Algal Research 27 (2017) 295–310

Contents lists available at ScienceDirect

Algal Research

journal homepage: www.elsevier.com/locate/algal

Modeling and analysis of flux distribution and bioproduct formation in *Synechocystis* sp. PCC 6803 using a new genome-scale metabolic reconstruction

Chintan J. Joshi[a], Christie A.M. Peebles[a,b], Ashok Prasad[a,b,*]

[a] Chemical and Biological Engineering, Colorado State University, Fort Collins, CO 80523, United States
[b] School of Biomedical Engineering, Colorado State University, Fort Collins, CO 80523, United States

ARTICLE INFO

Keywords:
Metabolic engineering
Dynamic flux balance analysis
Cyanobacteria
Biofuels
Genome-scale models

ABSTRACT

Cyanobacteria are prokaryotes capable of performing oxygenic photosynthesis. This makes them an attractive candidate for genetic engineering to produce commercially important chemicals. However, optimally harnessing this potential requires an understanding of metabolic regulation in cyanobacteria under photoautotrophic conditions. Here we present an updated genome-scale metabolic network reconstruction (iSynCJ816) of *Synechocystis* sp. PCC 6803. This updated model, containing 816 genes and 1045 reactions, builds upon previously published models. New features include an unconstrained photo-respiratory reaction mechanism and a mechanism to account for changes in energy absorption from light at different wavelengths. We used Flux Balance Analysis (FBA) to calculate the flux distribution within iSynCJ816 and compared in silico predictions with values obtained by previous in vivo metabolic flux analyses in *Synechocystis* sp. PCC 6803. A qualitative growth comparison of 167 gene-deletion mutants with experimental studies resulted in an accuracy rate between 70 and 79%. We used the model to estimate both the maximum theoretical yield of each metabolite and the feasibility of engineering *Synechocystis* to increase CO_2 fixation. We found that it is theoretically possible to increase CO_2 fixation by up to 35% from wild-type levels. We also carried out a dynamic flux balance analysis of fluxes throughout a light-dark cycle and obtained results that qualitatively matched experimental observations.

1. Introduction

Cyanobacteria are the only oxygenic prokaryotes capable of converting abundantly available carbon dioxide and sunlight into chemical energy via photosynthesis. As primary producers in aquatic environments, they play an important role in CO_2 assimilation and oxygen evolution. Geological and geochemical research has indicated that oxygen evolution by photosynthetic cyanobacteria is primarily responsible for the presence of significant levels of molecular oxygen in Earth's atmosphere, a process that began approximately 3 billion years ago [1]. Today, cyanobacteria are primary photosynthesizers, accounting for nearly 30% of Earth's photosynthetic productivity [2]. Their photosynthetic capabilities, along with other useful properties such as fast generation times relative to plants, small genomes that are amenable to transformation and synthesis of large numbers of useful metabolites have made them an attractive platform for genetic modifications to produce commercially important chemicals such as biofuels [3], pharmaceuticals [4], and nutraceuticals [5].

Among the gamut of cyanobacterial strains, *Synechocystis* sp. PCC 6803 is a model organism that has been extensively studied [6] with the aim of developing it as a chassis for metabolic engineering purposes. However, rational strategies for metabolic engineering require tools to help us understand the complexity of the metabolic network, to account for the metabolic trade-offs and bottlenecks. Genome-scale metabolic reconstructions, along with simulation techniques like FBA, provide a set of valuable tools to rationally engineer microorganisms like *Synechocystis* to meet commercial and societal needs. Because genome-scale reconstructions integrate available information into a systems framework, models ideally build upon previous work and address new questions in each iteration. Once a comprehensive reconstruction is obtained, various sets of methods developed by computational systems biologists are applied to validate and make predictions. One such well-known computational framework is FBA, a constraint-based modeling method used to calculate the flow of carbon through various metabolites to optimize reaction flux at steady state [7]. In the past, FBA has been utilized to predict flux distribution through wild-type [8–12] and mutant strains [13–15], epistatic interactions [16–20], futile cycles [21], and gene essentiality analysis [22–24].

* Corresponding author at: Chemical and Biological Engineering, Colorado State University, Fort Collins, CO 80523, United States.
E-mail address: ashokp@engr.colostate.edu (A. Prasad).

http://dx.doi.org/10.1016/j.algal.2017.09.013
Received 29 September 2016; Received in revised form 8 September 2017; Accepted 9 September 2017

Since the first cyanobacterial genome-scale metabolic reconstruction was published in 2005 [25], there have been nine more reconstructions [12,25–33]. They have progressed from being qualitatively useful to quantitatively useful models. Each subsequent reconstruction was more detailed than the previous one due to the availability of better data; the most recently published models included 677 genes [33]. More recently, a combined model was built by using a computational tool to aggregate three recent models which yielded a network with 778 reactions [34]. However, this combined model does not appear to have included a gene-reaction association table and has not been subjected to manual curation or validation with experimental data. This is important since some of the recently published models have been built independently of each other and contain inconsistencies among the models.

In this paper, we present a new genome-scale metabolic reconstruction of *Synechocystis* sp. PCC 6803. This reconstruction reconciles inconsistencies in the gene annotation within previously published reconstructions by mining databases such as KEGG [35–37] and Cyanobase [38,39]. Molecular mechanisms of the photosynthetic network around the thylakoid membrane have been included to facilitate a better understanding of respiratory and photosynthetic interactions. It should be noted that these molecular mechanisms have been included in only the most recent reconstruction [40]. Our work greatly improves upon earlier reconstructions by including thermodynamic analysis of > 500 reactions and by testing for thermodynamically infeasible loops (TILs). The inclusion of thermodynamic information leads us to remove or constrain reactions that participated in TILs.

