Cooperative Research and Development Agreement (CRADA) Final Report

Report Date:
March 28, 2024

In accordance with Requirements set forth in the terms of the CRADA, this document is the CRADA Final Report, including a list of Subject Inventions. It is to be forwarded to the DOE Office of Scientific and Technical Information upon completion or termination of the CRADA, as part of the commitment to the public to demonstrate results of federally funded research.

Parties to the Agreement: UC San Diego, PNNL and Lawrence Berkeley National Laboratory

CRADA number: FP00008811

CRADA Title: Advanced Algal Biofoundries for the Production of Polyurethane Precursors

Responsible Technical Contact at Berkeley Lab: Nathan Hillson

Name and Email Address of POC at Partner Company(ies):
Stephen Mayfield, UCSD (retired) smayfield@ucsd.edu
Ryan Simkovsky, UCSD rsimkovsky@ucsd.edu
Jeremy Zucker, PNNL jeremy.zucker@pnnl.gov

Sponsoring DOE Program Office(s):
DOE EERE BETO Conversion

LBNL Report Number:
LBNL-2001581

OSTI Number:
[SPO to complete]

Joint Work Statement Funding Table showing DOE funding commitment:

DOE Funding to LBNL	$185,750
DOE Funding to PNNL	$185,750
Participant Funding to LBNL	-
Participant In-Kind Contribution Value	$2,198,500
Total of all Contributions	$2,570,000

Provide a list of publications, conference papers, or other public releases of results, developed under this CRADA:
(Publications must include journal name, volume, issue, Digital Object Identifier)

Y. Torres-Tiji, H. Sethuram, A. Gupta, J. McCauley, J.V. Dutra-Molino, R. Pathania, L. Saxton, K. Kang, N. J. Hillson, S. P. Mayfield. (2024) “Bioinformatic prediction and high throughput in vivo screening to identify cis-regulatory elements for the development of algal synthetic promoters”. bioRxiv 2024.03.20.585996; doi: https://doi.org/10.1101/2024.03.20.585996

The above manuscript was also submitted or publication (pending) at ACS Synthetic Biology.

Provide a detailed list of all subject inventions, to include patent applications, copyrights, and trademarks:
(Patents and patent applications are to include the title and inventor(s) names. When copyright is asserted, the Government license should be included on the cover page of the Final Report)

None.

Executive Summary of CRADA Work:

[Excerpted verbatim from the “Abstract / Executive Summary” section of the attached project final report].

The primary goal of the BEEPs project was to develop a process that could accelerate the development of algae as bioproduction platforms, from initial chemical product concept to an economically viable market supply. Under this program we elected to develop strains of algae that could generate polyurethane precursors, while simultaneously developing basic genetic tools to enable improved algal production systems. This program was specifically designed to incorporate National Laboratories as a means to utilize the expertise and facilities for new bio-production platforms. To that end, we designed a program to collaborate with the Agile BioFoundry at Lawrence Berkeley National Laboratory (LBNL) and computational platforms at Pacific Northwest National Laboratory (PNNL). In addition to these National Laboratory partners, we also had academic partners from UC Davis and Georgia Tech, as well as commercial partners Algenesis Materials and BASF. To achieve these goals, we initially focused on developing the genetic tools and high throughput screening technologies necessary to generate and assess production of polymer precursors (succinic acid) in algae and cyanobacteria, including advanced promoters and biosensors. In parallel, we computationally identified potential production bottlenecks and then used the developed genetic tools to increase production rates and yields. Constant feedback of data was used in conjunction with machine learning, high-throughput cell sorting, and synthetic biology, for additional targeted metabolic engineering. Multiple rounds of tool design, building, testing, and learning were supplied to partners at PNNL and LBNL to develop new models and tools that could expedite bioproduction platform development and increase yield performance. We targeted, and achieved, the FOA requirement yield metric of 20 g/L, as a milestone and deliverable from at least one of our engineered strains for the production of succinic acid.

Summary of Research Results:

[Excerpted verbatim from the "Overall Project Performance" section of the attached project final report].

Overall, the project met its main goals of developing algae strains that had high levels of succinic acid production, actually reaching 34 g/L, well above the FOA recommended target. In addition, we were able to generate a number of new genetic tools that are now available for the algae community, and developed a system to identify new synthetic promoters in any algae strain, which appears to work well. The computational methods from PNNL for the metabolic models of succinic acid production cyanobacteria and algae were already developed before the project started, and we were able to quickly add our specific species to those models and generate predictions for gene knockouts and overexpression. The majority of those predictions turned out to be of little help, but a few worked well and increased SA production. The computational analysis and machine learning for the development of synthetic promoters work at a functional level, meaning we were able to generate new synthetic promoters that worked, but failed at the theoretical level, in that we were unable to generate a set of rules that might allow us to *de novo* design new synthetic promoters. The Agile BioFoundry and our partners at LBNL were able to consistently deliver DNA constructs on time and on budget, which allowed us to accelerate our timelines to include more than the intended two rounds of promoter design and testing. The other academic partners had mixed results: the UC Davis group delivered on developing strains of cyanobacteria producing SA at ever increasing levels and developed several new protocols for their growth that also added to the overall success of the project. The Georgia Tech group was unable to deliver any bio-sensors that could detect SA or other target chemicals at a meaningful level. This was partly due to the high-risk nature of that part of the project (biosensors), and partly due to the COVID-19 pandemic, which left the Georgia Tech group unable to hire new students or techs for almost 2 years of the three-year program.
In addition to the pandemic, we also faced very long delays due to the CRADA process required by the National Laboratories, before we could even start any discussions; let alone work. It took almost 9 months to sign the CRADA between UCSD and PNNL, a process which UCSD found to be both cumbersome and much too slow. Cumbersome because PNNL put in provisions that were either meaningless to the project we were pursuing, or were so one sided that no one, including UCSD, could sign. As an example, PNNL put in a provision that said all IP would be owned by PNNL, even though there were 3 academic partners and another national lab, LBNL, doing the majority of the work, and innovation, on the project. In addition, each round of negotiation took weeks or even months to get a response back from the National Labs. To be fair, UCSD was equally slow on their response time, but UCSD never put in provisions that required further negotiations. In conclusion, I believe the basic idea behind BEEPS – to get the National Labs more engaged with the larger research community, so that all of the US can benefit from their expertise – is a good idea. The secondary idea of getting Universities and companies access to the Agile BioFoundery is also a really good idea – it's a national resource that more universities and companies should be able to use, and the DOE should subsidize that use – it will help propel the bioeconomy in this country! However, the National Lab CRADA process is a national disgrace and needs to be completely reinvented, period; there is no place for arcane processes like that in the highly collaborative scientific world of the 21st century.

APPENDIX A (Reference Only)

*This appendix has been developed by DOE to assist DOE Labs in drafting the **Executive Summary** and **Summary of Research Results** sections of the CRADA Final Report.*

Executive Summary of CRADA Work:

Include a discussion of 1) how the research adds to the understanding of the area investigated; 2) the technical -effectiveness of the materials, methods or techniques investigated or demonstrated, and their economic feasibility, if known; and 3) how the project is otherwise of benefit to the public. The discussion should be a minimum of one paragraph and written in terms understandable by an educated layman.

Summary of Research Results:

- *INCLUDE, IF APPLICABLE: "This product contains Protected CRADA Information, which was produced on [DATE] under CRADA No. [##-####] and is not to be further disclosed for a period of [up to and not to exceed] five (5) years from the date it was produced except as expressly provided for in the CRADA."*
- *Summarize project activities for the entire period of performance, including original hypotheses, approaches used, problems encountered, any departure from planned methodology, and an assessment of their impact on the project results. Incorporate technical data, e.g. facts, figures, analyses, and assumptions used during the life of the project to support the technical conclusions of the work. It is acceptable to incorporate the technical data by reference to other publicly available sources, such as a publications or other reports, but not websites. Provide a comparison of the actual accomplishments with the goals and objectives of the project. Where possible, the summary should cover each task listed in the Statement of Work (SOW) and should note any deviations from the project plan, or lack of technical data.*
- *Identify products, potential applications, and technology transfer activities developed under the CRADA, including those completed and anticipated at the time of the report. These include, but are not limited to: 1) networks or collaborations fostered; 2) technologies/techniques/methodologies; 3) other products that reflect the results of the project, such as commercial products, internet sites, data or databases, physical collections, audio or video, software, models, educational aid or curricula, and instruments or equipment.*

Note: Recommended characteristics of Scientific and Technical Information reports can be found at https://www.osti.gov/stip/attributes

FINAL TECHNICAL REPORT

Department Of Energy

Office of Energy Efficiency & Renewable Energy

&

Bioenergy Technologies Office

EERE Award Number: DE-EE0008491

Name of Recipient: University of California, San Diego

Project Title: Advanced Algal Biofoundries for the Production of Polyurethane Precursors

Principal Investigator: Stephen Mayfield, PhD

Support period: 10/01/2018 - 12/31/2022

Submitted: November 22, 2023

Abstract / Executive Summary

The primary goal of the BEEPs project was to develop a process that could accelerate the development of algae as bioproduction platforms, from initial chemical product concept to an economically viable market supply. Under this program we elected to develop strains of algae that could generate polyurethane precursors, while simultaneously developing basic genetic tools to enable improved algal production systems.

This program was specifically designed to incorporate National Laboratories as a means to utilize the expertise and facilities for new bio-production platforms. To that end, we designed a program to collaborate with the Agile BioFoundry at Lawrence Berkeley National Laboratory (LBNL) and computational platforms at Pacific Northwest National Laboratory (PNNL). In addition to these National Laboratory partners, we also had academic partners from UC Davis and Georgia Tech, as well as commercial partners Algenesis Materials and BASF. To achieve these goals, we initially focused on developing the genetic tools and high throughput screening technologies necessary to generate and assess production of polymer precursors (succinic acid) in algae and cyanobacteria, including advanced promoters and biosensors. In parallel, we computationally identified potential production bottlenecks and then used the developed genetic tools to increase production rates and yields. Constant feedback of data was used in conjunction with machine learning, high-throughput cell sorting, and synthetic biology, for additional targeted metabolic engineering. Multiple rounds of tool design, building, testing, and learning were supplied to partners at PNNL and LBNL to develop new models and tools that could expedite bioproduction platform development and increase yield performance. We targeted, and achieved, the FOA requirement yield metric of 20 g/L, as a milestone and deliverable from at least one of our engineered strains for the production of succinic acid.

Overall Project Performance

Overall, the project met its main goals of developing algae strains that had high levels of succinic acid production, actually reaching 34 g/L, well above the FOA recommended target. In addition, we were able to generate a number of new genetic tools that are now available for the algae community, and developed a system to identify new synthetic promoters in any algae strain, which appears to work well. The computational methods from PNNL for the metabolic models of succinic acid production cyanobacteria and algae were already developed before the project started, and we were able to quickly add our specific species to those models and generate predictions for gene knockouts and overexpression. The majority of those predictions turned out to be of little help, but a few worked well and increased SA production. The computational analysis and machine learning for the development of synthetic promoters work at a functional level, meaning we were able to generate new synthetic promoters that worked, but failed at the theoretical level, in that we were unable to generate a set of rules that might allow us to *de novo* design new synthetic promoters. The Agile BioFoundry and our partners at LBNL were able to consistently deliver DNA constructs on time and on budget, which allowed us to accelerate our timelines to include more than the intended two rounds of promoter design and testing. The other academic partners had mixed results: the UC Davis group delivered on developing strains of cyanobacteria producing SA at ever increasing levels and developed several new protocols for their growth that also added to the overall success of the project. The Georgia Tech group was unable to deliver any bio-sensors that could detect SA or other target chemicals at a meaningful level. This was partly due to the high-risk nature of that part of the project (biosensors), and partly due to the COVID-19 pandemic, which left the Georgia Tech group unable to hire new students or techs for almost 2 years of the three-year program.

In addition to the pandemic, we also faced very long delays due to the CRADA process required by the National Laboratories, before we could even start any discussions; let alone work. It took almost 9 months to sign the CRADA between UCSD and PNNL, a process which UCSD found to be both cumbersome and much too slow. Cumbersome because PNNL put in provisions that were either meaningless to the project we were pursuing, or were so one sided that no one, including UCSD, could sign. As an example, PNNL put in a provision that said all IP would be owned by PNNL, even though there were 3 academic partners and another national lab, LBNL, doing the majority of the work, and innovation, on the project. In addition, each round of negotiation took weeks or even months to get a response back from the National Labs. To be fair, UCSD was equally slow on their response time, but UCSD never put in provisions that required further negotiations.

In conclusion, I believe the basic idea behind BEEPS – to get the National Labs more engaged with the larger research community, so that all of the US can benefit from their expertise – is a good idea. The secondary idea of getting Universities and companies access to the Agile BioFoundery is also a really good idea – it's a national resource that more universities and companies should be able to use, and the DOE should subsidize that use – it will help propel the bioeconomy in this country! However, the National Lab CRADA process is a national disgrace and needs to be completely reinvented, period; there is no place for arcane processes like that in the highly collaborative scientific world of the 21st century.

Acknowledgement

This material is based upon work supported by the U.S. Department of Energy's Office of Energy Efficiency and Renewable Energy (EERE) under the Bioenergy Technologies Office Award Number DE-EE0008491.

Table of Contents

Overview of Tasks and Primary Goals

Task	Title	Key Objective
1	Process and Data Validation	Validate key performance parameters (KPP) and quantitative metrics for production of each target molecule.
2	Generate Baseline Algal Production Platforms	Produce initial polyurethane precursor (PUP) production strains in a green alga and a cyanobacterium.
3	Design and Construct Advanced Synthetic Promoters - Round 1	Design and synthesize a first round of synthetic promoter libraries for expression of a fluorescent reporter gene.
4	Screening Advanced Promoters - Round 1	Generate, screen, and assess first round of expression libraries in *Chlamydomonas* using a fluorescent reporter gene.
5	Biosensors for Target Molecules	Design, generate, and optimize biosensors for 1+ target molecules.
6	Metabolic Characterization & Optimization of Production Platforms	Develop omics-based metabolic models and screening tools to predict and generate optimized production strains.
7	RB-TnSeq Mutagenesis Analysis & Optimization of Synechococcus Production Platform	Generate and analyze bar-coded mutant libraries to improve production strains.
8	Constructing Advanced Synthetic Promoters - Round 2	Design and synthesize a second round of synthetic promoter libraries for expression of a fluorescent reporter gene and/or production pathway expression.
9	Screening Advanced Promoters - Round 2	Generate, screen, and assess second round of expression libraries in *Chlamydomonas*.
10	Integration of Optimizations and Generation of Most Promising Production Strains	Generate optimized production strains.
11	Scaled Production Platform Yield Testing	Evaluate production strains at scaled growth under varying conditions.
12	TEA/LCA Modeling of Algal Bioproduction	Evaluate TEA/LCA of production strains comparing varying growth conditions.

Project Objectives and Deliverables

Objective 1: Develop polyurethane precursor production platforms in green algae and cyanobacteria.
Deliverable 1: Synthetic biology tools for production of at least one polymer precursor in green algae and/or cyanobacteria and the resulting modified strain(s).
Objective 2: Develop biosensors capable of detecting polymer precursors from algal production platforms.
Deliverable 2: Biosensor system(s) and quantitative characterization of biosensors detection capabilities.
Objective 3: Establish high-throughput screening methods using biosensors and/or mass spectrometry to rapidly evaluate polymer precursor productions strains.
Deliverable 3: Methods for rapidly identifying improved production strains from genetically diverse pools.
Objective 4: Develop a metabolic knowledge base through metabolic modeling, -omics studies, and RB-TnSeq to identify metabolic bottlenecks in production strains
Deliverable 4: Metabolic models, mutagenesis methods, large data sets, and machine learning methods to identify pathways for metabolic engineering for bioproduction optimization.
Objective 5: Improve upon synthetic promoter libraries for increased expression of enzymes and improved product yields using multiple rounds of design, build, test/screen, and learn/model steps.
Deliverable 5: Validated synthetic promoter libraries for advanced production of polymer production in algae.
Objective 6: Integrate knowledge gained from prior objectives into most promising production strains.
Deliverable 6: Most promising algal production strains based on previous results and objectives.
Objective 7: Evaluate yields of polyurethane precursors from most promising production strains under heterotrophic production conditions on cellulosic sugars and photosynthetic production.
Deliverable 7: Productivities associated with production strains under industrially relevant growth conditions.
Objective 8: Conduct economic analyses using data sets obtained from Objective 7.
Deliverable 8: TEA/LCA reports comparing the economic costs of heterotrophic and photosynthetic production.

Table of Abbreviations

Acronym	Meaning
AA	Azelaic Acid
AFDW	Ash Free Dry Weight
BDO or 1,4-BDO	1,4-butanediol
Ble^{R}	Bleomycin resistance
CTS	Chloroplast Transit Sequence
DIC	Differential Interference Contrast
DOE	Department of Energy
ER	Endoplasmic Reticulum
FACS	Fluorescently Activated Cell Sorting
FOA	Funding Opportunity Announcement
FP	Fusion Peptide
GCMS	Gas Chromatography Mass Spectrometry
GFP	Green Fluorescence Protein
GM	Genetically Modified
GPCR	G protein-coupled receptor
HR	Homology region
HSM	High Salt Media
KPP	Key Performance Parameters
LCA	Life-Cycle Assessment
LCI	Life cycle inventory
MCFA	Medium-chain fatty acids
MSP	Minimum sustainable price
NMR	Nuclear Magnetic Resonance
NTD	N-terminal Domain
OR	Olfactory Receptor
ORP	Open-raceway pond
PDO or 1,3-PDO	1,3-propanediol
PUP	Polyurethane Precursor
RBD	Receptor Binding Domain
RDB	Renewable diesel blendstock
sfGFP	Super folder GFP
SP	Synthetic Promoter
TAP	Tris-Acetate-Phosphate
TEA	Techno-Economic Analysis

Task 1.0: Process and Data Validation

Key performance parameters (KPPs), protocols, and other relevant metrics for performing, evaluating, and validating the awarded project DE-EE0008491 were established based upon current literature or prior laboratory experience by the participating groups. These were presented to and discussed with the site validation team during the on-site visit on December 4 – 5, 2019. As requested by the validation team, protocol demonstrations, including detection of succinate via HPLC and LC-MS, scaled photosynthetic growth of *S. elongatus* at the 20-L scale, and scaled heterotrophic growth of both *Chlamydomonas reinhardtii* and a genetically engineered strain of *S. elongatus* at the 6 and 2-L scales, respectively, were prepared and performed for the validation team. Based upon feedback from the validation team, an updated Technical Data Form was provided to the validation team that included an expanded verification table that included current and target KPPs for each potential production strain proposed for the project. The on-site visit and subsequent discussions resulted in the initial validation report.

Milestone 1.ML.1: Validation of baseline production values in published platforms. With DOE site visit team, validate and update parameters in the technical data sheet and provide documentation for a validation report.

Baseline production values were confirmed with the on-site validation team and updates to the technical data sheet were provided for the production of the validation report.

Go/No-Go Decision Point GN.1 Validation: If parameter validation fails or site visit team finds that the project is not feasible and these obstacles are found to be insurmountable, the project will be terminated.

Based upon the validation report, the program manager submitted award documents to approve advancing the project from Budget Phase 1 to Budget Phase 2.

Task 2.0: Generate Baseline Algal Production Platforms

Photosynthetic production of succinate from *Synechococcus elongatus* PCC 7942: The objective of this subtask is to establish baseline levels of succinate production from *S. elongatus* under photosynthetic conditions. It has been previously shown by Lan & Wei that an engineered *S. elongatus* strain overexpressing the *gabD* and *kgd* genes can produce ~50 mg/L of succinate in the media while a strain overexpressing *gabD, kgd*, *ppc*, and *gltA* genes can produce ~400 mg/L succinate. *gabD* encodes a succinate semialdehyde dehydrogenase, while *kgd* encodes an alpha-ketoglutarate decarboxylase. Together, these genes are sufficient to generate succinate in the media under phototrophic conditions. *ppc* encodes a phosphoenolpyruvate carboxylase and *gltA* encodes a citrate synthase. These additional two genes are believed to enhance the carbon flow into the TCA cycle where succinate is produced. To verify and reproduce these production values, the succinate producing strain LAN1 was acquired from Lan & Wei and used to verify our HPLC and LC-MS methods for detecting and quantifying succinate production. Quantification of succinate production from these strains at the tube, flask, and 20-L carboy scale have all confirmed the production of succinate from these strains, and have shown that after three days of induction the amount of succinate produced is on average 50 mg/L, consistent with the published values. We also observed several peaks eluting separately from the succinate peak, indicating a number of by-products that will need to be identified and removed in order to enhance succinate yields. In addition, because we wished to generate a succinate production system sufficiently distinct from that of Lan & Wei, vectors have been produced that encode and inducibly express *gabD* and *kgd* genes from the original *S. sp.* PCC 7002 coding sequences, a vector that we have dubbed pSuc2.

This vector has been transformed into our WT-strain in order to check strain background impacts. We know from experiments using a fumarate knockout described in Subtask 6.5 that there are background differences between our WT and the one sent to us by Lan & Wei. In the experiments reported for Subtask 6.5, our WT carrying the fumarate knockout cannot be easily transformed with the succinate production vector because it appears that this combination causes a synthetic lethality that is not observed with the Lan & Wei strain. Transformations with pSuc2 into our WT *S. elongatus* PCC 7942 strain demonstrates that the vector using the native PCC 7002 encodings of the genes, as compared to the original vector used to generate LAN1, reproducibly results in fewer transformants that also appear to generally have a growth defect. This phenotypic evidence leads us to believe that the basal leakiness of this vector is sufficient to drive production of succinate over wild type levels and reproduce the metabolic growth defect reported by Lan & Wei for LAN1 when that strain is induced for expression. This interpretation would also indicate that these gene encodings are more productive for succinate than those originally used by Lan & Wei, as the vector backbone and transcriptional and translational regulatory elements are identical in the two vectors. To test this hypothesis, we performed a preliminary HPLC analysis on the supernatants from an induced LAN1 strain and an induced WT+pSuc2 strain. Our preliminary analysis, however, demonstrated that LAN1 still produced more succinate in the supernatant than WT+pSuc2. This result could be for several reasons as multiple differences exist between these two strains: 1) The background genotype of LAN1 may be different from our WT strain through key mutations that allow succinate production and secretion, 2) the 7002 version of gabD is not as productive as that from E. coli, or 3) the codon optimization of 7002 kgd is necessary for full functional expression of the gene in 7942. Further investigations on this phenomenon are reported under Task 6.0.

Heterotrophic or photomixotrophic production of succinate from *S. elongatus* PCC 7942: The UC Davis lab's goal for this project is to generate a strain of *Synechococcus elongatus* PCC 7942 capable of producing succinate photomixotrophically. It has been previously shown

that an engineered *S. elongatus strain* overexpressing the *gabD*, *kgd*, *ppc*, and *gltA* genes can produce ~400 mg/L succinate photoautotrophically. First, we constructed a plasmid containing *gabD* and *kgd* (pAL1847). pAL1847 was then used as a template to assemble a plasmid containing all four genes (pAL1832). Both plasmids target neutral site three (NSIII). The initial plasmid, pAL1847 (NSIII-lacIq Ptrc:gabD-kgd GentR-NSIII'), has been assembled using single ligation independent-cloning (SLIC) in a strain of *E. coli,* XL-1 Blue, and introduced into *S. elongatus*. We have confirmed the successful integration of the fragment (NSIII-lacIq Ptrc:gabD-kgd GentR-NSIII') into the genome. For this vector, the gabD and kgd genes were PCR amplified from the genome of *Synechococcus sp.* strain PCC 7002. The vector fragment for pAL1847 was PCR amplified from a previous plasmid, pAL1136. Sanger sequencing was used to confirm sequence of pAL1847 prior to its transformation into *S. elongatus* PCC 7942. An alternative plasmid (NSIII-*lacI*^{q} Ptrc:*gabD-kgd-ppc-gltA* Gent^{R}-NSIII') that contains two additional genes; *ppc* and *gltA* (pAL1830) was also generated using SLIC and confirmed via Sanger sequencing. The *ppc* gene was PCR amplified from the genome of *S. elongatus*. The *gltA* gene was amplified from the genome of *Escherichia coli* strain MG1655. As confirmed at UCSD, these strains produced and secreted succinate.

