UNIVERSITÀ DEGLI STUDI DI PADOVA

FACOLTÀ DI INGEGNERIA
DIPARTIMENTO DI PRINCIPI E IMPIANTI DI INGEGNERIA
CHIMICA "I. SORGATO"

TESI DI LAUREA IN INGEGNERIA CHIMICA E DEI PROCESSI INDUSTRIALI

SYSTEMS BIOLOGY APPROACHES FOR CHEMICAL ENGINEERING

Relatore: Prof. Alberto Bertucco
Correlatore: Dr. Daniel Segrè

Laureando: FRANCESCO GATTO

ANNO ACCADEMICO 2010-2011

Alla memoria di Steve Jobs,

Riassunto

La crescente spinta a ricercare alternative rinnovabili ai prodotti chimici derivati dal petrolio ha indotto a considerare le biotecnologie più che una frontiera un obiettivo indispensabile della moderna ingegneria chimica. Le applicazioni industriali dei sistemi biologici sono ancora esigue, ma spaziano dall'industria alimentare agli impianti di trattamento delle acque inquinate. In quest'ottica, l'ingegneria chimica deve dotarsi di nuovi strumenti per un'analisi efficace ma rapida della qualità delle proposte che le biotecnologie offre a fronte della domanda di nuove vie chimiche per sviluppare determinati prodotti.

Al fine di condurre quest'analisi, la biologia dei sistemi è una branca emergente della biologia che abbraccia la filosofia di considerare i sistemi biologici nel loro insieme, enumerando le interazioni che si sviluppano tra le parti e quantificando la dinamica delle variazioni nel sistema, siano esse locali o globali, a fronte di certe perturbazioni nell'ambiente di crescita, nel genoma, ecc. Lo studio dei sistemi biologici è successivo alla ricostruzione *in silico* dei network metabolici di diversi microorganismi a partire dal loro stesso genoma. Queste ricostruzioni, note anche con l'acronimo di GENRE (*genome-based metabolic network reconstructions*), comprendono per ciascun microorganismo una quantità rappresentativa di metaboliti e di reazioni che li correlano, come in un grafo. Successivamente, con l'ausilio della bioinformatica, la biologia dei sistemi ha adottato delle tecniche per l'analisi di questi network metabolici, tra le quali la *flux balance analysis* (FBA). Questo metodo permette di predire la mappa dei flussi metabolici, intesi come le velocità delle reazioni enzimatiche che hanno luogo all'interno di una cellula, in un regime stazionario definito dall'ottimizzazione di una certa funzione cellulare. La qualità delle predizioni è risultata tanto migliore tanto più rappresentativa la funzione cellulare massimizzata: generalmente il principio evolutivo per cui il metabolismo di una cellula viene ottimizzato affinché sia massimizzata la sua duplicazione ha prodotto i risultati più rimarchevoli.

In questo lavoro, i metodi della biologia dei sistemi basati sulla FBA sono stati esaminati e modificati per rispondere alle esigenze di problemi tipici dell'ingegneria chimica. Un nuovo approccio per la progettazione di bioreattori tubulari, che sfrutta le predizioni della FBA per ricavare il valore di variabili chiave quali le velocità di reazioni, è stato proposto per bypassare la cronica mancanza di dati cinetici relativi ai sistemi biologici, spesso rappresentati da equazioni cinetiche multiparametriche. In seguito, sono state studiate varie applicazioni nell'ambito dell'ingegneria delle reazioni

chimiche, o ingegneria metabolica nel caso sia applicata ai sistemi biologici, intesa come il processo di ottimizzazione di una reazione di interesse al fine di aumentarne la produttività. In particolare si è posta enfasi sull'effetto di singoli *knock-out* genici nella massimizzazione di un dato flusso e si è sviluppato un nuovo metodo per verificare come addizionando certi metaboliti al medio di crescita si possa accrescere ulteriormente la velocità di una data reazione.

Questi metodi della biologia dei sistemi per l'ingegneria chimica sono stati applicati con successo a un caso di studio, la produzione di precursori del biodiesel, cioè i triacilgliceroli (TAG), da parte della microalga *Chlamydomonas reinhardtii*, il cui network metabolico è stato efficacemente ricostruito di recente. Sfruttando il nuovo approccio, dopo aver simulato le condizioni per cui la microalga prima cresce e poi accumula triacilgliceroli, si è realizzata la progettazione di un fotobioreattore, producendo i profili delle concentrazioni dei vari componenti, riportando l'andamento della produttività con il tempo di residenza e valutando il consumo di nutrienti. In seguito, si è tentato mediante singoli *knock-out* genici dapprima di ottimizzare la produzione di TAG, e successivamente, dati gli scarsi risultati, la sintesi di idrogeno gassoso. Per quest'ultimo caso è risultato che la deprivazione da zolfo comporterebbe l'attivazione della deidrogenasi responsabile della formazione di H_2, un esito che riproduce pur per ragioni diverse una scoperta che risale agli inizi del 2000. Infine, con l'applicazione della nuova tecnica di miglioramento del mezzo di crescita, si sono delineate due strategie che potrebbero consentire un'accelerazione della velocità di crescita della biomassa. Risulta infatti che l'incorporazione diretta nel metabolismo di urocanato, intermedio nel catabolismo dell'L-istidina, e di N-acetil-L-glutammato, intermedio nella biosintesi dell'arginina, potrebbe incrementare la formazione di biomassa dal momento che la loro disponibilità nel medio di crescita consentirebbe alla cellula di sintetizzare precursori essenziali alla cellula (come l'acido α-chetoglutarico) durante la loro digestione.

In conclusione, in questo lavoro si sono evidenziati i vantaggi che la biologia dei sistemi ha apportato nella descrizione e analisi dei sistemi biologici, con particolare enfasi sulle proprietà dei network metabolici cellulari. Una migliore descrizione del metabolismo consente all'ingegnere chimico di apprendere quale sia la migliore tra le innumerevoli opzioni che la natura offre per supplire alla domanda di certi composti chimici, e la sua analisi gli fornisce gli strumenti adatti per ottimizzarla.

Premessa

Il presente lavoro è stato realizzato nell'ambito del programma di scambio internazionale stipulato tra l'Università degli Studi di Padova e la Boston University, MA, USA. Esso è il risultato dell'incontro tra le ricerche del gruppo GreEn Development del prof. Bertucco, che studia la sintesi di olio microalgale come possibile biocarburante, e del gruppo del Dr. Segrè, che si occupa di dinamica e evoluzione dei network biochimici. La tesi altro non è che un tentativo di coniugare le due linee ricerca, nell'intenzione che ciascuna disciplina possa giovare delle conoscenze e competenze offerte dall'altra.

L'idea che l'ingegneria chimica possa servirsi della biologia dei sistemi è invero datata, considerato che tra i pionieri di questa branca della biologia figura il prof. Bernhard Ø. Palsson, che è ingegnere chimico. Tuttavia spesso l'applicabilità dei risultati ottenuti con le simulazioni dei network metabolici è stata trascurata, e questo è stato lo stimolo a riesaminare i metodi e le potenzialità che questa materia può offrire all'ingegneria chimica, ricercando al contempo nuovi approcci per risolvere delle comuni problematiche.

Nel primo capitolo sono riassunti gli strumenti che la biologia dei sistemi ha sviluppato per l'analisi dei network metabolici ricostruiti, scendendo nel dettaglio della *flux balance analysis*, discutendone tecniche avanzate, tra cui alcune di ingegneria metabolica.
Nel secondo capitolo vengono esaminate le possibili applicazioni dei predetti strumenti a due campi dell'ingegneria chimica, la progettazione di bioreattori e l'ingegneria di reazioni biochimiche, introducendo nuovi approcci dove possibile e richiamandone di consolidati negli altri casi.
Nel terzo capitolo si espone il caso di studio a cui quest'incontro tra biologia dei sistemi e ingegneria chimica viene applicato, cioè la produzione di precursori del biodiesel, i triacilgliceroli (TAG), da parte della microalga *Chlamydomonas reinhardtii*, di cui si illustra l'ottima ricostruzione metabolica recentemente disponibile.
Nel quarto capitolo si approfondisce, per il caso di studio, la progettazione di un fotobioreattore.
Nel quinto capitolo, infine, si analizzano i tentativi realizzati *in silico* di ingegnerizzare la produzione di TAG da parte della microalga. Seguono quindi le conclusioni.

Contents

Introduction

Problems of organization, phenomena not resolvable into local events, dynamic interactions manifested when parts of a system are isolated, are some of the issues which induced contemporary science to recognize the importance of "wholeness" (Chong and Ray, 2002). A cell is an example of a system which synthesizes all the previous problems. Systems biology is thus the branch of biology whose aim is the system-level understanding of any biological system, starting from the cell to the ecosystems (Kitano, 2002). Modern progresses in molecular biology boosted the collection of comprehensive data over increasingly complex biological systems. The recent emergence of systems biology compared to the early pronouncement of its theory is due to the necessity to accumulate sufficient descriptions of the parts to enable a reasonable reassembly of the whole. However, although systems biology is in its infancy, its potential benefits such as the ability to explain robustness or to predict metabolic behaviours are enormous in both scientific and practical terms.

Biological systems are becoming of central interest also in fields which go beyond science and medicine scopes, such as chemical engineering. Industrial production of some compounds requires steps that can be catalyzed by highly specific enzymes (for example, in syrup production, enzymes are employed to convert starch into glucose). In the last decades, more complex biological systems have been considered for large scale chemicals production. Among these, photosynthetic algae, with particular emphasis on microalgae, have been regarded to as a possible biofuel resource for several benefits they owe over other current technologies.

Systems biology approaches can be productively coupled to investigate some chemical engineering applications, such as bioreactor design and metabolic engineering, defined as the direct improvement of product formation or cellular properties by modifications of the metabolic network of reactions, or introduction of new ones by means of genetic engineering (Stephanopoulos *et al*., 1998). It is the objective of this work to understand the potentialities of combining systems biology approaches to chemical engineering problems. A case study regarding the production of biodiesel precursors, namely triacylglycerols (TAGs), by *Chlamydomonas reinhardtii*, a single celled microalgae, was investigated as an example. Two main applications were considered: the design of a tubular photobioreactor and the strategies for TAGs production optimization by means of single gene knock-outs and growth medium refinement.

Chapter 1

Systems biology approaches to metabolism analysis

Systems biology is the branch of biology which focuses on the study of complex biological systems using a top-bottom approach, as opposed to traditional biology which rather describes single molecular components involved in a specific biological activity. Among the first challenges of systems biology was the description and analysis of biological network as a whole, in particular the cell. The availability of complete genome sequences for a variety of species has allowed an increasing number of metabolic network reconstructions, progressively more precise as the annotation process developed. The applications for these so-called genome-based metabolic reconstruction (GENREs) are broad, and include contextualization of high-throughput data, strategies of metabolic engineering, directing hypothesis-driven discovery, interrogation of multi-species relationships, and network property discovery (Oberhardt *et al*., 2009). Due to the intrinsic complexity of these networks, most of the analysis must be carried out computationally, thus leading to a tight relationship between systems biology and bioinformatics tools.

1.1 Introduction to Flux Balance Analysis

While the availability of complete genomes for several species has revealed the entire set of molecular components involved in cellular activities, their functional integration in a highly developed and evolved system, such as a cell, calls for better instruments and models to be fully understood and analyzed. Molecular components in a cell indeed relate to each other in a very complex manner, which departs from a simple network of reactions. Cells comprise regulatory networks as well as signaling networks, not to mention the inherent physical chemistry of the cell, which is far to be totally understood. Thus approaches that focus on the systemic properties of the network are required.

It is the aim of systems biology to head off from the reductionist approach, which characterized biology in the last centuries, and explore an integrated approach, able to describe the interrelatedness of gene function and the role of each gene in the context of multi-genetic cellular functions. To deal with such systems, engineering is thus called to

design and provide mathematical or computer models which should, at least in principle, reproduce the cellular metabolism in a dynamic fashion. However, the reconstruction of dynamic models of a cell has been hampered by the current lack of kinetic information, in particular the enzymatic kinetic constants. To cope with this deficiency, new approaches have been derived to shed light on the relative importance of various metabolic events. Above all, *flux balance analysis* (FBA) has proved to give excellent insights of a metabolic network without using such kinetic data.

Flux balance analysis takes advantage of three fundamental assumptions. First, cellular activity can be described in terms of a metabolic network. In other words, the conversion of substrates to products and essential components for cellular growth is processed by a web of interrelated reactions which take place in the cell and are constrained by their stoichiometry. Most of these reactions are carried out by enzymes which had been previously identified to be present in the cell by a systematic annotation of cell genome. Other reactions can be artificially introduced in order to describe uptake and secretion of metabolites, as well as not yet identified reactions which are in turn essential for a complete description of cell metabolism. A second assumption is that all reactions occur at steady-state: it derives from the need of cell homeostasis, which ultimately reflects the fact that a particular metabolite should not accumulate within the cell along with time. Lastly, as a third hypothesis, all reactions occur at a rate such that one (or more) cellular activity is kept at its optimum. The rate at which each reaction is carried out is commonly referred to as *flux*. Frequently, it is assumed that growth is optimized, a hypothesis which originates from an evolutionary principle of optimality, the ultimate goal of a microorganism being a faster replication.

1.2 Mathematical modeling of FBA

FBA is a mathematical approach for analyzing the flow of metabolites through a metabolic network (Orth *et al.*, 2010). The main objective of this technique is to compute a feasible distribution of fluxes (i.e. reaction rates) within the cell. As stated above, FBA does not require any knowledge regarding the actual kinetics of the reactions occurring in the cell metabolism. Indeed, FBA assumes that all internal metabolite concentrations do not vary in time, that reactions are constrained by stoichiometry (and either thermodynamic reversibility or irreversibility), and that all reaction rates are optimized with respect to some cellular activities. FBA needs these concepts to be incorporated in a mathematical model. The analysis is carried out in multiple steps (Kauffman *et al*., 2003).

1.2.1 System definition

First, all metabolic reactions and metabolites are identified. Even if in a certain environmental condition only a subset of reactions will be active, the mathematical modeling is supposed to predict the pathway flux, neglecting an explicit consideration of the regulatory mechanisms. Ideally, the complete network can be inferred by the annotated genome sequence. The product of each gene is annotated by homology searches resulting in a list of enzymes leading each to a specific reaction. However, generally few enzymes will be identified in a known pathway, implying that the process is iterated until the network is complete. Sometimes manual addition of other reactions, based on physiological or biochemical data, is required to close the mass balance. It is important to note that genome-based metabolic reconstructions are time-consuming researches. It can take up to two years, depending on the size of the network. An extensive catalog of reconstructions currently available is listed in the BiGG database (Schellenberger, J. *et al.,* 2010), freely accessible from the academic site of the University of California, San Diego at http://bigg.ucsd.edu/. Other reconstructions can be retrieved in specific literature.

An example of a simple network, which will serve as an example for the next steps, is represented in Figure 1.1. Note that intracellular reactions are shown in black while exchange reactions, which are usually manually curated, are in red.

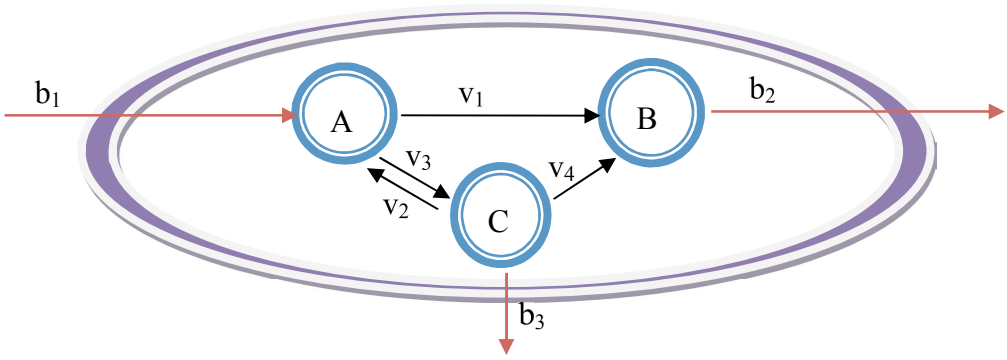

Figure 1.1. *A simple network of reactions occurring in a cell, the metabolites being A, B and C. Intracellular fluxes are shown in black, while exchange fluxes are shown in red.*

1.2.2 Mass balance

Each identified metabolite must satisfy a dynamic mass balance. For instance, for the network represented in Figure 1.1, three dynamic mass balances must be solved:

$$\begin{cases} \frac{dC_A}{dt} = -v_1 - v_2 + v_3 + b_1 \\ \frac{dC_B}{dt} = v_1 + v_4 - b_2 \\ \frac{dC_C}{dt} = v_2 - v_3 - v_4 - b_3 \end{cases} \quad (1.1)$$

where C_A, C_B and C_C are the concentration of metabolites A, B and C respectively and v or b are the rates of the respective reaction in the network, i.e. the fluxes.

This gives rise to a set of coupled ordinary differential equations. The differential equations can be represented using a matrix notation, where **S** is the stoichiometric matrix and **v** is the vector of fluxes. Thus, the matricial product **S·v** should be equal to **t**, which is the vector of metabolite concentration variation in time, i.e. the left hand side of each equation in (1.1). Since FBA assumes steady-state, **t** is equal to zero. Thus for the network in Figure 1, the matricial product **S·v** can be written as:

$$\begin{bmatrix} -1 & -1 & 1 & 0 & 1 & 0 & 0 \\ 1 & 0 & 0 & 1 & 0 & -1 & 0 \\ 0 & 1 & -1 & -1 & 0 & 0 & -1 \end{bmatrix} \begin{bmatrix} v_1 \\ v_2 \\ v_3 \\ v_4 \\ b_1 \\ b_2 \\ b_3 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \end{bmatrix} \quad (1.2a)$$

or shortly,

$$\boldsymbol{S} \cdot \boldsymbol{v} = 0 \quad (1.2b)$$

It should be noted that **S** is not a square matrix, i.e. there are more reactions than metabolites in the network. This case is by far the most common in metabolic reconstructions. Indeed, evolution has provided organisms the capability to reproduce and survive in several environmental conditions. The adaptability of a species to an environment is reflected in its metabolic network in terms of *silent phenotypes*, that is the availability of different pathways which are not active in standard conditions. For example it is well known that under aerobic conditions the more efficient glycolitic pathway is active in *E. coli* for ATP production (this stands for many other organisms as well). However, as soon as the availability of oxygen is poor, fermentation will be preferred over glycolysis. This means that metabolic networks comprise more reactions than metabolites, and therefore the corresponding stoichiometric matrix is never square (Palsson, 2006). As there are more reactions (hence fluxes) than metabolites, the steady-state solution for the flux distribution is underdetermined. Thus, additional constraints are needed to reduce the solution space dimension and ultimately search for a unique steady-state flux distribution.

1.2.3 Defining constraints

Additional constraints, including those related to the availability of nutrients (i.e. fluxes that transport a certain metabolite from the extracellular environment into the intracellular network, like reaction b_1 in Figure 1.1) or to the maximal fluxes that can be supported by enzymatic pathways, can be introduced as inequalities:

$$\alpha_i \leq v_i \leq \beta_i \tag{1.3}$$

For example, fluxes such as b_1 in the network of Figure 1.1, which are usually referred as exchange fluxes, can be constrained to a maximal value. Usually these values are determined experimentally (Varma and Palsson, 1994). Inequalities as in (1.3) can also be used to force irreversibility of a given flux, namely imposing $\alpha_i = 0$. In this fashion, also thermodynamic constraints can be implemented in FBA. Constraints can also define a set of fluxes that have been suppressed by, say, knocking out a specific gene which is known to encode the enzyme(s) catalyzing that reaction. In this case, $\alpha_i = \beta_i = 0$. Sometimes measured values of fluxes are available experimentally, for example via ^{13}C-isotope labeling or other techniques (Park *et al.*, 2010), so that feasible values for α_i and β_i can be selected. However, if no information is known for a given reaction, it is usually assumed to be unbounded (i.e. reversible) such that $-\alpha_i = \beta_i = \infty$ or, for numerical purposes, $-\alpha_i = \beta_i = 1000$.

1.2.4 Optimization

Since biological metabolic networks always comprise more fluxes than metabolites, even an extensive use of constraints rarely leads to an exactly determined solution for problem (1.2). Thus, to identify a flux distribution for the network, it is assumed that the network itself is optimized with respect to a certain objective. According to this approach, it is imposed that the network must achieve a specific goal by redistributing the fluxes toward that objective, without violating mass balance constraints. The nature of this goal must be carefully selected, as it must represent as closely as possible the actual behavior of the system under the examined conditions (Sendìn, 2008).

An evolutionary principle of optimality is usually adopted, which means that the flux distribution within a cell is tailored to pursue the maximum growth rate. The objective function (Z) is thus defined as the growth flux, or biomass objective function (BOF), that is a fictitious reaction which resembles all the metabolites that, in the correct proportions, are needed by the cell to form a unit mass of biomass. Other objective functions have been explored as well, such as maximization of ATP or overproduction of a certain metabolite (Palsson, 2006). More details regarding the definition and the

construction of the biomass reaction are given in literature (Thiele and Palsson, 2010; Feist and Palsson, 2010).

If the objective function Z is linear, then an optimal set of fluxes subject to the mass balance $\mathbf{S}\cdot\mathbf{v} = 0$ and linear inequalities $\alpha_i \leq v_i \leq \beta_i$ can be obtained. Under these conditions, the optimization problem is a linear programming problem which can be solved using commercial as well as open-source solvers.

1.3 Advanced techniques in Flux Balance Analysis

In general, the solution obtained by FBA is only as good as the constraints and the defined objective function used to identify it. Nevertheless several simulations carried out in the past, having $Z = \max(BOF)$ as objective, have been shown to be consistent with experimental data (Varma and Palsson, 1994; Edwards *et al.*, 2001). However, different approaches should be explored to characterize the effects of shifts in the steady-state in general.

1.3.1 Incorporation of explicit regulatory constraints

Under some conditions, such as a hostile environment or genetic mutations, the behavior of a microorganism will depart from the maximization of biomass. FBA alone cannot account for this explicitly because it does not implement any rules for network regulation. A first attempt to overcome this deficiency was the implementation of regulatory constraints with a Boolean logic approach (Covert *et al.*, 2001). These constraints arise due to changes in the environment, so that a flux will be temporarily constrained to a certain value based on the initial conditions of the cellular system. A standard FBA is then carried out, maximizing the growth rate. The result is the elimination of many unfeasible pathways from being taken in consideration while performing standard FBA. Namely Boolean rules interdict the use of some cellular pathways which require two or more inconsistent regulatory event to occur simultaneously.

1.3.2 Exploration of alternative classes of objective function

Cellular regulation can be accounted also by choosing a more appropriate objective function. The main difference with the previous approach is that the selection of an optimal objective function follows from the knowledge of the system response to a given event rather than by modeling the response with rules that must be incorporated in the mathematical model.

In this fashion, bacteria subjected to a genetic perturbation (typically a gene knock-out) have been assumed to pursue a metabolic flux distribution similar to the wild-type, rather than to regulate all fluxes toward an optimal growth configuration, as FBA suggests. Thus, the objective function switches from growth maximization to minimization of metabolic adjustment (MOMA), i.e. the mutant remains initially as close as possible to the wild-type optimum in terms of fluxes (Segrè *et al.*, 2002). MOMA can be used to improve the prediction efficiency of FBA for studying *E. coli* mutants.

Alternative classes of objective function can also be investigated to better describe flux distributions even in wild-type species, thus questioning the actual truthfulness of maximization of growth as the best objective function. Schütz *et al.* (2007) for instance systematically compared the prediction efficiency of different BOFs for wild-type *E. coli*, and they concluded that ATP-yield maximization rather than growth maximization better describes the functional state of the cell under several environmental conditions.

1.3.3 Pareto analysis: multi-objective FBA

By assuming a single objective function, FBA reports an optimal flux distribution according to one principle of optimality. For instance, if growth maximization is assumed, the distribution of the fluxes within the cell metabolism is driven by the production of the single metabolites which, in the correct ratio, will form a unit mass of biomass. Even if successful applications of this rationale have produced excellent predictions for some microorganisms (such as *E. coli* in a study by Edwards *et al.*, 2001), it is likely that nature has not a single goal (Sendìn *et al*., 2009). Thus, a more realistic approach would be to consider the simultaneous optimization of two or more objectives. These criteria are often conflicting, leading to a so called Pareto analysis of the network.

The choice of the functions to optimize must reflect the knowledge of the system and should interpret the environmental conditions of the simulation. Sendìn *et al.* for example investigated three common cellular functions for a more realistic prediction of *E. coli* central carbon metabolism. These three functions were maximization of biomass growth, maximization of ATP production and minimization of the overall intracellular fluxes. As Pareto analysis generally leads to a surface (Pareto frontier) over which the weighted sum of the objectives is at optimum, experimental data are necessary to assess which Pareto-optimal points best represent the cell metabolism in the conditions under investigation. It should be noted that experiments must be conducted to perform a complete Pareto analysis, but these experiments could just assess the value of a number

of fluxes equal to number of objectives (if commensurable) minus one (the latter being given automatically by intersecting the Pareto frontier).

1.4 Tools for metabolic engineering

Metabolic engineering in microorganisms has received increasing attention due to its potential for production of renewable fuels and chemicals (Tepper and Shlomi, 2009). Computational modeling of metabolism proved to be useful to guide experimental attempts by anticipating the effect of genetic engineering and metabolism modification. In the context of FBA, where the kinetic techniques have been bypassed by a computational prediction of metabolic fluxes given some constraints such as mass balance, several algorithms have been developed to understand how to detour a flux distribution toward the production of a certain metabolite. Prior to that, it is important to note that FBA itself can give some information over this goal. A simple approach consists in changing the objective function so to maximize the flux that leads to the desired compound. This simulation can help to verify whether such option is feasible, if biomass can still grow, which nutrients are needed in the process and finally what pathway are involved or not.

A more sophisticated attempt for direct metabolic engineering was first performed by Burgard *et al.* (2003), who realized a method, called OptKnock, which searches for sets of gene knockouts that lead to the overproduction of preferred metabolites. More recently, Tepper and Shlomi (2009) improved the method to account for competing pathways, with a novel algorithm called RobustKnock. Other examples of methods for metabolic engineering in the context of FBA are OptStrain (Pharkya *et al.*, 2004), which allows the incorporation of novel enzyme-coding genes not originally present in the species under analysis, and OptReg (Pharkya and Maranas, 2006), that can simulate up- and down-regulation of metabolic enzymes as well.

All these tools can help to identify strategies for metabolic engineering, whether these strategies include the design of an optimal medium or techniques of genetic engineering for gene deletion or regulation. However a deep analysis of the consequences of a certain strategy on the cell metabolism should be carried out to make sure that the result of the simulation does not violate fundamental laws of regulation. Such analysis should be sustained by means of comparison with specific literature and, eventually, by experimental validation.

Chapter 2

Application of systems biology approaches to chemical engineering problems

Systems biology approaches have proved to successfully describe in quantitative terms the metabolism of several microorganisms, which can in turn be exploited to produce valuable chemicals. For instance, lactic acid, ethanol and citric acid are industrially important biotech products which derive directly from bacteria central carbon metabolism (Bro *et al.*, 2005). Engineering of such reactions for a profitable production of both consolidated and new biotech chemicals is a primary aim of a chemical engineer. Interestingly, flux balance analysis and other advanced techniques can provide essential information to this end. Among all, FBA can predict the rate of biomass growth, the input requirements, and the rate of production of certain secondary by-products. Advanced techniques of FBA can either improve such predictions or individuate novel strategies for optimizing the production rate of a certain metabolite over others. In this chapter, it is explored a new and original way to integrate the results obtained by using systems biology approaches such as FBA for solving common chemical engineering problems.

2.1 Bioreactor design

The key requirement for the design of any industrial chemical reactor is the kinetics of the processed reactions (Sinnott, 2005). By understanding the rate at which the reaction proceeds and how it depends on operative or environmental factors, it is possible to provide sufficient residence time for the desired reaction to accomplish the required degree of conversion. Such information is difficult to retrieve in literature, especially those regarding particularly attractive processes. Moreover, experimental campaigns for the acquisition of kinetics data are usually expensive, and not suitable in the stage of conceptual design of the process. In biotechnology, an accurate description of the rate of reaction *per se* can be a hard task, considered that even the simplest model of enzymatic reaction rate, namely the Michaelis-Menten equation, relies on at least two parameters. In addition, even when these values are available, they must be known for each enzymatic reaction taking place in the biological network, but the quantity of these data

is still scarce and the methods to ascertain them are difficult to apply yet (Oberhardt *et al.*, 2009). Thus, a quick and economical way to obtain these parameters is highly desirable.

