

UC San Diego

UC San Diego Electronic Theses and Dissertations

Title

Identification of transcription factor interaction subnetwork and Modeling for metabolic-protein secretion in *Chlamydomonas reinhardtii*

Permalink

<https://escholarship.org/uc/item/3k55k8hf>

Author

Ren, Bijie

Publication Date

2020

Peer reviewed|Thesis/dissertation

UNIVERSITY OF CALIFORNIA SAN DIEGO

Identification of transcription factor interaction subnetwork and Modeling for metabolic-protein secretion in

Chlamydomonas reinhardtii

A dissertation submitted in partial satisfaction of the requirements for the degree

Doctor of Philosophy

in

Biology with a Specialization in Quantitative Biology

by

Bijie Ren

Committee in charge:

Professor Stephen Mayfield, Chair
Professor Susan Golden
Professor Terry Hwa
Professor Nathan Lewis
Professor Bing Ren

2021

Copyright

Bijie Ren, 2021

All rights reserved

The dissertation of Bijie Ren is approved, and it is acceptable in quality and form for publication on microfilm and electronically.

University of California San Diego

2021

DEDICATION

*This dissertation is dedicated to my cool lab mates,
excellent committee members,
outstanding collaborators and amazing classmates.*

TABLE OF CONTENTS

Dissertation Approval Page.....	iii
Dedication	iv
Table of Contents	v
List of Abbreviations	vi
List of Figures	vii
List of Tables	ix
Acknowledgements	x
Vita	xii
Abstract of the Dissertation	xiv
Introduction	1
Chapter 1: Oral delivery of a functional algal expressed TGF- β mimic halts colitis in a murine DSS model.....	4
Chapter 2: Discovery of two master regulators for lipid accumulation <i>Chlamydomonas reinhardtii</i> through Yeast Two-Hybrid sequencing.....	39
Chapter 3: Genome scale reconstruction of Metabolic-Protein secretion modeling for <i>Chlamydomonas reinhardtii</i> ...	67
Conclusion	94

LIST OF ABBREVIATIONS

IBD – Inflammatory bowel diseases

TGM – TGFb mimic protein

TAP – tris acetate phosphate

IBMX – 3-isobutyl-1-methylxanthine

Y2H – Yeast 2 hybridization

CreY2Hseq – Cre mediated Yeast two hybridization sequencing

TrTGM – Truncated TGM

RBG – IL22

COBRA - COntstraint-Based Reconstruction and Analysis

CHO - Chinese Hamster Ovary

LIST OF FIGURES

Figure 1.1 Key immune cells induced in Inflammatory bowel disease.....	8
Figure 1.2. Expression of <i>Heligmosomoides polygyrus</i> TGM1 in <i>Chlamydomonas reinhardtii</i>	17
Figure 1.3. FLAG-purified algal TGM1 and its bioactivity on TGF- β -responsive cell assays	19
Figure 1.4. Effect of algal TGM1 on mucosal immune cell populations.....	21
Figure 1.5. Amelioration of in vivo DSS mouse colitis with orally administered algal TGM1.....	22
Supplementary Figure 1.1 Nucleotide sequence of TGM1 as codon optimized for <i>C. reinhardtii</i> nuclear transcription	29
Supplementary Figure 1.2 Expression and Purification of Algal TGM1	30
Supplementary Figure 1.3 In vitro assays for TGF- β activity.....	30
Supplementary Figure 1.4 Physiological responses to orally administered algal TGM1.....	31
Supplementary Figure 1.5 TGM1 administration during DSS-induced colitis.....	32
Figure 2.1. Construction of <i>Chlamydomoans reinhardtii</i> Transcription factor AD library and BD library.....	42
Figure 2.2. Transcription factor interaction subnetwork identified by Y2Hseq assay.....	44
Figure 2.3 Lipid analysis of 4533 wild type, TF24 mutant, TF20 mutant and 5373 sta6- mutant strain.	46
Figure 2.4 Time course RNAseq data of transcription factor mutants and 4533 wild type under Nitrogen starvation.....	49
Supplementary Figure 2.1. Control assay for TF-AD Y8800 strain mating with CRY8930 strain transformed with empty BD vector.....	58
Supplementary Figure 2.2. Control assay for TF-BD strain mating with strain transformed with empty AD vector.....	58
Supplementary Figure 2.3 Total biomass for lipid mass spec analysis.....	59
Supplementary Figure 2.4 Unique pathways only upregulated in tf20m 48hours after nitrogen starvation.....	59
Supplementary Figure 2.5 Time course PAGE analysis for GO enrichment in wt strain.....	59
Supplementary Figure 2.6 GO enrichment (according to Biological process) in tf20m strain 4 hours versus 0 hour after Nitrogen starvation.	60
Supplementary Figure 2.7 Time course PAGE analysis for differential GO enrichment (according to Biological process) in wt and tf20m strain 48 hours versus 0 hour after Nitrogen starvation.	60
Supplementary Figure 2.8 Time course PAGE analysis for differential GO enrichment (according to Cellular component) in wt and tf20m strain 48 hours versus 0 hour after Nitrogen starvation.	61

Supplementary Figure 2.9 Time course PAGE analysis for differential GO enrichment (according to Molecular function) in wt and tf20m strain 48 hours versus 0 hour after Nitrogen starvation.	61
Supplementary Figure 2.10 Unique pattern of changed pathway enriched in tf24m from PAGE analysis of time course RNAseq data for all three strains.	62
Supplementary Figure 2.11 Uniquely downregulated genes in fatty acid metabolism in tf24m and tf20m strains.	62
Supplementary Figure 2.12 Unique pathway downregulated in tf20m strain 48 hours after Nitrogen starvation.	63
Supplementary Figure 2.13 Positive colonies obtained from different subgroups of TF-AD library mating with TF-BD library.	63
Figure 3.1 Distribution of subsystems and reactions in model iCre1355.	70
Figure 3.2 Genome scale reconstruction of protein secretion pathway.	71
Figure 3.3. Anti-correlation between growth rate and protein secretion rate.	72
Figure 3.4. Protein secretion rate depends on nitrogen concentration and nitrogen source.	74
Figure 3.5 Growth rate of different mutant strains versus wild type strain.	77
Figure 3.6 Time course Western blot of transgenic strain induced by IBMX.	79
Supplementary Figure 3.1 Energy cost vs Riboseq transcript abundance from human kidney and CHO cell.	86
Supplementary Figure 3.2 Suppression of recombinant protein secretion when algae cells grew in CO ₂ box with a higher growth rate.	86
Supplementary Figure 3.3 Time course measure of secreted proteins, growth rate, Chlorophyll, Nile red staining for lipid content and RNA.	87
Supplementary Figure 3.4 Recombinant protein expression slowed growth rate and lowered cell density during stationary phase.	87
Supplementary Figure 3.5 Time course measure of growth rate, Chlorophyll, Nile red staining for lipid content and RNA for TGM strain and wt strain in TAP medium.	88
Supplementary Figure 3.6 Nile red staining for lipid content of cells from different nitrogen groups.	88
Supplementary Figure 3.7 Time course western blot of secreted protein in the culture.	89
Supplementary Figure 3.8 Comparison of protein secretion yield with saturated Nitrogen vs limited Nitrogen predicted by genome-scale model.	89
Supplementary Figure 3.9 Time course measurement for Chlorophyll, Nile red and bulk RNA for cells from different nitrogen groups.	90
Supplementary Figure 3.10 Differential expression change between Urea medium and Ammonium medium.	90

LIST OF TABLES

Supplementary Table 1.1 Protein yields of secreted algal material.....26

Supplementary Table 1.2 Glycosylation site analysis of algal TGM127

ACKNOWLEDGEMENTS

I would like to acknowledge Professor Stephen Mayfield for his genius mentorship, insightful view and great support, especially for offering enormous opportunity in this exciting field and supporting me to do what interests me most.

I would like to acknowledge my undergrad students Song Cao, Yiqun Jiang, Nan Hu, Ashley Dung, Stephanie Domes and all members of the Mayfield lab for their nice help throughout my Ph.D program.

I would like to acknowledge Professor Susan Golden and Professor Terry Hwa for all excellent suggestions and mentorship starting from Bootcamping. Their outstanding guidance in the qBio courses and committee meeting help me to learn the great personalities for researching and mentoring.

I would like to acknowledge Professor Nathan Lewis and Jahir Gutierrez for the unquantifiable measure of support and guidance about cell modeling project, who also set up a role model for me to learn. I would like to acknowledge Professor Jagroop Pandhal and Dr. Caroline Evans for their great work in the O glycosylation of algal whole proteome. It is their support that helped me in an immeasurable way.

I would like to acknowledge my collaborators Professor Rick Maizels, Dr. Danielle Smyth, Dr. Madeleine White for the fantastic work in the algal therapeutic protein project. Their hard work and continuous help make it possible to develop a new therapeutic method to inflammatory disease and enable our project publishable.

Finally, I would like to thank Professor Bing Ren and other committee members I mentioned above for their great commitment and time, unique insights and excellent suggestions. Without them I cannot get through my Ph.D life.

Chapter 1, in part is currently being prepared for submission for publication of the material. Bijie Ren, Danielle Smyth, Madeleine White, Francis Fields, Rick Maizels, and Stephen Mayfield. Oral delivery of a functional

algal expressed TGF- β mimic halts colitis in a murine DSS model. The dissertation author was the primary investigator and author of this paper.

Chapter 2, in part is currently being prepared for submission for publication of the material. Bijie Ren, Song Cao, Stephen Mayfield. Discovery of two master regulators for lipid accumulation *Chlamydomonas reinhardtii* through Yeast Two-Hybrid sequencing. The dissertation author was the primary investigator and author of this paper.

Chapter 3, in part is currently being prepared for submission for publication of the material. Bijie Ren, Terence Hwa, Jahir M. Gutierrez, Frank Fields, Nathan Lewis, Stephen Mayfield. Genome-scale reconstruction of Metabolic-Protein secretion modeling for *Chlamydomonas reinhardtii*. The dissertation author was the primary investigator and author of this paper.

VITA

2008-2012 Bachelor of Science, China Agricultural University, Beijing, China
2014-2015 Advanced study program, Massachusetts Institute of Technology, Boston, MA
2015-2021 Doctor of Philosophy, Division of Biological Sciences, University of California San Diego,
La Jolla, CA

PUBLICATIONS

Bijie Ren, Madeleine White, Danielle J. Smyth, Francis Fields, Rick M. Maizels, and Stephen Mayfield. Oral delivery of a functional algal-expressed TGF- β mimic halts colitis in a murine DSS model. *Journal of Biotechnology*(submitted)
Anthony J Berndt; Tressa N Smalley; **Bijie Ren**; Amr Badary; Ashley Sproles; Francis Fields; Yasin Torres-Tiji; Vanessa Heredia; Stephen P Mayfield. Recombinant production of a functional SARS-CoV-2 spike receptor binding domain in the green algae *Chlamydomonas reinhardtii*. (BioRxiv)
Bijie Ren, Song Cao, Arthur Grossman, Stephen Mayfield. Discovery of two master regulators for lipid accumulation *Chlamydomonas reinhardtii* through Yeast Two-Hybrid sequencing assay. To be submitted to *Nature communication*.
Bijie Ren, Terry Hwa, Jahir Gutierrez, Frank Fields, Nathan Lewis, Stephen Mayfield. Genome-scale reconstruction of protein production/secretion pathway in algae reveals bimodal correlation between growth rate and protein production.
Bijie Ren, Yuansheng Zhou, Arthur Grossman, Stephen Mayfield. Systematic Identification of Transcription regulatory subnetwork for *C.reinhardtii* by high throughput sequencing. (ready to submit)
Paul Gamble, **Bijie Ren**, Shashwat Kishore, Adrian Cantu, Marina Mizar, Ben Busby. Biollante: A Machine learning tool to detect contamination in plant genomes. *BioRxive*.
Amr Badary, **Bijie Ren**, Yasin Torres-tiji, Stephen Mayfield. Understanding the relevance of the 5' UTR – Analysis and design of improved genetic tools for recombinant gene expression. (ready to submit)

FIELDS OF STUDY

Major field: Biology

Studies in Quantitative Biology

Professors Stephen Mayfield, Nathan Lewis, Terence Hwa

ABSTRACT OF THE DISSERTATION

Identification of transcription factor interaction subnetwork and Modeling for metabolic-protein secretion in
Chlamydomonas reinhardtii

by

Bijie Ren

Doctor of Philosophy in Biology with a Specialization in Quantitative Biology

University of California San Diego, 2021

Professor Stephen Mayfield, Chair

The world faces global warming, high cost of therapeutic proteins and natural products, serious pollution in developing countries, and potentially energy crisis in next 100 years. One potential solutions to solve the above problems is to discover useful genetic elements and enzymes from novel abundant microorganisms in nature, build a specific cell factory and perform state-of-art bioprogramming. However, it's time-consuming to discover and identify these useful molecular elements from these less-studied organisms.

Chlamydomonas reinhardtii is a simple unicellular photosynthetic microalgae, which exhibit unique benefits to address global warming, climate change, development of renewable energy, and reduce costs of therapeutic proteins. Here we tried a preliminary pipeline to show how we could quickly discover transcription factor (TF) interaction mechanism, model metabolic flux and protein secretion, and engineer the genome of *Chlamydomonas reinhardtii*. Such a pipeline will help us understand the fundamental regulation of photosynthetic cell, and enable us to build a reliable whole cell model, which will be used for bioprogramming of algae to make many natural products.

Introduction

Recombinant proteins have brought unquantifiable amounts of benefits to the world from clinical research to biological drugs. However, the cost of therapeutic proteins and antibodies are still extremely expensive. Yeast and Chinese Hamster Ovary (CHO) cells are two available platforms that express such complicated enzymes and antibodies. However, CHO cell production remains very expensive due to the high cost of serum medium, while the addition of numerous mannose modification on yeast recombinant proteins can trigger allergic response in humans.

The eukaryotic algae *Chlamydomonas reinhardtii* has been reported to be able to express bioactive mammalian proteins and complex human antibodies (Tran et al, 2009, Rasala et al 2012). Total cost of algae expressed protein can be orders of magnitude lower than similar proteins produced from CHO cell. In addition, algae endoplasmic reticulum (ER) and plastids both much more chaperones that are required for folding complex mammalian protein and antibodies than that in yeast. Moreover, *Chlamydomonas reinhardtii* has been found to be “generally regarded as safe” (GRAS) for human consumption by the FDA (<http://www.cfsan.fda.gov>). Based on these advantages, we used algae to express a novel anti-inflammatory cytokine, TGM, from a parasite worm and tested its efficiency on Treg cell induction and potential anti-inflammatory effects on a mouse model of Inflammatory bowel disease.

Chapter 1 is an exciting report of production of this novel parasite anti-inflammatory cytokine expressed in algae, which has the potential to be utilized as a cost-effective new therapeutic strategy for inflammatory bowel disease. This paper describes the design and expression of recombinant TGM protein in the culture supernatant of algae, high throughput identification of strains expressing TGM protein, and assessing efficiency of TGM in T cell line and when delivered orally in a mouse Inflammatory bowel disease model. We were able to demonstrate that the algal host successfully produced the TGM1 recombinant protein, and that the recombinant protein is biologically active in both cell-based assays and when delivered orally to mice.

Chapter 2 presents a high throughput sequencing assay to identify the transcription factor interactions from the *Chlamydomonas reinhardtii*. Here we applied a recently published CrY2H-seq method to uncover the transcription factor (TF) interactome of *Chlamydomonas reinhardtii*, which lies in the center of its transcriptional regulation and bridges various biological processes. By analyzing conditional transcriptomics data in the context of TF interactions, we illustrated how transcription factors respond to different environmental stimuli to coordinate different gene regulation pathways in *Chlamydomonas reinhardtii*. The result provides a comprehensive picture of the organism’s gene regulatory network that can be utilized towards its future genetic engineering.

Chapter 3 attempts to build a genome scale metabolic-protein secretion model for *Chlamydomonas reinhardtii*. In this chapter, we integrated a comprehensive pathway for protein production/PTM/secretion into the current metabolic model iCre1355. This refined reconstruction enables us to calculate energy cost of each protein under Autotrophic/Mixotrophic/Heterotrophic conditions and shows that *Chlamydomonas* has different growth capacity and maximal protein secretion yields under each condition. Moreover, there is an anti-correlation between protein secretion yield and overall growth rates, due to the limited growth capacity of these cells, which is demonstrated by our independent biophysical model and experimental data. In addition, this reconstruction model also indicates that if we knock out unnecessary subsystems in the cell, the energy and resources saved could be used in other subsystems, including the protein secretion pathway. We also identify that addition of cAMP will increase the lipid accumulation without affecting the secretion of proteins of interest, which provides a new strategy to utilize algae as a cost-effective cell factory for biofuels, therapeutic antibodies, or cytokines.

CHAPTER 1:

Oral delivery of a functional algal expressed TGF- β mimic halts colitis in a murine DSS model.

Oral delivery of a functional algal expressed TGF- β mimic halts colitis in a murine DSS model.

Bijie Ren, Danielle J. Smyth, Madeleine White, Francis Fields, Rick M. Maizels, Stephen Mayfield

Highlights (bullet points)

- An immune-regulatory cytokine from a parasitic worm expressed in green algae induces an anti-inflammatory response when fed to mice
- Recovery from inflammatory colitis is accelerated by the algal-expressed protein
- Oral administration offers an attractive route for the novel therapy

Abstract

Inflammatory bowel disease (IBD) is a set of immunological disorders which can generate chronic pain and fatigue associated with the inflammatory symptoms. The treatment of IBD remains a significant hurdle, with current therapies being only partially effective or having significant side effects, suggesting that new therapies that elicit different modes of action and delivery strategies are required. TGM1 is a TGF- β mimic that was discovered from the intestinal helminth parasite *Heligmosomoides polygyrus* and is thought to be produced by the parasite to suppress the intestinal inflammation response to help evade host immunity, making it an ideal candidate to be developed as a novel anti-inflammatory bio-therapeutic. Here we utilized the expression system of the edible green algae *Chlamydomonas reinhardtii* in order to recombinantly produce active TGM1 in a form that could be ingested. *C. reinhardtii* robustly expressed TGM1, and the resultant recombinant protein is biologically active as measured by regulatory T cell induction. When delivered orally to mice, the algal expressed TGM1 is able to ameliorate weight loss, lymphadenopathy, and disease symptoms in a mouse model of DSS-induced colitis, demonstrating the potential of this biologic as a novel treatment of IBD.

Key words: Algae, Helminth, Treg, TGM, TGF- β , Colitis

1. Introduction

Inflammatory Bowel Disease (IBD) is a chronic immunological disorder of the gastrointestinal tract that affects millions of people in the world (de Souza and Fiocchi, 2016; Kaser et al., 2010). Every year, hundreds of thousands of new patients suffer from this disease, with a high risk of becoming seriously sick and debilitated, creating an enormous socioeconomic burden (Burisch et al., 2019). The two principal types of IBD are Crohn's disease (CD) (Baumgart and Sandborn, 2012) and ulcerative colitis (UC) (Danese and Fiocchi, 2011). While diagnosis is well-developed, no curative treatments for these diseases are available, with surgical resection a last resort. IBD can be triggered by poorly understood physiological states, and by numerous environmental factors including diet, stress, and infections. Although the detailed mechanism of IBD development is unclear, it is well acknowledged that IBD is associated with a failure of immune regulatory mechanisms, resulting in chronic inflammation and production of pro-inflammatory cytokines, leading to tissue destruction and tumorigenesis (Burisch et al., 2019; Neurath et al., 2002; **Figure 1.1**).

Key inflammatory cytokines have been targeted for IBD treatment, with the most effective therapy found to be antibody against tumor necrosis factor (TNF); however, this is not effective for patients with few immune cells expressing membrane-bound TNF (Wong et al., 2008). An alternative strategy is to deliver recombinant anti-inflammatory cytokines. These include IL-10 (Saraiva et al., 2020), IL-22 (Sugimoto et al., 2008), IL-27 (McLean et al., 2017) and TGF- β (Ihara et al., 2017), which are produced by mucosal immune cells and/or epithelium, which exhibit both potent anti-inflammatory and tissue-repair effects, through regulatory T cell generation and enhancing epithelial cell function. However, the paracrine manner in which these cytokines act, creates a challenge in delivering pharmacologic amounts of these factors only to the gastrointestinal mucosa. In animal models, pathology can be suppressed by pre-treatment with anti-inflammatory cytokines such as IL-10 (Huber et al., 2011), while established disease is ameliorated by oral delivery of IL-10-expressing transgenic *Lactococcus lactis* (Steidler et al., 2000). In clinical trials, systemic IL-10 administration showed no benefit at doses low enough to avoid significant side effects (Buruiana et al., 2010; Schreiber et al., 2000), and published reports on oral delivery have been limited to safety trials in healthy individuals (Baat et al., 2006).

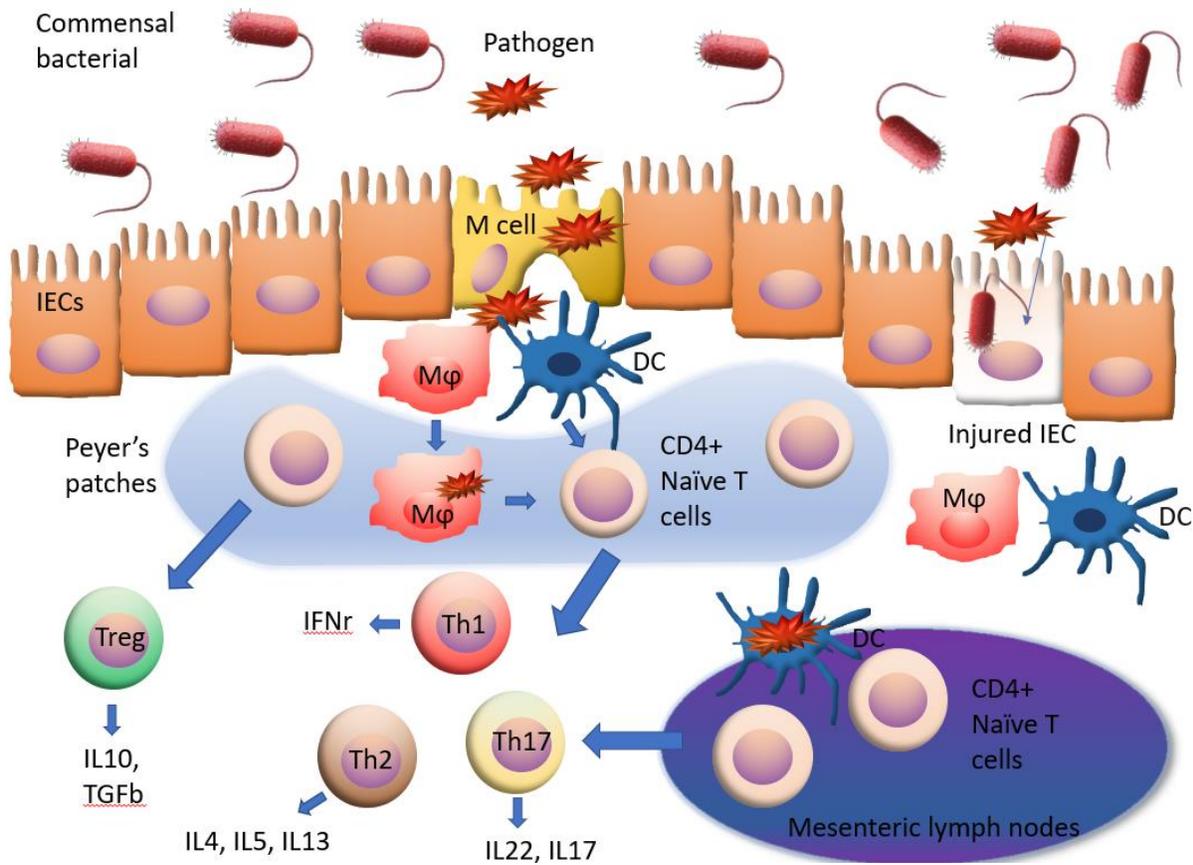


Figure 1.1 Key immune cells induced in Inflammatory bowel disease. DC, dendritic cell; Mphi, microphage cell; IEC, intestinal epithelial cell; Th, helper T cells; Treg, regulatory T cells.

TGF- β has wide-ranging effects in immune suppression and wound repair (David and Massague, 2018) and plays a prominent role in maintaining homeostasis of the mucosal immune system (Ihara et al., 2017; Konkel and Chen, 2011). In colitis, elevated TGF- β expression dampens inflammation, and one therapeutic avenue has been to boost TGF- β efficacy by anti-sense RNA abrogation of the Smad7 inhibitor of the TGF- β signaling pathway (Monteleone et al., 2015). Transgenic *Bacteroides ovatus* which can be induced to release TGF- β have also been shown to downmodulate colitis in mouse models (Hamady et al., 2011).

An alternative strategy to ameliorate inflammatory bowel diseases has emerged from studies with parasitic helminth worms (Weinstock and Elliott, 2013). Globally, there is an inverse relationship between the prevalence of intestinal helminth parasites and the incidence of IBD (Varyani et al., 2017); in mouse models, infections with diverse helminth species can abrogate colitis (Hunter et al., 2005; Leung et al., 2012; Smith et al., 2007), and deliberate infection of IBD patients has been advocated as a new therapy for disease (Summers et al., 2005). One species

associated with anti-inflammatory effects *in vivo*, *Heligmosomoides polygyrus*, has been shown to secrete a potent TGF- β mimic, TGM1 (Johnston et al., 2017; Smyth et al., 2018). Parasite TGM1 and mammalian TGF- β have no sequence similarity, however TGM1 ligates the mammalian TGF- β receptor (TBR1), initiating downstream Smad signaling and inducing suppressive Foxp3⁺ regulatory T cell (Tregs) (Cook et al., 2021; White et al., 2021). We are therefore investigating whether oral therapy with helminth-derived TGM, rather than live infection with helminth parasites, would be the therapy of choice for IBD.

In this study, we tested the idea of producing recombinant TGM1 in the edible green algae *Chlamydomonas reinhardtii*, in order to be able to feed TGM1 directly to mice and assess the protein's action from within the gastrointestinal tract, and with the view to developing a bio-therapeutic for use in humans to treat a range of gastrointestinal inflammatory diseases. Compared to other cell-based expression systems, the use of algae as a cellular factory has several benefits, including cost effective cultivation under photosynthetic growth, and rapid low cost scaling from lab to commercial scale, rendering it orders of magnitude less costly than conventional mammalian cell culture technologies (Dove, 2002). Additionally, unlike other low-cost cell systems such as yeast or bacteria, algae have the cellular machinery in both the endoplasmic reticulum (ER) and chloroplast, that are required for assembling complex mammalian proteins and antibodies (Gregory et al., 2013; Patra et al., 2015; Ravi et al., 2018; Tran et al., 2013). We have previously demonstrated that *C. reinhardtii* can be engineered to express a number of bioactive proteins (Gimpel et al., 2015; Rasala et al., 2012; Tran et al., 2009) and that the biomass can be safely consumed orally in mice and humans (Fields et al., 2020), and shown to be able to deliver recombinant protein cargoes to the intestinal tract of mice (Barrera et al., 2015). We have now expressed an active novel anti-inflammatory cytokine TGM1 in *C. reinhardtii* and further show that the algal TGM1 when given orally is able to regulate immune cells and protect mice from DSS colitis weight loss.

2. Materials and Methods

2.1. Plasmid construction for recombinant TGM1 protein expression

To express recombinant TGM1 using *C. reinhardtii*, we designed a codon-optimized TGM1 construct with a FLAG tag based on the previously described secretion vector pBR9 (Rasala et al., 2013), to maximize protein expression (**Figure 1.2 A**). The N-terminal ars1 algae signal peptide allows for the secretion of TGM1 into the medium. Additionally, the AR1 promoter and rbcS 3' UTR ensure that the transformed construct locates to the nucleus and that protein is continuously expressed under light conditions. pBR9 also encodes the Bleomycin selection gene Ble2A to

isolate transformed algae. As previous work had shown that domains 1-3 of TGM1 were sufficient for biological activity (Smyth et al., 2018), a similar construct was made to also express truncated TGM1 (TrTGM1).

The sequence for mature TGM1 (or trTGM1) was adapted to the nuclear codon usage of *C. reinhardtii* according to the Kazusa DNA Research Institute's database (<http://www.kazusa.or.jp/codon/cgi-bin/showcodon.cgi?species=3055>) as previously described (Rasala et al., 2013), and is presented in **Supplementary Figure 1**. Synthetic genes plus an 8-aa FLAG tag (DYKDDDDK) were synthesized by Genewiz and cloned into pBR9 as *XhoI/BamHI* fragments. The resultant constructs, named pBR9-TGM1 and pBR9-TrTGM1, were transformed into wild type CC-1690 *C. reinhardtii* by electroporation with high-level expressing transformants being selected by dot-blot screening for further large-scale expression and purification (**Figure 1.2 B**).

2.2. *C. reinhardtii* transformation and screening

2.2.1 Transformation

C. reinhardtii wild type strain CC-1690 was used for secreted TGM1 protein expression. *C. reinhardtii* was grown in tris-acetate-phosphate (TAP) (Gorman and Levine, 1965) liquid medium at room temperature (RT) on a rotary shaker set at 100 rpm, under constant light intensity ($100 \mu\text{mol photons m}^{-2} \text{s}^{-1}$) to a concentration of approximately $1 \times 10^6 \text{ cells mL}^{-1}$. Cells were harvested by centrifugation (2,000 g) and then resuspended with MAX Efficiency® Transformation Reagent for Algae (Thermo Fisher Scientific Inc., MA, Cat #A24229). One million cells in 300 μl transformation reagent were mixed with 5 μg of linearized plasmid digested with *KpnI* and incubated on ice for 30 mins. The cell suspension was placed into a prechilled disposable electroporation cuvette with a 4-mm gap (Bio-Rad Labs., Hercules, CA, Cat #1652081) for 5 mins on ice. A BioRad Gene Pulser Xcell with the following settings (Voltage 500 V, Capacity 50 μF and Resistance 800 Ω) was used for electroporation. Electroporated *C. reinhardtii* cells were recovered with 10 mL of 20 mM sucrose in TAP medium for 24 hrs, and then plated on TAP agar plates with 15 $\mu\text{g mL}^{-1}$ Zeocin and incubated at RT for 10 days.

2.2.2 Dot blot screening

A dot blot assay (Rupprecht et al., 2010) was modified to screen transformed colonies in order to select clones with higher TGM1 secretion yields. Nitrocellulose membrane of 7 x 7cm was cut to fit the Fisherbrand™ Petri Dish with a 10cm diameter (Fisher Scientific, Cat #FB0875713) and gridded to 100 small squares with similar size for algae colony growth. The gridded Nitrocellulose membrane was placed onto each TAP agar plate in 10cm-diameter Petri dish. 100 transformed of the colonies were randomly picked and patched with the same biomass onto the nitrocellulose

membrane in a grid formation and incubated under light at RT for three days. After three days all colonies were washed off the nitrocellulose membrane with TBST (Tris-buffered saline plus 0.05 % Tween-20). The nitrocellulose membrane was blocked with 5 % non-fat milk powder in TBST for 1 hour and then incubated with Monoclonal ANTI-FLAG® M2-Alkaline Phosphatase antibody (Sigma-Aldrich, St. Louis, MO, USA, Cat #A9469; final dilution 1:5000) for 1 hour. The membrane was then washed with TBST 3 x 5 mins and visualized in nitro-blue tetrazolium (NBT) / 5-bromo-4-chloro-3'-indolyphosphate p-toluidine salt (BCIP) (Sigma-Aldrich, St. Louis, MO, USA, Cat #B5655) added directly to the membrane for development. Developed dot blot intensity was correlated with yield of secreted TGM1 and the transgenic clones with highest protein secretion were selected for further protein expression and purification.

2.2.3 Western blot and Coomassie Blue staining

Transgenic TGM1 strains were grown in Tris-acetate-phosphate (TAP) (Gorman and Levine, 1965) liquid medium. Cells were cultured at RT, on rotary shaker set at 100 rpm, under constant light intensity ($100 \mu\text{mol photons m}^{-2} \text{s}^{-1}$), to a cell concentration around $6 \times 10^6 \text{ cells mL}^{-1}$. Supernatants were harvested by centrifuging cells at 7,800g, added to 4 volumes of prechilled acetone (Sigma-Aldrich Cat # 650501) and incubated at 4 °C overnight. Precipitated protein pellets were obtained by centrifuging acetone-treated supernatants at 17,000 g for 10 mins at 4 °C and air dried for 5 mins. ddH₂O was used to resuspend the dried protein pellet and 4X Laemmli sample buffer (BioRad, Cat # 1610747) with 10 % β-mercaptoethanol (Sigma-Aldrich Cat # M6250) was then added as 1:4 ratio (sample buffer: sample). Reduced samples were heated at 65 °C for 10 mins and loaded onto 10 % Mini-PROTEAN® TGX™ Precast Protein Gels (BioRad, Cat # 4561036). The molecular weight marker used was AccuRuler RGB Plus Prestained Protein Marker (Biopioneering, Cat # PM-001). A voltage of 120 was applied and the gels were run for 1 hr. One gel was stained in Coomassie SimplyBlue™ SafeStain (Thermos Fisher Scientific, Cat # LC6060) for 1 hour, followed by destaining in ddH₂O. Another gel was transferred to nitrocellulose membrane (BioRad, Cat # 1620112). The membrane was blocked with 5 % non-fat milk in TBST for 1 hour at RT. The membrane was probed with Monoclonal ANTI-FLAG® M2-Alkaline Phosphatase antibody as described above for 1 hour and then washed with TBST 3 x 5 mins and NBT/BCIP used to visualize the protein band on the membrane.

2.3. *C. reinhardtii* bulk expression conditions for algal TGM1

Transgenic TGM1 and TrTGM1 strains were grown with 1 L TAP liquid medium in 2 L Erlenmeyer flasks. These culture flasks were maintained at RT on a rotary shaker set at 100 rpm, under constant light intensity ($100 \mu\text{mol}$

photons $\text{m}^{-2} \text{s}^{-1}$), to a cell concentration of approximately 6×10^6 cells mL^{-1} . Supernatants were harvested by centrifugation (7,800 g) and then filtered by Whatman Grade GF/A Fine Retention Filter paper (GE Healthcare Life Science, 1820-021). Filtered supernatant was further processed to obtain recombinant protein with either FLAG affinity purification or ammonium sulfate precipitation (**Figure 1.2 C, D**). Wild type algae strain CC-1690 was grown under the same conditions and treated in the same manner as the transgenic TGM1 strain and used as an algal protein control for assays and *in vivo* mouse models.