In addition to these heterologous expression vectors, several upstream modifications were made to the succinate strain so that the strain could produce succinate photomixotrophically. For the glucose strain, we used the modifications that were used to improve photomixtrophic 2,3-butanediol production (doi: 10.1038/ncomms14724). One modification installs a sugar transporter gene, *galP* from *Escherichia coli* and necessary to import glucose into *S. elongatus*. Further modifications involve the overexpression of *gnd* and *zwf*, which encode a 6-phosphogluconate dehydrogenase and a glucose-6-phosphate dehydrogenase, respectively. These two enzymes are key targets for redirecting the carbon flux from glucose to the oxidative pentose phosphate pathway. The final modification involves a regulatory protein CP12 encoded by the *cp12* gene which will be deleted and replaced with the *prk* and *rbcLS* expression cassette to further improve carbon dioxide fixation. For the xylose strain, the *xylE* gene encoding a xylose transporter along with the *xylA* and *xylB* genes encoding D-xylose isomerase and xylulokinase, respectivey (xylose pathway), were introduced into the succinate strain (DOI: 10.1128/AEM.03326-12). Since *S. elongatus* does not harbor the *xylEAB* genes, the integration of the *xylEAB* genes from *E. coli* is essential to utilize xylose. A final regulatory modification was tested as part of Task 6.0 by knocking out a key circadian clock gene, *cikA*. CikA is responsible for phosphatase activity on RpaA, a key regulator of circadian transcriptional activity. By removing this phosphatase activity, we decouple chemical production from the day-night cycle, thereby enhancing the window of production activity to include nighttime.

The UC Davis group evaluated succinate production in their base strain and achieved modest titers of succinate when adjusting to a pH of 7. This base strain differs from that evaluated by the UCSD cyanobacterial lab in that the production pathway was inserted into Neutral Site III, rather than Neutral Site I, and the order of the the *gltA* and *ppc* genes is reversed compared to the UCSD lab strain, with *gltA* being encoded last of the four base strain genes and *ppc* being encoded third.

Succinate as a primary target molecule to be produced in *Chlamydomonas reinhardtii*: It has been reported that *Chlamydomonas* excretes succinate under sulfur deplete, nitrogen deplete, or anaerobic conditions. Because natural production and excretion of succinate represents a baseline production strain for this project, we are focusing on investigating the natural secretion of succinate from *Chlamydomonas* under nitrogen-deplete versus nitrogen-replete conditions. For this experiment, starch-deficient strains that show the most promise for generating succinate (CC-5372, CC-5374, and CC-5129) and reference wild-type strains (CC-

1690 and CC-124) were put under nitrogen-deplete or nitrogen-replete conditions for five days following a three-day initial growth and washing into the appropriate media. Supernatants were collected for HPLC-based detection and quantification of succinate to determine extracellular succinate yields. The three starchless mutants produced extracellular succinate in the 150 – 175 mg/L range, as compared to the WT strains yields in the single digit mg/L range.

Milestone 2.ML.1: Baseline production vectors constructed; sequence verified in *E. coli.* 1+ vectors destined for *Chlamydomonas* and 1+ for *Synechococcus* to engineer precursor production.

As described above, baseline production vectors and strains for succinate production have been generated in *E. coli* and sequence verified.

Generating decanedioic acid and decanediol in *Chlamydomonas reinhardtii*: The biochemistry lab at UCSD pursued the strategy of generating DDA and DDO through the heterologous expression of a medium-chain thioesterase (TE) that can interact with the native acyl-carrier protein (ACP) in *Chlamydomonas* that is responsible for Type II fatty acid synthesis in order to release the growing fatty acid chain at a shorter chain length than is typical in the organism. In order to improve yields, we are seeking to mutate the protein binding interface of the TE to improve interactions with the native ACP. To determine the amino acids contributing to this protein-protein interface and predict desired mutations, investigations were performed to determine the solution state NMR structure of the *C. reinhardtii* ACP (CrACP) or to produce homology models that will enable protein-protein docking simulations. These could then enable designed protocols and strategies for the production and chemoenzymatic modification of the *Chlamydomonas reinhardtii* ACP (CrACP) from expression constructs in *E. coli*. We have optimized the growth of the enzyme and established protocols to generate single “apo” species using purified thioesterase to hydrolyze acylated carrier proteins and ACP hydrolase (ACPH) to unload the phosphopantetheine cofactor. This apo ACP can be modified chemoenzymatically with crosslinkers and substrate mimics for study. As well, we have optimized the purification of the *C. reinhardtii* thioesterase (CrTE), we emphasized the need for a protocol that can yield pure samples for analysis in a single Nickel affinity purification step straight from grown pellets. This protocol will be essential in the future for screening the activity of designed mutants, allowing easily parallelized purifications with no need for relatively slow size exclusion purification.

After gaining confidence in protocols for the high yield growth of CrACP, we grew uniformly labeled C^{13} and N^{15} CrACP. This was purified and prepared chemoenzymatically using *Bacillus subtilis* SFP, a phosphopantetheinyl transferase, to generate a single population of the fully holo-form of the protein for 3-dimensional NMR experiments that are underway now at the UCSD Biomolecular NMR facility using a Bruker Avance 600 Mhz NMR to determine the solution state structure of CrACP. The protocols for these modifications were developed previously working with less precious unlabeled protein.

We have performed initial 1H-15N HSQC and 1H-13C HSQC experiments to establish the spectra of the carrier protein and ensure the spectra resolves as a single species. Before more advanced experiments, which require more time and protein, it was ensured that we could resolve the CrACP as a single species. We performed experiments characterizing the spectra collected from two protein species: apo and holo CrACP. We found that the holo-CrACP did not resolved as a single species, excluding further experiments. ApoCrACP was found to resolve as a single species and was thus chosen for moving forward.

We have performed initial 3D assignment experiments of the amide backbone of CrACP. We have acquired a high-resolution 3D HNCA dataset, which represents our first step in assigning the backbone of the *Chlamydomonas reinhardtii* ACP. A full backbone assignment will be essential for our future experiments in determining the structure of the ACP. Furthermore, the assignments will open the door to NMR titrations experiments of WT and engineered enzymes (Williamson, 2013).

Because no experimentally resolved structures of either CrACP or CrTE have been resolved to date, we produced models of CrACP and CrTE via comparative modeling (homology modeling) using Schrodinger's Biologics Suite 2019-3 and Prime. These models will allow us to construct a model of the CrACP•CrTE complex. To generate the CrACP model, we performed a BLAST search to identify homologous carrier proteins. The 10 carrier protein homologs most similar to CrACP were used as template structures for model construction. The sequence of CrACP was aligned to each template sequence using Prime STA, a program that generates alignments using both sequence as well as secondary sequence prediction for both the sequences of CrACP and the template carrier proteins. Prime STA is designed to generate alignments in cases where the protein of interest and templates have low sequence identify (20-50%). An analogous method blast and PRIME STA alignment was performed to generate an initial CrTE structure. Models of CrACP and CrTE were built using an energy-based (i.e., force field; OPLS 2005) method, which preserves the coordinates of the sidechains of conserved residues and of the backbone atoms of aligned regions of the template and target sequences. Subsequently, the geometries of all side chains were optimized and non-template residues were subjected to energy minimization. Lastly, insertions and deletions in the target proteins were built and removed, respectively, yielding models of CrACP. The geometries of non-template loops in the homology model of CrTE were refined using loop refinement protocols implemented in Schrodinger's Prime software. Loop geometries were sampled as recommended by Schrodinger. Schrodinger's protein preparation wizard was used to add hydrogens and predict the protonation states of titratable residues. UcFatB1 and many other acyl-ACP thioesterases exist as functional dimer. We attempted to construct a model of the dimeric CrTE complex using PISA. This attempt proved unsuccessful, likely owing to the quality of the model; UcFatB1 and CrTE share ~25% sequence identify. Consequently, protein-protein docking simulations to predict a CrACP•CrTE were performed using the monomeric CrTE generated via homology model. Protein-protein docking simulations were performed using the Cluspro 2.0 protein-protein docking webserver (Kozakov 2013, Kozakov, 2017, Vajda 2017). The structures of CrACP, UcFatB1, and CrTE were prepared for the simulations using Schrodinger's protein preparation wizard in order to add missing side chains, hydrogens, and predict the protonation states of titratable residues. In all simulations, the carrier protein and thioesterase were treated as ligand and receptor, respectively. The three most promising decoys (docked complexes generated via computer simulations) of CrACP•CrTE and CrACP•UcFatB1 were subjected to further refinement using RosettaDock, [Gray 2003, Wang 2005] which perform so-called "local" protein-protein docking simulations, wherein predicted complexes are further optimized, by performing small rigid body displacements as well as sidechain conformational sampling. These local protein-protein docking simulations were performed using ROSIE, a web front-end to the Rosetta 3.x software suite. [Lyskov 2008, Chaudhury 2011, Lyskov 2013].

We have begun preparing the homology models of CrACP and CrTE for molecular dynamics (MD) simulations. Models have been solvated and will be subjected to 50 ns MD simulation for

further refinement and to generate an ensemble of coordinate data for protein-protein docking simulation.In order to develop a starting structure for engineering the homology modeled *Chlamydomonas reinhardtii* ACP (CrACP), the protein sequence was docked to the 5X04 *Umbellularia californica* thioesterase (UcTE). The high resolution Fast fourier transform based ICM-Molsoft docking software was used. These generated structures were then examined. The interface of the UcTE-CrACP proteins is non-native, as such there is not a strong interface. However, the most stable structure which was possible at the interface was selected. This interface suffers from unfavorable electrostatics, with the CrACP having a very negative charge and the UcTE having a much more negative charge than the native *Chlamydomonas reinhardtii* thioesterase. These docked models were used for engineering of the *Umbellularia californica* thioesterase (UcTE) to increase its interaction with the *Chlamydomonas reinhardtii* ACP (CrACP) using a mutagenesis protocol. Briefly, the models were fed into the Rosetta design program, with many rounds of random mutagenesis performed with refinement of the interface of mutants and scoring. The best scoring mutants of the thioesterase interface were chosen for performing point mutations in order to assess experimentally the performance of the redesign.

During the course of these experiments, it was identified that a transit peptide tag left on the CrACP could be leading to a loss of stability and yield *in vitro*. As such this transit peptide was cleaved, yielding a more stable and better expressing construct of the CrACP. This ACP was chemoenzymatically loaded with reactive α-bromo crosslinking probes. It is relevant to note that this was previously low yielding and the preparation of crosslinking probes had hampered crystallography attempts. However, small scale crosslinking experiments with the newly prepared ACPs proved successful and allowed for further development of the methodology. We have achieved crosslinking between the *Chlamydomonas* ACP and TE using crosslinking probes containing an α-bromine warhead. We have shown in previous studies that these probes are able to crosslink enzymes that utilize a nucleophilic cysteine as part of the active site. In order to make the *Chlamydomonas* TE compatible with these probes, the reported nucleophilic aspartate was mutated to a cysteine, allowing for consistent crosslinking with ACP loaded with probes containing a six-carbon long acyl chain. However, denaturing PAGE gels of crosslinking reactions consistently showed a band corresponding to unreacted TE. As acyl-ACP TEs have been reported to form homodimers as the functional enzyme, it is possible that this band is TE forming a dimer to an ACP-TE crosslinked complex. To answer this, we utilized a TEV protease site in the TE construct to cleave the C-terminal His-tag, thus leaving only the ACP capable of binding Ni^{2+}. We then applied crosslinking reactions to Ni^{2+} immobilized on resin, washed the resin with buffer, and finally eluted with buffer containing imidazole. Denaturing PAGE gels of fractions only showed TE bands with a corresponding crosslink band in the same sample, thus confirming that one ACP crosslinks a TE homodimer. Currently we are optimizing conditions for crosslinking in order to prepare a pure concentrated sample for crystal trials.

In order to test the ability of TEs to hydrolyze different chain-length fatty acids off of the *Chlamydomonas* acyl-ACP, we are further developing an *in vitro* HPLC-based approach to detect different states of ACP. In short, the enzyme AasS from *Vibrio harveyi* is used to load acyl chains ranging from C6 to C18 on ACP, followed by purification through gel filtration. TE can then be incubated with purified acyl-ACP and the reaction terminated by the addition of TFA. The extent of TE-catalyzed hydrolysis will then be monitored using HPLC.

The UCSD biochemistry lab has developed the crosslinking technology to the extent that we have begun crystallography work and conducted crystal trials. We have performed several Thermoflour screens to identify optimal conditions to induce crystallization of the crosslinked

ACP-TE, evaluating the influence of buffer, salt, and additives on protein stability. This work has identified conditions that improve the stability of the complex, which may facilitate crystallization. Neutral pHs (~7.5), relatively high salt concentration (~500 mM), and the addition of phosphate so far yield the best results.

We have also produced HPLC evidence that the *U. californica* TE can hydrolyze dodecanoyl-CrACP *in vitro*. This result contradicts previously reported data, although those data were collected from *in vivo* experiments (Blatti, et al). Consequently, there remains the possibility that endogenous fatty acid synthases outcompete the heterologously expressed UcTE for medium chain length fatty acids *in vivo.* Inaba *et al.* found that the heterologous expression of a cognate plant ACP-TE pair in *C. reinhardtii* does yield "slight but significant increase" in triacylglyercols-containing medium chain fatty acids (12:0 and 14:0), suggesting that the low affinity between CrACP and UcTE may inhibit the production of medium chain length fatty acids.

The UCSD biochemistry lab has also been investigating the ketosynthases involved in fatty acid biosynthesis in *Chlamydomonas*. These enzymes catalyze the condensation step of fatty acid chain elongation, and so would seem likely candidates to compete with TEs for activity on acyl-ACPs. Thus far, we have expressed and purified two of the three putative KSs in Chlamydomonas, and have tested crosslinking on one of those with successful results. Based on these results, we are currently generating ACP-KS (ketosynthase) complex samples that will also go to crystal trials as soon as they are ready.

References

a. Blatti, J. L. *et al.* Manipulating Fatty Acid Biosynthesis in Microalgae for Biofuel through Protein-Protein Interactions. *Plos One* **7**, e42949 (2012).
b. Inaba, Y., Nakahigashi, K., Ito, T. & Tomita, M. Alteration of fatty acid chain length of Chlamydomonas reinhardtii by simultaneous expression of medium-chain-specific thioesterase and acyl carrier protein. *Phycol Res* **65**, 94–99 (2017).

Testing green algal capacities to grow on NREL-provided hydrolysates: Because *Chlamydomonas reinhardtii* does not naturally grow on sugars, without further genetic modifications currently under investigation, we have been investigating a few potential green algae strains of *Chlamydomanas* for the heterotrophic production part of this project. Currently, the two most promising are *Chlamydomonas debaryana* UTEX 231 and *Chlamydomonas assymetrica* UTEX 227, both of which grow well on the NREL-supplied hydrolysate sugars or on media supplemented with 40 g/L pure glucose, yielding ~0.3 – 0.5 grams of biomass per gram of glucose. This value decreases significantly when grown on the NREL sugars, which contains a table of yield values for media supplied with different concentrations NREL sugars ranging from 40 – 80 g/L. Our conclusion from this data is that both UTEX 227 and UTEX 231 grow on the NREL hydrolysate, but only consume the glucose component of the sugars, leaving the remaining unusable sugars in the media. We can deduce that only the glucose is consumed by these strains by comparing the drop in yields: about 50% of the NREL sugars are glucose and we observed the yields drop by 50% relative to a pure glucose media. This is both good and bad news: although we can use the NREL hydrolysate for growth of *Chlamydomonas*, it also means that cultures will develop a buildup of unusable sugars that will ultimately lead to osmotic stress and slowing or arresting of growth, which was observed as lower yields at higher concentrations of the NREL sugars. Ongoing experiments are being conducted to optimize the C:N ratios for the media containing the NREL hydrolysates to optimize growth and production. Simultaneously, we are developing strategies to remove or better utilize the remaining sugars in

the hydrolysate to prevent the growth inhibition by the other sugars. Although we are waiting for sequencing data to tell us which of the two alternative UTEX strains is the likely best strain for further genetic modification for succinate production according to our discoveries in *C. reinhardtii*, we are currently pursuing two paths for developing a production strain capable of growing on a glucose sugar stream: (1) analyze the transciptome of these strains and determine if we can adapt our current vectors to these strains to genetically modify them, as transformation with these vectors has so far been unsuccessful, and (2) isolate the glucose transporter gene from UTEX 227 or UTEX 231 and express the transporter in *C. reinhardtii* to give it the ability to grow on the NREL sugars. These experiments will directly inform the experiments of Task 11 for scaled growth of an optimized production strain.

Testing cyanobacterial capacities to grow on NREL-provided hydrolysates: *S. elongatus* was similarly grown on media containing varying concentrations of the NREL hydrolysates to determine that the strain would tolerate the presence of sugars and utilize those sugars when genetically modified to do so. Therefore, the UCSD cyanobacterial lab grew WT *S. elongatus* and two strains that the UC Davis lab had previously constructed to produce 2,3-BDO while importing and catabolizing glucose, AL2456 and AL3384, on BG11 media containing sugar concentrations from 0 – 80 g/L. For most flasks, IPTG was added at a final concentration of 0.1 mM to induce expression of the exogenously encoded glucose transporter and the 2,3-BDO production pathway. Although the highest concentrations of sugars allowed for fungal or bacterial contamination, the WT strain tolerated all sugar concentrations except 80 g/L. The genetically modified strains tolerated up to 20 g/L, but we hypothesize that the intolerance of 50 g/L was due to production of BDO rather than the presence of the sugars. Analysis of dry biomass from each of the cultures demonstrated that the utilization of the NREL sugars could enhance biomass productivity up to 8-fold over the growth observed for WT in the same sugar media. These experiments indicate both tolerance to and capacity to grow on the NREL-provided hydrolysate, a conclusion that will directly inform the experiments of Task 11 for scaled growth of an optimized production strain.

Production of BDO from *Synechococcus elongatus* PCC 7942: The UCSD cyanobacteria lab generated a vector that theoretically enables insertion of an IPTG-inducible BDO production pathway into neutral site I of *S. elongatus*. This vector was constructed in two steps. First, an intermediate vector was constructed and sequence verified that encodes *kgd* from *Synechococcus* sp. PCC 7002, the same gene used in the succinate production vectors (e.g. pSuc2) to generate succinate semialdehyde (SSA) from alpha-ketoglutarate, and *4hbd* from *Porphyromonas gingivalis*, which converts SSA into 4-hydroxybutarate – the first intermediate towards 1,4-BDO production. To generate this vector, the *4hbd* gene sequence, which was reported as the highest producing one by Genomatica, was synthesized as two gBlocks, stitched together via overlap extension PCR, and then cloned with fragments from the pSuc2 vector via SLIC. A second intermediate vector was generated that encode the three remaining genes necessary for BDO production: *Cat2*, *ald*, and *adh1*. We have synthesized these three ORFs (again using the highest producing sequences as reported by Genomatica) as 6 overlapping gBlocks and have stitched them together into two amplicons that were individually cloned into cloning vectors. Through sequential addition of these amplicons to the first intermediary BDO production vector, we were able to generate the desired pBDO vector and confirmed the entire protein coding region and promoter with Sanger sequencing. This vector was shared with the UC Davis lab so that the production pathway could be cloned into their desired integration and expression vectors for generating 1,4-BDO using heterotrophic strains of *S. elongatus* PCC 7942 that are capable of growing on glucose or xylose. This vector was used

to transform four background strains, as was performed for the succinate production vectors: WTcm, WTcm+Fko, Lcm, and Lcm+Fko. Transformation results demonstrated that WTcm could not be transformed with the vector, while the remaining backgrounds produced segregated, double recombinant clones. We noted, though, that the WTcm+Fko colonies were small compared to the Lcm and Lcm+Fko colonies, indicating that the mutation previously identified in the sodium-dependent bicarbonate transporter SbtA in the LAN1 background allows for increased tolerance to the leaky expression of the BDO production pathway from the IPTG-inducible Ptrc promoter. Interestingly, these results also indicate that the knockout of the gene required to reversibly convert succinate to fumarate also allows tolerance to the presence of the BDO production pathway. We hypothesize that because the 1,4-BDO production pathway has 3 NAD(P)H utilizing steps, that it causes both a consumption of carbon and redox potential that imbalances the metabolism of the cell without one or both of these modifications.

Three independent clones of each of these transformants were grown in liquid culture, induced for expression through addition of IPTG, and were analyzed for the presence of 1,4-BDO using GC-MS, HPLC, and LC-MS methods previously developed in collaboration with Algenesis to analyze polyurethane biodegradation samples for 1,4-BDO generation. An interesting observation was that addition of IPTG after three days of growth caused growth to stall for all pBDO strains, while analogous succinate production strains and parental strains continued to grow after IPTG addition. We hypothesize that either the strain is producing BDO or an intermediate chemical at such as rate as to prevent carbon utilization for growth or BDO or the accumulated intermediate are cytotoxic to the cells. Another alternative is that the production enzymes are promiscuous and are having cytotoxic effects. GC-MS analysis was unable to detect the presence of 1,4-BDO in the supernatants of the pBDO production strains. Spike in additions of 1,4-BDO to these samples indicated that the lack of 1,4-BDO detection was not a problem with the methodology, but instead due to a true lack of BDO production. To determine if anything of interest was present in these supernatants, we analyzed the same supernatant samples using the HPLC analysis method developed for analyzing succinate from the succinate production strains. HPLC analysis revealed a very large peak present near to where succinate typically appears, but repeatedly shifted relative to succinate (5.67 +/- 0.02 for succinate vs. 5.50 +/- 0.06 for pBDO samples' peak). Although our previous experiments indicate that this type of shift may occur for succinate due to pH and media changes, we do not believe that this is the explanation for the appearance of this peak, due to a lack of evidence of the kind of pH change necessary for this type of shift. Instead, we believe that this peak represents the accumulation and secretion of one of the intermediates along the pathway. To attempt to identify the molecule comprising this novel peak, we have run these same samples on the ECAL lab's Orbitrap LC-MS/MS machine to acquire a similar chromatogram trace and both the parental MS and fragmentation MS/MS fingerprints for these samples. We have uploaded the data into the GNPS database server and are currently using its molecular networking algorithms to identify molecules of interest in general, but the identity of this novel peak in specific. We have supplied this data to the group at PNNL, who has provided predictions to explain the observed results, as described for Subtask 6.4.

Production of BDO from *Synechococcus elongatus* PCC 7942 (mixotrophic): The UC Davis lab, having received the pBDO vector from the UCSD cyanobacterial lab, and has constructed several strains to produce 1,4-butanediol (1,4-BDO) in strains of PCC 7942 that are capable of growing mixotrophically on glucose or xylose.

Task 3.0: Design and Construct Advanced Synthetic Promoters - Round 1

The green algae group at UCSD has computationally identified specific sequences, elements, and structural motifs found in the 5'-untranslated regions (UTRs) of chloroplast mRNAs that affect mRNA translation. For this analysis, secondary structure predictions of these 5'-UTRs was carried out using RegRNA2 to predict if there are any identifiable structural motifs in each 5'UTR and RNA-fold to predict the global folding structure of each 5'-UTR. Over-represented motif sequences in 5'-UTRs were identified using MEME and RSAT::plants on the collection of 5'-UTRs. Shine-Delgarno (SD) sequence motifs were also analyzed. We found that the SD motif GGAGA is found in a number, but not all, of the mRNA UTRs from the chloroplast genome, including the psbA 5'-UTR. We scanned SD-like motifs GGAGG or GGAGA or GGAG in all chloroplast 5'-UTRs of protein coding genes. These computational analysis results were collected and sent to Pacific Northwest National Laboratory and Lawrence Berkeley National Laboratory to act as a starting point for the analysis of known promoters and the design and construction of a synthetic library of promoters to increase protein expression in either the nucleus or the chloroplast of *Chlamydomonas*.

The computational design of the novel synthetic algal promoters was split into three subtasks: 1) Proximal design focused on the first 160 bp upstream of the TSS, 2) Distal design focused on the remaining 90 bp, and 3) the addition of adaptors 20 bp upstream and 20 bp downstream of the promoter design for cloning of the promoters into a barcoded vector library.