Here it is proposed a new technique to design a bioreactor by integrating kinetic data retrieved from simulations of the biological system using FBA. This study investigated primarily and ideal tubular reactor (*plug flow reactor* or PFR), but it can be easily extended to ideal completely stirred tank reactor (or Chemostat or CSTR) or other more advanced reactor models.

2.1.1 A classic approach: the biological tubular reactor

In biochemical engineering, a biological reactor differs from an enzymatic reactor because substrates are converted by the biomass into products and biomass itself. So, apart from producing valuable compounds, in a bioreactor a biological system normally grows. The classic approach to deal with such systems is to imagine reactors as black boxes, as depicted in Figure 2.1.

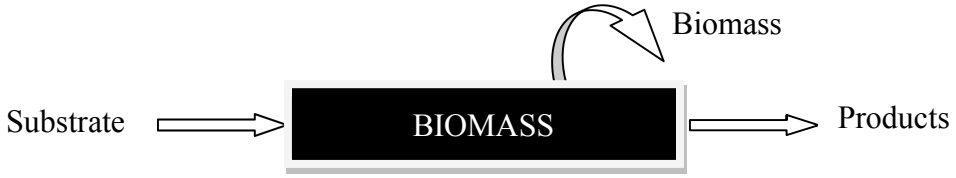

Figure 2.1. *A biological reactive system. The simplest approach to model biological reactors is to assume the black box model. Thus biomass is seen as a vehicle through which substrates are converted to products and other biomass (growth).*

The biomass is modeled as a reactor which converts substrates into products and biomass. In principle, there is a reaction rate for each substrate converted and product formed (including biomass). However, usually only the biomass producing reaction is modeled in detail. According to this approach, the growth is described taking advantage of some well known equations, like the Malthus equation (2.1) or the Monod equation (2.2):

$$r_X = \mu C_X \tag{2.1}$$

$$r_X = \frac{\mu C_S C_X}{k_M + C_S} \tag{2.2}$$

These two models describe quantitatively the change of biomass provided some parameters are known (μ, the maximum rate of growth, and k_M, the Monod constant) and some external conditions (such as C_S and C_X, the concentration of limiting substrate

and of biomass respectively): in this fashion, r_X is the rate of biomass formation or, in other terms, the reaction rate of the growth reaction. Once the rate of biomass formation is characterized, all other reactions, whether they represent the conversion of a substrate or the formation of a product, are related to it in terms of yield. In biological reactive systems, yield is commonly defined both for substrates and products as:

$$Y_{\frac{X}{S}} = -\frac{dC_X}{dC_S} \tag{2.3}$$

$$Y_{\frac{P}{S}} = -\frac{dC_P}{dC_S} \tag{2.4}$$

The yield describes the variation in biomass or product concentration given some variation in substrate concentration. They are defined as stated by equation (2.3) and (2.4) because their value is simple to determine experimentally. If a trend for the yield can be traced, then the reaction rate for substrate conversion or product formation is given, respectively, by:

$$r_S = -\frac{1}{Y_{\frac{X}{S}}} r_X \tag{2.5}$$

$$r_P = \frac{Y_{\frac{P}{S}}}{Y_{\frac{X}{S}}} r_X \tag{2.6}$$

Once all reaction rates have been established, mass balances can be written around the reactor. By solving the mass balances, the residence time can then be assessed and all values necessary for a preliminary design of the reactor are available.

The form of the mass balances depends on the reactor configuration. For tubular reactors, mass balances are differential. An equation must be written for each component of the reaction. In a general biological system with one substrate, one product and biomass, under steady state operation, the system of mass balances is given by (2.7):

$$\begin{cases} \frac{dC_X}{d\tau} = r_X \\ \frac{dC_S}{d\tau} = r_S \\ \frac{dC_P}{d\tau} = r_P \end{cases} \tag{2.7}$$

where τ is called residence time. It is defined as the ratio of reactor volume V over the average incoming volumetric flow rate $\dot{V}$. The residence time represents the time interval needed to fill the reaction volume at the given influent volumetric flow rate. In reactor design, given a certain productivity, which is proportionally related to the volumetric flow rate, once the residence time is found then the volume for the reactor is

straightforward. To obtain the residence time, all equations in (2.7) must be simultaneously satisfied. To solve the system of equations, the boundary conditions for each variable must be given and an arbitrary value for product conversion or biomass growth must be chosen. Only three variables are left unknown, namely the residence time and the outlet concentration of substrate and biomass (or product, depending on the previous choice). In conceptual design, usually biomass/product conversion is left parametric with the residence time.

In conclusion, the key parameter for bioreactor design is the residence time. The residence time is obtained by solving simultaneously a mass balance equation for each component in the reactive system. The solution of this system of equations exists only if a representative form for each reaction rate is available, if all parameters values are known and if an initial set of conditions on the components' concentration is given.

2.1.2 A new approach: solving mass balance equations using FBA

The classic approach suffers of several time (and money) consuming hurdles. Most of these hurdles involve the reaction kinetics description. The approach here proposed consists in substituting the black box with a genome-based metabolic reconstruction (GENRE) of the biological reactive system, as depicted in Figure 2.2. A GENRE recollects all the metabolic reactions occurring within the cell and it includes exchange reactions for nutrients uptake and secondary products secretion too.

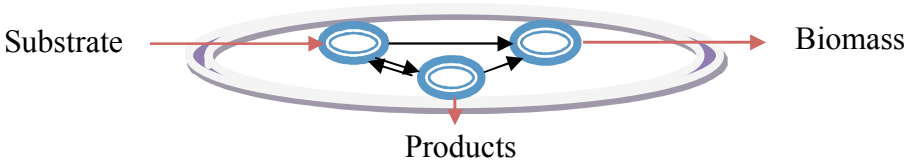

Figure 2.2. *A biological reactive system. A new approach to model biological reactors is to integrate its metabolic network. All fluxes, whether external or internal, are estimated using FBA, choosing an appropriate objective function, like biomass growth.*

To estimate numerically the reaction rates of all the reactions included in a GENRE, FBA is applied. In the context of bioreactor design, internal reaction rates are irrelevant but the exchange reaction rates are essential. They actually represent the rates at which a nutrient is absorbed by the cell or a metabolite is secreted. Moreover, all GENREs provide a fictitious reaction for biomass growth, namely the biomass producing reaction rate or biomass objective function (BOF). Actually, FBA usually estimates all reaction rates (or fluxes, as more commonly found in literature) based on optimizing the biomass producing reaction (see §1.2.4 for further details). Once all the fluxes, with particular emphasis on the external fluxes, have been computed, the system of equation (2.7) can

be solved by inserting the value of the flux in place of the proper reaction rate expression. However this step requires some care. As it will be seen in Chapter 3 (§3.1.1.3), fluxes are expressed in terms of mmol g_{DW}^{-1} h^{-1}. In standard chemical reaction engineering, reaction rates are given in mol L^{-1} s^{-1}. Thus the conversion factor from fluxes to reaction rates is the biomass concentration C_X:

$$r_i = v_i C_X \tag{2.8}$$

where v_i is the *i*-th flux (internal or external) calculated via FBA and r_i is the corresponding reaction rate for the material balance. All quantities are expressed at the residence time τ. The equations in the system (2.7) can be solved either numerically or analytically. If FBA was applied consistently, that is that the objective function has been carefully selected, constraints have been set correctly and the metabolic reconstruction is fairly complete, then in principle the flux distribution should not depend from the residence time. Moreover, it should not depend on the concentrations themselves. However, the biomass concentration does depend on the residence time, and it appears in each differential equation since it is the conversion factor according to eq. (2.8). The system of equations (2.7) can be rewritten as:

$$\begin{cases} \frac{dC_X}{d\tau} = v_X C_X \\ \frac{dC_S}{d\tau} = v_S C_X \\ \frac{dC_P}{d\tau} = v_P C_X \end{cases} \tag{2.9}$$

To find a solution to this problem, a method called *dynamic FBA* (dFBA) can be applied. This technique, which was first introduced by Varma and Palsson in 1994, solves all the equations simultaneously at every instant $d\tau$. In this infinitesimal interval, fluxes calculated in FBA are assumed to be constant and equations are solved analytically. In the subsequent instant $d\tau$, C_{X0} is updated to the last calculated value. In addition, initial conditions for the rest of the equations are updated to the last calculated value. The procedure is iterated until a certain conversion of either the product or the biomass is achieved or, alternatively, after an arbitrary number of iterations. The sum of all infinitesimal intervals dτ produces the overall residence time τ. As said before, at every infinitesimal interval, each differential equation of (2.9) are solved analytically, since fluxes are constant. Thus, at every $d\tau$:

$$\begin{cases} C_X = C_{X0} e^{v_X d\tau} \\ C_S = C_{S0} + v_S C_{X0} e^{v_X d\tau} d\tau \\ C_P = C_{P0} + v_P C_{X0} e^{v_X d\tau} d\tau \end{cases} \tag{2.10}$$

where C_{X0}, C_{S0}, C_{P0} are the last calculated concentration values. By stating that, it is implied that at the first iteration the initial conditions for every concentration must be given. Note that the value of $d\tau$ is arbitrary, but in order to achieve a realistic simulation, it should be significantly smaller than the relaxation time of the system. The iterative procedure that performs FBA at each step by progressively updating the initial values of a set of concentration, generally inputs and outputs, is the core of dFBA.

This method has been applied with some modifications for a case study, the production of triacylglycerols by *Chlamydomonas reinhardtii*. The fully detailed example is developed in Chapter 4.

2.1.2.1 dFBA and numerical methods comparison

dFBA is not a numerical scheme, since even though the system is discretized in infinitesimal instants $d\tau$, all equations of system (2.9) are solved analytically. In principle, since fluxes are given as numerical values by FBA, a method for numerical integration of ordinary differential equations, such as explicit Euler method, is more suitable. In this fashion, the time derivative in all equations of (2.9) is approximated by a finite difference $d\tau$, which leads to:

$$\begin{cases} C_{X_{n+1}} = C_{X_n} + d\tau(\ {}_{X_n}C_{X_n}) \\ C_{S_{n+1}} = C_{S_n} + d\tau(v_{S_n}C_{X_n}) \\ C_{P_{n+1}} = C_{P_n} + d\tau(v_{P_n}C_{X_n}) \end{cases} \tag{2.11}$$

where n is the iteration index. Clearly, at $n = 0$, boundary conditions must be given as dFBA. At the nth iteration, FBA must be run to retrieve the values of the fluxes (v_{i_n}) needed to solve equations in (2.11). In Appendix D, it is shown qualitatively that results are not sensibly affected by the choice of the method (whether dFBA or numerical integration).

2.1.3 Advantages

Besides the obvious pros of this new approach, which bypasses the need of collecting experimental data for the growth reaction kinetics and the product/substrate yield value, several other benefits should be accounted for.

First, several environmental conditions can be simulated. The advantages are enormous. In the assumption that only environments where the biological reactive system under study can grow may be simulated with realistic predictions, then this new approach can provide a quick quantitative estimation of the growth rate variation, as well as of all the other external fluxes necessary to compute the system of equations (2.9). If the black

box approach was used, both parameters in the Monod equation (2.2) and the yields (2.3) and (2.4) should be experimentally estimated for each substantially different environmental condition. On the other hand, FBA can switch medium conditions by simply adjusting the constraints on the external fluxes. For example, to switch from aerobic to anaerobic, it is sufficient to set to zero the constraint for the availability of oxygen in the environment, as described in §1.2.3. Same can be said for the availability of organic carbon (heterotrophy) or inorganic carbon (autotrophy). In general, both growth rate and substrate absorption are significantly influenced by the environment, so this approach can help to distinguish more favorable conditions for the bioreactor goal, by taking in account at the same time which nutrients are needed and in which quantity.

In the second place, genetic engineering strategies, such as gene knock-out or up-down regulation, can be tested. These techniques represent a huge possibility for reactor optimization that only biotech applications can take advantage of. FBA, or other advanced algorithms such as MOMA or ROOM (Shlomi *et al.*, 2005), can be employed to calculate the flux distribution after a gene perturbation has been applied. These results may be used to update the solution of the system of equations (2.9) at every step of iteration. However, it should be noted that predictability of the above methods is still poor, even though they can provide useful information for conceptual design without carrying out any experiment (Zhao and Kurata, 2005).

2.1.4 Limitations

Even though the inherent flexibility of this new approach is appealing, result predictions should be used with care. In particular, it should be noted that GENREs lack a reconstruction of the regulatory network. Thus, even unfeasible growing conditions for the biological reactive system can produce results in the simulation. Moreover, as regards genetic manipulation, it is not guaranteed that a certain gene knock-out or regulation is practicable, as cell adaptation is not accounted for and some manipulations can lead to cell death. Because of this, experiments are essential to confirm model predictions. Therefore, it is recommended to extend simulations within the framework of conditions that are known to be viable either from literature or experimental evidence. In particular, it is suggested to validate experimentally those simulations which turned to be considerably attractive. It should be noted that even by applying this protocol the advantages of this new approach are still evident, as it can help to select a few smart alternatives among all other reasonable options.

2.2. Biochemical reaction engineering

Traditional chemical engineering employs catalysts to enhance a reaction rate. In biological reactive systems, the catalysts are the enzymes, which can increase a kinetic constant as up as 10^{15} times. Thus, areas of further optimization must go beyond the catalysts themselves and investigate a strategy to optimally take advantage of those enzymes that carry out a desired reaction. Two methods have been studied, computational metabolic engineering and design of *ad hoc* medium.

2.2.1 Computational metabolic engineering

Systematic methods to identify gene manipulations that lead to a metabolite overproduction within a biological reactive system are possible thanks to specific algorithms that work in the GENREs environment (see §1.4). Due to the inherent difficulty to predict gene essentiality without a valuable regulatory network, results should be validated experimentally.

A possible strategy to achieve some useful indications for metabolic engineering, is to systematically test a single gene knock-out (KO) in the GENRE by optimizing biomass growth and a by-product synthesis simultaneously. To perform this, a freely available algorithm which works in the MATLAB™ framework is OptKnock (Burgard *et al.*, 2003). If a gene-protein-reaction association map is available within the GENRE, OptKnock deletes a single gene at the time, by constraining its upper and lower bound to zero. Then, it solves a mixed-integer linear problem which calls for a flux distribution which maximizes a target flux, provided that the biomass objective function is at optimum. The target flux usually represents a fictitious exchange reaction which secretes the desired metabolite. OptKnock verifies the range in which the target flux can span, that is the minimum and maximum rate at which the metabolite can be secreted.

In results analysis, the KO-gene set which proved to be unable to sustain growth is usually defined as gene essentiality test. Obviously, by knocking-out any of these genes, the cell will die, at least according to the simulation. To develop a genetic engineering strategy, it should be kept in mind that two situations may occur. First, a simulated gene KO which induces the target flux to span from zero to any upper bound does not guarantee any secretion or overproduction of the target metabolite. In other terms, the target flux may turn to be null in experiments. To improve the prediction accuracy, RobustKnock (Tepper and Shlomi, 2010) accounts for the presence of competing pathways in the network and predicts a minimal, guaranteed production rate of the target chemical. Secondly, many KOs are redundant, as they affect reactions belonging to the same pathway, thus leading to equal consequences on the flux distribution.

Therefore, it is recommended to analyze the GENRE in order to evaluate the pathways affected by the KOs, rather than the reactions, so to reduce notably the size of KOs strategies that may prove successful.

2.2.2 Design of ad hoc *media*

The metabolic flux distribution within a cell is sensitive to the environment. For the purposes of chemical reaction engineering, this can significantly change the rates of many internal fluxes. For instance, some microorganisms' ability to grow is greatly enhanced in heterotrophic conditions rather than in autotrophic, or viceversa. The presence or absence of a certain nutrient from the membrane surroundings can determine alteration of some cellular functions and consequently modify microorganisms' behavior and viability. In computational synthetic biology, this fact has been used to design artificial media capable of sustaining growth for a microbial community but unable to do the same for one or all individual species (Klitgord and Segrè, 2010).

Given this premise, it might be possible to design *in silico* a medium which can boost a certain cellular function over others. A simple and novel method has been developed to check whether a single biological system, such as a cell, can enhance a target flux provided that a certain compound is present in the environment. The rationale beyond this technique consists in assuming that if the cell is given the possibility to soak up a metabolite rather than synthesizing it, then the corresponding anabolic pathway may be bypassed. By doing so, the flux distribution could be detoured to boost the previously chosen target flux. For example, if the biomass producing reaction is selected as target flux, pyrimidines (uracil, UTP, and cytosine, CTP), which are essential precursors for biomass, may be supplemented by the environment. If so, the pyrimidine biosynthesis pathway can be inactivated, implying a better exploitation of nutrients and a minor cofactor usage. As a consequence, biomass may result to grow faster (Figure 2.3).

The algorithm searches for all metabolites naturally present in the metabolic network. At each iteration step, it assumes that one of these chemicals is individually added in the growth minimal medium. It further assumes that the cell is capable of absorbing it, by including a fictitious exchange reaction in the GENRE. When the simulation concludes, it is possible to evaluate which artificially added compounds increased the target flux, and the needed quantity. Since the algorithm does not check whether the cell itself can effectively absorb the artificially added compound, evidences in literature or by means of experiments are essential to further prove applicability of the simulation results according to this idea.

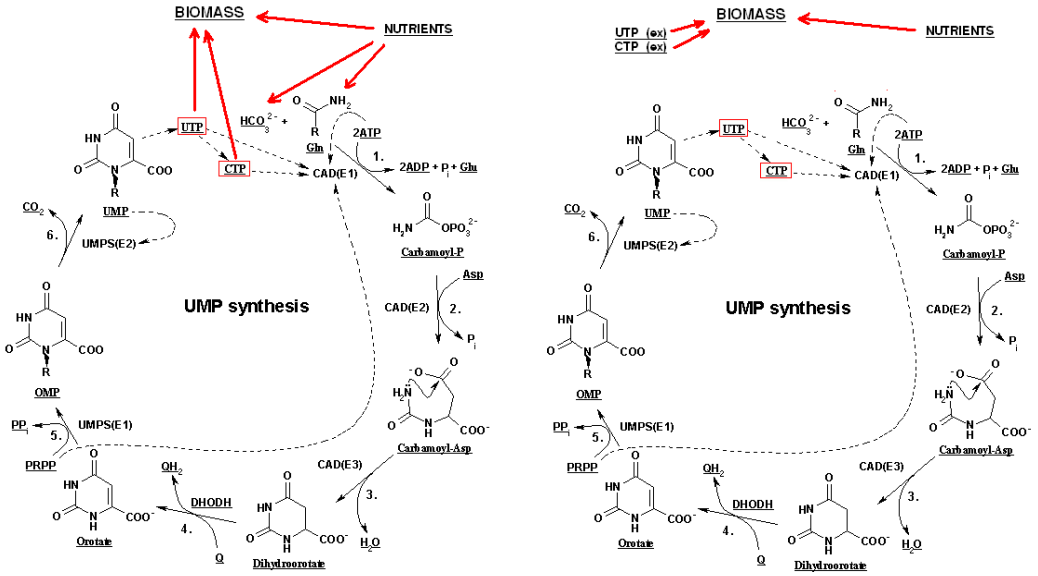

Figure 2.3. *Pyrimidine biosynthesis pathway. Pyrimidines, UTP and CTP (in red boxes), are essential to grow biomass. On the left, a cell standard behavior to produce biomass is depicted. Nutrients are converted via some pathways (red arrows) to biomass but part of them enters the pyrimidine synthesis cycle with extensive cofactor usage. On the right, UTP and CTP are provided by the environment. In this fashion, nutrients will bypass the pyrimidine cycle, possibly enhancing biomass growth.*

2.3 Conclusions

This chapter highlighted the advantages that follow from integrating systems biology approaches with chemical engineering. Common problems of biochemical engineering, such as bioreactor design and chemical reaction engineering can greatly benefit from the whole-istic perspective of systems biology. The reconstruction of genome-based metabolic networks (GENRE) and the emergence of flux balance analysis both contributed at quantitatively define the reactions taking place in a modelled cell. These data can be successfully applied to bypass the need to collect kinetics data in the stage of bioreactor conceptual design and to develop new strategies to engineer reactions of interest without turn to traditional catalysis.

Chapter 3

A case study: Flux Balance Analysis of *Chlamydomonas reinhardtii*

Among the greatest challenges of modern technology, the research for profitable and renewable fuels has become a major interest, urged by the instability of petroleum fuel costs, the vicinity of peak oil, and the increasing danger due to actual atmospheric CO_2 concentration levels (Radakovits *et al.*, 2010a). Since the 1980s, microalgae have appeared to be an attractive solution for their ability to grow rapidly and accumulate high levels of triacylglycerols (TAGs), which can be processed as feedstock for biofuel production (Hu *et al.*, 2008). The advantages are numerous, such as the substantial amount of TAGs which may be produced (up to 50-60% of dry biomass), elevated growth rates, easy adaptation to harsh environments, such as waste or salt water. As a consequence, they do not affect fresh-water resources, thus they do not compete with other food aimed cultivations. Moreover, if grown autotrophically, they consume carbon dioxide to sustain photosynthesis, so they contribute to greenhouse effect mitigation.

From this chapter on, systems biology approaches introduced in Chapter 1 will be investigated to solve some chemical engineering problems regarding the bioproduction of TAGs from a microalgal species. In this chapter, an overview of the GENRE used in the subsequent simulations will be given in detail. For the purposes of this case study, a reconstructed metabolic network of an oil producing microalgae was necessary. So far, *Chlamydomonas reinhardtii* is the only one whose genome has been mapped (Merchant *et al.*, 2007) and thus whose metabolic network has been reconstructed and refined (Boyle and Morgan, 2009, Chang *et al.*, 2011). It is proved that some strains of *C. reinhardtii* do accumulate TAGs so that they may serve as possible feedstock for biofuel production (Siaut *et al.*, 2011, Work *et al.*, 2010, Wang *et al.*, 2009, Li *et al.*, 2010, James *et al.*, 2010).

3.1. The *Chlamydomonas reinhardtii* model

A genome-based metabolic reconstruction of *Chlamydomonas reinhardtii*, called iRC1080, has been recently published (Chang *et al.*, 2011). It comprises 1080 genes,

associated with 2190 reactions and 1068 unique metabolites distributed across 10 compartments. The model is among the largest metabolic models available to date and include the most extensive description of lipid metabolism. It represents a major advance with respect to the previously available model for *C. reinhardtii*, which included 523 reactions and 470 metabolites (Boyle and Morgan, 2009). Metabolic pathways included in the model are presented in Table 3.1*a*. Other reaction classes, which do not actually account for a metabolic activity but are necessary for the complete description of the network (such as exchange reactions or the biomass objective function) are also listed in Table 3.1*b*. A pictorial representation of the network has been generated with iPath 2.0 (Yamada *et al.*, 2011), which is based on a KeGG metabolic map (http://www.genome.jp/kegg/). KeGG is a bioinformatics resource which contains, among others, a collection of manually drawn pathway maps. The generated map provides a fair description of the reconstructed metabolism of *C. reinhardtii* (Figure 3.1).

Table 3.1a. *Metabolic pathways accounted in iRC1080 model.*

Metabolic pathways	
Amino acids biosynthesis, metabolism, degradation	Methane metabolism
Amino sugar and nucleotide sugar metabolism	N-Glycan and O-Glycan biosynthesis
Ascorbate and aldarate metabolism	Nicotinate and nicotinamide metabolism
Biosynthesis of steroids	Nitrogen metabolism
Biosynthesis of unsaturated fatty acids	Oxidative phosphorylation
Butanoate metabolism	Pantothenate and CoA biosynthesis
Carbon fixation	Pentose phosphate pathway
Carotenoid biosynthesis	Photosynthesis
Cofactor recycling	Polyamine metabolism
Fatty acid biosynthesis	Porphyrin and chlorophyll metabolism
Fatty acid elongation in mitochondria	Propanoate metabolism
Fatty acid metabolism	Protein synthesis
Folate biosynthesis and one carbon pool by folate	Purine and pyrimidine metabolism
Fructose and mannose metabolism	Pyruvate metabolism
Galactose metabolism	Retinol metabolism
Glutamate metabolism	Riboflavin metabolism
Glutathione metabolism	Selenoamino acid metabolism
Lipid metabolism	Sucrose metabolism
Glycolysis / Gluconeogenesis	Starch metabolism
Glyoxylate metabolism	Sulfur metabolism
High-mannose type N-glycan biosynthesis	TCA cycle
Inositol phosphate metabolism	Terpenoid backbone biosynthesis
Linoleic acid metabolism	Urea degradation

Table 3.1*b*. *Subsystems accounted in iRC1080 model.*

Other subsystems
ATP maintenance
Biomass formation
Exchange/Demand
Spectral decomposition
Transport

3.1.1 Understanding the iRC1080 model

The iRC1080 model is a mathematical model which includes some basic features of all GENREs, such as the stoichiometric matrix of the metabolic network, and some advanced features, such as the matrix for gene-protein-reaction (GPR) associations and a comprehensive description of non-metabolic reactions, as those in the subsystems of Table 3.1b. In order to explore all the characteristics of this model, a general overview is provided in the following sections.

3.1.1.1. Network architecture

As described with more detail in §1.2.2, the core of a GENRE is the stoichiometric matrix **S**, in which each column represents a reaction and each row represents a compound. The stoichiometric matrix is a connectivity matrix, and it corresponds to a network (Palsson, 2006). To understand this, each row in the matrix should be thought as a node of the network. Namely, a certain compound can be produced or consumed by one or more reactions, and the type and number of reactions are given by the non-null entries in the corresponding row. On the other hand, each column should be thought as a link of the network. That is, a certain reaction relates two or more compounds, and the type and number of compounds are given by the non-null entries in the corresponding column.

Each column identifies univocally a reaction, and a comprehensive list of all reactions and their corresponding entry in the matrix is always provided in any GENREs, included iRC1080. It should be noted that *all the possible* reactions are incorporated in the matrix **S**. So, even non-metabolic reactions, such as exchange reactions, as well as fictitious reactions, such as the biomass objective function, are part of the matrix. Usually these reactions employ an unconventional stoichiometry to better describe their functions. A more detailed description of these reactions is given in §3.1.1.3.

Since *C. reinhardtii* is an eukaryote, the cell is highly compartmentalized. As a result, many metabolites can be present in more than one compartment at the same time. They can move in or out of a compartment by intracellular transport, or be secreted or adsorbed from the extracellular matrix by extracellular transport. Thus, in the model, to distinguish the same metabolite from a compartment to another an appropriate suffix is added to the metabolite formula (e.g. h2o[h] stands for water in chloroplast). A comprehensive list of suffixes used in the model is provided in Appendix B. The main consequence in the network modeling is that a same metabolite can be actually be associated to more than one row in the matrix **S**, depending on the compartment at which belongs. So, even if only 1068 unique metabolites are present in the reconstructed network, the actual number of rows of the stoichiometric matrix is 1707.

In conclusion, *C. reinhardtii* metabolic network can be synthesized in terms of the stoichiometric matrix of the reactions which participate in its reconstructed metabolism. Given that the reconstruction comprises 1707 metabolites distributed in ten compartments and linked by 2190 reactions, the dimensions of the matrix ***S*** is therefore 1707x2190.

