Characterization of Aureochromes in the diatom *Phaeodactylum tricornutum*

Dissertation submitted for the degree of

Doctor of Natural Science

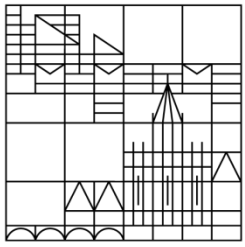
(Dr. rer. nat.)

Presented by

Manuel Serif

At the

Universität Konstanz



Faculty of Sciences

Department of Biology

Date of the oral examination: October 16th, 2017

First referee: Prof. Dr. Peter G. Kroth

Second referee: Prof. Dr. Erika Isono

Abstract

Aureochromes are a novel group of blue-light dependent transcription factors restricted to Stramenopiles. Phylogenetic analysis of Aureochromes of the model diatom *Phaeodactylum tricornutum* revealed four distinct clades of these proteins, allowing molecular and functional studies on the individual Aureochromes in this alga.

Isoform-specific antisera were validated successfully for individual quantification of PtAUREO protein levels. Using RT-qPCR, a mainly light-independent circadian rhythm was detected for *PtAureo1a* and *PtAureo1c*, a mainly light-dependent circadian rhythm for *PtAureo1b*, whereas *PtAureo2* seemed not to be regulated in a circadian manner.

For functional studies, a reproducible protocol for a highly efficient TALEN-based genome editing approach was established in *P. tricornutum*, yielding genetically homogenous colonies, forgoing the need of re-plating before the screening process. Knockout strains for each PtAUREO isoform have been generated. A lack of PtAUREO1a resulted in an increased xanthophyll cycle pool size and a reduced Chlorophyll *a* content per cell, similar to a previously characterized RNAi knockdown strain. Furthermore, a reduction in cell size was discovered. Interestingly, non-photochemical quenching (NPQ) was strongly reduced in the knockout strain, whereas it had been increased in the knockdown strains. Mono-allelic knockout strains, which eventually became overexpression strains by upregulation of the functional allele over time, showed an increased NPQ capacity, indicating that this contrasting phenotype may not be due to off-targets effects. Western Blots showed a significant reduction in protein level of the photoprotective Lhcx1 in the knockout strains. A shift of red-light adapted wild type or PtAUREO1a knockout strain to blue light analyzed by RNA-seq showed that short-term adaptation to blue light was almost completely blocked: While over 70% of transcripts were significantly up- or downregulated in the wild type after 10 min exposure to blue light, less than 3% were found to be regulated in the knockout strain. Furthermore, around 25% of the transcriptome was differentially regulated in the red-light condition, providing further evidence for a blue-light independent function of PtAUREO1a. While all other photoreceptors and transcription factors were found to be expressed in the knockout strain, their regulation pattern upon blue-light exposure was found to be disturbed. Additionally, transcriptional induction of the photoprotective Lhcx proteins was found to be repressed. In conclusion, PtAUREO1a seems to be a master switch of the short-term light acclimation and photoprotection pathway and its expression needs to be tightly regulated.

Zusammenfassung

Aureochrome sind eine Gruppe Blaulicht-abhängiger Transkriptionsfaktoren, die nur in Stramenopilen vorkommen. Phylogenetische Analysen der Aureochrome der Modell-Diatomee *Phaeodactylum tricornutum* zeigten vier distinkte Gruppen dieser Proteine, was molekulare und funktionelle Studien der einzelnen Aureochrome in dieser Alge ermöglicht.

Isoform-spezifische Antiseren wurden erfolgreich für die individuelle Quantifizierung von PtAUREO-Proteingehalten validiert. Mittels RT-qPCR konnte ein lichtunabhängiger circadianer Rhythmus für *PtAureo1a* und *PtAureo1c*, sowie ein lichtabhängiger circadianer Rhythmus für *PtAureo1b* gezeigt werden, während *PtAureo2* konstitutiv exprimiert wurde.

Für funktionale Studien wurde ein effizientes TALEN-basiertes Genom-Editierungsprotokoll für *P. tricornutum* etabliert, welches die Generierung genetisch homogener Klone ohne erneutes Ausplattieren vor dem Screening-Prozess ermöglicht. Knockout-Stämme für jede PtAUREO Isoform wurden generiert. Das Fehlen von PtAUREO1a resultierte, ähnlich wie in zuvor charakterisierten RNAi Knockdown-Stämmen, in einem erhöhten Xanthophyll-Zyklus Pool und reduziertem Chlorophyll *a* Gehalt pro Zelle. Zusätzlich wurde eine Reduktion der Zellgröße festgestellt. Interessanterweise wiesen die Knockout-Stämme eine verringerte non-photochemical quenching (NPQ) Kapazität auf, während Knockdown-Stämme erhöhtes NPQ zeigten. Monoallelische Knockout-Stämme, welche mit der Zeit durch Hochregulierung des intakten Allels schließlich zu Überexpressions-Stämmen wurden, zeigten erhöhtes NPQ, was darauf hinweist, dass der gegensätzliche Phänotyp nicht auf off-target Effekten beruhen könnte. Western Blots zeigten einen signifikant reduzierten Gehalt des photoprotektiven Lhcx1 Proteins. Ein durch RNA-seq analysierter Shift von Rotlicht-adaptiertem Wildtyp oder einem PtAUREO1a Knockout-Stamm zu Blaulicht zeigte, dass die kurzfristige Anpassung an Blaulicht fast komplett blockiert war: Während im Wildtyp über 70% der Transkripte nach 10 min Blaulicht-Behandlung signifikant unterschiedlich reguliert waren, waren dies im Knockout-Stamm lediglich unter 3%. Weiterhin waren über 25% des Transkriptoms im Rotlicht-Zustand unterschiedlich reguliert, was einen weiteren Hinweis auf eine Blaulicht-unabhängige Funktion von PtAUREO1a darstellt. Während alle anderen Photorezeptoren und Transkriptionsfaktoren im Knockout-Stamm exprimiert wurden, war deren Regulation nach Blaulicht-Exposition gestört. Weiterhin wurde die Transkription der photoprotektiven Lhcx-Gene gehemmt. PtAUREO1a scheint daher ein Masterswitch für die kurzfristige Adaptation an Lichtbedingungen sowie Schutz vor Lichtstress zu sein, dessen Expression durch die Zelle streng reguliert werden muss.

Table of contents

Abstract ..... **I**
Zusammenfassung ..... **II**
1 General introduction **1**
1.1 Diatoms and the model organism Phaeodactylum tricornutum 1
1.2 Photoprotection in P. tricornutum 2
1.3 Photoreceptors in P. tricornutum 4
1.3.1 Current state of Aureochrome research in P. tricorntum 6
1.4 Aims 7
2 Circadian rhythm of PtAureo transcripts and evaluation of isoform-specific antisera **8**
2.1 Abstract 9
2.2 Introduction 9
2.3 Material & Methods 11
2.3.1 Cultivation of algae 11
2.3.2 Transcript analysis 11
2.3.3 Prediction of dimerization capabilities 11
2.3.4 Protein extraction from P. tricornutum 12
2.3.5 Western Blot 12
2.3.6 Immunoprecipitation 12
2.3.7 Construction of plasmids 12
2.3.8 Overexpression of PtAUREOs in E. coli 13
2.4 Results 13
2.4.1 Analysis of Aureochrome transcript levels throughout the day 13
2.4.2 Prediction of Aureochrome dimerization capabilities 14
2.4.3 Validation of isoform-specific antibodies against PtAUREOs 16
2.5 Discussion 19
2.6 Acknowledgements 20
3 Establishing Bind-n-Seq for use with Aureochromes **21**
3.1 Abstract 22
3.2 Introduction 22
3.3 Materials and methods 24
3.3.1 Bind-n-Seq 24
3.3.2 Identification of genes putatively regulated by PtAUREO1a or PtAUREO1c 25
3.4 Results 25
3.4.1 Generation of DNA binding motifs for PtAUREO1a/1c via Bind-n-Seq 25
3.4.2 Identification of genes containing the DNA binding motifs in their promoters 27
3.5 Discussion 30
3.6 Acknowledgements 31
4 Generation of TALEN-mediated gene knockouts in the diatom P. tricornutum **32**
4.1 Abstract 33

4.2 Introduction ... 33
4.3 Material & Methods ... 35
4.3.1 Assembly of the TALEN plasmids ... 35
4.3.2 Cultivation of algae ... 37
4.3.3 Nuclear transformation of P. tricornutum ... 37
4.3.4 DNA isolation ... 37
4.3.5 Allele-specific PCR ... 38
4.3.6 Southern Blotting ... 38
4.3.7 Protein Isolation and Immunoblotting ... 39
4.3.8 Chlorophyll a determination ... 39
4.3.9 Pigment extraction ... 40
4.3.10 Measurement of non-photochemical quenching (NPQ) ... 40
4.3.11 Determination of cell size by microscopy ... 40
4.4 Results & discussion ... 40
4.4.1 Generation of the TALEN constructs ... 40
4.4.2 Screening of the obtained transformants and statistical evaluation ... 44
4.4.3 Homogeneity of the generated mutants ... 47
4.4.4 Phenotypic characterization of PtAUREO1a knockout mutants ... 49
4.5 Conclusion ... 52
4.6 Acknowledgements ... 52
5 Loss of PtAUREO1a results in blocked short term adaptation to blue light ... 53
5.1 Abstract ... 54
5.2 Introduction ... 54
5.3 Material & Methods ... 56
5.3.1 Cultivation and harvesting of cells for RNA-seq ... 56
5.3.2 Sample preparation for RNA-seq ... 56
5.3.3 Bioinformatical analysis of the sequencing data ... 56
5.3.4 Protein Isolation and Immunoblotting ... 57
5.4 Results & discussion ... 57
5.4.1 Lhcx1 levels are strongly reduced in PtAUREO1a knockout strains ... 57
5.4.2 Transcriptional characterization of a red to blue light shift ... 58
5.5 Discussion ... 66
5.6 Acknowledgements ... 69
6 Generation of knockout strains for PtAUREO1b, PtAUREO1c and PtAUREO2 ... 70
6.1 Abstract ... 71
6.2 Introduction ... 71
6.3 Material & Methods ... 72
6.3.1 Generation of TALEN plasmids ... 72
6.3.2 Cultivation and transformation of P. tricornutum ... 72
6.3.3 Allele-specific PCR ... 73
6.3.4 Western Blot ... 73
6.3.5 Light microscopy ... 73
6.3.6 Scanning electron microscopy ... 73
6.4 Results ... 74

6.4.1 Generation of PtAUREO1c knockout strains ... 74
6.4.2 Generation of PtAUREO2 knockout strains ... 76
6.4.3 Generation of PtAUREO1b knockout strains ... 77
6.4.4 Preliminary characterization of PtAUREO1b knockout strain 5 ... 78
6.5 Discussion ... 83
6.6 Acknowledgements ... 84
7 Knockout of Lhcx1 leads to loss of NPQ under low light conditions ... 85
7.1 Abstract ... 86
7.2 Introduction ... 86
7.3 Material & Methods ... 88
7.3.1 Generation of TALEN constructs ... 88
7.3.2 Cultivation of Algae ... 88
7.3.3 Nuclear transformation of P. tricornutum ... 88
7.3.4 Protein Isolation and Immunoblotting ... 89
7.3.5 Southern Blot ... 89
7.3.6 Measurement of non-photochemical quenching (NPQ) ... 89
7.4 Results & Discussion ... 89
7.5 Conclusion ... 93
7.6 Acknowledgements ... 94
8 General discussion ... 95
8.1 PtAUREO1a seems to act as a master switch for light regulation ... 95
8.1.1 The expression level of PtAUREO1a needs to be tightly regulated ... 100
8.2 What is the role of the other PtAUREOs? ... 101
8.3 TALEN and its potential applications ... 103
8.4 Conclusions and outlook ... 105
A. Supplementary data ... 107
B. Author contributions ... 114
C. Acknowledgements ... 116
D. List of publications ... 118
E. Bibliography ... 119

1 General introduction

1.1 Diatoms and the model organism *Phaeodactylum tricornutum*

Diatoms, which belong to the Stramenopiles, are a group of unicellular microalgae that play an important role in the nitrogen, phosphorous and silica cycle and are thought to account for up to 20% of global carbon fixation (Nelson *et al.*, 1995, Yool and Tyrrell, 2003, Armbrust, 2009). Diatoms are estimated to comprise well over 100000 different species, making it one of the most diverse groups of eukaryotic microorganisms (Mann and Droop, 1996), and they are widespread in most aquatic habitats. Diatoms arose from a process termed secondary endocytobiosis, i.e. the engulfment of a eukaryotic red alga by a heterotrophic host cell (Delwiche, 1999). Their plastids are surrounded by four membranes, of which the outermost is continuous with the endoplasmatic reticulum (cER) (Kroth and Strotmann, 1999, Keeling, 2013), which is interpreted as evidence for secondary endocytobiosis. One of the most characteristic features of diatoms is their silica cell wall (frustule) composed of two halves (valves) arranged like a petri dish. Diatoms can be split into two groups according to the shape of their silica frustules: the radially symmetrical centrics (e.g. *Thalassiosira pseudonana*) and the bilaterally symmetrical pennates (e.g. *Phaeodactylum tricornutum*) (Round *et al.*, 1990). As the frustules are very stable and chemically inert, possess a very high surface-to-volume ratio, can be produced/harvested cheaply and vary in their shape and properties between species, their use as a solid surface support for catalysts is intensely investigated (Hamm *et al.*, 2003, Gordon *et al.*, 2009, Dolatabadi and de la Guardia, 2011, Kröger and Brunner, 2014). As many diatoms contain larger amounts of lipids (up to 50% of their dry weight), which even can be increased by genetic manipulation (Trentacoste *et al.*, 2013, Levitan *et al.*, 2015), they are also suitable for the production of biodiesel and/or bioplastics (Chisti, 2007, Mata *et al.*, 2010, Roesle *et al.*, 2014).

The pennate diatom *Phaeodactylum tricornutum* (see Figure 1-1 for a light-microscopic image) is a common model organism for diatoms due to the availability of its genome sequence as well as different transformation techniques (Apt *et al.*, 1996, Bowler *et al.*, 2008, Niu *et al.*, 2012, Miyahara *et al.*, 2013, Zhang and Hu, 2014, Karas *et al.*, 2015). Thus, gene localization studies as well as reverse genetics approaches can be applied to gain insight into

gene functions of *P. tricornutum*. Until now, RNAi-based silencing as well as protein overexpression were the main techniques used to investigate the function of proteins (Coesel *et al.*, 2009, De Riso *et al.*, 2009). Limitations of the RNAi system include presence of residual target protein, i.e. only knockdown instead of knockout, as well as potential instability of the knockdown (Lavaud *et al.*, 2012). However, two systems allowing targeted genome editing, TALEN (Transcription activator-like effector nuclease) and CRISPR/Cas9 (Clustered regularly interspaced short palindromic repeats), have been successfully applied to *P. tricornutum* recently, allowing creation of stable knockout strains, thereby greatly improving the molecular toolbox available (Daboussi *et al.*, 2014, Weyman *et al.*, 2015, Nymark *et al.*, 2016).

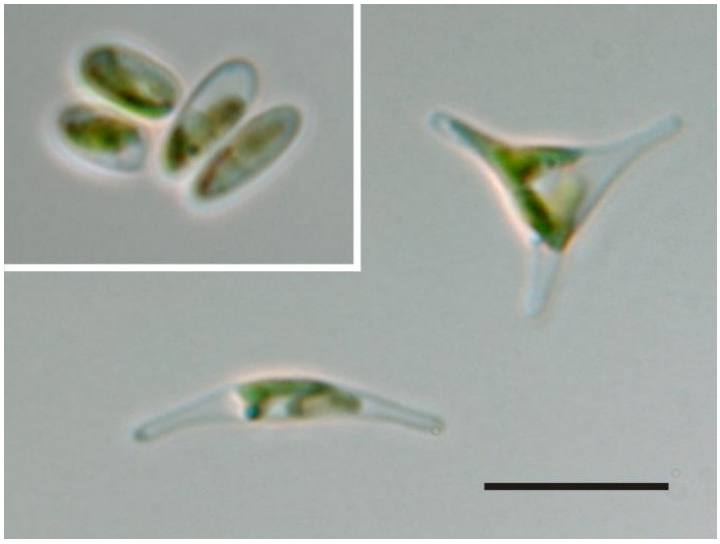


Figure 1-1: Light-microscopic images of the diatom *P. tricornutum* showing its three morphotypes: oval (upper left), fusiform (lower half) and triradiate (upper right). The strain used in this thesis, UTEX646 (UTEX, Austin, USA), is predominantly fusiform. The black bar corresponds to 10 µm. Picture reproduced with permission of Ansgar Gruber (Universität Konstanz, Germany).

1.2 Photoprotection in *P. tricornutum*

Diatoms need to cope with large variations of light quality and quantity (MacIntyre *et al.*, 2000, Ragni and D'Alcalà, 2004). Thus, complex regulatory networks are needed to ensure a sufficient photosynthetic capability under low light conditions and to avoid irreversible photodamage under high light conditions. Absorption of light by Chlorophyll (Chl) molecules in the antenna of the photosystems is required to fuel the photosynthetic machinery. However, the excited Chl molecule can also return to the ground state via triplet Chl *a* producing singlet oxygen, a highly toxic reactive oxygen species (Triantaphylidès and Havaux, 2009). Light absorption by Chl is several orders of magnitude faster than the rate-limiting step of

photosynthesis, the carboxylation reaction by RuBisCo (Wilhelm and Selmar, 2011). Thus, excess light energy, which cannot be used for photochemistry, needs to be dissipated via Chl fluorescence or as heat (termed non-photochemical quenching, NPQ) to avoid photodamage (Müller *et al.*, 2001) (see Figure 1-2).

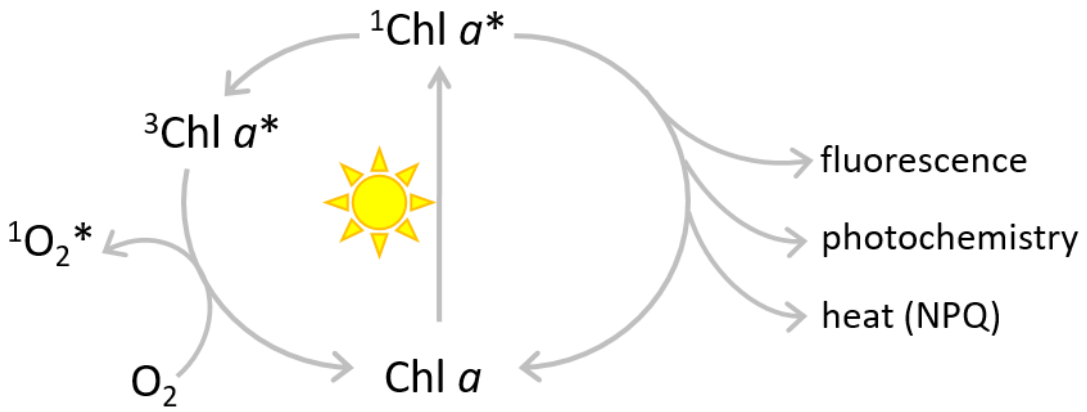


Figure 1-2: Possible fates of excited chlorophyll (Chl). Absorption of light by Chl *a* converts the molecule from the ground state to the excited ^{1}Chl *a** state. To get back to the relaxed state, the energy is either dissipated as fluorescence, used for photosynthesis (photochemistry), or dissipated as heat (called non-photochemical quenching, NPQ). Excited ^{1}Chl *a** can also lead to generation of $^1O_2{}^*$, a highly toxic reactive oxygen species (ROS), through the intermediate ^{3}Chl *a** state. Adapted from: (Müller *et al.*, 2001).

NPQ has been shown to be the major short-term adaptation for photoprotection in diatoms (Lavaud and Lepetit, 2013). In intermittent light conditions, *P. tricornutum* was found to have an up to 5-fold higher NPQ capacity than the model plant *Arabidopsis thaliana* (Ruban *et al.*, 2004). This high NPQ capacity compared to other photosynthetic organisms is thought to contribute to their ecological success. There are three prerequisites, which are required for NPQ formation in diatoms: The presence of a transthylakoidal pH gradient, the enzymatic conversion of the xanthophyll cycle pigment diadinoxanthin to diatoxanthin, which is induced by the transthylakoidal pH gradient, and the presence of Lhcx proteins (Bailleul *et al.*, 2010, Goss and Jakob, 2010, Goss and Lepetit, 2015). Two different modes of regulation of high light acclimation and photoprotection are known: direct light signaling via photoreceptors or indirect light signaling (termed "retrograde signaling"). One of these potential indirect triggers is the redox state of the plastoquinone pool (PQ) (Pfannschmidt, 2003), which was only recently identified to be used for regulation of nuclear-encoded genes involved in photoprotection of *P. tricornutum* (Lepetit *et al.*, 2013). Furthermore, blue light was found to be essential for high-light acclimation and photoprotection, indicating the involvement of blue-light regulated photoreceptors as well (Schellenberger Costa *et al.*, 2013a).

1.3 Photoreceptors in *P. tricornutum*

Three classes of photoreceptors are known in diatoms: The red-light regulated Phytochromes, as well as the blue-light regulated Cryptochromes and Aureochromes (Depauw *et al.*, 2012). The blue-light regulated Phototropins of the green lineage are not found in diatoms. In contrast to the land plant model organism *Arabidopsis thaliana*, whose genome encodes five Phytochromes and two Cryptochromes and two Phototropins (Sullivan and Deng, 2003), the genome of *P. tricornutum* encodes only a single Phytochrome but four Cryptochromes and four Aureochromes. This strong overrepresentation of blue-light photoreceptors is most likely an adaptation to the underwater light field, as blue light penetrates the water column much deeper than red light (Depauw *et al.*, 2012). While Phytochromes and Cryptochromes are widespread throughout the kingdoms of life (Lin and Todo, 2005, Ulijasz and Vierstra, 2011), Aureochromes were first discovered in the Xanthophyte *Vaucheria frigida* in 2007, where they control morphogenesis and maturation of sex organs (Takahashi *et al.*, 2007). So far, they have only been found in Stramenopiles (Ishikawa *et al.*, 2009, Schellenberger Costa *et al.*, 2013b). Photoreceptors require a chromophore binding domain, which undergoes a conformational change upon illumination with the corresponding wavelength, as well as a signaling domain which becomes activated or inhibited due to the light-induced conformational change (see exemplary domain arrangements of Phytochromes, Phototropins, Cryptochromes and Aureochromes shown in Figure 1-3). In the case of Phytochromes and Phototropins, this signaling domain is commonly a kinase domain (Christie *et al.*, 1998, Yeh and Lagarias, 1998, Briggs *et al.*, 2001, Li *et al.*, 2011a). In Cryptochromes, the exact mechanism of action of the C-terminal extension domain, which is thought to be the signaling domain, is not yet clear. However, strong conformational changes and phosphorylation upon blue-light illumination have been observed (Shalitin *et al.*, 2002, Shalitin *et al.*, 2003, Yu *et al.*, 2007). Thus, these photoreceptors likely require interaction with other proteins for their signaling cascade. Aureochromes possess an N-terminal bZIP (basic region leucine zipper) domain typically found in transcription factors, whereas a single C-terminal LOV (light-oxygen-voltage) domain acts as the photosensory domain via a FMN (Flavin mononucleotide) cofactor (Takahashi *et al.*, 2007). Thus, they are the only class of photoreceptors found in photosynthetic organisms which can directly influence transcription of other proteins.

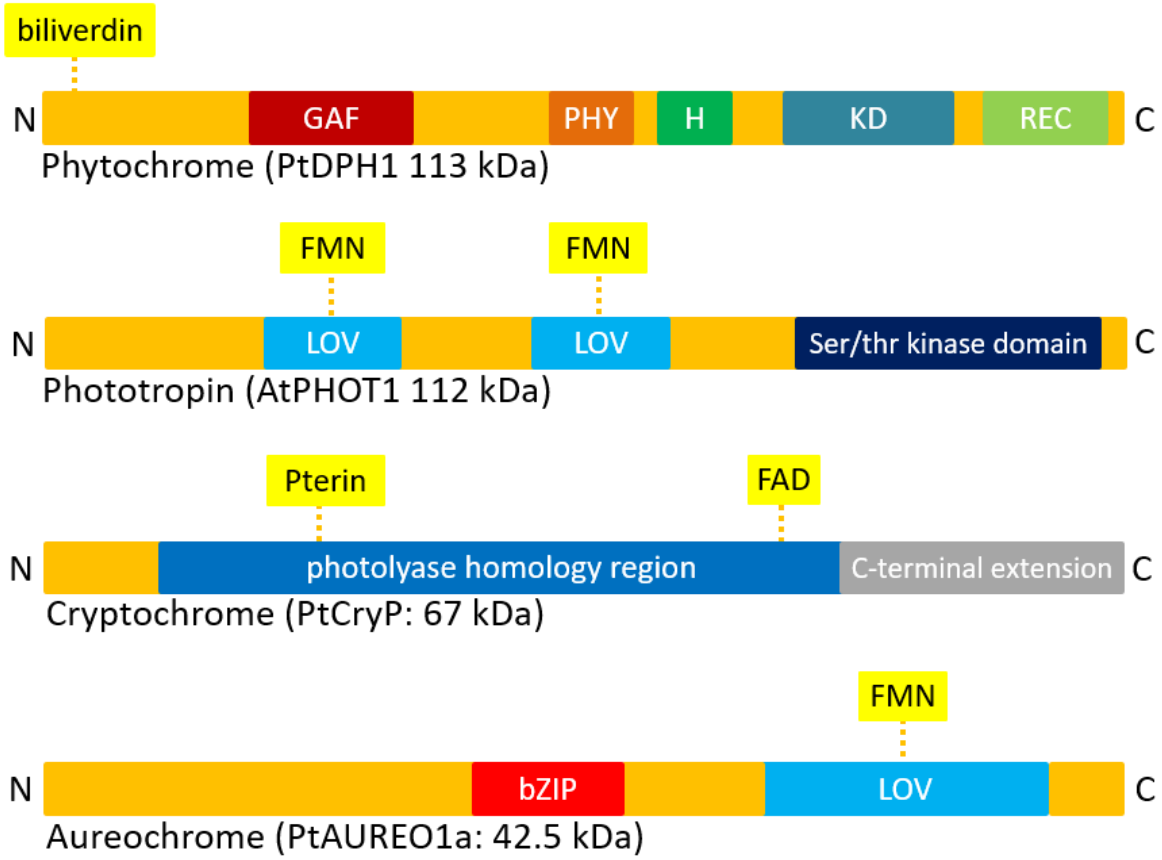


Figure 1-3: Schematic overview of the domain structure of Phytochromes, Phototropins, Cryptochromes and Aureochromes. The molecular weight of a member of each group found in *P. tricornutum* as well as a Phytochrome isoform of *Arabidopsis thaliana* is given. Conserved domains are indicated as colored boxes and the respective chromophores are indicated by yellow boxes with a dashed orange line representing the chromophore-binding amino acid. Notably, the PtDPH1 does not contain an N-terminal Per-Arnt-Sim (PAS) domain (Fortunato *et al.*, 2016). GAF: cyclic di-GMP phosphodiesterase adenyl cyclase / Fhla domain; PHY: Phytochrome domain; H: Histidine Kinase A domain; KD: Histidine kinase-like ATPase c domain; REC: Response receiver domain; LOV: light-oxygen-voltage domain; bZIP: basic region leucine zipper.

Interestingly, the domain arrangement is opposite of what is found in Phototropins (see Figure 1-3) and most other LOV-domain proteins, where an N-terminal LOV domain and a C-terminal effector domain is present (Crosson *et al.*, 2003). Hence, Aureochromes are of interest for optogenetics (Mitra *et al.*, 2012). Illumination with blue light was shown to lead to unfolding of the Jα-helix of the LOV domain, allowing dimerization (Herman *et al.*, 2013). As a bZIP dimer is required to bind to DNA (O'Shea *et al.*, 1989, Vinson *et al.*, 1989, O'Shea *et al.*, 1991), formation of either homo- or heterodimers is required for Aureochrome function.

1.3.1 Current state of Aureochrome research in *P. tricorntum*

Phylogenetic analysis of all available Aureochrome sequences revealed four distinct clades of Aureochromes, of which *P. tricornutum* possesses one isoform each: PtAUREO1a (JGI Protein ID 49116), PtAUREO1b (49458), PtAUREO1c (49742) and PtAUREO2 (56060) (Schellenberger Costa *et al.*, 2013b). Thus, *P. tricornutum* is an ideal model to study Aureochrome functions. A summary of their known biochemical/biophysical properties and functions is given in Table 1-1. Similarly to what has been observed in *Vaucheria frigida* for VfAUREO2, PtAUREO2 seems to be unable to bind the FMN cofactor, making its role and regulation unclear (Takahashi *et al.*, 2007, Banerjee *et al.*, 2016a). All isoforms possess a nuclear localization signal (NLS) and nuclear localization was confirmed for three out of four isoforms via expression of GFP fusion proteins (Schellenberger Costa *et al.*, 2013a). So far, the isoform PtAUREO1a has been the main focus of research: Using Yeast 1 Hybrid screens it could be demonstrated that PtAUREO1a interacts with the promoter of the diatom-specific cyclin2 (dsCYC2), which plays a major role in the light-dependent G1 checkpoint of the cell cycle (Huysman *et al.*, 2013). Additionally, RNAi-based knockdown lines of PtAUREO1a, which retained about 50% of wild type protein levels, showed a 'hyper high-light' acclimation phenotype under low light conditions manifested by decreased Chl *a* concentration per cell, increased non-photochemical quenching (NPQ) and increased xanthophyll cycle pigment pool, indicating that PtAUREO1a might be a repressor of light acclimation (Schellenberger Costa *et al.*, 2013b). Therefore, the four Aureochromes of *P. tricornutum* seem to perform at least partially non-overlapping functions, and thus can be categorized independently by reverse genetics approaches.

Table 1-1: Known biophysical/biochemical properties as well as known cellular localization and functions within the cell. Experimental localization via GFP-fusion proteins are indicated in bold, otherwise localization is inferred from predictions. Sources for the respective properties are indicated by uppercase numbers: [1]: (Schellenberger Costa *et al.*, 2013b); [2]: (Herman *et al.*, 2013); [3]: (Banerjee *et al.*, 2016a); [4]: (Huysman *et al.*, 2013). n.d.: not determined.

	Protein ID (JGI)	Size [kDa]	Isoelectric point (pI)	Cellular localization	FMN binding	Known physiological role	Known targets
PtAUREO1a	49116	41.5	5.26	**nucleus**[1]	yes[2]	photoacclimation[1]	dsCYC2[4]
PtAUREO1b	49458	46.7	5.83	**nucleus**[1]	n.d.	-	-
PtAUREO1c	56742	35.8	5.04	nucleus[1]	n.d.	-	-
PtAUREO2	56060	48.9	4.92	**nucleus**[1]	no[3]	-	-

1.4 Aims

The aim of this thesis was to further characterize the function of Aureochromes in *P. tricornutum* using several different approaches. Transcript pattern of the Aureochrome isoforms throughout the day were generated by qPCR and isoform-specific antisera were validated for their potential applications (Chapter 2). An *in vitro* based next generation sequencing approach (Bind-n-Seq) was employed to identify the DNA binding motif using heterologously overexpressed PtAUREO isoforms (Chapter 3). A TALEN-based reverse genetics approach was established to generate knockout mutants, thereby overcoming the limitations of RNAi (Chapter 4). The phenotype of the generated PtAUREO1a knockout mutants was further characterized, as well as genes directly influenced by the respective Aureochrome identified using a transcriptomics approach (Chapter 5). Additional TALEN knockout mutants were generated for all other isoforms for future research (Chapter 6). Lhcx1, a protein known to be involved in the photoprotection mechanism NPQ and seemingly regulated by PtAUREO1a, was knocked out in *P. tricornutum*, followed by a preliminary characterization of the resulting phenotype (Chapter 7).

2 Circadian rhythm of *PtAureo* transcripts and evaluation of isoform-specific antisera

Serif, M.* & Kroth, P.G.

*Corresponding Author: manuel.serif@uni-konstanz.de

Plant Ecophysiology, Fachbereich Biologie, Universität Konstanz, D-78457 Konstanz, Germany

Keywords: *Phaeodactylum tricornutum*, Aureochromes, blue light-dependent transcription factor

Parts of this chapter were used in the following publication:

Banerjee, A., Herman, E., Serif, M., Maestre-Reyna, M., Hepp, S., Pokorny, R., Kroth, P.G., Essen, L.O. and Kottke, T. (2016b) Allosteric communication between DNA-binding and light-responsive domains of diatom class I aureochromes. *Nucleic Acids Res.*, **44**, 5957-5970.

2.1 Abstract

The genome of the diatom *P. tricornutum* encodes for 4 isoforms of Aureochromes, a novel class of blue-light dependent transcription factors, which seemingly perform partially non-overlapping functions. To gain insight into the regulation of their expression levels a qPCR analysis was performed using a previously generated cDNA library of cells grown either in a day night rhythm or complete darkness. The Aureochromes seem to be differentially regulated throughout the day: While *PtAureo2* does not appear to be regulated at all throughout the day, *PtAureo1b* was found to be primarily regulated by light, whereas *PtAureo1a* and *PtAureo1c* showed a primarily light-independent circadian regulation. Additionally, bioinformatics prediction of dimerization capabilities of Aureochromes based on the bZIP domain revealed that PtAUREOs might not only form homodimers but also heterodimers. PtAUREO1a was found to be the most promiscuous, potentially forming both homodimers and heterodimers with all other isoforms. Furthermore, to allow individual characterization of all isoforms, four isoform-specific antisera were validated. A protocol allowing lysis of multiple samples in parallel in a bead mill was established, which was found to be more reliable and less time consuming for screening purposes than previously established lysis methods via ultrasonication or French press. Validation of the antisera by Western Blot using both crude extract of *P. tricornutum* as well as heterologously expressed PtAUREOs showed very little cross-reactivity between the isoforms when *E. coli* protein extracts were used, and little background when used with *P. tricornutum* crude extract. Thus, they can be used to reliably quantify expression levels of the individual isoforms. When Co-Immunoprecipitation was attempted, however, very little protein could be pulled down, indicating low binding affinity for the protein in its native state. Additionally, it could be confirmed that a single amino acid exchange in PtAUREO2 (M301V) restores FMN binding capability lost in this isoform.

2.2 Introduction

Aureochromes are a novel class of light-dependent transcription factors, which were first identified in 2007 in the Xanthophyte *Vaucheria frigida*, where the two isoforms, Aureo1a and Aureo2, are involved in the induction of branching and development of sexual organs, respectively (Takahashi *et al.*, 2007). Since then, Aureochromes have been identified in many other Stramenopiles and seem to be restricted to this specific group of algae (Ishikawa *et al.*,

2009). The genome of the model diatom *P. tricornutum* encodes four different Aureochromes, which were originally classified due to their homology to the two Aureochromes of *V. frigida*: PtAUREO1a (JGI Protein ID: 49116), PtAUREO1b (49458), PtAUREO1c (56742) and PtAUREO2 (56060). However, using additional sequences of other organisms to construct a phylogenetic tree, a more complex picture evolved with four distinct clades of Aureochromes, of which *P. tricornutum* possesses one of each (Schellenberger Costa *et al.*, 2013b). This finding is an indication that the four Aureochromes might have different functions in the cell. As bZIP transcription factors require dimerization to be able to bind to its target sequence (O'Shea *et al.*, 1989, Vinson *et al.*, 1989, O'Shea *et al.*, 1991), the different isoforms might form homo- or heterodimers with each other when the isoforms are co-expressed. So far, however, very little is known about their expression patterns: Shifting cells that were acclimated to continuous low intensity white light to higher light intensities caused down-regulation of *PtAureo1a* and *PtAureo2* transcripts, whereas exposing cells to 48 h of darkness lead to an increase of these transcripts (Nymark *et al.*, 2009, Nymark *et al.*, 2013). Re-exposure to white light caused transcript levels to decrease again, which was also observed upon re-exposure to blue light, but not upon re-exposure to green or red light (Nymark *et al.*, 2013, Valle *et al.*, 2014). Thus, a certain co-regulation of *PtAureo1a* and *PtAureo2* transcripts upon the different treatments was observed in previous studies, however, no data is available for fluctuation of transcript levels under physiological conditions, i.e. throughout a day-night cycle. Thus, no clear conclusion can be drawn whether the Aureochrome isoforms are co-expressed and are therefore able to interact with each other.

One of these isoforms, PtAUREO2, was shown to be unable to bind the light-sensitive cofactor flavin mononucleotide (FMN), and thus does not seem to be regulated by light (Banerjee *et al.*, 2016a). Previous studies on VfAUREO2 of *V. frigida*, which also lacks FMN, could show that silencing leads to a phenotype (Takahashi *et al.*, 2007), thus it is likely to still possess a function and might be regulated in a different manner than the other Aureochromes. Based on modelling of the PtAUREO2 LOV domain using the crystal structure of PtAUREO1a, a single amino acid exchange (M301V) was suggested to be sufficient to restore Flavin binding of PtAUREO2 (Banerjee *et al.*, 2016a). To study the function of these proteins in more detail, highly specific antisera for each isoform would be a valuable tool, both for screening knockdown/knockout mutants, estimating expression levels, as well as identifying potential interaction partners by Co-Immunoprecipitation or Chromatin Immunoprecipitation. Hence, antisera against all four isoforms were raised and validated for

their potential applications. Additionally, qPCR was performed to investigate their expression patterns throughout the day, and their capability to form homo- or heterodimers was analysed by a bioinformatic approach based on their bZIP domains.

2.3 Material & Methods

2.3.1 Cultivation of algae

The *P. tricornutum* (Bohlin) strain UTEX646 was obtained from the culture collection of algae of the University of Texas (UTEX, Austin, USA). *P. tricornutum* was grown axenically in liquid F/2 medium without added silica and 16.5 ‰ salt content or on solid f/2 media which contained additionally 1.2% (w/v) Bacto Agar (BD, Sparks, MD, USA). Cells in liquid f/2 medium were cultivated in a 16h/8h light/dark cycle in Erlenmeyer flasks under continuous shaking at 20°C and an illumination of 35 µmol photons m^{-2} s^{-1} (Osram Lumilux L58W/840, Munich, Germany).

2.3.2 Transcript analysis

The expression patterns of four Aureochrome isoforms in *P. tricornutum* were investigated to screen for their light-dependent differential expression using cDNA generated earlier (Lepetit *et al.*, 2013). In short, cells had been pre-adapted to 16 h of daily illumination with low light (LL) and had been either kept under the same condition or transferred to continuous darkness for one illumination period. Transcript levels were assayed every 3 h for 33 h under LL or darkness using these pre-adapted cells. HGPRT (hypoxanthine-guanine phosphoribosyltransferase; JGI Protein ID 35566), RPS (ribosomal protein S1; 45451) and TBP (TATA-binding protein; 10199) were chosen as housekeeping genes based on previous studies {Sachse, 2014 #222}. Statistical significant changes were calculated using the software REST 2009 (Pfaffl *et al.*, 2002).

2.3.3 Prediction of dimerization capabilities

The prediction of homo- and heterodimerization capabilities of the PtAUREOs is based on coiled-coil interaction of the leucine zipper domain (termed salt bridge rules). Positive and/or negative interactions were visualized with DrawCoil (Grigoryan and Keating, 2008).

2.3.4 Protein extraction from *P. tricornutum*

40 ml of culture were harvested in mid-exponential phase by centrifugation (5min at 5500g and 4°C). Cell pellets were stored frozen at -20°C until further use. Cells were resuspended in 500 µl of lysis buffer (4 M urea, 1.5 M thiourea, 1% SDS, 20 mM Tris HCl pH 8) and lysed using a bead mill (FastPrep Fp120, Thermo Scientific, Karlsruhe, Germany) at maximum speed for 6 times 20 seconds. Between cycles samples were cooled on ice for 1 min. Lysates were clarified by two centrifugation steps at 20000 g and 4°C for 30 min each.

2.3.5 Western Blot

Peptide-specific antibodies against PtAUREO1b, PtAUREO1c and PtAUREO2 were generated by Agrisera (Vännas, Sweden). The peptides chosen for immunization of the rabbits and their relative position in the respective protein are given in Table 2-1. SDS-PAGE was performed according to (Laemmli, 1970). Subsequent immunoblots were done as described in (Schellenberger Costa *et al.*, 2013b).

Table 2-1: Peptides chosen for generation of the antisera for PtAUREO1b, PtAUREO1c and PtAUREO2.

Target protein	Peptide sequence	Position within the protein
PtAUREO1b	(H_2N)-CNPVTKKQKTDEQSQ-$(CONH_2)$	87-100
PtAUREO1c	(H_2N)-CNGNGHDGASKKRSADDFD-$(CONH_2)$	64-81
PtAUREO2	(H_2N)-CVARPEPGDPEHDKG-$(CONH_2)$	401414

2.3.6 Immunoprecipitation

For immunoprecipitation, up to 500 ml of culture was used as starting material per sample. The antibodies raised against the respective isoforms were coupled to protein A/G magnetic beads (Thermo Fisher, Schwerte, Germany) and immunoprecipitation under non-denaturing conditions was performed according to the manufacturer's instructions.

2.3.7 Construction of plasmids

Codon-optimized sequences of the isoforms *PtAureo1a* (JGI ID 49116), *PtAureo1b* (49458), *PtAureo1c* (56742) and *PtAureo2* (56060) were synthesized and cloned into pET28a(+) (Merck Millipore, Darmstadt, Germany) by Eurofins MWG Operon (Ebersberg, Germany) using the restriction enzymes NdeI and SalI, encoding for an N-terminally His-tagged fusion protein. The coding sequence of pET28a(+)-PtAUREO2 was amplified using primers

Aureo2-Cterm-HIS_for/rev (see Table 2-2) and ligated into pET28a(+) after restriction digest with NcoI and XhoI in order to create a C-terminally His-tagged fusion construct. Afterwards, a single point mutation (corresponding to A977G of the gene in *P. tricornutum*, resulting in an amino acid exchange M301V of the gene product) was introduced by SDM as described in (Banerjee *et al.*, 2016b) using primers Aureo2_M301V_for/rev (see Table 2-2).

Table 2-2: Primers used in this study. The mutated base pair introduced using primers A2_M301V_for/rev is written in lower case.

Primer name	Application	Sequence (5'→3')
A2-HIS_for	Exchange N-terminal	ACGTCAGAACCATGGCGCAGAACCTTCAAATGCC
A2-HIS_rev	to C-terminal His tag	TTGACTGTACTCGAGACCACCGCCGGATGAACGACCCGCG
A2_M301V_for	Generation of	CGTCCCAGCACAGCTTTgTGATTACGGATCCATCTC
A2_M301V_rev	PtAUREO2_M301V	GAGATGGATCCGTAATCAcAAAGCTGTGCTGGGACG

2.3.8 Overexpression of PtAUREOs in *E. coli*

His-tagged PtAUREO isoforms were overexpressed in *E. coli* BL21(DE3) Rosetta cells using the EnPresso B growth system (BioSilta, Finland) according to the manufacturer's instructions. Protein expression was induced by addition of 10 µM IPTG. Cells were lysed by French press and the lysate was clarified by centrifugation for 1h at 21000g and 4°C. Clarified lysates were analyzed by Western Blot as described above.

2.4 Results

2.4.1 Analysis of Aureochrome transcript levels throughout the day

To gain insight into the expression of the *PtAureo* genes, a quantitative transcript analysis was performed. Expression of the four Aureochrome genes was investigated to screen for differential expression of the four isoforms using a previously generated cDNA library (Lepetit *et al.*, 2013): Cells had been pre-adapted to 16 h of daily illumination with low light (LL) and had been kept either in the same condition or transferred to continuous darkness for one illumination period. Transcript levels were assayed every 3h over a period of 33h. HPRT (Hypoxanthine-guanine phosphoribosyltransferase; JGI Protein ID 35566), RPS (Ribosomal protein S1; 45451) and TBP (TATA-binding protein; 10199) were chosen as housekeeping genes, due to their confirmed stable expression under the conditions used (Sachse *et al.*, 2014). The transcript patterns are shown in Figure 2-1, correlating colors to log2-transformed

expression levels relative to the first time point of each dataset. *PtAureo2* expression appeared to be mainly regulated in a time- and light-independent manner, whereas all other Aureos showed differential expression patterns throughout the day. *PtAureo1a* was upregulated in both conditions, indicating in a primarily light-independent circadian regulation. However, upregulation was less pronounced in the illuminated cultures, indicating a weaker light-dependent influence on regulation as well. *PtAureo1b* transcripts were much more increased in the illuminated cultures compared to the cultures grown in darkness with peak expression (5fold increase in expression compared to midnight) around 3pm. *PtAureo1c* was strongly expressed during the day in both conditions, but illumination shifted the peak of expression from around 6pm to the late morning. Thus, it can be concluded that the transcript of the four PtAureo genes do not follow the same regulation patterns, with *PtAureo2* showing little signs of light/circadian regulation, whereas *PtAureo1b* is mainly light-regulated and *PtAureo1a/c* are clearly under circadian regulation.