For the proximal design, we used a neural network to embed both motifs and the full promoter sequence into a lower-dimensional space by applying the Levenshtein distance metric to the known and novel motif and motif locations identified in the promoter sequences. We then used the Hyg-Zeo and Hyg resistant promoters from the UCSD algae group's previous data set (derived from the promoter sequence analysis and prior work from a DOE-funded PEAKS project to screen 2,000 synthetic algal basis promoters of length 160 bp for Hygromycin-Zeocin antibiotic resistance) to train a categorical classifier, and used RNA-seq data from Hyg-Zeo promoters to train a continuous regressor. Accuracy was about 80% on the training set and 40% on the test set. We also applied Auto-sklearn to the promoter sequences and obtained a predictor that had an accuracy of about 77% on the training set and 30% on the test set. We then split the design into two parts:

1. Exploitation focused on finding sequences that the machine learning algorithms confidently predicted was a high expressor.
    a. Group comb passed both neural network predictors (using motifs and positions as input and using the full promoter sequence as input)
    b. Group opt passed the AutoML predictor based on sequences generated from a genetic algorithm optimization.
2. Exploration focused on finding sequences that we could learn from the most in the next round.
    a. Group rand explores space-filling random sequences that are predicted to be high expressors but because they were not sampled from the training distribution, they may turn out to be duds and
    b. Group nonh are sequences that are predicted to be low expressors but contain motifs that indicate it may be a high expressor in reality.

These applications produced the following number of promoter sequences, as associated with the output files generated:

1. 796 from comb_sequence.csv, 398 each for 160bp and 250bp
2. 750 from rand_comb_sequence.csv, 500 each for 160bp and 250bp
3. 750 from nonHyg_sequence.csv, 500 each for 160bp and 250bp

4. 704 from optAutoML sequence, 352 each for 160bp and 250bp

For distal design, we predicted chromatin opening motifs by training on native expression data and native histone and RNApol ChIP-Seq data. We then randomly added 2-3 motifs to 90 bp of group 1 promoters (which did not contain any motifs), but did contain the desired GC content. Promoter sequences were screened to remove undesired restriction cut sites (XbaI and NdeI) and then the final set of two oligo pools of promoter sequences was delivered to the group at LBNL for synthesis of the oligos by Twist.

To support the construction of the promoter library, an initial vector (pAES5) was designed based on previous promoter screening libraries performed under a DOE PEAKs project and constructed for cloning of promoter sequences to allow the nuclear-encoded expression of a fluorescent reporter protein, Clover. The sequence of this vector was supplied to the national lab groups to allow for incorporation of cloning sequences into the synthetic promoter library sequences to be designed and constructed. Based upon feedback from the LBNL group on this vector sequence, an upgraded reporter vector, called pBEEPS1, was generated to interrogate the promoter library in *Chlamydomonas*. Briefly, a synthetic promoter sequence is inserted upstream from a basal RBCS2 core promoter. The core promoter itself drives very little transgene activity but does contain a transcriptional start site. A mClover (Green Fluorescent protein derivative) reporter gene fused to a Zeocin-resistance cassette is then transcribed carrying endogenous RBCS2 5’ and 3’ untranslated regions as well as two introns derived from the RBCS2 gene. Unlike many other model organisms, inclusion of introns is vital for strong transgene expression in Chlamydomonas. A secondary Hygromycin resistance cassette is driven by an independent Beta-tubulin2 promoter, which provides an internal control for the silencing or expression inherent to the site of nuclear integration. Based on technical limitations encountered in the first round of promoter library analysis, improvements in a common adaptor region, allowances for exonuclease “end nibbling” during integration of linearized DNA in to the genome, and markedly more flexible and improved sequencing primers for interrogating the promoter library have been added. The promoters themselves contain common adaptors that will be used to amplify out of the Twist Bioscience provided library. An 8-base pair “randomer” barcode is added to each end of the promoter for later quantitation before insertion into a common cloning site through standard Gibson assembly methods and *E. coli* transformation. The DNA is then extracted from the *E. coli*, linearized using the NdeI and XbaI restriction enzyme sites, and transformed into *Chlamydomonas*. After algal transformation and selection for high expressing clones, common sequencing primers will be used to amplify both the barcode and the promoter sequence. The PCR products can then be analyzed using Illumina short read sequencing. This vector sequence and samples were provided by the UCSD green algae group to the LBNL group for promoter library construction.

The group at LBNL ordered the two oligo pools from Twist corresponding to the two different synthetic promoter length library designs from the PNNL group, as well as the necessary oligos for reamplifying the pool. These oligos were designed to introduce 8-mer barcodes and encode vector overlap sequences for downstream Gibson assembly 5’ to the barcode and overlap sequences with the 20 bp oligomer adapters 3’ to the barcode, allowing the pooled library to be cloned into the pBEEPS1 destination vector. The promoter library generated by Twist was amplified via PCR using these reamplification oligos to add 8-mer barcodes to each end of the promoter library oligomers. The vector overlap sequences were designed to work well with primers appropriate for NEB's Q5 hot start polymerase and HiFi assembly kit. In parallel, the pBEEPS1 vector backbone that was provided to LBNL was modified by PCR to amplify the whole vector, but to add common adaptors as primer tails. The barcoded oligo library was then cloned into the modified pBEEPS1 vector using Gibson assembly. Purified plasmid libraries and

E. coli transformants were generated. Four libraries in total were generated and labeled 200, 290, 200 GC, and 290 GC. The libraries labeled 200 contained the 160 bp promoters (plus two 20 bp adapters at the ends of the oligo), while those labeled 290 contained the 260 bp promoters. For those labeled with GC, GC buffer was used for PCR amplification to achieve better plasmid population distribution.

An MTA was successfully executed to transfer the plasmid libraries to UCSD at about the time that the COVID-19 shutdown in California was put into place, preventing the libraries from being shipped to UCSD in a timely manner. The libraries were also prepared for high-throughput sequencing in order to evaluate the diversity, coverage, and bias of the library sequences with respect to the *in silico* designed library. In spite of delays due to the COVID-19 shutdown, the plasmid libraries were prepared for high-throughput sequencing in order to evaluate the diversity, coverage, and bias of the library sequences with respect to the *in silico* designed library, and subsequently sequenced. Both a preliminary analysis and follow up analysis of the sequencing data demonstrated that the constructed library was diverse and suitable for the experiments in Task 4.0. In this analysis, we noted that the 290 libraries are both approximately equivalent while the 200GC library has noticeable superior representation compared to the 200 library. Because the difference between the GC and the non-GC library was in whether the High GC enhancer from NEB was added to the Q5-based amplification of the promoter library from Twist while the substrate DNA library was the same for each reaction and resulting library, the observed difference in diversity between the 200GC and 200 library may represent an issue with the high GC-content of the *Chlamydomonas* encoded sequences. Independent of the reason and the difference in diversity, the two 200 bp libraries were pooled just as the 290 bp libraries were pooled and used as two promoter libraries instead of four for the purposes of *Chlamydomonas* transformation in Task 4.0

Milestone 3.ML.1: DNA library for advanced synthetic promoters. Synthesized DNA library with 300+ synthetic promoters and corresponding GFP expression vectors.

As detailed above, both the oligomeric synthetic promoter library and the GFP expression vectors were successfully generated and verified.

Task 4.0 Screening Advanced Promoters – Round 1

Subtask 4.1. Generate, screen, and assess first round of expression libraries in *Chlamydomonas* using a fluorescent reporter gene.

As a part of the vector and experimental design performed prior to receiving the DNA-encoded promoter libraries, the strategy for screening the promoter libraries to select for high expressing cells was established as follows. The frequency of a certain promoter sequence with a *unique* barcode sequence recovered from the algae is used to count the level of expression of each promoter. We do this because (1) transformation in to *Chlamydomonas* inserts the transgenes randomly in to the genome, thus causing strong positional effects and requiring many insertions of the same sequence to normalize for these effects in order to generate a relatively quantitative readout of promoter activity and (2) multiple levels of clonal duplication bias can occur during the *E. coli* preparation, algae grow-out and selection, and the downstream PCR isolation of promoter sequences from the algae and Illumina library prep. Therefore, using an 8-mer random barcode, we can ensure that there are many orders of magnitude greater levels of barcode complexity than the number of promoters being tested. Thus, each promoter sequence is flanked by tens of thousands of unique barcodes in our library. This means that when a unique barcode-promoter pair is detected in the Illumina sequencing library it can be interpreted as a unique insertional event. The relative frequency with which we see uniquely barcoded promoters is proportional to how well they drive transgene transcription as we select for high expressers, which occurs as follows:

1. Selection Parameters for promoter library to isolate the strongest synthetic promoters.
    a. Level 1- Hygromycin-only selection. Since hygromycin resistance is driven by a beta-tublin promoter that works independently of the synthetic promoters we are testing, the data from this selection represent the full range of the promoter library that is “algae transformable”.
    b. Level 2- FACS enrichment of the top 10% mClover expressing of Hygromycin selected cells. Here we look for enrichment of strong synthetic promoters using fluorescence alone.
    c. Level 3- Zeocin+Hygromycin selection. The Ble^R gene that is fused to the mClover reporter confers resistance to Zeocin by stoichiometrically binding and sequestering the Zeocin molecule. Comparatively, the Hygromycin resistance gene catalytically modifies and inactivates the Hygromycin. There needs to be about 100 fold higher Ble^R protein expression than Hygromycin expression to confer antibiotic resistance. Therefore, selection of promoter library transformants on Zeocin media strongly selects, through survival, for the top 1-3% of the most active promoters.
    d. Level 4- FACS enrichment of the top 10% mClover expressing of Zeocin+Hygromycin selected cells. This isolates the most extreme promoters.
2. The promoters that are found to be consistently enriched in the FACS-enriched or Zeocin enriched datasets compared to the Hygromycin-only will be considered to be stronger promoters and verified in follow up experiments. All expression-associated sequence data will then be provided to the group at PNNL to design a second round of the synthetic promoter library.

Once the DNA plasmid expression libraries were received by UCSD from LBNL, the green algae group at UCSD transformed the libraries into *Chlamydomonas reinhardtii* using electroporation. To test the efficacy of transformation methods, the expression plasmids were linearized for some transformations through digestion to release the expression cassette from the backbone, while in others the circular plasmids were used for transformation. For each

transformation, clones were plated onto plates containing either Hyg or Hyg+Zeo antibiotics. Because of the standard level of expression of AphVIII from the β-tubulin promoter, all transformants where integration occurs in a non-silent region of the nuclear genome should grow on either plate, but only those with a designed promoter from the library that has reasonable enough expression to drive the expression of the *ble* gene should survive on the Hyg+Zeo plate. For each library, approximately 4000 transformants grew on the Hyg plates in total while approximately 600 transformants grew on the Hyg+Zeo plates in total. Thus, we were able to generate between 9,000 – 10,000 colonies from the promoter libraries in the first round of transformations.

The transformants were then scraped from the plates, pooled and grown in culture, and then analyzed and sorted using FACS. Prior to the library runs, an untransformed WT *Chlamydomonas reinhardtii* sample was run through the FACS in order to set the gating for WT levels of chlorophyll fluorescence and mClover fluorescence. Those cells that had higher than WT levels of mClover expression, were sorted for recovery. For both the 200 bp and 290 bp libraries, only approximately 1 to 2% of the Hyg-only selected transformants fell into the selected higher expression gate, while approximately 25 to 30% of the Hyg+Zeo selected transformants displayed increased mClover expression over WT. As stated above, that was as expected due to the different selection requirements associated with the different antibiotics.

Those cells that were selected via the FACS gating for mClover expression over WT were then plated and grown under Hyg selection. Approximately 800 colonies were picked, grown in 96 well plates, and analyzed for mClover fluorescence using a fluorescence plate reader. Values ranged from similar to WT (0.07 – 0.1) to ten-fold higher values.

In addition to the first library transformation, an additional 4 library transformations were generated and screened via FACs, using the same techniques as described above. The top 10% GFP-expressing constructs from each of these were then PCR-amplified to capture the promoter and barcode regions and Illumina sequenced. The sequencing results for the all libraries that were successfully sequenced were supplied to PNNL for analysis for determining the design of the next round of promoter sequences (Task 8.0). In addition, another 800 more single individual clones from the second and third transformation replicates were picked, grown in 96 well plates, and analyzed for fluorescence values using a Tecan plate reader. For the top performing clones, the promoters demonstrated 5 to 28-fold increases in GFP expression over WT. The top 20 clones for each library were sequenced by Sanger sequencing to identify the barcodes and promoters in these clones, though no clear patterns were observed as each individual promoter sequence was unique and did not share any apparent motifs.

Further, in order to better determine if these promoters respond to different growth conditions, all clones were grown and tested under different media and light regimes. Cells were tested on standard TAP medium under continuous light or continuous dark. Cells were also tested in HSM medium, TAP medium without nitrogen, or TAP medium with 1% salt, all under continuous light. This screen resulted in a wide dynamic range of promoter expression values that were at minimum 4 – 5 fold greater than WT and ranged above 15 fold higher than WT fold expression values. For the two conditions shown (TAP+light and TAP+Dark), strong (9 – 15 fold higher than WT) and super strong (>15 fold higher than WT) expression values were observed. Interestingly, the different length promoter libraries resulted in different responses to light versus dark in terms of the number of super strong promoters – the 200 bp library had more super strong promoters when in light than dark while the 290 bp library had more super strong promoters in the dark than in the light. This data demonstrates that our design, build, and test

methodology can not only produce higher expressing promoters, but can also produce promoters that are responsive to different ambient conditions.

Milestone 4.ML.1: First round of synthetic promoter and expression libraries integrated into *Chlamydomonas*, generating at least 10,000 individual transformants. 10,000+ individual transformants resulting from electroporation of 3.ML.1 expression library.
As described above, approximately 9,000 – 10,000 individual transformants were generated in the first round of library integration into *Chlamydomonas*. Those numbers were eventually exceeded by at least four-fold through four additional transformations.

Subtask 4.2. Bioinformatics analysis on promoter library resulting in predicted improvements to promoter library.

Prior to sequencing data being available from Subtask 4.1, similar promoter library sequencing data generated from a prior DOE-funded PEAKS project was supplied to the computational group at PNNL in order to develop the methods and algorithms for analyzing this type of sequencing data. The algorithms that the PNNL group implemented calculate sequence distance measurements, based on the Levenshtein distance metric, that allow for evaluation of general sequence similarity between highly expressing promoters as well as for motif analysis associated with highly expressing promoters. These metrics and the associated expression level data were then used to train a neural network to learn a tri-partite classification scheme based on the experimental data, with the three component data values being whether promoters were present in the different screening stages described for Subtask 4.1. Although other networks were tested, a feedforward architecture proved fastest and the most accurate. Using this trained network, data not used to train the algorithm was processed to predict expression levels and survivability in the different antibiotic-based screening stages. Based on these predictions, survival indexes were calculated, used to evaluate false positive and false negative rates, and will become the basis for evaluating or designing a new library of promoter sequences. Thus, the PNNL group was able to develop and tested the necessary software tools to rapidly perform the Subtask 4.2 analysis prior to receiving the data from Subtask 4.1. Secondly, the set of survival indexes allowed the PNNL group to evaluate the consistency and reproducibility between the rounds of FACS screens. This is an important statistical analysis for the design of the Promoter Library screen experiments because it is this evaluation that can tell the UCSD green algae group whether replicate screens are necessary (which they are according to the data) and whether more replicates will be better than fewer (the answer again is yes). Thus, the evaluation of these data sets was crucial in providing information for an improved experimental design for screening these libraries.

Based on these initial results, the PNNL group developed additional computational methods to describe the relationship between the number of replicates, the false-positive discovery rate, and the threshold defining a high-promoter in the Hyg/HygZeo testing. They also use data from multiple replicates to estimate survival index distributions for high-expressing promoters - not just point estimates. These analyses suggested that a replicate number of 5 is ideal for generating strong statistics, and thus five separate transformations were recommended and performed in Subtask 4.1 by the UCSD group. The PNNL group also compared round 3 promoter results across replicates, with different embedding methods, and with round 2 data. These resulted in the following discoveries:

3. The Hyg and HygZeo counts varied significantly between rounds and across replicates within round 3. For example, R^2 values were higher for Hyg counts than Hyg-Zeo counts, but none were greater than 0.08 - the results showed a low degree of correlation. Several

other statistical methods were used to study the same phenomenon, and those methods returned similar results.

4. This variability corresponded to large variations in the ML predictions resulting from each round of data. Given an initial set of 1e6 randomly generated promoters, the classifiers and regression generated from each round were used to predict the top 1e4 promoters (i.e., top 1% of the sample size). The overlap between sets predicted from different rounds was less than 10% (typically on the order of 1%) regardless of which kind of embedding (string- or motif-based) was used.
5. There was a similar level of variability within data sets when comparing the results produced by the different embedding methods.
6. We identified some general trends in the motifs used (and their locations) in predicted high-expressors, but the variability here was also considerable. Furthermore, approximately 68% of the original 1e6 motifs contained no motifs (from a previously specified list), and all of the top 1e4 promoter lists had a comparable proportion of promoters with no motifs.

Based on these results, further methods were defined to refine the scoring of the sequencing data provided and to train algorithms for deciding on the next round of promoter sequences to test. The first method was based on a gamma-distributed population-based approach, whereas this method focuses on cell-level, promoter-specific Bernoulli distributions resulting in binomial-distributed populations; the methods differ in their assumptions and their methods for combining data across multiple replicates. In addition to this, we defined another assessment approach designed to leverage the FACS-sorted data, wherein the results come from the top 10% of the promoters. This method leverages the assumptions and calculations of the binomial-based approach to provide additional information about the promoter-specific expression level distributions. It also allows us to compare the stringency of the HygZeo test with that of the FACS screening.

Since the gamma- and binomial-based analysis methods can also be applied to the FACS data, we stopped algorithm development, having five different tests that we can apply to our promoter data: two analyses each for the normal and FACS-sorted data, and one analysis that combines the two datasets. Moreover, these results can be used to train classifiers/regressors based on all five evaluation methods. This essentially provides a multi-model approach, and therefore increased robustness, in evaluating promoter data and predicting high-expressing promoters.

We initially applied these analysis tools to the first two replicates of the 200 bp sequence data for which good sequencing results were available. Of the five analysis methods, 28 promoters were common to the top third in every category. Note, however, that only 193 promoters had a sufficient number of Hyg and HygZeo counts in both replicates to do the survival index analysis (using the gamma distribution); only 52 promoters had a sufficient number of counts in the FACS data. 28 is therefore over half of that 52. Of those 28 promoters, 14 came from the ‘comb’ category, 2 came from the ‘rand’ category, 5 came from the ‘nonh’ category, and 7 came from the ‘opt’ category. This would seem to indicate that our prediction methods do better than random space-filling sequence generation methods. Within those 28 promoters, three motifs occurred three times; none occurred more often than this. Also, five locations were the start points for motifs in three different promoters; no locations were motif start points for more than three different promoters.

Upon receiving more data, PNNL analyzed the barcode count data from the sets of 200 bp and 290 bp promoter sequences, including the FACS-sorted data for the first four libraries. We identified promoters that survived in multiple replicates of the Hyg-Zeo screening process and

identified repeated motifs and motif locations in those promoters. Following this analysis, we used the barcode data to train eight predictors based on the previously-defined metrics and embedding methods and performed k-fold cross-validation on those training processes. The 200 bp sequences did much better than the 290 bp, so we proceeded with the next round of promoter prediction solely focusing on the 200 bp sequences.

We then randomly generated one million promoter sequences, ran them through the predictors, and aggregated the results across predictors to identify top promoters; we also trained separate predictors on the previous round of 200 bp data for screening purposes. To narrow this down to a set of 3,000 unique promoters that would be synthesized in the next round of experimentation, we selected the top 10,000 promoters, based on the aggregate predictor metric. From those 10,000, we selected:

1. The top 1,000 promoters based on estimates from each individual predictor (764 unique promoters, not 800, because some promoters were in the top 1,000 of more than one predictor). These promoters were essentially 'high-risk, high-reward' choices.
2. The top 1,100 promoters based on the aggregate of the previous round's predictors (1,034 unique promoters, not 1,100, because some of the 1,100 had already appeared in the 764)
3. The top 1,492 promoters based on the aggregate measure (1,202 unique promoters, not 1,492, because of overlap with the sets of 764 and 1,034 promoters); the 290 overlapping promoters were highlighted for special attention, as they represented the most promising candidates.

The list of ~3,000 promoters was supplied to LBNL for oligo synthesis as part of Task 8.0

In addition to this, we defined and performed promoter selection calculations – essentially, the probability of obtaining a certain number of unique promoters when sampling randomly from a set of colonies. These results help to support microwell plate fluorescence testing as a means to evaluate promoter strength by informing how many clones need to be screened to get a reasonable coverage to discover unique promoters.

Task 5.0: Biosensors for Target Molecules

Subtask 5.1: Design, generate, and optimize biosensors for 1+ target molecules

The goal of this task was to develop olfactory receptor (OR)-based sensors in the yeast *Saccharomyces cerevisiae* for the detection of terpenes, fatty acids and diacids. The sensors would be used to detect these chemicals when overproduced by the cyanobacteria *Synechococcus elongatus* directly in the supernatant. Such a tool would dramatically accelerate the engineering of cyanobacteria for the production of polyurethane precursors over existing mass spectrometry-based methods for small molecular detection.

Succinic acid sensor

The research group at Georgia Tech has expertise in the development of OR-based sensors in yeast, having developed OR-based sensors for medium-chain fatty acids (MCFAs)[1] capable of detecting MCFAs in microbial broth[2], as well as OR-based sensors for pinene, lilial and undecanal[3]. Additionally, the Georgia Tech group has generated sensors using G protein-coupled receptors (GPCRs) more broadly, including sensors for serotonin[4], histamine, and oleoylethanolamide[5]. Key differences between those previous sensors and the desired succinic acid sensor are the fact that: 1) succinic acid is an endogenous microbial metabolite, which may result in higher sensor background levels, 2) no OR is known to bind succinic acid, and 3) the only GPCR known to bind succinic acid couples via $G_{\alpha i/q}$ to the mammalian cell machinery and not via $G_{\alpha olf}$ or $G_{\alpha s}$, both of which we have successfully coupled to the yeast machinery.

To develop a succinic acid sensor, a two-pronged approach was taken. *First,* we screened 350 human ORs for activation in the presence of succinic acid to identify a starting OR scaffold to develop a succinic acid sensor with the desired linear and dynamic ranges. This high risk/high reward proposition was mitigated by the fact that a) ORs couple well to the yeast machinery and thus we were likely to have functional OR-based sensors, and b) the Georgia Tech group has previously deorphanized ORs, i.e. identified ligands for ORs with no known ligands. *Second*, we leveraged an existing GPCR known to bind succinic acid in mammalian cells[6, 7], GPR91, to develop a succinic acid sensor in yeast. The key challenge with this second strategy was the fact that GPR91 couples to $G_{\alpha i/q}$ in mammalian cells, and $G_{\alpha i/q}$-coupled GPCRs have not been shown to couple to the yeast machinery.

OR-based succinic acid sensor

We generated 333 OR-based sensors and screened them for activation in the presence of succinic acid. Briefly, in OR-based sensors, a human OR is expressed on the cell surface of the yeast sensor strain (W303 *Δste2, Δsst2 Δfar1*). Upon activation with succinic acid, the OR signals via the yeast G_{α} subunit (GPA1) to the yeast mating pathway, ultimately leading to the transcriptional activation of green fluorescence protein (GFP) which is read via flow cytometry (**Fig. 5.1A**). An OR-based sensor was considered a hit if there was a statistically significant increase in fluorescence ($P < 0.005$) in the presence of succinic acid vs the presence of the carrier solvent only (DMSO). As shown in **Figure 5.1B**, we identified eight OR-based sensor hits. A secondary assay failed to validate the eight OR-based sensors. None of the eight OR-based sensor resulted in a dose-response increase in signal upon addition of increasing concentrations of succinic acid (**Fig. 5.1C**). *In conclusion*, no OR was identified to detect succinic acid in an OR-based sensor system in yeast using this strategy.