3.1.1.2 Gene-protein-reaction associations matrix

While many current GENREs must simulate gene deletions by constraining the flux of a given reaction to zero, a more correct way to proceed is to first identify the existing correlation between genes and reactions. This is preferable because the correlation between a reaction and a gene may not be linear, i.e. it is not obvious that there is a single gene encoding a single enzyme which carries out a single exclusive reaction. The maps which relate encoding genes to one or multiple reactions are called gene-protein-reaction (GPR) associations. Examples of different types of GPR associations are shown in Figure 3.2. At the top level is the gene locus, at the second level is the translated peptide, at the third level the functional protein, and at the bottom level is the reaction. For instance, succinate dehydrogenase (*Sdh*) is an example of promiscuous enzyme, which means that it carries out more than one reaction, namely two. Four gene locus encode four different peptides, and all these gene products are needed for the functional enzyme *Sdh*. *Sdh*, in turns, carries out two different reactions: thus a simulation of gene deletion in which only one of the two reaction fluxes is constrained to zero omits to take in account that also the second reaction flux is unfeasible. D-Xylose ABC Transporter is an example of reaction which is carried out by a protein complex. This complex is formed by subunits each individually encoded by a different gene. Finally, glyceraldehyde 3-phosphate dehydrogenase is an example of a reaction which can be

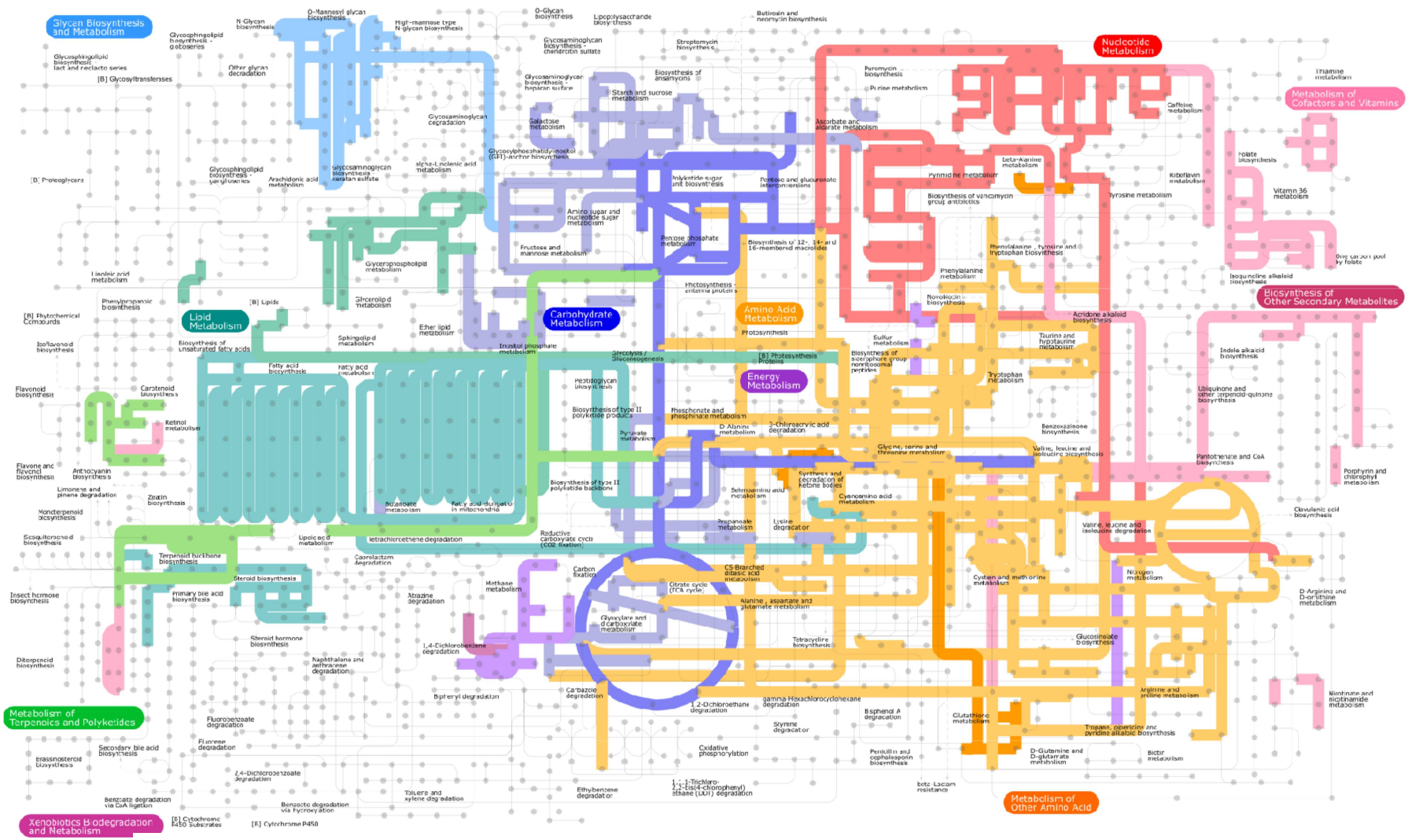

Figure 3.1. *A map generated with iPath2 representing the metabolic network reconstructed within the iRC1080 model, based on a KeGG pathway map. Colored edges correspond to reconstructed pathways. Each color is associated to one or more pathways linked to a specific cellular function (for example, violet stands for energy metabolism).*

carried out by two independent proteins (*GapA* and *GapC*). *GapA* and *GapC* are thus isoenzymes and they are encoded starting from different gene loci.

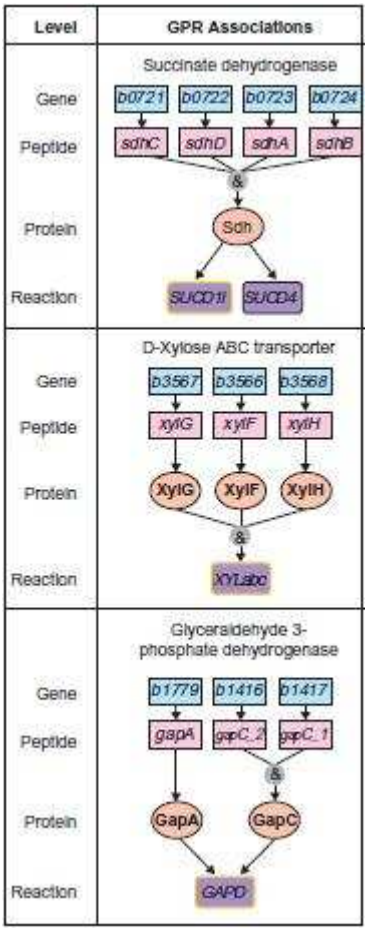

Figure 3.2. *Examples of GPR associations. On top succinate dehydrogenase, which displays a so called promiscuous enzyme. On the middle, D-Xylose ABC Transporter, which is carried out by a protein complex. On the bottom, glyceraldehyde 3-phosphate dehydrogenase, which is carried out by two different isoenzymes. Retrieved from Reed et al., 2003.*

GPR associations are provided in the iRC1080 model in terms of a GPR matrix, in which each row corresponds to a gene locus, while each column represents a reaction. Thus in order to perform a gene deletion simulation, the procedure will rather take into account all the consequences in the network by using the GPR matrix as guidance.

3.1.1.3 Non metabolic reactions

Non-metabolic reactions can be divided in four categories: transport reactions, exchange reactions, prism reactions and the objective function. The latter will be discussed with more detail in §3.1.1.4.

Transport reactions are available for those metabolites which are known to move from a certain compartment to another. The stoichiometry of these reactions is always 1:1, reflecting the fact that no actual reaction is occurring. To successfully describe the transport of metabolite *X* from compartment [*a*] to compartment [*b*], the transport reaction is modeled as:

$$X[a] \rightarrow X[b] \tag{3.1}$$

This type of transport reflects the physics of passive transport. To simulate active transport other metabolites can be part of the reaction (3.1), depending on whether the active transport is primary (i.e., it employs chemical energy in form of ATP) or secondary (i.e., it exploits electrochemical gradients induced by other metabolites).

Exchange reactions simulate metabolite uptake and secretion from/to the extracellular matrix. The former metabolites are usually the substrates (i.e. nutrients) while the latter are known cellular secondary products (for example, acetate). These reactions are inherently unbalanced. The uptake/secretion of metabolite *X* from the extracellular environment is modeled as:

$$X[e] \rightarrow \tag{3.2}$$

If the flux of reaction (3.2) is positive then the metabolite is secreted by the cell. Vice versa, the metabolite is absorbed. Exchange reactions are essential to give physical dimensions to the fluxes, as these values are usually easy to obtain experimentally. Generally, fluxes are expressed in terms of mmol $g_{DW}^{-1}\, h^{-1}$. Given an objective function to optimize, say growth, which is known to reproduce the actual behavior of the cell, usually just one or few metabolites are limiting to achieve that optimum state. For example glucose uptake is the limiting step for *E. coli* growth on aerobic rich medium. Thus, by constraining the uptake flux of the limiting nutrient to assume an experimentally determined value, quantitative results for growth flux can be produced by the model (Varma and Palsson, 1994).

Prism reactions are a novel way of describing photons acquisition by the cell, introduced in the iRC1080 model for the first time. They can be thus seen as a subcategory of exchange reactions. Spectral decomposition has been studied for several light sources, taking in account both effective light spectral ranges and biomass yield on light. In this fashion, it is possible to set the photon flux of a specific light source, in terms of $\mu mol\ m^{-2}\ s^{-1}$. Successively, the actual absorption of a particular range of photons wavelength is implemented in the model in terms of stoichiometric coefficients, which vary according to assumed conversion efficiency. Finally, only photons having

an active interval of wavelengths can carry out light-driven reactions, whereas all the other ranges are reflected. More details on the derivation of prism reactions are given in Supplemental Text 1, taken from Chang *et al.*, 2011. An example of prism reaction is equation (3.3) which represents the spectral decomposition of solar light (lithosphere).

$$\begin{aligned} \text{photonVis[e]} \rightarrow\ & + 0.000043\ \text{photon298[c]} + 0.088253\ \text{photon437[u]} \\ & + 0.187391\ \text{photon438[u]} + 0.114958\ \text{photon450[h]} \\ & + 0.195815\ \text{photon490[s]} + 0.185664\ \text{photon646[h]} \\ & + 0.083616\ \text{photon673[u]} + 0.094405\ \text{photon680[u]} \end{aligned} \tag{3.3}$$

Equation (3.3) states that each unit flux of photons entering the system is decomposed along the wavelength bandwidth spanning from 298 nm to 680 nm. The actual absorption for each class of photons in terms of flux is equal to its stoichiometric coefficient. Note that whether each class of photons will display in reality a non-zero flux depends on the simulation constraints. It is expected that photons in the violet/blue and red wavelength will always be used in growth on light, as they drive reactions of photosystems complex I & II.

3.1.1.4 The objective function

The objective function is a key requirement in FBA as it states what the microorganism is trying to do in a given environment. An appropriate objective function is thus fundamental to predict a functional state of metabolism which is physiologically realistic (Feist and Palsson, 2010). The cumulative data over time have suggested that in nutritionally rich and nutritionally scarce environments optimal biomass yield or growth rates are meaningful objective functions. It is assumed that a maximal growth rate phenotype is likely to occur after adaptive evolution or prolonged laboratory experimentation. Thus, as indicated by Chang *et al.* in their paper on iRC1080, a biomass objective function is suitable even for the case of *C. reinhardtii*. However, this rationale might not be realistic in elementally limited environments (Feist and Palsson, 2010).

Three objective functions are available in the iRC1080 model to account for autotrophic, heterotrophic and mixotrophic growth conditions. The biomass formation equations were derived according to previously reported methods (Chavali *et al*, 2008; Foster *et al*, 2003). These equations encompass all precursors necessary for *C. reinhardtii* to grow in a given condition. The classes of precursors included in the BOF are proteins, DNA, RNA, carbohydrates, fatty acids, glycerol, lipids, chlorophylls, and

xanthophylls. Stoichiometric coefficients for each of these metabolites are estimated based on their proportion in a unit of dry weight biomass. The coefficients are then normalized to result in a mass unit of biomass. The complete list of precursors and coefficients for the three biomass equations is provided in Supplementary Table S9, Chang *et al.*, 2011.

The biomass formation equation also incorporates a model-based value of growth associated ATP maintenance (Boyle and Morgan, 2009). This value, abbreviated as GAM, accounts for the energy necessary to replicate a cell, in the form of an ATP hydrolysis reaction: 1 ATP + 1 $H_2O \rightarrow$ 1 ADP + 1 P_i + 1H^+ (Thiele and Palsson, 2010). Growth associated ATP maintenance was estimated to be of 29.89 $mmol_{ATP}$ g_{DW}^{-1}. In contrast with GAM, non-growth associated ATP maintenance, or NGAM, is the supposed energy (in the form of ATP hydrolysis reaction) necessary to maintain a cell.

3.1.1.5 Constraints

The constraints inherently incorporated in iRC1080 can be distinguished in experimentally derived constraints and in thermodynamic constraints. Experimentally derived constraints include the above mentioned (§3.1.1.4) values for GAM and NGAM, effective incident photon flux values for all reported light sources, maximum starch degradation rate in both dark/light environment, and maximum nutrient uptake. A complete list of these values is available in Supplementary Table S6, Chang *et al.*, 2011. Notably, the model includes an experimental validation for maximum CO_2 uptake rate (-11.16 mmol g_{DW}^{-1} h^{-1}) and maximum O_2 photo-evolution rate (8.28 mmol g_{DW}^{-1} h^{-1}), that is oxygen produced via photosynthesis but not used for any metabolic functions within the cell, which diffuses out of the cell. On the other hand, thermodynamic constraints state whether a reaction is irreversible or not. This is based on biochemical knowledge rather than by a systematic evaluation of the thermodynamic state of the cell. In the model, if a reaction is irreversible, either the upper or the lower bound is set to zero (depending on the reaction direction).

A proper implementation of the constraints is essential to provide high quality predictions. In the following section, it is addressed the question on how the constraints can affect and modify the model for successive simulations.

3.1.2 Setting up the iRC1080 model

In order to perform FBA, the model must be set up properly, according to the simulation objectives. As regards iRC1080 model, three standard conditions can be simulated, autotrophic, heterotrophic and mixotrophic. Unless stated, all simulations carried out in

this work concern autotrophic conditions. If a simulation required some modifications in the autotrophic standard conditions, it will be specified in the input list. Medium conditions, specific cell constraints and the objective function are three important issues to address while setting up any genome-based metabolic network model.

3.1.2.1 Medium conditions

In FBA, specific availability of a certain nutrient in the growth medium is imposed through constraints. As these compounds are external with respect to the cell, in order to be incorporated in the metabolic network, an exchange reaction (see §3.1.1.2) must be added. A lower or upper bound on that reaction must then be decided upon the knowledge possessed about the medium and the organism capabilities. According to the fact that nutrients are usually absorbed by the cell, in order to obtain a positive flux in the exchange reaction the nutrient is commonly implemented as a product. However in the iRC1080 model all exchange reactions have the nutrients as a substrate, thus a negative flux actually means uptake of the compound from the medium. By setting the lower bound to be minus infinite, the model allows unlimited availability of the nutrient in terms of flux, i.e. it can be incorporated in the metabolic network in any amount. On the other hand, if the lower bound is set to zero, the nutrient is absent from the medium, i.e. it cannot be incorporated in the metabolic network at all.

Essential compounds for growth comprise a carbon source, a nitrogen source, a phosphate source, a sulfur source, oxygen, water and metals. As regards *C. reinhardtii*, the carbon source can be carbon dioxide (CO_2) in autotrophic conditions, acetate (Ac, $CH_3CO_2^-$) in heterotrophic or both of them in mixotrophic ones. Ammonia (NH_3) and nitrates (NO^{-3}) are popular nitrogen sources, inorganic phosphate (Pi, PO_3^{-4}) provides the phosphate source, while sulfur is supplemented via sulfate (SO_2^{-4}). Among the metals accounted in the model for cell growth are magnesium (Mg^{2+}), iron (Fe^{2+}, Fe^{3+}) and sodium (Na^+). If all nutrients are present in the medium, it is referred to as *rich medium*. Experimental data are usually preferable to predict the maximum amount of a nutrient that can be absorbed in a certain time by the cell, i.e. the exchange flux. Chang *et al.* suggested the values in Table 3.2 as lower bounds for the model, and these values are default in iRC1080 (as in §3.1.1.5). Again, it is important to stress that depending on the magnitude of these values, physical dimensions can be assigned to fluxes.

A fine *in silico* description of available nutrients in the medium is crucial to reproduce or compare experimental data correctly. As regards *C. reinhardtii*, most experiments are carried in so called TAP medium (Gorman and Levine, 1965) or HS medium (Bonnel and Raby, 1958), for heterotrophic and autotrophic conditions respectively. The main

difference is the presence of acetate in the former, and its absence in the latter. By fixing constraints as in Table 3.2, either a TAP medium or a HS medium can be fairly reproduced.

Table 3.2. *Maximum uptake rate of nutrients in iRC1080.*

Exchange reaction	Reaction in iRC1080	Lower bound [mmol gDW^{-1} h^{-1}]	Notes
Maximum proton uptake rate	EX_h(e)	-10	
Maximum water uptake rate	EX_h2o(e)	-10	
Maximum phosphate uptake rate	EX_pi(e)	-10	
Maximum ammonia uptake rate	EX_nh4(e)	-10	Either ammonia or nitrate are set non-zero at each time
Maximum nitrate uptake rate	EX_no3(e)	-10	
Maximum sulfate uptake rate	EX_so4(e)	-10	
Maximum ferrous ion uptake rate	EX_fe2(e)	-10	
Maximum ferric ion uptake rate	EX_fe3(e)	-10	
Maximum magnesium ion uptake rate	EX_mg2(e)	-10	
Maximum sodium ion uptake rate	EX_na1(e)	-10	
Maximum oxygen uptake rate	EX_o2(e)	-10	
Maximum carbon dioxide uptake rate	EX_co2(e)	-11.16	Only in autotrophic conditions
Maximum acetate uptake rate	EX_ac(e)	-10	Only in heterotrophic conditions

3.1.2.2 Specific cell constraints

Specific cell constraints are experimental constraints (§3.1.1.5) that better characterize cell metabolism by setting a bound to otherwise unbounded fluxes. This approach compensates for the lack of a regulatory network and typically involves processes occurring within the cell that saturate for reasons which go beyond the mass balance constraint. As regards specific cell constraints implemented in iRC1080, Chang *et al.* suggested the bounds in Table 3.3.

3.1.2.3 Tailored objective functions

In iRC1080, the three objective functions are included. Unless stated diversely, the autotrophic BOF is set as default. In Appendix C.1 and C.2, a table which reports the necessary precursors to form 1 g of dry weight biomass of *C. reinhardtii* in that condition is given. The stoichiometric coefficients in the BOF are calculated so to reproduce the correct proportion of each precursor in a mass unit of *C. reinhardtii* biomass, and thus they can be thought as the amount in mmol of that precursor which make up in 1 g of dry weight; in other terms, they have physical dimensions [mmol g_{DW}^{-1}]. With respect to the above cited paper values, stoichiometric coefficients for

RNA precursors had to be reassigned (according to data provided by Boyle and Morgan, 2009), in order to normalize all coefficients to 1 g.

Table 3.3 – *Specific cell constraints for some cellular processes implemented in iRC1080. Key: Pi – inorganic phosphate; Starch – 300 repeat units of amylase (80) and amylopectin (220); g1p – Glucose 1 phosphate; glc-A – α-glucose; glc-B – β-glucose; Photon-438- a photon in the violet/blue region, between 378 nm and 482 nm; Photon-673- a photon in the red region, between 659 nm and 684 nm; pq – Oxidized plastiquinone; pqh2 – Reduced plastiquinone.*

Cellular process	Reaction in iRC1080	Lower (mmol g_{DW}^{-1} h^{-1})	Upper (mmol g_{DW}^{-1} h^{-1})	Description
Non-growth associated ATP maintenance	ATP + H_2O -> ADP + Pi + H^+	0.183	0.183	Energy (in the form of ATP hydrolysis reaction) necessary to maintain a cell.
Starch degradation of glc-A (Dark)	49 H_2O + 250 Pi + Starch -> 250 g1p + 50 glc-A	0	$1.15 * 10^{-4}$ | $1.72 * 10^{-4}$	Degradation of starch by phosphorylase, amylase, dextrinase, maltase (chloroplast), 1:5 glc-A:g1p in dark aerobic | anaerobic conditions.
Starch degradation of glc-B (Dark)	49 H_2O + 250 Pi + Starch -> 250 g1p + 50 glc-B	0	$1.15 * 10^{-4}$ | $1.72 * 10^{-4}$	Degradation of starch by phosphorylase, amylase, dextrinase, maltase (chloroplast), 1:5 glc-B:g1p in dark aerobic | anaerobic conditions.
Starch degradation of glc-A (Light)	74 H_2O + 225 Pi + Starch -> 225 g1p + 75 glc-B	0	$4.35 * 10^{-5}$ | $6.53* 10^{-5}$	Degradation of starch by phosphorylase, amylase, dextrinase, maltase (chloroplast), 1:3 glc-A:g1p in light aerobic | anaerobic conditions.
Starch degradation of glc-B (Light)	74 H_2O + 225 Pi + Starch -> 225 g1p + 75 glc-B	0	$4.35 * 10^{-5}$ | $6.53* 10^{-5}$	Degradation of starch by phosphorylase, amylase, dextrinase, maltase (chloroplast), 1:3 glc-B:g1p in light aerobic | anaerobic conditions.
Maximum oxygen photoevolution	2 H_2O + 4 Photon-438 + 2 pq -> $O_{2,pe}$ + 2 pqh2 2 H_2O + 4 Photon-673 + 2 pq -> $O_{2,pe}$ + 2 pqh2	0	8.28	Maximum production of oxygen via photosynthesis as a result of the two reactions.

3.1.2.4 Default settings: the *standard configuration*

If iRC1080 is set up taking in account all the previous considerations with respect to medium conditions (with ammonia taken as the default nitrogen source), specific cell constraints and the normalized autotrophic biomass objective function, it is said to be in *standard configuration*. While reproducing simulation results, it must be made sure that

only uptake reactions which refer to a rich yet autotrophic medium are left unbounded (unless diversely specified in the simulation data). This means that all nutrients reported to be part of a rich medium are available for the cell, except for organic carbon sources. For instance, default settings of iRC1080 provide starch in excess from the environment, but this compound is not naturally present in a standard medium and thus it should be regarded as a supplementary nutrient. Note that if set in standard configuration, only the selected light source and its photon flux must be specified. This choice follows from the fact that light, as it will be explained in detail in the following section, is the limiting uptake flux for growth in autotrophic conditions.

3.2 Robustness analysis

Robustness analysis is the study of how an optimal state of a mathematical model output is sensitive, qualitatively or quantitatively, to different sources of variation in the input of the model. It can used to assess the limiting nutrient in GENREs. In this case, the mathematical model is the genome-based metabolic reconstruction of *C. reinhardtii*, namely iRC1080. The inputs are the nutrient fluxes, while the outputs are the growth rate and eventual by-product formation rate. For this particular microorganism, no by-products are expected to be secreted besides oxygen produced via photosynthesis. Therefore the growth rate is the only output for which robustness analysis has been carried out. The protocol followed to perform robustness analysis is described in Palsson, 2006 (§16.3, pg 213), with some modifications. Briefly, FBA is performed at each step by constraining the input flux to assume values up to an arbitrary constant, which is an upper bound for the input flux. The upper bound is incremented at each step until the whole span of values allowed for the input flux is covered. As a result, optimal growth rate is calculated by assuming that a single nutrient at each step is limited to a certain amount, so that the model can absorb at most that quantity (but it is not forced to take up that quantity, as in standard robustness analysis). By performing robustness analysis, it is possible to infer in which interval a nutrient is limiting for growth, when all other compounds are present in excess. Results were obtained by running the code `robustnessAnalysis.m`, available within the COBRA Toolbox (Becker *et al.*, 2007), which operates in MATLAB™, using the iRC1080 model in the standard configuration. Results are displayed in Figure 3.3.

While performing robustness analysis, only major sources were considered as input flux, namely water, ammonium, oxygen, carbon dioxide and solar light (lithosphere). Phosphate and sulfate were neglected to facilitate readability, as they both proved to be poorly limiting (data not shown). Two main results may be inferred from robustness analysis. First, in autotrophic conditions, light is by far the most limiting nutrient.

Secondly, even in light saturation conditions (above Φ_{sat} = 183.5 µmol m^{-2} s^{-1}) and rich medium, growth rate is limited to a maximum value of μ_{sat} = 0.1563 h^{-1}. The former result is a renowned behavior of microalgae, and nutrients often play a minor role in limiting growth (Zonneveld, 1998). The latter can be explained in terms of oxygen photo-evolution saturation. It should be recalled that the amount of oxygen produced via photosynthesis (and in particular by the PSII reaction) is constrained to an upper bound. This constraint has been introduced by Chang *et al.* since photosynthetically evolved O_2 cannot effectively drive other cellular processes and mostly diffuses out of the cell. Indeed, too much accumulation of O_2 formed via photosynthesis leads to photo-oxidative damage *in vivo*. Thus *in silico*, if light is provided in excess, photosynthetically evolved O_2 accomplishes a maximum production after which photons in excess are discarded by the simulation as they were useless and consequently reflected by the cell.

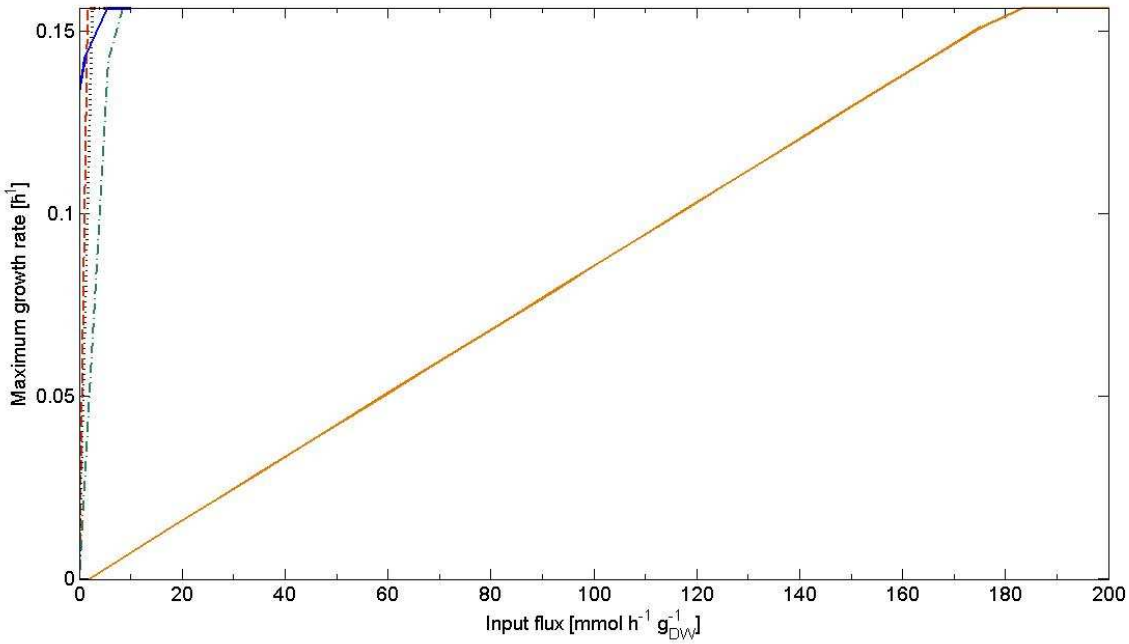

Figure 3.3. *Robustness analysis for iRC1080 model of* C. reinhardtii *set in standard conditions. Each line represents a different input flux (i.e. nutrient), spanning within its range of applicability. Only major nutrients were assessed.*
Key: — Water - - Ammonium ····· Oxygen - · - Carbon dioxide —Solar light (lithosphere).

3.3 Flux balance analysis

Flux balance analysis has been carried out for iRC1080 model set in standard configuration. Solar light (lithosphere) was chosen as light source and its photon flux has been varied from 0 to 300 µmol m^{-2} s^{-1}. Flux balance analysis was performed using the code `optimizeCbmodel.m`, available within the COBRA Toolbox, which operates in MATLAB™. In Figure 3.4, calculated specific growth rates (solid line) are

compared with experimental values, retrieved from the papers of Li *et al.* (2010), Geoffroy *et al.* (2007), Boyle and Morgan (2009) and Janssen *et al.* (1999). These values are collected in Table 3.4.

Table 3.4 – *Experimental specific growth rates of* C. reinhardtii *grown in HS medium.*

Photon flux [µmol m^{-2} s^{-1}]	Specific growth rate [h^{-1}]	Reference
240	0.16 ± 0.02	Janssen *et al.*, 1999
100	0.092 ± 0.008	Geoffroy *et al.*, 2007
65	0.059 ± 0.001	Boyle and Morgan, 2009
40	0.024	Li *et al.*, 2010

FBA of iRC1080 proved a striking ability to reproduce experimental results. It should be noted that Janssen *et al.* (1999) stated their intention to calculate specific growth rate of *C. reinhardtii* in light saturation conditions. The model not only reproduces growth saturation due to the constraint imposed on the maximum amount of photosynthetically evolved oxygen, but it also predicts a value for specific growth rate which is very close to the experiments and, in any case, within experimental error.

