Probing the mechanism of cyclic electron flow in the green alga *Chlamydomonas reinhardtii* using recombineering and CRISPR-based tools

University of Sheffield

Faculty of Science
Department of Molecular Biology and Biotechnology

A thesis submitted for the degree of Doctor of Philosophy

Thomas Z. Emrich-Mills

30th September 2022

Probing the mechanism of cyclic electron flow in the green alga *Chlamydomonas reinhardtii* using recombineering and CRISPR-based tools
Thomas Z. Emrich-Mills

Abstract

Understanding the mechanism of photosynthesis could aid in vital efforts to increase crop yields for the growing population. Our understanding of how photosynthesis may be improved or may be affected by the changing climate is hampered by missing links in our knowledge of the photosynthetic electron transport chain. The mechanism of cyclic electron transfer remains mysterious, with the electron transfer intermediates involved in the major form of CET (dependent on the small effector protein PGR5) yet to be discovered. This PGR5-dependent CET is the main route by which stromal electrons return to the plastoquinone pool in the thylakoid membranes and is crucial for regulating photosynthesis in plants under fluctuating environmental conditions.

In this thesis, we describe the development of genetic tools that were employed to make specific changes to the components of PGR5-dependent CET in the green alga *Chlamydomonas reinhardtii*, a useful model organism with highly similar photosynthetic machinery and regulatory mechanisms to those found in crop plants. Genetic and molecular techniques for modifying Chlamydomonas are developing quickly, making this organism a useful chassis for production of high value biomolecules as well. However, difficulties in making precise edits to the nuclear genome limit progress. In addition, Chlamydomonas genes are complex and difficult to clone, and are often poorly expressed when reintroduced to cells. These limitations represent a barrier for reverse genetics attempts to engineer changes in the genome.

We adapted a method used for the *in vivo* retrieval of DNA fragments in bacterial research for Chlamydomonas genes, known as recombineering. Evidence for the effectiveness of our recombineering method is presented in Chapter 3. The method was tested on ~200 large, complex genes from the Chlamydomonas CO_2-concentrating mechanism. These included STT7, a gene encoding an effector protein for CET implicated in activating the PGR5-dependent pathway. STT7 was tagged and purified in a recombinant strain of Chlamydomonas, forming the basis for future structural work on CET. The recombineering method was published in January 2021 and tools and methods are available to the community.

CRISPR-based gene editing tools are becoming widespread in the Chlamydomonas research. We harnessed a simple commercial CRISPR system to modify specific genes involved in CET, first by insertion-deletion of native STT7 in Chap-

ter 3, then by targeted insertion of modified genes for the ferredoxin:$NADP^+$ reductase, FNR. Evidence is presented in Chapter 4 for the generation of mutant strains expressing modified FNR tethered to PSI. The intended effect was to prevent tethered FNR from making directly contact with the thylakoid membrane but to still maintain ferredoxin:$NADP^+$ reductase activity. We hypothesized that loss of native FNR might disrupt CET, while the tethered FNR might leave linear electron flow intact. This hypothesis is tested over two chapters: Chapter 4 covers the generation of mutant strains, testing of their growth phenotypes, and proteomic comparisons; Chapter 5 covers measurements of photosynthesis using chlorophyll fluorescence and absorption spectroscopy techniques. Though we present evidence that loss of native FNR leads to modified electron transport in the mutants, including evidence of a deficiency in photosynthetic control, we argue against the hypothesis that loss of native FNR disrupts CET. We present a new theory to rationalise the mutant phenotypes, in which tethered FNR outcompetes other electron sinks in the stroma, causing an unbalanced stromal NADP pool and reduced photosynthetic efficiency.

Acknowledgements

I would first like to thank my supervisor, Matt, for choosing me as a student and giving me the opportunity to work at Sheffield. I have greatly enjoyed the work I undertook here, which has included countless unfinished projects, unexplored avenues, and hair-brained experiments that I would not have been able to consider without the freedom he gave me in the lab. In his 2013 companion, *Letters to a Young Scientist*, the late E. O. Wilson stressed the importance of curiosity in research, and of continuously seeking to experiment and tinker with subjects that interest you. Being in the Johnson lab gave me that space to experiment, some of which led to the results presented in this thesis, all in an environment of collegiality and generous expertise. Such an experience, I expect, will turn out to have been a rare privilege. Also a privilege: thanks to Matt for taking me to Île des Embiez in August 2021 for a week of Chlamydomonas talks in the greatest conference location imaginable.

To Luke Mackinder, with whom I began my career in science and who provided me with a stable base to build the skills that got me this PhD, I would like to give my heartfelt thanks. Without my time in the Mackinder Lab this thesis would unquestionably have been worse. I am grateful to Dr Gary Yates, Dr Charlotte Walker and Justin Lau for their friendship that has endured well past the end of my time in York. I would have far more grey hairs had it not been for their valuable contribution to the results in Chapter 3.

It has been a pleasure to be part of the lab over the past four years and witness the transitioning of PhDs into postdocs and postdocs into lecturers during my time here. This thesis would not have been possible, let alone rather enjoyable, without the guidance and feedback from my friends and colleagues at the University of Sheffield. To Prof Neil Hunter for his guidance and for enabling the flow of science in this lab. To Stuart Casson for his clear-headed advice and for stepping in as my adviser mid-way through the PhD. Thanks is owed to Mo Brindley for her kindness and help in and out of the lab. To Liz, without whom I would have had nothing to work with, a huge thank you is owed for all her hard work keeping the lab in order, as well as outside the lab for keeping the midweek fun with her boardgame nights. A big thank you is due to the members of the E6 and E14 offices, I looked forward to coming to work each day because of them. A thanks is owed to every mentor I have learned from while in Sheffield, including Craig for AFM and EM, Matt and George for protein purification,

Will for microscopy, and Gustaf for PAM fluorometry and coding. A special thank you to Matt Proctor, Andy Hitchcock and George Sutherland; without them I really would have been screwed. Though not always smooth sailing being the only person working on Chlamydomonas in the lab, their advice and guidance made me feel part of the lab proper. Together they have given me an understanding of research far beyond my research topic, and if I have a future in science, it will probably involve heavy emulation of them.

Outside the lab I would like to thank Emily, Holly, Pete and Federica for organising and corralling me into an entrepreneurship competition in my third year; one of my only extracurricular activities during this PhD, and one that certainly helped make the COVID lockdowns more enjoyable. I would also like to thank my favourite collaborators, Dr Felix Buchert and colleagues in the Hippler lab at the University of Münster. Prof Hippler's guidance and enthusiasm and Felix's stoic pragmatism and generosity of time and effort have improved Chapter 5 immeasurably, and made the complex techniques involved a joy to learn.

For steadying me in those moments of uncertainty and fatigue over the past four years, including through the COVID-19 pandemic, a huge thank you is owed to my sister Lizzie, my brother and sister-in-law, Luke and Beth, and my mother and father for their belief and support. I hope I have done them proud with this work, and that they will find it at least legible if not always intelligible. Last and most, a huge thank you to Dr Margrete Langmyhr who has helped me more than I can describe here. To see her own inspiring transition to a credentialed Doctor of Philosophy in 2021 did wonders for the old self-belief. Long nights in the lab, writing, coding or worrying about deadlines; these would all hardly have seemed worthwhile without her understanding and support.

Declaration

I, the candidate, confirm that this thesis is my own work, except where work that has formed part of jointly authored publications has been included. My contribution and the contributions of my co-authors are explicitly indicated below. I confirm that appropriate credit has been given within the thesis where reference has been made to the work of others. I am aware of the University's Guidance on the Use of Unfair Means (www.sheffield.ac.uk/ssid/unfair-means). This work has not been previously presented for an award at this, or any other, university.

Much of the work described in Chapter 3 contributed to the following research publication: Emrich-Mills, T. Z., Yates, G., Barrett, J., Girr, P., Grouneva, I., Lau, C. S., Walker, C. E., Kwok, T. K., Davey, J. W., Johnson, M. P., and Mackinder, L. C. M. (2021). A recombineering pipeline to clone large and complex genes in Chlamydomonas. Plant Cell, 33(4), 1161–1181.

The publication is included as an appendix. The author contributions for the publication were as follows. I developed the initial recombineering pipeline, designed and assembled the original pLM099 recombineering plasmid and performed the genome wide analysis. G.Y., T.K.K. and I assembled additional recombineering plasmids. G.Y. and I optimized and performed the large-scale recombineering pipeline. G.Y. performed the microscopy and Venus quantification data. P.G. validated the pipeline using fosmids. J.B. performed the complementation experiments. J.B., I.G., C.S.L., C.E.W., and T.K.K. supported the development and implementation of the recombineering pipeline. J.W.D. wrote the BACSearcher code and provided bioinformatics support to me for the remaining code. L.C.M.M. conceived the idea and led the research. L.C.M.M. and M.P.J. received funding to support the work. L.C.M.M., G.Y. and I wrote the manuscript.

Contents

Acronyms and abbreviations

2-PG 2-phosphoglycolate.

3-PGA 3-phosphoglycerate.

ANOVA Analysis of variance.

ANR1 Anaerobic Response Protein 1.

APHVII Aminoglycoside phosphotransferase type VII gene, conferring resistance to hygromcyin.

APHVIII Aminoglycoside phosphotransferase type VIII gene, conferring resistance to paromomycin.

ATP Adenosine triphosphate.

BAC Bacterial artificial chromosome.

BLE Beta-lactamase gene, conferring ampicillin and zeocin resistance.

BPGA 1,3-bisphosphoglycerate.

BST Bestrophin-Like Protein.

CAH3 Carbonic Anhydrase 3.

CAS Calcium Sensor Protein.

Cas9 CRISPR-Assocation Nuclease 9.

CBB Calvin-Benson-Bassham.

CC-4533 "Chlamydomonas Centre" strain number 4533.

CCM CO_2-concentrating mechanism.

cDNA Complementary DNA; DNA generated from the reverse transcription of RNA.

CET Cyclic electron transfer (a.k.a., cyclic electron flow).

CHLG Chlorophyll Synthase.

Chls Chlorophylls *a* and *b*.

CMEF Chloroplast-to-mitochondria electron flow.

CRISPR Clustered regular interspersed short palindromic repeats.

Cytb_6f The cytochrome b_6f complex.

DCMU 3-(3,4-dichlorophenyl)-1,1-dimethylurea; a potent inhibitor of the Q_B site in photosystem II.

DNA Deoxyribonucleic acid.

DSB Double-stranded break (in DNA).

dsRNA Double-stranded DNA.

DTS Dynamic thylakoid stacking.

ECS Electrochromic shift.

EPYC1 Essential Pyrenoid Component 1.

ETC Electron transport chain.

ETR Electron transfer rate.

Fd (Fd^-/Fd^+) Ferredoxin (reduced and oxidised).

FDX2 Ferredoxin 2; the second most abundant isoform of ferredoxin in the Chlamydomonas stroma.

Flv Flavodiiron protein.

FNR Ferredoxin:$NADP^+$ Oxidoreductase, a.k.a., ferredoxin:$NADP^+$ reductase.

FQR Ferredoxin:plastoquinone reduction; the theoretical reduction of plastoquinone using electrons from stromal ferredoxin.

FRET Förster resonance energy transfer.

GAP Glyceraldehyde 3-phosphate; the primary carbohydrate product of oxygenic photosynthesis.

GAPDH Glyceraldehyde 3-phosphate Dehydrogenase.

GC Guanine-cytosine, used to refer to the percentage of a sequence of DNA composed of guanine and cytosine bases.

gDNA Genomic DNA.

HA Hydroxylamine.

HLA3 High Light Activated Protein 3.

HSD Honest significant difference.

HSP70A Heat shock Protein 70A gene.

HYDA Hydrogenase A.

ISA1 Isoamylase 1.

ISP Iron-sulphur protein (e.g., the Rieske Iron-Sulphur Protein, a subunit of cytb_6f).

KEA3 K^+ Efflux Antiporter 3.

LCI1 Low Carbon Inducible 1.

LCIA/B/C/9 Low Carbon Inducible A/B/C/9.

LET Linear electron transfer (a.k.a., linear electron flow).

LHCI Light Harvesting Complex I.

LHCII Light Harvesting Complex II.

mRNA Messenger RNA.

NADP Nicotinamide adenine dinucleotide phosphate; two-electron donor/acceptor in the chloroplast stroma.

NADPH Hydrogenated (reduced) form of NADP.

NDA2 Type II NADPH Dehydrogenase.

NDH-CET NDA2-dependent cyclic electron transfer in Chlamydomonas photosynthesis.

NHEJ Non-homologous end joining.

NPQ Non-photochemical quenching.

NT (NT4, NT7, NT12) "Native/tethered" strains derived from the transformation of UVM4 with a cassette expressing a PsaF-FNR-Strep fusion protein.

OEC Oxygen evolving complex.

ORF Open reading frame; a length of DNA defined by a start codon on the 5' end and a stop codon at the 3' end, e.g., encoding a gene.

P680 The special pair of chlorophylls in the reaction centre of photosystem II, with an absorption maximum at 680 nm.

P700 ($P700^+$) The special pair of chlorophylls in the reaction centre of photosystem I, with an absorption maximum of 700 nm. P700+ is the oxidised form.

PAM Protospacer adjacent motif (a short sequence of DNA flanking a CRISPR recognition sequence), or pulse amplitude modulation (a method of fluorescence detection in photosynthetic research).

PC Plastocyanin.

PCEF Pseudo-cyclic electron flow.

PCI Photosynthetic complex I.

PCR Polymerase chain reaction.

PETF Photosynthetic Electron Transfer F; the specific name for the dominant isoform of ferrodoxin in the Chlamydomonas stroma.

PGR-CET PGR5/PGRL1-dependent cyclic electron transfer in Chlamydomonas photosynthesis.

PGR5 Proton Gradient Regulation 5.

PGRL1 Proton Gradient Regulation-Like 1.

Pmf Proton motive force.

PQ Plastoquinone.

PSI Photosystem I.

PSII Photosystem II.

PTOX Plastid Terminal Oxidase.

Q_A/Q_B Quinone binding sites in the membrane intrinsic region of photosystem II, significant as the destination of electrons generated from the oxidation of water due to light-driven charge separation.

qE Energy-dependent quenching in fluorescence from photosynthetic organisms/thylakoid membranes.

RBCS2 Rubisco small subunit 2 gene.

RBMP1 Rubisco-Binding Membrane Protein 1.

RNA Ribonucleic acid.

RNAi RNA interference; a method for silencing translation of mRNA.

RNP Ribonucleoprotein; formed by the complexing of guide RNA with a CRISPR associated nuclease.

ROS Reactive oxygen species.

Rubisco Ribulose 1,5-bisphosphate Carboxylase Oxygenase.

RuBP Ribulose 1,5-bisphosphate.

SAGA1 Starch Granules Abnormal 1.

SDS-PAGE Sodium dodecylsulphate polyacrylamide gel electrophoresis.

sgRNA Single guide RNA; an RNA molecule complementary to a target region of DNA that can form a complex with a Cas9 enzyme.

SP Saturating pulse.

STT7 State Transition 7; the serine-threonine kinase implicated in LHCII phosphorylation in Chlamydomonas.

T (T7, T10, T34, T37) "Tethered-only" strains derived from the transformation of UVM4 with a cassette expressing a PsaF-FNR-Strep fusion protein.

TAB2 RNA binding protein involved in translation of PsaB.

TAP Tris-acetate-phosphate media, containing acetate as a carbon source for growing Chlamydomonas heterotrophically.

TE buffer Tris-EDTA buffer.

Tic62 Translocon at the Inner Envelope Membrane of Chloroplast 62.

TP Tris-phosphate media; used to grow Chlamydomonas phototrophically.

tRNA Transfer RNA.

TROL Thylakoid Rhodanase-Like Protein.

TRX Thioredoxin system.

uORF Upstream open reading frame.

UTR (5'UTR, 3'UTR) Untranslated region (often present at the 5' or 3' end of a gene, or both).

UVM4 Ultraviolet mutant 4.

VDE Violaxanthin De-Epoxidase.

β**DDM** *n*-dodecyl-β-D-maltoside; a commonly used detergent for solubilising thylakoid membranes.

Δ**pH** Delta pH; The difference in pH across the thylakoid membrane between the lumen and the chloroplast stroma.

$\Delta\psi$ Delta psi; The difference in charge across the thylakoid membrane between the lumen and the chloroplast stroma.

List of Figures

Chapter 1

Introduction

The global population is projected to pass eight billion in November this year and ten billion by the year 2050 (Gaigbe-Togbe et al. 2022). At the same time, land use is under intense pressure from the growing consumption of animal products, growing demand for biofuels (Ray et al. 2013), and loss of arable land by desertification (Cameron et al. 2015). To meet the demand, crop yields may need to double by 2050 (Ray et al. 2013). These factors are fuelling high interest and investment in genetic strategies to improve crop yields beyond traditional breeding methods. Transgenic herbicide-resistant crops first put into use in the USA thirty years ago have now become widespread, but initial improvements have been stymied by the evolution of parallel herbicide resistance in weeds (Duke 2015). More ambitious proposals to achieve higher and more consistent crop yields now aim for improved disease resistance, drought tolerance, salt tolerance, flood tolerance, and optimisation of photosynthesis (Bailey-Serres et al. 2019). Current genetic strategies range from tweaking existing traits to more drastic restructuring or *de novo* introduction of whole metabolic processes (Jackson et al. 2021). However, genetic engineering of crop plants is generally a slow process, taking many months to grow mutant plants. Of great benefit to crop research is the use of alternative model organisms that share key features to plants but are more amenable to genetic manipulation and testing. Microalgae play an important role in this effort, and one of the most widely used eukaryotic algal species in research is *Chlamydomonas reinhardtii* (Fig. 1.1A).

1.1 *Chlamydomonas reinhardtii*

Chlamydomona reinhardtii, hereafter referred to as Chlamydomonas, is a single-celled, bi-flagellar, soil-dwelling alga that can easily be cultured in the lab in liquid or as colonies on agar-supplemented media. The major lab strains in use today can all be traced back to the strain collection of Gilbert M. Smith in 1945, apparently descended from cells found in soil from a potato field in Massachusetts (Harris 2009, p11). Chlamydomonas is a facultative heterotroph, able to grow without relying on

photosynthesis if provided with an alternative carbon source. Cultures supplemented with acetate grow relatively quickly and can reach high cell densities. The three genomes of Chlamydomonas are fully sequenced and transformable, and extensive mutant libraries are available from several repositories. Cells can be maintained in a haploid state under normal growth, or diploidy can be induced to promote mating between the mating types, mt+ and mt-. Sequencing of the reference nuclear genome was completed by Merchant and colleagues in 2007 (Merchant et al. 2007), though subsequent sequencing showed the genomes of modern lab strains have diverged to a remarkable degree in the past eighty years (Gallaher et al. 2015). Key morphological and anatomical features are common to all strains though; cells are ~4-14 μm in length and contain a single chloroplast that takes up the anterior third of the cell volume and extends in a cup-shape around the cytosol (Fig. 1.1B). The Chlamydomonas chloroplast functions similarly to those in plant cells, finding them wide use as a model organism for studying photosynthesis and plant metabolism. Meanwhile their retention of many genes from the common ancestor of plants and animals makes them useful models for research on human ciliopathies and cell cycle control (Silflow and Lefebvre 2001; Breker et al. 2018). They are also widely used as a chassis for the production of high-value metabolites and therapeutic proteins (Dyo and Purton 2018), even showing potential as a future food source, including for extra-terrestrial agriculture (Mapstone et al. 2022).

1.1.1 The Chlamydomonas chloroplast

The chloroplast is the photosynthetic organelle in plants and algae. Despite having diverged from vascular plants almost a billion years ago (Kumar et al. 2017), Chlamydomonas and plant chloroplasts are anatomically similar and share many of the major photosynthetic components and mechanisms of regulation. Like vascular plants, the Chlamydomonas chloroplast is bound within a chloroplast envelope and contains an extensive system of thylakoid membranes. The medium surrounding the thylakoids is the chloroplast stroma, which contains copies of the chloroplast genome (plastome) and the machinery for transcribing and translating genes and assembling the protein machinery of photosynthesis. Each thylakoid membrane consists of a lipid bilayer enclosing a thin lumen in a flattened sack-shape. The thylakoids themselves are densely packed with membrane proteins – as much as 80% protein composition – and are arranged in a single, undulating, highly folded, continuous compartment (Fig. 1.1C). The degree of folding of thylakoids is heterogeneous throughout the chloroplast; parts are tightly appressed, containing anywhere between two and fifty stacks of membrane pairs, and are interspersed by regions of more loosely arranged, non-appressed membranes (demonstrated nicely in Wietrzynski et al. 2020) (Fig. 1.1C). These two types of membrane ultrastructure – appressed and non-appressed – are comparable to the more regularly structured

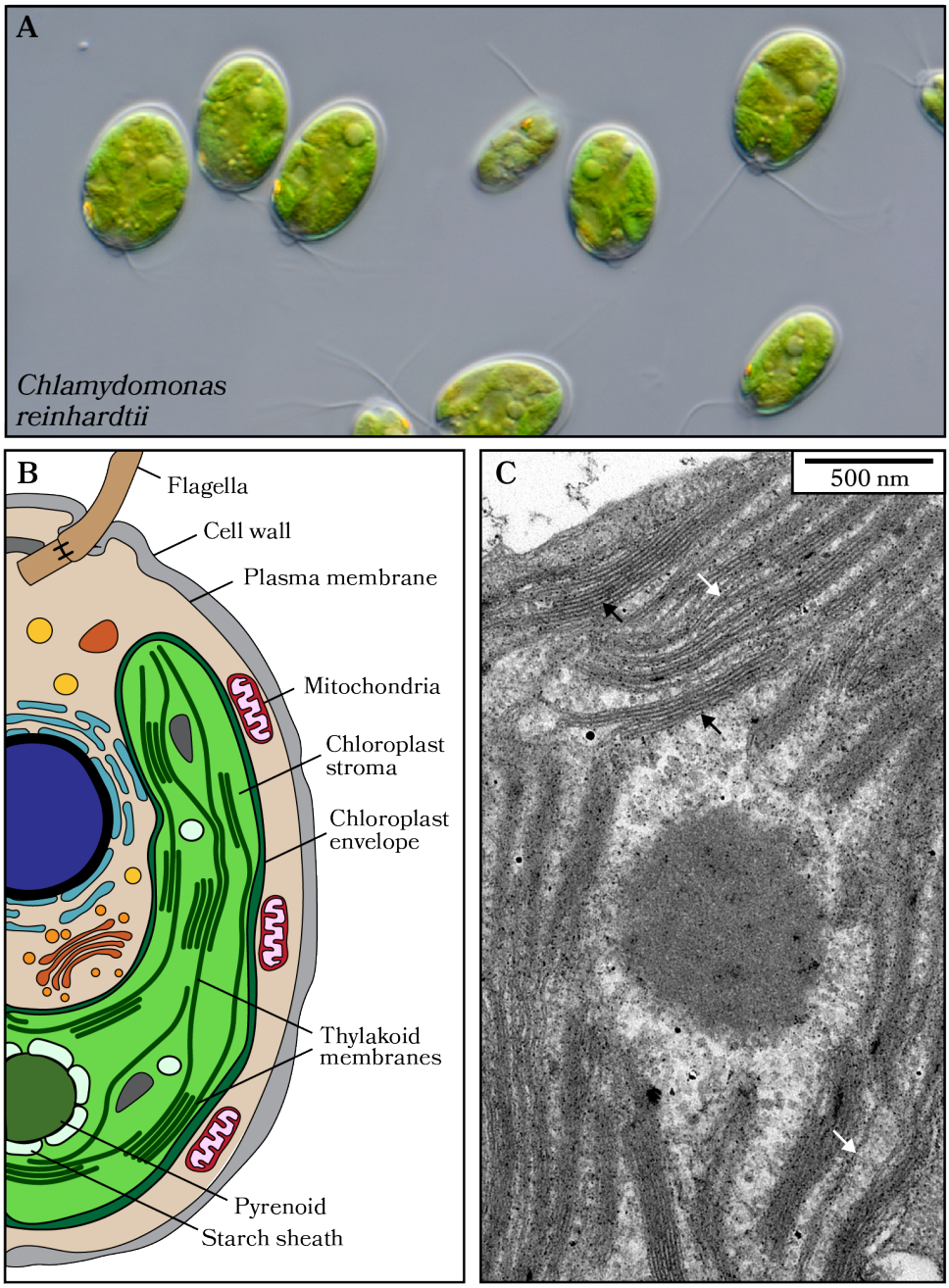


Figure 1.1: **The Chlamydomonas chloroplast. A.** Chlamydomonas cells under the microscope. Reproduced with permission from the photographer, Wim van Egmond. **B.** A cross section schematic of Chlamydomonas showing the lobed, cup-shaped chloroplast that extends from the base of the cell around the cystosol. The thylakoid membranes are arranged in the chloroplast into appressed and non-appressed regions. **C.** Transmission electron micrograph of the pyrenoid, a liquid-like microcompartment in the chloroplast, surrounded by thylakoid membranes. Examples of appressed thylakoid regions are indicated by black arrows and non-appressed regions by white arrows. Electron micrograph is author's own.

grana and stromal lamellae regions into which the thylakoids of vascular plants are arranged. The most noticeable difference to vascular plants is the presence of a sub-compartment of the Chlamydomonas chloroplast, the pyrenoid, a core component of the algal CO_2-concentrating mechanism that supercharges carbon fixation in photosynthesis.

1.2 Oxygenic photosynthesis

Photosynthesis is the process by which the radiant energy of light is stored by living organisms as chemical energy in the form of carbohydrates. The importance of this process for the biosphere cannot be overstated. Radiant energy is abundant (~240 W m^{-2} absorbed solar radiation at the earth's surface, Wild et al. 2013) but without photosynthetic organisms to perform its conversion to chemical bond energy, complex life could not have evolved and could not be sustained. Organisms that can use light to power their metabolism are called phototrophs, and the majority of these on earth produce oxygen as a byproduct of photosynthesis (> 80% of total biomass, Sánchez-Baracaldo and Cardona 2020). These are called oxygenic phototrophs and are subdivided into several taxa spread across the tree of life, including the prokaryotic cyanobacteria and the green lineage of plastid-containing, eukaryotic photosynthetic organisms, the archaeplastidae (Sánchez-Baracaldo and Cardona 2020). The entire green lineage likely stems from a common eukaryotic ancestor that contained a primitive form of a chloroplast, acquired by endosymbiosis from an ancient bacterium similar to the cyanobacteria that populate the oceans today (Douglas 1998). This common ancestor was likely a flagellated unicellular alga whose descendants diverged, among others, into the chlorophyta, encompassing most of the green algae including Chlamydomonas, and the streptophyta, which encompass the algal ancestors of the embryophytes, technically making all land plants a type of algae (Leliaert et al. 2012). There is enormous diversity across the oxygenic phototrophs in terms of body shape and size, life cycle, motility, degree of multicellularity and habitat. Their defining feature, though, relies on a core set of enzymes and reactions that are highly conserved across a staggering three billion years of evolution (Sánchez-Baracaldo and Cardona 2020). Oxygenic photosynthesis is itself an extensive and multifaceted process and so a broad field of study. Textbooks often divide photosynthesis into the reactions of light absorption and light-driven electron transfer, called the light reactions, and the down-stream reactions of carbon fixation, called the dark reactions. The light reactions split water into oxygen, protons and electrons, while the dark reactions use protons and electrons to reduce CO_2 to carbohydrates.

1.2.1 The light reactions of photosynthesis

The light reactions occur at the thylakoid membranes within the chloroplast and involve the absorption and conversion of light energy into electrical energy, culminating in an electron transfer chain (ETC) from H_2O to nicotinamide adenine dinucleotide phosphate ($NADP^+$) (Fig. 1.2A). Light is absorbed by chlorophyll and other pigment molecules arranged within the hydrophobic cores of light harvesting complexes (LHCs) embedded in the thylakoid membranes. The energy of an absorbed photon excites an electron on a pigment molecule in an LHC. This excitation energy is transferred between pigment molecules and between LHCs towards one of the two photosystems, PSI or PSII. PSII receives excitation energy from light harvesting complex II (LHCII). At the reaction centre of PSII, incoming excitation energy is used to extract electrons from water at the oxygen evolving complex (OEC), forming protons and molecular oxygen, which are released into the thylakoid lumen. The electrons are transferred across the thylakoid membrane and enter the electron transport chain (ETC). They are received by plastoquinone (PQ), which then diffuses through the membrane to the cytochrome b_6f complex (cytb_6f). Electrons are transferred from PQ to cytb_6f to plastocyanin (PC), a soluble protein bound at the lumenal side. Transfer of electrons from PQ to PC is coupled to the translocation of additional protons into the lumen. Reduced PC then diffuses through the lumen to photosystem I (PSI), which then transfers the electron across the membrane to the soluble stromal protein ferredoxin (Fd) using the energy of another absorbed photon, collected via the PSI-associated light harvesting complex I (LHCI). The ETC ends in the transfer of electrons from Fd to NADP+, reducing it to NADPH, one of the principal reactants of the downstream dark reactions (Fig. 1.2).

This sequence of electron transfers from PSII to $NADP^+$ is linear electron transfer (LET). The final Fd-to-$NADP^+$ reaction is catalysed by the key enzyme ferredoxin:$NADP^+$ oxidoreductase (FNR). FNR can also perform the reverse reaction, diverting electrons away from NADPH and making them available for further Fd-driven reactions like cyclic electron transfer (CET). In CET, electrons emerging from PSI are recycled from the stroma back into the PQ pool in the thylakoids instead of producing NADPH (Fig. 1.2A). Electrons cycled into the PQ pool can flow back through cytb_6f, PC and PSI, completing the circuit around PSI. Because CET also involves proton pumping at cytb_6f, CET can contribute to ATP production without NADPH production, whereas LET contributes to both.

The thylakoid lumen is more acidic than the chloroplast stroma and is kept compositionally distinct due to its small volume hindering the entry of bulky proteins. Under illumination the lumen becomes further acidified as protons are released from water at the OEC and translocated from the stroma by cytb_6f. This concentration of protons creates an osmotic potential across the membrane (ΔpH) and contributes to the electric potential ($\Delta\psi$). Together, ΔpH and $\Delta\psi$ comprise the proton motive force

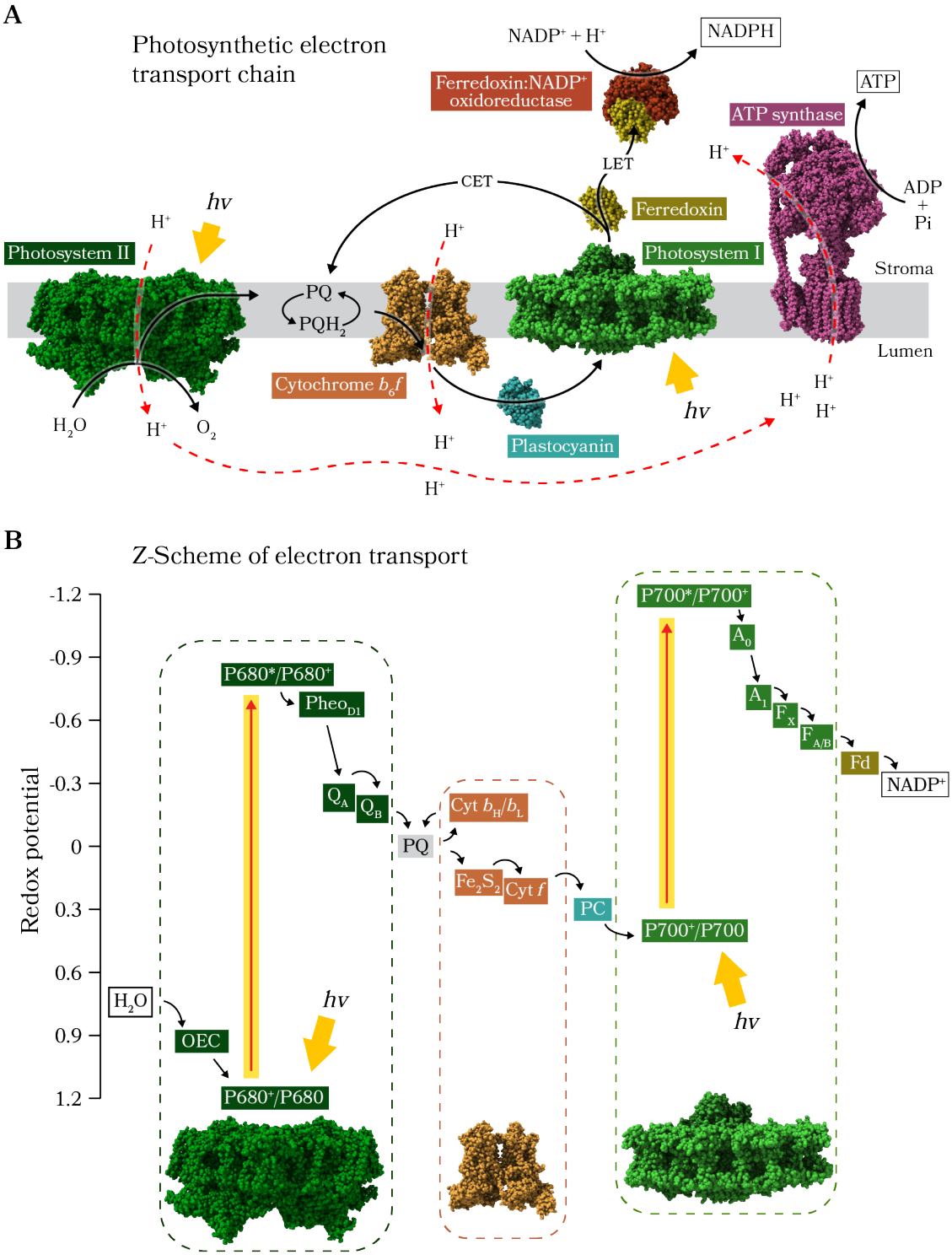

A
Photosynthetic electron transport chain
NADP⁺ + H⁺
NADPH
Ferredoxin:NADP⁺ oxidoreductase
ATP
ATP synthase
H⁺
LET
CET
ADP + Pi
hv
H⁺
Ferredoxin
Photosystem II
Photosystem I
Stroma
PQ
PQH2
Lumen
Cytochrome b6f
H2O
H⁺
O2
Plastocyanin
hv
H⁺
H⁺
H⁺
H⁺
B
Z-Scheme of electron transport
Redox potential
-1.2
-0.9
-0.6
-0.3
0
0.3
0.6
0.9
1.2
P680*/P680⁺
Pheo_D1
QA
QB
PQ
Cyt bH/bL
Fe2S2
Cyt f
PC
P700*/P700⁺
A0
A1
FX
FA/B
Fd
NADP⁺
H2O
OEC
hv
P680⁺/P680
P700⁺/P700
hv

Figure 1.2: **The light reactions of photosynthesis. A.** The canonical photosynthetic electron transport chain, culminating in the reduction of $NADP^+$ to NADPH (LET) or cycling of electrons back into the PQ pool through an unknown intermediate (CET). Proton translocation is coupled to electron transfer through PSII and cytb_6f (red dashed arrows). Protons in the lumen escape back through the ATP synthase, driving the phosphorylation of ADP to generate ATP. Protein models are represented by the following structures: Chlamydomonas PSII-LHCII (PDB 6KAC; Sheng et al. 2019), *Spinacia oleracea* cytb_6f (PDB 6RQF; Malone et al. 2019), *S. oleracea* plastocyanin (PDB 1YLB; Musiani et al. 2005), Chlamydomonas PSI-LHCI (PDB 6JO6; Suga et al. 2019), Chlamydomonas ferredoxin (PDB 6KV0; Ohnishi et al. 2020), *S. oleracea* FNR (PDB 1FNC; Bruns and Karplus 1995) and *S. oleracea* ATP synthase (PDB 6FKI; Hahn et al. 2018). **B.** The electron transfer intermediates of the ETC arranged according to redox potential, demonstrating the crucial input of energy required from the absorption of two photons to generate the highly reducing Fd and NADPH, which are then used to provide energy for metabolism.

(*pmf*), which can be harnessed to perform mechanical work by releasing protons back into the stroma down their electrochemical gradient. Protons can leak back through the membrane into the stroma if the *pmf* is high enough. Otherwise, protons are released through the ATP synthase, a valve that couples proton release to ATP production. The integrity of the thylakoid membranes and their ability to keep the stroma and lumen distinct is critical in maintaining the *pmf* and efficient ATP production.

1.2.2 The key photosynthetic components of the thylakoids

The key components of the light reactions are all bound to or associated with the thylakoid membranes. PSI, PSII and cytb_6f are all multi-subunit pigment-protein complexes with multiple transmembrane helices and hydrophobic cores that bind pigments, quinones and electron transfer intermediates. Most of the core subunits of these complexes are encoded in the chloroplast genome, but several are encoded in the nucleus and require post-translational processing and dedicated chaperones to enter the chloroplast. The structure and function of homologous subunits between plants and algae are highly similar even when encoded by different genomes, implying a remarkable conservation of the light harvesting and energy transfer mechanisms across billions of years. The following sections describe the major photosynthetic components of the Chlamydomonas thylakoids in more detail.

1.2.2.1 Pigments

Pigments are the light-active molecules that enable light-harvesting in photosynthesis. By far the majority of photosynthetically active pigments in the thylakoids are chlorophyll. Plants and algae possess two types of chlorophyll, *a* and *b*, that have

slightly different absorption spectra, allowing them to cover more of the electromagnetic spectrum together (Kan and Thornber 1976). Chlamydomonas also synthesises a range of carotenoids and xanthophylls that are important for extending the spectrum of light harvesting or are key components of electron transfer in the thylakoids. Chlorophyll, the primary light absorbing molecule for photosynthesis, consists of a hydrophobic phytol tail to enable protein binding and a chlorin macrocycle that provides a large surface area for light absorption. Electrons in each chlorin macrocycle are maintained in a specific conjugated π-system containing 24 delocalised electrons (Johnson 2016). All electrons orbiting any atomic nuclei can be excited from a ground state into a higher molecular orbital by the absorption of a photon whose energy matches the gap between the ground state, S_0, and an excited state, S_1 or S_2. However, electrons in conjugated systems have the property of their ground states and excited states being closer in energy than for non-conjugated systems. This allows the highly conjugated chlorophyll π-system to absorb low energy, red photons. Chlorophyll therefore shows a broad absorption peak in the red region of the electromagnetic spectrum, termed Q_Y. Chlorophyll also has an absorption peak in the UV-blue region called the Soret peak, resulting from the absorption of blue (higher energy) photons whose energy matches the gap between S_0 and a higher excited state, S_2. Upon absorption of a suitable photon by a chlorophyll molecule, an electron is briefly promoted to S_1 or S_2. Electrons in S_2 decay to S_1 extremely quickly (on a picosecond timescale), but the S_1 excited state survives for longer (~2 ns) before decaying back to S_0 (Johnson 2016). This longer lifetime makes possible the transfer of excitation energy to neighbouring chlorophylls. Energy transfer proceeds by two main methods in photosynthesis; exciton coupling between chlorophylls allows extremely efficient transfer if molecules are within ~1 nm, whereas longer range transfer is also possible by Förster resonance energy transfer (FRET) if molecules are within ~7 nm (reviewed in Jones and Bradshaw 2019).

1.2.2.2 Light harvesting complexes

Photosynthetically active chlorophyll *a* and *b* molecules are carefully arranged in the hydrophobic interiors of the light harvesting complexes (LHCs), their precise configurations being necessary to ensure efficient transfer distances for exciton coupling or FRET. The most abundant LHCs are LHCI and LHCII (Johnson 2016). The LHCs collect light for photosynthesis by forming antenna-like arrays around a photosystem in precise coordination to enable energy transfer. Energy occurs in series between coordinated chlorophylls within an LHC and between adjacent LHCs, enabling the transfer of energy from the site of absorption along a chain of pigment-protein complexes over several nanometres towards the reaction centre of either photosystem I (PSI) or photosystem II (PSII).

PSII receives energy from LHCII, probably the most abundant membrane protein

on earth. LHCII consists of a single polypeptide arranged as three transmembrane helices. Pigments and carotenoids are bound into the hydrophobic core of each monomer: eight chlorophyll *a*, six chlorophyll *b*, two carotenoids and one xanthophyll. LHCII is arranged into trimers in the membrane, wherein the monomer-monomer interfaces stabilise some of the chlorophylls and accommodate additional pigments. LHCI collects light exclusively for PSI and contains up to 20% of the chlorophyll in the thylakoid membrane (Harris 2009, p509). The LHCI monomer is structurally similar to LHCII and contains the same number of chlorophylls. LHCI is closely associated with PSI, and the LHCI-PSI interfaces help to accommodate additional chlorophylls.

1.2.2.3 Photosystem II

Photosystem II (PSII) is a large, multi-subunit, membrane-bound pigment-protein complex that is unique in nature for its ability to oxidise water and reduce bound plastoquinone (PQ) using the energy of absorbed light (Fig. 1.3A and B). Water oxidation takes place in the lumen at the oxygen evolving complex (OEC) and donates electrons to the donor side, whereas PQ reduction takes place on the stromal edge of the complex at the acceptor side. The structure of PSII is highly conserved between plants, algae, and cyanobacteria, and exists in the membrane as a heterodimeric core of proteins D1 and D2 that form the reaction centre, surrounded by chlorophyll binding proteins CP43 and CP47, peripheral antenna proteins CP26 and CP29 and trimeric LHCII antennae (Fig. 1.3A). For photochemistry to occur, excitation energy is transferred from CP43/47 to a special pair of chlorophylls in the reaction centre with an absorption maximum at 680 nm, known as P680. Upon excitation, a transient intermediate P680* is formed and a nearby molecule of pheophytin, $Pheo_{D1}$, is rapidly reduced on the timescale of a few picoseconds. The charge transfer can recombine back to P680 or can be quickly stabilized by electron transfer to a bound PQ (Q_A) to form a radical transmembrane pair, $P680^+Q_A^-$. $P680^+$ is reduced from the lumenal side in a four-step reaction mediated by a cluster of Mn atoms within the D1-bound oxygen evolving complex (OEC) via a nearby tyrosine (Y_Z) (Fig. 1.3C). Four sequential oxidations of the OEC by $P680^+$ builds an oxidation potential that is sufficient to oxidise water, releasing O_2 and protons into the lumen. Q_A^- is oxidised by a slower (microseconds) two-step reduction of a PQ at the Q_B site near the stromal side of the complex requiring the input of protons from the stroma. Once reduced, the association between PQH_2 and the Q_B site is weakened and PQH_2 can diffuse out of PSII and join the membrane-bound PQ pool (Fig. 1.2B). Overall, PSII-mediated H_2O:PQ oxidoreductase activity converts H_2O to reduced plastoquinone (PQH_2) as follows:

$$2H_2O + 2PQ + 4H^+_{(stroma)} \rightarrow O_2 + 4H^+_{(lumen)} + 2PQH_2 \tag{1.1}$$

1.2.2.4 The plastoquinone pool

Plastoquionone (PQ) is a lipid-soluble electron carrier that can diffuse within the thylakoid membranes, receiving electrons between PSII during LET and receiving electrons from the stroma during CET. The pool of available PQ needs to be maintained at a ratio near 50:50 oxidised/reduced so as to continue accepting and donating electrons efficiently. The redox state of the PQ pool is thus a crucial factor affecting photosynthetic efficiency. Though the name implies a single pool of molecules in flux, the PQ pool may be subdivided into distinct regions in the thylakoids, one that is mainly reduced photochemically (from PSII) and one that is mainly reduced non-photochemically (from the stroma) (Alric 2015). This is not to mention the smaller divisions caused by obstructions from the high concentration of protein complexes and supercomplex assemblies in the thylakoids.

1.2.2.5 The cytochrome b_6f complex

The cytochrome b_6f complex (cytb_6f) is a homodimeric membrane-bound PQ:PC oxidoreductase that transfers electrons from the plastoquinone pool to the lumenal one-electron carrier PC (Fig. 1.2). Cytb_6f sits in the centre of photosynthetic electrons transfer, connecting PSII and PSI and fed by electrons from both LET and CET. Each cytb_6f monomer consists of four major proteins, cytochrome f, cytochrome b_6, the Rieske iron sulphur protein (ISP) and subunit IV (Fig, 1.3D). These are stabilized by a 'picket fence' of transmembrane helices consisting of subunits PetG, L, M and N. In Chlamydomonas, cytb_6f also contains a loosely bound fifth transmembrane subunit, PetO, while cytb_6f from *Synechocystis spp.* contains a loosely bound stromal subunit, PetP (Proctor et al. 2022). The complex coordinates an iron-sulphur cluster (Fe_2S_2), two b-hemes (b_L and b_H) and two c-hemes (cyt f and c_i) per monomer, as well as one chlorophyll and one carotenoid with unknown functions (Stroebel et al. 2003).

Cytb_6f bears structural similarities to both the bacterial bc_1 complex and the mitochondrial cytochrome bc_1 (respiratory complex III), all of which perform quinol oxidation by an electron bifurcative process known as the Q-cycle, comprising two half-cycles (Crofts et al. 1983, Mitchell 1975). In the first half-cycle, PQH_2 from the PQ pool diffuses into the complex and is oxidised twice at the lumenal binding site Q_o, requiring the uptake of two protons from the stroma (Fig. 1.3E). One electron enters the high-potential chain of transfer reactions, which consists of an Fe_2S_2 in the ISP, then on to cyt f followed by PC. The other electron enters the low-potential chain, consisting of haems b_L and b_H, and on to haem c_i. Oxidation of another PQH_2 at the Q_o site is required for the second half-cycle. Here, again, one electron from PQH_2 enters the high-potential chain, leading to the reduction of another PC in the

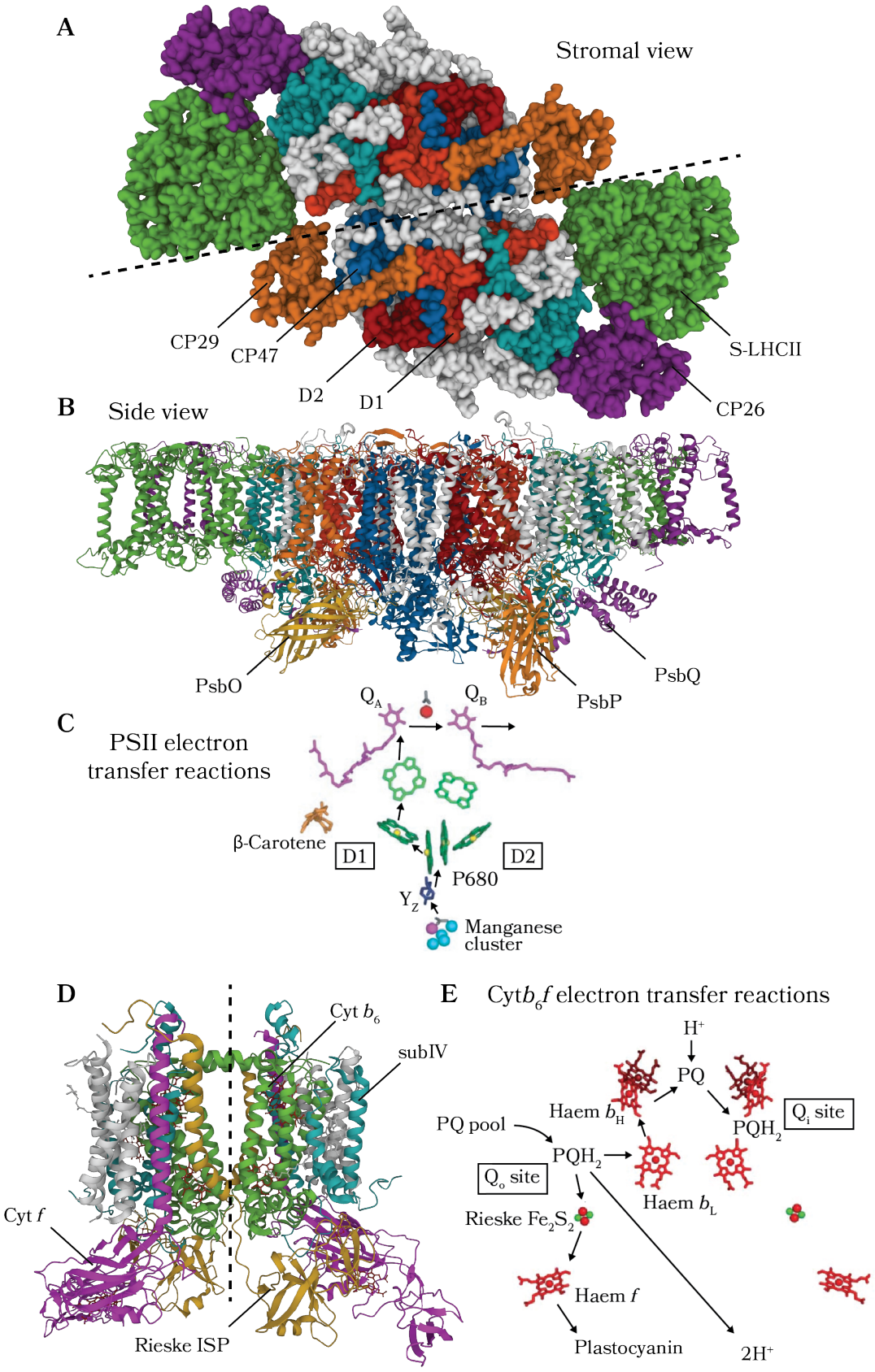

A
Stromal view
CP29
CP47
D2
D1
S-LHCII
CP26
B
Side view
PsbO
PsbP
PsbQ
C
PSII electron transfer reactions
Q_A
Q_B
β-Carotene
D1
D2
P680
Y_Z
Manganese cluster
D
Cyt b_6
subIV
Cyt f
Rieske ISP
E
Cytb_6f electron transfer reactions
H^+
PQ
Haem b_H
Q_i site
PQH_2
PQ pool
PQH_2
Q_o site
Haem b_L
Rieske Fe_2S_2
Haem f
Plastocyanin
$2H^+$

Figure 1.3: **Photosystem II and the cytochrome b_6f complex A.** Stromal view of Chlamydomonas PSII-LHCII (PDB 6KAC; Sheng et al. 2019), showing the PSII core subunits D1 and D2, chlorophyll binding protein CP47, peripheral binding proteins CP26/29 and a strongly bound LHCII trimer (S-LHCII). PSII and coupled antenna form a homodimeric supercomplex (symmetry indicated by dashed line). **B.** Side view of PSII-LHCII showing the oxygen evolving complex subunits PsbO/P/Q projecting into the lumen. Adapted with permission from Nelson and Ben-Shem (2004). **C.** Schematic of the electron transfer reactions within the PSII reaction centre, coordinated by the core subunits D1 and D2. The special pair of chlorophylls, P680, and two weakly coupled chlorophylls are shown as porphyrin rings. P680 forms a radical pair with the bound quinone Q_A upon absorption of light, triggering the reduction of Q_B and the almost simultaneous oxidation of the manganese cluster in the OEC (via tyrosine Y_Z). Fast transfer between Q_A and Q_B is facilitated by nearby HCO_3^- (grey) and Fe^{2+} (red sphere) ions. **D.** The cytochrome b_6f complex (cytb_6f) from *S. oleracea* (PDB 6RQf; Malone et al. 2019) – also a homodimeric complex – showing the four key subunits that coordinate the electron transfer intermediates, cyt f, cyt b_6, subunit IV and the Rieske iron sulphur protein. Additional subunits, shown in white, likely act to stabilise the complex. **E.** The electron transfer intermediates coordinated by cytb_6f, showing the relative positions of each haem and the electron transfer reactions of the Q-cycle. Haem c_i is shown in dark red. Adapted with permission from Nelson and Ben-Shem (2004).

lumen. The second electron enters the low-potential chain, joining the electron from the previous half-cycle by forming a highly reducing species made up of haems c_i and b_H, which reduces bound PQ at the stromal Q_i binding site, completing the Q-cycle. The sequential oxidation of two molecules of PQH_2 at the Q_o site liberates a net four protons into the lumen and regenerates one PQH_2, with the overall reaction as follows:

$$2PQH_{2(Qo)} + PQ_{(Qi)} + 2H^+_{(stroma)} \rightarrow 2PQ_{(Qo)} + PQH_{2(Qi)} + 4H^+_{(lumen)} + 2e^-_{(cytf \rightarrow PC)} \quad (1.2)$$

The oxidation of PQH_2 at the Q_o site is the rate limiting step for the whole photosynthetic ETC, occurring on a timescale of several milliseconds (Saif Hasan and Cramer 2012). However, the translocation of two protons into the lumen per electron transferred makes cytb_6f the major site of proton translocation in the ETC and the major contributor to ΔpH. In addition, transfer along the low-potential chain from PQH_2 to haems b_L and b_H shifts negative charge from the lumenal to the stromal side of the thylakoid membrane, contributing to the $\Delta\psi$ component of pmf while the Q-cycle is running.

Mechanistically, cytb_6f is still somewhat mysterious. According to the available high-resolution structures of cytb_6f (Malone et al. 2019, Proctor et al. 2022, Stroebel et al. 2003), the high-potential chain appears to require transfer of charge from

the Fe_2S_2 to cyt *f* across too large a distance to be physically possible, implying a significant conformational change is induced and reversed with each half-cycle to facilitate the transfer (Malone et al. 2021). In addition, evidence from *in silico* modelling of a potential Fd binding site on the stromal side suggests that cytb_6f may receive electrons directly from stromal Fd^- during CET, possibly via the second *c*-type heme, c_i (Nawrocki et al. 2019a). This theory of CET fits with the original Q-cycle mechanism proposed by Mitchell (1975), in which PQ reduction at the Q_i site was achieved via one electron from haem b_H and one from the stroma. This and competing theories regarding the site of electron entry to the PQ pool during CET are explored in more detail in Section 1.4 on page 27.

1.2.2.6 Plastocyanin

The small, soluble, copper-containing protein plastocyanin (PC) transfers electrons from cytb_6f to PSI (Fig. 1.2A). Oxidised PC (PC^+) receives an electron from cyt *f* of cytb_6f and diffuses freely through the lumen to enter the PC pool, while reduced PC (PC^-) associates with the donor side of PSI at the lumenal face of the PsaF subunit and functions as the primary electron donor to the PSI reaction centre's special pair of chlorophylls, P700. Though vascular plants use PC exclusively, in Chlamydomonas, the heme-containing protein cytochrome c_6 can substitute for PC under copper-deficient conditions (Merchant and Bogorad 1987). Constant turnover of the pool of available PC or cytochrome c_6 is required to maintain high rates of PSI reduction, with efficient net PC-to-PSI electron transfer taking place in less than a millisecond (Farah et al. 1995). This implies that individual PC molecules must quickly diffuse several tens of nanometres through the lumen between cytb_6f and PSI to transfer electrons.

1.2.2.7 Photosystem I

Photosystem I (PSI) is a large, multi-subunit protein complex that can oxidise PC or cytochrome c_6 and reduce ferredoxin (Fd) using the energy of absorbed light. PC oxidation takes place in the stroma at the PSI donor side, while Fd reduction takes place in the stroma at the PSI acceptor side. Like PSII, PSI consists of chlorophyll-binding subunits surrounding a heterodimeric reaction centre. The reaction centre comprises proteins PsaA and PsaB and contains a special pair of chlorophylls, P700, with an absorption maximum at 700 nm. Apart from the reaction centre, ten other transmembrane helix-containing subunits, PsaC-L, form the core complex, and are ringed on one side by eight bound LHCI that channel excitation energy to P700 in the reaction centre (Fig. 1.4A and B). Upon excitation, the transient intermediate P700* is produced, possessing one of the most negative redox potentials in nature (Baker et al. 2007) (Fig. 1.2B). P700* swiftly undergoes charge separation to a nearby chlorophyll, A_0, producing the radical pair $P700^+A_0^-$. This can either reduce a bound

quinone, A_1, or collapse back to P700 (charge recombination). The charge separation sequence from P700 to A_1 can occur along either of two identical routes coordinated by each monomer of the dimeric reaction centre (Fig. 1.4C). The routes converge at a cluster of Fe_4S_4 centres known as F_X, F_A and F_B, coordinated by PsaA/B and PsaC, respectively (Harris 2009, p559-560). Electron transfer from A_1 to F_X stabilizes the charge separation, followed by transfer to F_A and F_B, before reduction of a molecule of Fd at the stromal docking site, coordinated by PsaC, D and E (Sétif et al. 2002) (Fig. 1.4B). The $P700^+$ ion is simultaneously reduced at the PSI donor side by bound PC or cytochrome c_6 at a docking site coordinated by PsaF and PsaJ . In Chlamydomonas, loss of the PsaF subunit does not neutralize PSI, but results in ~80-fold drop in the rate of electron transfer from PC to $P700^+$ (Hippler et al. 1997) and consequently a much slower oxidation of the PC pool and cyt *f*.

1.2.2.8 Ferredoxin

Ferredoxin (Fd) is a small, soluble, stromal protein containing a Fe_2S_2 cluster that is present in several different isoforms in Chlamydomonas, of which the major isoform, FDX1, is thought to be the primary electron acceptor of PSI (Johnson 2016). Ferredoxin has a redox potential of approximately ~430 mV, making it the most electronegative cofactor in photosynthesis (Tagawa et al. 1963). Oxidised Fd (Fd^+) is reduced by the acceptor side of PSI due to the production of the highly reducing species P700* upon the absorption of excitation energy at the PSI reaction centre. Reduce Fd (Fd^-) can then diffuse freely within the stroma, where it is involved in electron transfer to several pathways in Chlamydomonas including LET and CET.

1.2.2.9 Ferredoxin:$NADP^+$ oxidoreductase

The final step in LET is the transfer of two electrons from two molecules of Fd^- to $NADP^+$ to produce NADPH, an important reductant for downstream carbon fixation. This step is catalysed by a 35 kDa enzyme, ferredoxin:$NADP^+$ oxidoreductase (FNR). FNR is a flavoprotein with its active site positioned between a NADP-binding domain and a bound molecule of flavin adenine dinucleotide (FAD). FNR is a member of a large, extended family of similar enzymes present across the tree of life, many of which catalyse the reverse reaction, producing reduced Fd from NAD(P)H (Aliverti et al. 2008). In the dark, FNR in photosynthetic organisms also primarily catalyses the reverse reaction, producing two Fd^- per NADPH consumed (Fulgosi and Vojta 2020). Although most NADPH is consumed in the chloroplast during carbon fixation, both NADPH and Fd^- are consumed in a wide variety of additional metabolic processes, from nitrogen fixation to hydrogen production (Aliverti et al. 2008, Ho et al. 2022). Thus, FNR has a key role in regulating the balance between NADPH-consuming pathways and Fd-consuming pathways.

FNR is a hydrophilic protein and is localized to the stroma. However, the degree of

association with the thylakoid membrane is still controversial. In cyanobacteria FNR is present in two isoforms, one soluble and one associated with the light-harvesting phycobilisomes bound to the thylakoids (Alcántara-Sánchez et al. 2017). In the model plant *Arabidopsis thaliana* (hereafter Arabidopsis), FNR isoforms FNR1 and FNR2 may both be bound to the membrane by tether proteins TROL and Tic62 (Kramer et al. 2021), though mutation of these tether proteins may not have a severe effect on LET (Lintala et al. 2014). Chlamydomonas contains a single isoform of FNR thought to reversibly associate with the thylakoid membrane at different locations, including as part of a supercomplex containing PSI and cytb_6f proposed to stimulate the rate of CET (Iwai et al. 2010). The putative role of FNR in CET is expanded in Section 1.5 on page 33 and is a key focus of Chapters 4 and 5.

1.2.2.10 ATP synthase

The ATP synthase is a multi-subunit, transmembrane proton valve that couples the movement of protons across the membrane to either phosphorylation of ADP plus inorganic phosphate (P_i) to ATP (in the forward reaction) or to the hydrolysis of ATP (in the reverse reaction). The forward reaction is powered by the *pmf* and so depends on proton translocation from stroma to lumen at PSII and cytb_6f. The chloroplastic ATP synthase is known as CF_0CF_1 due to its two structural distinct regions (Fig. 1.4D). The CF_0 region is embedded in the membrane and comprises subunits I-IV, while the CF_1 region protrudes into the stroma and comprises subunits α, β, δ, γ and ε in Chlamydomonas (Harris 2009, p640). Subunits I and II are anchored in the membrane and stabilise the interaction between CF_0 and CF_1. Subunit III has a high stoichiometry compared with I, II and IV, and is organized as a ring of transmembrane units called the c-ring, which is stabilized by IV (Fig. 1.4D). The c-ring binds lumenal protons, resulting in a slight conformational change to the structure of CF_0 with each binding. This movement is transmitted to CF_1 via an axel-like central stalk consisting of subunits δ, γ and ε. The stalk rotates a head group consisting of three pairs of α/β subunits in a hexagonal arrangement, with a catalytic site formed at the interface of each α/β pair. At each catalytic site, ATP synthesis occurs by a rotational catalytic mechanism in three stages, one with each 120° rotation of the head group: with the first, a molecule of ADP is bound; with the second, the cleft between the α and β subunits closes, fixing ADP to P_i; with the third, the synthesised ATP is released (Senior et al. 2002). Three catalytic sites per ATP synthase means that for each 360° turn of the head group, three ATP are synthesised.

1.2.2.11 A note on photosynthetic inhibitors

The details of the photosynthetic ETC were worked out over many decades aided by the use of inhibitors of different reactions in the chain. Several of these inhibitors are used in the work presented in this thesis to block the ETC at specific loci.

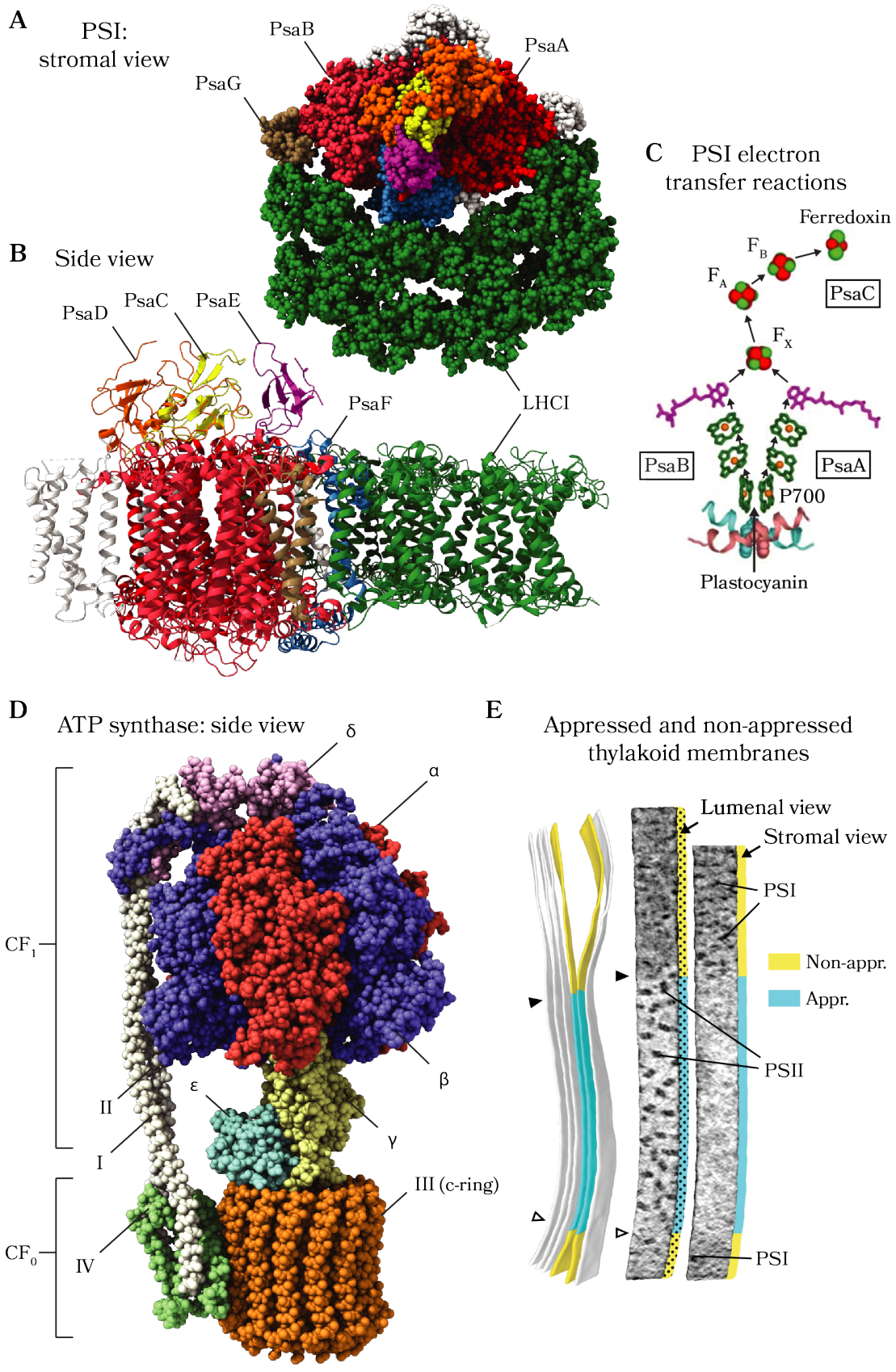

A
PSI:
stromal view
PsaB
PsaA
PsaG
B
Side view
PsaD
PsaC
PsaE
PsaF
LHCI
C
PSI electron
transfer reactions
Ferredoxin
F_B
F_A
PsaC
F_X
PsaB
PsaA
P700
Plastocyanin
D
ATP synthase: side view
δ
α
CF_1
β
II
ε
γ
I
III (c-ring)
CF_0
IV
E
Appressed and non-appressed
thylakoid membranes
Lumenal view
Stromal view
PSI
Non-appr.
Appr.
PSII
PSI

Figure 1.4: **Photosystem I, ATP synthase and their arrangement in the thylakoids. A.** Stromal view of Chamydomonas PSI-LHCI (PDB 6JO6; Suga et al. 2019). **B.** Side view of PSI, showing the three stromal subunits, PsaC/D/E, that coordinate the acceptor side electron transfer intermediates $F_A/F_B/F_X$. PsaF, a minor core subunit involved in docking plastocyanin on the lumenal side is shown in blue. Other minor core subunits PsaH-L are shown in white. **C.** They symmetrical pathways of electron transfer in the PSI reaction centre, illustrating transfer from PC on the lumenal donor side to Fd on the stromal acceptor side. Adapted with permission from Nelson and Ben-Shem 2004. **D.** The chloroplastic ATP synthase from *S. oleracea* (PDB 6FKI; Hahn et al. 2018). Three copies of the α and β subunits are present in each CF_1 headgroup. Subunit III forms a 14-unit c-ring in the membrane intrinsic CF_0 region, though the precise number in Chlamydomonas is yet to be ascertained. **E.** Cryo-electron tomograms from Chlamydomonas thylakoids illustrating the compositional differences between appressed and non-appressed membrane regions. Tightly appressed membranes (blue) are devoid of PSI and rich in PSII, as shown by the abundant protrusions into the thylakoid lumen, associated with the PSII OEC. Non-appressed membranes (yellow) are rich in PSI and ATP synthase, shown by protrusions of PsaC/D/E and ATP synthase headgroups into the stroma. Adapted with permission from Wietrzynski et al. 2020.

Hydroxylamine (HA) is used as an inhibitor of O_2 evolution at PSII and is thought to block electron transfer from water to P680* at the PSII donor side (Guiles et al. 1990). DCMU, a herbicide with the systematic name 3-(3,4-dicholorophenyl)-1,1-dimethylurea, is a specific inhibitor of PQ reduction by PSII. DCMU binds competitively and irreversibly to the PSII Q_B site after a single turnover of the reaction centre (Metz et al. 1986). Methyl viologen inhibits LET between PSI and NADPH, functioning as an electron acceptor at the PSI acceptor side, though may be primarily reduced by Fd^- rather than directly by F_A/F_B (Sétif 2015). Methyl viologen is therefore also used to inhibit CET as it functions to oxidise the stromal pool of Fd and, via the reverse reaction of FNR, NADPH (Sétif 2015). Lastly, antimycin A, a well-known inhibitor of respiration, is also a specific inhibitor for Fd-dependent reduction of PQ, blocking the major route of CET. The site of antimycin A inhibition in CET is still unclear since it is rather counterintuitively thought not to bind to cytb_6f in the same manner as to cytochrome bc_1 during its well-known inhibition of respiration. Instead, antimycin A may bind conditionally to cytb_6f under anoxic conditions (Buchert et al. 2022) or to a putative cytb_6f-interacting complex composed of PGR5 and PGRL1, two regulator proteins affecting the major route of CET.

1.2.3 Thylakoid heterogeneity and dynamics

PSII, PSI and ATP synthase are arranged heterogeneously in the Chlamydomonas thylakoid membranes. The tight gap between the stromal face of one membrane and the corresponding stromal face of an adjacent appressed membrane blocks entry of

ATP synthase and PSI into the appressed regions almost completely. Consequently, appressed regions are enriched in PSII while non-appressed regions are enriched in ATP synthase and PSI (and LHCI by association) (Fig. 1.4E). The evolutionary reason for the presence of highly appressed regions is unclear, but has been argued as necessary for the physical separation of the photosystems in vascular plants in order to prevent spillover of excitation energy from PSII to PSI, since PSI shows faster kinetics of excitation energy trapping (Trissl and Wilhelm 1993). Cytb_6f is present at roughly equal quantities in both regions but is thought to fulfil slightly different purposes in each; cytb_6f in the appressed regions receives electrons almost exclusively from PSII, whereas cytb_6f in the non-appressed regions may be sequestered mainly for CET, receiving the majority of its electrons from PSI via stromal carriers. This theory also extends to the PQ pool, which may be divided into appressed and non-appressed sub-domains (Joliot et al. 1992), one receiving the majority of its electrons from PSII while the other receives electrons from the stroma. A net partitioning of LET and CET into different spatial zones allows cells to regulate the ratio of LET:CET by modulating the net area of appressed and non-appressed thylakoids. These changes take place over several minutes in a process termed dynamic thylakoid stacking (DTS) (Hepworth et al. 2021). This is one of several important adaptational responses that enable cells to maintain homeostasis by upregulating CET when there is an increased demand for ATP relative to NADPH. In parallel and under similar conditions to DTS, cells can modulate the proportion of absorbed energy reaching the photosystems in the appressed and non-appressed regions. By far the majority of proteins in the thylakoid membranes are LHCII, and, like cytb_6f, LHCII is present in both appressed and non-appressed regions, funnelling absorbed radiant energy to the nearby reaction centres. LHCII in the non-appressed regions can reversibly associate with PSI and reversibly dissociate from PSII in the appressed regions. In this way, cells can balance the proportion of LHCII associated with each photosystem, modulating the amount of energy funnelled to PSI relative to PSII. Shifting the balance of excitation between PSI and PSII in this way is known as a state transition, and cells perform state transitions continuously to help keep the net activity of PSII and PSI balanced in the cell, which in turn helps maintain efficient production of ATP and NADPH in a fluctuating light-environment.

1.2.4 The dark reactions of photosynthesis

The dark reactions occur in the chloroplast stroma and are fuelled by atmospheric CO_2 and the major products of the light reactions, NADPH and ATP. These reactions involve the fixing of atmospheric carbon into organic molecules, followed by the reactions that regenerate the precursors for carbon fixation again. This cycle of reactions is the Calvin-Benson-Bassham (CBB) cycle (Fig. 1.5), the first step of which is catalysed by ribulose bisphosphate carboxylase oxygenase (Rubisco), estimated to

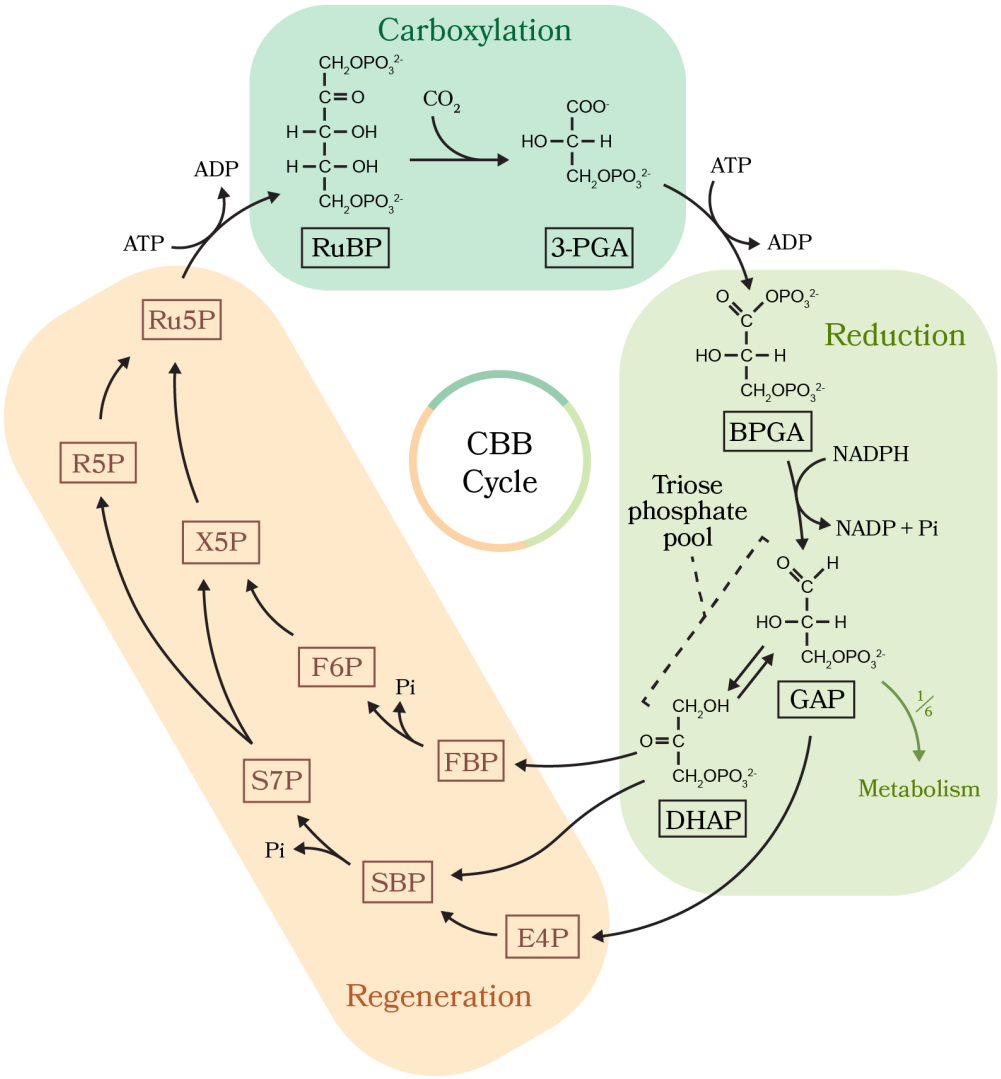


Figure 1.5: **The Calvin-Benson-Bassham (CBB) cycle.** Diagram outlining the reactions involved in the three phases of the CBB cycle: carboxylation, reduction and regeneration. During the carboxylation phase, atmospheric CO_2 reacts with ribulose 1,5-bisphosphate (RuBP) in the active site of Rubisco, briefly forming a six-carbon intermediate that quickly degrades to two molecules of 3-phosphoglycerate (3-PGA). During the reduction phase, the primary product of photosynthesis, glyceraldehyde 3-phosphate (GAP) is produced using NADPH generated in the light reactions, catalysed by the key NADPH-consuming enzyme GAPDH. GAP and dihydroxyacetone phosphate (DHAP) are in rapid equilibrium with each other and are collectively referred to as triose phosphate. Five out of every six molecules of triose phosphate produced in the reduction phase are used to regenerate RuBP in the regeneration phase via a series of intermediate sugar molecules. These include five steps via fructose 1,6-bisphosphate (FBP), fructose 6-phosphate (F6P), xylulose 5-phosphate (X5P) and ribulose 5-phosphate (Ru5P), and/or several additional steps involving the intermediates erythrose 4-phosphate (E4P), sedoheptulose 1,7-bisphosphate (SBP), sedoheptulose 7-phosphate (S7P) and ribose 5-phosphate (R5P).

be the most abundant protein on earth (Bar-On and Milo 2019). Rubisco catalyses carboxylation of the 5-carbon intermediate ribulose 1,5-bisphosphate (RuBP) into two 3-carbon molecules of 3-phosphoglycerate (3-PGA), sourcing the extra carbon atom from atmospheric CO_2. 3-PGA is then converted to 1,3-bisphosphoglycerate (BPGA) using the energy from the hydrolysis of ATP, catalysed by phosphoglycerate kinase. The enzyme glyceraldehyde 3-phosphate dehydrogenase (GAPDH) then converts BPGA to glyceraldehyde 3-phosphate (GAP) using electrons from NADPH. GAP is the primary carbohydrate product of photosynthesis and is a reactant in an array of metabolic processes. Every sixth GAP is released from the cycle for use in the synthesis of starch, amino acids, sucrose or other polysaccharides, while five GAP are used to regenerate RuBP, consuming further ATP and completing the second half of the cycle in five additional enzyme-catalysed steps (Fig. 1.5).