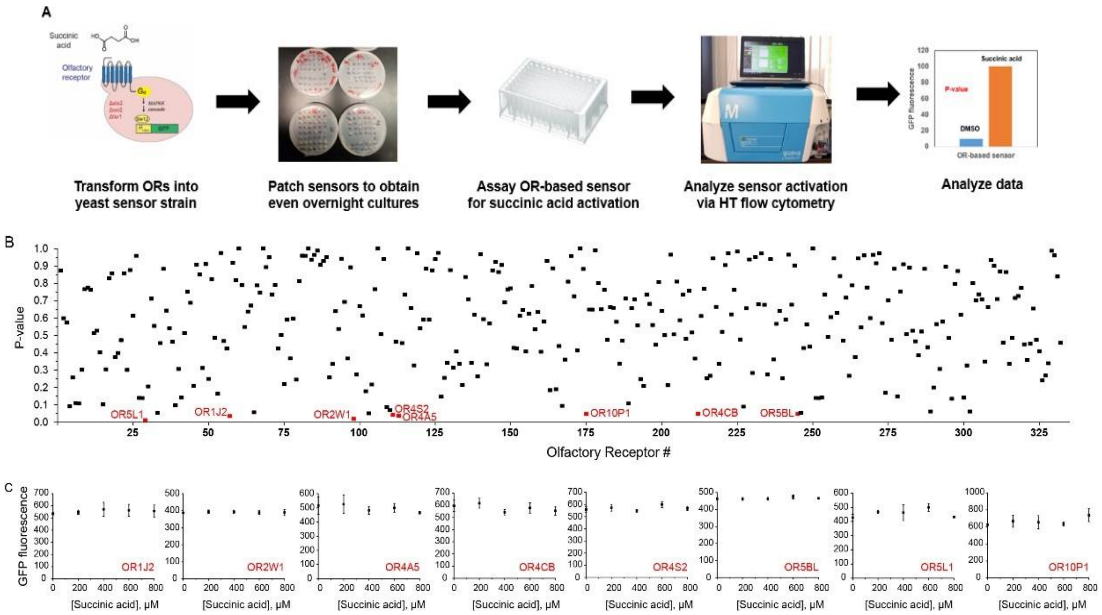

Figure 5.1. OR-based succinic acid sensor development. A. Workflow for screening OR-based sensors in yeast for succinic acid detection. OR-based sensor. Human OR (blue) is expressed on the yeast cell surface of the yeast sensor strain (W303 *Δste2, Δsst2 Δfar1*). Upon activation with succinic acid, the OR signals via the yeast Gα subunit and the yeast mating pathway (yellow) ultimately resulting in transcriptional activation of green fluorescence protein (GFP) which is read via flow cytometry. **B**. 333 OR-based sensors were tested in the presence of succinic acid or in the presence of the carrier solvent only (DMSO). Eight OR-based sensors (red) resulted in a statistically significant increase (P-value < 0.05) in signal after activation. **C.** Secondary assay of the 8 OR-based sensors hits. All hits were false positives, as they do not results in an increase in sensor signal upon increasing concentration of succinic acid.

GPR91-based succinic acid sensor

Gα subunit engineering. Previously, we have successfully coupled $G_{\alpha olf}$- and $G_{\alpha s}$-coupled human GPCRs via the yeast G_{α} subunit (GPA1) to the yeast machinery. The $G_{\alpha q/i}$-coupled

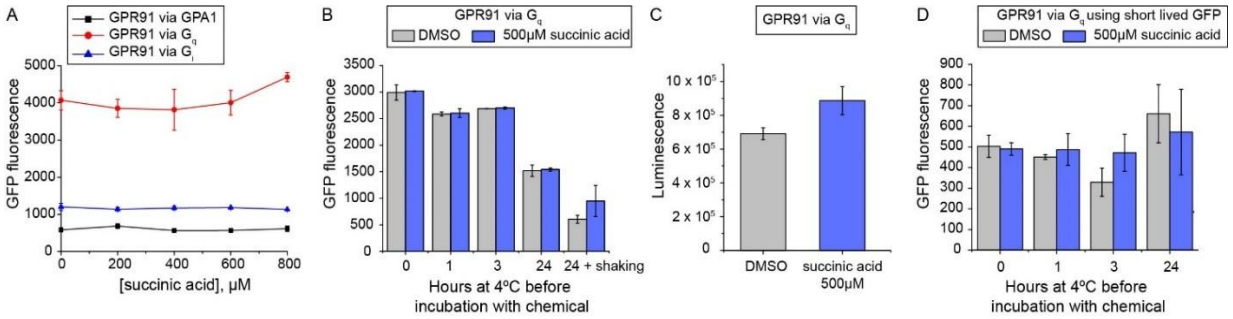

Figure 5.2. **Engineering the G_{α} subunit to improve the coupling of GPR91 to the yeast machinery**. **A**. GPR91 expressed in yeast sensor strain coupling via GPA1, $G_{\alpha q}$ or $G_{\alpha i}$ in the presence of succinic acid. **B**. GPR91 expressed in yeast sensor strain coupling via $G_{\alpha q}$ in the presence of DMSO or succinic acid after the cells were incubated at 4ºC between 0-24 hours. **C**. GPR91 expressed in yeast sensor strain coupling via $G_{\alpha q}$ using luciferase as the reporter in the presence of succinic acid or DMSO control. **D**. GPR91 expressed in yeast sensor strain coupling via $G_{\alpha q}$ using a short lived GFP (ubiquitinated) in the presence of DMSO or succinic acid after the cells were incubated at 4ºC between 0-24 hrs.

GPR91 did not did not couple via GPA1 to the yeast cell machinery (**Fig. 5.2A**). To couple GPR91 via either $G_{\alpha q}$ or $G_{\alpha i,}$ we constructed a new yeast sensor strain with the endogenous yeast Gα subunit, GPA1, deleted (W303, *Δste2, Δsst2, Δfar1, Δgpa1*). Using this new sensor strain, we observed a slight increase in GFP fluorescence when GPR91 was coupled to $G_{\alpha q}$ (**Fig. 5.2A**). Of note, GPR91 coupling via $G_{\alpha q}$ resulted a 4-fold higher background fluorescence than in the presence of $G_{\alpha i}$ or GPA1. This background fluorescence could be attributed to succinic acid being an endogenous metabolite. To address this challenge, we altered the sensor protocol to include cell washes and an incubation step at 4ºC to reduce the levels of succinic acid prior activation with exogenously added succinic acid. These changes resulted in a slight activation when the sensor was incubated at 4ºC for 24 hours prior to induction (**Fig. 5.2B**). However, the activation with succinic acid was not statistically significant. Hypothesizing that GFP had too high background fluorescence we switched reporters to luciferase, which has helped increase the dynamic range of previous GPCR-based sensors[8]. As shown in **Figure 5.2C**, addition of succinic acid did increase cell luminescence without any cell washes or incubation at 4 ºC. Finally, we also explored the use of a short lived GFP, by tagging GFP with ubiquitin, which would reduce the accumulation of GPF and lower the background signal. As shown in **Figure 5.2D**, there was an increase in signal in the presence of GFP after 3 hours of incubation at 4 ºC. However, the signal was not statistically significant.

Directed evolution of GPR91 to develop an improved succinic acid sensor. To further improve the coupling of GPR91 via $G_{\alpha q}$ to the yeast machinery, we then focused on evolving GPR91. The fluorescent output of the GPR91-based sensor allows us to screen very large GPR91 mutant libraries (> 10^6) using fluorescence activated cell sorting (FACS) (**Fig. 5.3**). Specifically, we generated three different GPR91 random mutagenesis libraries with a range of mutagenesis rates from 2-3 to 8-9 mutations per gene with library sizes of 10^5-10^6 members. The different libraries would allow searching of a wider sequence space.

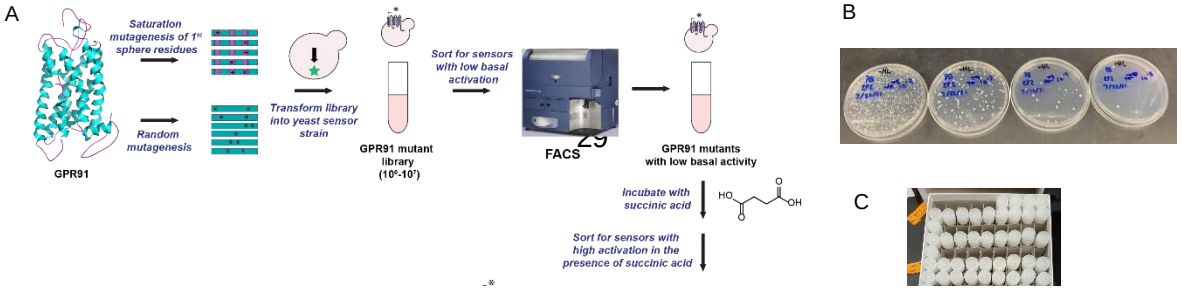

Figure 5.3. Succinic acid sensor evolution strategy. A. Workflow to evolve a GPR91-based succinic acid sensor. **B.** Sample picture of GPR91 random mutagenesis mutants transformed into the yeast sensor strain. **C.** Glycerol stock of GPR91-based sensor obtained via random mutagenesis.

To reduce the number of false positives, the libraries were screened in positive and negative sorts (**Fig. 5.4**). After three positive sorts of Library 3 (8-9 nucleotide mutations per gene), there was no significant increase fluorescence levels after addition of succinic acid. Similarly, after three negative sorts, we did not see a significant decrease in the sensor's background fluorescence.

Going forward, the sensor protocol will need to be optimized for FACS, which includes many more variables than the luciferase or flow cytometry assay. Specifically, FACS requires sorting of 10^6 cells rather than screening 10,000 via flow cytometry. For flow cytometry experiments, the cells are highly diluted (OD < 0.1) when they are interacting with the substrate (succinic acid); such a dilution factor is not possible with 10^6 cells due to the media volume required. We will therefore, in the future, take a step back to examine if the sensor will perform well at higher cell concentrations.

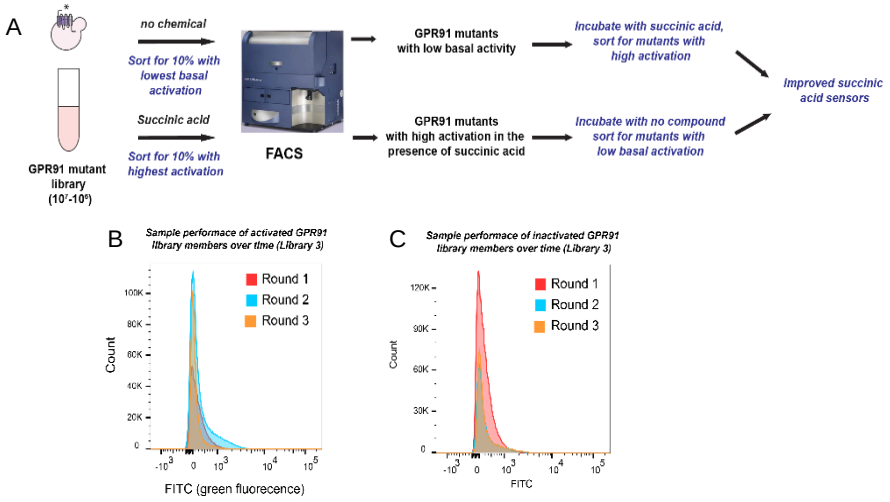

Figure 5.4. Sorting of GPR91-based sensor variants. A. GPR91-based sensor variants sorting strategy. **B**. and **C** Fluorescent activated cell sorting (FACS) results from library 3 activated in the presence of succinic acid (**B**) or inactivated in the absence of succinic acid.

Two-component system-based succinic acid sensor

The focus of the team was the production of succinic acid and its improvement. Given the paucity in identifying a succinic acid sensor using a GPCR-based system, we moved on to explore two-component system to detect succinic acid.

We started developing a succinic acid sensor based on a *Pseudomonas putida* DcuR/DcuS two-component system[9] that had resulted in succinic acid dependent cell growth in *E. coli*[10]. Given

that the ultimate application is rapid detection of succinic acid produced by cyanobacteria, we needed to switch to a high-throughput reporter, such as fluorescence rather than cell growth (**Fig. 5.5**). We observed that the background fluorescence of the DcuR/DcuS system was too high to detect succinic acid, likely due to the endogenous levels of succinic acid. We have attempted to reduce the levels of fluorescent background via multiple strategies, including using different media, shortening the chemical incubation time, testing different promoter strengths and vector copy number, but we have not yet been able to significantly reduce the level of background fluorescence.

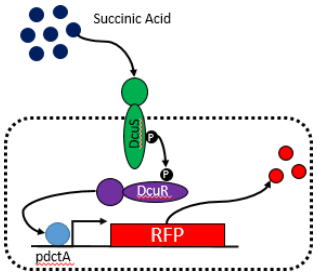

Figure 5.5. Schematic of DcuR/DcuS sensor. The red fluorescent protein (RFP)) reporter is driven by the dctA promoter.

In conclusion, the DcuS/DcuR system with the fluorescent reporter is too sensitive to succinic acid concentrations. Engineering of a DcuS with lower succinic acid affinity or a pdctA binding site with lower affinity for its transcription factor will be needed for the DcuS/DcuR system to provide a high-throughput read out of succinic acid.

Short chain alcohol sensors

Short chain alcohols were also a target as cyanobacteria was set to be engineered for the production of 1,3-propanediol (1,3-PDO) and 1,4-butanediol (1,4-BDO). Towards developing short chain alcohol sensors, we explored OR-based sensors and bacterial allosteric transcription factor-based sensors.

1,4-butanediol (1,4-BDO) Biosensor

OR-based 1,4-butanediol sensor. We leveraged the OR-based sensors constructed for the succinic acid screen, to screen ORs for activation with 1,4-BDO (**Fig. 5.6**). An OR-based sensor was considered a hit if there was a statistically significant increase in fluorescence ($P < 0.005$) in the presence vs. the absence of 1,4-BDO. We identified 23 OR-based sensors hits (**Fig. 5.6B**). *Going forward*, we will need to validate the 23 OR-based sensors hits via a secondary dose-response screen using 1,4-BDO (**Fig. 5.6C**).

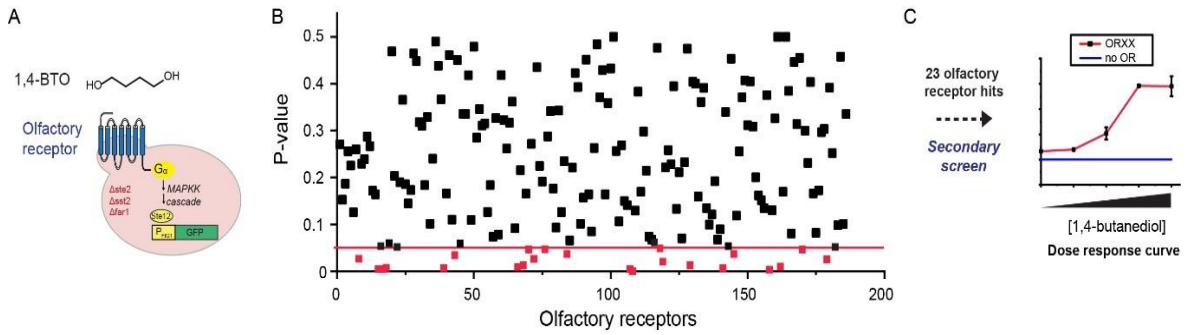

Figure 5.6. OR-based 1,4-BDO sensor development. **A.** OR-based sensor. Human OR (blue) is expressed on the yeast cell surface of the yeast sensor strain (W303 *Δste2, Δsst2 Δfar1*). Upon activation with 1,4-BDO, the OR signals via the yeast G_α subunit and the yeast mating pathway (yellow) ultimately resulting in transcriptional activation of green fluorescence protein (GFP) which is read via flow cytometry. **B**. 186 OR-based sensors were tested in the presence of 1,4-BDO or in the presence of the carrier solvent (DMSO). Twenty-three OR-based sensors (red) resulted in a statistically significant increase (P-value < 0.05) in signal after activation. **C.** In the future, a secondary assay of the 23 OR-based sensor hits will need to be performed to determine if any of the OR-based sensor hits is a *bona fide* 1,4-BDO sensor.

1,3-Propanediol (1,3-PDO) Biosensor

With no previous reports of a 1,3-propanediol sensor, we took a systems biology approach to identify a sensor for this compound. Specifically, we performed an RNA-sequencing experiment to identify *Escherichia coli* promoters that are upregulated in the presence of 1,3-PDO. We identified eight genes that are upregulated in the presence of PDO and we cloned their corresponding promoters upstream green fluorescent protein (**Fig. 5.7A**). As show in **Figure 5.7B**, seven of the eight promoters look promising, showing an increase in fluorescence upon addition of 6% 1,3-PDO. *Going forward*, we need to perform dose-response curves of the seven promoter hits to determine the extent to which they are responsive to PDO by running dose-response curves with different concentrations of 1,3-PDO.

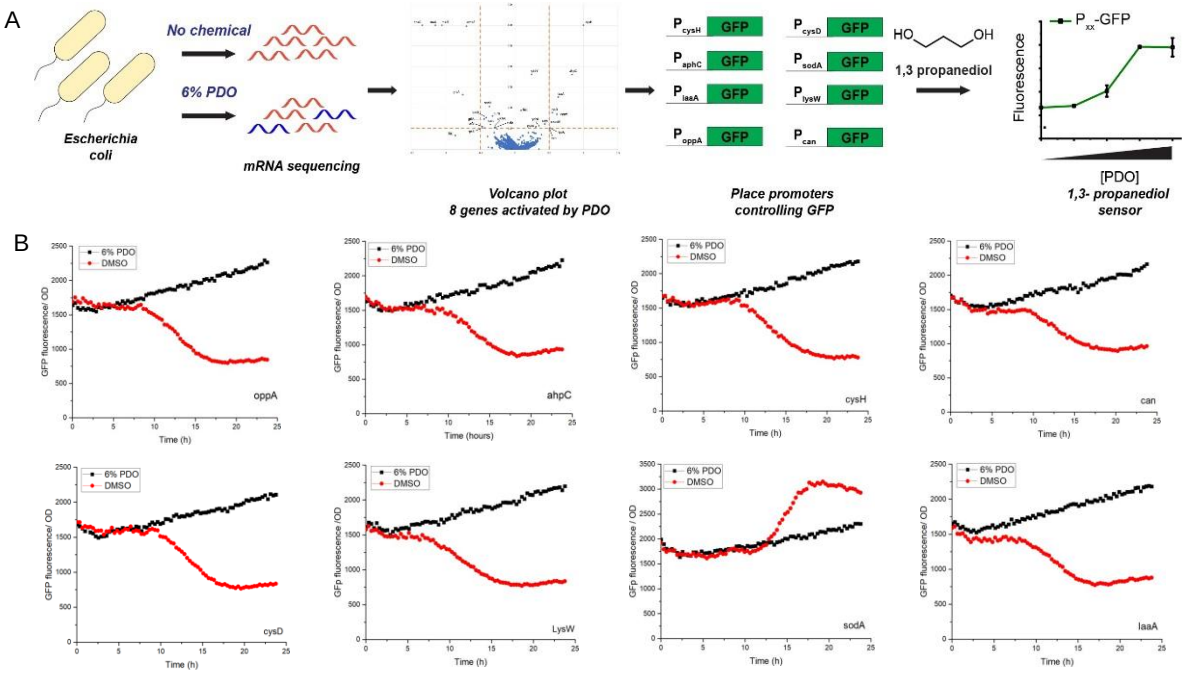

Figure 5.7. 1,3-propanediol sensor development. **A.** Workflow for the identification of promoters upregulated in the presence of 1,3-PDO. *Escherichia coli* was grown in the presence of 6% 1,3-PDO or in the absence of it. Via an RNA-seq genes, genes whose transcription was upregulated in the presence of PDO were identified (volcano plot). The promoters of the top 8 genes activated by 1,3-PDO were cloned upstream green fluorescence protein (GFP) to assess the dependence of GFP expression as a function of 1,3-PDO. **B**. Time curves of the PDO-dependent promoter hits were performed in the presence and absence of 1,3-PDO.

Unexpected challenges during the performance period

Task 5 started in June 2019 and was scheduled for 18 months, until December 2020. However, disruptions caused by the pandemic, including laboratory closure to non-essential research for 4 months (March-July 2019), limited laboratory capacity to observe social distancing (August 2019-December 2019), and a shortage in personnel (September 2021-December 2021) led to unexpected project delays. The stopping and restarting of **Task 3** throughout these disruptions and with different personnel, did not allow us to complete the experiments needed to publish the work. We expect that in the future, we will be wrapping up these projects leading to a publication after the completion of the award.

References for this section

1. Mukherjee, K., Bhattacharyya, S., Peralta-Yahya, P., GPCR-Based Chemical Biosensors for Medium-Chain Fatty Acids. *ACS Synth Biol* **2015**, 4, 1261-9.
2. Sarria, S., Bartholow, T. G., Verga, A., Burkart, M. D., Peralta-Yahya, P., Matching Protein Interfaces for Improved Medium-Chain Fatty Acid Production. *ACS Synth Biol* **2018**, 7, 1179-1187.
3. Yasi, E. A., Eisen, S. L., Wang, H., Sugianto, W., Minniefield, A. R., Hoover, K. A., Branham, P. J., Peralta-Yahya, P., Rapid Deorphanization of Human Olfactory Receptors in Yeast. *Biochemistry* **2019**, 58, 2160-2166.

4. Ehrenworth, A. M., Claiborne, T., Peralta-Yahya, P., Medium-Throughput Screen of Microbially Produced Serotonin via a G-Protein-Coupled Receptor-Based Sensor. *Biochemistry* **2017**, 56, 5471-5475.
5. Marquez-Gomez, P. L., Kruyer, N. S., Eisen, S. L., Torp, L. R., Howie, R. L., Jones, E. V., France, S., Peralta-Yahya, P., Discovery of 8-Hydroxyquinoline as a Histamine Receptor 2 Blocker Scaffold. *ACS Synth Biol* **2022**, 11, 2820-2828.
6. Trauelsen, M., Rexen Ulven, E., Hjorth, S. A., Brvar, M., Monaco, C., Frimurer, T. M., Schwartz, T. W., Receptor structure-based discovery of non-metabolite agonists for the succinate receptor GPR91. *Mol Metab* **2017**, 6, 1585-1596.
7. de Castro Fonseca, M., Aguiar, C. J., da Rocha Franco, J. A., Gingold, R. N., Leite, M. F., GPR91: expanding the frontiers of Krebs cycle intermediates. *Cell Commun Signal* **2016**, 14, 3.
8. Yasi, E. A., Allen, A. A., Sugianto, W., Peralta-Yahya, P., Identification of Three Antimicrobials Activating Serotonin Receptor 4 in Colon Cells. *ACS Synth Biol* **2019**, 8, 2710-2717.
9. Davies, S. J., Golby, P., Omrani, D., Broad, S. A., Harrington, V. L., Guest, J. R., Kelly, D. J., Andrews, S. C., Inactivation and regulation of the aerobic C(4)-dicarboxylate transport (dctA) gene of Escherichia coli. *J Bacteriol* **1999**, 181, 5624-35.
10. Dietrich, J. A., Shis, D. L., Alikhani, A., Keasling, J. D., Transcription factor-based screens and synthetic selections for microbial small-molecule biosynthesis. *ACS Synth Biol* **2013**, 2, 47-58.

Subtask 5.2: Evaluate biosensor function with green algae and/or cyanobacterial producer strain(s).

Supernatants from high performing, flask-based photosynthetic succinate production cyanobacterial strains that were evaluated as part of Task 6.0 were produced at UCSD and provided to the lab at Georgia Tech for evaluation with the biosensors.

Milestone 5.ML.1: A functional biosensor for at least one target production molecule evaluated on production strain(s). 1+ validated biosensors with linear ranges of detection in the mg/L to g/L ranges.

The data presented for Subtask 5.1 demonstrates the efforts made to generate a functional biosensor with a reasonable linear range for detection of succinate, as well as others for PDO and BDO.

Task 6.0: Metabolic Characterization & Optimization of Production Platforms

Subtask 6.1: Generate genome scale metabolic reconstructions for *Chlamydomonas* and *Synechococcus* with incorporation of target molecule production pathways.

