CHARACTERIZATION OF IRON-REGULATED PROTEINS AND THEIR TRANSCRIPTIONAL RESPONSE TO TRANSIENT IRON LEVELS IN THE OPEN OCEAN DIATOM *THALASSIOSIRA OCEANICA*

by

Joerg Behnke

Submitted in partial fulfilment of the requirements

for the degree of Doctor of Philosophy

Dalhousie University

Halifax, Nova Scotia

December 2019

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

ABSTRACT

Iron-requiring proteins are essential for numerous cellular processes, but low iron concentrations limit algal growth in over 30% of the ocean's surface waters. Iron-fertilization experiments induced diatom-dominated blooms, and transcriptional analysis of iron-deprived cultures revealed novel proteins, coined iron starvation-induced proteins (ISIP). The main objectives of this thesis were to study the transcriptional and physiological response of the open ocean diatom *Thalassasiosira oceanica* to iron addition and to characterize the ISIP proteins.

The first chapter introduces the background, explains the methods, and states the objectives of each chapter.

In the second chapter, the functional characterization of ISIP proteins is explored by reviewing and comparing the iron acquisition strategies in different algal species. Through *in-silico* analysis and cellular localization experiments, the study reveals differences in iron-uptake proteins, iron-acquisition strategies, and ISIP proteins. Furthermore, this study hypothesizes that ISIP3 is an iron storage protein.

The analysis of the high-resolution temporal response of iron-responsive genes to the addition of iron was the objective of the third chapter. Here, employing targeted transcriptomics, the dynamics of iron-regulated transcripts were followed after the addition of iron to iron-deprived cultures. Iron-uptake related transcripts were downregulated first, followed by an upregulation of transcripts for iron-containing proteins, and photosynthetic parameters recovered within a period of 6 h.

In the fourth chapter, N-linked glycosylated peptides were analyzed under high- and low-iron growth conditions. N-linked glycosylated peptides were captured using the solid-phase extraction of N-linked glycopeptides (SPEG) method and identified via LC-MS/MS, revealing 118 glycosylated peptides with mostly NXT-type motifs. Peptides recovered from low-iron cells confirmed the predicted topology of ISIP1a.

Chapter 5 reports on the development of a transformation system in *T. oceanica* to identify the cellular localization of YFP-tagged proteins. The transformation vector was designed using the Gibson® cloning technology, and cells were transformed via microparticle bombardment. *T. oceanica* was genetically transformable, confirmed by the expression of the resistance gene and growth in antibiotic-selective media.

Overall, the findings presented here greatly expand the understanding of the response of diatoms to changes in iron concentrations. The transcription kinetics will be useful for modeling efforts at the cellular and biogeochemical level.

LIST OF ABBREVIATIONS AND SYMBOLS USED

α	Alpha
ACN	Acetonitrile
actD	Actinomycin D
ALG	Asparagine-linked enzyme
ASW	Artificial seawater
β	Beta
blast	Basic local alignment search tool
BCA	Bicinchoninic acid
bp	Basepair
C	Carbon
ccdB	Coupled cell division B gene
CD	Conserved domain
CID	Collision-induced dissociation
Cl	Chlorine
CREGx2	Cellular-repressor-of-EA1-regulated-genes
Cu	Copper
DAPI	4′,6-diamidino-2-phenylindole

DMSO	Dimethyl sulfoxide
DNA	Desoxyribonucleic acid
Ds	*Dunaliella salina*
DTT	Dithiothreitol
DUF	Domain of unknown function
EDTA	Ethylenediaminetetraacetic acid
EfeUOP	Elemental ferrous iron uptake system
Ein2	Ethylene-insensitive protein 2
ER	Endoplasmic reticulum
ETR	Electron transport rate
FBA	Fructose-bisphosphate aldolase
FBP	Ferrichrome-binding protein
FDR	False discovery rate
Fe	Iron
FEA	Iron assimilating protein
FL3	Fluorescence light 3
FLDA	Flavodoxin
F_m	Maximal fluorescence

FRE	Iron reductase
FRO	Ferric reduction oxidase
FSC	Forward scatter
FTR	Iron (Fe) transporter
FucT	Fucosyltransferase
F_v	Variable fluorescence
GalT	Galactosylase
GC	Glucosidase
GFF	Glass fiber filter
GFP	Green fluorescence protein
Glc	Glucosyltransferase
GnT1	N-acetylglucosaminyltransferase 1
HCD	Higher-energy collisional dissociation
He	Helium
HNLC	High nutrient low chlorophyll
HPLC	High-pressure liquid chromatography
HSF	Heat shock factor
ID	Identity

IHF	Integration host factor
Int	Intergrase
IRT	Iron regulated transporter
ISIP	Iron starvation-induced protein
iTOL	Interactive tree of life
JGI	Joint Genome Institute
KEGG	Kyoto Encyclopedia of Genes and Genomes
LC	Liquid chromatography
LLO	Lipid linked oligosaccharide
Ln	Natural logarithm
µ	Micro
Man	Mannosidase
ManT	Mannosyltransferase
MCO	Multicopper oxidase
Met	Methionine
Mio	Million
Mito	Mitochondrion
ML	Measuring light

MMETSP	Marine Microbial Eukaryote Transcriptome Sequencing Project
Mn	Manganese
Mnth	Manganese transporter
mRNA	Messenger ribonucleic acid
MS	Mass spectroscopy
MUSCLE	Multiple sequence Comparison by log-expectation
N	Nitrogen
NADP	Nicotinamide adenine dinucleotide phosphate
NCBI	National Center for Biotechnology Information
Ni	Nickel
NO	Non-regulated energy dissipation
NPQ	Non-photochemical quenching
nr	non-redundant
NRAMP	Natural resistance-associated macrophage protein
nt	Nucleotide
Nucl	Nucleus
OST	Oligosaccharyltransferase
PC	Polycarbonate

PCR	Polymerase chain reaction
P-Dol	Dolichol-phosphate
PDS	Particle delivery system
PE	Photosystem efficiency
PETE	Plastocyanin
PETF	Ferredoxin
PC	Polycarbonate
POC	Particulate organic carbon
PON	Particulate organic nitrogen
Psi	pound per square inch
PSI	Photosystem I
PSII	Photosystem II
Pt	*Phaeodactylum tricornutum*
RNA	Ribonucleic acid
ROS	Reactive oxygen species
RT	Room temperature
σ^2	population variance
SAR	Stramenopiles, alveolates and Rhizaria

SOD	Superoxide dismutase
SP	Saturation pulse
SP	Signal peptide
SSC	Side scatter
Stt3	Staurosporine and temperature-sensitive 3
TCA	Tricarboxylic acid cycle
TFA	Trifluoroacetic acid
TM	Transmembrane
To	*Thalassiosira oceanica (T. oceanica*)
UCGT	UDP-glucose:glycoprotein glucosyltransferase
VIT1	Vacuolar iron transporter 1
Xis	Excisionase
YFP	Yellow fluorescent protein
YSL	Yellow stripe like proteins
ZIP	Zinc regulated iron-regulated protein
ZRT	Zinc regulated transporter
ZupT	Zinc uptake transporter

Units and prefixes were used according to the International System of Units.

ACKNOWLEDGMENTS

First, I want to thank Douglas Wallace and Julie LaRoche for the opportunity to conduct my PhD in Canada, making a childhood dream come true. Further, I want to thank Julie LaRoche for the chance to do cell biology work on diatoms in the context of iron limitation, which was my favorite topic to work on, and her great ideas and suggestions when things seem stuck. Over the years, I really enjoyed her positive attitude as well as the good atmosphere that she created in the lab and the office.

I want to thank my committee members and external examiners (Arunika Gunawardena, Tom MacRae, Michael McReith, Claudio Slamovits, John Archibald, Dion Durnford) for their time and support in making me a better scientist.

For financial support, I want to thank the German Academic Exchange Service (DAAD) and Dr. Patrick Lett.

This work would have been much harder if there wouldn't have been so many people surrounding me that contributed to a great work atmosphere. Here, I want to thank all LaRoche workgroup members of the last 8 years, with special thanks to Jenni Tolman, Dhwani Desai, and Ian Luddington. I also want to thank the Bertrand workgroup with special thanks to Erin Bertrand for many (many) very valuable suggestions, tips, general support, and the "bubble". I also want thank Scott McCain, Loay Jabre, and Eldon Rowland for good discussions, constantly providing new ideas and perspectives. Overall, I want to thank the Biology administrative staff for fast and efficient responses, whenever there was a problem.

The collaboration with Alejandro Cohen was a great experience and I want to thank Alejandro Cohen for his work on my data, the great atmosphere in his lab, and the great discussions.

My teaching assistance experience wouldn't have been the same without Mindy McCarville. I really enjoyed my time in her class, and I want to say thank you for that!

Work and life has to be in balance and sometimes the private life comes short. Therefore, a BIG thank you goes to my son Noah Behnke-Lussing, and my wife Andrea Lussing. Andrea Lussing always supported me, which needs a very special thank you, and Noah Behnke-Lussing is just the best. I also want to thank Frank Lussing and Barbara Lussing for their support and in making Nova Scotia easy to feel home in. In that regard, I want to thank all the great people in Nova Scotia that I met and became friends with. I also want to thank my family and friends at home in Kiel, Germany. Every time I went home, it felt as if I had never left and this is only possible with good friends and a great family. Special thanks go to my parents, Juergen and Martina Behnke, my sister, Christin Behnke, my brother, Dirk Behnke, my Opa, Bruno Behnke, and my Oma, Waltraut Behnke.

CHAPTER 1 INTRODUCTION

1.0 Diatoms

Diatoms are ubiquitous photosynthetic algae, and with over 100,000 species, diatoms are present in a wide variety of aquatic habitats ranging from benthic freshwater (Laird *et al.*, 2010) to open ocean surface waters (Carradec *et al.*, 2018). Diatoms dominate wide parts of the ocean and play a vital role in the biological carbon pump influenced by their unique silica shell (Tréguer *et al.*, 2018). The silica frustule protects diatoms from grazing pressure and increases their weight resulting in accelerated sinking velocity. Diatoms fix an estimated 20% of the global carbon, revealing their significant impact on the global climate (Falkowski *et al.*, 2000), and contribute substantially to carbon sequestration and export into the deep ocean (Smetacek *et al.*, 2012). Evolutionarily, the first diatoms appeared around 200 Mio years ago and divided into two groups around 70 Mio years ago, pennate and centric diatoms. Diatoms evolved through a secondary endosymbiosis event indicated by a complex plastid harboring four membranes. Sequence analysis points towards a red algal ancestor (Benoiston *et al.*, 2017). Genomic sequencing of diatoms revealed significant evolutionary distance between diatom species (Allen *et al.*, 2008; Armbrust *et al.*, 2010; Lommer *et al.*, 2012; Mock *et al.*, 2017), and led to unexpected findings like the identification of the ornithine-urea cycle which is usually found in metazoan but absent in green algae and plants (Allen *et al.*, 2011).

1.1 Iron in the ocean

Essential cellular processes, including oxidative respiration, nitrogen fixation, and photosynthesis, evolved using iron as a cofactor in numerous proteins. The incorporation

of iron is based on the broad redox potential (+500 mV–700 mV) between its two oxidation states, ferrous (Fe^{2+}) and ferric (Fe^{3}) (Koay *et al.*, 2008), making iron an excellent electron carrier. The primary sources for iron in the ocean are riverine input, coastal erosion, dust, and hydrothermal vents (Tagliabue *et al.*, 2017). The oxidative state of contemporary oceans leads to the immediate oxidation of Fe^{2+} to Fe^{3+}, which itself precipitates in the form of $Fe(OH)_x$. Oceanic chemistry changed vastly 1.5 – 2 billion years ago when photosynthetic cyanobacteria produced oxygen, creating a switch from a reductive to an oxidative environment, known as the great oxygenation event (Lyons *et al.*, 2014). It is estimated that 99% of the iron in the modern ocean is bound to organic substances (Gledhill & Buck, 2012) and a fraction of these organic compounds is actively produced by bacteria in the form of siderophores resulting in an improvement of the overall bioavailability of iron (Leventhal *et al.*, 2019). The rest is organic debris, such as humic substances and transparent exopolymers (Gledhill & Buck, 2012; Hassler *et al.*, 2012).

Based on the high demand for iron, low bioavailability, and the limited transport of iron into the open ocean, vast areas of the ocean's surface waters have very low iron concentration impacting phytoplankton growth. Most of the contemporary ocean is limited by nitrate or by iron, but co-limitations with copper, cobalt or cobalamin, zinc, manganese, or silicate have been reported (Moore *et al.*, 2013; Bertrand *et al.*, 2015; Pausch *et al.*, 2019).

Iron limitation of marine primary productivity was first demonstrated in the 1990s (John H. Martin & Fitzwater, 1988; Martin, 1990; Martin *et al.*, 1994), and has received scientific attention ever since (de Baar *et al.*, 2005; Boyd *et al.*, 2007; Tripathy & Jena,

2019). The first iron fertilization experiment was conducted in the equatorial Pacific (John H. Martin & Fitzwater, 1988), and subsequent fertilization experiments demonstrated a positive effect of iron on phytoplankton growth resulting in diatom dominated algal blooms in the Southern Ocean, equatorial Pacific, and the north Pacific (Boyd *et al.*, 2007). These regions are referred to as High Nutrients-Low Chlorophyll Regions (HNLC). Besides HNLC regions, open ocean gyres have low iron concentration but are mostly limited by macronutrients (Moore *et al.*, 2013).