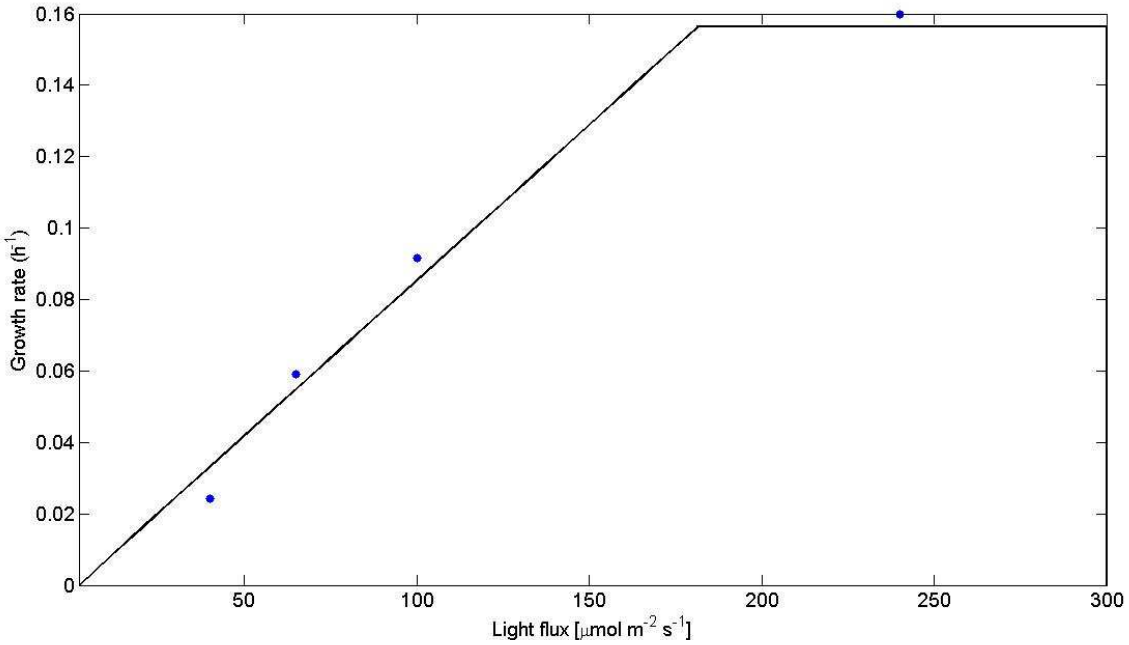

Figure 3.4. *Flux balance analysis of iRC1080 in standard conditions. Solid line represents the calculated specific growth rates. Dots represent experimental values as specified in Table 3.3.*

3.4 Conclusions

Both robustness analysis and flux balance analysis served mainly to test the capability of iRC1080 to correctly represent the flux distribution of the metabolic network of *C.*

reinhardtii within the conditions given in §3.1.2. Robustness analysis detected light photon flux as the most limiting nutrient and it identified two regions for growth, light-limited and photosynthetically evolved O_2-limited. Both results are experimentally confirmed. On the other hand, flux balance analysis reproduced, within experimental error, *C. reinhardtii* experimental specific growth rate at different photon fluxes. In conclusion, it can be stated that the model is predictive in standard configuration.

However, it should be noted that the predictability of the model relies heavily on the constraints imposed in §3.1.2. These constraints turn out to be essential to cope for the deficiency of a regulatory network. The only FBA fundamental assumptions, namely mass balance, steady state and optimization of a cellular function, are not sufficient to provide a realistic flux distribution of the metabolic network. Indeed, some experimental measures, with major emphasis on the values of Table 3.3 (and in particular the maximum rate of photo-evolved oxygen) and the biomass objective function as in Appendix C, were decisive to run reasonable simulations.

Chapter 4

Photobioreactor design for biofuel precursors production via *C. reinhardtii*

In conceptual design, a kinetic model normally is not available (Douglas, 1988). Generally, the cost associated with the reactor is not as important as the product cost distribution, so that a mere approximation is usually sufficient for preliminary cost estimation. This is particularly true for bulk chemicals production. However, biotech industry often requires residence time longer than standard in bulk chemical plants. Moreover, as regards biofuel precursors' production by microalgae, Radakovits *et al.* (2010a) noted that even modest improvements in conversion efficiency would drastically reduce the costs associated.

In this chapter, the concepts introduced in §2.1 will be applied for a preliminary design of a photobioreactor to produce triacylglicerols (TAGs) via *C. reinhardtii*. Since the goal of the photobioreactor is to grow biomass and at the same time to convert as much TAGs as possible, essential information about microalgae ability to accumulate TAGs must be collected and investigated in detail. Thus, in the following sections, an exhaustive description of the biochemistry underlying TAGs over-accumulation in *C. reinhardtii* will be given. Next, key parameters for photobioreactor design will be searched. Finally, some recommendations to improve prediction accuracy will be identified.

4.1 Introduction to reactor design for TAG bio-production

In biodiesel precursors' industrial production, the fundamental objectives for the reactor are both biomass growth and TAGs accumulation by *C. reinhardtii*. To simulate a reactor according to the principles in §2.1, a clear definition of cell functional state to achieve a precise goal must be given, in order to select both medium conditions and the objective function properly. As regards biomass growth, given the results of robustness analysis (§3.2) and flux balance analysis (§3.3), rich medium and light saturation determine a maximum specific growth rate. The prediction, obtained by setting the autotrophic BOF as objective function, was verified experimentally. However, TAG over-production in microalgae is still poorly understood (Work *et al*, 2010). In the following section, TAG accumulation will be better characterized for *C. reinhardtii* and this second goal will turn out to be conflicting with the previous one.

4.2 Triacylglycerols production by *C. reinhardtii*

Triacylglycerols (TAGs), also referred as neutral lipids or simply oil, occur naturally in microalgae biomass as fundamental constituents of the cell. They may be precursors of biodiesel, as they can be transesterificated in the presence of alcohols to methyl/ethyl-esthers. The way and the conditions under which microalgae over-accumulate neutral lipids are explored in the following sections.

4.2.1 Factors affecting TAGs content in microalgae

Even if TAGs are naturally present in microalgae, stored in cytosolic lipid bodies as a supplemental source of energy for the cell, in physiological conditions their content is very low, about 2-3%$_w$ of total dry weight (Li *et al.*, 2010). Neutral lipids start to be accumulated in stressed conditions. Though the occurrence and the extent to which TAGs are produced seem to be species/strain-specific, some common factors affect TAGs accumulation in several different microalgae. Among all, nitrogen deprivation displayed a critical general trend towards TAGs overproduction in numerous species. Green microalgae cultivated under nitrogen-depleted conditions showed an average increase of total lipid content from 25.5% to 45.7% of biomass dry weight (Hu *et al.*, 2008). Other causes were found to induce TAGs accumulation in some species, like high light intensity, stationary phase of growth and deficiency of other nutrients, for example phosphorus and sulfur. It is important to notice that if nitrogen is absent from the medium, biomass cannot grow at all, since fundamental components of the cell can no more be synthesized.

The reasons beyond microalgae behavior under stressed conditions are still debated. Apart from the evident physiological role of TAGs in serving as a carbon and energy resource for the cell, the TAG synthesis pathway may offer other significant advantages against stress. For instance, excess electrons derived from the over-reduced electron transport chain under high light or other stress conditions can damage cell macromolecules by inducing over-production of reactive oxygen species. Thus, the *de novo* TAG synthesis pathway would serve as an electron sink, since it consumes approximately 24 NADPH of the electron transport chain, more than any other macromolecule synthesis pathway within the cell metabolism. Other roles of TAG synthesis pathway in the excess light prevention, such as carotenoids over-production along a pathway which is coupled with TAG synthesis, have been reviewed in the past.

4.2.2 Experimental profile of TAG accumulation in C. reinhardtii

Several studies have confirmed that under stress conditions some strains of *C. reinhardtii* accumulate TAGs (Siaut *et al.*, 2011, Work *et al.*, 2010, Wang *et al.*, 2009,

Li *et al.*, 2010, James *et al.*, 2010). Most of them pointed out that over-production of TAGs occur in response to nitrogen deficiency. It is argued that starch synthesis is competitive with TAGs synthesis under these conditions, though it has been recently proved that no significant correlation can be assessed between oil and starch content in wild-type mutants, deprived of the starch synthesis pathway, and complemented strains (Siaut *et al.*, 2011). In addition, the same study found a general trend in oil accumulation, with maximum production between 2 and 3 days of nitrogen deficiency and achievement of a plateau after 5 days. This trend is more or less found in the above cited papers as well. In Figure 4.1, TAG content in time at quasi-saturated light intensity is reported, according to the works of Siaut *et al.* (2011). It should be noted that *C. reinhardtii* were grown in mixotrophic conditions, with acetate present in the medium. Maximum lipid accumulation rate has been interpolated and resulted $\mu_{TAGS,exp}^{max} = 0.088$ g g_{DW}^{-1} h^{-1}.

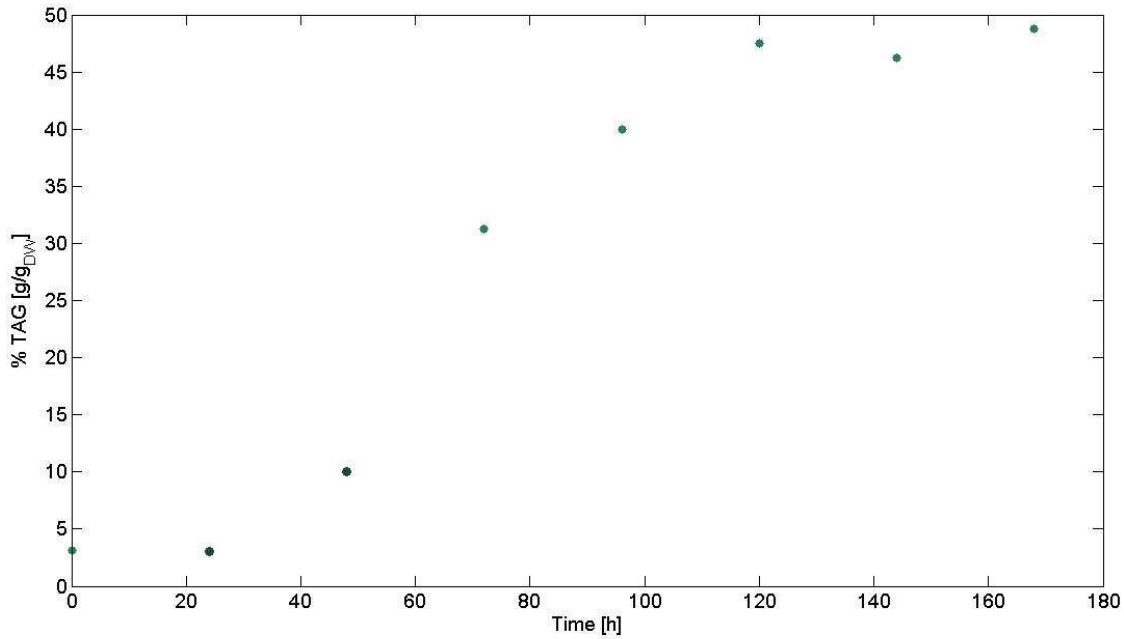

Figure 4.1. *TAG content time profile in* C. reinhardtii *grown at 150 µmol* m^{-2} s^{-1} *in TAP medium deprived of nitrogen sources. Data retrieved from Siaut* et al., *2011.*

4.3 Photobioreactor design principles

Since TAGs accumulation and biomass growth both depend on nitrogen availability on conflicting basis, the photobioreactor (PBR) should provide enough nitrogen to enhance *C. reinhardtii* duplication but it should deplete it on time to allow TAGs to be over-produced. In other terms, the PBR should be thought as divided in two parts. Initially a given amount nitrogen must be available to consent biomass growth. This section will be referred as biomass stage. When nitrogen is depleted, biomass will stop growing and conversely TAGs will start to accumulate. This section will be referred as TAGs stage.

It is clear that the higher the initial amount of nitrogen, the higher the biomass yield but, on the other hand, the residence time will increase. Thus a trade-off between productivity and reactor size arises. It is convenient to fix the nitrogen initial concentration, $C_{N,0}$, as a key parameter upon which productivity and residence time depend. A pictorial view of the proposed PBR is depicted in Figure 4.2.

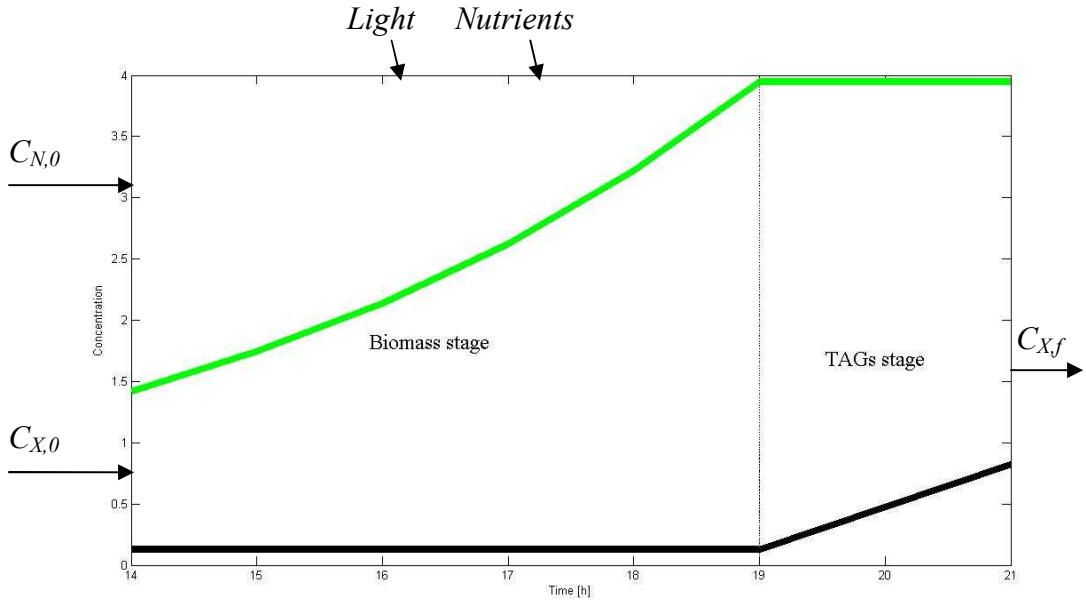

Figure 4.2. *A pictorial proposed scheme for the PBR, with concentration profiles for biomass (green solid line) and TAGs (black solid line). Nutrients (other than nitrogen) and light are provided in excess and are displayed by the arrows above the tube. Nitrogen and biomass enter as input in a given amount,* $C_{N,0}$ *and* $C_{X,0}$ *respectively. As output, biomass (*$C_{X,f}$*) exits from the PBR with a certain amount of TAGs accumulated.*

4.3.1 Biomass stage simulation

In the biomass stage, according to the robustness analysis results, growth is limited by the maximum amount of photo-synthetically evolved oxygen (§3.3). Thus, in order to accomplish maximum growth rate, rich medium and light saturation are imposed throughout the simulation. The objective function is the autotrophic biomass objective function, as it fairly represents *C. reinhardtii* functional state in physiological conditions, which is the case. According to the new approach introduced in §2.1.2, in the infinitesimal interval $d\tau$ each equation of the system (2.10) is solved by updating the initial conditions to the last available result for that concentration. Since nitrogen concentration is a key parameter, at the first iteration step the equation for this substrate is assigned an initial value equal to $C_{N,0.}$ All other substrates are neglected given the assumption of rich medium. As soon as nitrogen is depleted, i.e. by solving the corresponding mass balance equation its concentration becomes negative, the biomass stage simulation is stopped and the PBR enters the TAGs stage.

4.3.2 TAGs stage simulation

In the TAGs stage, nitrogen is depleted and according to microalgae response to this kind of stress, triacylglycerols begin to accumulate. In contrast with the previous stage, no functional state for the cell is well defined. The question is how to find an objective function which can represent it correctly. Apparently, TAGs accumulation may be the default pathways in algae under these environmental conditions (Hu *et al.,* 2008). Moreover, some studies suggest that neutral lipids rather than total lipid start to accumulate in response to nitrogen deprivation (Li *et al.*, 2010, Siaut *et al.*, 2011). If so, a tentative objective function could be the maximization of TAGs accumulation rate. A similar approach has already been investigated, to determine the biochemical production capabilities of particular cell (Palsson, 2006). In absence of a more comprehensive knowledge of the regulatory mechanisms which underlie *C. reinhardtii* response to nitrogen deficiency, the assumption of maximization of TAGs accumulation is adopted to simulate the TAGs stage.

To build such objective function, a fictitious reaction must be added, dubbed TAGs objective function. This has been derived from the autotrophic BOF by eliminating all non-TAGs metabolites. The ratios of each TAG in the BOF have been preserved, but coefficients have been normalized in order to give 1 g of TAGs per unit of dry weight biomass. In this fashion, the resulting flux can be compared with experimental results straightforwardly, since the latter data are usually given in percentage of dry weight on mass basis. A table with a complete list of the precursors which make up the TAGs objective function and the corresponding stoichiometric coefficient is presented in Appendix C.3.

The TAGs stage simulation relies heavily on the objective function chosen to best represent the metabolic flux distribution of *C. reinhardtii* after nitrogen depletion. To test whether simulation results are compatible with reality, *in silico* TAGs accumulation rate has been calculated both in heterotrophic and autotrophic conditions. Flux balance analysis has been carried out for iRC1080 model in standard configuration using the code `optimizeCbmodel.m`, available within the COBRA Toolbox, which operates in MATLAB™. The objective function was switched from the autotrophic BOF to the TAGs objective function before running the program. *In silico* heterotrophic conditions were achieved by setting the maximum acetate uptake rate to 10 mmol g_{DW}^{-1} h^{-1} and, conversely, the maximum carbon dioxide uptake rate to zero, as suggested by Table 3.2. In both autotrophic and heterotrophic simulations, TAGs accumulation rate has been left parametric with solar light (lithosphere), whose flux ranged from $\Phi_{min} = 0$ μmol m^{-2} s^{-1}

to Φ_{max} = 200 µmol m^{-2} s^{-1}. Results are displayed in Figure 4.3 and compared with the experimental value retrieved from Siaut *et al.* (see §4.2.2).

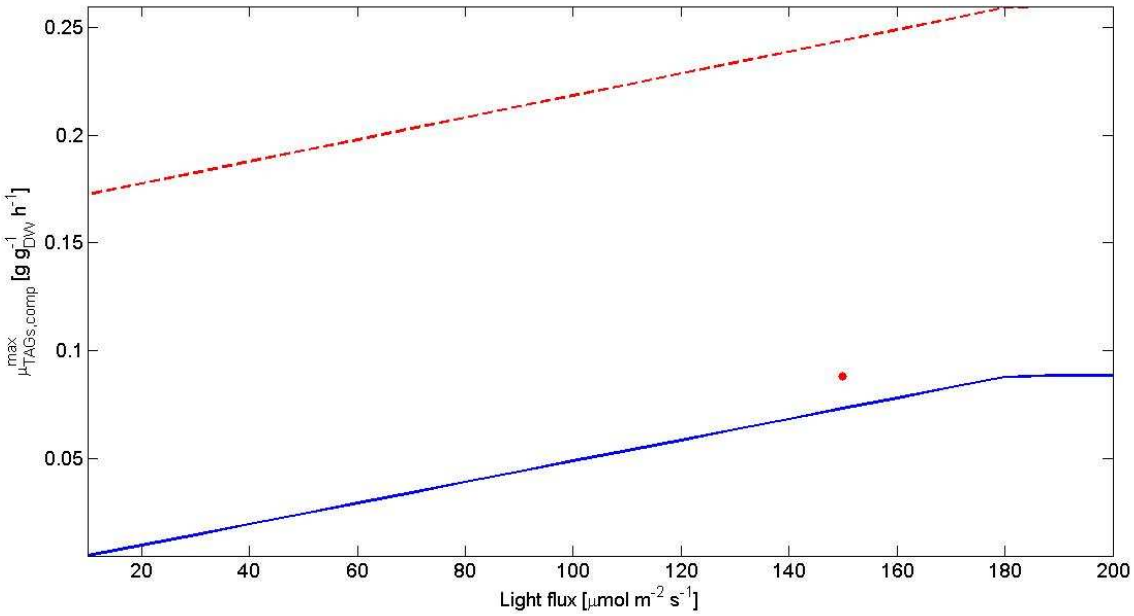

Figure 4.3. *TAGs accumulation rate as calculated* in silico *with iRC1080 model in both heterotrophic (dashed line) and autotrophic (solid line) conditions, using TAGs objective function as the optimized cellular function under nitrogen deprivation. The experimental data retrieved by Siaut* et al. *has been reported for comparison (red dot).*

It should be noted that the value obtained from the work of Siaut *et al*. refers to heterotrophic conditions, i.e. *C. reinhardtii* was grown in TAP medium. Therefore TAGs accumulation rate as predicted by *in silico* simulation using TAGs objective function is clearly overestimated. Several reasons may explain this result. First, neutral lipids accumulation may not be the only cellular function optimized under stress conditions. As cited before, starch accumulation or other pathways may be active as well, leading to a carbon source distribution within cell metabolism which actually detours from TAGs synthesis pathway. This leads to an accumulation rate which is smaller. Secondly, by analyzing simulation results it can be seen that acetate uptake rate is always at maximum. Thus, in nitrogen deficiency conditions, either the constraint is poorly representative or perhaps some metabolic mechanisms regulate acetate uptake to preserve its immediate consumption. In both cases TAGs accumulation rate would be affected negatively.

Since metabolism regulation under these conditions is still inadequately covered in literature, this solution for TAGs stage simulation has been chosen among other possible alternatives as it proved to attain a qualitative description of cell metabolism during nitrogen deprivation and semi-quantitative results.

4.4 Application of the new approach to PBR design

The new approach has been used to solve the system of equations (2.7) for the case of a PBR producing TAGs via *C. reinhardtii*. The system has been solved separately for the biomass stage and the TAGs stage. Computationally, the procedure is the same but the objective function switches from the autotrophic BOF to the TAGs objective function as soon as the nitrogen concentration goes below zero. The objective of the simulation is to predict the residence time needed to achieve a certain productivity given an initial amount of nitrogen. The total consumption of other nutrients, such as carbon dioxide or oxygen, has been estimated as well.

4.4.1 Simulating a PBR

The simulation was performed in MATLAB™, using iRC1080 model set in standard configuration (thus in autotrophic conditions), with some modifications. Taking the system of equations (2.10) as a reference, at each iteration step this set of equations must be solved. For the purposes of our analysis, the substrate *S* is the nitrogen source (i.e. ammonia) and the product *P* are TAGs. As regards nutrients, the initial conditions for all input fluxes other than ammonia are left unchanged throughout the simulation. In this fashion, at each step FBA is not constrained by their amounts but eventually by the maximum uptake rate as for Table 3.2. Similarly, solar light (lithosphere) is used at saturation conditions for the microalgae, i.e. $\Phi = 200$ μmol m^{-2} s^{-1}. On the other hand, the maximum ammonia uptake rate is initially imposed to be equal to $C_{N,0}$ and successively updated at each step by solving the corresponding equation in (2.10). By doing so, FBA can be constrained by the increasingly lack of nitrogen from the medium. As regards the choice of the solving method, dynamic FBA was preferred to numerical schemes in the following simulations. A comparison with results from the latter methods is provided in Appendix D. It is interesting to point out that results via dFBA are slightly more accurate than 1st order Euler method, but, on the other hand, Crank-Nicolson method seems to outperform dFBA.

Dynamic FBA prescribes to solve system (2.10) rather than (2.9). As described in §2.1, flux balance analysis must be run at each time step. FBA was performed using the code `optimizeCbmodel.m`, available within the COBRA Toolbox. At the first iteration, $C_{X,0}$ was set equal to 0.10 g_{DW} L^{-1} and $C_{P,0} = 0.12568$ g_{TAGs} g_{DW}^{-1}. A time step was selected to be $d\tau = 1$ h, to limit the computation time required. Once the flux distribution was obtained with FBA, the value of both BOF flux (v_X), TAGs objective function (v_P) and nitrate exchange flux (v_S) were extracted and used to solve the corresponding equations in (2.10). The output values for the concentration of substrate (C_S), product (C_P) and biomass (C_X) were successively recycled in the next time step as

initial values ($C_{S,0}$, $C_{P,0}$, $C_{X,0}$, respectively). When the algorithm found that the substrate concentration (C_S) was no more positive, the objective function for FBA switched from BOF to TAGs objective function and the procedure was repeated until TAGs concentration reached a reasonable amount equal to $C_{P,f}$ = 0.60 g_{TAGs} g_{DW}^{-1}. In Figure 4.4, concentration profile for substrate (red), product (green) and biomass (black) are shown in function of residence time, as a result of the above described procedure by setting $C_{N,0}$ = 4.0 mmol L^{-1}.

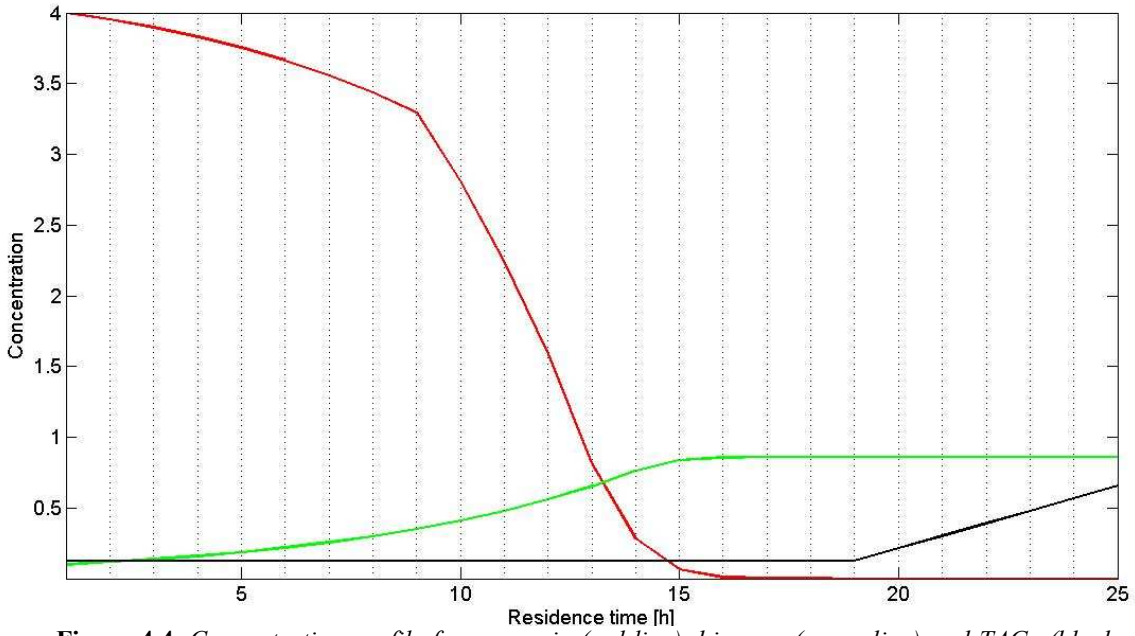

Figure 4.4. *Concentration profile for ammonia (red line), biomass (green line) and TAGs (black line). Physical dimensions for ammonia concentration is mmol L^{-1}, for biomass is g_{DW} L^{-1} and for TAGs is g_{TAGs} g_{DW}. Initial nitrogen concentration is $C_{N,0}$ = 4.0 mmol L^{-1} and $d\tau$ = 1h (as indicated by the vertical black dotted lines)*

4.4.2 TAGs outlet concentration and residence time

As introduced in §4.3, both outlet concentration Y [g_{TAGs} L^{-1}] and residence time τ [h] (from which reactor size can be deduced) depend on nitrogen initial concentration $C_{N,0}$ on opposite basis. Outlet concentration is defined as the mass amount of TAGs per unit of volume at the reactor exit:

$$Y = C_{X,f} C_{P,f} \tag{4.1}$$

where $C_{X,f}$ and $C_{P,f}$ are the biomass concentration [g_{DW} L^{-1}] and TAGs concentration [g_{TAGs} g_{DW}^{-1}], respectively, that exit from the reactor. Thus, in order to generate a trend for outlet concentration vs. residence time, the above described procedure has been repeated for different values of $C_{N,0}$. These details, along with the period required to deplete all available nitrogen (τ_{Nf}) and final biomass concentration (C_{Xf}), are reported in Table 4.1. In addition, Figure 4.5 displays outlet concentration (blue) and residence

time (green) distribution upon nitrogen initial concentration. It should be noted that outlet concentration varies linearly with nitrogen initial concentration, but residence time does not. Productivity is defined as outlet concentration per volume of reactor:

$$P = \frac{Y}{\tau} \tag{4.1}$$

Productivity represents the amount of TAGs that can be achieved per hour in a unit volume of reactor. According to the simulation, its trend along with nitrogen initial concentration is shown in Figure 4.6. It can be inferred that productivity increases with a larger availability of nitrogen at the reactor inlet, since biomass will grow to a higher concentration, but it tends to saturate, as residence time raises as well.

