2012

Reconstruction and Comparison of the Metabolic Potential of Cyanobacteria *Cyanothece* sp. ATCC 51142 and *Synechocystis* sp. PCC 6803

Rajib Saha  
*National University of Singapore, rsaha2@unl.edu*

Alex T. Verseput  
*The Pennsylvania State University*

Bertram M. Berla  
*Washington University*

Thomas J. Mueller  
*The Pennsylvania State University*

Himadri B. Pakrasi  
*Washington University*

*See next page for additional authors*

Follow this and additional works at: http://digitalcommons.unl.edu/chemengall
Reconstruction and Comparison of the Metabolic Potential of Cyanobacteria *Cyanothece* sp. ATCC 51142 and *Synechocystis* sp. PCC 6803

Rajib Saha1, Alex T. Verseput1, Bertram M. Berla2, Thomas J. Mueller1, Himadri B. Pakrasi2,3, Costas D. Maranas1*

1 Department of Chemical Engineering, The Pennsylvania State University, University Park, Pennsylvania, United States of America, 2 Department of Energy, Environmental, and Chemical Engineering, Washington University, St. Louis, Missouri, United States of America, 3 Department of Biology, Washington University, St. Louis, Missouri, United States of America

Abstract

Cyanobacteria are an important group of photoautotrophic organisms that can synthesize valuable bio-products by harnessing solar energy. They are endowed with high photosynthetic efficiencies and diverse metabolic capabilities that confer the ability to convert solar energy into a variety of biofuels and their precursors. However, less well studied are the similarities and differences in metabolism of different species of cyanobacteria as they pertain to their suitability as microbial production chassis. Here we assemble, update and compare genome-scale models (iCyt773 and iSyn731) for two phylogenetically related cyanobacterial species, namely *Cyanothece* sp. ATCC 51142 and *Synechocystis* sp. PCC 6803. All reactions are elementally and charge balanced and localized into four different intracellular compartments (i.e., periplasm, cytosol, carboxysome and thylakoid lumen) and biomass descriptions are derived based on experimental measurements. Newly added reactions absent in earlier models (266 and 322, respectively) span most metabolic pathways with an emphasis on lipid biosynthesis. All thermodynamically infeasible loops are identified and eliminated from both models. Comparisons of model predictions against gene essentiality data reveal a specificity of 0.94 (94/100) and a sensitivity of 1 (19/19) for the *Synechocystis* iSyn731 model. The diurnal rhythm of *Cyanothece* 51142 metabolism is modeled by constructing separate (light/dark) biomass equations and introducing regulatory restrictions over light and dark phases. Specific metabolic pathway differences between the two cyanobacteria alluding to different bio-production potentials are reflected in both models.

Introduction

Cyanobacteria represent a widespread group of photosynthetic prokaryotes [1]. By contributing oxygen to the atmosphere, they played an important role in the Precambrian phase [2]. Cyanobacteria are primary producers in aquatic environments and contribute significantly to biological carbon sequestration, O2 production and the nitrogen cycle [3–5]. Their inherent photosynthetic capability and ease in genetic modifications are two significant advantages over other microbes in the industrial production of valuable bioproducts [6]. In contrast to other microbial production processes requiring regionally limited cellulosic feedstocks, cyanobacteria only need CO2, sunlight, water and a few mineral nutrients to grow [6]. Sunlight is the most abundant source of energy on earth. The incident solar flux onto the USA alone is approximately 23,000 terawatts which dwarfs the global energy usage of 3.16 terawatts [7]. Cyanobacteria perform photosynthesis more efficiently than terrestrial plants (3–9% vs. 2.4–3.7%) [8]. The short life cycle and transformability of cyanobacteria combined with a detailed understanding of their biochemical pathways are significant advantages of cyanobacteria as efficient platforms for harvesting solar energy and producing bio-products such as short chain alcohols, hydrogen and alkanes [6].

The genus *Cyanothece* includes unicellular cyanobacteria that can fix atmospheric nitrogen. *Cyanothece* sp. ATCC 51142 (hereafter *Cyanothece* 51142) is one of the most potent diazotrophs characterized and the first to be completely sequenced [9]. Studies show that it can fix atmospheric nitrogen at rates higher than many filamentous cyanobacteria and also accommodate the biochemically incompatible processes of photosynthesis and nitrogen fixation within the same cell by temporally separating them [10]. *Synechocystis* sp. PCC 6803 (hereafter *Synechocystis* 6803), the first photosynthetic organism with a completely sequenced genome [11], is probably the most extensively studied model organism for photosynthetic processes [12]. It is also closely related to *Cyanothece* 51142 and shares many characteristics with all *Cyanothece* [9]. The genome of *Cyanothece* 51142 is about 35% larger than that of *Synechocystis* 6803 mostly due to the presence of nitrogen fixation and temporal regulation related genes in *Cyanothece* 51142 [9].


Editor: John Parkinson, Hospital for Sick Children, Canada

Received June 29, 2012; Accepted September 21, 2012; Published October 31, 2012

Copyright: © 2012 Saha et al. This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Funding: This study was supported by a grant from the United States Department of Energy, Biological and Environmental Research (DOE-BER), grant DE-SC0006870. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing Interests: The authors have declared that no competing interests exist.

* E-mail: costas@psu.edu
Synecocystis 6803 has been the subject of many targeted genetic manipulations (e.g., expression of heterologous gene products) as a photo-biological platform for the production of valuable chemicals such as poly-beta-hydroxybutyrate, isoprene, hydrogen and biofuels [12-20]. However, genetic tools for Cyanothece 51142 are still lacking thus hampering its wide use as a bio-production strain even though it has many attractive native pathways. For example, Cyanothece 51142 can produce (in small amounts) pentadecane and other hydrocarbons while containing a novel (though incomplete) non-fermentative pathway for producing butanol [21,22].

A breakthrough in solar biofuel production will require following one of two strategies: 1) obtaining photosynthetic strains that naturally have high-throughput pathways analogous to those in known biofuel producers, or 2) creating cellular environments conducive for heterologous enzyme function. Despite its attractive capabilities including nitrogen fixation and H2 production [19], unfortunately genetic tools are not currently available to efficiently test engineering interventions directly for Cyanothece 51142. Therefore, a promising path forward may be to use Synecocystis 6803 as a “proxy” (for which a comprehensive genetic toolkit is available) and subsequently transfer knowledge gained during experimentation with Synecocystis 6803 to Cyanothece 51142. This requires high quality metabolic models for both organisms. Comprehensive genome-wide metabolic reconstructions include the complete inventory of metabolic transformations of a given cyanobacterial system. Comparison of the metabolic capabilities of Cyanothece 51142 and Synecocystis 6803 derived from their corresponding genome-scale models will provide valuable insights into their niche biological functions and also open up new avenues for economical biofuel production.

Genome-scale models (GSM) contain gene to protein to reaction associations (GRPs) along with a stoichiometric representation of all possible biotransformations known to occur in an organism combined with a set of appropriate regulatory constraints on each reaction flux [23,24]. By defining the global metabolic space and flux distribution potential, GSMs can assess allowable cellular phenotypes under specific environmental conditions [23,24]. The first genome-scale model for Cyanothece 51142 was recently published [25]. The authors addressed the complexity of the electron transport chain (ETC) and explored further the specific roles of photosystem I (PSI) and photosystem II (PSII). In contrast, Synecocystis 6803 has been the target for metabolic model reconstruction for quite some time [12,26–32]. Most of these earlier efforts for Synecocystis 6803 focused on only central metabolism [26–28]. Knoop et al. [12] and Montagud et al. [30,31] developed genome-scale models for Synecocystis 6803, analyzed growth under different conditions, identified gene knockout candidates for enhanced succinate production and performed flux coupling analysis to detect potential bottlenecks in ethanol and hydrogen production. A more recent model describes in detail the photosynthetic apparatus, identifies alternate electron flow pathways and highlights the high photosynthetic robustness of Synecocystis 6803 during photoautotrophic metabolism [32]. All these efforts have brought about an improved understanding of the metabolic capabilities of Synecocystis 6803 and cyanobacterial systems in general.

