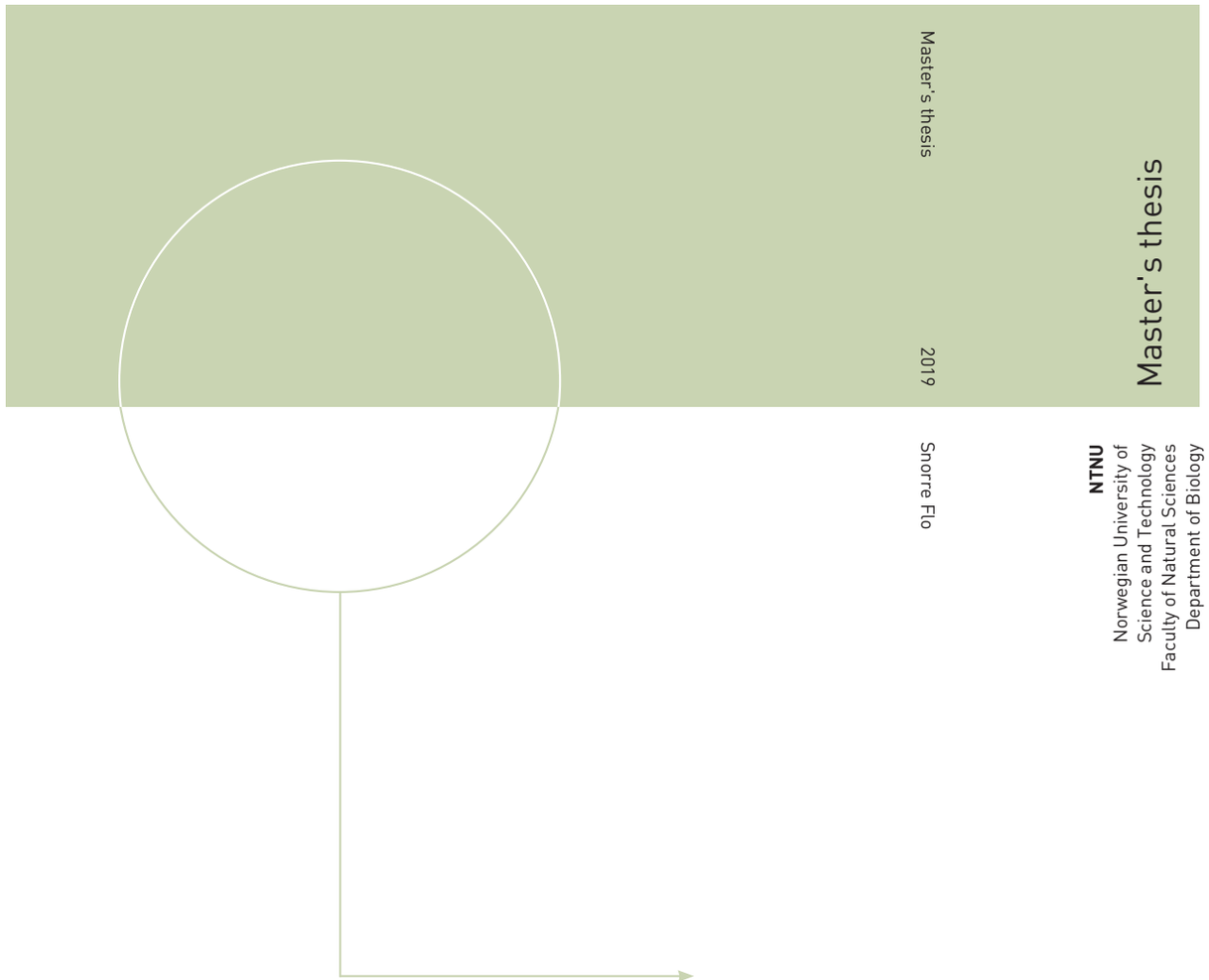


Master's thesis

2019

Snorre Flo

NTNU
Norwegian University of
Science and Technology
Faculty of Natural Sciences
Department of Biology

Snorre Flo

Functional studies and characterization of the light-harvesting systems of a marine alga

May 2019

NTNU
Norwegian University of
Science and Technology

NTNU
Norwegian University of
Science and Technology

NTNU

Norwegian University of
Science and Technology

Functional studies and characterization of the light-harvesting systems of a marine alga

Snorre Flo

Master's thesis in Cell- and Molecular Biology
Submission date: May 2019
Supervisor: Atle M. Bones
Co-supervisor: Marianne Nymark
Amit K. Sharma

Norwegian University of Science and Technology
Department of Biology

Acknowledgements

The majority of the work presented in this thesis was conducted at the Cell, Molecular Biology and Genomics (CMBG) group at the Department of Biology at the Norwegian University of Science and Technology (NTNU, Trondheim, Norway). Photobioreactor growth-trials were conducted at the Norwegian Centre for Plankton Technology at SINTEF Ocean (Trondheim, Norway) and various analyses at Trondheim Biological Station (TBS), also belonging to Dept. Biology (NTNU, Trondheim).

Foremost I would like to thank my supervisors Dr. Marianne Nymark, PhD student Amit K. Sharma and Prof. Atle M. Bones for their support, guidance and for allowing me to write my thesis on this intriguing creature. Especially I would like to thank Marianne Nymark and Amit Sharma for teaching me new techniques, for always being available to help me when I'm stuck, and with various daily tasks. Your encouraging comments have been key to the health of my algal cultures.

Thanks to the CMBG group for including me, and for providing valuable feedback on my thesis. A shout-out goes out to Assoc. Prof. Per Winge who has shared his vast knowledge. I have learned a lot.

A special thanks goes out to Matilde Chauton (SINTEF Ocean), Kjersti Andresen (TBS, Dept. Biology) and Glaucia Fragoso (TBS, Dept. Biology) for your competent guidance and helpfulness throughout various analyses and experiments.

Last but not least, I would like to thank my friends and family for their love and support. I look forward to the journeys ahead.

Trondheim

May 2019

Snorre Flo

Abbreviations

Alb3b	Albino 3b-insertase
ATP	Adenosine-triphosphate
AtpB	ATP-synthase subunit β
Cas	CRISPR-associated protein
Chl	Chlorophyll
CRISPR	Clustered regularly interspaced short palindromic repeats
Ddx	Diadinoxanthin
DES	De-epoxidation state index
diaCas9	*Streptococcus pyogenes* diatom-Cas9
DIN	Dissolved inorganic nitrogen
DIP	Dissolved inorganic phosphorous
DSB	Double-stranded break
Dtx	Diatoxanthin
EGT	Evolutionary gene transfer
E_k	Light-saturation index
E_{PAR}	Scalar irradiance of photosynthetically active radiation (PAR)
ETC	Electron transport chain
F_0	Minimal fluorescence
FCP	Fucoxanthin Chl *a*/*c* protein
FL	Fluorescence
F_m	Maximal fluorescence
Fuco	Fucoxanthin
gRNA	guide-RNA
HDR	Homology-directed repair
hifiCas9	High-fidelity *Streptococcus pyogenes* diatom-Cas9
HL	High-light
KO	Knockout
LHC	Light-harvesting complex
LHCF	Major Fucoxanthin Chl *a/c* binding protein
LHCFm	Multiple *lhcf* knockout mutant
LL	Low-light

MM	Mismatch
NHEJ	Non-homologous end-joining
NPQ	Non-photochemical quenching
PAM	Protospacer-adjacent motif
PQ	Photochemical quenching
PS	Photosystem
D1	Photosystem II protein D1
qE	pH- or energy-dependent component of NPQ
RC	Reaction-centre
RET	Resonance-energy transfer
$rETR_{max}$	Maximum relative electron transport rate
ROS	Reactive oxygen species
*Sp*Cas9	*Streptococcus pyogenes* Cas9
TLA	Truncated light-harvesting antenna
α	Light-utilization coefficient
Φ_{PSII}	Effective quantum yield of PSII photochemistry ($F_v´/F_m´$)
$\Phi_{PSIImax}$	Maximum quantum yield of PSII photochemistry (F_v/F_m)

Abstract

The CRISPR-Cas9 technology has presented a powerful tool for genome editing of microalgae. However, the Cas9 enzyme from *Streptococcus pyogenes* (SpCas9) is promiscuous and may insert unwanted modifications at off-target sites. To surpass this problem, several research groups have attempted at increasing its fidelity, one of which came up with the general strategy of replacing non-specific residues in the backbone (*Sp*Cas9-HF1) to lower the innate activity of Cas9:gRNA. The same modifications were made to create a high fidelity *Sp*Cas9 (hifiCas9) for use in diatoms. We then targeted multiple highly similar *LHCF* genes in the *Phaeodactylum tricornutum* genome to test the on- and off-target activities of hifiCas9. The findings were comparable to what has previously been reported in a human cell line, with a significantly reduced number of off-target mutagenesis events. However, a ~75% of the WT Cas9 on-target editing efficiency signified that activity was compromised for the sake of higher fidelity. Nevertheless, hifiCas9 may provide a good alternative to diaCas9, especially in applications that require high specificity.

A second reason for targeting the *LHCF* genes was that they code for important subunits of the light-harvesting apparatus situated in the thylakoids. So-called truncated light-harvesting antenna (TLA) mutants have decreased ability to absorb light, but have been shown to increase the overall productivity of other algal species when grown in light-limited cultures. Three *lhcf* mutants were selected for characterization based on their gradient-like phenotypes from brown to bright green, and preliminary sequencing data that portrayed one, three or five biallelic *lhcf* knockout mutations. It was found that the least perturbed *lhcf1* mutant was undistinctive from WT, and that a mutant with quintuple deletions (*lhcf1, 2, 3, 4* and *5*) was "over-the-top" causing a structurally deficient photosystem, and subsequently inefficient photosynthesis. The mutant showing the best photosynthetic performance was found to be LHCFm 6.1.11 – containing knockouts to *lhcf1*, *2* and *5*, and an in-frame gene-fusion product of *LHCF3* and *LHCF4*. LHCFm 6.1.11 was characterized by lowered Fuco:Chl *a*, decreased $Fuco_{RED}$ absorbance, and variable fluorescence indicated an increased light-saturation index (E_k), improved maximum relative transport rate ($rETR_{max}$), and lowered ability to modify the size of its light-harvesting antennas.

A final objective was set to test the performance of putative TLA mutants in light-limited cultures, but this experiment came short due to nutrient limitation. Instead we argue the importance of a rich medium, an increased light-path and monitoring of nutrient conditions – the latter which may be easily achieved by use of fluorometry.

Sammendrag

CRISPR-Cas9 har vist seg som et verdifullt verktøy i gen-redigering av mikroalger. En problemstilling knyttet til Cas9-basert gen-redigering er at Cas9:guide-RNA-komplekset er promiskuøst, og kan kutte også på gen-lokasjoner som ikke fullstendig komplementerer guide-RNAet. Det har blitt gjort flere forsøk på å løse dette problemet, deriblant har man laget modifiserte Cas9 enzymer ved å erstatte aminosyrer som interagerer uspesifikt med DNA. I denne oppgaven har vi testet ut et høy-fidelitets Cas9 enzym med de samme erstatningene (hifiCas9), for å undersøke hvor effektivt den kutter på både tilsiktede (on-target) og utilsiktede (off-target) DNA-lokasjoner. Det ble funnet at enzymet oppnådde tilsiktet effektivitet tilsvarende ~75% av villtype Cas9, og at utilsiktede mutasjoner var redusert betraktelig med kun én observasjon. Således er det tydelig at høyere fidelitet kommer på bekostning av lavere aktivitet. Likevel kan hifiCas9 være et godt alternativ til diaCas9, og spesielt i tilfeller som krever at det skilles mellom svært like DNA sekvenser.

De redigerte *LHCF*-genene koder for proteiner som inngår i lyshøstingsantennene (FCP komplekser) i *Phaeodactylum tricornutum*. Mutanter med trunkerte lyshøstingsantenner (TLA) har svekket evne til å absorbere lys, og kan derfor utnyttes i industrielle produksjonsprosesser der kulturene når høy tetthet og følgelig lys-begrensede forhold. TLA mutanter har vist seg å øke produktiviteten av kulturer bestående av andre algegrupper, men dette har ikke enda blitt vist for kulturer av diatoméer. Det ble derfor valgt ut tre lovende mutanter som viste en grønnaktig farge i kombinasjon med multiple tap av *LHCF*-gener. De ble karakterisert ved kvantifisering av vekstrater, pigment, FCP innhold, spektrale egenskaper og fotosyntetisk ytelse. Det ble vist at den minst påvirkede mutanten var nærmest lik villtypen, mens en mutant med tap av fem *LHCF*-gener var for svekket, og med ineffektiv fotosyntese som følge. Den mest lovende mutanten (LHCFm 6.1.11) hadde tap av *lhcf1*, *2* og *5* og et funksjonelt gen-fusjons-produkt av *LHCF3* og *LHCF4*. Denne mutanten hadde redusert mengde Fuco:Chl *a*, mindre $Fuco_{RED}$ absorbans, økt maksimal elektrontransportrate, og en redusert evne til å endre størrelsen på lyshøstingsantennen.

Et siste mål var å teste ut potensielle TLA mutanter i lys-begrensede fotobioreaktor-kulturer, men dette pilotforsøket kom til kort på grunn av nærings-begrensning tidlig i forsøksperioden. Derfor argumenteres det for viktigheten av et rikt medium, økt lys-lengde og overvåkning av nærings-forhold ved bruk av fluorometri.

Table of Contents

1 Introduction

1.1 General Introduction

1.1.1 Primary production

Photosynthetic organisms, or primary producers, are vital to the biosphere as we know it. Photosynthesis involves utilizing H_2O and light to process CO_2 into organic, energy-rich molecules, simultaneously producing O_2 as a by-product [1]. Primary producers in the ocean include multicellular algae (seaweeds and kelp) and single-celled phytoplankton (eukaryote algae and prokaryote cyanobacteria) [2]. These organisms provide organic carbon that most marine organisms depend on. Describing their productivity has therefore been the goals of many oceanographers – with various methodologies and results. Usually, primary productivity is measured as the net amount of carbon available to the next trophic level – net primary production (NPP), which equals the difference between autotrophic photosynthesis and respiration [3]. Forerunners such as Gaarder and Gran [4] used the oxygen evolution method to estimate phytoplankton productivity in the Oslo fjord. Since then, a myriad of technologies have been developed to obtain increasingly precise estimates of primary production, from utilizing stable isotopes of carbon (^{14}C-method) [5] to using fluorescence emission from the photosynthetically universal pigment chlorophyll *a* as a proxy [6]. Almost a century worth of development has provided us with means of expanding the area of study from the Oslo fjord to the vast ocean, with increased precision. Thanks to integrative approaches combining *in situ* sampling, remote satellite-imaging-derived data and computer modelling – it is now estimated that marine primary production rivals that of terrestrial ecosystems, and contributes equally to the global NPP – approximately 105 petagrams of carbon per year [3,7,8]. While the ecological group of marine primary producers is extremely diverse, one eukaryote group in particular, the single-celled diatoms (Bacillariophyceae) are photosynthetic workhorses estimated to annually assimilate as much organic carbon as the rainforests combined [9] – approximately one-fifth of the global NPP [3,10].

1.1.2 Diatoms in the environment

They occupy benthos and pelagos from equator to polar regions; incorporate as symbionts in corals [11] and sponges [12]; attaches to seaweeds, kelp and plants as epiphytes [13]; appoint cyanobacteria as nitrogen-fixing endosymbionts [14]; and some species endure the hypersaline environment of sea-ice brine channels [15]. Diatoms are distributed throughout earths marine and freshwater systems, but perhaps especially important in the ecosystems of cool and temperate polar oceans in which other photosynthetic groups are rare [16]. Sea-ice diatoms of the Arctic and Antarctic lay the base for entire ecosystems by hosting large seasonal blooms that drive the feeding and reproductive cycles of secondary producers such as copepods and krill, ultimately feeding marine animals in the upper trophic levels [17,18]. However, diatoms are perhaps best known for their intricate cell walls, or ‘frustules’,

made from amorphous silica ($(SiO_2)_nH_2O$), which were popularized by the likes of Ernst von Haeckel [19]. Besides their ornamental beauty, these frustules are believed to provide mechanical protection from grazers [20], and because diatoms are the only functional group to have developed frustules with this chemistry, they are also dominating the biogeochemical cycling of silica [21]. It has been estimated that every atom of silicon is re-mineralized 39 times in diatom frustules before it may reach the sea floor [22]. Another striking feature is that the principal majority of petroleum reserves are remnants of fossilized diatoms [23].

1.1.3 Pennates and centrics

Traditionally, diatoms are partitioned by apparent morphology into two main groups called pennates and centrics (Fig. 1.1a and b). Most diatoms belong to the pennates that are bilateral and elongated, while centric diatoms are recognizable for being radially symmetric [24]. In general, it is conceived that centric diatoms are more often pelagic, while pennates are usually benthic [25]. Moreover, the presence of a crevice, also known as a 'raphe', allow some pennates gliding movement [26]. Both pennates and centrics have a representative model species whose genome has been fully sequenced; *Phaeodactylum tricornutum* [27] and *Thalassiosira pseudonana* [28], respectively. Genome-wide studies of said species have indicated that diatoms developed approximately 200 million years ago, and diverged into pennates and centrics about 90 million years ago [27], after which both groups have radiated to perform the services of primary producers.

1.1.4 *Phaeodactylum tricornutum* – a pennate diatom model species

Phaeodactylum tricornutum is found throughout the world's estuaries but is not particularly important in the food chain [29]. Regardless, *P. tricornutum* was chosen to become a model species for pennate diatoms on the basis of the following desired traits. It's genome was estimated to be small (<20 megabases, Mb) [29], it divided rapidly, was easy to cultivate, and was susceptible to genetic transformation [30,31]. Additionally, an abundance of over 130,000 expressed sequence tags (ESTs) could be used to identify and annotate genes. Whole-genome sequencing proved that the size of its genome (27.4 Mb) surpassed that of its estimate, but housed 10,402 genes, had less introns per gene (0.79), and was therefore much more streamlined to that of the centric diatom model species *T. pseudonana* (32.4 Mb, 11,706 genes, 1.52 introns per gene) [27]. Moreover, *P. tricornutum* displays pleomorphism – an ability to change the size and shape of the cell in response to environmental conditions. Three distinct morphotypes exist (Fig. 1.1c-e); i) the fusiform cells are typically elongated and bilateral symmetric and approximate 25 µm in length. This is the dominating morphotype in most liquid *P. tricornutum* cultures [32]. ii) the triradiate cells are comparable to fusiform cells in size but have an additional protrusion. iii) the oval cells are shorter and appears to occur at increased frequency when growth conditions are suboptimal [33].

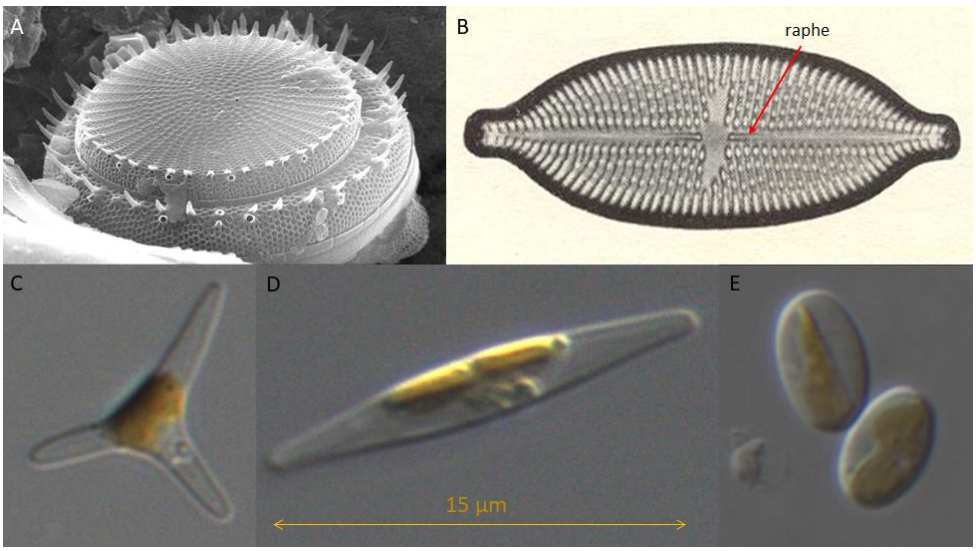


Fig. 1.1: Morphological differences between centrics, pennates and morphotypes of *P. tricornutum*. A) A centric diatoms representative in *Stephanodiscus hantzschii* (Grunow, published freely under WikiMedia Commons). B) A pennate diatom representative in *Navicula tuscula* (image courtesy of Hustedt (1956)). *P. tricornutum* exist in three distinct morphotypes; C) triradiate, D) fusiform and E) oval. All morphotypes have one large golden-brown chloroplast but vary in length. Fig. C-E are micrographs taken by Alessandra de Martino, published freely under the Creative Commons Attribution (CC-BY) license and reproduced from Vardi et al., [23].

1.1.5 Evolution and endosymbiotic history of diatoms

All photosynthetic eukaryotes acquired the ability to perform photosynthesis through endosymbiosis; evolutionary organellogenesis events that involved acquisition of a standalone cell (endosymbiont) by engulfment (phagocytosis), and its subsequent integration as a functional organelle within the host (exosymbiont) [1]. Primary endosymbiosis events, in which heterotrophic eukaryotes adopted ancestral cyanobacteria, gave rise to the primary plastid-containing red algal lineage (Rhodophyta) and green algal lineage (Chlorophyta), the latter from which all terrestrial plants descend. However, the plastids of diatoms and other brown algae (Chromalveolates) are of more complex origin. These are derived from secondary endosymbiosis events, in which a red algal cell was engulfed by an early eukaryote [35]. Molecularly speaking, primary and secondary plastids share in common that most of the their genomic DNA has been either lost or translocated to the host cell nuclear genome by evolutionary gene transfer (EGT) [1]. The remaining coding potential equals 120-130 genes in both diatoms [36] and land plants [37], which are likely detrimental to the functions of photosynthesis, transcription and translation.

There are nevertheless differences, and the evolutionary origin of primary and secondary plastids imposes distinct designs. For example, primary plastids are enveloped by two membranes, while secondary plastids have four [38]. Also, primary plastids possess differentiated domains in which photosystems II and I are spatially separated in appressed thylakoid stacks and the non-appressed lamellae, respectively [39]. This distribution has not been observed in secondary plasmids, although recent observations points out that photosystems II and I may be divided into specific sub-domains in

the thylakoid of *P. tricornutum* [39]. And as if the evolutionary history of diatoms was not sufficiently enigmatic, recent analyses have found that diatom genomes possess a high number of genes derived from a green algae [40,41]. Phylogenomic studies have indicated that *P. tricornutum* first possessed a prasinophyte-like green algal endosymbiont from an abundance of approximately 1700 green genes, constituting ~16% of the coding potential [40]. Thus, and in combination with a less pronounced abundance of red genes, it may be argued that diatoms have been through a series of secondary endosymbiosis events, by which they first obtained a green algal endosymbiont and 'green genes' by EGT, before reducing this plastid upon the entry of a novel red algal endosymbiont [40]. While this theory is still in development, it is arguably one to consider when pondering on why diatoms have gained their status as such a species-rich and ecologically important group of phytoplankton. Simply because they may have acquired two sets of genomes they could chose and pick from.

1.2 Genome editing in microalgae

1.2.1 Genome editing in algae

To enable studies that infer on the biology of diatoms, and to perform genetic engineering aiming to utilize algae industrially, it is essential that molecular biologists have an adequate set of molecular tools. Genome editing has been invaluable to biologists working with other study species, in that it offers the possibility of performing reverse genetics studies. In contrast to the traditional forward genetics approach, in which one seeks to infer the genetic basis of a phenotype, reverse genetics enables scientists to assess the traits of genetic sequences by comparing the phenotypes of a wildtype and a cell line whose gene(s) has been perturbed [1]. Perturbation is most commonly achieved by introducing changes to the gene itself or the controlling promoter region, by insertion or deletion or deletion of nucleotides (indels) or site-directed mutagenesis, respectively [1]. When changing the open reading frame (ORF) of a genetic sequence, the transcript changes with it and may encode a non-functional or a truncated protein. Likewise, when the promoter region is no longer viable for transcription initiation, there will be no product to exert the gene function. Alternatively, an endogenous exon may be replaced by an exogenous sequence as a means to remove the intrinsic trait, or to replace it with a new one (knock-in, KI). Even though these techniques have been employed for many years in other organismal systems, it has only recently become enabled in diatoms. There are several reasons why.

Only recently have the genomes of pennate (*P. tricornutum)* and centric (*T. pseudonana*) model-species of diatoms been sequenced [27,28]. Due to industrial and ecological petition for crop improvement, plants have been more extensively studied, and many more species are fully sequenced. Moreover, the problematics of crossing two individual cells by sexual reproduction limits reverse genetics approaches by eliminating the possibility of obtaining biallelic mutants by traditional hybridization. Subsequently, diatom biologists need tools that offer efficient biallelic modification, and methodologies that allow their expression *in vivo*.

1.2.2 Transformation of diatoms

At present, three methodologies have been developed to allow expression of molecular tools in diatoms. Biolistic transformation was first on the scene, after the successful introduction of selection-marker plasmid constructs in the diatoms *Cyclotella cryptica* and *Navicula saprophila* [42]. Although the method is disruptive in that it employs gold/tungsten beads coated with nucleic acids to shower-bomb a surface of cells, the methodology became an immediate success, and was repeated shortly after when *P. tricornutum* was equipped with the *sh ble*-gene (isolated from *Streptoalloteichus hindustanus)* to confer resistance to Zeocin and Phleomycin antibiotics [30,31]. Falciatore and co-workers [31] also demonstrated that diatoms may be transformed with two plasmid constructs simultaneously (co-transformation), essentially enabling larger quantities of DNA to be introduced. This is an intrusive method, disrupting the cell wall, cellular and organellar membranes and the general physiology of the cell, leading to random integration of exogenous DNA in the recipient cell genome, ultimately resulting in transgenicity [30,31]. Simultaneously, there is a high probability (~40%) of isolating transformants which only contain one of the applied vectors [31].

Electroporation applies an electrical field to increase the permeability of the cell membrane, essentially making it porous to allow more efficient uptake of DNA, protein or other molecules. Initially used in the 1980's to introduce DNA to mouse cells [43], it is only recently electroporation has been successfully applied to transform diatoms [44]. Compared with biolistic procedures, electroporation coax with avoiding irreversible genome-integration of the construct, decreased intrusiveness and higher transformation efficiency ~1.0 - 4.5 $\cdot$ 10^{-5} cells [44–46] as opposed to ~5.0 - 25.0 $\cdot$ 10^{-8} cells [47], while maintaining the possibility of co-transformation. Similar results were displayed by Karas and co-workers [48], who through conjugative plasmid delivery via *E. coli*, were able to stably express plasmid vectors as episomes (also not integrated in the recipient genome) in the nucleus of *P. tricornutum* and *T. pseudonana* at high transformation efficiencies (~4.0 $\cdot$ 10^{-4} cells in *P. tricornutum*). Furthermore, Sharma and co-workers [47] recently showcased that the episomal vector may be lost over time as a result of removing the selection pressure, thus providing scientists with the ability to temporarily express a desired effect before covering their tracks. With a seemingly robust choice of transformation systems, molecular genetics in diatoms is finally picking up the pace. Genome editing using methodologies with meganucleases [49] and TALENs [49,50], and more recently CRISPR-Cas9 [51] are now enabled.

1.2.3 CRISPR as a bacterial immune system

The Clustered, Regularly Interspaced, Short Palindromic Repeats (CRISPR) system has been embraced for its utility as a tool in molecular genetics. It was first discovered in the *E. coli* genome in 1987 [52], but the first hypotheses of its role in the bacterial immune system were not published until 2005 [53,54]. The onset of the so-called "CRISPR craze" came when Jinek and co-workers [55] showed that the Cas9

enzyme could be programmed by a chimeric RNA to cut at specific sites in the DNA *in vitro*. Since then, the scientific community has put much effort into refining its versatility for diverse applications – and to date (May 2019) there are over 5700 articles on PubMed with "CRISPR" in their titles (https://www.ncbi.nlm.nih.gov/pubmed?term=CRISPR%5BTitle%5D).

To understand its versatility, it is necessary to understand its evolution. It is now known that the CRISPR systems exist as an adaptive immune system that prevents bacteriophage and viral infections in bacteria and archaea [56]. In short, it has two main assets; first, its ability to effectively neutralize novel and familiar infections through cleavage of foreign nucleic acids. Second, its adaptive properties which allow short sequences of the viral genome to be incorporated and stored in the host-cell genome as part of a CRISPR array of protospacer sequences [53,54].

The most extensively studied CRISPR-Cas system is that of *Streptococcus pyogenes;* a bacteria of which genus is responsible for high rates of human morbidity and mortality worldwide [57]. Ironically, this disease-borne bacteria is now providing new possibilities in medicine and genetics, as the endonuclease of *S. pyogenes* (*Sp*Cas9) has been used as a molecular scissor in a wide range of cell lines including humans [58], mice [59], monkeys [60], yeast [61], plants [62], algae [51] and bacteria [63]. The increasing number of CRISPR variants are currently partitioned into two distinct classes and are thereon divided into 6 types [64]. In the class 2, type II system of *S. pyogenes*, a large single-peptide protein dubbed Cas9 (*Sp*Cas9) is accountable for cleavage of foreign nucleic acids. Its endonuclease activity is granted by two distinct domains; *RuvC* and *HNH*, each of which cleaves one strand of DNA [55]. However, the Cas9 protein itself is un-specific and requires the help of two separate RNA molecules to allocate and bind to a particular dsDNA sequence [65]. Transcription of the CRISPR array consisting of so-called protospacer sequences gives rise to CRISPR RNA (crRNA). While these are essentially the RNA products of sequences acquired during the adaptive immunity response, their function lies in providing affinity to the specific piece of DNA [66]. The second RNA molecule – trans-activating CRISPR RNAs (tracrRNA) combine with the sequence-specific crRNA through base-pairing, and links it up to the generalist Cas9 protein – forming an active ribonucleoprotein complex capable of cleaving dsDNA at specific sites [55]. Once activated, the ribonucleoprotein complex stochastically searches the DNA for sites matching its Protospacer Adjacent Motif (PAM) preference [67]. While some CRISPR systems require detection of large PAMs, *Sp*Cas9 contains a C-terminal PAM-interaction domain that recognizes a short three-nucleotide motif comprised of 5'-NGG-3' (in which N can be any nucleotide) [55,68,69]. If the Cas9-complex finds a matching DNA sequence with the appropriate PAM, the DNA melts into a complementary and a non-complementary DNA single-strand (ssDNA), ensuring that the target-specific crRNA may hybridize with the complementary strand [67]. Given that the heteroduplex crRNA:DNA are in fact an appropriate match, HNH and RuvC domains each cleaves one strand of DNA at the third base upstream of the PAM, leaving a blunt-end double-stranded break (DSB) [70].

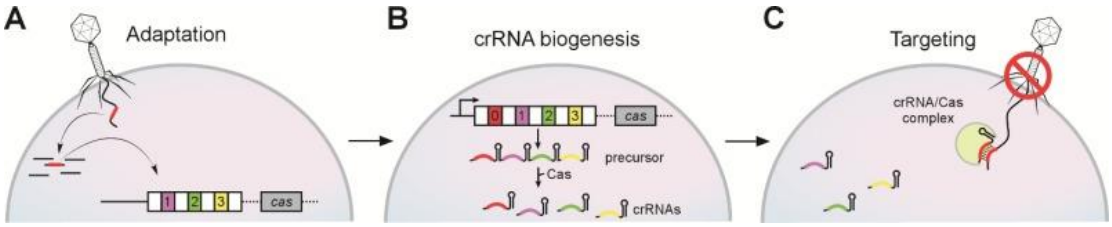


Fig. 1.2: CRISPR-Cas as an adaptive immune system. A) Immunity is acquired by adaptation as a sequence of the invading viral genome is incorporated into the host cell genome. B) Precursor CRISPR RNAs (crRNA) are transcribed from a storage of previously obtained CRISPR sequences (CRISPR loci). C) When under attack, crRNAs prime the Cas endonuclease, forming a complex capable of locating and neutralizing the infectious genome. Adapted from Marraffini [63].

1.2.4 The basics of microalgal genome editing using CRISPR-Cas9

Whereas the type II CRISPR-Cas9 system requires several components in nature, it has been streamlined to only require two components in the laboratory; The Cas9 endonuclease, and a chimeric substitution for the crRNA:tracrRNA duplex dubbed single-guide RNA (sgRNA), or simply guide RNA (gRNA) [55]. For transparency, we will hereon refer to the chimeric substitution as gRNA. In the microalgal genome editing methodology proposed by Nymark et al., [51], which was also adopted for this experimental work, a Cas9:gRNA-conferring plasmid is co-transformed together with a selection vector via biolistic transformation, leading to stable integration of both as part of the genomic DNA. The active Cas9:gRNA ribonucleoprotein complex continue to screen for viable target sites along the genome, starting with PAM authorization. Upon binding to a sequence of high affinity (i.e. a complementary site), the Cas9 proceeds with inserting nicks to both strands of DNA, leaving a DSB. Meanwhile, resistance to zeocin as conferred by the co-transformed selection vector pAF6 (Appendix B, Suppl. Fig. 3) allows for selection of transformants. Transcription of Cas9 is regulated by a *LHCF2* promoter and a *LHCF1* terminator, and the gRNA is controlled by U6 promoter and terminator. While the Cas9 protein itself is a generalist and does not require modification to target new sites, the systems targeting-properties are programmable by ligation of novel gRNA adapter sequences into the gRNA scaffold of the plasmid. These are variable 18-20 bp sequences, easily designed *in silico* with access to *P. tricornutum*'s genome sequence. Of course, the gRNA needs to be designed as to complement the PAM site preference of *Sp*Cas9, which, to some extent – limits the possibilities. However, since NGG is repeated throughout the genome, this is not a big challenge. Accordingly, targeting new genes may be easily accomplished by a methodology of design, manufacturing of adapter, followed by ligation into a Cas9 vector.

1.2.5 DSB repair

An essential characteristic of virtually all living organisms is the ability to maintain DNA integrity with repair mechanisms. In eukaryotes, DSBs are repaired by either of two major repair mechanisms (Fig.

1.3); Homology-Directed Repair (HDR) and Non-Homologous End-Joining (NHEJ) [71]. HDR utilizes a DNA donor template flanked by sequences that complement those flanking the cut site. While HDR may be provoked by providing the repair reaction with adequate exogenous templates, HDR also occur without influence from the experimenter, for instance by use of an intact sister chromatid as template [72]. The resulting change is a precise insert, and HDR may therefore be used to introduce specific properties or targeted replacements [73]. However, the dominating repair mechanism – given that HDR is not provoked – is NHEJ; a more error-prone pathway in which the cell attempts at repairing the DNA into a shape that allows for fusing the two threads together [74]. The cell does so by incorporating or removing nucleotides in a frantic manner, resulting in spontaneous insertion or deletion mutations (indels). Although large indels may be most critical, also small changes can alter the functionality of a target gene. So-called frameshift mutations are indels that shift the triplet code so that the transcribed product becomes truncated (when a premature stop-codon terminates transcription downstream of the indel) or non-functional/"knocked out" (KO) [1]. Subsequently, while an indel of 1 or 2 bases will be able to knock out the gene as the reading frame becomes shifted, an indel of 3 bases may not provide any functional change at all, since the product mRNA retains the correct code plus/minus a triplet. Thus, it can be generalized that indels who are non-dividable by 3 produce frameshifts, while indels dividable by 3 do not. However, if an indel becomes sufficiently large (i.e. if a large enough number of bases are removed or inserted), a gene may still be perturbed enough to partially or fully lose functionality, regardless of a possible in-frame triplet code. Thus, since at least 2/3 indels provide frameshift, genome editing with NHEJ-mediated repair is perhaps especially potent when the preferred outcome is a non-functional (knock-out/KO) gene-product [71]. When targeting several genetic loci, intentionally by multiplexing, or from Cas9:gRNA promiscuity, more peculiar effects may be seen. In these cases, NHEJ repair by ligation may cause large deletions or inversions (Fig. 1.3). Additionally, occurring DSBs on different alleles may cause the alleles to heterologous recombine [71].

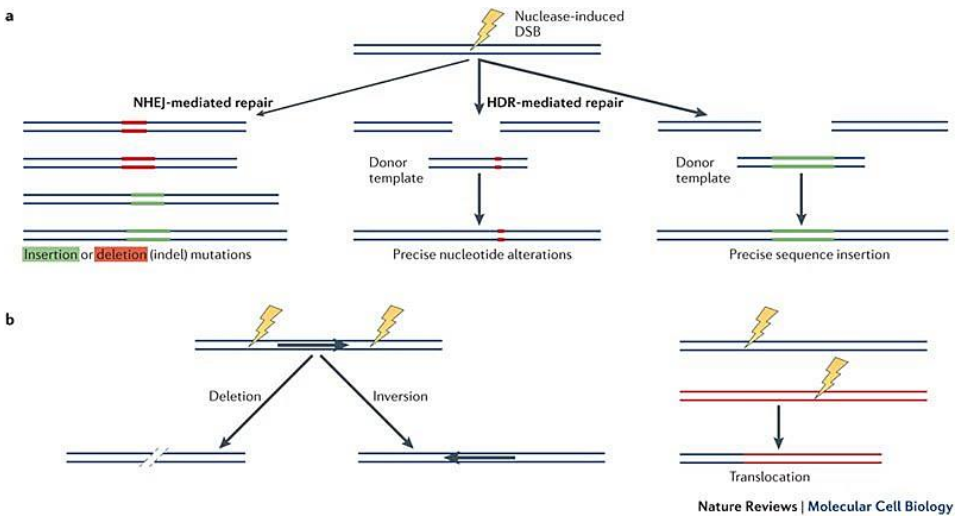


Fig. 1.3: The basal mechanisms of various repair mechanisms following a DSB. In eukaryotes, DNA integrity is restored through two main mechanisms: a) Non-Homologous End-Joining (NHEJ) and Homology-

Directed Repair (HDR). HDR occurs whenever there is a donor template available, and specific inserts may be facilitated by supplying a preferred sequence with ends homologous to those flanking the DSB. With no convenient template available, integrity is restored by frantically removing or inserting bases until the two ends may be re-joined (NHEJ). b) When multiple DSBs occur within the same time-space, additional effects may be achieved, such as the inversion, deletion or translocation of DNA. Adapted from Sander and Joung [71].