1.2 Low-iron adaptations

The low concentrations of iron in HNLC regions led to various adaptations including photochemical adaptations, specialized uptake proteins, replacement of iron-containing proteins, and a metabolic shift towards a mixotrophic lifestyle (Allen *et al.*, 2008; Lommer *et al.*, 2012; Marchetti *et al.*, 2012; Luxem *et al.*, 2017; Cohen, Gong, *et al.*, 2018). The report of differences in cellular iron demand of coastal and oceanic diatom species to maintain equivalent growth rates (Sunda *et al.*, 1991) led to investigations into the coping mechanisms of diatoms in low-iron concentrations. Soon after, it was shown that oceanic species living under chronic iron limitation have slower growth rates, smaller cell sizes, and lower metabolic iron demands (Sunda & Huntsman, 1995). The photosynthetic adaptations of *T. oceanica* showed an overall decrease in the iron-bearing photosystem I, resulting in a high PSII:PSI ratio (Strzepek & Harrison, 2004). The discovery of plastocyanin (PETE) (Peers & Price, 2006) and flavodoxin (FLDA1) (La Roche *et al.*, 1995) as replacements for the iron-containing proteins cytochrome c6 and ferredoxin, respectively, was a crucial step in the understanding of the acclimation to deficient iron concentrations. FLDA1, with a flavin as the electron carrier, replaces

ferredoxin in transferring electrons from photosystem I to the ferredoxin-$NADP^+$ reductase, and plastocyanin replaces cytochrome c_6. Flavodoxin is considered a marker for iron-limitation in the environment (Chappell *et al.*, 2015), and the analysis of the Tara Oceans dataset showed a strong negative correlation between iron concentration and FLDA expression (Carradec *et al.*, 2018). Other examples of protein replacements are fructose-bisphosphate aldolases (FBA). These proteins exist in two different forms that are either depending on a metal cofactor or a Schiff base. They are expressed according to their cofactor and the iron status of the cell (Lommer *et al.*, 2012). This strategy is described as the enzyme replacement strategy and is important to lower the iron demand in the cell.

The general reduction of cell size, growth rate, and protein content was reported for diatoms as a response to low-iron concentrations (Allen *et al.*, 2008; Lommer *et al.*, 2012) and was also shown in other algae (Botebol *et al.*, 2017; Luxem *et al.*, 2017). Besides decreasing the iron requirements, diatoms possess high-affinity iron uptake strategies to maintain sufficient iron uptake. Iron uptake is mediated by general uptake proteins such as ZIP (Guerinot, 2000) and NRAMP proteins (Nevo & Nelson, 2006) that take up divalent metals. Iron can also be taken up by a reductive high-affinity uptake system, as described in yeast (Terzulli & Kosman, 2010). This reductive high-affinity uptake system consists of an iron reductase, a multicopper oxidase (MCO), and an iron permease. *T. oceanica* upregulates an MCO and an iron reductase but lacks a permease (Lommer *et al.*, 2012). Besides the enzyme replacement and the iron uptake strategies, diatoms and other algae possess iron starvation-induced proteins (ISIP) that were first described in *Phaeodactylum tricornutum* (Allen *et al.*, 2008). Several ISIPs have been

identified, but only ISIP1 and ISIP2a in *P. tricornutum* have been functionally defined. ISIP1_*Pt* is involved in siderophore-mediated iron uptake (Kazamia *et al.*, 2018), and ISIP2a_*Pt* functions as a phytotransferrin (McQuaid *et al.*, 2018), taking up Fe^{3+}.

T. oceanica responded to iron limitation with a general metabolic shift seen in the downregulation of chloroplast transcripts and upregulation of mitochondrial transcripts. At the same time, carbon-compound degrading enzymes are upregulated, and an upregulation of proteins involved in endocytosis takes place (Lommer *et al.*, 2012; Smith *et al.*, 2016). Most of these adaptations are temporary and reversible. The diatom dominated iron-induced algal blooms in iron fertilization experiments brought great interest into the temporal response of diatoms. Analysis of transcript counts responding to iron addition demonstrated a downregulation for FLDA, ISIP3, and ferric reductases (FRE) after 3 h, but a more detailed temporal resolution was needed to show the exact expression profile (Kustka & Allen, 2007; Whitney *et al.*, 2011; Chappell *et al.*, 2015). In field-experiments, the response to the addition of iron elucidated differences between diatom genera of *Pseudo-nitzschia* and *Thalassiosira* but the upregulation of photosynthetic genes, as well as the downregulation of ISIP1 and ISIP3, was a common response (Marchetti *et al.*, 2012; Cohen *et al.*, 2017).

1.3 N-linked glycosylation in microalgae

Glycosylations are posttranslational modifications of proteins with a wide range of functions, and two main mechanisms exist. These two mechanisms are N-linked glycosylation when glycan structures are attached to asparagine residues and O-linked glycosylation, where glycan structures are attached to serine or threonine. The attachment of glycan structures occurs in the endoplasmic reticulum (ER), and one of the most

fundamental functions of glycosylation is protein folding control in the ER. The research of glycosylation in algae is still at the beginning, and no function has been described yet. The rising interest of microalgae for pharmaceutical recombinant protein production increased research efforts to understand N-linked glycosylation in microalgae. Most studies focused on glycan structures in *Phaedactylum tricornutum, Chlamydomonas reinhardtii, Botryococcus braunii, Porphyridium sp*., and *Chlorella vulgaris* (Baïet *et al.*, 2011; Levy-Ontman *et al.*, 2011; Mamedov & Yusibov, 2011; Schulze *et al.*, 2017; Mócsai *et al.*, 2019). Glycan structures are of special interest as they are often essential for the correct function of recombinant proteins (Vanier *et al.*, 2015). The process of glycosylation consists of multiple steps, and the attachment of the glycan, as well as the folding control, takes place in the ER. The overall process includes the maturation of a lipid-linked oligosaccharide (LLO) by well-conserved Asparagine-linked enzymes (ALG) along the membrane of the ER. The attachment of the oligosaccharide via an oligosaccharyltransferase (OST), a protein folding control, involving calreticulin or calnexin, and the modification of glycan structures in the Golgi apparatus on the basis of various enzymes follows. *P. tricornutum* has most of the ALG enzymes, the OST, enzymes involved in protein folding control, and enzymes involved in the maturation of the glycan structure in the Golgi apparatus (Baïet *et al.*, 2011). The folding control is a central function of N-linked glycosylation and may be vital for the survival of a cell.

The upregulation of chaperones and the remodeling of the cell surface proteome in *T. oceanica* under iron limitation (Lommer *et al.*, 2012) led to the hypothesis that there are transcriptional differences in the activity of enzymes involved in the N-linked

glycosylation pathway, as well as quantitative differences in glycosylated protein abundance between high- and low-iron cells.

1.4 Genetic transformation of diatoms

The impact of microalgae on marine biogeochemical cycles (Assmy *et al.*, 2013), fisheries, and the global climate (Tréguer *et al.*, 2018) are reasons for increasing research efforts to understand cellular concepts in microalgae. The results are used to improve oceanic modeling efforts and marine management, as well as to discover novel usable compounds. Besides genomic (Mock *et al.*, 2017), transcriptomic (Keeling *et al.*, 2014) and proteomic analysis (Smith *et al.*, 2016), the use of genetic transformation for silencing genes (De Riso *et al.*, 2009) or gene expression studies (Kazamia *et al.*, 2018) is established for eleven diatom species (Huang & Daboussi, 2017; Johansson *et al.*, 2019a). The majority of cell biology research in diatoms is carried out on *Phaeodactylum tricornutum,* despite its environmental insignificance and sparse distribution (Allen *et al.*, 2012; Schellenberger Costa *et al.*, 2013; Morrissey *et al.*, 2015; Kazamia *et al.*, 2018). The incorporation of foreign DNA into the nucleus or the chloroplast is mainly done via biolistic bombardment of vector DNA, but techniques such as bacterial transformation or multi-pulse electroporation are established in diatoms, allowing the incorporation of larger DNA fragments and higher transformation efficiencies (Doron *et al.*, 2016; Huang & Daboussi, 2017). The analyses of specific proteins via heterologous transformation can be useful, but the evolutionary distance between diatom species (Allen *et al.*, 2008; Bowler *et al.*, 2008; Lommer *et al.*, 2012; Mock *et al.*, 2017) and strain-specific responses to stressors (Hippmann *et al.*, 2017) shows that homologous transformation should be preferred.

1.5 Methods utilized in this study

1.5.1 Targeted transcriptomics using NanoString technologies

The NanoString technology (NanoString Technologies Inc., Seattle, Washington, USA) was used for the targeted transcriptome assay. This method provides absolute counts of mRNA transcripts and allows a fast and precise way to track changes in transcript abundance for up to 800 genes at a time (Tsang *et al.*, 2017). The design of target-specific probes is the first step. The total length of the target sequence is 100 bp with two separate probes binding to 50 bp of the target sequence. The first probe, probe A, hybridizes to 50 bp of the target sequence and to a reporter tag that contains six different fluorescent dyes allowing 800 combinations of color codes. The second probe, probe B, hybridizes to the target sequence and to a capture tag, which binds to streptavidin beads, allowing to wash off non-target transcripts and probes. The probes are not allowed to bind non-target sequences with more than 75 % identity to avoid false-positive binding (Kane, 2000).

1.5.2 Microparticle bombardment

Microparticle bombardment is a widely used technique to modify microalgae genetically (Huang & Daboussi, 2017). Here, a gene gun accelerates tungsten or gold beads coated with vector DNA onto plated cells. Several parameters can be adjusted to improve transformation efficiency in specific microalgal species. These parameters include the type and the size of the beads, the pressure used for the acceleration of the beads, and the distance between the beads and the plate with the cells. After the bombardment, the cells recover for 24 h on the plate or in suspension culture without antibiotics. After the

recovery, cells are plated on antibiotic-containing plates or kept as a suspension culture in antibiotic-containing media. After 2-4 weeks, single colonies can be picked off the plates (Kroth, 2007).

1.5.3 Pour plating

The pour plating technique represents an alternative plating technique for microalgae that do not grow on artificial seawater (ASW) plates. In the pour plating technique, the cells are added to agarose containing ASW before the medium is poured into the plates for solidification. The different compounds are autoclaved, mixed, and left to cool down to 32°C before adding the cells. The temperature plays a critical role as most cells do not survive hot temperatures, and agarose solidifies at 30°C. The plates contain 0.8 % extra clean agarose. The cleaning procedure of the agarose consists of two washes with MilliQ grade water, two washes with ethanol, and a final wash with acetone. Afterward, the agarose is dried and tightly sealed for storage (Turnsek & Dupont, 2017).

1.5.4 Solid-phase extraction of N-linked glycosylated peptides

The identification of N-linked glycosylated peptides is based on the oxidation of cis-diol groups of the sugar groups into aldehyde groups, which covalently bind to hydrazide coated beads and form hydrazone bonds. The protein lysate is digested with trypsin before the N-glycosylated peptides are captured onto the hydrazide-coated beads, and non-glycosylated peptides are washed off. The release of the N-linked glycosylated peptides from the beads is based on a PNGase F digest. The peptides are then dried and further processed and identified via LC-MS/MS (Tian *et al.*, 2007).

1.5.5 Gibson Assembly® technique

The Gibson Assembly® cloning technique (New England Biolabs, Ipswich, Massachusetts, USA) was used for the design of the *T. oceanica* specific vector. This technique allows the joining of up to five DNA fragments in one reaction. Prior to the Gibson Assembly, DNA fragments are amplified via polymerase chain reaction (PCR) to create a 40-50 bp overlap. During the Gibson Assembly reaction itself, an exonuclease, a DNA-polymerase, and a DNA-ligase are present. The DNA-exonuclease produces 5' single-stranded ends. The overlapping 5' single-stranded regions anneal, the DNA polymerase fills the gap, and the DNA-ligase finally seals the nick. Gibson Assembly vectors contain a coupled cell division B gene (ccdB), as well as antibiotic resistance genes for ampicillin and chloramphenicol. Therefore, ccdB resistant *Escherichia coli* cells are used on antibiotic-containing plates to achieve single colonies, containing the vector.

1.5.6 Gateway and TOPO® cloning

The TOPO® cloning technique allows the fast incorporation of a gene-of-interest into a vector. The joining of the gene-of-interest into the vector is based on a short four bp sequence that is added to the 3' end, resulting in a directional insertion into the TOPO® vector. The insertion itself is based on the activity of a topoisomerase I that binds to a specific sequence in the vector. Two topoisomerase I enzymes bind covalently to the linearized TOPO®. Vector, producing a 5' overhang that is complementary to the overhang sequence of the gene-of-interest. The gene-of-interest binds to the 5' overhang, and the topoisomerase I enzymes ligate the nick in the DNA and break off the vector, resulting in a circular vector that can be used for transformation. TOPO® vectors can be

used for the Gateway cloning system. This system allows a fast shuttling of the gene-of-interest between vectors. The final vector used for the desired transformation is called a destination vector. The exchange of the gene-of-interest involves an integrase (Int), the integration host factor (IHF), and an excisionase (Xis). After the gene-of-interest is successfully introduced into the destination vector, the vector is transformed into *E. coli* for amplification and purification.

