Chapter 4

Applications of genome-scale metabolic models of microalgae and cyanobacteria in biotechnology

Steinn Gudmundsson[1*], Lucía Agudo[2] and Juan Nogales[2]

[1] *Center for Systems Biology, University of Iceland, Sturlugata 8, 101 Reykjavik, Iceland. E-mail: steinng@hi.is*

[2] *Department of Environmental Biology, Centro de Investigaciones Biológicas-CSIC, Ramiro de Maeztu 9, 28040 Madrid, Spain. E-mail: l.agudo@cib.csic.es, jnogales@cib.csic.es*

[*] Corresponding author.

Abstract

Industrial production of biofuels and platform chemicals in microalgae and cyanobacteria holds great promise since only sunlight, water and atmospheric carbon dioxide are required. At present it is not economically feasible to use photosynthetic organisms for the production of large-volume, low cost compounds. If this is to change, product yields must improve significantly. This will most likely require extensive genetic engineering of phototrophs. Genome-scale models of microbial metabolism are increasingly being used in biological discovery and metabolic engineering. This chapter provides an overview of genome-scale metabolic models of cyanobacteria and microalgae and describes their applications in biotechnology, strain design and analysis of omics data.

Keywords: Systems biology; genome-scale metabolic models; cyanobacteria; microalgae; metabolic engineering; analysis of omics data

4.1 Introduction

Microalgae and cyanobacteria are unicellular microorganisms that are widely distributed in both terrestrial and aquatic ecosystems. They possess complex machineries for carrying out oxygenic photosynthesis, using light energy to convert water and atmospheric CO_2 into biomass. The considerable species variety and metabolic versatility has led to a great deal of interest in these microorganisms as biocatalysts. Ease of cultivation, rapid growth and their natural ability to accumulate lipids in large amounts has led to intense research on the use of microalgae for biofuel production. Microalgae are also natural producers of many valuable carotenoids such as astaxanthin, fucoxanthin and beta-carotene and have been used to produce such compounds on a commercial scale (Yen et al., 2013, Wijffels et al., 2013). Cyanobacteria have successfully been engineered to produce an extensive array of commodity compounds, including biofuels, platform chemicals and secondary plant metabolites such as limonene (Angermayr et al., 2015) but at present it is not economically feasible to use them in industrial production systems. In comparison to most of the heterotrophic organisms used in biotechnology, the biology of photosynthetic microorganisms is not well understood. A direct consequence is that the field of metabolic engineering in phototrophs is considerably less advanced. Titers and production rates have frequently been low and there have been recurrent problems with genetic instability of engineered strains. Nevertheless progress has been significant in recent years (Angermayr et al., 2015, Gimpel et al., 2013, Radakovits et al., 2010). This progress can be attributed to improved genetic tools and also to the fact that recent strain designs take properties of the photosynthetic metabolism into account rather than emulating heterotrophic designs directly. An example of the latter is an isobutanol producing *Synecochococcus* strain where a knockout in glycogen synthesis resulted in approx. 50% of the fixed carbon being redirected to isobutanol synthesis (Li et al., 2014). A study of the "chemical space“ covered by metabolically engineered cyanobacteria shows that only a small part of industrially important compounds has been covered so far (Gudmundsson and Nogales, 2015).

Systems biology is an emerging discipline which studies biological processes through detailed computer models. Systems biology has the potential to rapidly advance our knowledge about cyanobacteria and microalgae, which can then lead to increased applications in biotechnology. Among the many tools within systems biology, genome scale metabolic reconstructions (GENREs) play a vital role in connecting genotype to phenotype.

Metabolic GENREs are structured representations of the metabolic capabilities of a target organism that are based on existing biochemical, genetic and genotypic knowledge. In its most basic form, a GENRE is a list of stoichiometrically balanced metabolic reactions together with

descriptions of gene-protein relationships. A GENRE can be transformed into a mathematical model (genome-scale model, GEM) which represents the metabolic capabilities of a particular organism and can be used to predict phenotype from genotype, contextualize omics data etc. (O'Brien et al., 2015).

Systems metabolic engineering combines systems biology, GENREs in particular, with synthetic biology and rational and random mutagenesis for optimizing biotechnological processes (Lee et al., 2012). The adoption of systems metabolic engineering approaches has resulted in numerous success stories in recent years involving heterotrophic organisms (Lee and Kim, 2015).

Although the development of GENREs for photosynthetic organisms has lagged behind reconstructions of heterotrophic "work horses“ such as *S. cerevisiae* and *E. coli* (Monk et al., 2014, Schmidt et al., 2010), numerous genome-scale reconstructions of microalgae and cyanobacteria have been built in the last years. An overview of the construction and analysis of GENREs is given in the next section. This is followed by a discussion of their applications to biotechnology with special emphasis on metabolic engineering and the contextualization of omics data.

4.2 Construction and analysis of genome--scale metabolic models

The construction of GENREs can be divided into four main steps: i) construction of an initial draft model; ii) manual curation of the draft model; iii) conversion to a mathematical format; and iv) network evaluation and analysis (Figure 4.1). Detailed protocols describing how to build high-quality reconstructions have been published (Feist et al., 2009, Thiele and Palsson, 2010, Nogales, 2014).

4.2.1 Construction of an initial draft model

In order to carry out a metabolic reconstruction it is important to have a high quality genome annotation of the organism of interest. If this is not the case, gene properties will be ambiguously defined, eventually leading to erroneous model predictions. After obtaining an annotated sequence, a draft model can be created automatically using one of several existing tools, e.g. Model SEED (Henry et al., 2010) and RAVEN (Agren et al., 2013). Existing models which have been extensively evaluated are used during the automatic model building process in an attempt to improve the accuracy of the draft reconstruction. The draft reconstruction provides an approximation of the metabolic capabilities of the organism based on genetic content but specific properties of the organism, such as the subcellular location of reactions, cofactors and substrate specificity cannot be established with certainty because these properties are generally species specific.

4.2.2 Manual curation of the draft model

In order for a reconstruction to reflect the true metabolic capabilities of the target organism, a manual review of all the reactions included in the initial draft is required. This includes checking the mass and charge balance of individual reactions, gene associations of reactions and reaction directionality. Databases, such as KEGG (Kanehisa and Goto, 2000), BRENDA (Schomburg et al., 2002), MetaCyc (Caspi et al., 2006), the Transport Classification Database (Saier et al., 2016) and BiGG models (King et al., 2016) as well as organism specific database of cyanobacteria and algae such as Cyanobase (Nakao et al., 2010), Cyanomutant (Nakamura et al., 1999) and ChlamyCyc (May et al., 2009) are consulted during this step. In addition legacy biochemical data and the primary literature is used extensively (Nogales, 2014). Model accuracy depends on the careful assignment of reactions to intracellular compartments. Photosynthetic microorganisms have a large number of compartments compared to heterotrophic bacteria and there is limited experimental data available, therefore the assignment of protein localization within a network is difficult. Cellular localization is therefore often based on homologous proteins found in model organisms. In addition, software such as PredAlgo (Tardif et al., 2012) which analyzes transit peptide sequences can be used to assign proteins to compartments. Accurate assignment of reaction directionality is another essential model building step. In the absence of strain-specific biochemical data, the directionality can be inferred from standard Gibbs energies of reactions. When experimental values of standard Gibbs energy are not available, they can often be estimated computationally (Noor et al., 2013). Computational methods for assigning reaction directionality using thermodynamic information combined with network topology and heuristic rules have been developed as well (Kummel et al., 2006). In addition, since electrical potentials and pH vary between different cell compartments in photosynthetic organisms, the directionality of a particular reaction may depend on its cellular location. The manual curation of GENREs is extremely time-consuming, with a time span ranging from months to years, depending on the target organism. However, this step is critical in order to obtain an accurate metabolic reconstruction (Nogales, 2014).