1.2.6 Cas9-mediated genome-editing is promiscuous

So far, one of the major drawbacks encountered with CRISPR-Cas9 has been its inclination to generate mutations at sites closely resembling that of its intended target. These so-called off-target effects occur because the Cas9 endonuclease can tolerate imperfect nucleotide matches in the heteroduplex linkage between gRNA and DNA. The extent of promiscuity with Cas9 has become a concern in virtually all applications of genome editing since a tolerance of four [75] or even five [76] mismatched nucleotides entails many eligible off-target sites in a large genome. As a result, several protocols and software tools have been developed to allow *a priori* prediction of off-target sites based on complementarity of genomic DNA and gRNA, and/or similarity between genomic DNA and target sequence [76–78]. Moreover, initial studies have indicated that the potential for DSB is somewhat influenced by the position of mismatched nucleotides in relation to the PAM [55,79]. Thence, it is now conceived that cleavage is less probable when mismatched nucleotides are located in the 8-12 first nucleotides 5' to the PAM (seed region, Fig. 1.4a), while multiple mismatches may be tolerated given their location in the distal non-seed region (Fig. 1.4b) [69]. Therefore, it is highly beneficial to design gRNA carefully with respect to PAMs, potential off-target sites and their seed and non-seed mismatches.

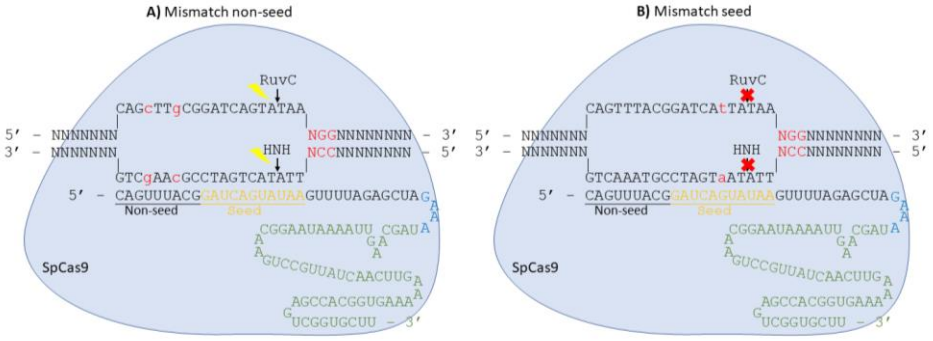


Fig. 1.4: An overview of the Cas9 ribonucleoprotein complex and the relation of gRNA sequence specificity for off-target mutagenic promiscuity. A) Mismatches in the non-seed region (PAM-distal) may not prohibit DSB. Conversely, mismatches in the seed-region (PAM-proximal, B) more often prohibits DSB. Blue nucleotides denote the artificial nucleotide tetraloop that combines crRNA and tracrRNA (green) to a chimeric gRNA. Note that gRNA is not responsible for PAM authorization. Modelled after Ma et al., [80].

1.2.7 Increasing the fidelity of Cas9

Fu and co-workers [81] theorized that the endonuclease activity of Cas9 is energy-driven, and that the complex interaction of Cas9:gRNA:DNA hold excess energy to what is required for cleavage of fully

complementary DNA sites. Moreover, structural studies have indicated the presence of several residues in the Cas9 peptide that interact with the phosphate backbone of DNA through direct hydrogen-bonds, amongst them N497, R661, Q695 and Q926 [68,82]. Fuelled by this information, Kleinstiver and co-workers [83] modified *Sp*Cas9 by replacing said DNA-contacting residues with less reactive Alanine residues to thereby lower the non-specific binding affinity of Cas9 to DNA. Using a human cell line, the modified High-Fidelity 1 (HF1) and a wildtype (WT) *Sp*Cas9 enzyme primed with 37 different gRNA, *in vivo* mutagenesis was detected through an assay- and targeted amplicon sequencing-based methodology. No off-target mutations were detected in cell lines transfected with the HF1 enzyme, and it retained >70% of the on-target efficiency recorded using WT *Sp*Cas9 [83]. Notably, this research provided a general strategy to improve precision of genome editing in other respective model species. Recently, Idokoh-Akoh and co-workers [84] made the same modifications in an SpCas9 variant intended for genome editing in chickens, and found that while SpCas9-HF1 increases the on-target editing efficiency, it also increases the frequency of single-allele mutations.

1.3 Photosynthesis and light-utilization

1.3.1 Photosynthesis – from light to organic molecules

Oxygenic photosynthesis governs the process that exploits photosynthetically available radiation (PAR, visible light ~[400, 700] nm) and water to generate organic carbon molecules from inorganic carbon dioxide (Fig. 1.5) [85,86]. The act of photosynthesis can roughly be divided into two types of reactions [1]. In the light reactions, quanta of light (photons) are harvested and utilized to drive electrons through an electron-transport chain (ETC) consisting of thylakoidal protein-complexes and intermediary electron acceptors. The molecules that makes photonic light-energy accessible to photosynthesis are pigments. Pigments are anchored to the thylakoid by a set of proteins that comprise the so-called light-harvesting complexes (LHCs) or "antennae" [87]. The purpose of these structures is to widen the area of absorption so that more light may be provided for photosynthesis [88]. In green algae and plants, these structures are further referred to as light-harvesting complex I or II (LHCI or LHCII) depending on which reaction centre (RC) they associate with (RCI or RCII, respectively) [89], and RC cores are composed of several proteins, including the D1, D2, Cyt b559 (both α- and β-subunits), CP43 and CP47 [90]. Collectively, LHCs and RCs comprise the photosystems I and II (PSI/PSII, Fig. 1.5). Light-harvesting in diatoms is highly similar, with the same core composition of RCs [90]. However, they do not seem to contain the strict functional organization that plants and green algae do. For instance, diatom LHCs are distinguished from plant LHCs on basis of what pigments they bind, and are therefore dubbed Fucoxanthin and Chl *a*/*c* protein (FCP) complexes. Moreover, no PSII:FCP super-complexes have been isolated from diatoms so far [91]. The FCP complexes of *P. tricornutum* are also expected to deliver energy to both PSI and PSII, and do not necessarily locate within the photosystems, as diatom FCP complexes have been postulated to be both peripheral and proximal to the RCs [91].

Nevertheless, photons harvested by the FCP complexes are transferred between pigments by resonance energy transfer (RET), before they are eventually delivered at the RCs, each of which contains a specialized Chl *a* dimer dubbed P700 (RCI) and P680 (RCII). P700 and P680 become excited whenever photons are received – manifested by an electron in the Chl *a* molecule jumping to a higher energy state [92]. For a moment, the centre goes from open and ready to accept photons, to being closed – ready to relay electrons into the ETC. Shortly after, the excited P680 transfers its excited electron to a pool of the primary electron acceptors Quinone A (Q_A), leading to P680 being short of an electron ($P680^+$) [89]. However, a neat, fine-tuned process solves this by replenishing the lost electron with electrons harvested from water. Namely from the manganese (Mn) water splitting enzyme of PSII that – upon P680 excitation – facilitate oxidation of water to molecular oxygen, two protons and two electrons [92]. When the Chl *a* dimer of P680 is replenished, the reaction centre returns to the open, relaxed state, once again ready to accept novel photons.

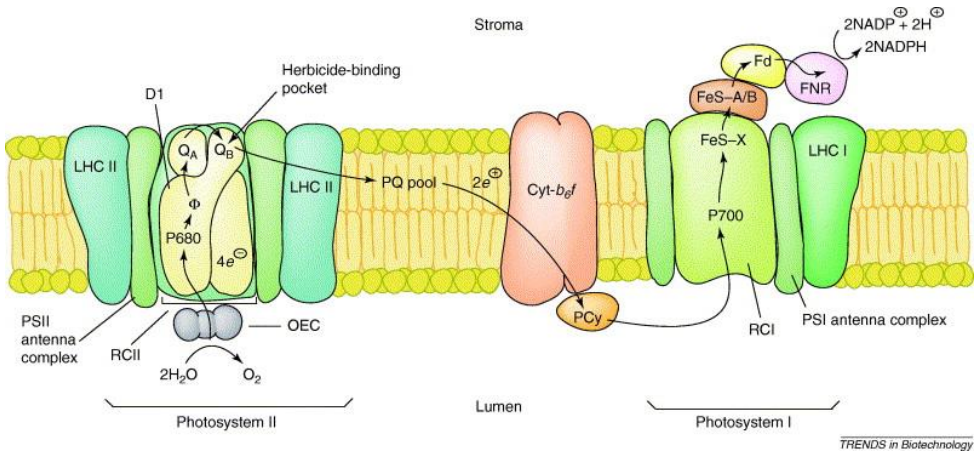


Fig. 1.5: A schematic presentation of the electron transport chain (ETC) in green algae and plants, including the structural organization of reaction centres (RCI/RCII), light-harvesting complexes (LHCI/II), water splitting enzyme (grey), and specialized Chl *a* dimers (P680/P700). Electrons flow from the water splitting enzyme, to P680 and through the primary electron acceptor (Quinone A (QA), Quinone B (QB)), the secondary electron acceptor Plastoquinone (PQ), via the Cytochrome b6f-complex (Cyt-b6f), Plastocyanine (PCy), the PSI reaction centre (RCI), FeS-clusters (FeS-X, FeS-A/B), Ferredoxin (Fd) and is eventually utilized by the Ferredoxin NADP+ reductase (FNR) to generate NADPH from NADP+ and H+. Note that the components which facilitate ATP synthesis are not presented here. Neither is the trans-thylakoidal pumping of protons from stroma to lumen. Moreover, the tight structural organization with LHCI and LHCII is not representative for diatoms, and Plastocyanine (PCy) is replaced by Cytochrome c6. Adapted from Giardi and Pace [89].

Electron-transport through the thylakoid promotes additional pumping of protons from the stroma, across the thylakoid membrane, and into the lumen. The resulting abundance of luminal protons create an electrochemical proton gradient that drives synthesis of ATP from ADP and inorganic phosphate (P_i) [1]. Moreover, NADP+ serves as the final electron acceptor in the ETC, accepting two electrons and a stromal proton to form NADPH. The follow-up dark reactions, more commonly known as the Calvin-Bensson-Bassham (CBB) cycle depend not directly on light, but on the reduction power of ATP and NADPH molecules that stem from the light-reactions. The end product of the CBB cycle,

Glyceraldehyde 3-Phosphate (G3P) – is manufactured from CO_2 over three steps (carboxylation, reduction, regeneration), involving 11 participating enzymes and 13 distinct reactions [93]. Once manufactured, G3P may be used for sustenance and/or as a precursor for sugars, fats and amino acids [1].

1.3.2 Diatom pigments

Evolution has given rise to a wide diversity in pigment chemistry, especially within the different algal phylogenies. Overall, there are three chemically and functionally distinct groups of pigments in the world of primary producers; chlorophylls, carotenoids and phycobilins [94]. In diatoms, only two of them apply; the chlorophylls (Chls), including Chl *a*, Chl c_1, c_2 and c_3, and the carotenoids, whom are known to perform roles in both light-harvesting (fucoxanthin – Fuco), and in photoprotection (β-carotene, diadinoxanthin – Ddx, diatoxanthin – Dtx).

1.3.2.1 Light-harvesting pigments of *P. tricornutum*

The primary role of the light-harvesting pigments is to absorb photons of wavelengths within the PAR spectrum. In *P. tricornutum* these include Chls *a*, c_1, c_2 and the xanthophyll carotenoid Fuco [94]. Crude structural differences because of chemistry or protein-related interactions result in distinct optical properties such as in the signatures of their absorbance (Fig. 1.6). Optical differences are also prevalent in pigment extracts (*in vitro*, Fig. 1.6a) and in living cells (*in vivo*, Fig. 1.6b) [95].

Chl *a* is the predominant pigment and is virtually found in all photosynthetic eukaryotes. It serves two main purposes, whereas one of them is to constitute the P680 and P700 Chl *a* dimers that ensure photochemical energy conversion. Additionally, Chl *a* serves in the FCP complexes as a major absorber within the spectral wavelength intervals (nm) of blue ([400-470]) and red ([620-700]) light [94]. Engaged as an accessory pigment, the purpose of Chl *c* is in diatoms somewhat similar to that of Chl *b* in green algae and plants (Viridiplantae) [94]. Chl *c* constist of a phytyl-lacking porfyrin-ring further distinguished into isoforms by the residues bound (Fig. 1.6c). While several isoforms of Chl *c* have been isolated from diatoms, mainly Chl c_1 and c_2 are prevalent in *P. tricornutum* [94]. Like Chl *a*, Chl *c* display a strong absorption peak in the blue-light bands of the spectrum with absorption maxima at 470 nm. However, this peak is slightly shifted towards the red [95], and in contrast, Chl *c* does not contribute to red light absorption (Fig. 1.6a and b).

Diatoms are well known for their utilization of carotenoids in light-harvesting, manifested by a high ratio of carotenoid to Chl in FCPs (4:5) compared to plant LHC proteins (4:14) [96,97]. The predominant carotenoid with this function is Fuco. The absorbance of Fuco overlaps with that of Chl *a* and *c* in the blue parts of the PAR-spectrum, but is more pronounced for light of green wavelengths ([470-540] nm). Moreover, the broad bandwidth of Fuco absorbance is caused by so-called bathochromic shifts – effectively causing Fuco molecules to absorb at different wavelengths when

bound to different FCP monomers. Thus, they may be further partitioned by the wavelength of light they harvest as $Fuco_{BLUE}$ (abs_{max} ~450 nm), $Fuco_{GREEN}$ (abs_{max} ~490) and $Fuco_{RED}$ (abs_{max} ~510 nm, Fig. 1.6b) [95].

1.3.2.2 Photoprotective pigments

All planktonic algae must be able to cope with variable light-conditions. With a sudden increase in irradiance arises the complications of photoinhibition; a state in which the components of the photosynthetic machinery become damaged. To prevent this, *P. tricornutum* contains Diadino- (Ddx) and Diatoxanthin (Dtx), pigments whose role is to remove excess energy through non-photochemical quenching of Chl fluorescence (NPQ, discussed below). These pigments absorb strongly at wavelengths between 400-500 nm (Fig. 1.6a) [94]. Additionally, diatoms have the carotenoid xanthophylls whom are utilized in photoprotection in green algae and higher plants (violaxanthin – Vtx, zeaxanthin – Zx and antheraxanthin – Ax), but these seem less important for the photoprotective response in diatoms, and may only accumulate over long-term high-light (HL) exposures [94].

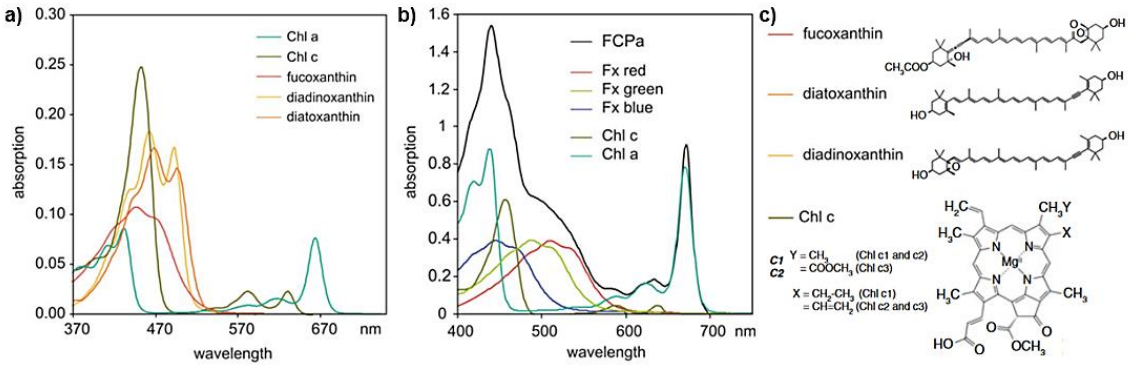


Fig. 1.6: Spectral absorbance and chemical composition of pigments in the marine diatom *C. meneghiniana*. Absorbance spectra were recorded for a) *in vitro* pigment-extract (acetone, 1 mM), and b) "*in vivo*" FCPa complexes. Chemical composition of Fuco, Dtx, Ddx and isomers of Chl c are presented in c). Note that both *C. meneghiniana* and *P. tricornutum* utilize the same pigments. Adapted from [95].

1.3.3 FCP proteins

The pigments are anchored to the thylakoid membrane by a superfamily of nucleus-encoded FCP proteins [98]. While their presence and importance in light-harvesting has been known for quite some time [99], it is only recently we have begun deciphering their properties. By solving the x-ray crystal structures of *P. tricornutum* FCPs, it was shown that each FCP monomer binds seven Chl *a*, seven Fuco, two Chl *c* and one Ddx [100]. The amino acid sequence and structure of FCPs show that they exhibit three trans-membrane α-helices, with C- and N-terminals in the lumen and stroma respectively [100,101], and that they enter basic trimeric complexes [95,102]. Moreover, the FCP complexes they form are further partitioned by the number (n) of FCP monomers they are comprised of as either FCPa

(trimers, n = 3) or FCPb; in which three trimers comprise a nonamer (n = 9) [103], or even as FCPo (hexamers, n = 6): in which FCP complexes consisted of six monomers or two trimers [104].

Although the FCPs of *P. tricornutum* were early found to exhibit highly similar sequences [105], of which has been later confirmed via whole genome sequencing [27], current consensus is that some FCPs perform tasks unrelated to light-harvesting, and they are thus sub-grouped by function into three [106], or potentially four [107] major groups. The LHCFs serves directly in light-harvesting as the major pigment scaffolds in the peripheral antennas directed towards the PSII reaction centre [106,108], and constitutes the largest LHC group with 17 genes currently annotated [27]. Previous global transcriptomic profiling by microarray have found that LHCFs are down-regulated upon exposure to HL, strongly supporting that LHCFs serve directly in light-harvesting [108]. LHCFs are moreover relatively small, consist of approximately ~200 amino acids amounting to a molecular weight of approximately 21 kDa (UniProt TrEMBL, https://www.uniprot.org/uniprot/?query=LHCF+AND+Phaeodactylum+tricornutum&sort=score).

The LHCX proteins are believed to play a role in photoprotection because they show high similarity to PsbS-homologues found in the green algae model species *Chlamydomonas reinhardtii* [109] and in plants [110]. Thus, like the PsbS proteins of plants and green algae, LHCX proteins are believed to be involved in activating mechanisms that ensure dissipation of surplus light [111,112]. Only 4 copies of LHCXs are found in the *P. tricornutum* genome [27]. The LHCRs comprise the second largest group with 14 genes, and are thought to function as proximal antenna proteins associated with PSI, due to shared similarity with rhodophyte proximal PSI antennas [109]. However, recent queries involving phylogenetic analysis of all antenna proteins in *P. tricornutum*, proposed a division of the LHCRs in two; LHCRI, which shares a role in light-harvesting, and LHCRII, of which role lies closer to that of LHCX – namely in photoprotection [107]. Yet another group, the LHCYs, which consist of previously unclassified FCP proteins is proposed to also act in light-harvesting [107].

1.3.4 Transport and insertion of FCPs

Since all LHC-proteins are encoded in the nuclear genome, they must be transported into the chloroplasts before they may be integrated in the thylakoid membrane. In green algae, translocation of LHC proteins into the thylakoid membrane are conducted over three distinct steps (Figure 1.5) [113]. First, peptides pass the outer and inner envelope membranes by interacting with the translocator of the outer chloroplast membrane (TOC) and translocator of the inner chloroplast membrane (TIC) complexes [1]. This requires detection of a chloroplast transit peptide (cTP) by a specific receptor domain associated with TOC, and subsequent GTP- and/or ATP-driven import [1]. LHCs are prevented from misfolding by chaperones in the chloroplast signal recognition particle (CpSRP) pathway and inserted by an Albino 3 insertase (ALB3.1/3.2) (Fig. 1.7).

While the processes of transport and insertion of thylakoidal proteins are relatively well known in green algae and plants, they remain somewhat enigmatic in diatoms. Stroma entry may be slightly more complicated in diatom chloroplasts that in contrast to green algae and higher plants consist of four double-layered membranes instead of two (Fig. 1.7a), and thus possibly several translocator complexes, each of which may require signal detection. Gruber et al. [114] demonstrated that diatom plastid proteins display two distinct N-terminal peptide sequences detrimental for such transport; a cTP that typically consist of 15-20 residues, immediately followed by a thylakoid transit peptide (tTP). Stromal entry culminates in cleavage off the cTP and exposure of a particular tTP motif starting with an aromatic amino acid such as phenylalanine (F), and in some cases tryptophan (W), tyrosine (Y) or leucine (L) [114]. The molecular role of this motif (ASA|FAP) is not yet known, but it is believed that the phenylalanine (F) is detrimental to either import or integration as it is a conserved property of plastid proteins in both diatoms and cryptophytes [114]. Similarly, the remaining mature peptide is recognized by proteins of the chloroplast signal recognition particle (CpSRP) pathway including chaperons that act to prevent misfolding [113], and integrated by a complex of proteins in which the Albino-3b insertase (Alb3b) is detrimental for insertion [115][115].

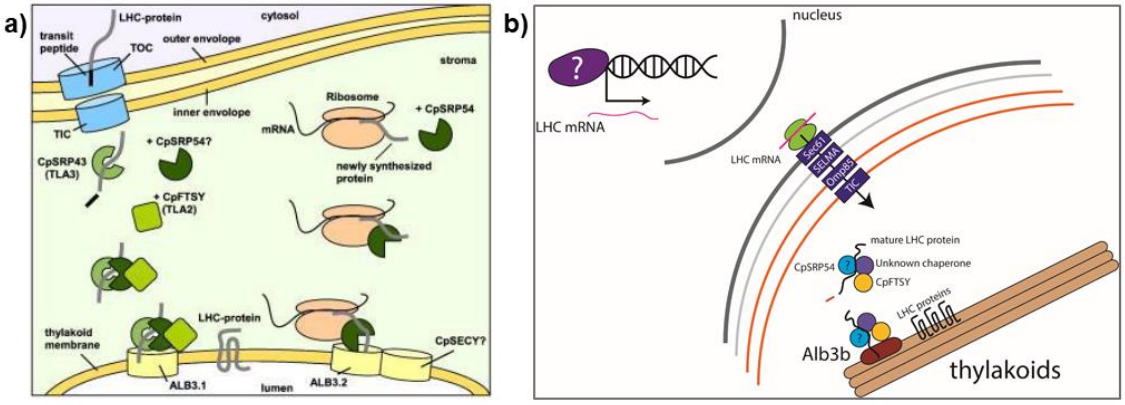


Fig. 1.7: Transport and integration of LHC proteins in green algae (a) and diatoms (b). Translocation and subsequent integration of LHC peptides requires a complex machinery and interpretation of atleast two signal-peptides in a) green algae. Translocation through the first two membranes/envelopes require detection of a chloroplast transit peptide (cTP) and subsequent interaction between TOC and TIC translocons. After import, cTP is cleaved off. Thylakoid membrane integration of protein requires cooperation between several stromal proteins in the CpSRP pathway and Albino insertases (ALB3.1/3.2) in the thylakoid membrane [113]. Modified from Kirst and Melis [113]. b) A similar but more complex translocation/integration machinery may be present in diatom chloroplast which exhibit additional membranes. The components that fascilitate this process are relatively unknown, but Alb3b is known to be detrimental for membrane insertion [115]. Modified from Nymark et al., [115].

1.3.5 The fates of absorbed light and variable fluorescence

Light absorption results in a single excited state Chl *a* (Es, ^{1}Chl*) that may dissipate its energy by returning to ground state (Gs, ^{1}Chl). Dissipation is achieved through either of the following pathways (Fig. 1.8a): its energy may be i) transferred towards RCs by resonance energy transfer and used to run

photochemical quenching (PQ), ii) dissipated as heat (NPQ) or iii) re-emitted as Chl fluorescence [116,117]. A last resort lies in iv) decay of $^{1}Chl^{*}$ through formation of triplet Chl ($^{3}Chl^{*}$), which is then quenched off by transferring its energy to molecular oxygen (O_2) forming the reactive oxygen species (ROS) singlet oxygen ($^{1}O_2^{*}$) [117]. In stable and possibly favourable conditions, 20-25 % of the absorbed light goes into photochemistry, 1-5 % is re-emitted as fluorescence, and the remaining 70-79 % are lost from formation of $^{3}Chl^{*}$ and heat dissipation [116,117]. However, the proportions of light going into these fates occur in equilibrial competition, such that an increase in the proportion of one lead to an inevitable proportional decrease in the others. Therefore, by measuring the Chl fluorescence, it is possible to gain information on the relative efficiency of the others (i.e. the proportion of absorbed light going into photosynthesis, heat or triplet Chl) [117]. In diatoms, Chl *a* fluorescence is predominantly emitted from PSII (usually ~95% of the total fluorescence), whereas the remaining fraction of 5% is emitted by the Chl *a* in PSI [118].

This is the foundation for variable fluorometries: methods that seeks to describe the photophysiology of cells based on the variable fluorescence of Chl *a*. So-called pulse-amplitude-modulated (PAM) fluorometers have become increasingly popular for this purpose. Because it is modulated, meaning that the light-source used to measure fluorescence is switched on and off at very high frequency, only fluorescence from Chls excited by the measuring light is detected, essentially allowing fluorescence yields to be tied to the intensity of illumination and without the interference of ambient light [117]. Moreover, re-emitted fluorescence can be distinguished by wavelength because the peak of the fluorescence spectrum is longer than that of the absorption. This allows for quantification of fluorescence alone by calibrating the detector to perceive light of wavelengths longer than that of the measuring light [116]. However, there are many pitfalls that relate to use of variable fluorescence to infer light utilization and effectivity. The 'light-doubling' technique [119] attempts at avoiding some of those pitfalls by measuring the efficiency of FL excitation when the contribution of photochemical quenching is transiently reduced to zero [116]. In practice, this is achieved by employing a high-intensity flash of light (known as a saturating flash) to close all RCs (i.e. all Q_A are reduced, unable to accept new electrons). Thus, the fluorescence intensity should amount to a value equivalent to that of a system lacking photochemical quenching (PQ) altogether, and thus be maximal (F_m). In comparison, minimal fluorescence (F_0) is achieved by incubation in darkness followed by fluorescence detection under measuring light. All RCs should be open given an adequate dark period, thus reaching the maximal probability of absorbed light going into PQ, and to subsequently yield the minimal fluorescence (F_0). By furthering the protocol with exposure to background light irradiation, one may assess the maximal (F_m') and minimal (F_o') and steady state fluorescence (F_t') yields in actinic light (′, Fig. 1.8b). Together, these parameters allow inferring on the photosynthetic efficiency and overall proportion of closed contra open RCs, usually presented as the effective (in actinic light, (F_m' - F_0') /

F_m´) or maximum (in darkness, (F_m - F_0) / F_m) quantum yields of PSII photochemistry (Φ_{PSII} and $\Phi_{PSIImax}$, respectively) [116].

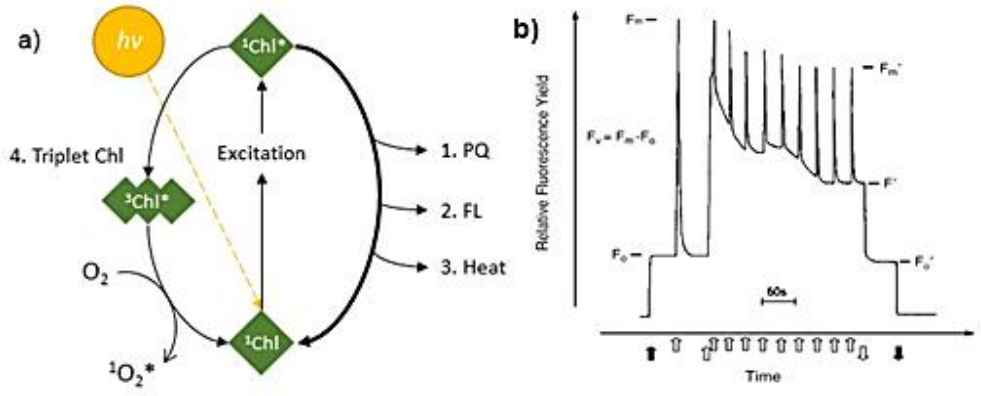


Fig. 1.8: The four fates of absorbed light (*hν* = photons, a), and a typical trace of fluorescence yields as obtained with a pulse-amplitude modulated (PAM) fluorometry protocol. a) Excited state Chl (^{1}Chl*) may return to ground state (^{1}Chl) through either of pathways involving 1. Photochemical quenching (PQ), 2. re-emission as fluorescence (FL), 3. dissipation of heat (Heat) or 4. decay of ^{1}Chl* by formation of triplet Chl (^{3}Chl*) and subsequent energy transfer to form singlet oxygen ($^1O_2^*$). Photon = *hν*. Modelled after [117]. b) A typical trace of fluorescence yields over time, when using a PAM light-induction protocol. Arrow colours distinguish between measuring light (black), saturating light-flash (grey) and actinic light (blank/white). Arrow direction signifies on (up) or off (down). Modified from Büchel and Wilhelm [120].

1.3.6 Non-photochemical quenching

Light-driven photosynthetic electron transport facilitates pumping of protons from the stroma and into the lumen. Whenever the cell absorb excessive amounts of light – the rate of electron transport may surpass that of what the CO_2-fixating machinery can process, leading to photoinhibition, photosystem inactivation and induced production of damaging reactive oxygen species (ROS) [94,117]. To circumvent the damaging effects of excessive absorption, diatoms employ NPQ processes. In theory it consists of three main components (qT, qI and qE), but the energy- and pH-dependent component (qE) has been shown to be of particular use to diatoms [121]. An integral part of the pH-dependent NPQ response is the xanthophyll cycle which involves de-epoxidation of Ddx to Dtx. The fundamental photoprotective action is executed by Dtx molecules through quenching of chlorophyll singlets (^{1}Chl*) and triplets (^{3}Chl*) thereby alleviating both energy load on RCs and production of ROS [122]. De-epoxidation is mediated by the Ddx de-epoxidase (DDE) enzyme that triggers upon detection of high-light driven ΔpH (high-proton in lumen, low-proton in stroma) [121]. Simultaneously, high-light driven ΔpH inhibits the reverse reaction mediated by Dtx epoxidase to a factor of 10 to 20 to that observed under low-light driven ΔpH [121,123]. Several methodologies are developed to tap into the relative photoprotective state of algal cells. At the metabolite level, HPLC-aided pigment analysis ensures

quantification of Dtx and Ddx. From this one can calculate the de-epoxidation state index (DES = Dtx / (Dtx + Ddx), indicating the relative level of Dtx to Ddx [124].

1.4 Industrial applications and limitations of diatoms

1.4.1 Where lies the potential of diatoms?

Diatoms are of interest in many human ventures and industries. Their intricate silicate frustules are influencing how nanotechnologists and biochemists design drug delivery systems. As such, silica nanoparticles have been successfully used as drug vehicles to deliver cancer therapy to both liver-tumors [125] and prostate cancer cell lines [126]. Since most of the currently estimated number of 200 000 species of diatoms build their frustules of silicate, a wealth of inspiration is yet to discover. Another application lies in phytoremediation. Diatoms produce oligomers of glutathione $(\gamma\text{-Glu-Cys})_n\text{-Gly}$, commonly referred to as phytochelatins (PCs), that aid in sequestering of heavy metals essential to their metabolism [127]. Actually, 10% of their mass may be owed to sequestered metals alone [128]. Thus, PCs may be applicable in detoxification of heavy metals which may otherwise accumulate to toxic levels. If phytoremediation is coupled with an otherwise obsolete nutrient source such as waste-water, diatoms may provide efficient means of N, P and metal recovery and/or production of profitable compounds from scraps [127].

The perhaps most promising application lies in aquaculture – an industry that require large quantities of marine fatty acids for fish-feed. Diatoms produce the long chain polyunsaturated fatty acids (LC-PUFAs) eicosapentaenoic acid (EPA, $20{:}5\Delta^{5,8,11,14,17}$, n-3) and docosahexaenoic acid (DHA, $22{:}6\Delta^{4,7,10,13,16,19}$, n-3). These are of essential nutritional value for growth and development of juveniles [129] and for cardiovascular health maintenance in adults [130], making them important constituents of fish targeted for human consumption. *P. tricornutum* is of particular interest in this regard, because it has a favourable fatty acid composition where 30% of the total fatty acid storage may be purposed for LC-PUFAs such as EPA and DHA [131]. Simultaneously, because omega-3 fatty acids are currently harvested from biologically and economically important fish-stocks such as the Peruvian anchovy, algal biomass may provide a low-trophic and cost-efficient alternative while simultaneously providing these fish-stocks relief.

1.4.2 Light-access as a limiting factor

One of the limiting factors to economically favourable, mass-scale cultivation of algae lies in their adaptations to a natural habitat with fluctuating light-conditions. In low-intensity light (LL), they tend to expand their absorption surface area by building up LHCs in efforts to stay competitive. Conversely, in high-intensity light, they do the opposite by downsizing the FCP:pigment complexes so as to avoid excessive absorption and subsequent photoinhibition. While these mechanisms may be advantageous for cells in fluctuating environments where the light-source is sufficient one day and virtually absent

the next, it imposes challenges to ventures seeking to cultivate them for their compounds. Most important is the notion that obtaining high levels of valuable compounds is dependent on reaching high biomass per volume, and that it is necessary to grow the culture into high density to do so. With every cell division, the culture becomes denser, and the cells will start shading each other in their attempts at competing for the light source. This is when the LL-acclimation kicks in, and the cells will start expanding their antennas to remain competitive, causing increasingly rapid light-attenuation, shorter light-path, and an ultimately smaller photosynthetically active culture volume where cells do not obtain the sufficient quanta of light for continued growth (Fig. 1.9). Simultaneously, only ~20% of the light absorbed is used in photosynthesis, whereas the remaining 80% is dissipated via NPQ [132]. Moreover, cells near the surface of a photobioreactor absorb quanta of light that far exceeds the rate of CO_2 fixation, leading to induced production of ROS, and subsequently losses in photosynthetic efficiency [117]. To circumvent these hurdles, several attempts have been made at creating algal cell lines in which the antennas are smaller, resulting in so-called truncated light-harvesting antenna (TLA) mutants.

1.4.3 Overcoming light-limitation by using TLA mutants

TLA mutants coax with being less prone to excessive absorption, thus avoiding the damaging effects of high irradiation while simultaneously maximizing the solar-to-energy conversion and photosynthetic production efficiencies in high-density cultures [132]. Kirst and co-workers [133] grafted a truncated TLA mutant of the green algae *Chlamydomonas reinhardtii*. By deletion of the *CpSRP43*-gene, they managed to perturb the CpSRP-component pathway that ensures proper folding and insertion of LHC-proteins in the thylakoidal membrane. The TLA mutant, which retained about 40% of the light-harvesting antennae surface area to that of the wildtype, displayed a distinctively increased light-intensity to reach saturation (E_k) of photosynthesis, and improved solar energy conversion efficiency [133]. A similar approach was executed by Shin et al., [134] whom by point-mutation perturbed a gene homologous to the CpSRP43-component in the green algae *Chlorella vulgaris*, effectively leading to reduced LHC-size and Chl *a* content. Kirst et al., [135] have also displayed that the TLA concept is transferable to cyanobacteria. By knocking out the CPC genes, of which function is to anchor phycocyanin pigments to the light-harvesting antennas, the group managed to obtain a mutant of *Synechocystis sp.* with doubled E_k, increased transmittance and increased biomass accumulation (1.57 times that of WT) [135].

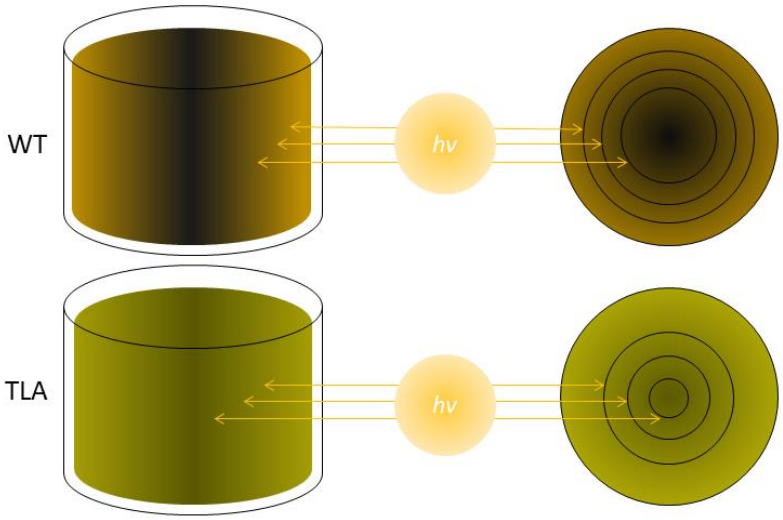


Fig. 1.9: The concept of increased light-penetration in diatom cultures with truncated light-harvesting antennas (TLA). WT-size antennas entail rapid attenuation of photons (*hv*), causing shorter light-path and subsequently loss of depth-penetration in a cylindrical culture. TLA mutants overcome these effects by having each cell absorb less, and transmit more, essentially freeing up more volume for continued phototrophic growth. Modelled after [136].

2 Objectives

Three main objectives were outlined for the thesis.

1. Investigate if a modified *Sp*Cas9 enzyme dubbed hifiCas9 (high-fidelity Cas9) avoids off-target mutagenesis when applied in genome editing of diatoms.
2. Characterize mutants with multiple *lhcf* knockouts and determine if they exhibit typical TLA traits.
3. Perform a pilot experiment to determine if TLA mutants could outperform WT cells when grown to high cell densities in photobioreactors.

3 Materials and methods

3.1 Strain and routine growth conditions

Phaeodactylum tricornutum (CCMP2561) were grown axenically in F/2 liquid and agar medium made from filtered and autoclaved seawater (Appendix A1). Transformed algae were grown selectively by adding the antibiotic Zeocin to the growth medium (Appendix A1). To ensure stable conditions between experiments, wildtype (WT) and mutant cell cultures were kept in a growth chamber at 23°C. The light-climate consisted of cool white light at an intensity of between 50 and 250 µmol photons m^{-2} s^{-1}, and a 16:8h day:night cycle. Generally, liquid cultures were kept without any mechanical shaking if not stated otherwise.