2.4. Purification of secreted algal TGM1 from *C. reinhardtii*

2.4.1. Anti-FLAG resin purification of algal TGM1 and TrTGM1

FLAG affinity chromatography was used to purify TGM1 and TrTGM1 for mass spectrometry, glycosylation analysis and *in vitro* cell-based assays for activity. Briefly, 1 mL of anti-FLAG affinity gel (Sigma-Aldrich Cat # A4596) was equilibrated in lysis buffer (50 mM Tris, 150 mM NaCl, 1 mM EDTA, pH 7.4) with 5 column volumes (CV) and then loaded with 2 L filtered supernatant. The resin was washed with 5 CV of lysis buffer and then eluted with 6 sequential 1 CV of elution buffer (100 mM Glycine, 400 mM NaCl, pH 3.5). The pH of the eluate was immediately adjusted to pH 7.5 with 50 μL 1 M Tris-HCL. 10ul of each purified protein sample (TGM1 and TrTGM1) and 50ng 45-KDa Recombinant Posi-Tag Epitope Tag Protein containing the epitope FLAG tag (BioLegend, Cat #931301) was loaded to SDS-PAGE gel and quantified by the western blot as previously described. The protein band density in western blot was quantified by Image Studio Lite (LI-COR Biosciences, Lincoln, NE, USA). The FLAG-purified TGM1 was then sent for mass spectrometry, glycosylation analysis and both FLAG-purified TGM1 and TrTGM1 were used for *in vitro* cell based activity assays.

2.4.2. Ammonium sulfate precipitated TGM1 and wild type algal proteins for *in vivo* mouse models.

For larger scale isolation of TGM1, ammonium sulfate was added slowly to filtered algae supernatant to a final salt concentration of 250 g.L^{-1} . The mixture was incubated at 4 °C overnight and then centrifuged at 7,800 g to harvest the precipitated protein pellet. The precipitated protein pellet was resuspended in DPBS (Thermo Fisher, Cat #14190144) and dialyzed twice with Thermo Scientific™ SnakeSkin™ Dialysis Tubing (10 K MWCO, Thermo Fisher, Cat # 68100) overnight in DPBS at 4 °C. To protect proteins from aggregation and increase their stability sterile glycerol was added into the samples with a final concentration of 2 % (v/v). 2ul of precipitated supernatant and 150ng of Recombinant Posi-Tag Epitope Tag Protein as a FLAG standard were loaded to SDS-PAGE gel and protein band pixel intensities were quantified by Image Studio Lite. Total protein concentration in precipitated

supernatant was quantified by Pierce™ BCA Protein Assay Kit (Thermo Fisher, Cat #23225) at 562nm against BSA standards.

Wild-type (WT) algal proteins used as a control in assays and mouse models was CC-1690 *C. reinhardtii* WT filtered supernatant processed with the same procedure as above and the final concentration of these supernatants were measured by Pierce™ BCA Protein Assay Kit.

2.5. Glycosylation analysis and mass spectrometry of FLAG-purified algal TGM1

For glycosylation analysis, 1 µg FLAG-affinity purified algal TGM1 samples were digested with either 5 U of PNGase F (NEB, Cat # P0704S) or 5 U EndoH (NEB, Cat # P0702S) at 37°C for one hour. Digested proteins were heat inactivated at 65°C for 10 mins and analyzed by SDS-PAGE. For more detailed investigation, untreated protein samples in 50 mM ammonium bicarbonate, 0.1 % w/v ProteaseMax surfactant were reduced and alkylated prior to proteolytic digestion with trypsin at a ratio of 1:50. Proteolysis was stopped, and surfactant hydrolyzed by the addition of 0.5 % trifluoroacetic acid (TFA). Sample desalting was performed using HyperSep Hypercarb solid-phase extraction tips followed by drying by vacuum centrifugation (Eppendorf).

Liquid Chromatography with tandem mass spectrometry (LC MS/MS) was performed by nano-flow liquid chromatography (U3000 RSLCnano) coupled to a hybrid quadrupole-orbitrap mass spectrometer (Q Exactive HF). Peptides were separated on an Easy-Spray C18 column (75 µm x 50 cm) using a 2-step gradient from 97 % solvent A (0.1 % formic acid in water) to 10 % solvent B (0.08 % formic acid in 80 % acetonitrile) over 5 min then 10 % to 50 % B over 75 min at 300 nL.min⁻¹. Mass spectra were acquired with automated data-dependent switching between full-MS and tandem MS/MS scans using stepped collision energy. Data analysis was performed using BioPharma Finder 4.0. All materials and equipment were supplied by Thermo Fisher Scientific unless otherwise stated.

2.6. *In vitro* cell based assay for recombinant TGM1 bioactivity

2.6.1. TGF-β Bioassay.

The TGF-β bioassay (MFB-F11) developed by Tesseur et al., (Tesseur et al., 2006) was performed as previously described (Smyth et al., 2018). Briefly, confluent MFB-F11 cells were added to each well of a 96-well round-bottomed plate. Wild type algal proteins or log dilutions (starting at 100 ng.mL⁻¹ of purified proteins such as HEK293T-expressed TGM1 (expressed and purified as previously described (Johnston et al., 2017)) or algal-expressed TGM1/TrTGM1 were then added to each well in a volume of up to 50 µL and incubated for 24 hrs at 37°C in an atmosphere of 5 % CO₂. Subsequently, 20 µL of supernatant was aspirated from each well, added to an ELISA

plate (NUNC) with 180 μL of reconstituted Sigma FastTM *p*-nitrophenyl phosphate substrate and incubated at RT in the dark for up to 4 hrs. Plates were read on at 405 nm on an Emax precision microplate reader (Molecular Devices). All conditions were set up in duplicate and repeated at least twice.

2.6.2. Foxp3⁺ Treg Induction Assay.

As described previously (White et al., 2021), a single cell suspension was prepared from the spleens of naïve Foxp3-GFP BALB/c transgenic mice (Fontenot et al., 2005), with contaminating red blood cells removed by resuspending the cells from one spleen in 2 ml of red blood cell lysis buffer (Sigma) and incubating at RT for 2 mins. Cells were then washed and resuspended in RPMI medium supplemented with 2 mM L-glutamine (Gibco), 100 U.mL⁻¹ penicillin (Gibco), 100 $\mu\text{g.mL}^{-1}$ streptomycin (Gibco), 10 % heat-inactivated fetal calf serum (FCS) (Gibco), and 1x MEM non-essential amino acid solution (Gibco). CD4⁺ T cells were enriched by magnetic sorting using the mouse naïve CD4⁺ T cell isolation kit (Miltenyi 130-104-453) on an autoMACS Pro Separator (Miltenyi) as per manufacturer's instructions. Cells were cultured at 4×10^5 per well in flat-bottomed 96-well plates (Costar) pre-coated with 10 $\mu\text{g.mL}^{-1}$ of anti-CD3 (clone 145-2C11; Invitrogen) with the addition of IL-2 (Miltenyi) at a final concentration of 400 U.mL⁻¹ and with or without retinoic acid (Sigma) at a final concentration of 1 nM. Purified mammalian TGF- β (Peprotech), HEK293T expressed TGM1 (expressed and purified as previously described, (Johnston et al., 2017)) or algal TGM1/TrTGM1 proteins were added at a concentration of 20-100 ng.mL⁻¹ and cells cultured at 37 °C in 5 % CO₂ for at least 72 hrs before being removed for flow cytometric analysis. All conditions were set up in triplicate and repeated at least twice.

2.7. Flow Cytometric Analysis.

For viability staining, LIVE/DEAD[®] fixable blue (Life Technologies) was diluted to 1:1000 in PBS; 100 μL was added to each sample of cells, which were then incubated in the dark for 20 mins at 4°C and washed twice in FACS buffer. To prevent non-specific antigen binding, cells were incubated with 50 μL of polyclonal IgG (Sigma) (diluted 1:50 in FACS buffer) for 10 mins at 4 °C and then washed twice in FACS buffer. Cells were surface stained with anti-CD3-BV711 (17A2, Biolegend), anti-CD4-PerCP/Cy5.5 (GK1.5, BioLegend) and anti-CD25-BV650 (PC61, Biolegend) to a total volume of 50 μL (diluted 1:200). Separate Foxp3 staining was not always required as cells were from Foxp3-GFP transgenic mice however; to assess transcription factor expression, cells were stained with anti-Foxp3-eF450 (FJK-16s, eBioscience) and anti-ROR γt -PE (Q31-378, BD Biosciences), both at 1:100 dilution, using the Transcription Factor Staining Buffer set (eBioscience) and following manufacturer's instructions. Single

stain controls were individually added to one drop of UltraComp eBeads (eBioscience). Samples were incubated for 20 mins at 4°C, washed twice in FACS buffer and then resuspended in 200 µL FACS buffer. All samples were acquired on a BD Biosciences Celesta and analyzed using FlowJo software (Tree Star).

2.8. Animal models

Female 6 - 8 week old C57BL/6J mice were purchased from Envigo (Huntingdon, UK) and housed in specific pathogen-free conditions and acclimatized in the animal unit for at least 1 week after arrival before experimental models were set up. All animal work was approved by the Ethical Review Board of the University of Glasgow and procedures were performed under a UK Home Office licence.

2.8.1. Mouse feeding trial.

For initial evaluation of algal expressed TGM1 on mouse health and whether it would be tolerated/remain active after being ingested, we added algal expressed recombinant TGM1 protein directly to drinking water and allowed mice to drink ad libitum. Body weight and food/water intake of the mice were monitored daily. For the control group, WT *C. reinhardtii* algal proteins or PBS/ 2 % (v/v) glycerol were added to drinking water at the same volume and timing as the algal TGM1. Mice were sacrificed and isolation of lymphocytes from the Peyer's patches (Mosconi et al., 2015), mesenteric lymph nodes (MLNs) (White et al., 2021) lamina propria of small (Webster et al., 2020) and large (Clay et al., 2020) intestines were analyzed by flow cytometry.

2.8.2. DSS colitis model.

The dextran sodium sulfate (DSS) chemical model of mouse model of colitis (Chassaing et al., 2014) was selected to determine the effects of algal expressed TGM1 in a murine model of acute colitis. Dextran sodium sulfate (36,000-50,000 MW, MP Biomedicals, Santa Ana, California, USA) was added to the drinking water at 2 % (w/v) for a total of 4 days to disrupt the gut mucosal barrier, resulting in acute colitis. Mice were then given normal drinking water for the remainder of the experiment in order to monitor weight recovery. *C. reinhardtii* expressed TGM1 (ammonium sulfate precipitated from the supernatant) was also added to the drinking water initially throughout the 4-day DSS administration and for an additional 3 days (equaling a total of 7 days of algal TGM1 administered to the mice).

Mice were weighed and scored daily for the entirety of the experiment using a Disease Activity Index (DAI) matrix which comprised of scores out of 4 for % weight loss, blood, stool consistency and general appearance. Each day scores for each parameter were summed to give a DAI total out of 16 for each mouse. Animals were euthanized

if they lost greater than 20 % of their initial weight. Peyer's patches and mesenteric lymph nodes were isolated for analysis by flow cytometry.

2.9. Statistical tests.

Graphs and statistics were analysed using Prism (GraphPad, San Diego, California, USA). Shown are the means \pm standard error (SEM), and one-way or two-way ANOVA or *t*-tests (paired or unpaired) were used where appropriate, with non-parametric Kruskal-Wallis tests being applied if data were not normally distributed. **P* <0.05

3. Results

3.1. Expression and glycan analysis of *C. reinhardtii* expressed TGM1

To express recombinant forms of the parasite TGF- β mimic TGM1 and TrTGM1 using *C. reinhardtii*, we chose to utilize a vector (pBR9) that would insert the gene cassette into the nucleus and secrete the recombinant protein continuously under light conditions (**Figure 1.2 A**). Our strategy of expression included screening recombinant algae for high-expressing clones by a modified dot-blot assay, and purification of TGM1 proteins via a FLAG-tag suitable for antibody affinity chromatography (**Figure 1.2 B**). This allowed us to test algal expressed TGM1 functionally, comparing it to the mammalian (HEK293T) expressed TGM1.

For both constructs, single bands were detected in the algal supernatant by anti-FLAG Western blot (**Figure 1.2 C**). Supernatant from transgenic TGM1 expressing strain was precipitated by ammonium sulfate and resuspended in DPBS. 2ul of precipitated supernatant and 150ng of Recombinant Posi-Tag Epitope Tag Protein as a FLAG standard were loaded to SDS-PAGE gel and quantified by Image Studio Lite. Total protein concentration in the precipitated supernatant was measured by Pierce™ BCA Protein Assay Kit (Thermo Fisher, Cat #23225) at 562nm against BSA standards. Quantification of the recombinant TGM1 against FLAG standard showed that the recombinant protein accumulated to approximately 5 % of total soluble protein (TSP) in the supernatant (**Figure 1.2 D, Supplementary Table 1.1**). Algal TGM1 was affinity purified using anti-FLAG M2 affinity resin, and the TGM1 amino acid sequence was confirmed using mass spectrometry (**Supplementary Figure 1.2 A**).

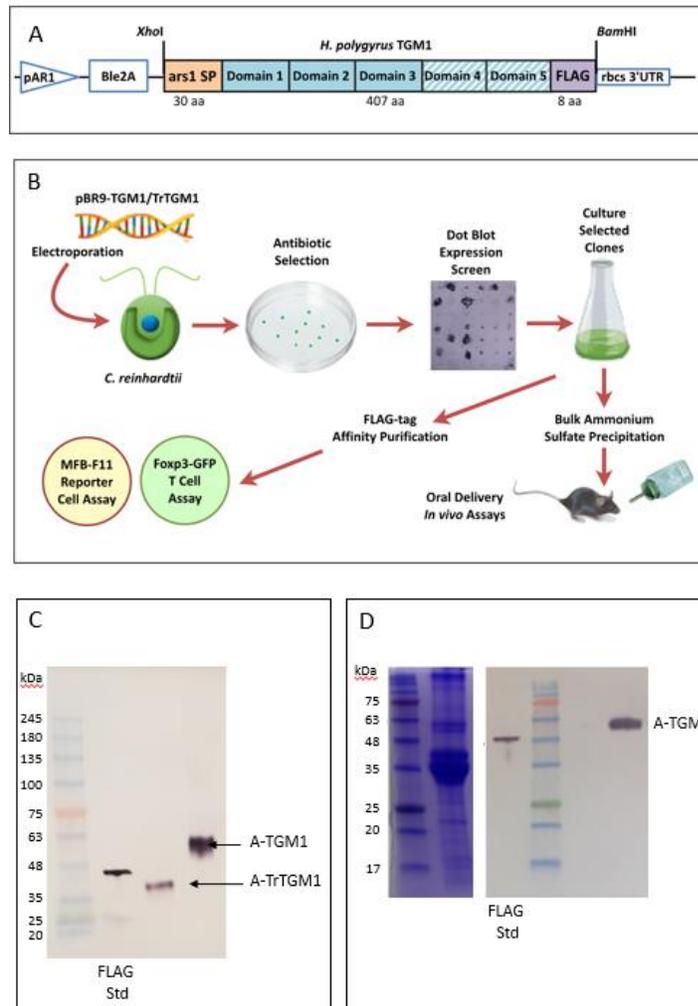


Figure 1.2 Expression of *Heligmosomoides polygyrus* TGM1 in *Chlamydomonas reinhardtii*. **A.** Construct design comprising pAR1, a constitutive promoter (670bp) used for high level protein expression in *Chlamydomonas*; Ble2A, an antibiotic resistance gene; Ars1 SP, a 30-aa secretion signal peptide (orange filled box) codon-optimized mature 5-domain TGM1 (blue boxes), or truncated TGM1 (TrTGM1, lacking Domains 4 and 5, hatched blue); FLAG tag (purple); and rbcS 3' UTR (246bp), a robust element to maintain algal mRNA stability. The sequence encoding the expressed protein was inserted into pBR9 through *XhoI* and *BamHI* restriction sites as indicated, and is given in full in Supplementary Figure 1. **B.** Schematic of experimental approach including dot blot screening to select high expressing clones, and purification of TGM1 by affinity chromatography (C) for *in vitro* bioassays (see Figure 2) or ammonium sulfate precipitation for *in vivo* administration (see Figures 3 and 4). **C.** Western blot of anti-FLAG antibody affinity-purified TGM1 and TrTGM1, stained with Monoclonal ANTI-FLAG® M2-Alkaline Phosphatase antibody. 10ul of each purified protein sample (TGM1 and TrTGM1) and 50ng 45-KDa Recombinant Posi-Tag Epitope Tag Protein containing the epitope FLAG tag (BioLegend, Cat #931301) was loaded to SDS-PAGE gel. Marker protein molecular weights are indicated. **D.** Coomassie Blue stained (left hand panel) and anti-FLAG Western blot (right hand panel) of ammonium sulfate-precipitated TGM1. 2ul of TGM1 sample and 150ng Recombinant Posi-Tag Epitope Tag Protein was loaded to SDS-PAGE gel. Marker protein molecular weights are indicated.

We next assessed the level of glycosylation of the recombinant TGM1 expressed by *C. reinhardtii*, as potential modification with plant-specific glycans, including core β 1,2-xylose and α 1,3-fucose, are reported to cause

immunogenicity of recombinant proteins and evoke human antibody responses (Bardor et al., 2003). We subjected FLAG-purified algal TGM1 to digestion with PNGase F and Endoglycosidase H. Only PNGase F appeared to show minor changes to the FLAG-purified algal TGM1 shown by the slight laddering of the protein on SDS-PAGE analysis after digestion (**Supplementary Figure 1.2 B**). Algal TGM1 was subjected to further glycosylation analysis by mass spectrometry, identifying 4 sites of *N*-glycosylation, and 1 of *O*-glycosylation (**Supplementary Table 2**). Although two previous studies found that 10-20% of glycosylated cellular proteins from *C. reinhardtii* carried one or two xylose additions to the mature oligomannoside chain (Lucas et al., 2020; Mathieu-Rivet et al., 2013), we could not detect any xylosylation of the recombinant TGM1 protein. In addition, we did not detect plant specific core α 1,3-fucose on recombinant TGM1, consistent with this modification occurring in a core xylose-dependent manner (Oltmanns et al., 2019). Overall, only mammalian-like N-glycans were detected by our mass spectrometry analysis, and the PNGase F digestion (**Supplementary Figure 1.2 B**) indicated that their abundance was at a very low level.

3.2. Bioactivity of TGM1 expressed in recombinant algae

We next tested FLAG-purified recombinant TGM1 and TrTGM1 produced in *C. reinhardtii* in two *in vitro* assays, a TGF- β reporter bioassay using MFB-F11 cells in which Smad signaling drives plasmid-encoded secretory alkaline phosphatase (Tesseur et al., 2006), and a Treg induction assay in which naive Foxp3-negative mouse splenocytes from transgenic mice encoding a Foxp3-GFP fusion protein, are induced to express Foxp3-GFP (White et al., 2021). In the reporter assay, the algal-expressed cytokines showed bioactivity (**Figure 1.3 A**), while wild-type algal products showed no effects (**Supplementary Figure 1.3 A**). We noted, however, that the full-length algal TGM1 was approximately ten-fold less potent than the equivalent expressed in mammalian HEK293T cells. Nevertheless, at higher concentrations, algal TGM induced a stronger signal than even mammalian TGF- β . In addition, we found that the truncated TGM-1, lacking domains 4 and 5, was significantly weaker than the full-length expressed in the same algal system when tested in the MFB-F11 system, although at the highest concentrations it was able to equal the signal induced by the mammalian cytokine.

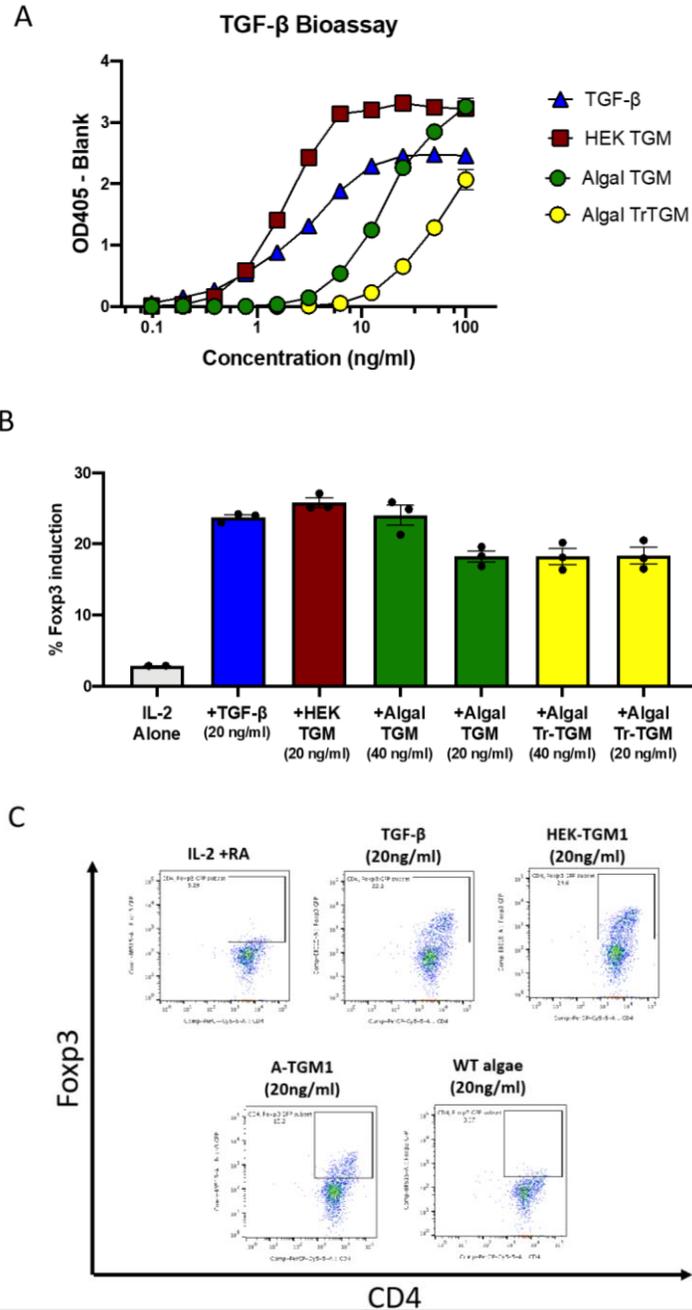


Figure 1.3 FLAG-purified algal TGM1 and its bioactivity on TGF- β -responsive cell assays. A. MFB-F11 TGF- β receptor binding assay. MFB-F11 cells encode a Smad-responsive reporter construct which induces secretion of alkaline phosphatase following ligation of the TGF- β receptors; alkaline phosphatase is detected by addition of *p*-nitrophenyl phosphate substrate generating a color reaction. Cells were incubated with mammalian TGF- β , HEK293 expressed TGM1, or algal secreted TGM1 / TrTGM1 at the indicated concentrations and separately for 24 hours, before recovery of supernatants for enzyme assay with *p*-nitrophenyl phosphate substrate. Data represent one of two similar experiments. **B.** Naïve T cells from transgenic mice encoding a Foxp3-GFP fusion protein (Fontenot et al., 2005) respond to TGF- β receptor ligation by expression of GFP that is detected by flow cytometry. Induction of primary mouse splenic CD4⁺ T cells to Foxp3-GFP expression. Data represent one of two similar experiments. **C.** Representative flow cytometry plots of Foxp3-GFP T cells treated with TGF- β or TGM1 from mammalian (HEK) or algal (A-TGM1) expression systems, together with IL-2 and retinoic acid (RA). Control cultures received IL-2+RA alone, or these mediators together with proteins from wildtype algae.

We next evaluated whether algal-produced TGM1 proteins could induce Foxp3⁺ Treg cells from naïve pre-sorted Foxp3-negative CD4⁺ T cells. Cell cultures were treated with mammalian TGF- β , HEK293T-expressed TGM1, or FLAG-affinity-resin purified algae TGM1, together with the growth factor IL-2. Three days later, flow cytometry was used to identify induced Foxp3-GFP⁺ Treg cells (**Figure 1.3 B**). In this system, both algal recombinant TGM1 and TrTGM1 proteins were able to induce Foxp3 to similar percentages as compared to TGF- β , and mammalian expressed TGM1, with the induction level being well above the background signal for the control group incubated with wild-type algal proteins alone (**Figure 1.3 C**). Because retinoic acid (RA) is a known enhancer of Foxp3 induction, similar assays were conducted with and without RA, but we found no significant increase in its presence with the different TGM preparations (**Supplementary Figure 1.3 B**).

3.3. Increase in mucosal Treg cell numbers following oral delivery of TGM1 to mice

We then explored whether purified algal TGM1 could be fed and tolerated by mice and if so, could induce any immunological changes in intestinal tissues. In view of the lower potency and reduced quantities available of the truncated TGM1 protein, these studies focused on the full-length construct. Given the dilute nature of the secreted algal TGM1, we administered TGM1 orally by addition to drinking water (at 100 $\mu\text{g}\cdot\text{mL}^{-1}$) and monitored water and food intake, and body weights, over time (**Supplementary Figure 1.4**).

Water intake was first ascertained and found to increase marginally above the average mouse consumption of 5 ml per day (indicated on **Supplementary Figure 4 A**); this equates to mice receiving approximately 0.5 mg of the algal TGM1 per day. Control mice received water with an equivalent concentration of WT algal secreting proteins including a matching volume of PBS/2 % glycerol used in adding TGM1 to drinking water. Over the course of an 8-day period, there was no significant difference measured between the control group with WT algae and group receiving the algal TGM1 water, with respect to water or food intake (**Supplementary Figure 4 A, B**) or weight (**Supplementary Figure 4 C**) indicating that algal TGM1 consumed in drinking water showed no detrimental effects on appetite or body weight.

At day 8, mucosal tissues (Peyer's patches, mesenteric lymph nodes, small intestine, colon,) were harvested and cells were isolated, enumerated and subjected to flow cytometric analysis (**Figure 1.4 A**). Peyer's patches are the specialized structures in the small intestinal epithelium that act as the first-line lymphoid organ for mucosal immune system, in which lymphocytes activated by epithelial dome microfold (M) cells that transport antigens from the

intestinal lumen (Da Silva et al., 2017). We noted with interest that both total cell counts, and the numbers of Foxp3⁺ Tregs, showed substantial increases in mice receiving TGM (Figure 1.4 B, C), although the frequency of Tregs within the total populations showed little change. Broadly similar patterns were observed in the mesenteric lymph nodes (MLNs, Figure 1.4 D, E), and in intestinal tissue preparations from the small intestine (Figure 1.4 F, G) and large intestine (Figure 1.4 H, I), although in the latter case total cell numbers declined slightly, resulting in an overall increase in Treg proportions.

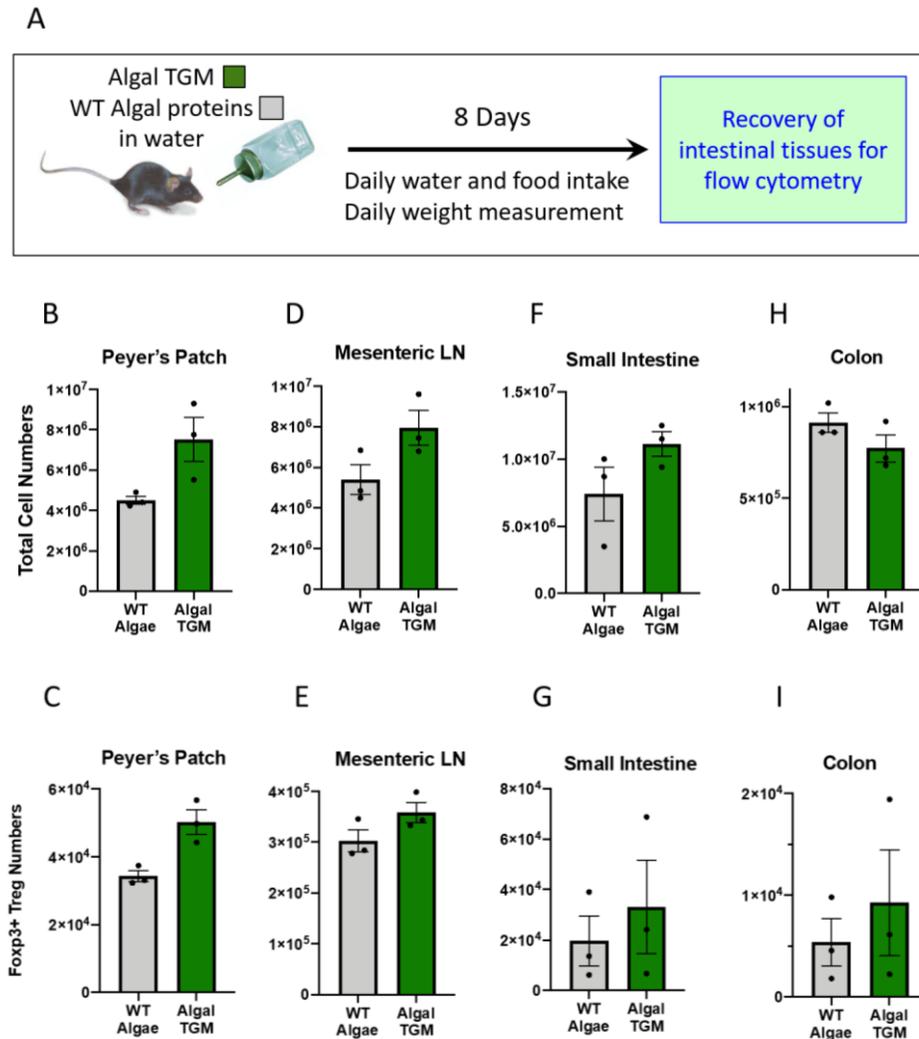


Figure 1.4 Effect of algal TGM1 on mucosal immune cell populations. **A.** Experimental design for algal TGM1 feeding trial. **B.** Cell counts and Foxp3⁺ Treg numbers in intestinal tissues following feeding of WT algal proteins or algal-expressed TGM1, in the Peyer's Patches (B, C), Mesenteric Lymph Nodes (D, E), Small Intestine (F, G) and Colon (H, I). Bars represent means and standard errors, with data from individual mice shown as solid symbols. Data represent one of two similar experiments.

3.4. Orally delivered algal TGM1 protects mice from DSS colitis weight loss.

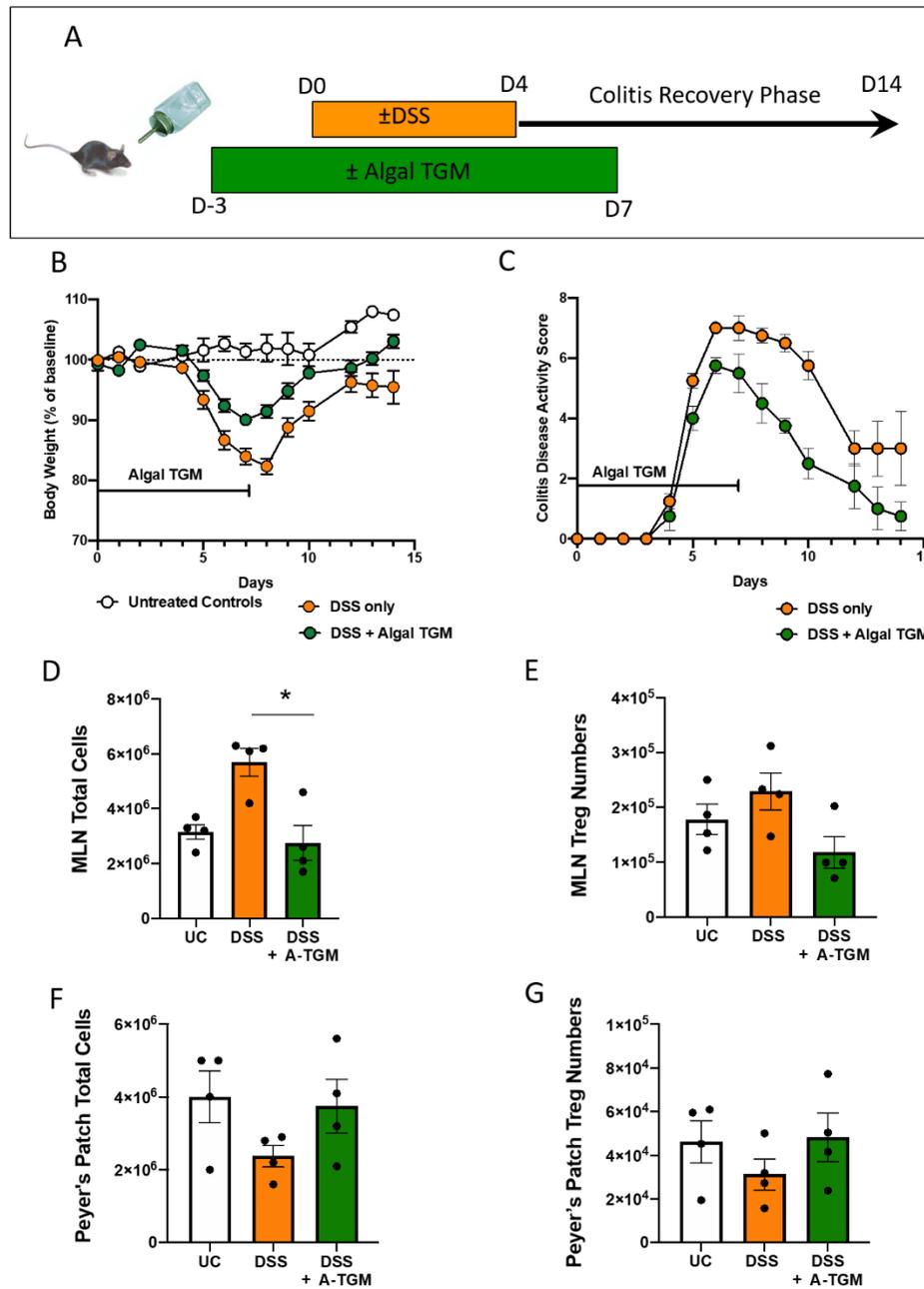


Figure 1.5 Amelioration of in vivo DSS mouse colitis with orally administered algal TGM1. **A.** Experimental design for DSS mouse feeding trial. DSS was fed to mice in the drinking water for 4 days with additional algal TGM1, followed by 3 days of standard drinking water with algal TGM1. The mice were allowed to recover for a further 7 days with standard drinking water only. **B, C.** Course of colitis in mice receiving DSS in drinking water, measured by (B) Weight change, and (C) Disease Activity Scores, in mice fed with algal TGM1, and controls receiving DSS alone. **D-G.** Total cells and Foxp3⁺ Treg numbers in MLNs (D, E) and Peyer's Patches (F, G) counted by flow cytometry. Bars represent means and standard errors, with data from individual mice shown as solid symbols. Data is from a single experiment.

To investigate anti-inflammatory effects of algal TGM1 in an *in vivo* mouse model of immunopathology, we used dextran sodium sulfate (DSS), a chemical to induce mucosal epithelial damage and disrupt mucosal barrier function (Chassaing et al., 2014). A percentage of 2 % DSS (weight/volume) was added to the drinking water, resulting in an acute colitis, and the water for the algal TGM group contained 100 $\mu\text{g}\cdot\text{mL}^{-1}$ TGM1 in addition to DSS. Mice were then allowed to drink freely from the DSS water for a 4 day period followed by a recovery period of 3 days on either standard drinking water (control group) or drinking water plus 100 $\mu\text{g}\cdot\text{mL}^{-1}$ TGM1 (algal TGM1 group), and an additional 7 days on standard drinking water to allow for mucosal repair (**Figure 1.5 A**). Control mice were given the equivalent volume of wild-type algal proteins or diluent (1 x PBS/ 2 % glycerol) in place of algal TGM1. Each mouse was weighed and scored daily to determine if algal TGM1 mitigated the weight loss effect of DSS.

