A Data Integration and Visualization Resource for the Metabolic Network of *Synechocystis* sp. PCC 6803[1][W]

Timo R. Maarleveld, Joost Boele, Frank J. Bruggeman, and Bas Teusink*

Life Sciences, Centrum voor Wiskunde en Informatica, 1098 XG Amsterdam, The Netherlands (T.R.M.); BioSolar Cells, 6700 AB Wageningen, The Netherlands (T.R.M.); Systems Bioinformatics, Amsterdam Institute for Molecules Medicines and Systems, Vrije Universiteit Amsterdam, De Boelelaan 1085, 1081 HV Amsterdam, The Netherlands (T.R.M., J.B., F.J.B., B.T.); and Kluyver Centre for Genomics of Industrial Fermentation/ Netherlands Consortium for Systems Biology, P.O. Box 5057, 2600 GA Delft, The Netherlands (F.J.B., B.T.)

Data integration is a central activity in systems biology. The integration of genomic, transcript, protein, metabolite, flux, and computational data yields unprecedented information about the system level functioning of organisms. Often, data integration is done purely computationally, leaving the user with little insight in addition to statistical information. In this article, we present a visualization tool for the metabolic network of *Synechocystis* sp. PCC 6803, an important model cyanobacterium for sustainable biofuel production. We illustrate how this metabolic map can be used to integrate experimental and computational data for *Synechocystis* sp. PCC 6803 systems biology and metabolic engineering studies. Additionally, we discuss how this map, and the software infrastructure that we supply with it, can be used in the development of other organism-specific metabolic network visualizations. In addition to the Python console package VoNDA (http://vonda.sf.net), we provide a working demonstration of the interactive metabolic map and the associated *Synechocystis* sp. PCC 6803 genome-scale stoichiometric model, as well as various ready-to-visualize microarray data sets, at http://f-a-m-e.org/synechocytis.

The recent advances in high-throughput experimental technologies and whole-genome sequencing resulted in an explosion of biological data availability. These genome-scale data sets are generally described as "omics" data sets, of which transcriptomics, metabolics, and fluxomics data are examples (Gstaiger and Aebersold, 2009; Wang et al., 2009; Grabherr et al., 2011; Kahn, 2011). Analyses and integration of these data types can reveal important information about the system level functioning of organisms (Berger et al., 2013), and these methods are having a large impact on biology. To gain additional insight into cellular behavior, data sets are also used to develop computational models at a genome-scale level. Both the central role of metabolism for maintenance of cellular integrity and growth and the availability of genome-scale data sets led to an increased interest in genome-scale reconstructions of metabolism. These reconstructions are based on the metabolic reactions that they can perform, given the content of their genome (Oberhardt et al., 2009; Thiele and Palsson, 2010). Essentially, all metabolic reactions are catalyzed by enzymes, but for most reactions, the enzyme kinetics are unknown. Consequently, genome-scale reconstructions are often solely based on the reaction stoichiometry, and they are therefore generally referred to as genome-scale stoichiometric models (GSSMs) of metabolism.

To leverage the biological information that is included in GSSMs and to compute physiological properties, various stoichiometric network analysis tools are available (Price et al., 2003; Maarleveld et al., 2013). These tools can be used to computationally study the growth characteristics of microorganisms in detail. For instance, flux balance analysis (FBA) and flux variability analysis (FVA) have been used to predict the internal flux distribution and its variability while optimizing for cellular growth yield (Mahadevan and Schilling, 2003; Orth et al., 2010; McCloskey et al., 2013). Stoichiometric computational methods can also directly be of practical value. For example, OptKnock (Burgard et al., 2003) can be used to predict which gene knockout strategies would result in increased production of metabolites of interest, such as biofuels.

The analysis of GSSMs results in large data sets that often contain thousands of reactions. These results come in the form of long lists of (reaction) identifiers and (flux) values. This makes the interpretation of these data, which often come in files comprising thousands of lines, a rather tedious process. Yet, to be of any biological value, interpretation of FBA results in terms of biological

[1] This work was supported by funding from the BioSolar Cells project, cofinanced by the Dutch Ministry of Economic Affairs (to T.R.M.), the Centre for Integrative Bioinformatics Vrije Universiteit (to B.T. and J.B.), the Amsterdam Institute for Molecules, Medicines, and Systems (to B.T. and J.B.), Nederlandse Organisatie voor Wetenschappelijk Onderzoek Vidi Project 864.11.011 (to F.J.B.), the Netherlands Genomics Initiative, and ZonMW (Zenith grant no. 40–41009–98–10038 to B.T.).

* Address correspondence to b.teusink@vu.nl.

The author responsible for distribution of materials integral to the findings presented in this article in accordance with the policy described in the Instructions for Authors (www.plantphysiol.org) is: Bas Teusink (b.teusink@vu.nl).

[W] The online version of this article contains Web-only data.
www.plantphysiol.org/cgi/doi/10.1104/pp.113.224394

functions, modules, and pathways is required. To achieve this, adequate visualization is paramount, and attempts to visualize genome-scale data have been made. These approaches can be classified into supervised (human-driven) and nonsupervised (fully automatic) ones. The latter approach, fully automated mapping of genome-scale models, results in (mathematical) graphs that are even more difficult to interpret than the raw data one set out to visualize on them, so we will not discuss them in detail here.

Human-driven visualization methods, on the other hand, are considerably more popular, as hand-drawn maps are markedly easier to interpret. Tools such as Kyoto Encyclopedia of Genes and Genomes (KEGG; Kanehisa and Goto, 2000), Cytoscape (Smoot et al., 2011), CellDesigner (Funahashi et al., 2003), and the Constraints Based Reconstruction and Analysis (COBRA) Toolbox (Schellenberger et al., 2011) each offer their own particular take on modeling and visualization. Nevertheless, because none of them was originally designed for visualizing genome-scale models or genome-scale data, they are of limited use for this purpose. This becomes clear when one attempts to share, extend, or adapt an existing map. Commonly, users who attempt this quickly find their progress obstructed by the need for special software or incompatible identifiers or simply by not being able to access an editable version of the desired image.

Understandably, these issues have thus far prevented the widespread communal use of visualization aids, as well as their use in combination with computational results. However, the increased availability of high-throughput data makes the lack of adequate graphical interpretation tools more dearly felt than ever. Here, we address this problem by presenting the first comprehensive data visualization tool, namely, a graphical map of metabolism that is capable of displaying any type of data that is associated with reaction identifiers, gene identifiers, or metabolite identifiers. Based on the increased interest in photosynthetic microorganisms and its importance as a model organism for metabolic engineering studies, we decided to base our graphical metabolic map on the metabolic reconstruction of *Synechocystis* PCC 6803 (hereafter, *Synechocystis*; Shastri and Morgan, 2005; Knoop et al., 2010; Nogales et al., 2012; Saha et al., 2012), the best characterized and most extensively studied cyanobacterium. The direct conversion of CO_2 into carbon compounds makes these microorganisms of great interest for the development of production techniques for third-generation biofuels (Chisti, 2007; Ducat et al., 2011; Angermayr et al., 2012). We illustrate the use of the map of this organism's metabolism and discuss how similar maps can be made for other organisms using the *Synechocystis* map as a template.