1.5.7 Photochemical measurements using the DUAL-Pam-100

The measurement of photosystem II-related parameters was done with a DUAL-PAM-100 (Heinz Walz GmbH, Effeltrich, Germany). The samples have to be dark-adapted for 10 min before measuring the normalized variable fluorescence (F_v/F_m). Once the cells are dark-adapted, all electron acceptor sites within PSII complexes are open, and the minimal fluorescence (F_0) is measured. A saturation pulse of a specific wavelength closes all electron acceptor sides in the PSII units, and the maximal fluorescence is measured (F_{max}). Based on these two parameters, the normalized variable fluorescence is calculated as $F_v/F_m = (F_m-F_0)/F_m$. The DUA-PAM-100 can also measure the electron transport rate (ETR), non-photochemical quenching (NPQ), and non-regulated heat dissipation (NO). These parameters are recorded during an induction assay using 11 saturation pulses with a 40 sec distance and increasing intensity (0; 15; 22; 31; 62; 104; 135; 225;3 48; 540; 834 µE).

1.5.8 Flow cytometry measurements

Flow cytometry is a powerful tool for counting cells and recording characteristics of the cell, including cell size, granularity, and fluorescence in relative units. The flow

cytometer used in this study was the Accuri C6 (Becton, Dickinson and Company (BD). The machine creates a thin stream of cells and measures each cell separately. Two lasers (blue (488 nm), red (670 nm)) are used for excitation, and four detectors record light intensities of different wavelengths with the FL3 (>670 nm LP) used for relative chlorophyll fluorescence. Two detectors measure characteristics of light scatter in the form of side scatter, used as an estimate for cell granularity, and forward scatter, used for the relative size of the cells.

1.6 Objectives

This thesis had two objectives, the first being the description of the iron-response of iron-limited *T. oceanica* cells to iron supply in a temporal high-resolution fashion and the second being the biochemical and functional characterization of iron starvation-induced proteins.

IRON UPTAKE PROTEINS IN ALGAE AND THE ROLE OF IRON-STARVATION INDUCED PROTEINS (ISIPS)

Three groups of iron starvation-induced proteins were discovered in 2008 (Allen *et al.*, 2008) and are suggested to play an essential role in the survival of iron-limited diatoms (Allen *et al.*, 2008; Lommer *et al.*, 2012; Marchetti *et al.*, 2012; Cohen *et al.*, 2017). In *P. tricornutum*, ISIP1 is involved in iron uptake through endocytosis of an iron-siderophore complex (Kazamia *et al.*, 2018) and ISIP2 transports Fe^{3+} with transferrin-like mechanisms (McQuaid *et al.*, 2018). The exact function of ISIP1 in the iron uptake process remains unknown, and there has been no function suggested for ISIP3.

The main objective of this work is to compare iron starvation-induced proteins with different well-characterized iron uptake proteins. Furthermore, the comparison of ISIP proteins between different algal species was used to confirm previous findings and reveal new hints towards the function of the uncharacterized ISIP3 proteins. ISIP proteins of 15 different algal species were used with the addition of *Arabidopsis thaliana* and *Saccharomyces cerevisiae*, as these two species have well-characterized iron uptake systems. The second objective was to the identification of differences and similarities in iron-binding sites, as well as the quantitative and structural properties of the iron-homeostasis related proteins across the 15 different algae. This work was based on *in-silico* analysis, focusing on conserved domains, transmembrane domains, and signal peptides. Protein alignments were performed to compare the position of binding sites, as well as the position of the identified structural components. Furthermore, heterologous transformations of *T. oceanica* genes into *P. tricornutum* were performed to localize the ISIP1 and ISIP3 proteins from *T. oceanica*.

SHORT-TERM RESPONSE TO IRON RESUPPLY IN IRON-LIMITED THALASSIOSIRA OCEANICA, AN OPEN OCEAN DIATOM, LEADS TO RAPID BUT DIFFERENTIAL DECAY OF IRON-RESPONSIVE TRANSCRIPTS

The discovery of iron-limitation in the ocean in the 1990s (Martin & Fitzwater, 1988; Martin, 1990b; Boyd *et al.*, 2007) led to various research projects aiming to understand the underlying concepts of survival in low-iron concentrations (Geider & La Roche, 1994; Sunda & Huntsman, 1995; Maldonado & Price, 2001; Peers & Price, 2006; Lommer *et al.*, 2012). Iron-fertilization experiments in the field resulted in mostly diatom-dominated algal blooms and led to laboratory work focusing on diatom cultures

(Kustka & Allen, 2007; Marchetti *et al.*, 2012; Chappell *et al.*, 2015; Cohen *et al.*, 2017). Some experiments were conducted to analyze the temporal response of iron-limited diatom species to iron addition, but significant transcripts were completely adjusted before the first measurement.

The main objective of this study was to obtain insight into the temporal response of iron-limited *T. oceanica* cells to the addition of iron on transcriptomic and physiological levels in a high-resolution fashion. A second objective was the analysis of mRNA half-lives and mechanisms influencing the stability of iron-responsive transcripts. Employing the NanoString technology, 54 transcripts were targeted to be measured in a temporal high-resolution fashion allowing the observation of transcript dynamics immediately after the addition of iron with five measuring points in the first hour after the addition of iron. A DUAL-PAM-100 was used for measurements of normalized variable fluorescence (F_v/F_m), as well as electron transport rate (ETR), photosystem efficiency (Y(II)), non-photochemical quenching (Y(NPQ)), and non-regulated heat dissipation (Y(NO). Cell characteristics such as cell size, granularity, and chlorophyll content were measured as relative units with a flow cytometer, and nitrogen per cell and carbon per cell were measured with a mass spectrometer.

IDENTIFICATION OF N-LINKED GLYCOSYLATION SITES AND TARGETED TRANSCRIPTOMICS REVEAL DIURNAL EXPRESSION OF MEMBERS OF THE N-LINKED GLYCOSYLATION PATHWAY AND A HIGH PERCENTAGE OF NXT-TYPE GLYCOSYLATION SITES IN THE OPEN OCEAN DIATOM THALASSIOSIRA OCEANICA

Iron-induced stress leads to substantial changes in the expression and translation of genes throughout all compartments of the cell (Lommer *et al.*, 2012; Smith *et al.*, 2016). Proteomic analysis of glycosylated proteins under varying iron concentrations are rare (Mathieu-Rivet *et al.*, 2013), and N-linked glycosylation is not only of interest for the understanding of adaptive mechanisms under iron-limitation but can also be used for the confirmation of predicted protein topologies. Furthermore, the successful expression of recombinant proteins for commercial purposes in diatoms (Vanier *et al.*, 2015, 2018) suggests a detailed understanding of the glycosylation machinery in diatoms.

The main objective of this study was to explore the qualitative and quantitative characteristics of N-linked glycosylated proteins and peptides under iron-limited and iron-replete conditions. Furthermore, an objective was the analysis of transcript dynamics of critical enzymes in the glycosylation pathway, providing insight into the N-linked glycosylation mechanism under iron-replete and iron-limiting conditions. A secondary objective was the use of glycosylation sites to confirm protein topologies. The N-linked glycosylated peptides were captured using the solid-phase extraction of N-linked glycosylated peptides. *T. oceanica* cells were grown in high-iron and low-iron conditions, with the high-iron conditions containing $^{15}NO_3$ as the only nitrogen source, allowing the quantification of the peptides present in both conditions.

SUCCESSFUL GENETIC TRANSFORMATION VIA MICROPROJECTILE BOMBARDMENT OF THE LOW-IRON ADAPTED OPEN OCEAN DIATOM THALASSIOSIRA OCEANICA LEADS TO ANTIBIOTIC RESISTANCE AND EXPRESSION OF THE GENE-OF-INTEREST

Genetic transformation of diatoms is a useful tool to gain insight into the function of specific proteins (Allen *et al.*, 2012; Kazamia *et al.*, 2018; McQuaid *et al.*, 2018). There are transformation systems for at least eleven diatom species described, but most cell biological research involving genetic transformation is done on *P. tricornutum*. The high diversity of diatom species with over 100,000 species (Armbrust, 2009) and the genetic evolutionary distance between diatom species (Lommer *et al.*, 2012; Mock *et al.*, 2017) supports the need for the development of further transformation systems.

The objective of this project was the development of a transformation system for *T. oceanica*, aiming to localize YFP-tagged ISIP proteins in *T. oceanica*. The project started with the design of a *T. oceanica*-specific vector using the Gibson Assembly® cloning technique. The introduction of the vector into *T. oceanic* cells was achieved via microparticle bombardment using a Biolistic® PDS-1000/He Particle Delivery System.

CHAPTER 2 IRON UPTAKE PROTEINS IN ALGAE AND THE ROLE OF IRON STARVATION-INDUCED PROTEINS (ISIP)

Joerg Behnke, Julie LaRoche

Submitted as:

Behnke, J. & LaRoche, J. (2019) Iron Uptake Proteins in Algae and the Role of Iron starvation-Induced Proteins (ISIPs). *Eur. J. Phycol.*

Contribution of authors:

Joerg Behnke:	Conducting experiments, data analysis, drafting of the manuscript
Julie LaRoche:	Planning and discussing the manuscript

2.0 Abstract

Iron is one of the most abundant elements on Earth, and it is essential for life. Despite the abundance of iron, its chemistry leads to very limited bioavailability, which has resulted in the evolution of a wide range of proteins involved in uptake and scavenging. However, because high intracellular concentrations of iron are toxic, iron homeostasis is essential for every organism. Approximately 30-40% of the ocean's surface is characterized by very low chlorophyll concentrations and high concentrations of nitrate and phosphate. These High Nutrient-Low Chlorophyll (HNLC) regions are also characterized by low concentrations of iron, and studies on low-iron adapted diatom strains revealed the presence of iron starvation-induced proteins (ISIP), a group of unrelated novel proteins that are strongly upregulated under iron limitation. First functional characterizations show that some of these ISIP proteins are involved in iron uptake. In the study presented here, the comparison of iron-uptake/storage proteins from 15 different algal species demonstrated that some well-characterized uptake proteins such as Natural Resistance Associated Macrophage Proteins (NRAMP) or the iron transporter1 (FTR1)-like permease have high similarities throughout the 15 algal species while others like zinc-regulated transporter (Zrt)- and iron-regulated transporter (Irt)-like proteins (ZIP) have a higher diversity amongst the algae analyzed here. The focus in this review is ISIP proteins, which are distinct from ZIP, NRAMP, and FTR1 permeases. ISIP2 proteins are transferrin-like proteins involved in Fe^{3+} uptake, and ISIP1 is responsible for endocytosis of siderophore bound iron. Further, the *in-silico* analysis combined with a localization study in a marine diatom suggests that ISIP3 acts as an iron storage protein. Overall, this analysis concludes that marine algae combine iron uptake strategies widespread in other

organisms with algal specific ISIPs as additional proteins for the utilization of diverse iron pools, thereby securing their success in iron-poor regions.

2.1 Introduction

Iron (Fe) is one of the most abundant elements on Earth and one of the most important trace metals for life. Iron is essential for all organisms, with a few exceptions (Archibald, 1983; Aguirre *et al.*, 2013), as an important cofactor in numerous proteins of central functions within the cell (Marchetti & Maldonado, 2016). The respiratory chain, photosynthesis (Raven et al., 1999) and nitrogen fixation (Whittaker *et al.*, 2011) are three fundamental biochemical processes with a high demand for iron (Weinberg, 1989; Geider & La Roche, 1994). Despite its high abundance in the Earth's crust, the concentration of iron is extremely low in the surface waters of up to 30% of the world's oceans (Morel & Price, 2003); given this, the dependence of cellular life on iron is surprising. However, the bioavailability of iron during the early origin of life was high in a reducing environment but changed drastically after the great oxygenation event in the late Archean/early Paleoproterozoic leading to the precipitation of $Fe(OH)_x$ (Ilbert & Bonnefoy, 2013; Lyons *et al.*, 2014).

