12-2013

Transcriptomic Analyses of the CO_2-Concentrating Mechanisms and Development of Molecular Tools for *Chlamydomonas reinhardtii*

Andrew J. Brueggeman
University of Nebraska-Lincoln, ajbrueggeman@huskers.unl.edu

TRANSCRIPTOMIC ANALYSES OF THE CO_2-CONCENTRATING MECHANISM AND DEVELOPMENT OF MOLECULAR TOOLS FOR *CHLAMYDOMONAS REINHARDTII*

by

Andrew J Brueggeman

A Dissertation

Presented to the Faculty of

The Graduate College at the University of Nebraska

In Partial Fulfillment of Requirements

For the Degree of Doctor of Philosophy

Major: Biochemistry

Under the Supervision of Professor Donald P. Weeks

Lincoln, Nebraska

December, 2013

TRANSCRIPTOMIC ANALYSES OF THE CO_2-CONCENTRATING MECHANISM AND DEVELOPMENT OF MOLECULAR TOOLS FOR *CHLAMYDOMONAS REINHARDTII*

Andrew J Brueggeman, Ph.D.

University of Nebraska, 2013

Advisor: Donald P. Weeks

Microalgae, such as *Chlamydomonas reinhardtii,* account for a large percentage of photosynthesis that occurs on the planet. Many algae possess a Carbon-Concentrating Mechanism, or CCM, that actively transports inorganic carbon (C_i) into the cell to create artificially high internal levels of CO_2, enhancing their rate of carbon fixation. The production of biofuels from algal sources can serve as both a renewable and carbon-neutral energy source. This thesis details research in Chlamydomonas, in the effort to both better understand the CCM in algae and improve laboratory and industrial manipulations with algae.

In the first chapter of this thesis, changes in the Chlamydomonas transcriptome in response to changes in environmental CO_2 levels are investigated. In addition to witnessing changes in the transcription of 38% of Chlamydomonas genes, numerous patterns of genetic response are detailed, including a transient reduction in overall transcription that is later alleviated as cells became acclimated to the new environment. Our data also reveal a vast bidirectional promoter system in Chlamydomonas, detail

markedly different responses of gene families, and elucidated several putative binding motifs for transcriptional regulators involved in responses to low CO_2.

The remaining two chapters detail efforts to develop tools for use in Chlamydomonas. In the second chapter, drawing upon prior knowledge of herbicide resistance in higher plants, we developed several herbicide-resistant genes for optimal function in Chlamydomonas. As a result, we now have resistance markers for three classes of herbicides. These genes can be used either as selectable markers in transgenic experiments or as tools for maintaining axenic cultures in large-scale biofuel production facilities. The final chapter details efforts to develop an in-house method of synthesizing Transcription Activator-Like Effector (TALE) proteins to target specific DNA sequences in Chlamydomonas. Final TALE constructs were fused with FokI nuclease domains to create TAL Effector Nucleases (TALENs) that can create DNA double-strand breaks at specific sites. With an efficient method of synthesizing TALEN genes *de novo*, efforts can proceed to utilize this technology to create specific modifications within the Chlamydomonas genome.

Table of Contents

Dedication

I would like to dedicate this thesis to my grandfather, George H. Johnson, who, just as I do, held a passion for both science and agriculture.

It is my hope and prayer that you will fully realize what a wonderful world this is; that you will try to understand, in so far as you can, how things work and why. So many, if not all, things around us are governed by scientific or natural laws and so can be understood.

-George Henry Johnson, Personal Memoirs, 1910 - 2004

Acknowledgements

I would like to acknowledge the support given to me by my advisor, Dr. Donald Weeks, during my time in graduate school. He has been instrumental in helping me develop the skills and knowledge to succeed as a scientist. I also would like to thank members of the Weeks lab, including Kempton Horken, Wen Zhi Jiang, Amy Knobbe, & Thomas Plucinak, for taking the time to answer questions and critique ideas I have had. Much of my computational work has been assisted by Dr. Istvan Ladunga, and I would like to thank him for taking time to teach me the skills needed for much of my work. In addition, I appreciate the guidance provided by all members of my committee, Dr. Weeks, Dr. Ladunga, Dr. Becker, Dr. Fromm, and Dr. Riekhof, for their support and scientific guidance as I proceeded through my degree.

The University of Nebraska has been my home for several years now, and I have received support from numerous individuals both within and outside my department. During my time Lincoln, I have called on professors throughout the University for counseling and expertise, and every person was extremely hospitable and helpful in addressing whatever question I had. To the countless members of the University and the town of Lincoln; your advice, assistance, and friendship have made my time as a Husker memorable.

My efforts here, of course, could not have been possible without the constant support of my family. I have been so very fortunate to have received encouragement from all of my family members during my pursuit of my Ph.D. My parents, Douglas and

Bonita, have shown nothing but support as I have toiled on my degree, sharing in excitement when I succeeded, and offering encouragement when I struggled.

List of Figures

List of Abbreviations

CO_2 ..... Carbon Dioxide

C_i ..... Inorganic Carbon

HCO_3^- ..... Bicarbonate

O_2 ..... Molecular Oxygen

Rubisco ..... Ribulose-1,5-bisphosphate carboxylase oxygenase

CA ..... Carbonic Anhydrase

PPM ..... Parts Per Million

Protox, PPO ..... Protoporphyrinogen IX Oxidase

ED ..... Effective Dose

EPSPS ..... 5-enolpyruvylshikimate-3-phosphate (EPSP) synthase

PEP ..... Phosphoenolpyruvate

AMPA ..... Aminomethylphosphonic acid

POEA ..... Polyoxyethyleneamine

GOX ..... Glyphosate Oxidoreductase

PDS ..... Phytoene Desaturase

FAD ..... Flavin Adenine Dinucleotide

DSB ..... Double-Stranded Break

NHEJ ..... Non-Homologous End Joining

HR ..... Homologous Recombination

ZFN ..... Zinc Finger Nuclease

TALEN ..... Transcription Activator-Like Effector (TALE) Nuclease

NLS ..... Nuclear Localization Signal

AD ..... Activation Domain

RVD ..... Repeat Variable Diresidue

dNTP..... Deoxynucleotide Triphosphate

CRISPR ..... Clustered Regularly Interspaced Short Palindromic Repeats

GMO..... Genetically Modified Organism

CHAPTER 1 - Literature Review

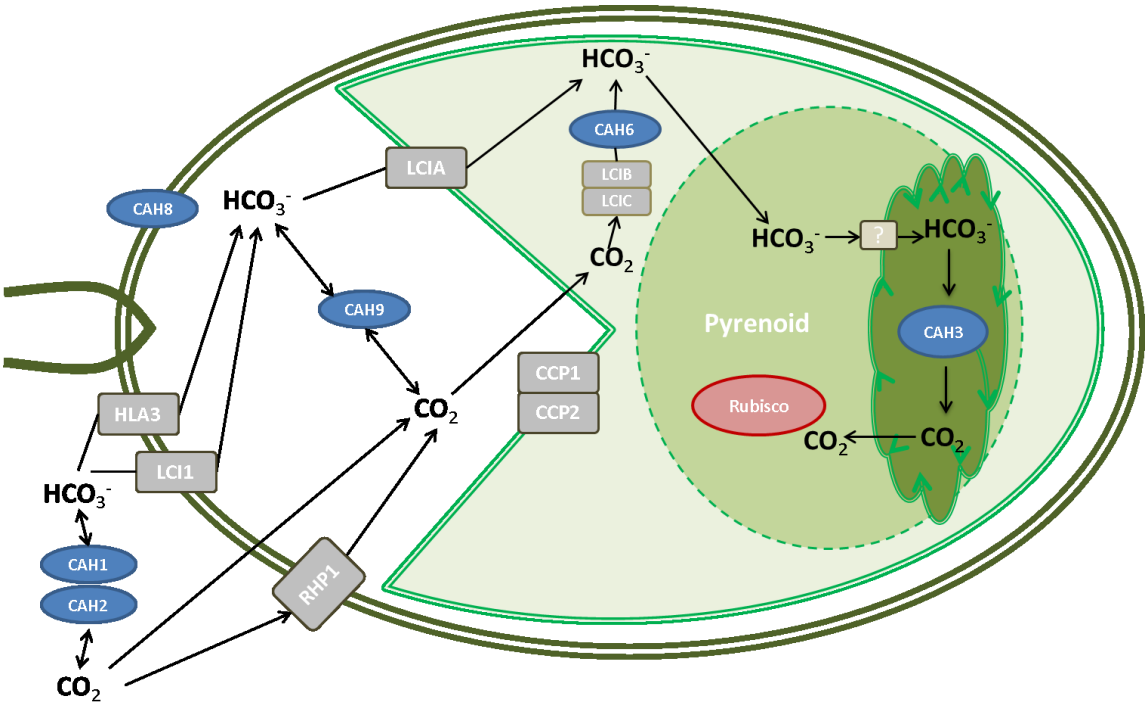
CHAPTER 1.1 - Chlamydomonas reinhardtii CCM

Photosynthetic observations with *Chlamydomonas reinhardtii* (hereafter, Chlamydomonas) revealed that the organism had higher photosynthetic affinity for CO_2 when placed in low-CO_2 (≤100ppm) environments, compared to high-CO_2 environments (30,000ppm)(Berry et al., 1976). This characteristic in low-CO_2 environments was accompanied by a reduction in the O_2 sensitivity of photosynthesis. However, studies of the ribulose-1,5-bis-phosphate carboxylase oxygenase (Rubisco) protein in Chlamydomonas showed no difference in its catalytic efficiency, ruling out the possibility of a more efficient form of Rubisco. This evidence, in addition to elevated levels of C_i found in Chlamydomonas in low-CO_2 conditions, pointed to the possibility that Chlamydomonas possessed a mechanism for concentrating C_i internally. Further studies revealed that, like cyanobacteria, Chlamydomonas increases the levels of CO_2 surrounding Rubisco by actively taking up C_i through a series of transporters coupled with interconversion of CO_2 and HCO_3^-, and localizes both the C_i and Rubisco to a microcompartment. Several components of the Chlamydomonas CCM have been elucidated, including regulatory factors, structural modifications, and various proteins, including carbon transporters and carbonic anhydrases. Overviews of the physiological mechanisms and known components of the Chlamydomonas CCM have been provided in several articles and reviews, including the most recent by Fang et al. (Fang et al., 2012). Here I will provide a brief summary of the Chlamydomonas CCM as it pertains to our analyses of transcriptome changes that occur when Chlamydomonas is shifted from

an environment replete in CO_2 to one in which CO_2 concentration becomes limiting to cell growth and survival.

In general, when exposed to C_i-limited environments, Chlamydomonas will utilize transporters and carbonic anhydrases to actively accumulate both CO_2 and HCO_3^-, utilizing various CO_2 and HCO_3^- transporters and energy to shuttle C_i into the thylakoid lumen. In the lumen, HCO_3^- is dehydrated to CO_2, which diffuses out of the lumen and into the pyrenoid structure (surrounding at least a portion of the luminal membrane structure), where it is used as a substrate by Rubisco.

Figure 1.1 - Chlamydomonas reinhardtii Carbon Concentrating Mechanism. The above representation of the Chlamydomonas CCM depicts most of the recognized CCM components. For brevity, some of the lesser-characterized putative CCM components have not been included in this figure. Blue ovals represent carbonic anhydrases, grey rectangles represent C_i transporters or scavengers, and a red oval represents Rubisco.



Ci Uptake

Like cyanobacteria, Chlamydomonas actively imports C_i as both CO_2 and HCO_3^-. As a charged molecule, HCO_3^- requires transporters to migrate across the membranes of Chlamydomonas (Spalding, 2008). The uncharged nature of CO_2 allows it to diffuse through the membranes (Prins and Elzenga, 1989). This diffusion may be assisted by aquaporins (Calamita et al., 1997)and Rhesus proteins (Soupene et al., 2004; Kustu and Inwood, 2006). However, maintenance by the cell of a higher internal concentration of C_i than the external environment is an endergonic reaction requiring energy from photosynthesis. Conversion of CO_2 to HCO_3^- by carbonic anhydrases in the cytoplasm and chloroplast stroma plays a significant role in preventing loss of C_i from the cell due to outward diffusion of CO_2.

Ci Uptake across the Plasma Membrane

Currently, two putative Chlamydomonas HCO_3^- transporters have been identified in the plasma membrane: HLA3 and LCI1 (Duanmu et al., 2009a; Ohnishi et al., 2010). The corresponding genes for these two transporters have both shown an increase in expression in response to carbon-depletion.

HLA3 has been annotated as an ATP Binding Cassette (ABC) transporter with homology to the multi-drug-resistance (MRP) sub-family of proteins (Im and Grossman, 2002). The presence of a complete ABC-HRP domain arrangement makes HLA3 a likely candidate as a single component transporter, not requiring dimerization or complex formation. Duanmu et al. (2009a) utilized RNAi to knock down transcript levels of HLA3.

The RNAi lines showed significant decreases in photosynthetic C_i affinity and uptake. This was especially notable under alkaline conditions (where HCO_3^- predominates as the C_i specie). Combinations of HLA3 knockdown with knockdown of other CCM components produced additional decreases in C_i affinity and uptake, further strengthening the suggestion that HLA3 plays a role as a HCO_3^- transporter in the plasma membrane. Like many CCM-associated genes, expression of HLA3 appears to be controlled by CIA5 (Xiang et al., 2001; Im and Grossman, 2002; Miura et al., 2004; Wang et al., 2005; Yamano et al., 2008; Fang et al., 2012). Mutants lacking the *CIA5* gene show no induction of HLA3 following a shift to low-CO_2 levels.

The second putative C_i transporter of the plasma membrane is LCI1. LCI1 has been characterized as localized to both the plasma and nuclear membranes (Ohnishi et al., 2010). LCI1 was artificially over expressed in the *lcr1* mutant (under conditions in which several components, including LCI1, fail to induce) to better characterize the potential function(s) of this protein. Increased LCI1 was shown to increase photosynthetic C_i affinity and uptake. Interestingly, however, was the resultant ambiguity regarding the carbon species that is transported by LCI1. LCI1 appeared to increase C_i affinity and accumulation at high pH, suggesting it may have a role as a HCO_3^- transporter. However, no differences were seen in light-dependent CO_2 gas exchange at different pH levels, leaving an ambiguity as to the C_i species preferred by this transporter. Like *HLA3*, *LCI1* was found to be a low-CO_2 inducible gene regulated by CIA5 (Burow et al., 1996; Im et al., 2003; Miura et al., 2004). Interestingly, studies by Yoshioka

et al. (2004) showed LCI1 to also be regulated by the gene *LCR1*, apparently downstream of *CIA5*.

C_i Uptake across the Chloroplast Inner Envelope Membrane

As noted above, following its uptake into the cell across the cellular membrane, C_i is largely hydrated to HCO_3^-, helping to reduce leakage of C_i from the cell. However, C_i still needs to enter the chloroplast to be utilized for carbon fixation. For HCO_3^-, several proteins may be critical for its transport into the chloroplast. Currently, four genes are being considered as putative chloroplast-localized bicarbonate transporters.

The first two putative bicarbonate transporters are CCP1 and CCP2 (Chloroplast Carrier Protein) (Chen et al., 1997; Pollock et al., 2004). The genes encoding these proteins are nearly identical in sequence, and are likely the result of a recent genome duplication event. Both have been shown to be induced by low-CO_2 conditions and to be regulated by CIA5. Unfortunately, no knockouts of these genes are currently available.

LCIA, or NAR1.2, is a member of the Formate/Nitrite Transporter (FNT) family of proteins. It shows perhaps the largest response to low-CO_2, being induced nearly 4000-fold, and is also under the control of CIA5 (Miura et al., 2004; Wang et al., 2005; Yamano et al., 2008; Fang et al., 2012). Studies in which LCIA was introduced into *Xenopus* oocytes showed LCIA to function as a low-affinity bicarbonate and a high-affinity nitrite transporter (Mariscal et al., 2006). While the subcellular location of LCIA has not yet been characterized, sequence analysis shows a putative chloroplast transit peptide. When NH_4^+ was the sole nitrogen source, LCIA showed a low-CO_2 induction response

(Mukherjee, 2013). Interestingly, when NO_3^- was the sole nitrogen source, LCIA maintained high expression in the presence of both high- and low-CO_2. This suggests the possibility that, in an environment where NO_3^- is the predominate nitrogen source, LCIA is required for NO_2^- transport, irrespective of CO_2 levels, whereas when NH_4^+ is present, LCIA can respond to CO_2 levels.

Lastly, YCF10, a plastid-encoded gene related to CemA of plant plastids, also has been proposed to function as a transporter of bicarbonate across the chloroplast inner membrane (Rolland et al., 1997). However, unlike CCP1, CCP2, and LCIA, this gene is not regulated by CIA5.

C_i Uptake across the Chloroplast Thylakoid Membrane

Interestingly, no transporters have been identified that may aid in C_i transport across the thylakoid membrane. However, a family of proteins (LCIB, LCIC, LCID, & LCIE) is believed to potentially facilitate C_i uptake and/or retention.

Spalding, Spreitzer, and Ogren (1983b) characterized a high CO_2-requiring mutant, *pmp-1* that was later shown to lack the LCIB protein and failed to accumulate C_i in low-CO_2. This mutant was believed to be deficient in C_i uptake. Interestingly, however, is the lack of transmembrane domains in this protein. It is believed that LCIB functions in complex with LCIC to retain CO_2 in the chloroplast or assist in C_i transport within the chloroplast (Yamano and Fukuzawa, 2009; Yamano et al., 2010). LCIB also functions epistatically downstream of CAH3, a carbonic anhydrase localized in the

thylakoid lumen, adding further mystery as to its precise function (Duanmu et al., 2009b).

Pyrenoid

The pyrenoid structure in Chlamydomonas functions as the destination for CO_2 and Rubisco, creating a microcompartment with elevated levels of CO_2 to enhance the carboxylation reaction. Rubisco is localized to the pyrenoid structure in many, if not most, eukaryotic algae, including Chlamydomonas (Rawat et al., 1996; Morita et al., 1997; Borkhsenious et al., 1998). This localization is initiated by low-CO_2, as Rubisco is evenly distributed across the chloroplast stroma in carbon-replete conditions. In addition, a starch sheath forms around the pyrenoid in low-CO_2 conditions. It is believed that Rubisco functions as part of the structure of pyrenoid, as mutants that lack Rubisco fail to form a pyrenoid structure (Rawat et al., 1996). Studies utilizing chimeric Rubisco enzymes (small subunits were taken from various plants and combined with the large subunit from Chlamydomonas) not only showed defects in the function of the CCM, but also the inability to form a pyrenoid (Genkov et al., 2010).

Because the pyrenoid is the primary site of carbon fixation, it is believed that one of its prime functions is the maintenance of artificially high levels of CO_2. The *cia6* mutant, lacking a putative methyltransferase, lacked an organized pyrenoid (Ma et al., 2011). This mutant, despite normal levels of Rubisco, showed limited growth in low-CO_2

environments - again highlighting the critical nature of this structure. The LCIB/LCIC complex has been found to surround the pyrenoid under low CO_2 conditions (Yamano et al., 2010). It is possible that this complex works in conjunction with CAH6 (a stromal-associated carbonic anhydrase) to act in limiting CO_2 leakage from the pyrenoid. The pyrenoid and its surrounding proteins may act as a physical barrier, an enzymatic barrier, or both, in limiting loss of CO_2.

Carbonic Anhydrases

To facilitate the accumulation and containment of C_i within the cell, Chlamydomonas utilizes a variety of carbonic anhydrases (CAs) and pH gradients. Robust analysis of the Chlamydomonas genome has uncovered 12 known CAs, including all major CA categories (α, β, and γ) (Moroney et al., 2011). Chlamydomonas possesses three α-CAs (CAH1, CAH2, CAH3), six β-CAs (CAH4, CAH5, CAH6, CAH7, CAH8, CAH9), and 3 γ-CAs (CAG1, CAG2, CAG3). CAs can function reversibly, catalyzing both the hydration of CO_2 and dehydration of HCO_3^-. When paired with the various pH conditions through the cell, these CAs function to create gradients to favor the retention of CO_2 and HCO_3^- molecules.

Periplasmic CAs

One of the first genes to be highly upregulated in response to a shift of cells to a low-CO_2 environment is *CAH1*, a periplasmic α-CA(Spencer et al., 1983; Coleman and Grossman, 1984; Moroney et al., 1985; Fukuzawa et al., 1990a). CAH1 was detected in the cell culture media, suggesting periplasmic localization (Kimpel et al., 1983). Later, immunoblots of cell wall fractions and immunogold localization using whole cells confirmed its periplasmic location in Chlamydomonas (Ishida et al., 1993; Moroney and Ynalvez, 2007). The use of cell membrane-impermeable CA inhibitors drastically reduced C_i affinity, suggesting a function of CAH1 in facilitating C_i uptake in cells (Berry et al., 1976; Badger et al., 1980; Moroney et al., 1985; Tsuzuki and Miyachi, 1989). Interestingly, however, was the finding that the *cah1* mutant, which lacks a functional *CAH1* gene, failed to show significant differences in C_i affinity, suggesting that a second periplasmic CA may play a role in the CCM (Van and Spalding, 1999). The CAH2 protein (an α-CA) is also periplasmically localized, and shares a high level of sequence identity with CAH1 (Rawat and Moroney, 1991). Unlike CAH1, however, CAH2 does not show an up-regulation in response to low-CO_2 levels, and is not under the control of CIA5 (Rawat and Moroney, 1991; Fang et al., 2012).

Stromal CAs

Currently, one CA, CAH6, is believed to be localized to the chloroplast stroma (Mitra et al., 2004). This CA is believed to function largely in the hydration of CO_2 entering the stroma from either the cytosol or the pyrenoid (Mitra et al., 2004; Moroney

et al., 2011). Because the pyrenoid is the exclusive location of Rubisco in the cell, it is possible that CAH6 serves largely to convert CO_2 that has leaked from the pyrenoid into HCO_3^-, which can be transported back into the pyrenoid. The alkaline environment of the stroma would favor this hydration reaction. This theory, however, is contingent on the exact location of CAH6 within the stroma. If CAH6 is localized in the stroma, but outside of the pyrenoid, then it is likely a CO_2 scavenger. If, however, CAH6 was present within the pyrenoid matrix, it might contribute to the dehydration of HCO_3^- ions, providing more CO_2 to Rubisco (Mukherjee, 2013). CAH6 shows only mild up-regulation in response to low-CO_2 levels, but appears to be regulated by CIA5 (Mitra et al., 2004; Ynalvez et al., 2008; Fang et al., 2012).

Thylakoid CAs

The final α-CA in Chlamydomonas to be discussed, CAH3, is associated with the thylakoid lumen (Karlsson et al., 1995; Karlsson et al., 1998; Moroney and Ynalvez, 2007). Unlike CAH1 and CAH2, CAH3 appears to be the sole thylakoid lumen-associated CA. This is supported by studies of the deletion mutant, *cah3*. Mutant strains of Chlamydomonas lacking the *CAH3* gene barely survived under low CO_2 conditions and, although C_i was over accumulated in the thylakoid lumen, insufficient levels of CO_2 were available for Rubisco (Spalding et al., 1983a; Moroney et al., 1986; Funke et al., 1997; Karlsson et al., 1998; Hanson et al., 2003). It is believed, then, that CAH3 is involved in the conversion of HCO_3^- to CO_2 within the lumen - a process favored by the acidic

environment of the lumen (Pronina and Semenenko, 1992; Karlsson et al., 1998; Moroney and Somanchi, 1999; Hanson et al., 2003). In the absence of CAH3, HCO_3^- ions entering the lumen would largely remain hydrated, limiting the amount of CO_2 available for Rubisco in the pyrenoid (Spalding et al., 1983a; Moroney et al., 1986). Unlike many CCM components, CAH3 shows only a mild up-regulation in response to low-CO_2 environments (Karlsson et al., 1998; Im et al., 2003; Wang and Spalding, 2006; Ynalvez et al., 2008; Yamano and Fukuzawa, 2009; Fang et al., 2012). Despite the lack of induction by low-CO_2, *CAH3* has been shown to be under control of the transcriptional regulator CIA5, as its expression is absent in *cia5* cells (Moroney et al., 1989; Marek and Spalding, 1991; Ynalvez et al., 2008).

Mitochondrial CAs

While the chloroplast appears to be paramount to the CCM, it should be noted that the mitochondria are not without a role in this mechanism. Two mitochondria-localized CAs, CAH4 & CAH5, were isolated from Chlamydomonas (Villand et al., 1997). Like CAH1 and CAH2, these proteins share a high level of sequence identity, and are also believed to be the result of a recent duplication event within the genome. The genes encoding these proteins are responsive to low-CO_2, as well as being under the control of CIA5 (Fujiwara et al., 1990; Ynalvez et al., 2008; Fang et al., 2012). The roles of CAH4 and CAH5 may be two-fold in Chlamydomonas. When subjected to low-CO_2 environments, mitochondria are found to migrate around the periphery of the chloroplast, and CAH4

and CAH5 may function in the capture of CO_2 escaping the chloroplast (rapidly converting these molecules to HCO_3^-, which can be better-retained by the cell) (Giordano et al., 2003; Moroney et al., 2011). Additionally, CAH4 and CAH5 may function in the rehydration of CO_2 produced by photorespiration, namely, the glycine decarboxylase complex (Giordano et al., 2003).

Additional CAs

The remaining three β-CAs in Chlamydomonas, CAH7, CAH8, and CAH9, are still being evaluated in regard to their potential role in the CCM (Moroney et al., 2011). CAH7 and CAH8 both possess C-terminal sequences with predicted hydrophobic regions, suggesting that these two proteins are located in the membrane (Ynalvez et al., 2008). Immunogold labeling of CAH8 further supports this notion, by showing CAH8 to be associated with the plasma membrane. However, the location of the CAH8's active site (periplasmic or cytosolic) is unclear. Periplasmic location of CAH8's active site would indicate a role in $CO_2 \rightarrow HCO_3^-$ conversion for cellular assimilation, similar to CAH1 and CAH2, while a cytosolic location would suggest a role in trapping CO_2 (in the form of bicarbonate) within the cell, such as CAH6. Unlike CAH7 & CAH8, CAH9 lacks any such hydrophobic extension, lending to the possibility that it is located in the cytosol (Moroney et al., 2011). All three of these β-CAs show little change in gene expression in low CO_2 (Ynalvez et al., 2008; Moroney et al., 2011; Fang et al., 2012). This lack of

induction may indicate the secondary role of these CAs in the CCM, or that they are constitutively present in the cell in sufficient amounts.

The third class of CAs present in Chlamydomonas, the γ-CAs, is associated with mitochondria (Sunderhaus et al., 2006; Klodmann et al., 2010). Like the γ-CAs of higher plants, these proteins are believed to be associated with Complex I of the Electron Transport Chain. Proteomic studies of mitochondria in Chlamydomonas lend evidence to support this position (Cardol et al., 2004; Cardol et al., 2005). While the importance of these CAs in mitochondrial function is clear, it remains to be seen if they play a significant role in the CCM.

Regulatory Responses in Chlamydomonas associated with the CCM

Detection of Carbon Levels

While the exact mechanism(s) has not yet been identified, Chlamydomonas cells are able to rapidly detect changes in external CO_2 levels. This detection does not appear to be influenced by HCO_3^- ion levels or the pH, suggesting some form of detection located within the plasma membrane (Matsuda and Colman, 1995; Bozzo and Colman, 2000; Giordano et al., 2005). Because of its nuclear location, it is unlikely that CIA5 is a direct sensor of CO_2 levels. Most likely, CIA5 response to changes in external CO_2 levels requires transduction of a signal from outside or within the cell membrane (Kohinata et

al., 2008). In addition to CO_2 levels, light is another factor implicated in CCM responses. Some genes, including CAH1, appear to be induced in the dark, while other CIA5-controlled genes, such as CAH4/CAH5, require light for induction (Villarejo et al., 1997). Considering that, by driving photosynthesis, light would cause depletion of internal carbon levels. In this regard, rescue (or lack thereof) of internal carbon levels could potentially provide a useful reflection of environmental CO_2 conditions. Subsequently, a reduction in C_i could affect the redox potential of the photosynthetic electron transport system, which could also serve as a signal to the cell (Spalding and Ogren, 1982; Spencer et al., 1983; Fukuzawa et al., 1990b; Villarejo et al., 1997). Alternatively, depletion of internal C_i would cause an increase in oxygenation reactions by Rubisco, and an accumulation of photorespiratory metabolites could function as signals to the cell. Lending support to this later hypothesis is evidence that several CO_2-responsive genes, including CAH1, CAH4, & CAH5, were not induced by low CO_2 when the glycolate pathway was inhibited (Spalding and Ogren, 1982; Ramazanov and Cardenas, 1992; Villarejo et al., 1997). Regardless of such speculation, no firm evidence exists in regard to the nature of the signal(s) used by Chlamydomonas to control the activity of the CCM in response to changes in external CO_2 concentrations.

Induction of the CCM

Study of carbon assimilation in Chlamydomonas showed significant differences in the alga's affinity for C_i in relation to environmental conditions; specifically,

Chlamydomonas showed higher affinity for C_i in carbon-depleted environments (≤390ppm CO_2) than in carbon-replete environments (1-5% CO_2) (Badger and Price, 1994). In addition, structural changes, including formation and localization of Rubisco to the pyrenoid, and migration of mitochondria to the membrane periphery, coincide with these low-CO_2 environments (Kuchitsu et al., 1988, 1991; Ramazanov et al., 1994; Geraghty and Spalding, 1996; Thyssen et al., 2001). Changes in gene expression also coincide with carbon-depleted environments. By the early 1990s, gene expression studies were being conducted to identify genes that showed a low-CO_2 response. Differential display (Im and Grossman, 2002) and EST-based microarray assays (Im et al., 2003; Miura et al., 2004; Wang et al., 2005; Yamano et al., 2008) were among some of the earlier works dedicated to identifying CO_2-responvie genes. These efforts have been expanded significantly by recent RNA-sequencing studies (Fang et al., 2012), including a chapter in this thesis.

As noted above, the CCM-deficient Chlamydomonas mutant, *cia5* (C_i acquisition mutant #5), isolated in 1989, has proven the most powerful tool yet available for studies of the regulation of the CCM. This mutant has led to insights into the function and control of the CCM that, otherwise, would not have been possible. The cia5 mutant, generated by UV irradiation, lacked a functional CCM, and was unable to survive in low-CO_2 environments (Moroney et al., 1989; Marek and Spalding, 1991). Independent efforts to complement this mutant led to isolation and characterization of the

CIA5/CCM1 gene (Fukuzawa et al., 2001; Xiang et al., 2001). This gene, containing two N-terminal zinc finger motifs, appeared to be a characteristic transcriptional activator. To date, researchers have been unable to show any ability of CIA5 to directly bind DNA (Kohinata et al., 2008). The regulatory role of this protein, however, is without question, as several CAs (CAH1, CAH3, CAH4, CAH5), transporters (HLA3, LCI1, CCP1, CCP2, LCIA), and additional CCM components fail to be induced in low CO_2 without the presence of CIA5. Interestingly, however, levels of CIA5 itself do not change in response to low-CO_2, suggesting that it serves sentinel in anticipation of the need for rapid changes in gene expression patterns when external CO_2 levels decrease or increase (Xiang et al., 2001; Moroney and Ynalvez, 2007; Kohinata et al., 2008; Wang et al., 2011). Another gene, *LCR1*, was found to act downstream of *CIA5* to control transcription of a subset of CO_2-responsive genes (Yoshioka et al., 2004). Mutant cells lacking a functional LCR1 failed to induce a subset of the CIA5-controlled low-CO_2-inducible genes. This suggests the possibility that CIA5 functions in activating multiple proteins that each regulate different subsets of CO_2-responsive genes.

References

Badger, M.R., and Price, G.D. (1994). The Role of Carbonic-Anhydrase in Photosynthesis. Annual Review of Plant Physiology and Plant Molecular Biology **45,** 369-392.

Badger, M.R., Kaplan, A., and Berry, J.A. (1980). Internal Inorganic Carbon Pool of Chlamydomonas reinhardtii: EVIDENCE FOR A CARBON DIOXIDE-CONCENTRATING MECHANISM. Plant Physiol **66,** 407-413.

Berry, J., Boynton, J., Kaplan, A., and Badger, M. (1976). Growth and photosynthesis of Chlamydomonas reinhardtii as a function of CO2 concentration. Carnegie Institute of Washington Yearbook **75,** 423-432.

Borkhsenious, O.N., Mason, C.B., and Moroney, J.V. (1998). The intracellular localization of ribulose-1,5-bisphosphate carboxylase/oxygenase in Chlamydomonas reinhardtii. Plant Physiol **116,** 1585-1591.

Bozzo, G.G., and Colman, B. (2000). The induction of inorganic carbon transport and external carbonic anhydrase in Chlamydomonas reinhardtii is regulated by external CO2 concentration. Plant Cell and Environment **23,** 1137-1144.

Burow, M.D., Chen, Z.Y., Mouton, T.M., and Moroney, J.V. (1996). Isolation of cDNA clones of genes induced upon transfer of Chlamydomonas reinhardtii cells to low CO2. Plant Molecular Biology **31,** 443-448.

Calamita, G., Kempf, B., Rudd, K.E., Bonhivers, M., Kneip, S., Bishai, W.R., . . . Agre, P. (1997). The aquaporin-Z water channel gene of Escherichia coli: structure, organization and phylogeny. Biol Cell **89,** 321-329.

Cardol, P., Vanrobaeys, F., Devreese, B., Van Beeumen, J., Matagne, R.F., and Remacle, C. (2004). Higher plant-like subunit composition of mitochondrial complex I from Chlamydomonas reinhardtii: 31 conserved components among eukaryotes. Biochimica Et Biophysica Acta-Bioenergetics **1658,** 212-224.

Cardol, P., Gonzalez-Halphen, D., Reyes-Prieto, A., Baurain, D., Matagne, R.F., and Remacle, C. (2005). The mitochondrial oxidative phosphorylation proteome of Chlamydomonas reinhardtii deduced from the genome sequencing project. Plant Physiol **137,** 447-459.

Chen, Z.Y., Lavigne, L.L., Mason, C.B., and Moroney, J.V. (1997). Cloning and overexpression of two cDNAs encoding the low-CO2-inducible chloroplast envelope protein LIP-36 from Chlamydomonas reinhardtii. Plant Physiol **114,** 265-273.

Coleman, J.R., and Grossman, A.R. (1984). Biosynthesis of Carbonic-Anhydrase in Chlamydomonas-Reinhardtii during Adaptation to Low Co2. Proceedings of the National Academy of Sciences of the United States of America-Biological Sciences **81,** 6049-6053.

Duanmu, D., Miller, A.R., Horken, K.M., Weeks, D.P., and Spalding, M.H. (2009a). Knockdown of limiting-CO2-induced gene HLA3 decreases HCO3- transport and

photosynthetic Ci affinity in Chlamydomonas reinhardtii. Proceedings of the National Academy of Sciences of the United States of America **106,** 5990-5995.

Duanmu, D.Q., Wang, Y.J., and Spalding, M.H. (2009b). Thylakoid Lumen Carbonic Anhydrase (CAH3) Mutation Suppresses Air-Dier Phenotype of LCIB Mutant in Chlamydomonas reinhardtii. Plant Physiol **149,** 929-937.

Fang, W., Si, Y.Q., Douglass, S., Casero, D., Merchant, S.S., Pellegrini, M., . . . Spalding, M.H. (2012). Transcriptome-Wide Changes in Chlamydomonas reinhardtii Gene Expression Regulated by Carbon Dioxide and the CO2-Concentrating Mechanism Regulator CIA5/CCM1. Plant Cell **24,** 1876-1893.

Fujiwara, S., Fukuzawa, H., Tachiki, A., and Miyachi, S. (1990). Structure and differential expression of two genes encoding carbonic anhydrase in Chlamydomonas reinhardtii. Proc Natl Acad Sci U S A **87,** 9779-9783.

Fukuzawa, H., Fujiwara, S., Tachiki, A., and Miyachi, S. (1990a). Nucleotide sequences of two genes CAH1 and CAH2 which encode carbonic anhydrase polypeptides in Chlamydomonas reinhardtii. Nucleic Acids Res **18,** 6441-6442.

Fukuzawa, H., Fujiwara, S., Yamamoto, Y., Dionisio-Sese, M.L., and Miyachi, S. (1990b). cDNA cloning, sequence, and expression of carbonic anhydrase in Chlamydomonas reinhardtii: regulation by environmental CO2 concentration. Proc Natl Acad Sci U S A **87,** 4383-4387.

Fukuzawa, H., Miura, K., Ishizaki, K., Kucho, K.I., Saito, T., Kohinata, T., and Ohyama, K. (2001). Ccm1, a regulatory gene controlling the induction of a carbon-concentrating mechanism in Chlamydomonas reinhardtii by sensing CO2 availability. Proc Natl Acad Sci U S A **98,** 5347-5352.

Funke, R.P., Kovar, J.L., and Weeks, D.P. (1997). Intracellular carbonic anhydrase is essential to photosynthesis in Chlamydomonas reinhardtii at atmospheric levels of CO2. Demonstration via genomic complementation of the high-CO2-requiring mutant ca-1. Plant Physiol **114,** 237-244.

Genkov, T., Meyer, M., Griffiths, H., and Spreitzer, R.J. (2010). Functional hybrid rubisco enzymes with plant small subunits and algal large subunits: engineered rbcS cDNA for expression in chlamydomonas. J Biol Chem **285,** 19833-19841.

Geraghty, A.M., and Spalding, M.H. (1996). Molecular and Structural Changes in Chlamydomonas under Limiting CO2 (A Possible Mitochondrial Role in Adaptation). Plant Physiol **111,** 1339-1347.

Giordano, M., Beardall, J., and Raven, J.A. (2005). CO2 concentrating mechanisms in algae: mechanisms, environmental modulation, and evolution. Annu Rev Plant Biol **56,** 99-131.

Giordano, M., Norici, A., Forssen, M., Eriksson, M., and Raven, J.A. (2003). An anaplerotic role for mitochondrial carbonic anhydrase in Chlamydomonas reinhardtii. Plant Physiol **132,** 2126-2134.

Hanson, D.T., Franklin, L.A., Samuelsson, G., and Badger, M.R. (2003). The Chlamydomonas reinhardtii cia3 mutant lacking a thylakoid lumen-localized

carbonic anhydrase is limited by CO2 supply to rubisco and not photosystem II function in vivo. Plant Physiol **132,** 2267-2275.

Im, C.S., and Grossman, A.R. (2002). Identification and regulation of high light-induced genes in Chlamydomonas reinhardtii. Plant J **30,** 301-313.

Im, C.S., Zhang, Z.D., Shrager, J., Chang, C.W., and Grossman, A.R. (2003). Analysis of light and CO2 regulation in Chlamydomonas reinhardtii using genome-wide approaches. Photosynthesis Research **75,** 111-125.

Ishida, S., Muto, S., and Miyachi, S. (1993). Structural-Analysis of Periplasmic Carbonic Anhydrase-1 of Chlamydomonas-Reinhardtii. European Journal of Biochemistry **214,** 9-16.

Karlsson, J., Hiltonen, T., Husic, H.D., Ramazanov, Z., and Samuelsson, G. (1995). Intracellular Carbonic-Anhydrase of Chlamydomonas-Reinhardtii. Plant Physiol **109,** 533-539.

Karlsson, J., Clarke, A.K., Chen, Z.Y., Hugghins, S.Y., Park, Y.I., Husic, H.D., . . . Samuelsson, G. (1998). A novel alpha-type carbonic anhydrase associated with the thylakoid membrane in Chlamydomonas reinhardtii is required for growth at ambient CO2. Embo Journal **17,** 1208-1216.

Kimpel, D.L., Togasaki, R.K., and Miyachi, S. (1983). Carbonic-Anhydrase in Chlamydomonas-Reinhardtii .1. Localization. Plant and Cell Physiology **24,** 255-259.

Klodmann, J., Sunderhaus, S., Nimtz, M., Jansch, L., and Braun, H.P. (2010). Internal Architecture of Mitochondrial Complex I from Arabidopsis thaliana. Plant Cell **22,** 797-810.

Kohinata, T., Nishino, H., and Fukuzawa, H. (2008). Significance of zinc in a regulatory protein, CCM1, which regulates the carbon-concentrating mechanism in Chlamydomonas reinhardtii. Plant and Cell Physiology **49,** 273-283.

Kuchitsu, K., Tsuzuki, M., and Miyachi, S. (1988). Changes of Starch Localization within the Chloroplast Induced by Changes in Co2 Concentration during Growth of Chlamydomonas-Reinhardtii - Independent Regulation of Pyrenoid Starch and Stroma Starch. Plant and Cell Physiology **29,** 1269-1278.

Kuchitsu, K., Tsuzuki, M., and Miyachi, S. (1991). Polypeptide Composition and Enzyme-Activities of the Pyrenoid and Its Regulation by Co2 Concentration in Unicellular Green-Algae. Canadian Journal of Botany-Revue Canadienne De Botanique **69,** 1062-1069.

Kustu, S., and Inwood, W. (2006). Biological gas channels for NH3 andCO(2): evidence that Rh (rhesus) proteins areCO(2) channels. Transfusion Clinique Et Biologique **13,** 103-110.

Ma, Y.B., Pollock, S.V., Xiao, Y., Cunnusamy, K., and Moroney, J.V. (2011). Identification of a Novel Gene, CIA6, Required for Normal Pyrenoid Formation in Chlamydomonas reinhardtii. Plant Physiol **156,** 884-896.

Marek, L.F., and Spalding, M.H. (1991). Changes in Photorespiratory Enzyme-Activity in Response to Limiting Co2 in Chlamydomonas-Reinhardtii. Plant Physiol **97,** 420-425.

Mariscal, V., Moulin, P., Orsel, M., Miller, A.J., Fernandez, E., and Galvan, A. (2006). Differential regulation of the Chlamydomonas Nar1 gene family by carbon and nitrogen. Protist **157,** 421-433.

Matsuda, Y., and Colman, B. (1995). Induction of Co2 and Bicarbonate Transport in the Green-Alga Chlorella-Ellipsoidea .2. Evidence for Induction in Response to External Co2 Concentration. Plant Physiol **108,** 253-260.

Mitra, M., Lato, S.M., Ynalvez, R.A., Xiao, Y., and Moroney, J.V. (2004). Identification of a new chloroplast carbonic anhydrase in Chlamydomonas reinhardtii. Plant Physiol **135,** 173-182.

Miura, K., Yamano, T., Yoshioka, S., Kohinata, T., Inoue, Y., Taniguchi, F., . . . Fukuzawa, H. (2004). Expression profiling-based identification of CO2-responsive genes regulated by CCM1 controlling a carbon-concentrating mechanism in Chlamydomonas reinhardtii. Plant Physiol **135,** 1595-1607.

Morita, E., Kuroiwa, H., Kuroiwa, T., and Nozaki, H. (1997). High localization of ribulose-1,5-bisphosphate carboxylase/oxygenase in the pyrenoids of Chlamydomonas reinhardtii (Chlorophyta), as revealed by cryofixation and immunogold electron microscopy. Journal of Phycology **33,** 68-72.

Moroney, J.V., and Somanchi, A. (1999). How do algae concentrate CO2 to increase the efficiency of photosynthetic carbon fixation? Plant Physiol **119,** 9-16.

Moroney, J.V., and Ynalvez, R.A. (2007). Proposed carbon dioxide concentrating mechanism in Chlamydomonas reinhardtii. Eukaryotic Cell **6,** 1251-1259.

Moroney, J.V., Husic, H.D., and Tolbert, N.E. (1985). Effect of Carbonic-Anhydrase Inhibitors on Inorganic Carbon Accumulation by Chlamydomonas-Reinhardtii. Plant Physiol **79,** 177-183.

Moroney, J.V., Tolbert, N.E., and Sears, B.B. (1986). Complementation Analysis of the Inorganic Carbon Concentrating Mechanism of Chlamydomonas-Reinhardtii. Molecular & General Genetics **204,** 199-203.

Moroney, J.V., Husic, H.D., Tolbert, N.E., Kitayama, M., Manuel, L.J., and Togasaki, R.K. (1989). Isolation and Characterization of a Mutant of Chlamydomonas-Reinhardtii Deficient in the Co2 Concentrating Mechanism. Plant Physiol **89,** 897-903.

Moroney, J.V., Ma, Y.B., Frey, W.D., Fusilier, K.A., Pham, T.T., Simms, T.A., . . . Mukherjee, B. (2011). The carbonic anhydrase isoforms of Chlamydomonas reinhardtii: intracellular location, expression, and physiological roles. Photosynthesis Research **109,** 133-149.

Mukherjee, B. (2013). Investigations of the role of putative inorganic carbon transporters in the carbon dioxide concentrating mechanisms of Chlamydomonas reinhardtii. In Department of Biological Sciences (Louisiana

State University Electronic Thesis & Dissertation Collection: Louisiana State University).

Ohnishi, N., Mukherjee, B., Tsujikawa, T., Yanase, M., Nakano, H., Moroney, J.V., and Fukuzawa, H. (2010). Expression of a Low CO2-Inducible Protein, LCI1, Increases Inorganic Carbon Uptake in the Green Alga Chlamydomonas reinhardtii. Plant Cell **22,** 3105-3117.

Pollock, S.V., Prout, D.L., Godfrey, A.C., Lemaire, S.D., and Moroney, J.V. (2004). The Chlamydomonas reinhardtii proteins Ccp1 and Ccp2 are required for long-term growth, but are not necessary for efficient photosynthesis, in a low-CO2 environment. Plant Molecular Biology **56,** 125-132.

Prins, H.B.A., and Elzenga, J.T.M. (1989). Bicarbonate Utilization - Function and Mechanism. Aquatic Botany **34,** 59-83.

Pronina, N.A., and Semenenko, V.E. (1992). Carbonic-Anhydrase Activity and Fatty-Acid Composition of Photosystem Deficient and High Co2 Required Mutants of Chlamydomonas-Reinhardtii. Photosynthesis Research **34,** 201-201.

Ramazanov, Z., and Cardenas, J. (1992). Involvement of Photorespiration and Glycolate Pathway in Carbonic-Anhydrase Induction and Inorganic Carbon Concentration in Chlamydomonas-Reinhardtii. Physiologia Plantarum **84,** 502-508.

Ramazanov, Z., Rawat, M., Henk, M.C., Mason, C.B., Matthews, S.W., and Moroney, J.V. (1994). The Induction of the Co2-Concentrating Mechanism Is Correlated with the Formation of the Starch Sheath around the Pyrenoid of Chlamydomonas-Reinhardtii. Planta **195,** 210-216.

Rawat, M., and Moroney, J.V. (1991). Partial Characterization of a New Isoenzyme of Carbonic-Anhydrase Isolated from Chlamydomonas-Reinhardtii. Journal of Biological Chemistry **266,** 9719-9723.

Rawat, M., Henk, M.C., Lavigne, L.L., and Moroney, J.V. (1996). Chlamydomonas reinhardtii mutants without ribulose-1,5-bisphosphate carboxylase-oxygenase lack a detectable pyrenoid. Planta **198,** 263-270.

Rolland, N., Dorne, A.J., Amoroso, G., Sultemeyer, D.F., Joyard, J., and Rochaix, J.D. (1997). Disruption of the plastid ycf10 open reading frame affects uptake of inorganic carbon in the chloroplast of Chlamydomonas. Embo Journal **16,** 6713-6726.

Soupene, E., Inwood, W., and Kustu, S. (2004). Lack of the Rhesus protein Rh1 impairs growth of the green alga Chlamydomonas reinhardtii at high CO2. Proceedings of the National Academy of Sciences of the United States of America **101,** 7787-7792.

Spalding, M.H. (2008). Microalgal carbon-dioxide-concentrating mechanisms: Chlamydomonas inorganic carbon transporters. Journal of Experimental Botany **59,** 1463-1473.

Spalding, M.H., and Ogren, W.L. (1982). Photosynthesis Is Required for Induction of the Co2-Concentrating System in Chlamydomonas-Reinhardii. Febs Letters **145,** 41-44.

Spalding, M.H., Spreitzer, R.J., and Ogren, W.L. (1983a). Reduced Inorganic Carbon Transport in a CO(2)-Requiring Mutant of Chlamydomonas reinhardii. Plant Physiol **73,** 273-276.

Spalding, M.H., Spreitzer, R.J., and Ogren, W.L. (1983b). Carbonic Anhydrase-Deficient Mutant of Chlamydomonas-Reinhardii Requires Elevated Carbon-Dioxide Concentration for Photoautotrophic Growth. Plant Physiol **73,** 268-272.

Spencer, K.G., Kimpel, D.L., Fisher, M.L., Togasaki, R.K., and Miyachi, S. (1983). Carbonic-Anhydrase in Chlamydomonas-Reinhardtii .2. Requirements for Carbonic-Anhydrase Induction. Plant and Cell Physiology **24,** 301-304.

Sunderhaus, S., Dudkina, N.V., Jansch, L., Klodmann, J., Heinemeyer, J., Perales, M., . . . Braun, H.P. (2006). Carbonic anhydrase subunits form a matrix-exposed domain attached to the membrane arm of mitochondrial complex I in plants. Journal of Biological Chemistry **281,** 6482-6488.

Thyssen, C., Schlichting, R., and Giersch, C. (2001). The CO2-concentrating mechanism in the physiological context: lowering the CO2 supply diminishes culture growth and economises starch utilisation in Chlamydomonas reinhardtii. Planta **213,** 629-639.

Tsuzuki, M., and Miyachi, S. (1989). The Function of Carbonic-Anhydrase in Aquatic Photosynthesis. Aquatic Botany **34,** 85-104.

Van, K., and Spalding, M.H. (1999). Periplasmic carbonic anhydrase structural gene (Cah1) mutant in Chlamydomonas reinhardtii. Plant Physiol **120,** 757-764.

Villand, P., Eriksson, M., and Samuelsson, G. (1997). Carbon dioxide and light regulation of promoters controlling the expression of mitochondrial carbonic anhydrase in Chlamydomonas reinhardtii. Biochemical Journal **327,** 51-57.

