Molecular Characterisation of Nucleotide Transporters in Diatoms

Dissertation

zur Erlangung des akademischen Grades eines Doktors der Naturwissenschaften (Dr. rer. nat)

an der Universität Konstanz, Fachbereich Biologie

vorgelegt von

Lili Chu

Tag der mündlichen Prüfung: 16.12.16

1. Prüfer: Prof Dr. Peter G. Kroth
2. Prüfer: Prof Dr. Wolfram Welte
3. Prüfer: Dr. Ilka Haferkamp (extern, TU Kaiserslautern)

Konstanz 2016

Summary

Diatoms harbour secondary plastids, surrounded by four membranes (instead of two like in higher plants), of which the outermost membrane is continuous with the ER and the nuclear envelope. The evolution of diatom plastids by secondary endocytobiosis was accompanied by transfers of genes from the endosymbiont to the host genome, which resulted in a rearrangement of the targeting pathways of plastid proteins, as well as of whole metabolic pathways, e.g. nucleotide biosynthesis.

To identify and characterise nucleotide translocators (NTTs) in the diatom *P. tricornutum*, the substrate spectrum, transport-mode and GFP-based localisation of *Pt*NTT5 were determined. This transporter accepts various purine nucleotides and integrates most likely into the outermost membrane of the plastid, which demonstrates that this transporter could play a major role in controlling the nucleotide pool within the ER lumen and that the ER membrane represents a selective barrier for nucleotides. Additionally, certain domains of the proteins were deleted to examine the influence on targeting and insertion processes. The truncations of *Pt*NTT5 revealed that this transporter protein might harbour targeting and tail-anchor domains and might be targeted and inserted into the membrane post-translationally.

The biochemical characterisation of NTT3 of the centric diatom *Thalassiosira pseudonana* (*Tp*NTT3), revealed an unexpected broad substrate spectrum of purine nucleotides, including cyclic mononucleotides, which has not been shown for any described NTT so far. The subcellular localisation of *Tp*NTT3 could not be determined with certainty, but might be ER associated. Accordingly, *Tp*NTT3 could be involved in signalling mechanisms connected to the secretory pathway, since cyclic mononucleotides are important second messengers.

In order to define the composition of glycans which might be attached to plastidial or secreted diatom proteins, metabolic glycoengineering (MGE) was applied to the model diatom *Phaeodactylum tricornutum*. Apparently, the chemical reporter group DIBO alkyne non-specifically labels unidentified structures in the centre of the cells. Furthermore, it was shown that diatoms do not utilise derivatives from mannosamine and therefore, other carbohydrates might be more suitable substrates.

For the investigation of sub-compartmental protein locations, proteins of interest are often fused to fragments of GFP, which are self-assembling to fluorescing units. We observed GFP fluorescence in transformed cell lines that, in theory, should not contain self-assembling GFP fragments in the same cellular compartment. This indicates that targeting of the utilised fusion proteins is strongly affected by the self-assembling GFP fragments, and that the application of this method requires extensive control experiments.

With the goal of controlling or fine tuning trans-gene expression of reporter proteins in *P. tricornutum*, the inducible nitrate reductase (nr) promoter was characterised by flow cytometry in a time resolved way. GFP expression was induced by the given media and increased already within the first hour, whereas a decrease after medium exchange was much slower. The nr promoter showed slight activity despite non-inducing conditions and furthermore, strong variations among independent cell lines and their expression patterns could be observed.

The studies on NTTs in diatoms demonstrate that transporters mediating the shuttle of purine nucleotides are present in the ER membranes. Furthermore, diatom NTTs have additional functions than solely the provision of energy to the plastid (unlike in plants). Instead, they are also involved in other physiological functions, such as nucleotide import into the plastids and possibly also signalling pathways.

Zusammenfassung

Kieselalgen besitzen sekundäre Plastiden, die von vier Membranen umgeben sind (anstelle von zwei wie in höheren Pflanzen), wobei die äußerste Membran kontinuierlich mit dem ER und der Kernhülle ist. Die Entwicklung von Kieselalgen-Plastiden durch sekundäre Endocytobiose ging einher mit dem Transfer von Genen aus dem Endosymbionten- in das Wirtsgenom, was eine Neuordnung der Transportwege plastidärer Proteine zur Folge hatte, sowie von ganzen Stoffwechselwegen, wie z. B. der Nukleotid-Biosynthese.

Zur Identifizierung und Charakterisierung von Nukleotidtranslokatoren (NTTs) in der Kieselalge *P. tricornutum* wurden das Substratspektrum, der Transport-Modus und die auf GFP-basierende Lokalisierungen eines Nukleotidtransporters von *P. tricornutum* (*Pt*NTT5) bestimmt. Dieser Transporter akzeptiert verschiedene Purinnukleotide und integriert wahrscheinlich in die äußerste Membrane der Plastiden, was zeigt, dass der Transporter eine wichtige Rolle in der Kontrolle des Nukleotidpools des ERs übernehmen kann und dass die ER Membran eine selektive Barriere für Nukleotide darstellt. Zudem wurde das Protein um bestimmte Domänen verkürzt um deren Einfluss dieser Domänen auf den Transport und die Membran Insertion zu untersuchen. Die Verkürzungen von *Pt*NTT5 zeigten, dass der Transporter möglicherweise Targeting- und Tail-Anchor-Domänen besitzt und dass er post-translational zu seiner Ziel Membran transportiert und inseriert wird.

Die biochemische Charakterisierung eines weiteren Nukleotidtransporters aus der zentrischen Kieselalge *Thalassiosira pseudonana* (*Tp*NTT3) zeigte ein unerwartet breites Substratspektrum, bestehend aus Purinnukleotiden, einschließlich zyklischer Mononukleotide, was bisher noch für keinen NTT gezeigt wurde. Die subzelluläre Lokalisation von *Tp*NTT3 konnte nicht eindeutig bestimmt werden, aber ist möglicherweise ER-assoziiert. Demnach könnte *Tp*NTT3 in Verbindung mit dem sekretorischen Weg an Signalübertragungsmechanismen beteiligt sein, da zyklische Mononukleotide wichtige Botenstoffe sind.

Um die Zusammensetzung von Polysacchariden, die an plastidären oder sekretierten Proteinen von Kieselalgen angeheftet sein könnten, zu bestimmen, wurde Metabolic Glycoengineering (MGE) in der Modell-Kieselalge *Phaeodactylum tricornutum* angewendet. Die chemische Reportergruppe DIBO-Alkin markiert dabei unbekannte Strukturen im Zentrum der Zellen. Des Weiteren wurde gezeigt, dass Kieselalgen offensichtlich keine Mannosamin-Derivate nutzen, andere Kohlenhydrate sind daher möglicherweise passendere Substrate für MGE in Kieselalgen.

Für die Untersuchung von Proteinlokalisierungen in Subkompartimenten wird das zu untersuchende Protein oft an GFP-Fragmente fusioniert, welche sich selbst zu fluoreszierenden Einheiten zusammensetzen können (self-assembling GFP). Wir beobachteten GFP-Fluoreszenz in transformierten Zelllinien, die theoretisch keine self-assembling GFP Fragmente im selben zellulären Kompartiment enthalten sollten. Daraus wird ersichtlich, dass der Transport der verwendeten Fusionsproteine durch die self-assembling GFP Fragmente stark beeinflusst ist und dass die Anwendung dieser Methode umfassende Kontrollexperimente erfordert.

Mit dem Ziel, Transgen-Expression von Reporterproteinen in *P. tricornutum* zu steuern und abzustimmen, wurde der induzierbare Nitratreduktase (nr) Promoter in zeit-auflösender Weise durch Durchflusszytometrie charakterisiert. GFP-Expression wurde durch das vorgegebene Medium induziert und stieg innerhalb der ersten Stunde, wohingegen eine Abnahme nach Wechsel des Mediums nur langsam erfolgte. Der nr Promoter zeigte weiterhin

geringe Aktivität auch unter inaktivierender Bedingungen. Die einzelnen Zelllinien unterscheiden sich voneinander, sowohl in der Expressionsstärke als auch in den Expressionsmustern.

Die Untersuchungen von NTTs in Kieselalgen haben gezeigt, dass Transporter in der ER Membran vorhanden sind, die den Transport von Purinnuklotiden ermöglichen. Des Weiteren haben Kieselalgen-NTTs zusätzliche Funktionen außer der Energiebereitstellung an die Plastiden. Sie sind darüber hinaus auch an anderen physiologischen Aufgaben beteiligt, wie am Nukleotidimport in die Plastiden und möglicherweise auch an Signalwegen.

Content

1 General Introduction

1.1 Diatoms

Diatoms (Bacillariophyceae) are unicellular, photoautotrophic eukaryotic organisms belonging to the group of Stramenopiles. Diatoms can be classified into two major groups, depending on the shape of their valves: the pennate diatoms (bilateral symmetrical, e.g. *Phaeodactylum tricornutum*) and the centric diatoms (radial symmetrical, e.g. *Thalassiosira pseudonana*). They represent an ecologically highly relevant group of phytoplankton, being abundant in the oceans and responsible for up to 20 % of the global carbon-fixation (Nelson 1995). Their ecological relevance, easy and cheap culturing conditions and the availability of molecular tools makes diatoms interesting organisms to study in order to engineer these organisms for biotechnological purposes (Kilian & Kroth 2006, Kroth 2007a). Due to the process of secondary endocytobiosis, which is the uptake of an eukaryotic alga by another eukaryotic host cell (Figure 1), several algal groups harbour so-called complex plastids (or secondary plastids), including haptophytes, cryptomonads, stramenopiles (to which diatoms belong to), and also apicomplexans, which are all derived by the engulfment of a red alga (Delwiche 1999, Keeling 2013, Sheiner & Striepen 2013). Resulting from this event, diatoms possess plastids surrounded by four membranes instead of two, as it is the case in higher plants or green algae with plastids evolved by primary endocytobiosis (the engulfment of a cyanobacterium by a eukaryotic host cell) (Delwiche & Palmer 1997, Martin & Müller 1998, Martin *et al.* 1998). The compartments from inside the plastid to outside are classified as: stroma, interenvelope space (IES), periplastidial space (PPS), chloroplast ER lumen (see Figure 3). The outermost compartment is connected with the endoplasmic reticulum and is therefore also called chloroplast ER (cER). The membranes of the endoplasmic reticulum (ER), the chloroplast ER membrane, and also the nuclear membrane are continuous. Resulting from secondary endocytobiosis, genes were transferred from the endosymbiont to the host nucleus and parts or whole metabolic pathways including involved enzymes were reduced, lost or relocated (Gould *et al.* 2008, Keeling 2010, Nisbet *et al.* 2004).

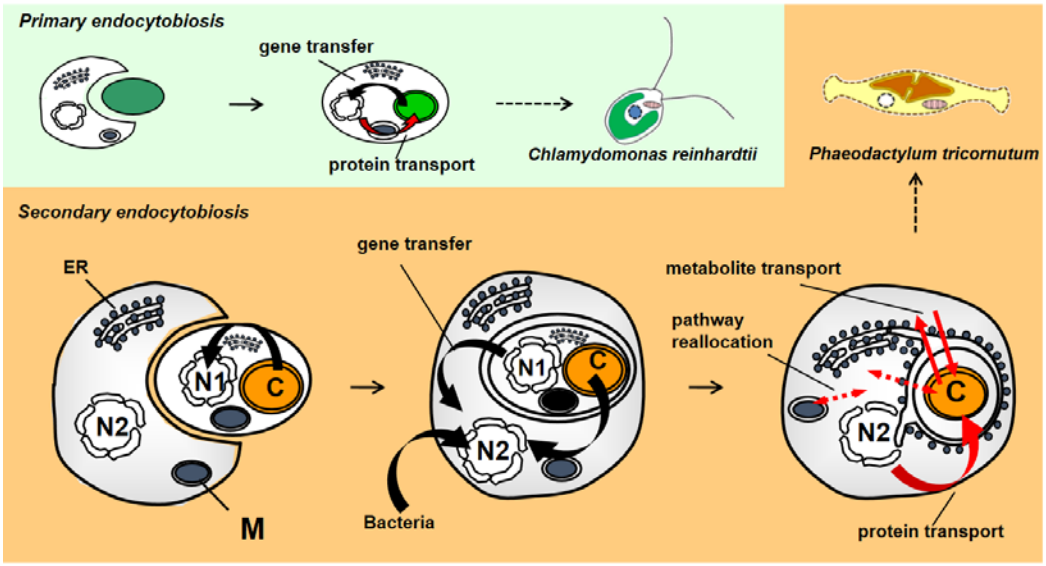


Figure 1: Land plants and algae evolved by primary endocytobiosis, which is the uptake of a cyanobacterium by a eukaryotic host cell. The evolved organisms harbour primary plastids. Diatoms evolved by secondary endocytobiosis. This is the uptake of a eukaryotic red alga which already contains plastids by another eukaryotic host cell. Subsequently, diatoms harbour so-called complex or secondary plastids, surrounded by four membranes, the outermost membrane being continuous with the ER and the nuclear membrane. Modified according to (Kroth 2007a)

1.2 Protein targeting and translocation into diatom plastids

Most of the proteins which are transported across the eukaryotic ER are translocated through a protein conducting channel (for reviews see (Rapoport 2007, Rapoport *et al.* 2004, Rapoport *et al.* 1996)). In prokaryotes there is a similar process, except that the proteins are transported across and are integrated into the plasma membrane (Rapoport 2007, Rapoport *et al.* 1996). The translocation happens in both cases by a conserved membrane-protein complex, called Sec61 in eukaryotes or SecY in bacteria (Rapoport 2007, Rapoport *et al.* 2004). The polypeptide chains can be transferred co-translationally to the channel by ribosomes, which is initiated by the recognition of a signal or transmembrane sequence and a signal-recognition particle (SRP). The SRP is then interacting with its membrane receptor and the polypeptide is released into the associated channel (Figure 2A, (Rapoport 2007)).

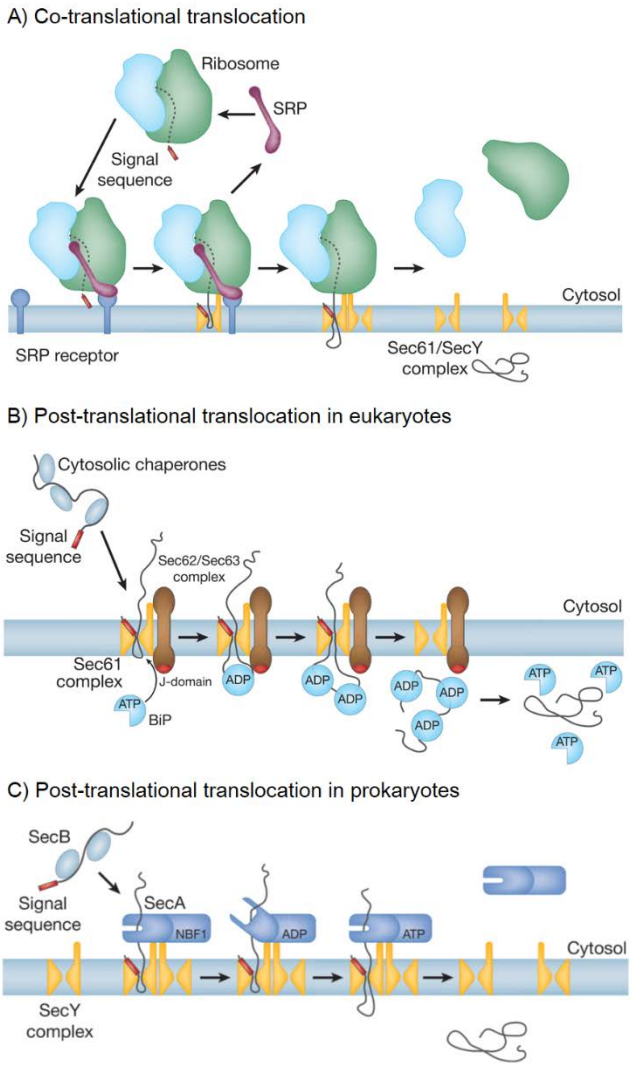


Figure 2: Different translocation processes across the ER membrane via Sec protein complexes (Rapoport 2007). A) Co-translational transport with interaction of a signal sequence, a signal recognition particle (SRP) and its receptor. B) Post-translational transport in eukaryotes (ratcheting mechanism). The ER luminal chaperone prevents the protein from moving back. C) Post-translational transport in prokaryotes (pushing mechanism). See Text for details.

Some proteins are transported after translation, post-translationally, probably due to a second post-targeting signal recognition event in the ER (Jungnickel & Rapoport 1995). It was postulated that this process is more common in bacteria and yeast, possibly due to fast-growing cells, where translocation would not keep pace with translation (Rapoport 2007). However, this mechanism is used mostly by soluble proteins, which possess only moderate hydrophobic domains and remain loosely folded after release from the ribosome (Huber *et al.* 2005a, Huber *et al.* 2005b, Ng *et al.* 1996). In yeast, and probably in all eukaryotes, post-translational transport requires the interaction of the protein complexes, Sec61 and Sec62/Sec63, and the luminal chaperone BiP (Deshaies *et al.* 1991, Panzner *et al.* 1995). The translocation starts with the binding of a signal sequence to the channel (Figure 2B, (Rapoport 2007)). The cytosolic chaperones release the substrate and the polypeptide is inserted into the channel with the help of Brownian motion and BiP inside the ER lumen preventing movements back to the cytosol (ratcheting mechanism) (Matlack *et al.* 1999). In prokaryotes the interaction partner is the cytosolic ATPase SecA (Figure 2C, (Rapoport

2007)). The translocation is starting with the binding of a cytosolic chaperone (SecB) (Randall *et al.* 1997), which is accepted by SecA. After insertion of the polypeptide chain into the channel, the substrate is translocated by a 'pushing' mechanism, which is not clear in detail, but possibly caused by the peptide-binding groove of SecA pushing the substrate towards the channel (Economou & Wickner 1994).

Due to gene transfers from the genome of the endosymbiont to the host nucleus, the majority of plastid proteins in diatoms is nucleus-encoded and depends on a targeting- and transport-mechanisms strongly correlated to their bipartite presequences (Apt et al. 2002, Lang et al. 1998). The characteristic bipartite presequence consists of a signal peptide and a transit peptide domain (Figure 3). The N-terminal region is necessary for transport to the ER, while the C-terminal region facilitates the protein to traverse the other envelope membranes (Apt *et al.* 2002). This import system includes the transport of plastidial proteins via the ER and therefore requires the presence of a signal peptide domain to allow the import into the ER (Apt et al. 2002, Bhaya & Grossman 1991). Once entered the ER lumen, the transition of the pre-protein could occur either by further transport through other transporter proteins present in the remaining membranes ('translocator model') or by vesicle trafficking across the periplastidial space ('vesicle shuttle model') (Cavalier-Smith 1999, Gibbs 1979, McFadden 1999, van Dooren 2001, Vugrinec *et al.* 2011). In both cases, the process is initiated by the co-translational import of the polypeptide chain into the ER lumen via the Sec protein translocator after recognition of a signal peptide (Kilian & Kroth 2005). The signal peptide is cleaved off after entering the ER lumen and further transport is possibly mediated by the 'translocator model' (Figure 3): the symbiont-derived ER-like machinery (SELMA) in the second outermost membrane (Hempel *et al.* 2009) would be the next transporter, followed by proteins similar to the Tic/Toc complex of land plant plastids and being situated in the two innermost membranes (Gruber *et al.* 2007). Here, a protein derived from the cyanobacterial Omp85 is thought to take over the translocation of proteins across the second innermost membrane in *P. tricornutum* (*Pt*Omp85) and to be the equivalent to the Toc75 protein in the outer envelope of chloroplasts (Bullmann *et al.* 2010). Transport across the innermost membrane is the same in both models and would be mediated by a Tic related translocon. The 'vesicle shuttle model' is based on electron microscopy (Gibbs 1979), where vesicles were found in the periplastidial space and would be responsible for transporting the proteins from one compartment or membrane to the next one (Figure 3).

For integral membrane proteins with multiple transmembrane domains, the translocation is thought to happen in a multistep process and co-translationally by releasing the polypeptide chain into the ER translocon Sec61 to subsequently insert the transmembrane domains into the phospholipid layer (for review see (Martinez-Gil *et al.* 2011, Shao & Hegde 2011)). The targeting process is also supposed to be dependent on the composition of the presequences and transport across the membranes would probably also occur either via the 'translocator model' or the 'vesicle shuttle model'.

Furthermore, since the protein transport into complex plastids requires the transition of the ER lumen, protein modifications would be feasible. Protein glycosylation during the transport could be possible, since N-glycosylation for example has been shown to occur in microalgae (Baiet *et al.* 2011, Mathieu-Rivet *et al.* 2014, Mathieu-Rivet *et al.* 2013). Accordingly, any plastidial protein could be involved in glycosylation, which might play a crucial role for protein import into plastids or integration processes of membrane proteins of complex plastids. Besides, it was already shown, that proteins can still be transported across translocation machineries in *P. tricornutum* despite bulky glycosylation sites (Peschke *et al.* 2013).

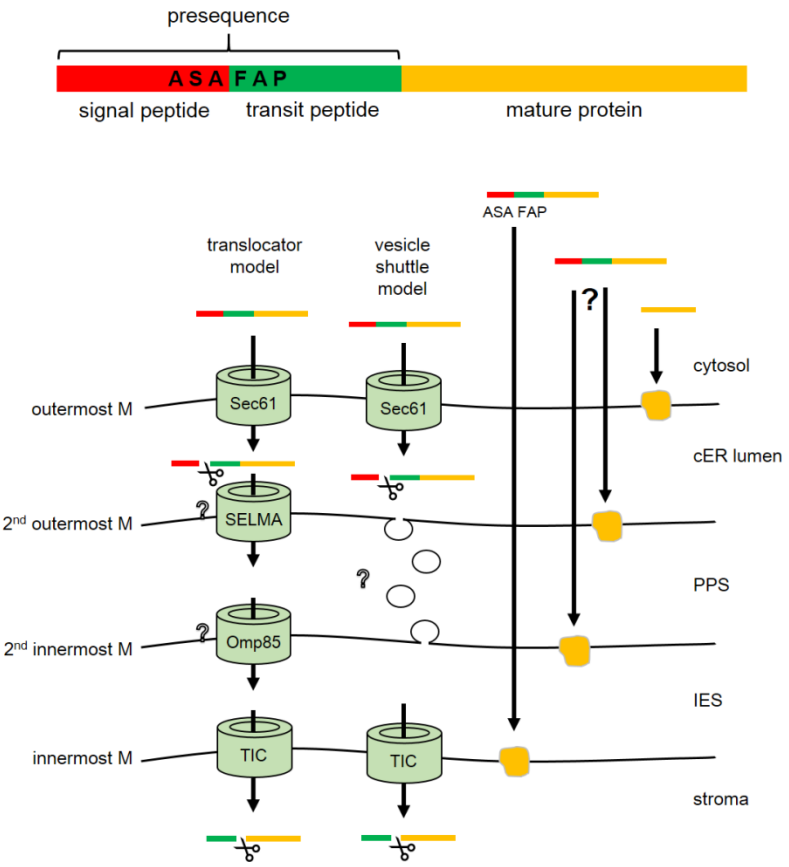


Figure 3: Schematic illustration of a characteristic bipartite presequence of nucleus-encoded plastid proteins in diatoms. The presequence consists of a signal (red) and a transit (green) peptide proceeding the mature protein (yellow). A so-called 'ASAFAP'-motif is crucial for import into the stroma. Depending on the composition of the presequence, the proteins are directed and transported across the membranes via different translocators into the subcompartments and inserted into the membranes of the plastid. (Gruber *et al.* 2007, Kilian & Kroth 2005, Lang *et al.* 1998)

1.3 Nucleotide metabolism in prokaryotes and eukaryotic organisms harbouring plastids

In land plants and green algae, the *de novo* biosynthesis of nucleotides takes place in the plastids (Ast *et al.* 2009, Zrenner *et al.* 2006). So far, *ntt* (nucleotide transporters) genes were identified and characterised in eukaryotic organisms harbouring plastids and in prokaryotes. In red algae and land plants plastids there are two types of nucleotide translocators (NTTs) with the same substrate specificity (antiport of ATP and ADP) (Möhlmann *et al.* 1998). Both isoforms are responsible for the energy provision of the plastid. One NTT is responsible for the transport of the newly synthesised nucleotides out of the organelle, whereas the second NTT is solely mediating the energy provision into the plastid when photosynthetic energy is missing (Heldt 1969, Linka *et al.* 2003, Möhlmann *et al.* 1998, Winkler & Neuhaus 1999).

Bacteria can harbour up to five NTTs, also for energy provision and net uptake of nucleotides (Haferkamp *et al.* 2004, Haferkamp *et al.* 2006b, Schmitz-Esser *et al.* 2004). It was shown, that bacteria living in eukaryotic cells can compensate the lack of a NAD^+ transporter gene by the interaction of a NAD^+/ADP exchanger and an ATP/ADP translocase facilitating the import of intact NAD^+ across the cytoplasmic membrane to exploit the host cell (Haferkamp *et al.* 2004). The substrate specificities in bacterial NTTs of *Protochlamydia amoebophila*

differ from plants: apart from ATP/ADP counter exchange (Schmitz-Esser *et al.* 2004) and NAD^+ transport (Haferkamp *et al.* 2004), the *Pam*NTTs are capable of using RNA nucleotides and transporting UTP, GTP and ATP coupled to a proton-gradient (Haferkamp *et al.* 2006b). For all of these NTTs several amino acid residues could be identified to be highly conserved also in plant NTTs and crucial for nucleotide transport (Haferkamp *et al.* 2006b).
In diatoms, the major role of NTTs is the supply of net nucleotides to the plastid. It was shown, that diatoms are able to regulate the ATP/NADPH ratio by energetic coupling of plastids and mitochondria in terms of importing mitochondrial ATP in case of reducing power generated in the plastid (Bailleul *et al.* 2015).
The majority of the steps for the *de novo* biosynthesis of nucleotides take place in the cytosol, which differs from plants and green algae (Ast *et al.* 2009, Zrenner *et al.* 2006). Thus, for building DNA and RNA in plastids, nucleotides need to be imported across four membranes surrounding the complex plastids of diatoms, which is most likely mediated via NTTs. The NTTs in diatoms are nucleus-encoded proteins most probably targeted to and integrating into plastidial membranes, where they facilitate the exchange and supply of nucleotides from the cytosol to the stroma (Ast *et al.* 2009). Putative diatom NTTs could be identified in the genomes of the model diatoms *P. tricornutum* (see Table 1) and *T. pseudonana* (Armbrust *et al.* 2004, Ast *et al.* 2009, Bowler *et al.* 2008). Not only the amount of NTT isoforms present in diatoms is higher compared to plants or bacteria (six isoform in *P. tricornutum* and eight isoforms in *T. pseudonana*), also the biochemical properties of these NTTs differs from plants and rather shows similarities to bacteria (Ast *et al.* 2009). As shown in Table 1, six different isoforms were found in *P. tricornutum* and their respective homologous NTTs in *T. pseudonana* was identified. The first two NTT isoforms of diatoms characterised so far (NTT1 and NTT2) were shown to integrate most likely into the innermost plastid membrane, where they facilitate the transport of a variety of nucleotides. Interestingly, NTT1 acts as a proton-dependent symporter for adenine nucleotides, whereas NTT2 is capable of transporting all nucleotide triphosphates, as well as their deoxy forms in an antiport-mode. Thus, diatoms seem to possess plastidial NTTs with bacterial features. The high amount of NTT isoforms could be correlated to the additional membranes surrounding the plastid, which also represent additional barriers for the nucleotides to pass.
Phylogenetically, the diatom NTT1 isoform apparently resembles red algal NTTs (Ast *et al.* 2009), most likely due to the gene transfer from the red algal endosymbiont. NTT2 however shares higher similarities with bacteria (Ast *et al.* 2009).
The origin of the remaining NTT isoforms is less obvious, but they are probably derived from a series of horizontal gene transfer events and subsequent gene duplication and modifications.

Table 1: List of putative nucleotide transporters in *P. tricornutum*. (Ast *et al.* 2009)

NTT	homologous NTT	JGI Protein ID	bipartite presequence (ASA-FAP?)	Substrates	presumed plastid localisation
***Pt*NTT1**	*Tp*NTT1	49533	TEA-FAP (yes)	ATP, ADP, AMP, dATP	innermost membrane
***Pt*NTT2**	*Tp*NTT2	45145	ISA-TSS (no)	ATP, GTP, CTP, TTP, UTP, dATP, dGTP, dCTP	innermost membrane?
***Pt*NTT3**	*Tp*NTT4	50189	-	?	one of the outer membranes?
***Pt*NTT4**	*Tp*NTT5	46794	-	?	one of the outer membranes?
***Pt*NTT5**	*Tp*NTT6/7/8	54110	-	?	one of the outer membranes?
***Pt*NTT6**	*Tp*NTT6/7/8	54907	VRA-LLP (yes)	?	innermost membrane?

1.4 Objectives of this thesis

The aim of this work was to study nucleotide translocation by the investigation of transporter proteins (NTTs) in the diatoms *Phaeodactylum tricornutum* and *Thalassiosira pseudonana*. NTTs should be examined in order to reveal the subcellular localisation, biochemical properties and to further understand targeting, transport- and insertion-processes of plastidial membrane proteins of complex plastids.

The NTT isoform 5 of *P. tricornutum* (*Pt*NTT5) was characterised by the heterologous expression of the transporter protein in *E. coli* strains, radioactively labelled transport measurements and GFP-based localisation studies (Chapter 2). Additionally, certain protein domains of *Pt*NTT5 were deleted to study the potential influence on the transport mechanism and the insertion processes into the plastid membrane. Moreover, the biochemical properties of NTT3 of the centric diatom *T. pseudonana* (*Tp*NTT3) were determined, as well as the subcellular localisation of the protein in this diatom (Chapter 3).

For further characterisation of plastidial proteins, different molecular tools were used. Metabolic glycoengineering (MGE) was applied to study protein glycosylation, which could possibly influence transport and/or integration mechanisms of plastid-targeted proteins (Chapter 4). Furthermore, GFP-based localisation studies of plastidial proteins initiated detailed investigations on a system utilising self-assembling GFP fragments (Chapter 5) and the examination on the promoter activity used in the study (Chapter 6).

2 Shuttling of (deoxy-) purine nucleotides between compartments of the diatom *Phaeodactylum tricornutum*

Lili Chu[1], Michelle Ast[2], Ansgar Gruber[1], Stephan Schmitz-Esser[3], Jacqueline Altensell[2], Horst Ekkehard Neuhaus[2], Peter G. Kroth[1], Ilka Haferkamp[2*]

[1] Pflanzliche Ökophysiologie, Universität Konstanz, 78457 Konstanz, Germany

[2] Pflanzenphysiologie, Technische Universität Kaiserslautern, 67653 Kaiserslautern, Germany

[3] Department of Animal Science, Iowa State University, Ames, IA, 50011, USA

*Corresponding author: haferk@rhrk.uni-kl.de

Keywords: complex plastid / diatom / endoplasmic reticulum (ER) / energy / nucleotide transport / nucleotide synthesis

Chu L, Gruber A, Ast M, Schmitz-Esser S, Altensell J, Neuhaus HE, Kroth PG & Haferkamp I (2016b): Shuttling of (deoxy-) purine nucleotides between compartments of the diatom *Phaeodactylum tricornutum*. New Phytol.

doi: 10.1111/nph.14126

2.1 Abstract

- Diatom plastids show several peculiarities when compared with primary plastids of higher plants or algae. They are surrounded by four membranes and depend on nucleotide uptake because, unlike in plants, nucleotide *de novo* synthesis exclusively occurs in the cytosol. Previous analyses suggest that two specifically adapted nucleotide transporters (NTTs) facilitate the required passage of nucleotides across the innermost plastid membrane. However, nucleotide transport across the additional plastid membranes remains to be clarified.
- Phylogenetic studies, transport assays with the recombinant protein as well as GFP-based targeting analyses allowed detailed characterization of a novel isoform (*Pt*NTT5) of the six NTTs of *Phaeodactylum tricornutum*.
- *Pt*NTT5 exhibits low amino acid similarities and is only distantly related to all previously characterized NTTs. However, in a heterologous expression system, it acts as a nucleotide antiporter and prefers various (deoxy-) purine nucleotides as substrates. Interestingly, *Pt*NTT5 is probably located in the endoplasmic reticulum, which in diatoms also represents the outermost plastid membrane.
- *Pt*NTT5, with its unusual transport properties, phylogeny and localization, can be taken as further evidence for the establishment of a sophisticated and specifically adapted nucleotide transport system in diatom plastids.

2.2 Introduction

All three lineages of photosynthetic organism, glaucophytes, rhodophytes, and chlorophytes (including land plants), according to the best supported hypothesis, gained their plastids in a process called primary endosymbiosis. During this primary endosymbiosis, a cyanobacterium was engulfed, retained and partially degraded by a heterotrophic host (Keeling 2013). Massive transfer of cyanobacterial genes to the host nucleus as well as metabolic connection and rearrangement of the consortium resulted in control and functional embedding of the symbiont and finally in the establishment of the photosynthetic organelle (Archibald 2015). Interestingly, a considerable number of the diverse algal lineages arose from secondary endosymbiosis, a process which occurred several times and which includes the fusion, and thus the genetic and metabolic connection of two eukaryotes, a host cell and an alga with a primary plastid (Archibald 2015, Keeling 2013).

Diatoms are an important group of algae with secondary plastids: They are of ecological relevance and of particular scientific interest because they are highly abundant in the oceans; they are among the most common species of the phytoplankton and contribute immensely to carbon fixation (Armbrust 2009, Falkowski & Oliver 2007). Genome analyses suggest that their rapid diversification, enormous success and predominance in the marine ecosystem might result from an extraordinary genetic composition/flexibility (Bowler *et al.*, 2008; Armbrust, 2009; Archibald, 2015).

The complex plastids of diatoms are of red-algal origin and possess four surrounding membranes with the outermost membrane being connected to the rough ER (Gould *et al.* 2008, Kroth 2002). Compared with plants, diatoms exhibit interesting metabolic peculiarities. For instance, nucleotide *de novo* synthesis is cytosolic in diatoms (Ast *et al.* 2009, Gruber *et al.* 2009) whereas in higher plants, the complete pathway of purine and several steps of pyrimidine nucleotide synthesis take place within the plastid (Witz *et al.* 2012, Zrenner *et al.* 2006). Accordingly, in diatom plastids, nucleotides have to be

transported across additional membranes and purine nucleotides even in an opposite direction (Ast *et al.*, 2009) compared with higher plant plastids (Witz *et al.* 2012, Zrenner *et al.* 2006).

Previous studies indicate that at least two nucleotide transporter (NTT)-type carriers mediate nucleotide provision to the diatom plastids (Ast *et al.* 2009), whereas in higher plant plastids, transport of purine nucleotides and intermediates of pyrimidine nucleotide synthesis apparently involves different types of carriers (Kirchberger *et al.* 2008, Leroch *et al.* 2005, Witz *et al.* 2012). Generally, NTT proteins are restricted to only a few organismic groups; they exist in certain obligate intracellular living bacteria and represent a hallmark of the orders *Chlamydiales* and *Rickettsiales* (Schmitz-Esser *et al.* 2004). *Chlamydiales* and *Rickettsiales* are metabolically impaired (Moulder, 1991;Moran, 2002), lack several anabolic pathways, including *de novo* nucleotide synthesis, and exploit the nucleotide and energy pool of the host via interaction of functionally diverse NTT-type proton-symporters and exchangers (Audia & Winkler 2006, Haferkamp *et al.* 2004, Haferkamp *et al.* 2006b, Knab *et al.* 2011, Krause *et al.* 1985, Tjaden 1999).

Among eukaryotes, NTTs were detected in algae (with primary and secondary plastids) and higher plants (Ast *et al.* 2009, Linka *et al.* 2003, Möhlmann *et al.* 1998, Tjaden *et al.* 1998b, Tyra *et al.* 2007) and in protists of the phylum *Microsporidia* (Heinz *et al.* 2014, Tsaousis *et al.* 2008), whereas other protists, fungi and animals apparently lack such *ntt* sequences in their genomes. NTTs of primary plastids were shown to operate exclusively as ATP/ADP exchangers (Linka *et al.* 2003, Möhlmann *et al.* 1998, Tjaden *et al.* 1998b) in the inner plastid envelope. Analyses of mutant plants clearly demonstrated their role in energy provision to the stroma, particularly when photosynthesis is insufficient or missing (Geigenberger *et al.* 2001, Reinhold *et al.* 2007, Reiser *et al.* 2004, Tjaden *et al.* 1998a). *Microsporidial* NTTs reside in the plasma membrane or in the mitosome (a mitochondrial relict unable to produce energy in form of ATP) and mediate ATP/ADP exchange or purine nucleotide transport (Heinz *et al.* 2014, Tsaousis *et al.* 2008).

Although, in recent years, the biochemical characteristics and physiological relevance of NTTs from several organisms were revealed, our knowledge about NTTs from complex plastids is comparably limited. This is because (1) metabolite transport into and/or out of complex plastids has not been investigated in detail, (2) diatoms possess an unusually high number of NTTs (six to eight isoforms) when compared to other algae or to higher plants (one to three isoforms), and finally because (3) the two NTTs from diatoms characterized to date (NTT1 and NTT2) exhibit functional properties previously not documented for any plastidial NTT and only rarely shown for NTTs from other organisms.

NTT1, and probably also NTT2, are located in the innermost envelope of the diatom plastid (Ast *et al.* 2009). The diatom NTT1 isoform acts as a proton-symporter and accepts mono-, di-, and triphosphorylated adenine nucleotides as substrates, whereas the NTT2 isoform transports various triphosphorylated nucleotides, including deoxy-forms, in a counter exchange mode (Ast *et al.* 2009). Physiological interaction of NTT1 and NTT2 has been suggested to deliver all nucleotides for DNA and RNA synthesis to the stroma. Although NTT1 and NTT2 proteins represent important components of the plastidial nucleotide uptake system in centric (*Thalassiosira pseudonana*) and pennate (*Phaeodactylum tricornutum*) diatoms, until now it has not been clarified how all required (deoxy-) nucleotides pass through the remaining membranes of the complex plastid (Ast *et al.* 2009).

In this context, it is important to keep in mind that nucleotide transport can connect the sites of nucleotide synthesis and consumption, and/or play a key role in the adjustment of the subcellular energy states. Just recently, plastids and mitochondria of diatoms were shown to be extraordinarily tightly connected by extensive energetic interaction (Bailleul *et al.* 2015).

This metabolic communication relies on the transport of reducing power from plastids to mitochondria as well as on re-shuttling of energy (from mitochondria to plastids), representing a main factor in the adaptation of cellular ATP/NADPH ratios for optimized photosynthetic performance, carbon fixation and growth (Bailleul *et al.* 2015). Accordingly, the analysis of putative NTT proteins will help to deduce their possible role in the nucleotide and/or energy metabolism.
In this study, we focus on the NTT isoform 5 from the diatom *P. tricornutum* (*Pt*NTT5). *Pt*NTT5 shows comparatively weak relationships to all previously characterized NTTs and here we report that it also is extraordinary in terms of its biochemical properties and subcellular localization.

2.3 Materials & Methods

2.3.1 *In silico* analyses of NTT proteins

Comparative amino acid sequence analysis of NTT proteins was performed using ClustalX (Thompson *et al.* 1997) and Genedoc (Nicholas *et al.* 1997). Presence or absence of putative pre-sequences, signal peptides, cleavage sites, chloroplast transit peptides or transmembrane domains was investigated with help of the programs SignalP 3.0, ChloroP 1.1, TargetP 1.1, TMHMM (http://www.cbs.dtu.dk/services/) (Emanuelsson *et al.* 2007) and ASAFind (http://rocaplab.ocean.washington.edu/tools/asafind) (Gruber *et al.* 2015) following the methods described by (Gruber & Kroth 2014).
NTT amino acid sequences were aligned with MAFFT (Katoh & Standley 2013), and phylogenetic trees were reconstructed with MEGA6 (Tamura *et al.* 2013) by using the neighbor-joining method and the Poisson correction, the parsimony bootstrap method, and the maximum likelihood method (using the Jones-Taylor-Thornton [JTT] amino acid substitution model); all trees were calculated with 1,000× bootstrapping. All positions containing gaps and missing data were eliminated. There were a total of 363 positions in the final data set.

2.3.2 Cultivation of *Phaeodactylum tricornutum*

P. tricornutum Bohlin 1897, strain 646 (available at UTEX Culture Collection of Algae, University of Texas, Austin, TX, USA; https://utex.org) was cultivated in f/2 medium at a concentration of 50% sea water (16,6 g of sea salt; Tropic Marin Dr. Biener, GmbH, Wartenberg, Germany) in 1L distilled water, adjusted to pH 7,0. Supplements were added like described in (Kroth 2007b). For agar plate cultivation, corresponding solid media were complemented with 1.2% Bacto Agar (Becton, Dickinson & Co., Le Pont de Claix, France). Cells were grown under continuous illumination at 35 µmol photons $m^{-2}s^{-1}$ or in day-night-cycle (16 h : 8 h, light : dark; for protein preparation and western-blotting).

2.3.3 Preparation of cDNA and generation of expression constructs

RNA isolation was performed with the RNeasy plant mini kit (Qiagen, Hilden, Germany) and Poly(A^+)-RNA was enriched from this preparation with the Oligotex kit (Qiagen). Production

of cDNA from mRNA templates was conducted with the reverse transcriptase SuperscriptII (Invitrogen, Carlsbad, USA) according to the manufacturer´s instruction. The coding sequence of *Pt*NTT5 was amplified via PCR and oligonucleotides (MWG, Ebersberg, Germany or Sigma Aldrich, Munich, Germany) that allow compatible insertion into the expression vectors. The amplification products were restricted with specific endonucleases, purified from the agarose gel with the NucleoSpin Extract II Kit (Macherey & Nagel, Düren, Germany).
For generation of the *E. coli* expression construct the coding sequence of full-length *Pt*NTT5 was inserted via the restriction sites NdeI and XhoI in frame with the Histidine tag into the Isopropyl-b-D-thiogalactoside inducible expression vector pET16b (Novagen, Heidelberg, Germany).
For the localization studies, the start codon of the eGFP gene of the pPha-T1-GFP plasmid (Gruber *et al.* 2007) (a modified version of the shuttle vector pPha-T1, GenBank AF219942 (Zaslavskaia *et al.* 2000) has been replaced by the codon "GGA" (coding for glycine) via site-directed mutagenesis to inactivate the potential translation start. The *Pt*NTT5 amplicon was ligated into the StuI site of this modified pPha-T1-GFP plasmid. Truncated versions of *Pt*NTT5 were produced via PCR or deletion-PCR with specific oligonucleotides (Supplementary Table S 1) and using the vector containing full-length *Pt*NTT5 with either N- or C-terminal GFP fusion as template.
Amplification products were purified and either inserted into the pPha-T1 shuttle vectors or re-ligated (deletion-PCR products). Correctness of the respective constructs was verified by restriction analyses (Fermentas, St. Leon-Rot, Germany) and sequencing (SEQ-IT, Kaiserslautern, Germany or GATC Biotech AG, Konstanz, Germany).

2.3.4 Heterologous expression in *Escherichia coli* and import measurements

The *E. coli* strain Rosetta 2(DE3)pLysS (Novagen, Heidelberg, Germany) was used for heterologous synthesis of *Pt*NTT5. To analyze transport properties of the recombinant *Pt*NTT5 either induced or non-induced *E. coli* cells harboring the expression plasmid pET16b-*Pt*NTT5 were incubated in phosphate buffer (KP_i) supplemented with the given concentrations of the respective [α^{32}P] labeled substrates. Import was conducted at 30°C and terminated by removal of external substrate using vacuum filtration and washing. Radioactivity was quantified by scintillation counting (Tricarb 2500; Canberra-Packard, Heidelberg, Germany).

2.3.5 Nuclear transformation and microscopy

Nuclear transformation of *P. tricornutum* was performed as described previously (Kroth 2007b). Cellular localization of GFP fusion proteins was analyzed with a confocal laser scanning microscope LSM 780 or LSM 510 META (Carl Zeiss, Oberkochen, Germany) using a Plan-Apochromat 63 × 1.4 oil immersion Nomarski differential interference contrast (DIC) objective (Carl Zeiss) or an epifluorescence microscope Olympus BX51 (Olympus Europe, Hamburg, Germany) equipped with a Zeiss AxioCam MRm digital camera (Carl Zeiss) and an Olympus PLN 40x objective (Olympus). Nucleic acids were stained with the dye Hoechst 33342 (CALBIOCHEM®, Behring Diagnostics, La Jolla, CA) to show the localization of the nucleus. Mitochondria were stained with the dye MitoTracker® Orange CM-H_2TMRos (Molecular Probes, Inc., Eugene, OR) to show localization of the mitochondria. Image processing was conducted using ZEN lite and AxioVision Rel. 4.7 (Carl Zeiss).

2.3.6 Western blot analyses and SDS-PAGE

Cells were harvested during exponential phase of growth by centrifugation (3000 g, 10 min, 4 °C) and the pellet was resuspended in 1 mL lysis buffer containing protease inhibitor "complete EDTA-free" (Roche, Mannheim, Germany), 50 mM Tris HCl pH 8, 1 mM EDTA, 8 M urea and 1 % SDS. A mixture of glass beads (0,1-1 mm diameter) was added and cells were homogenised in a Fast Prep FP 120 Bio 101 Savant (Qbiogene Inc., Carlsbad, CA, USA) at maximum speed 6 times 20 sec, 1 min on ice in between. Samples were again centrifuged to remove non-disrupted cells. Total protein concentration in the supernatant was determined by the 660 nm Pierce Protein Assay (Thermo Scientific, Schwerte, Germany). 5 µg of this protein extract were analyzed in a discontinuous, denaturing system with 10 % or 12 % separating polyacrylamide gels (Laemmli 1970) and further transferred onto a nitrocellulose membrane (Whatman Protran BA 79, Whatman Inc., Maidstone, Kent, UK) electrophoretically using a Trans-Blot Turbo (Bio-Rad, Hercules, California, USA). Page Ruler Prestained Protein Ladder (Thermo Scientific) was used as marker.
Immunodetection was performed using an anti-GFP antibody (Invitrogen, Molecular Probes, Eugene, USA) combined with anti-Rabbit IgG peroxidase antibody (Sigma Aldrich). Western blots were treated with Roti®-Block, Roti®-Blue and Roti®-Lumin plus like described in the manuals (Carl Roth GmbH & Co. KG, Karlsruhe, Germany). Chemoluminescence was documented with Amersham Hyperfilm ECL (GE Healthcare, Little Chalfont, UK) and Konica SRX-201 (Konica Minolta, Chiyoda, Japan).

2.4 Results

2.4.1 The evolutionary origin of *Pt*NTT5

To get first insights into the possible evolutionary origin of diatom NTTs, we performed a phylogenetic analysis with published and biochemically characterized NTTs from various organisms as well as NTT isoforms from *T. pseudonana* (*Tp*NTT1-8) and *P. tricornutum* (*Pt*NTT1-6).
The overall topology of the obtained phylogenetic tree (Figure 4) is in agreement with previous studies performed with larger data sets including also various NTTs of unknown function (Ast *et al.* 2009, Heinz *et al.* 2014, Tsaousis *et al.* 2008, Tyra *et al.* 2007).
Microsporidial isoforms generally constitute a separate branch among the NTTs (Figure 4; (Heinz *et al.* 2014, Tsaousis *et al.* 2008)). The isoforms of *Trachipleistophora hominis* and *Encephalitozoon cuniculi* form two distinct clusters. This is in line with recent analyses suggesting that the different NTT isoforms of contemporary microsporidians arose from repeated, lineage-specific gene duplications whereas the ancestor of microsporidians most likely possessed only a single NTT gene (Heinz *et al.* 2014).
It is widely accepted that the early primary photosynthetic eukaryote most likely acquired its nucleotide transporter gene from a donor closely related to extant chlamydiae (Greub & Raoult 2003, Schmitz-Esser *et al.* 2004, Tyra *et al.* 2007), and also in our analysis a monophyletic grouping of plastidial and chlamydial ATP/ADP transporters (bacterial NTT1 isoforms) becomes evident (Figure 4). Moreover, plastidial NTTs from higher plants, red algae and diatom NTT1 isoforms form a monophyletic cluster and the close relationship as

sister groups suggests that diatom NTT1 proteins originated from the red algal progenitor engulfed during secondary endosymbiosis (Figure 4; (Ast *et al.* 2009)).
Diatom NTT2 proteins and *Tp*NTT3 form a deep branch together with functionally diverse NTTs from rickettsiae (Ast *et al.* 2009) or even constitute a separate group among various bacterial and plastidial NTTs (Figure 4). The remaining diatom NTT isoforms also form a distinct subgroup of NTTs with two well-supported subclusters (Figure 4). One subcluster is formed by *Pt*NTT3, *Pt*NTT4 and the respective homologs from *T. pseudonana*, *Tp*NTT4 and *Tp*NTT5. *Pt*NTT5 and its possible homolog *Tp*NTT6 constitute the second subcluster together with *Pt*NTT6, *Tp*NTT7 and *Tp*NTT8. Therefore, these diatom NTTs seem to be only weakly related to other bacterial and plastidial NTTs and only the cluster formed by the microsporidial NTTs exhibits a higher phylogenetic distance.
The branching pattern indicates that the diatom NTT1 isoform originated from endosymbiotic gene transfer, whereas the ancestor of the remaining diatom NTTs cannot be resolved. It is possible that those proteins arose from the red algal derived NTT1 isoform via several gene duplication events. If this would be the case, the origin of these isoforms is apparently masked by considerable alterations in the amino acid sequences maybe due to functional constraints. Alternatively, the remaining diatom NTTs might result from horizontal gene transfer from bacterial (or maybe also microsporidial) donors, and subsequent gene duplications led to the establishment of the respective subgroups of NTTs in diatoms. Their distant relation to the plant NTTs, including the diatom NTT1 isoforms, and the fact that horizontal gene transfer is an important and common phenomenon in diatom evolution, might be taken to support the second scenario (Bowler *et al.* 2008).
In fact, the nuclear genome of diatoms appears as a patchwork and comprises a comparably high number of genes from different bacterial sources (Bowler *et al.* 2008). The higher GC-content of *Pt*NTT5 and *Pt*NTT6 (Table S 2), but most importantly the different numbers of introns (Figure S 1) suggest that the remaining NTT isoforms (with no or, in few cases, one or two introns) do not represent modified duplications of NTT1 (with three introns), but instead were gained by at least one or several independent horizontal gene transfers. The intron number of putative homologs also differs between the two diatom species and suggests a slightly higher tendency for intron gain in *T. pseudonana*. This is in line with comparative genome analyses documenting a higher number of introns in *T. pseudonana* when compared to *P. tricornutum* (Bowler *et al.* 2008). The diatom isoforms *Pt*NTT3-6 and *Tp*NTT4-8 are phylogenetically clearly separated from the previously characterized NTTs. This is why we were interested in a detailed analysis of *Pt*NTT5 as a first representative member of this particular group of diatom NTTs.

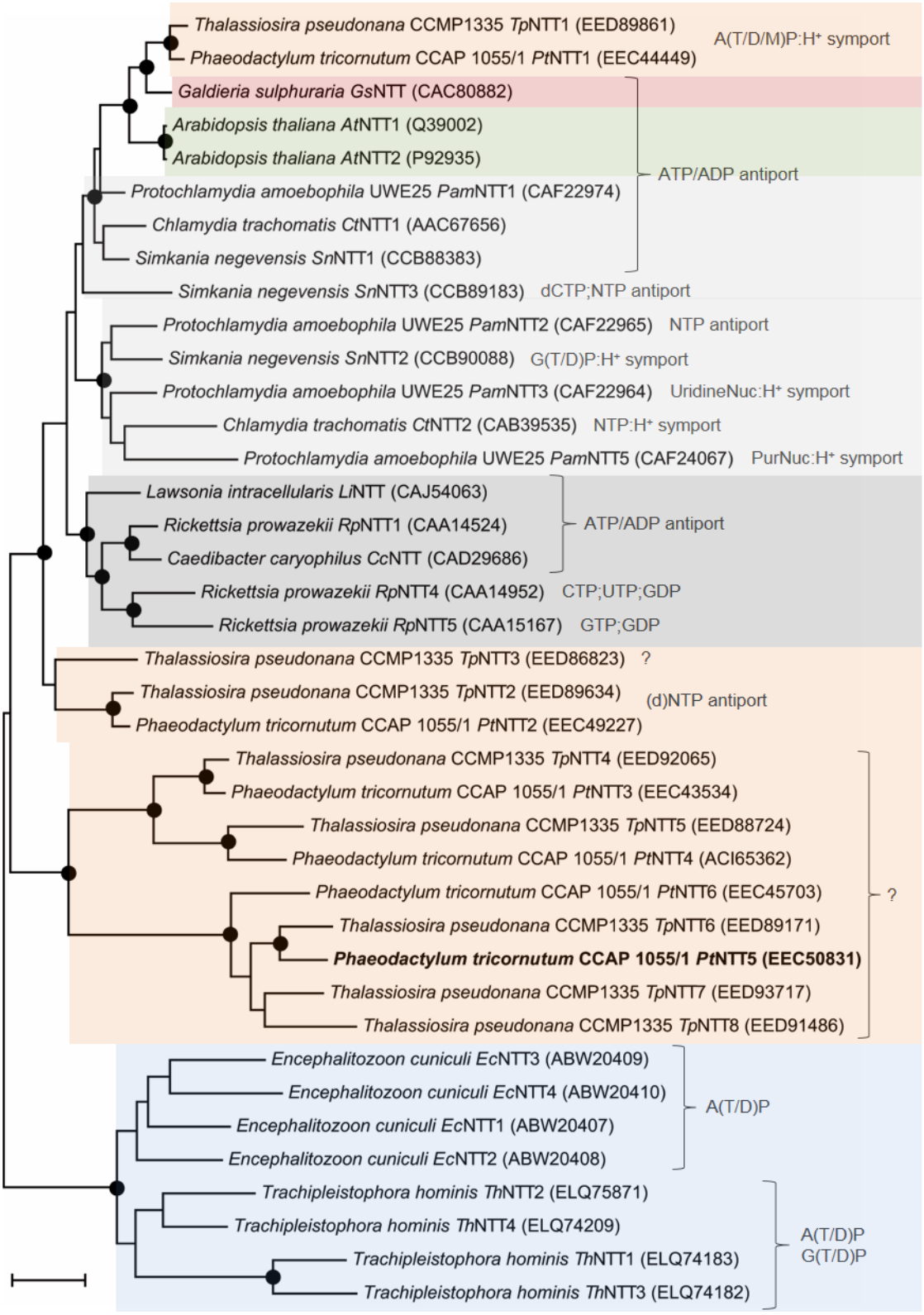


Figure 4: Phylogenetic relationships of NTTs from various organisms. An amino-acid-based phylogenetic tree calculated with MEGA6 using the maximum likelihood algorithm with the JTT model is shown. Black dots indicate nodes supported by maximum likelihood, maximum parsimony, and neighbor-joining bootstrap values (1,000× resampling) greater than 90 %. GenBank accession numbers are given in brackets. The bar represents 20% estimated evolutionary distance. Organismic groups are marked by shading: diatoms in orange, red alga in red, higher plant in green; Chlamydiales, Rickettsiales and Lawsonia intracellularis, grayscales; microsporidia, blue. Basic transport features were taken from (Ast *et al.* 2009, Heinz *et al.* 2014, Knab *et al.* 2011, Schmitz-Esser *et al.* 2008, Schmitz-Esser *et al.* 2004) and literature therein.

2.4.2 Peculiarities in the amino acid sequence of *Pt*NTT5

From our sequence alignment of *Pt*NTT5 and representative, biochemically characterized NTTs from diverse organisms. It became immediately evident that the amino acid sequences of NTTs from plants and algae possess an N-terminal extension when compared to the remaining NTTs (Figure S 2). These extensions generally contain information supporting plastidial targeting. In diatoms, nuclear-encoded proteins with a stromal destination exhibit a characteristic bipartite arrangement of the N-terminal extension (presequence) starting with a signal peptide followed by a transit peptide. Moreover, a conserved motif, the so called "ASAFAP" motif, mostly flanks/overlaps with the predicted signal peptide cleavage site (Gruber *et al.* 2007, Kilian & Kroth 2005). NTT1 and NTT2 from diatoms do show such a typically structured plastidial N-terminal targeting sequence and NTT1 isoforms even contain an authentic "ASAFAP" motif at the signal peptide (Figure S 2; (Ast *et al.* 2009)).
*Pt*NTT5 lacks a comparable N-terminal extension and in this context rather resembles microsporidial NTTs that are mainly located in the plasma membrane. *In silico* analyses also failed to detect a possible N-terminal signal peptide with a corresponding peptide cleavage site (Ast *et al.* 2009). Moreover, prediction of the subcellular localization of *Pt*NTT5 was unclear (Ast *et al.* 2009). However, the absence of characteristic plastidial targeting motifs (Figure S 2, (Ast *et al.* 2009)) implies that *Pt*NTT5 resides in a different compartment/membrane than isoforms 1 and 2. A detailed inspection of the *Pt*NTT5 sequence revealed a comparably high accumulation of positively charged amino acid residues at its far C-terminus; however, so far it is unclear whether this finding is of any relevance for correct protein targeting or membrane insertion.
*Pt*NTT5 generally exhibits only moderate amino acid sequence similarities (24 - 31 %) to previously described NTTs, including diatom NTT1 and NTT2 isoforms (Table S 3). Several regions that are conserved in plant and bacterial NTTs show higher variations in *Pt*NTT5 as well as in microsporidial NTTs (Figure S 2). Moreover, not all predicted transmembrane domains of *Pt*NTT5 did match with the hydrophobic regions of representative, characterized NTTs. Based on sequence similarities it is hard to judge whether *Pt*NTT5 is involved in plastidial nucleotide transport or whether it even fulfills a different physiological function.

2.4.3 *Pt*NTT5 accepts various purine nucleotides as substrates

We determined the functionality and biochemical properties of *Pt*NTT5 using transport measurements with radioactively labeled substrates and *E. coli* cells expressing the recombinant carrier. Initial studies with [α^{32}P]-ATP and [α^{32}P]-ADP revealed that *Pt*NTT5 in fact is able to transport nucleotides (Figure 5). ATP and ADP are efficiently imported into *E. coli* via *Pt*NTT5 whereas non-induced cells (control) showed no or only marginal accumulation of radioactively labeled nucleotides.
To determine the general substrate specificity of *Pt*NTT5, we conducted competition experiments with different non-labeled nucleotides or nucleotide derivatives. Apart from non-labeled ATP and ADP, also dATP and AMP, as well as the corresponding guanidine nucleotides markedly reduced [α^{32}P]-ATP uptake (Table 2). Among the tested purine nucleotides, GDP addition resulted in the lowest decrease of [α^{32}P]-ATP uptake, and the corresponding reduction was similar to that caused by TTP, but higher than that produced by UTP addition. To demonstrate that the given molecules do represent substrates competing with ATP for import, but not inhibitors of the transport protein, we performed uptake studies with the corresponding radiolabeled isotopes. All tested nucleotides were imported and most

of them exhibited rates comparable to the transport rates of ATP and ADP, however, GMP caused slightly lower, whereas deoxy purine nucleotides caused higher transport rates (Figure S 3). Therefore, a broad set of purine nucleotides, including dATP and dGTP, were identified as substrates of *Pt*NTT5.

To examine the biochemical properties of *Pt*NTT5 in more detail, we determined the substrate affinities (Michaelis constant, K_M) as well as the corresponding maximal transport velocities (V_{max}). For this purpose, transport measurements were conducted in presence of rising concentrations of the respective labeled nucleotides (Table 3). *Pt*NTT5 exhibits highest affinities for adenine nucleotides (~ 120 – 150 µM) followed by dATP and dGTP (~ 210 µM), whereas guanidine nucleotides were generally imported with lower affinity (~ 310 – 350 µM). GDP and ATP are imported with lowest maximal velocity (~ 30 and 37 nmol mg protein^{-1} h^{-1}) when compared to the remaining (deoxy-) nucleotides (~ 60 - 84 nmol mg protein^{-1} h^{-1}). These results demonstrate that not only tri-, di- and monophosphorylated adenine and guanidine nucleotides, but also dATP and dGTP represent relevant substrates of *Pt*NTT5.

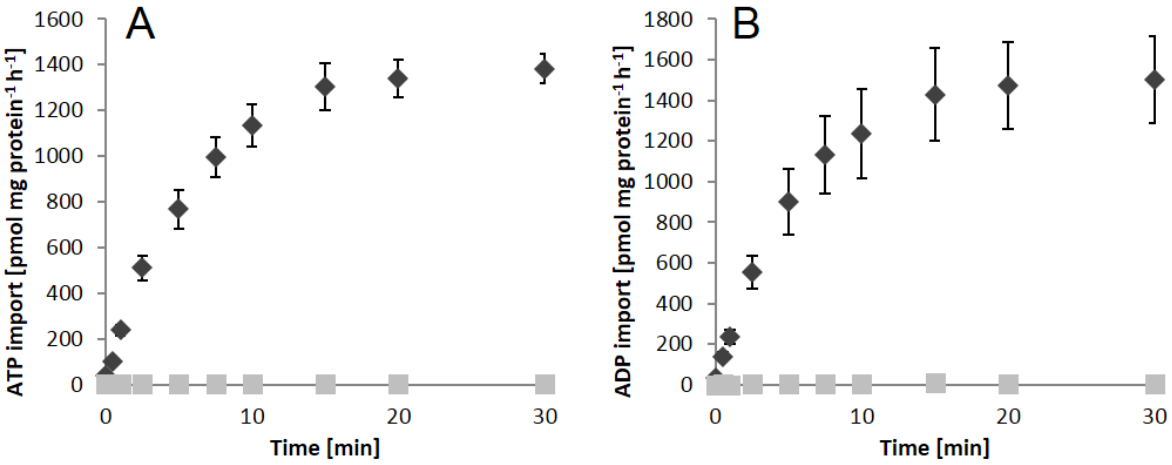


Figure 5**:** Time dependent uptake of ATP and ADP via *Pt*NTT5. Transport of [α^{32}P]-ATP (A) and [α^{32}P]-ADP (B) into Isopropyl-b-D-thiogalactoside - induced *E. coli* cells synthesizing *Pt*NTT5 (black symbols) in comparison to nucleotide uptake into corresponding non-induced *E. coli* cells (gray symbols). Data are the mean of at least three independent experiments, standard errors are given.

Table 2: Effects of various nucleotides and nucleotide derivates on [α^{32}P]-ATP import via *Pt*NTT5

Effector	Import [%]	SE [%]	Effector	Import [%]	SE [%]
(-)	100	0	CDP	71.5	2.9
ATP	**26.0**	**1.0**	TMP	99.7	4.5
ADP	**28.3**	**2.7**	dATP	**33.9**	**2.2**
AMP	**21.7**	**1.7**	dGTP	**30.1**	**1.8**
GTP	**43.0**	**1.5**	dCTP	70.0	1.8
GDP	53.6	1.8	TTP	50.5	6.2
GMP	**37.5**	**2.9**	cAMP	104.5	5.3
ITP	67.1	3.2	cGMP	91.2	4.2
IDP	81.9	3.4	NAD	88.2	5.5
IMP	82.5	5.0	PAP	93.6	4.3
UTP	62.8	2.3	PAPS	97.8	11.5
UDP	81.6	3.4	NADP	95.2	6.2
UMP	96.7	3.4	CoA	89.3	4.7
CTP	78.0	3.6	FAD	88.5	3.0

Uptake of [α^{32}P]-ATP by recombinant *Pt*NTT5 at a substrate concentration of 50 µM was set to 100% (control). Corresponding transport in present of non-labeled effectors (10-fold excess) was calculated according to the control. Import was stopped after 3 min. Nucleotide uptake rates are net values (minus control: ATP uptake into non-induced *E. coli* cells) and the mean of three independent experiments. Bold type is used to highlight significantly reduced import rates (< 50% residual activity).