Our map is also directly useful for other (micro)organisms, because the metabolic core (e.g. the carbohydrate, energy, amino acid, and nucleotide metabolism) is highly conserved between different microorganisms (Peregrín-Alvarez et al., 2009). Therefore, the map can easily be extended to cover additional pathways or organisms. This process will become progressively easier as more reactions are added to the map, as any added reactions can in turn be used in further mapping efforts. To demonstrate the usefulness of our graphical map, we have integrated it into the Flux Analysis and Modeling Environment (FAME; Boele et al., 2012) at http://f-a-m-e.org/synechocystis. Here, our map can be readily used for the analysis and visualization of the genome-scale model of *Synechocystis*, which also comes preloaded. The map and a Python console application of our visualization software Visualization of Network Data (VoNDA) can also be downloaded separately from http://vonda.sf.net. The console application can be used to visualize the output of popular constraint-based modeling software such as the COBRA Toolbox. It can also be used to take advantage of the recent development of constraint-based modeling software in Python. Using VoNDA in combination with the Python Simulator for Cellular Systems-Constraint Based Modeling (PySCeS-CBM; Olivier et al., 2005) or COBRApy (Ebrahim et al., 2013) allows for (interactive) modeling and visualization in Python.

RESULTS AND DISCUSSION

Metabolic Map of *Synechocystis*

We have developed a comprehensive metabolic map of *Synechocystis* (Fig. 1A). It is based on the metabolic reconstruction of Nogales et al. (2012), which we extended with reactions from Knoop et al. (2010) and other sources (Supplemental Data S1). The extended model contains 904 reactions and 816 metabolites. The associated map provides an overview of the reactions in the model and the resulting metabolic pathways. Together, the model and the map act as a readily extensible framework for human-friendly visualization of simulation results or experimental data (Fig. 1, B and C). Throughout this project, we aimed at offering an environment that can easily be extended with additional data and analysis types. To ensure its applicability to a maximum number of data and analysis types, as well as to maximize its potential for reuse and adaptation by the cyanobacterium community or other communities, a number of considerations were made during the development of the graphical map of the metabolism of *Synechocystis*.

An essential feature is that simulation results, such as FBA or FVA, can be displayed within the tool that generated them (Fig. 1, A and B). Moreover, experimental data, such as various flavors of omics' data sets, can be displayed (Fig. 1C). This was achieved by ensuring that the identifiers of the reactions, metabolites, and genes in the underlying model match those on the map. Such an annotation link is essential, and as long as experimental data sets use the same identifier classification, they can be visualized. In our map, the underlying identifiers can be easily detected by a mouseover of the element of interest in a Web browser.

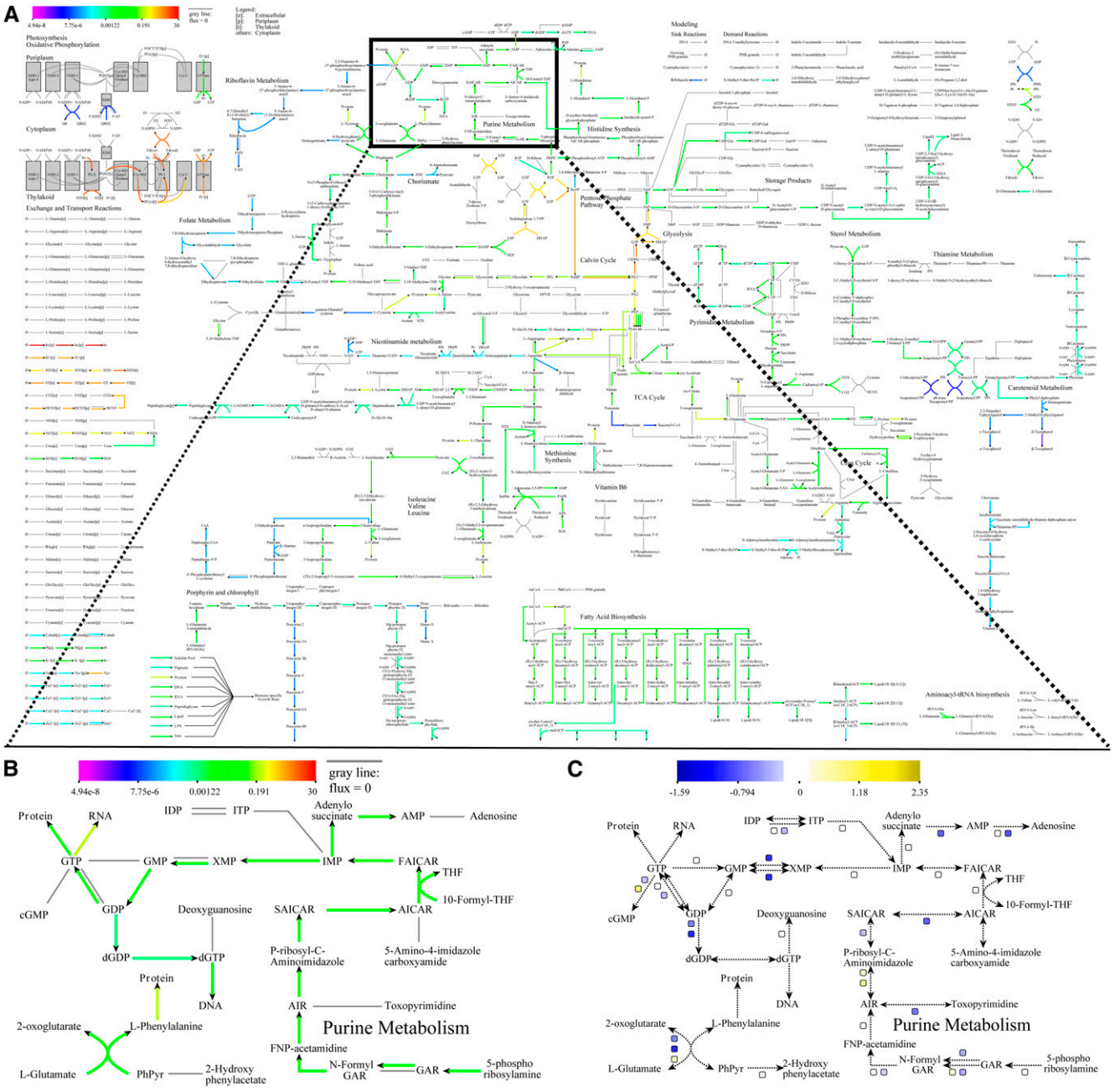

Figure 1. An overview of the graphical map of the metabolism of *Synechocystis* and its capabilities. Various elements (reactions, metabolites, and squares) are clickable and direct the user to the KEGG database or to run result details. A, When displaying FBA results, the arrows denoting reactions are colored according to the color scale in the top left corner. Gray reactions carry no flux. B, Detailed view of a section of the map when it is used to display FBA results. C, Detailed view of the same section of the map as in B, but now it is being used to display gene expression results. Here, the reaction lines are only displayed to indicate the general structure of the metabolic map, whereas the gene expression information is conveyed by colored boxes next to each reaction for which gene association information is available.

