UNIVERSITY OF CALIFORNIA, SAN DIEGO

Improving the genetic tractability of the green alga *Chlamydomonas reinhardtii*

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

in

Biological Sciences

by

Elizabeth Anne Specht

Committee in charge:

Professor Stephen Mayfield, Chair
Professor Jens Lykke-Andersen
Professor Steven Briggs
Professor Theresa Gaasterland
Professor Susan Golden

2014

The dissertation of Elizabeth Anne Specht is approved, and it is acceptable in quality and form for publication on microfilm and electronically:

_____

_____

_____
Chair

_____

_____

University of California, San Diego

2014

DEDICATION

This dissertation is dedicated to my family,

for their unwavering support

of all my academic and scientific endeavors.

TABLE OF CONTENTS

LIST OF ABBREVIATIONS

ARG7 – gene for arginosuccinate lyase

ARS2 – gene for arylsulfatase

HR – homologous recombination

HSM – high salt media

HSP70A – heat shock protein 70A

NHEJ – non-homologous end-joining

RBCS2 – rubisco small subunit

TAP – tris acetate phosphate

UTR – untranslated region

LIST OF FIGURES

LIST OF TABLES

ACKNOWLEDGEMENTS

My first acknowledgement for this work goes to my parents, who gave me all the support and tools I could ask for to thrive academically and to strive for achievement. Without my father's contagious enthusiasm for science and all things nerdy, and my mother's steadfast encouragement and support of my academic and personal pursuits, I could never have hoped to be delivering this doctoral dissertation. My sister has also been an invaluable advocate and supporter of my marathon academic journey; her unwavering appreciation for science and research have inspired me to push onward in my times of doubt.

My supporter through the daily grind and my weekly co-celebrant of positive results has been and will continue to be Dave Masica. He has been a source of inspiration, wisdom, encouragement, and – when needed – consolation. Dave has been extremely understanding of the demands grad school placed on me – exhibiting the kind of understanding that only comes from having been through it himself – and has supported me every step of the way, even from across the country. Our puppy Flacco deserves some acknowledgement as well, for providing me with endless comic relief and much-needed puppy cuddles during these final stressful months.

I want to thank by name the undergraduates I've had the privilege of mentoring, for all their dedication and willingness to contribute long hours to critical aspects of much of this work: Maya, Kevin, and Prema. The work described in this dissertation would not have been possible without these students, and I have had the opportunity to learn so much through teaching them.

I could never have made it through these five years without my friends and my labmates (and I am endlessly thankful that I joined a lab in which these are one and the same). I couldn't have asked for a more enthusiastic, comedic, supportive, and fun lab, and everyone I've worked with here has made at least a few very late nights in lab much easier to swallow. This lab serves as living proof of the merits of the "work hard, play hard" model. Of course, a resounding thank

you goes to Steve for welcoming me into his lab, throwing me opportunities left and right, and giving me the freedom to explore many opportunities while challenging me when I needed a push. I also sincerely thank my committee members for their invaluable feedback throughout the past several years, as their insightful suggestions have dramatically shaped this work. I am grateful for the entire experience I've had here, and I am thankful to feel so well-prepared for the next step in my academic and personal journey.

Chapter 1, in full, is a reprint of the material as it has been submitted for publication to Wiley Biotechnology Series *Industrial Biotechnology*. The dissertation author is the primary investigator and author of this chapter.

Chapter 2, in full, is a reprint of the material as it appears in *Biotechnology Letters* 2010. The dissertation author is the primary investigator and author of this paper.

Chapter 3, in full, is a reprint of the material as it appears in *Frontiers in Microbiology* 2014. The dissertation author is the primary investigator and author of this paper.

Chapter 4, in full, is a reprint of the material as it appears in *American Chemical Society Synthetic Biology* 2013. The dissertation author is the primary investigator and author of this paper. Reprinted with permission from *American Chemical Society Synthetic Biology*, copyright 2013 American Chemical Society.

Chapter 5, in full, is a reprint of the material as it has been submitted for publication to *Biotechnology Journal*. The dissertation author is the primary investigator and author of this paper.

Chapter 6, in full, is a reprint of material in preparation for submission for publication. The dissertation author is the secondary investigator and author of this paper.

Chapter 7, in full, is a reprint of material in preparation for submission for publication. The dissertation author is the primary investigator and author of this paper.

VITA

2009	Bachelor of Science, Department of Chemical and Biomolecular Engineering, Johns Hopkins University, Baltimore, MD
2009-2014	Graduate researcher, Division of Biological Sciences, University of California San Diego, La Jolla, CA
2014	Doctor of Philosophy, Division of Biological Sciences, University of California San Diego, La Jolla, CA

PUBLICATIONS

Specht EA, Nour-Eldin HH, Karunanithi PS, Ostrand J, Hoang KTD, Mayfield SP. High-throughput system for detecting improvements in homologous recombination efficiency in *Chlamydomonas reinhardtii*. **2014**. (In preparation.)

Nour-Eldin HH, **Specht EA**, Mayfield SP. ARG7-based cassette for high-level nuclear transgene expression in *Chlamydomonas reinhardtii*. **2014**. (In preparation.)

Specht EA, Nour-Eldin HH, Hoang KTD, Mayfield SP. An enhanced ARS2-derived nuclear reporter in *Chlamydomonas*. **2014**. (Submitted.)

Specht EA, Karunanithi KS, Ansari WS, Gimpel JA, Mayfield SP. Host organisms: algae. In *Industrial Biotechnology*; Wiley-Blackwell, Hoboken, NJ; 2014. (Submitted.)

Specht EA, Mayfield SP. Algae-based oral recombinant vaccines. Frontiers in Microbiology: Microbiotechnology, Ecotoxicology, and Bioremediation, 2014.

Rasala BA, Gimpel JA, Tran M, Hannon MJ, Miyake-Stoner SJ, **Specht EA**, Mayfield SP. Genetic engineering to improve algal biofuels production. In *Algae for Biofuels and Energy*; Springer, New York, NY; 2013.

Specht EA, Mayfield SP. Synthetic oligonucleotide libraries reveal novel regulatory elements in *Chlamydomonas* chloroplast mRNAs. American Chemical Society Synthetic Biology, 2013.

Gimpel JA, **Specht EA**, Georgianna DR, Mayfield SP. Advances in microalgae engineering and synthetic biology applications for biofuel production. Current Opinion in Chemical Biology, 2013.

Specht EA, Miyake-Stoner S, Mayfield S. Micro-algae come of age as a platform for recombinant protein production. Biotechnology Letters, 2010.

Masica DL, Schrier SB, **Specht EA**, Gray JJ. De novo design of peptide-calcite biomineralization systems. Journal of the American Chemical Society, 2010.

FIELDS OF STUDY

Major field: Biological Sciences

Studies in Molecular Biology
Professors Stephen Mayfield, Jens Lykke-Andersen, and Susan Golden

Studies in Computational Biology and Bioinformatics
Professors Steven Briggs and Theresa Gaasterland

ABSTRACT OF THE DISSERTATION

Improving the genetic tractability of the green alga
Chlamydomonas reinhardtii

by

Elizabeth Anne Specht

Doctor of Philosophy in Biological Sciences

University of California, San Diego, 2014

Professor Stephen P. Mayfield, Chair

Green algae present a unique platform for bioengineering and biomanufacturing; they grow rapidly, photosynthetically, and inexpensively and thus are suitable for large-scale cultivation, yet they are sophisticated eukaryotic cells with vast potential for introducing complex products, traits, or pathways. A long and growing list of publications has established that algae

are capable of producing large, intractable proteins that exceed the folding capacity of prokaryotic systems. More recently, studies involving metabolic engineering and systematic manipulation of the photosynthetic machinery have demonstrated that these algae are amenable to customization of complex endogenous processes such as photosynthetic anabolism and lipid metabolism.

While algae's successes and potential for industrial and therapeutic applications are covered extensively in this dissertation – including chapters on recombinant therapeutics, bioenergy applications, and oral vaccine development – these must be viewed in the context of the work that remains. Despite its moniker "the green yeast", even the model green alga *Chlamydomonas reinhardtii* falls short of other model organisms with regard to genetic tractability due to a relative lack of genetic engineering tools. The latter half of the dissertation addresses several of these shortfalls using innovative strategies inspired by synthetic biology approaches and high-throughput technologies.

In the chloroplast, gene targeting is routine but expression is regulated in translation. A better understanding of gene regulatory elements within transcripts was achieved at the intersection of a novel oligonucleotide library synthesis platform, high-efficiency seamless cloning, and next-generation sequencing technology. In the nuclear genome, a number of problems – including lack of facile reporters, robust promoters, and strong transgene expression cassettes – were addressed using optimized versions of endogenous genes, concomitantly alleviating concerns with genetically modified organism (GMO) regulations. Furthermore, the first reliable and reproducible strategy to measure incremental improvements in the gene targeting efficiency within the algal nuclear genome has been developed. This system is uniquely able to capture and characterize aberrant events at the recombination site, a phenomenon that had been predicted previously but proved difficult to elucidate unequivocally. Taken together, the advances described in this dissertation have significantly advanced the genetic malleability of the model alga *C. reinhardtii*, with potential application to additional vital algal species.

INTRODUCTION

Biotechnology, in its most basic sense, has been used by humans for thousands of years, since our ancestors' fateful discovery that yeast ferment sugars into alcohol. However, a new era of biotechnology has been ushered in by the field of molecular biology – the understanding at the molecular level how DNA sequences encode information conferring traits and phenotypes. With an understanding of the genetic code and the mechanisms by which this code is transcribed and translated to build the functional components of living cells, scientists began to hypothesize directed changes within these sequences to produce desirable traits and characteristics. These changes could initially be encouraged and selected for by induction of natural mutations and subsequent breeding, but it was not until the development of the first genetic tools that specific genetic changes could be produced by design and on demand.

Since the dawn of recombinant DNA technology, countless new tools and technologies have been pioneered to allow more precise seamless genome editing, more efficient cloning, more tightly controlled gene regulation, and even *de novo* synthesis of novel DNA constructs. These tools have been applied to model organisms ranging from simple prokaryotes like *E. coli* and single-cell eukaryotes like yeast to complex vertebrate organisms like zebrafish and mice. Among photosynthetic organisms, the model plant of choice has long been *Arabidopsis thaliana*, and several other agriculturally important species such as rice, corn, tomato, lettuce, and tobacco have routinely been engineered and transformed. Akin to the yeast of the eukaryotes, the plant kingdom also has single-celled model organisms that are simpler and more amenable to high-throughput approaches than multicellular terrestrial plants. These organisms are the unicellular algae, also known as microalgae. The most well-studied green alga is *Chlamydomonas reinhardtii*, which has long served as a model to study both photosynthesis and flagella function, and indeed is often touted as the "green yeast".

Recently, commercial interest in using microalgae for bio-manufacturing has spurred large amounts of research into advancing microalgae genome editing, strain selection, and

metabolic engineering on par with the research that has enabled large-scale recombinant production in hosts like yeast and bacteria. Prior to commercial realization of these applications for microalgae, an enhanced set of tools for complex genetic engineering is required, preferably that can be applied to a broad host range. After a discussion of several current and potential applications, my research towards both chloroplast and nuclear genome engineering is described, along with its ability to advance the field of algal engineering towards meeting these needs.

Chapter 1 provides a broad overview of the potential of algae as an industrial host organism for a range of valuable products and commodities. Importantly, because of their photosynthetic nature, their amenability to cultivation on non-arable land, and the ability of marine strains to circumvent the need for fresh water, algae are toted as an environmentally friendly alternative for producing vast ranges of bio-based products. Indeed, they have demonstrated utility already in the commercial food processing sector, and for supplementation in both human and aquaculture markets. Chapter 1 also explores the potential of algae for producing fungible, so-called "drop-in" biofuels, and the challenges associated with growth at a meaningful scale within this market. These applications are not limited to *Chlamydomonas*, and indeed this organism does not exhibit many of the robust growth characteristics required of a commercial strain, but nevertheless much of the research into the strain selection and engineering is being pioneered in *Chlamydomonas* and then translated into alternative production strains.

Chapter 2 introduces algae, and specifically *Chlamydomonas*, as a platform for producing recombinant proteins primarily for therapeutic and medicinal applications. To date, most of the recombinant proteins produced in algae have been expressed from the chloroplast genome, as it is more amenable to high transgene product accumulation. These proteins have included vaccine antigens (covered in more detail in Chapter 3), mammalian colostrum proteins, human antibody-toxin conjugates, and human hormones and growth factors. The algal chloroplast has proved to be an ideal platform for the production and high-level accumulation of these high-value products,

and is especially adept at folding complex and heavily disulfide-bonded proteins due to its sophisticated protein folding machinery. Ongoing advances in our understanding of chloroplast gene regulation (covered in Chapter 4) will continue to enhance the feasibility of this platform.

In Chapter 3, one particular recombinant therapeutic application is covered in depth: using microalgae to produce oral recombinant vaccines. The first recombinant vaccine, using yeast to express a subunit of the hepatitis B viral capsid protein, marked a new era of vaccines in which recombinant production eliminated risks associated with purifying the antigen from infected humans or animals. Recently, experts recognize the need to move towards orally delivered vaccines, particularly for diseases affecting large populations with limited financial resources and medical infrastructure and personnel. Oral vaccines have the potential to elicit both systemic and mucosal immunity, serving as an optimal line of defense at the site of entry of most pathogens, and their ease of administration increases compliance rates and vaccine schedule adherence. Algae are ideal hosts for producing recombinant oral vaccines because they are edible, obviating the expensive step of antigen purification, yet they avoid food supply contamination concerns associated with vaccine production in crop plants. Adjuvants and genetically-linked domains for enhanced immunogenicity are still areas in need of extensive investigation, but the successes described in Chapter 3 illustrate the promise and rapid progress in this field.

Nearly all of the therapeutic applications detailed in Chapters 2 and 3 rely on high-level transgene expression from the algal chloroplast genome. However, a number of factors currently hamper transgene expression levels within this organelle, including strong auto-attenuation when using endogenous regulatory elements to drive transgene expression. To gain a better understanding of specific gene regulatory elements involved in chloroplast translation to address this obstacle, Chapter 4 describes a synthetic biology approach to characterize novel sequence or structural elements in chloroplast transcript untranslated regions (UTRs) in a high-throughput, exhaustive, and non-biased manner. This strategy was validated using a previously well-studied

UTR – that of the photosystem I gene *psaD* – and was found to in fact provide additional information concerning the precise boundaries of elements that had previously been identified in a piecemeal fashion, often with artifacts introduced by limitations such as restriction site availability. The synthetic biology approach, by contrast, allowed study of the entire UTR in discrete segments and a relatively new seamless cloning technique eliminated spurious artifacts from the library construction and cloning steps. This method was then applied to a previously uncharacterized UTR – that of *psaA*, encoding another major subunit of photosystem I – and used to define positive and negative regulatory elements within the UTR, using millions of reads of data from next-generation sequencing of pooled clones with high and low expression levels of a reporter. This aggregate information was then used to construct a synthetic UTR, which was indeed able to drive higher expression, illustrating the utility of this approach for designing novel regulatory elements that may evade endogenous feedback. This was the first demonstration of using a non-endogenous UTR sequence to drive transgene expression within algal chloroplasts.

Although the chloroplast genome has been the site of most engineering performed to date, it is limited in utility with regards to metabolic engineering of pathways that exist in multiple cellular compartments, as gene products synthesized in the chloroplast cannot escape the organelle. Likewise, chloroplast-produced proteins cannot be exported and cannot be post-translationally modified, as they do not move through the endoplasmic reticulum for processing. Therefore, for more complex engineering, the nuclear genome must also be accessible for precisely controlled manipulation. Nuclear genome engineering in algae has encountered its own set of obstacles, including gene silencing of introduced transgenes, high variability between clones due to random gene integration, and a relative lack of well-characterized reporters, promoters and expression cassettes. Chapters 5 through 7 describe research that addresses many aspects of these challenges.

Chapter 5 is a brief but timely report of high utility to the algal community, describing an improvement upon an endogenous reporter gene that was previously isolated but not of general use due to overall low expression, rendering it valuable only for reporting the activity of highly inducible promoters. When cloned as a cDNA instead of a large genomic fragment, the ARS2 gene encoding the endogenous *C. reinhardtii* arylsulfatase gene is now much more suitable as a reporter due to ease of cloning, higher transformation efficiency, and its ability to express at much higher levels under a constitutive promoter. Furthermore, the arylsulfatase assay can be performed without expensive equipment as required for other available reporters such as fluorescent proteins. To demonstrate one application of this improved reporter, it was used to characterize the ARG7 promoter, which previously had not been used to drive expression of anything other than the native ARG7 gene. The improved ARS2 construct was able to show that the ARG7 promoter not only expresses to equally high levels as the previous best nuclear promoter (a chimeric HSP70-RBCS2 promoter/5' UTR) but also exhibits more robust expression, with a higher proportion of transformants expressing the reporter at intermediate levels. Thus, this endeavor not only added an enhanced nuclear reporter to the genetic engineering toolkit but also added a strong, robust promoter that can be used in conjunction with the HSP70-RBCS2 promoter in multi-gene cassettes, as required for complex metabolic pathway engineering.

Chapter 6 describes another pertinent improvement to a nuclear genetic engineering tool, a selectable marker, and its ability to overcome nuclear transgene silencing. The ARG7 gene encoding a key component in the pathway of arginine synthesis was previously used as a selectable marker in ARG7 mutant strains by complementing with the entire ARG7 genomic fragment. This construct was intractable for vector manipulation, and transformed with very low efficiency. The new cDNA-derived version, at less than 1/3 the size of the genomic fragment, transforms far better than the original construct. Furthermore, its utility was tested to drive transgene expression to high levels by transcriptional fusion to the gene of interest using a self-

cleaving viral peptide. This strategy had previously been shown to drive high transgene expression when coupled to the zeocin resistance gene *Sh-ble*, but alternative selections had not been tested, limiting the number of transgenes in a single strain that could be expressed using this strategy and necessitating the introduction of antibiotic resistance genes. When driven by its own promoter, the ARG7 gene was indeed able to drive high transgene expression via transcriptional fusion, providing an alternative to the *Sh-ble* system.

Even with improved genetic tools including better reporters, stronger promoters, and selectable markers to help overcome gene silencing and positional effect, nuclear genome engineering is still hampered by a major barrier: we are unable to efficiently target genes via homologous recombination (HR). Homologous recombination, a necessity in nearly all other model organisms routinely engineered, does occur in vegetative algal cells; however, its rate is so low that targeted insertions cannot practically be identified without selection. An increased HR rate would allow for facile identification of targeted gene knockouts, rather than laborious insertional library creation and screening to study loss of function of genes of interest. Furthermore, targeted insertion of transgenes into a pre-determined site would allow neutral sites to be verified and reliably targeted to prevent spurious off-target effects and to enable direct comparison between independent transformants without variability from positional effect. Prior attempts to measure the HR rate in *C. reinhardtii* have been met with low reproducibility and ambiguous results or an inability to verify true recombination. Furthermore, due to the extremely rare nature of HR events in algae, a high-throughput method of assessing recombination is essential, but none had previously been developed, confounding the problem of low reproducibility and making it impossible to exhaustively test approaches used in other model organisms for incrementally increasing HR efficiency.

In Chapter 7, a novel system is described that can reliably and reproducibly detect HR events in an extremely high-throughput fashion, screening hundreds of thousands of

transformants in a single experiment. The high fidelity of this system and the ability to amplify and sequence the precise region of recombination makes this the first method capable of unambiguously identifying and verifying recombination events. Furthermore, the inclusion of an intron within the homology allows us to capture even HR events in which the recombination is imprecise, resulting in insertions or deletions. These imprecise events have not been well-characterized in *C. reinhardtii* due to the inflexible nature of previous attempts to observe HR. Using this system, we were able to systematically assess two methods that previous published reports claimed increased HR efficiency in algae. We found that one method – transformation by particle bombardment instead of glass beads or electroporation – was unable to produce the HR rate claimed in the original report. On the other hand, a different approach of using single-stranded DNA did demonstrate potential with our system, as well as a novel method of incubating the DNA prior to transformation with a recombination protein. Because of the high-throughput screening capability of this system, even small incremental improvements to HR efficiency can be detected, enabling synergistic approaches of combining multiple strategies. In this way, this system may be able to catch algae up to the state of gene targeting in other organisms more rapidly, building upon advances pioneered in other models.

Altogether, these advances represent a significant step forward in the field of algal genetic engineering, both of the chloroplast genome and the nuclear genome. Indeed, the ability to introduce transgenes at precise locations and to regulate them in a controlled and predictable fashion is crucial for both genomes in order to achieve large-scale metabolic engineering of complex pathways, due to the intertwined evolutionary history of algae's multiple genomes. These novel genetic tools, as well as the approaches developed to discover them – including elements of synthetic biology and high-throughput screening – significantly expand the understanding of algal biology as well as its utility as a platform for practical bio-manufacturing and biofuel applications.

CHAPTER 1:

ALGAE AS INDUSTRIAL HOST ORGANISMS

Algae encompass an enormously diverse set of photosynthetic organisms, which have relatively recently been investigated for industrial-scale production of bio-products. Their fast growth and minimal nutritional requirements make them suitable candidates for large-scale applications such as biofuel production, yet they are also capable of producing complex, high-value products such as therapeutic proteins, vaccines, and nutritional supplements. As advances are made in the toolkit for algal genetic manipulation and metabolic engineering, algae are likely to emerge as a viable production platform for a wide range of products.

1.1 Introduction to algae as an industrial organism

1.1.1 Algal diversity

The term "algae" designates a wide range of photosynthetic organisms found throughout all aquatic environments, both freshwater and marine, as well as wet terrestrial environments like soil and symbiotically within lichens. They range from large macroalgae – such as seaweed and kelp – to both eukaryotic and prokaryotic microalgae. Red algae, brown algae, diatoms, cyanobacteria (blue-green algae) and green algae are all classified in distinct phyla with sometimes hundreds of genera within them (*1*). More than half a million species have been discovered, but very few have been cultured in the laboratory and even fewer have been genetically characterized. This vast, untapped reserve of biological diversity has spawned much interest in natural products discovery, as algae produce a large number of unique molecules as secondary metabolites (*2*). Beyond metabolic diversity, algae's vast range of ecological niches provides potential for identifying tolerance and resistance traits that are crucial for large-scale growth and robustness.

Some species of algae are already grown at industrial scale, mostly for harvesting food additives like agar and carrageenan or for direct consumption for their metabolites like omega3

fatty acids and anti-oxidant carotenoids. In some cultures, algae have long played a notable role in cuisine, mostly the seaweed macroalgae. Recently, significant investments have been made in exploring microalga potential for biofuels production, and several commercial and academic pilot facilities exist for developing the growth, harvesting, and lipid extraction technologies that will be required for industrial scale production.

1.1.2 Scalability and growth rate

One of the most advantageous traits of algae is their potential for growth at agricultural scale, but with much higher biomass accumulation per acre than terrestrial plants and crops. On the macroscopic scale, kelp is well-known for its extraordinary growth rate – up to half a meter a day. Microalgae also exhibit short doubling times, and cultures of eukaryotic green algae – the species most explored for large-scale biofuel production – can be grown to saturated density and harvested in as little as a week. Furthermore, algae are not subject to growing seasons to the same extent as most terrestrial plants, making year-round growth possible. Most commercial pilot facilities are situated on non-arable land, often in desert areas, where they do not compete with food crops for land or resources.

1.1.3 Genetic malleability

Although relatively few of the hundreds of thousands of algal species have been genetically transformed, there has been rapid progress in the last decade of research. Some algae, mostly in the cyanobacterial groups, are naturally competent for transformation (*3*). Extensive research has been done with these species – notably in the *Synechococcus*, *Synechocystis*, and *Anabaena* genera – to identify neutral sites and develop transformation vectors with appropriate gene regulatory elements (*4-6*). Section 2 below focuses on the genetic engineering of both the chloroplast and nuclear genomes of green algae, particularly the model species *Chlamydomonas*

reinhardtii. This is an active area of research, and new genetic engineering tools – including reporter genes, gene targeting methods, synthetic promoters and regulatory elements, and high-expression transgene cassettes – are being published each year. Recently there has been a strong push for tools with broad host species compatibility, such that the advances pioneered in the model organisms can be applied to strains more suitable for industrial-scale production.

1.2 Algal genetic engineering

At present, the species of green algae most commonly and easily transformed is *Chlamydomonas reinhardtii*. This model organism has been studied for decades, and thus the largest set of genetic tools exist for it. Our detailed discussion of algal genetic engineering will focus primarily on this organism, but recently many of these methods have been adapted to other species. Current progress towards transforming other species is covered in section 2.3. Green algae contain three genomes – nuclear, chloroplast, and mitochondrial – all of which have been transformed; however, almost all genome engineering has been targeted to the nuclear (Sec. 2.1) and the chloroplast (Sec. 2.2) genomes.

1.2.1 Nuclear genetic engineering

The complete nuclear genome of *C. reinhardtii* was made available in 2007 (*7*). While transformation of this genome is facile, there are several considerations that impact transgene expression from the nuclear genome, so it has not been as well exploited for transgenics as the chloroplast genome. Codon usage plays an important role in transgene expression, as the nuclear genome is highly GC rich (64% GC) (*8*). Codon optimization to reflect codon bias in the nuclear genome can substantially increase transgene expression (*9, 10*), and several additional advances have been made to increase expression (see Sec. 2.1.3). A distinct advantage of nuclear expression is the potential to add posttranslational modifications, including glycosylation, which

is not possible with chloroplast expression. Other potential benefits include targeting to cellular components and easier protein purification through secretion in media from cell-wall deficient strains (*11*).

1.2.1.1 Nuclear transformation methods

Nuclear transformation in *C. reinhardtii* was first demonstrated in the early 1980s using the yeast *ARG4* gene to complement an *arg7* mutation (*12*). Methods of nuclear transformation include particle bombardment or biolistic transformation (*13, 14*), agitation with glass beads in cell wall deficient strains (*15*), electroporation (*16*), agitation with silicon-carbide whiskers (*17*), and biologically-induced gene transfer with *Agrobacterium tumefaciens* (*18, 19*). *C. reinhardtii* is generally not susceptible to viral infection so gene transfer using viral agents has not been possible. Electroporation is commonly used for nuclear transformation due to its high transformation efficiency, up to over 10,000 transformants per microgram of DNA (*16*). Recently, droplet electroporation on microfluidic chips has been used to improve transformation efficiency by three orders of magnitude for wall-less mutants and over two orders of magnitude for wild type cells compared to bulk phase electroporation (*20*). Translocation by cell penetrating peptides (CPP), in particular the artificially synthesized pVEC (peptide vascular endothelial cadherin), is another potential method for successful delivery of macromolecules like DNA into cell wall-intact strains of *C. reinhardtii* (*21*).

1.2.1.2 Regulation of nuclear transgene expression

Integration of foreign DNA into the nuclear genome primarily occurs by non-homologous end joining (*22, 23*) because the rate of homologous recombination is very low under normal growth conditions (*24-26*). Packaging of transgenes in distinct chromosomal domains (*27*) and chromatin compaction through histone modifications such as methylation on

nucleosomes (*28*) are thought to contribute to epigenetic silencing, though the full range of silencing mechanisms is still under investigation. For non-silenced genes, nuclear expression is regulated predominantly at the transcriptional level, necessitating well-characterized robust promoters (*27*).

Nuclear promoters used to express transgenes include *β_2TUB* (*29*), *NIT1* (*30*), *PSAD* (subunit of photosystem I) (*31*), *LHCB1* (light harvesting complex of photosystem II) (*32*), and *LIP* (light-inducible protein gene) from *Dunaliella sp.* (*33*). The *HSP70A* (heat shock protein 70A) promoter acts as an expression enhancer (*34*), and *HSP70A-RBCS2* (rubisco small subunit) promoter fusions are the most effective to date (*35-39*). Recently, *NOS* (nopaline synthase from *Agrobacterium*) and *CaMV35S* (cauliflower mosaic virus 35S) have been used to drive transgene expression, although transformation efficiency is much lower than that of the *HSP70A-RBCS2* chimeric promoter (*40*).

1.2.1.3 Tools for increased nuclear gene expression

Several strategies have been pioneered for evading gene silencing to achieve higher nuclear transgene expression, with varied success. Endogenous heat shock factor 1 (*HSF1*) increases expression of transgenes by counteracting epigenetic silencing through scaffold formation and recruitment of RNA Pol II to *HSP70A* promoters (*39*). Insertion of one or more copies of the intron sequences from the *RBCS2* gene (*41, 42*) and use of linearized tandem vectors for co-expression (*43*) also significantly increase transgene expression. Recently, the viral 2A self-cleaving peptide from the foot-and-mouth disease virus has been used to drive high expression though transcriptional fusion to the selectable marker, while creating distinct polypeptides. This strategy is particularly effective with the zeocin resistance gene *ble* because the gene product binds the antibiotic stoichiometrically instead of enzymatically inactivating it, necessitating high expression of the resistance gene (and, thus, of the gene of interest) to survive

on zeocin (*44*). Transcriptional terminators have also been shown to influence transgene expression. The *psaD* terminator paired with its respective *psaD* promoter increased accumulation of the recombinant protein butyrylcholinesterase to 0.4% (*45*).

1.2.1.4 Selection and screening methods

The first selection markers were genes used to rescue auxotrophic mutants. *ARG7* for arginine biosynthesis (*46*) and *NIT1* for nitrate reductase (*47*) are well known examples. Antibiotic resistance-conferring genes were developed later, and include genes *sh-ble* for zeocin and phleomycin (*48*), *CRY1* for emetine (*49*), *aadA* for spectinomycin (*23*), *aphVIII* for kanamycin, paromomycin and neomycin (*50*), and *aph7* for hygromycin (*51*). Most transformation vectors include a selectable marker, but co-transformations can also be done using a selectable marker plasmid and a separate gene of interest plasmid; typically about a quarter of the selected transformants receive both plasmids.

A handful of nuclear reporters are available for a variety of chemical or fluorescent readouts. The endogenous arylsulfatase gene ARS2, which is tightly regulated and normally only expressed in sulfur starvation, cleaves a chromogenic compound for easy detection (*29*) and has been used to monitor inducible promoter activity (*52*). Luciferase-encoding genes from *Renilla* (*32*) and *Gaussia* (*53*) have both been used, but *Gaussia* luciferase may be more sensitive in *Chlamydomonas* (*54*). The *RSP3* gene encoding a radial spoke protein required for flagellar motility was used as a reporter in mutant strains lacking endogenous *RSP3* that are paralyzed under aerobic conditions (*55*). A suite of codon-optimized fluorescent protein genes have been characterized in *C. reinhardtii* as well, including several GFP variants, mCherry, tdTomato, and mCerulean (*56*). Many common reporters are listed in Table 1.1 below.

Table 1.1 Reporter genes for nuclear expression

Reporter	Gene	Source	Reference	Notes
Arylsulfatase	*ARS2*	*Chlamydomonas reinhardtii*	Davies et al. 1992 (*29*) ; de Hostos et al. 1988 (*57*)	Excreted into media and cleaves chromogenic substrates
Green Fluorescent Protein	*cGFP*	*Aequorea victoria*	Fuhrmann et al. 1999 (*9*)	GFP codon optimized for *C. reinhardtii*
Gaussia luciferase	*G-Luc*	*Gaussia princeps*	Shao and Bock, 2008 (*53*)	Codon-optimized luciferase
mCherry; tdTomato	*mCherry; tdTomato*	*Discosoma striata*	Rasala et al. 2013 (*58*)	RFP variants optimized for *C. reinhardtii*
Venus; mCerulean	*Venus; mCerulean*	*Aequorea victoria*	Rasala et al. 2013 (*58*)	GFP variants optimized for *C. reinhardtii*

1.2.1.5 Gene targeting and genome editing

Reverse genetics is critical for determining gene function by knocking out endogenous genes and observing phenotypic results. Since homologous recombination is a rare event (*26*), targeted knockout is not yet possible so knockouts are obtained through insertional mutagenesis (*59*). However, genes can be knocked down using RNA interference (RNAi) due the presence of the required machinery, including Dicer and Argonaute proteins, in algae (*60*). The two types of RNA involved in RNAi-mediated gene silencing are small interfering RNAs (siRNAs) and microRNAs (miRNAs). siRNAs are processed from double-stranded RNA (dsRNA) created through sense/antisense pairing or hairpin formation by inverted repeats (*61*) and have been used to knockdown gene expression (*62, 63*). Artificial miRNAs (amiRNAs) were created to mimic endogenous single-stranded transcripts of naturally-occurring miRNAs (*64*), promoting gene knockdown (*65*).

Genome editing is also not routinely possible in the nuclear genome. Recently, zinc finger nucleases have been employed in *C. reinhardtii* (*66*), but the process for constructing them for each target site requires significant effort and is costly. Other strategies such as TAL effector

nucleases and CRISPR/Cas genome editing have not yet been demonstrated in algae, though multiple groups are working on developing these technologies for algal systems.

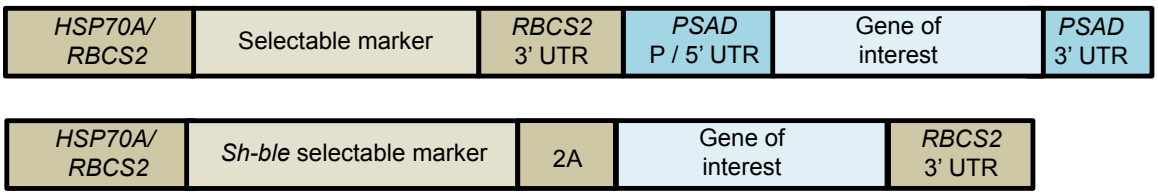


Figure 1.1 Schematic of two typical expression vectors for *Chlamydomonas* nuclear transformation. Above, a standard nuclear expression vector, including a selectable marker cassette and a gene of interest cassette, each with their own promoters, 5' UTRs, and 3' UTRs. Below, vector in which the gene of interest has been transcriptionally fused to zeocin resistance by the 2A self-cleaving peptide to drive high gene of interest expression.

1.2.2 Chloroplast genetic engineering

Due to the endosymbiotic evolutionary history of the chloroplast, it has its own complete genetic system, including a circular chromosome, and transcription and translation machinery including all tRNAs and prokaryotic-like ribosomes. Over time, many genes that originated in the chloroplast genome have moved into the nuclear genome, but their gene products are still targeted for chloroplast localization; there is substantial cross-talk between the nuclear and plastid genomes. In contrast to most terrestrial plants that contain many chloroplasts per cell, *C. reinhardtii* contains a single large chloroplast that occupies over 40% of cell volume (*67*). The chloroplast genome is a circular 203kb chromosome encoding 99 genes (*68*). Greater than 20% of the genome consists of repetitive DNA and different classes of short dispersed regions between the coding regions containing genes for photosynthetic proteins and transcriptional-translational machinery (*68*). Unlike the GC-rich nuclear genome, the chloroplast genome is only about 35% GC (*68*). The chloroplast contains about 80 copies of its genome, and stable transformation requires that all copies contain the transgenic construct or else reversion may be observed (*67*); this stable state is referred to as homoplasmy.

1.2.2.1 Chloroplast transformation methods

Particle bombardment remains the most effective method for transforming the chloroplast of *C. reinhardtii,* and was first demonstrated in the 1980s (*69*). DNA-coated gold or tungsten particles are bombarded at high velocity onto a plate with a lawn of algae. The particles penetrate both the cell wall and the chloroplast membranes, where the foreign DNA can access the chloroplast genome (*69*). The process is performed by a biolistic gene gun or in a vacuum chamber. Transformation by agitation with glass beads in cell-wall deficient strains has also been used, albeit with lower transformation efficiency compared to microparticle bombardment (*70*). Integration of transgenes occurs readily through homologous recombination, requiring flanking regions of endogenous sequence homologous to target insertion site (*69*). The minimal flanking sequence around the recombinant gene should be about 1kb to ensure successful integration (*71*). Since homologous recombination occurs in the chloroplast, knock down of endogenous genes is possible in addition to targeted gene insertion.

1.2.2.2 Gene regulation in the chloroplast

Unlike the nuclear genome, in the chloroplast most genes are predominantly regulated at the post-transcriptional level (*72*). The 5' and 3' untranslated regions (UTRs) largely influence translation rate (*73*) and mRNA stability (*74*) respectively. Gene expression in the chloroplast is regulated by post-transcriptional processes influencing RNA stability, RNA processing, cis- and trans-splicing, and translation (*75*).

Thus far, most studies have been devoted to identifying and utilizing endogenous chloroplast promoters and UTRs for heterologous gene expression (*73*). The most effective promoter and UTR for heterologous gene expression is from the *psbA* gene (encoding the D1 protein of Photosystem II), but transgene product accumulation is only significant in *psbA*-deficient strains because of auto-attenuation of translation in the presence of an intact

photosystem. However, photosynthesis can be restored by reintroducing the *psbA* coding region under the control of different regulatory elements, allowing transgene expression to be maintained (*76*). Other promoters and UTRs from *atpA, rbcL,* and *psbD* have been used to drive heterologous gene expression but with varied success (*77*). However, because of high recombination rates within the chloroplast, there is some risk of instability from introducing transgenes surrounded by endogenous regulatory sequences. Because of this and the auto-attenuation concerns, novel synthetic UTRs have been derived using discrete elements extracted from high-expressing members of a mutant UTR library (*78*).

1.2.2.3 Chloroplast selection and reporters

Chloroplast transformants can be conveniently selected by restoring photosynthesis in photosystem knockout strains, and *atpB* (*69*), *rbcL* (*79*), and *psbH* (*80*) have all been used as selectable markers in this manner. Antibiotic resistance is also commonly used for selection and can be conferred through heterologous gene expression or induced point mutations in endogenous genes. The *aphA6* gene (from *Acinetobacter baumannii*) for kanamycin resistance (*81*) and *aadA* (from *E. coli*) for spectinomycin and streptomycin resistance (*82*) have been used in multiple algal species. Point mutations in the 16S and 23S ribosomal RNAs confer resistance to streptomycin and erythromycin respectively (*83*), and mutations in the endogenous *psbA* gene can confer resistance to herbicides like atrazine (*84, 85*). Chloroplast reporters, like those used in the nucleus, include fluorescent proteins and luciferase variants (see Table 1.2). However, the same gene cannot be used in both compartments due to the stark difference in codon bias.

Table 1.2. Reporter genes for chloroplast expression

Reporter	Gene	Source	Reference
Renilla luciferase	*R-luc*	*Renilla reniformis*	Minko et al. 1999 (*86*)
GFP	*CrGFP*	*Aequorea victoria*	Franklin et al. 2002 (*87*)
Bacterial luciferase	*luxCt*	*Vibrio harveyi*	Mayfield and Schultz 2004 (*88*)
β-glucuronidase (GUS)	*uidA*	*Escherichia coli*	Klein et al. 1992 (*89*)

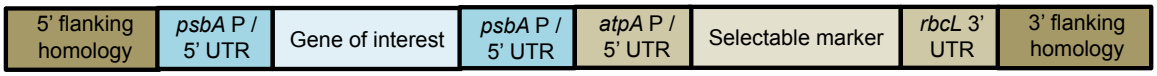


Figure 1.2 Schematic of a typical *Chlamydomonas* chloroplast transformation vector. The gene of interest and the selectable marker are expressed in two separate cassettes. Two sites are commonly targeted: the 3HB site is a neutral site in the chloroplast genome that tolerates transgene insertion, and the *psbA* site is replaced by a transgene for high expression in *psbA* knockouts. The flanking homology regions contain sequence homologous to the chloroplast genome at the intended site of integration.

1.2.3 Other transformable species

While *C. reinhardtii* has been the most popular species for pioneering genetic engineering work, in fact many species – encompassing a broad swath of the diversity among microalgae – have been successfully transformed. Most of these species have only been transformed in the nuclear genome, as penetrating the chloroplast often proves more difficult and requires accurate plastid genome sequence to obtain flanking regions for integration by homologous recombination.

1.2.3.1 Parameters for transformation optimization

Electroporation is generally the first technique tried in most algae for nuclear transformation, while chloroplast transformation often requires a more aggressive approach like biolistic particle transformation. With both of these methods, many parameters can be optimized to increase transformation efficiencies for each species. With electroporation, the voltage, resistance, and current can all be specified, as well as the buffers in which the cells are

transformed and subsequently recovered. With particle bombardment, the particle size and material are variable, as well as factors affecting the velocity of the particles such as the vacuum strength, gas pressure, and even the distance between the biolistic gun and the agar plate.

1.2.3.2 Species successfully transformed

For some species, notably blue-green algae (cyanobacteria), DNA is taken up directly from the media or through bacterial conjugation, as they are naturally competent for transformation. Even diatoms – which are generally surrounded by a tough silica coat – have been successfully penetrated and transformed after some troubleshooting. At least six species of diatoms have been transformed – all in the nuclear genome (*90*) – including *Phaeodactylum tricornutum* (*91*) and *Thalassiosira weissflogii* (*92*). Several species of *Chlorella* have been transformed as well, most notably *Chlorella vulgaris* (*93*). Green algae in the *Dunaliella* genus, which have significant commercial potential and have been transformed in both the nuclear and chloroplast genomes, include *Dunaliella salina* (*94, 95*) and *Dunaliella tertiolecta* (*96*). Two species of dinoflagellates have been transformed using silicon carbide whiskers (*97*), as well as two species of red algae, including chloroplast transformation in *Porphyridium* spp. (*98, 99*). These species represent widely diverse phyla and genera, and this broad range of transformability is a testament to the potential for applying the genetic engineering tools described in Sec. 2.1 and 2.2 to species beyond *C. reinhardtii*.

1.3 Therapeutic and nutraceutical applications

Algae are already harvested for natural products, including astaxanthin from *Haematococcus* for supplements and fish feed, and whole-cell *Chlorella* and *Spirulina* to be blended into specialty drinks and supplements. It is also likely that one of the first commercial applications of engineered algae will be for the therapeutic or nutraceutical industries. These

products have much higher value per unit mass than biofuels or bulk animal feed, so their production is economically viable even with relatively low expression levels of recombinant proteins. It is hoped that investment in these specialty products will spur further advancements in algal genetic engineering and in the downstream engineering needed to harvest and process at scale. These advancements can, in turn, eventually be applied to larger systems such as algal biofuel production, thereby essentially subsidizing the sizeable development costs of such a platform.

1.3.1 Protein therapeutics

Protein therapeutics comprise the fastest-growing sector of the booming pharmaceuticals industry (*100*), and current production platforms have their limitations. For example, mammalian cell culture is expensive and susceptible to contamination; prokaryotic systems cannot properly fold many complex human proteins; and yeast often hyperglycosylate proteins, making them unsuitable for many therapeutic applications. In the past two decades, interest in terrestrial crop plant medicinal production ("molecular pharming") has waxed and waned. Hostility towards this approach usually derives from concerns of gene flow or gene escape into other crops, leading to contamination of the food supply.

Besides the lack of pollen and the ability to grow in enclosed systems – alleviating gene flow concerns – algae have several other advantages over terrestrial plant-derived vaccines. They have a much more rapid strain generation time, allowing transformation and verification of a transgenic strain in a matter of weeks instead of months for crop plants. They also accumulate biomass more quickly per unit area, and can be harvested continually instead of seasonally. Finally, growth conditions can be better controlled, leading to less fluctuation in transgenic protein product content than is observed in crop plates from soil and climate variations. Several

human genes have been transformed into algal species, predominantly but not exclusively into *C. reinhardtii*.

1.3.1.1 Human antibodies and antibody-drug conjugates

Human antibodies are large, complex proteins that require proper folding and disulfide bond formation and precise incorporation of multiple chains, presenting difficulty for many production platforms. However, *C. reinhardtii* chloroplasts have been engineered to produce human antibodies or parts thereof since 2003, when a monoclonal antibody was produced as a single chain instead of a separate heavy and light chain. The antibody folded properly within the chloroplast and recognized its target antigen, herpes simplex virus (*101*). A few years later, the first antibody composed of separate heavy and light chains was produced inside algal chloroplasts; the chains assembled into full-length, functional antibodies *in vivo* (*102*).

Antibody-drug conjugates are a class of molecules in which a humanized or recombinant human antibody is attached to a drug molecule, which is often a small-molecule toxin. Several of these drugs have been approved as cancer treatments, wherein the antibody specifically recognizes cancer cells and is endocytosed, releasing the toxin and triggering cell death. Recently, algal chloroplasts have been shown to be a suitable platform for producing a novel class of antibody-drug conjugates in which the toxin is a protein that is genetically linked to the antibody, not a chemically-conjugated small molecule. These toxins include exotoxin A (*103*) and gelonin (*104*), both of which inhibit eukaryotic ribosomes, so they cannot be produced within the normal eukaryotic platforms that are required for proper folding of the antibody portion of the molecule. The chloroplast, in contrast, is not susceptible to these toxins because of the prokaryotic nature of its ribosomes, so it is a unique host in that is has the complexity to make antibodies yet the simplicity of a bacterial platform to evade the effects of the toxin domains. These molecules have

indeed been made in *C. reinhardtii* and tested in mice, exhibiting reduced tumor size and improved mouse survival over controls (*103, 104*).

1.3.1.2 Other chloroplast-produced protein therapeutics

Several human therapeutic proteins in addition to antibodies have been produced in algae. A human metallothionein protein – one of a class of proteins with metal-binding capacity and an ability to capture oxidative radicals – was introduced into the *C. reinhardtii* chloroplast and appeared to confer increased resistance to UV stress, though only temporarily (*105*). Vascular endothelial growth factor (VEGF) stimulates angiogenesis and its up-regulation is observed in many cancers, but it also plays a role in wound healing and is decreased in emphysema patients. VEGF was produced in *C. reinhardtii* at 2% total soluble protein, and was shown to be biologically active in a receptor-binding assay (*106*). In the same study, high mobility group protein B1 (HMGB1) – which is also involved in wound healing – was accumulated at similar levels and was able to recruit fibroblasts, demonstrating correct biological activity. Interestingly, erythropoietin was not well-expressed from the *C. reinhardtii* chloroplast although it has been expressed from the nuclear genome of another green alga (see Sec. 3.1.3), and insulin – the first approved recombinant protein drug – was also difficult to detect (*106*), indicating that there is still room for improvement in factors affecting recombinant protein expression in algae.

1.3.1.3 Nuclear-expressed protein therapeutics

Glycosylation can impact protein folding and stability, and is often essential for protein function or recognition. Proteins synthesized in the chloroplast are not glycosylated, necessitating nuclear expression for all therapeutics that require glycosylation, which constitute about 40% of all currently-approved protein therapeutics (*100*). Although nuclear transgene products do not accumulate to very high levels at present (see Sec. 2.1.2 and 2.1.3), there are some examples of

success. The hormone erythropoietin, which regulates blood cell differentiation and is used in anemia treatment, was expressed from the *C. reinhardtii* nuclear genome and secreted into the culture media, accumulating to 100 μg L^{-1} (*107*). In another species of green algae, *Chlorella ellipsoidea*, an antimicrobial peptide was expressed from the nuclear genome and exhibited antimicrobial activity against both gram-negative and gram-positive bacteria as well as a fungus (*108, 109*).

1.3.2 Nutraceuticals and nutritional supplements

The term nutraceutical – a portmanteau of "nutrition" and "pharmaceutical" – can refer to any supplement, health product, or formulated food intended to provide prevention against chronic disease or promote general well-being. In this sense, they are often considered similar to supplements. Both of these categories have current and potential future utility for industrial-scale algal production of wild type and/or engineered strains.

1.3.2.1 Omega-3 fatty acids

One of the most popular supplements is derived from fish oil or krill oil, delivering a high concentration of omega-3 fatty acids – in particular EPA and DHA – which are essential fatty acids and have notable importance for brain and cardiovascular health. Many algae, at the bottom of the food chain for these fish and krill, produce EPA (*110*) and DHA (*111*) themselves and are currently being grown for the purpose of supplementation (*112*). However, there has been significant interest in engineering other algal species – ones that are more amenable to large-scale growth and lipid extraction – to produce these valuable fatty acids (*113*). Often, only one or a few additional lipid biosynthetic enzymes need to be introduced (*114*), so this endeavor is also a useful exercise in metabolic engineering for aspirations of modifying more complex pathways (see biofuel lipid engineering in Sec. 4.1).

1.3.2.2 Source of organic selenium

Selenium, once thought to be toxic, is now known to be required in trace amounts by humans. Selenium in the form of selenocysteine is incorporated into at least 25 human proteins, several of which have only recently been discovered (*115*), and selenium must be consumed through the diet for normal function of these proteins. Organic selenium is less toxic and more readily absorbed than inorganic selenium, so the optimal source is to consume organisms that produce selenoproteins themselves. It was long thought that plants – unlike bacteria, archaea, and animals – did not contain any selenoproteins (*115*). However, they have now been definitively identified in *C. reinhardtii* by mass spectrometry, confirming that the opal codon UGA can indeed code for selenocysteine in some algae (*116*). Attempts to increase organic selenium concentrations in *C. reinhardtii* by introducing human selenoprotein transgenes have shown promising but not conclusive initial results (*117*).

1.3.2.3 Carotenoids from wild type and engineered algae

Dunaliella salina and *Haematococcus pluvialis* are the main sources for commercial production of non-synthetic beta-carotene (*118*) and astaxanthin, respectively. Other microalgae also show great potential for producing other relevant carotenoids such as lutein. These carotenoids are mainly used as animal feed additives and in supplements for human consumption (*119, 120*), and could become valuable co-products in the biofuel industry. Natural beta-carotene can reach a price between $300-3,000 per kilogram (*121, 122*), while astaxanthin can be sold for $2,500-10,000 per kilogram (*121, 123*).

Despite their high value, there are few reports on metabolic engineering for improved carotenoid production in algae. Several attempts have been made to engineer *C. reinhardtii* to produce astaxanthin by introducing carotenoid biosynthesis genes (*124-126*), but have been met with limited success. *H. pluvialis* has been stably transformed with an herbicide-resistant version

of its phytoene desaturase gene (*pds*), the second step of carotenoid synthesis, for enhanced astaxanthin accumulation. This approach led to 26% higher astaxanthin accumulation 48 hours after induction with high-light when compared to the wild-type strain (*127*).

More recently, the phytoene synthase gene (*psy*), the enzyme for the committed step in carotenoid synthesis, has been transformed into *C. reinhardtii* resulting in increased carotenoid accumulation. Transgenic strains overexpressing *psy* from *Dunaliella salina* and *Chlorella zoofingensis* accumulated 2.6 and 2.2-fold more lutein than the controls, respectively (*128, 129*). RNA interference technology has also been used to modify the carotenoid profile of *C. reinhardtii* and *D. salina* by silencing their phytoene desaturase genes (*pds*). Reduction of up to 93% and 72% of *pds* mRNA were achieved in *Chlamydomonas* and *Dunaliella*, respectively. However, the carotenoid content of *C. reinhardtii* did not change significantly and the effect on *D. salina* carotenoid content was not measured (*130, 131*).

1.3.3 Recombinant vaccines

Significant headway has been made in the past decade on expressing pathogen subunits in microalgae for low-cost (and perhaps eventually orally-deliverable) vaccines. Algae exhibit several advantages over other expression platforms for recombinant vaccine production. Unlike mammalian cells, algae are not susceptible to viral infection that could be passed to a patient; viral filtering is one of the most costly steps of vaccine purification from animal cell platforms. In fact, the species *C. reinhardtii* in which most vaccine antigen expression has been performed to date does not have a single identified viral pathogen – making it quite unique as a host.

1.3.3.1 Vaccine antigens produced in algae

The first recombinant vaccine antigens were produced in algae in 2003: a foot-and-mouth disease viral protein in the *C. reinhardtii* chloroplast (*132*) and the hepatitis B surface antigen in

D. salina (*133*). The hepatitis B surface protein, which is the antigen in the first ever FDA-approved recombinant vaccine, has also been produced via nuclear transformation in *C. reinhardtii* (*134*) and in *P. tricornutum* (*135*). With these exceptions, almost all other vaccine antigens have been produced in *C. reinhardtii* chloroplasts, including classical swine fever virus protein E2 (*136*), human glutamic acid decarboxylase 65 implicated in diabetes prevention (*137*), white spot syndrome virus VP28 protein (*138*), *Staphylococcus aureus* D2 protein (*139*), and a mutant version of the human papillomavirus E7 protein (*140*).

In addition, *C. reinhardtii* has garnered much attention for its unique suitability for producing malarial proteins in the chloroplast. Many of the surface proteins of malaria's causative pathogen, *Plasmodium*, are heavily disulfide bonded and also lack glycosylation (*141*). Disulfide bond formation is facilitated within the chloroplast compartment, and chloroplast-synthesized proteins lack post-translational modifications, making algal chloroplasts the only platform yet demonstrated to produce several properly-folded native *Plasmodium* proteins (*142*). The genes encoding *MSP1* from two different *Plasmodium* species and the protein *AMA1* from *Plasmodium berghei* were expressed in the *C. reinhardtii* nuclear genome and targeted to chloroplast starch granules via fusion to the major granule-associated protein (*143*). The *Plasmodium falciparum* surface proteins *Pfs*25, *Pfs*28, and *Pfs*48/45 – which are present on the sexual reproductive stages of the species most lethal to humans and must be accessible for reproduction and maturation within the mosquito gut – have also been expressed in the chloroplast (*142, 144, 145*). These antigens are suitable for a transmission-blocking vaccine because antibodies raised against them provide minimal protection to the individual, but they block transmission at the mosquito's next feeding.