The UCSD cyanobacteria and green algae labs accumulated data from published and unpublished genomics, transcriptomics, and metabolomics experiments on *S. elongatus* PCC 7942. These include experimental data sets from heterotrophic and phototrophic growth conditions. All experimental data was shared with the PNNL group to assist in PNNL's task of generating a baseline metabolic model for predicting production yields. The most complete previously published models were also acquired and supplied to PNNL. Further, the previously published flux-based analysis models of *S. elongatus* were modified by the UCSD cyanobacteria lab to include succinate production and secretion, as described for Task 2.0. All models, including this production-modified model, were supplied to the PNNL group.

Genome-scale models for *Chlamydomonas* and *Synechococcus* were generated with incorporation of succinate production pathways based on the previously published models and the production baseline strains described in Task 2.0. We then placed each model into the Memote continuous integration environment and generated baseline reports to determine areas of improvement. In addition, we generated BioCyc pathway/genome databases for *Chlamydomonas* and *Synechococcus* so that we could visualize multi-omics measurements by mapping it to metabolism. These models and maps were based on or adapted to include information from the following references:

a. Feist, A.M., Zielinski, D.C., Orth, J.D., Schellenberger, J., Herrgard, M.J., Palsson, B.Ø., 2010. Model-driven evaluation of the production potential for growth-coupled products of Escherichia coli. Metab. Eng. 12, 173–186. doi:10.1016/j.ymben.2009.10.003
b. Jensen, K., Broeken, V., Hansen, A.S.L., Sonnenschein, N., Herrgård, M.J., 2019. OptCouple: Joint simulation of gene knockouts, insertions and medium modifications for prediction of growth-coupled strain designs. Metab. Eng. Commun. 8, e00087. doi:10.1016/j.mec.2019.e00087
c. Kim, J., Reed, J.L., Maravelias, C.T., 2011. Large-scale bi-level strain design approaches and mixed-integer programming solution techniques. PLoS ONE 6, e24162. doi:10.1371/journal.pone.0024162

We applied the OptCouple algorithm [Jensen 2019] to the iJB785 *Synechococcus* genome-scale model [Broddrick 2018] to obtain 7 different engineered strains such that succinate production was coupled to growth. However, 6 of the 7 knockout sets (groups of genes to be knocked out to produce the predicted yield increases) contain essential genes as determined by RB-TnSeq [Rubin 2015], and the one engineered pathway that did not contain an essential gene knockout was already identified by the UCSD cyanobacteria lab. We also ran the OptKnock algorithm [Burgard 2003] to identify additional gene knockout strategies that are not necessarily growth-coupled, and predicted 9 additional engineered strains, 6 of which contain knockouts of essential genes. Lastly, we ran SimOptStrain [Kim 2011] to obtain 2 additional engineered strains that do not contain essential gene knockouts by adding a heterologous reaction.

For *Chlamydomonas*, we applied the OptCouple algorithm [Jensen 2019] to the iCre1355 *Chlamydomonas* genome-scale model [Shah 2015], but the only solution involved a mitochondrial transport activity whose transport protein is not known. When applying OptKnock [Burgard 2003], we identified an engineered strain that requires blocking oxygen diffusion into the chloroplast. Lastly, we discovered a flaw in the way the OptCouple algorithm is

implemented, and we also designed an objective-tilting algorithm [Bakker 2020] that generates provably better growth-coupled solutions than the OptCouple algorithm. The suggestions provided by OptCouple, OptKnock, and SimOptStrain were not as useful as we had hoped, but this was probably due to the fact that we used a fairly short time window to search for a solution. We also did not use any multi-omics measurements to constrain the wild-type fluxes.

In order to run the algorithms with longer time windows, we obtained access to a high-performance computing server at PNNL (Constance) and a faster mixed-integer programming solver (Gurobi). We revised the OptCouple, OptKnock, and SimOptStrain codes to run the algorithms on the Constance server using Gurobi. The revised codes were tested for parallel job submission successfully using small examples. We are now running the algorithms with *Synechococcus* and *Chlamydomonas* models with more processors and longer time windows. We manually curated the reaction important for succinate production in the *Synechococcus* metabolic model. For example, the reversibility of the malic enzyme reaction was changed from reversible to irreversible in the direction of decarboxylating, and the malate dehydrogenase reaction was removed due to the absence of the malate dehydrogenase gene in *S. elongatus* 7942. We applied the RNA-Seq data to more accurately simulate the heterologous growth of *Chlamydomonas*. The RNA-Seq data was mapped to the metabolic model enzymes using the gene to protein to reaction association from the *Chlamydomonas* metabolic model. We have updated the localization of a few key enzymes including isocitrate lyase in the metabolic model based on this analysis and literature. We are now running the OptForce algorithm from this wild-type flux estimation to obtain predictions on which genes to over or under-express.

References

a. Feist, A.M., Zielinski, D.C., Orth, J.D., Schellenberger, J., Herrgard, M.J., Palsson, B.Ø., 2010. Model-driven evaluation of the production potential for growth-coupled products of Escherichia coli. Metab. Eng. 12, 173–186. doi:10.1016/j.ymben.2009.10.003
b. Jensen, K., Broeken, V., Hansen, A.S.L., Sonnenschein, N., Herrgård, M.J., 2019. OptCouple: Joint simulation of gene knockouts, insertions and medium modifications for prediction of growth-coupled strain designs. Metab. Eng. Commun. 8, e00087. doi:10.1016/j.mec.2019.e00087
c. Kim, J., Reed, J.L., Maravelias, C.T., 2011. Large-scale bi-level strain design approaches and mixed-integer programming solution techniques. PLoS ONE 6, e24162. doi:10.1371/journal.pone.0024162
d. Burgard, A.P., Pharkya, P., Maranas, C.D., 2003. Optknock: a bilevel programming framework for identifying gene knockout strategies for microbial strain optimization. Biotechnol. Bioeng. 84, 647–657. doi:10.1002/bit.10803
e. Broddrick, J.T., Welkie, D.G., Jallet, D., Golden, S.S., Peers, G., Palsson, B.O., 2019. Predicting the metabolic capabilities of Synechococcus elongatus PCC 7942 adapted to different light regimes. Metab. Eng. 52, 42–56. doi:10.1016/j.ymben.2018.11.001
f. Imam, S., Schäuble, S., Valenzuela, J., López García de Lomana, A., Carter, W., Price, N.D., Baliga, N.S., 2015. A refined genome-scale reconstruction of Chlamydomonas metabolism provides a platform for systems-level analyses. Plant J. 84, 1239–1256. doi:10.1111/tpj.13059
g. Ranganathan, S., Suthers, P.F., Maranas, C.D., 2010. OptForce: an optimization procedure for identifying all genetic manipulations leading to targeted overproductions. PLoS Comput. Biol. 6, e1000744. doi:10.1371/journal.pcbi.1000744
h. Bakker, C., Kim, J., Zucker, J. (2020). Cell factory design formulations: Analytical and Computational Results. In preparation.

i. Rubin BE, Wetmore KM, Price MN, et al. The essential gene set of a photosynthetic organism. *Proc Natl Acad Sci U S A*. 2015;112(48):E6634–E6643. doi:10.1073/pnas.1519220112

Milestone 6.ML.1: Metabolic model-based predictions of pathway mutations for enhanced carbon flow towards target molecule production. 1+ predictions based on Task 6 & 7 experimental data integrated into metabolic models for native pathways to alter for increased production yields.

Metabolic models were established for both *Chlamydomonas* and *Synechococcus* and pathway mutations were predicted that would enhance the production of succinate. As described below for Subtask 6.5, these predictions in *S. elongatus* served as the basis of experimental validation.

Subtask 6.2: Metabolomics and transcriptomics comparing non-production and production strains from Task 2.1.

Production samples were collected from the top most performing strains determined through the HPLC analysis described for Subtask 6.5. These strains include WTcm+Fko+pGLKS5 #1, WTcm+Fko+pGLKS+T #5, Lcm+Fko+pSuc5 #1, Lcm+Fko+pSuc5 #2, WTcm+pLAN1-1+19B12 #1 (mutant in Succinate dehydrogenase C, Synpcc7942_0314), and LAN1 + 3F11 #2 (mutant in glucose-6-phosphate 1-dehydrogenase, Synpcc7942_2334), as well as the control strains LAN1 and LAN1+Fko. These strains were grown and induced to produce succinate at the 100 ml scale with higher amounts of IPTG (5 mM final instead of 0.1 mM as in the test tube experiments to ensure more consistent production across strains), as described for Subtask 6.5 below. Throughout the experiment, cell pellet samples were collected and frozen at -80°C. The samples are ready for preparation for Omics studies, including RNA-Seq, proteomics, and metabolomics. However, due to the timing of this task and the shutdown of facilities during the COVID-19 pandemic, these samples were never analyzed as intended, but could be in the future.

In addition to these sample collections, we have analyzed some of the supernatant samples that were analyzed by HPLC as part of Subtask 6.5 using the ECAL Lab's Orbitrap LC-MS/MS and the data collection protocol that we developed with Dr. Neal Arakawa, the director of the Environmental and Complex Analysis Laboratory (ECAL) facility at UCSD. We ran supernatants from our WT, LAN1, and LAN1+Fko strains, in addition to the BDO production strain described in Task 2. This experiment was performed in order to identify the molecules comprising the peaks other than the succinate peak by collecting LC-MS parental and LC-MS/MS fragmentation fingerprints throughout the HPLC run. The collected data has been uploaded into the GNPS database and continues to be analyzed using the molecular networking algorithms on the GNPS server. The goal is to identify molecules associated with peaks commonly observed in the chromatograms of our HPLC runs to identify chemicals and pathways that can be knocked out to improve carbon flow to succinate production.

Subtask 6.3: Head space CIMS analysis of non-production and production strains from Task 2.1.

To adapt the CIMS instrument for the purpose of high-throughput screening of samples, a headspace autosampler was purchased and arrived in the Chemical Analytics lab at UCSD. The intention was to use this device as a high to medium throughput front end device for

aerosolizing or derivatizing samples in order to rapidly and automatedly process samples to the CIMS device for real-time detection and quantification of succinate. However, its adaptation to act as the front end for the CIMS device was stalled by the COVID shutdowns. In spite of this, we were able to demonstrate early on in the project the ability to use the CIMS to detect succinate from an aerosolized sample. In this experiment, samples of 10 g/L succinate in MilliQ water were aerosolized and analyzed by thermal desorption chemical ionization mass spectrometry (TD-CIMS) using acetone as a reagent gas. Succinate at an m/z of 117 was observed, with an intensity approaching that of the reagent gas (typically the most common gas with the highest intensity observed in any CIMS analysis). Further experiments will be performed to determine the lower limit of detection, as well as the ability to detect succinate in media and in conditioned media prior to using this tool for screening clones.

An alternative that was explored is the use of headspace gas chromatography, HSGC, in conjunction with an FID or an MS detector. This is an excellent method for the analysis of volatile organic compounds (VOCs) in bio-oils including light aldehydes and ketones, ethers, esters, alcohols, carboxylic acids and aromatic compounds resulting in better precision and accuracy for the quantitative analysis. [TrAC Trends in Analytical Chemistry, 2020]. This technique is:

1. Very Simple –no to minimum sample prep
2. Rapid and Robust – enhance uptime
3. Minimal carry-over
4. Excellent accuracy and repeatability
5. Allows for investigating volatiles in complex matrices
6. Prevent non-volatiles from entering chromatographic system
7. Enables aggressive detection limits using a cold trap

The application of High Throughput Screening of succinic acid produced by algae (green and cyanobacteria) and transported to the media via HPLC has a duty cycle of approximately 30 minutes. This translates to a max of 24 samples per 12-hour run (shift). Every 12 hours the HPLC should be checked for solvent levels, leaks, and expected performance. At best, this means analyzing 48 sample per day.

Succinic acid is not very volatile, a prerequisite of HSGC. Chemical conversion to the diester increases its volatility significantly. This conversion is very straight forward, although pH dependent. To a liquid sample containing succinic acid, methanol and either an acid or base catalyst is added to produce the di ester – dimethylsuccinate, DMS. The increased volatility of the diester allows it to partition preferentially into the gas phase, eliminating the need to deal with matrix from the growth media or the esterification reagents (the methanol will elute very early). The current elution time for GC-DMS is 3.4 minutes, and this is far from optimized. The initial results indicate that the temperature profile can be more aggressive, which will reduce retention time. We believe this technique can result in a 10-minute duty cycle, which will increase the daily sample throughput to 144 per day.

Following the pandemic, the autosampling system was modified to function with the headspace CIMS instrumentation. Combined with prior sample analysis methods of vaporizing the sample for the CIMS machine, the autosampler allows for analysis time of each sample to be performed in 2 – 3 minutes per sample, which is much faster than the 15 minutes required on the HPLC and is more accurate due to using the mass spectrometry analysis specific to the molecular weight of the succinate. The detection range of this system was analyzed using samples ranging from 0.5 – 40 g/L of succinate. Succinate was detected at all levels. We evaluated the CIMS methodology as compared to a GC-MS methodology that was worked out for analyzing

succinate concentrations from vaporized samples. For the GC-MS methodology, succinate samples are derivatized in a sealed crimp vial at 100°C for 90 minutes with methanol and sulfuric acid to generate dimethyl succinate, which is volatile. The headspace gas can then be injected into a GC-MS using an autosampler. We easily observe DMS in such samples at high sensitivity, given the size of the DMS peak compared to the solvent carrier gas peak. The duty cycle for this method is 6.2 minutes after the derivatization. A minor artifact of the GC-MS method is that the DMS is sometimes fragmented, making quantitation non-absolute. For the thermal desorption CIMS methodology, the soft ionization technique prevents fragmentation and the succinic acid does not need to be derivatized in order to be observed by the methodology. To analyze the sample, it is first aerosolized, passed through a 130°C stainless steel tubing to vaporize the aerosols, and then soft-ionized in the CIMS with acetate as a reagent ion. With this method, succinate can be observed and quantified directly in negative mode. The duty cycle per sample is 3 minutes, approximately ten times less than that for our current HPLC method.

Subtask 6.4: Incorporation of metabolic analysis into the genome scale model and prediction of metabolic engineering mutations to optimize production platforms.

The metabolic model of *Synechococcus* was manually constrained to simulate limited flux in photosystems with low light. Model predicted flux was visualized on a metabolic map for different growth conditions including high vs low light, with or without glucose, and with or without xylose. The effects of glucose or xylose on the Calvin Benson Cycle flux and NADPH generating reaction flux were investigated to explain the improved growth of strains overexpressing oxidative pentose pathway genes and the poor growth of strains grown on xylose. The metabolic model of *Chlamydomonas* was further manually curated to remove TCA cycle reactions without any gene association or with incorrect compartment information based on the most recent genome annotation. The removed reactions include cytosolic citrate synthase, aconitase, and isocitrate dehydrogenase and chloroplast isocitrate dehydrogenase (no associated genes) as well as chloroplastic malate synthase (blocked reaction). A metabolic map was generated to visualize and compare RNA-Seq data and model-predicted flux distribution. From this analysis, a set of reactions with redox imbalance was identified and removed from the metabolic model (a cycle of reactions reducing pyrroloquinoline quinone). We performed OptCouple, OptKnock, SimOptStrain, and OptForce simulations to predict gene knockouts as well as up or down-regulation targets for improving succinate production in *Synechococcus* and *Chlamydomonas* under different growth conditions. The essentiality of predicted genes was evaluated using *Synechococcus* RB-TnSeq data and *Chlamydomonas* Library Project (CLiP) mutant data. We used the manually curated metabolic models of *Synechococcus* and *Chlamydomonas* to generate new predictions of pathway mutations for enhanced carbon flow towards target molecule production. Metabolic model-based predictions for succinate production were made for three different scenarios in *Synechococcus* (high light without glucose, low light with glucose, low light with glucose and limited flux in photosystems) and for two different scenarios in *Chlamydomonas* (high light without acetate and low light with acetate).

The metabolic model of the succinate-producing *Synechococcus* strain was altered to generate a model for 1,4-BDO production based on the production vector described in Task 2. Using this model, OPT-Force and OPT-Knock predictions were performed in order to predict modifications to the strain that can improve production of 1,4-BDO. Although a number of knock outs were predicted to increase flux moderately, a single over-expression prediction was made that predicts a large increase in production yields. This over-expression prediction is for the Acetyl-CoA synthetase (ACS) gene (Synpcc7942_1352), which is predicted to not only improve carbon flow towards the succinate production pathway through the phosphoketolase bypass but to also

generate acetyl-CoA that is otherwise consumed as part of the BDO production pathway. Based on this model, the group at PNNL believes that the intermediate that we have observed in our HPLC trace may be accumulation of 4HB due to a lack of sufficient acetyl-CoA to push the production system further to 1,4-BDO. This may explain the observed phenotypes associated with the BDO inductions described for Task 2.

Subtask 6.5: Generation and evaluation of metabolically engineered production strains based on modeling predictions.

Testing knockouts of SucD and a diacid exporter: The LAN1 strain acquired for Task 2.0 was further modified to test two hypothesizes arising from the work of Lan & Wei. The first hypothesis was that succinate production could be increased by preventing conversion of succinate to fumarate, which was also observed via HPLC as a co-product in the supernatant with succinate. Further, the peak observed in HPLC for fumarate often overlaps with succinate, making analysis of succinate production difficult. To test this first hypothesis, a transposon-insertional knockout in the SucD-encoding gene (Synpcc7942_1533) that had been previously generated and characterized for other purposes by Rubin, et al. and Broddrick, et al. was constructed in the LAN1 background and the induced supernatant was analyzed by HPLC. This strain showed elimination of fumarate from the media and increased yields of succinate in the media, supporting the hypothesis that this mutation increasing production yields. The second hypothesis that was tested is that a transporter encoded by the gene Synpcc7942_0366 is responsible for succinate transport out of the cell. Again, a transposon-insertional knockout was constructed and analyzed for succinate and fumarate production outside of the cell via HPLC. These experiments demonstrated a reduction in both succinate and fumarate from the media following IPTG induction relative to the original LAN1 strain, supporting the hypothesis that this transporter is responsible for the excretion of small diacids. In order to use the results of both of these experiments, vectors were constructed based upon the succinate producing vectors described above for Task 2.0 that also include the gene sequence for the transporter in order to test the hypothesis that constitutive expression of the transporter will increase succinate excretion and yields.

To utilize these findings to generate strains with increased succinate yield, we constructed a series of vectors based upon the pSuc2 vector described for Task 2 that included in various combinations inducibly-expressible copies of 7942's encodings of Synpcc7942_0366, *ppc,* and *gltA.* The inducible expression of Synpcc7942_0366 should allow for more secretion to the media of succinate, while the inducible expression of *ppc* and *gltA* should allow for increased production of succinate as was observed with the *C. glutamica* versions of these genes in LAN3. We also constructed a vector that included a single amino acid mutation in *gltA* that was previously shown in *E. coli* to remove inhibition by NAD and resulted in increased production of 1,4-BDO in a strain constructed by Genomatica. These vectors were subsequently transformed into the UCSD lab's WT PCC 7942 and into a WT PCC 7942 that had a knockout in the fumarate synthase Synpcc7942_1533 (WT+12A9 #1). Initial transformation results and growths of clones indicated that a number of these constructs are leaky and resulted in observable non-WT-like phenotypes in the absence of the inducer IPTG, consistent with the observations of metabolic defects in the Lan and Wei publication and as described above for Task 2.0. Given that all constructs were made with the same backbone and the same promoter system, the interpretation of these phenotypes is as a direct result of the gene combinations encoded on them. The most unusual phenotype was for the overexpression construct of the transporter alone, pTrans, which resulted in blobby colonies that had extremely elongated cells. Interestingly, this phenotype was not observed with the pSuc2+T or pSuc5 transformations, which also encode an overexpression of the transporter, indicating that the presence of the

other genes either inhibits its expression or mitigates the defect that overexpressing the transporter causes.

Testing Metabolic Predictions: Based on an initial prediction of 20 gene knockouts predicted by the metabolic modeling from Subtask 6.1, the UCSD cyanobacteria lab acquired the necessary transposon insertion mutation vectors to attempt to generate these individual mutants. Transformation of WT, LAN1, and WT+pLAN1 strains with knockout vectors for each of the predicted genes was performed. Every transformation, except for the control reactions with water, produced colonies with varying efficiency. This result was unexpected since 6 of those predicted genes were previously determined to be essential and 1 to be beneficial via Rb-TnSeq (Rubin, et al). Because mutations could be generated that were either not fully segregated or did not completely knock out the gene due to single recombination instead of double recombination, each of these strains were analyzed for double recombination and segregation. By streaking colonies on both Cm-containing plates and Km-containing plates, we were able to determine if clones were the result of single recombination, meaning the entire vector was inserted which results in no true knockout of the desired gene, or double recombination, which results in a true knockout of the desired gene in at least one copy of the chromosome. The results from this experiment demonstrated that, for all but 5 knockouts, the predicted essentiality from previous RB-TnSeq experiments (Rubin, et al.) was observed again – meaning that these strains did not fully segregate and still retained WT copies of the genes targeted for knockout. Interestingly, knocking out fumarate dehydrogenase (sp1007) or its neighbor gene sp1008 resulted in a somewhat sick or slow growing phenotype in this experiment, whereas the RB-TnSeq data states that this mutation is non-essential. Another interesting discrepancy is that knockout of a sodium-dependent bicarbonate transporter that was declared beneficial in the RB-TnSeq data set was here found to be non-essential. The only previously determined essential gene observed here to be non-essential was *pgk*, phosphoglycerate kinase.

At a later time, four additional mutants were predicted by the PNNL group to aide in succinate production. All four additional mutants were attempted as had been done for the previous 20. One KO vector for Synpcc7942_1501 encoding D-3-phosphoglycerate dehydrogenase did not produce any reliable clones. The other three KO vectors produced transformants, but all transformants proved to be single recombinants, meaning that the genes were not in fact knocked out. Thus, we determined that the 4 new predictions are for essential genes, all of which had been deemed essential previously from published RB-TnSeq data.

For those mutants that were not true mutants, we attempted to generate mutants in two additional background strains, WTcm+pLAN1-1 (D2-1) and Lcm+pLAN1-1 (F2-1), which were both shown in our previous work to produce succinate at ~300 mg/L/OD. Although we have not been able to generate a full set of mutations in each background, we now have mutations in either LAN1 or these two new background strains for 16 knockout vectors. Thus, we grew up and induced triplicate transformants for 33 different knockout strains, as well as the relevant parental strains, and collected supernatant samples for determining succinate yields. Triplicate 6 mL test tube cultures were induced with IPTG and supernatants were analyzed by HPLC. A number of knockouts displayed increased production, but did so in a strain background dependent manner. The only knockout that consistently increased production independent of the background strain was 19B12, which was a knockout of an alternative succinate to fumarate conversion gene than that in the Fko strains. However, altogether these data suggest a number of knockouts that can be made in a single cell to potentially combine to make an even higher production strain in Task 10.

Optimizing Succinate Production Process: At a time when we had generated a total of 10 succinate production vectors with varying combinations of heterologous production genes, we were enabled to test a number of hypotheses with respect to optimal succinate production and secretion, including the optimal gene set for generating succinate, whether overexpression of the putative diacid transporter would be tolerated and result in improved succinate secretion, and whether a single amino acid mutation of GltA would remove NAD-dependent inhibition and enable higher yields. In preparation for screening a number of novel strains derived from these vectors, we first performed a series of experiments with the LAN1 strains to determine the ideal conditions for enhanced succinate yield. Tested parameters included whether cultures were grown in 6-well plates, test tubes, or flasks; whether growth in the presence of CO2 enhanced yields proportionally to the enhanced biomass growth; and whether pH alterations to the media would lead to higher yields. Overall, we determined that production in higher pH media in test tubes yielded the highest production values. Specifically, the increased initial pH allowed for up to 4-fold improvement in yield in some cases. These results were consistent with the hypothesis that the putative transporter protein is instead a channel and the pH gradient between the inside of the cell and its environment drives the secretion of the small diacids. This is important to note for future scaled production, where steady removal of the diacid will continue to enable high diacid secretion while batch production will eventually lower the pH of the solution and prevent further secretion independent of production.