Table 3: Determination of apparent Michaelis (K_M) and maximal velocity (V_{max}) values of purine nucleotide transport mediated by nucleotide transporter 5 *Pt*NTT5 from *Phaeodactylum tricornutum* (*Pt*NTT)

	Adenine nucleotides			Guanidine nucleotides	
Substrate	K_M-values [µM]	V_{max}-values [nmol mg protein^{-1} h^{-1}]	Substrate	K_M-values [µM]	V_{max}-values [nmol mg protein^{-1} h^{-1}]
ATP	120.7 (13.4)	36.8 (3.7)	GTP	311.5 (19.6)	73.6 (9.6)
ADP	137.0 (1.8)	59.9 (4.8)	GDP	304.0 (23.1)	29.2 (2.4)
AMP	151.0 (8.9)	71.9 (8.3)	GMP	348.0 (31.0)	83.8 (9.1)
dATP	207.2 (17.6)	84.1 (6.4)	dGTP	211.3 (26.7)	82.1 (6.3)

Nucleotide uptake in the presence of rising substrate concentrations (5 µM – 1000 µM) was allowed for time spans in the linear phase of the corresponding transport at 50 µM. Apparent K_M values are given in µM, V_{max} values are given in nmol mg protein^{-1} h^{-1}. Data are the mean of at least three independent experiments. Corresponding standard errors are given in brackets.

2.4.4 *Pt*NTT5 is an antiporter

To reveal the NTT function in the cellular context, it is important to determine its mode of transport. In principle, NTT-type proton-symporters enable net gain of nucleotides, whereas NTT-type antiporters either allow energy uptake by the exchange of ATP and ADP or the adaptation of the composition of the respective nucleotide pools by exchanging internal with external (deoxy-) nucleotides.

To analyze whether recombinant *Pt*NTT5 is driven by the proton-gradient across the *E. coli* membrane, we investigated the influence of the protonophore m-chlorophenylhydrazone (CCCP) on ATP import (Figure 6A). Interestingly, nucleotide import is not inhibited, but rather slightly stimulated by moderate CCCP concentrations (at 50 - 100 µM CCCP: transport rate >100 %), while rather high concentrations (>250 µM CCCP) are required to cause visible reduction of transport when compared to the non-affected transport (set to 100 %). Even in presence of 750 µM CCCP *Pt*NTT5 showed quite high transport activity of 68 % (inhibition by 32 %). Proton-driven NTTs are generally known to be highly inhibited (by 65 - 80 %) already at moderate CCCP concentrations (10 - 100 µM CCCP: transport rates 20 - 35 %) (Ast *et al.* 2009, Haferkamp *et al.* 2006b, Knab *et al.* 2011), thus we can conclude that *Pt*NTT5 does not act as a proton symporter.

In a next step, we investigated the capacity of non-labeled import substrates to induce export of radioactively labeled nucleotides (previously loaded into *E. coli*) by a chase experiment. *E. coli* expressing *Pt*NTT5 were incubated in 50 µM [α^{32}P]-ATP and the radioactivity in the cells was quantified over time. After 5 min of import, significant amounts of [α^{32}P]-ATP accumulated in the cells and possible efflux was started by addition of non-labeled GTP or dGTP. Internal radioactivity markedly decreased, indicating that uptake of non-labeled nucleotides is accompanied by the export of interior label. This observation indicates that *Pt*NTT5 operates in an antiport mode (Figure 6B).

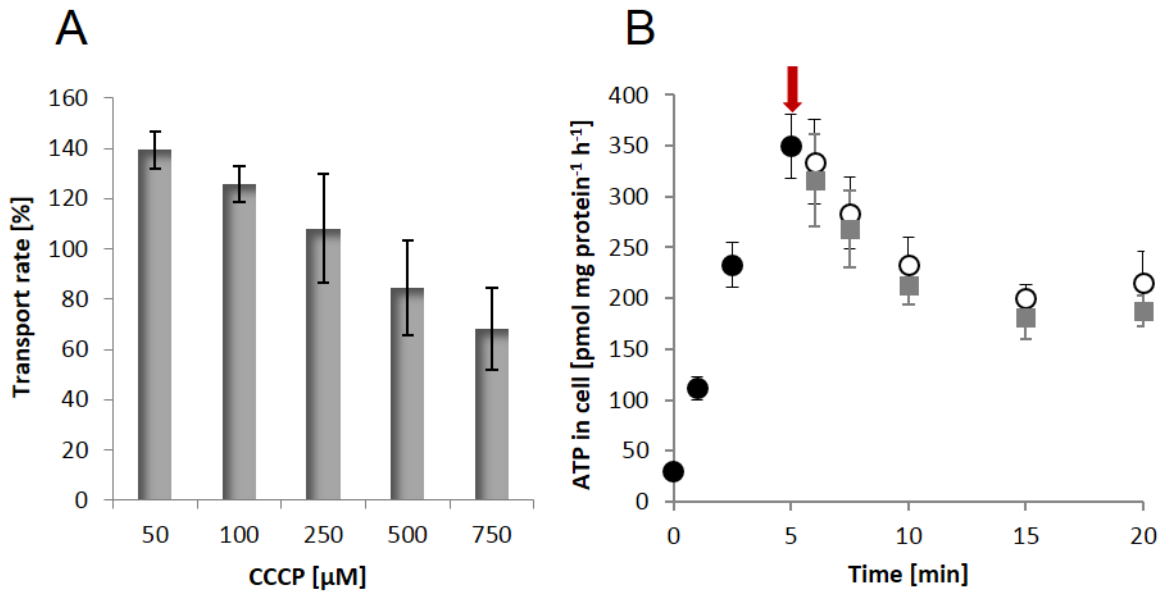


Figure 6**:** Analysis of the transport mode of *Pt*NTT5. (A) Influence of membrane potential depletion by protonophore addition on transport activity of *Pt*NTT5. Non-affected net [α^{32}P]-ATP import (import into induced cells minus rates of non-induced cells) was set to 100%. [α^{32}P]-ATP import in presence of rising CCCP concentrations was calculated as percentage of the non-affected transport. (B) Investigation of a possible counter exchange transport mode. Time dependent accumulation of 50 µM ATP in *E. coli* cells expressing *Pt*NTT5 (black circles). After 5 min of ATP import, possible efflux of internal label was induced by addition of 20 fold excess of non-labeled substrates (indicated by a red arrow). Efflux induction was tested with GTP (white circles) and dGTP (gray squares) as representative counter exchange substrates. Data are the mean of at least three independent experiments, standard errors are given.

2.4.5 GFP fusions suggest targeting of *Pt*NTT5 to the ER membrane

The amino acid sequence of *Pt*NTT5 lacks the N-terminal signal peptide typical for proteins with plastidial destination (stroma/innermost membrane) and also *in silico* analyses failed to clearly predict a specific subcellular targeting. We therefore expressed different GFP-fusion proteins in *P. tricornutum* to deduce the possible *in vivo* localization of the transporter. We fused GFP to either the N- or the C-terminus of *Pt*NTT5. Fluorescence of the N-terminal GFP-fusion labeled a network-like structure distributed throughout the cell, surrounding the plastid and the nucleus (Figure 7A and B). This fluorescence pattern is very characteristic for proteins located in the continuous ER/chloroplast ER (Apt *et al.* 2002). As a reference protein for the ER, we re-investigated a *P. tricornutum* cell line (generated by (Apt *et al.* 2002)), which expresses GFP fused to the presequence of the BiP (ER luminal binding protein precursor; Figure S 4). By staining intracellular DNA with the dye Hoechst 33342 we confirmed that the GFP signal (of the ER) is surrounding the nucleus (Figure 7A), as expected because of the continuity of the ER, nuclear envelope and cER lumen.

The C-terminal GFP-fusion protein does not label a cellular network but instead accumulates in rather diffuse patches close to the plastid/nucleus (Figure 7C). This pattern differs from that of ER targeted proteins (Figure S 4) and does not totally cover or surround the chlorophyll autofluorescence area of the plastid. Moreover, it did not overlap with the MitoTracker signal, which argues against a potential mitochondrial localization.

To investigate the subcellular localization of *Pt*NTT5 in independent systems we transiently expressed *Pt*NTT5 as C-terminal and N-terminal GFP-fusion (Grefen *et al.* 2010) in *Nicotiana benthamiana* and in *Arabidopsis thaliana*. Also in higher plant cells, *Pt*NTT5 with N-terminal GFP is targeted to a network-like structure (Figure S 5) and the GFP signal

overlapped with that of the co-expressed ER-marker protein (ER-rk; Nelson *et al.*, 2007). By contrast, *Pt*NTT5 with C-terminal GFP shows quite heterogeneous and ambiguous fluorescence patterns in both, tobacco and *A. thaliana* (Figure S 6). Diffuse GFP signals and punctuated structures appear in the cytosol and at the nucleus. The fluorescence also surrounds the plastids, however in a quite irregular manner with alternating regions of higher and lower intensity. In contrast to *Pt*NTT5 with N-terminal GFP which labels the entire ER network, the apparent ER localization of the C-terminal GFP-fusion protein is mainly restricted to the nuclear region and is not as constant and precise as observed with the N-terminal GFP construct.

The finding, that N - and C-terminal GFP-constructs cause different fluorescence patterns, led us to the assumption that the position of the marker protein affects subcellular targeting or membrane insertion. In order to identify possible targeting domains and to determine whether the C- or the N-terminal GFP-fusion reflects the localization of the native transporter, we performed targeting studies with truncated versions of *Pt*NTT5 carrying GFP either at the C-terminus or at the N-terminus (Figure S 7-S11).

Microscopic analyses of the respective transformed cell lines revealed that almost all fusion proteins caused GFP signals. Expression of selected GFP-fusions was additionally monitored by western blotting and immuno-staining using an GFP-antibody. This method not only verified the successful expression of the various fusion proteins (including one construct that did not cause visible GFP fluorescence), but also documents their different molecular weights (Figure S 12).

The fluorescence pattern resulting from the truncated versions of *Pt*NTT5 with C-terminal GFP resembled that of the corresponding full-length protein, no matter which part of the protein was deleted (Figure S 8 + Figure S 9). Therefore, no specific region(s) in the sequence of *Pt*NTT5 could be detected that may have caused the accumulation of GFP in proximity to the plastid, suggesting that this localization might be due to the C-terminal position of GFP. In contrast, C-terminally truncated *Pt*NTT5 proteins with N-terminal GFP caused different labeling patterns. Accordingly, the subcellular targeting of these fusion proteins depends on the length of the amino acid sequence of *Pt*NTT5. The GFP signal caused by the shorter versions of *Pt*NTT5 (up to 361 amino acids in length) appears somewhere in the cytosol, whereas longer versions (at least 390 amino acids in length) are targeted to the (chloroplast) ER (Figure S 10 + Figure S 11). Thus, the longer the fragment, the more the labeling pattern resembles that of the corresponding full-length protein.

These results suggest that adding GFP to the C-terminus of *Pt*NTT5 may prevent targeting to the ER by masking an as yet unknown motif or a domain required for (chloroplast) ER targeting. Furthermore, both a (nearly) full-length protein sequence and the unblocked C-terminal part of *Pt*NTT5 might be essential for correct targeting. In this context, it is important to mention that potential targeting information within the protein sequence (particularly the remarkable accumulation of positive charges) at the far C-terminus could not be clearly associated with ER-targeting, because also the correspondingly truncated *Pt*NTT5 versions with N-terminal GFP were detectable in the ER.

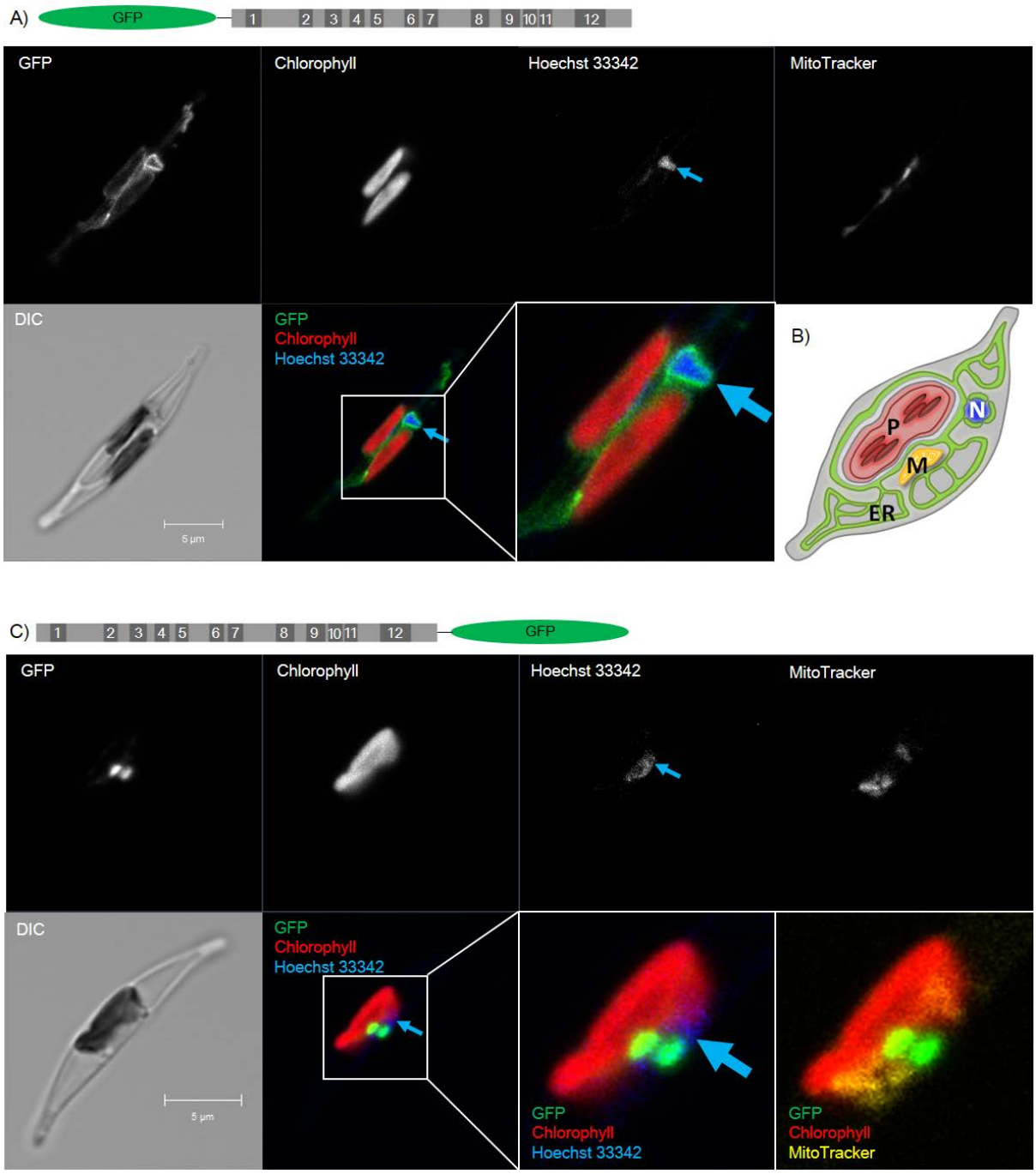


Figure 7: Cellular localization of *Pt*NTT5 in *P. tricornutum* cells. (A) GFP fused N-terminally to *Pt*NTT5. (B) Schematic illustration of a *P. tricornutum* cell depicting the stained organelles. (C) GFP fused C-terminally to *Pt*NTT5. GFP fluorescence in green; autofluorescence of the chlorophyll in red; Hoechst 33342 fluorescence in blue; MitoTracker fluorescence in yellow; Nomarski differential interference contrast (DIC) in grayscale. Position of the Hoechst stain is marked with an arrow. Scale bars: 5 µm. N = nucleus; M = mitochondrion; ER = endoplasmic reticulum; P = plastid.

2.5 Discussion

Diatoms possess the highest number of putative NTTs (up to eight) among all known organisms (Ast *et al.* 2009) which immediately raises questions regarding the phylogeny and physiological benefit of the different isoforms.

In phylogenetic studies, diatom NTT1 isoforms are most closely related to ATP/ADP transporters from red algae (Figure 4; (Ast *et al.* 2009)). Therefore, NTT1 was most likely introduced into the diatom genome via gene transfer from the red algal endosymbiont. The remaining diatom NTTs form separate clusters, more distantly related to the plastidial ATP/ADP transporters from plants and algae (Figure 4). Although the direct ancestor(s) of these NTTs cannot be assigned, their low number of introns and the high GC-content of at least *Pt*NTT5 and *Pt*NTT6 argue against a possible origin from the red algal-derived NTT1 isoform by gene duplication and subsequent modification (Table S 2 + Figure S 1). The origin of almost all diatom NTTs (apart from NTT1 isoforms) by horizontal gene transfer from bacterial sources thus appears more likely. The branching pattern might indicate that the group comprising diatom NTT2 and *Tp*NTT3 and the other group comprising *Pt*NTT3-6 and *Tp*NTT4-8 originated from two different events of horizontal gene transfer followed by gene duplications.

Although showing significant similarities to plastidial ATP/ADP transporters, diatom NTT1 isoforms act as proton-symporters and mediate net uptake of diverse adenine nucleotides across the innermost plastid envelope (Ast *et al.* 2009). Accordingly, the substrate spectrum and the transport mode of the ancestral red algal carrier were modified to fulfill a new function in diatom plastids. Diatom NTT2 isoforms catalyze the antiport of diverse (deoxy-) nucleotides (Ast *et al.* 2009). Interestingly, the functional properties of diatom NTT1 and NTT2 resemble that of NTTs in intracellular bacteria. These bacteria parasitize nucleotides from the host (via interaction of functionally different NTTs) and compensate their missing capacity for nucleotide synthesis (Audia & Winkler 2006, Haferkamp *et al.* 2004, Haferkamp *et al.* 2006b, Knab *et al.* 2011). Similarly, interaction of diatom NTT1 and NTT2 in the innermost plastid envelope was suggested to allow net gain of all substrates for DNA and RNA synthesis in the stroma (Figure 8, (Ast *et al.* 2009)). Adenine nucleotides imported via NTT1 can in turn (in form of ATP) drive the uptake of diverse (deoxy-) nucleotides via NTT2.

This finding was in some way unexpected because it contrasts to the strong expectation that all NTTs from plant and algae exclusively act as ATP/ADP exchangers, with a specific relevance for energy supply to the plastid. However, it is completely in line with the requirement for net nucleotide uptake into the complex diatom plastid. In diatoms, purine and pyrimidine nucleotides are produced exclusively in the cytosol and not, like in higher plants, mainly in the plastid stroma, which necessitates a specifically adapted nucleotide transport system in diatom plastids (Ast *et al.* 2009).

The intriguing question that arose from the previous study (Ast *et al.* 2009) is how nucleotides pass the remaining plastidial membranes. The recruitment of bacterial NTTs with appropriate properties might be a valuable source to establish a transport system in the different plastid membranes that connects nucleotide synthesis in the cytosol with nucleotide consumption in the plastid stroma. A further hypothesis could be that all plastid membranes, but not the innermost one, are selectively permeable for various molecules and allow more or less un-controlled passage of nucleotides. This assumption would imply that not all diatom NTTs are involved in plastidial nucleotide uptake but fulfill diverse physiological functions.

In this study, we performed a detailed analysis of *Pt*NTT5. In view of the low similarities of *Pt*NTT5 to previously described NTTs (Figure S 2), it was unclear whether this protein may actas a nucleotide transporter at all. Import experiments, however, clearly demonstrated its capacity for nucleotide transport. It is specific for tri-, di- and monophosphorylated purine

nucleotides, and dATP and dGTP are among the most favoured substrates (Figure 5, Table 3, Table 4, Figure S 3). Moreover, we discovered that transport activity of *Pt*NTT5 is not dependent upon a proton gradient but on the presence of suitable exchange substrates (Figure 6). Accordingly, *Pt*NTT5 represents a purine nucleotide antiporter.

The determined biochemical features of *Pt*NTT5 are not identical, but closely related to those of certain bacterial and microsporidial NTTs as well as to the NTT1 isoforms of diatoms (Ast *et al.* 2009, Audia & Winkler 2006, Haferkamp *et al.* 2006b, Heinz *et al.* 2014, Knab *et al.* 2011). This implies that not, or not only, the slight modification of the transport properties but rather other factors posed specific evolutionary pressure on *Pt*NTT5 which becomes reflected by the distant relation to bacterial, plastidial or microsporidial NTTs. In this context, it is important to consider that *Pt*NTT5 is apparently located in the ER membrane (Figure 7), and therefore in an environment not typical for NTTs. It is likely that considerable structural adaptations and hence alterations in the amino acid sequence were necessary to allow appropriate targeting and embedding of an NTT into the ER membrane. Similarly, the microsporidial NTTs also exhibit biochemical properties that are identical to or at least highly resemble those of NTTs from bacteria or primary plastids while residing in an NTT-atypical membrane. The amino acid sequences of microsporidial NTTs and *Pt*NTT5 apparently harbor interior targeting information. Although specific motifs or domains cannot be assigned at first glance, the corresponding sequence modifications might hamper the determination of the direct ancestor of these NTTs (Figure 4).

Phylogenetic analyses suggest that plastidial NTTs arose from an ancient chlamydial ATP/ADP transporter by horizontal gene transfer (Figure 4; (Greub & Raoult 2003, Schmitz-Esser *et al.* 2004, Tyra *et al.* 2007). The subsequent adaptation to a new targeting mode of the bacterial-type NTT towards the plastid envelope membrane, including membrane insertion, apparently did not require substantial modifications within the amino acid sequence (Figure S 2). The most striking difference between plastidial and chlamydial ATP/ADP transporters is the presence (addition) of an N-terminal sequence extension that directs the protein to the plastid and that is cleaved off after translocation (Neuhaus *et al.* 1997).

To investigate the subcellular localization of *Pt*NTT5 and, in particularly, to identify possible internal targeting information, we expressed full-length as well as truncated versions of this transporter with GFP either fused N- or C-terminally.

Shorter versions of the N-terminal GFP fusions ended up in the cytosol, whereas longer versions were targeted to the ER membrane (Figure 7A; Figure S 5, Figure S 10, Figure S 11). Therefore, a minimal protein length or even specific information in the C-terminal part of *Pt*NTT5 is required for its targeting to the ER. We therefore conclude that ER targeting of *Pt*NTT5 takes place post-translationally. We suggest that a translocation pathway similar to C-terminally anchored proteins might be used for post-translational insertion into the ER membrane. C-terminal- or tail-anchored (TA) proteins are transmembrane polypeptides that are held in the phospholipid bilayer by hydrophobic amino acids at the C-terminus (Borgese *et al.* 2003). Direct insertion of newly synthesized TA proteins may occur into mitochondrial membranes, but also in the membranes of ER, peroxisomes or the chloroplast (Borgese *et al.* 2007, Borgese & Fasana 2011). *In vitro* analyses revealed the existence of different insertion pathways for TA proteins: the unassisted spontaneous integration into lipid bilayers or the energy-requiring chaperone-mediated pathway (Borgese & Fasana 2011). Import competition assays suggest that the insertion pathway of TA and multi-span proteins may overlap (Dukanovic & Rapaport 2011, Otera *et al.* 2007).

In the case of C-terminal attachment of GFP to full-length or truncated versions of *Pt*NTT5, the GFP signals always appear in diffuse patches close to the plastid/nucleus (Figure 7C,

Figure S 8, Figure S 9). Therefore, this localization is independent of the used part or length of the *Pt*NTT5 sequence. Accordingly, no specific regions in the fusion protein, apart from the C-terminal GFP, could be identified that might be associated to the corresponding localization. The fluorescence pattern of the C-terminal GFP-fusions is indicative for a accumulation rather than for an insertion into a specific membrane. It is conceivable that the diffuse GFP signals are caused by degradation of the hydrophobic *Pt*NTT5 moiety and delivery of GFP. However, western blot analyses indicate that this is not the case (Figure S 12). Alternatively, one might envision that mal-targeted fusion proteins accumulate in membranous structures like in the Golgi apparatus or form aggregates by hydrophobic interaction in an as yet unknown compartment.

Although we cannot rule out completely the possibility that the diffuse patches reflect the correct localization, the GFP analyses in sum suggest that *Pt*NTT5 represents a membrane protein that is post-translationally inserted into the ER.

Activity of the (deoxy-) purine nucleotide shuttle *Pt*NTT5 in the ER membrane (which is continuous with the outermost plastid membrane) would presuppose the presence of suitable exchange substrates in the lumen. Adenine nucleotides are definitely present in the lumen because multiple energy (ATP) dependent reactions (e.g. protein import and folding, ATP hydrolysis via BiP) take place in the ER (Apt *et al.* 2002, Csala *et al.* 2007, Hirschberg *et al.* 1998). It is not clear which proteins catalyze net adenine nucleotide loading of the ER in all eukaryotes, but members of the mitochondrial carrier family are promising candidates (Csala *et al.* 2007, Hirschberg *et al.* 1998, Leroch *et al.* 2008). Independent of the nature of the carriers involved, the adenine nucleotides imported into the ER can drive the antiporter *Pt*NTT5 (Figure 8). When substrates of different phosphorylation states are exchanged (e.g. ATP against ADP) *Pt*NTT5 alters the organellar energy state. Interestingly, in contrast to transporters commonly known to be involved in energy metabolism, the substrate spectrum of *Pt*NTT5 comprises deoxy-nucleotides as additional important substrates. This feature can be considered as an important indicator of an additional and possibly major role of *Pt*NTT5 as a component of the nucleotide uptake system of the complex plastid: import of GTP, dATP and dGTP in exchange for the luminal adenine nucleotides not only alters the nucleotide composition of the ER but also allows the passage of building blocks for DNA and RNA synthesis across the outermost plastid envelope (Figure 8).

These findings indicate that further components of the nucleotide uptake system of diatom plastids await identification. It is unclear which factors mediate (deoxy-) pyrimidine nucleotide entry into the ER/translocation across the outermost plastid envelope and how various (deoxy-) nucleotides pass through membranes two and three (Figure 8). NTT proteins can be considered as important candidates for these tasks and three of the NTT isoforms found in *P. tricornutum* have not been characterized yet. The analysis of NTT isoforms from centric diatoms might additionally help to identify common principles or discrepancies in nucleotide transport of diatoms. In fact, at least slight differences in nucleotide transport can be expected because the centric *T. pseudonana* possesses a different number of putative NTT isoforms from the pennate *P. tricornutum* (Ast *et al.*, 2009). Interestingly, the identification of nucleotide transporters with extraordinary properties was recently shown to pave the way for sophisticated biotechnical approaches. Heterologous expression of *Pt*NTT2 allowed feeding artificial deoxy-nucleotides into *E. coli* and finally resulted in a semi-synthetic organism with an extended genetic alphabet (Malyshev *et al.* 2014).

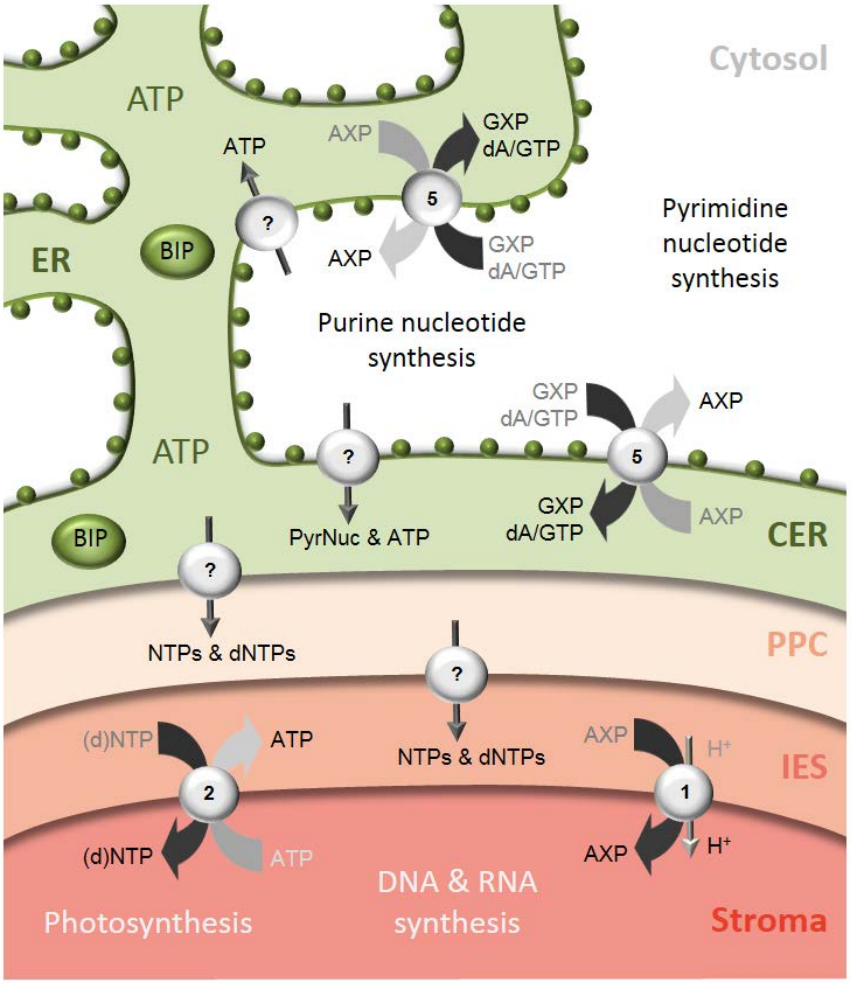


Figure 8: Schematic illustration of the proposed interaction of the biochemically characterized nucleotide transporters (NTTs) from *Phaeodactylum tricornutum*. Interaction of *Pt*NTT1 (grey circle, 1) and *Pt*NTT2 (grey circle, 2) in the innermost plastid membrane provides substrates for DNA and RNA synthesis in the stroma. Pyrimidine and purine nucleotide synthesis takes place in the cytosol. The existence of ATP-dependent reactions, such as activity of the luminal binding protein precursor (BIP) or protein folding, indicates that ATP is present in the ER and chloroplast ER (CER) lumen. However, net uptake of adenine nucleotides into the ER has generally not yet been clarified (gray circle, ?). *Pt*NTT5 (grey circle, 5) can use luminal adenine nucleotides (AXP) to drive the uptake of guanidine nucleotides (GXP), dATP and dGTP from the cytosol. Moreover, the antiport of substrates with different phosphorylation states has an impact on the organellar energy state. Passage of nucleotides (NTPs) and deoxynucleotides (dNTPs) across the remaining plastid membranes and their entry into the periplastidial compartment (PPC) and the inter envelope space (IES) as well as the entry of pyrimidine nucleotides (PyrNuc) into the ER are unclear (gray circles, ?).

Acknowledgments

We thank D. Ballert for help with the transformation and cultivation of *P. tricornutum*. This study was supported by the University of Konstanz, grants of the Deutsche Forschungsgemeinschaft (SFB 969 Project A4 to PGK, Reinhard Koselleck-Grant NE418/16-1 to HEN). We thank the Bioimaging Center (BIC), University of Konstanz, for access to the imaging core facilities.

3 New insights into nucleotide transport of the diatom *Thalassiosira pseudonana*

Lili Chu[1◇], Ansgar Gruber[1◇], Melanie Lorenz[2], Michelle Ast[2], Horst Ekkehard Neuhaus[2], Peter G. Kroth[1] and Ilka Haferkamp[2*]

[1] Pflanzliche Ökophysiologie, Universität Konstanz, 78457 Konstanz, Germany

[2] Pflanzenphysiologie, Technische Universität Kaiserslautern, 67653 Kaiserslautern, Germany

[◇]These authors contributed equally

*Corresponding author: haferk@rhrk.uni-kl.de

Keywords: chloroplast / complex plastids / energy / adenine nucleotide transport

3.1 Abstract

Because of their evolutionary origin by secondary endosymbiosis, diatoms harbour plastids surrounded by four membranes. Accordingly, metabolite transport into and out of these plastids is more complex than in plants with primary plastids. Moreover, in diatoms, main nucleotide synthesis is not located in the plastid, but in the cytosol and a set of **n**ucleo**t**ide **t**ransporters (NTTs, six to eight) with unusual properties is suggested to constitute an uptake system providing nucleotides to the stroma. Plants and algae with primary plastids possess only few NTTs exclusively acting as ATP/ADP exchangers in plastidial energy provision. However, diatom NTT1 and NTT2 isoforms mediate net import of adenine nucleotides or exchange various (deoxy)-nucleoside triphosphates probably across the innermost plastid membrane. NTT5 of the diatom *Phaeodactylum tricornutum* (*Pt*NTT5) acts as a shuttle for various adenine and guanosine nucleotides in the outermost membrane of the plastid (chloroplast ER membrane) and might modify the nucleotide composition in the ER. Here, we characterised a fourth diatom NTT isoform, *Tp*NTT3 from *Thalassiosira pseudonana*. *Tp*NTT3 transports an unexpectedly broad spectrum of substrates, namely various purine nucleotides, including cyclic mononucleotides and deoxy forms. These transport activities differ from all previously described activities of NTTs, which so far have not been known to transport cyclic mononucleotides. GFP-tagged *Tp*NTT3 accumulates in so-called clamp-like structures at the plastid. This fluorescence pattern was previously observed for proteins operating in silica-based cell wall formation and proposed to mark a sub-compartment of the ER involved in the pathway of silica deposition vesicle biogenesis. The current data suggest that *Tp*NTT3 is possibly not a further component of the plastidial nucleotide uptake system but might rather play a role in silicon metabolism. It has the capacity to provide energy, e.g. for phosphorylation of dominating biosilica proteins, and to alter the purine and cyclic nucleotide composition in the compartment.

3.2 Introduction

Glaucophytes, rhodophytes, and chlorophytes - the three main lineages of algae with primary plastids - originated from primary endosymbiosis (Keeling 2013), which is the engulfment of a cyanobacterium by a heterotrophic host cell (Keeling 2013). The endosymbiont was retained and partially degraded by the host. Furthermore, massive transfer of cyanobacterial genes to the host nucleus as well as metabolic connection and rearrangement of the consortium resulted in control and functional embedding of the symbiont and finally in the establishment of the photosynthetic organelle (Archibald 2015). In contrast, secondary endosymbiosis apparently occurred several times and involves the fusion of two eukaryotes, a heterotrophic host and an alga with primary plastids (Archibald 2015, Keeling 2013). This led to a considerable number of diverse algal lineages, most of them possess red-alga derived secondary plastids (heterokonts, cryptomonads, haptophytes, apicomplexa and peridin-containing dinoflagellates) and only few secondary plastids are apparently of green algal origin (those of euglenophytes or chlorarachniophytes) (Delwiche 1999). Secondary plastids - unlike primary plastids - are surrounded by three to four membranes, a probable consequence of their evolutionary history (Cavalier-Smith 2003). Similar to primary plastids, a residual genome (derived from the original cyanobacterial genome) is detectable in the stroma of secondary plastids whereas the nucleus of the eukaryotic symbiont is generally completely degraded. Notably, cryptomonads and chlorarachniophytes retain a remnant algal nucleus (nucleomorph) between membrane two and three of their complex plastids

(Archibald 2007, Gilson *et al.* 2006). This location is called the periplastidial space and represents the former cytosol of the eukaryotic endosymbiont (Gould *et al.* 2006, Gould *et al.* 2008, Sommer *et al.* 2007).

Diatoms became of particular scientific interest because of their high abundance in the oceans; they are among the most common types of the phytoplankton and immensely contribute to carbon fixation (Armbrust 2009, Falkowski & Oliver 2007). As a subgroup of heterokont algae diatoms harbour a secondary red-algae derived plastid surrounded by four membranes with the outermost connected to the ER (Gould *et al.* 2008, Kroth 2002). Genome sequencing and physiological studies revealed that pyrimidine and purine nucleotide *de novo* synthesis is apparently cytosolic, whereas in higher plants most steps of these pathways occur in the plastid stroma (Ast *et al.* 2009). Accordingly, diatom plastids rely on nucleotide uptake from the cytosol to fuel DNA and RNA synthesis in the stroma. First studies suggest that NTT-type transporters are involved in this process (Ast *et al.* 2009, Chu *et al.* 2016b).

Generally, NTT proteins are restricted to only few organismic groups. They are a hallmark feature of obligate intracellular living bacteria of the orders *Chlamydiales* and *Rickettsiales* (Schmitz-Esser *et al.* 2004). These bacteria lack several metabolic pathways, including *de novo* nucleotide synthesis. Bacterial NTT1 isoforms act as ATP/ADP transporters in energy parasitism, whereas other NTT isoforms (NTT2 to maximally 5) function as proton-symporters or nucleotide exchangers, allow to exploit the nucleotide pool of the host and supply substrates for DNA and RNA synthesis as well as NAD to the metabolically impaired bacterium (Audia & Winkler 2006, Haferkamp *et al.* 2004, Knab *et al.* 2011, Krause *et al.* 1985, Tjaden 1999). Among eukaryotes, NTTs were detected in algae and higher plants as well as in protists of the phylum Microsporidia, whereas fungi and animals completely lack comparable sequences. In Microsporidia, NTTs mediate the transport of adenine and guanosine nucleoside tri- and diphosphates and reside in the plasma membrane or in the mitosome - a mitochondrial relict unable for energy production (Heinz *et al.* 2014, Tsaousis *et al.* 2008). In higher plants and algae with primary plastids NTTs were shown to operate as ATP/ADP exchangers in the inner plastid envelope and to provide energy to the stroma under conditions, where photosynthesis is insufficient or missing (Geigenberger *et al.* 2001, Linka *et al.* 2003, Möhlmann *et al.* 1998, Reinhold *et al.* 2007, Reiser *et al.* 2004, Tjaden *et al.* 1998b)

Diatoms possess an unusually high number of putative NTTs (six to eight) when compared to algae or higher plants with primary plastids (one to maximally three) (Ast *et al.* 2009, Linka *et al.* 2003, Möhlmann *et al.* 1998, Reiser *et al.* 2004). Recent analyses of the first isoforms suggest that the NTTs of diatoms are not or not primarily involved in energy translocation but rather mediate nucleotide provision from the cytosol to the stroma (Ast *et al.* 2009).

Phylogenetic studies suggest that the gene encoding the diatom NTT1 isoform was introduced by the secondary endosymbiont and thus traces back to the red algal ATP/ADP transporter. However, NTT1 isoform acts as a proton-symporter and exhibits a slightly broader substrate spectrum of adenine nucleotides. The remaining diatom NTTs form two separate clusters with distant relation to functionally diverse NTTs from *Rickettsiales* and the purine nucleotide transporters from Microsporidia. It is hard to judge whether they evolved by gene duplication of diatom NTT1 that was considerably modified or whether they were introduced via horizontal gene transfer of a bacterial or microsporidial NTT (Ast *et al.* 2009, Chu *et al.* 2016b). Diatom NTT2 was shown to transport various triphosphorylated nucleotides including corresponding deoxy forms in a counter exchange mode. GFP-based localisation studies suggest that diatom NTT1 and most likely also NTT2 reside in the innermost envelope of the complex diatom plastid (Ast *et al.* 2009). Moreover, physiological interaction of NTT1

and NTT2 was proposed to deliver all nucleotides for DNA and RNA synthesis to the stroma (adenine nucleotides provided by NTT1 can in form of ATP drive the import of other (deoxy) nucleotides via NTT2). NTT5 of the diatom *Phaeodactylum tricornutum* (*Pt*NTT5) is suggested to reside in the ER which also forms the outermost plastid envelope and to catalyse the antiport of various purine nucleotides, including their deoxy forms (Chu *et al.* 2016b). Therefore, *Pt*NTT5 has the capacity to alter the composition and to broaden the spectrum of nucleotides in ER and the chloroplast ER (compartment between the two outermost plastid envelopes).
Although recent studies provided important insights into nucleotide transport of diatoms, the functional properties and subcellular localization of many NTT isoforms are still unknown. In this study, we focus on the NTT isoform 3 from the diatom *T. pseudonana* (*Tp*NTT3). This protein is closely related to diatom NTT2 isoforms and interestingly a direct homolog was not identified in *P. tricornutum*. Targeting studies and investigation of its biochemical properties allowed us to propose a possible physiological role of the NTT3 isoform from *T. pseudo*nana.

3.3 Materials and Methods

3.3.1 Culture conditions and transformation of diatoms

Thalassiosira pseudonana (Hustedt) Hasle et Heimdal clone CCMP1335 (National Center for Marine Algae and Microbiota (NCMA); Maine, USA) and *Phaeodactylum tricornutum* Bohlin 1897, strain 646 (UTEX Culture Collection of Algae, University of Texas, Austin, TX, USA) were cultivated in f/2 medium at a concentration of 50% sea water (16.6 g of sea salt; Tropic Marin Dr. Biener, GmbH, Wartenberg, Germany) in 1 L of distilled water, adjusted to pH 7.0. Supplements were added as described in Kroth (2007). For the cultivation of *T. pseudonana* medium was additionally supplemented with 5 g $Na_2SiO_3 \times 5\ H_2O$ and 150 µg Na_2SeO_3 per 1 litre f/2 medium. The cells were cultivated under alternating light conditions (16 h light/ 8 h dark) at 20 °C.

3.3.2 Preparation of cDNA and PCR

Poly(A^+)-RNA was isolated from *T. pseudonana* by the use of the RNeasy plant mini kit and the Oligotex kit (Qiagen, Hilden, Germany). RNA isolation and reverse transcription with SuperscriptII (Invitrogen, Carlsbad, USA) were performed according to supplier's instructions. PCR was conducted with oligonucleotides (MWG, Ebersberg, Germany or Sigma Aldrich, München, Germany) allowing compatible insertion into the expression vectors (Table S 4). Amplification products were gel-purified by using the NucleoSpin Extract II Kit (Macherey & Nagel, Düren, Germany) and inserted into the vector pBSK (Stratagene, Heidelberg, Germany) or directly into the expression and transformation vectors.

3.3.3 Generation of constructs for expression in *E. coli* or diatoms

For generation of the *E. coli* expression constructs, the coding sequence of *Tp*NTT3 (JGI protein ID 270249, (Ast *et al.* 2009)) was inserted in frame with the Histidine tag into the

IPTG-inducible expression vector pET16b (Novagen, Heidelberg, Germany). Presequence and full length sequence of *Tp*NTT3 were inserted into the *P. tricornutum* transformation vector GFP-pPha-T1 (Gruber 2008, Gruber *et al.* 2007), derived from the shuttle vector pPha-T1 (GenBank AF219942 (Zaslavskaia *et al.* 2000)), via *Stu*I and *Eco*RV restriction sites. For the generation of *T. pseudonana* transformation vectors, presequence and full length sequence of *Tp*NTT3 were inserted into the pTpfcp-GFP (Poulsen *et al.* 2006). Correctness of the constructs was verified by restriction analyses (Fermentas, St. Leon-Rot, Germany) and sequencing (NBC, TU Kaiserslautern or GATC Biotech AG, Konstanz, Germany).

3.3.4 Heterologous expression in *E. coli* and import measurements

Heterologous synthesis of *Tp*NTT3 in the *E. coli* strain BLR (DE3) and transport measurements were conducted according to previously reported methods (Ast *et al.* 2009, Chu *et al.* 2016b). To analyse transport properties of the recombinant *Tp*NTT3 either induced or non-induced (control) *E. coli* cells harbouring the corresponding plasmids, were incubated in phosphate buffer (KPi) containing 50 µM or the indicated concentrations of [α^{32}P] labelled substrates. Uptake was allowed at 30°C for the indicated time spans and terminated by removal of external substrate using vacuum filtration and washing (Haferkamp *et al.* 2006b) Radioactivity in the samples was quantified in a scintillation counter (Tricarb 2500; Canberra-Packard, Heidelberg, Germany).

3.3.5 Back-exchange analysis and thin layer chromatography

E. coli cells expressing *Tp*NTT3 were loaded with [α^{32}P] ATP for 5 min. Due to the metabolic activity of *E. coli* cells [α^{32}P] ATP is converted into labelled ADP, AMP and Pi. Non-imported radioactivity was removed by washing and radioactivity in the cells was quantified. Subsequently, cells were incubated in KPi (control) or in KPi containing 500 µM of the given import substrates. Back-exchange was carried out at 30°C for 2 min and terminated by rapid centrifugation. 10 µl of the supernatant were loaded onto a 0.5 mm poly (ethylene amine) cellulose thin layer chromatography plate and dried with a fan. Subsequent to the chromatography and radioautography, the nature of exported label was identified by comparison with radioactively labelled standards. Radioactively labelled positions were marked on the thin layer plate (Haferkamp *et al.* 2006b), cut out and quantified in a scintillation counter.

3.3.6 Nuclear transformation and microscopy

Nuclear transformation of *T. pseudonana* and *P. tricornutum* was performed as described previously (Kroth 2007b, Poulsen *et al.* 2006). Cellular localisation of GFP fusion proteins was analysed with a confocal laser scanning microscope LSM 510 META (Carl Zeiss MicroImaging GmbH, Göttingen, Germany) using a Plan-Apochromat 63 × 1.4 oil immersion Nomarski differential interference contrast (DIC) objective (Carl Zeiss).

3.4 Results

3.4.1 The transport substrates of *Tp*NTT3

Import measurements with radioactively labelled nucleotides on intact *E. coli* cells expressing the recombinant protein were conducted to experimentally clarify the biochemical features of *Tp*NTT3. Initial transport measurements with [α^{32}P]ATP revealed that *Tp*NTT3 mediates significant time-dependent uptake of ATP into *E.coli* (~ 300 pmol/mg protein in 30 min), whereas non-induced (control) cells showed no or only marginal accumulation of radioactivity (< 40 pmol/mg protein in 30 min) (Figure 9).
To investigate whether ATP is the only substrate of *Tp*NTT3, we conducted [α^{32}P]ATP import measurements in presence of 10-fold excess of various non-labelled nucleotides or nucleotide derivates. [α^{32}P]ATP import via *Tp*NTT3 was markedly reduced by a high number of added nucleotides (Table 4).

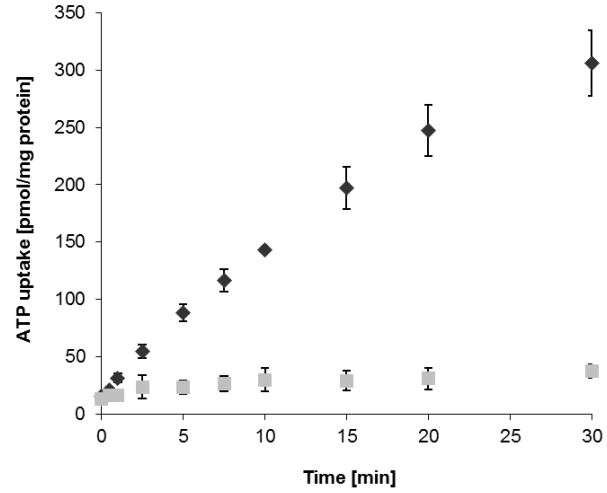


Figure 9: Time dependent uptake of [α^{32}P]ATP by *Tp*NTT3. ATP import (50 µM) mediated by *E.coli* cells expressing the recombinant *Tp*NTT3 (black rhombs) or by non-induced cells (grey squares). Data are the mean of at least three independent experiments, standard errors are given.

Table 4: Effects of various metabolites on [α^{32}P]-ATP uptake by *Tp*NTT3. Uptake of [α^{32}P]ATP by recombinant *Tp*NTT3 was measured at a substrate concentration of 50 µM, and non-labelled effectors were present in 10-fold excess. Import was stopped after 10 min. Rates of nucleotide uptake are net values (minus control: non-induced *E. coli* cells) given as percentage of the non-affected transport (set to 100%). Data are the mean of three independent experiments. Bold type is used to indicate significantly reduced import rates (< 50% residual activity).

Effector	**Import [%]**	**SE [%]**	**Effector**	**Import [%]**	**SE [%]**
None	100.0	/	UMP	59.5	+/- 3.7
ATP	**7.4**	+/- 1.1	**ITP**	**4.4**	+/- 1.2
GTP	**7.1**	+/- 2.0	**IDP**	**23.4**	+/- 1.6
UTP	64.0	+/- 2.0	**IMP**	**6.8**	+/- 1.2
CTP	93.3	+/- 1.2	**dITP**	**37.1**	+/- 0.3
dATP	**16.8**	+/- 5.6	**dAMP**	**46.6**	+/- 6.4
dGTP	**13.4**	+/- 2.1	**dGMP**	**28.6**	+/- 2.6
dCTP	81.1	+/- 6.2	**dIMP**	**36.9**	+/- 2.8
TTP	90.4	+/- 4.7	**cAMP 3,5**	**12.2**	+/- 2.1
ADP	**13.2**	+/- 1.4	**cGMP 3,5**	**18.1**	+/- 3.9
GDP	**13.7**	+/- 1.9	**cIMP 3,5**	**11.3**	+/- 1.6
AMP	**4.8**	+/- 2.5	cAMP 2,3	111.8	+/- 5.9
GMP	**16.6**	+/- 0.9	ADP-Glucose	80.0	+/- 3.1
dADP	67.2	+/- 5.4	UDP-Glucose	95.8	+/- 3.0
Adenosine	100.6	+/- 3.6	NAD	78.1	+/- 4.9
Guanosine	66.8	+/- 4.6	NADP	102.2	+/- 6.3
UDP	93.0	+/- 2.6			

Generally, purine nucleotides, including inosine nucleotides, mono- and triphosphorylated (deoxy)nucleosides and cyclic mononucleotides, caused a significant decrease in [α^{32}P]ATP transport, whereas pyrimidine nucleotides or the representative deoxynucleoside diphosphate dADP showed no or comparably low effects (~ 60 % or more residual transport activity). Moreover, deoxymononucleotides and dITP were not as efficient (~ 30 to 47 % residual activity) as the majority of purine nucleotides (~ 20 % or lower residual activity). It is important to mention that only cyclic 3',5'-AMP and cyclic 3',5'-GMP (that act in cellular signalling and protein regulation) led to transport reduction, whereas presence of cyclic 2',3'-AMP (product of mRNA degradation) did not affect [α^{32}P]ATP transport.
The observed transport reduction might be caused by inhibitory properties of the applied molecule or could be due to its competition during binding or translocation of [α^{32}P]ATP.
We performed import studies with selected radiolabelled purine nucleotides to investigate whether they represent substrates or inhibitors of *Tp*NTT3 (Figure 10). All nucleotides tested were imported. The rates of ADP, AMP, dAMP, dATP and dGTP uptake (~ 100 to 140 pmol mg protein^{-1}) were in the range of the ATP import (~ 115 pmol mg protein^{-1}) and transport of GTP, GDP and cAMP was slightly higher (~ 200 pmol mg protein^{-1}). GMP, dGMP (~ 80 pmol mg protein^{-1}) and finally cGMP (~ 30 pmol mg protein^{-1}) were transported with lowest rates. Therefore, we can conclude that *Tp*NTT3 accepts an extraordinarily wide range of purine nucleotides as substrates. We did not investigate transport with radiolabelled inosine nucleotides because most are not commercially available. However, all other nucleotides investigated were identified as substrates and therefore it seems likely that also the structurally related inosine nucleotides are accepted by *Tp*NTT3.
We also determined the maximal velocities (V_{max}-values) and affinities (K_M-values) of *Tp*NTT3 for the different substrates (Table 5). The amount of functional, recombinant protein in the *E.coli* membrane influences the V_{max} (the more protein, the higher the V_{max}) whereas K_M-values are independent of this factor. Therefore, K_M- and V_{max}-values of different substrates of one transporter can be compared. Comparisons of the K_M-values of different heterologously expressed NTTs is also possible but not comparison of their V_{max}-values. For determination of K_M-values and V_{max}-values we conducted import studies with rising concentrations of the labelled nucleotides, respectively. *Tp*NTT3 exhibits comparatively high affinities for ATP, ADP, AMP, GTP, GDP and cAMP (from ~ 20 to 57 µM) whereas dGTP, GMP and dATP are transported with intermediary (from ~ 80 to 140 µM) and dGMP and dAMP with rather low affinity (200 and 330 µM). Most substrates are transported with a maximal velocities ranging from 1.2 to 2.0 nmol mg protein^{-1} h^{-1}. However, dGMP, dGTP, cAMP and particularly dATP show higher V_{max}-values (2.5 to 3.8 and 6.8 nmol mg protein^{-1} h^{-1}). It should be noted that due to the low transport rate reliable determination of the transport parameters for cGMP was hampered.
The previous results suggest that *Tp*NTT3 generally prefers a broad spectrum of purine nucleotides as substrates and among the transported purine nucleotides it seems to be not highly selective. Therefore, *Tp*NTT3 immensely differs not only from the diatom isoforms NTT1 and NTT2 and *Pt*NTT5, but also from all plastidial or bacterial NTT-type carriers described so far.

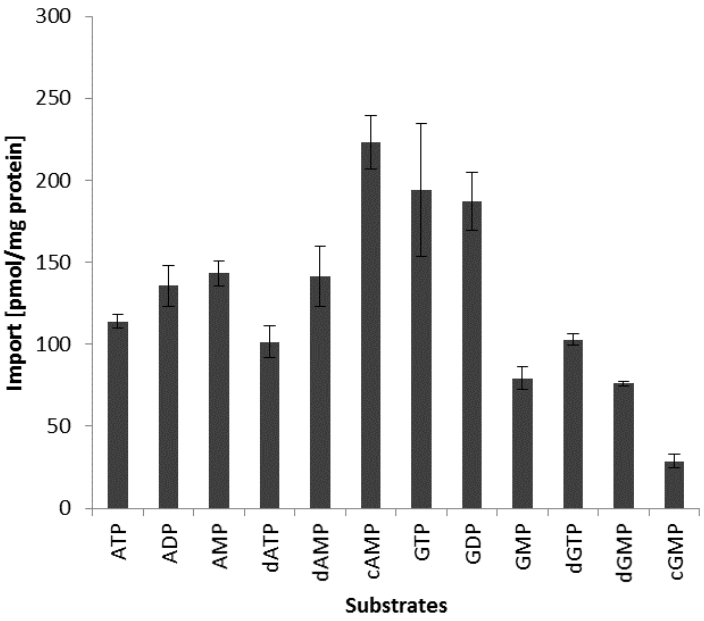


Figure 10: Import of radioactively labeled nucleotides via recombinant *Tp*NTT3. *E coli* expressing the recombinant *Tp*NTT3 and non-induced cells were incubated for 10 min in 50 µM of the given substrates. The displayed transport represents the net uptake via *Tp*NTT3 and was calculated by subtraction of the corresponding value of the control. Data are the mean of at least three independent experiments, standard errors are given.

Table 5: K_M and V_{max} values of nucleotide transport by recombinant *Tp*NTT3. Nucleotide uptake in the presence of rising concentrations of substrate (5 µM – 1000 µM) was allowed for time spans in the linear phase of corresponding transport at 50 µM. K_M values are given in µM, V_{max} values are given in nmol mg protein^{-1} h^{-1}. Data are the mean of at least three independent experiments. Corresponding standard errors are given in brackets.

Substrate	**K_M**	**V_{max}**
ATP	19.6 (+/- 2.7)	1.96 (+/- 0.29)
ADP	36.9 (+/- 6.5)	1.40 (+/- 0.15)
AMP	28.4 (+/- 2.3)	1.22 (+/- 0.06)
dATP	142.7 (+/- 8.6)	6.78 (+/- 0.77)
dAMP	332.5 (+/- 18.1)	1.89 (+/- 0.15)
cAMP	56.8 (+/- 2.4)	3.34 (+/- 0.18)
GTP	22.4 (+/- 1.7)	1.77 (+/- 0.17)
GDP	20.7 (+/- 1.9)	1.62 (+/- 0.14)
GMP	108.0 (+/- 7.1)	1.68 (+/- 0.15)
dGTP	82.7 (+/- 10.7)	3.79 (+/- 0.67)
dGMP	208.8 (+/- 16.8)	2.51 (+/- 0.16)

3.4.2 The transport mode of *Tp*NTT3

To propose the metabolic role of NTTs in the cellular context it is important to determine their transport mode. Proton-driven import allows net provision of substrates, whereas (1:1 stoichiometric) counter exchange of nucleotides does not result in gain of substrates. Most characterised NTTs that act in a counter exchange mode are ATP/ADP transporters. NTTs which are not specific for exclusively ATP and ADP either act as proton-driven nucleotide importers in net uptake or as antiporters balancing the cellular/organellar nucleotide pool. Moreover, antiporters that are not involved in energy provision (non-ATP/ADP transporters), generally interact with one or more proton-driven net importer(s) to allow net provision of different nucleotides to the organelle (or the parasitizing bacterium). To clarify whether recombinant *Tp*NTT3 is driven by a proton-gradient across the *E. coli* membrane, we investigated the influence of the protonophore CCCP on [α^{32}P]-ATP import. Transport was slightly reduced by moderate concentrations of CCCP (10 to 50 µM, > 80% residual activity) and also in presence of high CCCP concentrations (100 to 500 µM) a significant residual transport activity (> 60%) was still measurable (Figure 11A). Proton-driven NTTs are known to be significantly inhibited (~ 20-35% residual activity) by moderate CCCP concentrations and thus we can conclude that activity of *Tp*NTT3 is not strictly dependent upon the proton-gradient. By the help of a so called back exchange experiment, we investigated the capacity of non-labelled import substrates to induce export of radioactively labelled nucleotides (previously loaded into *E. coli* cells). *E. coli* cells expressing *Tp*NTT3 were incubated with [α^{32}P]-ATP for 5 min, subsequently non-labelled substrates were added in excess and efflux was documented (Figure 11B). The buffer medium already caused significant release of phosphate and AMP from the cells known to result from endogenous phosphate and adenosine export systems of *E. coli* (Ast *et al.* 2009). Moreover, also traces of ATP and ADP were detected. Addition of ATP, ADP or AMP significantly stimulated the efflux of radioactively labelled adenine nucleotides (Figure 11B). This analyses verifies that *Tp*NTT3 operates in an antiport mode.

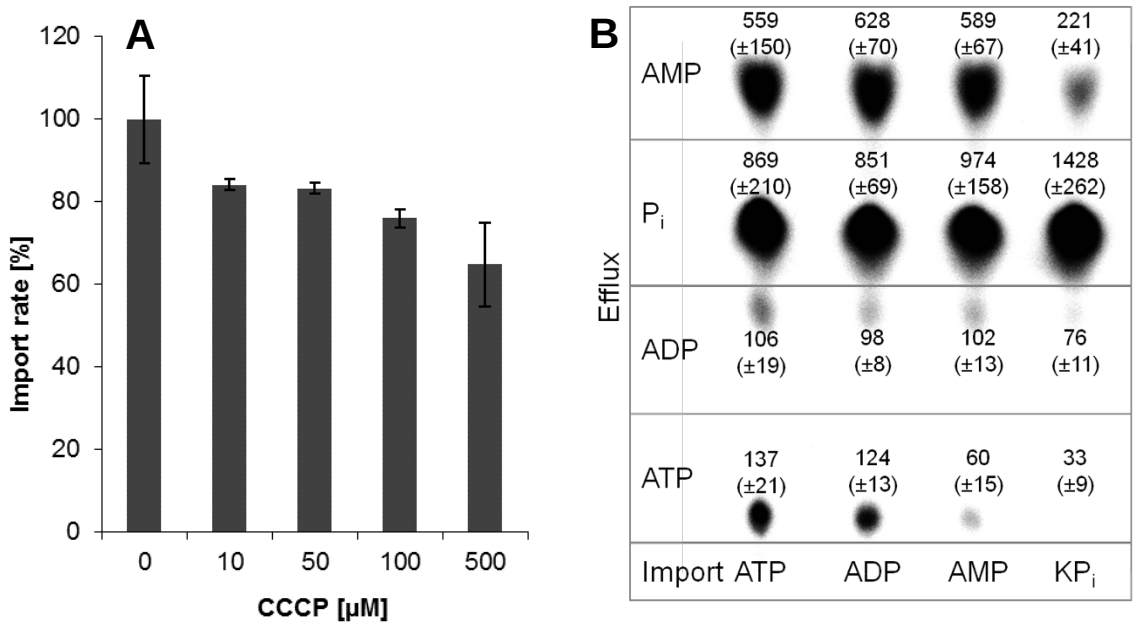


Figure 11: Analysis of the transport mode of *Tp*NTT3. (A) Effect of different concentrations of the protonophore CCCP on [α^{32}P] ATP uptake (50 µM] via recombinant *Tp*NTT3. Rates of ATP import are given as percentage of the control (non-affected transport = 100%). Standard errors are indicated. (B) Exchange-mediated efflux of intracellular radioactivity mediated by *Tp*NTT3. Subsequent to the thin layer chromatography and radioautography, the nature of exported label was identified by comparison with radioactively labelled standards. Radioactively labelled positions were marked on the thin layer plate. A representative thin layer chromatography result is given. Values are the mean of three independent experiments. Standard errors are indicated.

3.4.3 The subcellular localisation of *Tp*NTT3

In primary plastids, NTTs catalyse the counter exchange of ATP and ADP across the inner envelope membrane. Although exhibiting different biochemical features the diatom carriers NTT1 and NTT2 are also expected to reside in a plastidial membrane. The presence of a signal peptide with a bipartite presequence in *Tp*NTT3 is indicative for a plastidial localisation. Subcellular targeting of *Tp*NTT3 was analysed by the help of GFP-fusion constructs (full-length transporter, bipartite presequence) expressed in the diatom *T. pseudonana* as well as in the heterologous *P. tricornutum*. In *T. pseudonana* both, the chosen leader sequence and the full length protein, led to an accumulation of GFP in clamp-like and rarely in horseshoe-shaped structures closely associated to the plastid (Figure 12). The clamp-like green fluorescence pattern at the waistline of the plastid was also observed for GFP-fusions of a silaffin kinase proposed to be located in a specific plastid-associated subregion of the ER (Poulsen *et al.* 2013, Sheppard *et al.* 2009).

