**Systems Biology Research Group**
University of California San Diego

Final Scientific/Technical Report

# Optimization of Energy Flow Through Synthetic Metabolic Modules and Regulatory Networks in a Model Photosynthetic Eukaryotic Microbe

**Dr. Bernhard O. Palsson, Principal Investigator**

| | |
|---|---|
| **Award:** | SE-SC0008701 |
| **Lead Recipient:** | Dr. Bernhard O. Palsson |
| **Project Title:** | Optimization of Energy Flow Through Synthetic Metabolic Modules and Regulatory Networks in a Model Photosynthetic Eukaryotic Microbe |
| **Program Director:** | Dr. Bernhard O. Palsson |
| **Principal Investigator:** | Dr. Bernhard O. Palsson |
| **Contract Administrator:** | |
| **Date of Report:** | 11/30/2018 |
| **Reporting Period:** | 09/01/2012 – 08/37/2018 |

## Public Executive Summary

Diatoms are important oceanic photosynthetic microorganisms, responsible for approximately 20% of all carbon fixation on Earth. Additionally, diatoms are attractive metabolic engineering candidates. This investigation leveraged Systems Biology techniques to advance our understanding of and potential to engineer diatoms for bioproducts of societal interest, such as biofuels. Systems Biology combines computational and experimental techniques to understand an organism at the whole-cell level as well as characterize the interface between the organism and its environment. Our research resulted in major contributions to the understanding of systems-level light-driven metabolism.

Diatoms have a unique evolutionary history. As a result, existing understanding of photosynthetic metabolism may not be applicable to diatom physiology. We generated, validated, and published a high-quality metabolic model of diatom metabolism facilitating novel insights into these unique organisms. Additionally, we overcame a consistent challenge in the modeling of photosynthetic organisms by integrating light into metabolic models. This result allows realistic assessments of how the light environment affects cell physiology and metabolism, a necessary advancement for bioengineering. Next, we generated several methodologies for integrating large-scale datasets with models of metabolism. These “big data” approaches begin to address a constant challenge in biology, which is translating data into information and understanding. Finally, we built web-based databases for housing, distributing, and visualizing systems-level biology data. These projects created an online ecosystem for sharing, analyzing, and characterizing metabolic information and is currently being used by public and private institutions.

Photosynthetic bioengineering has the promise to create high-energy products with minimal inputs. Systems Biology has a long history of facilitating bioengineering of cellular metabolism. However, photosynthetic organisms are only sparsely present in this history. This project and its outputs have accelerated the application of proven Systems Biology techniques to light-driven metabolism. The resulting methodologies will accelerate public and private endeavors leveraging photosynthetic systems in the bioengineering of bioproducts such as biofuels or carbon sequestration.

## Acknowledgements

The authors wish to thank the Department of Energy, Office of Biological and Environmental Research for funding this research. Significant portions of this work were performed in collaboration with Dr. Andrew E. Allen and Dr. Chris Dupont of the J. Craig Venter Institute, La Jolla, CA; Dr. Graham Peers of Colorado State University, Fort Collins, CO; and Dr. Susan S. Golden, University of California San Diego, La Jolla, CA.

## Table of Contents

Public Executive Summary ..... 2
Acknowledgements ..... 2
Table of Figures/Tables ..... 3
Accomplishments and Objectives ..... 4
Project Activities ..... 9
Project Outputs ..... 10
Follow-On Funding ..... 10

## Table of Figures/Tables

Table 1. Key Milestones and Deliverables ..... 4
Table 2. Follow-on Funding Received ..... 4

## Accomplishments and Objectives

This award allowed the Systems Biology Research Group at UC San Diego to demonstrate a number of key objectives. The focus of the project was on building a computational framework for the reconstruction and analysis of phototrophic metabolism. Several tasks and milestones were laid out at the beginning of the project. The actual performance against the stated milestones is summarized here:

### Table 1. Key Milestones and Deliverables.

| Tasks | Milestones and Deliverables |
|---|---|
| Task 1: Construct a high-quality metabolic reconstruction of *Phaeodactylum tricornutum* | Q1: Metabolic reconstructions are bottom-up, biochemically, genetically, and genomically (BiGG) structured knowledge bases in an accessible format that allows iterative refinement of content. It is the framework upon which high-throughput data will be contextualized, metabolic processes analyzed, and synthetic modules hypothesized. Accomplishment of the genome-scale reconstruction goal requires the draft, curation, validation, and dissemination of the metabolic model. The draft leverages existing procedures and metabolic reconstructions to generate a starting point for model curation. Manual curation is a labor-intensive process where the body of knowledge for a given organism is consolidated into the reconstruction framework. Upon completion, the reconstruction is converted into a mathematical form and quantitative outputs are validated against experimental results. The primary metric for completion of this goal is the publication of the model in a peer-reviewed journal and making the reconstruction available in a publicly accessible database. |

| | |
|---|---|
| | **Actual Performance:** 100% completed.<br>The *Phaeodactylum tricornutum* metabolic reconstruction, iLB1027, was completed and presented in a peer-reviewed publication (doi: 10.1371/journal.pone.0155038). The reconstruction included a complete functional re-annotation of the updated genome; increasing the number of proteins with an annotated enzyme commission number by almost 300%. Bioinformatic approaches enabled compartmental localization for every gene in the organism; a necessary constraint in our efforts to engineer the system. The model also incorporated a new FTIR based biomass assay developed by our co-PIs. The biomass composition of a phototroph is highly variable across a culture duration. The FTIR method enabled parameterization of the model over the culture duration resulting in accurate predictions of biomass yields. Finding rate limiting steps in lipid metabolism was a desired output of the model. Recent investigations into diatoms indicated energetic coupling between the chloroplast and mitochondria. We were able to model this coupling and assess the impact on lipid generation. The results indicated an inverse relationship between lipid production and energetic coupling. The model also suggested a unique metabolic pathway responsible for transferring reducing power between compartments. This ornithine based redox shuttle, if confirmed, would be a novel energy pathway. The inverse relationship between lipid accumulation and energetic coupling may be the metabolic basis for low lipid accumulation in diatoms during exponential growth. It also suggests a mechanism by which to redirect reducing power away from the mitochondria towards chloroplast lipid production; a primary aim of our project. Overall, the metabolic reconstruction met all its stated objectives and will serve as the framework for future hypotheses and analysis. In addition, the curation of the literature required to accomplish this reconstruction was assembled into a resource. The so called bibliome for *Phaeodactylum tricornutum* has been formalized and presented in a peer-reviewed publication (doi:10.1016/j.algal.2016.06.020). |
| | |
| Task 2: An integrated environment for Constraint-Based Analysis of Systems Biology | Q1: Development and improvement of modeling tools and databases translates the reconstruction to biological knowledge. The primary milestones towards this need are generating a computational infrastructure for multiple ‘omic data integration and increasing the accessibility of tools and data. An integrated environment for constraint-based analysis of biological networks is necessary for the rapid generation and analysis of standardized, high quality genome-scale models. Especially in the investigation of sparsely characterized organisms, such as diatoms, development and accessibility of high-throughput data analysis tools are required. |