This paper introduces high-quality genome-scale models for Cyanothece 51142 (Cyt773) and Synecocystis 6803 (Syn731) (as shown in Table 1) that integrate all recent developments [25,32], supplements them with additional literature evidence and highlights their similarities and differences (see Files S1, S2, S7 and S8 for detailed description of these models). As many as 322 unique reactions are introduced in the Synecocystis iSyn731 model and 266 in Cyanothece iCyt773. New pathways include, among many, a TCA bypass [33], heptadecane biosynthesis [21] and detailed fatty acid biosynthesis in iSyn731 and comprehensive lipid and pigment biosynthesis and pentadecane biosynthesis [21] in Cyt773. For the first time, not only extensive gene essentiality data [34] is used to assess the quality of the developed model (i.e., Syn731) but also the allowible model metabolic phenotypes are contrasted against MFA flux data [35]. The diurnal rhythm of Cyanothece metabolism is modeled for the first time via developing separate (light/dark) biomass equations and regulating metabolic fluxes based on available protein expression data over light and dark phases [36].

Materials and Methods

Measurement of Biomass Precursors

**Growth conditions.** Wild-type Synecocystis 6803 and Cyanothece 51142 were grown for several days from an initial OD730 of ~0.05 to ~0.4. Synecocystis 6803 was grown in BG-11 medium [92] and Cyanothece 51142 in ASP2 medium [93] with (+N) or without (~N) nitrate. All cultures were grown in shake flasks with continuous illumination of ~100 μmol photons/m2/sec provided from cool white fluorescent tubes. Synecocystis was maintained at 30°C and Cyanothece at 25°C. For Synecocystis, the illumination was constant and doubling time was ~24 hours. Cyanothece alternated between 12 hours of light and 12 hours of darkness, with a doubling time of ~48 hours.

**Pigments.** 1 mL of cells of both Synecocystis 6803 and Cyanothece 51142 (from light and dark phases) was pelletted and extracted twice with 5 mL 80% aqueous acetone and the extracts pooled. Spectra of this extract and of a sample of whole cells were taken on a DW2000 spectrophotometer (Olis, GA, USA) against 80% acetone or BG-11 media as a reference. Chlorophyll a contents were calculated as reported [94] from the acetone extract. Total carotenoid concentrations were also calculated from the acetone extract according to a published method [95]. The relative amounts of different carotenoids included in the biomass equation were estimated according to known ratios [96]. Concentrations of phycocyanin were estimated from the spectra of intact cells [97]. All measurements were taken in triplicate.

**Amino acids.** Total protein contents were measured using a Pierce BCA Assay kit. Amino acid proportions were determined according to published shotgun proteomics data for both Cyanothece 51142 and Synecocystis 6803 across a range of conditions [84] according to the following procedure: From peptide-level data, each mass spectral observation of a peptide was taken as an instance of a particular protein. The amino acid composition of each protein was taken from data in Cyanobase (http://genome.kazusa.or.jp/cyanobase) and thus the ‘proteome’ was taken to include all of the proteins whose peptides were observed in our data set, in proportion according to how often their peptides were observed. Amino acid frequencies were averaged across the proteome by a weighting factor of number of observations divided by the number of amino acids in the protein, similar to RPKM normalization for next-gen sequencing [98].

**Other cellular components.** The compositions of other cellular components of Synecocystis 6803 and Cyanothece 51142 were estimated based on values in the literature. DNA and RNA contents for Synecocystis 6803 were reported by Shastr and Morgan [27]. The remaining biomass components of Synecocystis 6803 (i.e., lipid, soluble pool and inorganic ions) were extracted from the measurements carried out by Nogueles et al. [32]. For Cyanothece 51142, biochemical compositions of macromolecules such as lipids, RNA, DNA and soluble pool were extracted from the measurements reported by Vu et al. [25].
Model Simulations

Flux balance analysis (FBA) [99] was employed in both the model validation and model testing phases. *Cyanothece* *iCyt773* and *Synechocystis* *iSyn731* models were evaluated in terms of biomass production under several scenarios: light and dark phases, heterotrophic and mixotrophic conditions. Flux distributions for each one of these states were inferred using FBA:

\[
\text{Maximize } v_{\text{biomass}}
\]

Subject to

\[
\sum_{j=1}^{m} S_{ij} v_j = 0 \quad \forall \ i \in 1, \ldots, n
\]  
(1)

\[
v_{j,\text{min}} \leq v_j \leq v_{j,\text{max}} \quad \forall \ j \in 1, \ldots, m
\]  
(2)

Here, \( S_{ij} \) is the stoichiometric coefficient of metabolite \( i \) in reaction \( j \) and \( v_j \) is the flux value of reaction \( j \). Parameters \( v_{j,\text{min}} \) and \( v_{j,\text{max}} \) denote the minimum and maximum allowable fluxes for reaction \( j \), respectively. Light and dark phases in *Cyanothece* *51142* are represented via modifying the minimum or maximum allowable fluxes with the following constraints, respectively:

\[
v_{\text{Glyceral}} = 0 \quad \text{and} \quad v_{\text{Glycer}} = 0
\]  
(3)

\[
v_{\text{CO}_2} = 0, \ v_{\text{Glyceral}} = 0, \ v_{\text{light}} = 0 \quad \text{and} \quad v_{\text{cf}} = 0
\]  
(4)

Here, \( v_{\text{biomass}} \) is the flux of biomass reaction and \( v_{\text{Glyceral}}, v_{\text{Glycer}} \) and \( v_{\text{CO}_2} \) are the fluxes of glyceral, glycogen and carbon dioxide transport reactions and \( v_{\text{light}} \) and \( v_{\text{cf}} \) are the fluxes of light reactions and carbon fixation reactions. For light phase, constraint (3) was included in the linear model, whereas for dark phase constraint (4) was included.

Once the *Synechocystis* *iSyn731* model was validated, it was further tested for *in silico* gene essentiality. The following constraint(s) was included individually in the linear model to represent any mutant:

\[
v_{\text{mutant}} = 0
\]  
(5)

Here, \( v_{\text{mutant}} \) represents flux of reaction(s) associated with any genetic mutation.

Flux variability analysis [100] for the reactions (for which photoautotrophic \(^{13}\)C MFA measurements [35] were available) was performed based on the following formulation:

Maximize/Minimize \( v_j \)

### Table 1. *Synechocystis* 6803 *iSyn731* and *Cyanothece* 51142 *iCyt773* model statistics.

<table>
<thead>
<tr>
<th></th>
<th>Synechocystis 6803 <em>iSyn731</em> model</th>
<th>Cyanothece 51142 <em>iCyt773</em> model</th>
</tr>
</thead>
<tbody>
<tr>
<td>Included genes</td>
<td>731</td>
<td>773</td>
</tr>
<tr>
<td>Proteins</td>
<td>511</td>
<td>465</td>
</tr>
<tr>
<td>Single functional proteins</td>
<td>348</td>
<td>336</td>
</tr>
<tr>
<td>Multifunctional proteins</td>
<td>91</td>
<td>83</td>
</tr>
<tr>
<td>Isozymes</td>
<td>4</td>
<td>1</td>
</tr>
<tr>
<td>Multimeric proteins</td>
<td>32</td>
<td>22</td>
</tr>
<tr>
<td>Others*</td>
<td>36</td>
<td>23</td>
</tr>
<tr>
<td>Metabolic reactions</td>
<td>1,156</td>
<td>946</td>
</tr>
<tr>
<td>Transport reactions</td>
<td>972</td>
<td>761</td>
</tr>
<tr>
<td>GPR associations</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Gene associated (metabolic/transport)</td>
<td>827</td>
<td>686</td>
</tr>
<tr>
<td>Spontaneous</td>
<td>180</td>
<td>158</td>
</tr>
<tr>
<td>Nongene associated (metabolic/transport)</td>
<td>59</td>
<td>16</td>
</tr>
<tr>
<td>No protein associated</td>
<td>90</td>
<td>86</td>
</tr>
<tr>
<td>Exchange reactions</td>
<td>57</td>
<td>57</td>
</tr>
<tr>
<td>Metabolite*</td>
<td>996</td>
<td>811</td>
</tr>
<tr>
<td>Cytosolic</td>
<td>862</td>
<td>675</td>
</tr>
<tr>
<td>Carboxisomial</td>
<td>8</td>
<td>8</td>
</tr>
<tr>
<td>Thylakoidic</td>
<td>10</td>
<td>9</td>
</tr>
<tr>
<td>Periplasmic</td>
<td>59</td>
<td>62</td>
</tr>
<tr>
<td>Extracellular</td>
<td>57</td>
<td>57</td>
</tr>
</tbody>
</table>
| aOthers include proteins involve in complex relationships, e.g. multiple proteins act as protein complex which is one of the isozymes for any specific reaction.
| bSpontaneous reactions are those without any enzyme as well as gene association.
| cMetabolites represent total number of metabolites with considering their compartmental specificity.

doi:10.1371/journal.pone.0048285.t001
Subject to

\[ \sum_{j=1}^{m} S_{ij} v_j = 0 \quad \forall \ i \in \{1, \ldots, n\} \]  

(6)

\[ v_{j,\text{min}} \leq v_j \leq v_{j,\text{max}} \quad \forall \ j \in \{1, \ldots, m\} \]  

(7)

\[ v_{\text{Biomass}} \geq v_{\text{min}} \]  

(8)

Here, \( v_{\text{Biomass}} \) is the minimum level of biomass production. In this case we fixed it to be the optimal value obtained under light condition for the \( \text{Synechocystis Syn731} \) model.