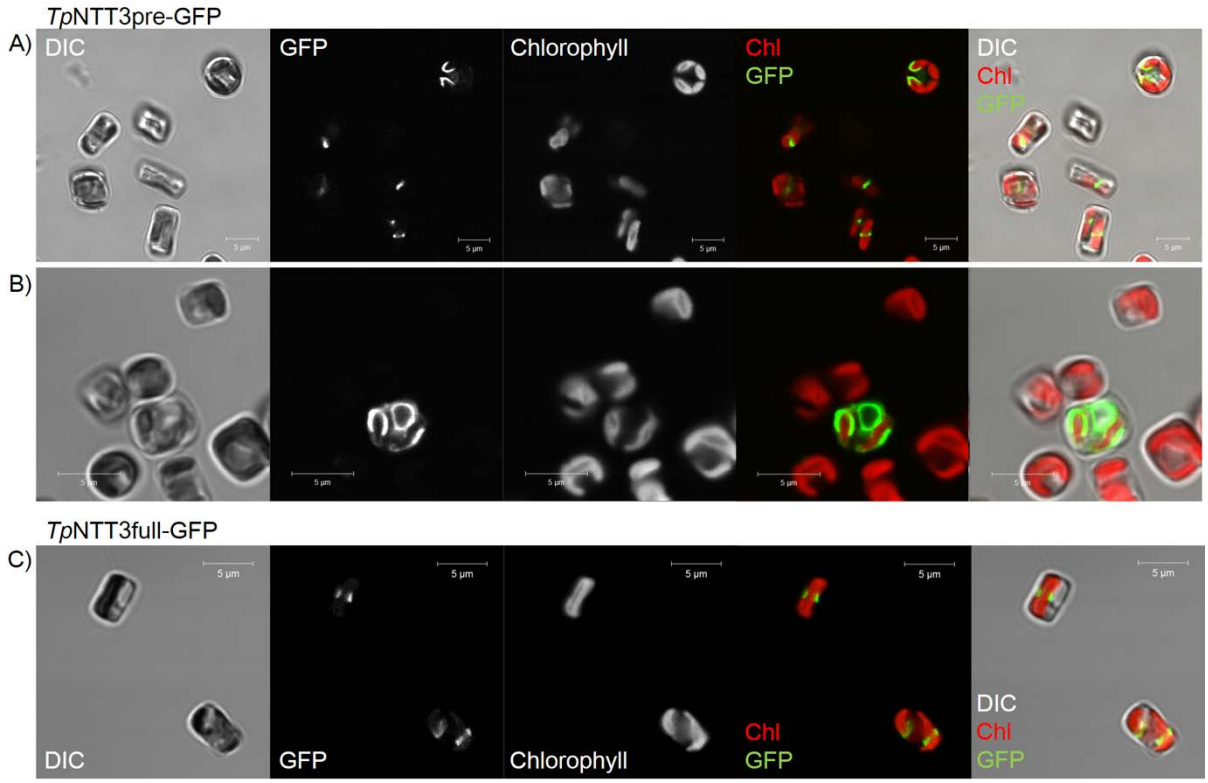


Figure 12: Cellular localisation of NTT3 in *T. pseudonana* cells. GFP fused C-terminally to *Tp*NTT3 presequence (A, B) and full length (C). GFP fluorescence in green; autofluorescence of the chlorophyll in red; Nomarski differential interference contrast (DIC) in grey scale. Scale bars: 5 µm

Also when heterologously expressed in *P. tricornutum*, the leader sequence and the full-length construct accumulated at the waistline of the plastid (Figure 13). The GFP fluorescence pattern in *P. tricornutum* resembles that of the 'blob'-like structures often observed for soluble GFP-fusion proteins targeted to the periplastidial space but appeared more diffuse with regions of higher and lower intensity. Interestingly, the full-length protein additionally labelled small dots which in rare cases were distributed throughout the cell. This punctual GFP-pattern might suggest an accumulation of GFP in small vesicular structures, maybe derived from the ER.

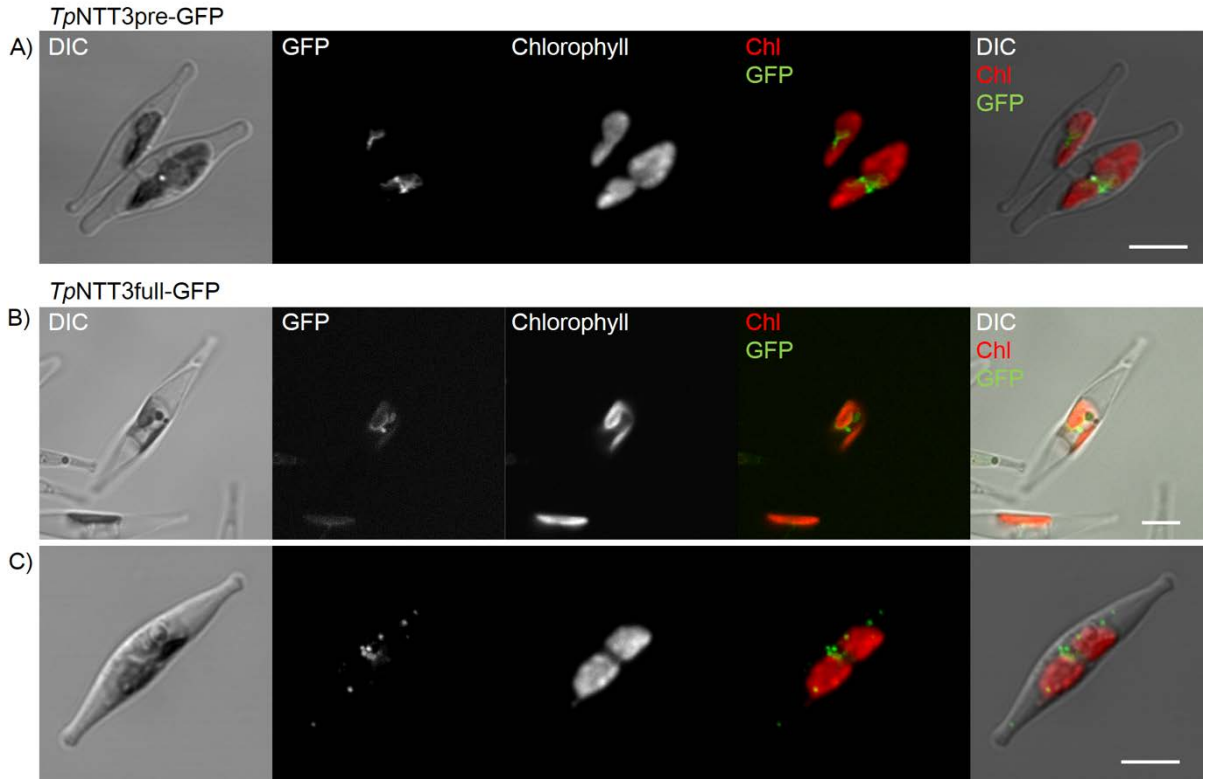


Figure 13: Cellular localisation of *Tp*NTT3 in *P. tricornutum* cells. GFP fused C-terminally to *Tp*NTT3 presequence (A) and full length (B, C). GFP fluorescence in green; autofluorescence of the chlorophyll in red; Nomarski differential interference contrast (DIC) in grey scale. Scale bars: 5 µm

3.5 Discussion

NTT-isoforms from plants with primary plastids (*Arabidopsis thaliana*, potato and also red algae) are quite well characterised (Linka *et al.* 2003, Reiser *et al.* 2004, Tjaden *et al.* 1998a). They were shown to exclusively catalyse a highly specific exchange of ATP and ADP across the inner plastid envelope. Investigations of mutant plants demonstrated that their NTTs play an essential role in energy provision to the plastid (Reiser *et al.* 2004, Trentmann *et al.* 2008). When compared to primary plastids of plants, diatom plastids are surrounded by two additional membranes. Moreover, diatoms encode six (*P. tricornutum*) to eight (*T. pseudonana*) putative NTTs in their genomes (Ast *et al.* 2009). The corresponding sets of diatom NTTs exhibit different versions of N-terminal extensions (no, short, bipartite targeting domain harbouring or lacking typical ASAFAP-motifs) (Ast *et al.* 2009). Therefore, it was primarily assumed that NTTs from diatoms might reside in the different plastidial membranes, where they mediate energy passage to the plastid stroma. However, characterisation of the first two NTT-isoforms from *P. tricornutum* (*Pt*NTT1 and *Pt*NTT2) and of the corresponding homologs from *T. pseudomana*, revealed that diatom NTTs functionally differ from all known plant NTTs (Ast *et al.* 2009). Diatom NTT1 proteins catalyse a proton-driven net uptake of adenine nucleotides and NTT2 isoforms transport (deoxy-)nucleoside triphosphates in a counter exchange mode. NTT1 isoforms are targeted to the innermost membrane and also for NTT2 proteins a similar localisation is proposed. In terms of their biochemical properties diatom NTT1 and NTT2 resemble bacterial NTT proteins that exploit nucleotides from the host cell and compensate missing nucleotide

biosynthesis of the metabolically impaired endosymbiont/parasite. In diatoms, net nucleotide uptake into the plastid is of particular importance because nucleotides are generated in the cytosol but required in the plastid, for e.g. DNA and RNA synthesis (Ast *et al.* 2009). The biochemical properties of diatom NTT1 and NTT2-isoforms, their plastidial localisation, and the requirement for nucleotide provision to the stroma strikingly argue for a role of these carriers in plastidial nucleotide uptake. NTT5 from *P. tricornutum* showed a different substrate spectrum comprising all adenine and guanine nucleotides, including dATP and dGTP (Ast *et al.* 2009, Chu *et al.* 2016b). Furthermore its plastidial localisation in the outermost membrane, which is connected with the ER, confirmed that this membrane represents a selective barrier for nucleotides.

We hypothesised that also the remaining NTT isoforms from diatoms might be involved in plastidial nucleotide import rather than in energy passage (as ATP/ADP exchangers). This is because all so far characterised diatom NTTs do not represent ATP/ADP transporters. Moreover, it is important to keep in mind that NTT1 isoforms represent the only NTTs from diatoms with considerable amino acid similarities to typical ATP/ADP transporters from primary plastids, however, they clearly catalyse a net uptake of adenine nucleotides (Ast *et al.* 2009). Diatom NTT2 isoforms and also NTT3 from *T. pseudonana* show at least moderate relation to various functionally different NTTs, whereas the remaining isoforms exhibit weakest similarities to all so far known NTTs (Ast *et al.* 2009).

In this study, we characterised *Tp*NTT3 and identified that this carrier also functionally differs from typical plant NTTs. This isoform exhibits an extraordinarily wide substrate spectrum previously not observed for any other NTT. Apart from adenine nucleotides, also guanosine nucleotides and probably even inosine nucleotides represent substrates, whereas pyrimidine nucleotides apparently are not transported. Among the guanosine and adenine nucleotides, the carrier accepts mono-, di- and triphosphorylated nucleosides, deoxy forms, as well as cyclic variants as substrates. It might be assumed that *Tp*NTT3 is quite unrestrictive in terms of purine nucleotide transport, however, certain (metabolically rather irrelevant) purine nucleotides, like 2'3'-cAMP or dADP, seem to be discriminated or to represent only inferior substrates. Therefore, we assume that transport of most if not all experimentally verified substrates is of physiological relevance.

GFP-based targeting studies with *Tp*NTT3 in the homologous host revealed a fluorescence pattern related to the so-called clamp-like structure previously documented for the silaffin kinase STK1 from *T. pseudonana* (Sheppard *et al.* 2009). Interestingly, silaffin proteins which are impaired in their targeting to their final destination, the silica deposition vesicles, also accumulate at this position (Sheppard *et al.* 2009). Silaffins play a key role in the silicon-dependent formation of the cell wall and represent the dominating organic matter in the silica deposition vesicles. These proteins are characterised by a high degree of phosphorylation, which is catalysed by silaffin kinases. The molecular mechanism of the biogenesis and formation of the silica deposition vesicles in diatoms is not completely clarified. However, the clamp-like structure was recently suggested to represent a component in the silaffin trafficking pathway from the ER via transport vesicles (and the Golgi apparatus) to the silicon deposition vesicles (Poulsen et al. 2013, Sheppard et al. 2009). Presence of *Tp*NTT3 in a compartment capable for vesicle formation might explain the fluorescent dots sporadically observed in *P. tricornutum* cells expressing *Tp*NTT3.

So far it is not clear whether the clamp-like structure labels a defined subcompartment associated to the ER, or whether heterogeneous accumulation of the soluble GFP-constructs in the ER lumen causes the observed pattern (Sheppard *et al.* 2009). Similarly, *Tp*NTT3 could reside in the membrane of this subcompartment, accumulates in specific regions of the

ER membrane or of the third plastidial membrane that separates the ER lumen from the periplastidial space.
One possible function of *Tp*NTT3 might be energy provision to fuel silaffin kinases with their cosubstrate ATP. However, in this context it is hard to explain why *Tp*NTT3 not only transports ATP and ADP but also accepts further purine nucleotides, deoxy-forms and particularly cyclic purine nucleotides as substrates, whereas certain structurally related purine nucleotides are specifically excluded from transport. Moreover, the substrate spectrum was probably newly developed to fulfil specific metabolic requirements of the diatom, because it differs from all other previously known NTTs. Therefore, we propose that *Tp*NTT3 might play an additional role in the passage of various purine nucleotides and maybe also in cAMP- and cGMP-based signalling. Interestingly, it was postulated that cyclic nucleotides are involved in the silicon metabolism of the diatom *Cylindrotheca fusiformis*, since it was shown that the levels of cAMP and cGMP change in correlation to silicon starvation (Aline *et al.* 1984, Borowitzka & Volcani 1977). Furthermore, transcriptome analysis of the silicon response of *T. pseudonana* identified several genes proposed to be involved in silica-based cell wall formation, among others GTP and cAMP dependent enzymes (Shrestha *et al.* 2012).
Particularly, the observed subcellular localisation but also the unusual substrate spectrum point to a role of *Tp*NTT3 in silicon metabolism. This is supported by the fact that a direct homologue of *Tp*NTT3 is missing in *P. tricornutum* (Ast *et al.* 2009), a diatom not necessarily depending on silicon (Martino *et al.* 2007). Situated in a membrane that flanks the ER or in a different compartment of the pathway of silica deposition vesicle formation, *Tp*NTT3 might supply energy for phosphorylation and also adjust the internal pool of purine and cyclic nucleotides. However, also a role of *Tp*NTT3 in the nucleotide uptake system of the plastid cannot be ruled out completely.

Acknowledgments
We thank D. Ballert for help with the transformation and cultivation of *T. pseudonana* and *P. tricornutum*. We thank the Bioimaging Center (BIC), University of Konstanz, for access to the imaging core facilities. This study was supported by the University of Konstanz, grants of the Deutsche Forschungsgemeinschaft (Project KR 1661/3-4 and SFB 969 Project A4 to PGK, the Graduate School Chemical Biology KoRS-CB, Project SPP 1131 “Life inside cells” NE 418/8-2 to HEN).

4 The application of metabolic glycoengineering via chemically modified N-acetyl-mannosamine in the diatom *Phaeodactylum tricornutum*

Lili Chu[1,*], Jochen Buck[1], Jana D. R. Schmidt[1], Anne-Katrin Späte[2], Verena Schart[2], Andrea Niederwieser[2], Valentin Wittmann[2], Ansgar Gruber[1], Peter G. Kroth[1]

[1] Fachbereich Biologie, Universität Konstanz, 78457 Konstanz

[2] Fachbereich Chemie, Universität Konstanz, 78457 Konstanz

*Author for correspondence: lili.chu@uni-konstanz.de

Keywords: diatom, fluorophore labelling, click reaction, sialic acids

4.1 Abstract

The attachment of carbohydrates to proteins is an important process for cellular recognition and protein binding or interaction. Benthic diatoms, unicellular algae living on submersed aquatic surfaces, are known to secrete polysaccharides, which are part of a surrounding matrix, consisting of various extracellular polymeric substances (EPS). Glycosylated proteins can be found on most eukaryotic cell surfaces and are also a part of the diatom EPS. However, it is only partially understood how and where the EPS is synthesised. To answer this question, metabolic glycoengineering (MGE) is a promising method. This method exploits the integration of modified monosaccharides into glycans by the native enzyme machinery, where they can be detected by subsequent labelling of the modified carbohydrates via a biorthogonal ligation reaction, for example the azide-alkyne cycloaddition (ACC), also known as click chemistry. This study is a first approach on the application of MGE to study glycans in the diatom *Phaeodactylum tricornutum*. We offered 1,3,4,6-Tetra-O-acetyl-N-3-azidoacetylmannosamine (Ac_4ManNAz) as a substrate to the cells during growth, and later exposed them to Alexa Fluor 488 DIBO alkyne to facilitate ACC. The cells were subsequently observed by fluorescence microscopy. Alexa488-labelled structures at the cell's periphery and an intracellular spot could be visualised. However, controls revealed that the binding of DIBO-Alexa488 occurs non-specifically. Hence, we assume that Ac_4ManNAz is not a suitable substrate taken up by the cells and not part of the glycan composition existing in diatoms.

4.2 Introduction

In all eukaryotic cells, one of the most ubiquitous, and most complex, post-translational modifications of proteins is glycosylation, which occurs in the endoplasmic reticulum (ER) (Friso & van Wijk 2015, Spiro 2002, Walsh *et al.* 2005). There are different kinds of glycosylation, depending on the glycopeptide linkage: N- and O-glycosylation, C-mannosylation, phosphoglycation, and glypiation (Spiro 2002). N-linked glycans, for example, occur commonly in eukaryotes, especially on the cell surface, where they fulfil important roles in cell surface recognition, interactions and binding events (Schwarz & Aebi 2011). Not only intra- but also extracellular compounds may contain or bear glycans.

Benthic diatoms are known to secrete various polysaccharides and glycosylated compounds, generating a so-called extracellular matrix, which consists of extracellular polymeric substances (EPS) (Hoagland *et al.* 1993). This matrix can be involved in the formation of biofilms, in attachment to certain surfaces and for diatom locomotion. It can also offer protection from environmental influences (Poulsen *et al.* 2014, Windler *et al.* 2015).

The biochemistry of N-glycosylation in eukaryotic microalgae has recently been investigated in the diatom *Phaodactylum tricornutum* (Baiet *et al.* 2011) and in the green alga *Chlamydomonas reinhardtii* (Mathieu-Rivet *et al.* 2013). N-glycans have been shown to occur in both organisms, which are phylogenetically very distant and which both gained interest regarding glycan engineering for biotechnological/-pharmaceutical purposes (for review see (Mathieu-Rivet *et al.* 2014). Furthermore, N-glycosylation seems to play a relevant role in the transport of proteins into plastids of diatoms (Peschke *et al.* 2013). Due to their evolutionary origin by eukaryote–eukaryote endosymbiosis diatoms possess plastids which are surrounded by four membranes. The outermost membrane is fused with the ER system, thus all nucleus-encoded plastid preproteins have to be transported across the chloroplast ER.

As reviewed by (Li & Chen 2012), sialic acids are commonly present as terminal residue in glycans of glycoconjugates on the surface of eukaryotic cells. Also in some bacteria, sialic acids are part of capsule-building oligo- or polysaccharides. In eukaryotes, N-acetylneuramic acid (Neu5Ac, the most common sialic acid in nature) is synthesised in the cytosol and activated by addition of cytidine 5'-monophosphate in the nucleus (Kean *et al.* 2004). The formation of glycoconjugates, which are secreted subsequently to the cell surface, is catalysed by sialyltransferases in the Golgi apparatus (Stanley 2011, Wee *et al.* 1998).

Metabolic glycoengineering (MGE) is an approach to label glycosylated proteins by introducing modified monosaccharide residues into cellular glycans. This method is applicable without genetic manipulation, instead it relies on the cellular native enzymatic machinery. MGE may be used to study disruption of glycan biosynthesis, chemical modification of cell surfaces, for probing metabolic flux inside cells and for identifying specific glycoprotein subtypes in the proteome (Dube 2003, Mahal *et al.* 1997). The addition of non-natural sugars either may exploit the ability of enzymes to process metabolites with slightly altered chemical structures, or may intercept biosynthetic pathways that would be maintained in the absence of the non-native substrate. Through incorporation of reactive functional groups (like azides) into glycoconjugates, these chemical 'tags' can serve as tools in pharmaceutical applications like drug delivery (recombinant expression of antibodies, hormones, vaccines etc.) (Campbell *et al.* 2007). Furthermore, MGE has greatly advanced diagnostics by localising and visualising glycans in living organisms, for example by conjugated fluorescent probes (Campbell *et al.* 2007).

Sialic acids and their precursor N-acetylmannosamine (ManNAc) are prominent targets for MGE, due to their terminal position in glycan structures (Du *et al.* 2009). N-acetylmannosamine as a part of glycoproteins in carbohydrate structures seems to be very rare and so far was only found to be integrated into glycoproteins of the gram-positive bacterium *Clostridium symbiosum* (Wittmann 2008). In vertebrates and invertebrates sialic acids are mainly found at the terminal parts of cell surface glycans (Wittmann 2008). Due to their location at the outermost periphery of sialo-glycoproteins, they are well accessible for chemical interactions. N-acetylmannosamine analogues are enzymatically converted into sialic acids analogues inside the cell, and incorporated into sialoglycans (Kayser *et al.* 1992). Being the key enzymes in biosynthesis of sialylated structures, sialyltransferases have frequently been identified and characterised (Li & Chen 2012).

Fluorophore-labelling of metabolically engineered glycoconjugates, based on click chemistry reactions, is currently possible in animal cells lines (Moller *et al.* 2012, Prescher & Bertozzi 2005). Even though this method might offer insights into the nature of glycosylated compounds and the utilisation of modified sugar derivatives, this method, to our knowledge, has not yet been established yet in diatoms or other algae.

4.3 Materials & Methods

4.3.1 Cultivation of *Phaeodactylum tricornutum*

Phaeodactylum tricornutum Bohlin (University of Texas Culture Collection, Austin, strain UTEX646) was cultivated in 16.6 PSU artificial seawater (Tropic Marine, Dr. Biener GmbH, Wartenberg/Angersbach, Germany) enriched with a modified f/2 nutrition described in (Rottberger *et al.* 2013a). Cells were kept in well plates at 20 °C and 35 µmol photons m^{-2} s^{-1} in alternating light conditions (16 h light/8 h dark) in f/2 medium containing 200 µM peracetylated mannosamine with azide-modification (Ac_4ManNAz). Control cells were grown without the addition of Ac_4ManNAz. After 3 days cells were harvested and prepared for labelling via click reaction.

4.3.2 Click-reaction

Cell density was determined by a Multisizer 3 Coulter Counter (Beckman Coulter) and $2*10^6$ cells were harvested by centrifugation (3000 rpm, 3 min), washed with fresh f/2 medium and resuspended in 250 µl f/2 medium.
For copper-free click reaction with azide-modified N-acetylmannosamine, 10 µM of Alexa Fluor 488 DIBO Alkyne (Invitrogen) was added and cells were kept for 30 min at RT. After 5 washing steps with fresh f/2 medium, cells were used for the *in vivo* detection of the reporter-group.

4.3.3 Fluorescence Microscopy

Cellular localisation of the incorporated and labelled Ac_4ManNAz was analysed with a confocal laser scanning microscope LSM 510 META or LSM 880 (Carl Zeiss, Oberkochen, Germany) using a Plan-Apochromat 63 × 1.4 oil-immersion objective (Carl Zeiss, Oberkochen, Germany), 505-530 BP filter and 650 LP filter. Cells were not prepared for fixation prior to observation. Image Processing was conducted using the software ZEN 2009 Light Edition and AxioVision Rel. 4.7 (Carl Zeiss, Oberkochen, Germany).

4.4 Results and Discussion

A number of modified carbohydrates for glycan labelling are available, here we chose to test the possible incorporation of mannosamine into glycosylated structures of the diatom *Phaeodactylum tricornutum*. This mannose derivative is a precursor compound for the synthesis of sialic acids, which occurs at the surface of cell membranes as a compound of glycolipids or glycoproteins (for review see (Ress & Linhardt 2004)).
Mannosamine was per-acetylated and the chemical reporter-group (azide) was attached. The product is then a 1,3,4,6-Tetra-O-acetyl-N-3-azidoacetylmannosamine (Ac_4ManNAz) (Figure 1A). The acetyl-groups increase hydrophobicity of the molecule and make it easier to cross cellular membranes. Azides are small functional groups that are metabolically stable, essentially inert in biological systems, and selectively reactive via Staudinger ligation or azide-alkyne [3+2] cycloaddition (Beckmann & Wittmann 2010, Prescher & Bertozzi 2006). Therefore organic azides (R-N_3) are ideal chemical reporters. In our study, a copper-free [3+2] cycloaddition reaction (Agard *et al.* 2004, Jewett & Bertozzi 2010) of Ac_4ManNAz with the alkyne 4-Dibenzocyclooctynol (DIBO, (Ning *et al.* 2008)) which was conjugated to the fluorophore Alexa 488 (DIBO-Alexa Fluor 488) was used to visualise the modified glycoconjugates.

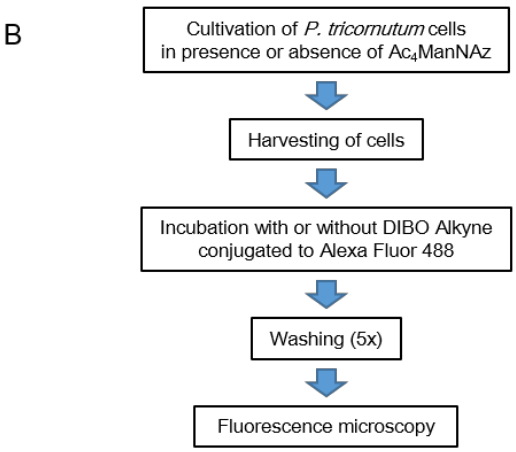


Figure 14. A: Copper-less azide/alkyne click reaction: Ac_4ManNAz (1,3,4,6-Tetra-O-acetyl-N-3-azidoacetylmannosamine) reacts with the alkyne DIBO (4-Dibenzocyclooctynol) conjugated to Alexa Fluor 488., resulting in labelled Ac_4ManNAz. B: Working scheme for fluorophore-labelling in the diatom *Phaeodactylum tricornutum*, based on modified sugar-derivatives and click reaction.

To apply MGE in biological systems, cells need to be cultivated in the presence of special monosaccharides (Figure 14). We have tested several approaches with variable combinations of modified monosaccharide and fluorophore to exclude unspecific labelling (Table 6).

Table 6: List of performed experiments. (+) marks the presence, (-) the absence in the respective experimental phase.

	Presence of Ac_4ManNAz during growth	Incubation with Alexa Fluor 488 DIBO Alkyne	Incubation with Alexa Fluor 488	Result
Labelling experiment	+	+	-	Signal at cell wall/cell periphery, also same signal as in negative control 1 (Fig. 2A, 3)
Negative control 1	-	+	-	Signal from internal structure of cell, one spot at cell center (Fig. 2B)
Negative control 2	+	-	-	No signal (Fig. 2C)
Negative control 3	-	-	-	No signal (Fig. 2D)
Negative control 4	-	-	+	No signal (Fig. 2E)

When the algae were grown in the presence of azide-modified N-acetyl-mannosamine and incubated with DIBO-Alexa Fluor 488, we could detect clear fluorescent signals, which were mostly restricted to the periphery of the cells (Figure 15A). Resulting from two independent experiments, we could observe slightly different phenotypes in the labelling experiment (Figure 16). The signal either surrounded the cell continuously, or was restricted to discrete spots along the outline of the cell (Figure 16). These differences might be due to different growth states of the cells, which change depending on the age of the individual cell and its phase within the cell cycle. Nevertheless, the signals were always detected at the periphery of the cell, which shows a clear difference to negative control 1. In this control, we could also detect weak fluorescent signals with the addition of DIBO-Alexa Fluor 488 without prior growth of the cells in the presence of Ac_4ManNAz (Figure 15B). The possibility that Alexa Fluor 488 is binding non-specifically to compounds within the cells could be excluded by experiments using free Alexa Fluor 488 (Figure 15E). In the control cells without the addition of DIBO-Alexa Fluor 488 (Figure 15C+D), we did not observe any fluorescence.

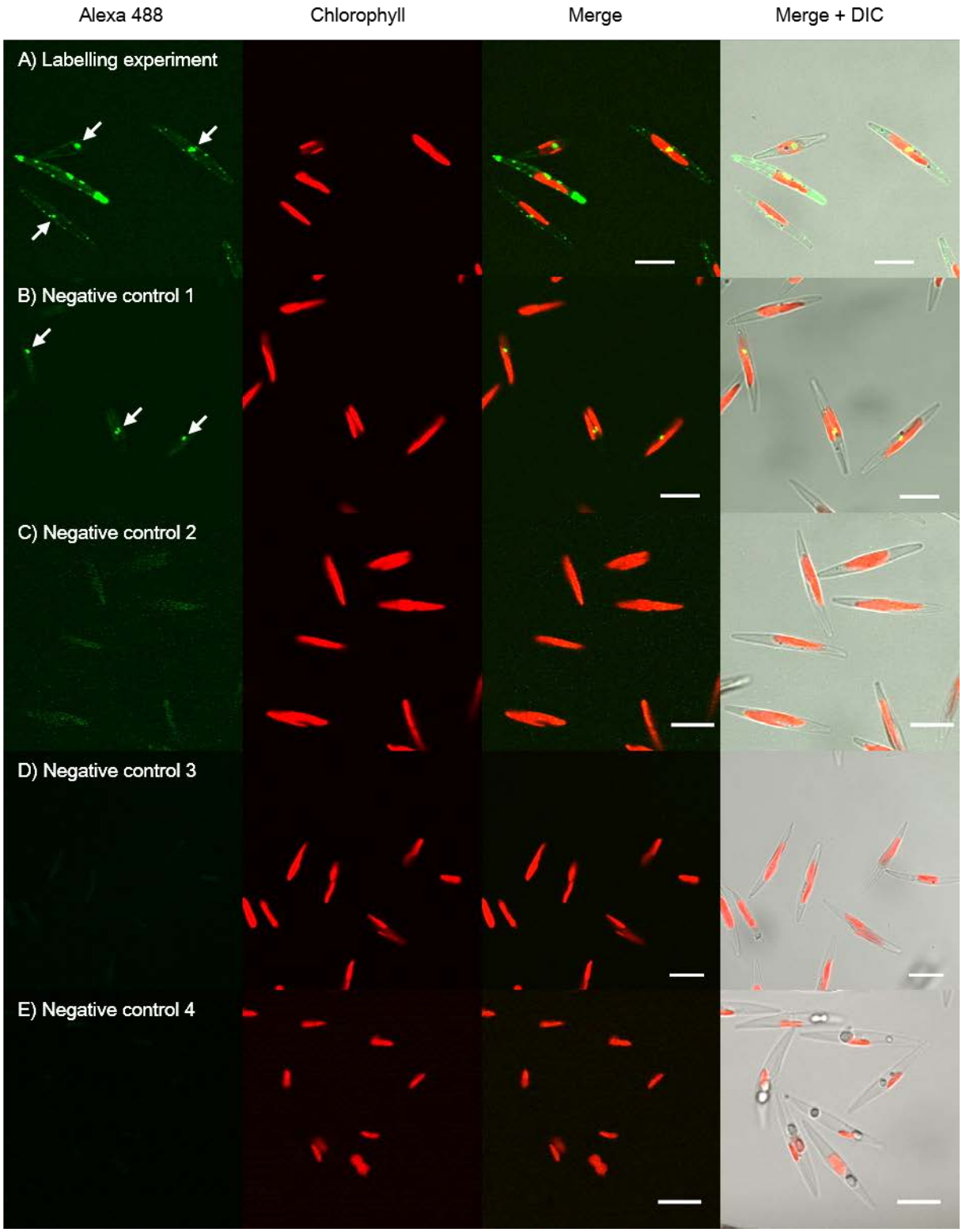


Figure 15: Cellular localisation of *P. tricornutum* cells incubated with Ac_4ManNAz. Alexa Fluor 488 fluorescence in green, chlorophyll autofluorescence in red and Nomarski differential interference contrast (DIC) in grey scale. Scale bars: 10 µm. See Table 6 and Figure 14 for details on the experimental design.

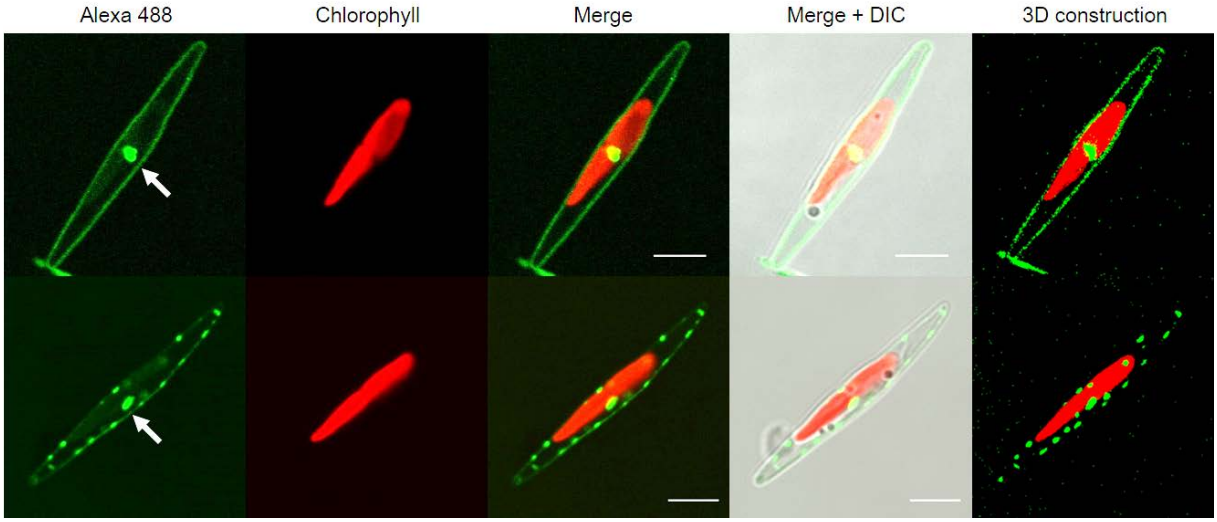


Figure 16: Detail images of single cells from the labelling experiment. Alexa Fluor 488 fluorescence in green, chlorophyll autofluorescence in red and Nomarski differential interference contrast (DIC) in grey scale. 3D construction of Z stack images show chlorophyll and Alexa Fluor 488 fluorescence. Scale bars: 5 µm.

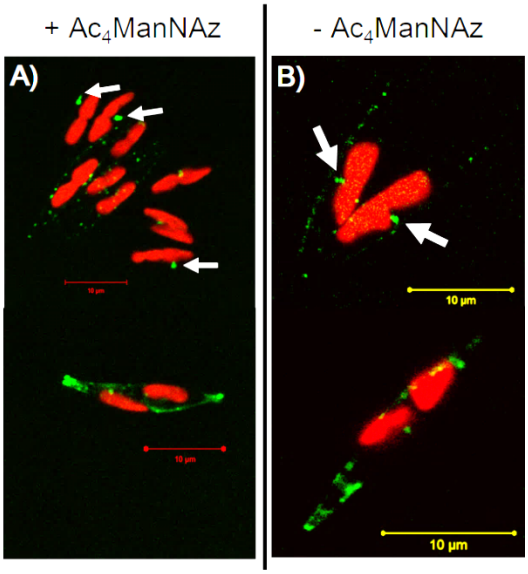


Figure 17: Detail images of single cells from a labelling experiment (A) in comparison with single cells from negative control 1 (B). Alexa Fluor 488 fluorescence in green, chlorophyll autofluorescence in red. Scale bars: 10 µm.

A third independent labelling experiment with longer incubation time of the cells with DIBO-Alexa488 (Figure 17A), showed strong similarities to the negative control 1, where the cells were incubated without Ac_4ManNAz but with DIBO-Alexa488 (Figure 17B). This indicates that the signal is apparently not due to the reaction of DIBO-Alexa488 with the azide-group, but rather due to the attachment of DIBO-Alexa488 to the surface of the cells. Further, DIBO-Alexa488 seems to be able to enter the cell and might bind non-specifically to other compounds, possibly to thiol-groups (Figure 18), e.g. glutathione, an important component in living organism for detoxification processes.

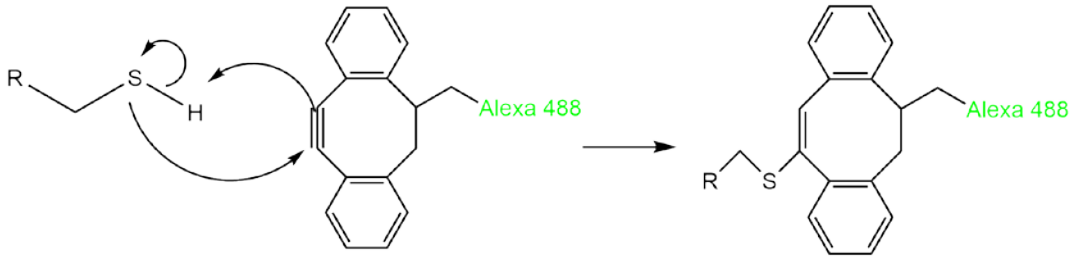


Figure 18: A possible side-reaction resulting from unspecific binding of DIBO-Alexa488 to thiol-groups.

In all observed cells, we could not detect intracellular signals other than the unspecific signal in the centre of the cells. This indicates that there is no detectable amount of Ac_4ManNAz inside the cells. There are three possible explanations for this absence of intracellular signal: i) the amount of Ac_4ManNAz might not be sufficient for subsequent fluorescence microscopic detection, ii) Ac_4ManNAz might enter the cells and reacts with DIBO-Alexa488, but is then degraded resulting in the fluorescent accumulation of DIBO-Alexa488 also present in negative control 1 or iii) Ac_4ManNAz might not be taken up by the cells in the first place, which is possibly due to the cell wall of the diatoms.
There is so far only one know example of a naturally occurring azide, namely in the dinoflagellate *Karenia brevis*, an organism known for the production of toxic red tides (Griffin 1994). Apart from this exception, azides are thought not to occur naturally, neither extra- nor intracellularly (Prescher & Bertozzi 2005). Therefore, to explain the intracellular signal observed in negative control 1, we have to conclude that Alexa Fluor 488 DIBO alkyne accumulates in a certain region of the cell (white arrows in Figure 15 + Figure 16 + Figure 17). This region could either function as a place for storage of compounds which form non-specific aggregates with Alexa Fluor 488 DIBO alkyne or might be involved in degradation/detoxification, possibly by reaction of the alkyne-group with other intracellular (thiol-) compound.

4.5 Conclusion

Based on our observation of fluorescence in the cell periphery and the undefined accumulation inside the cell of *Phaeodactylum tricornutum*, we conclude that DIBO-Alexa488 can enter the cell where it might accumulate. Additionally, this compound attaches to the surface of the diatoms where it causes unspecific signals. Apart from *in vivo* labelling experiments, alternative methods confirmed that sialic acids are not present in *P. tricornutum* (Baiet *et al.* 2011, Mathieu-Rivet *et al.* 2014, Schmidt 2016). Hence, an intracellular incorporation of Ac_4ManNAz into glycosylated compounds is not very likely. Accordingly, Ac_4ManNAz is probably not a suitable compound to trigger the uptake and the reaction with the fluorophore-labelled alkyne DIBO. However, it might be successful with other modified carbohydrates. Furthermore, if the cell wall of *P. tricornutum* represents the major barrier for a successful labelling, we would suggest to utilise modified cells which are capable of expressing hexose-transporters in the cell membrane (like in (Zaslavskaia *et al.* 2001)) to ensure the uptake of the modified carbohydrates in the first place.

Acknowledgements
We like to thank C. Río Bártulos and S. Kalkan for helpful discussions. We also like to thank the BioImagingCenter (BIC), Universität Konstanz for access to the confocal laser scanning microscope. This work was supported by the Universität Konstanz and by grants of the Deutsche Forschungsgemeinschaft (DFG/SFB969, TPA4).

5 The application of the self-assembling GFP and its limitations in the diatom *Phaeodactylum tricornutum*

Lili Chu*, Ansgar Gruber, Peter G. Kroth

Fachbereich Biologie, Universität Konstanz, 78457 Konstanz, Germany

* Author for correspondence: lili.chu@uni-konstanz.de

5.1 Abstract

To localise proteins intracellularly, the expression and localisation of GFP-fusion proteins is a commonly used method in molecular cell biology. However, to distinguish the GFP-phenotypes of strains expressing fusion proteins targeted to subcompartments of the same organelle, still represents a great challenge. In diatom plastids, which are surrounded by four membranes, in total five possible compartments can harbour soluble proteins, not counting possible locations for membrane proteins. The self-assembling GFP (saGFP) technique, utilising two GFP fragments with a strong affinity to each other, is supposed to circumvent these problems by the utilisation and co-expression of marker proteins together with the protein of interest. GFP fluorescence is then an indicator that both fragments are in the same compartment. We performed protein targeting experiments using saGFP in the diatom *Phaeodactylum tricornutum*, in order to test this method for plastidial marker proteins. A large number of unexpected GFP fluorescence in strains where GFP fragments had been targeted to different compartments, questioned the conclusiveness of this method. Control experiments demonstrate that molecular tools, like the promoter and marker sequences, need to be checked carefully when used for this approach. Furthermore, we found that also the protein targeting process was strongly affected in this system.

5.2 Introduction

Expression of the green fluorescent protein (GFP) is an efficient and frequently used tool in cell and molecular biology. From its inception in the 1990s to now, it is commonly used to achieve a series of different goals, ranging from localisation of proteins to protein purification and expression allowing insights into all kinds of cellular metabolic mechanisms. In 2008, Osamu Shimomura, Martin Chalfie and Roger Y. Tsien received the Nobel Prize in Chemistry “for the discovery and development of the green fluorescent protein, GFP” ("The Nobel Prize in Chemistry 2008", *Nobelprize.org*).

Fusing GFP to other proteins can reduce the folding properties of GFP (Peelle *et al.* 2001, Waldo 2003, Waldo *et al.* 1999). Hence, a robustly folded ‘superfolder’ GFP was developed (Cabantous *et al.* 2005, Pedelacq *et al.* 2006) in order to circumvent this risk. Cabantous et al. engineered self-associating GFP fragments, which can be used to tag proteins in living cells and which does not change the fusion protein solubility as shown for other split protein tags (Nixon & Benkovic 2000, Ullmann *et al.* 1967, Wehrman *et al.* 2002, Wigley *et al.* 2001). Cabantous et al. increased the brightness of the GFP-fluorescence roughly 80-fold through the evolution of GFP1-10 by DNA-shuffling (Stemmer 1994) and subsequent point mutations of the exceptionally stable ‘superfolder’ GFP. Furthermore, the GFP11 part was selected after testing solubility and least perturbation of folding kinetics. Neither of the parts alone is fluorescent. Due to their strong affinity to each other, the parts self-assemble as soon as they are situated in the same compartment and hence rearrange the barrel structure of GFP to regain full fluorescence. This property is used in the self-assembling GFP method, where each GFP fragment is fused to a protein of interest and allows to expand protein tagging and localisation experiments. This technique, also referred to as saGFP was first described by Cabantous *et al.* (2005). The termination ‘split GFP’ is also found in the literature, describing the same method. However, in this study the ‘self-assembling GFP’ is based on the affinity of the GFP-fragments to find the respective counterpart, whereas the ‘split GFP’ is based on the separation of two even GFP-halves, which can only reassemble, as soon as the fusion proteins they are attached to are in close proximity (bait and prey interaction). Accordingly, the split

GFP fragments can only restore GFP-fluorescence if the associated proteins interact and thus this method is rather used to detect protein-protein-interaction (Wilson *et al.* 2004).
This saGFP method was improved *in vivo* in bacteria regarding solubility, complementation, quantification, folding kinetics and stability (Cabantous *et al.* 2005, Cabantous & Waldo 2006, Pedelacq *et al.* 2006). The consequence is a complementation system which is supposed to minimise perturbation on protein folding or solubility, to maximise stability and to prevent misfolding of the target proteins.
In this study, we are focussing on the utilisation of the self-assembling GFP system to locate proteins that are located in different compartments, using appropriate marker proteins for these compartments. After introduction of the transgenes into the cells, both, marker protein and the protein of interest, will be expressed, resulting in assembled and fluorescent GFP if the protein of interest is located in the same compartment as the reference protein.
In particular the localisation of plastidial proteins in diatoms represents a challenge since diatom plastids are surrounded by four membrane, resulting in two additional compartments compared to land plant plastids (Delwiche & Palmer 1997). This fact impedes the exact localisation of proteins which are targeted to plastidial compartments and/or membranes, since the discrimination of the individual compartments and membranes surrounding the plastid is hard to define. The compartments (from outside to inside) of the diatom plastids are thereby classified as follows: cytosol, chloroplast ER lumen (cER), periplastidial space (PPS), interenvelope space (IES) and stroma.
The saGFP method has been applied in the diatom *Phaeodactylum tricornutum* in several studies in order to distinguish localisation of proteins expressed in the compartments surrounding the plastid (Bullmann *et al.* 2010, Felsner *et al.* 2010, Hempel *et al.* 2009, Hempel *et al.* 2010, Lau *et al.* 2016, Lau *et al.* 2015, Moog *et al.* 2015, Vugrinec *et al.* 2011): cells were transformed with gene constructs encoding the large GFP-fragment (GFP1-10) fused to a certain marker protein which is known to be targeted to a special compartment, whereas the small counterpart (GFP11) is fused to the protein of interest. GFP-fluorescence is supposed to be restored as soon as both fragments are expressed and situated in the same cellular compartment, revealing the localisation of the protein of interest.
In our study, we intended to test this system in *P. tricornutum* with respect to the choice and utilisation of appropriate reference proteins for compartments surrounding the diatom plastid.

5.3 Materials & Methods

5.3.1 Generation of expression constructs

PCR-amplifications of the coding sequence of the following proteins were conducted based on sequence analyses in the JGI genome database (http://genome.jgi-psf.org): phosphoglycerate kinase (PGK, Protein-ID 51125), ER luminal binding protein (BIP, Protein-ID 54246), 6-phosphogluconate dehydrogenase (6PGDH, Protein-ID 45333), plastidial type 1 signal peptidase (SP1, Protein-ID 17972), modified oxygen-evolving enhancer protein 1 precursor (mOEE1, Protein-ID 20331) and presequences were predicted using the CBS prediction server (SignalP3.0, http://www.cbs.dtu.dk/services/SignalP-3.0/ and TargetP1.1, http://www.cbs.dtu.dk/services/TargetP/). Oligonucleotides for amplification and screening were designed (Sigma Aldrich, St. Louis, MO, USA) and listed in Table S 5. cDNA and Plasmid-DNA from previous studies were used as template for amplification (see references

Table S 6). The correct targeting to the respective compartments was verified by the references listed in Table 8. Amplification products were purified from the agarose gel with the GeneClean Turbo Kit (MP Biomedicals, Illkirch, France) and inserted either into the pPha-T1 vector (PTV) (Zaslavskaia *et al.* 2000) (GenBank AF219942, lhcf1 promoter (Apt *et al.* 1996)) containing the eGFP gene (Clontech, Palo Alto, CA) (Gruber *et al.* 2007), or into the inducible vector pPha-T1-NR (PTV-NR) (Hempel *et al.* 2009) containing the nitrate-reductase (nr) promoter (Poulsen & Kröger 2005). The identity of each construct was verified by sequencing (GATC Biotech AG, Konstanz, Germany).

Table 7: Transformed cell lines of *Phaeodactylum tricornutum* used for time-dependent GFP-fluorescence measurements by flow cytometric and fluorescence microscopic analyses.

Cell line name	Description
lhcf1-GFP	*P. tricornutum* UTEX646, cell line resulting from genetic transformation with pPha-T1-GFP vector, containing the eGFP gene under control of the lhcf1 promoter region
nr-GFP	*P. triconutum* UTEX646, cell line resulting from genetic transformation with the pPha-NR-GFP vector, containing the eGFP gene under control of the nr promoter region
sa-nr-PGK+PGK	*P. triconutum* UTEX646, cell line resulting from genetic co-transformation with the cytosolic marker, using the vectors PGKpre-GFP1-10_PTV-NR + PGKpre-GFP11_PTV-NR
sa-nr-6PGDH+6PGDH	*P. triconutum* UTEX646, cell line resulting from genetic co-transformation with the PPS marker, using the vectors 6PGDHpre-GFP1-10_PTV-NR + 6PGDHpre-GFP11_PTV-NR
sa-nr-mOEE1+PGK	*P. triconutum* UTEX646, cell line resulting from genetic co-transformation with the stroma and the cytosolic marker, using the vectors mOEE1pre-GFP1-10_PTV-NR + PGKpre-GFP11_PTV-NR
sa-nr-6PGDH+BiP	*P. triconutum* UTEX646, cell line resulting from genetic co-transformation with the PPS and the ER marker, using the vectors 6PGDHpre-GFP1-10_PTV-NR + BiPpre-GFP11_PTV-NR

lhcf1 = chlorophyll *a/c*-binding light harvesting complex protein; nr = nitrate reductase; sa = self-assembling

5.3.2 Cultivation of *Phaeodactylum tricornutum*

P. tricornutum Bohlin (University of Texas Culture Collection, Austin, strain UTEX646, denoted “Pt4” by DeMartino et al. (Martino *et al.* 2007)) was grown in 16,6 PSU artificial seawater (Tropic Marine, Dr. Biener GmbH, Wartenberg/Angersbach, Germany) enriched with modified f/2 nutrition as described in (Rottberger *et al.* 2013a). Cells were grown under continuous illumination at 55 µmol photons m^{-2} s^{-1} at 18 °C either in liquid cultures on a horizontal shaker at 120 rpm or on solid medium containing 1.2 % Bacto Agar (Difco, Becton Dickinson and Company, Le Pont de Claix, France).

5.3.3 Biolistic transformation

Cells were transformed using the Biolistic PDS-1000/He Particle Delivery System (Bio-Rad) fitted with 1350 psi rupture discs as described in (Kroth 2007b). After transformation, cells were allowed to recover for 24 hours before being plated onto f/2 medium containing 75 µg/ml zeocin (Invitrogen) for selection. The plates were incubated at 22 °C under constant illumination (75 µmol photons m^{-2} s^{-1}).

5.3.4 Flow cytometry

Transformed cell lines were screened for GFP-expression using the flow cytometer Cell Lab Quanta SC MPL (Beckman Coulter). Cells were excited with a 488 nm laser, autofluorescence of chlorophyll was detected by a 670 long pass (LP) filter (FL3-670 nm), while GFP fluorescence was detected in a second channel by a 525/40 band pass (BP) filter (FL2-575 nm). To ensure that only living cells were analysed, chlorophyll autofluorescence was chosen as trigger criterion and data were gated for chlorophyll autofluorescence as well as electronic volume size range. The emission intensities of 10 000 cells per sample were collected and mean fluorescence intensities higher than 100 were counted as a positive signal. The background fluorescence intensity of wild type cells ranges from 60-70.
Time-dependent nitrate reductase induction experiments (see below) were performed by determination of GFP-fluorescence intensity using the flow cytometer BD FACSCalibur and the software BD CellQuestPro (BD Biosciences, CA, USA). For the detection of GFP, we used the 488 nm laser for excitation and the FL1 detector with a 530/30 BP filter for detection. For the detection of red fluorescence, we utilised in parallel the FL3 detector with a 650 LP filter. Emission intensities of 50 000 and 100 000 cells, respectively, per sample (triggered by side scatter) were collected and ungated median fluorescence intensities of each population were collected for the subsequent analyses. Dot plots and histograms were created and analysed using the Single Cell Analysis Software FlowJo (Tree Star, Inc. Ashland, OR).
Wild type cell cultures were used as negative control and a lhcf1-GFP transformed cell line as reference for a GFP-expressing cell line under the control of a nitrate-independent promoter.

5.3.5 Induction of nitrate reductase promoter

Cells were cultivated for 3 days in 50 mL liquid f/2 medium containing 0.882 mM NH_4Cl and under continuous light illumination at 55 µmol photons m^{-2} s^{-1} and 18 °C. Cells were harvested during exponential phase by centrifugation (3000 rpm, 10 min) and resuspended in f/2 medium containing 0.882 mM $NaNO_3$. Fluorescence was measured with the BD FACSCalibur as described above. Samples were taken every 3 hours, except the last sample. Microscopic analyses were performed in parallel.

5.3.6 Fluorescence Microscopy

Cellular expression GFP fusion proteins was analysed with an epifluorescence microscope Olympus BX51 (Olympus Europe, Hamburg, Germany) equipped with the Zeiss AxioCam MRm digital camera (Carl Zeiss, Oberkochen, Germany) and an Olympus PLN 40 × objective (Olympus Europe) or with a confocal laser scanning microscope LSM 510 META (Carl Zeiss) using a Plan-Apochromat 63 × 1.4 oil immersion Nomarski differential interference contrast (DIC) objective (Carl Zeiss). Image processing was conducted using AxioVision Rel. 4.7 and ZEN lite software (Carl Zeiss).

5.4 Results

5.4.1 Marker proteins for the respective plastidial subcompartment fused to GFP

We first chose appropriate marker proteins for each plastidial compartment and for the cytosol of the diatom *Phaeodactylumt tricornutum*. Previous localisation studies based on GFP full length fusion proteins confirmed the targeting properties of these proteins (see references listed in Table 8). The microscopic re-investigation of these cell lines is shown in Figure 19. PGK::GFP expression shows a characteristic cytosolic distribution of GFP throughout the whole cell and in addition an accumulation close to the plastid, possibly the nucleus (see (Gould *et al.* 2006)). BiP::GFP fusion proteins label a circular structure, which indicates the location of the nuclear membrane, next to a network-like structure distributed around the plastid as well as throughout the cell. This phenotype is representative for the chloroplast ER in diatoms, which is continuous with the plastid and thus surrounding both the plastid and the nucleus (Apt *et al.* 2002). The 6PGDH::GFP expressing cell line shows a so-called 'blob-like structure' (BLS, GFP-accumulation near to the plastid), which is supposed to represent the periplastidial space between the two inner and two outer membranes of the plastid (Gruber *et al.* 2009, Kilian & Kroth 2005). The GFP-signal of SP1::GFP matches entirely the chlorophyll-signal, which is also the case for mOEE1::GFP, and which indicates a location either in the stroma or a neighbouring compartment, eventually the interenvelope space. SP1 is reported to be expressed in the interenvelope space (Bullmann *et al.* 2010), which might be difficult to distinguish from the stroma, since the IES is surrounding the stroma and GFP expression in the IES would probably result in a similar GFP-pattern. OEE1 has been modified by a truncation of C-terminal domain of the presequence to prevent targeting to the thylakoid lumen, hence would result in a stromal localisation (Kilian & Kroth 2005). But in fact, both compartments are not explicitly distinguishable.

Table 8: Marker proteins used for GFP-localisation studies.

Marker protein	JGI Protein ID	Compartment	Reference
PGK (phosphoglycerate kinase)	51125	cytosol (C)	(Río Bártulos 2007)
BiP (ER luminal binding protein precursor)	54246	chloroplast ER lumen (cER)	(Apt *et al.* 2002)
6PGDH (6-phosphogluconate dehydrogenase)	45333	periplastidic space (PPS)	(Gruber *et al.* 2009)
SP1 (plastidal type I signal peptidase)	17972	interenvelope space (IES)	(Bullmann *et al.* 2010)
mOEE1 (modified oxygen-evolving enhancer protein 1 precursor)	20331	stroma (S)	(Kilian & Kroth 2005)

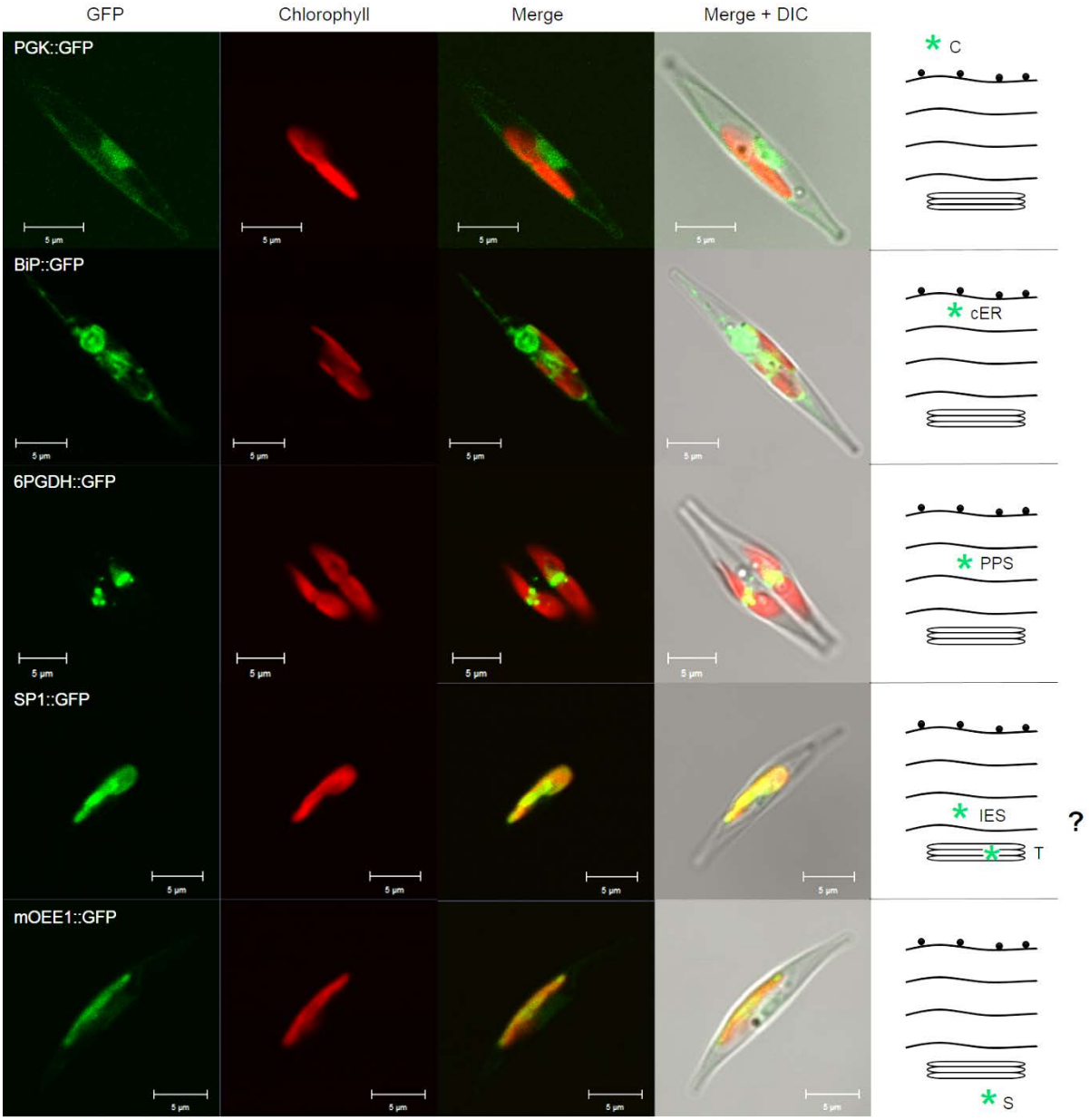


Figure 19: Cellular localisation of full length GFP fused to presequences of marker proteins listed in Table 8. GFP fluorescence in green; autofluorescence of chlorophyll in red: Nomarski differential interference contrast (DIC) in grey scale. Scale bars: 5 µm. C = cytosol; cER = chloroplast endoplasmic reticulum; PPS = periplastidial space; IES = interenvelope space; T = thylakoid; S = stroma.

5.4.2 The utilisation of the self-assembling GFP in *P. tricornutum*

The initial purpose for using saGFP were studies of the localisation of proteins situated in the membranes that are surrounding the plastid of *P. tricornutum*. Multiple GFP-signals with different co-transformations (data not shown) initiated testing of the reference proteins which were involved.

Therefore, we generated gene constructs encoding the presequences (according to (Kilian & Kroth 2005)) of each marker protein to GFP1-10 and also to its counterpart GFP11 (for amino acid sequences see Supporting Information, Chapter 5). We used SP1 full length sequence fused to GFP11, according to (Vugrinec *et al.* 2011). We co-transformed each marker labelled with GFP1-10 with another marker labelled with GFP11. These combinations were expected to result in no fluorescence since the GFP-parts should be targeted to different compartments. As a control, we also co-transformed each of the GFP-counterparts fused to the same marker, which should result in a GFP-signal in the respective compartment according to expectations. The expression of the gene constructs was controlled by the inducible nitrate reductase (nr) promoter, which is activated by the presence of nitrate (NO_3^-) and inactivated by the presence of ammonium (NH_4^+) (Hempel *et al.* 2009, Poulsen *et al.* 2006, Poulsen & Kröger 2005).

We investigated a total number of 480 transformant cell lines, whereof 227 cell lines were co-transformed with constructs encoding the presequences targeted to the same compartment. In these cases, the GFP-fragments should be able to reassemble, resulting in a GFP signal. Surprisingly, a large number of cell lines, 155 out of 227, were non-fluorescent, corresponding to 68%.

Expressed GFP-fragments targeted to different comparments due to the attached presequences, should not show any GFP fluorescence. However, within 253 cell lines, we found 21 transformant cell lines, corresponding to 8 %, which showed an unexpected GFP signal (Table 9, highlighted in grey). Out of these, we chose two cell lines (sa-nr-mOEE1+PGK and sa-nr-6PGDH+BiP), which were further analysed by nitrate-dependent induction of the promoter and subsequent flow cytometric analyses.

Table 9: Resulting amounts of cell lines after co-transformation of gene constructs encoding marker proteins fused to GFP1-10 and GFP11, respectively, in *Phaeodactylum tricornutum*.

	C: PGK-GFP11		**ER: BiP-GFP11**		**PPS: 6PGDH-GFP11**		**IES: SP1-GFP11**		**S: mOEE1-GFP11**	
	FL	NFL	FL	NFL	FL	NFL	FL	NFL	FL	NFL
C: PGK-GFP1-10	42% * (24/57)	58% (33/57)	13% * (3/23)	87% * (20/23)	0% (0/1)	100% * (1/1)	0% (0/89)	100% (89/89)		
ER: BiP-GFP1-10			39% * (12/31)	61% (19/31)			1.5% * (1/65)	98.5% * (64/65)		
PPS: 6PGDH-GFP1-10	0% (0/1)	100% * (1/1)	20% * (1/5)	80% * (4/5)	17% * (21/124)	83% (103/124)	0% (0/2)	100% * (2/2)		
IES: SP1-GFP1-10							100% * (5/5)	0% (0/5)		
S: mOEE1-GFP1-10	23% * (5/22)	77% * (17/22)	13.5% (5/37)	86.5% (32/37)	33% * (1/3)	67% * (2/3)	100% * (5/5)	0% (0/5)	100% * (10/10)	0% (0/10)

Numbers without brackets indicate the amount of transformants in %, showing GFP-fluorescence (FL) or no GFP-fluorescence (NFL). Numbers in brackets indicate the absolute amounts of transformants. Fluorescence was determined by flow cytometry. Highlighted in grey are the amounts of cell lines co-transformed with differently targeted constructs, but showing unexpected GFP-signals. Asterisks label co-transformations including cell lines from a previous approach where the several cell lines were pre-selected according to the expected result.

5.4.3 Observation of time-dependent cellular GFP-expression after induction of the nr promoter

As described above, GFP signals were also found in saGFP cell lines where no signal was expected. We assumed that signals are possibly accumulating non-specifically after a certain time period. In other studies using expression of saGFP under control of the nr promoter, a connection between expression time and subcellular localisation of GFP is indicated and an expression time of 6 hours after induction of the promoter is recommended (Hempel *et al.* 2009, Hempel *et al.* 2010). To test if the subcellular GFP-phenotype would change over time, we observed the development of the saGFP fluorescence intensity for 27 hours to see eventual intracellular changes in GFP-localisation (Figure 20). We examined the wild type and transformant cell lines as listed in Table 7: two cell lines were expressing the entire GFP molecule (lhcf1-GFP which is constitutively active, and nr-GFP which is inducible under the experimental settings) and the other four cell lines were co-expressing both saGFP counterparts.

The cells were first cultivated in ammonium-containing medium in order to keep the nr promoter in an inactive state. Measurements were started after transferring cells into medium containing nitrate (Figure 20, 0 h, indicated by the arrow).