4.2.3 Conversion to a mathematical format

The list of stoichiometrically balanced reactions is collected in a *stoichiometric matrix*, denoted by $\boldsymbol{S}$, where the rows correspond to metabolites and the columns to reactions. This matrix contains the stoichiometric coefficients of each metabolite in each reaction. An assumption of steady-state conditions where the concentration of internal metabolites stays constant, bypasses the need for specifying kinetic constants for each reaction. This results in a system of equations, $\boldsymbol{Sv} = \boldsymbol{0}$ where the vector $\boldsymbol{v}$ denotes reaction *fluxes*, turnover rates of individual metabolites which are the quantities of interest. The fluxes are subject to lower bound constraints, $\boldsymbol{v}_{min}$ and maximum capacity

constraints $\boldsymbol{v}_{max}$ such that $\boldsymbol{v}_{min} \leq \boldsymbol{v} \leq \boldsymbol{v}_{max}$. Irreversible reactions have lower bounds equal to zero. Gene knockouts are simulated by setting the flux in affected reactions to zero and heterologous gene insertion is simulated by adding corresponding enzymatic reactions to the model. The mathematical method most widely used for studying biochemical networks is flux balance analysis (FBA) (Palsson, 2015). The method identifies flux values which maximize a given cellular objective while simultaneously satisfying the above constraints. Frequently used objectives are the maximization of biomass or the production of a target compound, e.g. eicosapentaenoic acid. The former objective requires the construction of a species-specific biomass reaction which represents the conversion of metabolic compounds into biomass constituents, i.e. nucleic acids, proteins, lipids etc. It is therefore important to obtain experimental data on the macromolecular composition of the target organism. In addition, uptake and secretion rates of key metabolites from/to the media are needed for accurate model simulations.

4.2.4 Network evaluation and analysis

After obtaining a mathematical representation of a network, the next steps are to check for errors made during the reconstruction, the identification of gaps in metabolic pathways and the model validation (by comparing to experimental results). Computational algorithms, often built upon FBA, have been developed to aid with these tasks. Algorithms for identifying flux bounds, elucidation of objective functions, strain design, predictions of gene essentiality and many more have also been developed (Palsson, 2015, Maranas and Zomorrodi, 2016). Software packages implementing the algorithms are freely available for the Matlab (Schellenberger et al., 2011, Agren et al., 2013), Python (Ebrahim et al., 2013) and R (Gelius-Dietrich et al., 2013) programming languages.

Table 4.1a. Metabolic reconstructions of cyanobacteria. The number of genes, reactions (Rxns), metabolites (Mets) and compartments (Cmps) for each network are shown, together with an application example from each study.

TABLE 4.1a here

Table 4.1b. Metabolic reconstructions of microalgae. The number of genes, reactions (Rxns), metabolites (Mets) and compartments (Cmps) for each network are shown, together with an application example from each study.

TABLE 4.1b here

4.3 An overview of existing microalgae and cyanobacterial GENREs and their applications in biotechnology

A number of metabolic reconstructions of algae and cyanobacteria have been published in the last decade but they represent a limited number of different species. This is likely due to scarcity of experimental data. Approximately 116 cyanobacterial species have been sequenced to date but only ~30 microalgal genomes are available. Metabolic reconstructions for 5 cyanobacterial and 9 microalgae species have been published (Tables 4.1a and 4.1b). These numbers are in stark contrast to the approximately 2000 cyanobacteria (Nabout et al., 2013) and 35000 microalgae species (Borowitzka, 2013) which have been described to date.

The first metabolic reconstructions were based solely on existing biochemical knowledge, focusing mainly on central carbon metabolism and photosynthetic pathways but with the arrival of sequenced genomes, it became possible to build metabolic models at the genome-scale (Baroukh et al., 2015). A metabolic reconstruction of the cyanobacterium *Arthrospira platensis* was used to study exopolysaccharide production under photosynthetic conditions. The production was found to be associated with NADH production in the central metabolic pathways (Cogne et al., 2003). The microalgae *Chlorella pyrenoidosa* (Yang et al., 2000) was reconstructed in a similar fashion with the common goal of analyzing carbon and energy metabolism in varying light conditions.

Network reconstruction is an iterative process and models are gradually expanded and updated when new data become available. This is exemplified by the increasing scope of *Chlamydomonas reinhardtii* reconstructions (Manichaikul et al., 2009, Chang et al., 2011, Imam et al., 2015) and reconstructions of the cyanobacteria *Synechocystis* sp. PCC6803 (Shastri and Morgan, 2005, Nogales et al., 2012, Knoop et al., 2013). One of the main difficulties in modeling the metabolism of photoautotrophic microorganisms is the representation of photosynthesis. This involves both the light-dependent reactions representing photon capture, water splitting and the generation of ATP and NADPH; and the light-independent ("dark") reactions of CO_2 fixation and photorespiration. Several simplifying assumptions are usually made in modeling the light-dependent reactions. Photons are treated as substrates taking part in chemical reactions and the number of photons is assumed to be constant. Electron transfer via membrane proteins is modeled by including the proteins as separate species in oxidized and reduced forms. By including multiple compartments in the reconstruction, e.g. the thylakoid and cytoplasm it is possible to simulate proton gradient across membranes. The dark reactions utilize the ATP and NADPH generated by the light reactions in the Calvin cycle and the photorespiration pathway which recycles toxic metabolites generated by the reaction of oxygen with RuBisCO. The main challenges in modeling the dark reactions are due to

limited understanding of the photorespiratory processes and difficulties in modeling carbon concentration mechanisms.

The *Chlamydomonas reinhardtii* metabolic reconstruction by Chang et al. includes detailed modeling of light-driven metabolism, accounting for light intensity and the spectral properties of several different light sources. The latter was was achieved by including "prism reactions" which decompose light sources into their spectral components which then drive the photon-utilizing metabolic reactions (Chang et al., 2011). The light source modeling enables optimization of growth conditions, light source design and the study of light-driven metabolism. Shortly afterwards, Nogales and colleagues modeled photosynthetic electron flow pathways in detail, including many circular electron flow and accessory pathways, enabling the study of photosynthetic processes at the system level (Nogales et al., 2012). Alternate electron flow pathways, including cyclic electron flow, a hydrogenase and the Mehler reaction were found to assist in modulating the ATP/NADPH ratio. A high degree cooperativity of these pathways was found to be essential for optimal photoautotrophic metabolism under under changing environmental conditions. The photorespiratory process and diurnal cycle in *Synechocystis* were addressed in (Knoop et al., 2013). The modeling assumption that metabolites do not accumulate inside cells is frequently made (see previous section) but this is a problematic assumption for cells subject to environmental fluctuations such as light-dark cycles. The metabolic states change significantly during these cycles, with accumulation of metabolites such as glycogen and lipids during the light periods and their consumption during the dark periods. To address this issue Baroukh and colleagues developed a framework called DRUM. The metabolic network is decomposed into sub-networks and simple kinetics are used to describe each component. Differential equations were then used to describe substrate consumption, biomass production, product secretion and the accumulation of internal metabolites. DRUM was used to study the accumulation of lipids and carbohydrates in the microalgae *Tisochrysis lutea* under day/night cycles (Baroukh et al., 2014). The development of microalgae and cyanobacteria models over time, their main properties and highlights of the respective studies are summarized in Tables 4.1a and 4.1b.