| | |
|---|---|
| 2.1 Development of a computational infrastructure for multiple 'omic data integration and analysis | **Actual Performance:** 80% completed.<br>Accessibility and refinement of constraint-based (COBRA) modeling tools is critical for the rapid generation of standardized, high quality genome-scale models. The primary milestones towards this need are generating a computational infrastructure for multiple 'omic data integration and increasing the accessibility of tools and data. The steady-state assumption in flux balanced analysis (FBA) is violated by phototrophs due to circadian metabolic cycles. By integrating quantitative metabolomics data, we developed unsteady-state FBA (uFBA) allowing for analysis of non-steady state systems. The method development required a system with well-known metabolic capabilities and ample 13C metabolic flux data available. Since no phototroph meets these requirements, the model was validated in heterotrophic systems (*E. coli*, yeast, human red-blood cell). The validated protocol recapitulated metabolite pooling during time-course experiments; similar to expectations in phototrophic metabolism. However, since this method has yet to be applied to a phototrophic dataset it is only 60% complete.<br><br>The protocol has been presented in a peer-reviewed publication (doi:10.1038/srep46249). Integration of transcriptomics data along with protein-interaction data has been analyzed by clustering algorithms resulting in a hierarchal diatom transcriptional regulatory network. This work has been presented in a peer-reviewed publication (doi:10.1128/mSystems.00142-16). 100% complete. |
| | |
| Task 3: Maintenance, extension, and dissemination of COBRA methods and BiGG database | Q1: In the area of extending modeling tool accessibility, there are six subobjectives designed to increase standardization and dissemination. Accomplishment of the modeling tools and database goal requires extension of 'omics data accessibility and analysis, iterative improvements to modeling community standards, and network data visualization. The primary metric for completion is development and publication or dissemination of tools to meet these goals. |
| 3.1 Context-specific network tailoring | **Actual Performance:** 60% completed.<br>We developed a method that integrated proteomics data and genome-scale models to generate context-specific models, which was presented in a peer-reviewed publication (doi:10.1038/srep36734). Diatom datasets were not available at the time of the study, so the method was developed as a proof-of-concept using data for *E. coli*. The resulting method is applicable to all genome-scale models and will be leveraged in future efforts to characterize phototrophic metabolism. However, since this method has yet to be applied to a phototrophic dataset it is only 60% complete. |
| 3.2 Automated reconstruction procedures | **Actual Performance:**<br>Not addressed as the broader COBRA community has actively taken on the responsibility of archiving and updating existing methods. |
| 3.3 Web-portal Improvement | **Actual Performance:**<br>Not addressed as the broader COBRA community has actively taken on the responsibility of archiving and updating existing methods. |
| 3.4 SBML+COBRA extensions | **Actual Performance:** 100% completed.<br>A web-based COBRA SBML validator was developed and incorporated with the BiGG database (doi:10.1093/nar/gkv1049, see 3.6 below). |

| | |
|---|---|
| 3.5 Network visualization | **Actual Performance:** 100% completed.<br>To meet the visualization and data analysis needs of next-generation models; such as the diatom reconstruction, we developed Escher (escher.github.io). Escher is a web application for visualizing data on biological pathways. Users can rapidly design new pathway maps as Escher provides pathway suggestions based on data and genome-scale models. Users can also visualize data related to genes or proteins on the associated reactions and pathways. Thus, trends in common multiple 'omic data types can be identified. Escher harnesses the strengths of web technologies, so visualizations can be adapted, extended, shared, and embedded. Escher satisfies a previously un-met need in network analysis. It also paves the way for the development of automated reconstruction tools and high-throughput data analysis. Escher was presented in a peer-reviewed publication (doi:10.1371/journal.pcbi.1004321). |
| 3.6 BiGG database | **Actual Performance:** 100% completed.<br>To maximize the value of reconstructions, centralized stores of high-quality models must be established, models must adhere to standards, and components must be linked to databases. To meet these needs, we developed BiGG Models, presented in a peer-reviewed publication (doi: 10.1093/nar/gkv1049, http://bigg.ucsd.edu), a completely redesigned Biochemical, Genetic, and Genomic knowledge base. BiGG Models connects genome-scale models to genome annotations and external databases. Reaction and metabolite identifiers have been standardized to conform to community standards and enable rapid model comparisons. Furthermore, it provides a comprehensive application programming interface for access with modeling and analysis tools. As a resource for highly curated, standardized, and accessible models of metabolism, BiGG Models will facilitate systems biology studies and support knowledge-based analysis of experimental data; including those for the diatom reconstruction. |
| | |
| Task 4: Develop model-based platform for contextualization of high-throughput data | Q1: Contextualization of high-throughput diatom data is the application of modeling tools towards the understanding of diatom metabolism. Upon completion of the genome-scale metabolic model and the improvement of modeling tools and databases, existing diatom 'omics sets, such as an extensive RNAseq library, will be analyzed towards defining major metabolic nodes and regulatory mechanisms that form the basis of diatom physiology. The methods and insights generated during this analysis will drive hypothesis generation and experimental design for additional 'omics datasets. The intent is an iterative process of model-based hypothesis, experimental validation, and model based data analysis. Accomplishment of the contextualization of the high-throughput diatom data goal requires implementing the outputs of the previous two goals towards understanding and engineering of diatom metabolism. The primary metric for completion is the generation of unique insights into light driven metabolism that contributes to hypothesis generation and data analysis, supporting completion of project goals of the contributing co-PIs. |