We used the Scalable Vector Graphics (SVG) format to represent the map, as this allowed us to make it interactive. Specifically, we included hyperlinks that can direct the user toward sources of additional information (e.g. KEGG) or, when FAME is used to generate simulation results and display them on the map, to details about the run that generated the displayed results. Had we utilized one of the budding visualization standards such as the Systems Biology Graphical Notation (Le Novère et al., 2009) or Systems Biology Markup Language (SBML) layout (Gauges et al., 2006), this functionality would require the user to install additional third-party software

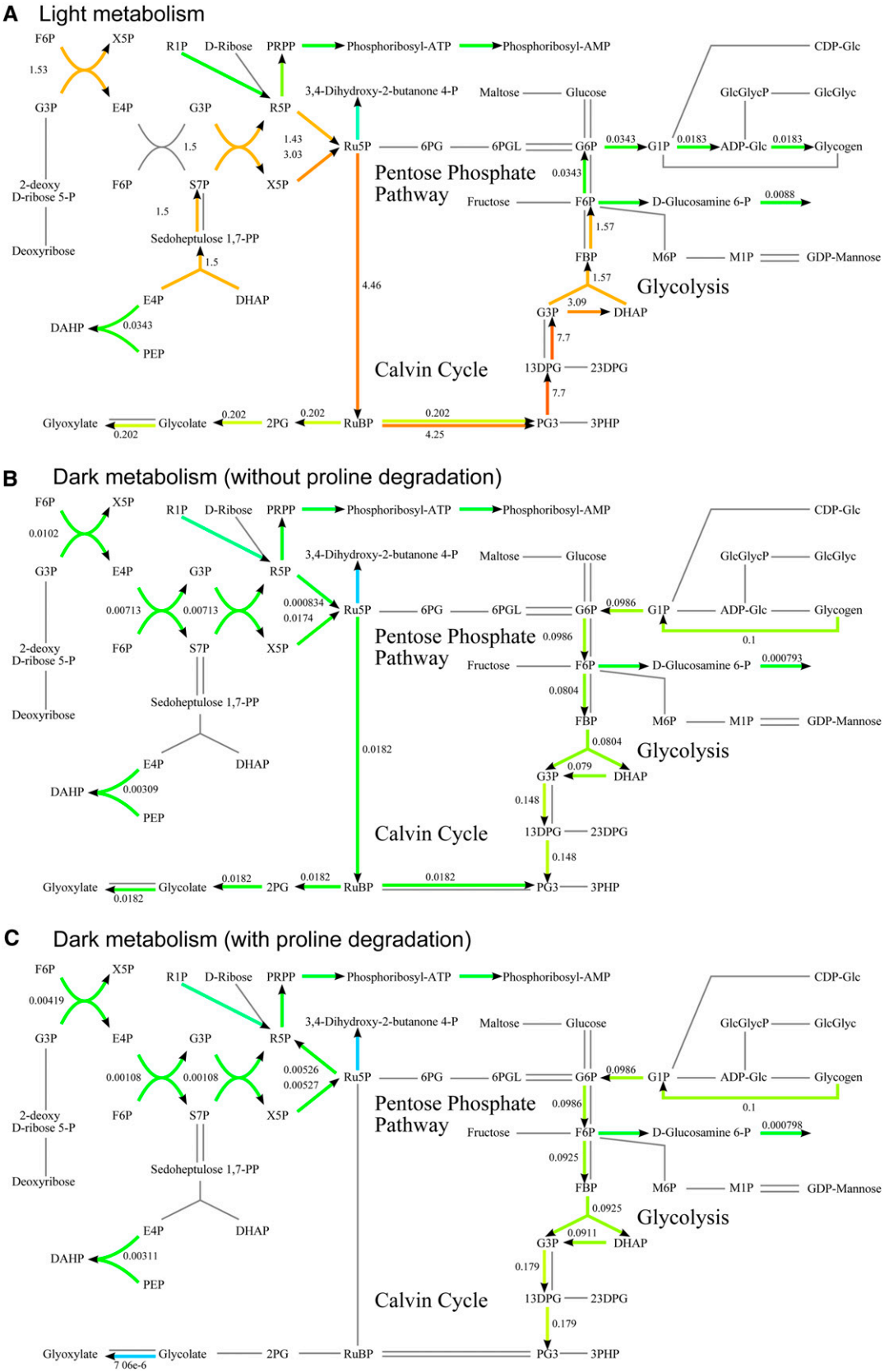

Figure 2. Example use cases of the map. Visualization of the difference in flux distribution when growing in the presence (A) or absence (B and C) of light. When light is present (A), the FBA solution harbors no surprises. However, when light is absent (B),

simply to view the image. Using SVG directly allows the user full graphical control over the image and allows direct embedding of dynamic, interactive elements that point to metabolite and reaction resources. The SVG format is commonly used on the Web, and popular image manipulation software (such as InkScape or Adobe Illustrator) can import SVG files. Therefore, these files can be used for figures in scientific publications. Using SVG also makes sure our map is readily extensible with new reactions or visualization applications. Moreover, the format is programmatically accessible, which allows for extraction and visualization of subnetworks of the entire metabolic map or development of a new metabolic map for another organism, using this map as a blueprint.

To demonstrate that our tool greatly simplifies the analysis of genome-scale data sets, we will provide two example use cases. One example concerns the physiological adaptation of cyanobacteria to changing environmental conditions. The other illustrates the expansion of the map with new reactions for biofuel production. Given GSSM constraints such as light availability, FBA is generally an accurate method to predict the optimal product or biomass yield. However, because these GSSMs are typically large, the associated metabolic flux distributions are large, which makes them difficult to analyze in detail. To exemplify this, FVA done on our GSSM predicts that, in an optimized GSSM, about 600 to 700 reactions can carry a flux during light and dark conditions. Our metabolic map allows for visualization of these FBA and FVA predictions, which allows for direct insight into, for instance, unexpected and interesting behavior.

Visualizing the Physiological Adaptation to Light and Dark Conditions

We will provide an example of unexpected behavior that can be found by modeling cyanobacteria in different environmental conditions in this section. Cyanobacteria such as *Synechocystis* are subject to a diurnal light-dark cycle and adjust their metabolism via circadian control. In the GSSM, we can simulate these conditions by changing the availability of light (photon influx) and adjusting the model constraints to indicate whether glycogen is to be produced as part of biomass (see Supplemental Data S2 for details).

In Figure 2, we illustrate FBA predictions of the metabolic activity under light and dark conditions. We limit ourselves to the visualization of the metabolic surroundings of Rubisco. Rubisco is a well-known key enzyme for carbon fixation during photoautotrophic growth, which carboxylates or oxygenates D-ribulose 1,5-bisphosphate. The carboxylation process (carbon fixation) yields two molecules of 3-phosphoglycerate, while the oxygenation process (photorespiration) results in one molecule each of 3-phosphoglycerate and 2-phosphoglycolate. 2-phosphoglycolate is toxic to *Synechocystis* and inhibits proper functioning of the Calvin-Benson cycle (Bauwe et al., 2012) and, as a result, reduces the conversion of solar energy into chemical energy. Nevertheless, both carbon fixation and photorespiration are essential for cyanobacteria in photoautotrophic growth conditions (Eisenhut et al., 2008).

We predicted steady-state flux values through this segment of metabolism by using FBA to optimize the biomass synthesis rate. Note that this rate optimization is carried out relative to other fixed nutrient uptake reaction rates and that we are effectively optimizing flux ratios (yields) rather than rates. Now, instead of working one's way through a list of about 1,000 predicted flux values for unexpected behavior, our map directly shows that both FBA and FVA predict that, in addition to carbon fixation, photorespiration is essential for obtaining an optimal growth yield in light and dark conditions (Fig. 2, A and B). The reason is that, in the metabolic model, photorespiration is the only way to produce glyoxylate, which is necessary for the production of the amino acids Cys, Gly, and Ser.