In addition to basic research into the photosynthetic metabolism, current GEMs find many uses in biotechnology, e.g. studying the production of biofuels and chemical building blocks. In the following we highlight some of the biotechnological applications of genome-scale models of microalgae and cyanobacteria. It is worth noting that almost all the compounds that have been studied to date *in silico* are low-value, high-volume compounds rather than high-value, low-volume compounds. It has been argued that the focus should be on the latter group while engineering methodologies are being developed and production costs are still high (Ducat et al., 2011).

4.3.1 Production of alcohols

The importance of the development of renewable energy is reflected in many of the publications on metabolic modeling of these organisms. By inserting ethanol and 1-propanol pathways into a *Synechocystis* GENRE and analyzing the network using FBA, Yoshikawa et al. found that nitrogen or phosphate limitation increased the production of both alcohols (Yoshikawa et al., 2011). Kamarainen et al. used FBA to study the maximum production potential of various biofuels in *Synechocystis*. They found that the energy yields per mol photon were more or less the same for ethanol, propane, butanol, heptadecane and octadecanol (Kamarainen et al., 2012). Vu and coworkers estimated theoretical yields for the production of a number of biofuels and platform chemicals in *Synechococcus* 7002 using a suit of computational methods (Vu et al., 2013) and identified knockout mutants that were predicted to increase production of the target compounds. Lower yields were predicted under dark conditions than under photoautotrophic conditions. In an attempt to systematically identify suitable strain designs for ethanol and isobutanol production in cyanobacteria, Erdrich and colleagues enumerated all potential knockout strategies in a *Synechocystis* GENRE (Erdrich et al., 2014). The methodology was based on the decomposition of the metabolic network into minimal functional units called elementary modes, an alternative to flux balance analysis. After identifying all the elementary modes supporting both photosynthetic growth and the production of the target compound, the smallest sets of reaction knockouts enabling the desired elementary modes were identified. Analysis of the knockout strategies revealed that yields could be increased by altering the ATP/NADPH ratio, e.g. by introducing ATP wasting mechanisms. A follow-up study by Knoop and Steuer (Knoop and Steuer, 2015) employed flux variability analysis to identify which reactions in the network had to be up- or down regulated in order to achieve the desired ATP/NADPH ratio. Hendry et al. reconstructed the metabolic network of *Synecococcus* sp. PCC7002 (Hendry et al., 2016). Using a method called Minimization of metabolic adjustments they predicted that 10% of the fixed carbon could be directed to the formation of ethanol and butanol. This was achieved by altering the ATP/NADPH ratio by knocking out glutamate dehydrogenase. Mueller and coworkers (Mueller et al., 2013) recently reconstructed five Cyanothece strains using the model from (Saha et al., 2012) as a template . The metabolic capabilities of the five strains with respect to production of higher alcohols were assessed using BLAST searches.

4.3.2 Production of bulk chemicals

In an *in silico* study on isoprene production in *Synechocystis*, Saha and coworkers inserted the non-native *ispS* gene which catalyzes the production of isoprene into a *Synechocystis* GENRE

(Saha et al., 2012). They used FBA to predict maximum isoprene yields by simulating conditions from an *in vivo* study on isoprene production (Bentley and Melis, 2012). The model underestimated the experimentally observed isoprene production by an order of magnitude. The discrepancy was attributed to a combination of factors, which included limitations of the biomass function and the possibility of reduced *in vivo* growth on the *ispS* transformant. A *Synechocystis* reconstruction was used to study carbon partitioning and *de novo* formation of α-ketoglutarate in two recombinant *Synechocystis* ethylene producing strains (Zavrel et al., 2016). Using experimental data to constrain the model, carbon partitioning for the two strains was estimated to be between 10% and 17%. Pathways for overproducing succinate from glycogen in *Synechocystis* under dark conditions were identified by screening a large number of non-native enzymatic reactions found in KEGG. Isocitrate lyase in particular was found to give a significant increease in succinate production (Shirai et al., 2016).

4.3.3 Production of fatty acids

The ability of some microalgae species to accumulate large quantities of lipids under nutrient limiting conditions has lead to major research efforts into maximizing lipid production for use in biofuels. To this end, a thorough study of algae metabolism at the systems level is needed. Functional annotation of the lipid metabolism uncovered during a *Chlamydomonas reinhardtii* reconstruction, revealed hypothetical latent genetic pathways which represent potential genetic engineering targets for lipid production (Chang et al., 2011). Muthuraj and colleagues studied neutral lipid accumulation in *Chlorella* sp. FC2 IITG during a switch from nutrient rich to nutrient limited conditions using a dynamic version of FBA to simulate light-dark cycles. The model predicted induction of lipid accumulation after approximately 72 h of photoautotrophic growth and there was good agreement between predicted and experimental results. The model was used to infer changes that took place in intracellular fluxes prior and during the accumulation of lipids (Muthuraj et al., 2013). *Chlorella protothecoides* is another promising candidate for biofuel production. Wu and coworkers constructed a metabolic model of this organism and used it to study the effects of light absorption and CO_2 uptake on the growth rate and oleic acid production (Wu et al., 2015). The optimal ratio of photon and CO_2 uptake was predicted to be around 8. Similar model simulations were carried out for a reconstruction of *Chlorella variabilis* (Juneja et al., 2016) where in addition to photon and CO2 uptake, the effects of ammonia and phosphorus uptake were also studied. A recent reconstruction of the diatom *Phaeodactylum tricornutum* which includes highly detailed modeling of lipid metabolism (Levering et al., 2016) can provide new insights into lipid production in this diatom.

4.3.4 Hydrogen production

GENREs have been used to gain insights into hydrogen production in both cyanobacteria and microalgae. The maximum theoretical yields of hydrogen in *Synechocystis* was estimated with FBA to be 0.17 mmol·gDW^{-1}·h^{-1} which indicates that there is a large room for improvement in optimizing hydrogen producing strains (Montagud et al., 2010). In a later study, Saha and coauthors found that two recent *Synechocystis* reconstructions gave theoretical estimates that were approximately 50% lower than experimentally observed values (Saha et al., 2012). They attributed the difference between *in silico* and *in vivo* results to a combination of modeling and experimental issues. Mass and charge imbalances in the model by Montagud et al. may play a role in the discrepancy between the two computational studies (Yoshikawa et al., 2011). The use of GEMs for enhancing hydrogen production in *Synechocystis* has recently been reviewed (Montagud et al., 2015). The *C. reinhardtii* reconstruction AlgaGEM predicted increased hydrogen production under mixotrophic conditions upon disruption of cyclic electron flow and that overexpression of the PFR1 gene would increase hydrogen yields under dark conditions (Dal'Molin et al., 2011).

4.4 Metabolic engineering

To obtain cyanobacteria and microalgae strains of industrial importance, extensive genetic engineering will be needed. The ability of GEMs to predict phenotypic properties under genetic and environmental perturbations makes them highly useful in metabolic engineering since the global impact of gene insertions and gene knockouts on metabolism can be predicted in advance. As a result, it is possible to search for optimal strain-designs *in silico*.