| | |
|---|---|
| 4.1 Analyzing light-driven metabolism through COBRA methods | **Actual Performance:** 100% completed.<br>A fundamental barrier to predictive modeling of phototrophic growth is the inability to characterize light uptake as a metabolic flux. To close this knowledge gap, we expanded our modeling efforts to a model cyanobacterial phototroph. The intent was to validate the modeling methodology in the prokaryotic system before applying it to the more complicated eukaryotic diatom. We developed a procedure to quantify light absorption based solely on incident light and cellular pigment composition. Combined with a parameter on maximum photosynthetic output, we accurately predicted the linear growth curve common to algal batch cultures. When applied to the model phototroph *Synechococcus elongatus*, these parameters yielded unique insights into obligate phototrophic metabolism and were presented in a peer-reviewed publication (doi:10.1073/pnas.1613446113). We have since applied this methodology to diatom metabolism with similar success. This manuscript is currently under review. Establishing the governing constraints on light-driven metabolism is a major contribution to systems biology of photoautotrophy. |
| 4.2 Model-based 'omics integration and analysis | **Actual Performance:** 80% completed.<br>We integrated diatom transcriptomics data along with protein-interaction data by clustering algorithms which resulted in a hierarchal diatom transcriptional regulatory network. This work has been presented in a peer-reviewed publication (doi:10.1128/mSystems.00142-16). 100% completed.<br>Additionally, we developed and presented in peer-reviewed publications methods that integrate metabolomics data and, in turn, predict biomarkers (doi:10.1371/journal.pcbi.1005424) and overall system stability (doi:10.1074/jbc.M117.804914). These efforts led to the generation of black-box models that can predict metabolic system dynamics presented in a peer-reviewed conference proceeding (doi:10.1109/CCTA.2017.8062584). Efforts to apply these methodologies to phototrophs is currently underway. However, since this method has yet to be applied to a phototrophic dataset it is only 60% complete. |
| 4.3 Metabolic engineering of light-driven metabolism | **Actual Performance:** 80% completed.<br>Generating actionable metabolic engineering strategies requires an understanding of the reference metabolic flux state. As this was not yet available for diatoms, we used the model phototroph *Synechococcus elongatus* PCC 7942 as a case study. Leveraging genome-scale modeling and photophysiology parameters, we accurately predicted quantitative intracellular fluxes for photoautotrophy at both low and high light. Then, this model was used to design and optimize a heterologous pathway for 2,3-butanediol production. This framework contextualized existing designs and suggested improvements. This study paved the way for successful *in silico* engineering of phototrophs and was presented in a peer-reviewed publication (doi:10.1016/j.ymben.2018.11.001).<br>However, since this method has yet to be applied to a diatom dataset it is only 80% complete. |