CPLEX solver (version 12.1, IBM ILOG) was used in the GAMS (version 23.3.3, GAMS Development Corporation) environment for implementing GapFind and GapFill [39] and solving the aforementioned optimization models. All computations were carried out on Intel Xeon E5450 Quad-Core 3.0 GH and Intel Xeon E5472 Quad-Core 3.0 GH processors that are part of the lionxj cluster (Intel Xeon E type processors and 96 GB memory) of High Performance Computing Group of The Pennsylvania State University.

Results and Discussion

Model Components

Biomass composition and diurnal cycle. The biomass equation approximates the dry biomass composition by draining all building blocks or precursor molecules in their physiologically relevant ratios. Most of the earlier genome-scale modeling efforts [12,29,30] of \( \text{Synechocystis 6803} \) contain approximate biomass equations completely or partially adopted from other species without direct measurements. This can adversely affect the accuracy of maximum biomass yield calculations, gene essentiality predictions and knockouts for overproduction.

Biomass composition for \( \text{Synechocystis Syn731} \) and \( \text{Cyanothece iCyt773} \) models were generated by defining all essential cellular biomass content values by experimental measurement or collection from existing literature (see ‘Materials and Methods’ for detail). Macromolecules present in both cyanobacteria such as protein, carbohydrates, lipids, DNA, RNA, pigments, soluble pool and inorganic ions were assigned to their corresponding metabolic precursors (e.g., L-glycine, glucose, 16C-lipid, ATP, dGTP, beta-carotene, coenzyme A and potassium respectively) (see File S3 for the complete list of biomass components). Based on the experimental measurements of precursor molecules needed to form a gram of the biomass, stoichiometric coefficients were assigned. For \( \text{Synechocystis 6803} \) we measured compositions of proteins and pigments and extracted compositions of the remaining biomass macromolecules from the model by Nagales et al. [32]. Thereby we developed biomass equations for three different conditions: photosautotrophic, mixotrophic and heterotrophic (see File S3). Experimental measurements (described in the Materials and Methods section and also in File S3) showed that biomass composition (i.e., mainly pigments) varies for \( \text{Cyanothece 51142} \) between light and dark conditions and nitrogen supplementation. Since pigments such as chlorophyll, carotenoids and phycoerythrobilin play important roles in photosynthetic processes their quantities are consequently higher under light conditions. In the presence of light \( \text{Cyanothece 51142} \) uses photosynthesis to store solar energy in the form of carbohydrates (i.e., glycogen), while in dark it expends that energy to fix nitrogen. Surprisingly, no significant change was measured in the carbohydrate pool between light and dark phases due to infinitesimal contribution of photosynthetically stored carbohydrates to total carbohydrate content in the biomass of \( \text{Cyanothece 51142} \). Aggregate quantities of the remaining biomass macromolecules for \( \text{Cyanothece 51142} \) such as lipids, RNA, DNA and soluble pool were extracted from the most recent \( \text{Cyanothece 51142} \) model by Vu et al. [25] to develop biomass equations for light and dark phases (File S3).

An earlier characterization study for \( \text{Cyanothece 51142} \) revealed that 113 proteins are expressed in higher abundance in the light phase while 137 are expressed in higher abundance in dark conditions [36]. The constructed model spans 26 light-specific proteins, associated with 36 reactions mainly involved in fatty acid, pigment, and amino acid metabolism and 11 dark-specific proteins accounting for 16 reactions from glycolysis, purine, pyrimidine, pyruvate, and amino acid metabolism (Files S4). Separate biomass equations as well as two regulatory structures for the model were derived in order to represent diurnal metabolic differences for \( \text{Cyanothece 51142} \) (File S4). In contrast, diurnal differences observed in \( \text{Synechocystis 6803} \) [37] are less pronounced (i.e., observed for only 54 genes) and less well functionally annotated (i.e., 32 genes with ‘unassigned’ functions). When compared to existing biomass equations of \( \text{Synechocystis 6803} \) [12,30] we found significantly lower values for the percent weight contribution of proteins towards the biomass pool (i.e., 52% for \( \text{Synechocystis 6803} \) and 53% for \( \text{Cyanothece 51142} \) vs. 84% [12] and 66% [30], respectively). The new protein biomass contribution is in better agreement with the previously reported value of 55% for \( \text{Cyanothece 51142} \) [38].

Identification and correction of network gaps. Upon ensuring biomass formation, GapFind [39] was applied to assess network connectivity and blocked metabolites. By applying GapFill [39] putative reconnection hypotheses were identified for blocked metabolites. Only the suggested modifications that were independently corroborated using literature sources and also did not lead to the introduction of thermodynamically infeasible cycle were included in the model. For \( \text{Synechocystis Syn731} \) model, GapFind [39] identified 207 blocked metabolites. Note that there exist 125 blocked metabolites in the \( \text{jN678} \) model [32], GapFill [39] identified unblocking hypotheses for 138 blocked metabolites. However, 88 of them led to the generation of infeasible thermodynamic cycles and thus were excluded. For only 5 blocked metabolites corroborating evidence for reconnection was obtained by adding 10 reactions (i.e., 2 metabolic, 4 transport and 4 exchange reactions). The added metabolic reactions have unknown gene associations (see File S1 for detailed information) while all 4 added transport reactions involve passive diffusion and thus are not associated with any specific gene(s) or protein(s). Ultimately, the 45 remaining blocked metabolites with GapFill suggested (but unconfirmed) reconnection mechanisms along with 69 blocked metabolites with no reconnection hypotheses were retained in the model \( \text{Syn731} \), while metabolites such as ubiquinone, which was proposed as an alternate substrate for succinate dehydrogenase [32] was excluded from \( \text{Syn731} \).

For the \( \text{Cyanothece iCyt773} \) model, 74 blocked metabolites were found after applying GapFind [39]. Note that there are 66 blocked metabolites in \( \text{iCce806} \) [25]. Two exchange reactions were added to allow the uptake of glucose and thiaminose ensuring biomass production under heterotrophic or mixotrophic conditions. Four blocked metabolites directly adopted from \( \text{iCce806} \) (during the draft model creation phase) were linked to five reactions with spurious gene associations and thus both metabolites and reactions were removed from \( \text{iCyt773} \). GapFill [39] suggested re-connection mechanisms for 52 blocked metabolites (out of a total of 70).
However, for 12 blocked metabolites the re-connection model modifications led to the creation of thermodynamically infeasible cycles and thus were discarded. Corroborating evidence for the reconnection of 30 blocked metabolites was identified through the addition of 19 GapFill suggested reactions (i.e., 8 metabolic, 7 transport and 4 exchange reactions). Of the eight added metabolic reactions we found direct literature evidence for five, homology-based evidence for one while two reactions are spontaneous (see File S2 for detailed information). All seven added transport reactions are through passive diffusion and thus are not connected with any specific gene(s) or protein(s). Ten remaining blocked metabolites with GapFill suggested reconnection hypotheses (along with 22 with no reconnection hypotheses) were left blocked in iCyt773 as no information to corroborate the GapFill suggested changes was found in the published literature and databases. For example, biotin is produced in Cyanothecae 51142; however, there is no literature evidence to support the presence of the initial step of the primary production pathway (i.e., conversion of pimeloyl-CoA from pimelate) and the intermediate step (i.e., biotransformation of 7,8-diamino-nonanoate from 8-amino-7-oxononanoate). This indicates that Cyanothecae 51142 may utilize a currently unknown pathway for producing biotin. The six other blocked metabolites are involved in the nonfermentative alcohol production pathway (as explained in model comparison section) known to be incomplete in Cyanothecae 51142. Table 2 summarizes the results related to connectivity restoration of Synechocystis iSyn731 and Cyanothecae iCyt773 models.