Table 4.1 – *Outlet concentration, residence time, time of nitrogen depletion and final biomass concentration are reported as a function of initial nitrogen concentration in a simulated PBR.*

Initial nitrogen concentration N_0 $[mmol\ L^{-1}]$	Outlet concentration $Y\ [g_{TAGs}\ L^{-1}]$	Residence time τ $[h]$	Time of nitrogen depletion τ_{Nf} $[h]$	Final biomass concentration $C_{Xf}\ [g_{DW}\ L^{-1}]$
0	0.06	5.5	0.0	0.10
4	0.52	22.4	17.0	0.87
8	1.58	27.6	22.2	2.61
12	1.89	28.6	23.2	3.14
16	2.18	29.4	24.0	3.61
20	2.63	30.4	25.0	4.36
24	3.34	31.7	26.3	5.53
28	3.59	32.2	26.8	5.95
32	4.09	32.9	27.5	6.78
36	4.33	33.2	27.8	7.17
40	4.45	33.4	28.0	7.38

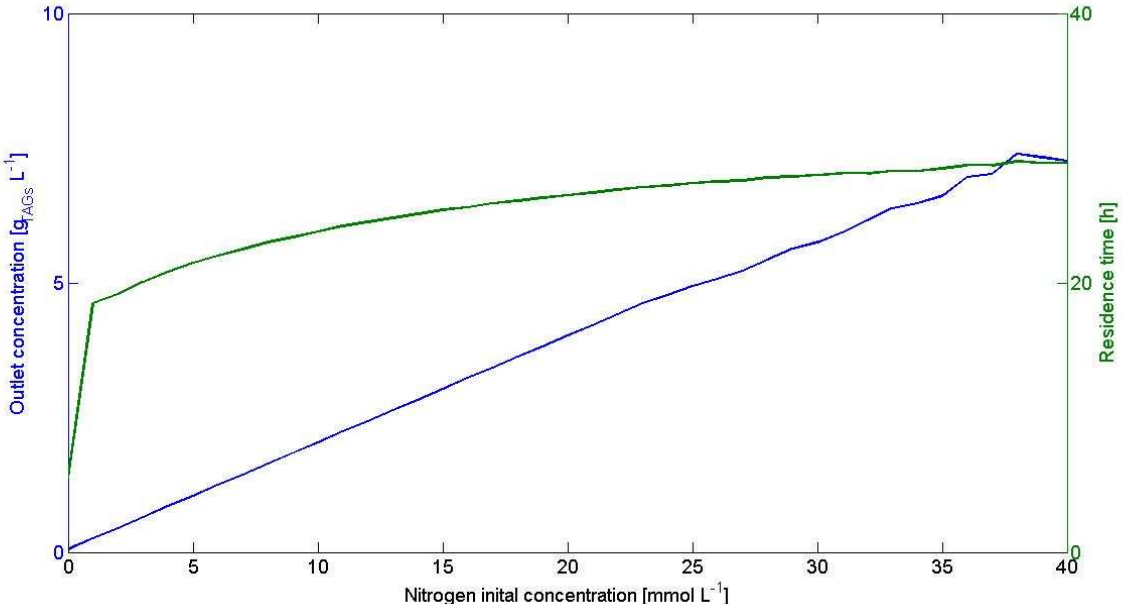

Figure 4.5. *Outlet concentration (blue) and residence time (green) as a function of nitrogen initial concentration for the above simulated photobioreactor varying $C_{N,0}$ from 0 mmol L^{-1} to 40 mmol L^{-1}.*

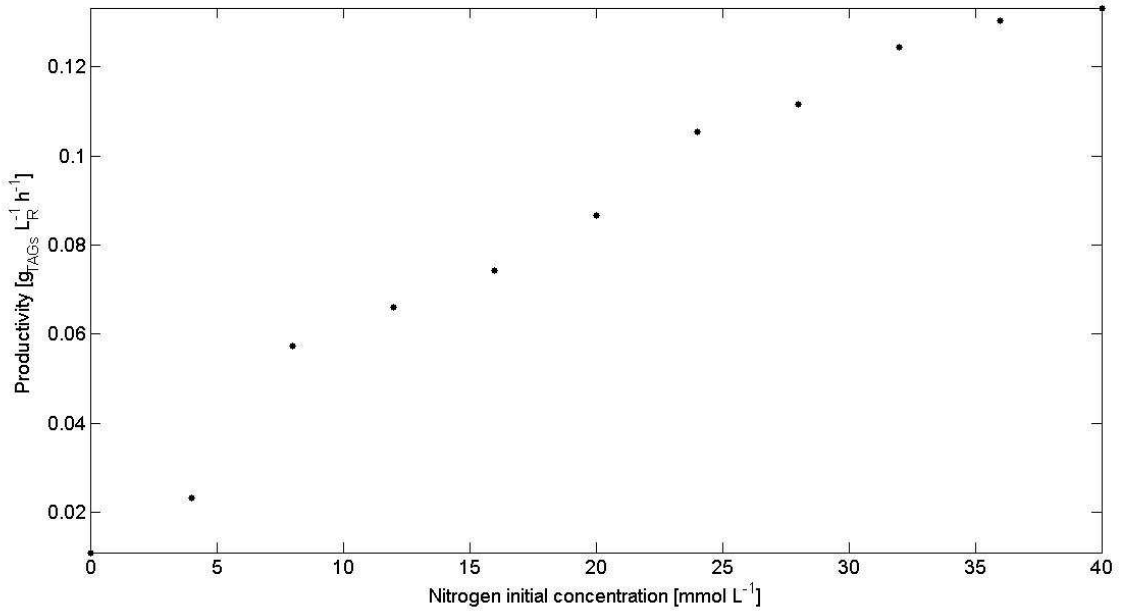

Figure 4.6. *Productivity as a function of nitrogen initial concentration for the above simulated photobioreactor varying $C_{N,0}$ from 0 mmol L^{-1} to 40 mmol L^{-1}.*

Figure 4.7 is based on Figure 4.5, but it represents the relationship between outlet concentration and residence time, which is generally more practical in reactor design.

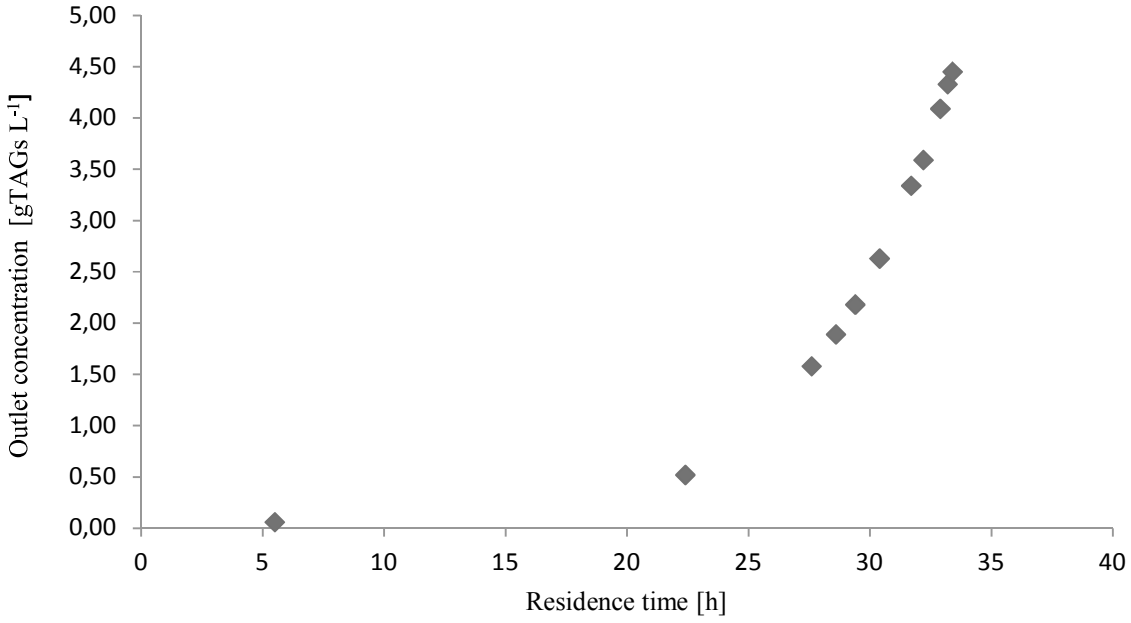

Figure 4.7. *Outlet concentration as a function of residence time for the above simulated photobioreactor.*

4.4.3 Nutrients consumption

Total nutrients consumption was also plotted against residence time and outlet concentration, in Figure 4.8 and 4.9, respectively. To evaluate the consumption of oxygen and carbon dioxide, the corresponding flux calculated with FBA was used to solve iteratively the first equation of system (2.10) at each time step. The initial condition for both concentrations was zero, so that at the end of the simulation the overall need of a certain nutrient was equal to the cumulated concentration. Oxygen concentration was split to assess metabolically consumed O_2, by taking in account the uptake flux of oxygen at each time step, against photoevolved O_2 ($O_{2,pe}$), evaluated at each iteration from the secretion flux of $O_{2,pe}$. As photoevolved O_2 is produced, it assumes negative values in Figure 4.8 and 4.9. As it can be inferred by the results, photoevolved oxygen always exceeds the need of metabolically consumed O_2, therefore suggesting that excess oxygen is not a key requirement for running the PBR. Such result has been confirmed experimentally by Heifetz *et al.* (2000). On the other hand, carbon dioxide is critical for cell viability. Figure 4.9 suggests that CO_2 consumption varies linearly with outlet concentration of TAGs. By interpolating data, it results that:

$$C_{CO_2,0} \cong 86.52\, Y \tag{4.2}$$

where $C_{CO_2,0}$ is the initial concentration of carbon dioxide [mmol L^{-1}] needed to achieve that outlet concentration (defined as above). The ratio is almost constant and stands for the amount of CO_2 (in mmol) required to produce 1 g of TAGs.

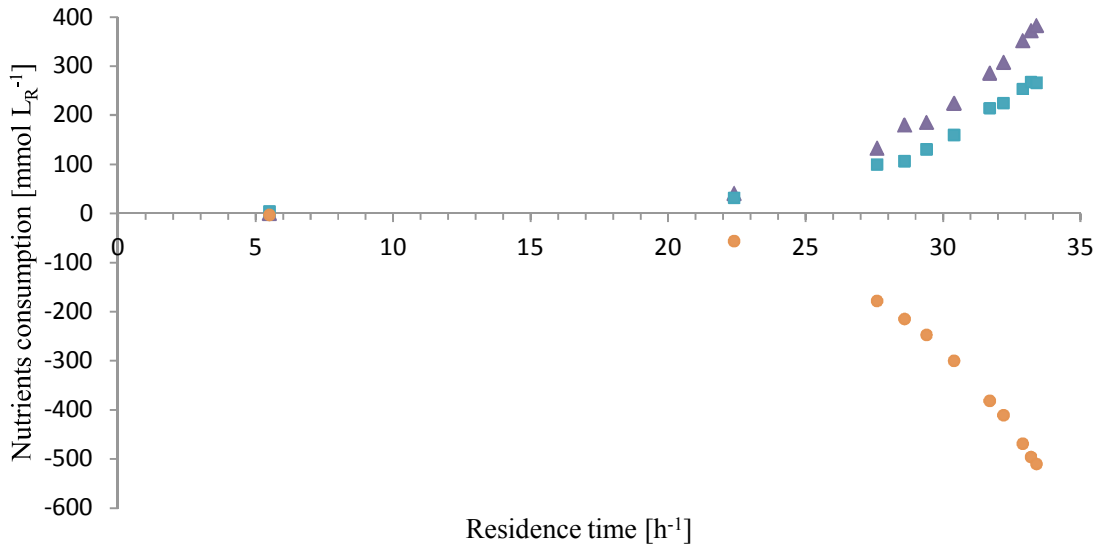

Figure 4.8. *Nutrients consumption in a PBR as function of residence time. Oxygen was split into metabolically consumed O_2 (blue squares) and photoevolved O_2, which is produced and thus assumes negative values (orange circles). Carbon dioxide is represented by violet triangles.*

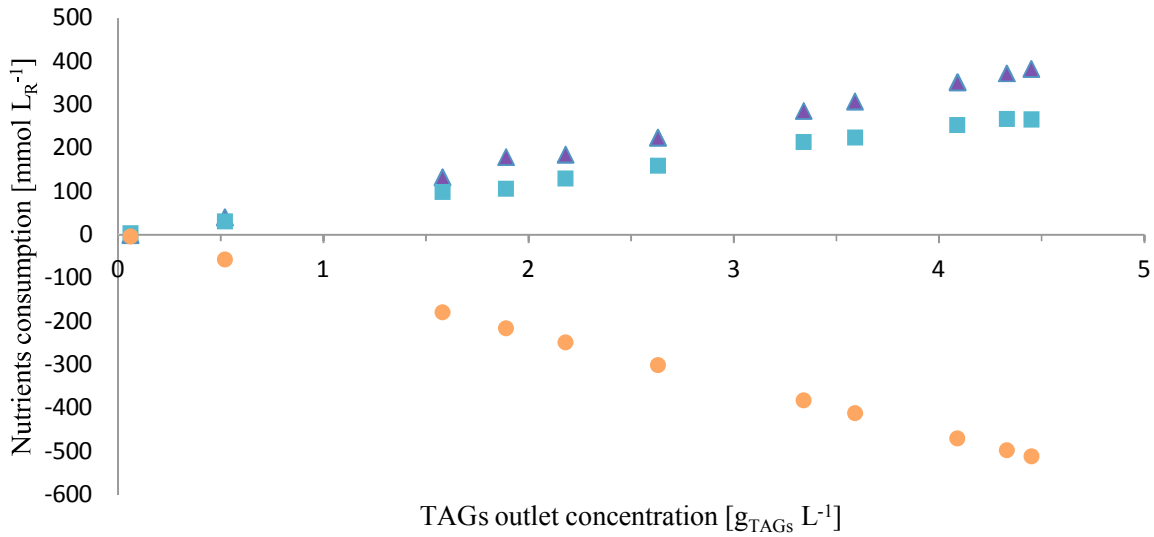

Figure 4.9. *Nutrients consumption in a PBR as function of TAGs outlet concentration. Key as for the figure above.*

4.4.4 Predictability accuracy improvement

In the case of TAGs production by *C. reinhardtii*, a critical step in PBR design was a realistic simulation of the TAGs stage. Though little experimental evidences are

available to validate a more precise value for the TAGs accumulation rate, still the *in silico* predicted rate appears to be somehow overestimated. Since nitrogen deprivation is the main reason for TAGs accumulation, a only partial adaptation of the cell to these stressed conditions could explain why during experiments microalgae have not optimized yet the TAGs synthesis pathway. If so, any optimization methods like FBA would fail in predicting a realistic flux distribution and more advanced techniques such as MOMA (§1.3.2) would be more appropriate. If not, a possible solution could be a different objective function. In this fashion, it would be allowed that microalgae under stress not only produce TAGs but also activate some other cellular functions. In other words, microalgae try to accomplish multiple goals to achieve survival. Among these, an often debated question in literature has been the role of starch synthesis during nitrogen deficiency. Siaut *et al.* (2011) for instance verified that *C. reinhardtii* strain cw15 after culture in nitrogen deprived media for 2 days over-accumulated both TAGs and starch. On the other hand, Li *et al.* proved that starchless BAFJ5 mutant of wild-type reference strain cc-1690 accumulated neutral lipids 3.5 to 8 fold higher than the wild-type.

To explore competition between TAGs synthesis and starch in *C. reinhardtii*, multi-objective FBA has been applied. If such competition exists, it could explain a lower TAGs accumulation rate than previously predicted. Indeed, in the above simulated TAGs stage of the PBR, no starch was being synthesized while TAGs were accumulating (data not shown). To perform multi-objective FBA, a Pareto analysis approach has been used. Thus, iRC1080 model was set in standard configuration and FBA was applied to maximize the TAGs objective function. Solar light (lithosphere) was selected and its flux varied from $\Phi = 0$ to 300 μmol m^{-2} s^{-1}. For each light flux value, FBA has been repeated by imposing an increasing minimum value of starch production from $v_{Starch} = 0$ to 0.3 g g_{DW}^{-1} h^{-1}. To do so, a lower bound for the following reaction has been updated at each simulation step:

$$300\ \mathrm{ADPGlc} + \mathrm{H_2O} \rightarrow 300\ \mathrm{ADP} + 300\ \mathrm{H^+} + \ \mathrm{Starch}^1 \qquad (4.3)$$

It results a Pareto frontier which represents the maximum TAGs accumulation rate at a certain solar light flux given a minimum amount of starch produced (Figure 4.9). As it can be inferred by Figure 4.10, starch synthesis competes with TAGs accumulation at any level of production for the whole range of light intensity.

[1] Key: ADPGlc – ADP Glucose

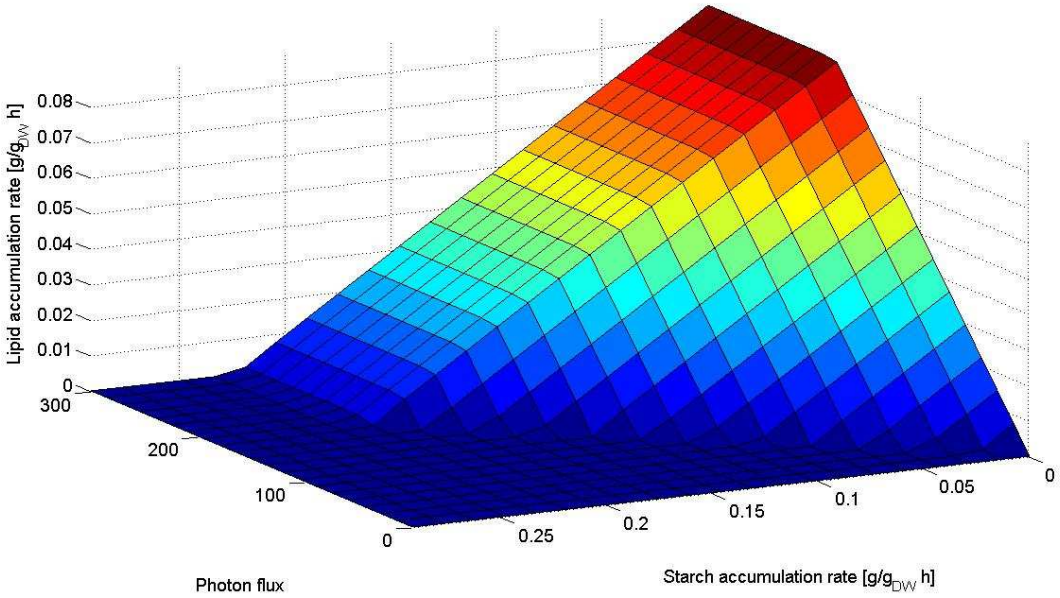

Figure 4.10. *Pareto frontier of* C. reinhardtii *under nitrogen deficiency conditions. At a certain photon flux, maximum lipid accumulation rate was estimated using FBA given a minimum amount of starch production.*

In saturating light conditions (Φ > 183.5 µmol m^{-2} s^{-1}), TAGs accumulation rate is limited linearly by the increasing production rate of starch. In conclusion, Pareto analysis proved that the two pathways are competitive, at least metabolically speaking. Thus an experimental value to constrain a minimum synthesis of starch could significantly improve the prediction of TAGs accumulation rate in PBR design.

4.5 Conclusions

A new approach for bioreactor design has been successfully applied. Through the application of systems biology techniques, among all FBA, a conceptual design of a photobioreactor for the production of biofuel precursors, namely triacylglycerols, by *C. reinhardtii* has been accomplished. The concentration profile for some key components such as biomass, TAGs and nitrogen has been shown in Figure 4.4. TAGs outlet concentration has been correlated to the reactor residence time in Figure 4.5. Total nutrients consumption, with particular emphasis on carbon dioxide, has been calculated and reported in Figure 4.8 and 4.9. In conclusions, most of data needed in process conceptual design of reactors could be inferred by properly setting the genome-based metabolic reconstruction of *C. reinhardtii* iRC1080 and by carefully performing the method described previously in §2.1.2.

The above-mentioned case study should be further studied to represent other different conditions, which were not a major goal of the present work. For instance, diverse

environmental conditions could be explored such as heterotrophy or mixotrophy and KO mutants as well. This is possible thanks to the intrinsic flexibility of the method, which finds in its main strength its diverse range of applicability.
Although previous results were found satisfactory under many points of view, some limitations hinder a more detailed design of the PBR. First of all, constant light conversion efficiency has been assumed. Not only light absorption is limited by growing biomass density, but also constant illumination is an abstraction which is not actually feasible for industrial application. Secondly, as already pointed out in §4.3.2., TAGs accumulation rate is always taken at its maximum during the TAGs stage. As examined by Siaut *et al.*, lipid accumulation reaches a maximum only after a few days of nitrogen deprivation. Moreover, TAGs accumulation rate predicted *in silico* is clearly overestimated with regards of the little available experimental data.

In conclusion, systems biology proved to provide precious instruments for bioreactor design, especially in the stage of process conceptual design, where discerning the most valuable process among multiple alternatives is usually a major challenge. However, this tool must be improved before quantitative information can be obtained out of it.

Chapter 5

Biochemical reaction engineering applied to biofuel precursors synthesis in *C. reinhardtii*

Genome sequencing and sequence manipulation techniques have paved the way to the engineering of microorganisms tailored to either enhance the yield of a valuable biological component or allow the use of living cells for synthetic applications in medicine, alternative energy, and materials science (Weeks and Chang, 2011). Several efforts have been made to genetically engineer microalgae to improve biofuel precursors production as well (Radakovits *et al*., 2010a). Along with some successful strategies (Radakovits *et al*., 2010b), such as expression of two thioesterases in *P. tricornutum* towards the production of lauric and mystiric acid, some pathway-targeted manipulation, like the ACCase enzyme overexpression which catalyzes the conversion of acetyl-CoA to malonyl-CoA in the early stage of fatty acid synthesis pathway, proved to fail (Dunahay *et al*., 1995). A systems approach to metabolic engineering is thus desirable to better understand those strategies which can actually turn into a secure overproduction of desired metabolites.

In this chapter, the concepts introduced in §2.2 will be applied to investigate about the strategy of metabolic engineering in *C. reinhardtii* towards the production of biofuel precursors, that are TAGs. First, a computational method called OptKnock, which finds its basis in FBA, will be performed to generate a range of diverse gene knock-outs which may lead to TAGs overproduction. Secondly, a simple but novel technique will be proposed to explore whether certain modified media can enhance biomass growth and therefore specific productivity of desired metabolites.

5.1 Computational metabolic engineering of *C. reinhardtii*

According to §2.2.1, to analyze systematically the effects of single gene KOs on the production of a certain metabolite, a genome-based metabolic reconstruction which includes a GPR association matrix must be available, which is the case of iRC1080. Next, the freely available algorithm OptKnock checks for every gene present in the matrix and it constraints all associated reactions bounds to zero. By performing a bilevel

linear programming optimization, the program asks the model to maximize a selected flux while optimizing biomass growth, in both cases solving iteratively FBA. A structured matrix is given as an output, in which each gene KO is matched to the range in which the selected flux can span.

5.1.1. Simulating gene KOs to induce a TAG overproduction

The above described procedure was applied to iRC1080 set in standard configuration. Solar light (lithosphere) was chosen and its flux imposed to be Φ = 300 µmol m^{-2} s^{-1}. The program `simpleOptKnock.m`, included within the COBRA Toolbox, was run by selecting a fictitious reaction which secretes a representative TAG (namely 16:0/18:1(9Z)/16:0) as target flux. To preserve readibility, results were displayed only for those KOs which produced at least a maximum secretion flux $v_{TAG*} = 1.2 \cdot 10^{-4}$ mmol g_{DW}^{-1} h^{-1}, that is v_{TAG*} = 0.1 mg g_{DW}^{-1} h^{-1}. Moreover, only KOs that did not cause cell death were reported. Thus, in Figure 5.1, the set of knocked-out gene which induced a meaningful target flux is represented. In this specific case, it must be noted that none of the KOs produced a unique value for the target flux, but it rather resulted in a range, indicated by the light red area comprised between the upper and lower red solid line.

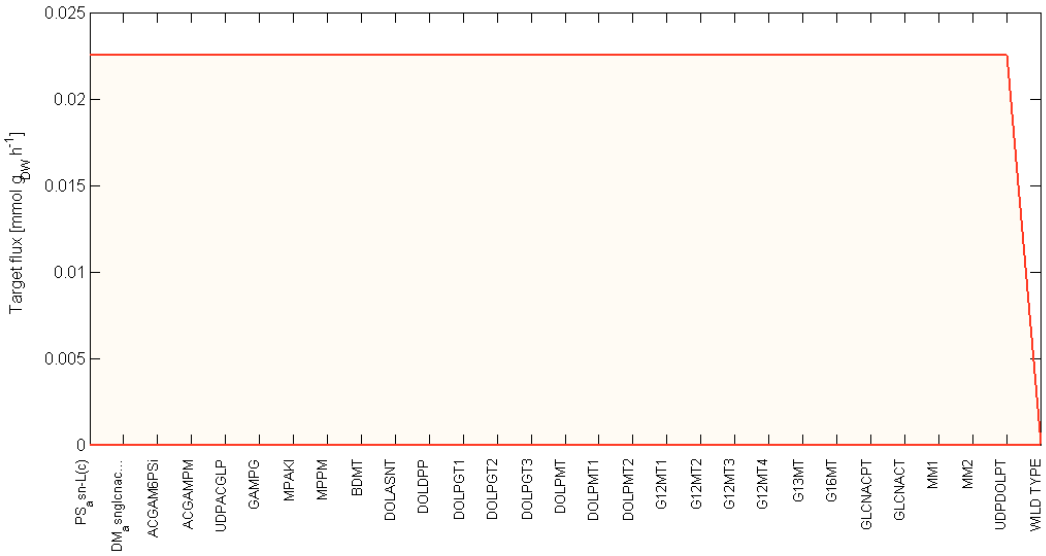

Figure 5.1. *Knocked-out genes that induced TAG overproduction in* C. reinhardtii *according to OptKnock simulation results. On the x-axis, genes abbreviations are reported. For each gene, the range of values that the target flux can take is displayed by the area comprised by the upper and lower red line. Wild type has been included as control (no KOs).*

5.1.2. Analysis of results for TAG overproduction

Because of the easy approach that OptKnock employs to describe a complex phenomenon like a genetic perturbation, a careful analysis of results has been carried out. First, it should be noted that all KOs are redundant, as they refer to the same

pathway. Thus, there are no appreciable differences in the deletion of any of the above genes, given that they all encode enzymes involved in that same pathway and therefore the consequences on cell metabolism are identical. Specifically, these enzymes participate at the N-glycan biosynthesis (also called amino sugar and nucleotide sugar metabolism). By knocking out any gene in the set, the pathway turns inactive. This result suggests that TAG overproduction could be induced by a more efficient exploitation of acetyl-CoA, a compound essential in the early stage of fatty acid synthesis, which is explicitly employed in the N-glycan biosynthesis pathway in the following reaction:

$$\mathrm{AcCoA} + \mathrm{GAM6P} \rightarrow \mathrm{CoA} + \mathrm{AcGAM6P} + \mathrm{H}^{+}\ ^{2} \tag{5.1}$$

This reaction is catalyzed by acetyl-CoA:D-glucosamine-6-phosphate N-acetyltransferase, encoded by ACGAMPSi in Figure 5.1. However, a further analysis of results revealed that this pathway is physiologically inactive if a specific glycoprotein[3] is absent from the medium. Considered that Figure 5.1 does not indicate a secure secretion of TAGs under these genetic perturbations, the analysis of results indicates that such metabolic engineering strategy is poor and, after all, negligible.

5.1.3. Simulating gene KOs to induce H_2 overproduction

The ability of *C. reinhardtii* to produce molecular hydrogen has been first elucidated by Melis and coworkers (2000). Under specific conditions, the green alga is able to carry out the following hydrogenase reaction:

$$2\ \mathrm{FDXRd} \leftrightarrow 2\ \mathrm{FDXOx} + \mathrm{H_2} + 2\mathrm{H}^{+}\ ^{4} \tag{5.2}$$

As photobiological production of H_2 by microorganisms is of prime interest in the current renewable fuel market (Melis *et al*., 2000), the same approach adopted for engineering TAGs production by *C. reinhardtii* has been applied to hydrogen synthesis. The procedure followed exactly the same instructions given in the previous section, besides that reaction (5.2) was selected to be the target flux of the simulation. Results are shown in Figure 5.2, in the same fashion as in Figure 5.1.

---

[2] Key: AcCoA – acetyl-coenzyme A, GAM6P – D-Glucosamine 6-phosphate, CoA – coenzyme A, AcGAM6P – N-Acetyl-D-glucosamine 6-phosphate, H^+ – proton.

[3] KEGG code: G00011, named Asn(-b1)GlcNAc(4-b1)GlcNAc(4-b1)Man[(6-a1)Man[(6-a1)Man(2-a1)Man](3-a1)Man(2-a1)Man](3-a1)Man(2-a1)Man(2-a1)Man.

[4] Key: FDXRd – reduced ferredoxin, FDXOx – oxidized ferredoxin.