The light and dark reactions of photosynthesis are often studied separately but are tightly linked and inter-dependent. The activities of at least four of the CBB cycle enzymes including GAPDH are dependent on the chloroplastic thioredoxin systems, which function to alter the thiol redox state of cysteine groups in a wide range of photosynthesis enzymes, thereby modifying their activity (Buchanan 2016). Two thioredoxin system (TRX) function concurrently in Chlamydomonas, the Fd-TRX and NADPH-TRX systems (Nikkanen and Rintamäki 2019), themselves driven by the redox state of the stromal Fd and NADPH pools. Given that the major source of Fd^- and NADPH is from PSI-catalysed reduction of Fd^+, the TRX systems provide the regulatory link between the light and dark reactions. Indeed, the fact that the CBB cycle is light-activated suggests a fundamental misnaming of the dark reactions, with Bob Buchanan – one of the key discoverers of thioredoxin-based regulation – leading the calls for a renaming of the dark reactions as the carbon reactions of photosynthesis instead (Buchanan 2016). This renaming would reflect *in vivo* evidence that the CBB cycle does not function efficiently in the dark, but rather must function during illumination to continuously process the stromal ATP and NADPH produced by light-driven reactions (Buchanan 2016).

1.2.5 The pyrenoid and the CO_2-concentrating mechanism

Unlike vascular plants, and indeed many other species of algae, part of the CBB cycle in Chlamydomonas takes place in a liquid-like phase-separated micro-compartment called a pyrenoid nestled amongst the thylakoid membranes (Mackinder 2018) (Fig. 1.6A). The pyrenoid primarily comprises a densely packed matrix of tetrahedrally coordinated Rubisco held in place by a scaffolding protein called essential pyrenoid component 1 (EPYC1) (Mackinder et al. 2016). The pyrenoid is encased in a starch sheath that is punctuated by specialized tubules that traverse the pyrenoid matrix (Engel et al. 2015). These pyrenoid tubules are continuous with the rest of the thylakoids outside the pyrenoid, with the interior of the tubules continuous with

the thylakoid lumen. Within each tubule are bundled several minitubules that allow a continuous passage for stromal molecules such as the CBB cycle enzymes and reactants to enter the pyrenoid matrix (Engel et al. 2015). The membranes of the pyrenoid tubules are compositionally unique and are enriched in CO_2 transporters on the surfaces facing the pyrenoid matrix (Mackinder 2018). These transporters work to inject the pyrenoid with CO_2 from the lumen, concentrating CO_2 in the vicinity of Rubisco for more efficient carbon fixation (Fig. 1.6B). Atmospheric CO_2 reaches the thylakoid lumen via a series of dedicated transmembrane pumps in the cell membrane, the chloroplast envelope and the thylakoids. The current understanding holds that external inorganic carbon is transported across the cell membrane in the form of HCO_3^- by HLA3 and LCI1, then across the chloroplast envelope by LCIA (Mackinder 2018). Alternatively, CO_2 that diffuses into the chloroplast stroma can be converted to HCO_3^- by the carbonic anhydrase pair LCIB/LCIC (Mackinder 2018). Stromal HCO_3^- is then transported into the thylakoid lumen via bestrophin-like proteins BST1-3 (Mukherjee et al. 2019) and converted to CO_2 by the carbonic anhydrase CAH3. Injection of CO_2 into the pyrenoid matrix is aided by an as-yet unidentified pump, with the leading candidate being another bestrophin-like protein, RBMP1 (Meyer et al. 2020). These pumps, along with the pyrenoid, make up the CO_2-concentrating mechanism (CCM) of Chlamydomonas (Fig. 1.6B): an evolutionary adaptation to life in an aqueous environment where carbon availability is often severely limited by the low solubility of CO_2.

1.2.6 Photorespiration

A starch sheath around the pyrenoid prevents leakage of CO_2 and keeps Rubisco and the other key enzymes densely packed together to dramatically increase the CBB cycle rate and limit the unproductive reaction of Rubisco with O_2, termed photorespiration. Photorespiration uses O_2 as the electron acceptor in the oxygenase reaction of Rubisco, which competes directly with CO_2 at the Rubisco active site. Here, Rubisco splits the 5-carbon RuBP into one molecule of 3-PGA and one molecule of 2-phosphoglycolate (2-PG) instead of two molecules of 3-PGA produced during the carboxylation reaction. The subsequent photorespiratory pathway can recycle 2-PG to regenerate RuBP for the CBB cycle but this is energetically costly: two molecules of ATP consumed per every two molecules of 2-PG recycled, as well as the release of one molecule of CO_2 in the mitochondria (Betti et al. 2016). The CCM in Chlamydomonas, as well as those in C4 plants and cyanobacteria, helps to limit this CO_2 release and unproductive ATP usage. In C3 plants, however, carbon is released at a rate of one molecule CO_2 per four molecules fixed due to photorespiration (Sage et al. 2012). Alleviation of the energetic and carbon costs of photorespiration by genetic engineering is a promising route to improving yield of C3 crops, including vital food sources such as rice and soybean (Mackinder 2018). However, there is some evidence

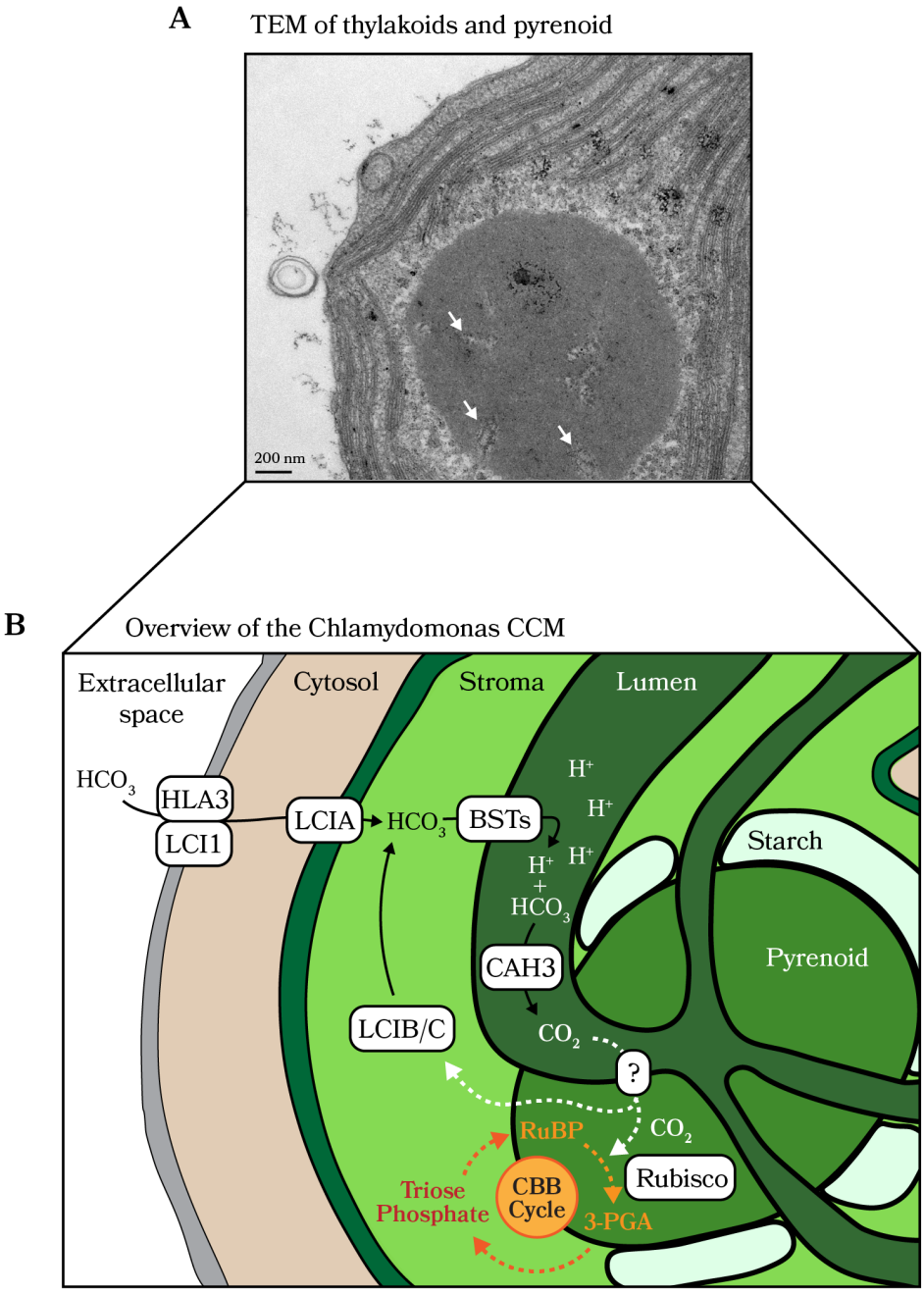


Figure 1.6: **The CO_2-concentrating mechanism.** **A.** Transmission electron micrograph of strain CC-4533 showing the pyrenoid. Pyrenoid tubules – modified thylakoid membranes continuous with the rest of the membranes in the chloroplast – can be seen within the pyrenoid matrix, indicated by white arrows. Electron micrograph is author's own. **B.** Simple schematic of the Chlamydomonas CCM, showing the key bicarbonate transporters that inject HCO_3^- into the thylakoid lumen, where it is converted to CO_2 using protons generated during photosynthesis. CO_2 is injected into the pyrenoid matrix via an as-yet-unidentified membrane pump or pore. The pyrenoid matrix mainly comprises Rubisco and is the site of the carboxylation phase of the CBB cycle, fuelled by the diffusion of RuBP into the pyrenoid and generating 3-PGA that diffuses back out into the stroma.

that photorespiration plays an important regulatory role in plants under fluctuating light conditions as a means of consuming excess stromal reductants (Huang et al. 2015). The products of photorespiration also feed into metabolic pathways involved in amino acid production and nitrogen fixation, so viewing photorespiration as purely wasteful may be an over-simplification (Betti et al. 2016). The high degree of inter-connectivity of regulatory and metabolic pathways provides an obstacle for artificial attempts to improve yields by bypassing photorespiration. In the wider view, any attempts to engineer improved photosynthesis for a new generation of crops face difficulties in avoiding upsetting photosynthetic regulatory processes. Plants possess a wide array of regulatory systems in chloroplasts that keep photosynthetic processes in balance, many of which are shared by Chlamydomonas. The following sections explores the triggers and effects of the major regulatory mechanisms in Chlamydomonas in more detail.

1.3 The need for regulation

The intensity and quality of irradiance reaching photosynthetic organisms from the sun can vary widely over the course of a day. Plants and green algae have evolved regulatory mechanisms to offset this variation, allowing the organisms to still meet their metabolic needs in suboptimal conditions. These include long-term adjustments to the stoichiometry of PSI and PSII by modifying gene expression in a process called developmental acclimation (Flannery et al. 2021). On shorter timescales, cells face two essential problems in maintaining productivity: the first is how to manage excess absorbed energy, the second is balancing the supply of NADPH and ATP from the thylakoids with CBB cycle demand.

1.3.1 Managing excess absorbed energy

Light is a dangerous resource on the nanoscale. Under high light conditions, the light harvesting complexes absorb more energy than can be processed by the downstream ETC. Excess absorbed energy results in an abundance of un-transferred excitons, increasing the chances of the formation of a chlorophyll triplet state in the reaction centre of PSII (^{3}P680) that can react with O_2 to form singlet oxygen ($^1O_2$). Meanwhile, light-driven electron transfer through the ETC can result in over-reduction of the stroma if the CBB cycle cannot consume the available NADPH fast enough, leaving electrons emerging from PSI more prone to react with O_2 to form superoxide radicals (O_2^-). Similarly, an over-reduced PQ pool in the thylakoids limits the rate at which electrons can emerge from the acceptor side of PSII, risking further $^1O_2$ production. $^1O_2$ and O_2^- are reactive oxygen species (ROS) and can damage the delicate Q_A/Q_B sites in PSII or F_A/F_B sites in PSI (Takagi et al. 2016) in a process called acceptor side limitation. The cumulative effect of extended acceptor

side limitation is a drop in photosynthetic efficiency called photoinhibition that can only be alleviated by spending time and energy to replace the damaged photosystems (Chaux et al. 2015) and adjust the proteome (Lempiäinen et al. 2022).

One strategy to avoid photoinhibition is to dissipate excess absorbed light before it reaches the photosystems. At the level of the light harvesting antenna, excitation energy can be rapidly dissipated as heat, termed energy dependent non-photochemical quenching (qE). This dissipation relies on the activation of a quencher pigment that can receive an exciton from chlorophyll and efficiently dissipate the received energy instead of transferring it. The identity of the quencher species is a topic of current research, but the activation of qE is known to be triggered by low lumenal pH brought on by the sharp acceleration of proton pumping driven by an upregulation of CET at the onset of high light (Munekage et al. 2002). This increased ΔpH results in the protonation of a PSII-associated subunit in plants called PsbS, and the activation of the lumenal enzyme violaxanthin de-epoxidase (VDE). VDE catalyses the conversion of the LHCII-integral pigment violaxanthin to zeaxanthin, which, together with PsbS, cause LHCII to aggregate and upregulate dissipation of excitation energy as heat. Chlamydomonas has a homologue of PsbS with 45% identity to the plant type that is similarly rapidly accumulated during high light treatment and can activate NPQ (Tibiletti et al. 2016). However, unlike plants, Chlamydomonas relies to an even greater extent on the upregulation of LHC-like proteins LHCSR1 and LHCSR3 under stress conditions, with LHCSR3 in particular contributing to the majority of qE induction (Peers et al. 2009, Tibiletti et al. 2016).

In addition to triggering qE, a high ΔpH also triggers down-regulation of the rate of electron flow through cytb_6f in a process called photosynthetic control. Cytb_6f sits at the crossroads of the photosynthetic ETC and represents the rate limiting step under normal growth conditions. This rate is further suppressed under stress conditions to protect the PSI acceptor side from oxidative damage by controlling the rate of electron donation from cytb_6f to PC (Joliot and Johnson 2011). The mechanism of this slow-down is a topic of ongoing research, but is thought to involve a modification to the Q-cycle (Tikhonov 2014). The prevailing theory postulates that a drop in the affinity of cytb_6f for PQH_2 is brought about by protonation of His128 of the Rieske ISP at the Q_o site. Below a lumenal pH of 6.2, His128 is constitutively protonated regardless of the oxidation state of the ISP Fe_2S_2 cluster (Fig. 1.3E), possibly inhibiting proton extraction from PQH_2 thereby slowing down the Q-cycle (Malone et al. 2021). While qE acts to protect PSII from excess light-induced ROS production, photosynthetic control acts to protect PSI by limiting electron transfer by limiting reduction of the PC pool. As a consequence, photosynthetic control also causes over-reduction of the PQ pool due to Q-cycle slow-down, leading to PSII acceptor side limitation.

1.3.2 Metabolic balancing and redox poise

The second main problem faced by cells is how to maintain metabolic balance in the stroma. The essence of the problem is in maintaining a stromal redox poise – an intermediate redox state between over-reduction and over-oxidation. Over-reduction can lead to ROS production at PSI, while over-oxidation can lead to under-production of ATP, causing the CBB cycle to stall and compromising cell growth. The CBB cycle in Chlamydomonas requires an input of CO_2, NADPH and ATP to power carbon fixation. For every three turns of the cycle, the carbon atoms of three molecules of CO_2 are fixed into GAP, requiring nine molecules of ATP and six molecules of NADPH: a ratio of 1.5:1. To fuel the CBB cycle with reactants, NADPH is produced during LET using two electrons from two molecules of Fd^-. In supplying these electrons, LET also translocates six protons into the lumen: two at PSII during the oxidation of water and four at cytb_6f during the Q-cycle. In Chlamydomonas, release of these protons through the ATP synthase generates ATP at a ratio of one per ~4.67 protons. This is because each full turn of the α/β head domain generates three ATP molecules and requires the sequential binding of a proton to all of the fourteen subunit-III monomers of the c-ring. This means that for every two electrons required to produce a single molecule of NADPH, six protons are translocated, and approx. 6 / 4.67 = 1.28 molecules of ATP are produced.

There is no evidence yet that Chlamydomonas can modulate the number of subunits in the c-ring to increase the efficiency of ATP production (Meyer Zu Tittingdorf et al. 2004). Alone, therefore, LET can supply the CBB cycle with ATP and NADPH at a ratio of 1.28:1, falling short of the required 1.5:1 ratio. Notably, 1.28:1 is itself an overestimation as it does not take into account uncontrolled proton leak through the membrane (Joliot and Joliot 2002). There is some uncertainty in the number of subunits in the c-ring (Harris 2009, p652), with some suggesting that Chlamydomonas has thirteen subunits (Meyer Zu Tittingdorf et al. 2004) while other photosynthetic eukaryotes contain fourteen, (e.g., spinach; Seelert et al. 2000). Even with thirteen subunits, however, Chlamydomonas would still run a shortfall at 1.39 ATP produced per NAPDH. Fluctuating conditions that require extra ATP production only exacerbate this problem, including the ATP required to power the CCM under low carbon conditions. Add to this the energetic costs of DNA and protein synthesis, nitrite reduction and other active chloroplastic processes, and it is clear that additional means of acquiring ATP without producing NADPH are required to make up the shortfall.

1.3.3 How is extra ATP generated?

Central to the problems of managing excess light and metabolic balancing is the generation of ΔpH. Regulatory processes either act to sense excess ΔpH as a trigger for qE and photosynthetic control or act to drive the production of ATP to make

up the ATP shortfall. As in vascular plants, Chlamydomonas has evolved several alternative electron transport pathways proposed to drive the production of ΔpH while diverting stromal electrons away from NADPH to redress the ATP/NADPH balance (reviewed succinctly in Alric and Johnson 2017). The most studied among the alternative electron transfer pathways and the major subject of this thesis is CET, in which stromal electrons return to the PQ pool via Fd or NADPH. To meet the ATP shortfall, CET is thought to function at a constant base rate during normal cell growth while, under high light or stress conditions, CET is upregulated to drive the extra production of ΔpH required to trigger qE and photosynthetic control (Chaux et al. 2015).

Alternatively, electrons emerging from PSI (and from PSII) can be used to reduce O_2 to H_2O in the stroma in an array of processes collectively termed O_2-photoreduction. O_2 is a well-known waste product of oxygenic photosynthesis and is produced by water-splitting at PSII, but it is also a major acceptor of stromal electrons emerging from the ETC. These include photorespiration, which competes with the CBB cycle as mentioned above, but also chlororespiration (Nawrocki et al. 2019c), the Mehler reactions around PSI (Miyake 2010), and the reactions catalysed by the stromal flavodiiron proteins (Flvs) (Chaux et al. 2017). These reactions all involve electron flow from water at the OEC to water in the stroma, and this water-to-water behaviour along with their lack of net O_2 evolution have led to their combined description as pseudo-cyclic electron flow (PCEF) (Heber 2002). In parallel, stromal electrons can be transported to the mitochondria in a process termed chloroplast-to-mitochondria electron flow (CMEF). Here, stromal NADPH is consumed by NADPH-malate dehydrogenase to produce malate, which can then be transported to the mitochondria, converted to NADH and used in respiration to produce ATP. CMEF can be viewed as a valve by which excess NADPH is processed to balance the ATP/NADPH ratio, as in CET and the PCEF pathways.

Together CET, PCEF and CMEF can produce extra ATP to power the cell and trigger photosynthetic regulation (Fig. 1.7A). However, they are also crucial in Chlamydomonas for powering the CCM, which likely consumes ΔpH by using lumenal protons for the conversion of HCO_3^- to CO_2 by CAH3 in the pyrenoid tubules (Burlacot et al. 2022, Mukherjee et al. 2019) (Fig. 1.6B). PCEF in particular is instrumental in driving the majority of ΔpH generation under very low CO_2 conditions, but there is likely a functional redundancy between the alternative pathways, with mutants in any one of the CMEF, PCEF or CET able to compensate for the loss of ATP production by upregulating the others (Dang et al. 2014). Under low oxygen (referred to hereafter as anoxic) conditions, however, PCEF is diminished, and so CET is upregulated to compensate (Alric 2014). Anoxia also triggers a form of anaerobic metabolism in Chlamydomonas in which cellular starch reserves are metabolised to generate ATP and pyruvate via glycolysis (reviewed in Yang et

al. 2015). Under this metabolism, stromal Fd^- can be consumed by the stromal hydrogenase HYDA to produce H_2 (Sawyer et al. 2017), either using electrons from pyruvate in the dark or from PSI in the light. Chlamydomonas mutants deficient in CET, though unable to efficiently trigger qE (Munekage et al. 2002, Munekage et al. 2004) or photosynthetic control (Suorsa et al. 2012), have found a promising use in biohydrogen production under anoxic conditions since CET no longer competes with HYDA for stromal electrons, resulting in a large boost in H_2 production (Steinbeck et al. 2015).

1.4 Cyclic electron transfer

CET is a physiologically significant process that represents a missing link in our knowledge of photosynthetic electron transfer. The current view is that CET functions both to maintain CBB cycle efficiency by balancing the ATP/NADPH ratio and to trigger the photoprotection of PSII (by qE) and PSI (by photosynthetic control). However, despite half a century of research, the precise mechanisms of CET are yet to be fully elucidated. The work of Arnon and colleagues at Berkeley between 1954 and 1964 provided the first experimental evidence for a slower form of Fd-dependent photophosphorylation in solutions of broken spinach chloroplasts that did not produce NADPH (Tagawa et al. 1963). Ever since, difficulties in measuring CET *in vivo* have hampered progress; CET produces no unique products and is intimately entangled with LET, utilizing the same machinery as LET, running in parallel to LET, and partially dependent on LET for a supply of electrons. The current consensus is that stromal electrons can re-enter the PQ pool by two main routes, either directly from the oxidation of NADPH via the NDH-dependent pathway or from Fd^- via the PGR5/PGRL1-dependent pathway.

1.4.1 NDH-dependent pathway

In cyanobacteria and vascular plants, the NDH-dependent pathway (NDH-CET) is catalysed by the NDH complex, now called photosynthetic complex I (PCI). PCI is a large, multi-subunit, proton-pumping PQ reductase that bears strong resemblance to the mitochondrial NADH:ubiquinone oxidoreductase (respiratory complex I), though receives electrons from Fd^- instead of NADPH (Yamamoto et al. 2011). In Chlamydomonas, though, PCI is not present and instead the pathway is catalysed by NDA2, a peripheral membrane protein with NADPH:PQ oxidoreductase activity present at low levels in Chlamydomonas thylakoids (Desplats et al. 2009). Measurements of its maximal rate suggest that NDH-CET cannot account for more than a fraction of total CET; the rate of non-photochemical PQ reduction for NDA2 was ascertained at ~2-4 e^- s^{-1} PSI^{-1} (Nawrocki et al. 2019a, Nawrocki et al. 2019c), while the maximum measured rate of CET is ~60 e^- s^{-1} PSI^{-1} in Chlamydomonas

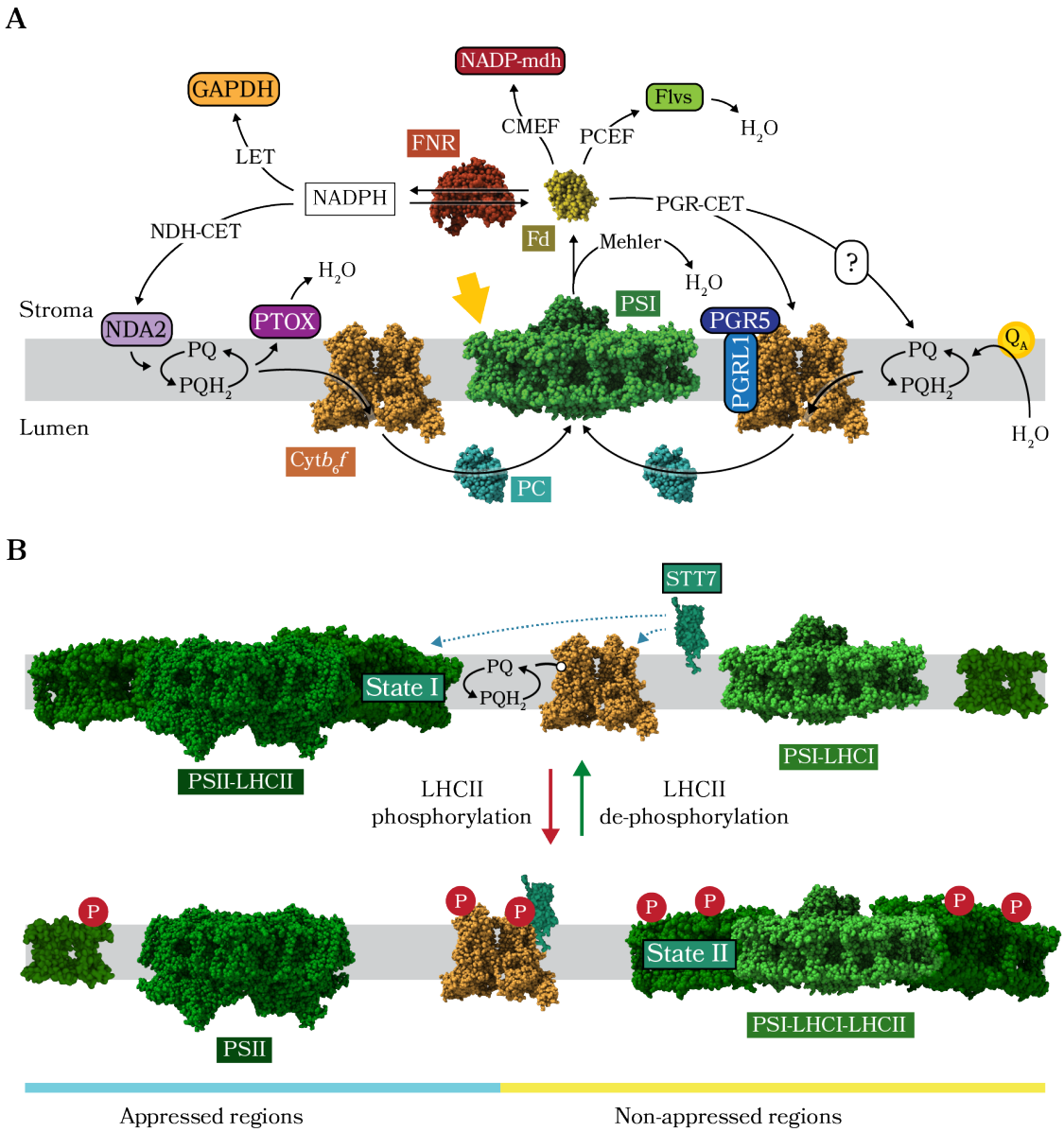


Figure 1.7: **Metabolic balancing in the chloroplast. A.** Diagram of alternative electron flows in Chlamydomonas. Electrons enter the PQ pool via the Q_A site of PSII or by CET, either via NDH-CET or the PGR-CET pathway, which involves and unknown intermediate. Electrons exit the PQ pool via chlororespiration (PTOX) or through cytb_6f to PC, then PC to PSI. Electrons exiting the acceptor side of PSI are received by Fd, which can transfer electrons to a wide variety of acceptors, including via the PCEF pathways (Mehler, Flvs), the CMEF pathway (NADP-MDH), to the CBB cycle (via FNR to NADPH), or recycled back into the PQ pool in CET. **B.** State transitions involve phosphorylation induced net migration of mobile LHCII trimers from PSII in state I to PSI in state II. This increases the light harvesting capacity of PSI in the non-appressed membrane regions while decreasing it for PSII in the appressed regions. Cytb_6f is present in both regions. The protein structure for STT7 is represented by *Micromonas spp.* STT7 (PDB 4IX3; Guo et al. 2013). Other protein stuctural models shown are the same as in Figure 1.2.

(Nawrocki et al. 2019a) and over 100 e^- s^{-1} PSI^{-1} in plants (Joliot et al. 2004). The current view, therefore, is that NDH-CET in Chlamydomonas mainly functions as part of the chlororespiratory pathway (Nawrocki et al. 2019b), a vestigial respiratory pathway present in chloroplasts that helps to ameliorate stromal reduction by transferring electrons from NADPH to O_2 via plastid terminal oxidase (PTOX) with no proton-pumping involved (Fig. 1.7A).

1.4.2 PGR5/PGRL1-dependent pathway

The PGR5/PGRL1-dependent pathway (PGR-CET) is the major CET pathway in plants and algae and is regulated by the small, soluble protein PGR5 and its binding partner, the membrane associated PGRL1. While the mechanism of NDH-CET is broadly agreed upon, the mechanism of PGR-CET and the details of the involvement of PGR5 and PGRL1 are a topic of intense research. Early research established the sensitivity of PGR-CET to the inhibitor antimycin A (Tagawa et al. 1963). This was later built upon by Bendall and Moss who observed a lack of direct effect of antimycin A on the turnover of haem b_H in vitro using isolated chloroplasts, implying the inhibitor does not specifically target the cytb_6f Q-cycle (Moss and Bendall 1984). The conclusion drawn then was the same as today; antimycin A must inhibit a reaction unique to CET, the reduction of PQ by Fd. This led to an ongoing hunt for an enzyme or set of enzymes that can perform Fd:PQ reduction (FQR) activity at a high enough rate to satisfy in vivo measurements (~60 e^- m^{-2} s^{-1}) and is sensitive to antimycin A.

The regulatory enzyme PGR5 was identified as a CET-effector protein in 2002 by observation of Arabidopsis mutants with severely deficient induction of qE (Munekage et al. 2002). Though clearly important for upregulation of CET for photoprotection, PGR5 itself is unlikely to possess FQR activity as it does not bind any cofactors (Hertle et al. 2013). What is more, the rate of P700 re-reduction after a flash of light was shown to be similar in wild type and PGR5-deficient mutants of Arabidopsis (Nandha et al. 2007). A more likely role for PGR5 is in the regulation or stabilization of a protein or complex that acts as the FQR, probably in complex with PGRL1. Though PGR5 is present at a lower level than PGRL1, their expression appears to be closely linked and they form a stable complex *in vivo* that can interact with cytb_6f, PSI and FNR (DalCorso et al. 2008). PGRL1 was confidently put forwards as the FQR in 2013 when an Arabidopsis PGRL1 appeared to show antimycin A-dependent FQR activity *in vitro* (Hertle et al. 2013) when in the presence of PGR5. However, a recent study with a PGRL1-deficient Chlamydomonas mutant showed no change in the maximal rate of PGR-CET, only the extent or duration, implying that it too is a regulatory protein (Nawrocki et al. 2019a). Supporting this, a 2021 study on the consequences of the loss of PGRL1 in Arabidopsis clearly demonstrated that PGRL1 is not required for CET to function or to enable growth under fluctuating light (Rühle et al. 2021).

So, what is the function of PGR5/PGRL1 in Chlamydomonas? Since CET is still observable in mutants of either protein, if only transiently, they are unlikely to be vital electron transfer intermediates. The consequences of their loss for the cell appear to show they have an important role as a means of upregulating CET under certain conditions, potentially through an interaction with cytb_6f. The question of the identity of the FQR is an enduring mystery. Mutagenesis studies have been instrumental in discovering the importance of PGR5, PGRL1 and NDA2, yet it has not been possible to isolate a mutant devoid of all CET but with LET left untouched (Nawrocki et al. 2019b). This may suggest there are in fact no unique players in the mechanism of CET, only a unique FQR reaction. If so, examining how the known players in CET – the CET effector proteins – interact may be key to discovering the FQR. This includes building a clearer understanding of what occurs at a molecular level when CET is upregulated.

1.4.3 Controlling the balance of LET and CET

Though PGR5/PGRL1 are likely involved, details of the mechanism by which upregulation of CET occurs are still unclear. Current competing theories mostly involve a compositional rearrangement of the thylakoid membranes that facilitates CET at the expense of LET. On theory builds on evidence from C4 plants that suggests two isoforms of Fd coexist in the chloroplast in competition with one another, one largely active in LET and the other in CET (Goss and Hanke 2014). The two isoforms are hypothesised to possess different binding affinities to FNR depending on whether FNR is bound to the membrane or soluble. This suggests a dynamic role for FNR in controlling the balance between LET and CET (Goss and Hanke 2014). Alternatively or in parallel, upregulation of CET might involve modulating the size of a pool of cytb_6f in the thylakoids sequestered for CET and excluded from LET. Indeed, one of the earliest recognised and clearest conditions under which upregulation of CET was observed was in parallel with state transitions (Finazzi et al. 1999), thought to occur along with a migration of cytb_6f to the non-appressed thylakoids (Vallon et al. 1991).

State transitions are controlled by the phosphorylation and dephosphorylation of thylakoid proteins (Vener et al. 1997, Zito et al. 1999). In Chlamydomonas, a transition from state I to state II is triggered by phosphorylation of LHCII by the peripheral thylakoid membrane protein state transition 7 (STT7) (Depège et al. 2003). This leads to a net decrease in the connectivity of LHCII with PSII in the appressed thylakoids regions and a net increase in their connectivity with PSI in the non-appressed regions, forming PSI-LHCI-LHCII supercomplexes (illustrated in Fig. 1.7B). STT7 itself is activated by the redox state of the PQ pool (Allen 1992) and functions as a negative feedback mechanism for PSII over-excitation: high PSII activity leads to over-reduction of the PQ pool, triggering STT7 activation,

phosphorylation of LHCII and the transition to state II, with a concomitant decrease in PSII excitation. Conversely, when the PQ pool is oxidised, thylakoid dephosphorylation is carried out by the combined activity of the constitutively active PPH1 and PBCP phosphatases in Chlamydomonas (Cariti et al. 2020), triggering net migration of LHCII to PSII. The portion of mobile LHCII in Chlamydomonas is particularly large compared to vascular plants; approximately 80% of the total LHCII antenna is mobile during transition from state I to state II, with a consequent rise of approx. 50% in both the PSI absorption cross-section and net energy efficiency (quantum yield) of PSI (Delosme et al. 1996). Since CET requires PSI excitation to cycle electrons from PC to Fd, a boost in PSI excitation appears at first to provide a simple explanation for the correlation between state II and high CET. The parallel migration of cytb_6f to the non-appressed, PSI-rich thylakoid regions (Vallon et al. 1991) may also be part of the explanation for a correlation between state II and high CET. However, mutant studies have now demonstrated that the two processes are extricable; mutants deficient in STT7 can still upregulate CET even without inducing a transition to state II (Takahashi et al. 2013).

STT7 activation and CET upregulation are nevertheless likely to be closely interconnected due the interaction of STT7 with CET effector proteins. The mechanism of STT7 activation is a topic of current research, but evidence that STT7 phosphorylates the PetO subunit of cytb_6f first suggested a direct interaction with the complex (Hamel et al., 2000). Current evidence from cytb_6f mutants suggests that STT7 binds transiently to subunit IV of cytb_6f and is activated by confirmational changes brought on by modification of the Q_o site in response to the PQ pool redox state (Harris 2009, p618, Dumas et al. 2017). In fact, recent research measured an inhibitory effect of antimycin A on cytb_6f function in vivo, but only when STT7 was present in the cell, suggesting a possible role for STT7 in the mechanism of PGR-CET itself (Buchert et al. 2022). STT7 has also been implicated in the activation of the lumenal carbonic anhydrase CAH3, part of the CCM, which is also thought to be phosphorylation controlled (Blanco-Rivero et al. 2012). Subsequent research demonstrated that induction of the CCM under very low CO_2 is correlated with upregulation of CET in Chlamydomonas, implying that cells boost CET to provide extra ATP for the CCM (Lucker and Kramer 2013), a process now known to occur concurrently with an upregulation of PCEF and CMEF as well (Burlacot et al. 2022). It appears, then, that state transitions and CET upregulation are also intimately connected to induction of the CCM, with the common factor likely being the need to generate extra ATP under scarce conditions (Burlacot et al. 2022).

The trigger for the upregulation of these processes may involve calcium signalling, with calcium dependency having recently been discovered for both CET (Terashima et al. 2012) and the CCM (Wang et al. 2016). Indeed, a physical interaction between the Ca^{2+}-sensor protein, CAS, and CET effector proteins was later reported as part of

a putative CET-driving supercomplex isolated from Chlamydomonas cells in state II.

1.4.4 Discovery of a CET-supercomplex

A key result was reported in 2018 with the characterization of a supercomplex isolated from the thylakoids of cells in anoxia-induced state II (Steinbeck et al. 2018). Evidence from electron micrographs and mass spectrometry appeared to show cytb_6f bound to PSI along with roughly stoichiometric quantities of FNR, PGRL1, CAS, and the anaerobic response protein, ANR1. This was considered a long-awaited follow-up to a 2010 publication reporting the isolation and biochemical characterization of a similar supercomplex containing PSI, cytb_6f (including PetO), FNR, PGRL1, ANR1 and CAS from cells chemically induced to enter state II by uncoupling the trans-thylakoid proton gradient (Iwai et al. 2010). Joliot and Joliot (2002) were the first to propose that the partitioning between LET and CET could be supercomplex controlled, and their idea built on an early envisioning of CET involving a thylakoid associated Fd and FNR bound to PSI and forming a direct interaction with cytb_6f (Carrillo and Vallejos 1983). The functional relevance of the putative CET-supercomplex is still a topic of debate, including whether close association of the proteins would work to increase CET rate. While the 2018 supercomplex was not measured *in vitro* (Steinbeck et al. 2018), the 2010 supercomplex displayed some slow activity for the oxidation of cytochrome f and reduction of haem b_H in solution (Iwai et al. 2010), though it was not clear whether the assembly was still intact during these measurements. It is also notable that this activity was found to be insensitive to antimycin A, and that no PGR5, STT7 or Fd were detected in the supercomplex preparations.

1.4.5 FNR involvement in CET

Though the purpose of the CET supercomplex is still unclear, the inclusion of FNR in the isolated samples supports a theory that FNR catalyses stromal electron donation to PQ via cytb_6f (Moss and Bendall 1984). That cytb_6f itself might be the site of re-entry of stromal electrons to the PQ pool was originally proposed by Peter Mitchell (Mitchell 1975) and has regained favour in the past three years. Cytb_6f is, after all, a well-known PQ reductase, and the original formulation of the Q-cycle postulated that PQ might be double-reduced at the Q_i site not by two electrons stored in the low potential chain, but by one electron from haem b_H and one electron from the stroma (Mitchell 1975). Further, Fd can be modelled into a putative binding site on the stromal edge of cytb_6f, bringing its Fe_2S_2 cluster in range of electron transfer to the Q_i site haems b_H and c_i. Early *in vitro* evidence ruled out the direct influence of Fd on the redox state of haem b_H in broken chloroplasts (Cox 1979), but the possibility still remains that FNR could bind at the stromal side of cytb_6f and facilitate PGR-CET (Clark et al. 1984, Shahak et al. 1981). This interaction may be stabilized by the

formation of a CET-supercomplex and/or by interaction with PGR5. Recent evidence suggests that PGR5-deficient Chlamydomonas operate a slightly less efficient Q-cycle during CET, with the conclusion being that PGR5 stabilises an interaction between FNR, Fd and the stromal face of cytb_6f during CET-upregulating conditions to efficiently perform the FQR reaction (Buchert et al. 2020). Gathering structural and mechanistic evidence for this modified, FNR-dependent, Fd-assisted Q-cycle is a topic of ongoing research.

1.5 Addressing key questions in CET with molecular biology

Understanding the mechanism and regulation of CET in Chlamydomonas is key to a full understanding of bioenergetics in this organism. This will benefit algae research itself, with applications in efficient H_2 photoproduction (Steinbeck et al. 2015). Equally it will benefit plant research owing to the similarities between plant and algal PGR-CET and the importance of qE and photosynthetic control for crop productivity (Foyer et al. 2012). Key questions remain, however, and our increasing ability to genetically manipulate photosynthesis in Chlamydomonas may prove central to addressing them.

1.5.1 The key questions

1.5.1.1 Is FNR the FQR?

Could a complex of cytb_6f, Fd and FNR be the elusive FQR? An important role for FNR in CET has long been hypothesised (Shahak et al. 1981). Studies with FNR-specific inhibitors showed decreased CET rate (reviewed in Cleland and Bendall 1992), and later strong binding of FNR to cytb_6f that survived several purification steps was observed (Clark et al. 1984). More recently, FNR was co-purified with cytb_6f (Zhang et al. 2001), and identified as part of the 2018 CET supercomplex (Steinbeck et al. 2018). However, *in vitro* measurements of a purified cytb_6f-FNR complex in the presence of PGR5, PGRL1 and Fd yielded only a slow Fd-dependent reduction of haem b_H (Hertle et al. 2013). Further in vitro work to stabilize and measure the putative complex are needed, but the purification has proved difficult to replicate and is an ongoing goal of CET research.

1.5.1.2 Does the PSI-cytb_6f supercomplex form *in vivo* to drive CET?

Close association between PSI and cytb_6f was first inferred from the co-migration of these proteins on sucrose gradients (Iwai et al., 2010), and then observed in electron micrographs of cross-linked samples (Steinbeck et al. 2018). This cross-linking likely precluded accurate measurements of *in vitro* CET on the sample, so

it is still unclear whether the supercomplex can function to perform CET. In very recent research, Joliot and colleagues presented spectroscopic data supporting the sequestering of pools of trapped or mobility-restricted PC in anoxic conditions that may indicate *in vivo* supercomplex formation (Joliot et al. 2022). This exciting result was not observed in STT7-deficient mutants, and forms the latest evidence for STT7-dependent, supercomplex-associated CET. Despite advances in *in situ* cryo-electron tomography, *in vivo* formation of the supercomplex is yet to be structurally observed. Such a result would open the door to a deeper understanding of the conditions in which it forms and dissolves, its dependency on STT7 and possibly the mechanism of FQR activity itself.

1.5.1.3 What is the mechanism of STT7 interaction/activation at cytb_6f?

STT7 phosphorylation is implicated in state transitions (Depège et al. 2003), CCM induction and modification of cytb_6f (Buchert et al. 2022, Hamel et al. 2000), clearly indicating an important regulatory role. Lack of a high-resolution structure of STT7 alone or in complex with LHCII, cytb_6f or CAH3 suggest it forms only labile, transient interactions, making mechanistic conclusions difficult to draw. Achieving the structure of a STT7-cytb_6f complex in particular would be a key step in understanding STT7 activation, perhaps helping to clarify the mechanism by which cytb_6f can sense the redox state of the PQ pool.

1.5.1.4 What is the mechanism of photosynthetic control?

Photosynthetic control is a major protective mechanism that plays an important role in determining growth rate in crops (Foyer et al. 2012). A key question here is how cytb_6f senses and responds to the drop in lumenal pH. The discovery of the importance of PGR5 for efficient photosynthetic control implies a close relationship between CET and the mechanism of the Q-cycle. The Q-cycle has proven a particularly difficult mechanism to probe, but recent evidence that PGR5-deficient mutants run an inefficient Q-cycle under anoxic conditions provides further reason to investigate (Buchert et al. 2020). If there is an Fd-assisted Q-cycle, is this involved in the mechanism of Q-cycle slow-down during photosynthetic control?

1.5.2 Chlamydomonas as a molecular testbed for engineering CET

Forty years of research and progress developing the molecular tools for genetic engineering has helped cement Chlamydomonas as one of the foremost eukaryotic microalgae in photosynthesis research. Complete sequences are available for the 110 Mb nuclear genome, the 205 kb chloroplast genome and the 15.8 kb mitochondrial genome. A host of resources support the constant expansion of the biotechnological toolbox for Chlamydomonas, including well-defined protocols

for mutagenesis, plasmid repositories, published biochemical assays and strain collections. This wealth of information has already led to discoveries relevant to CET research. Purification of cytb_6f from Chlamydomonas enabled early structural characterisation of the intact complex by protein crystallography in 2003, leading to the first identification of the Q_i site haem, c_i (Stroebel et al. 2003). Following this, modification of the gene encoding subunit IV of cytb_6f in Chlamydomonas enabled a clearer understanding of conditions for STT7 activation (Dumas et al. 2017). The following sections describe commonly used techniques and methods for engineering Chlamydomonas.

1.5.3 Engineering the chloroplast genome

The chloroplast genome was first assembled in 2002 and updated by Merchant in 2018, revealing 71 protein-coding genes, 29 tRNAs and 8 RNA genes (Gallaher et al. 2018, Maul et al. 2002). Though typical for a chloroplast genome, gene loss or gene transfer to the nucleus over evolutionary history has resulted in a genome thought to be only ~10% the size of the original endosymbiont progenitor of the chloroplast (Jackson et al. 2021). Still encoded in the chloroplast are many of the core polypeptides of the CET machinery, including the PSI core subunits PsaA/B, and cyt b_6, cyt f and subunit IV of cytb_6f.

The history of photosynthesis research in Chlamydomonas relies heavily on chloroplast engineering. This was first demonstrated 1988 using cell bombardment with DNA-coated microparticles (biolistics) to rescue a mutation in the ATP synthase gene, atpB (Boynton et al. 1988). Biolistics is a widespread technique and is often highly efficient but requires specialist equipment. An alternative, though less efficient method involves vortexing a suspension of cells and DNA with glass beads (Larrea-Alvarez et al. 2021). Either way, DNA integration occurs by homologous recombination into one or several of the ~40-100 copies of the genome contained in each Chlamydomonas chloroplast (Jackson et al. 2021). This necessitates subsequent rounds of single colony isolation and selection to drive transformed lines to homoplasmy, requiring 2-3 weeks in Chlamydomonas as opposed to several months in plants.

The facultative heterotrophy of Chlamydomonas is ideal for making severe photosynthetic mutations. Gene deletions or successful introduction of exogenous DNA can be selected for by appending antibiotic resistance cassettes, such as that encoding aminoglycoside adenine transferase (aadA), which confers resistance to spectinomycin (Goldschmidt-Clermont 1991). Restoration of photosynthetic mutants is also useful as a selection technique. This includes an efficient method to rescue the strain TN72, which contains a mutated psbH gene disrupted by the aadA cassette. Here, transformed genes are appended with an intact psbH gene to simultaneously replace the aadA cassette and repair the psbH locus, restoring PSII function and

so enabling selection by restoration of photosynthetic growth (Larrea-Alvarez et al. 2021). For modification of wild type strains, though, recent advances in marker-free cloning now enables multiple cycles of transformation and selection using the same resistance cassette. Here, the aadA cassette is excised by negative selection between each cycle, enabling its use for subsequent transformations (Jackson et al. 2022).

1.5.4 Engineering the nuclear genome

The nuclear genome in Chlamydomonas is much larger, containing 17,741 coding genes, encoding many of the key CET-effector proteins including FNR, PGR5, PGRL1, STT7, PetO, CAS, ANR1 and all six isoforms of Fd. Nuclear-encoded chloroplastic proteins are translated in the cytosol and targeted to the chloroplast by N-terminal transit peptides, which must be included on exogenous (transformed) genes for correct localisation. Transformation is achieved by biolistics, agitation with glass beads or most efficiently by electroporation (Shimogawara et al. 1998). Homologous recombination of transformed DNA is rare, and most nuclear transformations instead result in random integration into the genome (Zhang et al. 2014). Obtaining mutants in specific nuclear genes has therefore required UV, chemical or insertional mutagenesis followed by subsequent mapping of the mutations by sequencing to identify disrupted loci. Recent high-throughput efforts have resulted in a library of over 62,000 mutants covering 83% of protein-coding genes (Li et al. 2019).

Subsequent rescue of mutants expressing tagged or labelled versions of native genes (cisgenes) is hampered by locus-specific expression patterns, which make it unlikely that randomly integrated exogenous genes will show the same expression levels as their native counterparts. Successful expression usually requires appending an antibiotic resistance cassette to the transgene then maintaining selection pressure with antibiotics, though transformed genes may still fail to express even if adjacent to a resistance cassette (Caspari 2020). This may be due to the presence of highly effective native systems of foreign gene silencing (reviewed in Schroda 2019). The efficiency of gene silencing in Chlamydomonas is still a mystery, but likely occurs by any of three mechanisms; endonucleolytic cleavage of foreign DNA (Zhang et al. 2014), epigenetic suppression (Cerutti et al. 1997) or post-transcriptional gene silencing. A recent coordinated research effort to identify the causes of gene silencing led to the generation of two strains derived from UV mutagenesis that showed high expression of transgenes, UVM4 and UVM11 (Neupert et al. 2009). Both strains are now in widespread use as biotechnology workhorses, and recent research identified a mutation in a single gene (Cre10.g462200) encoding a histone deacetylase involved in epigenetic silencing as the cause of their phenotype (Neupert et al. 2020).

Post-transcriptional silencing systems in Chlamydomonas can still lead to poor expression but can be ameliorated by correct transgene design. For example, ensuring matching codon usage with the nuclear genome (Barahimipour et al. 2015) and

including native introns (Baier et al. 2018) yield substantial improvements in transgene expression. Harnessing of the post-transcriptional silencing system, known as RNA interference (RNAi), has also been used to good effect in Chlamydomonas for the targeted generation of mutants (Schroda 2006). RNAi as a phenomenon was first described in 1998 in *C. elegans* (Fire et al. 1998) and involves the expression of double-stranded RNA (dsRNA) with homology to the target mRNA, leading to its targeted degradation, thereby preventing translation. dsRNA homologous to a gene of interest can be introduced as an inverted repeat sequence on a plasmid or transgene into either the nuclear (Schroda et al. 1999) or chloroplast genomes (Charoonnart et al. 2019).

1.5.5 Targeted genome editing with CRISPR-Cas9

Recent advances in CRISPR-Cas9 technology have made reliable, targeted genome editing possible in Chlamydomonas. Initially discovered in prokaryotes, CRISPR (clustered regular interspersed short palindromic repeats) loci present in bacterial and archaeal genomes encode short sequences copied from virus DNA from past viral infections that are used to protect against future infections. This is achieved by complexing of the mRNA products of these short sequences (called guide sequences) with a Cas (CRISPR-associated) nuclease to form a guide RNA-Cas ribonucleoprotein. The mature ribonucleoprotein can cause a double stranded break in any accessible DNA with homology to the guide RNA providing it is flanked by a short recognition sequence called the protospacer adjacent motif (PAM) (Fig. 1.8). The discovery that these guide sequences could be customised for selective DNA targeting and the subsequent adaptation of CRISPR-Cas9 as a programmable genome editing tool by Doudna and Charpentier in 2012 (Jinek et al. 2012) has led to its widespread use for editing genomes in any transformable organism.

Successful adaptation for editing the Chlamydomonas nuclear genome was reported in 2014, with Cas9 and the guide RNA expressed from transformed plasmids (Jiang et al. 2014). A DNA-free method in which the ribonucleoprotein is complexed *in vitro* then electroporated directly into cells has now been developed for Chlamydomonas (Baek et al. 2016, Picariello et al. 2020), and demonstrated by successful targeting of photosynthetic genes (Tokutsu et al. 2019). Additional nucleases that recognise alternative PAM sequences, such as Cas12a, have been used to good effect as well (Ferenczi et al. 2021), but Cas9 from *Saccharomyces pyogenes* is still widely used due to its short, common PAM sequence of *NGG* (Fig. 1.8B).

To make targeted genome modifications with CRISPR-Cas9, exogenous DNA must be co-transformed or co-expressed with the CRISPR materials. Repair of CRISPR-induced double stranded breaks by homology directed repair is rare, usually occurring instead via the non-homologous end-joining pathway (NHEJ). The NHEJ pathway can be hijacked to insert a gene of interest during the repair process,

resulting in a targeted insertional mutant (Ghribi et al. 2020). Moreover, recent advances in our understanding of Chlamydomonas repair pathways has led to the recognition that single stranded DNA templates can facilitate relatively efficient homology directed repair of double stranded breaks and allow introduction of small, targeted sequence modifications (Akella et al. 2021, Ferenczi et al. 2021).

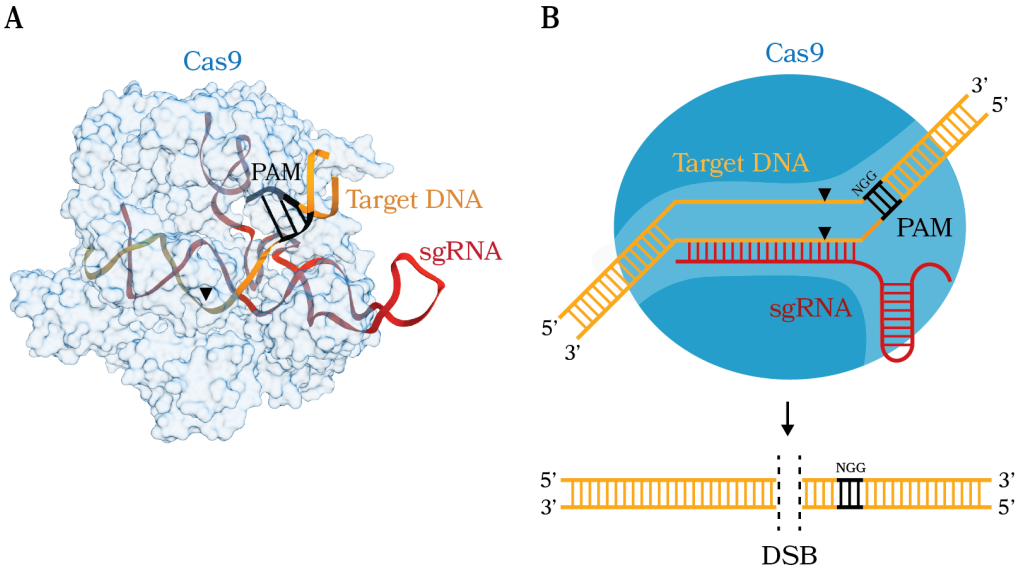


Figure 1.8: **Cas9-mediated DNA recognition and cleavage. A.** Structure of Cas9 from *S. pyogenes* in complex with the single-guide RNA (sgRNA; red) and a segment of target DNA (orange), shown as ribbons indicating the position of the ribose or deoxyribose backbones (PDB 5B2R; Hirano et al. 2016). The three-base PAM sequence and its reverse complement are shown in black, with the approximate position of the cleavage site indicated on the target DNA by the black arrow approximately four bases upstream. **B.** Diagram of the Cas9-sgRNA-DNA complex shown in **A**. Cleavage by Cas9 results in a double-stranded break (DSB) in the target DNA that can be repaired by the error-prone NHEJ pathway. In selected cases, the NHEJ pathway may integrate fragments of exogenous DNA as an insertion when repairing the DSB. This process is harnessed in Chapters 3 and 4 to generated directed insertion mutants of Chlamydomonas.

1.5.6 Cloning nuclear genes with recombineering

Both the localisation or purification of tagged target proteins and the rescue of mutants generally require cloning or synthesis of a gene of interest. Cloning from the nuclear genome by PCR faces several problems particular to Chlamydomonas. PCRs show reduced efficiency with very long or repetitive target sequences and are affected by the GC content of the template. The Chlamydomonas genome has a relatively

high GC content (~68% for coding regions) and an abundance of long and complex genes; the average gene is over 5000 bp long, contains ~7 introns and approx. one repeat-rich region per kilobase (Emrich-Mills et al. 2021). Synthesis or amplification of cisgenes without introns from complementary DNA (cDNA) is an important alternative that can avoid direct amplification from gDNA. This tactic can improve PCR efficiency by reducing the gene length, but exclusion of introns also decreases the expression efficiency of cisgenes, as mentioned above (Baier et al. 2018). cDNA-derived genes also miss out endogenous regulation by *cis* regulatory sequences and alternative splicing (Sarov et al. 2006). For cisgene cloning aimed at mutant rescue or native-like expression levels, a safe approach is to include the native promoter region, 5'UTR, coding sequence plus introns, 3'UTR and terminator region. Cisgenes built using DNA from large genomic clones such as bacterial artificial chromosomes (BACs), fosmids or cosmids offer a useful alternative to gDNA or cDNA. Recently, a BAC library containing the Chlamydomonas nuclear genome was made available to the research community (Brazelton et al. 2018). This resource made possible the use of homology-dependent, recombination-mediated genetic engineering, termed recombineering, to retrieve whole Chlamydomonas genes without PCR amplification. Similar to the well-established method of homology-dependent engineering in yeast genetics (Szostak et al. 1983), recombineering involves exchange of DNA between two sequences flanked by relatively short regions of homology, in some cases requiring less than 50 bp (reviewed in Court et al. 2002). gDNA retrieved from BACs contain all the native *cis* regulatory features, and can be inserted into a plasmid that can be propagated and manipulated in *E. coli* before re-introduction to Chlamydomonas.

1.6 Thesis objectives

This thesis attempts to contribute to the tools and experiments required to answer questions about photosynthesis in Chlamydomonas, with a focus on the CCM and CET.

The first aim of the thesis was to make tenable the cloning and modification of key genes involved in the CCM and CET by overcoming the technical hurdles caused by the complexity of the Chlamydomonas nuclear genome. This was done first by development of a cloning method based on recombineering and second by the use of CRISPR-Cas9 to make specific knock-down deletions of key CET genes, described in Chapters 3 and 4. The results of our recombineering method were published in 2021 and were presented that year at the International Conference on the Cell and Molecular Biology of Chlamydomonas in France, with the protocols and resources now available for the research community.

The second aim was to employ these tools to probe the mechanism of CET by engineering the key protein complexes involved. Recombineering was applied to the STT7 kinase with the goal of restoring state transitions in STT7-deficient cells

using a tagged construct. This may provide a springboard for future attempts to co-purify STT7 with cytb_6f under state II-activating conditions, with the eventual aim of characterising the transient interaction between the two proteins using cryo-electron microscopy. CRISPR-Cas9 was put to use building a mutant strain that expresses FNR tethered to the acceptor side of PSI. Here the aim was to maintain FNR for LET but exclude its ability to bind to cytb_6f. We hypothesised that this modification would reduce the efficiency of CET while maintaining or enhancing LET, perhaps leading to clues as to the FQR mechanism. The phenotypic consequences of tethering FNR to PSI are explored in Chapter 5.

The work in this thesis was undertaken across three laboratories. Many experiments in Chapter 3 were undertaken in the lab of Prof. Luke Mackinder at the University of York in collaboration with his postdoc Dr. Gary Yates. Genome analysis in Chapter 3, all lab experiments in Chapter 4 and many in Chapter 5 were performed at the University of Sheffield under the supervision of Prof. Matt Johnson. The remaining experiments in Chapter 5 were performed during visits to the lab of Prof. Michael Hippler at the University of Münster, Germany, in collaboration with his postdoc, Dr. Felix Buchert.

Chapter 2

Materials and methods

2.1 Materials, buffers and reagents

All materials and reagents were purchased from Sigma/Merck unless otherwise noted.

2.2 Strains and culturing conditions

2.2.1 *Chlamydomonas reinhardtii*

Chlamydomonas cells were cultured in Tris-phosphate (TP) or Tris-acetate-phosphate (TAP) liquid media with revised Hutner's trace elements (Kropat et al. 2011). All Chlamydomonas strains were maintained on TAP + 1.5% agar plates containing 20 μg ml^{-1} paromomycin or 25 μg ml^{-1} hygromycin where appropriate. The work in this thesis makes use of seven Chlamydomonas strains obtained or purchased from different sources:

- CMJ030 (CC-4533, the CLiP library strain) from Prof. Luke Mackinder.
- CLiP library mutants LMJ.RY0402.215132, LMJ.RY0402.236842 and LMJ.RY0402.101504 in the CC-4533 background, with putative insertions in LCIB, STT7 and FNR, respectively.
- The PGR5-deficient mutant *pgr5* and its wild type background strain, t222+, gifts from Prof. Michael Hippler.
- The high-expression strain UVM4, a gift from Prof. Ralph Bock via Prof. Alison Smith.

Additional strains containing mutations of interest were generated by transformation with exogenous DNA and are listed as described in the text. Cell concentration was quantified throughout this thesis using a LUNA-II automated cell counter (Logos Biosystems) and reported in cells per millilitre (cells ml^{-1}).

2.2.2 *Escherichia coli*

The major strains used in this thesis were JM109 (Promega) for standard cloning work and BL21 (New England Biosciences) for expression of plasmid-encoded genes. Chloramphenicol-resistant strains of DH10B containing fragments of the Chlamydomonas nuclear genome in the form of bacterial artificial chromosomes (BACs) were used for recombineering reactions (see Section 2.4.4) and were obtained from the Clemson University Genomics Institute and the Chlamydomonas Resource Center (University of Minnesota, USA). The spectinomycin-resistant strain DB3.1 (Thermo Fisher) was also used for standard cloning work but was specifically required for maintenance and propagation of the *ccdB*-containing plasmids as it expresses the antidote gene *ccdA*.

2.3 Transformation and selection

2.3.1 Chlamydomonas transformation

All three of the Chlamydomonas genomes are transformable, but transformation of the nuclear genome relies primarily on random integration of exogenous DNA. It is an ongoing focus of Chlamydomonas genetics to understand this process and to enhance the efficiency of targeted, recombination-mediated transformation.

In this thesis the Chlamydomonas nuclear genome was transformed by electroporation using a MicroPulser Electroporator (BioRad). Transformations were performed in 4 mm gap cuvettes (Geneflow) at 0.8 kV. A single pulse at 0.8 kV was usually sufficient to produce transformant colonies following recovery. Cells to be transformed were grown in low-medium light in TAP at 20°C, 120 rpm shaking until they reached mid-log phase. For transformations involving genome editing by CRISPR-Cas9, the temperature of the incubator was raised to 33°C for 3 h prior to transformation (see Section 2.3.3). Cells were then harvested at 1,000 x g for 5 min and resuspended to a concentration of 2 x 10^8 cells ml^{-1} in MAX Efficiency™ Transformation Reagent for Algae (Thermo Fisher), followed by a 5 min incubation in a water bath at 16°C. DNA to be transformed was obtained either by PCR amplification or by linearization of plasmid DNA by restriction digestion followed by heat inactivation at 80°C for 5-20 min. 5-10 µl containing 500-1,000 µg linearised DNA was added to 120 µl cells in a cuvette and mixed by gentle flicking and shaking. Cuvettes were dried with a tissue and pulsed at 0.8 kV, then cells were immediately added to 1 ml TAP + 40 mM sucrose in a 1.5 ml tube and stored at low light. Tubes were wrapped in a layer of tissue paper to protect from excess light and placed back in the growth incubator at 120 rpm for 16-20 h at 20°C. For transformations with Cas9, cells were instead recovered at 33°C and 1 µg ml^{-1} cyanocobalamin was added to the recovery solution to improve cell thermal tolerance (Xie et al. 2013).

2.3.2 Extraction of Chlamydomonas genomic DNA

Genomic DNA (gDNA) was extracted from whole cells using two methods. The first quick and 'dirty' method was used to prepare solutions of DNA suitable only for use in a colony PCR reaction, while the second method was used to prepare cleaner samples of DNA for whole genome sequencing or long-term storage. For the quick extraction, a colony or small scraping of Chlamydomonas cells was resuspended in 15 µl of 10 mM EDTA (pH 8) using a toothpick. The solution was incubated at 95-100°C for 10-20 min, then centrifuged at 7,000 x g for 2 min to clear cell debris. The resulting supernatant was stored at 4°C for up to two weeks.

For purification of cleaner gDNA, cells were processed according to the MasterPure™ DNA Purification Kit (Epicentre Technologies/Illumina). In brief, 1-2 x 10^6 cells were resuspended in 300 µl lysis solution containing 240 U Proteinase K (New England Biosciences) and incubated at 65°C for 15 min, then vortexed thoroughly. 10 µg RNase A (Thermo Fisher) was added, and the solution was incubated at 37°C for 30 min. The sample was cooled on ice and 175 µl precipitation reagent was added and mixed by vortexing. Cell debris was pelleted at 15,000 x g, 4°C for 10 min. The supernatant was added to 500 µl propan-2-ol and mixed by inverting, followed by another centrifugation step at 15,000 x g, 4°C for 10-20 min. The isopropanol-containing supernatant was carefully removed, and the DNA pellet washed twice with room-temperature 70% ethanol. The DNA was air-dried to remove all ethanol, followed by resuspension in Tris-EDTA buffer or water, then storage at 4°C or -20°C for several months.

2.3.3 CRISPR-mediated insertion-deletion

CRISPR-mediated mutations were made to the Chlamydomonas nuclear genome using the Alt-R CRISPR system (Integrated DNA Technology). This involves direct electrotransformation of Chlamydomonas cells with ribonucleoprotein complexes (RNPs) consisting of a modified Cas9 enzyme from Saccharomyces pyogenes (SpCas9) in complex with a custom single-guide RNA (sgRNA) with homology to the target locus. sgRNA sequences were designed using the online tools CRISPR-P 2.0 (http://crispr.hzau.edu.cn/CRISPR2/, accessed September 2022) or CRISPR direct (http://crispr.dbcls.jp/, accessed September 2022) and selected based on the fewest estimated off-target cleavage sites. Following transformation, recovery of the cells at 33°C increases the efficiency of SpCas9-induced double-stranded breakage at the target locus. In line with the current understanding of Chlamydomonas DNA repair, the cell's major repair pathway is the native error-prone non-homologous end-joining pathway (NHEJ).

The NHEJ pathway is not only prone to small deletions or insertions but can also incorporate large fragments of exogenous DNA into the breakage site during the repair process. Accordingly, to generate a specific insertion at the targeted locus,

exogenous linear DNA encoding an antibiotic resistance cassette was co-transformed with SpCas9-sgRNA RNPs into Chlamydomonas cells by electroporation. Successful reactions relied on the incorporation of the antibiotic resistance cassette into the breakage site and were selected for by plating the recovering cells on the appropriate antibiotic. Due to the nature of Chlamydomonas nuclear transformations, a large majority of antibiotic-resistant transformant colonies recovered after each reaction contained off-target insertions of the resistance cassette somewhere else in the genome. As such, extensive screening by colony PCR was required to identify colonies that contained a disruption at the intended target locus. Based on the number of colonies screened vs the number of colonies containing a disruption at the intended locus, the efficiency of this method was found to be 0.5-10% depending on the locus.

2.4 Cloning

Custom DNA constructs for all projects described in this thesis were either generated *in vitro* by enzyme-catalysed polymerization, restriction, ligation, or exonuclease reactions using a variety of techniques, generated *in vivo* using recombineering (see Section 2.4.4), or purchased from a repository or company. Custom DNA sequences were obtained in the form of *de novo* double-stranded fragments (gBlocks; Integrated DNA Technologies) where indicated, or by amplification from existing organisms using PCR. Plasmids were propagated, stored in and purified from *E. coli*. Purified plasmids and DNA fragments were stored at 4°C in nuclease-free TE buffer (Tris-EDTA buffer; 10 mM Tris, 1 mM EDTA, pH 8) or DPEC-treated water.

2.4.1 Polymerase chain reaction (PCR)

Fragments of DNA were amplified by PCR using two different polymerases and protocols depending on their downstream uses. To simply check the size of a DNA fragment or to check for the presence/absence of a DNA sequence in a sample (for example to compare between colonies of bacteria or Chlamydomonas) a low-cost polymerase derived from *Thermus aquaticus* was used for colony PCRs. To generate fragments for later cloning experiments, a high-fidelity polymerase was used in order to reduce errors and improve amplification speed and fidelity.

Note that each project in this thesis required unique modifications and troubleshooting to the PCR recipes and protocols given here in accordance with variation in template length, GC-content and predicted primer binding kinetics. Chlamydomonas colonies required a slightly different colony PCR protocol to bacterial colonies. High GC-content Chlamydomonas nuclear genes required different additives to the low GC-content chloroplast genes. To aid in PCR optimization, the NEB Tm Calculator (New England Biosciences) was used to find efficient starting conditions.

2.4.1.1 Colony PCR

For colony PCRs, fragments were amplified using MyTaq™ DNA Polymerase (Bioline). In general, these colony PCRs were performed in 15 µl volumes containing 7.5 µl 2X MyTaq Red Mix, 200 nM of each primer, 1-1.5% DMSO and 0.1-1 ng μl^{-1} template DNA or 0.5 µl cell solution/cell lysate. A generalized colony PCR program is given below:

Temperature	**Duration**	
95°C	5 min	
95°C 60°C 72°C	30 s 30 s 2 min	35x
72°C	4 min	

2.4.1.2 High-fidelity PCR

To amplify fragments with high-fidelity and specificity, Q5® High-Fidelity DNA polymerase (New England Biosciences) or Phusion HotStart II DNA Polymerase (Thermo Fisher) were used. In general, these PCRs were performed in 25 µl volumes containing 12.5 µl 2X Q5 master mix, 200-500 nM of each primer and 0.1-1 ng μl^{-1} template DNA. 1-9% DMSO was added for high GC-content amplicons. For extremely high-GC amplicons or problematic reactions, 1 M betaine was added in combination with 6% DMSO. Both these additives are thought to disrupt the formation of stable secondary structures in high-GC DNA templates. A generalized high-fidelity PCR program is given below:

Temperature	**Duration**	
98°C	30 s	
98°C 66°C 72°C	10 s 30 s 90 s	30x
72°C	2 min	

For difficult reactions, the annealing temperature was reduced stepwise by 0.2°C, e.g., from 66-60°C, to increase the chances of successful amplification in the early cycles of the reaction. For low GC-content templates, the extension temperatures were reduced from 72°C to 65°C and $MgCl_2$ concentration in the reaction mix was altered between 1-5 mM to modulate the melting temperature.

For purification of DNA amplicons directly from the PCR mix, a FastGene Gel/PCR Extraction Kit (Nippon Genetics Europe) was used, and the DNA was eluted in TE buffer.

2.4.2 *In vitro* DNA manipulation

Linear DNA fragments were primarily stitched together using Gibson assembly, restriction-ligation or by overlap-extension PCR. Circular fragments were annealed by ligation prior to transformation into *E. coli*.

2.4.2.1 Gibson assembly

Gibson assembly, described by Daniel Gibson in 2009, involves the stitching together of fragments with complementary ends, first by the action of an exonuclease that removes bases one at a time from the 5' end of the leading strand, creating a single strand overhang (Gibson et al. 2009). Subsequent annealing of the complementary overhangs between fragments at 50°C allows a DNA polymerase to fill up the remaining gaps in the overhang with free nucleotides, and finally a DNA ligase seals the nicks. The NEBuilder® HiFi DNA Assembly (New England Biosciences) Gibson assembly enzyme mix was used for stitching together fragments generated by either PCR or restriction digestion that contained 12-80 bp of complementary bases.

2.4.2.2 Restriction-ligation

Linear fragments were generated using a variety of restriction enzymes (FastDigest; Thermo Fisher). In general, restriction digestion reactions were performed at 37°C for 10-45 min in 20 µl reaction volumes containing 2-3 µl 10X FastDigest buffer, 0.2-2 µg DNA and 1-2 µl per restriction enzyme. Complementary overhangs generated by restriction digestion were ligated using T4 DNA ligase (New England Biosciences). In general, ligation reactions were performed in 10 µl volumes containing 1 µl 10X ligation buffer, 0.1-1 µg combined DNA fragments and 1 µl T4 Ligase. Reactions were incubated at a range of temperatures to increase efficiency: for simple ligations the reaction was incubated at room temperature for 30 min; for more complex ligations, the reaction tube was added to a small beaker of room temperature water and placed in the fridge overnight so that the temperature slowly dropped. 1 µl of 20 mM ATP solution was also added before placing in the fridge to increase ligase activity.

2.4.2.3 Overlap-extension PCR

Overlap-extension PCR involves stitching together two linear, overlapping fragments within a PCR using only a polymerase. The overlapping sections of each fragment anneal and act as their own primers resulting in a PCR amplicon spanning both fragments. For difficult assemblies, Gibson assembly or ligation reactions would

yield only a very low final concentration of successfully annealed products. In these cases, 0.5 µl of Gibson assembly or ligation reaction were often added directly to overlap-extension PCRs along with primers designed to amplify across the intended annealing site, thus increasing the concentration of successfully annealed products.

2.4.2.4 Agarose gel electrophoresis

Linear fragments of DNA were detected and quantified using agarose gel electrophoresis, formed by boiling a solution of 0.8-1% agarose in TAE buffer (40 mM Tris, 20 mM acetic acid, 1 mM EDTA, pH 8). For visualization of DNA fragments sizes, electrophoresis was performed at 90-110 V using a 1% agarose gel and imaged using. Gels were visualized using an Amersham™ Imager 600 (GE Healthcare) set to UV mode. For recovery of DNA from agarose gels, electrophoresis was performed at 75-90 V using a 0.8% agarose gel and DNA bands excised using a scalpel. DNA was extracted using a FastGene Gel/PCR Extraction Kit (Nippon Genetics Europe) and resuspended in TE buffer.

2.4.3 Transformation of *E. coli*

Following annealing of circular DNA products, 4-5 µl of Gibson assembly reaction or ligation reaction was cooled on ice and added gently to 50 µl of competent *E. coli* in a 1.5 ml tube. *E. coli* was transformed either by electroporation or by heat shock at 42°C.

For electrotransformation, cells were mixed with DNA and added to pre-cooled 1 mm gap electroporation cuvettes (Geneflow). Cells were pulsed at 1.8 kV in a MicroPulser Electroporator (BioRad) then immediately added to 750-900 µl room temperature super-optimal broth with catabolite repression (SOC) media (2% tryptone, 0.5% yeast extract, 10 mM NaCl, 10 mM $MgCl_2$, 10 mM $MgSO_4$, 2.5 mM KCl, 20 mM Glucose) followed by out-growth at 37°C for 1-2 h with vigorous shaking.

For heat shock transformation, cells were mixed with DNA and incubated for 5-30 min on ice before heat shock at 42°C for 30-45 s. Cells were immediately recovered for 1-2 min on ice before addition of 750-900 µl room temperature SOC media and out-growth at 37°C for 1-2 h with vigorous shaking.

Unless otherwise noted, out-grown cells were centrifuged at 7,000 x g for 2 min then resuspended by pipetting in 50 µl spent SOC media and spread gently onto pre-warmed LB agar plates containing appropriate antibiotics, then grown at 37°C for 16-20 h.

2.4.4 Recombination-mediated genetic engineering

Recombination-mediated genetic engineering (recombineering) was used in this thesis to clone sequences of Chlamydomonas nuclear DNA that were too long or

repetitive to efficiently amplify by PCR. Recombineering works by coopting the native recombination machinery of *E. coli* to edit or retrieve fragments of DNA from a bacterial artificial chromosome (BAC). Prior to the recombination reaction, a custom DNA construct must be introduced to the cells that contain homologous regions to the target DNA.