To validate the reconstruction, we applied constraints and environment-specific conditions to simulate flux distributions under autotrophic (growth on carbon dioxide and light) and heterotrophic (growth on glucose) growth. We compared the fluxes generated by our model with experimental fluxes that were previously obtained via metabolic flux analysis (MFA) under autotrophic and heterotrophic conditions [41,42]. To further validate the genome-scale reconstruction, we compared the *in silico* predicted growth/no-growth phenotypes of 167 metabolic gene knockouts with experimental results obtained from online databases and a detailed literature search. Our predictions showed better accuracy than previously published models.

We used FBA to simulate intracellular fluxes and demand fluxes of each metabolite (also called bioproduct production flux) under autotrophic conditions within the model to determine which metabolites could theoretically be produced with high productivities. Analysis of the data revealed that, by suitably maximizing bioproduct production flux, carbon fixation can be improved by up to 35%, and oxygen production can be improved by up to 12%. Interestingly principal component analysis of the flux distribution showed that flux distributions clustered together based on photosynthetic state (i.e. PSI/PSII ratio) and pentose phosphate pathway activity.

One of the main challenges in metabolic modeling of cyanobacteria is the intrinsic variability in metabolism due to the sinusoidal light-dark cycle of a typical day. FBA is a static method that assumes that metabolism is in steady state. There is significant interest to develop a dynamic version of FBA, called Dynamic Flux Balance Analysis (DFBA). DFBA applications and implementations have increased since the first formulation was released in 1994 [43]. DFBA has been applied to metabolic networks of *E. coli* [44], *S. cerevisiae* [45], *L. lactis* [46], *C. reinhardtii* [47], *H. sapiens* (red blood cells) [48], and *S. stipites* [45]. Most of the DFBA methods fall into three broad classes. (i) The first is a dynamic optimization approach [44] in which an objective function is optimized over the entire time trajectory. This is a non-linear programming problem that typically requires a significant amount of kinetic information, which is currently unavailable for most organisms. (ii) The second approach, which is perhaps the most popular, is a piecewise static optimization approach [44]. Here time is discretized into small steps, and the objective function is optimized at each time step. Consistency with the light-dark cycle is maintained by changing the external conditions at each time step by hand. However, this method implicitly assumes that the organism is unaware of the light-dark cycle and the metabolic network only responds to the current environment. As a result, this method cannot predict glycogen accumulation or other dynamically changing aspects of metabolism. (iii) The third approach involves embedding the FBA model into a kinetic model of the external fluxes. This has been called the direct approach (DA-DFBA) [13,49]. The kinetic equations concern processes like metabolite uptake and secretion, which are experimentally accessible. This method requires less information than the fully dynamic method, and at the same time is more predictive than the static method [47]. We therefore implemented DA-DFBA to simulate fluxes during the light-dark cycle using the iSynCJ816 model.

Previous attempts to simulate fluxes using a dynamic method include the recently published model iHK677 [33], in which the authors designed a biomass objective function that changes over time, reflecting the changing expression of metabolic genes over the light-dark cycle. However, this is not very different from the static approach in that the changes in biomass composition are hard constraints that have been imposed externally. A better treatment of dynamics should be able to predict the changes in composition itself.

The DA-DFBA optimization scheme we used is based on a lexicographic optimization scheme to guarantee the uniqueness of the flux distribution [47]. This scheme is equivalent to the assumption that the metabolic network of the organism has many hierarchically organized objectives, which is more plausible than the assumption of just one objective.

2. Materials and methods

2.1. Model reconstruction and enhancement

The initial draft of the *Synechocystis* sp. PCC 6803 metabolic network was extracted from online databases. The genomic and pathway-related information was extracted using MATLAB codes from online databases including KEGG [35–37], Cyanobase [38,39], METACYC [50], and an annotated genome sequence [51]. We performed a homology search with the BLASTp algorithm on NCBI (online) [52] on the extracted sequences present in Supplementary Table 1 against other genomes. We used an accepted identity cut-off of 44% and an accepted E-value of 10^{-50} for this homology search. The information concerning enzymes was extracted from BRENDA [53] and KEGG. The information specific to chemical species (metabolites) was taken from ChEBI [54] and PubChem [55]. The reaction-based information about photosynthetic machinery was adapted from previously published studies [56–59] and cyanobacteria textbooks [60,61]. We also utilized previously published metabolic reconstructions of *Synechocystis* sp. PCC 6803 [31–33] and enhanced the biochemical and genomic information of various pathways such as photorespiration, serine synthesis, electron transfer within and between light harvesting proteins, fatty acid synthesis, chlorophyll synthesis, amino acid metabolism, and purine and pyrimidine metabolism. A literature search also resulted in the addition of a newly discovered light-independent serine synthesis pathway [62]. In this way, the draft network was subjected to iterative manual gap-filling, consistent with the protocol described by Thiele and Palsson [63]. This model building process was used to generate the list of genes, reactions, and metabolites. Any further processing and analysis of the reconstruction and pathways included has been discussed in the Results and discussion section.

The metabolic reconstruction was made by assimilating all the above information in an SBML (.xml) file [Supplementary File 1]. The reconstruction was converted to a mathematical model in MATLAB 2013b and readable/writeable with SBML toolbox [64] and COBRA toolbox [65,66].