Villarejo, A., Martinez, F., and Ramazanov, Z. (1997). Effect of aminooxyacetate, an inhibitor blocking the glycolate pathway, on the induction of a CO2-concentrating mechanism and low-CO2-inducible polypeptides in Chlamydomonas reinhardtii (Chlorophyta). European Journal of Phycology **32,** 141-145.

Wang, Y., Sun, Z.H., Horken, K.M., Im, C.S., Xiang, Y.B., Grossman, A.R., and Weeks, D.P. (2005). Analyses of CIA5, the master regulator of the carbon-concentrating mechanism in Chlamydomonas reinhardtii, and its control of gene expression. Canadian Journal of Botany-Revue Canadienne De Botanique **83,** 765-779.

Wang, Y.J., and Spalding, M.H. (2006). An inorganic carbon transport system responsible for acclimation specific to air levels Of CO2 in Chiamlydomonas reinhardtii. Proceedings of the National Academy of Sciences of the United States of America **103,** 10110-10115.

Wang, Y.J., Duanmu, D.Q., and Spalding, M.H. (2011). Carbon dioxide concentrating mechanism in Chlamydomonas reinhardtii: inorganic carbon transport and CO2 recapture. Photosynthesis Research **109,** 115-122.

Xiang, Y.B., Zhang, J., and Weeks, D.P. (2001). The Cia5 gene controls formation of the carbon concentrating mechanism in Chlamydomonas reinhardtii. Proceedings of the National Academy of Sciences of the United States of America **98,** 5341-5346.

Yamano, T., and Fukuzawa, H. (2009). Carbon-concentrating mechanism in a green alga, Chlamydomonas reinhardtii, revealed by transcriptome analyses. Journal of Basic Microbiology **49,** 42-51.

Yamano, T., Miura, K., and Fukuzawa, H. (2008). Expression analysis of genes associated with the induction of the carbon-concentrating mechanism in Chlamydomonas reinhardtii. Plant Physiol **147,** 340-354.

Yamano, T., Tsujikawa, T., Hatano, K., Ozawa, S., Takahashi, Y., and Fukuzawa, H. (2010). Light and Low-CO2-Dependent LCIB-LCIC Complex Localization in the Chloroplast Supports the Carbon-Concentrating Mechanism in Chlamydomonas reinhardtii. Plant and Cell Physiology **51,** 1453-1468.

Ynalvez, R.A., Xiao, Y., Ward, A.S., Cunnusamy, K., and Moroney, J.V. (2008). Identification and characterization of two closely related beta-carbonic anhydrases from Chlamydomonas reinhardtii. Physiologia Plantarum **133,** 15-26.

Yoshioka, S., Taniguchi, F., Miura, K., Inoue, T., Yamano, T., and Fukuzawa, H. (2004). The novel Myb transcription factor LCR1 regulates the CO2-responsive gene Cah1, encoding a periplasmic carbonic anhydrase in Chlamydomonas reinhardtii. Plant Cell **16,** 1466-1477.

CHAPTER 1.2 - Herbicides

The use of herbicides in agriculture has been well established for several years. One of the first herbicides to be used in large-scale agricultural application was 2,4-D, which was developed during World War II. Since the introduction of 2,4-D, several herbicides have been developed, targeting various components of plant (and, occasionally, other organisms) metabolism. In this review, three classes of herbicides will be discussed; protoporphyrinogen oxidase-inhibiting compounds (PPO-inhibitors), including oxadiazon and oxyfluorfen, the phytoene desaturase-inhibitor norflurazon, and the EPSPS-inhibitor glyphosate.

Protoporphyrinogen Oxidase and PPO-Inhibitors

Compounds in Class 14 of the Weed Science Society of America herbicide grouping system are categorized as PPO-inhibitors. These compounds disrupt the function of protoporphyrinogen oxidase (Protox, PPO), an enzyme responsible for the conversion of protoporphyrinogen IX to protoporphyrin IX (Duke et al., 1991). Protoporphyrin IX is a substrate for both heme and chlorophyll biosynthesis. Inhibition of this enzyme leads to a buildup of the substrate, which subsequently leaks from the chloroplast to the cytosol and is oxidized by a non-specific plasma membrane-bound peroxidase (Ha et al., 2004). A result of Protox-inhibition, feedback signaling is altered and more carbon is directed towards the Protox pathway, thus exacerbating the buildup of protoporphyrin IX in the cytosol. Because protoporphyrin IX is not a natural

metabolite in the cytosol and because it possesses photosensitizing characteristics, exposure to light generates high levels of reactive oxygen species (Duke et al., 1991). These free radicals quickly cause peroxidation of unsaturated fatty acids in the cell membrane, as well as peroxidation of other cell constituents, such as lipids, proteins, and nucleic acids (Matringe et al., 1993; Dayan and Duke, 1997; Wakabayashi and Boger, 1999; Matsumoto, 2002; Ha et al., 2004). Because of the cellular damage occurring from peroxidation of components within the cell, plants treated with PPO-inhibitors exhibit bleaching of chloroplast pigments and subsequent inhibition of photosynthesis, formation of ethylene, ethane evolution due to membrane peroxidation, eventual tissue necrosis, and finally plant death (Wakabayashi and Boger, 1999; Matsumoto, 2002). Although PPO-inhibitors tend to be more effective in killing dicots than monocots, enzymatic studies have shown the enzyme target across these organisms to be equally sensitive to these herbicides (Li and Nicholl, 2005).

PPO-inhibitor herbicides have been in common use as both a pre-emergence herbicide (preventing germination of weed seeds) and post-emergence herbicide (eliminating growth of germinated weeds). The low dosages of these herbicides required for plant death and moderate cost make them suitable candidates for large-scale use. The research reported in this thesis focuses on two compounds within the class of PPO-inhibitors: oxadiazon (CAS: 19666-30-9) and oxyfluorfen (CAS: 42874-03-3). Oxadiazon, commonly sold under the trade name Ronstar (Bayer Environmental Science), is an oxadiazole herbicide first registered to be sold in the United States in

1978 (EPA, 2003). Oxyfluorfen, a diphenyl ether, was registered by the EPA shortly after (1979), and is marketed under the trade names Goal and Galigan (EPA, 2002). Both compounds were approved for agricultural and ornamental usage, although, since 1991, oxadiazon usage has been restricted to non-food applications (golf course fairways, roadsides, etc.). The stability of these compounds varies, with reports suggesting a relatively low half-life of oxadiazon of 1.5 days, while oxyfluorfen is reported to have a much longer half-life (291-603 days, depending on soil type) (EPA, 2002, 2003). These numbers are slightly disputed, as oxyfluorfen has been reported to be degraded quickly in light (Ha et al., 2004), and a field trial concluded that, after 60 days post-application, only 3% of oxyfluorfen persisted in the soil (Das et al., 2003). While some concerns exist about the risks of widespread use of herbicides, the low effective dosage of these herbicides (0.72 kg ai/ha oxyfluorfen, 0.75-1.25 kg ai/ha oxadiazon) (Qasem and Foy, 2006), broad herbicidal spectrum (Hao et al., 2011), limited persistence in soil (oxadiazon), and relative immobility (oxyfluorfen), make them attractive options for weed management.

The engineering of plants to be resistant to PPO-inhibitors has been explored extensively. Almost all photosynthetic eukaryotic organisms possess two forms of Protox; one form destined to the chloroplast (for chlorophyll biosynthesis) and the other to the mitochondria (for heme biosynthesis) (Watanabe et al., 2001). In contrast, non-photosynthetic organisms, such as mammals, tend to possess only a mitochondrial form of Protox (Morgan et al., 2004). While most forms of Protox are sensitive to PPO-

inhibitors, it has been found that simple overexpression of Protox can confer elevated levels of resistance to PPO-inhibitors in some plants. For example, in 2000 (Lermontova and Grimm), PPO-inhibitor-resistant lines of tobacco were generated by overexpression of the Arabidopsis plastidic Protox gene. This approach was also successful in generating PPO-inhibitor-resistant rice (Ha et al., 2004). Utilizing the *Myxococcus xanthus* Protox gene instead of that from Arabidopsis, researchers were also able to generate PPO-inhibitor-resistant rice (Jung et al., 2004). Characterization of a naturally PPO-inhibitor-resistant line of soybean showed that resistance was due to overexpression of the endogenous mitochondrial Protox (Warabi et al., 2001). Alteration of the amino acid composition of Protox can also generate PPO-inhibitor-resistant plants, but, in some cases, these mutations reduce enzymatic function (Li and Nicholl, 2005). No crop species naturally resistant to PPO-inhibitors or genetically engineered for PPO-inhibitor resistance are commercially available.

Analysis of the sensitivity of Chlamydomonas to PPO-inhibitors showed that it, too, was susceptible to the inhibitory effects of these compounds (Acifluorfen ED_{50} = 0.56 mg/L in Reboud (2002))(Oxadiazon ED_{90} = 6 µM, Oxyfluorfen ED_{90} = 0.3 µM in Kataoka et al. (1990)). In 1990, a mutant strain of Chlamydomonas was isolated and characterized that showed elevated levels of tolerance to PPO-inhibitor compounds. This strain, *rs-3*, showed a 15-fold increase in the effective dose $(ED)_{90}$ for oxadiazon (90 µM) and a 10.7-fold increase in the ED_{90} for oxyfluorfen (3.2 µM) (Kataoka et al., 1990). This mutant maintained normal levels of sensitivity for non-PPO-inhibitor herbicides,

indicating that the mutation was specific to Protox function. Later characterization of this mutant showed a one-bp alteration in the Protox gene, causing a valine to methionine change (Randolph-Anderson et al., 1998). Transformation of the cosmid containing this mutant gene conferred similar levels of resistance to WT cells as seen in the *rs-3* mutant, demonstrating the dominant nature of this mutation. Upon discovery of this gene, the Sumitomo Chemical Company (involved in the discovery and characterization of this mutant gene) filed a patent on the gene and its use. This resulted in the lack of availability of the gene for academic purposes. Moreover, no transgenic plants carrying a genetically engineered version of *rs-3* were developed for the marketplace. Continued interest in developing new selectable markers for genetic transformation of algal cells and a budding interest in finding compounds capable of controlling microbial contamination of commercial-scale algal ponds and raceways prompted our laboratory to reevaluate the *rs-3* gene and potentially engineer it for scientific and commercial purposes.

Glyphosate and EPSPS

Perhaps the most well-known herbicide in use today is glyphosate. As of 2011, the Class 9 herbicide glyphosate was one of the most commonly used herbicides in agriculture, forestry and horticulture (Grube et al., 2011). Glyphosate, or N(phosphonomethyl)glycine (CAS 1071-83-6) is a broad spectrum herbicide that was first introduced in 1974 (Monsanto, 2005). Since that time, it has grown in popularity

due to its broad range of effectiveness against weeds and low toxicity to non-target organisms (Williams et al., 2000). Glyphosate is sold under several trade names, though it is commonly referred to as Roundup. Glyphosate functions as a potent inhibitor of the enzyme 5-enolpyruvylshikimate-3-phosphate synthase (EPSPS), an enzyme of the shikimate pathway involved in the biosynthesis of aromatic amino acids (Vencill, 2002). Glyphosate binds EPSPS in the active site location normally reserved for the substrate phosphoenolpyruvate (PEP) (Boocock and Coggins, 1983; Schonbrunn et al., 2001; Eschenburg et al., 2003; Park et al., 2004; Funke et al., 2006). Glyphosate is considered slightly- to non-toxic to most animals, as only plants and some microorganisms possess EPSPS (Steinrucken and Amrhein, 1980). Inhibition of EPSPS by glyphosate causes a reduction in aromatic amino acids, coupled with deregulation of the shikimate pathway (Cerdeira and Duke, 2010). This deregulation causes a large-scale misallocation of carbon into the shikimate pathway, culminating with an accumulation of shikimic acid and its derivatives.

A beneficial characteristic of glyphosate application to soils is its resistance to leaching. Glyphosate possesses a strong absorptivity to soil particles (Duke et al., 2003). Once bound to soil particles, glyphosate has only minimal herbicidal activity, and is quickly degraded by microbes (Araujo et al., 2003). These microbes can metabolize glyphosate by two different routes (Pollegioni et al., 2011). One is through the C-P lyase pathway in which the carbon-phosphorous bond is cleaved to form phosphate and sarcosine. The other is through an oxidative reaction in which the carbon-nitrogen bond

of glyphosate is cleaved to form aminomethylphosphonic acid (AMPA) and glyoxylate. Metabolism of glyphosate is quite rapid, judging by its limited persistence in soil. The amount of time it takes for glyphosate and its primary degradation product, AMPA, to be 50% degraded (DT_{50}) is relatively short, at 9 and 32 days, respectively (Simonsen et al., 2008). While glyphosate is metabolized rather quickly in soil, its persistence in aquatic environments is considerably longer. Examination of aquatic environments treated with glyphosate found both glyphosate and AMPA to persist for several weeks (Giesy et al., 2000). Even though most non-plant aquatic organisms do not possess EPSPS, studies have indicated their sensitivity to glyphosate. Results from short-term toxicity tests indicated that non-target aquatic species were more sensitive to glyphosate formulations than most soil microbes tested. However, much of the lethality observed may have been due to the chemicals associated with the chemical formulation instead of the herbicide itself (Sihtmäe et al., 2013). Glyphosate is commonly used in formulations as a salt, as this enhances the efficacy of the herbicide. A common glyphosate salt, polyoxyethyleneamine (POEA) appeared to harbor higher toxicity to microalgae, protozoa, and crustaceans than pure glyphosate or other formulations (Tsui and Chu, 2003).

Two basic strategies have been utilized to generate glyphosate-resistant organisms (Pollegioni et al., 2011). The first is the expression of a glyphosate-insensitive form of EPSPS in the target species, and the second is utilization of a glyphosate-

metabolizing enzyme to detoxify the compound. Currently, two forms of EPSPS are known. The first, known as Class I EPSPSs, are naturally sensitive to glyphosate. Some of these forms of glyphosate have been modified to become glyphosate-resistant, such as the TIPS double mutant EPSPS (Pollegioni et al., 2011). The second form, known as Class II EPSPSs, includes the EPSPS from *Agrobacterium* sp. strain CP4. These proteins are structurally distinct from Class I, and are natively insensitive to glyphosate. Both classes of EPSPS have been employed in the creation of glyphosate-resistant plants. While the mutated Class I and native Class II EPSPSs are both resistant to glyphosate, Class II EPSPSs show a more robust tolerance to the herbicide, and have been favored in use in multiple agricultural crop plants. Glyphosate-detoxifying proteins have also been utilized in the development of glyphosate-resistant crops. Introduction of the glyphosate oxidase (GOX) gene from *Ochrobactrum anthropi* st. LBAA in parallel with the *Agrobacterium* sp. strain CP4 EPSPS successfully conferred glyphosate resistance to wheat (*Triticum* spp.), sugar beet (*Beta vulgaris*), and canola (*Brassica napus*) (Zhou et al., 1995; Mannerlöf et al., 1997; Monsanto, 2002). It should be noted, however, that the GOX gene alone has limited ability to impart glyphosate resistance in transgenic plants. In 2004, an enzyme was found in *Bacillus licheniformis* that showed weak use of glyphosate as a substrate in an acetylation reaction (Castle et al., 2004). The resultant product, N-acetylglyphosate, displayed none of the toxic characteristics of glyphosate. Use of DNA shuffling introduced 21 substitutions into this enzyme (glyphosate acetyltransferase, GAT), and improved its catalytic efficiency against glyphosate by 4

orders of magnitude. The modified GAT was shown to successfully confer resistance to glyphosate in E. coli, Arabidopsis, tobacco, and maize. Sometime later, the crystal structure of the modified GAT was solved, revealing that, of the 21 substitutions, four occurred within the active site (Siehl et al., 2007). The use of these two independent technologies (glyphosate-insensitive EPSPS and glyphosate-metabolizing enzymes) has been used by the academic community, both as a selectable marker in laboratory studies, and for maintenance of isogenic plant populations for field experiments and analysis.

Norflurazon and Phytoene Desaturase

A third class of herbicides used in studies presented in this thesis, herbicide class 12, targets the enzyme activity of phytoene desaturase (PDS), a key component in the biosynthesis of carotenoids. PDS is responsible for the conversion of phytoene to ζ-carotene. Carotenoids generated from this product participate in light harvesting reactions, as well as protect the chloroplast from deleterious effects of singlet oxygen that arises during photosynthesis (Sandmann and Boger, 1997). By inhibiting the activity of PDS, carotenoid production is halted. This results in a degradation of chlorophyll and destruction of chloroplast membranes (Böger and Sandmann, 1998). Destruction of chloroplast membranes produces an easily-recognized phenotype among photosynthetic organisms; namely, photobleaching of green tissues. Several compounds are classified as PDS-inhibiting, including fluridone (CAS: 59756-60-4) and norflurazon

(CAS: 27314-13-2). Fluridone, commonly marketed under trade name Avast!, has been frequently used as an herbicide in aquatic environments, while norflurazon (trade name Zorial) has been used as a selective pre-emergent herbicide (EPA, 1996; WS-DNR, 2012). Norflurazon was the first to be introduced into use in the United States (1974), with fluridone following shortly thereafter (1986). Studies of the mode of action of these compounds have shown that PDS inhibitors do not compete with phytoene for binding in PDS, but rather, that these compounds compete with the essential cofactor plastoquinone for its binding site on PDS (Norris et al., 1995; Breitenbach et al., 2001; Sandmann and Mitchell, 2001). The gene crtI, which catalyzes four desaturase steps, including that of PDS, is found in gram negative bacteria such as *Pantoea ananatis* (formerly *Erwinia uredovora*), and is naturally insensitive to PDS-inhibiting compounds (Misawa et al., 1993). While crtI is not a true classical plant-type PDS (it catalyzes four reactions, including the two desaturase reactions carried out by plant PDS), it can function in other organisms, and has been shown to confer herbicide resistance in transgenic plants (Misawa et al., 1993).

Several mutant organisms that are resistant to PDS-inhibiting herbicides have been characterized. Some of the first norflurazon-resistant mutants to be identified were found in cyanobacteria (Linden et al., 1989; Chamovitz et al., 1993). In particular, mutants of a marine (*Synechococcus* PCC 7942) and freshwater (*Synechocystis* PCC6803) cyanobacteria were generated that both displayed resistance to norflurazon (Chamovitz et al., 1991; Martinez-Ferez and Vioque, 1992). Despite sequence conservation in PDS

between the two species, the location of the mutation in each gene was different. The PDS mutant from *Synnechococcus* (termed NFZ 4) showed an ability to confer resistance to norflurazon when it was modified for expression in higher plants (Wagner et al., 2002). The first report of a higher plant to display resistance to norflurazon was in *Hydrilla verticillata* (Michel et al., 2004b). *Hydrilla*, an invasive species of aquatic plant, was first introduced into the United States in the 1950s, and for several years has been actively treated with fluridone to limit its spread (Arias et al., 2005). Interestingly, of the three characterized PDS mutants in *Hydrilla* populations that showed herbicide resistance, all three carried mutations in one position of the PDS gene. In each case residue Arg403 was mutated to Ser, Cys, or His. While most characterized norflurazon-resistant PDS mutants usually showed a reduction in the specific activity of the enzyme (Chamovitz et al., 1993), all three mutants of *Hydrilla* showed specific activities similar to the WT enzyme (Michel et al., 2004b). Further tests of these mutants showed that substitution of Arg with Thr conferred even higher levels of resistance to norflurazon (Arias et al., 2006). In addition, transformation of this mutant form of PDS into Arabidopsis generated norflurazon-resistant plants, highlighting the ability of this mutant to function across species. A patent has been placed on this technology for use both as a selectable marker and in conferring herbicide resistance in multiple plants (Michel et al., 2004a).

Not surprisingly, Chlamydomonas has shown sensitivity to PDS-inhibiting compounds. Norflurazon, for example, is reported to have an ED_{50} of 0.73 mg/L

(2.40μM)(Reboud, 2002). Chlamydomonas cells treated with PDS-inhibiting compounds show a reduction in carotenoid production and eventually become photobleached. If cells are grown on heterotrophic media and maintained in dark conditions, they are able to survive higher levels of norflurazon, though still retaining the characteristic white phenotype. Characterization of PDS mutants in Chlamydomonas built upon these earlier observations (Tran et al., 2012). One such mutant, *pds1-1*, displayed a partial inactivation of PDS; generating ~5% of the levels of carotenoids seen in WT cells. This mutant was found to contain an amino acid substitution occurring in the predicted dinucleotide (FAD)-dependent oxidoreductase/amine oxidase domain. Further mutation resulted in an enzymatically inactive PDS, named *pds1-2*. Interestingly, intragenic suppressors of *pds1-2* were successfully generated. Recent analysis (described in a chapter of this thesis) of the PDS protein across Chlamydomonas, *Hydrilla,* and the cyanobacteria *Synnechococcus* and *Synechocystis* show a high level of conservation at the amino acid sequence level. Of particular note, the locations of amino acid mutations that confer norflurazon-resistance in *Hydrilla* and *Synechocystis* are identical, and all four organisms show high conservation in the amino acid sequence surrounding this region, suggesting its importance in the function of PDS and the importance of the arginine residue in conferring herbicide resistance in mutant lines.

References

Araujo, A.S.F., Monteiro, R.T.R., and Abarkeli, R.B. (2003). Effect of glyphosate on the microbial activity of two Brazilian soils. Chemosphere **52,** 799-804.

Arias, R.S., Netherland, M.D., Scheffler, B.E., Puri, A., and Dayan, F.E. (2005). Molecular evolution of herbicide resistance to phytoene desaturase inhibitors in Hydrilla verticillata and its potential use to generate herbicide-resistant crops. Pest Manag Sci **61,** 258-268.

Arias, R.S., Dayan, F.E., Michel, A., Howell, J., and Scheffler, B.E. (2006). Characterization of a higher plant herbicide-resistant phytoene desaturase and its use as a selectable marker. Plant Biotechnol J **4,** 263-273.

Böger, P., and Sandmann, G. (1998). Carotenoid biosynthesis inhibitor herbicides - mode of action and resistance mechanisms. Pesticide Outlook **9,** 29-35.

Breitenbach, J., Zhu, C.F., and Sandmann, G. (2001). Bleaching herbicide norflurazon inhibits phytoene desaturase by competition with the cofactors. Journal of Agricultural and Food Chemistry **49,** 5270-5272.

Castle, L.A., Siehl, D.L., Gorton, R., Patten, P.A., Chen, Y.H., Bertain, S., . . . Lassner, M.W. (2004). Discovery and directed evolution of a glyphosate tolerance gene. Science **304,** 1151-1154.

Cerdeira, A.L., and Duke, S.O. (2010). Effects of glyphosate-resistant crop cultivation on soil and water quality. GM Crops **1,** 16-24.

Chamovitz, D., Pecker, I., and Hirschberg, J. (1991). The molecular basis of resistance to the herbicide norflurazon. Plant Molecular Biology **16,** 967-974.

Chamovitz, D., Sandmann, G., and Hirschberg, J. (1993). Molecular and biochemical characterization of herbicide-resistant mutants of cyanobacteria reveals that phytoene desaturation is a rate-limiting step in carotenoid biosynthesis. J Biol Chem **268,** 17348-17353.

Das, A.C., Debnath, A., and Mukherjee, D. (2003). Effect of the herbicides oxadiazon and oxyfluorfen on phosphates solubilizing microorganisms and their persistence in rice fields. Chemosphere **53,** 217-221.

Dayan, F.E., and Duke, S.O. (1997). Overview of protoporphyrinogen oxidase-inhibiting herbicides. Proceedings of the Brighton Crop Protection Conference: Weeds**,** 83-92.

Duke, S.O., Baerson, S.R., and Rimando, A.M. (2003). Herbicides: Glyphosate. In Encyclopedia of agrochemicals, J.R. Plimmer, D.W. Gammon, N.N. Ragsdale, and Knovel (Firm), eds (Hoboken, N.J.: Wiley-Interscience,), pp. 3 v. (x, 1638 p.).

Duke, S.O., Lydon, J., Becerril, J.M., Sherman, T.D., Lehnen, L.P., and Matsumoto, H. (1991). Protoporphyrinogen Oxidase-Inhibiting Herbicides. Weed Science **39,** 465-473.

EPA, U.S. (1996). R.E.D. Facts: Norflurazon, U.S.E.P. Agency, ed (U.S. Environmental Protection Agency, Prevention, Pesticides and Toxic Substances).

EPA, U.S. (2002). Reregistration Eligibility Decision (RED) - Oxyfluorfen, U.S.E.P. Agency, ed.

EPA, U.S. (2003). Reregistration Eligibility Decision for Oxadiazon, U.S.E.P. Agency, ed.

Eschenburg, S., Kabsch, W., Healy, M.L., and Schonbrunn, E. (2003). A new view of the mechanisms of UDP-N-acetylglucosamine enolpyruvyl transferase (MurA) and 5-enolpyruvylshikimate-3-phosphate synthase (AroA) derived from X-ray structures of their tetrahedral reaction intermediate states. J Biol Chem **278,** 49215-49222.

Funke, T., Han, H., Healy-Fried, M.L., Fischer, M., and Schonbrunn, E. (2006). Molecular basis for the herbicide resistance of Roundup Ready crops. Proceedings of the National Academy of Sciences of the United States of America **103,** 13010-13015.

Giesy, J.P., Dobson, S., and Solomon, K.R. (2000). Ecotoxicological risk assessment for Roundup (R) Herbicide. Reviews of Environmental Contamination and Toxicology, Vol 167 **167,** 35-120.

Grube, A., Donaldson, D., Kiely, T., and Wu, L. (2011). Pesticides Industry Sales and Usage: 2006 and 2007 Market Estimates, U.S.E.P. Agency, ed (Washington, DC).

Ha, S.B., Lee, S.B., Lee, Y., Yang, K., Lee, N., Jang, S.M., . . . Back, K. (2004). The plastidic Arabidopsis protoporphyrinogen IX oxidase gene, with or without the transit sequence, confers resistance to the diphenyl ether herbicide in rice. Plant Cell and Environment **27,** 79-88.

Hao, G.F., Zuo, Y., Yang, S.G., and Yang, G.F. (2011). Protoporphyrinogen Oxidase Inhibitor: An Ideal Target for Herbicide Discovery. Chimia **65,** 961-969.

Jung, S., Lee, Y., Yang, K., Lee, S.B., Jang, S.M., Ha, S.B., and Back, K. (2004). Dual targeting of Myxococcus xanthus protoporphyrinogen oxidase into chloroplasts and mitochondria and high level oxyfluorfen resistance. Plant Cell and Environment **27,** 1436-1446.

Kataoka, M., Sato, R., and Oshio, H. (1990). Isolation and partial characterization of mutant Chlamydomonas reinhardtii resistant to herbicide S-23142. Journal of Pesticide Science **15,** 449-451.

Lermontova, I., and Grimm, B. (2000). Overexpression of plastidic protoporphyrinogen IX oxidase leads to resistance to the diphenyl-ether herbicide acifluorfen. Plant Physiol **122,** 75-84.

Li, X., and Nicholl, D. (2005). Development of PPO inhibitor-resistant cultures and crops. Pest Manag Sci **61,** 277-285.

Linden, H., Sandmann, G., Chamovitz, D., Hirschberg, J., and Boger, P. (1989). iochemical characterization of herbicide resistant Anacystis mutants selected against the bleaching herbicide norflurazon. Pesticide Biochemistry and Physiology **36,** 46-51.

Mannerlöf, M., Tuvesson, S., Steen, P., and Tenning, P. (1997). Transgenic sugar beet tolerant to glyphosate. Euphytica **94,** 83-91.

Martinez-Ferez, I.M., and Vioque, A. (1992). Nucleotide sequence of the phytoene desaturase gene from Synechocystis sp. PCC 6803 and characterization of a new mutation which confers resistance to the herbicide norflurazon. Plant Molecular Biology **18,** 981-983.

Matringe, M., Camadro, J.M., and Brouillet, N. (1993). Protoporphyrinogen oxidase, the molecular target site for peroxidizing herbicides. Proceedings of the Brighton Crop Protection Conference: Weeds**,** 703-711.

Matsumoto, H. (2002). Inhibitors of protoporphyrinogen oxidase: a brief update. In Herbicide Classes in Development, P. Boger, K. Wakabayashi, and K. Hirai, eds (Berlin-Heidelberg: Springer), pp. 151-161.

Michel, A., Scheffler, B.E., Netherland, M., Dayan, F.E., and Arias de Ares, R.S. (2004a). Herbicide-Resistant Plants, and Polynucleotides and Methods for Providing Same.

Michel, A., Arias, R.S., Scheffler, B.E., Duke, S.O., Netherland, M., and Dayan, F.E. (2004b). Somatic mutation-mediated evolution of herbicide resistance in the nonindigenous invasive plant hydrilla (Hydrilla verticillata). Mol Ecol **13,** 3229-3237.

Misawa, N., Yamano, S., Linden, H., de Felipe, M.R., Lucas, M., Ikenaga, H., and Sandmann, G. (1993). Functional expression of the Erwinia uredovora carotenoid biosynthesis gene crtl in transgenic plants showing an increase of beta-carotene biosynthesis activity and resistance to the bleaching herbicide norflurazon. Plant J **4,** 833-840.

Monsanto. (2002). Safety assessment of roundup ready canola event GT73.

Monsanto. (2005). Backgrounder - History of Monsanto's Glyphosate Herbicides.

Morgan, R.R., Errington, R., and Elder, G.H. (2004). Identification of sequences required for the import of human protoporphyrinogen oxidase to mitochondria. Biochemical Journal **377,** 281-287.

Norris, S.R., Barrette, T.R., and DellaPenna, D. (1995). Genetic dissection of carotenoid synthesis in arabidopsis defines plastoquinone as an essential component of phytoene desaturation. Plant Cell **7,** 2139-2149.

Park, H., Hilsenbeck, J.L., Kim, H.J., Shuttleworth, W.A., Park, Y.H., Evans, J.N., and Kang, C. (2004). Structural studies of Streptococcus pneumoniae EPSP synthase in unliganded state, tetrahedral intermediate-bound state and S3P-GLP-bound state. Molecular Microbiology **51,** 963-971.

Pollegioni, L., Schonbrunn, E., and Siehl, D. (2011). Molecular basis of glyphosate resistance-different approaches through protein engineering. FEBS J **278,** 2753-2766.

Qasem, J.R., and Foy, C.L. (2006). Selective weed control in syrian marjoram (Origanum syriacum) with oxadiazon and oxyfluorfen herbicides. Weed Technology **20,** 670-676.

Randolph-Anderson, B.L., Sato, R., Johnson, A.M., Harris, E.H., Hauser, C.R., Oeda, K., . . . Boynton, J.E. (1998). Isolation and characterization of a mutant protoporphyrinogen oxidase gene from Chlamydomonas reinhardtii conferring resistance to porphyric herbicides. Plant Molecular Biology **38,** 839-859.

Reboud, X. (2002). Response of Chlamydomonas reinhardtii to herbicides: negative relationship between toxicity and water solubility across several herbicide families. Bull Environ Contam Toxicol **69,** 554-561.

Sandmann, G., and Boger, P. (1997). Phytoene desaturase as a target for bleaching herbicides. In Herbicide Activity: Toxicology, Biochemistry, and Molecular Biology, R.M. Roe, J.D. Burton, and R.J. Kuhr, eds (Amsterdam: IOS Press), pp. 1-10.

Sandmann, G., and Mitchell, G. (2001). In vitro inhibition studies of phytoene desaturase by bleaching ketomorpholine derivatives. J Agric Food Chem **49,** 138-141.

Schonbrunn, E., Eschenburg, S., Shuttleworth, W.A., Schloss, J.V., Amrhein, N., Evans, J.N., and Kabsch, W. (2001). Interaction of the herbicide glyphosate with its target enzyme 5-enolpyruvylshikimate 3-phosphate synthase in atomic detail. Proc Natl Acad Sci U S A **98,** 1376-1380.

Siehl, D.L., Castle, L.A., Gorton, R., and Keenan, R.J. (2007). The molecular basis of glyphosate resistance by an optimized microbial acetyltransferase. J Biol Chem **282,** 11446-11455.

Sihtmäe, M., Blinova, I., Künnis-Beres, K., Kanarbik, L., Heinlaan, M., and Kahru, A. (2013). Ecotoxicological effects of different glyphosate formulations. Applied Soil Ecology **72,** 215-224.

Simonsen, L., Fomsgaard, I.S., Svensmark, B., and Spliid, N.H. (2008). Fate and availability of glyphosate and AMPA in agricultural soil. J Environ Sci Health B **43,** 365-375.

Steinrucken, H.C., and Amrhein, N. (1980). The herbicide glyphosate is a potent inhibitor of 5-enolpyruvyl-shikimic acid-3-phosphate synthase. Biochem Biophys Res Commun **94,** 1207-1212.

Tran, P.T., Sharifi, M.N., Poddar, S., Dent, R.M., and Niyogi, K.K. (2012). Intragenic enhancers and suppressors of phytoene desaturase mutations in Chlamydomonas reinhardtii. PLoS One **7,** e42196.

Tsui, M.T., and Chu, L.M. (2003). Aquatic toxicity of glyphosate-based formulations: comparison between different organisms and the effects of environmental factors. Chemosphere **52,** 1189-1197.

Vencill, W.K. (2002). Herbicide Handbook. (Lawrence, KS, USA: Weed Science Society of America).

Wagner, T., Windhovel, U., and Romer, S. (2002). Bansformation of tobacco with a mutated cyanobacterial phytoene desaturase gene confers resistance to bleaching herbicides. Z Naturforsch C **57,** 671-679.

Wakabayashi, K., and Boger, P. (1999). General physiological characteristics and mode of action of peroxidizing herbicides. In Herbicide Classes in Development Mode of Action, Targets, Genetic Engineering, Chemistry, P. Boger, K. Wakabayashi, and K. Hirai, eds (Berlin, Heidelberg: Springer Berlin Heidelberg,), pp. 164-190.

Warabi, E., Usui, K., Tanaka, Y., and Matsumoto, H. (2001). Resistance of a soybean cell line to oxyfluorfen by overproduction of mitochondrial protoporphyrinogen oxidase. Pest Manag Sci **57,** 743-748.

Watanabe, N., Che, F.S., Iwano, M., Takayama, S., Yoshida, S., and Isogai, A. (2001). Dual targeting of spinach protoporphyrinogen oxidase II to mitochondria and chloroplasts by alternative use of two in-frame initiation codons. J Biol Chem **276,** 20474-20481.

Williams, G.M., Kroes, R., and Munro, I.C. (2000). Safety evaluation and risk assessment of the herbicide Roundup and its active ingredient, glyphosate, for humans. Regul Toxicol Pharmacol **31,** 117-165.

WS-DNR. (2012). Fluridone Chemical Fact Sheet (Wisconsin Department of Natural Resources).

Zhou, H., Arrowsmith, J.W., Fromm, M.E., Hironaka, C.M., Taylor, M.L., Rodriguez, D., . . . Fry, J.E. (1995). Glyphosate-tolerant CP4 and GOX genes as a selectable marker in wheat transformation. Plant Cell Reports **15,** 159-163.

CHAPTER 1.3 - TAL Effectors

The ability to knock out, or replace, genes of interest in an organism is one of the most powerful tools in molecular biology. Scientists studying model microorganisms have often relied on knock-out of a gene (and perhaps subsequent replacement) to ascertain its function. Even those working on crop enhancement have often discovered (after the fact) that certain beneficial traits were the result of a spurious gene deletion or inactivation. The ability to create deletions or insertions in specific genes was, however, significantly limited for many years in higher eukaryotes, due to the inability of scientists to precisely target genes of interest. Early methods relied on random mutagenesis (via chemical mutagenesis, neutron bombardment, etc.) and laborious screening of individual mutants for those displaying desired phenotypes. In some cases, scientists would expend considerable effort in an attempt to identify the gene(s) responsible for the observed phenotype. More modern approaches to elucidating gene function have followed a reverse genetics route. By knocking out genes of interest, scientists can look at the resultant phenotype to make conclusions regarding the function of the altered gene. With the random integration of gene fragments, either through transformation with foreign DNA or by inducing transposon movement, scientists could generate disruptions throughout a genome. While integration by these fragments into the genome was largely random, large numbers of gene knockouts could be created and studied. Nonetheless, the types and power of tools available to most scientists working with complex, multicellular organisms have been severely limited

compared to tools available to those utilizing organisms such as *Saccharomyces cerevisiae* (Rothstein, 1983) in which gene replacement by homologous recombination has paved the way to rapid progress. In such organisms, simply introducing a piece of DNA with regions of sequence identical to a genomic target has allowed for recombination of the foreign DNA into the genome. By engineering the foreign DNA to carry a deletion, insertion, or even a completely new gene, researchers were able to generate knock outs or gene replacements at precise locations in a genome. Only recently have new techniques begun to emerge that offer growing promise that facile and efficient techniques for gene knockout and gene replacement may soon be available to scientists wishing to modify higher eukaryotes for academic as well as commercial purposes.

Early tests in yeast (Rudin et al., 1989; Plessis et al., 1992) and mammalian cells (Rouet et al., 1994; Choulika et al., 1995) indicated that if researchers were able to make a double-stranded break (DSB) in the DNA of a target region, the level of homologous recombination between the target and the foreign DNA increased several orders of magnitude. Since these discoveries, there has been a huge increase in the attempts to develop a robust system of creating DSBs at specific sites in the genome.

The prototype enzymes for creating rare, but sequence-specific, DSBs were the homing nucleases I-*SceI* and HO (Carroll, 2011). While providing useful information on the efficiency of DSB repair, these enzymes were restricted in application due to their limited DNA recognition motif. A more robust tool would be one that could be created

de novo to target pre-determined DNA sequences. Early attempts were made to modify the recognition specificity of I-*SceI* to target new DNA sequences (Ashworth et al., 2006; Paques and Duchateau, 2007). While successful in certain limited cases, the difficulty in clearly associating protein structure modifications and DNA recognition makes engineering of these meganucleases challenging.

Zinc Finger Nucleases

The first development of a chimeric protein possessing both DNA recognition capabilities and DNA cleavage activity was fusion of a zinc finger protein with the cleavage domain of *FokI* nuclease (Li et al., 1992) to produce a zinc finger nuclease (ZFN). By replacing the native DNA recognition domain of *FokI* with a zinc finger molecule, scientists could construct proteins containing active *FokI* nuclease domains to target DNA sequences of their choosing for DSBs. Further studies discovered that the *FokI* nuclease domain required dimerization to cleave DNA (Bitinaite et al., 1998; Smith et al., 2000). By designing two of these zinc fingers-nuclease (ZFN) proteins to bind neighboring head-to-head DNA sequences, simultaneous binding of both proteins would allow dimerization of the *FokI* nuclease domains and cause cleavage of the DNA between the two target sites. A DSB in an organism's DNA is repaired by one of two methods, non-homologous end joining (NHEJ) or homologous recombination (HR) (Doyle et al., 2013). During NHEJ DNA repair, repair is sometimes imprecise, and occasional nucleotide deletions or insertions are introduced at the break site. If the

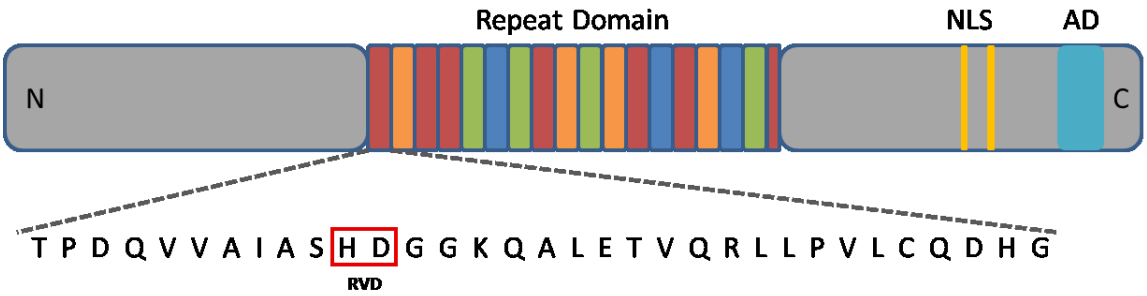
repair is within a coding region, the insertion or deletion often places the gene out of frame and a gene knockout is created. During DNA repair by HR, a DNA fragment possessing homology to the flanking regions of the DSB is incorporated in place of an existing DNA strand, allowing for introduction of a new DNA fragment. The application of ZFNs in targeted gene modification in various organisms was quickly tested, and proved to be a successful means of gene editing. Later, multiple species of plants, such as Arabidopsis (Osakabe et al., 2010; Zhang et al., 2010), tobacco (Townsend et al., 2009), maize(Shukla et al., 2009), and soybean (Curtin et al., 2011) were demonstrated to be compatible with ZFN technology. ZFNs are not, however, without limitations (DeFrancesco, 2011). Each "finger" domain of a zinc finger protein targets a three-nucleotide stretch of DNA sequence, and linkage of multiple domains, while extending the length of the DNA "target" sequence, does not always translate to an increase in specificity of the protein. Frequently, many ZFNs must be built and tested to identify one with the best binding ability. In addition, ZFNs are costly and can suffer from a low targeting efficiency and specificity, causing frequent off-target cleavage events throughout the genome. These factors, while not eliminating the usefulness of ZFNs in targeted gene editing, have left ample room for improvements in this technology.

The Discovery of TALE Effectors and Elucidation of Their Mechanism for Precise DNA Binding

A few years after the advent of ZFNs, a new class of DNA-binding proteins was studied to determine their mechanism for DNA recognition. These Transcription Activator-Like Effector (TALE) proteins were one of a myriad of effector proteins utilized by the bacterium *Xanthomonas* to induce pathogenicity in various host organisms (Sugio et al., 2007). These "effector" proteins are injected into host cells through Type III secretion systems. Importantly, each of several TALE effectors discovered were found to function by inducing transcription of a particular nuclear gene (Bogdanove et al., 2010). One particular set of TALE proteins from *Xanthomonas oryzae* were found to target, and induce expression of, a family of genes encoding sugar exporters known as SWEET proteins in rice (Yang et al., 2006; Antony et al., 2010; Yu et al., 2011; Verdier et al., 2012). Induction of these genes would direct photosynthetically-derived sucrose into the extracellular matrix, making this sugar easily available for consumption by the invading bacteria (Chen et al., 2012). While TALE proteins were commonly associated with plant susceptibility to bacterial blight, a few rare instances were noted in which the host plant appeared to garner resistance to the pathogen directly from exposure to the TALE protein. For example, the rice Xa27 gene was shown to be activated by a TALE upon infection by *Xanthomonas* (Gu et al., 2005) and this activation caused induction of a resistance response that halted bacterial infection. The prime question that evolved from these studies was: "How does a particular TAL Effector recognize as specific DNA sequence in its target gene?"

TALE proteins were shown to share a conserved N-terminus required for injection of a TAL Effector into the host cell by the bacterium's Type III secretion system as well as a C-terminus possessing nuclear localization signals (NLS) and an acidic domain involved in gene activation (AD). Most interesting, however, was the central domain of these proteins. The mid region of TALEs is composed of several highly-conserved repeats of 33-34 amino acids each, followed by a single truncated repeat composed of 20 amino acids. These repeats are highly conserved, with the exception of two amino acids in the 12^{th} and 13^{th} position. When analyzing these so-called repeat variable diresidues (RVDs) at the 12^{th} and 13^{th} position, it became clear that these amino acids, while not as conserved as the rest of the repeat unit, are limited in their makeup. In particular, four RVDs (HD, NG, NI, & NN), account for nearly 75% of the total RVDs found in nature (Moscou and Bogdanove, 2009). In independent investigations, two groups elucidated the mechanism for DNA recognition by TALEs (Boch et al., 2009; Moscou and Bogdanove, 2009). They discovered that each repeat unit was responsible for binding a single nucleotide, and the RVD gave the repeat specificity in regard to the nucleotide it bound. The most common RVD:nucleotide association (i.e., code) was determined to be HD:C, NG:T, NI:A, and NN:G/A. The structure and organization of a TAL Effector is depicted below.

Figure 1.2 - Transcription Activator-Like Effector (TALE) Proteins. The top figure depicts a native TAL Effector protein, with the repeat domain, nuclear localization signals (NLS), and activation domain (AD) labeled. The sequence composition of each repeat unit is depicted below the model, with the 12th and 13th residues (Repeat Variable Diresidue, RVD) marked in red. The chart highlights four RVDs and their corresponding nucleotide specificity.



RVD	Nucleotide
NI	A
HD	C
NN	A/G
NG	T

TALE Structure and Function

NMR analysis of a 1.5 repeat unit shed light onto the structural functionality of TALE DNA binding (Murakami et al., 2010). It was determined that the TALE repeat unit formed two antiparallel α-helices, similar to the fold found in tetratricopeptide repeats (TPRs). The RVD was found to be located within the turn between the two helices. Later, both a native TALE and a synthetic TALE were successfully crystalized bound to DNA, confirming early observations regarding the structure of repeats (Deng et al., 2012; Mak et al., 2012). This full TALE structure showed that the repeats within the TALE self-associate and form a right-handed superhelix that wraps around the major groove of the target DNA. Interestingly, only the second amino acid in the RVD makes base-specific contact with the DNA, while the first amino acid forms a stabilizing contact with the protein backbone. The crystal structure also provided insight into the mechanics of RVDs and methylated DNA; specifically that of cytosine methylated at the C5 position (5mC). The RVD that normally recognizes T (NG) also recognizes and binds 5mC, but not unmethylated C. Interestingly, the normal RVD for recognizing C (HD) was unable to bind 5mC. Experimental evidence had shown that every TALE DNA recognition site was flanked by a T nucleotide (position 0). Because none of the known TALEs contained a T-binding 34-amino acid repeat unit at the beginning of their repeat domains, it was surmised that the portion of the TALE N-terminus next to the repeat domain was responsible for binding this T nucleotide. Indeed, structural analysis determined that, despite lacking sequence similarity to the canonical repeats, the N-terminus of each

TALE repeat domain formed two repeat-like helical bundles, referred to as cryptic repeats (Gao et al., 2012). These cryptic repeats, designated as positions 0 and -1, engaged the T nucleotide directly upstream of the target sequence.

Methods for Producing Synthetic TALEs and TALENs

Soon after the discovery of TALE modularity, multiple efforts were made to assemble TALE repeats in an ordered fashion to bind to a pre-selected DNA sequence. Currently, two companies, Cellectis Bioresearch and Life Technologies, offer services for assembly of custom TALEs. However, commercial purchase of TALEs is cost prohibitive for most research groups, and many in-house methods of TALE assembly have been developed and released for use by individual laboratories. These general methods for TALE assembly usually fall into one of three categories: standard cloning assembly, Golden Gate assembly, or solid-phase assembly (Chen and Gao, 2013). Researchers can obtain individual components or multicomponent kits for several of these platforms from Addgene.

Standard methods for TALE construction rely on large libraries of TALE repeat units that can be selected and ligated in a directional fashion. In one method, researchers assembled TALE repeat units into groups of two (2-mers) (Huang et al., 2011). This was followed by a large ligation of the different 2-mers together into a vector containing the TALE N-terminus, C-terminus, and FokI nuclease domain. In parallel, a method termed REAL TALE assembly was published in which TALE repeats

were ligated in an additive fashion (Sander et al., 2011). Individual TALE repeats were initially synthesized with predetermined variations in the nucleotide sequence. Each repeat was flanked by one of two Type IIs restriction enzymes, BbsI and BsaI. Unlike standard restriction enzymes, these enzymes cut downstream of their DNA binding site. By positioning the enzyme binding site within a certain distance of the TALEs, digestion of two TALEs would yield fragments with unique overhangs, allowing for directional ligation of the fragments. Ligation of two repeats would eliminate the restriction sites on the internal junction, but external portions of the repeats would still possess the necessary restriction site for additional steps in TALE construction. In this manner, repeat units could be assembled to any desired length before being transferred to a vector containing the N- and C-termini of the TALE. A modification to this method was achieved by expanding the library of TALE repeats to include pre-assembled units containing two or more repeats (Reyon et al., 2012). A method of TALEN synthesis developed by researchers at Iowa State University utilized an internal BsmBI site present in each TALE repeat unit (Li et al., 2011b). By synthesizing repeats to contain various nucleotides within the cleavage site of BsmBI (another Type IIs restriction enzyme containing physically separated DNA binding and cleavage domains), repeats could be assembled in a directional manner. These methods, while successful for early attempts at TALE assembly, required significant time inputs, as DNA fragments had to be digested, purified, and then combined for ligation. Because of the cumbersome nature

of these methods, they were soon replaced by the more facile Golden Gate cloning system that involved fewer steps and hastened TALE assembly.

The Golden Gate method for DNA cloning, developed in 2008, utilized the unique characteristic of the Type IIs restriction enzyme BsaI, mentioned earlier, to allow for seamless ligation of DNA fragments (Engler et al., 2008). By having separate binding and cleavage domains, BsaI creates a staggered cut in DNA at 1 and 5bp downstream of its DNA binding site, creating a 4bp overhang that is unique to the DNA sequence flanking the binding site. In this manner, multiple BsaI sites could all yield unique overhangs, depending on the neighboring sequences at each site. In Golden Gate cloning, custom DNA primers are used to insert desired BsaI restriction enzyme sites into the ends of a DNA fragment of interest during PCR amplification. The introduced BsaI sites are 1bp upstream and 1bp downstream of the DNA sequence of interest. Similarly, the target plasmid is PCR modified to contain 4bp extensions, matching the ends of the insert, on both sides of the desired location for DNA insertion. Thus, like the insert, the target plasmid also contains strategically placed BsaI sites. When digested, the insert and target will produce complementary overhangs that, when combined, will result in a seamless ligation that lacks any remnant restriction site. Due to the compatibility of BsaI and T4 DNA ligase in reaction buffers, both enzymes can be combined along with insert DNA and the plasmid vector, and digestion and ligation can take place in a single reaction tube. Successful ligation of the insert into the target vector will result in an intact plasmid that lacks any BsaI sites. Thus, these plasmids are impervious to

additional rounds of digestion with BsaI. Following multiple cycles of digestion and ligation, the majority of ligated products will be the desired cloning product. Any other ligation combinations will continue to retain BsaI sites and can be digested subsequently, if necessary. Utilizing different antibiotic resistance markers on the insert and target plasmids helps to ensure that, following transformation of the ligation mixture into E. coli cells, plating cells on antibiotics specific to the insert and to the plasmid will yield cells carrying a copy of the final desired ligation product. The result is a highly efficient method to digest and ligate fragments at the same time, and produce a construct lacking additional nucleotides due to insertion of new restriction sites.

