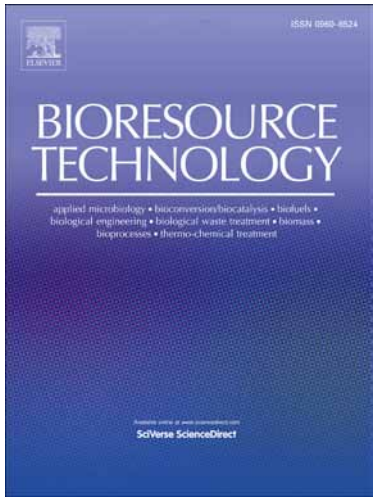


Metabolic model of *Synechococcus* sp. PCC 7002: prediction of flux distribution and network modification for enhanced biofuel production

John I. Hendry, Charulata B. Prasannan, Aditi Joshi, Santanu Dasgupta, Pramod P. Wangikar

PII: S0960-8524(16)30274-7
DOI: http://dx.doi.org/10.1016/j.biortech.2016.02.128
Reference: BITE 16178

To appear in: *Bioresource Technology*

Received Date: 2 January 2016
Revised Date: 25 February 2016
Accepted Date: 26 February 2016

Please cite this article as: Hendry, J.I., Prasannan, C.B., Joshi, A., Dasgupta, S., Wangikar, P.P., Metabolic model of *Synechococcus* sp. PCC 7002: prediction of flux distribution and network modification for enhanced biofuel production, *Bioresource Technology* (2016), doi: http://dx.doi.org/10.1016/j.biortech.2016.02.128

Metabolic model of *Synechococcus* sp. PCC 7002: prediction of flux distribution and network modification for enhanced biofuel production

John I Hendry[a], Charulata B Prasannan[a], Aditi Joshi[a], Santanu Dasgupta[d], Pramod P Wangikar[a,b,c,*]

[a]*Department of Chemical Engineering, Indian Institute of Technology Bombay, Mumbai-400076, India*
[b]*DBT-Pan IIT Center for Bioenergy, Indian Institute of Technology Bombay, Mumbai-400076, India*
[c]*Wadhwani Research Center for Bioengineering, Indian Institute of Technology Bombay, Mumbai-400076, India*
[d]*Reliance Technology Group, Reliance Industries Limited, Reliance Corporate Park, Ghansoli, Navi Mumbai-400701, India*

Abstract

Flux Balance Analysis was performed with the Genome Scale Metabolic Model of a fast growing cyanobacterium *Synechococcus* sp. PCC 7002 to gain insights that would help in engineering the organism as a production host. Gene essentiality and synthetic lethality analysis revealed a reduced metabolic robustness under genetic perturbation compared to the heterotrophic bacteria *E.coli*. Under glycerol heterotrophy the reducing equivalents were generated from tricarboxylic acid cycle rather than the oxidative pentose phosphate pathway. During mixotrophic growth in glycerol the photosynthetic electron transport chain was predominantly used for ATP synthesis with a photosystem I/photosystem II flux ratio higher than that observed under autotrophy. An exhaustive analysis of all possible double reaction knock outs was performed to reroute fixed carbon towards ethanol and butanol production. It was predicted that only ∼10% of fixed carbon could be diverted for ethanol and butanol production.

Keywords: Cyanobacteria, *Synechococcus* sp. PCC 7002, Genome scale metabolic network reconstruction, Flux balance analysis, Gene essentiality analysis, Synthetic lethality analysis, Minimization of Metabolic Adjustments (MOMA)

1. Introduction

Cyanobacteria are a group of gram negative bacteria capable of fixing CO_2 and performing oxygenic photosynthesis. Concerns over depletion of fossil fuel reserves and rising levels of atmospheric CO_2 levels have drawn attention towards cyanobacteria, a potential source of renewable energy (Parmar et al., 2011). Cyanobacteria grow significantly faster compared to other oxygenic

Estimation of intracellular fluxes or reaction rates is the first step towards rational design of pathways for enhanced production of chemicals of commercial interest (Feng et al., 2010). Flux Balance Analysis (FBA) is a widely used method for the estimation of intracellular fluxes (Orth et al., 2010). FBA is based on the stoichiometry of reactions in the metabolic network and does not require the kinetic parameters to predict fluxes. Thus, FBA is a useful alternative for detailed kinetic models for organisms that have little or no kinetic data available for their metabolic enzymes. Genome Scale Metabolic Model (GSMM), a computational representation of organism specific metabolism, is a prerequisite for FBA. GSMM construction and simulation involves the following essential steps: (i) Listing of the metabolic reactions from genome annotations, biochemical texts, organism specific pathway databases etc, (ii) organization of these reactions into a metabolic network, which can be broken down into pathways and subsystems, (iii) identification and filling of gaps in the network, (iv) simulation of the model with a set of constraints and objective function, and (v) model validation.

A number of methods have been developed to analyze the metabolic phenotype using GSMM. These are collectively known as Constraint Based Reconstruction Analysis (COBRA) methods (Lewis et al., 2012). Some of these help predict gene knockout and gene over-expression strategies for enhanced production of target biochemicals. Minimization of Metabolic Adjustments (MOMA) is one of these methods that can predict the flux distribution of a gene knockout mutant by minimizing the Euclidean distance between flux values of the mutant and wild type strains (Segrè et al., 2002). This method has been successfully applied to increase the rate of production of L-valine in *E.coli* (Park et al., 2007).

In this study, FBA was performed for *Synechococcus* sp. PCC 7002 under photoautotrophic, heterotrophic and mixotrophic conditions. Robustness of its metabolism was investigated using Gene Essentiality and Synthetic lethality analysis. For this, an updated GSMM for *Synechococcus* sp. PCC 7002 was built using the already available reconstruction (Vu et al., 2013) as a starting point. The model was validated by analyzing its prediction of growth rate and other important parameters such as photosynthetic quotient and Photosystem I (PSI)/Photosystem II (PSII) flux ratio using FBA. Further, validation was also carried out by comparing the predictions of gene essentiality analysis with the experimental data available in literature. Minimization of Metabolic

Adjustments (MOMA) was used to predict mutants with single and double reaction knockouts that overproduce ethanol and butanol. Analysis of the metabolism of these mutants also gave useful insights on the principles underlying flux redistribution for the enhanced production of these biochemicals.

2. Methods

2.1. Growth studies

Synechococcus sp. PCC 7002 was obtained from Pasteur Culture Collection. For the estimation of growth rates the organism was cultured in Sartorius glass fermenter (Biostat B Plus Univessal 2L) illuminated with external fluorescent lights (250μE). One litre (1L) of BG11+ASNIII (1:1 v/v) (initial pH = 8.2) media was inoculated to an initial OD_{730nm} of 0.1. The temperature was maintained at 38^{o}C. The culture was aerated at a rate of 0.2vvm (200ml per min) using compressed air. The agitation was maintained at 600rpm. Samples were collected in every 2 hour interval and the OD_{730nm} was measured.