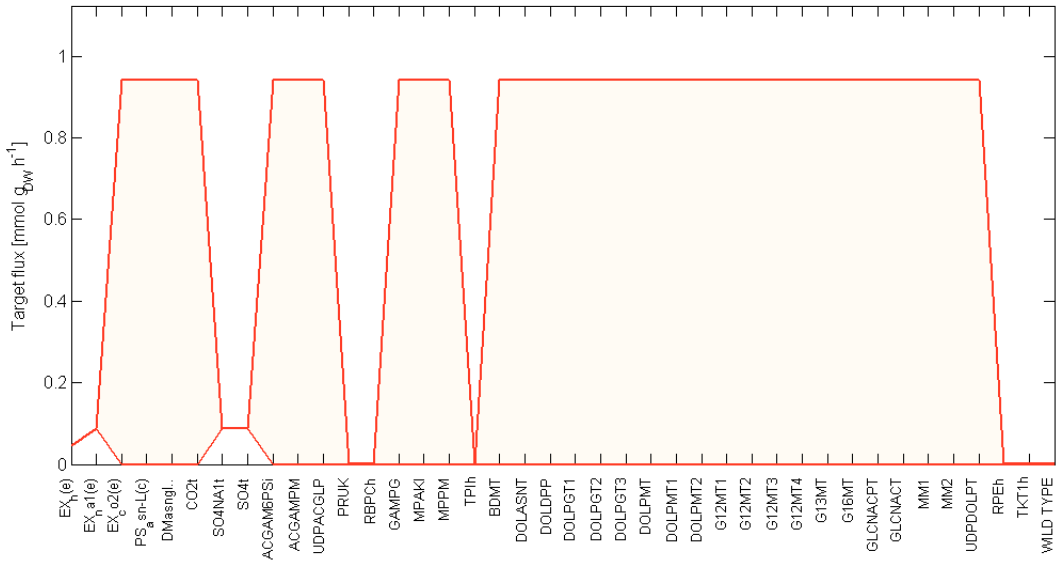

Figure 5.2. *Knocked-out genes that induced H_2 overproduction in* C. reinhardtii *according to OptKnock simulation results. On the x-axis, genes abbreviations are reported. For each gene, the range of values that the target flux can take is displayed by the area comprised by the upper and lower red line. Wild type has been included as control (no KOs).*

5.1.4. Analysis of results for H_2 overproduction

OptKnock simulation results of H_2 overproduction differ significantly from the case of TAGs. Again, N-glycan biosynthesis inactivation is a proposed strategy, though for the above considerations these KOs have been neglected. However, other interesting KOs resulted from the simulation, whose data are reported in Table 5.1.

Table 5.1 – *Data about KOs resulted in H_2 overproduction according to OptKnock simulation. For each KO gene, deleted reaction(s) and related pathway(s) are indicated. Key: ru5p-D - D-Ribulose 5-phosphate, rb15bp - D-Ribulose 1,5-bisphosphate, 3pg - 3-Phospho-D-glycerate, dhap - Dihydroxyacetone phosphate, g3p - D-glyceraldehyde 3-phosphate, xu5p-D - D-Xylulose 5-phosphate, r5p - α-D-Ribose 5-phosphate, s7p - Sedoheptulose 7-phosphate.*

Gene	Reaction	Formula	Pathway
CO2t	CO2 transport, extracellular	CO_2[e] <==> CO_2[c]	Transport
SO4NA1t	Sulfate:Na+ symporter, extracellular	SO_4^{2-} [e] + Na^+[e] <==> SO_4^{2-} [c] + Na^+ [c]	Transport
SO4t	Sulfate transport, extracellular	SO_4^{2-} [e] + H^+[e] <==> SO_4^{2-} [c] + H^+ [c]	Transport
PRUK	Phosphoribulokinase	ATP + ru5p-D -> ADP + H^+ + rb15bp	Carbon fixation
RBPCh	Ribulose-bisphosphate carboxylase	CO_2 + H_2O + rb15bp --> 2 3pg + 2 H^+	Carbon fixation
TPIh	Triosephosphate isomerase	dhap <==> g3p	Glycolysis
RPEh	D-Ribulose-5-Phosphate 3-Epimerase	ru5p-D <==> xu5p-D	Pentose phosphate pathway
TKT1h	Transketolase 1	r5p + xu5p-D <==> g3p + s7p	Pentose phosphate pathway

5.1.4.1 Sulfate transport inhibition strategy

KOs of genes which regulate the uptake of sulfate in cell metabolism indicate a net production of hydrogen roughly equal to v_{H2} = 0.0874 mmol g_{DW}^{-1} h^{-1}. There is no apparent link between elimination of sulfur from *C. reinhardtii* metabolism and activation of reaction (5.2). Sulfate (SO_4^{2-}) is a fundamental nutrient for the green alga, as it accounts for sulfation of lipids, polysaccharides and proteins and it is essential in the formation of cysteine and methionine (Stern and Harris, 2008, pg. 160-161). To be incorporated in these amino-acids as sulfur anion (S^{2-}), it must be reduced by means of ferredoxin:

$$HSO_3^- + 6\,\mathrm{FDXRd} \leftrightarrow 6\,\mathrm{FDXOx} + H_2S + 3H_2O + 6H^+ \qquad (5.3)$$

Thus, the activation of reaction (5.2) could be explained in terms of regeneration of oxidized ferredoxin which would otherwise accumulate in its reduced form when reaction (5.3) is inactivated.

Surprisingly, sulfur deprivation is nowadays the only ascertained condition which leads to hydrogen gas production by *C. reinhardtii* (Melis *et al.*, 2000). However, Melis and coworkers explained sustained hydrogen gas production upon sulfur deficiency in terms of regulatory mechanisms, rather than metabolic necessity, such as cofactors regeneration. Indeed, sulfur deprivation reversibly inactivates O_2 evolution, circumventing the severe O_2 sensitivity of the reversible hydrogenase, i.e. (5.2). On the other hand, given the complexity of cell metabolism, further analysis of the phenomenon could reveal new insights about hydrogen gas production upon sulfur deficiency.

5.1.4.2 Rubisco shunt inactivation strategy

PRUK, TPIh, RPEh, RBPCh, TKT1h all belong to the same pathway, namely the Rubisco shunt (Figure 5.3). By knocking out any of these enzymes, the Rubisco shunt would be inactivated. Rubisco shunt is an alternative pathway to glycolysis, compared to which it has a higher carbon conversion efficiency, yielding 20% more acetyl-CoA with 40% less carbon loss (Caspi *et al.*, 2007). In principle, Rubisco shunt is not essential for cell survival and thus, by erasing it, cell viability would not be affected. On the other hand, it is unclear the way its inactivation might induce reversible hydrogenase reaction to form H_2, as much as v_{H2} = 0.0015 mmol g_{DW}^{-1} h^{-1}. Since no evident correlation can be found, in the absence of more sophisticated analysis tools for such

wide networks, the inactivation of Rubisco shunt as a metabolic engineering strategy for hydrogen gas production should be carefully integrated with literature data and next validated experimentally.

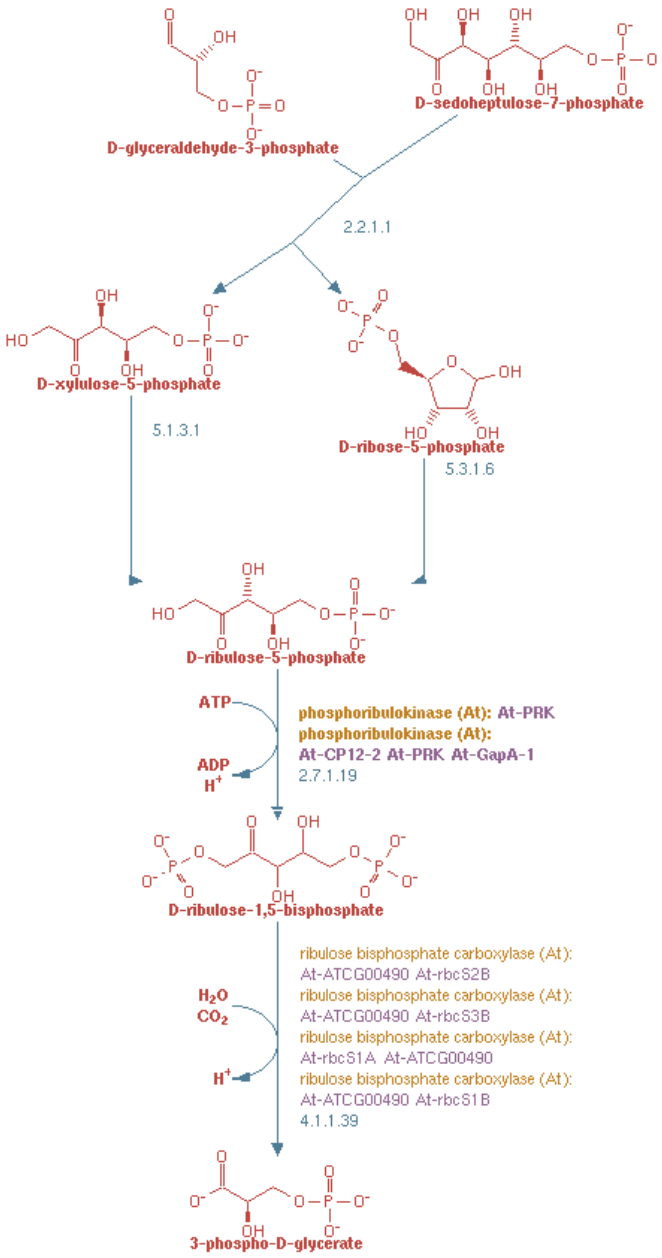

Figure 5.3. *Rubisco shunt pathway. Of all reactions within the pathway, only those KOs resulted from the previous simulation are represented. Retrieved from Metacyc.org (Caspi* et al.*, 2007).*

5.2 Design of *ad hoc* media

In principle, a properly designed medium could boost some cellular metabolic functions over others bypassing the activation of certain pathways required to form essential compounds for cell viability, since they would be now provided by the environment.

This means that a vital metabolite for biomass duplication could be directly incorporated in the metabolic network from the environment, and thus the cell would no more require to biosynthesize it. In the case of *C. reinhardtii* TAGs production, biomass growth has been selected as the target cellular function to maximize. An enhanced growth would increase outlet concentration of TAGs, minimizing the residence time required in the photobioreactior. On the other hand, by refining the initial medium composition, the costs associated with nutrients would increase as well.

To verify which compounds, if present in the environment, might enhance growth, the procedure described in §2.2.2 has been followed. FBA has been performed singularly for each metabolite comprised in iRC1080 model by adding a fictitious exchange reaction that simulates the possible uptake of that metabolite from the environment. iRC1080 was set in standard configuration (i.e. rich medium and photoautotrophic conditions) and solar light (lithosphere) at intensity $\Phi = 40$ μmol m^{-2} s^{-1} was selected. FBA has been run by using MATLAB™ script `optimizeCbmodel.m`, available within the COBRA Toolbox. The objective function to maximize was the autotrophic BOF. If no metabolites are supplemented to the rich medium, the physiological growth rate is $\mu = 0.0334$ h^{-1} (Figure 3.4). Only those metabolites that if available in the medium could achieve over 150% the physiological growth rate per mmol of absorbed metabolite were considered. Table 5.2 reports these metabolites along with the *in silico* predicted uptake flux and subsequent over-growth, both in absolute and relative values.

The simulation produced various results. Over 200 metabolites proved to boost biomass growth if only they could be incorporated in cell metabolism directly from the environment. Among these, 35 metabolites produced a relative over-growth 150% higher than physiological. Most of these belong to the class of lipids or saccharides. If incorporated in cell metabolism without need of being biosynthesized, they would slightly increase biomass growth (at most $\mu = 0.0450$ h^{-1}). However, they usually entail a drastic increase per unit of mmol absorbed, which goes from 216% to 2229%, higher than physiological growth. This may indicate that these metabolites are rate determining steps for biomass duplication. On the other hand, they are arguably difficult to introduce directly into the cell, since lipids and saccharides are normally big molecules which can be barely transported across cell membranes. The same can be said for fatty acids and nucleosides, although their potential increase in absolute growth rate is significantly higher. However, it is known that some membrane transporters of nucleosides are present in some organisms, like adenosine transporters in mammals (Thorn and Jarvis, 1996). Among the remaining metabolites, L-histidine catabolism intermediates as well as N-acetyl-L-glutamate have been analyzed in more detail in the following sections.

Table 5.2 –*Metabolites which if singularly present in the medium produced an* in silico *over growth per unit (mmol) of absorbed metabolite at least twice the physiological growth rate.*

Metabolite	Class	Over-growth [h^{-1}]	Uptake flux [mmol g_{DW}^{-1} h^{-1}]	% over-growth per unit of metabolite absorbed [$mmol^{-1}$ h^{-1}]
Triacylglycerol (16:0/18:1(11Z)/16:0)	Lipid	0,0385	0,0068	2228,82%
1-hexadecanoyl,2-(9Z)-octadecenoyl-sn-glycerol 3-phosphate	Lipid	0,0369	0,0065	1620,75%
1-(9Z)-octadecenoyl,2-hexadecanoyl-sn-glycerol 3-phosphate	Lipid	0,0369	0,0065	1620,47%
1,2-Diacyl-sn-glycerol (1-hexadecanoyl,2-(9Z)-octadecenoyl, 16:0/18:1(9Z))	Lipid	0,0369	0,0065	1619,49%
Stachyose	Saccharide	0,0406	0,0141	1517,73%
Manninotriose	Saccharide	0,0387	0,0135	1171,92%
Raffinose	Saccharide	0,0450	0,0313	1104,67%
Melibiitol	Saccharide	0,0409	0,0285	787,68%
ADPglucose	Saccharide	0,0368	0,0135	765,89%
Melibiose	Saccharide	0,0406	0,0283	758,86%
Epimelibiose	Saccharide	0,0406	0,0283	758,86%
Lactose	Saccharide	0,0406	0,0283	758,86%
1-hexadecanoyl-sn-glycerol 3-phosphate	Lipid	0,0408	0,0299	743,64%
Sucrose-6-phosphate	Saccharide	0,0398	0,0291	654,07%
α,α'-Trehalose 6-phosphate	Saccharide	0,0394	0,0288	625,53%
β-D-fructose 2-phosphate	Saccharide	0,0377	0,0276	466,15%
Galactosylglycerol	Lipid	0,0371	0,0258	425,64%
Tetradecanoate (n-C14:0)	Fatty acid	0,1550	0,8559	425,32%
Xanthosine	Nucleoside	0,4104	3,9800	283,60%
Inosine	Nucleoside	0,3755	3,8980	262,76%
Adenosine	Nucleoside	0,2436	2,4575	256,15%
N-Acetyl-L-glutamate	Arginine biosynthesis intermediate	0,4760	5,2610	251,88%
Guanosine	Nucleoside	0,2394	2,7063	227,87%
1-monoacylglycerol (16:0)	Lipid	0,2121	2,4750	216,17%
Urocanate	L-histidine catabolism intermediate	0,2121	2,7234	196,43%
4-Imidazolone-5-propanoate	L-histidine catabolism intermediate	0,2072	2,7076	192,18%
L-Histidinol	L-histidine catabolism intermediate	0,1484	1,8080	190,46%
L-Glutamate	Amino-acid	0,3847	5,6096	187,50%
Hexadecanoate (n-C16:0)	Fatty acid	0,1807	2,4283	181,64%
L-Glutamine	Amino-acid	0,2223	3,2850	172,18%

5.2.1. L-histidine catabolism intermediates

Metabolites that belong to this family are urocanate, 4-imidazolone-5-propanoate and L-histidinol. According to the above simulation results, they would enhance absolute growth as much as $\mu = 0.2121\ h^{-1}$, that is an over 6-fold increase with respect to the physiological growth rate in the same conditions. Table 5.3 summarizes some data regarding these compounds and their potential on biomass growth both in mass and molar basis.

Table 5.3 –*Intermediates in the L-histidine catabolism pathway which can potentially enhance* C. reinhardtii *growth in autotrophic conditions.*

Metabolite	Over-growth [h^{-1}]	MW [g mol^{-1}]	Uptake flux [mmol g_{DW}^{-1} h^{-1}]	[mg g_{DW}^{-1} h^{-1}]	% Over-growth per unit of metabolite absorbed	
					[$mmol^{-1}$ h^{-1}]	[mg^{-1} h^{-1}]
Urocanate	0,2121	138.12	2,7234	376	196,43%	1.42%
L-Histidinol	0,1484	141.17	1,8080	255	190,46%	1.35%
4-Imidazolone-5-propanoate	0,2072	156.14	2,7076	423	192,18%	1.23%

From Table 5.3, it is clear that the highest potential pertains to urocanate. The role that urocanate may play in boosting growth is correlated to its pathway. L-histidine catabolism is a linear pathway which can follow four different ways according to the organism. However, in all organisms in which this pathway is present, the early stage to degrade L-histidine is common and comprises the enzymatic conversion of the aminoacid into urocanate and, successively, 4-imidazolone-5-propionate (Figure 5.4). As regards *C. reinhardtii*, 4-imidazolone-5-propionate is then non-enzymatically transformed into 4-oxoglutaramate, which is in turn spontaneously converted into 2-oxoglutarate, also known as α-ketoglutaric acid (Stern and Harris, 2008). The latter metabolite is a key intermediate in the Krebs cycle. Hellio *et al.* proved and characterized the activity of both histidase (4.3.1.3) and urocanase (4.2.1.19), which are normally expressed when a more valuable and rapidly nitrogen source, such as ammonium, is being depleted (< 0.5 mM). When a primary nitrogen source is available, the L-histidine degradation pathway was shown to be deactivated. This experimental result was *in silico* confirmed with iRC1080 model under optimal growth in rich medium (data not shown). Thus, if urocanate was directly incorporated in *C. reinhardtii* metabolism, it could provide a valuable source to form α-ketoglutaric acid, incrementing Krebs cycle efficiency and thus boosting growth. Moreover, in the process of degrading L-histidine, important compounds for metabolism are formed, such as formate and ammonia, which can be recycled as a nitrogen source. Another question to address is why in optimal conditions *C. reinhardtii* should not take advantage of L-

histidine catabolism pathway. It can be hypothesized that histidine biosynthesis is *per se* energy or source intensive. In this fashion, an augmented production of the aminoacid for its successive degradation would be just a waste of energy or valuable sources. Even though the simulation suggested that a drastic increase in biomass growth rate would follow from the artificial introduction of urocanate in the cell metabolism, there are currently no proofs that the uptake of such metabolite is feasible in *C. reinhardtii*. Thus a series of experiments in urocanate replete medium should be conducted. The absorption of the compound by the cell should be sufficient *per se* to prove the hypothesis, since the two final reactions in the histidine degradation pathway are spontaneous (they are not enzymatic) and therefore should occur immediately.

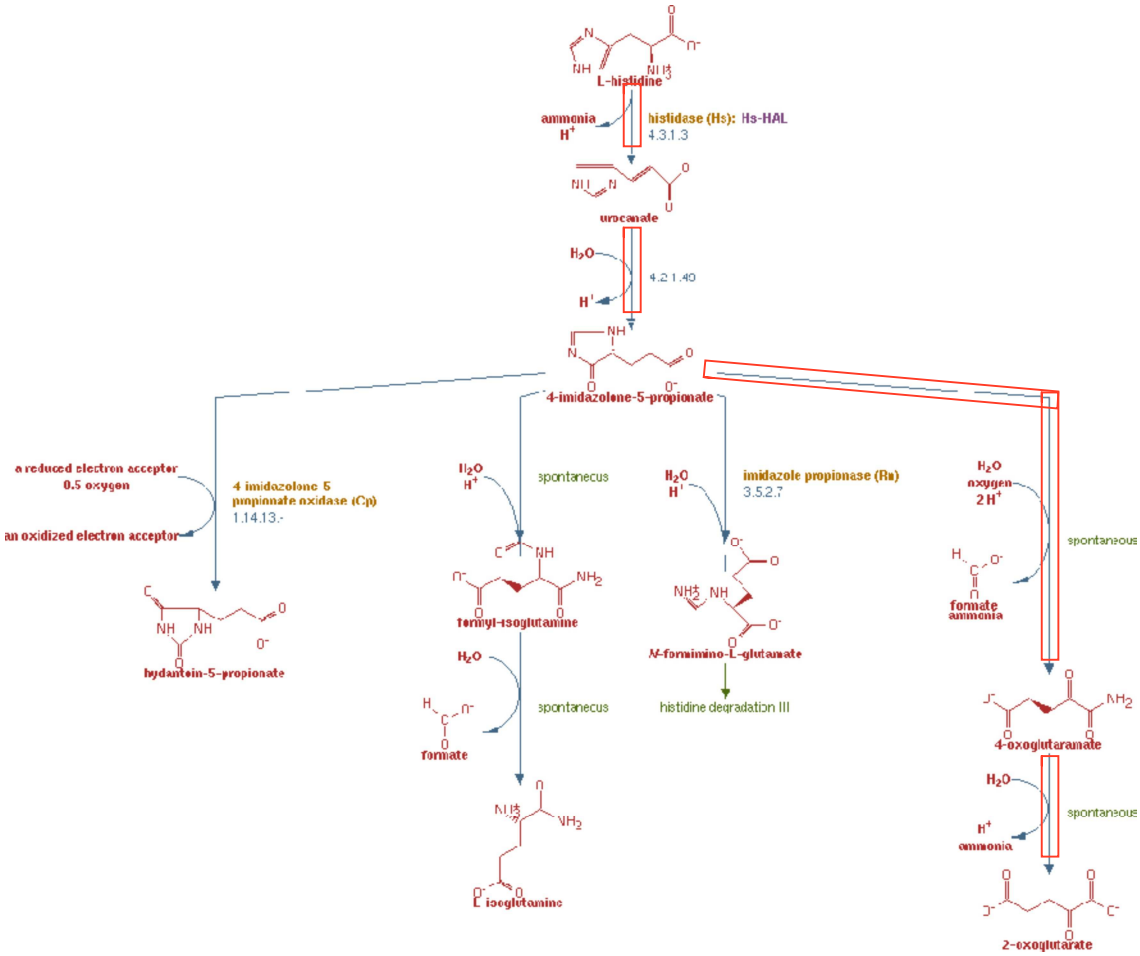

Figure 5.4. *L-histidine catabolism pathway. In* C. reinhardtii, *the path evidenced by red boxes is present, while the other paths are available for different organisms. Picture retrieved from Metacyc.org (Caspi* et al*, 2007).*

5.2.2. N-acetyl-L-glutamate

N-acetyl-L-glutamate (AcG) is an important intermediate in glutamate-derived aminoacids arginine and proline biosynthesis (Stern and Harris, 2008, pg. 134). As described by Figure 5.5, it is essential to form N-α-acetylornithine (AcO) which is a necessary step to produce ornithine, a precursor for arginine formation.

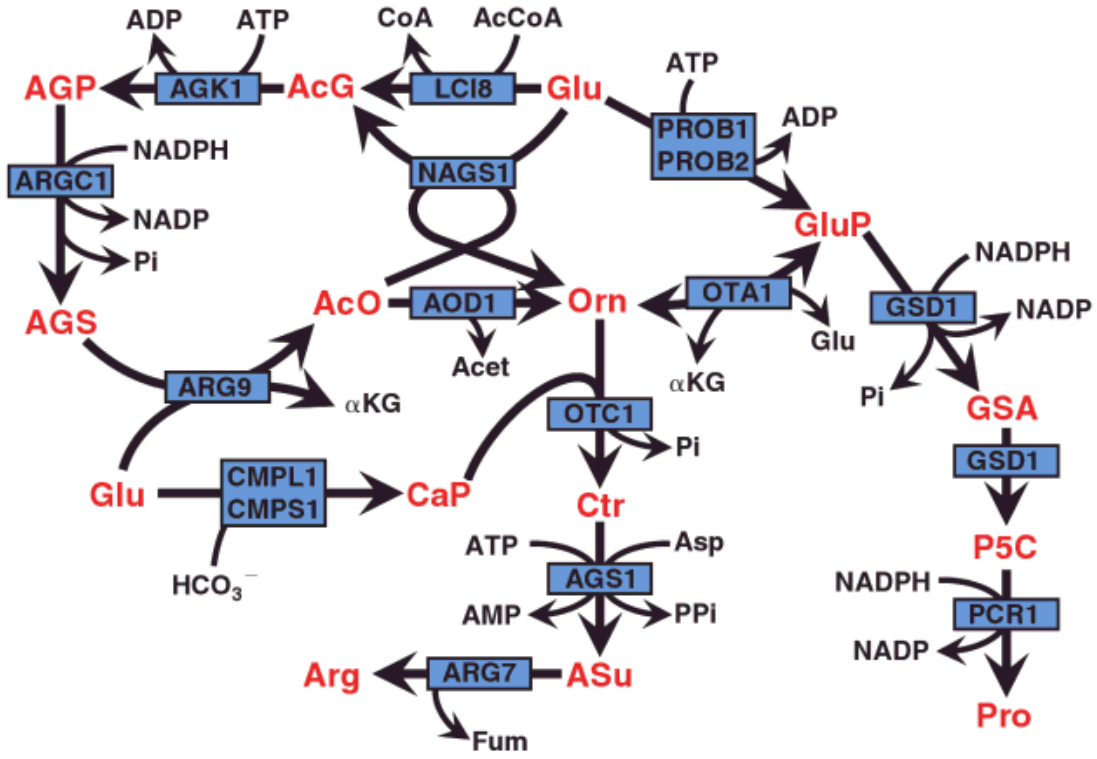

Figure 5.5. *Arginine and proline biosynthesis from glutamate in* C. reinhardtii *(Stern and Harris, 2006, pg. 134). Key: GluP – L-glutamyl-γ-phosphate, Pi – orthophosphate, GSA – glutamic-γ-semialdehyde, P5C – Δ^1-pyrroline-5-carboxylate, AcCoA – acetyl-coenzyme A; CoA – coenzyme A, AcG – N-acetyl-L-glutamate, AGP – N-acetylglutamyl-phosphate, AGS – N-acetylglutamate semialdehyde, AcO – N-α-acetylornithine, Orn – L-ornithine, CaP – carbamoyl-phosphate, Ctr – citrulline, PPi – pyrophosphate, ASu – L-arginino-succinate, Fum – fumarate. Blue boxes indicate enzyme abbreviation.*

According to the simulation results, direct incorporation of AcG in *C. reinhardtii* metabolism from the environment would accomplish the highest absolute growth rate, that is $\mu = 0.4760\ h^{-1}$. Considered that AcG molecular weight is $MW_{AcG} = 189.2\ g\ mol^{-1}$, its uptake flux is predicted to be $v_{AcG,in} = 5.26\ mmol\ g_{DW}^{-1}\ h^{-1}$ or $995.2\ mg\ g_{DW}^{-1}\ h^{-1}$. In mass units, the percent change in growth per unit of AcG absorbed with respect to the physiological growth rate is equal to $\%_{over\text{-}growth} = 1.33\%$, which is slightly less than the previous case with urocanate. This result can be explained in terms of a better exploitation of L-glutamate, which is no more needed to form AcG and, at the same time, reducing the necessity of expensive cofactors, such as acetyl-coenzyme A. Nevertheless, it can be argued why direct incorporation of arginine or other precursors

were not considered by the simulation to accomplish a similar result on growth rate. This can be explained in terms of the by-products formed by some enzymatic reactions occurring in AcG degradation to AcO, with particular emphasis on αKg, i.e. α-ketoglutaric acid. This justification does not hold for two other precursors, AGP and AGS, which in turn could prevent the waste of ATP and NADPH if they were incorporated in place of AcG. Indeed, the simulation actually contemplates this possibility, but given the elevated corresponding uptake flux, these compounds were neglected in results analysis.

An interesting point is the role of L-glutamate, which is reported to stimulate biomass growth as well, though to a slightly lesser extent than N-acetyl-L-glutamate (Table 5.2). In contrast, L-glutamate is a far easier compound to retrieve for a possible industrial application. Moreover, L-glutamate is known to be secreted by other microorganism under some conditions, such as *Corynebacterium glutamicum* (Hoischen and Krämer, 1990). Therefore, it might be possible to design synthetic communities of microorganisms which actually enhance *C. reinhardtii* growth without need of supplementing additional compounds to the initial medium.

