect divinylchlorophyllide a to chlorophyllide
a production) ~~IMPORTANT, following reaction not in sequence!!~~
·1.3.1.75a : divinylchlorophyllide a + NADPH + H+ <-> chlorophyllide a
+ NADP+
# not pres ! (would connect divinyl protochlorophyllide a to monovinyl
protochlorophyllide a production) ~~IMPORTANT, following reaction not
in sequence!!~~
·1.3.1.75b : divinyl protochlorophyllide a + NADPH + H+ <-> monovinyl
protochlorophyllide a + NADP+
1.3.7.7 : monovinyl protochlorophyllide a + Fd + 2 ADP + 2 phosphate
O4P -> chlorophyllide a + Fd* + 2 ATP
_BacChl : monovinyl protochlorophyllide a + phytyl diphosphate ->
bacterio-chlorophylls
_BacPheo : phytyl diphosphate + chlorophyllide a <-> bacterio-
pheophytins + Mg2+ + diphosphate
1.3.1.33b : monovinyl protochlorophyllide a + NADPH + H+ -> NADP+ +
chlorophyllide a
_dChla : divinyl protochlorophyllide a -> divinylchlorophyll a
_Chla : divinylchlorophyll a -> chlorophyll a
2.5.1.62a : chlorophyllide a + phytyl diphosphate -> chlorophyll a +
diphosphate
_Chlb : bacterio-chlorophyllide a + H2O -> bacterio-chlorophyllide b +
4 H+
_Pheo : chlorophyll a + 2 H+ -> pheophytins + Mg2+
# - Cob(II)yrinate a,c diamide pathway
2.1.1.107a : S-adenosyl-L-methionine + uroporphyrinogen-III -> S-
adenosyl-L-homocysteine + precorrin-1
2.1.1.107b : S-adenosyl-L-methionine + precorrin-1 -> S-adenosyl-L-
homocysteine + precorrin-2
_2.1.1.130 : S-adenosyl-L-methionine + precorrin-2 -> S-adenosyl-L-
homocysteine + precorrin-3A
# - Cobalt pathway
# not pres ! (would connect cobalt pathway to Cob(II)yrinate a,c
diamide production) ~~IMPORTANT, following reaction not in
sequence!!~~
·4.99.1.3 : precorrin-2 + Co2+ <-> cobalt-precorrin-2 + 2 H+
2.1.1.151 : S-adenosyl-L-methionine + cobalt-precorrin-2 <-> S-
adenosyl-L-homocysteine + cobalt-precorrin-3
2.1.1.131a : cobalt-precorrin-3 + S-adenosyl-L-methionine <-> cobalt-
precorrin-4 + S-adenosyl-L-homocysteine
2.1.1.133a : S-adenosyl-L-methionine + cobalt-precorrin-4 <-> S-
adenosyl-L-homocysteine + cobalt-precorrin-5A
3.7.1.12 : cobalt-precorrin-5A + H2O <-> cobalt-precorrin-5B +
acetaldehyde
2.1.1.195 : cobalt-precorrin-5B + S-adenosyl-L-methionine -> cobalt-
precorrin-6A + S-adenosyl-L-homocysteine
```

```
_1.3.1.54a : cobalt-precorrin-6A + NADPH + H+ <-> cobalt-precorrin-6B
+ NADP+
# not pres ! (would complete this pathway)~~IMPORTANT, following 2
reaction not in sequence!!~~
·RXN-8766 : cobalt-precorrin-6B + S-adenosyl-L-methionine -> cobalt-
precorrin-7 + S-adenosyl-L-homocysteine
·RXN-8767 : cobalt-precorrin-7 + S-adenosyl-L-methionine -> cobalt-
precorrin-8x + S-adenosyl-L-homocysteine + CO2
5.4.1.2a : cobalt-precorrin-8x <-> cobyrinate
6.3.5.11 : cobyrinate + 2 L-glutamine + 2 ATP + 2 H2O <->
cob(II)yrinate a,c-diamide + 2 L-glutamate + 2 ADP + 2 phosphate O4P
# - Non-Cobalt pathway
2.1.1.131b : precorrin-3B + S-adenosyl-L-methionine -> precorrin-4 +
S-adenosyl-L-homocysteine
2.1.1.133b : S-adenosyl-L-methionine + precorrin-4 -> S-adenosyl-L-
homocysteine + precorrin-5
# not pres ! (would complete this pathway)~~IMPORTANT, following
reaction not in sequence!!~~
·2.1.1.152 : precorrin-5 + S-adenosyl-L-methionine + H2O <->
precorrin-6A + S-adenosyl-L-homocysteine + acetate
1.3.1.54b : precorrin-6A + NADPH + H+ -> precorrin-6B + NADP+
2.1.1.132 : precorrin-6B + 2 S-adenosyl-L-methionine -> precorrin-8x +
2 S-adenosyl-L-homocysteine + CO2
5.4.1.2b : precorrin-8x -> hydrogenobyrinate
6.3.5.9 : hydrogenobyrinate + 2 L-glutamine + 2 ATP + 2 H2O -> 2 L-
glutamate + hydrogenobyrinate a,c-diamide + 2 ADP + 2 phosphate O4P
6.6.1.2 : hydrogenobyrinate a,c-diamide + Co2+ + ATP + H2O ->
cob(II)yrinate a,c-diamide + ADP + phosphate O4P + 2 H+
# - coenzyme B12 pathway
# not pres ! (would connect cob(II)yrinate a,c-diamide to
cob(I)yrinate a,c-diamide production) ~~IMPORTANT, following reaction
not in sequence!!~~
·1.16.8.1 : 2 cob(II)yrinate a,c-diamide + FMN -> 2 cob(I)yrinate a,c-
diamide + FMNH2
2.5.1.17a : cob(I)yrinate a,c-diamide + ATP -> adenosyl-cobyrinate
a,c-diamide + PPPi
6.3.5.10 : adenosyl-cobyrinate a,c-diamide + 4 L-glutamine + 4 ATP + 4
H2O -> 4 L-glutamate + adenosyl-cobyrate + 4 ADP + 4 phosphate O4P
6.3.1.10a : ATP + adenosyl-cobyrate + (R)-1-amino-2-propanol O-2-
phosphate -> ADP + phosphate O4P + adenosylcobinamide
# not pres ! (would connect L-threonine O-3-phosphate to
adenosylcobinamide) ~~IMPORTANT, following 2 reaction not in
sequence!!~~
·RXN-8626 : L-threonine + ATP <-> L-threonine O-3-phosphate + ADP
·4.1.1.81 : L-threonine O-3-phosphate -> (R)-1-amino-2-propanol O-2-
phosphate + CO2
6.3.1.10b : ATP + adenosyl-cobyrate + (R)-1-Aminopropan-2-ol -> ADP +
phosphate O4P + adenosylcobinamide
2.5.1.17b : cobinamide + ATP -> adenosylcobinamide + PPPi
2.7.1.156 : adenosylcobinamide + ATP -> adenosyl-cobinamide phosphate
+ ADP
2.7.7.62 : adenosyl-cobinamide phosphate + GTP -> adenosylcobinamide-
GDP + diphosphate
2.7.8.26 : adenosylcobinamide-GDP + alpha-ribazole -> coenzyme B12 +
GMP

# Riboflavin metabolism
3.5.4.25 : 3 H2O + GTP -> diphosphate + 2,5-diamino-6-(ribosylamino)-
4-(3H)-pyrimidinone 5'-phosphate + formate
3.5.4.26 : H2O + 2,5-diamino-6-(ribosylamino)-4-(3H)-pyrimidinone 5'-
phosphate -> 5-amino-6-(5'-phosphoribosylamino)uracil + ammonia H3N
```

```
1.1.1.193 : 5-amino-6-(5'-phosphoribosylamino)uracil + NADPH + H+ ->
5-amino-6-(5'-phosphoribitylamino)uracil + NADP+
2.5.1.78 : 5-amino-6-ribitylamino-2,4(1H,3H)-pyrimidinedione + 3,4-
dihydroxy-2-butanone-4-P -> 6,7-dimethyl-8-(1-D-ribityl)lumazine +
phosphate O4P + 2 H2O
DIOHBUTANONEPSYN-RXN : D-ribulose-5-phosphate -> 3,4-dihydroxy-2-
butanone-4-P + formate
2.5.1.9 : 2 6,7-dimethyl-8-(1-D-ribityl)lumazine -> 5-amino-6-
ribitylamino-2,4(1H,3H)-pyrimidinedione + riboflavin
_lumazine-spont : 6,7-dimethyl-8-(1-D-ribityl)lumazine + UQ -> 7-
Hydroxy-6-methyl-8-ribityl lumazine + UQH2
2.7.1.26 : riboflavin + ATP -> FMN + ADP
2.7.7.2 : FMN + ATP -> FAD + diphosphate
_benzimidazole-spont : 2 riboflavin + 2 CO2 -> 4 dimethylbenzimidazole
+ 3 O2
# not pres ! would end producing alpha-ribazole ~~IMPORTANT, following
4 reactions not in sequence!!~~
·2.4.2.21 : nicotinate nucleotide + dimethylbenzimidazole <->
nicotinate + alpha-ribazole-5'-P
·3.1.3.73 : H2O + alpha-ribazole-5'-P <-> alpha-ribazole + phosphate
O4P
·2.4.2.1 : nicotinate + ribose-1-phosphate <-> nicotinate D-
ribonucleoside + phosphate O4P
·2.7.1.1 : nicotinate D-ribonucleoside + ATP -> nicotinate nucleotide
+ ADP

# REDOX compounds
_1.8.1.9 : thioredoxin disulfide + NADPH + H+ <-> thioredoxin + NADP+
# not pres ! would produce thioredoxin~~IMPORTANT, following reaction
not in sequence!!~~
·_Thiored : thioredoxin disulfide + 2 Fd* + 2 H+ <-> 2 Fd +
thioredoxin
1.11.1.6 : 2 H2O2 -> 2 H2O + O2
_1.15.1.1 : 2 O2.- + 2 H+ <-> O2 + H2O2

# Pyruvate metabolism
1.1.1.38 : NAD+ + malate -> NADH + H+ + CO2 + pyruvate
4.1.1.31 : phosphoenolpyruvate + H2O + CO2 -> phosphate O4P +
oxaloacetate
1.2.4.1 : pyruvate + lipoate acetyltransferase N6-(lipoyl)lysine <->
lipoate acetyltransferase N6-(S-acetyldihydrolipoyl)lysine + CO2
2.3.1.12 : acetyl-CoA + lipoate acetyltransferase N6-
(dihydrolipoyl)lysine <-> coenzyme A + lipoate acetyltransferase N6-
(S-acetyldihydrolipoyl)lysine
1.8.1.4a : lipoate acetyltransferase N6-(dihydrolipoyl)lysine + NAD+ -
> lipoate acetyltransferase N6-(lipoyl)lysine + NADH + H+
# not pres ! (would produce methylglyoxal) ~~IMPORTANT, following
reaction not in sequence!!~~
·4.2.3.3 : dihydroxy-acetone phosphate -> methylglyoxal + phosphate
O4P
4.4.1.5 : S-lactoyl-glutathione <-> methylglyoxal + glutathione
3.1.2.6 : S-lactoyl-glutathione + H2O -> glutathione + D-lactate
1.1.1.28 : NAD+ + D-lactate -> NADH + H+ + pyruvate

# Glutathione metabolism
3.4.11.2a : cysteinylglycine + H2O -> L-cysteine + glycine
3.4.11.2b : R-S-cysteinylglycine + H2O -> R-S-cysteine + glycine
6.3.2.3 : glycine + L-gamma-glutamylcysteine + ATP -> glutathione +
phosphate O4P + ADP
_2.3.2.2a : glutathione -> L-glutamate + cysteinylglycine
```