Using the Golden Gate cloning strategy, TALE repeats are often assembled in small groups before they are combined into the final TALE. Cermak et al. (2011) used BsaI to construct multiple TALE repeats, ranging in size from 1-10 total repeats. These units were then ligated with other larger units into a vector containing the N- and C-termini of the TALE protein. This general 2-step method of TALE construction (small groups of repeats first, followed by full TALE construction) has been utilized by several research groups - highlighting the robust nature of Golden Gate cloning for TALE construction (Geissler et al., 2011; Morbitzer et al., 2011; Weber et al., 2011; Zhang et al., 2011).

While Golden Gate cloning of TALEs has proven successful for many labs, construction of each TALE still requires significant time, and can be a hindrance for labs interested in producing large (>10) numbers of TALEs. With this in mind, several

research groups have explored methods to enhance TALE assembly for large-scale production. One method, coined FLASH (for Fast Ligation-based Automatable Solid-phase High-throughput) (Reyon, Tsai, et al., 2012), utilizes a library of plasmids each carrying 1 of 376 effector repeat combinations. Each plasmid contains 1, 2, 3, or 4 repeat units. The first repeat is connected to a biotin molecule and is bound to a streptavidin column. Repeats are ligated in a semi-iterative fashion, depending on the 1-, 2-, 3-, or 4-mer selected, and washing of the column eliminates any unbound units. After final assembly, proteolytic digestion releases fully-constructed TALE repeats from the biotin tag. This method has proven highly effective for rapid construction of large numbers of TALEs. In a similar fashion, Wang et al. (2012) developed an apparatus for TALE assembly based on adhesion of TALEs to magnetic beads. This chip-based method allowed for assembly of over 100 unique TALE constructs, ranging in size from 16-20 repeat units. Briggs et al. (2012) utilized an iterative approach to TALE synthesis, in which repeat units were added individually to a growing chain. Following addition of a single repeat, a hairpin nucleotide was introduced to “cap” any incompletely extended chains. In this manner, final products contained a higher percentage of full-length chains. Most recently, a novel method of TALE assembly was published that bypasses a critical step of standard cloning reactions. In this method, termed Ligation Independent Cloning (LIC), individual repeat units are treated with T4 polymerase supplemented with only 1 type of dNTP (Schmid-Burgk et al., 2013). The exonuclease activity of T4 polymerase will chew back the 3’ end of the repeat until the dNTP can be incorporated,

leaving a large (10-30bp) overhang. When repeats containing overhangs of variable size are combined, treatment with T4 DNA ligase is unnecessary, as the association of repeat units due to complementary base pairing is strong enough to withstand the physical strains of electroporation. After these constructs are introduced into *E. coli* cells, the endogenous machinery of the cell will ligate the plasmids, allowing researchers to recover a fully intact assembled TALE.

Application of TALEs to Gene Editing

With facile methods for TALE assembly available, TALEs could be engineered to bind essentially any DNA sequence. TALEs are unique to other DNA-binding proteins in that, with the exception of zinc fingers, only TALEs display modularity in their DNA-binding characteristics (Doyle et al., 2013). Like ZFs, TALEs can also be fused to a *FokI* nuclease domain to create TALE-Nucleases (TALENs). TALENs can be used to create double strand DNA breaks (DSBs) at any of several locations in most genes. Following initial demonstrations that TALENs can accurately target specific DNA sequences both *in vitro* and *in vivo* (Christian et al., 2010; Li et al., 2011a), designer TALENs have been shown to selectively disrupt gene sequences in several model organisms, including Saccharomyces (Li et al., 2011a), *C. elegans* (Wood et al., 2011), Drosophila (Liu et al., 2012), mouse (Sung et al., 2013), rat (Tesson et al., 2011; Tong et al., 2012), zebrafish (Huang et al., 2011; Sander et al., 2011; Bedell et al., 2012; Cade et al., 2012; Dahlem et al., 2012; Moore et al., 2012), human (cell lines, somatic and pluripotent) (Hockemeyer

et al., 2011; Miller et al., 2011), Xenopus (Lei et al., 2012) and Arabidopsis (Cermak et al., 2011). Adoption for use in non-model organisms also has been successful, showing the robust nature of TALENs [silkworm (Ma et al., 2012), livestock (Carlson et al., 2012), rice (Li et al., 2013), tobacco (Mahfouz et al., 2011; Zhang et al., 2013), and brachypodium (Shan et al., 2013)]. In addition to gene knockout by error-prone NHEJ, success has also been achieved in using TALENs to cause DSBs to stimulate homologous recombination of exogenously supplied DNA fragments with endogenous gene sequences for partial or complete gene replacements (Bedell et al., 2012; Zhu et al., 2013; Zu et al., 2013)

The utility of TALEs is not limited to gene editing (knockout and replacement). Activation and repression of genes have also been achieved using this technology. The native, C-terminal, activation domain (AD) of a TALE is sufficient for activation of genes in certain plants (Morbitzer et al., 2010; Li et al., 2013), while in more distant species such as human and mammalian cells, gene activation was highest when the native AD was replaced with the VP16 AD or it tetrameric derivative VP64 (Geissler et al., 2011; Zhang et al., 2011). For gene repression, TALEs alone are sometimes able to repress genes simply by binding tightly to target DNA sites (Blount et al., 2012). However, TALEs showed the highest efficacy when fused to previously-identified repressor domains. In Arabidopsis, researchers were able to repress targeted genes by fusing TALEs to an EAR-repression domain (SRDX) (Mahfouz et al., 2012). In mammalian systems, fusion of the

TALE to an mSin interaction domain (SID) proved effective in reducing transcription of the target gene (Cong et al., 2012).

CRISPR

Within the last year, a new form of programmable DNA binding has been reported. This new system, named CRISPR (Clustered Regularly Interspaced Short Palindromic Repeats), was discovered in bacteria as a defense mechanism against viral invasion (Bhaya et al., 2011; Jinek et al., 2012). In the CRISPR system, a protein, referred to as Cas, associates with two RNAs, a CRISPR RNA (crRNA) and a tracrRNA, that self-associate via complementary base pairing (Jinek et al., 2012). This RNA-protein complex then binds a specific piece of DNA through base pairing of the crRNA (in nature, DNA from an invading virus) and the Cas protein causes a DSB though use of two on-board nuclease domains. Following initial characterization of this system, it was found that by programming the crRNA to target a selected DNA sequence (such as a gene of interest), one could facilitate targeted gene editing in much the same manner as achieved with ZFNs and TALENs. An important improvement was the discovery that the crRNA and tracrRNA genes could be fused into a "single guide" RNA (sgRNA) gene that, along with a single copy of a Cas gene, was all that was required to produce a functioning gene targeting system either *in vitro* or *in vivo* (Jinek et al., 2012). Two publications in Science in January of 2013 (Cong et al., 2013; Mali et al., 2013b) provided compelling evidence that the new CRISPR/Cas/sgRNA system was functional in human cell lines and

suggested that the system likely would function in cells from most other organism as well.

The simplicity of the two-gene CRISPR system (one gene encoding the Cas9 protein and the other a hybrid sgRNA composed of the crRNA/tracrRNA chimera) offers significant convenience compared to TALENs in regard to the speed and ease of their construction. That is, only one Cas gene is needed for all experiments with a single type of organism and this gene can be permanently integrated into the genome of a host cell. To target a specific gene for editing, only a single small sgRNA gene needs to be delivered to the host cell with Cas9. CRISPR has been rapidly adopted by many laboratories as a means of targeted gene editing, and its efficacy has been shown in numerous organisms (Pennisi, 2013). Like TALEs, the CRISPR system can be repurposed for gene activation and repression. Whether the CRISPR or TALE systems will be favored by laboratories for gene editing rests not only on construction, but also specificity. Both TALENs and CRISPR show some off-target effects, though the magnitude of these effects appears variable and likely organism-specific. It is thought by some that TALENs may provide better specificity than CRISPR give that TALENs can be engineered to recognize longer target sequences (40bp) than CRISPR (15-20bp) (Doyle et al., 2013). However, recent development of an ingenious new strategy for targeted CRISPR/Cas9/sgRNA gene disruption may eliminate most “off-site” DSB DNA cleavage (Mali et al., 2013a). In this strategy, Cas9 molecules having only one functional nuclease domain (i.e., DNA nickases) are expressed along with two sgRNA genes that target DNA sequences in close

proximity to one another in the target gene. Creation of two nearby single strand DNA cuts leads to deletion of the intervening DNA fragment and likely inactivation of the target gene following NHEJ DNA repair. Because nontargeted Cas9/sgRNA binding and DNA nicking will be widely scattered throughout the genome and because any nicks generated are easily repaired without nucleotide insertions or deletions, problems with "off-target" DSB DNA cleavage by the Cas9/sgRNA system are virtually eliminated.

The Future of TALEs and TALENs

The efficiency of TALEs and TALENs in disrupting gene function appears to be variable, with some groups reporting very high levels of efficacy, with others experiencing only low levels of successful gene editing. Chen & Gao (2013) report a wide range of efficiencies testing TALENs in plant cell protoplast (with 50% of TALENs having no detectable nuclease activity). Delivery of TALENs is also critical to their efficacy with some organisms, such as wheat, that are recalcitrant to transformation (Chen and Gao, 2013).

The future applications of TALEN technology appear to be expansive. TALENs have shown their potential in creating gene modifications (knockout, replacement, activation, repression) in multiple organisms and should provide a reliable, robust tool for use in numerous academic and commercial settings. TALEN-based advances in targeted gene mutagenesis provide researchers with a way to precisely modify a gene without permanently introducing foreign DNA into an organism. For agricultural crops in

which simple cross breeding can eliminate TALENs from the plant genome, TALEN technology will likely prove invaluable for crop development (Kuzma and Kokotovich, 2011; Li et al., 2012). Not only can this technology increase the speed of crop breeding, but because “genetically enhanced” crops can be produced without permanent retention of foreign DNAs, TALEN technology likely will stimulate modifications in the current mechanisms for regulating genetically modified organisms (GMOs). In the longer term, if scientists can effectively eliminate off-site cleavage by TALENs, the technology may one day extend into medical applications.

References

Antony, G., Zhou, J., Huang, S., Li, T., Liu, B., White, F., and Yang, B. (2010). Rice xa13 recessive resistance to bacterial blight is defeated by induction of the disease susceptibility gene Os-11N3. Plant Cell **22,** 3864-3876.

Ashworth, J., Havranek, J.J., Duarte, C.M., Sussman, D., Monnat, R.J., Jr., Stoddard, B.L., and Baker, D. (2006). Computational redesign of endonuclease DNA binding and cleavage specificity. Nature **441,** 656-659.

Bedell, V.M., Wang, Y., Campbell, J.M., Poshusta, T.L., Starker, C.G., Krug, R.G., 2nd, . . . Ekker, S.C. (2012). In vivo genome editing using a high-efficiency TALEN system. Nature **491,** 114-118.

Bhaya, D., Davison, M., and Barrangou, R. (2011). CRISPR-Cas systems in bacteria and archaea: versatile small RNAs for adaptive defense and regulation. Annu Rev Genet **45,** 273-297.

Bitinaite, J., Wah, D.A., Aggarwal, A.K., and Schildkraut, I. (1998). FokI dimerization is required for DNA cleavage. Proc Natl Acad Sci U S A **95,** 10570-10575.

Blount, B.A., Weenink, T., Vasylechko, S., and Ellis, T. (2012). Rational diversification of a promoter providing fine-tuned expression and orthogonal regulation for synthetic biology. PLoS One **7,** e33279.

Boch, J., Scholze, H., Schornack, S., Landgraf, A., Hahn, S., Kay, S., . . . Bonas, U. (2009). Breaking the code of DNA binding specificity of TAL-type III effectors. Science **326,** 1509-1512.

Bogdanove, A.J., Schornack, S., and Lahaye, T. (2010). TAL effectors: finding plant genes for disease and defense. Curr Opin Plant Biol **13,** 394-401.

Briggs, A.W., Rios, X., Chari, R., Yang, L., Zhang, F., Mali, P., and Church, G.M. (2012). Iterative capped assembly: rapid and scalable synthesis of repeat-module DNA such as TAL effectors from individual monomers. Nucleic Acids Res **40,** e117.

Cade, L., Reyon, D., Hwang, W.Y., Tsai, S.Q., Patel, S., Khayter, C., . . . Yeh, J.R. (2012). Highly efficient generation of heritable zebrafish gene mutations using homo- and heterodimeric TALENs. Nucleic Acids Res **40,** 8001-8010.

Carlson, D.F., Tan, W., Lillico, S.G., Stverakova, D., Proudfoot, C., Christian, M., . . . Fahrenkrug, S.C. (2012). Efficient TALEN-mediated gene knockout in livestock. Proc Natl Acad Sci U S A **109,** 17382-17387.

Carroll, D. (2011). Genome engineering with zinc-finger nucleases. Genetics **188,** 773-782.

Cermak, T., Doyle, E.L., Christian, M., Wang, L., Zhang, Y., Schmidt, C., . . . Voytas, D.F. (2011). Efficient design and assembly of custom TALEN and other TAL effector-based constructs for DNA targeting. Nucleic Acids Res **39,** e82.

Chen, K., and Gao, C. (2013). TALENs: customizable molecular DNA scissors for genome engineering of plants. J Genet Genomics **40,** 271-279.

Chen, L.Q., Qu, X.Q., Hou, B.H., Sosso, D., Osorio, S., Fernie, A.R., and Frommer, W.B. (2012). Sucrose efflux mediated by SWEET proteins as a key step for phloem transport. Science **335,** 207-211.

Choulika, A., Perrin, A., Dujon, B., and Nicolas, J.F. (1995). Induction of homologous recombination in mammalian chromosomes by using the I-SceI system of Saccharomyces cerevisiae. Mol Cell Biol **15,** 1968-1973.

Christian, M., Cermak, T., Doyle, E.L., Schmidt, C., Zhang, F., Hummel, A., . . . Voytas, D.F. (2010). Targeting DNA double-strand breaks with TAL effector nucleases. Genetics **186,** 757-761.

Cong, L., Zhou, R., Kuo, Y.C., Cunniff, M., and Zhang, F. (2012). Comprehensive interrogation of natural TALE DNA-binding modules and transcriptional repressor domains. Nat Commun **3,** 968.

Cong, L., Ran, F.A., Cox, D., Lin, S., Barretto, R., Habib, N., . . . Zhang, F. (2013). Multiplex Genome Engineering Using CRISPR/Cas Systems. Science **339,** 819-823.

Curtin, S.J., Zhang, F., Sander, J.D., Haun, W.J., Starker, C., Baltes, N.J., . . . Stupar, R.M. (2011). Targeted mutagenesis of duplicated genes in soybean with zinc-finger nucleases. Plant Physiol **156,** 466-473.

Dahlem, T.J., Hoshijima, K., Jurynec, M.J., Gunther, D., Starker, C.G., Locke, A.S., . . . Grunwald, D.J. (2012). Simple methods for generating and detecting locus-specific mutations induced with TALENs in the zebrafish genome. PLoS Genet **8,** e1002861.

DeFrancesco. (2011). Move over ZFNs. Nat Biotechnol **29,** 681-684.

Deng, D., Yin, P., Yan, C., Pan, X., Gong, X., Qi, S., . . . Shi, Y. (2012). Recognition of methylated DNA by TAL effectors. Cell Res **22,** 1502-1504.

Doyle, E.L., Stoddard, B.L., Voytas, D.F., and Bogdanove, A.J. (2013). TAL effectors: highly adaptable phytobacterial virulence factors and readily engineered DNA-targeting proteins. Trends Cell Biol **23,** 390-398.

Engler, C., Kandzia, R., and Marillonnet, S. (2008). A one pot, one step, precision cloning method with high throughput capability. PLoS One **3,** e3647.

Gao, H., Wu, X., Chai, J., and Han, Z. (2012). Crystal structure of a TALE protein reveals an extended N-terminal DNA binding region. Cell Res **22,** 1716-1720.

Geissler, R., Scholze, H., Hahn, S., Streubel, J., Bonas, U., Behrens, S.E., and Boch, J. (2011). Transcriptional activators of human genes with programmable DNA-specificity. PLoS One **6,** e19509.

Gu, K., Yang, B., Tian, D., Wu, L., Wang, D., Sreekala, C., . . . Yin, Z. (2005). R gene expression induced by a type-III effector triggers disease resistance in rice. Nature **435,** 1122-1125.

Hockemeyer, D., Wang, H., Kiani, S., Lai, C.S., Gao, Q., Cassady, J.P., . . . Jaenisch, R. (2011). Genetic engineering of human pluripotent cells using TALE nucleases. Nat Biotechnol **29,** 731-734.

Huang, P., Xiao, A., Zhou, M., Zhu, Z., Lin, S., and Zhang, B. (2011). Heritable gene targeting in zebrafish using customized TALENs. Nat Biotechnol **29,** 699-700.

Jinek, M., Chylinski, K., Fonfara, I., Hauer, M., Doudna, J.A., and Charpentier, E. (2012). A programmable dual-RNA-guided DNA endonuclease in adaptive bacterial immunity. Science **337,** 816-821.

Kuzma, J., and Kokotovich, A. (2011). Renegotiating GM crop regulation. Targeted gene-modification technology raises new issues for the oversight of genetically modified crops. EMBO Rep **12,** 883-888.

Lei, Y., Guo, X., Liu, Y., Cao, Y., Deng, Y., Chen, X., . . . Zhao, H. (2012). Efficient targeted gene disruption in Xenopus embryos using engineered transcription activator-like effector nucleases (TALENs). Proc Natl Acad Sci U S A **109,** 17484-17489.

Li, L., Wu, L.P., and Chandrasegaran, S. (1992). Functional domains in Fok I restriction endonuclease. Proc Natl Acad Sci U S A **89,** 4275-4279.

Li, T., Huang, S., Zhou, J., and Yang, B. (2013). Designer TAL effectors induce disease susceptibility and resistance to Xanthomonas oryzae pv. oryzae in rice. Mol Plant **6,** 781-789.

Li, T., Liu, B., Spalding, M.H., Weeks, D.P., and Yang, B. (2012). High-efficiency TALEN-based gene editing produces disease-resistant rice. Nat Biotechnol **30,** 390-392.

Li, T., Huang, S., Jiang, W.Z., Wright, D., Spalding, M.H., Weeks, D.P., and Yang, B. (2011a). TAL nucleases (TALNs): hybrid proteins composed of TAL effectors and FokI DNA-cleavage domain. Nucleic Acids Res **39,** 359-372.

Li, T., Huang, S., Zhao, X., Wright, D.A., Carpenter, S., Spalding, M.H., . . . Yang, B. (2011b). Modularly assembled designer TAL effector nucleases for targeted gene knockout and gene replacement in eukaryotes. Nucleic Acids Res **39,** 6315-6325.

Liu, J., Li, C., Yu, Z., Huang, P., Wu, H., Wei, C., . . . Jiao, R. (2012). Efficient and specific modifications of the Drosophila genome by means of an easy TALEN strategy. J Genet Genomics **39,** 209-215.

Ma, S., Zhang, S., Wang, F., Liu, Y., Xu, H., Liu, C., . . . Xia, Q. (2012). Highly efficient and specific genome editing in silkworm using custom TALENs. PLoS One **7,** e45035.

Mahfouz, M.M., Li, L., Shamimuzzaman, M., Wibowo, A., Fang, X., and Zhu, J.K. (2011). De novo-engineered transcription activator-like effector (TALE) hybrid nuclease with novel DNA binding specificity creates double-strand breaks. Proc Natl Acad Sci U S A **108,** 2623-2628.

Mahfouz, M.M., Li, L., Piatek, M., Fang, X., Mansour, H., Bangarusamy, D.K., and Zhu, J.K. (2012). Targeted transcriptional repression using a chimeric TALE-SRDX repressor protein. Plant Molecular Biology **78,** 311-321.

Mak, A.N., Bradley, P., Cernadas, R.A., Bogdanove, A.J., and Stoddard, B.L. (2012). The crystal structure of TAL effector PthXo1 bound to its DNA target. Science **335,** 716-719.

Mali, P., Yang, L., Esvelt, K.M., Aach, J., Guell, M., DiCarlo, J.E., . . . Church, G.M. (2013a). RNA-Guided Human Genome Engineering via Cas9. Science **339,** 823-826.

Mali, P., Aach, J., Stranges, P.B., Esvelt, K.M., Moosburner, M., Kosuri, S., . . . Church, G.M. (2013b). CAS9 transcriptional activators for target specificity screening and paired nickases for cooperative genome engineering. Nat Biotechnol **31,** 833-838.

Miller, J.C., Tan, S., Qiao, G., Barlow, K.A., Wang, J., Xia, D.F., . . . Rebar, E.J. (2011). A TALE nuclease architecture for efficient genome editing. Nat Biotechnol **29,** 143-148.

Moore, F.E., Reyon, D., Sander, J.D., Martinez, S.A., Blackburn, J.S., Khayter, C., . . . Langenau, D.M. (2012). Improved somatic mutagenesis in zebrafish using transcription activator-like effector nucleases (TALENs). PLoS One **7,** e37877.

Morbitzer, R., Romer, P., Boch, J., and Lahaye, T. (2010). Regulation of selected genome loci using de novo-engineered transcription activator-like effector (TALE)-type transcription factors. Proc Natl Acad Sci U S A **107,** 21617-21622.

Morbitzer, R., Elsaesser, J., Hausner, J., and Lahaye, T. (2011). Assembly of custom TALE-type DNA binding domains by modular cloning. Nucleic Acids Res **39,** 5790-5799.

Moscou, M.J., and Bogdanove, A.J. (2009). A simple cipher governs DNA recognition by TAL effectors. Science **326,** 1501.

Murakami, M.T., Sforca, M.L., Neves, J.L., Paiva, J.H., Domingues, M.N., Pereira, A.L., . . . Benedetti, C.E. (2010). The repeat domain of the type III effector protein PthA shows a TPR-like structure and undergoes conformational changes upon DNA interaction. Proteins **78,** 3386-3395.

Osakabe, K., Osakabe, Y., and Toki, S. (2010). Site-directed mutagenesis in Arabidopsis using custom-designed zinc finger nucleases. Proc Natl Acad Sci U S A **107,** 12034-12039.

Paques, F., and Duchateau, P. (2007). Meganucleases and DNA double-strand break-induced recombination: perspectives for gene therapy. Curr Gene Ther **7,** 49-66.

Pennisi, E. (2013). The CRISPR Craze. Science **341,** 833-836.

Plessis, A., Perrin, A., Haber, J.E., and Dujon, B. (1992). Site-specific recombination determined by I-SceI, a mitochondrial group I intron-encoded endonuclease expressed in the yeast nucleus. Genetics **130,** 451-460.

Reyon, D., Khayter, C., Regan, M.R., Joung, J.K., and Sander, J.D. (2012). Engineering designer transcription activator-like effector nucleases (TALENs) by REAL or REAL-Fast assembly. Curr Protoc Mol Biol **Chapter 12,** Unit 12 15.

Rothstein, R.J. (1983). One-step gene disruption in yeast. Methods Enzymol **101,** 202-211.

Rouet, P., Smih, F., and Jasin, M. (1994). Introduction of double-strand breaks into the genome of mouse cells by expression of a rare-cutting endonuclease. Mol Cell Biol **14,** 8096-8106.

Rudin, N., Sugarman, E., and Haber, J.E. (1989). Genetic and physical analysis of double-strand break repair and recombination in Saccharomyces cerevisiae. Genetics **122,** 519-534.

Sander, J.D., Cade, L., Khayter, C., Reyon, D., Peterson, R.T., Joung, J.K., and Yeh, J.R. (2011). Targeted gene disruption in somatic zebrafish cells using engineered TALENs. Nat Biotechnol **29,** 697-698.

Schmid-Burgk, J.L., Schmidt, T., Kaiser, V., Honing, K., and Hornung, V. (2013). A ligation-independent cloning technique for high-throughput assembly of transcription activator-like effector genes. Nat Biotechnol **31,** 76-81.

Shan, Q., Wang, Y., Chen, K., Liang, Z., Li, J., Zhang, Y., . . . Gao, C. (2013). Rapid and efficient gene modification in rice and brachypodium using TALENs. Mol Plant **6,** 1365-1368.

Shukla, V.K., Doyon, Y., Miller, J.C., DeKelver, R.C., Moehle, E.A., Worden, S.E., . . . Urnov, F.D. (2009). Precise genome modification in the crop species Zea mays using zinc-finger nucleases. Nature **459,** 437-441.

Smith, J., Bibikova, M., Whitby, F.G., Reddy, A.R., Chandrasegaran, S., and Carroll, D. (2000). Requirements for double-strand cleavage by chimeric restriction enzymes with zinc finger DNA-recognition domains. Nucleic Acids Res **28,** 3361-3369.

Sugio, A., Yang, B., Zhu, T., and White, F.F. (2007). Two type III effector genes of Xanthomonas oryzae pv. oryzae control the induction of the host genes OsTFIIAgamma1 and OsTFX1 during bacterial blight of rice. Proc Natl Acad Sci U S A **104,** 10720-10725.

Sung, Y.H., Baek, I.J., Kim, D.H., Jeon, J., Lee, J., Lee, K., . . . Lee, H.W. (2013). Knockout mice created by TALEN-mediated gene targeting. Nat Biotechnol **31,** 23-24.

Tesson, L., Usal, C., Menoret, S., Leung, E., Niles, B.J., Remy, S., . . . Cost, G.J. (2011). Knockout rats generated by embryo microinjection of TALENs. Nat Biotechnol **29,** 695-696.

Tong, C., Huang, G., Ashton, C., Wu, H., Yan, H., and Ying, Q.L. (2012). Rapid and cost-effective gene targeting in rat embryonic stem cells by TALENs. J Genet Genomics **39,** 275-280.

Townsend, J.A., Wright, D.A., Winfrey, R.J., Fu, F., Maeder, M.L., Joung, J.K., and Voytas, D.F. (2009). High-frequency modification of plant genes using engineered zinc-finger nucleases. Nature **459,** 442-445.

Verdier, V., Triplett, L.R., Hummel, A.W., Corral, R., Cernadas, R.A., Schmidt, C.L., . . . Leach, J.E. (2012). Transcription activator-like (TAL) effectors targeting OsSWEET genes enhance virulence on diverse rice (Oryza sativa) varieties when expressed

individually in a TAL effector-deficient strain of Xanthomonas oryzae. New Phytologist **196,** 1197-1207.

Wang, Z., Li, J., Huang, H., Wang, G., Jiang, M., Yin, S., . . . Xi, J.J. (2012). An integrated chip for the high-throughput synthesis of transcription activator-like effectors. Angew Chem Int Ed Engl **51,** 8505-8508.

Weber, E., Gruetzner, R., Werner, S., Engler, C., and Marillonnet, S. (2011). Assembly of designer TAL effectors by Golden Gate cloning. PLoS One **6,** e19722.

Wood, A.J., Lo, T.W., Zeitler, B., Pickle, C.S., Ralston, E.J., Lee, A.H., . . . Meyer, B.J. (2011). Targeted genome editing across species using ZFNs and TALENs. Science **333,** 307.

Yang, B., Sugio, A., and White, F.F. (2006). Os8N3 is a host disease-susceptibility gene for bacterial blight of rice. Proc Natl Acad Sci U S A **103,** 10503-10508.

Yu, Y., Streubel, J., Balzergue, S., Champion, A., Boch, J., Koebnik, R., . . . Szurek, B. (2011). Colonization of rice leaf blades by an African strain of Xanthomonas oryzae pv. oryzae depends on a new TAL effector that induces the rice nodulin-3 Os11N3 gene. Mol Plant Microbe Interact **24,** 1102-1113.

Zhang, F., Cong, L., Lodato, S., Kosuri, S., Church, G.M., and Arlotta, P. (2011). Efficient construction of sequence-specific TAL effectors for modulating mammalian transcription. Nat Biotechnol **29,** 149-153.

Zhang, F., Maeder, M.L., Unger-Wallace, E., Hoshaw, J.P., Reyon, D., Christian, M., . . . Voytas, D.F. (2010). High frequency targeted mutagenesis in Arabidopsis thaliana using zinc finger nucleases. Proc Natl Acad Sci U S A **107,** 12028-12033.

Zhang, Y., Zhang, F., Li, X., Baller, J.A., Qi, Y., Starker, C.G., . . . Voytas, D.F. (2013). Transcription activator-like effector nucleases enable efficient plant genome engineering. Plant Physiol **161,** 20-27.

Zhu, H., Lau, C.H., Goh, S.L., Liang, Q., Chen, C., Du, S., . . . Wang, S. (2013). Baculoviral transduction facilitates TALEN-mediated targeted transgene integration and Cre/LoxP cassette exchange in human-induced pluripotent stem cells. Nucleic Acids Res.

Zu, Y., Tong, X., Wang, Z., Liu, D., Pan, R., Li, Z., . . . Lin, S. (2013). TALEN-mediated precise genome modification by homologous recombination in zebrafish. Nat Methods **10,** 329-331.

CHAPTER 2 - Activation of the Carbon Concentrating Mechanism by CO_2 Deprivation Coincides with Massive Transcriptional Restructuring in *Chlamydomonas reinhardtii*

Plant Cell (2013). ***24(5)****, 1860-75*

Andrew J. Brueggeman[a], Dayananda S. Gangadharaiah[a], Matyas F. Cserhati[b], David Casero[c,d], Donald P. Weeks[a], and Istvan Ladunga[b]

[a]Department of Biochemistry, University of Nebraska, Lincoln, Nebraska 68588-0665
[b]Department of Statistics, University of Nebraska, Lincoln, Nebraska 68588-0665
[c]Department of Molecular, Cell, and Developmental Biology, University of California, Los Angeles, California 90095
[d]Institute of Genomics and Proteomics, University of California, Los Angeles, California 90095

Abstract

A CO_2-concentrating mechanism (CCM) is essential for the growth of most eukaryotic algae under ambient (392 ppm) and very low (<100 ppm) CO_2 concentrations. In this study, we used replicated deep mRNA sequencing and regulatory network reconstruction to capture a remarkable scope of changes in gene expression that occurs when *Chlamydomonas reinhardtii* cells are shifted from high to very low levels of CO_2 (≤100 ppm). CCM induction 30 to 180 min post-CO_2 deprivation coincides with statistically significant changes in the expression of an astonishing 38% (5884) of the 15,501 nonoverlapping *C. reinhardtii* genes. Of these genes, 1088 genes were induced and 3828 genes were downregulated by a log2 factor of 2. The latter indicate a global reduction in photosynthesis, protein synthesis, and energy-related biochemical pathways. The magnitude of transcriptional rearrangement and its major patterns are robust as analyzed by three different statistical methods. De novo DNA motif discovery revealed new putative binding sites for *Myeloid oncogene* family transcription factors potentially involved in activating low CO_2–induced genes. The $(CA)_n$ repeat ($9 \leq n \leq 25$) is present in 29% of upregulated genes but almost absent from promoters of downregulated genes. These discoveries open many avenues for new research.

Introduction

In nature, *Chlamydomonas reinhardtii* and other eukaryotic algae depend on a CO_2-concentrating mechanism (CCM) to supply sufficient inorganic carbon (C_i; CO_2 or bicarbonate) for photosynthesis-fueled cell growth and proliferation. Mutant cells lacking key components of the CCM molecular machinery or its regulatory system (Moroney and Ynalvez, 2007; Duanmu et al., 2009a; Yamano and Fukuzawa, 2009) do not grow or grow poorly unless supplied with high concentrations of CO_2 (e.g., >10,000 ppm) that are well above the ambient level of ~392 ppm. Because the diffusion rate of CO_2 in aqueous environments is ~10,000 times slower than in air, most natural populations of microalgae exist in CO_2-limited conditions. This is especially true for dense algal populations growing under abundant sunlight. Under such conditions, CO_2 concentrations can become very low (<100 ppm) and cells induce the CCM to maximal levels. CO_2 starvation induces the transcription of numerous genes encoding proteins closely associated with the CCM and its activities (Moroney and Ynalvez, 2007; Duanmu et al., 2009a; Yamano and Fukuzawa, 2009). Indeed, *C. reinhardtii* and most other eukaryotic algae have developed a finely tuned regulatory system that suppresses expression of CCM-related genes under conditions of replete CO_2 (i.e., >0.1% CO_2) and activates expression of these genes when CO_2 becomes limiting (Moroney and Ynalvez, 2007; Duanmu et al., 2009a; Yamano and Fukuzawa, 2009; Fang et al., 2012). Previous studies using RNA gel blot analyses and microarray analyses have revealed a number of CCM-associated genes and other CO_2-responsive genes whose transcription is tied to

the physiological changes that accompany cell acclimation to CO_2 stress conditions (Im and Grossman, 2002; Im et al., 2003; Miura et al., 2004; Wang et al., 2005; Yamano et al., 2008; Yamano and Fukuzawa, 2009).

Here, we report an extensive global analysis of the massive transcriptional changes evoked by the deprivation of C_i in *C. reinhardtii*. We measured these transcriptional events using replicated deep RNA sequencing (RNA-Seq) on the Illumina platform. The highly reproducible RNA-Seq experiments not only confirm earlier observations based on array analyses quoted above but also extend the list of differentially expressed genes from a few hundred to over 4000.

We report the discovery of an extensive system of head-to-head (HTH; also called divergent) gene pairs, many of them sharing bidirectional or connected promoters. HTH conformation and bidirectional or shared promoters frequently perform the highly accurate coregulation of gene pairs encoding subunits of the same protein complex or two proteins of similar or related functions. Here, we focus on those HTH, coregulated, gene pairs that are most relevant to the CCM. Advanced computational techniques also have allowed an extensive evaluation of potential regulatory elements in promoter regions in CO_2- responsive genes and the discovery of new elements shared by several of the most highly stimulated CO_2-responsive genes. We also report a previously unrecognized pattern of expression for many genes that suggests a significant, but transient, decrease in gene transcription immediately after a shift to very low CO_2 conditions (ASVLCO_2). Finally, we employ a vastly expanded pool of

transcriptomic data to strengthen earlier observations of metabolic and physiological changes that occur when CO_2 becomes limiting in the environment, including significant decreases in transcripts encoding proteins involved in photosynthesis, cytoplasmic, chloroplastic, and mitochondrial protein synthesis, energy use, protein transport, and other Gene Ontology (GO) categories (Ashburner et al., 2000).

Results and Discussion

Overview

C_i deprivation is a major stress that evokes a dramatic transcriptional response in algae. Using EST-based macroarrays, the Fukuzawa laboratory (Miura et al., 2004; Kohinata et al., 2008; Yamano et al., 2008) and Grossman and Weeks laboratories (Wang et al., 2005) pioneered the transcriptional profiling of C_i deprivation. In initiating our studies, our hypothesis was that revolutionary progress in sequencing technology and statistical methodology would allow us to discover a large number of activated or repressed biological processes and individual genes that may have escaped detection using EST arrays. To test this hypothesis, we performed deep RNA-Seq using the Illumina Genome Analyzer II platform at the Joint Genome Institute (JGI) of the Department of Energy. In total, the eight samples collected at four time points (0, 30, 60, and 180 min after C_i deprivation) produced 98.3 million uniquely mapped sequencing reads (12.3 million 71-base-long reads per sample; see Methods). When no more than two mismatches were allowed in the anchor regions, ~38% of the reads did not map uniquely or contained more than two base errors due to sequencing errors, genomic variability, alternative

splicing (Labadorf et al., 2010), a number of recently duplicated genes (Villand et al., 1997), and repetitive DNA elements (Merchant et al., 2007). Even with this conservative approach, RNA-Seq represents a major advance from micro- and macroarrays: It provides an unprecedentedly high coverage of transcripts, eliminates cross-hybridization effects, does not rely on the commercial availability of arrays, and is more robust against errors in predicted exon structures (Margulies et al., 2005). The significantly increased performance of RNA-Seq has been shown specifically for *C. reinhardtii* (Gonzalez-Ballester et al., 2010).

The high technological reproducibility of the RNA-Seq measurements performed at the Department of Energy's JGI is shown by the strong correlations of transcript levels between biological replicates (0.958, 0.965, 0.939, and 0.973 for 0, 30, 60, and 180 min time points after carbon deprivation, respectively). These high Pearson correlation coefficients indicate reproducible and multiplicative (linear) biases and that the nonlinear bias is miniscule. Note that linear, multiplicative, and reproducible bias does not alter fold change values by multiplying the transcript levels both in the numerator and the denominator. Such biases include sequencing reads that match imperfectly to the genome or the transcriptome (Li et al., 2010), amplification, and sequencing biases (Dohm et al., 2008). Additive effects, such as unreal exons, may reduce the extent of differential expression. These effects are due to imperfect gene models, such as those predicted by the Augustus method (Stanke and Waack, 2003) and alternative splicing (Labadorf et al., 2010). Such additive effects remain our primary concern. Recently

duplicated genes pose further challenges in mapping the 71-base-long sequencing reads to the transcriptome because these reads contain erroneous base calls, particularly at their 3' ends. Such gene pairs include major effectors of CO_2 concentration, such as four carbonic anhydrases (*CAHs*), *CAH1*-*CAH2* and *CAH4*-*CAH5*, that are recent duplicates. The pair *CAH4*-*CAH5*, for example, contains exons that are over 90% identical (Villand et al., 1997).

To avoid mappings to the duplicated genes, rigorous procedures (with one or two mismatches in the anchor regions that connect two exons) are necessary. However, this rigor also drastically reduces the coverage of all genes due to both sequencing errors and polymorphisms. Reduced coverage reduces the number of significantly differentially expressed genes. Therefore, we performed the mapping with both one and two allowed mismatches in the anchor region, as implemented in the tophat program (Trapnell et al., 2009). With one allowed mismatch, fewer but more accurate transcript levels were obtained than with two mismatches. For example, our data, as expected from earlier studies, demonstrated that *CAH1* is strongly upregulated at 3 h ASVLCO_2. However, in our initial analyses allowing two mismatches, *CAH2*, which had earlier been reported not to respond to CO_2, was falsely classified as upregulated. When reanalyzed using only one mismatch per read, the vast majority of reads in the 60- and 180-min time points were mapped to the *CAH1* gene, with few being attributable to *CAH2*. Quantitative RT-PCR (qRT-PCR) confirmed the results of the more rigorous alignments (see Supplemental Figure 1 online).

The lists of differentially expressed genes may be influenced by the choice of statistical methodology. Therefore, we analyzed our data using three different computational tools, *edgeR*, *DESeq*, and *baySeq*. Because differential expression of a large number (~16k) of genes is estimated using very few replicates (samples), statistical tools derived from large-sample asymptotic theory do not work. In particular, small sample size affects the correction for overdispersion (greater variability than expected based on Poissonian or other simple models), modeling the empirical distributions, and calculating statistical significance. To solve these issues, *edgeR* (Robinson et al., 2010) shrinks genewise dispersion estimates toward a constant value using an empirical Bayesian model and performs Fisher's exact test. *DESeq* (Anders and Huber, 2010) uses nonparametric regression models to fit the negative binomial variance as a function of the mean, assuming a locally linear relationship between overdispersion and mean expression levels. *baySeq* (Hardcastle and Kelly, 2010) is free of this assumption and uses a fully empirical Bayesian approach to estimate the posterior probabilities. We compared the numbers of overlapping differentially expressed genes reported by the *edgeR*, *DESeq*, and *baySeq* packages at 180 versus 0 min ASVLCO$_2$ (see Supplemental Figure 2 online). Because the exact test implemented in *edgeR* calculates lower false discovery rate (FDR) q-values than the other two methods (Lopez et al., 2011), at FDR ≤ 0.01, *edgeR*, *DESeq*, and *baySeq* reported 4222, 2364, and 3248 differentially expressed genes, respectively. The lists of differentially expressed genes are more consistent when the FDR threshold is elevated to 0.05, a still conservative level. All three methods

reported differential expression for as many as 3141 genes (see Supplemental Figure 2 online). An additional 702 genes were jointly reported by both *edgeR* and *baySeq*, and a further 95 genes were called jointly by *edgeR* and *DESeq*. Because of the high overlaps with other methods, and its wider acceptance, below we limit our discussions to the results obtained by the *edgeR* tool.

Our results reproduced the observed induction of major CCM-associated genes published by Miura et al. (2004), Wang et al. (2005), Yamano et al. (2008), Yamano and Fukuzawa (2009), as well as in the companion article (Fang et al., 2012). In addition, we report a large number of genes that have not been associated with the CCM previously. Some of the notable similarities and differences in gene sets of our studies and those presented in our companion paper (Fang et al., 2012) are discussed throughout this section (with special attention to the potential causes of observed differences provided near the end of this section).

Major Transcriptional Changes

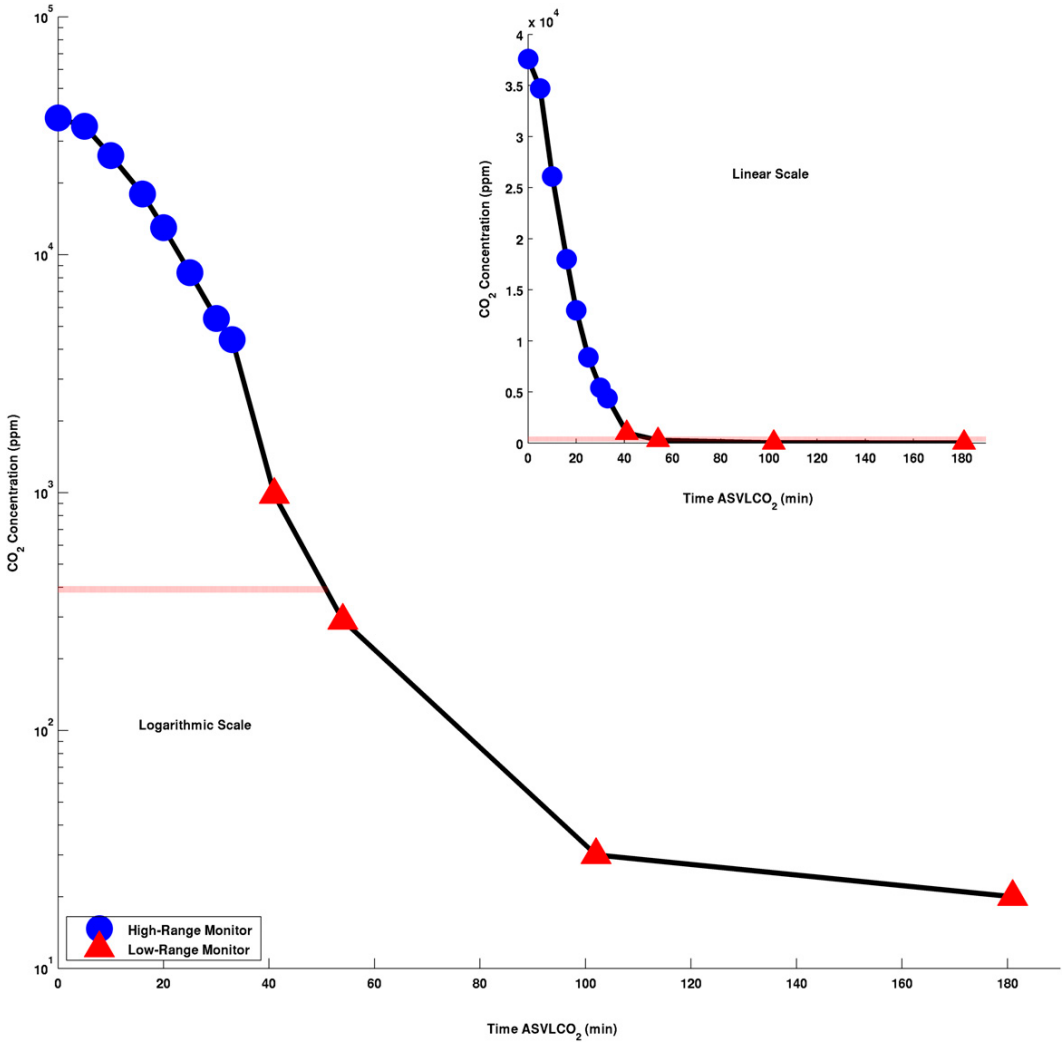
Our results greatly extend many aspects of earlier observations of differentially expressed genes following CO_2 deprivation. This is indicated by relatively similar lists of induced genes published previously (Miura et al., 2004; Wang et al., 2005; Yamano et al., 2008; Yamano and Fukuzawa, 2009) and by us (see Supplemental Data Set 1 online). In addition, RNA-Seq and modern statistical methodologies allowed us to discover an unexpectedly high 5884 genes that are differentially regulated at either 30, 60, or 180

min ASVLCO$_2$ relative to the 0 min control [FDR ≤ 0.01 and abs(log$_2$ (fold change)) ≥ 1] or 3828 genes [FDR ≤ 0.001 and abs(log$_2$(fold change)) ≥ 2].

Robust temporal expression patterns emerged under our conditions for imposition of CO_2 deprivation. We found that the transcriptional response becomes widespread only after 30 min and increases (or decreases) for many, but not all, genes. The relatively slow onset of significant transcript changes is likely coupled to the relatively slow decline in CO_2 concentrations employed in our experiments (Figure 1). At 30 min after deprivation, we found only 37 upregulated and five repressed genes relative to the 0 min control (FDR ≤ 0.001 and abs[log2(fold change) ≥ 2; or in absolute, nonlogarithmic scale, a fourfold increase or decrease) (see Supplemental Data Sets 2 and 3 online). At an hour ASVLCO$_2$, 409 genes are upregulated and 1663 genes are repressed (see Supplemental Data Sets 2 and 3 online). At 3 h ASVLCO2, 981 genes are induced and 1188 genes are repressed. These numbers are approximately doubled at the more typical thresholds (FDR q ≤ 0.01 and abs[log$_2$(fold change) ≥ 1; see Supplemental Data Sets 2 and 3 online). To measure transcript levels by a different method, we performed qRT-PCR analyses on the same RNA samples that were submitted for Illumina sequencing. Three different genes induced by CO_2 deprivation (Low CO_2 Induced A (*LCIA*), *CAH5*, and *LCIB*) displayed similar expression patterns between RNA-Seq and qRT-PCR, while a fourth gene, *CAH2*, previously reported as not responding or responding negatively to CO_2 depletion (Moroney et al., 2011), showed moderate decreases in transcript levels using both RNA-Seq and RT-PCR measurements

(see Supplemental Figure 1 online). Also to be noted (as described above) is the strong correlation of transcript levels in data obtained from biological replicates used for RNA-Seq analyses. Together, these observations confirm the high technological reproducibility of deep RNA-Seq as well as the reproducibility of our biological samples.

Figure 2.1 - Measurements of CO2 Levels Following a Shift of C. reinhardtii Cells from 5% to 100ppm CO2. Two CO_2 monitors were used in the fermenter: One was calibrated for high CO_2 concentrations (circles), and the other was calibrated to low CO_2 concentrations (triangles). CO_2 concentrations are plotted on a $\log_{10}$ scale. The horizontal red line represents the 392 ppm concentration of the atmosphere. The reduction in CO_2 is represented both in a linear (inset) and logarithmic scale.



Gene Sets Affected by CO_2 Deprivation

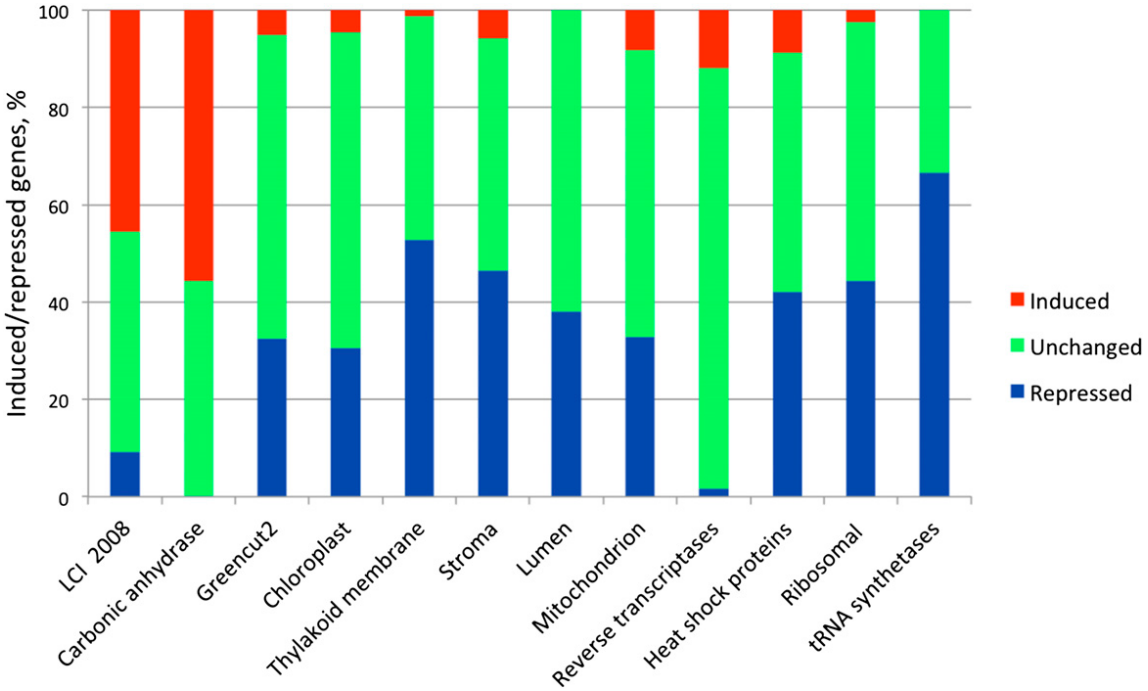
For brevity, unless otherwise noted, we limit our discussion in this section to transcript differences in cells maintained in high CO_2 to transcripts in cells 3 h ASVLCO_2. High-level overviews of the massive transcriptional response were obtained using GO categories (Ashburner et al., 2000). PLAZA, an online platform for plant comparative genomics (Proost et al., 2009), assigns GO categories to ~7700 *C. reinhardtii* genes. GO categories, like pathways or genes colocalized within a chromosomal band, allow examination of transcriptional changes at the level of gene sets as opposed to individual genes. These gene sets may be enriched in upregulated or downregulated genes. To avoid the subjectivity of interpretations, the statistical significance of enrichment is calculated using gene set enrichment analyses (GSEAs; Subramanian et al. (2005); see Methods). GSEA is capable of handling high-throughput processing of large databases of gene sets, such as GO or the KEGG Database of Biochemical Pathways (Okuda et al., 2008).