Since the red autofluorescence of the chlorophyll resulted in a slight background fluorescence (bleed-through) when detected by the green fluorescence channel, also the wild type displayed a value between 5 and 7 (Table S 7), whereas GFP-expressing cell lines showed higher values, up to 310 (nr-GFP after 27 hours, Table S 7). An increase of green fluorescence caused by a spill-over of chlorophyll autofluorescence could be excluded by determining red

fluorescence intensities whereby no change within one cell line could be observed throughout the experiment (Figure S 15 and Table S 7).
At the starting point, the nr-GFP and sa-nr-mOEE1+PGK cell line showed a slightly increased value compared to the background fluorescence of the wild type cells (40 %), indicating that a small GFP amount might be already present.
The GFP intensity measurements showed that after 12 hours, the nr-GFP cell line, containing the inducible nr promotor expressing full length GFP, had an intensity comparable to the lhcf1-GFP cell line which showed constantly a high fluorescence intensity throughout the experiment and is controlled by the constitutive lhcf1 promotor (25-fold higher than the wild type). The fluorescence intensity level even exceeded the lhcf1-GFP cell line after 15 hours (33-fold higher than the wild type), suggesting that the nr promoter (once induced) is stronger than the lhcf1 promoter (Chu *et al.* 2016a).
The cell lines expressing self-assembling GFP, which are also controlled by the nr promoter, showed a generally lower and a more decelerated development of GFP fluorescence intensity than the full length GFP expressing nr cell line. This indicates that the self-assembling process is a rather slow mechanism. The cell lines differ strongly between each other: sa-nr-PGK+PGK showed the strongest GFP signal (after 27 hours even 50 % stronger than lhcf1-GFP) and sa-nr-6PGDH+BiP showed only a faint GFP-signal (compare Figure 20 + Figure S 14 and Table S 7). Among the saGFP cell lines, sa-nr-PGK+PGK and sa-nr-6PGDH+6PGDH, which both expressed the GFP counterparts targeted to the same compartment, showed a similar increase in GFP expression compared to the nr-GFP cell line (highest increase between 3 and 6 hours after induction), resulting in an increase of fluorescence intensity up to threefold.
However, even when targeted to different compartments (sa-nr-mOEE1+PGK and sa-nr-6PGDH+BiP), the GFP-fragments surprisingly were still assembled and fluorescence was visible, even though the development of GFP fluorescence intensity seemed to be slower: the highest rate of fluorescence intensity increase (70 %) appeared between 6 and 9 hours after induction (sa-nr-mOEE1+PGK). The cell line sa-nr-6PGDH+BiP showed in general only a low rise of fluorescence intensity (only up to 30 % within the first 3 hours) throughout the measurements, ending in a 2-fold higher intensity compared to the wild type after 27 hours.
Furthermore, the flow cytometric measurements revealed that, despite the preceding cultivation in ammonium-containing medium, the cell lines nr-GFP and sa-nr-mOEE1+PGK either contained residual GFP-molecules which were not entirely degraded or that the nr promoter was not completely switched off (Table S 7).

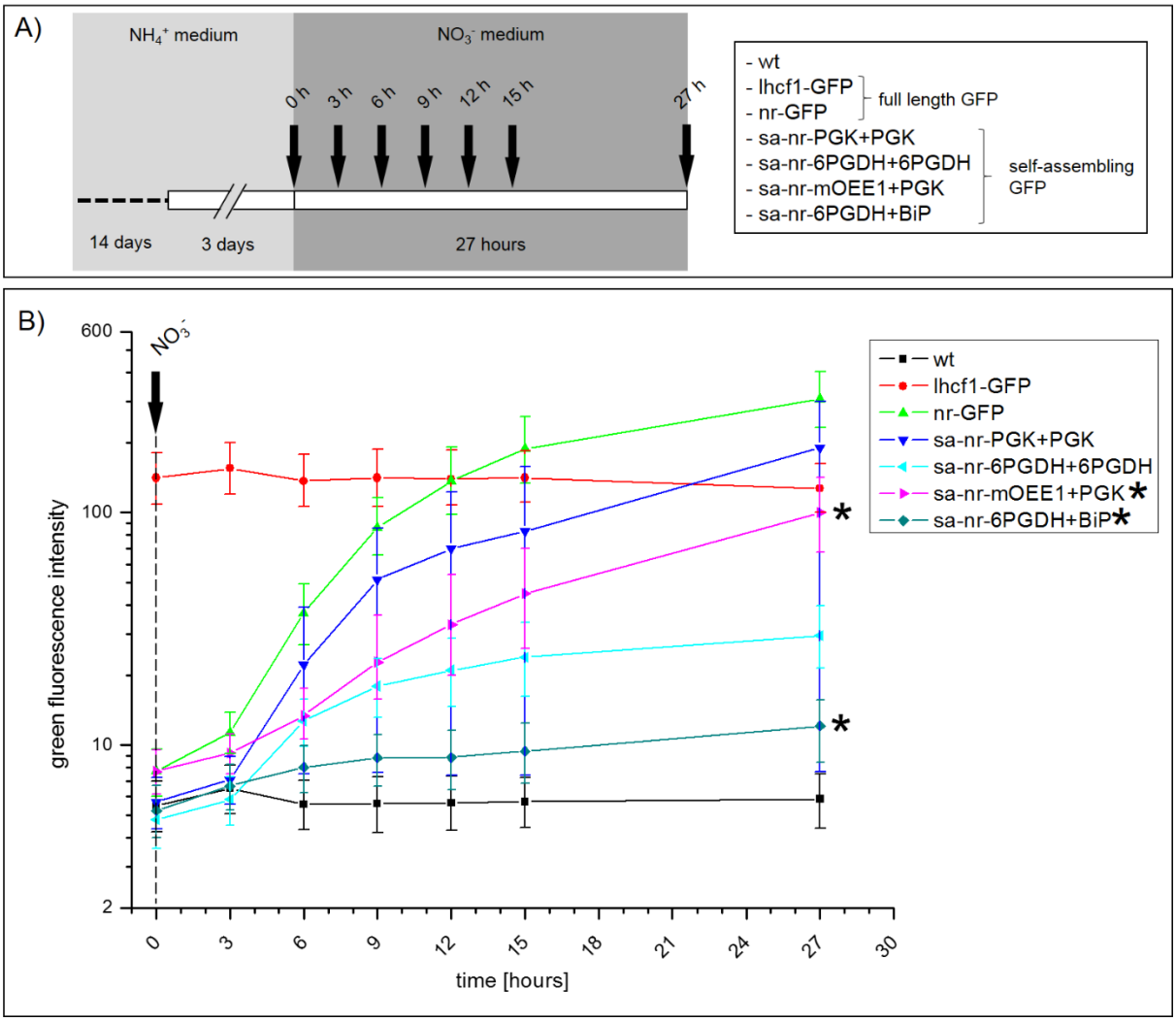


Figure 20: A) Time scale for flow cytometry sampling. *P. tricornutum* cell lines were kept in ammonium-medium (NH_4^+) prior to the start of the experiment by transferring the cell lines into nitrate-medium (NO_3^-) (0 h) and cultivation for 27 hours at continuous light. Samples were taken at the indicated time intervals. B) Median green fluorescence intensities determined for *P. tricornutum* wild type and GFP-expressing cell lines. Intensities were determined using the flow cytometer BD FACSCalibur at indicated time intervals. Bars represent the interquartile range (IQR). Asterisks label cell lines that were transformed with differently targeted marker proteins, resulting in unexpected GFP fluorescence.

5.4.4 Phenotype of the GFP-expressing cell lines

Considering the phenotype of the transformed cell lines, both full length GFP expressing cell lines (lhcf1-GFP and nr-GFP) showed a cytosolic GFP signal, according to expectations (Figure 21). The cell line sa-nr-PGK+PGK showed the same phenotype. Since the GFP-fragments were targeted to the cytosol due to PGK presequence, this result has been also as expected. Surprisingly, we found the same phenotype in the sa-nr-6PGDH+6PGDH cell line, which was supposed to result in a GFP signal in the PPS (compare Figure 19). The same applies to cell lines that expressed differently targeted GFP-fragments, resulting in an unexpected GFP-signal: sa-nr-mOEE1+PGK, sa-nr-6PDGH+BiP (both Figure 21) and sa-nr-PGK+BiP (Figure 22A). This suggests that in these cell lines, saGFP is assembled in the cytosol and import of the protein is blocked.
GFP signals were also detected in plastid-associated locations as shown for sa-nr-mOEE1+SP1 and sa-nr-BiP+SP1 (Figure 22B+C), indicating that import and preassembling may occur in the subcompartments of the plastid. Since the GFP pattern differs from stromal autofluorescence (indicated by the white arrows), we assume that GFP might be located in the interenvelope space due to the marker protein SP1.
A time-dependency of the expression of saGFP fragments under control of the nr promoter could not be confirmed, since the observation via fluorescence microscopy did not show any changes in the subcellular GFP-pattern throughout the experiment.

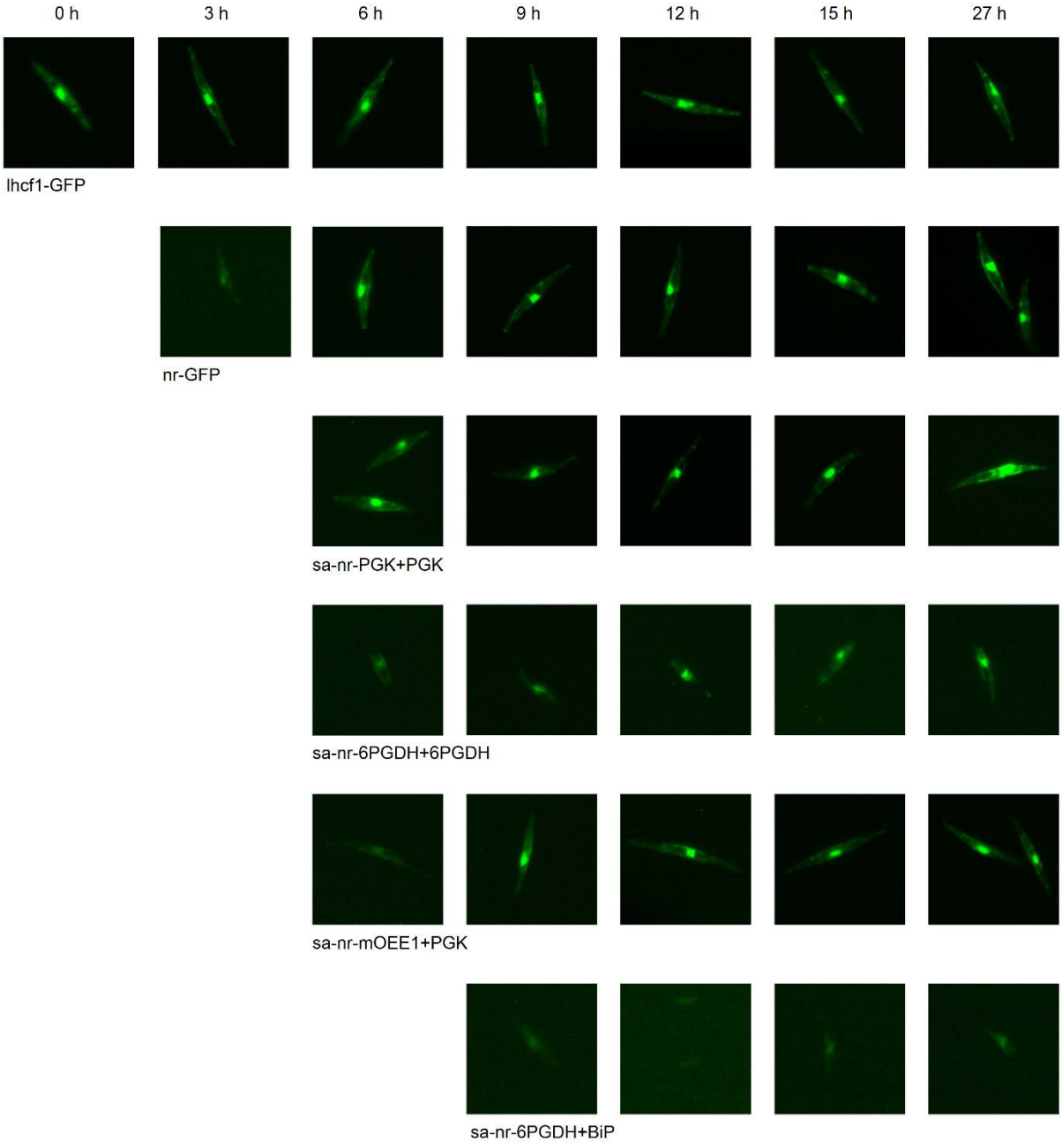


Figure 21: Microscopic images of the GFP fluorescence kinetics of full length GFP and self-assembling GFP expressing *P. tricornutum* cell lines under inducible conditions. Development of subcellular GFP expression was observed via fluorescence microscopy after transfer of the cells into nitrate medium (o h) and cultivation for 27 hours (27 h) at continuous light. Pictures were taken at the indicated time intervals with 40× magnification. For chlorophyll and merged images see Supplementary Data (Figure S4-S10).

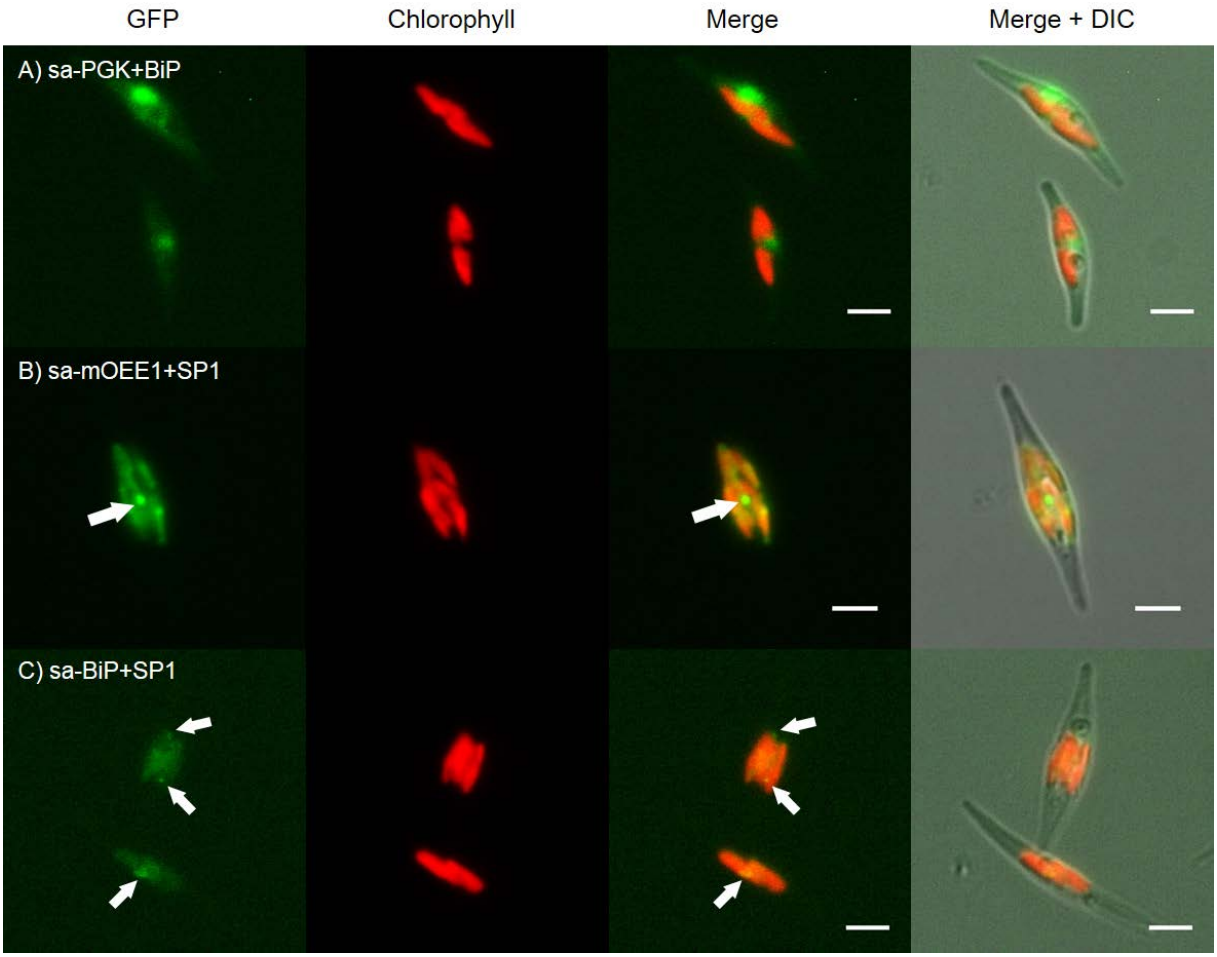


Figure 22: Cellular localisation of self-assembling GFP in *P. tricornutum* cells. Cells were co-transformed with constructs encoding differently targeted marker protein presequences (except SP1) fused to either GFP1-10 or GFP11. GFP fluorescence in green; autofluorescence of chlorophyll in red: Nomarski differential interference contrast (DIC) in grey scale. Scale bars: 5 µm.

5.5 Discussion

5.5.1 The objectives for using saGFP in *P. tricornutum*

There are several reports about the application of the self-assembling GFP complementation assay in mammalian cells, for examples see (Chun *et al.* 2007, Kaddoum *et al.* 2010, Kamiyama *et al.* 2016, Milech *et al.* 2015, Van Engelenburg & Palmer 2010). All of them are based on Cabantous *et al.* (2005), utilising saGFP in plant protoplasts ((Groß *et al.* (2011), Sommer *et al.* (2011)) for determination of protein topology, (Bionda *et al.* (2016), Bionda *et al.* (2010), Sommer *et al.* (2013))), or recently even in live multicellular organisms, like zebrafish (Nasu *et al.* 2016).

An alternative method for GFP-localisations, immuno-decoration of electron microscopy samples with gold particles, can be problematic for distinguishing subcompartments of the plastid: this technique requires the utilisation of beads attached to secondary antibodies that bind to primary antibodies, tagging the protein of interest. However, these antibodies confer a spatial distance to the actual protein. Since the compartments and membranes of the

diatom plastids are just several nm apart from each other, this is a clear drawback. By electron microscopy, only the position of the electron dense beads can be detected, the real position and orientation of the protein remains uncertain.
Therefore, the self-assembling GFP system seems to represent a promising method in order to achieve a distinct result regarding protein localisation within the compartments and membranes surrounding complex plastid. This system benefits from the strong affinity of the two GFP-fragments to reassemble and restore fluorescence as soon as they are in the same compartment. The usefulness of this method on protein localisation in secondary plastids was already shown for the apicomplexan parasite *Toxoplasma gondii*, (van Dooren *et al.* 2008). Here, preceding immune labelling approaches showed rather ambiguous results: the membrane protein *Tg*Tic20 was shown to be apparently membrane associated to the plastid of *T. gondii* (apicoplast), but the final localisation was suggested after application of the saGFP method and resulted in the localisation of the protein in the innermost membrane of the apicoplast.
Different from the *in vivo* approaches mentioned above, in diatoms the introduction of transgenes is stable, meaning that the saGFP fragments can be solidly expressed in diatoms as shown by several studies (Bullmann *et al.* 2010, Felsner *et al.* 2010, Hempel *et al.* 2009, Hempel *et al.* 2010, Lau *et al.* 2016, Lau *et al.* 2015, Moog *et al.* 2015, Vugrinec *et al.* 2011).

5.5.2 Reliability of the marker proteins

We used reference proteins and their presequences which were shown by previous studies to be reliably targeted to plastidial subcompartments of diatoms (Table 8). They were first fused to full length GFP to re-examine the cellular localisation in *P. tricornutum* (Figure 19). As expected, GFP was detected in different subcompartments associated to the plastid and in the cytosol, respectively. However, we could not clearly distinguish between the interenvelope space and the stroma. The marker protein for the IES, the SP1 (plastidial type I signal peptidase), is known to be located in thylakoids in plant protoplasts (Bullmann *et al.* 2010). Even though the SP1 homolog in *P. tricornutum* apparently shows a clear distinction from the stroma (Bullmann *et al.* 2010), we cannot fully rule out, that this protein might also be expressed in the thylakoids of *P. tricornutum*, especially since the SP1 fused to full length GFP did not show clear differences to the subcellular GFP-pattern characteristic for stroma localisation (mOEE1, Figure 19).
Furthermore, the GFP-signals, the so-called 'blob-like structures' (BLS), are assumed to be characteristic for the periplastidial space (Kilian & Kroth 2005). However, a GFP expression in the periplastidial compartment which surrounds the plastid would rather be expected to result in a more evenly distribution of GFP covering the chlorophyll autofluorescence, like it is expected for the interenvelope space. Instead we observe an accumulation of GFP molecules apparently in the centre of the plastid. These BLS were suggested to be vesicular structures within the periplastidial space tightly associated to the plastid (Kilian & Kroth 2005). This association was also confirmed with the help of electron micrographs by Flori *et al.* (2016), proposing that a continuity between the second innermost membrane of the plastid and this vesicular network exists. These vesicles might finally explain the structures which were previously described as BLS, because they appear at the same spot. However, the function as well as the synthesis process of these vesicles are still not clear. Besides, the question why proteins apparently tend to accumulate at that spot of the cell remains unsolved.
We tested the self-assembling GFP method for plastidial compartments in *P. tricornutum* by fusing the presequences of the marker proteins to both saGFP fragments (GFP1-10 and

GFP11). However, we found unexpected GFP-signals (around 8 %) with differently targeted presequences.
The high amount of cell lines, which didn't show any detectable GFP fluorescence (68 %), albeit transformed with GFP-fragments fused to the same marker protein, could be avoided by the application of two different antibiotic resistance genes contained in each of the transformation constructs. In doing so, the integration and expression of the antibiotic resistance gene contained in the transformation vector would be required for the growth of transformed cell lines. Accordingly, the probability to receive cell lines expressing both gene constructs is more likely than the utilisation of solely one antibiotic resistance gene. Another possibility to increase the transformation efficiency is the utilisation of one construct, containing two transgene cassettes to enable dual translation of both genes (first described in Hempel *et al.* (2009)). However in doing so, the size of the transformation vector would be enlarged, which also increases the possibility of DNA-strand breaks and might in return negatively influence the transformation efficiency.

5.5.3 Post-translational targeting might lead to mislocalisation of the saGFP-fragments

The sa-nr-mOEE1+PGK cell line was peculiar, because whereas one GFP-fragment was supposed to be directed to the stroma, the other fragment should be expressed into the cytosol, meaning four membranes would separate them. Nonetheless, this cell line showed clear GFP-fluorescence.
An explanation for this unexpected signal in the cytosol could be mislocation, caused by an early assemblage of the GFP-fragments due to their strong affinity. This might happen when the targeting of the proteins would take place post-translationally and not co-translationally like it was assumed for the majority of plastidial proteins in diatoms, because the N-termini of the nuclear-encoded plastidial precursor proteins consist of a signal and a transit peptide utilised for targeting (Baya & Grossmann 1991; Lang et al. 1998). Post-translational transport across the translocon situated in the ER (Sec61) is possible with the help of chaperones and a temporary loosely folding of the protein (Rapoport 2007). The folding of both GFP-fragments in the cytosol would facilitate their assemblage, which is not reversible due to the high stability of superfolder GFP. Consequently, GFP would be fully restored, but the access to the targeting domain of the protein might be restricted. This would perturb or even entirely block the targeting mechanisms for these proteins and result in a GFP location in the cytosol, independent on which marker presequence was used.
Alternatively, if the targeting domain would still be accessible and the assembled GFP-molecules would be translocated, the following mechanisms could be considered:
- In case the saGFP-fragments should reassemble prior to the targeting, the fully restored GFP-molecule would be both situated in the cytosol and in the expected compartment due to the respective targeting sequence. However, the cytosolic amount of GFP molecules might be present in a higher extent so that the (lower) GFP fluorescence in one of the plastidial compartments would not be visible.
- Another explanation could be that, after assemblage of the saGFP-fragments, one GFP-fragment is dragged along together with its counterpart across a membrane. This might be the case for sa-nr-BiP+SP1 (cER and IES marker, Figure 22C). The IES marker protein, plastidal type I signal peptidase, used in our study was a full length protein fused to GFP11 (Vugrinec *et al.* 2011) instead of the presequence alone. This is also the case for the cell line sa-nr-mOEE1+SP1 (Figure 22B), where a GFP-signal is also presumably detected in the IES. In accordance, it is possible that for unknown reasons the full length protein might have a

stronger targeting function than the presequence. This is possibly due to a proper folding and access to the targeting domain for chaperones or other accompanying proteins. A location of GFP in one of the inner compartments implies a protein translocation similar to the piggyback mechanism, which is known to occur especially in peroxisomal protein import (for reviews see (Thoms (2015), van der Klei and Veenhuis (2006))), but also in context of nuclear protein import (Freitas & Cunha 2009). This process describes an import pathway in which proteins that lack a targeting or localisation signal are co-imported into an organelle as long as they are bound to other proteins which contain the appropriate targeting sequence. Until now, the piggyback mechanism is known to be a rare process. In the case of the peroxisomal co-import of the lactate dehydrogenase subunits, it was shown that both subunits can contain a targeting signal, one of them hidden in the 3'UTR due to translational readthrough (Wanders & Waterham 2006). In our case we have two proteins which do contain signal peptides, assuming that the process of piggybacking would be rather simplified. Furthermore the affinity of the two GFP-counterparts is very high and might enhance this effect. Taken these aspects together, we have to suppose that the targeting function of the presequence might be negligible, possibly due to incomplete folding which would be a reasonable explanation since we have to keep in mind that we deal with an artificially truncated protein.
A further scenario deals with the case that the targeting process would be fulfilled prior to assemblage of the saGFP-fragments in the cytosol: in accordance to the respective protein targeting sequences, the saGFP-fragments would be located in the correct compartments. After a certain expression time, the amount of present GFP-fragment molecules in each compartment would increase and accumulate. The high amount of GFP-fragments would then also increase the probability of an assemblage of the respective GFP-counterparts due to their strong affinity to each other. Accordingly, the GFP-fragment which is originally targeted to an inner compartment would then be 'fished' by the counterpart and thus being trapped in the outermost compartment.
A similar problem for false signals caused by the strong affinity of fragmented fluorescent molecules, has been reported when using split-YFP for protein-interactions, albeit not the self-assembling, but the split system was utilised (two even halves of YFP, each coupled to potential protein interaction partners): the affinity of the fragments for assembling is so strong that it leads to signals even though the proteins of interest do not interact (Horstman *et al.* 2014).
The mislocation of saGFP could be confirmed by microscopic analyses of several transformed cell lines and their phenotypes (see Table 10). Following cell lines showed unexpected cytosolic GFP-fluorescence: sa-nr-6PGDH+6PGDH (PPS marker), sa-nr-mOEE1+PGK (stroma and cytosol marker, Figure 21), sa-nr-6PGDH+BiP (PPS and cER marker, Figure 21) and sa-nr-PGK+BiP (cytosol and cER marker, Figure 22A).
We also observed GFP-signals, which apparently were not cytosolic, but rather plastid-associated: sa-nr-mOEE1+SP1 (stroma and IES marker, Figure 22B) and sa-nr-BiP+SP1 (cER and IES marker, Figure 22C). In both cases, the cellular GFP-pattern differs from stromal autofluorescence (indicated by the white arrows), thus we assume GFP has to be located in the IES according to SP1.

5.6 Conclusions

In summary we propose different possible scenarios explaining the assembling of fragmented GFP despite their different destinations, resulting in different subcellular localisations (summarised in Table 10):

I) Targeting and import of the proteins is blocked: the saGFP-fragments reassemble in the cytosol resulting in GFP-fluorescence.
II) Targeting and import occurs, the saGFP fragments are situated in the correct compartments: after a certain expression time, one fragment is getting fished by the other counterpart and thus is trapped in an outer compartment surrounding the plastid.
III) The saGFP-fragments reassemble in the cytosol, targeting and import occurs: one GFP-counterpart is dragged along with the other part which is targeted into an inner compartment ('piggyback'). One of the marker proteins is preferred.
IV) The saGFP-fragments reassemble in the cytosol, targeting and import occurs: a certain amount of assembled GFP is targeted to the plastid (both of the marker proteins might be considered). However, the process of assemblage in the cytosol might be faster than the targeting process. Thus the cytosolic GFP-fluorescence would possibly overlay the lower plastidial fluorescence.
All of these scenarios require a post-translational folding and/or import mechanism. Based on our study and reports on the same or similar methods, we suggest the following approaches if utilising the self-assembling GFP system: a) the application of full length sequences of verified marker proteins, b) utilisation of another GFP instead of the superfolder GFP to avoid irreversible assemblage of the fragments, c) utilisation of another switchable promoter sequence in order to prevent residual gene expression, and d) the utilisation of a different transformation system, for instance bacterial conjugation (Karas *et al.* 2015) to exclude random insertion loci in the host genome which might affect the protein expression pattern.
Taking all these aspects into account, there might be reason for applying the self-assembling GFP system. Nevertheless this system has to be treated with caution and we need to consider carefully the extent of required control experiments. The risk of protein mislocation is concerning, especially regarding the distinction of protein locations for subcompartments surrounding the plastid in diatoms. For future research in correlation with this system in any case, controls are of increasing importance and absolutely indispensable.

Table 10: List of *P. tricornutum* transformant cell lines expressing full length GFP or saGFP showing a GFP-phenotype classified as 'cytosol' or 'plastid' with the putative descriptive mechanism. See text for details.

Cell line name	gene expression product	expected subcellular GFP-phenotype	subcellular GFP-phenotype	possible mechanism
lhcf1-GFP	expression of full length GFP under control of the lhcf1 promoter	cytosol	cytosol	as expected
nr-GFP	expression of full length GFP under control of the nr promoter	cytosol	cytosol	as expected
sa-nr-PGK +PGK	expression of self-assembling GFP fragments under control of the nr promoter: PGK-GFP1-10 (cytosol marker) + PGK-GFP11 (cytosol marker)	cytosol	cytosol	as expected I
sa-nr-6PGDH +6PGDH	expression of self-assembling GFP fragments under control of the nr promoter: 6PGDH-GFP1-10 (PPS marker) + 6PGDH-GFP11 (PPS marker)	plastid	cytosol	I
sa-nr-mOEE1 +PGK	expression of self-assembling GFP fragments under control of the nr promoter: mOEE1-GFP1-10 (stroma marker) + PGK-GFP11 (cytosol marker)	no fluorescence	cytosol	I/II/IV ?
sa-nr-6PGDH +BiP	expression of self-assembling GFP fragments under control of the nr promoter: 6PGDH-GFP1-10 (PPS marker) + BiP-GFP11 (cER marker)	no fluorescence	cytosol	I/IV ?
sa-nr-PGK+BiP	expression of self-assembling GFP fragments under control of the nr promoter: PGK-GFP1-10 (cytosol marker) + BiP-GFP11 (cER marker)	no fluorescence	cytosol / plastid?	I/II/IV ?
sa-nr-mOEE1 +SP1	expression of self-assembling GFP fragments under control of the nr promoter: mOEE1-GFP1-10 (stroma marker) + SP1-GFP11 (IES marker)	no fluorescence	plastid	II/III ?
sa-nr-BiP+SP1	expression of self-assembling GFP fragments under control of the nr promoter: BiP-GFP1-10 (cER marker) + SP1-GFP11 (IES marker)	no fluorescence	plastid	III

lhcf1 = chlorophyll *a/c*-binding light harvesting complex protein; nr = nitrate reductase; sa = self-assembling; PPS = periplastidial space; cER = chloroplast ER; IES = interenvelope space

Funding

This work was supported by the Universität Konstanz, Deutsche Forschungsgemeinschaft DFG/SFB969 TPA4 and the Graduate School for Chemical Biology (KoRS-CB).

Acknowledgements

We thank D. Ballert, C. Río Bártulos, D. Bova, D. Ewe, S. Vugrinec, V. Spegg, R. Fahrner, M. Halder for experimental help and the FlowKon Core Facility, Universität Konstanz, for access to the flow cytometer. We also like to thank the Bioimaging Center (BIC), University of Konstanz, for access to the imaging core facilities

6 Rapid induction of GFP expression by the nitrate reductase promoter in the diatom *Phaeodactylum tricornutum*

Lili Chu[1*], Daniela Ewe[1†], Carolina Río Bártulos[1], Peter G. Kroth[1], Ansgar Gruber[1]

[1] Fachbereich Biologie, Universität Konstanz, 78457 Konstanz, Germany

[†] Present address: Algatech Centre, Microbiological Institute, CAS, 37981 Třeboň, Czech Republic

*Author for correspondence: lili.chu@uni-konstanz.de

Chu L, Ewe D, Río Bártulos C, Kroth PG & Gruber A (2016a): Rapid induction of GFP expression by the nitrate reductase promoter in the diatom *Phaeodactylum tricornutum*. PeerJ 4, e2344.

doi: 10.7717/peerj.2344

6.1 Abstract

An essential prerequisite for a controlled transgene expression is the choice of a suitable promoter. In the model diatom *Phaeodactylum tricornutum*, mainly two endogenous promoters are currently used for genetic transformation: the light dependent lhcf1 (derived from the gene encoding a fucoxanthin chlorophyll *a*/*c* binding protein) and the nitrate-driven nr promoter (derived from the gene for a nitrate reductase). In this study, we investigated the time dependent expression of the green fluorescence protein (GFP) reporter under control of the nitrate reductase promoter in independently genetically transformed *P. tricornutum* cell lines following induction of expression by change of the nitrogen source in the medium via flow cytometry, microscopy and western blotting. In all investigated cell lines, GFP fluorescence started to increase 1 h after change of the medium, the fastest increase rates were observed between 2 and 3 h. Fluorescence continued to increase slightly for up to 4 h even after transfer of the cells to ammonium medium after 24 h. The subsequent decrease of GFP fluorescence was much slower than the increase, probably due to the stability of GFP. The investigation of several cell lines transformed with nr based constructs revealed that, also in the absence of nitrate, the promoter may show residual activity. Furthermore, we observed a strong variation of gene expression between independent cell lines, emphasising the importance of a thorough characterisation of genetically modified cell lines and their individual expression patterns.

6.2 Introduction

Diatoms play a significant role in almost all aquatic ecosystems (Armbrust 2009, Falkowski & Oliver 2007). Besides their ecological and biogeochemical importance, diatoms also gained large interest with respect to potential biotechnological or biopharmaceutical exploitation (Hamilton *et al.* 2015, Hempel *et al.* 2011a, Hempel *et al.* 2011b, Hempel & Maier 2012, Roesle *et al.* 2014, Vanier *et al.* 2015). Diatoms are interesting biotechnological targets because of their easy and cheap culturing requirements, asexual reproduction and fast growth, resulting in efficient biomass production (Bozarth *et al.* 2009, Kilian & Kroth 2006, Kroth 2007a). An important prerequisite for biotechnological applications is the availability of protocols for genetic transformation of different diatom species. Meanwhile, a variety of techniques for the introduction of foreign DNA into the genomes of diatoms have been established: biolistic transformation by particle bombardment (Apt *et al.* 1996, Falciatore *et al.* 1999), electroporation (Miyahara *et al.* 2013, Niu *et al.* 2012, Zhang & Hu 2013) or gene transfer via conjugating bacteria (Karas *et al.* 2015).

To regulate transgene expression, the choice of a suitable promoter is essential. Although there is some information available regarding the functionality of heterologous regulatory DNA sequences or promoters in diatoms (Brakemann *et al.* 2008, Dunahay *et al.* 1995, Falciatore *et al.* 1999, Kadono *et al.* 2015, Sakaue *et al.* 2008), all the widely utilised transformation systems for the diatoms *Phaeodactylum tricornutum* (Falciatore *et al.* 1999, Zaslavskaia *et al.* 2000), *Cylindrotheca fusiformis* (Fischer *et al.* 1999) and *Thalassiosira pseudonana* (Poulsen *et al.* 2006) employ promoters derived from the respective target organism. The availability of genome sequence data for several diatoms (Armbrust *et al.* 2004, Bowler *et al.* 2008, Lommer *et al.* 2012) allows the identification of putative promoter sequences that can be tested regarding their usability. Four different endogenous promoters are commonly used in the model organism *P. tricornutum*: the light driven lhcf1 promoter, derived from a gene encoding a chlorophyll *a*/*c*-binding light

harvesting complex protein (Apt *et al.* 1996) (formerly known as “FCPA” (Bhaya & Grossman 1993), and now referred to as “LHCF1” (Durnford 2003)), the nitrate reductase (nr) promoter (first demonstrated by Poulsen & Kröger in *C. fusiformis* (Poulsen & Kröger 2005), now also applied in *T. pseudonana* (Poulsen *et al.* 2006) and *P. tricornutum* (Hempel *et al.* 2009)), the histone h4 promoter, which is used to obtain stable levels of transcription (De Riso *et al.* 2009) and the ef2 promoter, which has been described recently as a new tool for constitutive gene expression (Seo *et al.* 2015). The lhcf1 promoter is light dependent and inactive in the dark (Nymark *et al.* 2013). Therefore, if controlled induction of the transgene expression is required, this promoter is not suitable for photoautotrophic organisms like diatoms that do not grow in darkness. The nr promoter activity in *C. fusiformis* is affected by the available nitrogen source (Poulsen & Kröger 2005). Transcripts are present during cultivation with nitrate (NO_3^-) as nitrogen source, while no nr transcripts are detected in the presence of ammonium (NH_4^+) in the medium (Poulsen & Kröger 2005). If the medium does not contain any nitrogen source, the respective transcripts accumulate without being translated (Poulsen & Kröger 2005). Accordingly, the expression of transgenes can be controlled easily by the choice of the nitrogen source added (Poulsen & Kröger 2005).
For *in vivo* localisation studies, proteins of interest are usually genetically fused to reporter genes (mostly fluorescent proteins) and then are integrated via vector DNA into the nuclear genome. In the case of *P. tricornutum*, various constructs in which selectable antibiotic markers and reporter genes are driven by the lhcf1 promoter have been reported (Apt *et al.* 1996, Poulsen & Kröger 2005, Siaut *et al.* 2007, Zaslavskaia *et al.* 2000). Since the artificially introduced vector DNA integrates randomly into the genome, each resulting transformed cell line is different, due to the random genomic position of the promoter-transgene construct, as well as possible disruption of wild type genes at the insertion site. Accordingly, the intensity of transgene expression can vary in the different transformed cell lines (Apt *et al.* 1996, Poulsen & Kröger 2005, Zaslavskaia *et al.* 2000). In this study, we chose several nr-GFP expressing cell lines of *P. tricornutum* in order to compare the promoter’s efficiency among different cell lines. We induced and inactivated the nr promoter by changing the nitrogen source in the medium, observed the development of the gene product GFP by flow cytometric analyses in comparison to a cell line expressing GFP under control of the lhcf1 promoter. Hereby, it became obvious that the cell lines differ from each other regarding the quantity of the expressed target protein. Interestingly, the nr promoter cannot be fully down regulated by the choice of the nitrogen source.

6.3 Materials & Methods

6.3.1 Culture conditions

Phaeodactylum tricornutum Bohlin (University of Texas Culture Collection, Austin, strain UTEX646) (denoted “Pt4” by DeMartino et al. (Martino et al. 2007)) was grown in artificial half-concentrated seawater (16.6 g L^{-1}, Tropic Marine, Dr. Biener GmbH, Wartenberg/Angersbach, Germany) enriched with f/2 nutrition as described in (Guillard 1975) and additionally buffered with Tris pH 8 (2 mM, half of the concentration suggested for “f-1” by (Guillard & Ryther 1962)). Growth media were supplemented with different nitrogen compounds: $NaNO_3$ (0.882 mM, medium A, corresponding to the original f/2 nutrition) or NH_4Cl (0.882 mM, medium B). Cells were grown at 18-20 °C and 75 µmol photons m^{-2} s^{-1},

either in liquid culture in Erlenmeyer flasks on a horizontal shaker (120 rpm), or on f/2 agar plates with solid media containing 1.2 % (w/v) Bacto Agar (Becton, Dickinson and Company, Le Pont de Claix, France).

6.3.2 Transformation vector and plasmid constructions

Standard cloning procedures were performed for plasmid construction (Sambrook *et al.* 1989). The *P. tricornutum* transformation vector pPha-T1 (GenBank accession number AF219942.1) (Zaslavskaia *et al.* 2000) was utilised for cloning. The EcoRV restriction site was used for StuI/eGFP insertion, giving rise to the plasmid pPha-T1-GFP (Gruber *et al.* 2007). A second plasmid was constructed based on the *P. tricornutum* transformation vector pPha-NR (GenBank accession number JN180663.1; (Stork *et al.* 2012)). The eGFP gene (Clontech, Palo Alto, CA) was cloned downstream of the nr promoter sequence using the EcoRV restriction site.

6.3.3 Biolistic transformation

Cells were transformed using the Biolistic PDS-1000/He Particle Delivery System (Bio-Rad, Hercules, California, USA) fitted with 1350 psi rupture discs as described in (Kroth 2007b). After transformation, cells were allowed to recover for 24 hours before being plated onto f/2 medium containing 75 µg/mL zeocin (Invitrogen, Molecular Probes, Eugene, USA) for selection. The plates were incubated at 22 °C under constant illumination (75 µmol photons m^{-2} s^{-1}). One of the resulting transformed cell lines of the transformation with the plasmid pPha-T1-GFP and six of the transformed cell lines expressing GFP under control of the nitrate reductase promoter were chosen based on GFP expression screened by flow cytometry (see below) as well as fluorescence microscopy (see below).

6.3.4 Determination of cell density

The cell numbers were determined using the Multisizer 3 (Beckman Coulter, Brea, CA, USA) as described in (Rottberger *et al.* 2013b), over a time period of 11 days. Samples were taken once per day.

6.3.5 Western blot analyses and SDS-PAGE

For western blot analyses, selected *P. tricornutum* cell lines were grown at 18 °C and 75 µmol photons m^{-2} s^{-1} in continuous light. Wild type cells, the lhcf1-GFP and nr-GFP transformant cell lines were grown in medium A containing $NaNO_3$, and in medium B containing NH_4Cl. Cells were harvested during exponential phase by centrifugation (3000 g, 10 min, 4 °C) and the pellets were resuspended in 1 mL lysis buffer containing protease inhibitor “complete EDTA-free” (Roche, Mannheim, Germany), 50 mM Tris HCl pH 8, 1 mM EDTA, and 1 % (w/v) SDS. A mixture of glass beads (0.1-1 mm diameter) was added and cells were homogenised in a MP FastPrep-24™ 5G (MP Biomedicals, Santa Ana, CA, USA) at a speed of 6 m/s for 4 times 20 sec, with 1 min breaks on ice in between the homogenisation pulses. Samples were centrifuged again (20 000 g, 30 min, 4 °C), and supernatant was transferred into a new tube and further used for SDS-PAGE. Total protein concentration was determined using the 660 nm Pierce Protein Assay (Thermo Scientific, Rockford, USA) and the spectrophotometer Ultrospec™ 8000 (GE Healthcare, Little Chalfont, UK). Each well of the gels was loaded with 3 µg of each protein extract. Proteins were separated by SDS-PAGE in 12 % acrylamide gel (Laemmli 1970) and transferred electrophoretically onto a nitrocellulose membrane (Amersham Protran 0.1 NC, GE Healthcare), using

Chromatography paper (Whatman™ 3MM Chr, GE Healthcare) and a Trans-Blot Turbo (Bio-Rad) at 1.3 A and 25 V for 12 min. Page Ruler Prestained Protein Ladder (Thermo Scientific, Schwerte, Germany), primary antibody α-GFP (catalog number A-6455, Invitrogen), diluted 1:10000, and secondary antibody α-Rabbit IgG (catalog number A0545, Sigma Aldrich, Munich, Germany), diluted 1:20000, were utilised. Roti®-Block, Roti®-Lumin plus (Carl Roth GmbH & Co. KG, Karlsruhe, Germany) and InstantBlue™ (Expedeon, San Diego, CA, USA) were applied as described in their manuals. Immunodetection was performed with the Odyssey® Fc Imaging System (LI-COR Biosciences, Lincoln, NE, USA).

6.3.6 Induction of nitrate reductase promoter

Before fluorescence was measured, the cell lines were kept for several days in liquid ammonium-containing medium in multi-well plates under continuous light illumination at 75 µmol photons m^{-2} s^{-1} and 18 C to make sure that the nr promoter was switched off and no GFP is visible in the nr-GFP transformed cell lines. For the measurements, cultures were inoculated in fresh ammonium-containing medium and were harvested after 3 days during exponential phase by centrifugation (3000 g, 10 min). The fluorescence measurements (BD FACSCalibur, BD Biosciences, CA, USA) were started after the cells were washed once by resuspension in fresh nitrate medium and another centrifugation step (time point 0 h). The samples were taken after different time intervals to observe changes in fluorescence intensity. After 24 hours in nitrate-containing medium, the nr promoter was switched off again by harvesting, washing and transferring the cells back into ammonium-medium. The cells were observed for another 10 days (264 hours in total) with samples taken at the indicated intervals. Microscopic analyses were performed in parallel to verify potential GFP fluorescence.

6.3.7 Flow cytometry

Flow cytometric analyses were performed using the flow cytometer BD FACSCalibur (BD Biosciences) and the Software BD CellQuestPro (BD Biosciences). For the detection of GFP, we used the 488 nm laser for excitation and the FL1 detector with a 530/30 BP filter for detection. For the detection of red fluorescence, we utilised in parallel the FL3 detector with a 650 LP filter. The emission intensities of 100 000 cells per sample (triggered by side scatter) were collected and ungated median fluorescence intensities of each population were collected for the subsequent analyses. Dot plots and histograms were created and analysed using the Single Cell Analysis Software FlowJo (Tree Star, Inc. Ashland, OR). The raw data exported from the flow cytometer BD FACSCalibur and the FlowJo workspace applied for data analyses and figure preparation are contained in Supplemental Data S1.
Wild type cell cultures were used as negative control and a lhcf1-GFP transformed cell line as reference for a GFP-expressing cell line under the control of a nitrate-independent promoter.

6.3.8 Fluorescence Microscopy

Cellular expression of GFP fusion proteins was analysed with an epifluorescence microscope Olympus BX51 (Olympus Europe, Hamburg, Germany), a Zeiss AxioCam MRm digital camera (Carl Zeiss, Oberkochen, Germany) and an Olympus PLN 40× objective (Olympus Europe, Hamburg, Germany). Image processing was conducted using the Software AxioVision Rel. 4.7 (Carl Zeiss, Oberkochen, Germany).

6.4 Results

We generated genetically transformed cell lines of *Phaeodactylum tricornutum* expressing eGFP, under the control of the nr promoter. We selected six nr strains with varying intensities of GFP fluorescence using fluorescence microscopy. Additionally, as a reference for a nitrate-independent promoter, we chose a GFP-expressing cell line under control of the lhcf1 promoter. (Table 11)

Table 11: Wild type and GFP expressing cell lines of *P. tricornutum* used for time-dependent GFP-fluorescence measurements by flow cytometry and fluorescence microscopy. Protein IDs refer to wild type strain CCMP632 (denoted "Pt1" by (De Martino *et al.* 2007)), which was sequenced by the U.S. Department of Energy Joint Genome Institute (http://genome.jgi.doe.gov/Phatr2/Phatr2.home.html) (Bowler et al. 2008).

Cell line name	Description
wt	wild type *Phaeodactylum tricornutum* Bohlin (University of Texas Culture Collection, Austin, strain UTEX646) (denoted "Pt4" by DeMartino et al. (Martino *et al.* 2007))
lhcf1-GFP	*P. triconutum* UTEX646, genetically transformed with the pPha-T1-GFP plasmid (Gruber *et al.* 2007), derived from pPha-T1 (GenBank AF219942.1, (Zaslavskaia *et al.* 2000)), which contains the 442 bp 5'-flanking region of the *P. triconutum* lhcf1 gene (equivalent protein in Pt1, ID 18049) (Apt *et al.* 1996) as promoter
nr-GFP_3 **nr-GFP_4** **nr-GFP_5** **nr-GFP_6** **nr-GFP_9** **nr-GFP_10**	*P. triconutum* UTEX646, genetically transformed cell lines expressing the eGFP gene with a construct derived from the pPha-NR vector (GenBank JN180663.1, (Stork *et al.* 2012)), which contains the 422 bp 5'-flanking region of the *P. triconutum* nr gene (equivalent protein in Pt1, ID 54983) (Hempel *et al.* 2009) as promoter (see text for details)

To determine the velocity of GFP synthesis induced by a change of nitrogen source in the medium, we set up a time dependent experiment with sampling at different time points (Figure 23). We determined the green and red fluorescence intensities with the flow cytometer BD FACSCalibur in regular time intervals after induction of the nr promoter by transferring the cells from ammonium- to nitrate-medium (Figure S 23 and Table S 8 + Table S 9).

Dot plots of green versus red channel signals show that the signal level in the red channel is similar between GFP expressing cell lines and wild type cell lines. This means that the chosen filter sets efficiently separated the green GFP fluorescence from the red chlorophyll autofluorescence and that the signal in the green channel did not result from spill over of the chlorophyll autofluorescence (Figure S 24 and Table S 9).

Furthermore, we checked for possible physiological effects of the medium changes on *P. tricornutum* wild type cells by PAM (pulse-amplitude modulation) fluorometry and by measuring the pH of the medium (Table S 10 + Table S 11). The pH of the medium changed only slightly and remained around pH 8 throughout the whole experiment (Table S 10). Ratios of variable fluorescence to maximal fluorescence (Fv/Fm) changed slightly between nitrate and ammonium media, while non-photochemical quenching (NPQ) stayed constant (Table S 11). Growth measurements confirmed that cell growth was similar between all investigated cell lines and independent from the nitrogen source present in the media. All

cultures entered the stationary growth phase after ~120 h (Figure S 25). The red autofluorescence intensities remained on a similar level in all of the cell lines until ~168 h of the experiment and dropped slightly towards the end of the experiment (Table S 9), indicating changes in the pigment content while the cells enter the stationary growth phase.

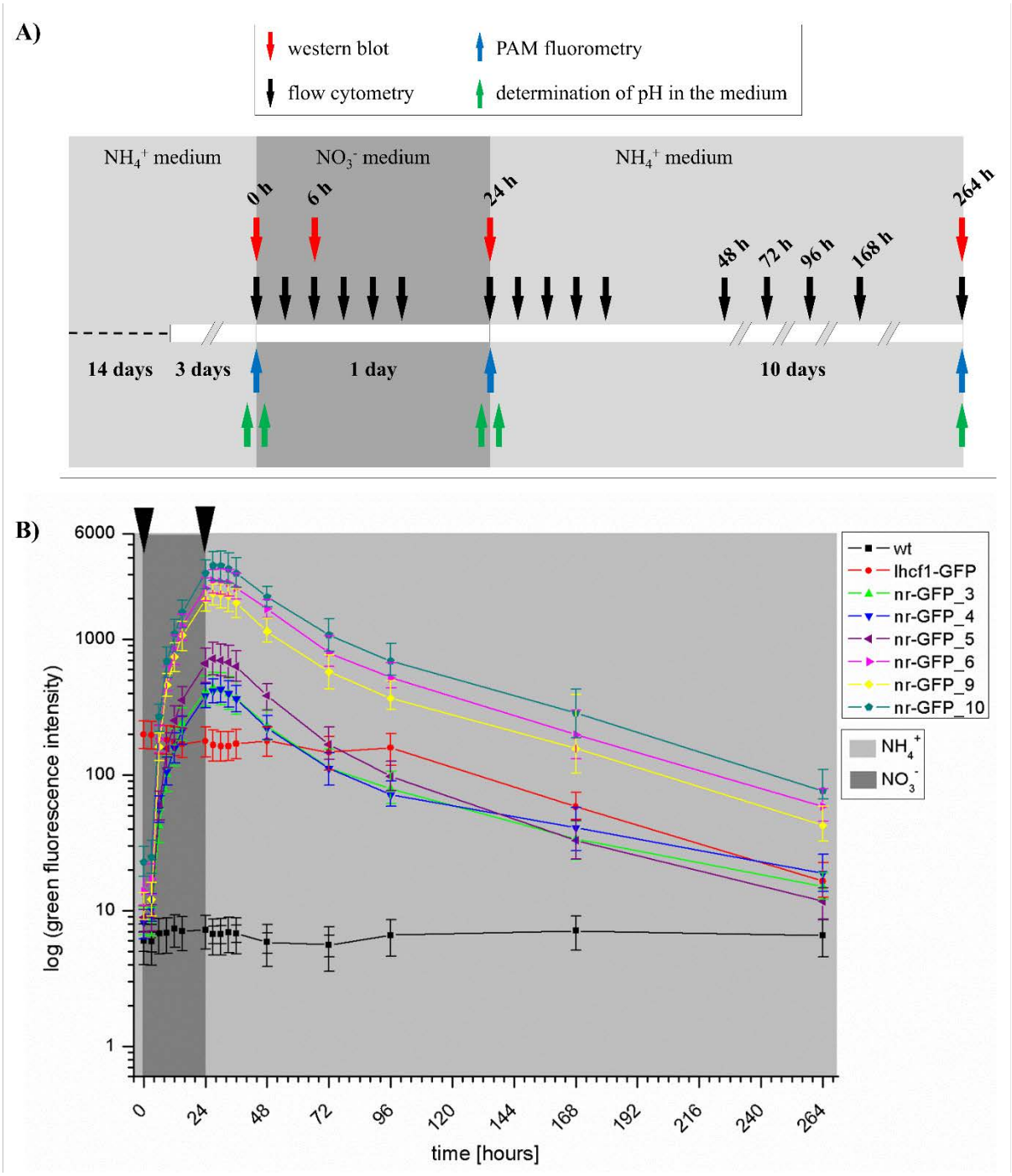


Figure 23: A) Time scale of the experimental setup. Arrows indicate the time points of sampling. Western blots and flow cytometry was performed using *P. tricornutum* wild type cell line and GFP-expressing cell lines. PAM (Pulse amplitude modulation) fluorometry of cell cultures and pH of the media was determined using wild type cell lines. h = hours. B) Median green fluorescence intensities determined for *P. tricornutum* wild type and GFP-expressing cell lines. Intensities were determined using the flow cytometer BD FACSCalibur at indicated time intervals. Arrows indicate washing steps and medium change. Error bars represent the interquartile range (IQR).

At the beginning of the induction experiment, the green fluorescence signals of all genetically transformed cell lines were higher than in the wild type cells (Figure S 23 and Table S 8A). In case of the nr-GFP cell lines, the green fluorescence signal indicated the presence of a certain GFP-level in the cells even during cultivation in ammonium-medium (Figure 23 and Table S 8), which could also be shown for several cell lines via western blots (Figure S 26). In case of the lhcf1-GFP reference cell line, the level of green fluorescence did not change throughout the induction experiment, with the exception of a decrease in fluorescence first measured after 96 hours of the experiment (Figure 23 and Table S 8B). Simultaneously decreasing autofluorescence signals indicated that this might be related to ageing of the culture (Table S 9B).

Upon the start of the experiment, after ammonium acclimated cells had been transferred into nitrate-medium, green fluorescence increased in all nr-GFP cell lines throughout the cultivation time, with the highest rates of increase between 3 and 6 hours after the change of the medium (Table S 8A). The GFP signal was not detectable microscopically until about 6 h after the medium change. Throughout the rest of the experiment, microscopic GFP fluorescence detection subjectively did not change, although the flow cytometer indicated an increase of GFP fluorescence intensity of up to twelve-fold between 6 and 24 hours after induction (Figure S 27 and Table S 8A).

After all cell lines had been incubated for 24 hours in nitrate-medium, they were transferred back into ammonium-containing medium and subsequent measurements showed decreasing green fluorescence in nearly all nr-GFP cell lines 3-6 hours after the medium change (27-30 h, Table S 8B). The rate of decrease of green fluorescence was much slower than the rate of increase. Even after 264 hours, the green fluorescence intensities in the nr-GFP cell lines were higher than at the beginning of the experiment. Due to the age of the cultures, the measurements after 168 and 264 hours generally show low fluorescence in both the green and red channel (including the positive control lhcf1-GFP, Table S 8B + Table S 9B). Since the chlorophyll autofluorescence of the cells was also lower than at the beginning of the measurements, the low green fluorescence signal does not specifically imply a decrease in the steady state levels of fluorescing GFP, but could also reflect the state of the cells in late stationary growth phase (Table S 8 + Table S 9).

An independent repetition of the fluorescence measurements with a higher temporal resolution (Figure 24A), confirmed that GFP fluorescence was not immediately decreasing after transferring the nr cell lines back into ammonium-medium (Figure 24B). To the contrary, we could still observe a slight increase of green fluorescence for the next 3 hours (27 h), albeit the rate of increase was immediately lower after the medium change. A first decrease of GFP intensity could be observed after 3 hours (27 h) in the cell lines nr-GFP_5, _6 and _9. However, from this time point, the intensities remained constant in all of the nr cell lines before finally decreasing after 7 hours (31 h).

In order to assess the range of fluorescence intensities obtained in cell lines resulting from independent genetic transformation events with identical constructs, we repeated transformation of wild type *P. tricornutum* with the lhcf1-GFP construct and screened all resulting zeocin resistant colonies for GFP expression. The genetic transformation resulted in 45 cell lines, most of these showed detectable GFP fluorescence (Figure S 28). The majority of the cell lines showed lower relative GFP fluorescence values than the nr-GFP cell lines (compare to Figure 23 and Figure 24).

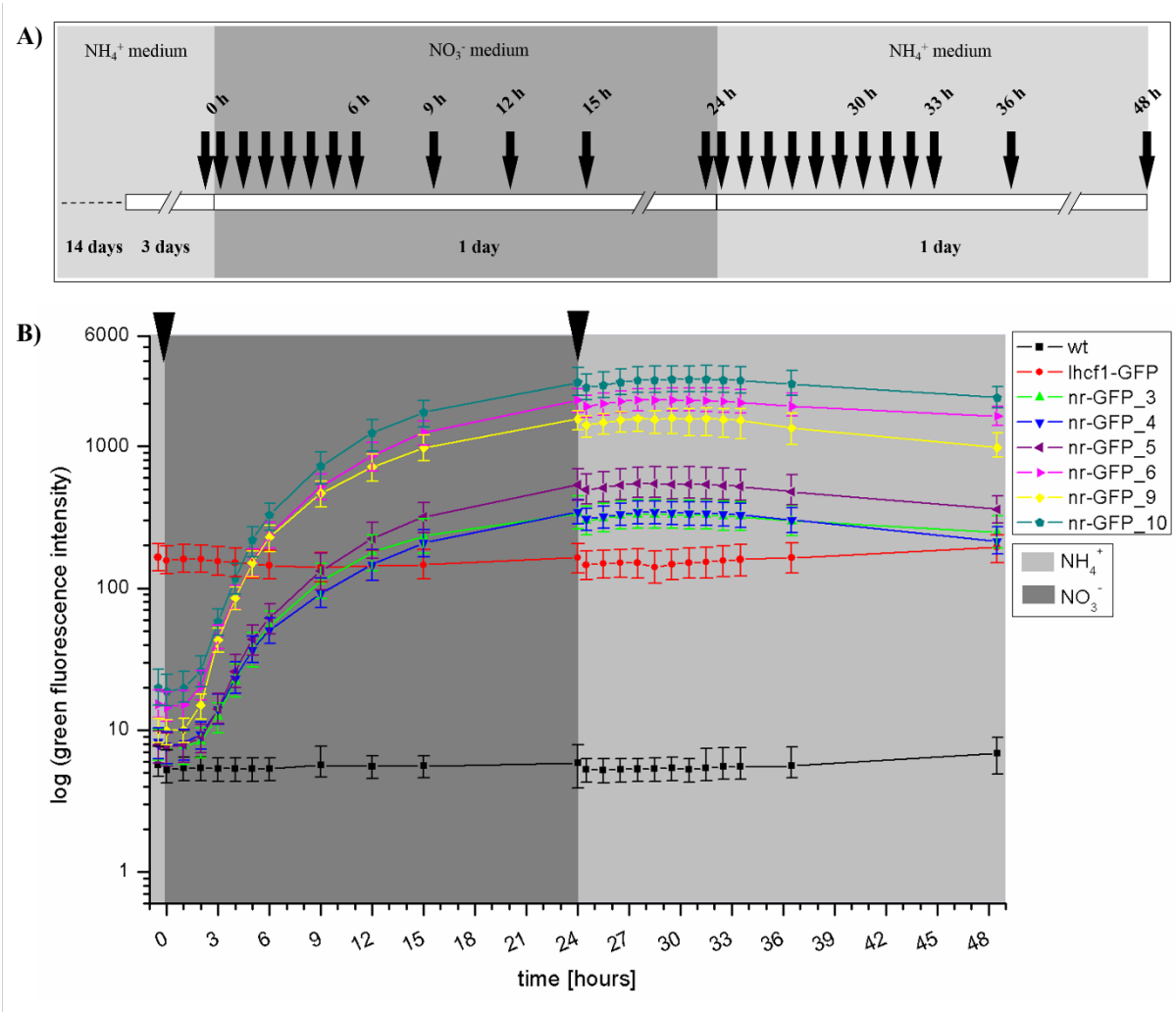


Figure 24: A) Time scale of flow cytometry sampling using *P. tricornutum* wild type cell line and GFP-expressing cell lines. h = hours. B) Median green fluorescence intensities determined for *P. tricornutum* wild type and GFP-expressing cell lines. Intensities were determined using the flow cytometer BD FACSCalibur at indicated time intervals. Arrows indicate washing steps and medium change. Error bars represent the interquartile range (IQR).

6.5 Discussion

Transgene expression in the diatom *Phaeodactylum tricornutum* in current biotechnological approaches is usually based on the lhcf1 or nr promoters (Apt *et al.* 1996, Hempel *et al.* 2009, Zaslavskaia *et al.* 2000). In wild type cells, the lhcf1 promoter drives the expression of a light harvesting complex protein, a member of a large multigene family (Durnford 2003, Sturm *et al.* 2013). Light dependency of transcription was confirmed in *P. tricornutum* for LHCF2 (FCPB) (Lepetit *et al.* 2013, Russo *et al.* 2015, Sturm *et al.* 2013) and further studies on gene expression showed similar patterns for LHCF2 (Protein ID 25172) and LHCF1 (Protein ID 18049) (Maheswari *et al.* 2010, Maheswari *et al.* 2009, Nymark *et al.* 2009, Nymark *et al.* 2013, Valle *et al.* 2014). In our study the lhcf1-GFP *P. tricornutum* strain served as a nitrate-independent reference strain. Using flow cytometry, GFP as reporter gene could be traced accurately, this approach consequently enables the characterisation of additional promoters for basic or applied research.

The nr promoter activity is inducible by a change of the nitrogen source in the media (Poulsen & Kröger 2005). However, we also observed a slight presence of GFP in all of the nr-GFP cell lines before the promoter was induced by nitrate. This implies that the nr promoter is 'leaky' in the presence of ammonium and a certain background level of the gene product is maintained.

Our data indicate that the GFP fluorescence in most of the investigated nr-GFP cell lines is stronger compared to the lhcf1-GFP cell line (compare Figure 23 + Figure 24 + Figure S 28). This strength of the nr promoter might be useful for the overexpression of trans-genes, when high protein levels are required. However, excessive protein expression sometimes comes with the risk of disturbing cellular homeostasis (in biotechnological applications) or the experiments themselves (for instance in protein localisation studies). In such situations, selection of transformed cell lines with low expression levels or careful adjustments of the ammonium:nitrate ratio in the media would be required.