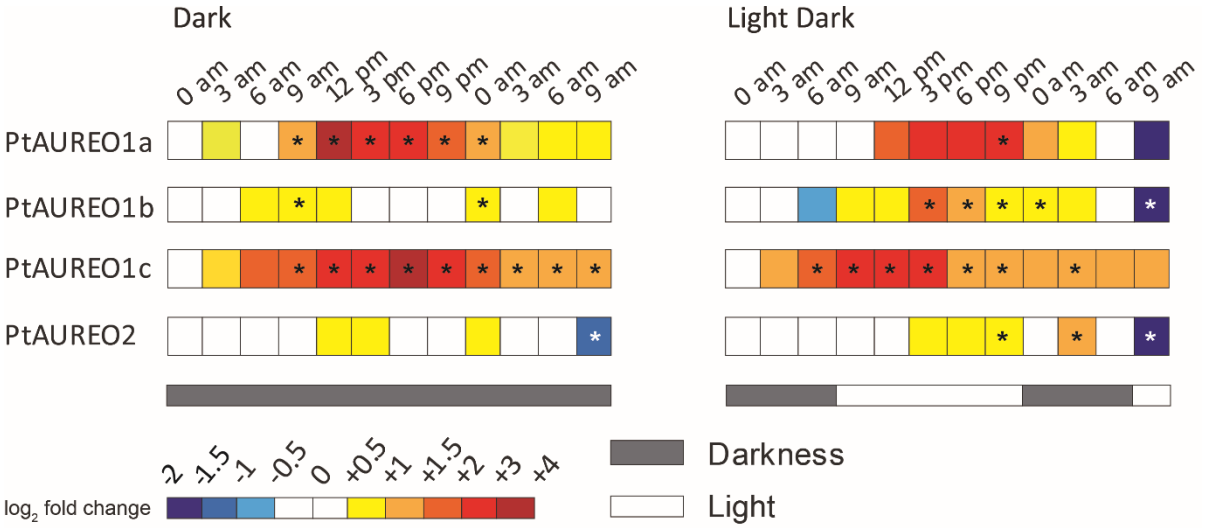


Figure 2-1: Transcript patterns of Aureochrome genes during complete darkness (left) or during a 16h light 8h dark night-day-night cycle (right), relative to the first timepoint of each dataset, depicted as log_2 fold changes. Transcripts of *PtAureo1a* and *PtAaureo1c* show a primarily light-independent circadian rhythm, whereas *PtAureo1b* shows a primarily light-dependent transcriptional control. Transcription of *PtAureo2* seems to be regulated independently of both light and time. Stars indicate significance according to the software REST 2009 (Pfaffl *et al.*, 2002).

2.4.2 Prediction of Aureochrome dimerization capabilities

The finding of differential expression patterns indicates that the PtAUREOs might have distinct functions from each other. However, co-expression of PtAUREO1a/c hints towards synergistic functions as well, potentially via heterodimerization. A bioinformatics prediction using DrawCoil (Grigoryan and Keating, 2008), which is based on the salt bridge interaction

ruleset of coiled coil interactions (Vinson *et al.*, 1993), revealed that PtAUREO1a and PtAUREO2 are possible candidates for formation of bZIP homodimers and, interestingly, PtAUREO1a could also form heterodimers with the other three isoforms (see Figure 2-2). Hence, it might have a key role among the Aureochromes, influencing the activity of the other isoforms. Moreover, formation of PtAUREO1c heterodimer with PtAUREO1b and PtAUREO2 also seems possible, however they might be less stable. As this prediction only takes the bZIP domain into account, more detailed *in vitro* or *in vivo* interaction studies are required to fully elucidate the possibility and occurrence of these homo- and heterodimers.

Figure 2-2: Prediction of potential homo- and heterodimerization capabilities of the leucine zipper domain of the Aureochromes using DrawCoil. The prediction is based on the salt bridge rule (Vinson *et al.*, 1993): Blue dashed lines show stabilizing positive interactions via salt bridges of positively and negatively charged amino acid side chains in close proximity in the coiled coil, whereas red lines indicate destabilizing negative interactions of side groups with the same charge. Models were generated using DrawCoil 1.0 (http://www.grigoryanlab.org/drawcoil/ (Grigoryan and Keating, 2008))

2.4.3 Validation of isoform-specific antibodies against PtAUREOs

For screening of PtAUREO knockdown or knockout cell lines highly specific antibodies are required which show very little cross-reactivity with the other isoforms or other unrelated proteins of *P. tricornutum*. An antiserum against PtAUREO1a had already been raised and had been successfully used for Western Blots of RNAi-induced PtAUREO1a knockdown strains (Schellenberger Costa *et al.*, 2013b). Antisera against the other three isoforms were raised as well (Agrisera, Vännas, Sweden) and needed to be validated. The amino acid sequences of the four isoforms, the location of the conserved domains as well as the peptide chosen for immunization is depicted in Figure S2-1. Previous studies on PtAUREO1a used either French press or sonication to lyse cells. However, the use of the French press is not suitable for processing many samples, e.g. derived from different cultivation conditions and/or screening of transformed cell lines, due to time-consuming sample processing and cleaning procedure in between samples. Sonication also does not allow processing of multiple samples at once, and preparations varied widely in quality (data not shown). As an alternative, lysis with 2% SDS for 30 min at room temperature, which was successfully used for extraction of the cryptochrome CPF1 (Coesel *et al.*, 2009), was also found not to be reliable for use with Aureochromes. Finally, the use of a bead mill and a strong lysis buffer containing urea, thiourea and SDS was found to be the preferred method, as there was very little variation between sample preparations and allowed lysis of multiple samples in parallel in a short time. Additionally, the use of chaotropic agents is known to destabilize DNA protein interactions, thus increasing solubility which facilitates extraction of Aureochromes. Using this protocol, each antibody produced a single band at the expected size (1a: 41.5 kDa, 1b: 46 kDa, 1c: 36 kDa, 2: 48 kDa) with seemingly little cross-reactivity among the different isoforms (see Figure 2-3).

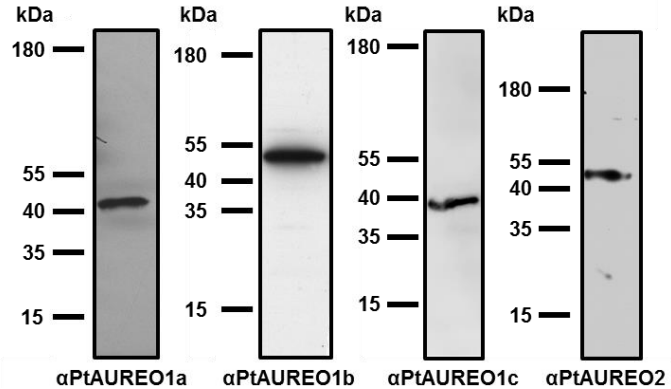


Figure 2-3: Validation of the isoform-specific antisera using crude protein extract of *P. tricornutum*. Each antisera detects a single protein corresponding to the expected molecular weight (PtAUREO1a: 41.5 kDa; PtAUREO1b: 46 kDa; PtAUREO1c: 36 kDa; PtAUREO2: 48 kDa)

To test whether the antibodies can also be used to detect their target protein in a native state Immunoprecipitation was tried under non-denaturing conditions and the elute analyzed by SDS-PAGE followed by silver staining (see Figure 2-4). A lower and upper band corresponding to the light and heavy chains of the antibody used for precipitation, as well as a third band presumably corresponding to the respective Aureochrome is visible. However, when analyzed by mass spectrometry, no Aureochrome could be identified, presumably due to a too low amount. Thus, the antibodies seem to have a strongly reduced affinity to the protein in its native state, at least under the conditions assayed.

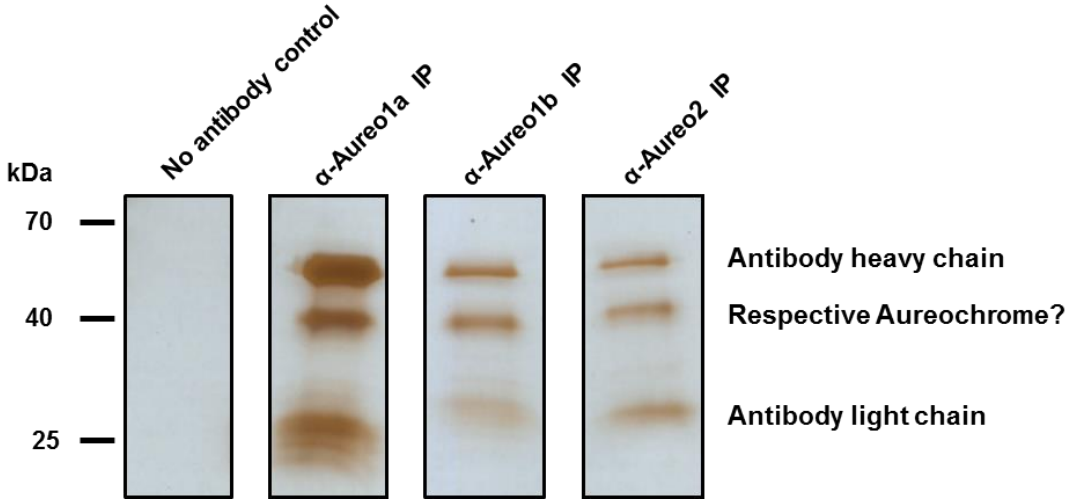


Figure 2-4: Immunoprecipitation of PtAUREO1a/1b/2 using the antibodies raised against the respective isoforms, followed by silver staining. The upper and lower band correspond to the heavy and light chain of the antibody, whereas the middle band could correspond to the respective Aureochrome. Analysis by mass spectrometry did not yield Aureochromes, presumably due to a too low amount of protein.

To confirm that the detected bands correspond to the protein of interest and are not the result of an unintended cross reaction, *E. coli* overexpression constructs were generated for all four isoforms. As FMN binding of isoform PtAUREO2 was hypothesized to be restored by a single point mutation (Banerjee *et al.*, 2016a), an additional overexpression construct, PtAUREO2_M301V, was generated by site-directed mutagenesis. Cell pellets as well as clarified cell lysates of PtAUREO1a-, PtAUREO1b- and PtAUREO1c-overexpressing cell lines were colored yellow instead of the standard brownish color, indicating high amounts of Flavin cofactors in the cell and therefore FMN-bound Aureochromes (see Figure 2-5). Wild type PtAUREO2-overexpressing cells on the other hand did not show this phenotype after induction, as expected, whereas overexpression of PtAUREO2_M301V lead to yellow coloring, confirming the hypothesis raised by Banerjee et al 2016. Thus, a single base pair change turned this photoreceptor into a non-light regulated protein. Whether it is regulated in

a different manner or performs a constitutive function should thus be studied in more detail in further studies.

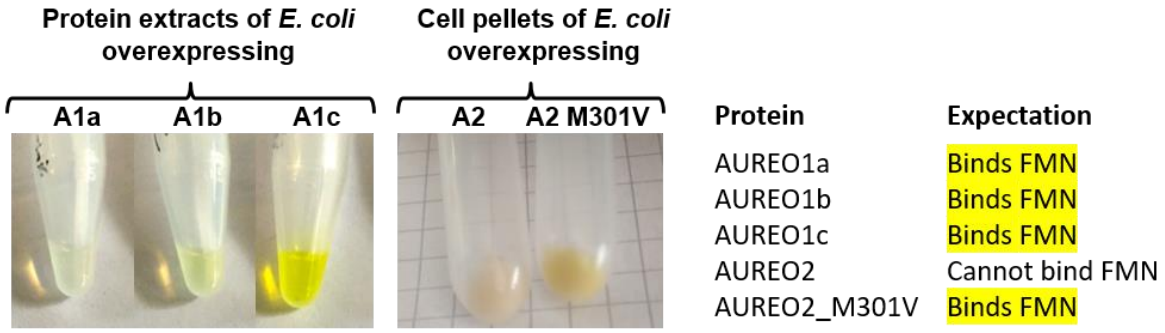


Figure 2-5: Cell pellets or crude extracts of E. coli strains overexpressing PtAUREOs. Expectation for FMN binding capability of the different isoforms is indicated. A yellow color can be seen in protein extracts of PtAUREO1a/1b/1c as well as cells overexpressing PtAUREO2 M301V, but not wild type PtAUREO2.

Analysis of *E. coli* protein extracts prior to and after induction of PtAUREO expression with IPTG via SDS-PAGE followed by Coomassie staining (see Figure 2-6) showed appearance of bands not present prior to induction. Thus, expression of the proteins was confirmed. Western Blots using the crude protein extracts and all four antibodies showed a strong band when using the corresponding protein extract and very little cross-reactivity for the other isoforms, making them a valuable tool for the characterization of PtAUREO expression levels and screening of knockdown/knockout strains.

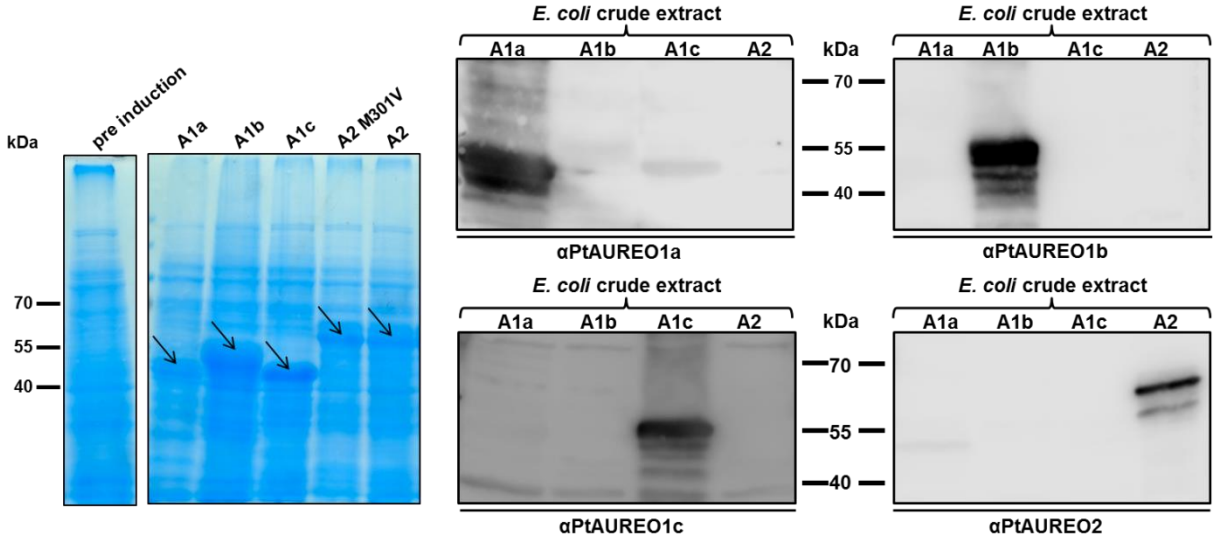


Figure 2-6: Validation of the isoform-specific antisera using *E. coli* strains expressing the different PtAUREO isoforms. A strong band (marked by an arrow) appears in each lane after induction of PtAUREO expression with IPTG (left). Each antisera labels the respective isoform with very little crossreactivity (right).

2.5 Discussion

Several different approaches have been utilized to gain more insight into the function of Aureochromes in *P. tricornutum*. Transcription of the Aureochromes seems to be differentially regulated (see Figure 2-1). While transcription of *PtAureo1b* seems to be light induced, *PtAureo1a/1c* show a circadian regulation, and *PtAureo2* seems to be not regulated by time of day and/or presence of light. Furthermore, the bioinformatic prediction of dimerization capabilities provides an indication that the PtAUREOs might form different homo- and/or heterodimers (see Figure 2-2). PtAUREO1a and PtAUREO1c might not only be co-regulated, but could potentially form heterodimers as well. Previous studies showed that there are four distinct clades of Aureochromes, and that knockdown of PtAUREO1a by about 50% results in a distinct light acclimation phenotype, indicating non-overlapping functions of the different isoforms (Schellenberger Costa *et al.*, 2013b). The findings presented here are another indication that the different isoforms perform partially non-overlapping functions, although some functions might require several isoforms in form of a heterodimer.

Additionally, a reliable and fast-throughput lysis procedure for detection of Aureochromes by Western Blot using a bead mill and chaotropic lysis buffer was established. Using this method, antisera raised against the different isoforms only detect the target protein in Western Blot without unspecific bands and little to no cross-reactivity amongst the different isoforms (see Figure 2-3 and Figure 2-6). The antibodies seem, however, primarily suitable for denatured protein under the conditions assayed. Thus, the predicted heterodimers could unfortunately not be verified with protein extracts by Co-Immunoprecipitation. This problem could potentially be overcome to a certain extent by upscaling of both amount of antibody and/or cell extract used, as well as assaying more binding conditions. Therefore, if native proteins are required, e.g. for Co-Immunoprecipitation or Chromatin Immunoprecipitation, the introduction of tagged variants (e.g. 3x-Flag or GST tag) of the Aureochromes into the genome of *P. tricornutum* should be considered. However, wild type protein should be knocked out and expression levels need to be similar to the wild type protein in order to avoid artifacts (Kolodziej *et al.*, 2009). Therefore, these strains should be generated by use of TALEN and/or CRISPR, which have recently been established for *P. tricornutum* (Daboussi *et al.*, 2014, Weyman *et al.*, 2015, Nymark *et al.*, 2016, Serif *et al.*, 2017).

Lastly, FMN-binding was restored in PtAUREO2 by a M301V point mutation (see Figure 2-5). As the transcript level of *PtAureo2* was the only isoform to not be strongly regulated in a

circadian rhythm, the question remains whether it is in a constitutively active or inactive state or if it is regulated by other factors. Hence, studying the function of this non-light regulated Aureochrome and how it is regulated is of interest for future work.

2.6 Acknowledgements

The authors are grateful for Sabine Sturm and Ansgar Gruber for generation of cDNA for the transcript analysis prior to this study, Vincent Spegg for help with the transcript analysis, and Marc Halder and Zeno Riester for help with overexpression of PtAUREOs in *E. coli*. This work was supported by the University of Konstanz, the graduate school Biological Sciences (GBS), and the DFG (grant KR 1661/8-2 to PGK).

3 Establishing Bind-n-Seq for use with Aureochromes

Serif, M.*[+], Jacobs, M.J.[+], & Kroth, P.G.

*Corresponding Author: manuel.serif@uni-konstanz.de

[+]Both authors contributed equally

Plant Ecophysiology, Fachbereich Biologie, Universität Konstanz, D-78457 Konstanz, Germany

Keywords: *Phaeodactylum tricornutum*, Aureochromes, blue light-dependent transcription factor, Bind-n-Seq

Parts of this chapter were used in the following publication:

Jacobs, M.J. (2015) Molecular characterization of Aureochromes in the diatom *Phaeodactylum tricornutum*. Universität Konstanz (Master's Thesis)

3.1 Abstract

Aureochromes are a class of light-dependent transcription factors recently identified in Stramenopiles. The genome of the model diatom *P. tricornutum* encodes four different isoforms. Except for the isoform PtAUREO1a, not much is known yet about their function, and no experimental evidence for their DNA binding motifs exists. To identify the DNA binding motif of two of the four isoforms, PtAUREO1a and PtAUREO1c, an *in vitro* approach termed Bind-n-Seq was applied. While the 8 bp binding motifs obtained contain the ACGT core motif typically found in bZIP transcription factors and feature similarities to the known binding motif of VfAUREO1, the enrichment of these sequences was found to be too low to be statistically significant. Nonetheless, a bioinformatical approach was employed to identify promoter regions containing these motifs. The motif for PtAUREO1a was found 1321 times and the motif of PtAUREO1c 321 times, a higher amount then expected for a single transcription factor. Only the motif for PtAUREO1a was found to be significantly overrepresented in the promoter sequences compared to randomized sequences of the same length and GC content. As eukaryotic transcription factor binding motifs seem to have, on average, a length of 9.9 bp, it is possible that only a partial motif was identified, and improvement of the method conditions might result not just in an increase in enrichment but also in a longer binding motif.

3.2 Introduction

Regulation of gene expression strongly depends on the function of transcription factors, which bind to cis-regulatory elements of the gene and either initiate or inhibit translation of the target gene. Several different conserved DNA binding domains have been identified, wich allow recognition of a vast amount of different sequences varying between 5 to >30 bp in length, with an average of 9.9 bp for eukaryotes (Stewart *et al.*, 2012). Transcription factors make up a high amount of the total number of genes in eukaryotes: More than 1500 transcription factors are encoded on the genome of the model plant *Arabidopsis thaliana*, corresponding to over 5% of its proteome (Riechmann *et al.*, 2000). In comparison, the genomes of the model diatoms *Phaeodactylum tricornutum* and *Thalassiosira pseudonana* encode for 212 and 258 known transcription factors, respectively (Rayko *et al.*, 2010), corresponding to approximately 2% of their proteomes. Thus, there seems to be a trend of increasing content of transcription factors with increasing organism complexity, with humans

being estimated to have 10% of its proteome dedicated to transcription factors (Levine and Tjian, 2003). Despite their importance, however, transcription factors remain a poorly characterized protein family: The JASPAR database (Mathelier *et al.*, 2016), for example, lists less than 250 plant transcription factors and about 500 vertebrate transcription factors with a known DNA binding motif.

Several different approaches are suitable for identifying the function of individual transcription factors. One possibility is the use of reverse genetics approaches, i.e. knockdown or knockout of the protein of interest, to identify phenotypes caused by the loss of the target protein, thereby inferring genes regulated by the transcription factor. However, it is very difficult to distinguish between direct and indirect effects using this approach. To identify only genes directly regulated by the transcription factor and not by downstream processes, the DNA binding motif and/or DNA binding site needs to be identified. The most common approach, Chromatin Immunoprecipitation and its variations, depend on highly specific antibodies which need to be tightly validated before use (Landt *et al.*, 2012). This need can be circumvented by introduction of tagged variants, however, wild type protein should be knocked out and expression levels need to be similar to the wild type protein in order to avoid artifacts (Kolodziej *et al.*, 2009). A high-throughput analysis of in vitro protein-DNA interactions, termed Bind-n-Seq, has been established as an alternative method (Zykovich *et al.*, 2009). Using tagged variants overexpressed in *E. coli* no antibody is required. Instead, small purification columns, for example Nickel beads for His-tagged proteins, can be used. The protein-DNA complexes are then purified and sequenced by next-generation sequencing methods to obtain DNA binding motifs. As the binding location within the genome is unknwon, unlike with ChIP-seq, the DNA binding motifs need to be mapped onto the promoter sequences of the organism to identify genes directly regulated by the protein of interest, requiring a bioinformatics approach not yet established for diatoms.

Aureochromes, which were recently discovered in the xanthophyte *Vaucheria frigida*, are a novel class of blue-light dependent transcription factors utilizing a bZIP domain to bind to their target motif on the DNA (Takahashi *et al.*, 2007). So far, they have only been identified in Stramenopiles (Ishikawa *et al.*, 2009). The genome of the model diatom *P. tricornutum* encodes for four different Aureochrome isoforms. Only one isoform, PtAUREO1a, has been characterized using knockdown strains (Schellenberger Costa *et al.*, 2013b). Additionally, the diatom-specific cyclin 2 (*dsCYC2*) controlling the onset of the cell cycle after dark arrest, has been identified by Yeast 2 Hybrid assays to be directly regulated by PtAUREO1a (Huysman

et al., 2013). However, no experimental evidence is available for its DNA binding motif. In this study, we attempted to use the Bind-n-Seq methodology for two Aureochrome isoforms of *P. tricornutum*: PtAUREO1a and PtAUREO1c. The putative binding motifs we observed were then mapped onto the promoterome of *P. tricornutum* by a bioinformatics approach to identify potential candidates for regulation by Aureochromes.

3.3 Materials and methods

3.3.1 Bind-n-Seq

N-terminally His-tagged PtAUREO1a and PtAUREO1c were kindly provided by Elena Herman and Tilman Kottke (Universität Düsseldorf, Germany). The purification procedure employed is described in (Herman *et al.*, 2013). Bind-n-Seq was conducted according to the principle described in (Zykovich *et al.*, 2009) using HPLC-purified oligonucleotides with a 23bp random region (Sigma Aldrich, Steinheim, Germany). The different conditions used are given in Table 3-1. NaCl was used in the binding buffer instead of $ZnCl_2$, as the protocol described by Zykovich et al. had been adapted for Zinc finger transcription factors. The binding reaction was performed at room temperature for 2h. For purification using magnetic His-tag beads (Expedeon, Swavesey, UK), samples were washed 6 times for 5min with 300 mM NaCl, 50 mM $NaPO_4$ pH8. Protein-DNA complexes were eluted using wash buffer containing either 1 M NaCl or 250 mM imidazole.

Table 3-1: Overview of the different conditions used for Bind-n-Seq. The barcode sequence was used to identify the different conditions after next-generation sequencing as all samples were pooled. The differences in the concentrations of PtAUREO1a and PtAUREO1c used are due to a slightly decreased percentage of FMN-bound protein in the PtAUREO1c protein preparation.

Sample barcode	Target protein	Elution condition
AAA	BSA	Imidazole
AGC	BSA	NaCl
TAG	PtAUREO1a (15 nM)	Imidazole
GCT	PtAUREO1a (15 nM)	NaCl
CTA	PtAUREO1c (20 nM)	Imidazole
TCC	PtAUREO1c (20 nM)	NaCl
ATG	PtAUREO1a+1c (10 nM each)	Imidazole
GAA	PtAUREO1a+1c (10 nM each)	NaCl

Protein-bound DNA was amplified in 50 µl reactions containing 0.5 µl of template DNA by 20 cycles with Phusion DNA polymerase (Thermo Fisher, Schwerte, Germany) and purified using the Geneclean Turbo kit (MP Biomedicals, Eschwege, Germany) according to the manufacturer's instructions. DNA concentration was determined using a Qubit fluorometer (Thermo Fisher). A total of 100 ng of amplified DNA from each condition was pooled into 30 µl of water and used for library preparation. Samples were analyzed on an Ion PGM next-generation sequencer (Thermo Fisher) by Marion Eisenhut and Andreas Weber (Universität Düsseldorf, Germany).

3.3.2 Identification of genes putatively regulated by PtAUREO1a or PtAUREO1c

Binding motifs were generated from sequencing data using the perl program Mermade 1.03 (https://github.com/KorfLab/Mermade) with default settings. The promoter regions (-1000 to +100) of genes from the "optimized gene catalog" of *P. tricornutum* (Gruber *et al.*, 2015) were extracted and searched for occurrence of the putative binding motifs. Annotations of the corresponding genes were extracted from the JGI genome portal (Nordberg *et al.*, 2014) and the DiatomCyc database (Fabris *et al.*, 2012). Overrepresentation or depletion of the motifs was tested for by comparing the frequency of occurrence of a motif within the promoter regions of the genome with randomized sequences with the same ACGT content and length. Sequence logos were created with Weblogo (Crooks *et al.*, 2004). The scripts written for this purpose by Mirta Jacobs are explained in detail in her Master's thesis (Jacobs, 2015).

3.4 Results

3.4.1 Generation of DNA binding motifs for PtAUREO1a/1c *via* Bind-n-Seq

Characterization of transcription factor binding sites and/or motifs are an important step of identifying the function of a transcription factor. As *in vivo* methods like Chromatin Immunoprecipitation generally require a highly specific antiserum recognizing native protein (Landt *et al.*, 2012), which was shown to be problematic using the antibodies validated in Chapter 2, an *in vitro* approach called Bind-n-Seq (Zykovich *et al.*, 2009) using heterologously expressed PtAUREOs was tried instead. Purified PtAUREO1a and PtAUREO1c with an N-terminal His tag were kindly provided by Elena Herman and Tilman

Kottke (Universität Bielefeld, Germany). BSA, PtAUREO1a, PtAUREO1c and a combination of both PtAUREO isoforms due to their known interaction capability (Banerjee *et al.*, 2016b) was incubated with the 23 bp random region oligonucleotides for identification of the binding motif by Bind-n-Seq. Additionally, elution from the Nickel beads after purification was tried with either 1 M NaCl or 250 mM imidazole to either separate the DNA from the protein or the DNA-protein complexes from the beads, respectively. Interestingly, elution with 1 M of NaCl was sufficient to elute the PtAUREO1c-bound DNA, but not the DNA bound to PtAUREO1a, as evidenced by a strongly decreased amount of DNA obtained for PtAUREO1a when eluted with NaCl (see Figure 3-1). This indicates a stronger affinity of PtAUREO1a to the DNA under the conditions assayed. Therefore, PtAUREO1a-DNA complexes and PtAUREO1a/PtAUREO1c-DNA complexes were eluted with imidazole instead.

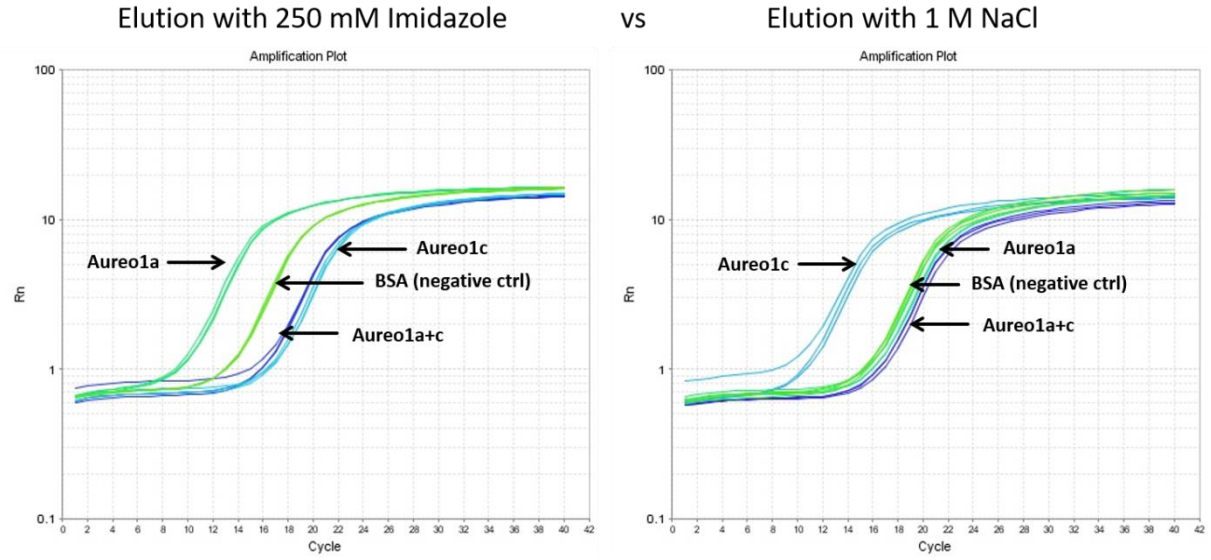


Figure 3-1: Raw data of qPCR using DNA eluted from the bead-protein-DNA complex using either 250 mM of imidazole to disrupt the bead-protein interaction (left panel) or 1 M NaCl to disrupt the protein-DNA interaction (right panel). The fewer PCR cycles are required for the fluorescence intensity to cross the threshold the higher the initial DNA content of the sample.

Samples were then amplified by PCR to obtain enough material for sequencing. Library preparation and sequencing was performed in cooperation with Marion Eisenhut and Andreas Weber (Universität Düsseldorf, Germany). Analysis of the sequencing data was performed using the software Mermade 1.03 (https://github.com/KorfLab/Mermade) using standard parameters. The motifs obtained for PtAUREO1a and PtAUREO1c were 8 bp in length and showed the characteristic ACGT core domain of bZIP transcription factors and strong similarity with the binding motif of VfAUREO1a (see Figure 3-2). However, enrichment of the motifs found was only about 50%, compared to a 2- to 10-fold enrichment reported by the creators of the Bind-n-Seq method (Zykovich *et al.*, 2009). Thus, those preliminary motifs

might not be biologically relevant. No motif was found to be enriched for the PtAUREO1a/PtAUREO1c mixture conditions, independent of elution conditions.

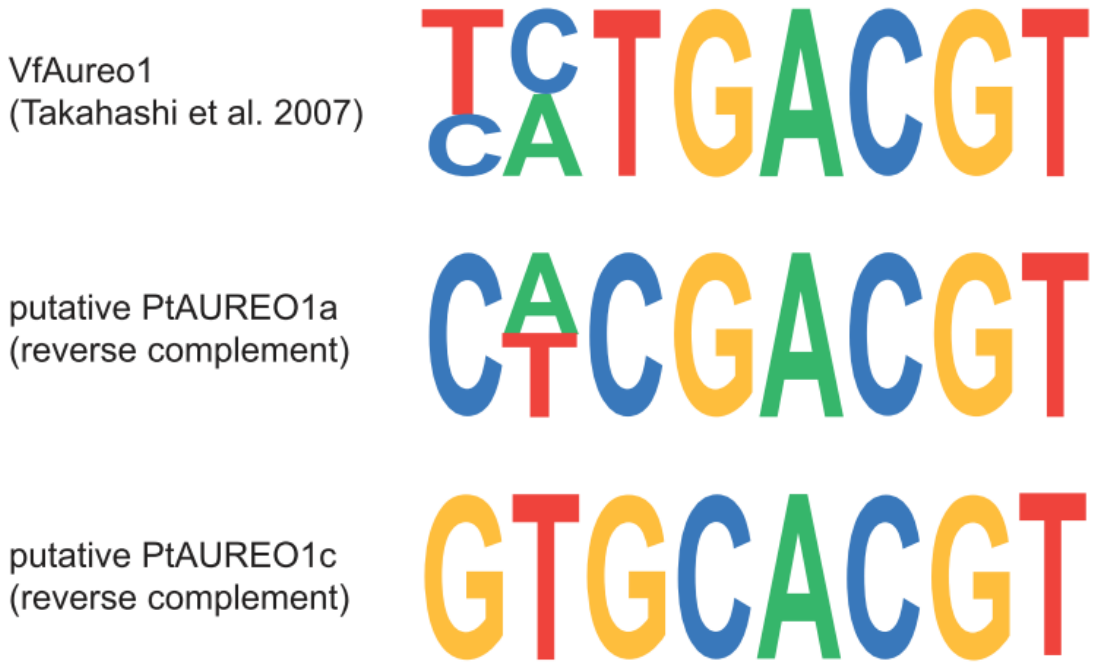


Figure 3-2: DNA binding motifs obtained by the Bind-n-Seq approach for PtAUREO1a and PtAUREO1c. As no information for the orientation of the motifs can be determined from this *in vitro* based assay, the motifs obtained are shown in reverse complement due to the higher similarity to the motif reported for VfAureo1 (Takahashi *et al.*, 2007).

3.4.2 Identification of genes containing the DNA binding motifs in their promoters

A bioinformatics approach was used to identify promoter regions which contain these putative binding motifs to find out whether this approach is feasible. Extraction of these promoter regions requires a gene catalog with as little redundancy as possible to avoid multiple extraction of the same promoter regions. Thus, the start codons of the open reading frames were determined using an optimized *P. tricornutum* gene catalog, which contains 10814 gene models that match some or all of the following criteria: subsequent, unique, not longer than 10 kb, start with a start codon and end with a stop codon and has no internal stop codon, has the longest total length compared to other models at the same position in the genome and is supported by EST data (Gruber *et al.*, 2015). Using a promoter region of -1000 to +100 from the start codon 1321 hits were found for PtAUREO1a and 321 for PtAUREO1c. The hits for PtAUREO1a were found to be distributed on 1188 different promoters, with 1065 promoters containing one motif, 113 containing two motifs and 10 containing three motifs. For

PtAUREO1c, the hits were spread on 306 different promoters with 291 promoters containing a single motif and 15 promoters containing two motifs. Additionally, 20 genes were found which contain a binding motif for both PtAUREOs. A complete list of the promoters identified including annotation data if available is shown in Supplemental dataset S3-1 (see attached CD-ROM). This number of potentially regulated genes for a single transcription factor is much higher than we had expected, indicating that the motif might be incomplete. To see whether this number could be decreased by identification of hot spots at certain distances from the start codon, occurrences of the motif were plotted relative to their position in the chosen promoter region, however no clear pattern was obtained (data not shown). As an alternative approach, sequence logos were generated 50 bp upstream and downstream of the putative binding motifs in promoters containing either one or two binding motifs to see whether the motif could be extended (see Figure 3-3A-E). For promoters containing only a single site no additional nucleotides were found to be enriched. In promoters containing two binding motifs of the respective PtAUREO, however, an additional G and T nucleotide were found to be enriched at the 3' end, extending each motif by 2 bp. Interestingly, this turns the PtAUREO1c motif into a palindrome (ACGTCGACGT). However, as the enrichment of the original motifs themselves is not very high, and no experimental evidence for these elongated motifs is available, not too much confidence should be put in these elongated motifs.

Furthermore, we wanted to find out whether the motifs are overrepresented within the promoter sequences. Thus, we compared the number of occurrences within the promoter sequences to 1000 randomizations of the promoter sequence with the same length and GC content separately for each possible orientation of the binding motif. The z-value, which is a multiple of the standard deviation, was used as a measure of significance. If the z-value of a single sample (the number of hits of one motif in the promoter sequences) exceeds the standard deviation of the control (the number of this of the same motif in the randomized promoter sequences) by more than three times a result is considered significant. Within the randomized promoter sequences each motif occurred about 160 times on average (see Figure 3-4A/B). The motifs identified for PtAUREO1c occurred almost the same amount of times within the promoter sequences and was thus found not to be enriched. The motifs for PtAUREO1a, however, had z-values between 10 and 17, and were thus strongly enriched within the promoter regions.

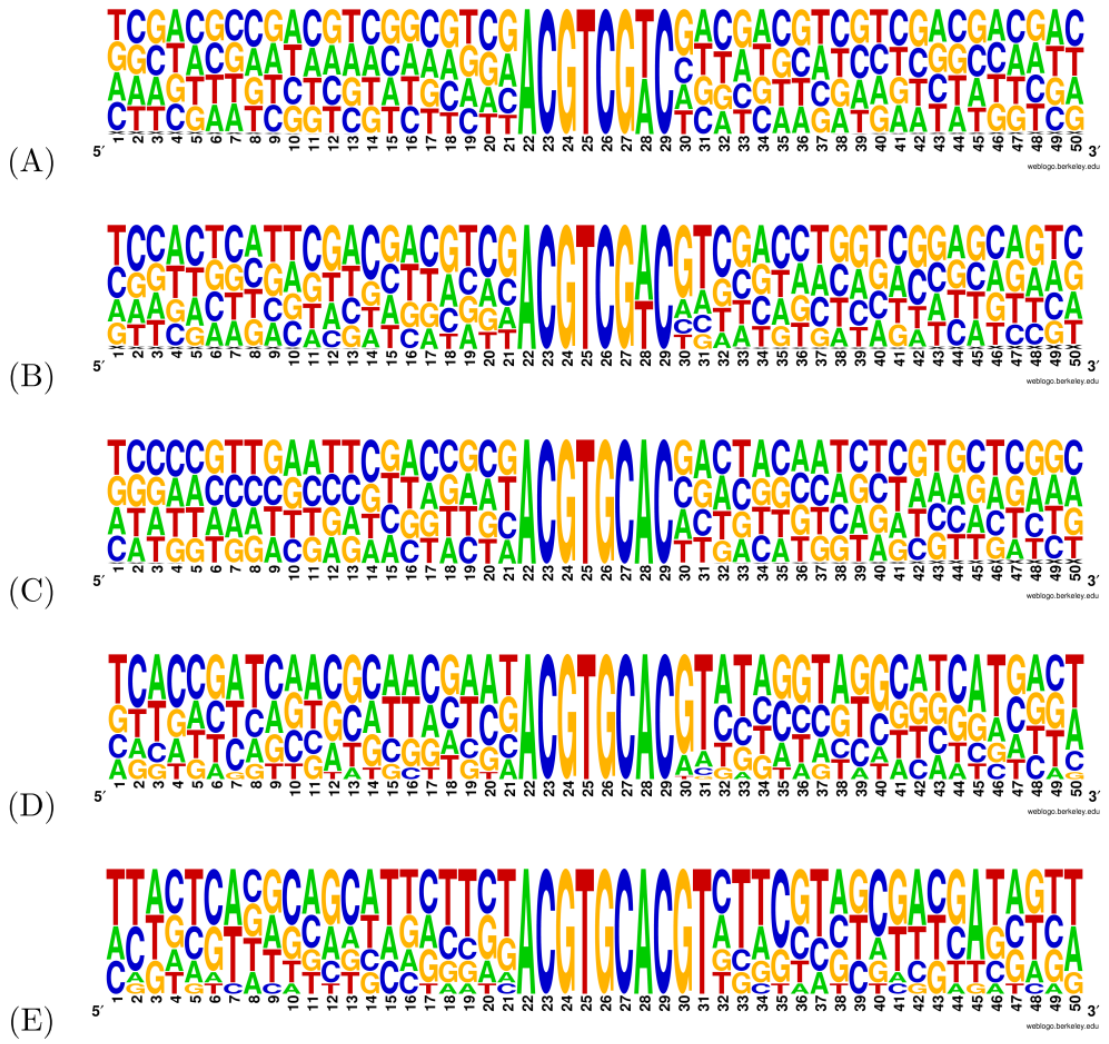


Figure 3-3: Sequence logos of the binding motifs of PtAureo1a and PtAureo1c. 50 nucleotides around a motif were taken from the according promoter-sequence and a frequency plot was created with Weblogo (Crooks *et al.*, 2004). A: all 1321 occurrences of the PtAureo1a motif, B: 145 motifs that occurred in promoters, which contained the PtAureo1a motif in both directions, C: all 321 occurrences of the PtAureo1c motif, D: 30 motifs that occurred in promoters, that contained two PtAureo1c motifs, E: 11 motifs of the palindromic PtAureo1c.

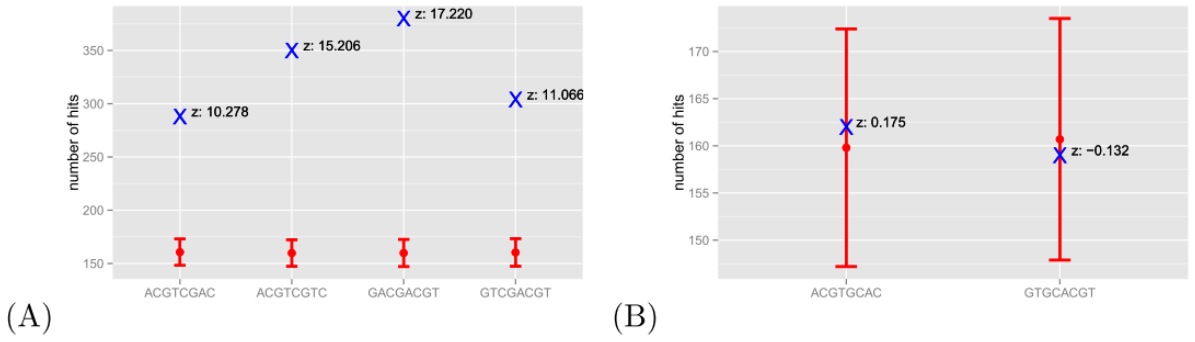


Figure 3-4: Significant enrichment or depletion of the motifs of PtAUREO1a or PtAUREO1c within the promoter regions of *P. tricornutum* compared to 1000 randomizations with the same length an GC content. The mean number of hits (red dot) for each possible version of the motif in 1000 randomizations of the promoter-sequences is shown, with the standard deviation (red bar). The blue cross indicates the z-value of the occurrences of each motif in the original promoter-sequences in comparison to the randomizations. A: significance of the PtAureo1a motif, B: significance of the PtAureo1c motif.

While in principle the analysis by bioinformatics to determine binding sites from a given motif was shown to be possible, a longer motif than 8 bp seems to be necessary to derive biologically relevant information from it, as the numbers obtained are far too high for manual validation of the potential target genes by EMSA. Thus, either the methodology of Bind-n-Seq must be better adapted for the use with Aureochromes or an alternative approach is required to determine significant binding motifs.

3.5 Discussion

Binding motifs of 8 bp length were determined for PtAUREO1a and PtAUREO1c using Bind-n-Seq. While these motifs show strong similarities to the motif reported for VfAureo1a and feature the bZIP core recognition sequence ACGT, the enrichment of the motifs was very low compared to what was reported by (Zykovich *et al.*, 2009) for the method. Thus, the author's confidence in the motifs is not very high, and the experiment should be repeated to obtain more meaningful motifs. Important findings which might lead to an improvement of the experiment were made, which were unavailable at the time this study was done: The addition of 10 mM of $MgCl_2$ to the binding buffer for electrophoretic mobility shift assays (EMSA) was found to be important for sequence-specific binding to the DNA. In addition, blue-light conditions were established for which PtAUREO1a binds to a DNA fragment containing the *dsCYC2* promoter sequence with high affinity, but has a strongly reduced affinity to an unspecific DNA fragment. Finally, PtAUREO1a concentration-dependent EMSA assays revealed the optimal protein to DNA ratio for similar experiments (Heintz and Schlichting, 2016). Thus, repetition of the experiment with these conditions in mind might lead to an increased enrichment and potentially slightly longer motif, and therefore statistically and biologically significant sequences. Alternatively, the TALEN or CRISPR approaches recently established in diatoms (Daboussi *et al.*, 2014, Weyman *et al.*, 2015, Nymark *et al.*, 2016, Serif *et al.*, 2017) could be adapted to introduce a tag into the gene sequence of the Aureochromes to allow the use of ChIP-seq with a commercially available antibody without artifacts due to overexpression of the target protein. This approach would offer the advantage of directly mapping sites on the genome to which the Aureochromes bind to instead of inferring it from motifs it bound to *in vitro*.