```
_2.3.2.2b : L-alanine + glutathione -> (5-L-glutamyl)-L-alanine +
cysteinylglycine
_2.3.2.2c : L-aspartate + glutathione -> (5-L-glutamyl)-L-aspartate +
cysteinylglycine
_2.3.2.2d : L-arginine + glutathione -> (5-L-glutamyl)-L-arginine +
cysteinylglycine
_2.3.2.2e : L-asparagine + glutathione -> (5-L-glutamyl)-L-asparagine
+ cysteinylglycine
_2.3.2.2f : L-lysine + glutathione -> (5-L-glutamyl)-L-lysine +
cysteinylglycine
_2.3.2.2g : L-threonine + glutathione -> (5-L-glutamyl)-L-threonine +
cysteinylglycine
_2.3.2.2h : L-serine + glutathione -> (5-L-glutamyl)-L-serine +
cysteinylglycine
_2.3.2.2i : L-cysteine + glutathione -> (5-L-glutamyl)-L-cysteine +
cysteinylglycine
_2.3.2.2j : L-glutamate + glutathione -> (5-L-glutamyl)-L-glutamate +
cysteinylglycine
_2.3.2.2k : L-glutamine + glutathione -> (5-L-glutamyl)-L-glutamine +
cysteinylglycine
_2.3.2.2l : L-proline + glutathione -> (5-L-glutamyl)-L-proline +
cysteinylglycine
_2.3.2.2m : L-tryptophan + glutathione -> (5-L-glutamyl)-L-tryptophan
cysteinylglycine
_2.3.2.2n : L-phenylalanine + glutathione -> (5-L-glutamyl)-L-
phenylalanine + cysteinylglycine
_2.3.2.2o : L-tyrosine + glutathione -> (5-L-glutamyl)-L-tyrosine +
cysteinylglycine
_2.3.2.2p : L-histidine + glutathione -> (5-L-glutamyl)-L-histidine +
cysteinylglycine
_2.3.2.2q : L-valine + glutathione -> (5-L-glutamyl)-L-valine +
cysteinylglycine
_2.3.2.2r : L-isoleucine + glutathione -> (5-L-glutamyl)-L-isoleucine
+ cysteinylglycine
_2.3.2.2s : L-leucine + glutathione -> (5-L-glutamyl)-L-leucine +
cysteinylglycine
_2.3.2.2t : L-methionine + glutathione -> (5-L-glutamyl)-L-methionine
+ cysteinylglycine
3.5.2.9 : 5-oxoproline + 2 H2O + ATP -> L-glutamate + phosphate O4P +
ADP
_2.5.1.18 : glutathione -> R-S-glutathione
_2.3.2.2u : R-S-glutathione + H2O -> R-S-cysteinylglycine + L-
glutamate
1.11.1.9 : H2O2 + 2 glutathione -> glutathione disulfide + 2 H2O
_1.8.1.7a : glutathione disulfide + NADPH + H+ -> 2 glutathione +
NADP+
_1.8.1.7b : glutathione disulfide + NADH + H+ -> 2 glutathione + NAD+
# not pres ! would close g-glutamyl cycle ~~IMPORTANT, following 20
reactions not in sequence!!~~
·6.3.2.2 : ATP + L-glutamate + L-cysteine -> ADP + diphosphate + L-
gamma-glutamylcysteine
·2.3.2.4a : (5-L-glutamyl)-L-alanine -> L-alanine + 5-oxoproline
·2.3.2.4b : (5-L-glutamyl)-L-aspartate -> L-aspartate + 5-oxoproline
·2.3.2.4c : (5-L-glutamyl)-L-arginine -> L-arginine + 5-oxoproline
·2.3.2.4d : (5-L-glutamyl)-L-asparagine -> L-asparagine + 5-oxoproline
·2.3.2.4e : (5-L-glutamyl)-L-lysine -> L-lysine + 5-oxoproline
·2.3.2.4f : (5-L-glutamyl)-L-threonine -> L-threonine + 5-oxoproline
·2.3.2.4g : (5-L-glutamyl)-L-serine -> L-serine + 5-oxoproline
·2.3.2.4h : (5-L-glutamyl)-L-cysteine -> L-cysteine + 5-oxoproline
·2.3.2.4i : (5-L-glutamyl)-L-glutamate -> L-glutamate + 5-oxoproline
·2.3.2.4j : (5-L-glutamyl)-L-glutamine -> L-glutamine + 5-oxoproline
```

```
·2.3.2.4k : (5-L-glutamyl)-L-proline -> L-proline + 5-oxoproline
·2.3.2.4l : (5-L-glutamyl)-L-tryptophan -> L-tryptophan + 5-oxoproline
·2.3.2.4m : (5-L-glutamyl)-L-phenylalanine -> L-phenylalanine + 5-
oxoproline
·2.3.2.4n : (5-L-glutamyl)-L-tyrosine -> L-tyrosine + 5-oxoproline
·2.3.2.4o : (5-L-glutamyl)-L-histidine -> L-histidine + 5-oxoproline
·2.3.2.4p : (5-L-glutamyl)-L-valine -> L-valine + 5-oxoproline
·2.3.2.4q : (5-L-glutamyl)-L-isoleucine -> L-isoleucine + 5-oxoproline
·2.3.2.4r : (5-L-glutamyl)-L-leucine -> L-leucine + 5-oxoproline
·2.3.2.4s : (5-L-glutamyl)-L-methionine -> L-methionine + 5-oxoproline
·2.3.2.4t : (5-L-glutamyl)-L-glicine -> L-glicine + 5-oxoproline
1.1.1.284a : S-(hydroxymethyl)glutathione + NAD+ -> S-
formylglutathione + NADH + H+
1.1.1.284b : S-(hydroxymethyl)glutathione + NADP+ <-> S-
formylglutathione + NADPH + H+
3.1.2.12 : S-formylglutathione + H2O -> formate + glutathione

# Fructose and mannose metabolism
4.1.2.13c : fructose-1-phosphate <-> dihydroxy-acetone phosphate +
glyceraldehyde
2.7.1.4 : beta-D-fructose + ATP <-> beta-D-fructose-6-phosphate + ADP
5.3.1.8 : D-mannose-6-phosphate <-> beta-D-fructose-6-phosphate
5.4.2.8 : alpha-D-mannose 1-phosphate <-> D-mannose-6-phosphate
2.7.7.13 : alpha-D-mannose 1-phosphate + GTP -> GDP-D-mannose +
diphosphate
4.2.1.47 : GDP-D-mannose -> H2O + GDP-4-dehydro-6-deoxy-D-mannose
1.1.1.271 : NADP+ + GDP-L-fucose <-> NADPH + H+ + GDP-4-dehydro-6-
deoxy-D-mannose
2.4.1.83 : dolichyl-phosphate + GDP-D-mannose -> dolichyl beta-D-
mannosyl phosphate + GDP
2.7.7.22 : GDP + alpha-D-mannose 1-phosphate -> phosphate O4P + GDP-D-
mannose

# Aminosugars metabolism
2.6.1.16 : beta-D-fructose-6-phosphate + L-glutamine <-> L-glutamate +
D-glucosamine 6-phosphate
3.5.99.6 : D-glucosamine 6-phosphate + H2O -> beta-D-fructose-6-
phosphate + ammonia H3N
3.5.1.25 : N-acetyl-D-glucosamine-6-phosphate + H2O <-> D-glucosamine
6-phosphate + acetate
4.2.1.126 : N-acetylmuramic acid 6-phosphate + H2O <-> N-acetyl-D-
glucosamine-6-phosphate + (R)-lactate
5.4.2.10 : D-glucosamine 6-phosphate -> D-glucosamine 1-phosphate
2.3.1.157 : D-glucosamine 1-phosphate + acetyl-CoA -> N-acetyl-
glucosamine-1-phosphate + coenzyme A
2.7.7.23 : N-acetyl-glucosamine-1-phosphate + UTP <-> UDP-N-acetyl-D-
glucosamine + diphosphate
2.5.1.7 : UDP-N-acetyl-D-glucosamine + phosphoenolpyruvate <-> UDP-
GlcNAc-enolpyruvate + phosphate O4P
_1.1.1.158 : NADP+ + UDP-N-acetylmuramate <-> NADPH + H+ + UDP-GlcNAc-
enolpyruvate
_5.1.3.14a : N-acetyl-D-mannosamine + UTP <-> UDP-N-acetyl-D-
glucosamine + diphosphate
_5.1.3.14b : UDP-N-acetyl-D-mannosamine <-> UDP-N-acetyl-D-glucosamine
5.1.3.9 : N-acetyl-D-glucosamine-6-phosphate <-> N-acetyl-D-
mannosamine-6-phosphate

# Peptidoglycan biosynthesis
6.3.2.8 : L-alanine + UDP-N-acetylmuramate + ATP -> UDP-N-
acetylmuramoyl-L-alanine + phosphate O4P + ADP
```

```
6.3.2.9 : UDP-N-acetylmuramoyl-L-alanine + D-glutamate + ATP -> UDP-N-
acetylmuramoyl-L-alanyl-D-glutamate + phosphate O4P + ADP
6.3.2.13 : UDP-N-acetylmuramoyl-L-alanyl-D-glutamate + meso-
diaminopimelate + ATP -> UDP-N-acetylmuramoyl-L-alanyl-D-glutamyl-
meso-2,6-diaminoheptanedioate + phosphate O4P + ADP
6.3.2.10a : UDP-N-acetylmuramoyl-L-alanyl-D-glutamyl-meso-2,6-
diaminoheptanedioate + D-alanyl-D-alanine + ATP -> UDP-N-
acetylmuramoyl-L-alanyl-D-glutamyl-meso-2,6-diaminoheptanedioate-D-
alanyl-D-alanine + phosphate O4P + ADP
2.7.8.13a : UDP-N-acetylmuramoyl-L-alanyl-D-glutamyl-meso-2,6-
diaminoheptanedioate-D-alanyl-D-alanine + undecaprenyl phosphate ->
undecaprenyl-diphospho-N-cetylmuramoyl-L-alanyl-D-glutamyl-meso-2,6-
diaminopimeloyl-D-alanyl-D-alanine + UMP
2.4.1.227a : undecaprenyl-diphospho-N-acetylmuramoyl-L-alanyl-D-
glutamyl-meso-2,6-diaminopimeloyl-D-alanyl-D-alanine + UDP-N-acetyl-D-
glucosamine -> undecaprenyl-diphospho-N-acetylmuramoyl-(N-
acetylglucosamine)-L-alanyl-D-glutamyl-meso-2,6-diaminopimeloyl-D-
alanyl-D-alanine + UDP
_6.3.1.2b : undecaprenyl-diphospho-N-acetylmuramoyl-(N-
acetylglucosamine)-L-alanyl-D-glutamyl-meso-2,6-diaminopimeloyl-D-
alanyl-D-alanine + ATP + ammonia H3N -> undecaprenyl-diphospho-N-
acetylmuramoyl-(N acetylglucosamine)-L-alanyl-D-glutaminyl-meso-2,6-
diaminopimeloyl-D-alanyl-D-alanine + ADP + phosphate O4P
6.3.2.10b : UDP-N-acetylmuramoyl-L-alanyl-gamma-D-glutamyl-L-lysine +
D-alanyl-D-alanine + ATP -> UDPMurAc(oyl-L-Ala-D-gamma-Glu-L-Lys-D-
Ala-D-Ala) + phosphate O4P + ADP
2.7.8.13b : UDPMurAc(oyl-L-Ala-D-gamma-Glu-L-Lys-D-Ala-D-Ala) +
undecaprenyl phosphate -> MurAc(oyl-L-Ala-D-gamma-Glu-L-Lys-D-Ala-D-
Ala)-diphospho-undecaprenol + UMP
2.4.1.227b : MurAc(oyl-L-Ala-D-gamma-Glu-L-Lys-D-Ala-D-Ala)-diphospho-
undecaprenol + UDP-N-acetyl-D-glucosamine -> undecaprenyl-diphospho-N-
acetylmuramoyl-(N-acetylglucosamine)-L-alanyl-gamma-D-glutamyl-L-
lysyl-D-alanyl-D-alanine + UDP
_6.3.1.2c : undecaprenyl-diphospho-N-acetylmuramoyl-(N-
acetylglucosamine)-L-alanyl-gamma-D-glutamyl-L-lysyl-D-alanyl-D-
alanine + ATP + ammonia H3N -> undecaprenyl-diphospho-N-
acetylmuramoyl-(N-acetylglucosamine)-L-alanyl-D-isoglutaminyl-L-lysyl-
D-alanyl-D-alanine + ADP + phosphate O4P
6.3.2.10c : UDP-N-acetylmuramoyl-L-alanyl-D-glutamyl-L-lysine + D-
alanyl-D-alanine + ATP -> UDP-N-acetylmuramoyl-L-alanyl-D-glutamyl-L-
lysyl-D-alanyl-D-alanine + phosphate O4P + ADP
2.7.8.13c : UDP-N-acetylmuramoyl-L-alanyl-D-glutamyl-L-lysyl-D-alanyl-
D-alanine + undecaprenyl phosphate -> undecaprenyl-diphospho-N-
acetylmuramoyl-L-alanyl-D-glutamyl-L-lysyl-D-alanyl-D-alanine + UMP
2.4.1.227c : undecaprenyl-diphospho-N-acetylmuramoyl-L-alanyl-D-
glutamyl-L-lysyl-D-alanyl-D-alanine + UDP-N-acetyl-D-glucosamine ->
undecaprenyl-diphospho-N-acetylmuramoyl-(N-acetylglucosamine)-L-
alanyl-D-glutamyl-L-lysyl-D-alanyl-D-alanine + UDP
_6.3.1.2d : undecaprenyl-diphospho-N-acetylmuramoyl-(N-
acetylglucosamine)-L-alanyl-D-glutamyl-L-lysyl-D-alanyl-D-alanine +
ATP + ammonia H3N -> undecaprenyl-diphospho-N-acetylmuramoyl-(N-
acetylglucosamine)-L-alanyl-D-glutaminyl-L-lysyl-D-alanyl-D-alanine +
ADP + phosphate O4P
# not pres ! would complete peptidoglycan biosynthesis ~~IMPORTANT,
following 3 reactions not in sequence!!~~
·2.3.2.10a : undecaprenyl-diphospho-N-acetylmuramoyl-(N-
acetylglucosamine)-L-alanyl-D-glutaminyl-meso-2,6-diaminopimeloyl-D-
alanyl-D-alanine + glycyl-tRNAgly <-> undecaprenyl-diphospho-N-
acetylmuramoyl-(N-acetylglucosamine)-L-alanyl-D-glutaminyl-meso-2,6-
diaminopimeloyl-(glycyl)5-D-alanyl-D-alanine + tRNAgly
```