2.2. Genome scale metabolic model

A previously published *isyp702* (Xu et al., 2013) model was used as a starting point to build an updated genome scale metabolic model for *Synechococcus* sp. PCC 7002. The network was further expanded and updated by exhaustive search of literature, Cyanobase (Nakao et al., 2010), pathway databases KEGG (Kanehisa and Goto, 2000) and Biocyc (Caspi et al., 2007), and annotations from RAST (Aziz et al., 2008). The fatty acid desaturation pathway was detailed to reflect the scheme reported for *Synechococcus* sp. PCC 7002 in literature (Murata et al., 1992). A detailed pathway accounting for the biosynthesis of various carotenoids found in this organism was modeled based on the study in Zhu et al. (2010). Based on an LC MS/MS based study on *Synechococcus* sp. PCC 7002 (Baran et al., 2010), 16 new transport reactions were added into the model. The updated model also incorporates seven compartments: Carboxysome, Cytoplasm, Cytoplasmic membrane, Extracellular space, Periplasmic space, Thylakoid lumen and Thylakoid membrane. The reactions were assigned to one or more of the 9 metabolic subsystems based on functional hierarchy used in KEGG BRITE. For the reaction imported directly from the published

model *isyp702* the Gene Protein Reaction (GPR) associations were cross checked and updated. This includes 54 reactions for which the EC numbers were newly assigned and 34 reactions for which the EC numbers were updated to their current EC number. The updated model includes 742 reactions and 728 genes. The model in SBML /Excel format is available with the Supplementary Information.

2.3. Formulation of biomass objective function

Maximum biomass formation was assumed to be the objective function for the calculation of flux distribution using the Flux Balance Analysis (FBA). The stoichiometric coefficients in the biomass equation are the molar fractions of the individual components in 1 gram of biomass (Table S5). The carbohydrate content of the biomass was estimated using phenol sulfuric acid method (DuBois et al., 1956). Glycogen was estimated using the method detailed in Krishnakumar et al. (2013). Glycogen was extracted using 40% KOH and digested using 2N HCl. The released sugars were estimated using glucose oxidase assay (Glucose GOD PAP kit, Biolab Diagnostics, Boisar, India). The total lipid was extracted using Bligh and Dryer method (Bligh and Dyer, 1959) and the lipid content was measured gravimetrically. DNA and RNA were first extracted with Triazol reagent and then quantified by measuring the OD at 260nm (De Mey et al., 2006). The chlorophyll and total carotenoid contents of *Synechococcus* sp. PCC 7002 were obtained from the literature (Xu et al., 2013) . The relative proportion of individual carotenoids were obtained from Takaichi and Mochimaru (2007). The contents of protein, phycocyanobilins, soluble pool and inorganic ions were obtained from Vu et al. (2013). The relative proportion of the constituents of soluble pool, inorganic ions and the content of the peptidoglycan were adopted from the *iJN678* model (Nogales et al., 2012) published for *Synechocystis* sp. PCC 6803.

2.4. Flux balance analysis

Flux balance analysis enables the prediction of flux through the Genome scale metabolic network using the knowledge of reaction stoichiometry derived from the network. In flux balance analysis the fluxes are estimated by solving the following optimization problem (Orth et al., 2010):

$$\text{Maximize} : Z_{\text{Obj}} = v_{\text{biomass}} \tag{1}$$

Subjected to the constraints:

$$S.v = 0 \tag{2}$$

$$\alpha_j \leq v_j \leq \beta_j \tag{3}$$

Where S is the stoichiometric matrix, in which the element S_{ij} is the stoichiometric coefficient of i th metabolite in the j th reaction of the metabolic network. v refers to the vector of fluxes through the reactions in the metabolic network. Assumption of metabolic steady state, where the concentration of the individual metabolites remains constant over time, results in equation 2. Since the number of metabolites is lesser than the number of reactions the equation 2 represents an under determined system of linear algebraic equations. Therefore a plurality of solutions exists for the system of linear algebraic equations represented by equation 2. Flux values are estimated by maximizing or minimizing a biologically relevant objective function with the above set of linear algebraic equations as constraints. The α_j and β_j are the lower and upper bounds on individual fluxes dictated by the reversibility and thermodynamics of the corresponding reactions.

In this study, the fluxes were estimated by optimizing the objective function and also minimizing the overall flux values (taxicab norm) using linear programming. Carbon limited photoautotrophic flux distribution was estimated in a 2 step strategy. In the first step, the model was simulated for maximum growth rate by fixing the CO_2 uptake rate at 7.54 mmol/gDW/h, which is the net CO_2 conversion rate into biomass under the experimental conditions used in this study. The the net CO_2 conversion rate was estimated from the growth rate and the carbon content of the biomass. In the second step, growth rate was fixed at the value estimated in the previous step and the CO_2 uptake rate remained fixed at 7.54 mmol/gDW/h. The flux distribution was then obtained by minimizing the photon uptake rate. Heterotrophic flux distribution was estimated by fixing the glycerol uptake rate at 10 mmol/gDW/hr and maximizing the biomass using FBA under complete absence of light. For mixotrophic metabolism, the glycerol uptake rate was set at 5mmol/gDW/hr and the biomass was maximized in the presence of unlimited light. Photons consumed per C fixed, % Carbon conversion, ratio of flux through RUBISCO to CO_2 uptake flux, ATP/NADPH ratio and PSI/PSII flux ratio for individual biomass precursors were calculated for the above conditions. This was done by maximizing flux through the sink reaction for the respective metabolite using

FBA assuming no growth. All the simulations were performed using the COBRA Toolbox running in the Matlab environment (Schellenberger et al., 2011).

2.5. Gene essentiality and synthetic lethality analysis

The minimum gene set needed for survival was enumerated using previously reported methods (Pál et al., 2006). Starting from the complete network, genes chosen randomly were sequentially knocked out one after the other. If the removal of a gene results in a viable network then it is permanently removed. If the removal of a gene results in a non-viable network then it is retained. The process is continued until no genes could be removed without compromising the network viability. The process is repeated 500 times starting with the complete network and genes that were part of the all 500 final minimal networks were classified as essential genes. The rest were all classified as non-essential genes. Those non-essential genes that were present only in some of the final minimal networks are classified as Synthetic lethal genes. The network viability criteria is 0.01 μ_w. Sensitivity and specificity of the model is calculated using equations 4 and 5 obtained from the Saha et al. (2012).

$$\text{Specificity} = \frac{\text{GG}}{\text{GG} + \text{NGG}} \tag{4}$$

$$\text{Sensitivity} = \frac{\text{NGNG}}{\text{NGNG} + \text{GNG}} \tag{5}$$

Where, GG is the number of case where there was both *in silico* and *in vivo* growth. NGG is the number of cases where there was no growth *in silico* but there was *in vivo* growth. NGNG is the number of cases where there was no growth *in silico* and *in vivo*. GNG is the number of cases where there was *in silico* growth but no growth *in vivo*.