GPR associations and elemental and charge balancing. GPR associations connect genotype to phenotype by linking gene(s) that code for the protein(s) that catalyze a particular reaction. They are important to trace correctly as they provide the means to target at the gene level any change in the network desired at the reaction level. This is critical because genes may catalyze multiple reactions in multiple pathways. Many earlier models for Synechocystis 6803 do not provide in detail complex GPR associations, rather list only gene(s) and enzyme(s) involved in a specific reaction [12,30,31]. For both iCyt773 and iSyn731 models, we included comprehensive GPR associations (see Table 1 for detail information). All four intracellular compartments (i.e., periplasm, cytosol, thylakoid lumen and carboxysome) were assumed to have the same pH (7.2) and thus were assume charge balanced.

Under high light intensity in photoautotrophic conditions, Cyanothecae iCyt773 model produces 0.026 mole biomass/mole carbon fixed whereas Synechocystis iSyn731 yields 0.021 mole biomass/mole carbon fixed. These yields are almost identical to the ones calculated using the most recent models of Cyanothecae 51142 [25] and Synechocystis 6803 [32]. Experimental measurements of biomass yields are in the same order of magnitude with model predictions for the two organisms (i.e., 0.072 [25,40] and 0.082 mole biomass/mole carbon fixed [41]), respectively.

Comparison of iSyn731 Model Predicted Flux Ranges Against Experimental Measurements

We superimposed photoautotrophic flux measurements [33] for Synechocystis 6803 onto iSyn731 model to assess if the measurements are consistent with the model and whether the biomass maximization assumption correctly apportions fluxes to the metabolic network. For each reaction that was assigned a flux we calculated the flux-range under the maximum biomass assumption. Table 3 and Figure 1 summarize the obtained results for a basis of 100 millimole of CO2 plus H2CO3 uptake [33]. In seven (out of thirty one) cases the measured flux is fully contained within the model predicted ranges obtained upon maximizing biomass formation implying model consistency with MFA measurements. In contrast, under the maximum biomass assumption for thirteen fluxes the ranges underestimate and for four fluxes the ranges overestimate the experimentally deduced flux ranges while for seven fluxes the model derived flux ranges partially overlap with the experimental ones.

Perhaps the most informative discrepancy is for the CO2 fixing RuBisCO (RBC) reaction, which has a measured flux range of (123.00 to 132.00) vs. the model-calculated range of (102.49 to 106.33). In both cases the increased RBC flux (in comparison to the basis of 100 millimole of CO2 plus H2CO3 uptake) is needed to counteract the carbon loss due to the CO2 releasing reactions such as isocitrate dehydrogenase (ICD) and pyruvate dehydrogenase (PDH). We find that flux ranges, under the maximum biomass production assumption, of reactions such as glucose 6-phosphate dehydrogenase (G6PD), 6-phosphogluconolactonase (6PGL) and phosphogluconate dehydrogenase (6PGD) in oxidative pentose phosphate (OPP) pathway are negligible (0.00 to 0.03). In contrast, the experimentally derived range for OPP is (12 to 21). This is approximately equal to the difference between the model-predicted vs. experimentally deduced RBC reaction range implying the persistence of OPP flux even under the photoautotrophic condition [33] despite the presence of a more efficient NADPH production route through photosynthesis as predicted by the model (under max biomass). The high values Young et al. [35] obtained for the OPP fluxes were surprising as OPP is not a very efficient route for cyanobacteria to generate reducing power. This may reflect some inherent biological constraint that is not captured by the optimality assumption.

Model predicted lower flux ranges for RBC are propagated to seven other reactions in the Calvin cycle (i.e., phosphoglycerate kinase (PGK), glyceraldehyde-3-phosphate dehydrogenase (13PDG), triose-phosphate isomerase (TPI), transketolase (TKT1), ribose-5-phosphate isomerase (RPI), ribulose 5-phosphate 3-epimerase (RPE) and phosphoribulokinase (PRK). The remain-

<table>
<thead>
<tr>
<th>Table 2. Summary of connectivity restoration in Synechocystis 6803 iSyn731 and Cyanothecae 51142 iCyt773 models.</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
</tr>
<tr>
<td>Number of blocked metabolites</td>
</tr>
<tr>
<td>Number of metabolites with GapFill [39] suggested reconnection strategies</td>
</tr>
<tr>
<td>Number of metabolites whose reconnection forms a cycle</td>
</tr>
<tr>
<td>Number of metabolites with validated reconnection mechanisms</td>
</tr>
<tr>
<td>Number of added reactions to the model</td>
</tr>
</tbody>
</table>

doi:10.1371/journal.pone.0048285.t002

PLOS ONE | www.plosone.org 5 October 2012 | Volume 7 | Issue 10 | e48285
ing six reaction fluxes with lower model predicted fluxes compared to measurements [35] are all in the TCA cycle (i.e., citrate synthase (CS), aconitate (ACONT), isocitrate dehydrogenase (ICD), succinate dehydrogenase (SUCD) and malic enzyme (ME1 and ME2) reactions). Even under the max biomass assumption, SUCD is not required to carry any flux due to the presence of other succinate dehydrogenases (as part of respiratory chain) in the iSyn731 model. Furthermore, in contrast with experimental observations, under the maximum biomass assumption, the model predicts no flux through the malic enzyme (ME) reactions presumably because it is a less energy-efficient route (i.e., phosphoenolpyruvate → oxaloacetate → malate → pyruvate) for pyruvate generation than the pyruvate kinase (PYK) reaction [35].

There are nine reactions with experimentally derived ranges completely subsumed within the ones derived under the maximum biomass assumption. Five of them are in the Calvin cycle (i.e., fructose-bisphosphate aldolase (FBA), fructose-bisphosphatase (FBP), Sedoheptulose 1,7-bisphosphate D-gluceraldehyde-3-phosphate-lyase (SBGPL), sedoheptulose-bisphosphatase (SBP) and bidirectional transaldolase (TAL)). The first four reactions are essential with experimentally deduced flux ranges of (53.00 to 66.00) for FBA and FBP and (29.00 to 43.00) for SBGPL and SBP. In contrast, the calculated flux ranges (−0.08 to 73.17) for FBA and SBGPL and (0.00 to 73.17) for FBP and SBP imply that they are in silico non-essential. As depicted in Figure 1, these reactions are involved in the production of sedoheptulose 7-phosphate (S7P) from fructose 1,6-bisphosphate (FDP). An alternative production route for S7P is afforded in the model through the bidirectional transaldolase (TAL) reaction from fructose 6-phosphate (F6P) alluding to an explanation for the wider flux ranges derived using the model. Experimental and model predicted flux ranges for TAL are (−6.00 to 9.00) and (−35.93 and 37.32), respectively. Upon restricting the TAL flux ranges in the calculations to the ones found experimentally, the flux variability analysis shrinks the flux ranges for FBA and FBP to (28.22 to 43.27) and (28.22 to 43.33) and for SBGPL and SBP to (29.82 to 44.87) and (29.82 to 44.87), respectively which are very close to the experimentally measured ranges. This is indicative that in addition to the maximization of biomass formation, additional restrictions (e.g., photosynthetic efficiency and relative selectivity of RuBISCO for carboxylation...
(over oxidation) limit the range of fluxes that the aforementioned glycolytic fluxes may span in vivo. Note that the presence of experimentally measured fluxes is important to test the model and the adopted maximization principle. We were fortunate in this case to have access to such data as for most organisms they are absent.

Phosphoglycerate mutase (PGM) and enolase (ENO) reactions have very similar model derived and experimentally obtained flux ranges. Model-predicted flux values of the remaining two reactions, pyruvate kinase (PYK) and pyruvate dehydrogenase (PDH), could reach as low as zero due to the metabolic flexibility that the iSyn731 model possesses by having alternate enzymes with different cofactor specificities. The max biomass flux range of fumarase (FUM) is found to be (2.54 to 1.25), compared to the experimentally measured (1.70 to 2.00). Therefore, it appears that under the phototrophic condition, the forward direction is kinetically favorable. By restricting the reaction to be irreversible the model predicted a FUM flux range of (0.00 to 1.49) which is close to the experimentally derived one (see Figure 1). However, contrary to MFA measurements these reactions (FUM and ME) are dispensable for in silico biomass production.