```
·2.3.2.10b : undecaprenyl-diphospho-N-acetylmuramoyl-(N-
acetylglucosamine)-L-alanyl-D-isoglutaminyl-L-lysyl-D-alanyl-D-alanine
+ glycyl-tRNAgly <-> undecaprenyl-diphospho-N-acetylmuramoyl-(N-
acetylglucosamine)-L-alanyl-D isoglutaminyl-L-lysyl-(glycyl)5-D-
alanyl-D-alanine + tRNAgly
·2.3.2.10c : undecaprenyl-diphospho-N-acetylmuramoyl-(N-
acetylglucosamine)-L-alanyl-D-glutaminyl-L-lysyl-D-alanyl-D-alanine +
glycyl-tRNAgly <-> undecaprenyl-diphospho-N-acetylmuramoyl-(N-
acetylglucosamine)-L-alanyl-D-glutaminyl-L-lysyl-(glycyl)5-D-alanyl-D-
alanine + tRNAgly
6.3.2.4 : 2 D-alanine + ATP -> D-alanyl-D-alanine + phosphate O4P +
ADP
3.6.1.27 : undecaprenyl diphosphate + H2O -> undecaprenyl phosphate +
phosphate O4P
_peptidoglycan_syfa : undecaprenyl-diphospho-N-acetylmuramoyl-(N-
acetylglucosamine)-L-alanyl-D-glutaminyl-meso-2,6-diaminopimeloyl-
(glycyl)5-D-alanyl-D-alanine -> peptidoglycan + undecaprenyl
diphosphate
_2.4.1.129 : [GlcNAc-(1-4)-Mur2Ac(oyl-L-Ala-gamma-D-Glu-L-Lys-D-Ala-D-
Ala)]n-diphosphoundecaprenol + GlcNAc-(1-4)-Mur2Ac(oyl-L-Ala-gamma-D-
Glu-L-Lys-D-Ala-D-Ala)-diphosphoundecaprenol -> [GlcNAc-(1-4)-
Mur2Ac(oyl-L-Ala-gamma-D-Glu-L-Lys-D-Ala-D-Ala)]n+1-
diphosphoundecaprenol + undecaprenyl diphosphate
3.4.16.4 : (Ac)2-L-lys-D-alanyl-D-alanine + H2O <-> D-alanine + (Ac)2-
L-lys-D-alanine
3.5.1.28 : N-acetylmuramoyl-Ala + H2O <-> N-acetylmuramate + L-alanine

# additional glycan biosynthesis
2.3.1.129 : (R)-3-hydroxymyristoyl-ACP + UDP-N-acetyl-D-glucosamine ->
UDP-3-O-(3-hydroxymyristoyl)-N-acetylglucosamine + an [acyl-carrier
protein]
3.2.1.24 : [2Man-(1->6)-(1->3)- Man-(1->4)- GlcNAc] + H2O <-> [Man-(1-
>3)- Man-(1->4)- GlcNAc] + D-mannose

# Starch and sucrose metabolism
2.4.1.14 : beta-D-fructose-6-phosphate + UDP-D-glucose -> sucrose-6-
phosphate + UDP
3.2.1.26 : sucrose-6-phosphate + H2O <-> beta-D-fructose + alpha-D-
glucose-6-phosphate
2.4.1.25 : maltotriose + maltose -> maltotetraose + beta-D-glucose
_2.4.1.1 : a glycogen_n1 + phosphate O4P -> a glycogen_n + alpha-D-
glucose 1-phosphate
2.7.7.27 : alpha-D-glucose 1-phosphate + ATP -> ADP-D-glucose +
diphosphate
_a-glucansyf : 2 ADP-D-glucose -> a 1,4-alpha-D-glucan_n
_2.4.1.21a : ADP-D-glucose + a 1,4-alpha-D-glucan_n -> ADP + a 1,4-
alpha-D-glucan_n1
_glycsyf : 2 ADP-D-glucose + a 1,4-alpha-D-glucan_n1 -> a glycogen_n
_2.4.1.21b : ADP-D-glucose + a glycogen_n -> ADP + a glycogen_n1
_2.4.1.18 : a glycogen_n1 + a 1,4-alpha-D-glucan_n <-> a glycogen_n +
a 1,4-alpha-D-glucan_n1
5.4.2.2 : alpha-D-glucose 1-phosphate <-> alpha-D-glucose-6-phosphate
4.2.1.45 : CDP-D-glucose -> H2O + CDP-4-dehydro-6-deoxy-D-glucose
2.4.1.12 : 1,4-beta-D-glucan_n + UDP-D-glucose <-> 1,4-beta-D-
glucan_n1 + UDP
_3.2.1.21a : cellobiose <-> 2 beta-D-glucose
_3.2.1.21b : 1,4-beta-D-glucan_n <-> beta-D-glucose

# Biosynthesis of steroids
# - NON-Mevalonate pathway - produces Isopentenyl-PP
```

```
2.2.1.7 : pyruvate + D-glyceraldehyde-3-phosphate -> 1-deoxy-D-
xylulose 5-phosphate + CO2
1.1.1.267 : 2-C-methyl-D-erythritol-4-phosphate + NADP+ <-> 1-deoxy-D-
xylulose 5-phosphate + NADPH + H+
2.7.7.60 : 2-C-methyl-D-erythritol-4-phosphate + CTP -> 4-(cytidine
5'-diphospho)-2-C-methyl-D-erythritol + diphosphate
2.7.1.148 : 4-(cytidine 5'-diphospho)-2-C-methyl-D-erythritol + ATP ->
2-phospho-4-(cytidine 5'-diphospho)-2-C-methyl-D-erythritol + ADP
4.6.1.12 : 2-phospho-4-(cytidine 5'-diphospho)-2-C-methyl-D-erythritol
-> 2-C-methyl-D-erythritol-2,4-cyclodiphosphate + CMP
1.17.7.1 : 1-hydroxy-2-methyl-2-(E)-butenyl 4-diphosphate + H2O +
thioredoxin disulfide <-> 2-C-methyl-D-erythritol-2,4-cyclodiphosphate
+ thioredoxin
1.17.1.2a : 1-hydroxy-2-methyl-2-(E)-butenyl 4-diphosphate + NADPH +
H+ -> isopentenyl diphosphate + NADP+ + H2O
1.17.1.2b : isopentenyl diphosphate + NADP+ + H2O -> (E)-4-hydroxy-3-
methylbut 2-en-1-yl diphosphate + NADPH + H+
# - Isopentenyl-PP to Squalene
5.3.3.2 : isopentenyl diphosphate -> dimethylallyl diphosphate
2.5.1.29a : isopentenyl diphosphate + (E,E)-farnesyl diphosphate ->
all-trans-geranyl-geranyl diphosphate + diphosphate
2.5.1.1 : dimethylallyl diphosphate + isopentenyl diphosphate ->
geranyl diphosphate + diphosphate
2.5.1.84 : geranyl diphosphate + 7 isopentenyl diphosphate <-> all-
trans-nonaprenyl diphosphate + 7 diphosphate
2.5.1.85 : geranylgeranyl diphosphate + 5 isopentenyl diphosphate ->
all-trans-nonaprenyl diphosphate + 5 diphosphate

# - Vit E production (no pres)
1.3.1.83 : all-trans-geranyl-geranyl diphosphate + 3 H+ + 3 NADPH ->
phytyl diphosphate + 3 NADP+
_VitE-spont : beta-tocopherol -> alpha-tocopherol

# Lipoic acid metabolism
_2.3.1.181a : octanoyl-[acyl-carrier protein] + a protein -> a
protein-N-6-octanoyl-lysine + an [acyl-carrier protein]
_2.8.1.8a : a protein-N-6-octanoyl-lysine + 2 SO + 2 S-adenosyl-L-
methionine -> a protein-N-6-lipoyl-lysine + 2 L-methionine + 2 5'-
deoxyadenosine
_2.8.1.8b : octanoyl-[acyl-carrier protein] + 2 SO + 2 S-adenosyl-L-
methionine -> lipoyl-[acyl carrier protein] + 2 L-methionine + 2 5'-
deoxyadenosine
_2.3.1.181b : lipoyl-[acyl carrier protein] + a protein -> a protein-
N-6-lipoyl-lysine + an [acyl-carrier protein]
_2.7.7.63a : ATP + lipoate -> diphosphate + lipoyl-AMP
# not pres ! would allow lipoate input~~IMPORTANT, following 3
reactions not in sequence!!~~
·transp_lipoate : lipoate_extrac -> lipoate
_2.7.7.63b : lipoyl-AMP + a protein -> a protein-N-6-lipoyl-lysine +
AMP

# Carotenoid biosynthesis
_2.5.1.32 : 2 all-trans-geranyl-geranyl diphosphate -> phytoene +
diphosphate
_1.3.5.5a : phytoene + thioredoxin disulfide <-> phytofluene +
thioredoxin
_1.3.5.5b : phytofluene + thioredoxin disulfide  <-> zeta-carotene +
thioredoxin
1.3.5.6a : zeta-carotene + O2 + thioredoxin -> neurosporene + 2 H2O +
thioredoxin disulfide
```

```
1.3.5.6b : O2 + thioredoxin + neurosporene -> 2 H2O + thioredoxin
disulfide + trans-lycopene
CrtLb-a : neurosporene -> beta-zeacarotene
CrtLb-b : beta-zeacarotene -> 7,8-dihydro-beta-carotene
CrtLb-c : trans-lycopene -> gamma-carotene
CruA : trans-lycopene -> gamma-carotene
CrtLb-d : gamma-carotene -> beta-carotene
_CrtRa : beta-carotene + NADH + H+ + O2 -> beta-cryptoxanthin + NAD+ +
H2O
_CrtRb : beta-cryptoxanthin + NADH + H+ + O2 -> zeaxanthin + NAD+ +
H2O
# not pres ! would complete the carotenoid biosynthesis ~~IMPORTANT,
following 2 reactions not in sequence!!~~
·CrtOa : beta-carotene -> echinenone
·CrtOb : echinenone -> canthaxanthin
_CrtRc : echinenone + NADH + H+ + O2 -> 3-hydroxyechinenone + NAD+ +
H2O
_CrtRd : 3-hydroxyechinenone + NADH + H+ + O2 -> adonixanthin + NAD+ +
H2O
_CrtRe : canthaxanthin + NADH + H+ + O2 -> phoenicoxanthin + NAD+ +
H2O
_CrtRf : phoenicoxanthin + NADH + H+ + O2 -> astaxanthin + NAD+ + H2O
_gCar : gamma-carotene <-> (2'S)-deoxymyxol 2'-(2,4-di-O-methyl-alpha-
L-fucoside)
_CrtRg : (2'S)-deoxymyxol 2'-(2,4-di-O-methyl-alpha-L-fucoside) <->
(3R,2'S)-myxol 2'-(2,4-di-O-methyl-alpha-L-fucoside)
5.2.1.13 : 7,9,7',9'-tetracis-lycopene -> trans-lycopene