The massive downregulation of several fundamental metabolic processes is highly statistically significant as shown by our GSEAs using GO annotations. We found significant repression of translation, ribosomal activities, RNA processing, intracellular protein transport, transport from the endoplasmic reticulum to the Golgi apparatus, nucleic acid binding, ATP synthesis, protein kinase activity, photosynthesis, oxidation reduction processes, protein folding, and unfolded protein binding, etc. (see Supplemental Data Set 4 online). Decreased metabolic activity may reflect the

stress/survival mode of metabolism, which is necessary to cope with the stress of CO_2 starvation. Upregulated GO categories are less abundant than downregulated ones, biased due to the limited annotations of the CCM- and other plant-specific biological processes in contrast with the basic metabolic processes present in most eukaryotes. Induced GO categories include nucleosome, carbon use, calcium ion transport, and proteolysis-related gene sets (see Supplemental Data Set 4 online). Unexpectedly, five flagella-related categories also are upregulated. GO analyses of several other biochemical pathways provided little additional information, possibly due to limited annotations.

To complement GO categories, we also studied other major functional and subcellular categories (Figure 2.2). One of these categories is a set of 595 plant-specific (greencut2) genes, which are conserved through diverse representatives of the plant kingdom but have no known relatives in animals, fungi, or prokaryotes (Karpowicz et al., 2011). As many as 193 greencut2 genes were repressed, while only 30 were activated (FDR q-value < 10^{-256} as calculated by the permutation test implemented in GSEA; Figure 2.2). To examine whether carbon deprivation reduces photosynthetic activity, we analyzed the 393 genes that code for proteins localized in the chloroplast. Of these nuclear or chloroplast genes, 120 were downregulated and only 18 were induced (FDR q < 10^{-256}).

Figure 2.2 - Differential Gene Expression in Major Functional and Subcellular Categories in C. reinhardtii Cells before or after a Shift from High to Low CO2. The percentage of induced genes [red bars; $\log_2$(fold change) ≥ 2], repressed genes [blue bars; $\log_2$(fold change) ≤ -2] statistically significant at the FDR $q \leq 0.01$ level, or unchanged [green bars; $\log_2$(fold change) between +2 and -2 and/or not significant]. LCI refers to low CO_2-induced genes according to Yamano et al. (2008); *greencut2* refers to conserved plant and diatom genes that have no close relatives in other kingdoms and in prokaryotes other than cyanobacteria (Karpowicz et al., 2011).



Using DNA arrays containing oligonucleotides derived from ESTs, the Fukuzawa laboratory (Miura et al., 2004; Yamano et al., 2008; Yamano and Fukuzawa, 2009) and the Grossman and Weeks laboratories (Wang et al., 2005) published several lists of genes differentially regulated by C_i deprivation. In our much more extensive analyses, we reproduced the induction of 40 previously reported genes (38%; see Supplemental Data Set 1 online). *LCI1* (*Formate/nitrite transporter1.2* [*NAR1.2*]), a putative plasma membrane anion transporter (Ohnishi et al., 2010), is among the most dramatically induced genes (fold change: $2^{12} \approx 4000$). Fourteen previously designated low CO_2-induced genes decreased their expression, including chloroplast geranylgeranyl pyrophosphate synthase *(geranylgeranyl diphosphate synthase* (*GGPS*) or *LCI14*), *LCI21*, *LCI25*, and *Light Harvesting Complex1* (*LHCSR1*) (at FDR ≤ 0.01 and log*2*(fold change) ≤ -1).

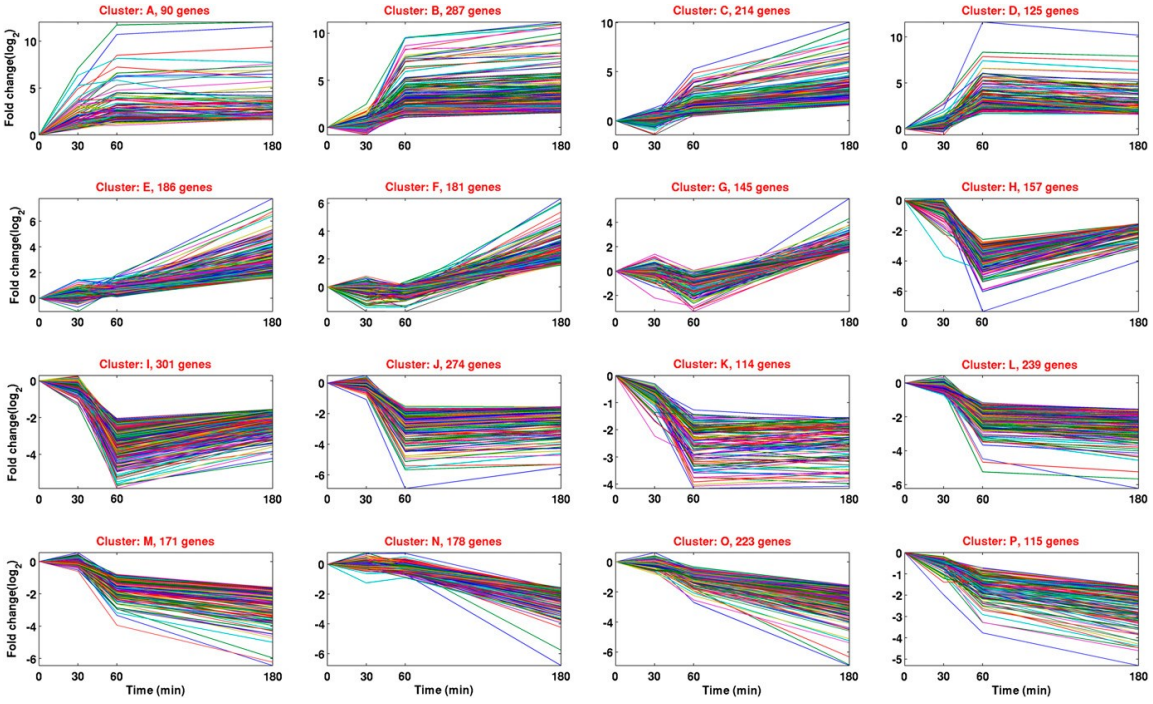
Yamano and Fukuzawa (2009) hypothesized that thylakoid membrane proteins may be involved in the regulation of CCM. We successfully reproduced the earlier reported expression patterns of genes coding for low CO_2–inducible chloroplast envelope proteins, including oxygen-evolving proteins, plastid division protein, a number of carbonic anhydrases, photosystem II stability/assembly factor *photosystem II stability/assembly factor* (*HCF136*), and peptidyl-prolyl *cis-trans* isomerase (see Supplemental Data Sets 2 and 3 online).

A Shift of *C. reinhardtii* Cells from High to Very Low CO_2 Triggers Transcription of Several Known Genes Encoding CCM-Associated Proteins and Several New CO_2-Reponsive Genes

Examination of transcript abundance in this RNA-Seq study for the 106 genes identified in earlier analyses of induced gene transcription triggered by shifts from high to low CO_2 conditions (see Supplemental Data Set 1 online) confirms that most are highly induced within an hour ASVLCO_2. Among the most highly induced genes at 3 h are several well-recognized CCM-associated or CO_2- responsive genes, including *LCIA*, encoding a putative bicarbonate transporter (4000-fold); *LCI1*, a gene regulated by the CO_2- responsive LCR1 transcription factor (see below) and encoding a transmembrane protein possibly involved in C_i transport (3000-fold); *CCP1*, a C_i Accumulating5 (CIA5)/CCM1-regulated gene encoding a chloroplast envelope protein (2000-fold); *CAH1*, encoding a periplasmic carbonic anhydrase (660-fold); *LCIE* (188-fold); *CCP2*, encoding a protein closely related to CCP1 (119-fold); *LCID*, encoding a protein highly similar to LCIB (see below) (89-fold); *LCR1*, encoding a *Myeloid oncogene* (*Myb*)- like transcription factor that regulates expression of *CAH1* and *LCI1* (52-fold); *High Light Acclimation3* (*HLA3*), encoding a plasma membrane–localized bicarbonate transporter (40-fold); and *LCIB*, encoding a putative chloroplast CO_2-scavenging protein (26-fold) (see Supplemental Data Sets 1 and 2 online). Using cluster analysis of gene expression, we found that expression patterns of these genes during the 3 h time course employed in this study fall primarily into the patterns displayed in clusters A to D (Figure 2.3). In

these clusters, mRNA levels markedly increase by an hour ASVLCO$_2$ and remain high at 3 h.

Figure 2.3 - Cluster Analysis of Gene Expression Patterns in C. reinhardtii Cells Subjected to a Change from High to Low CO2 Concentrations. Sixteen clusters were identified by the *k*-means algorithm. Shown are only the genes that are differentially expressed by a factor of four or more between at least two time points [FDR $q \leq 0.01$ and $abs(\log_2[\text{fold change}]) \geq 2$]. Time comparisons on the *x* axis: 0 represents 0-min time point expression levels versus 0 time point expression levels, 30 represents 30-min time point versus 0 time point, 60 represents 60-min time point versus 0 min time point, and 180 represents 180-min time point versus 0 time point. Fold changes in gene expression levels are represented on the *y* axis in $\log_2$ terms.



Strikingly, our deep RNA-Seq analyses increased the number of low CO_2-induced (LCI) genes to 1875 (see Supplemental Data Set 2 online), well beyond the 106 previously recognized LCI genes listed in Supplemental Data Set 1 online. The annotation of *C. reinhardtii* genes has recently been improved (Castruita et al., 2011; Lopez et al., 2011). Still, the function of a number of proteins encoded by differentially regulated genes remains unknown, including 56 out of the 100 most highly induced genes (see Supplemental Data Set 2 online).

A General, but Transient, Decrease in Transcription of Most Genes after a Shift to Very Low CO_2 Conditions

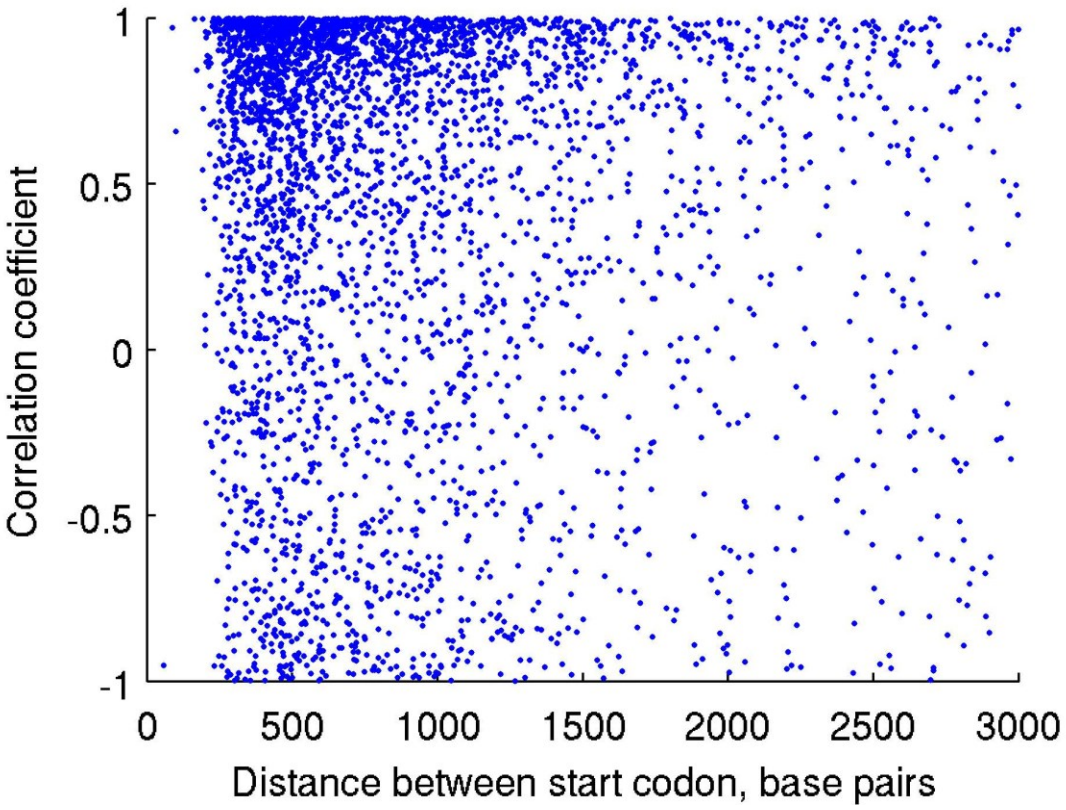
Past studies of genes whose transcription changes in response to decreased CO_2 levels have focused primarily on those genes whose products are components of the CCM and whose transcription is steadily increased in the first few hours ASVLCO_2. Expression patterns typical of such genes are shown in clusters A to D of Figure 2.3. However, 2399 genes are first downregulated but then induced by 3 h ASVLCO_2, forming a distinctive check mark (√) signature of transcript levels. Accordingly, the groups of genes in boxes F to I all display a significant decrease in transcript numbers at 30 and/or 60 min with a subsequent increase at 180 min. Indeed, if one examines the data of Supplemental Data Set 3 online that displays ~2400 genes downregulated at 180 min ASVLCO_2, nearly all of the last 1740 genes display a striking decrease in transcript levels at 60 min, but a marked recovery of transcript levels at 180 min (i.e., the check mark expression pattern; see Supplemental Figure 3 online). This observation suggests that, in

general, there is a rapid decrease in gene transcription immediately ASVLCO$_2$. Thus, our studies have revealed that *C. reinhardtii* cells not only rapidly sense a decrease in CO_2 concentrations and respond, as expected, by upregulating CCM-associated genes, but also transition into stress/survival mode by downregulating thousands of genes not directly related to the CCM. This observation opens a new area of study related to transcriptional regulatory networks that orchestrate the rapid and appropriate responses of *C. reinhardtii* cells to changes in CO_2 abundance.

Transient Gene Induction following CO_2 Reduction

Antithetic to the check mark pattern of expression discussed above are patterns for 139 genes that are rapidly induced (by a factor of 2 or higher; FDR $q \leq 0.01$) during the first hour ASVLCO$_2$ but markedly decrease in transcript levels 2 h later. This caret (∧) expression pattern (see Supplemental Figure 4 online) is observed for the low CO_2–induced bestrophin-like protein (516308) regulated by the transcriptional activator CIA5 and, interestingly, for 20 flagellar genes as well. Further experiments will determine if among these genes displaying this transient induction pattern are those encoding transient regulators that control acclimations to long-term restricted CO_2 availability. To assess the functions of such transiently expressed genes, mutations or artificial alterations of gene expression (e.g., using RNA interference techniques or gene replacement techniques) will be needed.

Figure 2.4 - Several HTH Genes Are Coregulated by Bidirectional or Interacting Promoters. Each dot represents a HTH gene pair where the distance between the start codons of the two genes is represented by the horizontal coordinate and the coexpression of the two genes is represented by the vertical coordinate of the dot. Coexpression is measured by the Pearson correlation coefficients between the two HTH genes for the transcript levels at 0, 30, 60, and 180 min after carbon deprivation.



Coregulation of Functionally Related Gene Pairs in a HTH Conformation

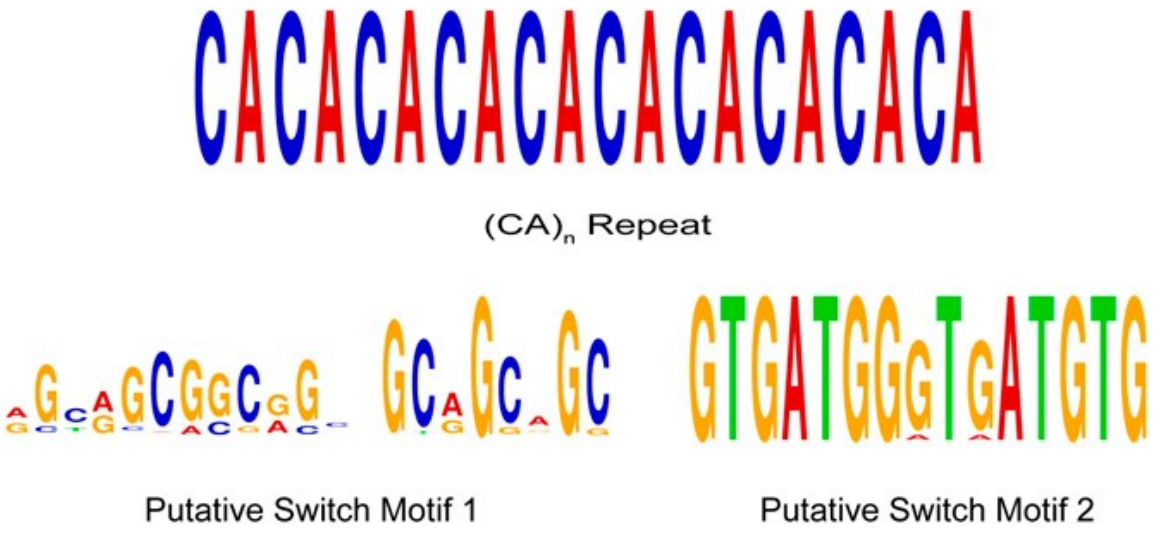
We discovered an extensive system of HTH (also called divergent) gene pairs, many of which are apparently regulated by bidirectional promoters. In general, when the distance between two start codons is small (in *C. reinhardtii*, <300 to 500 bp), coregulation is most likely due to a single bidirectional promoter that regulates transcription in both directions. Bidirectional promoters often perform usually tight coregulation of the upstream Crick strand and the downstream Watson strand genes (Trinklein et al., 2004). Bidirectional promoters facilitate the stoichiometric production of subunits of a particular protein complex or members of a particular biochemical or signaling pathway. Such mechanisms have been described in the human pilot-ENCODE project (Birney et al., 2007), yeast (Li et al., 2011b), mouse (Li et al., 2011a), flies, and other organisms. In *C. reinhardtii*, however, publications are scant regarding the 8852 genes that evolved into the HTH conformation. The few reported coregulated HTH gene pairs include the *Argonaut1* and *Dicer-like1* genes (Casas-Mollano et al., 2008) and two mitochondrial, β-type carbonic anhydrases CAH4 and CAH5 (then termed *ca1* and *ca2*) reported by Villand et al. (1997). *CAH4* and *CAH5*, like *CAH1* and *CAH2*, recently formed inverted repeats, share very high sequence identity even in introns and promoter regions, and are upregulated under CO_2 starvation.

Strong transcriptional correlations within HTH gene pairs may indicate fundamental regulatory mechanisms in *C. reinhardtii*. Here, we focus on CCM-related HTH gene pairs. Coregulation, frequently by a postulated bidirectional promoter, in

many of the 4276 HTH gene pairs in *C. reinhardtii* is indicated by two observations. First, the median correlation coefficient between transcript levels of the HTH gene pairs is as high as 0.674 compared with 0.522 in all other gene pairs ($P < 10^{-256}$, Wilcoxon-Mann-Whitney test; Figure 2.4; see Supplemental Figure 5 online), and for 25% of HTH gene pairs, this correlation exceeds 0.917. For other conformations, the correlation in the top 25% of non-HTH gene pairs is 0.8650 (for the whole distributions, $P < 10^{-256}$, Wilcoxon-Mann-Whitney test; see Supplemental Figure 5 and Supplemental Data Set 5 online). The high correlation of the top 25% even in the non-HTH genes indicates the existence of other mechanisms of coregulation, including locus control regions, microRNAs, and the binding of similar transcription factors (reviewed in Ladunga (2010)). The second indication for the bidirectional regulation is the short distance between the start codons of HTH gene pairs (median length: 831 ± 3228 bp SD). By contrast, intergenic regions of tail-to-tail, head-to-tail, and tail-to-head gene pairs tend to be more than twice as long (median: 1712 ± 5595 bp, $P < 10^{-256}$, Wilcoxon-Mann-Whitney test; see Supplemental Figure 5 online). Short intergenic regions of tail-to-tail, head-to-tail, and tail-to-head gene pairs do not necessarily indicate coregulation.

The correlation coefficient for the transcript levels exceeds 0.9 in 1188 pairs and 0.8 in 1711 pairs (Figure 2.4; see Supplemental Figure 5 and Supplemental Data Set 5 online). Such correlation exists for both induced and repressed gene pairs.

Figure 2.5 - Putative Transcription Factor Binding Sites and Switch Motifs. Putative switch motifs were found in intergenic regions between HTH gene pairs with negatively correlated transcription. Motifs are shown as sequence logos where the horizontal axis indicates sequence position and the vertical axis shows information content, related to conservation.



Several CCM and Related Pathway Genes Exist in HTH Conformation

Among the 1845 significantly induced nonoverlapping genes [FDR $q \leq 0.01$, $\log_2(\text{fold}) \geq 1$], 212 genes evolved into HTH conformations, in which both genes are upregulated, and an additional 1116 HTH pairs, in which one gene is upregulated. The former include 52 genes previously recognized as regulated by CIA5/CCM1 and associated with CCM (see Supplemental Data Set 1 online). In three pairs, *CCP1*/*LCIE*, *CCP2*/*LCID*, and *CAH4*/*CAH5*, each gene has been reported as CCM related. The sequence similarity, tight chromosomal linkage, and highly coordinated expression of *LCID*/*CCP2* and *LCIE*/*CCP1* (see Supplemental Figure 6 online) suggest recent gene pair duplications. Another nine known CCM-related genes form pairs with genes that were not reported as CCM related before: *LCI1*/*Flagellar Associated Protein292 (FAP292)*, *LCI6/*513965, *CAH1/*CLR18, *LHCSR1/*525344, *LCI7/*525344, *LCI7/*523113, *Early Light Inducible4 (ELI4)/*519915, *LCI31/*517052, and *GGPS/*512500. Transcript levels in seven of these nine pairs are correlated with $r > 0.6$.

By contrast, gene expression is negatively correlated in 1048 HTH gene pairs, which presents an enigmatic mechanism of transcription regulation (see Supplemental Data Set 6 online). Of these, the transcript levels of 287 pairs had a correlation of -0.8 or lower (see Supplemental Figure 2.5 and Supplemental Data Set 6 online). Apparently, the extent of negative correlation does not depend on the distance between the two start codons, in sharp contrast with the extent of positive correlations in transcript levels. The most likely explanation for negatively correlated HTH genes is the steric

occlusion between overlapping promoter regions, where the direct and indirect binding of transcriptional regulators and the RNA polymerase II to one promoter region excludes binding to the other promoter region. This phenomenon is termed as promoter occlusion or promoter interference (Nakajima et al., 1993). Alternatively, in a minority of negatively correlated HTH gene pairs, switch motifs may account for negatively correlated transcription. We postulated that switch motif(s) might also account for the induction of one gene and the repression of the other. It is also possible that insulator motifs separate the two adjacent but unidirectional and oppositely oriented promoter regions. Using Multiple Expectation Maximization for Motif Elicitation (MEME) and MAST (Bailey et al., 2009), we found two novel putative switch motifs (Figure 2.5). Putative switch motif 1 is a degenerate but 21-bp-long motif, while putative switch motif 2 is a well-conserved, 15-bp-long motif (Figure 2.5). Putative switch motif 1 is present in 44 and switch motif 2 is found in nine intergenic regions between negatively correlated HTH genes. Neither switch motif was found in promoter regions of positively correlated HTH genes.

Regulation of Specific Sets of CCM Genes

Carbonic Anhydrases

There are three known *C. reinhardtii* α-carbonic anhydrase genes, six known β-carbonic anhydrases, and three putative γ-carbonic anhydrases (Moroney et al., 2011) (see Supplemental Data Set 7 online). Four of the nine α- and β-CA genes (*CAH1*, *CAH4*/*CAH5*, and *CAH7*) were activated (twofold or more) (Figure 2.2; see Supplemental

Data Sets 2 and 7 online) and none were repressed significantly (FDR ≤ 0.01). Transcripts encoding *CAH1*, an α-type, periplasmic CA and *CAH4*/*CAH5*, recently duplicated, mitochondrial, β-type CAs were induced by over 2^9 (512-fold), similar to RT-PCR observations by Ynalvez et al. (2008). These results are confirmed by the companion article (Fang et al., 2012), including the activation of *CAH1* and *CAH4*/*CAH5*. As a notable difference to the observations of Ynalvez et al. (2008) and Fang et al. (2012), we found *CAH3* slightly upregulated. In the *cia5* mutant, Fang et al. (2012) found that *CAH5*, the most highly upregulated CA gene in wild-type cells, is actually repressed. Moreover, in cia5, *CAH1*, *CAH4*, *CAH6*, and *CAH8* were weakly induced (Ynalvez et al., 2008), suggesting the possible existence of a minor CCM regulatory mechanism independent of CIA5.

Levels of *CAH2*, encoding another periplasmic α-type carbonic anhydrase (Rawat and Moroney, 1991) either remained stable or slightly declined in previous RNA gel blot and macroarray transcriptome analyses (Miura et al., 2004; Wang et al., 2005; Yamano et al., 2008; Ynalvez et al., 2008). These observations were corroborated in our transcriptome analysis and further confirmed by qRT-PCR analyses (see Supplemental Figure 1 online). These observations strengthen the conclusion that although *CAH2* is possibly a recent duplication of the *CAH1* gene (or vice versa), it apparently does not have a role in CCM under limiting CO_2 conditions (Fujiwara et al., 1990).

Because of its possible localization either inside or immediately outside of the plasma membrane, a recently discovered CA, *CAH8*, has been considered to be a

potential contributor to C_i uptake in *C. reinhardtii* (Ynalvez et al., 2008; Moroney et al., 2011). The lack of significant induction of the *CAH8* gene ASVLCO2 by both the Moroney group (Ynalvez et al., 2008) and this study (see Supplemental Data Set 7 online) would suggest that if this is so, *CAH8* either contributes little to enhanced C_i uptake during induction of the CCM or that the *CAH8* enzyme is present constitutively in nonlimiting quantities. Quantitative measurements of *CAH8* concentrations and activities and/or the availability of *CAH8* mutants or knockdown lines are needed to differentiate between these alternatives.

Three putative γ -CA genes are not induced by CO_2 starvation (see Supplemental Data Set 7 online). These putative CAs appear to be strongly coupled with the Complex 1 mitochondrial electron transport chain (Klodmann et al., 2010) and therefore may contribute little to the changes in cellular metabolism triggered by CO_2 deprivation.

LCIB and LCIC

Although no longer considered as directly involved in C_i transport (because of the lack of transmembrane domains), the soluble chloroplast stromal protein, *LCIB* (Wang and Spalding, 2006), and the complex of *LCIB* with its close family member, *LCIC* (Yamano et al., 2010), have been implicated as involved in CO_2 retention in the chloroplast or in assisting the C_i transport system within the chloroplast. The novel CO_2-requiring phenotypes of mutants containing defective *LCIB* genes (Van et al., 2001), the unexpected close functional relationship between *LCIB* and *CAH3* (Duanmu et al., 2009b), and the close physical association of *LCIB* (and supposedly the *LCIB*/*LCIC*

complex) with the ribulose-1,5-bis-phosphate carboxylase/oxygenase–rich pyrenoid body when cells are exposed to CO_2-limiting conditions (Yamano et al., 2010; Wang et al., 2011) have strongly implicated these proteins as key players in the *C. reinhardtii* CCM. Consistent with this involvement is an observed 20- to 30-fold increase in *LCIB* and *LCIC* transcript levels in cells moved from a CO_2-replete to a CO_2-depleted condition (see Supplemental Data Set 1 online).

HLA3 and Other Putative C_i Transporters

Transcripts encoding *HLA3*, an ATP-energized HCO_3^- transporter (Duanmu et al., 2009a), increase at 60 min by 22-fold and at 180 min by 40-fold (see Supplemental Data Set 2 online). Increases of *LCIA* (*NAR1.2*) transcription are a remarkable 3300- fold at 60 min and 4000-fold at 180 min ASVLCO_2. The list of potential C_i transporters also includes the *CCP1* gene (induced 2000-fold) and the *CCP2* gene (induced 120-fold), two long recognized CO_2-responsive genes whose products have been implicated in potential C_i transport across the chloroplast envelopes. Because these four are among the most highly induced genes (see Supplemental Data Set 2 online), CCM assembly or augmentation in response to CO_2 deprivation likely requires the de novo assembly of a number of C_i transporters at several critical subcellular locations. Transporters in the plasma membrane include HLA3 (Duanmu et al., 2009a) and LCI1 (Ohnishi et al., 2010); in the chloroplast envelope, they include LCIA (Duanmu et al., 2009a), CCP1, and CCP2 (Pollock et al., 2004). Because evidence exists for a critical role of CAH3 in converting HCO_3^- to CO_2 in the lumen of the thylakoid membranes of the chloroplast (Moroney et

al., 2011) and the diffusion of the resulting CO_2 to ribulose-1,5-bis-phosphate carboxylase/oxygenase in the pyrenoid body, an argument has been made for the existence of a transporter or channel to shuttle bicarbonate from the alkaline chloroplast stroma to the acidic lumen of the thylakoid membrane. A future quest for this hypothetical HCO_3^- transporter/channel among the strongly induced genes encoding transmembrane proteins may prove rewarding.

In our RNA-Seq study, *GGPS* (*LCI14*) (part of the ~15-element chlorophyll a biosynthesis pathway in higher plants) (Robinson et al., 2010), *LCI25* (encoding a protein homologous to stress induced one-helix protein in *Arabidopsis thaliana*), and *LHCSR1* (encoding a stress-related chlorophyll a/b binding protein) are all downregulated ASVLCO_2. Because both *GGPS* and *LHCSR1* are induced by the transcriptional regulator CIA5/CCM1 (Wang et al., 2005), their downregulation may be due to other transcription factors or noncoding RNAs.

Mitochondrial Functions

Our transcriptome analyses suggest that carbon deprivation suppressed most genes involved in classical mitochondrial functions. Transcripts from 24 of the 73 annotated mitochondrial proteins were suppressed and only six were increased (see Supplemental Data Sets 2 and 3 online; Figure 2.2). On the whole, the decreased transcription of genes producing mitochondrial components is reflective of the decreased metabolic activity of cells exposed to CO_2-limiting conditions.

Regulation of the Response to Very Low CO_2

Despite considerable efforts, most regulators of the CCM remain unknown or incompletely characterized. Indeed, very few *C. reinhardtii* transcription factor binding sites have been determined experimentally (Kucho et al., 2003; Yoshioka et al., 2004; Sommer et al., 2010; Kropat et al., 2011). Inferences from other photosynthetic organisms are also limited due to the scarcity of known transcription factor binding sites (Yilmaz et al., 2011). Previous studies and our companion article (Fang et al., 2012) have focused on CIA5/CCM1, the master regulator of the transcriptional response to low CO_2, and the putative transcription factor LCR1, itself regulated by CIA5 (Kucho et al., 2003). The CIA5/CCM1 orthologs in *C. reinhardtii* and *Volvox carteri* share 51% identity at the amino acid level (Yamano et al., 2011). This conservation extends to two zinc finger domains (a mutation of which gives rise to the CCM defective phenotype of *cia5*), a multifunctional protein–protein interaction domain characteristic of transcriptional regulators (Plevin et al., 2005), putative nuclear export and localization signals, and a sumoylation site. We reported earlier (Wang et al., 2005) that regardless of the ambient CO_2 levels, the CIA5/CCM1 gene is constitutively expressed and the protein is also constitutively localized in the nucleus. Given that there is no evidence for CIA5/CCM1 binding to DNA (Kohinata et al., 2008), our hypothesis is that the atypical zinc-finger domains (Kohinata et al., 2008) of CIA5/CCM1 do not bind directly to DNA but regulate other transcription factors as a transcriptional activator via direct or indirect interactions. We postulate that, once activated, these latter factors may bind to *cis*-

regulatory elements in the promoter regions of CO_2-responsive genes to enhance transcription.

To examine this hypothesis, we searched for potential transcription factor binding sites in the putative promoter regions of two gene sets. The first gene set is the union of low CO_2– induced genes identified from previous publications and/or from our study. The second set is the potential CIA5/CCM1-regulated genes identified by the comparative transcriptomic studies of the *cia5* mutant and wild-type *C. reinhardtii* cells in our companion article (Fang et al., 2012). For these and other sets of promoter regions, we performed DNA sequence motif discovery (see Methods). We found several relatively conserved DNA sequence motifs that, as discussed below, are candidates for transcription factor binding sites. Because none of the binding sites is present in more than a quarter of the low CO_2–induced genes, the CCM is apparently regulated by multiple transcription factors, cofactors, and possibly other regulators. From its high level in the regulatory hierarchy, CIA5/CCM1 appears to regulate many or most subordinate regulators, but not all of them.

LCR1, whose binding (enhancer) sites are essential for the induction of the *CAH1* gene, is a prime example of a CIA5/CCM1-regulated transcription factor. Using gel mobility shift assays, the Fukuzawa laboratory identified two enhancer elements (EE's) in the promoter region of *CAH1* (Kucho et al., 2003; Yoshioka et al., 2004). These elements, EE-1 (5'-AGATTTTCACCGGTTGGAAGGAGGT-3') and EE-2 (5'-CGACTTACGAA-3'), share the DNA sequence GANTTNC (referred to here as Motif 1A). Later, this motif

was shown to be the binding site for *LCR1* (Miura et al., 2004). *LCR1* is an outlier of the *1R-Myb* subfamily; its *Myb* domain is only ≈20% identical to its relatives. We confirmed the extremely strong induction of the gene coding for the transcription factor *LCR1* following CO_2 deprivation (see Supplemental Data Set 2 online). We found for the GANTTNC motif (and its reverse complement) in the upstream 500-bp regions upstream of the start codon of the 893 genes that were upregulated in our experiments using very stringent upregulation criteria: $\log_2$FC ≥ 1 and FDRq ≤ 0.0001 3 h afterCO_2 deprivation. A total of 476 promoter regions (53%) matched this low information content motif or its reverse complement (see Supplemental Table 1 online). Surprisingly, however, 697 of the 1227 downregulated genes (78%) and 7141 of the 13,380 genes (53%) that were not significantly differentially regulated by the above criteria also matched Motif 1A. To increase the specificity of the low information content Motif 1A, we searched for its supersets EE-1 and EE-2 in the promoter regions of upregulated genes but found no exact matches other than the previously identified single copy in the *CAH1* promoter region. Therefore, we performed de novo motif discovery using a strategy described in detail earlier (Ladunga, 2010). Although no chromatin immunoprecipitation, protein binding array, or protein–protein interaction data are available for *C. reinhardtii*, diverse algorithms for promoter sequence analysis, each with different advantages, knockout mutants, and expression data allowed us to obtain insight into the regulatory network of CCM-associated genes (see Methods).

Putative Novel cis-Regulatory elements: (CA)n Repeats

From the promoter regions of all upregulated genes in our experiments, the MEME tool (Bailey et al., 2009) reported new motifs. The most remarkable is a $(CA)_n$ repeat (Motif 2), where n ranges from 9 to over 20. $(CA)_n$ is present in 29% of the promoters of upregulated genes ASVLCO2, but it is almost absent from the promoters of downregulated genes. Indications of potential transcription activating mechanisms associated with $(CA)_n$ repeats can be found in mammalian and yeast systems. In the human genome, 19.4 CA repeats occur per mega base pair, representing the most common simple-sequence repeat motif (Waterston et al., 2002). Among all dinucleotides, the structure of CA dinucleotides are the most stable regardless of their environment, and the stability of the DNA structure enhances the stability of the chromatin as well (Fujii et al., 2007). This stable chromatin structure allows the binding of numerous transcriptional activators and splicing regulators. Heterogeneous nuclear RiboNuclearProtein L (hnRNP L) in mammals binds to an intronic polymorphic CA repeat region in the human endothelial nitric oxide synthase gene (Hui et al., 2005). hnRNP L has a role in determining alternative splicing and splicing efficiency. A homologous nuclear ribonucleoprotein exists in *C. reinhardtii* as well (Johnston et al., 1999). In humans, the length of the $(CA)_n$ repeats positively correlates with the expression of the interferon γ gene (Pravica et al., 2000) and the integrin α2 gene. Similarly, alleles with longer $(CA)_n$ repeats in the upstream regulatory region of the integrin α2 gene enhance the specific binding of poly (ADP-ribose)polymerase-1, a molecular nick sensor, and KU80 polypeptide (Ku80)/70, two components of transcription coactivator complexes

(Cheli et al., 2010). *C. reinhardtii* genes encode for proteins remotely similar to Poly(ADP-ribose) polymerase- 1 (PARP-1), including the chloroplast tscA maturation factor (Protein ID: 525840, XP_001694431). The Ku80-like domain is also present in two proteins (509698, XP_001702579; 515111, XP_001699098). These observations suggest the possibility that the *C. reinhardtii* $(CA)_n$ motif may recruit transcription activator complexes similar to the mammalian PARP-1 Ku80/70–containing complexes.

Transcription Factors That May Bind to the (CA)n Motif

Searching against all known motifs in the TRANSFAC Professional Database (Matys et al., 2006), we found that $(CA)_n$ and similar sequences are bound both in vivo and in vitro by yeast Ras-related protein1 (Rap1), another *Myb* family repressor-activator protein. In yeast, the general transcription factors Rap1 and autonomously replicating sequence-binding factor1 (Abf1) can create local nucleosome-free regions by evicting nucleosomal histones, thereby facilitating the binding of more specific transcription factors (Castruita et al., 2011). Rap1 was observed to bind to 5'-CACACCCACACACC-3' motifs as well, and even low affinity binding sites for Abf1 and Rap1 play a role in determining nucleosome occupancy (Ganapathi et al., 2007). Rap1 is also implicated in silencing at telomeres and silent mating type loci in yeast (Iglesias et al., 2011). A BLAST search could not detect any homologs of Rap1 in the *C. reinhardtii* proteome. However, more sensitive hidden Markov model searches as implemented in the *hmmsearch* tool (Johnson et al., 2010) against the PFAM Database of protein domains (Punta et al., 2012) indicated that, at the very least, 55 *C. reinhardtii* proteins

carry some combinations of the *Myb-, Sant-, BRCT domain of the BRAC1 oncogene*, or homeodomain-related domains that are characteristic of Rap1. Of these, 35 genes are slightly upregulated at 180 min after CO_2 deprivation. Based on domain similarity with Rap1 and the upregulation of their genes, our best candidate for $(CA)_n$ binding is the predicted protein 515479 (XP_001692722), conserved in *V. carteri*, *Chlorella vulgaris* NC64A, *Ostreococcus tauri*, *Physcomitrella patens*, and *Micromonas pusilla*. Other candidates include two *Myb2*-like transcription factors (Protein IDs 511319 [XP_001690083] and 519163 [XP_001699726]), each of them harboring as many as 11 to 12 *Sant* DNA binding domains conserved across many eukaryotes, even in mammals. *Myb2* transcription factors, associated with calmodulins, coregulate salt and dehydration response in *Arabidopsis* (Yoo et al., 2005) and numerous other biological processes. LCR1, despite its robust induction, is a less likely candidate because it shares only the *Myb* domain with RAP1. In summary, the presence of $(CA)_n$ repeat elements in 29% of the promoters of upregulated genes and their complete absence from the promoters of downregulated genes (see Supplemental Data Set 3 online) combined with knowledge of the activator role these elements in the human integrin α2 and interferon γ genes discussed above, suggest that the $(CA)_n$ motif in *C. reinhardtii* may also serve as binding sites for transcriptional activator proteins.

A fourth member of the *Myb* family of transcription factors is the also induced *Myb11* (Protein ID 511940, XP_002945989). Its *Arabidopsis* ortholog regulates the biosynthesis of flavonols that absorb potentially damaging UV-B radiation (Stracke et al.,

2010b; Stracke et al., 2010a). In *Arabidopsis*, intense light conditions induce *Myb11*, even under normal CO_2 conditions (Stracke et al., 2010b). In *C. reinhardtii*, because relatively intense light appears to be necessary for the activation of carbon-concentrating mechanisms (Yamano et al., 2008), *Myb11* and other light-induced *Myb* transcription factors may be necessary for low CO_2 response.

The discovery of these new, putative, transcription factors should facilitate the discovery of the transcription factor network(s) that regulate several CCM-associated genes. The development of synthetic gene promoters containing these and other putative transcription factor binding sites (motifs) and the isolation of the transcription factors that bind these sites are but two approaches by which the *C. reinhardtii* community can begin to understand the detailed mechanisms by which algal cells sense and respond to external changes, such as altered CO_2 abundance.

Biological and Experimental Reproducibility

The companion study by Fang et al. (2012) focusing on both the effects of CO_2 deprivation and the role of the CIA5 gene in regulating the *C. reinhardtii* CCM nicely complement this article. Their findings support many of our prime observations and conclusions in regard to areas in which our studies overlap (i.e., in comparison of transcriptome changes associated with CO_2 deprivation). Nonetheless, the two studies were conducted completely independently and involved a number of significant differences in conditions (detailed below). Thus, it is not unexpected in the comparison of our two data sets there are differences in the patterns and the magnitude of gene

expression levels (i.e., it is to be expected that there will be considerably lower interlaboratory reproducibility than intralaboratory reproducibility). Key differences in the two studies include the strains employed, transcript level measurement technologies, the tempo and mode of reducing CO_2 levels, temperature, illumination, and cell densities.

Tempo and Mode of CO_2 Deprivation. In our earlier microarray studies, we rapidly shifted cells from high to ambient CO_2 levels (Wang et al., 2005). In the Fukuzawa laboratory's EST-based macroarray studies (Miura et al., 2004; Yamano et al., 2008; Yamano and Fukuzawa, 2009), CO_2 levels were shifted gradually. In this study, we initially maintained high CO_2 levels by sparging with 5% CO_2. At 0 time, we switched the input to a mixture of nitrogen and 100 ppm CO_2, which, within the rapidly stirred fermenter, resulted in a slow decline in CO_2 concentration: at 50 min, to ambient levels; at 75 min, to 100 ppm; and finally, at 180 min, to ~20 ppm (Figure 2.1). By contrast, Fang et al. (2012) measured the effect of CO_2 deprivation 1 h later than we did, 4 h ASVLCO$_2$. Given the rapid increases and decreases in transcript levels seen for many genes during the 3-h time course of CO_2 starvation (Figure 2.3), it is expected the extra hour of CO_2 deprivation in the studies of Fang et al. (2012) might well lead to at least some of the differences observed.

Light and Temperature. Our *C. reinhardtii* cultures grew in a rapidly stirred fermenter at 25°C under white light illumination at 200 mmol photons s^{-1} m^{-2}, 2 times higher than in our companion article (Fang et al., 2012). In *C. reinhardtii* cultures, a light

intensity of 200 mmol photons s^{-1} m^{-2} is sufficiently high to potentially cause oxidative damage to pigments, such as chlorophyll (Peers et al., 2009), damage that is much less likely to occur at a light intensity of 100 mmol s^{-1} m^{-2}. In addition, Fang et al. (2012) maintained cultures at room temperature, whereas our cultures were warmed to a constant 25°C.

Strains of C. reinhardtii May Differ in Splicing and Genomic Polymorphisms. A high proportion of sequencing reads from a nonstandard strain cannot be mapped to the reference transcriptome and genome or to the probes of a microarray, and fewer reads may bias the estimates of differential gene expression. For example, when one or more exons of a multiexon gene are not expressed but the others are, this bias decreases the expression difference that is calculated for a gene model incorrect for the strain.

Quantitative Analysis of Reproducibility

In assessing the potential causes of differences between results in our study and those in our companion article by Fang et al. (2012), we feel the above-mentioned variations in strains and culture conditions likely account for the bulk of differences in gene expression levels and patterns. Because previous studies reported almost exclusively genes induced by CO_2 deprivation (Miura et al., 2004; Wang et al., 2005; Yamano et al., 2008; Yamano and Fukuzawa, 2009), our analysis will necessarily need to be limited to induced genes. To ensure conservative estimates of differential gene expression, our analyses based on the *edgeR* tool (Robinson et al., 2010), an FDR

threshold of 10^{-4} is applied. Our experiments reproduced 166 of the 393 genes (42%) upregulated by CO_2 deprivation in our companion study (Fang et al., 2012) as shown by a Venn diagram (see Supplemental Figure 1 online). We also reproduced 40 genes (38%) upregulated in previous experiments (Miura et al., 2004; Wang et al., 2005; Yamano et al., 2008; Yamano and Fukuzawa, 2009; Yamano et al., 2010), and Fang et al. (2012) reproduced 39 genes (37%). Importantly, 27 of the 106 previously published genes (25%) were reproduced by both Fang et al. (2012) and us. Indeed, the expression patterns of genes most critical to CCM are similar across these three gene sets (see Supplemental Data Set 1 online). In the order of fold change 180 versus 0 min, these genes include carbonic anhydrases 1, 4, 5; *NAR1* (*LCIA*), low carbon–induced genes (*LCI*) *1, 23, 11, D, B, 15, C, 22, 31*, and *19*, as well as *CCP1* and *2*, *LHCSR2* and *3*, *HLA3*, *guanine deaminase 1* (*GUD1*), *Chloroplast DnaJ-like protein3* (*CDJ3*), *ELI4* (*LCI16*), and *Alanine Aminotransferase1* (*AAT1*).

Perspectives

The RNA-Seq data presented here relating changes in environmental CO_2 abundance to changes in gene expression levels greatly expand our knowledge and understanding of the CCM compared with data obtained in more limited RNA gel blot and microarray studies conducted earlier. Some general conclusions made in the earlier investigations have been confirmed, such as the expected decline in photosynthetic capacity as CO_2 availability becomes limited. Other observations, such as the marked downregulation of a vast number of genes immediately after CO_2 decrease have not

been noted earlier and may provide the foundation for a more detailed examination of how algal cells perceive and respond so rapidly to changes in environmental conditions that demand prompt changes in cellular metabolism and physiology. Likewise, our studies revealed an unexpectedly large number of gene pairs involved in similar cellular activities that are oriented in a HTH conformation, where a shared bidirectional promoter coordinately regulates the response to changes in CO_2 levels. Investigations of common regulatory elements now become a potentially fruitful area of investigation. Our results from a time-course study of changes in the transcriptomes at critical time points after imposition of CO_2 starvation builds on results from the accompanying publication of the Spalding laboratory (Fang et al., 2012) regarding the pivotal regulatory role of CIA5/CCM1 in response to CO_2 deprivation.

Our data confirm that synthesis of many of the key components of the CCM are triggered when CO_2 levels decrease and reveal an even greater number of candidate genes whose function may be essential or helpful to assembly of a fully functional CCM. These genes become potential targets for chromatin immunoprecipitation and deep sequencing (Wang et al., 2011) or for inactivation through RNA interference techniques (Cerutti et al., 2011), insertional gene inactivation (Ermilova et al., 2000), or targeted gene knockout using zinc-finger nucleases or newly developed techniques, such as TAL effector nucleases (Li et al., 2011a). The vast array of genes induced or repressed as a result of CO_2 changes point to a sophisticated set of regulatory networks that must

effectively govern multiple cellular responses, an area of study that appears fruitful for future exploration.

Methods and Materials

C_i Deprivation

Chlamydomonas reinhardtii wild-type strain CC124 was used for analysis. Briefly, cells were grown in 2 liters of Tris Phosphate medium at 25°C and 3% CO2 to a density of 1 X 10^6 cells/mL (Harris, 1989) before being transferred to a 3-liter autoclavable glass bioreactor (Applikon Biotechnology) that was connected with EZ control for analysis of temperature, pH, and dissolved oxygen. The bioreactor was illuminated with a light intensity of 200 µmol photons m^{-2} s^{-1}, and an input gas containing 5% CO_2 was introduced. Algal cells were allowed to equilibrate with the new environment for 1 h. Following a sampling of the culture, the input gas for the bioreactor was shifted to 100 ppm CO_2, which was monitored in the culture using two CO_2 transmitters (Vaisala; models GMT221 and GMT222). Samples were taken at 15, 30, 60, and 180 min following the shift to 100 ppm CO_2. During the experiment, pH was maintained at 7.2 using 3 M KOH.

RNA Preparation

Cellular samples were taken from the bioreactors in 100-mL volumes and transferred directly to a sterilized 2-liter Erlenmeyer flask submerged in an ice water bath. The flask was agitated within the ice water for 2 min to rapidly decrease the

temperature of the algal culture to reduce the degree of transcriptional changes following removal from the bioreactor. The cellular samples were then centrifuged at 2000g for 5 min at 4°C. The supernatant was discarded and the cellular pellet was frozen by immersing the centrifuge tube in liquid nitrogen. Samples were stored at -80°C until RNA extraction.