Nonetheless, a bioinformatics approach was employed to find out whether identification of genes which contain the motif in their promoter region proves successful. The motif identified

for PtAUREO1a was found to be overrepresented significantly within the promoter regions of *P. tricornutum* compared to 1000 randomized promoter regions with the same length and GC content, whereas the motif identified for PtAUREO1c was found not to be significantly overrepresented or depleted (see Figure 3-4). Due to the very high number of hits (1321 for PtAUREO1a and 321 for PtAUREO1c), however, it is unlikely that all these sites are targeted *in vivo*, and are too high in number to be verified experimentally. As transcription factor binding motifs seem to be, on average, 9.9 bp in eukaryotes (Stewart *et al.*, 2012), almost 2 bp more than the ones experimentally identified, it is possible that the sequences identified are indeed incomplete binding motifs. Interestingly, extending the sequence logo for promoters with two binding motifs for the respective PtAUREO present lead to the discovery of two additional conserved nucleotides at the 3' end (G and T, see Figure 3-3), however, no experimental evidence for these elongated motifs is available. A motif length of 10 bp would strongly reduce the amount of target sites found, most likely resulting in a low enough number of sites to allow validation of the individual sites obtained by EMSA. Hence, it is highly unlikely that the even shorter TGACGT motif reported for VfAUREO1 (Takahashi *et al.*, 2007) is the whole motif and thus should not be used for identification of potentially Aureochrome-regulated genes in diatoms.

3.6 Acknowledgements

Purified PtAUREO1a and PtAUREO1c with an N-terminal His tag was kindly provided by Elena Herman and Tilman Kottke (Universität Bielefeld, Germany). Library preparation and sequencing was performed in cooperation with Marion Eisenhut and Andreas Weber (Universität Düsseldorf, Germany). Ansgar Gruber provided helpful advice for the bioinformatic analysis. This work was supported by the University of Konstanz, the graduate school Biological Sciences (GBS), and the DFG (grant KR 1661/8-2 to PGK).

4 Generation of TALEN-mediated gene knockouts in the diatom *P. tricornutum*

Serif, M.[+], Lepetit, B[+], Weißert, K., Kroth, P.G. & Rio Bartulos, C*

*Corresponding Author: carolina.rio@gmail.com

[+] Both authors contributed equally

Plant Ecophysiology, Fachbereich Biologie, Universität Konstanz, D-78457 Konstanz, Germany

Keywords: *Phaeodactylum tricornutum*, TALEN, PtAUREO1a knockout, blue light-dependent transcription factor

Serif, M., Lepetit, B., Weißert, K., Kroth, P.G. and Rio Bartulos, C. (2017) A fast and reliable strategy to generate TALEN-mediated gene knockouts in the diatom Phaeodactylum tricornutum. *Algal Research*, **23**, 186-195.

4.1 Abstract

Reverse genetics techniques are powerful tools for studying gene functions. In the model diatom *Phaeodactylum tricornutum*, RNAi-mediated knockdown of genes still is the most commonly used reverse genetics technique. Due to the diploidic life cycle missing reproduction in lab cultures, many commonly used techniques to create knockout instead of knockdown lines are not applicable in *P. tricornutum*. These limitations can be overcome by using genome editing approaches like TALEN (Transcription activator-like effector nucleases), and/or CRISPR/Cas9 (clustered regularly interspaced short palindromic repeats), allowing the introduction of targeted mutagenesis events. Both techniques have recently been adapted exemplarily for diatoms, however, no concise guidelines exist yet for routine utilization of these tools and the subsequent characterization of the mutants. We therefore have adapted a cost-effective TALEN generation system previously established for mammalian cells for the use in *P. tricornutum*, allowing the assembly of TALENs in about two weeks. We further provide protocols for: a) choosing a TALEN target site in order to avoid potentially ineffective and/or off-target prone TALEN constructs, b) efficient transformation of *P. tricornutum* with both TALEN constructs, utilizing two antibiotics resistance markers, c) effective screening of the transformants. In order to test our system, we chose the blue-light dependent transcription factor Aureochrome 1a (PtAureo1a) as a target gene due to the known phenotype of previously characterized *P. tricornutum* RNAi knockdown strains. Our TALEN approach appears to be highly efficient: targeted mutation events were detected in 50% of all transformants obtained, whereas 21% of the transformants were found to be bi-allelic knockout lines. Furthermore, most TALEN transformed cell lines were found to be genetically homogeneous without the need for re-plating, which greatly facilitates the screening process.

4.2 Introduction

Diatoms are unicellular microalgae belonging to the Stramenopiles. They play an important role in global carbon fixation as well as for the nitrogen, phosphorous and silica cycles (Yool and Tyrrell, 2003). They are widespread in most aquatic habitats, where they need to cope with large variations of light quality and quantity (MacIntyre *et al.*, 2000, Ragni and D'Alcalà, 2004). As diatoms may contain larger amounts of lipids (up to 50% of dry weight), which even can be increased by genetic manipulation (Trentacoste *et al.*, 2013, Levitan *et al.*, 2015),

they are suitable for the production of biodiesel and/or bioplastics (Chisti, 2007, Mata *et al.*, 2010, Roesle *et al.*, 2014). The pennate *Phaeodactylum tricornutum* has become a model system for diatoms because of the availability of the genome sequence as well as genetic transformation techniques (Zaslavskaia *et al.*, 2000, Bowler *et al.*, 2008, Niu *et al.*, 2012, Miyahara *et al.*, 2013, Zhang and Hu, 2014, Karas *et al.*, 2015), allowing reverse genetics approaches. Because of their diplontic life cycle and the lack of of sexual reproduction in the lab, many methods for genetic manipulation like random mutagenesis or crossing are not available. Accordingly, knockdown via RNAi is currently the most commonly used approach for genetic manipulations (De Riso *et al.*, 2009). Recently, two new genetic tools for directed genome editing were developed, which allow induction of targeted DNA double-strand breaks to knock out genes irreversibly: the TALEN (Transcription activator-like effector nucleases) (Christian *et al.*, 2010, Li *et al.*, 2011b, Wood *et al.*, 2011) and the CRISPR/Cas9 (clustered regularly interspaced short palindromic repeats) systems (Cong *et al.*, 2013, DiCarlo *et al.*, 2013, Hwang *et al.*, 2013). Both approaches depend on nucleases that are guided to a specific DNA target sequence, and subsequently induce the formation of a DNA double strand break. The sequence-specific DNA binding of TALEN proteins is based on multiple 34 amino acid repeat modules, termed repeat variable di-residues (RVD), each binding specifically to one of the four nucleotides. All modules together (termed "targeting sequence") allow the recognition of a specific DNA sequence, so that in principle any DNA sequence of interest within the genome can be targeted (Schornack *et al.*, 2006, Moscou and Bogdanove, 2009). The catalytic domain of the endonuclease FokI, being fused to this targeting sequence, is only active as a dimer. Hence, for successful FokI activity two TALEN proteins are required to bind in the right orientation and in close proximity onto the DNA double strand to induce double strand break formation. The mandatory binding of both TALENs strongly increases the targeting specificity (Bitinaite *et al.*, 1998). The double strand break can be subsequently repaired by cellular mechanisms based either on homologous recombination (HR) or on non-homologous end joining (NHEJ). While NHEJ occurs during the whole cell cycle, HR is mainly restricted to the late S and G2 phase (Lieber, 2010). Both of these DNA repair mechanisms can be used to induce targeted mutations: HR can be exploited by introduction of foreign DNA with a strong homology to the DNA sequence surrounding the target site, which the cell uses as template to repair the double strand break (Zu *et al.*, 2013). If no HR template is used, the high error rate of NHEJ can be exploited to generate small random insertions/deletions until the target site is inaccessible for TALEN (Gong *et al.*, 2005).

The successful application of TALEN and CRISPR/Cas9 in *P. tricornutum* has been published recently (Daboussi *et al.*, 2014, Weyman *et al.*, 2015, Nymark *et al.*, 2016), however, a number of potential pitfalls have not yet been sufficiently addressed. Therefore, in this report, we describe how the TALEN approach can be optimized to obtain cell lines with targeted mutations at a high frequency and how to minimize the risk of potential off-target binding of TALENs. Additionally, we show that a thorough screening process is required for correctly distinguishing bi-allelic and mono-allelic knockout mutants, as well as to prove that a specific cell line is genetically homogeneous and does not contain different genotype lines. Although the CRISPR approach can be more easily adapted for a specific target site, we chose to improve the strategy of utilization of TALENs in diatoms because, based on research in other organisms (Wang *et al.*, 2015), it offers the potential for reduced off-target effects. We developed a diatom-specific TALEN system by combining two previously published *P. tricornutum* transformation plasmids, pPha-T1 (Genebank ID: AF219942.1, (Zaslavskaia *et al.*, 2000)) and pPha-NR (Genebank ID: JN180663.1, (Stork *et al.*, 2012)), and the TALEN plasmids developed by the Zhang lab for mammalian systems (Sanjana *et al.*, 2012). We adapted the Zhang protocol (Sanjana *et al.*, 2012) for TALENs assembly and verification for diatoms, allowing the creation of TALENs in less than two weeks. We furthermore developed concise guidelines for all steps of this method, from target site design to screening mutated lines, which allows any molecular biology lab equipped with a particle gun or an electroporator to produce knockout mutants of *P. tricornutum*. In order to evaluate our TALEN system, we generated TALENs targeting the *PtAureo1a* gene, encoding the blue-light photoreceptor Aureochrome 1a (PtAUREO1a), because of the availability of a specific antiserum as well as the known phenotype of RNAi-silenced strains, such as lower chlorophyll a (Chl *a*) content per cell and a higher relative amount of xanthophyll cycle pigments (Schellenberger Costa *et al.*, 2013b).

4.3 Material & Methods

4.3.1 Assembly of the TALEN plasmids

The plasmid kit used for building TALENs was a gift from Dr. Feng Zhang (Addgene, Cambridge, MA, USA; kit #1000000019) (Sanjana *et al.*, 2012). A HindIII restriction site was introduced upstream of the start codon of the TALEN plasmids using primer pair

TALEN_HindIII+_for/rev (see Table SI). The *P. tricornutum* expression plasmids pPha-T1 (GenBank AF219942, (Zaslavskaia *et al.*, 2000)) and pPha-NR (GenBank JN180663, (Stork *et al.*, 2012)) were modified using site-directed mutagenesis with the primer pairs PTV_BSAI1719SD_for/rev and PTV_BSAI2888SD_for/rev (see Table S4-1) to remove the two BsaI restriction sites. Furthermore, the Zeocin resistance gene (*Sh ble*) from the pPha-NR vector was exchanged with the Nourseothricin resistance gene (*nat*) gene from the pNat vector (Zaslavskaia *et al.*, 2000), creating the pPha-NR-Nat vector. The different backbones of the TALEN expression plasmids from (Sanjana *et al.*, 2012) (pTALEN_v2_NG (detects T), pTALEN_v2_NI (detects A), pTALEN_v2_NN (detects G or A), pTALEN_v2_HD (detects C)) were excised from the respective plasmids using a HindIII/SacI double digest, and ligated into both pPha T1 and pPha-NR-Nat to create *P. tricornutum* specific TALEN plasmids pM9_fcpA_NG, pM9_fcpA_NI, pM9_fcpA_NN, pM9_fcpA_HD and pM9_NR_NG, pM9_NR_NI, pM9_NR_NN, pM9_NR_HD, respectively. The correct integration of the TALEN backbone was verified by Sanger sequencing (GATC, Konstanz, Germany). These plasmids specifically designed for producing TALENs in *P. tricornutum* will be made available at Addgene (www.addgene.org).

We created TALEN pairs with a total of 20 repeat variable di-residues (RVD) each. Both the first RVD, which is always NG (targeting T), and the last RVD, which is only a half monomer, are already included in the respective plasmids. Based on these prerequisites, possible target sites for the *PtAureo1a* gene (JGI Protein ID 49116) were generated using the TAL Effector Nucleotide Targeter 2.0 (Doyle *et al.*, 2012) using a fixed repeat array length of 19 per TALEN (the first NG is not counted by the software), a spacer length of 15 to 22 bp between the two TALENs and the *P. tricornutum* Refseq ID (GCF_000150955.2) to predict possible off-target effects. Target sites were chosen according to the following parameters (summarized in Figure 4-1): The target site should be an exon region in or upstream of the first functional domain, have no predicted off-targets and contain few NN repeat variable di-residues (RVDs), which lack specificity (targeting G/A) compared to the other RVDs. Lastly, cutting efficiency of good candidates was estimated using SAPTA (Scoring Algorithm for Predicting TALEN Activity; search type used: "Score individual TALEN pair(s), optimal spacer length is assumed in score calculation") (Lin *et al.*, 2014).

Cloning and insertion of the target sequences into our TALEN plasmids were performed as described in (Sanjana *et al.*, 2012). In short, a RVD monomer library was constructed via PCR from the plasmids pHD_v2, pNG_v2, pNI_v2 and pNN_v2 (Addgene

Kit # 1000000019, (Sanjana *et al.*, 2012)). The monomers were then digested and ligated in a Golden Gate type reaction into hexamers, amplified via PCR and purified. Last, the three hexamers and the plasmid containing the backbone were digested and ligated in a second Golden Gate reaction (a summary of the TALEN assembly procedure can be found in Figure 1). The resulting plasmids were sequenced (GATC, Konstanz, Germany) to verify correct integration and order of the 20 targeting TALE monomers.

4.3.2 Cultivation of algae

The *P. tricornutum* (Bohlin) strain UTEX646 was obtained from the culture collection of algae of the University of Texas (UTEX, Austin, USA). *P. tricornutum* was grown axenically in liquid F/2 medium without added silica and 16.5‰ salt content or on solid F/2 media which contained additionally 1.2% (w/v) Bacto Agar (BD, Sparks, MD, USA). Cells in liquid F/2 medium were cultivated in a 16 h/8 h light/dark cycle in Erlenmeyer flasks under continuous shaking at 20 °C and an illumination of 35 µmol photons m^{-2} s^{-1} (Osram Lumilux L58W/840, Munich, Germany). Plated cultures were cultivated under continuous illumination at 75 µmol photons m^{-2} s^{-1} (Osram Biolux L30W/965).

4.3.3 Nuclear transformation of *P. tricornutum*

Nuclear transformation of *P. tricornutum* was performed using a Bio-Rad Biolistic PDS-1000/He Particle Delivery System (Bio-Rad, Hercules, CA, USA) fitted with 900/1100/1350 psi rupture disks as described previously (Apt *et al.*, 1996, Zaslavskaia *et al.*, 2000, Kroth, 2007). 10^8 cells per plate were bombarded with 1.25 µg of each plasmid. For selective cultivation of *P. tricornutum* transformants, 75 mg mL^{-1} Zeocin (Invitrogen, Carlsbad, CA, USA) and 150 mg mL^{-1} Nourseothricin (ClonNat, Werner Bioagents, Jena, Germany) were added to the solid F/2 media (Apt *et al.*, 1996, Zaslavskaia *et al.*, 2000).

4.3.4 DNA isolation

Genomic DNA was isolated using the nexttec™ 1step DNA isolation from tissues & cells kit (Biozym, Hessisch Oldendorf, Germany) according to the manufacturer's instructions. A cell pellet corresponding to 10 ml of culture in mid exponential growth phase was used as starting material. Incubation at 56 °C and 1200 rpm was done either 6 h or overnight. Concentration

of genomic DNA was measured by Nanodrop 2000 UV/VIS Spectrometer (Thermo Fisher, Schwerte, Germany).

4.3.5 Allele-specific PCR

DNA sequences for the two alleles of the *PtAureo1a* gene (JGI ID: 49116) were deduced from the whole genome shotgun sequencing (WGS) database (NCBI) by alignment of the individual sequence reads. Allele-specific primers for PCR were derived which include an allele-specific difference on the 3' terminal base, thereby preventing polymerases without proofreading function from amplifying the respective other allele. PCR was performed using either Taq B polymerase (Biozym, Hessisch Oldenburg, Germany) or HiDi polymerase (myPOLS, Konstanz, Germany) according to the manufacturer's instructions. An extension time of 4 min and an annealing temperature of 52°C (Taq B) or 55°C (HiDi polymerase) were used to amplify the TALEN target site of both *PtAureo1a* alleles from isolated genomic DNA using primers Aureo1a_for and Aureo1a_rev (see Table S4-1). PCR products were separated on 1% agarose gels and PCR products were isolated using the Geneclean Turbo Kit (MP Biomedicals, Eschwege, Germany) according to the manufacturer's instructions. Purified DNA was analyzed by Sanger sequencing using primer Aureo1a_for (see Table S4-1) (GATC, Konstanz, Germany, or Source Bioscience, Berlin, Germany). If sequencing results indicated mixed populations due to small insertions and/or deletions in one or both alleles, the PCR product was sub-cloned using the pGEM-T system (Promega, Mannheim, Germany).

4.3.6 Southern Blotting

Isolated genomic DNA was digested using each of the following restriction enzymes overnight according to the manufacturer's instructions: BamHI, BsrGI, HindIII (Thermo Fisher, Schwerte, Germany). Samples of 400 ng digested DNA were separated on 0.8% agarose gels. The agarose gel was incubated for 10 min in denaturation solution (0.5 M NaOH, 1 M NaCl), followed by 10 min incubation in neutralization solution (0.5 M Tris-HCl, 3 M NaCl, pH 7.5). A dry blot was then performed overnight: The gel was placed downside-up onto an acrylic glass plate and the positively charged nylon membrane (Roche, Mannheim, Germany; 11471240001), three layers of Whatman paper (3MM Chr, 3030917, VWR, Darmstadt, Germany) and absorbent paper were placed on top. The blotting setup was then weighed down. The 400 bp DIG-labeled probe was synthesized using the PCR DIG Probe Synthesis Kit (Roche, 11636090910) using an *PtAureo*1a-containing plasmid

(Schellenberger Costa *et al.*, 2013b) as template and primers Aureo1a_probe_for / rev (see Table S4-1). Hybridization occurred overnight using DIG Easy Hyb (Roche, 11603558001) at 50 °C. Post-hybridization steps were performed using the DIG Block and Wash Buffer set (Roche, 11585762001), but with a shortened 68 °C washing step (2x10min). The Anti-DIG-AP antibody (Roche, 11093274910) was used at a 1:20.000 fold dilution; the alkaline phosphatase substrate used was CDP-Star (Roche, 12041677001). The blots were developed using X-ray films (Amersham Hyperfilm ECL, GE Healthcare, Munich, Germany) after 15 to 60 min incubations by a Konica SRX-201 Developer.

4.3.7 Protein Isolation and Immunoblotting

For protein extraction, cell pellets were resuspended in lysis buffer (4 M urea, 1.5 M thiourea, 1% SDS, 20 mM Tris pH 8) supplemented with protease inhibitor (Complete^{TM} EDTA-free, Roche) according to the manufacturer's instructions. A spatula tip of 1 mm, 0.5 mm and 0.1 mm diameter beads was added and the cells were lysed in a Savant FastPrep FP120 bead mill (Thermo Scientific, Karlsruhe, Germany) six times for 20 s with cooling on ice for 1 min between each cycle. Cell debris and residual beads were removed by two centrifugation steps at 18000 g and 4 °C for 30 min.

Proteins were separated in a 10% polyacrylamide gel by SDS-PAGE according to Laemmli et al. (Laemmli, 1970). Each lane was loaded with protein extract of either wild type or mutant cell lines. After blotting, the nitrocellulose membrane (Amersham Protran 0.1 µm NC, GE Healthcare) was cut between 35 and 40 kDa and the top half (40-250 kDa) was used to detect PtAUREO1a, whereas the bottom half (0-35 kDa) was used to detect the D1 loading control. Immunoblots using a custom-made antiserum specific against PtAUREO1a (Agrisera AB, Vännas, Sweden) were performed as described in (Schellenberger Costa *et al.*, 2013b), whereas the D1-specific antiserum (AS05-084, Agrisera AB) was used according to the manufacturer's instructions. Blots were developed using an Odyssey FC Imaging System (Li-Cor, Bad Homburg, Germany).

4.3.8 Chlorophyll *a* determination

Chlorophyll *a* (Chl *a*) was isolated using 10% methanol and 90% acetone successively and its concentration was determined spectrophotometrically using the formula of Jeffrey and Humphrey (Jeffrey and Humphrey, 1975).

4.3.9 Pigment extraction

Extraction of pigments and subsequent analyses via HPLC were done as described in (Jakob *et al.*, 1999). Samples were analyzed on a calibrated Hitachi LaChrom Elite HPLC system equipped with a Nucleosil 120-5 C18 column (Macherey-Nagel, Düren, Germany).

4.3.10 Measurement of non-photochemical quenching (NPQ)

Cell suspensions in mid-exponential phase were adjusted to a Chl *a* content of 1 µg ml^{-1} and NPQ was measured with an AquaPen-C AP 100 (Photon Systems Instruments, Brno, Czech Republic) using light pulses with an intensity of 2100 µmol photons m^{-2} s^{-1} applied every 20 s to induce maximal fluorescence and 700 µmol photons m^{-2} s^{-1} of actinic light to induce NPQ.

4.3.11 Determination of cell size by microscopy

Exponentially growing cells were analyzed using an Olympus BX51 epifluorescence microscope equipped with a Zeiss AxioCam MRm digital camera system (Carl Zeiss Microscopy GmbH, Göttingen, Germany). Length and width of 50 cells were determined for each cell line and the cell volume was approximated as described in (Rottberger *et al.*, 2013) using the formula $V = (\pi/12) * d^2 * h$ (d: diameter of the cell; h: length of the cell).

4.4 Results & discussion

4.4.1 Generation of the TALEN constructs

In order to perform a cost-effective and easy assembly of the TALEN targeting sequence, we chose a system developed previously for mammalian systems, allowing the complete assembly and sequence verification of individual TALEN plasmids within two weeks (Sanjana *et al.*, 2012). In contrast to previous publications on *P. tricornutum* (Daboussi *et al.*, 2014, Weyman *et al.*, 2015) describing the expression of TALENs in a single transformation plasmid, we decided to clone the two TALEN backbones into two separate plasmids, one with a constitutive fcpA promoter and *sh ble* gene conferring resistance to Zeocin (pPha-T1), and one with an inducible nitrate reductase promoter (Chu *et al.*, 2016) and a *nat* gene conferring resistance to Nourseothricin (pPha-NR-Nat). This design (shown schematically in Figure 1) has several advantages: the expression of both TALENs from two plasmids, as compared to a

single plasmid, reduces the size of the plasmids (from 13 to 14 to 7 to 8 kb), which facilitates cloning procedures and, based on results with other systems (Hanahan, 1983, Ohse *et al.*, 1995), may increase the transformation efficiency. Furthermore, introducing two plasmids at the same time usually results in high rates of co-transformed *P. tricornutum* cells even without a second selection marker (Zhang and Hu, 2014). However, when using two different antibiotic resistances on the two plasmids encoding the individual TALENs, a higher selection stringency is achieved by screening for strains that have integrated both plasmids. Larger plasmids instead are more prone to random DNA double strand breaks induced by the tungsten particles used for the particle bombardment (Krysiak *et al.*, 1999). Additionally, the inducible promoter system allows switching off the expression the TALEN, once favorable mutations have been demonstrated, which decreases the probability of off-target DNA modifications, and allows checking for lethal mutations by inducing the expression of the TALENs (by exchanging ammonia by nitrate) only after the initial round of antibiotics selection.

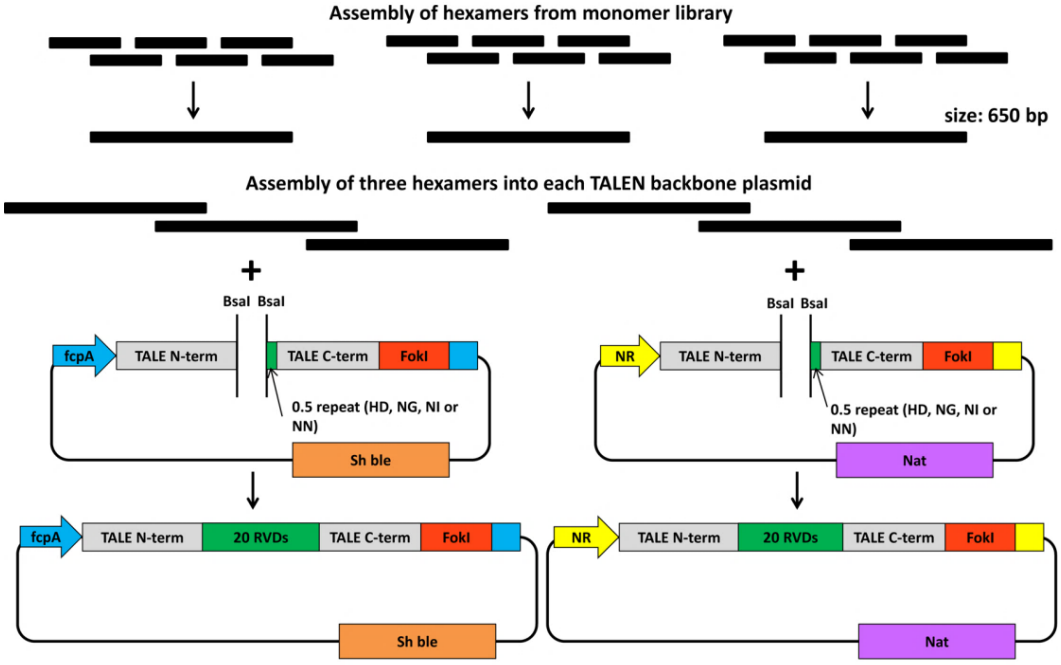


Figure 4-1: Schematic overview of the TALEN assembly process (according to [33]) and the resulting plasmids. Six monomers are assembled into hexamers corresponding to the desired target site in a golden-gate type reaction from a library consisting of 72 monomers (18 positions of the TALEN target site and 4 different RVDs), allowing assembly of multiple fragments in the desired order in a single step. Three of these hexamers are then ligated in a second golden gate-type reaction into the plasmid containing the first RVD (always NG) and the last half RVD (HD, NG, NI or NN, depending on the target site). Correct insertion of the target sequence needs to be verified by colony PCR (insert size: 2.2 kbp). Clones indicated to be positive by colony PCR should be verified by restriction digest with AfeI (expected fragment lengths: 5 kbp, 2.2 kbp and 165 bp; exemplary shown in Figure S4-1) as well as sequencing of the inserted fragment. fcpA: FcpA (=Lhcf1) promoter; *Fok*I: endonuclease; N-term/C-term: N and C terminus, respectively; Nat: nourseothricin resistance cassette; NR: nitrate reductase promoter; RVD: repeat variable di-residue; Sh ble: Zeocin resistance cassette

The recommended workflow for designing and generating TALEN constructs is presented in Figure 4-2. In order to avoid allele-specific polymorphisms in the TALEN target sites, which could hinder correct binding, allele-specific gene sequences were deduced from the *P. tricornutum* whole genome shotgun sequences (WGS), thereby identifying 11 different allele-specific polymorphisms in the *PtAureo1a* gene for strain Pt1 (CCAP 1055/1). In the next step, a combination of two online tools was used to choose the best potential target sites. TALE-NT 2.0 (Doyle *et al.*, 2012) was used to predict target sites in the gene of interest and potential off-targets for each TALEN pair based on the *P. tricornutum* RefSeq sequence (GCF_000150955.2). A total of 191 potential TALEN pairs were suggested within the first 700 bp of the *PtAureo1a* gene, of which only 29 were targeting an exon region and had no predicted potential off-targets. These TALEN pairs were then sorted by the frequency of the less specific NN RVD, and the *in vitro* cutting efficiency was estimated by SAPTA (Scoring Algorithm for Predicting TALE(N) Activity) (Lin *et al.*, 2014). According to the SAPTA guidelines, composite scores above 30 are recommended for a high rate of gene modifications. Additionally, the scores of the individual TALENs should have similar values (e.g., 20 and 25 are better than 5 and 40, although resulting in a similar composite score). A SAPTA analysis of the TALEN pair chosen for *PtAureo1a* (left target site: TCCCTCCTTAAGGAAGAGAA; right target site: TCGCCCAAGTGCGAACGAAT; spacer length: 19 bp) resulted in a composite score of 43.13 and scores for the individual TALENs of 25.44 and 20.15, respectively.

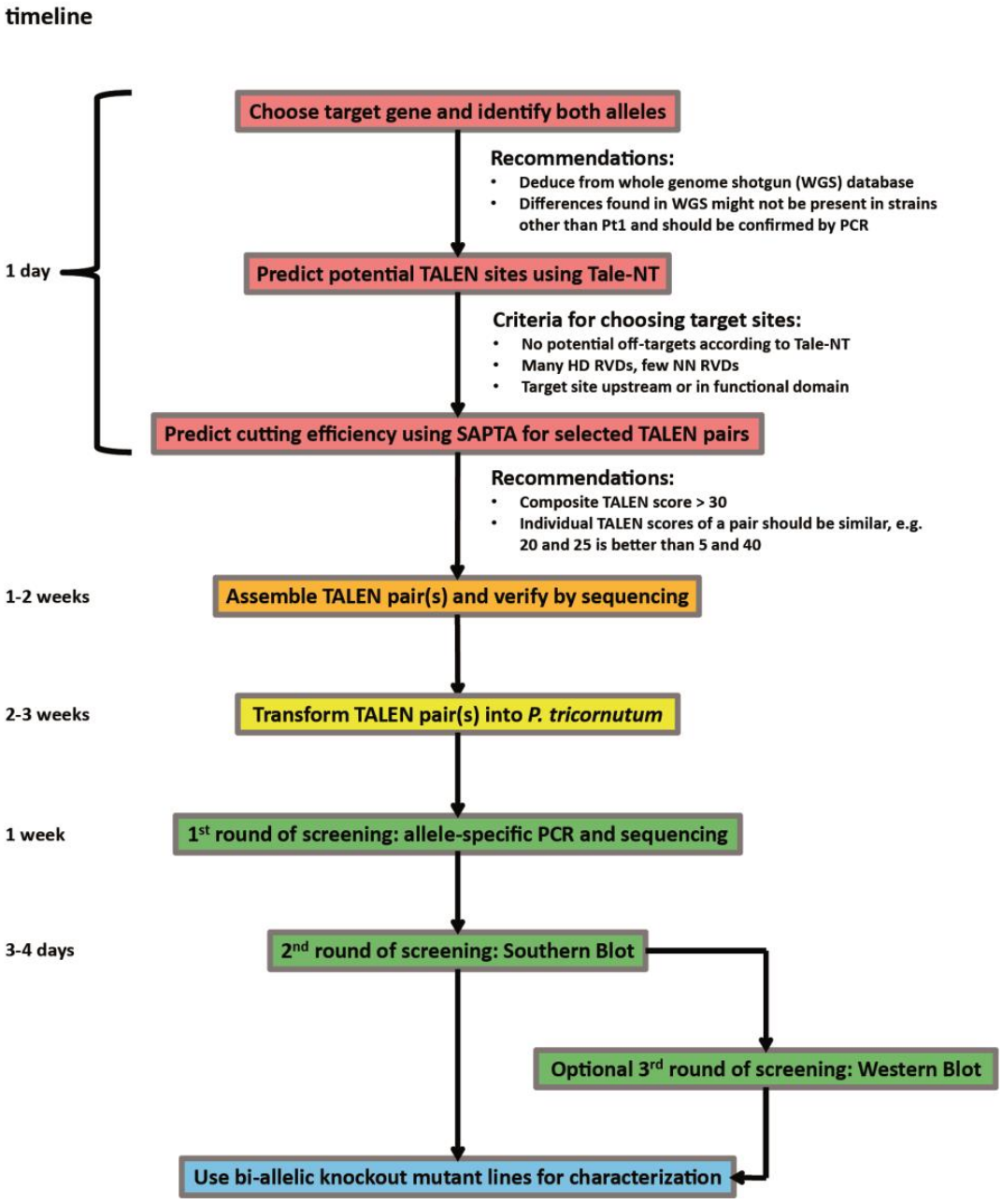


Figure 4-2: Workflow and recommendations for design and assembly of TALEN constructs, as well as screening of the obtained transformants. An estimated timeframe is given for each step.

The nucleases of the TALEN pair cleave at position 679, which is part of the leucine zipper domain of PtAUREO1a (see Figure 4-3, the TALENs are symbolized by a grey line and scissors symbols, the predicted cutting site by a dashed line). Thus, random mutations in this area could abolish DNA binding or may result in premature termination of protein translation. TALENs were assembled as described in Materials and Methods and verified by sequencing. We have applied the prediction tools to previously published TALEN target sites for *P. tricornutum* (Daboussi *et al.*, 2014, Weyman *et al.*, 2015), which had been constructed before these tools became available, and found potential off-targets for most of them. While

no direct evidence is available whether these potential off-targets are actually targeted by the TALENs *in vivo*, the search tool used here is much better suited than simple BLAST searches due to the complexity of the target sequence containing bi-specific RVDs and a gap of variable length.

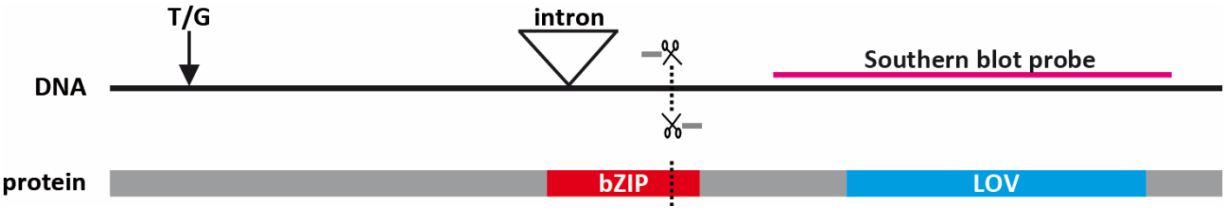


Figure 4-3: Schematic drawing of *PtAureo1a* and its gene product PtAUREO1a including conserved domains (bZIP in red and LOV in blue). The TALEN recognition sites are indicated by a grey line and the FokI endonuclease domains are symbolized by scissors. The predicted cutting site of the TALEN pair within the gene and its relative location within the gene product and its conserved domains is indicated by dashed lines. The allele-specific difference of *PtAureo1a* at position 81 (T/G) in strain Pt4 is indicated by an arrow, a 102 bp intron region by an inverted triangle shape and the binding site of the Southern blot probe by a magenta-colored line.

4.4.2 Screening of the obtained transformants and statistical evaluation

A total of 24 *P. tricornutum* colonies were obtained after transformation. While several allele-specific differences were identified (based on the sequenced *P. tricornutum* strain Pt1 (CCAP 1055/1)), the two alleles could not be amplified separately from Pt4 (UTEX 646), the strain used in this study. As the availability of the genome sequence greatly facilitates both TALEN design and screening, the use of wild type strain Pt1 is recommended for creation of *P. tricornutum* knockout mutants, unless previous work may require the use of another strain. We could solve the problem by identifying a mixed trace peak (T/G) at position 81 of the *PtAureo1a* gene in the Pt4 wild type cells as well as in the mutants, which allowed distinguishing both alleles. Using PCR, several small deletions (e.g., clone 6: 29 bp deletion in allele 1) or insertions (e.g., clone 14: 143 bp insertion in allele 2) were detected in these transformants, which are summarized in Table 4-1. No PCR products could be obtained for cell lines 8 and 9, presumably due to large insertions or deletions at the target site in both alleles. Furthermore, in eight transformant cell lines only one allele could be amplified by PCR, indicating mutation events in the respective other alleles. In case of clone 11, sub-cloning of the PCR fragment was necessary to demonstrate small deletions in each allele (7 and 18 bp, respectively). In summary, 10 clones were shown by PCR and sequencing to have no targeted mutagenesis events, whereas five clones (5, 6, 14, 16 and 21) were found to be at least mono-allelic mutants and clone 11 was found to be a bi-allelic mutant. As not all clones

could be reliably screened by PCR, we introduced a second screening process by employing Southern Blots of genomic DNA. The restriction enzymes BamHI, BsrGI and HindIII were chosen for digestion of the genomic DNA prepared from each of the 24 mutants. For each of the chosen restriction enzymes, Southern Blots of genomic DNA of wild type strains showed a single band in the range of 3 to 5 kbp (see Figure 4-4). In contrast, ten of the 24 transformants (1, 2, 3, 6, 8, 9, 13, 14, 16 and 18) showed distinct shifts of up to a few kbp (marked with an arrow). Out of these ten, eight strains (1, 2, 3, 6, 13, 14, 16 and 18) retained one band at the size of the wild type fragment, indicating that these are either heterozygous mutants or that small insertion or deletion events occurred in the second allele, which are not visible in Southern Blots, but which can be detected by PCR and sequencing (strain 16, see above). While the shifted band of strain 14 seems to indicate a shorter fragment, the 143 bp insertion detected by PCR (see above) introduced a new HindIII site, thereby causing a shorter fragment size. Two cell lines (8 and 9) did not show a band corresponding tothe wild type fragment size.

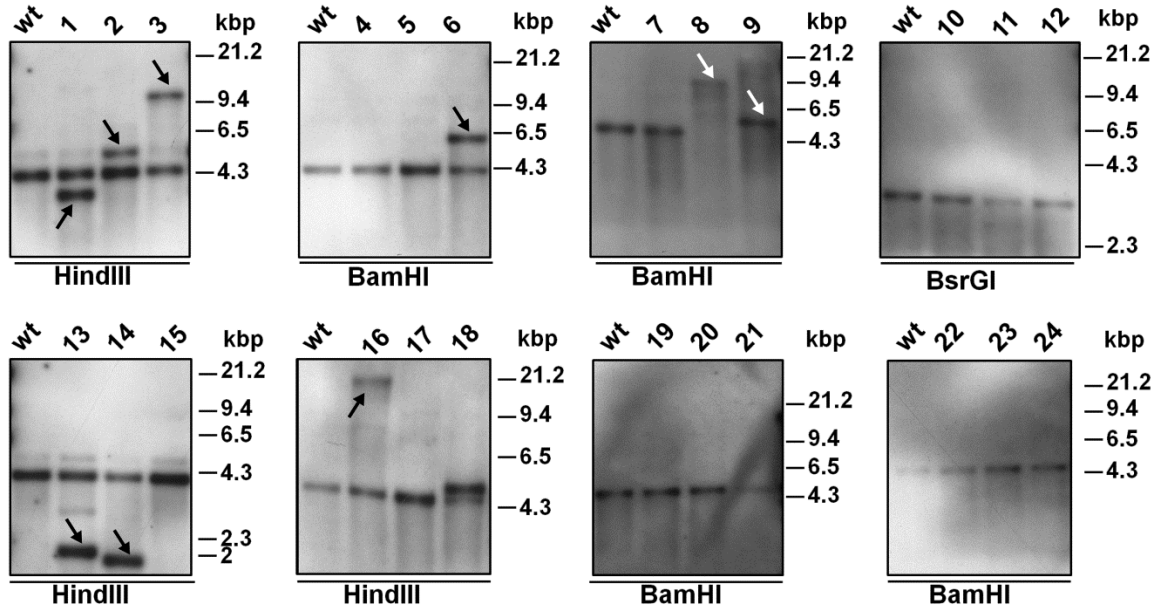


Figure 4-4: Southern Blots using digested genomic DNA of either wild type or transformed cell lines. Shifts of the DNA fragments can be caused either by insertion or deletion events. Shifted DNA bands due to insertions or deletions are marked with arrows.

An overview of the screening results for all mutants is given in Table 4-1. We conclude that 50% of all mutants had at least mono-allelic mutations induced by TALEN, whereas clones 6, 8, 9, 11 and 16 are bi-allelic mutant cell lines (20.83% of all mutants obtained). As only the combination of sequencing of PCR-amplified target genes and Southern Blots using a target gene specific probe allowed us to identify all mutated strains, we highly recommend

screening TALEN-transformants using both methods whenever possible. The workflow for screening the obtained mutants is presented in Figure 4-2.

Table 4-1: Evaluation of the screening of the two alleles via PCR and Southern Blot. Detected mutagenic events are indicated in bold.

Clone	Allele 1 (PCR)	Allele 2 (PCR)	Southern Blot*	conclusion
1	no change	**no amplification**	**1**	**mono-allelic mutation**
2	no change	no change	**1**	**mixed culture; mono-allelic mutation**
3	no change	**no amplification**	**1**	**mono-allelic mutation**
4	no change	no change	0	both alleles unchanged
5	no change	**2bp deletion**	0	**mono-allelic mutation**
6	**29bp deletion**	**no amplification**	**1**	**bi-allelic mutation**
7	no change	no change	0	both alleles unchanged
8	**no amplification**	**no amplification**	**2**	**bi-allelic mutation**
9	**no amplification**	**no amplification**	**2**	**bi-allelic mutation**
10	no change	no change	0	both alleles unchanged
11	**7bp deletion**	**18bp deletion**	0	**bi-allelic mutation**
12	no change	no change	0	both alleles unchanged
13	**no amplification**	no change	**1**	**mono-allelic mutation**
14	no change	**143bp insertion**	**1**	**mono-allelic mutation**
15	no change	no change	0	both alleles unchanged
16	**no amplification**	**3bp insertion**	**1**	**bi-allelic mutation**
17	no change	no change	0	both alleles unchanged
18	**no amplification**	no change	**1**	**mono-allelic mutation**
19	no change	no change	0	both alleles unchanged
20	no change	no change	0	both alleles unchanged
21	**692C→A****	no change	0	**mono-allelic mutation**
22	no change	**no amplification**	0	**mono-allelic mutation?**
23	no change	**no amplification**	0	**mono-allelic mutation?**
24	no change	no change	0	both alleles unchanged

***: Southern Blot results were grouped into three categories:**
0: indistinguishable from the wild type.
1: one band is at the height of the wild type band, the other is shifted due to large insertion or deletion events.
2: there is no band at the height of the wild type due to insertion or deletion events in both alleles.
****: a point mutation in clone 21 at position 692 downstream of the translation start is indicated**

4.4.3 Homogeneity of the generated mutants

As the average time required by the TALEN proteins to conduct a successful mutation in *P. tricornutum* is unknown, we have tested the homogeneity of the obtained cell lines with regard to double mutations. One mono-allelic knockout line (14) and four bi-allelic knockout lines (6, 8, 9 and 11) were spread onto individual agar plates, and three colonies of each line were re-isolated from single cell colonies. One re-isolated colony of cell line 9 did not grow properly in liquid medium, and thus was withdrawn from further analysis. Southern Blots and allele-specific PCRs were repeated in order to prove that the re-isolated clones were genetically identical regarding the mutated target site. Sequencing of PCR products showed no differences between re-isolated clones derived from cell lines 6, 11 and 14 and the sequences obtained prior to re-isolation. No PCR products could be obtained for re-isolated colonies of knockout lines 8 and 9 due to large insertion events. Southern Blots also showed identical band patterns for each mutant derived from the same clone, indicating homogeneity of the cell lines (see Figure 4-5).

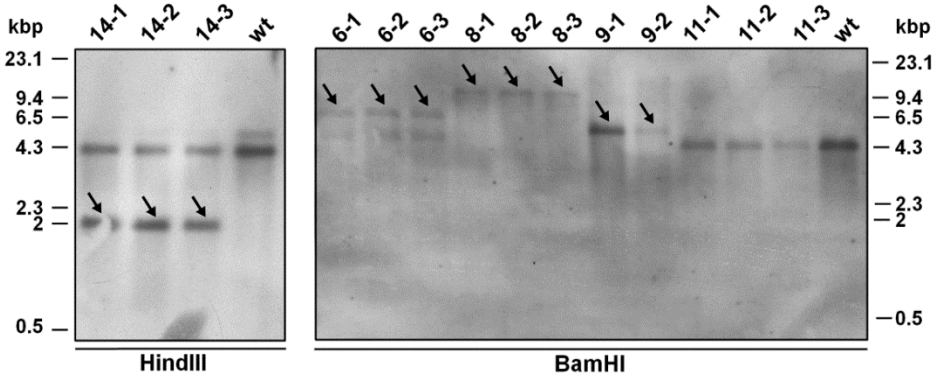


Figure 4-5: Southern Blot using digested genomic DNA of either wild type or transformed cell lines. Shifts of the DNA fragments can be caused either by insertion or deletion events. Shifted DNA bands due to insertions and deletions are marked with arrows.

As a final demonstration that both alleles had been effectively knocked-out, we studied the respective protein levels of the four bi-allelic lines and the mono-allelic knockout line by Western Blots using a specific PtAUREO1a-specific antiserum (Schellenberger Costa *et al.*, 2013b). An antiserum labeling the plastid-encoded D1 protein of photosystem II was used as a loading control. Western Blots using extracts of wild type cells or from the mono-allelic knockout mutant showed a single band, which corresponds to the molecular weight of 41.5 kDa of PtAUREO1a (see Figure 4-6). Knockout mutants derived from cell lines 6, 8 and 9 did not show any detectable bands at the expected size. All strains derived from cell line 11, however, showed a slightly truncated PtAUREO1a protein matching to the 18 bp deletion

(corresponding to a 6 amino acid deletion) observed by PCR in one of the alleles. Thus, clones 6-1, 8-1, 9-1 and 11-1 were chosen to characterize the phenotype of PtAUREO1a knockout mutants.