2.6. Minimization of Metabolic Adjustment (MOMA)

Reaction deletions for the enhanced production of ethanol and butanol under photoautotrophic condition (Carbon limited) were predicted using Minimization of Metabolic Adjustment (MOMA) algorithm (Segrè et al., 2002). MOMA predicts the flux distribution and the growth rate of a mutant given the flux distribution of the wild type. This is done by minimizing the euclidean distance between the wild type flux values estimated using FBA and the mutant flux values. The

quadratic programming problem solved by MOMA is mathematically represented as:

$$\text{Minimize} : D(w, v) = \sqrt{\sum_{i=1}^{N} (w_i - v_i)^2}. \quad (6)$$

Where w_i and v_i are the flux values of i th reaction in the wild type and mutant respectively. N is the total number of reactions in the model. The wild type flux values were estimated under carbon limited photoautotrophic condition. The CO_2 uptake rate was fixed at 7.54 mmol/gDW/hr and the photon flux was unconstrained.

The reactions in model were first classified as essential and non-essential by predicting the growth rate of in-silico reaction deletion mutants using FBA. If the deletion of a reaction reduces the growth rate below 10% of the wild type growth rate then the reaction is classified as essential. Only the non-essential reactions were used for the analysis using MOMA. These reactions were knocked out one at a time in the model and the flux distribution of these in-silico mutants were predicted under carbon limited photoautotrophic condition using MOMA. The reaction knock outs that gave positive flux through ethanol and butanol secretion reactions were chosen as candidates for further analysis. The theoretical maximum production rate from these in-silico mutants were predicted using FBA by maximizing the product production with growth rate fixed at the value predicted by MOMA algorithm. To identify double reaction deletion mutant, two reaction were deleted at a time (~60000 pairs) and the flux distributions and growth rates were predicted using MOMA. The double reaction knockout mutants that showed ethanol and butanol secretion rates higher than the maximum value observed in the single reaction deletion analysis were chosen for the further analysis of theoretical maximum production rate.

3. Results and Discussion

3.1. Validation of the reconstruction

The genome scale metabolic model was able to synthesize all the necessary precursors for biomass synthesis under photoautotrophy, dark heterotrophy, dark anoxic, heterotrophy and mixotrophy conditions. Sink reactions were added for 3-(methylthio)propanoate, 2-oxo-4-methylthiobutanoate and 4-Hydroxy-benzyl alcohol, as these metabolites were not consumed by any other reaction in

the model. Under heterotrophy the reaction Ferredxoin- NADP+ oxidoreductase was reversed to enable the production of reduced ferredoxin. This reaction was the only source of reduced ferredoxin under these conditions in the model. The model predicted a growth rate of 0.18 h^{-1} under autotrophy. This is within the range of values estimated experimentally in the photobioreactor ($0.18 \pm 0.07 h^{-1}$) . The PSI/PSII ratio predicted by the model was 1.4. For photosynthetic organisms, based on theoretical estimates, this value is expected to be above 1.2 (Allen, 2003). The model estimated a photosynthetic quotient of 1.5. This value lies within the range of values estimated for other photosynthetic organisms (Shastri and Morgan, 2005). Thus, the model was able to predict the growth rate and other parameters related to photosynthesis with a good degree of accuracy.

3.2. Gene essentiality analysis

Further validation of the model was carried out through Gene essentiality analysis. The model predicted essentiality of different genes were validated against the corresponding gene knockout data available in literature for this organism. In this study two rounds of Gene essentiality analysis were performed. In the first round, gene essentiality data obtained from the literature for specific genes of *Synechococcus* sp. PCC 7002 was used. In total 4 essential and 34 non-essential genes were obtained from the reported literature. Only 20 of these genes (2 essential and 18 non-essential genes) were represented in the model. Out of the 18 non-essential genes, the model predicted 16 correctly. The two gene wrongly predicted as essential genes were SYNPCC7002_A0095 and SYNPCC7002_A0322. SYNPCC7002_A0095 codes for ADP-glucose phosphorylase necessary for glycogen synthesis. Glycogen is not essential for the *in vivo* growth of *Synechococcus* sp. PCC 7002 under photoautotrophic condition. However in the model glycogen is part of the biomass equation and therefore it is essential for *in silico* growth. The second gene SYNPCC7002_A0322 codes for the extrinsic protein of the PSII complex PsbU involved in high temperature acclimation. Temperature acclimation is beyond the scope of this modeling frame work. Therefore this wrong prediction is due to the limitation of this modeling frame work. The model was able to predict the 2 essential genes correctly. From this analysis carried out with limited number of genes the specificity of the model was calculated to be 0.89 ($GG = 16$, $NGG = 2$) and sensitivity was

1 ($NGNG = 2$, $GNG = 0$). Since the number of reported gene knockout mutants are fewer in *Synechococcus* sp. PCC 7002, the initial round of Gene essentiality analysis involved only 20 genes. The second round of Gene essentiality analysis was based on the essential and non-essential genes in *Synechocystis* sp. PCC 6803 compiled by Saha et al. (2012). The homologues of these genes in *Synechococcus* sp. PCC 7002 were identified and their essentiality was predicted using the model. The model predictions were validated against the reported data for *Synechocystis* sp. PCC 6803 (Saha et al., 2012). Since metabolic network of these two organisms are different, the conclusions from this analysis was to be taken with a caution. Out of 57 non-essential genes used in this analysis the model predicted only 44 genes as non-essential. The rest of them were predicted to be essential. Most of these 13 genes code for subunits of PSI and PSII and the rest were part of pigment metabolism. The PSI and PSII comprise a number of individual subunits and essentiality of these subunits are not well characterized. Since pigments are part of the biomass equation, they are always essential for the in-silico growth. However, in-vivo, these pigments are mostly not essential for growth. The model predicted 12 out of 13 essential genes correctly. Based on this study the specificity of the model was calculated to be 0.77 ($GG = 44$, $NGG = 13$) and the sensitivity was 0.92 ($NGNG = 12$, $GNG = 1$).