1.3.3.2 Animal efficacy studies

While most of these studies provided immunological data demonstrating proper folding or biological activity *in vitro*, at least six of them have been tested in animals. The classical swine fever virus vaccine was purified and injected subcutaneously, where it triggered a systemic immune response (*136*). An *S. aureus* subunit vaccine was able to protect 80% of mice from a lethal challenge with the pathogen, even when administered orally instead of parenterally (see Sec. 3.3.3 below) (*139*). The malarial proteins *MSP1* and *AMA1* were able to reduce parasite load and prolong survival in mice exposed to the species of *Plasmodium* that infects rodents, *P. berghei* (*143*). The first study of algal-produced *Pfs*25 found that the sera of mice immunized with the purified protein was able to drastically reduce the number of oocysts in mosquito midguts when fed in a bloodmeal, indicating that the sexual reproduction was halted by the presence of the antibodies raised against the algal-produced antigen (*142*). Mice also produced systemic IgG antibodies to these antigens when injected with purified protein, and mucosal antibodies in the gut when fed whole transgenic algae (*144*). Recently, both purified protein and crude algal extracts from *C. reinhardtii* expressing a human papillomavirus antigen were able to confer 70% survival after tumor challenge in mice (*140*). All of these are promising preliminary results, and add authenticity to the *in vitro* data indicating that algal-produced antigens exhibit superb biological activity.

1.3.3.3 Recombinant oral vaccines

Recombinant oral vaccines present a variety of advantages over injectable vaccines. Oral immunization can induce mucosal immunity, which provides protection at the site of entry for almost all pathogens. With particular significance in areas with limited financial resources, medical personnel, and infrastructure, oral vaccines can be much cheaper because the antigens may not require extensive purification, and they do not require skilled medical staff for safe injection. For some diseases, such as malaria, the reduced costs of production, processing,

shipping, and administration using a freeze-dried transgenic algal oral vaccine are nearly a necessity to become a viable option for the at-risk population for the disease. Two studies tested the stability of antigens within lyophilized algae from six to twenty months at room temperature or higher and suggest that this may indeed possible (*139, 144*).

Oral immunization with the classical swine fever vaccine produced no response (*136*), but oral administration of the starch granule-bound malarial antigens protected mice nearly as well from *P. berghei* challenge as parenteral administration (*143*). In order to increase both uptake by the intestinal epithelial cells and immunogenicity, antigens can be fused to subunits from enteric pathogens that allow binding to and uptake by cells in the gut. One such subunit is the beta subunit of cholera toxin, CtxB, which binds to the GM1 ganglioside on intestinal cells. The *S. aureus* fusion to CtxB was fed to mice in the form of lyophilized whole algal cells; this oral route of administration elicited both a mucosal response – measured by IgA titers in fecal samples – and a systemic IgG response to both the *S. aureus* and the CtxB domains (*139*).

However, the same route of administration with a malarial antigen fused to the CtxB subunit was unable to elicit systemic immunity to the malarial antigen, though systemic CtxB immunity was triggered. Mucosal response was observed for both antigen subunits (*144*). The design of this chimeric molecule differed from the *S. aureus* construct in the presence of a furin protease cleavage site between the two domains. It is possible that after internalization and cleavage, the two halves experienced different intracellular fates, leading to different antigen presentation. It may also be that the malarial antigen is inherently less immunogenic than that from *S. aureus*, in which case stronger adjuvants may be needed.

1.4 Bioenergy applications

Much of the recent interest in algae as a host organism for industrial-scale production revolves around its potential for biofuel production. Ancient algae are the source of the biomass

that converted into today's petroleum, so many algae naturally produce the types of hydrocarbons that are suitable for fungible, or "drop-in", fuel alternatives. Algae are considered a promising alternative to other plant-based biofuels for several reasons. Unlike alternatives like corn ethanol or sugarcane ethanol, algal biofuel production will not compete with the food supply, alleviating food-versus-fuel pricing and availability concerns. Algal biofuel crops also will not compete with food crops for land because algae can be grown in outdoor ponds on non-arable land. Finally, the high biomass accumulation per acre compared to other potential fuel crops means that far less land is needed to produce a given volume of fuel.

1.4.1 Altering lipid metabolism

Algal lipid metabolism is believed to be largely similar to that of higher plants in that the same enzymes and subcellular compartments are used in the generation of the primary fatty acid types. The primary fatty acids palmitic, stearic, and stearadonic acid are formed in the plastid by the action of the FAS enzyme and Stearoyl ACP desaturase, and then transported to the endoplasmic reticulum for further modification and incorporation into triacylglycerol (TAG) molecules (*146*). The primary fatty acids of algae are photosynthetic and cellular membrane lipids, but there are also a number of other products such as wax esters, sterols, prenyls, and other hydrocarbons (*147*). The traditional means of obtaining high levels of storage lipids or polysaccharides has been through the induction of stress response via changes in nutrients or growth conditions (*148, 149*). Metabolic engineering has more recently been used with some success to induce accumulation of both of these compounds.

1.4.1.1 Changes in nutrient availability

The most effective and well-known nutrient modification for altering lipid metabolism and storage is the limitation of nitrogen. When nitrogen is limited in the media, algae are unable

to continue protein synthesis and therefore growth and replication slows and eventually halts. Photosynthesis will continue, but the algae respond by dramatically increasing their lipid content using the energy provided by photosynthesis or a carbon source such as sugars or acetate (*150-152*). Silicon, a nutrient vital to the growth of diatoms, also mimics the effects of nitrogen starvation when it becomes limited (*153*), as does phosphorus (*154*). Iron has also been shown to affect lipid levels, but inversely: excess iron in the media positively influences both biomass and lipid content (*152, 155-157*). Carbon dioxide supplied in high levels also supports more robust cell growth and the accumulation of lipids, as carbon dioxide is required for photosynthesis and also for the formation of acetyl-CoA, a substrate of fatty acid biosynthesis (*158*).

1.4.1.2 Changes in temperature

Lipids are known to play a role in regulating the fluidity of membranes in response to temperature, with more unsaturated lipids being required to maintain fluidity at low temperatures (*159*). Increased polyunsaturated fatty acid levels have been associated with growth in low temperatures for algae (*160*). It appears that these modifications are particularly important in the photosynthetic membranes of algae as an adaptation to the cold, and if more saturated fatty acids are desirable then the algae must be shifted to higher growing temperatures (*161*).

1.4.1.3 Changes in light intensity

Growth of cultures in various light conditions can also affect the proportion of total lipids to membrane lipids, although these trends are more difficult to interpret due to large variations between species. Some studies suggest that low light levels tend to favor the production of photosynthetic membrane lipids, while high light levels favor more production of non-polar storage lipids such as TAG (*162, 163*). On the other hand, high light levels can reduce some polyunsaturated fatty acid levels, but high light irradiance can favor the production of

polyunsaturated fats in the polar membrane lipids, possibly due to a role in the mitigation of free oxygen radicals from photosynthesis under high light conditions (*164, 165*).

1.4.1.4 Genetic manipulation of lipid content

Since the majority of lipid regulatory proteins appear to be nuclear-encoded, a primary limitation of successful genetic manipulation of lipid metabolism involves obtaining high-level recombinant protein expression from the nucleus (see Sec. 2.1.2 and 2.1.3). Early attempts to increase lipid metabolism aimed to increase the levels of substrates available for lipid production. Algae, like higher plants, utilize malonyl-CoA to synthesize lipids, and malonyl-CoA is used to add to lipid chains two carbons at a time. The first attempts to increase lipid accumulation were therefore focused on increasing malonyl-CoA levels by expressing the gene encoding acetyl-CoA carboxylase (ACCase) (*166*). Increasing ACCase levels by 2 to 3 fold in the diatom *Cyclotella cryptica* did not result in any increase in lipid levels, indicating that the rate-limiting step of lipid production in algae is not the production of lipid biosynthesis substrates. The lessons learned from early attempts like these seemed to point to regulatory proteins as being the primary bottlenecks in the production of lipids from algae.

Among the simplest modifications capable of producing increased lipid accumulation was a mutation of the sta6 gene that eliminated accumulation of this starch biosynthesis enzyme in *C. reinhardtii.* This mutant exhibited much higher levels of lipid production under nitrogen starvation, presumably due to the inability of this mutant to store energy as starch (*167*). One of the most impressive genetic manipulations applied to microalgae for the purpose of biofuel production involves the expression of two heterologous thioesterases in *Phaeodactylum tricornutum,* which resulted in the production of increased proportions of shorter chain fatty acids (*168*). Producing shorter chain fatty acids is a crucial advance in the production of biodiesel with excellent cold flow properties. Relatively few of the many lipid biosynthesis metabolic

modifications pioneered in higher plants have been transferred to algae so far, but a dramatic expansion in the field of algal metabolic engineering is anticipated in the coming years due to improved genetic tool availability.

1.4.2 Increasing photosynthetic efficiency

Photosynthesis typically only captures the energy from about 1% of incident light, leaving ample room for improvement of biomass yields, culture density, and growth rates. Several strategies for increasing photosynthetic efficiency have been employed in algae, as strains with modified photosynthetic apparatuses are quick to make and can be rapidly screened and characterized, at least compared to terrestrial plants.

1.4.2.1 Improving light utilization

Microalgae have evolved large light-harvesting complexes (LHCs) for maximizing light absorption in low-light environments. In laboratory culture where light is saturating, excess energy is dissipated through heat and fluorescence quenching in the LHCs. Energy that cannot be dissipated results in direct photodamage and the production of reactive oxygen species, leading to photoinhibition. The large size of the LHCs also limits light penetration into the culture, reducing the achievable cell density (*169*).

To overcome this, a single RNAi construct has been introduced to silence all twenty LHC protein isoforms of *C. reinhardtii*. These cells have lower mRNA and protein accumulation for all LHC genes, and 68% less chlorophyll than the parental strain, resulting in 290% higher light transmittance in the culture. Furthermore they present less fluorescence quenching, which leads to an increase in photosynthetic quantum yield. Under high-light conditions, transformed cells were less susceptible to photo-inhibition and grew at a faster rate; however they did not reach a higher cell density (*170*), which is considered a highly desirable trait for large-scale algae production.

Similar results were achieved by down-regulating LHC expression using a translational repressor of the LHC protein family, NAB1 (*171, 172*). A constitutive version of NAB1 was over-expressed in *C. reinhardtii*, producing a similar but less striking phenotype than the RNAi strain: chlorophyll was only reduced by 20% (*173*). Nonetheless, over-expressing a single repressor is notably easier than silencing twenty LHC isoforms.

1.4.2.2 Improving photosynthetic efficiency

Photoinhibition caused by excess sunlight is regarded as a major limiting factor for primary productivity in shallow production ponds under high irradiances (*174*). One of the main targets of photo-damage is Photosystem II (PSII), the multi-protein complex that performs the light-driven oxidation of water. Degradation of the D1 subunit of PSII is significantly increased when light is in excess (*175*). However, it is possible to select for mutant versions of the algal D1 protein that can evolve up to ~4.5 more oxygen *in vivo* under high-light conditions (50% midday sunlight) compared to the control (*176*). This was achieved by transforming error-prone PCR-amplified D1 coding sequences followed by selection under ionizing radiation. Unfortunately the mutant strains perform slightly worse under laboratory light conditions (10% midday sunlight) (*176*), suggesting that this strategy may not translate to increased biomass yield for commercial algae production.

The feasibility of using heterologous genes as interchangeable parts (biobricks) for synthetic biology in *C. reinhardtii* chloroplasts has also been tested using the gene coding for the D1 protein (*psbA*) (*177*). This work showed that introducing fully heterologous genes is not the best approach for engineering the chloroplast, but that heterologous coding sequences driven by endogenous regulatory regions are a viable method for modifying the algal photosynthetic machinery (*177*). Furthermore, two isoforms of the D1 proteins from *Synechoccocus sp.* PCC 7942 (cyanobacteria) expressed in *C. reinhardtii* reconstituted the low-light and high-light

phenotypes associated with each D1 isoform. Interestingly, *C. reinhardtii* expressing the cyanobacterial low-light isoform yielded 11% more dry weight biomass than the strains expressing the high-light isoform or the endogenous D1 protein, which is a highly desirable trait for biofuel production (*178*).

1.4.2.3 Expanding the photosynthetically active spectrum

Microalgae can only utilize the light spectra between 400-700 nm for photosynthesis, also known as photosynthetically active radiation (PAR). Expanding this spectrum could enhance primary productivity without additional inputs. Some newly discovered cyanobacterial species can synthesize two novel types of light-harvesting pigments, chlorophyll-d and chlorophyll-f, which can absorb light frequencies of up to 750 nm. It has been proposed that inclusion of these pigments by metabolic engineering and synthetic biology into microalgae and plants would increase the amount of photon flux available for photosynthesis by 19% (*179*), but this has not yet been demonstrated experimentally.

1.4.3 Modifying carbon assimilation

In algae grown phototrophically, all new biomass – including lipids – derives from the fixation of CO_2 into ribulose-1,5-biphosphate (RuBP) to form 3-phosphoglycerate, catalyzed by the enzyme ribulose-1,5-biphosphate carboxylase/ oxygenase (Rubisco). Several other enzymes are also required to regenerate RuBP in the Calvin cycle. Considerable amounts of ATP and NADPH are required for this process, and are supplied by the light-driven activity of photosystems I and II (*180*). Several studies have shown that the activity of Rubisco is the major bottleneck for carbon flux through the Calvin cycle when CO_2 is not enriched in the media and under high-light or high-temperature conditions, and all of these conditions are present in industrial-scale ponds in desert areas (*180, 181*).

1.4.3.1 Engineering Rubisco

Rubisco is the first enzyme in the CO_2 fixation pathway and is regarded as a "slow" catalyst; large amounts of Rubisco are required for a sustainable carboxylation rate, and it has an affinity for oxygen, which is used in a counter-productive competing reaction (*182*). *C. reinhardtii* is an ideal host for engineering Rubisco, since there are Rubisco-deficient strains that can complete their life cycle heterotrophically (*182*). The small and large sub-units of *C. reinhardtii* Rubisco (*rbcS* and *rbcL* genes) were mutated to resemble spinach Rubisco, and then transformed into an *rbcS*-*rbcL* deficient algae strain, resulting in a holoenzyme that resembled its plant counterpart. Compared to the *C. reinhardtii* Rubisco, the mutant version had a 17% increase of the *in vitro* CO_2/O_2 specificity factor (Ω) and a 53% decrease of the *V*max of carboxylation (*V*c) (*183*). Similarly, the *rbcS* genes of *Arabidopsis* and sunflower were introduced into *C. reinhardtii* while preserving the endogenous large subunit gene (*rbcL*). The Ω was improved by up to 11% while maintaining the *V*max; however, the cells displayed lower growth rates and lacked pyrenoids, presumably due to mis-targeting of Rubisco (*184*).

C. reinhardtii rbcL has also been subjected to PCR-based gene shuffling with oligonucleotides representing the natural diversity of this gene. Three rounds of gene shuffling and three rounds of strain selection resulted in a Rubisco with up to 20% and 56% increase in Ω and *V*c, respectively. Some of the enriched mutations were then incorporated into the tobacco *rbcL* and resulted in 14% Ω and 15% *V*c increases (*185*). Another interesting strategy is to tune Rubisco abundance according to the environmental culture conditions to optimize carbon, energy, and nitrogen utilization. *C. reinhardtii* strains with varying levels of Rubisco have been engineered by manipulating the *rbcL* mRNA maturation factor MRL1. Rubisco could be reduced to 15% of wild-type levels while maintaining phototrophic growth (*186*). An inducible promoter for MRL1 could potentially be used to tune Rubisco accumulation according to culture conditions like light intensity or CO_2 concentration.

1.5 Other industrial applications

Algae are being explored for many additional industrial-scale applications beyond therapeutics and biofuels. Some of these involve alternative uses for algal biomass left over after lipid extraction, while others require genetically engineered strains.

1.5.1 Animal feedstocks

As much as two thirds of the grain produced in the U.S. is used to feed livestock. To reduce costs and increase sustainability of livestock production, there is recent interest in using fast-growing, high-protein algal species as animal feed. Some species are already used extensively as aquaculture feed (*187*), and algal facilities for biofuels or other industrial uses will soon produce a large excess of residual biomass. Some of this biomass can be added back into the media for subsequent cultures to recycle the nutrients, but it is also a rich source of protein that is valuable as a feed additive for both livestock and farmed fish.

1.5.1.1 Algal protein extracts as animal feedstocks

After biofuel-relevant lipids have been extracted from algal biomass, the resulting protein-rich algal paste is a valuable co-product for animal feeds. It has been suggested that this co-product could subsidize the current cost of algal-based biofuels (*188*). Nutritional analysis indicates that algal protein provides a balanced source of amino acids, suitable for many animals from mammals to fish (*189*). However, this co-product makes some chemical extraction methods for lipids unfeasible, as they may leave residual harmful solvents in the algal paste. Mechanical extraction, which is a lower-yielding method of extracting fuel lipids, could be used instead, but a cost-benefit analysis will need to be incorporated into these considerations.

1.5.1.2 Engineered algae to enhance immune function

Recently, algae have been genetically engineered to express proteins from mammalian colostrum, the mother's first milk, which contains many proteins that stimulate immune function and protection in the offspring (*76*). In large-scale animal rearing, offspring are usually weaned from the mother's milk or given reconstituted formula at young age, leaving their immune systems inadequately developed. As a result, low-dose antibiotic administration has become standard practice in commercial farming operations, accounting for the vast majority of all antibiotics sold. Unsurprisingly, widespread antibiotic resistance has emerged and is often directly traceable to livestock farms (*190-192*).

As an alternative approach, critical colostrum proteins can be introduced into feed or formula using engineered algae, perhaps reducing the need for low-dose antibiotic administration. The colostrum protein mammary-associated serum amyloid, M-SAA, has been expressed from the *C. reinhardtii* chloroplast genome and accumulates to nearly 10% of total soluble protein (*76*). This protein stimulates mucin production in the gut of offspring, protecting them from enteric disease. Bioactive peptides derived from chimeric milk proteins have also recently been produced in *C. reinhardtii* chloroplasts (*193*). The ability to introduce multiple transgenes into algae allows multiple critical colostrum proteins to be introduced using a single algal strain, and this work is already underway (M. Tran, personal communication).

1.5.2 Industrial enzymes

Large quantities of relatively pure enzymes are required for many industrial processes, from food processing to paper milling to biofuels. With the recent push towards green manufacturing, many processes that have previously utilized energy-intensive steps like heat and pressure exposure, or steps with harsh chemicals like strong acids or bases, are turning to biological alternatives involve enzymes found in nature. Because of their potential for cheap,

large-scale growth, algae are being investigated as host organisms for producing these types of enzymes.

1.5.2.1 Enzymes for processing cellulosic biofuels

While algae have attracted attention from biofuels developers and researchers for their ability to accumulate high lipid levels, there may also be a role for them in downstream processing of other biofuel feedstocks. Cellulosic biofuels use the large quantities of carbohydrate polymers in terrestrial plant tissue (namely cellulose, but also other components of the associated matrix) as a source of sugars for fermentation into ethanol. Some cellulosic crops are grown for the explicit purpose of serving as biofuel feedstocks, such as high-yielding grasses like switchgrass, but another source is residual plant matter left over after harvesting food crops.

For any of these feedstocks, a mix of enzymes must be added to effectively break down the cellulose, hemicellulose, and lignin into their constituent fermentable sugars because no single enzyme can completely degrade these polymers (*194*). While some work has focused on engineering cellulosic crops to contain more easily degraded polymers (*195*), this strategy is limited because of the essential structural role they serve within plants; therefore, even engineered feedstocks require additional degradation enzymes. These enzymes include cellobiohydrolase, endoglucanase, beta-glucosidase, and xylanase. Xylanase – which is currently used in other industries like paper milling – has been produced in multiple algal species, including *D. salina* (*96*) and *C. reinhardtii* (*196*), and can even be expressed from the nuclear genome and targeted for secretion into the culture media, further reducing production costs of these enzymes.

1.5.2.2 Enzymes for food processing

Xylanase is also used in animal feeds, along with several other enzymes to enhance digestibility and maximize nutritional and caloric yield, including mannanase, alpha-

galactosidase, and phytase. Phytase releases absorbable inorganic phosphate from phytic acid, one of the main storage forms of phosphorus in grains. As phytic acid, it is not bioavailable to non-ruminant livestock. This not only makes it difficult for the animals to obtain sufficient phosphorus from a grain-based diet; it also contributes to environmental pollution because the phosphorus is excreted as waste, and fuels toxic microbial blooms downstream in the run-off. In fact, an "environmentally friendly" engineered pig that produces phytase in its saliva was created and tested (*197*), but funding was withdrawn before it completed the regulatory approval process.

Alternatively, these enzymes can be added to animal feed by mixing with edible transgenic algae. Alpha-galactosidase from two genetic sources, as well as two phytase enzymes, have been expressed and shown to be bioactive in both *D. salina* and *C. reinhardtii* (*96, 198*). In fact, chickens fed algae producing phytase have been shown to have lower phytic acid content in their manure (*198*). Mannanase was also produced in these algal species but did not exhibit bioactivity, which may be due to the lack of glycosylation because it was produced within the chloroplast (*96*).

1.6 Industrial-scale algal production

Growing algae at industrial scale requires addressing a host of concerns that are not encountered at the bench. The media inputs must be minimized to remain cost-effective; evaporative water losses must be taken into account; energy-efficient harvesting methods become critical; and crop protection strategies to evade grazers and pathogens must be developed. Lab-scale growth ranges from less than a milliliter, for cultures grown in 96-well plate high-throughput format, to around 20 L. In enclosed environments, algae can be grown in 100 L bags or equivalent contained systems. Pilot-scale ponds range from hundreds to tens of thousands of liters, providing some information about concerns that appear at scale, especially with outdoor growth. But industrial production will likely occur in ponds harboring hundreds of thousands to

millions of liters, with some applications – such as high-value therapeutics or vaccines – suitable for thousand-liter fermenters. Christenson and Sims (*199*) have compiled a list of all commercial facilities to date, including the location, production type (pond, raceway, tubular photobioreactor, etc.), and scale.

1.6.1 Enclosed photobioreactors and fermenters

At present, much of the preliminary culturing, harvesting, and analysis of algae is being done in enclosed bioreactors. This is due in part to tight regulations governing outdoor cultivation of genetically engineered strains, but this contained growth also exhibits distinct advantages, particularly with regard to parameter control and optimizing yields.

1.6.1.1 Optimizing photobioreactor design

Enclosed systems alleviate the challenges associated with pathogens and predators; allow for tighter controls on parameters such as media component concentrations, light intensity, and temperature; and are ideal for transgenic strains that may pose an environmental threat if they escape into the environment. However, they are much more expensive to build and maintain than open ponds, so these systems are more suitable for high-value products like therapeutics or nutritional supplements.

As mentioned in Sec. 4.2, photosynthetic organisms typically have a relatively narrow range of light intensity in which they perform best: too much light causes photoinhibition, and too little limits growth. This makes the design of photobioreactors especially critical. The most common design is tubular, though flat panels, manifolds, and serpentine designs have also been tested (*200, 201*). Besides expense, there are several drawbacks of enclosed bioreactors including biofouling or growth on the walls limiting light penetrance, and buildup of dissolved oxygen, both of which negatively impact productivity (*201*).

1.6.1.2 Modifying trophic conversion pathways for heterotrophic growth

Growing microalgae under heterotrophic or mixotrophic conditions (in the presence of an organic carbon source) has several advantages. Fermentation systems have already been widely studied and successfully applied in industry, and the culture conditions are highly controlled and reproducible. Moreover, heterotrophic cultures of microalgae achieve higher cell densities, thus resulting in lower harvesting costs (*200, 202*). For example, heterotrophic cultivation of *Chlorella protothecoides* for biofuel production resulted in 4-fold higher lipid accumulation than autotrophic cultures (*203*).

Some algal species are naturally heterotrophic, meaning that they can survive without light if provided an appropriate carbon source, so these are candidates for growth in existing industrial fermentation facilities. Indeed, many of the high-expressing transgenic *C. reinhardtii* strains are necessarily non-photosynthetic (see Sec. 2.2.2). However, many microalgae species are strict autotrophs – a problem that could be circumvented by genetic engineering. *Volvox carteri, C. reinhardtii,* and the diatoms *Phaeodactylum tricornutum* and *Cylindrotheca fusiformis* have been transformed with a hexose transporter (*HUP1*), resulting in strains that can survive on glucose in the dark (*204-207*). Despite glucose incorporation, the extent of conversion to full heterotrophy is variable between these four algae. Interestingly, transformed *C. reinhardtii* cells produce 150% more hydrogen than the parental strain upon induction in the presence of glucose (*206*). Trophic conversion is a good proof of concept for microalgae metabolic engineering, but despite the advantages of heterotrophic culture, it might not to be optimal for large-scale algae bio-product production. Additional costs of adding a carbon source and the requirement for enclosed bioreactors given the higher risk of contamination are major drawbacks compared to phototrophic systems.

1.6.2 Open pond growth

Realistically, open ponds will be necessary for growing algae at the cost and scale necessary to compete with cheap commodities like fuel and animal feed. However, they come with a profound set of unique challenges – both biological and abiotic. Fortunately, progress is being made on these fronts due to the availability of functional commercial and academic pilot-scale open pond facilities.

1.6.2.1 Media and water optimization

The ideal large-scale culture media contains minimal quantities of the required salts and trace elements, and no carbon source – which reduces cost as well as contamination risk. Used media after harvest is often recycled, but may eventually reduce productivity and invite contamination (*208*). One cost-cutting and environmentally beneficial proposal is to use diluted wastewater as a feedstock, since it contains high levels of nitrogen and phosphorus (*209, 210*). This would serve the dual purpose of supporting algal growth and removing these excess nutrients from the wastewater, which otherwise serves as fertilizer for unwanted algal growth in the open ocean, ultimately resulting in oxygen-depleted dead zones. Water consumption is a significant concern due to evaporative loss from pond surfaces, which can also affect the nutrient concentrations and pH, as can dilution of the culture from rainfall.

Another concern in large-volume open ponds is achieving adequate mixing. Mixing serves two purposes: it increases gas exchange – allowing atmospheric CO_2 to enter the media and oxygen to escape – and it keeps algae in circulation from the bottom of the pond to the surface, modulating light exposure to prevent photoinhibition while maintaining productivity. In smaller ponds, lifters blowing forced air can be used to mix the culture; however, this is energy intensive, so paddlewheels are typically used in large raceway ponds (*211*).

1.6.2.2 Efficient harvesting methods

At small scale, centrifugation is typically used to separate algal biomass from the culture media. However, due to the huge volume of water relative to algal biomass (typically only a few grams of dried biomass per liter of culture), large-scale harvesting is a difficult, expensive, and energy-intensive process, and centrifugation is not a viable option. One strategy is to use a filamentous algal species; indeed, this attribute has been used in large-scale *Spirulina* growth for years (*212*). Flocculants can also be used to induce aggregation of individual cells. Chemical flocculants – which utilize ions, metal precipitation, or charged polysaccharide or polyacrylamide polymers – are effective but expensive and can be difficult to remove from the algal biomass (*213*).

Bioflocculants are also being explored, including the introduction of bacteria (*214*) and filamentous fungi (*215, 216*) to induce algal aggregation. Some algal species, including the green algae *Dunaliella salina* (*217*) and *Chlorella vulgaris* (*218*) and the diatom *Phaeodactylum tricornutum* (*219*), autoflocculate in response to high pH and/or depletion of CO_2, which may serve as a useful feature for self-harvesting batch cultures that naturally aggregate upon reaching saturation.

1.7 Conclusions and potential of algal platforms

Significant progress is being made each year in the areas of algal genetic engineering and tool development, allowing for rapid advancement in achieving the complex engineering of production strains with multiple desirable traits. Bioprospecting for novel species in environments of interest for commercial growth – coupled with the power of inexpensive full-genome sequencing and assembly – also holds promise for discovering new strains and traits that are adaptable to meet the needs of industrial-scale growth, efficient harvesting, and resistance to biological and abiotic stresses. It is likely that algal platforms will not be utilized for a single

product, but that a single strain and a single harvest will contribute to multiple processing streams, akin to large-scale agriculture and animal husbandry. Algae exhibit several unique advantages over currently available industrial hosts, making them valuable additions to the commercial organism repertoire as fast-growing, inexpensive, highly tractable bio-production platforms.

Acknowledgements

Chapter 1, in full, is a reprint of the material as it has been submitted for publication to Wiley Biotechnology Series *Industrial Biotechnology*. The dissertation author is the primary investigator and author of this chapter.

References

1. T. A. Norton, M. Melkonian, R. A. Andersen, Algal biodiversity. *Phycologia* **35**, 308 (1996).

2. F. Metting Jr, Biodiversity and application of microalgae. *Journal of industrial microbiology* **17**, 477 (1996).

3. R. Barten, H. Lill, DNA uptake in the naturally competent cyanobacterium, synechocystis sp. Pcc 6803. *FEMS Microbiology letters* **129**, 83 (1995).

4. T. Thiel, H. Poo, Transformation of a filamentous cyanobacterium by electroporation. *Journal of bacteriology* **171**, 5743 (Oct, 1989).

5. G. Grigorieva, S. Shestakov, Transformation in the cyanobacterium synechocystis sp. 6803. *FEMS Microbiology Letters* **13**, 367 (1982).

6. Y. Xu, R. M. Alvey, P. O. Byrne, J. E. Graham, G. Shen, D. A. Bryant, in *Photosynthesis research protocols*. (Springer, 2011), pp. 273-293.

7. S. S. Merchant, S. E. Prochnik, O. Vallon, E. H. Harris, S. J. Karpowicz, G. B. Witman, A. Terry, A. Salamov, L. K. Fritz-Laylin, L. Marechal-Drouard, W. F. Marshall, L. H. Qu, D. R. Nelson, A. A. Sanderfoot, M. H. Spalding, V. V. Kapitonov, Q. H. Ren, P. Ferris, E. Lindquist, H. Shapiro, S. M. Lucas, J. Grimwood, J. Schmutz, P. Cardol, H. Cerutti, G. Chanfreau, C. L. Chen, V. Cognat, M. T. Croft, R. Dent, S. Dutcher, E. Fernandez, H. Fukuzawa, D. Gonzalez-Ballester, D. Gonzalez-Halphen, A. Hallmann, M. Hanikenne, M. Hippler, W. Inwood, K. Jabbari, M. Kalanon, R. Kuras, P. A. Lefebvre, S. D. Lemaire, A. V. Lobanov, M. Lohr, A. Manuell, I. Meir, L. Mets, M. Mittag, T.

Mittelmeier, J. V. Moroney, J. Moseley, C. Napoli, A. M. Nedelcu, K. Niyogi, S. V. Novoselov, I. T. Paulsen, G. Pazour, S. Purton, J. P. Ral, D. M. Riano-Pachon, W. Riekhof, L. Rymarquis, M. Schroda, D. Stern, J. Umen, R. Willows, N. Wilson, S. L. Zimmer, J. Allmer, J. Balk, K. Bisova, C. J. Chen, M. Elias, K. Gendler, C. Hauser, M. R. Lamb, H. Ledford, J. C. Long, J. Minagawa, M. D. Page, J. M. Pan, W. Pootakham, S. Roje, A. Rose, E. Stahlberg, A. M. Terauchi, P. F. Yang, S. Ball, C. Bowler, C. L. Dieckmann, V. N. Gladyshev, P. Green, R. Jorgensen, S. Mayfield, B. Mueller-Roeber, S. Rajamani, R. T. Sayre, P. Brokstein, I. Dubchak, D. Goodstein, L. Hornick, Y. W. Huang, J. Jhaveri, Y. G. Luo, D. Martinez, W. C. A. Ngau, B. Otillar, A. Poliakov, A. Porter, L. Szajkowski, G. Werner, K. M. Zhou, I. V. Grigoriev, D. S. Rokhsar, A. R. Grossman, C. Annotation, J. A. Team, The chlamydomonas genome reveals the evolution of key animal and plant functions. *Science* **318**, 245 (Oct 12, 2007).

8. S. E. Franklin, S. P. Mayfield, Prospects for molecular farming in the green alga chlamydomonas reinhardtii. *Curr Opin Plant Biol* **7**, 159 (Apr, 2004).

9. M. Fuhrmann, W. Oertel, P. Hegemann, A synthetic gene coding for the green fluorescent protein (gfp) is a versatile reporter in chlamydomonas reinhardtii. *Plant Journal* **19**, 353 (Aug, 1999).

10. J. Neupert, D. Karcher, R. Bock, Generation of chlamydomonas strains that efficiently express nuclear transgenes. *The Plant journal : for cell and molecular biology* **57**, 1140 (Mar, 2009).

11. K. J. Lauersen, H. Berger, J. H. Mussgnug, O. Kruse, Efficient recombinant protein production and secretion from nuclear transgenes in chlamydomonas reinhardtii. *Journal of Biotechnology* **167**, 101 (Aug 20, 2013).

12. J. D. Rochaix, J. Vandillewijn, Transformation of the green-alga chlamydomonas-reinhardii with yeast DNA. *Nature* **296**, 70 (1982).

13. R. Debuchy, S. Purton, J. D. Rochaix, The argininosuccinate lyase gene of chlamydomonas reinhardtii: An important tool for nuclear transformation and for correlating the genetic and molecular maps of the arg7 locus. *EMBO J* **8**, 2803 (Oct, 1989).

14. S. P. Mayfield, K. L. Kindle, Stable nuclear transformation of chlamydomonas reinhardtii by using a c. Reinhardtii gene as the selectable marker. *Proc Natl Acad Sci U S A* **87**, 2087 (Mar, 1990).

15. K. L. Kindle, High-frequency nuclear transformation of chlamydomonas-reinhardtii. *Proceedings of the National Academy of Sciences of the United States of America* **87**, 1228 (Feb, 1990).

16. K. Shimogawara, S. Fujiwara, A. Grossman, H. Usuda, High-efficiency transformation of chlamydomonas reinhardtii by electroporation. *Genetics* **148**, 1821 (Apr, 1998).

17. T. G. Dunahay, Transformation of chlamydomonas-reinhardtii with silicon-carbide whiskers. *Biotechniques* **15**, 452 (Sep, 1993).

18. S. V. Kumar, R. W. Misquitta, V. S. Reddy, B. J. Rao, M. V. Rajam, Genetic transformation of the green alga - chlamydomonas reinhardtii by agrobacterium tumefaciens. *Plant Sci* **166**, 731 (Mar, 2004).

19. P. T. Pratheesh, M. Vineetha, G. M. Kurup, An efficient protocol for the agrobacterium-mediated genetic transformation of microalga chlamydomonas reinhardtii. *Molecular biotechnology*, (Nov 6, 2013).

20. B. Qu, Y. J. Eu, W. J. Jeong, D. P. Kim, Droplet electroporation in microfluidics for efficient cell transformation with or without cell wall removal. *Lab Chip* **12**, 4483 (Nov 7, 2012).

21. A. Suresh, Y. C. Kim, Translocation of cell penetrating peptides on chlamydomonas reinhardtii. *Biotechnol Bioeng* **110**, 2795 (Oct, 2013).

22. L. W. Tam, P. A. Lefebvre, Cloning of flagellar genes in chlamydomonas-reinhardtii by DNA insertional mutagenesis. *Genetics* **135**, 375 (Oct, 1993).

23. H. Cerutti, A. M. Johnson, N. W. Gillham, J. E. Boynton, A eubacterial gene conferring spectinomycin resistance on chlamydomonas reinhardtii: Integration into the nuclear genome and gene expression. *Genetics* **145**, 97 (Jan, 1997).

24. N. J. Gumpel, J. D. Rochaix, S. Purton, Studies on homologous recombination in the green alga chlamydomonas reinhardtii. *Curr Genet* **26**, 438 (Nov-Dec, 1994).

25. J. A. E. Nelson, P. A. Lefebvre, Targeted disruption of the nit8 gene in chlamydomonas-reinhardtii. *Mol Cell Biol* **15**, 5762 (Oct, 1995).

26. B. Zorin, P. Hegemann, I. Sizova, Nuclear-gene targeting by using single-stranded DNA avoids illegitimate DNA integration in chlamydomonas reinhardtii. *Eukaryot Cell* **4**, 1264 (Jul, 2005).

27. H. Cerutti, A. M. Johnson, N. W. Gillham, J. E. Boynton, Epigenetic silencing of a foreign gene in nuclear transformants of chlamydomonas. *Plant Cell* **9**, 925 (Jun, 1997).

28. J. A. Casas-Mollano, B. R. Jeong, J. Xu, H. Moriyama, H. Cerutti, The mut9p kinase phosphorylates histone h3 threonine 3 and is necessary for heritable epigenetic silencing in chlamydomonas. *Proc Natl Acad Sci U S A* **105**, 6486 (Apr 29, 2008).

29. J. P. Davies, D. P. Weeks, A. R. Grossman, Expression of the arylsulfatase gene from the beta 2-tubulin promoter in chlamydomonas reinhardtii. *Nucleic Acids Res* **20**, 2959 (Jun 25, 1992).

30. M. Ohresser, R. F. Matagne, R. Loppes, Expression of the arylsulphatase reporter gene under the control of the nit1 promoter in chlamydomonas reinhardtii. *Curr Genet* **31**, 264 (Mar, 1997).

31. N. Fischer, J. D. Rochaix, The flanking regions of psad drive efficient gene expression in the nucleus of the green alga chlamydomonas reinhardtii. *Mol Genet Genomics* **265**, 888 (Jul, 2001).

32. M. Fuhrmann, A. Hausherr, L. Ferbitz, T. Schodl, M. Heitzer, P. Hegemann, Monitoring dynamic expression of nuclear genes in chlamydomonas reinhardtii by using a synthetic luciferase reporter gene. *Plant Mol Biol* **55**, 869 (Aug, 2004).

33. S. Park, Y. Lee, J. H. Lee, E. Jin, Expression of the high light-inducible dunaliella lip promoter in chlamydomonas reinhardtii. *Planta*, (Sep 17, 2013).

34. M. Schroda, D. Blocker, C. F. Beck, The hsp70a promoter as a tool for the improved expression of transgenes in chlamydomonas. *Plant J* **21**, 121 (Jan, 2000).

35. M. Lodha, M. Schulz-Raffelt, M. Schroda, A new assay for promoter analysis in chlamydomonas reveals roles for heat shock elements and the tata box in hsp70a promoter-mediated activation of transgene expression. *Eukaryot Cell* **7**, 172 (Jan, 2008).

36. J. X. Wu, Z. L. Hu, C. G. Wang, S. F. Li, A. P. Lei, Efficient expression of green fluorescent protein (gfp) mediated by a chimeric promoter in chlamydomonas reinhardtii. *Chinese Journal of Oceanology and Limnology* **26**, 242 (Aug, 2008).

37. Z. Y. Qiu, L. H. Wang, Q. Zhou, Effects of bisphenol a on growth, photosynthesis and chlorophyll fluorescence in above-ground organs of soybean seedlings. *Chemosphere* **90**, 1274 (Jan, 2013).

38. M. Heitzer, B. Zschoernig, Construction of modular tandem expression vectors for the green alga chlamydomonas reinhardtii using the cre/lox-system. *Biotechniques* **43**, 324 (Sep, 2007).

39. D. Strenkert, S. Schmollinger, M. Schroda, Heat shock factor 1 counteracts epigenetic silencing of nuclear transgenes in chlamydomonas reinhardtii. *Nucleic Acids Res* **41**, 5273 (May 1, 2013).

40. E. Diaz-Santos, M. de la Vega, M. Vila, J. Vigara, R. Leon, Efficiency of different heterologous promoters in the unicellular microalga chlamydomonas reinhardtii. *Biotechnol Progr* **29**, 319 (Mar-Apr, 2013).

41. S. D. Lumbreras V, and Purton S, Efficient foreign gene expression in chlamydomonas reinhardtii mediated by an endogenous intron. *The Plant Journal*, 441 (1998).

42. A. Eichler-Stahlberg, W. Weisheit, O. Ruecker, M. Heitzer, Strategies to facilitate transgene expression in chlamydomonas reinhardtii. *Planta* **229**, 873 (Mar, 2009).

43. M. Heitzer, B. Zschoernig, Construction of modular tandem expression vectors for the green alga chlamydomonas reinhardtii using the cre/lox-system. *BioTechniques* **43**, 324 (Sep, 2007).

44. B. A. Rasala, P. A. Lee, Z. X. Shen, S. P. Briggs, M. Mendez, S. P. Mayfield, 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**, (Aug 24, 2012).

45. A. Kumar, V. R. Falcao, R. T. Sayre, Evaluating nuclear transgene expression systems in chlamydomonas reinhardtii. *Algal Research* **2**, 321 (2013).

46. R. Debuchy, S. Purton, J. D. Rochaix, The argininosuccinate lyase gene of chlamydomonas-reinhardtii - an important tool for nuclear transformation and for correlating the genetic and molecular maps of the arg7 locus. *Embo Journal* **8**, 2803 (Oct, 1989).

47. K. L. Kindle, R. A. Schnell, E. Fernandez, P. A. Lefebvre, Stable nuclear transformation of chlamydomonas using the chlamydomonas gene for nitrate reductase. *J Cell Biol* **109**, 2589 (Dec, 1989).

48. D. R. Stevens, J. D. Rochaix, S. Purton, The bacterial phleomycin resistance gene ble as a dominant selectable marker in chlamydomonas. *Mol Gen Genet* **251**, 23 (Apr 24, 1996).

49. J. A. E. Nelson, P. B. Savereide, P. A. Lefebvre, The cry1 gene in chlamydomonas-reinhardtii - structure and use as a dominant selectable marker for nuclear transformation. *Mol Cell Biol* **14**, 4011 (Jun, 1994).

50. I. Sizova, M. Fuhrmann, P. Hegemann, A streptomyces rimosus aphviii gene coding for a new type phosphotransferase provides stable antibiotic resistance to chlamydomonas reinhardtii. *Gene* **277**, 221 (Oct 17, 2001).

51. P. Berthold, R. Schmitt, W. Mages, An engineered streptomyces hygroscopicus aph 7" gene mediates dominant resistance against hygromycin b in chlamydomonas reinhardtii. *Protist* **153**, 401 (Dec, 2002).

52. J. M. Quinn, J. Kropat, S. Merchant, Copper response element and crr1-dependent ni2+-responsive promoter for induced, reversible gene expression in chlamydomonas reinhardtii. *Eukaryot Cell* **2**, 995 (Oct, 2003).

53. N. Shao, R. Bock, A codon-optimized luciferase from gaussia princeps facilitates the in vivo monitoring of gene expression in the model alga chlamydomonas reinhardtii. *Curr Genet* **53**, 381 (Jun, 2008).

54. O. Ruecker, K. Zillner, R. Groebner-Ferreira, M. Heitzer, Gaussia-luciferase as a sensitive reporter gene for monitoring promoter activity in the nucleus of the green alga chlamydomonas reinhardtii. *Molecular Genetics and Genomics* **280**, 153 (Aug, 2008).

55. X. Sun, S. Perera, N. Haas, P. A. Lefebvre, C. D. Silflow, Using an rsp3 reporter gene system to investigate molecular regulation of hydrogenase expression in chlamydomonas reinhardtii. *Algal Research* **2**, 341 (2013).

56. B. A. Rasala, D. J. Barrera, J. Ng, T. M. Plucinak, J. N. Rosenberg, D. P. Weeks, G. A. Oyler, T. C. Peterson, F. Haerizadeh, S. P. Mayfield, Expanding the spectral palette of fluorescent proteins for the green microalga chlamydomonas reinhardtii. *The Plant journal : for cell and molecular biology* **74**, 545 (May, 2013).

57. E. L. de Hostos, R. K. Togasaki, A. Grossman, Purification and biosynthesis of a derepressible periplasmic arylsulfatase from chlamydomonas reinhardtii. *J Cell Biol* **106**, 29 (Jan, 1988).

58. B. A. Rasala, D. J. Barrera, J. Ng, T. M. Plucinak, J. N. Rosenberg, D. P. Weeks, G. A. Oyler, T. C. Peterson, F. Haerizadeh, S. P. Mayfield, Expanding the spectral palette of fluorescent proteins for the green microalga chlamydomonas reinhardtii. *Plant Journal* **74**, 545 (May, 2013).

59. D. Gonzalez-Ballester, W. Pootakham, F. Mus, W. Yang, C. Catalanotti, L. Magneschi, A. de Montaigu, J. J. Higuera, M. Prior, A. Galvan, E. Fernandez, A. R. Grossman, Reverse genetics in chlamydomonas: A platform for isolating insertional mutants. *Plant Methods* **7**, 24 (2011).

60. J. A. Casas-Mollano, J. Rohr, E. J. Kim, E. Balassa, K. van Dijk, H. Cerutti, Diversification of the core rna interference machinery in chlamydomonas reinhardtii and the role of dcl1 in transposon silencing. *Genetics* **179**, 69 (May, 2008).

61. M. Schroda, Rna silencing in chlamydomonas: Mechanisms and tools. *Curr Genet* **49**, 69 (Feb, 2006).

62. M. Fuhrmann, A. Stahlberg, E. Govorunova, S. Rank, P. Hegemann, The abundant retinal protein of the chlamydomonas eye is not the photoreceptor for phototaxis and photophobic responses. *J Cell Sci* **114**, 3857 (Nov, 2001).

63. M. Fuhrmann, A. Stahlberg, E. Govorunova, S. Rank, P. Hegemann, The abundant retinal protein of the chlamydomonas eye is not the photoreceptor for phototaxis and photophobic responses. *Journal of Cell Science* **114**, 3857 (Nov, 2001).

64. H. Cerutti, X. Ma, J. Msanne, T. Repas, Rna-mediated silencing in algae: Biological roles and tools for analysis of gene function. *Eukaryot Cell* **10**, 1164 (Sep, 2011).

65. T. Zhao, W. Wang, X. Bai, Y. Qi, Gene silencing by artificial micrornas in chlamydomonas. *Plant J*, (Nov 28, 2008).

66. I. Sizova, A. Greiner, P. Hegemann, Zinc finger nucleases generate DNA double strand brakes and modification in chlamydomonas reinhardtii. *Febs Journal* **280**, 60 (Jul, 2013).

67. S. E. Franklin, S. P. Mayfield, Prospects for molecular farming in the green alga chlamydomonas. *Current opinion in plant biology* **7**, 159 (Apr, 2004).

68. J. E. Maul, J. W. Lilly, L. Y. Cui, C. W. dePamphilis, W. Miller, E. H. Harris, D. B. Stern, The chlamydomonas reinhardtti plastid chromosome: Islands of genes in a sea of repeats. *Plant Cell* **14**, 2659 (Nov, 2002).

69. J. E. Boynton, N. W. Gillham, E. H. Harris, J. P. Hosler, A. M. Johnson, A. R. Jones, B. L. Randolph-Anderson, D. Robertson, T. M. Klein, K. B. Shark, J. C. Sanford, Chloroplast transformation in chlamydomonas with high velocity microprojectiles. *Science* **240**, 1534 (Jun 10, 1988).

70. K. L. Kindle, K. L. Richards, D. B. Stern, Engineering the chloroplast genome - techniques and capabilities for chloroplast transformation in chlamydomonas-reinhardtii. *Proceedings of the National Academy of Sciences of the United States of America* **88**, 1721 (Mar, 1991).

71. D. Dauvillee, L. Hilbig, S. Preiss, U. Johanningmeier, Minimal extent of sequence homology required for homologous recombination at the psba locus in chlamydomonas reinhardtii chloroplasts using pcr-generated DNA fragments. *Photosynthesis Research* **79**, 219 (2004).

72. W. Zerges, Translation in chloroplasts. *Biochimie* **82**, 583 (Jun-Jul, 2000).

73. J. Marin-Navarro, A. L. Manuell, J. Wu, P. M. S, Chloroplast translation regulation. *Photosynth Res* **94**, 359 (Nov-Dec, 2007).

74. D. L. Herrin, J. Nickelsen, Chloroplast rna processing and stability. *Photosynthesis Research* **82**, 301 (2004).

75. J. D. Rochaix, Post-transcriptional regulation of chloroplast gene expression in chlamydomonas reinhardtii. *Plant Mol Biol* **32**, 327 (Oct, 1996).

76. A. L. Manuell, M. V. Beligni, J. H. Elder, D. T. Siefker, M. Tran, A. Weber, T. L. McDonald, S. P. Mayfield, Robust expression of a bioactive mammalian protein in chlamydomonas chloroplast. *Plant biotechnology journal* **5**, 402 (May, 2007).

77. S. P. Fletcher, M. Muto, S. P. Mayfield, Optimization of recombinant protein expression in the chloroplasts of green algae. *Transgenic Microalgae as Green Cell Factories* **616**, 90 (2007).

78. E. A. Specht, S. P. Mayfield, Synthetic oligonucleotide libraries reveal novel regulatory elements in chlamydomonas chloroplast mrnas. *ACS synthetic biology* **2**, 34 (Jan 18, 2013).

79. H. C. Chen, A. Melis, Marker-free genetic engineering of the chloroplast in the green microalga chlamydomonas reinhardtii. *Plant Biotechnology Journal* **11**, 818 (Sep, 2013).

80. J. M. Bateman, S. Purton, Tools for chloroplast transformation in chlamydomonas: Expression vectors and a new dominant selectable marker. *Molecular & general genetics : MGG* **263**, 404 (Apr, 2000).

81. J. M. Bateman, S. Purton, Tools for chloroplast transformation in chlamydomonas: Expression vectors and a new dominant selectable marker. *Mol Gen Genet* **263**, 404 (Apr, 2000).

82. M. Goldschmidt-Clermont, Transgenic expression of aminoglycoside adenine transferase in the chloroplast: A selectable marker of site-directed transformation of chlamydomonas. *Nucleic Acids Res* **19**, 4083 (Aug 11, 1991).

83. E. H. Harris, B. D. Burkhart, N. W. Gillham, J. E. Boynton, Antibiotic resistance mutations in the chloroplast 16s and 23s rrna genes of chlamydomonas reinhardtii: Correlation of genetic and physical maps of the chloroplast genome. *Genetics* **123**, 281 (Oct, 1989).

84. E. Przibilla, S. Heiss, U. Johanningmeier, A. Trebst, Site-specific mutagenesis of the d1 subunit of photosystem ii in wild-type chlamydomonas. *The Plant cell* **3**, 169 (Feb, 1991).

85. J. M. Erickson, M. Rahire, P. Bennoun, P. Delepelaire, B. Diner, J. D. Rochaix, Herbicide resistance in chlamydomonas reinhardtii results from a mutation in the chloroplast gene for the 32-kilodalton protein of photosystem ii. *Proc Natl Acad Sci U S A* **81**, 3617 (Jun, 1984).

86. I. Minko, S. P. Holloway, S. Nikaido, M. Carter, O. W. Odom, C. H. Johnson, D. L. Herrin, Renilla luciferase as a vital reporter for chloroplast gene expression in chlamydomonas. *Molecular and General Genetics* **262**, 421 (Oct, 1999).

87. S. Franklin, B. Ngo, E. Efuet, S. P. Mayfield, Development of a gfp reporter gene for chlamydomonas reinhardtii chloroplast. *Plant J* **30**, 733 (Jun, 2002).

88. S. P. Mayfield, J. Schultz, Development of a luciferase reporter gene, luxct, for chlamydomonas reinhardtii chloroplast. *Plant J* **37**, 449 (Feb, 2004).

89. U. Klein, J. D. De Camp, L. Bogorad, Two types of chloroplast gene promoters in chlamydomonas reinhardtii. *Proc Natl Acad Sci U S A* **89**, 3453 (Apr 15, 1992).

90. Y. Gong, H. Hu, Y. Gao, X. Xu, H. Gao, Microalgae as platforms for production of recombinant proteins and valuable compounds: Progress and prospects. *Journal of industrial microbiology & biotechnology* **38**, 1879 (Dec, 2011).

91. K. E. Apt, P. G. Kroth-Pancic, A. R. Grossman, Stable nuclear transformation of the diatom phaeodactylum tricornutum. *Molecular & general genetics : MGG* **252**, 572 (Oct 16, 1996).

92. A. Falciatore, R. Casotti, C. Leblanc, C. Abrescia, C. Bowler, Transformation of nonselectable reporter genes in marine diatoms. *Marine biotechnology* **1**, 239 (May, 1999).

93. K.-C. Chow, W. Tung, Electrotransformation of chlorella vulgaris. *Plant Cell Reports* **18**, 778 (1999).

94. S. Feng, L. Xue, H. Liu, P. Lu, Improvement of efficiency of genetic transformation for dunaliella salina by glass beads method. *Molecular biology reports* **36**, 1433 (Jul, 2009).

95. C. Tan, S. Qin, Q. Zhang, P. Jiang, F. Zhao, Establishment of a micro-particle bombardment transformation system for dunaliella salina. *Journal of microbiology* **43**, 361 (Aug, 2005).