## Project Activities

Diatoms are important oceanic photosynthetic microorganisms, responsible for approximately 20% of all carbon fixation on Earth. Additionally, diatoms are attractive metabolic engineering candidates. This investigation leveraged Systems Biology techniques to advance our understanding of and potential to engineer diatoms for bioproducts of societal interest, such as biofuels. We generated, validated, and published a high-quality metabolic model of diatom metabolism facilitating novel insights into these unique organisms. Additionally, we overcame a consistent challenge in the modeling of photosynthetic organisms by integrating light into metabolic models. This result allows realistic assessments of how the light environment affects cell physiology and metabolism, a necessary advancement for bioengineering. Next, we generated several methodologies for integrating large-scale datasets with models of metabolism. These "big data" approaches begin to address a constant challenge in biology, which is translating data into information and understanding. Finally, we built web-based databases for housing, distributing, and visualizing systems-level biology data. These projects created an online ecosystem for sharing, analyzing, and characterizing metabolic information and is currently being used by public and private institutions. This project and its outputs have accelerated the application of proven Systems Biology techniques to light-driven metabolism.

## Project Outputs

### *A. Journal Articles*


1. King ZA, Dräger A, Ebrahim A, Sonnenschein N, Lewis NE, Palsson BO (2015) Escher: A Web Application for Building, Sharing, and Embedding Data-Rich Visualizations of Biological Pathways. PLoS Comput Biol 11(8): e1004321. https://doi.org/10.1371/journal.pcbi.1004321

2. King, Z.A., Lu, J., Dräger, A., Miller, P., Federowicz, S., Lerman, J.A., Ebrahim, A., Palsson, B.O., Lewis, N.E., Drager, A., Miller, P., Federowicz, S., Lerman, J.A., Ebrahim, A., Palsson, B.O., Lewis, N.E., 2016. BiGG Models: A platform for integrating, standardizing and sharing genome-scale models. Nucleic Acids Res 44, D515-22. https://doi.org/10.1093/nar/gkv1049

3. Levering, J., Broddrick, J., Zengler, K., 2015. Engineering of oleaginous organisms for lipid production. Curr. Opin. Biotechnol. 36, 32–39. https://doi.org/http://dx.doi.org/10.1016/j.copbio.2015.08.001

4. Levering, J., Broddrick, J., Dupont, C.L., Peers, G., Beeri, K., Mayers, J., Gallina, A.A., Allen, A.E., Palsson, B.O., Zengler, K., 2016. Genome-scale model reveals metabolic basis of biomass partitioning in a model diatom. PLoS One 11, e0155038. https://doi.org/10.1371/journal.pone.0155038

5. Bordbar, A., Yurkovich, J.T., Paglia, G., Rolfsson, O., Sigurjónsson, Ó.E., Palsson, B.O., 2017. Elucidating dynamic metabolic physiology through network integration of quantitative time-course metabolomics. Sci. Rep. 7, 46249. https://doi.org/10.1038/srep46249

6. Yang, L., Yurkovich, J.T., Lloyd, C.J., Ebrahim, A., Saunders, M.A., Palsson, B.O., 2016. Principles of proteome allocation are revealed using proteomic data and genome-scale models. Sci. Rep. 6, 36734. https://doi.org/10.1038/srep36734

7. Broddrick, J.T., Rubin, B.E., Welkie, D.G., Du, N., Mih, N., Diamond, S., Lee, J.J., Golden, S.S., Palsson, B.O., 2016. Unique attributes of cyanobacterial metabolism revealed by improved genome-scale metabolic modeling and essential gene analysis. Proc. Natl. Acad. Sci. 113, E8344–E8353. https://doi.org/10.1073/pnas.1613446113

8. Gallina, A.A., Layer, M., King, Z., Levering, J., Palsson, B.O., Zengler, K., Peers, G. (2016). A Phaeodactylum tricornutum literature database for interactive annotation of content. Algal Research. 18. 241-243. 10.1016/j.algal.2016.06.020.