3.3. Redundancy in the Metabolic network and Metabolic robustness

Based on the gene essentiality analysis 368 genes (out of 728 genes included in the model) were found to be essential for the photoautotorophic growth. Under heterotrophic and mixotrophic growth in glycerol the number of essential genes were 335 and 348. Out of the 368 genes essential under the autotrophic conditions, 35 genes were essential only under the autotrophic condition. Most of these 35 genes code for the components of PSI and PSII. The rest code for enzymes RUBISCO, ribulose 5-phosphate 3-epimerase, ribose-5-phosphate isomerase etc involved in CO_2 fixation. The 2 genes that were essential only under heterotrophic condition code for Triose phosphate isomerase and Ferredxoin- NADP+ oxidoreductase. The former catalyze the reaction that acts as the entry point for glycerol into the central carbon metabolism. The latter was the source of reduced ferredoxin under heterotrophic condition. The 15 genes essential only under the mixotrophic condition predominantly comprised genes encoding the subunits of NADH/NADPH

plastoquinine oxidoreductase. The smaller number of essential genes when compared to the total genes in the network indicate that the metabolic network has redundant reactions or pathways. Such redundancy makes the metabolic network robust against gene perturbation. It was observed that larger number of genes were essential under autotrophic conditions than that was observed under heterotrophic or mixotrophic condition. This is consistent with previous observations in *Synechocystis* sp. PCC 6803.

Altogether 333 genes were found to be essential under all the three conditions (Figure 1). This constitutes 45.7% of the total genes included in the model. The number of essential genes predicted by the model depends on the cellular constituents covered by the biomass equations and the redundant reaction or pathways inherent to the metabolic network of the organism. Similar studies with a model of *Synechocystis* sp. PCC 6803 (iJN678) reported that only 38% of the total metabolic genes included in the model are essential (Nogales et al., 2012) as opposed to 45.7 % observed for *Synechococcus* sp. PCC 7002 in this study. This indicates that the metabolic network of *Synechococcus* sp. PCC 7002 has lesser redundancy compared to *Synechocystis* sp. PCC 6803. Since, the biomass equation included in both these models are comparable it is possible to conclude that the metabolic network of *Synechococcus* sp. PCC 7002 is less robust against gene perturbation compared to *Synechocystis* sp. PCC 6803.

The number of synthetic lethal genes is considered to be a measure of metabolic robustness. The products of synthetic lethal genes are isozymes or enzymes catalyzing reactions in different pathways with redundant/complementary essential functions (Suthers et al., 2009). Highest number of synthetic lethal genes was observed under mixotrophic condition (215) followed by the heterotrophic (196) and autotrophic conditions (176). Thus the metabolic robustness was highest under mixotrophic condition since both photoautotrophic and heterotrophic metabolism occurs simultaneously under mixotrophic condition. The metabolic robustness is the lowest under autotrophic condition.

3.4. Photoautotrophy

The central carbon flux map for the autotrophic condition showed an incomplete Tricarboxylic acid (TCA) cycle (Figure 2a). The reactions 2-oxoglutarate decarboxylase (2OGDC) and Succinate-

semialdehyde dehydrogenase (SSALY) carried a negligible flux. The incomplete TCA cycle was predominantly used to supply precursors for biosynthesis. The flux through oxidative pentose phosphate pathway was also close to zero under photoautotrophic condition. The flux through reductive pentose phosphate pathway was very high to replenish the Ribulose-1,5-bisphosphate used for fixing the CO_2. Overall the flux distribution predicted under photoautotrophic condition was similar to what was reported for other cyanobacteria (Shastri and Morgan, 2005; Knoop et al., 2013). The model also predicted that the electrons were cycled around PSI through the NDH-1 complex. The cyclic electron flow through the NDH-1 was reported to give higher quantum yield when compared to the other alternative electron flow paths through the NDH-2 and FQR (Nogales et al., 2012). The minimum photon uptake rate required for the optimum growth rate was estimated to be 109.63 mmol/gDW/h. The quantum yield was calculated as 0.069. This value is close to and slightly lesser than the value reported for *Synechocystis* sp. PCC 6803 (0.072) (Nogales et al., 2012). Further, the theoretical yields of different biosynthetic precursors were analyzed (Table S3). Quantum yield (i.e. Photons consumed per unit CO_2 fixed) was high for the intermediates from the reductive pentose phosphate pathway branch of the central carbon metabolism. Among the amino acids, quantum yield was highest for glycine (12.7 mol/mol CO_2 fixed) and lowest for serine (7.6 mol/mol CO_2 fixed).

3.5. Heterotrophy in Glycerol

Glycerol is the only reported organic carbon source that supports heterotrophic growth of *Synechococcus* sp. PCC 7002 (Xu et al., 2006). The heterotrophic growth using glycerol was simulated assuming a glycerol uptake rate of 10mmol/gDW/h. Glycerol gets converted to glycerol-3-phosphate by the enzyme Glycerol kinase and then to dihydroxy acetone phosphate (DHAP) by NADP+ dependant-glycerol dehydrogenase. The DHAP then enters the glycolysis pathway through triose phosphate isomerase. The predicted flux distribution (Figure 2b) was significantly different from what was observed under dark respiratory metabolism in cyanobacteria (Knoop et al., 2013). During dark respiration using endogenous glycogen the reducing equivalents get generated from the complete oxidation of G6P through the oxidative pentose phosphate pathway. No cyclic flux was observed through the TCA cycle. Under glycerol heterotrophy, however, flux

through oxidative pentose phosphate pathway was close to zero. The TCA cycle was complete with significant flux through 2OGDC and SSALY. This shows that the bulk of reductants needed for biosynthesis are generated by the complete oxidation of glycerol through TCA cycle rather than through the oxidative pentose phosphate pathway. The higher flux through the TCA cycle was also supported by a flux value of same order through the PEP carboxylase reaction. The reductive pentose phosphate pathway was not active under the heterotrophic condition. There was a limited flux through transketolase to produce E4P and XU5P. E4P is a precursor for the synthesis of amino acids. The XU5P is used to produce R5P which is a precursor for nucleic acids and histidine. The biomass yield was calculated as 0.46g of biomass/g of glycerol.

3.6. Mixotrophy in Glycerol

Under mixotrophic metabolism the TCA cycle was incomplete with zero flux through 2OGDC and SSALY (Figure 2c). However there was a higher flux through the TCA branches as similar to that under heterotrophic metabolism. Flux through the PEP carboxylase reaction was of the same order as observed in the TCA cycle branches. Significant flux was observed through RUBISCO to fix the CO_2 evolved from the decarboxylation reactions in the metabolic network. Thus the reductive branch of the pentose phosphate pathway was also active to replenish the RUBP used in fixing the CO_2. ^{13}C-MFA based study of the mixotrophic metabolism of *Cyanothece* sp. ATCC 51142 under nitrate sufficient condition (Alagesan et al., 2013) also showed similar flux distribution in the central carbon metabolism. The PSI/PSII ratio under mixotrophic condition was 2.5, which was higher than that observed under the autotrohic growth. Thus, the flux through cyclic elecron flow is much higher than that through the linear electron flow. The PSI/PSII ratio calculated for the biosynthesis of individual biomass precursors were also higher than that observed under photoautotrophic condition. This may be due to the excess of reducing equivalents available under mixotrophy (Ludwig and Bryant, 2012). Thus, the photosynthetic electron transport is predominantly used for the ATP synthesis.