### Table 3. Comparison of 13C MFA flux measurements [35] vs. model-predicted flux ranges.

<table>
<thead>
<tr>
<th>Reaction</th>
<th>Flux measurements by Young et al., 2011 [35]</th>
<th>Flux ranges predicted by iJN678 model (With max biomass)</th>
<th>Flux ranges predicted by iSyn731 model (With max biomass)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>95% LB</td>
<td>95% UB</td>
<td>LB</td>
</tr>
<tr>
<td>RBC</td>
<td>123.00</td>
<td>132.00</td>
<td>109.02</td>
</tr>
<tr>
<td>PGK</td>
<td>219.00</td>
<td>237.00</td>
<td>187.11</td>
</tr>
<tr>
<td>13PGD</td>
<td>219.00</td>
<td>237.00</td>
<td>187.11</td>
</tr>
<tr>
<td>GAPDH</td>
<td>90.00</td>
<td>99.00</td>
<td>74.98</td>
</tr>
<tr>
<td>FBA</td>
<td>53.00</td>
<td>66.00</td>
<td>−0.17</td>
</tr>
<tr>
<td>FBP</td>
<td>53.00</td>
<td>66.00</td>
<td>0.00</td>
</tr>
<tr>
<td>PGI</td>
<td>15.00</td>
<td>24.00</td>
<td>0.68</td>
</tr>
<tr>
<td>G6PD</td>
<td>12.00</td>
<td>21.00</td>
<td>0.00</td>
</tr>
<tr>
<td>6PGL</td>
<td>12.00</td>
<td>21.00</td>
<td>0.00</td>
</tr>
<tr>
<td>6PGD</td>
<td>12.00</td>
<td>21.00</td>
<td>0.00</td>
</tr>
<tr>
<td>PRK</td>
<td>123.00</td>
<td>132.00</td>
<td>109.02</td>
</tr>
<tr>
<td>SBGPL</td>
<td>29.00</td>
<td>43.00</td>
<td>−0.17</td>
</tr>
<tr>
<td>SBP</td>
<td>29.00</td>
<td>43.00</td>
<td>0.00</td>
</tr>
<tr>
<td>TAL</td>
<td>−6.00</td>
<td>9.00</td>
<td>−36.74</td>
</tr>
<tr>
<td>TKT1</td>
<td>37.20</td>
<td>37.50</td>
<td>36.57</td>
</tr>
<tr>
<td>RPI</td>
<td>35.40</td>
<td>35.70</td>
<td>35.18</td>
</tr>
<tr>
<td>TKT2</td>
<td>35.40</td>
<td>35.70</td>
<td>37.25</td>
</tr>
<tr>
<td>RPE</td>
<td>75.50</td>
<td>76.20</td>
<td>73.83</td>
</tr>
<tr>
<td>PGM</td>
<td>22.90</td>
<td>23.60</td>
<td>26.83</td>
</tr>
<tr>
<td>ENO</td>
<td>23.40</td>
<td>23.80</td>
<td>26.84</td>
</tr>
<tr>
<td>PYK</td>
<td>7.90</td>
<td>11.10</td>
<td>0.00</td>
</tr>
<tr>
<td>PDH</td>
<td>11.50</td>
<td>12.00</td>
<td>0.00</td>
</tr>
<tr>
<td>CS</td>
<td>3.00</td>
<td>3.40</td>
<td>2.15</td>
</tr>
<tr>
<td>ACCONT</td>
<td>3.00</td>
<td>3.40</td>
<td>2.15</td>
</tr>
<tr>
<td>ICD</td>
<td>3.00</td>
<td>3.00</td>
<td>2.15</td>
</tr>
<tr>
<td>SUCD</td>
<td>0.00</td>
<td>0.40</td>
<td>0.00</td>
</tr>
<tr>
<td>FUM</td>
<td>1.70</td>
<td>2.00</td>
<td>−5.44</td>
</tr>
<tr>
<td>MDH</td>
<td>1.90</td>
<td>5.20</td>
<td>5.35</td>
</tr>
<tr>
<td>ME1</td>
<td>3.70</td>
<td>6.90</td>
<td>0.00</td>
</tr>
<tr>
<td>ME2</td>
<td>3.70</td>
<td>6.90</td>
<td>−0.00</td>
</tr>
<tr>
<td>PPC</td>
<td>9.90</td>
<td>13.20</td>
<td>11.74</td>
</tr>
</tbody>
</table>

doi:10.1371/journal.pone.0048285.t003

iSyn731 Model Testing Using in vivo Gene Essentiality Data

The quality of model iSyn731 for Synechocystis 6803 was tested using experimental data on the viability (or lack thereof) of single gene knockouts. We used the CyanobMutants database [34,42] that includes in vivo gene essentiality data for 119 genes (i.e., 19 essential and 100 nonessential) with metabolic functions in iSyn731 model. Cases that were flagged with incomplete segregation in the database were omitted in iSyn731 model comparisons. We examined the feasibility of biomass production for the model iSyn731 by comparing the maximum biomass formation upon imposing the gene knockout with the maximum theoretical yield of the wild-type organism. A threshold of 10% of
the maximum theoretical yield was used as a cutoff [43]. Comparisons between in vivo and in silico results led to four possible outcomes, as previously delineated by Kumar et al., GG, GNG, NGG and NGNG [43]. Initially, the model correctly predicted 18 out of 19 essential genes (i.e., 10 NGNG and 1 GG) and 74 out of 100 non-essential genes (i.e., 73 GG and 27 NGG). We next explored the causes of these discrepancies and attempted to mitigate them whenever possible.

The single GNG case corresponds to mutant ΔchlA2 exhibiting no growth under aerobic conditions [44]. The ChlA system is a Mg-protoporphyrin IX monomethyl ester (MPE) cycle system that is responsible for forming the isocyclic ring (E-ring) in chlorophylls under aerobic conditions [44]. The model allowed for the BchE and ChlA2 systems (alternate cyclase systems) to complement the loss of the ChlA1 system leading to an in silico viable mutant. However, the same literature source [44] suggested that both BchE and ChlA2 systems are unlikely to be active under aerobic conditions and thus rescue mutant ΔchlA2. This prompted the introduction of a regulatory restriction in iSyn731 model where only ChlA1 reactions were active under aerobic conditions as MPE while ChlA2 and BchE system reactions were deactivated. Using these regulatory restrictions resolves the single GNG inconsistency.

Twenty (out of 27) NGG cases were associated with Photosystem I (PSI), Photosystem II (PSII) and other photosynthesis reactions. While reconstructing the model, we assumed that all genes involved in photosynthetic reaction system were essential to the functioning of the overall system. Published literature [45–49] suggests that genes involved in photosynthetic reactions form complex interdependencies. We used NCBI COBALT multiple sequence alignment tool [50] to construct a phylogenetic tree of orthologous genes associated with each photosystem along with BLASTp searches to identify putative complementation relationships between genes to explain the inconsistencies between the predicted in silico and in vivo growth. Genes deemed homologous (i.e., lie adjacent in the phylogenetic tree) were linked with “OR” GPR relations implying that the loss of one gene can be complemented by the other. Seven out of twenty NGG cases (i.e., psaD, psaI and pshA2 for PSI and PSII and cpcC2, cpcC1, cpcD, and cpcD for other photosynthesis reactions) were resolved by modifying the corresponding GPR relations using an OR relation [46,47,51–54]. However, no phylogenetically adjacent or related (or homologous) genes were found for the remaining 13 NGG cases (pseA, pseD2, pseO, pseU, pseV, pse26, pseX, pse27, petE, cpaA, cpaB, cpaE, cpaF) [48,49,51,54–58]. For these cases, the genes were deemed nonessential to the functioning of the reactions in question (i.e., photosynthesis reactions) and thereby the corresponding GPRs were modified to show an OR relation between each of these genes and an ‘unknown gene’, similar to what was previously performed in the refinement of the ∆MM904 model [59] (see File S5 for detailed information).