There are several examples in the literature of GEMs being used for optimizing strain designs (Machado and Herrgard, 2015). To the best of our knowledge, all the computationally designed strains that have been experimentally verified to date have involved heterotrophs but the same computational methods can just as easily be applied to metabolic models of phototrophs. Algorithms for strain design can roughly be classified as belonging to one of three types, *de novo* pathway assembly, enzyme modulation and enzyme deletion. An example of the first type is OptStrain (Pharkya et al., 2004) which can be used to design pathways for non-native compounds as well as to increase yields of native compounds. OptStrain could also be used to search for pathways for converting toxic byproducts into non-toxic metabolites. The algorithm works by searching a database of enzymatic reactions from multiple organisms such as KEGG, for inclusion in a given metabolic model in order to reach a pre-specified production target. The OptForce algorithm (Ranganathan et al., 2010) searches for sets of enzymes which must be up-regulated, down-

regulated or knocked out in order to achieve a minimum level of target product. OptKnock (Burgard et al., 2003) is a prototypical example of the third group of algorithms which search for enzyme knockouts to couple the production of a target compound to growth. Growth-coupling is a desirable feature in engineered strains since it mitigates the problems of selection and low genetic stability and adaptive laboratory evolution can then be used to increase production of the target compound.

In silico knockout studies using metabolic models of the cyanobacteria *Synechocystis* PCC6803 (Nogales et al., 2013, Erdrich et al., 2014) and *Synechococcus* PCC7002 (Vu et al., 2013) found that a large number of knockouts is required to couple product secretion to growth, typically 8 - 13 enzymes whereas only 3 - 5 knockouts were needed in *E. coli* (Feist et al., 2010). The phototroph knockout strains identified in (Nogales et al., 2013, Erdrich et al., 2014) revealed that blocking the alternative electron flow pathways of photosynthesis is frequently required to obtain growth coupling. These results suggest that fundamentally different strategies are needed for obtaining growth coupling in phototrophs. In light of these results, it should be noted that the accuracy of knockout predictions is expected to decrease with increasing number of knockouts due to cumulative effects of model limitations and modeling errors. Using a GENRE of *Arthrospira platensis* NIES-39, Yoshikawa and coworkers studied the effects of nitrate depletion on glycogen production and identified *zwf* as a possible knockout target for increasing glycogen production. An *in-silico* gene knockout simulation identified NADPH dehydrogenase and cytochrome-c oxidase as possible targets for improving ethanol yields (Yoshikawa et al., 2015).

Model-driven strain design for algae and cyanobacteria is still in its infancy but the successful applications of this technology to heterotrophic organisms, together with advances in genetic tools for phototrophs, strongly suggest that systems metabolic engineering will play an important role for cyanobacteria and microalgae in the near future.

4.5 Contextualization of OMICS data

Modern biology experiments frequently involve data collection on a large scale in the form of “omics” data, i.e. genomic, transcriptomic, lipidomic, proteomic and metabolomic data sets. The arrival of omics technologies enables the monitoring of molecular components at the cellular level and at the genome-scale. Analyzing large and often disparate data sets has been a challenge, requiring the development of new computational approaches (De Smet and Marchal, 2010). Genome-scale models provide a much needed biological context for analyzing omics data sets and can be used to obtain a mechanistic interpretation of the data. Several methods for the integration of omics data with GEMs have been proposed in recent years (Hyduke et al., 2013, Machado and Herrgard, 2014). Most of the methods can be classified into two main categories, i) a “switch“

approach which simply remove reactions from the model whose corresponding genes have expression values below a user-specified threshold; and ii) a “valve” approach, which constrains the reaction fluxes based on relative gene expression levels. Other approaches include a method based purely on the network topology of GEMs (Patil and Nielsen, 2005).

Model integration methods have been used for the construction of tissue-specific metabolic models of higher plants, including arabidopsis (Mintz-Oron et al., 2012) and maize (Simons et al., 2014). Omics data from microalgae and cyanobacteria have also been analyzed with the aid of GEMs. For instance, the inclusion of high-resolution time series transcriptomic data in a *C. reinhardtii* GEM identified the dynamic changes in central pathways in response to nitrogen starvation and light availability (Imam et al., 2015). The model was used to study how carbon flux was funneled towards biosynthesis of triacylglycerols instead of biomass during nitrogen starvation. In a study on the phototrophic metabolism of *Chlorella protothecoides*, ^{13}C-label profiling was used to constrain fluxes in a GEM of this oleaginous green alga. The results revealed negligible photorespiratory fluxes and low activity of the tricarboxylic acid cycle under phototrophic conditions (Wu et al., 2015). A *Synechocystis* network in conjunction with transcriptomics data was used to identify metabolic hot-spots serving as regulatory hubs during changes in light availability (Montagud et al., 2010). Nogales and Agudo used transcriptomics data derived from autotrophic conditions to constrain a GEM of *Synechocystis* (Nogales and Agudo, 2015) and identified those photosynthetic pathways which are potentially active under optimal light conditions only. Transcriptomic and physiological data were used in combination with a genome-scale model to study the circadian rhythm in *Synechocystis* (Saha et al., 2016). Relative flux changes in the major cellular processes, such as CO_2 fixation, the pentose phosphate pathway and the TCA cycle were found to correlate well with relative changes in expression levels of the associated genes. A tool for data integration and visualization has recently been developed for the *Synechocystis* network (Maarleveld et al., 2014). By providing a clear graphical map of metabolic and transcriptomic data, this tool is a valuable addition for future systems biology and metabolic engineering studies.

.6 Outlook

High-quality, comprehensive, genome-scale models of photosynthetic microorganisms are a powerful tool for studying poorly characterized biological processes. By expanding our biological knowledge and providing means to rapidly carry out experiments on a large-scale, *in silico*, the models are likely play a vital role in future biotechnology research efforts. Future progress will be expedited by increasing the scope and accuracy of existing models (King et al., 2015). In addition to modeling photosynthesis and metabolism in more detail (Steuer et al., 2012), the next generation of

photosynthetic models needs to go beyond metabolism. By combining metabolic models with models of gene expression, so-called ME-models enable accurate predictions of growth, macromolecular composition, metabolic fluxes, enzyme efficiencies under nutrient limitation and more (Thiele et al., 2012, Lerman et al., 2012). By taking costs of enzyme synthesis into account and the complex interactions of metabolism and expression, ME-models could lead to improved strain designs compared to metabolic-only models. Another promising direction of model expansion is the integration of protein structure with metabolic models (Chang et al., 2013). Such models enable predictions of the effects of enzyme structural properties on cell metabolism and could be used to gain insights into the complex photosynthetic macrostructures such as the photosystems and light-harvesting complexes.

The development of next generation GEMs will not only improve our understanding of phototrophs but will also enable optimized model-driven designs for biotechnological applications, bridging the gap between photoautotrophic and heterotrophic biocatalysts and significantly contribute to biosustainability efforts.

References

AGREN, R., LIU, L., SHOAIE, S., VONGSANGNAK, W., NOOKAEW, I. & NIELSEN, J. 2013. The RAVEN toolbox and its use for generating a genome-scale metabolic model for Penicillium chrysogenum. *PLoS Comput Biol,* 9**,** e1002980.

ANGERMAYR, S. A., GORCHS ROVIRA, A. & HELLINGWERF, K. J. 2015. Metabolic engineering of cyanobacteria for the synthesis of commodity products. *Trends Biotechnol,* 33**,** 352-61.

BAROUKH, C., MUNOZ-TAMAYO, R., STEYER, J. P. & BERNARD, O. 2014. DRUM: a new framework for metabolic modeling under non-balanced growth. Application to the carbon metabolism of unicellular microalgae. *PLoS One,* 9**,** e104499.

BAROUKH, C., MUNOZ-TAMAYO, R., STEYER, J. P. & BERNARD, O. 2015. A state of the art of metabolic networks of unicellular microalgae and cyanobacteria for biofuel production. *Metab Eng,* 30**,** 49-60.