In general, the major ocean gyres have very low iron and macronutrient concentrations. The subarctic north Pacific, the equatorial Pacific, and the Southern Ocean are regions referred to as High Nutrient-Low Chlorophyll (HNLC) regions because iron limits phytoplankton growth, resulting in excess concentrations of phosphate and nitrate. In the early 1990s, iron addition experiments in HNLC regions induced algal growth dominated by diatoms (Martin *et al.*, 1991; Boyd *et al.*, 2007). Subsequent *in-situ* iron addition experiments in several other HNLC regions established iron as a primary nutrient limiting

productivity in these areas (Boyd *et al.*, 2007; Strzepek *et al.*, 2011; Assmy *et al.*, 2013; Duprat *et al.*, 2016). Culturing experiments explored the effect of iron limitation on diatoms, and demonstrated differences between coastal and open ocean species, with the latter being better adapted to scavenge iron, resulting in higher growth rates under iron limiting conditions (Strzepek & Harrison, 2004; Peers *et al.*, 2005; Marchetti *et al.*, 2006; Peers & Price, 2006). Transcriptomic studies focusing on iron-limited diatoms (Marchetti *et al.*, 2006; Allen *et al.*, 2008; Lommer *et al.*, 2012; Nunn *et al.*, 2013) and green algae (Naumann *et al.*, 2007; Lelandais *et al.*, 2016) led to the discovery of three novel groups of proteins, coined iron starvation-induced proteins 1-3 (ISIP1-3), although they are structurally unrelated. ISIP proteins are of particular interest, as ISIP1 and ISIP2 are involved in iron uptake (Kazamia *et al*., 2018; McQuaid *et al*., 2018), and ISIP proteins are strongly upregulated under iron limitation. However, the response of algae to trace metal limitation is complex and recent laboratory work on *Thalassiosira oceanica,* and *Ostreococcus tauri* revealed different adaptation mechanisms to trace-metal limitation even between closely-related strains of the same species (Botebol *et al.*, 2017; Hippmann *et al.*, 2017; Luxem *et al.*, 2017). The interplay of iron with other trace metals was demonstrated in colimitation studies showing synergistic effects between Fe/Cu (Peers *et al.*, 2005) and Fe/Mn (Peers & Price, 2004; Pausch *et al.*, 2019).

In order to fully understand the complex role of iron in terms of its biological and biogeochemical cycling, it is important to consider iron speciation, iron concentrations, and ligand concentrations (Tagliabue *et al.*, 2017). The chemical forms of iron found in the ocean are operationally defined by size fractionation into particulate (>0.45 µm) and dissolved (<0.45µm). Dissolved iron is further categorized into colloidal (0.45-0.2 µm)

and soluble iron (<0.2 μm) (Wu, 2001). Iron appears in two oxidation states as ferrous (Fe^{2+}) ions, the reduced, more soluble, but thermodynamically unstable form, and ferric (Fe^{3+}) ions, the oxidized form (Figure 2.1). The solubility of Fe^{2+} is generally dependent on temperature, oxygen level, and pH of the surrounding environment. In warm oxygenated ocean water, Fe^{2+} is only stable for seconds, whereas, in cold suboxic waters, Fe^{2+} can last for several hours, up to days (Hoffmann *et al.*, 2012). As most of the modern ocean is well oxygenated, free Fe^{2+} is instantly oxidized to Fe^{3+}, which then, unless bound to organic ligands, will precipitate as ferric-hydroxide, making this form of iron only conditionally available for uptake. It is estimated that more than 99% of iron ions are bound to organic ligands (Gledhill & Buck, 2012), thereby keeping iron available for microorganisms that can access iron-ligand complexes (Kenshi, 2001). Iron chemistry follows a diel cycle based on photoreduction of Fe^{3+} as well as photoreduction of organic ligands (Butler & Theisen, 2010) (Figure 2.1). Although rising global temperature and acidification of the ocean will undoubtedly impact iron chemistry in future ocean scenarios, initial experimental studies have yielded conflicting results. One study analyzing iron chemistry in the context of temperature and pH demonstrated that decreasing pH and increasing temperature have opposing effects on the Fe^{2+} oxidation rate, resulting in longer/shorter half-life time for Fe^{2+} in the upper ocean, respectively (Samperio-Ramos *et al.*, 2016). In contrast, others demonstrated in laboratory experiments with phytoplankton an overall negative effect on the bioavailability of iron following a decrease in pH (Shi *et al.*, 2010).

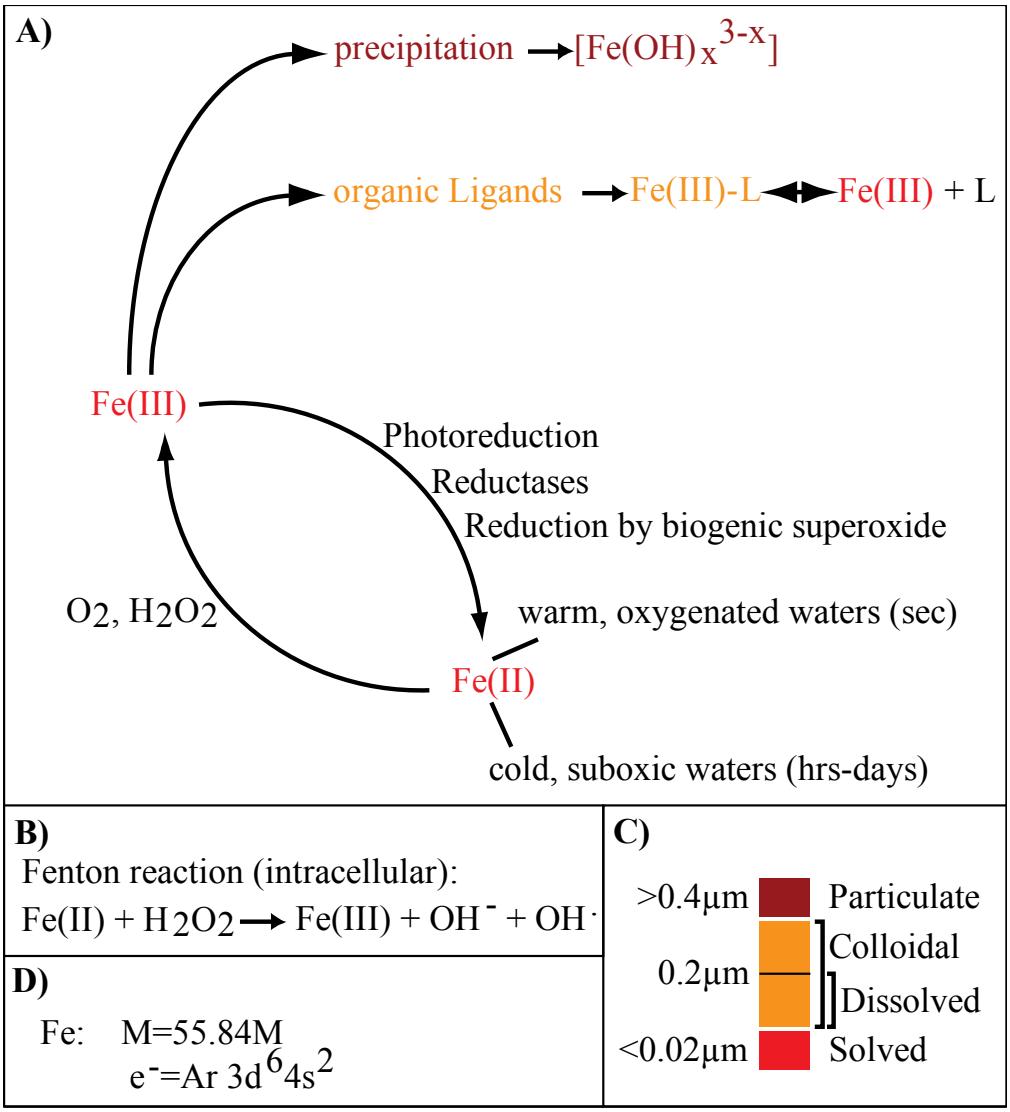


Figure 2.1: Overview of iron chemistry and iron speciation in the ocean.

A) Mechanisms of oxidation and reduction of iron are shown. B) Fenton reaction leading to intracellular production of reactive oxygen species (ROS). C) Operational classification of iron in the ocean based on filter size. The colors are used to show iron speciation in A). D) Molecular weight and electron configuration of iron.

2.1.1 Transport of ferrous iron (Fe^{2+})

Fe^{2+} uptake is carried out by proteins involved in high-affinity uptake systems and general divalent metal ion transporters. The two most widely distributed eukaryotic divalent metal-ion uptake protein families are Natural Resistance Associated Protein

(NRAMP) and Zinc Regulated Transporter (ZRT)-Iron Regulated Transporter (IRT)-like proteins (ZIP). The comparable bacterial proteins that are most likely the ancestors of the eukaryotic proteins are manganese transporter MntH and zinc uptake transporter ZupT (Lau *et al.*, 2016). In bacteria, these proteins are either located in the periplasmic space or are transmembrane proteins spanning the inner membrane of gram-negative bacteria. With 11 transmembrane domains (TM), the MntH protein belongs to the NRAMP family and can take up Fe^{2+} and Mn^{2+} (Courville *et al.*, 2004). The bacterial ZupT transporter consists of 8 TM domains and belongs to the ZIP family. ZIP proteins interact with a wide range of substrates and cannot solely be considered a Fe^{2+} transporter (Grass *et al.*, 2005). Bacteria possess a third specialized Fe^{2+} uptake system, the elemental ferrous iron uptake system (EfeUOP). This system consists of three proteins: an iron permease EfeU with homology to the iron-permease FTR1 in yeast, and two additional proteins with unidentified functions that are located in the periplasmic space (Große *et al.*, 2006).

In *Arabidopsis thaliana,* one of the best-studied plant models, three different protein families are involved in Fe^{2+} uptake. The first two families belong to the IRT(ZIP)- and NRAMP-protein family, each showing specific tissue localization and protein expression patterns depending on the metal concentration in soil (Kim & Guerinot, 2007; Ding *et al.*, 2017). IRT1_*At*, for example, is upregulated under low iron conditions and is part of a high-affinity uptake system (Vert *et al.*, 2002). NRAMP proteins in *A. thaliana* are often intracellular, and function in iron-homeostasis rather than iron uptake (Thomine & Ji, 2015), but they are also described as high-affinity transporters for manganese (Cailliatte *et al.*, 2010). NRAMP proteins are also found in humans as Divalent-Metal-Cation Transporter (DMT-1) (Fuqua *et al.*, 2012) and in yeast as Suppressor of Mif (SMF)

(Cohen *et al.*, 2000). The third protein family in *A.thaliana* is Yellow Stripe1-like proteins (YSL), which have a wide range of functions. YSL proteins take up Fe^{2+} directly and are involved in iron uptake via nicotianamine complexes (DiDonato *et al.*, 2004) as well as phytosiderophore complexes in general (Curie *et al.*, 2001).

2.1.2 Uptake of ferric iron (Fe^{3+}) and siderophore mediated uptake

Three main mechanisms have evolved for Fe^{3+} uptake. Transferrin proteins are capable of direct uptake of Fe^{3+}, and present in mammals, plants, and algae (Paz *et al.*, 2007; Bai *et al.*, 2016; McQuaid *et al.*, 2018). Another strategy involves the production of Fe^{3+} -binding organic ligands. In bacteria, these organic ligands, called siderophores, are synthesized from amino acids via non-ribosomal peptide synthases. Siderophores are small molecules of less than 1.5 kDa with a high affinity for Fe^{3+}. The selectivity for Fe^{3+} prevents the binding of other dissolved divalent metal ions. Specific transporter proteins are responsible for uptake of the siderophore-iron complex, and iron is released from the complex via reduction to Fe^{2+} (Harrington & Crumbliss, 2009). In general, siderophores are weak acids forming octahedral structures that chelate Fe^{3+} via electronegative oxygen atoms (Hider & Kong, 2010). They are classified as catechols, hydroxamates, or α-hydroxy carboxylates, according to their chemical structure. Plants secrete phyto-siderophores, which are small acidic molecules synthesized by different enzymes such as nicotianamine synthases (Masuda *et al.*, 2017). Algae do not produce siderophores, but some algal species can use iron-siderophores produced by other organisms (Soria-Dengg & Horstmann, 1995; Maldonado & Price, 2001; Kazamia *et al.*, 2018).

2.1.3 Reductive high-affinity iron uptake systems

The second strategy for Fe^{3+} uptake involves widespread high-affinity uptake systems that are found in plants, algae, fungi, and humans. The main components of a reductive high-affinity uptake system include an iron-reductase (FRE), a multicopper oxidase (MCO), and an iron-permease. The high-affinity iron uptake system of the yeast *Saccharomyces cerevisiae* is one of the best-studied (Kosman, 2003). The genome of *S. cerevisiae* contains eight iron-reductases, one plasma membrane iron-permease, and two MCOs (Figure 2.2). The first step of a high-affinity uptake system is the reduction of Fe^{3+} by FRE, creating a pool of soluble Fe^{2+} in close vicinity to the cell. The accumulated pool of Fe^{2+} is re-oxidized to Fe^{3+} via an MCO. The iron permease then internalizes the oxidized Fe^{3+}. The relevant enzymes are in close vicinity to each other to ensure the effectiveness of the system (Kwok *et al.*, 2006). Higher plants, capable of a reductive high-affinity uptake, are missing the MCO (Thomine & Lanquar 2010), but oxidation plays a vital role in iron uptake in yeast and algae (Maldonado *et al.*, 2006; Terzulli & Kosman, 2010; Thomine & Lanquar, 2010).