During cultivation in nitrate free ammonium containing medium, the nr promoter was reported to be inactive (Poulsen & Kröger 2005). The details of the process of inactivation are not entirely clear. It is suggested that ammonium addition inhibits the uptake of nitrate by the cells (Cresswell & Syrett 1979). Our data indicate, that inactivation of the promoter is a slow process, since we could observe a first decrease of GFP intensity in several nr cell lines only 3 h after change of the medium. The residual increase of GFP fluorescence within the first 3 h could be caused by translation of remaining GFP transcripts in the cell. Between 3 and 7 h, the GFP amount appears to stay on a static level, indicating that no additional GFP molecules are synthesised. The subsequent decrease of GFP amount may either occur by dilution of the GFP amount per cell during every cell division, or by degradation of the GFP.

Since the cells were kept in ammonium-medium for several days, they already reached the stationary growth phase. It was demonstrated that the addition of any particular nutrient alone does not lead to a positive growth response (Rottberger *et al.* 2013a), however, a combination of added nutrients might be sufficient to trigger cell division. Hence, if depletion of GFP in a culture is required, we suggest transferring the cells into fresh media containing ammonium. Alternatively, the use of a destabilised green fluorescent protein might be helpful (see (Dantuma *et al.* 2000, Houser *et al.* 2012, Kitsera *et al.* 2007, Li *et al.* 1998)), especially to track rapid changes in gene expression and protein turnover.

Complete nitrogen deprivation in *P. tricornutum* reportedly results in a decrease of photosynthetic capacity and chlorophyll content, and in a simultaneous accumulation of neutral lipids (Alipanah *et al.* 2015, Valenzuela *et al.* 2012, Yu *et al.* 2009). Also in our experiments the intensity of the red plastid autofluorescence decreased not before 96 hours in ammonium-medium (Table S 9B). We also observed accumulation of lipid droplets within the cells towards the end of the cultivation interval (Figure S 27), which is a typical phenotype for nitrogen starvation (Ge *et al.* 2014, Yang *et al.* 2013).

The characterisation of several nr-GFP and lhcf1-GFP cell lines revealed that transgene expression does not necessarily lead to the same expression pattern among the cell lines, although they were transformed with identical DNA. In addition to external factors, like medium, light and growth phase, each transgene promoter activity also depends on the position of the vector DNA insertion within the host genome. It has been shown for cells transformed by particle bombardment that vector DNA integration varies in the number and position of insertions (Zaslavskaia *et al.* 2000) leading to so far unpredictable variations in transgene expression (Zaslavskaia et al. 2000, this study). In the future, this might be overcome by using targeted gene insertions via TALEN (Daboussi *et al.* 2014) or by episomal plasmid delivery via bacterial conjugation (Karas *et al.* 2015).

Funding
This work was supported by the Universität Konstanz, and by grants of the Deutsche Forschungsgemeinschaft (DFG/SFB969, TPA4) and the German-Israel Science Foundation (GIF) to P.G.K.).

Acknowledgements
We like to thank D. Ballert; T. Sonntag, E. Xie for experimental help, B. Lepetit for helpful discussions, the FlowKon Core Facility, Universität Konstanz, for access to the flow cytometer FACSCalibur and S. Zauner, Philipps-University of Marburg, for providing the pPha-NR vector.

7 General Discussion

7.1 Nucleotide transporters

So far, nucleotide transporter proteins (NTTs) have been identified and characterised in bacteria, microsporidia and in plastid-harbouring eukaryotes. They are most likely located in the inner envelope membrane of the plastid in plants and green algae or in the plasma membranes of bacteria and microsporidia, respectively, as well as in membranes surrounding the microsporidian remnant mitochondrion, called mitosome (Ast *et al.* 2009, Haferkamp *et al.* 2004, Haferkamp *et al.* 2006b, Heinz *et al.* 2014, Linka *et al.* 2003, Möhlmann *et al.* 1998, Neuhaus *et al.* 1997, Schmitz-Esser *et al.* 2004, Tjaden *et al.* 1998a, Tjaden *et al.* 1998b, Tjaden 1999, Tsaousis *et al.* 2008, Tyra *et al.* 2007).
In *Rickettsiales* and *Chlamydiales* up to five NTT isoforms were identified (Audia & Winkler 2006, Haferkamp *et al.* 2004, Haferkamp *et al.* 2006b). They contain amino acid motifs conserved among known NTTs which are essential for nucleotide transport. Like in plant plastids, also *Protochlamydia amoebophila* harbours one NTT (PamNTT1) capable of ATP/ADP exchange (Schmitz-Esser *et al.* 2004). However, *Pam*NTT2 transport RNA nucleotides and *Pam*NTT4 is responsible for NAD^+-uptake in exchange for ADP (Haferkamp *et al.* 2004), whereas *Pam*NTT3 and *Pam*NTT5 perform ATP, GTP and UTP import coupled to a proton-gradient (Haferkamp *et al.* 2006b). Hence, NTTs in chlamydia apparently combine both, energy provision to the cells and net nucleotide uptake.
Microsporidia are obligate intracellular living parasites and harbour up to four NTTs. In *Encephalitozoan cuniculi*, the NTTs apparently serve for ATP acquisition from the host cell to the parasite: one NTT was identified to be located in the mitosome membrane, whereas the other three isoforms are situated in the plasma membrane (Tsaousis *et al.* 2008). All of the four NTTs in *Trachipleistophora hominis* apparently reside in the plasma membrane, where they function as transporters for net nucleotide uptake in order to exploit the host cell to compensate their own reduced metabolism (Heinz *et al.* 2014).
Primary plastids from higher plants and algae possess one to two NTT isoforms which are well characterised: they facilitate the counter-exchange for ATP/ADP and play a crucial role in energy provision to the organelle (Linka *et al.* 2003, Winkler & Neuhaus 1999).
The nucleotide metabolism in diatoms, harbouring complex plastids, differs from that of plants and green algae, with primary plastids (Ast *et al.* 2009): (i) the nucleotide *de novo* synthesis takes place in the cytosol instead of in the plastid, (ii) the amount of putative nucleus-encoded NTTs is higher (six isoforms in the pennate diatom *P. tricornutum* and eight in the centric diatom *T. pseudonana*), and (iii) apparently the biochemical properties of diatom NTTs facilitate both energy provision and net nucleotide uptake to the plastid. NTT1 acts as a proton-dependent adenine nucleotide symporter, whereas NTT2 is capable of transporting all triphosphorylated nucleotides including their deoxy forms (Ast *et al.* 2009). Thus, diatom NTTs seem to have features of bacterial and of plant NTTs. Situated in the innermost membrane of the plastid, NTT1 and NTT2 would represent important component for the net nucleotide uptake to the diatom plastid. However, for a complete nucleotide shuttling system across the four plastid envelope membranes, additional components would be needed. This lack of knowledge inspired further investigations of the other NTT isoforms, to gain further insights into nucleotide trafficking in diatom cells.

7.2 The physiological function of the diatom *Pt*NTT5 and *Tp*NTT3

A detailed biochemical investigation on the possible substrates of the NTT isoform 5 of *P. tricornutum* (Chu *et al.* 2016b), revealed that the transport properties of *Pt*NTT5 differ from those of diatom NTT1 and NTT2 (Table 13). *Pt*NTT5 prefers purine nucleotides, including tri- , di- and monophosphorylated nucleotides and even the deoxy-form of ATP and GTP. The transporter acts in a counter exchange mode. Thus, facilitating the antiport of such a broad substrate spectrum and situated in the outermost membrane of the diatoms plastid, *Pt*NTT5 could broaden the nucleotide composition inside the ER lumen from adenine nucleotides to guanine nucleotides, dATP and dGTP. Furthermore, it reveals that the outermost membrane represents a selective barrier for nucleotides.
Therefore, diverse purine nucleotides and their deoxy-forms would then have crossed the first barrier on their way into the stroma, where they finally enter DNA and RNA synthesis via interaction of NTT1 and NTT2 in the innermost membrane. Thus *Pt*NTT5, in cooperation with NTT1 and NTT2, could also contribute to the net nucleotide uptake of the plastid postulated by (Ast *et al.* 2009).
The analysis of the substrates of *Tp*NTT3, which is the only NTT isoform from the centric diatom *T. pseudonana* that has no homologous NTT in *P. tricornutum*, also revealed a broad spectrum: tri-, di, monophosphorylated purine nucleotides next to deoxy forms represent substrates of *Tp*NTT3 and interestingly, also cyclic adenine nucleotide monophosphate is among its preferred substrates. Cyclic nucleotides represent important second messengers in the signalling pathways of eukaryotes. In the diatom *Cylindrotheca fusiformis* it was postulated that cyclic nucleotides might play a role in silicon metabolism (Aline *et al.* 1984, Borowitzka & Volcani 1977), since cAMP and cGMP concentrations increased during recovery from silicon deprived conditions. A correlation between *Tp*NTT3 and silicon metabolism could explain, why a homologue of *Tp*NTT3 is not found in *P. tricornutum* (Ast *et al.* 2009), because this diatom is not depending on silica cell wall formation and uses silicic acids only facultatively (De Martino *et al.* 2007). The GFP-localisation studies of *Tp*NTT3 (Chapter 3) could not reveal the exact location of this transporter. However, the subcellular GFP pattern in *T. pseudonana* showed high similarities to results of localisation studies of silaffins, highly phosphorylated proteins involved in the silica-based formation of the cell wall (Poulsen *et al.* 2013, Sheppard *et al.* 2009). Silaffins lacking targeting information for further transport to the silica deposition vesicles (SDV) accumulate in a specific region in or tightly associated to the ER (Poulsen *et al.* 2013). This position might represent the place where phosphorylation of the silaffins occurs, since GFP-fusion of the representative silaffin kinase tpSTK1 show an identical fluorescence pattern (Sheppard *et al.* 2009) as the trapped (targeting impaired) silaffins. The tpSTK1 protein was shown to act as a serine/threonine kinase that prefers ATP as phosphate group donor (Sheppard et al., 2010). Interestingly, serine/threonine kinase are often regulated by cAMP and cGMP. Therefore, it is possible that *Tp*NTT3 is indeed located in a membrane nearby the silaffin kinases where it supplies the substrate for phosphorylation but also the factors for their regulation of enzymatic reactions. Thus, situated in a membrane enclosing the ER or accumulated at a potential sub-region at the ER, *Tp*NTT3 might represent an important component involved in the delivery of silica to the cell wall of diatoms.
It is obvious that *Pt*NTT5 and *Tp*NTT3 not only show a variable substrate specificity, when compared to NTT1 and NTT2, but also a different subcellular localisation (Figure 25). The cER lumen is the place where proteins undergo a quality check and post-translational modifications, like phosphorylation, and where they are introduced into the secretory pathway in order to fulfil their functions in other compartments (Csala *et al.* 2006, Ellgaard & Helenius 2003, Vitale *et al.* 1993). Hence, supplying energy-rich metabolites would be

crucial to stimulate these reaction, possibly NTTs, such as *Pt*NTT5 and *Tp*NTT3 might contribute to this process (in the form of (d)ATP, (d)GTP, ADP and GDP). If *Tp*NTT3 would also be situated in the ER membrane, this transporter could additionally shuttle cyclic AMP and GMP, which are important second messengers and would not be correlated to a net nucleotide uptake to supply nucleotides for DNA/RNA synthesis (see Chapter 3). Corresponding acceptors for G protein-coupled receptors (GPCR) were identified in diatoms (in *P. tricornutum* five, and in *T. pseudonana* two putative GPCR isoforms), indicating that a signalling pathway, possibly similar to the machinery in mammals, does exist in diatoms (Port *et al.* 2013). Furthermore, the presence of an N-terminal signal peptide is predicted for GPCR2 in *T. pseudonana* (data not shown) and would imply, that this receptor is present in the secretory pathway, possibly in the ER lumen, where *Tp*NTT3 would provide suitable substrates.

Taken together, both *Pt*NTT5 and *Tp*NTT3 represent NTTs which are most likely located in membranes enclosing the ER. Also, both of them regulate a quite broad substrate spectrum, indicating that a high diversity of purine nucleotides is required for several reaction steps inside the ER. The mechanism of the net adenine nucleotide loading of the ER is not entirely solved, but proteins of the mitochondrial carrier family might be the responsible transporter proteins that mediate the uptake of ATP (Csala *et al.* 2007, Hirschberg *et al.* 1998, Leroch *et al.* 2008). The biochemical features of both *Pt*NTT5 and *Tp*NTT3, would allow provision of energy-rich metabolites (in form of tri- and diphosphorylated purine nucleotides) and would in general broaden the spectrum and composition of nucleotides available in the ER lumen. From here, the nucleotides could either be further transporter via other NTT isoform, or via a vesicle shuttle system across the PPS. The nucleotides would be then available in the IES, where they could fuel NTT1 and NTT2 (adenine nucleotides and (deoxy) nucleotide triphosphates).

The remaining NTT isoforms, which have been not characterised yet (*Pt*NTT3, *Pt*NTT4, *Pt*NTT6, *Tp*NTT4-8), represent possible candidates for taking over the transport between the cER lumen and the IES, as well as the net pyrimidine nucleotide uptake.

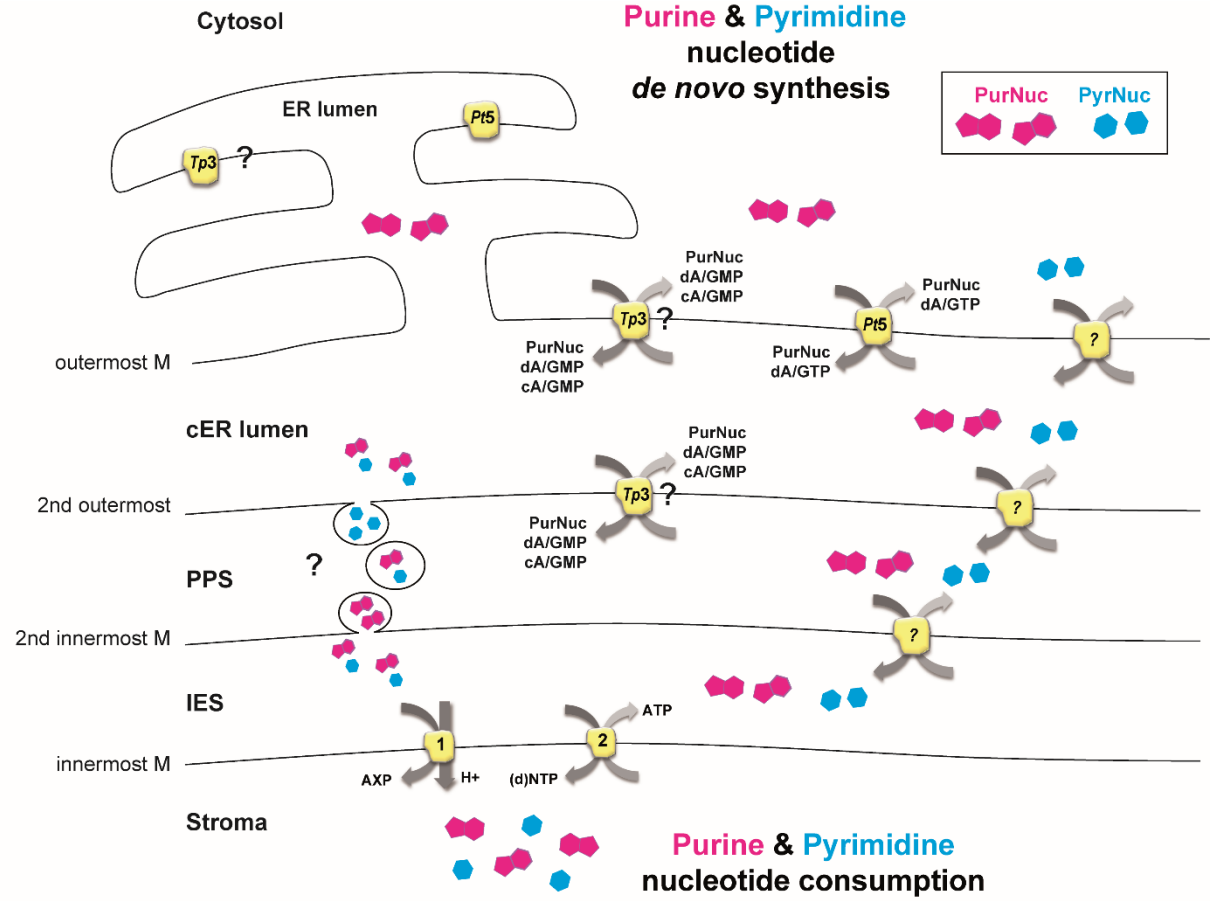


Figure *25*: Schematic illustration of the putative nucleotide metabolism in diatom plastids, based on the biochemical examination and localisation of diatom NTTs. Purine (PurNuc) and pyrimidine (PyrNuc) nucleotides are synthesised in the cytosol. PurNuc can be shuttled by *Pt*NTT5 and *Tp*NTT3, situated in the cER membrane. *Tp*NTT3 could also be situated in the 2nd outermost membrane. PurNuc and PyrNuc can be transported via NTTs of the 2nd outermost and 2nd innermost membrane or via vesicle shuttle.

7.3 What is the origin of metabolite transporter proteins in complex plastids?

Next to nucleotide translocators (NTTs), also nucleotide sugar translocators (NSTs) and phosphate translocators (PTs) are involved in the provision of energy-rich metabolites to the plastid. For plant plastids, it is known that a large number of different types of PTs are located within the inner envelope membrane, where they supply metabolites involved in the carbon cycle for the green lineage plastids (Facchinelli & Weber 2011): sugar phosphate translocators (xylulose-5 (XTPs), glucose-6 (GTPs)), phospho-enol-pyruvate translocators (PPTs) and triose phosphate translocators (TPTs). In the red lineage, PPTs and TPTs could be identified (Weber & Linka 2011), whereas GPTs and XPTs were apparently lost (Weber *et al.* 2006) and only little is known about the respective carbohydrate transport across those membranes into the plastids. Phylogenetic analyses show that the plastid translocators form a monophyletic tree, indicating that proteins from red and green algae as well as land plants apparently had a single origin (Weber *et al.* 2006). It is suggested that several plastidial PTs in secondary plastids originated from TPTs which might play an important role in the connection of the endosymbiont with the host metabolism. The presence of TPTs in complex plastids was reported in the cryptophyte *Guillardia theta* (Haferkamp *et al.* 2006a). Two different isoforms catalyse the exchange of DHAP (dihydroxyacetone phosphate) and PEP (phosphoenolpyruvate) against P_i (inorganic phosphate) and show extraordinary localisation: one is suggested to integrate into one of the two inner membranes, the other is possibly integrating into the ER membrane (cER) (Haferkamp *et al.* 2006a), which is similar to the unexpected ER membrane localisation received for *Pt*NTT5 and *Tp*NTT3.

In the diatom *P. tricornutum*, in total nine putative TPTs have been identified (Kilian & Kroth 2005, Lau *et al.* 2015, Moog *et al.* 2015). A typical bipartite targeting signal (BTS) could only be predicted for TPT2/4a/4b and a classic signal peptide was predicted for TPT1/8/10 (Moog *et al.* 2015). Among these possibly plastid-associated TPTs, only four were finally suggested to integrate into plastidial membranes. The proteins with plastid targeting signal were localised in the innermost and second outermost membrane, respectively (Moog *et al.* 2015). Among TPTs with a classic signal peptide, TPT1 was integrating in the outermost (ER) membrane, while the localisation of TPT8 was not clear (ER or PPS) and TPT10 resulted in a localisation distinctly outside the complex plastid (Moog *et al.* 2015). Accordingly, also regarding the TPTs, evident signal domains do not necessarily direct the proteins to the ER, implying that other factors may influence the targeting and localisation, as it is the case for NTTs. The localisation studies by Moog et al. (2015) are based on GFP fusion proteins (both full length GFP and saGFP), hence might be worth to be confirmed by additional detailed investigations, e.g. with reposition of the GFP and truncations. Moog et al. (2015) propose that the TPTs are of monophyletic origin and that they are located in either the inner membrane or in one of the two outer membranes in diatoms plastids. Thus, TPTs of the red algal endosymbiont were probably reused and integrated into both, the membranes of the host and the endosymbiont.

The phylogenetic analysis indicate the same for diatom NTTs, which are probably also derived from one or several independent events of horizontal gene transfer from the endosymbiont to the host genome. Especially regarding the complex membranous structure of diatom plastids, it is not clear how host proteins would be targeted to the symbiont without the previous establishment of a functional transport and targeting system. Several possibilities reach from an early-stage 'wrong' dual targeting (Gile *et al.* 2015, Martin 2010) to the hypothesis that the host directly transferred transporters to the symbiont by an unknown transport way prior to the stabilisation of the metabolic linkage (Moog *et al.* 2015). A similar idea was also discussed by Keeling in the so-called targeting-ratchet model (TRM)

(Keeling 2013). This would have been possibly facilitated by the symbiont-specific ERAD-like machinery (SELMA), derived from the cell membrane of the symbiont, and transporters for the inner and outer envelope of the chloroplast (TIC/TOC complex), which were also originated from the endosymbiont. Since *Pt*NTT5 and *Tp*NTT3 are most likely integrated into membrane surrounding the ER or into ER-associated membranes. The *ntt* genes in diatoms are either originated from the red algal endosymbiont or by gene transfer from the bacterial group of chlamydiales or rickettsiales. The ER is of host origin, so NTTs that integrate into this membrane could be possibly derived from the host genome, whereas other NTTs (1 and 2), which integrate into the innermost membrane, would be descended from the endosymbiont. Recycling of the transporter sequences as it is suggested for the TPTs (Moog *et al.* 2015) seems unlikely, due to the distant phylogenetic relation and the different numbers of present introns (Chu *et al.* 2016b). All these hints indicate that the different diatom NTTs evolved differently. NTT1 isoforms were acquired from the endosymbiont, whereas the remaining NTTs might have arisen from one or several horizontal gene transfers, and each transporter protein is possibly specific for different substrates.

7.4 Targeting and insertion processes of the membrane protein *Pt*NTT5

Starting with the *in silico* studies of *Pt*NTT5, it was obvious that this transporter differs from NTT1 and NTT2. The lack of a characteristic plastid presequence, low sequence similarities to any NTT investigated so far and an accumulation of positively charged amino acids emphasised the peculiarity of *Pt*NTT5. The latter indicated ER localisation, since it is known that ER proteins contain several positively charged amino acids at the N-terminus, flanked by hydrophobic residues (for review see (Zimmermann *et al.* 2011)). Interestingly, this accumulation was present at the C-terminus of *Pt*NTT5. Moreover, C-terminal GFP fusion apparently perturbs the targeting and/or folding processes, since all truncated versions of *Pt*NTT5 ended in patchy accumulations inside the cells, maybe associated to the plastid, but not very likely for an insertion into a membrane surrounding the plastid. Anyhow, *Pt*NTT5 harbours hydrophobic transmembrane (TM) domains, thus it is conceivable that it has to integrate into a membranous environment. It has been reported that a vesicle-like network exists in the periplastidial space of *P. tricornutum*, but its function remains to be unknown (Flori *et al.* 2016). One possible function could be the temporary storage of defective proteins prior to degradation, which can be kept either in a soluble state inside these vesicles or inserted into the vesicle membrane, if the proteins are hydrophobic like in the case of *Pt*NTT5. In contrast, N-terminal GFP fusion in combination with C-terminal truncations of the protein changed the localisation, indicating that these features are crucial for a proper targeting function. Additionally, the targeting process of *Pt*NTT5 was suggested to take place post-translationally (Chu *et al.* 2016b), and thus would differ from other nucleus-encoded plastid proteins, like the fucoxanthin chlorophyll proteins (FCPs), which were suggested to be targeted co-translationally depending on their precursor polypeptides (Bhaya & Grossman 1991). Surprisingly, despite the lack of a signal peptide domain and even with an N-terminal attachment of GFP, *Pt*NTT5 was directed to the ER membrane. This implies that the transporter protein was first translated prior to the insertion into the ER membrane (see Figure 26A).

It would be conceivable that *Pt*NTT5 harbours a so-called C-tail anchor (TA) domain. Such domains contain hydrophobic amino acids, which help in holding transmembrane polypeptides within the phospholipidlayer, being reported to be found in mitochondrial, ER and chloroplast membranes (Borgese *et al.* 2007, Borgese *et al.* 2003, Borgese & Fasana 2011). TA proteins normally contain only one C-terminal transmembrane domain, which

functions in both targeting to organelles and insertion into the membrane in a post-translational manner (Borgese *et al.* 2003). Several *in vitro* studies revealed that insertion pathways of TA proteins can range from spontaneous unassisted insertion into the membranes to energy requiring mechanisms, mediated by chaperones or protein complexes (Abell *et al.* 2007, Borgese & Fasana 2011, Shao & Hegde 2011). Furthermore, it was shown that insertion pathways of TA proteins and multi-span proteins might overlap (Dukanovic & Rapaport 2011, Otera *et al.* 2007). For *Pt*NTT5, a similar scenario is possible. After translation, the protein would be kept in an unfolded state, with the help of cytosolic chaperones. A C-terminal post-targeting signal could be present in *Pt*NTT5, facilitating transport to the ER. This mechanism could proceed without the interaction of a signal recognition particle (SRP), the protein is then drawn towards the ER membrane by the hydrophobic core of a signal peptide-like domain at the C-terminus. A similar phenomenon without the presence of a SRP was also described by (Jungnickel & Rapoport 1995), where the translocon Sec61 complex was demonstrated to be sufficient for binding the ribosome to the ER membrane, due to the interaction of the hydrophobic part of the signal sequence with the Sec61 complex and phospholipids (Martoglio *et al.* 1995). Since Sec61 is known to mediate co- and post translational translocation across the ER membrane for soluble proteins (Rapoport 2007), there might be similar mechanism for hydrophobic proteins, either involving the same or a translocation system homologous to the Sec system.
After translocation across a Sec-like channel, *Pt*NTT5 could then be folded from the inside of the cER lumen into the cER membrane, possibly stepwise and assisted by chaperones, starting from the C-terminus due to the hydrophobic domains.
The truncations of the C-terminal part of *Pt*NTT5 could define the location of a potential targeting domain. Possibly the ninth TM domain contains a crucial targeting domain, since the shorter versions which lack this area were not even entering the secretory pathway and the subcellular GFP fluorescence was observed in the cytosol. This is a further indication for the post-translational insertion of *Pt*NTT5.
In contrast, the C-terminal GFP attachment to *Pt*NTT5 showed a clearly different GFP-pattern: GFP seemed to be accumulated in or associated to the plastid. Most likely, the recognition of the targeting domain was still taking place in that case and the protein was also translocated across a Sec-like translocon (Figure 26B). However, the insertion of the membrane protein would be impaired, possibly by the GFP-molecule impeding the integration process of the membrane proteins or resulting in an unsteady insertion into the membrane due to restricted folding properties. A consequence would be the degradation of the protein. The symbiont-derived ERAD (ER-associated degradation) - like machinery (SELMA) is suggested to be situated in the second outermost membrane of diatoms and being part of a preprotein transport system across this membrane (Hempel *et al.* 2009). Furthermore, components of the SELMA are possibly involved in ubiquitination and proteasomal subunits are present in the periplastidial space (Sommer *et al.* 2007, Stork *et al.* 2012). Accordingly, it is very likely that the PPS is a place in which proteins that are misled to plastidial compartments are degraded. Soluble proteins destined for degradation could be released from the cER lumen to the cytosolic proteasome via the ERAD translocation machinery, which is most likely retained in diatoms from the host ER. Instead highly hydrophobic membrane proteins could be guided through the SELMA towards the PPS and might be ubiquitinated in parallel, followed by a direct insertion into the PPS-membrane in order to keep the hydrophobic proteins in a membranous environment prior to degradation. Alternatively, after translocation across the SELMA, proteins could form aggregates and be introduced into the degradation process. This could explain how the GFP fusion proteins

could be guided to this location and why GFP accumulates inside the cells in patchy structures, probably near to the plastid or associated to plastidial membranes.

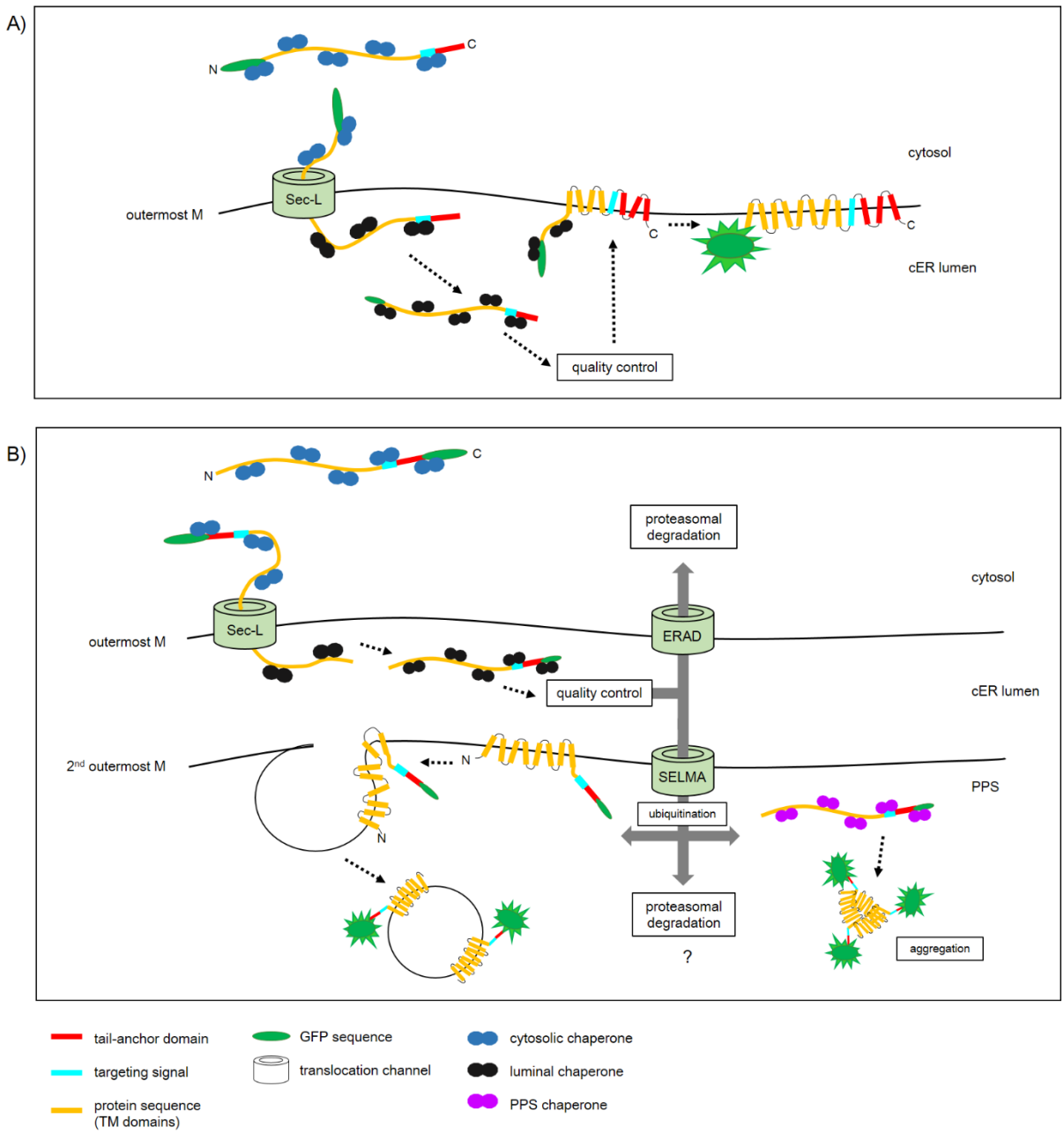


Figure 26: Schematic illustration of a putative post-translational targeting and insertion process for *Pt*NTT5 GFP fusion proteins. A) N-terminal GFP fusion: the protein is kept in an unfolded state with the help of chaperones and is translocated through a Sec-like translocon into the cER lumen. The C-terminal targeting and tail-anchor domain help to insert the protein into the ER membrane and to keep it within the membrane. B) C-terminal GFP fusion: the protein is also translocated into the cER lumen across a Sec-like translocon. After unsuccessful integration of the protein into the ER membrane, it is determined for degradation. Therefore, it could be guided across the SELMA into the PPS. For defective soluble proteins it is feasible to release them directly into the cytosol for proteasomal degradation. The hydrophobic proteins in the PPS might be ubiquitinated prior to integration into vesicle-like structures. The formation of vesicles would not require the ATP-dependent chaperone interaction and thus could keep the proteins temporarily in a membranous environment before they will be introduced into the degradation process. Alternatively, the proteins could form aggregates, which are finally also degraded. TM = transmembrane domain; Sec-L = Sec-like translocation channel; ERAD = ER-associated degradation system; SELMA = symbiont-derived ERAD-like machinery; ER = chloroplast ER; PPS = periplastidial space

7.5 What tools could help to locate proteins in plastidial subcompartments of diatoms?

The localisation of the protein of interest is an important aspect, which can lead to final conclusions in the intracellular role and function of a protein in the organism. A common method for protein localisation in diatoms, is the utilisation of fluorescence-based reporter genes which are attached to the protein and stably expressed in the cells. The fusion protein can then be traced *in vivo* via microscopic analyses. The exact localisation of the fluorescence signal represents a great challenge since the membranes and subcompartments of the plastid are in close proximity to each other. GFP-based localisation studies tracing plastidial proteins were shown to be error-prone (Chapter 5). The application of full length GFP might not only result in similar subcellular GFP-patterns, but might also hinder or perturb the folding of the protein. In particular, for membrane proteins the folding and integration into the respective membrane could be significantly affected by the fusion of GFP. With regard to *Pt*NTT5, it was shown that C-terminal GFP attachment apparently led to mistargeting of the protein (Chu *et al.* 2016b). Accordingly, the utilisation of smaller self-assembling GFP fragments should not only prevent the problem of misfolding and/or -targeting due to a bulky and large GFP-molecule, but should also give either clear distinct signals (if the protein is co-localised with the marker protein in the respective compartment) or should give no signal at all. However, unambiguous results after application of the saGFP system as it was obviously obtained for the plastidial TPTs (Moog *et al.* 2015), could not be confirmed for the NTTs. First attempts to localise *Pt*NTT5 with saGFP resulted in multiple signals in combination with the ER- and the PPS-marker (data not shown). The same applied to the experiments with the different marker proteins, which also resulted in a series of false signals despite differently targeted saGFP-fragments (Chapter 5). The targeting function was obviously inactive, probably due to the strong affinity of the GFP-fragments for each other. As a result the assembled GFP molecules were either trapped in one of the outer compartments or were dragged along into one of the inner compartments, similar to the 'piggyback-mechanism' reported for peroxisomal and nuclear proteins (Freitas & Cunha 2009, Thoms 2015, van der Klei & Veenhuis 2006). Another explanation for this phenomenon, could be the fact that the transgenes are stably, but also randomly, integrated into the diatom genome, where they are constantly expressed. One consequence could be over-expression artefacts, displayed by unexpected GFP-signals. Besides, the saGFP technique was developed and initially studied in bacterial *in vitro* systems (Cabantous *et al.* 2005, Cabantous & Waldo 2006). The method was used to study protein localisation and proteins topology in plant protoplasts or mammalian cells (Bionda *et al.* 2016, Bionda *et al.* 2010, Chun *et al.* 2007, Groß *et al.* 2011, Kaddoum *et al.* 2010, Kamiyama *et al.* 2016, Milech *et al.* 2015, Nasu *et al.* 2016, Sommer *et al.* 2013, Sommer *et al.* 2011, Van Engelenburg & Palmer 2010). In diatoms (Bullmann *et al.* 2010, Felsner *et al.* 2010, Hempel *et al.* 2009, Hempel *et al.* 2010, Lau *et al.* 2016, Lau *et al.* 2015, Moog *et al.* 2015, Vugrinec *et al.* 2011) and apicomplexans (van Dooren *et al.* 2008), it is in particular used to distinguish between the compartments and membrane of the complex plastid. It was suggested to induce the saGFP expression in *P. tricornutum* cells for 6 hours in order to take microscopic images (Hempel *et al.* 2009), indicating that saGFP expression is a time-dependent process with the risk of changing the subcellular GFP-pattern which was also mentioned in one study (Lau *et al.* 2015). However, the development of saGFP by flow cytometric and microscopic observations over time, using the same gene cassette transformed in *P. tricornutum* cells, could not confirm this correlation, since the GFP-phenotype remained unchanged independent on the time interval (Chapter 5).

Consequently, one should consider alternatives for GFP-based localisation studies. Another powerful tool to identify the correct subcellular localisation of proteins, is immunogold

labelling, which was also considered for the investigation of plastid-targeted nucleotide translocators (NTTs) of diatoms. An important prerequisite to perform immunogold labelling is the generation of a suitable antibody, which is a time-consuming and challenging process. Even when an antibody should have been raised successfully, NTTs (and generally membrane proteins) are rather present in only low abundance. Hence, these proteins are hard to detect and to label. Alternatively, the attempts to achieve homologous over-expression and purification of NTTs by using e.g. His-tags, are not successful by date. These findings indicate that NTTs are poorly expressed and that the His-tag is probably not accessible for interactions with antibodies. Furthermore, electron micrographs do not guarantee the proper distinction of the membranes, which are only in an nm-range distance from each other (Bailleul *et al.* 2015, Tanaka *et al.* 2015). The separation, preparation and enrichment of the different plastidial subcompartments and their membranes would be an important and crucial requirement for the establishment of future studies on plastid proteins. So far, these attempts were partly successful: the separation of semi-intact plastids in the centric diatom *T. pseudonana* is possible, but the isolated plastids lack parts of the stroma which is due to damage of the plastid envelopes (Schober 2014). Besides the conservation of the outer membranes and especially the outermost membrane which is continuous with the ER is yet not achievable. Nonetheless, the isolation of semi-intact plastids and the first investigations on the plastidial proteome (Schober 2014), represent important steps into the direction of studying the process of protein (de-)glycosylation and the influence on plastid proteins.

To confirm ER localisation of *Pt*NTT5, available chemical fluorophore-based ER trackers were utilised, but without the required success: co-localisation studies with the ER-Tracker Blue-White DPX (Molecular Probes, Eugene, OR, USA) in *P. tricornutum* did not lead to a conclusive labelling of the organelle in this organism. When applied as recommended by the manufacturer, no staining was received, possibly due to difficulties of diffusion of the dye through the cell wall. When applied at higher concentrations (5 µM, 5-fold higher compared to the recommended maximum concentration), the dye accumulated in lipid droplets of older cells, whereas in younger cells (which do not contain lipid droplets), no staining could be achieved. Additionally, another ER stain ($DiOC_6(3)$) was applied. However, DiOC6(3) is known to stain not only ER (Terasaki 1989), but also mitochondria (Koning *et al.* 1993), which was also the case in *P. tricornutum* cells. Apart from this ambiguous labelling characteristics, $DiOC_6(3)$ fluorescence is green and the dye is therefore not suitable for co-localisation experiments with GFP.

Hence, it is still not trivial to reveal the exact localisation of plastidial proteins in diatoms. The false positive GFP-signals, received when applying saGFP, initiated the investigation of the system in more detail, especially regarding the promoter activity. The examination of cell lines expressing full length GFP under control of the same promoter (nitrate reductase promoter) in active and inactive conditions revealed that the activity of the nr promoter apparently could not be entirely switched off (Chu *et al.* 2016a). One conclusion of the promoter study was, that the inducible nr promoter might achieve a higher yield in protein expression when compared to the constitutive lhcf1 promoter, but there might be an unspecific background expression, either due to the leakiness of the nr promoter or to residual nitrate amounts inside the cells. This background signal is undesirable, especially in correlation with saGFP, which could explain the false signals. According to this, the choice of another promoter, which leads to weaker expression of the protein and which is entirely inactive at a given conditions, is highly recommended. Since in diatoms, the availability of endogenous promoters is restricted, the only inducible promoter apart from the nr promoter to date, is the flavodoxin (fld) promoter (see Table 12). Preliminary work in order to characterise this promoter, demonstrate that this promoter is very sensitive to the iron

content in the medium: it is only switched on by iron-deprived conditions and otherwise stays in an inactive state (Buck 2016, Yoshinaga *et al.* 2014).

Table 12: Available endogenous promoters for the utilisation of genetic engineering in diatoms. JGI Protein IDs were obtained using the EST database http://www.diatomics.biologie.ens.fr/EST3/seq.php

name	**protein**	**JGI Protein ID**	**induction/ expression**	**reference**
lhcf1 (fcpA)	fucoxanthin chlorophyll a/c binding protein A	18049	induced by light, constitutive expression	(Apt *et al.* 1996)
lhcf2 (fcpB)	fucoxanthin chlorophyll a/c binding protein B	25172	induced by light, constitutive expression	(Apt *et al.* 1996)
nr	nitrate reductase	54983	induced by nitrate, inactive by ammonium	(Stork *et al.* 2012)
h4	histone 4	34971	constitutive	(De Riso *et al.* 2009, Siaut *et al.* 2007)
ef2	elongation factor 2	35766	constitutive	(Seo *et al.* 2015)
fld	flavodoxin	23658	only induced upon iron-limitation	(Yoshinaga *et al.* 2014)

7.6 Conclusion and Outlook

For *P. tricornutum*, preliminary GFP-localisations of the remaining NTT isoforms (*Pt*NTT3, *Pt*NTT4 and *Pt*NTT6) (Figure 27) indicate that they might be located in plastid associated membranes, as the GFP-signal is found in close proximity to the plastid. However, since we cannot rule out that a C-terminal GFP fusion might lead to perturbation in the folding and/or targeting mechanism, as it was already shown for *Pt*NTT5, an additional N-terminal GFP-fusion would possibly reveal a different localisation. The same does apply to the NTT isoforms of *T. pseudonana (Tp*NTT4-8).

The heterologous expression of the remaining NTT isoforms in *E. coli* in order to identify the substrate spectra and transport properties were not successful so far. Hence, the applied expression system and the respective conditions might be suitable for the previous NTTs but not for the remaining isoforms. Accordingly, the crucial parameters need to be identified and consequently to be changed and adapted to achieve the expression of those NTTs. These parameters can vary from the used promoter and/or protein sequences to variations in the culture conditions and can be followed up to a change of the applied expression system itself (Haferkamp & Linka 2012).

In conclusion, this work showed that NTTs are present in other membranes of the diatoms plastid, beside the innermost, in particular the ER membrane. Hence, the outermost membrane represents a selective barrier for purine nucleotide transport. Further investigations on the other NTT isoforms are necessary to understand how nucleotides, and especially pyrimidine nucleotides, find their way from the cytosol across the plastidial membrane to enter the stroma, where they are crucial building blocks for DNA and RNA synthesis. The studies on *Pt*NTT5 and *Tp*NTT3 revealed that the nucleotide metabolism in diatoms is indeed more complex regarding both the transport capacity and the location of NTTs (Table 13) and might explain the outnumbered presence of NTT isoforms, especially in accordance to the complex membrane structure of the plastid. Moreover, the physiological functions of diatoms NTTs indicates that these transporters might be crucial not only for the net nucleotide uptake but also for the supply of energy-rich metabolites to plastidial compartments and even potentially in signalling pathways.

Furthermore, the targeting of proteins to the ER membrane in *P. tricornutum* can occur post-translationally and is not solely depending on a characteristic N-terminal signal peptide domain. Instead, internal domains (which still need to be identified) can take over the targeting function. Moreover, the attempts to reveal the protein localisation based on GFP need to be handled with caution, because solely the C-terminal GFP-fusion might not result in the correct protein localisation, N-terminal fusion and/or internal GFP attachment should be applied as well. Additionally, the choice of suitable molecular tools, especially promoter sequences, can change the course of a study, thus needs to be checked for required expression conditions and be adapted the aim of the study.

Table 13: Putative NTTs isoforms identified for the diatoms *P. tricornutum* and *T. pseudonana*, respectively.

NTT	**homologous NTT** [1]	**JGI Protein ID** [1]	**Substrates**	**presumed GFP-based subcellular localisation**	**suggested origin** [1, 2] **closest relation**
***Pt*NTT1**	***Tp*NTT1**	49533	ATP, ADP, AMP, dATP [1]	plastidial innermost membrane [1]	plants
***Pt*NTT2**	***Tp*NTT2**	45145	ATP, GTP, CTP, TTP, UTP, dATP, dGTP, dCTP [1]	plastidial innermost membrane? [1]	bacteria
***Pt*NTT3**	***Tp*NTT4**	50189	?	plastid-associated?	diatom ? (microsporidia)
***Pt*NTT4**	***Tp*NTT5**	46794	?	plastid-associated?	diatom ? (microsporidia)
***Pt*NTT5**	***Tp*NTT6/7/8**	54110	ATP, ADP, AMP, GTP, GDP, GMP, dATP, dGTP [2]	plastidial outermost membrane/ER membrane [2]	diatom ? (microsporidia)
***Pt*NTT6**	***Tp*NTT6/7/8**	54907	?	plastid-associated?	diatom ? (microsporidia)
***Tp*NTT3**	?	270249	ATP, ADP, AMP, GTP, GDP, GMP, dATP, dGTP, dAMP, dGMP, cAMP, cGMP (Chapter 3)	plastidial outermost membrane / ER membrane / ER sub-compartment ? (Chapter 3)	diatoms ? (bacteria)

[1] (Ast *et al.* 2009)
[2] (Chu *et al.* 2016b)

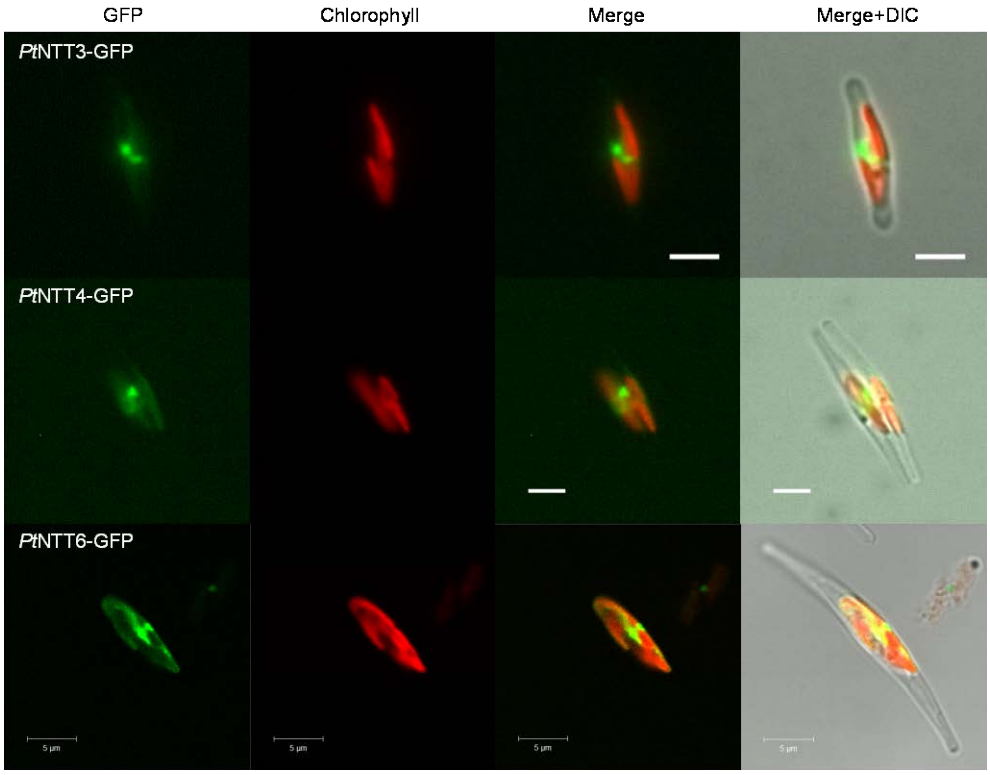


Figure 27: Subcellular localisation of *Pt*NTT isoforms fused C-terminally to GFP and expressed in *P. tricornutum* cells. GFP fluorescence, green; autofluorescence of the chlorophyll, red; Nomarski differential interference contrast (DIC), greyscale. Bars, 5 µm.

A. Supplementary Data

Supporting Information, Chapter 2

Table S 1: Oligonucleotides used to generate *Pt*NTT5 GFP fusion constructs.

Construct	Forward 5‘ -> 3‘	Reverse 5‘ -> 3‘	Method
Pt5full::GFP	ATGATGGAAATTGGTAGCCAAGTTC	GCGCTTCTTGTTCTTGATGGC	PCR
Pt5_1-159::GFP	GTCAGCCATGATGGAAATTGGTAGCC	ACCGCTGCGATCCAGAAGCTTTCC	„
Pt5_1-195::GFP	GTCAGCCATGATGGAAATTGGTAGCC	TACCGATCCAATCATGACGCCATGC	„
Pt5_1-229::GFP	GTCAGCCATGATGGAAATTGGTAGCC	TGTCAAGACAGACGCCATAAAGGACC	„
Pt5_1-256::GFP	GTCAGCCATGATGGAAATTGGTAGCC	GGAGACGGCGAATCCGGAGAG	„
Pt5_1-295::GFP	GTCAGCCATGATGGAAATTGGTAGCC	AGTAAAGCCGTACTTGTCCGCCATTTG	„
Pt5_130-547::GFP	CCGGCTGCCTTTCCCTTC	GCTACCAATTTCCATCATCCTGACG	del-PCR *
Pt5_160-547::GFP	CCACGGGGTGCCTTGAC	GCTACCAATTTCCATCATCCTGACG	„*
Pt5_196-547::GFP	CCCGCGGTCAAATTCATTTCTG	GCTACCAATTTCCATCATCCTGACG	„*
Pt5_230-547::GFP	CCCAACCAGAGCGCTACC	GCTACCAATTTCCATCATCCTGACG	„*
Pt5_257-547::GFP	CCGCTCGTCAACCGTGTC	GCTACCAATTTCCATCATCCTGACG	„*
Pt5_296-547::GFP	CCCAAGGACACGCACAAG	GCTACCAATTTCCATCATCCTGACG	„*
GFP::Pt5full	ATGATGGAAATTGGTAGCCAAGTTC	TTAGCGCTTCTTGTTCTTGATGGC	PCR
GFP::Pt5_1-542	TAAGCTGATCATCGACTAATTCGAGCTCGG	GATGGCTTTTGCCGGCCG	del-PCR
GFP::Pt5_1-531	TAAGCTGATCATCGACTAATTCGAGCTCGG	GTGGGCCTTGAGGTAGACTGTTTC	„
GFP::Pt5_1-470	TAAGCTGATCATCGACTAATTCGAGCTCGG	GCGGTAGCCAAAGACCCCAATC	„
GFP::Pt5_1-390	TAAGCTGATCATCGACTAATTCGAGCTCGG	TGGTGGCAGGATACCAAACTGC	„
GFP::Pt5_1-361	TAAGCTGATCATCGACTAATTCGAGCTCGG	GTCCGGGATTGCCGTGCC	„
GFP::Pt5_1-295	TAAGCTGATCATCGACTAATTCGAGCTCGG	AGTAAAGCCGTACTTGTCCGCCATTTG	„
GFP::Pt5_1-256	TAAGCTGATCATCGACTAATTCGAGCTCGG	GGAGACGGCGAATCCGGAGAG	„
GFP::Pt5_1-229	TAAGCTGATCATCGACTAATTCGAGCTCGG	TGTCAAGACAGACGCCATAAAGGACC	„
GFP::Pt5_1-195	TAAGCTGATCATCGACTAATTCGAGCTCGG	TACCGATCCAATCATGACGCCATGC	„
GFP::Pt5_1-159	TAAGCTGATCATCGACTAATTCGAGCTCGG	ACCGCTGCGATCCAGAAGCTTTCC	„
GFP::Pt5_1-129	TAAGCTGATCATCGACTAATTCGAGCTCGG	GAAGCCAGCGTACCCCGTGGAG	„

Constructs marked with an asterisk comprised the original start region (encoding the first six amino acids of the N-terminus) of full-length *Pt*NTT5 to support correct translation initiation. Cloning was done by PCR and insertion into PTV-GFP (PCR) or via deletion PCR and re-ligation (del-PCR).

Table S 2: GC-content of diatom NTTs.

T. pseudonana	**GC content [%]**	***P. tricornutum***	**GC content [%]**
NTT1	46	NTT1	51.3
NTT2	45.7	NTT2	48.5
NTT3	50.5	NTT3	52.1
NTT4	42.9	NTT4	47.4
NTT5	48.6	NTT5	56
NTT6	48.6	NTT6	58.9
NTT7	45.8		
NTT8	45.4		
Average	46.9	Average	48.8
Average (coding)	48		

To allow comparison the average GC contents of the respective genomes are given.

Table S 3: Amino acid similarities of diverse functionally characterised NTTs (in %).

	Plant *A.t.*	diatom NTT1	diatom NTT2	***Pt*NTT5**
Plant *A.t.* (ATP/ADP antiport)	(-)	58-59	41-44	**26**
Red alga *G.s.*(ATP/ADP antiport)	55-56	61-63	42-43	**24**
diatomNTT1 (AdeNuc:H^+symport)	58-59	(-)	44-46	**26-28**
diatom NTT2 (deoxy-)Nucantiport)	41-44	44-46	(-)	**26-27**
*Pt*NTT5	24-26	26-28	26-27	**(-)**
Chlamydiales(ATP/ADP antiport)	55-56	53-59	44-49	**29-31**
Chlamydiales(different functions)	41-49	40-52	36-45	**25-29**
Rickettsiales(ATP/ADP antiport)	42-45	47-51	41-44	**26-27**
Rickettsiales (different functions)	41-43	42-47	37-43	**25-29**
Microsporidia (ATP/ADP antiport or PurNuc transport)	23-33	25-35	24-32	**24-26**

*Pt*NTT5 exhibits highest similarity (29 %) to *Rp*NTT4 from *Rickettsia prowazekii* (CTP, UTP, GDP transport) and to *Sn*NTT2 from *Simkania negevensis* (GT/DP:H^+symport). *A.t.* = *Arabidopsis thaliana*; *G.s.* = *Galdieria sulphuraria*.

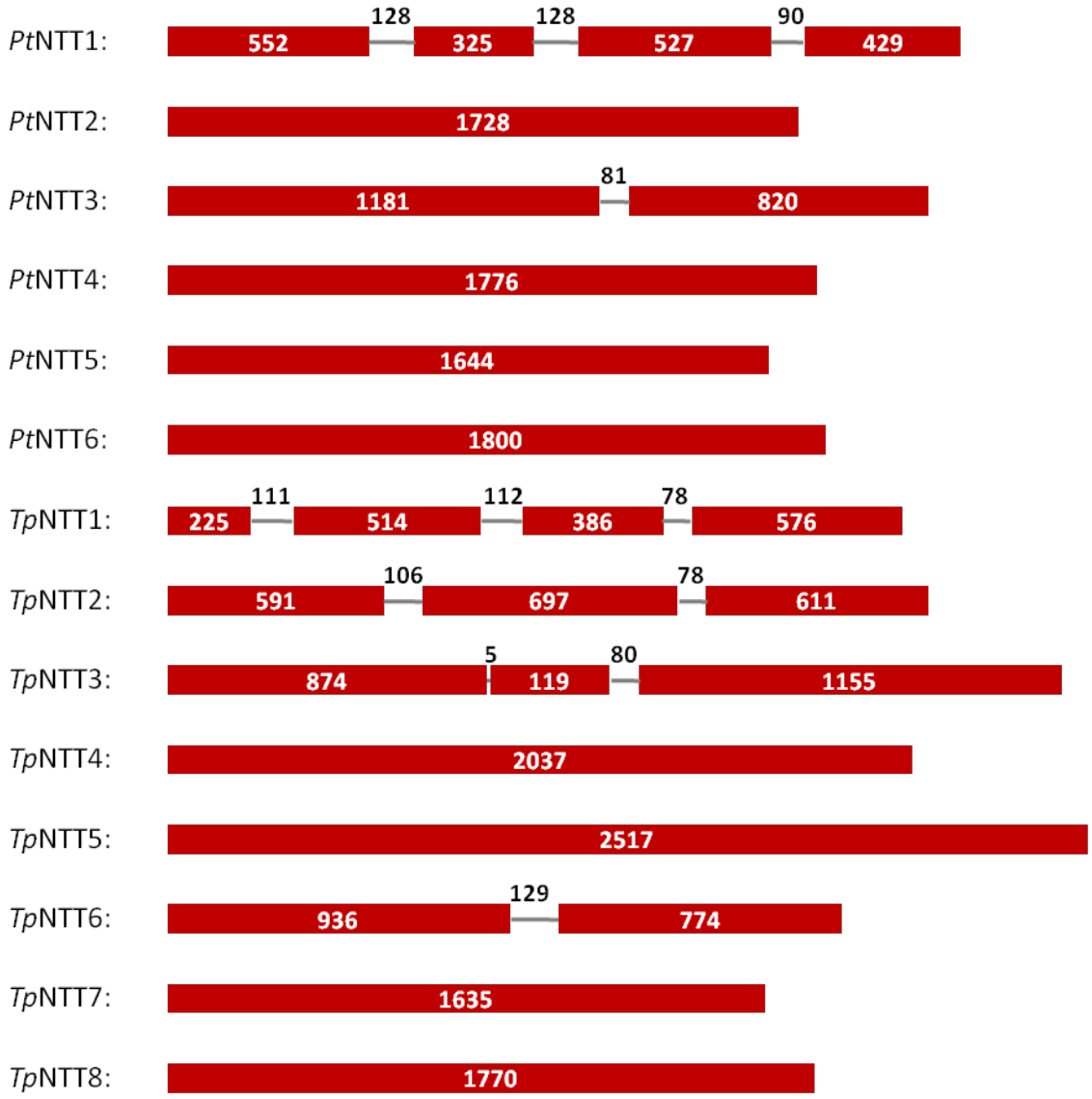
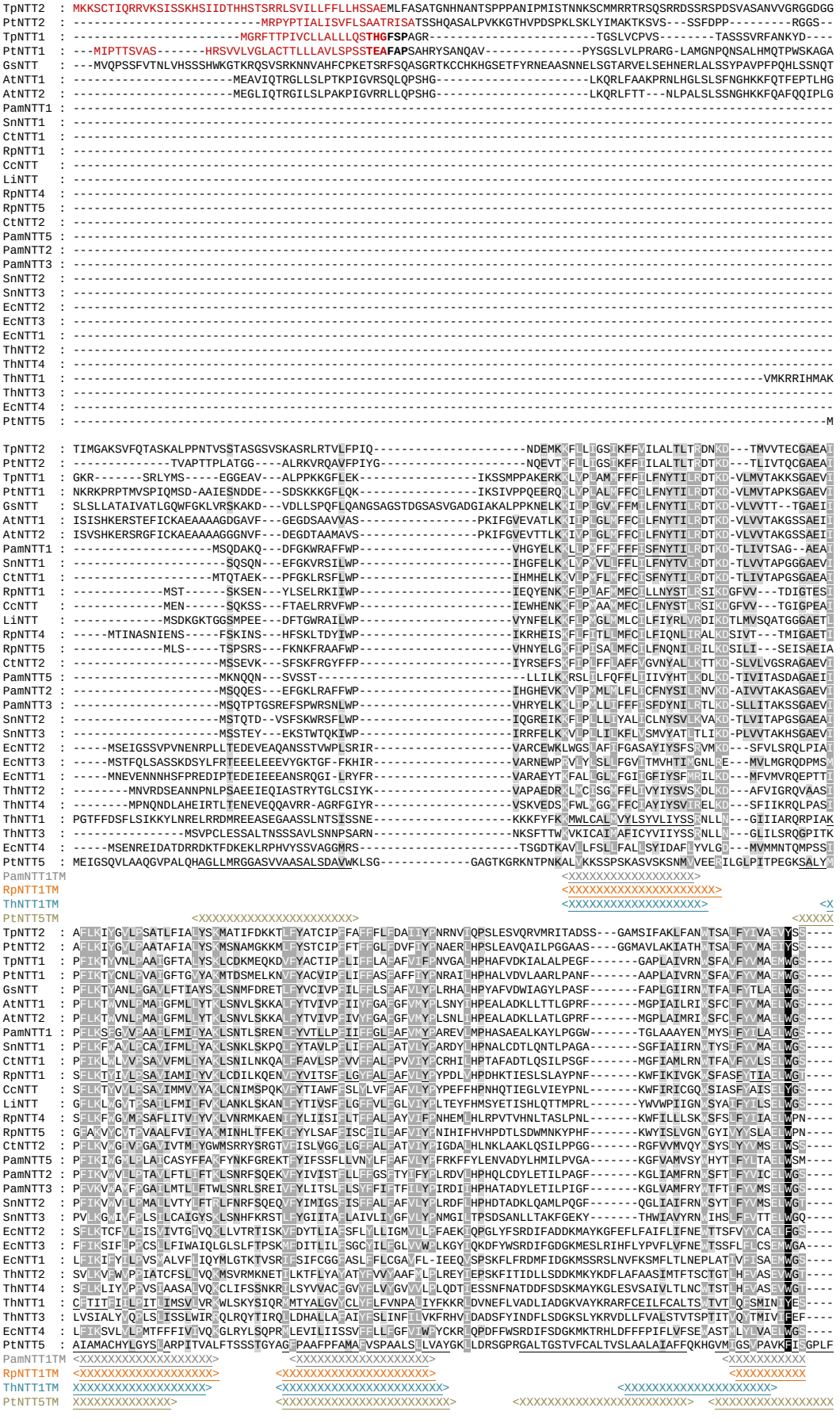


Figure S 1: Exon-intron structure of diatom NTTs. Exons are displayed as red boxes, introns as a grey line and the respective sizes (number of nucleotides) are given. Information was taken from the JGI genome portal.



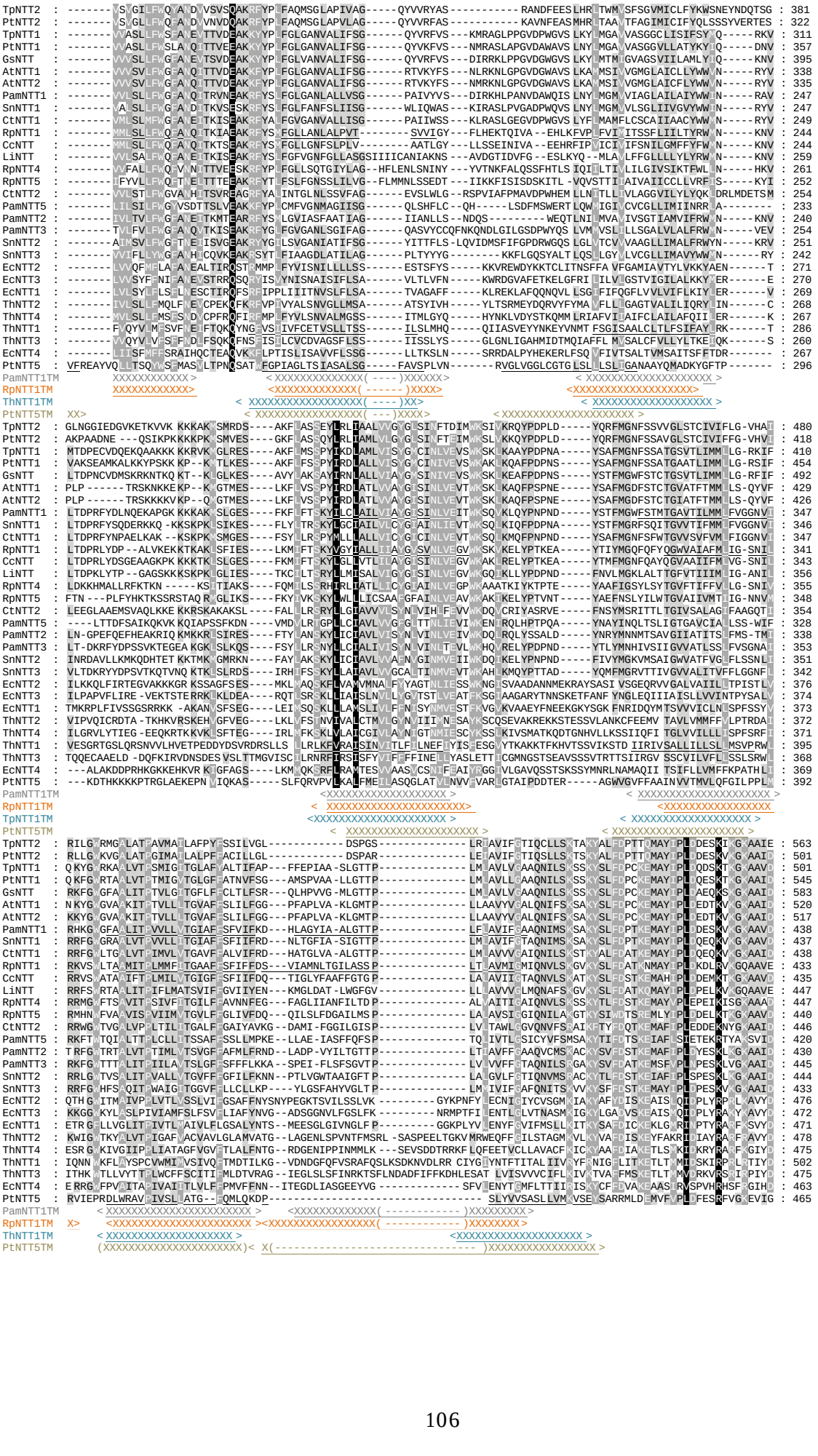


Figure S 2: Amino acid sequence alignment of biochemically characterised NTTs. The degree of conservation is indicated by different shading. The signal peptide of the bipartite targeting sequence is marked in red and the corresponding cleavage site in bold. Putative transmembrane domains of representative NTTs are underlined and additionally indicated below the corresponding regions.