Detailed protocols for both batch scale and large scale recombineering of Chlamydomonas nuclear genes is presented in the supplemental data of Emrich-Mills et al. 2021, included as an appendix. Briefly, for each target region, a recombineering cassette was amplified from pLM099 (see Chapter 3, Fig. 3.2 on page 73) by PCR using primers that contained 50 bp arms with homology to the 50 bp immediately upstream and downstream of the target region. Next, a strain of DH10B *E. coli* containing a suitable fragment of the Chlamydomonas nuclear genome in the form of a BAC was identified using a custom python script available as a program called BAC-Searcher (https://github.com/TZEmrichMills/Chlamydomonas_recombineering, accessed September 2022). The strain was first made recombination competent by transformation with the plasmid pRed, which encodes the viral recombinogenic *Red*αβγ genes and *recA* under the control of an arabinose inducible promoter. The pRed containing transformants were grown overnight to recover, then induced by addition of L-arabinose to a total of 12 mM. Cells were then transformed with the purified custom recombineering cassette and recovered in SOC for 1.5 h to allow for the recombineering reaction to take place before plating to produce single colonies after 16-18 h incubation at 37°C.

To screen for correct recombineering, three or more colonies were grown and their plasmids purified by miniprepping. Plasmids were digested using both single and double-cutting enzymes and analysed by agarose gel electrophoresis. Plasmids that produced the correct banding patterns after digestion were sequence by Sanger sequencing to confirm correct assembly and absence of truncated or mutated sequences.

2.4.5 Purification of plasmid DNA from *E. coli*

Plasmids were purified from *E. coli* using either a FastGene Plasmid Mini Kit (Nippon Genetics Europe) or a PureLink™ HiPure Plasmid Maxiprep Kit (Invitrogen) depending on the desired quantity of plasmid. For routine minipreps, 3-10 ml LB was inoculated with a colony of *E. coli* and incubated for 16-20 h at 37°C with vigorous shaking. The recovered plasmid was eluted in TE buffer and stored at 4°C. For maxipreps, 100-200 ml cells were harvested and processed depending on the expected plasmid copy number per cell. Plasmid concentration was measured using a Nanodrop spectrophotometer.

2.4.6 DNA sequence analysis

Plasmids and PCR amplicons were analysed by Sanger sequencing using the SupremeRun service from Eurofins Genomics and custom primers.

2.5 Isolation of proteins from thylakoid membranes

Thylakoid membranes were prepared under different conditions depending on downstream experiments. However, a set of four basic steps applies to all thylakoid preparation described in this thesis: (1) Resuspension of fresh or frozen cells to a high concentration in buffer containing 1-2 cOmplete™ Protease Inhibitor Cocktail tablets per 50 ml (Roche/Merck), (2) mechanical cell breakage, (3) a slow centrifugation step to pellet and discard unbroken cells and free starch, (3) a fast centrifugation step to pellet chloroplasts and thylakoids and discard soluble proteins, (4) resuspension of the pelleted membranes in buffer and quantification by measuring the concentration of chlorophyll in the sample.

Cell breakage was performed by sonication or by French press. Breakage was performed in a buffer that usually included $MgCl_2$, which keeps thylakoids tightly appressed, as well as a cushioning agent such as sucrose or glycerol. NaCl was always present for the purpose of screening surface charges on thylakoid proteins to prevent unwanted reactions during preparation. The buffering agents used were 10-50 mM of HEPES, Tris, Bis-Tris or sodium phosphate depending on the desired pH.

Proteins were purified from Chlamydomonas cells using two main methods: (1) native proteins were purified using sucrose gradient centrifugation, anion exchange chromatography and/or size exclusion chromatograpy; (2) affinity-tagged proteins were purified using affinity chromatography, including immobilized metal affinity chromatography (IMAC), followed by size exclusion chromatography. Between applications of these techniques, protein-containing solutions were concentrated by centrifugation using membranes with various molecular weight cut-offs (MWCO) depending on the target protein size. Proteins purified from Chlamydomonas cells in this thesis were thylakoid membrane proteins and first required the preparation of thylakoid membranes, followed by extraction from their native membrane environment using detergents or copolymers according to the methods detailed in the sections below. Proteins were protected from the activity of endogenous proteases during purification by addition of protease inhibiting chemicals. Serine, cysteine and metalloproteases were inhibited by the addition of cOmplete™ Protease Inhibitor Cocktail tablets (Roche). Where indicated, 6-aminocaproic acid was also added, which outcompetes lysine residues at the binding sites of lysine proteases.

2.5.1 Measurement of chlorophyll concentration

The combined concentration of chlorophyll A and B in a sample, [Chls], was calculated using two methods according to Porra et al. 1989.

Firstly, [Chls] was measured in 80% acetate using absorption measurements at 645, 663 and 750 nm. 4 µl of sample material was added to 2 ml of freshly made 80% acetone solution, vortexed vigorously for 10 s then centrifuged at 15,000 x g for 2 min to pellet and discard the cell debris. The supernatant containing pigments was added to a quartz cuvette and absorbance was measured using a Cary60 UV-Vis spectrophotometer (Agilent Technologies). [Chls] was calculated according to the equation:

$$[\mathrm{C}hls] = 8.88\mathrm{A}_{645} + 3.67\mathrm{A}_{663} - 12.55\mathrm{A}_{750} \tag{2.1}$$

Alternatively, [Chls] was measured in 80% methanol using absorption at 652, 665 and 750 nm. Cells were diluted 50x in a 200 µl volume of water, then added to 800 µl pure methanol, vortexed as above, then centrifuged for 2 min. The absorption of the supernatant was measured at the above wavelengths and the [Chls] according to the following equation (which includes a correction for the 50x dilution):

$$[\mathrm{C}hls] = 1.106\mathrm{A}_{652} + 0.122\mathrm{A}_{665} - 1.228\mathrm{A}_{750} \tag{2.2}$$

2.5.2 Purification of PSI from Chlamydomonas

Native photosystem I was purified from Chlamydomonas by anion exchange chromatography followed by sucrose density centrifugation and size exclusion chromatography. The method used was derived from a protocol by Ramesh and Webber 2004. Briefly, cells were grown to mid-log phase in 2 L TAP under ~50 µE m^{-2} s^{-1} white fluorescent light, agitated with a magnetic stir bar and bubbled with filtered air. Cells were collected at 4,000 x g for 5 min at room temperature and resuspended in HEPES buffer (50 mM HEPES pH 7.2, 5 mM $CaCl_2$, 5 mM $MgCl_2$, 20% glycerol, 1 tablet of cOmplete™ Protease Inhibitor Cocktail tablet (Roche) per 50 ml). Cells were broken with 2 passes through a French press at 9,000 psi, and unbroken cells were pelleted at 3,000 x g. Membranes were pelleted from the cell lysate at 100,000 x g for 20 min using a TLA-100.2 ultracentrifuge rotor (Beckman). The pellet was resuspended gently in 30 ml HEPES buffer using a fine-hair paintbrush and centrifuged a second time. The pellet was resuspended in 2 ml HEPES buffer and diluted to 0.5 mg ml^{-1} [Chls]. Membranes were solubilized with 0.6% βDDM for 15 min at 4°C with gentle shaking. Unsolubilised material was cleared at 125,000 x g for 20 min in a TLA-100.2 rotor. Solubilised membranes were applied to a custom 150 ml DEAE Sepharose column pre-equilibrated with HEPES buffer + 0.02% βDDM and pumped through at approx. 6 ml min^{-1}. The

column was washed with 50 ml HEPES buffer + 0.02% βDDM + 7 mM extra $CaCl_2$, followed by another wash of the same volume with 5 mM added $MgSO_4$. The column was eluted by application of a 5-25 mM $MgSO_4$ wash gradient in HEPES buffer + 0.02% βDDM + 7 mM $CaCl_2$. Fractions were collected in 3 ml aliquots and analysed using a Cary60 UV-Vis spectrophotometer (Agilent Technologies). Fractions with a maximum absorbance between 675 and 677 nm were pooled, diluted three-fold with LOAD buffer (25 mM HEPES pH 7.2, 12 mM $CaCl_2$, 5 mM $MgCl_2$) + 0.02% βDDM and concentrated down to approx. 500 µl using a 100 kDa MWCO spin concentrator at 3,000 x g.

The concentrated sample was loaded onto a 10-35% continuous sucrose gradient in LOAD buffer + 0.02% βDDM. Gradients were centrifuged in a SW41 ultracentrifuge rotor (Beckman) at 207,500 x g for 26 h. The sample separated into distinct bands representing monomeric and trimeric LHCII and PSI, with PSI running the furthest. Band composition was confirmed using a Cary60 UV-Vis spectrophotometer (Agilent Technologies), with the PSI-enriched band showing a distinct shift of the red maximum peak from 673 to 677 nm relative to the LHCII-containing bands. The PSI-containing band was removed and diluted ten-fold in LOAD buffer + 0.02% βDDM and concentrated by spin concentration as before. Concentrated samples were aliquoted and supplemented with 5% ethylene glycol before flash freezing and storage at -80°C.

2.5.3 Purification of STT7 from Chlamydomonas

STT7 was purified from Chlamydomonas cells by affinity chromatography using a C-terminal Strep-tag® II tag. Cells from strain cTE038 (see Chapter 3, Fig. 3.5 on page 82) were grown in 8 L TAP media supplemented with 2 mg ml^{-1} paromomycin and 2.5 mg ml^{-1} hygromycin under ~100 µE m^{-2} s^{-1} fluorescent light, bubbled with filtered air. Once at mid-log phase, cells were pelleted and resuspended in TP for the final 16 h of growth, and extra paromomycin and hygromycin was added to boost expression. To harvest, cells were pelleted at 4,000 x g for 5 min and resuspended in BREAK buffer (40 mM Tris pH 7.4, 5 mM $MgCl_2$, 50 mM NaCl, 10 mM NaF, 5% glycerol, 5 mg ml^{-1} 6-aminocaproic acid and one cOmplete™ Protease Inhibitor Cocktail tablet (Roche) per 50 ml). Cells were broken with a single pass through a French press at 8,000 psi, then unbroken cells were pelleted for 15 min at 3,000 x g. Unbroken cells were pooled and re-broken to increase yield. Lysate was centrifuged at 75,600 x g for 45 min at 4°C to pellet thylakoid membranes. The pellet was resuspended in BREAK buffer with a fine-hair paintbrush and homogenised by douncing, followed by a second spin at 75,600 x g for 30 min. These pellets were resuspended and homogenised in SOL buffer (20 mM Tris pH 8, 200 mM NaCl). The optical density of the homogenised membrane solution was measured, and the solution was diluted using 10% styrene maleic acid (SMA) and SOL buffer to a final

concentration of OD_{680} = 17, 4.5% SMA, 300 ml volume. The solution was brought to room temperature to facilitate solubilisation by SMA, then stirred in the dark for 1 h. Unsolubilised membranes were then pelleted by centrifugation at 60,000 x g for 30 min, 4°C. The supernatant containing solubilised membranes was passed once through a 0.22 μm filter using a syringe, then applied to several 100 kDa MWCO spin concentrators and centrifuged at 3000 x g for 2 h until the total sample volume had decreased to 200 ml. All solubilised membranes were then filtered once through a custom gel filtration column containing Sephadex® G-25 resin (Sigma) to remove free SMA, which can decrease binding efficiency of StrepTactin resin in the next step. Filtered membranes were applied to a 5 ml StrepTactin 4Flow (IBA Lifesciences) gravity column. The column was washed with 250 ml WASH buffer (20 mM Tris pH 8, 300 mM NaCl), then resuspended by gentle pipetting and washed with a further 250 ml WASH buffer. Finally, the column resin was resuspended in 10 ml SOL buffer + 50 mM biotin (pH 8) and incubated at room temperature for 20 min. The eluate was extracted and stored at 4°C, and the column was resuspended in a further 10 ml for a second elution, this time overnight at 4°C. The next day, both eluates were combined and concentrated in a 10 kDa MWCO spin concentrator until the volume had been reduced to ~1 ml. The resulting sample was applied to SDS-PAGE as described in Section 2.6.1.

2.6 Polyacrylamide gel electrophoresis (PAGE) and immunoblotting

2.6.1 Sodium dodecyl-sulphate (SDS-PAGE)

SDS-PAGE was used to separate samples of solubilized protein by peptide molecular weight.

Purified protein samples were prepared for SDS-PAGE by solubilisation in Laemmli sample buffer (60 mM Tris pH 6.8, 2% SDS, 10% glycerol, 0.01% w/v bromophenol blue) followed by heating at 100°C for 5 min. For samples enriched in membrane proteins such as PSI, the heating step was omitted to avoid heat-induced aggregation, and samples were instead sonicated for 20-60 s using a FB 11021 Ultrasonic Bath (Fisherbrand) and incubated at room temperature for 30 min. Samples were centrifuged at 15,000 x g for 2 min and the supernatant loaded onto an SDS gel.

Whole-cell samples of Chlamydomonas were solubilised for SDS-PAGE as follows. Cell pellets were solubilised in TPEB (Total protein extraction buffer; 180 mM Tris pH 6.8, 6% SDS, 35% glycerol, 100 mM β-mercaptoethanol) followed by 20-60 s sonication using a FB 11021 Ultrasonic Bath (Fisherbrand), 30 s vortexing at top speed, then heating and agitation for 5-10 min at 1,700 rpm in an Eppendorf ThermoMixer® C at 80-90°C. Lysed cell samples were centrifuged at 21,000 x g

for 10 min at 4°C and the supernatant (containing soluble proteins) mixed with NuPAGE™ LDS Sample Buffer (Thermo Fisher) or 2 x Laemmli buffer before loading onto an SDS gel. For comparison of peptide presence/absence between strains, samples were normalised by cell count before processing.

SDS-PAGE was performed using the NuPAGE™ system (Thermo Fisher) using different gel formulations and running buffers depending on the size of the peptides of interest. For whole-cell peptides and smaller peptides (< 60 kDa) of interest, 12% Bis-Tris 1 mm Mini Protein Gels were run in MES SDS Running Buffer at 120-180 V for 1-1.5 h. To distinguish larger peptides of interest (> 60 kDa) such as STT7, 7% Tris-Acetate 1 mm Mini Protein Gels were run in SDS Tris-Acetate Running Buffer at 110-140 V for 1-1.5 h. All gels were purchases pre-cast from Thermo Fisher.

2.6.2 Western blotting

Proteins and peptides were transferred from BN or SDS gels onto nitrocellulose or polyvinylidene difluoride (PVDF) membranes by semi-dry transfer using a Power Blotter System (Invitrogen) in Power Blotter 1-Step™ transfer buffer. For transfer from BN gels, transfer was performed twice per gel at 1.3 A for 10 min onto nitrocellulose membranes; the first transfer removes most of the G250 dye, the second captures most of the protein from the gel. For transfer of large peptides from SDS gels, transfer was performed at 1.3 A for 7-10 min. For small (< 15 kDa) peptides, the transfer buffer was supplemented with 2 mM $ZnCl_2$ (Maillet et al. 2001) and transfer was performed at 1.3 A for 3-5 min onto 0.2 μm pore-size PVDF membranes (Invitrogen).

Following transfer, membranes were washed with TBS (Tris buffered saline; 50 mM Tris pH 7.6, 150 mM NaCl) and blocked in TBS + 2 mg ml^{-1} TWEEN-20 + 5% milk powder for 1 h at room temperature with gentle agitation. For small peptides, 2% milk powder was used instead to prevent loss of peptides from membrane. Blocked membranes were washed with TBS and incubated with gentle agitation in TBS + 0.5 mg ml^{-1} TWEEN-20 containing the appropriate primary antibody, either at room temperature for 1-4 h or at 4°C for 16-20 h depending on the peptide sample and sensitivity of the antibody used.

Following incubation in primary antibody, membranes were washed three times for 5 min each with TBS + 0.5 mg ml^{-1} TWEEN-20, then incubated with the appropriate secondary antibody in TBS + 0.5 mg ml^{-1} TWEEN-20 for 1 h at room temperature. Secondary antibodies used were either conjugated to horse radish peroxidase (HRP) or to a fluorescent dye (goat anti-rabbit IgG cross-adsorbed secondary antibody, Alexa Fluor™ 488; Thermo Fisher). Membranes incubated in HRP-conjugated secondary antibodies were washed three times for 5 min each with TBS + 0.5 mg ml^{-1} TWEEN-20 before soaking for 5 min in Westar Sun chemiluminescent substrate solution (Cyanagen). Membranes were visualized using

an Amersham™ Imager 600 (GE Healthcare) in fluorescence mode if using the Alexa Fluor-conjugated secondary antibody or in chemiluminescence mode if using HRP-conjugated secondary antibodies.

2.7 Spectroscopic analysis

2.7.1 Low-temperature fluorescence emission spectroscopy (77 K)

Whole cells or samples of purified thylakoids were analysed by low temperature fluorescence emission spectroscopy at 77 K. Briefly, samples were prepared by dilution in TP to ~5 mg ml^{-1} Chls, then added to a 1 cm polymethyl methacrylate cuvette or a PCR tube. Samples were plunged into liquid nitrogen and added to a custom glass vacuum chamber placed in the beam of a FluoroLog FL3-22 spectrofluorimeter (Jobin Yvon). Fluorescence emission spectra were recorded by excitation at 435 nm and monitoring between 600-800 nm. Spectra were normalised to the peak at ~680 nm, associated with fluorescence from PSII, while the broader peak between ~700-740 nm is associated with PSI fluorescence. To assess a strain's ability to perform state transitions, cells were prepared in oxic or anoxic conditions prior to freezing, which triggers state I or state II, respectively (Buchert et al. 2020). For oxic conditions, cells were equalised in TP media to a concentration equivalent to 5 μg ml^{-1} Chls (combined chlorophyll *a* and *b*), then incubated in the dark for 20 min with vigorous aeration. For anoxic conditions, cells of the same concentration were mixed with oxygen scavengers: 50 mM glucose, 30 U ml^{-1} catalase and 10 U ml^{-1} glucose oxidase. Cells were then overlaid with 500 μl mineral oil to prevent gas exchange, and cells were incubated in the dark for 20 min without shaking to deplete oxygen.

2.7.2 Plate reader fluorescence emission measurements

Chlorophyll and Venus fluorescence emissions were monitored using a Clariostar plate reader (BMG Labtech) to compare fluorescence between high- and low-CO_2 grown cells. Cells were grown in TP bubbled with filtered air containing the appropriate concentration of CO_2 until they reached mid-log phase. 180 μl cells were added to wells of a 96-well plate, with four replicates per strain. Fluorescence emissions were measured for chlorophyll by excitation passed through a 625 ± 17 nm filter, while emission was detected via a 692 ± 25 nm filter. Venus emissions were detected by excitation passed through a 504 ± 5 nm filter, while emission was detected via a 540 ± 12 nm filter. Venus fluorescence was normalised by chlorophyll fluorescence for each well, then the Venus background fluorescence detected from wild type cells was subtracted. Normalised values under low CO_2 for replicates were averaged together then divided by the average under high CO_2, yielding values used

for the estimation of the change in fluorescence between high and low CO_2 for each strain tested.

2.7.3 Time-resolved chlorophyll fluorescence spectroscopy

Chlorophyll fluorescence emission spectroscopy is an important and widely used method of monitoring the metabolic state of plants and algae. Chlorophyll fluorescence was monitored by pulse amplitude modulated fluorescence in an IMAGING-PAM fluorescence system (Walz, Germany) and analysed using ImagingWin software. Cells were prepared for measurements using the IMAGING-PAM in either oxic or anoxic conditions. For oxic conditions, cells were equalised in TP media to a concentration equivalent to 5 μg ml^{-1} Chls (combined chlorophyll *a* and *b*), then 500 μl was added to wells of a 48-well microplate and incubated in the dark for 20 min with vigorous aeration. For anoxic conditions, 500 μl cells in a microplate were mixed with oxygen scavengers as in Section 2.7.1. Wells were overlaid with 300 μl mineral oil to prevent gas exchange, and cells were incubated in the dark for 20 min (no shaking) to deplete oxygen (Buchert et al. 2020). To assess the fluorescence parameters 1-qP and NPQ, saturating pulse-induced fluorescence was processed using ImagingWin software and values derived according to the following equations:

$$qP = \frac{F'_m - F_t}{F'_m - F'_o} \tag{2.3}$$

$$\mathrm{NPQ} = \frac{F^o_m - F'_m}{F'_m} \tag{2.4}$$

In Chlamydomonas, Fo' cannot be obtained using far red light, so it is instead estimated using the following equation (Oxborough and Baker 1997):

$$F'_o = \frac{F_o}{(F_v/F^o_m) + (F_o/F^o_m)} \tag{2.5}$$

2.7.4 Measurements of the electrochromic shift

Under illumination, the net voltage resulting from proton translocation across the thylakoid membranes modifies the microenvironment of nearby photosynthetic pigments. This modification results in a shift in absorption spectra, measurable as an increase in the absorption signal at ~515-520 nm relative to ~545-550 nm. Absorption was measured either using P515/535 module of the DUAL-PAM-100 system (Walz, Germany) as the 515 minus 550 nm difference signal or using a JTS-150 spectrophotometer (BioLogic) as the 520 minus 546 nm difference signal. The JTS-150 system belonged to Prof. Michael Hippler and was used as part of a collaboration with his lab at the University of Münster. Absorption changes were

normalised between runs and strains to the ECS ΔI/I signal induced by a short, saturating pulse in the presence of the inhibitors DCMU and HA (to deactivate PSII). When using the JTS-150, the saturating flash used for normalisation was provided by a separate laser module containing a DCM exciton laser dye pumped by a frequency doubled Nd:YAG laser (Minilite II, Continuum) (Buchert et al. 2020).

To monitor the magnitude of the ECS in Chlamydomonas, cells were grown in TAP or TP at ~60-70 μE m^{-2} s^{-1} with gentle bubbling under either constant light or under day/night cycling until they reached early- to mid-log phase. Cells were harvested by centrifugation at 4,000 x g for 5 min, then resuspended in a 4 ml cuvette in 1.5-2 ml of TP + 15-20% ficoll (to prevent sedimentation) to ~20 μg ml^{-1} Chls. Anoxia was induced using oxygen scavengers: 50 mM glucose, 30 U ml^{-1} catalase and 10 U ml^{-1} glucose oxidase, then overlaid with mineral oil to prevent gas exchange. The cuvette was then placed in the path of the detector module, with transmitted light filtered through appropriate interference filters with FWHM of 20 nm (Buchert et al. 2020). Measurements of proton motive force (*pmf*), proton flux ($v_H{}^+$) and proton conductivity ($g_H{}^+$) were made as described by Kramer et al. (2004) using the slope and extent of the relaxation of the ECS signal during a dark interval. Dark interval kinetics were modelled with a biexponential function using the SSbiexp package in R. The time constant of the slope for dark interval signals was estimated using the reported values from the SSbiexp function, with the overall time constant (half-life, or τ) being composed of the two τ values for the individual curves that make up the biexponential model. $g_H{}^+$ was calculated as $1/\tau \times 1{,}000$. *Pmf*, $g_H{}^+$ and $v_H{}^+$ are related according to the following equation (Kramer et al. 2004):

$$v_H^+ = pmf \times g_H^+ \tag{2.6}$$

Measurements of electron transfer rate were made as described in Nawrocki et al. 2019a. Briefly, the JTS-150 was programmed to provide 1 s high light, during which multiple 5 ms dark intervals were inserted. Recordings of the ECS during these dark intervals were processed to calculate the gradient of the signal during the latter 4 ms of each interval. The first millisecond is excluded to avoid confounding factors introduced by charge recombination. Simultaneously, the gradient immediately prior to each dark interval was recorded as well. Together, the gradient in the light minus the dark is an estimation for the electron transfer rate at each time point during the 1 s of high light. A timepoint is defined as the moment when each dark interval begins. For all measurements of electron transfer rate in this thesis, cells were adapted for 5 min in the cuvette prior to measurement to a light regime of ~21 μE m^{-2} s^{-1} background light interspersed every 10 s with 1 s high light (~550 μE m^{-2} s^{-1}).

2.7.5 Measurements of P700 oxidation

The net appearance of the PSI cation, $P700^+$, was measured by monitoring the absorption intensity over time at 705 nm minus 740 nm using a JTS-150 spectrophotometer (BioLogic, France) with appropriate interference filters (705 nm, FWHM 10 nm; 740 nm, FWHM 15 nm) (Buchert et al. 2020). As above, absorption changes were normalised between runs and strains to the ECS ΔI/I signal induced by a short, saturating pulse in the presence of the inhibitors DCMU and HA (to deactivate PSII). This required switching back to ECS mode on the JTS-150 after P700 measurements were complete (see Section 2.7.4 on page 55).

Cells were grown in TP to mid-log phase as before and resuspended to 20 μg ml^{-1} Chls in fresh TP + 20% Ficoll. Anoxia was induced using oxygen scavengers: 50 mM glucose, 30 U ml^{-1} catalase and 10 U ml^{-1} glucose oxidase, then overlaid with mineral oil to prevent gas exchange. Cells were dark adapted for 20 min prior to measurement. For all measurements of P700 oxidation in this thesis, cells were adapted for 5-10 min in the cuvette prior to measurement to a light regime that matched the eventual measurement light regime, e.g., 10 cycles of light protocol, with the 11th cycle being used for measurement. These adaptation periods help to stabilise and regularise any light induced changes to the P700 signal, decreasing variability between repeats. At the end of a set of measurements, a saturating pulse was administered then the actinic light removed. In the presence of the inhibitors DCMU and HA, this has the effect of first completely oxidising the P700 present in the sample, then allowing re-reduction in the dark. These two extremes of P700 oxidation were used to calculate acceptor and donor side limitation as described previously (Alric et al. 2010, Buchert et al. 2020), illustrated in Fig. 5.9 in Chapter 5 on page 131.

2.8 Confocal fluorescence microscopy

All fluorescence microscopy in this thesis was performed at the University of York in collaboration with Prof. Luke Mackinder, and the results were published in Emrich-Mills et al. 2021. The following section is adapted from that publication.

Cells were prepared for microscopy by growth in TP to early-log phase, then added to a clear-bottom μ-slide 8-well (Ibidi, UK) and fixed in place using a drop of 34°C TP + 1% low-melting-point agarose. Images were acquired using a Zeiss LSM880 confocal microscope on an Axio Observer Z1 invert, equipped with a 63× 1.40 NA oil planapochromat lens. Images were analyzed using ZEN 2.1 software (Zeiss, USA) and FIJI. Excitation and emission filter settings for chlorophyll fluorescence were 561 nm excitation, 665–705 nm emission. For other fluorophores, settings were as follows: 514 nm excitation and 525–550 nm emission for Venus and mNeon-Green, 561 nm excitation and 580–600 nm emission for mScarlet-i.

2.9 Proteomics

Mass spectrometry was performed largely as described in Flannery et al. 2021 with minor modifications.

2.9.1 Proteolytic digestion

Thylakoid membranes (50 µg protein, Bio-Rad DC assay) were solubilised in 15 µl 1% (w/v) sodium laurate (SL), 100 mM triethylammonium bicarbonate (TEAB) pH 8.5 by sonication for 5 min in ice cold water (Lin et al. 2013). Complete solubilisation was verified by the absence of a pellet after centrifugation at 15,700 x g for 2 min. Proteins were reduced by the addition of 1.5 µl 100 mM tris(2-carboxyethyl) phosphine-HCl and incubation at 37°C for 30 min. Cys thiol groups were derivatised by the addition of 1.5 µl 200 mM methyl methanethiosulphonate in isopropanol and incubation at room temperature for 10 min. Samples were adjusted to 50 µl with 1% (w/v) SL and 100 mM TEAB pH 8.5 and proteolytic digestion was carried out by treatment with 2 µg pre-mixed trypsin/endoproteinase Lys-C (Promega) for 3 h at 37°C. Extraction of SL was performed as previously described (Lin et al. 2013) by adding an equal volume of ethyl acetate and acidification with 10 µl 10% (v/v) trifluoroacetic acid (TFA). The samples were vortexed for 1 min and then centrifuged at 15,700 x g for 5 min to accelerate phase separation. The peptide-containing lower phase was isolated, dried by vacuum centrifugation and dissolved in 50 µl 0.5% (v/v) TFA and 3% (v/v) acetonitrile before desalting with C18 spin columns (Thermo Fisher Scientific) according to the manufacturer's protocol. The peptides were again dried by vacuum centrifugation and stored at -20°C.

2.9.2 Mass spectrometry and protein identification

For analysis by nano-LC-MS/MS, peptides were dissolved in 0.5% (v/v) TFA and 3% (v/v) acetonitrile, and 400 ng of each sample was analysed in triplicate in randomised order. Peptides were resolved on an EASY-Spray PepMap RSLC C18 column (Thermo Fisher Scientific, 50 cm x 75 µm ID, 2 µm particle diameter at 40°C) with the following gradient profile delivered at 300 nl min by a Dionex RSLCnano chromatography system (Thermo Fisher Scientific): 97% solvent A (0.1% (v/v) formic acid in water) to 10% solvent B (0.08% (v/v) formic acid in 80% (v/v) acetonitrile) over 5 min, then 10% to 50% solvent B over 3 h. The mass spectrometer was a Q Exactive HF hybrid quadrupole-Orbitrap system (Thermo Fisher Scientific) programmed for data-dependent acquisition with profile full MS scans at 120,000 resolution and a maximum of ten centroid product ion scans at 30,000 resolution per cycle. Protein identification and label-free quantification were performed by searching the MS data files against the *Chlamydomonas reinhardtii* reference proteome database (www.uniprot.org/proteomes/UP000006906, downloaded November 2021) using the

MaxLFQ algorithm embedded within FragPipe (v. 16.0) (Yu et al. 2021). Fixed and variable modifications were set to methanethio-Cys and Met sulphoxide respectively, with the 'match-between-runs' option also selected. All other parameters were as per default.

2.10 Bioinformatics used for genome analysis

Genome analysis results presented in this thesis in Chapter 3 were published in Emrich-Mills et al. 2021. The following section is adapted from that publication. All analysis was performed using Excel, python and R.

2.10.1 Data sources

All Chlamydomonas nuclear gene sequences were extracted from the CC-503 reference genome database Phytozome (version 5.5; phytozome-next.jgi.doe.gov/, accessed September 2022). Gene sequences for Arabidopsis (TAIR10 assembly) and *Triticum aestivum* (bread wheat; International Wheat Genome Consortium) were all obtained from EnsemblPlants BioMart (http://plants.ensembl.org/info/data/biomart/index.html, accessed September 2022). Sequence data for *Saccharomyces cerevisiae* (baker's yeast; S288C reference genome, 2015 release) were obtained from the Saccharomyces Genome Database (https://www.yeastgenome.org/, accessed September 2022). Gene sequences for yeast genes were appended to include all annotated UTRs and introns, resulting in more comparable database to those of Chlamydomonas, Arabidopsis and wheat.

2.10.2 Analysis of gene complexity

Repetitive genes were assessed using the NCBI WindowMasker program (Morgulis et al. 2006), which detects a broad range of repeat types and outputs a single value, termed global repeats. Default WindowMasker settings were used throughout, though options for use of the DUST and WinMask modules were selected; DUST detects shorter repetitive regions including tandem and inverted repeats, while WinMask detects families of longer repetitive regions that occur throughout the genome. Information for gene length, transcripts and the presence of UTRs were obtained from Phytozome Biomart. Data on the presence of upstream open reading frames (uORFs) was obtained from the results of a BLASTP analysis performed by (Cross 2016) and adapted to provide the per-gene values using python. Data on the presence of secondary structures in ssDNA were generated using RNAfold (Lorenz et al. 2011, downloaded from https://github.com/ViennaRNA/ViennaRNA April 2020). Default settings were used for all runs.

2.10.3 Analysis of primers

Primer suitability of a genome-wide set of ATG-Stop primers (that bind at the start codon and stop codon of each gene) was assessed using the check_primers algorithm from Primer3 (Rozen and Skaletsky 2000) using the default settings for Primer3Plus (Untergasser et al. 2007). Minor modifications were made to the settings to remove Tm constraints. Output was analysed using python. The majority of rejected primers produced one of the following three reasons for rejection: (1) "high end complementarity" for primer pairs, which implies a high likelihood that the 3'-ends of the forward and reverse primers will anneal, enabling amplification of a short, heterogeneous primer-dimer (cross-dimer); (2) "high end complementarity" for single primers, which implies a high likelihood that a primer's 3'-end will bind to that of another identical copy, self-priming to form a homogenous primer-dimer (self-dimer); and (3) "high any complementarity" for single primers, which implies a high likelihood of self-annealing without necessarily self-priming, relevant to both the inter-molecular annealing of identical copies and to instances of hairpin formation resulting from intra-molecular annealing. Primers rejected for these three reasons in Fig. 3.1C in Chapter 3 on page 69 are labelled as cross-dimers, self-dimers, and hairpins, respectively.

2.10.4 Analysis of gene ontology

Enrichment analysis was used to help make sense of the proteomics results presented in Fig. 4.7 in Chapter 4 on page 102. This was done by subjecting protein IDs from Uniprot to a PANTHER over-representation test using PANTHER v17.0 via the Gene Ontology Resource (http://geneontology.org/, accessed Sep. 2022). Results were identified and grouped according to the predicted biological process contributed to by each protein. The Bonferroni correction for multiple testing options was selected, otherwise all settings were as default.

Chapter 3

Development of a recombineering method for Chlamydomonas

3.1 Summary

The development of modern PCR-based and gene synthesis-based cloning methods have broadened the molecular toolbox for Chlamydomonas. However, PCR-based methods still suffer from problems, exacerbated by repetitive sequences and very long templates. Recombination-mediated cloning is a useful alternative to PCR and functions by homologous recombination to produce precise edits in a gene of interest, for example introducing a fluorescent tag. Recombineering can circumvent PCR-specific problems when cloning very long, complex genes with their native regulatory features intact. We set out to quantify the sequence features in the Chlamydomonas nuclear genome that pose difficulties to PCR-based cloning, including gene size and frequency of repetitive sequences. We addressed this problem by adaptation of a recombineering method from *C. elegans* for Chlamydomonas, which we tested on ~200 genes encoding components of the CCM. Our method is far more tolerant of long and complex sequences, with no effect of gene size found on our success rates. A small subset of genes tested for localisation and expression show native-like regulation, possibly due to our inclusion of the native promoter region in recombineered genes. Two recombineered genes were used to rescue mutants, including the chloroplast kinase STT7, which restored state transitions in a mutant and can be purified from Chlamydomonas.

The work presented in this chapter was started in the lab of Prof. Luke Mackinder at the University of York and completed at the University of Sheffield. The work was published in April 2021 in The Plant Cell (Emrich-Mills et al. 2021) under the terms of a Creative Commons CC BY license, which permits its reproduction here. The publication is accessible using the digital object identifier 10.1093/plcell/koab024 and

is included as an appendix at the end of this thesis. Co-author contributions are stated explicitly in the figure legends.

3.2 Background

Our ever-expanding understanding of chloroplast bioenergetics has inspired research into the potential for improving crop yields by introducing a pyrenoid-based CCM into vascular plants (Mackinder 2018). If correctly functional, this would help limit the loss of fixed carbon and nitrogen by photorespiration by concentrating CO_2 in the vicinity of tightly packed Rubisco, increasing its carboxylation efficiency. Modelling studies suggest that expression of a functional CCM in major food crops could result in a 60% increase in carbon fixation (Long et al. 2015). Realising this will likely require researchers to identify all the protein components of a CCM and understand how they behave and interact, a task usually reliant on reverse genetic approaches, i.e., localisation of proteins and complementation of mutants to help understand how genotype links to phenotype. One major problem faced by researchers attempting to understand the Chlamydomonas CCM arises from difficulties in reliably manipulating the nuclear genome of Chlamydomonas, which houses a large majority of the genes involved in chloroplast bioenergetics. This is indicative of a wider problem in Chlamydomonas genetics: the Chlamydomonas nuclear genome is replete with sequence features that make cloning difficult; a high GC-content, high prevalence of repetitive regions and long, intron-rich genes make for difficult cloning of certain targets by conventional, PCR-based methods (Merchant et al. 2007).

3.2.1 Modern methods for cloning nuclear genes

Optimised protocols are available for long-range or GC-rich PCRs using modern high-fidelity polymerases (e.g., Q5® or Phusion®, NEB), with an advertised maximum amplicon length of 30 kbp. In practice, though, amplification from complex templates like Chlamydomonas gDNA can prove problematic for targets > 5 kbp, requiring extensive optimisation for each target. PCR from gDNA can be avoided in some cases by whole or partial gene synthesis and reassembly using modern cloning strategies such as Golden Gate assembly. The release of a Modular Cloning kit for Chlamydomonas in 2018 outfitted researchers with a wide selection of codon-optimised genetic elements in a useful cloning framework, enabling the rapid construction of standardised transgene libraries (Crozet et al. 2018). Transgenes for MoClo must be devoid of type II restriction sites to enable Golden Gate assembly, so coding regions are usually synthesised *de novo*. This often necessitates extensive modification to the nucleotide sequence, both to decrease total length by removing native introns or to modify repetitive regions and to lower GC content

by codon optimisation to enable synthesis. However, inclusion of introns has been recognised as an important factor affecting efficient cis- or transgene expression in Chlamydomonas (Baier et al. 2018). Accordingly, a recently released online tool enables easy design of artificial or high-expression introns throughout a synthesised coding sequence (Jaeger et al. 2019). In tandem, the generation of high-expression mutant strains UVM4 and UVM11 by suppression of epigenetic gene silencing has facilitated isolation of transformant clones showing high expression of a transgene (Neupert et al. 2009, Neupert et al. 2020).

Though this growing toolbox has broadened the scope of Chlamydomonas genetic manipulation, PCR-based and synthesis-based cloning still face problems. Modern polymerases are still prone to errors in repetitive regions (Clarke et al. 2001), low efficiency with very GC-rich regions (~70%) (Orpana et al. 2012), and are limited in primer placement. For very long genes, synthesis in parts and re-stitching can be expensive and time-consuming. For example, cloning SAGA1 (StArch Granules Abnormal 1), a 16.7 kbp gene target, required over 12 months of work, included multiple gene synthesis failures and ultimately had to be assembled from three synthesised fragments with 14 introns removed due to repetitive regions (Itakura et al. 2019). In parallel, genome annotations may contain errors misidentifying the gene start site by ignoring upstream transcription start sites corresponding to possible alternative open reading frames (Cross 2016). Therefore, cloning from the wrong ATG may result in accidental truncation of an important targeting peptide. Lastly, extensively modified coding sequences are less likely to yield native-like expression for localisation studies due to loss of *cis* regulatory elements. The gold standard for native-like expression of cloned genes therefore remains the capture of the entire native promoter, native coding sequence with native introns, and native terminator region.

3.2.2 Recombination-mediated cloning as an alternative

Cloning based on homologous recombination provides a useful alternative to PCR. Recombination can occur between homologous regions of almost any nucleotide composition and are not subject to the same guidelines for thermodynamic properties used in primer design (Yu et al. 2000). One of the most efficient forms of homologous recombination for engineering *E. coli* is lambda-red recombineering (reviewed in Murphy 1998), which harnesses the natural recombinogenic properties of the λ prophage. Lambda-red recombineering occurs *in vivo* and relies on four key proteins to function. The first, the product of *recA*, must be present to enable recombination in a host cell (Yu et al. 2000). The remaining three are derived from the λ prophage genome, called Exo, Gam and Bet. Exo is an exonuclease that strips away one of the strands of the DNA to be recombined. Gam protects the single-stranded intermediate by inhibiting host machinery that degrades ssDNA (Yu et al.

2000). Bet binds along the ssDNA intermediate (Maresca et al. 2010) and facilitates ligation and polymerisation with the homologous strand.

Although lambda-red recombineering has not been shown to function efficiently within eukaryotic cells, it can be harnessed for modification of eukaryotic genes housed in *E. coli* on bacterial artificial chromosomes (BACs). Lamda-red recombineering can be used in two main methods. In the first, insertion of a novel sequence into a BAC takes place by homologous recombination between two custom homology regions. This method has recently been put to use for recombineering genes from *C. elegans* and the model plant *A. thaliana* (hereafter Arabidopsis) (Brumos et al. 2020, Sarov et al. 2012). For example, in Brumos et al. (2020), the authors present a pipeline for tagging Arabidopsis genes from a BAC library, cleverly combining insertional recombineering with recombinase-mediated cassette exchange, (e.g., the Cre-loxP or Flp-FRT recombinase systems, reviewed in Wang et al. 2011). The pipeline employed by Brumos and colleagues proceeded as follows: (1) a BAC is edited to add a tag and selectable marker to a gene of interest by recombineering; (2) the selectable marker is removed by FLP-mediated excision; (3) the tagged gene is exchanged into a smaller vector also using Flp-FRT, which is then purified from the BAC host (Brumos et al. 2020).

The second method for Lambda-red recombineering is gap-repair recombination. By reversing the orientation of a pair of homologous sequences, a section of host DNA can be recombined into the 'gap' in a linearised plasmid. This results in a re-circularised plasmid that can be easily purified from host cells. Retrieval of a target gene by gap-repair recombination has a number of advantages over a BAC editing approach (Poser et al. 2008), and achieves the same cassette-exchange effect shown by Brumos et al. (2020) without the use of Flp-FRT. In both methods, avoiding the need to purify the entire BAC is a key consideration, as BAC purification can be timely and low yielding. For adaptation to Chlamydomonas genes, the gap-repair set-up has four further advantages: (1) avoiding BAC purification means less DNA to transform, minimising DNA fragmentation during Chlamydomonas transformation (Zhang et al. 2014); (2) BACs contain many genes, with additional copies of adjacent genes to the gene of interest potentially having an unwanted phenotypic impact on transformed Chlamydomonas lines; (3) the backbone of the available BAC collection lacks a suitable Chlamydomonas selection marker, therefore additional BAC editing to insert a suitable selection marker (Aksoy and Forest 2019) or plasmid co-transformation strategies would be required for selection; (4) a typical BAC engineering approach would require two recombination steps, which would increase cloning time, decrease efficiency and add further challenges due to the repetitive nature of the Chlamydomonas genome. In pursuit of a simple, efficient method for parallel cloning, we set out to develop a recombineering pipeline based on gap-repair recombination for Chlamydomonas nuclear genes.

3.3 Results

3.3.1 A genome analysis highlights the challenges affecting PCR

To underline the necessity of a recombineering approach for cloning Chlamydomonas genes, we set out to assess the scale of the problems facing current cloning methods. To do this we performed a genome analysis to quantify sequence features of the Chlamydomonas genome thought to have a negative impact on PCR-based cloning, or that present a barrier to gene synthesis-based cloning. We limited our analysis to measurements of gene size, repeat frequency, intron frequency, and frequencies of UTRs and splice variants in the 17,741 protein-coding genes in the nuclear genome. To aid our analysis, we used data for cloning success from a recent study by Mackinder et al. (2017), in which the authors attempted to clone 624 genes from gDNA by PCR amplification of the native coding sequence (ATG-stop). The authors used a modern, proof-reading polymerase (Phusion Hot Start II, Thermo), and a set of primers designed to anneal at the start and end of the ATG-stop region. The result was a 48% success rate. Comparison of sequence features between the successful and unsuccessful targets revealed a strong dependency of success on gene size and repeat frequency.

3.3.1.1 Gene size

Chlamydomonas genes are unusually long amongst model organisms. Fig. 3.1A on page 69 shows comparative histograms for Chlamydomonas, Arabidopsis, the model eukaryote *S. cerevisiae* (yeast), and *T. aestivum* (bread wheat), which houses a very large (~17 Gb), hexaploid genome. Here, gene size is measured from the start of the 5'UTR to the end of the 3'UTR. Chlamydomonas has a higher proportion of long genes and fewer short genes than the three other genomes tested, along with a considerably higher average gene size for Chlamydomonas (5,322 bp versus 1,430 bp for yeast, 2,187 bp for Arabidopsis and 3,521 bp for chromosome-assigned genes in wheat). Also notable is the high proportion (~10%) of genes longer than 10 kbp in Chlamydomonas. This pattern of unusually long genes may reflect an abundance of introns and untranslated regions.

Statistical analysis shows a significant effect of gene size on PCR cloning success from Mackinder et al. (2017), with the average cloned region ~2.3 kbp long while the average uncloned region was ~4.5 kbp (Mann-Whitney U = 16306, $p < 0.001$, two-tailed). By extrapolating gene size distributions of the successfully PCR-cloned genes from Mackinder et al. (2017), it is possible to predict the proportion of the whole genome clonable by ATG-stop PCR. This extrapolation is presented in Fig. 3.1B and suggests ~68% of Chlamydomonas genes would be challenging to clone due to gene size. This may be caused by a drop-off in PCR cloning efficiency for templates > 3 kbp, apparent from (Fig. 3.1B). Though the largest amplified target in Mackinder et al.

(2017) was 8 kbp, and genes at least as large as 9.7 kbp have been cloned before (Kobayashi et al. 2016), these cases appear to be highly gene specific. Alternative approaches to clone larger genes include testing a broad range of PCR conditions and DNA polymerases, amplification in fragments and re-stitching, cloning from cDNA, and gene synthesis. However, these processes can be time consuming and costly for large numbers of targets, often have low success rates and may still result in poor expression.

3.3.1.2 Gene complexity

Chlamydomonas possesses a complex nuclear genome, taken here to mean that PCR-based cloning from gDNA is complicated by sequence features. Firstly, GC content is high, at ~68% in coding regions. Data from Mackinder et al. (2017) shows that the average GC content for successfully cloned targets by ATG-Stop PCR cloning was 61.4%, while the average for unsuccessful targets was 64.3% – a value exceeded by over 41% of Chlamydomonas nuclear genes. Secondly, cloning from the nuclear genome is complicated by an abundance of repetitive regions, even if those regions themselves are extremely simple in nucleotide composition at a sequence level (e.g., a run of AAAA...). Analysis of sequence data from Mackinder et al. (2017) for 624 Chlamydomonas genes indicated a strong association between cloning success and repeat frequency; the average PCR-cloned ATG-Stop region contained 6.1 repeats per kbp whereas the average uncloned region contained 7.5 repeats per kbp (Mann-Whitney $U = 24110$, $p < 0.001$, two-tailed).

Repetitive regions are categorised into tandem repeats, (e.g., mononucleotide, AAA; dinucleotide; ATATAT; trinucleotide, ATCATCATC, etc.), and inverted repeats, (i.e., sequences followed immediately or several bases later by their reverse complement). Beyond these, larger and more interspersed repeats comprise regions that are repeated several times throughout a gene or elsewhere, causing problems for gene synthesis and sequencing. These were measured using WindowMasker, a program designed to detect a range of repeat types simultaneously (Morgulis et al. 2006) and provide one combined value, referred to as global repeats. Analysis of all Chlamydomonas genes using WindowMasker revealed an average of 38.9 global repeats per gene (6.8 per kbp) whereas Arabidopsis contains 13.7 (5.5 per kbp) and yeast contains 6.0 (4.2 per kbp). Within genes, Chlamydomonas sequences are more repetitive between their start and stop codons than in their untranslated regions, although at least one repeat was detected by WindowMasker in 36.6% of 5'UTRs and 87.6% of 3'UTRs. Fig. 3.1A expands the histrogram of gene sizes to show per-kilobase repeat frequency for Chlamydomonas, yeast, Arabidopsis and wheat. Chlamydomonas genes > 3 kbp long were the most repetitive of the four genomes. Uniquely, Chlamydomonas genes showed a slight trend of increasing repeat-content as gene length increased, likely compounding the probability of PCR failure for very

long targets.

3.3.1.3 Unsuitable primers

A major limitation of ATG-stop PCR cloning is the limited flexibility of primer design. We analysed primer suitability from a genome-wide set of computer-designed primer pairs for ATG-Stop cloning, from which the primers for the 624 targets in Mackinder et al. (2017) were taken. The average primer in the dataset had a high predicted melting temperature (Tm) of 69.2°C and a high GC content of 64.2%. Primer Tm and GC content are expected to be high in comparison to other organisms with less GC-rich genomes, however, many primers also breached recommended thresholds pertaining to length, secondary structure formation, repetitive sequences or 3' GC content using Primer3Plus. The Primer3Plus check_primers module (Untergasser et al. 2007) utilises a strict set of thresholds for primer suitability. Analysis of our genome-wide primer set resulted in the rejection of over 60% of individual primers in the dataset, even when the program was set to ignore predicted annealing temperatures (Fig. 3.1C, orange bars). Under these settings, only 13% of pairs were free from detectable issues in both primers. Interestingly, there was a high GC content mismatch between forward and reverse primers with a considerably higher GC content of reverse primers (Fig. 3.1C, inset).

3.3.1.4 Mis-annotation of start sites

Another challenge associated with PCR or gene synthesis is incorrectly annotated gene models that lead to cloning of a non-biologically relevant sequence. The analysis of transcript models in the Chlamydomonas genome shows that additional ATGs upstream of the annotated start site are highly prevalent (Cross 2016) (Fig. 3.1D). Cross categorized these potential upstream open reading frames (uORFs) into three classes: class 1 uORFs initiate in-frame with the annotated start site, potentially producing an N-terminal extension relative to the annotated gene model; class 2 uORFs initiate out-of-frame with the annotated start site and terminate within the coding sequence; and class 3 uORFs initiate and terminate within the 5'UTR. Data on the presence of Kozak sequences preceding class 1 uORFs suggests that approximately half are the correct translation initiation site *in vivo* (Cross 2016). In a PCR-based approach where a constitutive promoter is used, cloning from the wrong ATG may result in an out-of-frame or truncated product, potentially removing essential signal sequences for correct targeting. 57 of the 298 successfully cloned genes from Mackinder et al. (2017) contained a class 1 in-frame ATG upstream of the cloned region, therefore ~10% of cloned regions may have encoded truncated protein products.

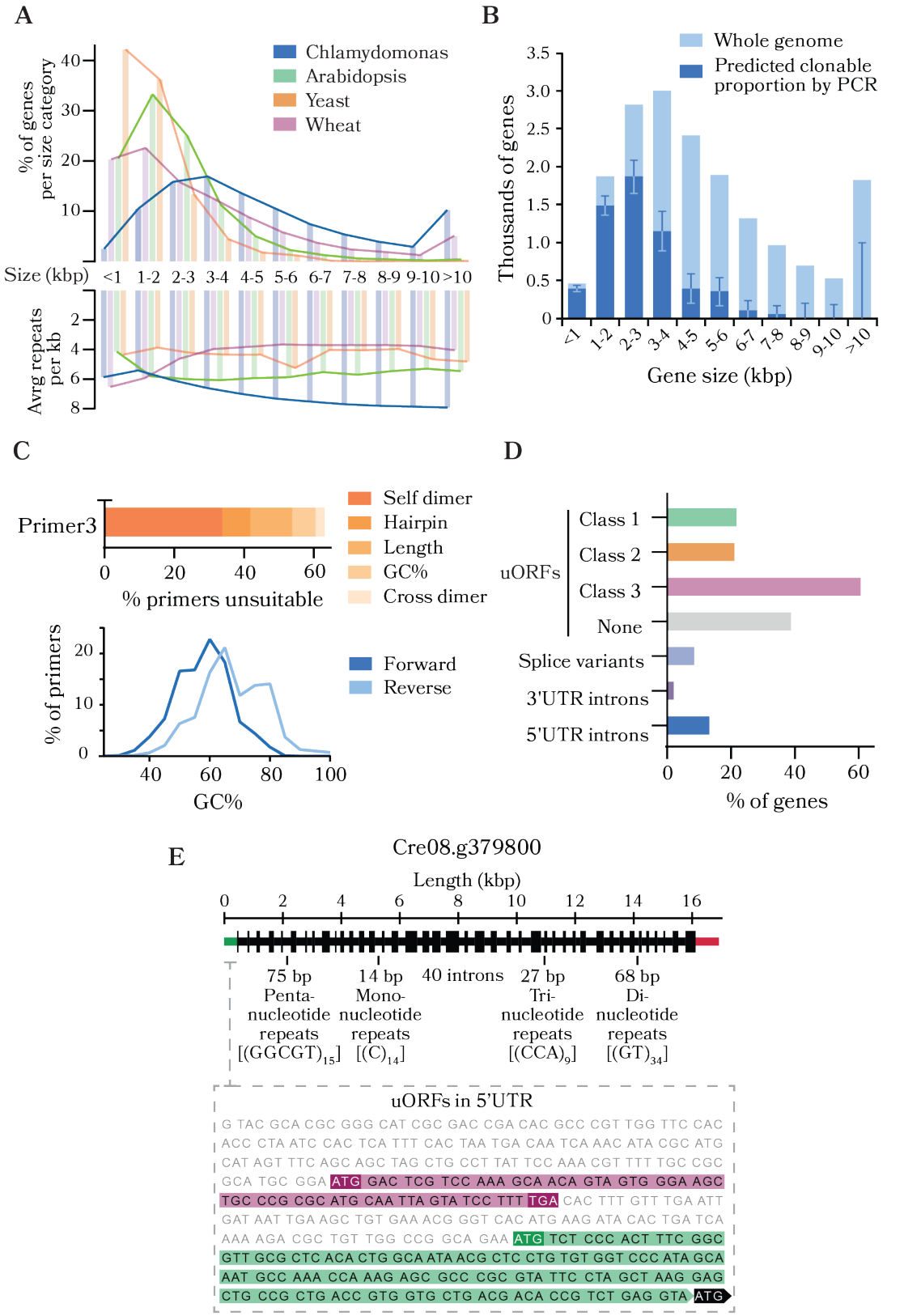

A
% of genes per size category
Chlamydomonas
Arabidopsis
Yeast
Wheat
Size (kbp)
<1 1-2 2-3 3-4 4-5 5-6 6-7 7-8 8-9 9-10 >10
Avrg repeats per kb
B
Thousands of genes
Whole genome
Predicted clonable proportion by PCR
Gene size (kbp)
C
Primer3
% primers unsuitable
Self dimer
Hairpin
Length
GC%
Cross dimer
% of primers
Forward
Reverse
GC%
D
uORFs
Class 1
Class 2
Class 3
None
Splice variants
3'UTR introns
5'UTR introns
% of genes
E
Cre08.g379800
Length (kbp)
75 bp Penta-nucleotide repeats [(GGCGT)15]
14 bp Mono-nucleotide repeats [(C)14]
40 introns
27 bp Tri-nucleotide repeats [(CCA)9]
68 bp Di-nucleotide repeats [(GT)34]
uORFs in 5'UTR
G TAC GCA CGC GGG CAT CGC GAC CGA CAC GCC CGT TGG TTC CAC
ACC CTA ATC CAC TCA TTT CAC TAA TGA CAA TCA AAC ATA CGC ATG
CAT AGT TTC AGC AGC TAG CTG CCT TAT TCC AAA CGT TTT TGC CGC
GCA TGC GGA ATG GAC TCG TCC AAA GCA ACA GTA GTG GGA AGC
TGC CCG CGC ATG CAA TTA GTA TCC TTT TGA CAC TTT GTT TGA ATT
GAT AAT TGA AGC TGT GAA ACG GGT CAC ATG AAG ATA CAC TGA TCA
AAA AGA CGC TGT TGG CCG GCA GAA ATG TCT CCC ACT TTC GGC
GTT GCG CTC ACA CTG GCA ATA ACG CTC CTG TGT GGT CCC ATA GCA
AAT GCC AAA CCA AAG AGC GCC CGC GTA TTC CTA GCT AAG GAG
CTG CCG CTG ACC GTG GTG CTG ACG ACA CCG TCT GAG GTA ATG

Figure 3.1: **Analysis of Chlamydomonas genes reveals the scale of the problems affecting PCR-based cloning of nuclear DNA sequences. A.** Histograms arranging the 17,741 genes of the Chlamydomonas nuclear genome by size, binned by 1,000 bp. The number of genes in the Chlamydomonas, Arabidopsis, yeast and wheat genomes within each 1 000 bp category are displayed above the x-axis, revealing more large genes in Chlamydomonas. Below, the frequency of repeats detectable by the program WindowMasker for genes in each size category are shown, revealing a trend of high repeat frequency for Chlamydomonas. **B.** Again arranged by size, the proportion of Chlamydomonas genes in each 1,000-bp size category we predict to be clonable by ATG-stop PCR are shown in dark blue. This prediction is an extrapolation from results in Mackinder et al. (2017). Error bars are 95% confidence intervals. **C.** The results of analysing a genome-wide set of ATG-stop PCR primers using Primer3. Over 60% of primers were deemed unsuitable, even when Tm was removed as a factor. Below, then inset shows a disparity in the GC content of primers, reflecting particularly high GC content at the 3' end of nuclear genes. **D.** The percentage of nuclear genes containing detectable upstream open reading frames, splice variants, or introns. **E.** An example gene displaying many of the problems associated with PCR-based cloning; long, repetitive, replete with introns and containing a large 5' and 3'UTR. The 5'UTR sequence is expanded below, showing a possibly misannotated start site due to a class 1 uORF (green) along with just one of seven detected class 3 uORFs (magenta). Luke Mackinder recorded data for PCR cloning success presented in panel **B** and generated the genome-wide set of primers analysed in panel **C**.

3.3.1.5 Introns, UTRs and splice variants

Chlamydomonas genes have a relatively high intron frequency, providing a further challenge for PCR-based cloning by increasing gene length. The average gene contains 7.3 introns with an average intron length of 373 bp compared to an average exon length of 190 bp. 94% of genes contain introns between their start and stop codons, 13% of genes contain one or more introns in their 5'UTRs and 3.4% have introns in their 3'UTRs. ATG-Stop cloning would omit introns in UTR regions, potentially missing critical regulatory information. Furthermore, approximately 9% of genes are annotated with two or more transcript models that result from alternative splicing (Fig. 3.1D). This variation would be missed through cloning from cDNA or through gene synthesis that excludes native introns.

Retrieval of gDNA sequences from BACs can circumvent many of these sequence features simultaneously. Recombineering can avoid issues of high GC content and repetitive sequences, provided that the homology regions are themselves unique. In theory, recombineering should be able to clone very large sections of DNA. Recombineering can therefore also avoid issues of mis-annotation by capturing upstream regions that contain all the native 5' *cis* regulatory features, including alternative start sites. Fig. 3.1E shows a gene from chromosome 8 that exhibits several examples of the sequence features presented in this genome

analysis. Cre08.g379800 is > 16 kbp long with 40 introns, contains mono-, di-, tri- and pentanucleotide repeat regions of ≥ 9 repeats. It also contains a potential misannotated upstream ATG in the 5'UTR that could initiate a class 1 uORF, as well as seven class 3 uORFs (Cross 2016). Cre08.g379800 was successfully cloned and tagged by recombineering.

3.3.2 Development of a recombineering pipeline

In developing our recombineering pipeline, we adapted a method designed for *C. elegans* by Sarov et al. (2006) to clone genes with their upstream regions intact and a 3' tandem fluorescent-affinity tag. The pipeline relies on gap-repair recombination to capture a target region contained on a BAC (or fosmid or cosmid), facilitated by the sequential introduction of two plasmids into the BAC clone. The first, pRed, contains a temperature sensitive origin of replication (pSC101) and is used to make a target BAC clone recombination competent by expressing the lambda viral Bet, Gam and Exo proteins (*Red*αβγ), as well as *recA* (Sarov et al. 2006). The second houses a tagging cassette that is introduced into the BAC clone as linear DNA and retrieves the target region by gap-repair recombination.

The simplicity of our pipeline enables completion in four days using only generic reagents. The final recombineered product is a plasmid containing the target region (typically including the native promoter, 5'UTR and open reading frame) recombined in-frame with a downstream fluorescent protein and terminator (Fig. 3.2A). Our pipeline has four key steps: (1) *E. coli* harbouring a BAC containing the gene of interest is made recombination competent by transformation with the pRed vector (Fig. 3.2B); (2) Expression of *Red*αβγ and *recA* is induced by addition of arabinose followed by transformation with a linear tagging cassette including 50 bp homology arms to the target gene (Fig. 3.2C); (3) kanamycin provides selection for successfully recombineered products, and a shift to 37°C inhibits pRed replication to minimise further undesired recombination (Fig. 3.2D); and (4) plasmid products are isolated and verified by restriction digest and junction sequencing (Fig. 3.2E).

3.3.2.1 Design of a series of gene tagging cassettes

We designed and built a series of tagging cassettes to facilitate gene tagging using a variety of tags and housing a range of antibiotic resistance cassettes. The original cassette, named pLM099 (Fig. 3.2A), was constructed by six-fragment Gibson assembly and contained eight key features: (1) the codon optimized YFP CrVenus for localisation studies, (2) a 3xFLAG tag for downstream immunoblotting, (3) the PSAD terminator as a general regulatory element, (4) the paromomycin selection marker (AphVIII) for transformation into Chlamydomonas, (5) the p15A medium-copy-number origin of replication to prevent over-production of the recombineered product once retrieval into a plasmid is achieved, (6) the kanamycin resistance gene

(Kan^R) for selection in *E. coli*, (7) the ccdB counter selection gene to minimise false positives due to pLM099 carryover (Bernard and Couturier 1992), and (8) the I-SceI restriction site to allow linearisation prior to transformation into Chlamydomonas. I-SceI has an 18 bp recognition site not found within the Chlamydomonas reference genome, allowing for functionality with any genomic target.

Amplification of the tagging cassette from pLM099 is performed using primers containing 50 bp homology arms corresponding to regions flanking the target gene (Fig. 3.2A). Accordingly, one primer binds at least 2,000 bp upstream of the start codon to encompass the native 5' promoter and UTR region and the other primer binds at the 3' end of the coding region (immediately upstream of the stop codon). The annealing site of the reverse primer can easily be altered to amplify a cassette from pLM099 that can clone genes without a fluorescent tag or with only the 3xFLAG tag. To further expand the functional application of our recombineering pipeline we designed additional backbone vectors (pLM160, 161, 162 and 459) that enable protein tagging with the fluorophores mScarlet-i-3xFLAG (Bindels et al. 2016), mNeonGreen-3xFLAG (Shaner et al. 2013) and mTurquoise2-3xHA (Goedhart et al. 2012) and that allow selection with hygromycin or zeocin (Fig. 3.2A). These make possible multi-protein labelling within the same cell and also permit complementation of paromomycin-resistant mutants from the Chlamydomonas Library Project (CLiP) (Li et al. 2019). The presence of either 3xFLAG or 3xHA tag enables use of the vectors for affinity purification to explore interacting partners of tagged proteins. Different fluorophore pairs (i.e., mNeonGreen and mScarlet-i) could also be used for FRET based studies to explore protein-protein interactions. In addition, by choosing different primer binding sites, all vectors can be used for cloning genes without fluorescence tags or with only short affinity tags (3xFLAG and 3xHA).

3.3.2.2 Small scale application of the pipeline

We initially tested our pipeline on 12 targets. For all targets except one, plasmids isolated from most picked colonies gave a correct banding pattern after restriction digest. Sequencing then confirmed the successful cloning of 11 out of our 12 targets. To further expand the capabilities of our pipeline we tested whether we could successfully recombineer large and complex gene from fosmids (available from the Chlamydomonas Resource Center). We targeted SAGA1 (fosmid VTP41289), a gene that had previously been highly challenging to gene synthesize (Itakura et al. 2019) and was not available in the BAC library (Brazelton et al. 2018). Restriction digest of recombineered plasmids purified from three colonies all showed the correct digestion pattern. Sequencing confirmed that the 19,601 bp target region, that included 2,913 bp upstream of the predicted SAGA1 start codon, was successfully cloned. Similarly, we successfully cloned both STT7 (fosmid VTP75523),

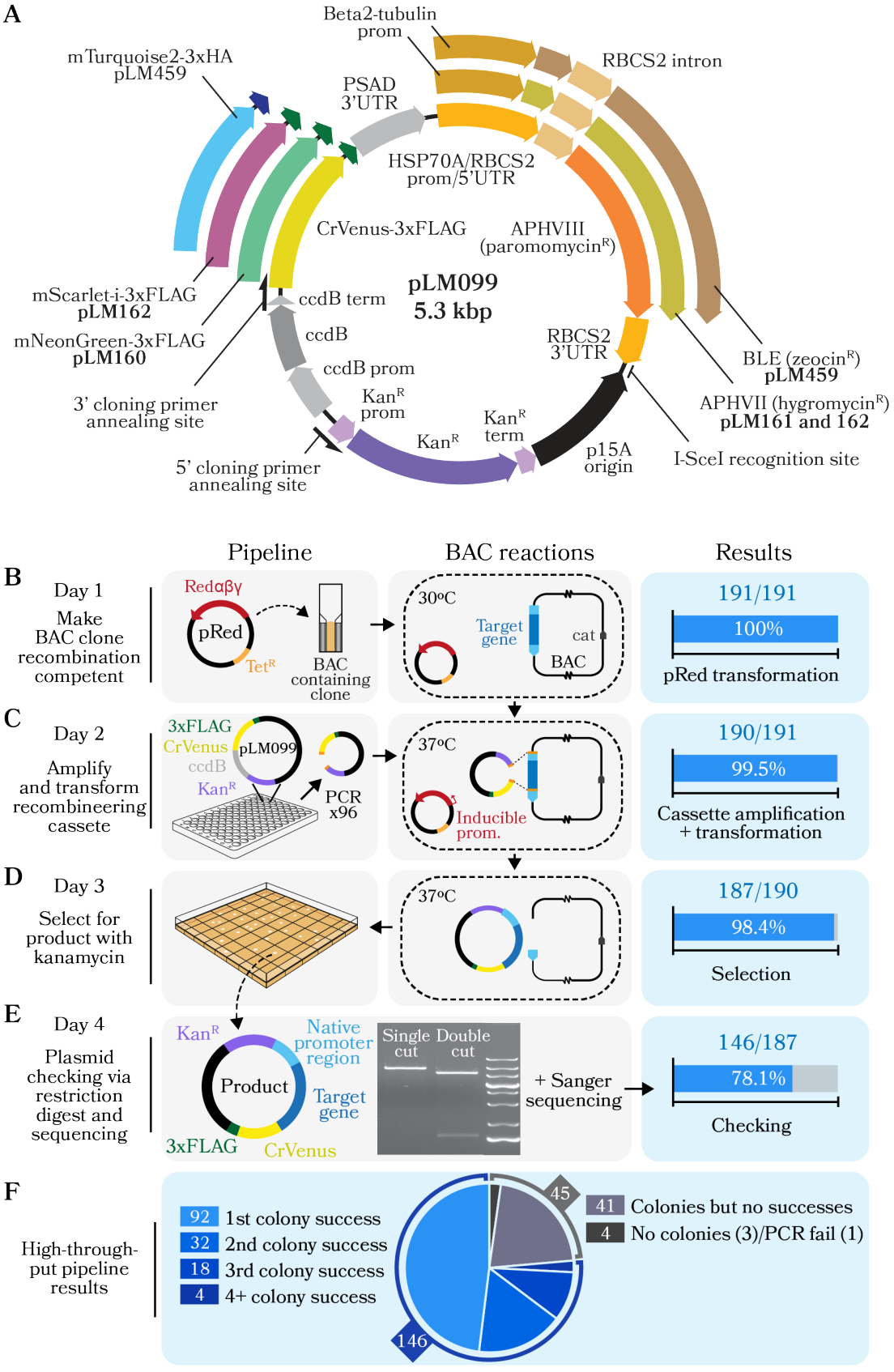

A
Beta2-tubulin prom
mTurquoise2-3xHA pLM459
RBCS2 intron
PSAD 3'UTR
HSP70A/RBCS2 prom/5'UTR
CrVenus-3xFLAG
APHVIII (paromomycin^R)
pLM099
5.3 kbp
mScarlet-i-3xFLAG pLM162
ccdB term
ccdB
mNeonGreen-3xFLAG pLM160
ccdB prom
RBCS2 3'UTR
BLE (zeocin^R) pLM459
3' cloning primer annealing site
Kan^R prom
Kan^R term
APHVII (hygromycin^R) pLM161 and 162
Kan^R
p15A origin
I-SceI recognition site
5' cloning primer annealing site
Pipeline
BAC reactions
Results
B
Day 1
Make BAC clone recombination competent
Redαβγ
pRed
Tet^R
BAC containing clone
30°C
Target gene
cat
BAC
191/191
100%
pRed transformation
C
Day 2
Amplify and transform recombineering cassete
3xFLAG
CrVenus
ccdB
Kan^R
pLM099
PCR x96
37°C
Inducible prom.
190/191
99.5%
Cassette amplification + transformation
D
Day 3
Select for product with kanamycin
37°C
187/190
98.4%
Selection
E
Day 4
Plasmid checking via restriction digest and sequencing
Kan^R
Native promoter region
Product
Target gene
3xFLAG
CrVenus
Single cut
Double cut
+ Sanger sequencing
146/187
78.1%
Checking
F
High-through-put pipeline results
92 1st colony success
32 2nd colony success
18 3rd colony success
4 4+ colony success
45
41 Colonies but no successes
4 No colonies (3)/PCR fail (1)
146

Figure 3.2: **Recombineering pipeline in Chlamydomonas. A.** Sequence features of plasmid pLM099, from which pLM160/1/2 and pLM459 were derived to expand the spectral palette of encoded fluorophores and resistance cassettes. **B.** Day 1 of the recombineering pipeline, involving transformation of BAC-containing cells with the plasmid pRed, encoding *Red*αβγ and *recA*. **C.** On day 2, recombineering cassettes were amplified from pLM099 with gene-specific primers. These linear PCR-products were transformed into the transformants from the previous day, then recombination induced by a shift to 37°C. Transformants were then recovered and plated on kanamycin-containing LB agar. **D.** On day 3, colonies from day 2's transformation were picked and grown in small-scale cultures for plasmid checking. **E.** On day 4, cultures were miniprepped and digest patterns ascertained using I-SceI and a double cutter, chosen based on the predicted sequence. **F.** The success rate from a large-scale attempt at parallel recombineering 191 genes encoding components of the Chlamydomonas CCM. Gary Yates, James Barrett, Irina Grouneva, Chun Sing Lau and Charlotte Walker aided in development and implementation of the pipeline. Recombineered products were checked for correct digestion (panel **E**) in collaboration with Gary Yates.

the key state transition kinase, and the chlorophyll synthase, CHLG (fosmid VTP23797), into pLM160 for downstream localisation and affinity purification using the mNeonGreen-3xFLAG tag.

3.3.2.3 High-throughput application of the pipeline

Confident that our recombineering method was robust we pursued the development of a large-scale pipeline that would allow the parallel tagging of genes with most steps achievable in 96-well format. We shortlisted 191 genes that could be mapped to a clone from the Chlamydomonas BAC library. 81 genes previously targeted for ATG-stop cloning in 2017 were retried here by recombineering, this time with > 2,000 bp upstream sequence included. 41 of these were previously unsuccessful by PCR and 40 were previously successful but included here to compare the effect of retaining the native promoter. The latter included five targets that contain a class 1 uORF and so may have previously produced misleading localization data due to expression of a truncated protein. Selection of the remaining 110 targets was guided by pyrenoid transcriptomics (Brueggeman et al. 2012) and pyrenoid proteomics data (Mackinder et al. 2016). *E. coli* strains containing the correct BACs were recovered from the BAC library and processed in parallel using 96-format culturing plates. We randomly selected a small subset of BAC clones to check for the presence of our target genes by PCR, all of which were correct.

To optimise efficiency, we successively ran the pipeline three times removing successful targets once confirmed. 100% of our 191 target BAC lines were made recombination competent and out of the 191 target genes, one gene-specific tagging cassette failed to amplify, possibly due to the formation of secondary

structure(s) within the 50 bp homology regions of the primers. Of the 190 that amplified successfully, 187 yielded colonies after selection with kanamycin. Validation by enzymatic digestion confirmed that 146 of these lines contained correct recombineering plasmid products (Fig. 3.2F). Recombineering plasmid products from the 146 successful lines were extracted and their junctions confirmed by Sanger sequencing. Our high-throughput pipeline had an overall efficiency of 76%. 26 target genes that were unsuccessful by PCR in 2017 were successfully cloned here by recombineering, and all five previously successful targets containing class 1 uORFs retried here were successful.

In addition, for 14 out of the 146 correctly recombineered lines in our high-throughput pipeline, use of an alternative BAC from the library yielded success after an initial failure. We found that for approximately half of the target genes it was necessary to validate multiple colonies by enzymatic digest in order to rule out false positives; beginning with the 187 colony-producing lines from our high-throughput pipeline, picking just a single colony gave a 49% success rate, screening a second colony increased the success rate to 66% and a third colony gave a 76% success rate. For a small proportion of targets screening > 3 colonies led to identification of a correctly recombineered product (Fig. 3.2F). Taken together with our 12 initial targets, we successfully cloned 157 out of 203 target regions from BACs using our recombineering pipeline, achieving an efficiency of 77%.

3.3.3 Analysis of successes reveals good tolerance of large and complex targets

We analysed our 203 target regions for differences in gene size and repeat frequency between successful and unsuccessful reactions. Cloning success rate versus gene size is shown as a histogram in Fig. 3.3A, revealing no obvious trend. This is reflected in a lack of statistical differences between the median lengths (Mann-Whitney U = 3303, p = 0.38, two-tailed); average successful targets were 7.26 kbp while average unsuccessful targets were 7.32 kbp long. This contrasts clearly with the strong effect of target size on cloning success for PCR-based cloning from Mackinder et al. (2017) (shown in Fig. 3.3A and Fig. 3.1B). Furthermore, our recombineering results produced a remarkably high efficiency for cloning very long genes; the longest recombineered region was 22,773 bp, with successful cloning achieved for 24/30 targets > 10 kbp long.

We then compared our cloning success rate to the per-kilobase repeat frequency. For our recombineering pipeline, the average successful target contained 6.48 repeats per kbp while the average unsuccessful target contained 6.81, which did not amount to a statistically significant difference (Mann-Whitney U = 3129, p = 0.17, two-tailed). In contrast, there is a clear negative effect on PCR-based cloning success for targets with > 4.8 repeats per kbp (Fig. 3.3B). For the most repetitive targets involved in

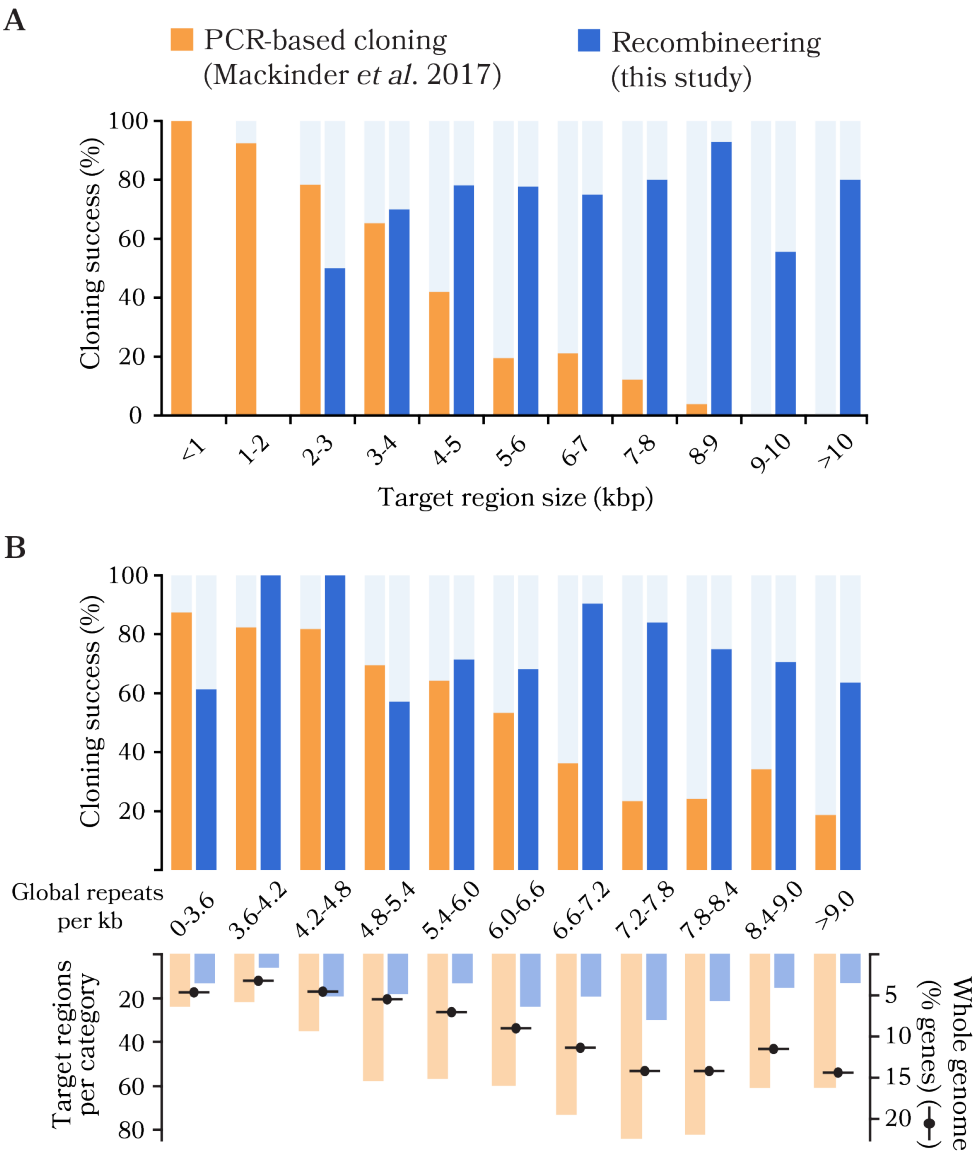


Figure 3.3: **Size and complexity tolerance of our recombineering pipeline vs PCR. A.** Cloning success rates for gene targets arranged by size (binned by 1,000 bp) shows no clear size-dependency on recombineering success, contrasting markedly with PCR success. **B.** Cloning success rates for gene targets arranged by repeat frequency (binned by 0.6 repeats per kbp) shows a slight negative trend for PCR success, but no clear trend for recombineering. Below the x-axis, the genome-wide distribution of genes is shown, revealing that most Chlamydomonas genes contain between 7 and 9 repeats per kbp. These highly repetitive genes were associated with low PCR success rates but high recombineering success rates.

our analysis (those with > 9 repeats per kbp), our recombineering cloning efficiency remained above 60%; an efficiency over three times higher than for ATG-stop cloning (Fig. 3.3B). The genome-wide distribution of repeat frequencies shows that most nuclear genes contain between 7 and 9 repeats per kbp, presenting a technical challenge for PCR-based cloning (Fig. 3.3B). The success rate of our recombineering method for genes with these repeat frequencies suggests many of these repetitive genes could now be more easily clonable.