All DSS mice started losing weight after administration of DSS, however mice administered algal TGM1 were protected from dramatic weight loss during DSS treatment and returned to their weight levels as measured before DSS treatment (**Figure 1.5 B**). Compared to mice given DSS alone, the group also receiving wild-type algal proteins showed no protection, and for clarity data from this group are not shown. In the absence of algal TGM1, mice suffered a 20 % weight loss 3 - 4 days after DSS administration while the weight loss of DSS + algal TGM1 group was only 10 % in the same time window. In addition, mice not receiving TGM1 could not fully recover their weight loss, and continued to exhibit a 10 % deficit after switching back to normal drinking water (**Figure 1.5 B**). In parallel with attenuated weight loss, algal TGM1 also ameliorated disease symptoms accompanying DSS administration, reaching a lower severity and returning closer to baseline more quickly (**Figure 1.5 C**). Taken together, these results indicate that algal TGM1 is effective at protecting mice from DSS-induced intestinal injury and weight loss.

DSS-induced intestinal inflammation is known to provoke lymphadenopathy with significant increases in the size, weight, and cellular content of the mesenteric lymph nodes (MLNs) (Rehal et al., 2018). Consistent with this previous work, we also found that total MLN cell numbers were greatly increased in mice treated with DSS, but remained at normal levels in the algal TGM recipients (**Figure 1.5 D**). MLN Treg numbers were not higher in TGM-treated mice (**Figure 1.5 E**) which may reflect more intense inflammation in the DSS alone animals. Interestingly, DSS treatment reduced Peyer's patch numbers (**Figure 1.5 F, G**) although not in those mice receiving TGM, suggesting a protective effect. We also investigated stimulation of ROR γ t+ T cells, which are associated with gut inflammation and increases permeability of the epithelial barrier. There was an increase in all DSS-treated mice seen in the MLN (**Supplementary Figure 1.5 B**) but not the Peyer's patches (**Supplementary Figure 1.5 C**).

4. Discussion

Immuno-modulatory parasites are increasingly recognized as treasure chests of novel, biologically active products with therapeutic potential, particularly for inflammatory disorders which are increasingly prevalent in industrialized societies (Johnston et al., 2009; Kahl et al., 2018; Maizels et al., 2018). Among the most severe of these “diseases of modernity” are inflammatory bowel diseases, for which no curative therapy is yet available (Kaser et al., 2010). The discovery of the immune suppressive protein, TGM1, which mimics TGF- β , from a helminth parasite which itself can suppress intestinal inflammation (Leung et al., 2012), led us to test the ability of recombinant TGM1 to inhibit IBD in the well-characterized mouse model of DSS. Because the parasite, *H. polygyrus*, resides in the intestinal lumen, we reasoned that oral delivery would replicate its natural pathway of host modulation; and to test this hypothesis, we produced quantities of the TGM1 protein in the alga *C. reinhardtii*.

Expression of recombinant proteins in algae has been reported to be a cost-effective method to produce protein biologics, including human therapeutics such as antibodies, growth hormones, cytokines, and vaccines (Gimpel et al., 2015; Gregory et al., 2013; Patra et al., 2015; Rasala et al., 2012; Ravi et al., 2018; Tran et al., 2009). Algae have also been shown to be an effective vehicle to deliver vaccines (Gregory et al., 2013) and recombinant therapeutic proteins (Barrera et al., 2015) to the gut, as *C. reinhardtii* has been granted GRAS status by the FDA and is safe to consume (Fields et al., 2020). We therefore tested expression of TGM1 in this expression system and found that it produced sufficient quantities of soluble, mature recombinant TGM1 protein. We then used two established assays for TGF- β function which confirmed fully functional properties for the recombinant TGM1, including the induction of the Foxp3⁺ phenotype of immunosuppressive regulatory T cells (Tregs). We noted that algal TGM1 required a higher concentration for equivalent effect to TGM1 produced in mammalian cells, but that is nevertheless reached a greater maximum signal than did TGF- β itself.

To evaluate the safety and physiological effect of algal TGM1 at steady state, it was administered to mice in the absence of any other intervention. No adverse effects were found, and there was a generalized enhancement of immune cell numbers, including Tregs, which was most marked in the Peyer’s patches, the localized immune structures in the small intestinal wall that serve as the primary interface between luminal contents and the mucosal immune system.

We then tested TGM1 in the DSS induced model of colitis, in which mice exhibit immunopathogenic phenotypes including weight loss, diarrhea and rectal bleeding, resembling clinical symptoms of human disease

(Okayasu et al., 1990; Wirtz et al., 2007). Along with the destruction of mucosal layer by DSS, commensal bacteria can cross the impaired epithelial barrier, stimulating the intestinal immune system to secrete pro-inflammatory cytokines and triggering inflammation. One corollary of intestinal inflammation is enlargement of the mesenteric lymph nodes which we observed were greatly increased in DSS-treated mice but not in those also receiving algal-expressed TGM orally. This suggests that TGM1 is able to reduce the inflammatory response of immune cells in this model.

Although TGM1 given to mice at steady-state tended to increase Treg numbers, a different outcome was observed in the DSS colitis model. One possibility is that inflammation itself elicits an influx of Tregs, such that at certain times they may be more abundant than in non-diseased cases; in IBD patients for example there are fewer circulating Tregs, but more in inflamed mucosal tissues (Maul et al., 2005; Wang et al., 2011). An alternative explanation is that the disease suppresses the *de novo* (TGF- β -dependent) induction of Tregs, for example through increased inflammatory cytokine production and/or loss of dendritic cell subsets able to drive Treg expansion (Boschetti et al., 2017). It will be important in future studies of TGM1 modulation to tease out its effects on dendritic cells which play critical roles in driving and balancing intestinal inflammation (Hart et al., 2005).

Over recent years, a number of potential new targets for IBD therapies have been explored. For example, DCs and macrophages in IBD patients produce a higher level of activated IL-1; unfortunately however, antibody against IL-1 has no benefit for IBD treatment (Kojouharoff et al., 1997). Similarly, a monoclonal antibody targeting IFN- γ had no effect on Crohn's disease patients (Reinisch et al., 2006). In the case of UC, patients express a more Type 2 profile (Fuss et al., 1996; Heller et al., 2005; Neurath, 2014), but anti-IL-13 antibody tralokinumab was not reported to have shown great clinical activity in patients with UC (Danese et al., 2015). For patients with either Crohn's disease or UC, Th17 cells are abundant in inflamed intestine terminal ileum and produce IL17A, IL17F, IL26 and IFN γ . Nevertheless, treatment of secukinumab, an antibody to IL17A, was ineffective for Crohn's disease and had severe side effects (Hueber et al., 2012). The most effective strategy to target pro-inflammatory cytokine is antibody against TNF which unfortunately is not effective for patients with few immune cells expressing membrane-bound TNF (Wong et al., 2008).

In contrast, treatment with recombinant anti-inflammatory cytokines has not been widely reported in IBD patients. Here we show for the first time that an algal produced TGF- β mimic, TGM1, acts in a similar manner to mammalian expressed TGM1 and limits disease intensity and lymph node inflammation when given orally in mouse

model of colitis. In the case of TGF- β , caution has been expressed as the mammalian cytokine is associated with Th17 differentiation when IL-6 is present in vitro (Veldhoen et al., 2006). However, recently TGF- β has been observed to be associated with a non-inflammatory Th17 subset (Lee et al., 2020), and our own recent work demonstrates that TGM1 is effective in inducing Foxp3 expression in human Th17 memory cells (Cook et al., 2021) In addition, with respect to concerns that TGF β promotes tissue fibrosis, TGM1 has been shown to be significantly less fibrogenic than mammalian TGF- β (Johnston et al., 2017).

In conclusion, among a number of potential therapeutic cytokine therapy strategies, recombinant algal TGM1 offers several advantages including the oral mode of delivery, allowing direct action on inflamed mucosal surfaces while minimizing any systemic off-target effects. It is also very economical, compared to the cost of anti-TNF therapy and limited efficiency of other anti-inflammatory cytokines. It may also be beneficial to identify if recombinant algal TGM1 can be combined with other therapeutics, perhaps also expressed in algae, or antibodies against other immunomodulatory cytokines (for example TNF) which may further increase their therapeutic efficacy for patients with diverse immunopathological backgrounds.

Supplementary Material

Supplementary Table 1. Protein yields of secreted algal material. Total protein amount of each sample was quantified by Pierce™ BCA Protein Assay Kit at 562nm and standardized with 250ug/ml, 500ug/ml, 750ug/ml and 1000ug/ml bovine serum albumin. Algal TGM1 concentration in precipitated supernatant were quantified by western blot against 150ng Recombinant Posi-Tag Epitope Tag Protein containing the epitope FLAG tag. Protein band pixel intensities were analyzed by Image Studio Lite.

	Total protein concentration	Volume	sTGM concentration in total protein samples
Proteins purified from WT supernatant	10510ng/ul	250mL	0
Protein purified from TGM1 transgenic strain supernatant	11850ng/ul	220mL	226ng/ul

Supplementary Table 2. Glycosylation site analysis of algal TGM1. A tryptic digest of algal TGM1 was analyzed by mass spectrometry for peptide identification and glycosylation site analysis. Data were analyzed using BioPharma Finder 4.0 software. Glycopeptides with a confidence value of > 80 %, and delta mass of 5 ppm were accepted. Glycosylated residues are highlighted by underlining of asparagine (N) or threonine (T). Cysteine residues alkylated as part of the sample processing are indicated by italics. Of the glycopeptides identified, a nonglycosylated counterpart peptide was only identified for N118.

<i>N or O glycosylation</i>	<i>Glycosylation site</i>	<i>Identification</i>	<i>Peptide sequence</i>	<i>Glycan</i>
N	N118	C112-K127	CSPLPT <u>N</u> DTVSF E YLK	M5
N	N118	C112-K127	CSPLPT <u>N</u> DTVSF E YLK	M5
N	N118	C112-K127	CSPLPT <u>N</u> DTVSF E YLK	M4
N	N118	C112-K127	CSPLPT <u>N</u> DTVSF E YLK	M4
N	N118	R111-K127	RCSPLPT <u>N</u> DTVSF E YLK	M7
N	N118	C112-K127	CSPLPT <u>N</u> DTVSF E YLK	M5
N	N118	C112-K127	CSPLPT <u>N</u> DTVSF E YLK	M5
N	N118	C112-K127	CSPLPT <u>N</u> DTVSF E YLK	None
N	N118	C112-K127	CSPLPT <u>N</u> DTVSF E YLK	None
N	N118	C112-K127	CSPLPT <u>N</u> DTVSF E YLK	None
N	N137	F136-K147	F <u>N</u> ITVHPDASGK	A2G2
N	N137	F136-K147	F <u>N</u> ITVHPDASGK	A2G2
N	~N137	L115-K147	LPTNDTVSF E YLKATVNP <u>G</u> IIF <u>N</u> ITVHPDASG K	A2G0F
N	N285	P281-T296	PLEAN <u>E</u> SVHYEYFTMT	A1G1F
N	N285	P281-T296	PLEAN <u>E</u> SVHYEYFTMT	A1G1F
N	N285	P281-T296	PLEAN <u>E</u> SVHYEYFTMT	A1G0F
N	N285	P281-T296	PLEAN <u>E</u> SVHYEYFTMT	A1G0F

N	N285	P281-T294	PLEAN<u>E</u>SVHYEYFTMT	A2G1B
N	N297	S287-K302	SVHYEYFTMTN<u>E</u>TDKK	A2G0B
N	N297	S287-K302	SVHYEYFTMTN<u>E</u>TDKK	A2G0B
N	N297	V288-K302	VHYEYFTMTN<u>E</u>TDKK	A2G1F
N	~N297	K280-K311	KPLEANESVHYEYFTMTN<u>E</u>TDKKGPPAKV GK	A2G0F
N	~N297	K280-K311	KPLEANESVHYEYFTMTN<u>E</u>TDKKGPPAKV GK	A2G0F
O	~T187	N160-C188	NFPTDSNVQGHIGMCYNAEWQFSSTP<u>T</u>C	GalNAc -3G
O	~T187	N160-C188	NFPTDSNVQGHIGMCYNAEWQFSSTP<u>T</u>C	GalNAc -3G

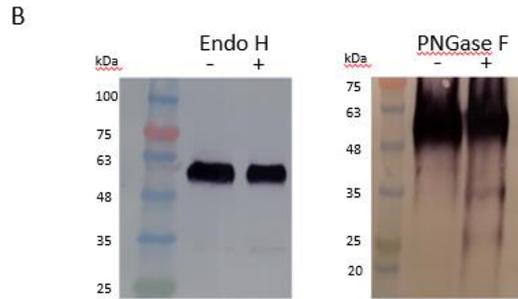
Supplementary Figure Legends



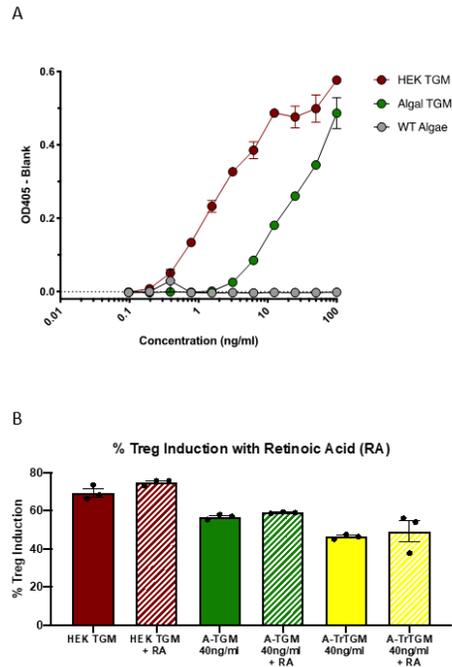
Supplementary Figure 1.1 Nucleotide sequence of TGM1 as codon optimized for *C. reinhardtii* nuclear transcription. The *C. reinhardtii* ars1 Signal Peptide secretion sequence (90bp, 1-90, in orange) immediately precedes the 5-domain TGM1 mature sequence, labeled in blue, with the darker color for Domains 1-3 (741 bp, 91-831), and lighter blue for Domains 4 and 5 (480 bp, 832-1311) which are absent from the TrTGM1 construct. The 8-aa FLAG tag with a stop codon (27bp, 1312-1338), immediately after the TGM1 sequence, is labeled in mauve.

A

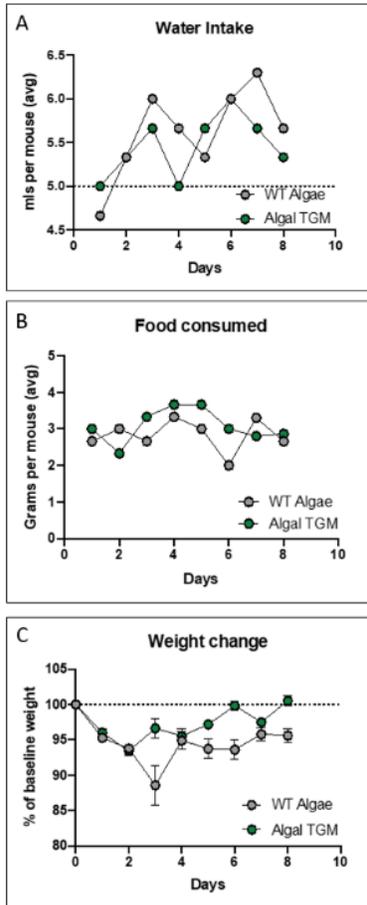
1	DDSGGMPFS	EAATYLYVAE	GFENIIPAQ	IDNSGMFDY	THVERFUEGL	NGEDTTGWFF	GIGLASQWNY	YESVQECDDR
81	EGSPLPTNDG	VSFEYLEAVY	NFGIIPNITV	HPDASGEVPE	LTYILEICEN	PFPGDNVQGN	IIGMEYNAEN	QFSSTPTCPA
161	SGCFPLPDDG	IVFYEYGYA	GDHNTVQPVV	TLDSSGNTPS	PTHAREECPA	LSQKADPGEF	VAICYRSOTT	QESHMETYLN
241	IGEGPDPFCK	PLEANESVHY	EYFTMTNETD	KEGPPAPVVG	ESGEYVENTC	VKVCSEWPFY	TCSTGGPIFG	ECIGATWNPY
321	ALMECINARQ	CSSDOLFDEL	GFEFVIVREG	EGSSSYLDGF	ARFYATQSKV	IARQGGKTVR	LECSNGEWHE	PGTETVHRQT
401	KDGIRTL							



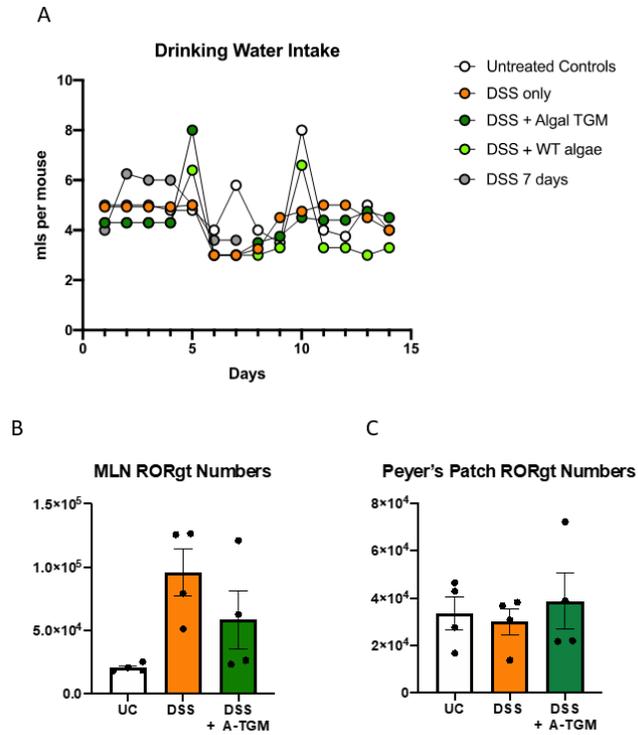
Supplementary Figure 1.2 Expression and Purification of Algal TGM1. **A.** Mass spectrometry coverage of FLAG-purified algal TGM1. Amino acid sequences highlighted in red are peptides identified from mass spectrometry; blue amino acids are cysteine residues that are alkylated post reduction; green amino acid sequences are the potential glycosylation sites. The lysine (K) and arginine (R) residues are underlined as potential and identified sites of tryptic cleavage. **B.** Anti-FLAG Western blot of FLAG-purified algal TGM1 digested with (+) and without (-) Endo H and with (+) and without (-) PNGase F endoglycosidases. Marker protein molecular weights are indicated.



Supplementary Figure 1.3 In vitro assays for TGF- β activity. **A.** MFB-F11 assay including WT algae protein controls, HEK-expressed and algal-expressed TGM1. **B.** Foxp3 induction in CD4⁺ T cells from a Foxp3-GFP transgenic reporter mouse by HEK-expressed and algal-expressed TGM1 and TrTGM1 in the presence or absence of IL-2 and retinoic acid (RA).



Supplementary Figure 1.4 Physiological responses to orally administered algal TGM1. Mice given WT algal proteins, or algal-expressed TGM1 as shown in Figure 3A were monitored over 8 days for water intake (A), food consumed (B), and weight change (C).



Supplementary Figure 1.5 TGM1 administration during DSS-induced colitis. **A.** Water intake over 14 days in each experimental group. Expression of ROR γ t in Tregs from the MLN (**B**) and Peyer's Patches (**C**). UC = untreated control animals.

Acknowledgements

Chapter 1, in part is currently being prepared for submission for publication of the material. Bijie Ren, Danielle Smyth, Madeleine White, Francis Fields, Rick Maizels, and Stephen Mayfield. Oral delivery of a functional algal expressed TGF- β mimic halts colitis in a murine DSS model. The dissertation author was the primary investigator and author of this paper.

The authors gratefully acknowledge the Flow Core Facility and the Wolfson Research Facility, University of Glasgow for their support and assistance. This work was supported by the Kenneth Rainin Foundation through Synergy and Innovator Grants (Refs 2015-964 and 2016-3067), the U.S. Department of Energy (DE-EE0008246), the Wellcome Trust through Investigator Awards to RMM (Ref 106122 and 219530), the Wellcome Trust core-funded Wellcome Centre for Integrative Parasitology (Ref: 104111), and by the Medical Research Council Confidence-in-Concept scheme. The authors declare no competing financial interests.

References

- Bardor, M., Faveeuw, C., Fitchette, A.C., Gilbert, D., Galas, L., Trottein, F., Faye, L., Lerouge, P., (2003) Immunoreactivity in mammals of two typical plant glyco-epitopes, core $\alpha(1,3)$ -fucose and core xylose. *Glycobiology* 13, 427-434.
- Barrera, D.J., Rosenberg, J.N., Chiu, J.G., Chang, Y.N., Debatis, M., Ngoi, S.M., Chang, J.T., Shoemaker, C.B., Oyler, G.A., Mayfield, S.P., (2015) Algal chloroplast produced camelid VH H antitoxins are capable of neutralizing botulinum neurotoxin. *Plant Biotechnol J* 13, 117-124.
- Baumgart, D.C., Sandborn, W.J., (2012) Crohn's disease. *Lancet* 380, 1590-1605.
- Boschetti, G., Kanjarawi, R., Bardel, E., Collardeau-Frachon, S., Duclaux-Loras, R., Moro-Sibilot, L., Almeras, T., Flourie, B., Nancey, S., Kaiserlian, D., (2017) Gut inflammation in mice triggers proliferation and function of mucosal Foxp3+ regulatory T
- Braat, H., Rottiers, P., Hommes, D.W., Huyghebaert, N., Remaut, E., Remon, J.P., van Deventer, S.J., Neiryneck, S., Peppelenbosch, M.P., Steidler, L., (2006) A phase I trial with transgenic bacteria expressing interleukin-10 in Crohn's disease. *Clinical gastroenterology and hepatology : the official clinical practice journal of the American Gastroenterological Association* 4, 754-759.
- Burisch, J., Kiudelis, G., Kupcinkas, L., Kievit, H.A.L., Andersen, K.W., Andersen, V., Salupere, R., Pedersen, N., Kjeldsen, J., D'Inca, R., Valpiani, D., Schwartz, D., Odes, S., Olsen, J., Nielsen, K.R., Vegh, Z., Lakatos, P.L., Toca, A., Turcan, S., Katsanos, K.H., Christodoulou, D.K., Fumery, M., Gower-Rousseau, C., Zammit, S.C., Ellul, P., Eriksson, C., Halfvarson, J., Magro, F.J., Duricova, D., Bortlik, M., Fernandez, A., Hernandez, V., Myers, S., Sebastian, S., Oksanen, P., Collin, P., Goldis, A., Misra, R., Arebi, N., Kaimakliotis, I.P., Nikuina, I., Belousova, E., Brinar, M., Cukovic-Cavka, S., Langholz, E., Munkholm, P., Epi, I.B.D.g., (2019) Natural disease course of Crohn's disease during the first 5 years after diagnosis in a European population-based inception cohort: an Epi-IBD study. *Gut* 68, 423-433.
- Buruiana, F.E., Sola, I., Alonso-Coello, P., (2010) Recombinant human interleukin 10 for induction of remission in Crohn's disease. *Cochrane Database Syst Rev*, CD005109.
- Chassaing, B., Aitken, J.D., Malleshappa, M., Vijay-Kumar, M., (2014) Dextran sulfate sodium (DSS)-induced colitis in mice. *Curr Protoc Immunol* 104, 15 25 11-15 25 14.
- Clay, S.L., Bravo-Blas, A., Wall, D.M., MacLeod, M.K.L., Milling, S.W.F., (2020) Regulatory T cells control the dynamic and site-specific polarization of total CD4 T cells following *Salmonella* infection. *Mucosal Immunol* 13, 946-957.
- Cook, L., Reid, K., Hakkinen, E., de Bie, B., Tanaka, S., Smyth, D.J., White, M., Wong, M.Q., Huang, Q., Gillies, J.K., Ziegler, S.F., Maizels, R.M., Levings, M., (2021) Induction of stable human FOXP3⁺ T regs by a parasite-derived TGF β mimic. *Immunology and Cell Biology in press*.
- Da Silva, C., Wagner, C., Bonnardel, J., Gorvel, J.P., Lelouard, H., (2017) The Peyer's patch mononuclear phagocyte system at steady state and during infection. *Front Immunol* 8, 1254.
- Danese, S., Fiocchi, C., (2011) Ulcerative colitis. *N Engl J Med* 365, 1713-1725.
- Danese, S., Rudzinski, J., Brandt, W., Dupas, J.L., Peyrin-Biroulet, L., Bouhnik, Y., Kleczkowski, D., Uebel, P., Lukas, M., Knutsson, M., Erlandsson, F., Hansen, M.B., Keshav, S., (2015) Tralokinumab for moderate-to-severe UC: a randomised, double-blind, placebo-controlled, phase IIa study. *Gut* 64, 243-249.
- David, C.J., Massague, J., (2018) Contextual determinants of TGF β action in development, immunity and cancer. *Nat Rev Mol Cell Biol* 19, 419-435.

- de Souza, H.S., Fiocchi, C., (2016) Immunopathogenesis of IBD: current state of the art. *Nat Rev Gastroenterol Hepatol* 13, 13-27.
- Dove, A., (2002) Uncorking the biomanufacturing bottleneck. *Nat Biotechnol* 20, 777-779.
- Fields, F.J., Lejzerowicz, F., Schroeder, D., Ngoi, S.M., Tran, M., McDonald, D., Jiang, L., Chang, J., T., Knight, R., Mayfield, S., (2020) Effects of the microalgae *Chlamydomonas* on gastrointestinal health. *J Functional Foods* 65, 103738.
- Fontenot, J.D., Rasmussen, J.P., Williams, L.M., Dooley, J.L., Farr, A.G., Rudensky, A.Y., (2005) Regulatory T cell lineage specification by the forkhead transcription factor Foxp3. *Immunity* 22, 329-341.
- Fuss, I.J., Neurath, M., Boirivant, M., Klein, J.S., de la Motte, C., Strong, S.A., Fiocchi, C., Strober, W., (1996) Disparate CD4⁺ lamina propria (LP) lymphokine secretion profiles in inflammatory bowel disease. Crohn's disease LP cells manifest increased secretion of IFN-gamma, whereas ulcerative colitis LP cells manifest increased secretion of IL-5. *J Immunol* 157, 1261-1270.
- Gimpel, J.A., Hyun, J.S., Schoepp, N.G., Mayfield, S.P., (2015) Production of recombinant proteins in microalgae at pilot greenhouse scale. *Biotechnol Bioeng* 112, 339-345.
- Gorman, D.S., Levine, R.P., (1965) Cytochrome f and plastocyanin: their sequence in the photosynthetic electron transport chain of *Chlamydomonas reinhardtii*. *Proc Natl Acad Sci U S A* 54, 1665-1669.
- Gregory, J.A., Topol, A.B., Doerner, D.Z., Mayfield, S., (2013) Alga-produced cholera toxin-Pfs25 fusion proteins as oral vaccines. *Appl Environ Microbiol* 79, 3917-3925.
- Hamady, Z.Z., Scott, N., Farrar, M.D., Wadhwa, M., Dilger, P., Whitehead, T.R., Thorpe, R., Holland, K.T., Lodge, J.P., Carding, S.R., (2011) Treatment of colitis with a commensal gut bacterium engineered to secrete human TGF-beta1 under the control of dietary xylan 1. *Inflamm Bowel Dis* 17, 1925-1935.
- Hart, A.L., Al-Hassi, H.O., Rigby, R.J., Bell, S.J., Emmanuel, A.V., Knight, S.C., Kamm, M.A., Stagg, A.J., (2005) Characteristics of intestinal dendritic cells in inflammatory bowel diseases. *Gastroenterology* 129, 50-65.
- Heller, F., Florian, P., Bojarski, C., Richter, J., Christ, M., Hillenbrand, B., Mankertz, J., Gitter, A.H., Burgel, N., Fromm, M., Zeitz, M., Fuss, I., Strober, W., Schulzke, J.D., (2005) Interleukin-13 is the key effector Th2 cytokine in ulcerative colitis that affects epithelial tight junctions, apoptosis, and cell restitution. *Gastroenterology* 129, 550-564.
- Huber, S., Gagliani, N., Esplugues, E., O'Connor, W., Jr., Huber, F.J., Chaudhry, A., Kamanaka, M., Kobayashi, Y., Booth, C.J., Rudensky, A.Y., Roncarolo, M.G., Battaglia, M., Flavell, R.A., (2011) Th17 cells express interleukin-10 receptor and are controlled by Foxp3⁻ and Foxp3⁺ regulatory CD4⁺ T cells in an interleukin-10-dependent manner. *Immunity* 34, 554-565.
- Hueber, W., Sands, B.E., Lewitzky, S., Vandemeulebroecke, M., Reinisch, W., Higgins, P.D., Wehkamp, J., Feagan, B.G., Yao, M.D., Karczewski, M., Karczewski, J., Pezous, N., Bek, S., Bruin, G., Mellgard, B., Berger, C., Londei, M., Bertolino, A.P., Tougas, G., Travis, S.P., Secukinumab in Crohn's Disease Study, G., (2012) Secukinumab, a human anti-IL-17A monoclonal antibody, for moderate to severe Crohn's disease: unexpected results of a randomised, double-blind placebo-controlled trial. *Gut* 61, 1693-1700.
- Hunter, M.M., Wang, A., Hirota, C.L., McKay, D.M., (2005) Neutralizing anti-IL-10 antibody blocks the protective effect of tapeworm infection in a murine model of chemically induced colitis. *J Immunol* 174, 7368-7375.
- Ihara, S., Hirata, Y., Koike, K., (2017) TGF-β in inflammatory bowel disease: a key regulator of immune cells, epithelium, and the intestinal microbiota. *J Gastroenterol* 52, 777-787.

- Johnston, C.J.C., Smyth, D.J., Kodali, R.B., White, M.P.J., Harcus, Y., Filbey, K.J., Hewitson, J.P., Hinck, C.S., Ivens, A., Kemter, A.M., Kildemoes, A.O., Le Bihan, T., Soares, D.C., Anderton, S.M., Brenn, T., Wigmore, S.J., Woodcock, H., Chambers, R.C., Hinck, A.P., McSorley, H.J., Maizels, R.M., (2017) A structurally distinct TGF- β mimic from an intestinal helminth parasite potently induces regulatory T cells. . Nature Communications 8, 1741.
- Johnston, M.J.G., Macdonald, J.A., McKay, D.M., (2009) Parasitic helminths: a pharmacopeia of anti-inflammatory molecules. Parasitology 136, 125-147.
- Kahl, J., Brattig, N., Liebau, E., (2018) The untapped pharmacopeic potential of helminths. Trends Parasitol 34, 828-842.
- Kaser, A., Zeissig, S., Blumberg, R.S., (2010) Inflammatory bowel disease. Annual review of immunology 28, 573-621.
- Kojouharoff, G., Hans, W., Obermeier, F., Mannel, D.N., Andus, T., Scholmerich, J., Gross, V., Falk, W., (1997) Neutralization of tumour necrosis factor (TNF) but not of IL-1 reduces inflammation in chronic dextran sulphate sodium-induced colitis in mice. Clin Exp Immunol 107, 353-358.
- Konkel, J.E., Chen, W., (2011) Balancing acts: the role of TGF- β in the mucosal immune system. Trends Mol Med 17, 668-676.
- Lee, J.Y., Hall, J.A., Kroehling, L., Wu, L., Najar, T., Nguyen, H.H., Lin, W.Y., Yeung, S.T., Silva, H.M., Li, D., Hine, A., Loke, P., Hudesman, D., Martin, J.C., Kenigsberg, E., Merad, M., Khanna, K.M., Littman, D.R., (2020) Serum Amyloid A Proteins Induce Pathogenic Th17 Cells and Promote Inflammatory Disease. Cell 180, 79-91 e16.
- Leung, J., Hang, L., Blum, A., Setiawan, T., Stoyanoff, K., Weinstock, J., (2012) *Heligmosomoides polygyrus* abrogates antigen-specific gut injury in a murine model of inflammatory bowel disease. Inflamm Bowel Dis.
- Lucas, P.L., Mathieu-Rivet, E., Song, P.C.T., Oltmanns, A., Loutelier-Bourhis, C., Plasson, C., Afonso, C., Hippler, M., Lerouge, P., Mati-Baouche, N., Bardor, M., (2020) Multiple xylosyltransferases heterogeneously xylosylate protein N-linked glycans in *Chlamydomonas reinhardtii*. Plant J 102, 230-245.
- Maizels, R.M., Smits, H.H., McSorley, H.J., (2018) Modulation of host immunity by helminths: the expanding repertoire of parasite effector molecules. Immunity 49, 801-818.
- Mathieu-Rivet, E., Scholz, M., Arias, C., Dardelle, F., Schulze, S., Le Mauff, F., Teo, G., Hochmal, A.K., Blanco-Rivero, A., Loutelier-Bourhis, C., Kiefer-Meyer, M.C., Fufezan, C., Burel, C., Lerouge, P., Martinez, F., Bardor, M., Hippler, M., (2013) Exploring the N-glycosylation pathway in *Chlamydomonas reinhardtii* unravels novel complex structures. Mol Cell Proteomics 12, 3160-3183.
- Maul, J., Loddenkemper, C., Mundt, P., Berg, E., Giese, T., Stallmach, A., Zeitz, M., Duchmann, R., (2005) Peripheral and intestinal regulatory CD4⁺ CD25^(high) T cells in inflammatory bowel disease. Gastroenterology 128, 1868-1878.
- McLean, M.H., Andrews, C., Hanson, M.L., Baseler, W.A., Anver, M.R., Senkevitch, E., Staniszewska, A.K., Smith, C., Davies, L.C., Hixon, J., Li, W., Shen, W., Steidler, L., Durum, S.K., (2017) Interleukin-27 is a potential rescue therapy for acute severe colitis through interleukin-10-dependent, T-cell-independent attenuation of colonic mucosal innate immune responses. Inflamm Bowel Dis 23, 1983-1995.
- Monteleone, G., Neurath, M.F., Ardizzone, S., Di Sabatino, A., Fantini, M.C., Castiglione, F., Scribano, M.L., Armuzzi, A., Caprioli, F., Sturniolo, G.C., Rogai, F., Vecchi, M., Atreya, R., Bossa, F., Onali, S., Fichera, M., Corazza, G.R., Biancone, L., Savarino, V., Pica, R., Orlando, A., Pallone, F., (2015) Mongersen, an oral SMAD7 antisense oligonucleotide, and Crohn's disease. N Engl J Med 372, 1104-1113.