RNA extraction was performed using TRIzol LS reagent (Life Technologies). Briefly, cellular pellets composed of ~5 x 10^7 cells were resuspended and incubated for 5 min at room temperature with Trizol LS reagent. Chloroform was added to the samples, agitated for 15 s by tube inversion, and allowed to incubate at room temperature for 15 min. Samples were centrifuged at 12,000g for 15 min at 4°C. The aqueous phase was recovered and transferred to new tubes. Isopropanol was added to the aqueous phase, mixed, and centrifuged at 12,000g for 10 min at 4°C to pellet RNA. Isopropanol was removed and the RNA pellet was washed with 75% ethanol. Following further centrifugation (7500g for 5 min at 4°C), ethanol was removed and the RNA pellets were allowed to air dry for 5 min. RNA pellets were solubilized in 500 mL nuclease-free water. RNA was further purified by precipitation with lithium chloride. An equal volume (500 µL) of 4 M LiCl was added with mixing, and samples were incubated at -20°C for 60 min. Following this incubation, samples were centrifuged at 16,000g for 20 min at 4°C. The aqueous phase was removed and discarded and the RNA pellet was washed twice with 75% ethanol. Following air drying, the pellet was resuspended in 50 mL of nuclease-free

water. RNA samples were then analyzed using a NanoDrop 2000 spectrophotometer (NanoDrop Products/Thermo Fisher Scientific) to verify RNA quantity and purity.

Preliminary Analysis of RNA Samples

To confirm induction of the carbon-concentrating mechanism, preliminary analysis of RNA samples was performed using qRT-PCR. RNA samples were prepared for analysis using the Plexor Two-Step qRT-PCR system (Promega). qRT-PCR analysis was performed using a 7500 Real-Time PCR System (Life Technologies). The genes *LCIA* (AB168092), *LCIB* (XM_001698292), and mitochondrial carbonic anhydrase (*CAH4*, XM_001695951) were chosen for analysis as they have been observed to increase in expression during carbon deprivation (Miura et al., 2004; Yamano et al., 2008). *CAH2* (X54488) was also selected as a control gene reported as displaying a moderate decrease in expression in response to carbon deprivation. *CIA5*/*CCM1* (AF317732) was used as a positive control as it shows constitutive expression during carbon deprivation. Fluorescently labeled primer pairs were designed for each of the aforementioned genes. Quantitative PCR analysis was performed using a 7500 Real-Time PCR System by measuring the threshold cycle (C_t) of each gene. Using the C_t values of CIA5/CCM1 for each RNA sample as a baseline control, the change in C_t for each gene could be used to calculate the fold change response of each gene throughout the time course (see Supplemental Figure 1 online).

RNA Sequencing, Mapping, and the Analyses of Gene Expression

From the qRT-PCR data, it was determined that four time points should be analyzed by RNA-Seq. A 15-min ASVLCO$_2$ sample was omitted from RNA-Seq as qRT-PCR analysis of this sample showed limited induction of the aforementioned genes. *C. reinhardtii* equilibrated at ~5% CO2 was used as the 0 time control, and three time points ASVLCO$_2$ (30, 60, and 180 min) were also analyzed. To provide for biological replicates, RNA samples from two individual bioreactor runs were analyzed. In total, eight RNA samples were submitted for RNA-Seq. Prior to submission, RNA samples were treated with DNase and resuspended in 8.3 mM Tris-HCl and 4.2 mM EDTA. RNA-Seq was performed at the JGI using an Illumina Genome Analyzer II.

Sequencing reads were mapped to the *C. reinhardtii* version 4 genome (Department of Energy JGI) as well as to the processed Augustus5 (Stanke and Waack, 2003) exon structure predictions using the *tophat* and *cufflinks* software (Trapnell et al., 2009). No more than two mismatches per sequencing read were allowed. Analyses of differential expression including FDR calculations were performed using three independent Bioconductor packages: *edgeR* (Robinson et al., 2010), *DESeq* (Anders and Huber, 2010), and *baySeq* (Hardcastle and Kelly, 2010).

Time series analysis of the transcriptional response was performed by *k*-means cluster analysis (MacQueen, 1967). This method partitions fold change patterns into *k* clusters where each fold change time series belongs to the cluster with the nearest mean. The clusters are iteratively refined. Such analyses have been used to identify temporal expression patterns of a large numbers of genes.

Gene Ontology Analyses

Complex functional patterns of the differentially regulated genes emerge at the level of photosynthetic categories, low CO_2–regulated genes, and plant and diatom genes that have no close relatives in other kingdoms or in prokaryotes other than cyanobacteria (Karpowicz et al., 2011). We extended these analyses to GO (Ashburner et al., 2000), a system for the hierarchical annotation of homologous gene and protein sequences in multiple organisms using a common, controlled vocabulary. GO allows the practical, high-throughput interpretation of experiments including RNA-Seq. To avoid the subjectivity inherent in the ad hoc interpretations for less than obvious patterns, a rigorous method was employed to assess the statistical significance of expression patterns, called GSEA (Subramanian et al., 2005). Briefly, GSEA ranks genes by fold changes and calculates enrichment scores for each set. Then, primarily upregulated gene sets are assigned high positive enrichment scores and primarily downregulated sets are assigned low negative scores. For the statistical significance of these enrichment scores, FDR (Benjamini and Hochberg, 1995) is calculated.

De Novo Motif Discovery of Putative Transcription Factor Binding Sites

Our complex strategy for the discovery and limited confirmation of the transcriptional regulatory network was described earlier (Ladunga, 2010). Even in the almost complete absence of chromatin immunoprecipitation, protein binding array, or protein–protein interaction observations for algae, an array of motif discovery algorithms for promoter sequence analysis, each complementing the others, knockout

mutants of transcription factors, and RNA-Seq data allowed us to better understand the CCM regulatory network. A key tool is the MEME package (Bailey et al., 2009) for the identification of statistically overrepresented variable sequence motifs. We searched all promoter regions for all motifs represented as positional weight matrices in the commercial version of the TRANSFAC Database (Matys et al., 2006) using its advanced search tool. Conversely, all identified motifs were queried against the TRANSFAC motifs.

Accession Numbers

Sequence data from this article can be found in the GenBank/EMBL data libraries under the following accession numbers (sequences are from *C. reinhardtii* unless otherwise noted): *AAT1*, XM_001698466; *Saccharomyces cerevisiae Abf1*, M29067; *Argonaut1*, XM_001694788; *CAH1*, D90206; *CAH2*, X54488; *CAH3*, U40871; *CAH4*, XM_001695951; *CAH5*, XM_001700718; *CAH6*, AY463239; *CAH7*, EU045569; *CAH8*, EU045570; *CAH9*, XM_001700857; *CAH10*, XM_001703185; *CAH11*, AY538680; *CAH12*, AY463241; *CCP1*/*LCIE*, XM_001692145; *CCP2*/*LCID*, XM_ 001692236; *CDJ3*, XM_001700205; *CIA5*/*CCM1*, AF317732; *CLR18*, XM_001692143; *Dicer-like1*, EU707797; *ELI4*/*LCI16*, XM_001694932; *FAP292*, XM_001703500; *GGPS*/*LCI14*, XM_001694932; *GUD1*, XM_001695571; *HCF136*, XM_001696142; *HLA3*, XM_001699988; *human integrin α2*, NM_002203; *human interferon γ2*, NM_005534; *Ku80-like domain*, XM_001702527; *LCI1*, XM_001703335; *LCI11*, XM_001697911; *LCI15*, XM_001691879; *LCI19*, XM_001698768; *LCI21*, XM_001692906; *LCI22*, XM_001702321; *LCI23*, XM_001695392; *LCI25*, XM_001696602; *LCI31*, XM_001691740; *LCI6*, XM_001693972,

AB168091; *LCI7*, XM_001696515; *LCIA*, AB168092; *LCIB*, XM_001698292; *LCIC*, AB168094; *LCID*, DQ657195; *LCIE*, DQ649007; *LCR1*, AB168089; *LHCSR1*, XM_001696086; *LHCSR2*, XM_001696012; *LHCSR3*, XM_001696086; *Arabidopsis thaliana Myb11*, AF062863; *Myb2*, XM_001690031; *human PARP1*, NM_001618; *Saccharomyces cerevisiae Rap1*, M18068; *tcsA*, XM_001694379. All RNA-Seq data are available at the National Center for Biotechnology Information Sequencing Read Archive at http://trace.ncbi.nlm.nih.gov/Traces/sra/?study=SRP004215.

Supplemental Data

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

Supplemental Figure 1. qRT-PCR Support for RNA Sequencing Results.

Supplemental Figure 2. Overlaps among the Lists of Differentially Expressed Genes Reported by the *edgeR*, *DESeq*, and *baySeq* Tools.

Supplemental Figure 3. Genes Showing a Transient Decrease in Response to CO_2 Deprivation.

Supplemental Figure 4. Genes Showing a Transient Increase in Response to CO_2 Deprivation.

Supplemental Figure 5. Spacing and Expression of HTH Conformation Genes.

Supplemental Figure 6. Comparison of Transcript Levels from Two Closely Linked Head-To-Head Gene Pairs.

Supplemental Figure 7. Overlaps Among Sets of Genes Induced by CO_2 Deprivation.

Supplemental Table 1. Widespread DNA Sequence Motifs as Putative Transcription Factor Binding Sites in the Upstream Regions of Genes.

Supplemental Data Set 1. Comparison of Differential Regulation Observations between RNA-Seq (This Study) and Arrays.

Supplemental Data Set 2. The Statistically Significantly Upregulated Genes for 180 versus 0, 60 versus 0, and 30 versus 0 Min ASVLCO$_2$.

Supplemental Data Set 3. The Statistically Significantly Downregulated Genes for 180 versus 0, 60 versus 0, and 30 versus 0 Min ASVLCO$_2$.

Supplemental Data Set 4. Gene Ontology Categories (Ashburner et al., 2000) Significantly Enriched in Upregulated and Downregulated Genes, Respectively.

Supplemental Data Set 5. The Most Highly Correlated HTH Gene Pairs Potentially Regulated by Bidirectional Promoters.

Supplemental Data Set 6. Anticorrelated Expression in HTH Gene Pairs with a Pearson Correlation Coefficient $r \leq -0.8$.

Supplemental Data Set 7. Expression Profiles of the 12 Carbonic Anhydrase Genes.

Acknowledgements

We thank Christa Pennachio, Erika Lindquist, and Feng Chen at JGI for sequencing, Paul Blum and Kemp Horken for help with the biofermenter, Martin H. Spalding, Wei Fang, Sabeeha S. Merchant, and Matteo Pellegrini for the access to their results and for reviewing the article, and Jean-Jack M. Riethoven and Kartik Vedalaveni

for assistance. D.P.W. acknowledges support from the Department of Energy JGI (Award JGI 2010 CSP-146), the National Science Foundation (Grants MCB-0952533 and EPSCoR-1004094), and Consortium for Algal Biofuels Commercialization (Department of Energy Award DE-EE0003373). I.L. and D.P.W. received support from an interdisciplinary grant from the University of Nebraska–Lincoln.

References

Anders, S., and Huber, W. (2010). Differential expression analysis for sequence count data. Genome Biol **11,** R106.

Ashburner, M., Ball, C.A., Blake, J.A., Botstein, D., Butler, H., Cherry, J.M., . . . Sherlock, G. (2000). Gene ontology: tool for the unification of biology. The Gene Ontology Consortium. Nat Genet **25,** 25-29.

Bailey, T.L., Boden, M., Buske, F.A., Frith, M., Grant, C.E., Clementi, L., . . . Noble, W.S. (2009). MEME SUITE: tools for motif discovery and searching. Nucleic Acids Res **37,** W202-208.

Benjamini, Y., and Hochberg, Y. (1995). Controlling the False Discovery Rate - a Practical and Powerful Approach to Multiple Testing. Journal of the Royal Statistical Society Series B-Methodological **57,** 289-300.

Birney, E., Stamatoyannopoulos, J.A., Dutta, A., Guigo, R., Gingeras, T.R., Margulies, E.H., . . . Consortium, E.P. (2007). Identification and analysis of functional elements in 1% of the human genome by the ENCODE pilot project. Nature **447,** 799-816.

Casas-Mollano, J.A., Rohr, J., Kim, E.J., Balassa, E., van Dijk, K., and Cerutti, H. (2008). Diversification of the core RNA interference machinery in Chlamydomonas reinhardtii and the role of DCL1 in transposon silencing. Genetics **179,** 69-81.

Castruita, M., Casero, D., Karpowicz, S.J., Kropat, J., Vieler, A., Hsieh, S.I., . . . Merchant, S.S. (2011). Systems Biology Approach in Chlamydomonas Reveals Connections between Copper Nutrition and Multiple Metabolic Steps. Plant Cell **23,** 1273-1292.

Cerutti, H., Ma, X.R., Msanne, J., and Repas, T. (2011). RNA-Mediated Silencing in Algae: Biological Roles and Tools for Analysis of Gene Function. Eukaryotic Cell **10,** 1164-1172.

Cheli, Y., Williams, S.A., Ballotti, R., Nugent, D.J., and Kunicki, T.J. (2010). Enhanced Binding of Poly(ADP-ribose)polymerase-1 and Ku80/70 to the ITGA2 Promoter via an Extended Cytosine-Adenosine Repeat. PLoS One **5**.

Dohm, J.C., Lottaz, C., Borodina, T., and Himmelbauer, H. (2008). Substantial biases in ultra-short read data sets from high-throughput DNA sequencing. Nucleic Acids Res **36**.

Duanmu, D., Miller, A.R., Horken, K.M., Weeks, D.P., and Spalding, M.H. (2009a). Knockdown of limiting-CO2-induced gene HLA3 decreases HCO3- transport and photosynthetic Ci affinity in Chlamydomonas reinhardtii. Proceedings of the National Academy of Sciences of the United States of America **106,** 5990-5995.

Duanmu, D.Q., Wang, Y.J., and Spalding, M.H. (2009b). Thylakoid Lumen Carbonic Anhydrase (CAH3) Mutation Suppresses Air-Dier Phenotype of LCIB Mutant in Chlamydomonas reinhardtii. Plant Physiol **149,** 929-937.

Ermilova, E.V., Zalutskaya, Z.M., Gromov, B.V., Hader, D.P., and Purton, S. (2000). Isolation and characterisation of chemotactic mutants of Chlamydomonas reinhardtii obtained by insertional mutagenesis. Protist **151,** 127-137.

Fang, W., Si, Y.Q., Douglass, S., Casero, D., Merchant, S.S., Pellegrini, M., . . . Spalding, M.H. (2012). Transcriptome-Wide Changes in Chlamydomonas reinhardtii Gene Expression Regulated by Carbon Dioxide and the CO2-Concentrating Mechanism Regulator CIA5/CCM1. Plant Cell **24,** 1876-1893.

Fujii, S., Kono, H., Takenaka, S., Go, N., and Sarai, A. (2007). Sequence-dependent DNA deformability studied using molecular dynamics simulations. Nucleic Acids Res **35,** 6063-6074.

Fujiwara, S., Fukuzawa, H., Tachiki, A., and Miyachi, S. (1990). Structure and differential expression of two genes encoding carbonic anhydrase in Chlamydomonas reinhardtii. Proc Natl Acad Sci U S A **87,** 9779-9783.

Ganapathi, M., Singh, G.P., Sandhu, K.S., Brahmachari, S.K., and Brahmachari, V. (2007). A whole genome analysis of 5 ' regulatory regions of human genes for putative cis-acting modulators of nucleosome positioning. Gene **391,** 242-251.

Gonzalez-Ballester, D., Casero, D., Cokus, S., Pellegrini, M., Merchant, S.S., and Grossman, A.R. (2010). RNA-Seq Analysis of Sulfur-Deprived Chlamydomonas Cells Reveals Aspects of Acclimation Critical for Cell Survival. Plant Cell **22,** 2058-2084.

Hardcastle, T.J., and Kelly, K.A. (2010). baySeq: Empirical Bayesian methods for identifying differential expression in sequence count data. Bmc Bioinformatics **11**.

Harris, E.H. (1989). The Chlamydomonas sourcebook : a comprehensive guide to biology and laboratory use. (San Diego: Academic Press).

Hui, J.Y., Hung, L.H., Heiner, M., Schreiner, S., Neumuller, N., Reither, G., . . . Bindereif, A. (2005). Intronic CA-repeat and CA-rich elements: a new class of regulators of mammalian alternative splicing. Embo Journal **24,** 1988-1998.

Iglesias, N., Redon, S., Pfeiffer, V., Dees, M., Lingner, J., and Luke, B. (2011). Subtelomeric repetitive elements determine TERRA regulation by Rap1/Rif and Rap1/Sir complexes in yeast. EMBO Rep **12,** 587-593.

Im, C.S., and Grossman, A.R. (2002). Identification and regulation of high light-induced genes in Chlamydomonas reinhardtii. Plant J **30,** 301-313.

Im, C.S., Zhang, Z.D., Shrager, J., Chang, C.W., and Grossman, A.R. (2003). Analysis of light and CO2 regulation in Chlamydomonas reinhardtii using genome-wide approaches. Photosynthesis Research **75,** 111-125.

Johnson, L.S., Eddy, S.R., and Portugaly, E. (2010). Hidden Markov model speed heuristic and iterative HMM search procedure. Bmc Bioinformatics **11**.

Johnston, S.D., Lew, J.E., and Berman, J. (1999). Gbp1p, a protein with RNA recognition motifs, binds single-stranded telomeric DNA and changes its binding specificity upon dimerization. Mol Cell Biol **19,** 923-933.

Karpowicz, S.J., Prochnik, S.E., Grossman, A.R., and Merchant, S.S. (2011). The GreenCut2 Resource, a Phylogenomically Derived Inventory of Proteins Specific to the Plant Lineage. Journal of Biological Chemistry **286,** 21427-21439.

Klodmann, J., Sunderhaus, S., Nimtz, M., Jansch, L., and Braun, H.P. (2010). Internal Architecture of Mitochondrial Complex I from Arabidopsis thaliana. Plant Cell **22,** 797-810.

Kohinata, T., Nishino, H., and Fukuzawa, H. (2008). Significance of zinc in a regulatory protein, CCM1, which regulates the carbon-concentrating mechanism in Chlamydomonas reinhardtii. Plant and Cell Physiology **49,** 273-283.

Kropat, J., Hong-Hermesdorf, A., Casero, D., Ent, P., Castruita, M., Pellegrini, M., . . . Malasarn, D. (2011). A revised mineral nutrient supplement increases biomass and growth rate in Chlamydomonas reinhardtii. Plant Journal **66,** 770-780.

Kucho, K., Yoshioka, S., Taniguchi, F., Ohyama, K., and Fukuzawa, H. (2003). Cis-acting elements and DNA-binding proteins involved in CO2-responsive transcriptional activation of Cah1 encoding a periplasmic carbonic anhydrase Chlamydomonas reinhardtii. Plant Physiol **133,** 783-793.

Labadorf, A., Link, A., Rogers, M.F., Thomas, J., Reddy, A.S.N., and Ben-Hur, A. (2010). Genome-wide analysis of alternative splicing in Chlamydomonas reinhardtii. Bmc Genomics **11**.

Ladunga, I. (2010). An Overview of the Computational Analyses and Discovery of Transcription Factor Binding Sites. Computational Biology of Transcription Factor Binding **674,** 1-22.

Li, B., Ruotti, V., Stewart, R.M., Thomson, J.A., and Dewey, C.N. (2010). RNA-Seq gene expression estimation with read mapping uncertainty. Bioinformatics **26,** 493-500.

Li, T., Huang, S., Jiang, W.Z., Wright, D., Spalding, M.H., Weeks, D.P., and Yang, B. (2011a). TAL nucleases (TALNs): hybrid proteins composed of TAL effectors and FokI DNA-cleavage domain. Nucleic Acids Res **39,** 359-372.

Li, T., Huang, S., Zhao, X., Wright, D.A., Carpenter, S., Spalding, M.H., . . . Yang, B. (2011b). Modularly assembled designer TAL effector nucleases for targeted gene knockout and gene replacement in eukaryotes. Nucleic Acids Res **39,** 6315-6325.

Lopez, D., Casero, D., Cokus, S.J., Merchant, S.S., and Pellegrini, M. (2011). Algal Functional Annotation Tool: a web-based analysis suite to functionally interpret large gene lists using integrated annotation and expression data. Bmc Bioinformatics **12**.

MacQueen, J. (1967). Some methods for classification and analysis of multivariate observations. In Proc. Fifth Berkeley Symp. on Math. Statist. and Prob. (Univ. of Calif. Press), pp. 281-297.

Margulies, M., Egholm, M., Altman, W.E., Attiya, S., Bader, J.S., Bemben, L.A., . . . Rothberg, J.M. (2005). Genome sequencing in microfabricated high-density picolitre reactors. Nature **437,** 376-380.

Matys, V., Kel-Margoulis, O.V., Fricke, E., Liebich, I., Land, S., Barre-Dirrie, A., . . . Wingender, E. (2006). TRANSFAC (R) and its module TRANSCompel (R): transcriptional gene regulation in eukaryotes. Nucleic Acids Res **34,** D108-D110.

Merchant, S.S., Prochnik, S.E., Vallon, O., Harris, E.H., Karpowicz, S.J., Witman, G.B., . . . Team, J.A. (2007). The Chlamydomonas genome reveals the evolution of key animal and plant functions. Science **318,** 245-251.

Miura, K., Yamano, T., Yoshioka, S., Kohinata, T., Inoue, Y., Taniguchi, F., . . . Fukuzawa, H. (2004). Expression profiling-based identification of CO2-responsive genes regulated by CCM1 controlling a carbon-concentrating mechanism in Chlamydomonas reinhardtii. Plant Physiol **135,** 1595-1607.

Moroney, J.V., and Ynalvez, R.A. (2007). Proposed carbon dioxide concentrating mechanism in Chlamydomonas reinhardtii. Eukaryotic Cell **6,** 1251-1259.

Moroney, J.V., Ma, Y.B., Frey, W.D., Fusilier, K.A., Pham, T.T., Simms, T.A., . . . Mukherjee, B. (2011). The carbonic anhydrase isoforms of Chlamydomonas reinhardtii: intracellular location, expression, and physiological roles. Photosynthesis Research **109,** 133-149.

Nakajima, K., Ikenaka, K., Nakahira, K., Morita, N., and Mikoshiba, K. (1993). An Improved Retroviral Vector for Assaying Promoter Activity - Analysis of Promoter Interference in Pip211 Vector. Febs Letters **315,** 129-133.

Ohnishi, N., Mukherjee, B., Tsujikawa, T., Yanase, M., Nakano, H., Moroney, J.V., and Fukuzawa, H. (2010). Expression of a Low CO2-Inducible Protein, LCI1, Increases Inorganic Carbon Uptake in the Green Alga Chlamydomonas reinhardtii. Plant Cell **22,** 3105-3117.

Okuda, S., Yamada, T., Hamajima, M., Itoh, M., Katayama, T., Bork, P., . . . Kanehisa, M. (2008). KEGG Atlas mapping for global analysis of metabolic pathways. Nucleic Acids Res **36,** W423-W426.

Peers, G., Truong, T.B., Ostendorf, E., Busch, A., Elrad, D., Grossman, A.R., . . . Niyogi, K.K. (2009). An ancient light-harvesting protein is critical for the regulation of algal photosynthesis. Nature **462,** 518-U215.

Plevin, M.J., Mills, M.M., and Ikura, M. (2005). The LxxLL motif: a multifunctional binding sequence in transcriptional regulation. Trends in Biochemical Sciences **30,** 66-69.

Pollock, S.V., Prout, D.L., Godfrey, A.C., Lemaire, S.D., and Moroney, J.V. (2004). The Chlamydomonas reinhardtii proteins Ccp1 and Ccp2 are required for long-term growth, but are not necessary for efficient photosynthesis, in a low-CO2 environment. Plant Molecular Biology **56,** 125-132.

Pravica, V., Perrey, C., Stevens, A., Lee, J.H., and Hutchinson, I.V. (2000). A single nucleotide polymorphism in the first intron of the human IFN-gamma gene: Absolute correlation with a polymorphic CA microsatellite marker of high IFN-gamma production. Human Immunology **61,** 863-866.

Proost, S., Van Bel, M., Sterck, L., Billiau, K., Van Parys, T., Van de Peer, Y., and Vandepoele, K. (2009). PLAZA: A Comparative Genomics Resource to Study Gene and Genome Evolution in Plants. Plant Cell **21,** 3718-3731.

Punta, M., Coggill, P.C., Eberhardt, R.Y., Mistry, J., Tate, J., Boursnell, C., . . . Finn, R.D. (2012). The Pfam protein families database. Nucleic Acids Res **40,** D290-D301.

Rawat, M., and Moroney, J.V. (1991). Partial Characterization of a New Isoenzyme of Carbonic-Anhydrase Isolated from Chlamydomonas-Reinhardtii. Journal of Biological Chemistry **266,** 9719-9723.

Robinson, M.D., McCarthy, D.J., and Smyth, G.K. (2010). edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. Bioinformatics **26,** 139-140.

Sommer, F., Kropat, J., Malasarn, D., Grossoehme, N.E., Chen, X.H., Giedroc, D.P., and Merchant, S.S. (2010). The CRR1 Nutritional Copper Sensor in Chlamydomonas Contains Two Distinct Metal-Responsive Domains. Plant Cell **22,** 4098-4113.

Stanke, M., and Waack, S. (2003). Gene prediction with a hidden Markov model and a new intron submodel. Bioinformatics **19,** Ii215-Ii225.

Stracke, R., Jahns, O., Keck, M., Tohge, T., Niehaus, K., Fernie, A.R., and Weisshaar, B. (2010a). Analysis of PRODUCTION OF FLAVONOL GLYCOSIDES-dependent flavonol glycoside accumulation in Arabidopsis thaliana plants reveals MYB11-, MYB12-and MYB111-independent flavonol glycoside accumulation. New Phytologist **188,** 985-1000.

Stracke, R., Favory, J.J., Gruber, H., Bartelniewoehner, L., Bartels, S., Binkert, M., . . . Ulm, R. (2010b). The Arabidopsis bZIP transcription factor HY5 regulates expression of the PFG1/MYB12 gene in response to light and ultraviolet-B radiation. Plant Cell and Environment **33,** 88-103.

Subramanian, A., Tamayo, P., Mootha, V.K., Mukherjee, S., Ebert, B.L., Gillette, M.A., . . . Mesirov, J.P. (2005). Gene set enrichment analysis: A knowledge-based approach for interpreting genome-wide expression profiles. Proceedings of the National Academy of Sciences of the United States of America **102,** 15545-15550.

Trapnell, C., Pachter, L., and Salzberg, S.L. (2009). TopHat: discovering splice junctions with RNA-Seq. Bioinformatics **25,** 1105-1111.

Trinklein, N.D., Aldred, S.F., Hartman, S.J., Schroeder, D.I., Otillar, R.P., and Myers, R.M. (2004). An abundance of bidirectional promoters in the human genome. Genome Research **14,** 62-66.

Van, K.J., Wang, Y.J., Nakamura, Y., and Spalding, M.H. (2001). Insertional mutants of Chlamydomonas reinhardtii that require elevated CO2 for survival. Plant Physiol **127,** 607-614.

Villand, P., Eriksson, M., and Samuelsson, G. (1997). Carbon dioxide and light regulation of promoters controlling the expression of mitochondrial carbonic anhydrase in Chlamydomonas reinhardtii. Biochemical Journal **327,** 51-57.

Wang, Y., Sun, Z.H., Horken, K.M., Im, C.S., Xiang, Y.B., Grossman, A.R., and Weeks, D.P. (2005). Analyses of CIA5, the master regulator of the carbon-concentrating mechanism in Chlamydomonas reinhardtii, and its control of gene expression. Canadian Journal of Botany-Revue Canadienne De Botanique **83,** 765-779.

Wang, Y.J., and Spalding, M.H. (2006). An inorganic carbon transport system responsible for acclimation specific to air levels Of CO2 in Chiamlydomonas reinhardtii. Proceedings of the National Academy of Sciences of the United States of America **103,** 10110-10115.

Wang, Y.J., Duanmu, D.Q., and Spalding, M.H. (2011). Carbon dioxide concentrating mechanism in Chlamydomonas reinhardtii: inorganic carbon transport and CO2 recapture. Photosynthesis Research **109,** 115-122.

Waterston, R.H., Lander, E.S., and Sulston, J.E. (2002). On the sequencing of the human genome. Proceedings of the National Academy of Sciences of the United States of America **99,** 3712-3716.

Yamano, T., and Fukuzawa, H. (2009). Carbon-concentrating mechanism in a green alga, Chlamydomonas reinhardtii, revealed by transcriptome analyses. Journal of Basic Microbiology **49,** 42-51.

Yamano, T., Miura, K., and Fukuzawa, H. (2008). Expression analysis of genes associated with the induction of the carbon-concentrating mechanism in Chlamydomonas reinhardtii. Plant Physiol **147,** 340-354.

Yamano, T., Fujita, A., and Fukuzawa, H. (2011). Photosynthetic characteristics of a multicellular green alga Volvox carteri in response to external CO2 levels possibly regulated by CCM1/CIA5 ortholog. Photosynthesis Research **109,** 151-159.

Yamano, T., Tsujikawa, T., Hatano, K., Ozawa, S., Takahashi, Y., and Fukuzawa, H. (2010). Light and Low-CO2-Dependent LCIB-LCIC Complex Localization in the Chloroplast Supports the Carbon-Concentrating Mechanism in Chlamydomonas reinhardtii. Plant and Cell Physiology **51,** 1453-1468.

Yilmaz, A., Mejia-Guerra, M.K., Kurz, K., Liang, X.Y., Welch, L., and Grotewold, E. (2011). AGRIS: the Arabidopsis Gene Regulatory Information Server, an update. Nucleic Acids Res **39,** D1118-D1122.

Ynalvez, R.A., Xiao, Y., Ward, A.S., Cunnusamy, K., and Moroney, J.V. (2008). Identification and characterization of two closely related beta-carbonic anhydrases from Chlamydomonas reinhardtii. Physiologia Plantarum **133,** 15-26.

Yoo, J.H., Park, C.Y., Kim, J.C., Heo, W.D., Cheong, M.S., Park, H.C., . . . Cho, M.J. (2005). Direct interaction of a divergent CaM isoform and the transcription factor, MYB2, enhances salt tolerance in Arabidopsis. Journal of Biological Chemistry **280,** 3697-3706.

Yoshioka, S., Taniguchi, F., Miura, K., Inoue, T., Yamano, T., and Fukuzawa, H. (2004). The novel Myb transcription factor LCR1 regulates the CO2-responsive gene Cah1, encoding a periplasmic carbonic anhydrase in Chlamydomonas reinhardtii. Plant Cell **16,** 1466-1477.

CHAPTER 3 - Development of Herbicide-Resistance Technologies for Use in the Green Alga *Chlamydomonas reinhardtii*

Abstract

Three herbicide resistance genes were developed and tested for use as dominant selectable markers in genetic transformation of *Chlamydomonas reinhardtii* and as potential tools for the protection of commercial-scale algal production facilities against contamination by a large variety of organisms that are sensitive to the broad-spectrum herbicides for which resistance genes have been developed. A synthetic, codon-optimized and intron-containing glyphosate acetyltransferase (GAT) gene, when fitted with a strong Chlamydomonas promoter, conferred a 2.7x-fold increase in tolerance to the EPSPS-inhibitor, glyphosate, in transgenic cells compared to progenitor WT cells. A mutated protoporphyrinogen oxidase (protox, PPO) gene from Chlamydomonas, previously shown to produce an enzyme insensitive to PPO-inhibiting herbicides, was modified for use as a transgene. The modified gene generated transgenic lines able to tolerate up to 136x higher levels of the PPO-inhibitor oxyfluorfen than nontransformed cells. By comparing mutant forms of phytoene desaturase (PDS) from various norflurazon-resistant organisms, we identified a promising point mutation and introduced it into the Chlamydomonas PDS gene. Cells transformed with the resultant mutant gene were able to tolerate up to 40x higher levels of norflurazon than nontransgenic cells. All three herbicide resistance genes produced transgenic cells with

high efficiency following electroporation, demonstrating their suitability as dominant selectable markers for genetic transformation of Chlamydomonas and, potentially, other eukaryotic algae. Lack of significant effects of normally lethal doses of the herbicides on the growth rates of transgenic Chlamydomonas cells suggests that use of one or more of the herbicide resistance genes in combination with appropriate herbicide concentrations may prove effective in large-scale production facilities in suppressing growth of organisms sensitive to the herbicides.

Introduction

Interest in utilizing algae as a source of biofuels has led to intensified research with these organisms, ranging from studies of the basic biology of algae to design and engineering of industrial production facilities. Numerous algal strains have been identified as candidates for biofuel production based on their lipid production, growth rates, and suitability for genetic manipulation. The importance of maintaining axenic or near-axenic algal cultures is critical, as contaminating organisms will reduce the production potential of algal bioreactors or raceways. In parallel, molecular work with algae requires a means of selecting for transgenic events. For both purposes, the availability of new genes that confer a selective advantage to the growth of desired algae over non-desired algae is critical. With this dual goal in mind, we initiated the present study to potentially exploit herbicide resistance genes that had not previously been widely used in algal systems as either dominant selectable markers for genetic transformation or for "crop protection" in large-scale algal cultures.

One of the most widely used herbicides in agriculture today is glyphosate. Introduced for use as a broad-spectrum herbicide in 1974 (Monsanto, 2005), glyphosate has become one of the best-selling herbicides worldwide. This is largely due to the development of transgenic crop plants resistant to glyphosate, beginning with the introduction of glyphosate-resistant soybean in 1996 (Padgette et al., 1995; Dill et al., 2008). Glyphosate's mode of action involves inhibition of enolpyruvylshikimate-3-phosphate synthase (EPSPS), an enzyme involved in the biosynthesis of aromatic amino

acids. Glyphosate has been adopted for use in both rural and urban environments, as it exhibits minimal human and environmental toxicity (Smith and Oehme, 1992; Williams et al., 2000). Over the years, substantial research has been dedicated to the discovery and development of genes that can provide resistance to glyphosate. Early efforts focused on mutant and natural forms of EPSPS that could maintain catalytic function in the presence of glyphosate. Two forms predominated. The first was an EPSPS composed of mutant domains from *E. coli* or *Zea mays,* and the second was a natively glyphosate-insensitive EPSPS from *Agrobacterium* CP4. Both forms were used in the development of glyphosate resistant crops, with the *Agrobacterium* EPSPS favored, as it showed higher tolerance to glyphosate than the *E. coli/Zea mays* double mutants (Barry et al., 1994). An alternative mode of glyphosate resistance can be achieved by detoxification of glyphosate. This has been achieved either by using enzymes capable of oxidizing glyphosate or enzymes capable of acetylating glyphosate. An example of the former type of enzyme is glyphosate oxidase (Barry et al., 1992). An example of the latter is an acetyltransferase from *Bacillus* that was modified via DNA shuffling to enhance specificity for glyphosate to yield an enzyme that was able to confer glyphosate-resistance to a variety of organisms (Castle et al., 2004).

Another family of herbicides that are attractive for potential use in algal cultures is the PPO-inhibitors. This class, which includes compounds such as oxadiazon, oxyfluorfen, and acifluorfen, functions by inhibiting protoporphyrinogen oxidase (PPO, protox). Protox, the last common enzyme in the biosynthetic pathway for heme and

chlorophyll production, catalyzes the oxidation of protoporphyrinogen IX to protoporphyrin IX (Beale and Weinstein, 1990). Inhibition of protox leads to an accumulation of the substrate, protoporphyrinogen IX, which is exported from the chloroplast to the cytoplasm, where it is oxidized by a non-specific plasma membrane-bound peroxidase (Ha et al., 2004). Accumulation of the oxidized product, protoporphyrin IX, in the cytoplasm leads to formation of singlet oxygen, resulting in membrane peroxidation (Duke et al., 1991). Compounds that inhibit protox have been used as herbicides in crop systems for many years. Certain crop species, including rice, soybean and tobacco, have been engineered to tolerate normally lethal doses of PPO-inhibitors (Lermontova and Grimm, 2000; Warabi et al., 2001; Ha et al., 2004; Jung et al., 2004).

The third class of herbicides addressed in this work target the enzyme activity of phytoene desaturase (PDS). PDS functions in the carotenoid biosynthesis pathway, and inhibition of this enzyme causes degradation of chlorophyll and the chloroplast membrane, resulting in photobleaching of green tissues (Böger and Sandmann, 1998). To date, three mutant forms of PDS have been identified from various organisms that provide tolerance to PDS-inhibiting herbicides. Two of these mutants were discovered in the photosynthetic cyanobacteria *Synechococcus* and *Synechocystis* (Chamovitz et al., 1991; Martinez-Ferez and Vioque, 1992), while the most recently-discovered norflurazon-resistant PDS mutant was found in *Hydrilla verticillata* (Michel et al., 2004b). Three independent norflurazon-resistant Hydrilla strains were characterized,

and surprisingly, all three showed different amino acid substitutions at the same location in PDS (histidine, serine, or cysteine in place of arginine at position 304). Engineered substitutions at this site showed that a fourth amino acid (threonine) conferred even higher resistance to norflurazon than the three naturally-occurring substitutions (Arias et al., 2006).

Previous work has shown Chlamydomonas to be susceptible to all three of the above-mentioned herbicide classes (Reboud, 2002). To date, no information is publically available detailing efforts to develop glyphosate resistance genes for algae. In 1990, a mutant strain of Chlamydomonas was isolated that showed elevated levels of tolerance to PPO-inhibitor compounds (Kataoka et al., 1990). This strain, *rs-3*, showed a 15-fold increase in the effective dose $(ED)_{90}$ for killing Chlamydomonas with oxadiazon (90µM) and a 10.7-fold increase in the ED_{90} for oxyfluorfen (3.2µM). Later characterization of this mutant showed a single nucleotide alteration in the Protox gene, causing a valine to methionine amino acid change (Randolph-Anderson et al., 1998). Upon discovery of this gene, the Sumitomo Chemical Company (involved in the discovery and characterization of this mutant gene) filed a patent on the gene and its use. This resulted in the lack of availability of the gene for academic purposes. Moreover, no transgenic plants carrying a genetically engineered version of *rs-3* were developed for the marketplace. Expiration of the Sumitomo patent provided an opportunity for development of the *rs-3* gene as a selectable marker and as a potential source of herbicide resistance for algae grown in commercial settings. Finally, analysis of the Chlamydomonas PDS gene shows significant

conservation in amino acid content with the PDS proteins from Hydrilla and the two cyanobacteria noted above. As detailed below, this conservation includes the location of the mutation that confers norflurazon resistance in Hydrilla and Synechocystis. This suggests that a similar mutation in the Chlamydomonas PDS may also be successful in conferring resistance to norflurazon.

The goal of this project was to develop a series of gene constructs conferring resistance to the herbicides glyphosate, oxadiazon/oxyfluorfen and norflurazon that could be used as selectable markers for the genetic transformation of Chlamydomonas and other eukaryotic algae and as a potential tool for maintaining axenic or near-axenic algal cultures in commercial settings.

Results

Analysis of Herbicide Tolerance in Wild Type and Mutant Chlamydomonas Strains

As a test of the effectiveness of various herbicides to kill eukaryotic algae, we subjected four strains of Chlamydomonas to various growth tests in the presence of three different classes of herbicides represented by glyphosate, oxadiazon/oxyfluorfen and norflurazon. The four strains used for this study, CC124, CC1010, CC503, & CC3491 (Chlamydomonas Resource Center), were selected based on their different mating types (mt+, mt-), and their cell wall composition (walled or wall-less). CC124 and CC1010 both possess intact cell walls, and are mt- and mt+, respectively. The other two strains, CC3491 (mt-) and CC503 (mt+), are deficient in producing intact cell walls. Such wall-less strains are used in laboratory research because there is no need to use autolysin

(Buchanan et al., 1989; Harris, 1989) to remove cell walls prior to genetic transformation using either electroporation (Shimogawara et al., 1998) or glass-bead vortexing (Kindle, 1990) techniques. Each algal strain was grown in the presence of multiple concentrations of herbicides in either liquid TAP media or as colonies on Petri dishes containing 1% agar and TAP medium. Results of these tests are presented in Figure 3.2.

All four algal strains showed sensitivity to glyphosate. When pH was controlled (glyphosate can significantly lower media pH) the dose required for complete inhibition of growth ranged from 15 mM to 30 mM in liquid cultures and 7.5mM to 15 mM on plates. An approximate two-fold increase in glyphosate sensitivity was noted in wall-less compared to walled cells. Mating type, conversely, had no significant effect in regard to glyphosate sensitivity. Growth condition also appeared to have an effect on toxicity, as cells showed an increased sensitivity to glyphosate on plates compared with growth in liquid culture.

Two PPO-inhibitors, oxadiazon and oxyfluorfen (Hao et al., 2011), were also effective in preventing growth of all four algal strains, at levels much lower than glyphosate (Figure 3.2). Lethal doses of oxadiazon were identical for all strains (4.0 µM) in liquid media and ranged from 2.5 µM to 5.0 µM in solid media. Oxyfluorfen showed a higher level of toxicity to Chlamydomonas cells than oxadiazon. Oxyfluorfen was effective at roughly 30-fold lower concentrations (lethal doses: 0.07 µM to 0.11 µM in liquid media, 0.06 µM to 0.12 µM in solid media) than oxadiazon. Unlike glyphosate, neither the cell wall nor the growth condition significantly affected the toxicity level of

oxadiazon and oxyfluorfen on cell growth. Mating type again showed little effect in influencing toxicity levels.

Norflurazon was used for study of PDS inhibition in Chlamydomonas. All four Chlamydomonas strains displayed sensitivity to norflurazon, producing a characteristic white phenotype when grown in complete darkness. When grown in the presence of light, the four strains failed to grow on elevated levels of norflurazon in both liquid (1.5-3.5µM) and solid (3.0-4.0µM) media. When each strain was compared in relation to its cell wall composition or mating type, it was noted that in liquid media, wall-less strains were susceptible to significantly lower levels of norflurazon than walled.

Figure 3.1 – Partial alignment of PDS gene sequences across species. Labels correspond to phytoene desaturase associated with each species. PCC7942, *Synechococcus elongatus* PCC7942; PCC6803, *Synechocystis* sp. PCC6803; Hydrilla, *Hydrilla verticillata*; Cre, *Chlamydomonas reinhardtii*. Locations of amino acid mutations that confer herbicide resistance are marked. Substitutions that confer herbicide resistance are as follows; Synechococcus: V403G, Synechocystis: R195C, Hydrilla: R304H, R304S, R304C, or R304T.

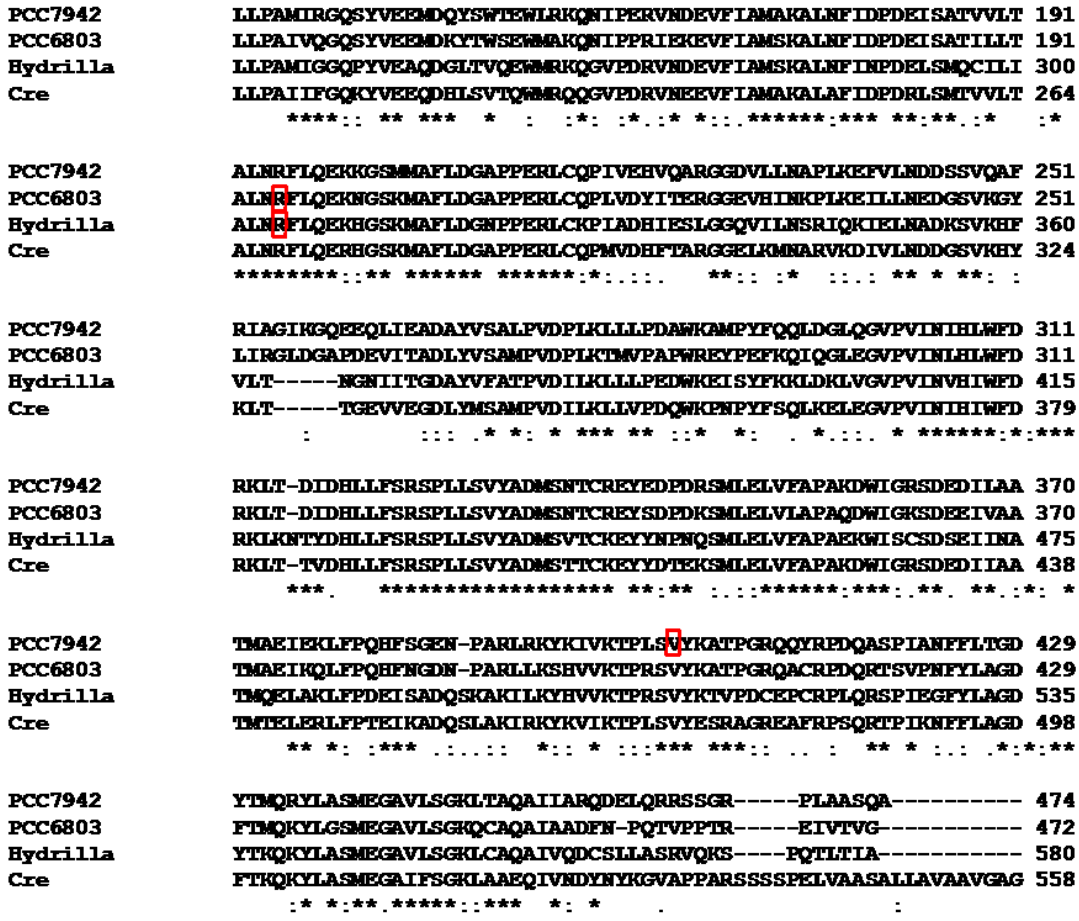


Figure 3.2 – Determination of minimal concentrations of four herbicides required to kill walled and wall-less strains of Chlamydomonas. Glyphosate, two PPO-inhibiting-compounds, oxadiazon and oxyfluorfen, and norflurazon were assayed for toxicity to Chlamydomonas. Four strains of Chlamydomonas were assayed for comparison of herbicide sensitivities: CC3491 (wall-less, mt-), CC503 (wall-less, mt+), CC124 (walled, mt-), CC1010 (walled, mt+).

	Glyphosate (mM)		Oxadiazon (µM)		Oxyfluorfen (µM)		Norflurazon (µM)	
Strain	**Liquid**	**Agar**	**Liquid**	**Agar**	**Liquid**	**Agar**	**Liquid**	**Agar**
CC3491	15.0	7.5	4.0	2.5	0.11	0.06	1.5	3.0
CC503	15.0	7.5	4.0	2.5	0.10	0.12	1.5	4.0
CC124	30.0	15.0	4.0	2.5	0.08	0.12	3.5	4.0
CC1010	30.0	15.0	4.0	5.0	0.07	0.12	2.5	4.0
Range	15.0-30.0	7.5-15.0	4.0	2.5-5.0	0.07-0.11	0.06-0.12	1.5-3.5	3.0-4.0

Herbicide Resistance Constructs and Transformation Efficiencies

The first gene tested for generating glyphosate resistance in Chlamydomonas was the class II EPSPS gene from *Agrobacterium* strain CP4, as it has been used to generate glyphosate resistance in multiple organisms. However, despite numerous attempts and modifications, this gene was found unsuitable for generating glyphosate resistant lines of Chlamydomonas.

The second route explored for creating glyphosate resistance in Chlamydomonas was detoxification of glyphosate via acetylation. The glyphosate acetyltransferase selected, originally recovered from *Bacillus lichenformis* and optimized for activity via DNA shuffling (Castle et al., 2004), was placed under control of the Chlamydomonas PsaD promoter and terminator. In addition, the first intron from Rbcs2 was inserted to improve gene expression (Lumbreras et al., 1998). This construct was labeled pGlyR1 (Figure 3.3). Multiple tests confirmed this gene as able to produce glyphosate resistance in Chlamydomonas.

Construct pGlyR1 showed variable transformation efficiency over a range of glyphosate ratios. When pH of the selection media was controlled (an important prerequisite when the acid form of glyphosate was employed), pGlyR1 achieved a median transformation efficiency of >4000 cfu/ug DNA when cells were selected in the presence of 10 mM glyphosate. Of concern, however, was the ability of wild type cells to also survive at this concentration of glyphosate after a mock transformation (ie. electroporation with plasmid DNA lacking the glyphosate resistance gene). At 10mM

glyphosate, background growth reached a level of 3000 cfu/ug DNA, although these colonies were easy to differentiate from pGlyR1-transformed cells, as they took several days longer to become visible on plates. 40mM glyphosate was sufficient to eliminate growth of any wild type cells, but pGlyR1 transformation recovery was reduced to 420 cfu/ug DNA. A suitable level of glyphosate was determined to be 20mM. At this concentration, transformation with pGlyR1 generated a large number of transformants (1899 cfu/ug DNA) while growth of cells subjected to mock transformations was significantly reduced (10 cfu/ug DNA) (Figure 3.4).

Three constructs were analyzed for generating resistance to PPO-inhibitors: pOxR1, pOxR2, pOxR3 (Figure 3.3). All three constructs contained a fusion of part of the wild type protox gene (including promoters and introns) with a cDNA containing the latter half of the protox gene, which contains the *rs-3* mutation. The three constructs differed only in the length of the promoter (complete details of the constructs can be found in Methods).

All three constructs successfully generated PPO-inhibitor-resistant lines of Chlamydomonas. Similar to pGlyR1 and glyphosate, the OxR constructs showed a gradual decrease in transformation efficiency as the level of herbicide was increased. But, as expected, this was accompanied with a decrease in the number of “background” colonies recovered following transformations with plasmid DNA lacking the oxadiazon/oxyfluorfen-resistance gene. A level of 15 µM oxadiazon was sufficient to eliminate survival of nontransformed cells, while still allowing survival of newly

transformed cells. With oxyfluorfen, however, the amount necessary to prevent survival of nontransformed cells (0.20 µM) also prevented the survival of cells transformed with any of the three OxR constructs. A concentration of 0.17 µM oxyfluorfen eliminated most nontransformed cells (8 cfu/ug DNA) while still allowing survival of transformed cells.