As "dark photorespiration" is not generally considered plausible, an alternative was proposed by Knoop et al. (2010). They added the conversion of L-Pro to Hyp (identified by a BLAST search (Altschul et al., 1997)) to provide a second, perhaps more plausible mechanism for glyoxylate synthesis in dark conditions. After inclusion of this reaction, our computational analysis showed that in light conditions no activity was predicted for the conversion of L-Pro to Hyp due to a lower carbon efficiency. According to FBA simulations, glyoxylate synthesis via L-Pro breakdown would be more efficient in dark conditions (Fig. 2C).

In addition to directing fixated carbon toward the production of precursors for (immediate) growth (e.g. nucleic acids or amino acids), *Synechocystis* also synthesizes the storage compound glycogen (Fig. 2A). During photoautotrophic growth, this results in a suboptimal growth strategy, as a lower growth rate will be obtained. In the absence of light, *Synechocystis* consumes the stored glycogen via either fermentation or respiration, depending on the availability of O_2. In the absence of light and presence of O_2, FBA predicts the breakdown of previously stored glycogen and energy generation via respiration. This enables the organism to maintain its integrity and grow under dark conditions. If the maintenance requirements of *Synechocystis* in the absence of light were known, modeling could help determine the minimal glycogen amount that should be stored during the light phase. This could

Figure 2. (*Continued.*)
FBA predicts that photorespiration is active, which is unlikely. C, Adding a reaction that converts L-Pro to Hyp returns the model to its expected state under dark conditions. Note that some glyoxylate is still produced from other precursors than glycolate (not shown here) to enable the production of amino acids.

represent an optimal strategy for this organism: optimize glycogen storage during the day to grow as fast as it can and then survive the night on the stored glycogen.

Exploring the Properties of *Synechocystis* Photosynthesis

In addition to photorespiration, photosynthesis is another key part of the metabolism of *Synechocystis* that is essential for growth. We studied the flexibility of this metabolic subnetwork by maximizing the specific growth rate as a function of the HCO_3^- and photon uptake rates (Fig. 3A). The specific growth rates we predict are in agreement with experimentally measured rates (Shastri and Morgan, 2005). We subsequently selected three combinations of HCO_3^- and photon uptake rates for further exploration (Fig. 3A). Briefly, these are (1) a state in which carbon and light are both limiting for growth (carbon- and light-limiting state [CLLS]), (2) a state in which light, but not carbon, limits growth (light-limiting state [LLS]), and (3) a state in which carbon, but not light, limits growth (carbon-limiting state [CLS]). We give a detailed overview of the characteristics of these conditions in Supplemental Data S2 and of the FBA predictions under each condition in Table I. In this section, we will discuss some analysis results that illustrate the practical application of our tool.

Visualization of FBA predictions on our metabolic map immediately shows that the model predicts different photosynthesis strategies under different conditions, indicating the flexibility of this system. The predictions in the CLLS (Fig. 3B) and LLS (Fig. 3C) are conceptually identical, although the predicted absolute fluxes are different. However, the two conditions can be distinguished on the basis of carbon secretion: the CO_2 that cannot be fixed in the LLS is released by the cell.

By comparing the visualization of photosynthetic fluxes, it is readily apparent that a different strategy is exploited in the CLS (Fig. 3D). Excess photons are used to transfer protons from the cytosol to the thylakoid, a process that is carried out by quinol oxidase (Cytochrome bd). These protons are then used to generate ATP. This explains the relatively low flux through PSI compared with PSII (J_{PSI}/J_{PSII}) and relatively high ATP formation flux compared with NADPH formation flux (J_{ATP}/J_{NADPH}); more details can be found in Table I. The resulting ATP is subsequently used in metabolism. FVA results (not shown) suggest that quinol oxidase is not essential for obtaining a maximal biomass yield. For example, the Mehler reaction can be used to oxidize the excess NADPH to NADP+ (Supplemental Fig. S1A; in that case, $J_{PSI}/J_{PSII} = 1$ and $J_{ATP}/J_{NADPH} = 1.28$, indicating linear electron flow).

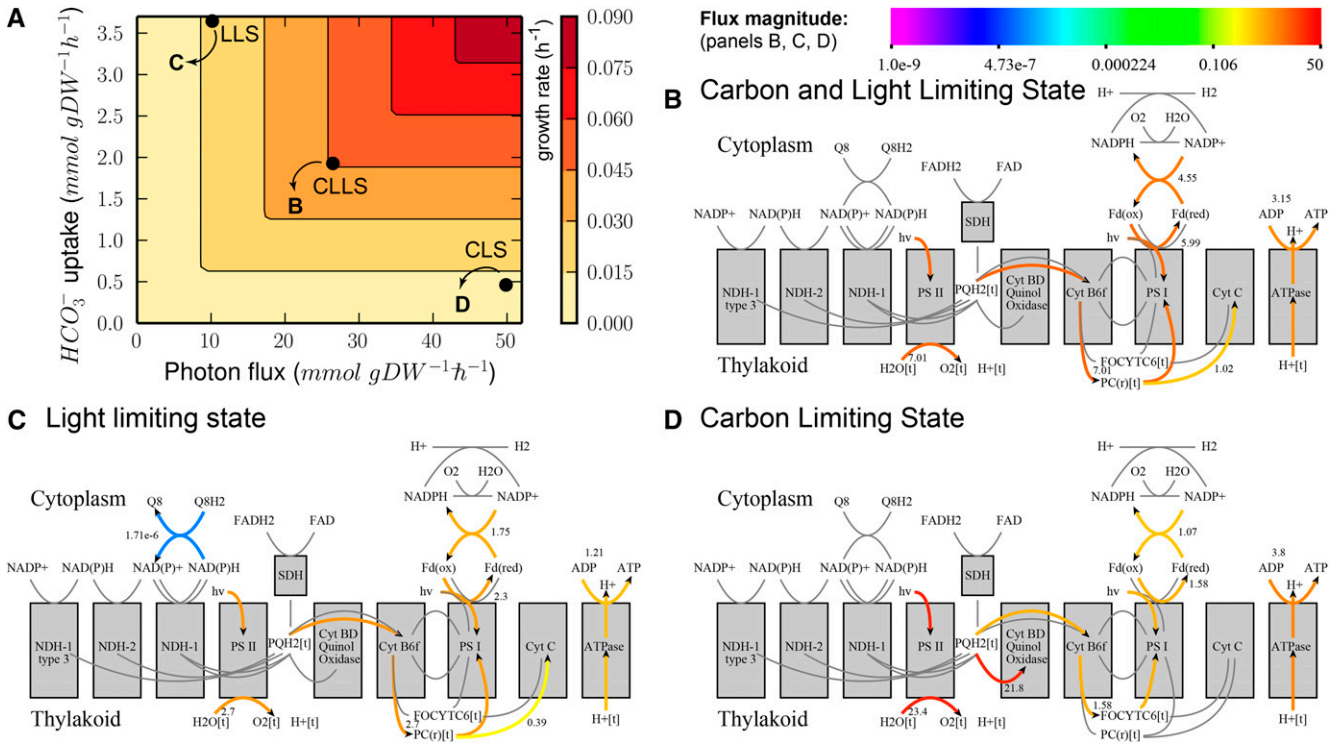

Figure 3. We used FBA to explore the properties of the *Synechocystis* model and map under various conditions, i.e. CLLS, LLS, and CLS. A, Predicted growth rate as a function of carbon and light uptake rates. The black dots in the phase plane indicate the conditions (in terms of photon flux and HCO_3^- uptake) that are shown in panels B to D. B to D, Visualization of predicted FBA fluxes for the three scenarios (CLLS, LLS, and CLS). Arrows indicate flux direction; the color bar indicates flux magnitude.