96. D. R. Georgianna, M. J. Hannon, M. Marcuschi, S. Q. Wu, K. Botsch, A. J. Lewis, J. Hyun, M. Mendez, S. P. Mayfield, Production of recombinant enzymes in the marine alga dunaliella tertiolecta. *Algal Research-Biomass Biofuels and Bioproducts* **2**, 2 (Jan, 2013).

97. M. R. Te, D. J. Miller, Genetic transformation of dinoflagellates (amphidinium and symbiodinium): Expression of gus in microalgae using heterologous promoter constructs. *The Plant Journal* **13**, 427 (1998).

98. M. Lapidot, D. Raveh, A. Sivan, S. M. Arad, M. Shapira, Stable chloroplast transformation of the unicellular red alga porphyridium species. *Plant Physiol* **129**, 7 (May, 2002).

99. A. Minoda, R. Sakagami, F. Yagisawa, T. Kuroiwa, K. Tanaka, Improvement of culture conditions and evidence for nuclear transformation by homologous recombination in a red alga, cyanidioschyzon merolae 10d. *Plant & cell physiology* **45**, 667 (Jun, 2004).

100. G. Walsh, Biopharmaceutical benchmarks 2010. *Nat Biotechnol* **28**, 917 (Sep, 2010).

101. S. P. Mayfield, S. E. Franklin, R. A. Lerner, Expression and assembly of a fully active antibody in algae. *Proc Natl Acad Sci U S A* **100**, 438 (Jan 21, 2003).

102. M. Tran, B. Zhou, P. L. Pettersson, M. J. Gonzalez, S. P. Mayfield, Synthesis and assembly of a full-length human monoclonal antibody in algal chloroplasts. *Biotechnology and bioengineering* **104**, 663 (Nov 1, 2009).

103. M. Tran, C. Van, D. J. Barrera, P. L. Pettersson, C. D. Peinado, J. Bui, S. P. Mayfield, Production of unique immunotoxin cancer therapeutics in algal chloroplasts. *Proc Natl Acad Sci U S A* **110**, E15 (Jan 2, 2013).

104. M. Tran, R. E. Henry, D. Siefker, C. Van, G. Newkirk, J. Kim, J. Bui, S. P. Mayfield, Production of anti-cancer immunotoxins in algae: Ribosome inactivating proteins as fusion partners. *Biotechnology and bioengineering* **110**, 2826 (Nov, 2013).

105. Y. K. Zhang, G. F. Shen, B. G. Ru, Survival of human metallothionein-2 transplastomic chlamydomonas reinhardtii to ultraviolet b exposure. *Acta biochimica et biophysica Sinica* **38**, 187 (Mar, 2006).

106. B. A. Rasala, M. Muto, P. A. Lee, M. Jager, R. M. Cardoso, C. A. Behnke, P. Kirk, C. A. Hokanson, R. Crea, M. Mendez, S. P. Mayfield, Production of therapeutic proteins in algae, analysis of expression of seven human proteins in the chloroplast of chlamydomonas reinhardtii. *Plant biotechnology journal* **8**, 719 (Aug, 2010).

107. A. Eichler-Stahlberg, W. Weisheit, O. Ruecker, M. Heitzer, Strategies to facilitate transgene expression in chlamydomonas reinhardtii. *Planta* **229**, 873 (Mar, 2009).

108. L. L. Bai, W. B. Yin, Y. H. Chen, L. L. Niu, Y. R. Sun, S. M. Zhao, F. Q. Yang, R. R. Wang, Q. Wu, X. Q. Zhang, Z. M. Hu, A new strategy to produce a defensin: Stable production of mutated np-1 in nitrate reductase-deficient chlorella ellipsoidea. *Plos One* **8**, e54966 (2013).

109. Y. Chen, Y. Wang, Y. Sun, L. Zhang, W. Li, Highly efficient expression of rabbit neutrophil peptide-1 gene in chlorella ellipsoidea cells. *Curr Genet* **39**, 365 (Jul, 2001).

110. Z. Y. Wen, F. Chen, Heterotrophic production of eicosapentaenoic acid by microalgae. *Biotechnology advances* **21**, 273 (Jul, 2003).

111. C. N. Kuratko, N. Salem, Docosahexaenoic acid from algal oil. *European Journal of Lipid Science and Technology*, (2013).

112. R. Winwood, C. Jacobsen, N. Skall Nielsen, A. Frisenfeldt Horn, A. Moltke S√ ∏rensen, Algal oil as a source of omega-3 fatty acids. *Food enrichment with omega-3 fatty acids*, 389 (2013).

113. M. L. Hamilton, R. P. Haslam, J. A. Napier, O. Sayanova, Metabolic engineering of phaeodactylum tricornutum for the enhanced accumulation of omega-3 long chain polyunsaturated fatty acids. *Metabolic engineering* **22C**, 3 (Dec 10, 2013).

114. R. Vaezi, J. A. Napier, O. Sayanova, Identification and functional characterization of genes encoding omega-3 polyunsaturated fatty acid biosynthetic activities from unicellular microalgae. *Marine drugs* **11**, 5116 (2013).

115. S. Gromer, J. K. Eubel, B. L. Lee, J. Jacob, Human selenoproteins at a glance. *Cellular and molecular life sciences : CMLS* **62**, 2414 (Nov, 2005).

116. L. H. Fu, X. F. Wang, Y. Eyal, Y. M. She, L. J. Donald, K. G. Standing, G. Ben-Hayyim, A selenoprotein in the plant kingdom. Mass spectrometry confirms that an opal codon (uga) encodes selenocysteine in chlamydomonas reinhardtii gluththione peroxidase. *The Journal of biological chemistry* **277**, 25983 (Jul 19, 2002).

117. Q. Hou, S. Qiu, Q. Liu, J. Tian, Z. Hu, J. Ni, Selenoprotein-transgenic chlamydomonas reinhardtii. *Nutrients* **5**, 624 (Mar, 2013).

118. R. Raja, S. Hemaiswarya, R. Rengasamy, Exploitation of dunaliella for beta-carotene production. *Applied microbiology and biotechnology* **74**, 517 (Mar, 2007).

119. A. C. Guedes, H. M. Amaro, F. X. Malcata, Microalgae as sources of carotenoids. *Marine Drugs* **9**, 625 (Apr, 2011).

120. L. Dufosse, P. Galaup, A. Yaron, S. M. Arad, P. Blanc, K. N. Chidambara Murthy, G. A. Ravishankar, Microorganisms and microalgae as sources of pigments for food use: A scientific oddity or an industrial reality? *Trends in Food Science & Technology* **16**, 389 (2005).

121. J. N. Rosenberg, G. A. Oyler, L. Wilkinson, M. J. Betenbaugh, A green light for engineered algae: Redirecting metabolism to fuel a biotechnology revolution. *Current opinion in biotechnology* **19**, 430 (Oct, 2008).

122. M. A. Borowitzka, High-value products from microalgae—their development and commercialisation. *Journal of Applied Phycology*, 1 (2013).

123. J. A. Del Campo, M. García-González, M. G. Guerrero, Outdoor cultivation of microalgae for carotenoid production: Current state and perspectives. *Applied Microbiology and Biotechnology* **74**, 1163 (2007).

124. E. Fukusaki, T. Nishikawa, K. Kato, A. Shinmyo, H. Hemmi, T. Nishino, A. Kobayashi, Introduction of the archaebacterial geranylgeranyl pyrophosphate synthase gene into chlamydomonas reinhardtii chloroplast. *Journal of bioscience and bioengineering* **95**, 283 (2003).

125. R. León, I. Couso, E. Fernández, Metabolic engineering of ketocarotenoids biosynthesis in the unicelullar microalga chlamydomonas reinhardtii. *Journal of Biotechnology* **130**, 143 (2007).

126. K. H. Wong, The University of Hong Kong (2006).

127. J. Steinbrenner, G. Sandmann, Transformation of the green alga haematococcus pluvialis with a phytoene desaturase for accelerated astaxanthin biosynthesis. *Applied and environmental microbiology* **72**, 7477 (2006).

128. B. F. Cordero, I. Couso, R. Leon, H. Rodriguez, M. A. Vargas, Enhancement of carotenoids biosynthesis in chlamydomonas reinhardtii by nuclear transformation using a phytoene synthase gene isolated from chlorella zofingiensis. *Applied microbiology and biotechnology* **91**, 341 (Jul, 2011).

129. I. Couso, M. Vila, H. Rodriguez, M. A. Vargas, R. Leon, Overexpression of an exogenous phytoene synthase gene in the unicellular alga chlamydomonas reinhardtii leads to an increase in the content of carotenoids. *Biotechnology progress* **27**, 54 (Jan-Feb, 2011).

130. G. Sun, X. Zhang, Z. Sui, Y. Mao, Inhibition of pds gene expression via the rna interference approach in dunaliella salina (chlorophyta). *Marine biotechnology* **10**, 219 (May-Jun, 2008).

131. M. Vila, I. Couso, R. Leon, Carotenoid content in mutants of the chlorophyte chlamydomonas reinhardtii with low expression levels of phytoene desaturase. *Process Biochemistry* **43**, 1147 (Oct, 2008).

132. M. Sun, K. X. Qian, N. Su, H. Y. Chang, J. X. Liu, G. F. Chen, Foot-and-mouth disease virus vp1 protein fused with cholera toxin b subunit expressed in chlamydomonas reinhardtii chloroplast. *Biotechnology Letters* **25**, 1087 (Jul, 2003).

133. D. G. Geng, Y. Q. Wang, P. Wang, W. B. Li, Y. R. Sun, Stable expression of hepatitis b surface antigen gene in dunaliella salina (chlorophyta). *Journal of Applied Phycology* **15**, 451 (Nov, 2003).

134. R. E. Soria-Guerra, Jocelín I. Ramírez-Alonso, Alejandro Ibáñez-Salazar, Dania O. Govea-Alonso, Luz M. T. Paz-Maldonado, Bernardo Bañuelos-Hernández, Schuyler S. Korban, Sergio Rosales-Mendoza, Expression of an hbcag-based antigen carrying angiotensin ii in chlamydomonas reinhardtii as a candidate hypertension vaccine. *Plant Cell, Tissue and Organ Culture* (2013).

135. F. Hempel, J. Lau, A. Klingl, U. G. Maier, Algae as protein factories: Expression of a human antibody and the respective antigen in the diatom phaeodactylum tricornutum. *Plos One* **6**, e28424 (2011).

136. D. M. He, K. X. Qian, G. F. Shen, Z. F. Zhang, Y. N. Li, Z. L. Su, H. B. Shao, Recombination and expression of classical swine fever virus (csfv) structural protein e2 gene in chlamydomonas reinhardtii chroloplasts. *Colloids and Surfaces B-Biointerfaces* **55**, 26 (Mar, 2007).

137. X. Wang, M. Brandsma, R. Tremblay, D. Maxwell, A. M. Jevnikar, N. Huner, S. Ma, A novel expression platform for the production of diabetes-associated autoantigen human glutamic acid decarboxylase (hgad65). *BMC biotechnology* **8**, 87 (2008).

138. R. Surzycki, K. Greenham, K. Kitayama, F. Dibal, R. Wagner, J. D. Rochaix, T. Ajam, S. Surzycki, Factors effecting expression of vaccines in microalgae. *Biologicals : journal of the International Association of Biological Standardization* **37**, 133 (Jun, 2009).

139. I. A. J. Dreesen, G. Charpin-El Hamri, M. Fussenegger, Heat-stable oral alga-based vaccine protects mice from staphylococcus aureus infection. *Journal of Biotechnology* **145**, 273 (Feb, 2010).

140. O. C. Demurtas, S. Massa, P. Ferrante, A. Venuti, R. Franconi, G. Giuliano, A chlamydomonas-derived human papillomavirus 16 e7 vaccine induces specific tumor protection. *Plos One* **8**, (Apr, 2013).

141. D. C. Gowda, E. A. Davidson, Protein glycosylation in the malaria parasite. *Parasitology today* **15**, 147 (Apr, 1999).

142. J. A. Gregory, F. Li, L. M. Tomosada, C. J. Cox, A. B. Topol, J. M. Vinetz, S. Mayfield, Algae-produced pfs25 elicits antibodies that inhibit malaria transmission. *Plos One* **7**, e37179 (2012).

143. D. Dauvillee, S. Delhaye, S. Gruyer, C. Slomianny, S. E. Moretz, C. d'Hulst, C. A. Long, S. G. Ball, S. Tomavo, Engineering the chloroplast targeted malarial vaccine antigens in chlamydomonas starch granules. *Plos One* **5**, e15424 (2010).

144. J. A. Gregory, A. B. Topol, D. Z. Doerner, S. Mayfield, Alga-produced cholera toxin-pfs25 fusion proteins as oral vaccines. *Applied and Environmental Microbiology* **79**, 3917 (Jul, 2013).

145. C. S. Jones, T. Luong, M. Hannon, M. Tran, J. A. Gregory, Z. Shen, S. P. Briggs, S. P. Mayfield, Heterologous expression of the c-terminal antigenic domain of the malaria vaccine candidate pfs48/45 in the green algae chlamydomonas reinhardtii. *Applied microbiology and biotechnology* **97**, 1987 (Mar, 2013).

146. I. Khozin-Goldberg, Z. Cohen, Unraveling algal lipid metabolism: Recent advances in gene identification. *Biochimie* **93**, 91 (Jan, 2011).

147. I. A. Guschina, J. L. Harwood, Lipids and lipid metabolism in eukaryotic algae. *Progress in Lipid Research* **45**, 160 (2006).

148. P. G. Roessler, Environmental control of glycerolipid metabolism in microalgae: Commercial implications and future research directions. *Journal of Phycology* **26**, 393 (1990).

149. W. L. Yu, W. Ansari, N. G. Schoepp, M. J. Hannon, S. P. Mayfield, M. D. Burkart, Modifications of the metabolic pathways of lipid and triacylglycerol production in microalgae. *Microb Cell Fact* **10**, 91 (2011).

150. Y. Q. Li, M. Horsman, B. Wang, N. Wu, C. Q. Lan, Effects of nitrogen sources on cell growth and lipid accumulation of green alga neochloris oleoabundans. *Appl Microbiol Biot* **81**, 629 (Dec, 2008).

151. A. P. Dean, D. C. Sigee, B. Estrada, J. K. Pittman, Using ftir spectroscopy for rapid determination of lipid accumulation in response to nitrogen limitation in freshwater microalgae. *Bioresource Technology* **101**, 4499 (Jun, 2010).

152. M. Chen, H. Tang, H. Ma, T. C. Holland, K. Y. Ng, S. O. Salley, Effect of nutrients on growth and lipid accumulation in the green algae dunaliella tertiolecta. *Bioresour Technol* **102**, 1649 (Jan, 2011).

153. P. G. Roessler, Effects of silicon deficiency on lipid-composition and metabolism in the diatom cyclotella-cryptica. *Journal of Phycology* **24**, 394 (Sep, 1988).

154. X. Li, H. Y. Hu, K. Gan, Y. X. Sun, Effects of different nitrogen and phosphorus concentrations on the growth, nutrient uptake, and lipid accumulation of a freshwater microalga scenedesmus sp. *Bioresource Technology* **101**, 5494 (Jul, 2010).

155. Z. Y. Liu, G. C. Wang, B. C. Zhou, Effect of iron on growth and lipid accumulation in chlorella vulgaris. *Bioresource Technology* **99**, 4717 (Jul, 2008).

156. S. Sriharan, D. Bagga, M. Nawaz, The effects of nutrients and temperature on biomass, growth, lipid production, and fatty-acid composition of cyclotella-cryptica reimann, lewin, and guillard. *Applied Biochemistry and Biotechnology* **28-9**, 317 (Spr, 1991).

157. C. Yeesang, B. Cheirsilp, Effect of nitrogen, salt, and iron content in the growth medium and light intensity on lipid production by microalgae isolated from freshwater sources in thailand. *Bioresource Technology* **102**, 3034 (Feb, 2011).

158. C. Yoo, S. Y. Jun, J. Y. Lee, C. Y. Ahn, H. M. Oh, Selection of microalgae for lipid production under high levels carbon dioxide. *Bioresource Technology* **101**, S71 (Jan, 2010).

159. L. Vigh, F. Joo, Modulation of membrane fluidity by catalytic-hydrogenation affects the chilling susceptibility of the blue-green-alga, anacystis-nidulans. *Febs Letters* **162**, 423 (1983).

160. N. Sato, N. Murata, Temperature shift-induced responses in lipids in the blue-green alga, *anabaena variabilis* the central role of diacylmonogalactosylglycerol in thermo-adaptation. *Biochimica et Biophysica Acta* **619**, 353 (1980).

161. D. V. Lynch, G. A. J. Thompson, Low temperature-induced alterations in the chloroplast and microsomal membranes of *dunaliella salina*. *Plant Physiology* **69**, 1369 (1982).

162. S. V. Khotimchenko, I. M. Yakovleva, Lipid composition of the red alga *tichocarpus crinitus* exposed to different levels of photon irradiance. *Phytochemistry* **66**, 73 (2005).

163. A. Sukenik, Y. Carmeli, T. Berner, Regulation of fatty acid composition by irradiance level in the eustigmatophyte *nannochloropsis* sp. *Journal of Phycology* **25**, 686 (1989).

164. D. M. Orcutt, G. W. Patterson, Effect of light intensity upon lipid composition of *nitzschia closterium* (*cylindrotheca fusiformis*). *Lipids* **9**, 1000 (1974).

165. Constant.G, K. Bloch, Effect of light intensity on lipid composition of eugena gracilis. *Journal of Biological Chemistry* **242**, 3538 (1967).

166. T. G. Dunahay, E. E. Jarvis, S. S. Dais, P. G. Roessler, Manipulation of microalgal lipid production using genetic engineering. *Applied Biochemistry and Biotechnology* **57-58**, (1996).

167. Z. T. Wang, N. Ullrich, S. Joo, S. Waffenschmidt, U. Goodenough, Algal lipid bodies: Stress induction, purification, and biochemical characterization in wild-type and starchless chlamydomonas reinhardtii. *Eukaryot Cell* **8**, 1856 (Dec, 2009).

168. R. Radakovits, P. M. Eduafo, M. C. Posewitz, Genetic engineering of fatty acid chain length in phaeodactylum tricornutum. *Metabolic Engineering* **13**, 89 (Jan, 2011).

169. D. R. Ort, A. Melis, Optimizing antenna size to maximize photosynthetic efficiency. *Plant Physiol* **155**, 79 (Jan, 2011).

170. J. H. Mussgnug, S. Thomas-Hall, J. Rupprecht, A. Foo, V. Klassen, A. McDowall, P. M. Schenk, O. Kruse, B. Hankamer, Engineering photosynthetic light capture: Impacts on improved solar energy to biomass conversion. *Plant Biotechnology Journal* **5**, 802 (Nov, 2007).

171. J. H. Mussgnug, L. Wobbe, I. Elles, C. Claus, M. Hamilton, A. Fink, U. Kahmann, A. Kapazoglou, C. W. Mullineaux, M. Hippler, J. Nickelsen, P. J. Nixon, O. Kruse, Nab1 is an rna binding protein involved in the light-regulated differential expression of the light-harvesting antenna of chlamydomonas reinhardtii. *The Plant cell* **17**, 3409 (Dec, 2005).

172. L. Wobbe, O. Blifernez, C. Schwarz, J. H. Mussgnug, J. Nickelsen, O. Kruse, Cysteine modification of a specific repressor protein controls the translational status of nucleus-encoded lhcii mrnas in chlamydomonas. *Proc Natl Acad Sci U S A* **106**, 13290 (Aug 11, 2009).

173. J. Beckmann, F. Lehr, G. Finazzi, B. Hankamer, C. Posten, L. Wobbe, O. Kruse, Improvement of light to biomass conversion by de-regulation of light-harvesting protein translation in chlamydomonas reinhardtii. *Journal of Biotechnology* **142**, 70 (Jun 1, 2009).

174. R. Ritchie, A. Larkum, Modelling photosynthesis in shallow algal production ponds. *Photosynthetica* **50**, 481 (2012).

175. N. Keren, A. Krieger-Liszkay, Photoinhibition: Molecular mechanisms and physiological significance. *Physiologia Plantarum* **142**, 1 (May, 2011).

176. G. Rea, M. Lambreva, F. Polticelli, I. Bertalan, A. Antonacci, S. Pastorelli, M. Damasso, U. Johanningmeier, M. T. Giardi, Directed evolution and in silico analysis of reaction centre proteins reveal molecular signatures of photosynthesis adaptation to radiation pressure. *Plos One* **6**, (Jan 13, 2011).

177. J. A. Gimpel, S. P. Mayfield, Analysis of heterologous regulatory and coding regions in algal chloroplasts. *Applied Microbiology and Biotechnology* **97**, 4499 (May, 2013).

178. D. J. Vinyard, J. Gimpel, G. M. Ananyev, M. A. Cornejo, S. S. Golden, S. P. Mayfield, G. C. Dismukes, Natural variants of photosystem ii subunit d1 tune photochemical fitness to solar intensity. *The Journal of biological chemistry* **288**, 5451 (Feb 22, 2013).

179. M. Chen, R. E. Blankenship, Expanding the solar spectrum used by photosynthesis. *Trends in plant science* **16**, 427 (Aug, 2011).

180. C. A. Raines, Increasing photosynthetic carbon assimilation in c-3 plants to improve crop yield: Current and future strategies. *Plant Physiol* **155**, 36 (Jan, 2011).

181. D. C. Ducat, P. A. Silver, Improving carbon fixation pathways. *Current Opinion in Chemical Biology* **16**, 337 (Aug, 2012).

182. S. M. Whitney, R. L. Houtz, H. Alonso, Advancing our understanding and capacity to engineer nature's co2-sequestering enzyme, rubisco. *Plant Physiol* **155**, 27 (Jan, 2011).

183. R. J. Spreitzer, S. R. Peddi, S. Satagopan, Phylogenetic engineering at an interface between large and small subunits imparts land-plant kinetic properties to algal rubisco. *Proc Natl Acad Sci U S A* **102**, 17225 (Nov 22, 2005).

184. T. Genkov, M. Meyer, H. Griffiths, R. J. Spreitzer, Functional hybrid rubisco enzymes with plant small subunits and algal large subunits engineered rbcs cdna for expression in chlamydomonas. *Journal of Biological Chemistry* **285**, 19833 (Jun 25, 2010).

185. G. H. Zhu, I. Kurek, L. Liu, Engineering photosynthetic enzymes involved in co2-assimilation by gene shuffling. *Chloroplast: Basics and Applications* **31**, 307 (2010).

186. X. Johnson, Manipulating rubisco accumulation in the green alga, chlamydomonas reinhardtii. *Plant Molecular Biology* **76**, 397 (Jul, 2011).

187. J. R. Benemann, Microalgae aquaculture feeds. *Journal of Applied Phycology* **4**, 233 (1992).

188. J. J. Milledge, Commercial application of microalgae other than as biofuels: A brief review. *Reviews in Environmental Science and Bio/Technology* **10**, 31 (2011).

189. E. W. Becker, Micro-algae as a source of protein. *Biotechnology advances* **25**, 207 (Mar-Apr, 2007).

190. A. G. Mathew, R. Cissell, S. Liamthong, Antibiotic resistance in bacteria associated with food animals: A united states perspective of livestock production. *Foodborne pathogens and disease* **4**, 115 (Summer, 2007).

191. N. Peak, C. W. Knapp, R. K. Yang, M. M. Hanfelt, M. S. Smith, D. S. Aga, D. W. Graham, Abundance of six tetracycline resistance genes in wastewater lagoons at cattle feedlots with different antibiotic use strategies. *Environmental microbiology* **9**, 143 (Jan, 2007).

192. T. W. Alexander, J. L. Yanke, T. Reuter, E. Topp, R. R. Read, B. L. Selinger, T. A. McAllister, Longitudinal characterization of antimicrobial resistance genes in feces shed from cattle fed different subtherapeutic antibiotics. *BMC microbiology* **11**, 19 (2011).

193. N. Campos-Quevedo, S. Rosales-Mendoza, L. M. a. T. Paz-Maldonado, L. Mart√ ≠nez-Salgado, J. C. Guevara-Arauza, R. E. Soria-Guerra, Production of milk-derived bioactive peptides as precursor chimeric proteins in chloroplasts of chlamydomonas reinhardtii. *Plant Cell, Tissue and Organ Culture (PCTOC)*, 1 (2013).

194. T. R. Brown, R. C. Brown, A review of cellulosic biofuel commercial-scale projects in the united states. *Biofuels Bioproducts & Biorefining-Biofpr* **7**, 235 (May-Jun, 2013).

195. M. Abramson, O. Shoseyov, Z. Shani, Plant cell wall reconstruction toward improved lignocellulosic production and processability. *Plant Science* **178**, 61 (2010).

196. B. A. Rasala, P. A. Lee, Z. Shen, S. P. Briggs, M. Mendez, S. P. Mayfield, 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 (2012).

197. C. Forsberg, J. Phillips, S. Golovan, M. Fan, R. Meidinger, A. Ajakaiye, D. Hilborn, R. Hacker, The enviropig physiology, performance, and contribution to nutrient management advances in a regulated environment: The leading edge of change in the pork industry. *Journal of Animal Science* **81**, E68 (2003).

198. S. M. Yoon, S. Y. Kim, K. F. Li, B. H. Yoon, S. Choe, M. M. Kuo, Transgenic microalgae expressing escherichia coli appa phytase as feed additive to reduce phytate excretion in the manure of young broiler chicks. *Applied microbiology and biotechnology* **91**, 553 (Aug, 2011).

199. L. Christenson, R. Sims, Production and harvesting of microalgae for wastewater treatment, biofuels, and bioproducts. *Biotechnology advances* **29**, 686 (Nov-Dec, 2011).

200. C. Y. Chen, K. L. Yeh, R. Aisyah, D. J. Lee, J. S. Chang, Cultivation, photobioreactor design and harvesting of microalgae for biodiesel production: A critical review. *Bioresource Technology* **102**, 71 (Jan, 2011).

201. T. M. Mata, A. n. A. Martins, N. S. Caetano, Microalgae for biodiesel production and other applications: A review. *Renewable and Sustainable Energy Reviews* **14**, 217 (2010).

202. G. Q. Chen, F. Chen, Growing phototrophic cells without light. *Biotechnology letters* **28**, 607 (2006).

203. X. Miao, Q. Wu, Biodiesel production from heterotrophic microalgal oil. *Bioresource technology* **97**, 841 (Apr, 2006).

204. A. Hallmann, M. Sumper, The chlorella hexose/h+ symporter is a useful selectable marker and biochemical reagent when expressed in volvox. *Proc Natl Acad Sci U S A* **93**, 669 (Jan 23, 1996).

205. L. A. Zaslavskaia, J. C. Lippmeier, C. Shih, D. Ehrhardt, A. R. Grossman, K. E. Apt, Trophic conversion of an obligate photoautotrophic organism through metabolic engineering. *Science* **292**, 2073 (Jun 15, 2001).

206. A. Doebbe, J. Rupprecht, J. Beckmann, J. H. Mussgnug, A. Hallmann, B. Hankamer, O. Kruse, Functional integration of the hup1 hexose symporter gene into the genome of c. Reinhardtii: Impacts on biological h(2) production. *Journal of biotechnology* **131**, 27 (Aug 1, 2007).

207. H. Fischer, I. Robl, M. Sumper, N. Kröger, Targeting and covalent modification of cell wall and membrane proteins heterologously expressed in the diatom cylindrotheca fusiformis (bacillariophyceae). *Journal of Phycology* **35**, 113.

208. N. R. Moheimani, M. A. Borowitzka, Limits to productivity of the alga pleurochrysis carterae (haptophyta) grown in outdoor raceway ponds. *Biotechnology and bioengineering* **96**, 27 (Jan 1, 2007).

209. J. B. Park, R. J. Craggs, Algal production in wastewater treatment high rate algal ponds for potential biofuel use. *Water science and technology : a journal of the International Association on Water Pollution Research* **63**, 2403 (2011).

210. J. B. Park, R. J. Craggs, A. N. Shilton, Wastewater treatment high rate algal ponds for biofuel production. *Bioresource technology* **102**, 35 (Jan, 2011).

211. M. Borowitzka, N. Moheimani, in *Algae for biofuels and energy,* M. A. Borowitzka, N. R. Moheimani, Eds. (Springer Netherlands, 2013), vol. 5, pp. 133-152.

212. A. Vonshak, S. Boussiba, A. Abeliovich, A. Richmond, Production of spirulina biomass: Maintenance of monoalgal culture outdoors. *Biotechnology and bioengineering* **25**, 341 (Feb, 1983).

213. D. Vandamme, I. Foubert, K. Muylaert, Flocculation as a low-cost method for harvesting microalgae for bulk biomass production. *Trends Biotechnol* **31**, 233 (Apr, 2013).

214. R. J. Powell, R. T. Hill, Rapid aggregation of biofuel-producing algae by the bacterium bacillus sp. Strain rp1137. *Applied and environmental microbiology* **79**, 6093 (Oct, 2013).

215. J. Zhang, B. Hu, A novel method to harvest microalgae via co-culture of filamentous fungi to form cell pellets. *Bioresource technology* **114**, 529 (Jun, 2012).

216. W. Zhou, M. Min, B. Hu, X. Ma, Y. Liu, Q. Wang, J. Shi, P. Chen, R. Ruan, Filamentous fungi assisted bio-flocculation: A novel alternative technique for harvesting heterotrophic and autotrophic microalgal cells. *Separation and Purification Technology*, (2013).

217. A. Besson, P. Guiraud, High-ph-induced flocculation-flotation of the hypersaline microalga dunaliella salina. *Bioresource technology* **147**, 464 (Nov, 2013).

218. D. Vandamme, I. Foubert, I. Fraeye, B. Meesschaert, K. Muylaert, Flocculation of chlorella vulgaris induced by high ph: Role of magnesium and calcium and practical implications. *Bioresource technology* **105**, 114 (Feb, 2012).

219. K. Spilling, J. Sepp√ § l√ §, T. Tamminen, Inducing autoflocculation in the diatom phaeodactylum tricornutum through co2 regulation. *Journal of Applied Phycology* **23**, 959 (2011/12/01, 2011).

CHAPTER 2:

MICROALGAE COME OF AGE AS A PLATFORM FOR RECOMBINANT PROTEIN PRODUCTION

Biotechnol Lett (2010) 32:1373–1383
DOI 10.1007/s10529-010-0326-5

REVIEW

Micro-algae come of age as a platform for recombinant protein production

Elizabeth Specht · Shigeki Miyake-Stoner · Stephen Mayfield

Received: 31 March 2010 / Accepted: 7 June 2010 / Published online: 17 June 2010

Abstract A complete set of genetic tools is still being developed for the micro-alga *Chlamydomonas reinhardtii*. Yet even with this incomplete set, this photosynthetic single-celled plant has demonstrated significant promise as a platform for recombinant protein expression. In recent years, techniques have been developed that allow for robust expression of genes from both the nuclear and plastid genome. With these advances, many research groups have examined the pliability of this and other micro-algae as biological machines capable of producing recombinant peptides and proteins. This review describes recent successes in recombinant protein production in *Chlamydomonas*, including production of complex mammalian therapeutic proteins and monoclonal antibodies at levels sufficient for production at economic parity with existing production platforms. These advances have also shed light on the details of algal protein production at the molecular level, and provide insight into the next steps for optimizing micro-algae as a useful platform for the production of therapeutic and industrially relevant recombinant proteins.

E. Specht · S. Miyake-Stoner · S. Mayfield (✉)
San Diego Center for Algae Biotechnology,
Department of Biological Sciences, University
of California, San Diego, 9500 Gilman Dr., La Jolla,
CA 92093-0368, USA
e-mail: smayfield@ucsd.edu

Keywords *Chlamydomonas* · Micro-algae · Recombinant protein expression · Therapeutic protein production

Introduction

As genetically accessible photosynthetic organisms, algae are now recognized for their potential as a platform for recombinant protein expression where large scale and reduced material costs are important. The most-commonly used eukaryotic model alga, *Chlamydomonas reinhardtii,* has recently been shown to be able to fill this role, and this review will primarily discuss the technical and biological advances made in recombinant protein production in this alga. Algae have now come of age as a platform for recombinant protein expression.

All three genomes (chloroplast, mitochondrial, and nuclear) can be transformed in *Chlamydomonas*, and each has distinct transcriptional, translational, and post-translational properties that make them distinct. Each of these genomes has been fully sequenced, providing a wealth of information and a strong foundation for targeted manipulation (Maul et al. 2002; Popescu and Lee 2007; Merchant et al. 2007). Recent efforts have primarily focused on understanding and improving gene transformation, mRNA transcript accumulation, and protein accumulation of nuclear and chloroplast recombinant genes.

Biotechnol Lett (2010) 32:1373–1383

Chlamydomonas is an excellent system for reasons beyond its genetic and metabolic malleability. This alga has a rapid doubling time (about 10 h), it is easily scaled in homogenous culture as an aqueous microbe, it can be grown either photoautotrophically or with acetate as a reduced carbon source, and it has a controllable and rapid sexual cycle (about 2 weeks) with stable and viable haploids.

Advantages of algal protein production

Protein production in plants provides a number of advantages not found in other production platforms. First, a major advantage that all plant protein production systems have over cell culture systems (including bacteria, yeast, and mammalian cell culture) is the potential for significant reduction in cost. It is estimated that protein production in transgenic plants can be as much as four orders of magnitude less expensive than production in mammalian cell culture, on a per gram of unpurified protein basis (Dove 2002). Secondly, plant-produced proteins are not susceptible to viral or prion contamination that can harm humans, as is always a concern with animal cell culture (Chebolu and Daniell 2009). Third, as eukaryotes, algae and other plants possess the chaperones and cellular machinery required to fold complex human proteins that bacteria and yeast may not be able to process properly (Franklin and Mayfield 2004). Finally, many species of green algae are considered GRAS (generally regarded as safe) (Rosenberg et al. 2008), meaning that if the protein can be expressed in a bioavailable form, purification steps could potentially be eliminated altogether.

Algae possess a number of advantages over transgenic plant systems for the production of recombinant proteins. They can be grown in contained bioreactors, reducing the risk of contamination of the production system by airborne contaminants, and also protecting the environment from any potential flow of transgenes into the surrounding ecosystem. Growth in containment also greatly reduces the potential for loss of the crop due to predation or pathogen attack. Algae progress from initial transformation to large-scale protein production in a matter of weeks, compared to timescales on the order of months or years in higher plants such as corn or tobacco (Franklin and Mayfield 2004). As micro-algae are all a single cell type, there should also be less variation in recombinant protein accumulation, making downstream processing more uniform.

Production of recombinant proteins in chloroplasts also possesses several unique attributes. At present transgenic proteins can accumulate to much higher levels in the chloroplast than when expressed from the nuclear genome, mainly because plastids lack gene silencing mechanisms and other mechanisms that reduce recombinant protein production from nuclear encoded genes (Bock 2007). Chloroplasts can be transformed with multiple genes in a single event, due to the availability of multiple insertion sites, as well as an ability to process polycistronic transcripts, allowing an entire gene cassette to be regulated by a single promoter (Rymarquis et al. 2006; Bock 2007). Additionally, proteins produced within the chloroplast are not glycosylated (Franklin and Mayfield 2005), which can prove useful in many applications such as producing antibodies that are similar to native antibodies in their ability to recognize their antigen, but whose lack of glycosylation prevents them from recruiting killer cells (Tran et al. 2009). In fact, it is estimated that over two-thirds of the therapeutic human monoclonal antibodies in the testing pipeline do not require glycosylation for therapeutic function (Dove 2002).

Genetic tools and techniques

Transformation techniques

The plastid genome can be reliably transformed through homologous recombination using bombardment by DNA-coated gold or tungsten particles (Koop et al. 2007). Nuclear transformation in algae can also be achieved by biolistic bombardment, but the preferred methods are electroporation or agitation with glass beads using a cell-wall defective strain (Eichler-Stahlberg et al. 2009; Leon and Fernandez 2007). New transformation techniques using the Cre/lox recombination system have been demonstrated to recombine in the nuclear genome of *Chlamydomonas* (Heitzer and Zschoernig 2007). Robust in vivo recombinant reporters, including GFPs (Fuhrmann et al. 1999; Franklin et al. 2002) and luciferases (Mayfield and Schultz 2004; Shao and Bock 2008), have been developed for tracking both nuclear and

Biotechnol Lett (2010) 32:1373–1383

chloroplast gene expression. Techniques that have been employed previously on higher plants, such as transformation by *Agrobacterium tumefaciens,* have also been demonstrated to work with *Chlamydomonas* (Kumar et al. 2004). However, nuclear transformants still generally fail to accumulate recombinant proteins to the levels observed in plastids, most likely due to nuclear silencing mechanisms (see *Nuclear Gene Silencing* below).

Codon optimization

As with other expression systems, codon optimization has played a large role in the success of recombinant protein expression in both the chloroplast and nuclear genomes of *Chlamydomonas* (Heitzer et al. 2007). The nuclear genome and the plastid genome have highly divergent codon usage, with the chloroplast preferring an A or T in the wobble position while the nuclear genome prefers a G or C (Nakamura et al. 2000). Using GFP, early work showed that codon optimization to reflect the genome bias could increase transgene protein accumulation 5-fold in the nucleus (Fuhrmann et al. 1999) and up to 80-fold in the chloroplast (Franklin et al. 2002). Today, recombinant genes are universally codon optimized for improved protein expression in almost every system (Xia 2007; Puigbo et al. 2007, 2008).

Chloroplast gene regulation

Promoter and regulatory mRNA untranslated region (UTR) choices are both crucial factors in transgene expression levels. In chloroplasts, the most successful promoter to date in algae is the *psbA* promoter in combination with the *psbA* UTRs (Manuell et al. 2007; Surzycki et al. 2009). However, there are two caveats with using the *psbA* regulatory elements. First, they appear to be highly auto-attenuated; if any of the *psbA* gene product (D1 protein) is present, it will strongly decrease expression of any recombinant coding sequence under the control of its 5′-UTR. Secondly, since D1 is essential for the function of photosystem II, *psbA* knockouts are nonphotosynthetic. This would clearly negate the benefits of using a photosynthetic organism for protein production. There is evidence that reintroducing an attenuated *psbA* gene at a new locus elsewhere in the plastid genome can restore photosynthesis while only mildly reducing recombinant protein production (Manuell et al. 2007).

Other UTRs in use for transgene expression include those from the endogenous *atpA*, *rbcL*, and *psbD* genes (Fletcher et al. 2007; Hallmann 2007). These have been used with varying levels of success, though as with the *psbA* promoter, it is unclear why certain regulatory elements engender high expression levels with some genes but not others (Marin-Navarro et al. 2007). While endogenous promoters have been primarily used, other exogenously induced expression systems have been explored in the chloroplast. It has been demonstrated that inducible systems, such as the lac operon system from *E. coli,* can be engineered into the *Chlamydomonas* chloroplast (Kato et al. 2007), and more recently a riboswitch was shown to work to regulate translation in *Chlamydomonas* chloroplasts using a small molecule for induction (Croft et al. 2007), similar to their use in many other expression platforms (Suess 2005; Winkler and Breaker 2005).

While regulation of gene expression occurs at both the transcriptional and translational level in the nucleus, it appears that most regulation is post-transcriptional in the plastid (Marin-Navarro et al. 2007). However, many of the activators and suppressors of mRNA splicing and processing in the chloroplast are indeed encoded by nuclear genes (Boudreau et al. 2000; Somanchi et al. 2005; Raynaud et al. 2007; Schwarz et al. 2007; Loiselay et al. 2008). A multi-component copper-induced system has been designed as a switch for chloroplast protein expression. This system utilizes the nuclear-encoded Nac2 chloroplast protein necessary for stable accumulation of *psbD* RNA by acting on its 5′-regulatory region. By transforming a copper induced cytochrome c6 promoter fused to the *Nac2* coding sequence into a *Nac2* deficient strain, proteins encoded with the *psbD* regulatory region only accumulate in the presence of copper (Surzycki et al. 2007). Additionally, nuclear gene products are required for splicing of group I introns in the chloroplast 23S rRNA and *psbA* genes, and can potentially be used to regulate plastid gene expression (Li et al. 2002).

Post-transcriptional control in the chloroplast is mediated by both cis- and trans-acting-elements. The 5′-UTR of chimeric chloroplast mRNAs was shown to significantly impact recombinant protein production,

Biotechnol Lett (2010) 32:1373–1383

while 3′-UTRs had little if any effect (Barnes et al. 2005). Cis-acting elements in 3′-UTRs include UG-repeats that are used in circadian protein regulation (Kiaulehn et al. 2007) and inverted repeat sequences that contribute to mRNA processing (Goldschmidt-Clermont et al. 2008). Recent work has identified a complex comprising an RNA stabilizing factor and a translational activator that appears to be specific to the D2 protein of photosystem II (Schwarz et al. 2007). Post-transcriptional control mechanisms can also have synergistic or antagonistic effects. Kasai et al. (2003) used the GUS reporter controlled by various endogenous 3′- and 5′-UTRs to determine that there is an inverse correlation between protein accumulation and transcript stability, suggesting a feedback mechanism. Together all of these data suggest that regulation of mRNA translation has the greatest impact on recombinant protein accumulation, and is clearly an area where additional research seems likely to identify mechanisms for increased recombinant protein accumulation. For a comprehensive review of chloroplast translation regulation, see Marin-Navarro et al. 2007.

Nuclear gene regulation

In the nucleus, the most commonly used promoters are those from the *HSP70A, psaD*, and *rbcS2* genes (Schroda et al. 2002; Fischer and Rochaix 2001; Berthold et al. 2002). Chimeric promoters have also demonstrated high levels of transcription and expression (Schroda et al. 2000; Fischer and Rochaix 2001; Wu et al. 2008). Additionally, placing endogenous intronic sequences in transgenes has been shown to enhance expression regardless of orientation or position. The first intron from the endogenous rbcS2 gene has shown particular efficacy in increasing mRNA and protein accumulation and is now commonly used to enhance recombinant gene expression (Lumbreras et al. 1998).

In the nucleus, gene expression can be induced by a number of factors, including heat-shock or metal addition (Wu et al. 2008; Ferrante et al. 2008). The Nit1 promoter suppresses transcription in the presence of ammonia, but induces transcription when cells are grown in nitrate- or nitrite-containing media (Ohresser et al. 1997). More recently, iron-deficiency response elements (FeREs) in nuclear gene promoters have been characterized in *Chlamydomonas* (Fei and Deng 2007). Signaling cascades triggered by photo-oxidative stress in the chloroplast can also activate transcription of specific nuclear genes, indicating that gene regulation in *Chlamydomonas* is not necessarily localized to the site of signal production (Fischer et al. 2007). More recently, signaling molecules such as Mg-protoporphyrin and heme produced in the plastid have been shown to activate transcription of nuclear genes such as HSP70A through an interaction with the plastid response element (PRE) (von Gromoff et al. 2008). While these inducible systems provide great insight into gene regulatory strategies, there is still significant work to be done before robust production of recombinant proteins can be routinely achieved from nuclear genes in algae.

Nuclear gene silencing

Transgene silencing is a significant obstacle for recombinant protein expression in *Chlamydomonas* nuclear transformants, but recent work is helping to overcome this problem (Casas-Mollano et al. 2007, 2008a, b). Neupert et al. (2009) have developed strains with impaired transgene silencing by using UV mutagenesis and selection on media that permits higher antibiotic tolerance proportional to higher expression of the transgene product, to select strains with improved protein accumulation. However, their most impressive yields of exogenous protein accumulation are only 0.2% of total soluble protein (TSP), as compared to nearly 10% TSP obtained in plastids (Manuell et al. 2007). It has been postulated that gene silencing may be difficult to eliminate because it may have evolved as a protective measure against intracellular pathogens or viruses (Rosenberg et al. 2008; Neupert et al. 2009). It appears that combating gene silencing will be a major hurdle before recombinant proteins can be expressed at economically viable levels from nuclear transgenes in *Chlamydomonas*.

Current successes in algal protein production

Plastid transformation in higher plants has identified a few recombinant proteins that accumulate to very high levels, reported up to as high as 70% of total protein for some antibiotic proteins in tobacco leaves (Oey et al. 2009), but in general recombinant protein expression is highly variable. For an extensive review

Biotechnol Lett (2010) 32:1373–1383

of vaccine production in plants, see Davoodi-Semiromi et al. (2009). However, very recent work in *Chlamydomonas* has demonstrated that fully bioactive proteins can indeed be produced to appreciable levels in green algae. A recent technique has been exploited to improve protein accumulation and stability by expression of cleavable fusions to highly expressed endogenous proteins (Muto et al. 2009), or to highly expressed recombinant proteins in the chloroplast (Rasala et al. 2010).

The first demonstration of mammalian protein expression in the chloroplast was of a large single-chain antibody (HSV8-lsc) directed against glycoprotein D of the herpes simplex virus (Mayfield et al. 2003). The protein was soluble, suggesting that it was correctly folded, and electrophoresis indicated the formation of the dimer by disulfide bond formation. This work was followed up by the expression of a single chain fragment variable antibody that accumulated to 0.54% TSP (Mayfield and Franklin 2005). More recently a full-length human IgG1 monoclonal antibody, directed against anthrax protective antigen 83 (83K7C), was expressed in the chloroplast of *Chlamydomonas* (Tran et al. 2009). Unlike the previously expressed lsc antibody, this antibody was assembled in the chloroplast from separately expressed light chain and heavy chain proteins, and it could be purified at 100 μg per 1 g dry algal biomass, and was found to have binding activity identical to that of the same antibody expressed in a traditional mammalian cell culture system (Tran et al. 2009).

A host of non-antibody recombinant proteins have been expressed in the chloroplast for therapeutic purposes. The human metallothionein-2 gene product, which is considered to have anti-radiation function, was expressed and demonstrated to improve the survivorship of transgenic algae compared to wild type algae (Zhang et al. 2006). The human tumor necrosis factor-related apoptosis-inducing ligand (TRAIL) protein, known to induce apoptosis in virus-infected and tumor cells, accumulated in the chloroplast at 0.43–0.67% TSP based on densitometric analysis of Western blot (Yang et al. 2006). The expression of human glutamic acid decarboxylase (hGAD65) was also achieved in the chloroplast of *Chlamydomonas* (Wang et al. 2008). This protein is an important autoantigenic marker in type I diabetes, and the algae-derived protein remained immunologically active, accumulating at 0.25–0.3% TSP. Human erythropoietin (Epo), used in the treatment of anemia, was fused to an export sequence and expressed from the nuclear genome (Eichler-Stahlberg et al. 2009). Protein was detected in and isolated from the culture medium, although with uncharacterized post-translational modification closely matching the mass of endogenous human Epo. Also, biologically active bovine mammary-associated serum amyloid (M-SAA) was expressed in the chloroplast (Manuell et al. 2007). Notably, the accumulation of this soluble protein was above 5% TSP as quantified by Western blot, with levels of expression determined to be more than twice that amount using ELISA quantification.

A fusion protein containing the foot-and-mouth disease virus VP1 gene and the cholera toxin B subunit (CTBVP1) was produced in the chloroplast as well (Sun et al. 2003). The protein was reported to accumulate to 3% total protein by ELISA quantification, and retained GM1-ganglioside binding activity and antigenicity. Another viral protein, classical swine fever virus E2 structural protein, was expressed in the chloroplast (He et al. 2007). ELISA quantification indicated the accumulation of the E2 protein to 1.5–2% TSP, and retained immunological activity. Similarly, along with a list of other recombinant proteins, the white spot syndrome virus protein 28 (VP28) was expressed in the chloroplast (Surzycki et al. 2009). VP28 was reported to accumulate to a striking 10.5% TSP, although no data was presented to show how this level of expression was determined. Recently, the D2 fibronectin-binding domain from a *Staphylococcus aureus* protein was fused to the B subunit of cholera toxin, and expressed in the chloroplast (Dreesen et al. 2010). The transgenic algae were fed to mice, and induced resistance against lethal doses of *S. aureus*, presumably by eliciting a systemic antigenic response to the *S. aureus* peptide. This is the first demonstration of the functional possibility of orally delivered vaccines from algal production.

More recently, Rasala et al. (2010) attempted the expression of a set of seven recombinant proteins in the chloroplast and were met with very good success (four out of the seven genes expressed at economically viable levels). This work demonstrates that recombinant protein expression in algal chloroplasts is on par with any other expression platform, and

Biotechnol Lett (2010) 32:1373–1383

shows that expression of complex mammalian proteins is as likely to be achieved in algae as it is in any eukaryotic system. These data are summarized in Table 1.

Potential future applications

Oral vaccines

Traditional vaccines are normally produced from an attenuated or killed form of the pathogenic organism itself. An alternative approach is to produce a pathogen antigen as a recombinant protein, and this is now used for some select vaccines like the hepatitis A vaccine (Powdrill and Johnston 1991). As algae contain very sophisticated protein folding machinery that bacteria and other prokaryotes lack, algae can be used to produce complex eukaryotic proteins that cannot be easily produced on large scale in bacterial culture without costly denaturation and refolding steps. Algae are also ideal for producing vaccines against pathogens which exhibit little or no glycosylation, such as those from the parasites plasmodium (Gowda and Davidson 1999).

Algae are also especially suited for the production of oral vaccines on a large scale due to their GRAS status. Once the process of oral delivery of a vaccine protein is refined, algae could produce inexpensive oral vaccines, making vaccination an accessible form of disease management for a whole host of third-world diseases. Furthermore, the oral delivery method and the option to store doses at ambient temperature would allow vaccines to be transported and administered to remote populations without the need for expensive refrigeration or highly trained medical personnel (Chebolu and Daniell 2009).

Evidence for the feasibility of plastid-produced vaccines has been provided by several groups who have produced antigens that elicit similar immune responses as the actual pathogen when injected in standard vaccine adjuvants (Tregoning et al. 2003; Koya et al. 2005; Molina and Shoenfeld 2005; Chebolu and Daniell 2009). Furthermore, fusions of antigens with the cholera toxin B subunit show promise for eliciting immune response from mucosal delivery alone, as the cholera toxin B subunit allows a fused protein to penetrate the intestinal lining (Sun et al. 2003; Harakuni et al. 2005).

Discovery of novel bioactive molecules

Screening of diverse algal species for novel bioactive compounds has become a field of intense interest lately, as algae are a rich source of secondary metabolites that often have industrial or nutritional implications. A wide range of functional molecules produced by algae—everything from antioxidants to pigments used in laboratory analytical techniques—have been isolated (Plaza et al. 2009). High-throughput screening methods comprising pressurized liquid extraction; functional characterization of antioxidant or antimicrobial activity; and chemical characterization by HPLC and GC-MS have been optimized to streamline the screening process (Plaza et al. 2010).

Protein secretion

Figure 1 illustrates the basic structure of *Chlamydomonas reinhardtii*, as well as the cellular locations of recombinant protein accumulation depending on the gene construct and transformation method. Plastid transformation results in accumulation of the transgene product in the single large chloroplast. Nuclear gene products accumulate in the cytosol by default, but nuclear transformation with appropriate signal sequences allows for targeting to the endoplasmic reticulum and Golgi for export (Griesbeck et al. 2006) or for localization to the cell membrane, which may be sufficient for antigenic recognition in the case of an oral vaccine (Eichler-Stahlberg et al. 2009).