3.2 Mutagenesis

3.2.1 Plasmid Constructs

Three plasmids were used to generate mutants of *P. tricornutum*; i) the SpCas9 enzyme adapted for genome editing of diatoms (pKS diaCas9_sgRNA, Addgene ID: 74923, Appendix B, Suppl. Fig. 1), ii) the modified version of pKS diaCas9_sgRNA dubbed High-Fidelity Cas9 (hifiCas9, Suppl. Fig. 2), and iii) a selection vector (pAF6, Suppl. Fig. 3) – conferring resistance to the antibiotic zeocin. All plasmids were previously constructed by employees in the CMBG group. Fundamentally, both Cas9 plasmid constructs are identical, except for 4 specific alterations to the Cas9-encoding gene of the hifiCas9 sequence (Fig. 3.1). Namely, the triplets encoding Asparagine (N, 540), Arginine (R, 704) and two Glutamine (Q, 738 and 969) residues were manipulated to encode Alanine (A).

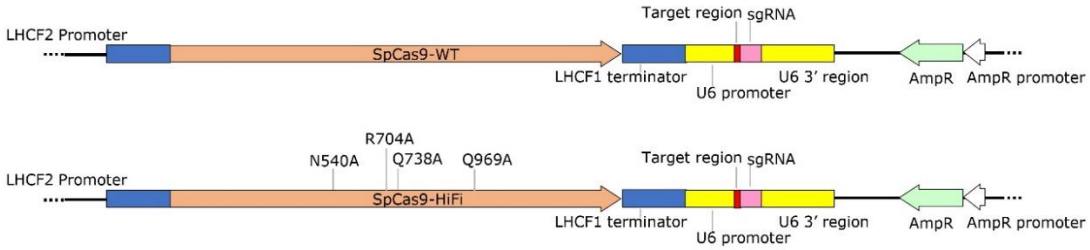


Fig. 3.1: The difference between dia- (SpCas9-WT) and hifiCas9 (SpCas9-HiFi). Four amino acids were replaced with Alanine (N540A, R704A, Q738A and Q969A). Transcription of the *Cas9* gene was under control of the *LHCF2* promoter and terminator. The chimeric gRNA (sgRNA) was under control of the *U6* promoter. Resistance to ampicillin enabled efficient amplification by cloning and was conferred by *AmpR* under control of the *AmpR* promoter. For a complete overview of plasmid structures, see Appendix B. Figure courtesy of Amit Sharma.

To test whether these modifications led to increased fidelity, both enzymes were accompanied by the same gRNA target-sequences (mPAM1, mPAM2). Both mPAM1 and mPAM2 were designed *in silico* by Associate Professor Per Winge. These target a subset of the LHCF gene family, but with varying

affinity (Table 3.1). So-called on-target, or fully complimented target sites of mPAM1 includes *LHCF1, 3* and *4*. Off-targets include *LHCF2, 5* and *11*. The same targeting pattern was achieved using mPAM2, even though the target sequence and location of mismatched nucleotides within are spatially distinct. In the experimental work carried out by the writer, mutagenesis was achieved using only mPAM1 gRNA. Data from cell lines derived from mutagenesis with the second gRNA (mPAM2) was obtained from co-supervisor Amit Sharma to increase the basis from which to assert the precision and efficiencies of hifiCas9. Thus, both enzymes were tested in combinations with both gRNA over a total of four variations. The 5'-3' DNA sequence of both gRNAs and their on- and off-targets are presented in Table 3.1. All on- and off-target sites were identified in the Phatr3 genome by Assoc. prof. Per Winge using a string-based pearl-script search algorithm (local program).

Table 3.1: Overview of DNA sequences targeted by mPAM1 and mPAM2 gRNAs. Mismatched nucleotides are labelled red and summed up (m #). PAM-sites (NGG) are underlined. Target sequences are described with their chromosomal location (Chr.), gene names and Accession IDs corresponding to *P. tricornutum* annotation *3* – Phatr3, available at Ensembl (http://protists.ensembl.org/Phaeodactylum_tricornutum/Info/Index).

	DNA sequence (5' to 3')	m #	Chr.	Gene	Accession ID (Phatr3)
mPAM1 targets	T**TC**TC**G**CAACGTAACGGA**A**AA**A**GG	5	9	Pred. Protein	*Phatr3_J12752*
	TGA**GTA**CAACGT**C**ACGGAGA**G**GG	5	5	Pred. Protein	*Phatr3_J44831*
	CCA**C**CTCAACGTAAC**C**GAGACGG	4	4	Exon: *CDKC1*	*Phatr3_EG02630*
	TGA**GT**TCAACGTAACG**A**AGACGG	3	24	Exon: *LHCF11*	*Phatr3_J51230*
	TGATCTCAACGTAACG**A**AGACGG	1	24	Exon: *LHCF5*	*Phatr3_J30648*
	TGAT**T**TCAACGTAACGGAGACGG	1	2	Exon: *LHCF2*	*Phatr3_J25172*
	TGATCTCAACGTAACGGAGACGG	0	2	Exon: *LHCF4*	*Phatr3_J50705*
	TGATCTCAACGTAACGGAGACGG	0	2	Exon: *LHCF1*	*Phatr3_J18049*
	TGATCTCAACGTAACGGAGACGG	0	2	*LHCF3*	*Phatr3_J50705*
mPAM2 targets	G**A**AGACGGTCGAACTT**T**TCCTGG	2	24	Exon: *LHCF11*	*Phatr3_J51230*
	G**A**AGACGGTCGAACTT**T**TCCTGG	2	24	Exon: *LHCF5*	*Phatr3_J30648*
	GGAGACGGTCGAACTT**T**TCCTGG	1	2	Exon: *LHCF2*	*Phatr3_J25172*
	GGAGACGGTCGAACTTCTCCTGG	0	2	Exon: *LHCF4*	*Phatr3_J25168*
	GGAGACGGTCGAACTTCTCCTGG	0	2	*LHCF3*	*Phatr3_J50705*
	GGAGACGGTCGAACTTCTCCTGG	0	2	Exon: *LHCF1*	*Phatr3_J18049*

3.2.2 Plasmid amplification

Stock solutions of pKS diaCas9_mPAM1, pKS hifiCas9_mPAM1 and pAF6 plasmids were amplified prior to mutagenesis. DNA concentrations were identified using the NanoDrop™ 1000 Spectrophotometer (Thermo Scientific™). The plasmids (~250 µg) were separately reintroduced to thawed, rubidium-chloride competent DH5α *E. coli* by heat-shock transformation (42°C, 40 s). Cells were suspended in liquid LB medium (500 µL, Appendix A2, Suppl. Table 3) and incubated on a shaker (37°C, 1 h). Suspension (80-100 µL) was plated out on prewarmed (37°C, 1 h) LB agar plates w. ampicillin (100 µg mL^{-1}, Suppl. Table 3) and incubated overnight (12 h, 37°C). Single-celled colonies were picked and inoculated in liquid LB-medium w. ampicillin (100 µg mL^{-1}), and grown on a shaker overnight (12 h, 37°C). Plasmids were isolated and purified with the E. Z. N. A ® Plasmid DNA Mini Kit I (Q-Spin, Omega Bio-tek®) according to the associated protocol [137]. Plasmid concentrations were detected using NanoDrop™, and extracts were stored at -20°C.

3.2.3 Biolistic Transformation of *P. tricornutum*

A culture of WT *P. tricornutum* (CCMP2561) was grown axenically for two weeks in sterile F/2 liquid medium. Growth conditions consisted of a continuous photoperiod (24:0 day:night cycle), 23°C, and irradiance (E_{PAR}) of 200 µmol photons m^{-2} s^{-1}. The cells were kept in exponential phase by dilution with new medium every second day. F/2 agar plates with and without zeocin (100 µg mL^{-1}) were prepared and stored in darkness at 4°C. The cell density of the axenic *P. tricornutum* culture was detected using the NovoCyte™ benchtop Flow Cytometer (ACEA Biosciences, San Diego, CA), and used to calculate the specific volume of culture needed to reach a cell concentration of $1.0 \cdot 10^8$ cells per plate [138]. The adequate volume was spun down by centrifugation (10 min, 4500 G), the resulting cell pellet was resuspended in F/2 liquid medium and $1.0 \cdot 10^8$ cells were plated out on each F/2 agar plate. Inoculated plates were incubated in a growth chamber for 24 h under the same growth conditions until biolistic transformation.

The biolistic transformation executed in this work utilized the Biolistic® PDS-1000/He Particle Delivery System (Bio-Rad®) in accordance with the provided manual [139]. The method relies on bombarding cells with DNA-coated tungsten- or gold-beads dubbed microcarriers. Microcarriers are loaded onto a macrocarrier and placed inside a chamber, on top of a rupture disk that can withstand a certain pressure. Inoculated plates are placed underneath, in a separate chamber. After all components are ready, gas is removed from the chamber to create a partial vacuum. Then, pressure is built-up using helium gas, causing the rupture disk to burst. Because of the vacuum in the chamber, the macrocarrier is forced towards the underneath-residing cells. However, only microcarriers are allowed into the chamber, as a stopping screen, in this instance a steel grid, rejects the macrocarrier/rupture disk entry.

Tungsten particles were coated in DNA by adding in sequence pAF6 (2.5 µg), either pKS diaCas9_mPAM1 or pKS hifiCas9_mPAM1 (2.5 µg), $CaCl_2$ (50 µL, 2.5 M) and spermidine (20 µL, 0.1 M) to an aliquot of M10 tungsten particles (3 µg, Bio-Rad®) while vortexing vigorously. The suspension was vortexed for another 3 minutes, before being allowed to settle on ice (1 min). Microcarriers were pelleted by microcentrifugation (2 s), and the liquid was discarded. Then, 140 µL cool ethanol (70% v/v, -20°C) was added to promote precipitation of DNA. The ethanol was discarded, and the process was repeated with 140 µL ethanol (100% v/v, -20°C). Finally, the pellet was resuspended in 48 µL ethanol (100% v/v, -20°C), and gently vortexed. Coated microcarriers were applied to macrocarriers, and five bombardments of 10 µL suspension were carried out for every aliquot, using rupture disks that withstand up to 1550 PSI. After bombardment, plates were incubated for 24 h in a growth chamber with continuous light (E_{PAR} ~120 µmol photons m^{-2} s^{-1}) and 23°C.

3.3 Isolation of mutants

3.3.1 Selection of transformants

After a 24 h recovery period, cells were resuspended in F/2 liquid medium (1.5 mL). Resuspension (500 µL) was transferred and used to inoculate three F/2 agar plates w. zeocin (100 µg mL^{-1}). The plates were then incubated in a growth chamber with continuous light (E_{PAR} ~120 µmol m^{-2} s^{-1}) and 23°C for 3-4 weeks to allow transformant colonies to form.

3.3.2 Lysate preparation

After recovery, colonies were transferred to F/2 agar plates w. zeocin (100 µg mL^{-1}) to increase cell mass. Alternatively, when transformant numbers were high, cells were inoculated in 24-well plates w. F/2 liquid medium w. zeocin (100 µg mL^{-1}). Cell mass was used to prepare lysates for screening. An appropriate abundance of cells was transferred from agar to lysis buffer (20 µL, Triton X-100 (1%), Tris-HCl (20 mM, pH 8.2), EDTA (2 mM)). Alternatively, lysate was prepared from F/2 liquid medium culture by centrifugation (2 min, 18000 G) and by subsequently adding the cell pellet to lysis buffer (20 µL). Suspensions were vortexed vigorously for 20 s before being allowed to settle on ice (15 min). Suspensions were then incubated at high temperature (85°C, 10 min), followed by 30 minutes in room temperature (23°C). Autoclaved MQ water (80 µL) was added and suspensions were centrifuged briefly in a microcentrifuge for 5 s. Final lysate suspensions were stored in a freezer (-20°C).

3.3.3 Detecting putative mutants

Co-transformants containing both plasmid vectors were identified by using PCR and a primer-set that amplified a fraction (544 bp) of the Cas9-conferring sequence and the gRNA. Importantly, these reactions detected the presence of a Cas9-conferring plasmid in the transformants, but since the amplified sequence only covers a fraction of the plasmid, it cannot describe whether the plasmid was intact or not. Nevertheless, this methodology eased the workload significantly by eliminating any transformants that did not contain the Cas9-vector. The PCR reactions were conducted as presented in Appendix C1 and C3, by using stock DreamTaq™ Polymerase and Green Buffer (Thermo Scientific™), M13 reverse primer (M13_R, Appendix C2) and LHCFm_PAM1 forward primer (LHCFm_PAM1_F, Appendix C2). Amplicons were identified by conventional 1% agarose gel electrophoresis (Appendix C4, Suppl. Table 7) of PCR product (6 µL), along with the GeneRuler™ 1 kb DNA Ladder (Thermo Scientific™). Colonies with amplified Cas9:gRNA fragments were cultured further in liquid F/2 medium, while negatives were omitted from further study.

3.3.4 Detecting mutations by targeted amplicon screening

To screen for mutants of interest, *LHCF* genes fully complementary to the gRNA (*LHCF1, 3* and *4*) were amplified by PCR as previously described, but now using the colourless DreamTaq™ buffer

(Thermo Scientific™) (Appendix C1). PCR amplicons were purified by use of the enzymatic ExoSAP-IT™ DNA clean-up kit (USB Corporation, Cleveland, OH) as presented in Appendix C5, and sent for sanger sequencing (Eurofins GATC, Köln, Germany). The resulting sequence data was analysed using Chromas (Technelysium Pty Ltd, Brisbane, Australia) or SnapGene (GSL Biotech, Chicago, IL) software. Moreover, an algorithm that permits genotyping of several sequences from sanger-derived .ab1-files known as CRISP-ID (available at: http://crispid.gbiomed.kuleuven.be/, [140]) was used to distinguish alleles. Putative mutations were identified from incongruent sequences with multiple peaks in and around the target region.

3.3.5 Isolating single-celled colonies from primary clones

Although the cultures were initially inoculated from primary single-celled colonies, sequencing data indicated that distinct subpopulations of mutants and WT resided in the cultures, a possible side-effect of prolonged Cas9 activity, and/or cell division taking place before mutagenesis. Thus, it was necessary re-isolate single-celled mutant cultures. These were obtained by inoculation of low cell density serial-diluted liquid cultures onto F/2 plates w. zeocin (50 µg mL^{-1}). Single-celled colonies were picked after a 3-4 weeks growth period and re-suspended in F/2 liquid media w. zeocin (100 µg mL^{-1}). Once more, targeted amplicon sequencing of *LHCF1*, *3* and *4* was conducted to verify mutant identity.

3.3.6 Confirming the indels by TOPO®-TA-based cloning and Sanger-sequencing

To confirm the sequence of previously discovered *LHCF* indels, new PCR products were processed using the TOPO® TA Cloning® Kit (Sequencing Invitrogen, Thermo Scientific™), and re-sequenced. The methodology exploits the single 3'end Adenine (A) of amplicons produced by *Taq*-polymerase. DNA topoisomerase I, an enzyme with both cutting and ligation activity, ensures that a single PCR product is ligated into a linearized TA-vector with overhanging Thymine (T), before plasmid circularity is restored. Upon ligation of PCR product, the *ccB* gene, of which is lethal to *E. coli*, becomes unfunctional. Thus, only *E. coli* cells with the recombinant TA-vector may be able to survive the following selection phase.

All *LHCF* sequences that were previously found to contain mutations were selected for re-sequencing via the TOPO-TA methodology (Invitrogen™). PCR reactions (50 µL) were conducted as described in Appendix C, and with different combinations of primers (Appendix C2). Resulting amplicons were confirmed by gel electrophoresis. DNA was purified from the PCR-products by use of the Wizard® SV Gel and PCR Clean-Up System (Promega®) and the concentration was measured by NanoDrop™ (Thermo Scientific™). Purified amplicons were ligated into TOPO® TA cloning® vectors (Invitrogen™) by incubating a solution of purified PCR product (2.0-3.5 µL), salt solution (1 µL), TOPO® TA vector (5 ng) and nuclease-free water (1.5-2.0 µL) at room temperature (22°C) for 10 minutes. PCR product was added in a volume that ensured adequate DNA quantity (~50 ng), while nuclease-free water was added to adjust the total reaction volume to 6 µL. During ligation, 100 µL

aliquots of DH5α *E. coli* were thawed on ice, before heat-shock transformation (40 s, 42°C) was conducted with the full TOPO® TA reaction volume (6 µL). Transformed bacteria (90 µL) were inoculated onto LB agar w. ampicillin (100 mM) and incubated overnight (37°C) to grow. To obtain sequences from both alleles, 5 single-celled colonies were picked from each plate. The colonies were suspended in LB medium and grown on a shaker overnight (37°C). The following day, cells were pelleted by centrifugation (15 min, 4000 g), and plasmids were isolated and purified by column isolation (E.Z.N.A® Plasmid Mini-Kit I, Q-Spin, Omega Bio-Tek). The DNA concentration (ng μL^{-1}) was measured by NanoDrop™ (Thermo Scientific™). Sequencing reaction solutions (10 µL) were prepared with the M13 reverse primer (2.5 µL, 10 mM) and purified TOPO®-TA plasmid (5.0-7.5 µL, ~750 ng). MQ water (2.5 µL) was added to adjust the volume to 10 µL given a high DNA concentration (>150 ng μL^{-1}). The resulting solutions were sent for sequencing to GATC (Eurofins GATC, Köln, Germany), and analysed using Chromas and/or SnapGene software.

3.3.7 Checking the effect of re-editing

The Cas9:gRNA complex may retain some of its affinity to edited target genes given that the introduced indel is small. Re-editing, in which a continuously expressed Cas9:gRNA ribonucleoprotein-complex binds and re-edits the sequence can therefore create novelties in cell lines that are already sequenced. To infer on the putative effect of re-editing, a selection of edited *LHCFs* with small indels were amplified by PCR, isolated by TOPO®-TA-based cloning and sent for sequencing.

3.4 Characterization of mutants

Three LHCFm mutants were picked for in-depth characterization; LHCFm 1.10, 6.1.11 and 15.1, where 'm' denotes multiple. Namely, these were picked because of their light- to dark-green phenotypes, and because they contained several *lhcf* KO mutations, possibly leading to significantly truncated light-harvesting antennas (TLAs).

3.4.1 Phenotypes

WT and LHCFm mutants were grown in F/2-medium for 5 days in continuous light of low irradiation (~50 µmol photons m^{-2} s^{-1}). Temperature was kept stable at 23°C. Cells were pelleted by centrifugation (4000 G, 13 min) and diluted to reach a uniform cell density of 10^8 cells mL^{-1}. Photographs were taken using a Canon EOS 650D mirror reflex camera.

3.4.2 Variable fluorescence

To infer on parameters that define photosynthetic efficiency under phototrophic growth, variable fluorescence was assessed using PAM fluorometry. Cells were grown for 2 weeks in F/2 medium and at continuous high-light (HL, 225 µmol photons m^{-2} s^{-1}) and low-light (LL, 50 µmol photons m^{-2} s^{-1}). The light source was a cool white LED (Solarmass EQUILIGHT, Appendix E, Suppl. Fig. 5). The

temperature was kept stable at 15°C. Cells were treated with new medium and shaken into suspension once every day. A day prior to analysis, the cell density was identified by hemocytometry, and used to prepare cultures (3 biological replicates of each cell line and light-intensity, total sample size = 24) that would reach a cell density of 0.8-1.5 · 10^6 cells mL^{-1} overnight. Fluorescence kinetics were assessed using the Pulse Amplitude Modulated fluorometer PHYTO-PAM (Waltz, Germany), and the associated PhytoWin software (Waltz, Germany). A Peltier cell (US-T/S, Walz) kept the temperature constant (15±0.2°C) during incubations. A protocol for fluorometric analyses was devised as presented in Appendix H (Suppl. Table 10). Plastic cuvette samples were subjected to a three-minute incubation in darkness before starting the analysis to ensure that all RCs were open. All samples were prepared with the least amount of time outside of the growth room to ensure that the cultures retained their light-acclimation physiology.

The effective quantum yield of PSII photochemistry (Φ_{PSII}) was calculated for each irradiance using the variable fluorescence ($F_v` = F_m` - F_0`$) and the maximum fluorescence ($F_m`$) in actinic light (`) (Eq. 1). The maximum quantum yield of PSII photochemistry ($\Phi_{PSIImax}$) was calculated using the variable fluorescence ($F_m - F_0$) and the maximum fluorescence (F_m) yields in darkness (Eq. 2). These parameters describe the proportion of absorbed light going into PSII photochemistry, and its maximum capacity when all RCs are open (in darkness) [141]. Relative electron transport rate (rETR) was obtained by multiplying the values of Φ_{PSII} with the associated irradiance (E_{PAR}) [142,143] (Eq. 3). Moreover, the recognizable photosynthesis vs irradiance (E) – or so-called PE-curve was plotted by taking the rETR as a function of the irradiance [144]. Traditionally, three informative parameters may be extracted from the curves of the PE plots. The slope, or α – is the light-utilization coefficient, and reports the photosynthetic efficiency at low irradiances [144,145]. The maximum relative electron transport rate ($rETR_{max}$, often abbreviated P_{max}) reports on the maximum photochemical capacity of photosystem II (PSII), may be interpreted as an estimate for the rate of photosynthesis, and is read off as the asymptote (i.e. its maxima) approximated by the PE-curve when photosystems become fully saturated [144]. Lastly, the light saturation index (E_k), denotes the irradiance at which photochemistry becomes saturated [141]. The E_k is therefore extracted from the irradiance at which the α intersects $rETR_{max}$ ($E_k = rETR_{max} / \alpha$). Conveniently, these were automatically exported from the PhytoWin software. Non-photochemical quenching (NPQ) was calculated using the Stern-Volmer equation as adapted for use in diatoms (F_m^{max} instead of $F_m`$, Eq. 4) where F_m^{max} is the F_m with the highest fluorescent yield, and $F_m´$ is the maximum fluorescence yield at a each irradiance [146].

$$\Phi_{PSII} = \frac{F_v`}{F_m`} = \frac{(F_m` - F_0`)}{Fm`} \qquad \text{Eq. 1}$$

$$\Phi_{PSIImax} = \frac{F_v}{F_m} = \frac{(F_m - F_0)}{F_m} \qquad \text{Eq. 2}$$

$$rETR = \Phi_{PSII} \cdot E_{PAR} \quad \text{Eq. 3}$$

$$NPQ = \left(\frac{{F_m}^{max}}{F_{m'}}\right) - 1 \quad \text{Eq. 4}$$

3.4.3 Lysate for pigment and protein analyses

Samples intended for HPLC and western blot analyses were grown and harvested simultaneously. LHCFm 1.10 and WT cultures were grown in triplicates in continuous LL (45-50 µmol photons m^{-2} s^{-1}) and HL (~250 µmol photons m^{-2} s^{-1}). The temperature was kept at 23°C. Cultures were diluted every 48h to keep the cells in exponential phase and avoid nutrient deprivation. After 2 weeks of growth, cells were harvested for both pigment and protein analyses.

3.4.4 Pigment composition by HPLC

Approximately 35 million cells were collected onto Whatman GF/F filters by vacuum-filtration. Two technical replicate samples were taken for each biological replicate. Filters were immediately stored at -20°C. At a later date, pigments were extracted in methanol (1.6 mL, 100%) over 24 h, and re-filtered (0.45 µm) to get rid of particles. HPLC analysis was conducted by senior technician Kjersti Andresen with 77 µL extraction volume on a HP 1100 series HPLC system (Hewlett-Packard, Wilmington, IL, USA) following the protocol described by Rodríguez et al., [147]. Pigment signals were detected at 440 nm.

The obtained pigment data (ng $filter^{-1}$) were correcting for by cell number (cells mL^{-1}), filtration volume (mL) and molecular weight (g mol^{-1}) of pigments – obtaining pigment quantity with the specific unit of mol $cell^{-1}$. Because cultures of LHCFm 6.1.11 and 15.1 were grown in a separate experiment by co-supervisor Amit Sharma, the pigment data from these had to be combined with the pigment data from LHCFm 1.10 and WT. However, even though the conditions of growth were approximately equal, other effects may surface when grown separately. Thus, only the pigment ratios were deemed comparable. The ratio of Fuco:Chl *a* + Chl *a*-like (mol:mol) was calculated from the amount of Fuco and the total amount of Chl *a* and Chl *a*-like derivates (slightly shifted retention time to that of intact Chl *a*). All Fuco:Chl *a* + Chl *a*-like ratios are now referred to as Fuco:Chl *a*. The de-epoxidation state index (DES = Dtx/(Dtx+Ddx)) was calculated in conformity with Ruban et al., [148].

3.4.5 FCP contents by western blots

Approximately 70 million cells of WT and LHCFm 1.10 were sampled by filtering the adequate volume (25-50 mL) of cell culture onto Durapore® Membrane Filters (0.65 µm). Harvested cells were re-suspended in F/2 medium (1 mL) and vigorously vortexed (10 s). Suspended cell samples were centrifuged at max speed (20150 G) for 1 min, and the resulting pellets were immediately snap-frozen in liquid nitrogen and stored at -80°C. Pre-cooled steal beads were added to frozen samples. Then, the

TissueLyser II (Qiagen) system was used to mechanically disrupt algae pellets (2 min, 25 Hz). Lysation buffer (700 µL, 50 mM Tris (pH 6.8), 2% SDS) was added to each tube, before another disruption step was conducted (8 min, 25 Hz). Lysates were centrifuged at low gravity (100 g) for 30 min at 4°C, and the resulting supernatant harvested for protein concentration detection, and subsequent western blot reactions.

3.4.5.1 Total protein concentration by the DC protein assay

The total protein concentrations of lysates were measured by colorimetry using the DC protein assay protocol (Bio-Rad®). A solution of reagents A (alkaline copper tartrate) and S were prepared by adding 20 µL of reagent S for every mL of reagent A. 40 µL of lysate and stock solutions of gamma-globulin (0, 0.2, 0.4, 0.8, 1.47 µg μL^{-1}) were separately added to 200 µL of reagent solution and 1600 µL of buffer B (diluted Folin). Reaction solutions were mixed by gently vortexing for 2 s, and left to incubate at room temperature (RT) for 15 minutes. The full reaction volumes were loaded into plastic cuvettes. Protein concentrations were measured at 750 nm with the SmartSpec™ Plus Spectrophotometer (Bio-Rad®), using stock solutions of gamma-globulin to generate a standard-curve ($R^2 = 0.995$).

3.4.5.2 Preparation of SDS-polyacrylamide gels

Because of the relatively small mass of LHCF proteins (20 < 23 kDa according to UniProtKB; https://www.uniprot.org/uniprot/?query=lhcf+AND+phaeodactylum+tricornutum&sort=score), we aimed at using dense 15% sodium-dodecyl-sulphate (SDS) polyacrylamide (PA) separation gels. These were prepared by adding the reagents presented in Appendix I (Suppl. Table 11). Liquid acrylamide mix (29.2% acrylamide, 0.8% N,N'-methylene-bis-acrylamide) and TEMED (N,N,N',N-Tetramethyl-ethylenediamine) were added in a fume-hood. 7 mL of reaction volume was initially used to set up the main body of the separation gel (1.5 mm thickness). Isopropanol (1 mL) was used to remove any air-bubbles at the top surface of the forming gels, before leaving them to polymerize for an hour. The isopropanol was then removed by using an absorbing filter-paper. A 15% acrylamide stacking gel was made adding in order the same reagents, although at different proportions and with the exception of now using a low-pH Trizma-base (pH 6.8). The stacking gel mixture was added to fill the upper cavity of the gel cast. A mould was applied to form wells in the stacking gel, and the construct left to solidify for another hour. Finished acrylamide gels were orderly covered in moist paper, wrapped in plastic (to prevent drying) and stored in a fridge (4°C) for later use.

3.4.5.3 Separation of proteins

Three SDS-polyacrylamide gels were set up; two for western blot, and one for a Coomassie Blue stain. Protein lysates from LHCFm 6.1.11 and 15.1 grown in the same LL- and HL-conditions were obtained from co-supervisor Amit Sharma. WT and mutant (LHCFm 1.10, 6.1.11 and 15.1) protein lysates were thawed on ice and diluted in lysis buffer (50 mM Tris (pH 6.8), 2% SDS) to approximately 1 µg μL^{-1}.

Additionally, a three-levelled serial dilution (100%, 50% and 25%) was made for both LL- and HL-acclimated WT protein lysates (100, 50 and 25%) by dilution with lysis buffer. All 60 µL protein lysates were diluted in 20 µL 4X SDS-loading buffer (200 mM Tris-Hcl (pH 6.8), 400 mM DTT, 8% SDS, 0.4% Bromophenol blue, 40% Glycerol). The samples were boiled (10 min., 95°C) to ensure that proteins were truly denatured and linearized, and then put on ice. SDS-PAGE was set up with a visible Precision Plus™ Dual Colour Standard Ladder (4 µL, BioRad®) and a MagicMark™ XP Western Protein Standard (Invitrogen™, 5 µL) and with sample volumes containing 10 µg protein. Proteins were separated over 1h and 30 minutes at 100 V in running buffer (25 mM tris-base, 250 mM glycine, 0.1% SDS).

3.4.5.4 Staining Coomassie-gel

The gel intended for Coomassie staining was transferred to MQ-water and boiled for 1 minute in a microwave (900W). The water was replaced by approximately 20 mL of SimplyBlue™ SafeStain (Invitrogen™). The staining solution was left to incubate overnight on a slowly rotating shaker (50 rpm). A day after, the staining solution was discarded, and replaced yet again by MQ-water. The gel was washed twice more in MQ-water before it was pictured and analysed.

3.4.5.5 Blotting

The gels intended for western blotting were incubated for 10 minutes in transfer buffer (tris base (25 mM), glycine (192 mM), methanol (20%, v/v), SDS (0.05%, w/v), and prepared in accordance with the Mini-PROTEAN® Tetra Mini Trans-Blot® system (Bio-Rad®). Filter paper (3 mm) and nitrocellulose membrane (0.2 µm, Bio-Rad®) were cut to the dimensions of the gel (5.6 · 8.3 cm) and incubated with foam-pads (Bio-Rad®) in transfer buffer for 10 minutes. Blotting "Sandwiches" were made by assembling in order foam-pad, filter paper, nitrocellulose membrane, gel, filter paper and foam-pad, and by finally enclosing it in a Mini Trans-Blot® cassette. The sandwiches were put into a buffer tank (Bio-Rad Mini Trans-Blot® system) containing ice-cold transfer buffer and a cooling unit. A 100V current was applied for 60 minutes to allow transfer of proteins from gel to nitrocellulose membrane. The nitrocellulose membranes were then transferred and incubated in a 1x phosphate-buffered saline blocking buffer w. 5% skimmed milk (PBS-T, 8 mM Na_2HPO_4, 150 mM NaCl, 2mM KH_2PO_4, 3mM KCl, 5% skimmed milk (w/v, Oxoid®)) for 2 hours to prevent unspecific binding of antibodies.

Two primary antibodies were used, one for each membrane. FCPs in the first membrane (which are presented in the results) were targeted by α-cmFCP: a polyclonal rabbit antibody created from an amino acid sequence identical in LHCF1-11 (GDFRNGYIDFGWDSFD) in *P. tricornutum* [96], that was obtained by correspondence with the Büchel laboratory (Institute for Molecular Bioscience, Frankfurt University, Germany). The other membrane (which is provided in Appendix F) was set up using an anti-FCP antibody made against a 19 kDa FCP of *Heterosigma akashiwa* (Agrisera®, Catalog #: AS174116) that can potentially bind to several different FCPs in *P. tricornutum*. Both primary

antibody:FCP reactions were set up by incubating the membranes in PBS-T buffer w. 5% skimmed milk (w/v, Oxoid®) and either of the two primary antibodies (1:500). Both membranes were incubated on a rolling-motion device at ~4°C overnight.

3.4.5.6 Washing and visualizing protein bands

Nitrocellulose membranes were washed four times in PBS-T (20 mL, 5 min) at RT on a roller. The blots were then incubated for another two hours (at RT) in 10 mL of PBS-T w. 5% skimmed milk and secondary antibody (1:10000, polyclonal Goat anti-rabbit IgG (H+L), HRP Conjugate, Invitrogen™, Catalog #: 65-6120). After coupling with the secondary antibody, both membranes were washed another four times in PBS-T buffer as previously described. A 1:1 solution of Super Signal™ West Pico Chemiluminescent Substrate (Thermo Scientific™) was prepared from luminol and hydrogen peroxide (1:1, v/v) and spread out (1.8 mL) on the membrane. The membrane was left to incubate for 10 minutes at RT before signals were detected in a G-box (ECL, Syngene®).

3.4.5.7 Displaying reference proteins

To control that the wells were loaded at equal protein quantities, another two primary antibodies were used to display proteins which should be uniformly expressed in all cell lines. The ATP synthase subunit β (atpB, Mw = 51.621 kDa (Uniprot: A0T0D2) and the Photosystem II protein D1 (D1, Mw = 39.651 kDa (Uniprot: A0T0G9)) were targeted using primary antibodies manufactured by Agrisera®. AtpB was targeted using a KLH-conjugated polyclonal rabbit synthetic peptide (1:4000, Agrisera®, Catalog #: AS05 085-10), and D1 was targeted using a KLH-conjugated polyclonal hen synthetic peptide (1:20000, Agrisera®, Catalog #: AS01 016). Both antibodies were diluted in PBS-T w. 5% skimmed milk and applied separately by incubating overnight as previously described. For each new primary antibody, the processes of washing and coupling w. secondary antibodies were repeated. The bands were identified by using the same solution (1:1, v/v) and volume (1.8 mL) of Super Signal™ West Pico Chemiluminescent Substrate (Thermo Scientific™) and inspected in G-box (Syngene®).

3.4.6 Absorbance

Cultures of WT and mutants LHCFm 1.10, 6.1.11 and 15.1 were acclimated to HL and LL intensities for two weeks and diluted every two- or three days to maintain a cell density of below 1.5 million cells mL^{-1}. After acclimation, culture samples were taken for spectrophotometric analyses. A Hitachi U-3010 UV-visible Spectrophotometer (Hitachi) equipped with an integrating sphere was used to take measurements. Absorption was measured twice for each sample, once with the quartz cuvette near/proximal the pinhole (A_p), and once 5 mm distal to the pinhole (A_d). This to ensure that side-scattered light is corrected for when translating the absorbance from transmitted and emitted light. Essentially, this was achieved by correcting the proximal absorbance (A_p) by subtracting a coefficient (α) of the difference between proximal and distal absorbance (i.e. scattered light, $D = A_p - A_d$). The

corrected spectral absorbance (A_c) of samples would then be computable with the given equation (Eq. 5), where α (~[0.6, 1.2]) was a variable used to force the smallest datapoint of A_c to approximate zero. Because this correction affects all the points in the curve to a similar extent (all datapoints are subtracted by a coefficient αD), it does not change the relationship between absorbance at different wavelengths.

$$A_c = A_p - \alpha D,\ \alpha \sim [0.6, 1.2] \qquad \text{Eq. 5}$$

Because the relation of curves to one another are of interest rather than the intensities, a min-max normalization was performed. All datapoints were scaled to a value between 0 and 1 by a function of the minimum (usually a datapoint around 715 nm) and maximum absorbance (usually the Chl *a* soret band maximum) (Eq. 6). Furthermore, absorbance difference spectra (A_{DS}) were calculated by subtracting the normalized datapoints of mutant spectra (A_c^{Mut}) from those of the WT (A_c^{WT}) of equal acclimation (e.g. HL or LL, Eq. 7).

$$A = \frac{(A_c - A_{min})}{(A_{max} - A_{min})} \qquad \text{Eq. 6}$$

$$A_{DS} = {A_c}^{WT_{LL}} - {A_c}^{Mut_{LL}} \qquad \text{Eq. 7}$$

3.4.7 Growth curves

LL- and HL-acclimated cultures of WT and 1.10 were inoculated to 50000 cells mL^{-1} and grown for a week at HL (~250 µmol photons m^{-2} s^{-1}) and LL (50 µmol photons m^{-2} s^{-1}) intensities. The day-length was continuous (24h light) with the spectral irradiance of cool white LED light (Appendix E, Suppl. Fig. 6). Temperature was kept stably at 23°C. Cell densities (CD) were counted each day by flow cytometry (NovoCyte™ flow cytometer, ACEA Biosciences) The specific growth rate (K'), cell divisions per day (divisions day^{-1}) and cell doubling time (h) in exponential phase were calculated as presented in Eq. 8-10 [149].

$$K' = \frac{LN\left(\frac{CD_{stop\ exponential}}{CD_{start\ exponential}}\right)}{Days} \qquad \text{Eq. 8}$$

$$\frac{Divisions}{day} = \frac{K'}{LN(days)} \qquad \text{Eq. 9}$$

$$Doubling\ time\ (h) = \left(\frac{1}{\frac{Divisions}{day}}\right) \cdot 24 \qquad \text{Eq. 10}$$

3.5 High-density growth trials in 1L photobioreactors

3.5.1 Selecting mutants for high-density cultivation

An important end-goal of the thesis was to perform a pilot-experiment that tested the performance of TLA mutants in light-deprived photobioreactors. Because light-deprivation is a principal limiting factor to the industrial productivity of such cultures, it was hypothesized that cultures who absorb less obtain higher densities, and higher biomass in light-limited conditions. Three of the LHCFm mutants that exhibited phenotypes with a changed coloration were selected for further trials; LHCFm 1.10, which during the pre-experimental acclimation phase had developed a phenotype resembling the WT, and another two LHCFm mutants generated with the same construct (diaCas9:mPAM1) dubbed LHCFm 15.1 and 6.1.11, which were obtained from co-supervisor Amit Sharma. In addition, three TLA mutants with KO of the *Albino 3b-insertase* (*alb3b 14.8*, *16.7* and *19.7*) were obtained from co-supervisor Dr. Marianne Nymark. These have a green phenotype, reduced FCP protein content, and are known to exhibit approximately 65% and 35% of the Fuco and Chl *a* content of WT [115].