BENTLEY, F. K. & MELIS, A. 2012. Diffusion-based process for carbon dioxide uptake and isoprene emission in gaseous/aqueous two-phase photobioreactors by photosynthetic microorganisms. *Biotechnol Bioeng,* 109**,** 100-9.

BOROWITZKA, M. A. 2013. Species and Strain Selection. *In:* BOROWITZKA, A. M. & MOHEIMANI, R. N. (eds.) *Algae for Biofuels and Energy.* Dordrecht: Springer Netherlands.

BOYLE, N. R. & MORGAN, J. A. 2009. Flux balance analysis of primary metabolism in Chlamydomonas reinhardtii. *BMC Syst Biol,* 3**,** 4.

BURGARD, A. P., PHARKYA, P. & MARANAS, C. D. 2003. Optknock: a bilevel programming framework for identifying gene knockout strategies for microbial strain optimization. *Biotechnol Bioeng,* 84**,** 647-57.

CASPI, R., FOERSTER, H., FULCHER, C. A., HOPKINSON, R., INGRAHAM, J., KAIPA, P., KRUMMENACKER, M., PALEY, S., PICK, J., RHEE, S. Y., TISSIER, C., ZHANG, P. & KARP, P. D. 2006. MetaCyc: a multiorganism database of metabolic pathways and enzymes. *Nucl. Acids Res.,* 34**,** D511-516.

CHANG, R. L., GHAMSARI, L., MANICHAIKUL, A., HOM, E. F., BALAJI, S., FU, W., SHEN, Y., HAO, T., PALSSON, B. O., SALEHI-ASHTIANI, K. & PAPIN, J. A. 2011. Metabolic network reconstruction of Chlamydomonas offers insight into light-driven algal metabolism. *Mol Syst Biol,* 7**,** 518.

CHANG, R. L., XIE, L., BOURNE, P. E. & PALSSON, B. O. 2013. Antibacterial mechanisms identified through structural systems pharmacology. *BMC Syst Biol,* 7**,** 102.

COGNE, G., LEHMANN, B., DUSSAP, C. G. & GROS, J. B. 2003. Uptake of macrominerals and trace elements by the cyanobacterium Spirulina platensis (Arthrospira platensis PCC 8005) under photoautotrophic conditions: culture medium optimization. *Biotechnology and Bioengineering,* 81**,** 588-93.

COGNE, G., RUGEN, M., BOCKMAYR, A., TITICA, M., DUSSAP, C. G., CORNET, J. F. & LEGRAND, J. 2011. A model-based method for investigating bioenergetic processes in autotrophically growing eukaryotic microalgae: application to the green algae Chlamydomonas reinhardtii. *Biotechnol Prog,* 27**,** 631-40.

DAL'MOLIN, C. G., QUEK, L. E., PALFREYMAN, R. W. & NIELSEN, L. K. 2011. AlgaGEM--a genome-scale metabolic reconstruction of algae based on the Chlamydomonas reinhardtii genome. *BMC Genomics,* 12 Suppl 4**,** S5.

DE SMET, R. & MARCHAL, K. 2010. Advantages and limitations of current network inference methods. *Nat Rev Microbiol,* 8**,** 717-29.

DUCAT, D. C., WAY, J. C. & SILVER, P. A. 2011. Engineering cyanobacteria to generate high-value products. *Trends Biotechnol,* 29**,** 95-103.

EBRAHIM, A., LERMAN, J. A., PALSSON, B. O. & HYDUKE, D. R. 2013. COBRApy: COnstraints-Based Reconstruction and Analysis for Python. *BMC Syst Biol,* 7**,** 74.

ERDRICH, P., KNOOP, H., STEUER, R. & KLAMT, S. 2014. Cyanobacterial biofuels: new insights and strain design strategies revealed by computational modeling. *Microb Cell Fact,* 13**,** 128.

FEIST, A. M., HERRGARD, M. J., THIELE, I., REED, J. L. & PALSSON, B. O. 2009. Reconstruction of biochemical networks in microorganisms. *Nat Rev Microbiol,* 7**,** 129-43.

FEIST, A. M., ZIELINSKI, D. C., ORTH, J. D., SCHELLENBERGER, J., HERRGARD, M. J. & PALSSON, B. O. 2010. Model-driven evaluation of the production potential for growth-coupled products of Escherichia coli. *Metab Eng,* 12**,** 173-86.

FU, P. 2009. Genome-scale modeling of Synechocystis sp. PCC 6803 and prediction of pathway insertion. *Journal of Chemical Technology & Biotechnology,* 84**,** 473-483.

GELIUS-DIETRICH, G., DESOUKI, A. A., FRITZEMEIER, C. J. & LERCHER, M. J. 2013. Sybil--efficient constraint-based modelling in R. *BMC Syst Biol,* 7**,** 125.

GIMPEL, J. A., SPECHT, E. A., GEORGIANNA, D. R. & MAYFIELD, S. P. 2013. Advances in microalgae engineering and synthetic biology applications for biofuel production. *Curr Opin Chem Biol,* 17**,** 489-95.

GUDMUNDSSON, S. & NOGALES, J. 2015. Cyanobacteria as photosynthetic biocatalysts: a systems biology perspective. *Mol Biosyst,* 11**,** 60-70.

HAMILTON, J. J. & REED, J. L. 2012. Identification of functional differences in metabolic networks using comparative genomics and constraint-based models. *PLoS One,* 7**,** e34670.

HENDRY, J. I., PRASANNAN, C. B., JOSHI, A., DASGUPTA, S. & WANGIKAR, P. P. 2016. Metabolic model of Synechococcus sp. PCC 7002: Prediction of flux distribution and network modification for enhanced biofuel production. *Bioresour Technol*.

HENRY, C. S., DEJONGH, M., BEST, A. A., FRYBARGER, P. M., LINSAY, B. & STEVENS, R. L. 2010. High-throughput generation, optimization and analysis of genome-scale metabolic models. *Nat Biotechnol,* 28**,** 977-82.