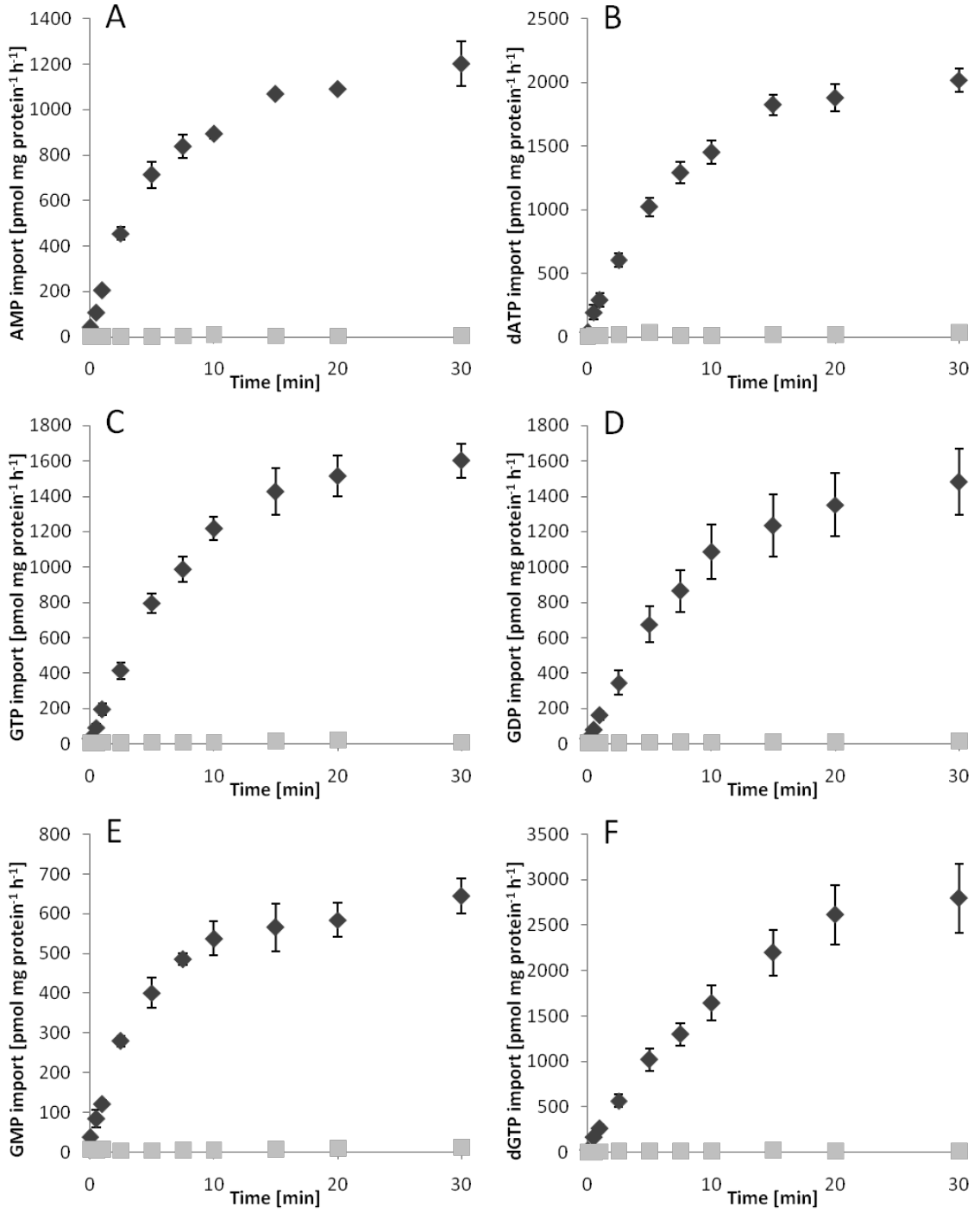


Figure S 3: Time dependent uptake of purine nucleotides via *Pt*NTT5. Transport of [α^{32}P]-AMP (A), [α^{32}P]-dATP (B), [α^{32}P]-GTP (C), [α^{32}P]-GDP (D), [α^{32}P]-GMP (E), [α^{32}P]-dGTP (F), into IPTG-induced *E. coli* cells synthesising *Pt*NTT5 (black rhombs) in comparison to nucleotide uptake into corresponding non-induced *E. coli* cells (grey squares). Data are the mean of at least three independent experiments, standard errors are given

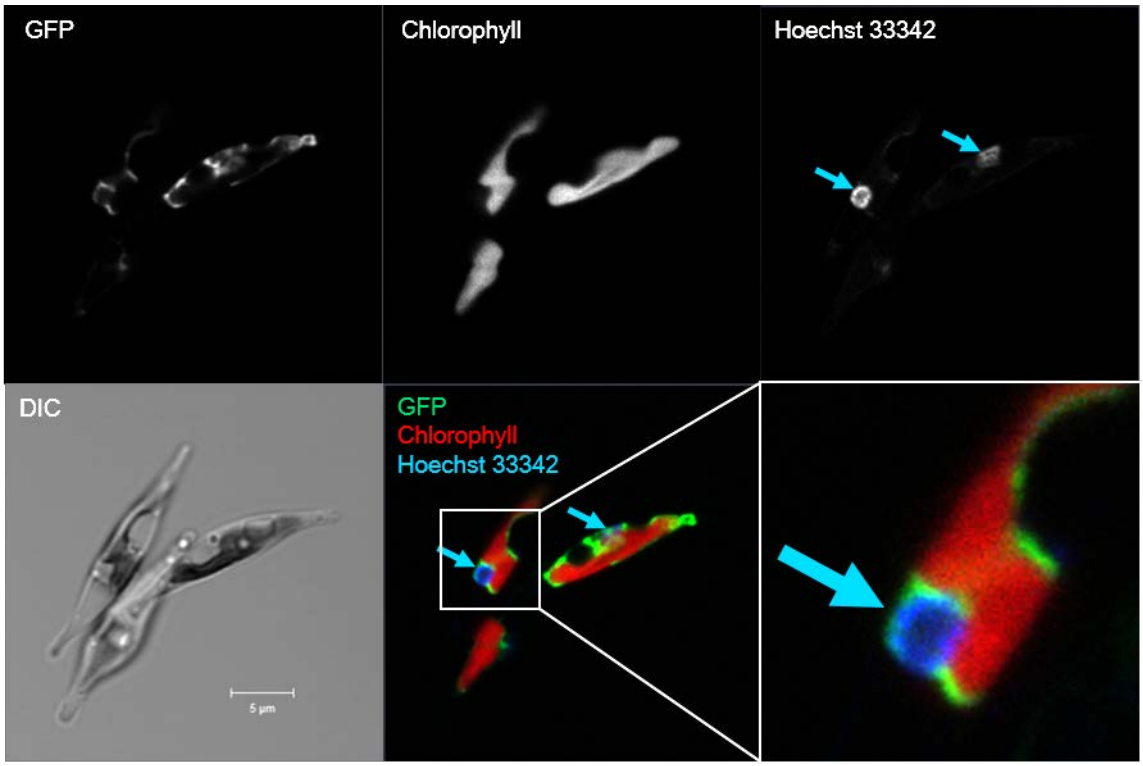


Figure S 4: Cellular localisation of *Pt*BiP::GFP in *P. tricornutum* cells. Cell line expressing the GFP fusion of the ER located luminal binding protein precursor (BiP). GFP fluorescence in green; autofluorescence of the chlorophyll in red; Hoechst 33342 fluorescence in blue; Nomarski differential interference contrast (DIC) in grey scale. Scale bars: 5 µm.

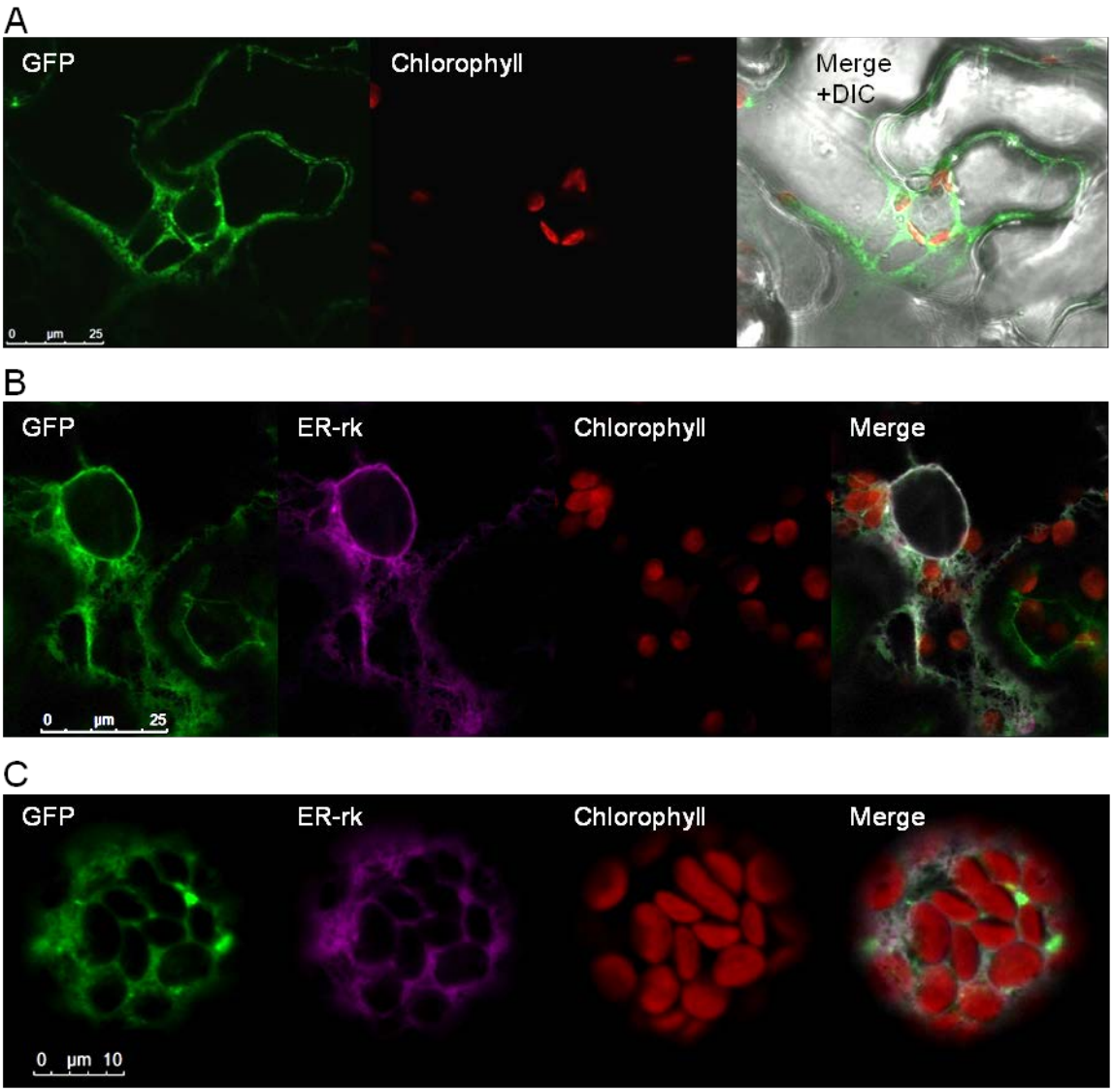


Figure S 5: Cellular localisation of *Pt*NTT5 with N-terminal GFP in plant cells. *Pt*NTT5 was inserted into the vector pUBN-GFP-Dest (N-terminal GFP fusion protein expression under control of the ubiqutin-10 gene promoter). Transient expression was analysed in *Nicotiana benthamiana* leaf tissue (Agrobacteria-mediated transformation, infiltration) and in Arabidopsis protoplasts (polyethylene glycol-mediated DNA uptake). Tobacco epidermal cells (A); tobacco leaf tissue (B); Arabidopsis leaf protoplasts (C). GFP fluorescence in green; autofluorescence of the chlorophyll in red; fluorescence of the ER-marker protein ER-rk in magenta; Nomarski differential interference contrast (DIC) in grey scale. Scale bars: Tobacco 25 µm; Arabidopsis 10 µm.

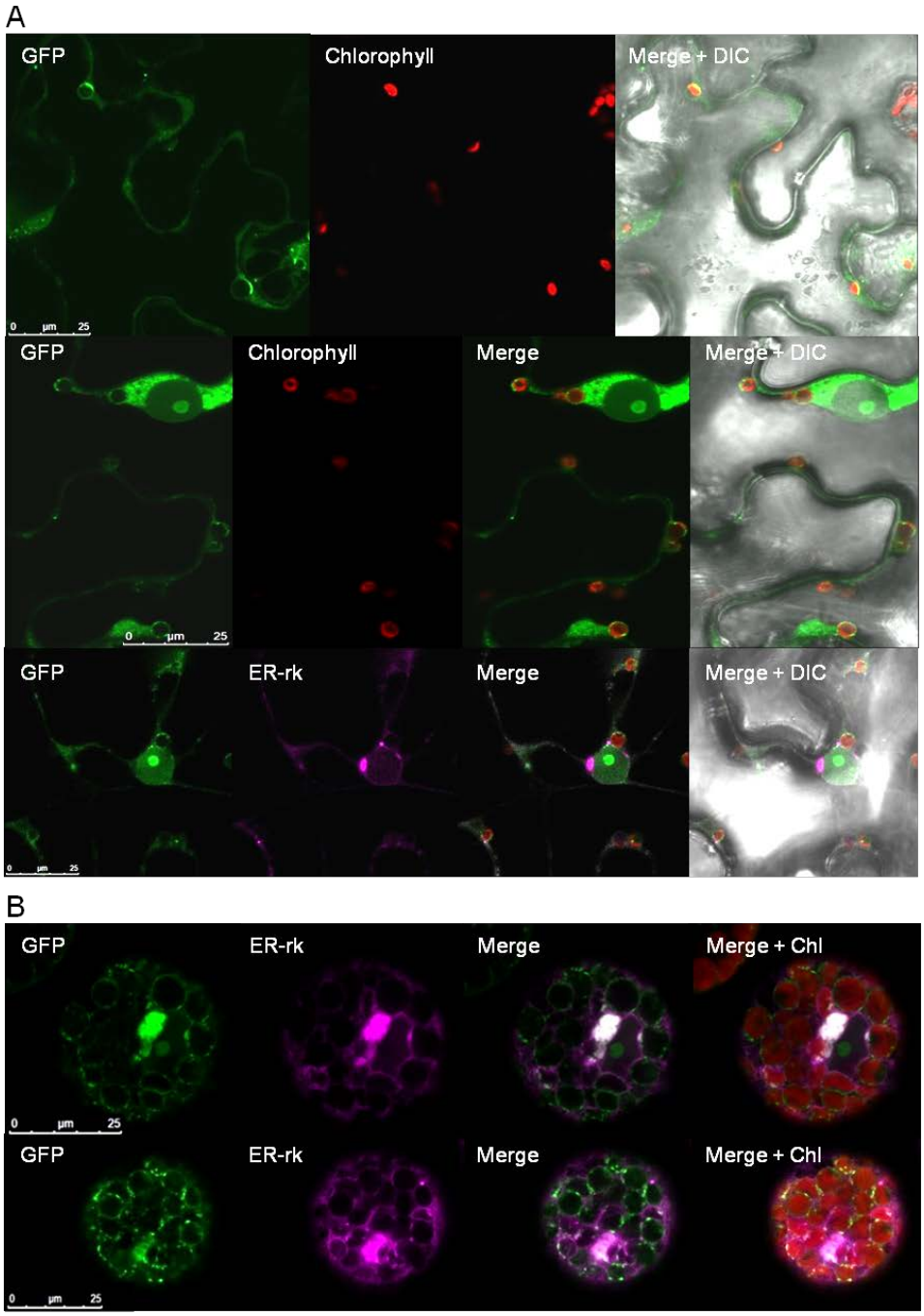


Figure S 6: Cellular localisation of *Pt*NTT5 with C-terminal GFP in plant cells. *Pt*NTT5 was inserted into the vector pUBC-GFP-Dest (C-terminal GFP fusion protein expression under control of the ubiqutin-10 gene promoter). Plant cell transformation was conducted as given in Figure S 5. Tobacco epidermal cells (A), Arabidopsis leaf protoplasts (B). GFP fluorescence in green; autofluorescence of the chlorophyll (Chl) in red; ER-rk fluorescence in magenta; Nomarski differential interference contrast (DIC) in grey scale. Scale bars: 25 µm

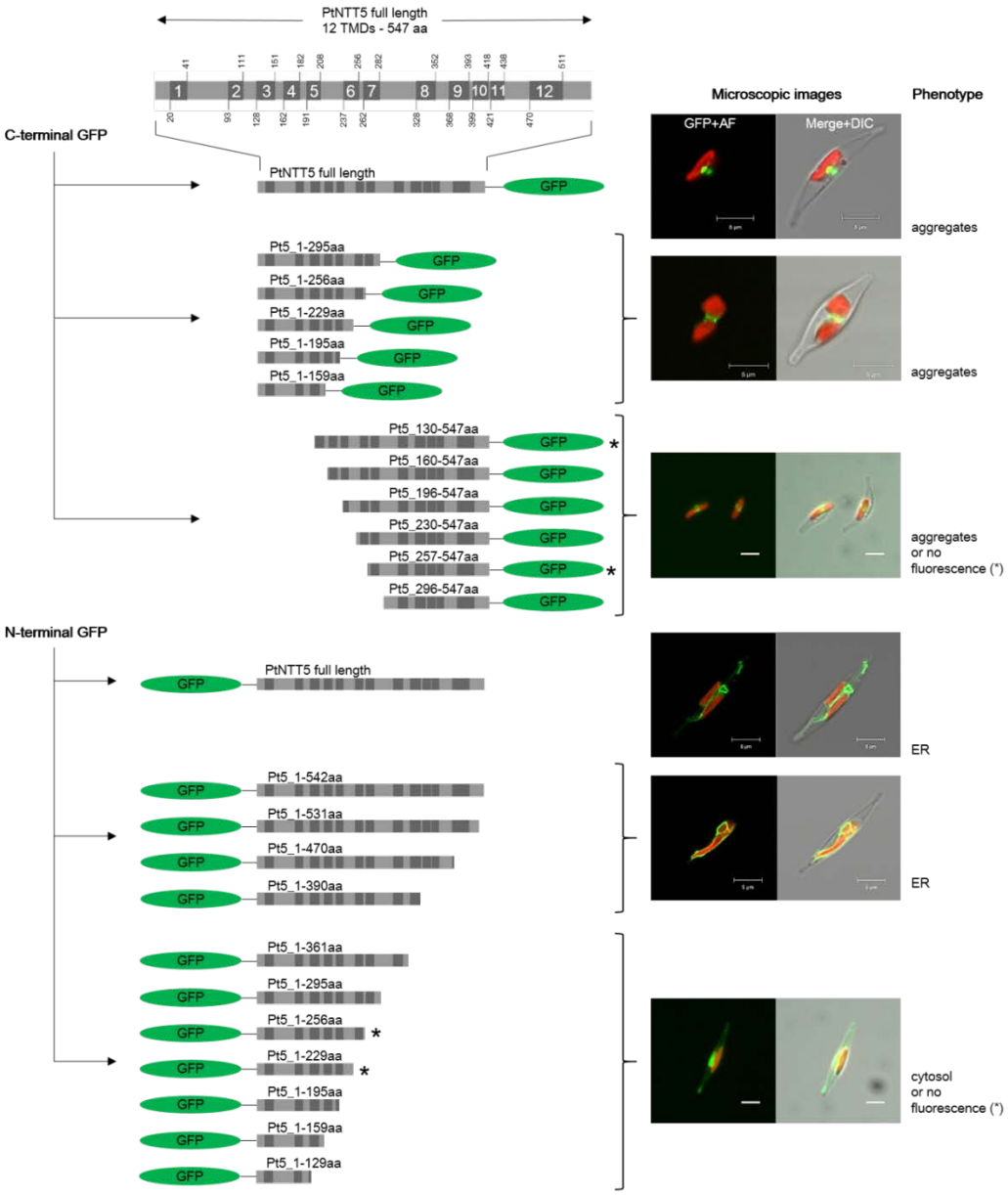


Figure S 7: Schematic illustration of GFP-fusion proteins of full-length and truncated *Pt*NTT5. *Pt*NTT5 (Pt5) harbours 12 predicted transmembrane domains (TMDs) (Figure S 2) and consists of 547 amino acids (aa). Given numbers of the truncated versions indicate the respective amino acid positions of *Pt*NTT5 (full length). GFP was fused C- or N-terminally to the target protein and transformed into *P. tricornutum* cells. The phenotypes were documented by microscopy and representative images were selected (for all microscopic images see Supplementary Figure S 8 - S 11). AF = autofluorescence of the chlorophyll; DIC = Nomarski differential interference contrast.Scale bars: 5 µm.

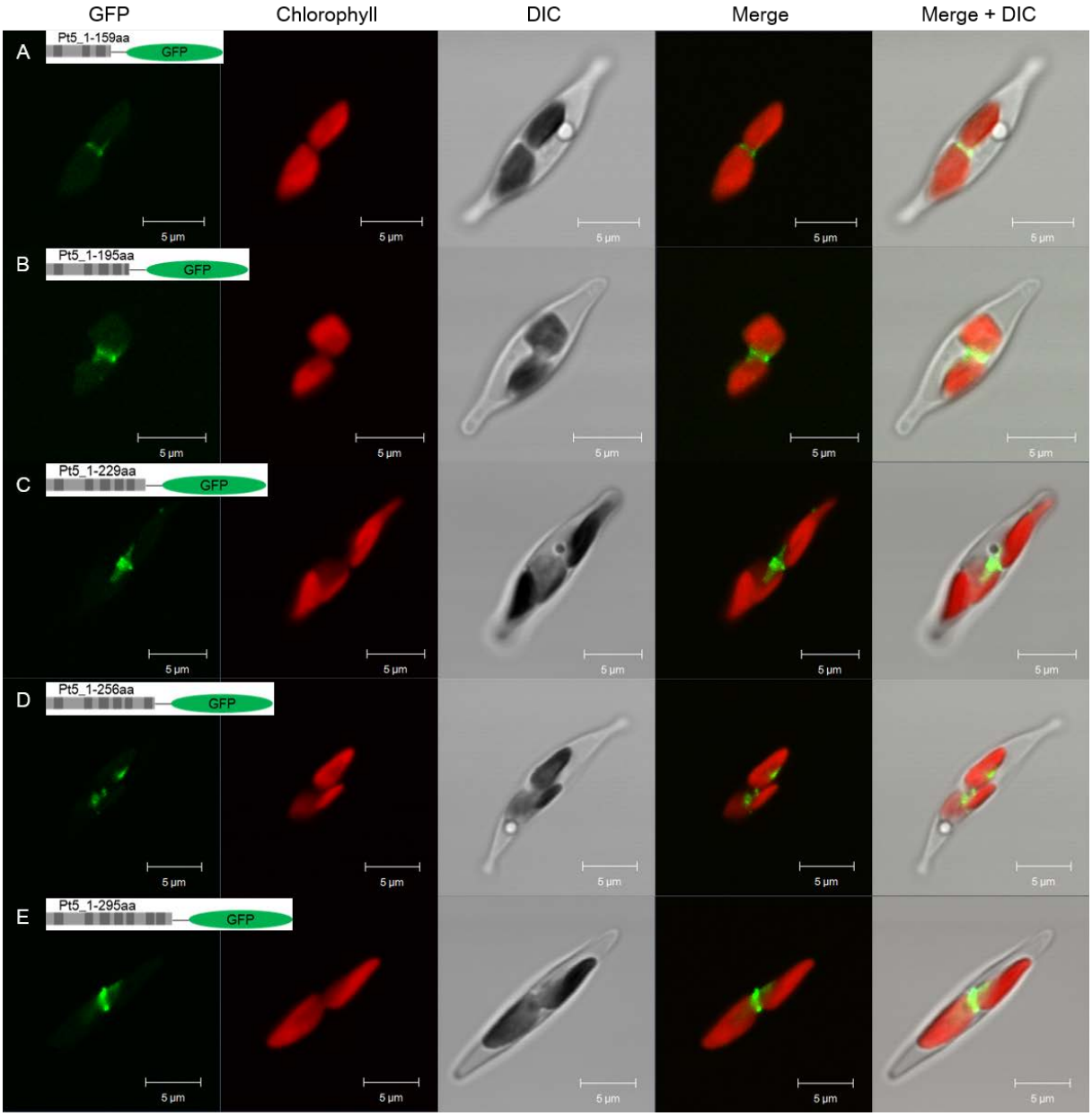


Figure S 8: Localisation of *Pt*NTT5 in *P. tricornutum* (C-terminal deletion, C-terminal GFP-fusion). GFP-fusion constructs with C-terminal fusion of GFP and C-terminal deletion of *Pt*NTT5 (Pt5) transmembrane domains. GFP fluorescence in green, autofluorescence of the chlorophyll in red, Nomarski differential interference contrast (DIC) in white. A: Pt5_1-159; B: Pt5_1-195; C: Pt5_1-229; D: Pt5_1-256; E: Pt5_1-295. Scale bars: 5 µm.

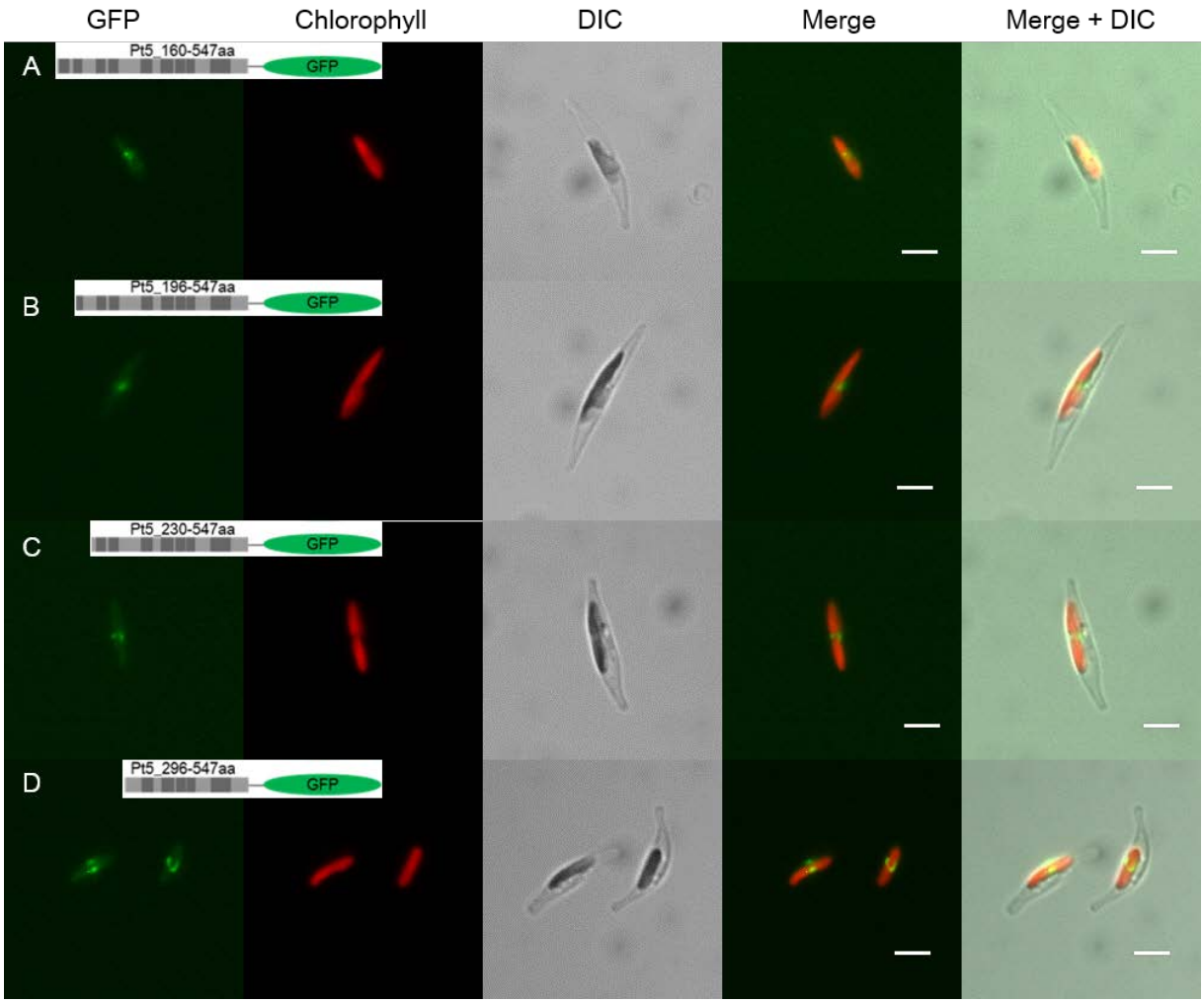


Figure S 9: Localisation of *Pt*NTT5 in *P. tricornutum* (N-terminal deletion, C-terminal GFP-fusion). GFP-fusion constructs with C-terminal fusion of GFP and N-terminal deletion of *Pt*NTT5 (Pt5) transmembrane domains. GFP fluorescence in green, autofluorescence of the chlorophyll in red, Nomarski differential interference contrast (DIC) in white. A: Pt5_160-547; B: Pt5_196-547;C: Pt5_230-547; D: Pt5_296-547. Scale bars: 5 µm.

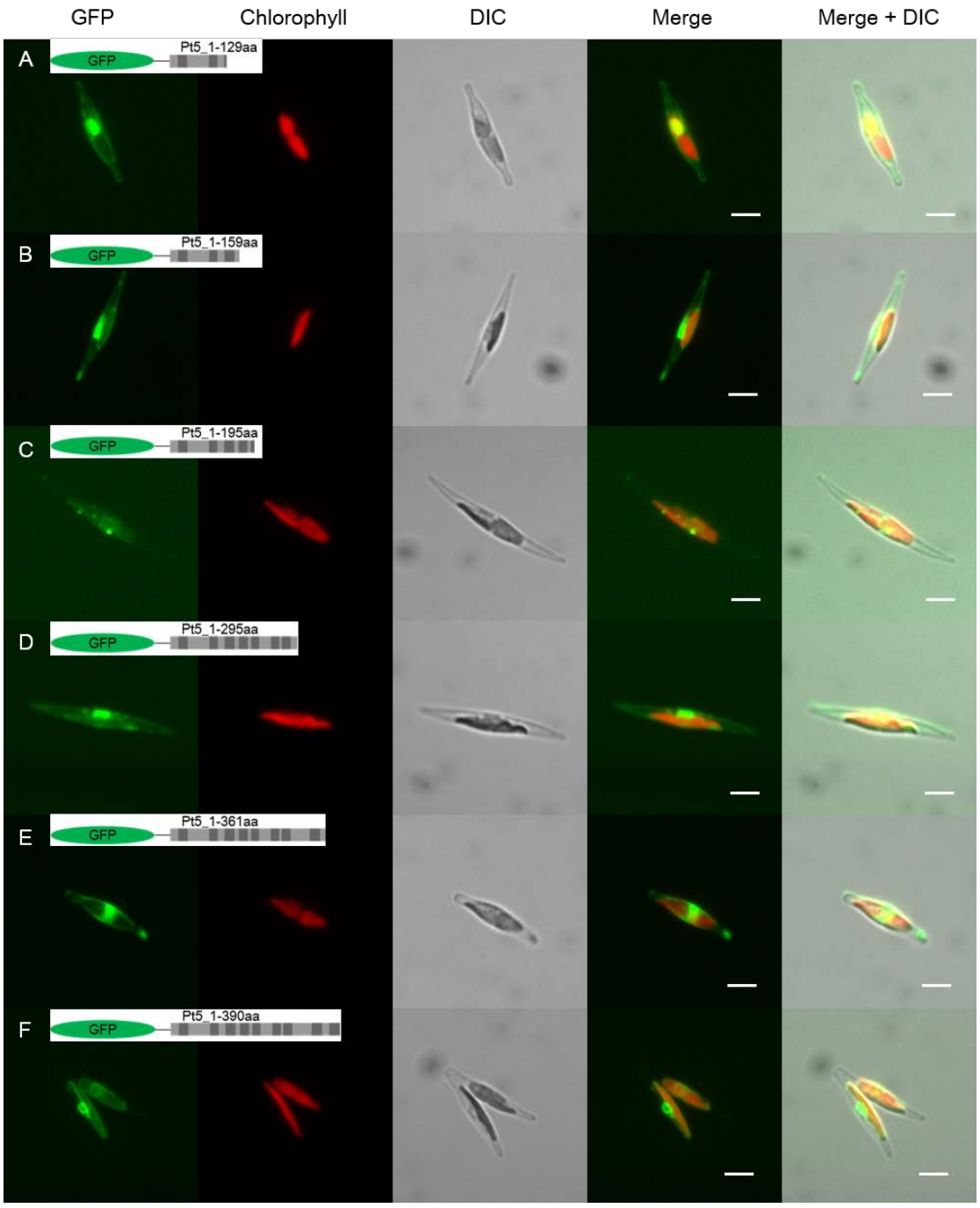


Figure S 10: Localisation of *Pt*NTT5 in *P. tricornutum* (C-terminal deletion, N-terminal GFP-fusion). GFP-fusion constructs with N-terminal fusion of GFP and C-terminal deletion of *Pt*NTT5 (Pt5) transmembrane domains. GFP fluorescence in green, autofluorescence of the chlorophyll in red, Nomarski differential interference contrast (DIC) in white. A: GFP::Pt5_1-129; B: GFP::Pt5_1-159;C: GFP::Pt5_1-195;D: GFP::Pt5_1-295;E: GFP::Pt5_1-361;F: GFP::Pt5_1-390. Scale bars: 5 µm.

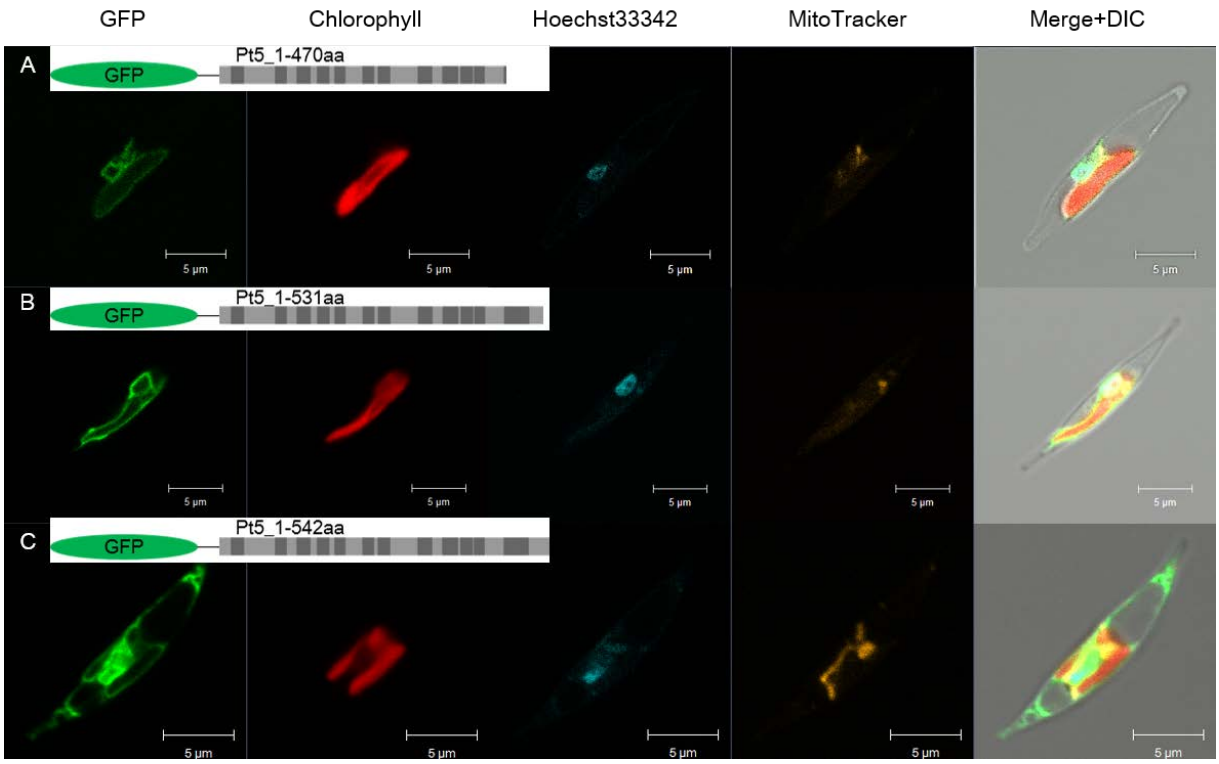


Figure S 11: Localisation of *Pt*NTT5 in *P. tricornutum* (C-terminal deletion, N-terminal GFP-fusion). GFP-fusion constructs with N-terminal fusion of GFP and C-terminal deletion of *Pt*NTT5 (Pt5) transmembrane domains. GFP fluorescence in green; chlorophyll autofluorescence in red; Hoechst 33342 fluorescence in blue; MitoTracker fluorescence in orange, Nomarski differential interference contrast (DIC) in white. A: GFP::Pt5_1-470; B: GFP::Pt5_1-531; and C: GFP::Pt5_1-542. Scale bars: 5 µm.

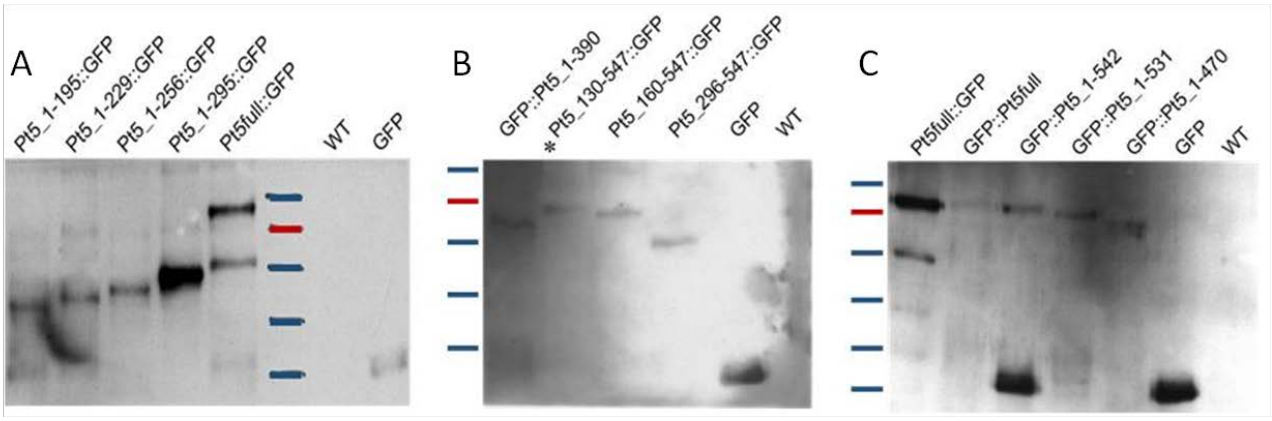


Figure S 12: Expression analysis of selected C-terminal and N-terminal GFP fusions of *Pt*NTT5. Expression of *Pt*NTT5 and selected truncated versions was analysed by Western-blotting and immunodetection with an anti-GFP antibody. (A) Detection of C-terminally truncated *Pt*NTT5 carrying C-terminal GFP. (B) Detection of a C-terminally truncated *Pt*NTT5 carrying N-terminal GFP and of various N-terminally truncated versions of *Pt*NTT5 carrying C-terminal GFP. (C) Detection of C-terminally truncated *Pt*NTT5 carrying N-terminal GFP. Prestained protein ladder shows reference protein sizes in kDa. Numbers of the truncated versions indicate the respective amino acid positions of the full-length *Pt*NTT5 (Pt5full). Pt5 = PtNTT5; WT = wildtype; ∗ = transformants not showing any microscopically detectable GFP-fluorescence. Calculated protein sizes (in kDa): Pt5full::GFP and GFP::Pt5full = 85.3; Pt5_1-195::GFP = 46.8; Pt5_1-229::GFP = 50.7; Pt5_1-256::GFP = 53.3; Pt5_1-295::GFP = 57.3; GFP::Pt5_1-390 = 67.6; Pt5_130-547::GFP = 72.7; Pt5_160-547::GFP = 69.7; Pt5_296-547::GFP = 55.6; GFP::Pt5_1-542 = 84.7; GFP::Pt5_1-531 = 83.4; GFP::Pt5_1-470 = 76.8; GFP = 26.8.

Supporting Information, Chapter 3

Table S 4: List of oligonucleotides used in this study to generate overexpression constructs for *E. coli* and transformation constructs for *P. tricornutum* or *T. pseudonana*

Oligonucleotide	**Sequence 5‘ -> 3‘**
NTT3*Tp*(Nde)_s	CATATGAGGGCACCGACAACGAGCGTAGC
NTT3*Tp*_as	TTATAACTTTGGCTTCTTACCCCGTCC
NTT3*Tp*(Stu)_s	AGGCCTGCCACCATGAGGGCACCGACAACGAGC
NTT3*Tp*(Stu)_as	AGGCCTTAACTTTGGCTTCTTACCCCG
TpNTT3_for	GCCATCATGAGGGCACCGAC
TpNTT3ful_revo	TAACTTTGGCTTCTTACCCCGTCC
TpNTT3pre_rev	GTGGGGTAGTTCTGCTTTGGAG

Supporting Information, Chapter 5

Table S 5: Oligonucleotides used in this study to generate transformation constructs.

Marker protein	Sequence 5' -> 3'
PGK_for	AGTCACACCATGGCTTCCGAC
PGK_rev	TTCTTTGGGTAGCAAGGCGGGC
BIP_for	GAAAGAACATGATGTTCATGAGAATTGCC
BiP_rev	CCAGCCTCTTCGGCACGAATGGAGGGAGC
PGDH_for	ATGTTCCTCACAATCCTACTG
PGDHpre_rev	TATCTCGTTGGTGGAAAGTG
SP1_for	ATGCCAACAATCAGCGATCAC
SP1pre_rev	GTAGGTCTTGACGTCTTCTCGG
PtOEE1_for	ATGAAGTTCACTGCCG
PtOEE1_rev	CTTGCCGACATTAGCG

Table S 6: List of transformation constructs used in this study.

construct	description	source
lhcf1-GFP-PTV	*P. tricornutum* transformation vector pPha-T1 (GenBank AF219942, Zaslavskaia et al. 2000), containing the 442 bp 5'-flanking region of the lhcf1 gene (ID 18049) as promoter and full length eGFP	(Gruber *et al.* 2007)
nr-GFP-PTV-NR	*P. tricornutum* transformation vector pPha-NR (GenBank JN180663.1, Stork et al. 2012), containing the 422 bp 5'-flanking region of the nr gene (ID 54983) (Poulsen & Kröger 2005) as promoter and full length eGFP	(Chu *et al.* 2016a)
GFP1-10 PTV-NR	*P. tricornutum* transformation vector (PTV) (Zaslavskaia *et al.* 2000), containing nr promoter region (Hempel *et al.* 2009, Poulsen & Kröger 2005) and GFP1-10 (Cabantous *et al.* 2005)	(Vugrinec 2011)
GFP11 PTV-NR	*P. tricornutum* transformation vector (PTV) (Zaslavskaia *et al.* 2000), containing nr promoter region (Hempel *et al.* 2009, Poulsen & Kröger 2005) and GFP11 (Cabantous *et al.* 2005)	(Vugrinec 2011)
PGKpre GFP1-10 PTV-NR	presequence of PGK (cytosol marker) fused to GFP1-10, inserted into PTV-NR	This study
PGKpre GFP11 PTV-NR	presequence of PGK (cytosol marker) fused to GFP11, inserted into PTV-NR	This study
BiPpre GFP1-10 PTV-NR	presequence of PtBiP (CER marker) fused to GFP1-10, inserted into PTV-NR	(Vugrinec 2011), modified
BiPpre GFP11 PTV-NR	presequence of PtBiP (CER marker) fused to GFP11, inserted into PTV-NR	(Vugrinec *et al.* 2011)
6PGDHpre GFP1-10 PTV-NR	presequence of 6PGDH (PPS marker) fused to GFP1-10, inserted into PTV-NR	This study
6PGDHpre GFP11 PTV-NR	presequence of 6PGDH (PPS marker) fused to GFP11, inserted into PTV-NR	This study
SP1preGFP1-10 PTV-NR	presequence PtSP1 (IES marker) fused to GFP1-10, inserted into PTV-NR	(Vugrinec *et al.* 2011), modified, site directed mutagenesis
SP1full GFP11 PTV-NR	full length sequence PtSP1 (IES marker) fused to GFP11, inserted into PTV-NR	(Vugrinec *et al.* 2011); site directed mutagenesis
mOEE1pre GFP1-10 PTV-NR	modified presequence of PtBiP (stroma marker) fused to GFP1-10, inserted into PTV-NR	(Vugrinec *et al.* 2011)
mOEE1pre GFP11 PTV-NR	modified presequence of PtBiP (stroma marker) fused to GFP11, inserted into PTV-NR	This study

Table S 7: Fluorescence intensity data of *P. tricornutum* wild type (wt) and GFP-expressing cell lines. Given data show ungated 25% quartile (25%), Median (M), 75% quartile (75%) and interquartile range (IQR) of the fluorescence intensities determined for each population.

Green fluorescence												
	0 h				3 h				6 h			
Sample	25%	**M**	75%	IQR	25%	**M**	75%	IQR	25%	**M**	75%	IQR
wt	4,23	**5,49**	6,97	2,74	5,09	**6,52**	8,19	3,1	4,32	**5,57**	7,05	2,73
lhcf1-GFP	109	**141**	181	72	120	**155**	200	80	106	**137**	178	72
nr-GFP	6,01	**7,74**	9,63	3,62	9,19	**11,3**	13,8	4,61	27	**37**	49,2	22,2
sa-nr-PGK+PGK	4,35	**5,69**	7,27	2,92	5,55	**7,14**	8,97	3,42	7,51	**22,2**	39,1	31,59
sa-nr-6PGDH+6PGDH	3,59	**4,77**	6,18	2,59	4,52	**5,84**	7,34	2,82	9,9	**12,7**	15,7	5,8
sa-nr-mOEE1+PGK	6,17	**7,72**	9,53	3,36	7,54	**9,26**	11,2	3,66	10,6	**13,4**	17,6	7
sa-nr-6PGDH+BiP	4,01	**5,23**	6,71	2,7	5,26	**6,68**	8,25	2,99	6,23	**8,03**	9,96	3,73
	9 h				12 h				15 h			
Sample	25%	**M**	75%	IQR	25%	**M**	75%	IQR	25%	**M**	75%	IQR
wt	4,22	**5,61**	7,33	3,11	4,3	**5,65**	7,36	3,06	4,42	**5,72**	7,24	2,82
lhcf1-GFP	106	**141**	187	81	108	**140**	186	78	111	**141**	185	74
nr-GFP	65,8	**86,3**	116	50,2	98,5	**137**	192	93,5	134	**188**	259	125
sa-nr-PGK+PGK	7,61	**51,5**	85,8	78,19	7,44	**69,9**	123	115,56	7,42	**83,4**	158	150,58
sa-nr-6PGDH+6PGDH	13,2	**18**	23,6	10,4	14,7	**21**	28,8	14,1	16,2	**23,9**	33,6	17,4
sa-nr-mOEE1+PGK	15,7	**22,7**	36,2	20,5	20	**33**	54	34	26,1	**44,9**	70,3	44,2
sa-nr-6PGDH+BiP	6,69	**8,83**	11,1	4,41	6,44	**8,85**	11,6	5,16	6,88	**9,44**	12,4	5,52
	27 h											
Sample	25%	**M**	75%	IQR								
wt	4,39	**5,87**	7,55	3,16								
lhcf1-GFP	100	**127**	162	62								
nr-GFP	233	**308**	404	171								
sa-nr-PGK+PGK	7,68	**190**	301	293,32								
sa-nr-6PGDH+6PGDH	21,4	**29,6**	39,7	18,3								
sa-nr-mOEE1+PGK	67,4	**100**	141	73,6								
sa-nr-6PGDH+BiP	8,41	**12,1**	15,6	7,19								

Red fluorescence												
	0 h				3 h				6 h			
Sample	25%	**M**	75%	IQR	25%	**M**	75%	IQR	25%	**M**	75%	IQR
wt	15,6	**19,5**	24,1	8,5	17	**21,3**	26,7	9,7	17,1	**21,1**	25,8	8,7
lhcf1-GFP	21,4	**27,2**	32,9	11,5	22,3	**28,8**	35,2	12,9	22,5	**29**	35,4	12,9
nr-GFP	18,5	**23,5**	28,6	10,1	20,5	**26,1**	31,7	11,2	20,6	**26,3**	31,5	10,9
sa-nr-PGK+PGK	16,3	**20,7**	25,4	9,1	18,1	**23,1**	28,5	10,4	18,3	**23,1**	28,5	10,2
sa-nr-6PGDH+6PGDH	14,6	**18,4**	22,3	7,7	15,9	**20,2**	24,7	8,8	15	**19,2**	23,6	8,6
sa-nr-mOEE1+PGK	23,6	**29,7**	37,5	13,9	25,7	**32,3**	40,8	15,1	26,3	**32,9**	41,1	14,8
sa-nr-6PGDH+BiP	13,9	**17,5**	21,8	7,9	15	**18,8**	23,7	8,7	15,5	**19,1**	23,9	8,4
	9 h				12 h				15 h			
Sample	25%	**M**	75%	IQR	25%	**M**	75%	IQR	25%	**M**	75%	IQR
wt	18,2	**22,9**	28,1	9,9	18,3	**23**	28,1	9,8	17,6	**21,9**	26,6	9
lhcf1-GFP	23,9	**30,9**	38	14,1	23,7	**30,1**	36,9	13,2	23,3	**29,2**	36	12,7
nr-GFP	20,6	**26,3**	31,8	11,2	19,3	**25**	31	11,7	18,8	**24,2**	30,2	11,4
sa-nr-PGK+PGK	18,4	**23,3**	28,8	10,4	17,9	**22,6**	28,3	10,4	17,5	**21,7**	27,3	9,8
sa-nr-6PGDH+6PGDH	14,6	**18,6**	23,3	8,7	14,2	**17,8**	22,9	8,7	14,2	**17,4**	22,5	8,3
sa-nr-mOEE1+PGK	26	**32,4**	40,2	14,2	25,3	**31,1**	39,3	14	24,7	**29,9**	38,3	13,6
sa-nr-6PGDH+BiP	15,7	**19,2**	23,8	8,1	15,9	**19,7**	24,5	8,6	15,8	**19,2**	23,8	8
	27 h											
Sample	25%	**M**	75%	IQR								
wt	17,2	**21,6**	26,7	9,5								
lhcf1-GFP	22,8	**28,2**	34,7	11,9								
nr-GFP	20,5	**25,7**	31,4	10,9								
sa-nr-PGK+PGK	18,1	**22,2**	27,6	9,5								
sa-nr-6PGDH+6PGDH	15,4	**18,4**	23,1	7,7								
sa-nr-mOEE1+PGK	25,2	**30,3**	37,2	12								
sa-nr-6PGDH+BiP	16,1	**19,9**	24,7	8,6								

PGKpre-GFP1-10
MASDMPKLAPGATRKRNVFDVIEALQKQSAKTILVRVDFNVPMNSDGKITDDSRIRGALPTIKAVVNAKCNAVLVSHMGRPK
LVQKAADDEETRQQRHELSLKPVADHLAKLLDQEVLFGDDCLHAQSTIRELPAEGGGVCLLENLRFYKEEEKNGEDFRKTLA
SYADGYVNDAFGTSHRAHASVAGVPALLadggsgggsGFP1-10

PGKpre-GFP11
MASDMPKLAPGATRKRNVFDVIEALQKQSAKTILVRVDFNVPMNSDGKITDDSRIRGALPTIKAVVNAKCNAVLVSHMGRPK
LVQKAADDEETRQQRHELSLKPVADHLAKLLDQEVLFGDDCLHAQSTIRELPAEGGGVCLLENLRFYKEEEKNGEDFRKTLA
SYADGYVNDAFGTSHRAHASVAGVPALLadggsgggsGFP11

BiPpre-GFP1-10
MIFMRIAVAALALLAAPSIRAEEggsgggsGFP1-10

BiPpre-GFP11
MIFMRIAVAALALLAAPSIRAEEggsgggsGFP11

6PDGHpre-GFP1-10
MFLTILLFAAFCTTTLESIQTLSTNEIadggsgggsGFP1-10

6PGDHpre-GFP11
MFLTILLFAAFCTTTLESIQTLSTNEIadggsgggsGFP11

SP1pre-GFP1-10
MPTISDHTHACRAANLASSPQRRMVLLLALCLSLVAPSVTAFRPSTPPAVRSSVFVSSRTKISWADASSSRRKQILGMSDNK
NDSPVDRKPASEPGLKASLRNKWEDETPASEPDAGFLEGIKRWFNSDEGREDVKTYadggsgggsGFP1-10

SP1full-GFP11
MPTISDHTHACRAANLASSPQRRMVLLLALCLSLVAPSVTAFRPSTPPAVRSSVFVSSRTKISWADASSSRRKQILGMSDNK
NDSPVDRKPASEPGLKASLRNKWEDETPASEPDAGFLEGIKRWFNSDEGREDVKTYFISLFLALLLRFTIIEPRFIPSLSMY
PTFEVGDQLAVEKVTKRIKPFYRTEVVVFQPPQAFRDIVENQYGDKSKGKEALIKRIVAVEGDKVEIKNGKLLINDIEQEEA
YTAEDAQYAFGPVRVPPENVLVLGDNRNHSLDGHIWGFLPTKNVIGRAVFVYWPPWRVGNGGMFadggsgggsGFP11

mOEE1pre-GFP1-10
MKFTAACSLALVASASAFAPIPSVSRTTDLSMSLQKDLANVGKSRMLEadggsgggsGFP1-10

mOEE1pre-GFP11
MKFTAACSLALVASASAFAPIPSVSRTTDLSMSLQKDLANVGKadggsgggsGFP11

Figure S 13: Amino acid sequences of the proteins used in this study. **BOLD**: signal peptide predicted by SignalP3.0 neutral networks and hidden Markov models. UNDERLINED: estimated transit peptide domain. lower case: spacer domain. GREY: conserved motif at signal peptide cleavage site. BLACK: fragments of green fluorescent protein.

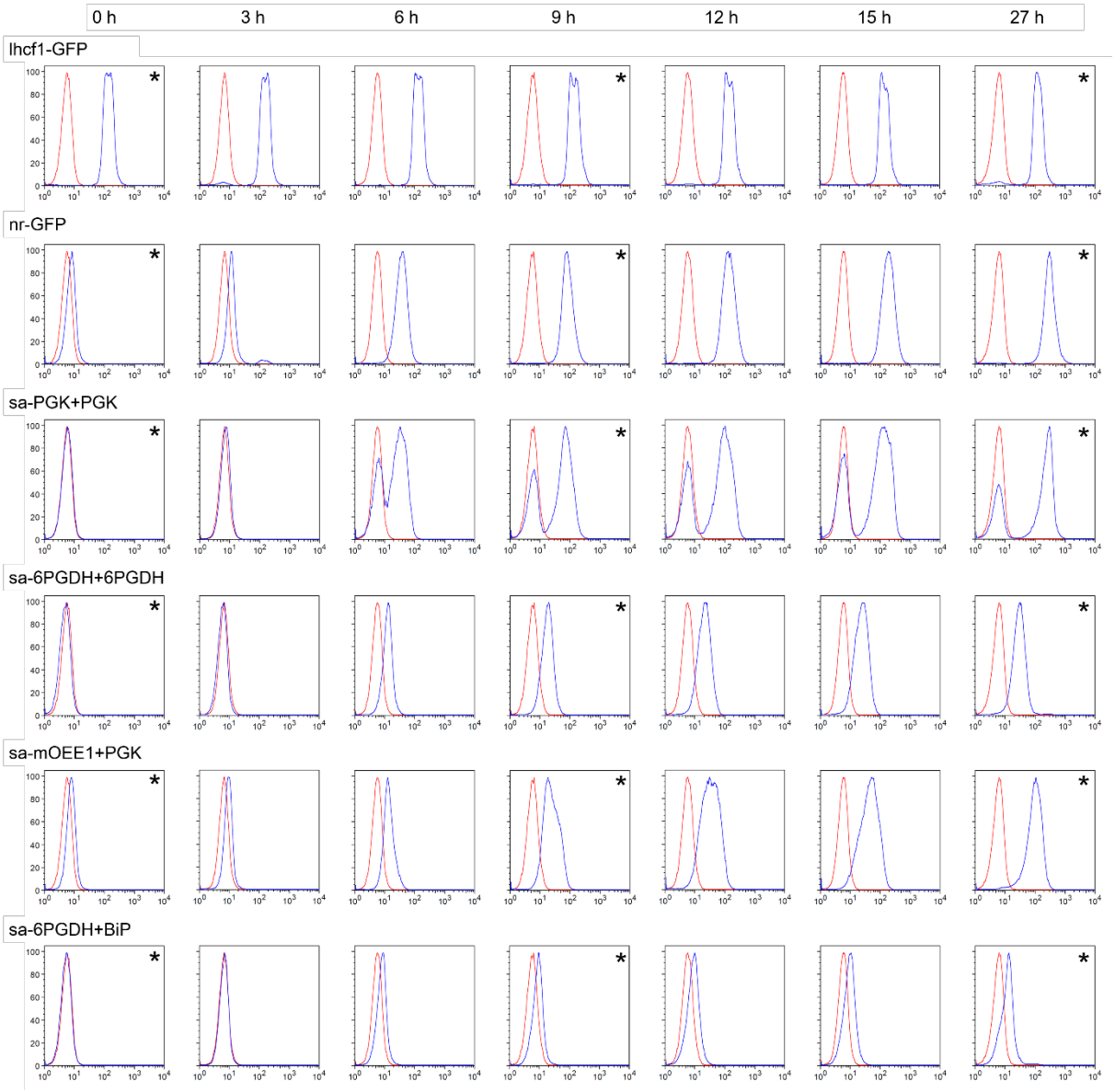


Figure S 14*:* Kinetic of the GFP fluorescence of GFP and self-assembling-GFP transformant cell lines under inducible conditions, measured with flow-cytometry. Green fluorescence of *P. tricornutum* wild type (red) and transformed cell lines (blue). The green fluorescence intensity has been plotted (X-axis) versus the percentage of cell counts detected by scattered light (Y-axis). h = hours; sa = self-assembling; * = corresponding dot plot shown in Figure S 15.

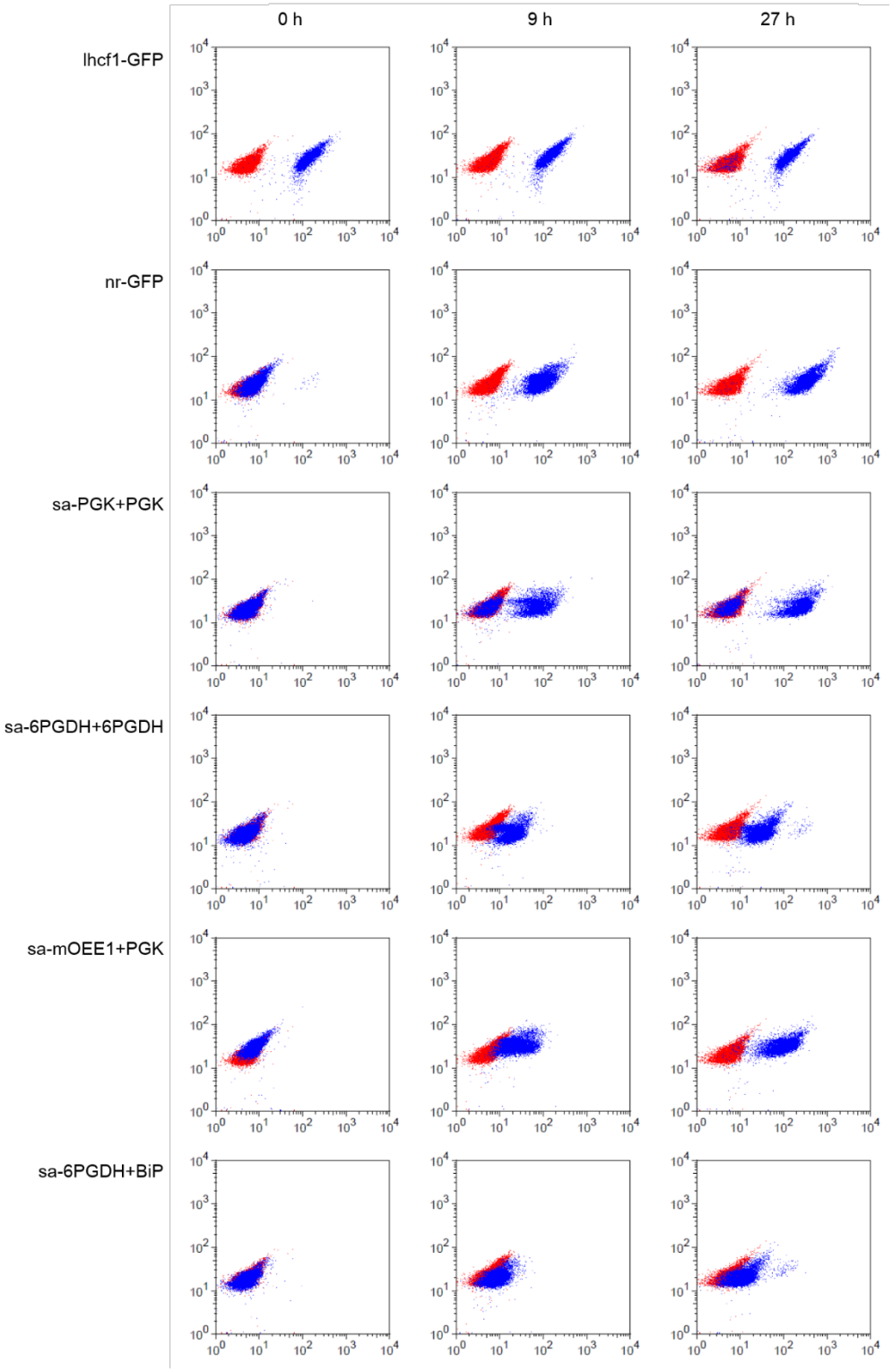


Figure S 15: Corresponding dot plots of the GFP fluorescence signal marked in Figure S 14. The green fluorescence intensity has been plotted (X-axis) versus the autofluorescence of chlorophyll (Y-axis); wild type cell line is shown in red, transformant cell lines in blue.