To ascertain which of the three *rs-3* constructs, pOxR1, pOxR2, or pOxR3, possessed the highest efficiency in generating PPO-inhibitor-resistant cells, the three plasmids were compared in regard to their abilities to generate transformants able to survive in the presence of 15 µM oxadiazon and 0.17 µM oxyfluorfen. Each construct was used for transformation of Chlamydomonas CC3491 cells in triplicate experiments and a median transformation efficiency for each construct was ascertained (Figure 3.4). The number of PPO-inhibitor-resistant constructs generated from initial selection on oxadiazon was comparable to the number recovered on oxyfluorfen. All constructs successfully generated PPO-inhibitor-resistant cells, and Analysis of Variance indicated no statistically significant difference in transformation efficiency between the three constructs. Median transformation rates, however, suggest that pOxR3, which contains the shortest of the three promoter lengths (139bp) may be less efficient in generating oxadiazon- and oxyfluorfen-resistant cells than the other two constructs. In the presence of oxyfluorfen, pOxR3 generated 338 cfu/µg DNA, whereas pOxR1 and pOxR2 generated 556 and 858 cfu/µg DNA, respectively (Figure 3.4). This pattern of transformation efficiency and was also seen in the presence oxadiazon. Electroporation

of Chlamydomonas often results in fragmentation and partial degradation of the foreign DNA prior to its integration into the genome. It is possible, then, that the additional promoter length found in pOxR1 and pOxR2 acts as a buffer during the electroporation event, resulting in a higher number of integrated transgenes containing the requisite minimal promoter length.

Chlamydomonas phytoene desaturase shares significant amino acid sequence identity with PDSs from several organisms, including *Hydrilla*, *Synechococcus elongatus* PCC7942, and *Synechocystis* sp. PCC6803. – all of which can be rendered insensitive to norflurazon by specific single amino acid substitutions. The location of two of these mutations, *Synechocystis* R195 and *Hydrilla* R304, occur in a highly conserved region of the protein. When aligned with Chlamydomonas PDS (Figure 3.1), significant sequence similarity can be seen in this particular region, including the aforementioned arginine that was mutated in *Synechocystis* and *Hydrilla*. To test if a similar substitution would confer norflurazon resistance to Chlamydomonas, we modified the cDNA of PDS to encode Thr at position 268 in place of the conserved Arg. The modified cDNA was then placed under the control of the Chlamydomonas PsaD promoter and terminator, and the first intron of Rbcs2 was inserted into the cDNA sequence to enhance expression of the transgene (Lumbreras et al., 1998). The resultant plasmid was labeled pNorR1 (Figure 3.3).

When delivered into Chlamydomonas cells by electroporation, pNorR1 generated numerous norflurazon-resistant colonies (Figure 3.4). However, in initial

experiments it was noted that Chlamydomonas cells undergoing transformations with plasmid DNA lacking the mutant PDS gene often produced several colonies on norflurazon-containing plates. We determined a variable lighting pattern to be most effective in generating the highest number of transformed cells while limiting growth of wild type cells. Following a 24 hour recovery, transformed cells were plated onto TAP media containing 4 uM norflurazon and incubated at 25°C under reduced light (5 µmol photons s^{-1} m^{-2}) for 7 days. Following this incubation, the plates were placed under direct light (50-100 µmol photons s^{-1} m^{-2}) for an additional 11 days. Under these conditions, pNorR1 generated 550 cfu/µg DNA (Figure 3.4). Screening of these colonies via PCR confirmed that 90% harbored the PDS transgene. The 10% of colonies that did not show an intact transgene were then assayed to determine whether the norflurazon^{R} phenotype was permanent and heritable. These colonies were resuspended in TAP media and spotted to TAP plates containing 0 and 4 µM norflurazon. None of the non-transformed lines assayed showed continued viability on norflurazon. In contrast, colonies containing the pNorR1 transgene showed robust growth when spotted to TAP containing 4 µM norflurazon. These results indicate that the growth of non-transformed cells following electroporation can be managed with a variable lighting pattern, and any non-transformed cells can be eliminated with a secondary screen on new TAP + norflurazon plates.

Figure 3.3 - Herbicide-Resistance Constructs. GlyR, OxR, and NorR constructs were used to generate resistance to glyphosate, oxadiazon/oxyfluorfen, and norflurazon, respectively. The following gene components are labeled as follows: PsaDP/PsaDT(promoter and terminator regions of Chlamydomonas PsaD), I1(first intron from Chlamydomonas Rbcs2), GAT(synthetic glyphosate acetyltransferase cDNA, divided into parts 1/2 and 2/2), PDS-mut(Chlamydomonas PDS cDNA, containing R268T mutation, divided into parts 1/2 and 2/2), 5'/Ex1/In1/Ex2/In2/Ex3(5' UTR, exon 1, intron 1, exon 2, intron 2, and exon 3 of Chlamydomonas protox gene), Protox *rs-3*(2/2)(second half of Chlamydomonas protox cDNA, containing *rs-3* mutation), VTE5 Ex1/VTE5 5'/Pro(exon 1 and 5' UTR for Chlamydomonas VTE5 gene, and shared promoter sequence separating VTE5 and protox genes in Chlamydomonas

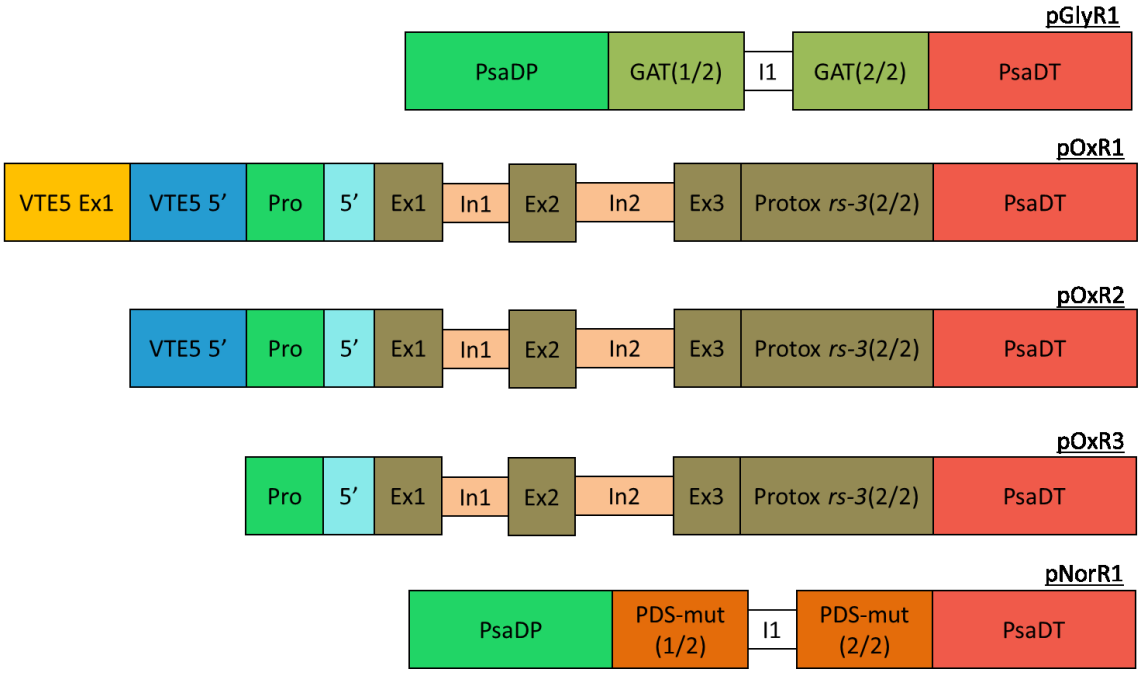


Figure 3.4 – Transformation efficiency of herbicide-resistance genes. Transformation efficiency was reported as the minimum, median, and maximum number of colony forming units (cfu) per µg of plasmid DNA used. Background indicates the number of cfu/µg that was false positives (ie. nontransformed cells).

			Transformation Efficiency			
Plasmid	**Gene**	**Herbicide Selection**	**Minimum**	**Median**	**Maximum**	**Background**
pGlyR1	GAT	Glyphosate (20mM)	440	1899	4160	10(0.5%)
pOxR2	protox rs-3	Oxyfluorfen (0.17µM)	212	858	2120	8(0.9%)
pNorR1	PDS (R268T)	Norflurazon (4µM)	500	550	700	55(10.0%)
pSP124s	Bleomycin	Zeocin (7.5mg/L)	20	796	1764	-

Level of Herbicide Resistance Imparted by Transgenes

After generating various herbicide-resistant lines of Chlamydomonas, we analyzed these lines to determine the maximum levels of herbicide to which they are tolerant. Tests were conducted both in liquid media and on agar plates. For liquid media, cell growth was monitored daily by counting cells with a hemocytometer. For growth on agar, standardized numbers of cells were spotted in serial dilutions onto agar plates containing various levels of herbicide, and plates were monitored visually for growth.

Glyphosate resistance: For tests on solid media, multiple transgenic lines were spotted onto agar plates containing various levels of glyphosate and assayed for their growth. We found that all pGlyR1-transformed lines tested were able to survive plating on levels of glyphosate up to 40 mM in both liquid and agar media (maximum solubility of glyphosate in water is 50 mM) (Figure 3.5). However, several lines showed reduced rates of growth on agar media as glyphosate levels were increased. Liquid tests of transgenic lines showed similar results. Notably, while the transgene appeared to have only a minimal effect on growth of the algae in liquid TAP media (0.48% increase in doubling time compared to wild type), the cells show a 51% increase in required doubling time at levels of glyphosate necessary to prevent growth of the WT progenitor strains (CC3491). At the level required to eliminate all growth of wild-type algae (30 mM), transgenic lines show a 180% increase in doubling time, illustrating the ability of

the GAT gene to impart glyphosate resistance to Chlamydomonas cells, but not eliminate deleterious effects in growth.

Oxyfluorfen resistance: All three OxR constructs, although displaying differential transformation efficiencies, behaved similarly when tested for maximum herbicide tolerances. Tests of the transgenic lines on agar plates were conducted with both oxadiazon and oxyfluorfen. For oxadiazon, transgenic lines of pOxR1, pOxR2, and pOxR3 were able to survive on levels up to 750 µM in liquid media and 1.5mM on agar, representing a 188-600 fold increase compared to wild type cells. These lines were also able to survive on levels of oxyfluorfen up to 5.0 µM on agar and 15 µM in liquid media, an approximately 83-136 fold increase compared to wild type cells. Quantitative growth tests in liquid media were limited to oxyfluorfen (Figure 3.5). All lines showed similar levels of tolerance to oxyfluorfen. Because it showed the highest transformation efficiency of the three constructs, only pOxR2 will be discussed in this section. Compared to the WT progenitor (CC3491), pOxR2 transgenic lines showed no significant differences in their rate of growth in non-selective media. These transgenic lines did, however, show a reduction in their rate of growth when oxyfluorfen was added to the media, although this reduction was substantially less than the reduction observed with pGlyR1-transformed cells grown in the presence of glyphosate. At 1.5 µM oxyfluorfen, lines showed 177% increases in doubling times, and at 5.0 µM, growth of transgenic lines was nearly halted. However, at 0.11 µM oxyfluorfen (the maximum lethal dose for all Chlamydomonas strains tested), transgenic lines showed only moderate increases in

doubling times (10.3%), illustrating the potential for oxyfluorfen to maintain axenic cultures while only slightly impacting growth of the transgenic line.

Norflurazon resistance: Cells transformed with the pNorR1 plasmid showed a 40x increase in tolerance to norflurazon (lethal doses of 60 µM in liquid media and 120 µM on solid media). When compared to CC3491, the transgene imparted minor deleterious effects on growth. Transgenic pNorR1-containing lines showed an average increase of ~17% in doubling time compared to wild type cells, though statistical analysis showed this increase to not be statistically significant. Like prior experiments with other herbicide-resistant genes, transgenic cells containing pNorR1, while able to survive in significantly higher levels of norflurazon than wild type strains, did experience reductions in their rate of growth (Figure 3.5). At the level of herbicide necessary to kill the WT progenitor cell line (CC3491), transgenic cells showed only a 2.6% increase in doubling time compared to transgenic cells in TAP media. At 3.5µM, the level of norflurazon necessary to kill off all strains of Chlamydomonas, transgenic cells exhibited a 12.9% increase in doubling time. To be noted, however, was the propensity of norflurazon to promote growth of fungi in our media. Tests with norflurazon were routinely forced to be repeated due to fungal contamination, and empirical observations suggested that norflurazon actually promoted fungal growth, possibly serving as a carbon source.

Figure 3.5 – Growth analysis of transgenic lines. Transgenic lines were assayed for their ability to survive in increasing levels of herbicide in both liquid and solid (agar) media. Lethal dose "increase" indicates the tolerance of transgenic lines to their respective herbicides in comparison to nontransformed CC3491. Transgenic lines were also assayed for their rate of growth in the presence of increasing levels of herbicide. The rate of growth of transgenic lines in was analyzed by comparing the doubling time (amount of time required for the cell density to double) of cells in TAP to that in the presence of herbicide necessary to kill WT cells. Doubling Time "increase" describes the additional amount of time required for the cells to grow in the presence of herbicide compared to growth in TAP

		Lethal Dose (Liquid Media)			Lethal Dose (Solid Media)		
Plasmid	Herbicide Resistance	Wild Type	with Transgene	Increase	Wild Type	with Transgene	Increase
pGlyR1	Gly^R	15 mM	40 mM	2.7x	7.5 mM	40 mM	5.3x
pOxR2	Oxy^R	0.11 µM	15 µM	136x	0.06 µM	5 µM	83x
pNorR1	Nor^R	1.5 µM	60 µM	40x	3.0 µM	120 µM	40x

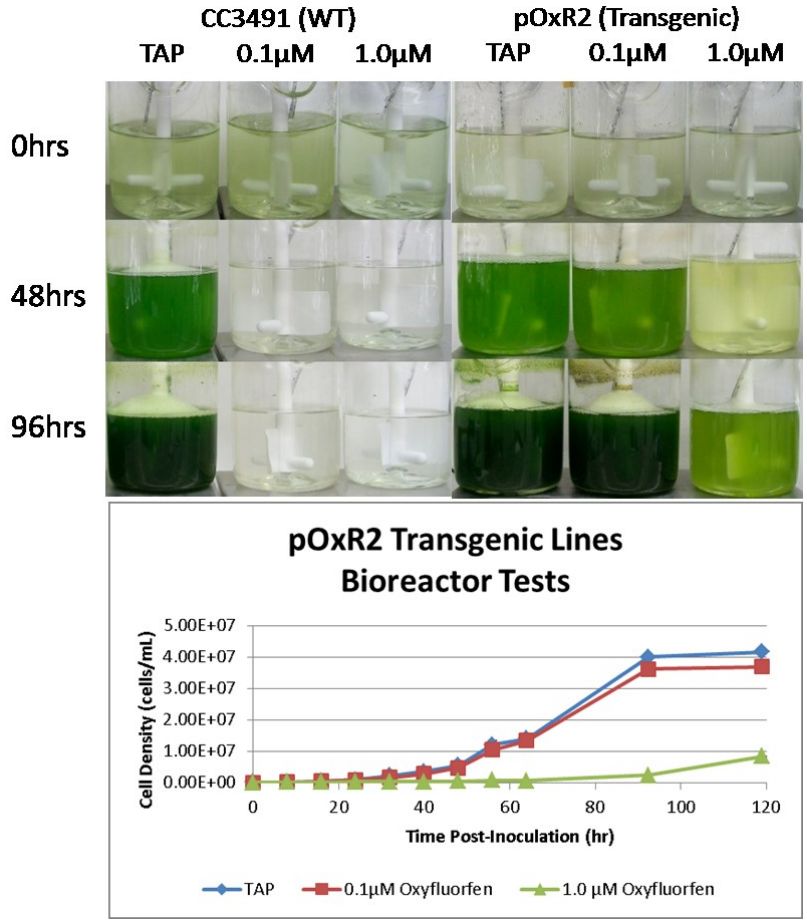
		Doubling Time (Liquid Media)		
Plasmid	Herbicide Resistance	TAP	At WT Lethal Dose	Increase
pGlyR1	Gly^R	9.17	13.89	51.5%
pOxR2	Oxy^R	8.66	9.54	10.3%
pNorR1	Nor^R	8.98	9.21	2.6%

Large-Scale Growth of Oxyfluorfen-Resistant Cells

To examine the potential application of herbicides in maintaining axenic cultures in large-scale algal production, we analyzed the growth of wild type and transgenic lines of Chlamydomonas in a bioreactor setting. Analysis of pOxR2-transformed cells was selected, as these lines exhibited nearly normal growth rates when cultured in the presence of oxyfluorfen at concentrations sufficient to kill wild type algae.

Levels of 0.0, 0.1 and 1.0 µM oxyfluorfen were tested in the bioreactors. At 0.1 µM oxyfluorfen, transgenic cells grew at rates nearly identical to that in non-selective media (after 120 hours, cells in non-selective media had reached stationary phase growth, and culture density of cells grown in 0.1 µM oxyfluorfen were 89% of those in non-selective media)(Figure 3.6). Transgenic cells grew at 1.0 µM, as expected from earlier tests, but growth was visibly reduced, and cell counts indicated a nearly 80% reduction in cell density compared to herbicide-free media. It should be noted that CC3491 was unable to grow in the presence of 0.1 µM oxyfluorfen during the bioreactor tests, suggesting 0.1 µM oxyfluorfen is completely adequate to maintain axenic cultures while avoiding the deleterious effects of higher herbicide concentrations on growth of the transgenic line.

Figure 3.6 –Large-Scale Growth Tests with pOxR2-transformed Oxyfluorfen Resistant Chlamydomonas. pOxR2-transformed lines were grown alongside their WT progenitor strain (CC3491) to examine large-scale growth. All lines were grown in TAP as well as in the presence of two levels (0.1 µM and 1.0 µM) of oxyfluorfen.



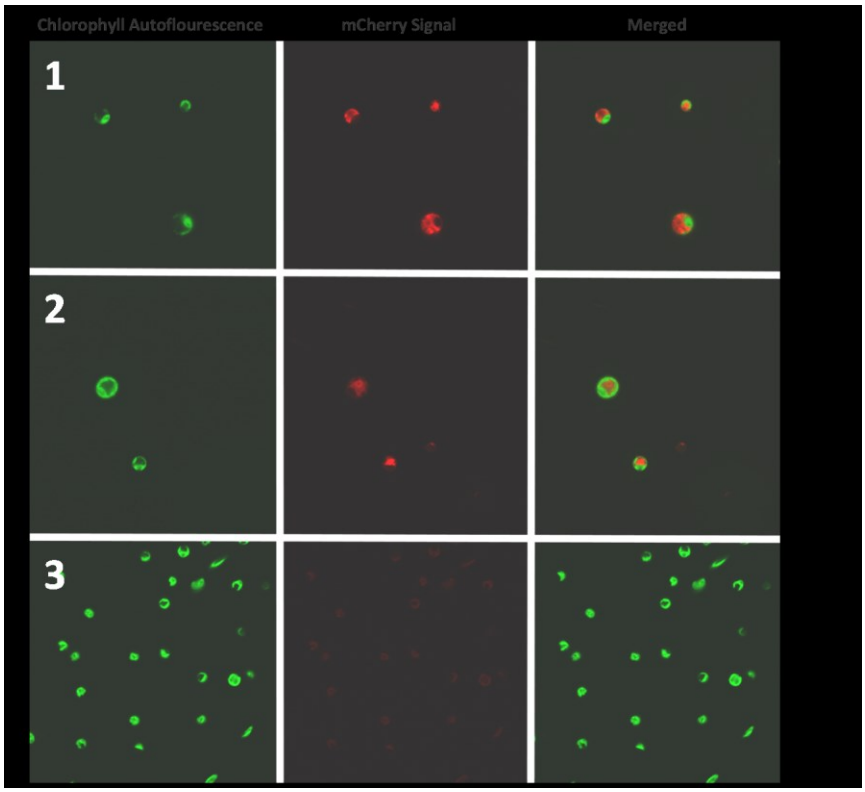
Efficiency of Adjacent Gene Recovery

To assess the ability of the herbicide-resistant markers to serve as a vehicle for recovery and expression of adjacently located genes of interest (GOI) in Chlamydomonas, constructs were built placing a GOI (in this case, a fusion of Ble (blecomycin) with the gene encoding the fluorescent protein, mCherry) adjacent to a herbicide-resistance gene. Transformants were initially selected on media containing either the herbicide or zeocin (for which the Ble gene confers resistance), and spotted to media containing both the herbicide and zeocin. From this, we were able to ascertain the frequency of transformation events in which both genes, the herbicide-resistance gene and Ble, were co-integrated into the cell. For constructs containing Ble adjacent to GAT (construct pGATBAM), we witnessed an average co-transformation of ~37% (i.e., of colonies initially selected on one chemical, 37% also showed cross-resistance to the other chemical). With the protox *rs-3* mutant (construct pRSBAM), the rate of co-transformation was lower, averaging ~13%. The PDS mutant in pNorR1 was not used in this study. For verification of these results, we analyzed zeocin^R/herbicide^R cells using a confocal microscope to detect expression of mCherry fluorescence (Figure 3.7). Utilizing confocal microscopy, we were able to detect mCherry fluorescent in our transformed cells. mCherry signal was comparable to that of a positive control (construct KV1, containing the Ble-mCherry fusion).

As an alternative method of adjacent selection, we coupled two herbicide-resistance genes, GAT and protox *rs-3*, to mCherry via the foot-and-mouth-disease

(FMDV) 2A peptide. This peptide has been shown to function in multiple organisms, including Chlamydomonas, to facilitate a self-cleaving reaction during translation of the polypeptide (Donnelly et al., 2001; Rasala et al., 2012). The result is production of two separate and functional proteins from a single mRNA. The 2A technology thus avoids issues associated with producing functional fusion proteins, while maintaining the convenience of single gene transcription. Unfortunately, we were unable to generate herbicide-resistant transgenic lines using these constructs, suggesting these genes are incompatible with the 2A peptide.

Figure 3.7 – Use of herbicide-resistance genes in adjacent gene selection. Herbicide-resistance genes were inserted adjacent to a Bleomycin-mCherry fusion gene on plasmids pGATBAM and pRSBAM. Transformants were initially selected on media containing either the appropriate herbicide (pGATBAM: glyphosate, pRSBAM: oxyfluorfen) or zeocin (resistance conferred by the Bleomycin resistance gene). Transformants were then transferred to media containing both inhibitors to determine the percentage of transformants containing both genes (gene coupling). Positive cell lines were then analyzed for mCherry expression. Row 1 and 3 represent positive and negative controls for mCherry expression, and row 2 represents one of the cell lines selected transformed with pRSBAM and selected on oxyfluorfen.



Discussion

Use of GAT, protox *rs-3*, and mutant PDS genes for genetic transformation of Chlamydomonas resulted in successful recovery of herbicide resistance cells. The GAT gene (construct pGlyR1) generated higher numbers of transgenic cells following electroporation (1899 cfu/µg DNA) than protox *rs-3* (construct pOxR2) (858 cfu/µg DNA, respectively) or the mutant form of PDS (construct pNorR1) (550 cfu/µg DNA), suggesting it may serve as a better selectable marker in laboratory research. However, the GAT gene was unable to completely eliminate the toxic effects of glyphosate as evidenced in the markedly reduced growth rates of transgenic cell growth in glyphosate-containing media.

Overall, the protox *rs-3* mutant gene perhaps proved the most favorable for selecting genetically transformed Chlamydomonas and for growing transgenic Chlamydomonas in bioreactors. Transformation efficiency with this gene was comparable to that of other antibiotic-based selectable markers routinely used Chlamydomonas research laboratories. For instance, use of the plasmid pSP124s, which confers resistance to zeocin, is routinely used in our laboratory and generates zeocin-resistant colonies at rates of 796 cfu/µg DNA. While the number of oxyfluorfen-resistant colonies generated from this transgene was less than the number generated by GAT gene, the resultant colonies were more uniform in size and easy to isolate for subsequent analyses. Small and mid-scale tests also suggest the potential for this gene to serve as a useful tool in maintenance of axenic algal cultures in large-scale production

facilities. Low dosages of oxyfluorfen (0.1 µM) were sufficient in preventing growth of wild type cells while imparting only minor effects in the growth of transgenic algae, making it an attractive option for use in large-volume algal ponds or raceways. Norflurazon, and use of the pNorR1 construct to confer resistance, while exhibiting lower inhibition of growth rate on transgenic cells compared to oxyfluorfen and pOxR2, is a less ideal candidate, as it appears to promote fungal growth within the cultures.

Methods and Materials

Chemicals

All algal cell growth experiments were performed in Tris Acetate Phosphate (TAP) media (Harris, 1989). Glyphosate was procured from Shanghai Majin (China). Additional herbicides (oxadiazon, oxyfluorfen, norflurazon) were procured from Chem-Service, Inc. (West Chester, PA, USA).

Chlamydomonas culture growth and maintenance

Four strains of Chlamydomonas were used for experiments: CC3491, CC503 (both lacking a cell wall), CC124, and CC1010 (both possessing a cell wall). Strains were maintained on TAP media (Harris, 1989) containing 20 g/L agar at 50 µmoles photons s^{-1} m^{-2}. For most experimental procedures, strains were inoculated into liquid cultures of TAP medium and maintained with shaking at 100 RPM under light at 200 µmoles photons s^{-1} m^{-2}.

Algal Growth Tests

Transgenic and wild-type lines of Chlamydomonas were inoculated in 100 µL volumes of TAP media (or TAP media supplemented with herbicide for transgenic lines) and allowed to grow for 48 hours at 200 µmoles photos s^{-1} m^{-2} lighting. For spot testing on agar plates, cells were washed in TAP, and a 3µL aliquot of three dilutions (333 cells/µL, 33 cells/µL, and 3 cells/µL) was spotted onto plates containing various levels of herbicide. Each line was also spotted onto TAP plates containing no herbicide as a measurement of maximum growth rate. Plates were incubated at 25°C under 200 µmoles photons s^{-1} m^{-2} lighting until colonies could be observed, usually 4-8 days. Growth of cell lines on herbicide-containing plates was compared with their growth on herbicide-free plates to estimate growth rate. For liquid tests, cell densities were normalized, and an equivalent amount of cells were dispensed into 100 µL or 3 mL TAP media volumes containing various levels of herbicide, or herbicide-free TAP media as a control. Cultures were grown with shaking under identical growth conditions as plates. Cell growth rates were calculated by comparing cell density of lines in herbicide-containing media to cell densities in herbicide-free media. Cell density was determined with a hemocytometer.

Construct Designs - GAT and EPSPS Gene Constructs

Two genes, the synthetic glyphosate acyltransferase (GAT) (PMID: 15155947) and 5-enolpyruvylshikimate-3-phosphate synthase (EPSPS) from *Agrobacterium* sp. strain CP4 (PMID: 8598558) were synthesized by Genscript (NJ) to be codon-optimized for expression in Chlamydomonas. The first intron of the Chlamydomonas Rbcs2 gene

was incorporated into both genes to act as an enhancer element for expression (Lumbreras et al., 1998). The genes were then placed individually under the control of the Chlamydomonas PsaD promoter and terminator. This final construction of the GAT gene was named pGlyR1. The EPSPS gene was further modified by being fused to two different 36nt chloroplast-targeting transit peptides. The two transit peptides were isolated from the Chlamydomonas genes, Rubisco activase and ferredoxin (Leon et al., 2007), leading to the final two constructs, pHRTP20 and pHRTP22, respectively.

Construct Designs - Protox Gene Constructs

For expression of the *rs-3* mutant form of protox, a hybrid expression cassette was constructed, using portions of the wild type and *rs-3* mutant protox gene. Chlamydomonas protox (Cre09.g396300) cDNA was synthesized by Genscript (NJ) containing the 1 bp mutation specific to the *rs-3* mutant. The last 1124 nucleotides of this cDNA, which include the point mutation conferring resistance to PPO-inhibitors, was ligated to a 5' portion of the wild type protox gene. This portion included the first two exons and introns of the protox gene, and a portion of the third exon. Introns were included in this construct to enhance expression of the transgene. The gene fusion was placed under the control of the endogenous protox promoter and the PsaD terminator. In the Chlamydomonas genome, protox is arranged in a head-to-head orientation with the gene VTE5, with the 5' UTRs of the two genes separated by 139nt, presumably functioning as a transcriptional promoter for both genes. Upstream segments of 139, 304, and 452nt were isolated, representing the promoter, VTE5 5' UTR, and 1st VTE5

exon, respectively. These fragments were incorporated into three constructs, each representing a variable promoter length. The final constructs contained the following components: “pOxR3” - 139nt promoter, “pOxR2” - 443nt promoter, containing the VTE5 5’ UTR, “pOxR1” - 895nt promoter, containing the 1st exon and 5’ UTR of VTE5.

Construct Designs - PDS Gene Constructs

The cDNA encoding PDS from Chlamydomonas (Cre12.g509650) was amplified from a cDNA pool using primers designed from the cDNA sequence, adding NdeI and EcoRI sites to the 5’ and 3’ ends, respectively. The resultant PCR product was further modified. First, a 2 bp change was made to the nucleotides 802 and 803 (CG to AC), resulting in an amino acid substitution at position 268 (Arg to Thr). NdeI and EcoRI sites allowed for directional ligation of the cDNA into a plasmid containing the promoter and terminator of the Chlamydomonas PsaD gene. In a similar fashion to the GAT gene, the first intron from the Chlamydomonas Rbcs2 gene was integrated into the GAT gene to enhance gene expression (Lumbreras et al., 1998). The resultant plasmid was labeled “pNorR1”.

Transformations

Chlamydomonas cells were transformed with DNA following the electroporation protocol outlined by Shimogawara et al. (1998). Briefly, cells were grown in TAP medium to a density of 1-3x10^{6} cells/mL. Cells were harvested via centrifugation and resuspended in TAP media containing 60mM sucrose to a density of 4x10^{8} cells/mL. An aliquot of 1x10^{8} cells was combined with 1 μg of the plasmid DNA of interest in a 0.4cm

electroporation cuvette (BioExpress, www.BioRad.com). Following incubation of the cuvette in a 16°C water bath for 5 minutes, each sample was pulsed at 0.75kv, 25uF, and no resistance using a Gene Pulser II (www.BioRad.com). Cells were recovered in 10 mL of TAP+60mM sucrose for 24 hours with low light (100 µmol photons m^{-2} s^{-1}) and minimal shaking. Transformants were then plated on various TAP plates containing the desired types and levels of herbicide. Plates were incubated for 6-10 days under 200 µmol photons m^{-2} s^{-1} lighting to allow herbicide-resistant colonies to develop. Of exception were pNorR1-transformed cells, which were incubated for 7 days under shaded conditions (5 µmol photons m^{-2} s^{-1}) and then transferred to higher light (50-100 µmol photons m^{-2} s^{-1}) for an additional 11 days.

Bioreactor Growth Tests

Transgenic and wild-type Chlamydomonas were grown as seed cultures in 500 mL Erlenmeyer flasks to mid-log phase (~$1x10^6$ cells/mL). To prevent transgene silencing, transgenic strains were grown in the presence of the minimal amount of herbicide necessary to kill nontransgenic wild type cells. Upon reaching the desired culture density, cells were pelleted via centrifugation (2,000xg, 5 minutes). The cell pellet (containing ~$2x10^8$ cells) was washed with TAP and repelleted, and the final pellet resuspended in a 1L bioreactor (BellCo, #1965-81005) containing TAP media and a preselected amount of herbicide. Algal cultures in the bioreactors were then illuminated at 200 µmol photons m^{-2} s^{-1}. Cultures were agitated by a combination of stirring (100 rpm) and bubbling with filtered air (6.5 units, monitored using a flow meter - Cole

Palmer PMR1-010285). Cell density was measured every 8 hours using a hemocytometer.

Determinations of Lethal Doses and Doubling Time Increases

Lethal dosage of each herbicide was determined for each strain (wild type and transgenic) tested. Strains were grown in various ranges of herbicide, and monitored visually for growth. The lethal dose was determined as the median concentration of herbicide that completely prevented growth of the algal strain. For quantification of the herbicidal effects on growth rate, cell growth was measured via cell counting with a hemocytometer until the cells reached late log phase growth (~$1x10^7$ cells/mL). An average doubling time was then determined for these cells, using the following equation: $T_D = \Delta T*(\log(2)/\log(C_2/C_1))$, where ΔT = the amount of time between cell counts, and C_1 and C_2 are the cell densities for the 1^{st} and 2^{nd} counts, respectively. Doubling times for cells in each level of herbicide were compared to determine the impact of herbicide on cell growth rate.

Acknowledgements

I would like to thank several individuals for their contributions in this work. Critical intellectual contributions were made by Dr. Donald Weeks in both the conceptualization and designs for much of this work. Dr. Kempton Horken provided invaluable expertise throughout this project, particularly in providing information for transformation of Chlamydomonas and subsequent screening procedures. For adjacent selection experiments, I would like to thank Thomas Plucinak for providing plasmids

containing the bleomycin-myCherry fusion genes. Daniel Kuehler and Andrew Blazek both provided assistance in the execution of several experiments and in analysis of data sets throughout the duration of this project. Financial support for this project was provided by the Department of Energy JGI (Award JGI 2010 CSP-146), National Science Foundation (grants MCB-0952533 and EPSCoR-1004094) and the Consortium for Algal Biofuels Commercialization (Department of Energy Award DE-EE0003373).

References

Arias, R.S., Dayan, F.E., Michel, A., Howell, J., and Scheffler, B.E. (2006). Characterization of a higher plant herbicide-resistant phytoene desaturase and its use as a selectable marker. Plant Biotechnol J **4,** 263-273.

Barry, G.F., Kishore, G.M., Padgette, S.R., and Stallings, W.C. (1994). Glyphosate-Tolerant 5-Enolpyruvylshikimate-3-phosphate Synthases (United States).

Barry, G.F., Kishore, G.M., Padgette, S.R., Taylor, M.L., Kolacz, K.H., Weldon, M., . . . Hallas, L. (1992). Inhibitors of amino acid biosynthesis: strategies for imparting glyphosate tolerance to crop plants. In Biosynthesis and Molecular Regulation of Amino Acids in Plants, B.K. Singh, H.E. Flores, and J.C. Shannon, eds (Mockville, MD: American Society of Plant Physiologists), pp. 139-145.

Beale, S.I., and Weinstein, J.D. (1990). Tetrapyrrole metabolism in photosynthetic organisms. In Biosynthesis of Heme and Chlorophylls, H.A. Dailey, ed (New York: McGraw-Hill), pp. 287-391.

Böger, P., and Sandmann, G. (1998). Carotenoid biosynthesis inhibitor herbicides - mode of action and resistance mechanisms. Pesticide Outlook **9,** 29-35.

Buchanan, M.J., Imam, S.H., Eskue, W.A., and Snell, W.J. (1989). Activation of the cell wall degrading protease, lysin, during sexual signalling in Chlamydomonas: the enzyme is stored as an inactive, higher relative molecular mass precursor in the periplasm. J Cell Biol **108,** 199-207.

Castle, L.A., Siehl, D.L., Gorton, R., Patten, P.A., Chen, Y.H., Bertain, S., . . . Lassner, M.W. (2004). Discovery and directed evolution of a glyphosate tolerance gene. Science **304,** 1151-1154.

Chamovitz, D., Pecker, I., and Hirschberg, J. (1991). The molecular basis of resistance to the herbicide norflurazon. Plant Molecular Biology **16,** 967-974.

Dill, G.M., Cajacob, C.A., and Padgette, S.R. (2008). Glyphosate-resistant crops: adoption, use and future considerations. Pest Manag Sci **64,** 326-331.

Donnelly, M.L., Hughes, L.E., Luke, G., Mendoza, H., ten Dam, E., Gani, D., and Ryan, M.D. (2001). The 'cleavage' activities of foot-and-mouth disease virus 2A site-directed mutants and naturally occurring '2A-like' sequences. J Gen Virol **82,** 1027-1041.

Duke, S.O., Lydon, J., Becerril, J.M., Sherman, T.D., Lehnen, L.P., and Matsumoto, H. (1991). Protoporphyrinogen Oxidase-Inhibiting Herbicides. Weed Science **39,** 465-473.

Ha, S.B., Lee, S.B., Lee, Y., Yang, K., Lee, N., Jang, S.M., . . . Back, K. (2004). The plastidic Arabidopsis protoporphyrinogen IX oxidase gene, with or without the transit sequence, confers resistance to the diphenyl ether herbicide in rice. Plant Cell and Environment **27,** 79-88.

Hao, G.F., Zuo, Y., Yang, S.G., and Yang, G.F. (2011). Protoporphyrinogen Oxidase Inhibitor: An Ideal Target for Herbicide Discovery. Chimia **65,** 961-969.

Harris, E.H. (1989). The Chlamydomonas sourcebook : a comprehensive guide to biology and laboratory use. (San Diego: Academic Press).

Jung, S., Lee, Y., Yang, K., Lee, S.B., Jang, S.M., Ha, S.B., and Back, K. (2004). Dual targeting of Myxococcus xanthus protoporphyrinogen oxidase into chloroplasts and mitochondria and high level oxyfluorfen resistance. Plant Cell and Environment **27,** 1436-1446.

Kataoka, M., Sato, R., and Oshio, H. (1990). Isolation and partial characterization of mutant Chlamydomonas reinhardtii resistant to herbicide S-23142. Journal of Pesticide Science **15,** 449-451.

Kindle, K.L. (1990). High-frequency nuclear transformation of Chlamydomonas reinhardtii. Proc Natl Acad Sci U S A **87,** 1228-1232.

Leon, R., Couso, I., and Fernandez, E. (2007). Metabolic engineering of ketocarotenoids biosynthesis in the unicelullar microalga Chlamydomonas reinhardtii. J Biotechnol **130,** 143-152.

Lermontova, I., and Grimm, B. (2000). Overexpression of plastidic protoporphyrinogen IX oxidase leads to resistance to the diphenyl-ether herbicide acifluorfen. Plant Physiol **122,** 75-84.

Lumbreras, V., Stevens, D.R., and Purton, S. (1998). Efficient foreign gene expression in Chlamydomonas reinhardtii mediated by an endogenous intron. Plant Journal **14,** 441-447.

Martinez-Ferez, I.M., and Vioque, A. (1992). Nucleotide sequence of the phytoene desaturase gene from Synechocystis sp. PCC 6803 and characterization of a new mutation which confers resistance to the herbicide norflurazon. Plant Molecular Biology **18,** 981-983.

Michel, A., Arias, R.S., Scheffler, B.E., Duke, S.O., Netherland, M., and Dayan, F.E. (2004). Somatic mutation-mediated evolution of herbicide resistance in the nonindigenous invasive plant hydrilla (Hydrilla verticillata). Mol Ecol **13,** 3229-3237.

Monsanto. (2005). Backgrounder - History of Monsanto's Glyphosate Herbicides.

Padgette, S.R., Kolacz, K.H., Delannay, X., Re, D.B., LaVallee, B.J., Tinius, C.N., . . . Kishore, G.M. (1995). Development, Identification, and Characterization of a Glyphosate-Tolerant Soybean Line. Crop Sci. **35,** 1451-1461.

Randolph-Anderson, B.L., Sato, R., Johnson, A.M., Harris, E.H., Hauser, C.R., Oeda, K., . . . Boynton, J.E. (1998). Isolation and characterization of a mutant protoporphyrinogen oxidase gene from Chlamydomonas reinhardtii conferring resistance to porphyric herbicides. Plant Molecular Biology **38,** 839-859.

Rasala, B.A., Lee, P.A., Shen, Z., Briggs, S.P., Mendez, M., and Mayfield, S.P. (2012). Robust expression and secretion of Xylanase1 in Chlamydomonas reinhardtii by

fusion to a selection gene and processing with the FMDV 2A peptide. PLoS One **7,** e43349.

Reboud, X. (2002). Response of Chlamydomonas reinhardtii to herbicides: negative relationship between toxicity and water solubility across several herbicide families. Bull Environ Contam Toxicol **69,** 554-561.

Shimogawara, K., Fujiwara, S., Grossman, A., and Usuda, H. (1998). High-efficiency transformation of Chlamydomonas reinhardtii by electroporation. Genetics **148,** 1821-1828.

Smith, E.A., and Oehme, F.W. (1992). The biological activity of glyphosate to plants and animals: a literature review. Vet Hum Toxicol **34,** 531-543.

Warabi, E., Usui, K., Tanaka, Y., and Matsumoto, H. (2001). Resistance of a soybean cell line to oxyfluorfen by overproduction of mitochondrial protoporphyrinogen oxidase. Pest Manag Sci **57,** 743-748.

Williams, G.M., Kroes, R., and Munro, I.C. (2000). Safety evaluation and risk assessment of the herbicide Roundup and its active ingredient, glyphosate, for humans. Regul Toxicol Pharmacol **31,** 117-165.

CHAPTER 4 - Synthesis of Custom-Designed TAL Effector Nuclease (TALEN) Genes for Targeted Gene Modifications in *Chlamydomonas reinhardtii*

Abstract

We have developed a method of constructing custom-designed TAL effector nuclease (TALEN) proteins in our lab for studies in the green alga *Chlamydomonas reinhardtii*. Our interest in this project was to create a laboratory-based system of arranging the 34-AA 'repeat' units of a TALEN in a predetermined fashion so that we could synthesize TALENs to target any selected DNA sequence. To accomplish this goal, we generated a "repeat library", comprised of repeat units containing RVDs to bind one of four nucleotides (A, C, G, or T) and modified with Type IIs restriction sites flanking the 5' and 3' ends of the repeat. By utilizing the degeneracy of the genetic code, we were able to create repeat units that all coded for identical proteins, but varied in the nucleotide composition of their first and last (1^{st} and 34^{th}) codons. When digested with the Type IIs enzymes (BsaI and BsmBI), the repeat units possessed unique nucleotide overhangs that allowed for directional ligation of eight repeat units into an ordered eight repeat (8-mer) set. 8-mers were then combined to form an intact, 24-repeat, TALEN, utilizing the N- and C-termini of the endogenous TALE AvrXa7 as a "shell" for insertion of the repeats. Pairs of TALENs were designed to bind 24.5bp DNA sequences (effector binding elements, EBEs) on opposite DNA strands separated by a 19bp DNA "spacer". Binding of both TALENs would allow dimerization of the FokI nuclease

domains of each protein, causing cleavage of the DNA strand within the spacer region. Several TALENs were synthesized using this strategy, and *in vitro* assay of the TALEN pairs confirmed ability of the synthetic TALEN proteins to bind DNA with specificity to the EBEs and cause cleavage of the DNA.

Introduction

The ability to tailor DNA binding domains to interact with pre-selected DNA sequences is a crucial tool for molecular biology (Klug, 2010). Advances in the understanding of DNA-binding zinc finger proteins have led to creation of zinc finger nucleases (ZFNs), chimeric proteins with separate DNA-binding and DNA-cutting domains, allowing scientists to target precise locations in an organism's genome for knockout, editing, or replacement (Durai et al., 2005). Shortly after the advent of ZFNs, a new class of DNA-binding proteins was identified. These proteins, referred to as Transcription Activator-Like Effectors (TALEs), harbor a unique protein structure containing several identical 33-34 amino acid repeats. The amino acid composition of these repeats is highly conserved, with the exception of two residues in the 12^{th} and 13^{th} position of each repeat, which show more, but still limited, variability. Elucidation of the DNA-binding characteristics of these proteins revealed that each of the repeats within a TALE bound to a single nucleotide, and the composition of the 12^{th} and 13^{th} amino acid (termed the Repeat Variable Diresidue, RVD) determined the nucleotide to which the repeat bound (Boch et al., 2009; Moscou and Bogdanove, 2009). The simplicity of the TALE code for DNA specificity made these proteins an attractable alternative to the DNA recognition domain of ZFNs, which suffer from a lack of modularity and interchangeability of DNA-binding units. By coupling the same FokI nuclease domains used in ZFNs to the DNA recognition regions of TALEs, scientists were soon able to create TAL Effector Nucleases (TALENs) that, like ZFNs, could be used in pairs to bind

closely spaced DNA sequences and create double strand DNA breaks (DSBs) between the sequences. Such DSBs are targets for error prone nonhomologous end joining (NHEJ) DNA repair that often results in gene inactivation (knockout) (See Chapter 1). Many labs have since pursued the construction and application of TALENs for targeted gene knockouts and gene editing in a wide range of organisms (see Doyle et al. (2013) for review).

As each repeat of a TALE is >100bp in length and 14 to 20 repeats are needed for each artificial TALEN, custom synthesis of multiple repeats by gene synthesis companies is cost prohibitive for many research groups. The in-house construction of synthetic TALENs that bind to pre-selected DNA sequences has been of interest to many research laboratories, and several groups have developed methods to systematically construct multiple repeat units in an ordered fashion (see Chapter 1 of this thesis for discussion of the various methods of TALE synthesis available). Shortly after the elucidation of TALE binding modularity, our research group set out to develop an in-house method to synthesize TALEs for the use of gene editing in the green alga *Chlamydomonas reinhardtii*. The following section discusses the components and construction designs developed by our lab to assemble TALE repeats in an ordered fashion and to assay the resultant synthetic TALENs for DNA binding and cleavage specificity.

Results

Designs for TALEN repeat assembly call for directional ligation of eight repeat units in a single reaction. A library was constructed of repeat units containing different

RVDs (HD, NG, NI, NN) and PCR amplified to contain various DNA sequences adjacent to a Type IIs restriction site. Each repeat was modified via PCR with 12 different extensions to allow for ligation in 12 possible positions (Figure 4.1). As TALEN repeats were assembled in 8-mers, each repeat was required to be potentially ligated into one of six "internal" positions (positions 2-7 within each 8-mer) or six "external" positions (positions 1 or 8 of 8-mer 1, 2, or 3). Final TALENs contained a total of 24 repeats, requiring construction and ordered assembly of three 8-mers; dictating the need for six "external" options for each repeat.

For construction of eight-repeat units, a nine-fragment ligation was necessary (eight repeat units + one acceptor vector) (Figure 4.2, Figure 4.3). To improve the efficiency of this potentially unfavorable ligation, digested DNA fragments were quantified, and equimolar amounts of each fragment was used for ligation. Ligations were often allowed to proceed overnight at 4°C. Colony PCR was used to screen cells transformed with the ligation mixture to select cells bearing plasmids containing all eight repeats (Figure 4.4). This allowed for rapid screening of large numbers of colonies, which was often necessary, as the ligation efficiency of each 8-mer varied from 1.5-37.5%. This variation in ligation efficiency was likely due to multiple variables, including unequal amounts of the nine digested DNA fragments, slightly different ligation conditions, and variable amounts/activity of the DNA ligase.

Figure 4.1 - TALEN Repeat Library. The chart represents 48 constructs comprising the 'TALEN Repeat Library'. Each construct is comprised of one repeat unit containing one of four RVDs. The ends of each repeat unit are modified to allow for directional ligation of eight repeat units (to produce an 8-mer of defined amino acid sequence). The 1st and 8th repeat of each 8-mer can have one of three modifications ((1), (2), or (3)), depending on the position of the 8-mer (1st, 2nd, or 3rd) in the final TALE(N).

Label	**RVD / Target Nucleotide**	**Positions within 8-mer / 8-mer 1,2, or 3**		**Label**	**RVD / Target Nucleotide**	**Positions within 8-mer / 8-mer 1,2, or 3**
1(1)-A	NI/A	1/1		5-A	NI/A	5/*
1(1)-C	HD/C	1/1		5-C	HD/C	5/*
1(1)-G	NN/G	1/1		5-G	NN/G	5/*
1(1)-T	NG/T	1/1		5-T	NG/T	5/*
1(2)-A	NI/A	1/2		6-A	NI/A	6/*
1(2)-C	HD/C	1/2		6-C	HD/C	6/*
1(2)-G	NN/G	1/2		6-G	NN/G	6/*
1(2)-T	NG/T	1/2		6-T	NG/T	6/*
1(3)-A	NI/A	1/3		7-A	NI/A	7/*
1(3)-C	HD/C	1/3		7-C	HD/C	7/*
1(3)-G	NN/G	1/3		7-G	NN/G	7/*
1(3)-T	NG/T	1/3		7-T	NG/T	7/*
2-A	NI/A	2/*		8(1)-A	NI/A	8/1
2-C	HD/C	2/*		8(1)-C	HD/C	8/1
2-G	NN/G	2/*		8(1)-G	NN/G	8/1
2-T	NG/T	2/*		8(1)-T	NG/T	8/1
3-A	NI/A	3/*		8(2)-A	NI/A	8/2
3-C	HD/C	3/*		8(2)-C	HD/C	8/2
3-G	NN/G	3/*		8(2)-G	NN/G	8/2
3-T	NG/T	3/*		8(2)-T	NG/T	8/2
4-A	NI/A	4/*		8(3)-A	NI/A	8/3
4-C	HD/C	4/*		8(3)-C	HD/C	8/3
4-G	NN/G	4/*		8(3)-G	NN/G	8/3
4-T	NG/T	4/*		8(3)-T	NG/T	8/3

* repeat unit can be used for any 8-mer (1, 2, or 3)

Figure 4.2 - TALE Repeat Assembly. In standard cloning reactions, a total of eight repeats are ligated at one time. For illustration, only two repeats are shown. Digestion with BsaI creates unique overhangs that allow for directional ligation. Shown above, the "A" repeat can only ligate to the "T" repeat in one orientation, resulting in a seamless DNA fragment encoding two repeats.