Other algal transformation efforts

Genetic manipulation of algae is no longer limited to *Chlamydomonas reinhardtii*. Recent successes in generating transgenic algae are growing in number. Exogenous genes have been expressed in the unicellular charophyte alga, *Closterium peracerosum–strigosum–littorale* complex (Abe et al. 2008). The nuclear genome of volvocine alga *Gonium pectorale* has been stably transformed (Lerche and Hallmann 2009), as has the chlorophyceae *Haematococcus pluvialis* (Kathiresan and Sarada 2009), by co-cultivation with *Agrobacterium* (Kathiresan et al. 2009). Some successes have been made by transient transformation of marine chlorarachniophyte *Lotharella amoebiformis* (Hirakawa et al. 2008), chlorophyta alga *Ulva pertusa*

Table 1 Recent successes in therapeutic protein production in algae

Gene expressed	Function	Expression level achieved	Application	Source
HSV8-lsc	First mammalian protein expressed, antibody	Detectable	Pharmaceutical	Mayfield et al. (2003)
CTB-VP1	Cholera toxin B subunit fused to foot and mouth disease VP1	3% TSP	Vaccine	Sun et al. (2003)
HSV8-scFv	Classic single-chain antibody	0.5% TSP	Pharmaceutical	Mayfield et al. (2005)
hMT-2	Human metallothionine-2	Detectable	Pharmaceutical, UV-protection	Zhang et al. (2006)
hTRAIL	Human tumor necrosis factor-related apoptosis-inducing ligand (TRAIL)	~0.67% TSP	Pharmaceutical	Yang et al. (2006)
M-SAA	Bovine mammary-associated serum amyloid	~5% TSP	Therapeutics, oral delivery	Manuell et al. (2007)
CSFV-E2	Swine fever virus E2 viral protein	~2% TSP	Vaccine	He et al. (2007)
hGAD65	Diabetes-associated anutoantigen human glutamic acid decarboxylase 65	~0.3% TSP	Diagnostics and therapeutics	Wang et al. (2008)
ARS2-crEpo-his6	Human erythropoietin fused to ARS2 export sequence w/6xhis tag	100 μg/l culture	Pharmaceutical, protein export	Eichler-Stahlberg et al. (2009)
83K7C	Full-length IgG1 human monoclonal antibody against anthrax protective antigen 83	0.01% dry algal biomass	Therapeutics	Tran et al. (2009)
IgG1	Murine and human antibodies (LC and HC)	Detectable	Therapeutics	Tran et al. (2009)
VP28	White spot syndrome virus protein 28	~10.5% TSP	Vaccine	Surzycki et al. (2009)
CTB-D2	D2 fibronectin-binding domain of Staphylococcus aureus fused with the cholera toxin B subunit	0.7% TSP	Oral vaccine	Dreesen et al. (2010)
10NF3, 14FN3	Domains 10 and 14 of human fibronectin, potential antibody mimics	14FN3: 3% TSP 10FN3: detectable	Therapeutics	Rasala et al. (2010)
M-SAA-Interferon β1	Multiple sclerosis treatment fused to M-SAA	Detectable	Therapeutics	Rasala et al. (2010)
Proinsulin	Blood sugar level-regulating hormone, type I diabetes treatment	Detectable	Therapeutics	Rasala et al. (2010)
VEGF	Human vascular endothelial growth factor isoform 121	2% TSP	Therapeutics	Rasala et al. (2010)
HMGB1	High mobility group protein B1	2.5% TSP	Therapeutics	Rasala et al. (2010)

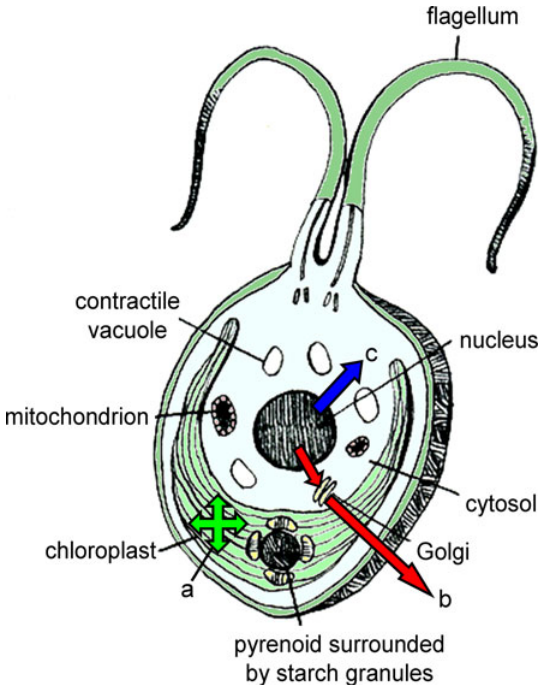


Fig. 1 *Chlamydomonas reinhardtii* as a versatile recombinant protein production platform. Protein expressed from the chloroplast genome is accumulated inside the single large chloroplast (*a*). The reducing environment of the chloroplast allows for proper folding of heavily disulfide bonded proteins, which is not easily accomplished in bacterial production platforms. Protein expressed from the nuclear genome accumulates in the cytosol (*b*), unless it is given an export signal sequence. In this case, it is sent to the endoplasmic reticulum for translocation and processing and then moves to the Golgi apparatus for packaging and export to the extracellular media (*c*)

(Kakinuma et al. 2009), and red alga *Cyanidioschyzon merolae* (Ohnuma et al. 2008). Methods have also been improved for previously transformed alga, such as *Dunaliella salina* (Feng et al. 2009), and cyanobacterial genetics have also been extensively explored, but will not be discussed in this review.

Future challenges

A significant obstacle for algal protein expression systems is the lack of production systems optimized for large-scale growth and harvesting of algae under photoautotrophic conditions. The main limitations in photobioreactor size arise from inhibited gas exchange and light penetration in large cultures, especially at the high cell densities required to keep costs low (Ugwu et al. 2008). An alternative approach, which has been discussed extensively by Chen and Chen (2006), is to grow algae heterotrophically in conventional fermentation bioreactors. This is certainly an economically viable option for high-value products such as therapeutics or enzymes, and indeed is currently used in much of the microalgae products industry (Chen and Chen 2006) with yields reported as high as 83 g dry weight per liter of culture for some species (de Swaaf et al. 2003). Maximum yields for *Chlamydomonas reinhardtii* are around 1.5 g dry biomass per liter when grown in continuous flow on acetate media (Chen and Johns 1996).

The scale and cost requirements for algal biofuels will likely necessitate photoautotrophic open cultivation systems. These systems are being improved, as are processes for efficiently harvesting the algae. For example, new methods have been optimized for harvesting algae using microbes capable of flocculating 90% of the algal mass with no deleterious effect on algal viability (Lee et al. 2009). More efficient methods and systems for large-scale growth are critical if algal-derived biofuels are to become a reality. If these issues can be resolved, algae represent a far superior source of biofuel than terrestrial plants. It is estimated algae can produce up to ten times as much oil per acre than any current terrestrial crop (Cooney et al. 2009). For a detailed recent review on the potential of algal biofuels, see Mata et al. 2010.

Conclusion

Several decades of work in *Chlamydomonas* has elucidated a better understanding of the transcriptional and translational machinery and regulation of the cell, ultimately generating improved methods for transgenic expression of recombinant proteins in algae. However, only in the past few years have these advances been used to successfully express large, fully bioactive, therapeutically relevant proteins at a level sufficient for economically viable large-scale production. Extensive research on optimal transformation constructs and gene optimization have greatly increased yields of recombinant protein, though further work is still needed to address nuclear gene

silencing, plastid auto-attenuation, and to optimize reactor design for large-scale use. Algae have proven their utility and tractability as a production system for therapeutic or industrial proteins and peptides, and algae now seem poised to become the "green" alternative to current mammalian, yeast, or bacterial recombinant protein production systems.

Acknowledgements This work was supported by a grant from the National Institute of Allergy and Infectious Diseases to SPM (AI059614). ES is supported by a National Science Foundation pre-doctoral fellowship.

References

Abe J, Hiwatashi Y, Ito M et al (2008) Expression of exogenous genes under the control of endogenous HSP70 and CAB promoters in the *Closterium peracerosum-strigosum-littorale* complex. Plant Cell Physiol 49:625–632

Barnes D, Franklin S, Schultz J et al (2005) Contribution of 5′- and 3′-untranslated regions of plastid mRNAs to the expression of *Chlamydomonas reinhardtii* chloroplast genes. Mol Genet Genomics 274:625–636

Berthold P, Schmitt R, Mages W (2002) An engineered *Streptomyces hygroscopicus* aph 7″ gene mediates dominant resistance against hygromycin B in *Chlamydomonas reinhardtii*. Protist 153:401–412

Bock R (2007) Plastid biotechnology: prospects for herbicide and insect resistance, metabolic engineering and molecular farming. Curr Opin Biotechnol 18:100–106

Boudreau E, Nickelsen J, Lemaire SD et al (2000) The Nac2 gene of *Chlamydomonas* encodes a chloroplast TPR-like protein involved in psbD mRNA stability. Eur Mol Biol Organ J 19:3366–3376

Casas-Mollano JA, van Dijk K, Eisenhart J et al (2007) SET3p monomethylates histone H3 on lysine 9 and is required for the silencing of tandemly repeated transgenes in *Chlamydomonas*. Nucleic Acids Res 35:939–950

Casas-Mollano JA, Jeong B-r, Xu J et al (2008a) The MUT9p kinase phosphorylates histone H3 threonine 3 and is necessary for heritable epigenetic silencing in *Chlamydomonas*. Proc Natl Acad Sci 105:6486–6491

Casas-Mollano JA, Rohr J, Kim E-J et al (2008b) Diversification of the core RNA interference machinery in *Chlamydomonas reinhardtii* and the role of DCL1 in transposon silencing. Genetics 179:69–81

Chebolu S, Daniell H (2009) Chloroplast-derived vaccine antigens and biopharmaceuticals: expression, folding, assembly and functionality. Plant-Produced Microbial Vaccines, pp 33–54

Chen G-Q, Chen F (2006) Growing phototrophic cells without light. Biotechnol Lett 28:607–616

Chen F, Johns MR (1996) Heterotrophic growth of *Chlamydomonas reinhardtii* on acetate in chemostat culture. Process Biochem 31:601–604

Cooney M, Young G, Nagle N (2009) Extraction of bio-oils from microalgae. Sep Purif Rev 38:291–325

Croft MT, Moulin M, Webb ME et al (2007) Thiamine biosynthesis in algae is regulated by riboswitches. Proc Natl Acad Sci 104:20770–20775

Davoodi-Semiromi A, Samson N, Daniell H (2009) The green vaccine: a global strategy to combat infectious and autoimmune diseases. Hum Vaccin 5:488–493

de Swaaf ME, Pronk JT, Sijtsma L (2003) Fed-batch cultivation of the docosahexacaoic-acid-producing marine alga *Crypthecodinium colunii* on ethanol. Appl Microbiol Biotechnol 61:40–43

Dove A (2002) Uncorking the biomanufacturing bottleneck. Nat Biotechnol 20:777–779

Dreesen IAJ, Charpin-El Hamri G, Fussenegger M (2010) Heat-stable oral alga-based vaccine protects mice from *Staphylococcus aureus* infection. J Biotechnol 145:273–280

Eichler-Stahlberg A, Weisheit W, Ruecker O et al (2009) Strategies to facilitate transgene expression in *Chlamydomonas reinhardtii*. Planta 229:873–883

Fei XW, Deng XD (2007) A novel Fe deficiency-responsive element (FeRE) regulates the expression of atx1 in *Chlamydomonas reinharditii*. Plant Cell Physiol 48:1496–1503

Feng S, Xue L, Liu H et al (2009) Improvement of efficiency of genetic transformation for *Dunaliella salina* by glass beads method. Mol Biol Rep 36:1433–1439

Ferrante P, Catalanotti C, Bonente G et al (2008) An optimized, chemically regulated gene expression system for *Chlamydomonas*. PLoS ONE 3(9):e3200

Fischer N, Rochaix JD (2001) The flanking regions of PsaD drive efficient gene expression in the nucleus of the green alga *Chlamydomonas reinhardtii*. Mol Genet Genomics 265:888–894

Fischer BB, Krieger-Liszkay A, Hideg E et al (2007) Role of singlet oxygen in chloroplast to nucleus retrograde signaling in *Chlamydomonas reinhardtii*. FEBS Lett 581:5555–5560

Fletcher SP, Muto M, Mayfield SP (2007) Optimization of recombinant protein expression in the chloroplasts of green algae. In: Leon R, Galvan A, Fernandez E (eds) Advances in experimental medicine and biology, vol 616: transgenic microalgae as green cell factories. Landes Bioscience and Springer Science Business Media LLC, New York, NY, pp 90–98

Franklin SE, Mayfield SP (2004) Prospects for molecular farming in the green alga *Chlamydomonas reinhardtii*. Curr Opin Plant Biol 7:159–165

Franklin SE, Mayfield SP (2005) Recent developments in the production of human therapeutic proteins in eukaryotic algae. Expert Opin Biol Ther 5:225–235

Franklin S, Ngo B, Efuet E et al (2002) Development of a GFP reporter gene for *Chlamydomonas reinhardtii* chloroplast. Plant J 30:733–744

Fuhrmann M, Oertel W, Hegemann P (1999) A synthetic gene coding for the green fluorescent protein (GFP) is a versatile reporter in *Chlamydomonas reinhardtii*. Plant J 19:353–361

Goldschmidt-Clermont M, Rahire M, Rochaix JD (2008) Redundant cis-acting determinants of 3′ processing and RNA stability in the chloroplast rbcL mRNA of *Chlamydomonas*. Plant J 53:566–577

Gowda DC, Davidson EA (1999) Protein glycosylation in the malaria parasite. Parasitol Today 15:147–152

Griesbeck C, Kobl I, Heitzer M (2006) *Chlamydomonas reinhardtii*. Mol Biotechnol 34:213–223

Hallmann A (2007) Algal transgenics and biotechnology. Transgenic Plant J 1:81–98

Harakuni T, Sugawa H, Komesu A et al (2005) Heteropentameric cholera toxin B subunit chimeric molecules genetically fused to a vaccine antigen induce systemic and mucosal immune responses: a potential new strategy to target recombinant vaccine antigens to mucosal immune systems. Infect Immun 73:5654–5665

He D-M, Qian K-X, Shen G-F et al (2007) Recombination and expression of classical swine fever virus (CSFV) structural protein E2 gene in *Chlamydomonas reinhardtii* chloroplasts. Colloids Surf B Biointerfaces 55:26–30

Heitzer M, Zschoernig B (2007) Construction of modular tandem expression vectors for the green alga *Chlamydomonas reinhardtii* using the Cre/lox-system. Biotechniques 43:324–328

Heitzer M, Eckert A, Fuhrmann M et al (2007) Influence of codon bias on the expression of foreign genes in microalgae. Transgenic Microalgae as Green Cell Factories, pp 46–53

Hirakawa Y, Kofuji R, Ishida K-i (2008) Transient transformation of a chlorarachniophyte alga, *Lotharella amoebiformis* (Chlorarachniophyceae), with *uidA* and *egfp* reporter genes. J Phycol 44:814–820

Kakinuma M, Ikeda M, Coury D et al (2009) Isolation and characterization of the rbcS genes from a sterile mutant of *Ulva pertusa* (Ulvales, Chlorophyta) and transient gene expression using the rbcS gene promoter. Fish Sci 75:1015–1028

Kasai S, Yoshimura S, Ishikura K et al (2003) Effect of coding regions on chloroplast gene expression in *Chlamydomonas reinhardtii*. J Biosci Bioeng 95:276–282

Kathiresan S, Sarada R (2009) Towards genetic improvement of commercially important microalga *Haematococcus pluvialis* for biotech applications. J Appl Phycol 21:553–558

Kathiresan S, Chandrashekar A, Ravishankar GA et al (2009) *Agrobacterium*-mediated transformation in the green alga *Haematococcus pluvialis* (Chlorophyceae, Volvocales). J Phycol 45:642–649

Kato K, Marui T, Kasai S et al (2007) Artificial control of transgene expression in *Chlamydomonas reinhardtii* chloroplast using the lac regulation system from *Escherichia coli*. J Biosci Bioeng 104:207–213

Kiaulehn S, Voytsekh O, Fuhrmann M et al (2007) The presence of UG-repeat sequences in the 3′-UTRs of reporter luciferase mRNAs mediates circadian expression and can determine acrophase in *Chlamydomonas reinhardtii*. J Biol Rhythms 22:275–277

Koop H-U, Herz S, Golds T (2007) The genetic transformation of plastids. In: Bock Ralph (ed) Topics in current genetics, vol 19: cell and molecular biology of plastids. Springer, Berlin/Heidelberg, pp 457–510

Koya V, Moayeri M, Leppla SH et al (2005) Plant-based vaccine: mice immunized with chloroplast-derived anthrax protective antigen survive anthrax lethal toxin challenge. Infect Immunol 73:8266–8274

Kumar SV, Misquitta RW, Reddy VS et al (2004) Genetic transformation of the green alga *Chlamydomonas reinhardtii* by *Agrobacterium tumefaciens*. Plant Sci 166:731–738

Lee AK, Lewis DM, Ashman PJ (2009) Microbial flocculation, a potentially low-cost harvesting technique for marine microalgae for the production of biodiesel. J Appl Phycol 21:559–567

Leon R, Fernandez E (2007) Nuclear transformation of eukaryotic microalgae—historical overview, achievements and problems. In: Leon R, Fernandez E (eds) Transgenic microalgae as green cell factories, vol 616. Landes Bioscience and Springer Science Business Media LLC, New York, pp 1–11

Lerche K, Hallmann A (2009) Stable nuclear transformation of *Gonium pectorale*. BMC Biotechnol 9:64

Li F, Holloway SP, Lee J et al (2002) Nuclear genes that promote splicing of group I introns in the chloroplast 23S rRNA and *psbA* genes in *Chlamydomonas reinhardtii*. Plant J 32:467–480

Loiselay C, Gumpel NJ, Girard-Bascou J et al (2008) Molecular identification and function of cis- and trans-acting determinants for petA transcript stability in *Chlamydomonas reinhardtii* chloroplasts. Mol Cell Biol 28:5529–5542

Lumbreras V, Stevens DR, Purton S (1998) Efficient foreign gene expression in *Chlamydomonas reinhardtii* mediated by an endogenous intron. Plant J 14:441–447

Manuell AL, Beligni MV, Elder JH et al (2007) Robust expression of a bioactive mammalian protein in *Chlamydomonas* chloroplast. Plant Biotechnol J 5:402–412

Marin-Navarro J, Manuell AL, Wu J et al (2007) Chloroplast translation regulation. Photosynth Res 94:359–374

Mata TM, Martins AA, Caetano NS (2010) Microalgae for biodiesel production and other applications: a review. Renew Sustain Energy Rev 14:217–232

Maul JE, Lilly JW, Cui L et al (2002) The *Chlamydomonas reinhardtii* plastid chromosome: islands of genes in a sea of repeats. Plant Cell 14:2659–2679

Mayfield SP, Franklin SE (2005) Expression of human antibodies in eukaryotic micro-algae. Vaccine 23:1828–1832

Mayfield S, Schultz J (2004) Development of a luciferase reporter gene, luxCt, for *Chlamydomonas reinhardtii* chloroplast. Plant J 37:449–458

Mayfield SP, Franklin SE, Lerner RA (2003) Expression and assembly of a fully active antibody in algae. Proc Natl Acad Sci 100:438–442

Merchant SS, Prochnik SE, Vallon O et al (2007) The *Chlamydomonas* genome reveals the evolution of key animal and plant functions. Science 318:245–250

Molina V, Shoenfeld Y (2005) Infection, vaccines and other environmental triggers of autoimmunity. Autoimmunity 38:235–245

Muto M, Henry Ryan E, Mayfield Stephen P (2009) Accumulation and processing of a recombinant protein designed as a cleavable fusion to the endogenous Rubisco LSU protein in *Chlamydomonas* chloroplast. BMC Biotechnol 9:26

Nakamura Y, Gojobori T, Ikemura T (2000) Codon usage tabulated from international dna sequence databases: status for the year 2000. Nucleic Acids Res 28:292

Neupert J, Karcher D, Bock R (2009) Generation of *Chlamydomonas* strains that efficiently express nuclear transgenes. Plant J 57:1140–1150

Oey M, Lohse M, Kreikemeyer B et al (2009) Exhaustion of the chloroplast protein synthesis capacity by massive expression of a highly stable protein antibiotic. Plant J 57:436–445

Ohnuma M, Yokoyama T, Inouye T et al (2008) Polyethylene glycol (PEG)-mediated transient gene expression in a red alga, *Cyanidioschyzon merolae* 10D. Plant Cell Physiol 49:117–120

Ohresser M, Matagne RF, Loppes R (1997) Expression of the arylsulphatase reporter gene under the control of the nit1 promoter in *Chlamydomonas reinhardtii*. Curr Genet 31:264–271

Plaza M, Herrero M, Cifuentes A et al (2009) Innovative natural functional ingredients from microalgae. J Agric Food Chem 57:7159–7170

Plaza M, Santoyo S, Jaime L et al (2010) Screening for bioactive compounds from algae. J Pharm Biomed Anal 51:450–455

Popescu CE, Lee RW (2007) Mitochondrial genome sequence evolution in *Chlamydomonas*. Genetics 175:819–826

Powdrill TF, Johnston JM (1991) Immunologic priming with recombinant hepatitis A virus capsid proteins produced in *Escherichia coli*. J Virol 65:2686–2690

Puigbo P, Guzman E, Romeu A et al (2007) OPTIMIZER: a web server for optimizing the codon usage of DNA sequences. Nucleic Acids Res 35:W126–W131

Puigbo P, Bravo I, Garcia-Vallve S (2008) E-CAI: a novel server to estimate an expected value of codon adaptation index (eCAI). BMC Bioinformatics 9:65

Rasala BA, Muto M, Lee PA et al (2010) Production of therapeutic proteins in algae, analysis of expression of seven human proteins in the chloroplast of *Chlamydomonas reinhardtii*. Plant Biotechnol J 8:1–15

Raynaud C, Loiselay C, Wostrikoff K et al (2007) Evidence for regulatory function of nucleus-encoded factors on mRNA stabilization and translation in the chloroplast. Proc Natl Acad Sci 104:9093–9098

Rosenberg JN, Oyler GA, Wilkinson L et al (2008) A green light for engineered algae: redirecting metabolism to fuel a biotechnology revolution. Curr Opin Biotechnol 19:430–436

Rymarquis LA, Higgs DC, Stern DB (2006) Nuclear suppressors define three factors that participate in both 5′ and 3′ end processing of mRNAs in *Chlamydomonas* chloroplasts. Plant J 46:448–461

Schroda M, Blcker D et al (2000) The HSP70A promoter as a tool for the improved expression of transgenes in *Chlamydomonas*. Plant J 21:121–131

Schroda M, Beck CF, Vallon O (2002) Sequence elements within an HSP70 promoter counteract transcriptional transgene silencing in *Chlamydomonas*. Plant J 31:445–455

Schwarz C, Elles I, Kortmann J et al (2007) Synthesis of the D2 protein of photosystem II in *Chlamydomonas* is controlled by a high molecular mass complex containing the RNA stabilization factor Nac2 and the translational activator RBP40. Plant Cell 19:3627–3639

Shao N, Bock R (2008) A codon-optimized luciferase from *Gaussia princeps* facilitates the in vivo monitoring of gene expression in the model alga *Chlamydomonas reinhardtii*. Curr Genet 53:381–388

Somanchi A, Barnes D, Mayfield SP (2005) A nuclear gene of *Chlamydomonas reinhardtii*, *Tba1*, encodes a putative oxidoreductase required for translation of the chloroplast *psbA* mRNA. Plant J 42:341–352

Suess B (2005) Engineered riboswitches control gene expression by small molecules. Biochem Soc Trans 33:474–476

Sun M, Qian KX, Su N et al (2003) Foot-and-mouth disease virus VP1 protein fused with cholera toxin B subunit expressed in *Chlamydomonas reinhardtii* chloroplast. Biotechnol Lett 25:1087–1092

Surzycki R, Cournac L, Peltier G et al (2007) Potential for hydrogen production with inducible chloroplast gene expression in *Chlamydomonas*. Proc Natl Acad Sci 104:17548–17553

Surzycki R, Greenham K, Kitayama K et al (2009) Factors effecting expression of vaccines in microalgae. Biologicals 37:133–138

Tran M, Zhou B, Pettersson PL et al (2009) Synthesis and assembly of a full-length human monoclonal antibody in algal chloroplasts. Biotechnol Bioeng 104:663–673

Tregoning JS, Nixon P, Kuroda H et al (2003) Expression of tetanus toxin Fragment C in tobacco chloroplasts. Nucleic Acids Res 31:1174–1179

Ugwu CU, Aoyagi H, Uchiyama H (2008) Photobioreactors for mass cultivation of algae. Bioresour Technol 99:4021–4028

von Gromoff ED, Alawady A, Meinecke L et al (2008) Heme, a plastid-derived regulator of nuclear gene expression in *Chlamydomonas*. Plant Cell 20:552–567

Wang XF, Brandsma M, Tremblay R et al (2008) A novel expression platform for the production of diabetes-associated autoantigen human glutamic acid decarboxylase (hGAD65). BMC Biotechnol 8:87

Winkler WC, Breaker RR (2005) Regulation of bacterial gene expression by riboswitches. Annu Rev Microbiol 59:487–517

Wu JX, Hu ZL, Wang CG et al (2008) Efficient expression of green fluorescent protein (GFP) mediated by a chimeric promoter in *Chlamydomonas reinhardtii*. Chin J Oceanol Limnol 26:242–247

Xia X (2007) An improved implementation of codon adaptation index. Evol Bioinform Online 3:53–58

Yang Z, Li y, Chen F et al (2006) Expression of human soluble TRAIL in *Chlamydomonas reinhardtii* chloroplast. Chin Sci Bull 51:1703–1709

Zhang Y-K, Shen G-F, Ru B-G (2006) Survival of human metallothionein-2 transplastomic *Chlamydomonas reinhardtii* to ultraviolet B exposure. Acta Biochim Biophys Sin 38:187–193

Acknowledgements

Chapter 2, in full, is a reprint of the material as it appears in *Biotechnology Letters* 2010. The dissertation author was the primary investigator and author of this paper.

CHAPTER 3:

ALGAE-BASED ORAL RECOMBINANT VACCINES

frontiers in
MICROBIOLOGY

MINI REVIEW ARTICLE
published: 17 February 2014
doi: 10.3389/fmicb.2014.00060

Algae-based oral recombinant vaccines

Elizabeth A. Specht and Stephen P. Mayfield*

California Center for Algae Biotechnology, University of California at San Diego, La Jolla, CA, USA

Edited by:
Germán Leandro Rosano, Instituto de Biología Molecular y Celular de Rosario, Argentina

Reviewed by:
Edward Rybicki, University of Cape Town, South Africa
Ruth Elena Soria-Guerra, Universidad Autónoma de San Luis Potosí, Mexico
Pal Maliga, Rutgers University, USA

***Correspondence:**
Stephen P. Mayfield, California Center for Algae Biotechnology, University of California at San Diego, Bonner Hall 2150, MC 0368, 9500 Gilman Drive, La Jolla, CA 92093, USA
e-mail: smayfield@ucsd.edu

Recombinant subunit vaccines are some of the safest and most effective vaccines available, but their high cost and the requirement of advanced medical infrastructure for administration make them impractical for many developing world diseases. Plant-based vaccines have shifted that paradigm by paving the way for recombinant vaccine production at agricultural scale using an edible host. However, enthusiasm for "molecular pharming" in food crops has waned in the last decade due to difficulty in developing transgenic crop plants and concerns of contaminating the food supply. Microalgae could be poised to become the next candidate in recombinant subunit vaccine production, as they present several advantages over terrestrial crop plant-based platforms including scalable and contained growth, rapid transformation, easily obtained stable cell lines, and consistent transgene expression levels. Algae have been shown to accumulate and properly fold several vaccine antigens, and efforts are underway to create recombinant algal fusion proteins that can enhance antigenicity for effective orally delivered vaccines. These approaches have the potential to revolutionize the way subunit vaccines are made and delivered – from costly parenteral administration of purified protein, to an inexpensive oral algae tablet with effective mucosal and systemic immune reactivity.

Keywords: oral vaccines, recombinant subunit vaccines, microalgae, plant-produced vaccines, algal engineering

INTRODUCTION

Infectious diseases directly account for nearly 25% of deaths worldwide, and are a predominant cause of morbidity and mortality in the developing world (Fauci et al., 2005). Even for diseases for which vaccines exist, limited access – due to financial as well as infrastructural or medical personnel limitations – is a major contributor to this high infectious disease burden. Many developing world diseases do not yet have vaccines, in part because traditional vaccine production costs present a significant investment hurdle, considering the financial capacity of the intended consumers. Both cost and ease of administration are challenges that must be tackled to address this undue burden on global health and productivity.

Oral vaccination has many distinct advantages over parenteral administration, but has proven difficult to achieve thus far, reflected by the scarcity of licensed oral vaccines. Perhaps the most significant benefit of oral vaccination is the ability to elicit both mucosal and systemic immunity. As most human pathogens enter via mucosal surfaces – either nasally, orally, or by sexual transmission – mucosal immunity can serve as a first line of defense to prevent infection before it reaches the bloodstream (Mason and Herbst-Kralovetz, 2012). Oral vaccines also obviate the need for trained medical personnel to administer them and reduce the risks of infection associated with needles. They also have higher compliance from patients, owing to the lack of fear and resistance associated with injections. Both of these latter aspects are important considerations for successful vaccination campaign coverage in remote or resource-limited settings.

Plant-produced vaccines have two critical advantages: much lower cost than traditional recombinant vaccine platforms, and improved safety because of insusceptibility to mammalian pathogen contamination. The batch costs of plant-produced vaccines may be as much as a thousand times less than traditional animal cell culture or even bacterial or yeast cell culture, though it has been noted that this will not translate directly to per-dose cost because downstream sales, packaging, and distribution costs are similar regardless of production method (Rybicki, 2009). The current status of plant-produced vaccines in pre-clinical and early phase human clinical trials has been extensively reviewed (Lossl and Waheed, 2011; Mason and Herbst-Kralovetz, 2012; Rosales-Mendoza et al., 2012a,b; Guan et al., 2013; Jacob et al., 2013); despite positive preliminary data, none have made it through to licensing. The only licensed plant-produced vaccine is a veterinary injectable vaccine against Newcastle disease virus in poultry, made from purified antigen expressed in cultured tobacco cells. Dow AgroSciences received Food and Drug Administration (FDA) approval for the vaccine in 2006, but only as a demonstration that plant-produced vaccines can meet the stringent regulatory requirements for approval; it is not currently for sale (Rybicki, 2009).

Plant cells are of particular interest for oral vaccines because their rigid cell walls provide exceptional antigen protection through the stomach into the intestines, where they can access the gut-associated lymphoid tissue (Kwon et al., 2013). Expression within chloroplasts or other storage organelles may also provide additional protection (Khan et al., 2012). While vaccine antigens have been transformed into many edible species including lettuce, tomato, potato, and tobacco, expression in stable transformed crop plants has suffered from low yields, typically less than 1% of total soluble protein (TSP; Lossl and Waheed, 2011). Yields have been increased by transient expression using recombinant viral vectors or *Agrobacterium* infection, but this expression is typically unstable (Rybicki, 2009). Even using these strategies, the most consistently

high-yielding host species is tobacco, which is inedible and therefore would require purification prior to vaccine administration (Lossl and Waheed, 2011).

ALGAE AS A RECOMBINANT PROTEIN PRODUCTION PLATFORM

Green microalgae have proven to be highly useful protein production platforms for a variety of industrial and therapeutic applications, particularly for complex or heavily disulfide-bonded proteins. The chloroplast provides a unique enclosed compartment that facilitates folding (Chebolu and Daniell, 2009), and transgene products have been shown to accumulate to high levels in the algal chloroplast – as high as 10% of TSP (Manuell et al., 2007; Surzycki et al., 2009). Unlike prokaryotes, chloroplasts of algae contain much of the same sophisticated cellular folding machinery as other eukaryotic organisms like yeast. While the algal nuclear genome can also be transformed, to date most transgene expression has been from the chloroplast genome due to reduced gene silencing and higher protein accumulation.

The green alga model organism *Chlamydomonas reinhardtii* has been used to produce a number of human and animal therapeutically relevant proteins, including full-length human antibodies (Tran et al., 2009), signaling molecules such as vascular endothelial growth factor (Rasala et al., 2010), and structural proteins like fibronectin (Rasala et al., 2010). Though expression levels are highly variable by gene, improvements in codon optimization (Franklin et al., 2002; Surzycki et al., 2009) and characterization of ideal gene regulatory elements (Rasala et al., 2011; Specht and Mayfield, 2013) continue to increase levels of transgene expression. *C. reinhardtii*'s success and future potential as a therapeutic protein production platform has been recently reviewed (Rasala and Mayfield, 2011).

ADVANTAGES OF AN ALGAL VACCINE PRODUCTION HOST

Unicellular green algae possess all the positive attributes of plant systems, plus several unique advantages over terrestrial plants as vaccine production hosts. Algal biomass accumulation is extremely rapid, and the entirety of the biomass can be utilized for vaccine production, unlike plants that expend energy producing supporting tissues that do not contain the vaccine antigen or cannot be harvested easily. Algae are also not restricted by growing season or local soil fertility, and concerns of cross-contamination of nearby food crops are non-existent. Enclosed bioreactors can be used for higher biomass yields and to reduce concerns of environmental escape (Franconi et al., 2010), and media can be recycled to minimize water and nutrient loss. The 2002 discovery of transgenic viral capsid protein-expressing maize in food harvests of nearby corn and soybean crops effectively halted efforts to produce vaccines in edible crop plants, making a food crop-based oral vaccine highly unlikely (Rybicki, 2009). Green algae such as *C. reinhardtii* are generally recognized as safe (GRAS) by the FDA, resurrecting hope that unprocessed edible vaccines can be produced in a photosynthetic organism.

Crop plants can contain hundreds of chloroplasts per cell, and each chloroplast harbors dozens of copies of its plastid genome. In contrast, *C. reinhardtii* contains a single chloroplast that occupies about half of the volume of the cell (Franklin and Mayfield, 2005), making stable homoplasmic transformed lines much easier to obtain (a few weeks versus several months) and allowing for increased yields of plastid-expressed vaccine antigens, which account for nearly all antigens expressed to date in algae. This genomic stability, combined with the ability to tightly regulate growth conditions inside contained bioreactors, allows for more consistent expression levels than terrestrial plants, which can vary by several-fold.

Finally, algae can be easily preserved by lyophilization, and two studies of algal-produced vaccine antigens have verified that dried algae stored at room temperature for 6 months (Gregory et al., 2013) or even 20 months (Dreesen et al., 2010) exhibit nearly equivalent antigen effectiveness as freshly harvested algae, though storage at 37° did begin to cause a loss of activity over time (Gregory et al., 2013). The algal cell wall appears sufficient to withstand harsh conditions within the stomach, as very little antigen degradation was observed after whole cells were incubated with pepsin at pH 1.7 (Dreesen et al., 2010). These observations indicate that algae are an ideal host for vaccine transport without cold-chain supply, and that the cells provide adequate protection for antigens en route to the intestinal mucosal lymph tissue, obviating the additional expense associated with encapsulation.

ALGAL VACCINE PROGRESS

The first reported algal-synthesized vaccine antigen was a chimeric molecule comprising the foot-and-mouth disease virus structural protein VP1 and the beta subunit of cholera toxin (CTB), a known mucosal adjuvant (Sun et al., 2003). This antigen had been previously expressed in plants and had demonstrated oral immunity in mice (Wigdorovitz et al., 1999), but advancement of trials was hindered by low expression levels. In *C. reinhardtii*, 3–4% TSP was reported, but higher yields may be possible because the strains examined were not completely homoplasmic (Sun et al., 2003).

The next report of an algal-produced vaccine antigen showed the first *in vivo* data for efficacy conferring immunity. The classical swine fever virus (CSFV) surface protein E2 was expressed from the *C. reinhardtii* chloroplast genome, and total protein extracts were administered subcutaneously with Freud's adjuvant or orally by gavage with no adjuvant. Subcutaneous immunization reportedly induced a significant immune response, but no data for this result was shown. No systemic or mucosal immune response was detected after the oral immunization, and it was suggested that a mucosal adjuvant may be necessary for oral administration to be effective (He et al., 2007).

Wang et al. (2008) expressed the human glutamic acid decarboxylase, a known Type 1 diabetes autoimmune antigen, which reacted with sera from non-obese diabetic mice. Surprisingly, detectable expression was achieved using a non-codon-optimized gene. A more thorough investigation of the factors affecting vaccine antigen expression in algae found that indeed codon optimization is critical for high yield. It has also been noted that yield is highly variable among individual transformants despite the fact that chloroplast transformation proceeds by homologous recombination, eliminating positional effects within the genome (Surzycki et al., 2009).

Oral immunization was finally shown to be effective when the antigen of interest was fused to the B subunit of CTB, which forms a pentameric structure and binds the GM1 ganglioside for internalization into intestinal cells. After feeding freeze-dried algae repeatedly to mice, fecal IgA and systemic IgG antibody titers reached similarly high levels for both the intended *Staphylococcus aureus* antigen and CTB. Significantly, within a week of finishing the 5-week oral vaccination, 80% of immunized mice survived a lethal challenge with *S. aureus* that killed all control mice within 48 h (Dreesen et al., 2010).

Two studies earlier this year reported relatively low yields of two additional algal-produced antigens, but they are still promising compared to previous literature using alternative systems. A human papillomavirus E7 protein, while only accumulated to 0.12% TSP, expressed similar to or better than in other plant systems and did not require fusion to a stabilizing protein to achieve consistent expression. Furthermore, the algal chloroplast-produced E7 was soluble, whereas the plant-produced E7 was found predominantly in the insoluble fraction using multiple solubilization buffers. While the antibody titer elicited by affinity purified protein was much higher, a crude algal extract was shown to be equally effective at preventing tumor development and promoting mouse survival (Demurtas et al., 2013). A chimeric antigen intended to prevent hypertension, consisting of a fusion between angiotensin and a Hepatitis B antigen as a carrier, was the first algal vaccine to be expressed from the nuclear genome without chloroplast targeting. While it only accumulated to 0.05% TSP, it was detectable by Western blot from algal TSP extracts (Soria-Guerra et al., 2014).

Since 2010, several studies have shown that malarial transmission-blocking vaccines can be produced in *C. reinhardtii*. Transmission-blocking vaccines target surface proteins that appear on the sexual and gamete stages of *Plasmodium*, the causative pathogen of malaria. There is some evidence that these vaccines may provide partial protection to individuals, but the main benefit of vaccination with a transmission-blocking vaccine is derived from herd immunity preventing the spread of the disease. Therefore, it is especially critical that transmission-blocking vaccines can be delivered easily and at extremely low cost, to reach threshold coverage of the huge populations living in malaria-endemic regions. One difficulty of producing these *Plasmodium* surface proteins is that they contain multiple EGF-like domains that are heavily disulfide-bonded, rendering them difficult to fold and therefore difficult to accumulate to high levels without forming insoluble aggregates (Gregory et al., 2012). Interestingly, *Plasmodia* appear to not glycosylate their proteins (Gowda and Davidson, 1999), making algal chloroplasts suitable hosts as the chloroplast also does not contain glycosylation machinery.

A total of six algae-produced malarial antigens or fragments thereof – *Pfs*25, *Pfs*28, *Pfs*48/45, *Pf*MSP1, *Pb*MSP1, and *Pb*AMA1 – have been shown to fold properly and exhibit antibody recognition akin to that of the native *Plasmodium* surface proteins (Dauvillée et al., 2010; Gregory et al., 2012; Jones et al., 2013). Algal chloroplast-produced *Pfs*25 was able to completely prevent malaria transmission, indicated by a total absence of *Plasmodium* oocysts in mosquito midguts after feeding on immunized mouse sera. Furthermore, feeding lyophilized algae expressing *Pfs*25 fused to CTB elicited a mucosal response to both antigens (Gregory et al., 2013). However, systemic IgG response was only observed for the CTB. This is in contrast with the *S. aureus* D2 protein fused to CTB, where systemic immunity was elicited for both domains (Dreesen et al., 2010), suggesting that either the furin protease cleavable linker between the *Pfs*25 and CTB domains prevented *Pfs*25 from being presented to the systemic immune system, or perhaps that *Pfs*25 is inherently less immunogenic. In a different strategy, truncated versions of the malarial proteins AMA1 and MSP1 were fused to the major protein constituent of the chloroplast starch granules, the granule-bound starch synthase (GBSS). Though they were expressed from the nuclear genome, reasonable accumulation was achieved because the proteins were targeted to and sequestered within the chloroplast starch granules. Both oral and injected vaccination using purified starch from these strains reduced parasite load and prolonged mice survival after challenge with *Plasmodium berghei*; in the case of an injected vaccine consisting of both antigens, 30% of mice survived the otherwise-lethal infection (Dauvillée et al., 2010).

All vaccines produced in algae to date are summarized in **Table 1**, along with reported yields and significant pre-clinical findings. Most work thus far has been performed in the green alga model organism *C. reinhardtii*, though one of the earliest reports of an algal-produced hepatitis B antigen was in the marine alga *Dunaliella salina* (Geng et al., 2003) and hepatitis B antigen has also been produced in the diatom *Phaeodactylum tricornutum* (Hempel et al., 2011). In recent years the algal genetic toolkit has been expanded to other algal species, including other green algae, diatoms, and cyanobacteria (Ducat et al., 2011; Georgianna and Mayfield, 2012; Qin et al., 2012), with a goal of broad host range compatibility. Already, over 20 species of algae – including dinoflagellates, red algae, and diatoms – have been transformed, and a suite of promoters and selectable markers have been characterized for many species (see Gong et al., 2011, for a comprehensive review). While the first generation of algal vaccines has been predominantly pioneered in *Chlamydomonas,* these advances can readily be applied to alternative algal species that may be more suitable for large-scale vaccine production.

FUTURE POTENTIAL FOR ALGAL-BASED ORAL RECOMBINANT VACCINES

From the research available to date, it is clear that algae can produce complex vaccine antigens, and that *Chlamydomonas*-produced antigens can elicit immunogenic responses that are appropriate for their intended roles as vaccines. It is also clear that identifying alternative mucosal adjuvants to complement these antigens is critical, whether for co-administration with algal-produced antigens or for incorporation into chimeric fusion proteins. It has been suggested that antigenic fusions with CTB, one of the preferred adjuvants, may interfere with the CTB subunit's ability to form the pentameric structure essential for strong GM1 ganglioside binding (Sun et al., 2003). Many alternatives to CTB are under investigation for oral vaccination in other production platforms, including CpG-containing oligodeoxynucleotides, saponins, and subunits from heat-labile enterotoxin and ricin toxin (Pelosi et al., 2012).

Table 1 | Summary of algal-produced vaccines and significant findings.

Antigen(s)	Host/integration site	Yield	Significant findings	Citation
Foot-and-mouth disease virus (FMDV) structural protein VP1 fused to cholera toxin B subunit (CTB)	*C. reinhardtii* chloroplast genome	3 to 4% TSP	Protein accumulation was higher than reported in previous plant studies and detectable by Western and ELISA. Binding to GM1 ganglioside was weak but statistically significant. Strains were not completely homoplasmic, so higher yields may be possible.	Sun et al. (2003)
Hepatitis B surface antigen (HBsAg)	*D. salina* nuclear genome	Up to 3ng/mg soluble protein (0.0003% TSP)	Expression level was quantified by ELISA and reported to be "high" but the results reflect very low expression.	Geng et al. (2003)
Classical swine fever virus (CSFV) structural protein E2	*C. reinhardtii* chloroplast genome	1.5 to 2% TSP	Subcutaneous immunization of algal extracts induced systemic response, but oral immunization did not result in either systemic or mucosal response.	He et al. (2007)
Human glutamic acid decarboxylase 65 (hGAD65)	*C. reinhardtii* chloroplast genome	0.25 to 0.3% TSP	Purified hGAD65 reacted with sera from diabetic mice, showing antigenicity. Higher yields may be obtained by codon-optimizing the human gene, which does not mimic the chloroplast AT content.	Wang et al. (2008)
White spot syndrome virus VP28 protein	*C. reinhardtii* chloroplast genome	Variable, ranging from 0.2 to 20.9% total cellular protein (0.1 to 10.5% TSP)	The authors report high variability in expression levels, despite same-site integration by homologous recombination in the plastid genome. Stronger resistance to antibiotic selection might be useful for identifying high-expressing strains.	Surzycki et al. (2009)
S. aureus D2 fibronectin-binding domain fused to cholera toxin B subunit	*C. reinhardtii* chloroplast genome	Up to 0.7% TSP	Within whole cells, the antigen can withstand low pH and pepsin treatment, and is stable at room temperature for 20 months. Mice fed with whole algae showed mucosal IgA and systemic IgG responses to both CTB and D2, and 80% survived lethal *S. aureus* challenge.	Dreesen et al. (2010)
P. falciparum MSP1, *P. berghei* MSP1, and *P. berghei* AMA1, C-terminal domains only, fused to granule-bound starch synthase (GBSS)	*C. reinhardtii* nuclear genome, targeted to the chloroplast starch granules	0.2 to 1.0 μg of protein per mg of purified starch	The *P. falciparum* gene expressed significantly better than either *P. berghei* gene. A single dose was as effective as three doses in reducing parasite load after *P. berghei* challenge in vaccinated mice. Oral immunization with purified starch protected equally well as parenteral administration until day 40.	Dauvillée et al. (2010)
Hepatitis B surface antigen (HBsAg) fused to GFP or ER retention signal	*P. tricornutum* nuclear genome	0.7% TSP	By inhibitory ELISA, algal-produced HBsAg binds antibodies slightly stronger than commercially available HBsAg produced in yeast.	Hempel et al. (2011)

(Continued)

Table 1 | Continued

Antigen(s)	Host/integration site	Yield	Significant findings	Citation
P. falciparum surface proteins *Pfs*25 and *Pfs*28	*C. reinhardtii* chloroplast genome	Not quantified; visible by Western and Coomassie after affinity purification	Conformation-specific antibodies and CD spectroscopy confirmed proper folding of heavily disulfide-bonded antigens. Sera from *Pfs*25-inoculated mice completely blocked replication within mosquitoes, and therefore malaria transmission.	Gregory et al. (2012)
P. falciparum surface protein *Pfs*25 fused to cholera toxin B subunit	*C. reinhardtii* chloroplast genome	Up to 0.09% TSP	Mice fed lyophilized algae showed mucosal IgA response to both CTB and *Pfs*25, but only systemic IgG response to CTB. A furin protease site separated the two domains.	Gregory et al. (2013)
P. falciparum surface protein *Pfs*48/45 C-terminal domain	*C. reinhardtii* chloroplast genome	Not quantified; visible by Western blot after affinity purification	The antigenic C-terminal domain was recognized by Western blot and ELISA with conformation-specific antibodies. It appeared to accumulate in the insoluble fraction.	Jones et al. (2013)
Human Papillomavirus type 16 E7 protein, attenuated mutant (E7GGG)	*C. reinhardtii* chloroplast genome	Up to 0.12% TSP. His-tagged version obtained only 0.02% TSP.	In contrast to higher plant expression, *C. reinhardtii* obtained much higher yields and most of it was soluble. The crude algal extracts elicited lower IgG response than purified protein, but both preparations allowed 70% survival after tumor challenge.	Demurtas et al. (2013)
Angiotensin II fused to Hepatitis B virus capsid antigen (HBcAg)	*C. reinhardtii* nuclear genome	Up to 0.05% TSP	The chimeric antigen was visible by Western blot and quantitated by ELISA, but no *in vivo* or immunological activity assays were performed.	Soria-Guerra et al. (2014)

Future work should empirically explore many combinations of antigens, mucosal adjuvants, and even testing multiple linkers and potential translocation domains. As has been noted previously, expression, uptake, and antigenicity are all difficult to predict in the context of plant-produced oral vaccine antigens (Rybicki, 2009), so a high-throughput system like algae is extremely valuable for rapidly testing many versions of potential chimeric vaccine molecules. Furthermore, many antigens will require proper post-translational modifications such as glycosylation to be recognized properly; more work needs to be done to increase expression levels from the nuclear genome, as glycosylation does not occur in the chloroplast.

It has been suggested that the first licensed plant-produced human vaccines likely will not be the first ones tested in humans, many of which targeted pathogens like Hepatitis B for which a relatively inexpensive vaccine already exists (Rybicki, 2009). Stepping stones along the way to human vaccines may include reagents for cheaper diagnostics and development of veterinary vaccines. Several human studies with plant-made vaccines have also indicated a role for oral boosting of an existing immune response conferred by traditional injectable vaccines (Mason and Herbst-Kralovetz, 2012). An algal-produced human vaccine production platform will likely come to fruition as an alternative for very expensive vaccines like HPV, or for novel vaccines against diseases for which no alternative currently exists (Martinez et al., 2012). The cost and logistical considerations of storage, delivery, and administration in resource-limited settings indicate that plant or algal production may be the only feasible option for large-scale inexpensive vaccination, and thus this avenue deserves increased attention from research funding agencies and investment from the pharmaceutical industry as well.

AUTHOR CONTRIBUTIONS

Elizabeth A. Specht and Stephen P. Mayfield wrote and revised the manuscript. Elizabeth A. Specht developed **Table 1**.

ACKNOWLEDGMENTS

This work was funded by a Department of Energy, Consortium for Algal Biofuels Commercialization grant, DE-EE0003373; and by the California Energy Commission, California Initiative for Sustainable Large Molecule Fuels, 500–10–039. Elizabeth A. Specht was supported by a National Science Foundation graduate research fellowship. We thank Prema Karunanithi for her careful proofreading of the manuscript.