- Mosconi, I., Dubey, L.K., Volpe, B., Esser-von Bieren, J., Zaiss, M.M., Lebon, L., Massacand, J.C., Harris, N.L., (2015) Parasite proximity drives the expansion of regulatory T cells in Peyer's patches following intestinal helminth infection. *Infect Immun* 83, 3657-3665.
- Neurath, M.F., (2014) Cytokines in inflammatory bowel disease. *Nat Rev Immunol* 14, 329-342.
- Neurath, M.F., Weigmann, B., Finotto, S., Glickman, J., Nieuwenhuis, E., Iijima, H., Mizoguchi, A., Mizoguchi, E., Mudter, J., Galle, P.R., Bhan, A., Autschbach, F., Sullivan, B.M., Szabo, S.J., Glimcher, L.H., Blumberg, R.S., (2002) The transcription factor T-bet regulates mucosal T cell activation in experimental colitis and Crohn's disease. *J Exp Med* 195, 1129-1143.
- Okayasu, I., Hatakeyama, S., Yamada, M., Ohkusa, T., Inagaki, Y., Nakaya, R., (1990) A novel method in the induction of reliable experimental acute and chronic ulcerative colitis in mice. *Gastroenterology* 98, 694-702.
- Oltmanns, A., Hoepfner, L., Scholz, M., Zinzius, K., Schulze, S., Hippler, M., (2019) Novel insights into N-glycan fucosylation and core xylosylation in *C. reinhardtii*. *Front Plant Sci* 10, 1686.
- Patra, K.P., Li, F., Carter, D., Gregory, J.A., Baga, S., Reed, S.G., Mayfield, S.P., Vinetz, J.M., (2015) Alga-produced malaria transmission-blocking vaccine candidate Pfs25 formulated with a human use-compatible potent adjuvant induces high-affinity antibodies that block *Plasmodium falciparum* infection of mosquitoes. *Infect Immun* 83, 1799-1808.
- Rasala, B.A., Barrera, D.J., Ng, J., Plucinak, T.M., Rosenberg, J.N., Weeks, D.P., Oyler, G.A., Peterson, T.C., Haerizadeh, F., Mayfield, S.P., (2013) Expanding the spectral palette of fluorescent proteins for the green microalga *Chlamydomonas reinhardtii*. *Plant J* 74, 545-556.
- Rasala, B.A., Lee, P.A., Shen, Z., Briggs, S.P., Mendez, M., Mayfield, S.P., (2012) Robust expression and secretion of Xylanase I in *Chlamydomonas reinhardtii* by fusion to a selection gene and processing with the FMDV 2A peptide. *PLoS One* 7, e43349.
- Ravi, A., Guo, S., Rasala, B., Tran, M., Mayfield, S., Nikolov, Z.L., (2018) Separation options for phosphorylated osteopontin from transgenic microalgae *Chlamydomonas reinhardtii*. *Int J Mol Sci* 19.
- Rehal, S., Stephens, M., Roizes, S., Liao, S., von der Weid, P.Y., (2018) Acute small intestinal inflammation results in persistent lymphatic alterations. *Am J Physiol Gastrointest Liver Physiol* 314, G408-G417.
- Reinisch, W., Hommes, D.W., Van Assche, G., Colombel, J.F., Gendre, J.P., Oldenburg, B., Teml, A., Geboes, K., Ding, H., Zhang, L., Tang, M., Cheng, M., van Deventer, S.J., Rutgeerts, P., Pearce, T., (2006) A dose escalating, placebo controlled, double blind, single dose and multidose, safety and tolerability study of fontolizumab, a humanised anti-interferon gamma antibody, in patients with moderate to severe Crohn's disease. *Gut* 55, 1138-1144.
- Rupprecht, K.R., Nair, R.K., Harwick, L.C., Grote, J., Beligere, G.S., Rege, S.D., Chen, Y.Y., Lin, Z., Fishpaugh, J.R., (2010) Development of a dot-blot assay for screening monoclonal antibodies to low-molecular-mass drugs. *Anal Biochem* 407, 160-164.
- Saraiva, M., Vieira, P., O'Garra, A., (2020) Biology and therapeutic potential of interleukin-10. *J Exp Med* 217.
- Schreiber, S., Fedorak, R.N., Nielsen, O.H., Wild, G., Williams, C.N., Nikolaus, S., Jacyna, M., Lashner, B.A., Gangl, A., Rutgeerts, P., Isaacs, K., van Deventer, S.J., Koningsberger, J.C., Cohard, M., LeBeaut, A., Hanauer, S.B., (2000) Safety and efficacy of recombinant human interleukin 10 in chronic active Crohn's disease. Crohn's Disease IL-10 Cooperative Study Group. *Gastroenterology* 119, 1461-1472.
- Smith, P., Mangan, N.E., Walsh, C.M., Fallon, R.E., McKenzie, A.N.J., van Rooijen, N., Fallon, P.G., (2007) Infection with a helminth parasite prevents experimental colitis via a macrophage-mediated mechanism. *J Immunol* 178, 4557-4566.

- Smyth, D.J., Harcus, Y., White, M.P.J., Gregory, W.F., Nahler, J., Stephens, I., Toke-Bjølgerud, E., Hewitson, J.P., Ivens, A., McSorley, H.J., Maizels, R.M., (2018) TGF- β mimic proteins form an extended gene family in the murine parasite *Heligmosomoides polygyrus*. *Int J Parasitol* 48, 379-385.
- Steidler, L., Hans, W., Schotte, L., Neiryneck, S., Obermeier, F., Falk, W., Fiers, W., Remaut, E., (2000) Treatment of murine colitis by *Lactococcus lactis* secreting interleukin-10. *Science* 289, 1352-1355.
- Sugimoto, K., Ogawa, A., Mizoguchi, E., Shimomura, Y., Andoh, A., Bhan, A.K., Blumberg, R.S., Xavier, R.J., Mizoguchi, A., (2008) IL-22 ameliorates intestinal inflammation in a mouse model of ulcerative colitis. *The Journal of clinical investigation* 118, 534-544.
- Summers, R.W., Elliott, D.E., Urban, J.F., Jr., Thompson, R.A., Weinstock, J.V., (2005) *Trichuris suis* therapy for active ulcerative colitis: a randomized controlled trial. *Gastroenterology* 128, 825-832.
- Tesseur, I., Zou, K., Berber, E., Zhang, H., Wyss-Coray, T., (2006) Highly sensitive and specific bioassay for measuring bioactive TGF- β . *BMC Cell Biol* 7, 15.
- Tran, M., Van, C., Barrera, D.J., Pettersson, P.L., Peinado, C.D., Bui, J., Mayfield, S.P., (2013) Production of unique immunotoxin cancer therapeutics in algal chloroplasts. *Proc Natl Acad Sci U S A* 110, E15-22.
- Tran, M., Zhou, B., Pettersson, P.L., Gonzalez, M.J., Mayfield, S.P., (2009) Synthesis and assembly of a full-length human monoclonal antibody in algal chloroplasts. *Biotechnol Bioeng* 104, 663-673.
- Varyani, F., Fleming, J.O., Maizels, R.M., (2017) Helminths in the gastrointestinal tract as modulators of immunity and pathology. *Am J Physiol Gastrointest Liver Physiol* 312, G537-G549.
- Veldhoen, M., Hocking, R.J., Atkins, C.J., Locksley, R.M., Stockinger, B., (2006) TGF β in the context of an inflammatory cytokine milieu supports de novo differentiation of IL-17-producing T cells. *Immunity* 24, 179-189.
- Wang, Y., Liu, X.P., Zhao, Z.B., Chen, J.H., Yu, C.G., (2011) Expression of CD4⁺ forkhead box P3 (FOXP3)⁺ regulatory T cells in inflammatory bowel disease. *J Dig Dis* 12, 286-294.
- Webster, H.C., Andrusaitė, A.T., Shergold, A.L., Milling, S.W.F., Perona-Wright, G., (2020) Isolation and functional characterisation of lamina propria leukocytes from helminth-infected, murine small intestine. *J Immunol Methods* 477, 112702.
- Weinstock, J.V., Elliott, D.E., (2013) Translatability of helminth therapy in inflammatory bowel diseases. *Int J Parasitol* 43, 245-251.
- White, M.P.J., Smyth, D.J., Cook, L., Ziegler, S.F., Levings, M., Maizels, R.M., (2021) The parasite cytokine mimic *Hp*-TGM potently replicates the regulatory effects of TGF- β on murine CD4⁺ T cells. *Immunology and Cell Biology* (Submitted).
- Wirtz, S., Neufert, C., Weigmann, B., Neurath, M.F., (2007) Chemically induced mouse models of intestinal inflammation. *Nat Protoc* 2, 541-546.
- Wong, M., Ziring, D., Korin, Y., Desai, S., Kim, S., Lin, J., Gjertson, D., Braun, J., Reed, E., Singh, R.R., (2008) TNF α blockade in human diseases: mechanisms and future directions. *Clin Immunol* 126, 121-136.

**Chapter 2: Discovery of two master regulators for lipid accumulation *Chlamydomonas reinhardtii* through
Yeast Two-Hybrid sequencing**

Discovery of two master regulators for lipid accumulation in *Chlamydomonas reinhardtii* through Yeast Two-Hybrid sequencing

Bijie Ren, Song Cao, Stephen Mayfield

Abstract

Chlamydomonas reinhardtii has become an essential model organism in the field of synthetic biology. As a single-cell green alga, it is often genetically engineered to produce valuable biopharmaceuticals and biofuels. Further exploration of *Chlamydomonas reinhardtii*'s potential in both enhanced production of existing compounds, and de-novo synthesis of novel molecules, requires a systematic investigation into the organism's gene regulation mechanism, which is currently understudied. Here we applied a recently published CrY2H-seq method to uncover the transcription factor (TF) interactome of *Chlamydomonas reinhardtii*, which lies in the center of its transcriptional regulation and bridges various biological processes. The result identifies two master regulators involved in lipid accumulation by suppressing different downstream pathways separately when cells suffered from nitrogen deprivation, which provides insights into Nitrogen-allocation and Carbon-allocation under nitrogen starvation for algae cells. By analyzing conditional transcriptomics data in the context of TF interactions and more TFs amplified from algae, we could illustrate how transcription factors respond to different environmental stimuli to coordinate different gene regulatory pathways in *Chlamydomonas reinhardtii*.

Introduction

As an increasingly important organism in synthetic biology, *Chlamydomonas reinhardtii* has been applied in the manufacture of numerous biomolecules including biodegradable material and mammalian therapeutic proteins (Tran et al., 2009; Rasala et al., 2012; Gregory et al., 2013; Ravi et al., 2018; Berdnt et al., 2021). However, the dynamics of its underlying gene regulatory networks and pathways are poorly characterized, while a more sophisticated engineering of the organism's cellular components will require a systematic study of the corresponding transcriptional regulatory network (Long et al, 2008). Here we sought to render a high level view of *Chlamydomonas reinhardtii*'s transcriptional regulation through investigating its conditional gene expression profiles and transcription factor (TF) interactome, which often provides useful information of how TF interactions bridge different biosynthetic

pathways under different conditions to fine-tune gene expression and adapt to a specific growth environment (Bemer et al, 2016).

Large scale protein-protein interaction is still a challenge for scientists to understand regulatory mechanism inside cell. Different methods have been developed in past decades to investigate such protein-protein interaction (Papachristou et al, 2018, Brückner et al, 2019). However, most of them cannot balance the cost, sensitivity, scale, high-throughput manner and false positive issue. In this manuscript, we used a new published CrY2Hseq (Cre reporter-mediated yeast two-hybrid coupled with next-generation sequencing) method to identify Transcription factor interaction for *Chlamydomonas reinhardtii* (Trigg et al, 2017). Integration of Cre-loxp system allows us to use next generation sequencing method to directly sequence all of the binary interaction pairs instead of doing iterative screening and tracking millions of colonies (Kühn et al 2002).

114 *Chlamydomonas reinhardtii* Transcription factors and regulators were amplified and used to build TF-AD and TF-BD library for CrY2H-seq. Novel interactions among Transcription factors were identified, and several experiments and analysis of all reported RNAseq data for *Chlamydomonas reinhardtii* were performed to validate such interactions. We identified one transcription factor cluster that may play a very important role to sense cAMP level, degrade ribosome, induce lipid and aromatic molecules accumulation under nitrogen starvation for the first time.

Results

Construction of *Chlamydomonas* Transcription factor AD and BD library

Several groups have already predicted *Chlamydomonas reinhardtii* transcription factors and regulators based on conserved eukaryotic transcription factor domain (Guo et al., 2007; Zhang et al., 2011; Jin et al., 2014). In our previous work, 92 *Chlamydomonas* TFs have been amplified from *Chlamydomonas reinhardtii* cDNA (Muff et al, 2015). Based on their prediction from PlantTFDB database (Tian et al, 2019), we designed primers for rest transcription factors. Then we analyzed 546 RNAseq datasets from 294 conditions to identify differential expression of Transcription factors (**Figure 2.1A**). Differential expression profile of TFs was used to find the optimal condition for maximal transcription of each TF. We incubated algae under different conditions, including constant light, constant dark, diurnal cycle, TAP medium with different nutrient depletion, and TAP medium with different nitrogen sources. mRNAs were extracted from algae under each condition and used to build cDNA library pool. Meanwhile, we also purchased cDNA library from ChlamyCollection center

(ChlamyCollection(<https://www.chlamycollection.org/products/cdna-libraries/>), which were used as templates to amplify rest *Chlamydomonas reinhardtii* Transcription factors. In the end, we got 114 *Chlamydomonas* TFs and cloned them into the Y2H vector pADlox and pBDlox separately by Gateway assembly or restriction enzyme digestion. Sequence of each amplified TF is attached in the supplementary file. Then TF-pADlox plasmid from AD library was transformed to Y8800 strain one by one (**Figure 1C**), and Trp gene in the pADlox plasmid confers positive yeast strain transformed with its plasmid ability to survive on SC Trp- plate. TF-pBDlox plasmid from BD library was transformed to CRY8930 strain individually (**Figure 1D**), and Leu gene in the pBDlox plasmid enables yeast strain ability to survive on SC Leu- plate.

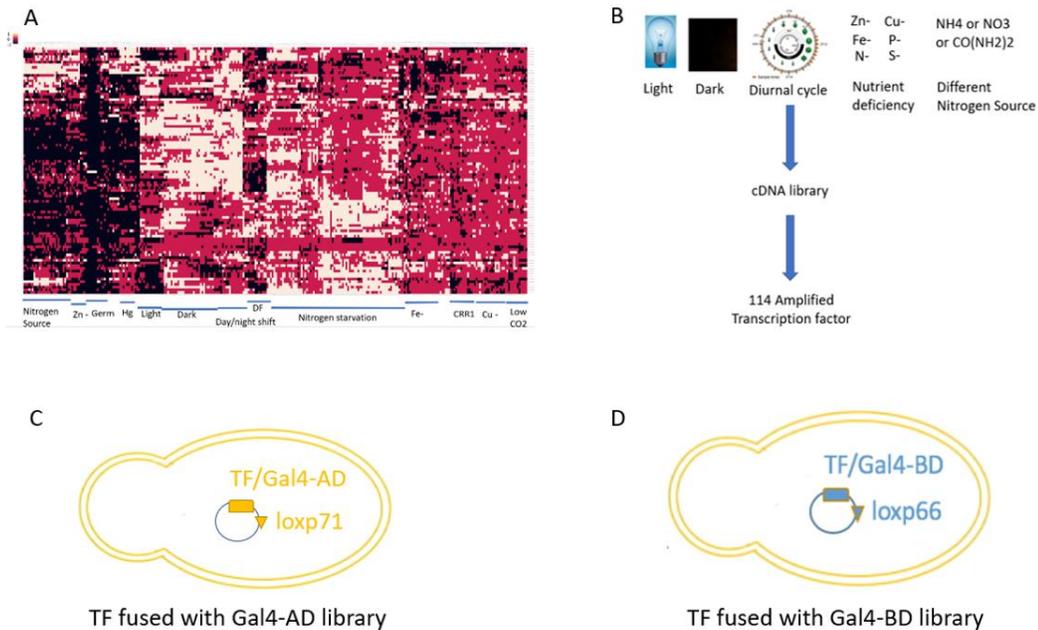


Figure 2.1. Construction of *Chlamydomoans reinhardtii* Transcription factor AD library and BD library. **A.** Differential RNA expression profile of algal Transcription factors (TFs) from all published RNAseq datasets. **B.** Amplification of TFs from homemade RNA library and cDNA library from ChlamyCollection center (<https://www.chlamycollection.org/products/cdna-libraries/>). **C.** Each TF was ligated to Gal4-AD vector individually and transformed to yeast Y8800 strain one by one to build the TF-AD library. **D.** Each TF was ligated to Gal4-BD vector and transformed to CRY8930 strain separately to build the TF-BD library.

Quality control for false positive TF for CRY2H-seq assay

For each library, we tested if each TF could induce false positive problem during Y2H assay. We mated each TF-AD Y8800 strain to CRY8930 strain transformed with empty pBDlox plasmid. After yeast mating, cells were

incubated on SC Trp-/Leu-/His- plate with 1mM 3AT for one week at 30 degrees Celsius. There are four TF-AD that could survive on SC Trp-/Leu-/His- plate (**Supplementary Figure 1**), suggesting that they have ability to bind to yeast Gal1 promoter. With the same strategy, each TF-BD CRY8930 strain was mated to Y8800 strain transformed with empty pADlox plasmid. Diploid cells were incubated on SC Trp-/Leu-/His- plates with 1mM 3AT at 30°C. Three days later, 25 TF-BD clones generated leaky His3 gene expression, allowing them to grow on this selection plate. In addition, we increased the concentration of 3AT to 10mM and 20mM (**Supplementary Figure 2**), 15 TF-BD clones can still grow on 20mM plate. These 25 false-positive TF-BD suggest that they may have the ability to recruit yeast RNA polymerase directly or through a mediator activator. In our later Y2H assay, we remove all false positive TF-AD and TF-BD strains from our library. Equal amount of each TF-AD Y8800 strain was pooled together to build the yeast TF-AD library, and equal amount of each TF-BD CRY8930 strain was pooled together for TF-BD library.

Y2H breeding and sequencing for TF-TF interaction with a 10X higher selection pressure

20 OD600 of Yeast TF-AD library and TF-BD library were mixed and incubated at 30 degrees for 5 hours. Hygromycin in SC Trp-/Leu- medium was used to treat mating cells overnight to kill unmated cells. Enriched diploid cells were plated on SC Trp-/Leu- plate with 10mM 3AT for three days at 30 degrees C. In any diploid cell, if one TF could bind to another TF, then such complex will bind to pGal1 and pGal4 promoter and initiate transcription of His3 and Cre protein (**Figure 2A**). Cre protein will then fuse TF-pADlox plasmid and TF-pBDlox plasmid together through loxp71 and loxp66 site. This hybrid region can be amplified by using one primer from pADlox vector and another primer from pBDlox plasmid. Here we pooled all diploid cells from selection plate, extracted all plasmids from cell lysate and use unique primer AD-F and BD-R to amplify this hybrid region. Then PCR product was purified and sent to UCSD core facility for sequencing.

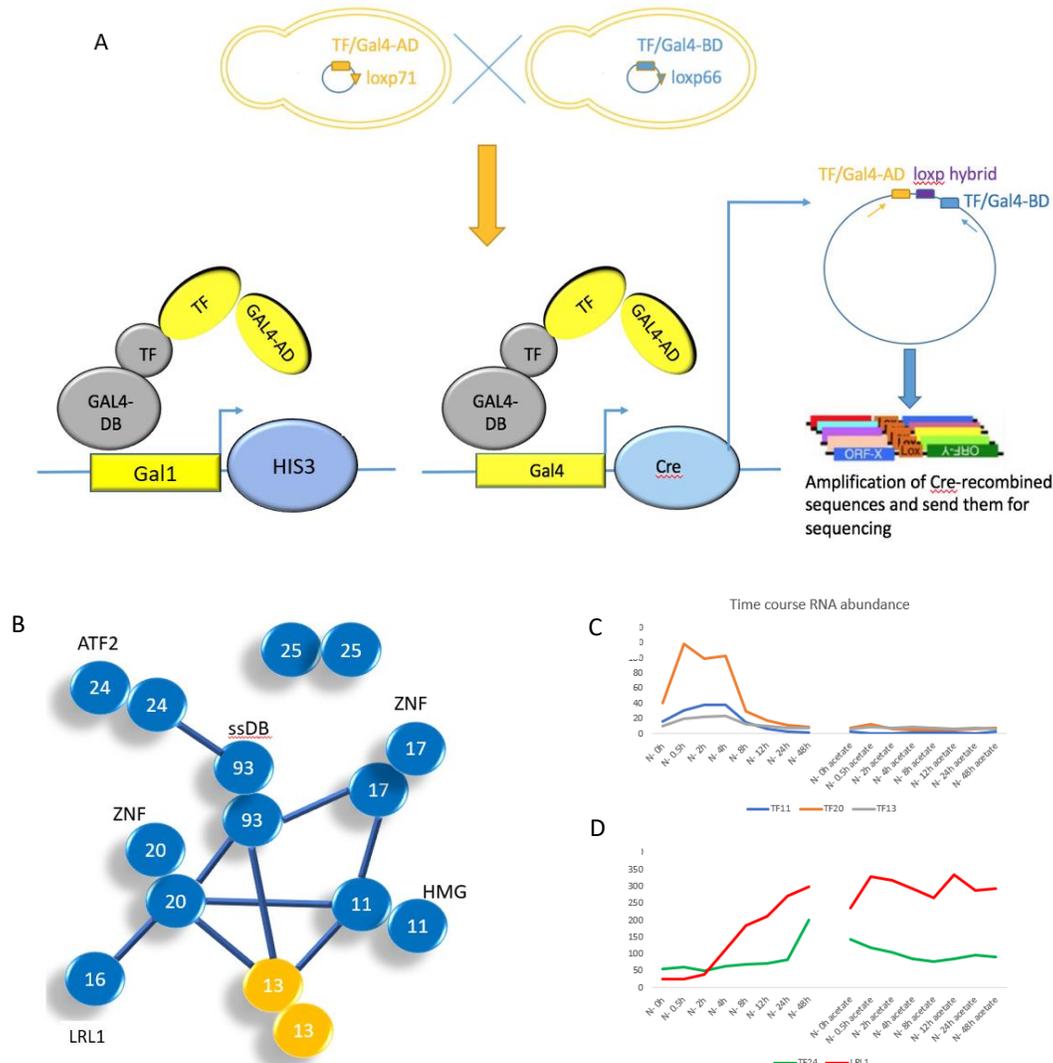


Figure 2.2. Transcription factor interaction subnetwork identified by Y2Hseq assay. **A.** Pooled Yeast AD library strains were mixed and mated with BD library strains. If a TF in AD library can bind to another TF in BD library, then Cre and His3 genes will be expressed in the mated cell and fuse two plasmids via loxp sites. By amplifying fusion region with primers from two plasmids separately, which are indicated as yellow and blue arrow, we can amplify only the hybrid region and identify individual TF-TF interactions from sequencing reads. **B.** Subnetwork obtained from CrY2H-seq assay with 10mM 3AT. RNAseq time course data under nitrogen starvation with or without acetate were displayed next to the TF interaction map. #13 is a Polymerase while rest blue proteins are predicted as Transcription factors. **C.** Time course transcript change after nitrogen starvation(left) and nitrogen starvation with acetate(right) for TF11, TF13, TF20. **D.** Time course transcript change after nitrogen starvation(left) and nitrogen starvation with acetate(right) for TF24 and TF16(LRL1).

Based on the sequencing result, we did alignment to all TF ORFs, yeast genome and vector sequence, performed quality filter and identified a small subnetwork of *Chlamydomonas* TF-TF interaction network (**Figure 2.2**). The reason why we only identified 10 TF interactions, is because we used 10mM 3AT instead of conventional 1mM 3AT to make sure there are fewer false positives, although our library already excludes all false positive TFs.

From this TF interaction subnetwork, only TF16(LRL1) was discovered to be a master regulator for lipid accumulation (Hidayati et al., 2019). We reviewed their expression pattern under nitrogen starvation and sulfur starvation, and found that most of them exhibit similar expression pattern under Nitrogen starvation and sulfur starvation can be divided into three categories within 48 hours window after nitrogen starvation: early effector, intermediate effector and late effector (**Figure 2C, 2D**). We took one further step to investigate if any of these TFs, like TF16 (LRL1), also participate in the lipid regulation under stress condition.

Lipid accumulation was depressed under Nitrogen starvation in tf24 mutant and tf20 mutant strains.

Mutant strains tf24m and tf20 were obtained from CLIP library (Zhang et al., 2014; Li et al., 2016). Those mutants were generated by random insertion through electroporation of a paromomycin resistance cassette CIB1 (Sizova et al., 2001; Zhang et al., 2014; Li et al., 2016). Difference of triacylglycerol (TAG) accumulation was observed 1 day after Nitrogen starvation between transcription factor mutant strains and wild type strains (**Figure 2.3A**). Lipid accumulation was slightly increased in tf20 mutant strain 24 hours after incubation in TAP N- medium (**Figure 2.3B**), suggesting that loss of TF20 cannot fully suppress the lipid accumulation. Based on the published RNAseq and Chipseq data (Goodenough et al., 2014), TF20 was induced at 2~6 hours after nitrogen starvation, indicating that it acts as an early mediator in lipid accumulation to recruit some master regulator like LRL1(Hidayati et al., 2019). In addition, lipid level of tf24 mutant strain in N- medium was always similar to that in the TAP medium, indicating that such loss of lipid synthesis under Nitrogen starvation is complete in tf24 mutant and therefore TF24 is a master regulator to control downstream lipid-synthesis-related genes. All three strains grown in TAP medium or 48h after Nitrogen starvation were sent for mass spec analysis (**Figure 2.3E**). The total fatty acid methyl esters (FAMES) were enriched in the 4533 wt strain under nitrogen starvation, while there is no difference of FAMES in tf20 mutant strain before and after Nitrogen starvation. Interestingly, although some fatty acids were increased in tf24 mutant strain, total FAMES were decreased after nitrogen starvation. The overall FAMES content change from mass spec data is consistent with our Nile red staining assay.

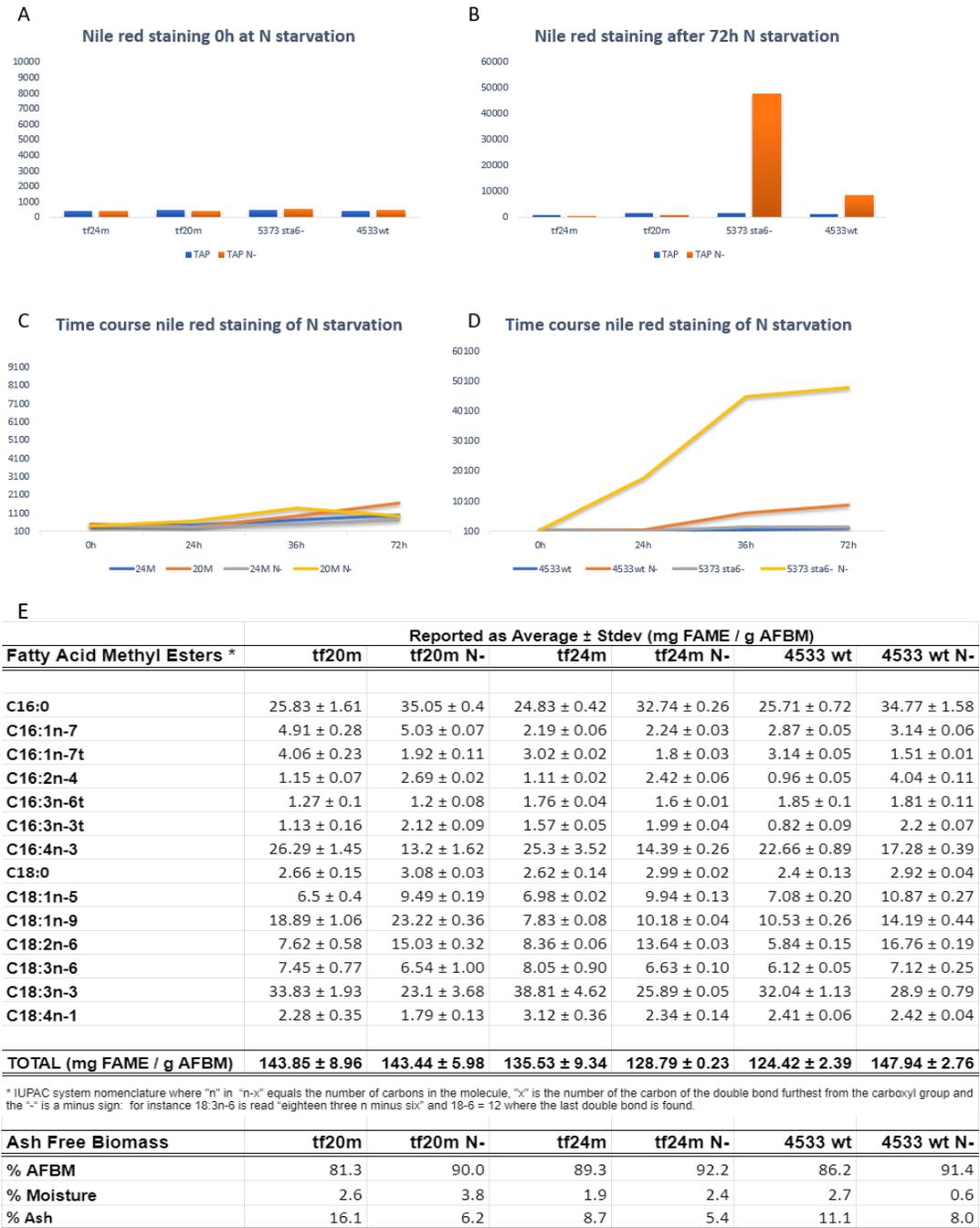


Figure 2.3. Lipid analysis of 4533 wild type , TF24 mutant, TF20 mutant and 5373 sta6- mutant strain. Nile red staining was measured with excitation wavelength 530nm and emission wavelength 575nm at Time point 0 h(A), and 72h(B) after Nitrogen starvation. Time course mutant samples of tf24m and tf24m were measured by Nile red staining at different time points after Nitrogen starvation or in control TAP medium(C). Two positive control strains 4533 wild type and 5373 sta6 mutant strain were harvested at different time points after Nitrogen starvation or in control TAP medium(D). Fatty acid methyl esters analysis from Mass spec results for tf mutant strain and 4533 wt strain 48hours after nitrogen starvation(N-) or in control TAP medium(E).

Time course RNAseq data under nitrogen starvation suggests TF20 is a repressor for ribosome biogenesis while TF24 is a potential regulator in the nitrogen metabolism.

To further investigate the potential regulation mechanism for each TF, we did a time course RNAseq data after Nitrogen starvation for each strain. Differential gene expression patterns were observed in tf20m, 4533wt and 24m strain from different time points after nitrogen starvation (**Figure 2.4A**). And the number of upregulated genes in tf20m and tf24m 48 hours after nitrogen starvation were much higher than that in wt strain (**Figure 2.4B**), while the overall downregulated genes 48 hours after Nitrogen starvation was less in tf20m and tf24m strains (**Figure 2.4C**), although there is no differential change for the housekeeping genes among these groups(**Supplementary Figure 2.3**). Consistent with that, downregulated KEGG Pathway from wt strain 48h_versus_0h after Nitrogen starvation (**Figure 2.4H**) were correspondingly upregulated in either tf20m (**Figure 2.4I, Supplementary Figure 2.4**) or tf24m strain (**Figure 2.4J**) 48h_vs_0h after Nitrogen starvation. For example, Carbon fixation in photosynthetic organism, Pentose phosphate pathway, Biosynthesis of alkaloids, Aminoacyl-tRNA biosynthesis, Glyoxylate and dicarboxylate metabolism, Biosynthesis of plant hormones, SNARE interaction in vesicular transport and Ribosome were downregulated in wt strain 48 hours after Nitrogen starvation when compared to the 0h at Nitrogen depletion (**Figure 2.4H**). But these pathways were upregulated in tf20m strain 48 hours after Nitrogen starvation versus 0h at Nitrogen deprivation (**Figure 2.4I**). In addition, Oxidative phosphorylation, Nitrogen metabolism and One carbon pool by folate and SNARE interactions in vesicular transport were downregulated in wt strain 48 hours after Nitrogen starvation (**Figure 2.4H**), but upregulated in tf24m strain 48 hours after Nitrogen starvation compared to the basal level before Nitrogen starvation (**Figure 2.4J**). Moreover, degradation of Ethylbenzene, Benzoate, 1- and 2-Methylnaphthalene, Limonene and pinene, 3-Chloroacrylic acid, 1,2-Dichloroethane, and downstream pathway Phenylalanine, tyrosine and tryptophan biosynthesis were specifically upregulated in tf24m mutant 48 hours after Nitrogen starvation. These results suggest that TF20 is a master regulator for Ribosome suppression after Nitrogen starvation while TF24 act as a repressor to Oxidative phosphorylation pathway and degradation of aromatic molecules after Nitrogen starvation.

We then did GO enrichment for those differentially expressed genes in each strain. Interestingly, many Gene Ontology groups were enriched in the differential expressed genes from wt strain after Nitrogen starvation (**Figure2.4G, Supplementary Figure 2.5**), but only ribosome related genes for translation were uniquely and significantly upregulated in the tf20m strain (**Figure 2.4H, Supplementary Figure 2.6**), suggesting that TF20 may play a role to suppress overall protein translation and biosynthesis under nitrogen starvation. Consistent with our GO

enrichment analysis, comparative Parametric Analysis of Gene set Enrichment (PAGE) analysis of time course RNAseq data for wt and tf20m strains reveals that Biological processes protein translation(GO: 0006412) and protein folding(GO:0006457), organic substance catabolic process (GO:1901575), Proteolysis involved in cellular protein catabolic process(GO: 0051603) and DNA replication (GO: 0006260) were uniquely downregulated in wt strain 48 hours after nitrogen starvation vs 0h, while purine ribonucleotide metabolic process (GO: 0009119) and part of light reaction (GO:0019684) and light harvesting complex (GO:0009765) in photosystem were downregulated in both wt and tf20m 48 hours after nitrogen starvation vs 0h (**Supplementary Figure 2.7**). GO enrichment in cellular component and molecular function was consistent with this PAGE analysis for GO enrichment in biological process: enzymes in ER and ribosome were uniquely upregulated in tf20m strain 48 hours after N deprivation (**Supplementary Figure 2.8**), and specific enzymes with following molecular functions: ribosome(GO: 0005840), translation initiation factor activity(GO:0003743) and GTP binding(GO: 0005525) related machinery were uniquely upregulated in the tf20m 48hours after nitrogen starvation (**Supplementary Figure 2.9**). Besides the above opposite pattern of pathway in wt and tf20m strains, cellular amino acid biosynthetic process (GO:0008652), response to oxidative stress (GO: 0006979) and electron transport chain (GO: 0022900) were uniquely upregulated in tf20m strain (**Supplementary Figure 2.7**).