Evaluating Current Strains: Using the higher pH production conditions, we examined several of the strains we previously generated from the initial 8 vectors constructed. We observed that most of the strains tested did not produce succinate at levels higher than background or that of WT. In some experiments LAN1's production appeared lower than in previous experiments and LAN1+Fko produced much higher yields than LAN1, while not producing fumarate into the media. Interestingly, our WT+pLAN1 strain, which should be genetically identical to LAN1, produced as much as or more than the LAN1+Fko strain. Overall, these results led to two conclusions: 1) The transformed strains did not appear to be producing as expected and may not have been as expected and 2) the parental strains do appear to be different and have an impact on production or stability of production. For the first, we followed up via PCR to check for segregation of the strains and confirmed that the exogenous genes were present. Thus, to test the second we isolated genomic DNA and submitted it for Illumina sequencing to determine the differences between the parental strains. The resequencing data, as analyzed using BreSeq, highlighted 3 SNPs and one possible deletion in the LAN strain as compared to the WT strain. The primary SNP and/or deletion that we found most compelling is that of the Synpcc7942_1475 gene, the sodium dependent bicarbonate transporter gene predicted by the metabolic modeling that we demonstrated here as being non-essential.

Generating and evaluating new transformants: To overcome both the parental strain background issue and the inability to readily evaluate segregation via PCR, we decided to first construct new parental strains from WT and LAN1, or their respective fumarate knockout strains WT+Fko and LAN1+Fko, by transforming these strains with the insertion vector 7H1-C2. This vector replaced either the WT NS1 or the LAN1 NS1 production pathway with a NS1 containing a Cm-resistance gene. Strains were verified by both PCR and segregation tests on antibiotic plates (e.g. LAN1 is SpSm resistant, but Lcm is not). These four new parental strains (WTcm, WTcm+Fko, Lcm, and Lcm+Fko) were then transformed with each of the 10 succinate production vectors described below to generate 40 new strains that were tested for NS1 segregation via antibiotic tests. For each strain, we were able to recover at least 3 clones that were fully segregated, though we noted that some vectors had different transformation efficiencies, segregation efficiencies, or phenotypes that were consistent with the results above for the previous round of transformations. In total, we produced over 120 novel strains.

For each of the desired parental strain and vector transformation combinations, we picked 3 double recombinant, segregated clones (as determined by antibiotic testing), grew the clones in 6 ml test tube cultures, and induced expression of the pathway with greater than 0.1 mM IPTG. After six days of induction, cultures were analyzed for contamination via OMNI plates, for culture density via optical density, for general health of the culture via fluorescence scanning, and for succinate yields in the supernatant via HPLC analysis. Based on these results, we made several important observations: 1) Vectors with the *E. coli gabD* (e.g. pLAN1 & pGLKS) secrete succinate in all strains, 2) Vectors with the 7002 *gabD* only secrete succinate in Lcm backgrounds, and not in WT backgrounds, 3) Overexpression of the putative transporter or diacid channel alone does not secrete succinate, 4) Knocking out fumarate production from succinate (sp1533, Fko) improves succinate yields in all cases and prevents fumarate production (not presented here, but was observed in the HPLC traces, 5) Addition of native gltA and ppc did not appear to enhance production in Lcm backgrounds, which is the only background where these vectors produced succinate, and 6) Some strains did not produce consistently across replicates and we are repeating those strains now to determine if the biological clonal replicates are truly as different from each other as the data suggests.

Of these observations, the most interesting to us was that the parental background strain had such a profound impact on whether or not the expression of 7002 *gabD* gene resulted in any succinate production in the supernatant (observations #1 & #2). This genetic background issue may also have contributed to observation #4, though this result may also be due to native regulation of these native genes that prevents their overexpression from having any true causal effect. To test this latter hypothesis, we ordered as gBlocks the *Corynebacterium glutamicum* versions of these genes as used by Lan & Wei and successfully regenerated the pLAN3 vector so that we can truly compare our results and methods to those of Lan & Wei.

We provided the PNNL group with all of the above data and asked them how the metabolic model could explain how a SNP or knockout of the sp1475 gene would result in the observed succinic production differences. The model demonstrated that, contrary to our belief that as a bicarbonate transporter this protein would increase carbon concentrations in the cell, the sp1475 protein product, SbtA, is actually acting as a carbonate exporter under these growth conditions. Thus, disruption of this gene would actually result in increased carbon concentrations and an increased pH gradient between the inside and outside of the cell. Both of these would result in increased succinate production and secretion. Unfortunately, this does not say why this would have a profound difference for the 7002 GabD versus the *E. coli* GabD. We hypothesize that either the *E. coli* GabD gene is better at utilizing the increased carbon or the 7002 GabD is post-translationally regulated by the photosynthetic state of the cell in a way that the *E. coli* GabD would not be having evolved in a heterotrophic organism.

In order to test this phenomenon further and to ensure that we could test the impacts of additional genes used in our production pathways (e.g. *kgd*, *gltA*, and *ppc*), we constructed 6 new vector constructs that utilize the *E. coli gabD* gene, including the original LAN3 construct using gBlocks. All 6 new vectors were transformed into our four background strains (WTcm, WTcm+Fko, Lcm, and Lcm+Fko) and double recombinant, segregated clones were identified and grown up. Triplicate clones were grown and induced for expression with 0.1 mM IPTG addition in 6 mL test tube cultures. 0.2 µm filtered supernatant samples were collected and analyzed by HPLC. Quantitation and comparison of the succinate production associated with each of the production strains were performed. A number of strains produced significantly higher amounts of succinate than the LAN1 strain, while only one clone appeared to produce more than LAN1+Fko. Surprisingly, no vector consistently produced higher levels of succinate in

a background strain independent manner. Similarly, we noted a number of replicates not behaving consistently or as expected. For example, all LAN3 clones failed to produce succinate while publications said that this should be the highest producing strain. Based on this data, we chose six strains that had the highest overall yield or highest yield per OD to retest at a 100 ml flask scale for consistency of production. These strains were two biological replicates of Lcm+Fko+pSuc5 (G11 #1 and #2), WTcm+Fko+pGLKS5 (strain 54), WTcm+Fko+pGLKS+T (strain 71), LAN1+UGS-3F11 (strain 162/C3 #2; mutant in glucose-6-phosphate 1-dehydrogenase, Synpcc7942_2334), and WTcm+pLAN1-1+UGS-19B12 (strain 233; mutant in Succinate dehydrogenase C, Synpcc7942_0314). We also included two baseline production controls of LAN1 and LAN1+Fko. Because we previously noted that a number of biological replicates did not behave or produce succinate consistently as expected, we altered the media to include higher concentrations of IPTG than previously used (5 mM instead of 0.1 mM) and increased the initial pH of the production cultures to pH 11. Triplicate replicates of each strain were allowed to grow for six days prior to induction, after which succinate production was monitored daily for nine days. As the data demonstrated, succinate production over time was fairly consistent between strains tested, with all strains except 233 and 71 reaching yields of approximately 150 - 200 mg/L. It is unclear at this time why 233 and 71 no longer produced succinate in this experiment, but could be related to media changes. Interestingly, LAN1 and G11 #1 had boosts in productivity earlier than other cultures. When we look at the normalized production of succinate per cell (yield/OD), however, strains became a little more discernable, with the G11 strains clearly out-performing the others in terms of maximal yield/OD.

Sodium-dependent carbonate transporter (SbtA): The resequencing of the LAN1 genome demonstrated that this genome contains a polymorphism in the coding sequence of the Synpcc7942_1475 (SbtA) gene that causes an amino acid change in a key protein binding region. This gene was also predicted as a knockout target by the modeling group. To test if the deletion of this gene's function results in the phenotypic differences associated with the LAN background (i.e. the capacity to produce succinate using the 7002 *gabD* gene), we introduced a transposon insertional knockout vector (UGS 26A12) into background strains and succinate production strains to determine the impact of this gene on succinate production. For example, we previously demonstrated that the WTcm+pSuc2 strain is unable to secrete succinate while the Lcm+pSuc2 strain produces as much succinate as the Lcm+pLAN1 strain. Thus, we wanted to determine if the addition of UGS 26A12 to WTcm+pSuc2 allows it to produce succinate at levels comparable to Lcm+pSuc2. A total of 9 strains were transformed with the UGS 26A12 vector and triplicate clones were grown, induced for production, and analyzed by HPLC as described above and below for the alternative production vector strains. Analysis of these strains demonstrated that addition of the SbtA knockout **did not** result in production of succinate from the otherwise non-producing WTcm production strains. Of all such non-strains tested, only one that had 26A12 added to it, a single WTcm+pSuc2+26A12 clone, had only slightly above detectable levels of succinate. However, for a number of strains that were previously capable of producing succinate, the addition of the 26A12 mutation resulted in increased production over the parental strain for a number of the tested clones, indicating that the metabolic model prediction was correct in this mutation leading to increased succinate production.

Photomixotrophic production of succinate from *S. elongatus* PCC 7942: Modifications to the base strain reported in Task 2.0 above include the addition of sugar metabolism in 7942, regulatory gene knockouts, and the overexpression of enzymes involved in the oxidative pentose phosphate (OPP) pathway Previous work done by our lab shows that sugar assimilation through the OPP increases product titers and carbon fixation in 7942. The increase in carbon fixation is achieved due to the additional carbon from sugars increasing the substrate availability for RuBisCO, which is the key carbon fixation enzyme in the Calvin-Benson-

Bassham (CBB) cycle. We tested the effects of glucose feeding in our base strain by installing a glucose importer encoded for by the gene *galP* and by overexpressing key enzymes in the OPP pathway encoded by the genes *gnd* and *zwf* (Strain 2). The overexpression of these three genes resulted in enhanced succinate titers when compared to our base strain. HPLC analysis confirmed that Strain 3, containing the succinate production pathway (*gabD-kgd-ppc-gltA*), the glucose import and OPP overexpression cassette (*galP-gnd-zwf*), and a knockout of the regulatory protein Cp12, was previously shown prior to quarantine to produce 1.1 g/L succinate with a rate of 0.2 g/L/d from glucose and CO_2 under continuous light conditions.

STRAIN NUMBER	GENOTYPE
2	NSI: *Ptrc*:galP-*zwf-gnd spec*^{R} NSIII: *Ptrc*:*gabD-kgd-ppc-gltA gent*^{R}
3	Strain 2 + *cikA*::*cm*^{R}
4	Strain 2 + *cp12*::P_L*lac01*:*prk-rbcLS*
5	Strain 2 + NSII: *Ptrc*:*rpe cm*^{R}

Further experiments at UC Davis performed succinate production experiments using Strain 2 (NSI: Ptrc:galP-zwf-gnd specR; NSIII: Ptrc:gabD-kgd-ppc-gltA gentR) to test changes in mixotrophic production conditions. The conditions tested were induced with IPTG with and without added bicarbonate (+B+I and -B+I) as well as no IPTG without bicarbonate (-B-I). The preculture from colonies was used to inoculate cultures at OD730 of 0.6 in 10 ml BG11 including 0.1 mM IPTG (for +B+I and –B+I), 10 mg/L thiamine, 20 mM NaHCO3 (for +B+I) and appropriate antibiotics. A target concentration of 5 g/L glucose was achieved in this production media by mixing 14.28 ml of NREL lysate containing ~360 g/L glucose to make 1 L of media. Every 24 h, 10% of the culture volume was removed and the volume was replaced with production media containing 5 g/L NREL lysate and 200 mM NaHCO3 (for +B+I). The media was adjusted to an initial pH of 10.5 for all conditions and was not adjusted for the duration of the experiment. In all conditions the pH initially dropped to 10 after 24 hours. For the condition with bicarbonate the pH then slowly rose to 11 by day 5, whereas in the conditions without bicarbonate the pH continued to drop to around 9.5. The culture supernatant samples were analyzed using HPLC equipped with a Fast Acid column from Bio-Rad, (100 x 7.8 mm, prepacked HPLC organic acid and alcohol column, hydrogen form, 9 µm particle size, 8% cross linkage) using a mobile phase of 5 mM sulfuric acid. Using this amethod, the NREL lysates increase the width of the void peak, obscuring the succinate peak. However, a peak can be seen at the elution time for succinate at an intensity indicative of ~1 g/L concentration for +B+I on day 3. Our analytical methods will continue to be developed using the HPLC. Additionally, we will use a succinate assay kit to quantify titer while we work on procuring an adequate column for this analysis. All cultures were induced at an OD_{730} of 0.6. The condition containing bicarbonate grew well for 3 days before beginning to experience a decline in OD which can be attributed to the increase in pH for this condition.

Finally, the UC Davis lab completed a series of production experiments using the succinate producing strain of *Synechococcus elongatus* PCC 7942 (NSI: *P*trc: *gabD-kgd, Spec*^{R}; NSII: *P*trc: *galP-zwf-gnd, Kan*^{R}). During these experiments we compared two different pH conditions, 7.0 and 10.5. Approximately 600 mg/L of succinate was produced after 6 days in the cultures maintained at pH 10.5 without sodium bicarbonate. However, glucose consumption was not observed at pH 10.5. Although glucose consumption was observed at pH 7.0, the strain produced only 50 mg/L after 6 days. We found that addition of sodium bicarbonate (50-200 mM)

was lethal to the cells at pH 10.5, but improved cell growth at pH 7.0. HEPES buffer is usually used for a production experiment at pH 7. HEPES had a slight detrimental effect on growth at pH 10.5 but no effect on the cultures at pH 7.0.

In the UCSD cyanobacteria lab, parallel experiments were performed thanks to the UC Davis lab providing the Ptrc:galP-zwf-gnd vector for addition to the phototrophic succinate production strains. To date, a glucose-consuming, succinate strain (LAN1+ Ptrc:galP-zwf-gnd) and a glucose-consuming control strain (WT+ Ptrc:galP-zwf-gnd) have been produced and tested for growth and succinate production using 40 g/L, 20 g/L, 10 g/L and 5 g/L NREL cellulosic hydrolysates. Strains did not survive at 40 g/L or 20 g/L, which was inconsistent with the tolerance data presented in Task 2.0, where 20 g/L was tolerated by strains. We believe this discrepancy was due to the media being adjusted to pH 11 or 10.5 in this experiment, whereas in the pH of the media was previously adjusted to 8.5 for the previous experiment. Nonetheless, strains grow well in BG-11+5 mM IPTG + pH 10.5 media containing either 10 g/L or 5 g/L of the NREL sugars when shaking under 250 µE lights. Strains continue to grow even after 6 days of induction. A notable difference to the experiments performed in the UC Davis lab is that bicarbonate is not added in these experiments, so that the growth observed here is more akin to the -B+I culture of the UC Davis lab. We observed that, upon testing the pH on day 3, the pH of all of the WT+ Ptrc:galP-zwf-gnd cultures had dropped to approximately 9 – 9.5, while all of the LAN1+ Ptrc:galP-zwf-gnd cultures had dropped to approximately 6 – 6.5, and yet continued to grow. We believe that this pH difference is due to the secretion of succinate, which, if true, indicates that a fairly large amount of succinate has been secreted to the media.

To meet our required yield levels for our stage gate Go/No-Go goal of 1 g/L, we continued to work with and optimize the production conditions of our baseline mixotrophic strains: a glucose-consuming, succinate strain (LAN1+ Ptrc:galP-zwf-gnd) and a glucose-consuming control strain (WT+ Ptrc:galP-zwf-gnd). In our repeat experiments, we first tested the impact of maintaining or correcting the pH of the culture to pH 10.5 on a daily basis versus allowing the culture pH to drift away from an initial pH of 10.5 as the culture proceeded from an initial OD of ~0.5 in media containing 10 g/L of the NREL cellulosic hydrolysate. Without daily pH adjustments, we observed that the pH of the succinate producing cultures rapidly dropped to below pH 6, causing the cultures to die after approximately 7 to 9 days, while the non-succinate producing culture initially dropped and then recovered to above pH 8. Thus, we presume this pH drop is both due to glucose metabolism and the production of succinate, though tolerance tests tell us that the concentration of succinate alone could not have caused toxicity. Further, upon analyzing the succinate concentrations of the pH-adjusted cultures as compared to the unadjusted cultures, we noted a decrease in yield of greater than 60% for those strains that were not maintained at a high pH. This is consistent with the hypothesis that alkaline conditions enable secretion of succinate and the drop in pH observed for the unadjusted cultures reduced their secretion yields. Following this initial 11-day production experiment, pH-adjusted cultures were spun down to pellet the cells, those cells were then washed with standard BG-11, and then resuspended in fresh media with 10 g/L of the NREL cellulosic hydrolysate. Cultures were allowed to produce for another 11 days before a second media replacement was performed and the cultures were allowed to produce for another 8 days. Throughout the 30-day experiment, succinate producing cultures continued to grow until an OD of 8 was reached at approximately Day 18. Interestingly, the control strain that did not produce succinate grow significantly more slowly and never reached the densities of the succinate producing strains, consistent with previous publications that secretion of compounds can drive photosynthesis or metabolism in general faster and that product generation does not inhibit, but instead may enhance, biomass accumulation.

Throughout this experiment, we observed succinate peaks in our HPLC analysis that far exceeded previous values and caused the succinate peak to overlap with a cyanobacterial-specific peak that previously eluted prior to the succinate peak. This overlap disrupts our ability to accurately quantify the yields. To overcome this problem, we acquired an academic license for the AIST software PFchrom, which can deconvolve overlapping peaks and accurately estimate their peak areas. Applying this software to our time course data demonstrated that the cultures reached a yield of approximately 2.7 g/L in the first round of production, over 4 g/L in the second round of production, and approximately 5 g/L in the third round of production. In the first round, productivity was linear throughout at a value of 0.27 g/L/D. In the second round, productivity was generally linear as well with an average productivity of 0.4 g/L/D. In contrast, the third round of production appeared to proceed slowly at first, followed by an exponential production of succinate and a flattening of productivities after 6 days of production. While each round exceeded the 1 g/L goal of the intermediate verification, the total succinate produced, 12 g per L of culture, is over halfway to the final goal of the project.

Subtask 6.5 – Photosynthetic and heterotrophic production of succinate from *Chlamydomonas*: For enhancing succinic acid production under photosynthetic conditions, we introduced a sugar transporter derived from *Chlorella* into *Chlamydomonas reinhardtii*, where it is expressed in the periplasm. The sugar transporter is linked to a fluorescent reporter gene, mClover. High expressors of the sugar transporter were screened using high-throughput fluorescence intensity measurements and western blots for mClover. The recombinant *Chlamydomonas reinhardtii* was inoculated in tap media+20 g/L glucose, as well as alternative *Chalmydomonas* species, were investigated for their ability to produce succinate in the media. Overall, we found that *Chlamydomonas debaryana* UTEX 231 produced higher amounts of succinate than other strains, including *Chlamydomonas asymmetrica* UTEX 227.

Fifty-four strains of green algae were screened for their ability to grow on glucose and/or xylose. 24 strains were shown to be capable of growing on glucose, while none were able to grow on xylose alone. These strains were also capable of growing on the NREL hydrolysates as well. Strains capable of growing on the NREL hydrolysates were checked for succinate secretion using our HPLC methodology, which indicates that a number of strains naturally produce > 1 g/L of succinate into the media. This was also demonstrated to be the case with pure glucose added to the media. However, overlaps with other peaks prevented absolute quantitation of the NREL hydrolysate cultures, as observed for the mixotrophic production of succinate from cyanobacteria. Attempts to use the ASIT software to deconvolve these HPLC traces using our standard method of analysis failed due to too much overlap between the succinate peak and a peak associated with the green algae. We therefore explored alternative methods of quantitation. We found success using a hydrophilic interaction liquid chromatography (HILIC) column, which was able to cleanly separate succinate from other peaks observed in the green algae. Analysis of older cyanobacterial supernatant samples containing succinate also found good separation of succinate from native cyanobacteria peaks, even at the higher concentrations of succinate previously observed in the mixotrophic experiment. Thus, we have opted to switch all future analyses to this novel method as it appears to be better for quantifying succinate in both algae and cyanobacteria samples. Finally, we generated reproducible standard curves to convert succinate peak area to concentration in the presence of F5+glucose media, the media primarily used to grow the algae heterotrophically. However, in some cases, we observed inconsistent results with different batches of media in which some detectable and quantifiable amount of succinate appeared to be present in the media itself. To eliminate the possibility of biological contamination and determine the source of the sporadic succinate peak, which could be accounted for each time an experiment is performed by including appropriate media blanks, we examined different components comprising the recipe of the F5 media using

the HILIC column. This examination resulted in determining that the yeast extract is the source of the contaminating succinate peak, which makes sense in that our biosensor work is also confounded by the native production of succinate in yeast. Because yeast extract is poorly homogeneous with respect to the analytical sensitivity of the HPLC assay, future experiments will attempt to avoid the use of yeast extract for media preparation whenever possible.

After establishing our new methodology using a hydrophilic interaction liquid chromatography (HILIC) column, we reevaluated all strains. Although no strains reached the same levels of yields as had been previously believed, a few of the strains had yields in the 0.5 – 1 g/L range. These strains have been chosen for further analysis, as stated above, for metabolomics analysis. However, it was not clear from these analyses that this line of inquiry would prove fruitful and it was decided to down-select away from the *Chlamydomonas* succinate production strains.

Task 7.0: RB-TnSeq Mutagenesis Analysis & Optimization of *Synechococcus* Production Platform

Subtask 7.1: Generation of RB-TnSeq libraries in *Synechococcus* with production pathways

Randomly bar-coded transposon sequences (RB-TnSeq) and the related interaction RB-TnSeq (IRB-Seq) experiments are extremely powerful screens and selections that provide a wealth of information from relatively quick and easy to perform experiments. However, to properly perform these experiments requires considerable design and planning to ensure that all appropriate controls and the best possible experimental conditions are used. For this experiment, the existing RB-TnSeq library constructed in PCC 7942 was originally going to be transformed with just one of the succinate production vectors described above for Task 6.0. Originally, this vector was going to be the pSuc4 vector. However, due to the results in Task 6.0 where we determined that pSuc4 does not allow expression in the lab's WT background due to the presence of the 7002 *gabD* gene and because the presence of two native S7942 genes likely leads to the incorporation of the genes into the chromosome in a location other than Neutral Site I, we determined that pSuc4 would not be an ideal candidate vector for generating the IRB-Seq library. This caused us to pause and plan out a much more ambitious, but fundamentally complete, set of experiments to evaluate the current RB-TnSeq library for tolerance to the chemicals being produced and for generating additional IRB-Seq libraries, including a proper empty vector control.

Four library freeze stocks were revived according to published practices to produce 1.2 L of the RB-TnSeq library at an OD_{750} of ~0.8. This pooled library was then sampled for four replicate Time 0 (T0) samples and then used to inoculate RB-TnSeq chemical tolerance studies and a series of IRB-Seq transformations. For the chemical tolerance studies, 33 total flasks containing 40 ml of culture were inoculated at an OD of ~0.1. These 33 flasks represent triplicate cultures for 11 different culture conditions, including standard BG11, BG11 adjusted to pH 11 with NaOH, and the pH-adjusted media with varying concentrations of succinate, 1,4-BDO, or PDO. The chemical concentrations tested were either 1 g/L, 5 g/L, or 20 g/L. Samples were taken every other day after inoculation and pellets were frozen at -80°C until all samples were collected and could be processed for barcode generation for Illumina sequencing. Interestingly, no library cultures could tolerate 1,4-BDO at 20 g/L, while all other cultures continued to grow in the tested media.

In addition, this pooled library was used to perform transformations of seven vectors in total, each introduced individually, into the RB-TnSeq library. These vectors include the pLAN1, pLAN3, pGLKS5, and pBDO vectors for succinate or BDO production. As controls for these vectors, we also transformed with two separate empty vectors and a constitutive sfGFP production vector as a control for protein production. Transformation efficiency experiments were initially performed to approximate the technical requirements of each transformation for obtaining the appropriate number of colonies. These initial tests indicate that some of the vectors (pLAN3 and pBDO) would not transform well. Consistent with our transformation efficiency experiments, the transformations with pLAN3 and pBDO did not produce as many colonies as the other transformations. Nonetheless, we were able to grow colonies for all transformations and pool the appropriate plates together to generate triplicate pooled libraries for each transformation – 21 pooled cultures altogether. Samples from each of these pooled cultures were taken to allow us to sequence for barcodes and determine those mutations that have been lost due to the transformation of the heterologous expression vector. Samples were also frozen down to generate revivable freeze stocks for each of the pooled libraries so that the

libraries could be screened at a later time for production yields. Subsequent to this, duplicate cultures of each of the pooled library were inoculated, grown, and placed under production conditions (BG11 + 5 mM IPTG + pH 10.5; 42 cultures total). Samples were allowed to incubate under production conditions for approximately one week, during which time samples were taken every other day for HPLC analysis and for barcode sequencing. HPLC analysis of the samples clearly demonstrated that the pooled library transformed with pLAN1 was capable of producing succinate, while none of the other pooled libraries produced succinate.