Table I. *A summary of the results of various (flux balance) analyses on the Synechocystis model*

When optimizing for biofuel production, we fixed the growth rate to 10% of its maximum. The constraints that were used for each of these simulations can be found in Supplemental Data S2. J_{hv}, Photon flux; KO AEF, knockout of AEF pathways; $J_{Rubisco,Co2}/J_{hv}$, Rubisco carboxylase flux divided by the photon flux; J_{23BD}, 2,3-butanediol flux; $\mu(h^{-1})$, growth rate.

Simulated Conditions	J_{PSI}/J_{PSII}	J_{ATP}/J_{NADPH}	$J_{Rubisco,CO2}/J_{hv}$	μ (h^{-1})	J_{23BD}
CLLS	0.99	1.6	0.085	0.045	0
CLS	0.058	7.2	0.011	0.012	0
LLS	0.99	1.6	0.085	0.017	0
KO AEF	1.0	1.28	0.075	0.040	0
Biofuel	0.99	1.3	0.13	0.0045	0.427
Biofuel CLS	0.058	8.1	0.014	0.0012	0.113
Biofuel LLS	0.99	1.3	0.13	0.0017	0.203
Biofuel KO AEF	1.0	1.28	0.11	0.0045	0.37

In addition to the photosynthetic linear electron flow, different alternative electron flow (AEF) pathways exist that allow for different ATP/NADPH ratios (Kramer and Evans, 2011). From Figure 3, B to D, we can immediately conclude that these AEF pathways are active during (these instances of) conditions of optimal growth yield. When the AEF pathways are removed from the model, the predicted growth yield is lower, indicating that AEF pathways are essential if maximum yield is to be attained under these conditions (Supplemental Fig. S1B).

Apart from plotting model predictions on the map, we can use the tool for visualization of many more data types. In Figure 4, we use microarray data under conditions of carbon limitation (similar to the conditions modeled above). Visualizing this type of data allows for convenient identification of genes that are over- or underexpressed compared with the wild type in a pathway-specific manner. For instance, compared with the wild type, our visualization shows that many genes associated with photosynthesis and respiration are initially underexpressed (Fig. 4A) upon CO_2 limitation. The transcriptomics data are from a time series, whereas predictions from FBA assume steady state. Therefore, these data sets are not directly comparable.

After 12 h of CO_2 limitation, the situation most likely represents a steady-state condition, and it can therefore be compared with FBA predictions. When this comparison is made, 73% of the photosynthetic gene expression data agree with predicted fluxes, while 27% do not match the flux predictions. For instance, genes associated with linear electron flow were underexpressed (Fig. 4B). FBA results visualized in Figure 3, B and D, show that lower fluxes are predicted in most reactions associated with the linear electron flow. By contrast, a decrease in ATPase gene expression was found under CLS conditions, whereas FBA/FVA predicted a higher flux through this enzyme. The gene expression data also indicate increased expression of NADH dehydrogenase 1 (NDH1) and NDH1 type 3. While FBA does not predict a flux through these reactions in CLS conditions, FVA results indicate that a flux through NDH1-related pathways would decrease the growth yield in LLS but not in CLS conditions. Thus, the NDH1 gene expression data, too, are consistent with the predicted solution space for photosynthesis fluxes in *Synechocystis*.

Expansion of the Map with a Biofuel Production Pathway

To illustrate how the metabolic map can be extended or modified to other microorganisms, we added a biofuel production pathway to the map and model. Recently, the development of a *Synechococcus elongatus* PCC 7942 strain that produces the bulk chemical 2,3-butanediol (23BD) was reported (Oliver et al., 2013). The production of 23BD using cyanobacteria offers several benefits compared with other higher alcohols. For instance, 23BD has a low host toxicity, which allows for a higher biofuel product concentration. Moreover, the common metabolite pyruvate is used as a substrate for 23BD production, and NADPH can be utilized as a reducing agent; no conversion of NADH to

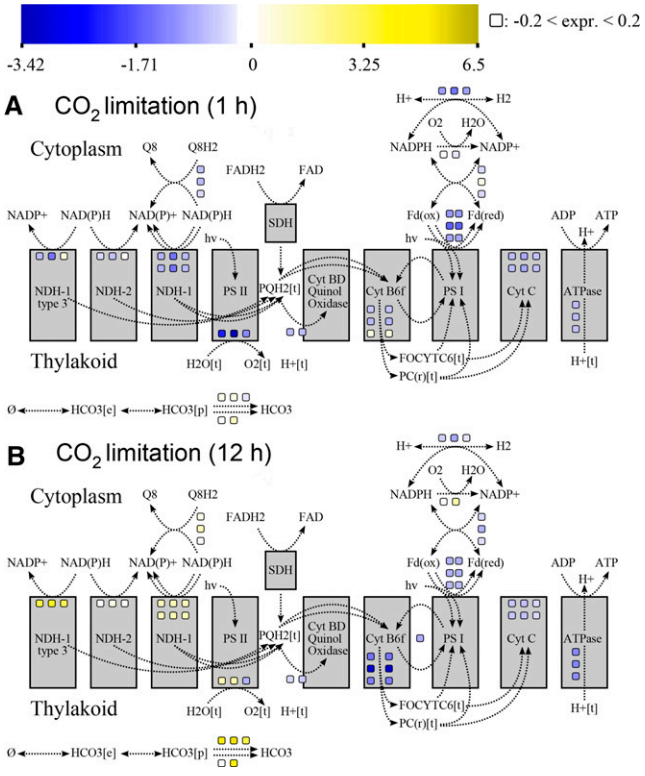

Figure 4. To demonstrate the applicability of our map to gene-associated ("omics"-sized) data, we used it to visualize microarray data for various CO_2-limiting growth conditions. The colored squares indicate differential expression relative to the wild type; yellow corresponds to overexpression, and blue corresponds to underexpression. The sections depict differential gene expression after 1 (A) and 12 h (B) of CO_2 limitation compared with wild-type *Synechocystis*.

NADPH is necessary. Finally, catabolic enzymes from O_2-tolerant organisms can be used for the production of 23BD, which is important because O_2 is produced by the photosynthetic machinery of *Synechocystis*. We therefore extended our map with a pathway that produces 23BD (see "Materials and Methods;" Supplemental Fig. S2).

We performed FBA on the model after adding the 23BD synthesis pathway under the same conditions as above (i.e. CLLS, LLS, and CLS; Fig. 5). Ideally, cyanobacteria can be used as cell factories (Angermayr et al., 2012), i.e. a nongrowing population that catalyzes the conversion of CO_2 into products of interest. To account for biomass replenishment and for the fact that a little growth will always occur, we fixed the growth rate to a basal growth level of 10% of its maximum specific growth rate. We then optimized for 23BD production (Fig. 5A). Given these conditions, our GSSM predicted 23BD production rates up to 0.90 mmol g^{-1} dry weight h^{-1}. Interestingly, the ratio of Rubisco carboxylase activity over the photon uptake flux was about 1.5 times higher when we optimized the flux to 23BD compared with optimizing the specific growth rate (Table I). Our metabolic map allowed for a straightforward identification of different factors (Supplemental Fig. S3) that contributed to this difference. Firstly, HCO_3^- is not entirely converted into CO_2 when biomass is synthesized. Instead, a substantial part (7%) of HCO_3^- is used for the synthesis of malonyl-CoA, an essential precursor for fatty acid biosynthesis. Secondly, the two reactions converting pyruvate to R-acetoin (the precursor of 23BD) release CO_2, which is subsequently fixed in the Calvin cycle. Lastly, it takes fewer photons to optimize the flux to 23BD (compared with optimizing for biomass production).