The remaining seven NGG cases are associated with a variety of metabolic functions. One such case is the ∆modBC mutant corresponding to the sole ABC molybdate transporter in the model. Literature evidence [60] revealed that a related cyanobacterium, Anabaena variabilis ATCC 29413, could continue to grow despite the loss of its molybdate ABC transporter due to the presence of another low affinity molybdate transporter or an inducible sulfate transport system that can serve as a low affinity molybdate transporter when required. We found the same gene coding for the sulfate transporter in A. variabilis (cysI) in the iSyn731 model allowing the resolution of the discrepancy by adding a cysI-linked alternate molybdate transporter. Another NGG case is mutant ΔcrtO that cannot produce echeinone (a biomass component) in iSyn731 with no effect on observed growth. Therefore, it appears that iSyn731 cannot capture the flexibility of Synechocystis 6803 metabolism [61] when echeinone production is restricted. The remaining five NGG cases are spread across many metabolic pathways. The Δmod mutant eliminates the copper ABC transporter without affecting growth, which alludes to the existence of another unknown mode of copper uptake not present in iSyn731 [62,63]. The ΔcrtG mutant eliminates a reaction for the production of phylloquinone while mutant Δpdf affects the production of homogentisate, a precursor for both tocopherols and plastoquinone. Finally, the Δocr3 mutant affects the production of both plastoquinone and tocopherol [64] and the viable Δocr4 mutant restricts the production of chorismate (a precursor to aromatic amino acids) and also restricts carboxysome formation [65–67]. These six inconsistencies between the model predictions and growth data imply that the cyanobacterium can co-opt another metabolic process to (partially) complement for the gene loss. Unlike the case of the ∆modBC mutant, we have found no plausible mechanism for the six remaining mutants.

After resolving the discrepancies, as described above, iSyn731 correctly predicted all 19 essential genes (i.e., 19 NGNG and 0 GNG) and 94 (out of 100) non-essential genes (i.e., 94 GG and 6 NGG). Figure 2 shows our results and comparisons against two other available Synechocystis 6803 models by Knoop et al. [12] and Nogales et al. [32]. We used the Cyanobacteria Mutant database [34] to identify 114 genes (i.e., 19 essential and 95 nonessential) having metabolic functions in the jN678 model by Nogales et al. [32]. Out of 114 genes the jN678 model correctly predicted 18 essential genes (i.e., 18 NGNG and 1 GNG) and 69 non-essential genes (i.e., 69 GG and 26 NGG). The model by Knoop et al. [12] was tested for 51 mutants but we found that only 43 (i.e., 7 essential and 36 non-essential) of them were reported to have complete segregation [34]. Of these 43, Knoop et al.’s [12] model correctly predicted 5 essential genes (i.e., 5 NGNG and 2 GNG) and 32 nonessential genes (i.e., 32 GG and 4 NGG). The specificity and sensitivity of each of these three models were also calculated and displayed at the bottom of Figure 2.

All 114 genes tested for jN678 were also present in the iSyn731 model. 26 GG and one GNG cases present in jN678 model correspond to NGG and GNG cases that were either fixed or still present in iSyn731 as discussed before. Lethal mutant Δppa is correctly predicted as NGNG in iSyn731 but deemed GNG in Knoop et al. [12] model. This was because ppa in iSyn731 codes for the degradation of both triphosphate into diphosphate and diphosphate to phosphate. Only the latter activity is linked to ppa in the Knoop et al’s model. Out of 4 NGG cases in [12], two involve ΔcymA and ΔcymB mutants. Both these genes are involved in the ABC transporter system for bicarbonate from periplasm to cytosol. iSyn731 avoids this inconsistency as it contains an alternate sodium and bicarbonate co-transport system.

Model Comparisons

Synechocystis 6803 model comparisons. The iSyn731 model integrates the description in the photosystems of the model presented by Nogales et al. [32] and adds additional detail. One notable difference is that iSyn731 uses a separate photon for each reaction center (i.e., PSI and PSII) as they are optimized for different ranges of wavelength [68], whereas jN678 [32] uses a single photon shared by both photosystem reactions. As many as 322 new reactions (see Figure 3A), are added in iSyn731 distributed across many pathways. Most of the additions are in the lipid and fatty acid metabolism to support the synthesis of measured fatty acids and lipids present in the biomass equation.
This list includes myristic acid (14-carbon saturated fatty acid) and lauric acid (12-carbon saturated fatty acid). iJN678 [32] contained four reactions exhibiting unbounded flux (i.e., two duplicate glycine cleavage reactions and two duplicate leucine transaminase reactions). They form a thermodynamically infeasible cycle (see Figure 4A for leucine transaminase reactions) that was resolved in iSyn731 by eliminating redundant functions. In addition, the glycine cleavage system was recast in detail by abstracting the separate action of the four enzymes (named the T-, P-, L-, and H-proteins) that ultimately catalyze the demethylamination of glycine.

iSyn731 improves upon iJN678 [32] by eliminating lumped reactions whenever a multi-step description is available and expands the range of functions carried out with alternate cofactors. As many as twelve reactions with an enoyl-[acyl-carrier-protein] reductase function were linked with not only NADP but also with the more rare NAD cofactor specificity. Another important difference between iSyn731 and iJN678 [32] is the cellular...
location of the CO₂ fixation (i.e., ribulose-1,5-bisphosphate carboxylase/oxygenase (RuBisCO) enzyme). Literature [69,70] shows that cyanobacteria possess a micro-compartment (i.e., carboxysome) encapsulating RuBisCO and carbonic anhydrase (CA) enzymes. Syn731 adds carboxysome as a cellular compartment and also all necessary transport reactions [69,70]. Recently, Zhang and Bryant [33] hypothesized the existence of a functional TCA cycle in most cyanobacterial species using a 2-ketoglutarate to succinate bypassing step. Syn731 allows for a complete TCA cycle using the bypassing step. In addition, Syn731 contains an intact heptadecane biosynthesis pathway as recently described [21] unlike earlier Synechocystis 6803 models [12,29,30,32] (see Figure 5A for distribution of unique reactions in Syn731).

**Cyanothecce 51142 model comparisons.** The iCyt773 model for Cyanothecce 51142 improves upon the iCce806 model [25]. iCyt773 segregates reactions into the periplasm, thylakoid lumen, carboxysome, and cytoplasm compartments thus introducing an additional 60 transport reactions compared to iCce806 [25]. Unlike iCce806 [25], iCyt773 does not track macromolecule synthesis for DNA, RNA, and proteins to maintain consistency with the Synechocystis 6803 model. This difference accounts for 69 genes present in iCce806 [25] but absent from iCyt773. iCce806 [25] contained 15 reactions which formed five cycles that could carry unbounded metabolic flux (i.e., thermodynamically infeasible cycles). All these cycles were eliminated by restricting...
A

\[
\begin{align*}
\text{α-Ketoglutarate} & \quad \text{LEUTA} \quad \text{L-Glutamate} \\
\text{L-Leucine} & \quad \text{LEUTA} \quad 4\text{-Methyl-2-oxopentanoate} \\
\text{α-Ketoglutarate} & \quad \text{L-Leucine} \quad \text{L-Glutamate}
\end{align*}
\]

B

\[
\begin{align*}
\text{Ferredoxin (reduced)} & \quad \text{NAD} \quad \text{NADPH} \\
\text{L-Alanine} & \quad \text{Pyruvate} \\
\text{Glyoxalate} & \quad \text{Glycine} \\
\text{L-Serine} & \quad \text{Hydroxyproline} \\
\text{Pyruvate} & \quad \text{L-Alanine}
\end{align*}
\]

\[
\begin{align*}
\text{Ferredoxin (oxidized)} & \quad \text{NAD} \quad \text{NADPH} \\
\text{L-Alanine} & \quad \text{Pyruvate} \\
\text{Glyoxalate} & \quad \text{Glycine} \\
\text{L-Serine} & \quad \text{Hydroxyproline} \\
\text{Pyruvate} & \quad \text{L-Alanine}
\end{align*}
\]

\[
\begin{align*}
\text{13DGP} & \quad \text{G3P} \\
\text{GMP} & \quad \text{GDP} \\
\text{ATP} & \quad \text{ADP} \\
\text{dADP} & \quad \text{dTATP} \\
\text{GDP} & \quad \text{GMP}
\end{align*}
\]

\[
\begin{align*}
\text{Glycerate} & \quad \text{Tartarate semialdehyde} \\
\text{Glycerate} & \quad \text{Tartarate semialdehyde}
\end{align*}
\]
reaction directionality and eliminating reactions that were linear combinations of others (coded by the same gene) (see Figure 4B).