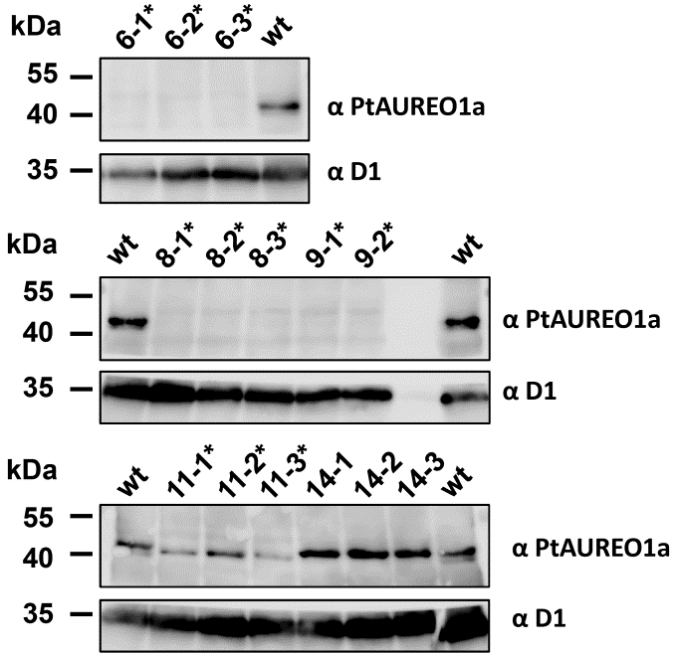


Figure 4-6 Western Blot of wild type and two mono-allelic as well as four bi-allelic cell lines (marked with a *) after re-isolation using the PtAUREO1a antiserum. The expected molecular weight of PtAUREO is 41.5 kDa. A D1-specific antiserum was used as a loading control.

Interestingly, the mono-allelic knockout line 14 no longer showed a reduced level of PtAUREO1a protein, as it had done prior to re-isolation (data not shown), although the TALEN induced mutations were still present, thus excluding allele repair/recombination (compare Figure 4-4 and Figure 4-5). We therefore conclude that *PtAureo1a* is a haplosufficient gene, i.e., one functional gene copy is sufficient to produce wild type-like protein levels when given some time for adaptation. Instability of knockdown strains has already been observed for RNAi-based silencing lines in *P. tricornutum* (Lavaud *et al.*, 2012) including the previously studied PtAUREO1a RNAi knockdown strain (data not shown). Thus, if knockdown cell lines are required, a variation of the amount of the targeted protein may occur over time and the TALEN-induced mono-allelic knockout strains should be tested regularly.

The generation of homogeneous bi-allelic knockout mutants without re-isolation is in strong contrast to the previously published diatom TALEN systems, which did not yield any homogeneous clones without replating (Daboussi *et al.*, 2014, Weyman *et al.*, 2015). Homogeneous bi-allelic knockout mutations require the TALEN proteins to induce the mutagenesis events in both alleles prior to the first cell division on selective media. We

routinely replated transformed cells after 24 h, hence this time span seems to be sufficient for the TALENs to operate. The reason why previously published systems (Daboussi *et al.*, 2014, Weyman *et al.*, 2015) did not obtain genetically homogeneous transformants, despite replating the cells onto selection media after 48 h, is unclear. One explanation could be the carefully chosen target site based on the newly available prediction techniques, which both avoids off-targets and thus minimizes binding events to unintended target sites, and allows for selection of TALEN pairs with a high cutting efficiency. Additionally, it was shown that in human cells DNA repair via HR can take up to 7 h to complete and only occurs in late S and G2 phase, whereas NHEJ works much faster (Mao *et al.*, 2008, Lieber, 2010). Thus, when using an HR-based system as in (Weyman *et al.*, 2015), more time prior to selection might be required in order to directly obtain genetically homogeneous mutants. Nonetheless, we highly recommend re-isolation of knockout strains after the screening process, which requires much less effort than re-isolation of clones prior to the screening process as described in previous studies, in order to avoid the potential risk of mixed cultures.

While CRISPR/Cas9 can be adapted to a specific target site much faster, the use of a TALEN based knockout system has other pronounced benefits. The targeting region of the guideRNA of CRISPR is only 20 bp in length and has been shown to tolerate several mismatched base pairs (Fu *et al.*, 2013, Hsu *et al.*, 2013, Kuscu *et al.*, 2014). TALEN on the other hand is much more robust due to the longer targeting site of two individual TALEN monomers separated by a 12-21bp spacer sequence to allow double strand break formation (Miller *et al.*, 2011). Additionally, TALEN seems less prone to mismatched base pairing indicated by a reduced off-target frequency relative to targeted mutagenesis events compared to CRISPR using a randomized DNA library (Pattanayak *et al.*, 2013, Guilinger *et al.*, 2014a), as has been summarized in (Stella and Montoya, 2016).

4.4.4 Phenotypic characterization of PtAUREO1a knockout mutants

Previously characterized PtAUREO1a knockdown strains generated *via* RNAi showed a significantly reduced chlorophyll *a* (Chl *a*) content per cell compared to the wild type strain under medium light conditions (Qphar 30; 30 µmol photons m^{-2} s^{-1} absorbed photosynthetic radiation) (Schellenberger Costa *et al.*, 2013b). Hence, we determined the Chl *a* content per cell and performed a pigment composition analysis via HPLC (see Table 4-2). Similar to the respective silencing strains, the four bi-allelic PtAUREO1a knockout mutants showed a reduction in Chl *a* content of about 30% compared to the wild type ($p<0.005$, Student's t test)

and an increase of the xanthophyll cycle (XC, diadinoxanthin + diatoxanthin) pool size by >30% in clones 6, 8 and 9 and by 25% in clone 11 ($p<0.001$). Additionally, the ratio of Chl *c* and fucoxanthin relative to Chl *a* was found to be decreased by 10-17% and 5-10% respectively ($p<0.001$). Interestingly, the NPQ capacity was reduced by about 40% in the three knockout lines 6, 8 and 9 despite the increased amount of XC pigments, whereas cell line 11 expressing a truncated PtAUREO1a showed wild type NPQ levels (See Table 4-2; Student's t test, $p<0.001$). NPQ in diatoms depends not only on XC pigments, but also on the expression of the different Lhcx proteins (Bailleul *et al.*, 2010, Lepetit *et al.*, 2017). Hence, in the PtAUREO1a knockout mutants the amount of Lhcx proteins may be reduced. The reduced NPQ in the PtAUREO1a knockouts is in line with previous results, showing that blue light induces higher NPQ values in the wild type than red light (Schellenberger Costa *et al.*, 2013a), and hence stresses the importance of blue light photoreceptors for generating NPQ capacity. However, in PtAUREO1a silencing strains, an increased NPQ capacity had been observed, which was unexpected and the opposite effect of what we could observe in the knockout lines (Schellenberger Costa *et al.*, 2013b). These results highlight that in certain cases stable knockout lines may not only amplify phenotypic effects compared to silencing lines, but may also provide completely different results. Poor correlation between knockdown lines and knockout lines has already been observed in several other model organisms (Gao *et al.*, 2015, Kok *et al.*, 2015, Rossi *et al.*, 2015).

Table 4-2: Analysis of pigment composition and non-photochemical quenching (NPQ) capacity of wild type (wt) and the four bi-allelic PtAUREO1a knockout strains (6, 8, 9 and 11). Chlorophyll a (Chl *a*) was determined photometrically (wild type n=5, bi-allelic mutants n=3), whereas the other pigments were assayed by HPLC (n=2). Pigments were isolated from strains being in mid exponential growth phase and having comparable Chl *a* per culture-volume ratios. NPQ was measured by PAM fluorometry (n=3). Statistical significant differences compared to the wild type were calculated using Student's t-test (*: p<0.01; **: p<0.001). XC pool: xanthophyll cycle pool (diadinoxanthin + diatoxanthin).

	wt	6	8	9	11
Chl *a* [pg cell^{-1}]	0.54±0.02	0.32±0.04**	0.33±0.04**	0.34±0.07*	0.33±0.02**
XC pool [mmol (mol Chl *a*)$^{-1}$]	65.54±0.06	94.90±1.33**	87.36±0.31**	90.40±1.58**	82.33±0.44**
Chl *c* [mmol (mol Chl *a*)$^{-1}$]	143.33±0.14	130.54±1.27**	130.79±1.88**	119.45±0.55**	138.70±0.53**
Fucoxanthin [mmol (mol Chl *a*)$^{-1}$]	566.24±0.56	539.31±0.38**	538.98±2.89*	517.63±1.88**	555.38±0.10*
β-carotene [mmol (mol Chl *a*)$^{-1}$]	53.95±0.18	57.00±0.18*	54.66±0.14	54.25±0.66	53.97±0.01
Non-photochemical quenching	0.476±0.026	0.306±0.005**	0.261±0.001**	0.299±0.001**	0.456±0.017

Despite the involvement of PtAUREO1a in the cell cycle (Huysman *et al.*, 2013), no changes of growth rates in the bi-allelic knockout lines were found at the described cultivation conditions, indicating that a loss of PtAUREO1a can be compensated by the cell. However,

we could detect a difference in the cell morphology of *P. tricornutum*. Length and width of 50 cells per cell line were measured microscopically and the cell volume was estimated by assuming *P. tricornutum* cell shape can be regarded as two joined cones as described in (Rottberger *et al.*, 2013) (cell width, length and volume of each measured cell is shown in Table S4-2, an example for each cell line is shown in Figure S4-2). While cell width was found to be highly similar in all strains, cell length differed significantly between wild type and knockout cell lines (see Figure 4-7A/B). Mono-allelic knockout line 14 as well as clone 11 showed a 10% decrease in cell length, whereas bi-allelic knockout lines 6, 8 and 9 showed a 25% decrease in cell length (One-way ANOVA followed by Tukey's HSD, $p<0.01$). When approximating the volume of fusiform *P. tricornutum*, only the bi-allelic knockout lines showed significant decreases in cell volume compared to wild type cells. While cell volume was reduced by >20% in cell lines 6, 8 and 9, the reduction in cell line 11 was only 13% (see Figure 4-7C, $p<0.01$). Such a reduced cell length and volume had not been observed for the previously characterized PtAUREO1a silencing strains (Schellenberger Costa *et al.*, 2013b), confirming that knockout approaches may yield much more obvious phenotypes than RNAi-based silencing. We also have to note that, with respect to XC-pool content and cell volume, the three bi-allelic knockout lines 6, 8 and 9 showed a very similar phenotype compared to the wild type, while the bi-allelic knockout line 11 expressing a truncated PtAUREO1a protein showed a weaker phenotype (see Table 4-2 and Figure 4-7). Hence, the truncated PtAUREO1a protein in strain 11 seems to still confer some activity.

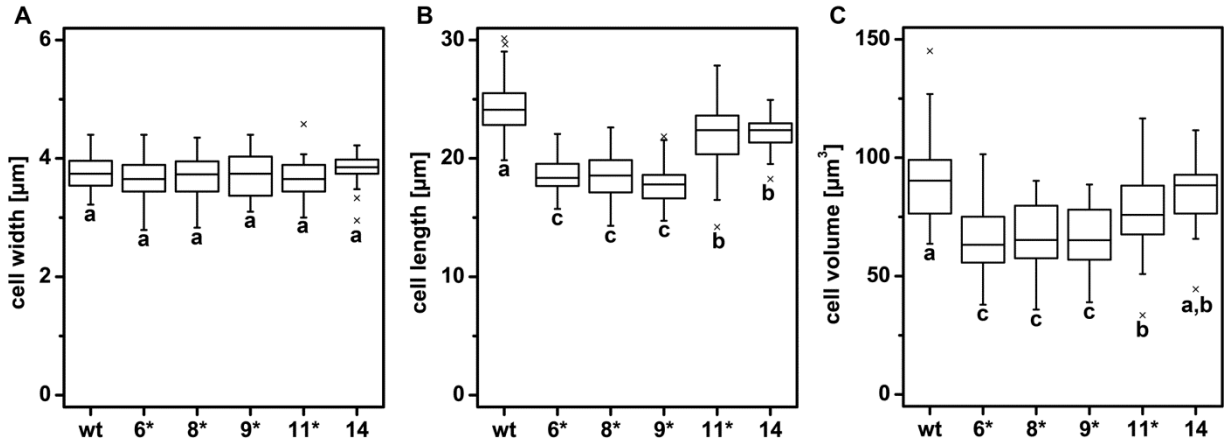


Figure 4-7: 50 cells of the wild type and each mutant strain (6, 8, 9, 11, 14) were examined microscopically and cell width (A) and length (B) was measured. Cell volume was approximated as two cone shapes (C). The bi-allelic knockout mutants are marked with a *. Whiskers represent the outermost data point within the 1.5-fold of the interquartile range; outliers are represented by an x. Different letters represent statistical significant differences in cell width, length or volume between strains (One-way ANOVA, followed by Tukey's HSD, $p<0.01$)

4.5 Conclusion

An efficient TALEN system has been established for the diatom *P. tricornutum*, which leads to a high frequency of targeted mutation events and yields mainly genetically homogeneous cell lines. Additionally, all important steps, from target site design, TALEN construction using an easy-to-follow and publically available modular construction system, up to screening of the mutants, have been addressed to avoid the potential pitfalls for knockout generation in *P. tricornutum*, like potential off-targets and inefficient TALEN proteins. The TALEN system presented here may expand the application of reverse genetics approaches for creation of *P. tricornutum* knockout mutants to a broader scientific community.

4.6 Acknowledgements

The authors are grateful to Doris Ballert for the help with the cultivation and transformation of *P. tricornutum*, Dr. Daniela Ewe for cloning of pPha-NR-Nat, as well as Simon Kienle and Florian Weeber for help in mutant screening. This work was supported by the University of Konstanz, the graduate school Biological Sciences (GBS), the EU FP7 Marie Curie Zukunftskolleg Incoming fellowship (grant no. 291784, to BL), the Gordon and Betty Moore Foundation GBMF 4966 (grant DiaEdit to PGK) and the DFG (grants KR 1661/8-2 to PGK, LE 3358/3-1 to BL).

5 Loss of PtAUREO1a results in blocked short term adaptation to blue light

Serif, M.[1], Mann, M.[2], Wilhelm, C.[2], Kroth, P.G.[1]

*Corresponding Author: manuel.serif@uni-konstanz.de

[1]Plant Ecophysiology, Fachbereich Biologie, Universität Konstanz, D-78457 Konstanz, Germany

[2]Institute of Biology, Department of Plant Physiology, University of Leipzig, D-04103 Leipzig, Germany

Keywords: *Phaeodactylum tricornutum*, TALEN, PtAUREO1a knockout, blue light-dependent transcription factor, photoacclimation

5.1 Abstract

Aureochromes are a novel type of blue-light dependent transcription factors restricted to the Stramenopiles. Four isoforms are encoded within the genome of the diatom model organism *Phaeodactylum tricornutum*, seemingly performing partially non-overlapping functions. While their biophysical properties are well characterized, very little is known about their biological function. Previously generated PtAUREO1a knockdown strains showed increased photosynthesis rates, a higher NPQ capacity, as well as a reduced Chl *a* content per cell whereas a decreased NPQ capacity was observed for TALEN-based knockout strains due to strongly reduced level of Lhcx1. Interestingly, a mono-allelic knockout strain, which reverted to an overexpression strain showed a phenotype similar to the knockdown strain, indicating that the level for PtAUREO1a needs to be strictly controlled within the cell. To identify genes directly regulated by PtAUREO1a, wild type and a knockout strain were cultured under red light and shifted to blue light for 10 and 60 min, respectively. Then, transcript levels were analyzed by RNA-seq. While a rearrangement of 85% of the transcriptome was detected for the wild type cells, less than 10% of transcripts were found to be regulated in the knockout strain. While all other photoreceptors were found to be expressed in the knockout strain, blue-light treatment only led to very minor changes in transcription levels of these photoreceptors in comparison to the wild type. Furthermore, a strong influence of PtAUREO1a on the regulation of transcription factors, especially on the classes of bZIP and zinc finger transcription factors was found not just under blue-light conditions but under red-light conditions as well, indicating PtAUREO1a to be active in absence of blue light as well. We therefore propose PtAUREO1a to be a master regulator for light acclimation and photoprotection, which most likely functions as a dual regulator with different activity in absence and presence of blue light.

5.2 Introduction

Diatoms, which belong to the Stramenopiles, are a group of unicellular microalgae that play an important role in global carbon fixation as well as for the nitrogen, phosphorous and silica cycles (Yool and Tyrrell, 2003). They are widespread in most aquatic habitats, where they need to cope with large variations of light quality and quantity (MacIntyre *et al.*, 2000, Ragni and D'Alcalà, 2004). Hence, sensing these parameters, either indirectly via retrograde signaling (Lepetit and Dietzel, 2015) or directly via photoreceptors, i.e., Cryptochromes,

Aureochromes, or Phytochrome (Depauw *et al.*, 2012), is of high importance to adapt to dynamic light conditions. The adaptation to higher light conditions has been shown to be mainly dependent on blue light in *P. tricornutum* (Schellenberger Costa *et al.*, 2013a).

Aureochromes are a novel class of blue-light dependent photoreceptors which first were identified in 2007 in the Xanthophyte *Vaucheria frigida*, where the two homologs, Aureo1a and Aureo2, are involved in the induction of branching and development of sexual organs, respectively (Takahashi *et al.*, 2007). Since then, Aureochromes have been identified in many other Stramenopiles and seem restricted to this diverse group of algae (Ishikawa *et al.*, 2009). Aureochromes possess two conserved domains: A C-terminal LOV (light oxygen voltage) domain, which is a blue-light sensing domain that binds FMN as a cofactor, and an N-terminal bZIP (basic region leucine zipper) domain, which is a DNA-binding domain (Takahashi *et al.*, 2007). Thus, they are thought to be blue-light dependent transcription factors. The genome of the diatom *P. tricornutum* encodes four different Aureochromes, which were originally classified due to their homology to the two Aureochromes of *V. frigida*: PtAUREO1a (JGI Protein ID: 49116), PtAUREO1b (49458), PtAUREO1c (56742) and PtAUREO2 (56060). However, using additional sequences of other organisms to construct a phylogenetic tree, a more complex picture evolved with four distinct clades of Aureochromes, of which *P. tricornutum* possesses one of each (Schellenberger Costa *et al.*, 2013b). While the biophysical properties of Aureochromes has been studied in great detail (Herman *et al.*, 2013, Herman and Kottke, 2015, Banerjee *et al.*, 2016a, Banerjee *et al.*, 2016b, Heintz and Schlichting, 2016), very little is known about their physiological function: Using Yeast 2 Hybrid assays, it could be shown that PtAUREO1a is involved in the cell cycle by interacting with the diatom-specific cyclin 2 (*dsCYC2*) promoter (Huysman *et al.*, 2010). RNAi-mediated PtAUREO1a knockdown strains showed a “hyper” high-light acclimated phenotype under low light conditions in form of a decreased Chl *a* content, as well as increased xanthophyll cycle pool and higher non-photochemical quenching (NPQ) capacity, providing further indications of partially non-overlapping functions of the Aureochrome isoforms (Schellenberger Costa *et al.*, 2013b). To provide more insight into the function of PtAUREO1a, we studied the recently generated TALEN-mediated PtAUREO1a knockout strains (Serif *et al.*, 2017) in more detail. Here, we compared the short-term reaction of the transcriptome to a shift from red to blue light with the wild type to identify genes directly regulated by PtAUREO1a activity.

5.3 Material & Methods

5.3.1 Cultivation and harvesting of cells for RNA-seq

Wild type and PtAUREO1a knockout strain 8 (Serif *et al.*, 2017) were cultured semi-continuously as described in (Schellenberger Costa *et al.*, 2013a). Light intensity of blue and red light was adjusted to yield the same amount of photosynthetically absorbed radiation (Qphar) according to (Gilbert *et al.*, 2000). Cells were adapted to 14h/10h light/dark red light (Qphar 10) for at least one week and were then shifted to blue light (Qphar 10). Samples were taken at timepoint t0 (red light) as well as 10 min and 60 min of blue light. 50 ml of cells were filtered onto 1.2 µm RTTP polycarbonate filters (Merck-Millipore, Darmstadt, Germany), flash-frozen in liquid nitrogen and stored at -80°C until further use. Culturing of cells was performed in cooperation with Marcus Mann and Christian Wilhelm (Universität Leipzig, Germany)

5.3.2 Sample preparation for RNA-seq

RNA was extracted from the filtered cell using PeqGOLD RNApure and PeqGOLD RNA extraction kit (VWR, Darmstadt, Germany). Contamination with residual DNA was removed by an on-column digest with the PeqGOLD DNase I digest kit (VWR). Extracted RNA was quality-controlled spectrophotometrically, by separation on agarose gel as well as on a 2100 bioanalyzer (Agilent, Waldbronn, Germany). cDNA libraries were prepared from 1 µg RNA using the TruSeqTM RNA Sample Prep Kit v2 (Illumina Inc., San Diego, USA). cDNA library generation and sequencing was performed with the Illumina HiSeq2000 in the paired end mode in cooperation with Marion Eisenhut and Andreas Weber (Universität Düsseldorf).

5.3.3 Bioinformatical analysis of the sequencing data

Transcript abundances were determined by mapping the reads against the *P. tricornutum* CCAP 1055/1 genome (Release 27, downloaded June 14, 2015) in Ensembl (http://protists.ensembl.org/Phaeodactylum_tricornutum/Info/Index, cDNA) using the CLC genomics workbench (https://www.qiagenbioinformatics.com/) with default parameters in cooperation with Marion Eisenhut and Andreas Weber (Universität Düsseldorf). The resulting table was analysed in R ((R Development Core Team, 2017), www.r-project.org,). Before principal component analyses, expression was normalized to tpkm (transcripts per kilobase

million) (Wagner *et al.*, 2012). Differential expression was determined using edgeR (Robinson *et al.*, 2010) on raw counts in GLM Fit mode with subsequent common and tagwise dispersion estimation. P-values were Benjamini-Hochberg corrected before further analysis.

5.3.4 Protein Isolation and Immunoblotting

Protein extraction and subsequent separation of proteins was done as described in (Serif *et al.*, 2017). After blotting, the nitrocellulose membrane was cut between 40 and 50 kDa and the top part (45-250 kDa) was used to detect the RbcL loading control, whereas the bottom half (0-30 kDa) was used to detect the Lhcx proteins. Immunoblots using an antibody specific against Lhcx proteins were done as described in (Lepetit *et al.*, 2017), whereas the RbcL-specific antibody (AS03-037, Agrisera AB, Vännas, Sweden) was used according to the manufacturer's instructions. Blots were developed using an Odyssey FC Imaging System (Li-Cor, Bad Homburg, Germany).

5.4 Results

5.4.1 Lhcx1 levels are strongly reduced in PtAUREO1a knockout strains

Previous studies using TALEN-mediated PtAUREO1a knockout strains revealed reduced Chl *a* content per cell, an increased xanthophyll cycle (XC) pigment pool size, and a decreased non-photochemical quenching (NPQ) capacity when grown in low intensity white light (Serif *et al.*, 2017). As NPQ in diatoms requires a transthylakoidal proton gradient, the conversion of the XC pigment diadinoxanthin to diatoxanthin and the presence of Lhcx proteins, it was speculated that the level of Lhcx1, the most prominent isoform under low light conditions, might be affected. Thus, we performed Western Blots against Lhcx proteins for both the three bi-allelic knockout strains without residual PtAUREO1a protein (strains 6, 8, 9), a bi-allelic knockout strain expressing a truncated version of PtAUREO1a (strain 11) as well as a mono-allelic knockout strain with 20% increased protein level compared to the wild type (strain 14, see Figure S5-1). Lhcx1 was found to be reduced by about 35% in strains 6, 8 and 9, which is consistent with the about 40% decrease in NPQ observed, whereas the two other strains showed a 20% increase in Lhcx1 levels. Correspondingly, NPQ was found to be increased by 25% in strain 14, whereas Chl *a* per cell was similarly reduced as in the

knockout strains (see Table 5-1). Thus, the overexpressing strain performs mostly like the previously published RNAi knockdown strains (Schellenberger Costa *et al.*, 2013b). The contrasting phenotype of knockout and overexpression strain, however, is a strong indication that the reduced NPQ is due to knockout of PtAUREO1a and not due to an off-target effect of the TALEN constructs.

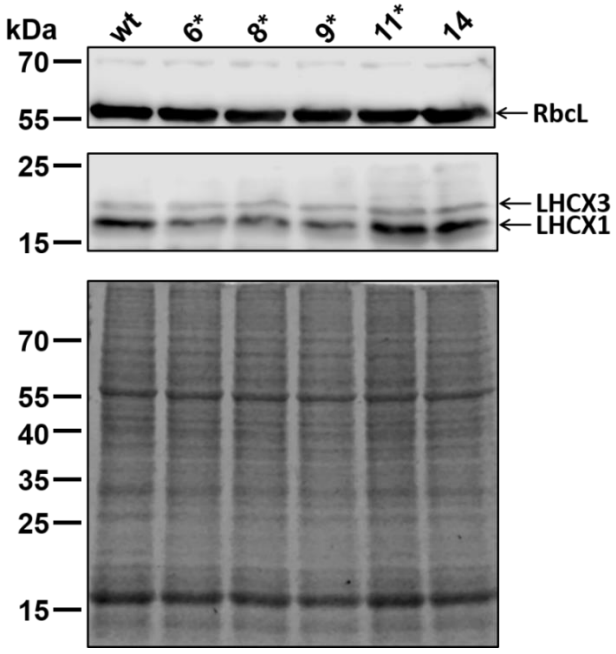


Figure 5-1: Western Blot against Lhcx proteins for the wild type (wt), four bi-allelic PtAUREO1a knockout strains (6, 8, 9, as well as strain 11 which expresses a truncated PtAUREO1a protein) and a mono-allelic knockout strain with 20% increased PtAUREO1a level (upper panel). Expression level of the large subunit of rubisco (RbcL) is used as a loading control. A duplicate coomassie gel is shown as additional loading control (lower panel).

Table 5-1: Comparison of Chl *a* content per cell and non-photochemical quenching (NPQ) capacity for the wild type (wt), four bi-allelic knockout strains (6, 8, 9, as well as strain 11 which expresses a truncated PtAUREO1a protein) and a mono-allelic knockout strain with 20% increased PtAUREO1a protein level. Statistical significant differences compared to the wild type strain were calculated using Student's t-test (*: p<0.01; **: p<0.001).

	wt	6*	8*	9*	11*	14
Chl *a* [pg cell^{-1}]	**0.54±0.02**	**0.32±0.04****	**0.33±0.04****	**0.34±0.07***	**0.33±0.02****	**0.39±0.02****
NPQ	**0.476±0.026**	**0.306±0.005****	**0.261±0.001****	**0.299±0.001****	**0.456±0.017**	**0.592±0.017***

5.4.2 Transcriptional characterization of a red to blue light shift

To further investigate the role of PtAUREO1a and to identify genes directly regulated by this transcription factor we performed an RNA-seq experiment: The wild type and the knockout (strain 8) were first grown semi-continuously under red light conditions, and then shifted to blue light. The light intensity was adjusted to yield the same amount of photosynthetically absorbed radiation (Qphar 10 µmol photons $m^{-2}s^{-1}$). Samples were taken at t0 (red light) as

well as after 10 min and 60 min of blue light illumination to distinguish both primary and secondary effects of PtAUREO1a. This cultivation was done in cooperation with Marcus Mann and Christian Wilhelm (Universität Leipzig, Germany). Three biological replicates of each shift were chosen for sequencing. The sequences obtained from Illumina Sequencing were mapped to the Phatr3 database (Ensemble Protists) and differential expression was calculated using the R-based EdgeR (Robinson *et al.*, 2010) package (FDR-corrected p value < 0.01). A principal component analysis (PCA) showed that the biological replicates of each sample cluster strongly together (see Figure 5-2). While the three different timepoints of the wild type appear strongly separated from each other, indicating strong changes in the transcriptome, the different timepoints of the knockout strain clustered very close to each other, indicating much smaller changes upon the blue light treatment. Additionally, the knockout samples clustered closest to the wild type adapted to red light, indicating that the differences between the strains were much more pronounced upon blue light treatment.

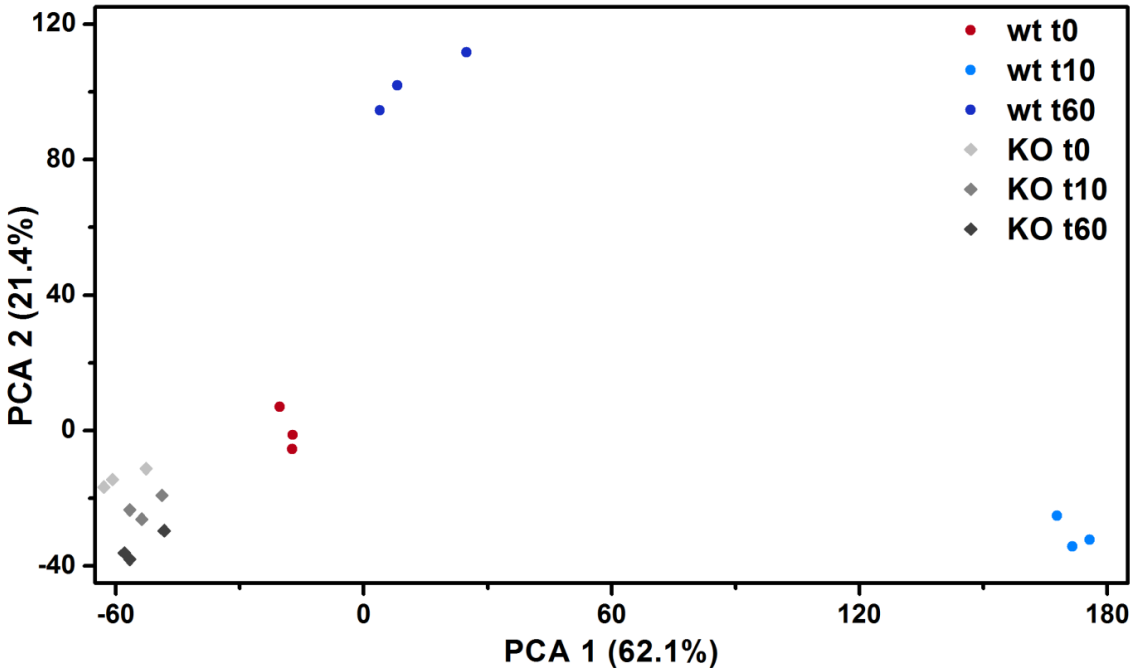


Figure 5-2: Principal component analysis (PCA) of the transcriptome data after mapping. The bigger the distance between data points the higher the differences between the corresponding samples. All replicates cluster closely together. Strong changes happen in the wild type (circles) upon illumination with blue light, whereas there seem to be much less changes in the knockout strain (diamonds).

A total of 10854 genes were found to be sufficiently expressed for analysis, of which less than 1400 genes (12.9%) seemed not to be regulated by the light treatment (see Table 5-2 for a summary, a table showing all transcripts and their regulation patterns is found on the attached CD-ROM in the Supplemental dataset S5-1). In the wild type, there is a strong initial reaction to the blue light treatment, causing almost 39% of genes to be significantly downregulated

and 35% to be significantly upregulated compared to the red-light conditions. The 60 min blue-light treatment led to a slightly weaker change in the transcriptome with almost 30.6% genes downregulated and 30.1% upregulated compared to the red-light conditions. In the PtAUREO1a knockout mutant, however, this strong reaction to the blue-light treatment was not observed. Instead, less than 3% of genes were either up- or downregulated after 10 min of blue light treatment compared to the red-light conditions, whereas about 3.2% of genes were upregulated and 5.7% downregulated after 60 min of blue light.

Table 5-2: Percentage of transcripts which are not regulated, upregulated or downregulated in the wt (wild type) or the PtAUREO1a strain 8 (KO) after 10 or 60 min of blue-light treatment (BL) compared to the red-light condition (RL), respectively, as well as the differences between wt and KO under RL conditions. Percentages correspond to a total of 10854 genes sufficiently expressed for analysis.

	wt t0 → t10	wt t0 → t60	KO t0 → t10	KO t0 → t60	t0 wt → KO
No regulation	26.16%	39.36%	97.33%	91.1%	72.44%
Upregulation	34.94%	30.07%	2.38%	3.22%	12.14%
Downregulation	38.88%	30.57%	0.29%	5.68%	15.42%

To better visualize the changes, UpSet plots were generated (see Figure 5-3A/B). Of the 3792 significantly upregulated transcripts after 10 min of blue-light treatment, 1701 were still significantly upregulated after 60 min of blue-light treatment. In contrast, 1386 transcripts were found not to be significantly different from the red-light condition anymore, whereas 705 were found to be downregulated after 60 min compared to the red-light condition. A highly similar result could be seen for the 4220 transcripts downregulated after 10 min, of which 1775 remained downregulated after 60 min of blue light, 1058 were not significantly different from the red-light condition anymore and 937 became upregulated compared to the red-light conditions. Thus, reactions to the blue-light treatment can be divided into several categories: (A) Transcripts that are upregulated after both 10 and 60 min, (B) transcripts that are downregulated after both 10 and 60 min, (C) transcripts that are only upregulated in the 10 min blue light condition, (D) transcripts that are only downregulated in the 10min blue light condition, (E) transcripts that are only upregulated after 60 min, (F) transcripts only downregulated after 60 min, (G) transcripts that are upregulated after 10 min and downregulated after 60min blue-light treatment and (H) transcripts that are downregulated after 10min and upregulated after 60min of blue-light treatment. In the PtAUREO1a knockout strain, however, not only the total number of regulated genes was much lower: While transcripts of categories A to F could be found, no transcripts could be identified which were regulated in different directions in the 10 min and 60 min blue-light treatment in comparison to the red-light condition.

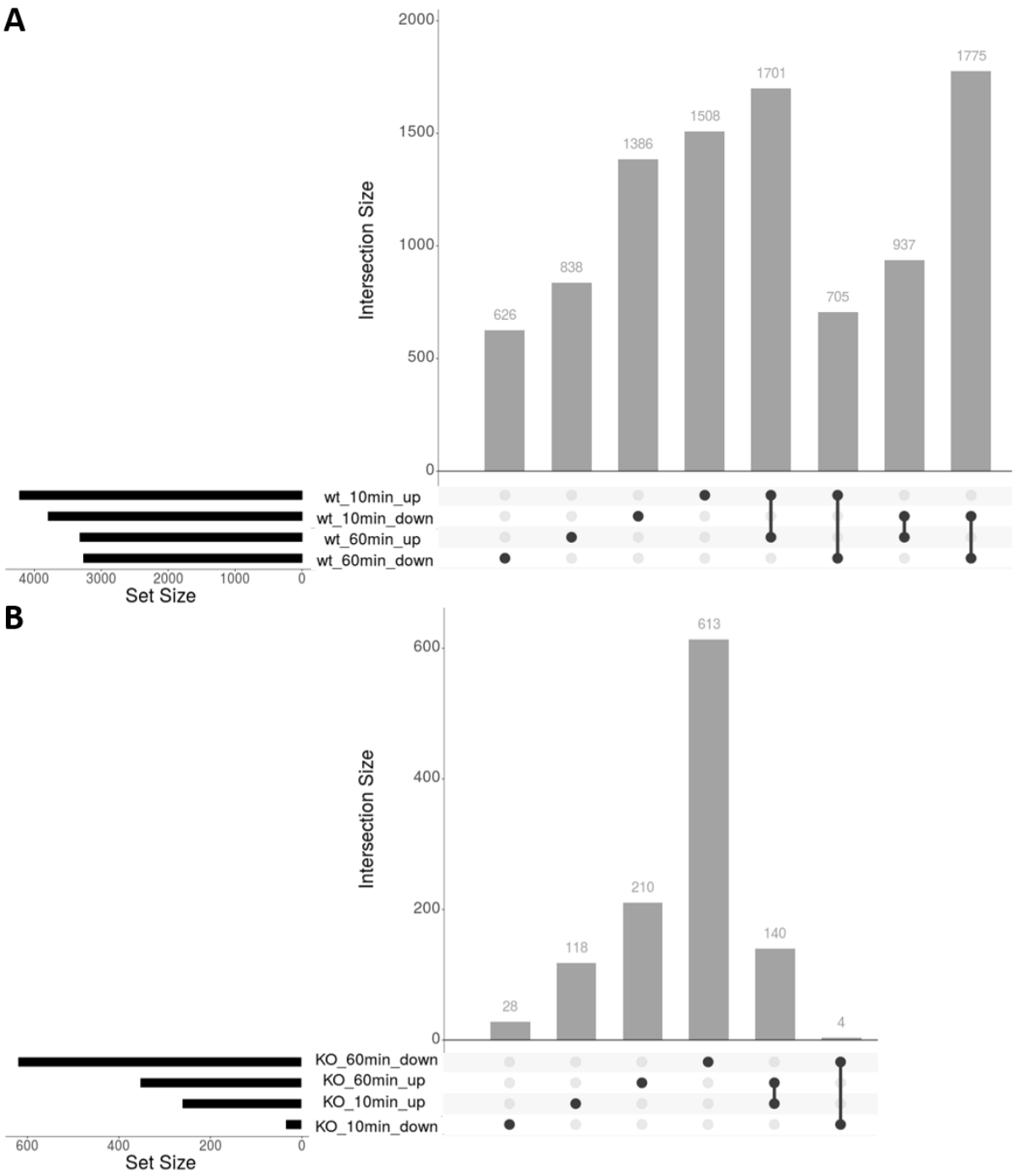


Figure 5-3: Visualization of overlap between significantly up- or downregulated transcripts for the different timepoints. A: Overlap of genes up- or downregulated after 10 and 60 min of blue-light treatment in the wild type (wt) relative to the red-light condition. B: Overlap of genes up- or downregulated after 10 and 60 min of blue-light treatment in the PtAUREO1a knockout strain (KO) relative to the red-light condition. The black horizontal bar graph shows the total number of genes significantly up- or downregulated at either 10 min or 60 min of blue-light treatment. The grey vertical bar graph shows the number of transcripts regulated significantly in one or more sets. The combination matrix below indicates the respective overlap between sets using black dots: A single black dot indicates significantly regulated transcripts only found in this set, whereas two connected black dots indicate transcripts regulated significantly in those two sets, but not in the other sets. Combinations of sets which do not share significantly regulated transcripts were omitted for clarity. Graphs were created using the Intervene web application (Khan and Mathelier, 2017).

Light acclimation was shown to be dependent on blue light in *P. tricornutum* and is strongly affected in both the PtAUREO1a knockdown and the knockout strain (Schellenberger Costa *et al.*, 2013a, Schellenberger Costa *et al.*, 2013b, Serif *et al.*, 2017). As the knockout strain seemed unable to react to the blue-light treatment, we had a closer look on the transcript changes of the photoreceptor genes found in *P. tricornutum* as well as the Lhcx genes, which play a major role in the photoprotective NPQ mechanism (Bailleul *et al.*, 2010). Figure 5-4A shows the relative changes of transcript abundance of the four Aureochrome genes. Interestingly, all four Aureochromes showed a strong reaction to the blue-light treatment in the wild type, with strong downregulation of *PtAureo1a*, *PtAureo1b* and *PtAureo2* after 10 and 60 min of blue-light treatment, whereas *PtAureo1c* was strongly upregulated after 10 min of blue-light treatment. In the PtAUREO1a knockout strain, on the other hand, transcription of the Aureochrome genes did not react to the blue-light treatment anymore, resulting in a strong deregulation of these photoreceptors. While the strongly reduced transcript levels of *PtAureo1a* in the PtAUREO1a knockout strain in comparison to the wild type might be an artifact due to the mutations caused by the TALENs, transcripts of *PtAureo1c* were found to be reduced almost 4-fold throughout the whole experiment in comparison to the wild type. A similar result was obtained for the other photoreceptors found in *P. tricornutum*, including four Cryptochromes (*CryP*, *CFP1*, *CPF2* and *CPF4*) as well as a Phytochrome (*DPH1*) (see Figure 5-4B). Transcripts of *CryP*, *CPF1* and *CPF2* were found to be strongly increased after 10 min of blue-light treatment in the wild type. While transcript levels of *CPF1* remained high after 60 min of blue-light treatment as well, transcription of *CPF2* decreased again to a level similar to the red-light state and *CryP* was found to be approximately 32-fold downregulated compared to the 10 min blue-light treatment and more than 8-fold downregulated compared to the initial red-light state. *CPF4*, on the other hand, was found to be slightly downregulated after 10 min of blue-light treatment. Interestingly, the red-light sensing Phytochrome was found to be regulated by the blue-light treatment, with a 4-fold reduction in transcript levels after 60 min. In the knockout strain, *CPF1* and *CPF2* were found to be regulated similarly to the wild type, however, to a much lower extent. *CryP* was found to be upregulated after 10 min, but remained upregulated, strongly contrasting the pattern of the wild type, whereas *CPF4* was found to be not regulated at all. For *DPH1*, no downregulation could be observed after 60 min of blue-light treatment.

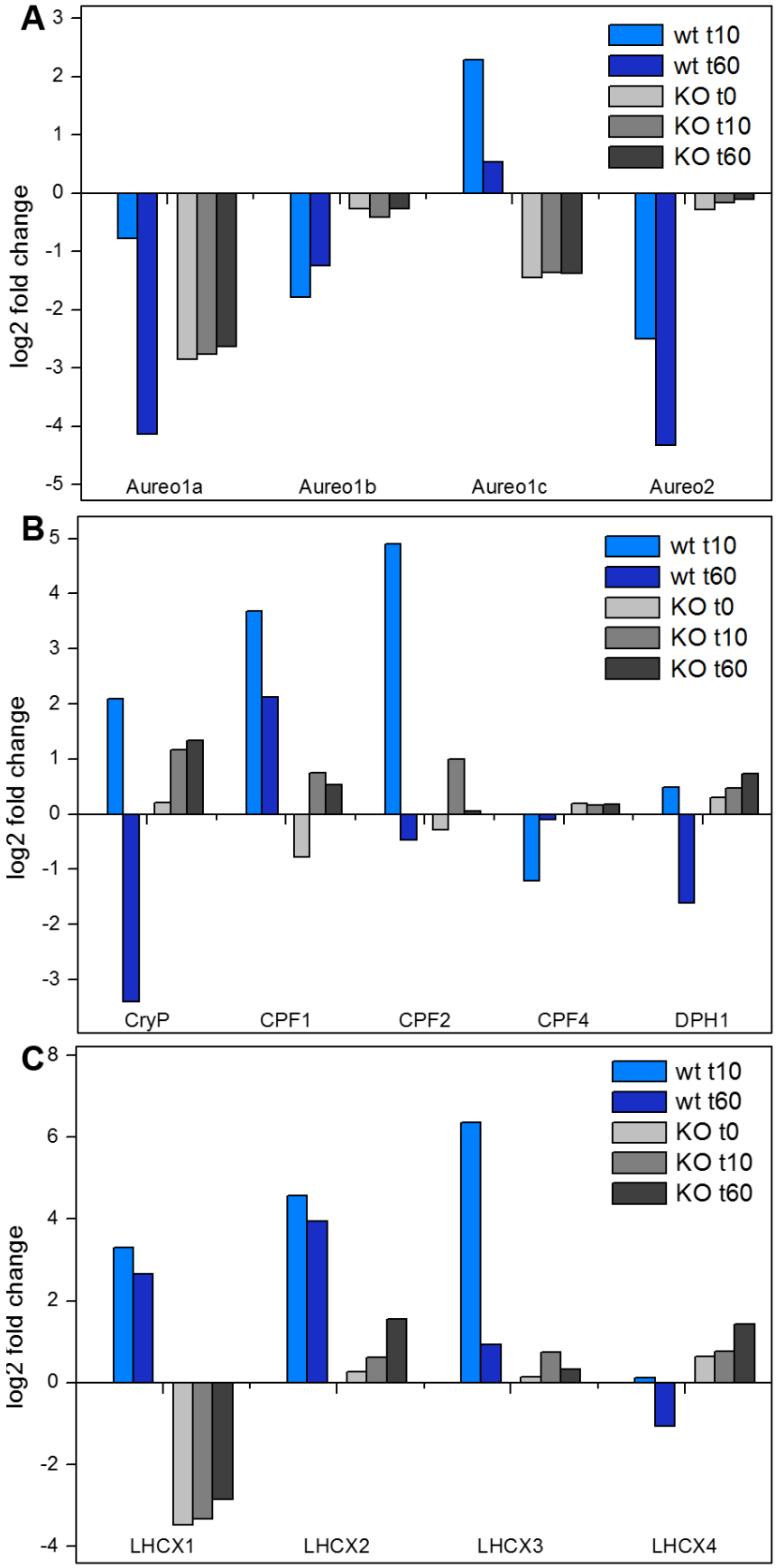


Figure 5-4: Changes in transcript levels upon shift to blue light in wild type (wt) and PtAUREO1a knockout strain (KO) of Aureochromes (panel A), other photoreceptors (panel B) and Lhcx isoforms (panel C) upon illumination with blue light, depicted as log2-fold change relative to wt t0 (red light).