3.3.4 Localisation of tagged proteins agrees with previous data

The aim of the pipeline was to enable parallel gene tagging for localisation studies, a powerful tool for estimating gene function in a multifaceted system like the CCM. It was therefore important to investigate whether localisation of proteins expressed from recombineered genes (with their native promoter and 3' Venus tag) agreed with localisation results derived from previous work. We selected a small subset of recombineered genes for confirmation of localisation by fluorescent microscopy, including PSAF, TAB2, ISA1, LCI9 and LCIB (3.4A-C). Recombineering products containing these genes were confirmed by sequencing, then transformed into strain CC-4533, the same strain used to generate the CLiP mutant library Li et al. 2019). Lines expressing these genes were also immunoblotted against the C-terminal 3xFLAG epitope to confirm fusion protein size (not shown).

PSAF (PSI subunit F), TAB2 (a protein linked to early PSI assembly) and ISA1 (Isoamylase 1) were present in the pyrenoid (Fig. 3.4A), supporting previous affinity purification mass spectrometry data (Mackinder et al. 2017) and pyrenoid proteomics data (Mackinder et al. 2016). PSAF is a core transmembrane subunit of photosystem I (see Fig. 1.4 on page 17 in Chapter 1), and so, as expected, shows strong colocalization with chlorophyll outside of the pyrenoid, however in addition it clearly localizes to the thylakoid tubules traversing the pyrenoid. Interestingly, in the pyrenoid tubules the chlorophyll signal is minimal, particularly at the "pyrenoid tubule knot" where the tubules converge (Engel et al. 2015). These data along with the localization of other PSI and PSII components to the pyrenoid tubules (Mackinder et al. 2017) suggest that the tubules contain both PSI and PSII but that chlorophyll-containing light harvesting complexes found within the pyrenoid may be quenched or at low abundance.

Tagged Cre17.g702500 (TAB2), a protein linked to early PSI assembly (Dauvillée et al. 2003) and previously identified as an interactor with PSBP4 at the periphery of the pyrenoid (Mackinder et al. 2017) was also enriched at the pyrenoid. Interestingly, the location of TAB2 is not just restricted to the pyrenoid periphery but is also found within the pyrenoid forming distinct small foci (Fig. 3.4A). This may indicate that early PSI assembly could be occurring within the pyrenoid as well as at the pyrenoid periphery (Uniacke and Zerges 2009).

ISA1 is a starch debranching enzyme that is essential for starch synthesis, with ISA1 deletion lines lacking both chloroplast and pyrenoid starch (Mouille et al. 1996). The presence of the pyrenoid starch sheath and its correct organization is known to be critical for correct CCM function; the absence of starch in an ISA1 knock out (4-D1) resulted in incorrect LCIB localization (correct localisation shown in Fig. 3.4C), retarded growth at very low CO_2 (0.01% v/v) and reduced inorganic carbon affinity (Toyokawa et al. 2020). Interestingly, in Toyokawa et al. (2020) the authors failed to attain localization data for an ISA1-mCherry fusion driven by the HSP70A/RBCS2 hybrid promoter.

3.3.5 Comparison of native promoter- and PSAD promoter-driven gene expression

To confirm that localization of proteins driven by their native promoter did not differ from those driven by the constitutive PSAD promoter we compared localization between native-LCIB-Venus and PSAD-LCIB-Venus. LCIB is an essential CCM component that shows dynamic relocalization to the pyrenoid periphery at CO_2 levels < 0.04% (v/v) (Yamano et al. 2010). LCIB expressed from its endogenous promoter was localized to the pyrenoid periphery at very low CO_2 (0.01% v/v), in full agreement with localization data when LCIB expression is driven by the constitutive PSAD promoter (Fig. 3.4C). Similarly, LCI9, a low-CO_2 inducible protein, was previously shown via PCR-based cloning to localize to the pyrenoid periphery. Recombineered LCI9 was 7,160 bp long including the native promoter region, and was cloned into pLM099, 160 and 162 (Fig. 3.2A) to assess localisation with CrVenus, mNeonGreen and mScarlet-i. All fluorophores displayed the same pyrenoid periphery localization pattern (Fig. 3.4B) and agree with the localization information obtained when LCI9 expression was driven from the PSAD promoter (Fig. 3.4B), supporting the use of 2,000 bp upstream regions as promoters for fusion protein expression.

As our pipeline retains the native promoter of the target gene, we hypothesized that fluorescence output would be representative of relative changes in protein abundance in response to environmental conditions. To test this, we grew lines with LCIB driven from either the constitutive PSAD promoter (PSAD-LCIB-Venus) or its native promoter (Native-LCIB-Venus). LCIB-Venus signal stayed relatively constant between high (3% v/v) and low (0.04% v/v) CO_2 when LCIB was expressed from the PSAD promoter but showed an approximate 8-fold increase between these conditions when the native promoter was used, with this change consistent across three independently transformed lines (mean ± SEM) (Fig. 3.4D). This agrees with previous immunoblotting data, in which a comparable fold increase was seen in LCIB abundance when cells were transferred from high CO_2 to low CO_2 (Yamano et al. 2010).

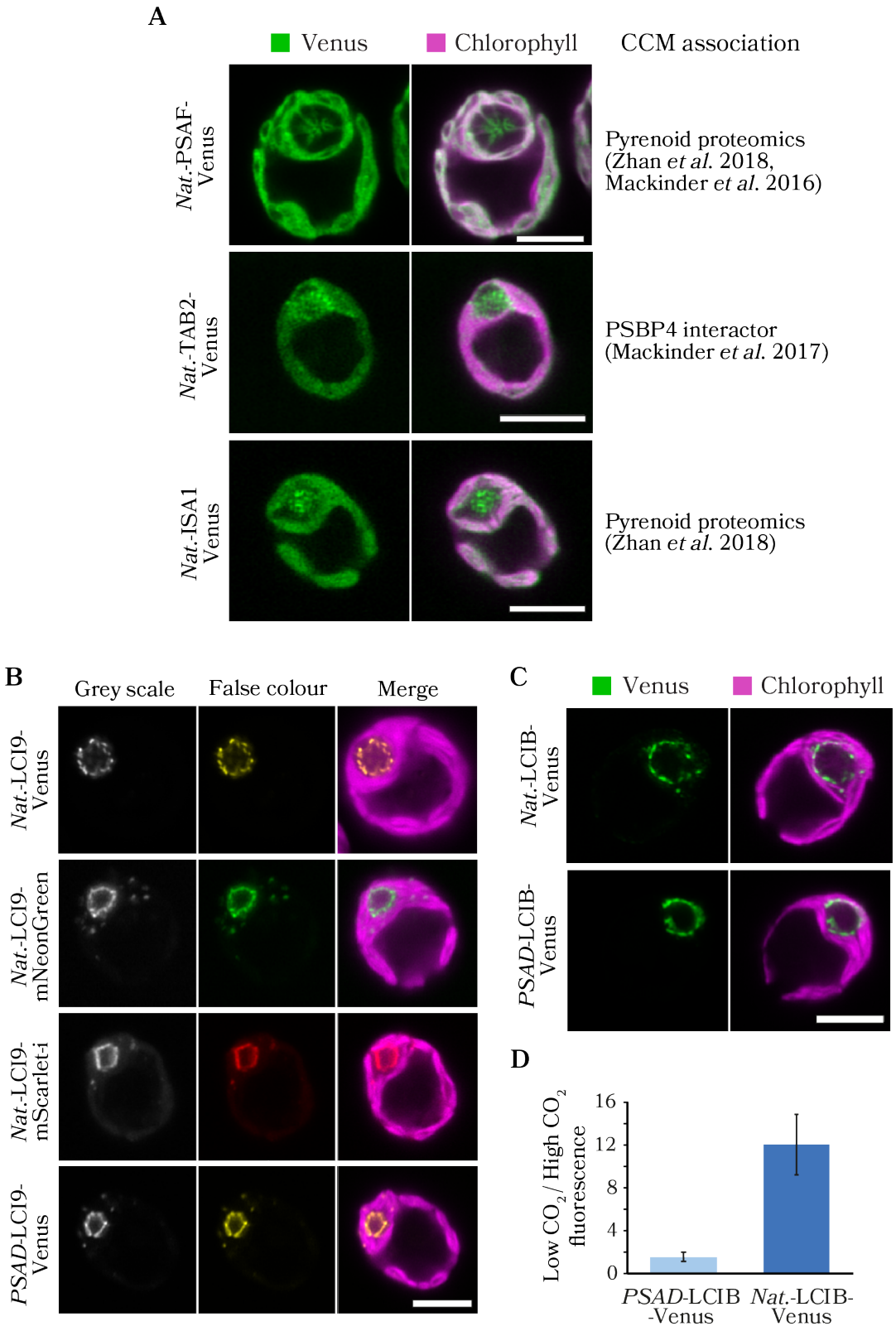

A
Venus
Chlorophyll
CCM association
Nat.-PSAF-Venus
Pyrenoid proteomics (Zhan et al. 2018, Mackinder et al. 2016)
Nat.-TAB2-Venus
PSBP4 interactor (Mackinder et al. 2017)
Nat.-ISA1 Venus
Pyrenoid proteomics (Zhan et al. 2018)
B
Grey scale
False colour
Merge
Nat.-LCI9-Venus
Nat.-LCI9-mNeonGreen
Nat.-LCI9-mScarlet-i
PSAD-LCI9-Venus
C
Venus
Chlorophyll
Nat.-LCIB-Venus
PSAD-LCIB-Venus
D
Low CO2 / High CO2 fluorescence
16
12
8
4
0
PSAD-LCIB -Venus
Nat.-LCIB-Venus

Figure 3.4: **Localisation data for a small subset of recombineered genes using fluorescent microscopy. A.** Confocal fluorescence microscopy images of transformed CC-4533, showing localisation data for recombineered three genes, each with their native promoter region and a C-terminal Venus tag. PSAF is the F subunit of PSI; TAB2 is a putative PSI assembly factor; ISA1 is a chloroplastic starch debranching enzyme. These data confirm previous proteomics and interactome data (listed to the right). **B.** Localisation data for LCI9 tagged with Venus, mNeonGreen, mScarlet-I and expressed from either the native promoter or the strong PSAD promoter plus 5'UTR, revealing a high level of agreements between imaged lines. **C.** LCIB expressed from either the native promoter region or PSAD promoter + 5'UTR, showing a high level of agreement in localisation. **D.** Relative change in LCIB-Venus fluorescence between high and low CO_2 detected using a plate reader. Three independently transformed lines were tested. Error bars are SEM. Gary Yates conducted the microscopy and plate reader fluorescence measurements. Scale bars are 5 μm throughout.

3.3.6 Mutant rescue with recombineered LCIB and STT7

Finally, we tested that our recombineering pipeline could be used to successfully complement mutants using two examples; first in a CLiP mutant deficient in LCIB (LMJ.RY0402.215132), and second in a mutant deficient in STT7 generated by CRISPR-Cas9. We first transformed *native-LCIB-Venus* (cloned into pLM161 that contains the *APHVII* gene conferring hygromycin resistance) into a CLiP *lcib* mutant (LMJ.RY0402.215132). Four transformants showing Venus fluorescence were selected for microscopy and growth phenotyping. All showed a typical pyrenoid peripheral localization pattern when grown at very low CO_2 (Fig. 3.5A) and all rescued the *lcib* mutant phenotype to varying degrees, with *lcib::LCIB-Venus-1* showing close-to-complete rescue (Fig. 3.5B).

Secondly, we aimed to rescue a CLiP *stt7* mutant (LMJ.RY0402.236842) that contains an insertion in the 5'UTR of STT7. The insertion was confirmed by PCR across the 5' of the gene, but Western blotting using the anti-STT7 antibody showed a weak but unmistakeable signal at ~77 kDa, implying this strain still accumulates some STT7 (Fig. 3.5C). We therefore set out to generate a mutant deficient in STT7 using CRISPR-Cas9 in the high expression strain UVM4, with the eventual aim of purifying tagged STT7. UVM4 cells were co-transformed with the hygromycin resistance cassette and a sgRNA-Cas9 RNP targeting STT7 exon 2. 87 transformant colonies were recovered on hygromycin in low light and screened by PCR across exon 2. Eight colonies yielded a larger amplicon or no amplicon, indicating the presence of an insertion; these were taken forward for Western blotting with anti-STT7. Four colonies showed no evidence of a band at ~77 kDa, one of which is shown in Fig. 3.5D. This colony, cTE035, was taken forward for rescue with recombineered STT7.

The 6.9 kbp ATG-stop region of STT7 was originally recombineered into pLM160

and sequenced to ensure correct assembly. PCR and restriction cloning was used to replace mNeonGreen-3xFLAG with a Strep-tag® II affinity tag for downstream purification, and to insert an upstream PSAD promoter for high expression. *PSAD*-STT7-Strep was transformed into cTE035 and transformants were screened as above, first by PCR then Western blotting (Fig. 3.5E). Three transformants (cTE038, 039 and 040) showing correct bands sizes on a Western blot were selected for phenotype testing by measuring the cells' ability to perform a transition to state II, which is dependent on STT7. This is demonstrated for cTE038 in Fig. 3.5F; cells of UVM4, cTE035 and cTE038 were prepared for measurements in anoxic conditions to trigger a transition to state II, then flash frozen for measurement of chlorophyll fluorescence at 77 K (see Methods, Section 2.7.1). The mutant cTE035 shows an impaired ability to enter state II relative to UVM4, observable as a smaller fluorescence peak at ~720 nm, which is associated with PSI chlorophyll fluorescence (Fig. 3.5F). Strong PSI fluorescence is strongly boosted in cTE038, indicating a transition to state II and implying overexpression of a functional STT7.

3.3.7 Purification of STT7 from Chlamydomonas

STT7 is a central regulatory protein in cellular bioenergetics, phosphorylating thylakoid proteins during a transition to state II (Depège et al. 2003). It is also implicated in both the upregulation of the CCM and CET by unknown mechanisms, and may be activated by interaction with the stromal domain of cytb_6f (Dumas et al. 2017). Having shown our recombineered STT7 could rescue state transitions, we set out to purify STT7 with the eventual aim of ascertaining the structure. The structure of the soluble kinase domain of a putative STT7 homologue from the green alga *Micromonas sp.* RCC299 was solved by x-ray crystallography in 2013, but the N-terminal domain was missing, and may be significantly differently to the N-terminal transmembrane domain in Chlamydomonas STT7 (Guo et al. 2013).

Eight litres of cTE038 cells expressing *PSAD*-STT7-Strep were grown in bubbled TAP media until mid-log phase, then switched to TP for 12 h. Cells were harvested, and the thylakoid membranes extracted and solubilised in styrene maleic acid copolymer solution to stabilise the protein. STT7-Strep was extracted from the thylakoids on a StrepTactin gravity column, then concentrated and prepared for SDS-PAGE. Fig. 3.5G shows the results of the purification, indicating a sample strongly enriched in ~77 kDa protein (left) shown by Western blotting to be STT7 (right). Due to time constraints on the project, structural and biochemical work was not pursued further. Instead, the purification results stand as evidence that successful purification of proteins expressed from recombineered genes is tenable and lays the groundwork for future structural and interaction work.

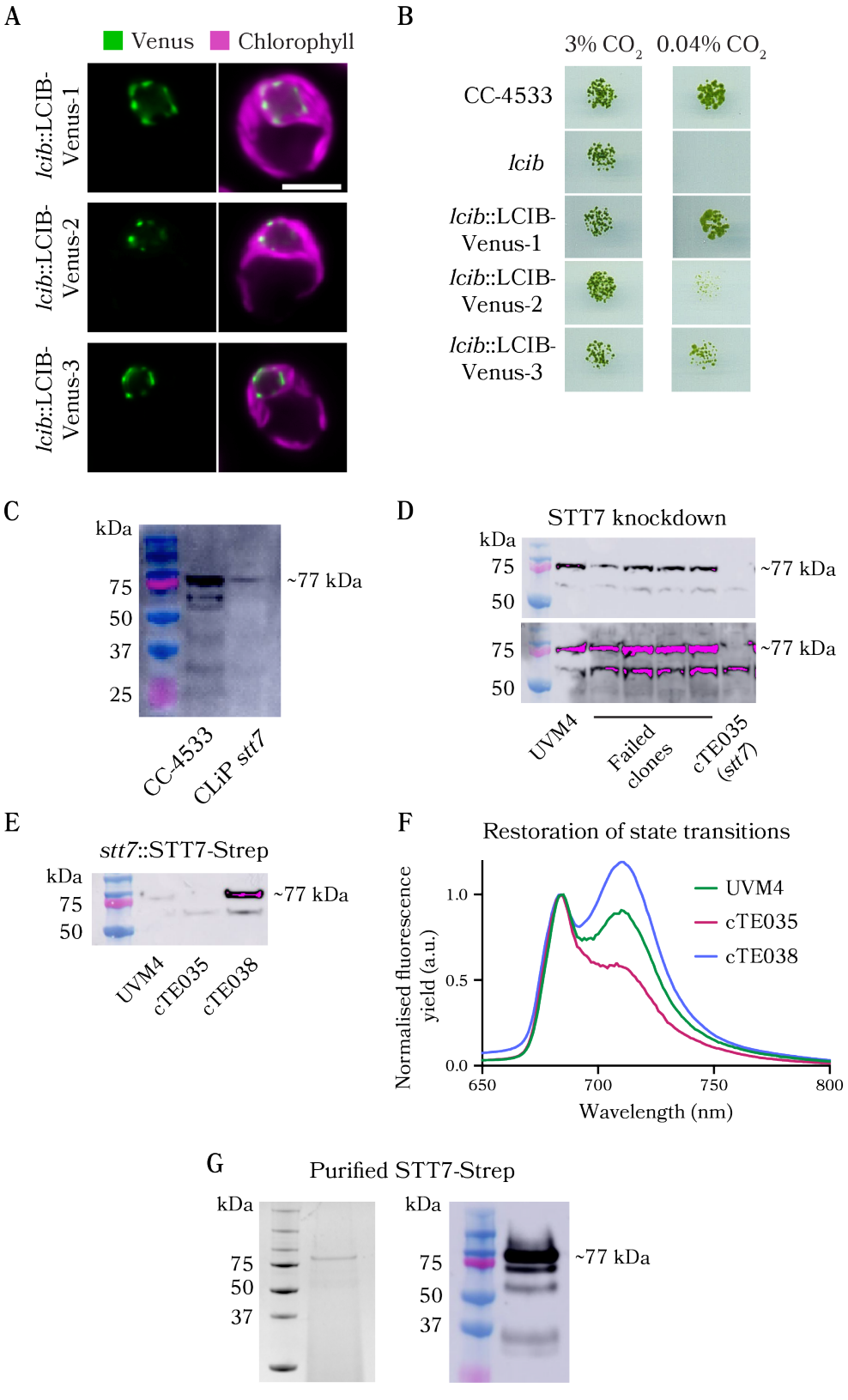

A
Venus
Chlorophyll
lcib::LCIB-Venus-1
lcib::LCIB-Venus-2
lcib::LCIB-Venus-3
B
3% CO2
0.04% CO2
CC-4533
lcib
lcib::LCIB-Venus-1
lcib::LCIB-Venus-2
lcib::LCIB-Venus-3
C
kDa
75
50
37
25
~77 kDa
CC-4533
CLiP stt7
D
STT7 knockdown
kDa
75
50
~77 kDa
75
50
~77 kDa
UVM4
Failed clones
cTE035 (stt7)
E
stt7::STT7-Strep
kDa
75
50
~77 kDa
UVM4
cTE035
cTE038
F
Restoration of state transitions
Normalised fluorescence yield (a.u.)
1.0
0.5
0.0
650
700
750
800
Wavelength (nm)
UVM4
cTE035
cTE038
G
Purified STT7-Strep
kDa
75
50
37
kDa
75
50
37
~77 kDa

Figure 3.5: **Recombineered genes rescue mutant phenotypes. A.** Fluorescence microscopy images of three independently transformed lines expressing recombineered LCIB in the CLiP library *lcib* mutant background. LCIB is a low-CO_2 inducible component of the CCM important for cell growth under limiting CO_2. Scale bar is 5 μm. **B.** Growth phenotypes of wild type, mutants and rescued LCIB lines grown under high and low CO_2, showing successful restoration of low-CO_2 growth the rescued strain *lcib*::LCIB-Venus-1. **C.** Western blot using the anti-STT7 antibody, indicating a low level of STT7 accumulation in a putative STT7-deficient mutant from the CLiP library. **D.** Western blots using the anti-STT7 antibody of five transformant lines in the UVM4 background resulting from a CRISPR-guided insertional mutagenesis experiment. Strain cTE035 shows no detectable accumulation of STT7 even when the blot is over-exposed. The additional band below ~77 kDa may be a non-specific target for the anti-STT7 antibody. **E.** Western blot showing very strong STT7 expression from strain cTE038, generated by transformation of cTE035 with recombineered STT7-Strep. **F.** Chlorophyll fluorescence spectra of anoxia-treated cells at 77 K normalised to the sharper peak at ~680 nm that is associated with chlorophyll fluorescence from PSII. Cells were subjected to oxygen depletion in the dark for 20 min, triggering a transition to state II (Buchert et al. 2020) prior to freezing. The broader peak at ~720 nm, associated with PSI fluorescence, is significantly depleted in cTE035 compared with UVM4 symptomatic of an inability to trigger state II. The rescued strain cTE038 displays a greatly increased PSI-associated fluorescence, likely due to overexpression of recombineered STT7-Strep. **G.** Coomassie-stained SDS gel showing purified STT7-Strep from cTE038, with a Western blot confirming that the band at ~77 kDa is STT7. Some fragmentation products of the purified protein are visible in the Western blot. James Barrett performed the LCIB complementation experiments in panels **A** and **B**.

3.4 Discussion

Our recombineering approach has enabled the efficient cloning of large and complex genes that could not be achieved via PCR-based cloning. It opens the door to a better understanding of the functional role of a large fraction of the Chlamydomonas genome though protein localization, protein-protein interaction studies, real-time monitoring of relative protein abundance and complementation of mutants (both random insertion mutants and CRISPR/Cas generated mutants). Our approach circumvents the challenges associated with cloning large, GC-rich and complex genes that are prevalent in Chlamydomonas. Indeed, the results from WindowMasker appear to confirm that Chlamydomonas has a particularly repeat-rich genome in comparison to other model organisms (Fig. 3.1A). While we recognise genome complexity can be assessed in multiple ways, we opted for simplicity in choosing to quantify repeat frequency using a single program. WindowMasker is particularly useful for comparisons between genomes as it detects repeats using only the genomes themselves (Morgulis et al. 2006), without referring to species-specific lists of known repetitive sequences, relied on by alternative programs such as RepeatMasker (Smit

et al. 2015).

Our cloning success rate was 77%, considerably higher than a previous PCR-based high-throughput cloning pipeline (48%) (Mackinder et al. 2017), which was itself inflated due to an enrichment of small target genes. Our comparisons to the success rate of PCR-based cloning are instructive but limited to ATG-stop cloning. PCR is an irreplaceable technique for cloning Chlamydomonas genes, and our method is intended to complement the capabilities of PCR-based cloning by sidestepping some of its major limitations (i.e., gene size, complexity, and unsuitable primers). For example, PCR was used for downstream editing of recombineered genes as described here for the adaptation of recombineered STT7 for high expression from a UVM4 background (see S Section 1.5.4 on page 36 in Chapter 1 for details). Our overall success rate is slightly lower when compared to recombineering pipelines in other organisms, with success rates of 89% achieved in *C. elegans* (Sarov et al. 2012) and 93% for Arabidopsis (Brumos et al. 2020). This reduced overall efficiency may be due to the complexity of the Chlamydomonas genome.

3.4.1 Possible reasons for recombineering failure

Restriction digest analysis of plasmids isolated from incorrectly assembled recombineering events suggested that cloning could fail due to a broad range of reasons. These include cassette recircularization, cassette duplication, cassette insertion into the BAC or retrieval of incorrect target regions. Increasing homology arm length, using alternative homology arms, and using fosmids are potential solutions to overcome incorrect recombineering for specific targets. However, the mechanisms contributing to the probability of failure for these reasons is unclear. Notably, batch-scale recombineering appears to show higher success rates than large-scale attempts. We expect a higher success rate when the pipeline is applied to small sample numbers since it is easier to optimise bacterial growth prior to transformation on a per-sample basis if there are fewer samples to manage. For example, since publication of the pipeline, we have successfully recombineered approximately fifty additional genes at batch-scale with a 100% success rate (Philipp Girr and Chun Sing Lau, personal communication).

Previous research has linked DNA secondary structural features to recombineering failure (Nelms and Labosky 2011). We reasoned that formation of strong hairpins or other secondary structures at or near the homology regions may disrupt lambda-red recombination. In Nelms and Labosky, 2011, the authors analysed a short region of DNA resistant to both PCR and recombineering using the secondary structure prediction program RNAfold (Lorenz et al. 2011), finding that the problematic sequence was flanked by two hairpin structures with high base-pairing probability. We therefore analysed our successful and unsuccessful targets using RNAfold at 37°C, the temperature of the recombination step in the pipeline (Fig. 3.6). However, to analyse multiple sequences in parallel we simply quantified the number of high-probability

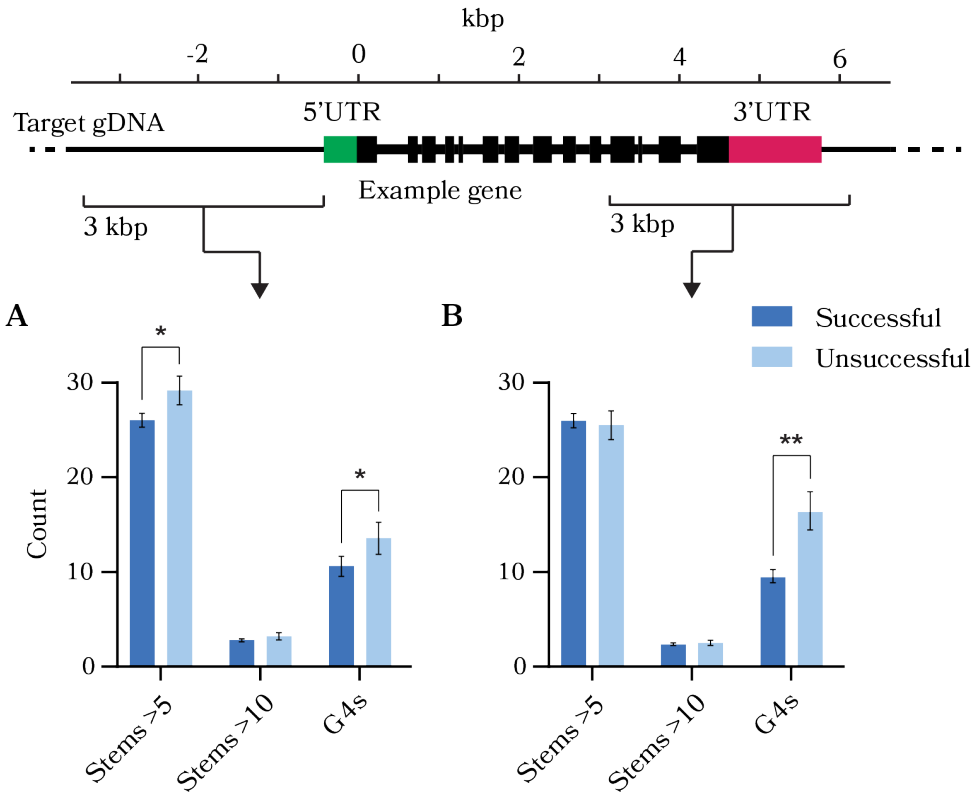


Figure 3.6: **Secondary structural features predicted by RNAfold differ between successful and unsuccessful recombineering targets. A.** RNAfold results from 3 kbp sequences surrounding the 5' end of each target region (centred ~2kbp upstream of the start codon). The average number of high-probability base pairs in unsuccessful targets involved in hairpin stems > 5 bp long (29.2) was significantly more than for successful targets ($t(188) = 2.01$, $p < 0.05$). The average number of predicted G4s in unsuccessful targets (10) was also significantly higher than for successful targets (Mann-Whitney $U = 2529$, $p < 0.05$, two-tailed). **B.** Results from querying 3 kbp sequences surrounding the 3' end of each target region (centred on the stop codon). The average number of predicted G4s in unsuccessful targets (12.5) was significantly higher than for successful targets (Mann-Whitney $U = 2332$, $p < 0.01$, two-tailed). Stems refer to the number of high-probability base pairs detected that were part of a run of > 5 or > 10 consecutive bases, indicating a hairpin stem. G4s are the number of g-quadruplexes predicted in the queried sequences by RNAfold. Plotted bars show means and SEM.

base pairs involved in consecutive runs of five or more throughout our target sequences. Many of our recombineering targets were longer than the maximum allowed length for analysis, so we split our datasets into the 3 kbp surrounding the 5' end of each target region (Fig. 3.6A) and the 3 kbp surrounding the stop codon (Fig. 3.6B). RNAfold also predicts the formation of g-quadruplexes (G4s) in ssDNA, which form among clusters of guanines by non-canonical base-pairing. G4s are thought to be resistant to polymerases (Jamroskovic et al. 2019), but we reasoned they may also disrupt recombineering. Fig. 3.6 shows the results of the analysis, revealing small but significant increases in the number of G4s and in the number of consecutive runs of five or more high probability base-pairs in our unsuccessful targets. Statistical values are reported in the figure legend. These two parameters alone may not fully explain why some of our targets failed, particularly since it has not yet been shown mechanistically how they could interrupt recombination. However, these data indicate that recombineering may not be truly independent of sequence complexity, and that secondary structure ought to be considered for difficult targets.

3.4.2 Benefits of including the native promoter

Due to the good tolerance of our method of gene size (Fig. 3.3A), we could include the native promoter of target genes. Harnessing this more native-like expression from recombineered genes, our localisation results were in good agreement with previous work, or even provided the first localisation data for a previously cloned gene in the case of ISA1 (Toyokawa et al. 2020). Interestingly, using a native promoter allowed relative protein abundance for LCIB to be tracked between conditions (Fig. 3.4D). Once validated, acquiring relative abundance data was straightforward and could be easily parallelised for multiple targets. In this way, our pipeline provides a highly complementary method to the recently released CLiP mutant collection (Li et al. 2019). Future experiments could include tracking relative protein abundance in 96-well libraries of tagged proteins in response to a perturbation (e.g., high to low CO_2 or O_2 transition). This could provide novel insights into cellular processes involved in the CCM and CET, aiding in our understanding of how bioenergetic processes are interconnected (Burlacot et al. 2022).

3.4.3 Limitations of our pipeline

Although our relative abundance data for LCIB appeared to closely reflect immunoblotting data, it should be noted that using a native promoter will not fully mimic native expression for several reasons. Firstly, this discrepancy can be due to insertional effects, either caused by integration into transcriptionally unfavourable regions of the genome that lack native *trans* regulatory features or by transcriptional silencing (Schroda 2019). At a protein level, fluorescent protein folding time could affect protein stability and turnover and the presence of the fused fluorescence protein

could affect function or multi-subunit assembly.

The largest contributing factor to non-native expression is likely the absence of 3' *cis* regulatory elements in our recombineered products. Whilst our approach allows the native promoter, 5'UTR region and open reading frame to be cloned, the native 3'UTR is not maintained. This could be addressed through a two-step recombineering pipeline where the tag is first inserted into the BAC at the desired location, markers could then be removed via a Flp-FRT recombinase system (Sarov et al. 2006, Brumos et al. 2020), and the edited target gene can then be retrieved into a final Chlamydomonas expression vector. When establishing our pipeline, we decided not to pursue this strategy in order to maximise the success rate by limiting the number of steps, with a focus on developing a simple, easy-to-apply approach. In addition, whilst we have focused on C-terminal tagging as this allows conservation of N-terminal transit peptides required for organelle targeting, our recombineering pipeline could be applied for N-terminal tagging by modification of our cloning vectors with a constitutive promoter and N-terminal tag.

Recombineering relies on the availability of a BAC library or similar, and another limitation we encountered was that only 86% of nuclear genes are covered by the BAC library. However, this value only takes into account ~73% of BACs, with the remaining BACs potentially incorrectly mapped to the current version of the Chlamydomonas genome. Our analysis suggests the true percentage of genes covered could be higher than 86% but confirming this may require a careful re-mapping of the library. A promising solution is cloning from fosmids, which are smaller DNA fragments compared to BACs, possibly helping to improve efficiency by reducing off-target recombination. In addition, the fosmid library is expected to have close to 100% genome coverage. We demonstrated that our pipeline can be successfully applied for cloning from fosmids using SAGA1, STT7 and CHLG. A Chlamydomonas fosmid library is now available, having been released in July 2020 from the Chlamydomonas Resource Center.

Finally, though our pipeline has proven useful for cloning genes for re-introduction into Chlamydomonas by nuclear transformation, future developments in CRISPR-based technologies may soon make recombineering obsolete. If large insertions can be made into the genome at specific loci with high efficiency, researchers may be able to tag any gene with any tag, achieving close-to-native expression and regulation of modified genes. Publication of our method has nonetheless provided a useful resource for the Chlamydomonas research community, and genes recombineered with our method are in use in several labs, including our collaborators in Münster, and have already been included in two publications (Mukherjee et al. 2019, Meyer et al. 2020).

Chapter 4

Tethering FNR to PSI

4.1 Summary

A tagging cassette was designed and constructed to express ferredoxin-$NADP^+$ reductase (FNR) tethered by its N-terminus to the stromal face of the F subunit of PSI (PsaF). The cassette was co-transformed with a gRNA-Cas9 ribonucleoprotein targeting exon 3 of PSAF, resulting in a strain in which every copy of PSAF was tagged with FNR. One of these lines was transformed again with a second ribonucleoprotein targeting exon 6 of FNR, resulting in strains that lack accumulation of native FNR but still express a PSI-tethered FNR. The growth phenotypes and thylakoid proteomes of two of these double-mutant lines were compared with the wild type, revealing a retarded growth rate and possible restructuring of PSI in response to the mutations.

4.2 Background

CET is an important regulatory feature of plant and algal photosynthesis. Our current knowledge of CET suggests it plays a dual role in Chlamydomonas, supplementing LET to balance the ATP/NADPH cycle while also providing a means to quickly increase ΔpH to trigger the important photoprotective processes of qE and photosynthetic control. CET effector proteins also interact with the state transition kinase STT7, implicated in CET upregulation under light limited conditions to drive the extra production of ATP during a transition to state II. Though the mechanism of electron transfer and the mechanism of upregulation under ATP-depleted conditions are still mysterious, the severe phenotypes displayed by mutants with a deficiency in CET effector proteins hints at its importance. In Chlamydomonas under anoxia, mutants deficient in PGR5 show light sensitivity, a PSI acceptor side limitation phenotype but also higher H_2 production via HYDA due to less competition from CET. In Arabidopsis, mutants deficient in PGR5 cannot trigger qE or photosynthetic control, leading to severe ROS-induced PSI acceptor side limitation and a reduced growth rate under fluctuating light.

CET clearly has a significant influence on cell bioenergetics. A deeper understanding of the mechanistic details of CET could be advantageous to crop researchers looking to improve yields of important species by tweaking photosynthesis. Since PGRL1 was recently excluded as the FQR in Arabidopsis (Rühle et al. 2021), the leading theory for the mechanism of CET is a return of electrons from Fd^- to PQ within cytb_6f via haem c_i, mediated by FNR, PGR5, PGRL1 and possibly STT7 (Buchert et al. 2020). The role of FNR is particularly interesting; FNR is currently thought to hold a pivotal role in supplying CET with reducing power by catalysing the interconversion of Fd^- and NADPH (Shahak et al. 1981). FNR can dynamically localise to its thylakoid-associated binding partners, Tic62 and TROL, in plants (Lintala et al. 2014), a process that may be light controlled (Kramer et al. 2021). FNR was also identified as part of the putative CET-supercomplex, which was reported to form under state II-inducing conditions (Iwai et al. 2010). In Arabidopsis, tethered FNR (to either Tic62 or TROL) was found to be crucial for high CET rates (Kramer et al. 2021) but equivalent tethering sites are yet to be identified in Chlamydomonas. Whether the CET supercomplex is obligatory for CET also remains unclear. In this case, supporting evidence for supercomplex-driven CET comes from mutants of LHCA2 and LHCA9 that show higher CET (Steinbeck et al. 2018). LHCA2/9 are peripheral PSI subunits thought to obstruct formation of the supercomplex when present, suggesting that some form of complex involving PSI is involved in enhanced CET. Could this involve a close association between cytb_6f and PSI, with FNR/Fd catalysing the transfer of electrons between them? Recent spectroscopic evidence suggests a functional supercomplex does indeed assemble under hypoxia to upregulate CET (Joliot et al. 2022). However, CET measurements were performed with an artificial stromal electron donor, methyl viologen (see Section 1.2.2.11 on page 15), and it is unclear whether FNR was involved in the methyl viologen-mediated FQR reaction.

Investigating the involvement of FNR in CET is hampered by our limited ability to perform reverse genetics to elucidate the role of FNR in CET. FNR is essential to LET, and LET is essential to organismal viability in most of the green lineage. Partial depletion of FNR has been achieved in tobacco and in Arabidopsis, with clear problems with photoprotection observed in the resulting plants in both studies (Lintala et al. 2012, Palatnik et al. 2003). The coming of age of genetic tools for Chlamydomonas represents an opportunity to advance our understanding of FNR by disentangling its roles in LET and CET using genetic engineering. This task is made easier due to the presence of only one known isoform of FNR in Chlamydomonas, while at least two are usually present in higher plants (Kramer et al. 2021), and three in maize (Goss and Hanke 2014). Unique amongst oxygenic photosynthetic model organisms, the facultative heterotrophy of Chlamydomonas (reviewed in Yang et al. 2015) may enable the deletion of FNR from wild type cells if grown in acetate-containing TAP media (tris-acetate-phosphate). Analysis of the essential proteins for mixotrophic growth (acetate plus light) conditions identified Chlamydomonas FNR

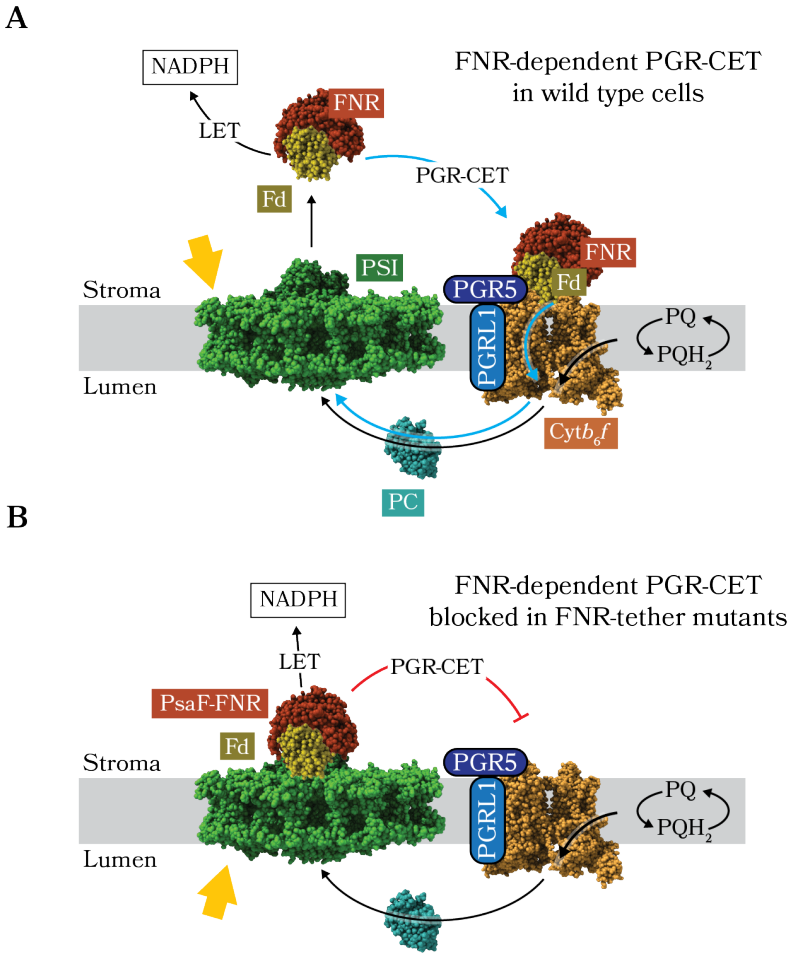


Figure 4.1: **Hypothetical model of altered photosynthetic electron flow in the FNR-tether mutants. A.** Schematic of photosynthetic electron flow in wild type Chlamydomonas thylakoids based on the theory that electrons re-enter the PQ pool at cytb_6f using an FNR-dependent, Fd-assisted Q-cycle (Buchert et al. 2020). **B.** We hypothesise that tethering FNR to the PSI acceptor side will permit LET as usual but block FNR-dependent PGR-CET, resulting in measurably less CET. Protein structural models are the same as those shown in Figure 1.2 on page 7.

as non-essential, or at least substitutable (Chapman et al. 2015). In the same study, CET was seen to be upregulated under mixotrophic conditions. Depending on the involvement of FNR in CET, therefore, mutations involving FNR may still produce a noticeable phenotype in acetate-containing media. To this end, we set out to design an experiment that would inhibit direct involvement of FNR in CET while maintaining its ability to participate in LET. This might be achieved by tethering FNR to the stromal face of PSI using a peptide linker. FNR tethered directly to PSI at the PSI acceptor side may still function to catalyse the reduction of $NADP^+$ by Fd but would be artificially prevented from freely associating with cytb_6f or other thylakoid components (Fig. 4.1).

4.3 Results

4.3.1 Design of a DNA construct for tethering FNR to PSI

Tethering FNR to PSI was accomplished by design of a DNA construct that could drive expression of a chimaeric fusion protein consisting of the PSI subunit PsaF and the FNR peptide connected by a short, flexible peptide linker (NGTGGSGGR). Like FNR, PsaF is encoded in the nucleus, and was identified as a suitable PSI subunit due to its proximity to the PsaA/B-coordinated PSI acceptor side. The PsaF peptide forms a transmembrane helix *in vivo* with its N-terminus in the lumen and C-terminus in the stroma, enabling tethering of FNR to the PsaF C-terminus. This orientation avoided any steric hindrance that might result from tethering a fusion partner to the FNR C-terminus, which is positioned close to the active site (Tejero et al. 2005). A serine-alanine linker and Strep-tag® II (WSHPQFEK) were added to the FNR C-terminus for downstream purification, chosen due to the tag's small size and comparable inertness (Kimple et al. 2013).

The PsaF-FNR construct design was based on the pLM005 vector family (Mackinder et al. 2017), which drives the expression of a gene of interest using the promoter and terminator from PSI subunit PsaD, and is prepended by the AphVIII gene, bestowing resistance to paromomycin. Construction first involved insertion of the ~1,200 bp native PsaF sequence into pLM005 by Gibson assembly, amplified from gDNA using PCR. Next, the FNR nucleotide sequence was analysed against a chloroplast protein database to estimate the length of the chloroplast transit peptide. The ~2,200 bp native FNR sequence (minus the transit peptide-encoding sequence) was amplified from gDNA using primers that included the codon-optimised linker peptide sequence and Strep-tag® II nucleotide sequence, then added into the vector in frame with the PsaF sequence by restriction-ligation cloning.

Lastly, a series of silent mutations were made by overlap extension PCR to remove Cas9 recognition sites in PsaF exon 3 and FNR exon 6. Suitable Cas9 cleavage sites were identified using online databases (see Section 2.3.3 in Chapter 2 on page 43)

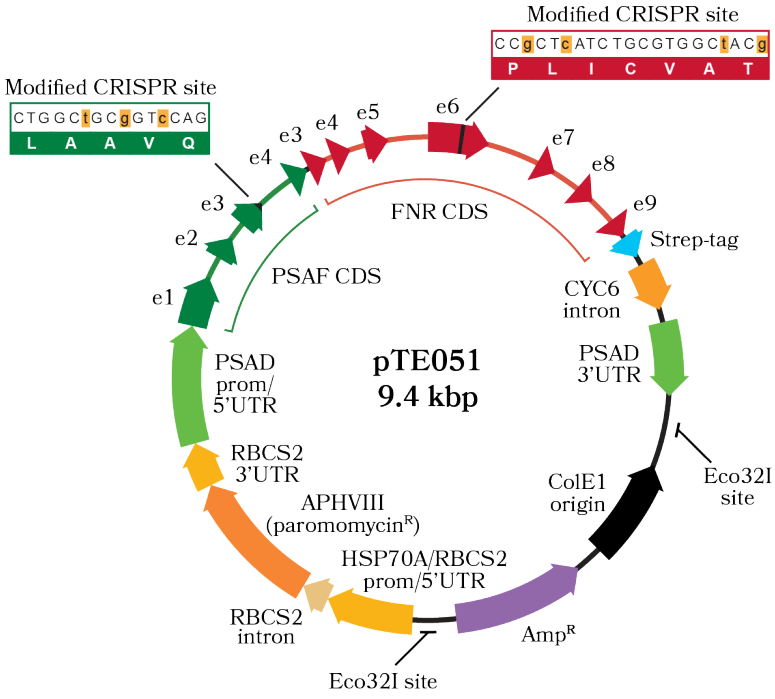


Figure 4.2: **Plasmid pTE051 encoding the AphVIII and PSAF-FNR-Strep expression cassette.** The double-cutter Eco32I was used to prepare the plasmid for transformation into UVM4. PSAF and FNR contain silent mutations at exon 3 and exon 6, respectively, to avoid recognition by Cas9. FNR is truncated at the 5' end to remove the predicted chloroplast transit peptide.

and chosen due to low predicted off-target cleavage sites. Three silent mutations were made to the PsaF exon 3 recognitions site, including removal of the PAM sequence. Codon restrictions around the FNR exon 6 recognition site meant the PAM sequence could not be modified, so four additional silent mutations were made throughout the remainder of the recognition site to compensate. The final vector, named pTE051 (Fig. 4.2), was sequenced to confirm correct assembly, then prepared for transformation by digestion with a blunt cutter. The resulting excised cassette was ~7,000 bp long and encoded the AphVIII cassette from pLM005 and the PsaF-FNR-Strep fusion protein driven by the PsaD promoter and terminator.

4.3.2 CRISPR-mediated one-step deletion/insertion with the PsaF-FNR cassette

We designed a cloning strategy to generate mutants deficient in native PsaF but expressing PsaF-FNR in a single transformation step. We achieved this by co-transformation with a sgRNA-Cas9 RNP targeting PsaF exon 3 and the pTE051-derived expression cassette. Ideally, NHEJ repair of the CRISPR-induced double stranded break would then incorporate the co-transformed DNA at the cut-site. Alternatively, alternate fragments of exogenous or endogenous DNA would be incorporated at the cut site disrupting the locus, while the expression cassette would be integrated into the genome elsewhere.

sgRNA corresponding to the PsaF exon 3 recognition site was complexed with Cas9 *in vitro* to produce RNP. The high-expression strain UVM4 was chosen for transformation as it harbours a mutation impairing epigenetic gene silencing (Neupert et al. 2020). UVM4 was co-transformed by electroporation with 61 pmol of RNP and 500 ng of expression cassette then selected on TAP + 1.5% agar plates containing 20 μg ml^{-1} paromomycin. After two weeks of recovery, 193 colonies were picked and their gDNA extracted for PCR screening. Disruption of the native PsaF locus was assessed first. Using primers that amplified across the native locus, 88 colonies failed to produce an amplicon at all, suggesting either PCR failure or the presence of a large insert (Fig. 4.3A). These were then assessed for the presence of the expression cassette. Primers were designed to amplify a ~1,000 bp sequence from the 3' end of the PsaF coding sequence to FNR exon 6. 72 colonies failed to amplify, but 16 colonies produced the correct amplicon indicating the presence of the central section of the PsaF-FNR cassette in the genome (Fig. 4.3A, (ii)). These 16 colonies were grown in TAP media for a week to achieve enough cells for Western blotting. Colonies were then blotted with an anti-FNR primary antibody; all colonies showed a clear band at ~35 kDa indicating the native FNR peptide, with eight colonies also showing a distinct band at ~53 kDa indicating expression of the PsaF-FNR fusion. Three colonies denoted as NT4, NT7 and NT12 (for 'native/tethered') were taken forward for Western blotting with anti-PsaF antibody; NT4 and NT12 showed no

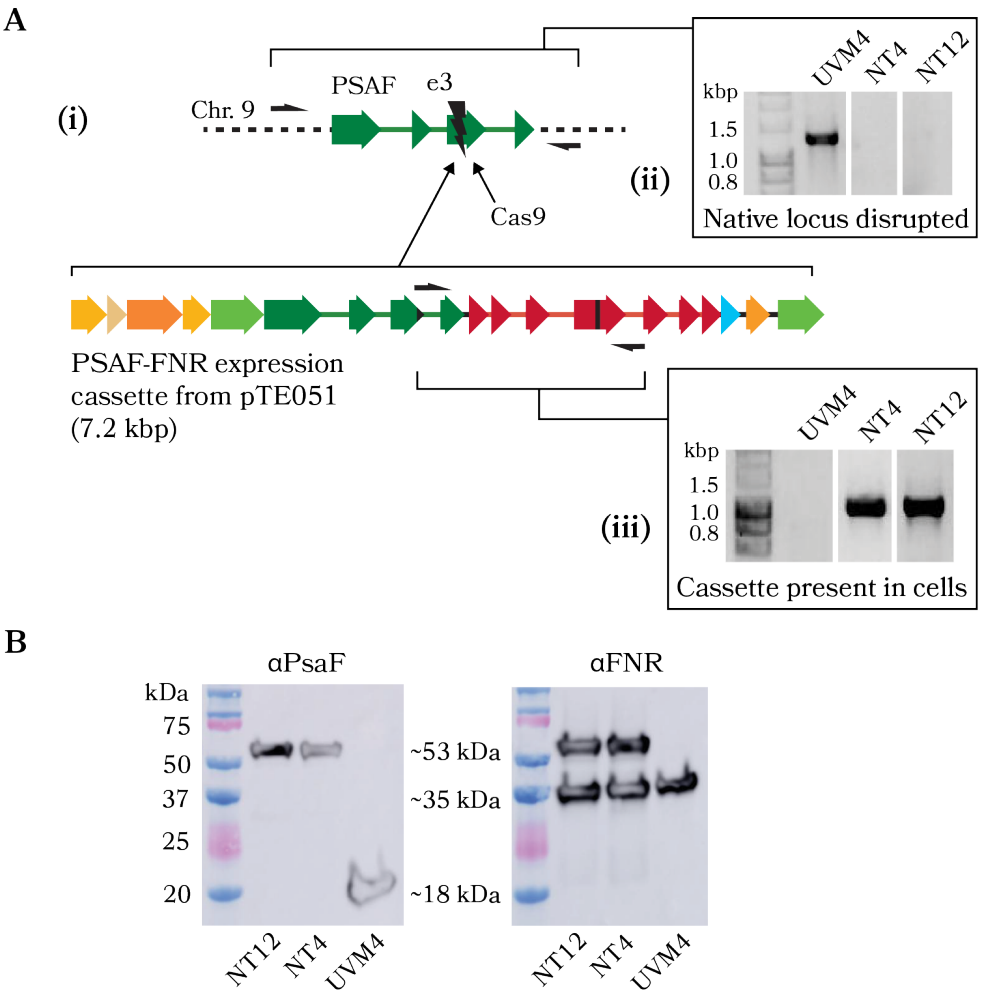


Figure 4.3: **Transformation of UVM4 with a PSAF-FNR expression cassette. A.** Results of the co-transformation of UVM4 with the PsaF-FNR expression cassette and RNP targeting PSAF exon 3. (i) shows the intended outcome of CRISPR-guided insertional mutagenesis, in which the PsaF-FNR expression cassette is inserted into native PSAF exon 3 as part of the NHEJ repair process. (ii) shows colony PCRs with NT4 and NT12. Absence of an amplicon across the PSAF locus indicated a large insertion at exon 3. (iii) shows successful amplification using primers specific to the PsaF-FNR cassette in lines NT4 and NT12, which indicated successful integration of the cassette into the genome. **B.** Western blots with anti-PsaF confirmed deletion of native PsaF in NT4 and NT12. Blots with anti-FNR confirmed expression of the PsaF-FNR peptide along with native FNR.

band at ~18 kDa but a clear band at ~53 kDa, indicating deletion of the native PsaF (Fig. 4.3B).

4.3.3 Deleting FNR

FNR is an essential protein for phototrophic growth but is thought to be non-essential during mixotrophic growth in acetate (Chapman et al. 2015). To investigate the essential nature of FNR we first tried to obtain mutants deficient in FNR alone. We obtained a mutant from the Chlamydomonas collection in Minnesota, USA (https://www.chlamycollection.org/, accessed Sep. 2022), LMJ.RY0402.101504, that contains a putative insertion in the 3'UTR of Cre11.g476750, encoding FNR. However, Western blotting of LMJ.RY0402.101504 using the anti-FNR primary antibody showed a clear signal at ~35 kDa, indicating this mutation does not deplete the expression level of FNR (Fig. 4.4A). Next, we tried making our own FNR-deficient mutant by targeted insertional mutagenesis. UVM4 cells were co-transformed with a sgRNA-Cas9 RNP targeting FNR exon 6 and the AphVII expression cassette excised from plasmid pLM006 (Mackinder et al. 2017), conferring resistance to hygromycin. Transformants were recovered on hygromycin in the dark to pre-empt any light-sensitivity caused by loss of FNR. 80 colonies were obtained and assessed by PCR, but all 80 produced an amplicon of the correct size across the native locus suggesting no insertions or deletions were present (result not shown). PCR products from nine colonies were sequenced to check for small frameshifts or indels but the FNR locus was unaltered. Strains were not tested by Western blotting due to time constraints, but the uniform, healthy appearance of the 80 colonies tentatively suggests that they all still express functional FNR.

We then applied the same method to colony NT12, expressing the PsaF-FNR fusion. TAP-grown NT12 cells were co-transformed with AphVII cassette and RNP targeting FNR exon 6, then recovered on 20 μg ml^{-1} paromomycin and 25 μg ml^{-1} hygromycin. 56 colonies were obtained and assessed by PCR using three primers in a single reaction, two that amplified outwards from either end of the AphVII cassette and one that bound in FNR exon 6 facing the CRISPR recognition site. Seven colonies produced amplicons of ~400 bp, indicating successful insertion of the AphVII cassette in either orientation within FNR exon 6 (Fig. 4.4B). These colonies were grown in TAP media in the dark and Western blotted using anti-FNR and anti-PsaF antibodies. Six colonies showed the correct banding pattern: an absence of native FNR, an absence of native PsaF, and a clear band at ~53 kDa with both antibodies (Fig. 4.4C). A faint ladder of bands around the size of native FNR could be seen in the Western blots using anti-FNR, possibly indicating bleed-through expression of native FNR, or a very low level of cleavage of PSAF-FNR. No evidence could be seen for native PsaF expression from whole cell blots. The six colonies were named T7, T10, T34, T37, T43 and T44 (for 'tethered' only). T7 and T10 were taken forward for isolation of PSI.

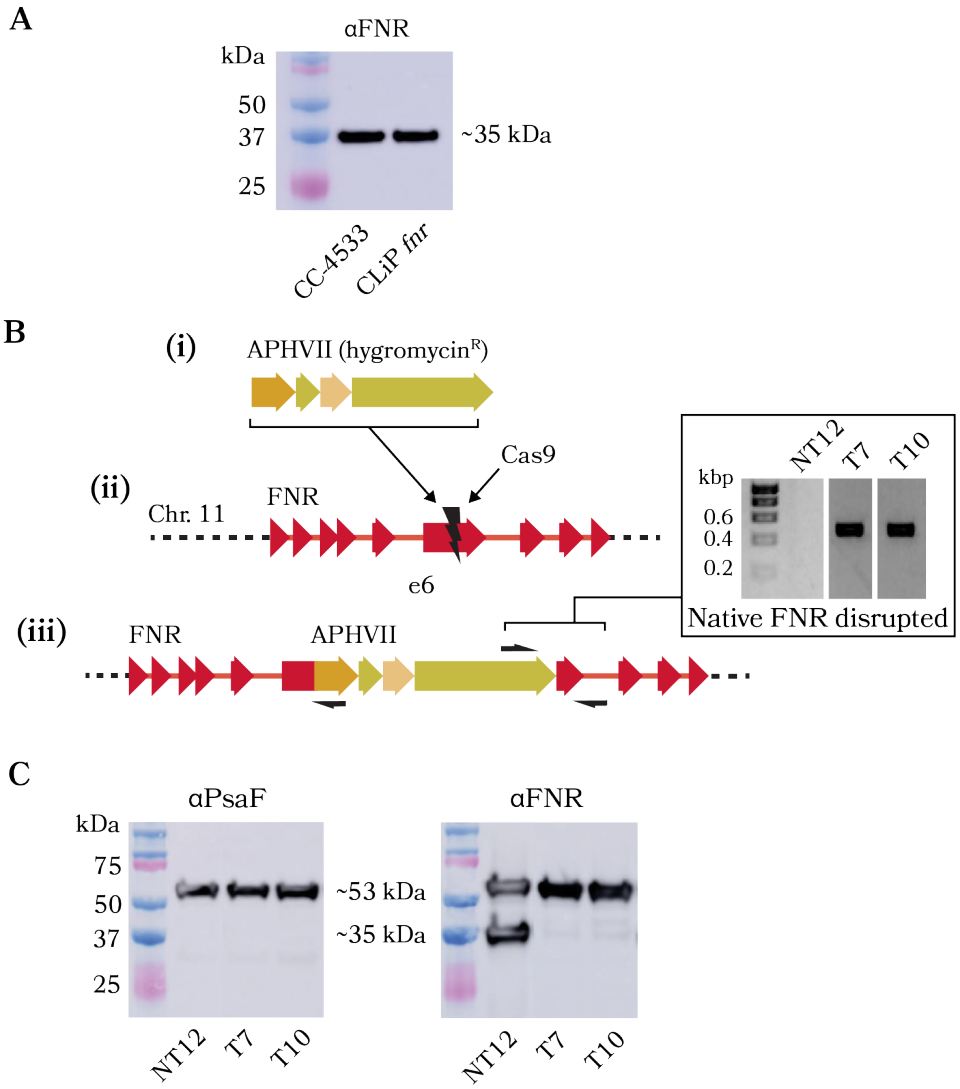


Figure 4.4: **Deletion of FNR. A.** Western blot using anti-FNR, demonstrating that CliP mutant LMJ.RY0402.101504 is not deficient in FNR. **B.** Results of the CRISPR-mediated deletion of native FNR from NT12. (i) shows the AphVII cassette used for transformation, encoding resistance to hygromycin. The intention was to select for insertion of AphVII into native FNR exon 6 (ii) as part of the NHEJ repair process. (iii) shows the intended result, along with evidence of successful PCR amplification across the boundary between native FNR and AphVII using a three-primer PCR that allowed for insertion of AphVII in either orientation. Amplicons in T7 and T10 are strong evidence of correct insertion. **C.** T7 and T10 Western blots using anti-PsaF and anti-FNR indicated deletion or thorough depletion of native FNR, showed the absence of native PsaF and the presence of PsaF-FNR.

4.3.4 Isolation of PSI-FNR

Strains UVM4, T7 and T10 were grown in TAP for purification of PSI by anion exchange and sucrose gradient separation (see Section 2.5.2 in Chapter 2 on page 50). UVM4 cells were grown under ~60 μE m^{-2} s^{-1} constant light while T7 and T10 were grown in the dark to avoid possible light sensitivity. Thylakoids were solubilised in βDDM; the resulting sucrose gradients showed distinct, high-molecular weight bands (Fig. 4.5A); banding patterns were the same for T7 and T10, but different in UVM4 likely owing to the different light conditions used to grow the cells. Absorption measurements of the lower bands from each gradient gave a distinct PSI-like spectrum (Fig. 4.5A). These bands were buffer exchanged out of sucrose and concentrated to 0.5 mg ml^{-1} chlorophyll then prepared for SDS-PAGE. SDS gels were stained with Coomassie dye, revealing a PSI-like banding pattern (Fig. 4.5B) including the distinct double-bands at ~50 kDa resulting from the reaction centre core subunits PsaA/B. Unfortunately, the strong staining of PsaA/B occluded any band at ~53 kDa that might indicate the presence of the PsaF-FNR fusion peptide in PSI complexes from T7 and T10. Instead, the presence of PsaF-FNR in the bands was indicated by Western blotting with anti-PsaF, anti-FNR and anti-Strep antibodies. These showed clear bands at ~53 kDa for strains T7 and T10, while UVM4 only showed a band at ~18 kDa with anti-PsaF (Fig. 4.5C).

PSI samples were next prepared for single particle electron microscopy. Samples were subjected to gel filtration and eluted in a mild HEPES buffer. Eluates were concentrated again and applied to microscopy grids, but unfortunately no suitable grids could be produced.

4.3.5 Growth phenotype testing

To determine the ability of the FNR-tether mutants to photosynthesise, we assessed growth rate in TP media, which lacks a carbon source. Liquid growth tests were performed with UVM4, NT12, T7 and T10 in TP media with gentle shaking under different light regimes (Fig. 4.6). To assess potential variation between transformants, we included two additional clones of NT12 (NT4 and NT7) and four additional clones of T7 and T10 (T34, T37, T43 and T44). Liquid growth tests under ~60 μE m^{-2} s^{-1} constant light in TP over the course of two weeks clearly indicated the mutants grew more slowly than UVM4 (Fig. 4.6A) but could still reach a relatively high cell density. There were no clear differences in mean growth rate between native/tethered mutants and tethered-only mutants. Clones from the same transformation did show some variation in growth patterns but these were consistent with the general slow-growth phenotype seen in NT12, T7 and T10. Similar results were obtained with just UVM4, T7 and T10 under 14 h:10 h day/night cycling (Fig. 4.6B). In the absence of evidence for *in vitro* PSI-FNR activity, the ability of T7 and T10 to grow in TP without native FNR is consistent with the suggestion that

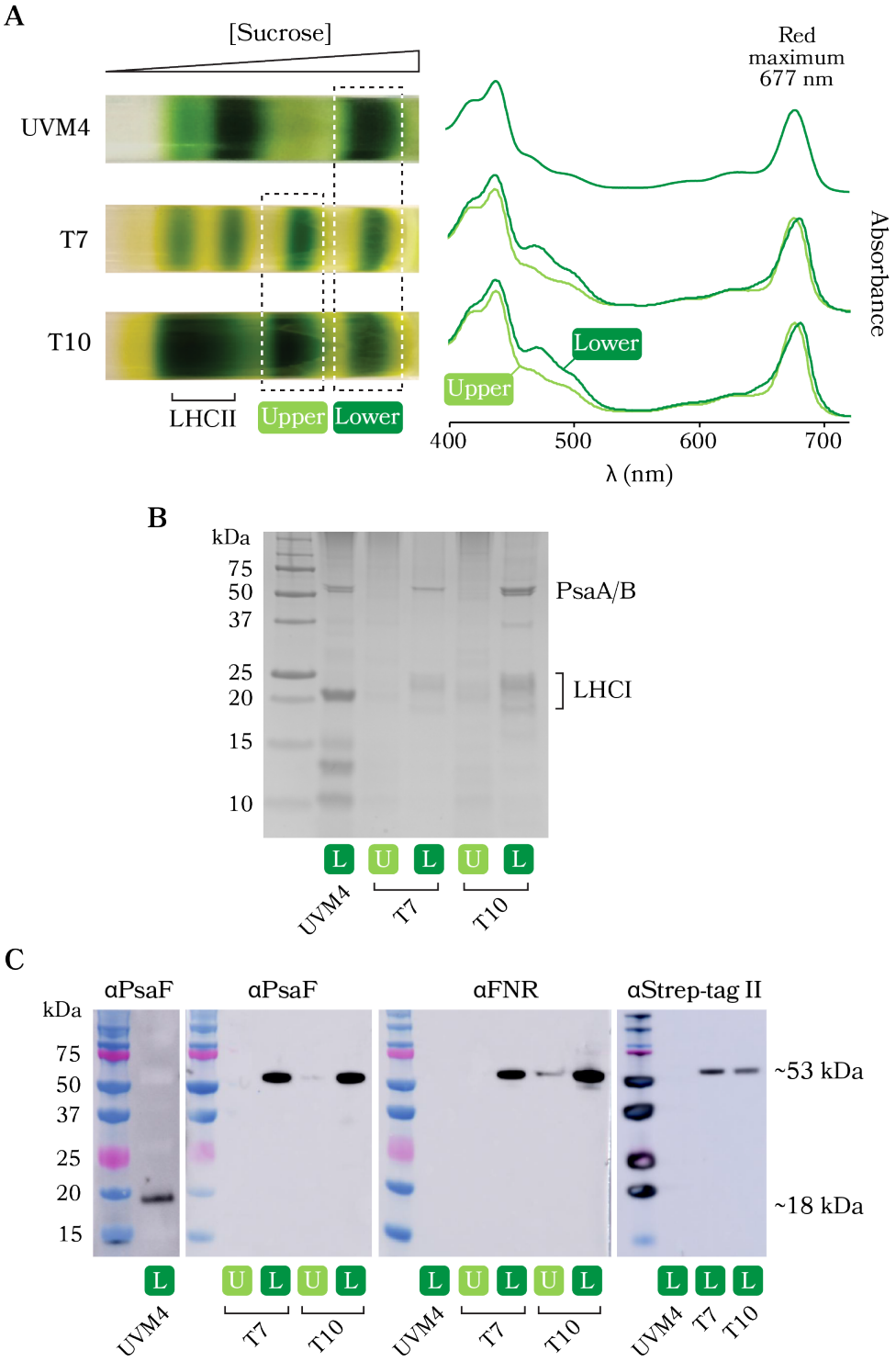

A
[Sucrose]
UVM4
T7
T10
LHCII
Upper
Lower
Red
maximum
677 nm
Absorbance
Lower
Upper
400
500
600
700
λ (nm)
B
kDa
75
50
37
25
20
15
10
PsaA/B
LHCI
L
U
L
U
L
UVM4
T7
T10
C
αPsaF
αPsaF
αFNR
αStrep-tag II
kDa
75
50
37
25
20
15
~53 kDa
~18 kDa
L
U
L
U
L
UVM4
T7
T10
L
U
L
U
L
UVM4
T7
T10
L
L
L
UVM4
T7
T10

Figure 4.5: **Purification of PSI from UVM4, T7 and T10. A.** Sucrose gradients and spectra indicate successful isolation of PSI-enriched bands from thylakoids of UVM4, T7 and T10. T7 and T10 cells were grown in the dark, likely pushing cells into a different state than UVM4, which was grown at ~60 μE m^{-2} s^{-1} constant light. This difference in state may explain the different banding patterns of PSI and PSII supercomplexes in T7 and T10 sucrose gradients. Spectra from the specific bands (upper, U; lower, L) are shown, indicating a strong enrichment in PSI. **B.** SDS gel of purification of bands from the sucrose gradients in panel A, showing clear PsaA/B banding pattern indicative of concentrated PSI in the lower (L) bands of T7 and T10. **C.** Western blots of the sucrose gradient bands in panel A using anti-PsaF, anti-FNR and anti-Strep. Presence of PsaF-FNR peptide in PSI particles from T7 and T10 is observed, but not from UVM4.

FNR-Strep is active in LET.

Next the effects of fluctuating light on growth rate were assessed. Cells were grown in TP for over two weeks under constantly cycling illumination; 25 s of low light (~24 μE m^{-2} s^{-1}) followed by 5 s of high light (~240 μE m^{-2} s^{-1}), with an average of ~60 μE m^{-2} s^{-1} over the full cycle. This extreme light regime retarded growth rate severely in all strains (Fig. 4.6C), though UVM4 nonetheless reached high cell density after 15 days. NT4, NT7 and NT12 showed extremely slow growth, implying an impaired ability to induce photoprotective mechanisms similar to the phenotype of PGR5-deficient cells (Yamamoto and Shikanai 2019). However, T7, T10, T34 and T37 showed no growth at all or even a loss of cells after two weeks. The small but significant difference in growth between NT12 and T7/T10 under fluctuating light may reveal specific sensitivity to photooxidative damage caused by the loss of native FNR.

4.3.6 Changes to the thylakoid proteome

We reasoned that changes in photosynthetic electron flow caused by a lack of native, diffusible FNR in T7 and T10 might trigger long-term expression changes in the thylakoid proteome. We set out to quantify and compare thylakoid proteins between UVM4, NT12, T7 and T10. Thylakoids from each strain were isolated from TAP-grown cells by sonication and suspended in Bis-Tris buffer. The protein concentration was assayed and equalised between samples before preparation for mass spectrometry. Membranes were processed for nanoLC-MS and analysed in triplicate for each sample according to Flannery et al. (2021). The data were queried against the Chlamydomonas reference proteome (see Section 2.9 on page 58 in Chapter 2) and a total of 4,030 Chlamydomonas proteins were identified across all four samples. Protein intensity (max LFQ abundance score) was estimated by label-free quantification and normalised by the intra-analysis sum of all detected protein intensities from each run. After checking intensity values from each replicate for

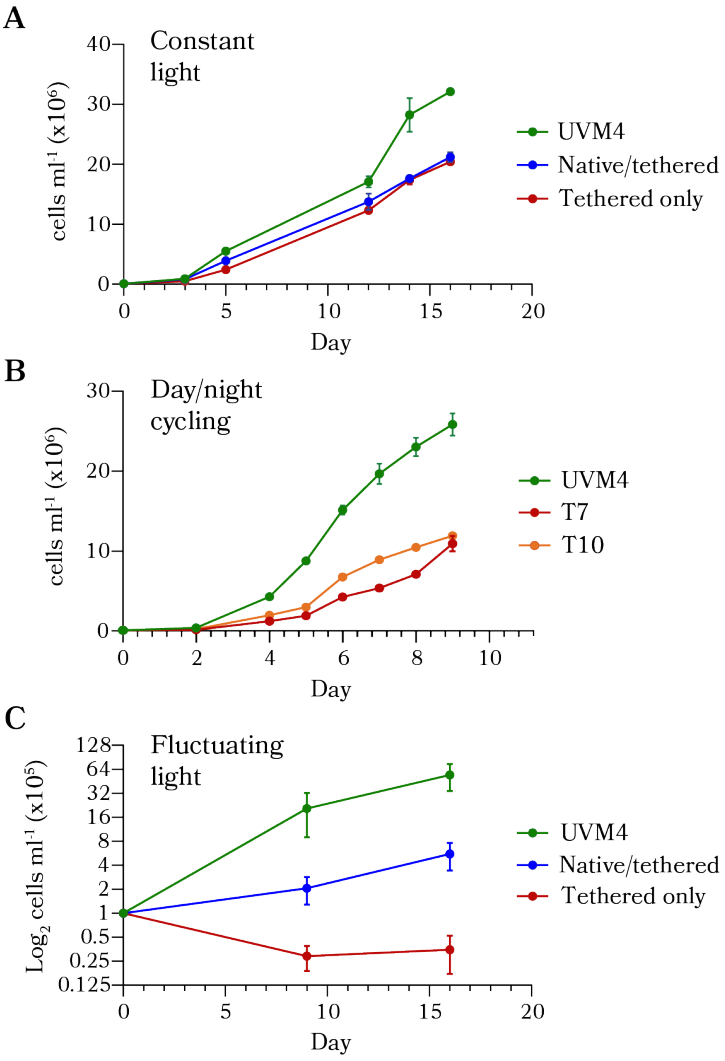


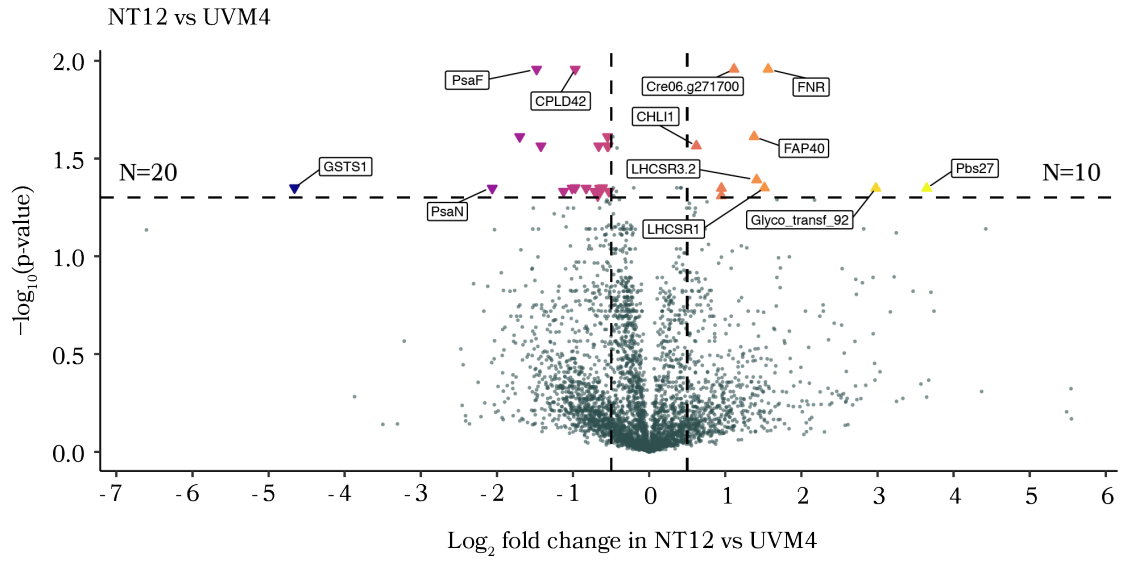
Figure 4.6: **Phototrophic growth curves for the FNR-tether mutants under different light regimes. A.** Growth in 3 ml TP under ~60 μE m^{-2} s^{-1} constant light, shaking at 120 rpm. Data for native/tethered FNR-containing strains are averaged across strains NT4, NT7 and NT12. Data for tethered-only strains are averaged across strains T7, T10, T34, T37, T43 and T44. **B.** Growth of UVM4, T7 and T10 in 3 ml TP under ~60 μE m^{-2} s^{-1} 14 h:10 h light/dark cycling, shaking at 120 rpm. **C.** Growth in 1 ml TP under fluctuating light, shaking at 180 rpm. The y-axis is scaled using $\log_2$ cell count to highlight the difference between the native/tethered strains and the tethered-only strains. The light regime fluctuated constantly between ~24 μE m^{-2} s^{-1} for 25 s and ~240 μE m^{-2} s^{-1} for 5 s. Data for native/tethered strains are averaged across strains NT4, NT7 and NT12. Data for tethered-only strains are averaged across strains T7, T10, T34 and T37. Cells were grown at 20°C. Error bars are SEM (three replicates per strain).

normality, missing values were imputed according to a normal distribution using Perseus software (Tyanova et al. 2016). The resulting values from the three technical replicates were averaged and compared between each sample by a multiple *t*-test in which the *p*-values were adjusted to control for the expected proportion of false positives (a.k.a., the false discovery rate; Benjamini and Hochberg 1995).