# Nucleotide sugars metabolism - present? necessary?
# not pres ! would complete the UDP-D-glucose biosynthesis
~~IMPORTANT, following 1 reactions not in sequence!!~~
·2.7.7.9 : UTP + alpha-D-glucose 1-phosphate -> diphosphate + UDP-D-
glucose
5.1.3.2a : UDP-D-glucose <-> UDP-D-galactose
3.13.1.1 : UDP-D-glucose + sulfite O3s -> UDP-6-sulfoquinovose + H2O
1.1.1.22 : UDP-D-glucose + H2O + 2 NAD+ -> UDP-D-glucuronate + 2 NADH
+ H+
2.7.7.24 : alpha-D-glucose 1-phosphate + dTTP -> dTDP-D-glucose +
diphosphate
5.1.3.2b : dTDP-D-glucose <-> dTDP-D-galactose
2.7.7.33 : alpha-D-glucose 1-phosphate + CTP -> CDP-D-glucose +
diphosphate
4.2.1.46 : dTDP-D-glucose -> H2O +dTDP-4-dehydro-6-deoxy-D-glucose
5.1.3.13a : dTDP-4-dehydro-6-deoxy-D-glucose -> dTDP-4-dehydro-6-
deoxy-L-mannose
5.1.3.13b : dTDP-4-dehydro-6-deoxy-D-glucose + GTP -> GDP-4-dehydro-6-
deoxy-L-mannose + dTDP + phosphate O4P
1.1.1.133a : NADP+ + dTDP-alpha-L-rhamnose <-> NADPH + H+ + dTDP-4-
dehydro-6-deoxy-L-mannose
1.1.1.133b : NADP+ + GDP-6-deoxy-L-mannose <-> NADPH + H+ + GDP-4-
dehydro-6-deoxy-L-mannose
5.4.99.9 : UDP-D-galactose <-> UDP-alpha-D-galacto-1,4-furanose

# Propanoate metabolism - present?
6.2.1.1b : propionyladenylate + CoA <-> AMP + propanoyl-CoA
2.7.2.1b : ATP + propanoate <-> ADP + propanoyl phosphate
6.2.1.1c : ATP + propanoate <-> diphosphate + propionyladenylate
_1.2.1.3e : 2-propyn-1-al + NAD+ + H2O -> NADH + H+ + propynoate
4.2.1.99 : (2S,3R)-3-hydroxybutane-1,2,3-tricarboxylate <-> (Z)-but-2-
ene-1,2,3-tricarboxylate + H2O
```

```
# Thiamine metabolism - present?
PYRIMSYF1-RXN : 5-aminoimidazole ribonucleotide <->
hydroxymethylpyrimidine
# not pres ! would complete the thiamine-phosphate biosynthesis
~~IMPORTANT, following reaction not in sequence!!~~
·2.7.1.49 : ATP + hydroxymethylpyrimidine <-> ADP +
hydroxymethylpyrimidine phosphate
2.7.4.7 : hydroxymethylpyrimidine phosphate + ATP <-> 4-amino-5-
hydroxymethyl-2-methylpyrimidine-pyrophosphate + ADP
1.4.3.19 : glycine <-> iminoglycine
ThiG : iminoglycine + L-tyrosine + [ThiS]-COSS-[ThiF] -> 4-methyl-5-
(beta-hydroxyethyl)thiazole phosphate
2.5.1.3 : 4-methyl-5-(beta-hydroxyethyl)thiazole phosphate + 4-amino-
5-hydroxymethyl-2-methylpyrimidine-pyrophosphate -> thiamine-phosphate
+ diphosphate
2.7.4.16 : thiamine-phosphate + ATP -> thiamine diphosphate + ADP
# not pres ! would complete the thiamine biosynthesis ~~IMPORTANT,
following reaction not in sequence!!~~
·2.7.1.89 : thiamine + ATP <-> thiamine-phosphate + ADP
3.5.99.2 : thiamine + H2O <-> hydroxymethylpyrimidine + 5-(2-
Hydroxyethyl)-4-methylthiazole + H+

# Biotin metabolism - present?
2.3.1.47 : pimeloyl-CoA + L-alanine -> CO2 + coenzyme A + 7-keto-8-
aminopelargonate
2.6.1.62 : S-adenosyl-L-methionine + 7-keto-8-aminopelargonate <-> S-
adenosyl-4-methylthio-2-oxobutanoate + 7,8-diaminopelargonate
6.3.3.3 : CO2 + 7,8-diaminopelargonate + ATP -> dethiobiotin +
phosphate O4P + ADP
2.8.1.6 : dethiobiotin + 2 S-adenosyl-L-methionine + hydrogen sulfide
-> biotin + 2 5'-deoxyadenosine + 2 L-methionine
6.3.4.15 : BCCP + biotin + ATP -> diphosphate + AMP + biotin-BCCP
(monomer)
RXN-7101 : 2 biotin-BCCP (monomer) <-> a biotin-BCCP (dimer)
_3.4.1.- : N6-D-biotinyl-L-lysine + BCCP <-> biotin-BCCP (monomer)

# Penicillin and cephalosporin biosynthesis - mostly not present
_3.5.2.6 : H2O + penicillin -> penicilloic acid

# Terpenoid biosynthesis - mostly not present
_2.5.1.31a : (E,E)-farnesyl diphosphate + isopentenyl diphosphate <->
all-trans-geranyl-geranyl diphosphate + diphosphate
_2.5.1.31b : all-trans-geranyl-geranyl diphosphate + 7 isopentenyl
diphosphate <-> di-trans,poly-cis-Undecaprenyl diphosphate + 7
diphosphate
2.5.1.10 : geranyl diphosphate + isopentenyl diphosphate ->
diphosphate + (E,E)-farnesyl diphosphate

# Vitamin B6 metabolism - present?
2.6.99.2 : 1-deoxy-D-xylulose 5-phosphate + 1-amino-propan-2-one-3-
phosphate -> pyridoxine-5'-phosphate + phosphate O4P + 2 H2O
_1.1.1.262 : O-phospho-4-hydroxy-L-threonine + NAD+ -> NADH + H+ +
(2S)-2-amino-3-oxo-4-phosphonooxybutanoate
RXN-8447 : (2S)-2-amino-3-oxo-4-phosphonooxybutanoate + H+ <-> 1-
amino-propan-2-one-3-phosphate + CO2
4.2.3.1b : O-phospho-4-hydroxy-L-threonine + H2O -> phosphate O4P + 4-
hydroxy-L threonine
_VitB6 : 4-hydroxy-L threonine -> pyridoxine
```

```
# Glycerophospholipid metabolism & Inositol phosphate metabolism &
Glycerolipid metabolism - mostly not present
1.1.1.94 : sn-glycerol 3-phosphate + NADP+ -> dihydroxy-acetone
phosphate + NADPH + H+
_2.3.1.15 : sn-glycerol 3-phosphate + acyl-CoA -> 1-acyl-sn-glycerol
3-phosphate + coenzyme A
2.3.1.51 : a fatty acyl CoA + a 2-lysophosphatidate -> an L-
phosphatidate + coenzyme A
2.7.7.41 : CTP + an L-phosphatidate -> a CDP-diacylglycerol +
diphosphate
2.7.1.107 : ATP + a 1,2-diacylglycerol -> an L-phosphatidate + ADP
2.7.8.5 : a CDP-diacylglycerol + sn-glycerol 3-phosphate -> an L-1-
phosphatidylglycerol-phosphate + CMP
3.1.3.25 : D-myo-inositol (3)-monophosphate + H2O -> myo-inositol +
phosphate O4P

# rest of reactions with cognate genes
3.1.4.- : 1-(1-Alkenyl)-sn-glycero-3-phosphoethanolamine + H2O <-> 1-
(1-Alkenyl)-sn-glycerol + ethanolamine phosphate
1.1.1.31 : (S)-3-hydroxyisobutyrate + NAD+ <-> (S)-methylmalonate
semialdehyde + NADH + H+
5.3.1.13 : D-arabinose-5-phosphate <->  D-ribulose-5-phosphate
2.5.1.55 : 3-Deoxy-D-manno-octulosonate-8-phosphate + phosphate O4P ->
phosphoenolpyruvate + D-arabinose-5-phosphate + H2O
3.1.3.45 : 3-Deoxy-D-manno-octulosonate-8-phosphate + H2O -> 3-Deoxy-
D-manno-octulosonate + phosphate O4P
2.7.7.38 : CTP + 3-Deoxy-D-manno-octulosonate -> diphosphate + CMP-3-
deoxy-D-manno-octulosonate
3.5.1.108 : UDP-3-O-(3-hydroxytetradecanoyl)-N-acetylglucosamine + H2O
-> UDP-3-O-(3-hydroxytetradecanoyl)-D-glucosamine + acetate
2.4.1.182 : UDP-2,3-bis(3-hydroxytetradecanoyl)glucosamine + 2,3-
bis(3-hydroxytetradecanoyl)-beta-D-glucosaminyl 1-phosphate -> UDP +
2,3-bis(3-hydroxytetradecanoyl)-D-glucosaminyl-(1-6)-beta-D-2,3-
bis(3-hydroxytetradecanoyl)-beta-D-glucosaminyl 1-phosphate
2.7.1.170 : ATP + 1,6-anhydro-N-acetyl-beta-muramate + H2O <-> ADP +
N-acetyl-beta-muramate 6-phosphate
1.1.1.6 : glycerol + NAD+ <-> dihydroxy-acetone phosphate + NADH + H+
4.1.99.5 : octadecanal + O2 + 2 NADPH + 2 H+ <-> heptadecane + formate
+ H2O + 2 NADP+
2.8.1.1 : thiosulfate + cyanide <-> sulfite O3s + thiocyanate
3.8.1.5 : 1,2-dichloroethane + H2O <-> 2-chloroethanol + HCL
4.1.99.12 : D-ribulose-5-phosphate -> L-3,4-dihydroxybutan-2-one-4-
phosphate + formate
3.4.13.22 : D-alanyl-D-alanine + H2O <-> 2 D-alanine
1.14.99.41 : all-trans-8'-apo-beta-carotenal + O2 <-> all-trans-
retinal + (2E,4E,6E)-2,6-dimethylocta-2,4,6-trienedial
1.12.98.1 : coenzyme-F420 + H2 <-> reduced coenzyme-F420
3.1.1.45 : cis-4-carboxymethylenebut-2-en-4-olide + H2O -> 2-
maleylacetate
2.5.1.77 : 3-(4-hydroxyphenyl)-pyruvate + 5-amino-6-ribitylamino-
2,4(1H,3H)-pyrimidinedione + 2 S-adenosyl-L-methionine + H2O -> 7,8-
didemethyl-8-hydroxy-5-deazariboflavin + 2 L-methionine + 2 5'-
deoxyadenosine + oxalate + ammonia H3N
1.7.1.13 : 7-aminomethyl-7-deazaguanine + 2 NADP+ <-> 7-cyano-7-
deazaguanine + 2 NADPH + 2 H+
2.10.1.1 : adenylated molybdopterin + molybdate <-> molybdenum
cofactor + AMP + H2O
_1.16.1.1 : Hg + NADP+ <-> Hg2+ + NADPH + H+
3.5.4.- : cyromazine + H2O -> N-cyclopropylammeline + ammonia H3N
```

```
2.3.1.28 : acetyl-CoA + chloramphenicol <-> coenzyme A +
chloramphenicol-3-acetate
4.2.1.35 : (R)-2-methylmalate -> 2-methylmaleate + H2O
3.1.3.71 : (2R)-O-phospho-3-sulfolactate + H2O -> (2R)-3-sulfolactate
+ phosphate O4P
2.1.1.104 : S-adenosyl-L-methionine + caffeoyl-CoA -> S-adenosyl-L-
homocysteine + feruloyl-CoA

###
# BIOMASS out of amino acids, nucleic acids, lipids and antenna
proteins
Alabm : L-alanine -> Ala
Aspbm : L-aspartate -> Asp
Argbm : L-arginine -> Arg
Asnbm : L-asparagine -> Asn
Lysbm : L-lysine -> Lys
Thrbm : L-threonine -> Thr
Serbm : L-serine -> Ser
Cysbm : L-cysteine -> Cys
Glybm : glycine -> Gly
Glubm : L-glutamate -> Glu
Glnbm : L-glutamine -> Gln
Probm : L-proline -> Pro
Trpbm : L-tryptophan -> Trp
Phebm : L-phenylalanine -> Phe
Tyrbm : L-tyrosine -> Tyr
Hisbm : L-histidine -> His
Valbm : L-valine -> Val
Ilebm : L-isoleucine -> Ile
Leubm : L-leucine -> Leu
Metbm : L-methionine -> Met
_a protein : 897 Ala + 526 Arg + 518 Asp + 374 Asn + 102 Cys + 576 Gln
+ 614 Glu + 702 Gly + 197 His + 628 Ile + 128 Leu + 417 Lys + 194 Met
+ 406 Phe + 512 Pro + 548 Ser + 580 Thr + 149 Trp + 294 Tyr + 638 Val
-> a protein

dATP : dATP -> dATPbm
dGTP : dGTP -> dGTPbm
dTTP : dTTP -> dTTPbm
dUTP : dUTP -> dUTPbm
dCTP : dCTP -> dCTPbm