The production pathway of 23BD we added to the model requires NADPH but not ATP (Supplemental Fig. S2B). Therefore, we expected a lower J_{ATP}/J_{NADPH} ratio compared with the optimization of specific growth rate. Our simulations show that when optimizing for 23BD production, the percentage of flux through PSI devoted to AEF pathways is much lower for both the CLLS and LLS (Fig. 3, B and C, versus Fig. 5, B and C). As a result, optimization of 23BD production required a lower J_{ATP}/J_{NADPH} ratio. This ratio is relatively close to the ratio that is obtained via a linear electron flow. Therefore, one may expect that the AEF knockout yields a similar flux toward 23BD.

However, surprisingly, an AEF knockout in the CLLS and LLS resulted in a decrease of 12% in 23BD production rate. As it turns out, the relatively small fraction of flux through the AEFs is eventually used for HCO_3^- uptake. Without the AEF pathways, less HCO_3^- can be imported into the cell, which results in a lower production of 23BD. By contrast, in the CLS, where light is available in excess, many metabolic routes will lead to an optimal FBA solution, even though the objective that is

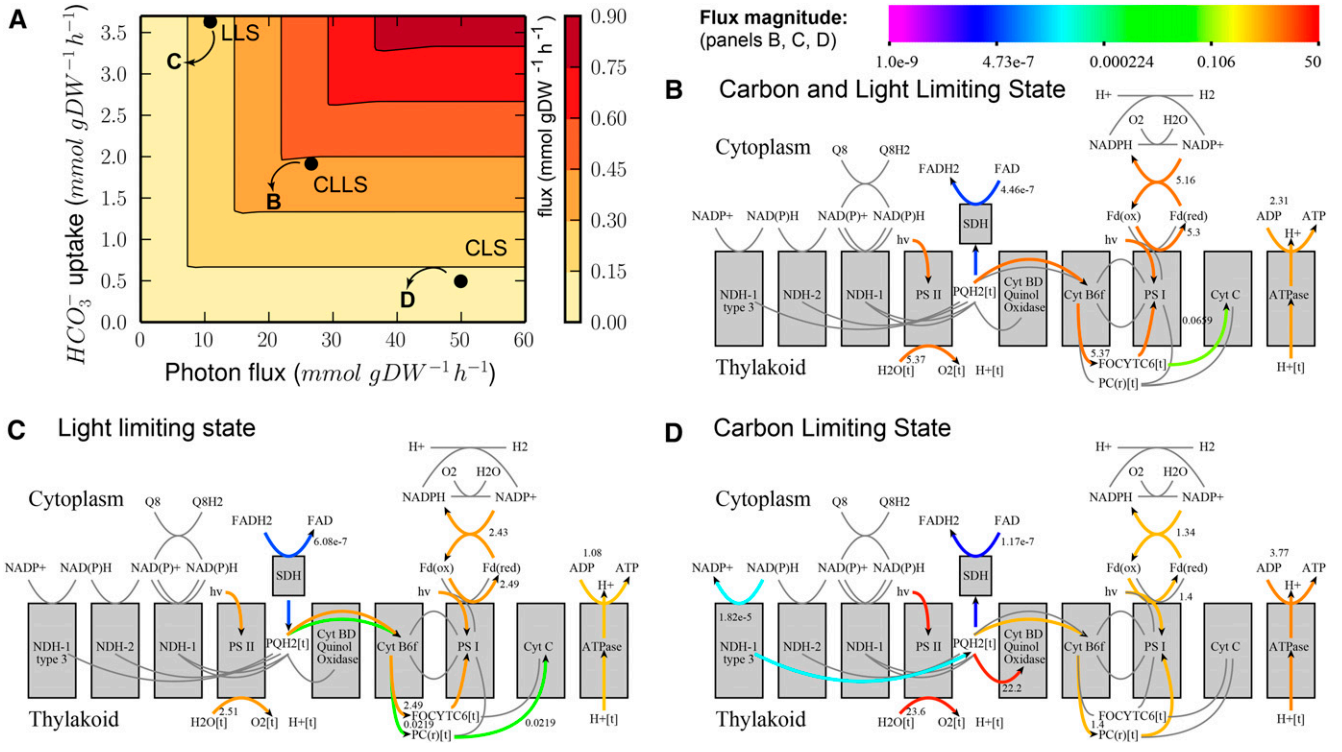

Figure 5. We used FBA to explore the properties of the *Synechocystis* model and map under various conditions (i.e. CLLS, LLS, and CLS), while optimizing for 23BD production. A, Predicted flux to 23BD as a function of carbon and light uptake rates. B to D, Visualization of predicted FBA fluxes for the three scenarios (CLLS, LLS, and CLS, respectively) indicated in A. Arrows indicate flux direction; the color bar indicates flux magnitude.

optimized for may differ; in the case of our simulations, the objective was either growth (Fig. 3D) or biofuel production (Fig. 5D). Consequently, the characteristics of the system in a CLS are quite similar between these two objectives (compare with Figs. 3D and 5D).

Implementation

With the development of the first interactive map of cyanobacterial metabolism, we provide the community with a graphical means of interpreting the large amounts of data generated by high-throughput experimental or computational methods. Given the absence of commonly available visualization options for results of analyses of genome-scale models, our main aim was the creation of an immediately useful map of *Synechocystis* metabolism. Visualization of heterogeneous data on this map is enabled by the standalone VoNDA package or the integration into FAME.

In addition to providing a pathway-oriented, human-friendly overview of the contents of the genome-scale metabolic model of *Synechocystis*, we included a number of features in our graphical map that will facilitate interpretation despite not being common practice in current-day visualization efforts. An important one is the choice to include multiple instances of common metabolites, such as ATP, ADP, NAD(P)(H), and other cofactors. These metabolites are included on the map in various locations, which greatly reduces the clutter that would arise if only one instance of each were drawn (graph-based visualization solutions tend to do this).

Other helpful features are hidden when they are not in use but deserve a brief explanation here. Space has been reserved on the map to display flux values for reactions and the direction in which the reaction takes place. An example is reaction-associated squares, which can be used for the visualization of gene expression data (Figs. 1C and 4). Each metabolic reaction has three such squares associated with it; if a reaction has more than three genes associated with it, we visualize the three genes with the highest differential expression value. Furthermore, when displaying run results generated in FAME, metabolite names link to a page with a list of fluxes that produce and consume the respective metabolite. Finally, when displaying run results, the lines that represent reactions are color coded to represent the magnitude of the flux through them (or the variability of this flux, in the case of FVA results), and an explanatory color bar is displayed at the top of the image.