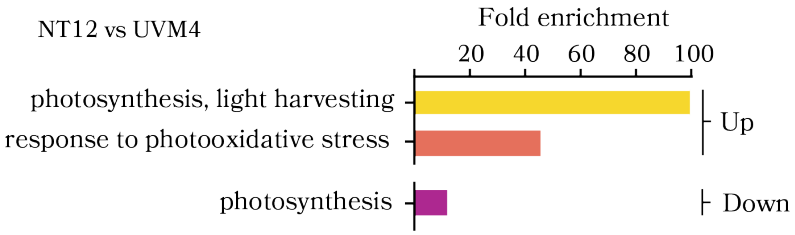
Protein fold-changes and adjusted *p*-values are presented as volcano plots in Fig. 4.7A and C for two comparisons: NT12 versus UVM4 and T7 versus NT12. Proteins are highlighted if they showed significantly different expression between strains ($p < 0.05$, three replicates). A threshold for the magnitude of change (fold change) was also set so that proteins are not highlighted unless they showed a positive or negative $\log_2$ fold change of 0.5 or more. This strategy identified 30 proteins in NT12 showing significant intensity changes from UVM4, and more than 200 proteins in T7 compared with NT12. PSAF and FNR are visible in Fig. 4.7 as two of the most significantly altered proteins in NT12 versus UVM4. Other salient up- and downregulated proteins are also highlighted, though many were unnamed or uncharacterised. To help us understand the functions of these highlighted proteins better, we performed enrichment analysis using the Gene Ontology Resource (http://geneontology.org/, accessed Sep. 2022), which produced a list of biological processes associated with the genes for these proteins. We identified the biological processes that were significantly over-represented in the highlighted proteins in comparison to all 4,030 detected proteins. The degree of over-representation is plotted below each volcano plot as fold enrichment (Fig. 4.7B and D).

Thylakoid proteins of interest to CET are highlighted further as a single heatmap in Fig. 4.8, along with proteins involved in alternative electron flows, state transitions, the CCM and representative subunits from the PSII core, cytb_6f, the PSI core and ATP synthase. PSII, cytb_6f and ATP synthase appear to be similarly expressed in all strains, with some minor exceptions. FNR and PsaF were expected to show significantly increased expression in NT12, T7 and T10 compared with UVM4 due to the strong PsaD promoter-driven expression of the PsaF-FNR fusion peptide. Surprisingly, PsaF appeared to be downregulated relative to UVM4 in all three mutants (Fig. 4.7A and Fig. 4.8). FNR was significantly increased in all three mutants, and interestingly was also slightly increased in T7 and T10 compared with NT12. However, native FNR is known to be only transiently associated with the thylakoids (Joliot and Johnson 2011, Mosebach et al. 2017), so the native FNR in NT12, as well as in UVM4, may have been partially lost during preparation of the thylakoid membranes, meaning these comparisons should be treated with caution. Nonetheless, higher FNR in the tethered-only mutants compared to NT12 suggests and upregulation of our PsaF-FNR fuson protein caused by the deletion of native FNR. Similarly, data for the soluble proteins PC, PGR5, FLVA and Fd (a.k.a., PETF) are not expected to be completely representative due to variation introduced by the thylakoid preparation procedure; these and other soluble proteins are indicated in

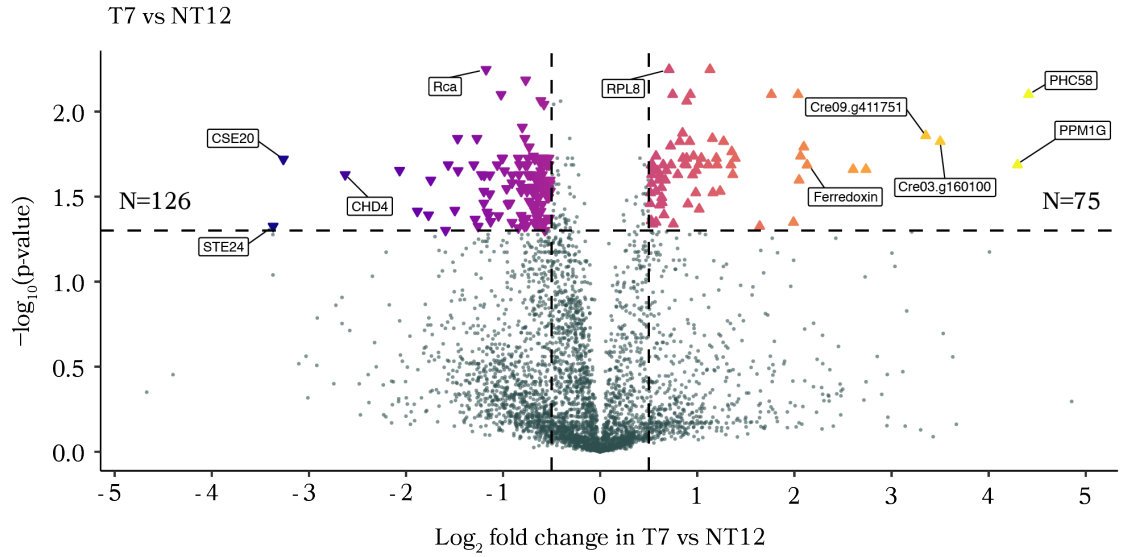
A


NT12 vs UVM4
PsaF
CPLD42
Cre06.g271700
FNR
CHLI1
FAP40
GSTS1
LHCSR3.2
Pbs27
N=20
N=10
PsaN
LHCSR1
Glyco_transf_92
−log10(p-value)
Log2 fold change in NT12 vs UVM4

B


NT12 vs UVM4
Fold enrichment
photosynthesis, light harvesting
response to photooxidative stress
photosynthesis
Up
Down

C


T7 vs NT12
Rca
RPL8
PHC58
Cre09.g411751
PPM1G
CSE20
CHD4
Ferredoxin
Cre03.g160100
N=126
N=75
STE24
−log10(p-value)
Log2 fold change in T7 vs NT12

D

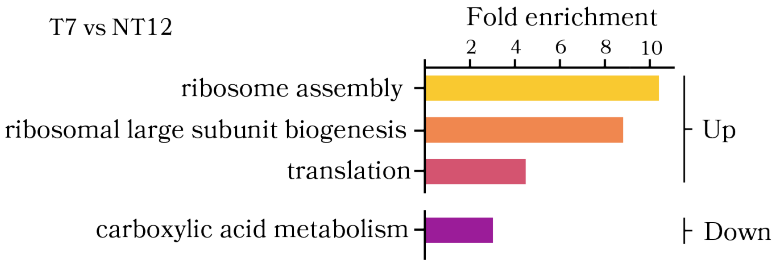

T7 vs NT12
Fold enrichment
ribosome assembly
ribosomal large subunit biogenesis
translation
carboxylic acid metabolism
Up
Down

Figure 4.7: **Significant protein intensity changes and fold enrichment of associated biological processes.** **A.** Volcano plot showing significant protein intensity changes between NT12 and UVM4. **B.** Enrichment analysis of biological processes for NT12 vs UVM4, showing a clear upregulation of photooxidative stress components LHCSR1 and LHCSR3.2. **C.** Volcano plot showing significant protein intensity changes between T7 and NT12. Many more proteins were significantly different between T7 and NT12 than between NT12 and UVM4. **D.** Enrichment analysis for T7 vs NT12. Note the difference in the scale of the x-axis to panel **B**, indicating fewer extreme changes to biological processes than between NT12 and UVM4. Samples were prepared in collaboration with Dr Phil Jackson. Mass spectrometry was performed by Phil Jackson and data analysis was aided by Jack Salisbury.

Fig. 4.8. The apparent depletion of PsaE peptide in the mutants is of particular note; PsaE stabilises the Fe_2S_2-containing PsaC/D subunit pair on the PSI acceptor side and is thought to interact with FNR in wild type cells (Andersen et al. 1992), so its loss may be caused by steric hindrance from the PsaF-FNR fusion. Also of note is the clear abundance of LHCSR proteins in NT12 and T7 versus UVM4; these proteins are involved in qE and upregulated in response to photooxidative stress. The ~45-fold enrichment in photooxidative stress components seen in Fig. 4.7B is entirely attributable to these LHCSR proteins. Increased photooxidative stress in NT12 may simply be a coincidence caused by improperly controlled growth conditions or could suggest a photoinhibition phenotype, potentially explaining the lack of growth of FNR-tether mutants under fluctuating light seen in Fig. 4.6C.

4.4 Discussion

The leading theory for the mechanism of Fd:PQ reduction (FQR) in CET postulates FNR-dependent electron transfer from stromal Fd to PQ in the Q_i site of cytb_6f (Buchert et al. 2020). If true, tethering FNR to PSI potentially represents a neat way to exclude FNR from catalysing FQR activity while maintaining its involvement in LET. The choice of PSAF and FNR as targets for mutation necessitated a nucleus-engineering strategy. This was successfully accomplished with the use of CRISPR-guided insertional mutagenesis, shown to be particularly efficient (~10%) for the targeted deletion of FNR using the hygromycin resistance cassette. The generation of these mutants showcases the promising potential of CRISPR-Cas9 for Chlamydomonas genetic engineering, and to our knowledge represents the first tethered FNR mutants in Chlamydomonas. However, several key questions need to be considered before we can confirm our system functions as intended.

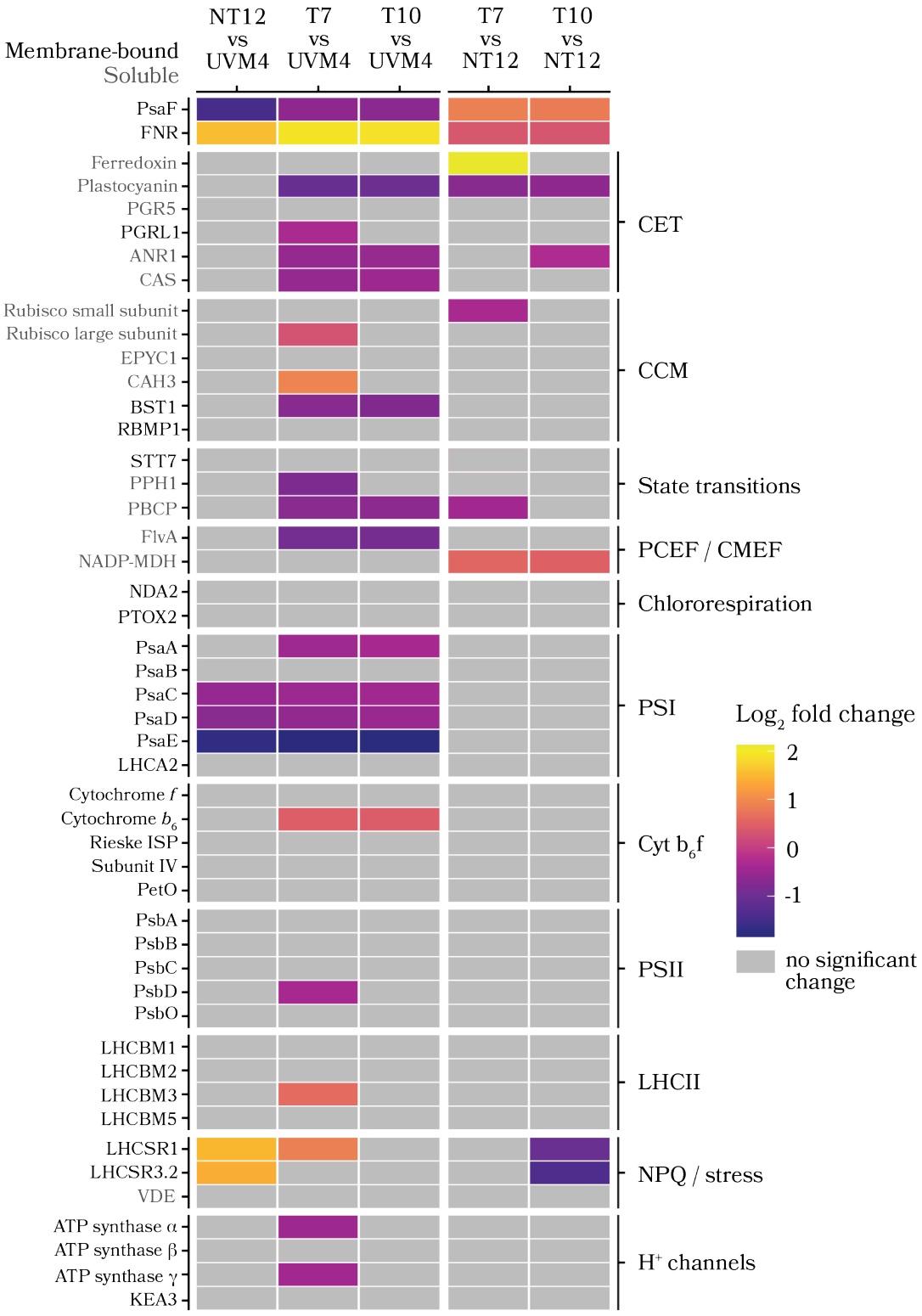

Membrane-bound
Soluble
NT12 vs UVM4
T7 vs UVM4
T10 vs UVM4
T7 vs NT12
T10 vs NT12
PsaF
FNR
Ferredoxin
Plastocyanin
PGR5
PGRL1
ANR1
CAS
CET
Rubisco small subunit
Rubisco large subunit
EPYC1
CAH3
BST1
RBMP1
CCM
STT7
PPH1
PBCP
State transitions
FlvA
NADP-MDH
PCEF / CMEF
NDA2
PTOX2
Chlororespiration
PsaA
PsaB
PsaC
PsaD
PsaE
LHCA2
PSI
Cytochrome f
Cytochrome b6
Rieske ISP
Subunit IV
PetO
Cyt b6f
PsbA
PsbB
PsbC
PsbD
PsbO
PSII
LHCBM1
LHCBM2
LHCBM3
LHCBM5
LHCII
LHCSR1
LHCSR3.2
VDE
NPQ / stress
ATP synthase α
ATP synthase β
ATP synthase γ
KEA3
H+ channels
Log2 fold change
2
1
0
-1
no significant change

Figure 4.8: **Relative protein intensities for key photosynthetic components.** Pairwise intensity changes for a subset of proteins involved in photosynthetic electron transport and regulation, showing comparisons between UVM4, NT12, T7 and T10. Protein intensities are shown on a $\log_2$ scale, where a $\log_2$ fold change of 1 is equivalent to a 100% increase, whereas a $\log_2$ fold change of -1 is equivalent to a 50% decrease. Proteins are grouped by function or structural partners listed on the right of the map. Soluble proteins are highlighted in grey; results for soluble proteins may be uninstructive due to unpredictable loss of soluble peptides during sample preparation. Lack of significant changes in PSII, cytb_6f and ATP synthase indicate these complexes accumulate to similar levels in all strains. Likewise, the PSI core subunits PsaA and PsaB are largely similar. Notably, PsaE is dramatically depleted in the mutants compared to UVM4, possibly indicating disruption caused by tethered FNR. Data analysis and processing was performed in collaboration with Jack Salisbury and Dr. Gustaf Degen.

4.4.1 Does PSI-FNR assemble *in vivo*?

The ideal tethering system would involve native-level accumulation of PSI and FNR, and native-like localisation of PSI-FNR in the membranes. Our system links the two genes as one open reading frame and drives expression using a non-native promoter and terminator, and potentially from a non-native locus as well. These changes likely have an impact on regulation, assembly, and transport of the fusion protein, potentially affecting correct incorporation of PsaF-FNR into PSI complexes. We observed lower levels of PsaF in our mutants compared with UVM4 and higher levels of FNR (Fig. 4.8). However, we also observed promising indications of successful assembly, e.g., the detection of FNR by immunoblotting of isolated PSI from strains T7 and T10. Subsequent size exclusion chromatography indicated that all the PSI particles in the sample were of similar sizes (results not shown), though the lack of structural data still leaves it is unclear whether every copy of purified PSI contained the FNR fusion. Previous work demonstrating that PSI can form *in vivo* without PsaF makes this unlikely (Farah et al. 1995), as well as observation that PsaF may be lost from PSI preparations during anion exchange (Hippler, personal communication). Still, the similar signal intensities observed from blotting purified PSI from UVM4 and T7 with anti-PsaF cautiously suggests there is no great difference in the PsaF content of PSI complexes between the strains. Additionally, the lack of PsaF signal or FNR signal at wild type molecular weights (~18 kDa and ~35 kDa, respectively) in T7 and T10 suggests the fusion peptide is not cleaved in vivo (Fig. 4.4C).

4.4.2 Is the PSI-FNR active in LET?

We were unable to collect in vitro kinetic data on the ability of our purified PSI-FNR complexes to catalyse PC to $NADP^+$ electron transfer due to time constraints.

However, successful growth in phototrophic conditions stands as indirect evidence of the functionality of PSI in NT12, T7 and T10 (Fig. 4.6A and B). The fact that even T7 and T10 can still grow phototrophically under constant light is promising evidence that the tethering of FNR to the PSI acceptor side does not abolish LET and suggests that the inclusion of the Strep tag does not completely disable the tethered FNR. T7 and T10 accumulate chlorophyll to a visually similar level as UVM4, suggesting the supply of NADPH required for many of the steps of chlorophyll biosynthesis is not severely compromised, as might be expected by loss of FNR-catalysed NADPH production (Lintala et al. 2012).

4.4.3 Are there any side-effects?

To our knowledge, this is the first example of an investigation into the role of FNR by expression of a tethered FNR fusion protein. This work builds on the promising results of a recent publication describing the tethering of the stromal hydrogenase, HYDA, to the PSI acceptor side (Kanygin et al. 2020). In that case, HYDA was fused to the PsaC subunit of PSI and used to demonstrate increased H2 production by outcompeting FNR for electrons at the PSI acceptor side. In a similar way, we hypothesised that our FNR-tether system might increase the efficiency of LET by concentrating FNR-bound Fd at the PSI acceptor side. On the contrary, the fact that NT12 grows slower than UVM4 despite having extra FNR may suggest a photosynthetic deficiency in NT12 (Fig. 4.6A and B). Additionally, the large changes in abundance of LHCSR proteins involved in response to photooxidative stress suggests cells are photoinhibited (Fig. 4.7A and 4.8), possibly because electron flow from PSI to stromal acceptors is limited. This would be consistent with previous work on FNR-deficient tobacco plants, which exhibited high photooxidative stress (Palatnik et al. 2003). There may be several reasons for this, due both to the intended mutation, (e.g., our tethered FNR may be a less efficient catalyst than free FNR, causing increased photooxidative stress and photoinhibition), and unintended mutations, (e.g., off-target effects of insertional mutagenesis).

All strains used in this work may house off-target mutations caused by mutagenesis, potentially introducing uncontrolled variation into the measurements. Firstly, UVM4 is itself a mutagenesis strain. The cause of the high transgene expression phenotype in UVM4 has been assigned to a loss-of-function mutation in a Sir2-type histone deacetylase (Neupert et al. 2020), however the strain may harbour additional mutations that may confound phenotypic measurements of photosynthesis in unexpected ways. Secondly, variation is introduced due to potential off-target effects of the CRISPR-mediated transformation of PsaF-FNR in NT12, and the further CRISPR-mediated deletion of FNR in T7 and T10. Though care was taken to find unique 20 bp Cas9 recognition sites during the project design, the guide RNA sequences used to target PsaF and FNR still showed predicted off-target recognition

of multiple 8 bp sub-sequences of the full recognition site throughout the genome. It is unclear whether efficient cleavage will have taken place at these sites, and we expect there is variability in the number of off-target mutations even amongst clones derived from the same transformation. As mentioned, we attempted to account for this variation in growth tests by including multiple clones of the native/tethered and tethered-only strains. Due to time constraints on the project we were unable to sequence the genomes of these strains, but this would be a valuable future objective to ensure there are no major off-target mutations compared to UVM4.

Risk of CRISPR off-target effects could have been avoided by developing an RNAi approach to knock down native PsaF and FNR. Though there are RNAi tools available for Chlamydomonas (Crozet et al. 2018), we did not pursue this approach because our CRISPR method appeared to yield the desired results, so we thought it prudent to push forward with the available mutants considering time constraints on the project. For the same reason, we did not attempt to explore recently published methods for encouraging homology-directed repair of CRISPR-induced cuts, instead relying on the messier but more easily achievable incorporation of exogenous DNA by NHEJ.

4.4.4 Do mutants T7 and T10 still perform PGR-CET?

Lastly, the most important remaining questions are whether CET is affected in T7 and T10, and what might be happening during photosynthesis at the level of PSI in these mutants? We cannot be certain that the peptide linker between PsaF and FNR in our fusion protein completely prevents interaction of FNR with cytb_6f, and the lack of structural data means we cannot estimate the range of movement of tethered FNR around PSI. Similarly, in the absence of *in vitro* measurements of CET using our mutants, FQR activity must be inferred indirectly instead.

If CET is impeded in T7 and T10, we might expect to see upregulation of mechanisms such as PCEF and CMEF to compensate for loss of CET, in accordance with previous data from PGRL1-deficient mutants (Burlacot et al. 2022). Mass spectrometry of samples from phototrophic cells was not undertaken due to time constraints but would provide a clearer picture of the consequences of the mutations on photosynthetic gene expression. Changes to the thylakoid proteome from TAP-grown cells, presented in Fig. 4.7 and 4.8, may still provide a general indication of the types of mechanisms affected by the mutations. The significant but marginal increase in NADP-MDH (NADP-malate dehydrogenase; a protein involved in CMEF) in T7 and T10 versus NT12 (Fig. 4.8) may indicate a compensatory mechanism, though this would need corroboration from measurements of mitochondrial activity to confirm. It is, however, worth nothing that intensity differences for soluble proteins may be less uninstructive than for thylakoid intrinsic proteins due to variability introduced during sample prep, as mentioned above.

Beyond inferences from changes in the proteome, FQR activity can be inferred by measuring changes in the redox state of key CET intermediates *in vivo* using spectroscopic techniques. The next chapter describes *in vivo* measurements on NT12, T7 and T10 to assess electron transfer and regulation, forming a clearer picture of what might be happening during photosynthesis in these mutants.

Chapter 5

Measuring photosynthesis in the FNR-tether mutants

5.1 Summary

The FNR-tether mutants NT12 and NT4 (expressing both native and tethered FNR) and T7 and T10 (expressing tethered FNR only) show slower growth and some putative alterations to the PSI acceptor side. FNR shows dynamic localisation in algae and is implicated in the mechanism of PGR-CET, potentially through a direct interaction with the stromal face of cytb_6f. We hypothesised that our tethered-only mutants would show lower CET due to a lack of native, mobile FNR. Cells expressing tethered FNR were assessed using three spectroscopic techniques to measure photosynthesis: (i) chlorophyll fluorescence was used to investigate changes in LHCII antenna size, photochemical activity of PSII and ability of cells to induce NPQ; (ii) light driven $\Delta\psi$ was assessed by measuring the electrochromic shift, which also enabled quantification of changes in photosynthetic electron transfer rate; and (iii) P700 oxidation changes were used to estimate PSI photochemical activity and disentangle the level of donor side and acceptor side limitation in mutant and wild type cells. The results show only scanty evidence for impaired CET in the tethered-only mutants, leading to rejection of the hypothesis. However, minor differences between UVM4 and the mutants led to our theorising a problem with photosynthetic control induced either by the FNR-tether or the lack of native FNR. Possible causes are discussed.

Some of the results presented here were gathered at the University of Münster, Germany, with the invaluable guidance and collaboration of Prof. Michael Hippler and his postdoc, Dr Felix Buchert. Contributions are stated explicitly in the figure legends.

5.2 Background

FNR catalyses the final step in LET, facilitating both the reduction of $NADP^+$ using electrons from two molecules of Fd^- and the reverse reaction to regenerate Fd^- from NADPH. This is an important role in cellular energetics, as evidenced by the essential role of FNR for photosynthesis in many organisms including Chlamydomonas (Chapman et al. 2015). In Chapter 4, we described the generation of two mutant strains expressing a PsaF-FNR fusion protein, with the aim of tethering FNR to the stromal face of PSI; first a strain expressing both native and tethered FNR (native/tethered; NT) and subsequently a strain deficient in native FNR (tethered-only; T). Previous evidence from angiosperms with multiple FNR isoforms suggests FNR can redistribute in response to environmental conditions, and that a loss of soluble FNR has consequences for photosynthetic electron flow (reviewed in Goss and Hanke 2014). One putative function of FNR that may requires dynamic redistribution is as an electron transfer mediator for CET, possibly via a direct interaction with cytb_6f or as part of a CET-supercomplex (Buchert et al. 2020, Steinbeck et al. 2018). If FNR is crucial to the mechanism of CET by direct interaction to the stromal face of cytb_6f, its tethering to PSI in Chlamydomonas with a short peptide linker could impair CET function by precluding its association with cytb_6f.

In this chapter we present a series of experiments aimed at assessing photosynthesis in UVM4 and our FNR-tether mutants. Formally expressed, our null hypothesis is that CET is unaffected in the tethered-only strains (T7, T10, etc.), with our alternative hypothesis being that a lack of native FNR in the tethered-only mutants results in impaired or dysfunctional CET. These hypotheses are composed relative to the native/tethered strains, since NT12 is the background strain for T7, T12 and T34, etc. This is an important distinction, since there may be confounding factors introduced by the tethered FNR that make comparisons between tethered-only strains and UVM4 difficult. Nonetheless, the experiments in this chapter include measurements of UVM4 alongside our FNR-tether mutants with the hope of improving our understanding of what may be happening in photosynthesis in both mutant genotypes.

CET is notoriously difficult to measure in Chlamydomonas. The stroma contains more available electron sinks than in angiosperms, making it more difficult to isolate the CET signal. Chlamydomonas PSI and LHCI also contain fewer far-red absorbing chlorophylls than angiosperms (Melkozernov et al. 2004), making it far more difficult to selectively drive CET using PSI-specific excitation without simultaneously exciting PSII. To side-step these issues, Chlamydomonas researchers have developed a set of spectroscopic techniques to disentangle CET from LET. These usually involve suspension of cells in buffer or TP media, often with an inert thickening agent to prevent sedimentation, followed by actinic (photosynthesis-driving) illumination

with red or blue LEDs, and a measuring light to track changes in fluorescence or absorption as the experiment progresses. We subjected our mutants to a series of such tests, first using the IMAGING-PAM and DUAL-PAM fluorescence systems from Walz, Germany, and then the JTS-150 spectrophotometer from BioLogic, France. Together, these systems enabled the use of three main techniques, each of which measure different but related phenomena occurring at the thylakoid membrane on the nanoscale during actinic illumination. These techniques are briefly introduced in the following sections.

5.2.1 Chlorophyll fluorescence

All chlorophyll-containing photosynthetic organisms fluoresce. Absorption of red light excites an electron in chlorophyll to the S_1 state. Absorption of a blue photon excites an electron to a higher energy S_2 state, which quickly relaxes to S_1. Electrons in the S_1 state then decay to S_0 through several competing pathways, including the use of absorbed energy for photochemistry, dissipation of energy as heat or re-emission of energy as a photon. Thylakoid-associated chlorophylls in light harvesting pigment-protein complexes constantly emit absorbed energy as red photons, allowing the fruitful use of chlorophyll fluorescence as a fluorescence microscopy marker for the thylakoids (Fig. 5.1A). The level (or yield) of chlorophyll fluorescence is highly responsive to changes in light intensity; if light intensity is increased, yield will increase accordingly within ~1 s. This fast rise (known as the Kautsky effect) is thought to be due mainly to light-driven electron flow quickly overwhelming the electron carrying capacity of the PQ pool (Maxwell and Johnson 2000). This causes a net over-reduction of the PSII acceptors, blocking further electron transfer from the PSII special pair P680 to the Q_A site quinone at the acceptor side of PSII, leading to a back reaction within PSII and dissipation or emission of the excitation energy instead (Maxwell and Johnson 2000). For this reason, a PSII reaction centre with a reduced Q_A site is referred to as closed, and the quantity of closed PSII centres is proportional to the fluorescence yield. While both PSI- and PSII-associated antenna can fluoresce, measurable chlorophyll fluorescence is dominated by signal from PSII and associated LHCII; PSI makes only a fractional contribution to fluorescence output due to having an extremely short fluorescence lifetime at ambient temperatures (Melkozernov et al. 2004).

Under sustained high illumination, a drop in fluorescence yield over several minutes is often seen, thought to be caused by two co-occurring phenomena: photochemical and non-photochemical quenching (Maxwell and Johnson 2000). Photochemical quenching is an increase in the activity of downstream acceptors (such as the CBB cycle enzymes), which increases the rate of electron flow through the ETC. This works to unblock the system, relieving some of the PQ pool over-reduction and allowing more PSII centres to open, leading to a drop in fluorescence. Non-

photochemical quenching describes a range of processes but is due mainly to an induction of qE (an increase in dissipation of absorbed energy as heat instead of fluorescence) and due to a transition to state II that decreases the connectivity of LHCII to PSII, decreasing the PSII-derived fluorescence. By monitoring fluorescence yield over time, researchers can assess the proportion of energy used for PSII photochemistry, can estimate the redox state of the PQ pool and can measure the ability of cells to induce qE.

5.2.2 The electrochromic shift

Light driven electron flow is strongly coupled to proton translocation into the lumen, generating a trans-thylakoid proton gradient that has a pH component (ΔpH) and a charge component ($\Delta\psi$), which together make up the *pmf* (Avenson et al. 2004). $\Delta\psi$ can be measured *in vivo* by absorption spectroscopy due to the Stark effect. This effect describes a slight change in the absorption spectrum of thylakoid-associated pigments during illumination due to the distorting effect of $\Delta\psi$ on the electrons in these pigments. The absorption shift is present in the green section of the electromagnetic spectrum, measurable as an increase in absorption at ~515-520 nm relative to ~545-550 nm, referred to as the electrochromic shift (ECS). In the light, $\Delta\psi$ is also contributed to by the movement of electrons; charge transfer within PSI and PSII during photochemistry moves negative charge towards the stroma, as does electron transfer within cytb_6f during the Q-cycle, which briefly moves electrons from the Q_o site on the lumen side to the low potential chain on the stromal side. Changes in the ECS signal are therefore reflective of light driven electron flow and proton translocation, providing a useful spectroscopic measure of photosynthetic activity. The ECS signal is complicated by competing signals from the conversion of violaxanthin to zeaxanthin by VDE (~505 nm) and from a broad peak at ~535 nm thought to be due to scattering caused by lumen acidification (Schreiber and Klughammer 2008). Fast changes occurring on the millisecond-to-second timescale are mostly independent of these factors and so can be used to estimate the ability of cells to form $\Delta\psi$.

Measured over time in the presence of actinic light, a flat ECS signal indicates a photosynthetic system in balance – called the steady state – with the production of $\Delta\psi$ continuously counter-balanced by the release of protons back through ATP synthase. When illumination is removed, the ECS signal in the light is collapsed in two ways, either very quickly (< 1 ms) by charge recombination events in the photosystems or, more slowly, by escape of excess protons back into the lumen via ATP synthase, other ion channels, or leakage through the lipid bilayer. An example ECS trace is shown in Fig. 5.1B, detailing the four main dynamic phases caused by changes in illumination. When actinic light is first introduced the signal rises sharply, referred to as the A-phase. The slower rise phase that follows is known as the B-phase, which

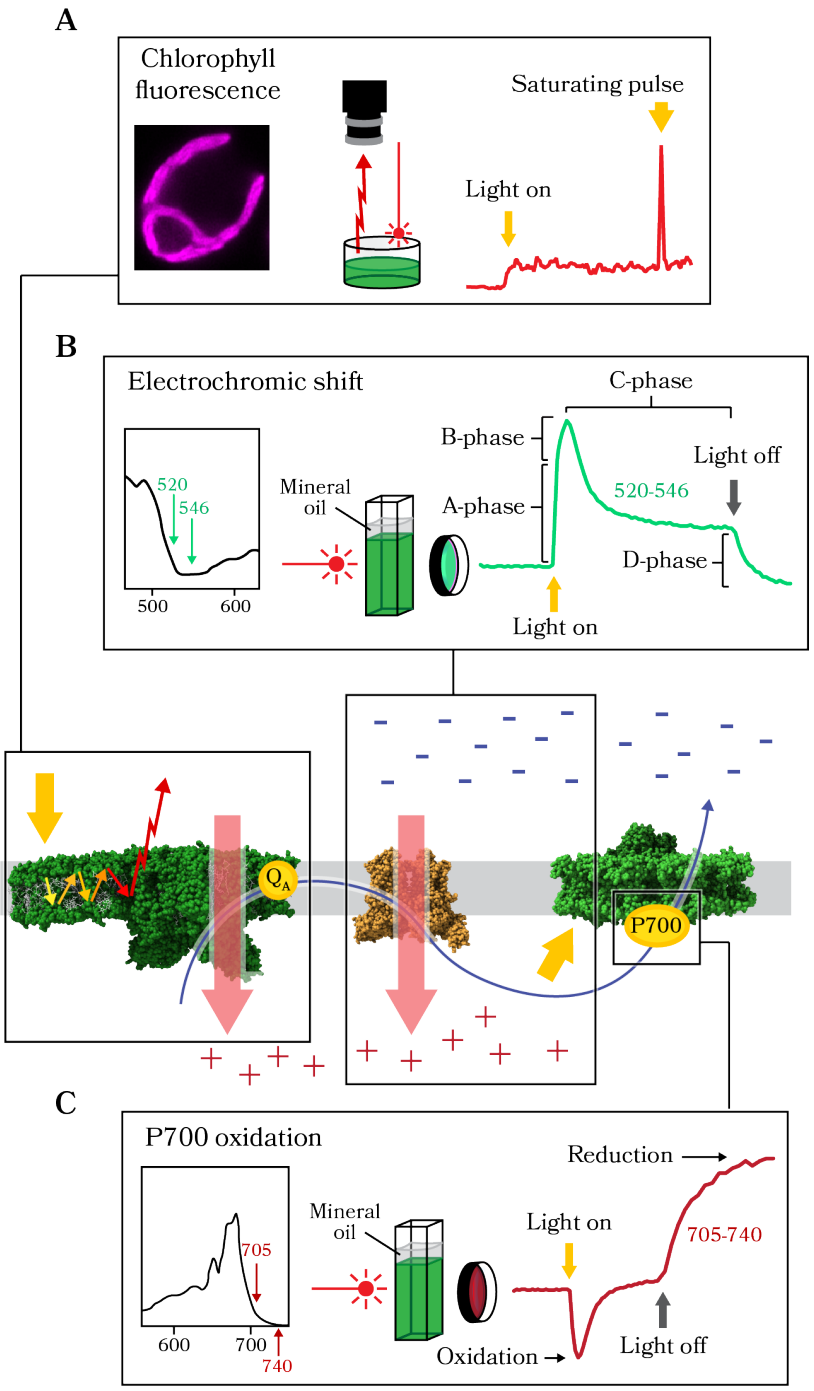

A
Chlorophyll fluorescence
Saturating pulse
Light on
B
Electrochromic shift
C-phase
B-phase
Light off
520
546
Mineral oil
A-phase
520-546
D-phase
500
600
Light on
Q_A
P700
C
P700 oxidation
Reduction
Mineral oil
Light on
705
705-740
600
700
740
Oxidation
Light off

Figure 5.1: **Chlorophyll fluorescence, the electrochromic shift and P700 oxidation A.** Chlorophyll fluorescence resulting from re-emission of absorbed light from PSII-associated chlorophyll molecules in the thylakoids of Chlamydomonas. The phenomenon is reliable enough to use a microscopy tool but can be monitored using time-resolved spectroscopy from a population of cells, providing information on PSII photochemical activity. **B.** An electrochromic shift (ECS) in the absorption spectrum of cells in the green region (~500-550 nm) is induced upon absorption of light. The cause is thought to be a change in the electrical micro-environment of certain pigments in the thylakoids caused by the small voltage that develops across the thylakoid membrane due to proton translocation and electron flow. Time-resolved ECS spectroscopy can be described by four phases during illumination and in darkness, explained in the text. Cells were overlaid with mineral oil during measurements in this chapter to encourage anoxia. **C.** Monitoring of absorption changes at 705 nm can reveal dynamics of the net appearance and disappearance of the PSI reaction centre cation $P700^+$. Appearance of $P700^+$ causes a drop in signal, associated with net PSI oxidation. Cells were overlaid with mineral oil during measurements in this chapter to encourage anoxia. Protein structural models are the same as those shown in Figure 1.2 on page 7.

decays slowly in the C-phase due mainly to ATP-synthase until a new steady-state is reached. When actinic light is removed, the signal decays exponentially, referred to as the D-phase, as the system settles back to the dark baseline. The initial rate of the D-phase (i.e., between 1-10 ms) is proportional to the rate of electron transfer in the light, and so can be used to compare the efficiency of photosynthetic electron transfer between strains. The vertical extent of the D-phase (from the light steady-state to the dark baseline) is proportional to the total *pmf* that was generated during illumination. When PSII activity is disabled by the addition of the donor side inhibitor HA and the acceptor side inhibitor DCMU, these D-phase kinetics provide information on the rate of CET, since electron flow from cytb_6f to PSI must originate in the stroma.

5.2.3 Measurement of P700 redox state

PSI fluorescence provides too weak a signal to enable its use as a monitor of PSI photochemistry. Instead, researchers assess PSI activity by following absorbance changes at ~705 nm or in the far red at ~820 nm, both relative to an isosbestic point on the spectrum. Under actinic illumination, an increase in absorption strongly indicates the net appearance of $P700^+$ cation in the sample, generated by the light-driven oxidation of PSI and the flow of electrons from P700 to stromal acceptors (Alric et al. 2010). With continued illumination this net oxidation disappears, and a steady state is achieved. When the light is turned off, the absorbance signal decays back to a dark baseline as the $P700^+$ are reduced by electrons from cytb_6f via PC (Fig. 5.1C).

Researchers use the kinetics of these changes in P700 oxidation to quantify the

donor and acceptor side limitation of PSI; these are measures of either the paucity of electrons flowing to P700 from PC (caused by a lack of electron flow through cytb_6f) or of the downstream blockage of electron flow into the stroma (caused by stromal over-reduction). For example, if a proportion of the population of $P700^+$ in a sample remains oxidised even in the illuminated steady state, this implies an insufficient availability of reduced PC^-, likely due to cytb_6f-dependent limitation in upstream electron flow. Alternatively, researchers can use a short, saturating pulse of strong illumination to attempt to fully oxidise the P700 present in the sample; if full oxidation cannot be achieved, this implies an insufficient availability of stromal acceptors for electrons emerging from PSI (i.e., the reaction centres are closed). Though values obtained from these experiments are relative, they can be useful for comparing the redox state of the stroma between strains. Additionally, in the presence of the PSII inhibitors DCMU and HA, the rate of re-reduction of $P700^+$ when illumination is removed is dependent on the rate of stromal electrons entering the PQ pool, due mainly to the FQR reaction of PGR-CET. Researchers can then compare the rate of CET between strains, with a higher rate of dark re-reduction implying a greater capacity for CET.

5.3 Results

5.3.1 Analysis of chlorophyll fluorescence shows minor differences at high light

We set out first to explore the effects of our FNR-tether mutation on basic photosynthetic parameters measurable using chlorophyll fluorescence. To measure changes in the fluorescence yield in response to light, we subjected cells to increasing light intensity over a 12-minute period (a.k.a., a fluorescence lightcurve) (Fig. 5.2). Lightcurves are a useful tool for comparing PSII activity between strains, providing indications for the photochemical activity of PSII and the relative rate of LET. Cells were suspended in TP to equal chlorophyll concentrations then placed under an IMAGING-PAM (Walz) and fluorescence recorded in response to blue actinic light. Fig. 5.2A shows the averaged results of three such independent experiments (mean ± SEM) using UVM4, NT12, NT4, T7 and T10. Each experiment was composed of the means of four technical replicates. Results from NT12 and NT4 were highly similar, likewise for T7 and T10, and were therefore plotted as overall averages for native/tethered and tethered strains for simplicity. A stepwise increase in fluorescence associated with increasing intensities of actinic light can be seen for all strains, with cells incubated at each light intensity for 2 min before being subjected to a saturating pulse to measure the maximum yield of PSII. Here the raw fluorescence yield traces have been normalised between strains by dividing by F_v/F_m (the maximum efficiency of PSII after dark adaptation), then aligning curves

such that F_o (the fluorescence yield in the absence of actinic light) is equal to 0, with these values ascertained from a saturating pulse of actinic light at the start of the lightcurve. Fig. 5.3A provides graphical information on how these values are calculated from the fluorescence yield signal, while more detailed equations are presented in section 2.7.3 in Chapter 2 on page 55. At the highest light intensities used (~90 and 175 $\mu E\ m^{-2}\ s^{-1}$), fluorescence yield flattens then begins to drop during illumination, and the saturating pulse response at 12.5 min is diminished compared with lower light intensities. Though UVM4 showed high variability between the three experiments at low light intensities, we observed a clearer trend of lower fluorescence output in the FNR-tether mutants at higher light, both between the saturating pulses and at the height of the pulses.

Lower fluorescence in the mutants could indicate smaller PSII antennae; fluorescence yield is affected by the state of the cells in that cells in state I have larger antenna around PSII and so emit more fluorescence. We therefore assessed the antenna size between UVM4, NT12 and T7 to check for major differences in cell state when subjected to the same treatment. 200 µl of dilute cell suspension were plunge frozen in liquid nitrogen and their absorption spectra measured at 77 K (Fig. 5.2A, inset). T7 and NT12 both showed a higher signal for the ~700-720 nm peak (associated with PSI) compared with the peak at ~680 nm (associated with PSII) (Fig. 5.2) in both oxia and anoxia, implying a relatively lower signal from PSII. This could be due either to fewer open PSII reaction centres or due to state transitions, with the mutants being pushed more into state II than UVM4.

We then repeated all measurements with cells placed in anoxia (see Methods), which has the effect of over-reducing the stroma, stimulating CET and inducing state II. Fig. 5.2B shows the effects of anoxia; the fluorescence yielded from each saturating pulse was much lower compared with the yield between the pulses indicating a strong shift to state II in all strains. However, a slight trend of lower fluorescence in the mutants remained, suggesting that the mutants are still pushed further into state II than UVM4. This was backed up by the 77 K spectra after induction of anoxia, in which T7 and NT12 still have a larger PSI-related signal relative to PSII than does UVM4.

Fast recordings of the fluorescence kinetics induced by each saturating pulse enabled the calculation of two important values, 1-*qP* and NPQ. 1-*qP* is an estimation of the proportion of closed PSII reaction centres, which increases along with light intensity as the PQ pool becomes over-reduced and blocks electron transfer from Q_A (Maxwell and Johnson 2000). NPQ is a measure of non-photochemical quenching, the triggering of which depends on the cell's ability to produce ΔpH. NPQ values reported here were derived from the loss of fluorescence yield induced by a second saturating pulse relative to an initial pulse, with the loss in yield likely due to quenching mechanisms induced by light treatment in between the two pulses.

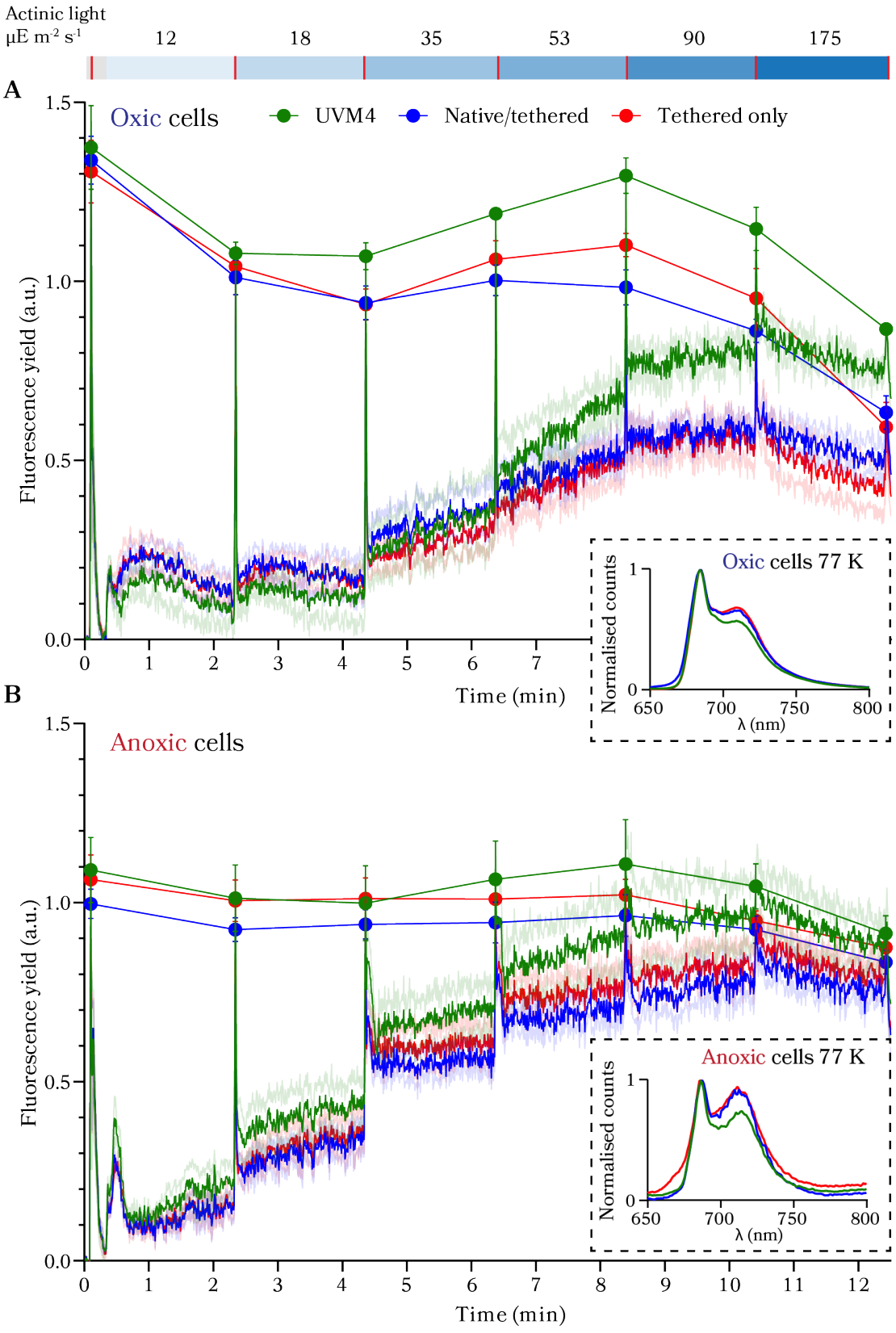

Actinic light
µE m-2 s-1
12
18
35
53
90
175
A
Oxic cells
UVM4
Native/tethered
Tethered only
Fluorescence yield (a.u.)
Time (min)
Oxic cells 77 K
Normalised counts
λ (nm)
B
Anoxic cells
Fluorescence yield (a.u.)
Time (min)
Anoxic cells 77 K
Normalised counts
λ (nm)

Figure 5.2: **Kinetics of fluorescence yield in response to increasing light. A.** Mean fluorescence yield of UVM4, NT4, NT12, T7 and T10 in oxic conditions ± SEM. NT4 and NT12 are averaged together, as are T7 and T10. Fluorescence yield in response to saturating pulses at the end of each 2-min period of light are shown as circles ± SEM. The inset shows whole-cell fluorescence emission spectra at 77 K, with the high peak at ~680 nm representing PSII-associated fluorescence and the lower shoulder at ~720 nm representing PSI-associated fluorescence. **B.** A repeat of the experiments shown in panel A under anoxia. Note higher fluorescence yields overall, and higher PSI-associated fluorescence at 77 K indicating relatively less PSII-associated fluorescence. 77 K spectra were recorded in collaboration with Felix Buchert.

Equations for 1-qP and NPQ are detailed in section 2.7.3 in Chapter 2 on page 55. Fig. 5.3A illustrates how the values used in these equations are calculated. Fo' is not possible to obtain using far red light in Chlamydomonas, so was instead estimated (Oxborough and Baker 1997) (see Section 2.7.3 in Chapter 2 on page 55). Fig. 5.3B and C shows calculated values for 1-qP and NPQ for each light intensity (mean ± SD), again showing averages for native/tethered and tethered-only genotypes by combining calculations for NT12 and NT4 and for T7 and T10. NPQ appeared to be slightly higher in oxic conditions in all strains but was so variable that no significant difference could be seen between strains. Despite the apparent difference in fluorescence yield from the kinetic traces in Fig. 5.2, 1-qP in the mutants appeared very similar to UVM4 in both oxia and anoxia. This suggested there was no clear difference in the redox state of the PQ pool between strains during these measurements.

5.3.2 FNR-tether mutants show normal PSI:PSII ratio

To follow up the observable differences in 77 K fluorescence yield between our mutants and UVM4, we assessed the ratio of open PSI:PSII reaction centres in our cells. In collaboration with Dr. Felix Buchert and Prof. Michael Hippler at the University of Münster, we measured the ECS signal of our FNR-tether mutants using a JTS-150 system (BioLogic). This system includes a laser module capable of nanosecond saturating pulses, which enabled us to induce near-to-single turnover photosystem excitation. When DCMU and HA were added to the cells, this allowed quantification of ECS changes in our cells normalised to the number of charge separations per second per PSI in the sample. Fig. 5.4A shows two example ECS traces from the JTS-150 resulting from a 6 ns flash, illustrating how the addition of DCMU and HA affects the A-phase of the ECS. By making two serial measurements on the same sample of cells, first without (Fig. 5.4A, (i)), then with DCMU and HA (Fig. 5.4A, (ii)), we could quantify the ratio of open PSI to open PSII reaction centres in our samples. Fig. 5.4B shows the results of six measurements of PSI:PSII ratio in

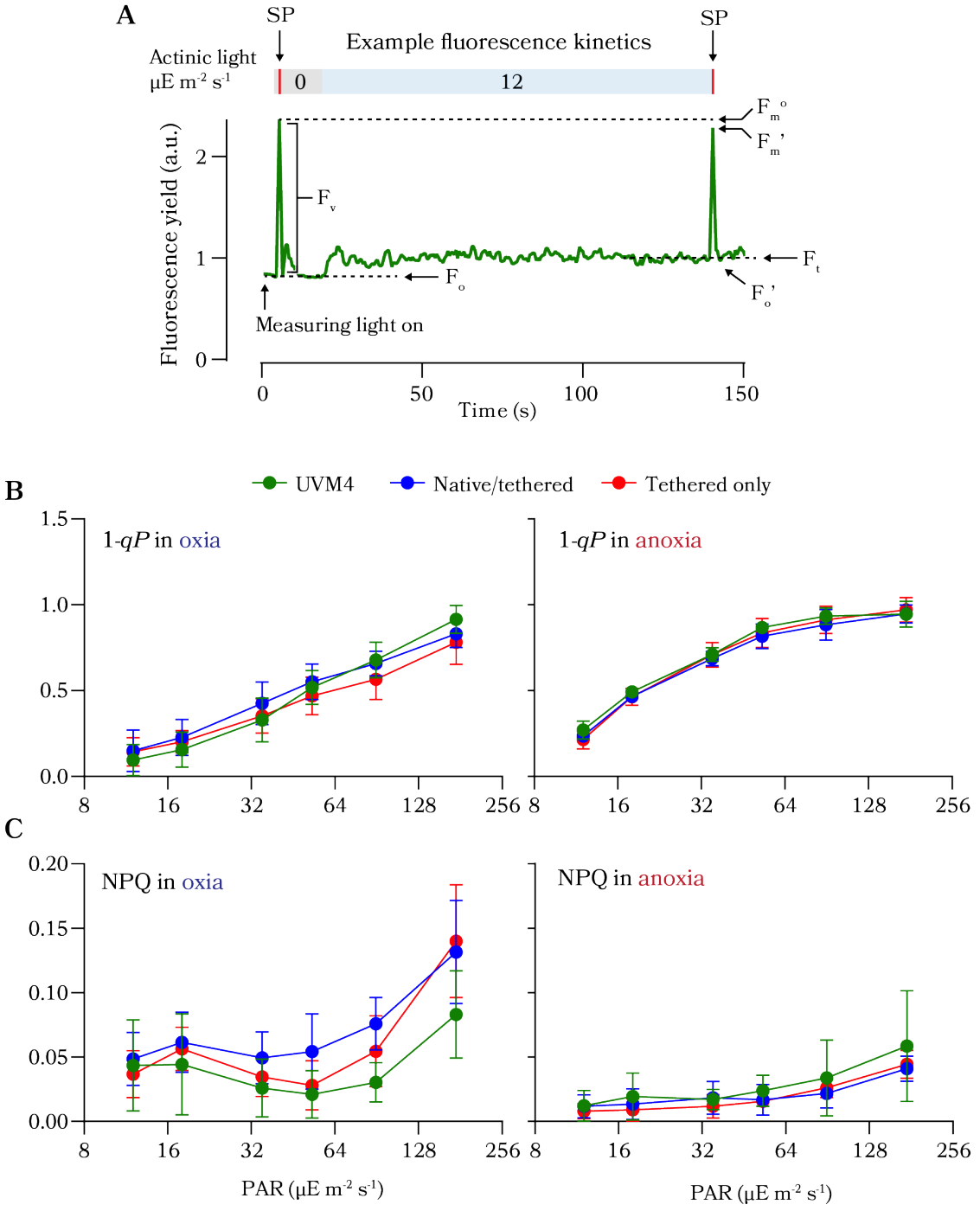


Figure 5.3: **Values derived from fluorescence kinetics A.** an example fluorescence trace showing fluorescence kinetics between two saturating pulses, showing a slight drop in pulse-induced yield of the second pulse relative to the first. Values indicated are used to estimate the behaviour of PSII acceptors, Q_A and Q_B, as well as to calculate the level of NPQ induced by the cells. **B.** 1-*qP* in oxia and anoxia calculated from the lightcurves in Figure 2. **C.** NPQ in oxia and anoxia calculated from the lightcurves in Figure 2. See methods for equations used. Note NPQ results showed high variation, but no differences were significant (*t*-test, $p > 0.05$).

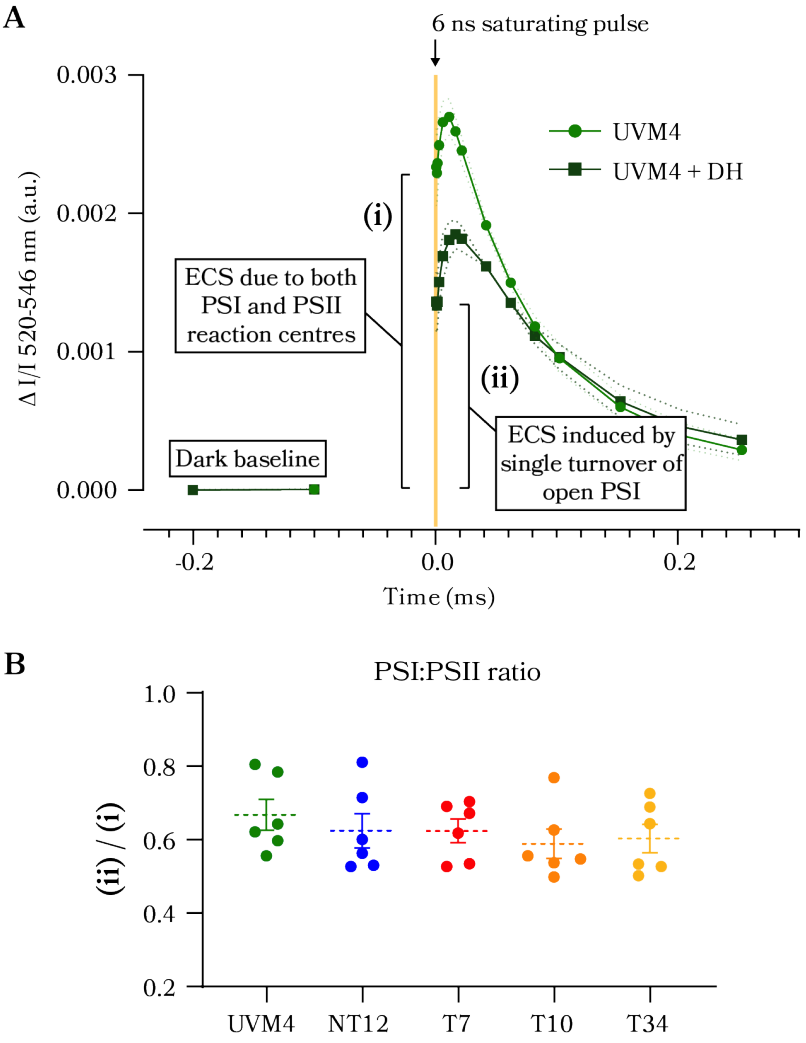


Figure 5.4: **ECS normalisation method and PSI:PSII ratio. A.** Example ECS data from UVM4 induced by a very short saturating pulse in the absence (i) and presence (ii) of DCMU and HA. DH = 10 μM DCMU and 1 mM HA. Data in the presence of DCMU (ii) is an estimation of the ECS signal induced by a single turnover of open PSI centres in the sample. **B.** The difference in ΔI/I absorption values of the first measured point after t = 0 (t ≈ 200 μs) in treated and untreated cells in panel A was used to calculate the PSI:PSII ratio. There were no significant differences in PSI:PSII ratio; all strains show ~0.6-0.7 open PSI per PSII. Error bars are SEM.

this way (means ± SEM). No significant differences could be ascertained between strains, indicating that the PSI-FNR complexes accumulate in our mutants to a similar level as does native PSI in UVM4 (~60-70% of open PSII) and are able to perform photochemistry to a similar degree under the conditions tested. This result was remarkable twice: first, as it suggested there was no significantly deleterious effect of tethering FNR to the PSI acceptor side on the number of functioning PSI centres relative to PSII, and second, because it showed that differences in 77 K spectra (Fig. 5.2) were not due to differences in PSI and PSII quantity, but suggested our mutants were indeed pushed further into state II than UVM4.

5.3.3 FNR-tether mutants show a trend of lower *pmf* in high light

Progressing logically, we reasoned that if our mutants were being pushed further into state II than UVM4, this may be caused by low ATP production. An ATP deficit could be a symptom of decreased CET, which affects the cells' ability to generate *pmf*, particularly under anoxia. We therefore endeavoured to quantify the ability of the FNR-tether mutants to generate *pmf*, this time by recording the ECS signal under eight stepwise increases in light intensity, known as a lightcurve. Each lightcurve consisted of a 2 min incubation at each light intensity followed by a 0.5 s dark interval, during which the ECS was recorded once per millisecond. Fig. 5.5A details an example of the changes undergone by the ECS signal during these light curves. The extent of decay of the ECS signal in the first 300 ms of each dark interval is often designated as proportional to the *pmf* generated by the cells during the preceding illumination (Avenson et al. 2004). Following each dark interval, the light intensity was increased to the next level in the light curve (Fig. 5.5A, (i)). The *pmf* can be estimated using two additional parameters, detailed in Fig. 5.5A, (ii): the proton flux, v_{H^+}, and the proton conductivity of the membrane, g_{H^+}. These are derived from the slope of each dark interval ECS decay as shown in Fig. 5.5A, (ii). Values for *pmf* from oxic and anoxic sample are presented in Fig. 5.5B for each light intensity (means ± SD). The results indicate a high degree of similarity between all strains tested (UVM4, NT12, NT4, T7, T10 and T34), particularly at lower light intensities. Anoxic samples produced lower *pmf* values than oxic samples, most evident at higher light intensities. Mean values appeared to trend towards lower *pmf* in our mutants, and indeed a small but significant difference was observed in anoxia between mean *pmf* values at the highest light intensity (One-way ANOVA with Tukey's posthoc HSD, 1,940 μE m^{-2} s^{-1}; $N = 10$, $p < 0.05$ where indicated; normality was confirmed using a Kolmogorov-Smirnov test). If representative of a physiological trend, lower *pmf* in our mutants would also be consistent with an ATP deficit, implying FNR-tether mutants struggled to produce enough ATP.

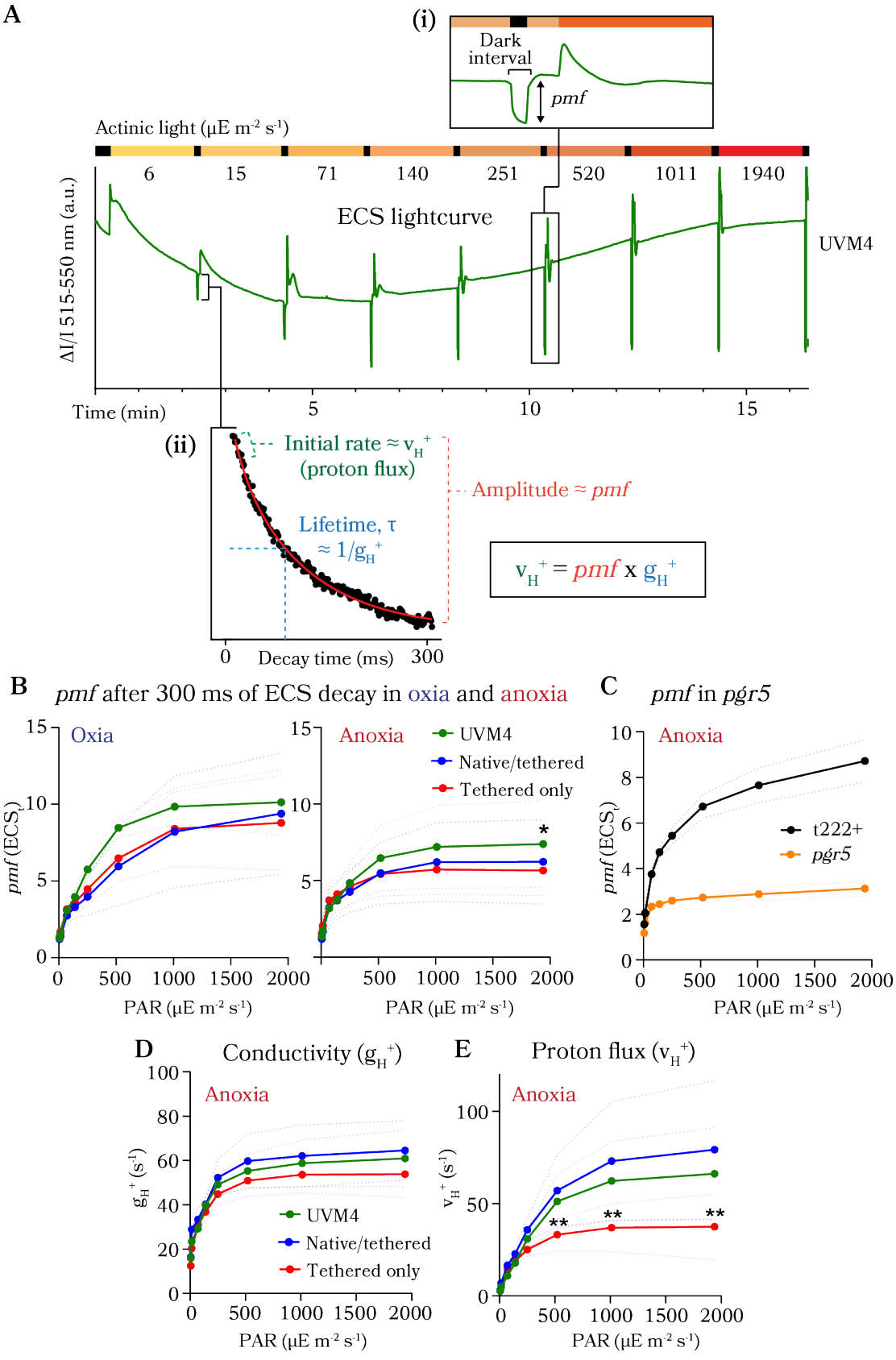

A
(i)
Dark interval
pmf
Actinic light (µE m-2 s-1)
6
15
71
140
251
520
1011
1940
ECS lightcurve
UVM4
ΔI/I 515-550 nm (a.u.)
Time (min)
5
10
15
(ii)
Initial rate ≈ vH+ (proton flux)
Amplitude ≈ pmf
Lifetime, τ ≈ 1/gH+
vH+ = pmf x gH+
0
Decay time (ms)
300
B
pmf after 300 ms of ECS decay in oxia and anoxia
Oxia
Anoxia
UVM4
Native/tethered
Tethered only
pmf (ECS)
PAR (µE m-2 s-1)
C
pmf in pgr5
Anoxia
t222+
pgr5
D
Conductivity (gH+)
Anoxia
gH+ (s-1)
E
Proton flux (vH+)
Anoxia
vH+ (s-1)

Figure 5.5: **Kinetics of the ECS signal in response to increasing light. A.** An example ECS trace in response to increasing actinic light, with a 1 s dark interval inserted after 2 min incubation at each light intensity. (i) shows a zoomed view of the dark interval, the vertical extent of which is proportional to the *pmf* generation by the cells in the sample. (ii) shows an example drop in ECS during a dark interval and demonstrates how values for proton flux (v_{H^+}) and conductivity (g_{H^+}) can be calculated from a bi-exponential model fit to the curve. **B.** Comparisons of the *pmf* after 300 ms of dark between UVM4, NT4, NT12, T7, T10 and T34 in oxia and anoxia. NT4 and NT12 were averaged together, as were T7, T10 and T34. Means ± SD are shown, $N = 10$. **C.** The results of repeating the experiments in panel **B** on cells of the PGR5-deficient mutant pgr5 and its wild type t222+. Means ± SD are plotted, $N = 3$. **D.** Values for conductivity and proton flux calculated using the method in panel A (ii) and derived from the anoxic data in panel B. Means ± SD are plotted. Proton flux was significantly lower in tethered only mutants at higher light intensities ($p < 0.01$ where indicated).

To clarify our data, we included two extra strains in our measurements for comparison: *pgr5*, a knock-down mutant deficient in PGR5 (kindly donated by Prof. Michael Hippler) and its progenitor wild type strain, t222+. *Pmf* in *pgr5* was drastically reduced compared to t222+, suggesting at least that the FNR-tether mutations do not affect the ETC in the same was as does the deletion of PGR5.

Pursuing the idea that our mutants may show a weakened version of the same trend, we attempted to disentangle the measured *pmf* values. Pmf is proportional to the influx of protons into the lumen (proton flux; v_{H^+}) and the efflux of protons through the ATP synthase (conductivity; g_{H^+}). The conductivity, g_{H^+}, is calculable from the half time of signal decay during a dark interval. We therefore needed to measure the shape of the signal decay for each strain at each light intensity and fit a decay model to estimate the half times. We found each decay could be modelled reasonably closely using a bi-exponential function using the SSbiexp package in R. Once modelled, we ascertained values for g_{H^+} (see Section 2.7.4 in Chapter 2 on page 55, and calculated v_{H^+} by multiplying g_{H^+} by the *pmf* for each curve. The resulting values are plotted as lightcurves in Fig. 5.5D and E. Though g_{H^+} was similar between strains, the tethered-only mutants showed a trend of lower v_{H^+} at high light (One-way ANOVA with Tukey's posthoc HSD, $N = 10$, $p < 0.01$ where indicated; normality was confirmed using a Kolmogorov-Smirnov test). This indicated that the tethered-only mutants were struggling to generate *pmf* in high light due to a low influx of protons into the lumen, likely linked to a defect in electron transfer. We reasoned this could be either caused by a deficiency in CET or a dysfunction in photosynthetic electron flow as a whole.

5.3.4 NPQ in FNR-tether mutants is impaired after high light treatment

Though the calculated values for NPQ in Fig. 5.3C showed no clear difference, we reasoned that longer light treatments might exacerbate a problem with *pmf* generation in our mutants. Since NPQ is triggered by the generation of ΔpH (a component of *pmf*), we returned to measurements of NPQ using chlorophyll fluorescence measurements. We subjected cells of UVM4, NT12, T7 and T10 to white light at 500 μE m^{-2} s^{-1} for 4 h, then measured NPQ under strong actinic illumination (~720 μE m^{-2} s^{-1} blue light) for 22 min, followed by 15 min darkness. Saturating pulses were taken throughout to record NPQ induction and relaxation. Fig. 5.6 shows the average NPQ for each strain (means ± SD) after 4 h high light treatment. The FNR-tether mutants showed severely impaired NPQ compared to UVM4 when pre-treated with high light. These data could be explained by both a lower *pmf* in the mutants and perhaps by a higher level of photoinhibition, damaging PSII and LHCII and diminishing fluorescence.

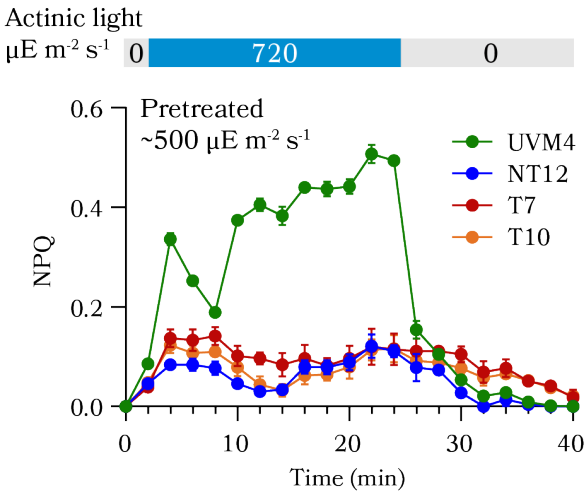


Figure 5.6: **Impaired NPQ after high light treatment.** NPQ response to incubation for 22 min at 720 μE m^{-2} s^{-1} blue actinic light after 4 h treatment with bright white light (~500 μE m^{-2} s^{-1}). Mutant NPQ response was severely disabled when pretreated with high light. Error bars are SD. Measurements aided by Felix Buchert.

5.3.5 Oxidation of PSI is slightly impaired in FNR-tether mutants

With several clues now pointing to a slight deficit in *pmf*, possibly caused by a defect in CET, we next aimed to measure the rate of CET directly. To do this we employed a third technique that tracks the appearance of the P700 cation using absorption changes at 705 nm minus 740 nm. Fig. 5.7A shows data recorded using a specific

light regime to assess P700 kinetics; cells were treated using 10 s cycles consisting of 9 s of low light followed by 1 s of high light and finishing with a saturating pulse to fully oxidise P700, then darkness to fully reduce P700. Note that in this experimental set up it was trivial to switch between ECS measurements and P700 measurements, enabling normalisation between P700 measurements and between strains by simply switching back to an ECS set-up and recording the response to a 6 ns flash, as described in Fig. 5.4, without having to replace or disturb the sample.

Net oxidation (an increase in the concentration of $P700^+$) is presented as a drop in signal, whereas net reduction is presented as a signal rise. Visible in Fig. 5.7A are the instances of P700 oxidation induced by the onset of high light, and the net re-reduction that occurs when the light is switched off, with the $P700^+$ signal decaying towards 0 (full reduction). When DCMU is added, full oxidation of P700 can be approximated using the extent of the signal immediately following a saturating flash. This is possible because in DCMU-treated cells, the ETC is effectively drained of electrons without contribution from PSII, allowing oxidation of almost every copy of PSI under strong illumination. Once a value for full oxidation is obtained, the signal trace in the presence of DCMU is then used to normalise the trace in the absence of DCMU, expressing any change in the concentration of $P700^+$ relative to the values for full oxidation and full reduction.

To compare the FNR-tether mutants with UVM4, cells were placed into anoxia to enhance CET, with the intention of making any differences between strains easier to spot. Fig. 5.7A, (i) shows the P700 signal in the presence of DCMU and HA in the dark immediately following the saturating pulse. Data for multiple biological replicates are shown (N = 5 for UVM4, N = 7 for mutants, mean ± SD plotted), and revealed no clear difference between the FNR-tether mutants and UVM4. Since the only electrons available to re-reduce P700 in the dark must come from the stroma when PSII is inactivated by DCMU, this observation implies that the rate of CET is similar between strains.