3.5.2 Pre-experimental acclimation phase

Mutants and WT cells were grown axenically in F/2 medium for two weeks prior to the experiment in order to prepare them for the growth conditions in the photobioreactors. The temperature was kept at 22°C, and the light-climate consisted of a cool white LED light (Solarmass EQUILIGHT, Appendix D, Suppl. Fig. 5) of medium-high irradiance intensity (~200 µmol photons m^{-2} s^{-1}) and a continuous photoperiod. Mixing was achieved using a shaking incubator (250 rpm). The cultures were kept in exponential phase (<1.5 million cells mL^{-1}) by diluting every 48 h.

3.5.3 Photobioreactor setup

Growth experiments were conducted at the Norwegian Centre for Plankton Technology at SINTEF Ocean in Trondheim, Norway. Cells were grown in two-spaced column glass flask photobioreactors with an inner growth chamber (1 L, 5 cm diameter), and a surrounding water-flow space (Fig. 3.2). The temperature of all eight flasks kept at 22°C by using a water-flow system with a mechanic temperature control, and by using the same water-body in a cycled loop. The displayed temperature was verified by independent measurements with a thermostat every two days. To ensure that nutrient availability was not a limiting factor until the end of the experiment, cells were grown in the original F medium [150] (Appendix A1). To address an emerging problem of higher light exposure for centred flasks, cardboard plates were mounted between centred flasks, so that all flasks received shared light from two diode plates (Fig. 3.2). Prior to start, and at day 14, the combined light exposure was measured at between 200 and 220 µmol photons m^{-2} s^{-1} in the centre of the flasks using a LI-250A Light Meter (LI-COR®, Lincoln, Nebraska, USA) with a spherical underwater quantum-sensor. The spectral irradiance of the light-source was that of a cool white LED as presented in Appendix E (Suppl. Fig. 4). Marking the start

of the experiment (0 hours), cells were inoculated in a concentration of 100 000 cells mL^{-1} in a total volume of 1 L of F medium (Appendix A1). Cultures were continuously aerated with atmospheric gas (375 mL min^{-1}, CO_2~ 300 ppm). Additional CO_2 influx was appointed after 96 hours to boost cells into continued exponential growth. All cultures received the same amount of CO_2, until the pH stabilized within an optimal range (pH~8) [151]. Equal quantities of nutrients to that of 1 L F-medium was added after 216 hours (WTA) and 264 hours (all other cultures) to encourage further growth.

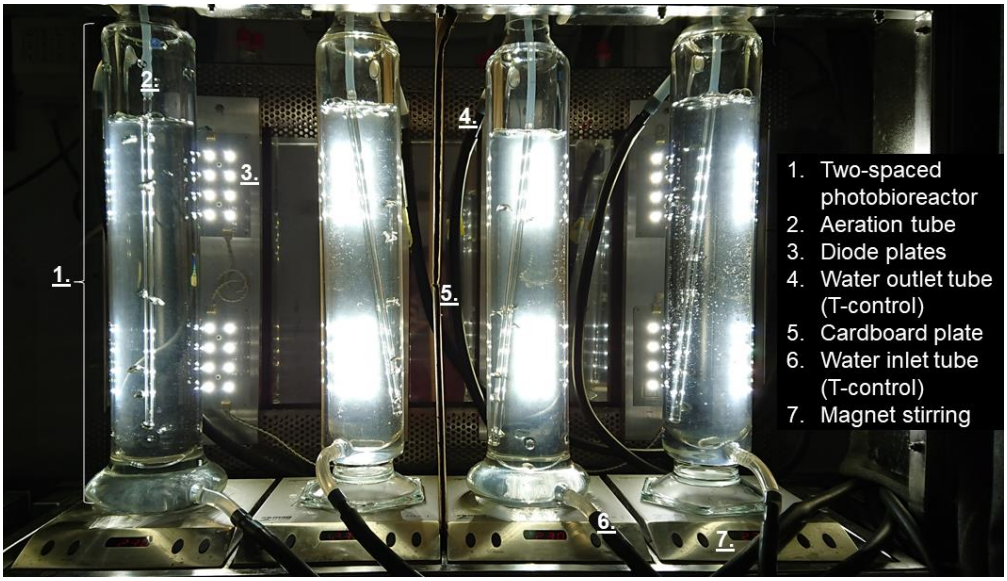


Fig. 3.2: The photobioreactor setup showing the main components in 4 out of 8 photobioreactors. Cylindrical two-spaced photobioreactors (1.) with aeration tubes (2.) providing input of atmospheric gas and CO_2 (optional), illumination from two diode plates (3.), a cardboard plate (5.) to ensure homogenous distribution of light and a magnet stirrer (7.) to ensure homogenous distribution of cells within the culture. The water flow system maintained a temperature of 22°C by cycling a shared water body from water inlet (6.) to water outlet (4.), via a control unit and back. Picture taken at day 1 (after 24 h). Note that the apparent differences in volume are caused by photobioreactor 2 and 4 (from the left) being a slightly different design, and thus taller.

3.5.4 Cell density and $\Phi_{PSIImax}$

Cell counts were performed daily using the BD Accuri Plus Flow Cytometer (BD Biosciences, San Jose, USA). Cell numbers at day 8 and 9 were discarded due to becoming increasingly unreliable, possibly caused by cells aggregating. Because of this, all samples for cell counts were fixed in lugol, and counted manually by hemocytometry from day 10 until end. Also, maximum quantum yields of PSII photochemistry ($\Phi_{PSIImax}$) were measured daily with the handheld AquaPen-C AP-C 100 pulse amplitude modulated (PAM) fluorometer (Photon Systems Instruments, Drásov, Czech Republic). Samples were subjected to 30 minutes incubation in darkness to ensure that all photosystems were open and active at the start of fluorometry, and to relax the qE-component of NPQ.

3.5.5 Nutrient analyses

Samples for particulate nutrient analysis were collected by vacuum-filtering onto Whatman GF/F filters (particulate phosphorous, PP) and acid-washed and heat-treated Whatman GF/F filters (particulate carbon and nitrogen, PC and PN respectively). Flow-through filtrate was utilized for analysis of dissolved inorganic nitrate + nitrite (DIN) and phosphate (DIP). In general, the amount of sample depended on cell density, and was chosen to reach an adequate colouring on the filter. All samples were immediately stored and kept at -20℃. All nutrient analyses were performed at Trondheim Biological Station (TBS, NTNU, Trondheim) supervised by senior technician Kjersti Andresen. Particulate nitrogen (PN) and carbon (PC) samples were prepared in fuming hydrochloric acid (HCl, 37%), enclosed in tin pellets and dried (48 h, 60℃) before analysis. Particulate phosphorous (PP) samples were prepared by adding distilled water (10 mL), potassium peroxidisulphate (max 0.001% N) and sulphuric acid (0.1 mL, 4 M) reagents in series. Reactions were autoclaved (30 min, 120℃, 200 kPa), re-filtered (0.45 µm) and analysed on a ECS 4010 CHNSO element analyser (Costech Instruments, Milan, Italy). Particulate nutrient quantities were calculated as picogram (pg) per cell by correcting for filtered volume (fv), dilution factor (df) and cell density (cd) (Eq. 11–13). Cellular N:P ratios were calculated from the molar concentrations of N and P per cell (Eq. 14). Dissolved inorganic nitrate and nitrite (DIN; NO_3^-, NO_2^-) and phosphate (DIP; PO_4^{-3}) samples were diluted and analysed in parallel as described by Hansen et al., [152]. The obtained unit of DIN and DIP (mg L^{-1}) was only corrected for by the dilution used during the analysis (df).

$$PP\left(\frac{pg}{cell}\right) = \frac{\left(P_{sample}\left(\frac{pg}{mL}\right) - P_{blank}\left(\frac{pg}{mL}\right)\right) \cdot 0.0121 \cdot df \cdot \left(\frac{1000}{fv}\right)}{cd\left(\frac{cells}{mL}\right)} \quad \text{Eq. 11}$$

$$PC\left(\frac{ng}{cell}\right) = \frac{C_{sample}\left(\frac{ng}{mL}\right) - C_{blank}\left(\frac{ng}{mL}\right)}{cd\left(\frac{cells}{mL}\right)} \quad \text{Eq. 12}$$

$$PN\left(\frac{ng}{cell}\right) = \frac{N_{sample}\left(\frac{ng}{mL}\right) - N_{blank}\left(\frac{ng}{mL}\right)}{cd\left(\frac{cells}{mL}\right)} \quad \text{Eq. 13}$$

$$N{:}P\left(\frac{mol}{mol}\right) = \left(\frac{PN\left(\frac{g}{cell}\right)}{Mw\left(\frac{g}{mol}\right)}\right) / \left(\frac{PP\left(\frac{g}{cell}\right)}{Mw\left(\frac{g}{mol}\right)}\right) \quad \text{Eq. 14}$$

3.5.6 Lipid analysis

Samples for lipid analysis were taken after three successive days of stable stationary phase (less than 15% change in cell density). 100 mL of cell culture was harvested and centrifuged (4000 g, 15 min). Pellets were dried in a freezer (-80℃) overnight. Total lipid content (% fats per. DW) was assessed using the Bligh & Dyer method [153] by laboratory technician Merethe Selnes at SINTEF Ocean, Trondheim.

4 Results

4.1 HifiCas9 is less promiscuous but also less active

Two Cas9-enzymes (pKS_hifi_diaCas9_mPAM1, pKS_diaCas9_mPAM1, Appendix B, Suppl. Fig. 1 and 2) were used for targeting the same subset of *LHCF*s in the *P. tricornutum* genome. Both constructs were introduced to the diatom cells by biolistic co-transformation with a selection-vector (pAF6, Appendix B, Suppl. Fig. 3). Transformants were identified by resistance to Zeocin antibiotics as conferred by pAF6, and putative mutants (PM) were identified by using PCR to amplify a part of the DNA encoding the Cas9 enzyme and the mPAM1 gRNA. Follow-up screens by PCR-based Sanger-sequencing identified putative mutations in *LHCF* genes. Mutations were confirmed in pure lines by TOPO-TA cloning of *LHCF* PCR products followed by Sanger-sequencing. The resulting sequence data was complemented by the work of co-supervisor Amit Sharma who have generated similar lines using mPAM1 and a second gRNA (mPAM2). Finally, the sequence data of both mPAMs were compiled, and used to describe the on- and off-target mutation efficiencies of both dia- and hifiCas9.

4.1.1 Transformation and mutation efficiencies

The overall efficiency of the transformation and subsequent mutation (Table 4.1) reports on the overall success of the project pipeline. The output of 248 transformants from $5.0 \cdot 10^8$ cells was well within the range of previously reported transformation efficiency values acquired with biolistic transformation (5-25 transformants per 10^8 cells). However, the number of transformants which were putative mutants and contained an at least partially intact Cas9:gRNA construct was low and only 3.5% and 11.8% for dia- and hifiCas9 constructs, respectively. Moreover, of the putative mutants only 15% (at a total of 3 mutants), were confirmed to contain mutations in target *LHCFs* (confirmed mutants, CM). Mutation efficiencies (M. eff.) were therefore low with only $6.7 \cdot 10^{-9}$ and $2.0 \cdot 10^{-8}$ confirmed mutants per cell for dia- and hifiCas9, respectively.

Table 4.1: Overview of transformation efficiencies obtained using constructs of mPAM1 gRNA with diaCas9 and HiFiCas9. Total number of transformants (t) are presented as the number of zeocin-resistant colonies (i.e. are transformed with pAF6). Transformation efficiency (T. eff.) is presented as the number of transformants (t) per number of cells used for biolistic transformation. PM represent putative mutants (cell lines with at least partially intact Cas9:gRNA constructs). CM represent confirmed mutants (i.e. cell lines with one or more confirmed mutations in either of the targeted *LHCF*s). The percentage of transformants which were putative mutants is presented as % t, and the percentage of putative mutants which were confirmed mutants is presented as % PM. Mutation efficiency (M. eff.) is the number of confirmed mutants per cell used for biolistic transformation.

Construct	t (+pAF6)	T. eff. (t cell^{-1})	PM	% t	CM	% PM	M. eff. (CM cell^{-1})
diaCas9:mPAM1	113	$4.10 \cdot 10^{-7}$	4	3.5	1	25.0	$6.67 \cdot 10^{-9}$
hifiCas9:mPAM1	135	$4.90 \cdot 10^{-7}$	16	11.8	2	12.5	$2.00 \cdot 10^{-8}$
Total	248	$4.50 \cdot 10^{-7}$	20	8.1	3	15.0	$1.33 \cdot 10^{-8}$

4.1.2 Overview of edited *LHCFs*

The sequencing data was compiled in detail, describing indels in singular alleles for all screened lines. However, when inferring the on- or off-target effects of dia- and hifiCas9 enzymes, it is arguably more interesting to infer whether DSBs occur or not, rather than describing the length or effect of indels. Table 4.2 presents the binomial records of edited (E) or un-edited (-) target-sequences of the screened mutant strains. No mutation events were identified at target sites with two or more mismatched nucleotides (m > 2). Thus, off-target gene editing events existed only for sites with 1 mismatched nucleotide. Only one off-target gene was found to contain indels in hifiCas9-derived lines (n_{tot} = 9), whereas 16 off-target genes were edited in diaCas9-derived lines (n_{tot} = 11). Simultaneously, a higher proportion of on-target genes are left un-edited, and hifiCas9-generated mutants are also more infrequent, as exemplified by a total output of four mutant lines generated with hifiCas9_mPAM1.

Table 4.2: Overview of genes edited utilizing different combinations of gRNAs (mPAM1, mPAM2) and Cas9 endonucleases (diaCas9, hifiCas9). Detected mutations are labelled "E" (Edited). On-target mutations are labelled green, while off-target mutations are labelled red. Target-sequences unaffected by Cas9 (i.e. un-edited WT sequences) are labelled with a "-". *LHCF* target-sites are accompanied with the number of mismatched nucleotides (m).

Cas9 enzyme	gRNA	Mutant strain	0 m			1 m		2/3 m
			LHCF1	*LHCF3*	*LHCF4*	*LHCF2*	*LHCF5*	*LHCF11*
diaCas9	mPAM1	6.1.11	E	E	E	E	E	-
		6.4	E	E	E	E	E	-
		15.1	E	E	E	E	E	-
		5.1.2	E	E	E	E	E	-
		1.10	E	E	E	E	-	-
		3.1	E	E	-	E	-	-
	mPAM2	5.4	E	E	E	E	-	-
		8.1	E	E	E	E	-	-
		8.5	E	E	E	E	-	-
		10.1	E	-	E	E	-	-
		3.1	E	E	E	E	-	-
hifiCas9	mPAM1	3.1	E	E	E	-	-	-
		5.1	E	E	E	-	-	-
		15.2	-	E	E	-	-	-
		14.2	E	-	-	-	-	-
	mPAM2	2.1	E	E	-	E	-	-
		6.1	E	E	-	-	-	-
		6.5	E	E	E	-	-	-
		4.1	E	E	-	-	-	-
		8.5	E	-	-	-	-	-

4.1.3 On- and off-target editing efficiency

In general, the number of edited genes was lower in lines generated with hifiCas9 (Fig. 4.1). Moreover, hifiCas9 completely avoided mutagenesis at mPAM1 off-target sites (*LHCF2*, *LHCF5*), as opposed to diaCas9_mPAM1 which promoted mutagenesis at these sites (Fig. 4.1a). At the on-target sites of mPAM1 (*LHCF1*, *LHCF3* and *LHCF4*), hifiCas9 induced mutagenesis in 75% of the screened lines (n=4). The overall on-target activity of hifiCas9 contra WT diaCas9 was 79% when primed with mPAM1 and 71% when primed with mPAM2. When primed with mPAM2 gRNA, both enzymes

produced indels at the in-seed 1 nucleotide off-target site of *LHCF2*, but hifiCas9 did so in only 1 out of 5 mutants, as opposed to 5 out of 5 when using diaCas9 (Fig. 4.1b). In the *LHCF2* (non-seed mismatch) and *LHCF5* (in-seed mismatch) off-target sites of mPAM1, mutations were detected in 100% and ~66% (respectively) of lines generated with diaCas9:mPAM1 (n = 6). Neither of *LHCF2* and *LHCF5* were edited when hifiCas9 was primed with mPAM1. No mutations were detected in *LHCF11*, where the mismatch count was 3 for mPAM1 and 2 for mPAM2.

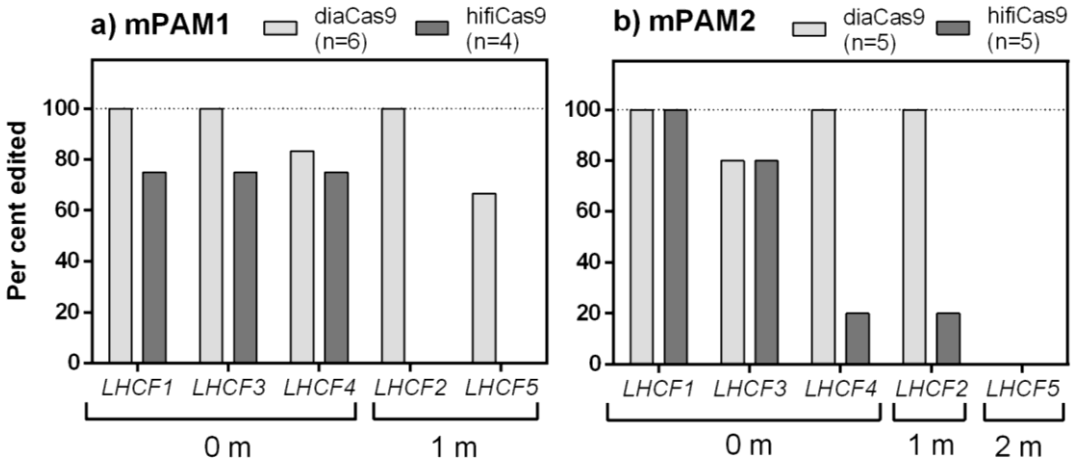


Fig. 4.1: The percentage of screened mutants in which editing events has occurred in *LHCF1-5*, and when using combinations of hifi- (dark grey) and diaCas9 (light grey) with a) mPAM1 and b) mPAM2 gRNAs. Both detected indels and sequences unamplified after numerous attempts are considered editing events. Sequences that were both impossible to amplify and intermediately located between two edited sites were deemed not edited as they may well have been deleted. Scoring is based on 1 (editing has occurred) and 0 (no occurrence of editing) and does not distinguish between bi- or mono-allelic mutagenesis. *LHCF* target sites are accompanied with the associated number of mismatched (m) nucleotides (0, 1 or 2), in which 0 m indicates on-target, and 1 and 2 m indicate off-target.

4.2 Characterization of LHCFm mutants

4.2.1 Selecting mutants for characterization

Three LHCFm mutant strains were selected for characterization on basis of preliminary sequencing data, and phenotypic differences from WT cells. Notably, the mutants displayed a gradient-like increase in greenness, whereof LHCFm 1.10 was light-brown and resembling the WT, LHCFm 6.1.11 was light-brown/dark-green, and LHCFm 15.1 was light-green (Fig. 4.2). Primary Sanger-based screens also indicated that these lines countained multiple *lhcf* knockouts which occurred in the same gradient-like pattern; LHCFm 1.10 with KO of *lhcf1*, LHCFm 6.1.11 with KO of *lhcf1*, *2* and *5*, and LHCFm 15.1 with KO of *lhcf1*, *2*, *3*, *4* and *5*.

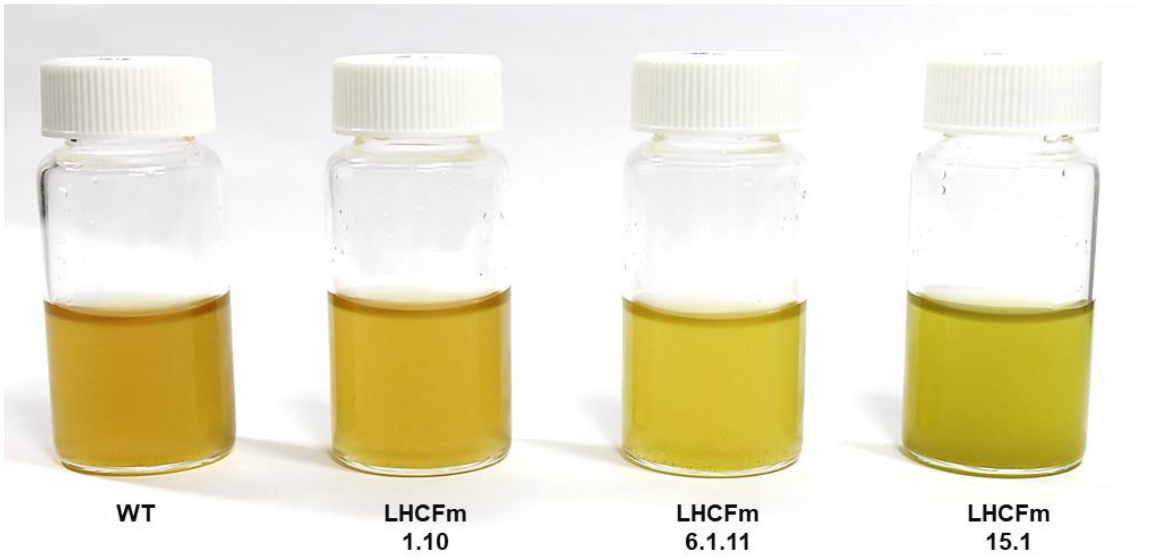


Fig. 4.2: Phenotypes of LL-acclimated *P. tricornutum* WT and LHCFm 1.10, 6.1.11 and 15.1 cultures. Cell density was 10^7 cells mL^{-1} in all samples. Photograph was taken with a Canon EOS 650D mirror reflex camera.

4.2.2 Summary of editing effects at DNA level

The LHCFm 1.10, 6.1.11 and 15.1 mutants were all generated using the standard diaCas9 coupled with mPAM1. As such, editing was expected to occur in on-target sites (*LHCF1*, *LHCF3*, *LHCF4*, m = 0), to some extent in off-target sites (*LHCF2*, *LHCF5*, m = 1) and possibly in the more elusive *LHCF11* (m = 3). The selection was of three mutants whom ended up with a gradient-like distribution of loss-of-function genes (Fig. 4.3). The less perturbed LHCFm 1.10 contained a knock-out of *lhcf1*, and an in-frame gene fusion-product of *LHCF3* and *LHCF4*. Both allelic copies of *LHCF2* were also confirmed to contain 1 bp deletions. However, in a later stage TOPO-TA based sequencing reaction, one of these copies were found to be WT – indicating that one of two non-functional *lhcf2* copies was re-edited and repaired to a perfect WT sequence. LHCFm 6.1.11 contained knock-outs of *lhcf1*, *lhcf2* and *lhcf5*, and a similar in-frame gene-fusion product of *LHCF3* and *LHCF4*. It is argued that the LHCFm 15.1 cell line lacks all copies of *lhcf1*, *lhcf2*, *lhcf3*, *lhcf4* and *lhcf5*. These genes have been subject to multiple attempts at PCR, with different lysates, enzymes and PCR-reaction protocols, but was never successfully amplified. It is therefore likely that the cell line has lost a large portion of these genes altogether, and/or recombined them in peculiar ways.

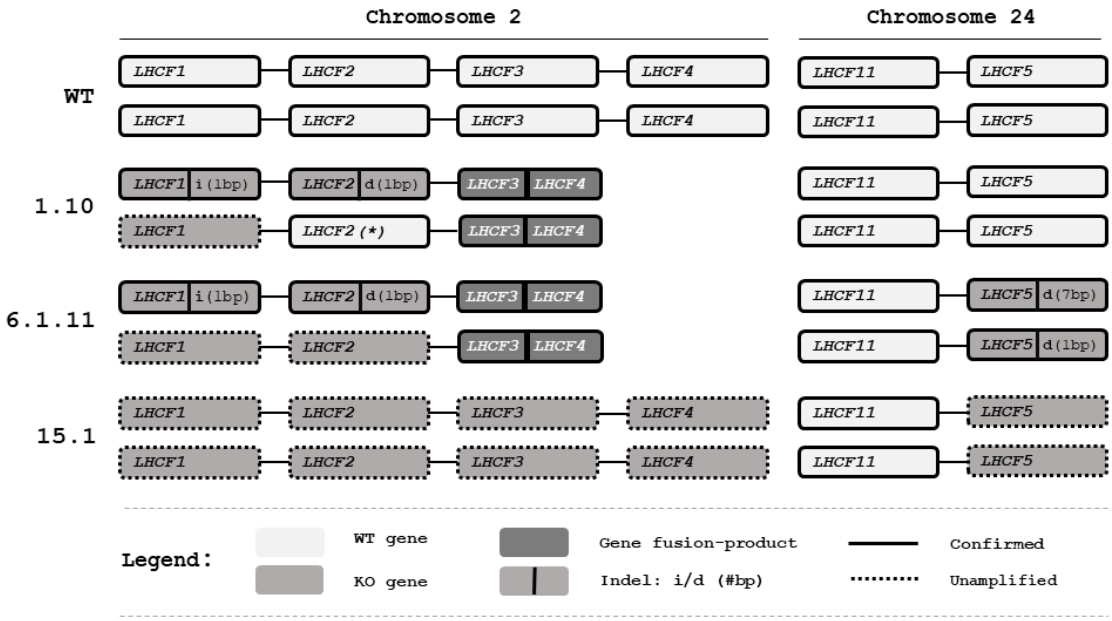


Fig. 4.3: Schematic overview of the genetic composition of targeted *LHCF* genes in LHCFm 1.10, 6.1.11 and 15.1. WT genes are white. Knocked out genes (KO genes) are light-grey and annotated by indel (i/d) and the base length of the indel (#bp) if known. In-frame gene fusion-products are labelled dark grey. Sequences confirmed by sequencing reactions are labelled with a solid border. Sequences we were unable to amplify by PCR have dashed borders. (*) denotes a re-edited copy of *LHCF2* in cell line LHCFm 1.10, in which a 1 bp deletion was repaired to a perfect WT sequence.

4.2.3 LHCF composition at the protein level

Western blots were conducted to semi-quantitatively assess the expression of LHCF1 through 11 at the protein level. Mutant protein lysates were compared to a dilution gradient (1:1, 1:2, 1:4) of WT protein lysates. The ATP-synthase β-unit (AtpB, 51.6 kDa, UniProtKB/A0T0D2) and photosystem II D1-protein (PsbA/D1, 39.6 kDa, UniProtKB/A0T0G9) were chosen as controls for loading quantity of lysates. While the loading quantity of WT controls were adequate, yielding a gradient of 100, 50 and 25% of the 10 ng protein quantity, only the LHCFm 1.10 mutant displayed an equal loading quantity to that of the WT 100 (Fig. 4.4). In fact, the intensities of LHCFm 6.1.11 and 15.1 control bands were more similar to the 1:2 dilution of WT (WT 50). To circumvent the problematics of inferring incorrect assessment of quantities, it was therefore attempted at comparing the LHCF bands of WT and LHCFm mutants which show a similar loading quantity (i.e. have equally pronounced control bands).

The controls of LHCFm 15.1 are arguably most equivalent to those of WT 50 in both LL and HL. Comparing the intensities of bands in the ~20 kDa range, it is evident that LHCFm 15.1 shows the most distinct loss of LHCFs at the protein level. Regardless of acclimation, the bands were un-detectable. LHCFm 6.1.11 should be compared to the loading quantity of WT 50 in HL. For LL-acclimated samples, the intensity of its controls fell somewhere between WT 50 and WT 25. Moreover, its LHCF bands rival that of the WT 50 intensity in both HL and LL, hinting that LHCFs were at least equally

expressed to WT in this cell line. LHCFm 1.10 showed similar band intensity to WT 100 in both LL and HL, and in loading controls and LHCF band alike.

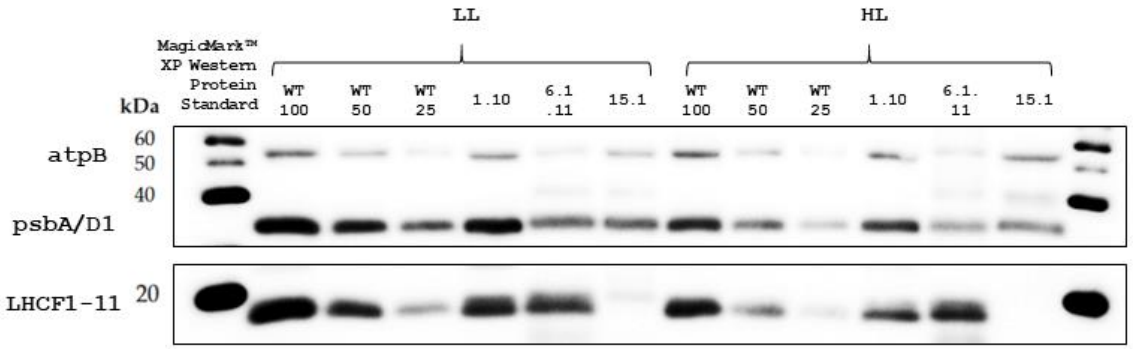


Fig. 4.4: Western blot analysis of LHCF1-11 proteins from WT and LHCFm mutant lines. LHCFm mutants 1.10, 6.1.11 and 15.1 were grown for two weeks in LL and HL before harvesting and isolation of total protein. Proteins separated on an SDS-PAGE gel were blotted on a nitrocellulose membrane, and polyclonal rabbit antibodies were used to target LHCF1-11, the ATP-synthase beta-unit (atpB) and photosystem II D1-protein (psbA/D1). The bands of a dilution gradient of WT protein lysates (WT 100, WT 50, WT 25) serves as reference for protein quantity at 100%, 50% and 25% of WT, respectively. MagicMark™ XP Western Protein standards (wells 1 and 14) report the protein mass (in kDa).

4.2.4 Effects of editing at the pigment level

With a lower amount of FCP proteins, less pigment should be anchored to the membrane. To test the effect of editing at the pigment level, cells were grown at HL and LL, harvested, and the pigments were extracted and quantified by HPLC. Because mutants LHCFm 6.1.11 and 15.1 were grown in a separate experiment, pigments were compared to one another, and reproduced as ratios of mol Fuco per. mol Chl *a* and DES indexes (Dtx / (Dtx + Ddx)) rather than mean values (mol cell^{-1}). In general, acclimation to LL or HL seemed to have little effect on the relative ratio of the light-harvesting pigments Fuco:Chl *a* (Fig. 4.5). However, and as expected from changed coloration and genetic composition of the LHCFm mutants, the level of Fuco:Chl *a* decreased with increased perturbance. The bright green LHCFm 15.1, which presumably had the most truncated antennas, showed the lowest number of Fuco:Chl *a* in HL ($\mu = 0.32$, $p = 0.0001$) and LL ($\mu = 0.39$, $p = 0.001$). The less green LHCFm 6.1.11, which compared to 15.1 – retained an in-frame *LHCF3-LHCF4* gene-fusion product, had slightly more Fuco:Chl *a* in HL ($\mu = 0.37$, $p < 0.001$) and LL ($\mu = 0.54$, $p = 0.228$). Lastly, LHCFm 1.10, of which phenotype is more brown than green, showed a slight decrease to Fuco:Chl *a*, although these values were non-significant when compared to WT.

The de-epoxidation state (DES, Dtx/Dtx+Ddx) reports on the relative expression of photoprotective pigments. As is evident from Fig. 4.5b, there was no pronounced cell line differences in the DES when grown in LL. In HL, the ratio of Dtx per Ddx were increased, leading to elevated DES index values. Moreover, there exist larger dissimilarities between the cell lines, in which WT DES ($\mu = 0.36$) is much higher than the DES values of LHCFm mutants ($\mu \sim [0.23, 0.25]$).

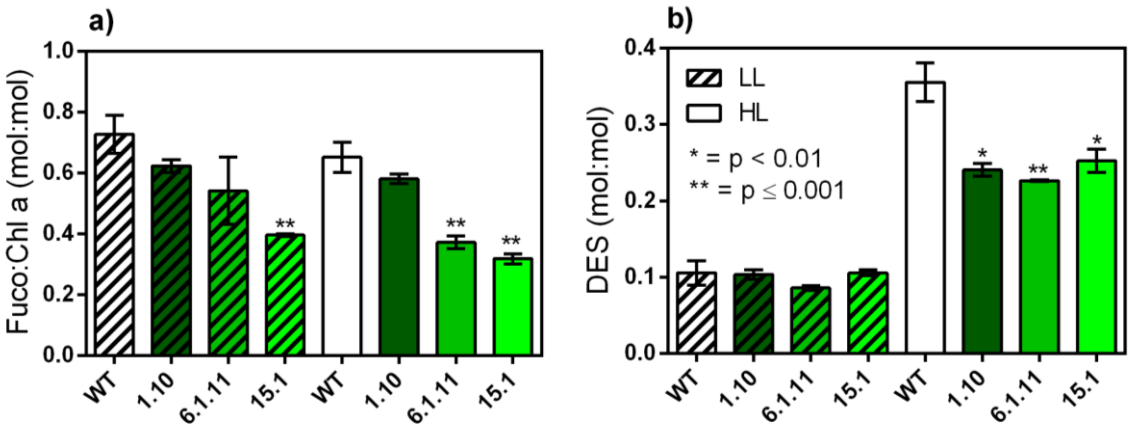


Fig. 4.5: Ratios of a) Fuco per. Chl *a* (Fuco:Chl *a*) and b) de-epoxidation state (DES, Dtx/(Dtx+Ddx)). All ratios are presented as means (μ) of n replicates of the relative unit mol:mol, with standard error of the mean (± SEM, n_{WT} = 9, $n_{1.10}$ = 6, $n_{6.1.11}$ = 3, $n_{15.1}$ = 3). Statistical significance was inferred using two-sided T. tests assuming heteroskedastic variance. Asterisks denotes significant differences with $p < 0.01$ (*) and $p \leq 0.001$ (**).

4.2.5 Proliferation rates

Cells were grown in triplicates for 7 days in LL (50 µmol photons m^{-2} s^{-1}) and HL (250 µmol photons m^{-2} s^{-1}) in F/2 medium and counted each day by Flow Cytometry. The maximum specific growth rate (k') was determined from cell numbers in exponential phase and used to calculate the maxima of divisions day^{-1} and doubling time (h, Table 4.3). WT, LHCFm 1.10 and 6.1.11 divided approximately equally fast in LL, whereas LHCFm 15.1 displayed slower growth. Cell division rate was increased in HL cultures to that of LL cultures. This was also true for LHCFm 1.10, but the two other mutants showed a decline in growth with increasing light intensity.

Table 4.3: The proliferation rates of WT, and LHCFm 1.10, 6.1.11 and 15.1 mutants. Maximum divisions day^{-1} and doubling time (h) are presented with mean values (μ) and standard error of the mean (SEM, n = 3).

	Divisions day^{-1}	±SEM	Doubling time (h)	±SEM
WT LL	1.93	0.04	12.46	0.26
1.10 LL	1.93	0.04	12.47	0.29
6.1.11 LL	2.02	0.02	11.87	0.13
15.1 LL	1.67	0.02	14.36	0.14
WT HL	2.11	0.16	11.53	0.95
1.10 HL	2.19	0.06	11.00	0.32
6.1.11 HL	1.82	0.10	13.30	0.71
15.1 HL	1.31	0.02	18.39	0.23

4.2.6 Decreased Chl *c* and $Fuco_{RED}$ absorbance

Cells acclimated to LL (50 µmol photons m^{-2} s^{-1}) and HL (250 µmol photons m^{-2} s^{-1}) were sampled and applied in spectrophotometry. The absorbance was calculated from detected light and corrected for by

a factor of the side-scattered light. Moreover, spectra were normalized by the max-min method and plotted against each other (Fig. 4.6a and b, respectively). Finally, difference spectra were calculated by subtracting the mutant absorbances from the WT (Fig. 4.6c and d). Differential spectra of all LHCFm mutants are presented in Suppl. Fig. 4 of Appendix D.

The spectra of mutants showed great similarity to that of the WT culture in both LL and HL. The major differences were identified in the LL and HL spectra of LHCFm 15.1 which showed decreased absorbance in the blue-green part of the spectrum (~450-570 nm). This decreased absorbance peaked at the wavelengths of 461 and 528 nm when LHCFm 15.1 was grown in LL, and at 460 and between 520 and 540 nm when grown in HL.

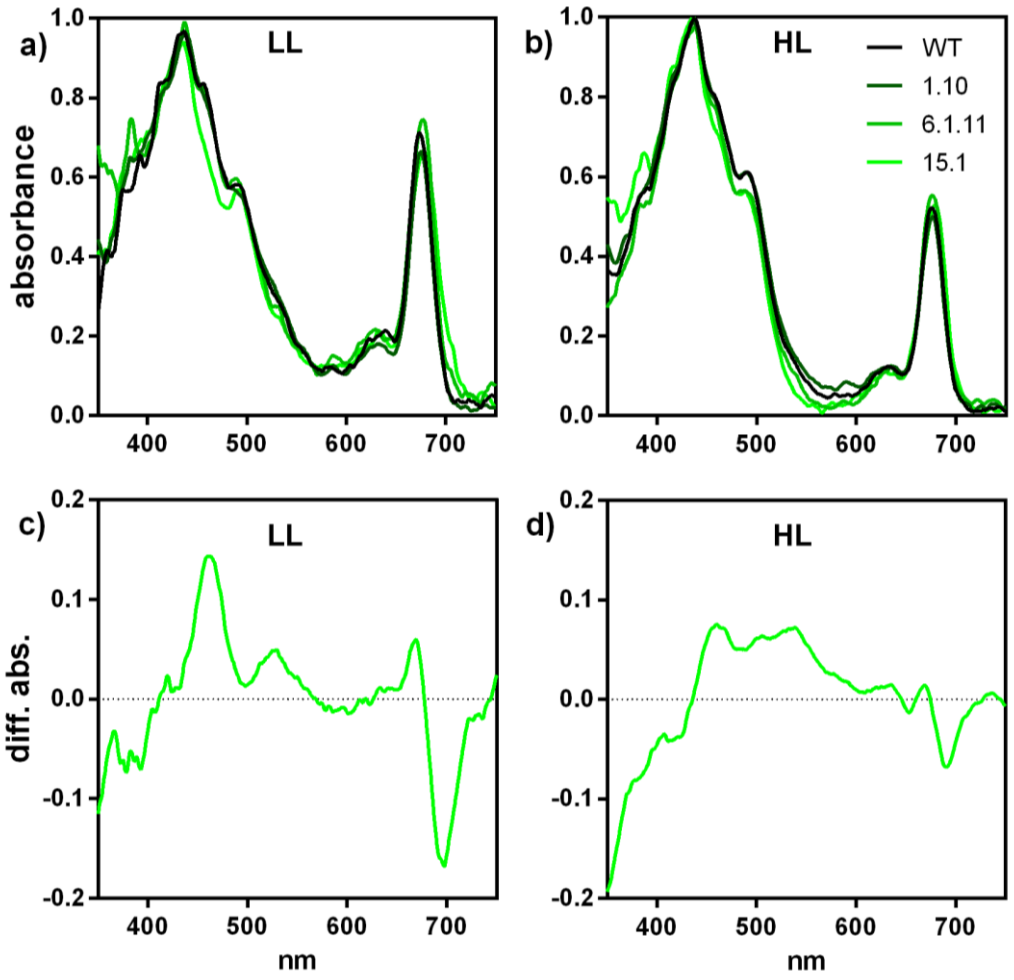


Fig. 4.6: Absorbance spectra of WT and LHCFm 1.10, 6.1.11 and 15.1 mutants and differential absorption of WT contra LHCFm 15.1. Absorbance was recorded for cells acclimated to a) LL (50 µmol photons m^{-2} s^{-1}) and b) HL (250 µmol photons m^{-2} s^{-1}). All absorbance-curves are corrected for side-scattered light and normalized by the min-max method. Differential absorption spectra (c, d) provides the difference in absorption by LHCFm mutant 15.1 and WT in LL (c) and HL (d).