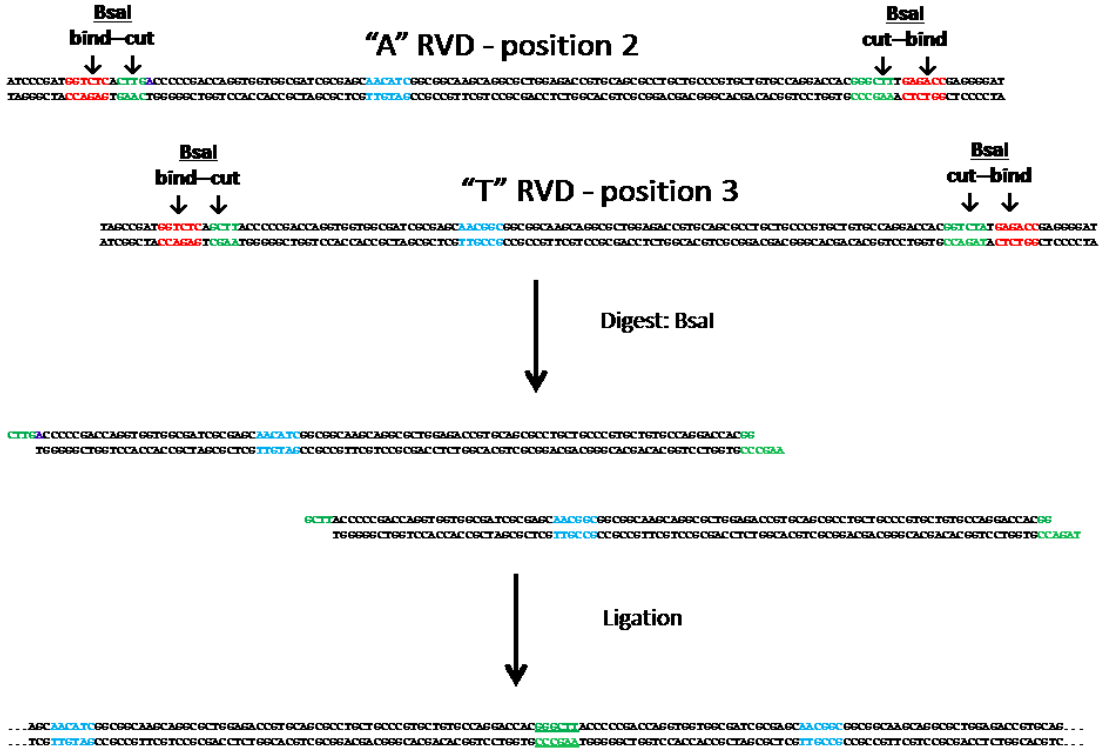


Figure 4.3 - 8-mer Assembly. Repeats have been modified via PCR to have, upon digestion with BsaI, unique overhangs complementary to only one fragment in the eight repeats. Digestion of repeat fragments with BsaI will create overhangs on each end of the repeats except those for later ligation (marked with X). Resultant fragment will consist of eight repeats ligated in a seamless fashion. External ligation points (marked X) will be used for later construction of full length TALE.

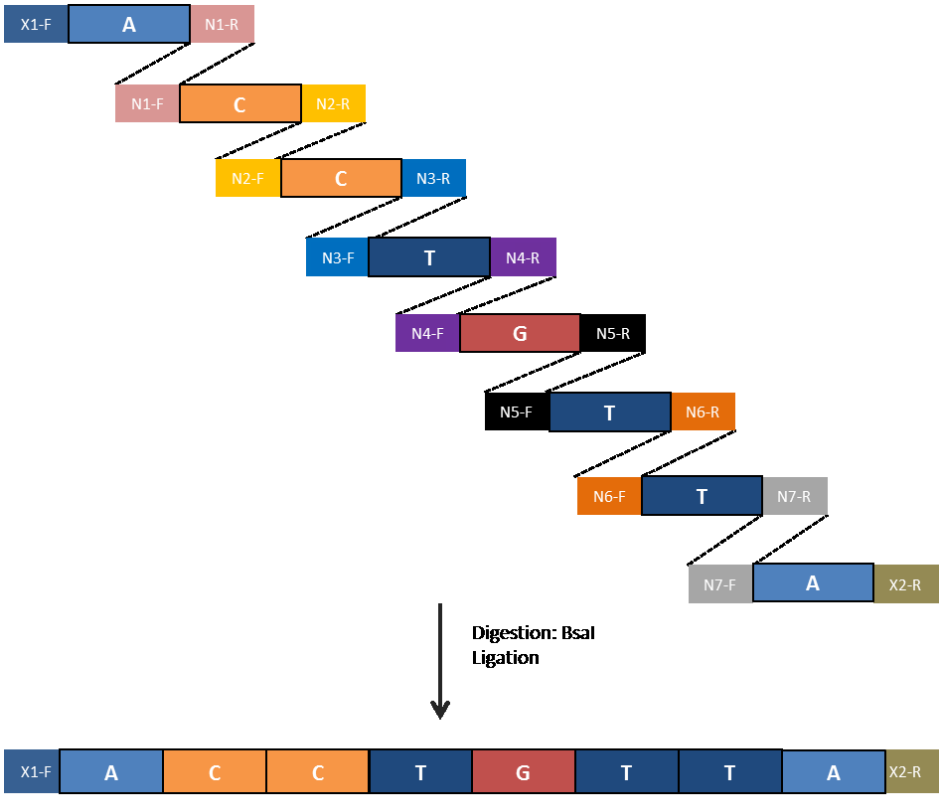
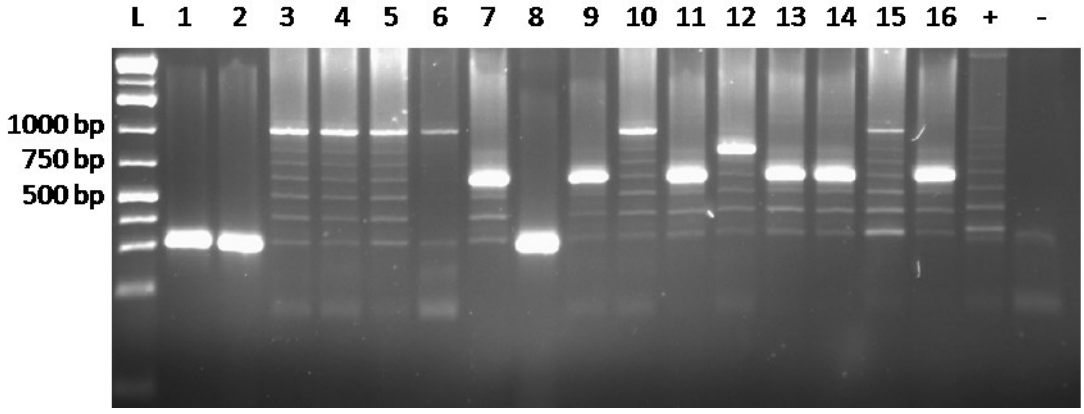


Figure 4.4 - Colony PCR Screen of Putative 8-mer Constructs. Example of colony PCR results from one set of ligations. Following overnight ligation of 8-mer, ligation mixture was transformed in DH5α E. coli cells and incubated overnight on LB plates containing 100 µg/mL ampicillin. Resultant colonies were picked and subjected to colony PCR. Previously-built 8-mer construct was used as a positive control, and failed constructs (containing 1 repeat) were used as a negative control. Due to self-priming of repeat fragments, positive samples often generated a laddering effect. In this figure, "L" designates the DNA ladder (Thermo Scientific GeneRuler 1kb DNA Ladder, SM0312). Labels "1-16" designate the 16 colonies selected for screening, and "+" and "-" designate the positive and negative controls, respectively. Six of the 16 samples (lanes 3, 4, 5, 6, 10 and 15) show a product corresponding to a full-length 8-mer.



Construction of the final TALEN required ligation of only four fragments (Figure 4.5); three 8-mer fragments and a “TALEN Shell”. The shell consisted of the cDNA of the endogenous TALE, AvrXa7, whose repeats were removed and replaced with cloning sites for the synthetic 24 repeats. Unsurprisingly, ligation of the final TALEN possessed a higher efficiency than construction of 8-mers. Due to this rate of success, and the difficulty in PCR amplifying assemblages of longer than eight repeat units at a time, colonies were screened by restriction enzyme digestion rather than colony PCR. Plasmids were digested with PvuII, which cuts numerous times within the N- and C-termini of each TALEN, allowing for easy identification of constructs containing a complete set of 24 repeats (Figure 4.6). Constructs containing 24 repeats often were recovered at rates of 80% to 100%.

Prior to use in Chlamydomonas, synthetic TALEN pairs were assayed for activity *in vitro* by incubating TALEN pairs with plasmid DNA containing the Effector Binding Element (EBE) for the TALEN pair being tested (Figure 4.7). All TALEN pairs assayed in this manner showed both activity and specificity in digestion of DNA.

Figure 4.5 - Final TALE Assembly. TALEN 8-mers and TALEN Shell were digested with BsmBI, producing unique overhangs. This allowed for a directional 4-part ligation, similar to the designs used in 8-mer construction.

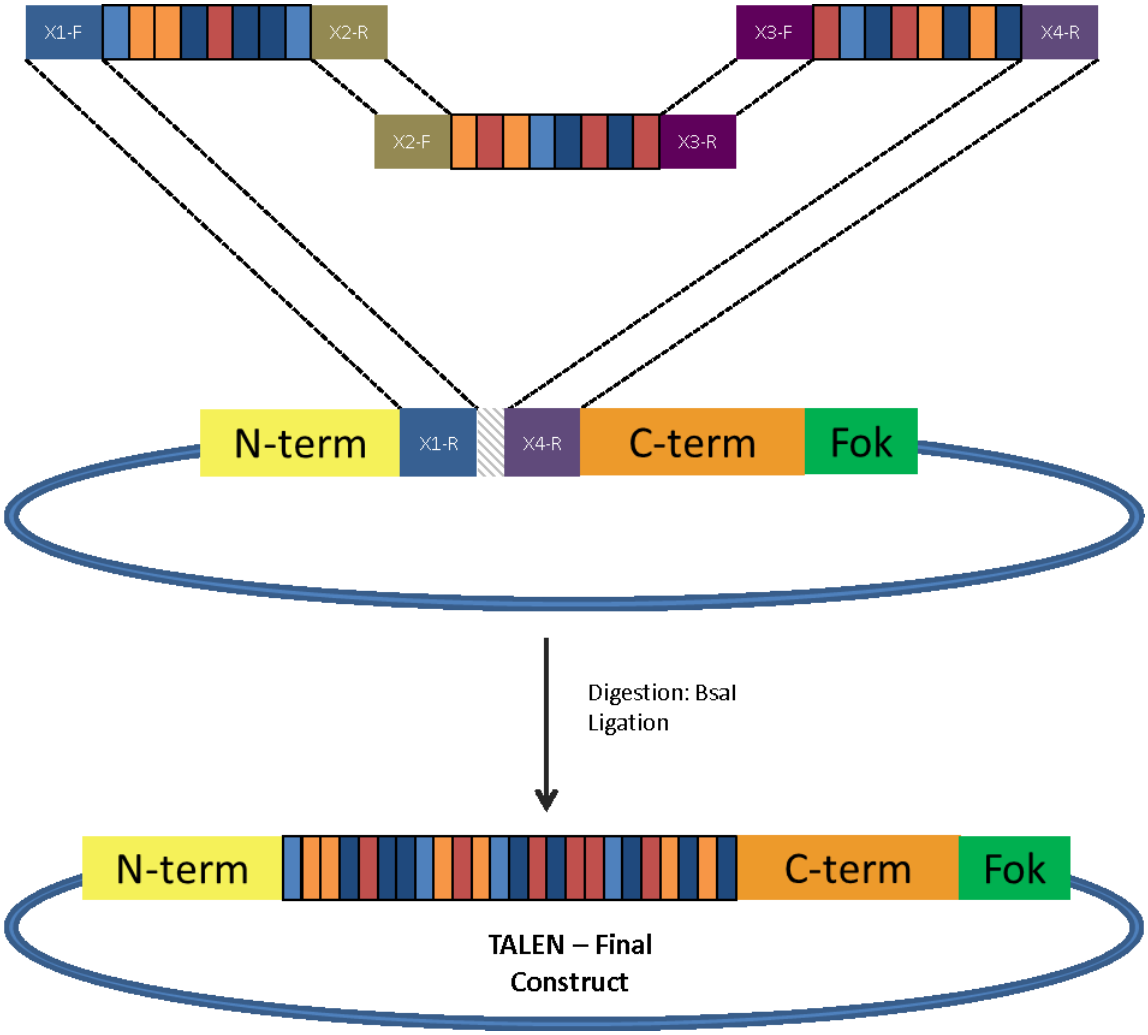


Figure 4.6 - Screen of Putative TALEN Constructs. Following ligation of 8-mers into an appropriate TALEN shell and transformation of E. coli cells, resultant colonies were grown overnight (3 mL LB + 100 µg/mL ampicillin) and plasmid DNA isolated. Plasmids were digested with PvuII and subjected to electrophoresis on a 1% agarose gel. Labels "1-6" and "7-12" represent samples 1-6 of the first TALEN and samples 1-6 of the second TALEN, respectively. The negative control (-) is an undigested sample of a previously-constructed TALEN, and the DNA ladder (L) used is Thermo Scientific GeneRuler 1kb DNA Ladder (SM0312).

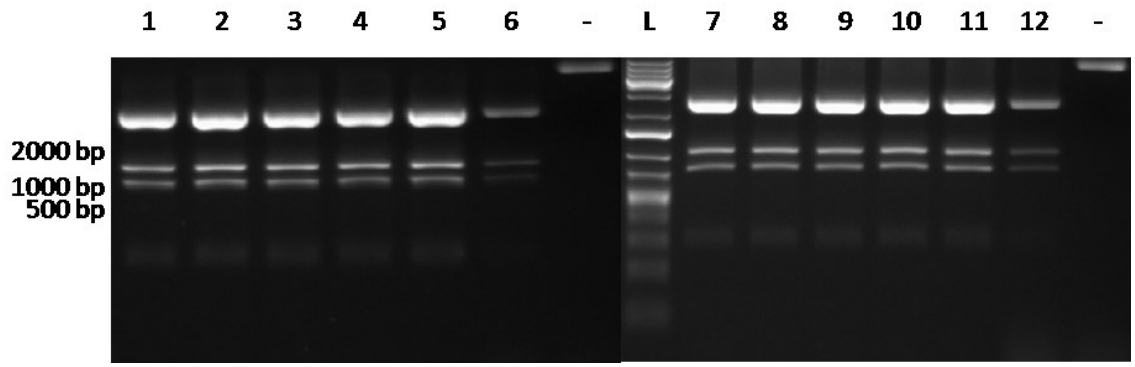
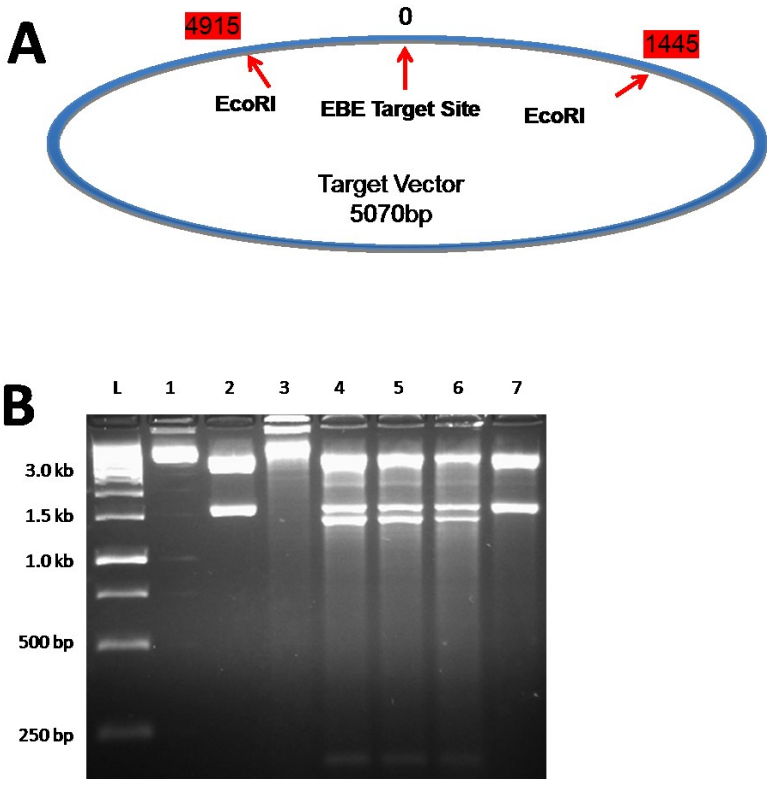


Figure 4.7 - Assay of Synthetic TALEN Activity *In vitro*. A: Schematic of TALEN EBE-containing plasmid. Highlighted are cuts sites for the synthetic TALEN pair (position zero) and EcoRI restriction sites. Restriction digest pattern: TALEN Pair: 5070bp (linear), EcoRI: 3740, 1600bp, TALEN Pair + EcoRI: 3740, 1445, 155bp. **B:** Analysis of TALEN-directed DNA cleavage patterns on a 1% agarose gel. Ladder is a 1kb DNA Ladder. To assess the requirement for two TALENs (and FokI dimerization) for DNA cleavasge, a single TALEN was used for the DNA digestion depicted in lane 7. Gel loading order: L: DNA Ladder, 1: Uncut, 2: EcoRI, 3: TALEN Pair, 4: TALEN Pair (0.2µL) + EcoRI, .5: TALEN Pair (0.5µL) + EcoRI, 6: TALEN Pair (1.0µL) + EcoRI, 7: TALEN 1 Only + EcoRI



Discussion

We have successfully created a method for the assembly of synthetic TALENs for target gene editing. All components have been codon-optimized for expression in Chlamydomonas. To date, multiple TALEN pairs targeting various Chlamydomonas genes and promoter regions, for gene knock out or gene activation, respectively, have been synthesized using these tools. Of those tested using the *in vitro* assay for TALEN activity, all TALEN pairs have shown specificity and precise activity in DNA binding and cleavage. The similarity in codon usage between Chlamydomonas and a number of higher eukaryotes with high G:C genomes makes the TALEN synthesis methods developed here attractive for use constructing TALENs for use in these plants and animals.

Methods and Materials

Repeat Library Synthesis

The amino acid sequence of TALE repeat unit we employed (with asterisks for the RVD amino acids in positions 12 and 13) was: LTPDQVVAIAS**GGKQALETVQRLLPVLCQDHG. Templates for repeats containing one of four RVDs (HD, NG, NI, NN) were synthesized by Genscript (New Jersey). Each RVD was selected for its ability to bind a specific nucleotide (HD:C, NG:T, NI:A, NN:A/G). Each fragment was reduced in that it lacked the first and last three nucleotides that code for leucine (Leu) and glycine (Gly), found in normal repeats. These repeats were termed ‘reduced repeats’.

PCR primers were then designed that amplified each reduced repeat, adding the missing codons coupled with a Type IIs restriction site (BsaI or BsmBI). A variety of codons were selected to code for Leu and Gly. The Type IIs restriction sites were aligned to the added nucleotides so that, once digested, they would expose the unique overhangs. By using different combinations of codons for Leu and Gly, we were able to create a library of repeats that, when digested with either of the Type IIs enzymes, would allow directional cloning. Each reduced repeat was amplified so that it could be ligated into one of 12 positions: internal (positions 2-7 of an 8-repeat series) or external (positions 1 or 8 of an 8-repeat series). For external repeats, additional uniqueness was required, as three 8-repeat series (8-mers) would eventually be ligated together to produce the complete 24-repeat series (24-mer). See Figure 4.1 for details on repeat library. All repeat units, following PCR amplification, were ligated into pAS2, a pBluescript SK+ vector modified with a SwaI site (a rare-cutting restriction enzyme with an 8bp recognition sequence) , via blunt end cloning.

Target Sequence Selection

Putative target sequences were analyzed for specific factors. All target sequences required an upstream thymidine residue flanking the 24 bp DNA target sequence. Due to the ambiguous binding nature of the NN RVD (bound equally well to both A and G), sequences were sought out that contained a low G-content (G composition of 30% or less within the target sequence). For TALEN dimerization, two TALE binding sites were selected. Due to TALEs binding in a 5’ to 3’ direction, TALE site

pairs had to be on opposite DNA strands with sufficient spacing (19bp) between the two DNA sequences to allow for effective dimerization of the two FokI nuclease subunits on the face-to-face TALEN molecules. After identification of a suitable 24bp target sequence (a so-called effector binding element or EBE) for one of the TALENs was made, repeat units of the synthetic TALEN were assembled first into three independent segments of 8 repeats each and then assembled into the desired 24 repeat TALEN.

8-mer Assembly

For 8-mer assembly, 8 repeat units were selected from the library corresponding to their desired position (1-8) within the 8-mer. For external positions (1,8), the position of the 8-mer in the final construct (8-mer 1, 2, or 3) affected which repeats were used for construction. As an example, if repeats targeting A and T were required for positions one and eight of the second 8-mer, repeats “1(2)-A” and “8(2)-T” would be selected from the repeat library. For internal repeats, which contained BsaI sites on both sides of the repeat, digestion with BsaI would yield a fragment size of 102bp, which was recovered from a 2% agarose gel. For external repeats, BsaI sites were present only on the internal side, i.e. the side that would ligate to repeat 2 or repeat 7. Thus, external repeats were digested with BsaI and a second restriction enzyme whose site was just outside of the repeat (BamHI, XhoI, EcoRI, or HindIII depending on the orientation of the repeat within the multiple cloning site). Digestion of external fragments with these enzymes produced a slightly larger fragment (130-140bp), which was also recovered from a 2% agarose gel. The vector pAS2, serving as a backbone for 8-mer construction,

was digested with the two non-BsaI enzymes used for the external repeats. Following digestion, the vector fragment (approximately 3kb) was recovered from a 2% agarose gel and treated with alkaline phosphatase to prevent self-ligation of partially digested vectors. A nine-part ligation was performed, comprised of equimolar amounts of the eight repeat fragments and the vector. The ligation mixture was incubated overnight at 4°C, and following incubation, electroporated into DH5α cells. Transformed cells were incubated overnight on LB + 100 µg/mL ampicillin plates. Resultant colonies were screened by colony PCR (Taq polymerase). Briefly, individual colonies were selected with a toothpick, submerged in a 40µL aliquot of LB, and then transferred to a Taq polymerase PCR reaction mixture containing 0.1% Tween20 and primers specific to the T7 and T3 promoters. Standard PCR conditions were used, with an annealing temperature of 45°C and an extension time of 70 seconds. Positive clones yielded a PCR product of 1006bp, with a characteristic 'laddered' appearance of minimal products every 102bp. DH5α lines containing positive clones were grown overnight in liquid LB + 100 µg/mL ampicillin, and plasmid DNA from resulting cells was collected using a Qiagen plasmid DNA mini-prep kit (www.Qiagen.com). Isolated plasmids were submitted for sequence analysis, utilizing primers targeting T7 (TAATACGACTCACTATAGGG) and T3 (CAATTAACCCTCACTAAAGG) sites. Plasmids containing 8-mers with the correct sequence were saved for construction of the final TALE.

Final TALE Assembly

For construction of full length TALEs containing 24 ordered repeats, each of the three previously constructed 8-mers were digested with BsmBI, exposing the Leu and Gly codons on the external repeats, generating an 818bp product, which was purified from a 1% agarose gel. A TALE 'shell' was selected for cloning based on the desired application. For gene editing or activation, 'pTALE-Shell' was used, while for gene repression, 'pTALR-Shell' was used. Both constructs contained a modified version of the TALE AvrXa7, codon-optimized for expression in Chlamydomonas. The AvrXa7 gene was modified to lack any endogenous repeats except for the final "truncated" TAL repeat found in all native TALEs. The remaining 'shell' contained coding regions for the N- and C-termini of AvrXa7, with BsmBI sites flanking the 3' end of the N-terminus coding region and 5' end of the C-terminus coding region. 'pTALR-Shell' differed from 'pTALE-Shell' in that it lacked the 3' 129 nucleotides of the C-terminus coding region, which coded for 43 amino acids comprising the activation domain of TALEs. BsmBI digestion of the shell produced a 4784bp fragment, which was recovered from a 1% agarose gel. A 4-part ligation comprised of the three 8-mers and the shell was performed overnight at 4°C. Ligation mixtures were electroporated into E. coli DH5α cells, and cells were incubated overnight on LB + 100 µg/mL ampicillin plates. Resultant colonies were grown overnight in liquid LB + 100 µg/mL ampicillin. Plasmid DNA was isolated from individual transformants and was analyzed by digestion with PvuII. PvuII sites are prevalent throughout the N- and C-termini of AvrXa7, but absent in the repeat units. Thus, if all 24 repeats were successfully ligated into the shell, digestion with PvuII would yield an easily

identifiable 2607bp band. Positive clones were subjected to sequencing to verify intactness and correct order of the 24 repeats. Two primers were used for sequencing the final TALE assembly, 'N-term_seq_3':F' (CCGGCCAGCTGCTGAAGA) and 'C-term_seq_5':R' (GCGGCTCAGCTGCGCCAC), which bound 74bp upstream and 65bp downstream of the repeat region, respectively.

Assembly of Final TALE Constructs

For gene editing, TALE constructs needed to be fused to a FokI nuclease domain, creating a TALEN. Several FokI domains were available for use, including homodimers and modified heterodimers. Restriction sites pre-incorporated onto shells allowed for quick attachment of the desired FokI domain and movement of the TALEN into an expression cassette.

***In vitro* Assay of Synthetic TALEN Activity**

Before *in vivo* experiments using TALENs in Chlamydomonas were conducted, TALEN pairs were assayed for *in vitro* DNA binding and cleavage activity. DNA oligonucleotides comprising the binding site of the TALEN pair (and the appropriate spacer sequence) were synthesized by Eurofins (Alabama) and cloned into a modified pBluescript SK+ vector. Following synthesis of the TALEN pair, each TALEN was cloned into a pET28b expression vector and electroporated into E. coli strain BL21(DE3). Cells containing the desired construct were incubated at 37°C until reaching a density of OD_{600} = 0.6, at which time expression of the protein was induced using IPTG. Bacterial cells were then lysed, and the TALE protein was purified by passing the supernatant

through a Ni^{+2}-NTA metal-affinity resin. Eluted protein was dialized against 50mM Tris (pH 7.5) and 2mM DTT. Final protein concentration was determined using Bradford's reagent (BioRad).

In vitro binding and cleavage activity of the TALEN pair was assayed by incubating the proteins with the pBluescript vector containing the DNA binding site. 0.5 µg of each TALEN was added to a solution containing 1 µg of the target DNA plasmid and a buffer allowing optimal function of the FokI nuclease. TALENs were incubated with the DNA at 37°C for 30 minutes and inactivated by heating the mixture to 75°C for 10 minutes. Following digestion with the TALEN pair, the plasmid DNA was subjected to further digestion with a common restriction enzyme (often EcoRI) that had one or two digestion sites at least 200bp away from the TALEN DNA binding site. Complete digestion with EcoRI provided a better resolution for ascertaining the activity and specificity of the TALEN pair.

Acknowledgements

There are several people I would like to acknowledge for their contributions to this work. Donald Weeks assisted in several of the designs for the TALEN library and construction methods, which helped significantly in the efficacy of this system. Wen Zhi Jiang provided invaluable assistance and expertise in the design and execution of the *in vitro* assay experiments. I would also like to thank members of the Weeks lab, particularly Kempton Horken, Thomas Plucinak, and Amy Knobbe for their invaluable help in troubleshooting various issues during this work. In addition, Daniel Kuehler, Tan

Nguyen, and Andrew Blazek provided assistance in much of the labors involved. Financial support for this project was provided by the Department of Energy JGI (Award JGI 2010 CSP-146), National Science Foundation (grants MCB-0952533 and EPSCoR-1004094) and the Consortium for Algal Biofuels Commercialization (Department of Energy Award DE-EE0003373).

References

Anders, S., and Huber, W. (2010). Differential expression analysis for sequence count data. Genome Biol **11,** R106.

Antony, G., Zhou, J., Huang, S., Li, T., Liu, B., White, F., and Yang, B. (2010). Rice xa13 recessive resistance to bacterial blight is defeated by induction of the disease susceptibility gene Os-11N3. Plant Cell **22,** 3864-3876.

Araujo, A.S.F., Monteiro, R.T.R., and Abarkeli, R.B. (2003). Effect of glyphosate on the microbial activity of two Brazilian soils. Chemosphere **52,** 799-804.

Arias, R.S., Netherland, M.D., Scheffler, B.E., Puri, A., and Dayan, F.E. (2005). Molecular evolution of herbicide resistance to phytoene desaturase inhibitors in Hydrilla verticillata and its potential use to generate herbicide-resistant crops. Pest Manag Sci **61,** 258-268.

Arias, R.S., Dayan, F.E., Michel, A., Howell, J., and Scheffler, B.E. (2006). Characterization of a higher plant herbicide-resistant phytoene desaturase and its use as a selectable marker. Plant Biotechnol J **4,** 263-273.

Ashburner, M., Ball, C.A., Blake, J.A., Botstein, D., Butler, H., Cherry, J.M., . . . Sherlock, G. (2000). Gene ontology: tool for the unification of biology. The Gene Ontology Consortium. Nat Genet **25,** 25-29.

Ashworth, J., Havranek, J.J., Duarte, C.M., Sussman, D., Monnat, R.J., Jr., Stoddard, B.L., and Baker, D. (2006). Computational redesign of endonuclease DNA binding and cleavage specificity. Nature **441,** 656-659.

Badger, M.R., and Price, G.D. (1994). The Role of Carbonic-Anhydrase in Photosynthesis. Annual Review of Plant Physiology and Plant Molecular Biology **45,** 369-392.

Badger, M.R., Kaplan, A., and Berry, J.A. (1980). Internal Inorganic Carbon Pool of Chlamydomonas reinhardtii: EVIDENCE FOR A CARBON DIOXIDE-CONCENTRATING MECHANISM. Plant Physiol **66,** 407-413.

Bailey, T.L., Boden, M., Buske, F.A., Frith, M., Grant, C.E., Clementi, L., . . . Noble, W.S. (2009). MEME SUITE: tools for motif discovery and searching. Nucleic Acids Res **37,** W202-208.

Barry, G.F., Kishore, G.M., Padgette, S.R., and Stallings, W.C. (1994). Glyphosate-Tolerant 5-Enolpyruvylshikimate-3-phosphate Synthases (United States).

Barry, G.F., Kishore, G.M., Padgette, S.R., Taylor, M.L., Kolacz, K.H., Weldon, M., . . . Hallas, L. (1992). Inhibitors of amino acid biosynthesis: strategies for imparting glyphosate tolerance to crop plants. In Biosynthesis and Molecular Regulation of Amino Acids in Plants, B.K. Singh, H.E. Flores, and J.C. Shannon, eds (Mockville, MD: American Society of Plant Physiologists), pp. 139-145.

Beale, S.I., and Weinstein, J.D. (1990). Tetrapyrrole metabolism in photosynthetic organisms. In Biosynthesis of Heme and Chlorophylls, H.A. Dailey, ed (New York: McGraw-Hill), pp. 287-391.

Bedell, V.M., Wang, Y., Campbell, J.M., Poshusta, T.L., Starker, C.G., Krug, R.G., 2nd, . . . Ekker, S.C. (2012). In vivo genome editing using a high-efficiency TALEN system. Nature **491,** 114-118.

Benjamini, Y., and Hochberg, Y. (1995). Controlling the False Discovery Rate - a Practical and Powerful Approach to Multiple Testing. Journal of the Royal Statistical Society Series B-Methodological **57,** 289-300.

Berry, J., Boynton, J., Kaplan, A., and Badger, M. (1976). Growth and photosynthesis of Chlamydomonas reinhardtii as a function of CO2 concentration. Carnegie Institute of Washington Yearbook **75,** 423-432.

Bhaya, D., Davison, M., and Barrangou, R. (2011). CRISPR-Cas systems in bacteria and archaea: versatile small RNAs for adaptive defense and regulation. Annu Rev Genet **45,** 273-297.

Birney, E., Stamatoyannopoulos, J.A., Dutta, A., Guigo, R., Gingeras, T.R., Margulies, E.H., . . . Consortium, E.P. (2007). Identification and analysis of functional elements in 1% of the human genome by the ENCODE pilot project. Nature **447,** 799-816.

Bitinaite, J., Wah, D.A., Aggarwal, A.K., and Schildkraut, I. (1998). FokI dimerization is required for DNA cleavage. Proc Natl Acad Sci U S A **95,** 10570-10575.

Blount, B.A., Weenink, T., Vasylechko, S., and Ellis, T. (2012). Rational diversification of a promoter providing fine-tuned expression and orthogonal regulation for synthetic biology. PLoS One **7,** e33279.

Boch, J., Scholze, H., Schornack, S., Landgraf, A., Hahn, S., Kay, S., . . . Bonas, U. (2009). Breaking the code of DNA binding specificity of TAL-type III effectors. Science **326,** 1509-1512.

Bogdanove, A.J., Schornack, S., and Lahaye, T. (2010). TAL effectors: finding plant genes for disease and defense. Curr Opin Plant Biol **13,** 394-401.

Böger, P., and Sandmann, G. (1998). Carotenoid biosynthesis inhibitor herbicides - mode of action and resistance mechanisms. Pesticide Outlook **9,** 29-35.

Boocock, M.R., and Coggins, J.R. (1983). Kinetics of 5-enolpyruvylshikimate-3-phosphate synthase inhibition by glyphosate. Febs Letters **154,** 127-133.

Borkhsenious, O.N., Mason, C.B., and Moroney, J.V. (1998). The intracellular localization of ribulose-1,5-bisphosphate carboxylase/oxygenase in Chlamydomonas reinhardtii. Plant Physiol **116,** 1585-1591.

Bozzo, G.G., and Colman, B. (2000). The induction of inorganic carbon transport and external carbonic anhydrase in Chlamydomonas reinhardtii is regulated by external CO2 concentration. Plant Cell and Environment **23,** 1137-1144.

Breitenbach, J., Zhu, C.F., and Sandmann, G. (2001). Bleaching herbicide norflurazon inhibits phytoene desaturase by competition with the cofactors. Journal of Agricultural and Food Chemistry **49,** 5270-5272.

Briggs, A.W., Rios, X., Chari, R., Yang, L., Zhang, F., Mali, P., and Church, G.M. (2012). Iterative capped assembly: rapid and scalable synthesis of repeat-module DNA such as TAL effectors from individual monomers. Nucleic Acids Res **40,** e117.

Buchanan, M.J., Imam, S.H., Eskue, W.A., and Snell, W.J. (1989). Activation of the cell wall degrading protease, lysin, during sexual signalling in Chlamydomonas: the enzyme is stored as an inactive, higher relative molecular mass precursor in the periplasm. J Cell Biol **108,** 199-207.

Burow, M.D., Chen, Z.Y., Mouton, T.M., and Moroney, J.V. (1996). Isolation of cDNA clones of genes induced upon transfer of Chlamydomonas reinhardtii cells to low CO2. Plant Molecular Biology **31,** 443-448.

Cade, L., Reyon, D., Hwang, W.Y., Tsai, S.Q., Patel, S., Khayter, C., . . . Yeh, J.R. (2012). Highly efficient generation of heritable zebrafish gene mutations using homo- and heterodimeric TALENs. Nucleic Acids Res **40,** 8001-8010.

Calamita, G., Kempf, B., Rudd, K.E., Bonhivers, M., Kneip, S., Bishai, W.R., . . . Agre, P. (1997). The aquaporin-Z water channel gene of Escherichia coli: structure, organization and phylogeny. Biol Cell **89,** 321-329.

Cardol, P., Vanrobaeys, F., Devreese, B., Van Beeumen, J., Matagne, R.F., and Remacle, C. (2004). Higher plant-like subunit composition of mitochondrial complex I from Chlamydomonas reinhardtii: 31 conserved components among eukaryotes. Biochimica Et Biophysica Acta-Bioenergetics **1658,** 212-224.

Cardol, P., Gonzalez-Halphen, D., Reyes-Prieto, A., Baurain, D., Matagne, R.F., and Remacle, C. (2005). The mitochondrial oxidative phosphorylation proteome of Chlamydomonas reinhardtii deduced from the genome sequencing project. Plant Physiol **137,** 447-459.

Carlson, D.F., Tan, W., Lillico, S.G., Stverakova, D., Proudfoot, C., Christian, M., . . . Fahrenkrug, S.C. (2012). Efficient TALEN-mediated gene knockout in livestock. Proc Natl Acad Sci U S A **109,** 17382-17387.

Carroll, D. (2011). Genome engineering with zinc-finger nucleases. Genetics **188,** 773-782.

Casas-Mollano, J.A., Rohr, J., Kim, E.J., Balassa, E., van Dijk, K., and Cerutti, H. (2008). Diversification of the core RNA interference machinery in Chlamydomonas reinhardtii and the role of DCL1 in transposon silencing. Genetics **179,** 69-81.

Castle, L.A., Siehl, D.L., Gorton, R., Patten, P.A., Chen, Y.H., Bertain, S., . . . Lassner, M.W. (2004). Discovery and directed evolution of a glyphosate tolerance gene. Science **304,** 1151-1154.

Castruita, M., Casero, D., Karpowicz, S.J., Kropat, J., Vieler, A., Hsieh, S.I., . . . Merchant, S.S. (2011). Systems Biology Approach in Chlamydomonas Reveals Connections between Copper Nutrition and Multiple Metabolic Steps. Plant Cell **23,** 1273-1292.

Cerdeira, A.L., and Duke, S.O. (2010). Effects of glyphosate-resistant crop cultivation on soil and water quality. GM Crops **1,** 16-24.

Cermak, T., Doyle, E.L., Christian, M., Wang, L., Zhang, Y., Schmidt, C., . . . Voytas, D.F. (2011). Efficient design and assembly of custom TALEN and other TAL effector-based constructs for DNA targeting. Nucleic Acids Res **39,** e82.

Cerutti, H., Ma, X.R., Msanne, J., and Repas, T. (2011). RNA-Mediated Silencing in Algae: Biological Roles and Tools for Analysis of Gene Function. Eukaryotic Cell **10,** 1164-1172.

Chamovitz, D., Pecker, I., and Hirschberg, J. (1991). The molecular basis of resistance to the herbicide norflurazon. Plant Molecular Biology **16,** 967-974.

Chamovitz, D., Sandmann, G., and Hirschberg, J. (1993). Molecular and biochemical characterization of herbicide-resistant mutants of cyanobacteria reveals that phytoene desaturation is a rate-limiting step in carotenoid biosynthesis. J Biol Chem **268,** 17348-17353.

Cheli, Y., Williams, S.A., Ballotti, R., Nugent, D.J., and Kunicki, T.J. (2010). Enhanced Binding of Poly(ADP-ribose)polymerase-1 and Ku80/70 to the ITGA2 Promoter via an Extended Cytosine-Adenosine Repeat. PLoS One **5**.

Chen, K., and Gao, C. (2013). TALENs: customizable molecular DNA scissors for genome engineering of plants. J Genet Genomics **40,** 271-279.

Chen, L.Q., Qu, X.Q., Hou, B.H., Sosso, D., Osorio, S., Fernie, A.R., and Frommer, W.B. (2012). Sucrose efflux mediated by SWEET proteins as a key step for phloem transport. Science **335,** 207-211.

Chen, Z.Y., Lavigne, L.L., Mason, C.B., and Moroney, J.V. (1997). Cloning and overexpression of two cDNAs encoding the low-CO2-inducible chloroplast envelope protein LIP-36 from Chlamydomonas reinhardtii. Plant Physiol **114,** 265-273.

Choulika, A., Perrin, A., Dujon, B., and Nicolas, J.F. (1995). Induction of homologous recombination in mammalian chromosomes by using the I-SceI system of Saccharomyces cerevisiae. Mol Cell Biol **15,** 1968-1973.

Christian, M., Cermak, T., Doyle, E.L., Schmidt, C., Zhang, F., Hummel, A., . . . Voytas, D.F. (2010). Targeting DNA double-strand breaks with TAL effector nucleases. Genetics **186,** 757-761.

Coleman, J.R., and Grossman, A.R. (1984). Biosynthesis of Carbonic-Anhydrase in Chlamydomonas-Reinhardtii during Adaptation to Low Co2. Proceedings of the National Academy of Sciences of the United States of America-Biological Sciences **81,** 6049-6053.

Cong, L., Zhou, R., Kuo, Y.C., Cunniff, M., and Zhang, F. (2012). Comprehensive interrogation of natural TALE DNA-binding modules and transcriptional repressor domains. Nat Commun **3,** 968.

Cong, L., Ran, F.A., Cox, D., Lin, S., Barretto, R., Habib, N., . . . Zhang, F. (2013). Multiplex Genome Engineering Using CRISPR/Cas Systems. Science **339,** 819-823.

Curtin, S.J., Zhang, F., Sander, J.D., Haun, W.J., Starker, C., Baltes, N.J., . . . Stupar, R.M. (2011). Targeted mutagenesis of duplicated genes in soybean with zinc-finger nucleases. Plant Physiol **156,** 466-473.

Dahlem, T.J., Hoshijima, K., Jurynec, M.J., Gunther, D., Starker, C.G., Locke, A.S., . . . Grunwald, D.J. (2012). Simple methods for generating and detecting locus-specific mutations induced with TALENs in the zebrafish genome. PLoS Genet **8,** e1002861.

Das, A.C., Debnath, A., and Mukherjee, D. (2003). Effect of the herbicides oxadiazon and oxyfluorfen on phosphates solubilizing microorganisms and their persistence in rice fields. Chemosphere **53,** 217-221.

Dayan, F.E., and Duke, S.O. (1997). Overview of protoporphyrinogen oxidase-inhibiting herbicides. Proceedings of the Brighton Crop Protection Conference: Weeds**,** 83-92.

DeFrancesco. (2011). Move over ZFNs. Nat Biotechnol **29,** 681-684.

Deng, D., Yin, P., Yan, C., Pan, X., Gong, X., Qi, S., . . . Shi, Y. (2012). Recognition of methylated DNA by TAL effectors. Cell Res **22,** 1502-1504.

Dill, G.M., Cajacob, C.A., and Padgette, S.R. (2008). Glyphosate-resistant crops: adoption, use and future considerations. Pest Manag Sci **64,** 326-331.

Dohm, J.C., Lottaz, C., Borodina, T., and Himmelbauer, H. (2008). Substantial biases in ultra-short read data sets from high-throughput DNA sequencing. Nucleic Acids Res **36**.

Donnelly, M.L., Hughes, L.E., Luke, G., Mendoza, H., ten Dam, E., Gani, D., and Ryan, M.D. (2001). The 'cleavage' activities of foot-and-mouth disease virus 2A site-directed mutants and naturally occurring '2A-like' sequences. J Gen Virol **82,** 1027-1041.

Doyle, E.L., Stoddard, B.L., Voytas, D.F., and Bogdanove, A.J. (2013). TAL effectors: highly adaptable phytobacterial virulence factors and readily engineered DNA-targeting proteins. Trends Cell Biol **23,** 390-398.

Duanmu, D., Miller, A.R., Horken, K.M., Weeks, D.P., and Spalding, M.H. (2009a). Knockdown of limiting-CO2-induced gene HLA3 decreases HCO3- transport and photosynthetic Ci affinity in Chlamydomonas reinhardtii. Proceedings of the National Academy of Sciences of the United States of America **106,** 5990-5995.

Duanmu, D.Q., Wang, Y.J., and Spalding, M.H. (2009b). Thylakoid Lumen Carbonic Anhydrase (CAH3) Mutation Suppresses Air-Dier Phenotype of LCIB Mutant in Chlamydomonas reinhardtii. Plant Physiol **149,** 929-937.

Duke, S.O., Baerson, S.R., and Rimando, A.M. (2003). Herbicides: Glyphosate. In Encyclopedia of agrochemicals, J.R. Plimmer, D.W. Gammon, N.N. Ragsdale, and Knovel (Firm), eds (Hoboken, N.J.: Wiley-Interscience,), pp. 3 v. (x, 1638 p.).

Duke, S.O., Lydon, J., Becerril, J.M., Sherman, T.D., Lehnen, L.P., and Matsumoto, H. (1991). Protoporphyrinogen Oxidase-Inhibiting Herbicides. Weed Science **39,** 465-473.

Durai, S., Mani, M., Kandavelou, K., Wu, J., Porteus, M.H., and Chandrasegaran, S. (2005). Zinc finger nucleases: custom-designed molecular scissors for genome engineering of plant and mammalian cells. Nucleic Acids Res **33,** 5978-5990.

Engler, C., Kandzia, R., and Marillonnet, S. (2008). A one pot, one step, precision cloning method with high throughput capability. PLoS One **3,** e3647.

EPA, U.S. (1996). R.E.D. Facts: Norflurazon, U.S.E.P. Agency, ed (U.S. Environmental Protection Agency, Prevention, Pesticides and Toxic Substances).

EPA, U.S. (2002). Reregistration Eligibility Decision (RED) - Oxyfluorfen, U.S.E.P. Agency, ed.

EPA, U.S. (2003). Reregistration Eligibility Decision for Oxadiazon, U.S.E.P. Agency, ed.

Ermilova, E.V., Zalutskaya, Z.M., Gromov, B.V., Hader, D.P., and Purton, S. (2000). Isolation and characterisation of chemotactic mutants of Chlamydomonas reinhardtii obtained by insertional mutagenesis. Protist **151,** 127-137.

Eschenburg, S., Kabsch, W., Healy, M.L., and Schonbrunn, E. (2003). A new view of the mechanisms of UDP-N-acetylglucosamine enolpyruvyl transferase (MurA) and 5-enolpyruvylshikimate-3-phosphate synthase (AroA) derived from X-ray structures of their tetrahedral reaction intermediate states. J Biol Chem **278,** 49215-49222.

Fang, W., Si, Y.Q., Douglass, S., Casero, D., Merchant, S.S., Pellegrini, M., . . . Spalding, M.H. (2012). Transcriptome-Wide Changes in Chlamydomonas reinhardtii Gene Expression Regulated by Carbon Dioxide and the CO2-Concentrating Mechanism Regulator CIA5/CCM1. Plant Cell **24,** 1876-1893.

Fujii, S., Kono, H., Takenaka, S., Go, N., and Sarai, A. (2007). Sequence-dependent DNA deformability studied using molecular dynamics simulations. Nucleic Acids Res **35,** 6063-6074.

Fujiwara, S., Fukuzawa, H., Tachiki, A., and Miyachi, S. (1990). Structure and differential expression of two genes encoding carbonic anhydrase in Chlamydomonas reinhardtii. Proc Natl Acad Sci U S A **87,** 9779-9783.

Fukuzawa, H., Fujiwara, S., Tachiki, A., and Miyachi, S. (1990a). Nucleotide sequences of two genes CAH1 and CAH2 which encode carbonic anhydrase polypeptides in Chlamydomonas reinhardtii. Nucleic Acids Res **18,** 6441-6442.

Fukuzawa, H., Fujiwara, S., Yamamoto, Y., Dionisio-Sese, M.L., and Miyachi, S. (1990b). cDNA cloning, sequence, and expression of carbonic anhydrase in Chlamydomonas reinhardtii: regulation by environmental CO2 concentration. Proc Natl Acad Sci U S A **87,** 4383-4387.

Fukuzawa, H., Miura, K., Ishizaki, K., Kucho, K.I., Saito, T., Kohinata, T., and Ohyama, K. (2001). Ccm1, a regulatory gene controlling the induction of a carbon-concentrating mechanism in Chlamydomonas reinhardtii by sensing CO2 availability. Proc Natl Acad Sci U S A **98,** 5347-5352.

Funke, R.P., Kovar, J.L., and Weeks, D.P. (1997). Intracellular carbonic anhydrase is essential to photosynthesis in Chlamydomonas reinhardtii at atmospheric levels

of CO2. Demonstration via genomic complementation of the high-CO2-requiring mutant ca-1. Plant Physiol **114,** 237-244.

Funke, T., Han, H., Healy-Fried, M.L., Fischer, M., and Schonbrunn, E. (2006). Molecular basis for the herbicide resistance of Roundup Ready crops. Proceedings of the National Academy of Sciences of the United States of America **103,** 13010-13015.

Ganapathi, M., Singh, G.P., Sandhu, K.S., Brahmachari, S.K., and Brahmachari, V. (2007). A whole genome analysis of 5 ' regulatory regions of human genes for putative cis-acting modulators of nucleosome positioning. Gene **391,** 242-251.

Gao, H., Wu, X., Chai, J., and Han, Z. (2012). Crystal structure of a TALE protein reveals an extended N-terminal DNA binding region. Cell Res **22,** 1716-1720.

Geissler, R., Scholze, H., Hahn, S., Streubel, J., Bonas, U., Behrens, S.E., and Boch, J. (2011). Transcriptional activators of human genes with programmable DNA-specificity. PLoS One **6,** e19509.

Genkov, T., Meyer, M., Griffiths, H., and Spreitzer, R.J. (2010). Functional hybrid rubisco enzymes with plant small subunits and algal large subunits: engineered rbcS cDNA for expression in chlamydomonas. J Biol Chem **285,** 19833-19841.

Geraghty, A.M., and Spalding, M.H. (1996). Molecular and Structural Changes in Chlamydomonas under Limiting CO2 (A Possible Mitochondrial Role in Adaptation). Plant Physiol **111,** 1339-1347.

Giesy, J.P., Dobson, S., and Solomon, K.R. (2000). Ecotoxicological risk assessment for Roundup (R) Herbicide. Reviews of Environmental Contamination and Toxicology, Vol 167 **167,** 35-120.

Giordano, M., Beardall, J., and Raven, J.A. (2005). CO2 concentrating mechanisms in algae: mechanisms, environmental modulation, and evolution. Annu Rev Plant Biol **56,** 99-131.

Giordano, M., Norici, A., Forssen, M., Eriksson, M., and Raven, J.A. (2003). An anaplerotic role for mitochondrial carbonic anhydrase in Chlamydomonas reinhardtii. Plant Physiol **132,** 2126-2134.

Gonzalez-Ballester, D., Casero, D., Cokus, S., Pellegrini, M., Merchant, S.S., and Grossman, A.R. (2010). RNA-Seq Analysis of Sulfur-Deprived Chlamydomonas Cells Reveals Aspects of Acclimation Critical for Cell Survival. Plant Cell **22,** 2058-2084.

Grube, A., Donaldson, D., Kiely, T., and Wu, L. (2011). Pesticides Industry Sales and Usage: 2006 and 2007 Market Estimates, U.S.E.P. Agency, ed (Washington, DC).

Gu, K., Yang, B., Tian, D., Wu, L., Wang, D., Sreekala, C., . . . Yin, Z. (2005). R gene expression induced by a type-III effector triggers disease resistance in rice. Nature **435,** 1122-1125.