However, in the absence of DCMU and HA, there was an observable difference between strains. In this case, the response of the FNR-tether mutants to 1 s of high light treatment was noticeably smaller than in UVM4. Fig. 5.7B and C show the signal changes averaged from across biological replicates (mean ± SD), each of which consists of the means of three technical replicates. Firstly, it is clear from Fig. 5.7B that each strain has a slightly different steady state level of oxidised P700 prior to the onset of high light, with UVM4 being the most oxidised in low light (~0.21 r.u.). All strains then show a fast net oxidation when high light is initiated, which decays quickly to a steady state within ~200 ms. The FNR-tether mutants show a smaller peak in oxidised P700 than UVM4 (~0.2 r.u for tethered-only, ~0.24 r.u. for native/tethered, ~0.35 r.u. for UVM4). We highlight this further in Fig. 5.7C, in which the oxidation curves induced by high light have been aligned to 0 at the onset of high

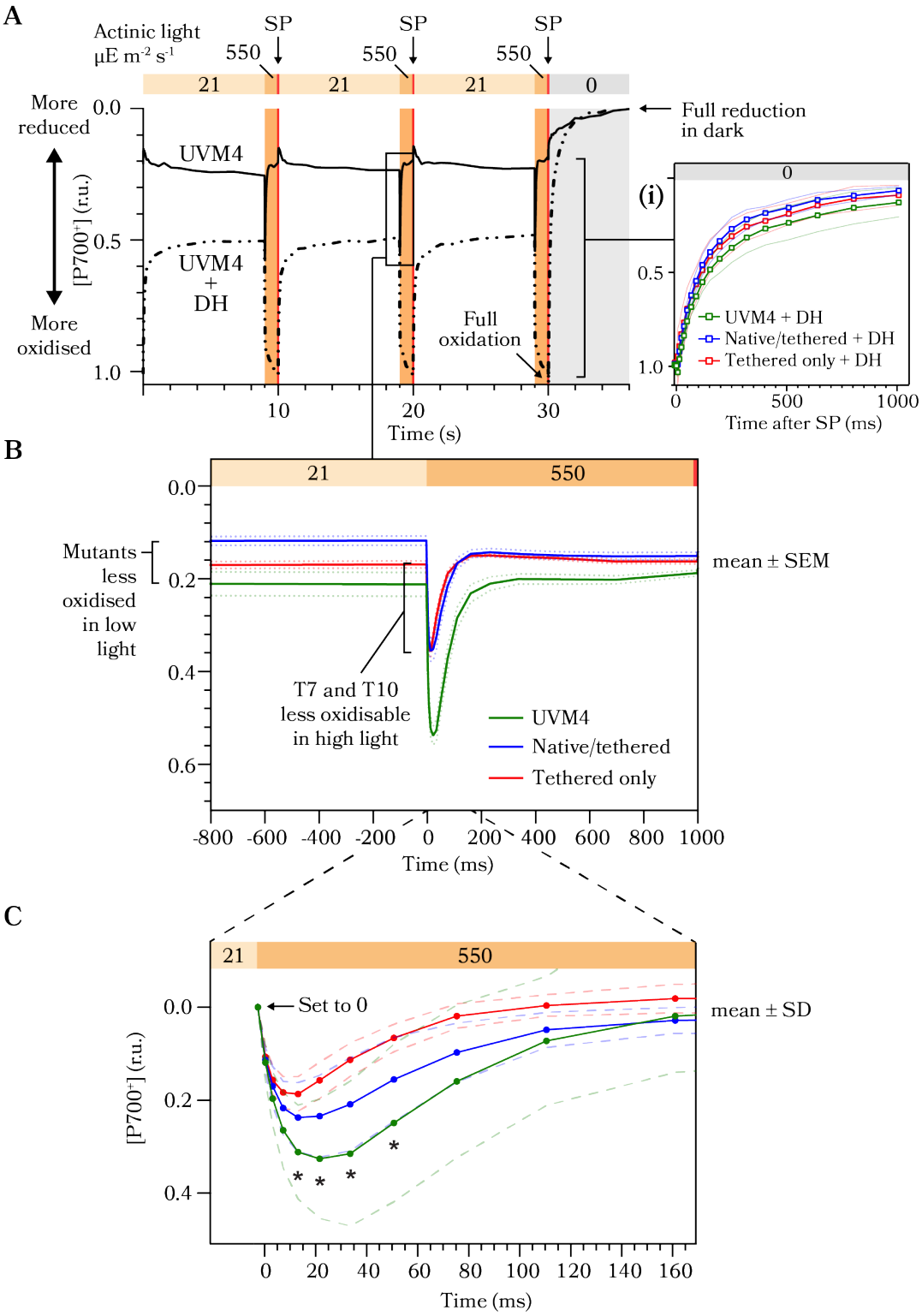

A
Actinic light
µE m-2 s-1
SP
550
21
0
More reduced
More oxidised
[P700+] (r.u.)
UVM4
UVM4 + DH
Full oxidation
Full reduction in dark
(i)
UVM4 + DH
Native/tethered + DH
Tethered only + DH
Time (s)
Time after SP (ms)
B
Mutants less oxidised in low light
T7 and T10 less oxidisable in high light
UVM4
Native/tethered
Tethered only
mean ± SEM
Time (ms)
C
Set to 0
mean ± SD
Time (ms)

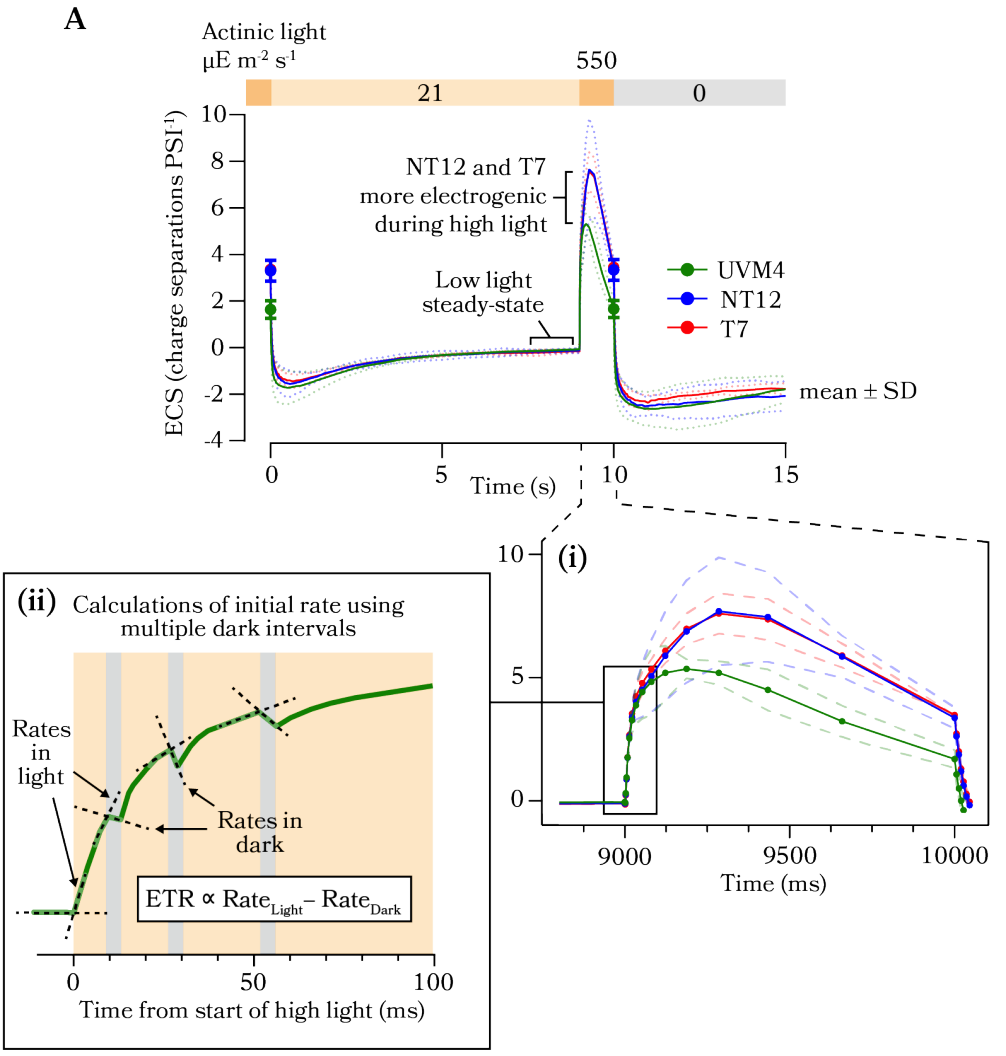
Figure 5.7: **P700 oxidation kinetics during 1 s high light and re-reduction kinetics in the dark. A.** Example P700 traces from UVM4 cells in the absence and presence of DCMU and HA demonstrating three 10-s light treatments, each consisting of 9 s of low light and 1 s of high light, ending with a saturating pulse of light. Traces are normalised between full oxidation and full reduction. After the third pulse, actinic light was removed and dark re-reduction of P700 was recorded. (i) shows a zoomed view of dark re-reduction for UVM4, NT4, NT12, T7 and T10 when treated with DCMU and HA to remove the contribution of electron flow from PSII. Means ± SD are plotted. DH = 10 μM DCMU and 1 mM HA. **B.** A zoomed view of oxidation caused by the onset of high light in UVM4, NT4, NT12, T7 and T10. The different y-values of each mutants during the moments before high light indicates slightly different steady-state oxidation states. **C.** A further zoom of the oxidation data in panel **B**, with traces now aligned to 0 at the onset of high light. UVM4 has a significantly greater response to high light than the mutants. No sig difference between native/tethered and tethered-only. Method development and measurements were made in collaboration with Felix Buchert.

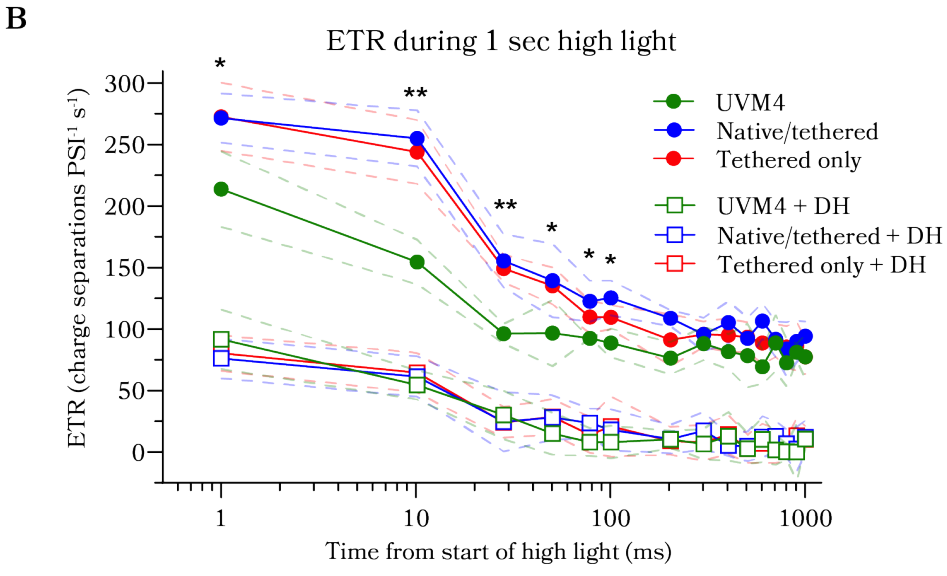
light to make comparison easier. The difference in oxidation, which peaks between ~15 to 50 ms after high light is initiated, is significantly different between UVM4 and the tethered-only mutants but not for the native/tethered mutants (one-way ANOVA with Tukey's posthoc HSD test, $N = 5$ for UVM4, $N = 7$ for mutants, $p < 0.05$ where indicated; normality was confirmed with a Kolmogorov-Smirnov test). These data indicate that the tethered-only mutants have a slightly diminished capacity to oxidise PSI at the onset of high light. One reason for this might be a higher rate of photosynthetic electron flow, meaning more electrons arriving at the donor side of PSI during high light, preventing wild type levels of P700 oxidation.

5.3.6 FNR-tether mutants show a faster initial rate of electron transfer

To complement our P700 oxidation data, we returned to measurements of the ECS which would enable us to quantify the photosynthetic electron transport rate (ETR). Using a similar light regime protocol to Fig. 5.7 we aimed to compare the ECS response between strains during 1 s of high light. The measurements were normalised using a flash-induced ECS pulse taken after the measurements were completed. Fig. 5.8A shows the ECS data from this 1-s high light protocol compared between UVM4, NT12 and T7 (mean ± SD). Though NT12 and T7 were highly similar, the results show an interesting difference in the generation of $\Delta\psi$ between the FNR-tether mutants and UVM4 during the 1 s of high light (expanded in Fig. 5.7A, (i)). The height of the A-phase and B-phase are both increased in the FNR-tether mutants, suggesting an increased, though short-lived, capacity for photosynthetic electron flow.

To pursue this, we made a modified series of ECS measurements during 1 s of high light. These measurements were carefully aligned with a series of brief (5 ms)


A
Actinic light
μE m-2 s-1
21
550
0
ECS (charge separations PSI-1)
NT12 and T7
more electrogenic
during high light
Low light
steady-state
UVM4
NT12
T7
mean ± SD
Time (s)
(i)
Time (ms)
(ii) Calculations of initial rate using
multiple dark intervals
Rates
in
light
Rates in
dark
ETR ∝ RateLight – RateDark
Time from start of high light (ms)


B
ETR during 1 sec high light
ETR (charge separations PSI-1 s-1)
UVM4
Native/tethered
Tethered only
UVM4 + DH
Native/tethered + DH
Tethered only + DH
Time from start of high light (ms)

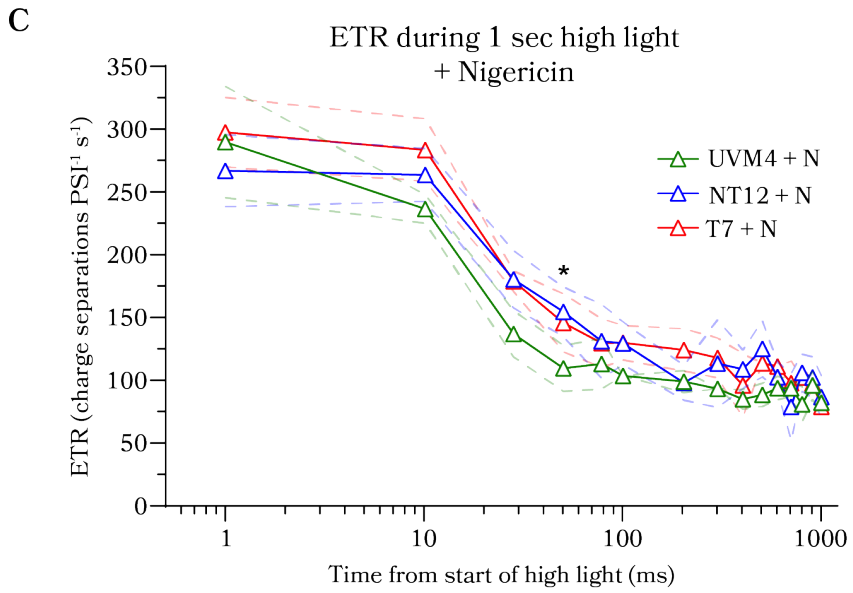


Figure 5.8: **Differences in ECS response to 1 s high light between UVM4 and FNR-tether mutants. A.** ECS traces for UVM4, NT4, NT12, T7 and T10 during a 10-s period of light (9 s low light, 1 s high light, followed by 5 s darkness. (i) shows a zoomed view of the ECS signal during 1 s high light, indicating the FNR-tether mutants generate a larger *pmf* than UVM4, possibly due to an increased rate of electron transfer (ETR) brough about by malfunctioning photosynthetic control. (ii) describes a method for ascertaining ETR using short dark intervals during the initial ~100 ms of high light, explained in the text. **B.** ETR rates during 1 s of high light calculated using the method shown in panel **A** (ii), in the absence and presence of DCMU and HA. UVM4 ETR was significantly lower than either mutant genotype for the first ~100 ms ($p < 0.05$ or 0.01 where indicated). The x-axis is scaled to $\log_{10}$ to provide higher resolution of the early timepoints. DH = 10 μM DCMU and 1 mM HA. **C.** A repeat of the experiment presented in panel **B** using only UVM4, T7 and NT12, this time in the presence of 100 μM nigericin. UVM4 ETR was still significantly lower than T7 or NT12 at ~50 ms ($p < 0.05$ where indicated). Method development and measurements were made in collaboration with Felix Buchert.

dark intervals placed at defined intervals throughout the 1 s of high light; at 10 ms, 30 ms, 50 ms, 70 ms, etc. (illustrated in Fig. 5.8A, (ii)). These short dark intervals were used to record the initial rate of decay (i.e., the gradient of the ECS signal) in the 4 ms preceding the dark interval (light) and 4 ms during the dark interval (dark). By subtracting the gradient value obtained in the light minus the dark, a value was obtained proportional to the ETR at each chosen timepoint throughout the 1 s high light (Nawrocki et al. 2019a). Fig. 5.8B plots the rate values derived from this method for anoxic cells of strains UVM4, NT4, NT12 (combined as native/tethered), T7 and T10 (combined as tethered-only), both in the presence and absence of DCMU and HA (mean ± SD).

In the presence of DCMU and HA, the ETR dropped precipitously in all strains and was in good agreement with previously published work using a similar protocol (Nawrocki et al. 2019a). However, no difference could be observed between FNR-tether mutants and UVM4, implying that the significant ETR difference in Fig. 5.8B requires electrons from PSII. This may be further evidence that CET functions similarly in the FNR-tether mutants and UVM4. In the absence of DCMU and HA, however, the results show a significantly higher ETR in the first ~100 ms of high light in both the native/tethered mutants and the tethered-only mutants compared with UVM4 (Student's t-test, $N = 4$ for UVM4, N = 7 for mutants, $p < 0.05$ or < 0.001 where indicated, normality confirmed with a Kolmogorov-Smirnov test). This is consistent with previous research showing that cells in state II have a higher initial rate of electron flow (Buchert et al. 2022), but also led us to reason that the FNR-tether mutants may have a defect in regulating ETR, possibly due to less photosynthetic control. Photosynthetic control is triggered by ΔpH, so a loss of photosynthetic control would be consistent with a lower *pmf* in high light.

To test this theory, we repeated the ETR measurements in the presence of the inhibitor nigericin. Nigericin is thought to dissipate ΔpH by forming K^+/H^+ exchanging pores in the thylakoid membrane (Walters and Horton 1991), so should cause proton leak through the thylakoids and raise the lumenal pH, in effect preventing cells from inducing photosynthetic control. If photosynthetic control was damaged in the mutants, we predicted that nigericin would have a smaller effect on ETR in the FNR-tether mutants than in UVM4. Fig. 5.8C shows the ETR in the presence of nigericin for UVM4, NT12 and T7. While UVM4 still showed a trend of slightly lower rates, which was significant at ~50 ms (Student's t-test, $N = 3$, $p < 0.05$ where indicated), the measured rates were far more similar between strains than in the absence of nigericin (Fig 5.8B).

5.3.7 FNR-tether mutants are acceptor-side limited in high light

We predicted that a loss of photosynthetic control could mean lower donor side limitation and higher acceptor side limitation of PSI. This is because faster flow

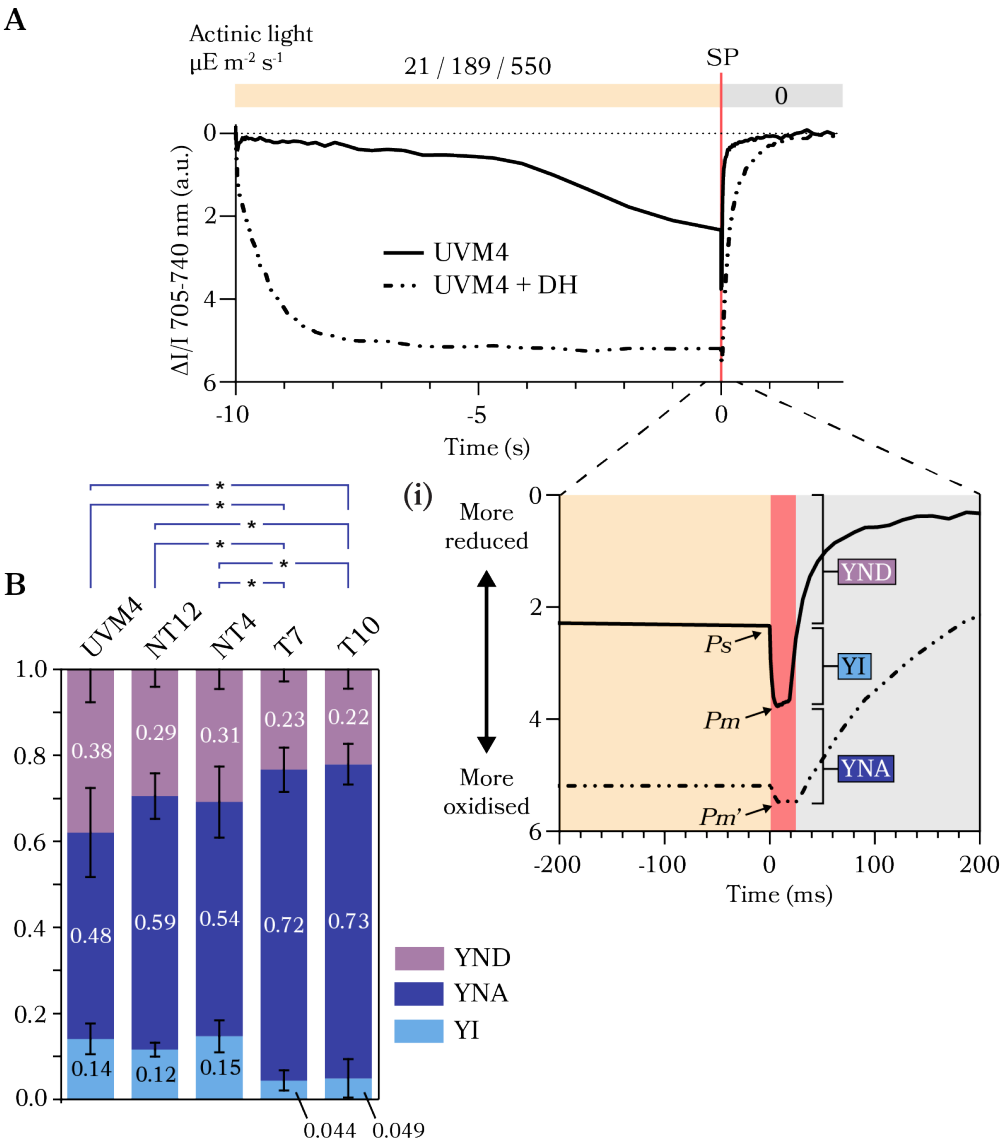


Figure 5.9: **Tethered-only mutants show higher acceptor side limitation than native/tether mutants or UVM4. A.** An example P700 oxidation trace from UVM4 in the absence and presence of DCMU and HA. Here, cells were incubated in light (low, medium or high) for 10 s before application of a saturating pulse to fully oxidise P700, then darkness to fully reduce P700. (ii) shows a zoomed view of the 700 kinetics surrounding the saturating pulse. The relative y-values in the absence and presence of DCMU can be used to estimate YND, YI and YNA as shown and described in the text. DH = 10 µM DCMU and 1 mM HA. **B.** Estimates for YND, YI and YNA calculated for UVM4, NT4, NT12, T7 and T10 using the method in panel A (i) after incubation for 10 s in high light (550 µE m^{-2} s^{-1}). T7 and T10 show significantly high YNA than UVM4, NT4 and NT12 ($p < 0.05$ where indicated). Error bars are SD. Method development and measurements were made in collaboration with Felix Buchert.

of electrons through cytb_6f to PSI implies an abundance of PC^- to reduce $P700^+$, shifting the problem onto the PSI acceptor side, where too many electrons exiting the PSI acceptor side could be overwhelming the stromal acceptor pool, leading to over-reduction. To explore this, we aimed to make a final set of P700 measurements that would allow estimation of donor and acceptors side limitation. Using the JTS-150 once more, we prepared a modified light treatment consisting of 10 s light followed by 2.5 s dark, then a saturating pulse to fully oxidise P700. The experiment was repeated on each strain four times, with each batch of cells per strains subjected to three consecutive measurements, first with 10 s low light treatment (21 μE m^{-2} s^{-1}), then medium light (18 μE m^{-2} s^{-1}) and high light (550 μE m^{-2} s^{-1}). Fig. 5.9A shows an example using this protocol, displaying the result from a sample of UVM4 cells measured first in the absence then the presence of DCMU and HA. Here, as before, the measurements were normalised using a flash-induced ECS pulse taken after both measurements were completed.

The example in Fig. 5.9A, (i) shows the key features of saturating-pulse induced oxidation changes that can be disentangled to calculate estimations for donor and acceptor side limitation (Klughammer and Schreiber 1994). The point *P*s refers to the steady-state P700 signal in the light just before application of the saturating pulse. The fact that *P*s does not equal 0 means that there is some concentration of P700+ present in the sample even during illumination, implying that the supply of electrons from cytb_6f (via PC) is insufficient to fully reduce the P700 in the cell. This is a measure of donor side limitation, termed YND. The difference between *P*s and the maximum oxidation level induced by a saturating pulse (*Pm*) is proportional to the total oxidisable PSI in the sample. This is a measure of the quantum yield of PSI, termed YI. The fact that *Pm* is not equal to the maximum oxidation signal in the presence of DCMU, *Pm'*, means that there is a population of PSI in the sample that cannot be oxidised even by a saturating pulse of actinic light. The difference between *Pm* and *Pm'*, then, is related to the proportion of closed PSI reaction centres in the sample, which implies a shortage of stromal electron acceptors to enable electron transfer from P700 to Fd. This is a measure of acceptor side limitation, termed YNA. These three values, YND, YI and YNA, are proportions that together equal 1 (Schreiber and Klughammer 2016).

Using this method, we quantified YND, YI and YNA for UVM4, NT12, NT4, T7 and T10. Data acquired from measurements made after treatment with low and medium light showed no clear trend (data not shown). However, after treatment with high light for 10 s, the tethered-only mutants showed significantly more YNA than UVM4 or the native/tethered mutants (Student's *t*-tests, $N = 4$, $p < 0.05$ where indicated) (Fig. 5.9B). T7 and T10 also showed significantly less YND than UVM4 ($p < 0.05$), and T7 (but not T10) showed significantly less YND than both NT4 and NT12 ($p < 0.05$). These trends are consistent with a high ETR in the FNR-tether mutants, possibly caused by mis-regulated electron flow, leading to over-reduction of

the stroma and higher YNA.

5.4 Discussion

5.4.1 Do we see evidence for impaired CET in the tethered-only mutants?

The first key question is whether the data presented here constitute evidence for damaged CET in the tethered-only mutants. In our view, only two clues from the presented evidence support this hypothesis.

1. A significant increase in acceptor side limitation is seen for the tethered-only mutants, compared with both UVM4 and with native/tethered mutants (Fig. 5.9B). This is the firmest spectroscopic confirmation that our mutants deficient in native FNR behave differently to those with both native and tethered FNR. A higher acceptor side limitation could indeed be caused by the loss or depletion of a stromal electron sink, which, in anoxia, could be CET. However, the effect is only observable in high light, and we might expect a similar effect across all light intensities if CET was significantly impaired.

2. Our FNR-tether mutants both showed a similar ETR to UVM4 when PSII was blocked by DCMU, but a higher ETR when it was not (Fig. 5.7B). As a proportion of total ETR, then, the ETR in the presence of DCMU is lower in our mutants than in UVM4 during the first 100 ms of high light. This method has been used recently to disentangle the rate of CET from LET (for example in Figure 5, Supplement 1 in Kramer et al. 2021). If valid, we could then say that CET is proportionally lower in the FNR-tether mutants than UVM4. However, this assertion is based on the shaky assumption that poisoning PSII allows one to access the full rate of CET without interference from LET. This is thought to be false (Nawrocki et al. 2019b), since electrons in CET are constantly lost to other stromal processes, meaning electrons from LET are always required to sustain the maximal rate of CET.

Conversely, then, do we have enough evidence to reject our hypothesis that CET is disturbed by deleting native FNR? Within the confines of our experimental system (i.e., in the presence of the PSI-FNR fusion complex instead of native PSI), there are three main pieces of evidence that recommend a rejection of the alternative hypothesis.

1. Though our FNR-tether mutants appeared to have a trend of lower *pmf* than UVM4 in high light, these data were far less dramatically different from the results obtained for *pgr5* relative to its wild type, t222+ (Fig. 5.3C). Taken alone, these data support the recognition that our mutants do not show altered

photosynthetic electron flow in the same way as brought about by the deletion of PGR5, suggesting, at least, that CET is not impacted to the same extent.

2. We observed no difference in the rate of $P700^+$ re-reduction in the dark (Fig. 5.7A, (i)). Using absorption spectroscopy to approximate the concentration of $P700^+$ in our cells, the decay (re-reduction) of $P700^+$ in the dark when DCMU was present should be proportional to the rate of CET, since only cycled electrons can reduce the PSI donor side when PSII is poisoned. That we see no clear difference strongly implies that CET functions at a similar rate in both FNR-tether genotypes, again suggesting that stromal electrons can still enter the PQ pool without significant impedance.

3. ETR patterns in the presence of DCMU were highly similar between all strains in anoxia (Fig. 5.8B). Though not explicitly indicative of a normal CET phenotype (for example, Chlamydomonas cells deficient in PGRL1 also show similar rates to wild type in Nawrocki et al. 2019a), a similar ETR under these conditions certainly implies that our mutants can still perform the FQR reaction. This therefore stands as strong evidence that FNR is not the FQR, or that our PsaF-FNR fusion protein can still perform FQR with the same rate as native FNR. This second option seems unlikely, but it also cannot yet be ruled out that some PsaF-FNR remains soluble, or that some very low level of native FNR is still accumulated in our tethered-only mutants that facilitates native-like FQR activity.

Together, then, these three key observations recommend the acceptance of the null hypothesis and a recognition that tethering FNR to PSI has not resulted in an inability to perform CET in our mutants. If this is true for wild type cells as well, then these results provide strong evidence that FNR is not the FQR.

5.4.2 What can be inferred about photosynthesis in our FNR-tether mutants?

The effects of our FNR-tether mutation and the subsequent deletion of native FNR remain somewhat mysterious. Rather remarkably, the similar measured ratios of open PSI:PSII reaction centres (Fig. 5.4), and indeed the generally small effects on electron flow offer tantalising clues that our PSI-FNR fusion complex is able to catalyse PC^--to-$NADP^+$ electron transfer. Nonetheless, some of the uncertainty associated with the FNR-tether mutation as mentioned in the previous chapter can be reiterated here: without further work we cannot be confident that every copy of PSI contains the PsaF-FNR peptide or that every copy of PsaF-FNR is integrated into a mature, functional PSI complex. Moreover, without structural clues, we cannot assert with confidence that our tethered FNR is physically prevented from

interacting with cytb_6f. There is a possibility, for example, that the putative PSI-cytb_6f supercomplex (Iwai et al. 2010, Steinbeck et al. 2018) is still able to form in our tethered-only mutants, and that the FAD moiety within tethered FNR is able to achieve electron transfer to haem c_i of cytb_6f. Though we did not pursue isolation of the PSI-cytb_6f supercomplex from our mutants due to time constraints, in vitro biochemical and structural data from a purified supercomplex would make a valuable addition to this data set and may help to clarify the mechanism of CET.

However, with the additional data presented here combined with that in Chapter 4, we can start to form a picture of what may be happening to photosynthesis in these mutants. The relatively high intensity of the PSI -associated peak in our 77 K fluorescence data suggest that the FNR-tether mutants are pushed further into state II than UVM4 in both oxia and anoxia. Proteomics results presented in Fig. 4.8 of Chapter 4 and ECS measurements (Fig. 5.4) strongly suggest that this is not due to a difference in the ratio of PSI:PSII reaction centres in each cell and is, in fact, due to state transitions. In agreement with a state II phenotype, we see some evidence for lower *pmf* at high light (Fig. 5.5B), which we find likely to be due to a lower proton influx (v_{H^+}), rather than a problem with ATP synthase conductivity (g_{H^+}) (Fig. 5.5D and E). The striking inability of the FNR-tether mutants to perform NPQ after 4 h high light treatment (Fig. 5.6) indicates a specific high light sensitivity that may be furtherly detrimental to the cells' ability to generate ΔpH, which is a trigger for NPQ. This high light sensitivity implies a photoinhibited phenotype, consistent with the growth phenotype under fluctuating light observed in Fig. 4.6 in Chapter 4, and consistent with the previously observed photoinhibited phenotype of FNR-depleted mutant tobacco (Palatnik et al. 2003). Next, using P700 oxidation and ECS measurements, we find no evidence of a change in CET in the presence of DCMU, but two pieces of evidence to suggest the rate of electron flow (ETR) is higher in our FNR-tether mutants (Fig. 5.7C and Fig. 5.8B), and that this could be explained by a loss of photosynthetic control. This theory was partially supported by preliminary results gathered in the presence of the inhibitor nigericin (Fig. 5.8C). As expected, in the presence of nigericin, ETR in UVM4 was faster, closing the gap between UVM4 and the FNR-tether mutants to a noticeable extent. The fact that UVM4 still showed a trend of lower ETR in the presence of nigericin implies that another factor is at play, but unfortunately time constraints on the project prevented us from following up these measurements with more repeats. Work is currently underway in collaboration with Dr. Buchert and Prof. Hippler to clarify these results. Lastly, a loss of photosynthetic control implies lower YND and higher YNA, which we do indeed see using the P700 oxidation data presented in Fig. 5.9. Notably there was no trend in YNA or YND after low light treatment, but a significant increase in YNA and YND after as little as 10 s high light.

Two key experiments stand out from this line of evidence: (i) in the presence of

DCMU, we see no difference in ETR or P700 reduction, suggesting the FNR-tether mutation does not significantly disrupt CET; and (ii) in the presence of nigericin, we mostly eliminate the ETR difference between strains, hinting that the FNR-mutants have a problem with photosynthetic control.

5.4.3 Why might photosynthetic control be impaired in the FNR-tether mutants?

Photosynthetic control describes the mechanism that slows down electron flow through cytb_6f in response to low lumenal pH, acting as a form of negative feedback regulation on high levels of photosynthetic electron flow. In our view, impaired photosynthetic control helps explain many of the differences seen between UVM4 and the FNR-tether mutants and could be part of the physiological consequences of tethering FNR to PSI. Assuming electrons emerging from PSI are still accessible to the primary electron acceptor Fd, then fusing FNR to PSI may work to increase the concentration of Fd in the vicinity of FNR. Alternatively, or perhaps in parallel, a tethered FNR may show different rates of forward and reverse reaction. Recent research in angiosperms, which contain multiple FNR isoforms, has demonstrated that different isoforms are tethered with different affinities to the thylakoid membrane-bound anchors Tic62 and TROL (Goss and Hanke 2014). Deletion of only a weakly bound or strongly bound isoform was suggested to disrupt the balance between forward and reverse reactions of FNR, suggesting that strongly tethered FNR may perform more of the reverse reaction (Goss and Hanke 2014). For example, Arabidopsis FNR has been shown to bind more strongly to the membrane in the dark, perhaps suggesting dynamic localisation based on the type of metabolism required by the cell (Benz et al. 2009). By forcing FNR near to PSI, we may have similarly disturbed the balance of the rates of forward and reverse reaction such that cells accumulate too much NADPH in the stroma. Furthermore, by then deleting native FNR in the tethered-only mutants we may have exacerbated the problem. Whether by an increased affinity for Fd or by upsetting the balance between forward and reverse reactions of FNR, the effect could be the overproduction of NADPH and the over-reduction of the stromal NADP pool. This would be consistent with an upregulation of NADPH-consuming processes in the stroma, possibly reflected in the increase of NADP-MDH seen in T7 and T10 compared with NT12 visible in the proteomics data presented in Fig 4.8 in Chapter 4 on page 104.

Previous work has shown that the NADP pool redox poise could be a triggering factor for feedback regulation of photosynthetic control (Hald et al. 2008), although the mechanism is still unclear. Using Arabidopsis depleted in either FNR or GAPDH (the primary consumer of NADPH in the CBB cycle), the authors reported a further interesting result wherein GAPDH deficient plants did not appear to suffer

stress caused by too much NADPH due to a feedback mechanism triggered by an unbalanced NADP pool (Hald et al. 2008), whereas FNR depleted mutants were proposed to lack this feedback (Joliot and Johnson 2011). This could also be the case in Chlamydomonas, with our tethered-only cells suffering from over-production of NADPH due to the tethered FNR and lack of native FNR to provide feedback regulation.

So, what might the consequences be of producing too much NADPH? The first might be to stimulate a transition to state II. At low light, too much NADPH might either over-reduce the electron transport chain carriers or might stimulate more NDH-CET via the NADPH:PQ reductase, NDA2. Either way, this could result in a more reduced PQ pool, activating STT7, which would then phosphorylate LHCII and trigger a migration of antenna complexes from PSII to PSI. The effects of state II in low light would be a stimulation of PSI photochemistry, increasing CET and producing extra ATP. The extra ATP helps counterbalance the ATP:NADPH ratio by boosting CBB cycle rate, consuming excess NADPH. If this efficient system of regulation proceeds in our FNR-tether mutants, then we would not expect to see an effect at low light. Indeed, we do see no differences in the photosynthetic parameters presented above when using low light intensities. In high light, though, an overproduction of NADPH would have worse consequences. When light is not a limiting factor then the cell's capacity to use stromal electrons becomes the bottleneck, with the limiting step likely being the carboxylation reaction of Rubisco. In this case, a transition to state II would not help, as stromal over-reduction would begin to clog-up the ETC, leading to a lack of proton translocation despite the larger PSI antenna. Couple this with lower photosynthetic control, and the problem would be exacerbated, with over-reduction being arrived at even faster. This may be reflected in the YNA calculations after just 10 s high light. Reduction of the ETC would slow down electron transfer, meaning *pmf* would also start to drop and LET would decrease. A scenario drawing together these ideas is shown as a guide in Fig. 5.10. This scenario might explain why the native/tether mutants grow slower generally and suffer worst at high light (Fig. 5.5 and Fig. 4.6 in Chapter 4). The tethered-only mutants appear to suffer to an even greater extent, possibly suggesting that the extra native FNR in the native/tether mutants was helping to partially alleviate the problem. This would agree with previous work on FNR-depleted Arabidopsis, which found that photoprotective mechanisms were permanently induced in the mutants (Lintala et al. 2012). Finally, though we did not measure it, we might expect an upregulation of the CCM in our mutants in high light as the cells attempt to clear the stromal overreduction by speeding up Rubisco.

So, what can this tell us about FNR in wild type cells? If the cause of an unbalanced NADP pool in our mutants really is due to the tether mutation, this may tell us that

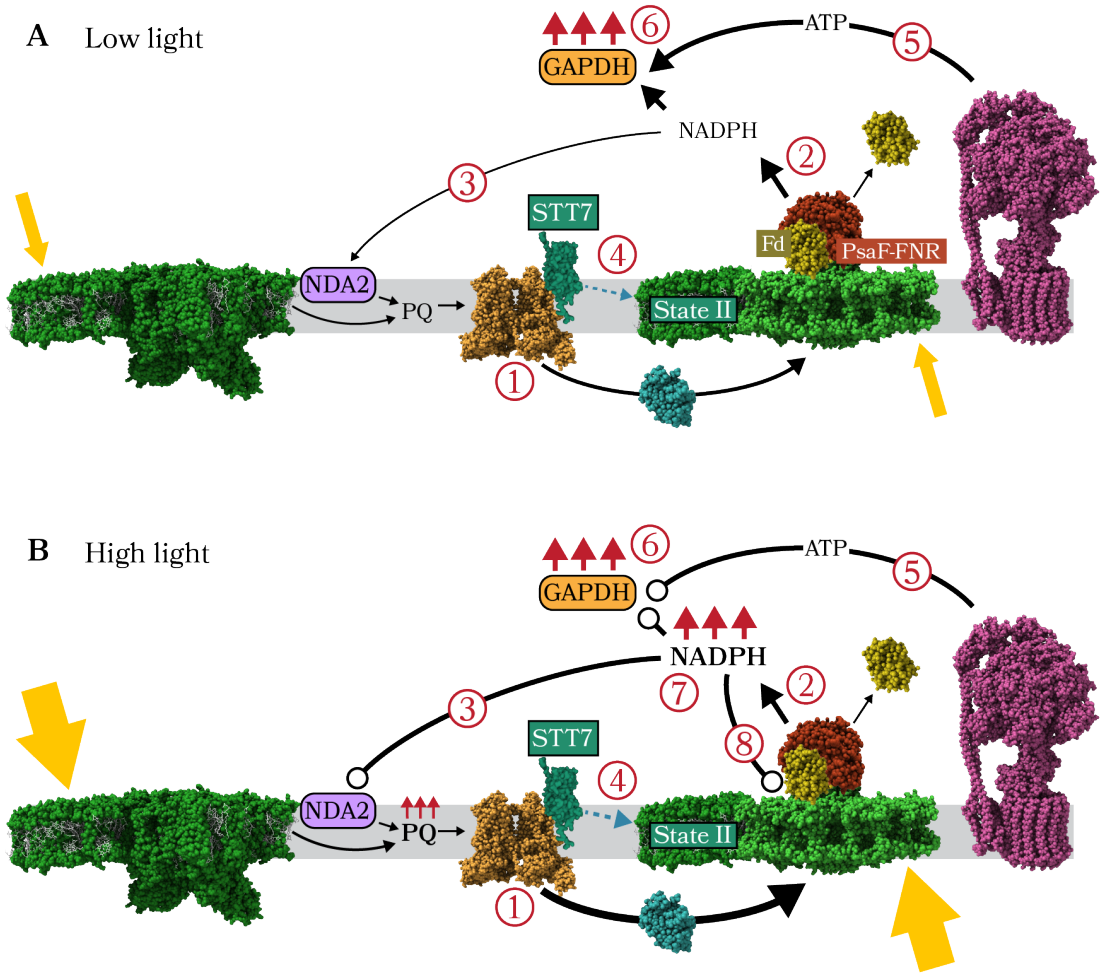


Figure 5.10: **Possible explanation for the high light sensitivity of the FNR-tether mutants based on impaired photosynthetic control and overproduction of NADPH. A.** In low light: (i) low photosynthetic control leads to low YND and high electron flow to PSI; (ii) tethered FNR causes overproduction of NADPH at the expense of reduced Fd; (iii) NADPH causes reduction of the PQ pool either by NDH-CET through NDA2 or by backing up the ETC; (iv) PQ pool reduction activates STT7, which triggers transition to state II; (v) state II increases ATP production via higher PGR-CET (data not shown); (vi) the extra ATP is used in the CBB cycle to fix CO_2, consuming the excess NADPH in the production of glyceraldehyde 3-phosphate by GAPDH. **B.** In high light, points (i) to (v) from panel **A** apply as before. Low photosynthetic control leads to low YND, while tethered FNR causes an overproduction of NADPH. Excess NADPH overwhelms the NDH-CET pathway (circle-ended arrows) and causes over-reduction of the PQ pool (red arrows), driving the transition to state II and triggering extra ATP production. This time, light is not a limiting factor, so upregulation of the CBB cycle reaches a plateau and becomes limiting. (vii) Excess NADPH is not consumed and builds up in the stroma, leading to (viii) eventual photoinhibition of the acceptor side of PSI, while backing up the ETC, over-reducing the PQ pool and leading to an eventual drop in photosynthetic efficiency and low *pmf*. Protein structural models are the same as those shown in Figure 1.2 on page 7 and Figure 1.7 on page 28.

tethered FNR outcompetes the other Fd-consuming electron sinks in the stroma, e.g., the hydrogenase HYDA or the Flv proteins. If this scenario is accurate, it would imply that a broader range of electron sinks in wild type cells enables them to oxidise the NADP pool, achieving redox poise and preventing stromal over-reduction and photoinhibition. Regulation of the NADP pool by FNR could then be seen as a crucial feature of photosynthetic regulation, with a lack of oxidised NADP having far-reaching consequences, including for photosynthetic control.

Designing experiments to measure CET in Chlamydomonas is difficult and requires careful measurement. Though care was taken to ensure reproducibility between replicates for the experiments presented here, small sample numbers mean caution must be used when drawing conclusions. This caution should apply particularly to P values derived from t-tests using less than five data points, since this is often not enough to establish whether the data are normally distributed. Additionally, large values for standard deviation may reflect uncontrolled variation in how the cells were prepared for measurement, for example due to differences in cell cycle position when measuring constant-light grown cells versus day/night synchronised cells. Time constraints on the project meant we could not supplement these measurements using additional techniques that could help us further understand the impact of the FNR-tether mutation on photosynthesis, including measuring O_2 evolution, CO_2 fixation, CCM induction and measurements of the redox state of the NADP pool. Nevertheless, if the data presented here are accurate, they not only represent a promising starting point for further work on these mutants, but they also highlight the importance of FNR for regulation of redox poise in the stroma, without which photosynthetic regulation cannot function properly.

Chapter 6

Concluding remarks

6.1 Thesis summary

The aims of this thesis were the development and use of genetic tools to answer mechanistic questions about CET. In Chapter 3 we described the successful adaptation of recombineering for use with Chlamydomonas genes, resulting in the cloning of several genes that have not been previously cloned or characterised. Including the native promoter region in cloned genes allowed native-like localisation and expression patterns in the case of the CCM component LCIB. Rescue of mutants using recombineered genes was also demonstrated, first for a CLiP mutant deficient in LCIB and then a CRISPR-generated mutant deficient in STT7. We demonstrated the efficacy of a four-day recombineering pipeline for cloning ~200 genes or regions in parallel and made available the tools and methods for the research community. This method was published in The Plant Cell in January 2021, and the publication is included in this thesis as an appendix. The publication includes links to extensive supplementary resources that can help any lab with basic biochemical equipment to set up a recombineering pipeline. The timing of the publication is complementary to the recent release of the Chlamydomonas fosmid and CLiP mutant libraries housed at the University of Minnesota (Li et al. 2019), as well as the refinement of Arabidopsis recombineering tools (Brumos et al. 2020). We designed the method to be robust and easy to perform. Though our success rates were not as high as those reported for *C. elegans* or Arabidopsis (Brumos et al. 2020, Sarov et al. 2012), we have since observed a higher success rate and more reliable results when recombineering genes in smaller batches, possibly due to the increased manageability of small sample sizes. It is hoped that use of the method will prevent length or complexity of any genes of interest from becoming a sticking point for future studies. Avoiding the delays and expense that accompany cloning problems, such as those reported for the cloning of the CCM gene SAGA1 (Itakura et al. 2019), would be a valuable boon to project efficiency. The usefulness of the method for the wider community will hopefully be proven in the long term by opening new avenues of scientific enquiry.

One such avenue is represented by the fledgling biochemical work we presented on the state transition kinase STT7. Downstream editing of recombineered STT7 to switch the affinity tag and promoter enabled Strep-based purification of STT7 from a UVM4 background. The STT7 kinase is a central regulatory protein in photosynthesis, providing the switch for a transition to state II (Depège et al. 2003) and possibly involved in the upregulation of CET (Buchert et al. 2022) and the CCM through phosphorylation of CAH3 (Blanco-Rivero et al. 2012). There is a pressing need to understand the activation of this protein, yet STT7 and its interaction with cytb_6f remain structurally uncharacterised from Chlamydomonas. The successful cloning and purification presented here open the door to possible future experiments such as co-purification of STT7 and cytb_6f, *in vitro* interaction studies with STT7 and cytb_6f, or interaction data using proximity labelling with tagged STT7.

In Chapter 4 we support recent publications showing that CRISPR-guided mutagenesis can work efficiently in Chlamydomonas. We show that deleting native FNR is possible after tethering FNR to PSI, and that our tethered FNR is cautiously presumed to be functional in LET. Experiments in Chapter 5 appeared to confirm that LET functions in the absence of native FNR, with similar photosynthetic parameters to wild type cells. We rejected our hypothesis that CET is disrupted in the tethered-only mutants, which are depleted in native FNR, with the caveats that further work is needed to show that the PsaF-FNR fusion protein cannot interact with cytb_6f. Small differences were observed in P700 oxidation and electron transfer rate in our mutants, with evidence pointing to an acceptor side limited phenotype under high light. We theorised this was due to an unbalanced NADP redox poise in the stroma, possibly caused by a distortion of the ratio of forward and reverse reactions of FNR caused by its tethering to PSI. These results cemented the reputation of FNR as a crucial maintainer of homeostasis. Accordingly, FNR is difficult to engineer in higher plants without severe phenotypes (Lintala et al. 2012), so the FNR-tether mutants described in Chapter 4 and 5 represent an exciting springboard for future work on FNR. Such work could involve an array of tethered positions for FNR, including tethering to PSAE, thought to be the docking site of FNR to PSI during CET, or tethering to cytb_6f followed by further testing on CET rates. Perhaps of equal interest, the FNR-tether mutants described here may express a unique, modified form of PSI that can perform LET without accumulating PsaC/D/E. These clues are worth further investigation, and work on these mutants is currently ongoing, with a likely first step being repeating proteomic analysis in phototrophic conditions.

6.2 Natural next steps for identifying the FQR

Building on observations from our FNR-tether mutants, we propose two lines of investigation that could be followed to address the question of the FQR mechanism in Chlamydomonas. The first is to attempt to block stromal electron transfer to cytb_6f

directly by steric hindrance. This could be achieved by a tethered, inactive Fd or perhaps by introduction of the PetP subunit from cyanobacteria, which appears to bind to the stromal face of cytb_6f (Proctor et al. 2022) and may bind successfully with Chlamydomonas cytb_6f as well. If successful blocking could be confirmed and experiments could be correctly controlled to account for possible disruption in binding of other cytb_6f effector proteins (e.g., STT7 and PETO), this could be a powerful tool for assessing the role of cytb_6f in CET.

Secondly, we envisage building on evidence from studies of different Fd isoforms in maize, in which the major isoform, FdI, is expressed in mesophyll cells while FdII is expressed in the bundle sheath cells (Kimata and Hase 1989). The recognition that maize mesophyll cells perform mostly LET while bundle sheath cells perform mostly CET has led to the suggestion that the Fd isoforms may be specifically active in LET or CET (Goss and Hanke 2014). Accordingly, the two major Fd isoforms in Chlamydomonas, PETF and FDX2 (Cre16.g658400), may also function differently in LET and CET even within the same organelle. This would support previous research suggesting FNR binds to PETF and FDX2 with different affinities, with a higher affinities reported for PETF when bound as part of the CET-supercomplex (Mosebach et al. 2017). Additionally, cells are thought to be unable to grow in the dark without FDX5 (Cre17.g700950), suggesting a specific role for different Fds in different metabolic modes (Yang et al. 2015). Selective deletion of each Fd isoform in Chlamydomonas may provide clues as to which isoform performs the majority of CET, leading to further discoveries in the mechanism of the FQR.

6.3 The future of CET research

CET remains an enigmatic missing link in our understanding of photosynthetic electron flow. As a regulatory process it is likely crucial for boosting *pmf* to trigger the photoprotective mechanisms qE and photosynthetic control (Chaux et al. 2015, Munekage et al. 2002). As a bioenergetic process it is likely more efficient than PCEF or CMEF because it recycles electrons, perhaps helping to explain why angiosperms appear to have dispensed with the Flv proteins (involved in PCEF) during their evolution. As a mechanism, however, we are yet to identified unique intermediates of CET despite over four decades of research. Chlamydomonas is not always the ideal model organism for studying crop plant CET. The nuclear genome can be difficult to work with, due to the sequence features, discussed at length in Chapter 3. The lack of efficient homology directed repair means transformations proceed by random integration into the nucleus, risking gene fragmentation and poor expression (Zhang et al. 2014). Widespread silencing also dampens expression of transgenes or can lead to phenotype loss after multiple rounds of propagation (Neupert et al. 2020). Beyond the genome, the Chlamydomonas stromal pool of electron acceptors is different to plants, with PCEF and the Flv proteins responsible for a large

portion of electron flow, especially under low CO_2 conditions (Burlacot et al. 2022). Moreover, Chlamydomonas PSI is spectroscopically distinct to PSI in plants despite sharing many structural features. Though small, this difference makes untenable the selective excitation of PSI using far red light, a technique used to good effect when measuring CET in plants (Kramer et al. 2021).

However, as pressure grows on the global food supply and the focus on developing novel agriculture technologies sharpens, valiant efforts have been made to overcome these technical hurdles, with possible benefits stretching beyond photosynthesis research. CRISPR technologies are developing quickly (Akella et al. 2021), bringing into view the eventual goal of creating scarless, targeted insertions or deletions in the nuclear genome with high efficiency. Further work may enable serial, efficient transformation experiments targeting multiple genes or whole pathways, aided by the release of genetic technologies designed for high-throughput work, e.g., MoClo (Crozet et al. 2018). Moreover, in an interesting synthesis of CET and biohydrogen research, the discovery that mutants deficient in PGR5 stimulate hydrogen production (Steinbeck et al. 2015) has helped build interest in the Chlamydomonas stroma as a pool of potentially interesting biochemical reactions that could be harnessed to produce high value biomolecules. While the facultative heterotrophy of Chlamydomonas has long been useful for analysis of knockout mutants of the key photosynthetic complexes, plastome engineering efforts now include addition of whole new pathways to the chloroplast genome (Jackson et al. 2021). These and parallel advances in chloroplast engineering have pushed Chlamydomonas to become a biotechnological workhorse as well, opening avenues into biopharmacology and nutraceutical production (Jackson et al. 2021). Studies of the more unique features of Chlamydomonas photosynthesis are now feeding back to efforts to improve crop plants, spurred on by preliminary successes in efforts to engineer a Chlamydomonas-like CCM in Arabidopsis (Atkinson et al. 2016, Atkinson et al. 2019). The stromal acceptors not found in plants are finding uses in plant research, such as the recent success in introducing Flvs from cyanobacteria into Arabidopsis, resulting in reduced photoinhibition (Vicino et al. 2021). The spectroscopic and biochemical techniques used to measure CET in Chlamydomonas are constantly improving, showcased by a wealth of recent publications (Buchert et al. 2020, Buchert et al. 2022; Joliot et al. 2022, Nawrocki et al. 2019a). It is thus an exciting moment for bioenergetics research in Chlamydomonas.

References

Akella, S., Ma, X., Bacova, R., Harmer, Z. P., Kolackova, M., Wen, X., Wright, D. A., Spalding, M. H., Weeks, D. P., & Cerutti, H. (2021). Co-targeting strategy for precise, scarless gene editing with CRISPR/Cas9 and donor ssODNs in Chlamydomonas. *Plant Physiology*, *187*(4), 2637–2655.

Aksoy, M., & Forest, C. (2019). One step modification of Chlamydomonas reinhardtii BACs using the RED/ET system. *Mediterranean Agricultural Sciences*, *32*(1), 49–55.

Alcántara-Sánchez, F., Leyva-Castillo, L. E., Chagolla-López, A., González de la Vara, L., & Gómez-Lojero, C. (2017). Distribution of isoforms of ferredoxin-NADP+ reductase (FNR) in cyanobacteria in two growth conditions. *The international journal of biochemistry & cell biology*, *85*, 123–134.

Aliverti, A., Pandini, V., Pennati, A., de Rosa, M., & Zanetti, G. (2008). Structural and functional diversity of ferredoxin-nadp+ reductases [Special Issue: Enzymology in Europe]. *Archives of Biochemistry and Biophysics*, *474*(2), 283–291.

Allen, J. F. (1992). Protein phosphorylation in regulation of photosynthesis. *Biochimica et Biophysica Acta - Bioenergetics*, *1098*(3), 275–335.

Alric, J. (2014). Redox and ATP control of photosynthetic cyclic electron flow in Chlamydomonas reinhardtii: (II) Involvement of the PGR5-PGRL1 pathway under anaerobic conditions. *Biochimica et Biophysica Acta - Bioenergetics*, *1837*(6), 825–834.

Alric, J. (2015). The plastoquinone pool, poised for cyclic electron flow? *Frontiers in Plant Science*, *6*(7), 540.

Alric, J., & Johnson, X. (2017). Alternative electron transport pathways in photosynthesis: a confluence of regulation. *Current Opinion in Plant Biology*, *37*, 78–86.

Alric, J., Lavergne, J., & Rappaport, F. (2010). Redox and ATP control of photosynthetic cyclic electron flow in Chlamydomonas reinhardtii (I) aerobic conditions. *Biochimica et Biophysica Acta - Bioenergetics*, *1797*(1), 44–51.

Andersen, B., Scheller, H. V., & Møller, B. L. (1992). The PSI-E subunit of photosystem I binds ferredoxin:NADP+ oxidoreductase. *FEBS Letters*, *311*(2), 169–173.

Atkinson, N., Feike, D., Mackinder, L. C., Meyer, M. T., Griffiths, H., Jonikas, M. C., Smith, A. M., & Mccormick, A. J. (2016). Introducing an algal carbon-concentrating mechanism into higher plants: location and incorporation of key components. *Plant Biotechnology Journal*, *14*(5), 1302–1315.

Atkinson, N., Velanis, C. N., Wunder, T., Clarke, D. J., Mueller-Cajar, O., & Mccormick, A. J. (2019). The pyrenoidal linker protein EPYC1 phase separates with hybrid Arabidopsis-Chlamydomonas Rubisco through interactions with the algal Rubisco small subunit. *Journal of experimental botany*, *70*(19), 5271–5285.

Avenson, T. J., Cruz, J. A., & Kramer, D. M. (2004). Modulation of energy-dependent quenching of excitons in antennae of higher plants. *Proceedings of the National Academy of Sciences of the United States of America*, *101*(15), 5530–5535.

Baek, K., Kim, D. H., Jeong, J., Sim, S. J., Melis, A., Kim, J. S., Jin, E., & Bae, S. (2016). DNA-free two-gene knockout in Chlamydomonas reinhardtii via CRISPR-Cas9 ribonucleoproteins. *Scientific Reports*, *6*(1), 1–7.

Baier, T., Wichmann, J., Kruse, O., & Lauersen, K. J. (2018). Intron-containing algal transgenes mediate efficient recombinant gene expression in the green microalga Chlamydomonas reinhardtii. *Nucleic Acids Research*, *46*(13), 6909–6919.

Bailey-Serres, J., Parker, J. E., Ainsworth, E. A., Oldroyd, G. E., & Schroeder, J. I. (2019). Genetic strategies for improving crop yields.

Baker, N. R., Harbinson, J., & Kramer, D. M. (2007). Determining the limitations and regulation of photosynthetic energy transduction in leaves. *Plant, Cell & Environment*, *30*(9), 1107–1125.

Barahimipour, R., Strenkert, D., Neupert, J., Schroda, M., Merchant, S. S., & Bock, R. (2015). Dissecting the contributions of GC content and codon usage to gene expression in the model alga Chlamydomonas reinhardtii. *Plant Journal*, *84*(4), 704–717.

Bar-On, Y. M., & Milo, R. (2019). The global mass and average rate of rubisco. *Proceedings of the National Academy of Sciences of the United States of America*, *116*(10), 4738–4743.

Benjamini, Y., & 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*(1), 289–300.

Benz, J. P., Stengel, A., Lintala, M., Lee, Y. H., Weber, A., Philippar, K., Gügel, I. L., Kaieda, S., Ikegami, T., Mulo, P., Soll, J., & Bölter, B. (2009). Arabidopsis Tic62 and Ferredoxin-NADP(H) Oxidoreductase Form Light-Regulated Complexes That Are Integrated into the Chloroplast Redox Poise. *The Plant Cell*, *21*(12), 3965.

Bernard, P., & Couturier, M. (1992). Cell killing by the F plasmid CcdB protein involves poisoning of DNA-topoisomerase II complexes. *Journal of Molecular Biology*, *226*(3), 735–745.

Betti, M., Bauwe, H., Busch, F. A., Fernie, A. R., Keech, O., Levey, M., Ort, D. R., Parry, M. A. J., Sage, R., Timm, S., Walker, B., & Weber, A. P. M. (2016). Manipulating photorespiration to increase plant productivity: recent advances and perspectives for crop improvement. *Journal of Experimental Botany*, *67*(10), 2977–2988.

Bindels, D. S., Haarbosch, L., Van Weeren, L., Postma, M., Wiese, K. E., Mastop, M., Aumonier, S., Gotthard, G., Royant, A., Hink, M. A., & Gadella, T. W. (2016). MScarlet: A bright monomeric red fluorescent protein for cellular imaging. *Nature Methods*, *14*(1), 53–56.

Blanco-Rivero, A., Shutova, T., Román, M. J., Villarejo, A., & Martinez, F. (2012). Phosphorylation controls the localization and activation of the lumenal carbonic anhydrase in Chlamydomonas reinhardtii. *PLOS ONE*, *7*(11).

Boynton, J. E., Gillham, N. W., Harris, E. H., Hosler, J. P., Johnson, A. M., Jones, A. R., Randolph-Anderson, B. L., Robertson, D., Klein, T. M., Shark, K. B., & Sanford, J. C. (1988). Chloroplast transformation in Chlamydomonas with high velocity microprojectiles. *Science*, *240*(4858), 1534–1538.

Brazelton, B., Silflow, C., & Lefebvre, P. (2018). A BAC library for Chlamydomonas reinhardtii. www.chlamycollection.org.

Breker, M., Lieberman, K., & Cross, F. R. (2018). Comprehensive discovery of cell-cycle-essential pathways in chlamydomonas reinhardtii. *Plant Cell*, *30*(6), 1178–1198.

Brueggeman, A. J., Gangadharaiah, D. S., Cserhati, M. F., Casero, D., Weeks, D. P., & Ladunga, I. (2012). Activation of the Carbon Concentrating Mechanism by CO_2 Deprivation Coincides with Massive Transcriptional Restructuring in *Chlamydomonas reinhardtii*. *The Plant Cell*, *24*(5), 1860–1875.

Brumos, J., Zhao, C., Gong, Y., Soriano, D., Patel, A. P., Perez-Amador, M. A., Stepanova, A. N., & Alonso, J. M. (2020). An Improved Recombineering Toolset for Plants. *The Plant Cell*, *32*(1), 100–122.

Bruns, C. M., & Karplus, A. P. (1995). Refined Crystal Structure of Spinach Ferredoxin Reductase at 1.7 Å Resolution: Oxidized, Reduced and 2'-Phospho-5'-AMP Bound States. *Journal of Molecular Biology*, *247*(1), 125–145.

Buchanan, B. B. (2016). The Path to Thioredoxin and Redox Regulation in Chloroplasts. *Annual Review of Plant Biology*, *67*(1), 1–24.

Buchert, F., Mosebach, L., Gäbelein, P., & Hippler, M. (2020). PGR5 is required for efficient Q cycle in the cytochrome b6f complex during cyclic electron flow. *Biochemical Journal*, *477*(9), 1631–1650.

Buchert, F., Scholz, M., & Hippler, M. (2022). Electron transfer via cytochrome b6f complex displays sensitivity to antimycin A upon STT7 kinase activation. *Biochemical Journal*, *479*(1), 111–127.

Burlacot, A., Dao, O., Auroy, P., Cuiné, S., Li-Beisson, Y., & Peltier, G. (2022). Alternative photosynthesis pathways drive the algal CO2-concentrating mechanism. *Nature*, *605*(7909), 366–371.

Cameron, D., Osborne, C., Horton, P., & Sinclair, M. (2015). A sustainable model for intensive agriculture. *Grantham Centre briefing note*, (December), 4.

Cariti, F., Chazaux, M., Lefebvre-Legendre, L., Longoni, P., Ghysels, B., Johnson, X., & Goldschmidt-Clermont, M. (2020). Regulation of light harvesting in chlamydomonas reinhardtii two protein phosphatases are involved in state transitions. *Plant Physiology*, *183*(4), 1749–1764.

Carrillo, N., & Vallejos, R. H. (1983). The light-dependent modulation of photosynthetic electron transport. *Trends in Biochemical Sciences*, *8*(2), 52–56.

Caspari, O. D. (2020). Introduction of a leaky stop codon as molecular tool in Chlamydomonas reinhardtii. *PLOS ONE*, *15*(8), e0237405.

Cerutti, H., Johnson, A. M., Gillham, N. W., & Boynton, J. E. (1997). Epigenetic silencing of a foreign gene in nuclear transformants of Chlamydomonas. *The Plant Cell*, *9*(6), 925–945.

Chapman, S. P., Paget, C. M., Johnson, G. N., & Schwartz, J. M. (2015). Flux balance analysis reveals acetate metabolism modulates cyclic electron flow and alternative glycolytic pathways in Chlamydomonas reinhardtii. *Frontiers in Plant Science*, *6*(6).

Charoonnart, P., Worakajit, N., Zedler, J. A., Meetam, M., Robinson, C., & Saksmerprome, V. (2019). Generation of microalga Chlamydomonas reinhardtii expressing shrimp antiviral dsRNA without supplementation of antibiotics. *Scientific Reports*, *9*(1), 1–8.

Chaux, F., Burlacot, A., Mekhalfi, M., Auroy, P., Blangy, S., Richaud, P., & Peltier, G. (2017). Flavodiiron Proteins Promote Fast and Transient O_2 Photoreduction in *Chlamydomonas*. *Plant Physiology*, *174*(3), 1825–1836.

Chaux, F., Peltier, G., & Johnson, X. (2015). A security network in PSI photoprotection: regulation of photosynthetic control, NPQ and O2 photoreduction by cyclic electron flow. *Frontiers in Plant Science*, *6*(10), 875.

Clark, R. D., Hawkesford, M. J., Coughlan, S. J., Bennett, J., & Hind, G. (1984). Association of ferredoxin-NADP+ oxidoreductase with the chloroplast cytochrome b-f complex. *FEBS Letters*, *174*(1), 137–142.

Clarke, L. A., Rebelo, C. S., Gonçalves, J., Boavida, M. G., & Jordan, P. (2001). PCR amplification introduces errors into mononucleotide and dinucleotide repeat sequences. *Journal of Clinical Pathology - Molecular Pathology*, *54*(5), 351–353.

Cleland, R. E., & Bendall, D. S. (1992). Photosystem I cyclic electron transport: Measurement of ferredoxin-plastoquinone reductase activity. *Photosynthesis Research*, *34*(3), 409–418.

Court, D. L., Sawitzke, J. A., & Thomason, L. C. (2002). Genetic engineering using homologous recombination. *Annual Review of Genetics*, *36*(1), 361–388.

Cox, R. P. (1979). Chloroplast cytochrome b-563, hydrophobic environment and lack of direct reaction with ferredoxin. *The Biochemical journal*, *184*(1), 39–44.

Crofts, A. R., Meinhardt, S. W., Jones, K. R., & Snozzi, M. (1983). The role of the quinone pool in the cyclic electron-transfer chain of Rhodopseudomonas sphaeroides A modified Q-cycle mechanism. *BBA - Bioenergetics*, *723*(2), 202–218.

Cross, F. R. (2016). Tying down loose ends in the Chlamydomonas genome: Functional significance of abundant upstream open reading frames. *G3: Genes, Genomes, Genetics*, *6*(2), 435–446.

Crozet, P., Navarro, F. J., Willmund, F., Mehrshahi, P., Bakowski, K., Lauersen, K. J., Pérez-Pérez, M. E., Auroy, P., Gorchs Rovira, A., Sauret-Gueto, S., Niemeyer, J., Spaniol, B., Theis, J., Trösch, R., Westrich, L. D., Vavitsas, K., Baier, T., Hübner, W., De Carpentier, F., … Lemaire, S. D. (2018). Birth of a Photosynthetic Chassis: A MoClo Toolkit Enabling Synthetic Biology in the Microalga Chlamydomonas reinhardtii. *ACS Synthetic Biology*, *7*(9), 2074–2086.

DalCorso, G., Pesaresi, P., Masiero, S., Aseeva, E., Schünemann, D., Finazzi, G., Joliot, P., Barbato, R., & Leister, D. (2008). A Complex Containing PGRL1 and PGR5 Is Involved in the Switch between Linear and Cyclic Electron Flow in Arabidopsis. *Cell*, *132*(2), 273–285.

Dang, K. V., Plet, J., Tolleter, D., Jokel, M., Cuiné, S., Carrier, P., Auroy, P., Richaud, P., Johnson, X., Alric, J., Allahverdiyeva, Y., & Peltier, G. (2014). Combined increases in mitochondrial cooperation and oxygen photoreduction compensate for deficiency in cyclic electron flow in Chlamydomonas reinhardtii. *Plant Cell*, *26*(7), 3036–3050.

Dauvillée, D., Stampacchia, O., Girard-Bascou, J., & Rochaix, J. D. (2003). Tab2 is a novel conserved RNA binding protein required for translation of the chloroplast psaB mRNA. *EMBO Journal*, *22*(23), 6378–6388.

Delosme, R., Olive, J., & Wollman, F. A. (1996). Changes in light energy distribution upon state transitions: An in vivo photoacoustic study of the wild type and photosynthesis mutants from Chlamydomonas reinhardtii. *Biochimica et Biophysica Acta - Bioenergetics*, *1273*(2), 150–158.

Depège, N., Bellafiore, S., & Rochaix, J. D. (2003). Role of chloroplast protein kinase Stt7 in LHCII phosphorylation and state transition in Chlamydomonas. *Science*, *299*(5612), 1572–1575.

Desplats, C., Mus, F., Cuiné, S., Billon, E., Cournac, L., & Peltier, G. (2009). Characterization of Nda2, a plastoquinone-reducing type II NAD(P)H dehydrogenase in chlamydomonas chloroplasts. *The Journal of biological chemistry*, *284*(7), 4148–4157.

Douglas, S. E. (1998). Plastid evolution: Origins, diversity, trends. *Current Opinion in Genetics and Development*, *8*(6), 655–661.

Duke, S. O. (2015). Perspectives on transgenic, herbicide-resistant crops in the United States almost 20 years after introduction. *Pest Management Science*, *71*(5), 652–657.

Dumas, L., Zito, F., Blangy, S., Auroy, P., Johnson, X., Peltier, G., & Alric, J. (2017). A stromal region of cytochrome b6f subunit IV is involved in the activation of the Stt7 kinase in Chlamydomonas. *Proceedings of the National Academy of Sciences of the United States of America*, *114*(45), 12063–12068.

Dyo, Y. M., & Purton, S. (2018). The algal chloroplast as a synthetic biology platform for production of therapeutic proteins. *Microbiology*, *164*(2), 113–121.

Emrich-Mills, T. Z., Yates, G., Barrett, J., Girr, P., Grouneva, I., Lau, C. S., Walker, C. E., Kwok, T. K., Davey, J. W., Johnson, M. P., & Mackinder, L. C. (2021). A recombineering pipeline to clone large and complex genes in Chlamydomonas. *Plant Cell*, *33*(4), 1161–1181.

Engel, B. D., Schaffer, M., Cuellar, L. K., Villa, E., Plitzko, J. M., & Baumeister, W. (2015). Native architecture of the chlamydomonas chloroplast revealed by in situ cryo-electron tomography. *eLife*, *2015*(4).

Farah, J., Rappaport, F., Choquet, Y., Joliot, P., & Rochaix, J. D. (1995). Isolation of a psaF-deficient mutant of Chlamydomonas reinhardtii: efficient interaction of plastocyanin with the photosystem I reaction center is mediated by the PsaF subunit. *The EMBO journal*, *14*(20), 4976–4984.

Ferenczi, A., Chew, Y. P., Kroll, E., von Koppenfels, C., Hudson, A., & Molnar, A. (2021). Mechanistic and genetic basis of single-strand templated repair at Cas12a-induced DNA breaks in Chlamydomonas reinhardtii. *Nature Communications*, *12*(1), 1–12.

Finazzi, G., Furia, A., Barbagallo, R. P., & Forti, G. (1999). State transitions, cyclic and linear electron transport and photophosphorylation in Chlamydomonas reinhardtii. *Biochimica et Biophysica Acta - Bioenergetics*, *1413*(3), 117–129.

Fire, A., Xu, S., Montgomery, M. K., Kostas, S. A., Driver, S. E., & Mello, C. C. (1998). Potent and specific genetic interference by double-stranded RNA in caenorhabditis elegans. *Nature*, *391*(6669), 806–811.

Flannery, S. E., Hepworth, C., Wood, W. H., Pastorelli, F., Hunter, C. N., Dickman, M. J., Jackson, P. J., & Johnson, M. P. (2021). Developmental acclimation of the thylakoid proteome to light intensity in Arabidopsis. *Plant Journal*, *105*(1), 223–244.

Foyer, C. H., Neukermans, J., Queval, G., Noctor, G., & Harbinson, J. (2012). Photosynthetic control of electron transport and the regulation of gene expression. *Journal of Experimental Botany*, *63*(4), 1637–1661.

Fulgosi, H., & Vojta, L. (2020). Tweaking Photosynthesis: FNR-TROL Interaction as Potential Target for Crop Fortification. *Frontiers in Plant Science*, *11*, 318.

Gaigbe-Togbe, V., Bassarsky, L., Gu, D., Spoorenberg, T., & Zeifman, L. (2022). *World Population Prospects* (tech. rep.). United Nations. New York.

Gallaher, S. D., Fitz-Gibbon, S. T., Glaesener, A. G., Pellegrini, M., & Merchant, S. S. (2015). Chlamydomonas Genome Resource for Laboratory Strains Reveals a Mosaic of Sequence Variation, Identifies True Strain Histories, and Enables Strain-Specific Studies. *The Plant Cell*, *27*(9), 2335–2352.

Gallaher, S. D., Fitz-Gibbon, S. T., Strenkert, D., Purvine, S. O., Pellegrini, M., & Merchant, S. S. (2018). High-throughput sequencing of the chloroplast and mitochondrion of *Chlamydomonas reinhardtii* to generate improved *de novo* assemblies, analyze expression patterns and transcript speciation, and evaluate diversity among laboratory strains and. *The Plant Journal*, *93*(3), 545–565.

Ghribi, M., Nouemssi, S. B., Meddeb-Mouelhi, F., & Desgagné-Penix, I. (2020). Genome editing by CRISPR-Cas: A game change in the genetic manipulation of chlamydomonas.

Gibson, D. G., Young, L., Chuang, R. Y., Venter, J. C., Hutchison, C. A., & Smith, H. O. (2009). Enzymatic assembly of DNA molecules up to several hundred kilobases. *Nature Methods*, *6*(5), 343–345.

Goedhart, J., Von Stetten, D., Noirclerc-Savoye, M., Lelimousin, M., Joosen, L., Hink, M. A., Van Weeren, L., Gadella, T. W., & Royant, A. (2012). Structure-guided evolution of cyan fluorescent proteins towards a quantum yield of 93%. *Nature Communications*, *3*(1), 1–9.

Goldschmidt-Clermont, M. (1991). Transgenic expression of aminoglycoside adenine transferase in the chloroplast: A selectable marker for site-directed transformation of chlamydomonas. *Nucleic Acids Research*, *19*(15), 4083–4089.

Goss, T., & Hanke, G. (2014). The End of the Line: Can Ferredoxin and Ferredoxin NADP(H) Oxidoreductase Determine the Fate of Photosynthetic Electrons? *Current Protein & Peptide Science*, *15*(4), 385–393.

Guiles, R. D., Yachandra, V. K., McDermott, A. E., Cole, J. L., Dexheimer, S. L., Britt, R. D., Sauer, K., & Klein, M. P. (1990). The S0 state of photosystem II induced by hydroxylamine: differences between the structure of the manganese complex in the S0 and S1 states determined by X-ray absorption spectroscopy. *Biochemistry*, *29*(2), 486–496.

Guo, J., Wei, X., Li, M., Pan, X., Chang, W., & Liu, Z. (2013). Structure of the catalytic domain of a state transition kinase homolog from Micromonas algae. *Protein and Cell*, *4*(8), 607–619.

Hahn, A., Vonck, J., Mills, D. J., Meier, T., & Kühlbrandt, W. (2018). Structure, mechanism, and regulation of the chloroplast ATP synthase. *Science*, *360*(6389).

Hald, S., Nandha, B., Gallois, P., & Johnson, G. N. (2008). Feedback regulation of photosynthetic electron transport by NADP(H) redox poise. *Biochimica et Biophysica Acta - Bioenergetics*, *1777*(5), 433–440.

Hamel, P., Olive, J., Pierre, Y., Wollman, F. A., & De Vitry, C. (2000). A new subunit of cytochrome b6f complex undergoes reversible phosphorylation upon state transition. *Journal of Biological Chemistry*, *275*(22), 17072–17079.

Harris, E. H. (2009). *The Chlamydomonas Sourcebook* (E. H. Harris, D. B. Stern, & G. Witman, Eds.; 2nd ed). Elsevier/Academic Press.

Heber, U. (2002). Irrungen, wirrungen? the mehler reaction in relation to cyclic electron transport in c3 plants. *Photosynthesis Research*, *73*(1), 223–231.

Hepworth, C., Wood, W. H., Emrich-Mills, T. Z., Proctor, M. S., Casson, S., & Johnson, M. P. (2021). Dynamic thylakoid stacking and state transitions work synergistically to avoid acceptor-side limitation of photosystem I. *Nature Plants 2021 7:1*, *7*(1), 87–98.

Hertle, A. P., Blunder, T., Wunder, T., Pesaresi, P., Pribil, M., Armbruster, U., & Leister, D. (2013). PGRL1 Is the Elusive Ferredoxin-Plastoquinone Reductase in Photosynthetic Cyclic Electron Flow. *Molecular Cell*, *49*(3), 511–523.

Hippler, M., Drepper, F., Farah, J., & Rochaix, J. D. (1997). Fast electron transfer from cytochrome c6 and plastocyanin to photosystem I of Chlamydomonas reinhardtii requires PsaF. *Biochemistry*, *36*(21), 6343–6349.

Hirano, S., Nishimasu, H., Ishitani, R., & Nureki, O. (2016). Structural Basis for the Altered PAM Specificities of Engineered CRISPR-Cas9. *Molecular Cell*, *61*(6), 886–894.

Ho, T. T. H., Schwier, C., Elman, T., Fleuter, V., Zinzius, K., Scholz, M., Yacoby, I., Buchert, F., & Hippler, M. (2022). Photosystem I light-harvesting proteins regulate photosynthetic electron transfer and hydrogen production. *Plant Physiology*, *189*(1), 329–343.

Huang, W., Hu, H., & Zhang, S.-B. (2015). Photorespiration plays an important role in the regulation of photosynthetic electron flow under fluctuating light in tobacco plants grown under full sunlight. *Frontiers in Plant Science*, *6*(8), 621.

Itakura, A. K., Chan, K. X., Atkinson, N., Pallesen, L., Wang, L., Reeves, G., Patena, W., Caspari, O., Roth, R., Goodenough, U., McCormick, A. J., Griffiths, H., & Jonikas, M. C. (2019). A Rubisco-binding protein is required for normal pyrenoid number and starch sheath morphology in Chlamydomonas reinhardtii. *Proceedings of the National Academy of Sciences of the United States of America*, *116*(37), 18445–18454.