Exploiting the Metabolic Map of *Synechocystis* for Other Microorganisms

To alleviate the effects of our design decisions on users that would have favored another approach, we created the entire map in SVG, an open-source image format. Our GSSM uses the BiGG nomenclature (Schellenberger et al., 2010), and therefore, our map is of direct value for all GSSMs that utilize this commonly used nomenclature. In addition, all identifiers on the map can be easily migrated to cognate identifiers in another nomenclature system (Lang et al., 2011; Bernard et al., 2014; Kumar et al., 2012). More importantly, the map can be modified to suit different species than *Synechocystis*. The addition of reactions and metabolites to the map can be done using either text or image manipulation software; when using an image editor, one must make sure that each new element is somehow assigned the correct identifier. We provide an online SVG editor in FAME, which allows for easy manipulation of the metabolic map, including any results superimposed on it. While the quality of any customized map will be strongly dependent on the knowledge and skill of whoever modified it and on the quality of the model it describes, this work presents the community with a solid basis from which to start this process. We provide an example of how one can go about this in "Materials and Methods" and in Supplemental Figure S2. Our demonstration of the expansion of the map and model with a biofuel production pathway not only illustrates the extensibility of our work, but also how adequate visualization can lead to novel biological insights.

Regardless of any discussion about its implementation, the presentation of the first map of metabolism that is both computer and human friendly will surely help advance our understanding of *Synechocystis* but also of other organisms to which the map can be adapted. By offering this resource not only as a flat file, but also as an integrated part of an interactive online analysis environment, we hope it will be of even greater use to the community.

CONCLUSION

The increased pace at which experimental data can be generated has fueled genome-scale research efforts, which presently call for genome-scale interpretation efforts. For the case of *Synechocystis*, an important model organism for biotechnological applications, we now enable the interpretation of these data from a biological (pathway) perspective by presenting the first comprehensive graphical map of its metabolism. This is the first results interpretation tool that is graphical, interactive, machine friendly, and, importantly, extensible. We expect that it will add great value to the community of cyanobacterium researchers but also to the scientific community as a whole.

MATERIALS AND METHODS

Adaptation of the *Synechocystis* Model and Creation of the Corresponding Map

We present a metabolic map and a corresponding stoichiometric model of *Synechocystis* sp. PCC 6803 metabolism consisting of 904 reactions, 816 species, and 686 genes. The model we present here is an extension of the iJN678 model described in Nogales et al. (2012). Pathways that were amended include the tricarboxylic acid cycle (Zhang and Bryant, 2011), Arg metabolism (Schriek et al., 2007, 2008, 2009), and Pro metabolism (Knoop et al., 2010). A full list of the modifications we made can be found in Supplemental Data S1. For

example, we modified the succinate dehydrogenase reactions in iJN678 to ensure that $FADH_2$ can be synthesized and degraded (originally, iJN678 was unable to degrade $FADH_2$). This modification also allowed us to extend Arg metabolism, where $FADH_2$ is synthesized. To facilitate the interpretation of results deriving from the model, we decomposed the biomass function into reactions toward nine major components (e.g. DNA, RNA, and protein). This allows users to track the uptake and destination of each of these components individually (Supplemental Fig. S2).

In spite of its usefulness to biotechnology, the biological relevance of 23BD to organisms is relatively low, which makes it a rare find on metabolic maps, and our initial version of the *Synechocystis* map was no exception. The first reaction, the conversion of two pyruvate molecules into 2-acetolactate and CO_2, is an essential metabolic reaction and was therefore already on our map. We added two new reactions to our map (R_ACLDC and R_23BDD), as well as the metabolites they connect (R-acetoin and 23BD). We also added exchange reactions to the map and model.

In total, we added 44 reactions and 22 species, removed seven reactions, and revised incorrect mass and charge balancing for 19 reactions. We refer to our new version of the model as iTM686 (Supplemental Data S3 and S4; the model can be accessed or downloaded in the SBML Level 3 format at http://f-a-m-e.org/synechocystis. The model's constraints and objectives are represented using features from the recently launched Flux Balance Constraints package (http://sbml.org/Documents/Specifications/SBML_Level_3/Packages/Flux_Balance_Constraints_(flux)).

Analyses Performed on the Model

We analyzed all instances of the *Synechocystis* model using FBA and FVA. As these analyses hinge upon the quasi-steady-state assumption, they can only predict steady-state flux distributions. In-depth descriptions of these computational methods are abundant in literature. We therefore refer to Price et al. (2003), Orth et al. (2010), and Santos et al. (2011) for an overview of the technical and methodological details. Throughout this article, wherever we mention performing an FBA, this FBA was performed with the additional objective of minimizing the absolute sum of fluxes to obtain a more biologically sensible solution. The details of all simulations, including the constraints and objectives that were used, are described in Supplemental Data S2.

Integration with FAME

To demonstrate the functionality of the map as well as to facilitate its use, we have added a *Synechocystis*-specific section to FAME (Boele et al., 2012). Visiting http://f-a-m-e.org/synechocystis causes the *Synechocystis* genome-scale model (iTM686) and our map to be loaded. The user is then able to view, edit, and run the model and to have analysis results visualized on the map we present in this article. Though the ability to visualize results on custom-made maps was already present in FAME, the URL described above preloads the model and map and eliminates the need for loading several files separately.

In addition, we have expanded FAME with an in-browser image editor and with the ability to visualize gene expression data from microarray profiling experiments on maps that support this kind of data. We have preloaded the results of 98 experiments from the CyanoExpress microarray data set (Hernandez-Prieto and Futschik, 2012) into the *Synechocystis* section of FAME. These data sets can be selected and superimposed on the map for analysis as a demonstration of the ability of the *Synechocystis* map to display gene expression data as well as flux analysis results.

Supplemental Data

The following materials are available in the online version of this article.

Supplemental Figure S1. Results of knock-out simulations.

Supplemental Figure S2. The addition of a 2,3-butanediol pathway to the Synechocystis map.

Supplemental Figure S3. Predicted import and export fluxes when optimizing for different objective functions.

Supplemental Data S1. A full list of modifications made to the genome-scale model iJN678.

Supplemental Data S2. FAME batch commands to reproduce the results presented in this article.

Supplemental Data S3. Synechocystis metabolic network file.

Supplemental Data S4. Synechocystis metabolic network overview.

ACKNOWLEDGMENTS

We thank Brett Olivier, Pascal van Alphen, Philipp Savakis, Andreas Angelmayr, and Klaas Hellingwerf for helpful discussions and tips for developing the data integration and visualization tool.

Received July 3, 2013; accepted January 8, 2014; published January 8, 2014.

LITERATURE CITED

Altschul SF, Madden TL, Schäffer AA, Zhang J, Zhang Z, Miller W, Lipman DJ (1997) Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. Nucleic Acids Res **25:** 3389–3402

Angermayr SA, Paszota M, Hellingwerf KJ (2012) Engineering a cyanobacterial cell factory for production of lactic acid. Appl Environ Microbiol **78:** 7098–7106

Bauwe H, Hagemann M, Kern R, Timm S (2012) Photorespiration has a dual origin and manifold links to central metabolism. Curr Opin Plant Biol **15:** 269–275

Berger B, Peng J, Singh M (2013) Computational solutions for omics data. Nat Rev Genet **14:** 333–346

Bernard T, Bridge A, Morgat A, Moretti S, Xenarios I, Pagni M (2014) Reconciliation of metabolites and biochemical reactions for metabolic networks. Brief Bioinform **15:** 123–135

Boele J, Olivier BG, Teusink B (2012) FAME, the Flux Analysis and Modeling Environment. BMC Syst Biol **6:** 8

Burgard AP, Pharkya P, Maranas CD (2003) Optknock: a bilevel programming framework for identifying gene knockout strategies for microbial strain optimization. Biotechnol Bioeng **84:** 647–657