Ha, S.B., Lee, S.B., Lee, Y., Yang, K., Lee, N., Jang, S.M., . . . Back, K. (2004). The plastidic Arabidopsis protoporphyrinogen IX oxidase gene, with or without the transit

sequence, confers resistance to the diphenyl ether herbicide in rice. Plant Cell and Environment **27,** 79-88.

Hanson, D.T., Franklin, L.A., Samuelsson, G., and Badger, M.R. (2003). The Chlamydomonas reinhardtii cia3 mutant lacking a thylakoid lumen-localized carbonic anhydrase is limited by CO2 supply to rubisco and not photosystem II function in vivo. Plant Physiol **132,** 2267-2275.

Hao, G.F., Zuo, Y., Yang, S.G., and Yang, G.F. (2011). Protoporphyrinogen Oxidase Inhibitor: An Ideal Target for Herbicide Discovery. Chimia **65,** 961-969.

Hardcastle, T.J., and Kelly, K.A. (2010). baySeq: Empirical Bayesian methods for identifying differential expression in sequence count data. Bmc Bioinformatics **11**.

Harris, E.H. (1989). The Chlamydomonas sourcebook : a comprehensive guide to biology and laboratory use. (San Diego: Academic Press).

Hockemeyer, D., Wang, H., Kiani, S., Lai, C.S., Gao, Q., Cassady, J.P., . . . Jaenisch, R. (2011). Genetic engineering of human pluripotent cells using TALE nucleases. Nat Biotechnol **29,** 731-734.

Huang, P., Xiao, A., Zhou, M., Zhu, Z., Lin, S., and Zhang, B. (2011). Heritable gene targeting in zebrafish using customized TALENs. Nat Biotechnol **29,** 699-700.

Hui, J.Y., Hung, L.H., Heiner, M., Schreiner, S., Neumuller, N., Reither, G., . . . Bindereif, A. (2005). Intronic CA-repeat and CA-rich elements: a new class of regulators of mammalian alternative splicing. Embo Journal **24,** 1988-1998.

Iglesias, N., Redon, S., Pfeiffer, V., Dees, M., Lingner, J., and Luke, B. (2011). Subtelomeric repetitive elements determine TERRA regulation by Rap1/Rif and Rap1/Sir complexes in yeast. EMBO Rep **12,** 587-593.

Im, C.S., and Grossman, A.R. (2002). Identification and regulation of high light-induced genes in Chlamydomonas reinhardtii. Plant J **30,** 301-313.

Im, C.S., Zhang, Z.D., Shrager, J., Chang, C.W., and Grossman, A.R. (2003). Analysis of light and CO2 regulation in Chlamydomonas reinhardtii using genome-wide approaches. Photosynthesis Research **75,** 111-125.

Ishida, S., Muto, S., and Miyachi, S. (1993). Structural-Analysis of Periplasmic Carbonic Anhydrase-1 of Chlamydomonas-Reinhardtii. European Journal of Biochemistry **214,** 9-16.

Jinek, M., Chylinski, K., Fonfara, I., Hauer, M., Doudna, J.A., and Charpentier, E. (2012). A programmable dual-RNA-guided DNA endonuclease in adaptive bacterial immunity. Science **337,** 816-821.

Johnson, L.S., Eddy, S.R., and Portugaly, E. (2010). Hidden Markov model speed heuristic and iterative HMM search procedure. Bmc Bioinformatics **11**.

Johnston, S.D., Lew, J.E., and Berman, J. (1999). Gbp1p, a protein with RNA recognition motifs, binds single-stranded telomeric DNA and changes its binding specificity upon dimerization. Mol Cell Biol **19,** 923-933.

Jung, S., Lee, Y., Yang, K., Lee, S.B., Jang, S.M., Ha, S.B., and Back, K. (2004). Dual targeting of Myxococcus xanthus protoporphyrinogen oxidase into chloroplasts and mitochondria and high level oxyfluorfen resistance. Plant Cell and Environment **27,** 1436-1446.

Karlsson, J., Hiltonen, T., Husic, H.D., Ramazanov, Z., and Samuelsson, G. (1995). Intracellular Carbonic-Anhydrase of Chlamydomonas-Reinhardtii. Plant Physiol **109,** 533-539.

Karlsson, J., Clarke, A.K., Chen, Z.Y., Hugghins, S.Y., Park, Y.I., Husic, H.D., . . . Samuelsson, G. (1998). A novel alpha-type carbonic anhydrase associated with the thylakoid membrane in Chlamydomonas reinhardtii is required for growth at ambient CO2. Embo Journal **17,** 1208-1216.

Karpowicz, S.J., Prochnik, S.E., Grossman, A.R., and Merchant, S.S. (2011). The GreenCut2 Resource, a Phylogenomically Derived Inventory of Proteins Specific to the Plant Lineage. Journal of Biological Chemistry **286,** 21427-21439.

Kataoka, M., Sato, R., and Oshio, H. (1990). Isolation and partial characterization of mutant Chlamydomonas reinhardtii resistant to herbicide S-23142. Journal of Pesticide Science **15,** 449-451.

Kimpel, D.L., Togasaki, R.K., and Miyachi, S. (1983). Carbonic-Anhydrase in Chlamydomonas-Reinhardtii .1. Localization. Plant and Cell Physiology **24,** 255-259.

Kindle, K.L. (1990). High-frequency nuclear transformation of Chlamydomonas reinhardtii. Proc Natl Acad Sci U S A **87,** 1228-1232.

Klodmann, J., Sunderhaus, S., Nimtz, M., Jansch, L., and Braun, H.P. (2010). Internal Architecture of Mitochondrial Complex I from Arabidopsis thaliana. Plant Cell **22,** 797-810.

Klug, A. (2010). The discovery of zinc fingers and their applications in gene regulation and genome manipulation. Annu Rev Biochem **79,** 213-231.

Kohinata, T., Nishino, H., and Fukuzawa, H. (2008). Significance of zinc in a regulatory protein, CCM1, which regulates the carbon-concentrating mechanism in Chlamydomonas reinhardtii. Plant and Cell Physiology **49,** 273-283.

Kropat, J., Hong-Hermesdorf, A., Casero, D., Ent, P., Castruita, M., Pellegrini, M., . . . Malasarn, D. (2011). A revised mineral nutrient supplement increases biomass and growth rate in Chlamydomonas reinhardtii. Plant Journal **66,** 770-780.

Kuchitsu, K., Tsuzuki, M., and Miyachi, S. (1988). Changes of Starch Localization within the Chloroplast Induced by Changes in Co2 Concentration during Growth of Chlamydomonas-Reinhardtii - Independent Regulation of Pyrenoid Starch and Stroma Starch. Plant and Cell Physiology **29,** 1269-1278.

Kuchitsu, K., Tsuzuki, M., and Miyachi, S. (1991). Polypeptide Composition and Enzyme-Activities of the Pyrenoid and Its Regulation by Co2 Concentration in Unicellular Green-Algae. Canadian Journal of Botany-Revue Canadienne De Botanique **69,** 1062-1069.

Kucho, K., Yoshioka, S., Taniguchi, F., Ohyama, K., and Fukuzawa, H. (2003). Cis-acting elements and DNA-binding proteins involved in CO2-responsive transcriptional activation of Cah1 encoding a periplasmic carbonic anhydrase Chlamydomonas reinhardtii. Plant Physiol **133,** 783-793.

Kustu, S., and Inwood, W. (2006). Biological gas channels for NH3 andCO(2): evidence that Rh (rhesus) proteins areCO(2) channels. Transfusion Clinique Et Biologique **13,** 103-110.

Kuzma, J., and Kokotovich, A. (2011). Renegotiating GM crop regulation. Targeted gene-modification technology raises new issues for the oversight of genetically modified crops. EMBO Rep **12,** 883-888.

Labadorf, A., Link, A., Rogers, M.F., Thomas, J., Reddy, A.S.N., and Ben-Hur, A. (2010). Genome-wide analysis of alternative splicing in Chlamydomonas reinhardtii. Bmc Genomics **11**.

Ladunga, I. (2010). An Overview of the Computational Analyses and Discovery of Transcription Factor Binding Sites. Computational Biology of Transcription Factor Binding **674,** 1-22.

Lei, Y., Guo, X., Liu, Y., Cao, Y., Deng, Y., Chen, X., . . . Zhao, H. (2012). Efficient targeted gene disruption in Xenopus embryos using engineered transcription activator-like effector nucleases (TALENs). Proc Natl Acad Sci U S A **109,** 17484-17489.

Leon, R., Couso, I., and Fernandez, E. (2007). Metabolic engineering of ketocarotenoids biosynthesis in the unicelullar microalga Chlamydomonas reinhardtii. J Biotechnol **130,** 143-152.

Lermontova, I., and Grimm, B. (2000). Overexpression of plastidic protoporphyrinogen IX oxidase leads to resistance to the diphenyl-ether herbicide acifluorfen. Plant Physiol **122,** 75-84.

Li, B., Ruotti, V., Stewart, R.M., Thomson, J.A., and Dewey, C.N. (2010). RNA-Seq gene expression estimation with read mapping uncertainty. Bioinformatics **26,** 493-500.

Li, L., Wu, L.P., and Chandrasegaran, S. (1992). Functional domains in Fok I restriction endonuclease. Proc Natl Acad Sci U S A **89,** 4275-4279.

Li, T., Huang, S., Zhou, J., and Yang, B. (2013). Designer TAL effectors induce disease susceptibility and resistance to Xanthomonas oryzae pv. oryzae in rice. Mol Plant **6,** 781-789.

Li, T., Liu, B., Spalding, M.H., Weeks, D.P., and Yang, B. (2012). High-efficiency TALEN-based gene editing produces disease-resistant rice. Nat Biotechnol **30,** 390-392.

Li, T., Huang, S., Jiang, W.Z., Wright, D., Spalding, M.H., Weeks, D.P., and Yang, B. (2011a). TAL nucleases (TALNs): hybrid proteins composed of TAL effectors and FokI DNA-cleavage domain. Nucleic Acids Res **39,** 359-372.

Li, T., Huang, S., Zhao, X., Wright, D.A., Carpenter, S., Spalding, M.H., . . . Yang, B. (2011b). Modularly assembled designer TAL effector nucleases for targeted gene knockout and gene replacement in eukaryotes. Nucleic Acids Res **39,** 6315-6325.

Li, X., and Nicholl, D. (2005). Development of PPO inhibitor-resistant cultures and crops. Pest Manag Sci **61,** 277-285.

Linden, H., Sandmann, G., Chamovitz, D., Hirschberg, J., and Boger, P. (1989). iochemical characterization of herbicide resistant Anacystis mutants selected against the bleaching herbicide norflurazon. Pesticide Biochemistry and Physiology **36,** 46-51.

Liu, J., Li, C., Yu, Z., Huang, P., Wu, H., Wei, C., . . . Jiao, R. (2012). Efficient and specific modifications of the Drosophila genome by means of an easy TALEN strategy. J Genet Genomics **39,** 209-215.

Lopez, D., Casero, D., Cokus, S.J., Merchant, S.S., and Pellegrini, M. (2011). Algal Functional Annotation Tool: a web-based analysis suite to functionally interpret large gene lists using integrated annotation and expression data. Bmc Bioinformatics **12**.

Lumbreras, V., Stevens, D.R., and Purton, S. (1998). Efficient foreign gene expression in Chlamydomonas reinhardtii mediated by an endogenous intron. Plant Journal **14,** 441-447.

Ma, S., Zhang, S., Wang, F., Liu, Y., Xu, H., Liu, C., . . . Xia, Q. (2012). Highly efficient and specific genome editing in silkworm using custom TALENs. PLoS One **7,** e45035.

Ma, Y.B., Pollock, S.V., Xiao, Y., Cunnusamy, K., and Moroney, J.V. (2011). Identification of a Novel Gene, CIA6, Required for Normal Pyrenoid Formation in Chlamydomonas reinhardtii. Plant Physiol **156,** 884-896.

MacQueen, J. (1967). Some methods for classification and analysis of multivariate observations. In Proc. Fifth Berkeley Symp. on Math. Statist. and Prob. (Univ. of Calif. Press), pp. 281-297.

Mahfouz, M.M., Li, L., Shamimuzzaman, M., Wibowo, A., Fang, X., and Zhu, J.K. (2011). De novo-engineered transcription activator-like effector (TALE) hybrid nuclease with novel DNA binding specificity creates double-strand breaks. Proc Natl Acad Sci U S A **108,** 2623-2628.

Mahfouz, M.M., Li, L., Piatek, M., Fang, X., Mansour, H., Bangarusamy, D.K., and Zhu, J.K. (2012). Targeted transcriptional repression using a chimeric TALE-SRDX repressor protein. Plant Molecular Biology **78,** 311-321.

Mak, A.N., Bradley, P., Cernadas, R.A., Bogdanove, A.J., and Stoddard, B.L. (2012). The crystal structure of TAL effector PthXo1 bound to its DNA target. Science **335,** 716-719.

Mali, P., Aach, J., Stranges, P.B., Esvelt, K.M., Moosburner, M., Kosuri, S., . . . Church, G.M. (2013a). CAS9 transcriptional activators for target specificity screening and paired nickases for cooperative genome engineering. Nat Biotechnol **31,** 833-838.

Mali, P., Yang, L., Esvelt, K.M., Aach, J., Guell, M., DiCarlo, J.E., . . . Church, G.M. (2013b). RNA-Guided Human Genome Engineering via Cas9. Science **339,** 823-826.

Mannerlöf, M., Tuvesson, S., Steen, P., and Tenning, P. (1997). Transgenic sugar beet tolerant to glyphosate. Euphytica **94,** 83-91.

Marek, L.F., and Spalding, M.H. (1991). Changes in Photorespiratory Enzyme-Activity in Response to Limiting Co2 in Chlamydomonas-Reinhardtii. Plant Physiol **97,** 420-425.

Margulies, M., Egholm, M., Altman, W.E., Attiya, S., Bader, J.S., Bemben, L.A., . . . Rothberg, J.M. (2005). Genome sequencing in microfabricated high-density picolitre reactors. Nature **437,** 376-380.

Mariscal, V., Moulin, P., Orsel, M., Miller, A.J., Fernandez, E., and Galvan, A. (2006). Differential regulation of the Chlamydomonas Nar1 gene family by carbon and nitrogen. Protist **157,** 421-433.

Martinez-Ferez, I.M., and Vioque, A. (1992). Nucleotide sequence of the phytoene desaturase gene from Synechocystis sp. PCC 6803 and characterization of a new mutation which confers resistance to the herbicide norflurazon. Plant Molecular Biology **18,** 981-983.

Matringe, M., Camadro, J.M., and Brouillet, N. (1993). Protoporphyrinogen oxidase, the molecular target site for peroxidizing herbicides. Proceedings of the Brighton Crop Protection Conference: Weeds**,** 703-711.

Matsuda, Y., and Colman, B. (1995). Induction of Co2 and Bicarbonate Transport in the Green-Alga Chlorella-Ellipsoidea .2. Evidence for Induction in Response to External Co2 Concentration. Plant Physiol **108,** 253-260.

Matsumoto, H. (2002). Inhibitors of protoporphyrinogen oxidase: a brief update. In Herbicide Classes in Development, P. Boger, K. Wakabayashi, and K. Hirai, eds (Berlin-Heidelberg: Springer), pp. 151-161.

Matys, V., Kel-Margoulis, O.V., Fricke, E., Liebich, I., Land, S., Barre-Dirrie, A., . . . Wingender, E. (2006). TRANSFAC (R) and its module TRANSCompel (R): transcriptional gene regulation in eukaryotes. Nucleic Acids Res **34,** D108-D110.

Merchant, S.S., Prochnik, S.E., Vallon, O., Harris, E.H., Karpowicz, S.J., Witman, G.B., . . . Team, J.A. (2007). The Chlamydomonas genome reveals the evolution of key animal and plant functions. Science **318,** 245-251.

Michel, A., Scheffler, B.E., Netherland, M., Dayan, F.E., and Arias de Ares, R.S. (2004a). Herbicide-Resistant Plants, and Polynucleotides and Methods for Providing Same.

Michel, A., Arias, R.S., Scheffler, B.E., Duke, S.O., Netherland, M., and Dayan, F.E. (2004b). Somatic mutation-mediated evolution of herbicide resistance in the nonindigenous invasive plant hydrilla (Hydrilla verticillata). Mol Ecol **13,** 3229-3237.

Miller, J.C., Tan, S., Qiao, G., Barlow, K.A., Wang, J., Xia, D.F., . . . Rebar, E.J. (2011). A TALE nuclease architecture for efficient genome editing. Nat Biotechnol **29,** 143-148.

Misawa, N., Yamano, S., Linden, H., de Felipe, M.R., Lucas, M., Ikenaga, H., and Sandmann, G. (1993). Functional expression of the Erwinia uredovora carotenoid biosynthesis gene crtl in transgenic plants showing an increase of beta-carotene biosynthesis activity and resistance to the bleaching herbicide norflurazon. Plant J **4,** 833-840.

Mitra, M., Lato, S.M., Ynalvez, R.A., Xiao, Y., and Moroney, J.V. (2004). Identification of a new chloroplast carbonic anhydrase in Chlamydomonas reinhardtii. Plant Physiol **135,** 173-182.

Miura, K., Yamano, T., Yoshioka, S., Kohinata, T., Inoue, Y., Taniguchi, F., . . . Fukuzawa, H. (2004). Expression profiling-based identification of CO2-responsive genes regulated by CCM1 controlling a carbon-concentrating mechanism in Chlamydomonas reinhardtii. Plant Physiol **135,** 1595-1607.

Monsanto. (2002). Safety assessment of roundup ready canola event GT73.

Monsanto. (2005). Backgrounder - History of Monsanto's Glyphosate Herbicides.

Moore, F.E., Reyon, D., Sander, J.D., Martinez, S.A., Blackburn, J.S., Khayter, C., . . . Langenau, D.M. (2012). Improved somatic mutagenesis in zebrafish using transcription activator-like effector nucleases (TALENs). PLoS One **7,** e37877.

Morbitzer, R., Romer, P., Boch, J., and Lahaye, T. (2010). Regulation of selected genome loci using de novo-engineered transcription activator-like effector (TALE)-type transcription factors. Proc Natl Acad Sci U S A **107,** 21617-21622.

Morbitzer, R., Elsaesser, J., Hausner, J., and Lahaye, T. (2011). Assembly of custom TALE-type DNA binding domains by modular cloning. Nucleic Acids Res **39,** 5790-5799.

Morgan, R.R., Errington, R., and Elder, G.H. (2004). Identification of sequences required for the import of human protoporphyrinogen oxidase to mitochondria. Biochemical Journal **377,** 281-287.

Morita, E., Kuroiwa, H., Kuroiwa, T., and Nozaki, H. (1997). High localization of ribulose-1,5-bisphosphate carboxylase/oxygenase in the pyrenoids of Chlamydomonas reinhardtii (Chlorophyta), as revealed by cryofixation and immunogold electron microscopy. Journal of Phycology **33,** 68-72.

Moroney, J.V., and Somanchi, A. (1999). How do algae concentrate CO2 to increase the efficiency of photosynthetic carbon fixation? Plant Physiol **119,** 9-16.

Moroney, J.V., and Ynalvez, R.A. (2007). Proposed carbon dioxide concentrating mechanism in Chlamydomonas reinhardtii. Eukaryotic Cell **6,** 1251-1259.

Moroney, J.V., Husic, H.D., and Tolbert, N.E. (1985). Effect of Carbonic-Anhydrase Inhibitors on Inorganic Carbon Accumulation by Chlamydomonas-Reinhardtii. Plant Physiol **79,** 177-183.

Moroney, J.V., Tolbert, N.E., and Sears, B.B. (1986). Complementation Analysis of the Inorganic Carbon Concentrating Mechanism of Chlamydomonas-Reinhardtii. Molecular & General Genetics **204,** 199-203.

Moroney, J.V., Husic, H.D., Tolbert, N.E., Kitayama, M., Manuel, L.J., and Togasaki, R.K. (1989). Isolation and Characterization of a Mutant of Chlamydomonas-Reinhardtii Deficient in the Co2 Concentrating Mechanism. Plant Physiol **89,** 897-903.

Moroney, J.V., Ma, Y.B., Frey, W.D., Fusilier, K.A., Pham, T.T., Simms, T.A., . . . Mukherjee, B. (2011). The carbonic anhydrase isoforms of Chlamydomonas reinhardtii: intracellular location, expression, and physiological roles. Photosynthesis Research **109,** 133-149.

Moscou, M.J., and Bogdanove, A.J. (2009). A simple cipher governs DNA recognition by TAL effectors. Science **326,** 1501.

Mukherjee, B. (2013). Investigations of the role of putative inorganic carbon transporters in the carbon dioxide concentrating mechanisms of Chlamydomonas reinhardtii. In Department of Biological Sciences (Louisiana State University Electronic Thesis & Dissertation Collection: Louisiana State University).

Murakami, M.T., Sforca, M.L., Neves, J.L., Paiva, J.H., Domingues, M.N., Pereira, A.L., . . . Benedetti, C.E. (2010). The repeat domain of the type III effector protein PthA shows a TPR-like structure and undergoes conformational changes upon DNA interaction. Proteins **78,** 3386-3395.

Nakajima, K., Ikenaka, K., Nakahira, K., Morita, N., and Mikoshiba, K. (1993). An Improved Retroviral Vector for Assaying Promoter Activity - Analysis of Promoter Interference in Pip211 Vector. Febs Letters **315,** 129-133.

Norris, S.R., Barrette, T.R., and DellaPenna, D. (1995). Genetic dissection of carotenoid synthesis in arabidopsis defines plastoquinone as an essential component of phytoene desaturation. Plant Cell **7,** 2139-2149.

Ohnishi, N., Mukherjee, B., Tsujikawa, T., Yanase, M., Nakano, H., Moroney, J.V., and Fukuzawa, H. (2010). Expression of a Low CO2-Inducible Protein, LCI1, Increases Inorganic Carbon Uptake in the Green Alga Chlamydomonas reinhardtii. Plant Cell **22,** 3105-3117.

Okuda, S., Yamada, T., Hamajima, M., Itoh, M., Katayama, T., Bork, P., . . . Kanehisa, M. (2008). KEGG Atlas mapping for global analysis of metabolic pathways. Nucleic Acids Res **36,** W423-W426.

Osakabe, K., Osakabe, Y., and Toki, S. (2010). Site-directed mutagenesis in Arabidopsis using custom-designed zinc finger nucleases. Proc Natl Acad Sci U S A **107,** 12034-12039.

Padgette, S.R., Kolacz, K.H., Delannay, X., Re, D.B., LaVallee, B.J., Tinius, C.N., . . . Kishore, G.M. (1995). Development, Identification, and Characterization of a Glyphosate-Tolerant Soybean Line. Crop Sci. **35,** 1451-1461.

Paques, F., and Duchateau, P. (2007). Meganucleases and DNA double-strand break-induced recombination: perspectives for gene therapy. Curr Gene Ther **7,** 49-66.

Park, H., Hilsenbeck, J.L., Kim, H.J., Shuttleworth, W.A., Park, Y.H., Evans, J.N., and Kang, C. (2004). Structural studies of Streptococcus pneumoniae EPSP synthase in unliganded state, tetrahedral intermediate-bound state and S3P-GLP-bound state. Molecular Microbiology **51,** 963-971.

Peers, G., Truong, T.B., Ostendorf, E., Busch, A., Elrad, D., Grossman, A.R., . . . Niyogi, K.K. (2009). An ancient light-harvesting protein is critical for the regulation of algal photosynthesis. Nature **462,** 518-U215.

Pennisi, E. (2013). The CRISPR Craze. Science **341,** 833-836.

Plessis, A., Perrin, A., Haber, J.E., and Dujon, B. (1992). Site-specific recombination determined by I-SceI, a mitochondrial group I intron-encoded endonuclease expressed in the yeast nucleus. Genetics **130,** 451-460.

Plevin, M.J., Mills, M.M., and Ikura, M. (2005). The LxxLL motif: a multifunctional binding sequence in transcriptional regulation. Trends in Biochemical Sciences **30,** 66-69.

Pollegioni, L., Schonbrunn, E., and Siehl, D. (2011). Molecular basis of glyphosate resistance-different approaches through protein engineering. FEBS J **278,** 2753-2766.

Pollock, S.V., Prout, D.L., Godfrey, A.C., Lemaire, S.D., and Moroney, J.V. (2004). The Chlamydomonas reinhardtii proteins Ccp1 and Ccp2 are required for long-term growth, but are not necessary for efficient photosynthesis, in a low-CO2 environment. Plant Molecular Biology **56,** 125-132.

Pravica, V., Perrey, C., Stevens, A., Lee, J.H., and Hutchinson, I.V. (2000). A single nucleotide polymorphism in the first intron of the human IFN-gamma gene: Absolute correlation with a polymorphic CA microsatellite marker of high IFN-gamma production. Human Immunology **61,** 863-866.

Prins, H.B.A., and Elzenga, J.T.M. (1989). Bicarbonate Utilization - Function and Mechanism. Aquatic Botany **34,** 59-83.

Pronina, N.A., and Semenenko, V.E. (1992). Carbonic-Anhydrase Activity and Fatty-Acid Composition of Photosystem Deficient and High Co2 Required Mutants of Chlamydomonas-Reinhardtii. Photosynthesis Research **34,** 201-201.

Proost, S., Van Bel, M., Sterck, L., Billiau, K., Van Parys, T., Van de Peer, Y., and Vandepoele, K. (2009). PLAZA: A Comparative Genomics Resource to Study Gene and Genome Evolution in Plants. Plant Cell **21,** 3718-3731.

Punta, M., Coggill, P.C., Eberhardt, R.Y., Mistry, J., Tate, J., Boursnell, C., . . . Finn, R.D. (2012). The Pfam protein families database. Nucleic Acids Res **40,** D290-D301.

Qasem, J.R., and Foy, C.L. (2006). Selective weed control in syrian marjoram (Origanum syriacum) with oxadiazon and oxyfluorfen herbicides. Weed Technology **20,** 670-676.

Ramazanov, Z., and Cardenas, J. (1992). Involvement of Photorespiration and Glycolate Pathway in Carbonic-Anhydrase Induction and Inorganic Carbon Concentration in Chlamydomonas-Reinhardtii. Physiologia Plantarum **84,** 502-508.

Ramazanov, Z., Rawat, M., Henk, M.C., Mason, C.B., Matthews, S.W., and Moroney, J.V. (1994). The Induction of the Co2-Concentrating Mechanism Is Correlated with the Formation of the Starch Sheath around the Pyrenoid of Chlamydomonas-Reinhardtii. Planta **195,** 210-216.

Randolph-Anderson, B.L., Sato, R., Johnson, A.M., Harris, E.H., Hauser, C.R., Oeda, K., . . . Boynton, J.E. (1998). Isolation and characterization of a mutant protoporphyrinogen oxidase gene from Chlamydomonas reinhardtii conferring resistance to porphyric herbicides. Plant Molecular Biology **38,** 839-859.

Rasala, B.A., Lee, P.A., Shen, Z., Briggs, S.P., Mendez, M., and Mayfield, S.P. (2012). Robust expression and secretion of Xylanase1 in Chlamydomonas reinhardtii by fusion to a selection gene and processing with the FMDV 2A peptide. PLoS One **7,** e43349.

Rawat, M., and Moroney, J.V. (1991). Partial Characterization of a New Isoenzyme of Carbonic-Anhydrase Isolated from Chlamydomonas-Reinhardtii. Journal of Biological Chemistry **266,** 9719-9723.

Rawat, M., Henk, M.C., Lavigne, L.L., and Moroney, J.V. (1996). Chlamydomonas reinhardtii mutants without ribulose-1,5-bisphosphate carboxylase-oxygenase lack a detectable pyrenoid. Planta **198,** 263-270.

Reboud, X. (2002). Response of Chlamydomonas reinhardtii to herbicides: negative relationship between toxicity and water solubility across several herbicide families. Bull Environ Contam Toxicol **69,** 554-561.

Reyon, D., Khayter, C., Regan, M.R., Joung, J.K., and Sander, J.D. (2012). Engineering designer transcription activator-like effector nucleases (TALENs) by REAL or REAL-Fast assembly. Curr Protoc Mol Biol **Chapter 12,** Unit 12 15.

Robinson, M.D., McCarthy, D.J., and Smyth, G.K. (2010). edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. Bioinformatics **26,** 139-140.

Rolland, N., Dorne, A.J., Amoroso, G., Sultemeyer, D.F., Joyard, J., and Rochaix, J.D. (1997). Disruption of the plastid ycf10 open reading frame affects uptake of inorganic carbon in the chloroplast of Chlamydomonas. Embo Journal **16,** 6713-6726.

Rothstein, R.J. (1983). One-step gene disruption in yeast. Methods Enzymol **101,** 202-211.

Rouet, P., Smih, F., and Jasin, M. (1994). Introduction of double-strand breaks into the genome of mouse cells by expression of a rare-cutting endonuclease. Mol Cell Biol **14,** 8096-8106.

Rudin, N., Sugarman, E., and Haber, J.E. (1989). Genetic and physical analysis of double-strand break repair and recombination in Saccharomyces cerevisiae. Genetics **122,** 519-534.

Sander, J.D., Cade, L., Khayter, C., Reyon, D., Peterson, R.T., Joung, J.K., and Yeh, J.R. (2011). Targeted gene disruption in somatic zebrafish cells using engineered TALENs. Nat Biotechnol **29,** 697-698.

Sandmann, G., and Boger, P. (1997). Phytoene desaturase as a target for bleaching herbicides. In Herbicide Activity: Toxicology, Biochemistry, and Molecular Biology, R.M. Roe, J.D. Burton, and R.J. Kuhr, eds (Amsterdam: IOS Press), pp. 1-10.

Sandmann, G., and Mitchell, G. (2001). In vitro inhibition studies of phytoene desaturase by bleaching ketomorpholine derivatives. J Agric Food Chem **49,** 138-141.

Schmid-Burgk, J.L., Schmidt, T., Kaiser, V., Honing, K., and Hornung, V. (2013). A ligation-independent cloning technique for high-throughput assembly of transcription activator-like effector genes. Nat Biotechnol **31,** 76-81.

Schonbrunn, E., Eschenburg, S., Shuttleworth, W.A., Schloss, J.V., Amrhein, N., Evans, J.N., and Kabsch, W. (2001). Interaction of the herbicide glyphosate with its target enzyme 5-enolpyruvylshikimate 3-phosphate synthase in atomic detail. Proc Natl Acad Sci U S A **98,** 1376-1380.

Shan, Q., Wang, Y., Chen, K., Liang, Z., Li, J., Zhang, Y., . . . Gao, C. (2013). Rapid and efficient gene modification in rice and brachypodium using TALENs. Mol Plant **6,** 1365-1368.

Shimogawara, K., Fujiwara, S., Grossman, A., and Usuda, H. (1998). High-efficiency transformation of Chlamydomonas reinhardtii by electroporation. Genetics **148,** 1821-1828.

Shukla, V.K., Doyon, Y., Miller, J.C., DeKelver, R.C., Moehle, E.A., Worden, S.E., . . . Urnov, F.D. (2009). Precise genome modification in the crop species Zea mays using zinc-finger nucleases. Nature **459,** 437-441.

Siehl, D.L., Castle, L.A., Gorton, R., and Keenan, R.J. (2007). The molecular basis of glyphosate resistance by an optimized microbial acetyltransferase. J Biol Chem **282,** 11446-11455.

Sihtmäe, M., Blinova, I., Künnis-Beres, K., Kanarbik, L., Heinlaan, M., and Kahru, A. (2013). Ecotoxicological effects of different glyphosate formulations. Applied Soil Ecology **72,** 215-224.

Simonsen, L., Fomsgaard, I.S., Svensmark, B., and Spliid, N.H. (2008). Fate and availability of glyphosate and AMPA in agricultural soil. J Environ Sci Health B **43,** 365-375.

Smith, E.A., and Oehme, F.W. (1992). The biological activity of glyphosate to plants and animals: a literature review. Vet Hum Toxicol **34,** 531-543.

Smith, J., Bibikova, M., Whitby, F.G., Reddy, A.R., Chandrasegaran, S., and Carroll, D. (2000). Requirements for double-strand cleavage by chimeric restriction enzymes with zinc finger DNA-recognition domains. Nucleic Acids Res **28,** 3361-3369.

Sommer, F., Kropat, J., Malasarn, D., Grossoehme, N.E., Chen, X.H., Giedroc, D.P., and Merchant, S.S. (2010). The CRR1 Nutritional Copper Sensor in Chlamydomonas Contains Two Distinct Metal-Responsive Domains. Plant Cell **22,** 4098-4113.

Soupene, E., Inwood, W., and Kustu, S. (2004). Lack of the Rhesus protein Rh1 impairs growth of the green alga Chlamydomonas reinhardtii at high CO2. Proceedings of the National Academy of Sciences of the United States of America **101,** 7787-7792.

Spalding, M.H. (2008). Microalgal carbon-dioxide-concentrating mechanisms: Chlamydomonas inorganic carbon transporters. Journal of Experimental Botany **59,** 1463-1473.

Spalding, M.H., and Ogren, W.L. (1982). Photosynthesis Is Required for Induction of the Co2-Concentrating System in Chlamydomonas-Reinhardii. Febs Letters **145,** 41-44.

Spalding, M.H., Spreitzer, R.J., and Ogren, W.L. (1983a). Carbonic Anhydrase-Deficient Mutant of Chlamydomonas-Reinhardii Requires Elevated Carbon-Dioxide Concentration for Photoautotrophic Growth. Plant Physiol **73,** 268-272.

Spalding, M.H., Spreitzer, R.J., and Ogren, W.L. (1983b). Reduced Inorganic Carbon Transport in a CO(2)-Requiring Mutant of Chlamydomonas reinhardii. Plant Physiol **73,** 273-276.

Spencer, K.G., Kimpel, D.L., Fisher, M.L., Togasaki, R.K., and Miyachi, S. (1983). Carbonic-Anhydrase in Chlamydomonas-Reinhardtii .2. Requirements for Carbonic-Anhydrase Induction. Plant and Cell Physiology **24,** 301-304.

Stanke, M., and Waack, S. (2003). Gene prediction with a hidden Markov model and a new intron submodel. Bioinformatics **19,** Ii215-Ii225.

Steinrucken, H.C., and Amrhein, N. (1980). The herbicide glyphosate is a potent inhibitor of 5-enolpyruvyl-shikimic acid-3-phosphate synthase. Biochem Biophys Res Commun **94,** 1207-1212.

Stracke, R., Jahns, O., Keck, M., Tohge, T., Niehaus, K., Fernie, A.R., and Weisshaar, B. (2010a). Analysis of PRODUCTION OF FLAVONOL GLYCOSIDES-dependent flavonol glycoside accumulation in Arabidopsis thaliana plants reveals MYB11-, MYB12-and MYB111-independent flavonol glycoside accumulation. New Phytologist **188,** 985-1000.

Stracke, R., Favory, J.J., Gruber, H., Bartelniewoehner, L., Bartels, S., Binkert, M., . . . Ulm, R. (2010b). The Arabidopsis bZIP transcription factor HY5 regulates expression of the PFG1/MYB12 gene in response to light and ultraviolet-B radiation. Plant Cell and Environment **33,** 88-103.

Subramanian, A., Tamayo, P., Mootha, V.K., Mukherjee, S., Ebert, B.L., Gillette, M.A., . . . Mesirov, J.P. (2005). Gene set enrichment analysis: A knowledge-based approach for interpreting genome-wide expression profiles. Proceedings of the National Academy of Sciences of the United States of America **102,** 15545-15550.

Sugio, A., Yang, B., Zhu, T., and White, F.F. (2007). Two type III effector genes of Xanthomonas oryzae pv. oryzae control the induction of the host genes OsTFIIAgamma1 and OsTFX1 during bacterial blight of rice. Proc Natl Acad Sci U S A **104,** 10720-10725.

Sunderhaus, S., Dudkina, N.V., Jansch, L., Klodmann, J., Heinemeyer, J., Perales, M., . . . Braun, H.P. (2006). Carbonic anhydrase subunits form a matrix-exposed domain attached to the membrane arm of mitochondrial complex I in plants. Journal of Biological Chemistry **281,** 6482-6488.

Sung, Y.H., Baek, I.J., Kim, D.H., Jeon, J., Lee, J., Lee, K., . . . Lee, H.W. (2013). Knockout mice created by TALEN-mediated gene targeting. Nat Biotechnol **31,** 23-24.

Tesson, L., Usal, C., Menoret, S., Leung, E., Niles, B.J., Remy, S., . . . Cost, G.J. (2011). Knockout rats generated by embryo microinjection of TALENs. Nat Biotechnol **29,** 695-696.

Thyssen, C., Schlichting, R., and Giersch, C. (2001). The CO2-concentrating mechanism in the physiological context: lowering the CO2 supply diminishes culture growth and economises starch utilisation in Chlamydomonas reinhardtii. Planta **213,** 629-639.

Tong, C., Huang, G., Ashton, C., Wu, H., Yan, H., and Ying, Q.L. (2012). Rapid and cost-effective gene targeting in rat embryonic stem cells by TALENs. J Genet Genomics **39,** 275-280.

Townsend, J.A., Wright, D.A., Winfrey, R.J., Fu, F., Maeder, M.L., Joung, J.K., and Voytas, D.F. (2009). High-frequency modification of plant genes using engineered zinc-finger nucleases. Nature **459,** 442-445.

Tran, P.T., Sharifi, M.N., Poddar, S., Dent, R.M., and Niyogi, K.K. (2012). Intragenic enhancers and suppressors of phytoene desaturase mutations in Chlamydomonas reinhardtii. PLoS One **7,** e42196.

Trapnell, C., Pachter, L., and Salzberg, S.L. (2009). TopHat: discovering splice junctions with RNA-Seq. Bioinformatics **25,** 1105-1111.

Trinklein, N.D., Aldred, S.F., Hartman, S.J., Schroeder, D.I., Otillar, R.P., and Myers, R.M. (2004). An abundance of bidirectional promoters in the human genome. Genome Research **14,** 62-66.

Tsui, M.T., and Chu, L.M. (2003). Aquatic toxicity of glyphosate-based formulations: comparison between different organisms and the effects of environmental factors. Chemosphere **52,** 1189-1197.

Tsuzuki, M., and Miyachi, S. (1989). The Function of Carbonic-Anhydrase in Aquatic Photosynthesis. Aquatic Botany **34,** 85-104.

Van, K., and Spalding, M.H. (1999). Periplasmic carbonic anhydrase structural gene (Cah1) mutant in Chlamydomonas reinhardtii. Plant Physiol **120,** 757-764.

Van, K.J., Wang, Y.J., Nakamura, Y., and Spalding, M.H. (2001). Insertional mutants of Chlamydomonas reinhardtii that require elevated CO2 for survival. Plant Physiol **127,** 607-614.

Vencill, W.K. (2002). Herbicide Handbook. (Lawrence, KS, USA: Weed Science Society of America).

Verdier, V., Triplett, L.R., Hummel, A.W., Corral, R., Cernadas, R.A., Schmidt, C.L., . . . Leach, J.E. (2012). Transcription activator-like (TAL) effectors targeting OsSWEET genes enhance virulence on diverse rice (Oryza sativa) varieties when expressed individually in a TAL effector-deficient strain of Xanthomonas oryzae. New Phytologist **196,** 1197-1207.

Villand, P., Eriksson, M., and Samuelsson, G. (1997). Carbon dioxide and light regulation of promoters controlling the expression of mitochondrial carbonic anhydrase in Chlamydomonas reinhardtii. Biochemical Journal **327,** 51-57.

Villarejo, A., Martinez, F., and Ramazanov, Z. (1997). Effect of aminooxyacetate, an inhibitor blocking the glycolate pathway, on the induction of a CO2-concentrating mechanism and low-CO2-inducible polypeptides in Chlamydomonas reinhardtii (Chlorophyta). European Journal of Phycology **32,** 141-145.

Wagner, T., Windhovel, U., and Romer, S. (2002). Bansformation of tobacco with a mutated cyanobacterial phytoene desaturase gene confers resistance to bleaching herbicides. Z Naturforsch C **57,** 671-679.

Wakabayashi, K., and Boger, P. (1999). General physiological characteristics and mode of action of peroxidizing herbicides. In Herbicide Classes in Development Mode of Action, Targets, Genetic Engineering, Chemistry, P. Boger, K. Wakabayashi, and K. Hirai, eds (Berlin, Heidelberg: Springer Berlin Heidelberg,), pp. 164-190.

Wang, Y., Sun, Z.H., Horken, K.M., Im, C.S., Xiang, Y.B., Grossman, A.R., and Weeks, D.P. (2005). Analyses of CIA5, the master regulator of the carbon-concentrating mechanism in Chlamydomonas reinhardtii, and its control of gene expression. Canadian Journal of Botany-Revue Canadienne De Botanique **83,** 765-779.

Wang, Y.J., and Spalding, M.H. (2006). An inorganic carbon transport system responsible for acclimation specific to air levels Of CO2 in Chiamlydomonas reinhardtii. Proceedings of the National Academy of Sciences of the United States of America **103,** 10110-10115.

Wang, Y.J., Duanmu, D.Q., and Spalding, M.H. (2011). Carbon dioxide concentrating mechanism in Chlamydomonas reinhardtii: inorganic carbon transport and CO2 recapture. Photosynthesis Research **109,** 115-122.

Wang, Z., Li, J., Huang, H., Wang, G., Jiang, M., Yin, S., . . . Xi, J.J. (2012). An integrated chip for the high-throughput synthesis of transcription activator-like effectors. Angew Chem Int Ed Engl **51,** 8505-8508.

Warabi, E., Usui, K., Tanaka, Y., and Matsumoto, H. (2001). Resistance of a soybean cell line to oxyfluorfen by overproduction of mitochondrial protoporphyrinogen oxidase. Pest Manag Sci **57,** 743-748.

Watanabe, N., Che, F.S., Iwano, M., Takayama, S., Yoshida, S., and Isogai, A. (2001). Dual targeting of spinach protoporphyrinogen oxidase II to mitochondria and

chloroplasts by alternative use of two in-frame initiation codons. J Biol Chem **276,** 20474-20481.

Waterston, R.H., Lander, E.S., and Sulston, J.E. (2002). On the sequencing of the human genome. Proceedings of the National Academy of Sciences of the United States of America **99,** 3712-3716.

Weber, E., Gruetzner, R., Werner, S., Engler, C., and Marillonnet, S. (2011). Assembly of designer TAL effectors by Golden Gate cloning. PLoS One **6,** e19722.

Williams, G.M., Kroes, R., and Munro, I.C. (2000). Safety evaluation and risk assessment of the herbicide Roundup and its active ingredient, glyphosate, for humans. Regul Toxicol Pharmacol **31,** 117-165.

Wood, A.J., Lo, T.W., Zeitler, B., Pickle, C.S., Ralston, E.J., Lee, A.H., . . . Meyer, B.J. (2011). Targeted genome editing across species using ZFNs and TALENs. Science **333,** 307.

WS-DNR. (2012). Fluridone Chemical Fact Sheet (Wisconsin Department of Natural Resources).

Xiang, Y.B., Zhang, J., and Weeks, D.P. (2001). The Cia5 gene controls formation of the carbon concentrating mechanism in Chlamydomonas reinhardtii. Proceedings of the National Academy of Sciences of the United States of America **98,** 5341-5346.

Yamano, T., and Fukuzawa, H. (2009). Carbon-concentrating mechanism in a green alga, Chlamydomonas reinhardtii, revealed by transcriptome analyses. Journal of Basic Microbiology **49,** 42-51.

Yamano, T., Miura, K., and Fukuzawa, H. (2008). Expression analysis of genes associated with the induction of the carbon-concentrating mechanism in Chlamydomonas reinhardtii. Plant Physiol **147,** 340-354.

Yamano, T., Fujita, A., and Fukuzawa, H. (2011). Photosynthetic characteristics of a multicellular green alga Volvox carteri in response to external CO2 levels possibly regulated by CCM1/CIA5 ortholog. Photosynthesis Research **109,** 151-159.

Yamano, T., Tsujikawa, T., Hatano, K., Ozawa, S., Takahashi, Y., and Fukuzawa, H. (2010). Light and Low-CO2-Dependent LCIB-LCIC Complex Localization in the Chloroplast Supports the Carbon-Concentrating Mechanism in Chlamydomonas reinhardtii. Plant and Cell Physiology **51,** 1453-1468.

Yang, B., Sugio, A., and White, F.F. (2006). Os8N3 is a host disease-susceptibility gene for bacterial blight of rice. Proc Natl Acad Sci U S A **103,** 10503-10508.

Yilmaz, A., Mejia-Guerra, M.K., Kurz, K., Liang, X.Y., Welch, L., and Grotewold, E. (2011). AGRIS: the Arabidopsis Gene Regulatory Information Server, an update. Nucleic Acids Res **39,** D1118-D1122.

Ynalvez, R.A., Xiao, Y., Ward, A.S., Cunnusamy, K., and Moroney, J.V. (2008). Identification and characterization of two closely related beta-carbonic anhydrases from Chlamydomonas reinhardtii. Physiologia Plantarum **133,** 15-26.

Yoo, J.H., Park, C.Y., Kim, J.C., Heo, W.D., Cheong, M.S., Park, H.C., . . . Cho, M.J. (2005). Direct interaction of a divergent CaM isoform and the transcription factor, MYB2, enhances salt tolerance in Arabidopsis. Journal of Biological Chemistry **280,** 3697-3706.

Yoshioka, S., Taniguchi, F., Miura, K., Inoue, T., Yamano, T., and Fukuzawa, H. (2004). The novel Myb transcription factor LCR1 regulates the CO2-responsive gene Cah1, encoding a periplasmic carbonic anhydrase in Chlamydomonas reinhardtii. Plant Cell **16,** 1466-1477.

Yu, Y., Streubel, J., Balzergue, S., Champion, A., Boch, J., Koebnik, R., . . . Szurek, B. (2011). Colonization of rice leaf blades by an African strain of Xanthomonas oryzae pv. oryzae depends on a new TAL effector that induces the rice nodulin-3 Os11N3 gene. Mol Plant Microbe Interact **24,** 1102-1113.

Zhang, F., Cong, L., Lodato, S., Kosuri, S., Church, G.M., and Arlotta, P. (2011). Efficient construction of sequence-specific TAL effectors for modulating mammalian transcription. Nat Biotechnol **29,** 149-153.

Zhang, F., Maeder, M.L., Unger-Wallace, E., Hoshaw, J.P., Reyon, D., Christian, M., . . . Voytas, D.F. (2010). High frequency targeted mutagenesis in Arabidopsis thaliana using zinc finger nucleases. Proc Natl Acad Sci U S A **107,** 12028-12033.

Zhang, Y., Zhang, F., Li, X., Baller, J.A., Qi, Y., Starker, C.G., . . . Voytas, D.F. (2013). Transcription activator-like effector nucleases enable efficient plant genome engineering. Plant Physiol **161,** 20-27.

Zhou, H., Arrowsmith, J.W., Fromm, M.E., Hironaka, C.M., Taylor, M.L., Rodriguez, D., . . . Fry, J.E. (1995). Glyphosate-tolerant CP4 and GOX genes as a selectable marker in wheat transformation. Plant Cell Reports **15,** 159-163.

Zhu, H., Lau, C.H., Goh, S.L., Liang, Q., Chen, C., Du, S., . . . Wang, S. (2013). Baculoviral transduction facilitates TALEN-mediated targeted transgene integration and Cre/LoxP cassette exchange in human-induced pluripotent stem cells. Nucleic Acids Res.

Zu, Y., Tong, X., Wang, Z., Liu, D., Pan, R., Li, Z., . . . Lin, S. (2013). TALEN-mediated precise genome modification by homologous recombination in zebrafish. Nat Methods **10,** 329-331.

APPENDIX A - Supplemental Material from Chapter Two

Supplemental material, including figures, tables, and datasets, for chapter two can be accessed from the journal Plant Cell or retrieved as supplementary files for this thesis. For direct download, visit http://www.plantcell.org/content/24/5/1860/suppl/DC1.

All Supplemental Figures and Tables are available in "Chapter 2 - Supplemental Figures and Tables" (pdf) (1,193 KB)

Supplemental Figure 1. qRT-PCR Support for RNA Sequencing Results.
Supplemental Figure 2. Overlaps among the Lists of Differentially Expressed Genes Reported by the *edgeR*, *DESeq*, and *baySeq* Tools.
Supplemental Figure 3. Genes Showing a Transient Decrease in Response to CO_2 Deprivation.
Supplemental Figure 4. Genes Showing a Transient Increase in Response to CO_2 Deprivation.
Supplemental Figure 5. Spacing and Expression of HTH Conformation Genes.
Supplemental Figure 6. Comparison of Transcript Levels from Two Closely Linked Head-To-Head Gene Pairs.
Supplemental Figure 7. Overlaps Among Sets of Genes Induced by CO_2 Deprivation.
Supplemental Table 1. Widespread DNA Sequence Motifs as Putative Transcription Factor Binding Sites in the Upstream Regions of Genes.

All Supplemental Data Sets are available in "Chapter 2 - Supplemental Data Sets" (pdf) (1,686 KB)

Supplemental Data Set 1. Comparison of Differential Regulation Observations between RNA-Seq (This Study) and Arrays.
Supplemental Data Set 2. The Statistically Significantly Upregulated Genes for 180 versus 0, 60 versus 0, and 30 versus 0 Min ASVLCO_2.
Supplemental Data Set 3. The Statistically Significantly Downregulated Genes for 180 versus 0, 60 versus 0, and 30 versus 0 Min ASVLCO_2.
Supplemental Data Set 4. Gene Ontology Categories (Ashburner et al., 2000) Significantly Enriched in Upregulated and Downregulated Genes, Respectively.
Supplemental Data Set 5. The Most Highly Correlated HTH Gene Pairs Potentially Regulated by Bidirectional Promoters.
Supplemental Data Set 6. Anticorrelated Expression in HTH Gene Pairs with a Pearson Correlation Coefficient $r \leq -0.8$.
Supplemental Data Set 7. Expression Profiles of the 12 Carbonic Anhydrase Genes.