REFERENCES

Chebolu, S., and Daniell, H. (2009). "Chloroplast-derived vaccine antigens and biopharmaceuticals: expression, folding, assembly and functionality," in *Plant-Produced Microbial Vaccines*, ed. A.V. Karasev. (Berlin: Springer-Verlag Berlin), 33–54. doi: 10.1007/978-3-540-70868-1_3

Dauvillée, D., Delhaye, S., Gruyer, S., Slomianny, C., Moretz, S. E., D'Hulst, C., et al. (2010). Engineering the chloroplast targeted malarial

vaccine antigens in *Chlamydomonas* starch granules. *PLoS ONE* 5:e15424. doi: 10.1371/journal.pone.0015424

Demurtas, O. C., Massa, S., Ferrante, P., Venuti, A., Franconi, R., and Giuliano, G. (2013). A *Chlamydomonas*-derived human papillomavirus 16 E7 vaccine induces specific tumor protection. *PLoS ONE* 8:e61473. doi: 10.1371/journal.pone.0061473

Dreesen, I. A., Charpin-El Hamri, G., and Fussenegger, M. (2010). Heat-stable oral alga-based vaccine protects mice from *Staphylococcus aureus* infection. *J. Biotechnol.* 145, 273–280. doi: 10.1016/j.jbiotec.2009.12.006

Ducat, D. C., Way, J. C., and Silver, P. A. (2011). Engineering cyanobacteria to generate high-value products. *Trends Biotechnol.* 29, 95–103. doi: 10.1016/j.tibtech.2010.12.003

Fauci, A. S., Touchette, N. A., and Folkers, G. K. (2005). Emerging infectious diseases: a 10-year perspective from the national institute of allergy and infectious diseases. *Emerg. Infect. Dis.* 11, 519–525. doi: 10.3201/eid1104.041167

Franconi, R., Demurtas, O. C., and Massa, S. (2010). Plant-derived vaccines and other therapeutics produced in contained systems. *Expert Rev. Vaccines* 9, 877–892. doi: 10.1586/erv.10.91

Franklin, S. E., and Mayfield, S. P. (2005). Recent developments in the production of human therapeutic proteins in eukaryotic algae. *Expert Opin. Biol. Ther.* 5, 225–235. doi: 10.1517/14712598.5.2.225

Franklin, S., Ngo, B., Efuet, E., and Mayfield, S. P. (2002). Development of a GFP reporter gene for *Chlamydomonas reinhardtii* chloroplast. *Plant J.Cell Mol. Biol.* 30, 733–744. doi: 10.1046/j.1365-313X.2002.01319.x

Geng, D. G., Wang, Y. Q., Wang, P., Li, W. B., and Sun, Y. R. (2003). Stable expression of hepatitis B surface antigen gene in *Dunaliella salina* (Chlorophyta). *J. Appl. Phycol.* 15, 451–456. doi: 10.1023/B:JAPH.0000004298.89183.e5

Georgianna, D. R., and Mayfield, S. P. (2012). Exploiting diversity and synthetic biology for the production of algal biofuels. *Nature* 488, 329–335. doi: 10.1038/nature11479

Gong, Y., Hu, H., Gao, Y., Xu, X., and Gao, H. (2011). Microalgae as platforms for production of recombinant proteins and valuable compounds: progress and prospects. *J. Ind. Microbiol. Biotechnol.* 38, 1879–1890. doi: 10.1007/s10295-011-1032-6

Gowda, D. C., and Davidson, E. A. (1999). Protein glycosylation in the malaria parasite. *Parasitol. Today* 15, 147–152. doi: 10.1016/S0169-4758(99)01412-X

Gregory, J. A., Li, F., Tomosada, L. M., Cox, C. J., Topol, A. B., Vinetz, J. M., et al. (2012). Algae-produced Pfs25 elicits antibodies that inhibit malaria transmission. *PLoS ONE* 7:e37179. doi: 10.1371/journal.pone.0037179

Gregory, J. A., Topol, A. B., Doerner, D. Z., and Mayfield, S. (2013). Alga-produced cholera toxin-pfs25 fusion proteins as oral vaccines. *Appl. Environ. Microbiol.* 79, 3917–3925. doi: 10.1128/AEM.00714-13

Guan, Z. J., Guo, B., Huo, Y. L., Guan, Z. P., Dai, J. K., and Wei, Y. H. (2013). Recent advances and safety issues of transgenic plant-derived vaccines. *Appl. Microbiol. Biotechnol.* 97, 2817–2840. doi: 10.1007/s00253-012-4566-2

He, D. M., Qian, K. X., Shen, G. F., Zhang, Z. F., Li, Y. N., Su, Z. L., et al. (2007). Recombination and expression of classical swine fever virus (CSFV) structural protein E2 gene in *Chlamydomonas reinhardtii* chroloplasts. *Colloids Surf. B Biointerfaces* 55, 26–30. doi: 10.1016/j.colsurfb.2006.10.042

Hempel, F., Lau, J., Klingl, A., and Maier, U. G. (2011). Algae as protein factories: expression of a human antibody and the respective antigen in the diatom *Phaeodactylum tricornutum*. *PLoS ONE* 6:e28424. doi: 10.1371/journal.pone.0028424

Jacob, S. S., Cherian, S., Sumithra, T. G., Raina, O. K., and Sankar, M. (2013). Edible vaccines against veterinary parasitic diseases-current status and future prospects. *Vaccine* 31, 1879–1885. doi: 10.1016/j.vaccine.2013.02.022

Jones, C. S., Luong, T., Hannon, M., Tran, M., Gregory, J. A., Shen, Z., et al. (2013). Heterologous expression of the C-terminal antigenic domain of the malaria vaccine candidate Pfs48/45 in the green algae *Chlamydomonas reinhardtii*. *Appl. Microbiol. Biotechnol.* 97, 1987–1995. doi: 10.1007/s00253-012-4071-7

Khan, I., Twyman, R. M., Arcalis, E., and Stoger, E. (2012). Using storage organelles for the accumulation and encapsulation of recombinant proteins. *Biotechnol. J.* 7, 1099–1108. doi: 10.1002/biot.201100089

Kwon, K. C., Verma, D., Singh, N. D., Herzog, R., and Daniell, H. (2013). Oral delivery of human biopharmaceuticals, autoantigens and vaccine antigens bioencapsulated in plant cells. *Adv. Drug Deliv. Rev.* 65, 782–799. doi: 10.1016/j.addr.2012.10.005

Lossl, A. G., and Waheed, M. T. (2011). Chloroplast-derived vaccines against human diseases: achievements, challenges and scopes. *Plant Biotechnol. J.* 9, 527–539. doi: 10.1111/j.1467-7652.2011.00615.x

Manuell, A. L., Beligni, M. V., Elder, J. H., Siefker, D. T., Tran, M., Weber, A., et al. (2007). Robust expression of a bioactive mammalian protein in *Chlamydomonas* chloroplast. *Plant Biotechnol. J.* 5, 402–412. doi: 10.1111/j.1467-7652.2007.00249.x

Martinez, C. A., Giulietti, A. M., and Talou, J. R. (2012). Research advances in plant-made flavivirus antigens. *Biotechnol. Adv.* 30, 1493–1505. doi: 10.1016/j.biotechadv.2012.03.004

Mason, H.S., and Herbst-Kralovetz, M. M. (2012). "Plant-derived antigens as mucosal vaccines," in *Mucosal Vaccines: Modern Concepts, Strategies, and Challenges*, ed. P. A. Kozlowski (Berlin: Springer-Verlag Berlin), 101–120. doi: 10.1007/82_2011_158

Pelosi, A., Shepherd, R., and Walmsley, A. M. (2012). Delivery of plant-made vaccines and therapeutics. *Biotechnol. Adv.* 30, 440–448. doi: 10.1016/j.biotechadv.2011.07.018

Qin, S., Lin, H., and Jiang, P. (2012). Advances in genetic engineering of marine algae. *Biotechnol. Adv.* 30, 1602–1613. doi: 10.1016/j.biotechadv.2012.05.004

Rasala, B. A., and Mayfield, S. P. (2011). The microalga *Chlamydomonas reinhardtii* as a platform for the production of human protein therapeutics. *Bioeng. Bugs* 2, 50–54. doi: 10.4161/bbug.2.1.13423

Rasala, B. A., Muto, M., Lee, P. A., Jager, M., Cardoso, R. M., Behnke, C. A., et al. (2010). Production of therapeutic proteins in algae, analysis of expression of seven human proteins in the chloroplast of *Chlamydomonas reinhardtii*. *Plant Biotechnol. J.* 8, 719–733. doi: 10.1111/j.1467-7652.2010.00503.x

Rasala, B. A., Muto, M., Sullivan, J., and Mayfield, S. P. (2011). Improved heterologous protein expression in the chloroplast of *Chlamydomonas reinhardtii* through promoter and 5' untranslated region optimization. *Plant Biotechnol. J.* 9, 674–683. doi: 10.1111/j.1467-7652.2011.00620.x

Rosales-Mendoza, S., Govea-Alonso, D. O., Monreal-Escalante, E., Fragoso, G., and Sciutto, E. (2012a). Developing plant-based vaccines against neglected tropical diseases: where are we? *Vaccine* 31, 40–48. doi: 10.1016/j.vaccine.2012.10.094

Rosales-Mendoza, S., Rubio-Infante, N., Govea-Alonso, D. O., and Moreno-Fierros, L. (2012b). Current status and perspectives of plant-based candidate vaccines against the human immunodeficiency virus (HIV). *Plant Cell Rep.* 31, 495–511. doi: 10.1007/s00299-011-1194-8

Rybicki, E. P. (2009). Plant-produced vaccines: promise and reality. *Drug Discov. Today* 14, 16–24. doi: 10.1016/j.drudis.2008.10.002

Soria-Guerra, R. E., Ramírez-Alonso, J. I., Ibáñez-Salazar, A., Govea-Alonso, D. O., Paz-Maldonado, L. M. T., Bañuelos-Hernández, B., et al. (2014). Expression of an HBcAg-based antigen carrying angiotensin II in *Chlamydomonas reinhardtii* as a candidate hypertension vaccine. *Plant Cell Tissue Organ Cult.* 116, 133–139. doi: 10.1007/s11240-013-0388-x

Specht, E. A., and Mayfield, S. P. (2013). Synthetic oligonucleotide libraries reveal novel regulatory elements in *Chlamydomonas* chloroplast mRNAs. *ACS Synth. Biol.* 2, 34–46. doi: 10.1021/sb300069k

Sun, M., Qian, K. X., Su, N., Chang, H. Y., Liu, J. X., and Chen, G. F. (2003). Foot-and-mouth disease virus VP1 protein fused with cholera toxin B subunit expressed in *Chlamydomonas reinhardtii* chloroplast. *Biotechnol. Lett.* 25, 1087–1092. doi: 10.1023/A:1024140114505

Surzycki, R., Greenham, K., Kitayama, K., Dibal, F., Wagner, R., Rochaix, J. D., et al. (2009). Factors effecting expression of vaccines in microalgae. *Biologicals* 37, 133–138. doi: 10.1016/j.biologicals.2009.02.005

Tran, M., Zhou, B., Pettersson, P. L., Gonzalez, M. J., and Mayfield, S. P. (2009). Synthesis and assembly of a full-length human monoclonal antibody in algal chloroplasts. *Biotechnol. Bioeng.* 104, 663–673. doi: 10.1002/bit.22446

Wang, X., Brandsma, M., Tremblay, R., Maxwell, D., Jevnikar, A. M., Huner, N., et al. (2008). A novel expression platform for the production of diabetes-associated autoantigen human glutamic acid decarboxylase (hGAD65). *BMC Biotechnol.* 8:87. doi: 10.1186/1472-6750-8-87

Wigdorovitz, A., Carrillo, C., Dus Santos, M. J., Trono, K., Peralta, A., Gomez, M. C., et al. (1999). Induction of a protective antibody response to foot and mouth disease virus in mice following oral or parenteral immunization with

alfalfa transgenic plants expressing the viral structural protein VP1. *Virology* 255, 347–353. doi: 10.1006/viro.1998.9590

Conflict of Interest Statement: The author Stephen P. Mayfield declares a financial interest in Triton Animal Health, a company making orally available nutritional supplements, and potentially orally available vaccines, should these prove to be biologically functional.

Received: 28 November 2013; paper pending published: 23 December 2013; accepted: 30 January 2014; published online: 17 February 2014.

*Citation: Specht EA and Mayfield SP (2014) Algae-based oral recombinant vaccines. Front. Microbiol. **5**:60. doi: 10.3389/fmicb.2014.00060*
This article was submitted to Microbiotechnology, Ecotoxicology and Bioremediation, a section of the journal Frontiers in Microbiology.

Acknowledgements

Chapter 3, in full, is a reprint of the material as it appears in *Frontiers in Microbiology* 2014. The dissertation author was the primary investigator and author of this paper.

CHAPTER 4:

SYNTHETIC OLIGONUCLEOTIDE LIBRARIES REVEAL NOVEL REGULATORY ELEMENTS IN *CHLAMYDOMONAS* CHLOROPLAST mRNAS

ACS SyntheticBiology

Research Article

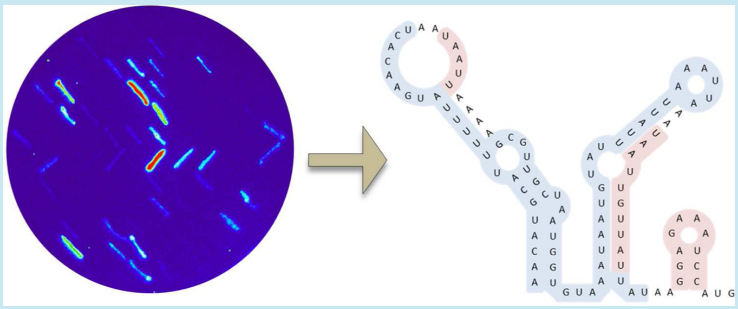
pubs.acs.org/synthbio

Synthetic Oligonucleotide Libraries Reveal Novel Regulatory Elements in *Chlamydomonas* Chloroplast mRNAs

Elizabeth A. Specht and Stephen P. Mayfield*

The San Diego Center for Algae Biotechnology, Division of Biological Sciences, University of California San Diego, La Jolla, California 92093, United States

Supporting Information



ABSTRACT: Gene expression in chloroplasts is highly regulated during translation by sequence and secondary-structure elements in the 5′ untranslated region (UTR) of mRNAs. These chloroplast mRNA 5′ UTRs interact with nuclear-encoded factors to regulate mRNA processing, stability, and translation initiation. Although several UTR elements in chloroplast mRNAs have been identified by site-directed mutagenesis, the complete set of elements required for expression of plastid mRNAs remains undefined. Here we present a synthetic biology approach using an arrayed oligonucleotide library to examine *in vivo* hundreds of designed variants of endogenous UTRs from *Chlamydomonas reinhardtii* and quantitatively identify essential regions through next-generation sequencing of thousands of mutants. We validate this strategy by characterizing the relatively well-studied 5′ UTR of the *psbD* mRNA encoding the D2 protein in photosystem II and find that our analysis generally agrees with previous work identifying regions of importance but significantly expands and clarifies the boundaries of these regulatory regions. We then use this strategy to characterize the previously unstudied *psaA* 5′ UTR and obtain a detailed map of regions essential for both positive and negative regulation. This analysis can be performed in a high-throughput manner relative to previous site-directed mutagenesis methods, enabling compilation of a large unbiased data set of regulatory elements of chloroplast gene expression. Finally, we create a novel synthetic UTR based on aggregate sequence analysis from the libraries and demonstrate that it significantly increases accumulation of an exogenous protein, attesting to the utility of this strategy for enhancing protein production in algal chloroplasts.

KEYWORDS: *oligonucleotide synthesis, translational regulation, chloroplast gene regulation, synthetic biology*

Microalgae have significant potential for mass production of industrial molecules such as biofuels, therapeutic proteins, or industrial enzymes, due to their ability to produce biomass at large scale in a rapid and cost-effective manner. However, current levels of recombinant protein accumulation in algae are well below that achieved in other production hosts,[1] making only the most valuable products economically suitable for algal production. The highest yields achieved to date in the model alga *Chlamydomonas reinhardtii* are around 10% of total soluble protein,[1] but most recombinant proteins accumulate to less than 1%.[2] To fully realize the potential of algae as a biotechnology platform, we need robust and controlled gene expression in both the nuclear and chloroplast genomes.

Although progress has been made understanding gene expression in algae, the regulatory processes that govern gene expression and protein accumulation in the chloroplast remain an area of active investigation. A better understanding of the regulatory processes that govern protein accumulation may enable us to design custom regulatory regions that overcome the current limitations in producing recombinant proteins.

Protein expression in *Chlamydomonas* chloroplasts is regulated primarily during translation, governed by regulatory sequences in the untranslated regions (UTRs) of mRNAs and by *trans*-acting factors that interact with elements in these UTRs. Chloroplast 3′ UTRs are processed at the distal end of a stem-loop structure that appears to protect the mature transcript from 3′ to 5′ exonucleases,[3] but these 3′ sequences appear to have little impact on protein expression[4] although

Received: July 30, 2012
Published: October 25, 2012

there is some evidence that UG repeats in 3′ UTRs are involved in circadian expression.[5] For chloroplast gene expression, most research has focused on the 5′ UTRs of plastid mRNAs, as these have been shown to exert significant control over protein accumulation, impacting both translation rates and mRNA stability.

One of the most well-studied UTRs, the *psbA* 5′ UTR has been shown to interact with a four-member protein complex,[6] and an RNA stem-loop and adjacent ribosome binding site have been identified as key RNA elements required for translation.[7] Several nuclear mutants have been found that affect *psbA* translation or stability, revealing the complex regulatory interactions between the nuclear genome and the plastid genome.[8] *Cis*-acting elements have also been identified in the 5′ UTRs of *petD*,[9] *rbcL*, *atpB*,[10] *psbD*,[11] and a few other genes, but the mechanisms by which these elements exert control over translation are not well understood. The details of plastid RNA elements and translational regulation have been obtained piecemeal through individual biochemical or mutant analyses, making it difficult to know how many elements are present in any mRNA, and how important any one element may be in controlling gene expression. Using a more systematic approach to characterize many chloroplast UTRs may elucidate recurring structures or sequences, leading to a more comprehensive understanding of plastid gene regulation.

In addition to gaining a better understanding of native mRNA elements involved in gene regulation, the thorough systematic analysis of UTR regulatory elements will allow us to design truly synthetic UTRs for driving expression of exogenous genes in transgenic algae. A notable characteristic of some of the most commonly used endogenous chloroplast regulatory regions is that they exhibit autoattenuation and therefore can only be used effectively in a strain in which the native gene has been deleted. For example, the *psbA* regulatory regions are capable of high levels of recombinant protein production, but the *psbA* gene product inhibits expression of additional *psbA* transcript. Therefore these regulatory regions are only useful for driving exogenous gene expression in a non-photosynthetic *psbA* knockout strain, eliminating the energetic benefit of using a photosynthetic organism to produce recombinant proteins at large scale. By identifying regions useful for strong positive regulation across many UTRs, we may be able to mix and match sequence elements from many UTRs to create novel synthetic elements that evade the negative feedback mechanisms involved in autoattenuation while maintaining high expression of the transgene of interest.

Here we present a systematic synthetic biology approach for identifying important 5′ UTR sequence elements involved in modulating chloroplast gene expression. We used large-scale oligonucleotide synthesis to create libraries of variant UTRs and cloned these libraries into vectors driving the expression of a codon-optimized luciferase reporter gene. By selecting pools of transformants that have high, medium, or low expression of the reporter, followed by next-generation sequencing of these different pools, we were able to quantitatively identify effector elements from each of the pooled groups. We validate this method by confirming previous partially characterized elements within the *psbD* 5′ UTR, though we find that these elements extend far beyond the region initially identified through site-directed mutagenesis, highlighting the superiority of a comprehensive, unbiased approach. We then extend the analysis to *psaA* to identify novel regulatory regions within its 5′ UTR. Finally, we demonstrate the predictive power of our method by creating a synthetic UTR, based on the aggregate data from the *psaA* 5′ UTR, that outperforms the wild type version.

RESULTS AND DISCUSSION

Design and Representation of Libraries. Two highly expressed chloroplast mRNA 5′ UTRs, from the *psaA* and *psbD* genes, were selected for variant analysis to identify potential regulatory elements. These genes encode subunits of photosystem I (*psaA*) and photosystem II (*psbD*). The prevalence of these photosystem proteins in the thylakoid membranes, as well as their high turnover as a result of photodamage, contributes to their inherent high expression. These UTRs were also selected because of their short length, ensuring that they could be synthesized with high fidelity to make a variant library with a manageable number of members.

The *psbD* mRNA is transcribed with a 74 nucleotide 5′ UTR that is cleaved to 47 nucleotides through a processing step that appears to be coupled with translation.[11,12] Previous analyses of the regulatory elements in the *psbD* 5′ UTR have been somewhat complicated by the introduction of multiple mutations in a single strain, insertion of restriction sites into the UTR, or use of the *psbD* protein itself as a reporter for expression.[11,13] In this last instance there is the additional complication of autoregulation of synthesis in which unassembled subunits of the photosystem complexes can significantly impact their own synthesis.[14] Because D2 is the limiting assembly partner in PSII, a mutant UTR that raises or lowers the amount of D2 produced will have significant downstream effects on the regulation of all other subunits of the PSII complex.[14] As a result of this impact on chloroplast gene regulation, the readout may reflect the downstream consequences of the mutant UTR instead of providing an unbiased analysis of the UTR's performance alone.

In our system we have reduced these epistatic effects as much as possible to allow an unbiased analysis that can be compared to the results of previous studies. By constructing libraries using the USER cloning system[15] to clone the synthesized UTRs directly downstream of their respective promoter, there are no areas containing engineered restriction site sequences that could potentially alter UTR activity. All of our UTRs were specifically designed and synthesized, so the entire UTR was scanned and all possible variants generated to cover the entire UTR. All nucleotide replacements were made with homopolymer adenosine and not with random combinations of nucleotides, to ensure that properties like local GC content were not unpredictably affected and so that new RNA secondary structures were not inadvertently introduced.

We used the chloroplast codon-optimized luciferase reporter *luxAB* for expression analysis[16] instead of the D2 protein as a reporter, and all constructs were examined in a photosynthetic strain to ensure that there are no translational regulation artifacts from defective photosystem assembly. The selection for transformation was restoration of photosynthesis, as constructs were transformed into a strain with a deletion in the *psbH* gene which was subsequently repaired to wild type by the transformation vector as previously described.[17] In all constructs, the 3′ UTR remained unaltered; it is the native 3′ UTR from the *psbA* transcript.

The *psaA* 5′ UTR is far less well-characterized than the *psbD* 5′ UTR. The *psaA* 5′ UTR appears to extend up to 238 nucleotides upstream of the initiation codon, based on reverse transcription data from the gene's original identification in

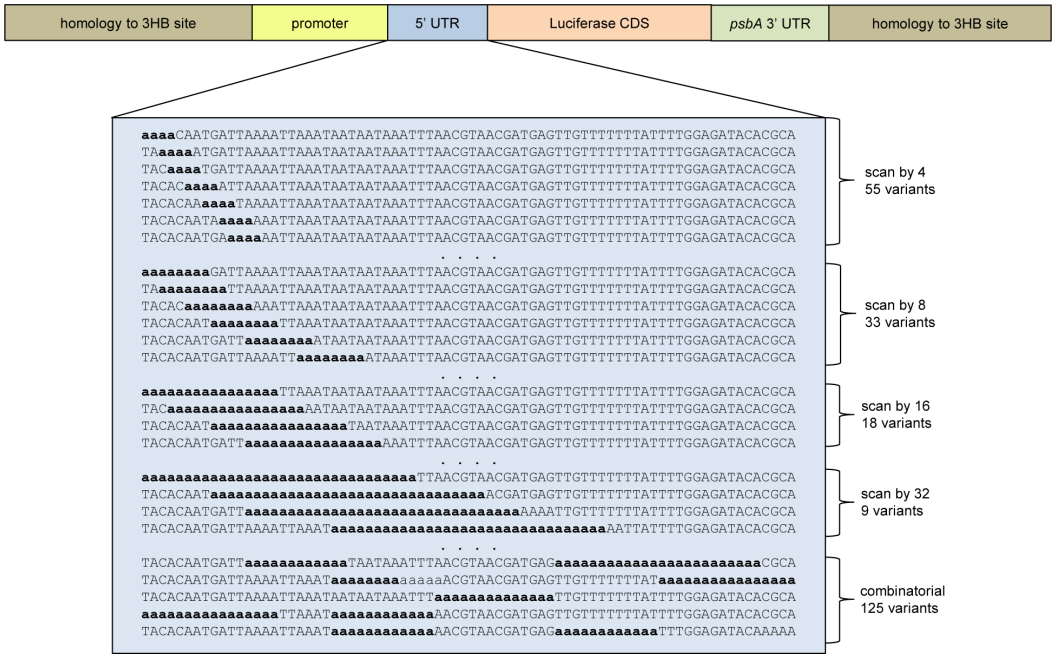


Figure 1. Schematic of the transformation vectors containing variant 5′ UTRs. The 3HB site is a silent site in the chloroplast genome near the *psbH* locus. The luciferase coding sequence has been codon-optimized for higher expression in the algal chloroplast. The 3′ UTR is kept as that from *psbA* regardless of the origin of the 5′ UTR, while the promoter correlates with the 5′ UTR. Representations of the *psbD* poly(A) scanning pools and combinatorial pools are illustrated below. Not all sequences are shown, but the total number of variants of each type is indicated; scans continue across the full length of the UTR, and the combinatorial pool contains all combinations of wild type sequence and poly(A) in 12-mer segments. The full list of sequences is available in the Supporting Information. Substitutions from wild type are shown in bold and lowercase.

Chlamydomonas.[18] However, subsequent chimeric constructs have relied on as little as 245 nucleotides upstream of the start codon to serve the function of both promoter *and* 5′ UTR,[19] suggesting that the promoter is only the seven preceding bases, which seems unlikely. For this study we manipulated only the 100 nucleotides immediately upstream of the start codon due to oligonucleotide synthesis length constraints. The full 245 nucleotides were included to serve as the promoter and UTR, but the first 145 nucleotides were kept as wild-type sequence while only the latter 100 were mutated.

Each UTR was manipulated in two ways. In the first permutation, the native sequence was scanned across and replaced with one of four different length stretches, either 4, 8, 16, or 32 nucleotides, of adenosine homopolymer. Each UTR in this set, 162 variants for *psaA* and 113 for *psbD*, had a single segment of sequence replaced by these poly(A) stretches. In the second set, the UTR was divided into segments of 12 nucleotides, and these segments were either left as native sequence or were replaced by 12 bases of poly(A) in all possible permutations. This set comprises 512 variants for *psaA* and 128 variants for *psbD* and captures combinatorial effects of removing sequence elements from parts of the UTR that are not immediately adjacent. In all of these UTRs, the total length of the UTR and the spacing remains unaltered to eliminate confounding effects of changing the spacing between important elements. This set also allows us to determine the minimal set of elements required for translation. Figure 1 provides representations of the library members, using mutated *psbD* sequences as examples, as well as a schematic for the algal transformation vector.

Poly(A) was chosen as the replacement sequence in both of these manipulations in keeping with the A-T rich nature of *Chlamydomonas* chloroplast UTRs and also to decrease the potential for forming structured elements by GC base pairing. There is some potential that addition of poly(A) stretches could enhance translation as there is evidence of a role for poly(A)-binding proteins in light-induced translational activation of chloroplast mRNAs.[20] In addition, the chloroplast homologue of the bacterial ribosomal protein S1, which binds stretches of poly(U) to assist in association of the ribosome, may in fact bind stretches of poly(A) instead, as demonstrated in higher plant chloroplasts.[21]

The synthetic oligonucleotides were converted into double-stranded fragments by second-strand synthesis and amplified by PCR using primers containing specific uracil residues such that the PCR fragments could be seamlessly cloned into transformation vectors by USER (uracil-specific excision reagent) cloning.[15] Illumina sequencing of bacterial plasmid minipreps from cells harboring the complete pool of algal transformation vectors revealed that all 241 designed *psbD* UTRs were unambiguously present in the DNA transformed into algae. One *psaA* sequence was potentially lost through the cloning steps from synthesized oligonucleotide to full transformation vector, but the other 673 out of 674 designed sequences were unambiguously identified in the final transformation DNA pool.

***psbD* 5′ UTR Element Identification.** Algal transformants containing a single variant from the library were characterized

as individual clones by assaying plates containing arrays of individual colonies by luminescence following addition of a luciferase substrate (see Figure 2). Each clone was qualitatively classified as high, intermediate, or low expression based on their luciferase signal. Clones with no visible expression were included in the low-expression pool. DNA from high-, low-, or medium-expression pools was sequenced using Illumina 150 bp paired-end reads (for further details see Methods). Reads were parsed by barcode and then mapped to the designed UTRs. Only those with a 100% match to the UTR reference sequences were retained for analysis. After removal of low quality and ambiguous reads, over 1.3 million reads from the *psbD* pools were analyzed.

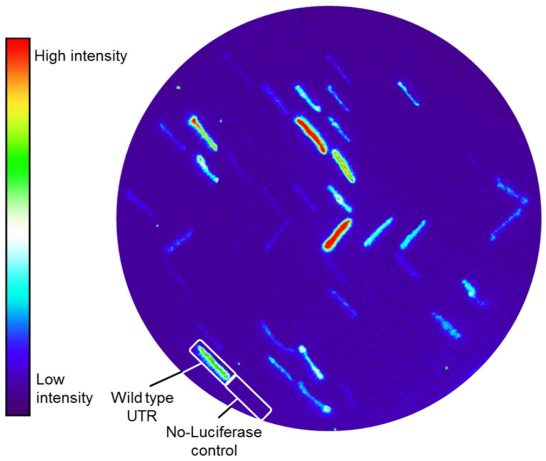


Figure 2. False-color heat-map image of one plate of 100 unique clones expressing luciferase. Indicated in the lower left is the patch of algae expressing luciferase driven by the wild-type *psaA* UTR. Immediately below is a patch of wild type *Chlamydomonas reinhardtii* that has not been transformed with luciferase; as expected, no luminescence is detected in this sample.

We tallied whether each reference sequence had an A or the wild type nucleotide at each position along the UTR and then multiplied by the number of sequence reads that mapped to each reference. After summing across all sequence reads, we generated a likelihood map for the probability that an A (mutation) or the wild type nucleotide was present at any position of the UTR, for each expression-level pool. For instance, there was a 96% chance that a clone in the high-expression pool had a T at the −43 position relative to the AUG start codon in *psbD*, whereas a clone in the low-expression pool had only a 79% chance of having a T at that position. All clones that did not have a T at this position have an A instead, as this was the only substitution made (and all mutations to another nucleotide were filtered out in the read mapping). Figure 3 graphically illustrates the results for the *psbD* 5′ UTR analysis.

We mapped the potential elements identified using this analysis onto a secondary structure map of the *psbD* 5′ UTR as predicted by Quikfold[22] and confirmed by RNaseH mapping.[13] Previously identified RNA stability elements[11] are shown in orange, while regions previously implicated in translational activation[11] are shown in green. In Figure 4b, we highlight in blue the regions where our data indicate the high-expression pool has greater than 90% conservation of the wild type sequence; in red are regions where this conservation is less than 90% in the high expression lines. These cut-offs were determined empirically by comparison with the previously identified elements, as χ-square tests using 3 × 2 contingency tables to determine the statistical significance of the differences between each pool at each position indicate that all the points are significant ($p \ll 10^{-6}$) due to the large read counts. Adenosine nucleotides that lie on the boundaries of these regions cannot be assigned either way, as all sequences have adenosine at these positions regardless of whether that region was mutated or left as wild type sequence.

Using this method, we confirmed two of the three previous characterized elements that were defined by biochemical and

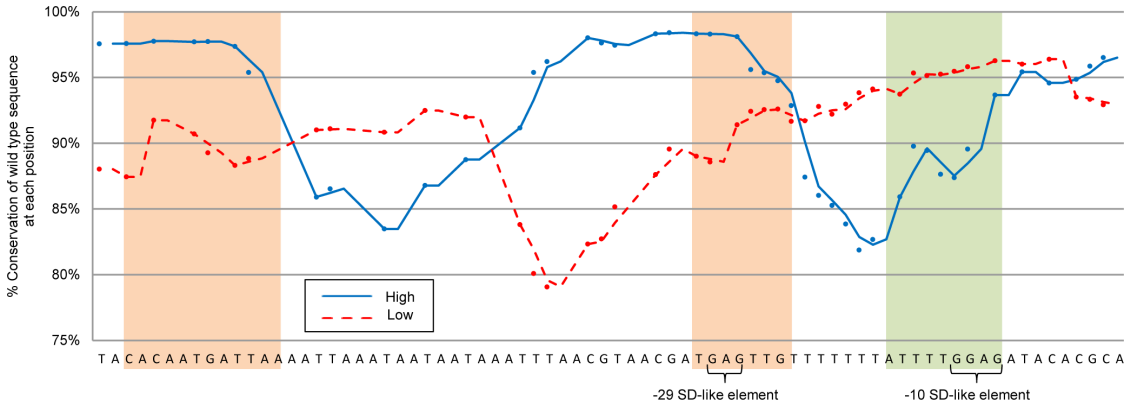


Figure 3. Conservation of wild-type sequence at each position along the *psbD* 5′ UTR. The ATG start codon is immediately downstream of the sequence shown. Blue represents aggregate data from the high-expression pool and red represents the low-expression pool. Trend lines are two-point moving averages of positions where the wild type nucleotide is not already an adenosine (i.e., only positions where mutation caused an actual nucleotide change are considered here). Sequences highlighted in orange have been previously identified as important for RNA stability. The sequence in green was previously implicated in translation, but later it was determined that it is merely the spacing it provides rather than the sequence itself that is important.

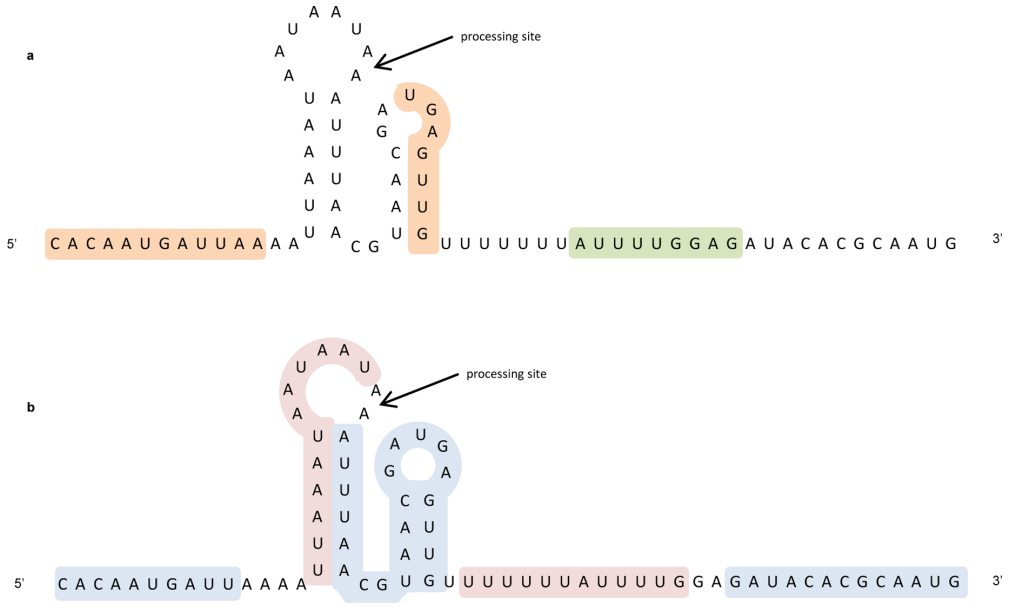


Figure 4. Depiction of the *psbD* 5′ UTR secondary structure. (a) The previously identified regulatory elements for message stability (orange) or translation (green) are highlighted. (b) Regions identified by our method as conserved in the high expression pool are highlighted in blue; regions that favor or tolerate mutation in the high expressers are highlighted in red. The previously described stem-loop containing the AUG is not depicted here because the stem requires additional nucleotides in the *psbD* coding sequence that are not present in the *lux* coding sequence, so this structure was not relevant to our studies. Furthermore, this secondary structure was shown to not exert much effect in a background in which the spacing between all other 5′ UTR elements remains unaltered.[13] Though these secondary structures have been verified experimentally *in vitro*,[13] it is possible that they do not exist *in vivo*, particularly upon protein binding. At the least, the stem-loop nearest the 5′ end cannot exist in the processed form of the transcript.

site-directed mutagenic studies of the *psbD* UTR. Our data significantly enlarge one of these elements and clarify the boundaries of these previously identified elements. These two regions, thought to be important in mRNA stability,[11] are both identified in our study as among regions highly conserved in our high-expression pool (see orange boxes in Figure 3).[11] The first element, at the 5′ end of the unprocessed mRNA, may form the binding site for a previously identified RNA-binding protein that interacts only with the unprocessed *psbD* 5′UTR and is critical for message stability,[23] and our data maps quite closely. The second stability element is implicated in processing and stable accumulation of the processed form of the transcript,[11] and our analysis shows that this element is likely twice as large as previously predicted. Our results indicate that this element extends all the way up to the processing site, which agrees with previous evidence that RNA-binding proteins act as protective caps at the ends of chloroplast transcripts to protect them from exonuclease degradation.[24] This processing site, therefore, may not be a single site of action by specific processing machinery; it may simply be the site where the cap protein no longer protects the 5′ end of the transcript from degradation.

The region indicated in green in Figure 3 had been identified, along with the adjacent uracil tract, to be important for translation.[11] Our results do not find this sequence to be important for expression, and this is supported by more recent studies that revealed that in fact neither the sequence nor the secondary structure of this U-rich region was essential for *psbD* translation; this region predominantly functions as a spacer for elements located on either side of it.[25] Since our mutational constructs do not alter spacing between unmutated regions, we do not observe the same importance of this region, in agreement with the work by Ossenbühl et al.[25] Again, this demonstrates the strength of our unbiased method for assessing the contribution of individual sequence elements without unintended secondary effects.

Some *Chlamydomonas* chloroplast genes contain a sequence resembling a Shine-Dalgarno consensus sequence at around −10 from the start of translation, but these elements are often dispensable. For example, in *atpB*, *atpE*, *rps4*, and *rps7*, eliminating the −10 SD-like sequence or replacing it with a canonical SD appears not to affect expression of any of these mRNAs.[26] In Figure 3 we see that indeed for *psbD* this −10 SD sequence is not important for expression, but that a SD-like GAG further upstream at position −29 is indeed strongly conserved in the high expression lines, indicating that this may be the true ribosome binding site as previously suggested.[11]

Our scanning mutagenesis analysis can also elucidate potential elements that function as negative regulatory elements. In order to determine whether a minimally conserved region is merely unimportant for high expression or whether mutations in that region are indeed favored in the high expression lines, we compared the high expression pool to the sum of all sequence reads obtained in all pools (high, medium, and low expression). Areas where the wild-type sequence conservation is *lower* for the high expression lines than for the sum of all reads indicate that mutations in these sequence

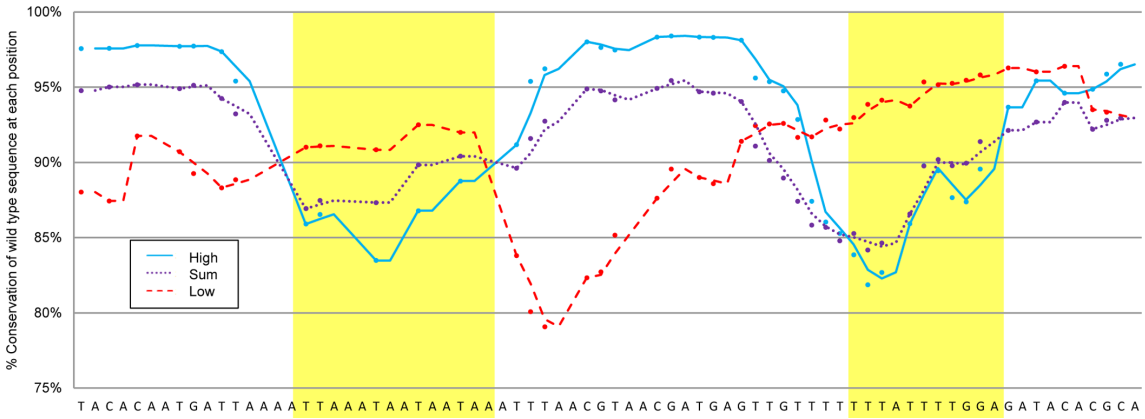


Figure 5. Overlaid plots of the high-expression pool (blue), low-expression pool (red), and the sum of all *psbD* UTR reads (purple). Highlighted in yellow are the only two regions where the following criteria are met: the high expressers show lower conservation than both the low-expresser pool and the sum of all reads, and the low expressers show higher conservation than both the high-expresser pool and the sum of all reads. These regions may be elements involved in negative regulation.

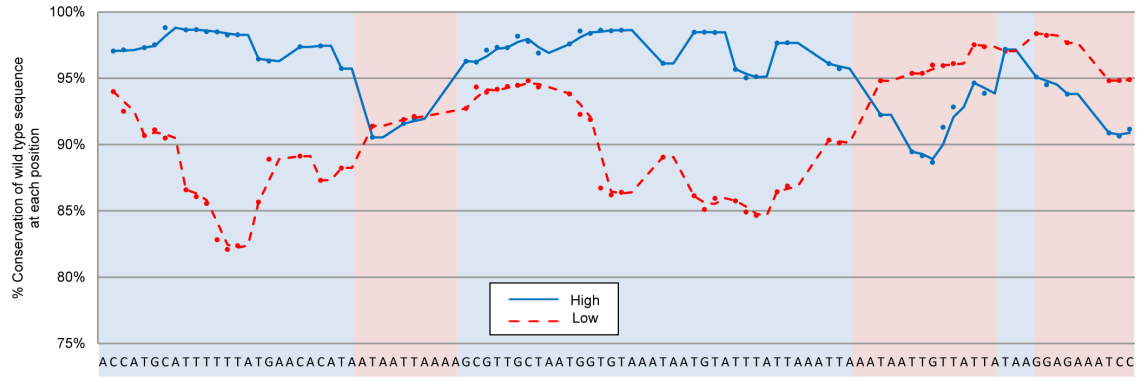


Figure 6. Conservation of wild type sequence at each position along the last 100 nucleotides of the *psaA* 5′ UTR. Blue represents the high-expression pool, and red represents the low-expression pool. Regions that exhibit greater than 95% wild-type sequence conservation in the aggregate data of the high expressers are designated with a blue background; regions with less than 95% conservation have a red background, indicating that they are either unimportant for high expression or indeed may be sites of negative regulation.

elements are in fact favored for high expression, suggesting that they may be regions where negative regulatory factors interact.

The plots of the mutation frequency for the high and low expression lines and the sum of all reads are shown in Figure 5, with the percent conservation of the wild type sequence at each position graphed along the entire *psbD* UTR. There are two regions where the high expression lines contain mutations at a rate much higher than observed in all the algal reads from the high, medium, and low pools combined. Lending additional evidence for the negative regulatory impact of these elements, the low-expression pool actually shows increased conservation of these regions above that observed in the sum of all reads.

This indicates that while some mutants certainly lose their ability to express the reporter due to mutation in an essential element, there is overall a significant bias toward retaining these two putative negative regulatory elements in the low-expression pool. Interestingly, one of these regions is upstream of the processing site, suggesting that the negative control it may exert must occur prior to processing. Note that the plot for the sum of all reads does not always fall exactly between the high- and low-expresser pools, as the majority of our clones were classified as intermediate expressers. In certain regions, these intermediate expressers conserved wild-type sequence below the levels observed in either the high or the low expressers. Further work will be needed to determine whether these regions may have dual functionality or are involved in more complex regulatory mechanisms.

Identification of Regulatory Elements within the *psaA* 5′ UTR. Phenotyping, sequencing, mapping, and data analysis for the *psaA* clones were performed exactly as described above for the *psbD* clones. We have no prior knowledge of important sequence elements in the *psaA* 5′ UTR, so all findings from the sequence analysis are novel. These results are based on greater than 12.8 million reads from *psaA* clones that were retained after all the quality and mapping constraints were applied.

As with the *psbD* results, overall we observe that increased sequence conservation in the high expression pool directly corresponds with poor conservation in the low expression pool, as we would expect. In other words, elements that are critical for expression are retained in the high-expression lines and have

been lost in the low-expression lines. In Figure 6, we designate with a blue background the regions that are highly conserved in the high expression lines and with a red background the regions that tolerate or favor alteration in the high expression lines. In this analysis, our empirical cutoff for defining these regions was raised to 95% conservation of the wild type sequence in the high expresser pool, since nearly all of the positions were conserved at greater than 90%. In Figure 7 we map these regions onto a predicted secondary structure of the *psaA* 5′ UTR.

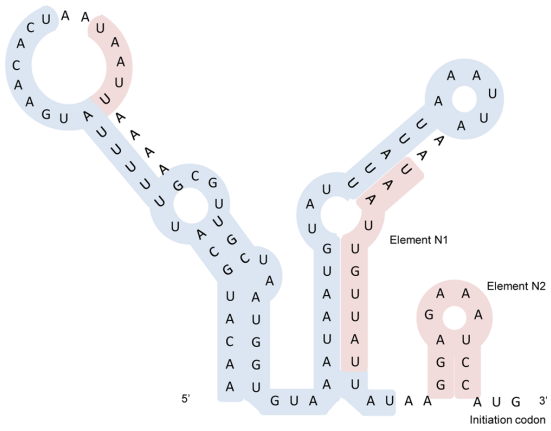


Figure 7. Depiction of the secondary structure of the mutagenized region of the *psaA* 5′ UTR. This structure is predicted by QuikFold version 2.3 at 25 °C but has not been confirmed experimentally. Regions identified as conserved in the high expression pool are highlighted in blue; regions that favor or tolerate mutation in the high expressers are highlighted in red.

In our analysis of the *psaA* 5′ UTR, we find a lack of conservation at the RBS-like GGAG immediately upstream of the start codon, similar to what we observed for the −10 GGAG in the *psbD* UTR. In fact, we observe a strong propensity to eliminate this sequence in the high-expressing clones. Again similar to the *psbD* UTR, in the *psaA* UTR we observe very strong conservation of a similar RBS-like sequence further upstream in the high expression lines. The GGUG at position −51 is among the most highly conserved regions across the entire analyzed region and forms a perfect complementary match with four nucleotides in the anti-Shine-Dalgarno sequence at the 3′ end of the 16s rRNA.[27] Though this site is significantly further upstream from the start codon than in the *psbD* mRNA, it has been noted that the spacing between ribosome binding sites and initiation codons is much more variable in chloroplasts than in bacteria,[28] and there are other *Chlamydomonas* chloroplast genes with similar spacing such as the −51 SD-like element in the *rps12* transcript.[29]

Aside from *cis*-acting secondary structure, many UTR elements act as recognition sites for nuclear-encoded *trans*-acting factors that promote RNA stability or translation initiation. There is evidence that the *psaA* 5′ UTR participates in a feedback loop, as exogenous genes under the control of the *psaA* 5′ UTR accumulate to higher levels in a strain with a *psaA* splicing defect than when the *psaA* gene product is being synthesized normally.[30] The sequence elements that participate in this feedback may have already been identified in this analysis, and we can test candidate mutated UTRs in this splicing defect background to determine which UTR elements are responsible.

Regions where negative regulators may interact, indicated by particularly low conservation in the high-expression pool and high conservation in the low-expression lines, are highlighted in Figure 8. The involvement of these two elements in negative regulation is strongly supported by preliminary analysis of several mutant-*psaA* 5′ UTR clones that express the reporter luciferase at levels higher than the wild-type *psaA* UTR. Twenty-four of the brightest individuals from the initial luciferase assays were grown in liquid culture and then assayed for *lux* expression using equal numbers of cells. Of these, 15 were found to express *lux* at higher levels than the wild-type *psaA* UTR; these individuals were sequenced with traditional Sanger sequencing. The results shown in Table 1 indicate that the putative negative regulatory elements identified by the

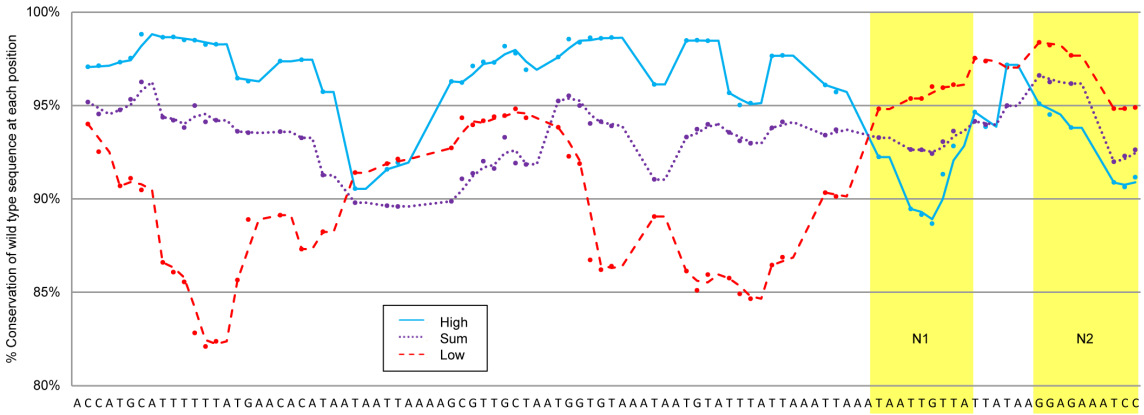


Figure 8. Overlaid plots of the high-expresser pool (blue), low-expresser pool (red), and the sum of all *psaA* UTR reads (purple). As in Figure 5 above, highlighted in yellow are the only two regions where high expressers show lower conservation than the sum of all reads *and* low expressers show higher conservation than the sum of all reads. These elements are candidates for recognition sites for negative regulatory factors, and preliminary analysis of brighter-than-wild-type *psaA* mutant UTR individuals supports this notion.

Table 1. Sequence Analysis of 15 Individual Clones Expressing Luciferase at Levels Higher than the *psaA* UTR

clone number	% of *psaA* UTR-driven luciferase expression	mutated in region N1	mutated in region N2
A1-2	115	no	yes
A1-8	126	yes	no
A1-15	106	no	yes
A1-25	159	no	yes
A1-26	238	yes	yes
A1-98	112	yes	yes
A2-20	104	no	yes
A2-56	122	yes	no
A2-58	137	no	yes
A2-81	107	yes	yes
A2-85	115	yes	yes
A4-37	116	yes	yes
A4-51	126	no	yes
A4-60	120	yes	yes
A5-32[a]	295	no	yes

[a]Clone A5-32 has a two-nucleotide deletion at the 3′ end in addition to the poly(A) substitution disrupting region N2.

aggregate sequence analysis are indeed highly mutated in all 15 of these individuals. This suggests that it is necessary to release inhibitory regulation conferred by these regions in order to achieve expression levels beyond those observed with the wild-type *psaA* UTR.

Ongoing work to more thoroughly characterize the highest-expressing candidates from the libraries will lend insight into the mechanisms underlying the varying expression levels we observe.

Constructing a Synthetic UTR. The initial aim of this work was to elucidate positive and negative regulatory elements in chloroplast 5′ UTRs using an unbiased scanning approach to better characterize the regulatory mechanisms of gene expression in algal chloroplasts. Identifying these key regulators of chloroplast gene expression allows us to develop tools for increased recombinant protein accumulation in algal plastids. In particular, this information also allows us to develop modified or wholly synthetic UTRs designed to overcome the limitations currently encountered using algae as a protein production platform for complex therapeutic molecules or industrial enzymes, or in metabolic engineering to produce high-value small molecules.

Designing a heterologous but highly effective UTR may avoid the negative feedback issues currently encountered when using a native UTR from a photosystem gene to drive recombinant protein production. Furthermore, homologous recombination in the chloroplast is so efficient that we have observed crossover events between endogenous UTRs and the introduced corresponding regulatory elements driving exogenous transgenes, resulting in loss of transgene expression over time (unpublished data). With multiple unique synthetic UTRs at our disposal, we can introduce multiple exogenous genes at once without recombination between repeated native UTRs. This will be especially important for introducing multiple genes, for example, novel metabolic pathways for producing biofuels or high-value secondary metabolites.

Taking the aggregate sequencing results from the *psaA* UTR variant library, we designed a simple synthetic UTR to drive expression of a useful industrial enzyme. The putative negative regulatory regions N1 and N2 indicated in Figure 8 were substituted with poly(A) stretches, and the remainder of the

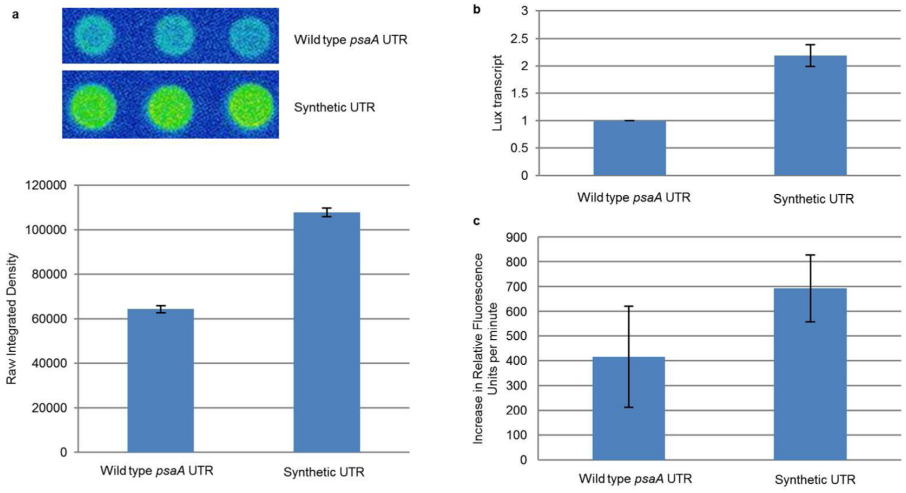


Figure 9. A designed synthetic UTR significantly exceeds activity from the wild-type UTR. (a) Results of a quantitative luciferase spotting assay are shown for the synthetic UTR compared to the wild-type *psaA* UTR, using the same intensity scale depicted in Figure 2. Below, this difference is quantitated by integrated density over the spots; the synthetic UTR produces 67% higher expression. (b) Fold change in transcript level relative to the *lux* transcript in the wild-type *psaA* UTR strain, as determined by quantitative RT-PCR. The error bar indicates standard deviation among four independent synthetic UTR clones. The synthetic UTR produces approximately 2.1 times as much transcript as the endogenous UTR. (c) Xylanase activity of four independent homoplasmic lines of each construct (synthetic UTR and wild-type *psaA* UTR) was analyzed in triplicate using a fluorogenic substrate. Activity is expressed as the increase in relative fluorescent units (RFU) per minute in the linear range between 2 and 6 min after substrate addition. Error bars indicate one standard deviation from the mean. In very close agreement with the luciferase assays, when driving xylanase the synthetic UTR on average results in 66% greater expression than the wild-type *psaA* UTR.

UTR was left as wild-type sequence. This synthetic UTR and the wild-type *psaA* UTR as a control were cloned into identical chloroplast transformation vectors driving luciferase expression. Quantitative luciferase spotting assays were used to determine the difference in expression, and quantitative RT-PCR was used to examine transcript level differences between the two projects. Indeed, the transcript level for luciferase mRNA is approximately 2-fold higher with the synthetic UTR compared to the endogenous *psaA* UTR, which is close to the observed increase in expression as determined by luciferase assay. This suggests that elements N1 and N2 may negatively affect mRNA stability, and that their replacement can enable the transcript to persist and therefore result in increased translation.

The synthetic UTR and endogenous control were also cloned into a chloroplast transformation vector driving a chloroplast codon-optimized xylanase II (*xyn2*) gene from *Trichoderma reesei*[31] but otherwise identical to the luciferase vector previously described. Xylanase is an industrially important enzyme for the pulp and paper industries and will become increasingly valuable for converting agricultural wastes and hemicellulosic feedstocks into animal feed and biofuels.[32] Based on xylanase activity in cell lysates, our designed synthetic UTR produces nearly 70% more enzyme than the wild-type *psaA* UTR. This correlates well with the effect observed in the luciferase strains, demonstrating that synthetic UTRs can be effective for increasing transgene expression from a number of genes of interest.

We have shown the utility of using a synthetic oligonucleotide library to generate a large variant library coupled with phenotype screening and next-generation sequencing to rapidly identify critical regulatory elements in plastid 5′ UTRs. There are many cases where this synthetic high-throughput technique could allow rapid identification of RNA elements required for gene expression. RNA binding assays have identified many binding partners for specific mRNAs or leader sequences, but in many instances the follow-up work to determine exactly where and how these factors interact with the UTR has not been done. Our approach can significantly accelerate this process and allow us to gather enough information from distinct UTRs to begin to view chloroplast gene regulation from a systems level, rather than on a gene-by-gene basis. There is already evidence that some RNA binding proteins (RBPs) such as RBP46 may interact with multiple mRNAs,[33] but the current information is not comprehensive enough to allow us to formulate consensus binding sites or estimate tolerance of variations on these elements.

New advances in computational prediction of RBP recognition sites may provide avenues for synergistic discovery of RBP−RNA interactions, further accelerating our understanding of gene regulation in the chloroplast. Future work to map *trans*-acting RPBs to each *cis*-element can contribute to predictability of RBP sites for engineering rationally designed RBPs. Recently, the maize chloroplast RNA-binding protein PPR10 has been engineered to bind novel RNA sequences,[34] akin to the combinatorial engineering of TAL Effectors that have garnered much attention as tools for genetic engineering. A better understanding of RNA-binding proteins and their recognition sequences may soon facilitate engineering both RNAs and their binding sites to enable highly optimized gene regulation.

This method is far more comprehensive, high-throughput, and unbiased than previous methods involving discrete mutated versions of UTRs. We can examine the entire length of UTRs from a single synthesized library and identify regions of these UTRs that may not be recognizable *a priori* as important for chloroplast gene expression. Furthermore, we demonstrate the practical utility of this method by introducing a novel synthetic UTR based on the aggregate sequencing data from the synthetic library of the previously uncharacterized *psaA* UTR. We show that this designed UTR exhibits increased protein accumulation compared to the wild-type *psaA* UTR from which it was derived, and that this enhanced expression is maintained when the synthetic UTR is used to drive expression of an industrial enzyme, demonstrating that the elements identified by our method will function with a variety of exogenous recombinant genes.