3.7. Prediction of reaction deletions for over production of Ethanol and Butanol

GSMMs are widely used to predict reaction knockouts that enhance the production of specific biochemicals of interest. Numerous algorithms have been developed to predict pathway engineer-

ing strategies using the GSMMs. Reaction deletions that lead to growth coupled product formation are more attractive strategy for metabolic engineering as they result very stable phenotype. However, cyanobacteria needs a large number of reaction knockouts to couple the product production with the growth under photoautotrophic condition (Vu et al., 2013; Erdrich et al., 2014). Therefore this study was confined to predicting single and double reaction deletions for enhanced production of ethanol and butanol under carbon limited photoautotrophic condition. Wild type *Synechococcus* sp. PCC 7002 cannot produce any of the above two products as they do not have the necessary enzymes. Hence heterologous pathways were added to the model to enable the production of these biochemicals. To allow ethanol production pyruvate dehydrogenase was added into the model. For butanol production, the pathway was adopted from Lan and Liao (2012). The single reaction deletions that resulted in higher production rate of ethanol and butanol are shown in table 2. Glutamate dehydrogenase was predicted to be the best knock out candidate for the enhancing the production of both ethanol and butanol. It was recently reported that for the photautotrophic metabolism of cyanobacteria all *in silico* predicted knockouts alter the ATP/NADPH balance (Erdrich et al., 2014). The photosynthetic linear electron transport generate ATP and NADPH in the ratio 1.29:1 (Allen, 2003). However, as per the model prediction, ATP and NADPH are required in the stoichiometric ratio of 2.33:1 for biomass synthesis. Cyclic electron transport around PSI and other alternative electron transport mechanisms are used to maintain this high ATP: NADPH ratio (2.33:1) by producing more ATP and oxidizing excess NADPH. Any alteration in the network that decreases this ratio will result in excess NADPH that could be detrimental for growth. Glutamate dehydrogenase is involved in the NADPH dependant assimilation of nitrate. Knocking out this reaction disturb the ATP/NADPH balance needed for biomass production because of excess NADPH. Consumption of NADPH for the production of the target biochemicals restore optimal ATP:NADPH ratio required for biomass synthesis. As observed in the table 2 the production rates were very low, however, the maximum production rate estimates correspond to a conversion of ~10% fixed carbon.

Further an exhaustive analysis of all possible double reaction deletion mutants was also carried out to identify mutants with production rates higher than that observed for single reaction deletion. Table 3 shows the best candidates identified. It was observed that all the high yielding

double reaction deletion candidates had Glutamate dehydrogenase as one of the two reactions to be deleted. In case of butanol producing mutants the other gene of each of these pairs were from the Citramalate pathway. This pathway is used for the synthesis of α-Ketobutyrate, a precursor for Isoleucine. In the absence of this pathway α-Ketobutyrate has to be generated from threonine by the action of Threonine dehydratase. This results in the wastage of excess ATP, there by altering the ATP/NADPH balance. In case of ethanol producing double reaction deletion mutants, glutamate dehydrogenase was knocked out in combination with NADPH sink reactions such as Lactate dehydrogenase and MEHLER reaction. The production rates predicted by MOMA for the double reaction deletion were significantly higher than the single reaction deletions. However it is to be noted that in case of both single and double reaction deletions the maximum possible production rates predicted by FBA were almost same. Thus, the model predicts that in *Synechococcus* sp. PCC 7002 a maximum of only ~10% of the fixed carbon could be redirected by deleting two reactions at a time.

4. Conclusion

An updated GENRE was developed for Synechococcus sp. PCC 7002 and validated using available experimental data. Only 45.7% of the genes in the model were predicted to be essential under all studied conditions. Under photoautotrophic conditions the TCA cycle was open and mostly used for synthesizing the precursors for amino acids. Under heterotrophic growth with glycerol, the TCA cycle was closed and acted as the main source of reducing equivalents. Analysis of double reaction knockouts predicted that only ~10% of the fixed carbon can be diverted for the production of ethanol and butanol.

Acknowledgement

The work was supported by a research grant provided by Reliance Industries Ltd., Mumbai and Department of Biotechnology, Government of India (Grant No: BT/EB/PAN IIT/2012).

Supporting Information

Table S1.pdf -List of essential and non-essential genes obtained from the literature for validating the model.

Table S2.xlsx -Absolute flux values of reactions in the central carbon metabolism photoautotrophic, heterotrophic and mixotrophic growth conditions.

Table S3.xlsx -Model based estimates of quantum yield (photon utilized per C fixed) for biomass precursors under photoautotrophy.

Table S4.xlsx -Comparision of model predicted PSI/PSII flux ratio under mixotrophy and photoautotrophy.

Table S5.xlsx -Constituents of the biomass equation used in the model.

Genome scale metabolic model isyp728.xml -Genome scale model in the SBML format.

Genome scale metabolic model isyp728.xls -Genome scale model in the Excel format.

References

A. Parmar, N. K. Singh, A. Pandey, E. Gnansounou, D. Madamwar, Cyanobacteria and microalgae: A positive prospect for biofuels, Bioresource Technology 102 (22) (2011) 10163–10172, ISSN 09608524, doi: 10.1016/j.biortech.2011.08.030.

M. Ludwig, D. A. Bryant, Synechococcus sp. Strain PCC 7002 Transcriptome: Acclimation to Temperature, Salinity, Oxidative Stress, and Mixotrophic Growth Conditions, Frontiers in Microbiology 3 (2012) 1–14, ISSN 1664-302X, doi:10.3389/fmicb.2012.00354.

Y. Xu, R. M. Alvey, P. O. Byrne, J. E. Graham, G. Shen, D. A. Bryant, Expression of Genes in Cyanobacteria: Adaptation of Endogenous Plasmids as Platforms for High-Level Gene Expression in Synechococcus sp. PCC 7002, in: Methods in molecular biology, vol. 684, ISBN 9781607619253, 273–293, doi:10.1007/978-1-60761-925-3_21, 2011.

Y. Xu, L. T. Guerra, Z. Li, M. Ludwig, G. C. Dismukes, D. A. Bryant, Altered carbohydrate metabolism in glycogen synthase mutants of Synechococcus sp. strain PCC 7002: Cell factories for soluble sugars., Metabolic engineering 16 (1) (2013) 56–67, ISSN 1096-7184, doi:10.1016/j.ymben.2012.12.002.