AMP : AMP -> AMPbm
UMP : UMP -> UMPbm
CMP : CMP -> CMPbm
GMP : GMP -> GMPbm

8C-lipidbm : octanoic acid -> 8C-lipidbm
10C-lipidbm : decanoic acid -> 10C-lipidbm
12C-lipidbm : dodecanoic acid -> 12C-lipidbm
14C-lipidbm : tetradecanoic acid -> 14C-lipidbm
16C-lipidbm : hexadecanoic acid -> 16C-lipidbm
18C-lipidbm : octadecanoic acid -> 18C-lipidbm
(9Z)16C-lipidbm : (9Z)-hexadecanoic acid -> (9Z)16C-lipidbm
(9Z)18C-lipidbm : (9Z)-octadecanoic acid -> (9Z)18C-lipidbm

PhCybibm : (3Z)-phycocyanobilin -> PhCybibm
Chlabm : chlorophyll a -> Chlabm
Carotenebm : trans-lycopene -> Lycopbm
Glycbm : a glycogen_n1 -> Glycbm
Zeaxbm : zeaxanthin -> Zeaxbm
```

```
Beta-carotenebm : beta-carotene -> Betacarotenebm
trans-lycopene -> Lycopbm

ATPbm : ATP -> ATPbm
ADPbm : ADP -> ADPbm

Biomass : 0.000459 a protein + 0.028 14C-lipidbm + 0.0042 16C-lipidbm
+ 0.00448 18C-lipidbm + 0.0066 (9Z)16C-lipidbm + 0.00625 (9Z)18C-
lipidbm + 0.53439 Glycbm + 0.00079 Zeaxbm + 0.000875 Betacarotenebm +
0.00820225 Lycopbm + 0.0057 Chlabm + 0.0285 PhCybibm + 0.0201156
dATPbm + 0.0201156 dTTPbm + 0.02538445 dGTPbm + 0.02538445 dCTPbm +
0.140389293 AMPbm + 0.140389293 UMPbm + 0.123745851 GMPbm +
0.123745851 CMPbm + 59.28 H2O + 59.28 ATP -> BM + 59.28 ADP + 59.28
phosphate O4P + 59.28 H+

_Growth : BM -> BM_out


###

# SPECIAL
# - autotrophic growth
_lightII : photonII_extrac -> photonII
_lightI : photonI_extrac -> photonI
CO2in : CO2_extrac <-> CO2
4.2.1.1b : CO2 + H2O <-> HCO3- + H+
_H2CO3transport : H2CO3_extrac <-> H2CO3
_H2CO3desprot : H2CO3 <-> HCO3- + H+

# maintenance
NADHmain : NADH -> NAD+ + H+
NADPHmain : NADPH -> NADP+ + H+

# - transport
Phosphate TRANS-RXN59G-90 : phosphate O4P_extrac + H+_extrac ->
phosphate O4P + H+
_3.6.3.27 : ATP + H2O + phosphate O4P_extrac -> ADP + 2 phosphate O4P
_Phosphate_out : phosphate O4P -> phosphate O4P_extrac
H2O : H2O_extrac <-> H2O
PROTONS : H+_extrac <-> H+
sulfate TRANS-RXN59G-407 : sulfate O4S_extrac <-> sulfate O4s
nitrate TRANS-RXN59G-237 : nitrate NO3_extrac + ATP + H2O -> nitrate
NO3 + ADP + phosphate O4P
ammonia H3N TRANS-RXN59G-178 : NH4+_extrac + H+_extrac <-> NH4+ + H+
_ammonia H3N : ammonia H3N <-> NH4+
iron TRANS-RXN59G-711 : Fe2+_extrac -> Fe2+
magnesium TRANS-RXN59G-340 : Mg2+_extrac + ATP + H2O -> Mg2+ + ADP +
phosphate O4P
magnesium TRANS-RXN59G-53 : Mg2+_extrac <-> Mg2+
cobalt TRANS-RXN59G-71 : Co2+_extrac + ATP + H2O -> Co2+ + ADP +
phosphate O4P
_COtransp : CO_extrac <-> CO
_H2O2transp : H2O2_extrac <-> H2O2
_O2in : O2_extrac -> O2
_O2out : O2 -> O2_extrac

potassium 3.6.3.12 : K+_extrac + ATP + H2O -> K+ + ADP + phosphate O4P
potassium TRANS-RXN59G-35 : K+_extrac -> K+
chloride TRANS-RXN59G-171 : chloride_extrac -> chloride
```

```
chloride_ATPase : chloride_extrac + ATP + H2O -> chloride + ADP +
phosphate O4P
sodium_symporter : Na+_extrac -> Na+
sodium_antiporter : Na+_extrac + H+ -> Na+ + H+_extrac
calcium TRANS-RXN59G-708 : Ca2+_extrac + H+ -> Ca2+ + H+_extrac
molybdate_transport : molybdate_extrac -> molybdate
_3.6.3.4 : ATP + H2O + Cu2+ -> ADP + phosphate O4P + Cu2+_extrac
_3.6.3.30 : ATP + H2O + Fe3+_extrac -> ADP + phosphate O4P + Fe3+
_3.6.3.25 : ATP + H2O + sulfate O4S_extrac -> ADP + phosphate O4P +
sulfate O4s

_H2 : H2 -> H2_extrac

sulfide TRANSP : hydrogen sulfide_extrac -> hydrogen sulfide
sulfite TRANSP : sulfite O3s_extrac -> sulfite O3s

_an [acyl-carrier protein] : an [acyl-carrier protein]_extrac <-> an
[acyl-carrier protein]
# not pres ! would complete the thiamine biosynthesis ~~IMPORTANT,
following two reactions not in sequence!!~~
·_a protein disulfide R1Protein2R2R3S2 : a protein + 2 hydrogen
sulfide <-> a protein disulfide
·_a protein dithiol H2R1Protein2R2R3S2 : a protein dithiol + NAD+ <->
a protein disulfide + NADH
_3.1.3.16 : a phosphoprotein + H2O <-> a protein + phosphate O4P
_2.7.11.1 : ATP + a protein -> ADP + a phosphoprotein

_RNAout : RNAn1 -> RNAn1_out
_DNAout : DNAn1 -> DNAn1_out
_glycout : a glycogen_n1 -> a glycogen_n1_out
_glycolaldehyde : glycolaldehyde -> glycolaldehyde_extrac

_eth : ethanol -> ethanol_extrac

-CONSTRAINTS
# all units should be mmol/g DCW/h
# carbon input normalized: input flux/number C

# - autotrophic
# -- theoretical values:
_lightI [0, 1.96]
_lightII [0, 1.96]
# -- for a doubling time of 8 hours, a 0.09 specific growth is needed,
# -- values resulting from the inverse optimization of light uptake:
CO2in [0, 1.99]
_H2CO3transport [0, 1.99]
4.2.1.1b [-10, 10]
·_Thiored [0, 100]

###
NADHmain [0.5, 1000]
NADPHmain [0.5, 1000]

# constraints used for light and CO2in minimization
#_Growth [0.0598681, 0.0598681]
#_1.18.1.2 [0.5, 0.5]


# - transport
Phosphate TRANS-RXN59G-90 [0, 1000]
_3.6.3.27 [0, 1000]
```

```
_Phosphate_out [0, 100]
H2O [-500, 500]
sulfate TRANS-RXN59G-407 [-104, 104]
_3.6.3.25 [0, 105]
nitrate TRANS-RXN59G-237 [0, 160]
ammonia H3N TRANS-RXN59G-178 [-1000, 1000]
_COtransp [-10, 10]
_H2O2transp [-100, 100]

# shut down transport for Glu and Gln
transport TRANS-RXN59G-639 [0, 0]
transport TRANS-RXN59G-245 [0, 0]


-EXTERNAL METABOLITES
BM_out

photonII_extrac
photonI_extrac
O2_extrac
H2O2_extrac

H2CO3_extrac
glycolaldehyde_extrac
an [acyl-carrier protein]_extrac
Fe2+_extrac
Mg2+_extrac
Co2+_extrac
CO_extrac
Cu2+_extrac
Fe3+_extrac

tRNAala
tRNAasp
tRNAarg
tRNAasn
tRNAlys
tRNAthr
tRNAser
tRNAcys
tRNAgly
tRNAGlu
tRNApro
tRNAtrp
tRNAphe
tRNAtyr
tRNAhis
tRNAval
tRNAile
tRNAleu
tRNAmet

RNAn1_out
DNAn1_out
a glycogen_n1_out
lipoate_extrac

L-glutamate_extrac
L-glutamine_extrac
nitrate NO3_extrac
```

```
sulfate O4S_extrac
NH4+_extrac

hydrogen sulfide_extrac
sulfite O3s_extrac

phosphate O4P_extrac
CO2_extrac
H2O_extrac
H+_extrac
K+_extrac
chloride_extrac
Na+_extrac
Ca2+_extrac
molybdate_extrac
H2_extrac
ethanol_extrac