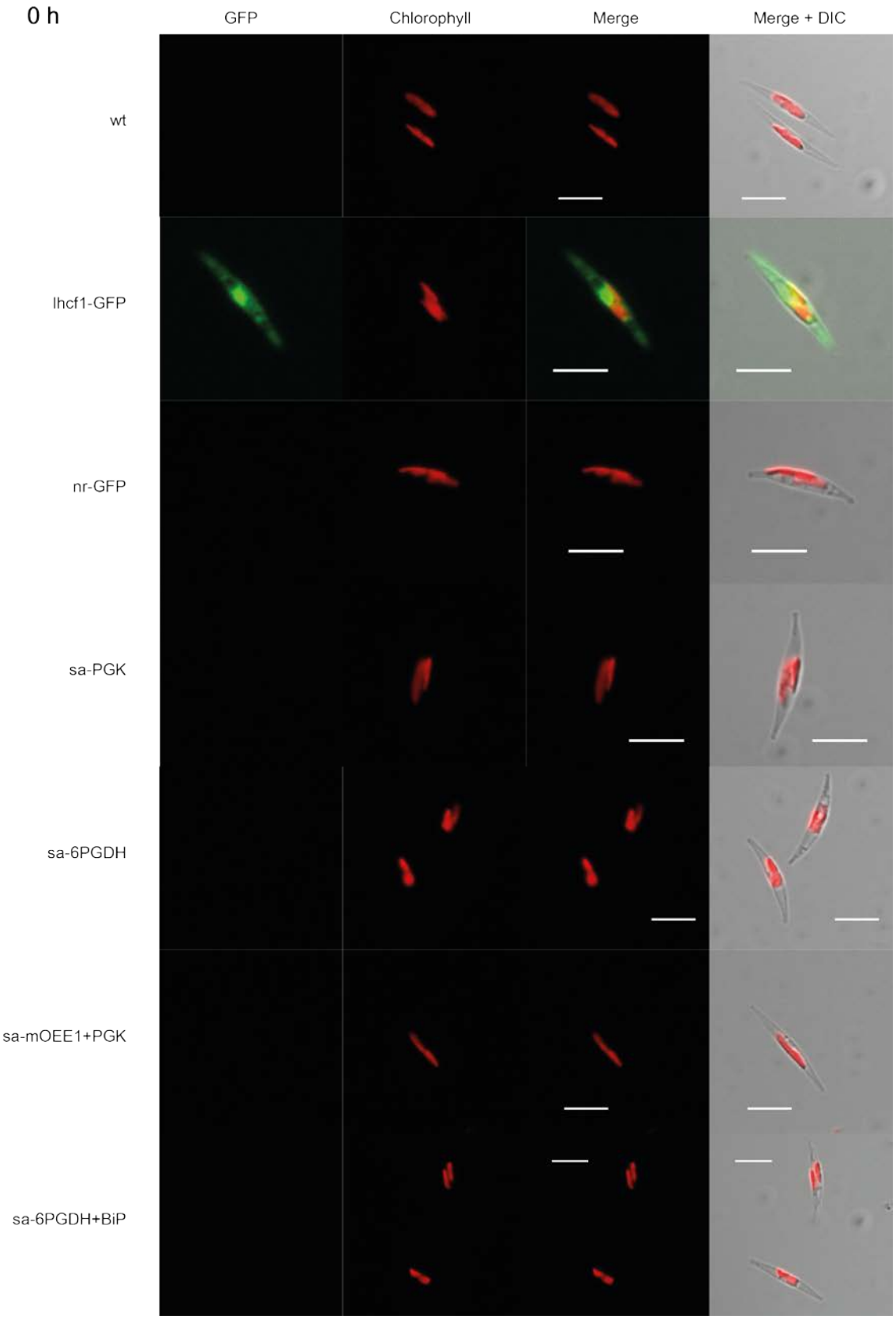


Figure S 16: Analysis of cellular GFP-expression in *P. tricornutum* wild type and transformant cell lines after transfer from NH_4- into NO_3-containing medium. Pictures were taken directly after the transfer with an epifluorescence microscope Olympus BX51 and a Zeiss AxioCam MRm digital camera. Green: GFP fluorescence; red: autofluorescence of the chlorophyll; white: Nomarski differential interference contrast (DIC). (Scale bars: 5 µm)

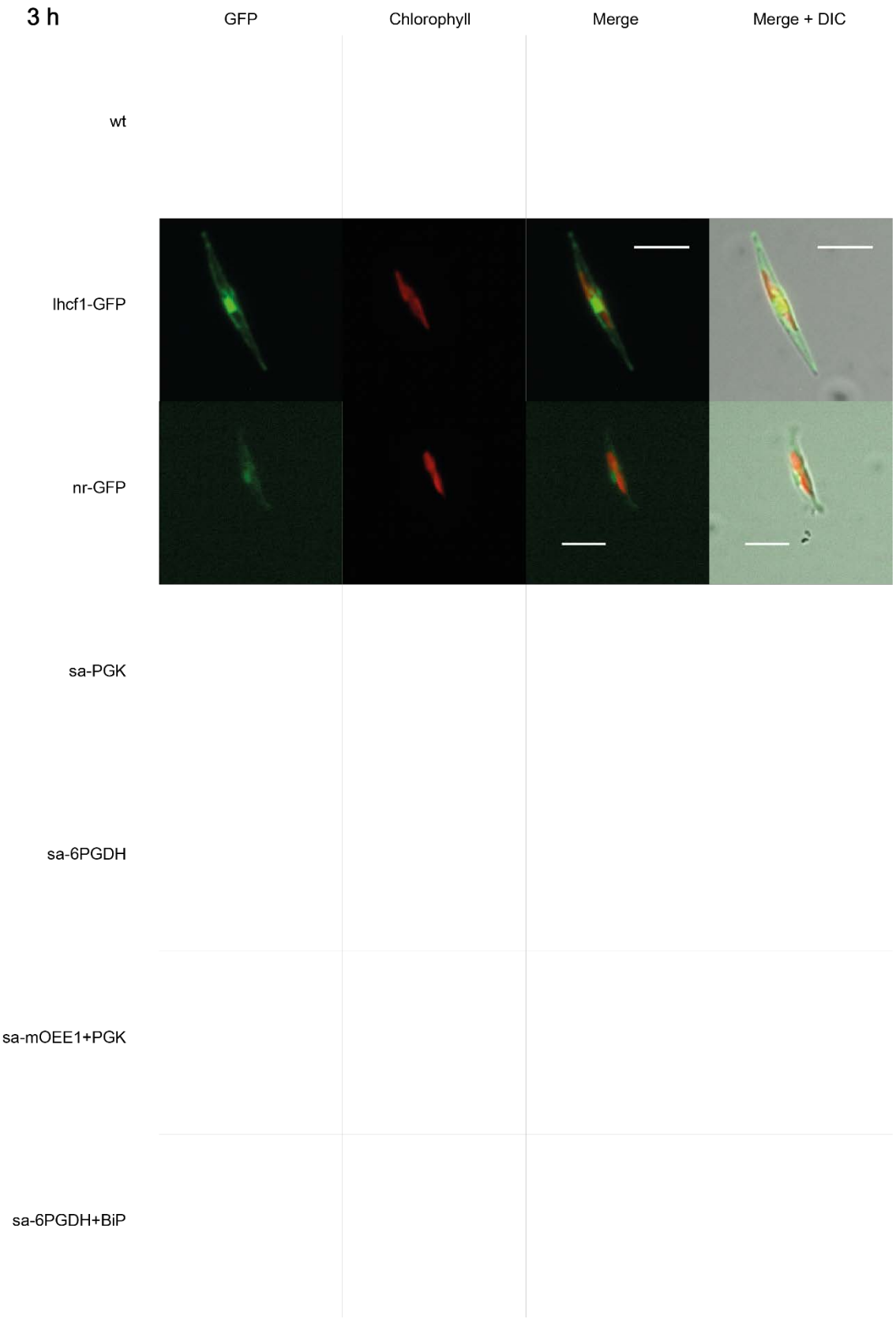


Figure S 17: Analysis of cellular GFP-expression in *P. tricornutum* wild type and transformant cell lines 3 hours after transfer from NH_4- into NO_3-containing medium. Pictures were taken directly after the transfer with an epifluorescence microscope Olympus BX51 and a Zeiss AxioCam MRm digital camera. Green: GFP fluorescence; red: autofluorescence of the chlorophyll; white: Nomarski differential interference contrast (DIC). (Scale bars: 5 µm)

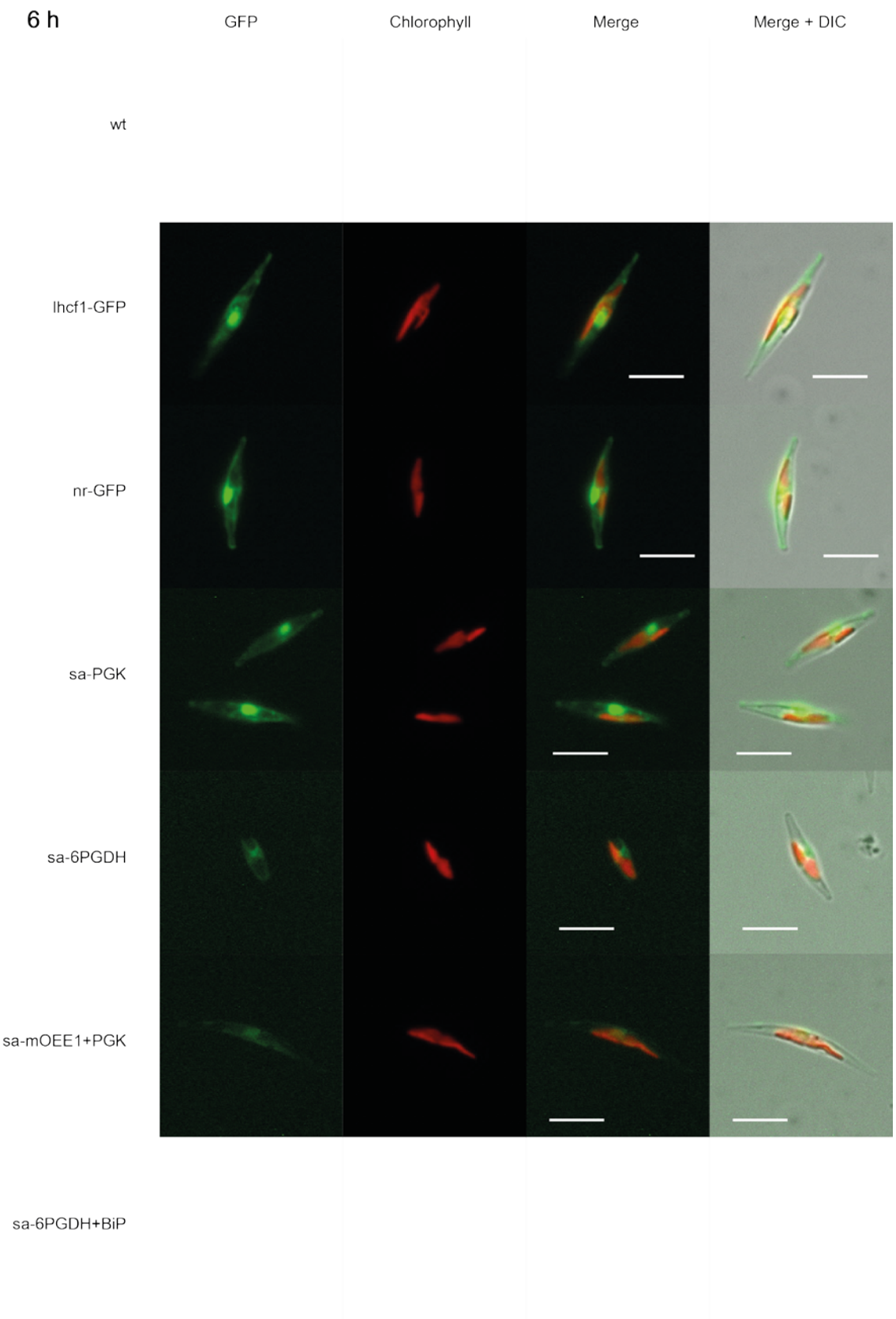


Figure S 18: Analysis of cellular GFP-expression in *P. tricornutum* wild type and transformant cell lines 6 hours after transfer from NH_4- into NO_3-containing medium. Pictures were taken directly after the transfer with an epifluorescence microscope Olympus BX51 and a Zeiss AxioCam MRm digital camera. Green: GFP fluorescence; red: autofluorescence of the chlorophyll; white: Nomarski differential interference contrast (DIC). (Scale bars: 5 µm)

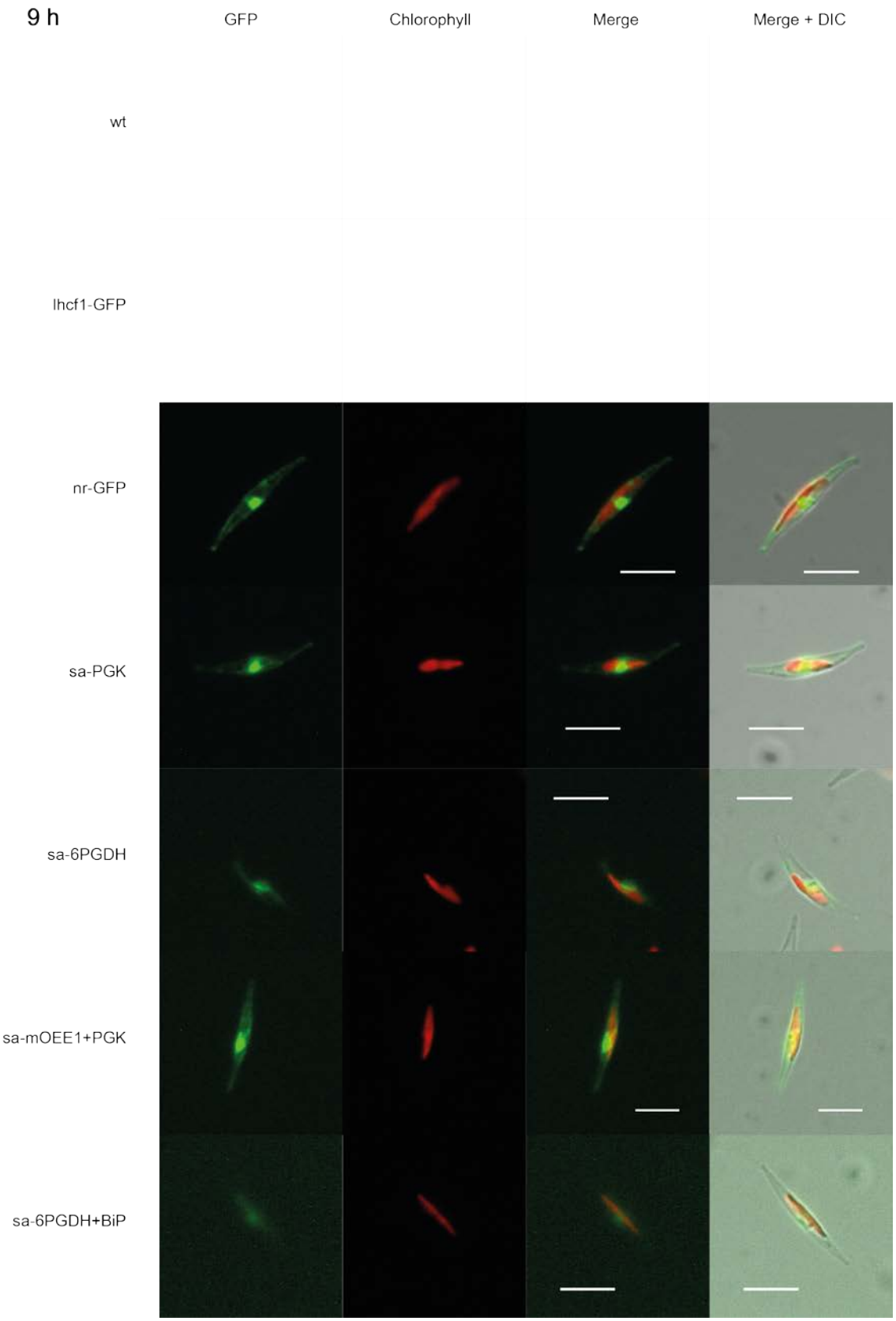


Figure S 19: Analysis of cellular GFP-expression in *P. tricornutum* wild type and transformant cell lines 9 hours after transfer from NH_4- into NO_3-containing medium. Pictures were taken directly after the transfer with an epifluorescence microscope Olympus BX51 and a Zeiss AxioCam MRm digital camera. Green: GFP fluorescence; red: autofluorescence of the chlorophyll; white: Nomarski differential interference contrast (DIC). (Scale bars: 5 µm)

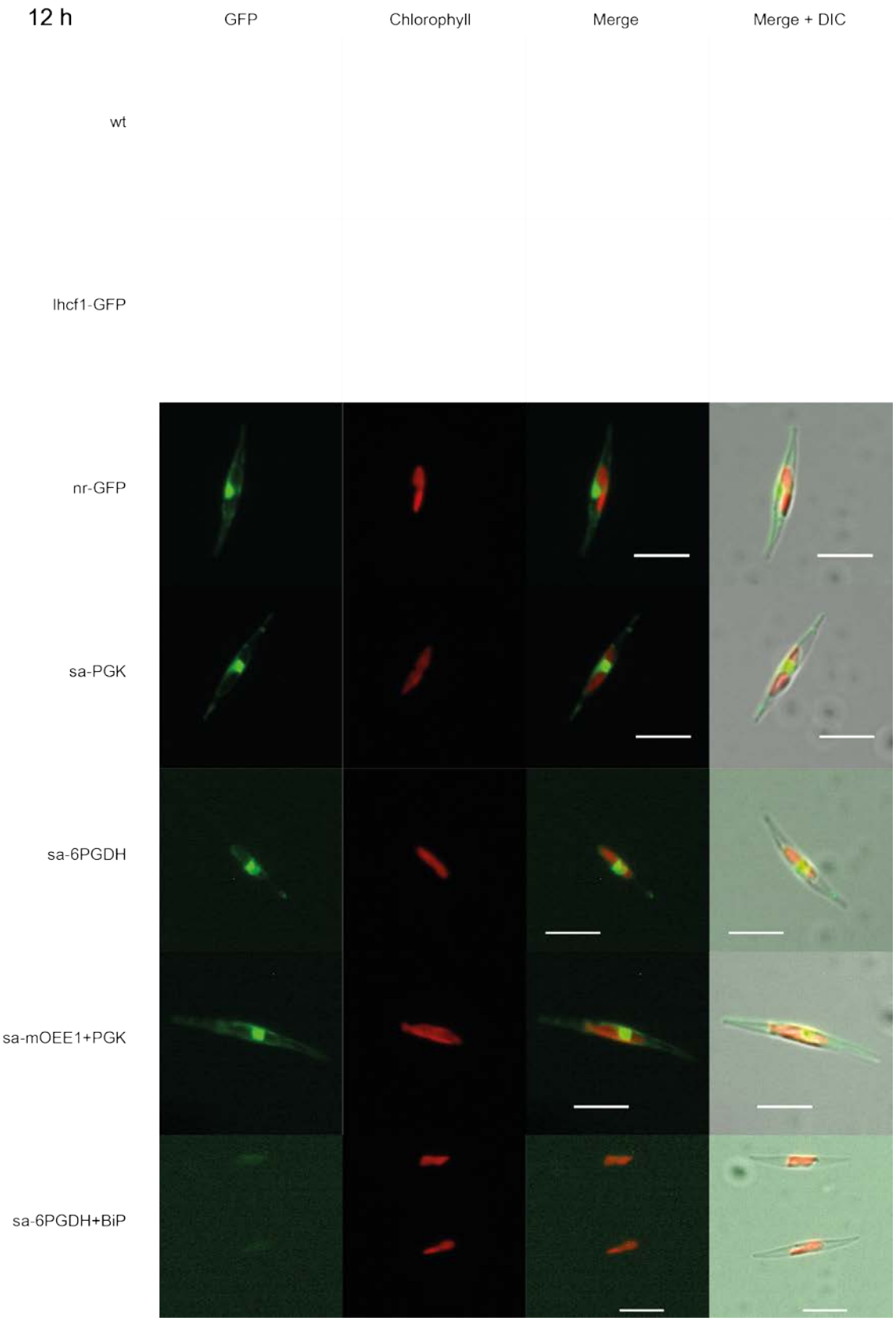


Figure S 20: Analysis of cellular GFP-expression in *P. tricornutum* wild type and transformant cell lines 12 hours after transfer from NH_4- into NO_3-containing medium. Pictures were taken directly after the transfer with an epifluorescence microscope Olympus BX51 and a Zeiss AxioCam MRm digital camera. Green: GFP fluorescence; red: autofluorescence of the chlorophyll; white: Nomarski differential interference contrast (DIC). (Scale bars: 5 µm)

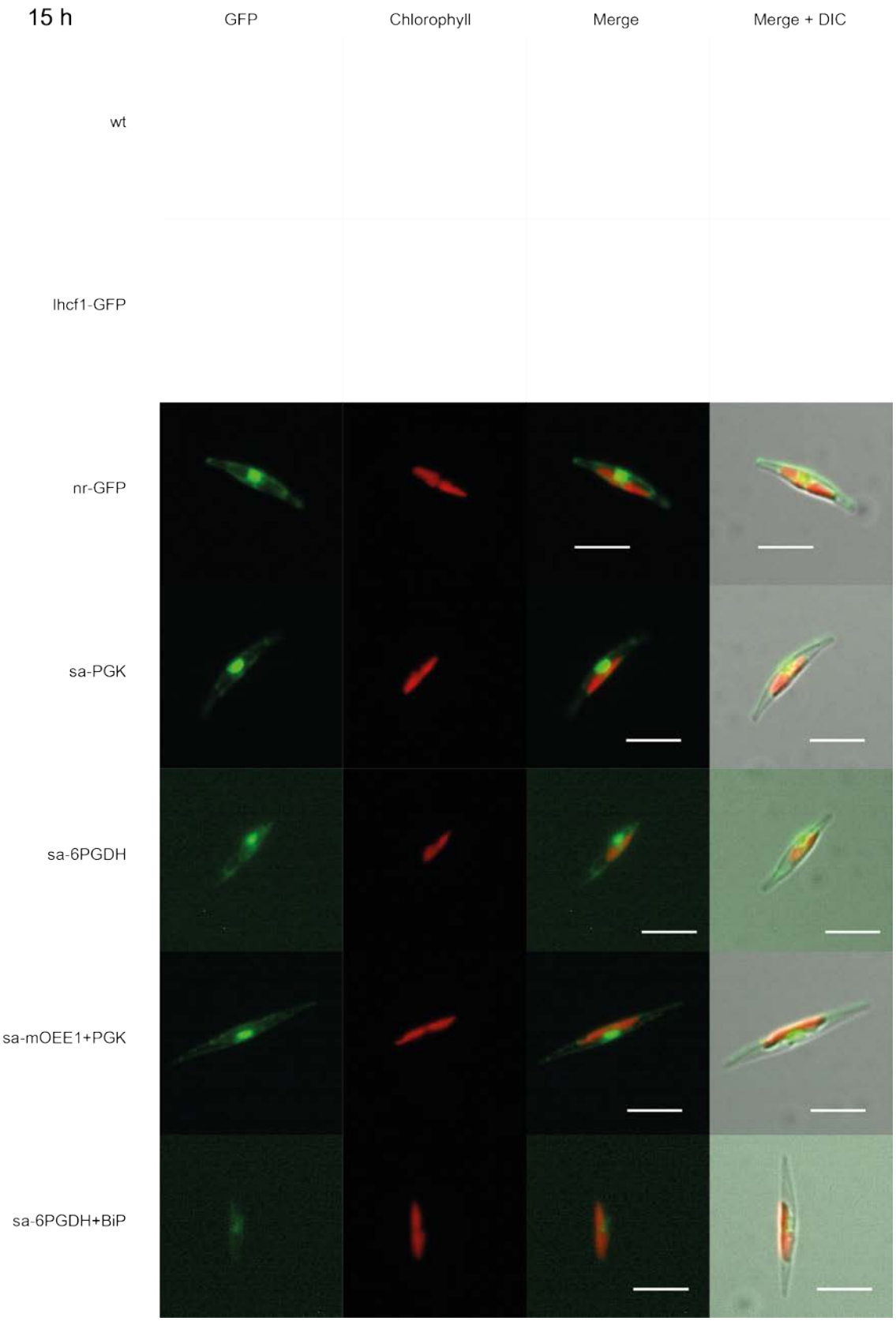


Figure S 21: Analysis of cellular GFP-expression in *P. tricornutum* wild type and transformant cell lines 15 hours after transfer from NH_4- into NO_3-containing medium. Pictures were taken directly after the transfer with an epifluorescence microscope Olympus BX51 and a Zeiss AxioCam MRm digital camera. Green: GFP fluorescence; red: autofluorescence of the chlorophyll; white: Nomarski differential interference contrast (DIC). (Scale bars: 5 µm)

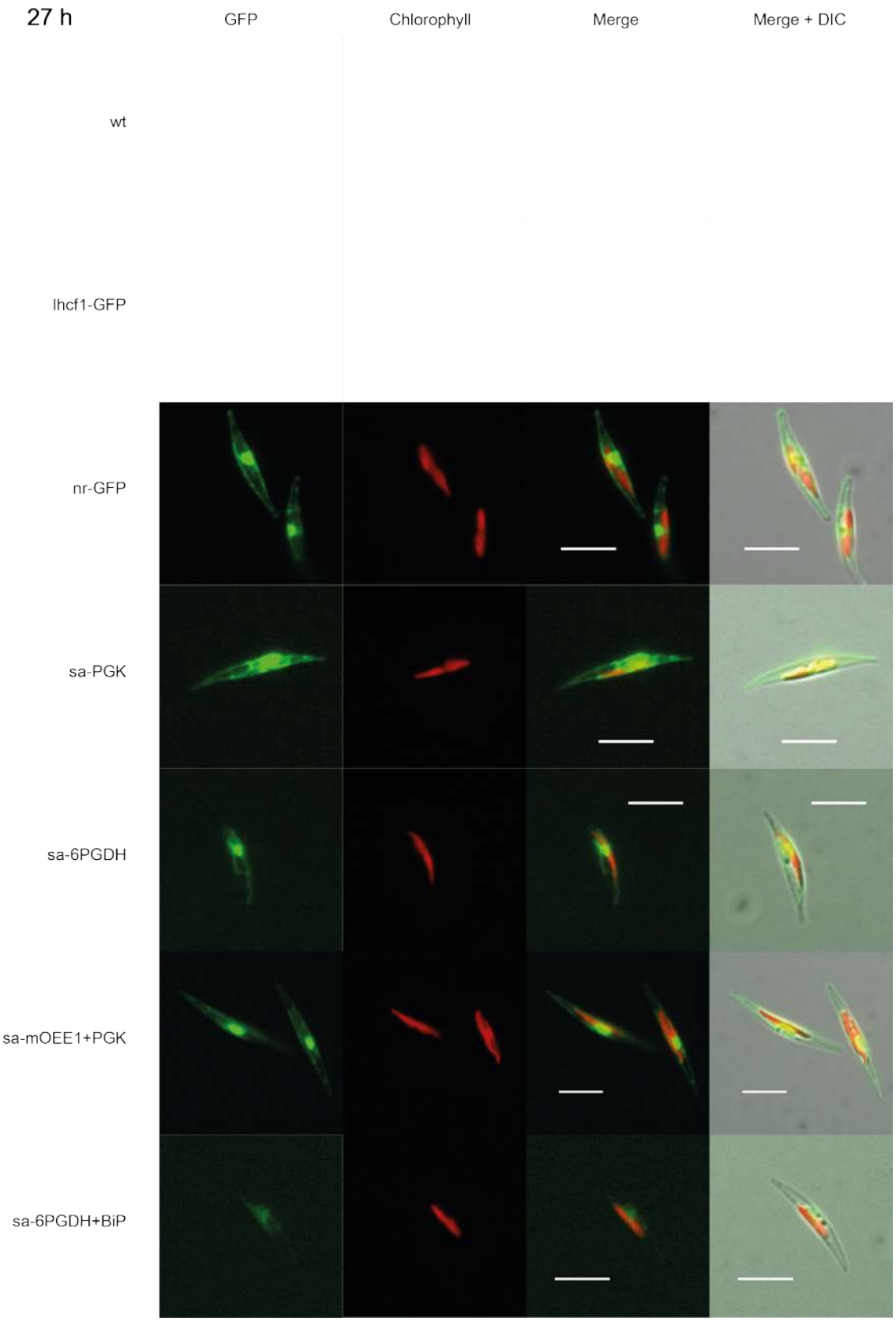


Figure S 22: Analysis of cellular GFP-expression in *P. tricornutum* wild type and transformant cell lines 27 hours after transfer from NH_4- into NO_3-containing medium. Pictures were taken directly after the transfer with an epifluorescence microscope Olympus BX51 and a Zeiss AxioCam MRm digital camera. Green: GFP fluorescence; red: autofluorescence of the chlorophyll; white: Nomarski differential interference contrast (DIC). (Scale bars: 5 µm)

Supporting Information, Chapter 6

Table S 8: Green fluorescence intensity data of *P. tricornutum* wild type (wt) and GFP-expressing cell lines (lhcf1-GFP; nr-GFP_3, _4, _5; _6; _9; _10). Given data show ungated 25% quartile (25%), Median (M), 75% quartile (75%) and interquartile range (IQR) of the green fluorescence intensities determined for each population (100 000 counts). A = medium A (nitrate); B = medium B (ammonium); h = hours.

A) Green fluorescence												
	0 h				3 h				6 h			
Sample	25%	**M**	75%	IQR	25%	**M**	75%	IQR	25%	**M**	75%	IQR
wt_A	5	**6**	8	3	5	**6**	8	3	5	**7**	8	3
lhcf1-GFP_A	157	**199**	250	93	155	**197**	247	92	143	**183**	233	90
nr-GFP_3_A	7	**9**	11	4	7	**9**	11	4	31	**43**	56	24
nr-GFP_4_A	6	**8**	10	4	8	**10**	13	5	44	**56**	70	25
nr-GFP_5_A	7	**9**	11	4	7	**9**	11	4	47	**61**	75	28
nr-GFP_6_A	11	**14**	17	6	13	**17**	22	9	145	**184**	231	86
nr-GFP_9_A	8	**11**	13	5	9	**12**	16	6	128	**162**	205	77
nr-GFP_10_A	18	**23**	30	13	19	**25**	32	13	208	**269**	331	123
	9 h				12 h				15 h			
Sample	25%	**M**	75%	IQR	25%	**M**	75%	IQR	25%	**M**	75%	IQR
wt_A	5	**7**	9	3	6	**7**	9	4	5	**7**	9	3
lhcf1-GFP_A	141	**181**	230	89	138	**175**	225	87	134	**170**	220	86
nr-GFP_3_A	75	**102**	138	63	116	**161**	219	103	179	**240**	314	135
nr-GFP_4_A	84	**106**	135	51	122	**158**	203	81	168	**214**	270	102
nr-GFP_5_A	113	**157**	204	91	182	**251**	321	139	268	**356**	447	179
nr-GFP_6_A	414	**511**	633	219	654	**847**	1073	419	977	**1253**	1531	554
nr-GFP_9_A	378	**461**	558	180	577	**744**	941	364	825	**1073**	1360	535
nr-GFP_10_A	521	**688**	876	355	809	**1099**	1401	592	1213	**1579**	1940	727
	24 h											
Sample	25%	**M**	75%	IQR								
wt_A	6	**7**	9	3								
lhcf1-GFP_A	136	**178**	226	90								
nr-GFP_3_A	321	**414**	534	213								
nr-GFP_4_A	312	**384**	468	156								
nr-GFP_5_A	482	**661**	861	379								
nr-GFP_6_A	1916	**2371**	2833	917								
nr-GFP_9_A	1615	**1961**	2354	739								
nr-GFP_10_A	2356	**3067**	3826	1470								

B) Green fluorescence												
	27 h				30h				33 h			
Sample	25%	**M**	75%	IQR	25%	**M**	75%	IQR	25%	**M**	75%	IQR
wt_B	5	**7**	8	3	5	**7**	8	3	5	**7**	9	3
lhcf1-GFP_B	126	**166**	214	88	126	**163**	208	82	127	**163**	209	82
nr-GFP_3_B	340	**444**	572	232	329	**431**	564	235	304	**401**	532	228
nr-GFP_4_B	338	**418**	511	173	342	**428**	536	194	315	**398**	496	181
nr-GFP_5_B	545	**722**	945	400	536	**699**	921	385	519	**676**	899	380
nr-GFP_6_B	2188	**2700**	3249	1061	2155	**2665**	3272	1117	2091	**2590**	3255	1164
nr-GFP_9_B	1777	**2163**	2552	775	1700	**2144**	2547	847	1601	**2094**	2537	936
nr-GFP_10_B	2703	**3498**	4416	1713	2681	**3481**	4471	1790	2569	**3301**	4323	1754
	36 h				48 h				72 h			
Sample	25%	**M**	75%	IQR	25%	**M**	75%	IQR	25%	**M**	75%	IQR
wt_B	5	**7**	8	3	5	**6**	7	3	4	**6**	7	3
lhcf1-GFP_B	132	**170**	218	86	137	**179**	226	89	109	**147**	191	82
nr-GFP_3_B	279	**367**	491	212	180	**231**	303	123	84	**113**	155	71
nr-GFP_4_B	291	**366**	458	167	182	**221**	273	91	84	**112**	145	61
nr-GFP_5_B	492	**630**	827	335	299	**385**	471	172	130	**168**	227	97
nr-GFP_6_B	1981	**2398**	3046	1065	1447	**1679**	1955	508	630	**796**	1064	434
nr-GFP_9_B	1443	**1859**	2346	903	952	**1142**	1421	469	430	**577**	763	333
nr-GFP_10_B	2424	**3053**	3992	1568	1723	**2054**	2447	724	813	**1077**	1412	599
	96 h				168 h				264 h			
Sample	25%	**M**	75%	IQR	25%	**M**	75%	IQR	25%	**M**	75%	IQR
wt_B	5	**7**	8	3	5	**7**	9	4	5	**7**	8	4
lhcf1-GFP_B	117	**159**	201	84	48	**58**	75	27	12	**17**	22	10
nr-GFP_3_B	61	**79**	106	45	24	**34**	45	21	12	**15**	20	8
nr-GFP_4_B	58	**72**	91	32	28	**41**	58	29	14	**19**	26	12
nr-GFP_5_B	77	**98**	127	50	24	**33**	46	22	9	**12**	15	6
nr-GFP_6_B	440	**522**	669	229	131	**198**	299	168	46	**59**	79	33
nr-GFP_9_B	304	**368**	493	189	103	**157**	234	131	33	**43**	58	26
nr-GFP_10_B	539	**691**	927	388	188	**286**	430	242	59	**77**	109	51

Table S 9: Red fluorescence intensity data of *P. tricornutum* wild type (wt) and GFP-expressing cell lines (lhcf1-GFP; nr-GFP_3, _4, _5; _6; _9; _10). Given data show ungated 25% quartile (25%), Median (M), 75% quartile (75%) and interquartile range (IQR) of the green fluorescence intensities determined for each population (100 000 counts). A = medium A (nitrate); B = medium B (ammonium); h = hours.

A) Red fluorescence												
	0 h				3 h				6 h			
Sample	25%	**M**	75%	IQR	25%	**M**	75%	IQR	25%	**M**	75%	IQR
wt_A	18	**22**	27	10	18	**23**	28	10	20	**25**	31	11
lhcf1-GFP_A	22	**28**	34	12	22	**28**	34	12	22	**29**	36	13
nr-GFP_3_A	20	**26**	32	12	20	**25**	32	12	21	**27**	33	12
nr-GFP_4_A	17	**21**	26	10	17	**21**	27	10	18	**23**	29	11
nr-GFP_5_A	20	**25**	31	11	20	**25**	30	11	20	**26**	32	12
nr-GFP_6_A	20	**25**	30	11	20	**25**	31	11	21	**27**	33	12
nr-GFP_9_A	16	**20**	24	8	16	**20**	24	8	16	**20**	25	9
nr-GFP_10_A	19	**24**	30	11	19	**24**	30	11	20	**25**	31	11
	9 h				12 h				15 h			
Sample	25%	**M**	75%	IQR	25%	**M**	75%	IQR	25%	**M**	75%	IQR
wt_A	21	**26**	32	11	21	**27**	34	13	21	**27**	32	11
lhcf1-GFP_A	23	**30**	37	14	24	**32**	39	15	25	**32**	40	15
nr-GFP_3_A	22	**27**	33	12	23	**29**	36	12	24	**30**	37	12
nr-GFP_4_A	18	**23**	28	10	20	**25**	31	12	21	**26**	33	12
nr-GFP_5_A	21	**26**	32	12	22	**28**	34	12	23	**29**	36	12
nr-GFP_6_A	22	**28**	33	12	22	**28**	34	12	23	**29**	35	12
nr-GFP_9_A	17	**21**	25	9	17	**21**	27	10	18	**23**	29	11
nr-GFP_10_A	20	**26**	31	11	21	**27**	33	12	22	**28**	34	12
	24 h											
Sample	25%	**M**	75%	IQR								
wt_A	22	**28**	34	12								
lhcf1-GFP_A	25	**33**	40	14								
nr-GFP_3_A	25	**31**	37	13								
nr-GFP_4_A	23	**30**	36	13								
nr-GFP_5_A	24	**31**	37	12								
nr-GFP_6_A	23	**29**	35	12								
nr-GFP_9_A	21	**26**	32	11								
nr-GFP_10_A	23	**29**	35	12								

B) Red fluorescence												
	27 h				30 h				33 h			
Sample	25%	**M**	75%	IQR	25%	**M**	75%	IQR	25%	**M**	75%	IQR
wt_B	20	**25**	31	11	20	**25**	31	11	20	**25**	31	10
lhcf1-GFP_B	23	**30**	37	13	24	**30**	37	14	24	**31**	37	13
nr-GFP_3_B	22	**29**	35	13	22	**28**	34	12	22	**28**	34	12
nr-GFP_4_B	21	**27**	34	13	22	**28**	36	14	22	**27**	34	13
nr-GFP_5_B	22	**28**	33	11	21	**27**	32	11	22	**28**	33	11
nr-GFP_6_B	21	**27**	33	12	21	**27**	32	11	22	**27**	33	11
nr-GFP_9_B	20	**25**	30	11	19	**25**	30	10	20	**25**	30	10
nr-GFP_10_B	21	**27**	33	12	21	**27**	33	12	22	**27**	33	11
	36 h				48 h				72 h			
Sample	25%	**M**	75%	IQR	25%	**M**	75%	IQR	25%	**M**	75%	IQR
wt_B	21	**26**	31	10	19	**24**	28	9	18	**23**	29	11
lhcf1-GFP_B	25	**32**	39	14	24	**30**	36	13	23	**30**	37	14
nr-GFP_3_B	23	**29**	35	12	22	**28**	33	11	20	**26**	32	13
nr-GFP_4_B	22	**28**	34	12	21	**26**	31	11	20	**25**	31	11
nr-GFP_5_B	23	**29**	34	11	21	**26**	31	10	21	**28**	33	12
nr-GFP_6_B	23	**28**	33	11	21	**26**	31	10	18	**24**	29	10
nr-GFP_9_B	20	**24**	30	10	18	**23**	27	9	16	**22**	27	10
nr-GFP_10_B	23	**28**	34	11	21	**26**	30	10	19	**25**	30	11
	96 h				168 h				264 h			
Sample	25%	**M**	75%	IQR	25%	**M**	75%	IQR	25%	**M**	75%	IQR
wt_B	22	**28**	35	13	18	**21**	28	10	16	**19**	24	8
lhcf1-GFP_B	28	**36**	43	15	20	**25**	32	12	16	**21**	26	10
nr-GFP_3_B	23	**30**	37	13	19	**23**	29	10	16	**19**	24	8
nr-GFP_4_B	22	**29**	35	13	19	**24**	32	13	16	**19**	26	11
nr-GFP_5_B	23	**29**	36	13	17	**20**	26	9	13	**17**	21	7
nr-GFP_6_B	21	**27**	33	12	17	**20**	25	8	14	**17**	20	6
nr-GFP_9_B	18	**23**	29	11	16	**19**	26	10	13	**16**	19	6
nr-GFP_10_B	22	**29**	35	13	18	**22**	30	11	15	**18**	23	8

Table S 10: pH-determination of the NH_4^+ (ammonium) and NO_3^- (nitrate) media for the cultivation of *P. tricornutum* wild type cell lines in triplicates. The cells were kept in NH_4^+-medium for 2 weeks before inoculation into fresh NH_4^+-medium. After 3 days cells were transferred into NO_3^--medium for 24 hours and subsequent back-transfer into NH_4^+-medium and cultivation for another 10 days (264 hours). Cell were removed by filtration prior to pH-measurements.

Sample description	**pH**
NH_4^+-medium without cells	8,18
	8,13
	8,16
NH_4^+-medium, 3 days cell cultivation	8,35
	8,34
	8,33
NO_3^--medium without cells	8,19
	8,22
	8,21
NO_3^--medium, freshly transferred cells (1 x washing)	8,14
	8,2
	8,18
NO_3^--medium, 24 hours cell cultivation	8,35
	8,78
	8,83
NH_4^+-medium, freshly transferred cells (1 x washing)	8,12
	8,1
	8,12
NH_4^+-medium, 264 hours cell cultivation	7,73
	7,77
	7,76

Table S 11: Pulse Amplitude Modulated (PAM) fluorometry. Wild type cells were cultured in triplicates and washed in the same procedure as the one used for the GFP fluorescence measurements. After washing, cells were resuspended in NH_4^+ or NO_3^- media at a cell density of $2*10^6$ cells/ml. PAM measurements were performed in Plastibrand (BRAND GmbH & Co. KG, Wertheim Germany) PMMA cuvettes with an AquaPen AP-C 100 (Photon Systems Instruments, spol. s r.o., Brno, Czech Republic) using the NPQ 2 protocol with actinic light at 700 µmol photons m^{-1} s^{-1} and saturating flashes at 2100 µmol photons m^{-1} s^{-1}, the blue measuring light was adjusted to 0.0099 µmol photons m^{-1} s^{-1}.

Time	Transfer of medium	Fv/Fm		NPQ	
0 h	$NH_4^+ \rightarrow NH_4^+$	0.565	±0.010	0.346	±0.021
	$NH_4^+ \rightarrow NO_3^-$	0.603	±0.017	0.351	±0.019
24 h	$NO_3^- \rightarrow NO_3^-$	0.623	±0.001	0.512	±0.093
	$NO_3^- \rightarrow NH_4^+$	0.585	±0.006	0.477	±0.036
264 h	$NH_4^+ \rightarrow NH_4^+$	0.284	±0.004	2.305	±0.115

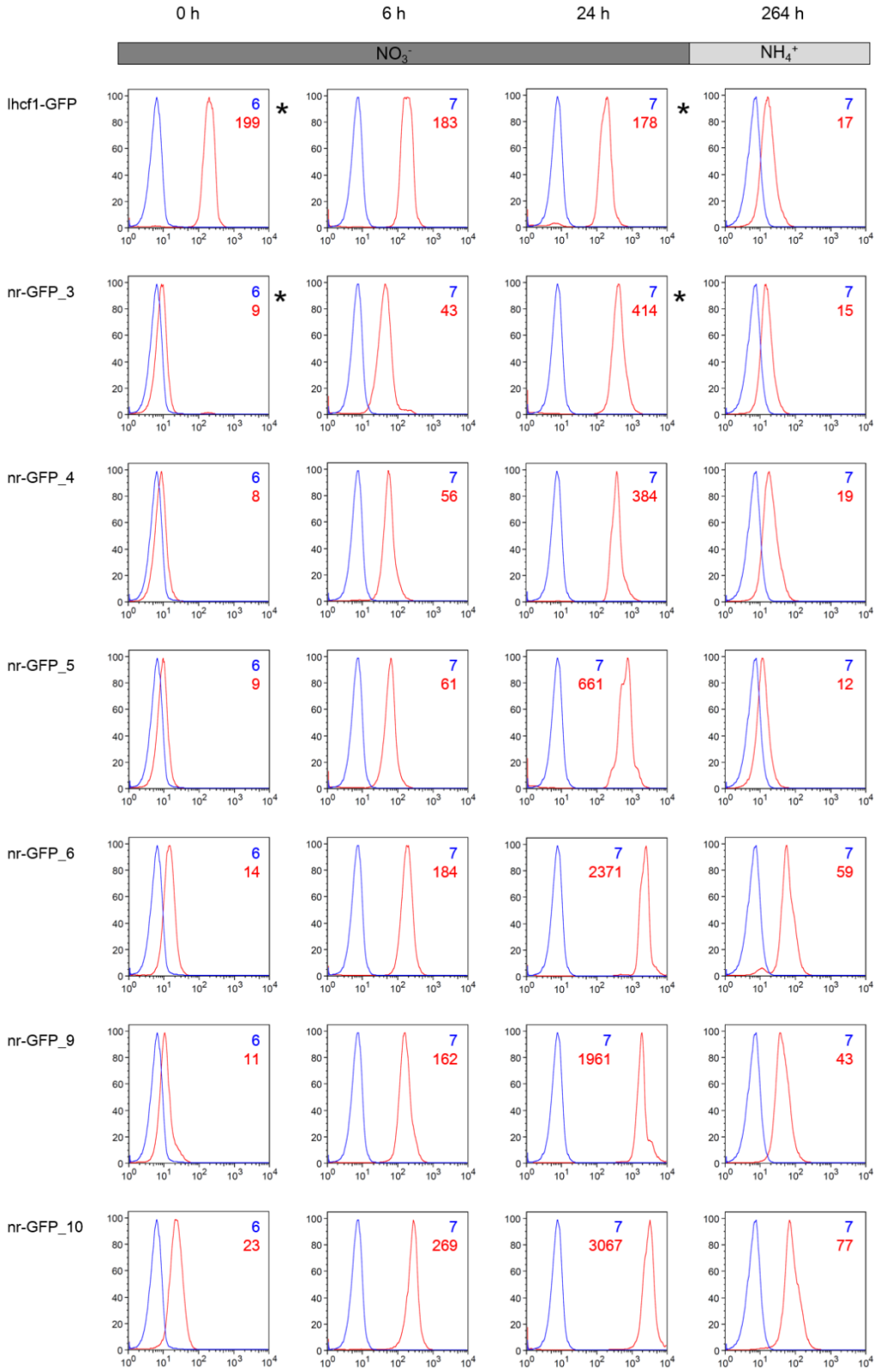


Figure S 23: Histograms showing green fluorescence development of *P. tricornutum* wild type (blue) and genetically transformed cell lines (red) after transfer into nitrate-medium. The green fluorescence intensity has been plotted in log-scale (X-axis) versus cell counts detected by scattered light (Y-axis). Numbers indicate median green fluorescence intensity of 100 000 cells of each transformed cell line (red) and of the wild type cell line (blue). h = hours. * = dot plot shown in Figure S 24

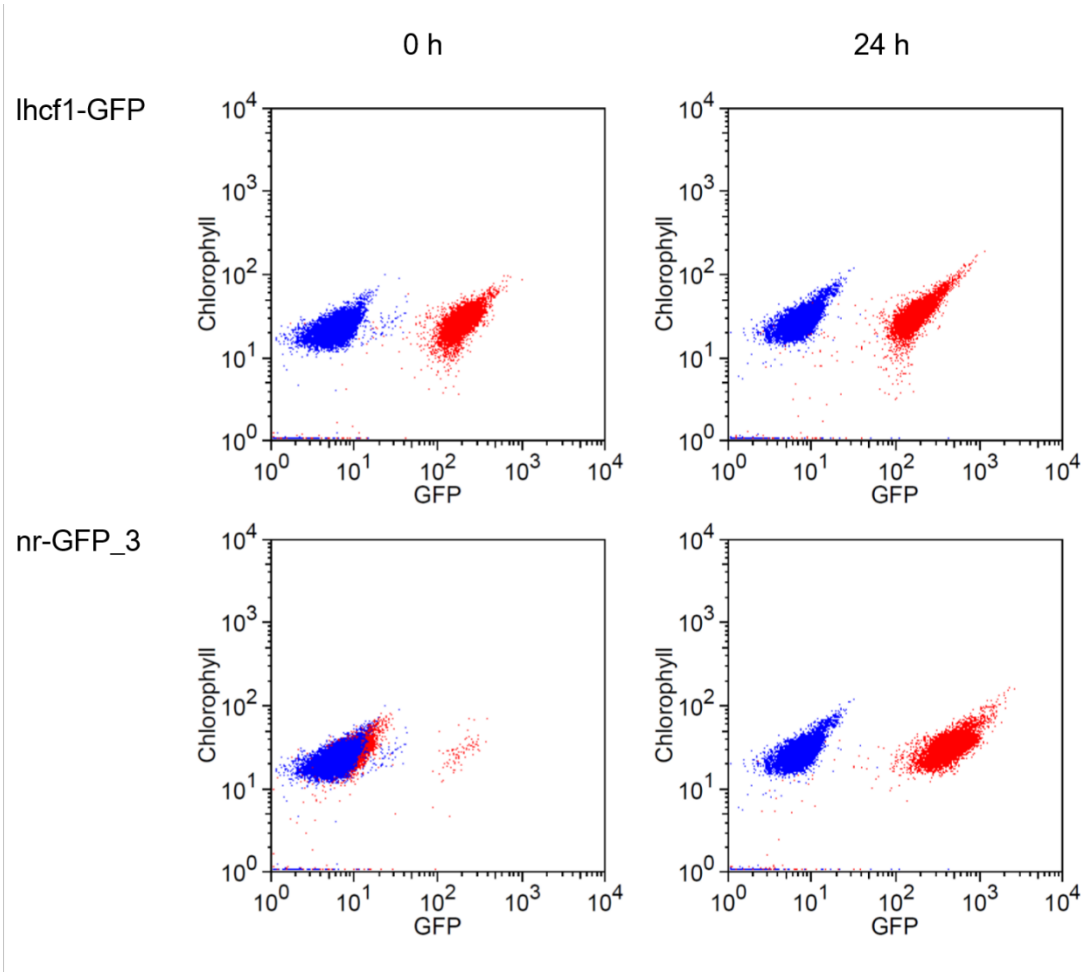


Figure S 24: Dot plots (compare Figure S 23) of *P. tricornutum* cell lines 0 hours (0 h) and 24 hours (24 h) after the cells were transferred into NO_3+-medium. Green fluorescence (X-axis) is plotted versus autofluorescence of chlorophyll (Y-axis). Data of each population (100 000 counts) of transformed cell lines are shown in red for transformed cell lines and in blue for wild type cell line.

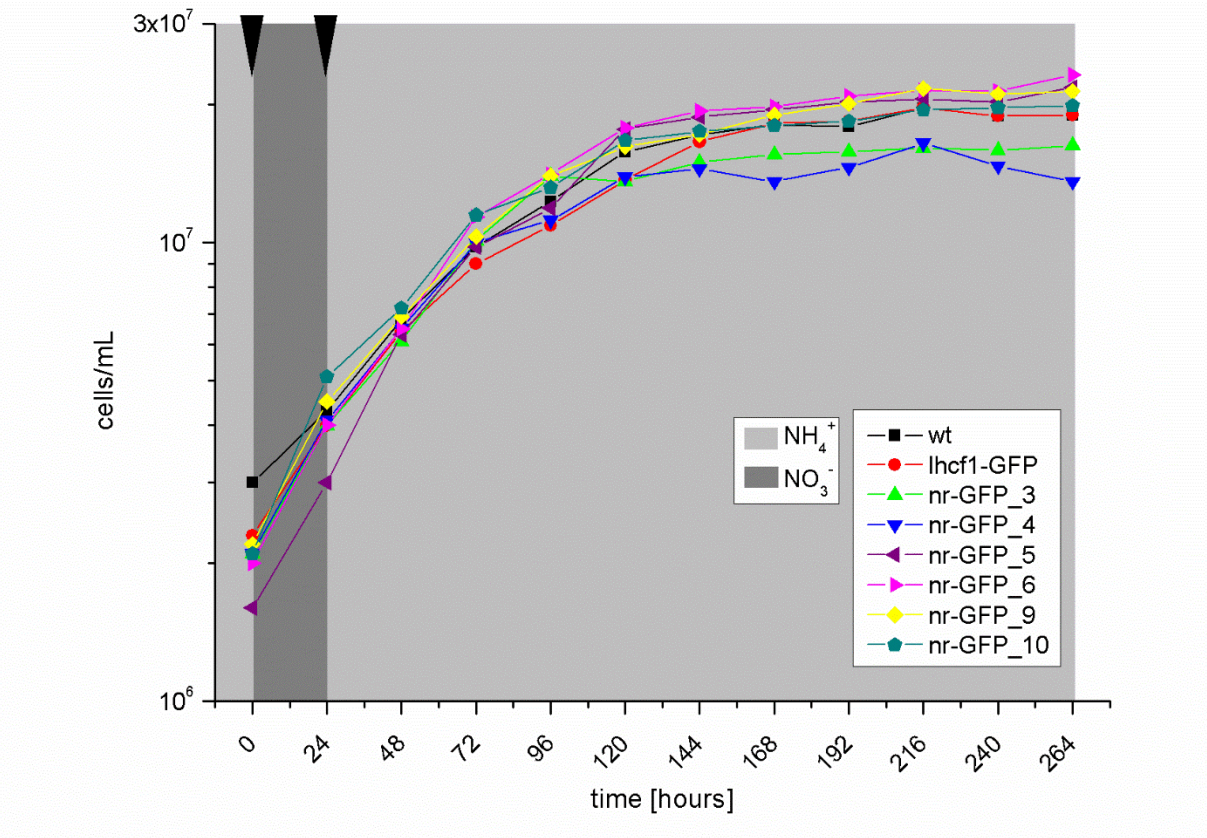


Figure S 25: Cell numbers determined for *P. tricornutum* wild type and transformed cell lines. All cell lines were kept in NH_4^+-medium (light grey) before transferred into NO_3^--medium (dark grey) for 24 hours and subsequent transfer back into NH_4^+-medium. Arrows indicate washing steps and medium change.

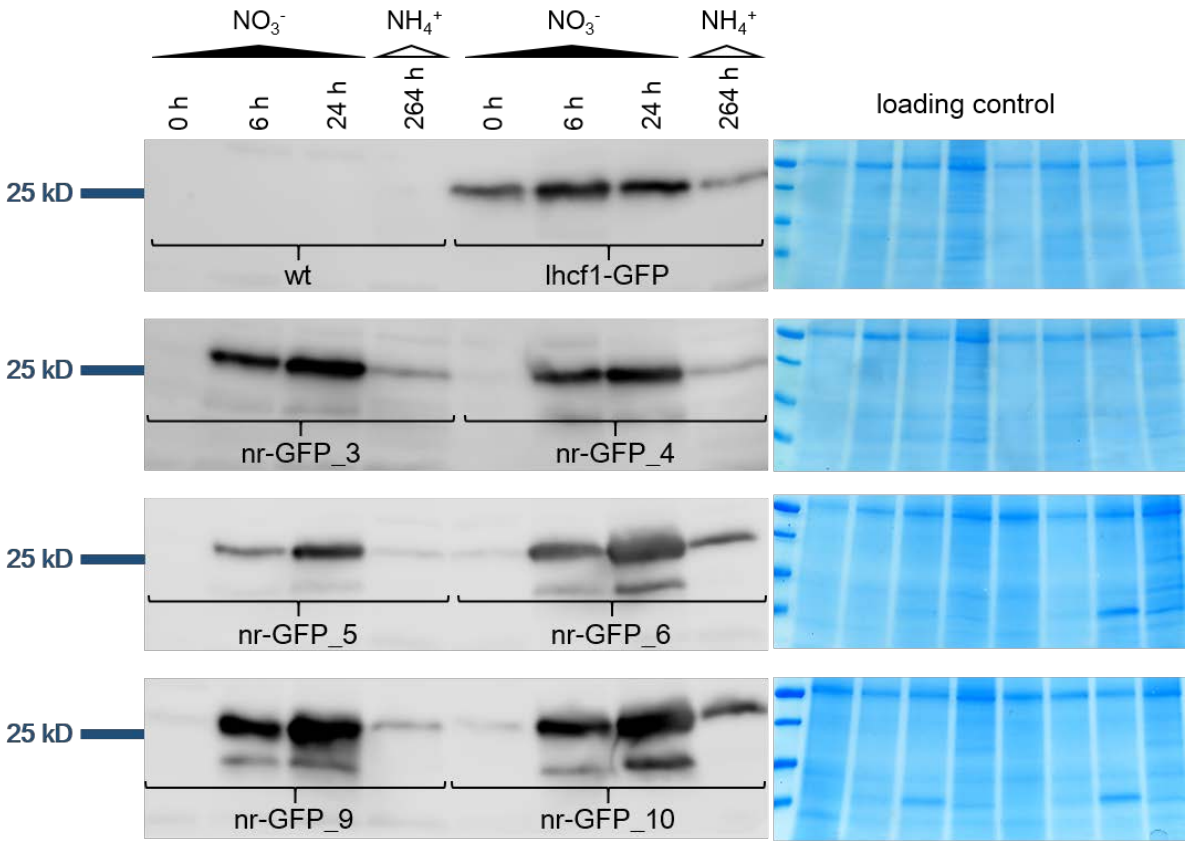


Figure S 26: Analysis of GFP protein expression by western blot. Total protein extract (3 µg) of *P. tricornutum* wild type (wt) and transformant cell lines (lhcf1-GFP; nr-GFP_3, _4, _5, _6, _9, _10) and GFP-antibody were used for detection. Left: western blot; right: loading control. GFP-expression by nr promoter is stronger compared to the lhcf1 promoter. GFP (27 kDa) even appears in the coomassie stained loading control of nr-GFP_6, _9 and _10.

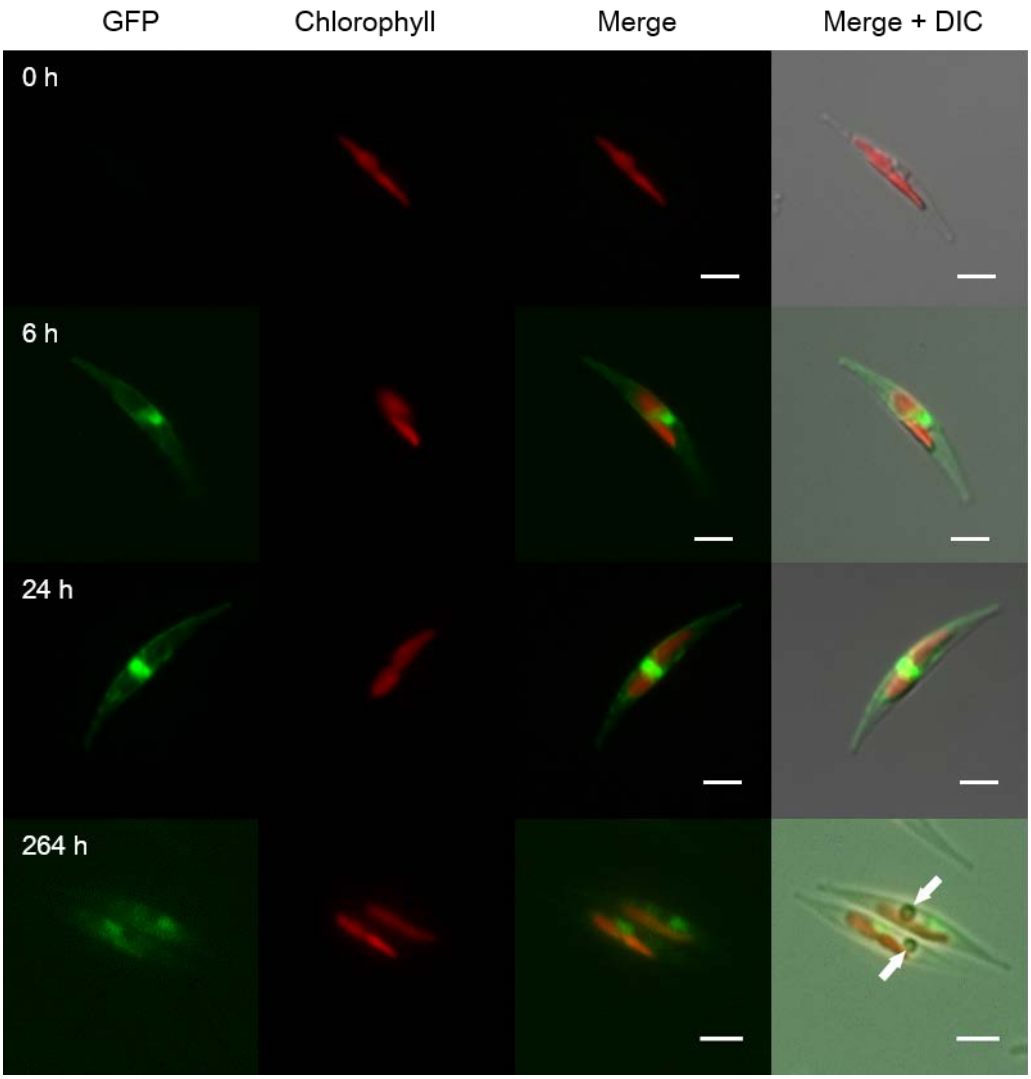


Figure S 27: Fluorescence micrographs of the *P. tricornutum* transformant cell line nr-GFP_3. Images were taken 0 hours (0 h), 6 hours (6 h), 24 hours (24 h) after transfer from NH_4^+-medium into NO_3^--medium, and after back-transfer into NH_4^+-medium (264 h). Arrows indicate lipid droplets. GFP fluorescence is shown in green, autofluorescence of chlorophyll in red, and Nomarski differential interference contrast (DIC) in grey scale. Scale bars: 5 µm.