X. Feng, L. Page, J. Rubens, L. Chircus, P. Colletti, H. B. Pakrasi, Y. J. Tang, Bridging the Gap between Fluxomics and Industrial Biotechnology, Journal of Biomedicine and Biotechnology 2010 (2010) 1–13, ISSN 1110-7243, doi:10.1155/2010/460717.

J. D. Orth, I. Thiele, B. Ø. Palsson, What is flux balance analysis?, Nature biotechnology 28 (2010) 245–248, ISSN 1087-0156, doi:10.1038/nbt.1614.

N. E. Lewis, H. Nagarajan, B. O. Palsson, Constraining the metabolic genotypephenotype relationship using a phylogeny of in silico methods, Nature Reviews Microbiology 10 (4) (2012) 291–305, ISSN 1740-1526, doi:10.1038/nrmicro2737.

D. Segrè, D. Vitkup, G. M. Church, Analysis of optimality in natural and perturbed metabolic networks., Proceedings of the National Academy of Sciences of the United States of America 99 (23) (2002) 15112–15117, ISSN 00278424, doi:10.1073/pnas.232349399.

J. H. Park, K. H. Lee, T. Y. Kim, S. Y. Lee, Metabolic engineering of Escherichia coli for the production of L-valine based on transcriptome analysis and in silico gene knockout simulation., Proceedings of the National Academy of Sciences of the United States of America 104 (19) (2007) 7797–802, ISSN 0027-8424, doi:10.1073/pnas.0702609104.

T. T. Vu, E. A. Hill, L. A. Kucek, A. E. Konopka, A. S. Beliaev, J. L. Reed, Computational evaluation of Synechococcus sp. PCC 7002 metabolism for chemical production, Biotechnology Journal 8 (2013) 619–630, ISSN 18606768, doi:10.1002/biot.201200315.

M. Nakao, S. Okamoto, M. Kohara, T. Fujishiro, T. Fujisawa, S. Sato, S. Tabata, T. Kaneko, Y. Nakamura, CyanoBase: the cyanobacteria genome database update 2010, Nucleic Acids Research 38 (Database Issue) (2010) D379–D381, ISSN 0305-1048, doi:10.1093/nar/gkp915.

M. Kanehisa, S. Goto, KEGG: Kyoto Encyclopaedia of Genes and Genomes, Nucl. Acids Res. 28 (1) (2000) 27–30, ISSN 0305-1048, doi:10.1093/nar/28.1.27.

R. Caspi, H. Foerster, C. A. Fulcher, P. Kaipa, M. Krummenacker, M. Latendresse, S. Paley, S. Y. Rhee, A. G. Shearer, C. Tissier, T. C. Walk, P. Zhang, P. D. Karp, The MetaCyc Database of metabolic pathways and enzymes and the BioCyc collection of Pathway/Genome Databases, Nucleic Acids Research 36 (Database Issue) (2007) D623–D631, ISSN 0305-1048, doi:10.1093/nar/gkm900.

R. K. Aziz, D. Bartels, A. A. Best, M. DeJongh, T. Disz, R. A. Edwards, K. Formsma, S. Gerdes, E. M. Glass, M. Kubal, F. Meyer, G. J. Olsen, R. Olson, A. L. Osterman, R. A. Overbeek, L. K. McNeil, D. Paarmann, T. Paczian, B. Parrello, G. D. Pusch, C. Reich, R. Stevens, O. Vassieva, V. Vonstein, A. Wilke, O. Zagnitko, The RAST Server: Rapid Annotations using Subsystems Technology, BMC Genomics 9 (2008) 75, ISSN 1471-2164, doi:10.1186/1471-2164-9-75.

N. Murata, H. Wada, Z. Gombos, Modes of Fatty-Acid Desaturation in Cyanobacteria, Plant and Cell Physiology 33 (7) (1992) 933–941, ISSN 00320781, doi:10.1017/S1473550408004175.

Y. Zhu, J. E. Graham, M. Ludwig, W. Xiong, R. M. Alvey, G. Shen, D. A. Bryant, Roles of xanthophyll carotenoids in protection against photoinhibition and oxidative stress in the cyanobacterium Synechococcus sp. strain PCC 7002., Archives of biochemistry and biophysics 504 (1) (2010) 86–99, ISSN 1096-0384, doi:10.1016/j.abb.2010.07.007.

R. Baran, B. P. Bowen, N. J. Bouskill, E. L. Brodie, S. M. Yannone, T. R. Northen, Metabolite identification in synechococcus sp. PCC 7002 using untargeted stable isotope assisted metabolite profiling, Analytical Chemistry 82 (21) (2010) 9034–9042, ISSN 00032700, doi:10.1021/ac1020112.

M. DuBois, K. A. Gilles, J. K. Hamilton, P. A. Rebers, F. Smith, Colorimetric Method for Determination of Sugars and Related Substances, Analytical Chemistry 28 (3) (1956) 350–356, ISSN 0003-2700, doi:10.1021/ac60111a017.

S. Krishnakumar, S. B. Gaudana, G. A. Viswanathan, H. B. Pakrasi, P. P. Wangikar, Rhythm of carbon and nitrogen fixation in unicellular cyanobacteria under turbulent and highly aerobic conditions, Biotechnology and Bioengineering 110 (9) (2013) 2371–2379, ISSN 00063592, doi:10.1002/bit.24882.

E. G. Bligh, W. J. Dyer, A rapid method of total lipid extraction and purification, Canadian Journal of Biochemistry and Physiology 37 (8) (1959) 911–917.

M. De Mey, G. Lequeux, J. Maertens, S. De Maeseneire, W. Soetaert, E. Vandamme, Comparison of DNA and RNA quantification methods suitable for parameter estimation in metabolic modeling of microorganisms, Analytical Biochemistry 353 (2006) 198–203, ISSN 00032697, doi:10.1016/j.ab.2006.02.014.

S. Takaichi, M. Mochimaru, Carotenoids and carotenogenesis in cyanobacteria: Unique ketocarotenoids and carotenoid glycosides, Cellular and Molecular Life Sciences 64 (2007) 2607–2619, ISSN 1420682X, doi:10.1007/s00018-007-7190-z.

J. Nogales, S. Gudmundsson, E. M. Knight, B. O. Palsson, I. Thiele, Detailing the optimality of photosynthesis in cyanobacteria through systems biology analysis, Proceedings of the National Academy of Sciences 109 (2012) 2678–2683, ISSN 0027-8424, doi:10.1073/pnas.1117907109.

J. Schellenberger, R. Que, R. M. T. Fleming, I. Thiele, J. D. Orth, A. M. Feist, D. C. Zielinski, A. Bordbar, N. E. Lewis, S. Rahmanian, J. Kang, D. R. Hyduke, B. Ø. Palsson, Quantitative prediction of cellular metabolism with constraint-based models: the COBRA Toolbox v2.0, Nature Protocols 6 (9) (2011) 1290–1307, ISSN 1754-2189, doi:10.1038/nprot.2011.308.