HONG, S.-J. & LEE, C.-G. 2007. Evaluation of central metabolism based on a genomic database ofSynechocystis PCC6803. *Biotechnology and Bioprocess Engineering,* 12**,** 165-173.
HUNT, K. A., FOLSOM, J. P., TAFFS, R. L. & CARLSON, R. P. 2014. Complete enumeration of elementary flux modes through scalable demand-based subnetwork definition. *Bioinformatics,* 30**,** 1569-78.
HYDUKE, D. R., LEWIS, N. E. & PALSSON, B. O. 2013. Analysis of omics data with genome-scale models of metabolism. *Mol Biosyst,* 9**,** 167-74.
IMAM, S., SCHAUBLE, S., VALENZUELA, J., LOPEZ GARCIA DE LOMANA, A., CARTER, W., PRICE, N. D. & BALIGA, N. S. 2015. A refined genome-scale reconstruction of Chlamydomonas metabolism provides a platform for systems-level analyses. *Plant J,* 84**,** 1239-56.
JUNEJA, A., CHAPLEN, F. W. & MURTHY, G. S. 2016. Genome scale metabolic reconstruction of Chlorella variabilis for exploring its metabolic potential for biofuels. *Bioresour Technol*.
KAMARAINEN, J., KNOOP, H., STANFORD, N. J., GUERRERO, F., AKHTAR, M. K., ARO, E. M., STEUER, R. & JONES, P. R. 2012. Physiological tolerance and stoichiometric potential of cyanobacteria for hydrocarbon fuel production. *Journal of Biotechnology,* 162**,** 67-74.
KANEHISA, M. & GOTO, S. 2000. KEGG: kyoto encyclopedia of genes and genomes. *Nucleic Acids Res,* 28**,** 27-30.
KIM, J., FABRIS, M., BAART, G., KIM, M. K., GOOSSENS, A., VYVERMAN, W., FALKOWSKI, P. G. & LUN, D. S. 2016. Flux balance analysis of primary metabolism in the diatom Phaeodactylum tricornutum. *Plant J,* 85**,** 161-76.
KING, Z. A., LLOYD, C. J., FEIST, A. M. & PALSSON, B. O. 2015. Next-generation genome-scale models for metabolic engineering. *Curr Opin Biotechnol,* 35**,** 23-9.
KING, Z. A., LU, J., DRAGER, A., MILLER, P., FEDEROWICZ, S., LERMAN, J. A., EBRAHIM, A., PALSSON, B. O. & LEWIS, N. E. 2016. BiGG Models: A platform for integrating, standardizing and sharing genome-scale models. *Nucleic Acids Res,* 44**,** D515-22.
KLANCHUI, A., KHANNAPHO, C., PHODEE, A., CHEEVADHANARAK, S. & MEECHAI, A. 2012. iAK692: a genome-scale metabolic model of Spirulina platensis C1. *BMC Syst Biol,* 6**,** 71.
KLIPHUIS, A. M., KLOK, A. J., MARTENS, D. E., LAMERS, P. P., JANSSEN, M. & WIJFFELS, R. H. 2012. Metabolic modeling of Chlamydomonas reinhardtii: energy requirements for photoautotrophic growth and maintenance. *J Appl Phycol,* 24**,** 253-266.
KNOOP, H., GRUNDEL, M., ZILLIGES, Y., LEHMANN, R., HOFFMANN, S., LOCKAU, W. & STEUER, R. 2013. Flux balance analysis of cyanobacterial metabolism: the metabolic network of Synechocystis sp. PCC 6803. *PLoS Comput Biol,* 9**,** e1003081.
KNOOP, H. & STEUER, R. 2015. A computational analysis of stoichiometric constraints and trade-offs in cyanobacterial biofuel production. *Front Bioeng Biotechnol,* 3**,** 47.
KNOOP, H., ZILLIGES, Y., LOCKAU, W. & STEUER, R. 2010. The metabolic network of Synechocystis sp. PCC 6803: systemic properties of autotrophic growth. *Plant Physiol,* 154**,** 410-22.
KRUMHOLZ, E. W., YANG, H., WEISENHORN, P., HENRY, C. S. & LIBOUREL, I. G. 2012. Genome-wide metabolic network reconstruction of the picoalga Ostreococcus. *J Exp Bot,* 63**,** 2353-62.
KUMMEL, A., PANKE, S. & HEINEMANN, M. 2006. Systematic assignment of thermodynamic constraints in metabolic network models. *BMC Bioinformatics,* 7**,** 512.
LEE, J. W., NA, D., PARK, J. M., LEE, J., CHOI, S. & LEE, S. Y. 2012. Systems metabolic engineering of microorganisms for natural and non-natural chemicals. *Nat Chem Biol,* 8**,** 536-46.

LEE, S. Y. & KIM, H. U. 2015. Systems strategies for developing industrial microbial strains. *Nat Biotechnol,* 33**,** 1061-72.
LERMAN, J. A., HYDUKE, D. R., LATIF, H., PORTNOY, V. A., LEWIS, N. E., ORTH, J. D., SCHRIMPE-RUTLEDGE, A. C., SMITH, R. D., ADKINS, J. N., ZENGLER, K. & PALSSON, B. O. 2012. In silico method for modelling metabolism and gene product expression at genome scale. *Nat Commun,* 3**,** 929.
LEVERING, J., BRODDRICK, J., DUPONT, C. L., PEERS, G., BEERI, K., MAYERS, J., GALLINA, A. A., ALLEN, A. E., PALSSON, B. O. & ZENGLER, K. 2016. Genome-Scale Model Reveals Metabolic Basis of Biomass Partitioning in a Model Diatom. *PLoS One,* 11**,** e0155038.
LI, X., SHEN, C. R. & LIAO, J. C. 2014. Isobutanol production as an alternative metabolic sink to rescue the growth deficiency of the glycogen mutant of Synechococcus elongatus PCC 7942. *Photosynth Res,* 120**,** 301-10.
MAARLEVELD, T. R., BOELE, J., BRUGGEMAN, F. J. & TEUSINK, B. 2014. A data integration and visualization resource for the metabolic network of Synechocystis sp. PCC 6803. *Plant Physiol,* 164**,** 1111-21.
MACHADO, D. & HERRGARD, M. 2014. Systematic evaluation of methods for integration of transcriptomic data into constraint-based models of metabolism. *PLoS Comput Biol,* 10**,** e1003580.
MACHADO, D. & HERRGARD, M. J. 2015. Co-evolution of strain design methods based on flux balance and elementary mode analysis. *Metabolic Engineering Communications,* 2**,** 85 - 92.
MANICHAIKUL, A., GHAMSARI, L., HOM, E. F., LIN, C., MURRAY, R. R., CHANG, R. L., BALAJI, S., HAO, T., SHEN, Y., CHAVALI, A. K., THIELE, I., YANG, X., FAN, C., MELLO, E., HILL, D. E., VIDAL, M., SALEHI-ASHTIANI, K. & PAPIN, J. A. 2009. Metabolic network analysis integrated with transcript verification for sequenced genomes. *Nat Methods,* 6**,** 589-92.
MARANAS, C. D. & ZOMORRODI, A. R. 2016. *Optimization methods in metabolic networks*.
MAY, P., CHRISTIAN, J. O., KEMPA, S. & WALTHER, D. 2009. ChlamyCyc: an integrative systems biology database and web-portal for Chlamydomonas reinhardtii. *BMC Genomics,* 10**,** 209.
MINTZ-ORON, S., MEIR, S., MALITSKY, S., RUPPIN, E., AHARONI, A. & SHLOMI, T. 2012. Reconstruction of Arabidopsis metabolic network models accounting for subcellular compartmentalization and tissue-specificity. *Proc Natl Acad Sci U S A,* 109**,** 339-44.
MONK, J., NOGALES, J. & PALSSON, B. O. 2014. Optimizing genome-scale network reconstructions. *Nat Biotechnol,* 32**,** 447-52.
MONTAGUD, A., GAMERMANN, D., FERNANDEZ DE CORDOBA, P. & URCHUEGUIA, J. F. 2015. Synechocystis sp. PCC6803 metabolic models for the enhanced production of hydrogen. *Crit Rev Biotechnol,* 35**,** 184-98.
MONTAGUD, A., NAVARRO, E., FERNANDEZ DE CORDOBA, P., URCHUEGUIA, J. F. & PATIL, K. R. 2010. Reconstruction and analysis of genome-scale metabolic model of a photosynthetic bacterium. *BMC Syst Biol,* 4**,** 156.
MONTAGUD, A., ZELEZNIAK, A., NAVARRO, E., DE CORDOBA, P. F., URCHUEGUIA, J. F. & PATIL, K. R. 2011. Flux coupling and transcriptional regulation within the metabolic network of the photosynthetic bacterium Synechocystis sp. PCC6803. *Biotechnol J,* 6**,** 330-42.
MUELLER, T. J., BERLA, B. M., PAKRASI, H. B. & MARANAS, C. D. 2013. Rapid construction of metabolic models for a family of Cyanobacteria using a multiple source annotation workflow. *BMC Syst Biol,* 7**,** 142.