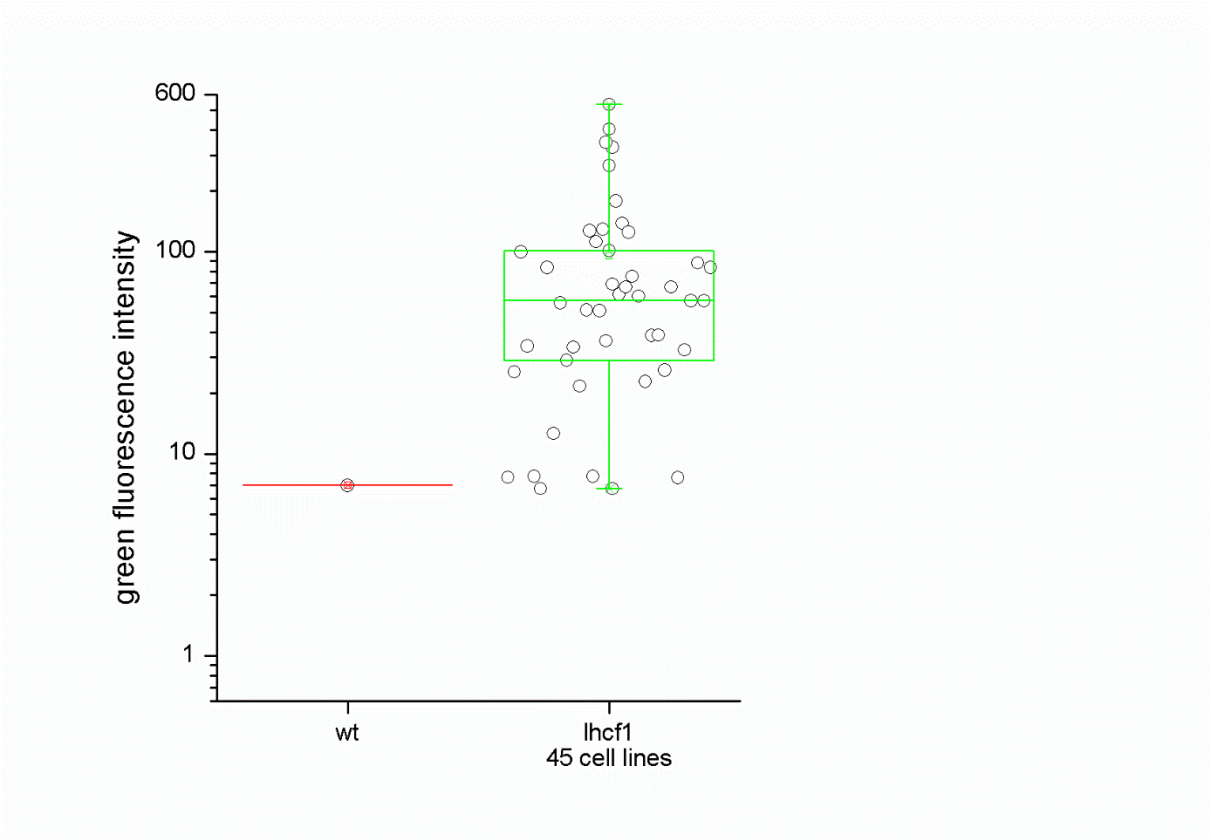


Figure S 28: Box plot of median green fluorescence intensities. The intensities were determined for a *P. tricornutum* wild type cell line (red) and 45 lhcf1-GFP transformed cell lines (green). Whiskers represent min-max-range.

Supplemental Data_S1: raw data exported from the flow cytometer BD FACSCalibur and the FlowJo workspace used for data analyses and figure preparation (see attached CD-ROM).

Supplemental Data_S1_A: raw data exported from the flow cytometer BD FACSCalibur applied for data analyses and preparation for Figure 1 and Tables S1 + S2 for the time period of 0-6 hours.

Supplemental Data_S1_B: raw data exported from the flow cytometer BD FACSCalibur applied for data analyses and preparation for Figure 1 and Tables S1 + S2 for the time period of 6-15 hours.

Supplemental Data_S1_C: raw data exported from the flow cytometer BD FACSCalibur applied for data analyses and preparation for Figure 1 and Tables S1 + S2 for the time period of 24-36 hours.

Supplemental Data_S1_D: raw data exported from the flow cytometer BD FACSCalibur applied for data analyses and preparation for Figure 1 and Tables S1 + S2 for the time point of 48 hours.

Supplemental Data_S1_E: raw data exported from the flow cytometer BD FACSCalibur applied for data analyses and preparation for Figure 1 and Tables S1 + S2 for the time point of 72 hours.

Supplemental Data_S1_F: raw data exported from the flow cytometer BD FACSCalibur applied for data analyses and preparation for Figure 1 and Tables S1 + S2 for the time point of 96 hours.

Supplemental Data_S1_G: raw data exported from the flow cytometer BD FACSCalibur applied for data analyses and preparation for Figure 1 and Tables S1 + S2 for the time point of 168 hours.

Supplemental Data_S1_H: raw data exported from the flow cytometer BD FACSCalibur applied for data analyses and preparation for Figure 1 and Tables S1 + S2 for the time point of 264 hours.

Supplemental Data_S1_I: raw data exported from the flow cytometer BD FACSCalibur applied for data analyses and preparation for Figure 1 and Tables S1 + S2 by using FlowJo.

Supplemental Data_S1_J: raw data exported from the flow cytometer BD FACSCalibur applied for data analyses and preparation for the detailed investigation shown Figure 2 for the time period of 0-2 hours.

Supplemental Data_S1_K: raw data exported from the flow cytometer BD FACSCalibur applied for data analyses and preparation for the detailed investigation shown Figure 2 for the time period of 2-6 hours.

Supplemental Data_S1_L: raw data exported from the flow cytometer BD FACSCalibur applied for data analyses and preparation for the detailed investigation shown Figure 2 for the time period of 6-15 hours.

Supplemental Data_S1_M: raw data exported from the flow cytometer BD FACSCalibur applied for data analyses and preparation for the detailed investigation shown Figure 2 for the time period of 24-26 hours.

Supplemental Data_S1_N: raw data exported from the flow cytometer BD FACSCalibur applied for data analyses and preparation for the detailed investigation shown Figure 2 for the time period of 27-30 hours.

Supplemental Data_S1_O: raw data exported from the flow cytometer BD FACSCalibur applied for data analyses and preparation for the detailed investigation shown Figure 2 for the time period of 31-36 hours.

Supplemental Data_S1_P: raw data exported from the flow cytometer BD FACSCalibur applied for data analyses and preparation for the detailed investigation shown Figure 2 for the time point of 48 hours.

Supplemental Data_S1_Q: raw data exported from the flow cytometer BD FACSCalibur applied for data analyses and preparation for Figure 1 and Tables S1 + S2 for the time period of 0-15 hours by using FlowJo.

Supplemental Data_S1_R: raw data exported from the flow cytometer BD FACSCalibur applied for data analyses and preparation for Figure 1 and Tables S1 + S2 for the time period of 24-36 hours by using FlowJo.

Supplemental Data_S1_S: raw data exported from the flow cytometer BD FACSCalibur applied for data analyses and preparation for Figure 1 and Tables S1 + S2 for the time point of 48 hours by using FlowJo.

B. Author Contributions

Chapter 2

LC designed and carried out the localisation studies and participated in the biochemical characterisation and in drafting and in finalizing the manuscript. AG carried out microscopic analyses and participated in designing the localisation studies and in drafting and finalizing the manuscript. MA participated in the biochemical characterisation. SSE designed and carried out the phylogenetic analyses. HEN participated in designing the biochemical studies. PGK conceived of the study and participated in designing the localisation studies and in finalizing the manuscript. IH conceived of the study, drafted the manuscript and participated in the biochemical characterisation and in finalizing the manuscript. All authors read and approved the manuscript.

Chapter 3

LC designed and carried out the localisation studies in *T. pseudonana* and *P. tricornutum* and participated in drafting and finalising the manuscript. AG designed and carried out the localisation studies in *P. tricornutum* and participated in drafting and finalising the manuscript. ML and MA participated in the biochemical characterisation. HEN participated in designing the biochemical studies. PGK conceived of the study and participated in designing the localisation studies and finalising the manuscript. IH conceived of the study, designed and participated in the biochemical characterisation, drafted and participated in finalising the manuscript. All authors read and approved the manuscript.

Chapter 4

LC designed and carried out the labelling experiment, drafted and participated in finalising the manuscript. JB carried out the labelling experiment and participated in finalising the manuscript. JDRS carried out the labelling experiment and participated in the design of the labelling experiment. AKS, VS, AN carried out the synthesis of Ac_4ManNAz and participated in the design of the labelling experiment. VW participated in designing and conceived of the study. AG participated in finalising the manuscript and conceived of the study. PGK conceived of the study, and participated on its design and coordination. All authors read and approved the final manuscript.

Chapter 5

LC designed and carried out the study, drafted and participated in finalising the manuscript. AG participated in designing the study and finalising the manuscript. PGK conceived of the study and participated in finalising the manuscript.

Chapter 6

LC conceived and designed the experiments, performed the experiments, analysed the data, wrote the paper, prepared figures and/or tables, reviewed drafts of the paper. DE performed the experiments, analysed the data, wrote the paper, reviewed drafts of the paper. CRB performed the experiments, contributed reagents/materials/analysis tools, reviewed drafts of the paper. PGK conceived and designed the experiments, analysed the data, wrote the paper, reviewed drafts of the paper. AG conceived and designed the experiments, performed the experiments, analysed the data, wrote the paper, reviewed drafts of the paper.

C. List of Publications

Chu L, Gruber A, Ast M, Schmitz-Esser S, Altensell J, Neuhaus HE, Kroth PG & Haferkamp I (2016b): Shuttling of (deoxy-) purine nucleotides between compartments of the diatom *Phaeodactylum tricornutum*. New Phytol.

doi: 10.1111/nph.14126

Chu L, Ewe D, Río Bártulos C, Kroth PG & Gruber A (2016a): Rapid induction of GFP expression by the nitrate reductase promoter in the diatom *Phaeodactylum tricornutum*. PeerJ 4, e2344.

doi: 10.7717/peerj.2344

Bibliography

Abell BM, Rabu C, Leznicki P, Young JC & High S (2007): Post-translational integration of tail-anchored proteins is facilitated by defined molecular chaperones. *Journal of Cell Science* **120**, 1743-1751.

Agard NJ, Prescher JA & Bertozzi CR (2004): A Strain-Promoted [3 + 2] Azide–Alkyne Cycloaddition for Covalent Modification of Biomolecules in Living Systems. *J Am Chem Soc* **126**, 15046-15047.

Aline RF, Reeves CD, Russo AF & Volcani BE (1984): Role of Silicon in Diatom Metabolism : Cyclic Nucleotide Levels, Nucleotide Cyclase, and Phosphodiesterase Activities during Synchronized Growth of Cylindrotheca fusiformis. *Plant Physiology* **76**, 674-679.

Alipanah L, Rohloff J, Winge P, Bones AM & Brembu T (2015): Whole-cell response to nitrogen deprivation in the diatom Phaeodactylum tricornutum. *J Exp Bot* **66**, 6281-6296.

Apt KE, Kroth-Pancic PG & Grossman AR (1996): Stable nuclear transformation of the diatom Phaeodactylum tricornutum. *Mol Gen Genet* **252**, 572-579.

Apt KE, Zaslavkaia L, Lippmeier JC, Lang M, Kilian O, Wetherbee R, Grossman AR & Kroth PG (2002): In vivo characterization of diatom multipartite plastid targeting signals. *Journal of Cell Science* **115**, 4061-4069.

Archibald JM (2007): Nucleomorph genomes: structure, function, origin and evolution. *Bioessays* **29**, 392-402.

Archibald JM (2015): Genomic perspectives on the birth and spread of plastids. *Proc Natl Acad Sci U S A* **112**, 10147-10153.

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

Armbrust EV, Berges JA, Bowler C, Green BR, Martinez D, Putnam NH, Zhou S, Allen AE, Apt KE, Bechner M, Brzezinski MA, Chaal BK, Chiovitti A, Davis AK, Demarest MS, Detter JC, Glavina T, Goodstein D, Hadi MZ, Hellsten U, Hildebrand M, Jenkins BD, Jurka J, Kapitonov VV, Kröger N, Lau WW, Lane TW, Larimer FW, Lippmeier JC, Lucas S, Medina M, Montsant A, Obornik M, Parker MS, Palenik B, Pazour GJ, Richardson PM, Rynearson TA, Saito MA, Schwartz DC, Thamatrakoln K, Valentin K, Vardi A, Wilkerson FP & Rokhsar DS (2004): The genome of the diatom Thalassiosira pseudonana: ecology, evolution, and metabolism. *Science* **306**, 79-86.

Ast M, Gruber A, Schmitz-Esser S, Neuhaus HE, Kroth PG, Horn M & Haferkamp I (2009): Diatom plastids depend on nucleotide import from the cytosol. *Proc Natl Acad Sci U S A* **106**, 3621-3626.

Audia JP & Winkler HH (2006): Study of the five Rickettsia prowazekii proteins annotated as ATP/ADP translocases (Tlc): Only Tlc1 transports ATP/ADP, while Tlc4 and Tlc5 transport other ribonucleotides. *J Bacteriol* **188**, 6261-6268.

Baiet B, Burel C, Saint-Jean B, Louvet R, Menu-Bouaouiche L, Kiefer-Meyer MC, Mathieu-Rivet E, Lefebvre T, Castel H, Carlier A, Cadoret JP, Lerouge P & Bardor M (2011): N-glycans of Phaeodactylum tricornutum diatom and functional characterization of its N-acetylglucosaminyltransferase I enzyme. *J Biol Chem* **286**, 6152-6164.

Bailleul B, Berne N, Murik O, Petroutsos D, Prihoda J, Tanaka A, Villanova V, Bligny R, Flori S, Falconet D, Krieger-Liszkay A, Santabarbara S, Rappaport F, Joliot P, Tirichine L, Falkowski PG, Cardol P, Bowler C & Finazzi G (2015): Energetic coupling between plastids and mitochondria drives CO2 assimilation in diatoms. *Nature* **524**, 366-369.

Beckmann HSG & Wittmann V (2010) Azides in Carbohydrate Chemistry. In *Organic Azides*. John Wiley & Sons, Ltd, pp. 469-490.

Bhaya D & Grossman A (1991): Targeting proteins to diatom plastids involves transport through an endoplasmic reticulum. *Mol Gen Genet* **229**, 400-404.

Bhaya D & Grossman AR (1993): Characterization of gene clusters encoding the fucoxanthin chlorophyll proteins of the diatom Phaeodactylum tricornutum. *Nucleic Acids Research* **21**, 4458-4446.

Bionda T, Gross LE, Becker T, Papasotiriou DG, Leisegang MS, Karas M & Schleiff E (2016): Eukaryotic Hsp70 chaperones in the intermembrane space of chloroplasts. *Planta* **243**, 733-747.

Bionda T, Tillmann B, Simm S, Beilstein K, Ruprecht M & Schleiff E (2010): Chloroplast import signals: the length requirement for translocation in vitro and in vivo. *J Mol Biol* **402**, 510-523.

Borgese N, Brambillasca S & Colombo S (2007): How tails guide tail-anchored proteins to their destinations. *Curr Opin Cell Biol* **19**, 368-375.

Borgese N, Colombo S & Pedrazzini E (2003): The tale of tail-anchored proteins: coming from the cytosol and looking for a membrane. *J Cell Biol* **161**, 1013-1019.

Borgese N & Fasana E (2011): Targeting pathways of C-tail-anchored proteins. *Biochim Biophys Acta* **1808**, 937-946.

Borowitzka LJ & Volcani BE (1977): Role of silicon in diatom metabolism. VIII. Cyclic AMP and cyclic GMP in synchronized cultures of Cylindrotheca fusiformis. *Arch Microbiol* **112**, 147-152.

Bowler C, Allen AE, Badger JH, Grimwood J, Jabbari K, Kuo A, Maheswari U, Martens C, Maumus F, Otillar RP, Rayko E, Salamov A, Vandepoele K, Beszteri B, Gruber A, Heijde M, Katinka M, Mock T, Valentin K, Verret F, Berges JA, Brownlee C, Cadoret JP, Chiovitti A, Choi CJ, Coesel S, De Martino A, Detter JC, Durkin C, Falciatore A, Fournet J, Haruta M, Huysman MJ, Jenkins BD, Jiroutova K, Jorgensen RE, Joubert Y, Kaplan A, Kröger N, Kroth PG, La Roche J, Lindquist E, Lommer M, Martin-Jezequel V, Lopez PJ, Lucas S, Mangogna M, McGinnis K, Medlin LK, Montsant A, Oudot-Le Secq MP, Napoli C, Obornik M, Parker MS, Petit JL, Porcel BM, Poulsen N, Robison M, Rychlewski L, Rynearson TA, Schmutz J, Shapiro H, Siaut M, Stanley M, Sussman MR, Taylor AR, Vardi A, von Dassow P, Vyverman W, Willis A, Wyrwicz LS, Rokhsar DS, Weissenbach J, Armbrust EV, Green BR, Van de Peer Y & Grigoriev IV (2008): The Phaeodactylum genome reveals the evolutionary history of diatom genomes. *Nature* **456**, 239-244.

Bozarth A, Maier UG & Zauner S (2009): Diatoms in biotechnology: modern tools and applications. *Appl Microbiol Biotechnol* **82**, 195-201.

Brakemann T, Becker F, Kroth P & Rhiel E (2008): Structural and functional characterization of putative regulatory DNA sequences of Fcp genes in the centric diatom Cyclotella cryptica. *Diatom Research* **23**, 31-49.

Buck J (2016): Improving Strategies for Genome Editing in Diatoms. *Universität Konstanz* **Master's Thesis**.

Bullmann L, Haarmann R, Mirus O, Bredemeier R, Hempel F, Maier UG & Schleiff E (2010): Filling the gap, evolutionarily conserved Omp85 in plastids of chromalveolates. *J Biol Chem* **285**, 6848-6856.

Cabantous S, Terwilliger TC & Waldo GS (2005): Protein tagging and detection with engineered self-assembling fragments of green fluorescent protein. *Nat Biotechnol* **23**, 102-107.

Cabantous S & Waldo GS (2006): In vivo and in vitro protein solubility assays using split GFP. *Nat Methods* **3**, 845-854.

Campbell CT, Sampathkumar SG & Yarema KJ (2007): Metabolic oligosaccharide engineering: perspectives, applications, and future directions. *Mol Biosyst* **3**, 187-194.

Cavalier-Smith T (1999): Principles of protein and lipid targeting in secondary symbiogenesis: euglenoid, dinoflagellate, and sporozoan plastid origins and the eukaryote family tree. *J Eukaryot Microbiol* **46**, 347-366.

Cavalier-Smith T (2003): Genomic reduction and evolution of novel genetic membranes and protein-targeting machinery in eukaryote-eukaryote chimaeras (meta-algae). *Philos Trans R Soc Lond B Biol Sci* **358**, 109-133; discussion 133-104.

Chu L, Ewe D, Río Bártulos C, Kroth PG & Gruber A (2016a): Rapid induction of GFP expression by the nitrate reductase promoter in the diatomPhaeodactylum tricornutum. *PeerJ* **4**, e2344.

Chu L, Gruber A, Ast M, Schmitz-Esser S, Altensell J, Neuhaus HE, Kroth PG & Haferkamp I (2016b): Shuttling of (deoxy-) purine nucleotides between compartments of the diatom Phaeodactylum tricornutum. *New Phytol.*

Chun W, Waldo GS & Johnson GV (2007): Split GFP complementation assay: a novel approach to quantitatively measure aggregation of tau in situ: effects of GSK3beta activation and caspase 3 cleavage. *J Neurochem* **103**, 2529-2539.

Cresswell RC & Syrett PJ (1979): Ammonium inhibition of nitrate uptake by the diatom, Phaeodactylum tricornutum. *Plant Science Letters* **14**, 321-325.

Csala M, Banhegyi G & Benedetti A (2006): Endoplasmic reticulum: a metabolic compartment. *FEBS Lett* **580**, 2160-2165.

Csala M, Marcolongo P, Lizak B, Senesi S, Margittai E, Fulceri R, Magyar JE, Benedetti A & Banhegyi G (2007): Transport and transporters in the endoplasmic reticulum. *Biochim Biophys Acta* **1768**, 1325-1341.

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

Dantuma NP, Lindsten K, Glas R, Jellne M & Masucci MG (2000): Short-lived green fluorescent proteins for quantifying ubiquitin/proteasome-dependent proteolysis in living cells. *Nat Biotech* **18**, 538-543.

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

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

Delwiche C & Palmer J (1997) The origin of plastids and their spread via secondary symbiosis. In *Origins of Algae and their Plastids* (Bhattacharya D ed.). Springer Vienna, pp. 53-86.

Delwiche CF (1999): Tracing the thread of plastid diversity through the tapestry of life. *The American Naturalist* **154**.

Deshaies RJ, Sanders SL, Feldheim DA & Schekman R (1991): Assembly of yeast Sec proteins involved in translocation into the endoplasmic reticulum into a membrane-bound multisubunit complex. *Nature* **349**, 806-808.

Du J, Meledeo MA, Wang Z, Khanna HS, Paruchuri VD & Yarema KJ (2009): Metabolic glycoengineering: sialic acid and beyond. *Glycobiology* **19**, 1382-1401.

Dube DB, C. R. (2003): Metabolic oligosaccharide engineering as a tool for glycobiology. *Current Opinion in Chemical Biology* **7**, 616-625.

Dukanovic J & Rapaport D (2011): Multiple pathways in the integration of proteins into the mitochondrial outer membrane. *Biochim Biophys Acta* **1808**, 971-980.

Dunahay TG, Jarvis EE & Roessler PG (1995): Genetic transformation of the diatoms Cyclotella cryptica and Navicula saprophila. *Journal of Phycology* **31**, 1004-1012.

Durnford DG (2003) Structure and Regulation of Algal Light-Harvesting Complex Genes. In *Photosynthesis in Algae* (Larkum AW, Douglas SE & Raven JA eds.). Kluwer Academic, Dordrecht, Netherlands, pp. 63-82.

Economou A & Wickner W (1994): SecA promotes preprotein translocation by undergoing ATP-driven cycles of membrane insertion and deinsertion. *Cell* **78**, 835-843.

Ellgaard L & Helenius A (2003): Quality control in the endoplasmic reticulum. *Nat Rev Mol Cell Biol* **4**, 181-191.

Emanuelsson O, Brunak S, von Heijne G & Nielsen H (2007): Locating proteins in the cell using TargetP, SignalP and related tools. *Nat Protoc* **2**, 953-971.

Facchinelli F & Weber AP (2011): The metabolite transporters of the plastid envelope: an update. *Front Plant Sci* **2**, 50.

Falciatore A, Casotti R, Leblanc C, Abrescia C & Bowler C (1999): Transformation of Nonselectable Reporter Genes in Marine Diatoms. *Mar Biotechnol (NY)* **1**, 239-251.

Falkowski PG & Oliver MJ (2007): Mix and match: how climate selects phytoplankton. *Nat Rev Micro* **5**, 813-819.

Felsner G, Sommer MS & Maier UG (2010): The physical and functional borders of transit peptide-like sequences in secondary endosymbionts. *BMC Plant Biol* **10**, 223.

Fischer H, Robl I, Sumper M & Kröger N (1999): Targeting and covalent modification of cell wall and membrane proteins heterologously expressed in the diatom Cylindrotheca fusiformis (Bacillariophyceae). *Journal of Phycology* **35**, 113-120.

Flori S, Jouneau PH, Finazzi G, Marechal E & Falconet D (2016): Ultrastructure of the Periplastidial Compartment of the Diatom Phaeodactylum tricornutum. *Protist* **167**, 254-267.

Freitas N & Cunha C (2009): Mechanisms and Signals for the Nuclear Import of Proteins. *Current Genomics* **10**, 550-557.

Friso G & van Wijk KJ (2015): Posttranslational Protein Modifications in Plant Metabolism. *Plant Physiol* **169**, 1469-1487.

Ge F, Huang W, Chen Z, Zhang C, Xiong Q, Bowler C, Yang J, Xu J & Hu H (2014): Methylcrotonyl-CoA Carboxylase Regulates Triacylglycerol Accumulation in the Model Diatom Phaeodactylum tricornutum. *Plant Cell* **26**, 1681-1697.

Geigenberger P, Stamme C, Tjaden J, Schulz A, Quick PW, Betsche T, Kersting HJ & Neuhaus HE (2001): Tuber physiology and properties of starch from tubers of transgenic potato plants with altered plastidic adenylate transporter activity. *Plant Physiol* **125**, 1667-1678.

Gibbs SP (1979): The route of entry of cytoplasmically synthesized proteins into chloroplasts of algae possessing chloroplast ER. *J Cell Sci* **35**, 253-266.

Gile GH, Moog D, Slamovits CH, Maier UG & Archibald JM (2015): Dual Organellar Targeting of Aminoacyl-tRNA Synthetases in Diatoms and Cryptophytes. *Genome Biol Evol* **7**, 1728-1742.

Gilson PR, Su V, Slamovits CH, Reith ME, Keeling PJ & McFadden GI (2006): Complete nucleotide sequence of the chlorarachniophyte nucleomorph: nature's smallest nucleus. *Proc Natl Acad Sci U S A* **103**, 9566-9571.

Gould SB, Sommer MS, Kroth PG, Gile GH, Keeling PJ & Maier UG (2006): Nucleus-to-nucleus gene transfer and protein retargeting into a remnant cytoplasm of cryptophytes and diatoms. *Mol Biol Evol* **23**, 2413-2422.

Gould SB, Waller RF & McFadden GI (2008): Plastid Evolution. *Annu Rev Plant Biol* **59**, 491-517.

Grefen C, Donald N, Hashimoto K, Kudla J, Schumacher K & Blatt MR (2010): A ubiquitin-10 promoter-based vector set for fluorescent protein tagging facilitates temporal stability and native protein distribution in transient and stable expression studies. *Plant J* **64**, 355-365.

Greub G & Raoult D (2003): History of the ADP/ATP-Translocase-Encoding Gene, a Parasitism Gene Transferred from a Chlamydiales Ancestor to Plants 1 Billion Years Ago. *Applied and Environmental Microbiology* **69**, 5530-5535.

Griffin RJ (1994): The medicinal chemistry of the azido group. *Prog Med Chem* **31**, 121-232.

Groß LE, Machettira AB, M. R, Schleiff E & Sommer MS (2011): GFP-based in vivo protein topology determination in plant protoplasts. *Journal of Endocytobiosis and Cell Research* **21**.

Gruber A (2008): Molecular Characterisation of Diatom Plastids. *Universität Konstanz* **Dissertation**.

Gruber A & Kroth PG (2014): Deducing intracellular distributions of metabolic pathways from genomic data. *Methods Mol Biol* **1083**, 187-211.

Gruber A, Rocap G, Kroth PG, Armbrust EV & Mock T (2015): Plastid proteome prediction for diatoms and other algae with secondary plastids of the red lineage. *Plant J*.

Gruber A, Vugrinec S, Hempel F, Gould SB, Maier UG & Kroth PG (2007): Protein targeting into complex diatom plastids: functional characterisation of a specific targeting motif. *Plant Mol Biol* **64**, 519-530.

Gruber A, Weber T, Bartulos CR, Vugrinec S & Kroth PG (2009): Intracellular distribution of the reductive and oxidative pentose phosphate pathways in two diatoms. *J Basic Microbiol* **49**, 58-72.

Guillard RRL (1975) Culture of Phytoplankton for Feeding Marine Invertebrates. In *Culture of Marine Invertebrate Animals: Proceedings — 1st Conference on Culture of Marine Invertebrate Animals Greenport* (Smith WL & Chanley MH eds.). Springer US, Boston, MA, pp. 29-60.

Guillard RRL & Ryther JH (1962): Studies of marine planktonic diatoms. I. Cyclotella nana Hustedt, and Detonula confervacea (cleve) Gran. *Can J Microbiol* **8**, 229-239.
Haferkamp I, Deschamps P, Ast M, Jeblick W, Maier U, Ball S & Neuhaus HE (2006a): Molecular and biochemical analysis of periplastidial starch metabolism in the cryptophyte Guillardia theta. *Eukaryot Cell* **5**, 964-971.
Haferkamp I & Linka N (2012): Functional expression and characterisation of membrane transport proteins. *Plant Biol (Stuttg)* **14**, 675-690.
Haferkamp I, Schmitz-Esser S, Linka N, Urbany C, Collingro A, Wagner M, Horn M & Neuhaus HE (2004): A candidate NAD+ transporter in an intracellular bacterial symbiont related to Chlamydiae. *Nature* **432**, 622-625.
Haferkamp I, Schmitz-Esser S, Wagner M, Neigel N, Horn M & Neuhaus HE (2006b): Tapping the nucleotide pool of the host: novel nucleotide carrier proteins of Protochlamydia amoebophila. *Mol Microbiol* **60**, 1534-1545.
Hamilton ML, Warwick J, Terry A, Allen MJ, Napier JA & Sayanova O (2015): Towards the Industrial Production of Omega-3 Long Chain Polyunsaturated Fatty Acids from a Genetically Modified Diatom Phaeodactylum tricornutum. *PLOS ONE* **10**, e0144054.
Heinz E, Hacker C, Dean P, Mifsud J, Goldberg AV, Williams TA, Nakjang S, Gregory A, Hirt RP, Lucocq JM, Kunji ER & Embley TM (2014): Plasma membrane-located purine nucleotide transport proteins are key components for host exploitation by microsporidian intracellular parasites. *PLoS Pathog* **10**, e1004547.
Heldt HW (1969): Adenine nucleotide translocation in spinach chloroplasts. *FEBS Lett* **5**, 11-14.
Hempel F, Bozarth AS, Lindenkamp N, Klingl A, Zauner S, Linne U, Steinbüchel A & Maier UG (2011a): Microalgae as bioreactors for bioplastic production. *Microbial Cell Factories* **10**, 1-6.
Hempel F, Bullmann L, Lau J, Zauner S & Maier UG (2009): ERAD-derived preprotein transport across the second outermost plastid membrane of diatoms. *Mol Biol Evol* **26**, 1781-1790.
Hempel F, Felsner G & Maier UG (2010): New mechanistic insights into pre-protein transport across the second outermost plastid membrane of diatoms. *Mol Microbiol* **76**, 793-801.
Hempel F, Lau J, Klingl A & Maier UG (2011b): Algae as protein factories: expression of a human antibody and the respective antigen in the diatom Phaeodactylum tricornutum. *PLOS ONE* **6**, e28424.
Hempel F & Maier UG (2012): An engineered diatom acting like a plasma cell secreting human IgG antibodies with high efficiency. *Microbial Cell Factories* **11**, 1-6.
Hirschberg CB, Robbins PW & Abeijon C (1998): Transporters of nucleotide sugars, ATP, and nucleotide sulfate in the endoplasmic reticulum and Golgi apparatus. *Annu Rev Biochem* **67**, 49-69.
Hoagland KD, Rosowski JR, Gretz MR & Roemer SC (1993): Diatom extracellular polymeric substances: function, fine structure, chemistry and physiology. *Journal of Phycology* **29**, 537-566.
Horstman A, Tonaco IA, Boutilier K & Immink RG (2014): A cautionary note on the use of split-YFP/BiFC in plant protein-protein interaction studies. *Int J Mol Sci* **15**, 9628-9643.
Houser JR, Ford E, Chatterjea SM, Maleri S, Elston TC & Errede B (2012): An improved short-lived fluorescent protein transcriptional reporter for Saccharomyces cerevisiae. *Yeast* **29**, 519-530.
Huber D, Boyd D, Xia Y, Olma MH, Gerstein M & Beckwith J (2005a): Use of thioredoxin as a reporter to identify a subset of Escherichia coli signal sequences that promote signal recognition particle-dependent translocation. *J Bacteriol* **187**, 2983-2991.
Huber D, Cha MI, Debarbieux L, Planson AG, Cruz N, Lopez G, Tasayco ML, Chaffotte A & Beckwith J (2005b): A selection for mutants that interfere with folding of Escherichia coli thioredoxin-1 in vivo. *Proc Natl Acad Sci U S A* **102**, 18872-18877.
Jewett JC & Bertozzi CR (2010): Cu-free click cycloaddition reactions in chemical biology. *Chem Soc Rev* **39**, 1272-1279.

Jungnickel B & Rapoport TA (1995): A posttargeting signal sequence recognition event in the endoplasmic reticulum membrane. *Cell* **82**, 261-270.

Kaddoum L, Magdeleine E, Waldo GS, Joly E & Cabantous S (2010): One-step split GFP staining for sensitive protein detection and localization in mammalian cells. *Biotechniques* **49**, 727-728, 730, 732 passim.

Kadono T, Miyagawa-Yamaguchi A, Kira N, Tomaru Y, Okami T, Yoshimatsu T, Hou L, Ohama T, Fukunaga K, Okauchi M, Yamaguchi H, Ohnishi K, Falciatore A & Adachi M (2015) Characterization of marine diatom-infecting virus promoters in the model diatom Phaeodactylum tricornutum. In *Sci Rep*. Macmillan Publishers Limited, p. 18708.

Kamiyama D, Sekine S, Barsi-Rhyne B, Hu J, Chen B, Gilbert LA, Ishikawa H, Leonetti MD, Marshall WF, Weissman JS & Huang B (2016): Versatile protein tagging in cells with split fluorescent protein. *Nat Commun* **7**, 11046.

Karas BJ, Diner RE, Lefebvre SC, McQuaid J, Phillips APR, Noddings CM, Brunson JK, Valas RE, Deerinck TJ, Jablanovic J, Gillard JTF, Beeri K, Ellisman MH, Glass JI, Hutchison Iii CA, Smith HO, Venter JC, Allen AE, Dupont CL & Weyman PD (2015): Designer diatom episomes delivered by bacterial conjugation. *Nat Commun* **6**.

Katoh K & Standley DM (2013): MAFFT multiple sequence alignment software version 7: improvements in performance and usability. *Mol Biol Evol* **30**, 772-780.

Kayser H, Zeitler R, Kannicht C, Grunow D, Nuck R & Reutter W (1992): Biosynthesis of a nonphysiological sialic acid in different rat organs, using N-propanoyl-D-hexosamines as precursors. *J Biol Chem* **267**, 16934-16938.

Kean EL, Münster-Kühnel AK & Gerardy-Schahn R (2004): CMP-sialic acid synthetase of the nucleus. *Biochimica et Biophysica Acta (BBA) - General Subjects* **1673**, 56-65.

Keeling PJ (2010): The endosymbiotic origin, diversification and fate of plastids. *Philos Trans R Soc Lond B Biol Sci* **365**, 729-748.

Keeling PJ (2013): The number, speed, and impact of plastid endosymbioses in eukaryotic evolution. *Annu Rev Plant Biol* **64**, 583-607.

Kilian O & Kroth PG (2005): Identification and characterization of a new conserved motif within the presequence of proteins targeted into complex diatom plastids. *Plant J* **41**, 175-183.

Kilian O & Kroth PG (2006): Molecular biology and genetic engineering in microalgae. *In: Algal cultures, analogues of blooms and applications (Ed.: D.V. Subba Rao) Science Publishers, Enfield (NH) USA*, 769- 799.

Kirchberger S, Tjaden J & Neuhaus HE (2008): Characterization of the Arabidopsis Brittle1 transport protein and impact of reduced activity on plant metabolism. *Plant J* **56**, 51-63.

Kitsera N, Khobta A & Epe B (2007): Destabilized green fluorescent protein detects rapid removal of transcription blocks after genotoxic exposure. *Biotechniques* **43**, 222-227.

Knab S, Mushak TM, Schmitz-Esser S, Horn M & Haferkamp I (2011): Nucleotide parasitism by Simkania negevensis (Chlamydiae). *J Bacteriol* **193**, 225-235.

Koning AJ, Lum PY, Williams JM & Wright R (1993): DiOC6 staining reveals organelle structure and dynamics in living yeast cells. *Cell Motil Cytoskeleton* **25**, 111-128.

Krause DC, Winkler HH & Wood DO (1985): Cloning and expression of the Rickettsia prowazekii ADP/ATP translocator in Escherichia coli. *Proc Natl Acad Sci U S A* **82**, 3015-3019.

Kroth P (2007a): Molecular biology and the biotechnological potential of diatoms. *Adv Exp Med Biol* **616**, 23-33.

Kroth PG (2002): Protein transport into secondary plastids and the evolution of primary and secondary plastids. *Int Rev Cytol* **221**, 191-255.

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

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

Lang M, Apt KE & Kroth PG (1998): Protein Transport into "Complex" Diatom Plastids Utilizes Two Different Targeting Signals. *The Journal of Biological Chemistry* **273**, 20973-30978.

Lau JB, Stork S, Moog D, Schulz J & Maier UG (2016): Protein-protein interactions indicate composition of a 480 kDa SELMA complex in the second outermost membrane of diatom complex plastids. *Mol Microbiol* **100**, 76-89.

Lau JB, Stork S, Moog D, Sommer MS & Maier UG (2015): N-terminal lysines are essential for protein translocation via a modified ERAD-system in complex plastids. *Mol Microbiol*.

Lepetit B, Sturm S, Rogato A, Gruber A, Sachse M, Falciatore A, Kroth PG & 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 Physiology* **161**, 853-865.

Leroch M, Kirchberger S, Haferkamp I, Wahl M, Neuhaus HE & Tjaden J (2005): Identification and characterization of a novel plastidic adenine nucleotide uniporter from Solanum tuberosum. *J Biol Chem* **280**, 17992-18000.

Leroch M, Neuhaus HE, Kirchberger S, Zimmermann S, Melzer M, Gerhold J & Tjaden J (2008): Identification of a novel adenine nucleotide transporter in the endoplasmic reticulum of Arabidopsis. *Plant Cell* **20**, 438-451.

Li X, Zhao X, Fang Y, Jiang X, Duong T, Fan C, Huang CC & Kain SR (1998): Generation of destabilized green fluorescent protein as a transcription reporter. *J Biol Chem* **273**, 34970-34975.

Li Y & Chen X (2012): Sialic acid metabolism and sialyltransferases: natural functions and applications. *Appl Microbiol Biotechnol* **94**, 887-905.

Linka N, Hurka H, Lang BF, Burger G, Winkler HH, Stamme C, Urbany C, Seil I, Kusch J & Neuhaus HE (2003): Phylogenetic relationships of non-mitochondrial nucleotide transport proteins in bacteria and eukaryotes. *Gene* **306**, 27-35.

Lommer M, Specht M, Roy AS, Kraemer L, Andreson R, Gutowska MA, Wolf J, Bergner SV, Schilhabel MB, Klostermeier UC, Beiko RG, Rosenstiel P, Hippler M & LaRoche J (2012): Genome and low-iron response of an oceanic diatom adapted to chronic iron limitation. *Genome Biol* **13**, R66.

Mahal LK, Yarema KJ & Bertozzi CR (1997): Engineering Chemical Reactivity on Cell Surfaces Through Oligosaccharide Biosynthesis. *Science* **276**, 1125-1128.

Maheswari U, Jabbari K, Petit J-L, Porcel BM, Allen AE, Cadoret J-P, De Martino A, Heijde M, Kaas R, La Roche J, Lopez PJ, Martin-Jezequel V, Meichenin A, Mock T, Schnitzler Parker M, Vardi A, Armbrust EV, Weissenbach J, Katinka M & Bowler C (2010): Digital expression profiling of novel diatom transcripts provides insight into their biological functions. *Genome Biol* **11**.

Maheswari U, Mock T, Armbrust EV & Bowler C (2009): Update of the Diatom EST Database: a new tool for digital transcriptomics. *Nucleic Acids Res* **37**, D1001-1005.

Malyshev DA, Dhami K, Lavergne T, Chen T, Dai N, Foster JM, Correa IR, Jr. & Romesberg FE (2014): A semi-synthetic organism with an expanded genetic alphabet. *Nature* **509**, 385-388.

Martin W (2010): Evolutionary origins of metabolic compartmentalization in eukaryotes. *Philos Trans R Soc Lond B Biol Sci* **365**, 847-855.

Martin W & Müller M (1998): The hydrogen hypothesis for the first eukaryote. *Nature* **392**, 37-41.

Martin W, Stoebe B, Goremykin V, Hapsmann S, Hasegawa M & Kowallik KV (1998): Gene transfer to the nucleus and the evolution of chloroplasts. *Nature* **393**, 162-165.

Martinez-Gil L, Sauri A, Marti-Renom MA & Mingarro I (2011): Membrane protein integration into the endoplasmic reticulum. *FEBS J* **278**, 3846-3858.

Martino AD, Meichenin A, Shi J, Pan K & Bowler C (2007): Genetic and phenotypic characterization of Phaeodactylum tricornutum (Bacillariophyceae) accessions1. *Journal of Phycology* **43**, 992-1009.

Martoglio B, Hofmann MW, Brunner J & Dobberstein B (1995): The protein-conducting channel in the membrane of the endoplasmic reticulum is open laterally toward the lipid bilayer. *Cell* **81**, 207-214.

Mathieu-Rivet E, Kiefer-Meyer MC, Vanier G, Ovide C, Burel C, Lerouge P & Bardor M (2014): Protein N-glycosylation in eukaryotic microalgae and its impact on the production of nuclear expressed biopharmaceuticals. *Front Plant Sci* **5**, 359.

Mathieu-Rivet E, Scholz M, Arias C, Dardelle F, Schulze S, Le Mauff F, Teo G, Hochmal AK, Blanco-Rivero A, Loutelier-Bourhis C, Kiefer-Meyer MC, Fufezan C, Burel C, Lerouge P, Martinez F, Bardor M & Hippler M (2013): Exploring the N-glycosylation pathway in Chlamydomonas reinhardtii unravels novel complex structures. *Mol Cell Proteomics* **12**, 3160-3183.

Matlack KE, Misselwitz B, Plath K & Rapoport TA (1999): BiP acts as a molecular ratchet during posttranslational transport of prepro-alpha factor across the ER membrane. *Cell* **97**, 553-564.

McFadden GI (1999): Plastids and protein targeting. *J Eukaryot Microbiol* **46**, 339-346.

Milech N, Longville BA, Cunningham PT, Scobie MN, Bogdawa HM, Winslow S, Anastasas M, Connor T, Ong F, Stone SR, Kerfoot M, Heinrich T, Kroeger KM, Tan YF, Hoffmann K, Thomas WR, Watt PM & Hopkins RM (2015): GFP-complementation assay to detect functional CPP and protein delivery into living cells. *Sci Rep* **5**, 18329.

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

Möhlmann T, Tjaden J, Schwöppe C, Winkler HH, Kampfenkel K & Neuhaus HE (1998): Occurrence of two plastidic ATP/ADP transporters in Arabidopsis thaliana L. *European Journal of Biochemistry* **252**, 353-359.

Moller H, Bohrsch V, Bentrop J, Bender J, Hinderlich S & Hackenberger CP (2012): Glycan-specific metabolic oligosaccharide engineering of C7-substituted sialic acids. *Angew Chem Int Ed Engl* **51**, 5986-5990.

Moog D, Rensing SA, Archibald JM, Maier UG & Ullrich KK (2015): Localization and Evolution of Putative Triose Phosphate Translocators in the Diatom Phaeodactylum tricornutum. *Genome Biol Evol* **7**, 2955-2969.

Nasu Y, Asaoka Y, Namae M, Nishina H, Yoshimura H & Ozawa T (2016): Genetically Encoded Fluorescent Probe for Imaging Apoptosis in Vivo with Spontaneous GFP Complementation. *Anal Chem* **88**, 838-844.

Nelson DMea (1995): Production and dissolution of biogenic silica in the ocean: Revised global estimates, comparison with regional data and relationship to biogenic sedimentation. *Global biogeochemical cycle* **9**, 359-372.

Neuhaus HE, Thom E, Mohlmann T, Steup M & Kampfenkel K (1997): Characterization of a novel eukaryotic ATP/ADP translocator located in the plastid envelope of Arabidopsis thaliana L. *Plant J* **11**, 73-82.

Ng DT, Brown JD & Walter P (1996): Signal sequences specify the targeting route to the endoplasmic reticulum membrane. *J Cell Biol* **134**, 269-278.

Nicholas K, Nicholas HJ & Deerfield DI (1997): GeneDoc: analysis and visualization of genetic variation. *Embnet.news 4*, 1–4.

Ning X, Guo J, Wolfert MA & Boons G-J (2008): Visualizing Metabolically Labeled Glycoconjugates of Living Cells by Copper-Free and Fast Huisgen Cycloadditions. *Angewandte Chemie International Edition* **47**, 2253-2255.

Nisbet RE, Kilian O & McFadden GI (2004): Diatom genomics: genetic acquisitions and mergers. *Curr Biol* **14**, R1048-1050.

Niu YF, Yang ZK, Zhang MH, Zhu CC, Yang WD, Liu JS & Li HY (2012): Transformation of diatom Phaeodactylum tricornutum by electroporation and establishment of inducible selection marker. *Biotechniques* **0**, 1-3.

Nixon AE & Benkovic SJ (2000): Improvement in the efficiency of formyl transfer of a GAR transformylase hybrid enzyme. *Protein Eng* **13**, 323-327.

Nymark M, Valle KC, Brembu T, Hancke K, Winge P, Andresen K, Johnsen G & Bones AM (2009): An integrated analysis of molecular acclimation to high light in the marine diatom Phaeodactylum tricornutum. *PLOS ONE* **4**, e7743.

Nymark M, Valle KC, Hancke K, Winge P, Andresen K, Johnsen G, Bones AM & 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.

Otera H, Taira Y, Horie C, Suzuki Y, Suzuki H, Setoguchi K, Kato H, Oka T & Mihara K (2007): A novel insertion pathway of mitochondrial outer membrane proteins with multiple transmembrane segments. *J Cell Biol* **179**, 1355-1363.

Panzner S, Dreier L, Hartmann E, Kostka S & Rapoport TA (1995): Posttranslational protein transport in yeast reconstituted with a purified complex of Sec proteins and Kar2p. *Cell* **81**, 561-570.

Pedelacq JD, Cabantous S, Tran T, Terwilliger TC & Waldo GS (2006): Engineering and characterization of a superfolder green fluorescent protein. *Nat Biotechnol* **24**, 79-88.

Peelle B, Lorens J, Li W, Bogenberger J, Payan DG & Anderson DC (2001): Intracellular protein scaffold-mediated display of random peptide libraries for phenotypic screens in mammalian cells. *Chem Biol* **8**, 521-534.

Peschke M, Moog D, Klingl A, Maier UG & Hempel F (2013): Evidence for glycoprotein transport into complex plastids. *Proc Natl Acad Sci U S A* **110**, 10860-10865.

Port JA, Parker MS, Kodner RB, Wallace JC, Armbrust EV & Faustman EM (2013): Identification of G protein-coupled receptor signaling pathway proteins in marine diatoms using comparative genomics. *BMC Genomics* **14**, 503.

Poulsen N, Chesley PM & Kröger N (2006): Molecular Genetic Manipulation of the Diatom Thalassiosira Pseudonana (Bacillariophyceae). *Journal of Phycology* **42**, 1059-1065.

Poulsen N & Kröger N (2005): A new molecular tool for transgenic diatoms: control of mRNA and protein biosynthesis by an inducible promoter-terminator cassette. *FEBS J* **272**, 3413-3423.

Poulsen N, Kröger N, Harrington MJ, Brunner E, Paasch S & Buhmann MT (2014): Isolation and biochemical characterization of underwater adhesives from diatoms. *Biofouling* **30**, 513-523.

Poulsen N, Scheffel A, Sheppard VC, Chesley PM & Kröger N (2013): Pentalysine clusters mediate silica targeting of silaffins in Thalassiosira pseudonana. *J Biol Chem* **288**, 20100-20109.

Prescher JA & Bertozzi CR (2005): Chemistry in living systems. *Nat Chem Biol* **1**, 13-21.

Prescher JA & Bertozzi CR (2006): Chemical technologies for probing glycans. *Cell* **126**, 851-854.

Randall LL, Topping TB, Hardy SJS, Pavlov MY, Freistroffer DV & Ehrenberg M (1997): Binding of SecB to ribosome-bound polypeptides has the same characteristics as binding to full-length, denatured proteins. *Proc Natl Acad Sci U S A* **94**, 802-807.

Rapoport TA (2007): Protein translocation across the eukaryotic endoplasmic reticulum and bacterial plasma membranes. *Nature* **450**, 663-669.

Rapoport TA, Goder V, Heinrich SU & Matlack KE (2004): Membrane-protein integration and the role of the translocation channel. *Trends Cell Biol* **14**, 568-575.

Rapoport TA, Jungnickel B & Kutay U (1996): Protein transport across the eukaryotic endoplasmic reticulum and bacterial inner membranes. *Annu Rev Biochem* **65**, 271-303.

Reinhold T, Alawady A, Grimm B, Beran KC, Jahns P, Conrath U, Bauer J, Reiser J, Melzer M, Jeblick W & Neuhaus HE (2007): Limitation of nocturnal import of ATP into Arabidopsis chloroplasts leads to photooxidative damage†. *The Plant Journal* **50**, 293-304.

Reiser J, Linka N, Lemke L, Jeblick W & Neuhaus HE (2004): Molecular physiological analysis of the two plastidic ATP/ADP transporters from Arabidopsis. *Plant Physiol* **136**, 3524-3536.

Ress DK & Linhardt RJ (2004): Sialic Acid Donors: Chemical Synthesis and Glycosylation. *Current Organic Synthesis* **1**, 31-46.

Río Bártulos C (2007): Eine mitochondriale Teil-Glykolyse bei Phaeodactylum tricornutum und anderen Heterokonten: Nachweis der Kompartimentierung mittels dem „green fluorescent protein“ und molekulare Phylogenie. *Technischen Universität Carolo-Wilhelmina zu Braunschweig* **Dissertation**.

Roesle P, Stempfle F, Hess SK, Zimmerer J, Río Bártulos C, Lepetit B, Eckert A, Kroth PG & Mecking S (2014): Synthetic Polyester from Algae Oil. *Angewandte Chemie International Edition* **53**, 6800-6804.

Rottberger J, Gruber A, Boenigk J & Kroth PG (2013a): Influence of nutrients and light on autotrophic, mixotrophic and heterotrophic freshwater chrysophytes. *Aquatic Microbial Ecology* **71**, 179-191.

Rottberger J, Gruber A & Kroth PG (2013b): Analysing size variation during light-starvation response of nutritionally diverse chrysophytes with a Coulter counter. *Algological Studies* **141**, 37-51.

Russo MT, Annunziata R, Sanges R, Ferrante MI & Falciatore A (2015): The upstream regulatory sequence of the light harvesting complex Lhcf2 gene of the marine diatom Phaeodactylum tricornutum enhances transcription in an orientation- and distance-independent fashion. *Mar Genomics* **24 Pt 1**, 69-79.

Sakaue K, Harada H & Matsuda Y (2008): Development of gene expression system in a marine diatom using viral promoters of a wide variety of origin. *Physiol Plant* **133**, 59-67.

Sambrook J, Fritsch EF & Maniatis T (1989) MOLECULAR CLONING A LABORATORY MANUAL SECOND EDITION VOLS. 1 2 AND 3, pp. XXXIX-+PAGINATION VARIES(VOL 1), XXXIII+PAGINATION VARIES(VOL 2), XXXII+PAGINATION VARIES(VOL 3).

Schmidt JDR (2016): Metabolic Oligosaccharide Engineering in Diatoms. *Universität Konstanz* **Bachelor Thesis**.

Schmitz-Esser S, Haferkamp I, Knab S, Penz T, Ast M, Kohl C, Wagner M & Horn M (2008): Lawsonia intracellularis contains a gene encoding a functional rickettsia-like ATP/ADP translocase for host exploitation. *J Bacteriol* **190**, 5746-5752.

Schmitz-Esser S, Linka N, Collingro A, Beier CL, Neuhaus HE, Wagner M & Horn M (2004): ATP/ADP Translocases: a Common Feature of Obligate Intracellular Amoebal Symbionts Related to Chlamydiae and Rickettsiae. *Journal of Bacteriology* **186**, 683-691.

Schober A (2014): Isolation and characterization of plastids from the diatom Thalassiosira pseudonana. *Universität Konstanz* **Master's Thesis**.

Schwarz F & Aebi M (2011): Mechanisms and principles of N-linked protein glycosylation. *Curr Opin Struct Biol* **21**, 576-582.

Seo S, Jeon H, Hwang S, Jin E & Chang KS (2015): Development of a new constitutive expression system for the transformation of the diatom Phaeodactylum tricornutum. *Algal Research* **11**, 50-54.

Shao S & Hegde RS (2011): Membrane protein insertion at the endoplasmic reticulum. *Annu Rev Cell Dev Biol* **27**, 25-56.

Sheiner L & Striepen B (2013): Protein sorting in complex plastids. *Biochim Biophys Acta* **1833**, 352-359.

Sheppard V, Poulsen N & Kröger N (2009): Characterization of an Endoplasmic Reticulum-associated Silaffin Kinase from the Diatom Thalassiosira pseudonana. *Journal of Biological Chemistry* **285**, 1166-1176.

Shrestha RP, Tesson B, Norden-Krichmar T, Federowicz S, Hildebrand M & Allen AE (2012): Whole transcriptome analysis of the silicon response of the diatom Thalassiosira pseudonana. *BMC Genomics* **13**, 499.

Siaut M, Heijde M, Mangogna M, Montsant A, Coesel S, Allen A, Manfredonia A, Falciatore A & Bowler C (2007): Molecular toolbox for studying diatom biology in Phaeodactylum tricornutum. *Gene* **406**, 23-35.

Sommer M, Rudolf M, Tillmann B, Tripp J, Sommer MS & Schleiff E (2013): Toc33 and Toc64-III cooperate in precursor protein import into the chloroplasts of Arabidopsis thaliana. *Plant Cell Environ* **36**, 970-983.

Sommer MS, Daum B, Gross LE, Weis BL, Mirus O, Abram L, Maier UG, Kuhlbrandt W & Schleiff E (2011): Chloroplast Omp85 proteins change orientation during evolution. *Proc Natl Acad Sci U S A* **108**, 13841-13846.

Sommer MS, Gould SB, Lehmann P, Gruber A, Przyborski JM & Maier UG (2007): Der1-mediated preprotein import into the periplastid compartment of chromalveolates? *Mol Biol Evol* **24**, 918-928.

Spiro RG (2002): Protein glycosylation: nature, distribution, enzymatic formation, and disease implications of glycopeptide bonds. *Glycobiology* **12**, 43r-56r.

Stanley P (2011): Golgi glycosylation. *Cold Spring Harb Perspect Biol* **3**.
Stemmer WP (1994): DNA shuffling by random fragmentation and reassembly: in vitro recombination for molecular evolution. *Proc Natl Acad Sci U S A* **91**, 10747-10751.
Stork S, Moog D, Przyborski JM, Wilhelmi I, Zauner S & Maier UG (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.
Sturm S, Engelken J, Gruber A, Vugrinec S, G Kroth P, Adamska I & Lavaud J (2013): A novel type of light-harvesting antenna protein of red algal origin in algae with secondary plastids. *BMC Evolutionary Biology* **13**, 159-159.
Tamura K, Stecher G, Peterson D, Filipski A & Kumar S (2013): MEGA6: Molecular Evolutionary Genetics Analysis version 6.0. *Mol Biol Evol* **30**, 2725-2729.
Tanaka A, De Martino A, Amato A, Montsant A, Mathieu B, Rostaing P, Tirichine L & Bowler C (2015): Ultrastructure and Membrane Traffic During Cell Division in the Marine Pennate Diatom Phaeodactylum tricornutum. *Protist* **166**, 506-521.
Terasaki M (1989): Fluorescent labeling of endoplasmic reticulum. *Methods Cell Biol* **29**, 125-135.
Thompson JD, Gibson TJ, Plewniak F, Jeanmougin F & Higgins DG (1997): The CLUSTAL_X windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools. *Nucleic Acids Res* **25**, 4876-4882.
Thoms S (2015): Import of proteins into peroxisomes: piggybacking to a new home away from home. *Open Biol* **5**.
Tjaden J, Möhlmann T, Kampfenkel K & Neuhaus Gudrun HaE (1998a): Altered plastidic ATP/ADP-transporter activity influences potato (Solanum tuberosum L.) tuber morphology, yield and composition of tuber starch. *The Plant Journal* **16**, 531-540.
Tjaden J, Schwöppe C, Möhlmann T, Quick PW & Neuhaus HE (1998b): Expression of a Plastidic ATP/ADP Transporter Gene inEscherichia coli Leads to a Functional Adenine Nucleotide Transport System in the Bacterial Cytoplasmic Membrane. *Journal of Biological Chemistry* **273**, 9630-9636.
Tjaden JW, H. H.; Schwoeppe, C.; van der Laan, M.; Moehlmann, T.; Neuhaus, H. E. (1999): Two nucleotide transport proteins in Chlamydia trachomatis, one for net nucleoside triphosphate uptake and the other for transport of energy. *Journal of Bacteriology* **181**, 1196-1202.
Trentmann O, Jung B, Neuhaus HE & Haferkamp I (2008): Nonmitochondrial ATP/ADP transporters accept phosphate as third substrate. *J Biol Chem* **283**, 36486-36493.
Tsaousis AD, Kunji ER, Goldberg AV, Lucocq JM, Hirt RP & Embley TM (2008): A novel route for ATP acquisition by the remnant mitochondria of Encephalitozoon cuniculi. *Nature* **453**, 553-556.
Tyra HM, Linka M, Weber AP & Bhattacharya D (2007): Host origin of plastid solute transporters in the first photosynthetic eukaryotes. *Genome Biol* **8**, R212.
Ullmann A, Jacob F & Monod J (1967): Characterization by in vitro complementation of a peptide corresponding to an operator-proximal segment of the beta-galactosidase structural gene of Escherichia coli. *J Mol Biol* **24**, 339-343.
Valenzuela J, Mazurie A, Carlson RP, Gerlach R, Cooksey KE, Peyton BM & Fields MW (2012): Potential role of multiple carbon fixation pathways during lipid accumulation in Phaeodactylum tricornutum. *Biotechnology for Biofuels* **5**, 1-17.
Valle KC, Nymark M, Aamot I, Hancke K, Winge P, Andresen K, Johnsen G, Brembu T & Bones AM (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.
van der Klei IJ & Veenhuis M (2006): PTS1-independent sorting of peroxisomal matrix proteins by Pex5p. *Biochim Biophys Acta* **1763**, 1794-1800.
van Dooren GG, Tomova C, Agrawal S, Humbel BM & Striepen B (2008): Toxoplasma gondii Tic20 is essential for apicoplast protein import. *Proc Natl Acad Sci U S A* **105**, 13574-13579.

van Dooren GGS, S.; Osafune, T.; McFadden, G. I. (2001): Translocation of proteins across the multiple membranes of complex plastids. *Biochimica et Biophysica Acta* **1541**, 34-53.

Van Engelenburg SB & Palmer AE (2010): Imaging type-III secretion reveals dynamics and spatial segregation of Salmonella effectors. *Nat Methods* **7**, 325-330.

Vanier G, Hempel F, Chan P, Rodamer M, Vaudry D, Maier UG, Lerouge P & Bardor M (2015): Biochemical Characterization of Human Anti-Hepatitis B Monoclonal Antibody Produced in the Microalgae Phaeodactylum tricornutum. *PLOS ONE* **10**, e0139282.

Vitale A, Ceriotti A & Denecke J (1993): The Role of the Endoplasmic Reticulum in Protein Synthesis, Modification and Intracellular Transport. *J Exp Bot* **44**, 1417-1444.

Vugrinec S (2011): Protein Targeting into Diatom Plastids. *Universität Konstanz* **Dissertation**.

Vugrinec S, Gruber A & Kroth PG (2011): Protein targeting into complex plastids - support for the translocator model. *Journal of Endocytobiosis and Cell Research*, 59-63.

Waldo GS (2003): Genetic screens and directed evolution for protein solubility. *Curr Opin Chem Biol* **7**, 33-38.

Waldo GS, Standish BM, Berendzen J & Terwilliger TC (1999): Rapid protein-folding assay using green fluorescent protein. *Nat Biotech* **17**, 691-695.

Walsh CT, Garneau-Tsodikova S & Gatto GJ, Jr. (2005): Protein posttranslational modifications: the chemistry of proteome diversifications. *Angew Chem Int Ed Engl* **44**, 7342-7372.

Wanders RJA & Waterham HR (2006): Biochemistry of Mammalian Peroxisomes Revisited. *Annual Review of Biochemistry* **75**, 295-332.

Weber AP, Linka M & Bhattacharya D (2006): Single, ancient origin of a plastid metabolite translocator family in Plantae from an endomembrane-derived ancestor. *Eukaryot Cell* **5**, 609-612.

Weber AP & Linka N (2011): Connecting the plastid: transporters of the plastid envelope and their role in linking plastidial with cytosolic metabolism. *Annu Rev Plant Biol* **62**, 53-77.

Wee EG, Sherrier DJ, Prime TA & Dupree P (1998): Targeting of active sialyltransferase to the plant Golgi apparatus. *The Plant Cell* **10**, 1759-1768.

Wehrman T, Kleaveland B, Her JH, Balint RF & Blau HM (2002): Protein-protein interactions monitored in mammalian cells via complementation of beta -lactamase enzyme fragments. *Proc Natl Acad Sci U S A* **99**, 3469-3474.

Wigley WC, Stidham RD, Smith NM, Hunt JF & Thomas PJ (2001): Protein solubility and folding monitored in vivo by structural complementation of a genetic marker protein. *Nat Biotech* **19**, 131-136.

Wilson CGM, Magliery TJ & Regan L (2004): Detecting protein-protein interactions with GFP-fragment reassembly. *Nat Meth* **1**, 255-262.

Windler M, Leinweber K, Bartulos CR, Philipp B, Kroth PG & Cock M (2015): Biofilm and capsule formation of the diatomAchnanthidium minutissimumare affected by a bacterium. *Journal of Phycology* **51**, 343-355.

Winkler HH & Neuhaus HE (1999): Non-mitochondrial ATP transport. *Trends in Biochemical Sciences* **24**, 64-68.

Wittmann V (2008): Glycoproteins: Occurrence and Significance. 1738--1770.

Witz S, Jung B, Furst S & Mohlmann T (2012): De novo pyrimidine nucleotide synthesis mainly occurs outside of plastids, but a previously undiscovered nucleobase importer provides substrates for the essential salvage pathway in Arabidopsis. *Plant Cell* **24**, 1549-1559.

Yang Z-K, Niu Y-F, Ma Y-H, Xue J, Zhang M-H, Yang W-D, Liu J-S, Lu S-H, Guan Y & Li H-Y (2013): Molecular and cellular mechanisms of neutral lipid accumulation in diatom following nitrogen deprivation. *Biotechnology for Biofuels* **6**, 1-14.

Yoshinaga R, Niwa-Kubota M, Matsui H & Matsuda Y (2014): Characterization of iron-responsive promoters in the marine diatom Phaeodactylum tricornutum. *Mar Genomics* **16**, 55-62.

Yu ET, Zendejas FJ, Lane PD, Gaucher S, Simmons BA & Lane TW (2009): Triacylglycerol accumulation and profiling in the model diatoms Thalassiosira pseudonana and Phaeodactylum tricornutum (Baccilariophyceae) during starvation. *Journal of Applied Phycology* **21**, 669-681.

Zaslavskaia LA, Lippmeier JC, Kroth PG, Grossman AR & Apt KE (2000): Transformation of the diatom Phaeodactylum tricornutum (Bacillariophyceae) with a variety of selectable marker and reporter genes. *J. Phycol.* **36**, 379-386.

Zaslavskaia LA, Lippmeier JC, Shih C, Ehrhardt D, Grossman AR & Apt KE (2001): Trophic conversion of an obligate photoautotrophic organism through metabolic engineering. *Science* **292**, 2073-2075.

Zhang C & Hu H (2013): High-efficiency nuclear transformation of the diatom Phaeodactylum tricornutum by electroporation. *Mar Genomics*.

Zimmermann R, Eyrisch S, Ahmad M & Helms V (2011): Protein translocation across the ER membrane. *Biochim Biophys Acta* **1808**, 912-924.

Zrenner R, Stitt M, Sonnewald U & Boldt R (2006): Pyrimidine and purine biosynthesis and degradation in plants. *Annu Rev Plant Biol* **57**, 805-836.