-OBJ
_Growth 1 1

-DESIGNOBJ
_Growth 1 1
```

Appendix 1.2

FBA and FVA of *i*Syf715 (Flux values and Minimum and Maximum values, respectively)

Objective = 0.0598681

Reaction name	Flux Value	Minimum	Maximum
2.7.1.2a	0	0	0
2.7.1.2b	0	0	0
5.3.1.9a	-0.191957	-0.331667	1.84266e-016
5.3.1.9b	-0.191957	-50.47	1.84266e-016
2.7.1.11	0	0	151.41
3.1.3.11	1.52122	0	151.41
4.1.2.13a	-1.52122	-50.47	0
5.3.1.1	2.82716	-1.916e-016	75.705
_1.2.1.59a	0	0	1000
_1.2.1.59b	0	0	1000
1.2.1.12	-6.98422	-1000	2.63678e-016
2.7.2.3	-6.98422	-126.175	2.63678e-016
5.4.2.1	0.548951	-75.705	25.235
4.2.1.11	0.548951	-75.705	25.235
2.7.1.40a	0	0	1000
2.7.9.2	0	0	151.41
2.3.3.1	0.0776041	0	60.564
4.2.1.3a	0.0776041	0	60.564
4.2.1.3b	0.0776041	0	60.564
1.1.1.42a	0.0776041	-1.20183e-016	0.0776042
1.1.1.42b	0.0776041	-1.20183e-016	0.0776042
_1.3.99.1	27.0365	-0.705	160.564
4.2.1.2	-0.10381	-60.564	0.317147
·4.1.3.1	0	0	60.564
·2.3.3.9	-0.10381	-60.564	23.2938
4.1.1.39a	0	0	12.6175
_3.1.3.18	0	0	12.6175
_1.1.3.15	0	0	12.6175
_4.1.1.2	0	0	0
1.1.1.49	0	0	50.47
3.1.1.31	0	0	50.47
1.1.1.44	0	0	50.47
5.1.3.1	-2.6352	-50.47	0
5.3.1.6	1.13138	0	25.235
5.3.1.9c	0	0	0
2.2.1.1a	-1.30594	-25.235	0
2.2.1.1b	-1.32927	-25.235	0
_2.2.1.2	0	0	25.235
4.1.2.14	0	0	0
4.1.2.4	0	0	0
1.1.1.47a	-0.169915	-1000	1000
1.1.1.47b	0.169915	-1000	1000
2.7.1.19	3.76658	0	75.705
4.1.2.13b	1.30594	0	25.235
3.1.3.37	1.30594	0	25.235
4.1.2.22	0	0	0
4.1.1.39b	3.76658	0	75.705
5.1.1.1	0	0	0

1.4.1.1	-0.128459	-121.128	0.209474
tRNA 6.1.1.7	0	0	0
2.6.1.44	0.10381	0	1000
1.4.3.16a	0	0	43.26
2.6.1.1a	-0.343353	-43.26	50.47
4.3.2.2a	0.0893555	-50.47	37.8525
6.3.4.4	0.0893555	-50.47	37.8525
2.1.3.3	0.0144542	0	0.0144542
6.3.4.5a	0.0144542	0	0.0144542
4.3.2.1	0.0144542	0	0.0144542
3.5.1.1	-0.0102773	-0.0102773	0
2.7.2.4	0.0499851	0	0.0499851
1.2.1.11	0.0499851	0	0.0499851
_1.1.1.3a	0.0385262	-1000	1000
_1.1.1.3b	0	-1000	1000
2.7.1.39	0.0331952	0	0.0331952
4.2.3.1a	0.0331952	0	0.0331952
4.2.1.52	0.0114589	0	0.0114589
_1.3.1.26a	-0.0114589	-1000	1000
_1.3.1.26b	0	-1000	1000
2.6.1.83	-0.0114589	-0.0114589	0
5.1.1.7	0.0114589	0	0.0114589
4.1.1.20	0.0114589	0	0.0114589
tRNA 6.1.1.12	0	0	0
tRNA 6.1.1.19	0	0	0
4.1.1.19	0	0	0
·3.5.3.12	0	0	0
3.5.1.53	0	0	0
tRNA 6.1.1.22	0	0	0
tRNA 6.3.5.6	0	0	0
tRNA 6.1.1.6	0	0	0
4.1.1.18	0	0	0
THREOSPON-RXN	0	0	0
·4.3.1.19a	0.0172571	0	0.0172571
tRNA 6.1.1.3	0	0	0
2.7.1.31	0	0	0
1.1.1.95	0	0	0
·_4.3.1.19b	-0.495065	-100.947	60.564
2.3.1.30	0.512628	0	100.947
2.5.1.47	0.0028029	0	0.0028029
2.5.1.49a	0.509825	0	100.947
2.1.2.1	-0.0367167	-60.564	0.418947
2.6.1.45	0	-121.128	1000
tRNA 6.1.1.11	0	0	0
2.6.1.51	0	-121.128	1000
CYSPON-RXN	0	0	0
2.6.1.1e	0.509825	0	100.947
2.8.1.2	0.509825	0	100.947
tRNA 6.1.1.16	0	0	0
tRNA 6.1.1.14	0	0	0
3.4.11.1	0	0	0
transport TRANS-RXN59G-639	0	0	0
2.7.2.11	0.0140695	0	0.0140695
1.2.1.41	-0.0140695	-0.0140695	0
SPONTPRO-RXN	0.0140695	0	0.0140695
_1.5.1.2a1	-0.0140695	-1000	1000
_1.5.1.2a2	0	-1000	1000

6.3.1.2	7.89757	0	302.84
5.1.1.3	0	0	0
1.4.7.1	0	0	100.947
·5.1.1.9	0	0	0
4.1.3.27	0.00409444	0	0.00409444
2.4.2.18	-0.00409444	-0.00409444	0
5.3.1.24	0.00409444	0	0.00409444
4.1.1.48	0.00409444	0	0.00409444
4.2.1.20a	0	-1000	0.00409444
_4.2.1.20b	0.00409444	0	1000
4.2.1.20c	0.00409444	0	1000
2.6.1.1b	0	0	0
tRNA 6.1.1.17	0.0163799	0	0.0163799
transport TRANS-RXN59G-245	0	0	0
tRNA _6.3.5.7	0	0	0
·tRNA 6.1.1.18	0	0	0
_1.5.1.2b	0	0	0
_1.5.1.2c	0	0	0
_2.6.1.1c	0	0	0
4.1.3.16	0	0	0
_spont1	0	0	0
tRNA 6.1.1.15	0	0	0
5.2.1.8	0	0	0
tRNA 6.1.1.2	0	0	0
2.5.1.54	0.0233301	0	0.0233301
4.2.3.4	0.0233301	0	0.0233301
4.2.1.10	0.0233301	0	0.0233301
1.1.1.25	-0.0233301	-0.0233301	0
2.7.1.71	0.0233301	0	0.0233301
2.5.1.19	0.0233301	0	0.0233301
4.2.3.5	0.0233301	0	0.0233301
5.4.99.5	0.0192356	0	0.0192356
4.2.1.51	0.0111567	0	0.0111567
2.6.1.9b	0.0111567	0	0.0111567
4.2.1.96	0	0	0
tRNA 6.1.1.20	0	0	0
·1.3.1.12	0.00807896	0	0.00807896
_2.6.1.1d	0.00807896	0	0.00807896
1.3.1.78	0	0	0
tRNA 6.1.1.1	0	0	0
2.4.2.17	-0.00541345	-0.00541345	-4.56036e-017
3.6.1.31	0.00541345	4.56036e-017	0.00541345
3.5.4.19	0.00541345	4.56036e-017	0.00541345
5.3.1.16	0.00541345	4.56036e-017	0.00541345
4.1.3.-	0.00541345	4.56036e-017	0.00541345
4.2.1.19	0.00541345	4.56036e-017	0.00541345
2.6.1.9a	0.00541345	4.56036e-017	0.00541345
·3.1.3.15	0.00541345	4.56036e-017	0.00541345
1.1.1.23a	0.00541345	4.56036e-017	0.00541345
1.1.1.23b	0.00541345	4.56036e-017	0.00541345
tRNA 6.1.1.21	0	0	0
_2.2.1.6d	0.0210493	-3.14411e-017	0.0210493
1.1.1.86a	-0.0210493	-0.0210493	0
4.2.1.9a	0.0210493	0	0.0210493
2.6.1.42a	0.0175319	0	0.0175319
2.6.1.66	0	0	0
tRNA 6.1.1.9	0	0	0

_2.2.1.6e	0.0172571	0	0.0172571
_1.1.1.85b	0	0	0
1.1.1.86b	0.0172571	0	0.0172571
4.2.1.9b	0.0172571	0	0.0172571
2.6.1.42b	-0.0172571	-0.0172571	0
tRNA 6.1.1.5	0	0	0
2.3.3.13	0.00351737	-3.14411e-017	0.00351737
4.2.1.33a	0.00351737	-3.14411e-017	0.00351737
4.2.1.33b	-0.00351737	-0.00351737	3.14411e-017
1.1.1.85a	0.00351737	-3.14411e-017	0.00351737
RXN-7800	0.00351737	-3.14411e-017	0.00351737
2.6.1.42c	0.00351737	-3.14411e-017	0.00351737
tRNA 6.1.1.4	0	0	0
2.3.1.31	0.00533102	0	0.00533102
2.3.1.46	0	0	0
2.5.1.49b	0.00533102	0	0.00533102
3.3.1.1	0.000341248	0	0.000341248
2.1.1.13	0.00567226	0	0.00567226
2.5.1.6	0.000341248	0	0.00443623
2.1.1.37	0	0	0
4.1.1.50	0	0	0
2.5.1.16a	0	0	0
2.4.2.28	0	0	0
5.3.1.23	0	0	0
4.2.1.109	0	0	0
3.1.3.77	0	0	0
1.13.11.54	0	0	0
1.13.11.53	0	0	0
tRNA 2.1.2.9b	0	0	0
tRNA 6.1.1.10	0	0	0
3.5.2.10	0	0	0
4.4.1.16	0	0	0
6.3.5.5	0.1795	0	0.221111
2.1.3.2	0.165046	0	0.221111
3.5.2.3	-0.165046	-0.221111	0
1.3.98.1a	-26.8715	-143.26	0.705
2.4.2.10	-0.165046	-0.221111	0
4.1.1.23	0.165046	0	0.221111
1.3.98.1b	27.0365	-0.705	143.26
3.1.3.5g	0	-151.41	1000
2.7.4.14a	0.156641	-1000	1000
_2.7.4.22	0	-1000	1000
2.4.2.9	0	0	0
3.6.1.19e	0	0	75.705
2.7.4.6e	0.156641	-1000	151.41
3.5.4.13a	0	0	1000
6.3.4.2	0.156641	0	151.41
2.7.4.6h	0	-1000	75.705
3.6.1.19f	0	0	75.705
2.7.4.9b	0	0	75.705
1.17.4.1e	0	0	1000
2.1.1.148	0	0	0
2.7.7.8c	0	0	151.41
2.7.7.6c	0	0	151.41
3.5.4.5a	0	-1000	151.41
3.5.4.1a	0	0	0
2.7.4.6f	-0.151915	-151.41	1000
2.7.4.14b	-0.00268256	-151.41	1000

1.17.4.1d	-0.149233	-1000	151.41
3.1.3.5h	0	-1000	151.41
2.7.4.6g	0.149233	-151.41	1000
2.7.4.14c	0	0	0
3.1.3.5i	0	0	0
3.5.4.5b	0	0	0
3.5.4.13b	0	0	1000
2.7.7.6d	0	0	151.41
2.7.7.8d	0	0	151.41
2.7.7.7c	0.147713	-151.41	151.41
3.1.3.5j	0	0	0
2.7.4.9a	0	0	0
2.7.4.6i	0	0	0
2.7.7.7d	-0.00120428	-0.00120428	0
_3.5.4.1b	0	0	0
3.6.1.13	0	0	0
2.7.6.1	0.174554	0	0.221111
2.4.2.14	0	0	0.209474
6.3.4.13	0	0	0.209474
2.1.2.2a	0	-1000	0.209474
6.3.5.3	-1.66973e-017	0	0.209474
6.3.3.1	-1.66973e-017	0	0.209474
6.3.4.18	-1.66973e-017	0	0.209474
5.4.99.18	-1.66973e-017	0	0.209474
6.3.2.6	-1.66973e-017	0	0.209474
4.3.2.2b	-1.66973e-017	0	0.209474
2.1.2.3a	0.00541345	4.56036e-017	0.209474
2.4.2.7c	0	0	0
3.5.4.10	-0.00541345	-1000	1000
2.1.2.3b	0	-1000	1000
3.1.3.5c	-7.77668	-302.84	-0.02
_2.7.4.3c	0	0	75.705
3.6.1.19a	0	0	75.705
3.6.1.19g	0	0	0
_1.17.4.1b	0	0	0
2.7.4.6j	0	0	151.41
2.7.4.6b	0	0	0
1.1.1.205	7.69274	0.02	302.84
3.5.4.4a	7.77668	0.02	302.84
3.1.3.5d	0	0	0
3.6.1.19b	0	0	0
6.3.5.2	7.69274	0.02	302.84
3.1.3.5a	7.77634	0.02	302.84
2.4.2.7a	0	0	0
2.7.4.3a	0.720697	-1.99	151.41
2.7.4.6a	-7.82911	-1000	1000
2.7.1.40b	0	-1000	1000
1.17.4.1a	7.82911	-151.39	302.84
2.7.4.3b	0	0	0.00409498
3.1.3.5b	0	-0.00409498	0
2.7.7.7a	-7.83032	-302.84	151.39
2.7.7.6a	0	0	151.41