5.3 Conclusions

Metabolic engineering is the field of chemical reaction engineering which deals with biological systems. As such, it can benefit of systems biology tools, such as FBA, to define a map of the metabolic flux distribution and analyze the effects of redirecting fluxes towards a desired reaction. Here, it has been reviewed some practical programs that are based on FBA and that can help to find new strategies to optimize biosynthesis of a target metabolite. These programs have been applied for the case study of TAGs production by *C. reinhardtii*. The optimization of TAGs formation has been explored through two methods, single gene knock-outs which lead to overproduction of desired compounds, and medium refinement by means of metabolites that the cell is no more required to synthesize in order to grow. The former method has been applied via a freely available algorithm called OptKnock. It produced no valuable results with regard to TAGs overproduction but gave interesting indications to stimulate H_2 synthesis, such as sulfur deprivation and inactivation of Rubisco shunt pathway. The latter method suggested that if an intermediate in L-histidine catabolism or arginine biosynthesis, like urocanate and N-acetyl-L-glutamate respectively, were present in the medium, then biomass growth would be greatly enhanced. Moreover, the results shed light on the role of L-glutamate in the microalgae metabolism and, as this amino acid is known to be secreted by other microorganisms, they suggested a way for the design of synthetic

communities aimed at providing *C. reinhardtii* the necessary nutrients for growth optimization.

In conclusion, this chapter collects the results that follow the application of those methods described in detail in §2.2 to an actual biological system of interest. Along with some poor if not disappointing conclusions (see §5.1.2), most of these applications revealed some indications that are not readily inferable at a first glance. Experimental validation is essential to prove these ambitious hypotheses. However the contribution of systems biology approaches in the formulation of the above mentioned metabolic engineering strategies is a precious indication.

Conclusions

Systems biology embraces the study of a biological system as a whole, by quantifying the interactions within the parts and enabling the analysis of variations within the network given some perturbations. As such, technical applications for biological systems which require the knowledge of the dynamics and potentiality of the network can greatly benefit from systems biology approaches. Chemical engineering appears to be one of the major disciplines which can take advantage of systems biology, as an increasing number of applications now deal with biotechnology, from fine chemicals production to renewable fuels.

In this work, diverse methods proper of the emergent systems biology were investigated and coupled to solve typical problems of chemical engineering. The application of flux balance analysis has been reviewed to cope for the lack of kinetic data in biological systems such as the cell, and a new approach that solves bioreactor design equations by using dynamic FBA has been proposed as an alternative to the black box approach (§2.1). Together with reactor design, chemical reaction engineering is another field which can profit by a careful application of FBA-based algorithms. In this fashion, the biological network can be perturbed in order to optimize the reaction flux of interest. Two main perturbations have been reviewed, that are single gene knock-out (§2.2.1) and refinement of growth medium, developed with a novel method. The results may constitute valuable indications of metabolic engineering strategies and can be used as a starting point for an experimental campaign aimed at engineering the biological system under study to maximize the desired reaction.

In order to validate the above mentioned systems biology approaches to chemical engineering problems, a case study has been considered, i.e. the production of biofuel precursors, namely triacylglycerols, by green microalga *C. reinhardtii*, whose metabolic network has been recently reconstructed *in silico*. The design of a photobioreactor has been efficiently conducted and it produced interesting results for conceptual design, such as the components concentration profile along the tube, the trend of outlet concentration with residence time, and the nutrients consumption (§4.4). Direct improvement of valuable products by modifications of *C. reinhardtii* metabolic network has been performed as well. Some already known experimental results have been predicted, such as sulfur deficiency to maximize H_2 biosynthesis, an experimental result which has been a major accomplishment of modern metabolic engineering (Mellis *et al.*, 2000). In terms of enriched growth medium, biomass growth was seen to be greatly

enhanced by direct incorporation in *C. reinhardtii* metabolic network of some minor intermediates, like urocanate and N-acetyl-L-glutamate. Also L-glutamate proved to have a key role in boosting growth rate, and given that this amino acid is known to be secreted by other microorganisms, this result may pave the way to the design of synthetic communities too.

As it is widely recognized that biotechnology owns the potential to sustain the transition from oil-derived chemicals to renewable alternatives, it is expected that chemical engineering will follow a similar trend. If so, this work has investigated how major applications like reactor design and metabolic engineering can be successfully coupled with systems biology approaches to find the most rational engineering of a biological system among the uncountable options that nature has provided.

Appendix A
Abbreviations

1. List of acronyms

Most of biochemical compounds in this work are indicated with an acronym. Here, only those acronyms that are found in the main text are considered. Acronyms that appear in captions are explained inside key of the caption.

AcG – N-acetyl-L-glutamate;
AcO – N-α-acetylornithine;
ATP – Adenosine triphosphate;
BOF – Biomass objective function;
CSTR – Continuously stirred tank reactor;
CTP – Cytosine triphosphate;
dFBA – Dynamic flux balance analysis;
FBA – Flux balance analysis;
GAM – Growth-associated ATP maintenance;
GENRE – Genome-based metabolic reconstruction;
GPR – Gene-protein-reaction (association);
iRC1080 – Latest *C. reinhardtii* genome scale metabolic model available;
KO – Knock out;
MOMA – Minimization of metabolic adjustment;
NGAM – Non-growth-associated ATP maintenance;
PBR – Photobioreactor;
PFR – Plug flow reactor;
TAGs – Triacylglycerols;
UTP – Uracil triphosphate;

2. List of variables

All variables in the main text are written in *italics* (except for vectors and matrices which are in **bold**). Subscript *0* stands for "initial", while subscript *f* stands for "final". Subscript * stands for "maximum".

2.1 Latin symbols

$b_{i,j}$ – Exchange flux of metabolite *i* in reaction *j* (generally [mmol g_{DW}^{-1} h^{-1}]);
C_i – Concentration of metabolite *i* [mol L^{-1}] or [mmol L^{-1}];

C_P – Product concentration [mmol L^{-1}];
C_S – Substrate concentration [mmol L^{-1}];
C_X – Biomass concentration [g_{DW} L^{-1}];
k_M – Monod constant [mol L^{-1}] or [mmol L^{-1}];
P – Productivity [g $L_{reactor}^{-1}$];
r_i – Reaction rate of formation/consumption of metabolite i [mol L^{-1} s^{-1}];
S – Stoichiometric matrix of the reconstructed metabolic network;
t – Vector of metabolite concentrations variation in time;
v – Vector of all fluxes within the reconstructed metabolic network;
$v_{i,j}$ – (General) flux of metabolite i in reaction j (generally [mmol g_{DW}^{-1} h^{-1}]);
$\dot{V}$ – Volumetric flow rate (generally [L h^{-1}]);
Y – TAGs outlet concentration [g_{TAGs} L^{-1}];
$Y_{i/j}$ – Yield of metabolite i over variation of metabolite j [-] or [g_{DW} $mmol^{-1}$];
Z – Objective function;

2.2 Greek symbols

α_i – Lower bound of flux i (generally [mmol g_{DW}^{-1} h^{-1}]);
β_i – Upper bound of flux i (generally [mmol g_{DW}^{-1} h^{-1}]);
μ – Specific growth rate or biomass duplication rate [h^{-1}];
τ – Residence time [h^{-1}];
Φ – Photon flux intensity [µmol m^{-2} s^{-1}];

Appendix B
Nomenclature conventions in iRC1080

List of suffixes used in iRC1080

In iRC1080 model, suffixes are used to refer a metabolite to the cellular compartment in which it can be found. A unique metabolite can participate to reactions taking place in multiple compartments.

[c] – Cytoplasm;
[e] – Extracellular;
[f] – Flagellum;
[g] – Golgi apparatus;
[h] – Chloroplast;
[m] – Mitochondria;
[n] – Nucleus;
[s] – Eyespot;
[u] – Thylakoid lumen;
[x] – Glyoxysome;

Appendix C
Objective functions

The objective function is a key requirement to run FBA. In the present work, either a real reaction of the reconstructed metabolic network or a fictitious reaction has been used in the simulations, according to the simulation goal. In this appendix, the autotrophic BOF and the TAGs objective function are explicitly decomposed in their components.

1. Autotrophic biomass objective function

In Table C.1., all precursors reported to make up 1 gram of dry biomass of *C. reinhardtii* in autotrophic conditions are listed along with their stoichiometric coefficient in the biomass objective function. These coefficients represent each precursor proportion in the dry cell (in mmol per gram of dry weight).

Table C.1. *Precursors in autotrophic BOF and their stoichiometric coefficients.*

Biomass precursor	Stoichiometric coefficient [mmol g_{DW}^{-1}]
Alanine	0.27375
Arginine	0.15020
Asparagine	0.06783
Aspartate	0.06783
Cysteine	0.00242
Glutamine	0.08116
Glutamate	0.08116
Glycine	0.10296
Histidine	0.00121
Isoleucine	0.03270
Leucine	0.08237
Lysine	0.01817
Methionine	0.00242
Phenylalanine	0.03392
Proline	0.04724
Serine	0.02059
Threonine	0.08237
Tryptophan	0.00121
Tyrosine	0.00121
Valine	0.05935
dATP	0.00218
dCTP	0.00388
dGTP	0.00388
dTTP	0.00218
ATP	0.03965*
CTP	0.07048*
GTP	0.07048*

UTP	0.03965*
Starch	0.00641
Mannose	0.32836
Arabinose	0.52414
Galactose	0.69700
Monogalactosyldiacylglycerol (1-(9Z,12Z,15Z)-octadecatrienoyl,2-(4Z,7Z,10Z,13Z)-hexadecatetraenoyl, 18:3(9Z,12Z,15Z)/16:4(4Z,7Z,10Z,13Z))	0.02845
Monogalactosyldiacylglycerol (1-(9Z,12Z,15Z)-octadecatrienoyl,2-(7Z,10Z,13Z)-hexadecatrienoyl, 18:3(9Z,12Z,15Z)/16:3(7Z,10Z,13Z))	0.00324
Monogalactosyldiacylglycerol (1-(9Z,12Z,15Z)-octadecatrienoyl,2-(4Z,7Z,10Z)-hexadecatrienoyl, 18:3(9Z,12Z,15Z)/16:3(4Z,7Z,10Z))	0.00324
Monogalactosyldiacylglycerol (1-(9Z,12Z)-octadecadienoyl,2-(4Z,7Z,10Z,13Z)-hexadecatetraenoyl, 18:2(9Z,12Z)/16:4(4Z,7Z,10Z,13Z))	0.00324
Monogalactosyldiacylglycerol (1-(9Z,12Z)-octadecadienoyl,2-(7Z,10Z,13Z)-hexadecatrienoyl, 18:2(9Z,12Z)/16:3(7Z,10Z,13Z))	0.00040
Monogalactosyldiacylglycerol (1-(9Z,12Z)-octadecadienoyl,2-(4Z,7Z,10Z)-hexadecatrienoyl, 18:2(9Z,12Z)/16:3(4Z,7Z,10Z))	0.00040
Monogalactosyldiacylglycerol (1-(9Z,12Z,15Z)-octadecatrienoyl,2-(7Z,10Z)-hexadecadienoyl, 18:3(9Z,12Z,15Z)/16:2(7Z,10Z))	0.00040
Monogalactosyldiacylglycerol (1-(9Z,12Z)-octadecadienoyl,2-(7Z,10Z)-hexadecadienoyl, 18:2(9Z,12Z)/16:2(7Z,10Z))	0.00040
Monogalactosyldiacylglycerol (1-(9Z,12Z)-octadecadienoyl,2-(7Z)-hexadecenoyl, 18:2(9Z,12Z)/16:1(7Z))	0.00020
Monogalactosyldiacylglycerol (1-(9Z,12Z)-octadecadienoyl,2-(9Z)-hexadecenoyl, 18:2(9Z,12Z)/16:1(9Z))	0.00020
Monogalactosyldiacylglycerol (1-(9Z,12Z,15Z)-octadecatrienoyl,2-hexadecanoyl, 18:3(9Z,12Z,15Z)/16:0)	0.00020
Monogalactosyldiacylglycerol (1-(9Z,12Z)-octadecadienoyl,2-hexadecanoyl, 18:2(9Z,12Z)/16:0)	0.00020
Digalactosyldiacylglycerol (1-(9Z,12Z,15Z)-octadecatrienoyl,2-(4Z,7Z,10Z,13Z)-hexadecatetraenoyl, 18:3(9Z,12Z,15Z)/16:4(4Z,7Z,10Z,13Z))	0.00027
Digalactosyldiacylglycerol (1-(9Z,12Z,15Z)-octadecatrienoyl,2-(7Z,10Z,13Z)-hexadecatrienoyl, 18:3(9Z,12Z,15Z)/16:3(7Z,10Z,13Z))	0.00074
Digalactosyldiacylglycerol (1-(9Z,12Z,15Z)-octadecatrienoyl,2-(4Z,7Z,10Z)-hexadecatrienoyl, 18:3(9Z,12Z,15Z)/16:3(4Z,7Z,10Z))	0.00074
Digalactosyldiacylglycerol (1-(9Z,12Z,15Z)-octadecatrienoyl,2-(7Z,10Z)-hexadecadienoyl, 18:3(9Z,12Z,15Z)/16:2(7Z,10Z))	0.00054
Digalactosyldiacylglycerol (1-(9Z,12Z)-octadecadienoyl,2-(7Z,10Z,13Z)-hexadecatrienoyl, 18:2(9Z,12Z)/16:3(7Z,10Z,13Z))	0.00027
Digalactosyldiacylglycerol (1-(9Z,12Z)-octadecadienoyl,2-(4Z,7Z,10Z)-hexadecatrienoyl, 18:2(9Z,12Z)/16:3(4Z,7Z,10Z))	0.00027
Digalactosyldiacylglycerol (1-(9Z,12Z)-octadecadienoyl,2-(7Z,10Z)-hexadecadienoyl, 18:2(9Z,12Z)/16:2(7Z,10Z))	0.00054
Digalactosyldiacylglycerol (1-(9Z)-octadecenoyl,2-(7Z,10Z,13Z)-hexadecatrienoyl, 18:1(9Z)/16:3(7Z,10Z,13Z))	0.00020
Digalactosyldiacylglycerol (1-(9Z)-octadecenoyl,2-(4Z,7Z,10Z)-hexadecatrienoyl, 18:1(9Z)/16:3(4Z,7Z,10Z))	0.00020
Digalactosyldiacylglycerol (1-(9Z)-octadecenoyl,2-(7Z,10Z)-hexadecadienoyl, 18:1(9Z)/16:2(7Z,10Z))	0.00040
Digalactosyldiacylglycerol (1-(9Z,12Z,15Z)-octadecatrienoyl,2-hexadecanoyl, 18:3(9Z,12Z,15Z)/16:0)	0.00240
Digalactosyldiacylglycerol (1-(9Z,12Z)-octadecadienoyl,2-(7Z)-hexadecenoyl, 18:2(9Z,12Z)/16:1(7Z))	0.00030
Digalactosyldiacylglycerol (1-(9Z,12Z)-octadecadienoyl,2-(9Z)-	0.00030

hexadecenoyl, 18:2(9Z,12Z)/16:1(9Z))	
Digalactosyldiacylglycerol (1-(9Z,12Z)-octadecadienoyl,2-hexadecanoyl, 18:2(9Z,12Z)/16:0)	0.00333
Digalactosyldiacylglycerol (1-(9Z)-octadecenoyl,2-(7Z)-hexadecenoyl, 18:1(9Z)/16:1(7Z))	0.00030
Digalactosyldiacylglycerol (1-(9Z)-octadecenoyl,2-(9Z)-hexadecenoyl, 18:1(9Z)/16:1(9Z))	0.00030
Digalactosyldiacylglycerol (1-(9Z)-octadecenoyl,2-hexadecanoyl, 18:1(9Z)/16:0)	0.00173
Diacylglyceryl-N,N,N-trimethylhomoserine (18:3(9Z,12Z,15Z)/18:4(5Z,9Z,12Z,15Z))	0.00041
Diacylglyceryl-N,N,N-trimethylhomoserine (18:3(9Z,12Z,15Z)/18:3(5Z,9Z,12Z))	0.00108
Diacylglyceryl-N,N,N-trimethylhomoserine (18:2(9Z,12Z)/18:4(5Z,9Z,12Z,15Z))	0.00008
Diacylglyceryl-N,N,N-trimethylhomoserine (18:2(9Z,12Z)/18:3(5Z,9Z,12Z))	0.00162
Diacylglyceryl-N,N,N-trimethylhomoserine (18:1(9Z)/18:4(5Z,9Z,12Z,15Z))	0.00027
Diacylglyceryl-N,N,N-trimethylhomoserine (18:1(11Z)/18:4(5Z,9Z,12Z,15Z))	0.00027
Diacylglyceryl-N,N,N-trimethylhomoserine (16:0/18:4(5Z,9Z,12Z,15Z))	0.00209
Diacylglyceryl-N,N,N-trimethylhomoserine (18:2(9Z,12Z)/18:2(9Z,12Z))	0.00040
Diacylglyceryl-N,N,N-trimethylhomoserine (18:3(9Z,12Z,15Z)/18:1(9Z))	0.00034
Diacylglyceryl-N,N,N-trimethylhomoserine (18:2(9Z,12Z)/18:1(9Z))	0.00007
Diacylglyceryl-N,N,N-trimethylhomoserine (18:3(9Z,12Z,15Z)/18:1(11Z))	0.00034
Diacylglyceryl-N,N,N-trimethylhomoserine (18:2(9Z,12Z)/18:1(11Z))	0.00007
Diacylglyceryl-N,N,N-trimethylhomoserine (18:1(9Z)/18:3(5Z,9Z,12Z))	0.00040
Diacylglyceryl-N,N,N-trimethylhomoserine (18:1(11Z)/18:3(5Z,9Z,12Z))	0.00040
Diacylglyceryl-N,N,N-trimethylhomoserine (16:0/18:3(5Z,9Z,12Z))	0.00390
Diacylglyceryl-N,N,N-trimethylhomoserine (18:1(9Z)/18:2(9Z,12Z))	0.00020
Diacylglyceryl-N,N,N-trimethylhomoserine (18:1(11Z)/18:2(9Z,12Z))	0.00020
Diacylglyceryl-N,N,N-trimethylhomoserine (18:1(9Z)/18:1(9Z))	0.00007
Diacylglyceryl-N,N,N-trimethylhomoserine (18:1(9Z)/18:1(11Z))	0.00007
Diacylglyceryl-N,N,N-trimethylhomoserine (18:1(11Z)/18:1(9Z))	0.00007
Diacylglyceryl-N,N,N-trimethylhomoserine (18:1(11Z)/18:1(11Z))	0.00007
Diacylglyceryl-N,N,N-trimethylhomoserine (16:0/18:2(9Z,12Z))	0.00111
Diacylglyceryl-N,N,N-trimethylhomoserine (16:0/18:1(9Z))	0.00008
Diacylglyceryl-N,N,N-trimethylhomoserine (16:0/18:1(11Z))	0.00008
Sulfoquinovosyldiacylglycerol (1-(9Z,12Z,15Z)-octadecatrienoyl,2-hexadecanoyl, 18:3(9Z,12Z,15Z)/16:0)	0.00078
Sulfoquinovosyldiacylglycerol (1-(9Z,12Z)-octadecadienoyl,2-hexadecanoyl, 18:2(9Z,12Z)/16:0)	0.00073
Sulfoquinovosyldiacylglycerol (1-(9Z)-octadecenoyl,2-hexadecanoyl, 18:1(9Z)/16:0)	0.00048
Sulfoquinovosyldiacylglycerol (1-(11Z)-octadecenoyl,2-hexadecanoyl, 18:1(11Z)/16:0)	0.00048
Sulfoquinovosyldiacylglycerol (dihexadecanoyl, n-C16:0)	0.00245
2'-O-all-cis-5,9,12-octadecatrienoyl-sulfoquinovosyldiacylglycerol (2'-18:3(5,9,12)/18:1(9Z)/16:0)	0.00118
2'-O-all-cis-5,9,12-octadecatrienoyl-sulfoquinovosyldiacylglycerol (2'-18:3(5,9,12)/18:1(11Z)/16:0)	0.00118
2'-O-all-cis-5,9,12-octadecatrienoyl-sulfoquinovosyldiacylglycerol (2'-18:3(5,9,12)/18:2(9Z,12Z)/16:0)	0.00118
2'-O-all-cis-5,9,12-octadecatrienoyl-sulfoquinovosyldiacylglycerol (2'-	0.00119

18:3(5,9,12)/18:3(9Z,12Z,15Z)/16:0)	
2'-O-all-cis-5,9,12,15-octadecatetraenoyl-sulfoquinovosyldiacylglycerol (2'-18:4(5,9,12,15)/18:1(9Z)/16:0)	0.00118
2'-O-all-cis-5,9,12,15-octadecatetraenoyl-sulfoquinovosyldiacylglycerol (2'-18:4(5,9,12,15)/18:1(11Z)/16:0)	0.00118
2'-O-all-cis-5,9,12,15-octadecatetraenoyl-sulfoquinovosyldiacylglycerol (2'-18:4(5,9,12,15)/18:2(9Z,12Z)/16:0)	0.00119
2'-O-all-cis-5,9,12,15-octadecatetraenoyl-sulfoquinovosyldiacylglycerol (2'-18:4(5,9,12,15)/18:3(9Z,12Z,15Z)/16:0)	0.00119
Phosphatidylglycerol (1-(9Z,12Z,15Z)-octadecatrienoyl,2-(3E)-hexadecenoyl, 18:3(9Z,12Z,15Z)/16:1(3E))	0.00123
Phosphatidylglycerol (1-(9Z,12Z,15Z)-octadecatrienoyl,2-hexadecanoyl, 18:3(9Z,12Z,15Z)/16:0)	0.00074
Phosphatidylglycerol (1-(9Z,12Z)-octadecadienoyl,2-(3E)-hexadecenoyl, 18:2(9Z,12Z)/16:1(3E))	0.00336
Phosphatidylglycerol (1-(9Z,12Z)-octadecadienoyl,2-hexadecanoyl, 18:2(9Z,12Z)/16:0)	0.00074
Phosphatidylglycerol (1-(9Z)-octadecenoyl,2-(3E)-hexadecenoyl, 18:1(9Z)/16:1(3E))	0.00045
Phosphatidylglycerol (1-(11Z)-octadecenoyl,2-(3E)-hexadecenoyl, 18:1(11Z)/16:1(3E))	0.00045
Phosphatidylglycerol (1-(9Z)-octadecenoyl,2-hexadecanoyl, 18:1(9Z)/16:0)	0.00061
Phosphatidylglycerol (1-(11Z)-octadecenoyl,2-hexadecanoyl, 18:1(11Z)/16:0)	0.00061
Phosphatidylethanolamine (18:2(9Z,12Z)/18:3(5Z,9Z,12Z))	0.00020
Phosphatidylethanolamine (18:1(9Z)/18:4(5Z,9Z,12Z,15Z))	0.00000
Phosphatidylethanolamine (18:1(9Z)/18:3(5Z,9Z,12Z))	0.00009
Phosphatidylethanolamine (18:1(11Z)/18:4(5Z,9Z,12Z,15Z))	0.00013
Phosphatidylethanolamine (18:1(11Z)/18:3(5Z,9Z,12Z))	0.00311
Phosphatidylethanolamine (18:0/18:4(5Z,9Z,12Z,15Z))	0.00065
Phosphatidylethanolamine (18:0/18:3(5Z,9Z,12Z))	0.00235
1-Phosphatidyl-D-myo-inositol (1-(11Z)-octadecenoyl,2-hexadecanoyl, 18:1(11Z)/16:0)	0.00192
1-Phosphatidyl-D-myo-inositol (1-(9Z)-octadecenoyl,2-hexadecanoyl, 18:1(9Z)/16:0)	0.00026
Triacylglycerol (16:0/18:1(11Z)/16:0)	0.00133
Triacylglycerol (16:0/18:1(9Z)/16:0)	0.00133
Triacylglycerol (18:0/18:1(9Z)/16:0)	0.00129
Triacylglycerol (18:1(11Z)/18:1(11Z)/16:0)	0.00129
Triacylglycerol (18:1(11Z)/18:1(9Z)/16:0)	0.00129
Triacylglycerol (18:1(9Z)/18:1(11Z)/16:0)	0.00129
Triacylglycerol (18:1(9Z)/18:1(9Z)/16:0)	0.00129
Triacylglycerol (16:0/18:1(11Z)/18:0)	0.00129
Triacylglycerol (16:0/18:1(9Z)/18:0)	0.00129
Triacylglycerol (18:0/18:1(9Z)/18:0)	0.00124
Triacylglycerol (18:1(11Z)/18:1(11Z)/18:0)	0.00125
Triacylglycerol (18:1(11Z)/18:1(9Z)/18:0)	0.00125
Triacylglycerol (18:1(9Z)/18:1(11Z)/18:0)	0.00125
Triacylglycerol (18:1(9Z)/18:1(9Z)/18:0)	0.00125
Triacylglycerol (16:0/18:1(11Z)/18:1(11Z))	0.00129
Triacylglycerol (16:0/18:1(9Z)/18:1(11Z))	0.00129
Triacylglycerol (18:0/18:1(9Z)/18:1(11Z))	0.00125
Triacylglycerol (18:1(11Z)/18:1(11Z)/18:1(11Z))	0.00125
Triacylglycerol (18:1(11Z)/18:1(9Z)/18:1(11Z))	0.00125

Triacylglycerol (18:1(9Z)/18:1(11Z)/18:1(11Z))	0.00125
Triacylglycerol (18:1(9Z)/18:1(9Z)/18:1(11Z))	0.00125
Triacylglycerol (16:0/18:1(11Z)/18:1(9Z))	0.00129
Triacylglycerol (16:0/18:1(9Z)/18:1(9Z))	0.00129
Triacylglycerol (18:0/18:1(9Z)/18:1(9Z))	0.00125
Triacylglycerol (18:1(11Z)/18:1(11Z)/18:1(9Z))	0.00125
Triacylglycerol (18:1(11Z)/18:1(9Z)/18:1(9Z))	0.00125
Triacylglycerol (18:1(9Z)/18:1(11Z)/18:1(9Z))	0.00125
Triacylglycerol (18:1(9Z)/18:1(9Z)/18:1(9Z))	0.00125
Triacylglycerol (16:0/18:1(11Z)/18:3(5Z,9Z,12Z))	0.00129
Triacylglycerol (16:0/18:1(9Z)/18:3(5Z,9Z,12Z))	0.00129
Triacylglycerol (18:0/18:1(9Z)/18:3(5Z,9Z,12Z))	0.00125
Triacylglycerol (18:1(11Z)/18:1(11Z)/18:3(5Z,9Z,12Z))	0.00126
Triacylglycerol (18:1(11Z)/18:1(9Z)/18:3(5Z,9Z,12Z))	0.00126
Triacylglycerol (18:1(9Z)/18:1(11Z)/18:3(5Z,9Z,12Z))	0.00126
Triacylglycerol (18:1(9Z)/18:1(9Z)/18:3(5Z,9Z,12Z))	0.00126
Triacylglycerol (16:0/18:1(11Z)/18:4(5Z,9Z,12Z,15Z))	0.00130
Triacylglycerol (16:0/18:1(9Z)/18:4(5Z,9Z,12Z,15Z))	0.00130
Triacylglycerol (18:0/18:1(9Z)/18:4(5Z,9Z,12Z,15Z))	0.00126
Triacylglycerol (18:1(11Z)/18:1(11Z)/18:4(5Z,9Z,12Z,15Z))	0.00126
Triacylglycerol (18:1(11Z)/18:1(9Z)/18:4(5Z,9Z,12Z,15Z))	0.00126
Triacylglycerol (18:1(9Z)/18:1(11Z)/18:4(5Z,9Z,12Z,15Z))	0.00126
Triacylglycerol (18:1(9Z)/18:1(9Z)/18:4(5Z,9Z,12Z,15Z))	0.00126
Acetic acid	0.03706
Propionic acid	0.03005
Butyric acid	0.02526
Glycerol	0.01208
Chlorophyll a	0.01008
Chlorophyll b	0.01655
Retinal (rhodopsin-bound)	0.00000
alpha-Carotene	0.00050
Antheraxanthin	0.00010
beta-Carotene	0.00141
Loroxanthin	0.00066
Lutein	0.00126
Neoxanthin	0.00055
Violaxanthin	0.00035
Zeaxanthin	0.00030
ATP maintenance (growth-associated)	29.89000

*Note: the precursors coefficients denoted with an asterisk * were fixed according to data previously reported by Boyle and Morgan (2009) in order to normalize all coefficients in the BOF to 1 g.*

2. Summary of the autotrophic BOF

In Table C.2, all precursors in Table C.1 are lumped into biochemical classes and their relative proportion in 1 g of dry biomass (both on molar and weight basis) is reported.

Table C.2. *Precursors in autotrophic BOF lumped into biochemical classes, and their proportion in 1* g_{DW} *(on molar and weight basis).*

Class		mg/g_{DW}	$mmol/g_{DW}$
Protein		149.1110	1.2101
RNA		109.5036	0.2203
DNA		5.8481	0.0121
Carbohydrates		575.2248	1.5559
Fatty acids		6.5842	0.0924
Glycerol		1.1128	0.0121
Chlorophyll		24.0248	0.0266
Retinal		0.0003	0.0001
Xanthophyll		2.9053	0.0051
Lipids		125.6851	0.1517
of which	ASQDCA	10.2246	0.0095
	DGDG	11.7306	0.0128
	DGTS	10.2246	0.0137
	MGDG	30.2777	0.0406
	PAIL	1.8230	0.0022
	PE	4.8349	0.0065
	PG	6.1030	0.0082
	SQDG	3.9630	0.0049
	TAG	46.5037	0.0533
Total biomass		1000	3.28

Note that by fixing precursors coefficients indicated by an asterisk in Table C.1, total biomass was normalized, as results from the second column of Table C.2.