Transcription factors (GO: 0003712) were uniquely downregulated in tf24m strain 48 hours after nitrogen starvation. Oxidoreductase activity (GO: 0016671) pathway and threonine-type peptidase (GO: 0004298) were uniquely upregulated in tf24m strain 48 hours after nitrogen starvation (**Supplementary Figure 2.10**). Combined with our KEGG pathway analysis for differential expressed genes, TF24 might act as a master regulator for specific TFs to repress oxidative phosphorylation and degradation of aromatic carbon chemicals after Nitrogen starvation.

N-rich subsystems, such as photosystem (GO: 0034357) and ribosome (BP GO: 0006412) related genes, were downregulated in wt strain 48 hours after nitrogen starvation versus 0 hour at nitrogen starvation (**Supplementary Figure 2.10**). Consistent with our lipid assay, metabolic genes in the fatty acids pathways were uniquely downregulated in the tf20m and tf24m strains but not in the wt strain (**Supplementary Figure 2.11, 2.12**).

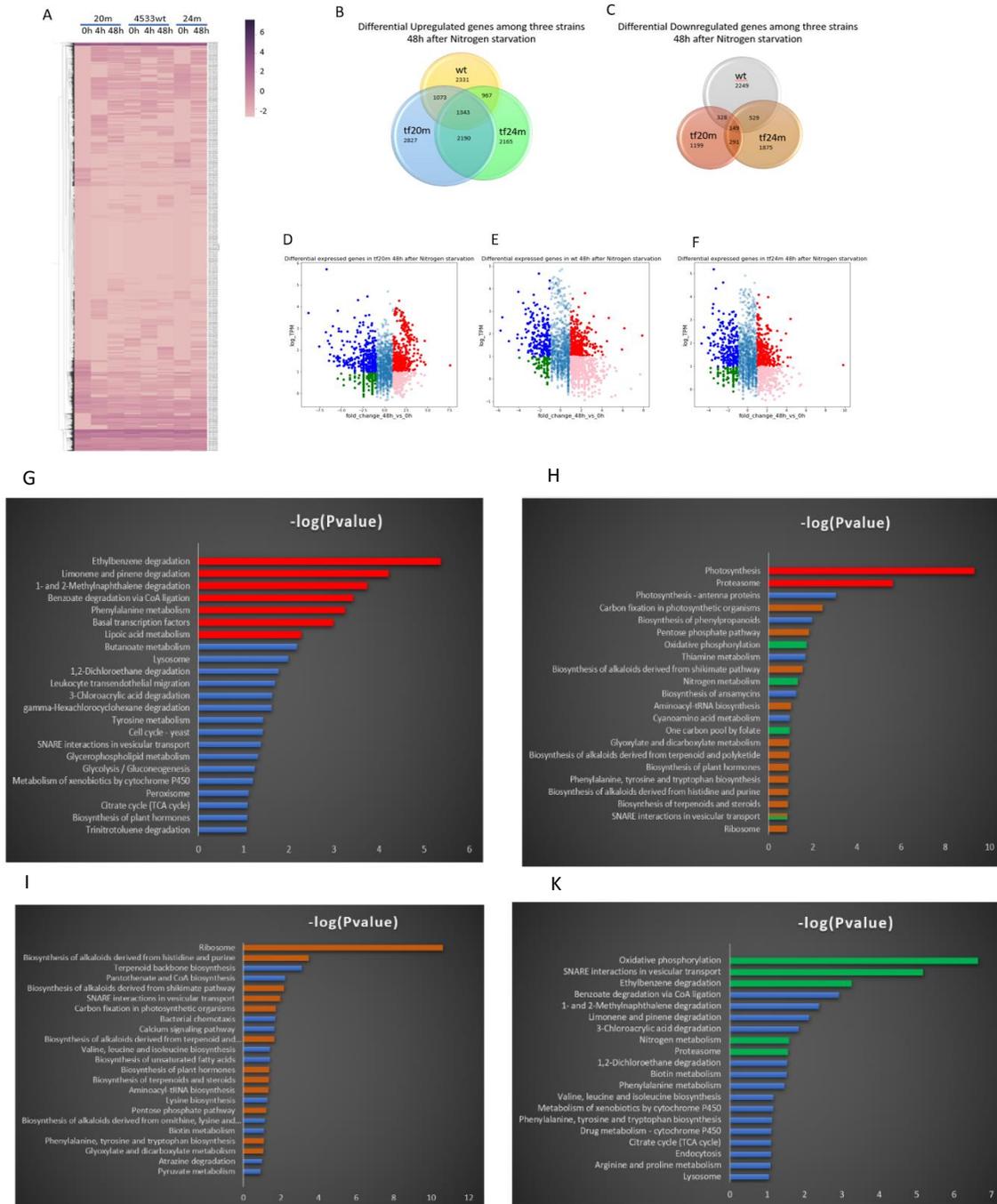


Figure 2.4. Time course RNAseq data of transcription factor mutants and 4533 wild type under Nitrogen starvation. **A.** Hierarchical clustering heatmap result for differential RNAseq data of *tf20m*, 4533 wt and *tf24m*. **B.** Venn diagram of Differential upregulated genes among three strains 48hour after Nitrogen starvation versus 0h at Nitrogen starvation. **C.** Venn diagram of Differential downregulated genes among three strains 48hour after Nitrogen starvation versus 0h at Nitrogen starvation. **D.** Differential expressed genes in *tf20m* 48 hour after Nitrogen starvation versus 0h at Nitrogen starvation. **E.** Differential expressed genes in 4533 wt strain 48 hour after Nitrogen starvation versus 0h at Nitrogen starvation. **F.** Differential expressed genes in *tf24m* 48 hour after Nitrogen starvation versus 0h at Nitrogen starvation. **G.** KEGG pathway enriched in upregulated genes 48 hours after Nitrogen starvation in wt strain. **H.** KEGG pathway enriched in downregulated genes 48 hours after Nitrogen starvation in wt strain. **I.** KEGG pathway enriched in upregulated genes 48 hours after Nitrogen starvation in *tf20m* strain. **J.** KEGG pathway enriched in upregulated genes 48 hours after Nitrogen starvation in *tf24m* strain.

Discussion

Nitrogen starvation in *Chlamydomonas reinhardtii* has been well studied by different omics techniques and groups, although only NRR1 identified as master regulator in lipid accumulation and NIT2 transcription factor for nitrogen assimilation (Fernández and Matagne et al., 1986; Camargo et al., 2007) under nitrogen starvation (Boyle et al., 2012). In addition, few TFs, such as PSR1 and LRL1, have been reported in lipid remodeling under phosphate deprivation. Here we showed that with a high throughput sequencing assay, we identified several algal Transcription factors interaction and discovered two master regulators that might be involved in regulation of ribosome and oxidative phosphorylation separately under Nitrogen starvation.

Metabolism and proteome reallocations have been reported to be significantly changed under nitrogen starvation (Fernández et al., 2009; Schmollinger et al., 2014, Park et al.,2015). Nitrogen assimilation pathway were upregulated with nitrogen deprivation in all three strains we tested and many other strains used in published papers(Romero et al., 1987; Fernández et al., 1989; Quesada et al., 1998; Camargo et al., 2007). Ammonium transporter AMT1, AMT8, AMT4, Urea transporter and Urea transporter DUR1, DUR3, DUR4, DUR5 were upregulated in all three strains and Purine transporter uracil permease UAPA6 and UAPA1, UAPA3 significantly upregulated and their downstream effectors UOX and XDH were upregulate as well among three test strains. Transcriptions of such nitrogen uptake and assimilation pathways have been reported to quickly induced from 1 to 24 h after nitrogen starvation in previous reports (Fernández et al., 2009; Schmollinger et al., 2014; Park et al.,2015).

Uptake of external nitrogen is then transported or reduced to intracellular ammonium, which will be catalyzed by Glutamine oxoglutarate aminotransferase (GOGAT) to yield Glutamine and Glutamic acid. These two amino acids can be used for the synthesis of purine, pyrimidines and chlorophyll, thus interconnected with Carbon metabolism and photosynthesis (Nunes-Nesi et al., 2010). However, purine ribonucleotide metabolic process(GO: 0009119) and photosynthesis were downregulated in all three tested strains 48 hours after nitrogen starvation vs 0h(Supplementary Figure 2.7), which is in line with the previous work(López García de Lomana et al., 2015, Schmollinger et al., 2014). The universal downregulation and degradation of photosystem is suggested to be sacrificed for more lipid and starch storage and to reduce excessive photo oxidative stress. We took a further analysis of these downregulation pattern in three strains and found that transcription of 15 photosystem related proteins and 6 antenna proteins were downregulated in wt strain, while only transcription of 6 antenna proteins were downregulated in tf20m strain and

transcription of 8 non antenna photosynthetic genes were downregulated in tf24m strain. This differential pattern may imply the gene-specific regulation for photosystem under nitrogen starvation.

Besides elevated mRNA level of Nitrogen uptake and assimilation and downregulation of purine biosynthesis and photosynthesis pathway, N-rich proteins, including ribosomes and Rubisco, were downregulated in wt strain after nitrogen starvation while the corresponding degradation pathways were upregulated in wt strain, consistent with the previous reports (Schmollinger et al., 2014, Siersma and Chiang, 1971; Martin and Goodenough, 1975). Such degradation could generate more free Nitrogen and save internal Nitrogen usage right after Nitrogen starvation. In addition, decrease of ribosome in both transcript and protein level under Nitrogen limitation suggest that there's a global reduced protein synthesis rate to further decrease the Nitrogen use (Schmollinger et al., 2014).

Given the change of C:N ration occurs not immediately but several hours after nitrogen starvation, downregulation of these macromolecules for reutilization of internal N pools right after Nitrogen depletion was suggested to be regulated by early sensor (Park et al., 2015). Here we showed that in our tf20m strain, contrast to the wt and tf24m strain, there is a striking upregulation of ribosome after nitrogen starvation, and based on expression pattern of TF20 after Nitrogen starvation from many other published datasets (Boyle, et al., 2012, Blaby, et al., 2013, Goodenough et al., 2014), it's possible that TF20 is the early sensor that determine the degradation of ribosome. But how TF20 regulates ribosome degradation needs to be further investigated.

However, with an upregulation of Ribosome mRNAs in tf20m strain after N depletion, the global Nitrogen usage should be changed systematically. We did observe that Biosynthesis pathways for alkaloids, sterols, terpenoids and aminoacyl-tRNA were upregulated in tf20m strain but downregulated in wt strain 48 hours after Nitrogen starvation. Although terpenoids and sterols don't contain Nitrogen atom, they can be part of alkaloids which contain at least one Nitrogen atom. Downregulation of alkaloids can save internal Nitrogen amount under Nitrogen limitation in wt strain. However, these N-rich sectors were upregulated in tf20m strain, suggesting that tf20m plays a critical role in suppressing N demand from the early phase of Nitrogen starvation. In addition, Pentose phosphate pathway, Glyoxylate metabolism, Carbon fixation and its downstream pathway Phenylalanine, tyrosine and tryptophan biosynthesis were uniquely upregulated in tf20m while they were downregulated in wt strain after nitrogen depletion when compared to nitrogen replete condition. Reduced Glyoxylate transcripts may suggest that cells try to suppress this pathway to generate more acetyl-CoA as a substrate for fatty acid biosynthesis (Miller et al., 2010, Park et al., 2015).

For tf24m, there is another unique upregulation pattern after nitrogen starvation in terms of transcript level. Oxidative phosphorylation was increased significantly 48 hours after nitrogen starvation in tf24m strain but downregulated in wt strain. Such result is consistent with previous work to show that respiratory oxidative phosphorylation was decreased to generate less unstable ATP when there's Nitrogen depletion (Chen et al., 2015). In addition, degradation of Ethylbenzene, Benzoate, 1- and 2-Methylnaphthalene, Limonene and pinene, 3-Chloroacrylic acid, 1,2-Dichloroethane, and TCA cycle were upregulated in tf24m mutant 48 hours after Nitrogen starvation. In addition, Phenylalanine metabolism and Phenylalanine, tyrosine and tryptophan biosynthesis were also upregulated, which are the downstream pathway connected to the aforementioned aromatic chemicals. All of these results suggest TF24 regulate the suppression of mitochondrion oxidative phosphorylation to transit cell state from a energy consumption state to an energy storage state under nitrogen starvation, and suppress the degradation of Carbon molecules with aromatic rings. It's unclear why algal cells prefer degradation of aromatic molecules under nitrogen starvation in tf24m. There are two hypotheses: One is that with an elevated oxidative phosphorylation in tf24m strain, energy supply and C/N balance was changed so that cells have to degrade more aromatic molecules to produce enough metabolite intermediate and amino acids for respiration and substrate for Nitrogen metabolism. The other is that wild type algal cells may accumulate these aromatic molecules for the energy storage under nitrogen starvation and avoid excess degradation of Carbon molecules. Although above discussed pathways were changed significantly in tf24m after N limitation, it's hard to believe that tf24m directly regulate so many different genes. From our GO enrichment, we also found that contrast to wt strain and tf20m, a subset of TFs (GO: 0003712) were downregulated only in tf24m strain after Nitrogen starvation, suggesting that tf24m may be a pioneer factor to regulate different pathways via specific TFs. More investigations need to be done to fully understand such regulation combined with other omics data in different knock-out strains.

Lipid accumulation and fatty acid synthesis are well known to be increased after nitrogen starvation in algal cells. C16:0, C16:3, C18:0, C18:1, C18:2 was accumulated in all three strains after nitrogen starvation, although their fold changes were higher in wt than that in TF mutant strains. C16:4 and C18:3 was reduced in all three strains, but such reduction was more significant in tf mutant strains when compared to wt strains. Moreover, C18:4 level didn't change in wt but got reduced in tf mutants. With a lower accumulation of C16:0, C16:3, C18:0 and higher reduction of C16:4, C18:3 and C18:4 in two tf mutant strains, total FAMES after nitrogen starvation were lower or equal to the level before nitrogen starvation. Acetyl-CoA carboxylase A (ACCCaseA, Cre12.g519100) was upregulated in wt and

tf20m strains but downregulated in tf24 mutant strain, while ACCaseC was upregulated in tf20m strain but downregulated in wt and tf24 mutant strains. Biotin carboxyl carrier protein (BCC1) were downregulated in wt and tf24m strains but upregulated in tf20m strain, while beta-carboxyltransferase (BCX1) was upregulated in all three strains. All differential expression of enzymes suggests that there is a more complicated post transcription regulation for the initiation of Acetyl-CoA to fatty acid biosynthesis pathway. Moreover, GPD1 was upregulated in all three strains but other GPD enzymes exhibited different pattern among three strains. Chloroplastidic desaturase $\Delta 12$ (CDD12) and Stearoyl-CoA desaturase $\Delta 9$ (SCD) were increased after nitrogen starvation. In addition, pathway analysis showed that enzyme for the third step for biosynthesis of unsaturated fatty acids was upregulated while the last step in this pathway was downregulated in tf24m strain only. Interestingly, there's increased level of enzyme in fatty acid degradation pathway but decreased level of enzyme in fatty acid biosynthesis 48 hours after N depletion in wt strain. But the opposite pattern of fatty acid biosynthesis and degradation was observed in tf20m and tf24m. Taken together, change in transcript level cannot be used to elucidate the accumulation of TAG and lipid body, but protein level and post translation modification should be considered to explain such lipid accumulation based on previous work (Park et al., 2015). And the detailed mechanism for different FAME synthesis is still not clear and needs to be investigated with specific enzyme knock-out strain in future.

This study represents the first characterization of Transcription factor interaction network in *C.reinhardtii* by high-throughput Y2Hseq, although only 110 algal predicted TFs were amplified from total 230 TFs. We tried to amplify the rest TFs by using libraries from Chlamy collection center, homemade cDNA or even purified gDNA, but most of the rest TFs are super long, usually larger than 4kb, and thus difficult to get amplified by PCR. In the future probably with more TFs synthesized in vitro or amplified by other groups, a more comprehensive TF interaction network can be generated. Among these TF interactions, we found two master regulators, which interact with a recently reported master regulator LRL1 for lipid accumulation, might be involved in suppression of ribosome and oxidative phosphorylation under nitrogen starvation separately for the first time.

Methods

Strain and plasmid

Yeast strains (CRY8930 and Y8800) and plasmids (pADlox and pDBlox) used for CrY2H-seq were obtained from the Arabidopsis Biological Resource Center (<https://abrc.osu.edu/>). The genotype of CRY8930 is MAT α leu2-

3,112 trp1-901 his3-200 ura3-52 gal4Δ gal80Δ PGAL2-ADE2 LYS2::PGAL1-HIS3 MET2::PGAL7-CRE-HPHMX6 cyh2 R. The genotype of Y8800 is MATa leu2-3,112 trp1-901 his3-200 ura3-52 gal4Δ gal80Δ PGAL2-ADE2 LYS2::PGAL1-HIS3 MET2::PGAL7-lacZ cyh2 R.

Transcription factor library construction

Predicted TF list was downloaded from PlantTFDB database(planttfdb.cbi.pku.edu.cn/index.php?sp=Cre) Primers were designed for each TF and synthesized by IDT. Core cDNA library and Stress 1 cDNA library were purchased from Chlamycollection(<https://www.chlamycollection.org/products/cdna-libraries/>). At the same time, homemade cDNA library was generated from *Chlamydomonas reinhardtii* cc1690 incubated under different conditions(light, dark, nitrogen/sulfur/phosphorus starvation, iron/copper deficiency). These cDNA libraries were used as templates for TF PCR amplification.

After PCR amplification, each TF was cloned into pADlox or pBDlox vector by Gateway assembly or restriction enzyme digestion. Recombinant plasmids were transformed to MAX Efficiency™ DH5α Competent Cells (ThermoFisher Scientific, Cat #18258012) for storage and amplification. *E.coli* colonies were selected by Ampicillin plate and verified by restriction enzyme digestion and sequencing. Plasmids were purified from right *E.coli* colony by GeneJet plasmid miniprep kit(ThermoFisher Scientific, Cat # K0502).

Then we transformed each TF-AD plasmid into Y8800 yeast competent cells by Lithium acetate transformation. With the same transformation protocol, we also transformed each TF-BD plasmid into CRY8930 strain to generate TF-BD library. Positive Y8800 colonies transformed with TF-AD were selected on SC Trp- plate(MilliporeSigma Supelco, Cat # Y1771-20G) and positive CRY8930 colonies transformed with TF-BD were selected on SC Leu- plate(MilliporeSigma Supelco, Cat # Y2021-20G). Colonies were inoculated in SC Trp- or SC Leu- medium for three days at 30 degree Celsius and stored in 30% glycerol in -80 freezer for downstream work. The simplified yeast plasmid transformation is shown below: single colony of yeast strain was inoculated in 5ml culture tube overnight at 30 degree Celsius. Yeast culture was inoculated into fresh YPD medium to make initial OD as 0.1~0.2. Six hours after incubation at 30 degree, cells were harvested by centrifugation at 1500rpm when OD reaches 0.6~0.8. Centrifuged cells were resuspended with 1M LiCl and centrifuged again. Repeat this step twice and resuspend cells with 1M LiCl/PEG buffer. Plasmids were incubated with competent yeast cells and heated salmon sperm ssDNA at 42 degree Celsius for 1hour. Later cells were centrifuged and washed with ddH2O. Resuspended cells in ddH2O was spotted on the SC Trp- or SC Leu- plate.

CrY2H-seq screening of transcription factor libraries

Each replicate screen consisted of mating ~ 20 OD600 of each TF clone library (pADlox in Y8800 and pDBlox in CRY8930). Based on cell titers of 2×10^7 cells/OD that we observed for each library, we estimated that each replicate screen would test the ~ 3.6 million possible protein combinations at 10-fold excess, assuming a 10% mating efficiency

Next-generation sequence analysis of CrY2H-seq screen libraries

Reads were mapped using Bowtie2-2.0.241 local alignment with default settings to a custom genome composed of *Chlamydomonas* TF coding sequences, the *Saccharomyces cerevisiae* genome, Gal4 AD and Gal4 DB domain sequences, and the empty CrY2H-seq plasmid sequences. A quality filter was applied requiring reads to map with at least 30 matching bases, allowing a maximum of 2 mismatches, 2 insertions or deletions, and 2 bases of trimming from the beginning of the read. Reads were then joined with their corresponding read pairs and included in the next analysis step only if both reads passed the first filter and mapped to *Chlamydomonas* TF ORF sequences. Clonal fragments were removed from read pairs if both reads in a fragment contained the same start positions. Paired reads for which each of the mates aligned to a different ORF and showed the same strand orientation (Cre recombination occurs such that ORFs on pADlox and pDBlox plasmids become inverted in a 3'-to-3' orientation) were included in further analysis. Fragments were further subjected to a size filter that required that the sum of the lengths of each read (start position of each read to the end of each ORF) and the lox region conformed to the expected library size of 400–600bp. Remaining fragments that mapped to Cre-recombined ORF junctions were totaled. Each screen had on average ~ 1.4 million fragments corresponding to ORF junction sites and ~ 16 million fragments mapping to gene bodies. Remaining data mapped to priming site region ORF junctions or did not align. Analysis scripts can be found in Supplementary Software. After applying the basal fragment cutoff mentioned above to all data sets, fragments were normalized by the median filtered fragments as follows: A scale factor for each replicate dataset was determined by dividing the filtered protein interaction fragments by the median filtered protein interaction fragments. The number of fragments per protein pair was multiplied by this scale factor and rounded down to the nearest integer to normalize protein interaction fragments.

Time course RNAseq analysis of wt strain, tf mutant strain under Nitrogen starvation

Cell cultures of either wt strain or tf mutant strains were grown in 0.2 L TAP medium in 1 L Erlenmeyer flasks at 30°C with shaking at 120 rpm under continuous white light illumination ($\approx 100 \mu\text{mol photons/m}^2 \text{ s}$) at

atmospheric carbon dioxide (0.03% CO₂) within a 120 rpm shaking incubator. Cultures with 3 x 10⁶ cells mL⁻¹ were pelleted at log phase by centrifugation at 3000×g, at room temperature for 10 min. Cell pellets were washed 3 times with fresh TAP nitrogen free medium and resuspended in TAP nitrogen free medium. Cells of each strain were harvested at 0h, 4h and 48h after nitrogen starvation separately for RNA extraction. PureLink™ Plant RNA Reagent (Invitrogen, Cat # 12322012) were used to for cell lysis with a ratio of 1ml reagent : 0.2g algae biomass. Equal volume of Chloroform – isoamyl alcohol mixture (Millipore Sigma, Cat # 25666-500ML) was added to isolate proteins and nucleotides. Extracted RNA in upper phase after 10mins centrifugation at 12,000g (4 °C) were harvested and precipitated with equal volume of Isopropanol (Millipore Sigma, Cat # 67-63-0). RNA pellets were obtained after 15mins centrifugation at 12,000g (4 °C) and washed with 75% ethanol. RNA was dried for 10mins, resuspended with DEPC treated water (Invitrogen, Cat # AM9920) and sent to UCSD core facility for sequencing.

RNAseq reads were mapped using GSNAP aligner to the reference genome of *Chlamydomonas reinhardtii*. Sequence alignments were counted with Hisat2 and analyzed with edgeR to generate TPM for comparative quantification among samples. Differential expressed genes (DEGs) between every two time points were calculated when their log₂ fold change is great than 2. DEG lists were analyzed by algal functional annotation tools to identify the differential KEGG pathway enrichment (Lopez et al., 2011). In addition, DEG lists were submitted to AgriGOv2 for the GO enrichment (Tian et al., 2017). Parametric Analysis of Gene set Enrichment (PAGE) analysis of whole time course TPM table was done in AgriGOv2 as well to systematically identify differential GO enrichment cross samples and time points.

Determination of total lipid content and Individual FAME profile

Cell cultures of either wt strain or tf mutant strains were grown in 4 L Tap medium in 2 L Erlenmeyer flasks (1 L in each 2 L flask) at 30°C with shaking at 120 rpm under continuous white light illumination ($\approx 100 \mu\text{mol photons/m}^2 \text{ s}$) at atmospheric carbon dioxide (0.03% CO₂) within a 120 rpm shaking incubator. Cultures were pelleted at log phase by centrifugation at 3000×g, at room temperature for 10 min. Cell pellets were washed 3 times either with fresh standard TAP medium or TAP nitrogen free medium. Then, half of the cells were resuspend in 2 L fresh standard TAP medium and the other half were resuspend in fresh TAP nitrogen free medium. Cultures were grown as 1 L each in 2 L Erlenmeyer flasks under the same ambient conditions as described above. After 48 h, cell cultures were harvested and subjected to Lyophilization. Total lipid content, individual fatty acid methyl esters

(FAMEs) Ash free biomass (AFBM) were determined following the same protocol and instruments performed by Smalley et al., 2020.

Imaging analysis

Nile red staining was performed for cc1690 wild type strain, TF20 mutant strain and TF24 mutant strain under nitrogen starvation or addition of cAMP.

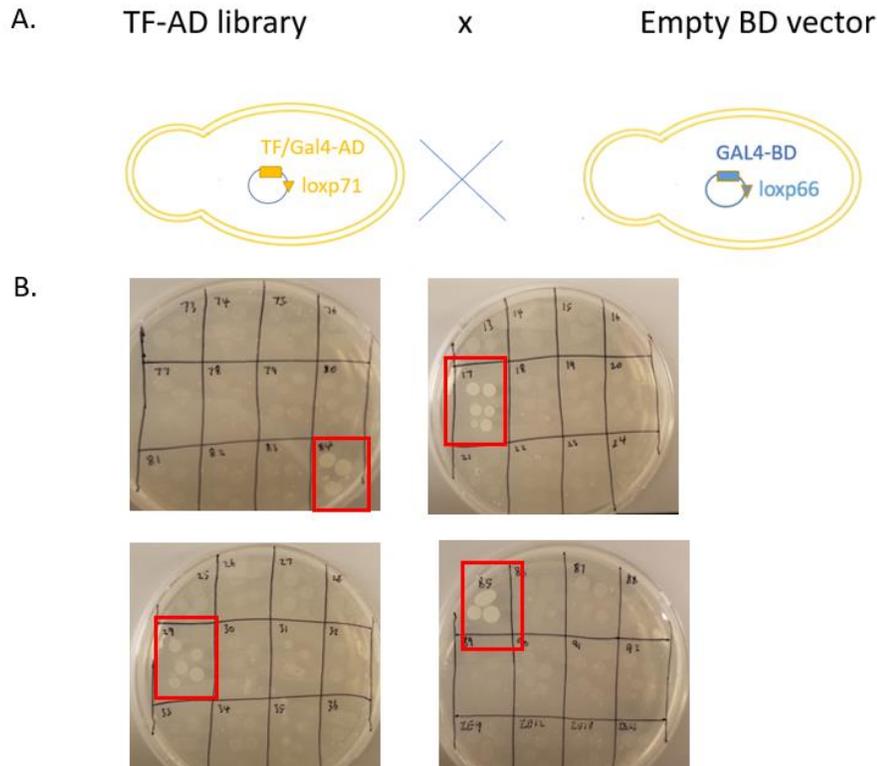
TF24_GFP-his_tag with native promoter was cloned into the pBR9-Ble expression vector and recombinant plasmid was transformed to cc1690 strain or TF24 mutant strain. CREB domain was deleted from TF24 and truncated TF24_GFP_his_tag with native promoter was cloned into the same vector and transformed to TF24 mutant strain.

Data-driven ontology construction for transcription factors

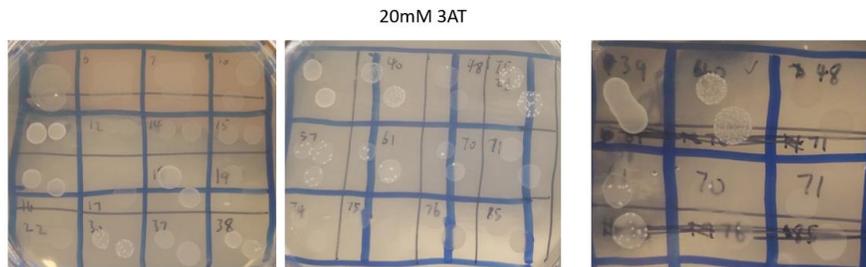
Coexpression is calculated using Pearson correlation from expression of 330 transcription factors under 294 conditions. From coexpression a fully connected networks is constructed with edge weights representing absolute coexpression values between nodes.

CliXO algorithm with $\alpha=0.04$ and $\beta=0.4$ is applied on the coexpression network for hierarchical community detection. The resultant communities in the hierarchy are labeled using gProfiler with `organism='creinhardtii'` and p-value threshold of 0.00001. Within gProfiler, we calculate Jaccard index (minimum=0.05) and sort the mapped annotations first descending by Jaccard index and then ascendingly by p-value. The top hit is then utilized as label for each community.

Supplementary figures



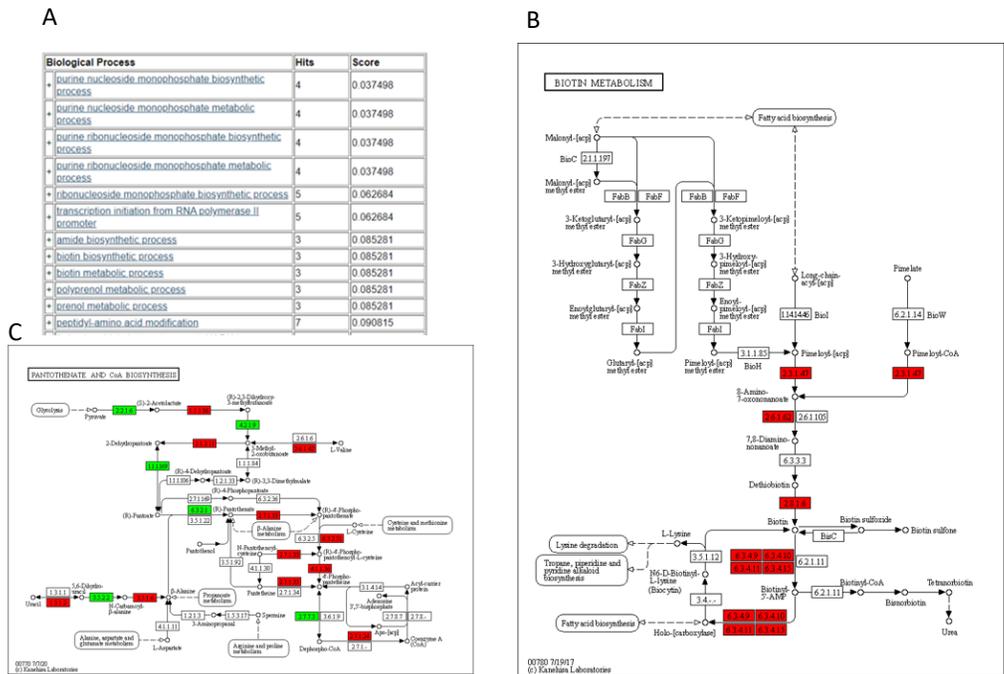
Supplementary Figure 2.1. Control assay for TF-AD Y8800 strain mating with CRY8930 strain transformed with empty BD vector. **A.** Diagram of yeast mating between TF-AD Y8800 strain and CRY8930 strain transformed with empty pBDlox plasmid. **B.** Only four TF-AD strains can generate false positive colonies on SC Trp-/Leu-/His- plate with 1mM 3AT when mating with control CRY8930 strain.



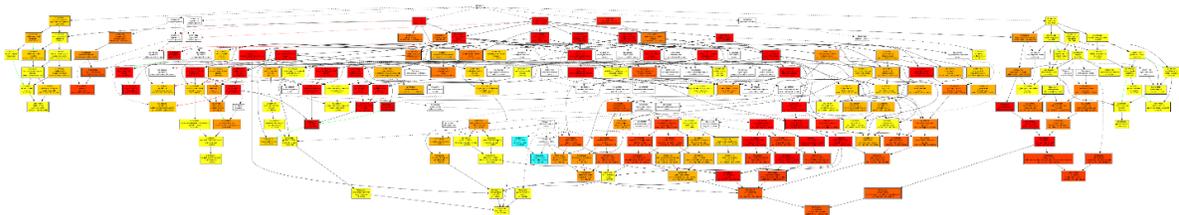
Supplementary Figure 2.2. Control assay for TF-BD strain mating with strain transformed with empty AD vector. 15 TF-BD strains can generate false positive colonies on SC Trp-/Leu-/His- plate with 20mM 3AT when mating with control Y8800 strain.

Ash Free Biomass	tf20m	tf20m N-	tf24m	tf24m N-	4533 wt	4533 wt N-
% AFBM	81.3	90.0	89.3	92.2	86.2	91.4
% Moisture	2.6	3.8	1.9	2.4	2.7	0.6
% Ash	16.1	6.2	8.7	5.4	11.1	8.0

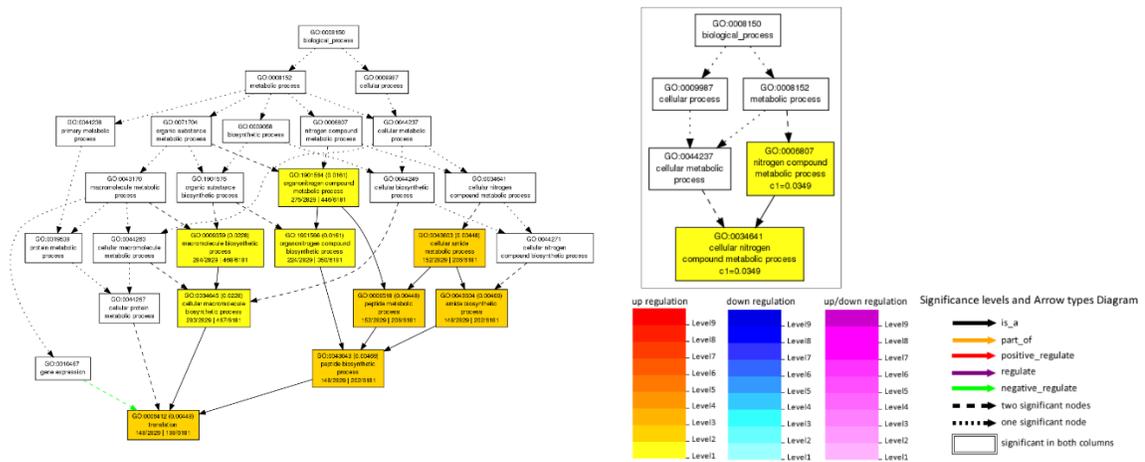
Supplementary Figure 2.3. Ash free biomass composition for lipid profile from Mass spec data for normalization



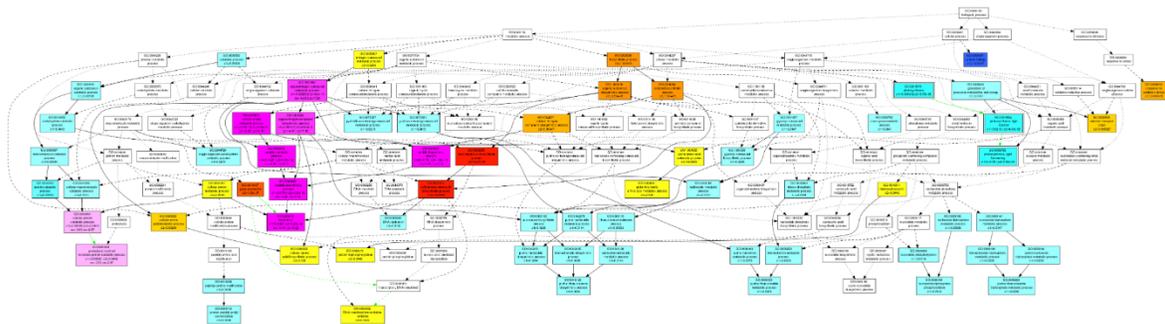
Supplementary Figure 2.4 Unique pathways only upregulated in tf20m 48hours after nitrogen starvation. **A.** GO enrichment of differential KEGG pathway according to Biological Process. **B.** Upregulated Biotin metabolism KEGG pathway. **C.** Upregulated Pantothenate and CoA biosynthesis uniquely upregulated in tf20m strain 48 hours after nitrogen depletion. Genes labeled as red are upregulated genes while genes labeled as white are genes not present in *C.reinhardtii*.