2.1.4 Iron uptake in algae

Algae possess diversified systems for iron uptake that include reductive high-affinity iron uptake systems (Maldonado & Price, 1999, 2001; Terzulli & Kosman, 2010), low-affinity metal transporters capable of iron transport (Allen *et al.*, 2008; Smith *et al.*, 2016), phytotransferrins (Morrissey *et al.*, 2015; McQuaid *et al.*, 2018), and the ability to take up siderophores (Kazamia *et al.*, 2018). Overall, our knowledge of iron uptake systems in marine algae is often solely based on *in-silico* analysis, and biochemical characterization of the diverse components is still lacking. This study compared proteins involved in iron

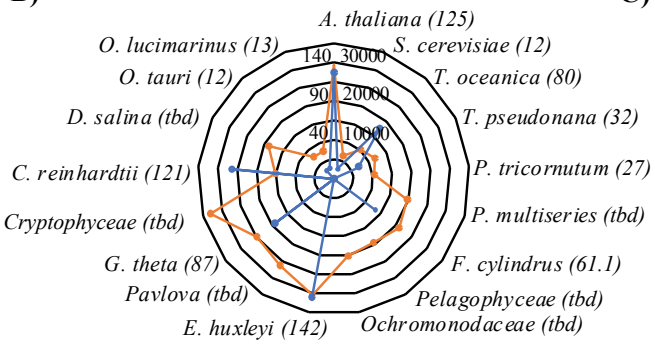
uptake and iron storage from 15 different algal species, ranging from the freshwater green algae *Chlamydomonas reinhardtii* to a collection of recently available genomes from arctic algae. Overall, our algae cover four major groups: the supergroup of stramenopiles, alveolates and Rhizaria (SAR), haptophytes, cryptophytes, and chlorophytes (Figure 2.2) (Palenik *et al.*, 2007; Bowler *et al.*, 2008; Curtis *et al.*, 2012; Lommer *et al.*, 2012; Read *et al.*, 2013; E.W Juergen *et al.*, 2017; Mock *et al.*, 2017). *A. thaliana* and *S. cerevisiae* were included to achieve a better comparison to well-characterized iron uptake systems. In order to maintain consistency with known iron uptake systems, the analysis included proteins of known function such as ferritin, FRE, NRAMP, permease, and ZIP, in addition to ISIP1, ISIP2, and ISIP3, the poorly characterized novel group of iron-related proteins found primarily in marine algae (Figure 2.2). This study combines recently published work on ISIP1 and ISIP2 with new results from a cellular localization study of members of the ISIP proteins.

A)

	ZIP	FRE	NRAMP	MCO	Ferritin	ISIP2	ISIP3	Permease	ISIP1	References
A. thaliana	**15**	**5**	**7**	**0**	**11**	**0**	**0**	**0**	**0**	uniprot.org
S. cerevisiae	**5**	**8**	**1**	**1**	**0**	**0**	**0**	**1**	**0**	uniprot.org
T. oceanica	**8**	**2**	**1**	**2**	**0**	**2**	**1**	**0**	**2**	Lommer et al. 2012
T. pseudonana	**7**	**4**	**1**	**1**	**0**	**1**	**1**	**2**	**0**	Armbrust et al. 2004
P. tricornutum	**10**	**5**	**0**	**0**	**1**	**2**	**1**	**0**	**1**	Bowler et al. 2008
P. multiseries	**10**	**1**	**1**	**1**	**2**	**0**	**0**	**0**	**0**	jgi
F. cylindrus	**8**	**1**	**4**	**11**	**1**	**2**	**1**	**1**	**1**	Mock et al. 2017
Pelagophyceae	**8**	**3**	**2**	**6**	**0**	**1**	**2**	**0**	**0**	Lovejoy, jgi
Ochromonodaceae	**8**	**1**	**1**	**1**	**0**	**2**	**0**	**0**	**0**	Lovejoy, jgi
E. huxleyi	**16**	**2**	**4**	**0**	**0**	**1**	**1**	**0**	**0**	Read et al. 2013
Pavlova	**8**	**3**	**1**	**2**	**0**	**2**	**1**	**2**	**0**	Lovejoy, jgi
G. theta	**16**	**0**	**4**	**0**	**1**	**1**	**1**	**2**	**0**	Curtis et al. 2012
Crytophyceae	**7**	**0**	**1**	**0**	**1**	**1**	**1**	**1**	**0**	Lovejoy, jgi
C. reinhardtii	**10**	**1**	**1**	**2**	**2**	**2**	**0**	**1**	**0**	Merchant et al. 2007
D. salina	**5**	**1**	**1**	**1**	**1**	**2**	**0**	**0**	**1**	Polle et al. 2017
O. tauri	**5**	**1**	**1**	**0**	**1**	**1**	**0**	**0**	**0**	Palenik et al. 2007
O. lucimarinus	**8**	**0**	**1**	**0**	**2**	**1**	**0**	**0**	**0**	Palenik et al. 2007
total	**154**	**37**	**32**	**28**	**23**	**21**	**10**	**10**	**5**	

Arabidopsis thaliana | *Saccharomyces cerevisiae* | SAR supergroup
Chlorophytes | Haptophytes | Cryptophytes

B)



C)

Function	Protein-Family	Binding sites
Fe^{2+}-uptake	ZIP	Poly-histidine sites
	NRAMP	DNAM
Fe^{3+}-uptake	Permease	[R/AEXXE]
	Transferrin	DYRYH
Siderophore-uptake	ISIP1	unknown
Unknown	ISIP3	unknown
Storage	Ferritin	EEEHEQE

Figure 2.2: Overview of species analyzed in this study with quantitative data regarding their iron uptake/storage related proteins.

A) The overview table shows quantities of proteins (decreasing total number from left to right) in each species analyzed in this study with the genome reference in the last column. B) Overview figure of genome size (blue line, scale 100-190 (exact number is in brackets beside the species name)) in Mbp and total number of genes (orange line, scale 10000-35000), Genome accession numbers for the species: *A. thaliana* GCA_000001735.1, *S. cerevisiae* GCA_000146045.2, *T. oceanica* AGNL01000000, *T. pseudonana* AAFD01000000, *P. tricornutum* ABQD00000000, *P. multiseries* PRJNA32659, *F.*

cylindrus LFJG00000000, *Pelagophyceae* PRJNA210205, *Ochromonodaceae* PRJNA171379, *E. huxleyi* AHAL01000000, *Pavlovales* PRJNA223446, *G. theta* AEIE00000000, *Cryptophyceae* CCMP2293 v1.0(jgi), *C. reinhardtii* ABCN01000000, *D. salina* NSFN01000000, *O. tauri* GCA_000214015.2, *O. lucimarinus* PRJNA337613 C) Table of iron uptake/storage proteins including their function and their respective iron binding sites.

2.2 Materials and methods

2.2.1 Culturing, RNA extraction, and vector design

Axenic *T. oceanica* (CCMP1005) and axenic *Phaeodactylum tricornutum* (CCMP2561) were grown in artificial seawater (ASW) f/2 (Goldman *et al.*, 1978) at 22ºC and 100 µmol photons $m^{-2}s^{-1}$ light in a 14/10 h day/night cycle. RNA from *T. oceanica* was extracted using the RNeasy plant mini kit (Qiagen, Inc., Valencia, CA, USA), and cDNA was generated with Superscript III (Invitrogen, Carlsbad, CA, USA). ISIP1 and ISIP3 were amplified with the following primers: ISIP1a-cDNA-Fw: 5'-$_{1}$ATGAAGTTCCCGGTCGCCCTC$_{21}$-3'ISIP1a-cDNA-Rv:5'-$_{1824}$CTAAG CTACCTGCTTTGAGCC$_{1804}$-3', ISIP3-cDNA-Fw: 5'-$_{1}$ATGTACCTTGTCACTGCCTTG$_{21}$-3',ISIP3-cDNA-Rv: 5'-$_{1380}$TCACAAGAACAAGAAGGTCTT$_{1360}$-3'. ISIP1 and ISIP3 were introduced into the pDEST vector (Andrew Allen, Scripps Institution, LaJolla, CA, USA) using the Gateway® system (Invitrogen, Carlsbad, CA, USA).

2.2.2 Transformation and microscopy of *P. tricornutum*

P. tricornutum cells were grown to mid- to late-exponential phase. $1x10^{8}$ cells were concentrated and placed in the middle of an ASW plate (1.2% Agarose). After 24 h, cells were bombarded with 1.1 µm (M17) tungsten beads (Bio-Rad, Hercules, CA, USA) at a pressure of 1550 psi with a PDS-1000/He system Genegun (Bio-Rad, Hercules, CA,

USA). After 24 h, cells were scraped off the plate and plated on zeocin-containing plates (100 µg/ml). Colonies were picked after 2-4 weeks and maintained in selective media (50 µg/ml zeocin) for further analysis.

2.2.3 Data collection and protein alignment

All protein sequences were retrieved from Uniprot or Joint Genome Institute (JGI, Walnut Creek, CA, USA). Proteins that were annotated as a transporter for metals other than iron are not included in the analysis. Genome sizes and the total number of genes were retrieved from NCBI and JGI (JGI, Walnut Creek, CA, USA). Proteins were aligned in MEGA7vs7.0.21 using MUSCLE with default settings (Kumar *et al.*, 2016), and the alignments were illustrated with Jalview (Waterhouse *et al.*, 2009).

2.2.4 *In-silico* analysis of protein sequences

Protein sequences were analyzed for transmembrane domains using Topcons (Shu *et al.*, 2015). Conserved domains were identified with the NCBI conserved domain search (Marchler-Bauer *et al.*, 2017). The graphical illustration of transmembrane domains and conserved domains was done with iTOL (Letunic & Bork, 2016). N-glycosylation sites were predicted with the NetNGlyc 1.0 Server (Chuang *et al.*, 2012), and cellular localization of the proteins was predicted with the TargetP 1.1 Server (Emanuelsson *et al.*, 2000). The presence of signal peptides was predicted with SignalP 4.1 Server (Petersen *et al.*, 2011), and chloroplast localization was analyzed with the ChloroP 1.1 Server (Emanuelsson *et al.*, 2000).

2.3 Results

The primary purpose of this analysis was to create an overview of iron uptake strategies and to provide insight into iron-binding sites found in iron uptake proteins. Within the result section, the two Fe^{2+} uptake proteins, ZIP, and NRAMP, are first analyzed, followed by the iron permease FTR1. The analysis of iron-binding sites in the iron storage protein ferritin and the results on the novel group of iron-related proteins, the ISIP proteins, will be the last part of this section.

2.3.1 ZIP (Zinc-regulated, Iron-regulated Protein)

The large number of ZIP proteins in each species suggests that this protein family must cover a range of functions beyond iron uptake at the cell surface. ZIP proteins are involved in general divalent metal uptake, function as high-affinity iron uptake proteins in *A. thaliana* (Castaings *et al.*, 2016), and they are often differentially expressed in *A. thaliana* in response to environmental stressors (Guerinot, 2000). They are widely distributed across all algal species in this study, but the specific functions are unknown. ZIP protein expression in diatoms showed a mixed response to iron addition in field samples, depending on the sample site and diatom genera (Cohen *et al.*, 2017; Lampe *et al.*, 2018). The divalent metal ion binding capacity is likely linked to the poly-histidine sites that are located between the third and fourth TM domain, as shown for *A. thaliana* (Ivanov & Bauer, 2017). Approximately one-third of the ZIP sequences in the dataset had a similar metal-binding motif to that found in *A. thaliana*, with several histidine residues in close vicinity to one another (Figure 2.3). These sequences are distributed across all algal species. The ZIP proteins in microalgae are diverse in primary structure, especially in terms of the position of TM domains and the characteristics of poly-histidine sites

(Figure 2.3). Further laboratory work is needed to demonstrate the specific functions of each of these proteins.