Subtask 7.2: Sequencing of RB-TnSeq libraries and evaluation of populations under production conditions

Samples from both the RB-TnSeq library for chemical tolerance and the IRB-Seq libraries for tolerance to generating the product, a grand total of 514 samples, were processed to extract genomic DNA, PCR amplify bar codes according to standard RB-TnSeq methods so that amplicons were generated with appropriate Illumina adapters, and PCR-purified for Illumina sequencing. Samples were prioritized according to the likelihood of informative data resulting from them and based upon the priority for succinate tolerance and production, so that a subset of 322 samples were submitted for sequencing.

General Analysis: Sequencing data on 322 samples, amounting to over 150 GB of raw sequencing data, was analyzed using modified scripts from the previously published RB-TnSeq analysis methodologies developed by Wetmore, et al. (doi: 10.1128/mBio.00306-15) and Rubin, et al. (doi: 10.1371/journal.pgen.1007301). We first ran scripts derived from those of Wetmore, et al. to extract bar codes and bar code counts from the raw sequencing data. This very large table of information was then passed through a novel script we wrote in R to perform the analysis developed by Rubin, et al. on this large a number of samples and experimental comparisons in a rapid manner. Rubin, et al.'s scripts were developed to determine gene fitness and false detection rates (FDR) scores using a nested linear mixed effects models when comparing a set of experimental replicates (typically 3) against a set of control experiments (typically 3 replicates), as normalized against a set of Time 0 control samples (typically 4 replicates). Our scripts were written to perform this comparison given any number of replicates per group and any number of comparisons. In total, to analyze the tolerance experiments, the transformation experiments, and the induced production experiments, a total of 73 different comparisons being made across various experimental conditions, each of which involved data from 10 – 18 experimental samples. In addition to producing the table of fitness and FDR scores for each non-essential gene in the genome, our scripts produce static and interactive volcano plots of these data for each comparison. We are currently enhancing these scripts to also compare data across comparisons. Another alteration of our scripts are the option to use the traditional arbitrary fitness threshold of 1 for determining gene hits or automatically setting the fitness threshold based on the highest absolute fitness value of the set of genes that did not show statistically significant FDR values (FDR > 0.05). For the data presented in this report, hits are based on the automatically determined FDR-dependent threshold.

Chemical Tolerance: The first set of experiments to be analyzed were those for analyzing the tolerance of the cyanobacteria to an elevated pH growth media and to varying concentrations of succinate, PDO, and BDO. For all of these experiments, samples were taken 2 and 6 days following culture inoculation. When comparing the effect of increased pH (11 vs. 7.5), we see only four gene knockouts that improve growth in pH 11 on Day 2. Three of these are plasmid-encoded genes and if we look at the genes classified as "neutral up"(FDR < 0.05 but fitness < 1 or auto-threshold), most of the higher fitness"neutral up" genes are plasmid-encoded. Only a single gene knockout inhibits growth in pH 11 as compared to pH 7.5, that in the gene 0401

(dgkA) a diacylglycerol kinase. After 6 days, a number 10 gene knockouts appear to be inhibiting growth at pH 11, including this same dgkA gene and an operon of conserved hypothetical genes (0416, 0417 – an ATPase, and 0418). By Day 6, the benefit of the plasmid gene knockouts appears to be no longer significant, but the initial non-plasmid knockout gene of 0083, a hypothetical protein, is still beneficial. Of interest to the analysis is that the co-cistronic genes of 0082 and 0084 do not display a significant phenotype, demonstrating the precision of this methodology for associating phenotype with genotype. In addition to 0083, 5 additional genes appear to be beneficial to growth at pH 11 when knocked out. Identifying the genes important for growth under this pH difference is important for then analyzing the further impact of the desired diacids and diols. Analyzing the impact of succinic acid on the TnSeq library, showing increasing succinate concentration going from left to right and increase time in going from top to bottom), demonstrates that time and concentration both impact the degree to which knockout mutations are critical to survival and growth in the presence of succinate, with more genes demonstrating both fitness and significance values that are over the hit-determination thresholds with increases in both concentration and time. Some of these gene knockout hits appear in multiple time points or concentration points, including 0916 (unknown function), 1396 (unknown function), 2144 (nuclease), and 1174 (psbJ, a photosystem II gene whose non-essential role is still unclear). Similar observations were observed for BDO and PDO tolerance, though with different gene knockouts allowing for enhanced growth or tolerance to these chemicals. For BDO, many of the genes associated with improved survival in pH 11 once again appeared as hits, including dgkA and 0416 – 0418. Novel gene knockouts appeared to be beneficial in high concentrations for both BDO and PDO, including 1628, a conserved hypothetical gene, and 0319, a rRNA methyltransferase.

Tolerance to presence of production genes (post-transformation): The RB-TnSeq library was separately transformed with five different production vectors to identify genes tolerant to the presence of the production vector, as compared to two empty control vectors. All plasmids used to perform these transformations had the same backbone for neutral site I (NS1) chromosomal integration and antibiotic resistance gene (*aadA*, resistance to spectinomycin and streptomycin) for selection of transformants. The two control vectors were an 1) empty vector without any promoter or expression cassette and 2) an empty Ptrc vector that expresses the lacQ gene to regulate a Ptrc promoter, though no gene was encoded for the Ptrc promoter to turn on. The expression vectors included a constitutive GFP expression vector, three Ptrc-regulated succinate production vectors (LAN1, LAN3, and pGLKS5), and Ptrc-regulated BDO production vector. From previous experiments, we know that this Ptrc promoter system is leaky, hence we expected some degree of expression in each of these systems that could lead to intolerance or impaired cell growth in the presence of some of the RB-TnSeq knockout mutations. When comparing the control vectors against each other and the GFP expression plasmid following just transformation, we see minimal impact on insertional mutant survivability, as expected for our controls. Only when comparing the GFP vector to the empty vector (but not the empty Ptrc vector) did we see significant fitness changes, with only one gene knockout improving growth (2182, a small hypothetical protein) and seven limiting growth. When we look at the succinate vectors, as compared to either of the two empty vectors, we see very different results depending on the vector of interest. For example, for LAN1 (two genes added for succinate production), most knockout insertions limited the growth or prevalence of the strain in the transformation pool as compared to either of the empty vectors, while the opposite was true for LAN3 (four genes added for succinate production and carbon flow to the TCA cycle). For pGLKS5 (five genes added), the distribution was more evenly distributed between negative and positive impacts. It is difficult to surmise the interpretation of these transformations given these results.

Induction: Each transformed library was then grown and split into parallel flasks, half of which were induced for expression and half of which were not. Thus, we could make the same type of comparisons as above for chemical tolerances or vector transformations, but this time between induced and uninduced cultures, using the initial growth cultures as the Time 0 controls. Inductions were sampled and examined 4 and 9 days following induction. For the empty vector controls and the succinate production strains, we observed very few significantly high fitness gene knockouts on either day of expression, indicating that few genes impact tolerance or survivability under these production conditions. One gene that appeared as a significant hit was 0319, whose knock out improved growth for succinate production but decreased growth for the empty Ptrc vector library. This gene is the same rRNA methyltransferase observed above that provided tolerance in the presence of high concentrations of BDO or PDO, so likely this is a gene whose knockout generally improves cell viability under stress conditions, potentially through changes in translation. In contrast, induction of the BDO system, which is known to stall growth of the cells possibly through accumulation of a toxic intermediate, identified a number of gene knockouts that improved growth of the strain under these conditions. It is not yet clear why so many gene knockouts help survival under these conditions, but likely it is to prevent expression of the pathway. These genes may thus be reasonable targets for understanding how the cell prevents heterologous pathway expression that causes undesired genetic instability or genetic drift that ultimately leads to decreased production yields.

Milestone 7.ML.1: Set of genes enhancing tolerance to precursor production based on RB-TnSeq analysis. Completed analysis of RB-TnSeq production library for gene mutations with enhanced fitness under production regime.

The above description for Subtask 7.2 outlines several genes that impact tolerance and potentially productivity for generating precursor molecules.

Subtask 7.3: Screening of RB-TnSeq libraries using biosensors and/or mass spectrometry methods for mutations leading to increased yields

Due to delays in Task 5.0 and Subtask 6.3, the technologies needed to perform this high-throughput screening of the RB-TnSeq and IRB-Seq libraries were not available during the course of this project. However, because the libraries were frozen and can be revived, these libraries are available for screening once the appropriate technologies are available.

Task 8.0: Constructing Advanced Synthetic Promoters - Round 2

The ~3,000 promoter sequences provided by PNNL as a result of Task 4.0, which included a subset of 300 that were present in all sets of promoters designed by different methods, were ordered for synthesis by Twist via LBNL. All were successfully synthesized and returned to LBNL for quality control assessments. Once the quality was determined to be as good or better than the first library constructed in Task 3.0, the LBNL group cloned the library into the same GFP expression vector, pBEEPS1, as was used for Task 3.0 and delivered the promoter library to UCSD.

Milestone 8.ML.1: Second DNA library for advanced synthetic promoters and production pathway expression. Synthesized DNA library with 300+ synthetic promoters and corresponding GFP and/or production pathway expression vectors.

Synthesized DNA library with 3,000 synthetic promoters and corresponding GFP expression vectors were generated and delivered to UCSD.

Go/No-Go Decision Point GN.2 Project-Wide: Based on data from Tasks 2 – 8, if no strains were generated capable of producing >1 g/L of a target molecule or could feasibly be generated based on combining genetic alterations, then the project will be terminated.

As detailed under Task 6, we successfully demonstrated the ability to generate > 1 g/L of succinate with multiple biological replicates of the strain LAN1+pAL1515, which is capable of generating and secreting succinate due to the presence of 2 exogenous genes (*kgd* and *gabD*), growth in pH 10.5 media, and the ability to ingest and utilize glucose due to the exogenous expression of a glucose transporter and two glucose catabolism genes. In any single round of growth and expression with 10 g/L NREL cellulosic hydrolysates added to the media, these strains resulted in 2 – 5 g/L of succinate, which combined resulted in the generation of ~12 g/L of succinate. Further, all data from Tasks 2 – 8 supported that the project was proceeding successfully.

This data and details of all tasks completed to date were presented to the BETO site visit team on April 12, 2021. Based on this presentation and all data supplied, the site visit team verified that this project met the Go/No-Go milestone and the program manager recommended proceeding to Budget Phase 3.

Task 9.0: Screening Advanced Promoters - Round 2

For the library comprising the 300 promoters (A-library) that appeared in each of the four design methods used, 7 transformations were performed. 8 transformations were performed using the larger 3,000 promoter library (B-library). For each transformation, a portion of the library was grown on Hyg antibiotic while another was grown on Hyg+Zeo. In total, over 81,000 transformant colonies resulted from these transformations.

These colonies were collected and analyzed using FACS to sort the cells by expression level prior to preparing sequencing samples to identify the barcodes and promoter sequences present in the top performing fractions. Three different parameters were analyzed: single selection using Hyg, double selection on both Hyg and Zeo antibiotics, and selecting the top 10 % top expressors from the whole library. DNA extractions were performed. 2-step PCRs were carried out to acquire the amplicons and add the bar codes necessary for Illumina sequencing, as previously described for Task 4.0. PCR success was confirmed by gel electrophoresis of the post-PCR amplicons. The samples were submitted for sequencing and the resulting sequencing data were analyzed to determine barcode and integration event counts. These counts were sent to the national labs to direct a third round of machine learning. Simultaneously, the resulting promoters that appeared in the top 10% of expressors were investigated for the presence of promoter motifs by a computational group at UCSD. We found little evidence of retention or accumulation of expected promoter motifs amongst the high expressing promoters. To test if this is a failure to investigate enough of sequence space or truly a result of finding improved promoter sequences, designs were assembled to digitally evaluate if the third round of promoter libraries should include purposefully designed promoters that have up to 5 copies of known promoter elements, as detailed below.

To further evaluate if multiple rounds of DBTL are generating promoters with increased expression or are more frequently predicted promoters with greater expression than the baseline synthetic promoter, but not necessarily increased individual promoter strengths, individual clones were picked, grown, and screened using plate-based fluorescence screening with a Tecan plate reader. High expressing clones were selected based on the fluorescence measurements, their promoter and bar codes were PCR amplified, and were submitted for Sanger sequencing. These same clones were also evaluated for expression strength under 5 different conditions: 1) Tap media + light, 2) Tap media in the dark, 3) HSM + light, 4) a salt water media + light, and 5) under nitrogen depletion + light. Under constant light conditions, it is clear that the average expression, as calculated as a fold increase over expression of the standard synthetic promoter AR1, has more than doubled for the B-library as compared to the first round, and has nearly doubled for the larger A-library as compared to the first round, indicating an overall improvement in the population of promoter under investigation. Further, the higher increase of the B-library over that of the A-library shows that combining all predictors during the design phase does result in a more successful pool of promoter sequences. This was consistent with an early observation at the time of the initial FACS analysis that the 300 promoter library provided a greater frequency of higher expressing individuals than the 3,000 library.

Interestingly, this same result was true when tested under alternative conditions of expression. When specifically evaluated for the frequency of strong promoters (fractionated for analysis as those with >10x, those with 7-10x, and those with 5-6x increases in expression over WT) in each library under varying conditions, the frequency of super strong promoters is increased in the second round of promoters when compared to the first round. However, the maximal strength of high-expression outliers ((maximal expression over AR1 of 2.49) has not necessarily increased substantially from the first

round (maximal expression over AR1 of 2.23). This is further evidence for investigating the lack of expected promoter elements in the next round of promoter libraries.

Milestone 9.ML.1: Second round of synthetic promoter and expression libraries integrated into *Chlamydomonas*, generating at least 10,000 individual transformants. 10,000+ individual transformants resulting from electroporation of 8.ML.1 expression library.

Over 81,000 individual transformants were generated through the repeated library electroporations, as detailed above.

After fully analyzing the sequencing data, we found little evidence of retention or accumulation of expected promoter motifs amongst the high expressing promoters, nor did we observe a more general increase in promoter strength amongst the library tested, as compared to previous libraries. Although more promoters among those in the library were found that had higher expression levels, sheer intensity of the maximal expression promoters did not increase significantly. To test if these promoters truly had an increase in expression levels, top promoters and the standard synthetic promoter were individually retransformed into *Chlamydomonas* cells and pooled libraries of transformants of individual promoters were analyzed for expression levels via FACs. These analyses demonstrated that the population of library-derived promoters were not significantly different in terms of promoter strength as compared to the previously published synthetic promoter when the top 10% of integration events for each promoter were compared to each other.

To test if improvements in promoter strength can be achieved through purposeful design of promoter sequences, as opposed to machine learning algorithm derived promoter sequences, we developed a new round of promoter designs based on the idea of generating promoters based on multiple copies of known promoter elements. Promoter elements were 8 – 10 bp long and one to three copies of each element were present in each 200 bp promoter design. A computational model of elements was developed based on the POWRS and STREME publications and a model of the background sequence between elements was developed based on the background probabilities from previously analyzed promoter sequences. In total, approximately 2,000 synthetic promoters were designed using this methodology. These promoters were analyzed using the promoter scoring algorithm at PNNL. This analysis indicated that these promoter designs were no stronger than those produced previously by the machine learning algorithms, indicating that either these designs are truly no better than those previously generated or the scoring algorithm does not recognize these promoter elements as being of significance to the expression system. We decided to test this hypothesis through production of the designed library, transformation, and evaluation of the integration events as performed previously.

To ensure that promoter strengths will be tested appropriately, a novel cloning vector was designed based on the cloning and sequencing strategies presented previously with some extra controls for the system. One such control was the use of a background-only promoter-containing vector called pLibrary-0 (pL0). This background vector acts as the control for comparison, as it does not contain any promoter elements and simply contains a 200 bp randomly generated background promoter sequence. As further controls, two more vectors were generated to act as positive controls: 1) pLR, which contains the RBCS2 promoter derived from the *Chlamydomonas* genome and 2) pLS, which contains the SAP11 synthetic promoter. These new constructs were generated to ensure that the expression from the promoter, which drives Ble and GFP expression to allow for zeomycin resistance and GFP fluorescence, do not influence the expression of the Hygromycin resistance cassette. In previous constructs, the *hyg* gene

was downstream of the promoter. In this new construct, *hyg* is both upstream and encoded in the opposite orientation from the promoter system.

To test these promoters for the expected expression strengths, these three vectors were transformed into *C. reinhardtii* and over 1000 colonies from each transformation were picked, pooled into liquid cultures, and analyzed using FACS. Positive and negative controls for this experiment included a clonal population of a previously transformed AR1 construct and WT cells, respectively. Cell populations analyzed on FACS were separated into top 10, 1, and 0.1 % fractions for analysis, which clearly demonstrated the expected relative promoter strengths of each of the promoters in each of these fractions. For each promoter, average promoter strength increased in correlation with the stringency of the selection criteria. Across promoters, pLR had the highest expression levels in each subpopulation (7 – 13 fold higher than WT), pLS and the positive control were approximately the same in expression level at about 5x higher than the WT negative control (~5 – 8 fold higher than WT), while pL0 was only slightly higher than the WT negative control (~1.6 – 1.9 fold higher than WT). Because the promoter designs meet specification and the control and cloning vectors acted as expected, sequences and vectors were sent to LBNL and Twist for synthesis and vector construction of a third round of promoters.

This third round of promoters was synthesized as previously described in Tasks 3.0 and 8.0 and were provided to UCSD. This third round of promoters were transformed into *Chlamydomonas reinhardtii*. Following transformation, cells were selected on either Hyg or Hyg+Zeo. Following selection, pools of transformants from the different pools of promoters and selections were sequenced using NGS. When selected with Hyg, which was just selected for integration, nearly all promoters that were designed were present in the sample and fairly distributed as per expectations. One particular promoter was overrepresented, which as it turned out was an accidental feature of the presence of one of the control vectors in the design of the library – thus indicating that the transformation well represented the initial synthetic promoter library. When subsequently analyzing the promoters after Hyg+Zeo selection, the distribution of promoters was very different, with the majority of the promoters being fractionally abundant (< 0.1% of the population for each promoter, highest being 0.7%) and a select few of approximately 50 – 100 out of the 2000 designed promoters being well-represented (>1 %) in the sequencing pool. Because Zeo selection selects for promoter strength, this demonstrates the expected outcome that only a few of the designed promoters would have high expression values. These pools are currently being analyzed by FACS and by picking individual colonies for testing in 96 well plates. A fourth and fifth round of promoters was quickly designed based on these results, submitted to LBNL, and synthesized within the remaining time frame of the project. These promoter libraries will continue to be tested as performed for the previous libraries.

Task 10.0: Integration of Optimizations and Generation of Most Promising Production Strains

Subtask 10.1: Generation of most promising production strains of *Chlamydomonas* based on baseline strain (Task 2), highest yield promoter systems (Tasks 8 & 9), and metabolic predictions (Task 6)

Due to successes with production in *Synechococcus*, and the lack of a *C. reinhardtii* strain that produced succinate at yields greater than 100 mg/L, as described in Tasks 2.0 & 6.0, we down-selected to focus on succinate production from *Synechococcus*. This also means that Subtasks 11.1, 11.3, 11.5 & 11.6, where co-cultures of green algae and cyanobacteria were to be performed, did not proceed as originally planned, as well as the economic analysis of those cultures in Subtasks 12.2.

Subtask 10.2: Generation of most promising production strains of *Synechococcus* based on baseline strain (Task 2), highest yield mutations (Task 7), and metabolic predictions (Task 6)

The UC Davis lab conducted a production experiment using the succinate producing cyanobacteria strain LAN1 (baseline strain) with the glucose consumption pathway (Strain 1), which was further modified to include a knockout of the gene encoding the enzyme that converts succinate to fumarate (*sdhB*). The knockout of *sdhB* has recently been shown to help accumulate succinate in *Synechococcus elongatus* PCC 11801 (Sengupta, S., Jaiswal, D., Sengupta, A. et al. Metabolic engineering of a fast-growing cyanobacterium Synechococcus elongatus PCC 11801 for photoautotrophic production of succinic acid. Biotechnol Biofuels 13, 89 (2020). https://doi.org/10.1186/s13068-020-01727-7), as well as through our own research efforts. The succinate production pathway is in neutral site 1 (Ptrc:*gabD-kgd;* SpecR), the glucose consumption pathway is in neutral site 2 (Ptrc:*galP-zwf-gnd,* KanR), and the gene *sdhB* has been replaced by a chloramphenicol resistance cassette (Strain 2). This experiment was carried out at pH=10.5 in BG-11 with no bicarbonate supplementation. High pH was previously found to be beneficial for succinate production. However, glucose import does not occur when the pH is maintained at 10.5. It is necessary to allow succinate production to lower the pH from 10 to 7 to activate glucose consumption. The pH is manually adjusted back to 10.5 each day to optimize succinate production. Approximately 700 mg/L of succinate was produced after 3 days. However, the *sdhB* KO with the succinate pathway (*gabD-kgd*) inhibited cell growth. The culture had early-stage chlorosis throughout preculture and the experiment

The photomixotrophic strain which harbors the glucose consumption pathway (*galP-gnd-zwf*) and the two gene succinate production pathway (*gabD-kgd*) achieves a productivity of 175 mg/L/day compared to 18 mg/L from the photoautotrophic strain.[1] Fumarate is a major side product in these strains. We have been able to remove fumarate production by deleting *sdhB*. However, the deletion of *sdhB* conferred a growth detriment under heterotrophic growth conditions not otherwise observed for autotrophic growth conditions. We hypothesized overexpression of *ppc* and *gltA* could rescue growth since *sdhB* was deleted without growth detriment in the *S. elongatus* PCC 11801 succinate strains with *ppc* and *gltA* overexpressed.[2] Our proposed modifications involve overexpressing various versions *ppc* alone and in conjunction with *gltA* in the *sdhB* locus. The genes *ppc* and *gltA* have been used with success to enhance succinate titer in two recent studies.[1,2] In the study from Lan & Wei, the *Corynebacterium glutamicum* versions of these genes were used to improve succinate titers.[1] *C. gltuamicum* is regarded as one of the major succinate producing organisms.[3] The other recent

succinate production study in *S. elongatus* PCC 11801 used the versions of *ppc* and *gltA* from 7942.[2] Due to the disparity in which versions of *ppc* and *gltA* have been used, we aim to create multiple strains housing different variants of these genes to gain a further understanding into which of these genes will confer the best productivity to 7942. The strains were constructed and tested for relative succinate production.

Table 1 Strain list

Strain	Genotype
Strain 1	NSI: *Ptrc:gabD-kgd*, *Spec*^{R} NSII: *Ptrc:galP-zwf-gnd*, *Kan*^{R}
Strain 2	As Strain 1 but *sdhB*::*Gent*^{R}
Strain 3	As Strain 1, but *sdhB*:: $P_L$lacO1: *ppc* (*C. glutamicum*), *Gent*^{R}
Strain 4	As Strain 1, but *sdhB*:: $P_L$lacO1: *ppc* (*E. coli* MG1655), *Gent*^{R}
Strain 5	As Strain 1, but *sdhB*:: $P_L$lacO1: *ppc* (*S. elongatus* PCC7942), *Gent*^{R}
Strain 6	As Strain 1, but *sdhB*:: $P_L$lacO1: *ppc* (*C. glutamicum*)–*gltA* (*C. glutamicum*), *Gent*^{R}
Strain 7	As Strain 1, but *sdhB*:: $P_L$lacO1: *ppc* (*C. glutamicum*)–*gltA* (*E. coli* MG1655), *Gent*^{R}
Strain 8	As Strain 1, but *sdhB*:: $P_L$lacO1: *ppc* (*C. glutamicum*)–*gltA* (*S. elongatus* PCC7942), *Gent*^{R}

The strains investigated demonstrate a couple of important learned lessons, some of which have been discussed above as part of Task 6.0. A novel finding was that while inclusion of the knockout of sdhB and fumarate production in the baseline mixotrophic strain did initially increase succinate production, it hindered growth of the strain under mixotrophic conditions that included bicarbonate at pH 7, and thus hindered production. When bicarbonate was not added and the culture was grown at pH 10.5, the ΔsdhB strain did produce significantly more succinate than Strain 1, but the stopped producing after 3 days and was again growth limited. This is in contrast to the photosynthetic growths, where this mutation did not hinder growth and enhanced succinate production 2-fold. However, addition of a *ppc* gene to the ΔsdhB strain recovered productivity and growth with or without bicarbonate. Testing *ppc* genes from different organisms showed that they all supported growth equally, but the one derived from *S. elongatus* (Strain 5) proved the most productive in terms of succinate yields. Interestingly, in contrast to strains grown previously without bicarbonate, the pH of the culture needed to be lowered daily to pH 7.5 to allow improved glucose uptake, rather than raised daily to 10.5 to ensure productivity. Adding a *gltA* gene to strain 5 hindered productivity when the gene was derived from *E. coli* or *S. elongatus*, but did not hinder productivity when derived from *C. glutamicum*. Growth was hindered by the *gltA* gene from *E. coli*, was the same with the gene from *S. elongatus*, and was improved by the *gltA* gene from *C. glutamicum*. pH adjustments were similar as for strain 5. Further minor improvements in growth and productivity were observed with deletion of the Cp12 transcription factor, which regulates the calvin benson cycle based on light conditions and metabolism, and simultaneous expression of *prk*, a gene involved in the calvin benson cycle. Finally, we performed high density culture growths in 2x BG-11 media and have observed increased cell biomass densities (> 2-fold) and much higher levels of succinate production (> 4-fold) than when performed under lower density conditions.