METHODS

Design and Synthesis of UTR Libraries. We designed 915 sequences as synthetic variants of the 5′ UTRs of two *Chlamydomonas reinhardtii* genes, *psbD* and *psaA*. The *psbD* and *psaA* genes were selected because of their high expression levels and relatively short 5′ UTRs. Library sequences were designed with two distinct goals, manifested in two sets of mutant UTRs for each gene. In one set, we scanned across the entire UTR, altering a single sliding window of nucleotides changing them from the wild type sequence to poly adenosine. The length of this window was either 4, 8, 16, or 32 nucleotides, and the window moved along the UTR by 1, 2, 4, or 8 nucleotides at a time, respectively to the length of the window. In the second set, we divide the UTRs into windows of 12 nucleotides each and then create all combinations of these windows with the background sequence being poly adenosine. This set was created to capture potential interactions between spatially disparate regions of the UTR and to determine the minimal set of elements required for translation in each of these UTRs. These strategies are represented in Figure 1, along with examples of each set of variant UTRs.

UTR variants were synthesized as single-stranded DNA oligonucleotides using the Agilent Oligo Library Synthesis (OLS) platform[35,36] and consolidated into a single pool. The 915 unique sequences for *psaA* and *psbD* were synthesized with 28-fold redundancy. In addition to the UTR sequence, the oligonucleotides include a portion of the *lux* gene downstream, as well as regions unique for each gene corresponding to the promoter region immediately upstream of the UTR, enabling PCR amplification of gene-specific subpools from the total OLS pool.

Cloning and Vector Creation. Using gene-specific primer sets, two subpools were amplified, one for the *psaA* UTRs and one for the *psbD* UTRs, using the original oligonucleotide library as template. The reverse primer was common to all oligos, annealing to the 3′ end of the oligos corresponding to the first 20 nucleotides of the *lux* gene. The forward primer defined the subpool, as it annealed to the endogenous promoter upstream of the UTR. All oligos had to be synthesized with equal length, so all were precisely 140 nucleotides, which included a minimum of 20 nucleotides on either side of the UTR for the amplification primers to anneal. These PCR products were then cloned using USER fusion technology[15] into a truncated version of the transformation vector, consisting of just the UTRs and *Chlamydomonas* codon-optimized luciferase reporter coding sequence[16] inserted in a pGEM backbone. Two fragments, one consisting of the UTR and one for the backbone, were amplified by PCR using the uracil-containing primers provided in Supplementary Table S3,

mixed in equimolar quantities along with the USER enzyme (New England BioLabs, Ipswich, MA, USA), and then incubated for 30 min at 37 °C followed by 20 min at 25 °C prior to transformation into DH5α chemically competent *E. coli*.

This step was necessary because the full chloroplast transformation vectors proved too large to amplify by PCR for the USER fusion technique.[15] Classical restriction ligation cloning techniques were not considered for this project because they would have left a restriction site scar between the promoter and UTR, which could have affected sequence or spatial elements important for translation.

After USER fusion cloning into the pGEM-based vectors, the entire luciferase expression cassette (Figure 1) was cloned into a chloroplast transformation vector[17] by restriction ligation using a *BamH*I site upstream of the promoter and a *Spe*I site downstream of the 3′ UTR. As with the first cloning step, this procedure was performed on gene-specific pools, not on individual clones.

To determine if there was bias in the original oligonucleotide library or if bias was introduced into the library at any of these cloning steps, we analyzed a pool of plasmids at each step with next-generation sequencing. In each cloning step we generated greater than 1,000-fold more colonies than unique library sequences, as determined by plating a small aliquot of the transformation before pooling the rest into liquid LB culture. Analysis of loss of library diversity is addressed in the results.

Algal Transformation, Strain Isolation, and Expression Analysis. Chloroplast transformation vectors were purified using the GeneJet bacterial miniprep kit (Fermentas Inc., Glen Burnie, MD, USA) from the pooled liquid bacterial culture described above, containing all variants of the library for either *psaA* (675 variants) or *psbD* (240 variants). As previously described, particle bombardment with gold particles (S550d, Seashell Technologies, San Diego, CA, USA) was used to transform the algal chloroplast of a *psbH*- strain[37] using a PDS-1000/HE biolistic transformation chamber (Bio-Rad, Hercules, CA, USA). Transformants were recovered on minimal media to select for restoration of the *psbH* locus following homologous recombination with the transformation vector.[17]

For the *psaA* library, 6979 individual algal clones were analyzed, giving 97.9% certainty that all 674 designed sequences were present in the algal library assuming unbiased selection with replacement. For the *psbD* library, 2704 clones were analyzed, giving 99.7% certainty that all 241 designed sequences were present in the library. Individual algal clones were replated three times on minimal media to drive the cells toward homoplasmy, reducing the effect of variable heteroplasmy on expression phenotype. Before assaying expression levels of the luciferase reporter, stronger expression was induced with a light shift of the plated cells. Plates were kept in the dark for 36 h and then shifted to bright light for 8 h prior to assaying. After adding 150 μL of a mixture of 3% decanal (Sigma, St. Louis, MO, USA) in heavy mineral oil (Sigma) to the plate lid and spreading it evenly with a sterile plastic inoculating loop, the plate was covered with the lid and the clones were incubated in the dark with the substrate for 5 min. Colonies were then assayed for luciferase expression using a 5-min exposure on an EG&G Berthold NightOwl Imager (Berthold Technologies, Bad Wildbad, Germany). Clones were pooled into three categories by visual estimation: high expression, low expression, and intermediate expression. Figure 2 shows a false-color example of the output of this assay for one sample plate. Expression was not normalized to cell count because the patches were assayed directly, but no notable differences in growth were observed among the clones, allowing the intensities to be directly compared between patches.

For a more quantitative phenotypic analysis of 24 of the brightest *psaA* library members and of the synthetic UTR, clones were grown in 6 mL of minimal media for 3 days, cell density was determined by hemocytometer, and the cultures were spun down and resuspended in HSM. Five microliters of each culture, containing a total of 3×10^6 cells, was spotted in triplicate onto minimal media plates alongside a strain with the wild-type *psaA* UTR driving luciferase prepared in the same manner. These equal-cell-count spots were allowed to grow for 2 days and then were subjected to a light-shift and assayed exactly as described above. Luciferase expression was quantitated using integrated density measurements on ImageJ software (available at http://rsbweb.nih.gov/ij/) across equal pixel areas; the average of the three spots for each clone was divided by the average for the three *psaA* UTR-driven spots on the same plate to obtain percentage of *psaA* UTR-driven expression. Those expressing luciferase at a level higher than the wild-type *psaA* control were sequenced with traditional Sanger sequencing by PCR-amplifying the UTR of interest using a forward primer in the promoter and a reverse primer in the *lux* gene (see Supplementary Table S3 for primer sequences). The PCR product was treated with Exonuclease I (New England BioLabs) and Shrimp Alkaline Phosphatase (SAP, Fermentas Inc.) (10 μL PCR product, 0.2 μL Exonuclease I, 0.2 μL SAP; 37 °C for 30 min followed by 85 °C for 15 min) and sent directly for sequencing using the reverse primer.

Next-Generation Sequencing and Data Analysis. Up to 250 colonies from a single expression pool were inoculated into each of 60 flasks of tris acetate phosphate (TAP) media such that each flask contained as many as 250 unique clones of roughly equivalent expression. These cultures were allowed to grow for 48 h, and then genomic DNA was extracted with a protocol modified from that of Newman et al.[38] Ten milliliters of log-phase culture were concentrated by centrifugation at 3000 rpm and resuspended in 0.5 mL of TEN buffer (10 mM Tris-HCl, 10 mM EDTA, 150 mM NaCl). This solution was transferred to a 1.5 mL tube and spun at 10,000 rpm for 10 s, and the cell pellet was resuspended in 150 μL of H_2O and 300 μL of SDS-EB (2% SDS, 400 mM NaCl, 40 mM EDTA, 100 mM Tris-HCl, pH 8.0) and vortexed. Then 350 μL of 1:1 phenol/chloroform was added, and the mixture was vortexed again. Phases were separated by a 5 min 14,000 rpm spin, and the aqueous phase was transferred to a new tube, where was it again extracted by adding 300 μL of pure chloroform, vortexing, and spinning as above to separate phases. The aqueous phase was again transferred to a new tube, added to 2 vol of ice-cold 100% ethanol, and kept on ice for 30 min. Then the tubes were spun for 10 min at 14,000 rpm, and the pellet was washed with 200 μL of 70% ethanol before drying and resuspending in 100 μL of EB.

From these DNA preps, we amplified subpools for sequencing by PCR using Kapa Biosystems Library Amplification Kit (Kapa Biosystems Inc., Woburn, MA, USA), which reduces PCR-based bias especially with A/T-rich templates.[39,40] Primers for this step had internal barcodes where the forward primer barcode designates the gene origin of the UTR, and the reverse barcode designates the pool (low, medium, high expression) from which the sample originated (see barcoding

primers in Supplementary Table S3). Standard Illumina adapters were ligated onto the PCR fragments according to Illumina product literature on Paired-End Sample Preparation, and all samples were sequenced on a single lane of an Illumina GAIIx (Illumina Inc., San Diego, CA, USA) at The Scripps Research Institute DNA Array Core using paired-end 150 bp reads. Because the nucleotide chain growth is blocked after the addition of each nucleotide, the Illumina platform is particularly well-suited for sequences with long homopolymer regions, compared to other next-generation sequencing technologies. Though some slippage in sequencing may still have occurred, we mitigated this risk with the paired-end reads: these allowed us to read a single sequence from both ends and determine the consensus sequence by aligning both reads.

Reads were analyzed in CLC Genomics Workbench (CLC bio, Aarhus, Denmark). First, subpools were isolated on the basis of the pairs of forward and reverse internal barcodes, and then reads were filtered and trimmed by quality. Reads were removed if any one of the following criteria were met: if P_{error} exceeded 0.05, if there were more than two ambiguous nucleotides, or if read length was less than 64 nucleotides. In cases where one read of a pair did not meet these stringency cutoffs, the remaining read was mapped to the reference sequences as a single read.

Trimmed reads were mapped to a set of reference sequences of the designed oligonucleotides synthesized on the Agilent OLS platform. Mapping tolerated zero mismatch along the full length of the UTR, and nonspecific mappings were ignored, such that all mapped reads unambiguously map to a single unique reference sequence. All subsequent positional mutation likelihood analysis was done in Microsoft Excel after importing read counts from each pool for each reference UTR sequence. Each reference was parsed into individual residues, and then references where the wild-type nucleotide was replaced by an adenosine were assigned a value of 1 at that position and a value of 0 at all positions identical to wild type. The resulting matrices of zeros and ones for each reference UTR were multiplied by the number of times a read mapped to that reference from each expression pool, producing a matrix of aggregate data on how frequently a given position was mutated to adenosine for the high, low, or intermediate expression lines. The data presented in Figures 3, 5, 6, and 8 are essentially inverse of these aggregate matrices: the graphs depict the likelihood that the wild-type sequence was conserved at a given position, rather than the likelihood that the position was mutated to adenosine. To determine statistical significance of these positional likelihoods among each of the three expression level pools, a 3 × 2 contingency table was constructed for each position, delineating the aggregate wild type and "A" read counts for each pool. Pearson's χ-square tests indicate that even at positions exhibiting the most similar % conservation of wild type sequence for all expression pools (i.e., the points where the plots most nearly overlap in Figures 3 and 6), the p-value approaches zero ($p \ll 10^{-6}$), indicating that the differences observed between the pools are highly significant at every position along both of these UTRs.

Synthesizing and Cloning the Synthetic UTR. The synthetic UTR was made by ordering one oligonucleotide primer (see primer LS191 in Supplementary Table S3) containing all the desired nucleotide substitutions (all nucleotides in the negative regulatory regions N1 and N2 substituted with adenosines) to PCR-amplify using the wild-type *psaA* UTR as template. This fragment was cloned into the luciferase expression vector previously described and simultaneously into an identical chloroplast transformation vector except with a chloroplast codon-optimized β-xylanase gene from *Trichoderma reesei*. Another fragment containing the wild-type *psaA* UTR was cloned into these same vectors, for comparison. The vectors were amplified in three fragments with uracil-containing primers for USER fusion cloning with the synthetic UTR fragment (see Supplementary Table S3 for primer sequences). All fragments were gel-purified and mixed in equimolar amounts with 1 μL of the USER enzyme in a 10 μL reaction and incubated as described above. These plasmids were transformed into the *psbH*- strain of *C. reinhardtii* and restreaked to homoplasmy on minimal media. Four homoplasmic lines of each, as verified by a homoplasmic PCR screen as described[41] (see primers in Supplementary Table S3), were assayed for luciferase expression using the same equal-cell spotting assay described above and for xylanase activity.

Xylanase Assays. The four homoplasmic lines of each construct were inoculated into 50 mL flasks of TAP media and allowed to grow to mid-log phase. Cells were harvested at 3,000 rpm, resuspended in 1 mL of lysis buffer (TBS plus 0.5% Tween-20), and sonicated for two intervals of 10 s each at 20% amplitude (S450D digital sonifier, Branson) on ice. Cellular debris was separated from the soluble fraction by centrifugation at 14,000 rpm at 4 °C for 20 min. Next 400 μL of the soluble lysate was transferred to a new tube, and a Bio-Rad DC Protein Assay was performed as described by the manufacturer (Bio-Rad). This quantitation of total soluble protein was used to normalize the volume of lysate used in the subsequent xylanase assays, such that each sample contained equal amounts of total protein.

The xylanase assays were performed according to the manufacturer using the EnzChek Ultra Xylanase assay kit (Invitrogen) in 96-well black flat-bottom plates and read on an Infinite 200 Pro platereader (Tecan, Männedorf, Switzerland) at 355 nm excitation and 455 nm emission. The temperature was held at 42 °C, and readings were taken every 2 min until the increase in fluorescence was no longer linear. In the linear range, the strength of xylanase expression was calculated as the increase in relative fluorescence units (RFU) per minute. Each homoplasmic line was analyzed in triplicate; the average of all four homoplasmic lines is represented in Figure 9 for each construct.

Quantitative PCR. From cultures grown to mid-log phase, RNA was extracted using the standard protocol for the Plant RNA Reagent (Invitrogen) and eluted at a concentration of approximately 500 ng/μL in nuclease-free water. Complementary DNA (cDNA) was synthesized using the Verso cDNA Kit (Thermo Scientific) as described by the manufacturer. This template was diluted 1:10 in nuclease-free water before adding to the qPCR reactions. qPCRs were all performed in triplicate, using a 126 bp amplicon in the luciferase gene and a 139 bp amplicon in the chlorophyll B gene as a control (see primers in Supplementary Table S3). SsoFast EvaGreen Supermix (Bio-Rad) was used to perform the qPCR, according to the manufacturer's recommendations for cDNA templates. The efficiency of the luciferase primer set was calculated to be >97% at an annealing temperature of 54.8 °C across serial dilutions from 1:5 to 1:625. Under the same conditions, the chlorophyll B primer set also gave an efficiency >97%, and melt curves reveal single products devoid of off-target amplification for both primer sets, so calculations for fold change in transcript level have assumed perfect exponential amplification.

■ ASSOCIATED CONTENT

Supporting Information

Sequences of all synthesized variants of the *psbD* and *psaA* UTRs and sequences of all primers used in this work. This material is available free of charge via the Internet at http://pubs.acs.org.

■ AUTHOR INFORMATION

Corresponding Author

*E-mail: smayfield@ucsd.edu.

Notes

The authors declare no competing financial interest.

■ ACKNOWLEDGMENTS

The authors would like to thank Agilent Technologies for their support in constructing the UTR libraries, and David Masica for his assistance writing code to generate and filter the mutagenized reference sequences. We gratefully acknowledge Kevin Hoang for his help patching, screening, and maintaining the transformant libraries and for his assistance with the genomic DNA preparations for sequencing. We also express many thanks to Steve Head and Lana Schaffer at The Scripps Research Institute DNA Array Core facility for their assistance with the Illumina sequencing and data analysis. Ryan Georgianna and Hussam Auis graciously reviewed the manuscript and provided valuable feedback, and we thank Vineet Bafna for his input on the data analysis and statistical methods. This work was supported by the U.S. Department of Energy (DE-EE0003373) and California Energy Commission (CILMSF #500-10-039). E.S. is supported by a National Science Foundation Graduate Research Fellowship.

■ ABBREVIATIONS

HSM, high salt minimal media; OLS, oligonucleotide library synthesis; RBS, ribosome binding site; SD, Shine-Dalgarno; TAP, tris acetate phosphate media; USER, uracil-specific excision reagent; UTR, untranslated region

■ REFERENCES

(1) Surzycki, R., Greenham, K., Kitayama, K., Dibal, F., Wagner, R., Rochaix, J. D., Ajam, T., and Surzycki, S. (2009) Factors effecting expression of vaccines in microalgae. *Biologicals 37*, 133−138.

(2) Specht, E., Miyake-Stoner, S., and Mayfield, S. (2010) Micro-algae come of age as a platform for recombinant protein production. *Biotechnol. Lett. 32*, 1373−1383.

(3) Blowers, A. D., Klein, U., Ellmore, G. S., and Bogorad, L. (1993) Functional in vivo analyses of the 3′ flanking sequences of the Chlamydomonas chloroplast Rbcl and Psab genes. *Mol. Gen. Genet. 238*, 339−349.

(4) Barnes, D., Franklin, S., Schultz, J., Henry, R., Brown, E., Coragliotti, A., and Mayfield, S. (2005) Contribution of 5′- and 3′-untranslated regions of plastid mRNAs to the expression of Chlamydomonas reinhardtii chloroplast genes. *Mol. Genet. Genomics 274*, 625−636.

(5) Kiaulehn, S., Voytsekh, O., Fuhrmann, M., and Mittag, M. (2007) The presence of UG-repeat sequences in the 3 ′-UTRs of reporter luciferase mRNAs mediates circadian expression and can determine acrophase in Chlamydomonas reinhardtii. *J. Biol. Rhythms 22*, 275−277.

(6) Yohn, C. B., Cohen, A., Danon, A., and Mayfield, S. P. (1996) Altered mRNA binding activity and decreased translational initiation in a nuclear mutant lacking translation of the chloroplast psbA mRNA. *Mol. Cell. Biol. 16*, 3560−3566.

(7) Mayfield, S. P., Cohen, A., Danon, A., and Yohn, C. B. (1994) Translation of the psbA mRNA of Chlamydomonas reinhardtii requires a structured RNA element contained within the 5′ untranslated region. *J. Cell Biol. 127*, 1537−1545.

(8) Boudreau, E., Nickelsen, J., Lemaire, S. D., Ossenbuhl, F., and Rochaix, J.-D. (2000) The Nac2 gene of Chlamydomonas encodes a chloroplast TPR-like protein involved in psbD mRNA stability. *EMBO J. 19*, 3366−3376.

(9) Higgs, D. C., Shapiro, R. S., Kindle, K. L., and Stern, D. B. (1999) Small cis-acting sequences that specify secondary structures in a chloroplast mRNA are essential for RNA stability and translation. *Mol. Cell. Biol. 19*, 8479−8491.

(10) Anthonisen, I. L., Salvador, M. L., and Klein, U. (2001) Specific sequence elements in the 5′ untranslated regions of rbcL and atpB gene mRNAs stabilize transcripts in the chloroplast of Chlamydomonas reinhardtii. *RNA 7*, 1024−1033.

(11) Nickelsen, J., Fleischmann, M., Boudreau, E., Rahire, M., and Rochaix, J. D. (1999) Identification of cis-acting RNA leader elements required for chloroplast psbD gene expression in chlamydomonas. *Plant Cell 11*, 957−970.

(12) Schwarz, C., Elles, I., Kortmann, J., Piotrowski, M., and Nickelsena, J. (2007) Synthesis of the D2 protein of photosystem II in Chlamydomonas is controlled by a high molecular mass complex containing the RNA stabilization factor Nac2 and the translational activator RBP40. *Plant Cell 19*, 3627−3639.

(13) Klinkert, B., Elles, I., and Nickelsen, J. (2006) Translation of chloroplast psbD mRNA in Chlamydomonas is controlled by a secondary RNA structure blocking the AUG start codon. *Nucleic Acids Res. 34*, 386−394.

(14) Wobbe, L., Schwarz, C., Nickelsen, J., and Kruse, O. (2008) Translational control of photosynthetic gene expression in phototrophic eukaryotes. *Physiol. Plant 133*, 507−515.

(15) Nour-Eldin, H. H., Hansen, B. G., Norholm, M. H., Jensen, J. K., and Halkier, B. A. (2006) Advancing uracil-excision based cloning towards an ideal technique for cloning PCR fragments. *Nucleic Acids Res. 34*, e122.

(16) Mayfield, S. P., and Schultz, J. (2004) Development of a luciferase reporter gene, luxCt, for Chlamydomonas reinhardtii chloroplast. *Plant J. 37*, 449−458.

(17) Bateman, J. M., and Purton, S. (2000) Tools for chloroplast transformation in Chlamydomonas: expression vectors and a new dominant selectable marker. *Mol. Gen. Genet. 263*, 404−410.

(18) Kuck, U., Choquet, Y., Schneider, M., Dron, M., and Bennoun, P. (1987) Structural and transcription analysis of two homologous genes for the P700 chlorophyll a-apoproteins in Chlamydomonas reinhardii: evidence for in vivo trans-splicing. *EMBO J. 6*, 2185−2195.

(19) Wostrikoff, K., Girard-Bascou, J., Wollman, F. A., and Choquet, Y. (2004) Biogenesis of PSI involves a cascade of translational autoregulation in the chloroplast of Chlamydomonas. *EMBO J. 23*, 2696−2705.

(20) Yohn, C. B., Cohen, A., Danon, A., and Mayfield, S. P. (1998) A poly(A) binding protein functions in the chloroplast as a message-specific translation factor. *Proc. Natl. Acad. Sci. U.S.A. 95*, 2238−2243.

(21) Franzetti, B., Carol, P., and Mache, R. (1992) Characterization and RNA-binding properties of a chloroplast S1-like ribosomal protein. *J. Biol. Chem. 267*, 19075−19081.

(22) Zuker, M. (2003) Mfold web server for nucleic acid folding and hybridization prediction. *Nucleic Acids Res. 31*, 3406−3415.

(23) Nickelsen, J., van Dillewijn, J., Rahire, M., and Rochaix, J. D. (1994) Determinants for stability of the chloroplast psbD RNA are located within its short leader region in Chlamydomonas reinhardtii. *EMBO J. 13*, 3182−3191.

(24) Pfalz, J., Bayraktar, O. A., Prikryl, J., and Barkan, A. (2009) Site-specific binding of a PPR protein defines and stabilizes 5′ and 3′ mRNA termini in chloroplasts. *EMBO J. 28*, 2042−2052.

(25) Ossenbuhl, F., and Nickelsen, J. (2000) cis- and trans-Acting determinants for translation of psbD mRNA in Chlamydomonas reinhardtii. *Mol. Cell. Biol. 20*, 8134−8142.

(26) Fargo, D. C., Zhang, M., Gillham, N. W., and Boynton, J. E. (1998) Shine-Dalgarno-like sequences are not required for translation of chloroplast mRNAs in Chlamydomonas reinhardtii chloroplasts or in Escherichia coli. *Mol. Gen. Genet. 257*, 271−282.
(27) Dron, M., Rahire, M., and Rochaix, J. D. (1982) Sequence of the chloroplast 16S rRNA gene and its surrounding regions of Chlamydomonas reinhardii. *Nucleic Acids Res. 10*, 7609−7620.
(28) Harris, E. H., Boynton, J. E., and Gillham, N. W. (1994) Chloroplast ribosomes and protein synthesis. *Microbiol. Rev. 58*, 700−754.
(29) Liu, X. Q., Gillham, N. W., and Boynton, J. E. (1989) Chloroplast ribosomal protein gene rps12 of Chlamydomonas reinhardtii. Wild-type sequence, mutation to streptomycin resistance and dependence, and function in Escherichia coli. *J. Biol. Chem. 264*, 16100−16108.
(30) Michelet, L., Lefebvre-Legendre, L., Burr, S. E., Rochaix, J. D., and Goldschmidt-Clermont, M. (2011) Enhanced chloroplast transgene expression in a nuclear mutant of Chlamydomonas. *Plant Biotechnol. J. 9*, 565−574.
(31) Torronen, A., Mach, R. L., Messner, R., Gonzalez, R., Kalkkinen, N., Harkki, A., and Kubicek, C. P. (1992) The two major xylanases from Trichoderma reesei: characterization of both enzymes and genes. *Biotechnology (N. Y.) 10*, 1461−1465.
(32) Subramaniyan, S., and Prema, P. (2002) Biotechnology of microbial xylanases: enzymology, molecular biology, and application. *Crit. Rev. Biotechnol. 22*, 33−64.
(33) Zerges, W., and Rochaix, J. D. (1994) The 5′ leader of a chloroplast mRNA mediates the translational requirements for two nucleus-encoded functions in Chlamydomonas reinhardtii. *Mol. Cell. Biol. 14*, 5268−5277.
(34) Barkan, A., Rojas, M., Fujii, S., Yap, A., Chong, Y. S., Bond, C. S., and Small, I. (2012) A combinatorial amino Acid code for RNA recognition by pentatricopeptide repeat proteins. *PLoS Genet. 8*, e1002910.
(35) Kosuri, S., Eroshenko, N., LeProust, E. M., Super, M., Way, J., Li, J. B., and Church, G. M. (2010) Scalable gene synthesis by selective amplification of DNA pools from high-fidelity microchips. *Nat. Biotechnol. 28*, 1295−U1108.
(36) LeProust, E. M., Peck, B. J., Spirin, K., McCuen, H. B., Moore, B., Namsaraev, E., and Caruthers, M. H. (2010) Synthesis of high-quality libraries of long (150mer) oligonucleotides by a novel depurination controlled process. *Nucleic Acids Res. 38*, 2522−2540.
(37) Kindle, K. L., Richards, K. L., and Stern, D. B. (1991) Engineering the chloroplast genome − Techniques and capabilities for chloroplast transformation in Chlamydomonas reinhardtii. *Proc. Natl. Acad. Sci. U.S.A. 88*, 1721−1725.
(38) Newman, S. M., Boynton, J. E., Gillham, N. W., Randolph-Anderson, B. L., Johnson, A. M., and Harris, E. H. (1990) Transformation of chloroplast ribosomal RNA genes in Chlamydomonas: molecular and genetic characterization of integration events. *Genetics 126*, 875−888.
(39) Quail, M. A., Otto, T. D., Gu, Y., Harris, S. R., Skelly, T. F., McQuillan, J. A., Swerdlow, H. P., and Oyola, S. O. (2011) Optimal enzymes for amplifying sequencing libraries. *Nat. Methods 9*, 10−11.
(40) Oyola, S. O., Otto, T. D., Gu, Y., Maslen, G., Manske, M., Campino, S., Turner, D. J., Maclnnis, B., Kwiatkowski, D. P., Swerdlow, H. P., and Quail, M. A. (2012) Optimizing Illumina next-generation sequencing library preparation for extremely AT-biased genomes. *BMC Genomics 13*, 1.
(41) Rasala, B. A., Muto, M., Lee, P. A., Jager, M., Cardoso, R. M., Behnke, C. A., Kirk, P., Hokanson, C. A., Crea, R., Mendez, M., and Mayfield, S. P. (2010) Production of therapeutic proteins in algae, analysis of expression of seven human proteins in the chloroplast of Chlamydomonas reinhardtii. *Plant Biotechnol. J. 8*, 719−733.

Acknowledgements

Chapter 4, in full, is a reprint of the material as it appears in *American Chemical Society Synthetic Biology* 2013. The dissertation author was the primary investigator and author of this paper. Reprinted with permission from *American Chemical Society Synthetic Biology*, copyright 2013 American Chemical Society.

CHAPTER 5:

AN ENHANCED ARS2-DERIVED NUCLEAR REPORTER IN *CHLAMYDOMONAS*

Abstract

The model alga *Chlamydomonas reinhardtii* has been used to pioneer genetic engineering techniques for high-value protein and biofuel production from algae. To date, most studies of transgenic *Chlamydomonas* have utilized the chloroplast genome, which exhibits targeted engineering, high transgene expression, and a sizeable suite of reporters and well-characterized expression constructs. The advanced manipulation of algal nuclear genomes has been hampered by limited strong expression cassettes and a lack of high-throughput reporters. We have improved upon an endogenous reporter gene – the ARS2 gene encoding an arylsulfatase enzyme – that was first cloned and characterized decades ago but has not been used extensively. The new construct, derived from ARS2 cDNA, expresses significantly higher levels of reporter protein and transforms more efficiently, allowing qualitative and quantitative screening using a rapid, inexpensive 96-well assay. We used this construct to screen a new transgene promoter from the ARG7 gene, and found that the ARG7 promoter can express the ARS2 reporter as strongly as the HSP70-RBCS2 chimeric promoter that currently ranks as the best available promoter, thus adding to the list of useful nuclear promoters. These tools enhance the efficiency and ease of genetic engineering within the *Chlamydomonas* nuclear genome, with potential application to other algal strains.

Practical Applications

The ARS2 reporter provides a high-throughput, inexpensive, and sensitive nuclear reporter in *Chlamydomonas reinhardtii*. A reporter construct using the ARS2 gene has been developed that is smaller, more efficiently transformed, and more highly expressed than previous ARS2 reporters. In contrast to the current publicly available ARS2 reporter construct, co-transformation is not required as a resistance marker has been included in the vector. In this study, the improved ARS2 reporter gene is used to characterize the ARG7 promoter, which has

not previously been explored for transgene expression. Thus, we introduce an improved robust nuclear reporter for nuclear engineering, as well as characterize a novel promoter for expressing transgenes. The improved ARS2 reporter is also amenable for other applications, such as screening *C. reinhardtii* mutants for reduced transgene silencing. Two ARS2 reporter constructs are generated in this study, and have been fully characterized and made available to the algal research community.

Introduction

Unicellular eukaryotic algae are ideal models for studying cellular processes or characteristics due to their fast growth, ease of culturing, efficient transformation, and short timescale for obtaining transformants. They are also amenable to a number of high-throughput analysis techniques such as fluorescence-activated cell sorting (FACS) and 96-well assays. Recently, microalgae have garnered attention as a protein production platform for obtaining properly folded, complex high-value proteins such as vaccine antigens (*1*), human antibody-drug conjugates (*2*), and many other therapeutic proteins (*3*). They also hold potential for biofuel production, but will likely require complex metabolic engineering in parallel with stringent strain selection to achieve robust, high-yielding strains approaching economic parity with fossil fuel sources (*4*). Due to its long history of study and robust transformation methods, the green microalga *Chlamydomonas reinhardtii* has emerged as the leading model organism to explore the potential of molecular farming (*5, 6*) and engineered algal strains for biofuels.

To date, much of the genetic engineering of *Chlamydomonas* has been achieved in the chloroplast genome, where transgene expression levels up to 10% of total soluble protein (*7*) have been reported. However, proteins synthesized in the chloroplast cannot be exported from the organelle, limiting its utility for metabolic engineering that requires coordinated action in the cytosol or other intracellular compartments. While transformation of the nuclear genome is

routine, high levels of transgene expression from this genome remains problematic for several reasons. First, foreign DNA integrates into the nuclear genome by non-homologous end-joining (NHEJ), creating high variability among transformants; furthermore, the level of expression achievable from nuclear-integrated transgenes is far below that observed in the chloroplast, likely due to epigenetic and RNA-mediated gene silencing mechanisms (*8, 9*), and finally there are a limited number of characterized nuclear promoters available for transgene expression, and none of these achieve the levels of transcription that are routinely achieved in other eukaryotic hosts.

Several promoters are currently used for nuclear expression in *C. reinhardtii* (*10*). However, only one – the heat shock protein 70 - rubisco small subunit chimeric promoter, HSP70-RBCS2 (*11*) – is suitable for high heterologous expression of recombinant genes. The remaining promoters are primarily used to drive expression of selectable marker genes, and in general cannot drive expression of genes of interest to high levels. One approach for remedying low expression of heterologous proteins has been pioneered, utilizing the foot-and-mouth-disease-virus 2A self-cleaving peptide to transcriptionally fuse heterologous gene expression to antibiotic resistance (*12*). While multiple proteins can be translated from the same transcript with this approach (*13*), there are indications that increasing the number of proteins leads to imbalanced expression levels (*14*), rendering this system undesirable for inserting multiple genes (for example, a metabolic pathway). While many promoters have been screened systemically for transgene expression within the chloroplast (*15*), such systematic approaches have seldom been applied to characterize constitutive nuclear promoters (*16*). A nuclear reporter amenable to inexpensive, high-throughput screening is essential for identifying novel promoter candidates to expand the genetic toolkit for more complex nuclear genetic engineering.

A handful of reporter genes have been codon-optimized for expression in the *Chlamydomonas* nuclear genome, including luciferase (*17, 18*), green fluorescent protein (*19*), and xylanase (*12*), and recently several additional fluorescent reporter proteins have been

characterized (*20*). However, even with these optimized reporters the signal is typically low and these proteins require sophisticated and expensive equipment to measure, such as a plate-reader or a fluorescence microscope with appropriate filter sets. For laboratories without access to this equipment, an arylsulfatase assay is far more suitable; the cost of reagents for a 96-well plate is less than $1, and no specialized equipment is necessary for qualitative detection.

The ARS2 gene in *C. reinhardtii* encodes a periplasmic arylsulfatase (*21*); the endogenous gene is only expressed under sulfur starvation, and its enzymatic activity can be detected by a simple and sensitive colorimetric assay. Moreover, a secretion signal allows detectable enzymatic activity in culture media, obviating cell lysis prior to assay even for strains with an intact cell wall. Thus, ARS2 has the prerequisites to become a robust quantifiable reporter gene for high-throughput screening. Previously, the publicly available ARS2 reporter has been available only as a large genomic clone with no resistance marker included in the vector; transformants were typically isolated by co-transformation with another plasmid carrying a selectable marker (*22*).

In this study we improve the ARS2 reporter by making the construct smaller, more efficiently transformed, and more strongly expressed by cloning the cDNA and optimizing the flanking regulatory elements. We demonstrate that this construct can be used to rapidly characterize promoter strength and robustness, and we use it to determine that the ARG7 promoter produces as much ARS protein as the HSP70-RBCS2 chimeric promoter yet is less susceptible to positional effects.

Materials and Methods

Strains and growth conditions

Algal strains cc1819, cc1820, cc1826, and cc1010, and plasmid pJD100 were obtained from the *Chlamydomonas* Resource Center (University of Minnesota). Cultures were grown in

constant light in tris-acetate-phosphate media (TAP); arginine auxotrophs were supplemented with 100 mg/L L-arginine (Sigma, St. Louis, MO). For sulfur starvation, $MgSO_4$, $CuSO_4$, $FeSO_4$, and $ZnSO_4$ were replaced with molar equivalents of the respective chlorides.

Cloning of ARS2 cDNA

RNA was isolated using the RNA Bee reagent (Amsbio) from a late-log culture ($5x10^6$ cells/ml) grown for 24 hours in sulfur-free TAP. cDNA was generated using the RevertAid Reverse Transcriptase kit (Fermentas) according to the manufacturer's instructions for GC-rich templates using oligo$(dT)_{18}$ primers. The PfuX7 polymerase (*23*) was used with the Phusion GC buffer (New England Biolabs) and 4% DMSO to amplify the ARS2 CDS and 3' UTR.

Plasmid Construction

Full length cDNA for ARS2 was assembled seamlessly from smaller fragments via USER fusion (*24*). ARS2 was amplified in two fragments: the 5' half of the coding sequence, and the 3' half of the coding sequence plus the 3' UTR. The ARG7 promoter and 5'UTR were amplified from cc1010 genomic DNA (see Supporting Information Table S1 for all primers). Fragments were assembled into the USER cloning-compatible vector Hcr1 (pBluescript into which the USER cloning cassette (GCTGAGGGTTTAATATTAAGACCTCAGC) has been inserted as previously described (*24*)) by USER fusion (*25*), using 100 ng of vector and equimolar amounts of each fragment. To insert the paromomycin resistance cassette into pJD100, a *Kpn*I-excised cassette was ligated into an existing *Kpn*I site upstream of the ARS2 cassette in pJD100. All plasmids (pHR15, pHR16, and P-pJD100) and sequence maps are available from the *Chlamydomonas* Resource Center (chlamycollection.org).

Transformation conditions

Strains were grown to mid log phase (3-5x10^6 cells/ml), harvested by centrifugation, and resuspended at 3x10^8 cells/ml in *Chlamydomonas* MaxEfficiency transformation reagent (Life Technologies). 250 µl of cells were added to 4mm cuvettes, along with 1 µg of plasmid linearized in the plasmid backbone using *NotI* (pHR15 and pHR16) or *BsaI* (P-pJD100). Cuvettes were incubated at room temperature for 5 min, then on ice for 5 min, and electroporated in a Gene Pulser (Bio-Rad) with the following settings: 500 V, 50 µF capacitance, 800 Ω resistance. Cuvettes were recovered at room temperature for 5 min, then transferred to TAP with 40mM sucrose for 14-18 hours, resuspended in 600 µl TAP, and plated onto selection.

Arylsulfatase assays

Colonies were inoculated into deep-well plates (Axygen Scientific) containing 0.5 ml TAP. Cultures were grown for 3-4 days to saturated density and centrifuged for 5 min at 2000 rpm. 100 µl of media was transferred to a flat-bottom clear plate (Corning). To each well, 25 µl of N-SO_4 reaction buffer (2M glycine-NaOH, pH 9, 50mM imidazole, 4mM α-naphthyl sulfate (Sigma-Aldrich)) was added, then incubated at 37°C for 1hr. Then 125 µl stopping buffer (0.2M sodium acetate, pH 4.8, 4% SDS) was added, followed by 25 µl of 10 mg/ml Fast Blue B Salt (Sigma-Aldrich). The color was allowed to develop for 1 min and read using an Infinite 200 Pro platereader (Tecan) for absorbance at 540nm. Three deep-well plates of each construct were analyzed, from separate transformations. The wild type control was normalized across all plates to reduce variability in overall intensity between plates. Experimental wells were then expressed as percentage of maximum signal.

96-well lysis of C. reinhardtii *by sonication for qualitative analysis*

In deep-well 96-well culture plates, cultures in stationary phase were allowed to settle to the bottom. To each well, 2 glass beads (1.5-3mm in diameter) were added and the culture plate was sonicated in an ice-cold sonicator bath (Branson 2510, Branson Ultrasonics) for 7 minutes.

Results

ARS2 cDNA constructs transform more efficiently than genomic constructs

Three new ARS reporter constructs were generated as a means to improve the ARS2 reporter system and to investigate the strength of ARG7 promoter for transgene expression. The genomic ARS2 clone pJD100 (*26*) does not contain a selectable marker, so transformants can only be obtained by co-transformation with another plasmid containing a selectable gene. The first improvement consisted of inserting a paromomycin resistance cassette into pJD100; this vector is called P-pJD100 and it eliminates the need for co-transformation.

The second improvement made to the ARS2 reporter system consisted of a novel construct called pHR15. In this vector, the coding sequence for ARS2 was inserted under the control of the HSP70-RBCS2 chimeric promoter/5'UTR and the endogenous ARS2 3'UTR. Additionally, the pHR16 plasmid was generated; pHR16 is identical to pHR15 but has the HSP70/RBCS2 promoter/5'UTR exchanged with 1425 bp upstream of the ARG7 gene in the nuclear genome, functioning as the promoter and 5' UTR (Figure 5.1). Similar to the P-pJD100 vector, both pHR15 and pHR16 utilize a paromomycin resistance cassette for selection. Compared to P-pJD100, pHR15 and pHR16 resulted in approximately two and three fold higher transformation efficiencies per pmol DNA, respectively (Table 5.1).

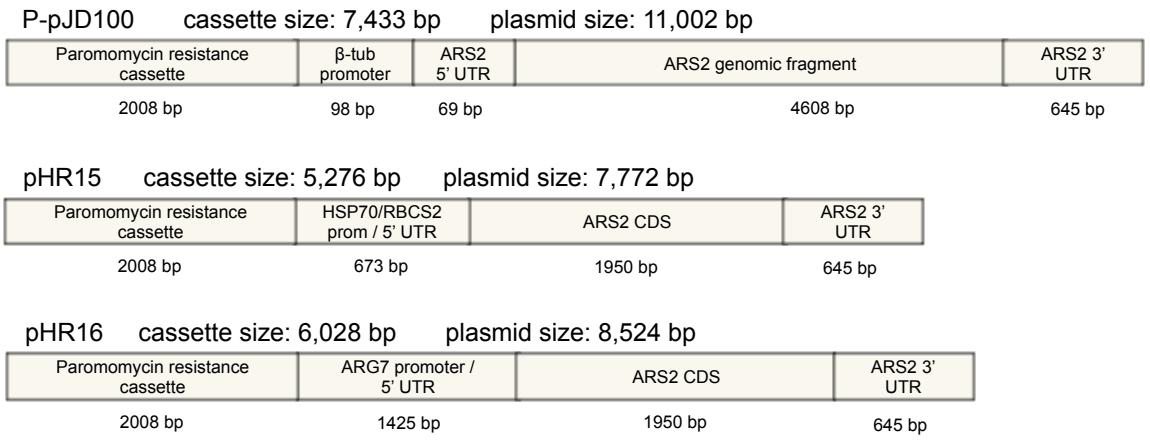


Figure 5.1 Schematic of constructs. P-pJD100 is identical to pJD100 except for the insertion of the paromomycin resistance cassette under the control of the *psaD* promoter and UTRs. pHR15 and pHR16 are identical except for the promoter and 5' UTR driving ARS2 expression, which are HSP70-RBCS2 and ARG7, respectively.

Table 5.1 Transformation efficiency of cDNA-derived ARS2 versus genomic ARS2 fragment.

Transformation vector	# colonies obtained per μg DNA	# colonies obtained per pmol DNA	Fold-change in per-pmol efficiency relative to P-pJD100
P-pJD100	1,127 ± 452	7,564 ± 3,034	-
pHR15	3,463 ± 867	16,412 ± 4,109	2.2
pHR16	4,550 ± 329	23,698 ± 1,714	3.1

ARS2 cDNA constructs express better than genomic constructs

ARS2 expression was determined for 285 transformants per plasmid via a colorimetric assay. These colorimetric assays can also be evaluated by eye; a representative 96 well plate for each plasmid is shown (Figure 5.2a-c). Qualitative visual examination indicates that both pHR15 and pHR16 produce substantially more arylsulfatase-expressing transformants than P-pJD100. Notably, the cell walled cc1010 wild type strain was used for all experiments and the assays were performed directly on culture media without cell lysis. To compare expression levels of ARS quantitatively, the assays were analyzed using a plate-reader. Background signal was normalized to zero for each plate, and the maximum normalized signal obtained across all three constructs defined the maximum arylsulfatase activity. Signal from individual transformants was then

binned into an expression level based on its signal relative to the maximum. The strains transformed with P-pJD100 predominantly occupied the lowest expression levels, and none of them exceeded 40% of the maximum expression observed (Figure 5.2d). pHR15 and pHR16 produced many transformants with several-fold higher arylsulfatase activity than the single best P-pJD100 transformant, and a much higher percentage of these transformants produced visible signal (Figure 5.2).

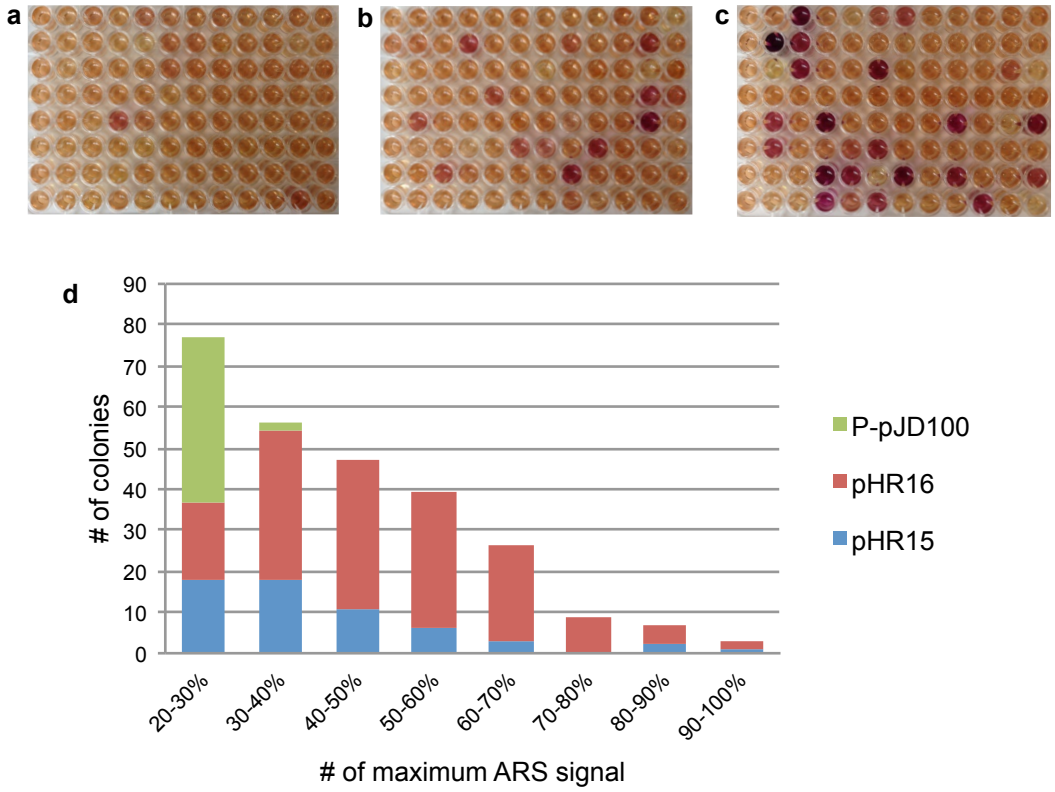


Figure 5.2 Relative strength and robustness of expression for P-pJD100, pHR15, and pHR16. Representative plates of colorimetric arylsulfatase assays for *C. reinhardtii* transformed with: **a**) P-pJD100; **b**) pHR15; **c**) pHR16. **d**) Quantification of colorimetric arylsulfatase assays. For each construct, 285 transformants were assayed and signals were binned into strength categories relative to the strongest signal obtained across all transformants. Only transformants that expressed at levels exceeding 20% of maximum expression are shown.

To investigate whether the cell wall of cc1010 inhibited export of ARS2 outside the periplasm into the media and thus reduce the sensitivity of the assay, ARS2 activity was

determined for 96 pHR15 transformants lysed by sonication to release cellular contents into the culture media. No difference was seen compared to the non-lysed plates (data not shown).

The cDNA-derived ARS2 reporter can be used to characterize novel promoters

More detailed comparison between pHR15 with the HSP70-RBCS2 promoter and pHR16 with the putative ARG7 promoter showed that they both give roughly equal numbers of transformants expressing at the highest level of expression (Figure 5.2d), so they are comparable in terms of expression level potential. However, pHR16 produced many more transformants at intermediate and intermediate-high levels of expression than pHR15, indicating that the ARG7 promoter is more robust with respect to positional effects and can generate more transformants at moderate to high expression levels.

Discussion

While single-celled algae can make use of a number of recently optimized fluorescent reporters (*27*) with FACS and plate-reader analysis, these reporters require expensive equipment that may not be readily available to all research labs. Our goal was to improve and expand upon reporters that can be used in any laboratory environment at low cost, to increase the accessibility of nuclear genome engineering in algae. The arylsulfatase screen is highly adaptable for use in labs with limited resources or equipment. If no strict quantitation is required, expression levels can easily be evaluated by eye to choose the top performers. It is also noteworthy that wild type controls never exhibit arylsulfatase activity even though the strain contains an intact ARS2 gene. This reflects the strong repression of the ARS2 promoter under normal culture conditions, making this an ideal reporter in any strain; an ARS2 knockout strain is not required.

The only vector previously available from the *Chlamydomonas* Resource Center containing the ARS2 gene was the vector pJD100, in which a 4.6 kb genomic fragment

containing ARS2 is driven by the beta-tubulin minimal promoter (*26*). The increased transformation efficiency per microgram of pHR15 and pHR16 compared to P-pJD100 was expected due to the reduced size of these plasmids (Figure 5.1). However, the increased transformation efficiency exceeds what would be expected from reduced plasmid size alone, as reflected by the per-picomole efficiencies (Table 5.1). Therefore there may be other factors affecting the transformation efficiency of these plasmids, or the correlation between plasmid size and propensity to enter the cell via electroporation is non-linear.

Improved expression of ARS2 was expected from pHR15, where the cDNA of ARS2 is controlled by the HSP70-RBCS2 promoter/5'UTR. This result confirms the superior strength of this chimeric regulatory promoter compared to the beta-tubulin promoter, which drives the genomic ARS2 fragment in the P-pJD100 plasmid (*28*). However, it should be noted that a direct comparison is not possible between P-pJD100 and pHR15 because part of the ARS2 5'UTR is included between the beta-tubulin promoter and the start codon in P-pJD100, and because P-pJD100 encodes the genomic fragment of ARS2. In comparison, in pHR15 the coding sequence of only the ARS2 cDNA has been placed under the control of the HSP70-RBCS2 promoter and 5'UTR. Close examination of the ARS2 5'UTR according to the Phytozome annotation (www.phytozome.net), shows that the original pJD100 is missing one fourth of the endogenous ARS2 5' UTR at the furthest 5' end. The complete ARS2 5'UTR has therefore not been characterized for constitutive expression; thus, it is possible that the 22 base pairs absent from the ARS2 5'UTR in P-pJD100 are required for message stability or translation initiation, which may help explain the low expression levels observed with the P-pJD100 construct.

While the genomic-derived version of the ARS2 gene has been used previously to detect promoter function, it was only used for inducible promoters that are strongly up-regulated under specific conditions (*29-31*). As shown in Figure 5.2a, the output from the beta-tubulin promoter is not sufficient to provide a robust range of signal for comparison with other constitutive

promoters. The cDNA-derived construct, on the other hand, can easily report differences in constitutive promoter strength. The ARG7 promoter appears as strong, and more robust to genomic context, or more resistant to nuclear silencing mechanisms, compared to the chimeric HSP70/RBCS2 promoter. This adds the ARG7 promoter to the limited list of strong promoters suitable for high transgene expression in the nuclear genome. As the minimum functional region has not been determined, it is possible that the ARG7 promoter could be further reduced in size while retaining its strength, further enhancing its desirability in transgenic constructs. There are potentially many more promoters suitable for heterologous expression yet to be identified, and the enhanced ARS2 construct provides an ideal platform for rapidly screening them for strength and robustness to further expand the algal nuclear genetic engineering toolkit.

In order for eukaryotic microalgae to continue to play a role in bioenergy production or recombinant bio-manufacturing, the ability to precisely engineer the nuclear genome is crucial. Nuclear genetic engineering thus far has been hampered by several factors, including a lack of molecular tools such as well-characterized promoters and reporters. Here, the endogenous *C. reinhardtii* arylsulfatase gene ARS2 has demonstrated utility for expanding the molecular toolkit by identifying additional promoters, and has potential for more applications such as screening strains for reduced nuclear silencing. Because genetic engineering advances pioneered in *C. reinhardtii* have now been applied to other algal species – including those with biofuel or industrial biotechnology potential – these tools may have broad significance for the algal engineering community at large.

Acknowledgements

We thank Dudley Page from Sabeeha Merchant's lab for the vector map and sequence for pJD100. This work was funded by the US Department of Energy, grant number DE-EE0003373, and the California Energy Commission Initiative for Large Molecule Sustainable Fuels,

agreement number 500-10-039. HHN-E was supported by the Danish Research Council for Technology and Production (FTP) grant 09-065827/274-08-0354, and EAS was supported by an NSF Graduate Research Fellowship.

Chapter 5, in full, is a reprint of the material as it has been submitted for publication to *Biotechnology Journal*. The dissertation author is the primary investigator and author of this paper.

References

1. J. A. Gregory, F. Li, L. M. Tomosada, C. J. Cox, A. B. Topol, J. M. Vinetz, S. Mayfield, Algae-produced pfs25 elicits antibodies that inhibit malaria transmission. *Plos One* **7**, e37179 (2012).

2. M. Tran, C. Van, D. J. Barrera, P. L. Pettersson, C. D. Peinado, J. Bui, S. P. Mayfield, Production of unique immunotoxin cancer therapeutics in algal chloroplasts. *Proc Natl Acad Sci U S A* **110**, E15 (Jan 2, 2013).

3. B. A. Rasala, M. Muto, P. A. Lee, M. Jager, R. M. Cardoso, C. A. Behnke, P. Kirk, C. A. Hokanson, R. Crea, M. Mendez, S. P. Mayfield, Production of therapeutic proteins in algae, analysis of expression of seven human proteins in the chloroplast of chlamydomonas reinhardtii. *Plant Biotechnol J* **8**, 719 (Aug, 2010).