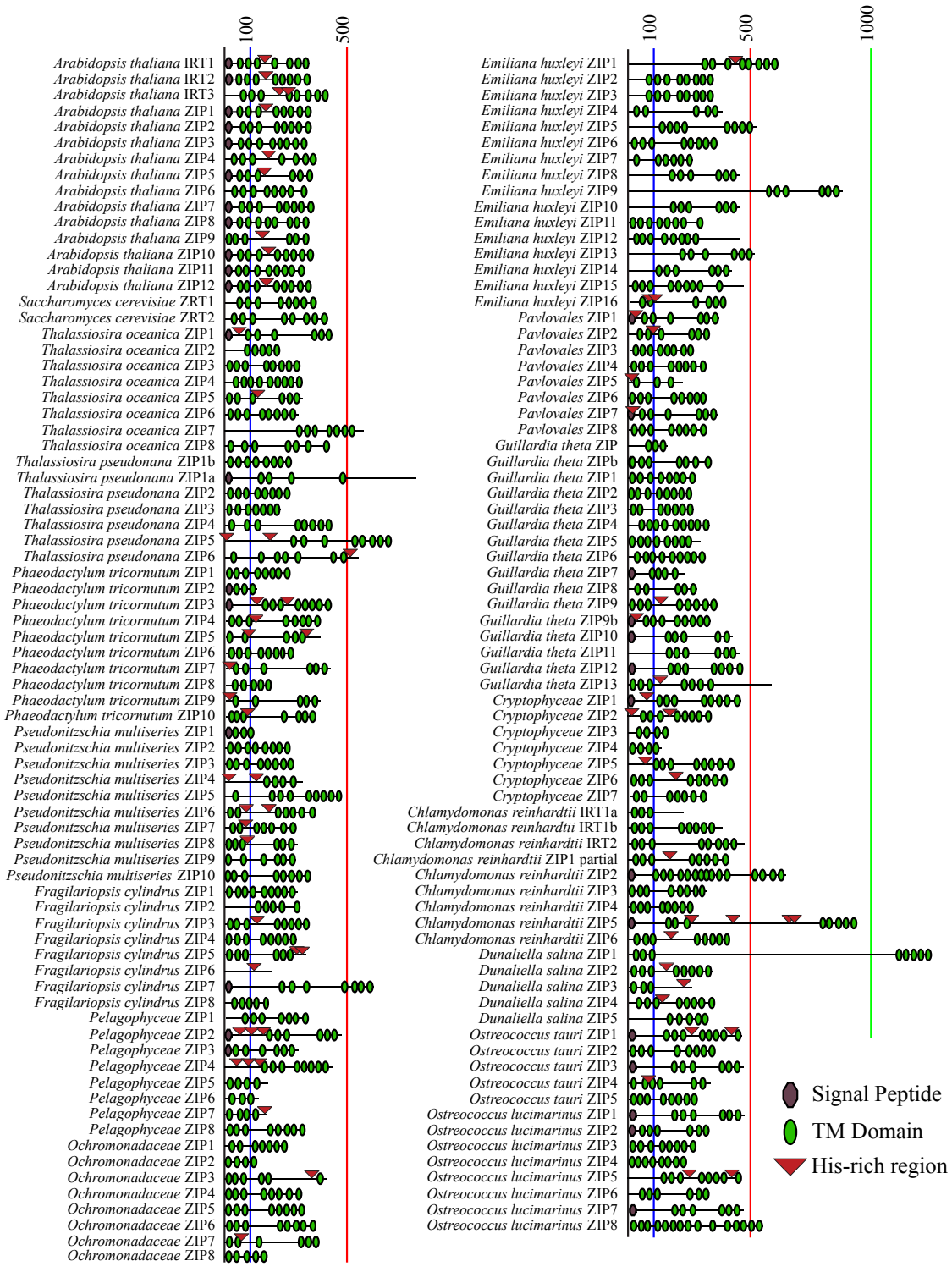


Figure 2.3: Overview of structural elements and histidine-rich sites in ZIP proteins.

Protein backbones for each protein are shown as black lines, representing the length of the protein as indicated by the blue, red, and green line. Transmembrane domains (green circles) and signal peptides (purple hexagons) are indicated along the backbone. Regions

with four or more histidine-residues within 10 amino acids are indicated by a red triangle. Protein backbone was generated with iTOL and manually curated for the display of poly-histidine sites. Signal peptide and transmembrane domains were predicted with Topcons (Shu *et al*., 2015).

2.3.2 NRAMP

Like ZIP proteins, NRAMP proteins function as general divalent metal ion transporters capable of Fe^{2+} uptake, as well as intracellular Fe^{2+} transport and were recovered in all algal species analyzed here except *P. tricornutum* (Figure 2.4). Specific functions of these proteins in microalgae have yet to be investigated at a physiological and biochemical level, but a role in iron uptake (Kustka & Allen, 2007) or iron release from vacuoles as shown for *A. thaliana* (Thomine *et al.*, 2003) is possible (Lampe *et al.*, 2018). A sequence alignment with the confirmed metal-binding site of the ortholog MnTh of the bacterium *Deiniococcus radiourans* (Bozzi *et al.*, 2016) shows a high conservation of binding sites in most NRAMP proteins recovered from the microalgal genomes used in this study. Four amino acids, D_{56}, N_{59}, $A_{227,}$ and M_{230}, are responsible for metal-coordination with D_{56} (91%), and N_{59} (85%), showing the highest conversation within the dataset. In *D. radiourans* the methionine M_{230}, found in 85% of the sequences (Figure 2.4), is known to inhibit the binding of highly abundant divalent metals such as Ca^{2+} and Mg^{2+} while still transporting Fe^{2+} and Mn^{2+} (Bozzi *et al.*, 2016). The structural analysis results in an average of 11 TMs for each of the sequences. The NRAMP from the recently sequenced genomes of the Arctic haptophyte *Pavlovales* has only three TMs, and none of the conserved binding domains are present, suggesting that this variant is most likely a pseudogene. In contrast, although the average length of NRAMP proteins is 504 amino acids comprising an average of 11 TM domains, the NRAMP1 of Pelagophyceae is much longer at 835 amino acids comprising 16 TM. The position of

binding motifs also differs from the other proteins, as the first binding domain, found in or close to the first TM in all other sequences, is present in the 5th TM in *Pelagophyceae*. Whether this sequence represents a truly functional NRAMP remains to be shown. The EIN2 protein from *A. thaliana* is the longest NRAMP in this dataset, with 1294 amino acids. Despite a long "tail", EIN2 appears structurally very similar to the other sequences, but its function is regulatory rather than divalent metal transport (Alonso *et al.*, 1999).

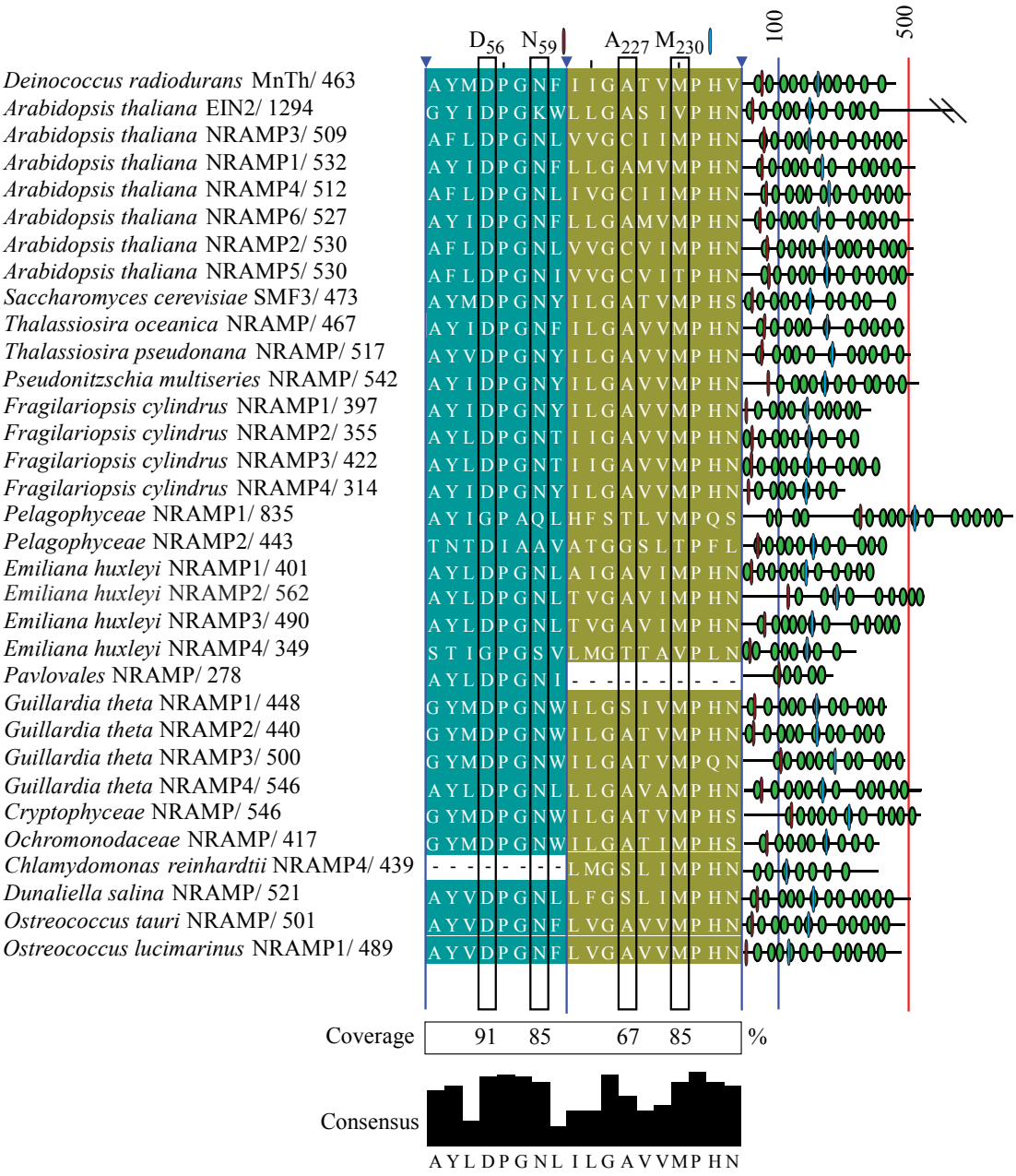


Figure 2.4: Protein alignment of metal-binding sites in NRAMP proteins and structural information.

Sequence alignment was generated with MUSCLE in Mega7.0 (Kumar et al., 2016), indicating the metal-binding sites in comparison to a bacterial NRAMP protein from *Deiniococcus radiourans* MnTh (Bozzi *et al.*, 2016). Black boxes highlight the amino acids (one-letter code) responsible for iron-binding. The colors in the alignment separate the two regions where the binding sites are present. The coverage and the consensus sequence are below the alignment. Backbones for each protein are shown as black lines representing the length of the protein, including the position of transmembrane domains (green circles, predicted by TOPCONS (Shu *et al.*, 2015)) and the two binding domains (red/blue line).

The length of the protein is indicated by blue and red lines representing 100 and 500 amino acids, respectively, as well as indicated next to the species name.

2.3.3 Permease (FTR1), ferric reductases (FRE) and multicopper oxidases (MCO)

The three components of a reductive high-affinity iron uptake system are found in all four groups of microalgae analyzed here, an expected outcome given the wide distribution of these proteins throughout several diatom classes (Groussman *et al.*, 2015). Despite structural similarities in MCO proteins and FRE proteins characterized in diatoms (Groussman *et al.*, 2015), the analysis of these two proteins is complicated by their presence in multiple metabolic pathways and a cell biology approach is required to conclusively determine localization and functions of these proteins (Schröder *et al.*, 2003; Sakurai & Kataoka, 2007). However, FRE proteins and MCO proteins are upregulated under low iron conditions in diatoms (Kustka & Allen, 2007; Allen *et al.*, 2008; Lommer *et al.*, 2012; Smith *et al.*, 2016) and iron uptake in diatoms is partially based on the reduction of iron (Maldonado & Price, 2001; Shaked *et al.*, 2005; Maldonado *et al.*, 2006). The absence of MCO and FRE in the genome of both cryptophytes included in this study suggests a solely non-reductive iron uptake system for both species.

The uptake of Fe^{3+} via the permease FTR1 in *S. cerevisiae* requires four protein motifs: two EXXE motifs with X being any amino acid, a DASE motif, and an EDWLE motif (Severance *et al.*, 2004). Functional characterization of FTR1 proteins in *S. cerevisiae* and *Candida albicans* revealed differences concerning the required motifs. Therefore, FTR1_*Ca* was included in the protein alignment in Figure 2.5. Whereas in *S. cerevisiae,* two EXXE motifs are essential for iron uptake, *C. albicans* requires only one EXXE motif for iron uptake (Fang & Wang, 2002). The amino acid in front of an EXXE motif must be an arginine or a lysine, and the EXXE motifs in yeast are located in the first and

fourth transmembrane domains, respectively (Fang & Wang, 2002; Severance *et al.*, 2004). The protein alignment in Figure 2.5 shows the location of all four essential motifs in *S. cerevisiae*. All sequences have the $EXXE_{16}$ motif with arginine in front. The $EXXE_{157}$ motif from *S. cerevisiae* and *C. albicans* is only present in the yeast sequences, as well as in *C. reinhardtii* and *G. theta*. The other species have two alternative EXXE motifs, $EXXE_{Alt\text{-}1}$ and $EXXE_{Alt\text{-}2}$. $EXXE_{Alt\text{-}1}$ has an arginine in front and is located in the fourth transmembrane domain, similar to $EXXE_{158}$. The EDLWE motif is not fully conserved across all sequences, but the glutamic acid residues at the end of the motif that are required for iron permeation are conserved in 91% of the sequences. The only motif absent from all sequences other than yeast is the DASE motif which has been discussed as a direct "connector" to the iron oxidase Fet3p (Severance *et al.*, 2004). This may be a yeast-specific motif that is replaced by a different motif, such as the $EXXE_{Alt\text{-}2}$, in other organisms based on their specific interaction partners.