The transcriptional changes observed for the *Lhcx* genes are shown in Figure 5-4C. In the wild type, a strong increase in transcript abundance was observed for *Lhcx1* (8-fold), *Lhcx2* (16-fold) and *Lhcx3* (64-fold) after 10 min of blue-light treatment, which remained increased for *Lhcx1* and *Lhcx2* after 60 min of blue light as well, whereas *Lhcx4* was found to be downregulated 2-fold after 60 min of blue-light treatment. In the PtAUREO1a knockout strain, however, transcript levels of *Lhcx1* were found to be reduced about 8-fold regardless of light treatment, whereas *Lhcx2*, *Lhcx3* and *Lhcx4* were found to be regulated much less pronounced compared to the wild type. This result fits very well to be strongly reduced protein level of *Lhcx1* and the reduced level of NPQ, which were observed for all three bi-allelic PtAUREO1a knockout strains when cultured in white light (see Figure 5-1 and Table 5-1). In conclusion, transcription of all photoreceptors and *Lhcx* genes were found to be strongly influenced by blue light in the wild type, whereas little regulation of the transcript levels of these genes was detectable in the PtAUREO1a knockout strain, indicating it to be a master switch for the light acclimation process.

As most photoreceptors were expressed similarly in wild type and knockout strain for the red-light treatment, they cannot be responsible for the initial "blindness" to blue light alone. Aureochromes are the only group of photoreceptors present in *P. tricornutum*, which can directly influence transcription levels. Thus, other photoreceptors require interaction with downstream factors, e.g., kinases or transcription factors, for their signal transduction. The genome of *P. tricornutum* encodes a total of 212 known transcription factors (Rayko *et al.*, 2010). Blue-light treatment had a very strong effect on the transcription level of most these transcription factors (see Table 5-3 for a summary, the complete list of transcription factors and their regulation patterns is found on the attached CD-ROM in the Supplemental dataset S5-2). 43% of them were found to be upregulated after 10 min while 33.6% were found to be downregulated compared to red light. Prolonged blue-light treatment resulted about 27% upregulated and 32.6% downregulated compared to the red-light treatment, slightly less pronounced than the 10 min blue-light treatment. Thus, a short term and a longer term regulation can be seen in the wild type, similarly to what was observed for the whole transcriptome. In the knockout strain, however, only 8.5% of transcripts were found to be upregulated and only about 1% of transcripts were found to be downregulated after 10 min of blue light. The prolonged blue-light treatment led to a slightly more pronounced change, with about 7.1% of transcripts found to be upregulated and 5.2% downregulated after 60 min. As the response of transcription factor transcripts to blue light was strongly impaired, we

additionally compared the transcript levels of wild type and knockout strain for the initial red-light sample to see whether some of them were not expressed or differentially expressed. While transcripts of all transcription factors found in the wild type under red light were also found in the knockout strain, about 19.3% were found to be downregulated in the knockout strain whereas 8.5% were found to be upregulated, indicating that loss of PtAUREO1a has strong effects on the transcript level of transcription factors in absence of blue light as well.

Table 5-3: Percentage of transcripts of 212 known *P. tricornutum* transcription factors which are not regulated, upregulated or downregulated in the wt (wild type) or the PtAUREO1a strain 8 (KO) between timepoints t0, t10 and t60, as well as between the t0 condition of both strains. A complete list with regulation patterns for each transcription factor is found on the attached CD-ROM in the Supplemental dataset S5-2.

	wt t0→ t10	wt t0→ t60	KO t0→ t10	KO t0→ t60	t0 wt→ KO
No regulation	23.59%	40.56%	90.57%	87.73%	72.17%
Upregulation	42.92%	26.89%	8.49%	7.08%	8.49%
Downregulation	33.59%	32.55%	0.94%	5.19%	19.34%

To identify whether certain groups of transcription factors were influenced more by the loss of PtAUREO1a than others, we examined the four biggest classes of transcription factors found in *P. tricornutum* in more detail: HSF (heat shock factors, 70 isoforms), zinc finger (40 isoforms), Myb (34 isoforms) and bZIP (basic region leucine zipper, 23 isoforms) transcription factors, corresponding to almost 79% of all transcription factors found in *P. tricornutum* (Rayko *et al.*, 2010). In the wild type, the shift from red to blue light led to a significant up- or downregulation of most of the members of these transcription factor classes after 10 min (see Table 5-4). The prolonged blue-light treatment resulted in more than 50% of transcripts of all major transcription factor classes to be up- or downregulated compared to the red-light conditions. HSFs were affected the strongest with over 54% of transcripts downregulated and 17.1% upregulated. In the knockout strain, however, almost no downregulation took place after 10 min of blue-light treatment and upregulated transcripts were also strongly reduced. Strikingly, upregulation of HSFs accounted for over 76% of the regulated transcripts observed, and no regulation of bZIP or zinc finger domain proteins, which together account for almost 30% of all transcription factors found in *P. tricornutum*, was detectable. Prolonged blue-light treatment resulted in almost no changes compared to the 10 min blue-light treatment with the exception of downregulation of three zinc finger transcripts. While regulation of all major transcription factor classes was found to be influenced by the knockout of PtAUREO1a, the strongest effect was detected for zinc finger and bZIP transcription factors, whose regulation upon blue-light treatment was almost

completely abolished. Furthermore, when the red-light conditions of both strains were compared, all groups were affected to a similar extent, but downregulation was much more pronounced for HSF and Myb transcripts compared to the other classes, where up- and downregulated transcripts were similarly distributed. Thus, PtAUREO1a seems to be not only active under blue-light conditions. Similarly, silencing of PtAUREO1a had been shown to cause not only a phenotype under blue-light conditions but under red-light conditions as well, resulting in increased photosynthesis rates, NPQ and decreased Chl *a* content per cell (Schellenberger Costa *et al.*, 2013b).

Table 5-4: Percentage of transcripts of major transcription factor classes, which are upregulated or downregulated in the wt (wild type) or the PtAUREO1a strain 8 (KO) between timepoints t0, t10 and t60, as well as between the t0 condition of both strains. HSF: heat shock factor; Zf: zinc finger; bZIP: basic region leucine zipper. A complete list with regulation patterns for each transcription factor is found on the attached CD-ROM in the Supplemental dataset S5-2.

	wt t0→ t10	wt t0→ t60	KO t0→ t10	KO t0→ t60	t0 wt→ KO
HSF no regulation	13 (18.6%)	22 (31.5%)	56 (80%)	56 (80%)	52 (74.3%)
HSF upregulation	28 (40%)	12 (17.1%)	13 (18.6%)	12 (17.1%)	5 (7.1%)
HSF downregulation	29 (41.4%)	36 (54.4%)	1 (1.4%)	2 (2.9%)	13 (18.6%)
Zf no regulation	11 (27.5%)	18 (45%)	40 (100%)	37 (92.5%)	28 (70%)
Zf upregulation	12 (42.5%)	13 (32.5%)	0 (0%)	0 (0%)	5 (12.5%)
Zf downregulation	17 (30%)	9 (22.5%)	0 (0%)	3 (7.5%)	7 (17.5%)
Myb no regulation	12 (35.3%)	16 (47%)	31 (91.2%)	31 (91.2%)	25 (73.6%)
Myb upregulation	12 (29.4%)	11 (32.4%)	2 (5.9%)	2 (5.9%)	1 (2.9%)
Myb downregulation	10 (35.3%)	7 (20.6%)	1 (2.9%)	1 (2.9%)	8 (23.5%)
bZIP no regulation	5 (21.7%)	11 (47.8%)	23 (100%)	23 (100%)	15 (65.2%)
bZIP upregulation	13 (56.6%)	2 (8.7%)	0 (0%)	0 (0%)	4 (17.4%)
bZIP downregulation	5 (21.7%)	10 (43.5%)	0 (0%)	0 (0%)	4 (17.4%)

5.5 Discussion

To characterize the function of PtAUREO1a, both RNAi-based knockdown and TALEN-mediated knockout strains had been generated previously. While both approaches resulted in a similar phenotype regarding Chl *a* content per cell, they behaved oppositely in their effect on the NPQ level. While NPQ was found to be increased in the knockdown strain, it was decreased in the knockout strain (Schellenberger Costa *et al.*, 2013b, Serif *et al.*, 2017). The comparison, however, could not be done in similar light conditions as the knockdown strain had reverted to wild type level of protein expression (data not shown). This behavior was, however, not only observed for the knockdown strain but also for a mono-allelic knockout strain, which over time turned into a slight overexpression strain (see Figure S5-1).

Interestingly, this strain showed a decrease in *Chl a* content per cell as well as an increased level of NPQ (see Table 5-1) similar to what had been observed for the RNAi-based knockdown strain (Schellenberger Costa *et al.*, 2013b). This finding indicates that the reduction of NPQ found in the knockout strains might not be due to an off-target effect of the TALEN system, but instead caused by the complete loss of PtAUREO1a. Knockout strains not mimicking the phenotypes of knockdown strains has been observed several times, however, the mechanism of this effect is unknown (Gao *et al.*, 2015, Kok *et al.*, 2015, Rossi *et al.*, 2015). Additionally, similar phenotypes of knockdown and overexpression strains are often observed when regulatory proteins like kinases, GTPases or transcription factors are targeted which may require hetero-oligomer formation with other proteins to perform their function (Bernick *et al.*, 2010, Park *et al.*, 2010). Thus, expression of PtAUREO1a needs to be tightly controlled to perform its intended function. This could be another indication that PtAUREO1a acts not only as a homodimer but also as a heterodimer.

Whereas much is known about the phenotype caused by PtAUREO1a deficiencies, very little is known about which genes are directly regulated by PtAUREO1a. The only gene identified so far is the diatom-specific cyclin 2 (*dscyc2*), which controls the onset of the cell cycle after dark arrest (Huysman *et al.*, 2013). Therefore, we performed an RNA-seq experiment with the wild type and one PtAUREO1a knockout strain based on a shift from red light to blue light of the same intensity (Qphar 10) with sampling after 10 and 60 min of blue-light treatment. This shift resulted in a major rearrangement of about 85% of the transcriptome among all conditions in the wild type (see Table 5-2 and Figure 5-3). While some transcripts were regulated in a similar manner after 10 and 60 min of blue-light treatment in comparison to the red-light condition, others were found to be only up- or downregulated at one of the timepoints or inversely regulated for the different timepoints, indicating that there is a short-term reaction and a long-term reaction to the blue-light stimulus. In the knockout strain, however, only about 10% of the transcripts were found to be affected by the blue-light treatment. This effect is much stronger than we anticipated, as 212 known transcription factors and nine different photoreceptors, of which eight are blue-light sensitive, are encoded within the genome of *P. tricornutum* (Rayko *et al.*, 2010, Depauw *et al.*, 2012). Thus, we examined whether these photoreceptors are expressed in the knockout strain. While some were found to be slightly downregulated under red-light conditions, all other photoreceptors seemed to be expressed in the knockout strain. Hence, a lack of photoreceptors other than PtAUREO1a cannot be used to explain the initial “blindness” to 10 min of blue-light

treatment. There is, however, bioinformatical evidence that PtAUREO1a might interact with the PtAUREO isoforms 1b/1c/2 as well as experimental evidence that it interacts with PtAUREO1c *in vitro* and with bZIP10 by Yeast 2 Hybrid assay (Huysman *et al.*, 2013, Banerjee *et al.*, 2016b). Therefore, loss of PtAUREO1a might also result in disruption of several bZIP heterodimers, leading to more pronounced effects. Additionally, transcription of all photoreceptor genes was found to be strongly influenced by blue light in the wild type, which was not observed in the knockout strain (see Figure 5-4). Thus, this deregulation of the photoreceptors upon blue-light exposure could be an explanation for the prolonged blue-light conditions. Interestingly, transcripts of Pt*Aureo1a*, Pt*Aureo1b*, Pt*Aureo2* and *CPF4* were found to be strongly downregulated within 10 minutes upon light exposure in the wild type, indicating them to have a role in short term acclimation to blue light, whereas Pt*Aureo1c*, *CPF1* and *CPF2* seem to have more of a role in longer term light acclimation. *CryP* and *DPH1* seemed to fall in-between those categories, with upregulation after 10 min but severe downregulation after 60 min of blue-light treatment.

As all photoreceptors were found to be expressed and could thus not be responsible for the lack of blue-light perception alone, we had a closer look on the regulation of the known transcription factors of *P. tricornutum*. As only Aureochromes can directly influence gene expression, the other photoreceptors might directly or indirectly regulate activity of other transcription factors Interestingly, almost 20 % were downregulated in the knockout strain under red-light conditions compared to the wild type, whereas 9% were found to be upregulated, indicating that loss of PtAUREO1a has strong effects on the transcript level of transcription factors in absence of blue light. Thus, PtAUREO1a seems to be not only active under blue-light conditions. As a pronounced red-light phenotype had been observed in the knockdown strain as well, it is highly unlikely to be an off-target effect of the TALEN mutagenesis (Schellenberger Costa *et al.*, 2013b). Furthermore, it had been shown previously that heterologously expressed PtAUREO1a protein binds to DNA in absence of light and a conformational shift in the bZIP domain of the DNA-protein complex can be induced by blue light, supporting this assumption (Banerjee *et al.*, 2016b, Heintz and Schlichting, 2016). Thus, in the absence of blue light, PtAUREO1a might bind to it its regulatory motifs with lower affinity, resulting in a weak activation/inhibition of its target genes, whereas the significantly increased DNA-binding affinity upon blue-light exposure results in stronger activation/inhibition. Alternatively, the conformational shift observed upon blue-light exposure might also influence the downstream regulatory factors which interact with

PtAUREO1a, thereby influencing regulation differently in dependence of blue light. While transcription factors are often thought to be only activators or inhibitors of transcription, there is increasing evidence of so-called “dual”-targeting transcription factors (Latchman, 2001, Madan Babu and Teichmann, 2003, Ma, 2005, Boyle and Despres, 2010).

In conclusion, expression of PtAUREO1a needs to be tightly regulated to perform its intended function. Loss of PtAUREO1a leads to a blocked short term adaptation to blue light and strongly reduced NPQ capacity through reduced levels of Lhcx1, indicating it to be a master switch in light acclimation and photoprotection. However, further studies are necessary to confirm the potential dual-regulatory function, which could explain the red-light phenotype observed for both the PtAUREO1a knockdown and the knockout strains.

5.6 Acknowledgements

The authors are grateful to Doris Ballert for the help with the cultivation and transformation of *P. tricornutum*, as well as Marion Eisenhut and Andreas Weber (Universität Düsseldorf) for performing the library preparation, sequencing and transcript mapping. This work was supported by the University of Konstanz, the graduate school Biological Sciences (GBS) and the DFG (grant KR 1661/8-2 to PGK).

6 Generation of knockout strains for PtAUREO1b, PtAUREO1c and PtAUREO2

Serif, M.*, Rio Bartulos, C., & Kroth, P.G.

*Corresponding Author: manuel.serif@uni-konstanz.de

Plant Ecophysiology, Fachbereich Biologie, Universität Konstanz, D-78457 Konstanz, Germany

Keywords: *Phaeodactylum tricornutum*, Aureochrome, blue-light dependent transcription factor, TALEN knockout

6.1 Abstract

Four isoforms of the blue-light dependent transcription factors are encoded on the genome of the diatom *Phaeodactylum tricornutum*. TALEN plasmids for the remaining isoforms PtAUREO1b, PtAUREO1c and PtAUREO2 have been designed, constructed and transformed into *P. tricornutum*. One bi-allelic knockout strain of PtAUREO1b has been identified, which showed an increase in average cell size as well as formation of laterally attached cell chains of up to 50 cells. These chains were found to be very resistant to mechanical stress. Scanning electron microscopy showed that the cells are closely attached at patch-like structures, indicating an influence of PtAUREO1b on cell division. However, biofilm formation usually not observed in the fusiform morphotype cannot be excluded to play a role as well. Additionally, several candidates for PtAUREO1c and PtAUREO2 knockout strains were identified, which need to be further confirmed experimentally.

6.2 Introduction

The genome of the diatom *Phaeodactylum tricornutum* encodes for four isoforms of Aureochrome blue-light dependent transcription factors. Phylogenetic analysis showed that the Aureochromes form four distinct clades, of which *P. tricornutum* possesses one isoform each. Additionally, RNAi-mediated knockdown of a single isoform, PtAUREO1a, by about 50% led to a strong phenotype (Schellenberger Costa *et al.*, 2013b). Both findings indicate that the isoforms perform at least partially non-overlapping functions. So far, most research regarding the function of Aureochromes in *P. tricornutum* has been focused on the isoform PtAUREO1a (Huysman *et al.*, 2013, Schellenberger Costa *et al.*, 2013b). TALEN approaches have been shown to be a powerful tool for genome editing in the diatom *P. tricornutum* (Daboussi *et al.*, 2014, Weyman *et al.*, 2015), leading to discovery of a more severe phenotype in PtAUREO1a knockout strains (Serif *et al.*, 2017). To gain insight into the function of the other Aureochrome isoforms, TALEN plasmids were constructed for each isoform to generate knockout strains. Several candidates for knockout strains for each of the isoforms were identified, which need to be further screened to confirm that they are indeed bi-allelic knockout strains.

6.3 Material & Methods

6.3.1 Generation of TALEN plasmids

TALEN target sites were designed and constructed as described in (Serif *et al.*, 2017). Only targets without any predicted off-targets according to prediction with TALE-NT and a SAPTA score >30 were considered. The target sites chosen and their respective predicted cutting site downstream of the start codon is shown in table Table 6-1.

Table 6-1: TALEN target sites chosen for each isoform after evaluation with TALE-NT and SAPTA. The predicted cutting site downstream of the start codon is given.

Target	Cutting site	Target site 1	Target site 2
PtAUREO1b	422	TCAGCAAACACAGAATGCTT	TTGTACCATCGCGGCACCCG
PtAUREO1c	562	TCTCCGCAAAAAGTTTCTAC	TCCTCCAACCCGTGAATCTG
PtAUREO2	443	TCTCTCAGTTTTCCTGCACT	TGACCATCATTGCCAGGCAT

6.3.2 Cultivation and transformation of *P. tricornutum*

The *P. tricornutum* (Bohlin) strain UTEX646 was obtained from the culture collection of algae of the University of Texas (UTEX, Austin, USA). *P. tricornutum* was grown axenically in liquid F/2 medium without added silica and 16.5 ‰ salt content or on solid F/2 media which contained additionally 1.2% (w/v) Bacto Agar (BD, Sparks, MD, USA). Cells in liquid F/2 medium were cultivated in a 16h/8h light/dark cycle in Erlenmeyer flasks under continuous shaking at 20°C and an illumination of 35 µmol photons m^{-2} s^{-1} (Osram Lumilux L58W/840, Munich, Germany). Cell densities of liquid cultures were determined using a Coulter Counter Multisizer3 (Beckman Coulter, Krefeld, Germany). Plated cultures were cultivated under continuous illumination at 75 µmol photons m^{-2} s^{-1} (Osram Biolux L30W/965).

Nuclear transformation of *P. tricornutum* was performed using a Bio-Rad Biolistic PDS-1000/He Particle Delivery System (Bio-Rad, Hercules, CA, USA) fitted with 900/1100/1350 psi rupture disks as described previously (Apt *et al.*, 1996, Zaslavskaia *et al.*, 2000, Kroth, 2007). 10^8 cells per plate were bombarded with 1.25 µg of each plasmid. For selective cultivation of *P. tricornutum* transformants, 75 mg mL^{-1} Zeocin (Invitrogen, Carlsbad, CA, USA) and 150 mg mL^{-1} Nourseothricin (ClonNat, Werner Bioagents, Jena, Germany) were added to the solid F/2 media (Apt *et al.*, 1996, Zaslavskaia *et al.*, 2000).

6.3.3 Allele-specific PCR

Allele-specific PCR was performed as described in (Serif *et al.*, 2017). The primers used for amplification from genomic DNA are given in Table 6-2. As *PtAureo1b* and *PtAureo1c* do not possess suitable SNPs for allele-specific PCR, non-allele specific primers were used.

Table 6-2: Primers used for amplification of the Aureochrome genes to screen for insertions/deletions caused by the TALEN and subsequent repair by the cell.

Target	Forward primer (5'→3')	Reverse primer (5'→3')
PtAureo1b	TTGAACGAAATCTTTGCCGAGTA	TGCGTTTGGCGTGTTCAC
PtAureo1c	CGAATCCCAGTTCCACTCTG	GGAAAGAGCCGACATGAGC
PtAureo2-Allele1	CGATAATGGTATCTGTAAATAGTTCGCTGA	GCCCAAAAAGGCGGATC
PtAureo2-Allele2	CGATAATGGTATCTGTAAATAGTTCGCTGT	GCCCAAAAAGGCGGATG

6.3.4 Western Blot

Western Blots for the Aureochrome isoforms were done as described in (Serif *et al.*, 2017). An antibody targeting the large subunit of Rubisco (Agrisera, Vännas, Sweden) was used as a loading control.

6.3.5 Light microscopy

Exponentially growing cells were analyzed using an Olympus BX51 epifluorescence microscope equipped with a Zeiss AxioCam MRm digital camera system (Carl Zeiss Microscopy GmbH, Göttingen, Germany).

6.3.6 Scanning electron microscopy

Cells were harvested in mid-exponential phase by centrifugation and washed once in fresh F/2 medium without supplements. Cells were then fixed using 8% formaldehyde, 1% glutardialdehyde, 0.5% caffeine, 0.02 M calcium chloride, 0.02 M magnesium chloride, 0.2 M sodium cacodylate buffer, pH 7 on ice for 30 min, followed by centrifugation and a second fixation step for 60 min on ice. The fixed cells were then washed three times in 0.01 M sodium chloride, 0.01 M magnesium chloride, 0.1 M sodium cacodylate buffer for 10 minutes. Afterwards, the cells were incubated in 1% osmium tetroxide in 0.1 M sodium cacodylate buffer in the dark for 1 h at 4°C, followed by three 10 min washing steps in 0.1 M sodium cacodylate buffer. Samples were then dehydrated using successively increased concentrations of ethanol (30%, 50%, 70%, 90%, 96%, absolute). The dehydrated samples

were then filtered onto 1 µm PTFE pore filtre membranes, followed by critical point drying in CO_2 and sputtering with platinum to a thickness of 6 nm (Balzers SCD030; Oerlikon Balzers, Balzers, Liechtenstein). The finished samples were imaged with a Zeiss AURIGA Crossbeam scanning electron microscope (Carl Zeiss Microscopy GmbH, Göttingen, Germany).

Sample preparation was done by Lauretta Nejedli and the scanning electron microscopy was performed by Michael Laumann of the Electron Microscopy Center (EMC) of the Universität Konstanz.

6.4 Results

To gain insight into the function of the so far mostly uncharacterized Aureochrome isoforms, PtAUREO1b, PtAUREO1c and PtAUREO2, TALEN constructs were generated against each target gene. The resulting transformants are in different states of the screening process, which needs to be completed for their use in further studies.

6.4.1 Generation of PtAUREO1c knockout strains

A total of 33 antibiotic-resistant colonies were obtained by transformation with the TALEN constructs against PtAureo1c. No allele-specific differences suitable for amplification could be identified for *PtAureo1c* in wild type strain Pt4. To screen for potential knockout strains Western Blots with PtAUREO1c-specific antiserum were performed, using an antiserum against the large subunit of Rubisco as a loading control (see Figure 6-1). A total of eight strains were identified which showed reduced protein levels of PtAUREO1c. While four of them (strains 17, 19, 20 and 21) showed residual PtAUREO1c protein, the other 4 strains did not show any band of the expected size (strains 11, 14, 26 and 32). Thus, strains 11, 14, 26 and 32 were presumed to be bi-allelic knockout strains.

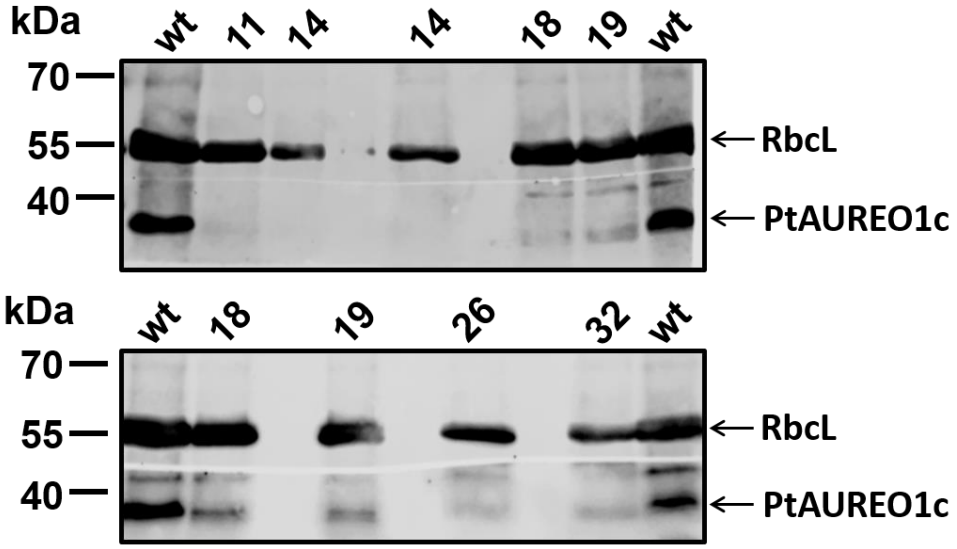


Figure 6-1: Western Blot of wild type and potential PtAUREO1c knockout strains using a PtAUREO1c antiserum. The expected molecular weight of PtAUREO1c is 36 kDa. An antiserum against the large subunit of Rubisco was used as a loading control. Different amounts of protein isolated from strain 14 were loaded due to the low protein concentration of this sample.

To confirm this assumption, genomic DNA was prepared and the PtAureo1c gene was amplified by PCR. Agarose gel electrophoresis showed that in strain 32 a smaller product could be amplified in addition to the fragment at the expected size, indicating a deletion (see Figure 6-2). Sequencing results of the obtained PCR products, however, did not indicate any mutations caused by the TALEN system in clones 11, 14 and 26. Strain 32, however, was found to have a 19 bp deletion in the first allele and a 604 bp deletion in the second allele, both of which result in frameshift mutations. As the *PtAureo1c* gene does not possess suitable allele-specific differences to distinguish both alleles, it is possible that the other allele indeed is mutated in those strains with a strongly reduced protein level, which still has to be confirmed by Southern Blots. However, as harvesting of cells for Western Blot and extraction of genomic DNA was performed independently, it cannot be excluded that cultures were mixed up during one of the screening steps. Thus, both screening steps should be repeated in addition to performing Southern Blot to correctly identify the knockout strains.

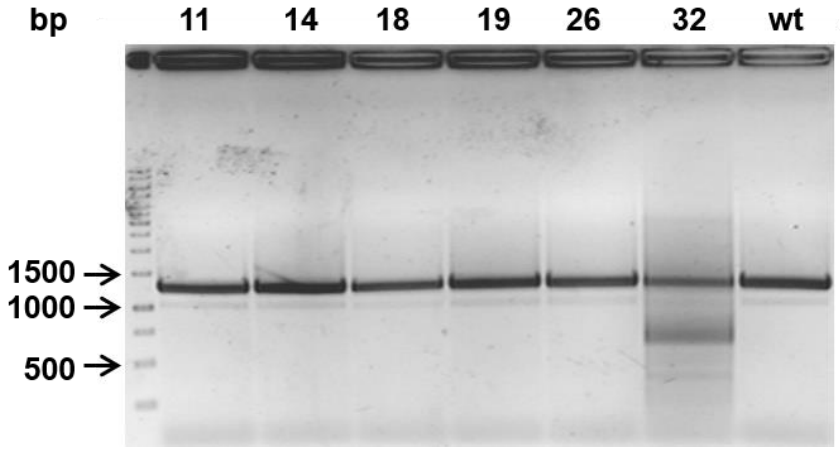


Figure 6-2: Amplification of the PtAureo1c gene in wild type (wt) and potential knockout strains (11, 14, 18, 19, 26 and 32). An additional smaller PCR product could be amplified for strain 32, indicating a deletion of about 500 bp.

6.4.2 Generation of PtAUREO2 knockout strains

A total of 17 antibiotic-resistant colonies were obtained by transformation with the TALEN constructs directed against *PtAureo2*. As the *PtAureo2* gene was found to be the only Aureochrome gene in strain Pt4 to possess several single nucleotide polymorphisms, allele-specific PCR was performed as a first screening step (see Figure 6-3). Two strains could be detected for which only amplification of a single allele was possible (strains 6 and 13), whereas for strain 16 no PCR product could be obtained for either of the alleles. Sequencing of the PCR products obtained for strains 6 and 13 showed no change in strain 6, whereas an 8 bp deletion was found in strain 13.

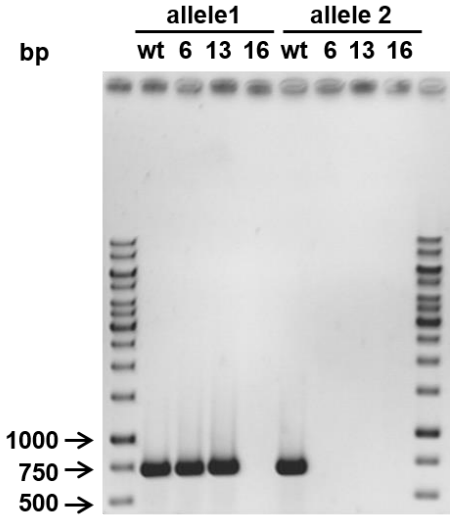


Figure 6-3: Allele-specific PCR followed by agarose-gel electrophoresis of the *PtAureo2* gene in wild type (wt) and putative knockout strains (6, 13 and 16). The expected fragment length is approximately 750 bp.

Thus, strain 6 is presumably a mono-allelic knockout strain, whereas strains 13 and 16 seem to be bi-allelic knockout strains. However, this finding needs to be confirmed by both Southern and Western Blots to prove the complete loss of PtAUREO2 before further studies can be made.

6.4.3 Generation of PtAUREO1b knockout strains

A total of 19 antibiotic-resistant colonies were obtained by transformation with the TALEN constructs against *PtAureo1b*. No allele-specific differences suitable for amplification could be identified for *PtAureo1b* in wild type strain Pt4. Thus, as a first round of screening Western Blot was performed (see Figure 6-4). In the wild type strain a single band was detectable at around 50 kDa, corresponding to PtAUREO1b. This band was not visible in strain 5, whereas only weak signals were detected for strain 7 and 10. Thus, strain 5 is indicated to be a bi-allelic knockout strain and strain 10 could be a mono-allelic knockout strain.

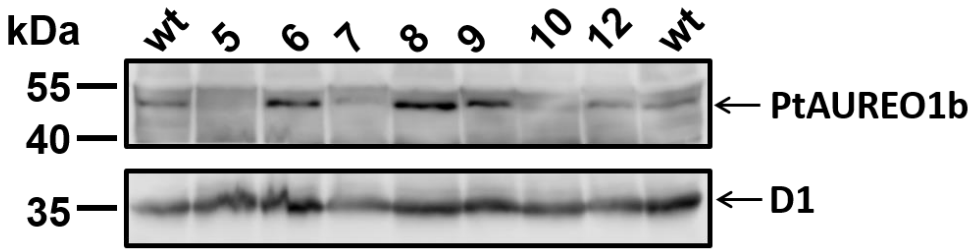


Figure 6-4: Western Blot of wild type (wt) and putative PtAUREO1b knockout strains using a PtAUREO1b antiserum. The expected molecular weight of PtAUREO1b is 46 kDa. An antiserum against D1 was used as a loading control.

To confirm that targeted mutagenesis events caused the loss of PtAUREO1b, PCR was performed (see Figure 6-5). In the wild type strain as well as strain 10, a single band was amplified. Sequencing results did not show mutations in strain 10. No PCR product could be amplified for strain 5, indicating larger insertions and/or deletion events in both alleles. To confirm that strain 5 is a bi-allelic knockout strain and strain 10 a mono-allelic knockout strain, Southern Blots need to be performed.

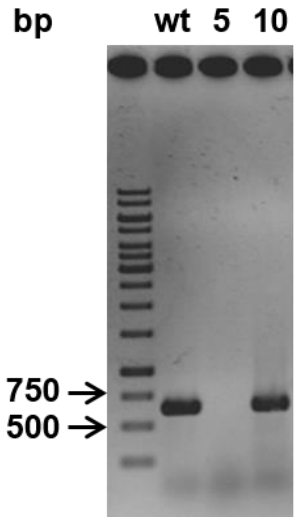


Figure 6-5: Amplification of the *PtAureo1b* gene of wild type (wt) and putative knockout strains (5 and 10) by PCR. No PCR product could be amplified for strain 5, indicating large insertion or deletion events in both alleles.

6.4.4 Preliminary characterization of PtAUREO1b knockout strain 5

During culturing of the PtAUREO1b knockout strain 5 for genetic characterization, measuring of the cell density using the Coulter Counter Multisizer 3 revealed a strange-looking size profile (see Figure 6-6): Wild type cells normally show a peak at a spherical volume of 4.5 to 4.9 µm depending on the age of the culture. In older cultures, a shoulder at around 5.8 µm can be seen. In this knockout strain, however, instead of a shoulder a second peak was visible at 5.8 µm, which is almost the same height of the primary peak, followed by a shoulder that reached a spherical volume of over 10 µm. As the cell size is calculated as a sphere, doubling the cell diameter results in an 8-fold increase in cell volume. In addition, while the peak around 4.8 µm was not shifted between wild type and mutant, the peak was more narrow, resulting in an average increase in cell size.

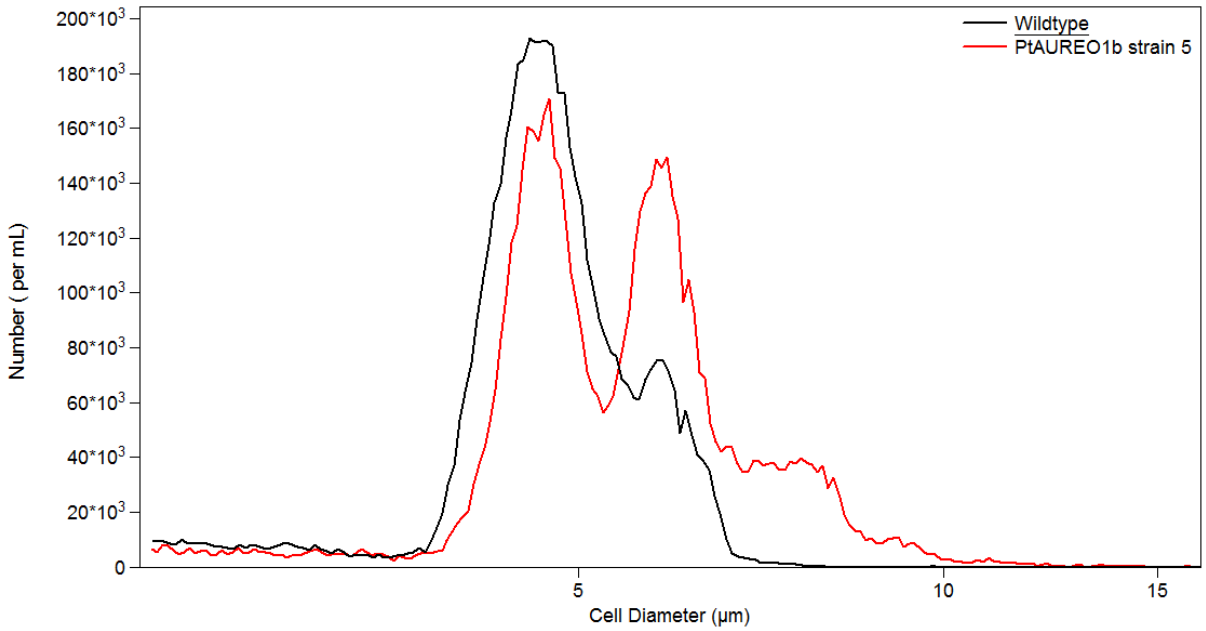


Figure 6-6: Cell size profile of wild type (depicted in black) and PtAUREO1a knockout strain 5 (depicted in red) measured using the Multisizer 3. The cell diameter given on the x axis corresponds to the water displacement volume of a sphere. The graph has been exported from the Multisizer 3 program using the print function.

Thus, we examined the cells under the light microscope (see Figure 6-7 A/B). For the wild type strain, mostly single fusiform cells were observed. In the knockout strain, however, in addition to single fusiform *P. tricornutum* cells, many cell pairs of two laterally attached cell pairs were found. These correspond presumably to the peak at 5.8 µm visible in the Coulter Counter measurement, as an increase in sphere diameter from 4.8 to 5.8 µm results in a doubling of the cell volume. Furthermore, longer chains of laterally attached cells could be found, which were usually multiples of 2 cells long. Most of these chains were between 4 and 10 cells long, while the longest chain observed was over 50 cells in length. To investigate whether these chains grow longer and longer, the cells were cultured for several weeks in liquid media. The size distribution pattern, however, remained largely unchanged throughout the different growth phases, whereas the total cell number increased (data not shown). Thus, these cell chains did not seem to grow indefinitely, indicating that separation of cells was not completely impaired. Additionally, we tested whether these chains could be disrupted by mechanical stress, i.e. vortexing, strong shaking in a bead beater (without addition of glass beads), or sonication. No treatment was found to disrupt the cell chains without destroying the cells itself, indicating the chains to be very stable.

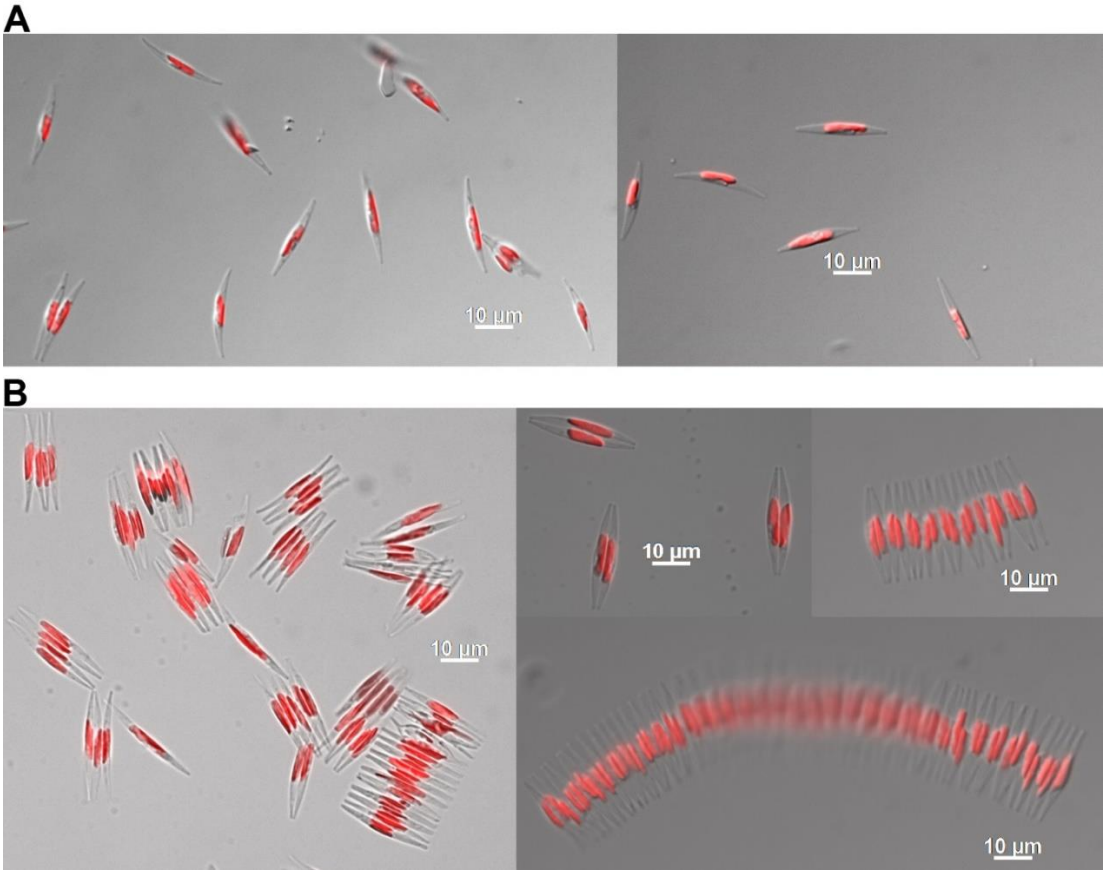


Figure 6-7: Light microscopy imaging of wild type (A) and PtAUREO1b knockout strain 5 (B). While the wild type mainly shows single cells, lateral attachment of cells is visible in the PtAUREO1b knockout strain. Scale bars representing 10 µm are shown for each picture. The autofluorescence of the chloroplast is shown in red.

As light microscopy could not give an indication as to whether this chain-forming phenotype is either caused by biofilm formation, which normally is not observed in fusiform cells, or a defect in cell division, we performed scanning electron microscopy to study how the cells are attached to each other (see Figure 6-8 for lower magnifications and Figure 6-9 for higher magnifications). Figure 6-8A shows an overview, many different chain lengths can be seen with very few single cells visible. Using higher magnification (see Figure 6-8B and Figure 6-9A/B), it became apparent that the cells were not physically attached over their whole length, but instead in a more patch-like manner. Additionally, the orientation of all cells within a chain seemed to be identical. Thus, it seems highly likely that the cells have a defect in cytokinesis. Unfortunately, artifacts were very abundant, which hindered any opportunity to take images at high magnifications to better visualize the composition of these patches. To see whether the chains are caused by fused membranes failing to divide transmission electron microscopy was attempted. Unfortunately, the sample preparation failed, resulting in no usable images. Thus, both electron microscopy approaches should be repeated and, eventually, the sample preparation process should be further optimized.

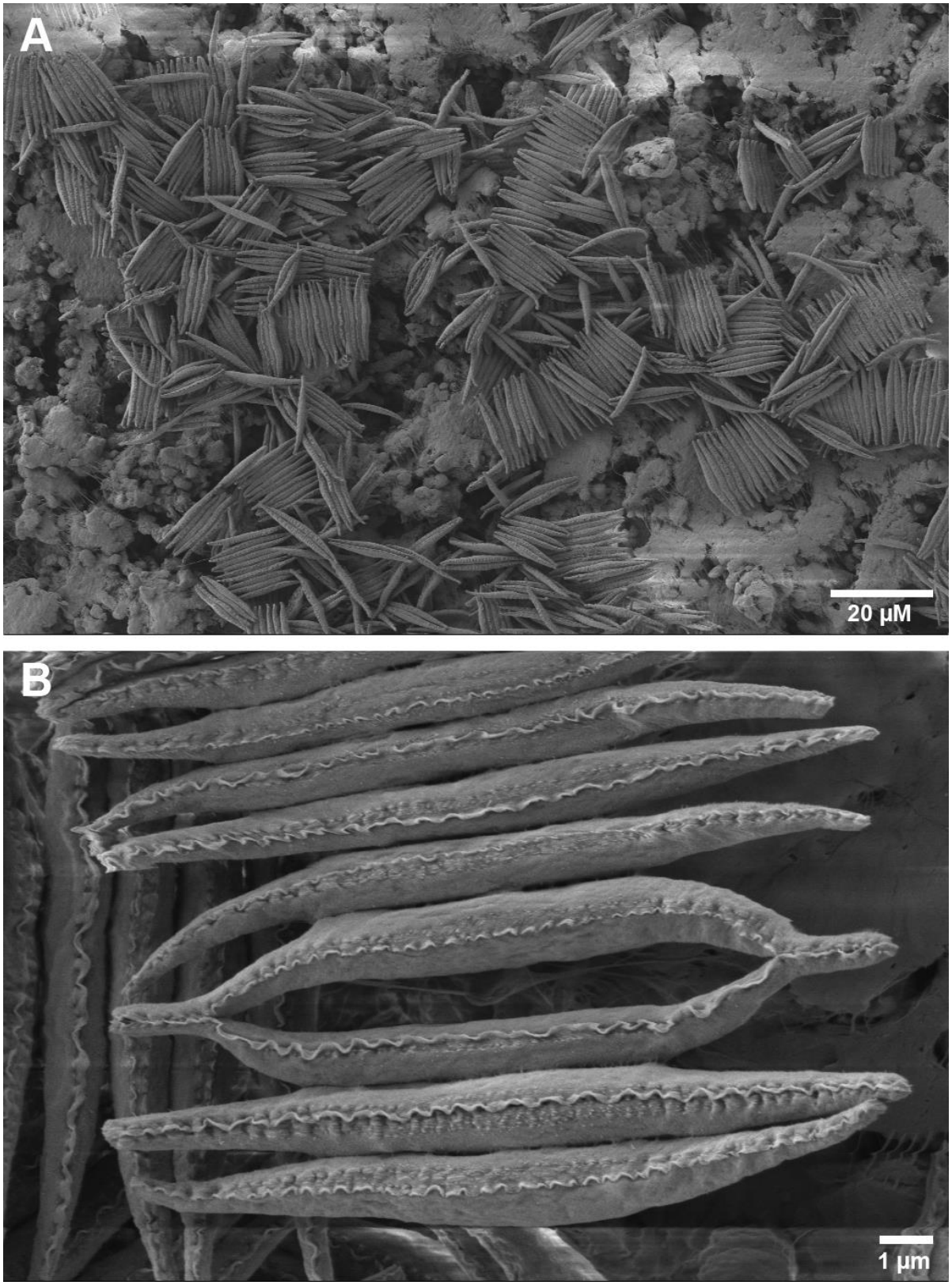


Figure 6-8: Scanning electron microscopy of strain PtAUREO1b K5. Pictures were taken at different magnifications (a scale bar is given in the lower right corner for reference). Cells seem not to be laterally attached along the whole length of the cell but only at small patches.