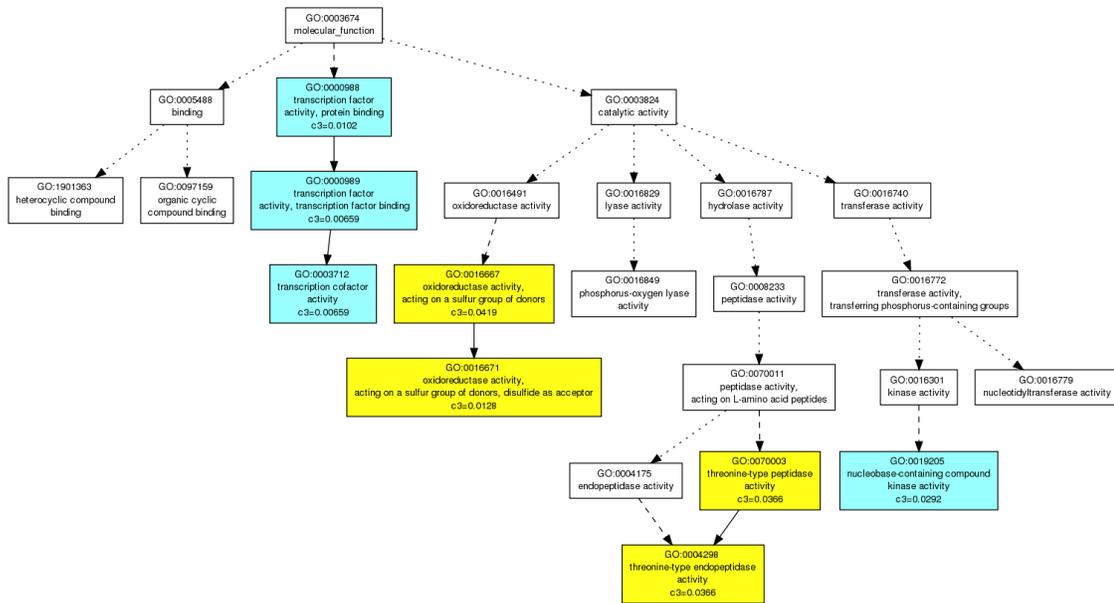


Supplementary Figure 2.5 Time course PAGE analysis for GO enrichment in wt strain. Differential GO enrichment in wt 48h versus 0h after Nitrogen starvation.

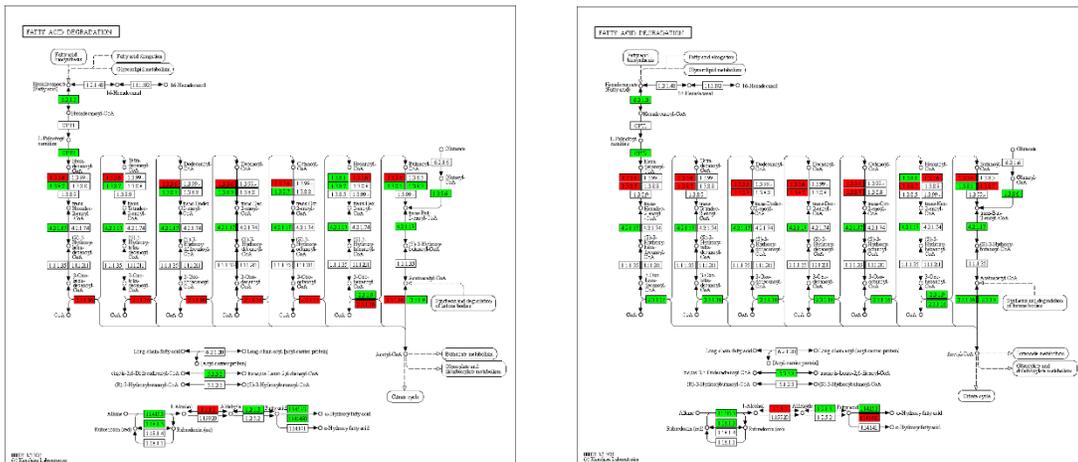


Supplementary Figure 2.6 GO enrichment (according to Biological process) in tf20m strain 4 hours versus 0 hour after Nitrogen starvation. Ribosome and translation related GO is highly upregulated in tf20m 4hours after nitrogen starvation

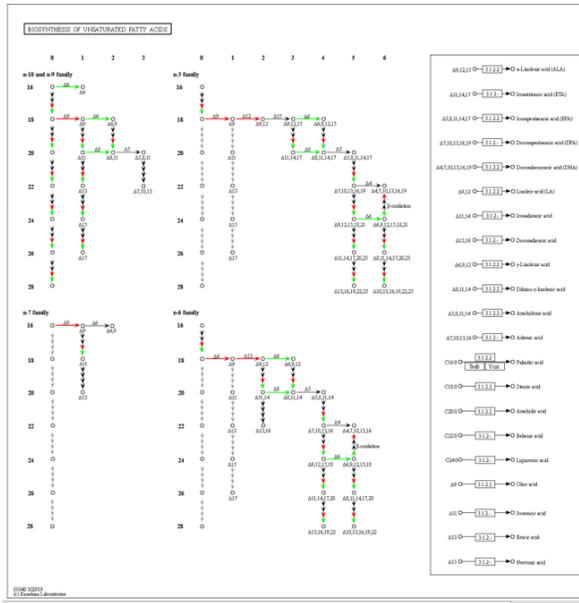




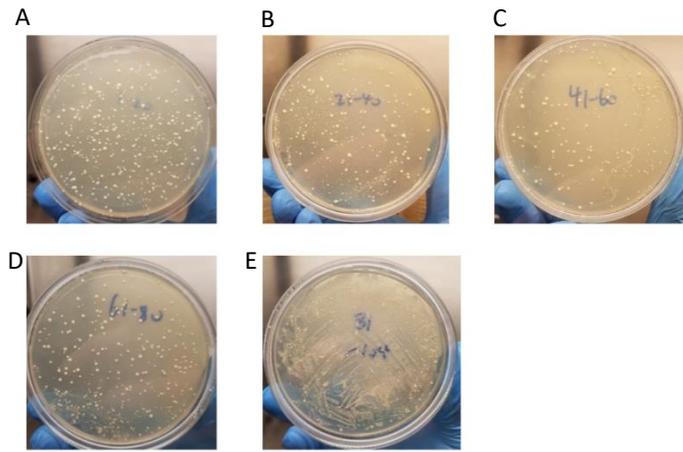
Supplementary Figure 2.10 Unique pattern of changed pathway enriched in tf24m from PAGE analysis of time course RNAseq data for all three strains. Transcription factors and nucleobase-containing compound kinase activity were downregulated while Oxidoreductase activity and threonine-type endopeptidase were upregulated specifically in tf24m strain 48 hours after nitrogen starvation.



Supplementary Figure 2.11 Uniquely downregulated genes in fatty acid metabolism in tf24m and tf20m strains. Red color indicates downregulation and green color represent presence of gene ortholog in *C. reinhardtii*.



Supplementary Figure 2.12 Unique pathway downregulated in tf20m strain 48 hours after Nitrogen starvation. Red color indicates upregulation and green color represent presence of gene ortholog in *C.reinhardtii*.



Supplementary Figure 2.13 Positive colonies obtained from different subgroups of TF-AD library mating with TF-BD library.

Acknowledgement

Chapter 2, in part is currently being prepared for submission for publication of the material. Bijie Ren, Song Cao, Stephen Mayfield. Discovery of two master regulators for lipid accumulation *Chlamydomonas reinhardtii* through Yeast Two-Hybrid sequencing. The dissertation author was the primary investigator and author of this paper.

This publication includes data generated at the UC San Diego IGM Genomics Center utilizing an Illumina NovaSeq 6000 that was purchased with funding from a National Institutes of Health SIG grant (#S10 OD026929). I would like to acknowledge my undergrad students Song Cao, Yiqun Jiang, Nan Hu, Ashley Dung, Stephanie Domes and all members of the Mayfield lab for their nice help throughout my Ph.D program.

References

- Miller Tran, Bin Zhou, Pär L Pettersson, Maria J Gonzalez, Stephen P Mayfield. Synthesis and assembly of a full-length human monoclonal antibody in algal chloroplasts. *Biotechnol Bioeng.* 2009 Nov 1;104(4):663-73 2.
- Specht, E., Miyake-Stoner, S. & Mayfield, S. Micro-algae come of age as a platform for recombinant protein production. *Biotechnol Lett* 32, 1373–1383 (2010).
- Rasala, B.A., Lee, P.A., Shen, Z., Briggs, S.P., Mendez, M., Mayfield, S.P., (2012) Robust expression and secretion of Xylanase1 in *Chlamydomonas reinhardtii* by fusion to a selection gene and processing with the FMDV 2A peptide. *PLoS One* 7, e43349.
- Gregory, J.A., Topol, A.B., Doerner, D.Z., Mayfield, S., (2013) Alga-produced cholera toxin-Pfs25 fusion proteins as oral vaccines. *Appl Environ Microbiol* 79, 3917-3925.
- Ravi, A., Guo, S., Rasala, B., Tran, M., Mayfield, S., Nikolov, Z.L., (2018) Separation options for phosphorylated osteopontin from transgenic microalgae *Chlamydomonas reinhardtii*. *Int J Mol Sci* 19.
- Anthony J Berndt; Tressa N Smalley; Bijie Ren; Amr Badary; Ashley Sproles; Francis Fields; Yasin Torres-Tiji; Vanessa Heredia; Stephen P Mayfield. Recombinant production of a functional SARS-CoV-2 spike receptor binding domain in the green algae *Chlamydomonas reinhardtii*. *bioRxiv* | ID: ppbiorxiv-428890.
- Long, T., Brady, S., et al. Systems Approaches to Identifying Gene Regulatory Networks in Plants. *Annu Rev Cell Dev Biol.* 2008; 24: 81–103.
- Bemer, M., et al. Cross-Family Transcription Factor Interactions: An Additional Layer of Gene Regulation. *Cell Trends Plant Sci.* <https://doi.org/10.1016/j.tplants.2016.10.007>
- Papachristou, E.K., Kishore, K., Holding, A.N. et al. A quantitative mass spectrometry-based approach to monitor the dynamics of endogenous chromatin-associated protein complexes. *Nat Commun* 9, 2311 (2018).
- Brückner, A., et al. Yeast Two-Hybrid, a Powerful Tool for Systems Biology. *Int J Mol Sci.* 2009 Jun; 10(6): 2763–2788.
- Trigg, S., Garza, R., MacWilliams, A. et al. CrY2H-seq: a massively multiplexed assay for deep-coverage interactome mapping. *Nat Methods* 14, 819–825 (2017).

- Kühn R., M. Torres R. (2002) Cre/loxP Recombination System and Gene Targeting. In: Clarke A.R. (eds) Transgenesis Techniques. Methods in Molecular Biology, vol 180. Springer, Totowa, NJ. <https://doi.org/10.1385/1-59259-178-7:175>
- Tian F, Yang DC, Meng YQ, Jin JP and Gao G. (2019). PlantRegMap: charting functional regulatory maps in plants. Nucleic Acids Research, gkz1020
- Guo AY, Chen X, Gao G, Zhang H, Zhu QH, Liu XC, Zhong YF, Gu X, He K, Luo J. PlantTFDB: a comprehensive plant transcription factor database. Nucleic Acids Res. 2008 Jan;36(Database issue):D966-9. doi: 10.1093/nar/gkm841. Epub 2007 Oct 12. PMID: 17933783; PMCID: PMC2238823.
- Zhang H, Jin J, Tang L, Zhao Y, Gu X, Gao G, Luo J. PlantTFDB 2.0: update and improvement of the comprehensive plant transcription factor database. Nucleic Acids Res. 2011 Jan;39(Database issue):D1114-7. doi: 10.1093/nar/gkq1141. Epub 2010 Nov 18. PMID: 21097470; PMCID: PMC3013715.
- Jin J, Zhang H, Kong L, Gao G, Luo J. PlantTFDB 3.0: a portal for the functional and evolutionary study of plant transcription factors. Nucleic Acids Res. 2014 Jan;42(Database issue):D1182-7. doi: 10.1093/nar/gkt1016. Epub 2013 Oct 29. PMID: 24174544; PMCID: PMC3965000.
- Nur A. Hidayati, Yui Yamada-Oshima, Masako Iwai, Takashi Yamano, Masataka Kajikawa, Nozomu Sakurai4, Kunihiko Suda, Kanami Sesoko, Koichi Hori, Takeshi Obayashi, Mie Shimojima, Hideya Fukuzawa and Hiroyuki Ohta. Lipid remodeling regulator 1 (LRL1) is differently involved in. The Plant Journal (2019) 100, 610–626
- Goodenough U, Blaby I, Casero D, Gallaher SD, Goodson C, Johnson S, Lee JH, Merchant SS, Pellegrini M, Roth R, Rusch J, Singh M, Umen JG, Weiss TL, Wulan T. The path to triacylglyceride obesity in the sta6 strain of *Chlamydomonas reinhardtii*. Eukaryot Cell. 2014 May;13(5):591-613. doi: 10.1128/EC.00013-14. Epub 2014 Feb 28. PMID: 24585881; PMCID: PMC4060482.
- Fernández E, Matagne RF. In vivo complementation analysis of nitrate reductase-deficient mutants in *Chlamydomonas reinhardtii*. Curr Genet. 1986;10(5):397-403. doi: 10.1007/BF00418413. PMID: 3442823.
- Camargo A, Llamas A, Schnell RA, Higuera JJ, González-Ballester D, Lefebvre PA, Fernández E, Galván A. Nitrate signaling by the regulatory gene NIT2 in *Chlamydomonas*. Plant Cell. 2007 Nov;19(11):3491-503. doi: 10.1105/tpc.106.045922. Epub 2007 Nov 16. PMID: 18024571; PMCID: PMC2174885.
- Boyle NR, Page MD, Liu B, Blaby IK, Casero D, Kropat J, Cokus SJ, Hong-Hermesdorf A, Shaw J, Karpowicz SJ, Gallaher SD, Johnson S, Benning C, Pellegrini M, Grossman A, Merchant SS. Three acyltransferases and nitrogen-responsive regulator are implicated in nitrogen starvation-induced triacylglycerol accumulation in *Chlamydomonas*. J Biol Chem. 2012 May 4;287(19):15811-25. doi: 10.1074/jbc.M111.334052. Epub 2012 Mar 8. PMID: 22403401; PMCID: PMC3346115.
- Romero L.C., Galván F., Vega J.M. (1987). Purification and properties of the siroheme-containing ferredoxin-nitrite reductase from *Chlamydomonas reinhardtii*. Biochim. Biophys. Acta 914: 55–63
- Fernández E, Schnell R, Ranum LP, Hussey SC, Silflow CD, Lefebvre PA. Isolation and characterization of the nitrate reductase structural gene of *Chlamydomonas reinhardtii*. Proc Natl Acad Sci U S A. 1989 Sep;86(17):6449-53. doi: 10.1073/pnas.86.17.6449. PMID: 2475871; PMCID: PMC297861.
- Quesada A, Gómez I, Fernández E. Clustering of the nitrite reductase gene and a light-regulated gene with nitrate assimilation loci in *Chlamydomonas reinhardtii*. Planta. 1998 Oct;206(2):259-65. doi: 10.1007/s004250050398. PMID: 9737004.
- Fernandez E, Galvan A. Nitrate assimilation in *Chlamydomonas*. Eukaryot Cell. 2008 Apr;7(4):555-9. doi: 10.1128/EC.00431-07. Epub 2008 Feb 29. PMID: 18310352; PMCID: PMC2292633.

- Hirasawa M, Tripathy JN, Sommer F, Somasundaram R, Chung JS, Nestander M, Kruthiventi M, Zabet-Moghaddam M, Johnson MK, Merchant SS, Allen JP, Knaff DB. Enzymatic properties of the ferredoxin-dependent nitrite reductase from *Chlamydomonas reinhardtii*. Evidence for hydroxylamine as a late intermediate in ammonium production. *Photosynth Res.* 2010 Feb;103(2):67-77. doi: 10.1007/s11120-009-9512-5. Epub 2009 Dec 29. PMID: 20039132; PMCID: PMC2833264.
- Fernández, E., Llamas, A. and Galván, A. (2009) Chapter 3 - Nitrogen assimilation and its regulation. In *The Chlamydomonas Sourcebook*, 2nd edn. (Harris, E.H., Stern, D.B. and Witman, G.B., eds). London: Academic Press, pp. 69–113.
- Schmollinger S, Mühlhaus T, Boyle NR, Blaby IK, Casero D, Mettler T, Moseley JL, Kropat J, Sommer F, Strenkert D, Hemme D, Pellegrini M, Grossman AR, Stitt M, Schroda M, Merchant SS. Nitrogen-Sparing Mechanisms in *Chlamydomonas* Affect the Transcriptome, the Proteome, and Photosynthetic Metabolism. *Plant Cell.* 2014 Apr;26(4):1410-1435. doi: 10.1105/tpc.113.122523. Epub 2014 Apr 18. PMID: 24748044; PMCID: PMC4036562.
- Park JJ, Wang H, Gargouri M, Deshpande RR, Skepper JN, Holguin FO, Juergens MT, Shachar-Hill Y, Hicks LM, Gang DR. The response of *Chlamydomonas reinhardtii* to nitrogen deprivation: a systems biology analysis. *Plant J.* 2015 Feb;81(4):611-24. doi: 10.1111/tpj.12747. PMID: 25515814.
- Nunes-Nesi A, Fernie AR, Stitt M. Metabolic and signaling aspects underpinning the regulation of plant carbon nitrogen interactions. *Mol Plant.* 2010 Nov;3(6):973-96. doi: 10.1093/mp/ssq049. Epub 2010 Oct 6. PMID: 20926550.
- López García de Lomana A, Schäuble S, Valenzuela J, Imam S, Carter W, Bilgin DD, Yohn CB, Turkarslan S, Reiss DJ, Orellana MV, Price ND, Baliga NS. Transcriptional program for nitrogen starvation-induced lipid accumulation in *Chlamydomonas reinhardtii*. *Biotechnol Biofuels.* 2015 Dec 2;8:207. doi: 10.1186/s13068-015-0391-z. PMID: 26633994; PMCID: PMC4667458.
- Siersma PW, Chiang KS. Conservation and degradation of cytoplasmic and chloroplast ribosomes in *Chlamydomonas reinhardtii*. *J Mol Biol.* 1971 May 28;58(1):167-85. doi: 10.1016/0022-2836(71)90239-7. PMID: 5088925.
- Martin NC, Goodenough UW. Gametic differentiation in *Chlamydomonas reinhardtii*. I. Production of gametes and their fine structure. *J Cell Biol.* 1975 Dec;67(3):587-605. doi: 10.1083/jcb.67.3.587. PMID: 1202015; PMCID: PMC2111648.
- Chen H, Hu J, Qiao Y, Chen W, Rong J, Zhang Y, He C, Wang Q. Ca(2+)-regulated cyclic electron flow supplies ATP for nitrogen starvation-induced lipid biosynthesis in green alga. *Sci Rep.* 2015 Oct 9;5:15117. doi: 10.1038/srep15117. PMID: 26450399; PMCID: PMC4598854.
- Blaby IK, Glaesener AG, Mettler T, Fitz-Gibbon ST, Gallaher SD, Liu B, Boyle NR, Kropat J, Stitt M, Johnson S, Benning C, Pellegrini M, Casero D, Merchant SS. Systems-level analysis of nitrogen starvation-induced modifications of carbon metabolism in a *Chlamydomonas reinhardtii* starchless mutant. *Plant Cell.* 2013 Nov;25(11):4305-23. doi: 10.1105/tpc.113.117580. Epub 2013 Nov 26. PMID: 24280389; PMCID: PMC3875720.
- Lopez D, Casero D, Cokus SJ, Merchant SS, Pellegrini M. Algal Functional Annotation Tool: a web-based analysis suite to functionally interpret large gene lists using integrated annotation and expression data. *BMC Bioinformatics.* 2011 Jul 12;12:282. doi: 10.1186/1471-2105-12-282. PMID: 21749710; PMCID: PMC3144025.
- Tian T, Liu Y, Yan H, You Q, Yi X, Du Z, Xu W, Su Z. agriGO v2.0: a GO analysis toolkit for the agricultural community, 2017 update. *Nucleic Acids Res.* 2017 Jul 3;45(W1):W122-W129. doi: 10.1093/nar/gkx382. PMID: 28472432; PMCID: PMC5793732.

Chapter 3: Genome scale reconstruction of Metabolic-Protein secretion modeling for *Chlamydomonas reinhardtii*

Genome scale reconstruction of Metabolic-Protein secretion modeling for *Chlamydomonas reinhardtii*

Bijie Ren, Terence Hwa, Jahir Gutierrez, Frank Fields, Nathan Lewis, Stephen Mayfield

Abstract

Eukaryotic algae have been developed as a novel platform for the expression of therapeutic antibodies and other complex proteins. However, little information is available for this platform in terms of protein translocation, post-translation modification (PTM) or secretion. In this manuscript, we integrated systematic pathway for protein production, post-translation modification and protein secretion into the current metabolic modeling of this green alga: iCre1355. This refined reconstruction enables us to calculate energy cost of any protein under Autotrophic/Mixotrophic/Heterotrophic condition and shows that *Chlamydomonas* has different growth capacity and maximal protein secretion yield under different condition. Moreover, there is an anti-correlation between protein yield and growth capacity, which are confirmed by our independent biophysical model and experimental data. In addition, this reconstruction model also indicates that if we knock out unnecessary subsystem in the cell, the free energy saved could be used in other subsystems, and lead to faster growth rate and higher protein secretion rate. We also identify that inhibition of cAMP degradation will increase the useful biolipid accumulation and boost protein of interest secretion, which provides a new strategy to utilize algae as a cost-effective cell factory for biofuel and therapeutic antibody or cytokine as potential coproducts.

Introduction

Global warming has a grave impact on our ecosystem due to increasing levels of atmospheric and oceanic CO₂. In addition, billions of petroleum-based plastic products have been produced every year in the world, and the vast majority of these cannot be recycled or reused and become an ecological burden to our planet. Microalgae is an efficient absorber for carbon source and was the major source for the production of petroleum, so it clearly has the potential to help solve both global warming issue and plastic pollution on a planetary scale. However, little is known about the molecular regulatory networks of microalgae and the cost of algae-based biofuel or bio-degradable plastic is not cheap enough to compete with petroleum and petroleum-based materials. Recently, *Chlamydomonas reinhardtii* has been harnessed for the production of therapeutic proteins and monoclonal antibodies. To help develop microalgae as an efficient production platform, we tried to understand how we might secrete high value proteins, and

simultaneously accumulate biolipid precursors inside cell, so that these co-products could act synergistically to drive the overall economics of algae as a production platform.

Genome-scale of metabolic models has guided scientists to analyze and redirect metabolic flux to product of interest (Alper et al., 2005; Park et al., 2007; Milne et al., 2009; Gutierrez et al., 2020). The *Chlamydomonas reinhardtii* iCre1355 model showed ability to predict growth rate and lipid accumulation (Imam et al., 2015). However, no models accounting for protein production and secretion pathway has been built for microalgae, thus limiting the ability for model-guided engineering of biotherapeutics production. However, with more sequencing data available for *Chlamydomonas* genome and transcriptome, annotated genes and peptides have increased to enable the construction of such a model for *Chlamydomonas reinhardtii*.

In this study, we developed a network reconstruction detailing the genes and mechanisms involved in protein production, translocation, PTM, and secretion in *Chlamydomonas reinhardtii*, and integrated this into the current genome-scale metabolic model iCre1355 (**Figure 3.1**). We used this model, iCre1477, to calculate energy cost of proteome synthesis and to predict how metabolic flux can be redirected into products of interest, such as recombinant therapeutic cytokine. Similar modeling of protein synthesis in CHO cells has shown how to increase yield of secreted antibodies by cell engineering (Gutierrez et al., 2020; Kol et al., 2020). Our model also indicates the similar relationships between protein yield and growth capacity found in *E. coli* (You et al., 2013; Basan et al., 2015). More interestingly, we found that green algae have differences in growth capacity and maximal protein yield under heterotrophic, mixotrophic or autotrophic condition. We predicted and experimentally validated that protein yield and growth rate are anti-correlated for these three conditions. Such a phenomenon was predicted by our top-down toy model as well. In addition, we showed that if we deleted “unnecessary subsystem” in *Chlamydomonas*, energy equivalents and other resources are liberated and can be redirected to other pathways, including protein yield. Finally, we show that inhibition of cAMP degradation via IBMX increased biolipid accumulation and protein yield without affecting growth rate. Such discovery of cAMP can be used in industry as a model to increase secretion of high value protein co-products, while increasing lipid accumulation, enabling significant reduction in the cost of algal biomass for biofuel or other lipid products.

Results

Reconstruction of algae secretory pathway and supplementation of metabolic pathway

A literature search on yeast, mouse, human and Arabidopsis allowed us to enumerate candidate proteins involved in protein production, ER translocation, folding, glycosylation, GPI biosynthesis, COPII/COPI transportation and other processes in the secretory pathway. Using Phytozome, blastp and Uniprot, we found algal homologs for each gene. From this we identified 122 candidate genes in *Chlamydomonas reinhardtii*, which we classified into different subsystems. Compared to the corresponding genes in yeast, algae cells contain more molecular chaperones for protein folding, consistent with previous work showing that algae exhibit superior ability to fold complicated proteins and antibodies (**Figure 3.2**)(Berlec et al., 2013; Spadiut et al., 2014; Tran et al., 2009). We found no other considerable difference between algae and yeast cell across the remaining subsystems. However, we note that we identified fewer candidate genes involved in ALP pathway in *Chlamydomonas reinhardtii*.

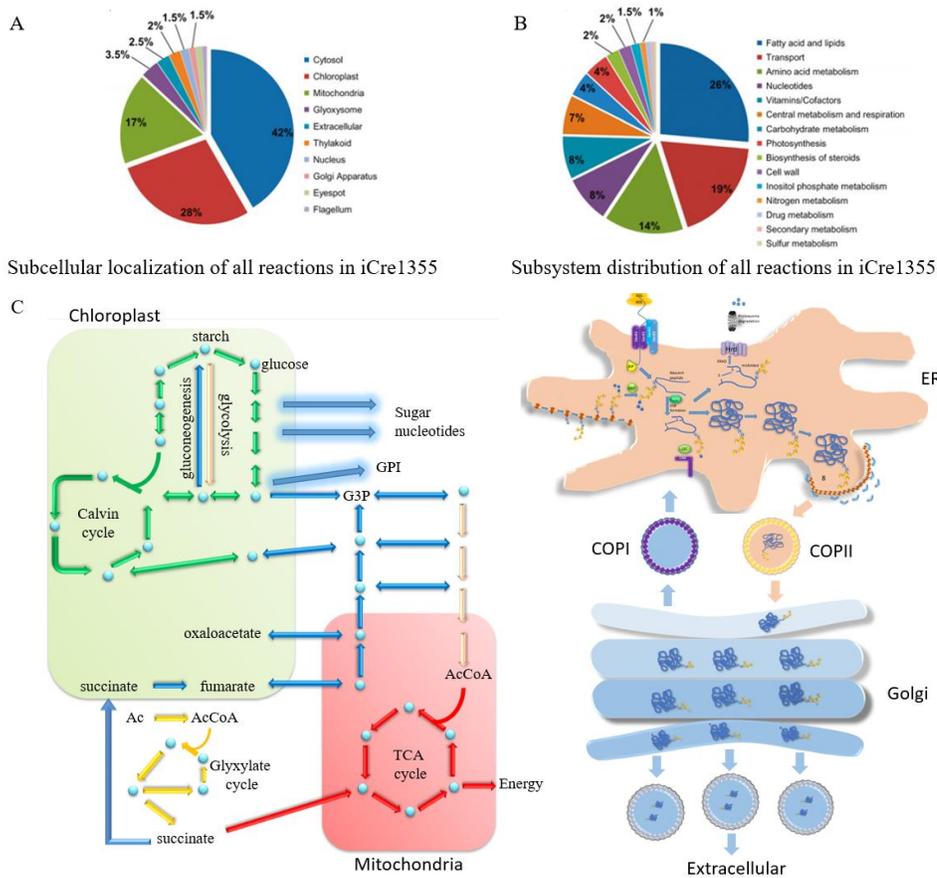


Figure 3.1 Distribution of subsystems and reactions in model iCre1355. **A.** Subcellular localization of all reactions in iCre1355. **B.** Subsystem distribution of all reactions in iCre1355. **C.** Integration of protein secretory pathway to genome-scale metabolic model iCre1355.

Each subsystem was added to the iCre1355 model separately and template reactions were defined. Based on the peptide information from Protein specific index matrix (PSIM) proposed in previous work (Feizi et al., 2013), we

classified all secretory pathway proteins in *Chlamydomonas* into different families and assigned a specific set of reactions to each family. Protein secretion cost were calculated and plotted versus published Riboseq data to investigate translation cost of each transcript (Chung et al., 2015). We also computed the energy cost for each secreted protein by calculating total cost from translation, translocation, vesicular transport and secretion. We found a negative correlation between ATP cost for secreted protein and ribosomal occupancy from the Ribo-Seq data (**Figure 3.2C**). This indicates that for proteins required high amount of energy equivalents are often expressed at lower levels, possibly to conserve resources. This is consistent with previously observations in human tissues and CHO cells, suggesting there is a fundamental mechanism for cells to regulate such energy homeostasis (**Supplementary Figure 3.1**).

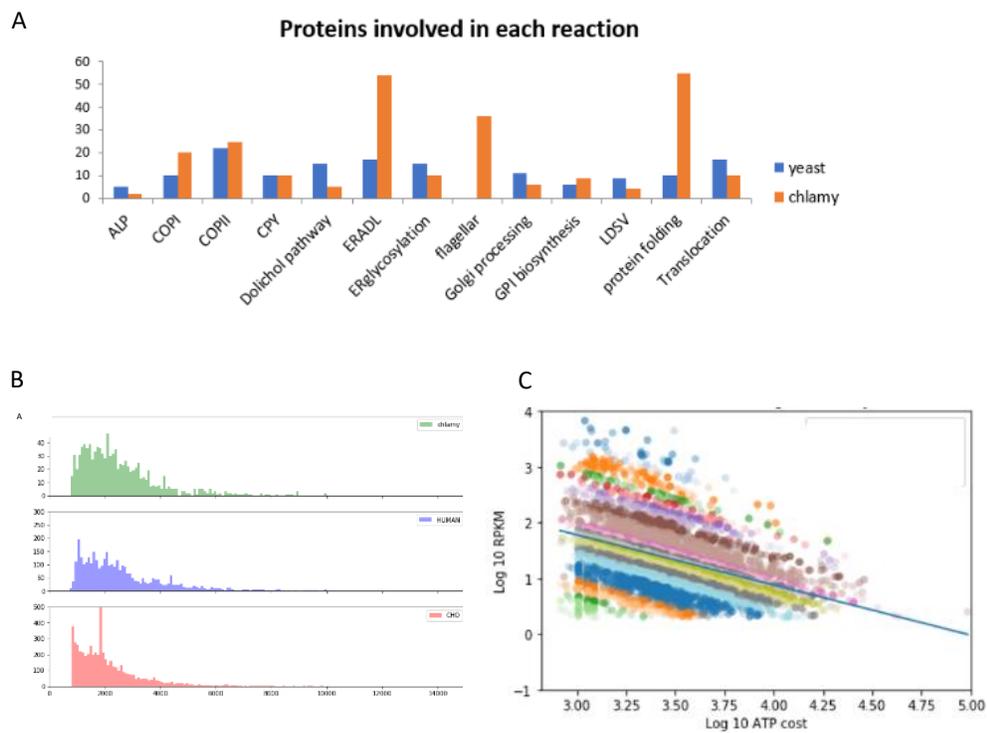


Figure 3.2 Genome scale reconstruction of protein secretion pathway. A. Total number of components in each subsystem from *Chlamydomonas reinhardtii* and yeast cell. **B.** Energy cost distribution of each protein in *Chlamydomonas reinhardtii*, human and CHO cells. **C.** Correlation between energy cost for each protein versus corresponding transcript abundance actively for translation from previously published Ribo-Seq data (Chung et al., 2015).

The model-predicted tradeoff between growth and protein secretion rate is observed experimentally under

To assess if the anti-correlation seen computation is observed experimentally, we overexpressed recombinant protein TGM in *Chlamydomonas reinhardtii* strain 1690 driven by constant AR1 Promoter (**Figure 3.2B**). This

transgenic strain was cultured under different phototrophic, mixotrophic and heterotrophic conditions to measure its growth rate and protein secretion yield. Growth rate and protein secretion rate of algae transgenic strain was monitored every three hours when cultured in CO₂ box and every 8 hours when cultured outside of CO₂ box (**Supplementary Figure 3.2**).