4.2.7 Variable fluorescence

The Chl *a* variable fluorescence was used to assess the relative electron transport rate (rETR) of WT and mutant lines in LL and HL acclimated cells. rETR was obtained by multiplying the quantum yield of PSII photochemistry (Φ_{PSII}) with the associated irradiance (E_{PAR}). Photosynthesis versus Irradiance (PE) curves were plotted with rETR as functions of increasing irradiance (E_{PAR}, Fig. 4.7). For all lines except LHCFm 15.1, rETR increased when cells were acclimated to HL. This disparity between LL- and HL-acclimated cells was most profound in the curves of WT and LHCFm 1.10, but also conspicuous in the curves of LHCFm 6.1.11, albeit to a lower degree. On the contrary, LHCFm 15.1 cells had a lower $rETR_{max}$ when acclimated to HL than when acclimated to LL, but this difference may be negligible as the curves were almost identical. Furthermore, 15.1 LL and HL curves are more or less united towards higher light-intensities (>800 umol photons m^{-2} s^{-1}).

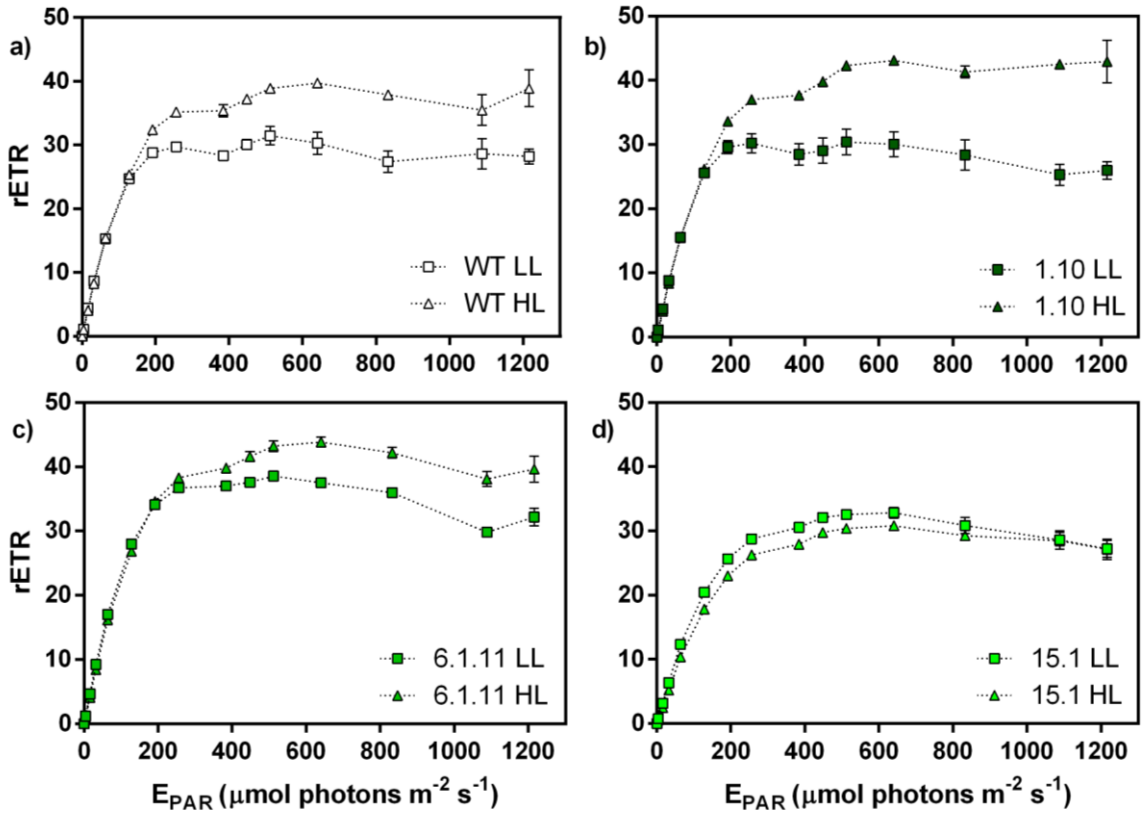


Fig. 4.7: Photosynthesis (rETR) versus Irradiance (E) curves for LL- and HL-acclimated cells of a) WT and LHCFm mutants b) 1.10, c) 6.1.11 and d) 15.1. LL-acclimated samples are shown as squares, and HL-acclimated samples as triangles. All datapoints are the mean of triplicates with standard error of the mean (±SEM, n = 3).

4.2.8 Light curve coefficients and photosynthetic efficiency

The light-utilization coefficient (α) equals the slope of the PE-curve, and reports on how quickly the samples reach light-saturation. No major differences were observed when comparing the α-values of

LL with HL (Fig. 4.8a). However, a major difference appears when comparing the α of LHCFm 15.1 with those of the other cell-lines. Its α is greatly reduced both in LL ($\alpha = 0.22$) and HL ($\alpha = 0.20$), when compared to WT and other LHCFm mutants ($\alpha \sim [0.29, 0.31]$).

The maximum relative electron transport rate ($rETR_{max}$) denotes the capacity of photosynthesis. In general, this capacity should be higher in cells acclimated to HL (Fig. 4.8b). However, this apprehension did not hold for all cell lines, and the LHCFm 15.1 mutant showed no great distinction between LL and HL $rETR_{max}$. In fact, these recordings showed a slightly decreased $rETR_{max}$ when 15.1 cells are acclimated to HL. In opposite fashion, LHCFm 6.1.11 stood out by having $rETR_{max}$ values significantly increased to that of the WT, both when acclimated to LL and HL.

The light saturation index (E_k) describes the light-intensity threshold at which the photosystems become fully saturated [154]. In theory, HL-acclimated cells will show an increased E_k to that of LL-acclimated cells. This because acclimation to increased irradiance favours a lowered ratio of antenna molecules per. reaction centre. While this seemed the case for both WT and LHCFm 1.10 (Fig. 4.8c), the effective difference between LL and HL E_k became increasingly small in LHCFm 6.1.11 and 15.1. In the latter, the E_k values of LL- ($\mu = 149$) and HL-acclimated cells ($\mu = 153$) were practically indifferent, especially if taking the standard error into account.

The maximum quantum yield of photosystem II ($\Phi_{PSIImax}$), often referred to as Fv/Fm, measures the proportion of absorbed light going into PSII photochemistry when all RCs are open. Thus, the maximum quantum yield of photosystem II ($\Phi_{PSIImax}$) allows us to measure how efficient PSII photochemistry is at its peak. From Fig. 4.8d it was apparent that WT and LHCFm mutants 1.10 and 6.1.11 shared the same PSII efficiency ($\mu \sim 0.67$ in LL, $\mu \sim 0.59$ in HL). Mutant LHCFm 15.1 however, had a greatly reduced $\Phi_{PSIImax}$ in both LL ($\mu = 0.43$) and HL ($\mu = 0.31$).

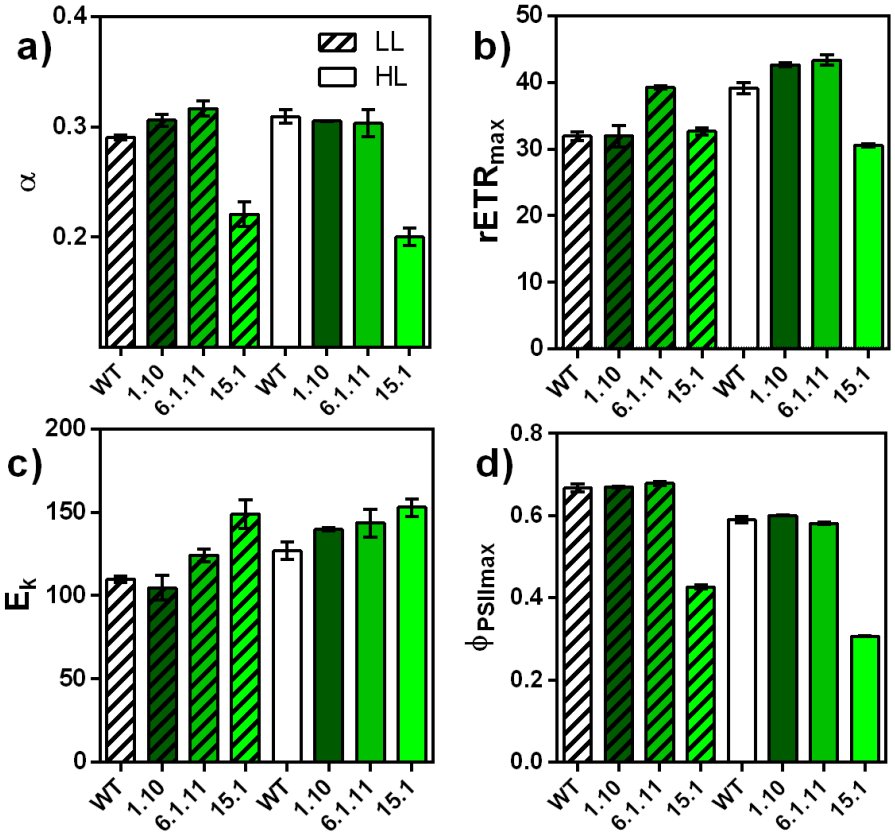


Fig. 4.8: Photosynthesis parameters for WT and LHCFm mutants grown in LL (dark grey pattern) and HL (no pattern). Variable fluorescence data gained by PhytoPAM fluorometry was used to calculate the following light-curve parameters: a) the light utilization coefficient (α), b) the maximum relative electron transport rate ($rETR_{max}$), c) the light saturation index (E_k) and d) the maximum quantum yield of photosystem II photochemistry ($\Phi_{PSIImax}$). All bars display the mean of triplicates with standard error of the mean (± SEM, n = 3).

4.2.9 Non-photochemical quenching (NPQ)

Data obtained by PAM-fluorometry of LL-acclimated cultures were used to calculate non-photochemical quenching (NPQ) of Chl fluorescence (Fig. 4.9). NPQ describes the capacity/need to dissipate excessively absorbed energy as heat. In all cell lines, NPQ increased as a function of increasing irradiance. LHCFm 1.10 NPQ matched the high values and steady increase as expressed by the WT. LHCFm 6.1.11 showed slightly lower, but steadily rising NPQ values. LHCFm 15.1 displayed a greatly decreased NPQ, which was very slowly rising with the irradiance.

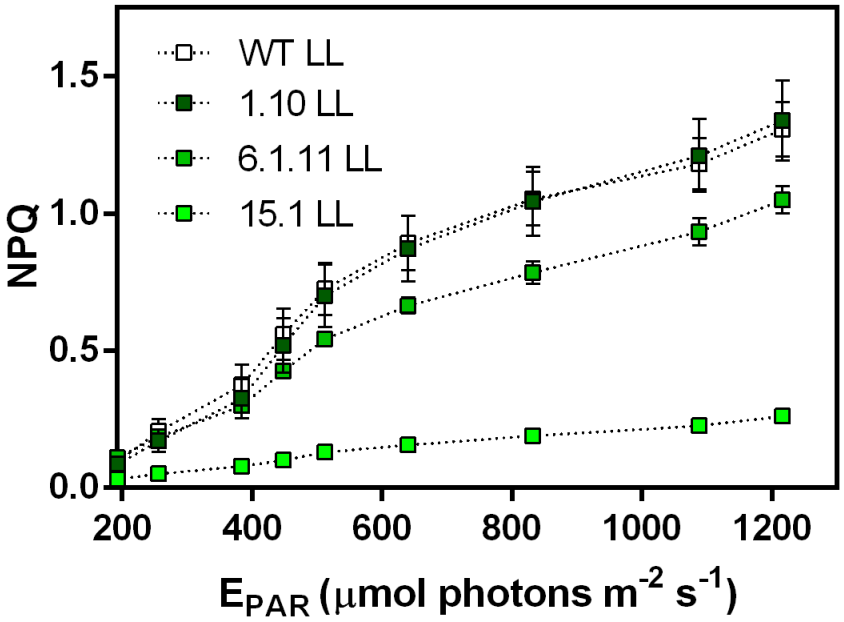


Fig. 4.9: Non-photochemical quenching (NPQ) in LL-acclimated WT and LHCFm mutants as a function of increasing irradiance (E_{PAR} ~ [192, 1216] µmol photons m^{-2} s^{-1}). All datapoints are the mean (µ) of triplicates, shown with standard error of the mean (±SEM, n = 3).

4.3 Growth trials in photobioreactors

Growth trials in photobioreactors were conducted to test the TLA-theory out in practice, and to screen for suitable TLA mutants. Six different cell lines were picked, of which three were the LHCFm 1.10, 6.1.11 and 15.1 mutants described previously, and three were *Albino 3b-insertase* (*alb3b*) KO mutants. The *alb3b*-mutants had green phenotypes (Fig. 4.10b), approximately 25% of the LHC contents of WT cells, and ~65% and 35% reduction to Fuco and Chl *a* (Nymark et al., Unpublished). The cultures were inoculated at 100 000 cells mL^{-1} and grown at equal light intensity (225 µmol photons m^{-2} s^{-1}) for 14 days.

4.3.1 Growth and density

Culture density was measured each day by flow cytometry or hemocytometry. The resulting growth curves (Fig. 4.10) resemble those of other microorganisms, although the lag-phase was surpassed by keeping the inoculate cultures in exponential phase for a week prior to the experiment. Thus, cultures jump-started into exponential phase. Both replicates cultures of WT cells grew quicker in the early stages, but the division rate declined already after 72 hours of cultivation. The mutants grew slower, having a less exponential, but steady growth from inoculation to 96 hours. At the influx of additional CO_2 (96 hours), all cultures saw a momentary rise in their growth rate, as if the exponential phase was somewhat extended. A transition from exponential to early stationary phase likely occurred in all cultures after 168 or 192 hours. After applying additional nutrients (at 216 hours for WTA, and 264

hours for all other cultures) cell densities continued rising, albeit at moderate rates compared to exponential phase. The final stages of the experiment were characterized by a slight decline of growth, and generally it was conceived that WT cultures met decline earlier, while mutants retained a slow increase.

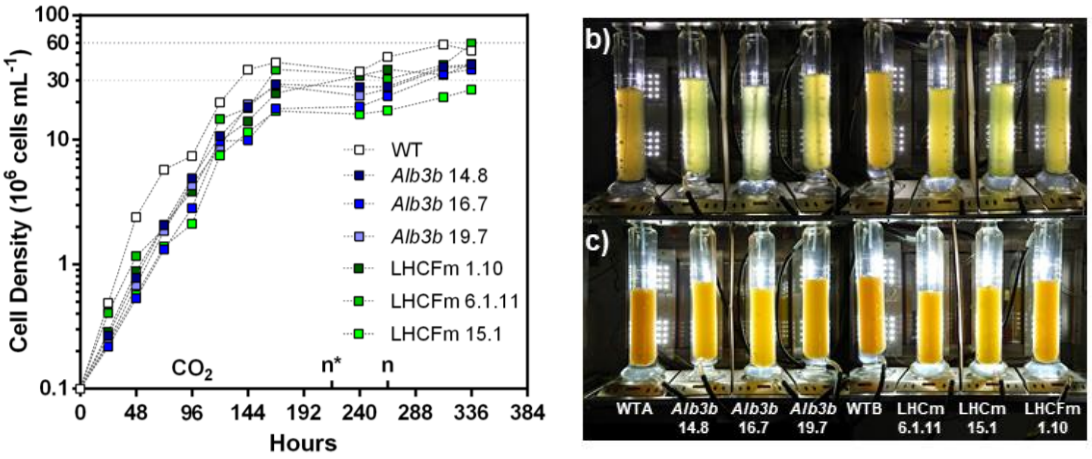


Fig. 4.10: Cell density and culture colouration over time. a) Evolution of cell density in WT (mean of two replicates) and mutant LHCFm (1.10, 6.1.11 and 15.1) and *alb3b* (14.8, 16.7 and 19.7) cultures over time. Dashed lines indicate injection of CO_2, or addition of nutrients (n). n* indicates an addition of nutrients to a declining WT (WTA) culture. n indicates addition of nutrients to the remaining seven cultures. b) and c) are pictures taken of the photobioreactors at day 5 (120 hours) and day 9 (216 hours), respectively.

The maximum cell density was identified along with the maximum specific growth rate (k') of the exponential phase. k' was used to calculate the maxima of divisions day^{-1} and doubling time (h, Table 4.4). Here, the rate of growth was distinctively increased in WT cells (Divisions day^{-1} = 2.29, Doubling time = 10.49 h) compared to any of the LHCFm or *alb3b* mutants (collectively; divisions day^{-1} ~ [1.24, 1.46], doubling time ~ [16.44, 19.37] h). The maximum cell density (cells mL^{-1}) was in general extremely high (Fig. 4.4), especially in WT (58.8 mill. cells mL^{-1}) and LHCFm 6.1.11 (59.1 mill. cells mL^{-1}) cultures.

Table 4.4: The maximum cell density, divisions day^{-1} and doubling time (h) of WT, a*lb3b* and LHCFm mutants during photobioreactor growth trials. Maximum cell density numbers are based on hemocytometry cell counts from final stages of the experiment. Divisions day^{-1} and doubling time (h) are calculated from the specific growth coefficient (k') in exponential phase. WT numbers are the average of 2 biological replicates.

Cell line	Max. cell density (mill. cells mL^{-1})	Divisions day^{-1}	Doubling time (h)
WT	58.8	2.29	10.49
alb3b 14.8	40.3	1.46	16.47
alb3b 16.7	36.7	1.24	19.37
alb3b 19.7	39.3	1.40	17.10
LHCFm 1.10	40.6	1.46	16.44
LHCFm 6.1.11	59.1	1.42	16.88
LHCFm 15.1	25.3	1.27	18.97

4.3.2 Dissolved Inorganic Nitrate + Nitrite and Phosphate

Liquid media filtrate was sampled for nutrient analyses throughout the experiment to get an idea of how DIN and DIP availability changed over time (Fig. 4.11). From an initial quantity of 36.01 mg L^{-1}, DIN became almost completely depleted by 144 hours in all cultures with exception of the *alb3b* 16.7 culture (DIN_{144} = 1.96 mg L^{-1}, Fig. 4.11a).

An initial quantity of 3.35 mg L^{-1} DIP was decreased to 0.003 (WTA) and 0.063 mg L^{-1} (WTB) after just 72 hours in WT cultures (Fig. 4.11b). At the same time (72 h), the DIP reservoir was close to 50% of its initial quantity in mutant cultures, with the exception of LHCFm 6.1.11, in which DIP removal looked more resembling to WT cultures. By 144 hours, DIP was practically exhausted in all cultures. Comparing the depletion times of DIN and DIP, it is apparent that DIP was taken up earlier, and that both were exhausted earliest in the WT cultures.

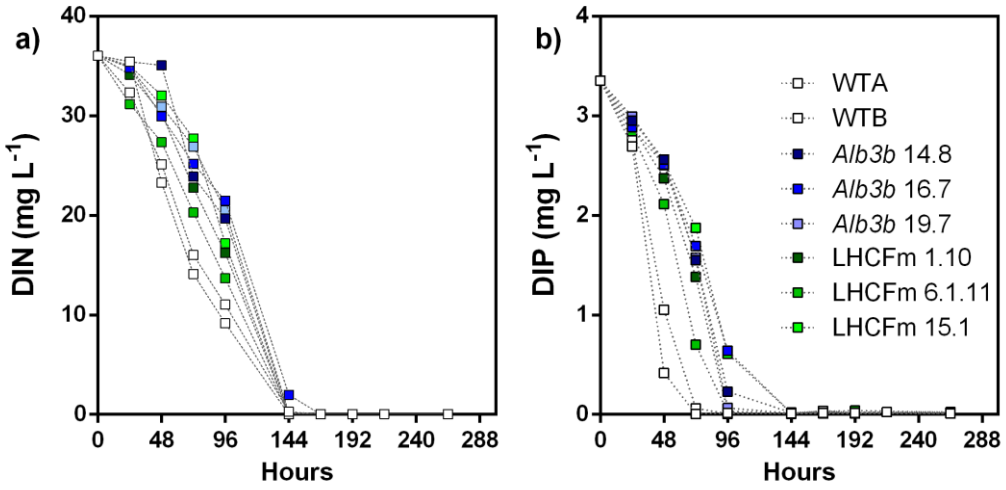


Fig. 4.11: Liquid media quantities of a) DIN and b) DIP over time. Notice the 10-fold difference in y-axis values. Data from both WT replicates are shown because they were sampled at different times.

4.3.3 Organic nutrients

Uptake, and re-mineralization of dissolved nutrients to organic biomolecules plays a fundamental role in all living organisms. These are the molecules that constitute DNA, RNA, proteins, pigments, fats, energy-rich carriers and carbohydrates. Their utilization is therefore key in proliferation, and when considering variations in their productivities.

Carbon quantity was measured by element analysis and corrected for by the harvested cell number. For most of the cultures, the temporal evolution of carbon per cell (pg $cell^{-1}$) was characterized by an intermediate minimum after 168 h (apart from *alb3b* 16.7; minimum at 96 h), and an early or late (264 h) maximum (Fig. 4.12a). WT cultures saw their maximum sequestered carbon at the early stage, while most of the mutants saw theirs in the final stages of growth (264 h). The evolution of nitrogen per cell

(Fig. 4.12b) appeared somewhat analogous to that of carbon. Although the masses of the nutrients were vastly different, the datapoints were arranged similarly – indicating that carbon and nitrogen composition correlated. Also, nitrogen was at higher quantities in early stages, then plummeted at an intermediate timepoint (168 h). However, as opposed to the carbon evolution, nitrogen remained at relatively low quantities in the final recordings (264 h), seeing small decreases or increases from the intermediary stage. Both final carbon and nitrogen quantity is highest in LHCFm 15.1 and lowest in WT cultures. Phosphorous levels varied greatly at early exponential phase, spanning from 0.40 to 0.84 pg cell^{-1} (Fig. 4.12c). During the exponential phase, phosphorous levels decreased rapidly, and stabilized at less than 0.1 pg cell^{-1} at the later stages of all cultures. Apparently, phosphorous was exhausted earliest in the WT cells. Lastly, the ratios of N:P were low to begin with but saw an increase in all cultures at an intermediate stage (168 h, Fig. 4.12d). WT cultures in particular were found to encroach extremely high N:P ratios (62, 93, 168 h), while mutant cultures displayed a mediocre increase into the N:P co-limitation range (25-33) [155].

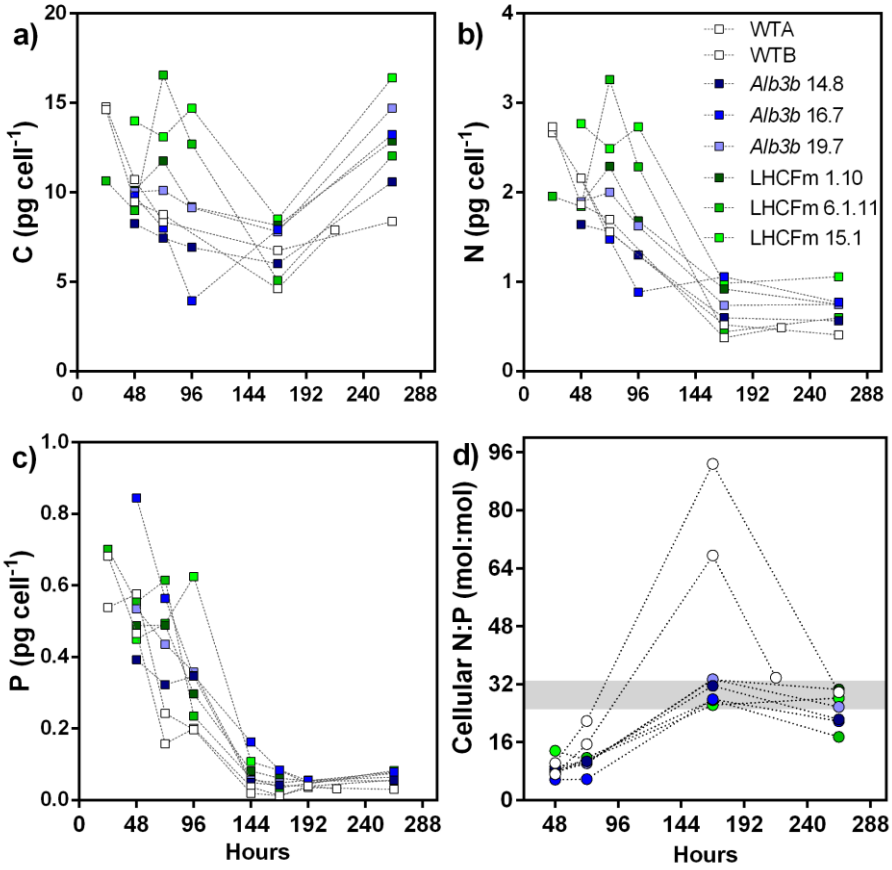


Fig. 4.12: Cellular organic a) C, b) N and c) P and d) N:P ratio over time. All quantified amounts were corrected for by the number of harvested cells. N:P ratios denote the molar ratio of N and P per cell. The critical range (25-33) for N and P co-limitation in *P. tricornutum* cultures [155] is labelled with a grey bar.

4.3.4 Total fats per dryweight

Total fat content was assessed using the Bligh & Dyer method [153]. Because fatty acid metabolism is expected to be affected by conditions of growth (nutrient access, pH, salinity etc.) and the status of the culture population (i.e. a CO_2-deprived culture might oxidise fats for energy), these values may differ not only due to the effects of a particular genetic manipulation. The data should therefore be approached with caution. In touch with this notion is the observation that there is no apparent trend. WT fatty acid content (μ = 13.6%, n = 4) was lower than that of *alb3b* 14.8 (μ = 18.6, n = 2) and slightly higher than that of *alb3b* 16.7 (μ = 13.2, n = 2). Likewise, fatty acids per dry weight was lower in cultures of LHCFm 15.1 (μ = 11.27, n = 2) and higher in LHCFm 6.1.11 (μ = 18.4, n = 2). The most profoundly different values were those of *alb3b* 14.8 and LHCFm 6.1.11, in which fatty acid content was increased to 18.6 and 18.4% of the DW, respectively.

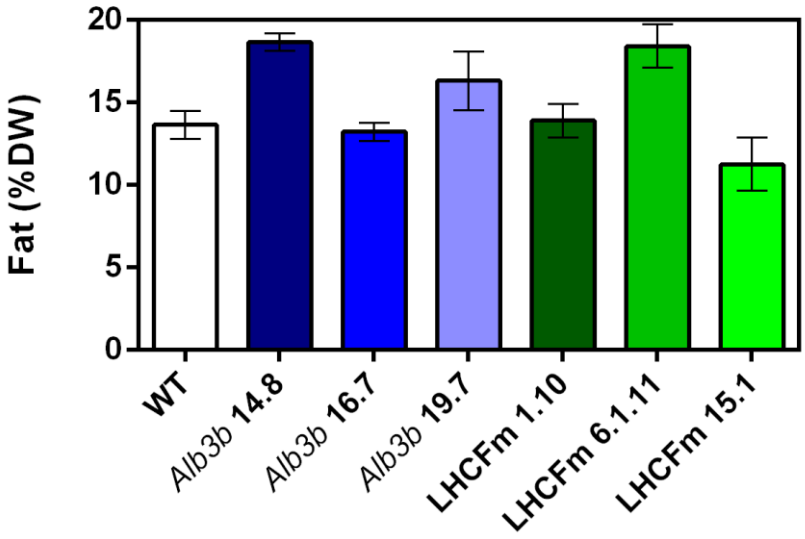


Fig. 4.13: Total fats as a per cent of dry weight (% DW). Dashed line indicates a "baseline" constituted by the WT % DW. Values are presented as the means of n mechanical replicates (n = 4 for WT cultures, and n = 2 for remaining cultures) with standard error of the mean (± SEM).

4.3.5 Maximum Quantum yield of PSII photochemistry ($\Phi_{PSIImax}$)

Besides a measure on the maximal proportion of absorbed light entering PSII photochemistry, $\Phi_{PSIImax}$ offers insight on the physiological state of algae. In general, fit and healthy cells whom retain a highly functional photosynthetic apparatus score higher [145]. *alb3b* mutants appeared to convey the most stabile $\Phi_{PSIImax}$ signals, maintaining a high $\Phi_{PSIImax}$ in the range of 0.49 to 0.70 throughout the experiment (Fig. 4.14). LHCFm 1.10 and 6.1.11 displayed WT-like curves, but their $\Phi_{PSIImax}$ fluctuations were not as pronounced as the WT. In comparison, the initial $\Phi_{PSIImax}$ values of LHCFm 15.1 were extremely low (0.44, 0 h), were surpassed 5 days later (0.49, 120 h) and reached a maximum at the final recording (0.50, 312 hours) that signifies a great increase from the onset of additional nutrients (0.22, 264 h).

While all cultures saw an increase in the $\Phi_{PSIImax}$ with additional CO_2 influx (96 h), the cultures responded particularly well with new nutrients (216 and 254 h). In fact, the results presented indicate that the steady decreases in $\Phi_{PSIImax}$ that started from 120 hours came to a sudden stop at nutrient onset.

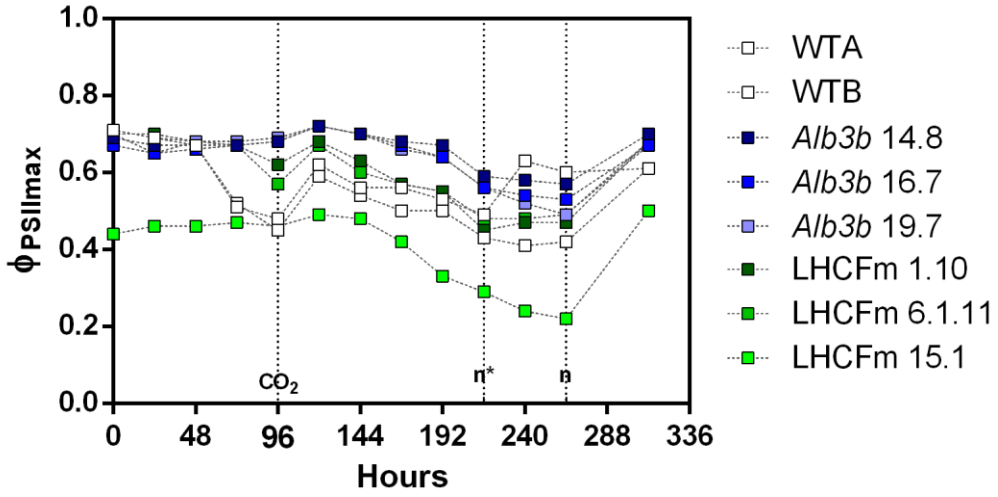


Fig. 4.14: The maximum quantum yield of photosystem II ($\Phi_{PSIImax}$) over time. Dashed lines labelled CO_2 or n indicate additional influx of CO_2 or nutrients, respectively. n* indicate additional nutrients to WTA.

5 Discussion

5.1 HifiCas9 – a trade-off between activity and high fidelity

The first objective of this thesis was to test whether a modified hifiCas9 would avoid off-target mutagenesis to a larger extent when applied in diatom genome editing. This effect was previously reported by Kleinstiver and co-workers [83] who tested different hifiCas9 modifications in a human cell line. They found that >70% of the on-target activity was retained while off-target events were much more infrequent when replacing four non-specific DNA-interacting residues of the Cas9 peptide with alanines. When transforming diatoms with two gRNAs and a hifiCas9 that followed the same general principles, the combined efficiency was lower at on-target sites (~75% to that of the diaCas9), and only one off-target event was identified. From these results it is argued that hifiCas9 does have higher fidelity, but that it comes at the cost of overall activity. Nevertheless, hifiCas9 may prove a useful alternative to the standard diaCas9, especially in applications that require high specificity. An in-depth discussion of the above described results is presented below.

5.1.1 Many transformants but few mutants.

The overall transformation efficiency achieved with biolistic transformation (~$4.5 \cdot 10^{-7}$ transformants $cell^{-1}$) was almost a 10-fold higher than those reported in recent literature (5-$25 \cdot 10^{-8}$ transformants $cell^{-1}$) [47]. In theory, 60-70% of these should have both constructs [31] and in this case both pAF6 and either of diaCas9 or hifiCas9. However, secondary screening by amplification of a Cas9:gRNA fragment showed that only 20 of the 248 transformants contained both selection vectors and at least one fragment of the Cas9-conferring plasmids. These numbers amount to a co-transformation efficiency of 8.1%. Why so few transformants contained both constructs are not clear. One possibility is that pAF6, which was added first during the preparation of microcarriers, may have saturated the tungsten particles, essentially blocking the Cas9-conferring plasmid copies from binding. Also, if by mistake pAF6 was added at a higher quantity than the Cas9-conferring plasmids, this blocking reaction might have been favoured. This would moreover coincide well with a relatively high number of transformants and a low number of mutants.

It is important to stress that detection of the Cas9:gRNA fragment does not confirm a fully operational Cas9:gRNA system. It is, however, an efficient method to lower the costs and labour of screening when multiplexing or when targeting multiple genes. By amplifying a piece of the Cas9:gRNA construct, we efficiently located the transformants which at least have had a copy of the Cas9-plasmid, without regards to its integrity as being intact or fragmented. As could be expected, mutations were not found in all putative mutants. In fact, only three of 20 putative mutants were found to contain any indel at the on-target sites, indicating that a clear majority of putative mutants contained fragmented or somehow

dysfunctional Cas9-systems, or that mutagenesis for some other reason did not take place. Importantly, these efficiencies did not favour one Cas9 enzyme over the other.

5.1.2 High fidelity or lack of activity?

When reviewing the data presented in Table 4.2 and Fig. 4.2, it is evident that hifiCas9 was active and capable of inducing mutagenesis at both on- and off-target sites. When primed with mPAM1, the hifiCas9 enzyme reproduced approximately 79% of the on-target efficiency of WT diaCas9 and did not induce any off-target mutations. Conversely, diaCas9:mPAM1, induced off-target editing of *LHCF2* in all cell lines (5/5) and *LHCF5* in 4 out of 6 cell lines, indicating that while hifiCas9 to a large extent avoided off-target editing, diaCas9 did not. Larger discrepancies were observed when using hifiCas9:mPAM2, which induced mutagenesis at the on-target site of *LHCF1* (in 5/5 cell lines) and *LHCF3* (4/5) at similar activities to that of diaCas9, but only to 20% of diaCas9 in the on-target site *LHCF4* (1/5). Summarized, these values amount to 71% of the on-target activity observed with diaCas9:mPAM2. With both gRNA, the on-target activity of hifiCas9 amounted to approximately 75% of the WT diaCas9 on-target efficiency, an efficiency that agrees well with the >70% activity reported by Kleinsteiver and co-workers [83]. These observations support that the modifications to hifiCas9 did in fact remove some of its innate catalytic energy, effectively leading to higher-fidelity coming at the cost of overall activity. Because the reduced activity is an expected outcome of lowering the DNA affinity of Cas9, the question becomes whether the lowered on-target activity is a tolerable downside with regards to increased precision.

Evidence was found that supports a lower, but still present off-target potential. When primed with mPAM2, hifiCas9 has induced mutagenesis at *LHCF2* at least once in the LHCFm 2.1 cell line. Additionally, it was impossible to amplify the *LHCF2* of LHCFm 6.1, but in this cell line there existed a fusion-product of *LHCF1* and *LHCF3*. Since *LHCF2* is spatially located between these target sites, it is conceivably more probable that *LHCF2* was deleted rather than edited. However, the occurrence of editing in *LHCF2* with hifiCas9:mPAM2, and absence of editing at *LHCF2* with hifiCas9:mPAM1 was peculiar. Our current understanding of off-target activity underlines the importance of the mismatched nucleotide location, emphasizing that non-seed mismatches may be tolerated to a larger extent than in-seed mismatches [69]. This makes the occurrence of hifiCas9:mPAM2-mediated editing of *LHCF2* more severe, simply because *LHCF2* was an off-target site with an in-seed mismatched nucleotide (see Table 2.1). Moreover, observations that support this current consensus are manifested in the activity of diaCas9 when primed with mPAM1, whereof ~100% of the cell lines have induced mutations in *LHCF2* (non-seed mismatch) and ~67% have mutations in *LHCF5* (in-seed mismatch). Furthering this investigation, one can ponder on why hifiCas9:mPAM2 instigates mutagenesis in *LHCF2*, while avoiding the same gene when primed with mPAM1. Based on the arguments just provided, *LHCF2* should be an “easier” target when guided by mPAM1, given that its mismatched nucleotide is located

in the non-seed region. However, this peculiarity may just be random effect as the hifiCas9:mPAM1 lines have been difficult to generate, such that only four pure lines were isolated.