2.7.7.8a	5.5203e-017	0	151.41
3.5.4.4b	0	0	0
3.1.3.5e	0	0	0
2.4.2.7b	0	0	0
2.7.4.8a	7.68533	0.02	302.84
3.6.1.19c	0	0	75.705

1.17.4.1c	-7.68533	-302.84	151.39
2.7.4.6c	0	-1000	1000
2.7.1.40c	-0.0893555	-1000	1000
2.7.4.8b	0	0	75.705
2.7.4.6d	7.69335	-1000	1000
2.7.1.40d	0.0080211	-1000	1000
3.1.3.5f	0	0	0
3.6.1.19d	0	0	75.705
3.1.7.2	0	0	0
3.6.1.11	0	0	0
2.7.7.6b	0	0	151.41
2.7.7.8b	0	0	151.41
2.7.7.7b	7.68381	-151.39	302.84
2.7.7.4	0	0	100.94
2.7.1.25	0	0	151.41
_2.7.7.53a	0	0	50.47
2.7.7.53b	0	-50.47	0
4.6.1.1a	0	0	0
4.6.1.1b	0	0	0
3.5.4.16	0	0	0
4.1.2.25	0	0	0
2.7.6.3	0	0	0
2.5.1.15a	0	0	0
2.5.1.15b	0	0	0
2.6.1.85	0	0	0
·6.3.4.3	0	0	151.41
·4.1.3.38	0	0	0
6.3.2.12	0	0	0
·1.5.1.3	0	0	0
6.3.2.17a	0	0	0
6.3.2.17b	0	0	0
6.3.2.17c	0	0	0
_4.2.3.12	0	0	0
2.1.2.2b	0	0	1000
_1.4.4.2	0.0478024	0	60.564
_2.1.2.10a	0.0478024	0	60.564
_1.8.1.4c	0.0478024	0	60.564
1.5.1.5	0.00541345	4.56036e-017	0.418947
3.5.1.10	0	0	151.41
3.5.4.9a	0.00541345	-1000	0.418947
6.3.3.2	0	0	151.41
2.1.2.10b	0	0	151.41
·1.5.1.20a	-0.00567226	-1000	1000
·1.5.1.20b	0	-1000	1000
2.1.2.9a	0	0	0
_PSII	0	0	0.98
_UQ	0	0	0.98
_citb6	27.0365	0	403.8
_1.10.9.1	0	0	0.98
_PSI	0	0	1.96
_Fd	1.96	0	1.96
_FNR	0.978294	-8.32667e-017	101.927
_1.18.1.2	0.978294	0	101.927
_3.6.3.14	40.0648	0.75	908.52
3.6.1.1	8.62266	0.02	302.84
2.7.4.1	0	0	0
_NADH2	0	0	50.47
_BidHyd	0	0	50.47

_cyclic	0	0	60.956
_NDH1	0	0	0.98
_1.6.5.3	0	0	302.82
_1.6.99.3	0	0	302.82
_quinol oxidase	0	0	67.1367
_cit c	27.0365	0	403.8
_1.9.3.1	13.0283	0	201.41
_PSIaltern	1.96	0	1.96
2.3.1.35	0.0144542	0	0.0144542
2.6.1.11	-0.0144542	-0.0144542	0
1.2.1.38	-0.0144542	-0.0144542	0
2.7.2.8	0.0144542	0	0.0144542
2.3.1.1	0	0	0
4.2.1.1a	0	-10.1106	11.99
_1.4.3.4	0	0	0
_1.2.1.3a	0	0	0
_1.2.1.3b	0	0	0
1.7.7.2	0	-8.96412	0
1.7.7.1	0	-8.96412	0
4.2.1.104	0	0	0
3.5.5.1	0	0	0
1.4.3.16b	0	0	8.85319e-017
_2.5.1.72	0	0	8.85319e-017
2.4.2.19	0	0	8.85319e-017
3.1.3.5k	0	0	151.41
2.7.7.1a	0	0	8.85319e-017
2.7.7.18a	0	0	8.85319e-017
6.3.1.5	0	0	0
2.7.7.1b	0	0	0
2.7.7.18b	0	0	0
3.1.3.5l	0	0	0
2.7.1.23	0	0	0
1.6.1.2	0	-1000	1000
·4.1.3.40	0	0	0
·octaprenylsyn	0	0	0
2.5.1.-a	0	0	0
4.1.1.-a	0	0	0
·OCTAPRENYLPHENOL-HYDROX-RXN	0	0	0
·2.1.1.64b	0	0	0
1.14.13.-a	0	0	0
2.1.1.201	0	0	0
1.14.13.-b	0	0	0
·2.1.1.64a	0	0	0
_UQsyf	0	0	0
5.4.4.2	0	0	0
2.2.1.9	0	0	0
4.2.99.20	0	0	0
4.2.1.113	0	0	0
6.2.1.26	0	0	0
4.1.3.36	0	0	0
2.5.1.74a	0	0	0
2.5.1.74b	0	0	0
2.1.1.163a	0	0	0
2.1.1.163b	0	0	0
3.1.2.28	0	0	0
1.8.4.8	0	0	0
·1.13.11.8	0	0	0

1.8.7.1	0	0	0.00813392
3.1.3.7	0	0	151.41
6.2.1.1a	0.517959	-1.99	100.947
1.2.1.3c	0	-1.99	1.21198e-017
1.1.1.1	0	0	1.99
2.7.2.1a	0	0	0
_1.1.1.86c	0	0	0.0210493
_4.2.1.9c	0	0	0.0210493
2.1.2.11	0	0	0
6.3.2.1	0	0	0
_1.2.1.3d	0	0	0
2.7.1.33	0	0	0
6.3.2.5	0	0	0
4.1.1.36	0	0	0
_2.7.7.3	0	0	0
_2.7.1.24	0	0	0
_6.4.1.2a	0	0	0.0197229
_6.4.1.2b	0.0197229	0	0.0197229
_6.3.4.14	0.0197229	0	0.0197229
2.3.1.39	0.0197229	0	0.0197229
2.3.1.41a	0.00296527	-1.93632e-018	0.00296527
2.3.1.41b	0.00296527	0	0.00296527
1.1.1.100a	0.00296527	0	0.00296527
4.2.1.-a	0.00296527	0	0.00296527
1.3.1.9a	0.00296527	0	0.00296527
2.3.1.41c	0.00296527	0	0.00296527
1.1.1.100b	0.00296527	0	0.00296527
4.2.1.-b	0.00296527	0	0.00296527
1.3.1.9b	0.00296527	0	0.00296527
2.3.1.41d	0.00296527	0	0.00296527
1.1.1.100c	0.00296527	0	0.00296527
4.2.1.-c	0.00296527	0	0.00296527
1.3.1.9c	0.00296527	0	0.00296527
2.3.1.41e	0.00296527	0	0.00296527
1.1.1.100d	0.00296527	0	0.00296527
1.3.1.9d	0.00296527	0	0.00296527
4.2.1.-d	0.00296527	0	0.00296527
2.3.1.41f	0.00296527	0	0.00296527
1.1.1.100e	0.00296527	0	0.00296527
4.2.1.-e	0.00296527	0	0.00296527
1.3.1.9e	0.00296527	0	0.00296527
2.3.1.41g	0.00296527	0	0.00296527
1.1.1.100f	0.00296527	0	0.00296527
4.2.1.-f	0.00296527	0	0.00296527
1.3.1.9f	0.00296527	0	0.00296527
2.3.1.41h	0.00128896	0	0.00128896
1.1.1.100g	0.00128896	0	0.00128896
4.2.1.-g	0.00128896	0	0.00128896
1.3.1.9g	0.00128896	0	0.00128896
2.3.1.179i	0.000642385	0	0.000642385
1.1.1.100h	0.000642385	0	0.000642385
4.2.1.-h	0.000642385	0	0.000642385
1.3.1.9h	0.000642385	0	0.000642385
_1.14.19.1a	0.000395129	0	0.000395129
_1.14.19.1b	0.000374176	0	0.000374176
·3.1.2.14a	0	0	0
·3.1.2.14b	0	0	0
·3.1.2.14c	0	0	0

·3.1.2.14d	0.00167631	0	0.00167631
·3.1.2.14e	0.000251446	0	0.000251446
·3.1.2.14f	0.000395129	0	0.000395129
·3.1.2.14g	0.000268209	0	0.000268209
·3.1.2.14h	0.000374176	0	0.000374176
2.3.1.179a	0	0	0
2.3.1.179b	0	0	0.00296527
2.3.1.179c	0	0	0.00296527
2.3.1.179d	0	0	0.00296527
2.3.1.179e	0	0	0.00296527
2.3.1.179f	0	0	0.00296527
2.3.1.179g	0	0	0.00296527
2.3.1.179h	0	0	0.00128896
2.3.1.180a	0	0	0
2.3.1.180b	0	0	0.00296527
2.3.1.180c	0	0	0.00296527
2.3.1.180d	0	0	0.00296527
2.3.1.180e	0	0	0.00296527
2.3.1.180f	0	0	0.00296527
2.3.1.180g	0	0	0.00296527
2.3.1.180h	0	0	0.00128896
6.2.1.3	0	0	0
tRNA 1.2.1.70	-0.0163799	-0.0163799	0
5.4.3.8	0.0163799	0	0.0163799
4.2.1.24	0.00818996	0	0.00818996
2.5.1.61	0.00204749	0	0.00204749
4.2.1.75	0.00204749	0	0.00204749
_uroporphyrinogen-I_spont	0	0	0
_uroporphyrin-I_spont	0	0	0
4.1.1.37b	0	0	0
_coproporphyrin-I_spont	0	0	0
_uroporphyrin-III_spont	0	0	0
4.1.1.37a	0.00204749	0	0.00204749
_1.3.3.3	0.00204749	0	0.00204749
_1.3.99.22	0	0	0.00204749
·1.3.3.4	0.00204749	0	0.00204749
4.99.1.1	0.00170624	0	0.00170624
2.5.1.-c	0	0	0
_COX15	0	0	0
_1.14.99.3	0.00170624	0	0.00170624
1.3.7.5	0.00170624	0	0.00170624
6.6.1.1	0.000341248	0	0.000341248
2.1.1.11	0.000341248	0	0.000341248
Zn-Bacterio-chlorophyll-a	0	0	0
1.14.13.81a	0.000341248	0	0.000341248
1.14.13.81b	0.000341248	0	0.000341248
1.14.13.81c	0.000341248	0	0.000341248
1.3.1.33a	0	0	0.000340664
·1.3.1.75a	0	-8.62589e-017	0.000340664
·1.3.1.75b	-2.1684e-019	-1.78551e-036	0.000340664
1.3.7.7	0	0	0.000340664
_BacChl	0	0	0
_BacPheo	0	0	0

1.3.1.33b	0	0	0.000340664
_dChla	0.000341248	0	0.000341248
_Chla	0.000341248	0	0.000341248
2.5.1.62a	0	0	0.000340664
_Chlb	0	0	0
_Pheo	0	0	0
2.1.1.107a	0	0	0
2.1.1.107b	0	0	0
_2.1.1.130	0	0	0
·4.99.1.3	0	0	0
2.1.1.151	0	0	0
2.1.1.131a	0	0	0
2.1.1.133a	0	0	0
3.7.1.12	0	0	0
2.1.1.195	0	0	0
_1.3.1.54a	0	0	0
·RXN-8766	0	0	0
·RXN-8767	0	0	0
5.4.1.2a	0	0	0
6.3.5.11	0	0	0
2.1.1.131b	0	0	0
2.1.1.133b	0	0	0
·2.1.1.152	0	0	0
1.3.1.54b	0	0	0
2.1.1.132	0	0	0
5.4.1.2b	0	0	0
6.3.5.9	0	0	0
6.6.1.2	0	0	0
·1.16.8.1	0	0	0
2.5.1.17a	0	0	0
6.3.5.10	0	0	0
6.3.1.10a	0	0	0
·RXN-8626	0	0	0
·4.1.1.81	0	0	0
6.3.1.10b	0	0	0
2.5.1.17b	0	0	0
2.7.1.156	0	0	0
2.7.7.62	0	0	0
2.7.8.26	0	0	0
3.5.4.25	0	0	0
3.5.4.26	0	0	0
1.1.1.193	0	0	0
2.5.1.78	0	0	0
DIOHBUTANONEPSYN-RXN	0	0	0
2.5.1.9	0	0	0
_lumazine-spont	0	0	0
2.7.1.26	0	0	0
2.7.7.2	0	0	0
_benzimidazole-spont	0	0	0
·2.4.2.21	0	0	0
·3.1.3.73	0	0	0
·2.4.2.1	0	0	0
·2.7.1.1	0	0	151.41
_1.8.1.9	0.514222	-51.45	100.947
·_Thiored	0	0	100
1.11.1.6	0	0	50.3525
_1.15.1.1	0	0	0
1.1.1.38	0	0	37.8525

4.1.1.31	0.420957	0	60.564
1.2.4.1	0	0	50.47
2.3.1.12	0	-50.47	0
1.8.1.4a	0	0	50.47
·4.2.3.3	0	0	75.705
4.4.1.5	0	-75.705	0
3.1.2.6	0	0	75.705
1.1.1.28	0	0	75.705
3.4.11.2a	3.3695e-018	0	75.705
3.4.11.2b	0	0	0
6.3.2.3	0	0	75.705
_2.3.2.2a	1.97901e-017	0	75.705
_2.3.2.2b	-1.46859e-017	0	50.47
_2.3.2.2c	0	0	50.47
_2.3.2.2d	0	0	50.47
_2.3.2.2e	0	0	50.47
_2.3.2.2f	0	0	50.47
_2.3.2.2g	0	0	50.47
_2.3.2.2h	0	0	50.47
_2.3.2.2i	0	0	50.47
_2.3.2.2j	0	0	50.47
_2.3.2.2k	0	0	50.47
_2.3.2.2l	0	0	50.47
_2.3.2.2m	0	0	0
_2.3.2.2n	0	0	50.47
_2.3.2.2o	0	0	50.47
_2.3.2.2p	0	0	50.47
_2.3.2.2q	0	0	50.47
_2.3.2.2r	0	0	50.47
_2.3.2.2s	0	0	50.47
_2.3.2.2t	0	0	50.47