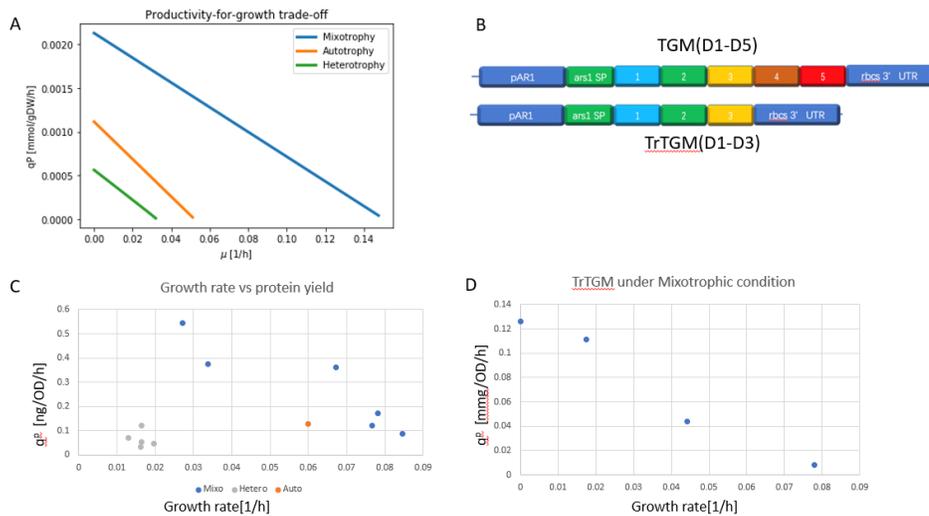


Figure 3.3 Anti-correlation between growth rate and protein secretion rate. Growth rate and proteins secretion rate has a broader range when algae cultured under mixotrophic condition than phototrophic condition. **A.** Growth vs protein productivity tradeoff curves were predicted for algae cell under Mixotrophy/Autotrophy/Heterotrophy condition. **B.** Expression of two recombinant proteins TGM and TrTGM with a constant promoter pAR1. A FLAG tag was added at the C terminal of each protein for western blot quantification. **C.** Dot plot of growth rate vs protein secretion rate for TGM protein. **D.** Dot plot of growth rate vs protein secretion rate of TrTGM. One raw data example of growth curve and time course western blot was shown in Supplementary Figure 3.2

Based on our genome-scale model prediction (**Figure 3A**), there is a difference between growth capacity and protein secretion yield when algae cells were incubated under heterotrophic, mixotrophic and autotrophic condition. Consistent with this prediction, our experimental data shows that algae under mixotrophic condition exhibited higher growth rate and TGM protein secretion yield (**Figure 3C, 3D**). Significant difference of growth rate and protein secretion rate was observed when cells were grown with same medium in or outside of CO₂ box(**Supplementary Figure 3.2**). To verify this discovery is not caused by integration of our transgene TGM expression cassette into the functional genes in genome, we also tested another transgenic strain expressing TrTGM. This strain also exhibited the

same anti-correlation between growth rate and protein secretion yield when cell growth is affected by different concentration of acetate or CO₂ (**Figure 3.3D**). Constant secretion rates of whole secretome during log phase were observed (**Supplementary figure 3.3A, B**). In addition, we measured chlorophyll, lipid content and total RNAs in two groups with different acetate concentration (1/4 of normal acetate(16.67mM) and 1/2 of normal acetate(16.67mM) used in TAP medium). But there were no differences of these macromolecules in these two groups (**Supplementary figure 3.3 C~G**).

To better investigate underlying mechanism behind this anti-correlation, we built a top-down physics model. If the cell is in the log phase and the given influx of amino acid is not in short, biomass accumulation depends on growth rate based on its definition (see equation 1 and 2). And the influx of amino acids J_a^{in} is equal to the protein synthesis rate (see equation 3) at steady state or log phase. Since influx of amino acids J_a^{in} can be defined by metabolic enzyme activity $k_a M_a$, where k_a is the enzymatic catalytic rate and M_a is total amount of metabolic enzymes for amino acid synthesis. Then given the constrained growth capacity $\Phi_{max} = \Phi_p + \Phi_a + \Phi_c$, where Φ_p is fraction of protein of interest over biomass and Φ_a , Φ_c is fraction of metabolic enzymes[for amino acids influx] and rest proteome over biomass. Based on these definitions, we can get equation $\Delta\Phi_p = \Phi_{max} - \Delta\Phi_c - \frac{\beta\lambda}{k_a}$, indicating anti-correlation between protein secretion and growth rate.

$$\begin{array}{l} \text{Biomass accumulation:} \\ \mu \text{ is growth rate} \end{array} \quad \frac{dN}{dt} = \mu N \quad (1)$$

$$\begin{array}{l} M \text{ is cell biomass} \end{array} \quad \frac{dM}{dt} = \mu M \quad (2)$$

$$\begin{array}{l} \text{Free amino acid pool:} \end{array} \quad \frac{da}{dt} = J_a^{in} - \beta \frac{dM}{dt} \quad (3)$$

β is fraction of total proteins over cell biomass

This discovery in eukaryotic algae cell is also consistent with previously published growth law in *E.coli* (Klump, Cell 2010). Combined with **Figure 2C, 2D, Figure 3** and **Supplementary Figure 2**, protein expression and secretion in the cell is an energetic activity, which makes algae cell suppress more expensive proteins to maintain their growth rate. This is supported by the fact that overexpression of recombinant protein reduces the cell density and growth rate (**Supplementary Figure 3.4**). And the reduced growth rate is dependent on fraction of recombinant protein over total biomass based on the above model, if our transgene was not integrated to the essential genes for cell growth. In addition, there is no significant difference in terms of chlorophyll, bulk RNA and lipid profile level between

TGM strain and wt strain (**Supplementary Figure 3.5**). However, a mild decrease of lipid content and a slightly increased bulk RNA was detected in TGM strain when compared to wt strain (**Supplementary Figure 3.5D,E**).

Nitrogen-dependent protein secretion rate and energy-dependent growth rate

To investigate the effect of influx of amino acids J_a^{in} , we tested algae growth with different nitrogen source and concentrations, and monitored corresponding protein secretion yield. The limited nitrogen and saturated nitrogen condition was defined by the lipid accumulation stained by Nile red for each condition (**Supplementary Figure 3.6**) (Boyle et al., 2012).

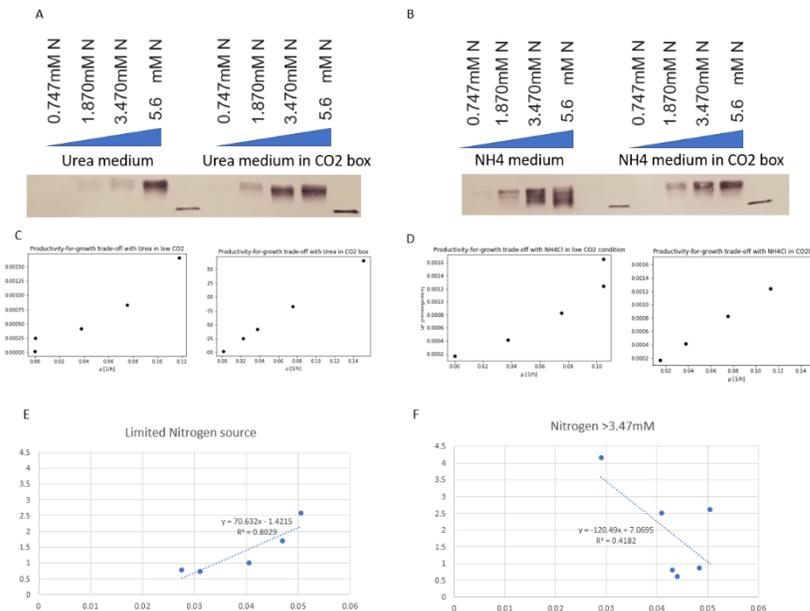


Figure 3.4 Protein secretion rate is dependent on Nitrogen when Nitrogen is limited. **A.** Different concentration of Urea was added to TAP N- medium for algae growth in CO₂ box (0.5% CO₂) or in normal lab condition under constant light. **B.** Different concentration of NH₄Cl was added to TAP N-free medium for algae grown in CO₂ box or in normal lab condition under constant light. 45-KDa Recombinant Posi-Tag Epitope Tag Protein (BioLegend, Cat #931301) was loaded as a FLAG standard in (**A,B**). TGM protein was detected with Monoclonal ANTI-FLAG® M2-Alkaline Phosphatase antibody (Sigma-Aldrich, St. Louis, MO, USA, Cat #A9469; final dilution 1:5000). **C.** Prediction of growth rate and protein secretion rate with low concentration of Urea. **D.** Model prediction of growth rate and protein secretion rate with low concentration of NH₄Cl. **E.** Positive correlation between protein secretion rate and growth rate with limited nitrogen. **F.** Anti-correlation between protein secretion rate and growth rate with saturated nitrogen (Nitrogen > 3.4mM)

With increasing of nitrogen (Urea or ammonium chloride) concentration, the protein secretion rate was increased as well (**Figure 4A, Figure 4b, Supplementary 3.7**). Our flux model also generated similar positive correlation between growth rate and protein secretion yield when nitrogen is limited (**Figure 3.4C and Figure 3.4D**).

However, the growth rate and protein secretion rate became anti-correlated when nitrogen was no longer a limit, which is line with the previous discovery in *E.coli* (You et al. 2013).

This is also predicted by our genome-scale modeling when nitrogen is not limited, protein secretion yield is lower in CO₂ box(0.5% CO₂) than that in low atmospheric CO₂ concentration(0.03% CO₂). However, if the nitrogen is limited, protein secretion yield is instead higher in CO₂ box (**Supplementary Figure 3.8**). To further investigate change of major macromolecules in different conditions, we measured chlorophyll and bulk RNA in groups with different nitrogen sources and concentrations (**Supplementary Figure 3.9**). Surprisingly, chlorophyll was constantly degrading in cells from all nitrogen limited groups and bulk RNA in TAP N-free medium with 0.747mM Urea was severely degraded, resembling the same phenomenon that chlorophyll, RNA and ribosome degradation in algal cells under Nitrogen starvation (Schmollinger et al., 2014).

Another interesting thing we found is that protein secretion rate drops significantly when nitrogen source is Urea. One hypothesis is that conversion of each molecule of Urea to Ammonium, which then will be served as a direct substrate for amino acid, requires two ATP energy cost. Based on the RNAseq result, we can clearly see that catabolic pathway were dramatically increased when algae incubated with Urea instead of Ammonium while anabolic pathway were decreased (**Supplementary Figure 3.10**). This response enabled cells to generate more energy to convert Urea to ammonium for the essential cell activities.

Universal toy model to explain correlation between Protein synthesis rate and cell growth rate

In order to explain the discrepancy of correlation between protein synthesis rate and cell growth rate under different nutrient condition, we further introduced the growth equation used in the previous paper (You et al., 2013) and did some modifications so that we could use that to explain not only protein synthesis change, but also explain the change of Phosphate-abundant PolyP body and Carbon-rich macromolecules, such as starch and lipid, under different nutrient limitations found in previous studies (Park et al., 2014; Boyle et al., 2012; Aksoy et al).

The key assumption is that the whole cell capacity for energy production and the theoretical maximal biomass is constrained, which means there is no other activated pathway to generate extra energy or precursors under each specific condition. If we could divide the maximal biomass to different sectors, and each sector is mainly built from one major nutrient source while there's a constant sector independent of growth rate(Equation 4). Here $\Delta\Phi_c$ is the growth-rate dependent Carbon sector defined by difference between the carbon sector minus the minimal constant

carbon sector. And the minimal constant carbon sector is growth rate independent, the same parameter introduced in previous work (You et al., 2013). With the similar definition, we can get nitrogen sector $\Delta\Phi_N$, and biomass sector $\Delta\Phi_{BM}$ without Carbon atom or Nitrogen atom. And sector $\Delta\Phi_N$ is mainly built by nitrogen atom (Equation 5). For nitrogen limitation we tested in this manuscript, other nutrients such as Carbon source acetate or sulfate source is sufficient for nitrogen sector and the Michaelis-Menten equation of carbon and sulfate is thus constant in Equation 5. k_N is the maximal catalytic rate for Nitrogen sector and $f_N(C, N, S)$ is the integrated Michaelis-Menten function of different nutrient sources. In addition, growth rate λ can be defined by the Monod equation to illustrate the relationship between growth rate and each nutrient concentration (Equation 6). Given the sufficient carbon and sulfate concentration in the nitrogen limitation assay, the growth rate is only a function of Nitrogen source. Based on the previous derivation, as cells are in the steady state during the log phase, synthesis rate of nitrogen sector is equal to the influx of amino acids (Equation 7). Combined with equation 5 and equation 6 under nitrogen limitation condition while other nutrients are saturated, we can get the linear dependence of growth rate for the nitrogen sector shown in equation 8. v_a is maximal catalytic rate for amino acid consumption, α and β is the simplified form of $v_a \frac{[aa]}{[aa]+K_a}$ to indicate the flux function of nitrogen and nutrient availability (You et al., 2013). And such function can be replaced with equation 6 to get term v_{nN} , which represents the Nitrogen-independent term $\lambda_{max} \frac{C}{C+K_C} \frac{S}{S+K_S}$.

However, when nitrogen is sufficient to synthesize extra precursors and activate other pathways but there's a limited carbon availability in the medium, then equation 5 and 6 and their derivation v_{nN} are now dependent on carbon source, so correlation of growth rate and nitrogen sector in equation 8 is not linear any more. However, if we derive the carbon sector by repeating the previous procedure to get the linear dependence of growth rate and carbon sector under carbon limitation condition, which is $\Delta\Phi_c = \lambda \left(\frac{1}{v_{nC}}\right)M$, taken together with the equation 4, we can get the anticorrelation of nitrogen sector and growth rate in equation 9.

$$\text{Constraint capacity: } \Delta\Phi_{BM} + \Delta\Phi_c + \Delta\Phi_N = \Phi_{max} \quad (4)$$

$$\text{Each sector: } \frac{dM_N}{dt} = k_N * f_N(C, N, S) \quad (5)$$

$$\text{Monod equation: } \lambda = \lambda_{max} \frac{[aa]}{[aa]+K_a} \frac{C}{C+K_C} \frac{S}{S+K_S} \dots \quad (6)$$

When N is limited and not enough to support precursors for other pathways:

$$\frac{dM_{ma}}{dt} = J_a^{in} - \frac{dM_N}{dt} = 0 \text{ in log phase} \quad (7)$$

$$\frac{dM_N}{dt} = J_a^{in} = v_a \frac{[aa]}{[aa] + K_s} M_a = \alpha \lambda \beta M = \lambda \left(\frac{1}{v_{nN}} \right) M \quad (8)$$

When N is enough for synthesize extra precursors for other pathways, while growth rate is affected by carbon availability, v_{nN} will be a function of carbon :

$$\Delta \Phi_{BM} + \Delta \Phi_N = \Phi_{max} - \Delta \Phi_c = \Phi_{max} - \lambda * \left(\frac{1}{v_{nC}} \right) \quad (9)$$

Knock out unnecessary subsystem increased growth rate

The assumption of constrained growth capacity is used for both genome-scale metabolic modeling and top-down biophysics modeling and critical for the model prediction. Given the importance of this assumption, we want to further investigate if the growth capacity of algae cell is limited or constrained in reality.

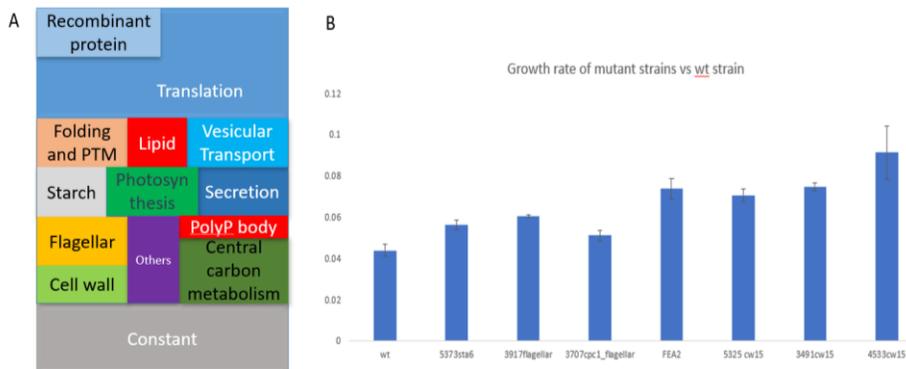


Figure 3.5. Growth rate of different mutant strains versus wild type strain. A. Schematic of constrained growth capacity under specific condition. **B.** Growth rate of each mutant cultured in TAP medium under constant light with atmospheric CO₂ (0.03%). Triplicate samples for each strain were measured.

To test this assumption, we cultured different mutant strains in the same TAP medium and monitor their growth rate. All these mutant strains grow faster than wild type strain (**Figure 3.5**). Cell wall mutant strains exhibit 50%~100% higher growth rate than wild type does while sta6 mutant strain exhibit only 30% increase in growth rate. Since phenotype of cell wall in normal TAP medium is not essential for the cell maintenance, removal of “unnecessary” cell wall significantly increase the cell growth rate. However, sta6 is a key gene in starch biosynthesis, which is important for cell activity, deletion of starch biosynthesis can only increase cell growth rate slightly. Consistent with that, FEA2 is the most abundant secreted protein in the medium and its knock-out could also release the burden on the cell. Interestingly, flagellar related genes are known to consume high amounts of energy for flagellar synthesis,

however, deletion of these genes doesn't increase growth rate a lot. Considering flagellar synthesis is a complicated subsystem including many flagellar-related genes and key hubs, one key hub seems not deplete the whole flagellar system and affect the cell growth rate too much.

Inhibition of cAMP degradation by IBMX boosts both protein secretion and lipid accumulation in algae cells.

It's interesting to find this anti-correlation between protein secretion and growth rate under carbon limit and positive correlation between them with limited nitrogen in the medium. And during nitrogen starvation, lipid accumulation was observed in algae cells (**Supplementary Figure 3.8**). Here we want to test the detailed mechanism underlying this correlation. It's reported that cAMP is a master regulator to coordinate bacterial proteome and metabolism (You et al., 2013) and its concentration is correlated with protein expression, we tested the effect of IBMX, which can inhibit degradation of cAMP by deactivate PDE protein, on the protein secretion and lipid accumulation (Choi et al., 2015).

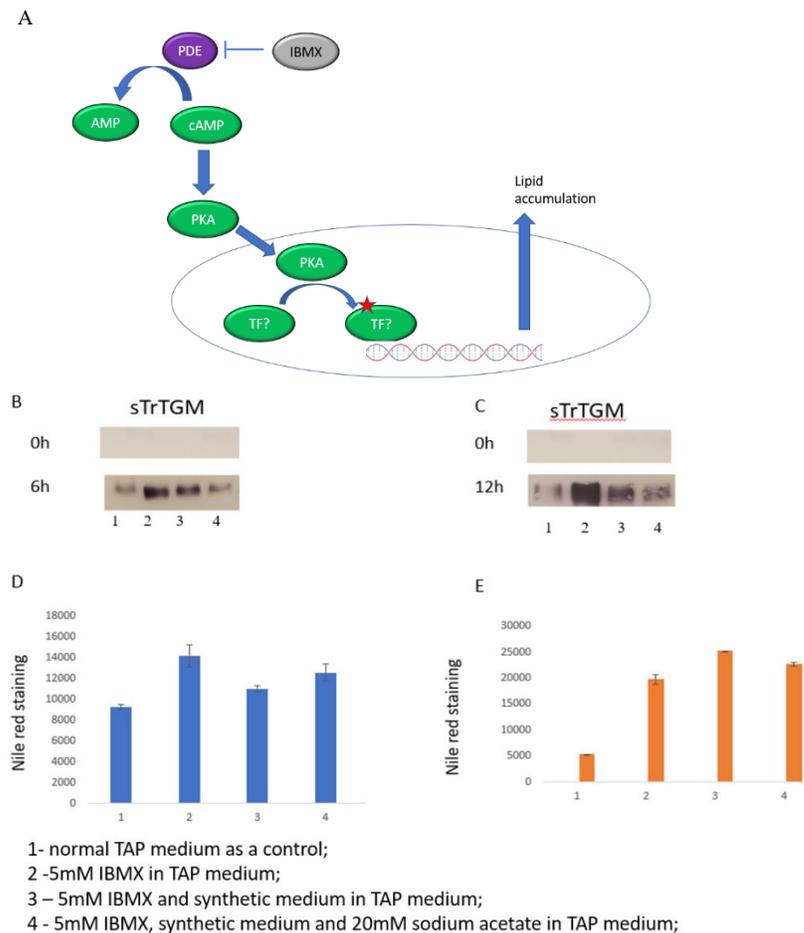


Figure 3.6 Time course Western blot of transgenic strain induced by IBMX. A. Diagram of effect of IBMX and possible downstream pathway to regulate lipid and proteome. Western blot(B) and Nile red staining(D) for cells 6 hours after different treatment. Lane 1 is normal TAP culture served as a control, Lane 2 is TAP culture supplemented with 5mM IBMX, Lane 3 is TAP culture supplemented with 5mM IBMX and 10mM synthetic amino acids medium, Lane 4 is TAP culture supplemented with 5mM IBMX, 10mM synthetic amino acids medium and extra 20mM sodium acetate. Western blot(C) and Nile red staining(E) of cells 12 hours after different treatment. 45-KDa Recombinant Posi-Tag Epitope Tag Protein(BioLegend, Cat #931301) was loaded as a FLAG standard. TGM protein was detected with Monoclonal ANTI-FLAG® M2-Alkaline Phosphatase antibody (Sigma-Aldrich, St. Louis, MO, USA, Cat #A9469; final dilution 1:5000). Nile red stainings for lipid profile per OD were measured by Tecan plate reader with an Excitation wavelength 550nm and Emission wavelength 640nm. 200ul of three triplicates of each group were measured in Corning Flat white 96 well plate.

Based on the time course western blot (Figure 3.6B,C), IBMX can boost protein secretion significantly, although rich nutrient could mitigate this IBMX-dependent boost effect while increase the growth rate. In addition, IBMX can induce lipid accumulation inside cell(Figure 3.6D,E), which offers a new mode to harvest expensive therapeutic protein in the supernatant while biolipids could be accumulated as a by-product for the Biofuel, which will significantly decrease the cost of renewable bio-based energy.

Discussion

Genome-scale reconstruction of protein secretion has been accomplished in yeast and CHO cell systems (Gutierrez et al., 2020; Feizi et al., 2013) as a guidance to engineer cells to express protein of interest. *C.reinhardtii*, a new FDA approved algae for the food, is reported to express human antibody and therapeutic proteins with normal biological activity (Tran et al., 2009; Berndt et al., 2021). Here we integrated the core protein secretory pathway to the previously published metabolic flux model iCre1355. Candidates involved in algal secretory pathways were found by using orthologs from human, mouse, CHO, yeast and Arabidopsis with BLAST tool. Translocation machinery in algal is conserved and there are much more chaperones in the algal ER than that in the yeast ER, suggesting that *C.reinhardtii* exhibits a higher potential to express more complex enzymes and antibodies than yeast. As a reference, only monoclonal antibody fragments can be produced in glycoengineered yeast strain (Berlec et al., 2013; Spadiut et al., 2014), while antibody generated in algal chloroplast can be up to 5% of total protein contents (Tran et al., 2009).

In addition, major N-glycosylation steps are also conserved in algae, ManGlcNAc2-dolichol PP were synthesized sequentially on the cytosolic side of the ER and then flipped into the inner lumen of the ER and further matured to Glc3Man5GlcNAc2-dolichol PP instead of Glc3Man5GlcNAc2-dolichol PP found in mammalian cells (Mathieu-Rivet et al., 2013). Such unique glycans were predicted based on previous bioinformatic analysis that enzymes ALG3, ALG9, ALG10, ALG12 are missing in the algae genome (Liu et al., 2021). Recent studies also showed that the most abundant N-glycan forms identified were Man2GlcNAc2, Man3GlcNAc2, Man4GlcNAc2, Man5GlcNAc2 in *C.reinhardtii* while only a trace amount of higher oligomannoside N glycans were detected (Mathieu-Rivet et al., 2013). Once Glc3Man5GlcNAc2-dolichol PP is transferred to Asn-X-Ser/Thr/Cys N-glycosylation consensus sites of the nascent proteins (Stanley et al., 2015), the glycosylated protein will go through the protein folding and quality control by the conserved machinery (UGGT, calnexin/CRT, GSI, GSII) and then get transported into the Golgi. Further glycosylation modifications in the Golgi are species dependent and thus generate various glycan structures. According to the mass spec studies on N-glycans in *C.reinhardtii*, N-glycan modification in algae resembles mammalian N-glycans but also harbor plant specific glycosylation group such as xylose and fucose (Liu et al., 2021; Schulze et al., 2018). Cryo-electron tomography has been applied to track and measure COPI, COPII and secretory vesicles in algae (Bykov et al., 2017; Donohoe et al., 2013). And each candidate participated in these

vesicles were found in algae and corresponding energy costs for assembly and disassembly were calculated based on the average of vesicle size and published data (Borner et al., Cheng et al., 2007; Takamori et al., 2006).

This genome-scale model and our experimental data revealed the anti-correlation between protein production rate and growth rate with different carbon concentration in the medium, but positive correlation between them when nitrogen was limited. The anti-correlation between protein production rate and growth rate when carbon source was changed suggests that more building blocks are synthesized to support cell division for faster growing cells and unnecessary recombinant proteins were suppressed. However, the detailed mechanism needs to be investigated in the future. On the other side, the positive correlation between growth rate and protein synthesis rate under nitrogen limitation implies that both parameters share similar function of nitrogen concentration when nitrogen source is low in the medium. In addition, a constant degradation of chlorophyll content per cell and much lower total RNA abundance per cell was observed only in cells from Nitrogen limited condition. Since 80% of RNAs are ribosomal RNA (rRNA), a decrease of bulk RNAs indicates the disassembly of ribosomes for translation (Lodish et al., 2000). The same response of these Nitrogen-rich macromolecules, including chlorophyll, Bulk RNA and ribosome degradation upon nitrogen deprivation was reported in algae and yeast cells (Huang et al., 2015; Schmollinger et al., 2014). And with a higher level of RNA degradation, there are more nucleosides accumulated in yeast. Based on these facts and our recent discovery that an early sensor algal TF containing cAMP responding domain plays an essential role for ribosome degradation upon nitrogen starvation (Ren et al., 2021), one hypothesis is that suppression or degradation of bulk RNA under nitrogen limitation may increase the cAMP accumulation, activating specific TF to degrade ribosome to suppress the demand of nitrogen. As a result, protein production rate was decreased in the end. With a global decreased level of protein synthesis, growth rate was also decreased and thus there is such a positive correlation between growth rate and protein secretion rate with limited nitrogen. The same scenario is true for the Nitrogen-rich chlorophyll, which also got degraded in nitrogen limited medium. However, lipid as a major carbon sink was accumulated to store energy and carbon substance in this case and there is an anti-correlation between lipid accumulation and growth rate with low nitrogen in the medium. Moreover, Polyphosphate bodies in algae cells were reported to be accumulated under nitrogen and sulfur starvation, and starch and lipid bodies were increased under nitrogen, sulfur and phosphate limitation (Park et al., 2014; Boyle et al., 2012; Aksoy et al). All these previous discoveries supported our model and hypothesis that when one nutrient element (e.g, Nitrogen) is absent or limited in

the environment, the algal cells will re-allocate resources and save active energy and other nutrient elements, such as Carbon and Phosphate, to corresponding sink starch, lipid and Polyphosphate bodies.

Methods

Reconstruction of the algae secretory pathway

The whole pathway from protein production to their different destination in algae was divided to subsystems: production, translocation, protein folding, ER glycosylation, Dolichol synthesis, ERAD, golgi glycosylation, GPI biosynthesis, COPII transportation, COPI transportation, ALP, CPY. The component involved in each subsystem was blasted in NCBI BLAST with yeast, human and mouse orthologs proteins and compiled from UniProt, Phytozome, KEGG and NCBI. All proteins and enzymes obtained was verified with Arabidopsis orthologs to double check their function or conserved domain. Candidates with at least 60% sequence similarity to yeast and mouse orthologs were kept if they were found in other database or mapped to the unique Arabidopsis protein. Specific reactions for each subsystem were de novo reconstructed and utilized as a template for algal proteins. In addition, necessary metabolic reaction such as GPI synthesis and GLYCOSPHINGOLIPID metabolism were missing in the iCre1355 and hence added to our new model.

Disulfide bond prediction

Dipro(Cheng et al., 2005; Baldi et al., 2006) is a cysteine disulfide bond predictor based on 2D recurrent neural network, support vector machine, graph matching and regression algorithms. It can predict if the sequence has disulfide bonds or not, estimate the number of disulfide bonds, and predict the bonding state of each cysteine and the bonded pairs. It yields the best accuracy and can handle any number of disulfide bonds where most of methods available so far only can handle less than 6 disulfide bonds. The proteome is processed in two steps. The first step is to use support vector machine to classify if the sequence has disulfide bonds or not. And then we used neural network and graph algorithm to predict the number of bonds, bond pattern.

Protein-specific information matrix (PSIM)

The PSIM contains the necessary information to construct a protein-specific secretory model from the template reactions in our reconstruction. The columns in the PSIM are presence of a signal peptide (SP), number of disulfide bonds (DSB), presence of Glycosylphosphatidylinositol (GPI) anchors, number of N-linked (NG) and O-linked (OG) glycans, number of transmembrane domains (TMD), subcellular location, protein length, and molecular

weight. For most proteins, the information in the PSIM was obtained from the Uniprot database. Computational tool GPI-SOM was used to predict GPI anchors. Disulfide bond was predicted as aforementioned machine learning algorithm. N glycosylation data was collected from previous published work (Mathieu-Rivet et al., 2013; Stanley et al., 2015; Schulze et al., 2018). And we utilized click chemistry to identify O-glycans in algae cells.

Detection of O-linked glycosylation sites in *Chlamydomonas* proteome

Algae cells were seeded at 500,000 cells/ml the day before treatment. Cells were treated with 100 μ M GalNAz alone for 2 h, then 500 ng/ml DMSO (vehicle) was added, and cells were harvested 4 h later. Protein amounts were quantified by Pierce™ BCA Protein Assay Kit (Thermo Fisher, Cat #23225) at 562nm against BSA standards, and 9 mg of wt total protein or 5mg of total protein from transgenic lysate was processed further. Alkyne-agarose beads (Click Chemistry Tools, Cat # 1032-2) were washed three times in click buffer. Protein samples were precleared with 300 μ l bead volume of washed alkyne-agarose beads with gentle rotation for 2 h at room temperature. After preclearing, supernatants were removed and combined with 100 μ l of equilibrated alkyne-agarose beads, 5 mM sodium ascorbate(Sigma aldrich, Cat # A4034-100G), 100 μ M Tris[(1-benzyl-1H-1,2,3-triazol-4-yl)methyl]amine (TBTA, Sigma aldrich, Cat #) and 1 mM CuSO₄(Sigma aldrich, Cat # C1297-100G). Reactions were rotated at room temperature for 2 h and then quenched by addition of 10 mM Ethylenediaminetetraacetic acid (EDTA, Sigma Aldrich, Cat # 60-00-4) Beads were washed sequentially with three 1 ml washes of each of the following: 1% Sodium dodecyl sulfate (SDS, Sigma Aldrich, Cat # 151-21-3), 20 mM Tris pH 7.4; 1% SDS, 10 mM dithiothreitol (DTT, Sigma Aldrich, Cat # D0632-5G), 20 mM Tris pH 7.4; 1X PBS (Thermo Fisher, Cat #14190144); 8 M urea (Sigma Aldrich, Cat # U5378-100G); 1X PBS; 6 M guanidine hydrochloride(Sigma Aldrich, Cat # G3272-25G); 1X PBS; 5 M NaCl; 1X PBS; 10X PBS; 1X PBS; 20% isopropanol; 20% acetonitrile(Sigma Aldrich, Cat # 34851-1L); 50 mM ammonium bicarbonate(Sigma Aldrich, Cat # A6141-25G). Washed beads were stored in 100 μ l 50 mM ammonium bicarbonate.

Construction of models and constraint-based analysis

In the Python Jupyter Notebook, we took each row from the PSIM as input to produce an expanded iCre1477 with the product-specific secretory pathway of the corresponding protein. Flux balance analysis and all other constraint-based analyses were done using the COBRA toolbox v2.049 in MATLAB R2019b and the Gurobi solver version 6.0.0.

Batch cultivation with different carbon concentrations

C. reinhardtii transgenic strain cc1690 expressing recombinant protein TGM were cultured in modified tris acetate-phosphate (TAP) medium with different concentration of sodium acetate (4.17mM, 8.34mM, 16.67mM, 30mM and 40mM). PH of modified TAP medium was titrated to 7.0 by HCl. OD was monitored and 1ml supernatant of culture was harvested every three hours.

Batch cultivation with different nitrogen concentrations

C. reinhardtii transgenic strain expressing recombinant protein TGM were cultured in tris acetate-phosphate(TAP) medium supplemented with different nitrogen sources(NH₄Cl, NaNO₃, Urea) at 0.747mM/1.87mM/3.47mM/5.6mM of Nitrogen correspondingly. For reference, molar concentration of Nitrogen in TAP medium is 7.47mM. Cells were grown in CO₂ box filled with 0.5% CO₂ or at atmospheric CO₂ concentration. Fresh cultures were changed every 6 hours after OD achieved to 0.6 to make it chemostat. OD was monitored and 1ml supernatant of culture was harvested every three hours. Lipid content was measured by Nile red staining assay.

Nile red staining assay and chlorophyll quantification

Stock solution of Nile red (9-diethylamino-5H-benzo[a]phenoxa-]phenoxazine-5-one, Thermo fisher, Cat # N1142) was prepared dissolving 0.5 mg of Nile red per ml in acetone. 1ul of Nile red stock solution was added to 10ul acetone and mixed to 1 ml cells. After 30mins incubation, chlorophyll and Nile red staining were measured by Tecan plate reader (Tecan, Männedorf, Switzerland) with an Excitation wavelength 550nm and Emission wavelength 640nm. 200ul of three triplicates of each group were measured in Corning Flat white 96 well plate. Chlorophyll was measured with excitation wavelength 440nm and Emission wavelength 680nm.

Bulk RNA extraction and quantification

Cells from each time point were pelleted by centrifugation at 3000×g, at room temperature for 10 min and mixed with PureLink™ Plant RNA Reagent (Invitrogen, Cat # 12322012) were used to for cell lysis with a ratio of 1ml reagent : 0.2g algae biomass. Equal volume of Chloroform/isoamyl alcohol mixture (Millipore Sigma, Cat # 25666-500ML) was added to isolate proteins and nucleotides. Extracted RNA in upper phase after 10mins centrifugation at 12,000g (4 °C) were harvested and precipitated with equal volume of Isopropanol (Millipore Sigma, Cat # 67-63-0). RNA pellets were obtained after 15mins centrifugation at 12,000g (4 °C) and washed with 75% ethanol. RNA was dried for 10mins, resuspended with DEPC treated water (Invitrogen, Cat # AM9920) and quantified by Nanodrop.

Titer determination

1ml supernatant at different time points were collected and 500ul of them were separately added with 2ml acetone for protein precipitation. Protein pellets were resuspended with 25ul PBS and then 9ul of 4X Protein loading buffer. 10ul of each sample along with a certain amount(5ng~20ng) of Recombinant Posi-Tag Epitope Tag Protein containing the epitope FLAG tag (BioLegend, Cat #931301) as a FLAG standard protein were loaded to the SDS-PAGE gel and transferred to Nitrocellulose membrane. Membrane was washed with TBST (Tris-buffered saline plus 0.05 % Tween-20) 5mins for three times and then blocked with 5 % non-fat milk powder in TBST for 1 hour and then incubated with Monoclonal ANTI-FLAG® M2-Alkaline Phosphatase antibody (Sigma-Aldrich, St. Louis, MO, USA, Cat #A9469; final dilution 1:5000) for 1 hour. The membrane was then washed with TBST 3 x 5 mins and visualized in nitro-blue tetrazolium (NBT) / 5-bromo-4-chloro-3'-indolyphosphate p-toluidine salt (BCIP) (Sigma-Aldrich, St. Louis, MO, USA, Cat #B5655) added directly to the membrane for development. Developed dot blot intensity was correlated with yield of secreted TGM1 and the transgenic clones with highest protein secretion were selected for further protein expression and purification.