MUTHURAJ, M., PALABHANVI, B., MISRA, S., KUMAR, V., SIVALINGAVASU, K. & DAS, D. 2013. Flux balance analysis of Chlorella sp. FC2 IITG under photoautotrophic and heterotrophic growth conditions. *Photosynth Res,* 118**,** 167-79.

NABOUT, J. C., DA SILVA ROCHA, B., CARNEIRO, F. M. & SANT'ANNA, C. L. 2013. How many species of Cyanobacteria are there? Using a discovery curve to predict the species number. *Biodiversity and Conservation,* 22**,** 2907-2918.

NAKAMURA, Y., KANEKO, T., MIYAJIMA, N. & TABATA, S. 1999. Extension of CyanoBase. CyanoMutants: repository of mutant information on Synechocystis sp. strain PCC6803. *Nucleic Acids Res,* 27**,** 66-8.

NAKAO, M., OKAMOTO, S., KOHARA, M., FUJISHIRO, T., FUJISAWA, T., SATO, S., TABATA, S., KANEKO, T. & NAKAMURA, Y. 2010. CyanoBase: the cyanobacteria genome database update 2010. *Nucleic Acids Res,* 38**,** D379-81.

NOGALES, J. 2014. A Practical Protocol for Genome-Scale Metabolic Reconstructions. *Springer Protocols Handbooks.* Totowa, NJ: Humana Press.

NOGALES, J. & AGUDO, L. 2015. A Practical Protocol for Integration of Transcriptomics Data into Genome-Scale Metabolic Reconstructions. *Springer Protocols Handbooks.* Totowa, NJ: Humana Press.

NOGALES, J., GUDMUNDSSON, S., KNIGHT, E. M., PALSSON, B. O. & THIELE, I. 2012. Detailing the optimality of photosynthesis in cyanobacteria through systems biology analysis. *Proc Natl Acad Sci U S A,* 109**,** 6.

NOGALES, J., GUDMUNDSSON, S. & THIELE, I. 2013. Toward systems metabolic engineering in cyanobacteria: opportunities and bottlenecks. *Bioengineered,* 4**,** 158-63.

NOOR, E., HARALDSDOTTIR, H. S., MILO, R. & FLEMING, R. M. 2013. Consistent estimation of Gibbs energy using component contributions. *PLoS Comput Biol,* 9**,** e1003098.

O'BRIEN, E. J., MONK, J. M. & PALSSON, B. O. 2015. Using Genome-scale Models to Predict Biological Capabilities. *Cell,* 161**,** 971-87.

PALSSON, B. 2015. *Systems biology : constraint-based reconstruction and analysis*.

PATIL, K. R. & NIELSEN, J. 2005. Uncovering transcriptional regulation of metabolism by using metabolic network topology. *Proc Natl Acad Sci U S A,* 102**,** 2685-9.

PHARKYA, P., BURGARD, A. P. & MARANAS, C. D. 2004. OptStrain: a computational framework for redesign of microbial production systems. *Genome Res,* 14**,** 2367-76.

RADAKOVITS, R., JINKERSON, R. E., DARZINS, A. & POSEWITZ, M. C. 2010. Genetic engineering of algae for enhanced biofuel production. *Eukaryot Cell,* 9**,** 486-501.

RANGANATHAN, S., SUTHERS, P. F. & MARANAS, C. D. 2010. OptForce: an optimization procedure for identifying all genetic manipulations leading to targeted overproductions. *PLoS Comput Biol,* 6**,** e1000744.

SAHA, R., LIU, D., HOYNES-O'CONNOR, A., LIBERTON, M., YU, J., BHATTACHARYYA-PAKRASI, M., BALASSY, A., ZHANG, F., MOON, T. S., MARANAS, C. D. & PAKRASI, H. B. 2016. Diurnal Regulation of Cellular Processes in the Cyanobacterium Synechocystis sp. Strain PCC 6803: Insights from Transcriptomic, Fluxomic, and Physiological Analyses. *MBio,* 7.

SAHA, R., VERSEPUT, A. T., BERLA, B. M., MUELLER, T. J., PAKRASI, H. B. & MARANAS, C. D. 2012. Reconstruction and comparison of the metabolic potential of cyanobacteria Cyanothece sp. ATCC 51142 and Synechocystis sp. PCC 6803. *PLoS One,* 7**,** e48285.

SAIER, M. H., JR., REDDY, V. S., TSU, B. V., AHMED, M. S., LI, C. & MORENO-HAGELSIEB, G. 2016. The Transporter Classification Database (TCDB): recent advances. *Nucleic Acids Res,* 44**,** D372-9.

SCHELLENBERGER, J., QUE, R., FLEMING, R. M., THIELE, I., ORTH, J. D., FEIST, A. M., ZIELINSKI, D. C., BORDBAR, A., LEWIS, N. E., RAHMANIAN, S., KANG, J., HYDUKE, D. R. & PALSSON, B. O. 2011. Quantitative prediction of cellular metabolism with constraint-based models: the COBRA Toolbox v2.0. *Nat Protoc,* 6**,** 1290-307.