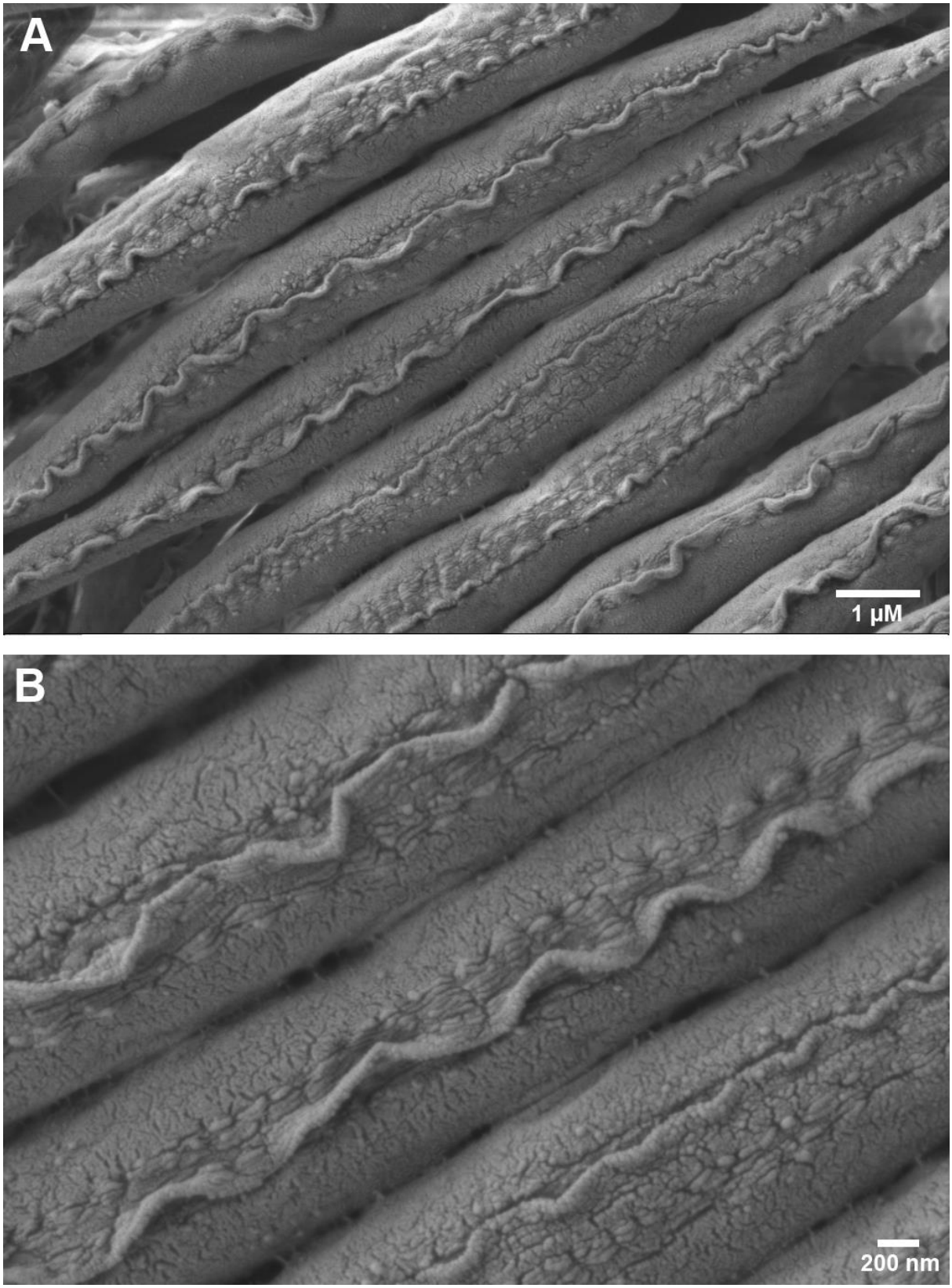


Figure 6-9: Scanning electron microscopy of strain PtAUREO1b K5. Pictures were taken at different magnifications (a scale bar is given in the lower right corner for reference). A closer look at the patch-like attachments is shown.

6.5 Discussion

Candidates for bi-allelic knockout strains have been identified for all three remaining PtAUREO isoforms, which still need to be confirmed before using them for further studies. Of these strains, only the PtAUREO1b knockout strain 5 has been preliminarily characterized, mainly due to its striking phenotype of chain formation. These chains were found to be very stable and could not be disrupted by mechanical stress without destroying the cells themselves. The average number of cells per chain did not increase during continued culturing in liquid media, however, indicating that separation of the cells was not completely impaired. Chain-forming *P. tricornutum* has been observed independently in wild type and transgenic lines, however, the cause of it remains unknown (Coughlan, 1962, Borowitzka *et al.*, 1977, Gherardi *et al.*, 2016). Unfortunately, only one bi-allelic PtAUREO1b knockout strain has been identified which shows this phenotype, therefore alterations due to the random integration of the plasmids into the genome and/or potential off-target effects of the TALEN cannot be excluded as the cause. Hence, the remaining transformants should be screened as well to find out whether this phenotype was caused randomly or is found in other bi-allelic knockout strains of PtAUREO1b as well. For further characterization of the strain it would be of interest to determine whether the chain formation is caused by biofilm formation, which is usually not observed in the fusiform morphotype, or instead is caused by a cytokinesis defect. The fresh water diatom *Achnanthidium minutissimum* forms a biofilm when grown xenically and/or in co-culture with a *Bacteroidetes* isolate (Windler *et al.*, 2015). Recently, pictures of these biofilms were taken using scanning electron microscopy (Leinweber and Kroth, 2015). These pictures show large amounts of extracellular matrix forming the biofilm, in which the diatom cells are only loosely attached to each other. The patch-like pattern of attachment seen in the SEM pictures of the PtAUREO1b knockout strain, however, looks different with cells closely attached and without a visible extracellular matrix, indicating that a cytokinesis defect may be more likely. Therefore, transmission electron microscopy should be employed to determine, whether the cell membranes are partially fused. Additionally, degradation of extracellular matrix compounds using a combination of proteases and gluconases could be used to prove whether biofilm formations plays a role as well.

6.6 Acknowledgements

The authors are grateful to Doris Ballert for the help with the cultivation and transformation of *P. tricornutum*, Dr. Carolina Rio Bartulos for help with light microscopy, as well as Marc Halder and Zeno Riester for help in mutant screening of PtAUREO1c. We also thank Lauretta Nejedli and Dr. Michael Laumann at the Electron Microscopy Center of the Universität Konstanz for cell preparation and SEM-microscopy. This work was supported by the University of Konstanz, the graduate school Biological Sciences (GBS), and the DFG (grant KR 1661/8-2 to PGK).

7 Knockout of Lhcx1 leads to loss of NPQ under low light conditions

Serif, M.*, Rio Bartulos, C., Kroth, P.G. & Lepetit, B.

*Corresponding Author: manuel.serif@uni-konstanz.de

Plant Ecophysiology, Fachbereich Biologie, Universität Konstanz, D-78457 Konstanz, Germany

Keywords: *Phaeodactylum tricornutum*, Lhcx1, Non-photochemical quenching (NPQ), TALEN knockout

7.1 Abstract

NPQ (non-photochemical quenching), the dissipation of excess light energy as heat, is an important photoprotective process to avoid photodamage under high light conditions. In diatoms, this process depends mainly on the presence of a transthylakoidal pH gradient, the conversion of the xanthophyll cycle pigment diadinoxanthin to diatoxanthin and the presence of Lhcx proteins. The genome of the model diatom *Phaeodactylum tricornutum* encodes four different isoforms, of which Lhcx1 is the most strongly expressed under low light conditions. Knockout of the blue-light dependent transcription factor PtAUREO1a led to a decrease of Lhcx1 protein level, an increased xanthophyll cycle pigment pool and strongly reduced NPQ capacity. Thus, we created Lhcx1 knockout strains using TALEN to investigate whether the isoform Lhcx1 is essential for NPQ formation under low light conditions. The generated knockout strains were found to have a severely reduced NPQ capacity under low light conditions. Thus, the other three Lhcx isoforms does not appear capable of compensating for the loss of Lhcx1, indicating the importance of this isoform for photoprotection under these conditions.

7.2 Introduction

Photosynthesis is a tightly regulated process, which needs to be adapted to the ever-changing light conditions of the environment. If light availability is low, the available light must be used very efficiently to convert it into sufficient energy for growth. Absorption of too much light, however, can cause irreversible photodamage by formation of free radicals, which can then attack all major cellular macromolecule classes (Triantaphylidès and Havaux, 2009). Thus, efficient perception of light intensity and subsequent regulation of photosynthesis and photoprotection mechanisms are of utmost importance to avoid cell damage, especially under strongly fluctuating light conditions present in aquatic habitats (MacIntyre *et al.*, 2000, Lavaud, 2007). Light changes can be either measured directly by photoreceptors, of which the model diatom *Phaeodactylum tricornutum* possesses several Cryptochromes, Aureochromes and a Phytochrome (Depauw *et al.*, 2012), or in an indirect manner by measuring physiological parameters directly influenced by light (Lepetit and Dietzel, 2015). One of these potential indirect triggers is the redox state of the plastoquinone pool (PQ) (Pfannschmidt, 2003), which was only recently identified to be used for regulation of nuclear-encoded genes involved in photoprotection of diatoms (Lepetit *et al.*, 2013). Several different mechanisms of

photoprotection are present in diatoms, however, one of the most important mechanisms to quickly react to high light intensity is the dissipation of excessively absorbed light energy as heat (Lavaud and Lepetit, 2013). This process, termed NPQ (non-photochemical fluorescence quenching), depends on three major components in diatoms: The presence of a transthylakoidal pH gradient, the enzymatic conversion of the xanthophyll cycle pigment diadinoxanthin to diatoxanthin, which is induced by the transthylakoidal pH gradient, and the presence of Lhcx proteins (Bailleul *et al.*, 2010, Goss and Jakob, 2010, Goss and Lepetit, 2015). The involvement of Lhcx proteins in photoprotection of diatoms has only been discovered a few years ago (Bailleul *et al.*, 2010, Zhu and Green, 2010). The genome of the model diatom *P. tricornutum* encodes four different Lhcx proteins, which differ in their expression patterns. While Lhcx2, Lhcx3 and Lhcx4 rather seem to participate in photoprotection under prolonged high light stress and/or nutrient deprivation (Nymark *et al.*, 2009, Taddei *et al.*, 2016, Lepetit *et al.*, 2017), Lhcx1 is already strongly expressed under low light conditions, gets induced further under high light (Lepetit *et al.*, 2013), and is expressed more strongly under blue light treatment compared to red light treatment (Schellenberger Costa *et al.*, 2013a).

Regulation of Lhcx1 gene expression occurs on a rather rapid timescale. For instance, using microarrays it was shown that the transcripts of Lhcx1 increase within half an hour of light treatment when light intensity is increased or dark arrested cells are re-exposed to light (Nymark *et al.*, 2009, Nymark *et al.*, 2013). Furthermore, a shift of red-light adapted cells to blue light conditions led to a strong increase of Lhcx1 transcripts within 10 min (see Chapter 5). This regulation was completely abolished in a PtAUREO1a knockout strain, indicating that Lhcx1 is directly regulated by PtAUREO1a, as synthesis of an intermediate protein in the signaling cascade would require more time. Consequently, knockout of the blue-light dependent transcription factor PtAUREO1a led to both a 50% reduction of Lhcx1 protein levels and NPQ capacity under low light conditions, (see Chapters 4 and 5). A similar result was obtained by direct silencing of Lhcx1, which led to a strong decrease of NPQ capacity, whereas overexpression resulted in strong increase of NPQ capacity (Bailleul *et al.*, 2010). All these results indicate a dominant role of the Lhcx1 protein in the NPQ mechanism of *P. tricornutum*. Thus, we attempted to create Lhcx1 knockout strains to investigate to what extent the complete lack of Lhcx1 leads to a loss of NPQ, or whether other mechanisms, which may be mechanistically based on the existence of the other Lhcx isoforms, could provide NPQ.

7.3 Material & Methods

7.3.1 Generation of TALEN constructs

TALEN plasmids were designed and constructed as described in (Serif *et al.*, 2017). Two target sites were chosen within the *lhcx1* gene, which are shown in Table 7-1.

Table 7-1: TALEN target sites chosen and their predicted cutting site downstream of the start codon is given. The name of the TALEN pair is the predicted cutting site within the *lhcx1* gene.

Name	Left TALEN target sequence	Right TALEN target sequence
40	TCGCTGCCACCATCCTTGCT	TTGTTTGGGCCGGAGCGAAA
179	TACACGACCGTGGTCCTTCT	TGGCTCCAACTAGATCTTCT

7.3.2 Cultivation of Algae

The *P. tricornutum* (Bohlin) strain UTEX646 was obtained from the culture collection of algae of the University of Texas (UTEX, Austin, USA). *P. tricornutum* was grown axenically in liquid F/2 medium without added silica and 16.5 ‰ salt content or on solid F/2 media which in addition contained 1.2% (w/v) Bacto Agar (BD, Sparks, MD, USA). Cells in liquid F/2 medium were cultivated in a 16h/8h light/dark cycle in Erlenmeyer flasks under continuous shaking at 20°C and an illumination of 35 (low light) or 100 (medium high light) µmol photons m^{-2} s^{-1} (Osram Lumilux L58W/840, Munich, Germany). Plated cultures were cultivated under continuous illumination at 75 µmol photons m^{-2} s^{-1} (Osram Biolux L30W/965).

7.3.3 Nuclear transformation of *P. tricornutum*

Nuclear transformation of *P. tricornutum* was performed using a Bio-Rad Biolistic PDS-1000/He Particle Delivery System (Bio-Rad, Hercules, CA, USA) fitted with 900/1100/1350 psi rupture disks as described previously (Apt *et al.*, 1996, Zaslavskaia *et al.*, 2000, Kroth, 2007). 10^8 cells per plate were bombarded with 1.25 µg of each plasmid. For selective cultivation of *P. tricornutum* transformants, 75 mg mL^{-1} Zeocin (Invitrogen, Carlsbad, CA, USA) and 150 mg mL^{-1} Nourseothricin (ClonNat, Werner Bioagents, Jena, Germany) were added to the solid F/2 media (Apt *et al.*, 1996, Zaslavskaia *et al.*, 2000).

7.3.4 Protein Isolation and Immunoblotting

Protein isolation subsequent Western Blots against Lhcx1 were performed as described in (Lepetit *et al.,* 2017). Blots were developed using an Odyssey Fc Imaging System (Li-Cor, Bad Homburg, Germany).

7.3.5 Southern Blot

Genomic DNA was isolated from *P. tricornutum* and used for Southern Blots as described in (Serif *et al.,* 2017). The DIG-labeled probe was generated using primers Lhcx1_for (ATGCTTGCTGTCGTTGGATTCC) and Lhcx1_rev (ATTCTCAAGGATTCCCTTCCCGTT) and a previously constructed Lhcx1 overexpression plasmid.

7.3.6 Measurement of non-photochemical quenching (NPQ)

Cell suspensions in mid-exponential phase were adjusted to a Chl *a* content of 1 µg ml^{-1} and NPQ was measured with an AquaPen-C AP 100 (Photon Systems Instruments, Brno, Czech Republic) using light pulses with an intensity of 2100 µmol photons m^{-2} s^{-1} applied every 20 s to induce maximal fluorescence and 700 µmol photons m^{-2} s^{-1} of actinic light to induce NPQ.

7.4 Results & Discussion

NPQ in diatoms is known to depend on the presence of a thylakoidal proton gradient, the de-epoxidation state and pool size of the xanthophyll cycle pigments and the presence of Lhcx proteins. Previously studied RNAi-mediated knockdown strains of Lhcx1 were found to have strongly decreased NPQ capacity compared to the wild type despite an unchanged de-epoxidation state of the xanthophyll cycle pool. TALEN-mediated PtAUREO1a knockout strains also showed a reduced level of NPQ capacity despite an increased xanthophyll cycle pigment pool (see Chapter 4), and expression levels of its gene product were found to be strongly reduced (see Chapter 5). To provide more insight into the role of Lhcx1 for NPQ, two TALEN pairs targeting position 40 or 179 downstream of the start codon were generated. The plasmids were transformed into strain Pt4 and the resulting transformants were screened with a combination of Western and Southern Blots. As a first screening step, Western Blots were performed using an antiserum recognizing all Lhcx isoforms. The strongest band visible at around 20 kDa corresponds to Lhcx1, however the isoform Lhcx3 can hardly be separated

on a mini gel run for a mini gel standard length (Figure 7-1A). Thus, strains which showed no or very little residual protein at the height of Lhcx1 (strains 40-1, 40-12 and 179-3) were chosen as candidates. Cells were spread on plates and single colonies were picked to ensure that genetically homogenous colonies were used for further genetic and phenotypic characterization.

Southern Blots of genomic DNA digested with the restriction enzyme SalI resulted in a single band at approximately 3.5 kbp for the wild type (Figure 7-1B). The re-isolated clone from strain 40-1 and 179-3 each showed a single shifted band and retained no band corresponding to the wild type fragment, thus indicating them to be bi-allelic knockout strains with insertions of up to 1 kbp, whereas the clone isolated from strain 40-12 showed a shifted band indicating a small deletion and another band indicating an insertion of over 5 kbp.

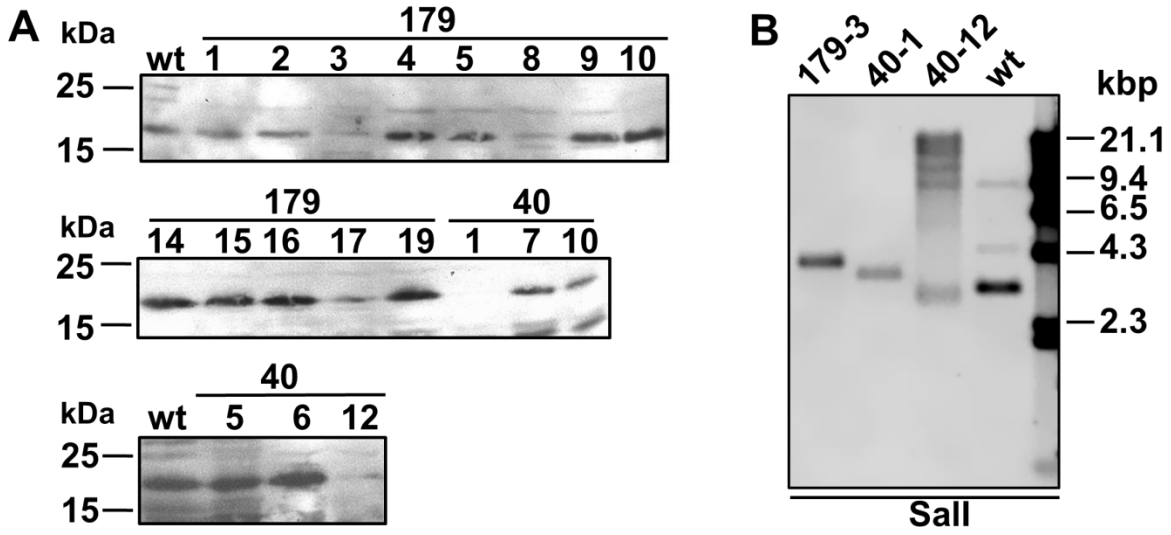


Figure 7-1: Screening of the potential LHCX1 knockout strains. A: Detection of LHCX1 via Western Blot of wild type and potential LHCX1 knockout strains. The expected molecular weight of Lhcx1 is around 20 kDa. B: Southern blot of wild type and three candidate knockout strains (40-1, 40-12 and 179-3). Shifted bands indicate big deletions or insertions.

To prove that the strains chosen are true knockout strains the cells were cultivated under medium light conditions (100 µmol photons m^{-2} s^{-1}), which induces a stronger expression of Lhcx1. Gels were run for longer time periods to separate Lhcx3 from Lhcx1. Western Blots with the Lhcx-specific antibody showed very high levels of Lhcx1 in the wild type compared to the blots done with cells cultivated at low light conditions (compare Figure 7-1A with Figure 7-2). In the knockout lines, no band corresponding to Lhcx1 was detectable, confirming that they are bi-allelic knockout lines. Interestingly, there was no major change in

Lhcx3 expression between wild type and knockout lines, confirming that the TALEN pairs specifically detect Lhcx1, but not the closely related Lhcx3.

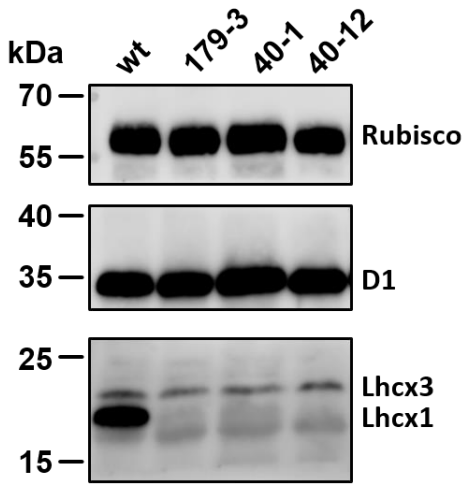


Figure 7-2: Western Blot of bi-allelic knockout strains under high-light conditions. The LHCX-specific antibody detects all isoforms of LHCX proteins. It strongly labels a band at 20 kDa in the wild type but not in the knockout strains. Antisera against D1 (35 kDa) and the large subunit of Rubisco (55 kDa) served as loading controls and confirmed equal sample loading and blotting.

To investigate whether the knockout lines can perform NPQ when exposed, cells in mid-exponential growth phase were analyzed by PAM fluorometry. These measurements can be divided in two phases: In the first phase, actinic light is used to expose the cells to high light intensity to induce NPQ, whereas in the second phase of the measurement, the cells are exposed to darkness to follow the relaxation of NPQ. Short pulses of saturating light are given at certain time points to induce the maximum fluorescence emitted by the chlorophylls due to a complete closure of the photosystem II. The intensities of these fluorescence spikes are used to calculate NPQ, whereby a decrease of the maximum fluorescence during actinic light exposure indicated the onset of NPQ. The raw fluorescence output of the NPQ measurement of wild type and an Lhcx1 knockout strain is shown in Figure 7-3. Whereas the wild type shows a strong decrease in relative fluorescence intensity due to NPQ during the initial 4 min of exposure to the actinic light and a subsequent recovery of fluorescence intensity in the dark phase, the knockout lines showed a very weak reduction of fluorescence intensity, which did not regenerate during the second part of the measurement.

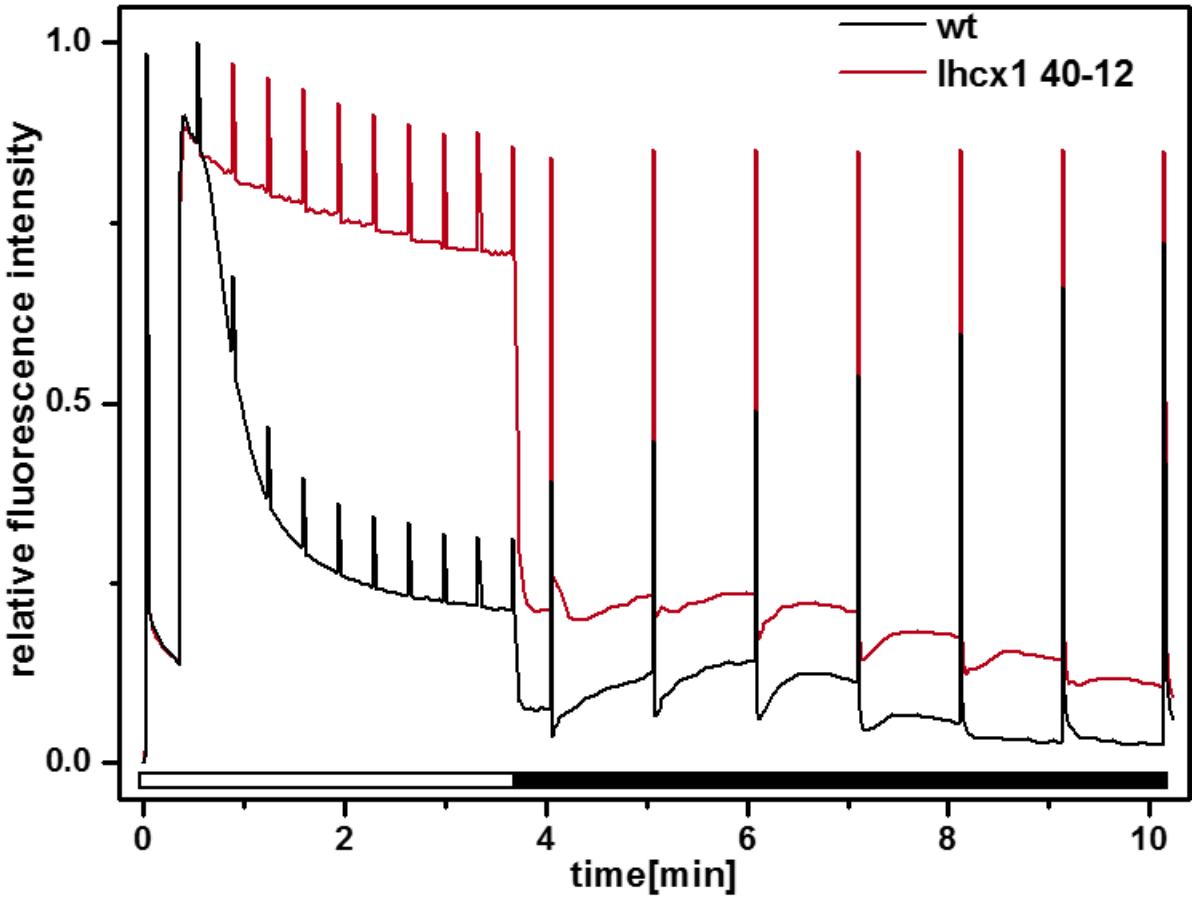


Figure 7-3: Raw data of NPQ measurement of either wild type (wt) or an Lhcx1 knockout strain (40-12). A white bar at the bottom indicates actinic light whereas a black bar indicates darkness except for the saturating light pulses, which cause the spikes of fluorescence in the measurement used for calculating NPQ.

The NPQ values calculated from the raw data over the course of the measurement are shown in Figure 7-4. The wild type showed a maximum NPQ of 0.87 after 3 min of actinic light whereas in the three knockout strains an NPQ of only 0.15 was measured. For the part of the measurement taking place in darkness the wild type quickly reduced NPQ to about 30% of the maximum level within 6 minutes. In the knockout strains, however, NPQ seemed to slightly increase within the dark phase of the measurement as well instead of falling. This indicates that the slight NPQ observed in the knockout strains does not belong to the rapid energy quenching process called qE, which depends on Lhcx proteins and the xanthophyll cycle, but rather is a photoinhibitory process possibly caused by degradation of photosystem II (Derks *et al.*, 2015, Goss and Lepetit, 2015). The observed phenotype in the knockout strains is much more pronounced than in the previously studied RNAi-mediated knockdown strains or the PtAUREO1a knockout strains (Bailleul *et al.*, 2010, Serif *et al.*, 2017), resulting in a decrease of over 80% of the NPQ capacity and essentially a complete loss of the qE component (qE null mutant). At least under low light culturing conditions, NPQ seems to be primarily

dependent on Lhcx1 (which works together with the de-epoxidation reaction), whereas the other isoforms seem to play only a minor role under the conditions assayed. The obtained qE null mutants will allow getting in-depth insights into the NPQ mechanism of diatoms for the first time.

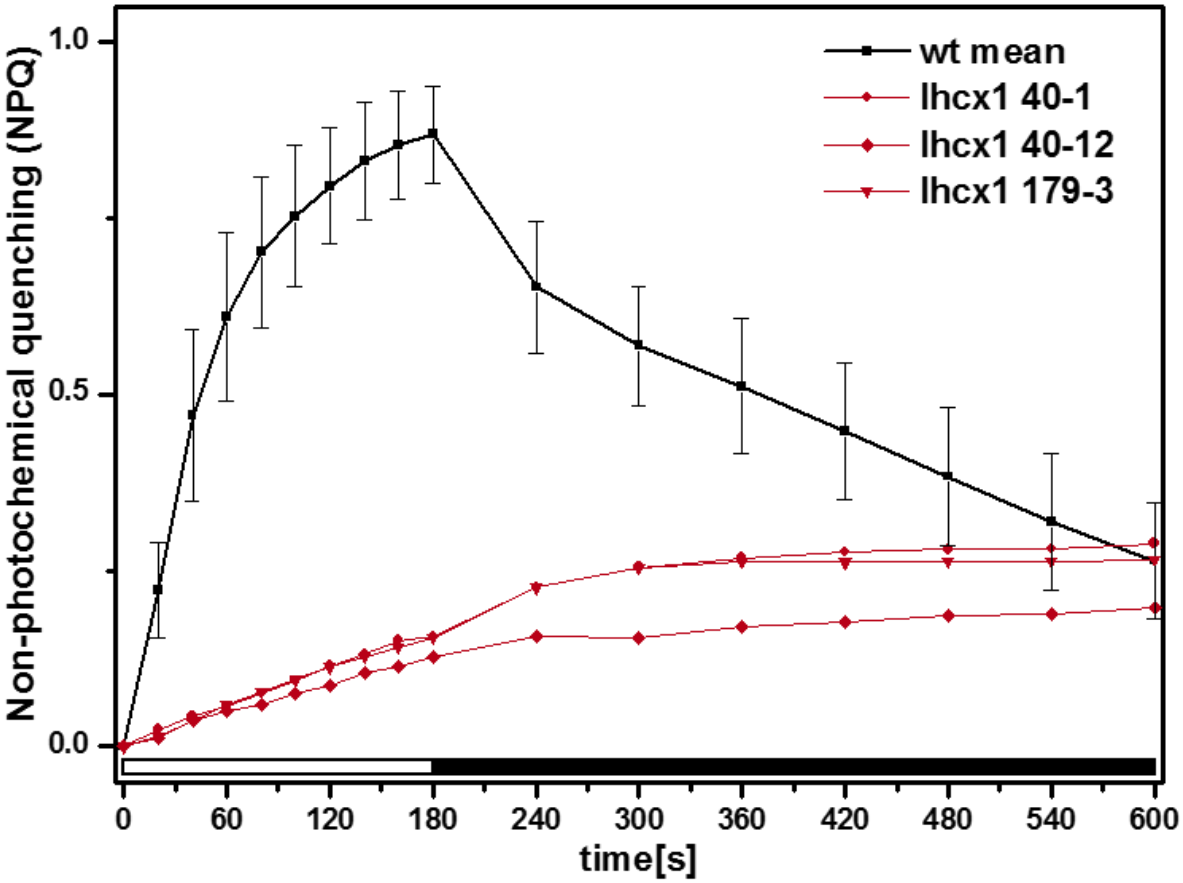


Figure 7-4: NPQ values calculated from the raw data shown in Figure 7-3. The wild type strain was measured in three replicates, whereas one measurement for each knockout strain is shown. A white bar at the bottom indicates actinic light, whereas a black bar indicates darkness except for the measuring light pulses.

7.5 Conclusion

A total of three bi-allelic Lhcx1 knockout strains were generated using two different TALEN target sites. All three strains cultivated under low light conditions show a severely reduced level of NPQ when exposed to light stress conditions and represent qE-null mutants under these conditions, which makes them an invaluable tool for future research on photoprotection.

7.6 Acknowledgements

The authors are grateful to Doris Ballert for the help with the cultivation and transformation of *P. tricornutum*, Annette Ramsperger for assembly of the TALEN constructs, as well as Teresa Hagen for help in mutant screening. This work was supported by the University of Konstanz, the graduate school Biological Sciences (GBS), the EU FP7 Marie Curie Zukunftskolleg Incoming fellowship (grant no. 291784, to BL), and the DFG (grants KR 1661/8-2 to PGK, LE 3358/3-1 to BL).

8 General discussion

Perception of light intensity and light quality, which can be performed by photoreceptors or via the photosystems, is essential for photosynthetic organisms to avoid photodamage under high light conditions and to generate sufficient energy under low light conditions. Aureochromes are a class of blue-light dependent transcription factors only found in Stramenopiles. Four different isoforms have been found to be encoded in the genome of the model diatom *P. tricornutum* (Takahashi *et al.*, 2007, Ishikawa *et al.*, 2009, Schellenberger Costa *et al.*, 2013a). While the biophysical characterization of these Aureochromes was well underway at the start of this thesis, very little was known about their function *in vivo*. The only known regulatory element which was known to interact with one Aureochrome, PtAUREO1a, is the diatom-specific cyclin 2 (*dsCYC2*) promoter, indicating a role in cell cycle progression (Huysman *et al.*, 2013). Furthermore, a PtAUREO1a knockdown strain had been generated, which showed an "hyper" high-light acclimation phenotype with reduced *Chl a* content per cell, increased photosynthesis and oxygen evolution rates as well as an increased NPQ capacity. Interestingly, while the effect was more pronounced at blue light conditions, this phenotype was also observed when cells were cultured in red light, indicating a blue-light independent effect of PtAUREO1a (Schellenberger Costa *et al.*, 2013b). To get further insight into the function of the Aureochromes, several different molecular approaches were applied.

8.1 PtAUREO1a seems to act as a master switch for light regulation

The studies on knockdown strains of PtAUREO1a revealed a clear phenotype, however, as the amount of PtAUREO1a only had been reduced, residual protein (approximately 50%) might have masked more severe phenotypes (Schellenberger Costa *et al.*, 2013b). Thus, the generation of a strain without any residual PtAUREO1a was essential to unravel PtAUREO1a function. Knockout of PtAUREO1a *via* TALEN was shown to lead to a decreased Chl *a* content per cell, a reduced NPQ capacity and a decrease in average cell size (see Chapter 4). Interestingly, despite the known interaction of PtAUREO1a with the *dsCYC2* promoter, no changes of the growth of PtAUREO1a knockout mutants were observed under the conditions

assayed. To identify genes directly regulated by PtAUREO1a and to investigate the molecular mechanisms behind the observed phenotype, wild type and knockout strain were adapted to red light and shifted to blue light, taking samples for RNA-seq after 0, 10 and 60 min (see Chapter 5). While the wild type cells featured a pronounced re-adjustment of their transcriptome upon blue-light illumination, the knockout strain seemed to be "blind" regarding the blue-light trigger (see Table 5-2). Furthermore, it became apparent that the transcripts of all photoreceptors were regulated after switching from red to blue light in the wild type, but not in the knockout strain, indicating a direct or indirect influence of PtAUREO1a on all other photoreceptors (see Figure 5-4). Furthermore, other transcription factors seemed to be strongly affected by the lack of PtAUREO1a as well. Regulation of transcription of bZIP and zinc finger transcription factors upon blue-light illumination was almost completely blocked in the knockout strain, and a deregulation of about 20% of these transcripts was also detected under red-light conditions (see Table 5-3). PtAUREO1a may not only potentially form heterodimers with the other PtAUREO isoforms as has been experimentally confirmed for PtAUREO1c (Banerjee *et al.*, 2016b), but also with some of the 19 other bZIP transcription factors encoded on the genome of *P. tricornutum* (Rayko *et al.*, 2010), thereby multiplying its effects. Unfortunately, very little is known about the function of individual transcription factors of *P. tricornutum*, thus not much information can be deduced from their respective regulation patterns. bZIP11 seems to be involved in regulation of carbonic anhydrases in response to the carbon dioxide concentration, whereas bZIP14 was shown to be involved in activation of the TCA cycle upon nitrogen starvation (Ohno *et al.*, 2012, Matthijs *et al.*, 2017). Furthermore, Yeast 2 Hybrid assays and Luciferase assays in tobacco protoplast cells strongly indicate a synergistic interaction between PtAUREO1a and bZIP10 to induce transcription of *dsCYC2*. However, it remains unclear whether heterodimerization occurs or if homodimers of both transcription factors occupy different DNA binding sites in close proximity to increase transcription rates (Huysman *et al.*, 2013). Transcription of *dsCYC2* was drastically reduced in the knockout strain even when compared to the red-light grown wild type cells, in which *dscyc2* should be transcribed at a very low level. As addition of the protein synthesis inhibitor CHX resulted in a delayed reduction of *dsCYC2* transcript levels after blue-light exposure, an unknown repressor of *dsCYC2* transcription is most likely transcribed upon blue-light exposure (Huysman *et al.*, 2013), which could be induced by an Aureochrome or a different class of transcription factor in conjunction with a cryptochrome.

The observed reduction in NPQ capacity (see Table 4-2) seems to be caused by downregulation and/or lack of upregulation upon blue-light exposure of the photoprotective Lhcx1 protein, as was shown by RNA-seq and by Western Blot in cells grown in white-light (see Figure 5-1 and Figure 5-4). As a strong induction of Lhcx transcription was detected after 10 min of blue-light illumination, a direct regulation by PtAUREO1a is highly likely and should be confirmed by Yeast 1 Hybrid assays. Interestingly, a similar phenotype was observed in the green algae *Chlamydomonas reinhardtii* phototropin knockout strains: Chl content and NPQ was found to be strongly reduced, and expression of LHCSR3, a homolog to Lhcx1, was strongly downregulated (Petroutsos *et al.*, 2016). As diatoms do no possess phototropins, which are only found in the green lineage, PtAUREO1a might have partially taken over its functions regarding light acclimation. The presumed functions of PtAUREO1a upon blue-light illumination inferred from earlier studies as well as the generated knockout strain are summarized in Figure 8-1.

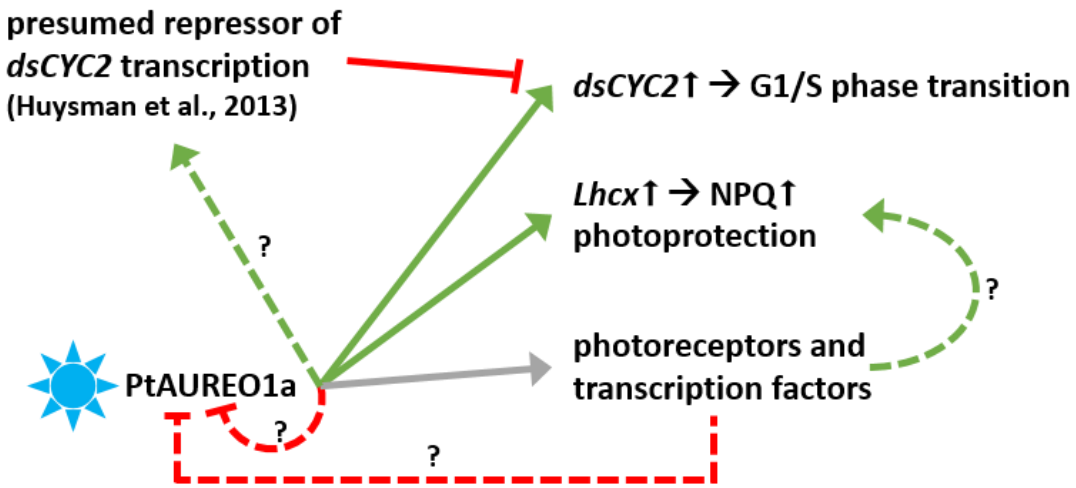


Figure 8-1: Influence of PtAUREO1a on transcription upon blue-light illumination. Green arrows represent induction of transcription, whereas red lines with a horizontal bar represent suppression of transcription. Dotted arrows/lines with a question mark indicate speculative regulatory function, for which no direct evidence is available.

Loss of PtAUREO1a leads to a weaker phenotype under red light as well, as had already been observed in RNAi-based knockdown strains (Schellenberger Costa *et al.*, 2013b). Hence, a blue-light independent function of PtAUREO1a seems very likely. Using *in vitro* assays, it has been shown that heterologously expressed PtAUREO1a homodimers bind to DNA light-independently, but a conformational shift can be detected when illuminated with blue light (Banerjee *et al.*, 2016b, Heintz and Schlichting, 2016). Taken together, these findings indicate two potential modes of action: An activator/repressor-only function, which becomes more pronounced upon blue-light illumination due to the increased binding affinity and/or attraction of different co-factors due to the conformational change. An alternative mode of action would be a dual regulatory function, i.e. repression of the transcription of target genes in absence of blue light and upregulation of transcription under blue-light conditions, or *vice versa* (see Figure 8-2). Thus, absence of PtAUREO1a would lead to a phenotype under red light conditions, which would be amplified under blue light conditions by a lack of further activation/suppression of transcription compared to the wild type. Dually regulating transcription factors were found to be very common (approximately 20% of transcription factors) in both *E. coli* and *B. subtilis* (Perez-Rueda and Collado-Vides, 2000, Moreno-Campuzano *et al.*, 2006), and they are also common in eukaryotes as well (Ma, 2005, Boyle and Despres, 2010). They can be divided into dual regulators, which have different activities depending on the context of the DNA binding site, or those that are depending on oligomerization states and posttranslational modifications, e.g., phosphorylation, sumoylation or acetylation (Boyle and Despres, 2010). As PtAUREO1a has its own blue-light dependent signaling domain and might form both homo- and heterodimers, the latter seems to be more likely.

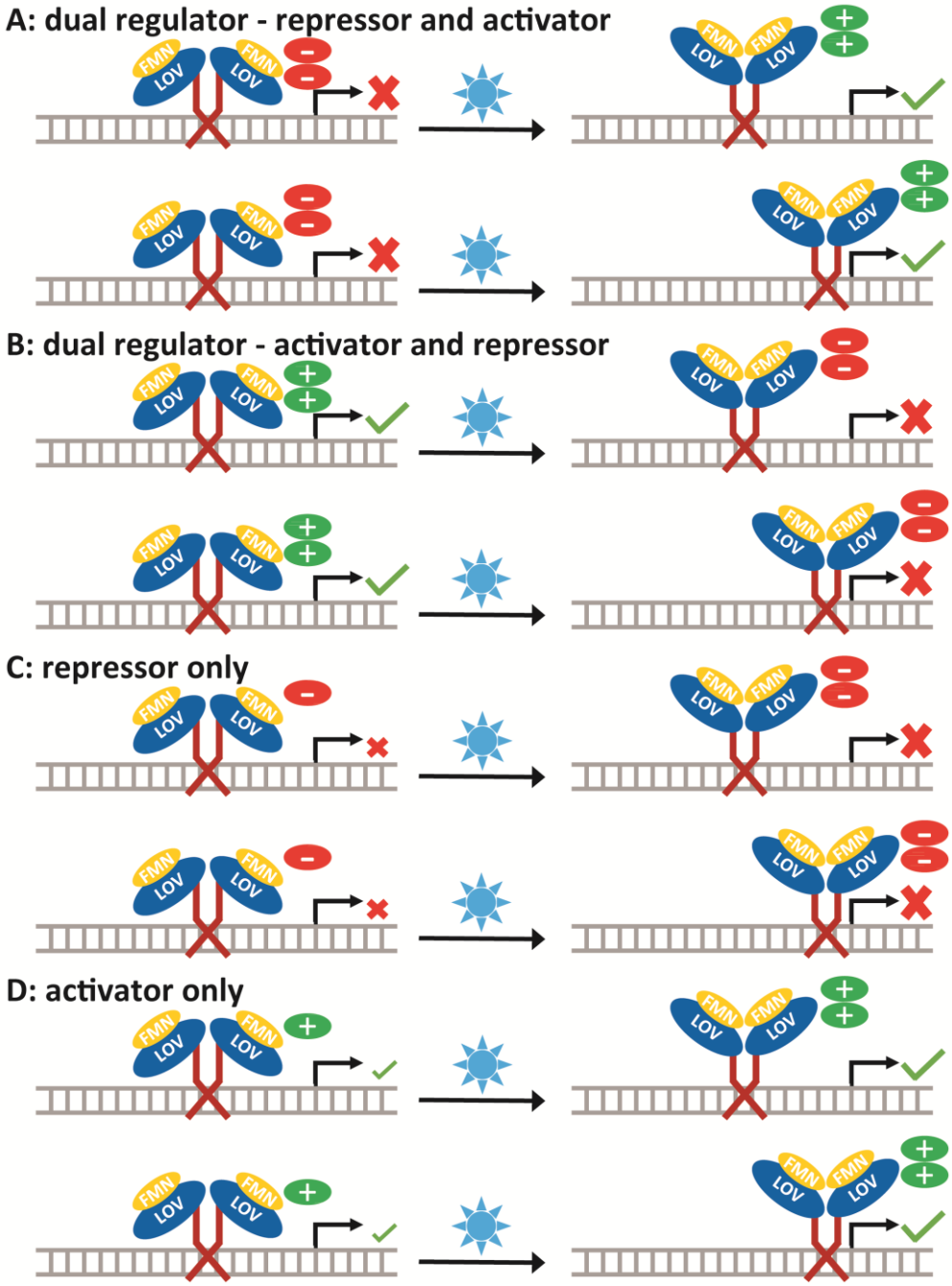


Figure 8-2: Potential mechanisms to explain the observed blue-light independent functions of PtAUREO1a. The conformational change upon blue-light illumination could either lead to attraction of different regulatory co-factors or a switch from one DNA-binding site to another. A/B: dual regulatory functions; C: repressor function only; D: activator function only. Putative co-activators are depicted as a green ellipse with a + sign, whereas putative co-repressors are depicted as a red ellipse with a – sign. Inhibitory effects on transcription are depicted with a red cross, stimulatory effects are depicted with a green check mark. Illumination with blue light is indicated by the cyan-colored sun.