4. D. R. Georgianna, S. P. Mayfield, Exploiting diversity and synthetic biology for the production of algal biofuels. *Nature* **488**, 329 (Aug 16, 2012).

5. S. Rosales-Mendoza, L. M. Paz-Maldonado, R. E. Soria-Guerra, Chlamydomonas reinhardtii as a viable platform for the production of recombinant proteins: Current status and perspectives. *Plant Cell Rep* **31**, 479 (Mar, 2012).

6. E. Specht, S. Miyake-Stoner, S. Mayfield, Micro-algae come of age as a platform for recombinant protein production. *Biotechnol Lett* **32**, 1373 (Oct, 2010).

7. A. L. Manuell, M. V. Beligni, J. H. Elder, D. T. Siefker, M. Tran, A. Weber, T. L. McDonald, S. P. Mayfield, Robust expression of a bioactive mammalian protein in chlamydomonas chloroplast. *Plant Biotechnol J* **5**, 402 (May, 2007).

8. H. Cerutti, A. M. Johnson, N. W. Gillham, J. E. Boynton, Epigenetic silencing of a foreign gene in nuclear transformants of chlamydomonas. *Plant Cell* **9**, 925 (Jun, 1997).

9. H. Cerutti, X. Ma, J. Msanne, T. Repas, Rna-mediated silencing in algae: Biological roles and tools for analysis of gene function. *Eukaryot Cell* **10**, 1164 (Sep, 2011).

10. C. a. K. Griesbeck, A., Ed., *Algae: An alternative to the higher plant system in gene farming*, (Springer, Netherlands, 2012), pp. 125-143.

11. M. Schroda, C. F. Beck, O. Vallon, Sequence elements within an hsp70 promoter counteract transcriptional transgene silencing in chlamydomonas. *Plant J.* **31**, 445 (Aug, 2002).

12. B. A. Rasala, P. A. Lee, Z. Shen, S. P. Briggs, M. Mendez, S. P. Mayfield, 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 (2012).

13. A. El Amrani, A. Barakate, B. M. Askari, X. J. Li, A. G. Roberts, M. D. Ryan, C. Halpin, Coordinate expression and independent subcellular targeting of multiple proteins from a single transgene. *Plant Physiology* **135**, 16 (May, 2004).

14. F. Geu-Flores, C. E. Olsen, B. A. Halkier, Towards engineering glucosinolates into non-cruciferous plants. *Planta* **229**, 261 (Jan, 2009).

15. B. A. Rasala, M. Muto, J. Sullivan, S. P. Mayfield, Improved heterologous protein expression in the chloroplast of chlamydomonas reinhardtii through promoter and 5' untranslated region optimization. *Plant biotechnology journal* **9**, 674 (Aug, 2011).

16. O. Ruecker, K. Zillner, R. Groebner-Ferreira, M. Heitzer, Gaussia-luciferase as a sensitive reporter gene for monitoring promoter activity in the nucleus of the green alga chlamydomonas reinhardtii. *Molecular Genetics and Genomics* **280**, 153 (Aug, 2008).

17. M. Fuhrmann, A. Hausherr, L. Ferbitz, T. Schodl, M. Heitzer, P. Hegemann, Monitoring dynamic expression of nuclear genes in chlamydomonas reinhardtii by using a synthetic luciferase reporter gene. *Plant molecular biology* **55**, 869 (Aug, 2004).

18. N. Shao, R. Bock, A codon-optimized luciferase from gaussia princeps facilitates the in vivo monitoring of gene expression in the model alga chlamydomonas reinhardtii. *Curr Genet* **53**, 381 (Jun, 2008).

19. M. Fuhrmann, W. Oertel, P. Hegemann, A synthetic gene coding for the green fluorescent protein (gfp) is a versatile reporter in chlamydomonas reinhardtii. *The Plant journal : for cell and molecular biology* **19**, 353 (Aug, 1999).

20. B. A. Rasala, D. J. Barrera, J. Ng, T. M. Plucinak, J. N. Rosenberg, D. P. Weeks, G. A. Oyler, T. C. Peterson, F. Haerizadeh, S. P. Mayfield, Expanding the spectral palette of fluorescent proteins for the green microalga chlamydomonas reinhardtii. *The Plant journal : for cell and molecular biology* **74**, 545 (May, 2013).

21. E. L. de Hostos, J. Schilling, A. R. Grossman, Structure and expression of the gene encoding the periplasmic arylsulfatase of chlamydomonas reinhardtii. *Mol Gen Genet* **218**, 229 (Aug, 1989).

22. J. P. Davies, D. P. Weeks, A. R. Grossman, Expression of the arylsulfatase gene from the beta 2-tubulin promoter in chlamydomonas reinhardtii. *Nucleic Acids Res* **20**, 2959 (Jun 25, 1992).

23. M. H. Norholm, A mutant pfu DNA polymerase designed for advanced uracil-excision DNA engineering. *BMC Biotechnol* **10**, 21 (2010).

24. H. H. Nour-Eldin, B. G. Hansen, M. H. Norholm, J. K. Jensen, B. A. Halkier, Advancing uracil-excision based cloning towards an ideal technique for cloning pcr fragments. *Nucleic Acids Res* **34**, e122 (2006).

25. F. Geu-Flores, H. H. Nour-Eldin, M. T. Nielsen, B. A. Halkier, User fusion: A rapid and efficient method for simultaneous fusion and cloning of multiple pcr products. *Nucleic Acids Res.* **35**, (Apr, 2007).

26. J. P. Davies, D. P. Weeks, A. R. Grossman, Expression of the arylsulfatase gene from the beta 2-tubulin promoter in chlamydomonas reinhardtii. *Nucleic Acids Res* **20**, 2959 (Jun 25, 1992).

27. B. A. Rasala, D. J. Barrera, J. Ng, T. M. Plucinak, J. N. Rosenberg, D. P. Weeks, G. A. Oyler, T. C. Peterson, F. Haerizadeh, S. P. Mayfield, Expanding the spectral palette of fluorescent proteins for the green microalga chlamydomonas reinhardtii. *Plant J*, (Mar 23, 2013).

28. M. Schroda, D. Blocker, C. F. Beck, The hsp70a promoter as a tool for the improved expression of transgenes in chlamydomonas. *Plant J* **21**, 121 (Jan, 2000).

29. M. Ohresser, R. F. Matagne, R. Loppes, Expression of the arylsulphatase reporter gene under the control of the nit1 promoter in chlamydomonas reinhardtii. *Curr Genet* **31**, 264 (Mar, 1997).

30. R. Loppes, M. Radoux, M. C. Ohresser, R. F. Matagne, Transcriptional regulation of the nia1 gene encoding nitrate reductase in chlamydomonas reinhardtii: Effects of various environmental factors on the expression of a reporter gene under the control of the nia1 promoter. *Plant molecular biology* **41**, 701 (Nov, 1999).

31. J. M. Quinn, P. Barraco, M. Eriksson, S. Merchant, Coordinate copper- and oxygen-responsive cyc6 and cpx1 expression in chlamydomonas is mediated by the same element. *Journal of Biological Chemistry* **275**, 6080 (Mar 3, 2000).

CHAPTER 6:

ARG7-BASED CASSETTE FOR HIGH-LEVEL NUCLEAR TRANSGENE EXPRESSION IN *CHLAMYDOMONAS REINHARDTII*

Abstract

Molecular tools for nuclear expression in *Chlamydomonas* are in need of improvement if algae are to truly emerge as the heterologous expression hosts of the future. Currently, the selection of transgene expression cassettes remains limited with some markers only available as large intractable genomic fragments. Here we improve the ARG7 nuclear marker by cloning its cDNA and showing superior transformation efficiencies to any existing ARG7-containing plasmid. The ability of the ARG7 promoter to drive expression of a reporter transgene that has been transcriptionally fused to the ARG7 cDNA is shown to be comparable to current state of the art expression cassettes. Serendipitously, we found that the arginine auxotroph strain CC1826 appear to be a superior host strain amenable to future studies. Finally, we show that placing the *Sh-ble* selection marker after the gene of interest results in dramatic increases in transgene expression level.

Introduction

Chlamydomonas reinhardtii has emerged, due to a number of reasons, as a leading model system to explore the potential of using micro-algae as low-cost hosts for producing recombinant proteins. Foremost, the suite of available molecular tools for modifying *C. reinhardtii* genetically is unparalleled in comparison to other micro-algae. However, expression of transgenes in the nuclear genome represents an area in need of development. Currently, seven different promoters are commonly used for nuclear expression in *C. reinhardtii* (*1*), however, only one (the Heat shock protein 70 - rubisco small subunit chimeric promoter, HSP70/RBCS2) is considered suitable for high heterologous expression, whereas the remaining promoters are mostly used to drive expression of selectable marker genes. For molecular pharming purposes transgenes are translationally fused to the strong Ble selection marker via the foot-and-mouth-disease-virus 2A self-cleavage peptide to ensure high expression (*2*). However, presence of antibiotic resistance

genes is not desirable in algal production strains that may be used directly for animal or human consumption. Our aim is to develop an alternate antibiotic resistance marker-free system, using an endogenous gene for selection, capable of driving transgene expression to levels comparable to those of the HSP70/RBCS2-ble-2A-GOI (gene of interest) system.

One of the important markers for nuclear transformation in *C. reinhardtii* is the endogenous and essential arginosuccinate lyase gene (ASL, EC 4.3.2.1). This enzyme performs the last step in the arginine biosynthesis pathway and rescues arginine requiring strains that have been mutated in the ARG7 locus (*3*). In the context of driving transgene expression, the ARG7 marker is notable because of a recent study where the ARG7 promoter alone conferred expression levels of arylsulfatase (ARS) comparable to those obtained with the HSP70/RBCS2 chimeric promoter (Specht *et al*., under review). This suggested that the ARG7 promoter in combination with the ASL gene could serve as a suitable driver for transgene expression in a translational fusion akin to the ble-2A constructs previously described.

The current publicly available pARG7.8 plasmid at the Chlamydomonas Resource Center contains a 7.8 kb genomic fragment, which has widely been used as a co-transformed selectable marker (*4-6*). In a previous attempt to improve the ARG7 marker, the cDNA fragment of the ASL gene was generated to simplify DNA manipulation procedures and increase transformation efficiency (*7*). Sequence analyses of the cloned cDNA revealed an extra exon and intron pair (*7*) upstream of the previously predicted 13 coding exons (*8*). Both cDNA versions – one containing all 14 exons and the other lacking the first exon complemented arginine auxotrophic strains but did not show transformation efficiencies significantly different from that of the much larger ARG7.8 genomic fragment (*7*). Notably, the cDNA versions were devoid of introns and had been placed under the control of an RBCS2 promoter, which was also devoid of introns (*7*).

In this study, we generate a transformation construct containing the ASL cDNA, which shows a superior transformation efficiency compared to the genomic pARG7.8 plasmid in several

arginine auxotrophic strains. Subsequently, using mCherry as a reporter, we use the HSP70/RBCS2 promoter-ble-2A-Gene of interest (GOI) system as a benchmark to show that ARG7 promoter-ASL-2A-GOI provides an alternate system to drive high-level transgene expression. Furthermore, we test whether transgene expression is influenced by the arrangement of the selection markers and GOI, further characterizing optimal alternative constructs for high-level transgene expression using transcriptional fusions.

Results

The cDNA-derived ARG7 marker exhibits higher transformation efficiency

The large size (nearly 8 kb) of the publicly available ASL marker complicates its inclusion in complex expression vectors and hampers transformation efficiency. We desired a smaller fragment to limit overall plasmid size. The cDNA sequence of ASL has been shown experimentally to contain 14 exons of which the first exon is suggested to be in the 5'UTR and hence non-coding (*7*)(*7*). We cloned the last 13 exons of the ASL gene from wild type *C. reinhardtii* cDNA. The resulting 1422 bp ASL coding sequence was placed under the control of its own 1425 bp upstream genomic fragment (identical to that controlling the expression of the ARG7 marker in the pARG7.8 plasmid and henceforth called the ARG7 promoter) to generate plasmid pHR11 (Figure 6.1).

The transformation efficiency of pHR11 was tested compared to that of pARG7.8 in three different arginine auxotrophic strains: CC1819, CC1820 and CC1826. For all three strains, colonies grew on normal TAP media due to complementation by either pHR11 or pARG7.8. In all three strains, pHR11 provided far greater numbers of colonies per μg DNA (Table 6.1). In CC1826, in particular, ~40 times more colonies were generated by pHR11 compared to pARG7.8. In comparison, pHR11 produced approximately 7- and 10-fold more colonies in CC1819 and CC1820, respectively.

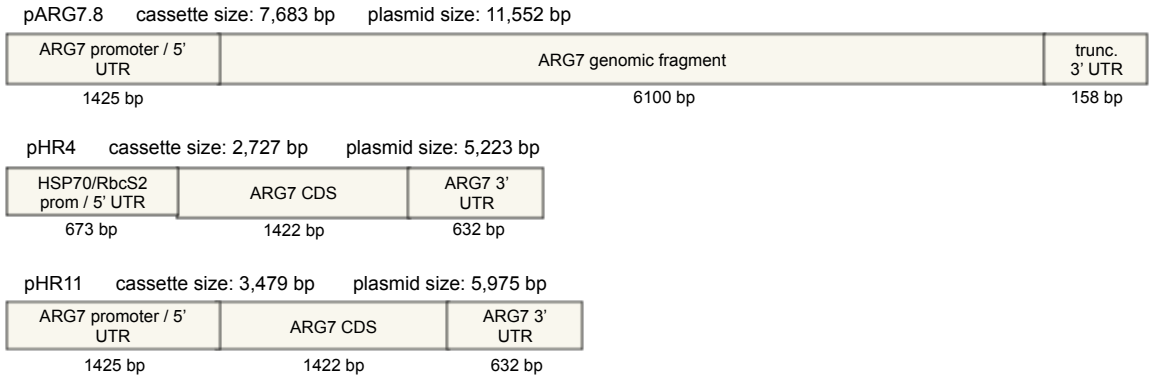


Figure 6.1 Schematic of genomic and cDNA-derived ARG7 selectable marker constructs. The first cDNA-derived construct, pHR4, is largely unable to produce colonies, while the pARG7.8 cassette can restore arginine synthesis but does not transform with high efficiency. pHR11 produces many more colonies than either other construct, and it is identical to pHR4 except for the promoter and 5' UTR regions.

Table 6.1 Transformation efficiencies of pHR11 and pARG7.8 in three arginine auxotrophic strains

Strain transformed	Colonies per µg pHR11	Colonies per µg pARG7.8
CC1819	5596 ± 455	824 ± 36
CC1820	8363 ± 735	887 ± 119
CC1826	832 ± 31	18 ± 4

The ARG7 promoter and ASL CDS function as molecular components for driving transgene expression

To test the ability of the ARG7 promoter and ARG7 gene to drive transgene expression, we linked the reporter gene mCherry to the ASL CDS via the 2A peptide and placed the ARG7 promoter upstream of the ARG7 CDS. In parallel, mCherry was linked to the ble gene in the HSP70/RBCS2-Ble-2A-GOI system, resulting in the following two plasmids: Hcr215: HSP70/RBCS2prom-Ble-2A-mCherry; and Hcr216: ARG7prom-ARG7cds-2A-mCherry (Figure 6.2).

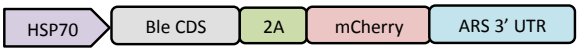
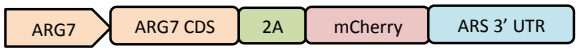
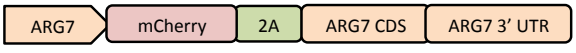
Plasmid name	Simplified construct map	Transformants / ug DNA		Plasmid size (bp)
		CC1010	CC1826	
Hcr215		152	121	5123
Hcr216		n/a	500	6771
Hcr218		36	90	4564
Hcr219		n/a	640	6755

Figure 6.2 Constructs for testing expression strength and robustness. Arrows at left of the schematics denote the promoter/5'UTR. "HSP70" is the chimeric HSP70-RBCS2 promoter/5'UTR (heat shock protein 70/rubisco small subunit 2). 2A is the E2A 24-residue self-cleaving peptide from foot-and-mouth disease virus. Transformation efficiency measured as number of colonies per ug DNA.

Hcr215 was used to transform both CC1010 and the arginine auxotrophic strain CC1826, whereas Hcr216 was only used to transform CC1826 because the selection is restoration of arginine prototrophy. mCherry expression was quantified to gauge the strength and robustness of the different expression cassettes. mCherry expression was determined quantitatively using mCherry-mediated fluorescence normalized to cell density. Normalized signals from individual transformants were displayed in relation to a threshold value of 2-fold higher signal than that obtained from non-transformed control strains (Figure 6.3).

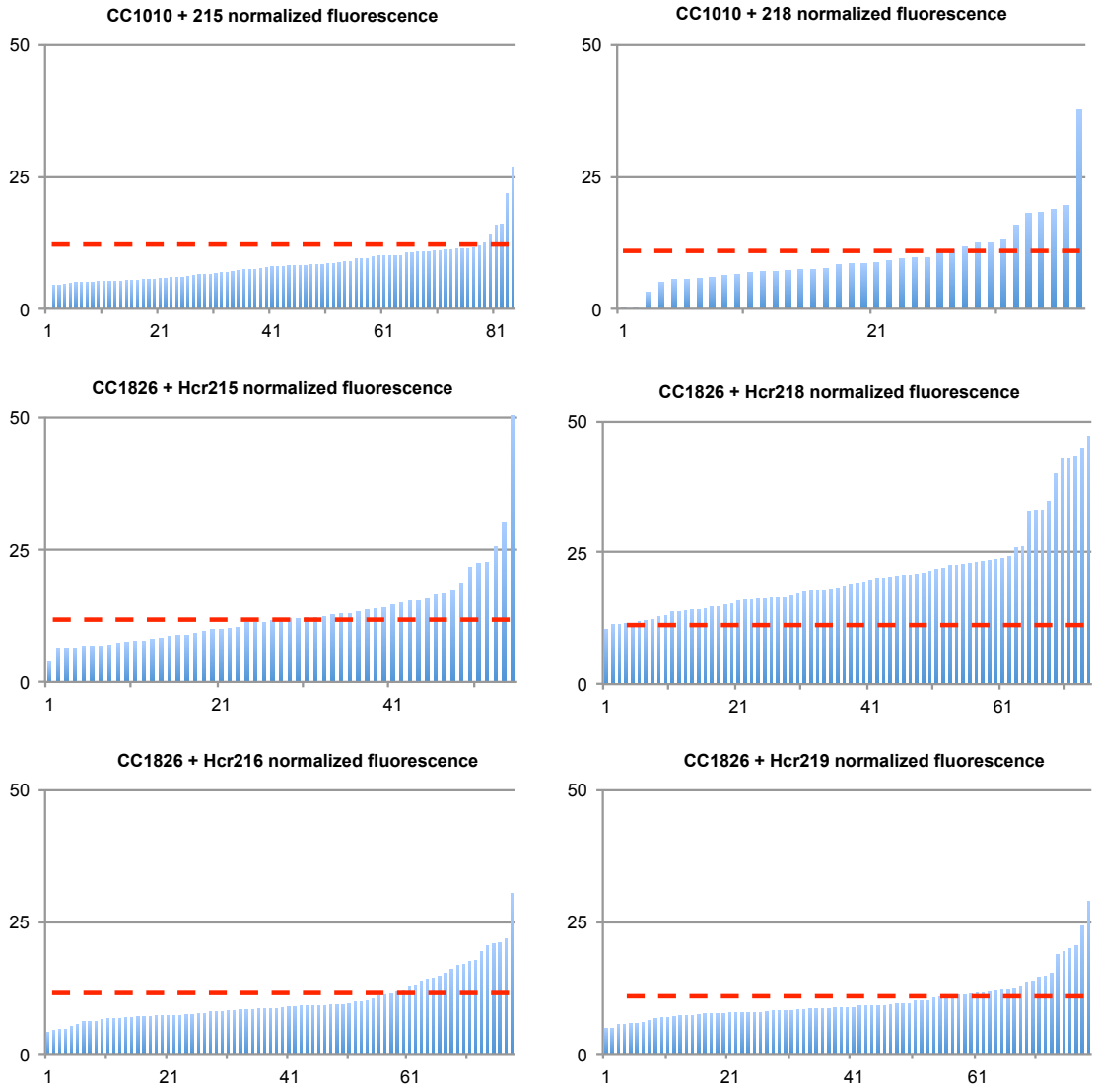


Figure 6.3 mCherry expression driven by Ble- and ARG-based expression cassettes. CC1010 or CC1826 strain transformed with mCherry expression cassettes. Bars represent fluorescent values normalized to cell density in relative fluorescent units. Each bar represents mCherry expression in an individual transformant. Dashed lines indicate threshold value defined as two times the normalized fluorescence values obtained in CC1010 transformed with water.

HSP70/RBCS2prom-Ble-2A-mCherry (Hcr215) was included as a benchmark to evaluate the ability of ARG7 based cassettes to drive transgene expression. In CC1826, Hcr215 generated almost twice as many transformants with signal above the threshold compared to when used to transform CC1010 (56% versus 30%) and 5 times as many when comparing signals that were >2x

threshold value (11% versus 2%) (Figures 6.3 and 6.4). Subsequently, Hcr215 in CC1826 was used as the benchmark for comparing to Hcr216 the ARG-based expression cassette. Transformation with Hcr216 in CC1826 resulted in 31% of the transformants with mCherry signals above threshold, while 6% exhibited signals >4x negative control levels (Figures 6.3 and 6.4).

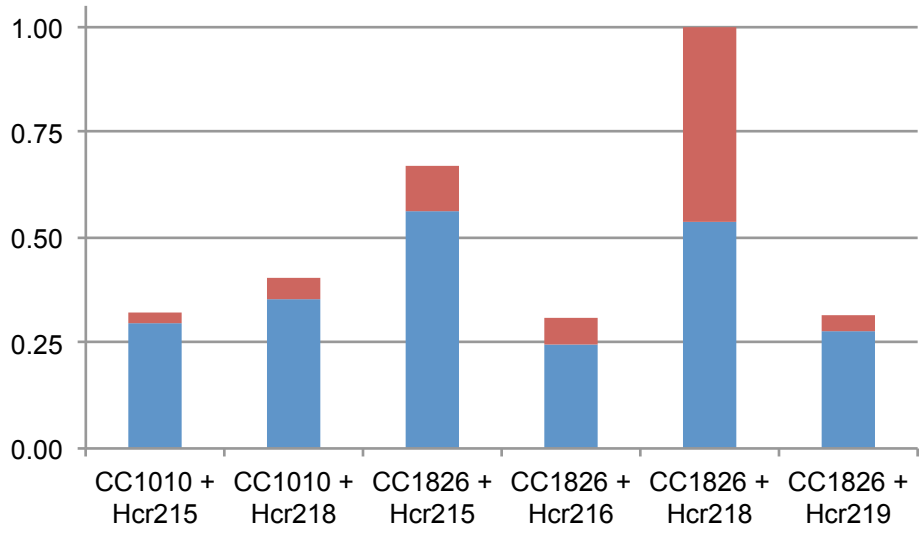


Figure 6.4 Proportion of mCherry-expressing transformants in relation to total transformants. CC1010 or CC1826 strain transformed with mCherry expression cassettes. Blue bars represent number of transformants showing values between 2-fold and 4-fold higher than negative control as a fraction of total transformants analyzed. Red bars represent number of transformants showing values at least 4-fold higher than negative control as a fraction of total transformants analyzed.

The order of the selection marker and the GOI affects transgene expression

To test the importance of the order of the selection gene versus the gene of interest in designing transcriptional fusion expression cassettes, two additional constructs were generated that are identical to Hcr215 and Hcr216 except that the order of reporter gene and selection marker is reversed: Hcr218: HSP70/RBCS2prom-mCherry-2A-Ble; and Hcr219: Arg7prom-mCherry-2A-ARG7cds (Figure 6.2). Hcr218 was used to transform both CC1010 and CC1826 whereas Hcr219 was used to transform only CC1826 because it requires an arginine auxotrophic

background in order to select transformants. Compared to Hcr215, Hcr218 performed similarly in CC1010 with approximately 40% of transformants exhibiting mCherry expression levels above threshold. However, for Hcr218 in CC1826 the increase in expression levels was remarkable. 100% of tested transformants had expression levels >2x negative control levels and almost 50% had expression levels >4x negative control levels. In contrast, no significant difference was observed between Hcr216 and Hcr219 when transformed into CC1826; whether the GOI was positioned upstream or downstream of the ARG7 selectable marker appeared not to affect expression levels of the GOI. Transformation efficiency varied between our constructs and depended on which strain was transformed. Generally, ble-containing constructs (Hcr215 and Hcr218) generated fewer transformants per ug of DNA compared to Hcr216 and Hcr219, which contain the ASL gene as selectable marker. When the order of ble selection marker and mCherry was reversed in Hcr218 with mCherry GOI placed upstream of the ble gene, transformation efficiencies were reduced to 20% of the Hcr215 construct. In comparison, transformation efficiency of ASL containing vectors was unaffected by the order of GOI and selection marker (Figure 6.2).

Discussion

In this study, we first improve upon the ARG7 selection marker, which is one of the oldest selection markers used for nuclear transformation of *C. reinhardtii*. One aim was to reduce the size of the marker in order to increase transformation efficiency and to facilitate molecular manipulations and inclusion in complex vector constructs. Both aims were achieved by the pHR11 vector, which contains the ASL-cDNA driven by its endogenous promoter/5'UTR. The 7-40 fold increase in transformation efficiency when using pHR11 compared to the pARG7.8 plasmid stands in contrast to the low transformation efficiencies obtained by the ASL cDNA-derived vectors generated in a previous study (*7*). As inclusion of introns in expression constructs

in general has been shown to enhance transgene expression (*9*), one explanation for the excellent transformation efficiency obtained by the pHR11 may be the inclusion of the first ASL intron in the construct. Alternatively, it has been shown that arg7 mutants harbor a second mutation that appears to inhibit arginine degradation and increases arginine uptake processes (*10*). It remains to be investigated whether this unknown mutation also affects the regulation of the arginine biosynthetic pathway, which may affect the strength of the ARG7 promoter/5'UTR driving the expression of the ASL-cDNA in the pHR11 construct. Finally, ASL may be subject to alternative splicing as the protein is able to complement arg7 arginine auxotrophic strains regardless of whether it is expressed with or without the first exon, which contains a predicted chloroplast targeting signal (*7*). As alternative splicing is sequence context dependent (*11*), high expression of ASL-cDNA may depend on the inclusion of both the endogenous promoter/5'UTR and the first intron.

Numerous arginine auxotrophic strains have been generated in *C. reinhardtii* and differences in transformation efficiencies among strains using the same construct have been observed previously. For example, complementation of arg7.8 mutants using pARG7.8 produces colonies faster than in arg7 and arg7.3 strains (*8*). It has been suggested that the arg7.8 mutant still produces some of the subunits of ASL, as it has been possible to complement it with truncated non-functional versions of the pARG7.8 plasmid (*12*). The three strains used in this study harbor different mutations in the ARG7.8 locus (CC1819:arg7.1; CC1820:arg7.2; and CC1826:arg7.8). The large differences in transformation efficiency of CC1819, CC1820 and CC1826 using pARG7.8 and pHR11 indicate that these strains require varying levels of ASL expression for complementation. For evaluating transgene expression we chose the CC1826 strain because complementation of this strain yielded the fewest transformants, indicating that this strain requires the highest expression of the ASL marker gene, which we hypothesized would be beneficial for high transgene expression via translational fusions to the ASL cDNA.

Previously, translational fusion of xylanase1 to the 3' end of the genomic ARG7 clone in the pARG7.8 plasmid (Beth Rasala, personal communication) via the 2A peptide rescued arginine phototrophy in the arginine auxotrophic strain CC3395 and showed detectable levels of xylanase activity (*2*). In this study, a high-throughput mCherry assay was used to show that the ARG7 promoter and ASL cDNA-based translational fusion expression cassette represents a viable alternative to the HSP70/RBCS2-ble containing system. Generally, the expression levels of mCherry achieved in this study from the ARG7 based expression cassette were largely comparable to the ble-based system (Figure 6.3). Importantly, from a comparable sample size the frequency of transformants with high transgene expression was similar when comparing the two expression cassettes (Hcr215 in CC1010 to Hcr216 in CC1826). To our knowledge CC1826 has not previously been used as a host for transgene expression. It is therefore notable that Hcr215 generated a higher proportion of high expressers in CC1826 than in CC1010, indicating that CC1826 may be particularly suitable for molecular pharming purposes. Previously, UV mutagenesis generated numerous independent *C. reinhardtii* strains capable of expressing transgenes up to 0.2% of total soluble protein (*13*), indicating that there is ample room for genetic improvement of nuclear transgene expression capacity. The higher expression levels obtained in CC1826, which was generated by UV mutagenesis, suggests that the large mutant strain collection available at the *Chlamydomonas* Resource Center may contain other strains suitable for high transgene expression.

When designing 2A transcription fusion expression cassettes, it is important to consider the order of selection gene and gene of interest. *Chlamydomonas* commonly cleaves and degrades introduced DNA molecules prior to genomic integration. Placing the GOI upstream of the selectable marker gene should theoretically ensure that all transformants express the GOI. The reduced transformation efficiency obtained by Hcr218 in CC1010 appears to confirm the vulnerability of the construct when the selection marker is exposed at the 3' terminus. However,

based on this reasoning, we anticipated that all transformants obtained from the Hcr218 vector would exhibit high expression of mCherry. This was only the case in CC1826, further stressing the potential suitability of this strain for transgene expression. Reasons for low GOI expression despite expression of the selectable marker may include the possibility that the construct was degraded at the 5' terminus into the GOI yet integrated into the genome downstream of a strong promoter, allowing expression of the remainder of the cassette.

In summary, we have generated a superior ARG7 marker with a small size which enables easy incorporation in complex expression vectors. This cassette is capable of complementing arginine auxotrophic strains with much higher transformation efficiency than previous arg7 complementation constructs. Moreover, we show that an ARG7prom-ASLcDNA-2A-GOI expression cassette represents a viable alternative to the HSP70/RBCS2-ble-2A-GOI system. Finally, we suggest that the arginine auxotrophic CC1826 strain may serve as a suitable expression host for both expression systems. Notably, the expression system generated in this study allows molecular pharming in *Chlamydomonas reinhardtii* without the use of antibiotic resistance markers, which is highly desirable.

Materials and Methods

Plasmids, strains and growth conditions

Arginine auxotrophic strains CC1819, CC1820, and CC1826, and wild type CC1010, were obtained from the *Chlamydomonas* Resource Center (chlamycollection.org, University of Minnesota). Strain names correspond to the Resource Center's nomenclature. Plasmid pArg7.8 was also obtained from the *Chlamydomonas* Resource Center. Cultures were grown with constant shaking under light in standard tris-acetate-phosphate media (TAP); arginine auxotrophs were supplemented with 100 mg/L L-arginine (Sigma, St. Louis, MO, USA).

RNA isolation, cDNA generation and PCR

RNA was isolated using the RNA Bee reagent (Amsbio, Lake Forest, CA, USA) following the manufacturer's instructions. cDNA was generated using the RevertAid Reverse Transcriptase kit (Fermentas, Glen Burnie, MD, USA) according to the manufacturer's instructions and recommendations for GC-rich templates. The best results were obtained with oligo$(dT)_{18}$ primers and PCRs were set up immediately. The PfuX7 polymerase (Norholm 2010) was used together with the Phusion polymerase GC buffer (New England Biolabs, Ipswich, MA, USA) and DMSO at a final concentration of 4% v/v.

Cloning ARG7

Fragments up to 1700bp could be amplified from cDNA. Full-length cDNA for ARG7 was assembled seamlessly from smaller fragments via the USER fusion technique. The ARG7 cDNA was amplified in three fragments: the promoter and 5' UTR (forward primer: LS160, GGGTTTAAUGTCGACGAGGAGGAGGTGC; reverse primer: LS161, ATGACGCUGAGACGAGGGG), the entire coding sequence (forward primer: LS162, AGCGTCAUGGCCCAGGCAGCTGC; reverse primer: LS163, ACCCGCTCAGUGCTGTCCCTCCGC), and the 3' UTR (forward primer: LS164, ACTGAGCGGGUCGGGGGAGGG; reverse primer: LS165, GGTCTTAAUATGTAACATGTTCTATTTAAAAAGACTGCCC). The three fragments were assembled into the USER- compatible cloning vector Hcr1 (pBluescript into which the USER cloning cassette (GCTGAGGGTTTAATTAAGACCTCAGC) has been has been inserted into the *EcoR*I and *Pst*I restriction sites as previously described (*14*) by USER fusion as previously described (*15*), using approximately 100 ng of vector and equimolar amounts of each fragment with 1 µl USER enzyme in a total reaction volume of 10 µl. Reactions were incubated at 37°C for 20 minutes followed by 25°C for 20 minutes.

Generation of reporter constructs

All reporter constructs (Hcr215, Hcr216, Hcr218, and Hcr219) were constructed with seamless USER cloning using fragments amplified from existing reporter plasmids. Both the *sh*-Ble gene (*2*) and the mCherry gene (*16*) were codon-optimized for expression in the *C. reinhardtii* nuclear genome. The 2A sequence is derived from equine rhinitis A virus and is known to cleave with very high efficiency relative to other 2A sequences (*17*). Constructs were sequenced fully after USER construction to ensure that no mutations were introduced in the USER PCR-amplification step.

Transformation conditions

Strains were grown to mid-late log phase (3-5 x 10^6 cells/ml) before harvesting by centrifugation at 2500 rpm and resuspension at $3x10^8$ cells/ml in *Chlamydomonas* MaxEfficiency transformation reagent (Life Technologies). 250 µl of the resuspension was added to 4mm cuvettes, along with 1 ug of linearized plasmid DNA. Cuvettes were incubated at room temperature for 5 min, on ice for 5 min, and electroporated in a Bio-Rad Gene Pulser Xcell with the following settings: 800 V, 25 µF, infinite resistance. Cuvettes were immediately placed in a room-temperature water bath for 5 min, and transferred to 50 ml conical tubes containing 10 ml of TAP plus 40mM sucrose. Cells were incubated with gentle shaking for 14-18 hours spun down (2000 rpm, 10 min), resuspended in 800 µl TAP with 40mM sucrose, and plated onto selection. Colonies appeared after about 5 days.

mCherry fluorescence assay

Transformants were grown for 1 week by inoculating 1ml TAP (without selection) in deep-well 96-well microtiter plates under constant shaking (200 rpm) covered by breathable sealing tape. On the day of the assay, 200 ul was transferred from each well to a standard clear

flat bottom 96 well microtiter plate for OD750 measurements. The remaining 800ul culture was centrifugated and one metal ball was added to each well. The plate was sonicated twice in a sonicator bath with an additional centrifugation between sonications. Following the second sonication 200ul supernatant was transferred to a black and flat bottom 96-well microtiter plate for fluorescent measurements on an ELISA reader with the following settings: excitation wave length: 570nm; filter: 610nm; emission: 620nm. Fluorescence output was the average of 5 readings in different spots within each well. A dilution series was generated to correlate OD to an arbitrary concentration range, which was used to normalize fluorescence readings.

Acknowledgements

This work was funded by the Department of Energy grant DE-EE0003373. EAS was supported by a National Science Foundation graduate research fellowship, and HHN-E was supported by the Danish Research Council for Technology and Production (FTP) grant 09-065827/274-08-0354.

Chapter 6, in full, is a reprint of material in preparation for submission for publication. The dissertation author is the secondary investigator and author of this paper.

References

1. C. Griesbeck, A. Kirchmayr, Algae: An alternative to the higher plant system in gene farming. *Molecular Farming in Plants: Recent Advances and Future Prospects*, 125 (2012).

2. B. A. Rasala, P. A. Lee, Z. Shen, S. P. Briggs, M. Mendez, S. P. Mayfield, 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 (2012).

3. R. Debuchy, S. Purton, J. D. Rochaix, The argininosuccinate lyase gene of chlamydomonas-reinhardtii - an important tool for nuclear transformation and for correlating the genetic and molecular maps of the arg7 locus. *Embo Journal* **8**, 2803 (Oct, 1989).

4. J. P. Davies, A. R. Grossman, Sequences controlling transcription of the chlamydomonas-reinhardtii beta(2)-tubulin gene after deflagellation and during the cell-cycle. *Mol Cell Biol* **14**, 5165 (Aug, 1994).

5. J. P. Davies, D. P. Weeks, A. R. Grossman, Expression of the arylsulfatase gene from the beta 2-tubulin promoter in chlamydomonas reinhardtii. *Nucleic Acids Res* **20**, 2959 (Jun 25, 1992).

6. M. Ohresser, R. F. Matagne, R. Loppes, Expression of the arylsulphatase reporter gene under the control of the nit1 promoter in chlamydomonas reinhardtii. *Curr Genet* **31**, 264 (Mar, 1997).

7. A. Auchincloss, A. Loroch, J.-D. Rochaix, The argininosuccinate lyase gene of chlamydomonas reinhardtii: Cloning of the cdna and its characterization as a selectable shuttle marker. *Molecular and General Genetics MGG* **261**, 21 (1999).

8. S. Purton, J. D. Rochaix, Characterization of the arg7 gene of chlamydomonas-reinhardtii and its application to nuclear transformation. *Eur J Phycol* **30**, 141 (May, 1995).

9. V. Lumbreras, S. Purton, Recent advances in chlamydomonas transgenics. *Protist* **149**, 23 (Feb, 1998).

10. R. Loppes, Strijker.Pj, Arginine metabolism in chlamydomonas-reinhardi - conditional expression of arginine-requiring mutants. *Mol Gen Genet* **116**, 248 (1972).

11. A. Labadorf, A. Link, M. F. Rogers, J. Thomas, A. S. N. Reddy, A. Ben-Hur, Genome-wide analysis of alternative splicing in chlamydomonas reinhardtii. *Bmc Genomics* **11**, (Feb 17, 2010).

12. W. Mages, O. Heinrich, G. Treuner, D. Vlcek, I. Daubnerova, M. Slaninova, Complementation of the chlamydomonas reinhardtii arg7-8 (arg2) point mutation by recombination with a truncated nonfunctional arg7 gene. *Protist* **158**, 435 (Oct 23, 2007).

13. J. Neupert, D. Karcher, R. Bock, Generation of chlamydomonas strains that efficiently express nuclear transgenes. *The Plant journal : for cell and molecular biology* **57**, 1140 (Mar, 2009).

14. H. H. Nour-Eldin, B. G. Hansen, M. H. Norholm, J. K. Jensen, B. A. Halkier, Advancing uracil-excision based cloning towards an ideal technique for cloning pcr fragments. *Nucleic Acids Res* **34**, e122 (2006).

15. F. Geu-Flores, H. H. Nour-Eldin, M. T. Nielsen, B. A. Halkier, User fusion: A rapid and efficient method for simultaneous fusion and cloning of multiple pcr products. *Nucleic Acids Res* **35**, e55 (2007).

16. B. A. Rasala, D. J. Barrera, J. Ng, T. M. Plucinak, J. N. Rosenberg, D. P. Weeks, G. A. Oyler, T. C. Peterson, F. Haerizadeh, S. P. Mayfield, Expanding the spectral palette of fluorescent proteins for the green microalga chlamydomonas reinhardtii. *The Plant journal : for cell and molecular biology* **74**, 545 (May, 2013).

17. M. L. Donnelly, L. E. Hughes, G. Luke, H. Mendoza, E. ten Dam, D. Gani, M. D. Ryan, The 'cleavage' activities of foot-and-mouth disease virus 2a site-directed mutants and naturally occurring '2a-like' sequences. *J Gen Virol* **82**, 1027 (May, 2001).

CHAPTER 7:

HIGH-THROUGHPUT SYSTEM FOR DETECTING IMPROVEMENTS IN HOMOLOGOUS RECOMBINATION EFFICIENCY IN *CHLAMYDOMONAS REINHARDTII*

Abstract

The ability to perform targeted gene knockouts and targeted gene integration is routine in most model organisms, and indeed is required for reverse genetics and for most complex multi-trait engineering. However, in the model green alga *Chlamydomonas reinhardtii*, introduced genes insert into the nuclear genome primarily by non-homologous end-joining, resulting in random integration and therefore high variability between transformants due to positional effect. Previous attempts to rigorously measure the basal rate of homologous recombination (HR) in this species have been met with difficulty; the extreme rarity of observed recombination events leads to a lack of reproducibility in the absence of a reliable high-throughput screening system, and attempts to verify the nature of the recombination events have often led to ambiguous results. We have devised a high-fidelity, high-throughput system for measuring HR efficiency using a combination of complex multi-cassette vectors utilizing a truncated selectable marker. Each experiment provides a direct measurement of both HR and non-homologous end-joining (NHEJ) events, which allows us to reproducibly measure the basal HR rate even when variables in culture conditions greatly affect overall transformation efficiency. Recombinants are characterized to provide the first comprehensive analysis of the nature of recombination events within *C. reinhardtii*, revealing that both double and single crossover events occur, often with errors at the recombination site. We then use this system to evaluate two approaches previously described in the literature as increasing the HR efficiency, as well as a novel approach that is able to increase HR efficiency. This system is valuable for high-throughput, systematic testing of myriad additional approaches for increasing HR efficiency, such that the most fruitful approaches can be potentially synergistically combined to achieve HR rates necessary for facile reverse genetics and gene targeting in *Chlamydomonas*.

Introduction

Upon DNA damage, two main repair pathways exist: homologous recombination (HR) also known as homology-directed repair, which requires an intact region of homology to template the repair; and non-homologous end-joining (NHEJ), wherein double-stranded DNA ends are ligated at random to repair the break. Nuclear transformation of *C. reinhardtii* typically is achieved by electroporation of a linearized, double-stranded plasmid vector. In nearly all cases, this linear piece of foreign DNA is recognized as a double-stranded break, and the NHEJ pathway is utilized to insert the DNA into the genome without location specificity. This method is useful in some cases, such as generation of insertional mutant libraries. However, for the vast majority of applications, integration of foreign DNA by homologous recombination is far preferable. Some model organisms, including many bacteria and even eukaryotes like *Saccharomyces cerevisiae*, naturally perform homologous recombination with high efficiency, making targeted gene knockouts and insertions straightforward. Indeed, this characteristic facilitated the early development of these model systems, particularly due to the ability to knock out genes for reverse genetics studies.

Chlamydomonas reinhardtii is often called the "green yeast"; however, its genetic malleability is significantly hampered compared to yeast due to its inability to perform homologous recombination efficiently within the nuclear genome. In *C. reinhardtii* chloroplasts, homologous recombination is the predominant method of foreign DNA integration. Due to the chloroplast prokaryotic ancestry, a RecA recombinase similar to that found in bacteria is encoded within the algal nuclear genome and the gene product is targeted to the chloroplast using a localization signal during translation (*1*). Transformation vectors for the chloroplast genome typically contain approximately 1kb of homology on either side of the transgene expression cassette, targeting the vector either to a neutral site or to a site for targeted knockout, often of the *psbA* gene. Random insertions are essentially never observed.

In the nuclear genome, on the other hand, the RecA recombinase is not present and thus cannot facilitate homologous recombination. *C. reinhardtii* does contain homologues of eukaryotic-like HR-associated genes, such as Rad51 – which has even been purified and shown to be active (*2*) – that do not display chloroplast-localization sequences. These HR-associated genes remain generally down-regulated, presumably because the cells typically exist vegetatively with haploid genomes; thus there is no homologous partner with which recombination could occur during most of the cell cycle.

However, several studies have shown that this machinery is not completely inactivated; HR is indeed possible in vegetative cells and can be detected through extensive screening facilitated by selection for the targeted locus, but the frequency relative to NHEJ is extremely low. The first conclusive evidence of HR within the genome was obtained using a "low-reverting" nit1 mutant, in which HR:NHEJ ratios were reportedly as high as 1:24 using the glass beads transformation method (*3*). These numbers have been questioned by a number of subsequent studies, but this report did demonstrate that their flagged nit1-complementing cassette had integrated into the target region in a handful of transformants, providing the first evidence of recombination in vegetative algal cells. It also revealed that the process of recombination in algae is not necessarily straightforward or predictable, and multiple insertions of the cassette were often observed at a single locus, hinting at the complexity of homology-based DNA integration. Similarly, a subsequent study two years later used NIT8 as the target locus and also observed complex rearrangements and multiple insertions (*4*). This study estimated the HR rate to be approximately one in 700, but their selection suffered from several confounding factors including the fact that their selection phenotype could result from disruption of multiple nit genes, not necessarily NIT8. Indeed, their presumptive recombinants' phenotypes often did not segregate with the selection used to disrupt NIT8 (*4*), confirming that the apparent NIT8 knockout phenotype was spurious.

Two studies have used restoration of arginine prototrophy in arg7-8 mutants (at the ARG7 locus, encoding arginosuccinate lyase) as a selection for complementation by recombination (*5, 6*). Using co-transformed truncated versions of a functional ARG7 cassette, it was shown that as little as 230 bp of homology can enable recombination between two simultaneously introduced plasmids (*5*). However, transformants exhibiting speculative recombination into the ARG7 site in the genome were not conclusively analyzed; for most colonies, circumstantial evidence for HR did not fit well with any of their predicted models of recombination. Despite the authors' speculation that the ARG7 gene may be a poor choice for genome targeting due to highly repetitive sequences and a potential propensity for rearrangement (*5*), this target site was used in a subsequent study over a decade later. The nature of the arg7-8 mutants was finally characterized and was found to be a single missense point mutation within the active site (*6*), making reversion a likely possibility for the appearance of arginine prototrophic transformants. Both groups claim that in their hands they did not observe this reversion spontaneously, but Mages *et al.* did witness other compensatory suppressor mutations, and therefore concluded that arginine prototrophic colonies with the desired point mutation reversion were "presumably" due to recombination (*6*).

Finally, to avoid the ambiguity associated with targeting endogenous genes, Zorin *et al*. developed a fully exogenous system using a truncated paromomycin resistance gene (*7*). However, their system was unable to report both HR and NHEJ events using the same DNA substrate; thus, calculation of the HR:NHEJ ratio relied on comparison between transformations of different molecular species, requiring significant assumptions to be made about the comparability of overall transformation efficiency. As with previous studies, they observed significant rearrangements at the site of recombination; rearrangements that affected the 5' end of their construct rendered their selectable marker nonfunctional, so they were not observed (*7*). Furthermore, although the HR prevalence was increased relative to NHEJ events because of the

reduction in NHEJ, HR events were still extremely rare – requiring 70 μg of DNA to achieve between one and five verifiable recombinants. In subsequent studies, this group found this approach to be unfeasible for targeting many genes due to the need to introduce the selection cassette in frame with the gene of interest (*8*). Thus, while the machinery exists for HR to occur in the nuclear genome, the rate at which it is observed is too low for practical applications without extensive, laborious screening to identify recombinants. Indeed, making insertional libraries and subsequently screening for a hit in a gene of interest has proven to be more feasible than performing knockouts by homologous recombination (*9*), but this method is far from facile.

Researchers of other model organisms in which HR does not frequently occur, such as mice and higher plants, have tackled this issue with a variety of approaches. One strategy has been developing site-specific recombination systems, such as Cre, FLP, and PhiC31 recombinase, which have been used extensively in higher eukaryotes (*10-12*). However, these systems require significant up-front work to obtain a recognition cassette insertion in the desired location prior to being able to reliably target that location. Recently, more involved systems such as designed zinc finger nucleases and TALENs (transcription activator-like effector nucleases) have been used with some success to cause targeted double-stranded breaks, though only zinc fingers have been demonstrated thus far in algae (*8*). These approaches also require extensive up-front work, as the nuclease must be designed and then the gene synthesized or constructed modularly for each desired DNA recognition site. Furthermore, several recognition sites are often targeted within one gene due to lack of predictability of the binding affinity, and therefore effectiveness, at a given site.

Most recently, the CRISPR-Cas bacterial immunity system has been adapted for genetic engineering applications (for a review, see (*13*)), requiring a small guide RNA unique to the target site in addition to the large (160 kD) Cas9 nuclease to cleave the target site. This system has been used with notable success in organisms ranging from bacteria to mammalian cells to

multicellular plants (*14-16*), but has not yet been demonstrated in green algae. Furthermore, all of these nuclease-driven approaches may have limited efficacy in an organism with very low basal HR machinery activity; they create a break at the target site to encourage DNA repair at that locus (which may be in the form of either NHEJ or HR), but they do not directly facilitate homologous recombination itself.

Fortunately, there are other options to test that may increase HR without such a significant time and resource investment, such as modifying culture conditions prior to transformation, modifying the substrate DNA, or pre-incubating with proteins to facilitate recombination. Even approaches as simple as testing the location at which substrate DNA is linearized, the nature and depth of the linearization overhang, and the length of homology require a standardized, reliable, high-throughput method of quantitative HR efficiency – something that had not yet been developed in *C. reinhardtii*. We describe such a system, which can reliably measure both HR and NHEJ events from a single transformation. This significantly reduces the risk of over- or under-estimating HR efficiency based on incorrect assumptions comparing between transformations using different DNA substrates. Furthermore, this is the first system described to date in algae that can capture and precisely characterize a wide range of recombination events, including insertions, deletions, and rearrangements at the recombination site while still making use of a selectable marker to identify recombinants. Using a high-throughput screening strategy, we consistently observe dozens of verifiable recombinants – compared to one or a few colonies with ambiguous Southern blot-based characterization reported in previous studies – lending significant credibility to the HR rates observed, and allowing incremental improvements in HR efficiency to be reliably detected. As proof of principle, we test three approaches for affecting the HR rate using this system, as well as the synergistic effects of the two most promising approaches.

Methods

Vector design and construction

All plasmids were designed *in silico* using Geneious sequence analysis software (Auckland, New Zealand) and constructed seamlessly with USER cloning as described previously (*17*), into the USER cloning-compatible vector Hcr1 (pBluescript into which the USER cloning cassette (GCTGAGGGTTTAATATTAAGACCTCAGC) has been inserted as previously described (*18*). Fragments were amplified from existing plasmids using a modified Phusion polymerase, X7, in 5X Phusion GC buffer (New England Biolabs) with 2M betaine (Sigma-Aldrich) to accommodate high GC content. Candidate colonies were sequenced in full to ensure no errors were introduced in the PCR amplification.

Algal strains and growth conditions

All strains used in this work (cc1820 and cc1690) were obtained from the *Chlamydomonas* Resource Center, www.chlamycollection.org (St. Paul, MN). The recipient strain, B12, described in this work is now available at the *Chlamydomonas* Resource Center as well. All strains were grown in standard TAP media supplemented with 100 µg/ml sterile-filtered L-arginine (Sigma-Aldrich) prior to transformation. Liquid cultures were grown in constant light with shaking.

Algal transformation by electroporation

Chlamydomonas recipient strain B12 cultures were grown in TAP to a density of 3-5x10^6 cells/ml prior to electroporation. They were then spun down and resuspended at 3x10^8 cells/ml in *Chlamydomonas* transformation reagent (Life Technologies, Carlsbad, CA). For each trial, 25 cuvettes were transformed in sets of five as follows: 1400 µl of resuspended cells were incubated with 5 µg of linearized DNA at room temperature for 5 minutes and transferred into five 4 mm

electroporation cuvettes, 270 ul per cuvette. The cuvettes were then incubated on ice for 5 minutes prior to electroporation in a BioRad Gene Pulser with the following settings: 500 V, 50 µF capacitance, 800 Ω resistance. After electroporation, cells were allowed to recover at room temperature for 5 minutes, then all five cuvettes were pooled into a single 50 ml flask of TAP supplemented with 100 µg/ml of L-arginine and 40 mM sucrose. The cultures recovered for 12-16 hours in constant light on a shaker. Cells were then spun down and resuspended in 3 ml TAP. From this resuspension, 20 µl was mixed with 500 µl of TAP and plated onto TAP containing 20 µg/ml paromomycin to obtain a measure of the overall transformation efficiency. The remaining resuspended cells were divided evenly among five TAP plates containing 15 µg/ml hygromycin and 100 µg/ml arginine, to test for homologous recombinants at the desired locus. Colonies were typically visible on paromomycin within 4-5 days of plating and visible on hygromycin within 5-6 days, and all colonies were counted within 10 days of plating.

Arylsulfatase assays

Only colonies that were PCR-positive for the 3' end of the recipient cassette using primers HR27 (AACAGGTGTTCCCGGAGCTGCTGCGC) and HR28 (GGTCTTAATCGCTTCAAATACGCCCAGCC) were screened for arylsulfatase activity. Colonies were inoculated into deep-well plates (Axygen Scientific) containing 0.5 ml TAP. Cultures were grown for 3-4 days to saturated density and centrifuged for 5 min at 2000 rpm. 100 µl of media was transferred to a flat-bottom clear plate (Corning). To each well, 25 µl of N-SO_4 reaction buffer (2M glycine-NaOH, pH 9, 50mM imidazole, 4mM α-naphthyl sulfate (Sigma-Aldrich)) was added, then incubated at 37°C for 1hr. Then 125 µl stopping buffer (0.2M sodium acetate, pH 4.8, 4% SDS) was added, followed by 25 µl of 10 mg/ml Fast Blue B Salt (Sigma-Aldrich). The color was allowed to develop for 1 min and read using an Infinite 200 Pro platereader (Tecan) for absorbance at 540nm.