Chisti Y (2007) Biodiesel from microalgae. Biotechnol Adv **25:** 294–306

Ducat DC, Way JC, Silver PA (2011) Engineering cyanobacteria to generate high-value products. Trends Biotechnol **29:** 95–103

Ebrahim A, Lerman JA, Palsson BO, Hyduke DR (2013) COBRApy: COnstraints-Based Reconstruction and Analysis for Python. BMC Syst Biol **7:** 74

Eisenhut M, Ruth W, Haimovich M, Bauwe H, Kaplan A, Hagemann M (2008) The photorespiratory glycolate metabolism is essential for cyanobacteria and might have been conveyed endosymbiontically to plants. Proc Natl Acad Sci USA **105:** 17199–17204

Funahashi A, Tanimura N, Morohashi M, Kitano H (2003) CellDesigner: a process diagram editor for gene-regulatory and biochemical networks. BIOSILICO **1:** 159–162

Gauges R, Rost U, Sahle S, Wegner K (2006) A model diagram layout extension for SBML. Bioinformatics **22:** 1879–1885

Grabherr MG, Haas BJ, Yassour M, Levin JZ, Thompson DA, Amit I, Adiconis X, Fan L, Raychowdhury R, Zeng Q, et al (2011) Full-length transcriptome assembly from RNA-seq data without a reference genome. Nat Biotechnol **29:** 644–652

Gstaiger M, Aebersold R (2009) Applying mass spectrometry-based proteomics to genetics, genomics and network biology. Nat Rev Genet **10:** 617–627

Hernandez-Prieto MA, Futschik ME (2012) CyanoEXpress: a web database for exploration and visualisation of the integrated transcriptome of cyanobacterium *Synechocystis* sp. PCC6803. Bioinformation **8:** 634–638

Kahn SD (2011) On the future of genomic data. Science **331:** 728–729

Kanehisa M, Goto S (2000) KEGG: Kyoto Encyclopedia of Genes and Genomes. Nucleic Acids Res **28:** 27–30

Knoop H, Zilliges Y, Lockau W, Steuer R (2010) The metabolic network of *Synechocystis* sp. PCC 6803: systemic properties of autotrophic growth. Plant Physiol **154:** 410–422

Kramer DM, Evans JR (2011) The importance of energy balance in improving photosynthetic productivity. Plant Physiol **155:** 70–78

Kumar A, Suthers PF, Maranas CD (2012) MetRxn: a knowledgebase of metabolites and reactions spanning metabolic models and databases. BMC Bioinformatics **13:** 6

Lang M, Stelzer M, Schomburg D (2011) BKM-react, an integrated biochemical reaction database. BMC Biochem **12:** 42

Le Novère N, Hucka M, Mi H, Moodie S, Schreiber F, Sorokin A, Demir E, Wegner K, Aladjem MI, Wimalaratne SM, et al (2009) The systems biology graphical notation. Nat Biotechnol **27:** 735–741

Maarleveld TR, Khandelwal RA, Olivier BG, Teusink B, Bruggeman FJ (2013) Basic concepts and principles of stoichiometric modeling of metabolic networks. Biotechnol J **8:** 997–1008

Mahadevan R, Schilling CH (2003) The effects of alternate optimal solutions in constraint-based genome-scale metabolic models. Metab Eng **5:** 264–276

McCloskey D, Palsson BO, Feist AM (2013) Basic and applied uses of genome-scale metabolic network reconstructions of *Escherichia coli*. Mol Syst Biol **9:** 661

Nogales J, Gudmundsson S, Knight EM, Palsson BO, Thiele I (2012) Detailing the optimality of photosynthesis in cyanobacteria through systems biology analysis. Proc Natl Acad Sci USA **109:** 2678–2683

Oberhardt MA, Palsson BO, Papin JA (2009) Applications of genome-scale metabolic reconstructions. Mol Syst Biol **5:** 320

Oliver JW, Machado IM, Yoneda H, Atsumi S (2013) Cyanobacterial conversion of carbon dioxide to 2,3-butanediol. Proc Natl Acad Sci USA **110:** 1249–1254

Olivier BG, Rohwer JM, Hofmeyr JH (2005) Modelling cellular systems with PySCeS. Bioinformatics **21:** 560–561

Orth JD, Thiele I, Palsson BO (2010) What is flux balance analysis? Nat Biotechnol **28:** 245–248

Peregrín-Alvarez JM, Sanford C, Parkinson J (2009) The conservation and evolutionary modularity of metabolism. Genome Biol **10:** R63

Price ND, Papin JA, Schilling CH, Palsson BO (2003) Genome-scale microbial in silico models: the constraints-based approach. Trends Biotechnol **21:** 162–169

Saha R, Verseput AT, Berla BM, Mueller TJ, Pakrasi HB, Maranas CD (2012) Reconstruction and comparison of the metabolic potential of cyanobacteria *Cyanothece* sp. ATCC 51142 and *Synechocystis* sp. PCC 6803. PLoS ONE **7:** e48285

Santos F, Boele J, Teusink B (2011) A practical guide to genome-scale metabolic models and their analysis. Methods Enzymol **500:** 509–532

Schellenberger J, Park JO, Conrad TM, Palsson BO (2010) BiGG: a Biochemical Genetic and Genomic knowledgebase of large-scale metabolic reconstructions. BMC Bioinformatics **11:** 213

Schellenberger J, Que RM, Fleming RM, Thiele I, Orth JD, Feist AM, Zielinski DC, Bordbar A, Lewis NE, Rahmanian S, et al (2011) Quantitative prediction of cellular metabolism with constraint-based models: the COBRA Toolbox v2.0. Nat Protoc **6:** 1290–1307

Schriek S, Aguirre-von-Wobeser E, Nodop A, Becker A, Ibelings BW, Bok J, Staiger D, Matthijs HC, Pistorius EK, Michel KP (2008) Transcript profiling indicates that the absence of PsbO affects the coordination of C and N metabolism in *Synechocystis* sp. PCC 6803. Physiol Plant **133:** 525–543

Schriek S, Kahmann U, Staiger D, Pistorius EK, Michel KP (2009) Detection of an L-amino acid dehydrogenase activity in *Synechocystis* sp. PCC 6803. J Exp Bot **60:** 1035–1046

Schriek S, Rückert C, Staiger D, Pistorius EK, Michel KP (2007) Bioinformatic evaluation of L-arginine catabolic pathways in 24 cyanobacteria and transcriptional analysis of genes encoding enzymes of L-arginine catabolism in the cyanobacterium *Synechocystis* sp. PCC 6803. BMC Genomics **8:** 437

Shastri AA, Morgan JA (2005) Flux balance analysis of photoautotrophic metabolism. Biotechnol Prog **21:** 1617–1626

Smoot ME, Ono K, Ruscheinski J, Wang PL, Ideker T (2011) Cytoscape 2.8: new features for data integration and network visualization. Bioinformatics **27:** 431–432

Thiele I, Palsson BO (2010) A protocol for generating a high-quality genome-scale metabolic reconstruction. Nat Protoc **5:** 93–121

Wang Z, Gerstein M, Snyder M (2009) RNA-seq: a revolutionary tool for transcriptomics. Nat Rev Genet **10:** 57–63

Zhang S, Bryant DA (2011) The tricarboxylic acid cycle in cyanobacteria. Science **334:** 1551–1553