3. TAGs objective function

In Table C.3., all precursors assumed to make up 1 gram of neutral lipids under nitrogen deficiency conditions are listed along with their stoichiometric coefficient in the objective function.

Table C.3. *Precursors in TAGs objective function and their stoichiometric coefficients.*

Biomass precursor	Stoichiometric coefficent [mmol g_{DW}^{-1}]
Triacylglycerol (16:0/18:1(11Z)/16:0)	23,8092913186678
Triacylglycerol (16:0/18:1(11Z)/18:0)	23,8093809868631
Triacylglycerol (16:0/18:1(11Z)/18:1(11Z))	23,8094951741635
Triacylglycerol (16:0/18:1(11Z)/18:1(9Z))	23,8094951741635
Triacylglycerol (16:0/18:1(11Z)/18:3(5Z,9Z,12Z))	23,8094468076432
Triacylglycerol (16:0/18:1(11Z)/18:4(5Z,9Z,12Z,15Z))	23,8092829474617
Triacylglycerol (16:0/18:1(9Z)/16:0)	23,8092913186678
Triacylglycerol (16:0/18:1(9Z)/18:0)	23,8093809868631
Triacylglycerol (16:0/18:1(9Z)/18:1(11Z))	23,8094951741635
Triacylglycerol (16:0/18:1(9Z)/18:1(9Z))	23,8094951741635
Triacylglycerol (16:0/18:1(9Z)/18:3(5Z,9Z,12Z))	23,8094468076432
Triacylglycerol (16:0/18:1(9Z)/18:4(5Z,9Z,12Z,15Z))	23,8092829474617
Triacylglycerol (18:0/18:1(9Z)/16:0)	23,8093809868631
Triacylglycerol (18:0/18:1(9Z)/18:0)	23,8094649995080
Triacylglycerol (18:0/18:1(9Z)/18:1(11Z))	23,8095757688329
Triacylglycerol (18:0/18:1(9Z)/18:1(9Z))	23,8095757688329
Triacylglycerol (18:0/18:1(9Z)/18:3(5Z,9Z,12Z))	23,8095293057068
Triacylglycerol (18:0/18:1(9Z)/18:4(5Z,9Z,12Z,15Z))	23,8093708484060
Triacylglycerol (18:1(11Z)/18:1(11Z)/16:0)	23,8094951741635
Triacylglycerol (18:1(11Z)/18:1(11Z)/18:0)	23,8095757688329
Triacylglycerol (18:1(11Z)/18:1(11Z)/18:1(11Z))	23,8096870435767
Triacylglycerol (18:1(11Z)/18:1(11Z)/18:1(9Z))	23,8096870435767
Triacylglycerol (18:1(11Z)/18:1(11Z)/18:3(5Z,9Z,12Z))	23,8096409827489
Triacylglycerol (18:1(11Z)/18:1(11Z)/18:4(5Z,9Z,12Z,15Z))	23,8094824167205
Triacylglycerol (18:1(11Z)/18:1(9Z)/16:0)	23,8094951741635
Triacylglycerol (18:1(11Z)/18:1(9Z)/18:0)	23,8095757688329
Triacylglycerol (18:1(11Z)/18:1(9Z)/18:1(11Z))	23,8096870435767
Triacylglycerol (18:1(11Z)/18:1(9Z)/18:1(9Z))	23,8096870435767
Triacylglycerol (18:1(11Z)/18:1(9Z)/18:3(5Z,9Z,12Z))	23,8096409827489
Triacylglycerol (18:1(11Z)/18:1(9Z)/18:4(5Z,9Z,12Z,15Z))	23,8094824167205
Triacylglycerol (18:1(9Z)/18:1(11Z)/16:0)	23,8094951741635
Triacylglycerol (18:1(9Z)/18:1(11Z)/18:0)	23,8095757688329
Triacylglycerol (18:1(9Z)/18:1(11Z)/18:1(11Z))	23,8096870435767
Triacylglycerol (18:1(9Z)/18:1(11Z)/18:1(9Z))	23,8096870435767
Triacylglycerol (18:1(9Z)/18:1(11Z)/18:3(5Z,9Z,12Z))	23,8096409827489
Triacylglycerol (18:1(9Z)/18:1(11Z)/18:4(5Z,9Z,12Z,15Z))	23,8094824167205
Triacylglycerol (18:1(9Z)/18:1(9Z)/16:0)	23,8094951741635
Triacylglycerol (18:1(9Z)/18:1(9Z)/18:0)	23,8095757688329
Triacylglycerol (18:1(9Z)/18:1(9Z)/18:1(11Z))	23,8096870435767
Triacylglycerol (18:1(9Z)/18:1(9Z)/18:1(9Z))	23,8096870435767
Triacylglycerol (18:1(9Z)/18:1(9Z)/18:3(5Z,9Z,12Z))	23,8096409827489
Triacylglycerol (18:1(9Z)/18:1(9Z)/18:4(5Z,9Z,12Z,15Z))	23,8094824167205

Appendix D
Comparison of dFBA with numerical schemes

Dynamic FBA has been the first method used to compare experimental results with fluxes predicted by FBA (Varma and Palsson, 1994). It is a numerical scheme, as it employs an iterative procedure in which FBA is run at each step to calculate flux values. However, it is not a numerical integration scheme, since at each iteration it does not solve the ordinary differential equations numerically, but analytically. In this appendix it is proved qualitatively that 1st order and 2nd order numerical schemes belonging to the family of Runge-Kutta methods produce very similar results to dFBA. In particular, it can be argued that dFBA is slightly more accurate than 1st order Euler method, but it appears to be outperformed by 2nd order Crank-Nicolson method.

The comparison has been carried out for the case of photobioreactor design as described in §4.4. iRC1080 model was tested in the same settings as indicated in §4.4. Three schemes were compared: dFBA, 1st Euler method and 2nd order Crank-Nicolson method. In all cases $d\tau$ = 0.1 h^{-1}. The concentration profiles have been reported in Figure D1, D2 and D3, which refer to the results obtained through dFBA, Euler method and Crank-Nicolson method respectively.

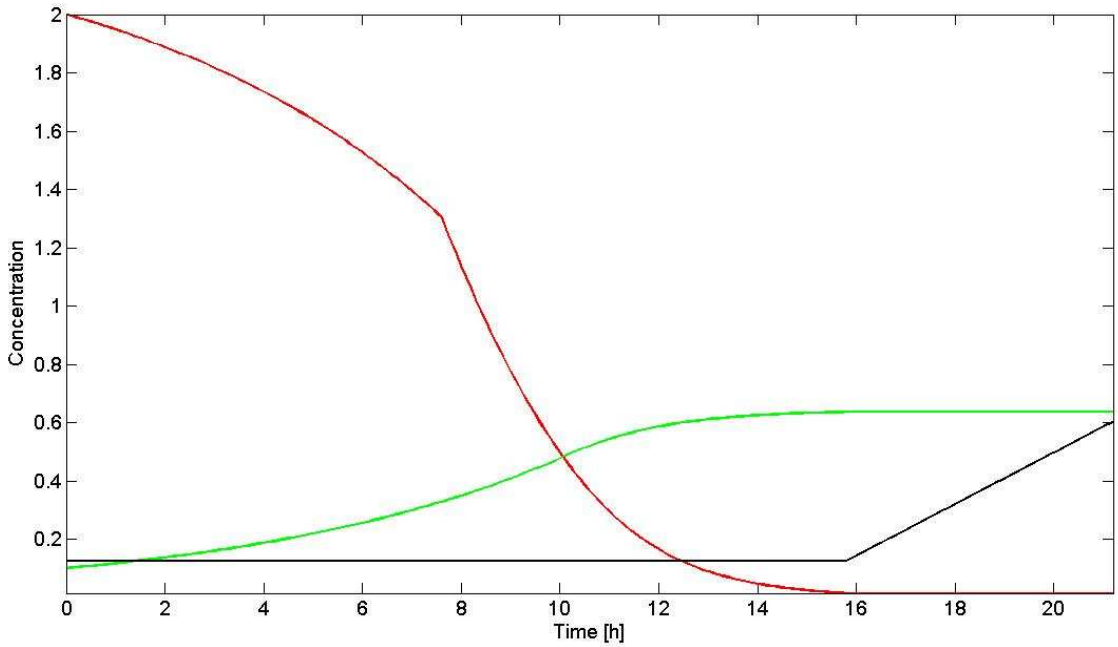

Figure D1. *Concentration profile by dFBA. Key as Fig. 4.4.*$C_{N,0}$ *= 2.0 mmol* L^{-1} *and* $d\tau$ *=0.1* h^{-1}.

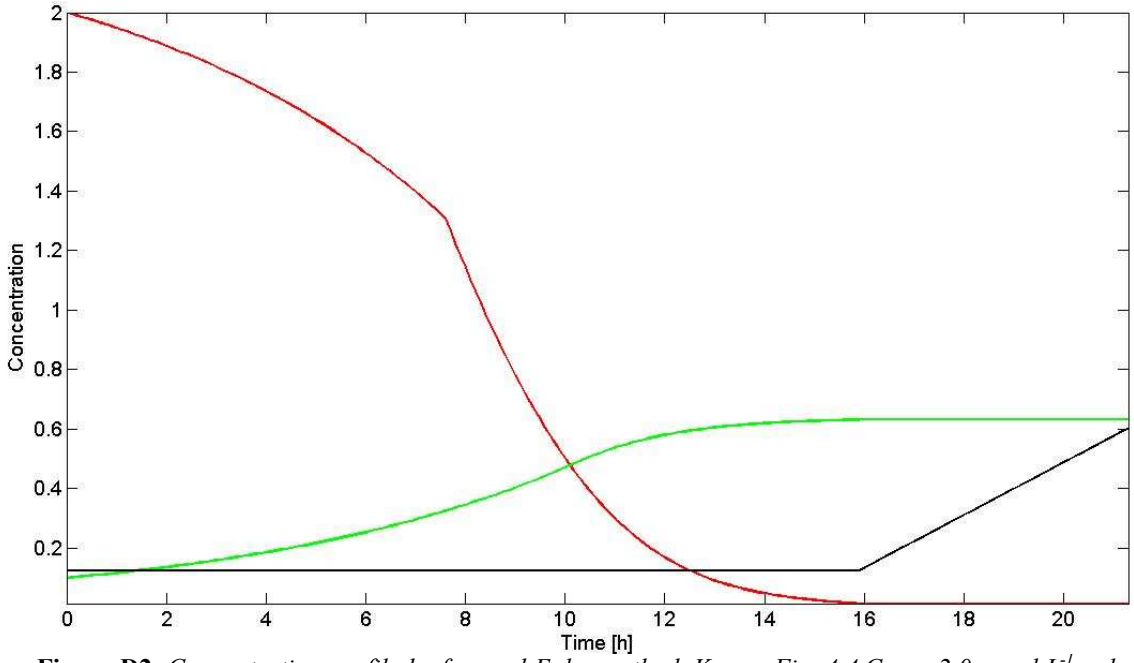

Figure D2. *Concentration profile by forawrd Euler method. Key as Fig. 4.4.*$C_{N,0}$ *= 2.0 mmol* L^{-1} *and* $d\tau = 0.1\ h^{-1}$.

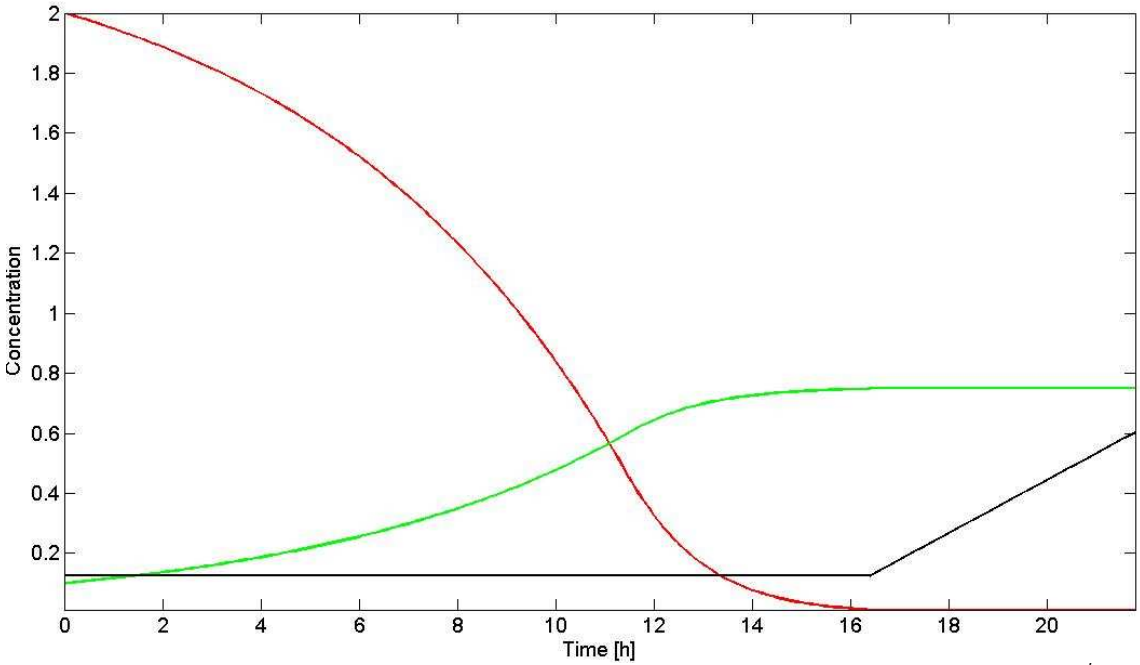

Figure D3. *Concentration profile by Crank-Nicolson method. Key as Fig. 4.4.*$C_{N,0}$ *= 2.0 mmol* L^{-1} *and* $d\tau = 0.1\ h^{-1}$.

Acknowledgements

My first acknowledgement goes to Dr. Daniel Segrè who accepted and helped me to join his lab at the Boston University. Not only, he has been essential for the genesis of this work, thanks to his suggestions and curiosity. I would like to thank all lab members for their help and sympathy, starting from Ed, Elisabetta, Varun, Christopher, Lina, Yasin, Bill, Niels, Sara, David, Brian, Hsuan-Chao and Qasim. I would like to thank also Prof. DeLisi, for taking contact with me when I was submitting my application for the above mentioned scholarship.

Then I would like to thank prof. Bertucco, who accepted for the second time to supervise my work though the inherent difficulty to manage (for the second time!) a research conducted so far away from his department. Together with prof. Bertucco, I have to acknowledge Dr.ssa Eleonora Sforza as well, whose precious suggestions and biochemical competences along with her continuous interest and passion for this research have helped me to walk on throughout my stay in Boston.

For their longstanding support and unconditioned trust, I must thank my parents, my family and my relatives, with my deepest and humblest regards.

Finally, it is immense and immeasurable my devotion to my endless source of perseverance, curiosity, creativity, passion and ultimately life, which are my friends. Thank you all.

Bibliography

1. Becker, S., Feist, A., Mo, M.L., Hannum, G., Palsson, B. Ø., Herrgard, M.J., "Quantitative prediction of cellular metabolism with constraint-based models: The COBRA Toolbox", Nature Protocols, **2**: *727-738* (2007).
2. Bonnel, P.H., Raby, C., Proceedings of 7th Congress of the International Society of Blood Transfusion, (1958).
3. Boyle, N. and Morgan, J.A., "Flux balance analysis of primary metabolism in *Chlamydomonas reinhardtii*", BMC Systems Biology, **3**: *4* (2009).
4. Bro, C., B. Regenberg, J. Forster and J. Nielsen, "In silico aided metabolic engineering of Saccharomyces cerevisiae for improved bioethanol production", Metabolic Engineering, **8**: *102–111* (2006).
5. Burgard, A.P., P. Pharkya and C.D. Maranas, "OptKnock: A Bilevel Programming Framework for Identifying Gene Knockout Strategies for Microbial Strain Optimization," Biotechnology and Bioengineering, **84**: *647-657* (2003).
6. Caspi R, Foerster H, Fulcher C. A, Kaipa P, Krummenacker M, Latendresse M, Paley S, Rhee S. Y, Shearer A. G, Tissier C, Walk T. C, Zhang P, Karp P. D., "The MetaCyc Database of metabolic pathways and enzymes and the BioCyc collection of Pathway/Genome Databases", Nucleic Acids Research, **36**: *D623-D631* (2007).
7. Chang, R. L., L. Ghamsari, A. Manichaikul, E. F. Hom, S. Balaji, W. Fu, Y. Shen, T. Hao, B. Palsson, K. Salehi-Ashtiani, and J. A. Papin, "Metabolic Network Reconstruction of Chlamydomonas Offers Insight into Light-Driven Algal Metabolism", Molecular Systems Biology, **7**: *518* (2011).
8. Chavali, A.K., J.D. Whittemore, J.A. Eddy, K.T. Williams, and J.A. Papin, "Systems analysis of metabolism in the pathogenic Trypanosomatid *Leishmania major*", Molecular Systems Biology, **4**: *177* (2008).
9. Chong, L., Ray, L.B., "Whole-istic Biology", Science, **295**: *1661* (2002).
10. Covert, M.W., Schilling C.H., Palsson B.Ø., "Regulation of gene expression in flux balance models of metabolism", Journal of Theoretical Biology, **213**(1): *73-88* (2001).
11. Dunahay, T. G., Jarvis, E. E., and Roessler, P. G., "Genetic transformation of the diatoms *Cyclotella cryptica* and *Navicula saprophila*", Journal of Phycology, **31**: *1004–1011* (1995).
12. Edwards, J.S., Ibarra, R.U., and Palsson, B.Ø., "In silico predictions of *Escherichi coli* metabolic capabilities are consistent with experimental data", Nature Biotechnology, ***19****: 125-130* (2001).

13. Feist, A.M., Palsson, B. Ø., "The biomass objective function", Current Opinion in Microbiology, **13**: *344-349* (2010).
14. Foster, J., Famili, I., Fu, P.C., Palsson, B., and Nielsen, J., "Genome-scale reconstruction of the *Saccharomyces cerevisiae* metabolic network", Genome Resources, **13**(2): *244-53* (2003).
15. Geoffroy L, Gilbin R, Simon O, Floriani M, Adam C, Pradines C, Cournac L, Garnier-Laplace J., "Effect of selenate on growth and photosynthesis of *Chlamydomonas reinhardtii*", Aquatic Toxicology, **83** (2):*149-58* (2007).
16. Gorman, D.S., Levine, R.P., "Cytochrome f and plastocyanin: their sequence in the photosynthetic electron transport chain of Chlamydomonas reinhardii", PNAS, **54**: *1665-1669* (1965).
17. Heifetz, P, Forster, B, Osmond, C., Giles, L.J., Boynton, J.E., "Effects of acetate on facultative autotrophy in *Chlamydomonas reinhardtii* assessed by photosynthetic measurements and stable isotope analysis", Plant Physiology, **122**: *1439-1445* (2000).
18. Hellio C., Veron B., Le Gal Y., "Amino acid utilization by *Chlamydomonas reinhardtii*: Specific study of histidine", Plant Physiology and Biochemistry, **42**(3): *257-264* (2004).
19. Hoischen C, Krämer R., "Membrane alteration is necessary but not sufficient for effective glutamate secretion in *Corynebacterium glutamicum*", Journal of Bacteriology, **172**(6): *3409–3416* (1990).
20. Hu, Q., Sommerfeld, M., Jarvis, E., Ghirardi, M., Posewitz, M., Seibert, M. and Darzins, A., "Microalgal triacylglycerols as feedstocks for biofuel production: perspectives and advances", The Plant Journal, **54**: *621–639* (2008)
21. James G.O., Hocart CH, Hillier W., Chen, H., Kordbacheh F., Price D and Djordjevic MA. "Fatty Acid Profiling of *Chlamydomonas reinhardtii* under Nitrogen Deprivation", Bioresearch Biotechnology, **102**: *3343-51* (2011).
22. Janssen, M.G.J., Kuijpers, T.C., Veldhoen, B., Brik Ternbach, M., Tramper, J., Mur, L.R., Wijffels, R.H., "Specific growth rate of Chlamydomonas reinhardtii and Chlorella sorokiniana under medium duration light/dark cycles: 13-87 s", Journal of Biotechnology, **70**: *323-333* (1999).
23. Li, Y., Han, D., Hu, G., Sommerfeld, M. and Hu, Q, "Inhibition of starch synthesis results in overproduction of lipids in *Chlamydomonas reinhardtii*", Biotechnology and Bioengineering, **107**: *258–268* (2010).
24. Kauffman K.J., Prakash, P., Edwards, J.S., "Advances in flux balance analysis", Current opinion in biotechnology, **14** (5): *491–496* (2003).
25. Kitano H., "Systems Biology: A Brief Overview", Science, **295**: *1662-1664* (2002).

26. Klitgord, N. and Segre', D., “Ecosystems biology of microbial metabolism”, Current Opinions in Biotechnology, **22**: *1-6* (2011).
27. Melis, A., Zhang, L., Forestier, M., Ghirardi, M. and Seibert, M., “Sustained photobiological hydrogen gas production upon reversible inactivation of oxygen evolution in the green alga *Chlamydomonas reinhardtii*”, Plant Physiology, **122**: *127-135* (2000).
28. Merchant, S.S., Prochnik, S.E., Vallon, O., Harris, E.H., Karpowicz, S.J., Witman, G.B., Terry, A., Salamov, A., Fritz-Laylin, L.K., Marechal-Drouard, L., Marshall, W.F., Qu, L.-H., Nelson, D.R., Sanderfoot, A.A., Spalding, M.H., Kapitonov, V.V., Ren, Q., Ferris, P., Lindquist, E., Shapiro, H., Grimwood, J., Schmutz, J., Lucas, S., Chlamydomonas community annotation team, JGI annotation team, Grigoriev, I.V., Rokhsar, D.S., Grossman, A.R., “The *Chlamydomonas* Genome Reveals the Evolution of Key Animal and Plant Functions”, Science, **318**: *245 – 250* (2007).
29. Oberhardt M.A., Palsson, B.Ø. and Papin, J.A., “Applications of genome-scale metabolic reconstructions”, Molecular Systems Biology, **5**: *320* (2009).
30. Orth, J.D., Thiele, I., Palsson, B.Ø., “What is flux balance analysis?”, Nature Biotechnology, **28**: *245-248* (2010).
31. Palsson, B.Ø., “Systems Biology: Properties of Reconstructed Networks”, Cambridge University Press, (2006).
32. Park J.M., Kim T.Y., Lee S.Y., “Constraints-based genome-scale metabolic simulation for systems metabolic engineering”, Biotechnology Advances, **27**: *979–988* (2009).
33. Pharkya, P., A.P. Burgard and C.D. Maranas, "OptStrain: A Computational Framework for Redesign of Microbial Production Systems", Genome Research, **14**(11): *2367-76* (2004).
34. Pharkya, P. and C.D. Maranas, "An Optimization Framework for Identifying Reaction Activation/Inhibition or Elimination Candidates for Overproduction in Microbial Systems", Metabolic Engineering, **8**: *1-13* (2006).
35. Radakovits, R., Eduafo, P., and Posewitz, M.C., **“**Genetic engineering of fatty acid chain length in *Phaeodactylum tricornutum*”, Metabolic Engineering, **13**:*89-95* (2010b).
36. Radakovits, R., Jinkerson, R. E., Darzins, A., Posewitz, M. C., “Genetic Engineering of Algae for Enhanced Biofuel Production”, Eukaryotic Cell, **9**: *486-501* (2010a).
37. Reed J.L., Vo T.D., Schilling C.H., and Palsson B.O., “An expanded genome-scale model of Escherichia coli K-12 (iJR904 GSM/GPR)”, Genome Biology, **4**(9): *R54.1-R54.12* (2003).

38. Sendín, J.O.H., Alonso, A. A., & Banga, J. R., "Multi-objective optimization of biological networks for prediction of intracellular fluxes", Advances in Soft Computing, **49**: *197-205* (2009).
39. Schellenberger, J., Park, J. O., Conrad, T. C., and Palsson, B. Ø., "BiGG: a Biochemical Genetic and Genomic knowledgebase of large scale metabolic reconstructions", BMC Bioinformatics, **11**: *213*, (2010).
40. Shlomi, T., Berkman, O., and Ruppin, E., "Regulatory on/off minimization of metabolic flux changes after genetic perturbations", PNAS, **102***: 7695-7700* (2005).
41. Schütz R, L Kuepfer L, Sauer U., "Systematic evaluation of objective functions for predicting intracellular fluxes in *E. coli.*", Molecular Systems Biology, **3**: *119* (2007).
42. Siaut M, Cuiné S, Cagnon C, Fessler B, Nguyen M, Carrier P, Beyly A, Beisson F, Triantaphylidès C, Li-Beisson Y, Peltier G., "Oil accumulation in the model green alga Chlamydomonas reinhardtii: characterization, variability between common laboratory strains and relationship with starch reserves", BMC Biotechnology, **11**: *7* (2011).
43. Sinnott, R.K, "Chemical Engineering Design", Vol. 6, Elsevier, (2005).
44. Stephanopoulos, G.N., Aristidou, A.A., Nielsen, J., "Metabolic Engineering: Principles and Methodologies", Academic Press, (1998).
45. Stern, D.B., Harris, E., "The Chlamydomonas Sourcebook: Organellar and metabolic processes", Academic Press, (2008).
46. Thiele, I. and Palsson, B.O., "A protocol for generating a high-quality genome-scale metabolic reconstruction", Nature Protocols, **5**: *93-121* (2010).
47. Thorn J.A., Jarvis S.M., "Adenosine transporters", General Pharmacology, 27: 613-*620* (1996).
48. Tepper, N., Shlomi, T., "Predicting metabolic engineering knockout strategies for chemical production: accounting for competing pathways", Bioinformatics, **26**: *536-543* (2010).
49. Varma A., Palsson B.O., "Stoichiometric flux balance models quantitatively predict growth and metabolic by-product secretion in wild-type *Escherichia coli* W3110", Applied and Environmental Microbiology, **60**: *3724-3731* (1994).
50. Wang Z, Ullrich N, Joo S, Waffenschmidt S, Goodenough U., "Algal lipid bodies: stress induction, purification, and biochemical characterization in wild-type and starchless *Chlamydomonas reinhardtii*", Eukaryotic Cell. **8**: *1856–1868* (2009).
51. Weeks, A.M. and Chang, M.C.Y., "Constructing de Novo Biosynthetic Pathways for Chemical Synthesis inside Living Cells", Biochemistry, **50**: *5404-5418* (2011).
52. Work V.H., Radakovits R., Jinkerson R.E., Meuser J.E., Elliott L.G., Vinyard D.J., Laurens L.M., Dismukes G.C., Posewitz M.C., "Increased lipid accumulation in the

Chlamydomonas reinhardtii sta7-10 starchless isoamylase mutant and increased carbohydrate synthesis in complemented strains", Eukaryot Cell, **9**(8): *1251-61* (2010).

53. Yamada, T., Letunic, I.,Okuda, S., Kanehisa, M. and Bork, P.,"iPath2.0: interactive pathway explorer", Nucleic Acids Research, **39 :** *W412-W415* (2011).
54. Zhao Q.,Kurata, B., "Comparison of the prediction abilities of FBA, MOMA, and ROOM for a *pykF* mutant of *E. coli"*, The 17th International Conference on Genome Informatics, *107-1-2*, Pacifico Yokohama, Japan (2006).
55. Zonneveld, C., "Light-limited microalgal growth: a comparison of modelling approaches", Ecological Modelling, **113:** *41–54* (1998).

Websites

- BiGG Database (Biochemical Genomic and Genetic knowledgebase): http://bigg.ucsd.edu/ - Last access: 05/09/2011 (with permission);
- BRENDA: The Comprehensive Enzyme Information System: www.brenda-enzymes.org/ - Last access: 05/09/2011;
- Chlamydomonas reinhardtii v4.0: a genome browser: http://genome.jgi-psf.org/Chlre4/Chlre4.home.html - Last access: 05/09/2011;
- iPath2: Interactive Pathway Explorer: http://pathways.embl.de/ - Last access: 05/09/2011;
- KEGG (Kyoto Encyclopedia of Genes and Genomes): www.genome.jp/kegg - Last access: 05/09/2011;
- MetaCyc: Encyclopedia of Metabolic Pathways: metacyc.org/ - Last access: 05/09/2011;