9. Yurkovich JT, Yang L, Palsson BO (2017) Biomarkers are used to predict quantitative metabolite concentration profiles in human red blood cells. PLoS Comput Biol 13(3): e1005424. https://doi.org/10.1371/journal.pcbi.1005424

10. Yurkovich, J.T., Zielinski, D.C., Yang, L., Paglia, G., Rolfsson, O., Sigurjónsson, Ó.E., Broddrick, J.T., Bordbar, A., Wichuk, X.K., Brynjólfsson, S., Palsson, S., Gudmundsson, S., Palsson, B.O., 2017. Quantitative time-course metabolomics in human red blood cells reveal the temperature dependence of human metabolic networks. J. Biol. Chem. 292, jbc.M117.804914. https://doi.org/10.1074/jbc.M117.804914

11. Yurkovich, J.T., Yang, L., Palsson, B.O., 2017. Utilizing biomarkers to forecast quantitative metabolite concentration profiles in human red blood cells, in: 2017 IEEE Conference on Control Technology and Applications (CCTA). IEEE, pp. 961–966. https://doi.org/10.1109/CCTA.2017.8062584

12. Yurkovich, J.T., Yurkovich, B.J., Dräger, A., Palsson, B.O., King, Z.A., 2017. A Padawan Programmer's Guide to Developing Software Libraries. Cell Syst. 5, 431–437. https://doi.org/10.1016/J.CELS.2017.08.003

13. Levering, J., Dupont, C.L., Allen, A.E., Palsson, B.O., Zengler, K., 2017. Integrated Regulatory and Metabolic Networks of the Marine Diatom Phaeodactylum tricornutum Predict the Response to Rising CO2 Levels. mSystems 2, e00142-16. https://doi.org/10.1128/mSystems.00142-16

14. Fang, X., Sastry, A., Mih, N., Kim, D., Tan, J., Yurkovich, J.T., Lloyd, C.J., Gao, Y., Yang, L., Palsson, B.O., 2017. Global transcriptional regulatory network for Escherichia coli robustly connects gene expression to transcription factor activities. Proc. Natl. Acad. Sci. U. S. A. 114, 10286–10291. https://doi.org/10.1073/pnas.1702581114

15. Broddrick, J.T., Welkie, D.G., Jallet, D., Golden, S.S., Peers, G., Palsson, B.O., 2018. Predicting the metabolic capabilities of Synechococcus elongatus PCC 7942 adapted to different light regimes. Metab. Eng. 52, 42–56. https://doi.org/10.1016/J.YMBEN.2018.11.001

### *B. Papers*
N/A

### *C. Status Reports*
Department of Energy, Office of Biological and Environmental Research RPT-0000001771

### *D. Media Reports*
N/A

### *E. Invention Disclosures*
N/A

### *F. Patent Applications*
N/A

### *G. Licensed Technologies*
N/A

### *H. Networks/Collaborations Fostered*
Food and Fuel for the 21st century, UC San Diego, La Jolla, CA.
California Center for Algae Biotechnology, La Jolla, CA.

Collaborations with the following Principal Investigators:
Dr. Andrew E. Allen, JCVI, La Jolla, CA.
Dr. Chris Dupont, JCVI, La Jolla, CA.
Dr. Susan S. Golden, UC San Diego, La Jolla, CA.
Dr. Graham Peers, Colorado State University, Fort Collins, CO.
Dr. Karsten Zengler, UC San Diego, La Jolla, CA.
Dr. Nathan E. Lewis, UC San Diego, La Jolla, CA.
Dr. Yusuke Matsuda, Kansai Gakuin University, Sanda, Hyogo, Japan.

### *I. Websites Featuring Project Work Results*
N/A

### *J. Other Products (e.g. Databases, Physical Collections, Audio/Video, Software, Models, Educational Aids or Curricula, Equipment or Instruments)*

https://escher.github.io
https://bigg.ucsd.edu
Genome-scale model of *P. tricornutum* (*i*LB1025) https://doi.org/10.1371/journal.pone.0155038
Genome-scale model of *S. elongatus* PCC7942 (*i*JB785) https://doi.org/10.1073/pnas.1613446113

### *K. Awards, Prizes, and Recognition*

N/A

## Follow-On Funding

N/A