**iCy773** contains 43 unique genes and 266 unique reactions (including transport and alternate cofactor utilizing reactions) as shown in Figure 3B. Figure 3B depicts the distribution of the new reactions across different pathways. Most of the additions are found in lipid and pigment biosynthesis pathways. The **iCy773** model captures in detail the lipid biosynthesis pathway composed of 73 reactions and links as many as 28 biomass precursor lipids (e.g., sulfoquinovosyldiacylglycerols, monogalactosyldiacylglycerols, digalactosyldiacylglycerols, and phosphatidylglycerols) directly to the biomass equation. The porphyrin and chlorophyll metabolism and carotenoid biosynthesis pathways were updated to include 24 reactions for the production of accessory pigments such as echinenone, an accessory pigment, and (3Z)-phycocyanobilin, a phycoobilin. Accessory pigments donate electrons to chlorophyll rather than directly to photosynthesis. Phycobilins are adapted for many wavelengths not absorbed by chlorophyll thus broadening the spectrum useful for photosynthesis. The variety of pigments in cyanobacteria is well documented [71–73] providing so far untapped avenues for engineering increased efficiency in photosynthesis and control of electron transfer processes in biological systems. Another new function in **iCy773** is L-Aspartate Oxidase. L-Aspartate Oxidase allows the deamination of aspartate, forming α-ketoglutarate and ammonia. The L-Aspartate Oxidase allows the deamination of aspartate, forming α-ketoglutarate and ammonia. The ammonia is then used by other reactions to produce other metabolites.

**iCy773** is L-Aspartate Oxidase. L-Aspartate Oxidase allows the deamination of aspartate, forming α-ketoglutarate and ammonia. The remaining 218 reactions describe secondary metabolism, fatty acid and lipid biosynthesis, amino acid biosynthesis, and biochemical characterization. As many as 670 reactions and 596 metabolites are shared by both models corresponding to 47% and 63% of the total reactome and metabolome, respectively (see Figure 3C). The higher degree of conservation of metabolites (as opposed to reactions) across the two cyanobacteria suggests that lifestyle adaptations tend to usher new enzymatic activities that most of the time make use of the same metabolic pool without introducing new metabolites. There are 486 reactions that are unique to **iSyn731** with no counterpart in **iCy773**. These reactions are not preferentially allotted to a handful of specific pathways. Instead they are spread over tens of different pathways. Primary metabolism reactions dispersed throughout fatty acid biosynthesis, lipid metabolism, oxidative phosphorylation, purine and pyrimidine metabolism, transport and exchange reactions account for 295 reactions. Secondary metabolism including chlorophyll and cyanochromycin metabolism, folate, terpenoid, phenylpropanoid and flavonoid biosynthesis accounts for the remaining 191 **iSyn731**-specific reactions. Interestingly, the 276 **iCy773**-specific reactions span the same set of diverse pathways implying that the two organisms have adopted unique/divergent biosynthetic pathways for similar metabolic needs. Fifty-eight span primary metabolism pathways such as purine and pyrimidine metabolism, fatty acid and lipid biosynthesis, amino acid biosynthesis. The remaining 218 reactions describe secondary metabolism such as terpenoid biosynthesis, chloropyll and cyanochromycin biosynthesis, plastoquinone and phyloquinone biosynthesis (see File S6 for detail information). The much larger set of unique **iSyn731**-specific reactions compared to **iCy773** reflect more complete genome annotation and biochemical characterization rather than augmented metabolic versatility.

A number of distinct differences in metabolism between the two organisms have been accounted for in the two models. For example, **iCy773** does not have the enzyme threonine ammonia-lyase, which catalyzes the conversion of threonine to 2-ketobutyrate and as a consequence lacks the traditional route for isoleucine synthesis. Instead it employs part of the alternative citramalate pathway for isoleucine synthesis with pyruvate and acetyl-CoA as precursors. Follow up literature queries revealed the existence of this alternative pathway in *Cyanothece* 51142 [22]. Ketobutyrate, an intermediate in the citramalate pathway, can be readily converted to higher alcohols, such as propanol and butanol, via a non-fermentative alcohol production pathway. Using the **iCy773** model, we determined that only 2-ketoacid decarboxylase is missing from these three-step processes. In contrast, **iSyn731** was found to have only the traditional route for isoleucine production with the citramalate pathway completely absent (see Figure 6A). In another example, the fermentative 1-butanol pathway is known to be incomplete in both organisms. By querying the developed models we can pinpoint exactly which steps are absent. Specifically, the conversion between 3-hydroxybutyryl-CoA and butanal is missing in both models. In addition to higher alcohols, higher alkanes (G13 and above) are important biofuel molecules as the main constituents of diesel and jet fuel [21].

**Figure 4. Schematics that illustrate the thermodynamically infeasible cycles and subsequent resolution strategies.** (A) Cycles present in JN678 [32], and (B) Cycles present in *Cce805* [25]. Blue colored lines represent the original reaction directionality whereas green ones denote modified directionality to eliminate cycle.

doi:10.1371/journal.pone.0048285.g004

**iSyn731 and iCy773 models comparison.** Figure 3C illustrates the total number of common and unique reactions and metabolites between **iSyn731** and **iCy773** models. The *Cyanothece* 51142 genome [9,77] is 1.5 times larger than the one for *Synechocystis* 6803 [11], nevertheless **iCy773** is smaller than **iSyn731** due to differences in the level of detail of annotation and biochemical characterization. As many as 670 reactions and 596 metabolites are shared by both models corresponding to 47% and 63% of the total reactome and metabolome, respectively (see Figure 3C). The higher degree of conservation of metabolites (as opposed to reactions) across the two cyanobacteria suggests that lifestyle adaptations tend to usher new enzymatic activities that most of the time make use of the same metabolic pool without introducing new metabolites. There are 486 reactions that are unique to **iSyn731** with no counterpart in **iCy773**. These reactions are not preferentially allotted to a handful of specific pathways. Instead they are spread over tens of different pathways. Primary metabolism reactions dispersed throughout fatty acid biosynthesis, lipid metabolism, oxidative phosphorylation, purine and pyrimidine metabolism, transport and exchange reactions account for 295 reactions. Secondary metabolism including chlorophyll and cyanochromycin metabolism, folate, terpenoid, phenylpropanoid and flavonoid biosynthesis accounts for the remaining 191 **iSyn731**-specific reactions. Interestingly, the 276 **iCy773**-specific reactions span the same set of diverse pathways implying that the two organisms have adopted unique/divergent biosynthetic pathways for similar metabolic needs. Fifty-eight span primary metabolism pathways such as purine and pyrimidine metabolism, fatty acid and lipid biosynthesis, amino acid biosynthesis. The remaining 218 reactions describe secondary metabolism such as terpenoid biosynthesis, chlorophyll and cyanochromycin biosynthesis, plastoquinone and phyloquinone biosynthesis (see File S6 for detail information). The much larger set of unique **iSyn731**-specific reactions compared to **iCy773** reflect more complete genome annotation and biochemical characterization rather than augmented metabolic versatility.
Figure 5. List of added reactions across pathways. (A) iSyn731 compared to iJN678 [32], and (B) iCyt773 compared to iCce806 [25]. doi:10.1371/journal.pone.0048285.g005
Recently reported [21] novel genes involved in the biosynthesis of alkanes in several cyanobacterial strains were incorporated in the models. Metabolic differences in *Cyanothece* 51142 and *Synechocystis* 6803 lead to the production of different alkanes (e.g., pentadecane in *Cyanothece* 51142 and heptadecane in *Synechocystis* 6803) (see Figure 6B).

Model *iCyt773*, in contrast to *iSyn731*, does not have a complete urea cycle as it lacks the enzyme L-arginine aminohydrolase catalyzing the production of urea from L-arginine. Literature sources [77,78] support this finding and explain the absence of a functional urea cycle as a consequence of the nitrogen-fixation ability of *Cyanothece* 51142 [79,80]. Because *Cyanothece* 51142 can fix nitrogen directly from the atmosphere and produce ammonium via the enzyme nitrogenase, genes corresponding to the activity of L-arginine aminohydrolase and urease (for breaking down urea) become redundant, explaining why they are not present in its genome [80]. In addition to nitrogen metabolism, *iCyt773* and *iSyn731* models reveal marked differences in anaerobic metabolic capabilities. Unlike *iSyn731*, *iCyt773* includes a L-lactate dehydrogenase activity that enables the complete fermentative lactate production pathway. On the other hand, *iSyn731* contains the anaerobic chlorophyll biosynthetic pathway using enzyme protoporphyrin IX cyclase (BchE) that is absent in *iCyt773*. Other differences in metabolism include lipid and fatty acid synthesis, fructose-6-phosphate shunt and nitrogen fixation. Model *iSyn731* traces the location of the double bond for unsaturated fatty acid synthesis pathways, as two separate isomers of unsaturated C18 fatty acids are part of the biomass description. *iCyt773* allows for the shunting of fructose-6-phosphate into erythrose-4-phosphate along with acetate and ATP using the fructose-6-phosphate phosphoketolase activity. Finally, both *iSyn731* and *iCyt773* contain multiple hydrogenases allowing both to produce hydrogen. However, only the latter has a nitrogenase activity that can fix nitrogen while simultaneously producing hydrogen.