SCHMIDT, B. J., LIN-SCHMIDT, X., CHAMBERLIN, A., SALEHI-ASHTIANI, K. & PAPIN, J. A. 2010. Metabolic systems analysis to advance algal biotechnology. *Biotechnol J,* 5**,** 660-70.
SCHOMBURG, I., CHANG, A. & SCHOMBURG, D. 2002. BRENDA, enzyme data and metabolic information. *Nucleic Acids Res,* 30**,** 47-9.
SHASTRI, A. A. & MORGAN, J. A. 2005. Flux balance analysis of photoautotrophic metabolism. *Biotechnol Prog,* 21**,** 1617-26.
SHIRAI, T., OSANAI, T. & KONDO, A. 2016. Designing intracellular metabolism for production of target compounds by introducing a heterologous metabolic reaction based on a Synechosystis sp. 6803 genome-scale model. *Microb Cell Fact,* 15**,** 13.
SIMONS, M., SAHA, R., AMIOUR, N., KUMAR, A., GUILLARD, L., CLEMENT, G., MIQUEL, M., LI, Z., MOUILLE, G., LEA, P. J., HIREL, B. & MARANAS, C. D. 2014. Assessing the metabolic impact of nitrogen availability using a compartmentalized maize leaf genome-scale model. *Plant Physiol,* 166**,** 1659-74.
STEUER, R., KNOOP, H. & MACHNE, R. 2012. Modelling cyanobacteria: from metabolism to integrative models of phototrophic growth. *J Exp Bot,* 63**,** 2259-74.
TARDIF, M., ATTEIA, A., SPECHT, M., COGNE, G., ROLLAND, N., BRUGIERE, S., HIPPLER, M., FERRO, M., BRULEY, C., PELTIER, G., VALLON, O. & COURNAC, L. 2012. PredAlgo: a new subcellular localization prediction tool dedicated to green algae. *Mol Biol Evol,* 29**,** 3625-39.
THIELE, I., FLEMING, R. M., QUE, R., BORDBAR, A., DIEP, D. & PALSSON, B. O. 2012. Multiscale modeling of metabolism and macromolecular synthesis in E. coli and its application to the evolution of codon usage. *PLoS One,* 7**,** e45635.
THIELE, I. & PALSSON, B. O. 2010. A protocol for generating a high-quality genome-scale metabolic reconstruction. *Nat Protoc,* 5**,** 93-121.
TRIANA, J., MONTAGUD, A., SIURANA, M., FUENTE, D., URCHUEGUIA, A., GAMERMANN, D., TORRES, J., TENA, J., DE CORDOBA, P. F. & URCHUEGUIA, J. F. 2014. Generation and Evaluation of a Genome-Scale Metabolic Network Model of Synechococcus elongatus PCC7942. *Metabolites,* 4**,** 680-98.
VU, T. T., HILL, E. A., KUCEK, L. A., KONOPKA, A. E., BELIAEV, A. S. & REED, J. L. 2013. Computational evaluation of Synechococcus sp. PCC 7002 metabolism for chemical production. *Biotechnol J,* 8**,** 619-30.
VU, T. T., STOLYAR, S. M., PINCHUK, G. E., HILL, E. A., KUCEK, L. A., BROWN, R. N., LIPTON, M. S., OSTERMAN, A., FREDRICKSON, J. K., KONOPKA, A. E., BELIAEV, A. S. & REED, J. L. 2012. Genome-scale modeling of light-driven reductant partitioning and carbon fluxes in diazotrophic unicellular cyanobacterium Cyanothece sp. ATCC 51142. *PLoS Comput Biol,* 8**,** e1002460.
WIJFFELS, R. H., KRUSE, O. & HELLINGWERF, K. J. 2013. Potential of industrial biotechnology with cyanobacteria and eukaryotic microalgae. *Curr Opin Biotechnol,* 24**,** 405-13.
WU, C., XIONG, W., DAI, J. & WU, Q. 2015. Genome-based metabolic mapping and 13C flux analysis reveal systematic properties of an oleaginous microalga Chlorella protothecoides. *Plant Physiol,* 167**,** 586-99.
YANG, C., HUA, Q. & SHIMIZU, K. 2000. Energetics and carbon metabolism during growth of microalgal cells under photoautotrophic, mixotrophic and cyclic light-autotrophic/dark-heterotrophic conditions. *Biochem Eng J,* 6**,** 87-102.
YEN, H. W., HU, I. C., CHEN, C. Y., HO, S. H., LEE, D. J. & CHANG, J. S. 2013. Microalgae-based biorefinery--from biofuels to natural products. *Bioresour Technol,* 135**,** 166-74.
YOSHIKAWA, K., AIKAWA, S., KOJIMA, Y., TOYA, Y., FURUSAWA, C., KONDO, A. & SHIMIZU, H. 2015. Construction of a Genome-Scale Metabolic Model of Arthrospira

platensis NIES-39 and Metabolic Design for Cyanobacterial Bioproduction. *PLoS One,* 10**,** e0144430.
YOSHIKAWA, K., KOJIMA, Y., NAKAJIMA, T., FURUSAWA, C., HIRASAWA, T. & SHIMIZU, H. 2011. Reconstruction and verification of a genome-scale metabolic model for Synechocystis sp. PCC6803. *Appl Microbiol Biotechnol,* 92**,** 347-58.
ZAVREL, T., KNOOP, H., STEUER, R., JONES, P. R., CERVENY, J. & TRTILEK, M. 2016. A quantitative evaluation of ethylene production in the recombinant cyanobacterium Synechocystis sp. PCC 6803 harboring the ethylene-forming enzyme by membrane inlet mass spectrometry. *Bioresour Technol,* 202**,** 142-51.

Species	Reference	Genes	Rxns	Mets	Cmps	Application example
Arthrospira platensis PCC8005	Cogne et al., 2003	n/a	121	134	1	Exopolysaccharide production
Arthrospira platensis C1	Klanchui et al., 2012	692	875	837	3	Identification of essential genes
Arthrospira platensis NIES-39	Yoshikawa et al., 2015	n/a	746	673	3	Glycogen and ethanol production
Synechocystis sp. PCC6803	Shastri and Morgan, 2005	n/a	70	46	1	Study of photoautotrophic metabolism
	Hong and Lee, 2007	86	56	48	2	Study of photoautotrophic metabolism
	Fu, 2009	633	831	704	1	Ethanol production
	Montagud et al., 2010	669	882	790	2	Metabolic engineering and omic data integration
	Knoop et al., 2010	337	380	291	1	Functional properties of phototrophic growth
	Montagud et al., 2011	811	956	911	2	Flux coupling in photoautotrophic metabolism
	Yoshikawa et al., 2011	393	493	465	2	Ethanol and 1-propanol production
	Nogales et al., 2012	678	863	795	3	Analysis of photosynthetic processes
	Saha et al., 2012	731	1156	996	4	Comparisons of Cyanothece and Synechocystis
	Knoop et al., 2013	677	759	601	4	Modelling of diurnal cycle
Cyanothece sp. ATCC 51142	Saha et al., 2012	773	946	811	4	Comparisons of Cyanothece and Synechocystis
	Vu et al., 2012	806	719	587	1	System analysis of light-driven metabolism
Synecococcus elongatus PCC7942	Triana et al., 2014	715	851	838	2	Analysis of flux partitioning
Synecococcus sp. PCC7002	Hamilton and Reed, 2012	611	552	542	2	Analysis of metabolic differences in cyanobacteria
	Vu et al., 2013	708	602	581	2	Biofuel production
	Hendry et al., 2016	728	742	696	7	Ethanol and butanol production

Table 4.1a. Metabolic reconstructions of cyanobacteria. The number of genes, reactions (Rxns), metabolites (Mets) and compartments (Cmps) for each network are shown, together with an application example from each study.

Species	Reference	Genes	Rxns	Mets	Cmps	Example application
Chlorella pyrenoidosa	Yang et al., 2000	n/a	67	61	1	Analysis of carbon and energy metabolism
Chlorella protothecoides	Wu et al., 2015	461	272	144	4	Fatty acid production
Chlorella sp. FC2	Muthuraj et al., 2013	n/a	154	114	1	Lipid accumulation during nutrient starvation
Chlorella variabilis NC64A	Juneja et al., 2016	526	1455	1236	6	Analysis of different light conditions
Chlamydomonas reinhardtii	Manichaikul et al., 2009	n/a	259	467	6	Model refinement via transcript verification exps.
	Boyle and Morgan, 2009	n/a	484	458	3	Estimation of intracellular fluxes
	Cogne et al., 2011	n/a	280	278	1	Analysis of light-driven respiration
	Chang et al., 2011	1080	2190	1068	9	Growth predictions for different light light sources
	Dal'Molin et al., 2011	2279	1725	1869	4	Hydrogen production
	Kliphuis et al., 2012	n/a	160	164	2	Analysis of biomass yield on light
	Imam et al., 2015	1355	2394	1845	10	Lipid accumulation during nitrogen starvation
Ostreococcus tauri	Krumholz et al., 2012	n/a	871	1014	1	Network comparisons
Ostreococcus lucimarinus	Krumholz et al., 2012	n/a	964	1100	1	Network comparisons
Tisochysis lutea	Baroukh et al., 2014	n/a	160	157	2	Dynamic modeling of intracellular processes
Phaeodactylum tricornutum	Hunt et al., 2014	680	318	335	4	Network decomposition
	Kim et al., 2016	607	849	587	5	Exploration of lower glycolysis in the mitochondria
	Levering et al., 2016	1027	4456	2172	6	Detailed modeling of lipid metabolism

Table 4.1b. Metabolic reconstructions of microalgae. The number of genes, reactions (Rxns), metabolites (Mets) and compartments (Cmps) for each network are shown, together with an application example from each study.

Figure 4.1: Reconstructing a genome scale metabolic model.