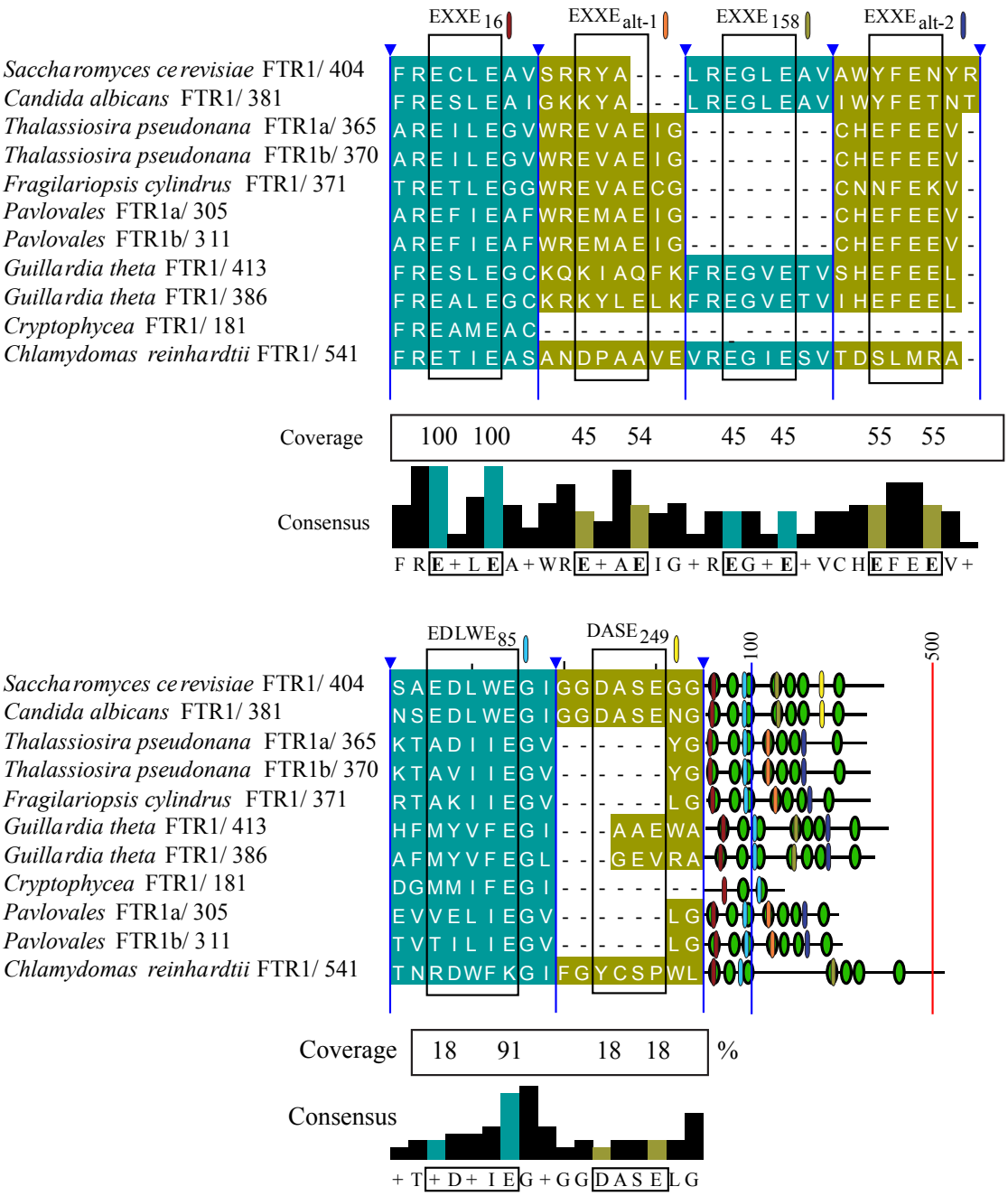


Figure 2.5: Protein alignment with binding motifs and overview of transmembrane domains in FTR1 permeases.

Sequence alignment (one-letter code) was done with MUSCLE in Mega7.0 (Kumar *et al.*, 2016). The alignment concerning iron-binding motifs from *S. cerevisiae* (Severance *et al.*, 2004) and *C. albicans* (Fang & Wang, 2002) are shown and indicated as black boxes. Colors in the alignment indicate different regions, which are also separated by a blue arrow. The consensus sequence and the coverage are shown below the alignment. The location of transmembrane domains is shown as green circles, predicted by TOPCONS (Shu *et al.*,

2015), and the different iron-binding motifs are indicated on the protein backbone, which is represented as a black line. Protein backbone was generated in iTOL (Letunic & Bork, 2016) and manually curated concerning the position of the binding motifs. Protein length is indicated next to the species name.

2.3.4 Ferritin

Ferritin is an iron storage protein, present in plants and animals, which provides a buffer for periods of low iron conditions or regulates the diel cycle of iron (Marchetti *et al.*, 2009; Botebol *et al.*, 2015; Lampe *et al.*, 2018). Iron binding sites in the ferritin of *P. multiseries* include several glutamic acid residues and one histidine residue (Pfaffen *et al.*, 2013). Three of these glutamic acid residues show very high conservation (87-96%) in the protein alignment (Figure 2.6). The glutamic acid residue GLU130, present only in *P. multiseries*, has been further suggested as an essential factor for the oxidation of iron, resulting in an overall shift in the function from storage to oxidation (Pfaffen *et al.*, 2015). Overall, the haptophyte *Emiliana huxelyi* and the newly sequenced *Pavlovales sp.* lack the iron storage protein ferritin, whereas all chlorophytes in this study contain ferritin. The high number of ferritin proteins in *A. thaliana* is based on sequences retrieved from uniprot. A blastp search in the most recent genome of *A. thaliana* resulted in only four ferritin proteins (Ferritin 1-4).

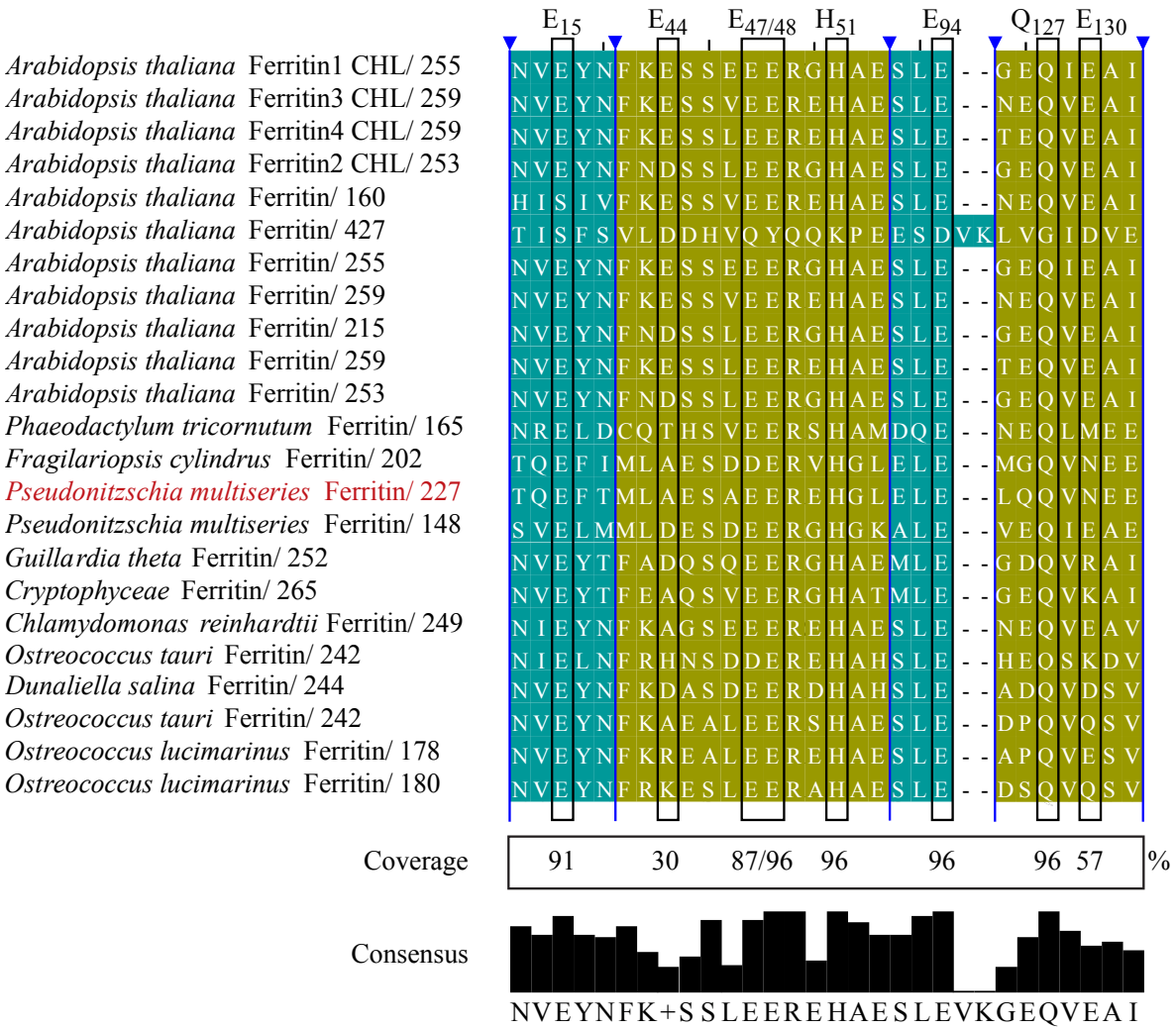


Figure 2.6: Ferritin

Protein sequence alignment (one-letter code) was done with MUSCLE in Mega7.0 (Kumar *et al*., 2016). The binding sites from *Pseudonitzschia multiseries* ferritin1 are highlighted in red. Colors in the alignment separate the different regions. The start and finish of each region are indicated by a blue arrow. The coverage and consensus sequence are shown below. The number following the species name is the protein length.

2.3.5 Iron Starvation Induced Proteins (ISIP)

The initial characterization of the ISIP genes, as iron-starvation induced genes, was made in the pennate diatom *P. tricornutum* (Allen *et al.*, 2008). ISIP proteins are separated into three groups, ISIP1, ISIP2, and ISIP3, with no similarity to each other, as demonstrated from primary amino acid sequences. Although ISIP proteins were first discovered in the

genomes of diatoms, they have since been reported in the genome of numerous marine algal species (Bowler *et al.*, 2008; Lommer *et al.*, 2012; Lelandais *et al.*, 2016; Mock *et al.*, 2017), as well as in metagenomic and metatranscriptomic analysis of algal communities (Marchetti *et al.*, 2012; Morrissey *et al.*, 2015; Kazamia *et al.*, 2018; Caputi *et al.*, 2019). The first functional characterization of ISIPs in *P. tricornutum* revealed that ISIP1 and ISIP2 are involved in two different iron uptake strategies (Morrissey *et al.*, 2015; Kazamia *et al.*, 2018; McQuaid *et al.*, 2018).

T. oceanica is the only species in this analysis that has two isoforms of ISIP1, ISIP1a, and ISIP1b. However, while ISIP1a is heavily transcribed under iron limitation, no transcript for ISIP1b could be detected, suggesting that the latter might have originated from a gene duplication with a loss of function (Lommer *et al.*, 2012). ISIP1 has been identified mostly in diatoms, but extensive metagenomic studies like the Tara Oceans dataset and the MMETSP databank revealed the presence of ISIP1 in chlorophytes, pelagophytes, haptophytes and dinoflagellates (Kazamia *et al.*, 2018). ISIP1 in *P. tricornutum* is involved in endocytosis mediated siderophore transport, leading an iron-siderophore complex to the chloroplast and supplying iron for photosynthesis (Kazamia *et al.*, 2018). The structural overview of three ISIP1 proteins from diatoms and p130b from *D. salina* shows that all four proteins have one transmembrane domain with an intracellular C-terminus. P130b has a high similarity to ISIP1 and is therefore included in the structural comparison (Figure 2.7). The three ISIP1 from diatoms have a short C-terminal intracellular domain with a putative endocytosis motif. In all four proteins, the extracellular N-terminal domain has a predicted β-propeller configuration. This extracellular domain contains two CXXC motifs in the diatom ISIP1 proteins (Figure

2.7). A heterologous transformation of ISIP1_*To* into *P. tricornutum* shows that ISIP1_*To* is located on the cell surface (Figure 2.8).