Southern blotting

Genomic DNA was extracted from each of the analyzed recipient strains and the parental strain as described (*19*). 2.5 µg of genomic DNA was digested with *Xho*I, which cuts only once in the recipient strain cassette, downstream of the probe. A 601 bp probe was amplified using primers LS138 (CGGCTACGTGCAGTGCAG) and LS163 (ACCCGCTCAGTGCTGTCCCTCCGC) and DIG-labeled dNTPs following the manufacturer's protocol (Roche). The digested genomic DNA was run on a 0.7% agarose gel and blotted onto nylon membrane overnight by capillary transfer, and then hybridized and exposed using the BIG High Primer Starter Kit II as per the manufacturer's instructions (Roche).

Sequencing and restriction analysis of recombination sites

Individual hygromycin-resistant clones were subjected to further analysis to characterize the nature of the recombination. These clones were grown in 5 ml cultures of TAP to saturation, then the genomic DNA was extracted as previously described (*19*). This was used as template for PCR spanning the intron with primers HH235 (CTACCTGGTCATGTCGCGGATGAC) and LS330 (CAGGAGGGTGTCCACGTCAG), using Phusion GC buffer and 2M betaine to enhance amplification of large fragments from the GC-rich nuclear genome. PCR-amplified fragments were gel-purified and cloned into pJet2.1 (CloneJET Kit, Thermo Scientific), and one colony per algal clone was sequenced via Sanger sequencing and aligned using Geneious sequence analysis software. The CloneJET plasmids containing the PCR-amplified recombination junctions were analyzed by restriction digest using *Not*I-HF and *Xba*I (New England Biolabs). The CloneJET backbone contains one of each of these sites, one on each side of the PCR insert. The ARG7 cassette contains a *Not*I site near the 3' end, but these enzymes are otherwise absent from both pHR18 and pHR23.

Algal transformation by particle bombardment

Chlamydomonas recipient B12 cultures were grown to 2-3x10^6 cells/ml, then spun down and resuspended at 3x10^7 cells/ml in TAP media. For each experiment, 0.5 ml of cells were plated onto paromomycin plates and hygromycin plus arginine plates; this experiment was performed three times, for a total of seven paromomycin and 32 hygromycin plus arginine plates analyzed. 60 µl of SD 500 gold particles (Seashell Technologies, San Diego) were coated with 10 µg of *EcoR*V-HF linearized, heat-killed pHR23 (Experiments P1 and P3) or 10 µg of circular pHR23 plasmid (Experiment P2) according to the particle manufacturer's instructions. Each gold particle preparation (was spread onto three carrier disks, for a total of 39 transformations. In these experiments, the number of theoretical transformants was equal to five times the number of colonies counted on the paromomycin plates, and the HR calculations were otherwise performed as described above.

Preparation of single-stranded DNA

pHR23 plasmid was digested using *Not*I-HF and *Sbf*I-HF, to provide one digestible end and one end with a 3' overhang greater than 4nt, which is resistant to Exonuclease III degradation. After heat-killing both enzymes, the digested plasmid was buffer-exchanged using 0.22 micro filter discs (Millipore) and digested with Exonuclease III for 1 hr at 37°C in NEB Buffer 1. Afterwards, the reaction was heat-killed for 20 min at 70°C. Due to the differences in DNA preparation compared to the standard pHR23 preparation, a new basal rate was established using pHR23 prepared as described above, with buffer added but no enzyme. Transformations were performed by electroporation, as described above.

Pre-incubation with Tth *RecA DNA-binding protein*

For each set of five transformations, 5 µg of linearized pHR23 was incubated with 2 µg of *Tth* RecA (New England Biolabs) with 1X RecA buffer in 100 µl total volume for 30 minutes at either 60°C, 75°C, or 90°C. The reaction was chilled on ice prior to electroporation, which was performed exactly as described above. For the combination of single-stranded DNA with *Tth* RecA, the DNA was prepared exactly as above for ssDNA and then incubated with *Tth* RecA at 60°C for 30 min in 1X RecA buffer immediately prior to electroporation.

Results

Drawing upon the perceived drawbacks of previous attempts to study HR in *C. reinhardtii*, we designed a novel system for detecting HR that meets the following criteria: accommodates high-throughput transformation and screening to rapidly evaluate hundreds of thousands of transformants; provides a high-fidelity phenotype for detecting HR without false positives from reversion events or locus heterogeneity; reports both HR events and NHEJ events from a single transformation for direct comparability; and tolerates imprecise recombination or rearrangements at the integration site, to capture a wider range of possible recombination events. Our system comprises two vectors (Figure 7.1) that recombine to report homologous recombination while simultaneously providing a measure of non-homologous end-joining, permitting a direct calculation of the HR:NHEJ ratio within a single experiment. HR is detected by reconstituting an exogenous resistance gene to the antibiotic hygromycin, which is known to elicit no spontaneous resistance. While designed for use with electroporation using a recently developed high-efficiency transformation reagent for high-throughput generation and screening of transformants, this system is amenable to any nuclear transformation technique. This is the first system developed in algae that can accommodate imprecise recombination – including rearrangements, loss of sequence, and introduction of concatamers or other exogenous sequence – without interfering with the selection.

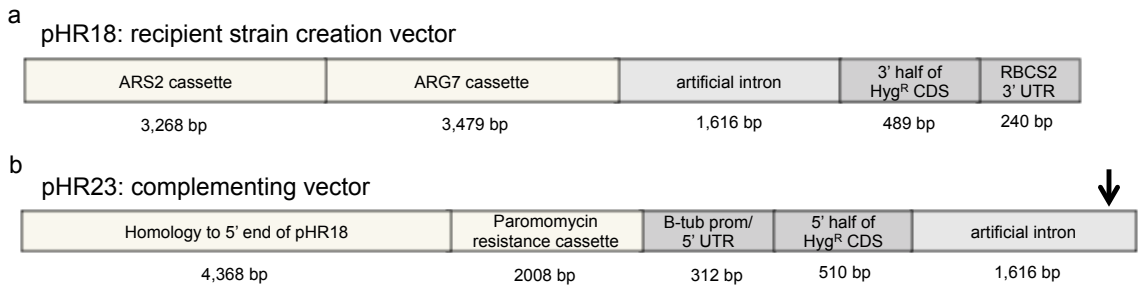


Figure 7.1 Schematic of vectors comprising the HR detection system. a) pHR18 is the vector used to create the recipient strains, and integrates at random into the nuclear genome. Selection is conferred by the ARG7 cassette, and the ARS cassette is used to screen for high expression of the recipient vector. **b**) pHR23 is the complementing vector designed to recombine with at the pHR18 locus within the nuclear genome to confer hygromycin resistance. The paromomycin resistance cassette allows measurement of the overall transformation efficiency, so that the observed homologous recombinants can be translated into a homologous recombination rate as a fraction of total transformants. The arrow on pHR23 indicates the *EcoR*V site, where the plasmid was linearized prior to electroporation.

Artificial introns can be introduced within selectable markers, and are properly recognized and spliced

The system we designed relies on recombination between two non-functional halves of the hygromycin resistance cassette to provide selection for rare recombination events. However, in designing this system, we wished to address two obstacles that had complicated previous studies. First, we desired a system with flexibility at the recombination site, such that the rearrangements and unpredictable recombination outcomes previously postulated could be reliably captured and examined in detail. Secondly, we wanted each half of the reconstituted selectable marker to be profoundly non-functional, thus avoiding the difficulties of previous studies that had relied upon a nearly-intact marker. The length of homology between the two complementary constructs was initially limited by the length of the selectable marker cassette. The entire hygromycin resistance cassette is less than 1.6 kb, and the coding sequence itself is only 1 kb. This leaves very little room for overlapping homology while ensuring that each truncated half is fully non-functional and is incapable of gaining functionality following fortuitous integration in frame with an actively expressed gene. Indeed, in previous systems to

detect HR, often only the promoter or a small portion of the coding sequence was removed (*3, 5, 7*), leaving open the possibility of trapping a promoter for selection marker expression in the absence of gene targeting. To circumvent this obstacle, we referenced the sequence logo of algal splice sites known or predicted to be constitutively recognized (*20*) and identified a site in the middle of the hygromycin resistance gene likely to be recognized as the exon-side halves of a constitutive splice site (AG upstream of the 5' splice site and GT downstream of the 3' splice site). A 1.6 kb intron containing canonical constitutive splice sites was obtained from a list of long, native *C. reinhardtii* introns (Mark Rogers, personal communication), cloned from genomic DNA, and inserted into the hygromycin resistance gene at the location described above.

This hygromycin cassette with the artificial intron was transformed into wild type *C. reinhardtii* and selected on hygromycin. While the efficiency was slightly lower than the identical cassette without the intron – likely due to the increased size with the intron present – the construct with the intron was indeed able to confer hygromycin resistance, indicating proper intron recognition and splicing. The truncated halves were also validated to be non-functional, by transforming each half of the hygromycin resistance cassette accompanied by the intron and selecting on hygromycin. No colonies were observed for either truncated construct, confirming that the truncations on their own are unable to confer hygromycin resistance. This established the hygromycin resistance cassette with the artificial intron as an ideal selection for identifying homologous recombinants.

The addition of this intron serves a secondary purpose as well: it allows imperfect recombination to still be detected, because insertions or deletions within this region will not disrupt the functionality of the recombined hygromycin resistance cassette, given that the splice sites remain intact. This is the first system described in *Chlamydomonas* in which instances of recombination with imperfect repair can be observed, due to the system's flexibility for allowing insertions and deletions within the intronic homology region.

The truncated selectable marker can recombine in vivo *to produce functional hygromycin resistance*

A pair of vectors (Figure 7.1) was designed such that recombination within the artificial intron described above would result in a complete hygromycin resistance cassette, providing an unambiguous, reliable read-out for HR. Once recombinants are selected on hygromycin, they can then be further analyzed to determine whether the recombination event involved a single or double crossover by testing for arginine auxotrophy. Prior to using these vectors to measure homologous recombination, pHR18 and pHR23 were transformed separately into wild type *C. reinhardtii*. As anticipated, no colonies were obtained on hygromycin, confirming that each half of the hygromycin cassette is non-functional on its own in the context of these constructs.

Previous reports have indicated that recombination occurs much more frequently between two co-transformed plasmids than between a plasmid and the genome (*3, 5*). Before ramping up to a high-throughput screen to detect recombination within the genome, we first tested the ability of our system to produce functional hygromycin resistance via co-transformation into a naïve wild-type strain, cc1690. In a single experiment, 312 colonies were obtained on paromomycin and then patched onto hygromycin. One of the patches grew on hygromycin, and the junction was PCR-amplified and sequenced to reveal that it was indeed hygromycin resistant as a result of recombination within the intron homology region. Sequencing also showed that a significant portion of the intron had been deleted (945 bp of the 1,616 bp intron were missin), demonstrating the utility of the intronic flexibility at the recombination site. Inspired by this demonstration of the system's ability to report recombination events between extrachromosomal plasmids, we proceeded to create recipient strains to measure recombination within the genome.

The system can measure homologous recombination rates with high fidelity and reproducibility

The vector pHR18, also called the recipient strain creation vector, was introduced into arginine auxotrophic strain cc1820, which contains a loss-of-function mutation in the ARG7 gene required for arginine biosynthesis (*21*). The recipient strain creation vector contains a cDNA-derived version of the ARG7 gene (Nour-Eldin *et al*., in preparation) for selection when transformed into an arginine auxotrophic strain. Colonies obtained on TAP were screened by two methods: first, they were screened by PCR using primers that anneal at the 3' end of the construct, to ensure that the full construct had integrated into the genome (Figure 7.2b). Clones that were PCR-positive for full integration were then inoculated into 96-deep-well plates and assayed for arylsulfatase activity in the media (Figure 7.2c). The ARS2 gene is endogenous to *C. reinhardtii* but is not expressed under normal conditions. However, the ARS2 included at the 5' end of the pHR18 construct is under the control of a constitutive promoter and serves as a reporter for semi-quantitative measurement of gene expression. The arylsulfatase assay ensured that the recipient cassette was highly expressed, not silenced, so that recombination at that locus would manifest as a detectable phenotype (hygromycin resistance). Using these screens, three strong candidate recipient strains were identified, referred to as B12, C6, and F5.

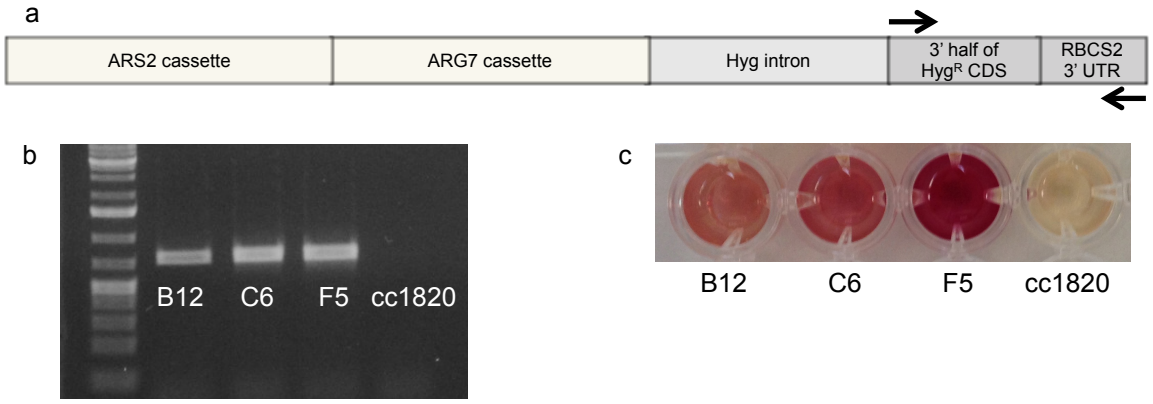


Figure 7.2 Criteria for selecting recipient strains. a) Arrows indicate the location of primers annealing to pHR18, used to screen candidate algal colonies for full integration of the vector. **b**) PCR verification that the entire pHR18 vector integrated into the genome in three candidate recipient strains, using the primers indicated in **a**. Parental strain cc1820 does not contain the cassette. **c**) Arylsulfatase assays on the candidate recipient strains, indicating that they all constitutively express the ARS2 gene, unlike cc1820.

All three candidate recipient strains were tested in preliminary experiments. Hypothesizing that the intensity of ARS2 expression may reflect physical accessibility of the recipient cassette's genomic locus, the basal rate of all three of these strains was tested, but we did not find a correlation between HR rate and ARS2 intensity (data not shown). Strain B12 was chosen for all subsequent experiments because its HR rate was slightly higher than that of the others (approximately two-fold difference). A high-throughput analysis strategy was developed such that thousands of transformants can be pooled and screened at once (Figure 3), with no additional screening effort required after the transformation and plating on selection.

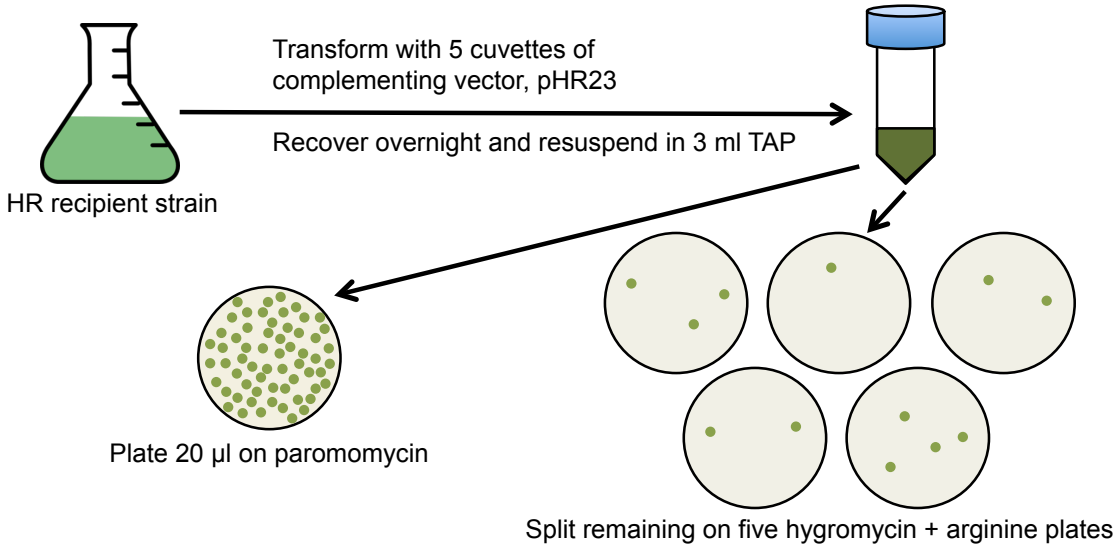


Figure 7.3 Schematic of the experimental protocol for high-throughput identification of homologous recombinants. Each recovery flask of five pooled transformations was resuspended in 3 ml of TAP media, and 20 µl was plated onto paromomycin, to measure the total transformation efficiency. The rest was split evenly onto five hygromycin plus arginine plates to select for homologous recombinants in which the truncated hygromycin resistance cassette was repaired by homologous recombination.

In order to obtain a reliable basal HR rate, especially given the variability observed in transformation efficiency between independent electroporations, the number of recombinant transformants must be compared to the overall transformation efficiency of a given transformation. The intact paromomycin cassette in pHR23 can be used to measure overall

transformation efficiency, as it will produce a paromomycin-resistant colony regardless of insertion site within the genome. Overall efficiency was measured directly by plating an aliquot of the pooled transformation recovery onto paromomycin. Thus, the number of theoretical transformants challenged on hygromycin can be calculated based on the number of colonies obtained on paromomycin, allowing the HR rate to be expressed as the ratio of homologous recombinant events to random insertion events. For each set of 5 cuvettes, the colonies on paromomycin (from 20 µl of transformation recovery) and on hygromycin (from the remaining 2.98 ml of transformation recovery) were counted and then the HR rate calculated as follows:

$$\frac{[\text{paro colonies}]}{20\ \mu l} \times 2980\ \mu l = [\text{theoretical transformants tested}]$$

$$\frac{[\text{observed hygromycin colonies}]}{[\text{theoretical transformants tested}]} = \text{HR efficiency rate}$$

We established that our system can reliably screen for homologous recombination events and that the basal HR rate is indeed reproducible across many transformations, indicating robustness to slight variations in culture and transformation conditions. Four independent experiments were performed in recipient strain B12, each with 25 electroporation cuvettes. Table 7.1 shows that across all four experiments, the basal homologous recombination rate is relatively stable, at $6.0 \pm 1.8 \times 10^{-5}$. Notably, this rate does not change even when the overall transformation efficiency is considerably lower, as in the fourth experiment listed in Table 7.1. This low transformation efficiency is likely attributable to the culture having been split multiple times in liquid instead of being grown directly from a plate, which is commonly observed to result in low transformation efficiency; however, the homologous recombination rate remained unchanged. All colonies obtained in the initial hygromycin selection were subsequently patched onto a 33% higher concentration of hygromycin (20 µg/ml instead of 15 µg/ml) to confirm that they were not by-products of abnormal plating density or faulty antibiotic strength, prior to calculating the HR

rate. Greater than 95% of all patched colonies survived on the challenge concentration of hygromycin, indicating that the presence of spurious colonies obtained in the initial screen is very low and thus reaffirming the high fidelity of this system.

Table 7.1: Reproducibility of basal HR efficiency measurement

Experiment number	Colonies obtained on paromomycin	Theoretical transformants screened on hygromycin	Colonies obtained on hygromycin	Homologous recombination rate
B1	1659	247,191	21	8.5×10^{-5}
B2	1929	287,421	16	5.6×10^{-5}
B3	2640	393,360	16	4.1×10^{-5}
B4	457	68,093	4	5.9×10^{-5}
Average basal homologous recombination rate in strain B12				$6.0 \pm 1.8 \times 10^{-5}$

Analysis of recombinant transformants reveals insertions, deletions, and rearrangements within the homology region

To characterize the nature of the recombination and provide some mechanistic elucidation, 44 of the hygromycin resistant clones were selected for further analysis. The system is designed such that three types of events can produce hygromycin-resistant transformants: a double crossover, resulting in the outcome depicted in Figure 7.4a; a single crossover within the intron after which the branch migrates to the end of the pHR23 template and then recombines non-homologously (Figure 7.4b); and a single crossover within the 5' homology region and subsequent non-homologous repair at the opposite end (Figure 7.4c). In the last instance, the site of linearization within pHR23 (see arrow in Figure 7.1b) is critical because the closing splice site is removed during linearization, allowing additional sequence downstream to be recognized as an extension of the intron.

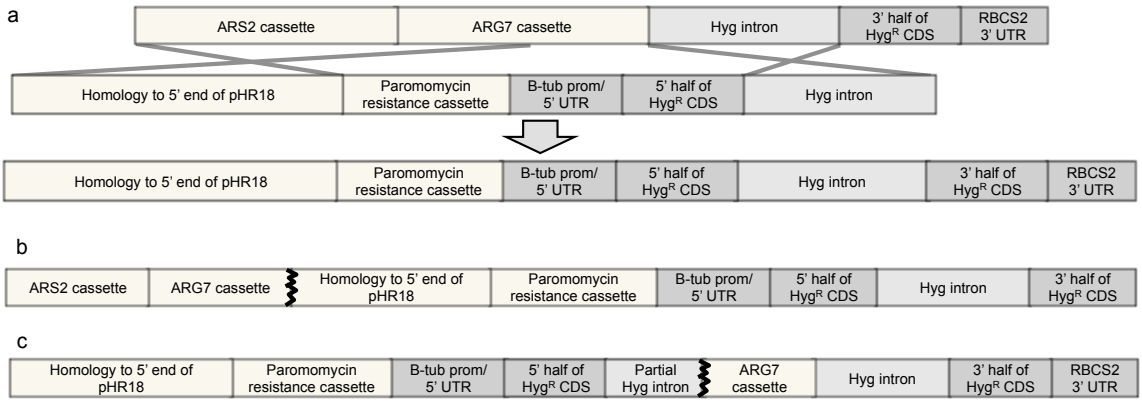


Figure 7.4 Predicted outcomes for achieving hygromycin-resistant transformants. a) Schematic of the expected recombination site after a perfect double recombination event. **b**) Schematic of the expected recombination site after a single recombination within the 5' homology region. The jagged border between the ARG7 cassette and the 5' homology indicates that this junction would have to be formed non-homologously after branch migration from the intron crossover reached the end of the pHR23 template. Bacterial backbone sequence (not pictured) may also be incorporated at this site. **c**) Schematic of the expected recombination site after a single recombination within the 3' homology region. Again, the jagged border indicates a region of non-homologous junction.

Primers that span the intron within the hygromycin resistance gene were used to amplify the recombination junction (Figure 7.5), and the resulting fragments were analyzed by restriction digest and, when possible, by Sanger sequencing. The fragments were cloned into a bacterial vector to assist with both restriction analysis and sequencing, as the PCR amplicons often exhibited low yield due to their large size and GC-rich nature. In approximately a third of the recombinants analyzed, the recombination site within the hygromycin resistance cassette appeared exactly as predicted from a crossover within that homology region, with no associated insertions or deletions. Restriction analysis with sites flanking the fragment as well as a site internal to the ARG7 cassette (Figures 7.5b and 7.5c) indicated that 16 of the 44 clones analyzed (36%) exhibited perfect recombination, producing fragments of the exact size predicted for a recombination event resembling Figure 7.5a. These transformants were all subsequently verified by Sanger sequencing. However, insertions, deletions, and rearrangements or concatamerizations

within the recombination site were also observed (Figure 7.5d and Table S1, supporting information).

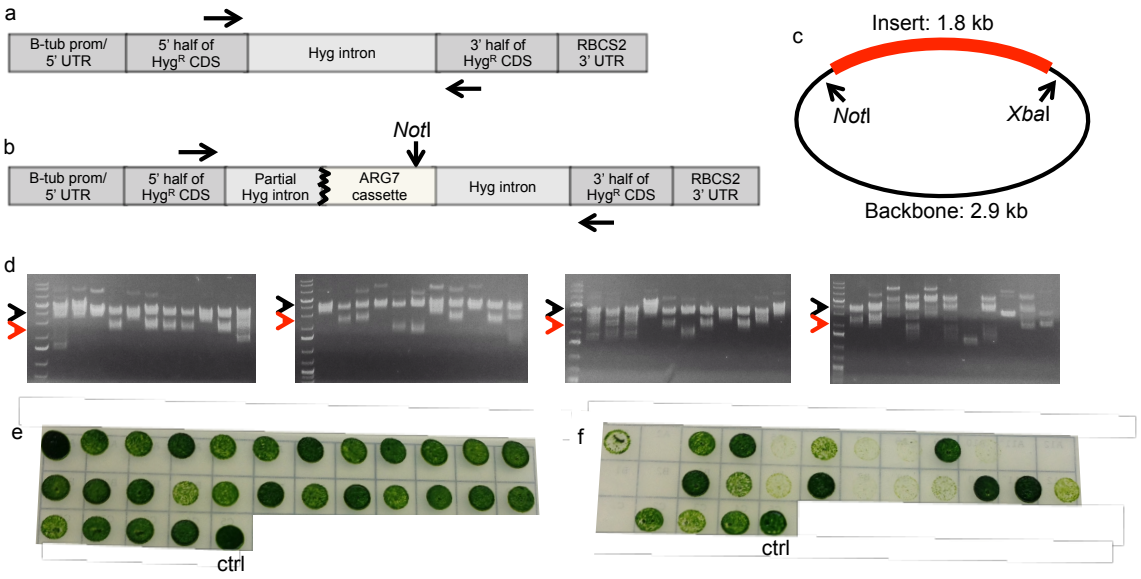


Figure 7.5 Characterization of the nature of individual recombination events. a) Anticipated form of the hygromycin cassette resulting from a double recombination event or single recombination within the intron. Arrows indicate the annealing site of primers used to amplify the recombination site. **b**) Anticipated form of the cassette resulting from a single recombination event within the 5' homology region, and primer binding sites for amplification. The vertical arrow indicates a *Not*I site near the 3' end of the ARG7 cassette. **c**) The CloneJet backbone has a *Not*I site and an *Xba*I site on either side of the insert. Expected bands are shown for the case of perfect recombination at the intron site, as depicted in **a**. **d**) Restriction digest analysis of CloneJet plasmids containing PCR products amplified by the primers indicated in **a** and **b**. A transformant exhibiting perfect recombination within the intron (as depicted in **5a**) would have 2.9 kb and 1.8 kb bands, indicated by black and red arrowheads respectively. A transformant resembling **b** would have 4.5 kb, 2.9 kb, and 2 kb fragments. Additional bands or bands of different sizes than predicted indicate concatamerization, rearrangements, or partial deletions. **e**) Growth of 28 independent hygromycin-resistant transformants on TAP media with arginine added, and **f**) without supplemented arginine. The last strain is a wild type arginine prototrophic control (ctrl).

Because these irregularities and repair errors occurred within the intron splice sites, they did not impair the functionality of the hygromycin resistance gene. Complete Sanger sequencing verification of several of these constructs proved difficult due to the concatamerizations indicated by multiple bands in Figure 7.5b, resulting in multiple annealing sites for all internal primers that could otherwise be used to obtain sequence coverage across the entire fragment. Those for which

full sequencing was obtained are fully described in Table S1 in the supporting information. In at least three transformants analyzed, a portion of the intron was deleted in the process of recombination; in at least one of these, this deletion was accompanied by a very small insertion of novel DNA not traceable to the vector (sequence R3-14 in the supporting information). At least one recombinant (sequence R3-07 in supporting information) fit the model from Figure 7.4c perfectly, in which the non-homologous junction between the partial intron and the ARG7 cassette experienced loss of sequence on both sides prior to repair. Several other recombinants exhibit integration of multiple concatamers or rearrangements resulting in multiple bands with restriction analysis.

Upstream of the hygromycin resistance cassette, imperfect recombination within the 5' homology region could abolish functionality of the ARG7 cassette without affecting hygromycin resistance. Because the homology region at this end was too long to span effectively with PCR for sequencing, 28 of the transformants were assayed phenotypically for arginine auxotrophy. 16 of these clones had mostly or entirely lost the ability to grow without supplemented arginine (Figures 7.5e and 7.5f), indicating that in about half of the recombinants, the ARG7 gene within the recipient cassette was disrupted. This does not necessarily indicate that homologous recombination occurred within the 5' homology region, as the ARG7 cassette could have been partially deleted during the non-homologous end-joining following a single crossover in the intron (as in Figure 7.4b).

Homologous recombination detection system can validate multiple approaches for improving HR efficiency

After establishing a reliable basal HR rate, we wished to evaluate two approaches previously reported to increase the HR rate in *C. reinhardtii*. In 1993, Sodeinde and Kindle reported that transformation by particle bombardment using a biolistic chamber increased the

prevalence of HR to as much as one in 24 transformants (*3*). We tested this assertion using the validated B12 strain with both linearized and circular pHR23 coated onto gold particles. Akin to the electroporations described above, a portion of the recipient strain culture was plated onto paromomycin prior to transformation by particle bombardment and the rest was plated onto hygromycin with arginine. The overall transformation efficiency was much lower than with electroporation, as expected; however, if the HR rate was truly on the order of one in 24 transformants, we should have easily been able to observe it within over 1,300 theoretical transformants screened. Surprisingly, no colonies were obtained on hygromycin from any of the four particle bombardment experiments (Table S2 in supporting information).

Secondly, we wished to validate previous reports of using single-stranded DNA to reduce non-homologous integration events. These studies typically used either asymmetric PCR or phage to produce the single-stranded DNA for transformation (*7*). Neither of these methods are ideal for creating single-stranded pHR23 due to the construct's size (over 8 kb for the minimum region; over 11 kb for the whole plasmid) and difficult amplification, attributable to high GC content throughout and repeat regions especially within the intron. Furthermore, previous attempts to transform heat-denatured DNA did not elicit a significant increase in HR efficiency (*6*). As an alternative, Exonuclease III was used to selectively degrade one strand of the double-stranded plasmid (*22*).

Because the restriction sites used to prepare this substrate for Exonuclease III were different than the standard preparation of pHR23 (which involves linearization by a single enzyme rather than two), the control for this experiment was the substrate prepared identically except without Exonuclease III treatment. The HR rate for this preparation of double-stranded pHR23 was significantly lower (7.2×10^{-6}) than the basal rate observed with the standard pHR23 preparation (6.0×10^{-5}). Transformation with the Exonuclease III-treated single-stranded DNA resulted in nearly a ten-fold improvement over the control, bringing the HR efficiency to $6.5 \times 10^{-}$

[5]. While this improvement is slightly less dramatic than previous studies have observed, it does indicate that single-stranded DNA may be a preferable substrate for encouraging HR prevalence relative to NHEJ.

Finally, we tested a novel approach for increasing HR efficiency in algae: incubating the substrate DNA with a purified bacterial RecA recombinase prior to electroporation. Three different temperatures were tested: 60°C, 75°C, and 90°C. 60°C and 75°C are within the optimal range of the *Tth* RecA, but we hypothesized that it should remain stable at even higher temperatures due to its origin from *Thermus thermophilus*. RecA was mixed with double-stranded linearized plasmid DNA, but RecA binds preferentially to single-stranded DNA; therefore, higher temperatures seemed more likely to transiently denature the ends of the plasmid to facilitate RecA binding, so incubation at 90°C was also tested.

Incubation with 1 µg RecA per five cuvettes at 75°C for half an hour prior to transformation resulted in an HR efficiency of 2.6×10^{-4}, which is approximately 4.3-fold better than the basal rate. Increasing the amount of RecA to 4 µg per five cuvettes increased the HR efficiency even further, to 3.9×10^{-4}, which is approaching an order of magnitude improvement over the basal HR rate.

Discussion

Previous studies have reported highly variable numbers for measurements of HR efficiency in *C. reinhardtii*, and these numbers likely depend on genomic target context, length of homology, growth conditions, and other factors. We have designed a system in which HR can be measured reliably, reproducibly, and in a high-throughput fashion, allowing side-by-side comparison of several approaches for increasing the efficiency of gene targeting within the *C. reinhardtii* nuclear genome. Effective approaches can be identified in the recipient strain described here, and subsequently applied to naïve strains for targeted knockout or insertion of

genes of interest in the desired genetic background. Indeed, introduction of single-stranded DNA has proven to be one effective approach, as is the pre-incubation of DNA with a thermostable purified RecA recombinase.

The vectors described in this work (pHR18 and pHR23) represent the final iteration of several generations of modification and testing to arrive at these designs. Throughout all generations, hygromycin was chosen as the truncated selectable marker for two reasons: first, the enzymatic mechanism of action of the hygromycin resistance gene product allows colonies to be obtained even with relatively low expression, in contrast to an antibiotic resistance like zeomycin resistance in which the *Sh-ble* gene must be expressed at high levels to obtain a viable colony. Secondly, very few false hygromycin-resistant background colonies are ever observed, in agreement with previous observations (*23*), making it an ideal selectable marker for obtaining reliable HR rates based on counting colonies without further validation or characterization.

In the first generation, a truncated green fluorescent protein (GFP) cassette was located at the 5' end of the vectors, to create an intact functional GFP cassette if double recombination occurred. We anticipated this would allow detection by flow cytometry for recombination at the 5' end, along with hygromycin selection for resistance at the 3' end. However, preliminary experiments using a control vector containing the entire GFP cassette with the inserted artificial intron were unable to detect GFP by microscopy, fluorescent plate-reader, or by flow cytometry with sufficient signal for use in a high-throughput screen. In future attempts, this strategy could be made feasible by using one of the recently developed nuclear fluorescent reporters with a higher signal to noise ratio than GFP (*24*), and/or through transcriptional linkage to a selectable marker (*25*).

In the second generation of vectors, the truncated GFP cassette was replaced with a truncated ARS2 cassette, constitutively expressing a cDNA-derived endogenous arylsulfatase (Specht *et al*., under review). We have previously characterized this cassette as a robust and

reliable reporter using a plate-reader assay requiring only 500 µl of liquid algal culture. However, the requirement for small liquid cultures of individual clones makes this reporter unfeasible for screening hundreds of thousands of transformants, but still provides a useful method for testing double recombination in the colonies obtained from hygromycin selection. This recipient vector was transformed into cc1820, and six PCR-positive strains were transformed with the complementing vector containing the other halves of the ARS and hygromycin resistance cassettes. After screening several hundred thousand transformants in the six different recipient strains, no hygromycin-resistant colonies were obtained. Due to strong gene silencing and positional effects commonly observed in the nuclear genome, there was concern that the lack of recombinants may indicate that the recipient cassette had not integrated into the genome in a locus capable of sufficient expression to obtain hygromycin resistance, even if recombination did occur. Furthermore, if silencing is partially a consequence of local chromatin remodeling, strains in which the recipient cassette had been silenced may have had limited access to that region of the genome, reducing the likelihood of recombination at that site. Indeed, one previous study of HR in *C. reinhardtii* addressed this issue by ensuring that prior to transformation, cells were grown under conditions known to induce transcription of the target gene (*3*).

To alleviate these concerns, we decided to forego the ability to distinguish double recombination from single recombination using the truncated ARS cassette, and instead used an intact ARS cassette in the recipient vector to select strains in which the recipient cassette was highly expressed. To facilitate recombination, 4.4 kb from the 5' end of the recipient vector (encompassing the full ARS cassette and part of the ARG7 cassette) was added onto the 5' end of the complementing vector. This system finally allowed expression screening of potential recipient strains, and indeed less than 15% of all PCR-positive candidates expressed arylsulfatase at levels visible to the naked eye in the ARS assay (data not shown). This is the final version (pHR18) of the recipient strain creation vector that was used in this work.

A suitable recipient strain was selected using PCR to screen for full integration of the recipient cassette and arylsulfatase assays to confirm that the integration locus was actively transcribed and expressed. Using the high-throughput screening strategy depicted in Figure 7.3, this strain was then transformed repeatedly in independent experiments. This is the first reliable, reproducible measurement of the basal rate of HR efficiency in *C. reinhardtii* using a systematic approach. The complementing vector pHR23 was introduced as a linearized plasmid to increase overall transformation efficiency. The linearization site near the end of the intron in the hygromycin resistance gene (arrow in Figure 7.1b) was chosen strategically: previous work in yeast and mammalian cells indicates that HR can be encouraged by linearizing the plasmid within the homology region (*26, 27*). Furthermore, eliminating the closing splice site allowed additional sequence inserted non-homologously to be recognized as an extension of the intron and therefore to be spliced out.

As stated previously, the introduction of an intron in the hygromycin cassette homology enabled imperfect recombination events to be captured, including those where DNA within the homology region was deleted or instances where additional DNA was inserted, often from integrating the rest of the plasmid after a single crossover. The variability in the nature of the recombination events (i.e. presence of insertions or deletions and single vs. double cross-over) was anticipated based on previous research in other organisms and in *C. reinhardtii*. Nelson and Lefebvre characterized recombinants in which one side inserted homologously and the rest of the plasmid integrated non-homologously – often with multiple insertions with varying degrees of degradation prior to integration – as well as genomic deletions upstream or downstream of the recombination site (*4*). Indeed we observe this same phenomenon, and the regions of degradation are often predictable based on the models for single crossover events depicted in Figure 7.4.

After confirming the validity of the system for obtaining and characterizing individual recombinants, we then tested three approaches for increased the HR rate. The lack of

recombinants obtained using particle bombardment was initially surprising because of strong disagreement with a previous report of very high (one in 24 transformants) HR efficiency using this transformation method (*3*). However, closer analysis of the literature reveals several possible explanations for a mistakenly high HR rate estimation. That study targeted the NIT1 locus in a *nit1* mutant strain whose molecular basis was uncharacterized but known to revert spontaneously (albeit with low frequency). The construct used to repair the non-functional NIT1 gene was comprised of essentially the entire NIT1 gene, missing only the promoter/5' UTR, and a small portion of the first exon. While the authors state that they did not observe any evidence to suggest colonies resulted from random placement of this fragment downstream of active promoters, it seems reasonable to conclude that some number of the colonies were attributable to promoter trapping. The fragment was not transformed into a wild-type strain as a control; if a similar number of colonies were obtained in the wild type, then the promoter trapping explanation would be very likely. Furthermore, the measurement for the overall transformation efficiency is not directly comparable because it was obtained by transforming a different plasmid altogether (the full, intact NIT1 gene) in separate transformations. Given that this construct appears to be at least 3 kb larger than the truncated version, the transformation efficiency difference due to size alone may have been considerable, mistakenly inflating the HR rate due to an unfairly low denominator in the HR to non-HR ratio. Due to the low overall transformation efficiency obtained with particle bombardment, we ceased to pursue this avenue further once it was apparent that we were not obtaining an HR rate anywhere near one in 24 transformants.

Fortunately, the previous reports of the effectiveness of single-stranded DNA for gene targeting appear to be more reproducible, though our findings indicate it is slightly less effective than previous studies have found. Zorin *et al*. found that single-stranded DNA reduced NHEJ by up to 100-fold, thus increasing the relative proportion of HR events (*3, 5, 7*). In comparison, our data indicate an approximate 10-fold increase in the HR:NHEJ ratio; however, Zorin *et al*. used

predominantly purified phagemid DNA, whereas we use Exonuclease III degradation of one strand to produce single-stranded DNA. Hence, their substrate was usually a circular single strand, while ours is linear. This may have contributed to higher instability of our construct, and limiting the time it remained intact within the cell and/or the preservation of the full region of homology. Another study attempted to increase HR efficiency by heat-denaturation to achieve single-stranded DNA but did not observe a significant difference, which may be attributable to re-annealing of the strands prior to or during transformation (*6*). We also attempted transformation with heat-denatured single-stranded DNA and similarly did not obtain a significant increase in HR efficiency (data not shown), and subsequently switched to the Exonuclease III method for producing single-stranded DNA. It is unclear why the double-stranded substrate for Exonuclease III (pHR23 cut with two restriction enzymes instead of linearized with a single enzyme) exhibited a much lower HR rate than the standard pHR23 preparation. One possible explanation is that restriction sites with long 3' overhangs suitable for protecting one strand from Exonuclease III degradation are rather limited within the pHR23 plasmid, so the site we chose was an *Sbf*I site immediately downstream of the intron rather than within it. Because of the observation mentioned previously that linearization within the homology is preferable for facilitating HR, this observation may indicate that simply linearizing near the homology is insufficient to provide this boost.

Finally, we attempted a novel method for increasing HR efficiency that had not previously been reported in *C. reinhardtii* but was based on literature in multicellular plants. In tobacco protoplasts, expression of a nuclear-targeted prokaryotic RecA conferred resistance to chemical mutagenesis, indicating that a prokaryotic recombinase could function within plant nuclei to repair DNA damage (*28*). We used a commercially available RecA rather than engineering a nuclear-targeted RecA gene into *C. reinhardtii* because we are interested in developing approaches that can be applied easily to any strain. A thermostable RecA recombinase

– *Tth* RecA from *Thermus thermophilus* – was used to tolerate incubation with the substrate DNA at a temperature likely to allow partial strand separation, enabling the RecA to bind to the single-stranded ends akin to its function *in vivo* where it binds the resected, single-stranded ends of DNA (see (*29*) for a review of the machinery involved in homology-directed repair). Recently it has been shown that RecA can also contribute to branch migration of Holliday junctions, potentially enabling incorporation of longer stretches of DNA from an introduced template (*30*). As the tobacco studies showed, higher plants do contain homologs of the machinery required to interact with and utilize a prokaryotic RecA, and this observation appears to hold true for algae due to the increased HR rate observed with the addition of a small amount of bacterial RecA protein. Additional approaches from plants may prove suitable in algae, as well as novel strategies such as altering growth conditions, synchronizing cell division, and others that are particularly amenable for single-celled organisms. With the availability of a reliable and high-throughput screen for measuring HR, the algal community can now rapidly assess the efficacy of many strategies to facilitate advanced engineering of the algal nuclear genome.

Acknowledgements

We thank Michael Hannon and Michael Mendez for many useful discussions during the design of this system, and Adam Labadorf and Mark Rogers from the Ben-Hur lab at Colorado State University for graciously providing a compiled list of constitutively spliced *Chlamydomonas* introns. We also thank Farzad Haerizadeh and his research team at Life Technologies for their support and collaboration throughout this work. This work was funded by the Department of Energy grant DE-EE0003373. EAS was supported by a National Science Foundation graduate research fellowship, and HHN-E was supported by the Danish Research Council for Technology and Production (FTP) grant 09-065827/274-08-0354.

Chapter 7, in full, is a reprint of material in preparation for submission for publication. The dissertation author is the primary investigator and author of this paper.

References

1. E. Nakazato, H. Fukuzawa, S. Tabata, H. Takahashi, K. Tanaka, Identification and expression analysis of cdna encoding a chloroplast recombination protein rec1, the chloroplast reca homologue in chlamydomonas reinhardtii. *Biosci Biotechnol Biochem* **67**, 2608 (Dec, 2003).

2. V. I. Shalguev, O. K. Kaboev, I. A. Sizova, P. Hagemann, V. A. Lanzov, Identification of chlamydomonas reinhardtii rad51c: Recombinational characteristics. *Mol Biol+* **39**, 98 (Jan-Feb, 2005).

3. O. A. Sodeinde, K. L. Kindle, Homologous recombination in the nuclear genome of chlamydomonas reinhardtii. *Proc Natl Acad Sci U S A* **90**, 9199 (Oct 1, 1993).

4. J. A. E. Nelson, P. A. Lefebvre, Targeted disruption of the nit8 gene in chlamydomonas-reinhardtii. *Mol Cell Biol* **15**, 5762 (Oct, 1995).

5. N. J. Gumpel, J. D. Rochaix, S. Purton, Studies on homologous recombination in the green-alga chlamydomonas-reinhardtii. *Curr Genet* **26**, 438 (Nov-Dec, 1994).

6. W. Mages, O. Heinrich, G. Treuner, D. Vlcek, I. Daubnerova, M. Slaninova, Complementation of the chlamydomonas reinhardtii arg7-8 (arg2) point mutation by recombination with a truncated nonfunctional arg7 gene. *Protist* **158**, 435 (Oct 23, 2007).

7. B. Zorin, P. Hegemann, I. Sizova, Nuclear-gene targeting by using single-stranded DNA avoids illegitimate DNA integration in chlamydomonas reinhardtii. *Eukaryot Cell* **4**, 1264 (Jul, 2005).

8. I. Sizova, A. Greiner, M. Awasthi, S. Kateriya, P. Hegemann, Nuclear gene targeting in chlamydomonas using engineered zinc-finger nucleases. *Plant J* **73**, 873 (Mar, 2013).

9. D. Gonzalez-Ballester, W. Pootakham, F. Mus, W. Q. Yang, C. Catalanotti, L. Magneschi, A. de Montaigu, J. J. Higuera, M. Prior, A. Galvan, E. Fernandez, A. R. Grossman, Reverse genetics in chlamydomonas: A platform for isolating insertional mutants. *Plant Methods* **7**, (Jul 27, 2011).

10. W. R. A. Brown, N. C. O. Lee, Z. Y. Xu, M. C. M. Smith, Serine recombinases as tools for genome engineering. *Methods* **53**, 372 (Apr, 2011).

11. D. P. Long, B. Tan, A. C. Zhao, L. X. Xu, Z. H. Xiang, [progress in cre/lox site-specific recombination system in higher eukaryotes]. *Yi Chuan* **34**, 177 (Feb, 2012).

12. C. Monetti, K. Nishino, S. Biechele, P. Z. Zhang, T. Baba, K. Woltjen, A. Nagy, Phic31 integrase facilitates genetic approaches combining multiple recombinases. *Methods* **53**, 380 (Apr, 2011).

13. T. Xu, Y. Li, J. D. Van Nostrand, Z. He, J. Zhou, Cas9-based tools for targeted genome editing and transcriptional control. *Applied and environmental microbiology*, AEM. 03786 (2014).

14. L. Cong, F. A. Ran, D. Cox, S. L. Lin, R. Barretto, N. Habib, P. D. Hsu, X. B. Wu, W. Y. Jiang, L. A. Marraffini, F. Zhang, Multiplex genome engineering using crispr/cas systems. *Science* **339**, 819 (Feb 15, 2013).

15. Z. Y. Feng, B. T. Zhang, W. N. Ding, X. D. Liu, D. L. Yang, P. L. Wei, F. Q. Cao, S. H. Zhu, F. Zhang, Y. F. Mao, J. K. Zhu, Efficient genome editing in plants using a crispr/cas system. *Cell Res* **23**, 1229 (Oct, 2013).

16. Z. S. Yu, M. D. Ren, Z. X. Wang, B. Zhang, Y. K. S. Rong, R. J. Jiao, G. J. Gao, Highly efficient genome modifications mediated by crispr/cas9 in drosophila. *Genetics* **195**, 289 (Sep, 2013).

17. F. Geu-Flores, H. H. Nour-Eldin, M. T. Nielsen, B. A. Halkier, User fusion: A rapid and efficient method for simultaneous fusion and cloning of multiple pcr products. *Nucleic Acids Res* **35**, e55 (2007).

18. H. H. Nour-Eldin, B. G. Hansen, M. H. Norholm, J. K. Jensen, B. A. Halkier, Advancing uracil-excision based cloning towards an ideal technique for cloning pcr fragments. *Nucleic Acids Res* **34**, e122 (2006).

19. S. M. Newman, J. E. Boynton, N. W. Gillham, B. L. Randolphanderson, A. M. Johnson, E. H. Harris, Transformation of chloroplast ribosomal-rna genes in chlamydomonas - molecular and genetic-characterization of integration events. *Genetics* **126**, 875 (Dec, 1990).

20. A. Labadorf, A. Link, M. F. Rogers, J. Thomas, A. S. N. Reddy, A. Ben-Hur, Genome-wide analysis of alternative splicing in chlamydomonas reinhardtii. *Bmc Genomics* **11**, (Feb 17, 2010).

21. R. Loppes, A new class of arginine-requiring mutants in chlamydomonas reinhardi. *Mol Gen Genet* **104**, 172 (1969).

22. A. J. H. Smith, Use of exonuclease-iii for preparing single stranded DNA for use as a template in the chain terminator sequencing method. *Nucleic Acids Res* **6**, 831 (1979).

23. V. G. Ladygin, The transformation of the unicellular alga chlamydomonas reinhardtii by electroporation. *Microbiology+* **72**, 585 (Sep-Oct, 2003).

24. B. A. Rasala, D. J. Barrera, J. Ng, T. M. Plucinak, J. N. Rosenberg, D. P. Weeks, G. A. Oyler, T. C. Peterson, F. Haerizadeh, S. P. Mayfield, Expanding the spectral palette of fluorescent proteins for the green microalga chlamydomonas reinhardtii. *The Plant journal : for cell and molecular biology* **74**, 545 (May, 2013).

25. B. A. Rasala, P. A. Lee, Z. Shen, S. P. Briggs, M. Mendez, S. P. Mayfield, 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 (2012).

26. M. R. Capecchi, Altering the genome by homologous recombination. *Science* **244**, 1288 (Jun 16, 1989).

27. T. L. Orrweaver, J. W. Szostak, R. J. Rothstein, Yeast transformation - a model system for the study of recombination. *P Natl Acad Sci-Biol* **78**, 6354 (1981).

28. B. Reiss, M. Klemm, H. Kosak, J. Schell, Reca protein stimulates homologous recombination in plants. *Proc Natl Acad Sci U S A* **93**, 3094 (Apr 2, 1996).

29. H. Y. Niu, S. Raynard, P. Sung, Multiplicity of DNA end resection machineries in chromosome break repair. *Gene Dev* **23**, 1481 (Jul 1, 2009).

30. M. J. Rossi, O. M. Mazina, D. V. Bugreev, A. V. Mazin, The reca/rad51 protein drives migration of holliday junctions via polymerization on DNA. *Proc Natl Acad Sci U S A* **108**, 6432 (Apr 19, 2011).

CONCLUSION

Genetically engineered algae hold the potential to revolutionize bio-manufacturing from the energy sector to the medical sector, as these low-cost, fast-growing photosynthetic organisms are amenable to both large-scale industrial cultivation and small-scale research. Advances in algal engineering have historically been pioneered in *Chlamydomonas reinhardtii* and then applied to other strains and species, and this will likely continue to be the case. Therefore, intensive research is still needed to develop advanced tools in *Chlamydomonas*, including better methods of precisely modulating transgene placement and expression, to enable complex metabolic and multi-trait engineering.

The research described in this dissertation represents several critical steps towards engineering algae on par with the current industrial workhorses like bacteria and yeast. Synthetic biology approaches have now been applied to algal chloroplast engineering, to design novel gene regulatory sequences that evade endogenous negative feedback mechanisms. These findings have immediate impact in the growing field of algal therapeutic protein production, as higher levels of protein accumulation in the chloroplast are necessary to achieve economic viability in a commercial setting. In addition, a whole suite of nuclear engineering tools have been improved and developed to enable more advanced metabolic engineering. An enhanced reporter has demonstrated its utility in screening nuclear promoters for high signal and robust expression, and a highly-transformable endogenous selectable marker has been optimized and adapted to drive high transgene expression, obviating antibiotic selection for obtaining high-expressing transformants. The fact that both of these nuclear tools are derived from endogenous *Chlamydomonas* genes is significant because of complex genetically modified organism regulations that restrict foreign gene introduction. Furthermore, the development of a novel, high-throughput strategy for identifying homologous recombination events in *Chlamydomonas* with high fidelity finally offers hope for reverse genetics in algae to study targeted knockouts of genes of interest. Homologous recombination also potentially allows for identification and utilization of

neutral sites or indeed expression hotspots, to address the overwhelming clonal variability associated with random integration of transgenes.

Altogether, this body of work offers a substantial improvement to the availability of genetic engineering tools for both the nuclear and chloroplast genomes of *Chlamydomonas*, and opens the field for more rapid development of additional tools. For example, the regulatory sequence analysis methods described in Chapter 4 can easily be applied to additional endogenous UTRs to build the library of positive element "parts" that can be incorporated into novel chloroplast regulatory regions with a synthetic biology approach. The enhanced nuclear reporter described in Chapter 5 is ideal for screening for additional strong nuclear promoters, as well as for screening mutant libraries for strains with reduced nuclear silencing. Finally, the high-fidelity screening system described in Chapter 7 will be critical for testing many more approaches for increasing the HR efficiency in algae, to approach levels suitable for targeted knockout or insertion without selection. Now that the screening system has been developed and validated, years' worth of approaches to increase HR efficiency in other organisms can be tested in algae much more rapidly.

With these tools, eukaryotic green algae are increasingly valuable both for basic biological research – especially as a single-celled model for photosynthesis research – and for advanced industrial applications. Indeed, several commercial start-up endeavors within just the past decade are expounding upon the advantages of using an algal platform, and millions of dollars of both private and public money have been invested in algal research. From a biological standpoint, the lessons learned through tool development within this organism provide a roadmap for branching out into even more intractable organisms, reducing the severe research limitations imposed by the lack of malleability of non-model organisms.