5.1.3 Key notes on high-fidelity Cas9

The results indicated that hifiCas9 has increased fidelity, but that it comes at a cost of overall enzymatic activity. The >70% on-target efficiency observed using SpCas9-HF1 [83] was surpassed by this work (~75%). However, as a consequence of the laborious and time-consuming process of transformation and PCR-based screening, this study became limited by the number of replicate cell lines. Thus, only a fraction of observations lies at the foundation of this indicator. In contrast, a high number of cell lines coupled with high-throughput screening allowed previous studies [83,84] to put the high-fidelity strategy through vigorous testing. Moreover, the spatial closeness of the *LHCFs* further complicated the screening process by introducing various mutation events such as fusion products and inversions.

With lowered overall activity, one can expect a lower number of mutants, but the chance of isolating mutants with off-target effects should be smaller, thus making hifiCas9 an alternative to the more promiscuous diaCas9. It is argued that it might be particularly beneficial when targeting genes that share high DNA sequence similarity to others, such as the *LHCFs* homologs that occur in 17 highly similar copies.

DSB and subsequent mutagenesis relies on a complex interaction between gRNA, Cas9, PAM-site [67], DNA [68], and is furthermore determined by the location of potential mismatched nucleotides within in-seed or non-seed target regions [55,79]. Modifications to Cas9 is therefore one of many possible steps towards increasing the fidelity of genome editing, and a refinement that enables complete off-target avoidance and maintains on-target activity would topple this strategy.

5.3 Characterization of LHCFm mutants and their similarity to TLA mutants

The second objective of this thesis was to characterize LHCFm mutants and to determine whether they exhibit traits that are typical of a cell line with TLA. Green algae and cyanobacterial TLA mutants have previously been produced by deletion of for example the CpSRP43 homolog in *C. rheinhardtii* [133], point mutation of the same homolog in *C. vulgaris* [134], by disruption of the CpSRP54 homolog in *C. rheinhardtii* [156] and by knock-in replacement of the CPC-operon in *Synechocystis sp.* [135].

Although the photosynthetic machinery of cyanobacteria, green algae and diatoms are inevitably different given the origins of "simple" (cyanobacteria, green algae) and "complex" diatom chloroplasts, the purpose of their antennas remain the same: to harvest light energy and funnel it to the RCs. Thus, some of the common photophysiological traits reported for these TLAs were used as references to what could be expected of diatom LHCFm mutants. Typical traits include pale coloration caused by lower amounts of antenna proteins and pigments per cell, and elevated photosynthetic activity and light-saturation levels [133,134,156].

5.3.1 Colouration may correlate with antenna size

The mutants chosen for characterization displayed a gradient-like increase in greenness with perturbance of *LHCFs*. LHCFm 15.1 was identified as the presumably most perturbed mutant from having a bright green phenotype and from the possibility of containing quintuple *lhcf* KOs. It was followed by LHCFm 6.1.11 which also displayed a green colouration (albeit less distinct), and with fewer KO mutations (*lhcf1, 2* and *5*). While LHCFm 1.10 was green at an early stage, it turned brownish, and resembled that of WT after 4 weeks of liquid medium cultivation (prior to all analyses). Initial screening confirmed that LHCFm 1.10 contained biallelic KO mutations in both *lhcf1* and *lhcf2*, and a gene fusion product of *LHCF3* and *LHCF4* as a result of a large deletion between these target-sites. However, one allelic copy of *LHCF2* was later found to have become re-edited to a perfect WT sequence. With the gene fusion product being in-frame and likely fully operational, LHCFm 1.10 became the least perturbed mutant furthered for characterization. Repair of *LHCF2* may be viewed as a plausible explanation for returning to a WT-like phenotype, but even though this is also supported by the fact that both LHCFm 6.1.11 and 15.1 are green and display KO mutations of *LHCF2*, it is argued that a green phenotype is not caused by KO of *LHCF2* alone.

The perhaps strongest reasoning for this are the green phenotypes of the *alb3b*-mutants [115] in which the antennas are truncated because of an *overall* reduced ability to insert FCPs. These are moreover described by having 25% of the LHCF content of WT cells, and by a ~35% and ~65% reduction to Chl *a* and Fuco, respectively [115]. Moreover, the "greenest" of them all; LHCFm 15.1 – is suggested to lack expression of *LHCF1, 2, 3, 4* and *5* completely, presumably causing profoundly truncated antennas. With this in mind, it seems more reasonable to attribute greenness to a general loss of FCPs in the thylakoid, as is the case for *alb3b* mutants and for the LHCFm mutants that exhibit a gradient-like

increase in greenness with increased number of *LHCF* KOs. However, the actual optical difference may rather be caused by a lack of Fuco per Chl *a*, since loss of FCPs cause a loss in proteins to anchor pigments to the antennas. The Fuco:Chl *a* ratios presented here clearly demonstrates a gradual decrease in cellular Fuco:Chl *a* from the non-perturbed and darkest brown WT cells, to the most perturbed and least brown LHCFm 15.1 mutant, in both HL and LL. As such is the case, colouration may also be a good indicator of overall "truncated-ness" in diatoms. Additionally, the colour of the mutants might also be affected by structural changes in the antenna, since it is known that the absorption properties of Fuco strongly depends on the protein environment [95,100,109]. This phenomenon is described in more detail below.

5.3.2 Decreased absorbance at blue-green wavelengths

The absorbance curves of LHCFm mutants and WT *P. tricornutum* appear relatively equal at both LL and HL. The only major differences were seen in the spectra of LHCFm 15.1. Apparently, LHCFm 15.1 absorb less photons in the blue-green range of the spectrum fitting with the visual appearance of this mutant. Its decreased absorbance (as seen in the differential spectra of Fig. 4.6) peaked at the spectral wavelengths of 461 and 528 nm in LL and at 460 and 539 nm in HL. By comparing with the reference spectra (Fig. 1.6b) obtained from isolated FCPs of *C. meneghiniana* [95], it can be theorized that the decreased absorbance corresponds to the abs. maxima of Chl *c* (~460 nm *in vivo*) and of $Fuco_{GREEN}$ or $Fuco_{RED}$ which are the only pigments to absorb broadly in the green (>500 nm) bands of the spectrum. In other words, decreased absorbance indicates that LHCFm 15.1 contained lower levels of some of these pigments. Even though Chl *c* quantities could not be rendered by HPLC, this notion lies in agreement with the presented ratios of Fuco:Chl *a* per. cell. Intriguingly, Fuco:Chl *a* decreased steadily from WT towards the green-most mutant LHCFm 15.1, which has a Fuco:Chl *a* ratio that is halved compared to that of WT in HL. In LL, this effect is slightly less profound.

Another potential source to lowered absorption at ~530 nm may be described by the effects of Fuco's bathochromic shift, in which the optical properties of Fuco depend on the FCPs it interacts with [109]. If for instance one or several of the FCPs that are knocked out in LHCFm 15.1 interacts with Fuco to change its optical properties to those of $Fuco_{GREEN}$ or $Fuco_{RED}$, it would make sense that a lowered absorption was recorded in the blue-green wavebands. Gundermann and co-workers [109] found that FCP trimers consisting of LHCF5 are the most abundant in the antennas of *P. tricornutum*. Moreover, a fraction consisting almost solely of LHCF5 displayed the most red-shifted absorption, indicating that LHCF5 caused the most extreme bathochromic shift of Fuco to $Fuco_{RED}$ [109]. Considering this information, it seems reasonable that LHCFm 15.1 display the least profound absorbance in the waveband of ~530 nm, which also happens to characterize $Fuco_{RED}$ absorption (~500-540 nm, See Fig. 1.6b). This notion is furthermore supported by a similar decrease in absorption (also at ~ 530 nm) in

LL-acclimated LHCFm 6.1.11 (See Appendix D, Suppl. Fig. 4 for the Differential absorbance spectra of all LHCFm mutants), which also contain a *lhcf5* KO mutation.

5.3.3 LHCF content is strongly reduced in LHCFm 15.1

Western blots were conducted to semi-quantitatively describe the expression of LHCFs at the protein level. Although it was evident that protein quantities were not uniformly applied (control bands differed in intensity), a dilution gradient of WT lysate allowed to compare these intensities nevertheless. As one would expect from a cell line with quintuple loss of *LHCFs*, LHCFm 15.1 displayed a significantly decreased expression of LHCFs. When compared to WT lysates of equal D1-band intensity (WT 50 in HL, WT 50/25 in LL), LHCFm 15.1 had a clearly distinguished reduction – LHCF bands were un-detectable in both LL- and HL-acclimated cultures. Moreover, the antibody used (α-cmFCP) is predicted to have affinity for LHCF1-11, but based on the described results it seems more likely that it only detects LHCF1-5.

In contrast to LHCFm 15.1, LHCFm 6.1.11 did not show a profound decrease even though it is assumed to be the second most perturbed LHCFm mutant with KO of *lhcf1*, *2* and *5*. A possible explanation might be that LHCFm 6.1.11 retained the potential to upregulate expression of other light-harvesting proteins to compensate for *lhcf1*, *2* and/or *5*, such as the in-frame *LHCF3:LHCF4* gene fusion product. The extent of compensation was not explored further in the experimental work that led to this thesis, but similar effects have previously been reported in *Arabidopsis thaliana*, in which up-regulation of other light-harvesting proteins (notably LHCA and CP26) was proven in the absence of two major LHCII-comprising proteins (LHCB1 and LHCB2) [157]. In contrast, expression of FCPs seemed to be uniform in both LL and HL (equal intensity), suggesting that while LHCFm 6.1.11 may be able to compensate for its losses, it was less able to modulate the size of its antennas as a response to high or low irradiances. As one of the most basal cues of light-acclimation, low irradiances are expected to promote expansion of the antennas by increasing the number of FCPs per RC [91]. Conversely, HL-acclimation may lead to a general reduction of FCP expression, causing a down-sized antenna [91]. Thus, while this effect is detectable in the increased LL band intensity of both WT and LHCFm 1.10, it is not detectable in either of LHCFm 6.1.11 or LHCFm 15.1, indicating that both cell lines are somewhat hampered from changing the amount of FCP in their antennas.

5.3.4 LHCFm 15.1 has lowered photosynthetic efficiency

Inferred by variable fluorescence, Φ_{PSII} describes the quantum efficiency of PSII photochemistry (F_v'/F_m') and may be interpreted as the proportion of absorbed light entering photochemistry [141]. The $\Phi_{PSIImax}$ describe the maximum efficiency of photochemistry as obtained by variable fluorescence measurements after a short dark incubation; when all RCs are open and ready to receive and utilize photonic energy [141]. The $\Phi_{PSIImax}$ is therefore a good indicator on the potential productivity of a particular cell line, and a high value is perhaps especially important for a high-performing TLA mutant

in biomass production [132] as it indicates a high probability for the PSII RCs to use the absorbed excitation energy for photosynthesis.

The $\Phi_{PSIImax}$ values of LHCFm 1.10 and 6.1.11 are practically equivalent to that of the WT when acclimated to both LL and HL, indicating that they are equally efficient. The most striking feature of the $\Phi_{PSIImax}$ values were those of LHCFm 15.1 which displayed a severely reduced quantum efficiency of approximately 50-60% of the WT. It is hypothesized that the reduced $\Phi_{PSIImax}$ values may either be attributed to a low F_m or a high F_0 through either of two scenarios: i) an elevated level of the RCs are damaged (closed) and unable to use the light energy in photosynthesis, so that the minimum fluorescence yield (F_0) is high even after dark incubation, or ii) the saturation flash does not saturate PSII RC, resulting in a lower F_m as a result of poor transfer of energy from FCPs to RC. Also, perhaps it is necessary to take a look at what is known about the LHCFs that are deleted from this cell line. Previous transcriptomic and proteomics analyses have showed that *LHCF1-5* are amongst the most expressed LHC genes/proteins in the genome of *P. tricornutum* [31,108,109] indicating that their role in light-harvesting is important relative to other LHCFs. Also, besides observing that trimers of LHCF5 were the most abundant, Gundermann and co-workers [109] have reported that the abundance of LHCF4 and LHCF5 varies in LL- and HL-treatment, indicating that they were differentially regulated to optimize or cope to changing light intensities [109]. In fact, while LHCF4 was found in increased abundance during HL, LHCF5 was down-regulated [109].

The nuances that make diatom FCPs differ from one another are to a large extent still unknown, and it is therefore difficult to attribute the changes in $\Phi_{PSIImax}$ to loss of a particular protein. However, by considering a greatly reduced $\Phi_{PSIImax}$, literature on the abundances and properties of LHCFs, and the genetic modifications this cell line-has obtained – it is argued that the antennas of LHCFm 15.1 are structurally deficit. A low $\Phi_{PSIImax}$ would make sense given antennas that were either too far apart, so that RET is hindered by an increased distance from one pigment to another. This would moreover explain scenario ii) by attributing a lowered F_m to quantum losses during the saturating flash. By having multiple *LHCFs* knocked out, the antennas may be lacking FCPs that are important in the trimers, such as LHCF5 or LHCF4 [109]. Conversely, LHCFm 6.1.11 (which also lack LHCF5 but seems to retain a functional *LHCF3*:*LHCF4* gene fusion) displayed equal $\Phi_{PSIImax}$ to the WT, indicating that LHCF3 and/or LHCF4 may contribute in maintaining the structural integrity of the antennas. Even though the nuances of the diatom FCPs to a large extent remains unknown, several reports have been published on the distinct functional roles of the corresponding LHC-proteins in *A. thaliana* and *C. rheinhardtii* [158–160]. Thus, it may not be too far-fetched with nuanced roles in *P. tricornutum* FCPs. It is shown that while LHCFm 6.1.11 managed to retain an $\Phi_{PSIImax}$ equivalent to the WT, the $\Phi_{PSIImax}$ of the LHCFm 15.1 cell line is greatly reduced. This indicates that at least one of the highly expressed LHCFs (in this case the LHCF3:LHCF4 gene fusion) are required to retain a functional photosynthetic response, and

that a complete lack of these LHCF copies puts LHCFm 15.1 "over the edge" – unable to perform efficient photosynthesis.

5.3.5 LHCFm mutants require more light to reach saturation of photosynthesis

Acclimation to LL is normally characterized by antenna expansion. As a consequence, LL-acclimated cells subjected to a fluorometric assessment will generally show characteristic signatures in the parameters that describe the PE-curves, such as a steeper α, and lowered maximum electron transport rate and E_k – simply because they exhibit larger antennas, and thus require less light to reach saturation. The PE-curves presented in this thesis showed how the LHCFm mutants are gradually becoming less able to HL-acclimate. From a smaller discrepancy between LL- and HL-acclimated PE-curves in LHCFm 6.1.11 to the virtually equal curves of LHCFm 15.1, it can be argued that KO of *LHCF*s have hampered their ability to switch between LL- and HL-configurations by fine-tuning the FCP:pigment contents of their antennas. The parameters that stem from the PE-curves (α, E_k, $rETR_{max}$) may be further assessed to describe why.

The $rETR_{max}$ (synonymous to P_{max}) reports on the relative electron transport rate at its maximum efficiency [141]. Therefore, $rETR_{max}$ can be used as a proxy for obtaining an estimate of how productive cells are at their peak. For a WT organism, the highest $rETR_{max}$ will be obtained in samples acclimated to HL; when the antennas are smaller and the light requirement for saturation becomes increased [116]. As such is the case, one would expect the $rETR_{max}$ values to be increased in TLA mutants of which antennas are already smaller. This has been exemplified previously, for instance by the curves of the *C. rheinhardthii* TLA strain *tla3* [161]. However, this is not necessarily the case for the LHCFm mutants. In fact, the LHCFm 15.1 mutant displayed an equal (LL-acclimated) or lower (HL-acclimated) $rETR_{max}$ to that of the WT. In contrast, LHCFm 6.1.11 display the most promising $rETR_{max}$ values with regards to its potential as a TLA mutant. It is increased from WT regardless of acclimation but is perhaps particularly profound when grown in LL (increased by 23%). Compared to the LHCFm 15.1 mutant, it is greatly increased in HL, with a 30% higher $rETR_{max}$.

One of the most reoccurring parameters in TLA mutant research is the light-saturation coefficient – E_k [132,135,156]. E_k describes the irradiance (µmol photons m^{-2} s^{-1}) at which the photosystem RCs become saturated. By having truncated antennas, TLA mutants are supposed to require more quanta of photons to reach saturation, and thus display increased E_k's [132]. Although not as profound as the 200% increased E_k of a *Synechocystis sp.* TLA mutant [135], this trait was also visible in the E_k's of the LHCFm mutants. Intriguingly, E_k increased in a gradient-like fashion from WT-levels to increased *lhcf* perturbance, with LHCFm 15.1 as the extreme, showing a 37% and 21% increased E_k to that of the WT in LL and HL, respectively. However, even though it is the highest, the nature of the PE curves has it that all these coefficients are correlated, so that the E_k is drawn to the irradiance from the intercept where the α meets the $rETR_{max}$. Thus, a high E_k may correlate with either a small α or a high rETRmax

– and in this case, it is safe to say that a small light-utilization coefficient was causative for a high E_k. This is moreover in agreement with a low $\Phi_{PSIImax}$. Another increased E_k was that of LHCFm 6.1.11, indicating that this mutant would also require more photons for saturation.

5.3.6 Do LHCFm mutants require less NPQ, or are they less capable of performing it?

NPQ describes the need or capacity to dissipate energy to alleviate pressure, and to avoid build-up of ROS [117]. The ΔpH-dependent qE-component of NPQ is particularly important in diatoms, and involves the conversion of Ddx to Dtx [112]. Therefore, the prevalence of Dtx contra Ddx (DES, Fig. 4.5b) often correlates well with the level of NPQ [162]. NPQ can be inferred from the decrease in F_m during the creation of rapid light curves by PAM fluorometry (Fig. 4.9), and indicates that absorbed energy is dissipated as heat instead of fluorescence.

The differences seen in NPQ capacity between WT and mutant lines (Fig. 4.9) with the highest values in WT and LHCFm 1.10, somewhat lower in LHCFm 6.1.11 and very low in LHCFm 15.1 could indicate i) unequal needs for photoprotection related to differences in antenna size and absorption efficiencies and/or ii) changes in the antenna structure affecting the capability for performing NPQ. A smaller antenna size, as indicated for LHCfm 6.1.11 and 15.1, decreases the absorption efficiency, and could subsequently lead to a lower NPQ since more light is necessary to saturate photosynthesis fitting with theory i) for an explanation of the differences in NPQ. However, the decline in growth rate in HL acclimated cells compared to LL observed for LHCFm 6.1.11 and 15.1, coupled with the incapability of the LHCFm 15.1 mutant to increase its photosynthetic performance after HL acclimation, indicates that these mutants are light-sensitive and thus in support of theory ii). Similar decreases in NPQ have been observed in plants lacking the major LHCII antenna complexes, but whether these changes occur as a result of general loss or structural deficiency is not known [163,164]. It is therefore difficult to attribute lower NPQ to either of the two theories, since a cell line lacking the major FCP complexes may be both structurally deficit and/or lacking the FCPs.

As a response to HL-acclimation, all mutant lines showed a two-fold increase in DES (Fig. 4.5b), whereas the DES increase was more profound in WT (>3 times), indicating that a WT sized antenna is accompanied by an increased need or capability for excessive energy dissipation in accordance with theory i). However, from our current understanding of the genetic composition and the NPQ measurements of these mutants, it would then also be expected that the DES index of LHCFm 1.10 was similar to WT, while LHCFm 6.1.11 and 15.1 had decreased or greatly decreased DES indexes, respectively. Thus, The DES index values in this study do not correlate well with the observed NPQ values. It must also be emphasized that the DES index values presented are not absolute numbers, and because the cell lines subjected to HPLC pigment analyses were grown in a separate experiment, it is not possible to directly compare the absolute quantities of pigments. Thus, the study lacks data to explain this peculiarity.

5.3.7 Which LHCFm-strain would be a suitable TLA mutant?

In general, it is desired to have a TLA mutant which is not able to LL-acclimate. Such a mutant would retain a small antenna even when grown into light-deprivation, thereby allowing a larger volume of a high-density culture to remain illuminated. LHCFm 1.10 falls short in this regard in virtually all the analyses employed. With only one *lhcf* KO it is deemed to be the least perturbed, and thus it behaves more like a WT than a TLA mutant. Compared to WT, it displayed approximately equally distinct PE-curves, equal absorbance spectra, a slight but non-significant decrease in the Fuco:Chl *a* ratio, unchanged FCP quantity, and an equally expressed NPQ of Chl fluorescence. In the other end of the spectrum it was shown that even though LHCFm 15.1 displayed the most profound TLA-like traits (highest E_k, lowest Fuco:Chl *a*, greatly reduced FCP contents, and a diminished ability to increase the size of its antenna), it is less fit due to a greatly lowered maximum quantum yield of PSII photochemistry ($\Phi_{PSIImax}$) and greatly decreased growth-rate regardless of light-intensity. The results provided indicated that the extent of its modifications was too extreme, leading to inefficient photosynthesis. As the "golden mean" of the three, LHCFm 6.1.11 displayed a combination of high E_k, $\Phi_{PSIImax}$ and an increased $rETR_{max}$. Moreover, it is thought to retain some of the modularity that LHCFm 15.1 was predicted to lack, possibly by having an in-frame *LHCF3:LHCF4* copy that may compensate for *lhcf1*, *2* and *5* KOs. Even though the western blot indicated that FCP contents was at least equal to that of WT, it had decreased NPQ of Chl fluorescence, decreased Fuco:Chl *a* contents, and a growth rate that rivalled the WT. On the basis of these characteristics, it is argued that the most promising TLA mutant was LHCFm 6.1.11.

5.5 A photobioreactor pilot-study came short due to nutrient deprivation

A pilot experiment was performed to screen for a putative TLA mutant and to detect whether any of the LHCFm or *alb3b*-mutants could outperform WT cells when grown to high densities in photobioreactors. It is important to stress that this was a pilot-experiment conducted to find out if the TLA concept may be applied in diatoms with either *lhcf* or *alb3b* KO mutations. LHCFm 1.10, 6.1.11 and 15.1 were selected based on a changed colouration to green or light brown and on preliminary sequencing data indicating multiple *lhcf* KOs. The KO mutants of *alb3b* (called line 14.8, 16.7 and 19.7) were selected based on typical diatom TLA characteristics with 65% and 35% decrease to the amount of Fuco and Chl *a*, respectively, a 75% decrease of LHCF proteins, and a bright-green colour display (Nymark et al., [115] .

At the end of the pilot experiment (after 2 weeks of growth), the maximum cell density of the LHCFm 6.1.11 mutant (59.1 mill. cells mL^{-1}) rivalled that of the WT (58.8 mill. cells mL^{-1}), but light deprivation was never reached. In fact, the light-intensity in the centre of these high-density cultures was measured to be around 50 µmol photons m^{-2} s^{-1}. Moreover, post-experiment analyses confirmed that the cultures were limited by access to DIP and DIN after 72–96 and 144 hours, respectively. The TLA mutants were therefore not tested in the light-limited conditions they were thought to potentially outperform WT cells, but were instead limited by nutrient access. For this reason, the following paragraphs are allocated to discuss findings, the apparent limitations and the measures that may be taken to avoid them.

5.5.1 WT vs. TLA – Wild and free vs. Slow and steady

Most KO mutants may be expected to display retarded growth in comparison to a healthy and intact WT cell. Likewise, LHCFm and *alb3b* mutants lack constituents of their antenna, and it was therefore expected that they would grow slower, especially in the exponential phase, and in particular when the growth conditions were optimal. This was also the case for the both LHCFm and *alb3b* mutants when grown in photobioreactors. Mutant cultures were lagging behind already after the first 24 hours, and were at least ~ three times less dense than the WT average (5.7 mill. cells mL^{-1}) at 72 hours ([1.3, 2.1] mill. cells mL^{-1}). This difference was particularly profound during the early exponential phase where WT cell density dominance reoccurred throughout the experiment until it was overtaken by LHCFm 6.1.11 after 336 hours (14 days). In general, the curves presented indicated that WT grows fast, showing a maximal growth rate in the exponential phase in accordance with literature [165], and when growth conditions improved at the onset of additional CO_2 (96 hours) or nutrients (216 and 248 hours) – but that it quickly reached some sort of decline – indicating that rapid growth comes at a cost of high nutrient and CO_2 consumption. These effects were also apparent given a linearized axis (Appendix G). The growth curves of LHCFm and *alb3b* mutants are contrasted by a slower and steadier rise, indicating a decreased rate of both nutrient uptake and photosynthetic efficiency. Based on the growth curves, it is probable that LHCFm 15.1 was too perturbed to obtain a high density in the first place given an

increased doubling time during the exponential phase (~19 h) and a maximum cell density of only 25 mill. cells mL^{-1}. LHCFm 6.1.11 however, even though also its doubling time in the exponential phase is lower (~17 h) compared of the WT (10.5 h), showed the most promising densities towards the end of the experiment, rivalling that of the WT average with a maximum density of 59 mill. cells mL^{-1}, further strengthening its position as the first candidate for continued diatom TLA trials.

5.5.2 All cultures became nutrient-deprived

Post-experiment analyses revealed that the nutrients of the rich F-medium were rapidly depleted. Its initial quantities of ~36.0 DIN and 3.4 mg L^{-1} DIP were reduced to zero only after 144 hours (day 6). Moreover, it was found that DIP is taken up prior to DIN, and that cultures with increased growth rates deplete their media more rapidly. Diatoms have been shown to exhibit increased P uptake as a response to P deprivation [166], but an opposite response has been observed in absence of N as *P. tricornutum* upon N deprivation enhanced its uptake of ammonia (NH_4^+), and decreased its uptake of Nitrate (NO_3^-) [167]. Since the DIN-component of the F-medium (Appendix A1) is NO_3^- and not NH_4^+, this notion may aid in explaining why DIN is depleted after DIP: simply because P deficiency promotes additional uptake, while N deficiency does not. As such, WT depletion of DIP occurs already after 72 hours, while depletion of DIN occurs at 144 hours. DIP depletion may moreover correlate with the decline in WT growth rate observed at the same time, but diatoms are known to be able to store substantial amounts of both Nitrate (NO_3^-) [168] and Phosphate (PO_4^{3-}) intracellularly [169], implying that continued growth may be accomplished even in the absence of DIN and DIP. Hence appears the relevance of cellular nutrient contents.

Also the cellular contents of organic C, N and P correlate with the growth because the most rapidly growing cultures display the lowest amounts of organic C, N and P. As such, WT cultures are generally on the bottom end of this range, in which cultures divide rapidly and therefore shares a finite pool of resources over a larger quantity of cells. Moreover, it is important to mention that the cellular contents of N and P appear to have stabilized between 144 to 264 hours. This indicates a timing of the onset stationary phase that corresponds well with the growth curves (start stationary phase ~144 h (WT), ~168 h (mutants)), and that minimal cellular N and P quantities limited continued growth. This observation also agrees with the drastic increases in C accumulation, as photosynthetic sequestering of carbon has been found unaffected in *P. tricornutum* grown in severe nitrate deprivation [170]. Moreover, large discrepancies are observed in the N:P ratios. While the Redfield ratio [171,172] denotes the optimal ratio of inorganic N:P in a culture medium, the critical ratio denotes the cellular content that a species approach when facing both N- and P-limitation [155]. As mentioned earlier, all mutant ratios fell within the 25-33 critical range of *P. tricornutum* [155] after 168 h, indicating that they may be co-limited by N and P. However, WT N:P ratios are extremely high, indicating that they foremost see P-deficiencies. This may be explained given the presented observation of DIP being more rapidly taken up by WT

cultures than by mutants, which would also make sense given *P. tricornutum*'s ability to increase P uptake. However, the key observation here is that growth rate decides rate of removal and the potential for limitation. Hence arises the problem of having separate cultures that grow at different speeds, which may moreover be inevitable because of the overall lower fitness of a KO mutant. It is thus paramount that a richer medium is used, so that both WT and mutant cultures would be allowed to reach light-limited cell densities while still in nutrient replete conditions. To further promote light-limitation over nutrient-limitation, it is proposed to use photobioreactors with larger diameters, so that the light is attenuated over a larger volume. Consequently, one would require a lower density of cells to achieve the light-limited conditions in which a TLA mutant is thought to excel in.

5.5.3 High $\Phi_{PSIImax}$ is re-established with additional nutrients and CO_2

The maximum quantum yield of PSII photochemistry ($\Phi_{PSIImax}$) values fluctuated over time but re-established at high yields when additional nutrients and CO_2 were appointed. Judging by Fig. 4.14, it appears that *alb3b* mutants were generally more fit, as they displayed high $\Phi_{PSIImax}$ values that varied between 0.49 and 0.70 throughout the experiment. Larger discrepancies were seen for the LHCFm mutants, of which 6.1.11 and 1.10 displayed similar curves to that of the WT, although their $\Phi_{PSIImax}$ seemed to change less over time. LHCFm 15.1 however seemed to be "chronically ill" – with very low initial $\Phi_{PSIImax}$ values (0.44) that gradually decreased to a mere 0.22, but then saw a twofold increase to 0.50 at the onset of additional nutrients (264 h).

While it is difficult to dissect which component of the F-medium could cause such a response, it can be tapped into by a look into literature. Of-course, because only N and P were analysed during this experiment, other elements such as Iron (Fe) and Sulphur (S) may be overlooked, and thus it is impossible to attribute decreases or increases in the $\Phi_{PSIImax}$ to a specific chemical limitation. However, Wykoff and co-workers [173] found that the $\Phi_{PSIImax}$ was significantly reduced when *C. rheinhardtii* was grown in P- and S-deficient media, and that S-deficiency was particularly causal to a reduced electron flow from the secondary electron acceptor (Pq), possibly clogging the ETR and thus a lowered $\Phi_{PSIImax}$.

The effect of N-limitation does however seem to provoke disagreement. Some findings indicate that N-limitation causes significant drops to Φ_{PSII} [174–176], while others refute such effects [170]. Some have found significant decreases to $\Phi_{PSIImax}$ [177,178], while some have found significant decreases to Φ_{PSII} but not to $\Phi_{PSIImax}$ when growing *P. tricornutum* in N-limited cultures [174]. With these divergent findings it is therefore difficult to determine how influential N-limitation was on the $\Phi_{PSIImax}$ of *P. tricornutum*. It was however quite evidently shown that DIP was depleted before DIN, indicating that *P. tricornutum* grown in F-medium will sooner become limited by P-starvation rather than N-starvation. It was furthermore argued that P-starvation was particularly severe in rapidly proliferating cultures, as manifested by extreme cellular N:P ratios. Nevertheless, it is argued that Φ_{PSII} measurements can be

used to probe on the relative conditions of the culture, hinting at when the medium is approximating a nutrient limitation. Thus, daily monitoring can aid in future trials by allowing increased probability of reaching light-limited densities, while avoiding nutrient-limited conditions.

6 Future aspects

This study indicated that the high-fidelity modifications proposed by Kleinstiver et al., [83] are adaptable to the *Sp*Cas9 enzyme used in diatom genome editing, but that more vigorous testing is required to satisfactory determine its on- and off-target efficiencies. This can be achieved with a larger pool of mutants, coupled with a more efficient method of screening. It is furthermore argued that such a study could benefit from targeting genomic DNA sequences that are spatially distinct, so that large NHEJ-mediated deletions does not affect other potential target sites (i.e. recording editing events in un-amplifiable deleted target sites). A second improvement could be to transform diatom cells by bacterial conjugation. Expressing the CRISPR-Cas9 system via episomes would avoid plasmid integration, subsequent re-editing to WT sequence, and thus also the possibility of inferring false negatives.

At present, little is known about the nuanced roles of FCPs in *P. tricornutum*, as opposed to the corresponding proteins of green algae and plants. This would also apply to the *alb3b*-insertase which has been indicated to interact with other proteins, for instance Cyt-b6f (Nymark et al., Unpublished data.). The results presented here indicate that LHCF3, LHCF4 and LHCF5 are important constituents of *P. tricornutum* FCP trimers, and that their absence may lead to an inefficient and structurally deficit antenna. From this it is argued that more basic research is needed to map out nuances and distinct roles of FCP proteins within the antennas, since such studies would generate information that can aid scientists in making an optimal diatom TLA. With our current knowledge, perhaps a more suitable way to generate fit and photosynthetically efficient TLA mutants would be through a general reduction of the antennas without complete removal of any of its integral components, so that TLAs are achieved without loss of particular functions. An alternative strategy could be to induce knock-down of LHCFs through RNA interference (RNAi). Such a strategy was earlier showcased by Oey and co-workers [179] who performed a triple knock-down of the *LHCBM1*, 2 and 3 in *C. rheinhardtii*, and subsequently improved both light-distribution and cell densities of photobioreactor cultures.

LHCFm 6.1.11 was the most promising cell line, with increased light-saturation requirements, lowered NPQ and ability to LL-acclimate, and higher $rETR_{max}$ to that of the WT. This notion was also solidified by growth trials where LHCFm 6.1.11 achieved a cell density comparable to the WT. It would therefore be interesting to test this cell line against the WT, preferentially with several replicate cultures. Such a study could benefit from this pilot-experiment and should consider using a richer culture medium, and photobioreactors with larger diameter. Daily measurements of nutrient levels and $\Phi_{PSIImax}$ could moreover aid in detecting other potential limitations such as N, P or CO_2 depletion, so that novel ingredients may be added preventively.

7 Conclusion

Multiple highly similar LHCFs were targeted to test whether a high-fidelity SpCas9 enzyme was better apt to avoid off-target DSBs when applied in diatom genome editing. The hifiCas9 enzyme was found to retain approximately 75% of the on-target activity reported for diaCas9 and did only induce off-target DSB at one occasion. In contrast, diaCas9 was highly promiscuous, and introduced DSBs in all cell lines at the 1 mismatch off-target site of *LHCF2*. Hence it is argued that the hifiCas9 enzyme may provide the fidelity that diaCas9 lacks, especially when targeting DNA sequences with high similarity to other genomic loci.

Through various analyses it was found that the brownish LHCFm 1.10 with only *lhcf1* knocked out behaved like WT, and that its bright green opposite LHCFm 15.1 was too perturbed; possibly causing an inefficient and structurally deficit antenna. As the golden mean, LHCFm 6.1.11 was proposed to be the most promising mutant, containing three *lhcf* knockouts (*lhcf1*, *2* and *5*) and an in-frame gene fusion of *LHCF3* and *LHCF4*. LHCFm 6.1.11 required more light to become saturated, had increased capacity for photosynthesis, an equal maximum quantum yield of PSII photochemistry to that of the WT, had decreased NPQ, a much lower Fuco:Chl *a* ratio and divided almost as rapidly as the WT when grown in HL. The PE-curves indicated a lowered ability to modulate the antenna size, enabling it to maintain a smaller antenna even under LL.

The overall goal of the photobioreactor growth trials was to screen for diatom TLA mutants which could apply to the increased productivity that the TLA mutants of algal groups coax with. However, because our findings conclude that light-deprivation was never reached, and that nutrient limitation instead kicked in – the principal criteria for TLA performance was never reached. Hence it is proposed that further trials are needed to test diatom TLAs, and that the findings of this thesis may aid in improving such an experimental design.