C. Pál, B. Papp, M. J. Lercher, P. Csermely, S. G. Oliver, L. D. Hurst, Chance and necessity in the evolution of minimal metabolic networks., Nature 440 (7084) (2006) 667–670, ISSN 0028-0836, doi:10.1038/nature04568.

R. Saha, A. T. Verseput, B. M. Berla, T. J. Mueller, H. B. Pakrasi, C. D. Maranas, Reconstruction and Comparison of the Metabolic Potential of Cyanobacteria Cyanothece sp. ATCC 51142 and Synechocystis sp. PCC 6803, PLoS ONE 7 (10) (2012) e48285, ISSN 19326203, doi:10.1371/journal.pone.0048285.

J. F. Allen, Cyclic, pseudocyclic and noncyclic photophosphorylation: New links in the chain, Trends in Plant Science 8 (1) (2003) 15–19, ISSN 13601385, doi:10.1016/S1360-1385(02)00006-7.

A. A. Shastri, J. A. Morgan, Flux balance analysis of photoautotrophic metabolism., Biotechnology progress 21 (2005)

1617–26, ISSN 8756-7938, doi:10.1021/bp050246d.

P. F. Suthers, A. Zomorrodi, C. D. Maranas, Genome-scale gene/reaction essentiality and synthetic lethality analysis., Molecular systems biology 5 (301) (2009) 301, ISSN 1744-4292, doi:10.1038/msb.2009.56.

H. Knoop, M. Gründel, Y. Zilliges, R. Lehmann, S. Hoffmann, W. Lockau, R. Steuer, Flux Balance Analysis of Cyanobacterial Metabolism: The Metabolic Network of Synechocystis sp. PCC 6803, PLoS Computational Biology 9 (6), ISSN 1553734X, doi:10.1371/journal.pcbi.1003081.

D. Xu, X. Liu, C. Guo, J. Zhao, Methylglyoxal detoxification by an aldo-keto reductase in the cyanobacterium Synechococcus sp. PCC 7002., Microbiology 152 (2006) 2013–21, ISSN 1350-0872, doi:10.1099/mic.0.28870-0.

S. Alagesan, S. B. Gaudana, A. Sinha, P. P. Wangikar, Metabolic flux analysis of Cyanothece sp. ATCC 51142 under mixotrophic conditions, Photosynthesis Research 118 (2013) 191–198, ISSN 01668595, doi:10.1007/s11120-013-9911-5.

P. Erdrich, H. Knoop, R. Steuer, S. Klamt, Cyanobacterial biofuels: new insights and strain design strategies revealed by computational modeling, Microbial Cell Factories 13 (2014) 128, ISSN 1475-2859, doi:10.1186/s12934-014-0128-x.

E. I. Lan, J. C. Liao, ATP drives direct photosynthetic production of 1-butanol in cyanobacteria, Proceedings of the National Academy of Sciences 109 (16) (2012) 6018–6023, ISSN 0027-8424, doi:10.1073/pnas.1200074109.

J. Oliveros, Venny. An interactive tool for comparing lists with Venn's diagrams., 2007.

Figure Captions

Figure 1: Comparison of essential genes under the computationally evaluated conditions. The Venn diagram was constructed using Venny 2.0.2 (Oliveros, 2007)

Figure 2: (a) Flux distribution predicted under Photoautotrophic condition. The CO_2 uptake was fixed at 7.54 mmol/gDW/hr. (b) Flux distribution predicted under Heterotrophic growth in glycerol. The glycerol uptake rate was fixed at 10mmol/gDW/hr. (c) Flux distribution under Mixotrophic growth in glycerol. The glycerol uptake rate was fixed at 5mmol/gDW/hr. The thickness of the arrows are proportional to the flux through corresponding reactions. Refer Table S2 in Supporting Information for the absolute value of fluxes. Abbreviations: 3PGA, 3-phoshoglycerate; ACA, acetyl-CoA; AKG, α-ketoglutarate; CO_2, carbon-di-oxide; DHAP, dihydroxyacetone phosphate; E4P, erythrose-4-phosphate; F6P, fructose-6-phosphate; FUM, fumarate; G1P, glucose-1-phosphate; G6P, glucose-6-phosphate; GAP, glyceraldehyde-3-phosphate; GLY, glycogen; ICI, isocitrate; MAL, malate; OAA, oxaloacetate; PEP, phosphoenolpyruvate; PYR, pyruvate; R5P, ribose-5-phosphate; RU5P, ribulose-5-phosphate; RUBP, ribulose-1,5-bisphosphate; S7P, sedoheptulose-7-phosphate; SBP, sedoheptulose-1,7-bisphosphate; SUCC, succinate.

Table 1: The relative content of different biomass components as used to formulate the biomass objective function in the model

Biomass component	**mg/mg Dry weight**	**Scaled value**	**Reference**
Protein	0.580	0.608	Vu et al. (2013)
Carbohydrates	0.151	0.158	This study
Lipids	0.039	0.041	This study
RNA	0.068	0.071	This study
DNA	0.002	0.002	This study
Phycocyanobilins	0.009	0.009	Vu et al. (2013)
chlorophyll	0.037	0.039	Xu et al. (2013)
soluble pool	0.025	0.026	Vu et al. (2013)
inorganic ions	0.009	0.009	Vu et al. (2013)
carotenoids	0.011	0.011	Xu et al. (2013)
peptidoglycan	0.025	0.026	Nogales et al. (2012)
	1.034	1.000	

Table 2: Single reaction knock out analysis. Reaction knock outs that result in positive flux through the product formation reactions were identified using MOMA. Those that gave the highest production rates are listed below.