3.5.2.9	0	0	50.47
_2.5.1.18	0	0	0
_2.3.2.2u	0	0	0
1.11.1.9	0	0	33.6467
_1.8.1.7a	0	0	33.6467
_1.8.1.7b	0	0	33.6467
·6.3.2.2	0	0	75.705
·2.3.2.4a	-1.46859e-017	0	50.47
·2.3.2.4b	0	-7.67119e-017	50.47
·2.3.2.4c	0	0	50.47
·2.3.2.4d	0	0	50.47
·2.3.2.4e	0	0	50.47
·2.3.2.4f	0	0	50.47
·2.3.2.4g	0	0	50.47
·2.3.2.4h	0	0	50.47
·2.3.2.4i	0	0	50.47
·2.3.2.4j	0	1.09732e-016	50.47
·2.3.2.4k	0	0	50.47
·2.3.2.4l	0	0	0
·2.3.2.4m	1.46859e-017	0	50.47
·2.3.2.4n	0	0	50.47
·2.3.2.4o	0	-4.56036e-017	50.47
·2.3.2.4p	0	0	50.47
·2.3.2.4q	0	0	50.47
·2.3.2.4r	0	3.14411e-017	50.47
·2.3.2.4s	0	0	50.47
·2.3.2.4t	0	0	0

1.1.1.284a	0	0	1000
1.1.1.284b	0	-1000	0
3.1.2.12	0	0	0
4.1.2.13c	0	0	0
2.7.1.4	0	0	75.705
5.3.1.8	0	0	0
5.4.2.8	0	0	0
2.7.7.13	0	0	0
4.2.1.47	0	0	0
1.1.1.271	0	0	0
2.4.1.83	0	0	0
2.7.7.22	0	0	0
2.6.1.16	0	-1.7436e-016	151.41
3.5.99.6	0	0	151.41
3.5.1.25	0	0	0
4.2.1.126	0	0	0
5.4.2.10	0	0	0
2.3.1.157	0	0	0
2.7.7.23	0	0	0
2.5.1.7	0	0	0
_1.1.1.158	0	0	0
_5.1.3.14a	0	0	0
_5.1.3.14b	0	0	0
5.1.3.9	0	0	0
6.3.2.8	0	0	0
6.3.2.9	0	0	0
6.3.2.13	0	0	0
6.3.2.10a	0	0	0
2.7.8.13a	0	0	0
2.4.1.227a	0	0	0
_6.3.1.2b	0	0	0
6.3.2.10b	0	0	0
2.7.8.13b	0	0	0
2.4.1.227b	0	0	0
_6.3.1.2c	0	0	0
6.3.2.10c	0	0	0
2.7.8.13c	0	0	0
2.4.1.227c	0	0	0
_6.3.1.2d	0	0	0
·2.3.2.10a	0	0	0
·2.3.2.10b	0	0	0
·2.3.2.10c	0	0	0
6.3.2.4	0	0	151.41
3.6.1.27	0	0	0
_peptidoglycan_syfa	0	0	0
_2.4.1.129	0	0	0
3.4.16.4	0	0	0
3.5.1.28	0	0	0
2.3.1.129	0	0	0
3.2.1.24	0	0	0
2.4.1.14	0	0	75.705
3.2.1.26	0	0	75.705
2.4.1.25	0	0	0
_2.4.1.1	0	0	151.41
2.7.7.27	0.191957	0	151.41
_a-glucansyf	0.0319929	-3.93972e-017	0.0552778
_2.4.1.21a	0	0	151.41
_glycsyf	0.0319929	-3.93972e-017	0.0552778

_2.4.1.21b	0.0639858	0	151.41
_2.4.1.18	0.0319929	-151.41	0.0552778
5.4.2.2	-0.191957	-75.705	1.84266e-016
4.2.1.45	0	0	0
2.4.1.12	0	0	0
_3.2.1.21a	0	0	0
_3.2.1.21b	0	0	0
2.2.1.7	0.00472587	0	0.00608043
1.1.1.267	-0.00472587	-0.00608043	0
2.7.7.60	0.00472587	0	0.00608043
2.7.1.148	0.00472587	0	0.00608043
4.6.1.12	0.00472587	0	0.00608043
1.17.7.1	-0.00472587	-0.00608043	0
1.17.1.2a	0.00472587	0	0.00608043
1.17.1.2b	0	0	0
5.3.3.2	0.00118147	0	0.00152011
2.5.1.29a	0.00118147	0	1000
2.5.1.1	0.00118147	0	0.00152011
2.5.1.84	0	0	0
2.5.1.85	0	0	0
1.3.1.83	0	0	0.000340664
_VitE-spont	0	0	0
_2.3.1.181a	0	0	0
_2.8.1.8a	0	0	0
_2.8.1.8b	0	0	0
_2.3.1.181b	0	0	0
_2.7.7.63a	0	0	0
·transp_lipoate	0	0	0
_2.7.7.63b	0	0	0
_2.5.1.32	0.000590734	2.69789e-036	0.000590734
_1.3.5.5a	0.000590734	2.69789e-036	0.000590734
_1.3.5.5b	0.000590734	2.69789e-036	0.000590734
1.3.5.6a	0.000590734	2.69789e-036	0.000590734
1.3.5.6b	0.000590734	2.69789e-036	0.000590734
CrtLb-a	0	0	0
CrtLb-b	0	0	0
CrtLb-c	9.96804e-005	0	9.96804e-005
CruA	0	0	9.96804e-005
CrtLb-d	9.96804e-005	0	9.96804e-005
_CrtRa	4.72958e-005	0	4.72958e-005
_CrtRb	4.72958e-005	0	4.72958e-005
·CrtOa	0	0	0
·CrtOb	0	0	0
_CrtRc	0	0	0
_CrtRd	0	0	0
_CrtRe	0	0	0
_CrtRf	0	0	0
_gCar	0	0	0
_CrtRg	0	0	0
5.2.1.13	0	0	0
·2.7.7.9	0	0	75.705
5.1.3.2a	0	0	0
3.13.1.1	0	0	0
1.1.1.22	0	0	0
2.7.7.24	0	0	0
5.1.3.2b	0	0	0
2.7.7.33	0	0	0
4.2.1.46	0	0	0

5.1.3.13a	0	0	0
5.1.3.13b	0	0	0
1.1.1.133a	0	0	0
1.1.1.133b	0	0	0
5.4.99.9	0	0	0
6.2.1.1b	0	0	0
2.7.2.1b	0	0	0
6.2.1.1c	0	0	0
_1.2.1.3e	0	0	0
4.2.1.99	0	0	0
PYRIMSYF1-RXN	0	0	0
·2.7.1.49	0	0	0
2.7.4.7	0	0	0
1.4.3.19	0	0	0
ThiG	0	0	0
2.5.1.3	0	0	0
2.7.4.16	0	0	0
·2.7.1.89	0	0	0
3.5.99.2	0	0	0
2.3.1.47	0	0	0
2.6.1.62	0	0	0
6.3.3.3	0	0	0
2.8.1.6	0	0	0
6.3.4.15	0	0	0
RXN-7101	0	0	0
_3.4.1.-	0	0	0
_3.5.2.6	0	0	0
_2.5.1.31a	0	-1000	0.00152011
_2.5.1.31b	0	0	0
2.5.1.10	0.00118147	0	0.00152011
2.6.99.2	0	0	0
_1.1.1.262	0	0	0
RXN-8447	0	0	0
4.2.3.1b	0	0	0
_VitB6	0	0	0
1.1.1.94	0	0	0
_2.3.1.15	0	0	0
2.3.1.51	0	0	0
2.7.7.41	0	0	0
2.7.1.107	0	0	0
2.7.8.5	0	0	0
3.1.3.25	0	0	0
3.1.4.-	0	0	0
1.1.1.31	0	0	0
5.3.1.13	0	0	0
2.5.1.55	0	0	0
3.1.3.45	0	0	0
2.7.7.38	0	0	0
3.5.1.108	0	0	0
2.4.1.182	0	0	0
2.7.1.170	0	0	0
1.1.1.6	0	0	0
4.1.99.5	0	0	0
2.8.1.1	0	0	0
3.8.1.5	0	0	0
4.1.99.12	0	0	0
3.4.13.22	0	0	151.41
1.14.99.41	0	0	0

1.12.98.1	0	0	0
3.1.1.45	0	0	0
2.5.1.77	0	0	0
1.7.1.13	0	0	0
2.10.1.1	0	0	0
_1.16.1.1	0	0	0
3.5.4.-	0	0	0
2.3.1.28	0	0	0
4.2.1.35	0	0	0
3.1.3.71	0	0	0
2.1.1.104	0	0	0
Alabm	0.0246491	0	0.0246491
Aspbm	0.0142344	0	0.0142344
Argbm	0.0144542	0	0.0144542
Asnbm	0.0102773	0	0.0102773
Lysbm	0.0114589	0	0.0114589
Thrbm	0.0159381	0	0.0159381
Serbm	0.0150587	0	0.0150587
Cysbm	0.0028029	0	0.0028029
Glybm	0.0192906	0	0.0192906
Glubm	0.0168724	0	0.0168724
Glnbm	0.0158282	0	0.0158282
Probm	0.0140695	0	0.0140695
Trpbm	0.00409444	0	0.00409444
Phebm	0.0111567	0	0.0111567
Tyrbm	0.00807896	0	0.00807896
Hisbm	0.00541345	0	0.00541345
Valbm	0.0175319	0	0.0175319
Ilebm	0.0172571	0	0.0172571
Leubm	0.00351737	0	0.00351737
Metbm	0.00533102	0	0.00533102
_a protein	2.74795e-005	0	2.74795e-005
dATP	0.00120428	0	0.00120428
dGTP	0.00151972	0	0.00151972
dTTP	0.00120428	0	0.00120428
dUTP	0	0	0
dCTP	0.00151972	0	0.00151972
AMP	0.00840484	0	0.00840484
UMP	0.00840484	0	0.00840484
CMP	0.00740843	0	0.00740843
GMP	0.00740843	0	0.00740843
8C-lipidbm	0	0	0
10C-lipidbm	0	0	0
12C-lipidbm	0	0	0
14C-lipidbm	0.00167631	0	0.00167631
16C-lipidbm	0.000251446	0	0.000251446
18C-lipidbm	0.000268209	0	0.000268209
(9Z)16C-lipidbm	0.000395129	0	0.000395129
(9Z)18C-lipidbm	0.000374176	0	0.000374176
PhCybibm	0.00170624	0	0.00170624
Chlabm	0.000341248	0	0.000341248
Carotenebm	0.000491053	0	0.000491053
Glycbm	0.0319929	0	0.0319929
Zeaxbm	4.72958e-005	0	4.72958e-005
Beta-carotenebm	5.23846e-005	0	5.23846e-005
ATPbm	0	0	0
ADPbm	0	0	0
Biomass	0.0598681	0	0.0598681

_Growth	0.0598681	0	0.0598681
_lightII	0	0	1.96
_lightI	1.96	0	1.96
CO2in	1.99	0	1.99
4.2.1.1b	-1.8105	-10	10
_H2CO3transport	1.99	0	1.99
_H2CO3desprot	1.99	-10	10.2211
NADHmain	0.5	0.5	50.97
NADPHmain	0.5	0.5	50.97
Phosphate TRANS-RXN59G-90	0.305095	0	100.442
_3.6.3.27	0	0	100.442
_Phosphate_out	0	0	100
H2O	-26.2401	-250.41	52.41
PROTONS	-15.8535	-356.31	149.41
sulfate TRANS-RXN59G-407	0	-104	50.47
nitrate TRANS-RXN59G-237	0	0	8.96412
ammonia H3N TRANS-RXN59G-178	0.838087	-151.41	151.41
_ammonia H3N	-0.838087	-151.41	151.41
iron TRANS-RXN59G-711	0	0	0
magnesium TRANS-RXN59G-340	0	0	151.41
magnesium TRANS-RXN59G-53	0.000341248	-151.41	0.000341248
cobalt TRANS-RXN59G-71	0	0	0
_COtransp	-0.00170624	-0.00170624	0
_H2O2transp	26.8715	-0.705	100
_O2in	0	0	1000
_O2out	13.8309	0	1000
potassium 3.6.3.12	0	0	0
potassium TRANS-RXN59G-35	0	0	0
chloride TRANS-RXN59G-171	0	0	0
chloride_ATPase	0	0	0
sodium_symporter	0	0	0
sodium_antiporter	0	0	0
calcium TRANS-RXN59G-708	0	0	0
molybdate_transport	0	0	0
_3.6.3.4	0	0	0
_3.6.3.30	0	0	0
_3.6.3.25	0	0	105
_H2	0	0	50.47
sulfide TRANSP	0.00813392	0	0.00813392
sulfite TRANSP	0	0	0.00813392
_an [acyl-carrier protein]	0.00593053	-1.93632e-018	0.00593053
·_a protein disulfide R1Protein2R2R3S2	0	0	0
·_a protein dithiol H2R1Protein2R2R3S2	0	0	0

_3.1.3.16	1.70962e-020	0	151.41
_2.7.11.1	0	0	151.41
_RNAout	-5.5203e-017	0	0
_DNAout	-3.08998e-016	0	0
_glycout	0	0	0.0552778
_glycolaldehyde	0	0	0
_eth	0	0	1.99

Vita

Name: Julián Triana Dopico.

Address: Department of Chemistry

Faculty of Forestry and Agronomy, University of Pinar del Río

Calle Martí, No. 270, Esquina a 27 de noviembre, Pinar del Río, Cuba

Tel: +34 48 779661.

Email Address: jtriana@upr.edu.cu, jtriana96@gmail.com.

Education: B.S., Biochemistry, University of Havana, Cuba, 2007.

M.S., Forestry Science, University of Pinar del Río, Cuba, 2011.