Iwai, M., Takizawa, K., Tokutsu, R., Okamuro, A., Takahashi, Y., & Minagawa, J. (2010). Isolation of the elusive supercomplex that drives cyclic electron flow in photosynthesis. *Nature*, *464*(7292), 1210–1213.

Jackson, H. O., Taunt, H. N., Mordaka, P. M., Kumari, S., Smith, A. G., & Purton, S. (2022). CpPosNeg: A positive-negative selection strategy allowing multiple cycles of marker-free engineering of the Chlamydomonas plastome. *Biotechnology Journal*, 2200088.

Jackson, H. O., Taunt, H. N., Mordaka, P. M., Smith, A. G., & Purton, S. (2021). The Algal Chloroplast as a Testbed for Synthetic Biology Designs Aimed at Radically Rewiring Plant Metabolism. *Frontiers in Plant Science*, *12*.

Jaeger, D., Baier, T., & Lauersen, K. J. (2019). Intronserter, an advanced online tool for design of intron containing transgenes. *Algal Research*, *42*, 101588.

Jamroskovic, J., Obi, I., Movahedi, A., Chand, K., Chorell, E., & Sabouri, N. (2019). Identification of putative G-quadruplex DNA structures in S. pombe genome by quantitative PCR stop assay. *DNA Repair*, *82*, 102678.

Jiang, W., Brueggeman, A. J., Horken, K. M., Plucinak, T. M., & Weeks, D. P. (2014). Successful transient expression of Cas9 and single guide RNA genes in Chlamydomonas reinhardtii. *Eukaryotic Cell*, *13*(11), 1465–1469.

Jinek, M., Chylinski, K., Fonfara, I., Hauer, M., Doudna, J. A., & Charpentier, E. (2012). A programmable dual-RNA-guided DNA endonuclease in adaptive bacterial immunity. *Science*, *337*(6096), 816–821.

Johnson, M. P. (2016). Photosynthesis. *Essays in Biochemistry*, *60*(3), 255.

Joliot, P., Béal, D., & Joliot, A. (2004). Cyclic electron flow under saturating excitation of dark-adapted Arabidopsis leaves. *Biochimica et Biophysica Acta - Bioenergetics*, *1656*(2-3), 166–176.

Joliot, P., & Johnson, G. N. (2011). Regulation of cyclic and linear electron flow in higher Plants. *Proceedings of the National Academy of Sciences of the United States of America*, *108*(32), 13317–13322.

Joliot, P., & Joliot, A. (2002). Cyclic electron transfer in plant leaf. *Proceedings of the National Academy of Sciences of the United States of America*, *99*(15), 10209–10214.

Joliot, P., Lavergne, J., & Béal, D. (1992). Plastoquinone compartmentation in chloroplasts. I. Evidence for domains with different rates of photo-reduction. *Biochimica et Biophysica Acta - Protein Structure and Molecular*, *1101*(1), 1–12.

Joliot, P., Sellés, J., Wollman, F.-A., & Verméglio, A. (2022). High efficient cyclic electron flow and functional supercomplexes in Chlamydomonas cells. *Biochimica et Biophysica Acta - Bioenergetics*, 148909.

Jones, G. A., & Bradshaw, D. S. (2019). Resonance Energy Transfer: From Fundamental Theory to Recent Applications. *Frontiers in Physics*, *7*.

Kan, K.-S., & Thornber, J. P. (1976). The Light-harvesting Chlorophyll a/b-Protein Complex of Chlamydomonas reinhardii. *Plant Physiology*, *57*(1), 47.

Kanygin, A., Milrad, Y., Thummala, C., Reifschneider, K., Baker, P., Marco, P., Yacoby, I., & Redding, K. E. (2020). Rewiring photosynthesis: A photosystem

I-hydrogenase chimera that makes H2: In vivo. *Energy and Environmental Science*, *13*(9), 2903–2914.

Kimata, Y., & Hase, T. (1989). Localization of ferredoxin isoproteins in mesophyll and bundle sheath cells in maize leaf. *Plant physiology*, *89*(4), 1193–1197.

Kimple, M. E., Brill, A. L., & Pasker, R. L. (2013). Overview of affinity tags for protein purification. *Current Protocols in Protein Science*, *73*(73).

Klughammer, C., & Schreiber, U. (1994). An improved method, using saturating light pulses, for the determination of photosystem I quantum yield via P700 absorbance changes at 830 nm. *Planta 1994 192:2*, *192*(2), 261–268.

Kobayashi, Y., Takusagawa, M., Harada, N., Fukao, Y., Yamaoka, S., Kohchi, T., Hori, K., Ohta, H., Shikanai, T., & Nishimura, Y. (2016). Eukaryotic Components Remodeled Chloroplast Nucleoid Organization during the Green Plant Evolution. *Genome Biology and Evolution*, *8*(1), 1–16.

Kramer, D. M., Avenson, T. J., & Edwards, G. E. (2004). Dynamic flexibility in the light reactions of photosynthesis governed by both electron and proton transfer reactions. *Trends in Plant Science*, *9*(7), 349–357.

Kramer, M., Rodriguez-Heredia, M., Saccon, F., Mosebach, L., Twachtmann, M., Krieger-Liszkay, A., Duffy, C., Knell, R. J., Finazzi, G., & Hanke, G. T. (2021). Regulation of photosynthetic electron flow on dark to light transition by ferredoxin:NADP(H) oxidoreductase interactions. *eLife*, *10*.

Kropat, J., Hong-Hermesdorf, A., Casero, D., Ent, P., Castruita, M., Pellegrini, M., Merchant, S. S., & Malasarn, D. (2011). A revised mineral nutrient supplement increases biomass and growth rate in Chlamydomonas reinhardtii. *The Plant Journal*, *66*(5), 770–780.

Kumar, S., Stecher, G., Suleski, M., & Blair Hedges, S. (2017). TimeTree: A Resource for Timelines, Timetrees, and Divergence Times. *Molecular Biology and Evolution*, *34*(7), 1812–1819.

Larrea-Alvarez, M., Young, R., & Purton, S. (2021). A simple technology for generating marker-free chloroplast transformants of the green alga Chlamydomonas reinhardtii. In *Methods in molecular biology* (pp. 293–304). Humana Press Inc.

Leliaert, F., Smith, D. R., Moreau, H., Herron, M. D., Verbruggen, H., Delwiche, C. F., & De Clerck, O. (2012). Phylogeny and Molecular Evolution of the Green Algae. *Critical Reviews in Plant Sciences*, *31*(1), 1–46.

Lempiäinen, T., Rintamäki, E., Aro, E.-M., & Tikkanen, M. (2022). Plants acclimate to Photosystem I photoinhibition by readjusting the photosynthetic machinery. *Plant, Cell & Environment*, *45*(10), 2954–2971.

Li, X., Patena, W., Fauser, F., Jinkerson, R. E., Saroussi, S., Meyer, M. T., Ivanova, N., Robertson, J. M., Yue, R., Zhang, R., Vilarrasa-Blasi, J., Wittkopp, T. M., Ramundo, S., Blum, S. R., Goh, A., Laudon, M., Srikumar, T., Lefebvre, P. A.,

Grossman, A. R., & Jonikas, M. C. (2019). A genome-wide algal mutant library and functional screen identifies genes required for eukaryotic photosynthesis.

Lin, Y., Huo, L., Liu, Z., Li, J., Liu, Y., He, Q., Wang, X., & Liang, S. (2013). Sodium laurate, a novel protease- and mass spectrometry-compatible detergent for mass spectrometry-based membrane proteomics. *PLOS ONE*, *8*(3).

Lintala, M., Lehtimäki, N., Benz, J. P., Jungfer, A., Soll, J., Aro, E. M., Bölter, B., & Mulo, P. (2012). Depletion of leaf-type ferredoxin-NADP + oxidoreductase results in the permanent induction of photoprotective mechanisms in Arabidopsis chloroplasts. *Plant Journal*, *70*(5), 809–817.

Lintala, M., Schuck, N., Thormählen, I., Jungfer, A., Weber, K. L., Weber, A. P., Geigenberger, P., Soll, J., Bölter, B., & Mulo, P. (2014). Arabidopsis tic62 trol mutant lacking thylakoid-bound ferredoxin-NADP$^+$ oxidoreductase shows distinct metabolic phenotype. *Molecular Plant*, *7*(1), 45–57.

Long, S. P., Marshall-Colon, A., & Zhu, X. G. (2015). Meeting the global food demand of the future by engineering crop photosynthesis and yield potential.

Lorenz, R., Bernhart, S. H., Höner zu Siederdissen, C., Tafer, H., Flamm, C., Stadler, P. F., & Hofacker, I. L. (2011). ViennaRNA Package 2.0. *Algorithms for Molecular Biology*, *6*(1), 1–14.

Lucker, B., & Kramer, D. M. (2013). Regulation of cyclic electron flow in Chlamydomonas reinhardtii under fluctuating carbon availability. *Photosynthesis Research*, *117*(1-3), 449–459.

Mackinder, L. C. M. (2018). The *Chlamydomonas* CO_2-concentrating mechanism and its potential for engineering photosynthesis in plants. *New Phytologist*, *217*(1), 54–61.

Mackinder, L. C., Chen, C., Leib, R. D., Patena, W., Blum, S. R., Rodman, M., Ramundo, S., Adams, C. M., & Jonikas, M. C. (2017). A Spatial Interactome Reveals the Protein Organization of the Algal CO2-Concentrating Mechanism. *Cell*, *171*(1), 133–147.e14.

Mackinder, L. C., Meyer, M. T., Mettler-Altmann, T., Chen, V. K., Mitchell, M. C., Caspari, O., Rosenzweig, E. S., Pallesen, L., Reeves, G., Itakura, A., Roth, R., Sommer, F., Geimer, S., Mühlhaus, T., Schroda, M., Goodenough, U., Stitt, M., Griffiths, H., & Jonikas, M. C. (2016). A repeat protein links Rubisco to form the eukaryotic carbon-concentrating organelle. *Proceedings of the National Academy of Sciences of the United States of America*, *113*(21), 5958–5963.

Maillet, C., Gupta, R., Schell, M., Brewton, R., Murphy, C., Wall, J., & Mullin, B. (2001). Enhanced capture of small histidine-containing polypeptides on membranes in the presence of $zncl_2$. *BioTechniques*, *30*(6), 1224–1230.

Malone, L. A., Proctor, M. S., Hitchcock, A., Hunter, C. N., & Johnson, M. P. (2021). Cytochrome b6f – orchestrator of photosynthetic electron transfer. *Biochimica et Biophysica Acta - Bioenergetics*, *1862*(5), 148380.

Malone, L. A., Qian, P., Mayneord, G. E., Hitchcock, A., Farmer, D. A., Thompson, R. F., Swainsbury, D. J., Ranson, N. A., Hunter, C. N., & Johnson, M. P. (2019). Cryo-EM structure of the spinach cytochrome b6f complex at 3.6 Å resolution. *Nature*, *575*(7783), 535–539.

Mapstone, L. J., Leite, M. N., Purton, S., Crawford, I. A., & Dartnell, L. (2022). Cyanobacteria and microalgae in supporting human habitation on Mars.

Maresca, M., Erler, A., Fu, J., Friedrich, A., Zhang, Y., & Stewart, A. F. (2010). Single-stranded heteroduplex intermediates in λ Red homologous recombination. *BMC Molecular Biology*, *11*(1), 54.

Maul, J. E., Lilly, J. W., Cui, L., DePamphilis, C. W., Miller, W., Harris, E. H., & Stern, D. B. (2002). The *Chlamydomonas reinhardtii* Plastid Chromosome. *The Plant Cell*, *14*(11), 2659–2679.

Maxwell, K., & Johnson, G. (2000). Chlorophyll fluorescence—a practical guide. *Journal of Experimental Botany*, *51*(345), 659–668.

Melkozernov, A. N., Kargul, J., Lin, S., Barber, J., & Blankenship, R. E. (2004). Energy coupling in the PSI-LHCI supercomplex from the green alga chlamydomonas reinhardtii. *Journal of Physical Chemistry B*, *108*(29), 10547–10555.

Merchant, S., & Bogorad, L. (1987). Metal ion regulated gene expression: use of a plastocyanin-less mutant of Chlamydomonas reinhardtii to study the Cu(II)-dependent expression of cytochrome c-552. *The EMBO Journal*, *6*(9), 2531–2535.

Merchant, S. S., Prochnik, S. E., Vallon, O., Harris, E. H., Karpowicz, S. J., Witman, G. B., Terry, A., Salamov, A., Fritz-Laylin, L. K., Maréchal-Drouard, L., Marshall, W. F., Qu, L. H., Nelson, D. R., Sanderfoot, A. A., Spalding, M. H., Kapitonov, V. V., Ren, Q., Ferris, P., Lindquist, E., … Zhou, K. (2007). The Chlamydomonas genome reveals the evolution of key animal and plant functions. *Science*, *318*(5848), 245–251.

Metz, J. G., Pakrasi, H. B., Seibert, M., & Arntzer, C. J. (1986). Evidence for a dual function of the herbicide-binding D1 protein in photosystem II. *FEBS Letters*, *205*(2), 269–274.

Meyer, M. T., Itakura, A. K., Patena, W., Wang, L., He, S., Emrich-Mills, T., Lau, C. S., Yates, G., MacKinder, L. C., & Jonikas, M. C. (2020). Assembly of the algal CO2-fixing organelle, the pyrenoid, is guided by a Rubisco-binding motif. *Science Advances*, *6*(46).

Meyer Zu Tittingdorf, J. M., Rexroth, S., Schäfer, E., Schlichting, R., Giersch, C., Dencher, N. A., & Seelert, H. (2004). The stoichiometry of the chloroplast ATP synthase oligomer III in Chlamydomonas reinhardtii is not affected by the metabolic state. *Biochimica et Biophysica Acta - Bioenergetics*, *1659*(1), 92–99.

Mitchell, P. (1975). The protonmotive q cycle: A general formulation. *FEBS Letters*, *59*(2), 137–139.

Miyake, C. (2010). Alternative electron flows (water-water cycle and cyclic electron flow around PSI) in photosynthesis: Molecular mechanisms and physiological functions. *Plant and Cell Physiology*, *51*(12), 1951–1963.

Morgulis, A., Gertz, E. M., Schaffer, A. A., & Agarwala, R. (2006). WindowMasker: window-based masker for sequenced genomes. *Bioinformatics*, *22*(2), 134–141.

Mosebach, L., Heilmann, C., Mutoh, R., Gäbelein, P., Steinbeck, J., Happe, T., Ikegami, T., Hanke, G., Kurisu, G., & Hippler, M. (2017). Association of Ferredoxin:NADP+ oxidoreductase with the photosynthetic apparatus modulates electron transfer in Chlamydomonas reinhardtii. *Photosynthesis Research*, *134*(3), 291–306.

Moss, D. A., & Bendall, D. S. (1984). Cyclic electron transport in chloroplasts. The Q-cycle and the site of action of antimycin. *BBA - Bioenergetics*, *767*(3), 389–395.

Mouille, G., Maddelein, M. L., Libessart, N., Talaga, P., Decq, A., Delrue, B., & Ball, S. (1996). Preamylopectin Processing: A Mandatory Step for Starch Biosynthesis in Plants. *The Plant Cell*, *8*(8), 1353–1366.

Mukherjee, A., Lau, C. S., Walker, C. E., Rai, A. K., Prejean, C. I., Yates, G., Emrich-Mills, T., Lemoine, S. G., Vinyard, D. J., Mackinder, L. C., & Moroney, J. V. (2019). Thylakoid localized bestrophin-like proteins are essential for the CO2 concentrating mechanism of Chlamydomonas reinhardtii. *Proceedings of the National Academy of Sciences of the United States of America*, *116*(34), 16915–16920.

Munekage, Y., Hashimoto, M., Miyake, C., Tomizawa, K. I., Endo, T., Tasaka, M., & Shikanai, T. (2004). Cyclic electron flow around photosystem I is essential for photosynthesis. *Nature*, *429*(6991), 579–582.

Munekage, Y., Hojo, M., Meurer, J., Endo, T., Tasaka, M., & Shikanai, T. (2002). PGR5 is involved in cyclic electron flow around photosystem I and is essential for photoprotection in Arabidopsis. *Cell*, *110*(3), 361–371.

Murphy, K. C. (1998). Use of bacteriophage λ recombination functions to promote gene replacement in Escherichia coli. *Journal of Bacteriology*, *180*(8), 2063–2071.

Musiani, F., Dikiy, A., Semenov, A. Y., & Ciurli, S. (2005). Structure of the Intermolecular Complex between Plastocyanin and Cytochrome *f* from Spinach. *Journal of Biological Chemistry*, *280*(19), 18833–18841.

Nandha, B., Finazzi, G., Joliot, P., Hald, S., & Johnson, G. N. (2007). The role of PGR5 in the redox poising of photosynthetic electron transport. *Biochimica et Biophysica Acta - Bioenergetics*, *1767*(10), 1252–1259.

Nawrocki, W. J., Bailleul, B., Cardol, P., Rappaport, F., Wollman, F. A., & Joliot, P. (2019a). Maximal cyclic electron flow rate is independent of PGRL1 in Chlamydomonas. *Biochimica et Biophysica Acta - Bioenergetics*, *1860*(5), 425–432.

Nawrocki, W., Bailleul, B., Picot, D., Cardol, P., Rappaport, F., Wollman, F.-A., & Joliot, P. (2019b). The mechanism of cyclic electron flow. *Biochimica et Biophysica Acta - Bioenergetics*, *1860*(5), 433–438.

Nawrocki, W., Buchert, F., Joliot, P., Rappaport, F., Bailleul, B., & Wollman, F. A. (2019c). Chlororespiration controls growth under intermittent light. *Plant Physiology*, *179*(2), 630–639.

Nelms, B. L., & Labosky, P. A. (2011). A predicted hairpin cluster correlates with barriers to PCR, sequencing and possibly BAC recombineering. *Scientific Reports*, *1*(1), 1–7.

Nelson, N., & Ben-Shem, A. (2004). The complex architecture of oxygenic photosynthesis. *Nature Reviews Molecular Cell Biology 2004 5:12*, *5*(12), 971–982.

Neupert, J., Gallaher, S. D., Lu, Y., Strenkert, D., Segal, N., Barahimipour, R., Fitz-Gibbon, S. T., Schroda, M., Merchant, S. S., & Bock, R. (2020). An epigenetic gene silencing pathway selectively acting on transgenic DNA in the green alga Chlamydomonas. *Nature Communications*, *11*(1), 1–17.

Neupert, J., Karcher, D., & Bock, R. (2009). Generation of *Chlamydomonas* strains that efficiently express nuclear transgenes. *The Plant Journal*, *57*(6), 1140–1150.

Nikkanen, L., & Rintamäki, E. (2019). Chloroplast thioredoxin systems dynamically regulate photosynthesis in plants. *Biochem J*, *476*(7), 1159–1172.

Ohnishi, Y., Muraki, N., Kiyota, D., Okumura, H., Baba, S., Kawano, Y., Kumasaka, T., Tanaka, H., & Kurisu, G. (2020). X-ray dose-dependent structural changes of the [2Fe-2S] ferredoxin from *Chlamydomonas reinhardtii*. *The Journal of Biochemistry*, *167*(6), 549–555.

Orpana, A. K., Ho, T. H., & Stenman, J. (2012). Multiple heat pulses during PCR extension enabling amplification of GC-rich sequences and reducing amplification bias. *Analytical Chemistry*, *84*(4), 2081–2087.

Oxborough, K., & Baker, N. R. (1997). Resolving chlorophyll a fluorescence images of photosynthetic efficiency into photochemical and non-photochemical components – calculation of qP and Fv-/Fm-; without measuring Fo-; *Photosynthesis Research 1997 54:2*, *54*(2), 135–142.

Palatnik, J. F., Tognetti, V. B., Poli, H. O., Rodríguez, R. E., Blanco, N., Gattuso, M., Hajirezaei, M. R., Sonnewald, U., Valle, E. M., & Carrillo, N. (2003). Transgenic tobacco plants expressing antisense ferredoxin-NADP(H) reductase transcripts display increased susceptibility to photo-oxidative damage. *The Plant journal : for cell and molecular biology*, *35*(3), 332–341.

Peers, G., Truong, T. B., Ostendorf, E., Busch, A., Elrad, D., Grossman, A. R., Hippler, M., & Niyogi, K. K. (2009). An ancient light-harvesting protein is critical for the regulation of algal photosynthesis. *Nature*, *462*, 518–521.

Picariello, T., Hou, Y., Kubo, T., McNeill, N. A., Yanagisawa, H. A., Oda, T., & Witman, G. B. (2020). TIM, a targeted insertional mutagenesis method utilizing CRISPR/Cas9 in Chlamydomonas reinhardtii. *PLOS ONE*, *15*(5).

Porra, R. J., Thompson, W. A., & Kriedemann, P. E. (1989). Determination of accurate extinction coefficients and simultaneous equations for assaying chlorophylls a and b extracted with four different solvents: verification of the concentration of chlorophyll standards by atomic absorption spectroscopy. *BBA - Bioenergetics*, *975*(3), 384–394.

Poser, I., Sarov, M., Hutchins, J. R., Hériché, J. K., Toyoda, Y., Pozniakovsky, A., Weigl, D., Nitzsche, A., Hegemann, B., Bird, A. W., Pelletier, L., Kittler, R., Hua, S., Naumann, R., Augsburg, M., Sykora, M. M., Hofemeister, H., Zhang, Y., Nasmyth, K., … Hyman, A. A. (2008). BAC TransgeneOmics: A high-throughput method for exploration of protein function in mammals. *Nature Methods*, *5*(5), 409–415.

Proctor, M. S., Malone, L. A., Farmer, D. A., Swainsbury, D. J., Hawkings, F. R., Pastorelli, F., Emrich-Mills, T. Z., Siebert, C. A., Hunter, C. N., Johnson, M. P., & Hitchcock, A. (2022). Cryo-EM structures of the Synechocystis sp. PCC 6803 cytochrome b6f complex with and without the regulatory PetP subunit. *Biochemical Journal*, *479*(13).

Ramesh, V. M., & Webber, A. N. (2004). Rapid isolation and purification of photosystem I chlorophyll-binding protein from Chlamydomonas reinhardtii. *Methods in molecular biology (Clifton, N.J.)*, *274*, 19–28.

Ray, D. K., Mueller, N. D., West, P. C., & Foley, J. A. (2013). Yield Trends Are Insufficient to Double Global Crop Production by 2050. *PLOS ONE*, *8*(6).

Rozen, S., & Skaletsky, H. (2000). Primer3 on the WWW for general users and for biologist programmers. *Methods in molecular biology (Clifton, N.J.)*, *132*, 365–386.

Rühle, T., Dann, M., Reiter, B., Schünemann, D., Naranjo, B., Penzler, J. F., Kleine, T., & Leister, D. (2021). PGRL2 triggers degradation of PGR5 in the absence of PGRL1. *Nature Communications*, *12*(1), 1–14.

Sage, R. F., Sage, T. L., & Kocacinar, F. (2012). Photorespiration and the evolution of C4 photosynthesis. *Annu Rev Plant Biol*, *63*, 19–47.

Saif Hasan, S., & Cramer, W. A. (2012). On rate limitations of electron transfer in the photosynthetic cytochrome b6f complex. *Physical Chemistry Chemical Physics*, *14*(40), 13853–13860.

Sánchez-Baracaldo, P., & Cardona, T. (2020). On the origin of oxygenic photosynthesis and Cyanobacteria. *New Phytologist*, *225*(4), 1440–1446.

Sarov, M., Murray, J. I., Schanze, K., Pozniakovski, A., Niu, W., Angermann, K., Hasse, S., Rupprecht, M., Vinis, E., Tinney, M., Preston, E., Zinke, A., Enst, S., Teichgraber, T., Janette, J., Reis, K., Janosch, S., Schloissnig, S., Ejsmont,

R. K., ... Hyman, A. A. (2012). A genome-scale resource for in vivo tag-based protein function exploration in C. elegans. *Cell*, *150*(4), 855–866.

Sarov, M., Schneider, S., Pozniakovski, A., Roguev, A., Ernst, S., Zhang, Y., Hyman, A. A., & Stewart, A. F. (2006). A recombineering pipeline for functional genomics applied to Caenorhabditis elegans. *Nature Methods*, *3*(10), 839–844.

Sawyer, A., Esselborn, J., Winkler, M., & Happe, T. (2017). Chlamydomonas: Hydrogenase and hydrogen production. In M. Hippler (Ed.), *Chlamydomonas: Biotechnology and biomedicine* (pp. 21–44). Springer International Publishing.

Schreiber, U., & Klughammer, C. (2008). New accessory for the DUAL-PAM-100: The P515/535 module and examples of its application. *1*(2), 1–10.

Schreiber, U., & Klughammer, C. (2016). Analysis of Photosystem I Donor and Acceptor Sides with a New Type of Online-Deconvoluting Kinetic LED-Array Spectrophotometer. *Plant and Cell Physiology*, *57*(7), 1454–1467.

Schroda, M. (2006). Rna silencing in chlamydomonas: Mechanisms and tools. *Current Genetics*, *49*(2), 69–84.

Schroda, M. (2019). Good news for nuclear transgene expression in Chlamydomonas.

Schroda, M., Vallon, O., Wollman, F. A., & Beck, C. F. (1999). A chloroplast-targeted heat shock protein 70 (HSP70) contributes to the photoprotection and repair of photosystem II during and after photoinhibition. *Plant Cell*, *11*(6), 1165–1178.

Seelert, H., Poetsch, A., Dencher, N. A., Engel, A., Stahlberg, H., & Müller, D. J. (2000). Proton-powered turbine of a plant motor. *Nature*, *405*(6785), 418–419.

Senior, A. E., Nadanaciva, S., & Weber, J. (2002). The molecular mechanism of atp synthesis by f1f0-atp synthase. *Biochimica et Biophysica Acta - Bioenergetics*, *1553*(3), 188–211.

Sétif, P. (2015). Electron-transfer kinetics in cyanobacterial cells: Methyl viologen is a poor inhibitor of linear electron flow. *1847*(2), 212–222.

Sétif, P., Fischer, N., Lagoutte, B., Bottin, H., & Rochaix, J. D. (2002). The ferredoxin docking site of photosystem I. *Biochimica et Biophysica Acta - Bioenergetics*, *1555*(1-3), 204–209.

Shahak, Y., Crowther, D., & Hind, G. (1981). The involvement of ferredoxin-NADP+ reductase in cyclic electron transport in chloroplasts. *Biochimica et Biophysica Acta*, *636*(2), 234–243.

Shaner, N. C., Lambert, G. G., Chammas, A., Ni, Y., Cranfill, P. J., Baird, M. A., Sell, B. R., Allen, J. R., Day, R. N., Israelsson, M., Davidson, M. W., & Wang, J. (2013). A bright monomeric green fluorescent protein derived from Branchiostoma lanceolatum. *Nature Methods*, *10*(5), 407–409.

Sheng, X., Watanabe, A., Li, A., Kim, E., Song, C., Murata, K., Song, D., Minagawa, J., & Liu, Z. (2019). Structural insight into light harvesting for photosystem II in green algae. *Nature Plants 2019 5:12*, *5*(12), 1320–1330.

Shimogawara, K., Fujiwara, S., Grossman, A., & Usuda, H. (1998). High-Efficiency Transformation of Chlamydomonas reinhardtii by Electroporation. *Genetics*, *148*(4), 1821–1828.

Silflow, C. D., & Lefebvre, P. A. (2001). Assembly and Motility of Eukaryotic Cilia and Flagella. Lessons from *Chlamydomonas reinhardtii*. *Plant Physiology*, *127*(4), 1500–1507.

Smit, A., Hubley, R., & Green, P. (2015). RepeatMasker Open-4.0. http://www.repeatmasker.org.

Steinbeck, J., Nikolova, D., Weingarten, R., Johnson, X., Richaud, P., Peltier, G., Hermann, M., Magneschi, L., & Hippler, M. (2015). Deletion of Proton Gradient Regulation 5 (PGR5) and PGR5-Like 1 (PGRL1) proteins promote sustainable light-driven hydrogen production in Chlamydomonas reinhardtii due to increased PSII activity under sulfur deprivation. *Frontiers in Plant Science*, *6*(10).

Steinbeck, J., Ross, I. L., Rothnagel, R., Gäbelein, P., Schulze, S., Giles, N., Ali, R., Drysdale, R., Sierecki, E., Gambin, Y., Stahlberg, H., Takahashi, Y., Hippler, M., & Hankamer, B. (2018). Structure of a PSI-LHCI-cyt b6f supercomplex in Chlamydomonas reinhardtii promoting cyclic electron flow under anaerobic conditions. *Proceedings of the National Academy of Sciences of the United States of America*, *115*(41), 10517–10522.

Stroebel, D., Choquet, Y., Popot, J. L., & Picot, D. (2003). An atypical haem in the cytochrome b6f complex. *Nature*, *426*(6965), 413–418.

Suga, M., Ozawa, S. I., Yoshida-Motomura, K., Akita, F., Miyazaki, N., & Takahashi, Y. (2019). Structure of the green algal photosystem I supercomplex with a decameric light-harvesting complex I. *Nature Plants 2019 5:6*, *5*(6), 626–636.

Suorsa, M., Järvi, S., Grieco, M., Nurmi, M., Pietrzykowska, M., Rantala, M., Kangasjärvi, S., Paakkarinen, V., Tikkanen, M., Jansson, S., & Aro, E. M. (2012). PROTON GRADIENT REGULATION5 is essential for proper acclimation of Arabidopsis photosystem I to naturally and artificially fluctuating light conditions. *Plant Cell*, *24*(7), 2934–2948.

Szostak, J. W., Orr-Weaver, T. L., Rothstein, R. J., & Stahl, F. W. (1983). The double-strand-break repair model for recombination. *Cell*, *33*(1), 25–35.

Tagawa, K., Tsujimoto, H. Y., & Arnon, D. I. (1963). Role of chloroplast ferredoxin in the energy conversion process of photosynthesis. *Proceedings of the National Academy of Sciences of the United States of America*, *49*(4), 567–572.

Takagi, D., Takumi, S., Hashiguchi, M., Sejima, T., & Miyake, C. (2016). Superoxide and singlet oxygen produced within the thylakoid membranes both cause photosystem I photoinhibition. *Plant Physiology*, *171*(3), 1626–1634.

Takahashi, H., Clowez, S., Wollman, F. A., Vallon, O., & Rappaport, F. (2013). Cyclic electron flow is redox-controlled but independent of state transition. *Nature Communications*, *4*.

Tejero, J., Pérez-Dorado, I., Maya, C., Martínez-Júlvez, M., Sanz-Aparicio, J., Gómez-Moreno, C., Hermoso, J. A., & Medina, M. (2005). C-terminal tyrosine of ferredoxin-NADP+ reductase in hydride transfer processes with NAD(P)+/H. *Biochemistry*, *44*(41), 13477–13490.

Terashima, M., Petroutsos, D., Hüdig, M., Tolstygina, I., Trompelt, K., Gäbelein, P., Fufezan, C., Kudla, J., Weinl, S., Finazzi, G., & Hippler, M. (2012). Calcium-dependent regulation of cyclic photosynthetic electron transfer by a CAS, ANR1, and PGRL1 complex. *Proceedings of the National Academy of Sciences of the United States of America*, *109*(43), 17717–22.

Tibiletti, T., Auroy, P., Peltier, G., & Caffarri, S. (2016). Chlamydomonas reinhardtii PsbS Protein Is Functional and Accumulates Rapidly and Transiently under High Light. *Plant Physiology*, *171*(4), 2717.

Tikhonov, A. N. (2014). The cytochrome b6f complex at the crossroad of photosynthetic electron transport pathways. *Plant Physiology and Biochemistry*, *81*, 163–183.

Tokutsu, R., Fujimura-Kamada, K., Matsuo, T., Yamasaki, T., & Minagawa, J. (2019). The CONSTANS flowering complex controls the protective response of photosynthesis in the green alga Chlamydomonas. *Nature Communications*, *10*(1).

Toyokawa, C., Yamano, T., & Fukuzawa, H. (2020). Pyrenoid starch sheath is required for LCIB localization and the CO2-concentrating mechanism in green algae. *Plant Physiology*, *182*(4), 1883–1893.

Trissl, H. W., & Wilhelm, C. (1993). Why do thylakoid membranes from higher plants form grana stacks? *Trends in Biochemical Sciences*, *18*(11), 415–419.

Tyanova, S., Temu, T., Sinitcyn, P., Carlson, A., Hein, M. Y., Geiger, T., Mann, M., & Cox, J. (2016). The Perseus computational platform for comprehensive analysis of (prote)omics data. *Nature Methods 2016 13:9*, *13*(9), 731–740.

Uniacke, J., & Zerges, W. (2009). Chloroplast protein targeting involves localized translation in Chlamydomonas. *Proceedings of the National Academy of Sciences of the United States of America*, *106*(5), 1439–1444.

Untergasser, A., Nijveen, H., Rao, X., Bisseling, T., Geurts, R., & Leunissen, J. A. (2007). Primer3Plus, an enhanced web interface to Primer3. *Nucleic Acids Research*, *35*(2).

Vallon, O., Bulte, L., Dainese, P., Olive, J., Bassi, R., & Wollman, F. A. (1991). Lateral redistribution of cytochrome b6/f complexes along thylakoid membranes upon state transitions. *Proceedings of the National Academy of Sciences of the United States of America*, *88*(18), 8262–8266.

Vener, A. V., Van Kan, P. J., Rich, P. R., Ohad, I., & Andersson, B. (1997). Plastoquinol at the quinol oxidation site of reduced cytochrome bf mediates signal transduction between light and protein phosphorylation: Thylakoid protein kinase deactivation by a single-turnover flash. *Proceedings of the*

National Academy of Sciences of the United States of America, *94*(4), 1585–1590.

Vicino, P., Carrillo, J., Gómez, R., Shahinnia, F., Tula, S., Melzer, M., Rutten, T., Carrillo, N., Hajirezaei, M. R., & Lodeyro, A. F. (2021). Expression of Flavodiiron Proteins Flv2-Flv4 in Chloroplasts of Arabidopsis and Tobacco Plants Provides Multiple Stress Tolerance. *International Journal of Molecular Sciences*, *22*(3), 1–21.

Walters, R. G., & Horton, P. (1991). Resolution of components of non-photochemical chlorophyll fluorescence quenching in barley leaves. *Photosynthesis research*, *27*(2), 121–133.

Wang, L., Yamano, T., Takane, S., Niikawa, Y., Toyokawa, C., Ozawa, S. I., Tokutsu, R., Takahashi, Y., Minagawa, J., Kanesaki, Y., Yoshikawa, H., & Fukuzawa, H. (2016). Chloroplast-mediated regulation of CO2-concentrating mechanism by Ca2+-binding protein CAS in the green alga Chlamydomonas reinhardtii. *Proceedings of the National Academy of Sciences of the United States of America*, *113*(44), 12586–12591.

Wang, Y., Yau, Y.-Y., Perkins-Balding, D., & Thomson, J. G. (2011). Recombinase technology: applications and possibilities. *Plant Cell Reports*, *30*(3), 267–285.

Wietrzynski, W., Schaffer, M., Tegunov, D., Albert, S., Kanazawa, A., Plitzko, J. M., Baumeister, W., & Engel, B. D. (2020). Charting the native architecture of chlamydomonas thylakoid membranes with single-molecule precision. *eLife*, *9*.

Wild, M., Folini, D., Schär, C., Loeb, N., Dutton, E. G., & König-Langlo, G. (2013). The global energy balance from a surface perspective. *Climate Dynamics*, *40*(11-12), 3107–3134.

Xie, B., Bishop, S., Stessman, D., Wright, D., Spalding, M. H., & Halverson, L. J. (2013). Chlamydomonas reinhardtii thermal tolerance enhancement mediated by a mutualistic interaction with vitamin B12-producing bacteria. *ISME Journal*, *7*(8), 1544–1555.

Yamamoto, H., Peng, L., Fukao, Y., & Shikanai, T. (2011). An Src Homology 3 Domain-Like Fold Protein Forms a Ferredoxin Binding Site for the Chloroplast NADH Dehydrogenase-Like Complex in *Arabidopsis*. *The Plant Cell*, *23*(4), 1480–1493.

Yamamoto, H., & Shikanai, T. (2019). PGR5-Dependent Cyclic Electron Flow Protects Photosystem I under Fluctuating Light at Donor and Acceptor Sides. *Plant physiology*, *179*(2), 588–600.

Yamano, T., Tsujikawa, T., Hatano, K., Ozawa, S.-i., Takahashi, Y., & 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*(9), 1453–1468.

Yang, W., Catalanotti, C., Wittkopp, T. M., Posewitz, M. C., & Grossman, A. R. (2015). Algae after dark: mechanisms to cope with anoxic/hypoxic conditions. *The Plant Journal*, *82*(3), 481–503.

Yu, D., Ellis, H. M., Lee, E. C., Jenkins, N. A., Copeland, N. G., & Court, D. L. (2000). An efficient recombination system for chromosome engineering in Escherichia coli. *Proceedings of the National Academy of Sciences of the United States of America*, *97*(11), 5978–5983.

Yu, F., Haynes, S. E., & Nesvizhskii, A. I. (2021). IonQuant enables accurate and sensitive label-free quantification with FDR-controlled match-between-runs. *Molecular and Cellular Proteomics*, *20*, 100077.

Zhang, H., Whitelegge, J. P., & Cramer, W. A. (2001). Ferredoxin:NADP+ Oxidoreductase Is a Subunit of the Chloroplast Cytochrome b6f Complex. *Journal of Biological Chemistry*, *276*(41), 38159–38165.

Zhang, R., Patena, W., Armbruster, U., Gang, S. S., Blum, S. R., & Jonikas, M. C. (2014). High-Throughput Genotyping of Green Algal Mutants Reveals Random Distribution of Mutagenic Insertion Sites and Endonucleolytic Cleavage of Transforming DNA. *The Plant Cell*, *26*(4), 1398–1409.

Zito, F., Finazzi, G., Delosme, R., Nitschke, W., Picot, D., & Wollman, F. A. (1999). The Qo site of cytochrome b6f complexes controls the activation of the LHCII kinase. *The EMBO Journal*, *18*(11), 2961–2969.

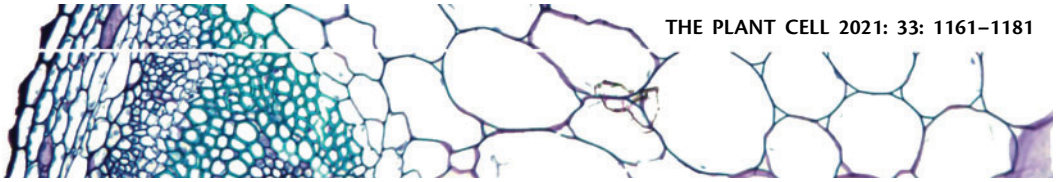
Appendix A

Appendix

doi:10.1093/plcell/koab024

THE PLANT CELL

THE PLANT CELL 2021: 33: 1161–1181



Large-Scale Biology

A recombineering pipeline to clone large and complex genes in Chlamydomonas

Tom Z. Emrich-Mills ,[1,2,†] Gary Yates ,[1,†] James Barrett ,[1] Philipp Girr ,[1] Irina Grouneva ,[1] Chun Sing Lau ,[1] Charlotte E. Walker ,[1] Tsz Kam Kwok ,[1] John W. Davey ,[1] Matthew P. Johnson [2] and Luke C.M. Mackinder [1,*,‡]

1 Department of Biology, University of York, York YO10 5DD, UK
2 Department Molecular Biology and Biotechnology, University of Sheffield, Sheffield S10 2TN, UK

*Author for correspondence: luke.mackinder@york.ac.uk (L.C.M.M.)
†These authors contributed equally (T.Z.E-M., G.Y.).
‡Senior author.

T.Z.E.M. developed the initial recombineering pipeline, designed and assembled the original pLM099 recombineering plasmid and performed the genome wide analysis. G.Y., T.Z.E.M., and T.K.K. assembled additional recombineering plasmids. T.Z.E.M. and G.Y. optimized and performed the large-scale recombineering pipeline. G.Y. performed the microscopy and Venus quantification data. P.G. validated the pipeline using fosmids. J.B. performed the complementation experiments. J.B., I.G., C.S.L., C.E.W., and T.K.K. supported the development and implementation of the recombineering pipeline. J.W.D. wrote the BACSearcher code and provided bioinformatics support to T.Z.E.M. for the remaining code. L.C.M.M. conceived the idea and led the research. L.C.M.M. and M.P.J. received funding to support the work. L.C.M.M., T.Z.E.M., and G.Y. wrote the manuscript.

The author responsible for distribution of materials integral to the findings presented in this article in accordance with the policy described in the Instruction for Authors (https://academic.oup.com/plcell) is Luke C.M. Mackinder (luke.mackinder@york.ac.uk).

Abstract

The ability to clone genes has greatly advanced cell and molecular biology research, enabling researchers to generate fluorescent protein fusions for localization and confirm genetic causation by mutant complementation. Most gene cloning is polymerase chain reaction (PCR) or DNA synthesis-dependent, which can become costly and technically challenging as genes increase in size, particularly if they contain complex regions. This has been a long-standing challenge for the *Chlamydomonas reinhardtii* research community, as this alga has a high percentage of genes containing complex sequence structures. Here we overcame these challenges by developing a recombineering pipeline for the rapid parallel cloning of genes from a Chlamydomonas bacterial artificial chromosome collection. To generate fluorescent protein fusions for localization, we applied the pipeline at both batch and high-throughput scales to 203 genes related to the Chlamydomonas CO_2 concentrating mechanism (CCM), with an overall cloning success rate of 77%. Cloning success was independent of gene size and complexity, with cloned genes as large as 23 kb. Localization of a subset of CCM targets confirmed previous mass spectrometry data, identified new pyrenoid components, and enabled complementation of mutants. We provide vectors and detailed protocols to facilitate easy adoption of this technology, which we envision will open up new possibilities in algal and plant research.

Introduction

The unicellular alga *Chlamydomonas reinhardtii* (hereafter Chlamydomonas) is a widely used model organism for studying photosynthesis, biofuel production, human ciliopathies, flagella-powered motility, and cell cycle control (Salomé and Merchant, 2019). The nuclear, chloroplast, and

Received May 07, 2020. Accepted January 18, 2021. Advance access publication February 2, 2021

mitochondrial genomes of Chlamydomonas have been sequenced, are well annotated and transformable. In addition, a variety of genetic resources are available to any institution, including a close-to-genome-saturating mutant library (Li et al., 2019), extensive-omics-based data, and a wealth of molecular tools developed over decades by a dedicated research community (Salomé and Merchant, 2019). These collections, data, and tools are a vital resource for studies that aim to understand fundamental biological processes, to guide engineering efforts such as improved photosynthetic efficiency, and to enable efficient biomolecule production.

Reverse genetic approaches in Chlamydomonas often depend on localizing target proteins to understand spatial distribution and the complementation of mutants to link genotype to phenotype. Both of these methods generally rely on cloning a gene of interest into a plasmid from genomic DNA (gDNA) by PCR, followed by amplification in *Escherichia coli* and reintroduction into Chlamydomonas cells. PCR-based cloning from gDNA presents its own challenges and limitations that are particularly problematic when working with Chlamydomonas nuclear genes, which generally have a high guanine-cytosine (GC) content (68% in coding regions), contain one or more introns, and can include complex repeating regions (Merchant et al., 2007). On the other hand, cloning from complementary DNA can result in low or no expression of target genes, most likely due to a lack of introns and regulatory elements (Lumbreras et al., 1998; Schroda, 2019). Some of the challenges associated with PCR-based cloning can be circumvented via whole or partial gene synthesis followed by re-assembly using cloning strategies such as Golden Gate. Although the falling costs of gene synthesis make this a viable option for some genes, for many others the need to include introns, high GC content and high gene complexity, typical of the Chlamydomonas nuclear genome, results in synthesis failure or is prohibitively expensive. For example, the cloning of *SAGA1* (encoding StArch Granules Abnormal 1), a 16.7-kbp gene target, required over 12 months of work, included multiple gene synthesis failures, and ultimately had to be assembled from three synthesized fragments with 14 introns removed due to repetitive regions (Itakura et al., 2019).

Improved Chlamydomonas target gene and foreign gene (collectively transgene) expression (e.g. *GFP*) have been achieved through strain optimization (Neupert et al., 2009), the development of systems with linked transgene and antibiotic resistance gene expression (Rasala et al., 2012; Onishi and Pringle, 2016), and an advanced understanding of transgene silencing (reviewed in Schroda, 2019). Furthermore, the release of the Chlamydomonas Golden Gate-based Modular Cloning kit has provided a cloning framework and a selection of genetic elements to enable laboratories to rapidly assemble and test transgene constructs (Crozet et al., 2018). Independent of background strain and expression system, it is now clear that inserting or maintaining introns, correct codon usage, and promoter sequence are all critical for robust transgene expression (Barahimipour et al., 2015; López-Paz et al., 2017; Baier et al., 2018; Weiner et al., 2018; Schroda, 2019). These considerations have made the cloning of Chlamydomonas target genes directly from gDNA the community standard for mutant complementation and fluorescent protein tagging. However, there are considerable technical hurdles to overcome when working with the expression of large Chlamydomonas genes, predominantly caused by inefficient amplification of gDNA due to gene size, GC content, and the complexity of target genes (Sahdev et al., 2007). Although modern polymerases have been engineered to overcome sequence challenges (Hommelsheim et al., 2014), they may still suffer from replication slippage events, which are exacerbated by the presence of repetitive regions (Levinson and Gutman, 1987; Clarke et al., 2001). In addition to considerations of size and complexity, cloning native genes based on current genome annotations can be complicated by the abundance of upstream transcription start sites corresponding to possible alternative open reading frames (ORFs; Cross, 2015) and, hence, potentially resulting in incorrect target gene cloning.

The results of a recent high-throughput localization study illustrate the challenges of PCR-based cloning of Chlamydomonas nuclear genes (Mackinder et al., 2017). In Mackinder et al. (2017), genes were PCR amplified from start site to stop site using gDNA as the template. Amplicons were then cloned in-frame via Gibson assembly with a fluorescent protein and a constitutive promoter and terminator, resulting in the successful cloning of 298 genes out of an attempted 624 (48% success rate), with most failures at the PCR amplification step. This relatively low success rate prompted us to develop a cloning platform based on recombination-mediated genetic engineering (recombineering) to allow Chlamydomonas genes to be cloned independently of size or sequence. Recombineering allows gene cloning to be performed by homologous recombination in *E. coli* without the need for PCR amplification of the template, making this technique predominantly independent of the size of the target region. Large-scale recombineering pipelines have been developed for bacterial artificial chromosome (BAC) and fosmid libraries from a broad range of organisms including the round worm *Caenorhabditis elegans* (Sarov et al., 2006), fruit fly (*Drosophila melanogaster*; Sarov et al., 2016), human and mice (Poser et al., 2008), and *Arabidopsis thaliana* (Brumos et al., 2020). Our pipeline involves making BAC-containing *E. coli* homologous recombination competent by introducing the recombinogenic viral proteins Red α, β, and γ from the bacteriophage lambda virus (Yu et al., 2000; Copeland et al., 2001) and then retrieving the target sequence by introducing 50-bp homology regions flanking a linearized plasmid.

We decided to apply our recombineering pipeline to an extended list of putative CO_2 concentrating mechanism (CCM) genes. The CCM enhances photosynthesis by increasing the concentration of CO_2 around Rubisco. To achieve this, Chlamydomonas actively accumulates inorganic carbon in the chloroplast and delivers it as CO_2 to tightly packed Rubisco within the pyrenoid (Wang et al., 2015). The pyrenoid is essential for CCM function in Chlamydomonas (Meyer et al., 2012; Mackinder et al., 2016). Due to the

photosynthetic turbocharging properties of pyrenoid-based CCMs, there is growing interest in engineering them into crop plants to boost yields (Mackinder, 2017; Rae et al., 2017). Recent studies have identified a large number of potential pyrenoid and CCM components (Mackinder et al., 2017; Zhan et al., 2018) that require functional characterization to understand their utility for future synthetic CCM engineering efforts. However, many of the genes encoding these components are challenging to clone due to size and sequence complexity, making localization and mutant complementation studies difficult.

Here, by applying our pipeline, we successfully cloned 157 CCM-related genes with their native promoters. Successful cloning using our system appears to be independent of target gene size, and many target genes have multiple complex features that would typically result in PCR failure. The average cloned region was 7.3 kbp, and target regions up to 22.7 kbp in size were successfully cloned. The inclusion of the native promoters ensures that any upstream ORFs (uORFs) have been incorporated. The localization of a subset of the proteins encoded by these genes allowed us to identify their diverse cellular locations, confirming previous interaction data (Mackinder et al., 2017) and pyrenoid proteomic data (Mackinder et al., 2016; Zhan et al., 2018). We developed a series of recombineering vectors with a range of different fluorophores, epitope tags, and selection markers that can be used for protein localization, protein interaction, mutant complementation, and relative protein abundance studies. Our method takes 4 days to implement, is accessible for any laboratory equipped for molecular biology, and requires no specialized reagents or equipment. The BAC library used in this work and all developed vectors are available from the Chlamydomonas Resource Center. Finally, we provide a detailed protocol, allowing this method to be rapidly adopted by research laboratories to clone nuclear Chlamydomonas genes.

Results

Analysis of the Chlamydomonas genome highlights the challenges affecting PCR-based cloning

Cloning Chlamydomonas genes for successful localization and complementation studies often requires the amplification of complete ORFs from gDNA, spanning from their start site to their stop site including any introns (ATG-Stop). To gain a better understanding of the challenges involved in cloning Chlamydomonas genes, we performed a whole-genome analysis of gene size, complexity, intron prevalence, splice variants, and ATG-Stop primer suitability in Chlamydomonas, including comparisons to available datasets and other organisms.

Gene size

A major limitation of PCR-based cloning is the size of the target amplicon. Analysis of ATG-Stop cloning data from Mackinder et al. (2017) for 624 genes using gDNA as a template and Phusion Hot Start II DNA polymerase (ThermoFisher Scientific, Waltham, MA, USA) revealed an association between cloning success and gene size; the average cloned ATG-Stop region was ~2.3 kbp long, while the average uncloned region was ~4.5 kbp (Mann–Whitney $U = 16{,}306$, $P < 0.001$, two-tailed). Extrapolation of PCR efficiency relative to target size from Mackinder et al. (2017) to whole genes in the Chlamydomonas genome (version 5.5) indicated that 68% of genes would be technically challenging to clone via PCR-based methods (Figure 1A), predominantly due to a severe drop off in amplification efficiency for genes >3-kbp long. The largest amplified target in Mackinder et al. (2017) was 8 kbp, and genes at least as large as 9.7 kbp have been cloned before (Kobayashi et al., 2015), but this appears to be highly gene-specific. Alternative approaches exist to clone larger genes, such as testing a broad range of PCR conditions and DNA polymerases, amplification in fragments and re-stitching together, cloning from cDNA, and gene synthesis. While some of these approaches avoid the challenges presented here, they can be time consuming, costly, have low success rates, and may still result in no or poor expression.

Gene complexity

High GC content and the presence of numerous repetitive regions can make PCR-based cloning challenging. Analysis of data from Mackinder et al. (2017) showed that the average GC content for successfully cloned targets by ATG-Stop PCR cloning was 61.4%, while the average GC content for unsuccessful targets was 64.3%—a value exceeded by over 41% of Chlamydomonas nuclear genes. To analyze the genome for repetitive regions, we determined the frequency of simple tandem repeats, inverted repeats, and larger, interspersed repeats between the start of the 5′-untranslated region (UTR) and the end of the 3′-UTR of each gene. Tandem repeats were assessed by counting individual regions that consist of consecutive mono-, di-, or trinucleotide repeats. Mononucleotide repeats shorter than 10 bp and regions of di- and trinucleotide repeats shorter than 20 bp were excluded. Some slight imperfections in the repeating pattern of a region were allowed, with regions that showed ⩾90% identity (such as GGGGGTGGGG) included. Of the 17,741 coding genes in the nuclear genome, 8,810 contain one or more mono-, di-, or trinucleotide repeats (Figure 1B). In terms of prevalence per kilobase, the average Chlamydomonas gene contains 0.21 tandem repeats, whereas Arabidopsis contains 0.16 and the yeast *Saccharomyces cerevisiae* contains 0.10. Interestingly, if polynucleotide repeats with higher period numbers are counted as well (from tetranucleotide repeats to tandem repeating units of hundreds of base pairs), these values increase 5-fold for Chlamydomonas (1.07 per kbp), 2.5-fold for Arabidopsis (0.39 per kbp), and 3-fold for yeast (0.3 per kbp), highlighting the repetitive nature of the Chlamydomonas genome. We assessed inverted repeats by counting regions over 10-bp long that are followed closely downstream by their reverse complement, with some mismatches allowed so that regions with ⩾90% identity were included. 14,454 genes

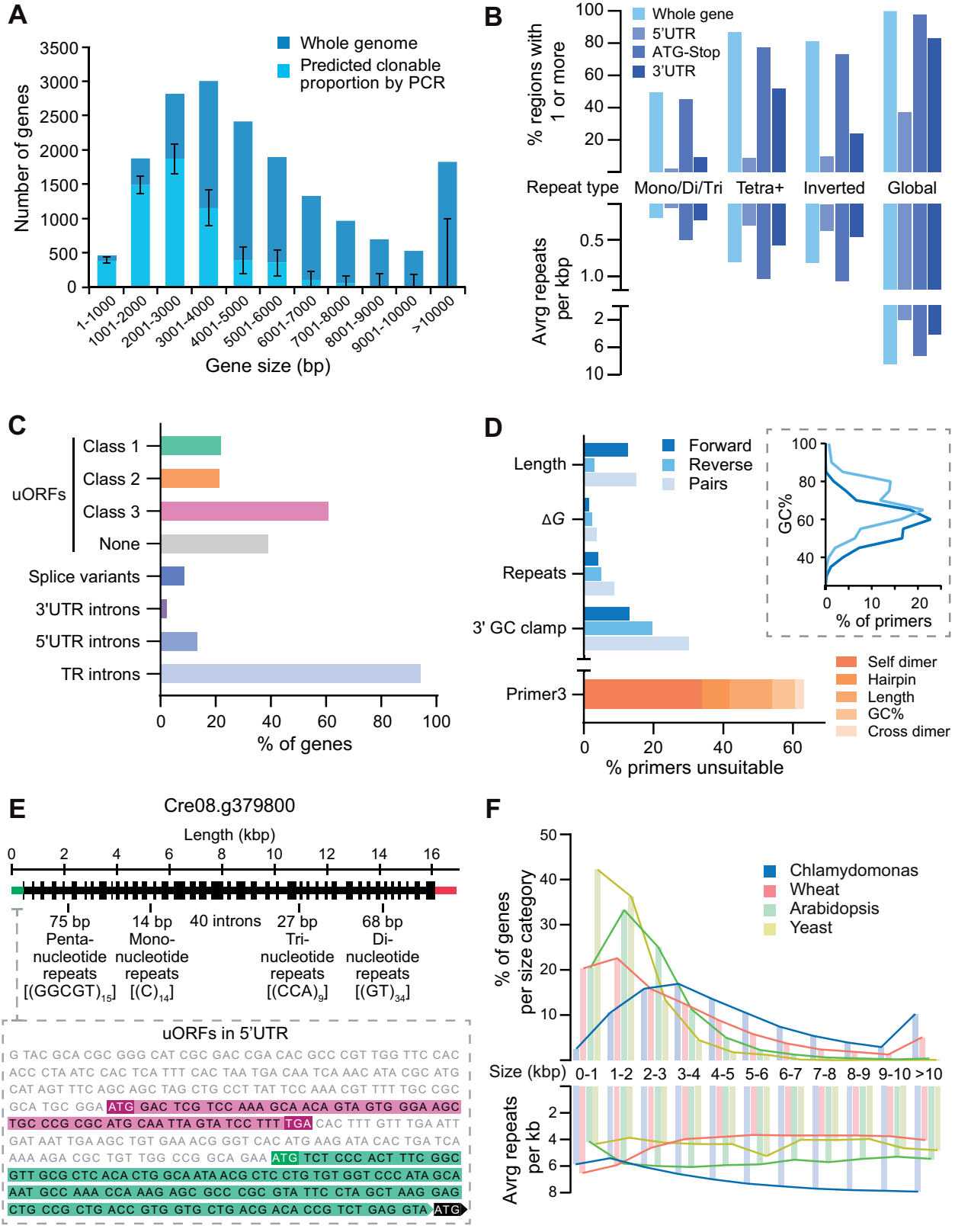


Figure 1 Chlamydomonas nuclear genes are often large, complex, or misannotated, affecting PCR-based cloning attempts and transgene expression success. A, The distribution of gene sizes for the 17,741 genes in the Chlamydomonas nuclear genome (dark blue). Gene sizes are measured from the start of the 5′-UTR to the end of the 3′-UTR. Within each size category, the predicted proportion amenable to PCR-based cloning is shown in light blue. These proportions were extrapolated from cloning success for 624 CCM-related genes from Mackinder et al. (2017) in which PCR-based cloning was used to amplify the ATG-Stop region of each gene, excluding any UTRs. The strong size-dependence of ATG-Stop cloning

contain one or more inverted repeats of this kind (Figure 1B), with an average of 0.93 repeats per kbp.

To further validate these findings, we analyzed nuclear gene sequences for repeats using WindowMasker, a program for detecting global repeats that include larger non-adjacent sequences as well as a diverse range of tandem repeats and inverted repeats (Morgulis et al., 2006). With this expanded detection range, Chlamydomonas genes contain an average of 38.9 repeats (6.8 per kbp), whereas Arabidopsis contains 13.7 (5.5 per kbp) and yeast contains 6.0 (4.2 per kbp). On average, Chlamydomonas genes are more repetitive between their start and stop codons than in their untranslated regions (Figure 1B), although at least one repeat was detected by WindowMasker in 36.6% of 5′-UTRs and 87.6% of 3′-UTRs. Crucially, analysis of sequence data from Mackinder et al. (2017) for 624 Chlamydomonas genes revealed an association between ATG-Stop PCR cloning success and repeat frequency: the average cloned ATG-Stop region contained 6.1 repeats per kbp, whereas the average uncloned region contained 7.5 repeats per kbp (Mann–Whitney $U = 24{,}110$, $P < 0.001$, two-tailed).

Mis-annotation of start sites

Another challenge associated with PCR-based and gene synthesis-based cloning is the presence of incorrectly annotated gene models, which leads to the cloning of a non-biologically relevant sequence. An analysis of transcript models in the Chlamydomonas genome showed that additional ATGs upstream of the annotated start site are highly prevalent (Cross, 2015; Figure 1C top 4 bars). Cross (2015) categorized these potential uORFs into three classes: Class 1 uORFs, which initiate in-frame with the annotated start site, potentially producing an N-terminal extension relative to the annotated gene model; Class 2 uORFs, which initiate out-of-frame with the annotated start site and terminate within the coding sequence; and Class 3 uORFs, which initiate and terminate within the 5′-UTR. Data from Cross (2015) on the presence of translation initiation sites (Kozak sequences) preceding class 1 uORFs suggest that approximately half are the correct translation initiation site *in vivo*. In a PCR-based approach where a constitutive promoter is used, cloning from the wrong ATG may result in an out-of-frame or truncated product, potentially removing essential signal sequences for correct targeting. Fifty-seven of the 298 successfully cloned genes from Mackinder et al. (2017)

Figure 1 (continued)

efficiency seen in 2017 indicates that 68% of the genome would be challenging to clone. 95% confidence intervals for the predicted clonable proportions of each size category were calculated using the Wilson score interval method. No genes over 8,000 bp are predicted to be clonable by PCR although only a handful of regions of these sizes were tested in 2017 giving rise to the large confidence intervals for these categories. B, Genome-wide sequence complexity, as indicated by the presence of one or more repetitive sequences and frequency of repeats per kbp in each gene (pale blue). Values are also given for repeats localized to the 5′-UTR (light indigo), ATG-Stop (indigo), and 3′-UTR (dark indigo) within each gene. Note that while all 17,741 genes contain a start-to-stop region, not all genes contain a 5′-UTR and/or 3′-UTR, so the percentages presented for these are relative to totals of 17,721 and 17,717, respectively. Simple repeats are shown in the left three categories. Mono/di/tri refers to tandem repeats with a period length of one, two or three; tetra + refers to all oligonucleotide tandem repeats with a period length of 4 or more and a total length $\geqslant$20 bp. Combining whole-gene counts for mono-, di-, tri-, and tetra + produces an average value of 1.07 tandem repeats per kbp. Inverted repeats refer to short (20–210 bp) sequences that have the potential to form secondary structures by self-complementary base pairing. About 836 genes were free from detectable tandem and inverted repeats under our criteria, most of which are small, with an average length of 1,766 bp. Global repeats refer to repetitive sequences masked by the National Centre for Biotechnology Information (NCBI) WindowMasker program (Morgulis et al., 2006), which includes both longer, non-adjacent sequences and shorter, simple repeats (see Methods section). All genes contained detectable repetitive regions using the default WindowMasker settings, with an average of 40.07 per gene. UTR data are based on gene models from Phytozome (version 5.5). C, Gene features that complicate correct transgene expression. Top four bars illustrate potential misannotation of functional start sites in the genome shown by the percentage of genes containing one or more uORFs of each class (see text). Note that some genes contain multiple classes of uORF. Shown below this is the percentage of Chlamydomonas genes with multiple transcript models (splice variants), and those containing introns in the UTRs and TRs (between start and stop codons). uORF data are from Cross (2015). Splice variant and intron data are based on gene models from Phytozome (version 5.5). D, Analysis of a set of ATG-Stop PCR primers designed to clone every gene in the genome from start to stop codon using gDNA as the template (Mackinder et al., 2017). Many primers are predicted to be unsuitable for efficient PCR, as shown by the percentage of forward (dark blue) and reverse (light blue) primers that breach various recommended thresholds associated with good primer design. Pairs (pale blue) are shown for which one or both primers breach the respective thresholds. Thresholds shown pertain to length, secondary structure stability, tandem repeats, and 3′-GC content. The inset shows the distribution of GC content of primers in the dataset, illustrating a clear trend in higher GC content at the 3′-end of coding sequences. Below this, the given reason for rejection of primers by the Primer3 check_primers module is shown in orange. Dimer and hairpin values refer to primers rejected for "high end complementarity" and "high any complementarity" errors, respectively. E, Annotated gene structure of Cre08.g379800. The gene encodes a predicted protein of unknown function but shows examples of several sequence features that contribute to sequence complexity. The unspliced sequence is 16,892 bases long with a GC content of 64.3%. The 41 exons are shown as regions of increased thickness, with 40 introns between them, the annotated 5′-UTR in green (left) and the 3′-UTR in red (right). Labels denote selected examples of simple repeats throughout the gene. The inset shows the 5′-UTR sequence, displaying examples of two classes of uORFs (see text); Class 3 is highlighted in magenta and Class 1 in green. For simplicity, only one of the seven class 3 uORFs are shown in full. Cre08.g379800 was successfully cloned and tagged using recombineering. F, A comparison of gene size and complexity between Chlamydomonas, bread wheat (*Triticum aestivum*), *A. thaliana* and *Saccharomyces cerevisiae*. Gene sizes were binned as in (A), and the average number of global repeats kbp masked by the NCBI WindowMasker program was counted for genes in each size category (Morgulis et al., 2006). Genes were measured from the start of the 5′-UTR to the end of the 3′-UTR.

contained a Class 1 in-frame ATG upstream of the cloned region; therefore, ~10% of cloned regions may have encoded truncated protein products.

Introns, UTRs, and splice variants

Chlamydomonas genes have a relatively high frequency of introns, providing a further challenge for PCR-based cloning. The average gene contains 7.3 introns with an average intron length of 373 bp compared to an average exon length of 190 bp. Ninety-four percent of genes contain introns between their start and stop codons, 13% of genes contain one or more introns in their 5′-UTRs, and 3.4% have introns in their 3′-UTRs. ATG-Stop cloning would omit introns in UTRs, potentially missing critical regulatory information. Furthermore, ~9% of genes are annotated with two or more transcript models that result from alternative splicing (Figure 1C). This variation would be missed by cloning from cDNA or by gene synthesis that excludes native introns.

Unsuitable primers

When performing ATG-Stop PCR cloning of either gDNA or cDNA, the flexibility of primer design is limited. Sequence analysis of a set of genome-wide primer pairs for ATG-Stop cloning (Mackinder et al., 2017) indicates that primers are frequently of poor quality and unsuitable for efficient PCR. The average primer in the dataset had a predicted melting temperature (T_m) of 69.2°C and an average GC content of 64.2%. The primer T_m and GC contents of Chlamydomonas genes are expected to be high compared to other organisms with less GC-rich genomes; however, many primers also breached the recommended thresholds pertaining to length, secondary structure formation, repetitive sequences, and 3′-GC content. Primers are shown in Figure 1D (blue bars) as having breached these four thresholds if (1) they were longer than 30 bp; (2) the free energy (ΔG) required to disrupt secondary structure formation (self-dimers, cross-dimers, or hairpins) was less than −9 kcal mol^{-1} at PCR-relevant annealing temperatures (66–72°C); (3) they contained mono- or dinucleotide repeats of 5 or more; or (4) their 3′-ends contained five or more consecutive G/C bases. A stricter set of thresholds is utilized by the Primer3 check_primers module (Rozen and Skaletsky, 2000), which resulted in the rejection of over 60% of individual primers in the dataset, even when the program was set to ignore predicted annealing temperatures (Figure 1D, orange bar). Under these settings, only 13% of pairs were free from detectable issues in both primers. Interestingly, there is a mismatch in GC content between forward and reverse primers, with reverse primers having a considerably higher GC content (Figure 1D, inset).

Many individual genes contain a range of the above features that result in challenges faced during PCR cloning or gene synthesis. Figure 1E shows a gene from chromosome 8 that exhibits several examples and was a target for recombineering. Cre08.g379800 is >16-kbp long with 40 introns and contains mono-, di-, tri-, and pentanucleotide repeat regions of ⩾9 repeats. It also contains a potential misannotated upstream ATG in the 5′-UTR that could initiate a Class 1 uORF, as well as seven Class 3 uORFs (Cross, 2015). Structural information for Cre08.g379800 was obtained from the version 5.5 gene model currently available on Phytozome.

To investigate whether the challenges faced in Chlamydomonas were similar in other organisms, we analyzed gene size and gene complexity relative to gene size for the model eukaryote *S. cerevisiae*, the model plant Arabidopsis, and the ~17 Gb hexaploid genome of bread wheat (*Triticum aestivum*). As shown in Figure 1F, Chlamydomonas has a higher proportion of long genes and fewer short genes than the three other genomes tested, along with a considerably higher average gene size (5,322 bp versus 1,430 bp for yeast, 2,187 bp for Arabidopsis, and 3,521 bp for chromosome-assigned genes in wheat). Unlike wheat, Arabidopsis, and yeast, Chlamydomonas genes show a trend of increasing complexity per kilobase for longer genes (Figure 1F), which might be in line with the observation that average UTR length increases with increasing gene length (Salomé and Merchant, 2019).

Development of our recombineering pipeline

To overcome the challenges associated with PCR-based cloning, we developed a high-throughput recombineering pipeline for large-scale parallel cloning of Chlamydomonas nuclear genes from BACs with their native promoter regions intact. During pipeline development, we decided to pursue a simplified 1-step DNA retrieval recombineering approach rather than a BAC editing approach (i.e. Poser et al., 2008; Brumos et al., 2020) for several reasons: (1) using a gene retrieval method allows all cloning to be performed in the BAC host *E. coli*, thereby avoiding the need for BAC purification, which can be timely and low yielding; (2) assembled constructs contain only the gene of interest, making them considerably smaller than the original BAC. This allows a medium copy origin of replication to be used, which improves the ease of handling, and the smaller constructs minimize DNA fragmentation during Chlamydomonas transformation (Zhang et al., 2014); (3) BACs contain many genes, and therefore, additional copies of genes adjacent to the gene of interest could have an unwanted phenotypic impact on transformed Chlamydomonas lines; (4) the backbone of the available BAC collection lacks a suitable Chlamydomonas selection marker. Therefore, additional BAC editing to insert a suitable selection marker (Aksoy and Forest, 2019) or inefficient and poorly understood plasmid co-transformation strategies would be required for selection; and (5) a typical BAC engineering approach would require two recombination steps, which would increase the time needed for the pipeline, decrease pipeline efficiency, and add further challenges due to the repetitive nature of the Chlamydomonas genome.

The simplicity of our pipeline allows the entire process to be completed in 4 days using only generic reagents. The final recombineered construct is a vector containing the target region (typically including the native promoter, 5′-UTR, and ORF) recombined in-frame with a downstream

fluorescent protein gene followed by the photosystem I subunit II (*PSAD*) terminator (see Figure 2 for a pipeline schematic and Supplemental Protocol 1 for details). Our pipeline has four key steps: (1) *E. coli* harboring a BAC containing the gene of interest is made recombination competent by transformation with the pRed vector containing the lambda viral *exo*, *beta*, and *gam* genes (*Red*$\alpha\beta\gamma$) and *recA* (Sarov et al., 2006; Figure 2A); (2) *Red*$\alpha\beta\gamma$ and *recA* are induced by arabinose treatment, followed by transformation with a linear tagging cassette including 50-bp homology arms to the target gene (Figure 2B); (3) kanamycin selection is performed to identify successful recombination events, and temperature-induced inhibition of the pRed pSC101 replication origin is performed to minimize further undesired recombination (Figure 2C); and (4) the plasmid is isolated and verified via restriction digest and junction sequencing (Figure 2D).

The original tagging cassette consists of the codon-optimized yellow fluorescent protein (YFP) gene *CrVenus*, a 3xFLAG tag, the *PSAD* terminator, the paromomycin selection marker (*AphVIII*), the p15A medium-copy-number origin of replication, and the kanamycin resistance gene (Kan^R). Amplification of the tagging cassette from pLM099 is performed using primers containing 50-bp homology arms corresponding to regions flanking the target gene; the forward primer is located at least 2,000-bp upstream of the start codon to encompass the native 5′-promoter and UTR and the reverse primer is located at the 3′-end of the coding region (immediately upstream of the stop codon). The annealing site of the reverse primer can easily be altered to amplify a cassette from pLM099 to clone genes without a fluorescent tag or with only the 3xFLAG tag (see Supplemental Protocol 1). To minimize false positives due to pLM099 carryover, pLM099 contains the control of cell death B (*ccdB*) counter-selection gene (Bernard and Couturier, 1992). In addition, the cassette includes an I-SceI restriction site. I-SceI has an 18-bp recognition site not found within the reference Chlamydomonas genome (strain CC-503) and allows the cassette to be linearized prior to transformation into Chlamydomonas.

We initially tested our pipeline on 12 targets. To ensure that the BAC library (available from the Chlamydomonas Resource Center; https://www.chlamycollection.org/) was correctly mapped, we performed PCR to check for the presence of the 5′- and 3′-ends of our target genes (Supplemental Figure 1A). We next implemented the pipeline using a small-scale batch protocol (Supplemental Protocol 1A). For all targets except one, plasmids isolated from most picked colonies gave a correct banding pattern after restriction digest (Supplemental Figure 1B). Sequence confirmation indicated that we successfully cloned 11 out of our 12 targets, resulting in a 92% success rate (Supplemental Figure 1C). To further expand the capabilities of our pipeline, we tested whether we could successfully recombineer a large and complex gene from a fosmid (available from the Chlamydomonas Resource Center). We targeted *SAGA1* (from fosmid VTP41289), which had previously been highly challenging to synthesize (see above; Itakura et al., 2019) and was not available in the BAC library. Recombineered plasmids purified from three colonies all showed the correct restriction digestion pattern (Supplemental Figure 1D). Sequencing confirmed that the 19,601-bp target region, including 2,913-bp upstream of the predicted *SAGA1* start codon, was successfully cloned. Confident that our recombineering method was robust, we pursued the development of a large-scale pipeline that would allow the parallel tagging of genes with most steps achievable in 96-well format.

Successful large-scale application of the recombineering pipeline

To test the efficiency of the pipeline, we shortlisted 191 genes that could be mapped to clones from the Chlamydomonas BAC library. To more easily identify BACs within the library that contain a target gene, we designed a Python script (BACSearcher; see Supplemental Protocol 2) and outputted the five smallest BACs for all targets in the genome (Supplemental Data Set 1), revealing that 86% of nuclear genes are covered by at least one BAC (87% if BACs are included that terminate within 3′-UTRs). BACSearcher can also be used for the automated design of primers containing 50′-bp homology regions to target genes in optimal positions; the script reports suitable 5′-homology regions 2,000- to 3,000-bp upstream of the annotated start codon and takes into account local DNA complexity features, including mono- and dinucleotide repeating runs and GC content. This feature can be easily modified to design 5′-homology regions further upstream of the target (see Supplemental Protocol 2A). The length of 50 bp is short enough to design into an oligonucleotide but long enough to be unlikely to share homology with more than one site within a BAC. Supplemental Data Set 1 includes sequences for the top five optimal 5′-homology regions for each target, all more than 2,000-bp upstream of the start codon, along with the corresponding 50 bp 3′-homology region. In addition, four pairs of primer sequences are included that can be used to check for the presence of each target in a BAC.