Reaction	μ_k/μ_w [a]	**Rate of production** [b] **(mmol/gDW/h)**	**Theoretical maximum production rate** [c] **(mmol/gDW/h)**	**% Carbon conversion to product** [d]
Butanol				
Glutamate Dehydrogenase	0.9	0.026	0.19	10
Glyceraldehyde-3-Phosphate Dehydrogenase (NAD)	0.97	0.0075	0.06	3.2
Ethanol				
Glutamate Dehydrogenase	0.91	0.0095	0.36	9.5

[a] μ_k is the growth rate of the mutant predicted by MOMA and μ_w is the growth rate of the wild type predicted by FBA

[b] Flux value for the product formation reaction predicted by MOMA

[c] obtained by maximizing the product formation using FBA while fixing the growth rate at the value predicted by MOMA

[d] This value corresponds to the theoretical maximum production rate

Table 3: Summary of results of double reaction knock out analysis. Reactions were knocked out two at a time and the fluxes through the product formation reactions were predicted using MOMA. Listed are reaction pairs that gave the highest production rates when knocked out

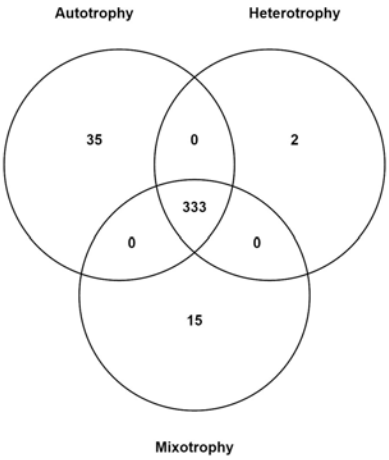
Reaction 1	**Reaction 2**	μ_k/μ_w[a]	**Rate of production**[b] **(mmol/gDW/h)**	**Theoretical maximum production rate**[c] **(mmol/gDW/h)**	**% Carbon conversion to product**[d]
Butanol					
2-methylmaleate Hydrolyase	Glutamate Dehydrogenase	0.88	0.047	0.2	10.6
(R)-2-methylmaleate Hydrolyase	Glutamate Dehydrogenase	0.88	0.047	0.2	10.6
Citramalate Synthase	Glutamate Dehydrogenase	0.88	0.047	0.2	10.6
Erythro methyl maleate Dehydrogenase	Glutamate Dehydrogenase	0.88	0.047	0.2	10.6
Ethanol					
Lactate Dehydrogenase	Glutamate Dehydrogenase	0.9	0.038	0.34	9
MEHLER	Glutamate Dehydrogenase	0.89	0.02	0.36	9.5

[a] μ_k is the growth rate of the mutant predicted by MOMA and μ_w is the growth rate of the wild type predicted by FBA

[b] Flux value for the product formation reaction predicted by MOMA

[c] obtained by maximizing the product formation using FBA while fixing the growth rate at the value predicted by MOMA

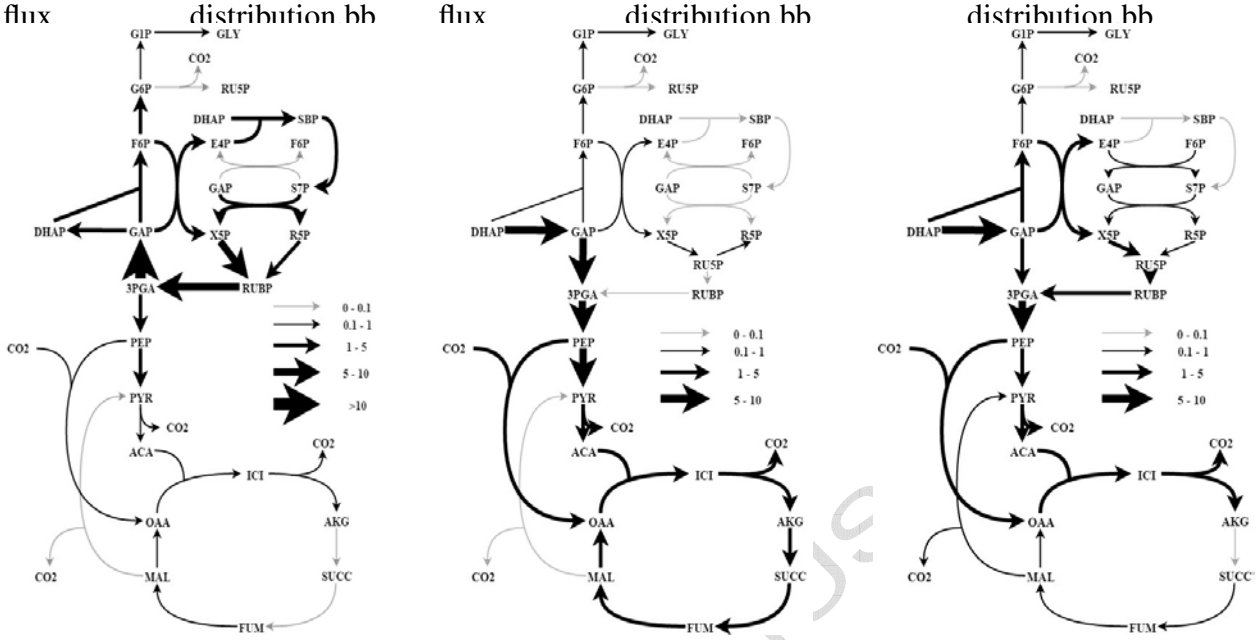
[d] This value corresponds to the theoretical maximum production rate



of coresets.jpeg of coresets.bb

Figure 1: Comparison of essential genes under the computationally evaluated conditions. The Venn diagram was constructed using Venny 2.0.2 (Oliveros, 2007)

flux distribution.jpeg flux distribution.jpeg flux distribution.jpeg flux distribution.jpeg flux distribution.jpeg flux distribution.jpeg



(a) (b) (c)

Figure 2: (a) Flux distribution predicted under Photoautotrophic condition. The CO_2 uptake was fixed at 7.54 mmol/gDW/hr. (b) Flux distribution predicted under Heterotrophic growth in glycerol. The glycerol uptake rate was fixed at 10mmol/gDW/hr. (c) Flux distribution under Mixotrophic growth in glycerol. The glycerol uptake rate was fixed at 5mmol/gDW/hr. The thickness of the arrows are proportional to the flux through corresponding reactions. Refer Table S2 in Supporting Information for the absolute value of fluxes. Abbreviations: 3PGA, 3-phoshoglycerate; ACA, acetyl-CoA; AKG, α-ketoglutarate; CO_2, carbon-di-oxide; DHAP, dihydroxyacetone phosphate; E4P, erythrose-4-phosphate; F6P, fructose-6-phosphate; FUM, fumarate; G1P, glucose-1-phosphate; G6P, glucose-6-phosphate; GAP, glyceraldehyde-3-phosphate; GLY, glycogen; ICI, isocitrate; MAL, malate; OAA, oxaloacetate; PEP, phosphoenolpyruvate; PYR, pyruvate; R5P, ribose-5-phosphate; RU5P, ribulose-5-phosphate; RUBP, ribulose-1,5-bisphosphate; S7P, sedoheptulose-7-phosphate; SBP, sedoheptulose-1,7-bisphosphate; SUCC, succinate.

Highlights for review

1) An updated GENRE was developed for Synechococcus sp PCC 7002.
2) Only 45.7% of the metabolic genes were essential under all studied conditions.
3) Under autotrophy TCA cycle was open and used to supply precursors for amino acids.
4) TCA cycle was main source of reducing equivalents under glycerol heterotrophy.
5) Double gene knock outs could divert only 10% of the fixed carbon.