8 Literature

1. **Alberts B, Johnson A, Lewis J, Morgan D, Raff M, Roberts K, et al.** *Molecular Biology of The Cell*. 6th ed. 711 Third Avenue, New York, NY 10017, US: Garland Science, Taylor & Francis Group, LLC; 2015.
2. **Sigman DM, Hain MP**. *The Biological Productivity of the Ocean*. Nat Educ Knowl. 2012;3(6):1–16.
3. **Field CB, Behrenfeld MJ, Randerson JT, Falkowski P**. *Primary Production of the Biosphere: Integrating Terrestrial and Oceanic Components*. Science (80-) . 1998 Jul 10;281(5374):237 LP – 240. Available from: http://science.sciencemag.org/content/281/5374/237.abstract
4. **Gaarder T, Gran H**. *Investigations of the production of plankton in the oslo fjord*. Cons Intebational pour l'exploration la Mar Rapp Proces-Verbaux. 1927;42:1–48.
5. **Nielsen ES**. *The Use of Radio-active Carbon (C14) for Measuring Organic Production in the Sea*. ICES J Mar Sci . 1952 Aug 1;18(2):117–40. Available from: http://dx.doi.org/10.1093/icesjms/18.2.117
6. **Regaudie-de-Gioux A, Lasternas S, Agustí S, Duarte CM**. *Comparing marine primary production estimates through different methods and development of conversion equations*. Vol. 1, Frontiers in Marine Science. 2014. p. 19. Available from: https://www.frontiersin.org/article/10.3389/fmars.2014.00019
7. **Bowler C, Vardi A, Allen AE**. *Oceanographic and Biogeochemical Insights from Diatom Genomes*. Ann Rev Mar Sci . 2009 Dec 14;2(1):333–65. Available from: https://doi.org/10.1146/annurev-marine-120308-081051
8. **Curtin TB, Belcher EO**. *Innovation in oceanographic instrumentation*. Oceanography. 2008;21:44.
9. **Nelson DM, Tréguer P, Brzezinski MA, Leynaert A, Quéguiner B**. *Production and dissolution of biogenic silica in the ocean: Revised global estimates, comparison with regional data and relationship to biogenic sedimentation*. Global Biogeochem Cycles . 1995 Aug 28;9(3):359–72. Available from: https://doi.org/10.1029/95GB01070
10. **Falkowski PG, Barber RT, Smetacek V**. *Biogeochemical Controls and Feedbacks on Ocean Primary Production*. Science (80-). 1998 Jul 10;281(5374):200 LP – 206. Available from: http://science.sciencemag.org/content/281/5374/200.abstract
11. **Méndez-Ferrer N, Hallock P, Jones DL**. *Photochemical Efficiencies of Diatom Symbionts in Hospite in Amphistegina Gibbosa (foraminifera) Across Seasons in the Florida Keys, USA*. J Foraminifer Res. 2018 Jan 1;48(1):4–16. Available from: http://dx.doi.org/10.2113/gsjfr.48.1.4
12. **Di Camillo CG, Cerrano C, Romagnoli T, Calcinai B**. *Living inside a sponge skeleton: the association of a sponge, a macroalga and a diatom*. Symbiosis. 2017;71(3):185–98. Available from: https://doi.org/10.1007/s13199-016-0426-7
13. **Letáková M, Fránková M, Poulíčková A**. *Ecology and Applications of Freshwater Epiphytic Diatoms — Review*. Cryptogam Algol. 2018 Feb 1;39(1):3–22. Available from: https://doi.org/10.7872/crya/v39.iss1.2018.3
14. **DeYoe HR, Lowe RL, Marks JC**. *Effects of Nitrogen and Phosphorus on the Endosymbiont Load of Rhopalodia Gibba and Epithemia Turgida (Bacillariophyceae)*. J Phycol. 1992;28(6):773–7.
15. **Thomas DN, Dieckmann GS**. *Oceanography and Marine Biology, An Annual Review, Biogeochemistry of Antarctic Sea Ice*. Vol. 40. London: CRC Press; 2002. 6 p.
16. **Armbrust EV**. *The life of diatoms in the world's oceans*. Nature . 2009 May 13;459:185. Available from: http://dx.doi.org/10.1038/nature08057
17. **Post E, Bhatt US, Bitz CM, Brodie JF, Fulton TL, Hebblewhite M, et al.** *Ecological consequences of sea-ice decline*. Science (80-). 2013;341(6145):519–24.
18. **Lizotte MP**. *The Contributions of Sea Ice Algae to Antarctic Marine Primary Production1*. Am Zool. 2001 Feb 1;41(1):57–73. Available from: http://dx.doi.org/10.1093/icb/41.1.57
19. **Haeckel E**. *Kunstformen der Natur / von Ernst Haeckel ; Fünfzig Illustrationstafeln mit beschreibendem Text*. Leipzig und Wien: Verlag des Bibliographischen Instituts; 1899.
20. **Hamm CE, Merkel R, Springer O, Jurkojc P, Maier C, Prechtel K, et al.** *Architecture and material properties of diatom shells provide effective mechanical protection*. Nature. 2003 Feb 20;421:841. Available from: http://dx.doi.org/10.1038/nature01416
21. **Struyf E, Smis A, Van Damme S, Meire P, Conley DJ**. *The Global Biogeochemical Silicon Cycle*. Silicon . 2009;1(4):207–13. Available from: https://doi.org/10.1007/s12633-010-9035-x
22. **Tréguer P, Nelson DM, Van Bennekom AJ, DeMaster DJ, Leynaert A, Quéguiner B**. *The Silica Balance in the World Ocean: A Reestimate*. Science (80-). 1995 Apr 21;268(5209):375 LP – 379. Available from: http://science.sciencemag.org/content/268/5209/375.abstract
23. **Vardi A, Thamatrakoln K, Bidle KD, Falkowski PG**. *Diatom genomes come of age*. Genome Biol. 2008 Jan 9;9(12):245. Available from: http://www.ncbi.nlm.nih.gov/pmc/articles/PMC2646270/
24. **Alverson AJ, Cannone JJ, Gutell RR, Theriot EC**. *The evolution of elongate shape in diatoms*. J Phycol. 2006 Jun 1;42(3):655–68. Available from: https://doi.org/10.1111/j.1529-8817.2006.00228.x
25. **Falciatore A, Bowler C**. *Revealing the molecular secrets of marine diatoms*. Annu Rev Plant Biol . 2002 Jun 1;53(1):109–30. Available from: https://doi.org/10.1146/annurev.arplant.53.091701.153921

26. **Poulsen NC, Spector I, Spurck TP, Schultz TF, Wetherbee R**. *Diatom gliding is the result of an actin-myosin motility system*. Cell Motil. 1999 Sep 1;44(1):23–33. Available from: https://doi.org/10.1002/(SICI)1097-0169(199909)44:1%3C23::AID-CM2%3E3.0.CO

27. **Bowler C, Allen AE, Badger JH, Grimwood J, Jabbari K, Kuo A, et al.** *The Phaeodactylum genome reveals the evolutionary history of diatom genomes*. Nature. 2008 Nov 13;456(7219):239–44. Available from: http://dx.doi.org/10.1038/nature07410

28. **Armbrust EV, Berges JA, Bowler C, Green BR, Martinez D, Putnam NH, et al.** *The Genome of the Diatom Thalassiosira Pseudonana: Ecology, Evolution, and Metabolism*. Science (80-). 2004 Oct 1;306(5693):79 LP – 86. Available from: http://science.sciencemag.org/content/306/5693/79.abstract

29. **Scala S, Carels N, Falciatore A, Chiusano ML, Bowler C**. *Genome Properties of the Diatom Phaeodactylum tricornutum*. Plant Physiol. 2002 Jul 1;129(3):993 LP – 1002. Available from: http://www.plantphysiol.org/content/129/3/993.abstract

30. **Apt KE, Grossman AR, Kroth-Pancic PG**. *Stable nuclear transformation of the diatom Phaeodactylum tricornutum*. Mol Gen Genet MGG. 1996;252(5):572–9. Available from: https://doi.org/10.1007/BF02172403

31. **Falciatore A, Casotti R, Leblanc C, Abrescia C, Bowler C**. *Transformation of Nonselectable Reporter Genes in Marine Diatoms*. Mar Biotechnol. 1999;1(3):239–51. Available from: https://doi.org/10.1007/PL00011773

32. **Valenzuela J, Mazurie A, Carlson RP, Gerlach R, Cooksey KE, Peyton BM, et al.** *Potential role of multiple carbon fixation pathways during lipid accumulation in Phaeodactylum tricornutum*. Biotechnol Biofuels. 2012;5.

33. **De Martino A, Meichenin A, Shi J, Pan K, Bowler C**. *Genetic and phenotypic characterization of Phaeodactylum tricornutum (Bacillariophyceae) accessions*. J Phycol. 2007;43(5):992–1009.

34. **Hustedt F**. *Kieselalgen*. 1956;

35. **Parker MS, Mock T, Armbrust EV**. *Genomic Insights into Marine Microalgae*. Annu Rev Genet . 2008 Nov 4;42(1):619–45. Available from: https://doi.org/10.1146/annurev.genet.42.110807.091417

36. **Oudot-Le Secq M-P, Grimwood J, Shapiro H, Armbrust EV, Bowler C, Green BR**. *Chloroplast genomes of the diatoms Phaeodactylum tricornutum and Thalassiosira pseudonana: comparison with other plastid genomes of the red lineage*. Mol Genet Genomics . 2007;277(4):427–39. Available from: https://doi.org/10.1007/s00438-006-0199-4

37. **Daniell H, Lin CS, Yu M, Chang WJ**. *Chloroplast genomes: Diversity, evolution, and applications in genetic engineering*. Genome Biol. 2016;17(1).

38. **Cavalier-Smith T**. *Genomic reduction and evolution of novel genetic membranes and protein-targeting machinery in eukaryote-eukaryote chimaeras (meta-algae)*. Philos Trans R Soc London Ser B Biol Sci . 2003 Jan 29;358(1429):109 LP – 134. Available from: http://rstb.royalsocietypublishing.org/content/358/1429/109.abstract

39. **Flori S, Jouneau P-H, Bailleul B, Gallet B, Estrozi LF, Moriscot C, et al.** *Plastid thylakoid architecture optimizes photosynthesis in diatoms*. Nat Commun . 2017 Jun 20;8:15885. Available from: http://dx.doi.org/10.1038/ncomms15885

40. **Moustafa A, Beszteri B, Maier UG, Bowler C, Valentin K, Bhattacharya D**. *Genomic Footprints of a Cryptic Plastid Endosymbiosis in Diatoms*. Science (80-). 2009 Jun 26;324(5935):1724 LP – 1726. Available from: http://science.sciencemag.org/content/324/5935/1724.abstract

41. **Prihoda J, Tanaka A, de Paula WBM, Allen JF, Tirichine L, Bowler C**. *Chloroplast-mitochondria cross-talk in diatoms*. J Exp Bot . 2012 Feb 1;63(4):1543–57. Available from: http://dx.doi.org/10.1093/jxb/err441

42. **Dunahay TG, Jarvis EE, Roessler PG**. *Genetic Transformation of the Diatoms Cyclotella Cryptica and Navicula Saprophila*. J Phycol. 1995;31(6):1004–12.

43. **Neumann E, Schaefer-Ridder M, Wang Y, Hofschneider PH**. *Gene transfer into mouse lyoma cells by electroporation in high electric fields.* EMBO J . 1982;1(7):841–5. Available from: http://www.ncbi.nlm.nih.gov/pmc/articles/PMC553119/

44. **Niu YF, Yang ZK, Zhang MH, Zhu CC, Yang WD, Liu JS, et al.** *Transformation of diatom Phaeodactylum tricornutum by electroporation and establishment of inducible selection marker*. Biotechniques. 2012;52(6).

45. **Zhang C, Hu H**. *High-efficiency nuclear transformation of the diatom Phaeodactylum tricornutum by electroporation*. Mar Genomics. 2014;16:63–6. Available from: http://www.sciencedirect.com/science/article/pii/S1874778713000597

46. **Miyahara M, Aoi M, Inoue-Kashino N, Kashino Y, Ifuku K**. *Highly Efficient Transformation of the Diatom Phaeodactylum tricornutum by Multi-Pulse Electroporation*. Biosci Biotechnol Biochem. 2013;77(4):874–6. Available from: http://www.ncbi.nlm.nih.gov/pubmed/23563551

47. **Sharma AK, Nymark M, Sparstad T, Bones AM, Winge P**. *Transgene-free genome editing in marine algae by bacterial conjugation – comparison with biolistic CRISPR/Cas9 transformation*. Sci Rep. 2018;8(1):14401. Available from: https://doi.org/10.1038/s41598-018-32342-0

48. **Karas BJ, Diner RE, Lefebvre SC, McQuaid J, Phillips APR, Noddings CM, et al.** *Designer diatom episomes delivered by bacterial conjugation*. Nat Commun. 2015;6:6925.

49. **Daboussi F, Leduc S, Maréchal A, Dubois G, Guyot V, Perez-Michaut C, et al.** *Genome engineering empowers the diatom Phaeodactylum tricornutum for biotechnology*. Nat Commun. 2014 May 29;5:3831. Available from: https://doi.org/10.1038/ncomms4831

50. **Serif M, Lepetit B, Weißert K, Kroth PG, Rio Bartulos C**. *A fast and reliable strategy to generate TALEN-mediated*

gene knockouts in the diatom Phaeodactylum tricornutum. Algal Res. 2017;23:186–95. Available from: http://www.sciencedirect.com/science/article/pii/S2211926416305550

51. **Nymark M, Sharma AK, Sparstad T, Bones AM, Winge P**. *A CRISPR/Cas9 system adapted for gene editing in marine algae*. Sci Rep. 2016;6(1):24951. Available from: http://www.nature.com/articles/srep24951

52. **Ishino Y, Shinagawa H, Makino K, Amemura M, Nakata A**. *Nucleotide sequence of the iap gene, responsible for alkaline phosphatase isozyme conversion in Escherichia coli, and identification of the gene product.* J Bacteriol. 1987 Dec 1;169(12):5429 LP – 5433. Available from: http://jb.asm.org/content/169/12/5429.abstract

53. **Pourcel C, Salvignol G, Vergnaud G**. *CRISPR elements in Yersinia pestis acquire new repeats by preferential uptake of bacteriophage DNA, and provide additional tools for evolutionary studies*. Microbiology. 2005;151(3):653–63.

54. **Mojica FJM, Díez-Villaseñor C, García-Martínez J, Soria E**. *Intervening Sequences of Regularly Spaced Prokaryotic Repeats Derive from Foreign Genetic Elements*. J Mol Evol. 2005;60(2):174–82. Available from: https://doi.org/10.1007/s00239-004-0046-3

55. **Jinek M, Chylinski K, Fonfara I, Hauer M, Doudna JA, Charpentier E**. *A Programmable Dual-RNA–Guided DNA Endonuclease in Adaptive Bacterial Immunity*. Science (80-). 2012 Aug 16;337(6096):816 LP – 821. Available from: http://science.sciencemag.org/content/337/6096/816.abstract

56. **Malzahn A, Lowder L, Qi Y**. *Plant genome editing with TALEN and CRISPR*. Cell Biosci. 2017 Dec;7(1):21.

57. **Carapetis JR, Steer AC, Mulholland EK, Weber M**. *The global burden of group A streptococcal diseases*. Lancet Infect Dis. 2005;5(11):685–94. Available from: http://www.sciencedirect.com/science/article/pii/S147330990570267X

58. **Mali P, Yang L, Esvelt KM, Aach J, Guell M, DiCarlo JE, et al.** *RNA-Guided Human Genome Engineering via Cas9*. Science (80-). 2013 Feb 15;339(6121):823 LP – 826. Available from: http://science.sciencemag.org/content/339/6121/823.abstract

59. **Wang H, Yang H, Shivalila CS, Dawlaty MM, Cheng AW, Zhang F, et al.** *One-Step Generation of Mice Carrying Mutations in Multiple Genes by CRISPR/Cas-Mediated Genome Engineering*. Cell. 2013 May 9;153(4):910–8. Available from: https://doi.org/10.1016/j.cell.2013.04.025

60. **Niu Y, Shen B, Cui Y, Chen Y, Wang J, Wang L, et al.** *Generation of Gene-Modified Cynomolgus Monkey via Cas9/RNA-Mediated Gene Targeting in One-Cell Embryos*. Cell. 2014 Feb 13;156(4):836–43. Available from: https://doi.org/10.1016/j.cell.2014.01.027

61. **Akhmetov A, Laurent JM, Gollihar J, Gardner EC, Garge RK, Ellington AD, et al.** *Single-step Precision Genome Editing in Yeast Using CRISPR-Cas9*. Bio-protocol. 2018 Mar 20;8(6):e2765. Available from: https://www.ncbi.nlm.nih.gov/pubmed/29770349

62. **Li J-F, Norville JE, Aach J, McCormack M, Zhang D, Bush J, et al.** *Multiplex and homologous recombination–mediated genome editing in Arabidopsis and Nicotiana benthamiana using guide RNA and Cas9*. Nat Biotechnol. 2013 Aug 8;31:688. Available from: https://doi.org/10.1038/nbt.2654

63. **Marraffini L**. *The CRISPR-Cas system of Streptococcus pyogenes: function and applications.* In: Ferretti JJ, Stevens DL, Fischetti VA E, editor. Streptococcus pyogenes : Basic Biology to Clinical Manifestations. Oklahoma City (OK); 2016. Available from: https://www.ncbi.nlm.nih.gov/books/NBK355562/

64. **Koonin E V, Makarova KS, Zhang F**. *Diversity, classification and evolution of CRISPR-Cas systems*. Curr Opin Microbiol. 2017;37:67–78. Available from: http://www.sciencedirect.com/science/article/pii/S1369527417300231

65. **Brouns SJJ, Jore MM, Lundgren M, Westra ER, Slijkhuis RJH, Snijders APL, et al.** *Small CRISPR RNAs Guide Antiviral Defense in Prokaryotes*. Science (80-). 2008 Aug 15;321(5891):960 LP – 964. Available from: http://science.sciencemag.org/content/321/5891/960.abstract

66. **Marraffini LA, Sontheimer EJ**. *CRISPR Interference Limits Horizontal Gene Transfer in Staphylococci by Targeting DNA*. Science (80-). 2008 Dec 19;322(5909):1843 LP – 1845. Available from: http://science.sciencemag.org/content/322/5909/1843.abstract

67. **Sternberg SH, Redding S, Jinek M, Greene EC, Doudna JA**. *DNA interrogation by the CRISPR RNA-guided endonuclease Cas9*. Nature. 2014 Jan 29;507:62. Available from: https://doi.org/10.1038/nature13011

68. **Anders C, Niewoehner O, Duerst A, Jinek M**. *Structural basis of PAM-dependent target DNA recognition by the Cas9 endonuclease*. Nature. 2014 Jul 27;513:569. Available from: http://dx.doi.org/10.1038/nature13579

69. **Jinek M, Jiang F, Taylor DW, Sternberg SH, Kaya E, Ma E, et al.** *Structures of Cas9 Endonucleases Reveal RNA-Mediated Conformational Activation*. Science (80-). 2014 Mar 14;343(6176):1247997. Available from: http://science.sciencemag.org/content/343/6176/1247997.abstract

70. **Gasiunas G, Barrangou R, Horvath P, Siksnys V**. *Cas9–crRNA ribonucleoprotein complex mediates specific DNA cleavage for adaptive immunity in bacteria*. Proc Natl Acad Sci. 2012 Sep 25;109(39):E2579 LP-E2586. Available from: http://www.pnas.org/content/109/39/E2579.abstract

71. **Joung JK, Sander JD**. *TALENs: a widely applicable technology for targeted genome editing*. Nat Rev Mol Cell Biol. 2013 Jan;14(1):49–55. Available from: http://dx.doi.org/10.1038/nrm3486

72. **Ye L, Wang C, Hong L, Sun N, Chen D, Chen S, et al.** *Programmable DNA repair with CRISPRa/i enhanced homology-directed repair efficiency with a single Cas9*. Cell Discov. 2018 Jul 24;4:46. Available from: https://www.ncbi.nlm.nih.gov/pubmed/30062046

73. **Ishizu T, Higo S, Masumura Y, Kohama Y, Shiba M, Higo T, et al.** *Targeted Genome Replacement via Homology-directed Repair in Non-dividing Cardiomyocytes*. Sci Rep. 2017;7(1):9363. Available from: https://doi.org/10.1038/s41598-017-09716-x

74. **Miyaoka Y, Berman JR, Cooper SB, Mayerl SJ, Chan AH, Zhang B, et al.** *Systematic quantification of HDR and NHEJ reveals effects of locus, nuclease, and cell type on genome-editing*. Sci Rep. 2016 Mar 31;6:23549. Available from: https://doi.org/10.1038/srep23549

75. **Pattanayak V, Lin S, Guilinger JP, Ma E, Doudna JA, Liu DR**. *High-throughput profiling of off-target DNA cleavage reveals RNA-programmed Cas9 nuclease specificity*. Nat Biotechnol. 2013 Aug 11;31:839. Available from: https://doi.org/10.1038/nbt.2673

76. **Fu Y, Foden JA, Khayter C, Maeder ML, Reyon D, Joung JK, et al.** *High-frequency off-target mutagenesis induced by CRISPR-Cas nucleases in human cells*. Nat Biotechnol. 2013 Jun 23;31:822. Available from: https://doi.org/10.1038/nbt.2623

77. **Ran FA, Hsu PD, Wright J, Agarwala V, Scott DA, Zhang F**. *Genome engineering using the CRISPR-Cas9 system*. Nat Protoc. 2013 Nov;8(11):2281–308.

78. **Cho SW, Kim S, Kim Y, Kweon J, Kim HS, Bae S, et al.** *Analysis of off-target effects of CRISPR/Cas-derived RNA-guided endonucleases and nickases*. Genome Res. 2014;24(1):132–41.

79. **Mali P, Aach J, Stranges PB, Esvelt KM, Moosburner M, Kosuri S, et al.** *Cas9 transcriptional activators for target specificity screening and paired nickases for cooperative genome engineering*. Nat Biotechnol. 2013;31(9):833–8.

80. **Ma X, Zhu Q, Chen Y, Liu Y-G**. *CRISPR/Cas9 Platforms for Genome Editing in Plants: Developments and Applications*. Mol Plant. 2016;9(7):961–74. Available from: http://www.sciencedirect.com/science/article/pii/S1674205216300314

81. **Fu Y, Sander JD, Reyon D, Cascio VM, Joung JK**. *Improving CRISPR-Cas nuclease specificity using truncated guide RNAs*. Nat Biotechnol. 2014 Mar 26;32(3):279–84. Available from: http://www.ncbi.nlm.nih.gov/pmc/articles/PMC3988262/

82. **Nishimasu H, Ran FA, Hsu PD, Konermann S, Shehata SI, Dohmae N, et al.** *Crystal structure of Cas9 in complex with guide RNA and target DNA*. Cell. 2014;156(5):935–49.

83. **Kleinstiver BP, Pattanayak V, Prew MS, Tsai SQ, Nguyen NT, Zheng Z, et al.** *High-fidelity CRISPR–Cas9 nucleases with no detectable genome-wide off-target effects*. Nature. 2016 Jan 28;529(7587):490–5. Available from: http://dx.doi.org/10.1038/nature16526

84. **Idoko-Akoh A, Taylor L, Sang HM, McGrew MJ**. *High fidelity CRISPR/Cas9 increases precise monoallelic and biallelic editing events in primordial germ cells*. Sci Rep. 2018;8(1):15126. Available from: https://doi.org/10.1038/s41598-018-33244-x

85. **Hill R, Bendall F**. *Crystallization of a photosynthetic reductase from a green plant*. Nature. 1960;187(4735):417.

86. **Rabinowitch EI, Govindjee**. *the Role of Chlorophyll in Photosynthesis*. Sci Am. 1965;213:74–83.

87. **Ogawa T, Obata F, Shibata K**. *Two pigment proteins in spinach chloroplasts*. BBA - Biophys Incl Photosynth. 1966;112(2):223–34.

88. **Kühlbrandt W, Wang DN, Fujiyoshi Y**. *Atomic model of plant light-harvesting complex by electron crystallography*. Nature. 1994;367(6464):614–21. Available from: https://doi.org/10.1038/367614a0

89. **Giardi MT, Pace E**. *Photosynthetic proteins for technological applications*. Trends Biotechnol. 2005;23(5):257–63.

90. **Vinyard DJ, Ananyev GM, Charles Dismukes G**. *Photosystem II: The Reaction Center of Oxygenic Photosynthesis*. Annu Rev Biochem. 2013;82(1):577–606. Available from: http://www.annualreviews.org/doi/10.1146/annurev-biochem-070511-100425

91. **Lepetit B, Goss R, Jakob T, Wilhelm C**. *Molecular dynamics of the diatom thylakoid membrane under different light conditions*. Photosynth Res. 2012;111(1):245–57. Available from: https://doi.org/10.1007/s11120-011-9633-5

92. **Johnson MP**. *Photosynthesis*. Essays Biochem. 2016/10/26. 2016 Oct 31;60(3):255–73. Available from: https://www.ncbi.nlm.nih.gov/pubmed/27784776

93. **Jensen E, Clément R, Maberly SC, Gontero B**. *Regulation of the calvin – Benson – Bassham cycle in the enigmatic diatoms: Biochemical and evolutionary variations on an original theme*. Philos Trans R Soc B Biol Sci. 2017;372(1728).

94. **Kuczynska P, Jemiola-Rzeminska M, Strzalka K**. *Photosynthetic Pigments in Diatoms*. Vol. 13, Marine Drugs. 2015.

95. **Gundermann K, Büchel C**. *Structure and Functional Heterogeneity of Fucoxanthin-Chlorophyll Proteins in Diatoms*. 2014;21–37.

96. **Juhas M, Büchel C**. *Properties of photosystem i antenna protein complexes of the diatom Cyclotella meneghiniana*. J Exp Bot. 2012;63(10):3673–82.

97. **Premvardhan L, Robert B, Beer A, Büchel C**. *Pigment organization in fucoxanthin chlorophyll a/c2 proteins (FCP) based on resonance Raman spectroscopy and sequence analysis*. Biochim Biophys Acta - Bioenerg. 2010;1797(9):1647–56. Available from: http://www.sciencedirect.com/science/article/pii/S0005272810005761

98. **Durnford DG, Deane JA, Tan S, McFadden GI, Gantt E, Green BR**. *A phylogenetic assessment of the eukaryotic light-harvesting antenna proteins, with implications for plastid evolution*. J Mol Evol. 1999;48(1):59–68.

99. **Friedman AL, Alberte RS**. *A Diatom Light-Harvesting Pigment-Protein Complex: Purification and Characterization*. Plant Physiol. 1984;76(2):483–9.

100. **Wang W, Yu L-J, Xu C, Tomizaki T, Zhao S, Umena Y, et al.** *Structural basis for blue-green light harvesting and energy dissipation in diatoms*. Science (80-). 2019;363(6427):eaav0365.

101. **Engelken J, Brinkmann H, Adamska I**. *Taxonomic distribution and origins of the extended LHC (light-harvesting complex) antenna protein superfamily*. BMC Evol Biol . 2010 Jul 30;10:233. Available from: https://www.ncbi.nlm.nih.gov/pubmed/20673336

102. **Büchel C**. *Fucoxanthin-Chlorophyll Proteins in Diatoms: 18 and 19 kDa Subunits Assemble into Different Oligomeric States*. Biochemistry. 2003 Nov 1;42(44):13027–34. Available from: https://doi.org/10.1021/bi0349468

103. **Beer A, Gundermann K, Beckmann J, Büchel C**. *Subunit Composition and Pigmentation of Fucoxanthin−Chlorophyll Proteins in Diatoms: Evidence for a Subunit Involved in Diadinoxanthin and Diatoxanthin Binding*. Biochemistry. 2006 Oct 1;45(43):13046–53. Available from: https://doi.org/10.1021/bi061249h

104. **Lepetit B, Volke D, Szabó M, Hoffmann R, Garab G, Wilhelm C, et al.** *Spectroscopic and Molecular Characterization of the Oligomeric Antenna of the Diatom Phaeodactylum tricornutum*. Biochemistry. 2007 Aug 1;46(34):9813–22. Available from: https://doi.org/10.1021/bi7008344

105. **Bhaya D, Grossman AR**. *Characterization of gene clusters encoding the fucoxanthin chlorophyll proteins of the diatom Phaeodactylum tricornutum*. Nucleic Acids Res. 1993 Sep 25;21(19):4458–66. Available from: https://www.ncbi.nlm.nih.gov/pubmed/8233779

106. **Zhu S-H, Green BR**. *Light-Harvesting and Photoprotection in Diatoms: Identification and Expression of L818-Like Proteins*. Photosynth Energy from Sun. 2008;261–4.

107. **Nymark M, Valle KC, Hancke K, Winge P, Andresen K, Johnsen G, et al.** *Molecular and Photosynthetic Responses to Prolonged Darkness and Subsequent Acclimation to Re-Illumination in the Diatom Phaeodactylum tricornutum*. PLoS One. 2013 Mar 8;8(3):e58722. Available from: https://doi.org/10.1371/journal.pone.0058722

108. **Nymark M, Valle KC, Brembu T, Hancke K, Winge P, Andresen K, et al.** *An Integrated Analysis of Molecular Acclimation to High Light in the Marine Diatom Phaeodactylum tricornutum*. PLoS One. 2009 Nov 3;4(11):e7743. Available from: https://doi.org/10.1371/journal.pone.0007743

109. **Gundermann K, Schmidt M, Weisheit W, Mittag M, Büchel C**. *Identification of several sub-populations in the pool of light harvesting proteins in the pennate diatom Phaeodactylum tricornutum*. Biochim Biophys Acta - Bioenerg. 2013;1827(3):303–10. Available from: http://www.sciencedirect.com/science/article/pii/S0005272812010742

110. **Roach T, Krieger-Liszkay A**. *The role of the PsbS protein in the protection of photosystems I and II against high light in Arabidopsis thaliana*. Biochim Biophys Acta - Bioenerg. 2012;1817(12):2158–65. Available from: http://www.sciencedirect.com/science/article/pii/S0005272812010341

111. **Bailleul B, Rogato A, de Martino A, Coesel S, Cardol P, Bowler C, et al.** *An atypical member of the light-harvesting complex stress-related protein family modulates diatom responses to light*. Proc Natl Acad Sci. 2010 Oct 19;107(42):18214 LP – 18219. Available from: http://www.pnas.org/content/107/42/18214.abstract

112. **Lepetit B, Sturm S, Rogato A, Gruber A, Sachse M, Falciatore A, et al.** *High Light Acclimation in the Secondary Plastids Containing Diatom Phaeodactylum tricornutum is Triggered by the Redox State of the Plastoquinone Pool*. Plant Physiol. 2013 Feb 1;161(2):853 LP – 865. Available from: http://www.plantphysiol.org/content/161/2/853.abstract

113. **Kirst H, Melis A**. *The chloroplast signal recognition particle (CpSRP) pathway as a tool to minimize chlorophyll antenna size and maximize photosynthetic productivity*. Biotechnol Adv. 2014;32(1):66–72. Available from: http://www.sciencedirect.com/science/article/pii/S0734975013001523

114. **Gruber A, Vugrinec S, Hempel F, Gould SB, Maier U-G, Kroth PG**. *Protein targeting into complex diatom plastids: functional characterisation of a specific targeting motif*. Plant Mol Biol. 2007;64(5):519–30. Available from: https://doi.org/10.1007/s11103-007-9171-x

115. **Nymark M, Volpe C, Hafskjold MCG, Kirst H, Vadstein O, Bones AM, et al.** *The chloroplast ALB3b insertase is required for assembly of the light harvesting antenna in diatoms*. Submitt to Plant Physiol.

116. **Maxwell K, Johnson GN**. *Chlorophyll fluorescence--a practical guide*. J Exp Bot. 2000;51(345):659–68. Available from: http://jxb.oxfordjournals.org/cgi/content/abstract/51/345/659

117. **Müller P, Li XP, Niyogi KK**. *Non-photochemical quenching. A response to excess light energy*. Plant Physiol. 2001 Apr;125(4):1558–66. Available from: https://www.ncbi.nlm.nih.gov/pubmed/11299337

118. **Johnsen G, Sakshaug E**. *Biooptical characteristics of PSII and PSI in 33 species (13 pigment groups) of marine phytoplankton, and the relevance for pulseamplitude-modulated and fast-repetition-rate fluorometry*. J Phycol. 2007;43(6):1236–51.

119. **Quick WP, Horton P**. *Studies on the induction of chlorophyll fluorescence in barley protoplasts. II. Resolution of fluorescence quenching by redox state and the transthylakoid pH gradient*. Proc R Soc London Ser B Biol Sci. 2006;220(1220):371–82.

120. **Büchel C, Wilhelm C**. *In vivo analysis of slow chlorophyll fluorescence induction kinetics in algae: progress, problems and perspectives*. Photochem Photobiol. 1993;58(1):137–48.

121. **Goss R, Ann Pinto E, Wilhelm C, Richter M**. *The importance of a highly active and ΔpH-regulated diatoxanthin epoxidase for the regulation of the PS II antenna function in diadinoxanthin cycle containing algae*. J Plant Physiol.

2006;163(10):1008–21. Available from: http://www.sciencedirect.com/science/article/pii/S0176161705003561

122. **Ballottari M, Mozzo M, Girardon J, Hienerwadel R, Bassi R**. *Chlorophyll Triplet Quenching and Photoprotection in the Higher Plant Monomeric Antenna Protein Lhcb5*. J Phys Chem B. 2013 Sep 26;117(38):11337–48. Available from: https://doi.org/10.1021/jp402977y

123. **Mewes H, Richter M**. *Supplementary Ultraviolet-B Radiation Induces a Rapid Reversal of the Diadinoxanthin Cycle in the Strong Light-Exposed Diatom Phaeodactylum tricornutum*. Plant Physiol. 2002 Nov 1;130(3):1527 LP – 1535. Available from: http://www.plantphysiol.org/content/130/3/1527.abstract

124. **Grouneva I, Jakob T, Wilhelm C, Goss R**. *The regulation of xanthophyll cycle activity and of non-photochemical fluorescence quenching by two alternative electron flows in the diatoms Phaeodactylum tricornutum and Cyclotella meneghiniana*. Biochim Biophys Acta - Bioenerg. 2009;1787(7):929–38. Available from: http://www.sciencedirect.com/science/article/pii/S0005272809000528

125. **Zhang B, Luo Z, Liu J, Ding X, Li J, Cai K**. *Cytochrome c end-capped mesoporous silica nanoparticles as redox-responsive drug delivery vehicles for liver tumor-targeted triplex therapy in vitro and in vivo*. J Control Release. 2014;192:192–201. Available from: http://www.sciencedirect.com/science/article/pii/S0168365914004465

126. **de Oliveira LF, Bouchmella K, Gonçalves K de A, Bettini J, Kobarg J, Cardoso MB**. *Functionalized Silica Nanoparticles As an Alternative Platform for Targeted Drug-Delivery of Water Insoluble Drugs*. Langmuir. 2016 Apr 5;32(13):3217–25. Available from: https://doi.org/10.1021/acs.langmuir.6b00214

127. **Bozarth A, Maier U-G, Zauner S**. *Diatoms in biotechnology: modern tools and applications*. Appl Microbiol Biotechnol. 2009;82(2):195–201. Available from: https://doi.org/10.1007/s00253-008-1804-8

128. **Rajamani S, Siripornadulsil S, Falcao V, Torres M, Colepicolo P, Sayre R**. *Phycoremediation of Heavy Metals Using Transgenic Microalgae*. Strategies. 2007;

129. **Voigt RG, Jensen CL, Fraley JK, Rozelle JC, Brown FR, 3rd, et al.** *Relationship between omega3 long-chain polyunsaturated fatty acid status during early infancy and neurodevelopmental status at 1 year of age*. J Hum Nutr Diet. 2002;15(2):111–20. Available from: https://doi.org/10.1046/j.1365-277X.2002.00341.x

130. **Calder PC**. *n−3 Polyunsaturated fatty acids and inflammation: From molecular biology to the clinic*. Lipids. 2003 Apr 1;38(4):343–52. Available from: https://doi.org/10.1007/s11745-003-1068-y

131. **Hamilton ML, Warwick J, Terry A, Allen MJ, Napier JA, Sayanova O**. *Towards the Industrial Production of Omega-3 Long Chain Polyunsaturated Fatty Acids from a Genetically Modified Diatom Phaeodactylum tricornutum*. PLoS One. 2015 Dec 14;10(12):e0144054–e0144054. Available from: https://www.ncbi.nlm.nih.gov/pubmed/26658738

132. **Melis A**. *Solar energy conversion efficiencies in photosynthesis: Minimizing the chlorophyll antennae to maximize efficiency*. Plant Sci. 2009;177(4):272–80. Available from: http://www.sciencedirect.com/science/article/pii/S0168945209001861

133. **Kirst H, Garcia-Cerdan JG, Zurbriggen A, Ruehle T, Melis A**. *Truncated photosystem chlorophyll antenna size in the green microalga Chlamydomonas reinhardtii upon deletion of the TLA3-CpSRP43 gene*. Plant Physiol. 2012/10/05. 2012 Dec;160(4):2251–60. Available from: https://www.ncbi.nlm.nih.gov/pubmed/23043081

134. **Shin W-S, Lee B, Kang NK, Kim Y-U, Jeong W-J, Kwon J-H, et al.** *Complementation of a mutation in CpSRP43 causing partial truncation of light-harvesting chlorophyll antenna in Chlorella vulgaris*. Sci Rep. 2017;7(1):17929. Available from: https://doi.org/10.1038/s41598-017-18221-0

135. **Kirst H, Formighieri C, Melis A**. *Maximizing photosynthetic efficiency and culture productivity in cyanobacteria upon minimizing the phycobilisome light-harvesting antenna size*. Biochim Biophys Acta - Bioenerg. 2014;1837(10):1653–64. Available from: http://www.sciencedirect.com/science/article/pii/S0005272814005362

136. **Formighieri C**. *Solar-to-fuel conversion in algae and cyanobacteria*. 2015; Available from: http://link.springer.com/10.1007/978-3-319-16730-5

137. *E.Z.N.A.® Plasmid DNA Mini Kit I Spin Protocol* . Omega Bio-tek Inc. 2017 [cited 2018 Jul 16]. p. 10–6. Available from: http://omegabiotek.com/store/wp-content/uploads/2013/05/D6942.D6943.D6945-January-2017-Online.pdf

138. **Kroth PG**. *Genetic transformation: A tool to study protein targeting in diatoms*. Methods Mol Biol. 2007;390:257–67. Available from: http://www.embase.com/search/results?subaction=viewrecord&from=export&id=L350183664%0Ahttp://dx.doi.org/10.1385/1-59745-466-4:257

139. *Biolistic ® PDS-1000/He Particle Delivery System*. Bio-Rad Laboratories. [cited 2018 Jul 16]. Available from: http://www.bio-rad.com/webroot/web/pdf/lsr/literature/M1652249.pdf

140. **Dehairs J, Talebi A, Cherifi Y, Swinnen J V**. *CRISP-ID: decoding CRISPR mediated indels by Sanger sequencing.* Sci Rep. 2016 Jul 1;6:28973. Available from: https://doi.org/10.1038/srep28973

141. **Sakshaug E, Bricaud A, Dandonneau Y, Falkowski PG, Kiefer DA, Legendre L, et al.** *Parameters of photosynthesis: Definitions, theory and interpretation of results*. J Plankton Res. 1997;19(11):1637–70.

142. **Schreiber U, Bilger W**. *Progress in Chlorophyll Fluorescence Research: Major Developments During the Past Years in Retrospect*. In: Behnke H-D, Lüttge U, Esser K, Kadereit JW, Runge M, editors. Berlin, Heidelberg: Springer Berlin Heidelberg; 1993. p. 151–73. Available from: https://doi.org/10.1007/978-3-642-78020-2_8

143. **Hofstraat JW, Peeters JCH, Snel JFH, Geel C**. *Simple determination of photosynthetic efficiency and photoinhibition of Dunaliella tertiolecta by saturating pulse fluorescence measurements*. Mar Ecol Prog Ser. 1994;103(1–2):187–96.

144. **Jassby AD, Platt T**. *Mathematical formulation of the relationship between photosynthesis and light for phytoplankton*. Limnol Oceanogr. 1976;21(4):540–7.

145. **Falkowski PG, Raven JA**. *Aquatic Photosynthesis*. Princeton University Press.; 2007.

146. **Serôdio J, Cruz S, Vieira S, Brotas V**. *Non-photochemical quenching of chlorophyll fluorescence and operation of the xanthophyll cycle in estuarine microphytobenthos*. J Exp Mar Bio Ecol . 2005;326(2):157–69. Available from: http://www.sciencedirect.com/science/article/pii/S0022098105002601

147. **Rodríguez F, Chauton M, Johnsen G, Andresen K, Olsen LM, Zapata M**. *Photoacclimation in phytoplankton: Implications for biomass estimates, pigment functionality and chemotaxonomy*. Mar Biol. 2006;148(5):963–71.

148. **Ruban A, Lavaud J, Rousseau B, Guglielmi G, Horton P, Etienne AL**. *The super-excess energy dissipation in diatom algae: Comparative analysis with higher plants*. Photosynth Res. 2004;82(2):165–75.