Our 191 targets were primarily chosen based on our 2017 association study for CCM components (Mackinder et al., 2017), transcriptomics (Brueggeman et al., 2012; Fang et al., 2012), and pyrenoid proteomics (Mackinder et al., 2016; Zhan et al., 2018). Eighty-one genes previously targeted in 2017 were retried here by recombineering, this time with more than 2,000-bp upstream sequences included. Forty-one of these were previously unsuccessfully cloned by PCR and 40 were successfully cloned but were included here to compare the effect of retaining the native promoter. These included five targets that contain a Class 1 uORF (Cross, 2015) and so may have previously produced misleading localization data due to expression of a truncated protein. Selection of the remaining 110 targets was guided by new pyrenoid proteome (Zhan et al., 2018) and CCM interactome data

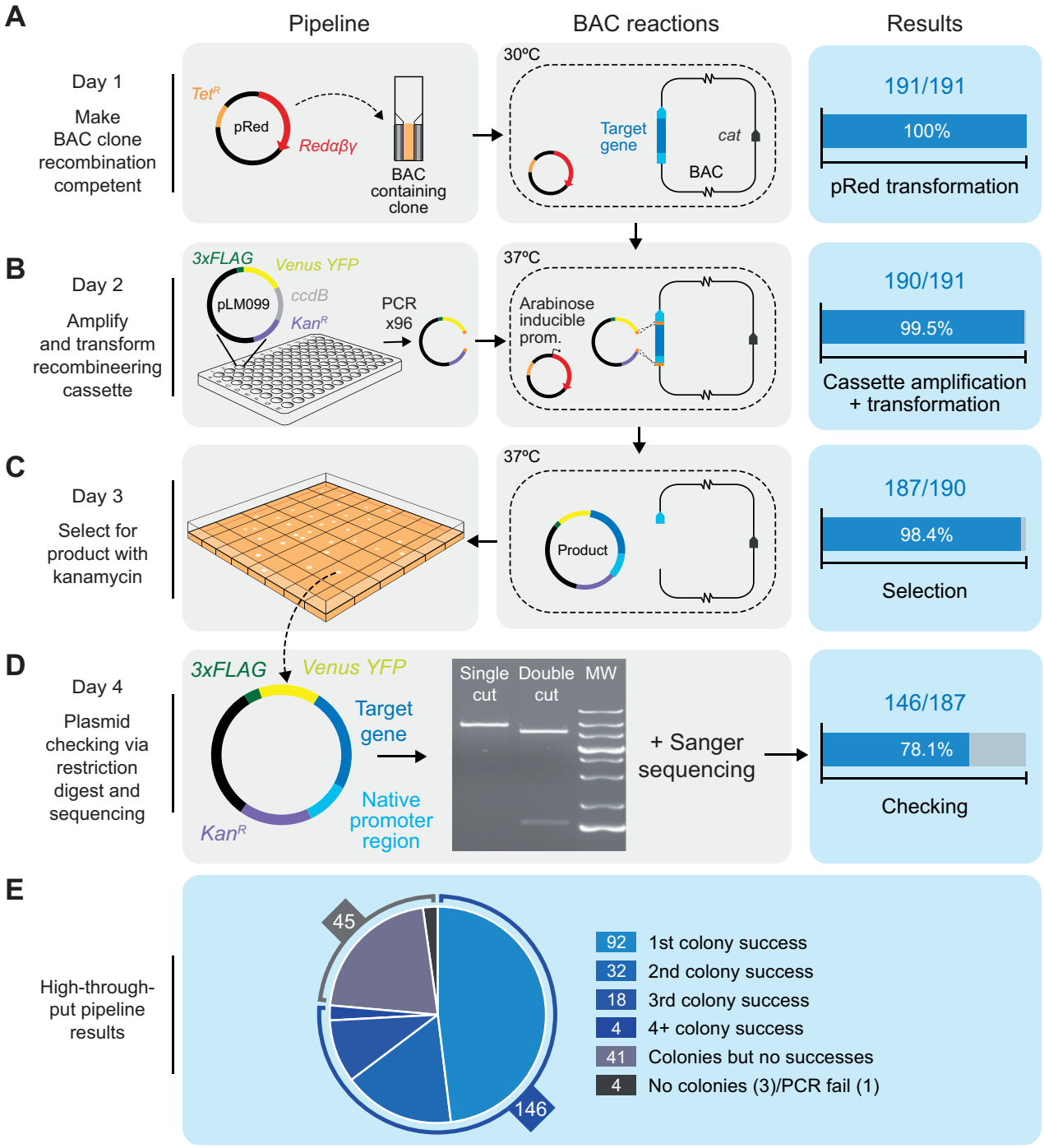


Figure 2 High-throughput recombineering pipeline for generating Venus-tagged fusion proteins with the native promoter regions intact. A, On Day 1, BAC clones containing target genes are made recombineering competent by transformation with the pRed plasmid, which encodes the viral recombinogenic *Redαβγ* genes and *recA* under the control of an arabinose inducible promoter. Transformation efficiency shown on the right-hand side relates to BAC clones that yielded colonies after selection with tetracycline and chloramphenicol. *Cat*: the chloramphenicol resistance gene in the backbone of every BAC clone in the BAC library. B, On or before Day 2, the recombineering cassette is amplified from pLM099 using primers that contain 50-bp homology arms complementary to regions flanking the target gene (shown in orange): one >2,000-bp upstream of the annotated ATG and one at the 3′-end of the coding sequence. On Day 2, BAC-containing cells are electrotransformed with the recombineering cassette after induction with L-arabinose. Recombination between the BAC and the cassette results in a plasmid product containing the target gene in frame with CrVenus-3xFLAG and under the control of its native promoter. Efficiency shown at this stage relates to PCRs that yielded efficient amplification of the recombineering cassette. C, On Day 3, colonies containing plasmid products are isolated. Efficiency at this stage relates to the number of transformations that yielded colonies after selection with kanamycin. D, On Day 4, plasmid products are extracted from cells, screened by enzymatic digestion and confirmed by sequencing. Efficiency shown at this stage relates to correct digest patterns with single and double cutting restriction enzymes. MW, molecular weight marker. E, Overall efficiency split into number of colonies screened via restriction digest. For 74% of target regions, the correct digest pattern was observed from plasmids isolated from the first, second or third colony picked per target. For 3% of targets, analyzing >3 colonies yielded the correct product.

(Mackinder et al., 2017). *Escherichia coli* strains containing the correct BAC, as identified by BACSearcher, were recovered from the BAC library and processed in parallel using 96-format culturing plates. To optimize the efficiency of our high-throughput pipeline, we successively ran the pipeline three times, removing successful targets once confirmed. A detailed protocol for the optimized high-throughput pipeline is provided in Supplemental Protocol 1B. In summary, 100% of our 191 target BAC lines were made recombination competent (Figure 2A) and out of the 191 target genes, one gene-specific tagging cassette failed to amplify (Figure 2B), likely due to the formation of secondary structure(s) within the 50-bp homology regions of the primers. Of the 190 that amplified successfully, 187 yielded colonies after selection with kanamycin (Figure 2C). Validation by enzymatic digestion confirmed that 146 of these lines contained correct recombineering plasmid products (Figure 2D). We extracted the recombineering plasmid products from the 146 successful lines and confirmed their junctions by Sanger sequencing. Our high-throughput pipeline had an overall efficiency of 76%, an average recombineered region length of 7,259 bp, and a maximum cloned length of 22,773 bp corresponding to gene Cre10.g427850 (Supplemental Data Set 2). Twenty-six target genes that were unsuccessfully cloned by PCR in 2017 were successfully cloned here by recombineering, and all five previously successful targets containing Class 1 uORFs retried here were successfully cloned.

During pipeline development, we found that optimizing bacterial growth prior to transformation with the recombineering cassette was critical (see notes in Supplemental Protocol 1). In addition, for 14 out of the 146 correctly recombineered lines in our high-throughput pipeline, the use of an alternative BAC from the library yielded success after an initial failure. For approximately half of the target genes, it was necessary to validate multiple colonies by enzymatic digest to rule out false positives; beginning with the 187 colony-producing lines from our high-throughput pipeline, picking just a single colony gave a 49% success rate, screening a second colony increased the success rate to 66%, and screening a third colony gave a 76% success rate. For a small proportion of targets, screening >3 colonies led to the identification of a correctly recombineered construct (Figure 2E). Restriction digest analysis of plasmids isolated from incorrectly assembled recombineering events suggested that cloning could fail for a broad range of reasons, including cassette recircularization, cassette duplication, cassette insertion into the BAC, or retrieval of incorrect target regions. Increasing homology arm length, using alternative homology arms, using alternative BACs, and using fosmids are potential solutions to overcome incorrect recombineering for specific targets. Supplemental Data Set 1 provides up to five options for homology arms and up to five available BACs per gene, and BACSearcher can be easily modified to increase homology arm length (see Supplemental Protocol 2A). Taken together with our 12 initial targets, we successfully cloned 157 out of 203 target regions from BACs using our recombineering pipeline, achieving an efficiency of 77%.

Cloning success is size independent and tolerant of sequence complexity

To investigate if our recombineering approach was gene size and complexity independent, we compared our successful targets against unsuccessful targets (Figure 3). Here we defined a target region to mean the ATG-stop ORF for PCR-based cloning and the ATG-stop ORF plus an upstream region of $>2{,}000$ bp designed to encompass the 5′-UTR and native promoter for recombineering. There was no significant difference in the lengths of these regions between cloned and uncloned targets for recombineering (Figure 3A; Mann–Whitney $U = 3{,}303$, $P = 0.38$, two-tailed), indicating that our method is target size independent. This contrasts to the clear effect of target size on cloning success for our previous PCR-based cloning data (Figure 3A; Mackinder et al., 2017). We then compared our cloning success to the number of simple and global repeats per kilobase in target regions. Our method appears to be far more tolerant of repetitive sequences than PCR-based cloning in terms of both the per-kilobase prevalence of simple and global repeats and the number of repeats per target region (Figure 3B and C). For our recombineering pipeline, there was no significant difference in the average number of repeats per kilobase between cloned and uncloned regions (Mann–Whitney $U = 3{,}129$, $P = 0.17$, two-tailed), whereas PCR-based cloning success was significantly reduced for targets with over ∼4.8 repeats per kbp (Figure 3B). For the most repetitive targets involved in our analysis (>9 repeats per kbp), our recombineering cloning efficiency remained above 60%, a rate over three times higher than that of PCR-based cloning (Figure 3B). Extrapolation of these data overlaid with the genome-wide distribution of repeat frequencies indicates that a large proportion of genes that are technically challenging for PCR-based cloning due to high repeat frequencies could be cloned by recombineering (Figure 3B).

Localization of Venus-tagged proteins

To assess the validity of the pipeline for localization studies, we transformed wild-type Chlamydomonas cells with a subset of linearized recombineering plasmid products tagged at the C-terminus with *CrVenus* (Figure 4A). Paromomycin-resistant colonies were directly screened for YFP fluorescence on transformation plates, picked, grown in Tris–phosphate (TP) minimal medium at air-levels of CO_2 (∼0.04%), imaged by fluorescence microscopy to examine the localization pattern (Figure 4B; Supplemental Figure 2), and immunoblotted against the C-terminal 3xFLAG epitope to confirm fusion protein size (Supplemental Figure 2A). We selected the genes for transformation based on previous affinity purification mass spectrometry data (Mackinder et al., 2017) and pyrenoid proteomics data (Mackinder et al., 2016; Zhan et al., 2018). The localization data support the proteomics data as we detected Photosystem I subunit F (PSAF), ISA1 (Isoamylase 1) and Chloroplast Stem-loop Binding Protein of 41 kDa B (CSP41B) in the pyrenoid. PSAF is a core transmembrane subunit of photosystem I. As expected, PSAF showed strong colocalization with chlorophyll outside

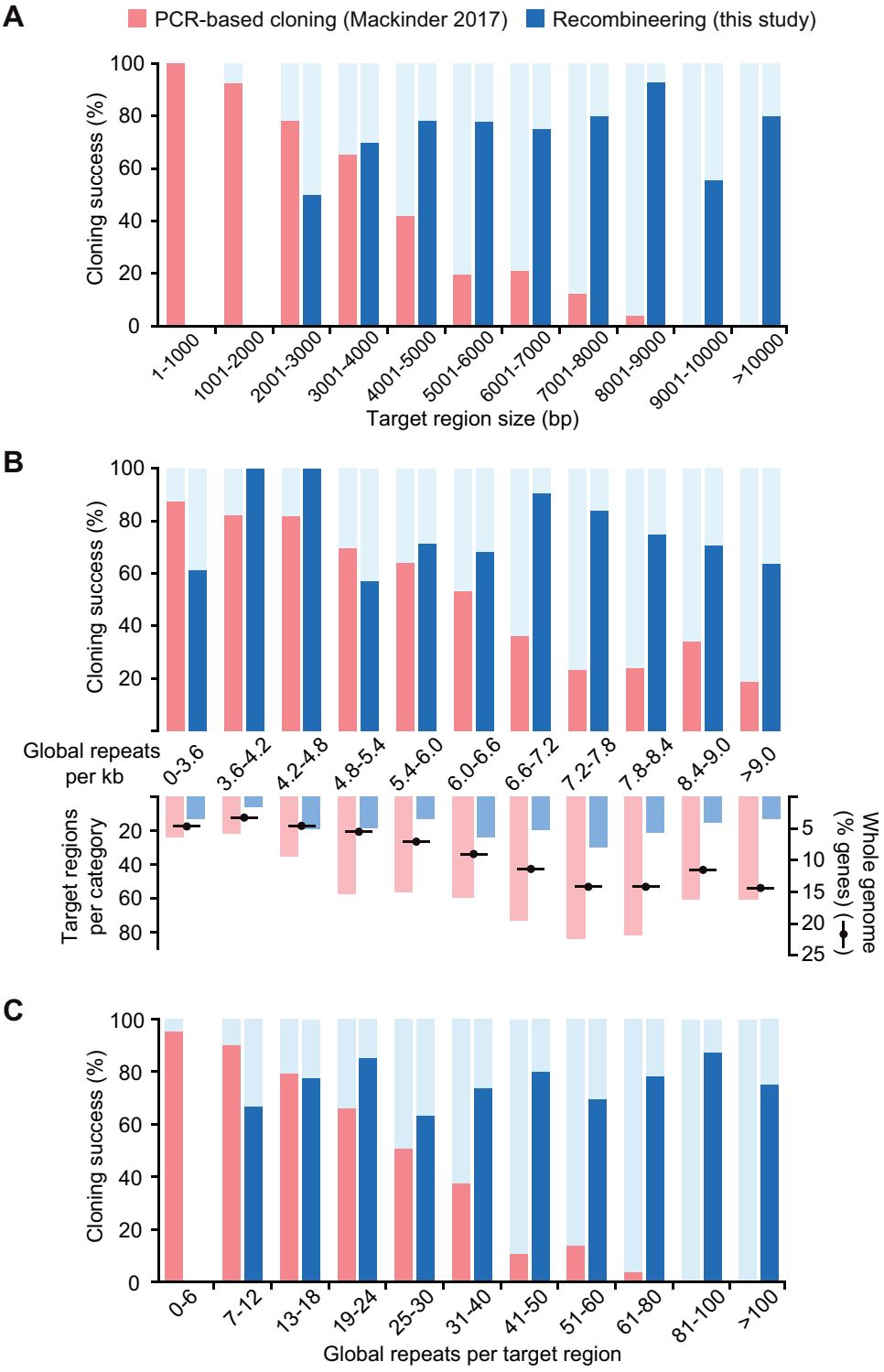


Figure 3 Our recombineering pipeline is target gene size independent and tolerant of sequence complexity. A, The size distribution of successfully PCR-cloned coding sequences (Mackinder et al., 2017; red) or recombineered regions (this study; blue) are shown. Regions cloned by recombineering include ~2 kbp of flanking DNA upstream of the annotated start codon to incorporate native 5′-promoter sequences. A severe drop in PCR-based cloning efficiency can be seen for templates >3-kbp long, whereas recombineering cloning efficiency does not show size dependency. No recombineering target regions were less than 2,000-bp long due to inclusion of native 5′-promoter sequences. B, As above but showing the dependence of cloning success on the per-kilobase frequency of repeats masked by the NCBI WindowMasker program with default settings (Morgulis et al., 2006). The number of target regions per repeat category is shown beneath this, overlaid with the percentage of Chlamydomonas genes in each category. The distribution of targets for this study and our previous PCR-based cloning attempt (Mackinder et al., 2017) gives a reasonably close representation of the whole-genome distribution. Almost a third of nuclear genes contain 7.2–8.4 repeats per kbp; this peak corresponds to a clear drop in PCR-based cloning efficiency, but to a high recombineering efficiency of 75–85%. Data for repeats per kbp were continuous and

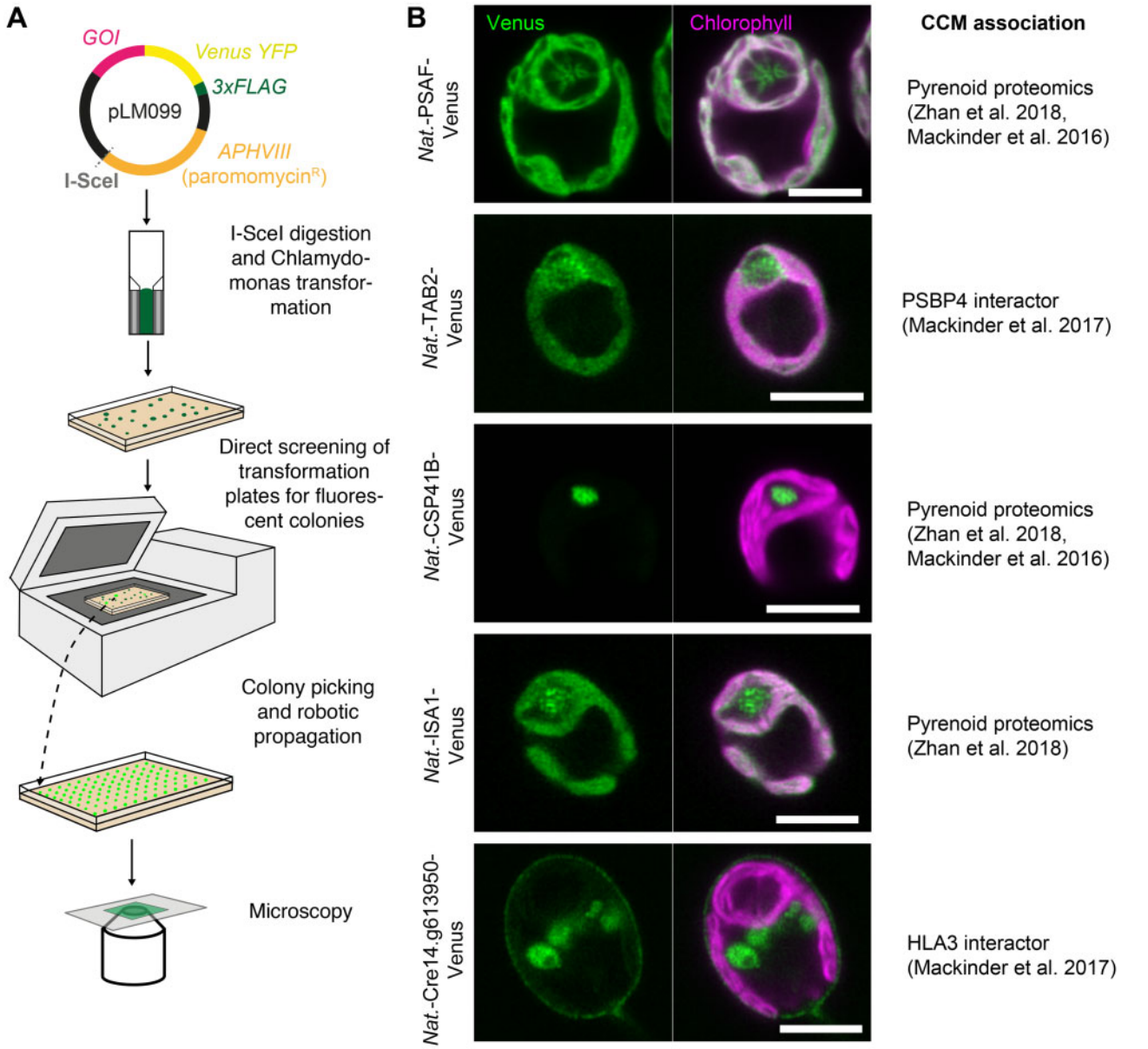


Figure 4 Transformation and localization of a subset of recombineered targets. A, Chlamydomonas transformation pipeline. The I-SceI cut site allows vector linearization prior to Chlamydomonas transformation via electroporation. Transformants are directly screened for fluorescence using a Typhoon scanner (GE Healthcare, San Diego, CA, USA) and then picked and propagated prior to imaging. GOI: gene of interest. B, The localizations of a subset of the recombineered target genes. Localizations agree with data from an affinity-purification followed by mass spectrometry study (Mackinder et al., 2017) or pyrenoid proteomics (Zhan et al., 2018 and/or Mackinder et al., 2016). Scale bars: 5 μm.

of the pyrenoid; however, in addition, it clearly localized to the thylakoid tubules traversing the pyrenoid. Interestingly, in the pyrenoid tubules, the chlorophyll signal was minimal, particularly at the "pyrenoid tubule knot," where the tubules converge (Engel et al., 2015). These data, along with the finding that other PSI and PSII components localized to the pyrenoid tubules (Mackinder et al., 2017), suggest that the tubules contain both PSI and PSII but that chlorophyll-

Figure 3 (continued)

there are no values present in more than one category. C, As above but showing the number of simple and global repeats masked by WindowMasker per template. Data are binned to provide a higher resolution for the lower value categories, since the targets for PCR-based cloning were enriched in targets with low numbers of repeats. As in (A), a severe negative trend in PCR-based cloning efficiency can be seen, reflecting a strong positive correlation between repeat number and region size. No negative association is present for recombineering cloning efficiency, likely illustrating the benefit of avoiding size- and complexity-associated polymerase limitations. No recombineering target regions contained fewer than six repeats.

containing light harvesting complexes found within the pyrenoid may be quenched or at low abundance. Tagged TAB2, a protein linked to early PSI assembly (Dauvillée et al., 2003) that was identified as an interactor with PSBP4 (a photosystem II subunit P-like protein) found within and at the periphery of the pyrenoid (Mackinder et al., 2017), was also enriched at the pyrenoid. Interestingly, the location of TAB2 was not just restricted to the pyrenoid periphery, but it was also found within the pyrenoid, forming distinct small foci (Figure 4B). This finding suggests that early PSI assembly could be occurring within the pyrenoid as well as at the pyrenoid periphery (Uniacke and Zerges, 2009).

CSP41B localized to the pyrenoid matrix, and analysis of the translated product of CSP41B showed that it belongs to a family of NAD-dependent epimerase/dehydratases (IPR001509) and contains a UDP-galactose 4-epimerase domain that may be involved in galactose metabolism. Its role in pyrenoid function is unclear. Localization of ISA1 showed that it was enriched in the pyrenoid, with an uneven distribution. ISA1 is a starch-debranching enzyme that is essential for starch synthesis, as *ISA1* deletion lines lack both chloroplast and pyrenoid starch (Mouille et al., 1996). The presence of pyrenoid starch and its correct organization are critical for correct CCM function (Itakura et al., 2019; Toyokawa et al., 2020), as the absence of starch in an *ISA1* knockout mutant (4-D1) led to incorrect localization of the low-CO_2-inducible (LCI) protein LCIB (see below), retarded growth at very low CO_2 levels (0.01% v/v), and reduced inorganic carbon affinity (Toyokawa et al., 2020). Interestingly, Toyokawa et al. (2020) failed to obtain localization data for an ISA1-mCherry fusion driven by the *HSP70A/RBCS2* hybrid promoter.

Cre14.g613950 encodes a protein belonging to the ABC transporter family that was identified as an interactor of HLA3 (high light-activated gene 3; Mackinder et al., 2017), a putative HCO_3^- transporter located in the plasma membrane (Duanmu et al., 2009; Gao et al., 2015). Like HLA3, Cre14.g613950 shows a typical plasma membrane localization pattern, with YFP signal at the cell periphery and signal typical of the Golgi network. However, immunoblotting against the C-terminal 3xFLAG tag of Cre14.g613950 in two independent transformants showed a smaller molecular weight band than predicted (Supplemental Figure 2). This potentially indicates that the gene model for Cre14.g613950 is incorrect or that the protein undergoes post-translational cleavage, as seen for other CCM-related proteins that transit via the secretory pathway (Fukuzawa et al., 1990; Tachiki et al., 1992).

Development of backbones with additional tags and markers

To further expand the functional application of our recombineering pipeline, we designed additional backbone vectors that enable protein tagging with the fluorophores mScarlet-i (Bindels et al., 2017), mNeonGreen (Shaner et al., 2013), and mTurquoise2 (Goedhart et al., 2012) and that allow selection with hygromycin or zeocin (Figure 5, A and B). These vectors can be used to complement Chlamydomonas Library Project (CLiP) mutants that have been generated using the *AphVIII* marker conferring paromomycin resistance (Li et al., 2016, 2019) and to express two or three differently tagged proteins within the same cell. For comparison, we tested these vectors on *LCI9* (Cre09.g394473), which encodes the low-CO_2 inducible protein LCI9 that, via PCR-based cloning, we previously showed localizes to the pyrenoid periphery (Mackinder et al., 2017). Recombineered *LCI9* was 7,160-bp long, including the native promoter region. All fluorophores displayed the same pyrenoid periphery localization pattern (Figure 5C), which agrees with the localization information obtained when LCI9 expression was driven by the *PSAD* promoter (Figure 5C bottom image; the *PSAD* promoter is here defined as the sequence spanning from 3 to 763-bp upstream of the *PSAD* start codon [Cre05.g238332], encompassing both the 5′-UTR and promoter region). These results further support the use of approximately 2,000-bp upstream regions as promoters for fusion protein expression.

To further confirm that the localizations of proteins driven by their native promoters do not differ from those driven by the constitutive *PSAD* promoter, we compared the localizations of *native*-LCIB-Venus and *PSAD*-LCIB-Venus. LCIB is an essential CCM component that shows dynamic relocalization to the pyrenoid periphery at CO_2 levels $<0.04\%$ (Yamano et al., 2010). LCIB expressed from its endogenous promoter localized to the pyrenoid periphery under ambient CO_2 levels (~0.04% v/v), in full agreement with the localization data when LCIB expression was driven by the constitutive *PSAD* promoter (Figure 5D).

Finally, we tested whether our recombineering pipeline could be used to successfully complement a CLiP mutant. We transformed *native-LCIB-Venus* (cloned into pLM161, containing the *APHVII* gene conferring hygromycin resistance) into a CLiP *lcib* mutant (LMJ.RY0402.215132) and examined the phenotypes of four transformants showing Venus fluorescence. All four transformants showed a typical pyrenoid peripheral localization pattern when grown at very low CO_2 levels, and all rescued the *lcib* mutant phenotype to varying degrees, with *lcib:LCIB-Venus-1* showing complete rescue (Supplemental Figure 3).

Maintaining the native promoter enables relative protein abundances to be monitored

As our pipeline retains the native promoter of the target gene, we hypothesized that fluorescence output would be representative of relative changes in protein abundance in response to environmental conditions. To test this, we grew lines with *LCIB* driven by either the constitutive *PSAD* promoter (*PSAD*-LCIB-Venus) or its native promoter (*Native*-LCIB-Venus). LCIB-Venus signal stayed relatively constant between high (3% v/v) and low (0.04% v/v) CO_2 conditions when LCIB was expressed from the *PSAD* promoter (*PSAD*-LCIB-Venus), but showed an approximately eight-fold

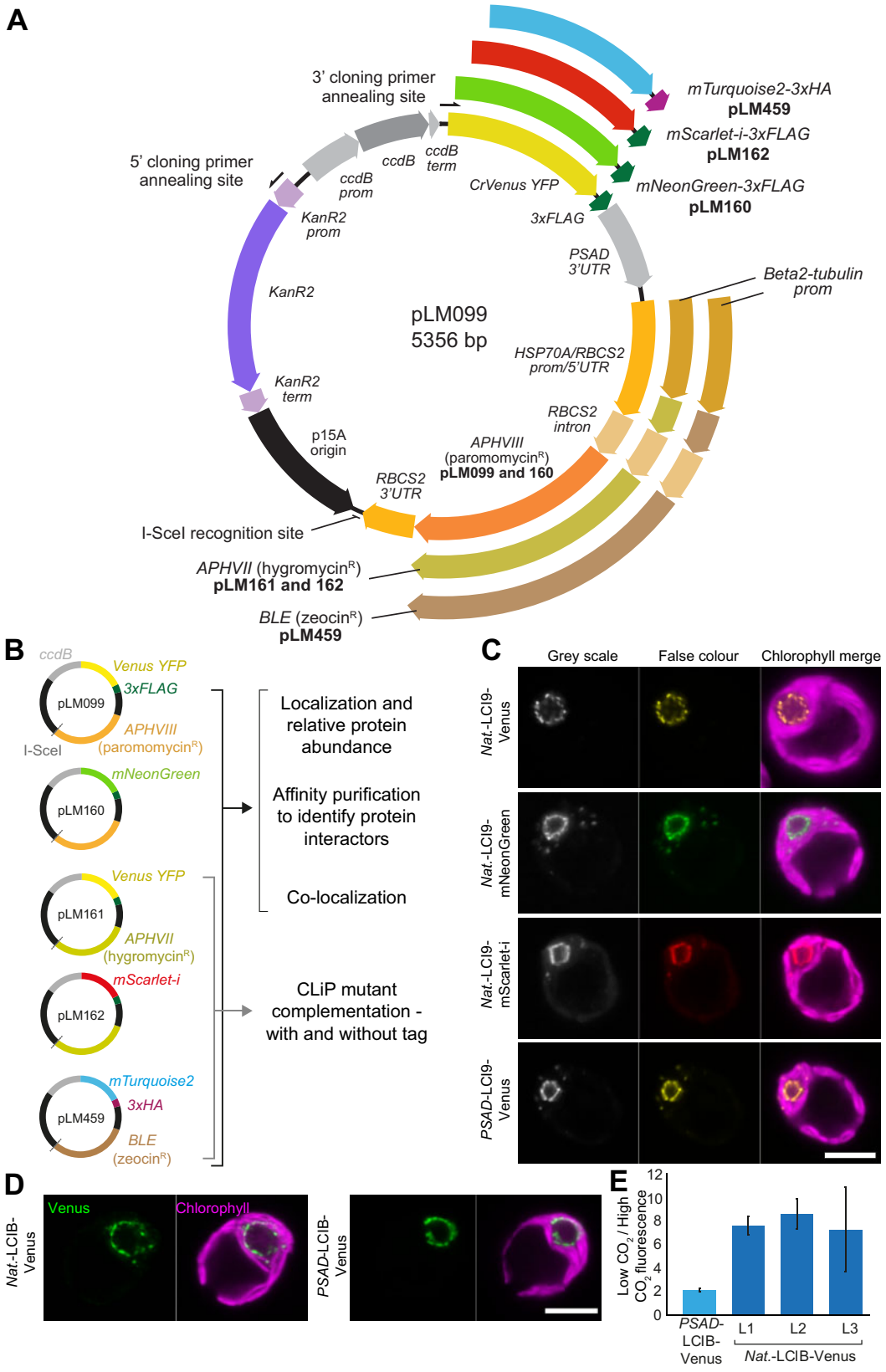


Figure 5 Development and application of different recombineering vectors to obtain novel biological insights into Chlamydomonas biology. A, Plasmid map for pLM099 and derivative recombineering vectors. PCR amplification with 5′- and 3′-cloning primers at the annealing sites shown results in a ~4.6-kbp linear cassette for recombineering target genes in-frame with a fluorescent protein and affinity tag. For each recombineering vector, the fluorescent protein sequence is preceded by a flexible linker (GGLGGSGGR) and followed by a tri-glycine linker prior to the affinity tag. The *PSAD* 3′-UTR terminates all four fluorescent protein-affinity tag cassettes. The *RBCS2* 3′-UTR terminates all three Chlamydomonas

increase between these conditions when the native promoter was used, with this change consistent across three independently transformed lines (Figure 5E). This finding agrees with previous immunoblotting data, in which a comparable fold increase was seen in LCIB abundance when cells were transferred from high CO_2 to low CO_2 conditions (Yamano et al., 2010). This indicates that our recombineering lines can be used to monitor relative protein abundance across different growth conditions.

DISCUSSION

We have established a rapid recombineering-based method to clone large and complex Chlamydomonas genes from BACs. Our approach circumvents the challenges associated with cloning large, GC-rich, and complex genes that are prevalent in Chlamydomonas. We demonstrated that the method could be applied for small batch cloning as well as 96-well high-throughput cloning. Our overall cloning success rate (combined batch and high-throughput results) was 77%, a value considerably higher than that of our previous PCR-based high-throughput cloning pipeline (48%), which was inflated due to an enrichment of small target genes. Our overall success rate is slightly lower compared to recombineering pipelines in other organisms, with success rates of 89% achieved in *C. elegans* (Sarov et al., 2012) and ~93% in Arabidopsis (Brumos et al., 2020). This reduced overall efficiency is likely due to the complexity of the Chlamydomonas genome (Figure 1), as DNA secondary structure was previously linked to recombineering failure (Nelms and Labosky, 2011). We expect a higher success rate when the pipeline is applied to a smaller number of samples, since it is easier to optimize bacterial growth prior to electrotransformation on a per-sample basis if there are fewer samples to manage. This may be evidenced by our successful cloning of 11 out of 12 targets in an initial batch-scale pipeline attempt (Supplemental Figure 1), although the sample size is insufficient to generalize with confidence.

To enable expression of multiple fluorophores simultaneously and for the complementation of CLiP mutants, we designed a series of vectors with modern fluorophores and varying selection markers and demonstrated their performance in Chlamydomonas (Figure 5). The presence of either 3xFLAG or 3xHA tag allows these vectors to be used for affinity purification to explore interacting partners of tagged proteins. Different fluorophore pairs (i.e. mNeonGreen and mScarlet-i) could also be used for FRET-based studies to explore protein–protein interactions. In addition, all vectors can be used to clone genes without fluorescence tags or with only short affinity tags (3xFLAG and 3xHA).

Due to the size independence of our method, we could maintain the native promoters of target genes. For two genes, *LCI9* and *LCIB*, there were no noticeable differences in protein localization between native promoter-driven expression and *PSAD* promoter-driven expression. Interestingly, using a native promoter allows relative protein abundance to be tracked between conditions (Figure 5E). Once validated, acquiring relative abundance data is straightforward and can be easily parallelized. This allows relative protein abundance to be tracked in real-time across a broad range of conditions. Future experiments could include tracking relative protein abundance in 96-well libraries of tagged proteins in response to a perturbation (i.e. high to low CO_2 transition). This would be highly supportive of available transcriptomic and proteomic data sets and would provide novel insights into cellular processes (Mettler et al., 2014; Zones et al., 2015; Strenkert et al., 2019). Although our relative abundance data for LCIB appear to closely reflect the immunoblotting data, it should be noted that using a native promoter may not always fully reflect native changes. This discrepancy could be due to insertional effects caused by integration into transcriptionally unfavorable regions of the genome and the absence of *cis*-regulatory regions in the recombineered construct or by transcriptional silencing (Schroda, 2019). At the protein level, fluorescent protein folding time could affect protein stability and turnover, and the presence of the fused fluorescence protein could affect function or multi-subunit assembly.

While our approach allows the native promoter, 5′-UTR, and ORF to be cloned, the native 3′-UTR is not maintained. This could be addressed through a two-step recombineering pipeline where the tag is first inserted into the BAC at the desired location and the markers are then removed via a Flp-*FRT* recombinase system (Sarov et al., 2006; Brumos et al., 2020). The edited target gene could then be retrieved into a final Chlamydomonas expression vector. When establishing our pipeline, we decided not to pursue this strategy to maximize the success rate by limiting the number of steps, with a focus on developing a simple, easy to apply approach. In addition, while we have focused on C-terminal tagging, as this allows N-terminal transit peptides required

Figure 5 (continued)

selection cassettes. The same *RBCS2* intron is present in all three Chlamydomonas selection cassettes but is only inter-exonic in the hygromycin and zeocin-resistance cassettes. B Additional vectors for tagging with different fluorophores and for complementation of Chlamydomonas library mutants generated using insertion of the *AphVIII* paromomycin resistant gene. C, Localization of LCI9 with different fluorescence protein tags. *LCI9* was recombineered with its native promoter (*Nat.*) using pLM099, pLM160, and pLM162. A previously developed line cloned by PCR and using the constitutively expressed *PSAD* promoter is shown for comparison (*PSAD*-LCI9-Venus). Scale bar: 5 µm. D, A comparison of the low CO_2 upregulated gene *LCIB* cloned with its native promoter via recombineering versus *LCIB* under the control of the constitutive *PSAD* promoter. Cells were grown and imaged at atmospheric CO_2 levels. Scale bar: 5 µm. E, Relative change in LCIB-Venus fluorescence between high (3% v/v) and low (0.04% v/v) CO_2 when expressed from the constitutive *PSAD* promoter versus expression from the native *LCIB* promoter. Data are shown for three independent native *LCIB* promoter lines (L1–L3). Error bars are standard error of the mean.

for organelle targeting to be conserved, our recombineering pipeline could be applied for N-terminal tagging by modifying our cloning vectors with a constitutive promoter and N-terminal tag.

The simplicity of our framework and vector design could be adopted for other organisms with relative ease provided a BAC or fosmid library and efficient transformation protocols are available. Multiple features of our recombineering cassette could make adaptation to different organisms relatively straightforward, such as the use of ccdB counter-selection and the rare I-SceI recognition site used for linearization of the recombineering cassette prior to transformation. For organisms in which selection with paromomycin, hygromycin, or zeocin is ineffective or the *AphVII*, *AphVIII*, or *BLE* genes included in the pLM099-derived cassettes cannot be utilized, alternative selection genes can be quickly incorporated by restriction-ligation using flanks containing KpnI and I-SceI recognition sites at the 5′- and 3′-ends, respectively.

One limitation we encountered was that only 86% of nuclear genes are covered by the BAC library. However, this value only takes into account ∼73% of BACs, with the remaining BACs potentially incorrectly mapped to the current version of the Chlamydomonas genome (see Supplemental Protocol 2B). Our analysis suggests that the true percentage of genes covered could be higher than 86%, but confirming this may require a careful re-mapping of the BAC library. A promising solution is cloning from fosmids. We demonstrated that our pipeline can be successfully used for cloning from fosmids, and a Chlamydomonas fosmid library is now available (released July 2020; Chlamydomonas Resource Center). The use of fosmids, with smaller DNA fragments compared to BACs, could help improve the efficiency of our pipeline by reducing off-target recombination between the PCR-amplified cassette and the BAC or by reducing recombination between two repetitive regions of the BAC. In addition, the fosmid library is expected to have close to 100% genome coverage.

Our recombineering approach allowed us to efficiently clone large and complex genes, which could not be achieved via PCR-based cloning. This method opens the door to a better understanding of the functional roles of a large fraction of the Chlamydomonas genome through protein localization, protein–protein interaction studies, real-time monitoring of relative protein abundance, and complementation of mutants (e.g. random insertion and CRISPR/Cas-generated mutants). In addition, it provides a highly complementary method to the recently released CLiP mutant collection.

Materials and methods

Plasmid and cassette construction

Fragments for pLM099 were amplified by PCR (Phusion Hotstart II polymerase, ThermoFisher Scientific, Waltham, MA, USA) from the following plasmids: Venus-3xFLAG, PSAD terminator and *AphVIII* from pLM005 (Mackinder et al., 2017); the p15A origin of replication from pNPC2; the Kan^R resistance gene from pLM007; the counter-selection *ccdB* gene from Gateway pDONR221 Vector (ThermoFisher Scientific). The resulting amplicons were gel purified (MinElute Gel Extraction Kit, Qiagen, Valencia, CA, USA) and assembled by Gibson assembly (see Figure 5A for detailed map). pLM160 was constructed from pLM099 to replace CrVenus with mNeonGreen (Shaner et al., 2013), and pLM161 was constructed from pLM099 to replace the paromomycin resistance gene (*AphVIII*) with the hygromycin resistance gene (*AphVII*). pLM162 was constructed from pLM161 with the synthetic fluorophore mScarlet-i (Bindels et al., 2017) replacing CrVenus. pLM459 was constructed from pLM161 to replace CrVenus with mTurquoise2 (Goedhart et al., 2012), the 3xFLAG with the 3xHA haemagglutinin tag, and *AphVII* with the zeocin resistance gene (*Sh ble*). Gene-specific cloning primers were designed to amplify a ∼4.6 kbp cassette from the recombineering vectors pLM099, 160, 161, 162, and 459 (Figure 5), excluding *ccdB*, and providing 50 bp of sequence homology to the target gene an average of ∼2,500-bp upstream of the 5′-UTR and directly upstream of the stop codon. This enables the retrieval of each target gene into the cassette in frame with a fluorescent tag and with the native promoter region intact. All oligonucleotide and plasmid sequences can be found in Supplemental Data Sets 3 and 4.

Culturing

Escherichia coli cells were cultured in lysogeny broth (LB) or yeast extract nutrient broth (YENB) at 37°C unless they contained the temperature-sensitive pSC101-BAD-gbaA-tet (pRed), in which case 30°C was used. All DNA for transformation was introduced by electroporation and transformants were recovered in super optimal broth with catabolite repression (SOC). DH10B cells containing fragments of the Chlamydomonas genome in the form of BACs were obtained from the Clemson University Genomics Institute (now distributed by the Chlamydomonas Resource Center, University of Minnesota, Minnesota, MN, USA). DB3.1 cells expressing the *ccdB* antidote gene *ccdA* were obtained from ThermoFisher Scientific and used for maintenance of the recombineering vectors.

Chlamydomonas wild-type cells (strain CC-4533) were cultured in Tris–acetate–phosphate (TAP) medium with revised Hutner's trace elements (Kropat et al., 2011) and illuminated by white fluorescent light. Assembled recombineering vectors were prepared for transformation into Chlamydomonas by restriction digest with I-SceI endonuclease (NEB). Transformation and selection of fluorescence lines were performed in accordance with Mackinder et al. (2017) using a Typhoon Trio fluorescence scanner (GE Healthcare, Madison, WI, USA). Viable Chlamydomonas transformants were screened for CrVenus and mNeonGreen fluorescence emission at 555/20 nm, and for mScarlet-i at 615/12 nm. Several strains emitting the strongest fluorescence for each line were picked. The average number of fluorescent colonies for recombineered Venus fusion proteins with their

native promoter was ~10%; however, this varied considerably between constructs (PSAF (10/134) 7%, TAB2 (6/44) 13.6%, CSP41B (6/43) 13.9%, ISA1 (25/297) 8%, Cre14.g613950 (2/22) 9%, LCI9 (6/25) 24%, and LCIB (6/19) 31.5%). Picked fluorescent strains were cultured in Tris–phosphate (TP) minimal medium under ambient CO_2 (~0.04%) conditions and imaged by fluorescent microscopy to visualize protein localization. To ensure that the determined localizations were not due to in-frame integration of a fluorophore-containing fragment of the cassette with another gene, we confirmed localization in at least two independent transformants and performed immunoblotting against the 3xFLAG epitope to confirm the expected fusion protein size.

For spot tests, cells were grown to ~8×10^6 cells mL^{-1} in TAP at ~50 μmol photons m^{-2} s^{-1}, washed with TP, and serial diluted in TP prior to spotting 1,000, 100, and 10 cells on TP 1.5% agar plates. Replica plates were incubated in 0.04% or 3% CO_2 chambers for 24 h at 50 μmol photons $m^{-2}s^{-1}$, then 24 h at 150 μmol photons m^{-2} s^{-1}, followed by 48 h at 300 μmol photons m^{-2} s^{-1} prior to imaging.

Protein extraction and immunoblotting

Lines expressing recombineered fusion proteins were cultured in 50 mL TAP medium containing 5 μg mL^{-1} paromomycin to a cell density of ~2×10^6 cells mL^{-1}. Cells were harvested by centrifugation at 5,000*g* for 10 min at room temperature. The supernatant was discarded, and the pellet was resuspended in 500 μL of protein extraction buffer (20 mM Tris–HCl pH 7.5, 5 mM $MgCl_2$, 300 mM NaCl, 5 mM Dithiothreitol [DTT], 0.1% Triton-X100, Roche protease inhibitor) and flash frozen in liquid nitrogen in 100 μL aliquots. Cells were thawed on ice and flash frozen again before a final thaw on ice. Samples were then centrifuged at 17,000*g* for 15 min at 4°C to separate the soluble and insoluble fractions. The soluble supernatant was transferred to a new tube, mixed 1:1 with 2× Laemmli buffer containing β-mercaptoethanol, and heated at 80°C for 10 min prior to sodium dodecyl sulfate–polyacrylamide gel electrophoresis (SDS–PAGE).

About 15–30 μL of each sample was loaded onto a 10% mini-protean TGX gel (Bio-Rad) and transferred to a polyvinylidene difluoride membrane via semi-dry transfer (10V, 60 min). Fusion proteins were immuno-detected using the monoclonal anti-flag M2 antibody (1:1,000; Sigma-Aldrich, St Louis, MO, USA; catalog # F1804), followed by Alexa-Fluor 555 goat anti-mouse secondary antibody (1:10,000; Invitrogen, Carlsbad, CA, USA; catalog # A-21422). The membrane was imaged using a Typhoon 5 Scanner.

Microscopy

Sample preparation for microscopy was performed as per (Mackinder et al., 2017). Images were acquired using a Zeiss LSM880 confocal microscope on an Axio Observer Z1 invert, equipped with a 63× 1.40 NA oil planapochromat lens. Images were analyzed using ZEN 2.1 software (Zeiss, San Diego, CA, USA) and FIJI. Excitation and emission filter settings were as follows: Venus and mNeonGreen, 514 nm excitation, 525–550 nm emission; mScarlet-i, 561 nm excitation, 580–600 nm emission; and chlorophyll, 561 nm excitation, 665–705 nm emission.

Plate reader assay

To monitor changes in fluorescence in response to CO_2, three independent *native*-LCIB-Venus lines, a single *PSAD*-LCIB-Venus line, and wild-type (WT) were grown in TP bubbled at low CO_2 (0.04% v/v) or high CO_2 (3% v/v) conditions at 300 μmol photons m^{-2} s^{-1}. Four samples per line were aliquoted into a 96-well plate and chlorophyll (excitation 625/34, emission 692/50) and Venus (excitation 504/10, emission 540/12) fluorescence was immediately measured using a BMG Labtech Clariostar Plate Reader. Venus fluorescence was normalized by chlorophyll, and WT background was then subtracted. The average fluorescence value under low CO_2 conditions was divided by the average fluorescence value under high CO_2 conditions for each line. Error was calculated by the propagation of variance across both low and high CO_2 values and is shown as the standard error of the mean.

Recombineering procedure for one-step subcloning and tagging

The following outlines the batch-scale recombineering protocol. Extended batch and multi-well plate-scale recombineering protocols are provided in Supplemental Protocol 1.

For each target, a recombineering cassette was amplified from plasmid pLM099 (Phusion Hotstart II polymerase, ThermoFisher Scientific, Waltham, MA, USA) using primers containing 50-bp homology arms, one homologous to a region upstream of the annotated start codon of the target gene, and one homologous to the 3′-end of the coding sequence (excluding the stop codon). The resulting PCR product was gel purified (MinElute Gel Extraction Kit, Qiagen, Valencia, CA, USA) and its concentration measured using a NanoDrop spectrophotometer. Upstream region lengths ranged from 1,000 to 4,000 bp from the start codon, with an average of ~2,500 bp. For two genes, Cre04.g220200 and Cre16.g678661, the first 50 bp of the 5′-UTR was used as the upstream homology region due to BAC coverage limitations.

The pRed plasmid pSC101-BAD-gbaA-tet was extracted from *E. coli* cells grown overnight at 30°C (Plasmid Mini Kit, Qiagen, Valencia, CA, USA) and its concentration measured by NanoDrop. *Escherichia coli* cells harboring a BAC containing the target gene were recovered from the Chlamydomonas BAC library and used to inoculate 20 mL of YENB medium containing 12.5 μg mL^{-1} chloramphenicol, followed by overnight growth in a 50 mL conical flask at 37°C with vigorous shaking. After 16 h of growth, 120 μL of the culture was used to inoculate 4 mL of fresh YENB containing 12.5 μg mL^{-1} chloramphenicol. This was grown for ~2 h at 37°C until an optical density (OD_{600}) of 2 was reached. About 2 mL of the culture was then incubated on ice for 2 min, followed by centrifugation at 5,000*g* for 10 min at 4°C. After removing the supernatant, the pellet was

placed back on ice and washed by resuspension in 1 mL of chilled 10% glycerol, followed immediately by centrifugation at 5,000*g* for 10 min at 4°C. The resulting supernatant was removed, and the pellet was placed back on ice and resuspended in 100 μL of 0.1 ng μL^{-1} pRed. This mixture was transferred to a pre-chilled 2 mm gap electroporation cuvette and electroporated at 2,500 V, 400 Ω, and 25 μF using a Gene Pulser II (Bio-Rad, San Diego, CA, USA). The electroporated cells were immediately recovered in 800 μL SOC and incubated at 30°C for 90 min with vigorous shaking. The whole outgrowth was added to 20 mL of YENB containing 12.5 μg mL^{-1} chloramphenicol and 5 μg mL^{-1} tetracycline and grown overnight at 30°C with vigorous shaking.

After 16 h of growth, 600 μL of culture was used to inoculate 4 mL of fresh YENB containing 12.5 μg mL^{-1} chloramphenicol and 5 μg mL^{-1} tetracycline. This was grown for 3 h at 30°C, or until reaching an OD_{600} >2, at which point 80 μL of 10% L-arabinose was added to induce pRed expression and growth was shifted to 37°C for 1 h with vigorous shaking. About 2 mL of the induced culture was incubated on ice for 2 min, then centrifuged at 5,000*g* for 10 min at 4°C, the supernatant removed, and the pellet placed back on ice. Cells were then washed in 10% glycerol, centrifuged at 5,000*g* for 10 min at 4°C, the supernatant removed, and the pellet placed back on ice. The pellet was resuspended in 100 μL of 5 ng μL^{-1} PCR product and transferred to a pre-chilled 2-mm gap electroporation cuvette, followed by electroporation as before. Electroporated cells were immediately added to 800 μL of SOC and recovered at 37°C for 90 min with vigorous shaking. About 450 μL of outgrowth was spread onto 1.5% LB-agar containing 25 μg mL^{-1} kanamycin, air-dried, and incubated overnight at 37°C. Selected colonies were used to inoculate 4 mL of LB containing 25 μg mL^{-1} kanamycin and grown for 16–18 h at 37°C with shaking. Recombineering products were extracted and validated by restriction digest using the appropriate enzymes, followed by Sanger sequencing using primers designed to amplify the junctions between the pLM099-derived cassette and the target region.

Statistics

Confidence intervals for Figure 1A were calculated using the Wilson score interval method based on the number of attempted and successfully cloned ATG-Stop amplicons per size category in Mackinder et al. (2017). Statistical differences in the distribution of sizes and repeat frequencies between successful and unsuccessful PCR and recombineering targets (presented in Figure 3) were assessed using the Mann–Whitney *U* test. A non-parametric test was chosen based on results of the Kolmogorov–Smirnov test for normality for recombineering targets. Test statistics are detailed in Supplemental Table 1.

Genome analysis

Chlamydomonas, Arabidopsis, yeast, and wheat nuclear genes were analyzed for gene size and sequence complexity. Gene sizes are defined from the start of the 5′-UTR to the end of the 3′-UTR. Note that in Figure 1A, the predicted clonable proportion of genes in each size category is based on cloning success for ATG-Stop regions, not full genes. Sequence complexity is defined in relation to intron prevalence, GC content, and the prevalence of various repeat regions. We designate regions containing a high frequency of repeats as being more complex than regions with a low frequency. This reflects the increased potential for cloning complications presented by sequences with large numbers of repetitive regions, though it differs from descriptions given by Morgulis et al. (2006). Sequences were analyzed for complexity using the freely available bioinformatics software detailed below (see Supplemental Protocol 3 for settings), and outputs were processed using custom python scripts (see Supplemental Protocol 4 for usage information). GC content was calculated using annotated bases only.

Sequence data sources: Unspliced Chlamydomonas nuclear gene sequences used for the analyses were generated using a custom python script (available in the associated GitHub repository) to extract whole-gene, 5′-UTR, ATG-Stop, and 3′-UTR sequences from the genome based on their start and end positions in the current gene models (Phytozome version 5.5). Chlamydomonas gene models are based on predictions using Augustus (annotation version u11.6) and refined using a range of RNA-seq datasets. Files containing the whole genome nucleotide sequence (version 5.0) and the annotation information for each of the 17,741 nuclear genes (version 5.5) were downloaded from Phytozome 12 and are provided as precursor files for running the BACSearcher script (see Supplemental Protocol 2). Sequence data for *A. thaliana* (TAIR10 assembly) and *Triticum aestivum* nuclear genes (International Wheat Genome Consortium assembly) were obtained from EnsemblPlants BioMart. Analysis was limited to the 105,200 chromosome-assigned wheat genes. Sequence data for *Saccharomyces cerevisiae* (S288C reference genome, 2015 release) were obtained from the Saccharomyces Genome Database. Gene sequences were appended to include all annotated UTRs and introns, resulting in a dataset that is more closely comparable to the unspliced gene data used for Chlamydomonas, Arabidopsis, and wheat.

Analysis of repeats: Repetitive regions in the nucleotide sequences analyzed in this work are categorized into simple and global repeats. We use the term simple repeats to refer to relatively short (tens to hundreds of bases) repetitive regions in a nucleotide sequence that display regular or semi-regular repeating patterns. We include consecutive repeating motifs of varying unit lengths, known as tandem repeats, as well as inverted patterns in which a short region is followed closely (or immediately, if palindromic) by its reverse complement sequence. Chlamydomonas genes were analyzed for tandem repeats using Tandem Repeats Finder (Benson, 1999). The default settings were modified to provide a cut-off for detection such that no repeats under 10 bp in length were reported (see Supplemental Protocol 3A). All Tandem Repeats Finder outputs were processed using a

custom python script and analyzed in spreadsheet format to generate mean values for the number of genes with either: (1) at least one mononucleotide repeat ⩾10 bp in length and with ⩾90% identity; (2) at least one di- or trinucleotide repeat ⩾20 bp in length with ⩾90% identity; (3) at least one tandem repeat ⩾20 bp in length, with a period length of four or more (tetra+), with ⩾90% identity; and (4) the mean number of repeats of these types per kilobase of sequence.

Chlamydomonas genes were analyzed for inverted repeats using the Palindrome Analyser webtool (Brázda et al., 2016), available at http://palindromes.ibp.cz/#/en/palindrome. The default settings were modified to report repeats with a maximum of 1 mismatch for every 10 bp of stem sequence, a maximum spacer length of 10 bp, and a maximum total length of 210 bp (see Supplemental Protocol 3B for settings). All Palindrome Analyser outputs were downloaded and analyzed in spreadsheet format to generate mean values for the number of genes containing one or more inverted repeats over 20 bp long with ⩾90% identity and the mean number of inverted repeats of this type per kilobase.

All nuclear genes from Chlamydomonas (Figure 1B), Arabidopsis, yeast, and wheat (Figure 1F), and recombineering target regions (Figure 3B and C) were analyzed for global repeats using the NCBI WindowMasker program (Morgulis et al., 2006). We use the term global repeats to denote the combined number of individual masked regions detected by the WindowMasker modules DUST and WinMask. DUST detects and masks shorter repetitive regions including tandem and inverted repeats, overlapping with and providing support for the Tandem Repeats Finder and Palindrome Analyser outputs. WinMask detects and masks families of longer repetitive regions that do not necessarily occur adjacently in the genome. Default settings were used throughout (see Supplemental Protocol 3C). These modules mask repetitive regions using only the supplied sequence as a template.

Chlamydomonas repeats localized to the 5′-UTRs, ATG-Stop regions, and 3′-UTRs were distinguished using positional information from Phytozome (genome annotation version 5.5). Repeats that spanned from a 5′-UTR across the start codon or across the stop codon into the 3′-UTR were not counted, although were included in the whole-gene repeat analyses described above.

uORFs, transcripts and intron analysis: Data on the presence of uORFs in Chlamydomonas transcripts were obtained from the results of a BLASTP analysis performed by Cross (2015) and adapted to provide the per-gene values. A list of Chlamydomonas transcripts was downloaded from Phytozome Biomart and used to identify the number of genes with more than one transcript model. Genomic data detailing the number and order of exons within each gene were also downloaded from Phytozome Biomart; this information was used to ascertain the number of genes containing introns in their translated and untranslated regions.

Primer analysis: To assess the impact of inefficient priming on PCR-based cloning, analysis was performed on a dataset of PCR primers designed to clone every gene in the Chlamydomonas genome from start to stop codon using gDNA as the template and generated such that the predicted T_m difference for each pair was not more than 5°C where possible. Primer sequences were then assessed against four thresholds pertaining to efficient priming, set in accordance with advice found in the Primer3 manual, support pages provided by IDT, and the Premier Biosoft technical notes. These thresholds relate to primer length, propensity for secondary structure formation, the presence of repeats, and the GC content of the 3′-end. Long primers can have a reduced amplification efficiency, secondary structure formation can reduce the number of primers available to bind to the intended template during a PCR, multiple repeats can increase the risk of mispriming, and a high 3′-end GC content can increase the risk of primer–dimer formation. Thresholds for each were set as follows: (1) primer length should not be more than 30 bp; (2) the ΔG required to disrupt predicted secondary structures should be above −9 kcal mol^{-1} at 66 or 72°C; (3) tandem single nucleotides or dinucleotide motifs should repeat no more than 4 times, and (4) the 3′-end should consist of no more than 4 G/C bases in a row. The number of primers in breach of each of these thresholds is shown in Figure 1D as a percentage of the dataset. The percentage of unsuitable primer pairs was calculated by counting pairs for which one or both primers breached one or more of these thresholds. T_m considerations were omitted from analysis since Chlamydomonas genes have an unusually high GC content, so primers designed to amplify gDNA are expected to have higher than recommended T_m's according to generic primer design guidelines. GC content was calculated using annotated bases only.

To complement these results, primers were analyzed using the check_primers algorithm from Primer3 (Rozen and Skaletsky, 2000). The settings used were the default settings for Primer3Plus (Untergasser et al., 2007)—an updated, online version of the Primer3 package—with minimal modifications that included removing the Tm constraints (see Supplemental Protocol 3D for full settings used). The output was analyzed with a custom python script that reported the primary reason for rejection of individual primers (see Supplemental Protocol 4C). T_m was removed as a constraint to allow for more detailed analysis of primer sequence parameters, since the default maximum allowable T_m for Primer3Plus is 63°C, which results in rejection of almost 90% of primers for this reason alone if used. 1.6% of primers were too long to be considered for analysis (>36 bp); these were included in Figure 1D (orange bar) as having been rejected for breaching the length constraint. The majority of rejected primers produced one of the following three reasons for rejection: (1) "high end complementarity" for primer pairs, which implies a high likelihood that the 3′-ends of the forward and reverse primers will anneal, enabling amplification of a short, heterogeneous primer-dimer (cross-dimer); (2) "high end complementarity" for single

primers, which implies a high likelihood that a primer's 3′-end will bind to that of another identical copy, self-priming to form a homogenous primer-dimer (self-dimer); and (3) "high any complementarity" for single primers, which implies a high likelihood of self-annealing without necessarily self-priming, relevant to both the inter-molecular annealing of identical copies and to instances of hairpin formation resulting from intra-molecular annealing. Primers rejected for these three reasons are labeled in Figure 1D (orange bar) as cross-dimers, self-dimers, and hairpins, respectively.

Note on differences between the Chlamydomonas BAC library strain and CLiP mutant strain: The Chlamydomonas BAC library was constructed using the genome reference strain CC-503, so researchers working with alternative strains need to take into account potential genomic divergence. For example, here we transformed recombineered DNA from the BAC library into CC-4533, the wild-type strain used for the CLiP mutant collection and a popular strain for studying the CCM. Genomic analysis of CC-4533 relative to CC-503 has revealed 653 instances of variation that may be disruptive to protein function, although only three of these are unique to CC-4533 compared to other common laboratory strains (Li et al., 2016). Two genes affected by this variation were successfully cloned using our recombineering pipeline: Cre06.g250650 in CC-4533 contains three short deletions relative to CC-503 with an uncertain impact on the protein, while Cre06.g249750 in CC-4533 contains a predicted inversion affecting the final three exons and part of the 3′-UTR.

BACSearcher python resource

Suitable BACs containing the target genes were identified using a python script that also identifies 50-bp binding sites for recombineering cloning primers and provides sequences for primers that can be used to check for the presence of a target gene within a BAC (see Supplemental Protocol 2). BACSearcher output is available for all 17,741 genes in the genome in Supplemental Data Set 1. For individual targets in our recombineering pipeline that were not covered by a BAC in the BACSearcher output, an alternative method was employed to search for BAC coverage. This method is detailed in Supplemental Protocol 2, along with usage and modification instructions for BACSearcher, including instructions on how to output suitable fosmids for all genes in the genome. BACSearcher resources can be found in the associated GitHub repository at https://github.com/TZEmrichMills/Chlamydomonas_recombineering.

Accession numbers

Sequence data from this article can be found in Phytozome, the Plant Comparative Genomics portal of the Department of Energy's Joint Genome Institute, under the following accession numbers: Cre11.g467712: SAGA1; Cre09.g412100: PSAF; Cre03.g155001: ISA1; Cre10.g435800: CSP41B; Cre17.g702500: TAB2; Cre10.g452800: LCIB; Cre09.g394473: LCI9.

All plasmid sequences are available in Supplemental Data Set 4 and have been deposited in GenBank with the following IDs: pLM099, MT737960; pLM160, MT737961; pLM161, MT737962; pLM162, MT737963; pLM459, MT737964. Plasmids are available from the Chlamydomonas Resource Center (https://www.chlamycollection.org/), as are the BAC and fosmid libraries. Full protocols for batch and high-throughput recombineering are available in Supplemental Protocol 1. Data used for the genome analyses presented in Figure 1 are available on request. The python computer code used for identifying BACs, fosmids and suitable homology regions for recombineering is available at https://github.com/TZEmrichMills/Chlamydomonas_recombineering.

Supplemental Data

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

Supplemental Figure 1. Batch-scale recombineering results.

Supplemental Figure 2. Validation of fluorescently localized lines.

Supplemental Figure 3. Complementation of the *lcib* CLiP mutant.

Supplemental Table 1. Mann-Whitney U test statistics.

Supplemental Protocol 1. Protocols for batch and large-scale recombineering.

Supplemental Protocol 2. BACSearcher usage.

Supplemental Protocol 3. Bioinformatics software usage.

Supplemental Protocol 4. Bioinformatics python analysis.

Supplemental Data Set 1. BACSearcher output.

Supplemental Data Set 2. Large-scale pipeline results summary.

Supplemental Data Set 3. Oligonucleotide sequences.

Supplemental Data Set 4. Plasmid sequences.

Acknowledgments

The authors would like to thank Guy Mayneord for programming advice, the University of York Biosciences Technology Facility for confocal microscopy and bioinformatics support, and Mihail Sarov (University of Dresden, Germany) for technical advice and providing plasmids pSC101-BAD-gbaA-tet and pNPC2.

Funding

This work was funded by UK Biotechnology and Biological Sciences Research Council Grants BB/R001014/1 and BB/S015337/1 (to L.C.M.M.); Leverhulme Trust Grant RPG-2017-402 (to L.C.M.M.); UKRI Future Leader Fellowship MR/T020679/1 (to L.C.M.M.); BBSRC DTP2 BB/M011151/1 (to T.Z.E.M. and M.P.J.); BBSRC DTP2 BB/M011151/1a (to J.B. and L.C.M.M.); University of York Biology Pump Priming award (to LCMM); and University of York Biology Start-up grant (to L.C.M.M.).

References

Aksoy M, Forest C (2019) One step modification of *Chlamydomonas reinhardtii* BACs using the RED/ET system. Mediterranean Agric Sci **32**: 49–55

Baier T., Wichmann J., Kruse O., Lauerse KJ (2018) Intron-containing algal transgenes mediate efficient recombinant gene expression in the green microalga *Chlamydomonas reinhardtii*. Nucleic Acids Res **46**: 6909–6919

Barahimipou R, Strenker D, Neuper J, Schrod M, Merchan SS, Boc R (2015) Dissecting the contributions of GC content and codon usage to gene expression in the model alga *Chlamydomonas reinhardtii*. Plant J **84**: 704–717

Benson G (1999) Tandem repeats finder: a program to analyze DNA sequences. Nucleic Acids Res **27**: 573–580

Bernard P, Couturier M (1992) Cell killing by the F plasmid CcdB protein involves poisoning of DNA-topoisomerase II complexes. J Mol Biol **226**: 735–745

Bindels DS, Haarbosch L, Van Weeren L, Postma M, Wiese KE, Mastop M, Aumonier S, Gotthard G, Royant A, Hink MA (2017) mScarlet: a bright monomeric red fluorescent protein for cellular imaging. Nat Methods **14**: 53

Brázda V, Kolomazník J, Lýsek J, Hároníková L, Coufal J, Šťastný J (2016) Palindrome analyser – a new web-based server for predicting and evaluating inverted repeats in nucleotide sequences. Biochem Biophys Res Commun **478**: 1739–1745

Brueggeman AJ, Gangadharaiah DS, Cserhati MF, Casero D, Weeks DP, Ladunga I (2012) Activation of the carbon concentrating mechanism by CO_2 deprivation coincides with massive transcriptional restructuring in *Chlamydomonas reinhardtii*. Plant Cell **24**: 1860–1875

Brumos J, Zhao C, Gong Y, Soriano D, Patel AP, Perez-Amador MA, Stepanova AN, Alonso JM (2020) An improved recombineering toolset for plants. Plant Cell **32**: 100

Clarke L, Rebelo C, Goncalves J, Boavida M, Jordan P (2001) PCR amplification introduces errors into mononucleotide and dinucleotide repeat sequences. Mol Pathol **54**: 351

Copeland NG, Jenkins NA, Court DL (2001) Recombineering: a powerful new tool for mouse functional genomics. Nat Rev Genet **2**: 769–779

Cross FR (2015) Tying down loose ends in the Chlamydomonas genome: functional significance of abundant upstream open reading frames. G3 **6**: 435–446

Crozet P, Navarro FJ, Willmund F, Mehrshahi P, Bakowski K, Lauersen KJ, Pérez-Pérez M-E, Auroy P, Gorchs Rovira A, Sauret-Gueto S, et al. (2018) Birth of a photosynthetic chassis: a MoClo toolkit enabling synthetic biology in the microalga *Chlamydomonas reinhardtii*. ACS Synth Biol **7**: 2074–2086

Dauvillée D, Stampacchia O, Girard-Bascou J, Rochaix, JD (2003) Tab2 is a novel conserved RNA binding protein required for translation of the chloroplast psaB mRNA. EMBO J **22**: 6378–6388

Duanmu D, Miller AR, Horken KM, Weeks DP, Spalding MH (2009) Knockdown of limiting-CO_2-induced gene HLA3 decreases HCO_3^- transport and photosynthetic Ci affinity in *Chlamydomonas reinhardtii*. Proc Natl Acad Sci U S A **106**: 5990–5995

Engel BD, Schaffer M, Kuhn Cuellar L, Villa E, Plitzko JM, Baumeister W (2015) Native architecture of the Chlamydomonas chloroplast revealed by in situ cryo-electron tomography. Elife **4**: e04889

Fang W, Si Y, Douglass S, Casero D, Merchant SS, Pellegrini M, Ladunga I, Liu P, Spalding MH (2012) Transcriptome-wide changes in *Chlamydomonas reinhardtii* gene expression regulated by carbon dioxide and the CO_2-concentrating mechanism regulator CIA5/CCM1. Plant Cell **24**: 1876–1893

Fukuzawa H, Fujiwara S, Yamamoto Y, Dionisio-Sese ML, Miyachi S (1990) cDNA cloning, sequence, and expression of carbonic anhydrase in *Chlamydomonas reinhardtii*: regulation by environmental CO_2 concentration. Proc Natl Acad Sci USA **87**: 4383

Gao H, Wang Y, Fei X, Wright DA, Spalding MH (2015) Expression activation and functional analysis of HLA3, a putative inorganic carbon transporter in *Chlamydomonas reinhardtii*. Plant J **82**: 1–11

Goedhart J, Von Stetten D, Noirclerc-Savoye M, Lelimousin M, Joosen L, Hink MA, Van Weeren L, Gadella TW, Royant A (2012) Structure-guided evolution of cyan fluorescent proteins towards a quantum yield of 93%. Nat Commun **3**: 1–9

Hommelsheim CM, Frantzeskakis, L., Huang, M., Ülker, B (2014) PCR amplification of repetitive DNA: a limitation to genome editing technologies and many other applications. Sci Rep **4**: 5052

Itakura AK, Chan KX, Atkinson N, Pallesen L, Wang L, Reeves G, Patena W, Caspari O, Roth R, Goodenough U (2019) A Rubisco-binding protein is required for normal pyrenoid number and starch sheath morphology in *Chlamydomonas reinhardtii*. Proc Natl Acad Sci USA **116**: 18445–18454

Kobayashi Y, Takusagawa M, Harada N, Fukao Y, Yamaoka S, Kohchi T, Hori K, Ohta H, Shikanai T, Nishimura Y (2015) Eukaryotic components remodeled chloroplast nucleoid organization during the green plant evolution. Genome Biol Evol **8**: 1–16

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

Levinson G, Gutman GA (1987) Slipped-strand mispairing: a major mechanism for DNA sequence evolution. Mol Biol Evol **4**: 203–221

Li X, Zhang R, Patena W, Gang SS, Blum SR, Ivanova N, Yue R, Robertson JM, Lefebvre PA, Fitz-Gibbon ST, et al. (2016) An indexed, mapped mutant library enables reverse genetics studies of biological processes in *Chlamydomonas reinhardtii*. Plant Cell **28**: 367

Li X, Patena W, Fauser F, Jinkerson RE, Saroussi S, Meyer MT, Ivanova N, Robertson JM, Yue R, Zhang R, et al. (2019) A genome-wide algal mutant library and functional screen identifies genes required for eukaryotic photosynthesis. Nature Genetics **51**: 627–635

López-Paz C, Liu D, Geng S, Umen JG (2017) Identification of *Chlamydomonas reinhardtii* endogenous genic flanking sequences for improved transgene expression. Plant J **92**: 1232–1244

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

Mackinder LC, Meyer MT, Mettler-Altmann T, Chen VK, Mitchell MC, Caspari O, Freeman Rosenzweig ES, Pallesen Lz, Reeves G, Itakura A, et al. (2016) A repeat protein links Rubisco to form the eukaryotic carbon-concentrating organelle. Proc Natl Acad Sci **113**: 5958–5963

Mackinder LCM (2017) The *Chlamydomonas* CO_2-concentrating mechanism and its potential for engineering photosynthesis in plants. New Phytologist **217**: 54–61

Mackinder LCM, Chen C, Leib RD, Patena W, Blum SR, Rodman M, Ramundo S, Adams CM, Jonikas MC (2017) A spatial interactome reveals the protein organization of the algal CO_2-concentrating mechanism. Cell **171**: 133–147.e114

Merchant SS, Prochnik SE, Vallon O, Harris EH, Karpowicz SJ, Witman GB, Terry A, Salamov A, Fritz-Laylin LK, Marechal-Drouard L, et al. (2007) The Chlamydomonas genome reveals the evolution of key animal and plant functions. Science **318**: 245–250

Mettler T, Mühlhaus T, Hemme D, Schöttler M-A, Rupprecht J, Idoine A, Veyel D, Pal SK, Yaneva-Roder L, Winck FV, et al. (2014) Systems analysis of the response of photosynthesis, metabolism, and growth to an increase in irradiance in the photosynthetic model organism *Chlamydomonas reinhardtii*. The Plant Cell **26**: 2310

Meyer MT, Genkov T, Skepper JN, Jouhet J, Mitchell MC, Spreitzer RJ, Griffiths H (2012) Rubisco small-subunit α-helices control pyrenoid formation in Chlamydomonas. Proc Natl Acad Sci USA **109**: 19474–19479

Morgulis A, Gertz EM, Schäffer AA, Agarwala R (2006) WindowMasker: window-based masker for sequenced genomes. Bioinformatics **22**: 134–141

Mouille G, Maddelein ML, Libessart N, Talaga P, Decq A, Delrue B, Ball S (1996) Preamylopectin processing: a mandatory step for starch biosynthesis in plants. Plant Cell **8**: 1353

Nelms BL, Labosky PA (2011) A predicted hairpin cluster correlates with barriers to PCR, sequencing and possibly BAC recombineering. Sci Rep **1**: 106

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

Onishi M, Pringle JR (2016) Robust transgene expression from bicistronic mRNA in the green alga *Chlamydomonas reinhardtii*. G3 **6**: 4115–4125

Poser I, Sarov M, Hutchins JR, Heriche JK, Toyoda Y, Pozniakovsky A, Weigl D, Nitzsche, A, Hegemann B, Bird AW, et al. (2008) BAC TransgeneOmics: a high-throughput method for exploration of protein function in mammals. Nat Methods **5**: 409–415

Rae BD, Long BM, Förster B, Nguyen ND, Velanis CN, Atkinson N, Hee WY, Mukherjee B, Price GD, McCormick AJ (2017) Progress and challenges of engineering a biophysical carbon dioxide-concentrating mechanism into higher plants. J Exp Bot **68**: 3717–3737

Rasala BA, Lee PA, Shen Z, Briggs SP, Mendez M, Mayfield SP (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

Rozen S, Skaletsky H (2000) Primer3 on the WWW for general users and for biologist programmers. *In* Bioinformatics Methods and Protocols. Springer, Berlin, pp. 365–386

Sahdev S, Saini S, Tiwari P, Saxena S, Saini KS (2007) Amplification of GC-rich genes by following a combination strategy of primer design, enhancers and modified PCR cycle conditions. Mol Cell Probes **21**: 303–307

Salomé PA, Merchant SS (2019) A series of fortunate events: introducing Chlamydomonas as a reference organism. Plant Cell **31**: 1682.

Sarov M, Schneider S, Pozniakovski A, Roguev A, Ernst S, Zhang Y, Hyman AA, Stewart AF (2006) A recombineering pipeline for functional genomics applied to *Caenorhabditis elegans*. Nat Methods **3**: 839–844

Sarov M, Murray JI, Schanze K, Pozniakovski A, Niu W, Angermann K, Hasse S, Rupprecht M, Vinis E, Tinney M (2012) A genome-scale resource for in vivo tag-based protein function exploration in *C. elegans*. Cell **150**: 855–866

Sarov M, Barz C, Jambor H, Hein MY, Schmied C, Suchold D, Stender B, Janosch S, K JV, Krishnan RT, et al. (2016) A genome-wide resource for the analysis of protein localisation in Drosophila. Elife **5**: e12068

Schroda M (2019) Good news for nuclear transgene expression in Chlamydomonas. Cells **8**: 1534

Shaner NC, Lambert GG, Chammas A, Ni Y, Cranfill PJ, Baird MA, Sell BR, Allen JR, Day RN, Israelsson M (2013) A bright monomeric green fluorescent protein derived from *Branchiostoma lanceolatum*. Nat Methods **10**: 407

Strenkert D, Schmollinger S, Gallaher SD, Salomé PA, Purvine SO, Nicora CD, Mettler-Altmann T, Soubeyrand E, Weber AP, Lipton MS (2019) Multiomics resolution of molecular events during a day in the life of Chlamydomonas. Proc Natl Acad Sci USA **116**: 2374–2383

Tachiki A, Fukuzawa H, Miyachi S (1992) Characterization of carbonic anhydrase isozyme CA2, which is the CAH2 gene product, in *Chlamydomonas reinhardtii*. Biosci Biotechnol Biochem **56**: 794–798

Toyokawa C, Yamano T, Fukuzawa H (2020) Pyrenoid starch sheath is required for LCIB localization and the CO_2-concentrating mechanism in green algae. Plant Physiol **182:** 1883–1893

Uniacke J, Zerges W (2009) Chloroplast protein targeting involves localized translation in *Chlamydomonas*. Proc Natl Acad Sci U S A **106**: 1439–1444

Untergasser A, Nijveen H, Rao X, Bisseling T, Geurts R, Leunissen JA (2007) Primer3Plus, an enhanced web interface to Primer3. Nucleic Acids Res **35:** W71–W74

Wang Y, Stessman DJ, Spalding MH (2015) The CO_2 concentrating mechanism and photosynthetic carbon assimilation in limiting CO_2: how *Chlamydomonas* works against the gradient. Plant J **82**: 429–448

Weiner I, Atar S, Schweitzer S, Eilenberg H, Feldman Y, Avitan M, Blau M, Danon A, Tuller T, Yacoby I (2018) Enhancing heterologous expression in *Chlamydomonas reinhardtii* by transcript sequence optimization. Plant J **94**: 22–31

Yamano T, Tsujikawa T, Hatano K, Ozawa S-i, Takahashi Y, Fukuzawa H (2010) Light and low-CO_2-dependent LCIB–LCIC complex localization in the chloroplast supports the carbon-concentrating mechanism in *Chlamydomonas reinhardtii*. Plant Cell Physiol **51**: 1453–1468

Yu D, Ellis HM, Lee EC, Jenkins NA, Copeland NG (2000) An efficient recombination system for chromosome engineering in *Escherichia coli*. Proc Natl Acad Sci USA **97**: 5978–5983

Zhan Y, Marchand CH, Maes A, Mauries A, Sun Y, Dhaliwal JS, Uniacke J, Arragain S, Jiang H, Gold ND (2018) Pyrenoid functions revealed by proteomics in *Chlamydomonas reinhardtii*. PLoS One **13**: e0185039

Zhang R, Patena W, Armbruster U, Gang SS, Blum SR, Jonikas MC (2014) High-throughput genotyping of green algal mutants reveals random distribution of mutagenic insertion sites and endonucleolytic cleavage of transforming DNA. Plant Cell **26**: 1398–1409

Zones JM, Blaby IK, Merchant SS, Umen JG (2015) High-resolution profiling of a synchronized diurnal transcriptome from *Chlamydomonas reinhardtii* reveals continuous cell and metabolic differentiation. Plant Cell **27**: 2743–2769