Estimation of protein secretion cost

We estimated the energetic cost of synthesizing and secreting all endogenous Chlamy cell proteins. These proteins were chosen for containing a signal peptide in their sequence and/or for being localized in the cell membrane (according to the UniProt database). The energetic cost (in units of number of ATP equivalents) of secreting each protein (length L) was computed using the same formula from published CHO cell protein secretion model (Gutierrez et al., 2020).

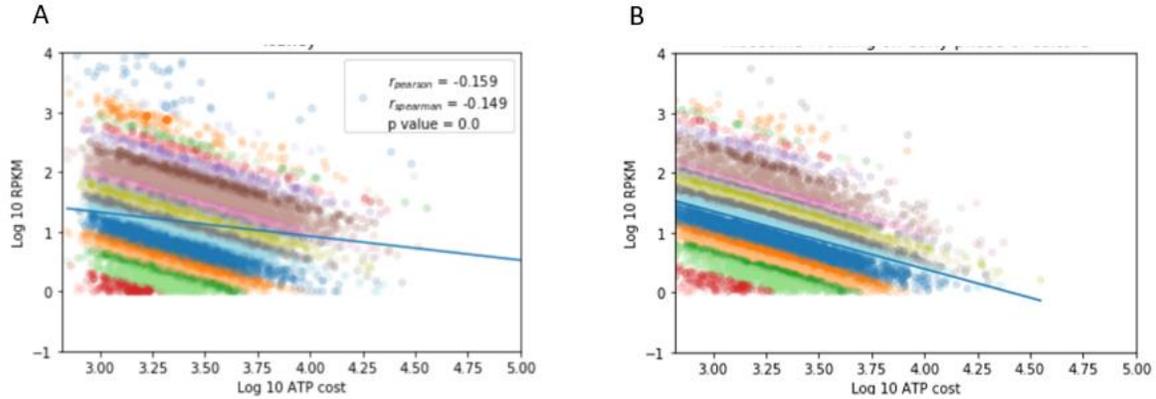
Constraints used in models and Pareto optimality frontiers

All models were constrained using different sets of experimental uptake rates. To construct Pareto optimality frontiers, we used the robustAnalysis function from the COBRA Toolbox v2.0 in Matlab 2019a using biomass as the control and secretion of the recombinant protein as the objective reactions, respectively.

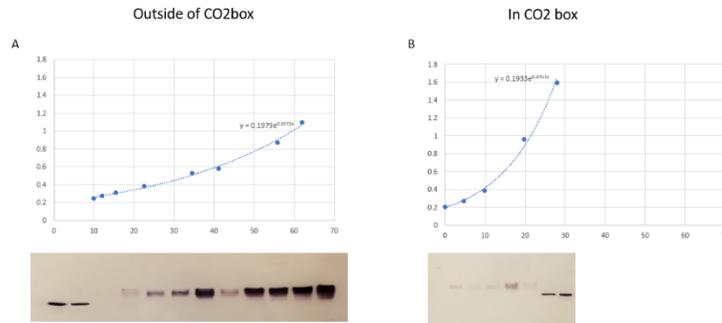
Analysis of gene expression versus protein cost

Ribosome-profiling data were used to quantify the ribosomal occupancy of each transcript in algae cells (Chung et al., 2015). A cutoff of 1 RPKM was used to remove genes with low expression (10,045 genes removed from day 3 analysis and 10,411 from day 6 analysis). We used Spearman correlation to assess the variation of expression levels with respect to protein ATP cost.

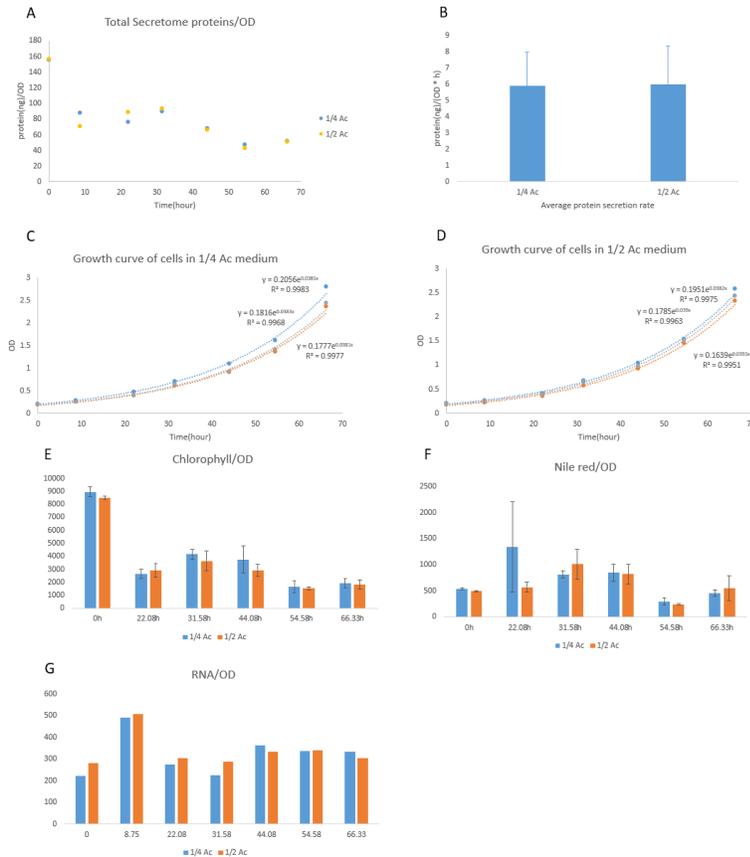
Supplementary Figures



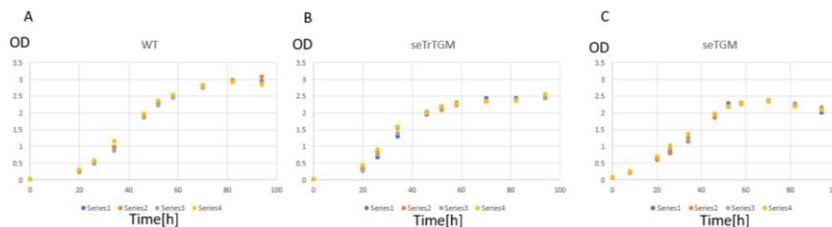
Supplementary Figure 3.1 Energy cost vs Riboseq transcript abundance from human kidney (A) and CHO cell(B). Scatter plot and Spearman correlation of protein cost(ATP equivalent) per protein versus transcript abundance from previous Ribo-Seq data(Kallehaug et al. 2017).



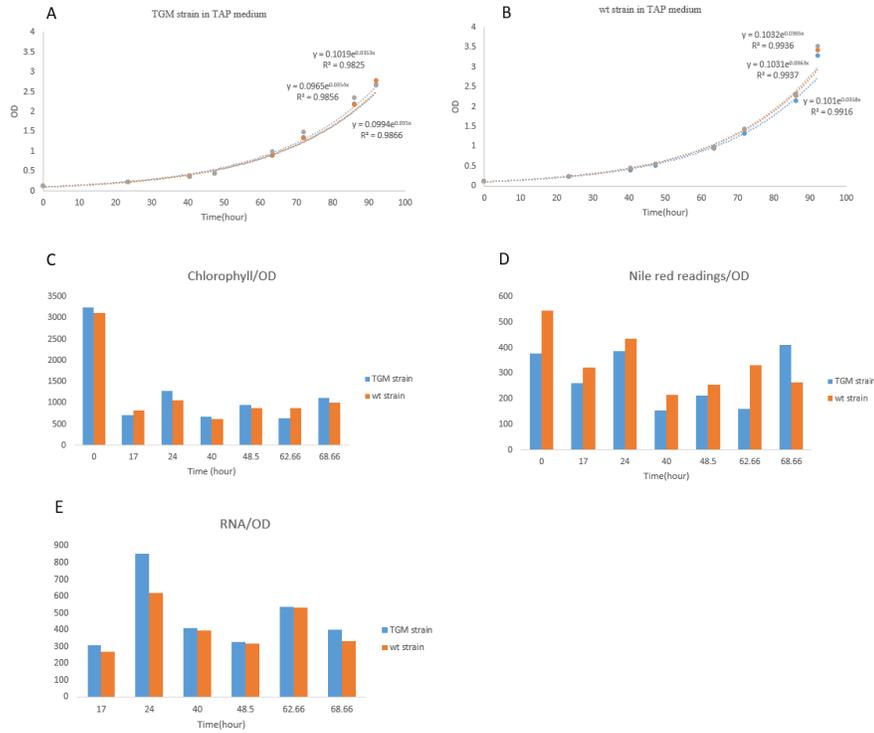
Supplementary Figure 3.2 Suppression of recombinant protein secretion when algae cells grew in CO2 box with a higher growth rate. **A.** Cells grown outside of CO2 box with atmospheric CO2 in TAP medium under constant light. **B.** Cells grown in the CO2 box (0.5% CO2) in TAP medium under constant light. 45-KDa Recombinant Posi-Tag Epitope Tag Protein(BioLegend, Cat #931301) was loaded as a FLAG standard. TGM protein was detected with Monoclonal ANTI-FLAG® M2-Alkaline Phosphatase antibody (Sigma-Aldrich, St. Louis, MO, USA, Cat #A9469; final dilution 1:5000).



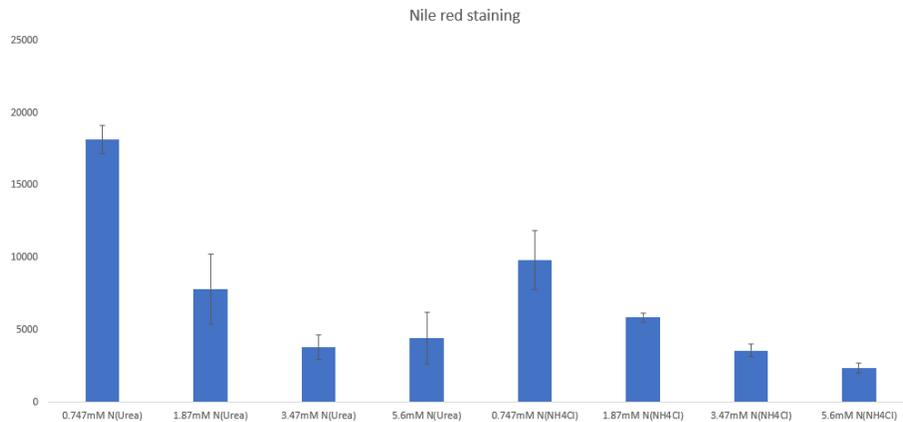
Supplementary Figure 3.3 Time course measure of secreted proteins, growth rate, Chlorophyll, Nile red staining for lipid content and RNA. **A.** Total protein content secreted per OD in supernatant. **B.** Total protein secretion rates for cells from TAP acetate-free medium with 1/4 of normal acetate concentration(16.67mM) and cells from TAP acetate-free medium with 1/2 of normal acetate concentration. **C.** Growth rate of cells from TAP acetate-free medium with 1/4 of normal acetate concentration(16.67mM). **D.** Growth rate of cells from TAP acetate-free medium with 1/2 of normal acetate concentration(16.67mM). **E.** Chlorophyll per OD for cells from two groups with different acetate concentration. **F.** Nile red readings representing lipid content per OD was measured by Tecan plate reader with an Excitation wavelength 550nm and Emission wavelength 640nm.. 200ul of three replicates of each group were measured in Corning Flat white 96 well plate. **G.** Bulk RNA per OD for cells from two groups with different acetate concentration.



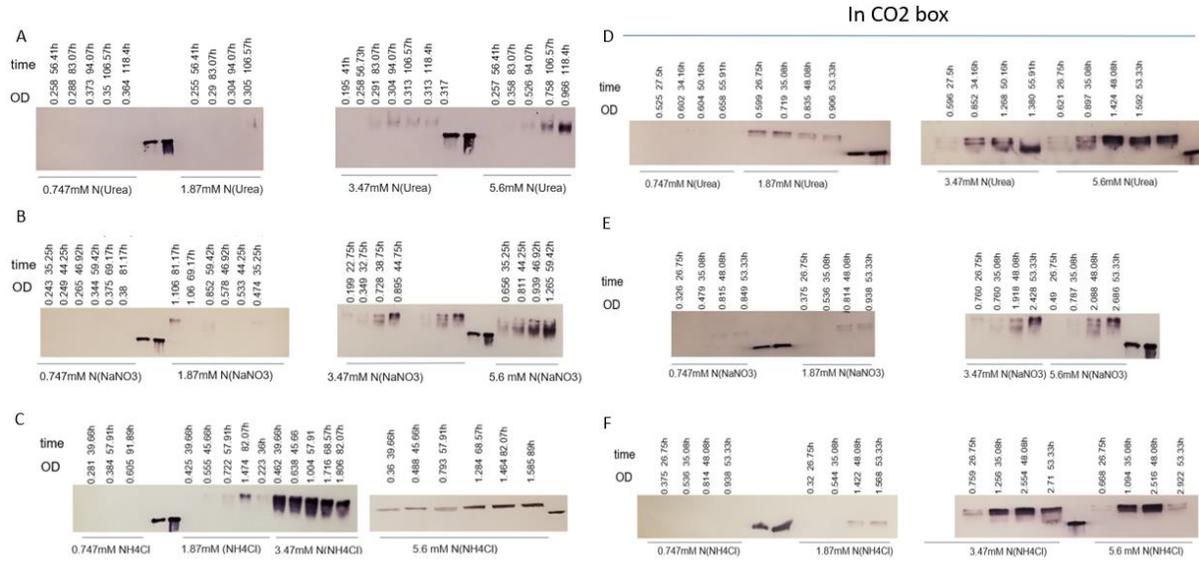
Supplementary Figure 3.4 Recombinant protein expression slowed growth rate and lowered cell density during stationary phase. (A) Growth rate of wild type strain. (B) Growth rate of transgenic sTrTGM strain. (C) Growth rate of transgenic sTGM strain. Cells were grown in TAP medium under constant light with atmospheric CO₂ and culture was not refreshed to allow cells go through whole log phase and saturation phase.



Supplementary Figure 3.5 Time course measure of growth rate, Chlorophyll, Nile red staining for lipid content and RNA. **A.** Growth rate of TGM strain in TAP medium. **B.** Growth rate of wt strain in TAP medium. **C.** Chlorophyll per OD for TGM strain and wt strain in TAP medium. **D.** Nile red stainings for lipid profile per OD for TGM strain and wt strain in TAP medium were measured by Tecan plate reader with an Excitation wavelength 550nm and Emission wavelength 640nm. 200ul of three triplicates of each group were measured in Corning Flat white 96 well plate. **E.** Bulk RNA per OD for TGM strain and wt strain in TAP medium.



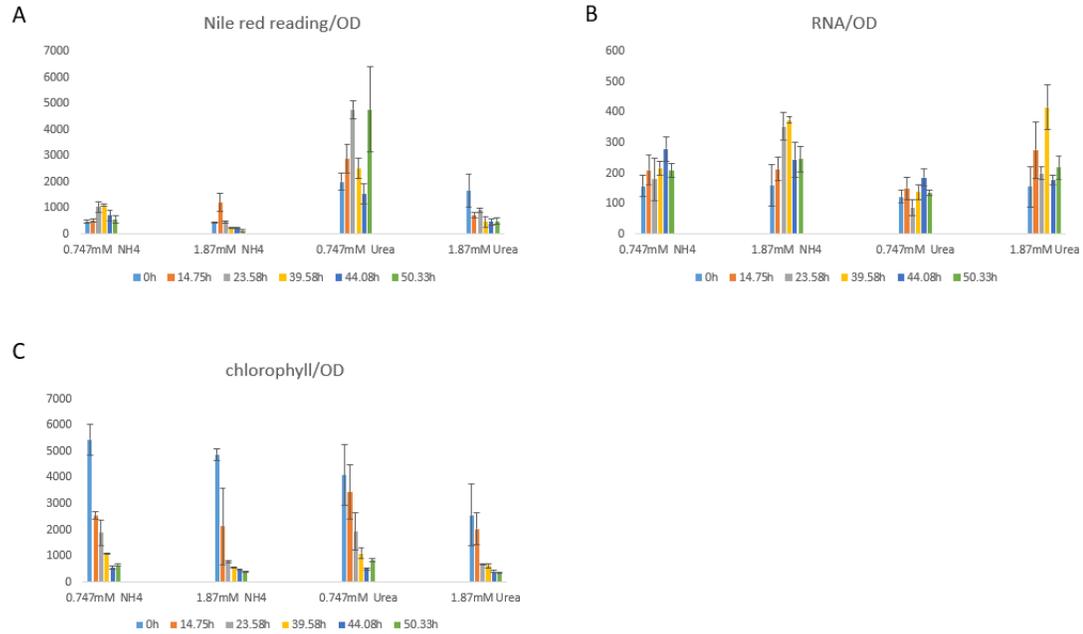
Supplementary Figure 3.6 Nile red staining for lipid content of cells from different nitrogen groups. Levels of Nile red fluorescence were measured by Tecan plate reader with an Excitation wavelength 550nm and Emission wavelength 640nm. 200ul of three triplicates of each group were measured in Corning transparent 96 well plate and calculated normalized to their OD. Lipid accumulation in algae when nitrogen is limited for 0.747mM Urea or NH₄Cl.



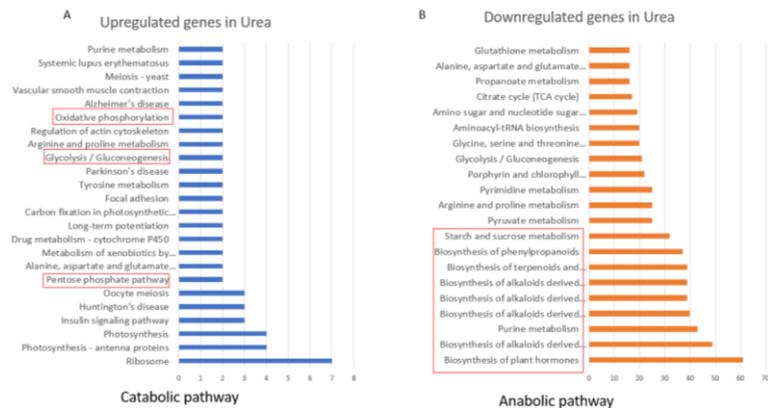
Supplementary Figure 3.7 Time course western blot of secreted protein in the culture. Different concentration of Urea(A,D), Sodium nitrate(B,E) and Ammonium Chloride(C,F) were incubated in normal lab condition or CO₂ box with 0.5% CO₂. 45-KDa Recombinant Posi-Tag Epitope Tag Protein (BioLegend, Cat #931301) was loaded as a FLAG standard. TGM protein was detected with Monoclonal ANTI-FLAG® M2-Alkaline Phosphatase antibody (Sigma-Aldrich, St. Louis, MO, USA, Cat #A9469; final dilution 1:5000).



Supplementary Figure 3.8 Comparison of protein secretion yield with saturated Nitrogen vs limited Nitrogen predicted by genome-scale model. Yellow color represents high protein secretion rate while black color indicates low protein secretion rate. Percentage of normal Nitrogen(7.47mM) used in TAP medium was given after each nitrogen source name. Different nitrogen concentration was simulated in CO₂ box(0.5% CO₂) or with low atmospheric CO₂(0.03% CO₂).



Supplementary Figure 3.9 Time course measurement for Chlorophyll, Nile red and bulk RNA for cells from different nitrogen groups. **A.** Chlorophyll per OD for cells from TAP N-free medium with 0.747mM or 1.87mM NH₄Cl or Urea. **B.** Nile red staining for lipid profile per OD from different nitrogen groups were measured by Tecan plate reader with an Excitation wavelength 550nm and Emission wavelength 640nm. 200ul of three triplicates of each group were measured in Corning Flat white 96 well plate. **C.** Bulk RNA per OD for cells from TAP N-free medium with 0.747mM or 1.87mM NH₄Cl or Urea.



Supplementary Figure 3.10 Differential expression change between Urea medium and Ammonium medium. **A.** Catabolic pathways were upregulated in cells from Urea medium. **B.** Anabolic pathways were downregulated in cells from Urea medium.

Acknowledgement

Chapter 3, in part is currently being prepared for submission for publication of the material. Bijie Ren, Terence Hwa, Jahir M. Gutierrez, Frank Fields, Nathan Lewis, Stephen Mayfield. Genome-scale reconstruction of Metabolic-Protein secretion modeling for *Chlamydomonas reinhardtii*. The dissertation author was the primary investigator and author of this paper.

I would like to acknowledge Professor Nathan Lewis and Jahir Gutierrez for the unquantifiable measure of support and guidance about cell modeling project, who also set up a role model for me to learn. I would like to acknowledge Professor Jagroop Pandhal and Dr. Caroline Evans for their great work in the O glycosylation of algal whole proteome. It is their support that helped me in an immeasurable way. I would like to acknowledge Professor Hwa's genius guidance for the top-down toy modeling and great mentoring during qBio program. I would like to acknowledge Dr. Christophe Magnan for the great help to predict disulfide bond numbers for Chlamydomons proteins.

Reference

- Alper H, Jin YS, Moxley JF, Stephanopoulos G. Identifying gene targets for the metabolic engineering of lycopene biosynthesis in *Escherichia coli*. *Metab Eng.* 2005;7:155–164.
- Park JH, Lee KH, Kim TY, Lee SY. Metabolic engineering of *Escherichia coli* for the production of L-valine based on transcriptome analysis and in silico gene knockout simulation. *Proc Natl Acad Sci U S A.* 2007;104:7797–7802.
- Milne CB, Kim PJ, Eddy JA, Price ND. Accomplishments in genome-scale in silico modeling for industrial and medical biotechnology. *Biotechnol J.* 2009;4:1653–1670.
- S. Imam, S. Schauble, J. Valenzuela, A.L.G. de Lomana, W. Carter, N.D. Price, et al. A refined genome-scale reconstruction of *Chlamydomonas* metabolism provides a platform for systems-level analyses *Plant J.*, 84 (2015), pp. 1239-1256).
- J.M. Gutierrez, A. Feizi, S. Li, T.B. Kallehauge, H. Hefzi, L.M. Grav, D. Ley, D.B. Hizal, M.J. Betenbaugh, B. Voldborg, et al. Genome-scale reconstructions of the mammalian secretory pathway predict metabolic costs and limitations of protein secretion *bioRxiv* (2018), 10.1101/351387
- Feizi A, Österlund T, Petranovic D, Bordel S, Nielsen J. Genome-scale modeling of the protein secretory machinery in yeast. *PLoS One.* 2013 May 7;8(5):e63284. doi: 10.1371/journal.pone.0063284. PMID: 23667601; PMCID: PMC3646752.
- Kol S, Ley D, Wulff T, Decker M, Arnsdorf J, Schoffelen S, Hansen AH, Jensen TL, Gutierrez JM, Chiang AWT, Masson HO, Palsson BO, Voldborg BG, Pedersen LE, Kildegaard HF, Lee GM, Lewis NE. Multiplex secretome engineering enhances recombinant protein production and purity. *Nat Commun.* 2020 Apr 20;11(1):1908. doi: 10.1038/s41467-020-15866-w. PMID: 32313013; PMCID: PMC7170862.
- Lewis, N. E. et al. Large-scale in silico modeling of metabolic interactions between cell types in the human brain. *Nature Biotech.* 28, 1279–1285 (2010).
- You C, H. Okano, S. Hui, Z. Zhang, M. Kim, C.W. Gunderson, Y.P. Wang, P. Lenz, D. Yan, T. Hwa Coordination of bacterial proteome with metabolism by cyclic AMP signalling. *Nature*, 500 (2013), pp. 301-306

- Basan M, Hui S, Okano H, Zhang Z, Shen Y, Williamson JR, Hwa T. Overflow metabolism in *Escherichia coli* results from efficient proteome allocation. *Nature*. 2015 Dec 3;528(7580):99-104. doi: 10.1038/nature15765. PMID: 26632588; PMCID: PMC4843128.
- Choi YE, Rhee JK, Kim HS, Ahn JW, Hwang H, Yang JW. Chemical Genetics Approach Reveals Importance of cAMP and MAP Kinase Signaling to Lipid and Carotenoid Biosynthesis in Microalgae. *J Microbiol Biotechnol*. 2015 May;25(5):637-47. doi: 10.4014/jmb.1408.08035. PMID: 25563422.
- Chung BY, Hardcastle TJ, Jones JD, Irigoyen N, Firth AE, Baulcombe DC, Brierley I. The use of duplex-specific nuclease in ribosome profiling and a user-friendly software package for Ribo-seq data analysis. *RNA*. 2015 Oct;21(10):1731-45. doi: 10.1261/rna.052548.115. Epub 2015 Aug 18. PMID: 26286745; PMCID: PMC4574750.
- Kallehauge TB, Li S, Pedersen LE, Ha TK, Ley D, Andersen MR, Kildegaard HF, Lee GM, Lewis NE. Ribosome profiling-guided depletion of an mRNA increases cell growth rate and protein secretion. *Sci Rep*. 2017 Jan 16;7:40388. doi: 10.1038/srep40388. PMID: 28091612; PMCID: PMC5238448.
- Reimo Zoschke, Kenneth P. Watkins, Alice Barkan. A Rapid Ribosome Profiling Method Elucidates Chloroplast Ribosome Behavior in Vivo. *The Plant Cell* Jun 2013, 25 (6) 2265-2275; DOI: 10.1105/tpc.113.111567.
- Boyle NR, Page MD, Liu B, Blaby IK, Casero D, Kropat J, Cokus SJ, Hong-Hermesdorf A, Shaw J, Karpowicz SJ, Gallaher SD, Johnson S, Benning C, Pellegrini M, Grossman A, Merchant SS. Three acyltransferases and nitrogen-responsive regulator are implicated in nitrogen starvation-induced triacylglycerol accumulation in *Chlamydomonas*. *J Biol Chem*. 2012 May 4;287(19):15811-25. doi: 10.1074/jbc.M111.334052. Epub 2012 Mar 8. PMID: 22403401; PMCID: PMC3346115.
- Lodish H, Berk A, Zipursky SL, et al. *Molecular Cell Biology*. 4th edition. New York: W. H. Freeman; 2000. Section 11.6, Processing of rRNA and tRNA. Available from: <https://www.ncbi.nlm.nih.gov/books/NBK21729/>
- Huang H, Kawamata T, Horie T, Tsugawa H, Nakayama Y, Ohsumi Y, Fukusaki E. Bulk RNA degradation by nitrogen starvation-induced autophagy in yeast. *EMBO J*. 2015 Jan 13;34(2):154-68. doi: 10.15252/embj.201489083. Epub 2014 Dec 2. PMID: 25468960; PMCID: PMC4337068.
- Schmollinger S, Mühlhaus T, Boyle NR, Blaby IK, Casero D, Mettler T, Moseley JL, Kropat J, Sommer F, Strenkert D, Hemme D, Pellegrini M, Grossman AR, Stitt M, Schroda M, Merchant SS. Nitrogen-Sparing Mechanisms in *Chlamydomonas* Affect the Transcriptome, the Proteome, and Photosynthetic Metabolism. *Plant Cell*. 2014 Apr;26(4):1410-1435. doi: 10.1105/tpc.113.122523. Epub 2014 Apr 18. PMID: 24748044; PMCID: PMC4036562.
- Anthony J Berndt; Tressa N Smalley; Bijie Ren; Amr Badary; Ashley Sproles; Francis Fields; Yasin Torres-Tiji; Vanessa Heredia; Stephen P Mayfield. Recombinant production of a functional SARS-CoV-2 spike receptor binding domain in the green alga *Chlamydomonas reinhardtii*. bioRxiv | ID: ppbiorxiv-428890.
- Berlec A, Strukelj B. Current state and recent advances in biopharmaceutical production in *Escherichia coli*, yeasts and mammalian cells. *J Ind Microbiol Biotechnol*. 2013 Apr;40(3-4):257-74. doi: 10.1007/s10295-013-1235-0. Epub 2013 Feb 6. PMID: 23385853.
- Spadiut O, Capone S, Krainer F, Glieder A, Herwig C. Microbials for the production of monoclonal antibodies and antibody fragments. *Trends Biotechnol*. 2014 Jan;32(1):54-60. doi: 10.1016/j.tibtech.2013.10.002. Epub 2013 Oct 31. PMID: 24183828; PMCID: PMC3906537.
- Gutierrez JM, Feizi A, Li S, Kallehauge TB, Hefzi H, Grav LM, Ley D, Baycin Hizal D, Betenbaugh MJ, Voldborg B, Fastrup Kildegaard H, Min Lee G, Palsson BO, Nielsen J, Lewis NE. Genome-scale reconstructions of the mammalian secretory pathway predict metabolic costs and limitations of protein secretion. *Nat Commun*. 2020 Jan 2;11(1):68. doi: 10.1038/s41467-019-13867-y. PMID: 31896772; PMCID: PMC6940358.

- Mathieu-Rivet E, Scholz M, Arias C, Dardelle F, Schulze S, Le Mauff F, Teo G, Hochmal AK, Blanco-Rivero A, Loutelier-Bourhis C, Kiefer-Meyer MC, Fufezan C, Burel C, Lerouge P, Martinez F, Bardor M, Hippler M. Exploring the N-glycosylation pathway in *Chlamydomonas reinhardtii* unravels novel complex structures. *Mol Cell Proteomics*. 2013 Nov;12(11):3160-83. doi: 10.1074/mcp.M113.028191. Epub 2013 Aug 2. PMID: 23912651; PMCID: PMC3820931.
- Xiaojuan Liu, Xihui Xie, Hong Du, Edmond Sanganyado, Wanna Wang, Muhammad Aslam, Jichen Chen, Weizhou Chen, Honghao Liang. Bioinformatic analysis and genetic engineering approaches for recombinant biopharmaceutical glycoproteins production in microalgae. *Algal Research*. Volume 55, 2021, ISSN 2211-9264, <https://doi.org/10.1016/j.algal.2021.102276>.
- Schulze S, Oltmanns A, Machnik N, Liu G, Xu N, Jarmatz N, Scholz M, Sugimoto K, Fufezan C, Huang K, Hippler M. N-Glycoproteomic Characterization of Mannosidase and Xylosyltransferase Mutant Strains of *Chlamydomonas reinhardtii*. *Plant Physiol*. 2018 Mar;176(3):1952-1964. doi: 10.1104/pp.17.01450. Epub 2017 Dec 29. PMID: 29288232; PMCID: PMC5841687
- Bykov YS, Schaffer M, Dodonova SO, Albert S, Plitzko JM, Baumeister W, Engel BD, Briggs JA. The structure of the COPI coat determined within the cell. *Elife*. 2017 Nov 17;6:e32493. doi: 10.7554/eLife.32493. PMID: 29148969; PMCID: PMC5716667.
- Donohoe BS, Kang BH, Gerl MJ, Gergely ZR, McMichael CM, Bednarek SY, Staehelin LA. Cis-Golgi cisternal assembly and biosynthetic activation occur sequentially in plants and algae. *Traffic*. 2013 May;14(5):551-67. doi: 10.1111/tra.12052. Epub 2013 Feb 25. PMID: 23369235; PMCID: PMC3622843.
- Borner GH, Antrobus R, Hirst J, Bhumbra GS, Kozik P, Jackson LP, Sahlender DA, Robinson MS. Multivariate proteomic profiling identifies novel accessory proteins of coated vesicles. *J Cell Biol*. 2012 Apr 2;197(1):141-60. doi: 10.1083/jcb.201111049. PMID: 22472443; PMCID: PMC3317806.
- Cheng, Y., Boll, W., Kirchhausen, T., Harrison, S. C. & Walz, T. Cryo-electron tomography of clathrin-coated vesicles: structural implications for coat assembly. *J. Mol. Biol.* 365, 892–899 (2007).
- Takamori S, Holt M, Stenius K, Lemke EA, Grønborg M, Riedel D, Urlaub H, Schenck S, Brügger B, Ringler P, Müller SA, Rammner B, Gräter F, Hub JS, De Groot BL, Mieskes G, Moriyama Y, Klingauf J, Grubmüller H, Heuser J, Wieland F, Jahn R. Molecular anatomy of a trafficking organelle. *Cell*. 2006 Nov 17;127(4):831-46. doi: 10.1016/j.cell.2006.10.030. PMID: 17110340.
- Park JJ, Wang H, Gargouri M, Deshpande RR, Skepper JN, Holguin FO, Juergens MT, Shachar-Hill Y, Hicks LM, Gang DR. The response of *Chlamydomonas reinhardtii* to nitrogen deprivation: a systems biology analysis. *Plant J*. 2015 Feb;81(4):611-24. doi: 10.1111/tpj.12747. PMID: 25515814.

CONCLUSION

This dissertation attempts to develop a preliminary pipeline of Discover-Model-Engineering for synthetic biology and unicellular factory based on green algae. We tried to set up a high throughput discovery pipeline to identify transcription regulation network for *C.reinhardtii*, and build a bottom up flux model to simulate metabolic and protein secretion. These methods and tools combined with CRISPR-Cas9 and transformation technology allow us to engineer algal cells to express many valuable products, as demonstrations of the future potential of this platform.

One example is listed in Chapter One, where we produced a novel recombinant therapeutic protein (TGM1) that can potentially act as a therapeutic for treatment of chronic inflammatory response in the gut. The potential for ER-retained TGM to be delivered orally using microalgae as a vehicle has the potential to greatly boost the anti-inflammatory effect and reduce cost of this novel therapeutic. These advances in medical biotechnology are already being pursued at the next level and several biotech companies have formed to continue the work and produce useful therapeutics to aid in relieving human diseases.

Chapter 2 introduces a new high throughput pipeline to identify transcription factor interactions, by utilizing a modified yeast two hybridization assay to quickly discover transcription factor interactions of green algal. This pipeline can be applied to other organisms and should also help us to further understand how cells coordinate metabolism and proteomes under different condition. For example, we identified two transcription factors that may repress ribosome biosynthesis and oxidative phosphorylation pathways separately under nitrogen starvation. Such work can be used to secrete therapeutic proteins under nitrogen starvation while increasing lipid accumulation at the same time, for use in biofuels or as biodegradable polymers.

Moreover, our refined reconstruction in chapter 3 enables us to calculate energy cost of each protein under Autotrophic/Mixotrophic/Heterotrophic condition and shows that *Chlamydomonas* has different growth capacity and maximal protein secretion yield under each condition. Interestingly, there's an anti-correlation between protein yield and growth capacity due to the limitation of resources for growth, which is proved by our independent biophysical model and experimental data. In addition, this reconstruction model also indicates that if we knock out unnecessary subsystem in the cell, free energy saved could be used in rest subsystem including protein secretion pathway. We also identify that addition of cAMP will increase the useful lipid accumulation without affecting protein of interest secretion, which provides a new strategy to utilize algae as a cost-effective cell factory for coproduction of biofuel and therapeutic antibody or cytokine.

The preliminary pipeline developed in this dissertation will be used to understand how algae cells work and proteins are secreted. Based on these discoveries, a new mode was proposed to harvest expensive proteins in the supernatant while accumulated lipids inside the cell biomass as a free by-product.