149. **Andersen RA, editor**. *Algal culturing techniques*. 1st ed. Academic Press; 2005. 596 p.

150. **Guillard RRL, Ryther JH**. *Studies of Marine Planktonic Diatoms: I. Cyclotella Nana Hustedt, and Detonula Confervacea (Cleve) Gran.* Can J Microbiol. 1962;8(2):229–39. Available from: http://www.nrcresearchpress.com/doi/abs/10.1139/m62-029

151. **Perez EB, Pina IC, Rodriguez LP**. *Kinetic model for growth of Phaeodactylum tricornutum in intensive culture photobioreactor*. Biochem Eng J. 2008;40:520–5.

152. **Hansen HP, Koroleff F**. *Determination of nutrients*. Methods of Seawater Analysis. 2007. (Wiley Online Books). Available from: https://doi.org/10.1002/9783527613984.ch10

153. **Bligh EG, Dyer WJ**. *A Rapid Method of Total Lipid Extraction and Purification*. Can J Biochem Physiol. 1959;37(1):911–7.

154. **Behrenfeld MJ, Prasil O, Babin M, Bruyant F**. *In search of a physiological basis for covariations in light-limited and light-saturated photosynthesis*. J Phycol. 2004;40(1):4–25.

155. **Geider RJ, Roche JLA**. *Redfield revisited: variability of C:N:P in marine microalgae and its biochemical basis*. Eur J Phycol. 2002/04/09. 2002;37(1):1–17. Available from: https://www.cambridge.org/core/article/redfield-revisited-variability-of-cratio-nratio-p-in-marine-microalgae-and-its-biochemical-basis/A32314CF87D1645F8420D8D471B7C05F

156. **Jeong J, Baek K, Kirst H, Melis A, Jin E**. *Loss of CpSRP54 function leads to a truncated light-harvesting antenna size in Chlamydomonas reinhardtii*. Biochim Biophys Acta - Bioenerg. 2017;1858(1):45–55. Available from: http://www.sciencedirect.com/science/article/pii/S0005272816306508

157. **Jenny A, Mark W, Robin G. W, Caroline A. H, Alexander V. R, Peter H, et al.** *Absence of the Lhcb1 and Lhcb2 proteins of the light-harvesting complex of photosystem II - Effects on photosynthesis, grana stacking and fitness*. Plant J . 2003;35(3):350–61. Available from: http://doi.wiley.com/10.1046/j.1365-313X.2003.01811.x

158. **Wobbe L, Bassi R, Kruse O**. *Multi-Level Light Capture Control in Plants and Green Algae*. Trends Plant Sci. 2016;21(1):55–68. Available from: http://www.sciencedirect.com/science/article/pii/S1360138515002526

159. **Mussgnug JH, Wobbe L, Elles I, Claus C, Hamilton M, Fink A, et al.** *NAB1 Is an RNA Binding Protein Involved in the Light-Regulated Differential Expression of the Light-Harvesting Antenna of Chlamydomonas reinhardtii* Plant Cell. 2005 Dec 1;17(12):3409 LP – 3421. Available from: http://www.plantcell.org/content/17/12/3409.abstract

160. **Ferrante P, Ballottari M, Bonente G, Giuliano G, Bassi R**. *LHCBM1 and LHCBM2/7 polypeptides, components of major LHCII complex, have distinct functional roles in photosynthetic antenna system of Chlamydomonas reinhardtii*. J Biol Chem. 2012;287(20):16276–88.

161. **Kirst H, García-Cerdán JG, Zurbriggen A, Melis A**. *Assembly of the Light-Harvesting Chlorophyll Antenna in the Green Alga Chlamydomonas reinhardtii Requires Expression of the TLA2-CpFTSY Gene*. Plant Physiol. 2012 Feb 1;158(2):930 LP – 945. Available from: http://www.plantphysiol.org/content/158/2/930.abstract

162. **Lavaud J, Lepetit B**. *An explanation for the inter-species variability of the photoprotective non-photochemical chlorophyll fluorescence quenching in diatoms*. Biochim Biophys Acta - Bioenerg. 2013;1827(3):294–302.

163. **Jahns P, Krause GH**. *Xanthophyll cycle and energy-dependent fluorescence quenching in leaves from pea plants grown under intermittent light*. Planta. 1994;192(2):176–82. Available from: https://doi.org/10.1007/BF00194450

164. **Havaux M, Dall'Osto L, Bassi R**. *Zeaxanthin Has Enhanced Antioxidant Capacity with Respect to All Other Xanthophylls in Arabidopsis Leaves and Functions Independent of Binding to PSII Antennae*. Plant Physiol . 2007 Dec 1;145(4):1506 LP – 1520. Available from: http://www.plantphysiol.org/content/145/4/1506.abstract

165. **Mann JE, Myers J**. *On Pigments, Growth, and Photosynthesis of Phaeodactylum Tricornutum*. J Phycol. 1968;4(4):349–55.

166. **Perry MJ**. *Phosphate utilization by an oceanic diatom in phosphorus-limited chemostat culture and in the oligotrophic waters of the central North Pacific*. Limnol Oceanogr. 1976;21(1):88–107.

167. **Dortch Q, Clayton JR, Thoreson SS, Bressler SL, Ahmed SI**. *Response of marine phytoplankton to nitrogen deficiency: Decreased nitrate uptake vs enhanced ammonium uptake*. Mar Biol. 1982;70(1):13–9.

168. **Lomas MW, Gilbert PM**. *Comparisons of nitrate uptake, storage, and reduction in marine diatoms and flagellates*. J Phycol. 2000;36(5):903–13.

169. **Goldberg ED, Walker TJ, Whisenand A**. *Phosphate Utilization By Diatoms*. Biol Bull. 1951;101(3):274–84.

170. **Osborne BA, Geider RJ**. *Effect of nitrate-nitrogen limitation on photosynthesis of the diatom Phaeodactylum tricornutum Bohlin (Bacillariophyceae)*. Plant Cell Environ. 1986;9(8):617–25.

171. **Redfield AC**. *The Biological Control of Chemical Factors in the Environment*. Sci Prog. 1960;11:150–70.

172. **Redfield AC**. *On the proportions of organic derivatives in sea water and their relation to the composition of plankton*. James Johnstone Meml Vol. 1933;176–92. Available from: http://cmore.soest.hawaii.edu/summercourse/2012/documents/bronk_05-30-12/Redfield_1934.pdf

173. **Wykoff DD, Davies JP, Melis A, Grossman AR**. *The Regulation of Photosynthetic Electron Transport during Nutrient Deprivation in Chlamydomonas reinhardtii* Plant Physiol. 1998 May 1;117(1):129 LP – 139. Available from: http://www.plantphysiol.org/content/117/1/129.abstract

174. **Li W, Gao K, Beardall J**. *Interactive Effects of Ocean Acidification and Nitrogen-Limitation on the Diatom Phaeodactylum tricornutum*. PLoS One. 2012;7(12).

175. **Kolber Z, Zehr J, Falkowski P**. *Effects of Growth Irradiance and Nitrogen Limitation on Photosynthetic Energy Conversion in Photosystem II*. Plant Physiol. 1988 Nov;88(3):923–9. Available from: https://www.ncbi.nlm.nih.gov/pubmed/16666405

176. **Berges JA, Charlebois DO, Mauzerall DC, Falkowski PG**. *Differential Effects of Nitrogen Limitation on Photosynthetic Efficiency of Photosystems I and II in Microalgae*. Plant Physiol. 1996 Feb;110(2):689–96. Available from: https://www.ncbi.nlm.nih.gov/pubmed/12226211

177. **Huete-Ortega M, Okurowska K, Kapoore RV, Johnson MP, Gilmour DJ, Vaidyanathan S**. *Effect of ammonium and high light intensity on the accumulation of lipids in Nannochloropsis oceanica (CCAP 849/10) and Phaeodactylum tricornutum (CCAP 1055/1)*. Biotechnol Biofuels. 2018 Mar 9;11:60. Available from: https://www.ncbi.nlm.nih.gov/pubmed/29541157

178. **Frada MJ, Burrows EH, Wyman KD, Falkowski PG**. *Quantum requirements for growth and fatty acid biosynthesis in the marine diatom Phaeodactylum tricornutum (Bacillariophyceae) in nitrogen replete and limited conditions*. J Phycol. 2013;49(2):381–8.

179. **Oey M, Ross IL, Stephens E, Steinbeck J, Wolf J, Radzun KA, et al.** *RNAi Knock-Down of LHCBM1, 2 and 3 Increases Photosynthetic H2 Production Efficiency of the Green Alga Chlamydomonas reinhardtii*. PLoS One. 2013 Apr 16;8(4):e61375. Available from: https://doi.org/10.1371/journal.pone.0061375

180. **Guillard RRL**. *Culture of Phytoplankton for Feeding Marine Invertebrates*. Culture of Marine Invertebrate Animals. 1975. 29–60 p. Available from: http://link.springer.com/10.1007/978-1-4615-8714-9_3

181. **Bertani G**. *Studies on lysogenesis. III. Superinfection of lysogenic Shigella dysenteriae with temperate mutants of the carried phage.* J Bacteriol. 1954;67(6):696–707.

9 Appendices

9.1 Appendix A: Medium and culture

9.1.1 Appendix A1: Diatom media and culture components

Guillard's F/2 medium [180] is a widely common enriched sea-water based medium used to grow marine algae. It is developed on the original "F medium" used by Guillard and Ryther (1962). Liquid medium was made by vacuum-filtering natural seawater from the Trondheim fjord through a sterile 0.22 µm filter (Whatman®). The water was autoclaved at 120°C for 20 minutes and allowed to cool in a cold room (~4°C). Components were added in succession (from top to bottom) through a sterile syringe filter (0.22 µm) as presented below (Suppl. Table 1). When making agar for diatom F/2 plates, approximately half of the filtered sea-water was substituted with MQ-water. BactoAgar (VWR Chemicals) was added, and the solution was only allowed to cool to ~60°C after autoclaving. Optionally, the antibiotic Zeocin was added to provide selection for transformed cells. 1.0 mL Zeocin for every 1 L medium yields a concentration of 100 µg mL^{-1}. Stock solution components are listed in Suppl. Table 2.

Supplementary Table 1: Guillard's F/2 medium for cultivating diatoms. Asterisk (*): Stock solution. Double Asterisk (**): Optional selection vector.

	F/2 liquid (1 L)	**F liquid (1 L)**	**F/2 agar (1 L)**	**unit**
Filtered (0.2 µm) Sea-water	995 (cold, autoclaved)	995 (cold, autoclaved)	500.0	mL
Milli-Q Water	-	-	490.0	mL
BactoAgar	-	-	10.0	g
$NaNO_3$ *	1.0	2.0	1.0	mL
$NaH_2PO_4 \cdot 2\ H_2O$ *	1.0	2.0	1.0	mL
$Na_2SiO_3 \cdot 9\ H_2O$ *	1.0	2.0	1.0	mL
Trace Metals *	1.0	2.0	1.0	mL
Vitamins *	0.5	1	0.5	mL
Zeocin (100 mg mL^{-1}) **	0.5	1	0.5	mL

Supplementary Table 2: Composition of stock solutions utilized used to prepare F/2 and F media. All components were pre-existing at the CMBG group, and prepared in the following concentrations.

Stock	Chemical component	Primary Stock Concentration ($g\ L^{-1}$)	F/2-medium Concentration (M)	F-medium Concentration (M)
Sodium Nitrate	$NaNO_3$	75.00	8.82E-04	1.76E-03
Monosodium Phospho-dihydrate	$NaH_2PO_4 \cdot 2\ H_2O$	5.00	3.62E-05	7.24E-05
Sodium Silicate Nonahydrate	$Na_2SiO_3 \cdot 9\ H_2O$	30.00	1.06E-04	2.12E-04
Trace Metals	$FeCl_3 \cdot 6\ H_2O$	3.15	1.17E-05	2.34E-05
	$Na_2EDTA \cdot 2\ H_2O$	4.36	1.17E-05	2.34E-05
	$CuSO_4 \cdot 5\ H_2O$	9.80	3.93E-08	7.86E-08
	$Na_2MoO_4 \cdot 2\ H_2O$	6.30	2.60E-08	5.20E-08
	$ZnSO_4 \cdot 7\ H_2O$	22.00	7.65E-08	1.53E-07
	$CoCl_2 \cdot 6\ H_2O$	10.00	4.20E-08	8.40E-08
	$MnCl_2 \cdot 4\ H_2O$	180.00	9.10E-07	1.82E-06
Vitamins	Thiamine HCl (vit. B^1)	0.20	2.96E-07	5.92E-07
	Biotin (Vit. H)	1.00	2.05E-09	4.10E-09
	Cyanocobalamin (Vit. B^{12})	1.00	3.69E-10	7.38E-10
Zeocin	$C_{55}H_{83}N_{19}O_{21}S_2Cu$	100.00	[50-100] $\mu g\ mL^{-1}$	

9.1.2 Appendix A2: LB Culture medium.

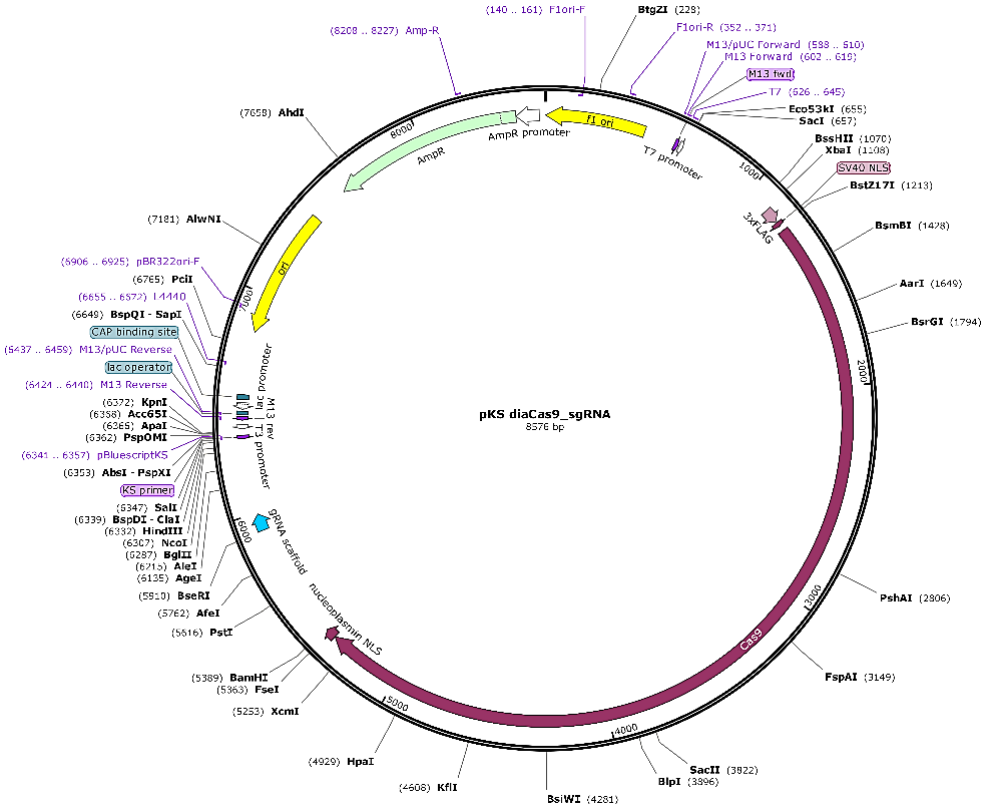
Luria Bertani's (LB) broth (liquid) and agar medium for cultivation of bacteria [181]. Prior to cloning and bacterial amplification of plasmids, LB medium was made by mixing Pancreatic Peptone (VWR Chemicals), Bacto™ Yeast-Extract (BP-Becton, Dickinson & Comp.), Sodium-Chloride (NaCl, Merck Emsure®) and BactoAgar (VWR Chemicals) with MQ water (Suppl. Table 3). The solution was autoclaved at 120°C for 20 minutes. When preparing agar, the solution was left to cool to ~60°C. If selection was needed, an appropriate quantity of antibiotic (ampicillin/zeocin) was added to the solution before pouring onto petri-plates under a sterile hood.

Supplementary Table 3: LB liquid and agar components used in cultivation of E. coli.

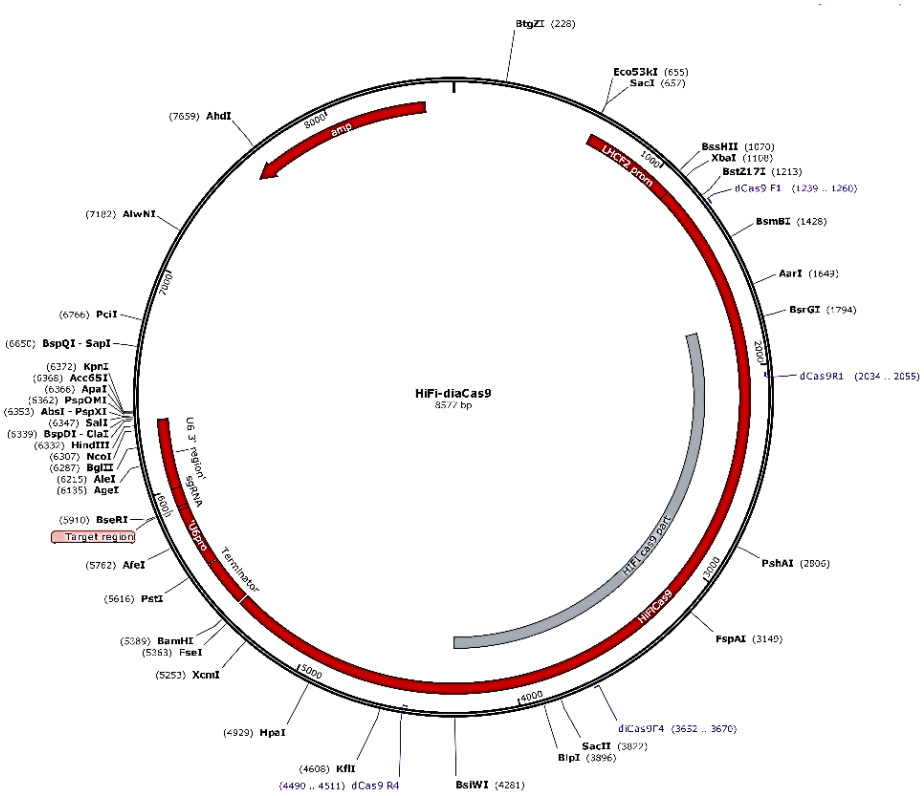
	LB liquid (1 L)	LB agar (1 L)	unit
Milli-Q water	990	990	mL
Peptone	10	10	g
Yeast extract	5	5	g
NaCl	5	5	g
BactoAgar	-	15	g
Ampicillin (100 $mg\ mL^{-1}$)	1	1	mL

9.2 Appendix B: Plasmid vector maps

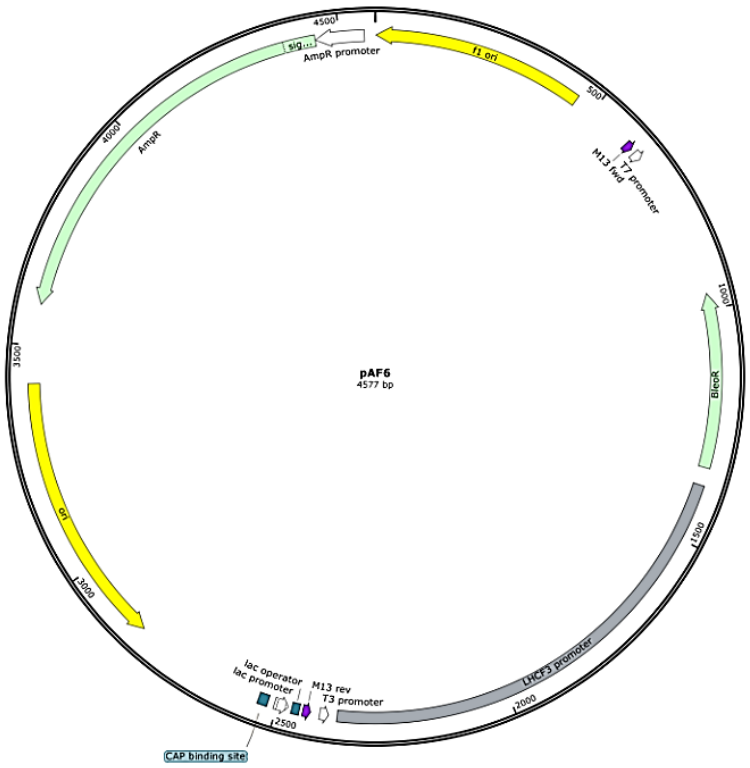
Three plasmids were utilized in the experimental work of this thesis: pKS_diaCas9_sgRNA (Suppl. Fig. 1), pKS_HiFi-diaCas9_sgRNA (Suppl. Fig. 2) and pAF6 (Suppl. Fig. 3). All maps were exported from SnapGene™ (GSL Biotech, Chicago, IL).



Supplementary Figure 1: The conventional Cas9-carrying vector pKS_diaCas9_sgRNA (8576 bp).



Supplementary Figure 2: The modified HiFi Cas9-carrying vector pKS_HiFi-diaCas9_sgRNA (8577 bp).



Supplementary Figure 3: The co-transformed selection vector pAF6 (4577 bp).

9.3 Appendix C: PCR

9.3.1 Appendix C1: PCR reaction setup

Screening for putative mutants was conducted through a series of PCR amplification reactions. The reactions were prepared using primers (10 mM), stock dNTPs (100 mM), MQ water, and stock DreamTaq™ buffer and polymerase (Suppl. Table 4). Green DreamTaq™ buffer was used for identification of Cas9 only, while the colorless buffer was used when PCR amplicons were to be sequenced. In case of high primer-dimer levels, PCR reactions were repeated with increased template quantity (4 or 7 µL).

Supplementary Table 4: List of components in PCR amplification reactions.

	Mastermix (1x reactions)	
Component	**20 µL**	**50 µL**
Milli-Q water	14.25	35.63
DreamTaq™ green/colorless buffer (stock)	2.00	5.00
dNTPs (100 mM)	0.50	1.25
FP (10 mM)	0.50	1.25
RP (10 mM)	0.50	1.25
DreamTaq™ Polymerase (stock)	0.25	0.63
Lysate template	2.00	5.00

9.3.2 Appendix C2: List of PCR primers

To limit the workload of screening, putative mutants were identified by positive amplification of a sequence within hifiCas9 and diaCas9 plasmids by using the M13R and LHCFm_PAM1F primers (Suppl. Table 5). Identified putative mutants were further screened by amplification of *LHCF*-sequences, with various combinations of LHCF forward (F) and reverse (R) primers.

Supplementary Table 5: List of primers used to identify pKS_diaCas9_sgRNA and pKS_HiFiCas9_sgRNA plasmids (LHCFm_PAM1F, M13R) and genomic Cas9-targeted genes (LHCF1-LHCF11). Forward and reverse orientations are labelled F and R, respectively.

Primer ID	Orientation	5'- Sequence	Purpose
LHCF1F	F	TGTTTTGACTGCAAGATCAGC	*LHCF*-sequencing
LHCF1R	R	CGGAAGTCGCCAACAAACTCT	
LHCF3F	F	GCTTGAACACGATCCTTGTGA	
LHCF3R	R	TGAACTGGGTCTCCTCGTCA	
LHCF4F	F	KCCCTCATAATCTCTGTTAGA	
LHCF4R	R	CACCATAAGGGCAAGAATTCC	
LHCF2F	F	CAAATCCTGGTTGAACACGTA	
LHCF2R	R	CCTGGTTAAGCTCAATGGCAC	
LHCF5F	F	AGTGCTATGCCTGAAATGTGC	
LHCF5R	R	TCTAACCGGAGGAACATCAAC	
LHCF11F	F	ACACCACCTGGATCTTCACAT	
LHCF11R	R	CAACTTTTCGTGGACCATCAA	
LHCFm_PAM1F	F	TCGATGATCTCAACGTAACGGAGA	Cas9-screening
M13R	R	CAGGAAACAGCTATGAC	

9.3.3 Appendix C3: PCR amplification protocol

The following PCR protocol was designed for amplification of 563-793 bp sequences with the DreamTaq™ polymerase (Suppl. Table 6). Elongation time was raised to 30s for sequences of increased length.

Supplementary Table 6: PCR protocol used for amplification.

#	Step	Temperature (°C)	Duration (mm:ss)	Cycles
1	Initial denaturation	95	2:00	1
2	Denaturation	95	0:30	
3	Annealing	62	0:30	28 (step #2-5)
4	Elongation	72	0:25-0:30	
5	Final elongation	72	10:00	1
6	Incubation	4	∞	-

9.3.4 Appendix C4: Agarose gel electrophoresis

Agarose gel electrophoresis was undertaken after PCR reactions to verify amplifications, and to identify unwanted primer dimers. A 1% agarose gel was utilized on most occasions (Suppl. Table 7).

Supplementary Table 7: Gel electrophoresis component list.

Component	1%	10%	unit
Agarose	1.0	10.0	g
TAE-buffer	100.0	100.0	mL
GelRed	2.5	2.5	µL

9.3.5 Appendix C5: ExoSAP-IT PCR product clean-up

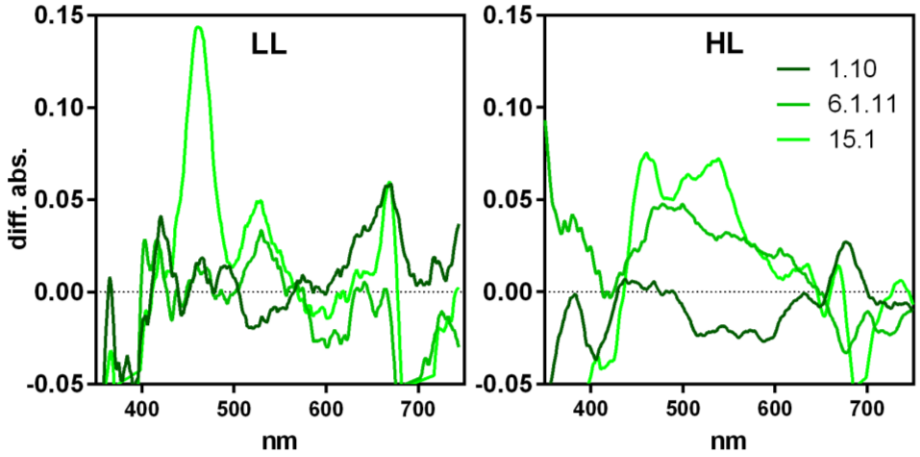
After PCR, sequences were purified by use of the ExoSAP-IT™ PCR Product Clean-up Reagent. The clean-up is enzymatic and utilizes an exonuclease that hydrolyzes single-stranded DNA at 37°C. PCR product (5 µL) was cleaned using 2 µL ExoSAP-IT reaction solution and by incubating at enzyme-activating temperature (37°C) followed by deactivation at 80°C (Suppl. Table 8).

Supplementary Table 8: ExoSAP-IT PCR product Clean-Up protocol.

Step	T (°C)	Minutes
1	37	15
2	80	15

9.5 Appendix D: Differential absorbance and pigment ratios

The differential absorbance was also calculated for the remaining LHCFm mutant lines (Suppl. Fig. 4).



Supplementary Figure 4: Differential absorbance of all LHCFm cell lines. Differences were calculated by subtracting the relative absorbance of mutants from WT absorbance.

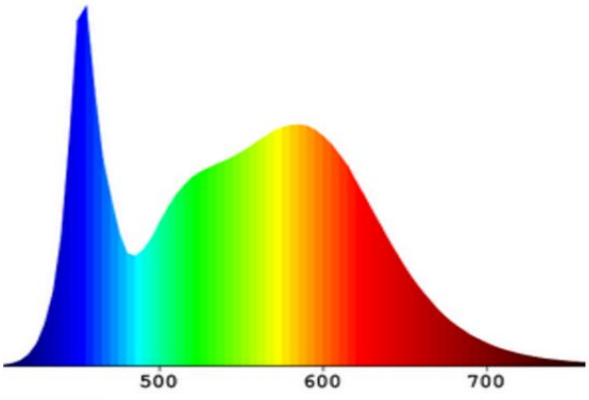
The Fuco:Chl a ratio and DES index values of Fig. 3.4 is based on the following ratio numbers (Suppl. Table 9).

Supplementary Table 9: Fuco:Chl *a* ratio and DES-index.

	Fuco:Chl *a* (mol:mol)		**DES (Dtx / (Dtx+Ddx)) (mol:mol)**	
	HL (µ±SEM)	**LL (µ±SEM)**	**HL (µ±SEM)**	**LL (µ±SEM)**
WT (n = 9)	0.65 ± 0.050	0.73 ± 0.062	0.36 ± 0.025	0.11 ± 0.016
1.10 (n = 6)	0.58 ± 0.016	0.62 ± 0.021	0.24 ± 0.008	0.10 ± 0.006
6.1.11 (n = 3)	0.37 ± 0.021	0.54 ± 0.110	0.23 ± 0.001	0.09 ± 0.003
15.1 (n = 3)	0.32 ± 0.016	0.40 ± 0.005	0.25 ± 0.015	0.11 ± 0.004

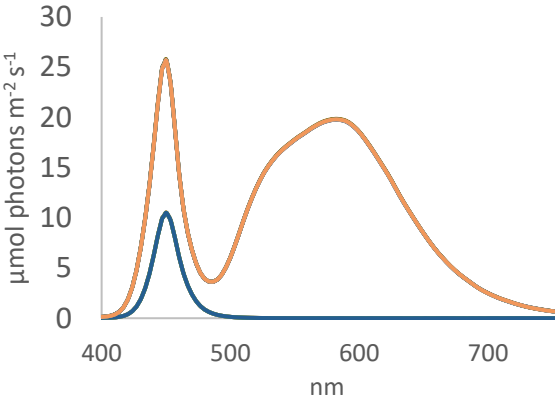
9.7 Appendix E: Light sources

Cultures of WT and mutant *P. tricornutum* were grown in different conditions. During routine and experimental work at the laboratory of CMBG, cells were grown in a cool white LED light source (Suppl. Fig. 5).



Supplementary Figure 5: Spectral irradiance distribution of the cool white LED light-source (Solarmass EQUILIGHT) used for routine and experimental work at the CMBG laboratory.

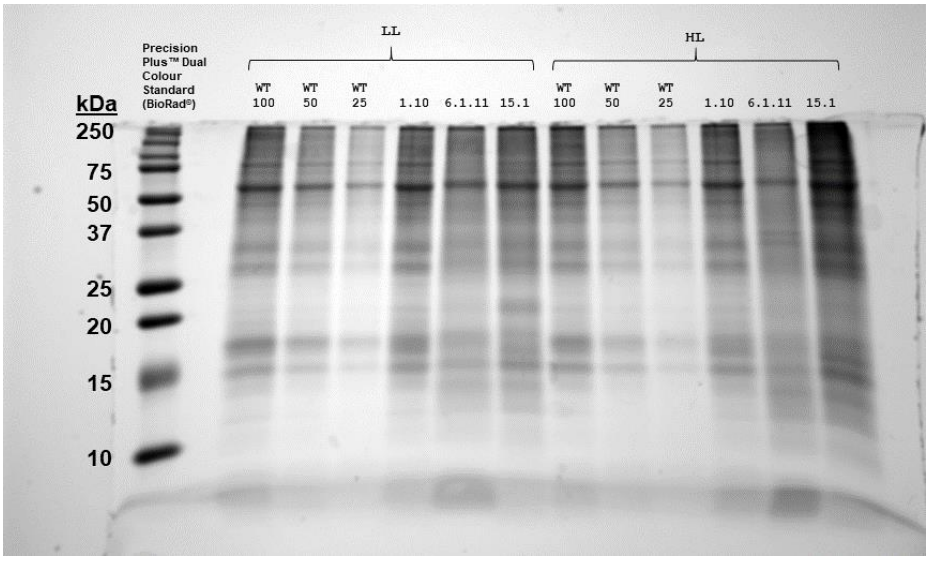
During the photobioreactor trials at SINTEF Brattørkaia, cells were grown using a cool white LED light-source with a spectral irradiance resembling (Suppl. Fig. 6) to that of routine-growth.



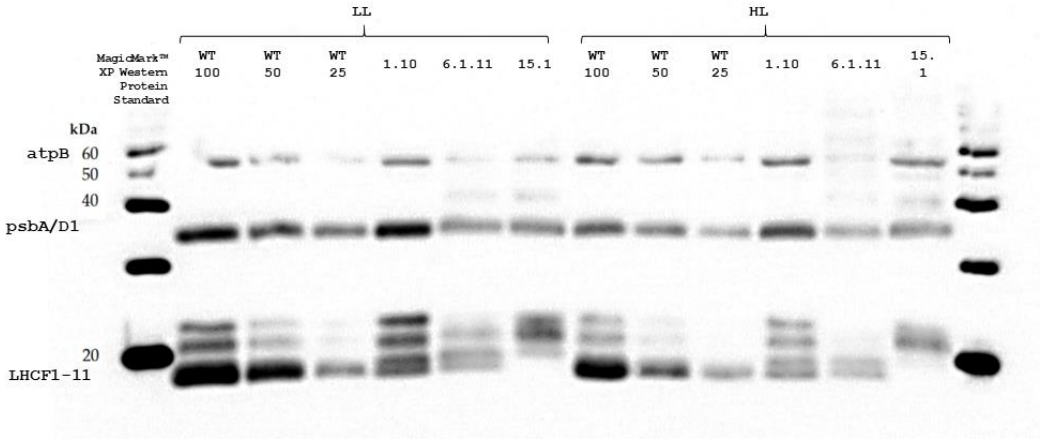
Supplementary Figure 6: Spectral irradiance distribution of the cool white LED light-source used during the photobioreactor trials at SINTEF.

9.9 Appendix F: Western blot and Coomassie gel

In total two western blots and a Coomassie gel were set up to infer the FCP contents of LHCFm mutants. The western blot membrane displaying α-cmFCP-targeted FCPs is presented in Fig. 3.3 of the main text. The Coomassie gel stained with SimplyBlue™ SafeStain (Invitrogen™) serves as a loading control and is presented in Suppl. Fig. 7. Another membrane displaying potentially several FCP-bands by use of the anti-FCP antibody made against a 19 kDa FCP of *Heterosigma akashiwa* (Agrisera®, Catalog #: AS174116) is presented in Suppl. Fig. 8.



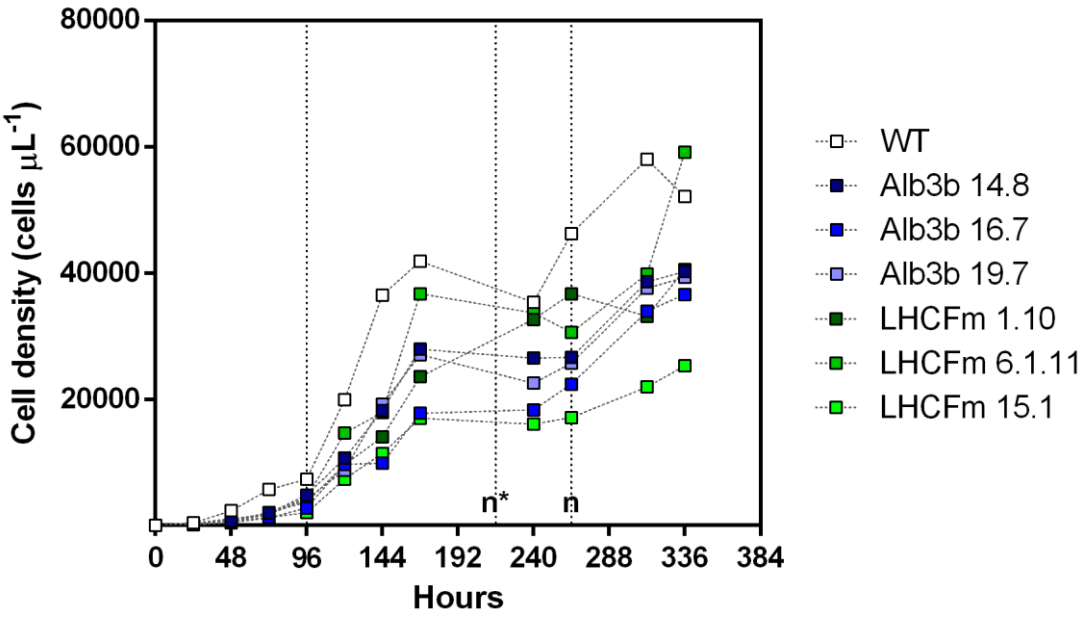
Supplementary Figure 7: Coomassie gel stained with SimplyBlue™ SafeStain (Invitrogen™).



Supplementary Figure 8: Western blot with the primary antibody made from the 19 kDa FCPs of *Heterosigma akashiwa* (Agrisera®, Catalog #: AS174116).

9.10 Appendix G: Linearized growth

The growth curves of cultures during the photobioreactor trials were also plotted on a linear axis (Suppl. Fig. 9).



Supplementary Figure 9: Growth curves of photobioreactor cultures given a linearized y-axis.

9.12 Appendix H: PhytoPAM protocol

Supplementary Table 10: Phyto-PAM protocol for obtaining variable fluorescence measurements. Note that the initial darkness-incubation time lasted 3 minutes (180 s).

Step	E_{PAR} (µmol photons m^{-2} s^{-1})	Time (s)
0	0 (Darkness)	180
1	4	30
2	32	30
3	64	30
4	128	30
5	192	30
6	256	30
7	384	30
8	448	30
9	512	30
10	640	30
11	832	30
12	1088	30
13	1216	30
	Tot.	570

9.14 Appendix I: SDS-PA gel composition

Supplementary Table 11: Reagents used to prepare SDS-polyacrylamide separation and stacking gels.

Gel type	Reagent	Volume (mL)
15% Acrylamide separation gel	H_2O	11.40
	1.5M Tris (pH 8.8)	12.50
	10% SDS	0.50
	10% APS	0.50
	30% Acrylamide mix	25.00
	TEMED	0.02
	Total volume (mL)	49.92
15% Acrylamide stacking gel	H_2O	13.60
	1.5M Tris (pH 6.8)	2.50
	10% SDS	0.20
	10% APS	0.20
	30% Acrylamide mix	3.40
	TEMED	0.02
	Total volume (mL)	19.92