1. Lan, E. I. & Wei, C. T. Metabolic engineering of cyanobacteria for the photosynthetic production of succinate. *Metab. Eng.* **38**, 483–493 (2016).

2. Sengupta, S. *et al.* Metabolic engineering of a fast-growing cyanobacterium *Synechococcus elongatus* PCC 11801 for photoautotrophic production of succinic acid. *Biotechnol. Biofuels* **13**, 1–18 (2020).
3. McKinlay, J.B., Vieille, C. & Zeikus, J.G. Prospects for a bio-based succinate industry. *Appl Microbiol Biotechnol* **76,** 727–740 (2007).

Uptake of both glucose & xylose: Because the NREL-supplied cellulosic hydrolysate is predominantly composed of glucose and xylose, we tested the capacity of PCC 7942 to be engineered to take up both sugars, grow, and produce a chemical product. The generated strain has three genes added to uptake and convert glucose to Ru5P, three genes to take up xylose and convert it to Xu5P, three genes to produce 2,3-BDO from pyruvate, and overexpress *prk* while simultaneously knocking out *cp12*. This strain was capable of growing to very high ODs, higher than those observed for our succinate production strains that could only utilize glucose or our 1,4-BDO strains that could only utilize xylose. Further, this strain produced high titers of 2,3-BDO – up to 4 g/L, while consuming both sugars. Another observation of note is that *S. elongatus* does not appear to be inhibited by increasing concentrations of the cellulosic hydrolysates (as shown previously), while *E. coli* is inhibited. This could be important for systems that can be contaminated due to the presence of sugars that this hydrolysate can inhibit contaminating or competing organisms.

Milestone 10.ML.1: Most promising production strains in green algae and cyanobacteria. 1+ strains of green algae and cyanobacteria, generated by integrating results and tools from prior tasks, ready to undergo scaled growth. Product yield of strain > 0.1 g/L or >1.0 g/L depending on product

The data above describes an optimized strain for increased succinate production and its productivity yields.

Task 11.0: Scaled Production Platform Yield Testing

Subtask 11.1: Heterotrophic scaled growth and yield evaluation of most promising production strains of *Chlamydomonas*

Due to successes with production in *Synechococcus*, and the lack of a *C. reinhardtii* strain that produced succinate at yields greater than 100 mg/L, as described in Tasks 2.0 & 6.0, we down-selected to focus on succinate production from *Synechococcus*. This also means that Subtasks 10.1, 11.3, 11.5 & 11.6, where co-cultures of green algae and cyanobacteria were to be performed, did not proceed as originally planned, as well as the economic analysis of those cultures in Subtasks 12.2.

Subtask 11.2: Heterotrophic scaled growth and yield evaluation of most promising production strains of *Synechococcus*

Scaled growth in a fermentor: The UCSD groups together attempted a growth of the succinate producing strain in a 6-L fermentor with glucose. In order to support enhanced growth, we attempted a preliminary experiment to grow the cells in 5x concentrated BG11 (demonstrated in the literature to allow growth of PCC 7942) in the fermentor. Unfortunately, either due to a combination of this new media or the increased pH using ammonium hydroxide instead of sodium hydroxide, the fermentation culture died within a day. Follow up experiments in flasks demonstrated that the production strain is tolerant of 5xBG11, though WT is not, but is less tolerant when the pH is adjusted above 10.5. Media optimization proved to be necessary to determine the ideal high nutrient media for future fermentation experiments.

Multiple subsequent attempts to optimize the fermentation culture conditions; however, the culture has crashed early on in the experiment. In one experiment, the issue was in trying to maintain a high pH with an ammonium hydroxide solution, as opposed to a sodium hydroxide solution. In a second experiment, we attempted to adjust the concentration of media components that we believed were either necessary for high growth or had accumulated to the point of making the culture too salty for the freshwater strain (another reason why moving to a marine strain is a promising path forward). To better get a handle on these factors and variables, we performed experiments in baffled flasks that simulate the shear forces and gas exchange of the fermentation tank while slowly increasing the concentration of components from a 1x media recipe to a 5x BG-11 media recipe, in the presence of 5 g/L glucose. In these baffle flasks, cultures survived throughout the experiment even to the end of the experiment when the media was close to a full 5x BG-11 solution. One caveat was that the flask was clearly contaminated with fungus at this point due to the presence of glucose and the constant sampling required for the experiment.

Continued succinate production experiments were performed in the baffled flasks and in a homemade bioreactor system in order to optimize media components for growth and production. After adaptation periods in 2x BG-11 and 3x BG-11, the culture was adapted to a 5x BG-11 culture media that had previously been published as supporting production by the UC Davis group. In the bioreactor system, the 5x BG-11 media supported photosynthetic growth at levels approximating 4 g/L in dry biomass, which are values only previously observed in the presence of glucose and IPTG. This concentrated media was subsequently used for 20 L carboy growths and 2 L bioreactor growths, both with and without glucose and IPTG addition. Interestingly, survivability in the presence of glucose and IPTG was somewhat hindered at these higher concentrations of BG-11, though it was unclear if this was due to the rapid pH drop associated with the metabolism of glucose or other factors. After a couple of trial runs, it

became clear that monitoring and adjusting pH continuously in those growths with glucose was crucial for stability of the culture due to the sudden drop in pH associated with either glucose consumption or succinate production.

Fermentation in the small bioreactors and fermenters proved quite successful, with the OD of the culture at 750 nm reaching 16, approximately 16 fold higher than that generally reached via photosynthesis in flask cultures. This OD correlated with a dry weight biomass of 30 g/L. Key to this high growth was proper inputs of pH adjustments and high flow of aeration into the culture. These fermentation systems were analyzed for succinate production, which ranged from 5 to 13 g/L, depending upon the date of sampling and the culture sampled. A caveat to the analysis is that the peak observed, which does correlate with the expected accumulation of succinate, has a retention time of 0.5 min slower than that of succinate standards, though this shift in retention time has been previously observed with past cultures, possibly due to pH of the sample.

In both bioreactors, glucose consumption and biomass accumulation were difficult to accurately quantitate due to volume losses due to evaporation. Nonetheless, glucose consumption, biomass accumulation, and succinate production were observed. However, succinate production did not appear to occur as expected and the hypothesis was postulated and tested that not enough IPTG was added to induce the succinate system, given the much higher density of cells present in the culture than in any other previous culture performed in any lab working with PCC 7942. Either IPTG was degrading in these cultures, or the number of cells is so high that the effective concentration of IPTG per cell is too low to substantially increase production. Addition of IPTG in both bioreactors increased succinate concentrations, indicating that in either case, IPTG concentrations must be increased for optimal succinate yields. This lesson was then applied to the baffled flask experiments. Cultures were again grown as previously described in 5xBG-11. Cultures were grown for 600 hours, at which time a succinate yield of 12 g/L was observed, as determined by HPLC and previously generated calibration curves. Based on the observations and experiments performed on the bioreactor cultures, IPTG was added and succinate production was assessed 24 hours later, along with final glucose consumption and biomass production. It was determined that 32.2 g/L of glucose was consumed in total, 18 g/L of biomass was generated, and a final succinate yield of 32.5 g/L was achieved. Thus, in 24 hours after adding IPTG, an additional 20 g/L of succinate was generated, indicating catalytic conversion of glucose to succinate. Following this experiment, a new culture was inoculated from the 32 g/L culture. In the first 150 hours, growth rates were nearly tripled from the previous culture.

Growths experiments were continued for both reproducibility and for scaled volume. One of the factors contributing the most towards the observed high growth and productivity was the large surface area to volume ratio of the baffled flask growths that were performed. That high SA:V ratio allowed for increased air exchange and mixing that allowed for rapid growth and production. However, this growth system means that scaling to larger volumes is challenging due to increased demands for space. Experiments were performed to increase the volume in the same vessels or in smaller vessels. Unfortunately, at reduced SA:V values, productivity measurements were not greater than 2 g/L of succinate produced in these cultures. It is clear that aeration is a major component of the enhanced productivity and we are working on novel strategies to increase that while scaling.

Ongoing cultures were also switched from pure glucose to the cellulosic hydrolysates (CH) provided by NREL. Cultures adapted well to an equivalent amount of CH as glucose used; previously we grew up to 40 g/L glucose and thus have tested 40 g/L CH or 80 g/L CH due to

CH being approximately 50% glucose. To date in these experiments, growth on CH matches that on pure glucose.

Mixotrophic growth at UC Davis: The overexpression of *ppc* variants installed into the *sdhB* locus on Strain 1 (NSI: *P*trc:*gabD-kgd, Spec*^{R} NSII: *P*trc:*galP-zwf-gnd, Kan*^{R}) were tested for productivity. Initially, we performed these experiments at high pH (10.5) without bicarbonate, since we had seen bicarbonate negatively affecting growth and production at high pH. At the high pH, Strain 3 (as Strain 1, but *sdhB*::$P_L$lacO1:*ppc* (*Corynebacterium glutamicum*), *Gent*^{R}) died by day 3 due to the pH dropping from 10 to 5. Approximately 150 mg/L succinate was produced at this point in time. Thus, we performed the experiments at an initial pH of 7 and added 20 mM sodium bicarbonate to the media. Of these *ppc* variants, Strain 5 (as Strain 1, but *sdhB*::$P_L$lacO1:*ppc* (*Synechococcus elongatus* PCC 7942)*, Gent*^{R}) produced 3.6 g/L succinate by day 5 and possessed a robust growth phenotype. Since these strains with *ppc* did not experience any bicarbonate sensitivity we will be proceeding forward with testing all strains under the neutral pH with 20mM bicarbonate condition. Strain 4 (as Strain 1, but *sdhB*::$P_L$lacO1:*ppc* (*E. coli* MG1655)*, Gent*^{R}) produced no succinate over the same experimental time course as the experiment with Strain 5. Once it is determined which *ppc* variant is best, several *gltA* variants will be expressed in conjunction with the best *ppc* to determine the optimal combination of these genes in the context of our succinate production system. We have already constructed these strains.

Milestone 11.ML.1: Preliminary evaluation of scaled production yields under heterotrophic growth. 1+ replicate growths and yield analysis of production strains grown for 180+ hrs at 2+ L scales.

Baffled flask, as well as bioreactor, growth and production analysis demonstrated that a yield of >30 g/L of succinate could be achieved.

Subtask 11.3: Photosynthetic scaled growth and yield evaluation of most promising production strains of *Chlamydomonas*

Due to successes with production in *Synechococcus*, and the lack of a *C. reinhardtii* strain that produced succinate at yields greater than 100 mg/L, as described in Tasks 2.0 & 6.0, we down-selected to focus on succinate production from *Synechococcus*. This also means that Subtasks 10.1, 11.1, 11.5 & 11.6, where co-cultures of green algae and cyanobacteria were to be performed, did not proceed as originally planned, as well as the economic analysis of those cultures in Subtasks 12.2.

Subtask 11.4: Photosynthetic scaled growth and yield evaluation of most promising production strains of *Synechococcus*

The UCSD cyanobacteria group initiated preliminary studies for scaling the succinate production strain previously used to demonstrate > 4 g/L yields. This strain was first used to attempt growth in 200-L raceway ponds. Twice, 4 L of cultures were grown in the lab and used to inoculate 20 L of BG11 media in the raceway ponds. In the first experiment, the culture survived well for

approximately 5 days, but upon raising the culture volume to 40 L, the culture died. We suspect that this culture death was due to photobleaching as the culture was not extremely dense and was not maintained under sufficient shading during this early growth period. In the second experiment, the inoculated 20-L culture died within a few days of inoculation due to the presence of a *Tetrahymena* predator that the group has previously isolated from the field station. These experiments highlight the sensitivity of PCC 7942 to growth conditions at the field station and indicate numerous difficulties that may arise in attempting growth and production of succinate from this strain. Media optimization may assist in this endeavor, as the UCSD cyanobacterial group has demonstrated that high pH or media with higher amounts of sodium carbonate and sodium bicarbonate can kill the *Tetrahymena*. However, some of these media optimizations, for example the enhanced carbonate alteration, are stressful to the PCC 7942 strain, impairs growth, and ultimately leads to culture death. To overcome this challenge, as part of Task 10.0, we have been attempting to adapt the strain to this media. As a second alternative, we have investigated growing a novel high biomass marine strain similar to PCC 7942, *Synechococcus sp.* PCC 11901, at the field station in its preferred MAD2 media (high salt content), which we have also demonstrated kills the *Tetrahymena*. We have successfully inoculated and grown this culture to high density (> 4.5 gDW/L) up to a scale of approximately 60 L in the raceway pond for over a month. Although the media was not initially adjusted for pH, the cyanobacteria have naturally adjusted the pH of the media to be greater than 10. Examination of the pond indicates that it is contaminated with a heterotrophic bacterium that is now accumulating as an odorous biofilm in the pond. So far, this bacterium does not appear to be detrimental to the PCC 11901 culture. We have also noted the presence of amoebae that are capable of grazing on the PCC 11901 strain when the culture is grown on a plate. The amoebae do not appear to be sufficiently deleterious to the culture to prevent growth or cause a crash of the pond. These preliminary experiments highlight the challenges ahead for scaling, but also suggest viable solutions, such as media changes, strain adaptation, and changing the chassis organism.

Milestone 11.ML.2: Preliminary evaluation of scaled production yields under photosynthetic growth. 1+ replicate growths and yield analyses of production strains grown for 180+ hrs at 20+ L scales.

20-L carboy growth and production were tested in Tasks 2.0 and in Tasks 1.0, demonstrating that cultures could grow under these conditions and produce succinate. However, growth of this strain in ponds proved to be difficult if not impossible due to the presence of predators. Alternative strategies have been designed to overcome these issues, which may be investigated in future projects.

Subtask 11.5: Heterotrophic scaled growth and yield evaluation of polycultures composed of mixtures of green algal and cyanobacterial production strains.

Due to successes with production in *Synechococcus*, and the lack of a *C. reinhardtii* strain that produced succinate at yields greater than 100 mg/L, as described in Tasks 2.0 & 6.0, we down-selected to focus on succinate production from *Synechococcus*. This also means that Subtasks 10.3, 11.1, 11.3 & 11.6, where co-cultures of green algae and cyanobacteria were to be

performed, did not proceed as originally planned, as well as the economic analysis of those cultures in Subtasks 12.2.

Subtask 11.6: Photosynthetic scaled growth and yield evaluation of polycultures composed of mixtures of green algal and cyanobacterial production strains.

Due to successes with production in *Synechococcus*, and the lack of a *C. reinhardtii* strain that produced succinate at yields greater than 100 mg/L, as described in Tasks 2.0 & 6.0, we down-selected to focus on succinate production from *Synechococcus*. This also means that Subtasks 10.1, 11.1, 11.3, & 11.5, where co-cultures of green algae and cyanobacteria were to be performed, did not proceed as originally planned, as well as the economic analysis of those cultures in Subtasks 12.2.

Task 12.0: TEA/LCA Modeling of Algal Bioproduction

Subtask 12.1: TEA/LCA evaluation comparing heterotrophic and photosynthetic production systems as compared to current bioproduction systems.

To initiate the work on this task, and to help inform the data generation requirements of Task 11 (scaled growth), a preliminary block flow diagram for the modeling of production has been developed. This model stems from one developed under the PEAK grant that had focused on intracellular high-value protein production in green algae. Alterations to the previous PEAK model derive from the need to realistically compare heterotrophic and phototrophic production, which at scale would be performed in different growth vessels (fermentation tank vs. pond or bioreactor, respectively), at different sizes (~10,000 L vs. possibly multiacres), and with different inputs. These differences had practical implications for the experiments performed here: fermentation was performed on the 2 – 6 L range in a sealed fermentor with concentrated media and NREL cellulosic hydrolysates, while phototrophic production was performed in 20 L carboys or in 200 L raceway ponds. If a platform production strain were to be ultimately grown in raceway ponds, for which a model has been established in collaboration with Global Algae Innovations for extrapolating productivity to agricultural scale values, the impact of contaminants and predators (see Task 11) must be realistically considered on both the growth of the culture and the production yield, as many contaminants can ingest succinate. Task 11 describes mitigation strategies of contamination for improved crop protection and protection of polyurethane precursors. For modeling purposes, however, the models and systems were focused on those processes for which the best data was generated during the course of this project – raceway ponds for photosynthetic production, 2 L flask growth for mixotrophic production, and 2 – 6 L fermentor growth for fermentative production. Furthermore, the differing productivities observed for fermentation and photosynthetic production has implications for recovery of the succinic acid product. The preferred method for succinic acid purification is electrodialysis, as this is the method previously used at scale by BioAmber for purification from fermentation broth and is capable of generating highly pure product. However, if the yield is too low, e.g. < 10 g/L, then a concentration or evaporation step is required prior to electrodialysis in the purification scheme. To allow this experimentally at lab scale, a 20-L rotary evaporator (rotovap) was acquired. A recent review paper (https://doi.org/10.1016/j.jclepro.2020.123954) describes a number of alternatives to this scheme, including membrane filtration systems that could in theory (not tested yet in practice) be put in line with the fermentation or growth culture and enable recycling of cells, media, and water back to the culture. Although these alternatives may be pursued in future work, the purification scheme used for TEA analysis was modeled on electrodialysis with a possible concentration step.

The preliminary LCA and TEA models were examined for production and processing of bio-succinic acid, with particular attention to production under fermentation conditions. Based on the literature, parameterized models were constructed for the production and processing of bio-succinic acid, with particular attention to production under fermentation conditions. Based on the literature, parameterized models were constructed for open pond systems, industrial fermentation, and industrial bioreactor systems. For the photosynthetic production of succinate in open pond systems, the model was built off of the model developed during the UCSD PEAK project, where parameter values were based on those supplied by a large grower, Global Algae Innovations, and supplemented with reference LCIs from the GaBi software. Characterization factors, sourced through TRACI 2.1, were added to assess the impact of model results. Results show a global warming potential of ~11.3 kg CO_2e per kg of succinic acid production. Other characterization factors included are total primary energy, total water consumption, human toxicity, ecotoxicity, fossil fuel resource depletion, particulate matter, ozone depletion, smog

formation, acidification, and eutrophication. Fermenter and bioreactor models were developed in SuperPro Designer, which are based on reference models for BioAmber's commercial scale succinic acid fermentation production system. For many components of the model, various literature sources were used to provide values for the parameterization of the model. Experimental results by UCSD were used to complete the model mass balance. Data for the techno-economic assessment was obtained from Global Algae. LCA and TEA models were revised and run with new data values for each situation being modeled: phototrophic growth in open ponds, mixotrophic growth in bioreactors, and fermentative growth in a fermenter. Data was compared against prior publications on production from yeast or heterotrophic bacteria, as well as petroleum-derived materials.

Running the model based on these data sets resulted in the following. The capital cost comes out to $303,316,913.18. When multiplied by the capital charge factor reduces to $40,250,154.38. The operating cost total is $31,035,299.39. The techno-economic assessment results show a total cost of $71,285,454, or $0.43 per kg of succinic acid harvested. The adjusted cost after the co-product credit for biomass converted to livestock feed is $61,586,882.34, with the credit being $0.31/kg .

It should be noted, however, that these comparisons should be taken with care because the back-end purification systems were not included in the model and the scaling of the heterotrophic systems are not yet optimized. The results of running the models with these inputs shows that the open pond system is closer to that of current bio or petroleum production systems in terms of global warming emissions and costs per kilogram produced by at least an order of magnitude. The results suggest that while open ponds currently are the most comparable systems for production as compared to petroleum systems, optimizations associated with scaling may allow for mixotrophic or heterotrophic systems to compete.

Milestone 12.ML.1: Determination of cost impacts of heterotrophic versus photosynthetic production systems. Completed TEA/LCA analysis using data from Tasks 11.1 - 11.4.

Completed as described above for Subtask 12.1.

Subtask 12.2: TEA/LCA evaluation comparing a mixed algae and cyanobacterial production system versus monoculture systems utilizing the entire carbon source from cellulosic hydrolysates.

As described above under Tasks 10 and 11, due to the lack of a sufficiently high-yielding green algae strain and the subsequent down-selection to optimizing and scaling just the cyanobacteria, Subtasks 11.5 and 11.6 were not performed and thus the data needed to perform Subtask 12.2 was not generated. The down-selection removed the capacity to perform this subtask.

Publications

1. Badary, Amr, Nora Hidasi, Simone Ferrari, and Stephen P. Mayfield. "Isolation and characterization of microalgae strains able to grow on complex biomass hydrolysate for industrial application." Algal Research 78 (March 2024): 103381. DOI: 10.1016/j.algal.2023.103381.

2. Diaz, Crisandra J., Kai J. Douglas, Kalisa Kang, Ashlynn L. Kolarik, Rodeon Malinovski, Yasin Torres-Tiji, João V. Molino, Amr Badary, and Stephen P. Mayfield. “Developing Algae as a Sustainable Food Source.” Frontiers in Nutrition 9 (January 19, 2023): 1029841. DOI: 10.3389/fnut.2022.1029841.

3. Gonzales, Jake N., Tanner R. Treece, Stephen P. Mayfield, Ryan Simkovsky, & Shota Atsumi. "Utilization of lignocellulosic hydrolysates for photomixotrophic chemical production in Synechococcus elongatus PCC 7942." Communications Biology 6, no. 1 (October 2023): 1022. DOI: 10.1038/s42003-023-05394-w.

4. Sproles, Ashley E., Anthony Berndt, Francis J. Fields, and Stephen P. Mayfield. “Improved High-Throughput Screening Technique to Rapidly Isolate Chlamydomonas Transformants Expressing Recombinant Proteins.” Applied Microbiology and Biotechnology 106, no. 4 (February 2022): 1677–89. DOI: 10.1007/s00253-022-11790-9.

5. Sproles, Ashley E., Francis J. Fields, Tressa N. Smalley, Chau H. Le, Amr Badary, and Stephen P. Mayfield. “Recent Advancements in the Genetic Engineering of Microalgae.” Algal Research 53 (March 2021): 102158. DOI: 10.1016/j.algal.2020.102158.

6. Tanner R. Treece, Marissa Tessman, Robert S. Pomeroy, Stephen P. Mayfield, Ryan Simkovsky, & Shota Atsumi. (2023). "Fluctuating pH for efficient photomixotrophic succinate production." Metabolic Engineering. DOI: 10.1016/j.ymben.2023.07.008.

7. Torres-Tiji, Yasin, Francis J. Fields, Yanping Yang, Vanessa Heredia, Svein Jarle Horn, Saumya R. Keremane, Moonsoo M. Jin, and Stephen P. Mayfield. “Optimized Production of a Bioactive Human Recombinant Protein from the Microalgae Chlamydomonas Reinhardtii Grown at High Density in a Fed-Batch Bioreactor.” Algal Research 66 (July 2022): 102786. DOI: 10.1016/j.algal.2022.102786.