Using *iSyn731* and *iCyt773* to Estimate Production Yields

We tested the recently developed models *iSyn731* and *iCyt773* by comparing the predicted maximum theoretical product yields with experimentally measured values for two very different metabolic products: isoprene and hydrogen. Isoprene, a volatile hydrocarbon and potential feedstock for biofuel, is mostly produced in plants under heat stress [13]. Cyanobacteria offer promising production alternatives as they can grow to high densities in bioreactors and produce isoprene directly from...
photosynthesis intermediates [13]. It was reported [13] that Synechocystis 6803 has all but one gene (encoding isoprene synthase) in the methyl-erythritol-4-phosphate (MEP) pathway for isoprene synthase from dimethylallyl phosphate (DMAPP). Upon cloning the isoprene synthase from kudzu vine (Pueraria montana) into Synechocystis 6803 isoprene production was demonstrated using sunlight and atmospheric CO$_2$ of 4.3$x10^{-3}$ mole isoprene/mole carbon fixed [81]. We calculated the maximum isoprene yield using Syn731 to be 3.63$x10^{-3}$ mole isoprene/mole carbon fixed upon adding the isoprene synthase activity to the model and simulating the conditions described in [41] under maximum biomass production. Similar isoprene yields were obtained with JN678 [32] while earlier models of Synechocystis 6803 [12,29-31] lack the MEP pathway (partially or completely) and thus do not support isoprene production. The underestimation of the experimentally observed isoprene yield by the model predicted maximum yield may be due to sub-optimal growth of the production strain, differences in the list of measured biomass components, missing isoprene-relevant reactions from the model or more likely a combination of the above factors.

Both Cyanothece 51142 and Synechocystis 6803 produce hydrogen by utilizing nitrogenase and hydrogenase activities, respectively [19]. Under subjective dark conditions [19] whereby (i) stored glycogen acts as a carbon source, (ii) photosynthesis harnesses light energy, and (iii) nitrogenase activity is not restricted, hydrogen production yield for Cyanothece 51142 was measured at 49.67 mole/mole glycogen consumed. Simulating the same conditions in iCyt773 and iCce806 [25] leads to maximum theoretical yields for hydrogen production of 48.43 mole/mole glycogen and 102.4 mole/mole glycogen, respectively. The entire amount of hydrogen produced in iCyt773 is due to the nitrogenase activity. In contrast, the predicted doubling of the maximum hydrogen yield in iCce806 is due to the utilization of the reverse direction of two hydrogen dehydrogenase reactions without any nitrogenase activity. Utilization of the nitrogenase reaction requires the use and recycling of more ATP than simply running the dehydrogenase reactions in reverse. However, it has been reported that hydrogen production in Cyanothece 51142 is primarily mediated by the nitrogenase enzyme [19] in the dark phase. This lends support to the reversibility of the dehydrogenase reactions (under dark condition) as present in the iCyt773 model. Experimental results for Synechocystis 6803 support up to 4.24 mole/mole glycogen consumed [19,82] of hydrogen production. iSyn731 predicts a maximum hydrogen theoretical yield of 2.28 mole/mole glycogen consumed while JN678 [32] yields a value of 2.00 mole/mole glycogen consumed. Again the factors outlined for isoprene production may explain the lower theoretical yields predicted by the two models. The small difference between the model predicted yields is due to the presence of one step lumped biotransformation between isocitrate and oxoglutarate via isocitrate dehydrogenase in JN678 [32]. iSyn731 describes this biotransformation in two steps (isocitrate $\rightarrow$ oxaloacetate $\rightarrow$ oxoglutarate) [83] generating an additional NADPH and subsequently more hydrogen via the hydrogenase reaction.

**Conclusion**

In this paper, we expanded upon existing models to develop two genome-scale metabolic models (Synechocystis Syn731 and Cyanothece iCyt773) for cyanobacterial metabolism by integrating all available knowledge available from public databases and published literature. All metabolite and reaction naming conventions are consistent between the two models allowing for direct comparisons. Systematic gap filling analyses led to the bridging of a number of network gaps in the two models and the elimination of orphan metabolites. Two separate biomass equations as well as two different versions of Cyanothece iCyt773 models were developed for light and dark phases to represent diurnal regulation. The development of two separate models for Cyanothece 51142 (i.e., light and dark) provides the two “end-points” for the future development of dynamic metabolic models capturing the temporal evolution [36,84-86] of fluxes during the transition phases DFBA [87]. Comparisons against available $^{13}$C MFA measurements for Synechocystis 6803 [35] revealed that the Syn731 model upon biomass maximization yields flux ranges that are generally consistent with experimental data. Discrepancies between the two identify metabolic nodes where regulatory constraints are needed in addition to biomass maximization to recapitulate physiological behavior. The ability of Syn731 to predict the fate of single gene knock-outs was further improved (specificity of 0.94 and sensitivity of 1.00) by reconciling in vivo growth predictions with in vivo gene essentiality data [34]. Similar analyses could also be carried out for Cyanothece iCyt773 model once such flux measurements and in vivo gene essentiality data become available.

It is becoming widely accepted that focusing on a single pathway at a time without quantitatively assessing the system-wide implications of genetic manipulations may be responsible for suboptimal production levels. By accounting for both primary and some secondary metabolism pathways, the Cyanothece iCyt773 model can be used to explore in silico the effect of genetic modifications aimed at increased production of useful biofuel molecules. By taking full inventory of Cyanothece 51142 metabolism (as abstracted in iCyt773), and applying available strain optimization techniques [88,89] optimal gene modifications could be pursued for a variety of targets in coordination with experimental techniques. In particular, the availability of a microaerobic environment in Cyanothece 51142 at certain times during the diurnal cycle can be exploited for the expression of novel pathways that are not usually found in oxygenic cyanobacterial strains that largely maintain an aerobic environment. However, the use of Cyanothece 51142 as a bio-production platform is currently hampered by the inability to efficiently carry out genetic modifications.

By systematically cataloguing the shared (and unique) metabolic content in iSyn731 and iCyt773, successful genetic interventions assessed experimentally for Synechocystis 6803 can be “translated” to Cyanothece 51142. For example, it has been reported [90,91] that overproduction of fatty alcohols can be achieved in Synechocystis 6803 upon cloning a fatty acyl-CoA reductase (far) from Jojoba (Simmondsia chinensis) and the over-expression of gene slr1609 coding for an acyl-ACP synthetase. By using models iSyn731 and iCyt773 we can infer that in addition to cloning far from Jojoba, over-expression of gene cce1133 coding for a native acyl-ACP synthetase would be needed to bring about the same over-production in Cyanothece 51142.

**Supporting Information**

**File S1** Synechocystis Syn731 model along with established GPR, metabolite, gene and protein information. (XLSX)

**File S2** Cyanothece iCyt773 model along with established GPR, metabolite, gene and protein information. (XLSX)

**File S3** Biomass component measurements and stoichiometry of biomass equation. (XLSX)

**File S4** Reactions with diurnal activation/inactivation.
References


Author Contributions

Conceived and designed the experiments: RS ATV BMB CDM. Performed the experiments: RS. Analyzed the data: RS ATV BMB CDM. Wrote the paper: RS ATV BMB CDM. Developed the models: RS ATV.

Acknowledgments

We acknowledge Thanura Elvitiqda and Lawrence Page for contributions during the amino acid and pigment measurements. We also thank Anthony P. Burgard and Alireza Zomorrodii for discussions during the reconstruction process.

PLOS ONE | www.plosone.org 16 October 2012 | Volume 7 | Issue 10 | e48285