8.1.1 The expression level of PtAUREO1a needs to be tightly regulated

Interestingly, PtAUREO1a knockdown *via* RNAi and knockout *via* TALEN let to partially opposite phenotypes. While previously studied knockdown strains showed an increase in NPQ capacity (Schellenberger Costa *et al.*, 2013b), the knockout strains showed a strong decrease (see Table 4-2). However, a mono-allelic knockout strain, which showed an increased level of PtAUREO1a, showed an increased NPQ capacity similar to the knockdown strain, indicating that the reduced NPQ capacity may not be an off-target effect eventually caused by the TALEN pair. Strongly different phenotypes in knockdown and knockout strains of the same protein have been previously observed in many model organisms (Gao *et al.*, 2015, Kok *et al.*, 2015). While potential off-target effects can never be excluded for both RNAi and genome editing techniques (Jackson *et al.*, 2003, Pattanayak *et al.*, 2013, Guilinger *et al.*, 2014a), these most likely should not account for all of the strong differences observed. A study by Rossi *et al.* on zebrafish, for instance, showed that knockdown of *egfl7* by either morpholinos or CRISPRi leads to severe vascular defects, which were not observed in *egfl7* knockout strains. Injecting the morpholinos into the knockout strain, however, did not result in the same severe phenotype observed for injection of the morpholinos into the wild type. This led to the identification of a set of genes upregulated in the knockout but not the knockdown strain that could rescue the phenotype. This finding indicates that there is a compensatory network to circumvent deleterious mutations, which is not active after translational or transcriptional knockdown (Rossi *et al.*, 2015). If such a compensatory network should be active in the case of PtAUREO1a, however, it does not lead to a rescued phenotype but instead to an opposite phenotype. Thus, a direct comparison of the knockdown and knockout strains would be of interest to identify the differently regulated genes, however, the knockdown strains do not show reduced protein levels anymore (data not shown) and would need to be generated again. Alternatively, the RNAi plasmid could be expressed in the knockout strains or wild type PtAUREO1a expressed at different levels in the knockout strain to exclude potential off-target effects of the RNAi or the TALEN construct, respectively.

Similarly, an identical phenotype in knockdown and overexpression strains has been observed frequently for regulatory proteins like kinases, GTPases, transcription factors and others (Guilherme *et al.*, 2004, Morino *et al.*, 2004, Strick and Elferink, 2005, Magadan *et al.*, 2006, Park *et al.*, 2009, Bernick *et al.*, 2010, Wang *et al.*, 2014). The main reason for this observation is supposed to be caused by protein heterooligomer formation, whose stoichiometry is disrupted by overexpression or downregulation of one of the protein

involved, or competition for the same subunit with other proteins (Park *et al.*, 2010). This finding is a further indication that PtAUREO1a might also form different heterodimers to regulate target genes. Thus, the PtAUREO1a protein level needs to be tightly regulated to perform its intended function of providing a balanced regulation of photoprotection parameters. While the RNA-seq data suggest a negative feedback of PtAUREO1a leading to inhibition of transcription of the *PtAureo1a* gene upon blue-light illumination (see Chapter 5), a mostly light-independent circadian rhythm of *PtAureo1a* transcripts was observed throughout the day-night cycle (see Chapter 2), indicating involvement of other transcription factors in its regulation as well. Thus, reverse ChIP (also called PICh), i.e. the use of DNA baits to capture interacting proteins followed by identification *via* mass spectrometry (Dejardin and Kingston, 2009), could be attempted to identify the regulators of this putative master switch of the light acclimation pathway.

8.2 What is the role of the other PtAUREOs?

While the PtAUREO1a protein seems to act as a master switch of short term adaptation to blue light and photoprotection (see Chapters 4 and 5), much less is known about the functions of the other Aureochromes. Thus, knockout strains of the other AUREO isoforms have been generated as well, which currently are in different states of screening. Suitable candidate cell lines have been identified for all of them that need to be confirmed by further screening steps (see Chapter 6). The known properties of the different isoforms regarding regulation, biophysical properties and their presumed role within the cell are summarized in Table 8-1.

The chain-forming phenotype discovered in a PtAUREO1b knockout strain (see Figure 6-7 to 6-9) indicates an involvement in cell cycle progression and/or cell division, causing an incomplete detachment of daughter cells. However, as this phenotype has also been observed independently in wild type and transgenic lines (Coughlan, 1962, Borowitzka *et al.*, 1977, Gherardi *et al.*, 2016), more knockout strains need to be identified to confirm this phenotype to be due to loss of PtAUREO1b and not due to random integration of one of the plasmids into the genome. Then, transcript levels of the major players of the cell cycle, i.e. the cyclins (*cyc*) and cyclin-dependent kinases (*cdk*), as well as the anaphase promoting complex (*apc*) and its activators *cdc20* and *cdh1* (Huysman *et al.*, 2010, Huysman *et al.*, 2014) should be assayed by qPCR using synchronized cells to determine potential deregulation of one or more of these genes. Alternatively, biofilm formation, which is usually not observed in the fusiform

morphotype, could be another reason for the observed phenotype. Enzymatic digestion of the cell chains with a combination of proteases and gluconases could be used to determine if biofilms are formed by the PtAUREO1b knockout strain.

While no visible phenotype has yet been detected for PtAUREO1c knockout mutants, this photoreceptor features some interesting characteristics. It seems to be co-expressed with PtAUREO1a in a mainly light-independent circadian rhythm when grown under low intensity white light conditions (see Figure 2-1) and was shown to be capable to form heterodimers with PtAUREO1a (Banerjee *et al.*, 2016b). On the other hand, its transcript was found to be strongly induced upon illumination of red-light adapted cells with blue light, whereas transcripts of all other Aureochrome isoforms were found to be downregulated (see Chapter 5). Thus, it seems to be mainly involved in longer term adaptation to blue-light exposure, but confirmed knockout strains are required to elucidate this function.

The role of PtAUREO2 is unclear yet. It has been shown to be unable to bind to the FMN cofactor due to steric constraints caused by a single point mutation (V301M). As it accordingly is not photo-switchable, several different possibilities can be imagined. As its expression remains constant throughout the day-night cycle, which is in contrast to all other Aureochromes (see Chapter 2), it could perform its functions either constitutively or be inactive. However, exposure of red-light adapted cells to blue light resulted in very strong downregulation of its transcript within 10 min, indicating it to be indirectly light-regulated, presumably by PtAUREO1a. Alternatively, PtAUREO2 might depend on heterodimerization with the other FMN-binding Aureochrome isoforms to form an active transcription factor. Finally, the mutated LOV domain might not be sensitive to light but to other triggers, e.g. the redox state or oxidative stress (Losi and Gartner, 2017). Thus, expression of mutated PtAUREO2 isoforms capable of FMN-binding or altered dimerization properties of the bZIP domain might be required to identify the role of this enigmatic protein.

Table 8-1: Summary of known biophysical properties, regulation and presumed role of the different PtAUREO isoforms. Dimerization states confirmed *in vitro* are indicated in bold (Banerjee *et al.*, 2016b), others are inferred from bioinformatical predictions (see Chapter 2).

	FMN	Circadian rhythm	Blue-light exposure	dimerization	Role
PtAUREO1a	Yes	Light-independent	Downregulation	**1a**/1b/**1c**/2	Light acclimation
PtAUREO1b	Yes	Light-dependent	Downregulation	1a/1b/2	Cell division?
PtAUREO1c	Yes	Light-independent	Upregulation	**1a/1c**	Unknown
PtAUREO2	No	None	Downregulation	1a/1b/2	Unknown

8.3 TALEN and its potential applications

The TALEN method for *P. tricornutum* established in Chapter 5 was found to be highly efficient and, in theory, can be used to target almost any site within the genome. Thus, the limitations of RNAi, which is the most commonly used reverse genetics approach in the diatom *P. tricornutum* until now, can be overcome: Many RNAi lines have been found to be instable, whereas bi-allelic mutations by TALEN are irreversible when the recognition site of both alleles has been mutated enough to prevent further binding of the TALEN pair in close proximity. As we very often could observe big insertions or deletions of several 100 bp in length, these mutations are irreversible. Additionally, residual protein present in knockdown lines can prevent the identification of more pronounced phenotypes. Interestingly, we observed that mono-allelic knockout strains did not change their genotype over the course of a year. As expression of the TALEN proteins was never assayed by Western Blot, it is possible that expression of one or both TALENs is strongly silenced by the cell *via* an unknown mechanism that also leads to the previously observed reversion of RNAi-based knockdown mutants (Lavaud *et al.*, 2012). Another possibility is that bi-allelic mutants arise, but are strongly outcompeted by the mono-allelic line, thereby resulting in a population of mostly mono-allelic cells. Hence, isolation from single cells might prove successful, however, the effort required might be too high when compared to screening of additional transformants.

While both mono-allelic and bi-allelic knockout lines were found to be genetically stable, the protein expression of mono-allelic PtAUREO1a knockout mutants unfortunately proved to be as instable as the respective RNAi lines. A pronounced upregulation was observed after the screening process, leading to wild type expression levels or even slight overexpression of the target protein (see Figure 4-6 and Figure S5-1). This effect, termed “haplosufficiency” has been observed in many other organisms before. Extensive studies on humans, yeast, *Drosophila melanogaster* and *Arabidopsis thalina* revealed that “haplosufficiency” is a much more common occurrence than haploinsufficiency (Deutschbauer *et al.*, 2005, Dang *et al.*, 2008, Cook *et al.*, 2012, Meinke, 2013). Thus, mono-allelic strains cannot be used as an alternative to RNAi for target proteins where complete loss of the protein would be lethal without determining protein levels before and after each experiment.

The TALEN system itself, while highly efficient, could still be improved. The cells from the current protocols continuously express a nuclease, which over time might lead to unintended off-target mutations within the genome. This could be prevented by either identifying

inducible promoter elements that are not active during standard culturing conditions or by expression from an episome, which would have the advantage that it is not stably integrated into the genome and accordingly is lost after the selection pressure is relieved (Karas *et al.*, 2015, Diner *et al.*, 2016). The potential applications surpass the use for generation of knockout mutants by random mutagenesis, especially by using homologous recombination templates (see Figure 8-3). One example would be the introduction of a tag into the protein which allows purification and/or detection of the protein *via* tag-specific antibodies, circumventing the need to produce expensive antisera directed against the proteins of interest. If a method requires the utilization of native protein, for example for Co-Immunoprecipitation or Chromatin Immunoprecipitation, the necessary generation and validation of high quality antisera is very laborious and not always successful (see Chapter 2). While it would be possible to introduce a tagged variant in addition to the native protein, the resulting overexpression however might cause artefacts (Kolodziej *et al.*, 2009). Additionally, targeted mutations of genes could be introduced by TALEN as well. For example, a strain expressing a PtAUREO1a protein unable to bind FMN or unable to propagate the conformational change caused by the absorption of blue light by the FMN co-factor would allow investigation of whether the protein performs a role under non-blue light conditions as well as blue-light conditions. Alternatively, the leucine zipper domain could be modified to only allow specific combinations of homo- or heterodimerization to elucidate which functions are performed by which type of dimer. Finally, the function of the long N-terminal extension of the Aureochromes, which varies considerably in length and has very low sequence homology between the isoforms, could be studied with N-terminal truncations of different lengths.

In addition to TALEN, similar potential approaches could be based on the CRISPR/Cas9 system as well, which was recently established for *P. tricornutu*m (Nymark *et al.*, 2016). In contrast to TALEN, Cas9 is insensitive to DNA methylation and also provides a more cost-effective approach for multiplexing by expression of several guide RNAs from a single plasmid into a strain expressing the Cas9 protein (Hsu *et al.*, 2013, Minkenberg *et al.*, 2017). Thus, multiple gene knockouts could be more easily achieved using this alternative genome-editing system. Furthermore, the problem of increased probability of off-target effects has been partly resolved: Cpf1, an alternative CRISPR protein, seems to tolerate fewer mismatches within its guide RNA than Cas9 (Kim *et al.*, 2016, Kleinstiver *et al.*, 2016b). Additionally, mutated Cas9 variants with increased specificity or with an inactivated catalytic domain (termed dCas) fused with FokI nuclease, which require binding of two dCas proteins

in close proximity, have also been reported to ameliorate this problem (Guilinger *et al.*, 2014b, Kleinstiver *et al.*, 2016a). For future studies, the choice of targeted nuclease should therefore be influenced by the kind and amount of different desired mutations.

Hence, TALEN, CRISPR, and other genome editing techniques represent very powerful tools for functional characterization of Aureochromes apart from the generation of knockout strains.

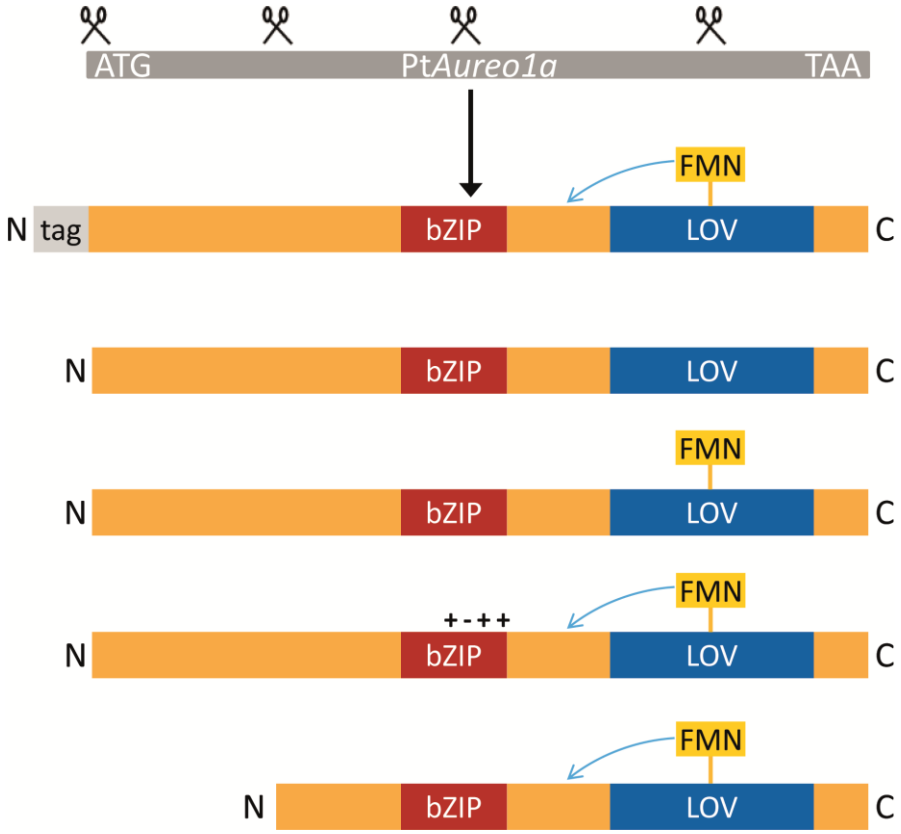


Figure 8-3: Potential applications of TALEN-based genome-editing besides generation of knockout strains using the Pt*Aureo1a* gene as an example: Introduction of variants fused with an epitope tag for Co-IP or ChIP, non-photoswitchable variants to study potential functions in absence of blue light, modification of the bZIP domain to influence dimerization specificity and truncated variants to elucidate the function of the N-terminal extension. Cutting sites for the TALENs are indicated by the scissors symbol. The bZIP domain is depicted in red, the LOV domain in blue and the epitope tag in light grey. The blue arrow indicates the signal transduction upon blue-light illumination. +/- symbols indicate modifications of the charged residues of the leucine zipper domain, which influence the dimerization properties of the Aureochrome.

8.4 Conclusions and outlook

There have been many different molecular approaches trying to gain more insight into the function of Aureochromes of *Phaeodactylum*: Specific antisera are available for all isoforms allowing quantification of protein expression levels of all PtAUREOs (see Chapter 2), TALEN has been established as a highly efficient tool to create knockout strains, which

should be further adapted to create strains with desired mutations using homologous recombination templates (see Chapter 4). Furthermore, analyses showed a strong influence of PtAUREO1a on the transcriptome, indicating a role as a master switch for light acclimation and photoprotection (see Chapter 5). The strongly reduced NPQ capacity observed in PtAUREO1a knockout strains seems to be caused by a direct regulation of the Lhcx1 expression by PtAUREO1a, which should be confirmed by Yeast 1 Hybrid assays. Additionally, the way in which PtAUREO1a may affect transcription in absence/presence of blue light remains to be elucidated. While the Bind-n-Seq approach to identify Aurechrome-binding sites did not lead to a conclusive result yet, it could be refined as a basis for further research (see Chapter 3).

The complex light regulation network of the Aureochromes in conjunction with other photoreceptors and/or interacting transcription factors still leaves many open questions that need to be answered to fully understand Aureochrome function. In contrast, the yet unknown regulatory pathway of the light-insensitive PtAUREO2 could potentially provide an interconnection between redox regulation and light-sensing pathways. Additionally, other secondary regulation effects *via* posttranslational modifications could further modulate activity or provide feedback loops to differentiate between short-term or longer-term light exposure or interconnect different regulatory pathways. Thus, large scale protein-protein interaction studies focused on bZIP transcription factors, potential co-activators/repressors, and other regulatory enzymes will provide essential information to unravel the complex interconnected regulatory pathways of the cell in response to blue-light triggers.

A. Supplementary data

Supplementary data, chapter 2

>PtAUREO1a
MTDNNKSLSAHAQAAVTNRGNPATLNLDDIFGDVMFTPDGDTVFMSEQKEELLNSGEREV
TTMASKATQDGQYQPVQQGGGLYTTQLYDNSKPALTMGVAGGINVQATAPVPYKSAPQAT
HHLQYAAPKKKSSSSSTSGSGSRSDRKMSEQQKVERRERNREHAKRSRIRKKFLLESLQQ
SVSLLKEENEKLKTSIRSHLGEEKADTLIDSANNNKTDVDGLLASSQGIANKVLDDPDFS
FIKALQTAQQNFVVTDPSLPDNPIVYASQGFLNLTGYSLDQILGRNCRFLQGPETDPKAV
ERIRKAIEQGNDMSVCLLNYRVDGTTFWNQFFIAALRDAGGNVTNFVGVQCKVSDQYAAT
VTKQQEEEEEAAANDDED

>PtAUREO1b
MDDFDLNEIFAEYFTNEFDDPLNAYTSSMANHNGAVTAVAMTQNDTVPTEGRKLAQTGGL
TLPTGGIRTTFHTAAIQKAPLITDGTNPVTKKQKTDEQSQQQHQQLLHVQNPLQQQQQQA
LAAAHNSAMQAHQQTQNASNMQHPGAAMVQGGMVSLPVGVGIRLGGIGGIAPATAQSATR
TSGVPGQFNMWGPGTGGMSEQAVAERRQRNREHAKRSRVRKKFMLESLQEQVREMQKQNQ
NLRMLVQEHIPEHAMKIIAECCTSSPLFEEMDGIDQTKGANLERADFSLMQSLTMGQQCF
VLSDPKLPDNPIVFASPGFYKLTGYTSREVLGRNCRFLQGPGTDAKAVDVIRKAVGTGSD
ATVCLLNYKADGTPFWNQFFIAALRDSDNCIVNYVGVQTEVEPQAGVSMLEDKVNAILPL
QTKDDSSE

>PtAUREO1c
MADQAKANPSGPSNTAPLFQMSDPANPSSTLDLDDLFNDIYYMDLPNGSNMDSDPTPVDV
PSSNGNGHDGASKKRSADDFDGDSLDGTGNKDLTEQQKLERRERNREHAKRSRLRKKFLL
ESLQEQIHGLEEQLDGLKSAIKKELPQQAEQIITRICGDKEKFTPLPMPSGFGPVKTLME
PDFRLMSALSGSQQNFAISDPTLPDNPIVYVSQGFLDLTGYTLDQVLGRNCRFLQGPGTD
QSAVEVIRKGITEGVDTSVCLLNYKADGTPFWNQFFVASLRDAENNVVNHVGVQCEVSKA
VVEKHMGEQKKAAAEAAKARPVTTTSS

>PtAUREO2
MAQNLQMPRFGRANNTSGADTWGDCDAFDVDMLTEYLLNDGTLTSSGVTFDFNMDGAAHL
SSTVSPENSEDGALPTVADSNEISAEVQKFAASHEGPSYVAVATSGMPIAPSPSLSFPAL
PTQVSMPGNDGQGLVFHHHQQQHANSTATSKRRRIDGMSGSALLAVSGGPTGSAFLGGDQ
LAAAAARAIQNQGRGRKKSQAQIDRRRERNRILARRTRLRKKFFFESLQKEIMDLQRENL
VLKELVKVNISGEEGKKILEGCNAAENLPSSVLEACGEENDMDSQDFNLVRSIQSSQHSF
MITDPSLQDNPIVFASDDFLKLTGYTREQVLGRNCRFLQGTETSQEKVNQIRKNLSEGED
VTVTLMNYTADGTPFWNKLFIAALRDAQNNIVNFIGVIVKVARPEPGDPEHDKGNESEQN
DGDESDEDAEDTVRAIEGAVTAAVAAAGRSS

Figure S2-1: Protein sequences of the four Aureochrome isoforms of *P. tricornutum*. Red background: bZIP domain; cyan background: LOV domain; yellow background: peptide chosen for generation of a specific antisera. Due to the overlap of the LOV domain of PtAUREO2 with the peptide chosen for immunization, the last amino acid of the LOV domain is indicated in light green.

Supplementary data, chapter 3

Supplemental dataset S3-1 (see attached CD-ROM): Table of hits for putative PtAUREO1a and PtAUREO1c binding motifs (1321 and 321, respectively) within the promoter regions. Protein ID, location of the gene within the genome, number of hits per gene, location and orientation of the binding motif (in relation to promotor) are given. If available, annotations (gene ontology, interpro domains, other domains and user annotations) are given.

Supplementary data, chapter 4

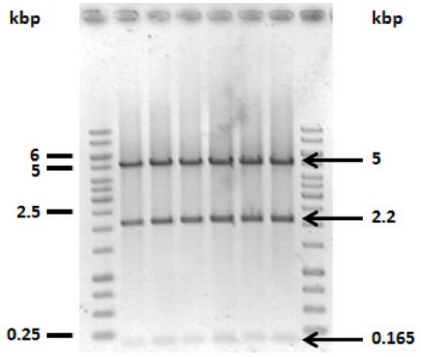


Figure S4-1: Verification of assembled TALEN plasmids *via* restriction digest using AfeI (expected fragment lengths: 5 kbp, 2.2 kbp and 165 bp).

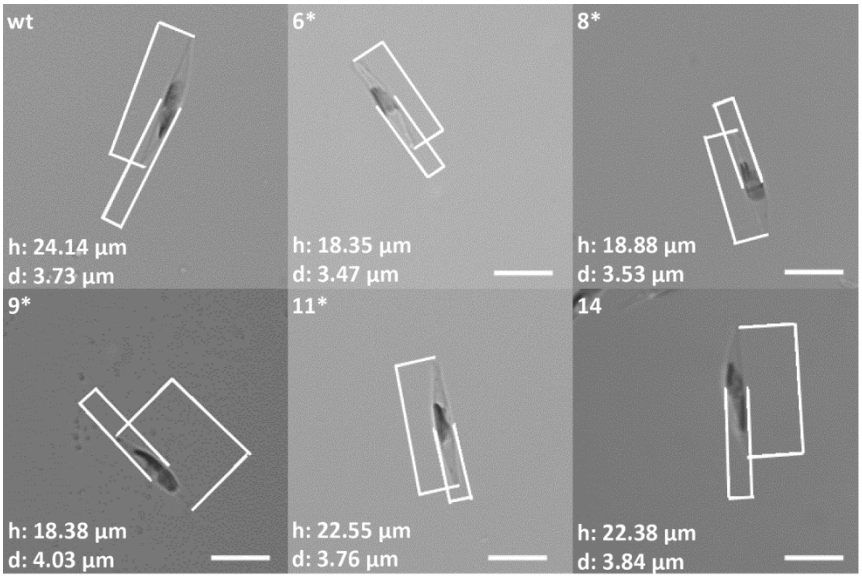


Figure S4-2: Cell length (h) and width (d) of 50 cells for each strain were determined microscopically. An example picture is given for each cell line assayed. The white line in the bottom-right corner is 10 µm in length.

Table S4-1: Primers used for this study not described elsewhere.

Primer Name	Application	Sequence (5'→3')
TALEN_HindIII+_for	Add HindIII site into TALEN plasmids from Addgene	CCTTGACCCTGGAAGCTTCCACTCCCACTG
TALEN_HindIII+_rev		CAGTGGGAGTGGAAGCTTCCAGGGTCAAGG
PTV_BSAI1719SD_for	Deletion of BsaI restriction site at position 1719 of P.t. plasmids	TTTCTCGAGGCCGGACTCCCTATAGTGAGT
PTV_BSAI1719SD_rev		ACTCACTATAGGGAGTCCGGCCTCGAGAAA
PTV_BSAI2888SD_for	Deletion of BsaI restriction site at position 2888 of P.t. plasmids	GCAATGATACCGCGGGACCCACGCTCACCG
PTV_BSAI2888SD_rev		CGGTGAGCGTGGGTCCCGCGGTATCATTGC
Aureo1a_probe_for	Synthesis of DIG-labeled Southern Blot probe	CAAGGCATCGCCAACAAGG
Aureo1a_probe_rev		GTGACTGTGGCGGCGTATTG
Aureo1a _for	PCR amplification of TALEN target site of PtAUREO1a in the mutants	ATGACCGACAACAACAAG
Aureo1a _rev		GTTGCTTGGTGACTGTG

Tables S4-2a-c show the raw data of microscopical cell size determination of *P. tricornutum* wild type (wt) and four bi-allelic transformant strains (6*, 8*, 9* and 11*), as well as a mono-allelic transformant strain (14). Figure S2 shows an example image of each strain.

Table S2a: Cell length [µm] of 50 cells of each cell line, which were measured microscopically.

wt	6*	8*	9*	11*	14
24.15	18.35	21.89	17.05	21.07	21.7
23.87	15.81	21.06	19.37	22.55	21.51
24.66	19.04	19.1	16.62	20.61	23.2
23.03	20.25	17.35	18.61	22.52	20.5
27.6	19.54	17.16	19.19	18.92	22.9
22.27	18.18	19.45	16.03	14.2	22.36
25.47	19.41	17.22	15.47	21.81	22.49
22.59	21.54	18.55	16.95	24.08	22.78
27.95	16.23	16.41	17.73	21.77	22.51
24.88	16.97	16.34	18.6	17.59	22.51
19.84	20.17	18.92	18.48	22.87	24.06
23.67	17.1	17.55	17.3	16.49	22.69
24.11	18.5	17.86	15.92	20.35	21.69
23.4	17.83	22.52	21.55	22.75	24.7
23.89	17.69	18.78	18.16	26.72	23
27.58	18.19	18.04	17.43	23.52	22.68
27.14	22.06	15.74	21.85	17.54	21.47
24.11	17.83	22.62	19.03	19.41	22.38
23.08	19.71	19.38	17.67	26.06	22.78
24.66	18.03	16.67	18.5	22.43	18.26
24.51	17.07	16.25	18.15	22.14	21.28
24.51	16.2	18.2	15.8	20.25	22.86
23.08	16.32	19.72	16.04	22.5	21.72
23.87	19.44	17.13	17.34	21.37	20.21
23.72	18.61	15.9	16.15	22.88	19.68
22.92	15.73	17.23	18.38	20.14	21.98
22.58	20.02	22.53	18.36	20.08	21.04
22.64	18.34	19.74	18.56	19.51	22.84
22.3	19.97	21.55	18.08	20.75	20.76
24.67	18.31	16.78	15.11	22.72	19.61
29.64	20.09	14.84	16.47	22.88	20.1
20.74	19.55	19.85	15.8	24.66	24.11
29.02	18.31	18.35	14.84	27.84	23.5
24.19	17.89	14.31	14.73	24.35	19.75
22.58	19.85	21.25	17.47	25.07	21.34
25.69	18.46	17.84	17.8	26.01	24.08
26.21	18.7	20.55	21.55	22.83	24.94
26.32	21.21	16.44	21.5	25.91	24.51
24.97	16.55	16.22	18.52	20.28	23.17
22.41	17.57	19.18	16.64	25.19	22.96
30.14	18.34	20.76	19.47	22.03	21.81
27.06	18.38	16.5	18.69	22.52	20.91
25.51	17.67	19.36	18.61	24.54	22.3
21.66	19.07	17.47	18.58	22.38	19.52
24.11	18.44	21.08	19.94	21.27	21.95
27.49	16.69	21.23	16.31	24.04	23.43
20.66	18.51	19.08	17.47	20.97	22.17
22.82	19.09	20.74	19.36	16.62	23.12
22.41	17.05	19.23	17.73	20.36	22.8
23.46	20.21	19.8	19.99	23.62	22.38
24.40±2.26	18.48±1.45	18.63±2.03	17.90±1.71	21.98±2.74	22.14±1.44

mean:

Table S4-2b: Cell width [µm] of 50 cells of each cell line, which were measured microscopically.

	wt	6*	8*	9*	11*	14
	3.73	3.47	3.8	3.74	3.74	4.01
	4.03	3.43	3.43	3.65	3.76	3.74
	3.44	3.56	3.96	3.13	3.54	3.98
	3.53	3.96	3.6	3.6	3.53	3.5
	3.74	3.76	3.27	4.12	3.8	3.67
	3.88	3.84	3.74	4.26	3	3.8
	3.68	4.12	4.14	3.1	3.93	4.07
	4.12	3.6	3.54	3.85	3.89	3.89
	4.03	3.88	3.4	4.22	3.4	3.89
	3.9	3.65	3.9	3.95	3.64	3.89
	3.74	3.78	4.03	3.93	3.96	3.96
	3.83	3.48	4.12	3.73	3.62	3.84
	3.8	3.7	3.54	3.85	3.28	3.95
	3.52	4.01	2.83	3.4	3.73	3.96
	3.76	3.18	3.42	3.14	3.35	3.85
	3.64	3.4	3.84	4.4	3.8	3.48
	4.12	3.89	2.95	3.8	3.44	3.81
	3.8	3.17	3.32	4.03	4.58	3.98
	4.01	3.02	3.89	3.14	3.54	4.22
	3.44	3.44	3.73	4.2	3.5	3.74
	3.87	3.37	3.6	4.32	3.56	3.8
	3.87	4.07	4.12	4.22	3.87	4.14
	3.56	3.61	4.18	3.93	3.5	3.5
	4.03	3.9	3.53	3.34	3.89	3.8
	3.5	2.79	3.57	4.37	3.44	4.21
	3.61	3.89	3.95	4.03	3.44	3.98
	3.28	4.4	3.78	4.03	3.5	4.03
	3.54	3.9	3.6	3.64	3.89	3.96
	3.5	3.58	3.93	3.78	3.57	3.67
	3.22	3.11	4.26	3.67	3.73	3.8
	3.54	4.17	3.58	3.73	3.69	3.78
	3.65	4.25	3.34	4.01	3.06	3.9
	4.37	3.44	3.67	3.7	4	4.06
	3.8	3.37	3.9	3.43	3.81	3.7
	3.28	4.01	3.88	3.17	4.03	3.52
	3.53	3.54	3.35	3.2	3.33	3.81
	4.25	3.57	3.61	3.35	3.84	3.98
	3.99	3.78	3.96	3.18	3.96	4.17
	3.73	3.07	3.18	4.21	3.65	3.33
	3.96	3.57	3.5	4.25	3.93	3.65
	4.01	4.18	3.33	3.44	3.92	4.03
	4.07	3.64	2.95	4.24	3.57	4.13
	3.67	3.66	3.89	3.8	3.29	3.89
	3.69	3.68	4.35	3.66	3.84	2.95
	3.4	3.17	3.92	3.31	3.69	3.8
	3.32	3.8	4.01	3.37	3.21	3.5
	3.76	3.87	3.9	3.84	3.35	3.9
	3.8	4.22	3.44	3.92	3.42	3.7
	3.96	3.8	4.03	3.3	4.07	3.9
	4.4	3.33	4.13	3.3	3.9	3.84
mean:	3.76±0.27	3.66±0.35	3.70±0.34	3.74±0.38	3.66±0.29	3.83±0.23

Table S4-2a: Cell volume [µm³] of 50 cells of each cell line, which were measured microscopically.

wt	6*	8*	9*	11*	14
87.96	57.84	82.75	62.44	77.16	91.35
101.49	48.70	64.87	67.56	83.46	78.77
76.40	63.17	78.41	42.63	67.62	96.21
75.13	83.14	58.87	63.14	73.47	65.74
101.07	72.32	48.04	85.28	71.52	80.75
87.77	70.18	71.22	76.16	33.46	84.53
90.30	86.26	77.27	38.92	88.19	97.53
100.39	73.08	60.86	65.77	95.39	90.24
118.84	63.97	49.66	82.66	65.88	89.18
99.07	59.19	65.07	75.98	61.02	89.18
72.65	75.45	80.45	74.72	93.89	98.78
90.90	54.22	77.99	63.01	56.57	87.59
91.15	66.30	58.59	61.78	57.32	88.60
75.90	75.06	47.22	65.22	82.86	101.40
88.42	46.83	57.51	46.88	78.50	89.25
95.67	55.05	69.64	88.34	88.91	71.91
120.61	87.39	35.86	82.60	54.34	81.59
91.15	46.91	65.27	80.91	106.59	92.81
97.16	47.06	76.78	45.61	85.50	106.21
76.40	55.86	60.72	85.44	71.93	66.87
96.10	50.75	55.13	88.68	73.46	80.45
96.10	70.25	80.88	73.66	79.40	102.58
76.58	55.68	90.20	64.86	72.16	69.66
101.49	77.41	55.88	50.64	84.66	76.40
76.07	37.92	53.05	80.74	70.88	91.32
78.20	62.32	70.38	78.15	62.39	91.15
63.60	101.47	84.28	78.06	64.40	89.46
74.28	73.03	66.98	64.38	77.29	93.77
71.52	67.01	87.14	67.63	69.23	73.20
66.97	46.36	79.72	53.28	82.76	74.13
97.24	91.46	49.79	59.99	81.56	75.19
72.34	92.45	57.97	66.51	60.45	96.01
145.09	56.72	64.70	53.19	116.62	101.41
91.45	53.19	56.98	45.37	92.54	70.78
63.60	83.56	83.75	45.96	106.59	69.22
83.81	60.56	52.41	47.72	75.51	91.51
123.94	62.39	70.11	63.31	88.13	103.43
109.70	79.34	67.49	56.92	106.37	111.58
90.95	40.84	42.94	85.94	70.73	67.26
92.00	58.62	61.51	78.69	101.85	80.08
126.88	83.89	60.27	60.32	88.62	92.73
117.35	63.76	37.59	87.96	75.14	93.37
89.95	61.97	76.70	70.35	69.54	88.34
77.21	67.61	86.54	65.16	86.40	44.47
72.97	48.51	84.80	57.19	75.82	82.98
79.33	63.09	89.37	48.49	64.85	75.14
76.47	72.58	75.98	67.44	61.61	88.28
86.27	89.00	64.25	77.88	50.89	82.86
92.00	64.46	81.76	50.55	88.29	90.79
118.91	58.67	88.42	56.99	94.05	86.40
90.94±17.49	65.66±14.42	67.28±14.05	66.02±13.63	77.72±15.90	85.64±12.40

mean:

Supplementary data, chapter 5

Supplemental dataset S5-1 (See attached CD-ROM): Table of transcripts for wild type and PtAUREO1a knockout strain for the red-to-blue light shift RNA-seq experiment and their log2-fold changes and the corresponding FDR-adjusted p-value (cutoff: $p<0.01$) for each comparison.

Supplemental dataset S5-2 (See attached CD-ROM): Table of transcription factors derived from (Rayko *et al.*, 2010) with log2-fold changes and the corresponding FDR-adjusted p-value for each comparison of the RNA-seq shift experiment.

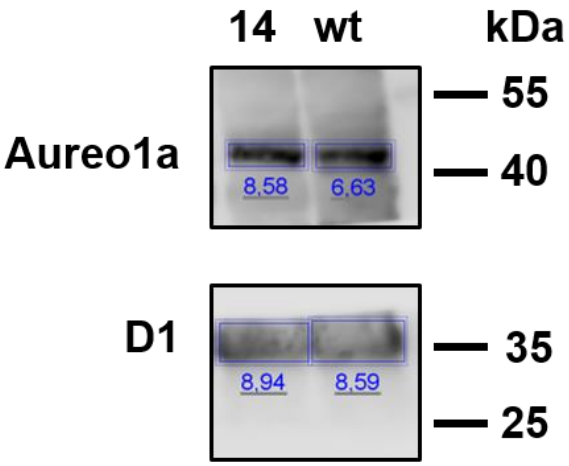


Figure S5-1: Quantification of PtAUREO1a expression level in wild type (wt) and a mono-allelic knockout strain (14) by Western Blot. Numbers below the bands correspond to signal intensity as calculated by the software LiCor Image Studio Lite (LiCor, Bad Homburg, Germany). The signal intensity of the D1 protein is used as an internal loading control.

B. Author contributions

Chapter 2

Conception and design of the experiments was done by MS and PGK. Transcript analysis, antibody validation and prediction of dimerization capabilities was carried out by MS. Data analysis was done by MS. The manuscript was drafted by MS and finalized by MS and PGK.

Chapter 3

Conception and design of the experiments was done by MS and PGK. The Bind-n-Seq experiment and elucidation of the binding motif was done by MS. The promoter analysis was done by MJ. Data analysis was done by MS and MJ. The manuscript was drafted by MS and finalized by MS, MJ and PGK.

Chapter 4

Conception and design of the experiments was done by MS, BL, PGK and CRB. Adaptation of the TALEN system to *P. tricornutum* was done by KW under supervision by CRB and BL. Generation and characterization of knockout strains was carried out by MS and BL. The data was analyzed by MS. The manuscript was drafted by MS and finalized by MS, BL, PGK and CRB.

Chapter 5

Conception and design of the experiments was done by MS, MM, CW and PGK. Cultivation of cells for the shift experiment was performed by MM. Extraction of RNA and quality control was performed by MS. Further data analysis after transcript mapping was carried out by MS and MM. The manuscript was drafted by MS and finalized by MS and PGK.

Chapter 6

Conception and design of the experiments was done by MS and PGK. Creation and genetic characterization of the knockout strains was done by MS. Light microscopy was performed by CRB. The manuscript was drafted by MS and finalized by MS, CRB and PGK

Chapter 7

Conception and design of the experiments was done by MS, CRB, PGK and BL. Creation and characterization of the knockout strains was done by MS and CRB. The manuscript was drafted by MS and finalized by MS, CRB, PGK and BL.

C. Acknowledgements

I would like to express my gratitude to the people who have contributed to the successful completion of this thesis:

Prof. Peter Kroth, for giving me the opportunity to do my PhD thesis in his group, the good supervision and guidance throughout. Nonetheless, you gave me the freedom to decide in which ways to develop this very interesting project.

All members of the Forschergruppe FOR1261, especially Marcus Mann and Prof. Christian Wilhelm (Universität Leipzig), Dr. Elena Herman and Dr. Tilman Kottke (Universität Bielefeld), Dr. Ankan Banerjee and Prof. Lars-Oliver Essen (Universität Marburg) and Dr. Marion Eisenhut and Prof. Andreas Weber (Universität Düsseldorf) for fruitful cooperation, as well as the DFG and the Universität Konstanz for funding and resources.

Caro, for being the not-so-secret self-proclaimed "democratic dictator" of the group who keeps the lab organized. Either you knew the answer to my questions or could tell me whom to ask instead, no matter whether the question was "How can I do that?", "Where is that?" or "Has anybody in our lab ever done that?". Your help and advice, especially regarding molecular biology, was greatly appreciated. It definitely wasn't "all equally"!

Bernard, for all the patience when explaining things about photosynthesis and photoprotection, your help with PAM fluorometry and pigment analysis *via* HPLC, and entrusting me to supervise your Vertiefungskurs project when you were on parental leave, allowing me to "inherit" it as a side project.

Doris, for transforming Phaeo with all my plasmids, as well as maintenance of the resulting clones. Without you diligently taking care of them every time they needed to be re-plated I am sure some important ones would have gone missing sooner or later…

The students who assisted me throughout my studies as part of their Bachelor's thesis, Master's thesis, Vertiefungskurs or as a "Hiwi": Teresa Hagen, Marc Halder, Mirta Jacobs, Simon Kienle, Dorothea Kleinpeter, Zeno Riester, Vincent Spegg and Florian Weeber. You did not only perform many experiments and provided valuable input, you also challenged and thereby (hopefully) improved my experimental planning skills.

Current and former members of the AG Kroth for the nice working atmosphere, fruitful discussions about science, as well as fun times inside and outside the lab: Alex for always having an open ear for scientific and non-scientific problems, going to concerts, enjoying barbecues and beers and the creation of Algaeman, Angelika for taking care of administrative things, Annette for midipreps, electroporations and shared trips to the special waste depot, Ansgar for helpful advice regarding bioinformatics, Chao for chinese dumplings and teaching me both useful and useless phrases in Chinese (somehow I can remember the useless stuff much better...), Guilan for hot pot and being funny without realizing it, Jan and Jens (nukular!) for pre-work coffee/breakfast/discussion sessions, Jochen for constantly "stealing" and re-labeling eppi racks, pens, scissors and more, as well as taking over my apartment during my time in Toulouse, Julia for telling me about the available PhD position and the shared interest of teasing Martin, Lachy for his cheerful Aussie nature as well as proofreading parts of my thesis due to his "nativespeakeredness", Lili for bao-baos, reverse-engineering the perfect cheesecake and selling me the best sofa ever, Loni for pre-work coffee/breakfast/discussion sessions and watching countless football games and Game of Thrones episodes together, Marc for lots of minipreps and stacking of pipette tips, as well as all the funny stories about things that have happened to him, Martin for supplying me with greek olive oil and oregano, Mirta for trips to the Summer Breeze festival and the design of Algaeman, Masha for her cheerful nature, Miriam for continued discussions about the definition of room temperature, Norico for introducing me to dinotoms, "Sachsias" for introducing me into my project as well as board game evenings and archery courses, and Shvaita for continuing the Aureochrome project.

I really enjoyed my time on M9 with its welcoming and friendly atmosphere between the different groups, and some legendary parties in the coffee corner or the botanical garden. The weekly badminton session (thanks to everybody who participated and special thanks to Matthias and Dietmar for their persistence in asking me to join until I finally agreed!) provided lots of fun and interesting discussions with beer and pizza afterwards.

Last but not least, I would like to thank my family, who always supported me and without them this thesis would never have been written.

D. List of publications

Publications

05/2016 (published) — Banerjee, A., Herman, E., Serif, M., Maestre-Reyna, M., Hepp, S., Pokorny, R., Kroth, P.G., Essen, L.O. and Kottke, T. (2016) "Allosteric communication between DNA-binding and light-responsive domains of diatom class I aureochromes". *Nucleic Acids Res.*, **44**, 5957-5970.
DOI: 10.1093/nar/gkw420

02/2017 (published) — Serif, M., Lepetit, B., Weißert, K., Kroth, P.G. and Rio Bartulos, C. (2017) "A fast and reliable strategy to generate TALEN-mediated gene knockouts in the diatom *Phaeodactylum tricornutum*". *Algal Research*, **23**, 186-195.
DOI: 10.1016/j.algal.2017.02.005

05/2017 (published) — Mann, M., Serif, M., Jakob, T., Kroth, P.G., Wilhelm, C. "PtAUREO1a and PtAUREO1b knockout mutants of the diatom *Phaeodactylum tricornutum* are blocked in photoacclimation to blue light". *Journal of Plant Physiology.*
DOI: 10.1016/j.jplph.2017.05.020

Presentations

25/02/2014 — "Light regulation in diatoms"
15. Wissenschaftliche Tagung der Sektion Phykologie 2014, Stralsund, Germany

08/03/2016 — "Molecular characterization of Aureochromes in the diatom *Phaeodactylum tricornutum*"
16. Wissenschaftliche Tagung der Sektion Phykologie 2016, Leipzig, Germany

Posters

07/07/2015 - 10/07/2015 — "TALEN-mediated knockout of AUREO1a and LHCX1 in *Phaeodactylum tricornutum*"
Molecular Life of Diatoms 2015, Seattle, USA

E. Bibliography

Apt, K.E., Kroth-Pancic, P.G. and Grossman, A.R. (1996) Stable nuclear transformation of the diatom Phaeodactylum tricornutum. *Mol Gen Genet*, **252**, 572-579.

Armbrust, E.V. (2009) The life of diatoms in the world's oceans. *Nature*, **459**, 185-192.

Bailleul, B., Rogato, A., de Martino, A., Coesel, S., Cardol, P., Bowler, C., Falciatore, A. and Finazzi, G. (2010) An atypical member of the light-harvesting complex stress-related protein family modulates diatom responses to light. *Proc Natl Acad Sci U S A*, **107**, 18214-18219.

Banerjee, A., Herman, E., Kottke, T. and Essen, L.O. (2016a) Structure of a Native-like Aureochrome 1a LOV Domain Dimer from Phaeodactylum tricornutum. *Structure*, **24**, 171-178.

Banerjee, A., Herman, E., Serif, M., Maestre-Reyna, M., Hepp, S., Pokorny, R., Kroth, P.G., Essen, L.O. and Kottke, T. (2016b) Allosteric communication between DNA-binding and light-responsive domains of diatom class I aureochromes. *Nucleic Acids Res*, **44**, 5957-5970.

Bernick, E.P., Zhang, P.J. and Du, S. (2010) Knockdown and overexpression of Unc-45b result in defective myofibril organization in skeletal muscles of zebrafish embryos. *BMC Cell Biol*, **11**, 70.

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

Borowitzka, M.A., Chiappino, M.L. and Volcani, B.E. (1977) Ultrastructure of a Chain-Forming Diatom Phaeodactylum-Tricornutum. *J Phycol*, **13**, 162-170.

Bowler, C., Allen, A.E., Badger, J.H., Grimwood, J., Jabbari, K., Kuo, A., Maheswari, U., Martens, C., Maumus, F., Otillar, R.P., Rayko, E., Salamov, A., Vandepoele, K., Beszteri, B., Gruber, A., Heijde, M., Katinka, M., Mock, T., Valentin, K., Verret, F., Berges, J.A., Brownlee, C., Cadoret, J.P., Chiovitti, A., Choi, C.J., Coesel, S., De Martino, A., Detter, J.C., Durkin, C., Falciatore, A., Fournet, J., Haruta, M., Huysman, M.J., Jenkins, B.D., Jiroutova, K., Jorgensen, R.E., Joubert, Y., Kaplan, A., Kroger, N., Kroth, P.G., La Roche, J., Lindquist, E., Lommer, M., Martin-Jezequel, V., Lopez, P.J., Lucas, S., Mangogna, M., McGinnis, K., Medlin, L.K., Montsant, A., Oudot-Le Secq, M.P., Napoli, C., Obornik, M., Parker, M.S., Petit, J.L., Porcel, B.M., Poulsen, N., Robison, M., Rychlewski, L., Rynearson, T.A., Schmutz, J., Shapiro, H., Siaut, M., Stanley, M., Sussman, M.R., Taylor, A.R., Vardi, A., von Dassow, P., Vyverman, W., Willis, A., Wyrwicz, L.S., Rokhsar, D.S., Weissenbach, J., Armbrust, E.V., Green, B.R., Van de Peer, Y. and Grigoriev, I.V. (2008) The Phaeodactylum genome reveals the evolutionary history of diatom genomes. *Nature*, **456**, 239-244.

Boyle, P. and Despres, C. (2010) Dual-function transcription factors and their entourage: unique and unifying themes governing two pathogenesis-related genes. *Plant Signal Behav*, **5**, 629-634.

Briggs, W.R., Christie, J.M. and Salomon, M. (2001) Phototropins: a new family of flavin-binding blue light receptors in plants. *Antioxid Redox Signal*, **3**, 775-788.

Chisti, Y. (2007) Biodiesel from microalgae. *Biotechnol Adv*, **25**, 294-306.

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

Christie, J.M., Reymond, P., Powell, G.K., Bernasconi, P., Raibekas, A.A., Liscum, E. and Briggs, W.R. (1998) Arabidopsis NPH1: a flavoprotein with the properties of a photoreceptor for phototropism. *Science*, **282**, 1698-1701.

Chu, L., Ewe, D., Rio Bartulos, C., Kroth, P.G. and Gruber, A. (2016) Rapid induction of GFP expression by the nitrate reductase promoter in the diatom Phaeodactylum tricornutum. *PeerJ*, **4**, e2344.

Coesel, S., Mangogna, M., Ishikawa, T., Heijde, M., Rogato, A., Finazzi, G., Todo, T., Bowler, C. and Falciatore, A. (2009) Diatom PtCPF1 is a new cryptochrome/photolyase family member with DNA repair and transcription regulation activity. *EMBO Rep*, **10**, 655-661.

Cong, L., Ran, F.A., Cox, D., Lin, S., Barretto, R., Habib, N., Hsu, P.D., Wu, X., Jiang, W., Marraffini, L.A. and Zhang, F. (2013) Multiplex genome engineering using CRISPR/Cas systems. *Science*, **339**, 819-823.

Cook, R.K., Christensen, S.J., Deal, J.A., Coburn, R.A., Deal, M.E., Gresens, J.M., Kaufman, T.C. and Cook, K.R. (2012) The generation of chromosomal deletions to provide extensive coverage and subdivision of the Drosophila melanogaster genome. *Genome Biol*, **13**, R21.

Coughlan, J. (1962) Chain Formation by Phaeodactylum. *Nature*, **195**, 831-832.

Crooks, G.E., Hon, G., Chandonia, J.M. and Brenner, S.E. (2004) WebLogo: a sequence logo generator. *Genome Res*, **14**, 1188-1190.

Crosson, S., Rajagopal, S. and Moffat, K. (2003) The LOV domain family: photoresponsive signaling modules coupled to diverse output domains. *Biochemistry*, **42**, 2-10.

Daboussi, F., Leduc, S., Marechal, A., Dubois, G., Guyot, V., Perez-Michaut, C., Amato, A., Falciatore, A., Juillerat, A., Beurdeley, M., Voytas, D.F., Cavarec, L. and Duchateau, P. (2014) Genome engineering empowers the diatom Phaeodactylum tricornutum for biotechnology. *Nat Commun*, **5**, 3831.

Dang, V.T., Kassahn, K.S., Marcos, A.E. and Ragan, M.A. (2008) Identification of human haploinsufficient genes and their genomic proximity to segmental duplications. *Eur J Hum Genet*, **16**, 1350-1357.

De Riso, V., Raniello, R., Maumus, F., Rogato, A., Bowler, C. and Falciatore, A. (2009) Gene silencing in the marine diatom Phaeodactylum tricornutum. *Nucleic Acids Res*, **37**, e96.

Dejardin, J. and Kingston, R.E. (2009) Purification of proteins associated with specific genomic Loci. *Cell*, **136**, 175-186.

Delwiche, C.F. (1999) Tracing the Thread of Plastid Diversity through the Tapestry of Life. *Am Nat*, **154**, S164-S177.

Depauw, F.A., Rogato, A., Ribera d'Alcala, M. and Falciatore, A. (2012) Exploring the molecular basis of responses to light in marine diatoms. *J Exp Bot*, **63**, 1575-1591.

Derks, A., Schaven, K. and Bruce, D. (2015) Diverse mechanisms for photoprotection in photosynthesis. Dynamic regulation of photosystem II excitation in response to rapid environmental change. *Biochim Biophys Acta*, **1847**, 468-485.

Deutschbauer, A.M., Jaramillo, D.F., Proctor, M., Kumm, J., Hillenmeyer, M.E., Davis, R.W., Nislow, C. and Giaever, G. (2005) Mechanisms of haploinsufficiency revealed by genome-wide profiling in yeast. *Genetics*, **169**, 1915-1925.

DiCarlo, J.E., Norville, J.E., Mali, P., Rios, X., Aach, J. and Church, G.M. (2013) Genome engineering in Saccharomyces cerevisiae using CRISPR-Cas systems. *Nucleic Acids Res*, **41**, 4336-4343.

Diner, R.E., Bielinski, V.A., Dupont, C.L., Allen, A.E. and Weyman, P.D. (2016) Refinement of the Diatom Episome Maintenance Sequence and Improvement of Conjugation-Based DNA Delivery Methods. *Front Bioeng Biotechnol*, **4**, 65.

Dolatabadi, J.E.N. and de la Guardia, M. (2011) Applications of diatoms and silica nanotechnology in biosensing, drug and gene delivery, and formation of complex metal nanostructures. *Trends Anal Chem*, **30**, 1538-1548.

Doyle, E.L., Booher, N.J., Standage, D.S., Voytas, D.F., Brendel, V.P., Vandyk, J.K. and Bogdanove, A.J. (2012) TAL Effector-Nucleotide Targeter (TALE-NT) 2.0: tools for TAL effector design and target prediction. *Nucleic Acids Res*, **40**, W117-122.

Fabris, M., Matthijs, M., Rombauts, S., Vyverman, W., Goossens, A. and Baart, G.J. (2012) The metabolic blueprint of Phaeodactylum tricornutum reveals a eukaryotic Entner-Doudoroff glycolytic pathway. *Plant J*, **70**, 1004-1014.

Fortunato, A.E., Jaubert, M., Enomoto, G., Bouly, J.P., Raniello, R., Thaler, M., Malviya, S., Bernardes, J.S., Rappaport, F., Gentili, B., Huysman, M.J., Carbone, A., Bowler, C., d'Alcala, M.R., Ikeuchi, M. and Falciatore, A. (2016) Diatom Phytochromes Reveal the Existence of Far-Red-Light-Based Sensing in the Ocean. *Plant Cell*, **28**, 616-628.

Fu, Y., Foden, J.A., Khayter, C., Maeder, M.L., Reyon, D., Joung, J.K. and Sander, J.D. (2013) High-frequency off-target mutagenesis induced by CRISPR-Cas nucleases in human cells. *Nat Biotechnol*, **31**, 822-826.

Gao, Y., Zhang, Y., Zhang, D., Dai, X., Estelle, M. and Zhao, Y. (2015) Auxin binding protein 1 (ABP1) is not required for either auxin signaling or Arabidopsis development. *Proc Natl Acad Sci U S A*, **112**, 2275-2280.

Gherardi, M., Amato, A., Bouly, J.P., Cheminant, S., Ferrante, M.I., d'Alcala, M.R., Iudicone, D., Falciatore, A. and Cosentino Lagomarsino, M. (2016) Regulation of chain length in two diatoms as a growth-fragmentation process. *Phys Rev E*, **94**, 022418.

Gilbert, M., Wilhelm, C. and Richter, M. (2000) Bio-optical modelling of oxygen evolution using in vivo fluorescence: Comparison of measured and calculated photosynthesis/irradiance (P-I) curves in four representative phytoplankton species. *J Plant Physiol*, **157**, 307-314.

Gong, C., Bongiorno, P., Martins, A., Stephanou, N.C., Zhu, H., Shuman, S. and Glickman, M.S. (2005) Mechanism of nonhomologous end-joining in mycobacteria: a low-fidelity repair system driven by Ku, ligase D and ligase C. *Nat Struct Mol Biol*, **12**, 304-312.

Gordon, R., Losic, D., Tiffany, M.A., Nagy, S.S. and Sterrenburg, F.A. (2009) The Glass Menagerie: diatoms for novel applications in nanotechnology. *Trends Biotechnol*, **27**, 116-127.

Goss, R. and Jakob, T. (2010) Regulation and function of xanthophyll cycle-dependent photoprotection in algae. *Photosynth Res*, **106**, 103-122.

Goss, R. and Lepetit, B. (2015) Biodiversity of NPQ. *J Plant Physiol*, **172**, 13-32.

Grigoryan, G. and Keating, A.E. (2008) Structural specificity in coiled-coil interactions. *Curr Opin Struct Biol*, **18**, 477-483.

Gruber, A., Rocap, G., Kroth, P.G., Armbrust, E.V. and Mock, T. (2015) Plastid proteome prediction for diatoms and other algae with secondary plastids of the red lineage. *Plant J*, **81**, 519-528.

Guilherme, A., Soriano, N.A., Bose, S., Holik, J., Bose, A., Pomerleau, D.P., Furcinitti, P., Leszyk, J., Corvera, S. and Czech, M.P. (2004) EHD2 and the novel EH domain binding protein EHBP1 couple endocytosis to the actin cytoskeleton. *J Biol Chem*, **279**, 10593-10605.

Guilinger, J.P., Pattanayak, V., Reyon, D., Tsai, S.Q., Sander, J.D., Joung, J.K. and Liu, D.R. (2014a) Broad Specificity Profiling of TALENs Results in Engineered Nucleases With Improved DNA Cleavage Specificity. *Nat Methods*, **11**, 429-435.

Guilinger, J.P., Thompson, D.B. and Liu, D.R. (2014b) Fusion of catalytically inactive Cas9 to FokI nuclease improves the specificity of genome modification. *Nat Biotechnol*, **32**, 577-582.

Hamm, C.E., Merkel, R., Springer, O., Jurkojc, P., Maier, C., Prechtel, K. and Smetacek, V. (2003) Architecture and material properties of diatom shells provide effective mechanical protection. *Nature*, **421**, 841-843.

Hanahan, D. (1983) Studies on transformation of Escherichia coli with plasmids. *J Mol Biol*, **166**, 557-580.

Heintz, U. and Schlichting, I. (2016) Blue light-induced LOV domain dimerization enhances the affinity of Aureochrome 1a for its target DNA sequence. *Elife*, **5**, e11860.

Herman, E. and Kottke, T. (2015) Allosterically regulated unfolding of the A'alpha helix exposes the dimerization site of the blue-light-sensing aureochrome-LOV domain. *Biochemistry*, **54**, 1484-1492.

Herman, E., Sachse, M., Kroth, P.G. and Kottke, T. (2013) Blue-Light-Induced Unfolding of the Jalpha Helix Allows for the Dimerization of Aureochrome-LOV from the Diatom Phaeodactylum tricornutum. *Biochemistry*.

Hsu, P.D., Scott, D.A., Weinstein, J.A., Ran, F.A., Konermann, S., Agarwala, V., Li, Y., Fine, E.J., Wu, X., Shalem, O., Cradick, T.J., Marraffini, L.A., Bao, G. and Zhang, F. (2013) DNA targeting specificity of RNA-guided Cas9 nucleases. *Nat Biotechnol*, **31**, 827-832.

Huysman, M.J., Fortunato, A.E., Matthijs, M., Costa, B.S., Vanderhaeghen, R., Van den Daele, H., Sachse, M., Inze, D., Bowler, C., Kroth, P.G., Wilhelm, C., Falciatore, A., Vyverman, W. and De Veylder, L. (2013) AUREOCHROME1a-mediated induction of the diatom-specific cyclin dsCYC2 controls the onset of cell division in diatoms (Phaeodactylum tricornutum). *Plant Cell*, **25**, 215-228.

Huysman, M.J., Martens, C., Vandepoele, K., Gillard, J., Rayko, E., Heijde, M., Bowler, C., Inze, D., Van de Peer, Y., De Veylder, L. and Vyverman, W. (2010) Genome-wide analysis of the diatom cell cycle unveils a novel type of cyclins involved in environmental signaling. *Genome Biol*, **11**, R17.

Huysman, M.J., Martens, C., Vyverman, W. and De Veylder, L. (2014) Protein degradation during the diatom cell cycle: annotation and transcriptional analysis of SCF and APC/C ubiquitin ligase genes in Phaeodactylum tricornutum. *Mar Genomics*, **14**, 39-46.

Hwang, W.Y., Fu, Y., Reyon, D., Maeder, M.L., Tsai, S.Q., Sander, J.D., Peterson, R.T., Yeh, J.R.J. and Joung, J.K. (2013) Efficient genome editing in zebrafish using a CRISPR-Cas system. *Nat Biotech*, **31**, 227-229.

Ishikawa, M., Takahashi, F., Nozaki, H., Nagasato, C., Motomura, T. and Kataoka, H. (2009) Distribution and phylogeny of the blue light receptors aureochromes in eukaryotes. *Planta*, **230**, 543-552.

Jackson, A.L., Bartz, S.R., Schelter, J., Kobayashi, S.V., Burchard, J., Mao, M., Li, B., Cavet, G. and Linsley, P.S. (2003) Expression profiling reveals off-target gene regulation by RNAi. *Nat Biotechnol*, **21**, 635-637.

Jacobs, M.J. (2015) Molecular characterization of Aureochromes in the diatom *Phaeodactylum tricornutum*. Universität Konstanz (Master's Thesis).

Jakob, T., Goss, R. and Wilhelm, C. (1999) Activation of diadinoxanthin de-epoxidase due to a chlororespiratory proton gradient in the dark in the diatom Phaeodactylum tricornutum. *Plant Biol*, **1**, 76-82.

Jeffrey, S.W. and Humphrey, G.F. (1975) New Spectrophotometric Equations for Determining Chlorophylls a, B, C1 and C2 in Higher-Plants, Algae and Natural Phytoplankton. *Biochem Physiol Pflanz*, **167**, 191-194.

Karas, B.J., Diner, R.E., Lefebvre, S.C., McQuaid, J., Phillips, A.P., Noddings, C.M., Brunson, J.K., Valas, R.E., Deerinck, T.J., Jablanovic, J., Gillard, J.T., Beeri, K., Ellisman, M.H., Glass, J.I., Hutchison, C.A., 3rd, Smith, H.O., Venter, J.C., Allen, A.E., Dupont, C.L. and Weyman, P.D. (2015) Designer diatom episomes delivered by bacterial conjugation. *Nat Commun*, **6**, 6925.

Keeling, P.J. (2013) The Number, Speed, and Impact of Plastid Endosymbioses in Eukaryotic Evolution. *Annu Rev Plant Biol*, **64**, 583-607.

Khan, A. and Mathelier, A. (2017) Intervene: a tool for intersection and visualization of multiple gene or genomic region sets. *BMC Bioinformatics*, **18**, 287.

Kim, D., Kim, J., Hur, J.K., Been, K.W., Yoon, S.H. and Kim, J.S. (2016) Genome-wide analysis reveals specificities of Cpf1 endonucleases in human cells. *Nat Biotechnol*, **34**, 863-868.

Kleinstiver, B.P., Pattanayak, V., Prew, M.S., Tsai, S.Q., Nguyen, N.T., Zheng, Z. and Joung, J.K. (2016a) High-fidelity CRISPR-Cas9 nucleases with no detectable genome-wide off-target effects. *Nature*, **529**, 490-495.

Kleinstiver, B.P., Tsai, S.Q., Prew, M.S., Nguyen, N.T., Welch, M.M., Lopez, J.M., McCaw, Z.R., Aryee, M.J. and Joung, J.K. (2016b) Genome-wide specificities of CRISPR-Cas Cpf1 nucleases in human cells. *Nat Biotechnol*, **34**, 869-874.

Kok, F.O., Shin, M., Ni, C.W., Gupta, A., Grosse, A.S., van Impel, A., Kirchmaier, B.C., Peterson-Maduro, J., Kourkoulis, G., Male, I., DeSantis, D.F., Sheppard-Tindell, S., Ebarasi, L., Betsholtz, C., Schulte-Merker, S., Wolfe, S.A. and Lawson, N.D. (2015) Reverse genetic screening reveals poor correlation between morpholino-induced and mutant phenotypes in zebrafish. *Dev Cell*, **32**, 97-108.

Kolodziej, K.E., Pourfarzad, F., de Boer, E., Krpic, S., Grosveld, F. and Strouboulis, J. (2009) Optimal use of tandem biotin and V5 tags in ChIP assays. *BMC Mol Biol*, **10**, 6.

Kröger, N. and Brunner, E. (2014) Complex-shaped microbial biominerals for nanotechnology. *Wiley Interdiscip Rev Nanomed Nanobiotechnol*, **6**, 615-627.

Kroth, P. and Strotmann, H. (1999) Diatom plastids: Secondary endocytobiosis, plastid genome and protein import. *Physiol Plant*, **107**, 136-141.

Kroth, P.G. (2007) Genetic transformation: a tool to study protein targeting in diatoms. *Methods Mol Biol*, **390**, 257-267.

Krysiak, C., Mazus, B. and Buchowicz, J. (1999) Generation of DNA double-strand breaks and inhibition of somatic embryogenesis by tungsten microparticles in wheat. *Plant Cell Tissue Organ Cult*, **58**, 163-170.

Kuscu, C., Arslan, S., Singh, R., Thorpe, J. and Adli, M. (2014) Genome-wide analysis reveals characteristics of off-target sites bound by the Cas9 endonuclease. *Nat Biotechnol*, **32**, 677-683.

Laemmli, U.K. (1970) Cleavage of structural proteins during the assembly of the head of bacteriophage T4. *Nature*, **227**, 680-685.

Landt, S.G., Marinov, G.K., Kundaje, A., Kheradpour, P., Pauli, F., Batzoglou, S., Bernstein, B.E., Bickel, P., Brown, J.B., Cayting, P., Chen, Y., DeSalvo, G., Epstein, C., Fisher-Aylor, K.I., Euskirchen, G., Gerstein, M., Gertz, J., Hartemink, A.J., Hoffman, M.M., Iyer, V.R., Jung, Y.L., Karmakar, S., Kellis, M., Kharchenko, P.V., Li, Q., Liu, T., Liu, X.S., Ma, L., Milosavljevic, A., Myers, R.M., Park, P.J., Pazin, M.J., Perry, M.D., Raha, D., Reddy, T.E., Rozowsky, J., Shoresh, N., Sidow, A., Slattery, M., Stamatoyannopoulos, J.A., Tolstorukov, M.Y., White, K.P., Xi, S., Farnham, P.J., Lieb, J.D., Wold, B.J. and Snyder, M. (2012) ChIP-seq guidelines and practices of the ENCODE and modENCODE consortia. *Genome Res*, **22**, 1813-1831.

Latchman, D.S. (2001) Transcription factors: bound to activate or repress. *Trends Biochem Sci*, **26**, 211-213.

Lavaud, J. (2007) Fast Regulation of Photosynthesis in Diatoms: Mechanisms, Evolution and Ecophysiology. *Funct Plant Sci Biotechnol*, **1**, 267-287.

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

Lavaud, J., Materna, A.C., Sturm, S., Vugrinec, S. and Kroth, P.G. (2012) Silencing of the violaxanthin de-epoxidase gene in the diatom Phaeodactylum tricornutum reduces diatoxanthin synthesis and non-photochemical quenching. *PLoS One*, **7**, e36806.

Leinweber, K. and Kroth, P.G. (2015) Capsules of the diatom Achnanthidium minutissimum arise from fibrillar precursors and foster attachment of bacteria. *PeerJ*, **3**, e858.

Lepetit, B. and Dietzel, L. (2015) Light signaling in photosynthetic eukaryotes with 'green' and 'red' chloroplasts. *Environ Exp Bot*, **114**, 30-47.

Lepetit, B., Gelin, G., Lepetit, M., Sturm, S., Vugrinec, S., Rogato, A., Kroth, P.G., Falciatore, A. and Lavaud, J. (2017) The diatom Phaeodactylum tricornutum adjusts nonphotochemical fluorescence quenching capacity in response to dynamic light via fine-tuned Lhcx and xanthophyll cycle pigment synthesis. *New Phytol*, **214**, 205-218.

Lepetit, B., Sturm, S., Rogato, A., Gruber, A., Sachse, M., Falciatore, A., Kroth, P.G. and Lavaud, J. (2013) High light acclimation in the secondary plastids containing diatom Phaeodactylum tricornutum is triggered by the redox state of the plastoquinone pool. *Plant Physiol*, **161**, 853-865.

Levine, M. and Tjian, R. (2003) Transcription regulation and animal diversity. *Nature*, **424**, 147-151.

Levitan, O., Dinamarca, J., Zelzion, E., Lun, D.S., Guerra, L.T., Kim, M.K., Kim, J., Van Mooy, B.A., Bhattacharya, D. and Falkowski, P.G. (2015) Remodeling of intermediate metabolism in the diatom Phaeodactylum tricornutum under nitrogen stress. *Proc Natl Acad Sci U S A*, **112**, 412-417.

Li, J., Li, G., Wang, H. and Wang Deng, X. (2011a) Phytochrome signaling mechanisms. *Arabidopsis Book*, **9**, e0148.

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

Lieber, M.R. (2010) The mechanism of double-strand DNA break repair by the nonhomologous DNA end-joining pathway. *Annu Rev Biochem*, **79**, 181-211.

Lin, C. and Todo, T. (2005) The cryptochromes. *Genome Biol*, **6**, 220.

Lin, Y., Fine, E.J., Zheng, Z., Antico, C.J., Voit, R.A., Porteus, M.H., Cradick, T.J. and Bao, G. (2014) SAPTA: a new design tool for improving TALE nuclease activity. *Nucleic Acids Res*, **42**, e47.

Losi, A. and Gartner, W. (2017) Solving Blue Light Riddles: New Lessons from Flavin-binding LOV Photoreceptors. *Photochem Photobiol*, **93**, 141-158.

Ma, J. (2005) Crossing the line between activation and repression. *Trends Genet*, **21**, 54-59.

MacIntyre, H.L., Kana, T.M. and Geider, R.J. (2000) The effect of water motion on short-term rates of photosynthesis by marine phytoplankton. *Trends Plant Sci*, **5**, 12-17.

Madan Babu, M. and Teichmann, S.A. (2003) Functional determinants of transcription factors in Escherichia coli: protein families and binding sites. *Trends Genet*, **19**, 75-79.

Magadan, J.G., Barbieri, M.A., Mesa, R., Stahl, P.D. and Mayorga, L.S. (2006) Rab22a regulates the sorting of transferrin to recycling endosomes. *Mol Cell Biol*, **26**, 2595-2614.

Mann, D.G. and Droop, S.J.M. (1996) Biodiversity, biogeography and conservation of diatoms. *Hydrobiologia*, **336**, 19-32.

Mao, Z., Bozzella, M., Seluanov, A. and Gorbunova, V. (2008) Comparison of nonhomologous end joining and homologous recombination in human cells. *DNA Repair (Amst)*, **7**, 1765-1771.

Mata, T.M., Martins, A.A. and Caetano, N.S. (2010) Microalgae for biodiesel production and other applications: A review. *Renew Sustainable Energy Rev*, **14**, 217-232.

Mathelier, A., Fornes, O., Arenillas, D.J., Chen, C.Y., Denay, G., Lee, J., Shi, W., Shyr, C., Tan, G., Worsley-Hunt, R., Zhang, A.W., Parcy, F., Lenhard, B., Sandelin, A. and Wasserman, W.W. (2016) JASPAR 2016: a major expansion and update of the open-access database of transcription factor binding profiles. *Nucleic Acids Res*, **44**, D110-115.

Matthijs, M., Fabris, M., Obata, T., Foubert, I., Franco-Zorrilla, J.M., Solano, R., Fernie, A.R., Vyverman, W. and Goossens, A. (2017) The transcription factor bZIP14 regulates the TCA cycle in the diatom Phaeodactylum tricornutum. *EMBO J*, **36**, 1559-1576.

Meinke, D.W. (2013) A survey of dominant mutations in Arabidopsis thaliana. *Trends Plant Sci*, **18**, 84-91.

Miller, J.C., Tan, S., Qiao, G., Barlow, K.A., Wang, J., Xia, D.F., Meng, X., Paschon, D.E., Leung, E., Hinkley, S.J., Dulay, G.P., Hua, K.L., Ankoudinova, I., Cost, G.J., Urnov, F.D., Zhang, H.S., Holmes, M.C., Zhang, L., Gregory, P.D. and Rebar, E.J. (2011) A TALE nuclease architecture for efficient genome editing. *Nat Biotechnol*, **29**, 143-148.

Minkenberg, B., Wheatley, M. and Yang, Y. (2017) CRISPR/Cas9-Enabled Multiplex Genome Editing and Its Application. *Prog Mol Biol Transl Sci*, **149**, 111-132.

Mitra, D., Yang, X. and Moffat, K. (2012) Crystal structures of Aureochrome1 LOV suggest new design strategies for optogenetics. *Structure*, **20**, 698-706.

Miyahara, M., Aoi, M., Inoue-Kashino, N., Kashino, Y. and Ifuku, K. (2013) Highly efficient transformation of the diatom Phaeodactylum tricornutum by multi-pulse electroporation. *Biosci Biotechnol Biochem*, **77**, 874-876.

Moreno-Campuzano, S., Janga, S.C. and Perez-Rueda, E. (2006) Identification and analysis of DNA-binding transcription factors in Bacillus subtilis and other Firmicutes--a genomic approach. *BMC Genomics*, **7**, 147.

Morino, C., Kato, M., Yamamoto, A., Mizuno, E., Hayakawa, A., Komada, M. and Kitamura, N. (2004) A role for Hrs in endosomal sorting of ligand-stimulated and unstimulated epidermal growth factor receptor. *Exp Cell Res*, **297**, 380-391.

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

Müller, P., Li, X.P. and Niyogi, K.K. (2001) Non-photochemical quenching. A response to excess light energy. *Plant Physiol*, **125**, 1558-1566.

Nelson, D.M., Treguer, P., Brzezinski, M.A., Leynaert, A. and Queguiner, B. (1995) Production and Dissolution of Biogenic Silica in the Ocean - Revised Global Estimates, Comparison with Regional Data and Relationship to Biogenic Sedimentation. *Glob Biogeochem Cycles*, **9**, 359-372.

Niu, Y.F., Yang, Z.K., Zhang, M.H., Zhu, C.C., Yang, W.D., Liu, J.S. and Li, H.Y. (2012) Transformation of diatom Phaeodactylum tricornutum by electroporation and establishment of inducible selection marker. *BioTechniques*, **52**, 1-3.

Nordberg, H., Cantor, M., Dusheyko, S., Hua, S., Poliakov, A., Shabalov, I., Smirnova, T., Grigoriev, I.V. and Dubchak, I. (2014) The genome portal of the Department of Energy Joint Genome Institute: 2014 updates. *Nucleic Acids Res*, **42**, D26-31.

Nymark, M., Sharma, A.K., Sparstad, T., Bones, A.M. and Winge, P. (2016) A CRISPR/Cas9 system adapted for gene editing in marine algae. *Sci Rep*, **6**, 24951.

Nymark, M., Valle, K.C., Brembu, T., Hancke, K., Winge, P., Andresen, K., Johnsen, G. and Bones, A.M. (2009) An integrated analysis of molecular acclimation to high light in the marine diatom Phaeodactylum tricornutum. *PLoS One*, **4**, e7743.

Nymark, M., Valle, K.C., Hancke, K., Winge, P., Andresen, K., Johnsen, G., Bones, A.M. and Brembu, T. (2013) Molecular and photosynthetic responses to prolonged darkness and subsequent acclimation to re-illumination in the diatom Phaeodactylum tricornutum. *PLoS One*, **8**, e58722.

O'Shea, E.K., Klemm, J.D., Kim, P.S. and Alber, T. (1991) X-ray structure of the GCN4 leucine zipper, a two-stranded, parallel coiled coil. *Science*, **254**, 539-544.

O'Shea, E.K., Rutkowski, R., Stafford, W.F., 3rd and Kim, P.S. (1989) Preferential heterodimer formation by isolated leucine zippers from fos and jun. *Science*, **245**, 646-648.

Ohno, N., Inoue, T., Yamashiki, R., Nakajima, K., Kitahara, Y., Ishibashi, M. and Matsuda, Y. (2012) CO(2)-cAMP-responsive cis-elements targeted by a transcription factor with CREB/ATF-like basic zipper domain in the marine diatom Phaeodactylum tricornutum. *Plant Physiol*, **158**, 499-513.

Ohse, M., Takahashi, K., Kadowaki, Y. and Kusaoke, H. (1995) Effects of plasmid DNA sizes and several other factors on transformation of Bacillus subtilis ISW1214 with plasmid DNA by electroporation. *Biosci Biotechnol Biochem*, **59**, 1433-1437.

Park, B., Ying, H., Shen, X., Park, J.S., Qiu, Y., Shyam, R. and Yue, B.Y. (2010) Impairment of protein trafficking upon overexpression and mutation of optineurin. *PLoS One*, **5**, e11547.

Park, S.Y., Cable, A.E., Blair, J., Stockstill, K.E. and Shannnon, K.B. (2009) Bub2 regulation of cytokinesis and septation in budding yeast. *BMC Cell Biol*, **10**, 43.

Pattanayak, V., Lin, S., Guilinger, J.P., Ma, E., Doudna, J.A. and Liu, D.R. (2013) High-throughput profiling of off-target DNA cleavage reveals RNA-programmed Cas9 nuclease specificity. *Nat Biotechnol*, **31**, 839-843.

Perez-Rueda, E. and Collado-Vides, J. (2000) The repertoire of DNA-binding transcriptional regulators in Escherichia coli K-12. *Nucleic Acids Res*, **28**, 1838-1847.

Petroutsos, D., Tokutsu, R., Maruyama, S., Flori, S., Greiner, A., Magneschi, L., Cusant, L., Kottke, T., Mittag, M., Hegemann, P., Finazzi, G. and Minagawa, J. (2016) A blue-light photoreceptor mediates the feedback regulation of photosynthesis. *Nature*, **537**, 563-566.

Pfaffl, M.W., Horgan, G.W. and Dempfle, L. (2002) Relative expression software tool (REST) for group-wise comparison and statistical analysis of relative expression results in real-time PCR. *Nucleic Acids Res*, **30**, e36.

Pfannschmidt, T. (2003) Chloroplast redox signals: how photosynthesis controls its own genes. *Trends Plant Sci*, **8**, 33-41.

R Development Core Team (2017) R: A language and environment for statistical computing. Vienna, Austria: R Foundation for Statistical Computing. Retrived from: www.R-project.org

Ragni, M. and D'Alcalà, M.R. (2004) Light as an information carrier underwater. *J Plankton Res*, **26**, 433-443.

Rayko, E., Maumus, F., Maheswari, U., Jabbari, K. and Bowler, C. (2010) Transcription factor families inferred from genome sequences of photosynthetic stramenopiles. *New Phytol*, **188**, 52-66.

Riechmann, J.L., Heard, J., Martin, G., Reuber, L., Jiang, C., Keddie, J., Adam, L., Pineda, O., Ratcliffe, O.J., Samaha, R.R., Creelman, R., Pilgrim, M., Broun, P., Zhang, J.Z., Ghandehari, D., Sherman, B.K. and Yu, G. (2000) Arabidopsis transcription factors: genome-wide comparative analysis among eukaryotes. *Science*, **290**, 2105-2110.

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

Roesle, P., Stempfle, F., Hess, S.K., Zimmerer, J., Bartulos, C.R., Lepetit, B., Eckert, A., Kroth, P.G. and Mecking, S. (2014) Synthetic Polyester from Algae Oil. *Angew Chem Int Ed*, **53**, 6800-6804.

Rossi, A., Kontarakis, Z., Gerri, C., Nolte, H., Holper, S., Kruger, M. and Stainier, D.Y. (2015) Genetic compensation induced by deleterious mutations but not gene knockdowns. *Nature*, **524**, 230-233.

Rottberger, J., Gruber, A. and Kroth, P.G. (2013) Analysing size variation during light-starvation response of nutritionally diverse chrysophytes with a Coulter counter. *Algol Stud*, **141**, 37-51.

Round, F.E., Crawford, R.M. and Mann, D.G. (1990) *Diatoms: Biology and Morphology of the Genera*: Cambridge University Press.

Ruban, A., Lavaud, J., Rousseau, B., Guglielmi, G., Horton, P. and Etienne, A.L. (2004) The super-excess energy dissipation in diatom algae: comparative analysis with higher plants. *Photosynth Res*, **82**, 165-175.

Sachse, M., Sturm, S., Gruber, A. and Kroth, P.G. (2014) Identification and evaluation of endogenous reference genes for steady state transcript quantification by qPCR in the diatom Phaeodactylum tricornutum with constitutive expression independent from time and light. *Endocytobiosis Cell Res*, **24**, 1-7.

Sanjana, N.E., Cong, L., Zhou, Y., Cunniff, M.M., Feng, G. and Zhang, F. (2012) A transcription activator-like effector toolbox for genome engineering. *Nat Protocols*, **7**, 171-192.

Schellenberger Costa, B., Jungandreas, A., Jakob, T., Weisheit, W., Mittag, M. and Wilhelm, C. (2013a) Blue light is essential for high light acclimation and photoprotection in the diatom Phaeodactylum tricornutum. *J Exp Bot*, **64**, 483-493.

Schellenberger Costa, B., Sachse, M., Jungandreas, A., Bartulos, C.R., Gruber, A., Jakob, T., Kroth, P.G. and Wilhelm, C. (2013b) Aureochrome 1a Is Involved in the Photoacclimation of the Diatom. *PLoS One*, **8**.

Schornack, S., Meyer, A., Romer, P., Jordan, T. and Lahaye, T. (2006) Gene-for-gene-mediated recognition of nuclear-targeted AvrBs3-like bacterial effector proteins. *J Plant Physiol*, **163**, 256-272.

Serif, M., Lepetit, B., Weißert, K., Kroth, P.G. and Rio Bartulos, C. (2017) A fast and reliable strategy to generate TALEN-mediated gene knockouts in the diatom Phaeodactylum tricornutum. *Algal Res*, **23**, 186-195.

Shalitin, D., Yang, H., Mockler, T.C., Maymon, M., Guo, H., Whitelam, G.C. and Lin, C. (2002) Regulation of Arabidopsis cryptochrome 2 by blue-light-dependent phosphorylation. *Nature*, **417**, 763-767.

Shalitin, D., Yu, X., Maymon, M., Mockler, T. and Lin, C. (2003) Blue light-dependent in vivo and in vitro phosphorylation of Arabidopsis cryptochrome 1. *Plant Cell*, **15**, 2421-2429.

Stella, S. and Montoya, G. (2016) The genome editing revolution: A CRISPR-Cas TALE off-target story. *Inside Cell*, **1**, 7-16.

Stewart, A.J., Hannenhalli, S. and Plotkin, J.B. (2012) Why transcription factor binding sites are ten nucleotides long. *Genetics*, **192**, 973-985.

Stork, S., Moog, D., Przyborski, J.M., Wilhelmi, I., Zauner, S. and Maier, U.G. (2012) Distribution of the SELMA translocon in secondary plastids of red algal origin and predicted uncoupling of ubiquitin-dependent translocation from degradation. *Eukaryot Cell*, **11**, 1472-1481.

Strick, D.J. and Elferink, L.A. (2005) Rab15 effector protein: a novel protein for receptor recycling from the endocytic recycling compartment. *Mol Biol Cell*, **16**, 5699-5709.

Sullivan, J.A. and Deng, X.W. (2003) From seed to seed: the role of photoreceptors in Arabidopsis development. *Dev Biol*, **260**, 289-297.

Taddei, L., Stella, G.R., Rogato, A., Bailleul, B., Fortunato, A.E., Annunziata, R., Sanges, R., Thaler, M., Lepetit, B., Lavaud, J., Jaubert, M., Finazzi, G., Bouly, J.P. and Falciatore, A. (2016) Multisignal control of expression of the LHCX protein family in the marine diatom Phaeodactylum tricornutum. *J Exp Bot*, **67**, 3939-3951.

Takahashi, F., Yamagata, D., Ishikawa, M., Fukamatsu, Y., Ogura, Y., Kasahara, M., Kiyosue, T., Kikuyama, M., Wada, M. and Kataoka, H. (2007) AUREOCHROME, a photoreceptor required for photomorphogenesis in stramenopiles. *Proc Natl Acad Sci U S A*, **104**, 19625-19630.

Trentacoste, E.M., Shrestha, R.P., Smith, S.R., Gle, C., Hartmann, A.C., Hildebrand, M. and Gerwick, W.H. (2013) Metabolic engineering of lipid catabolism increases microalgal lipid accumulation without compromising growth. *Proc Natl Acad Sci U S A*, **110**, 19748-19753.

Triantaphylidès, C. and Havaux, M. (2009) Singlet oxygen in plants: production, detoxification and signaling. *Trends Plant Sci*, **14**, 219-228.

Ulijasz, A.T. and Vierstra, R.D. (2011) Phytochrome structure and photochemistry: recent advances toward a complete molecular picture. *Curr Opin Plant Biol*, **14**, 498-506.

Valle, K.C., Nymark, M., Aamot, I., Hancke, K., Winge, P., Andresen, K., Johnsen, G., Brembu, T. and Bones, A.M. (2014) System responses to equal doses of photosynthetically usable radiation of blue, green, and red light in the marine diatom Phaeodactylum tricornutum. *PLoS One*, **9**, e114211.

Vinson, C.R., Hai, T. and Boyd, S.M. (1993) Dimerization specificity of the leucine zipper-containing bZIP motif on DNA binding: prediction and rational design. *Genes Dev*, **7**, 1047-1058.

Vinson, C.R., Sigler, P.B. and McKnight, S.L. (1989) Scissors-grip model for DNA recognition by a family of leucine zipper proteins. *Science*, **246**, 911-916.

Wagner, G.P., Kin, K. and Lynch, V.J. (2012) Measurement of mRNA abundance using RNA-seq data: RPKM measure is inconsistent among samples. *Theory Biosci*, **131**, 281-285.

Wang, P., Zhou, Z., Hu, A., Ponte de Albuquerque, C., Zhou, Y., Hong, L., Sierecki, E., Ajiro, M., Kruhlak, M., Harris, C., Guan, K.L., Zheng, Z.M., Newton, A.C., Sun, P., Zhou, H. and Fu, X.D. (2014) Both decreased and increased SRPK1 levels promote cancer by interfering with PHLPP-mediated dephosphorylation of Akt. *Mol Cell*, **54**, 378-391.

Wang, X., Wang, Y., Wu, X., Wang, J., Wang, Y., Qiu, Z., Chang, T., Huang, H., Lin, R.J. and Yee, J.K. (2015) Unbiased detection of off-target cleavage by CRISPR-Cas9 and TALENs using integrase-defective lentiviral vectors. *Nat Biotechnol*, **33**, 175-178.

Weyman, P.D., Beeri, K., Lefebvre, S.C., Rivera, J., McCarthy, J.K., Heuberger, A.L., Peers, G., Allen, A.E. and Dupont, C.L. (2015) Inactivation of Phaeodactylum tricornutum urease gene using transcription activator-like effector nuclease-based targeted mutagenesis. *Plant Biotechnol J*, **13**, 460-470.

Wilhelm, C. and Selmar, D. (2011) Energy dissipation is an essential mechanism to sustain the viability of plants: The physiological limits of improved photosynthesis. *J Plant Physiol*, **168**, 79-87.

Windler, M., Leinweber, K., Bartulos, C.R., Philipp, B. and Kroth, P.G. (2015) Biofilm and capsule formation of the diatom Achnanthidium minutissimum are affected by a bacterium. *J Phycol*, **51**, 343-355.

Wood, A.J., Lo, T.-W., Zeitler, B., Pickle, C.S., Ralston, E.J., Lee, A.H., Amora, R., Miller, J.C., Leung, E., Meng, X., Zhang, L., Rebar, E.J., Gregory, P.D., Urnov, F.D. and Meyer, B.J. (2011) Targeted Genome Editing Across Species Using ZFNs and TALENs. *Science*, **333**, 307.

Yeh, K.C. and Lagarias, J.C. (1998) Eukaryotic phytochromes: light-regulated serine/threonine protein kinases with histidine kinase ancestry. *Proc Natl Acad Sci U S A*, **95**, 13976-13981.

Yool, A. and Tyrrell, T. (2003) Role of diatoms in regulating the ocean's silicon cycle. *Glob Biogeochem Cycles*, **17**, 1103–1121.

Yu, X., Shalitin, D., Liu, X., Maymon, M., Klejnot, J., Yang, H., Lopez, J., Zhao, X., Bendehakkalu, K.T. and Lin, C. (2007) Derepression of the NC80 motif is critical for the photoactivation of Arabidopsis CRY2. *Proc Natl Acad Sci U S A*, **104**, 7289-7294.

Zaslavskaia, L.A., Lippmeier, J.C., Kroth, P.G., Grossman, A.R. and Apt, K.E. (2000) Transformation of the diatom Phaeodactylum tricornutum (Bacillariophyceae) with a variety of selectable marker and reporter genes. *J Phycol*, **36**, 379-386.

Zhang, C.Y. and Hu, H.H. (2014) High-efficiency nuclear transformation of the diatom Phaeodactylum tricornutum by electroporation. *Mar Genomics*, **16**, 63-66.

Zhu, S.H. and Green, B.R. (2010) Photoprotection in the diatom Thalassiosira pseudonana: role of LI818-like proteins in response to high light stress. *Biochim Biophys Acta*, **1797**, 1449-1457.

Zu, Y., Tong, X., Wang, Z., Liu, D., Pan, R., Li, Z., Hu, Y., Luo, Z., Huang, P., Wu, Q., Zhu, Z., Zhang, B. and Lin, S. (2013) TALEN-mediated precise genome modification by homologous recombination in zebrafish. *Nat Methods*, **10**, 329-331.

Zykovich, A., Korf, I. and Segal, D.J. (2009) Bind-n-Seq: high-throughput analysis of in vitro protein-DNA interactions using massively parallel sequencing. *Nucleic Acids Res*, **37**, e151.