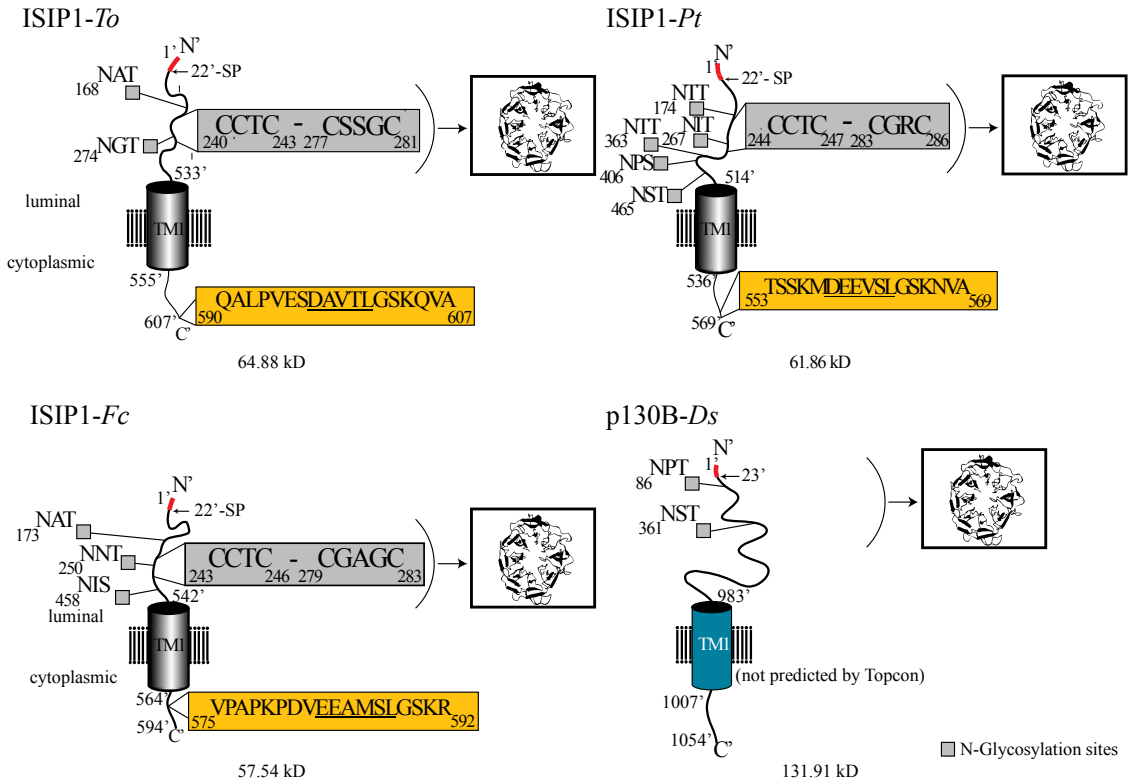


Figure 2.7: Structural overview of ISIP1 proteins.

Shown are ISIP1 proteins from three diatom species and p130B from *D.* salina. CPH modeler was used for structural information (in Box) (Nielsen *et al.*, 2010). Indicated are putatively N-glycosylation motifs, transmembrane domains, and signal peptides. NetNGlyc 1.0 Server and TOPCONS server were used, respectively (Petersen *et al.*, 2011; Chuang *et al.*, 2012; Tsirigos *et al.*, 2015). The signal peptide is indicated in red, and the cut-off is marked with an arrow. Grey boxes represent functional motifs, CXXC, with X being any amino acid. Orange boxes show endocytosis motifs (one-letter code). *To*: *Thalassiosira oceanica*; *Pt*: *Phaeodactylum tricornutum*; *Fc*: *Fragilariopsis cylindrus*; *Ds*: *Dunaliella salina*.

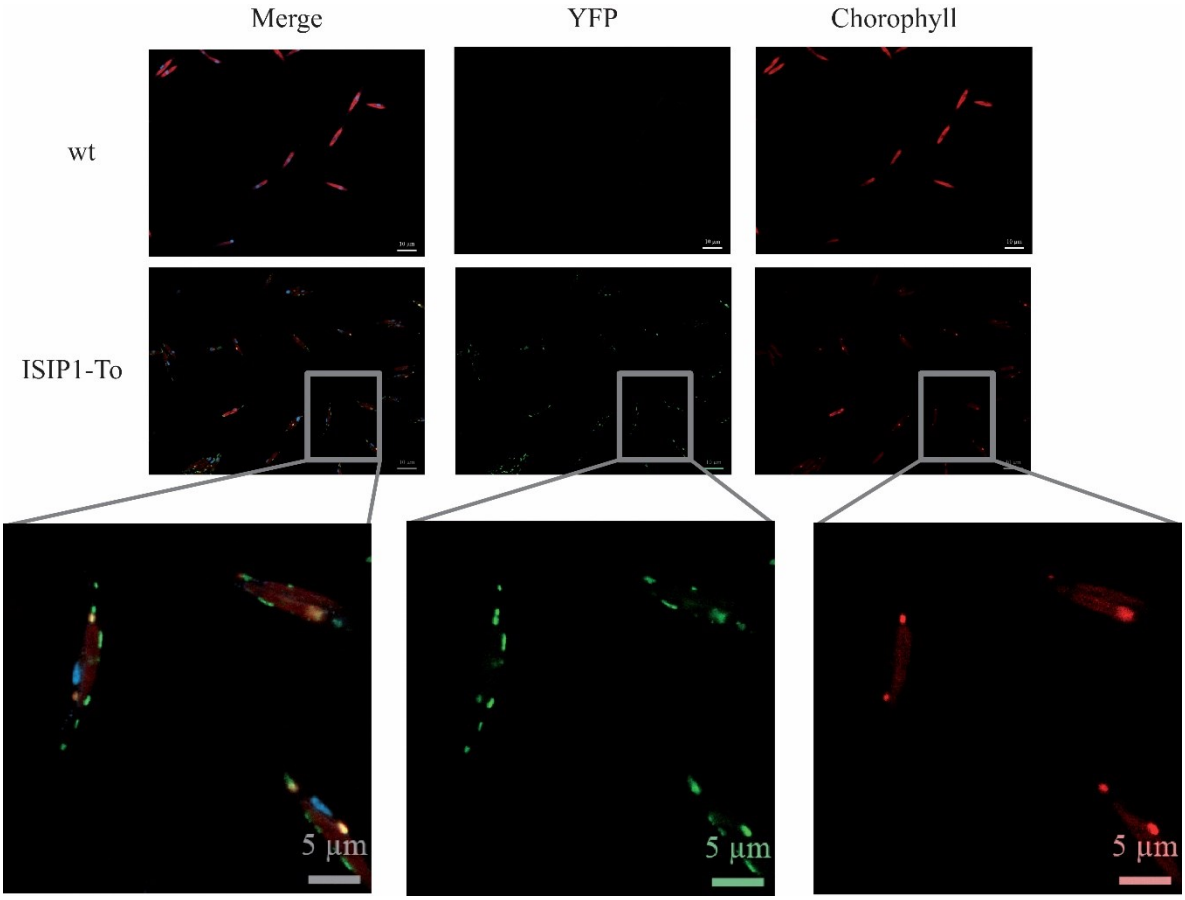


Figure 2.8: Heterologous transformation of ISIP1_*To* in *P. tricornutum*.

ISIP1_*To* locates at the cell surface. Nuclei are stained with DAPI and appear blue. Chlorophyll is shown in red and ISIP1-YFP in green. A panel of untransformed cells is shown (wt) on top of the transformed cells (ISIP1-*To*). Scale bars are indicated as 5 µm (magnified images) and 10 µm in others.

In contrast to ISIP1, which is predominantly found in diatoms, ISIP2 has a broader distribution in algae and is present in all algal species included in this study. ISIP2 contains two subdomains, a C-terminal subdomain, and an N-terminal subdomain (Lommer *et al.*, 2012). The C-terminal domain has similarity to the iron-assimilating proteins (FEA), which are involved in iron uptake in *C. reinhardtii* (Allen *et al.*, 2007). *T. oceanica* has two copies of the ISIP2 genes, a short version and an extended version containing an eight-fold duplication of the C-terminal domain. Similar extended proteins

with duplications of the ISIP2 C-terminal domain are found in *Fragilariopsis cylindrus*, *Thalassiosira pseudonana,* and *Ectocarpus silicolusos* (Morrissey *et al.*, 2015). The protein alignment of ISIP2, excluding the extended ISIP2 versions with domain duplications, provides information on signal peptides and transmembrane domains, demonstrating that most of the proteins possess a signal peptide and are expected to be either secreted, located on the cell surface, found in vesicles inside the cell, or targeted to the plastids (Figure 2.9). Implied by transcriptomic studies (Keeling *et al.*, 2014) and verified by PCR (data not shown), ISIP2 from *T. oceanica* has a signal peptide, but no transmembrane domain and is significantly longer than the other ISIP2 proteins analyzed in this study. In contrast, ISIP2a in *P. tricornutum* has a transmembrane domain and a signal peptide, and it is found at the cell surface as well as in vesicles (Morrissey *et al.*, 2015; McQuaid *et al.*, 2018). ISIP2a_*Pt* concentrates iron on the cell surface and functions as an independent iron uptake protein belonging to the transferrin family, which is supported by the presence of iron- and carbonate-binding sites equivalent to human transferrin (Morrissey *et al.*, 2015; McQuaid *et al.*, 2018). The protein alignment of ISIP2 sequences, along with two additional recently-described phytotransferrins from *P. tricornutum* (McQuaid *et al.*, 2018), resulted in 59% coverage for D_{48}, 71% for Y_{71}, only 27% for Y_{214}, and 45% for H_{294} as iron-binding sites. The carbonate binding site R_{183} was in 55% of the sequences present. Iron uptake studies of ISIP2a_*Pt* demonstrated a negative correlation to carbonate concentration similar to that seen for human transferrin (McQuaid *et al.*, 2018). One of the first algal transferrin-like proteins, Triplicate transferrin (Ttf_*Ds*), was described in *D. salina* and contained three transferrin domains, each similar to human transferrin. Although a previous alignment with human transferrin

showed the presence of several putative iron-binding sites (Fisher *et al.*, 1997), the Ttf_*Ds* aligned with only one tyrosine as iron-binding site in our alignment (Figure 2.9). The two transferrins from *D. salina* are the only proteins in this analysis with a predicted phosphonate-binding domain (PBD) and a transferrin domain (Figure 2.9). The FEA domain, similar to the C-terminal subdomain of ISIP2, is present in *C. reinhardtii*, *F. cylindrus, G. theta, Pelagophyceae,* and in both ISIP2 proteins from *Ochromonodaceae*.

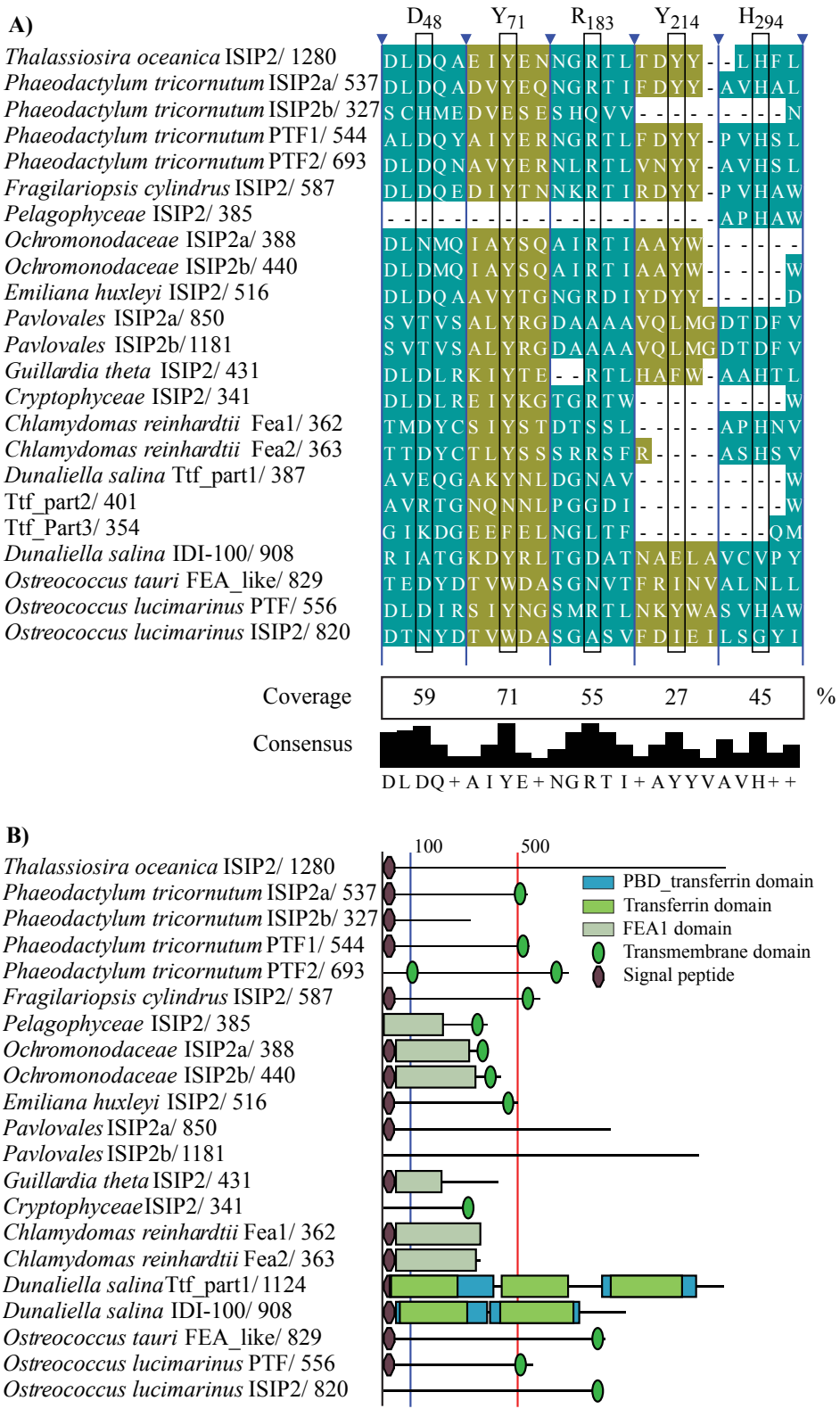
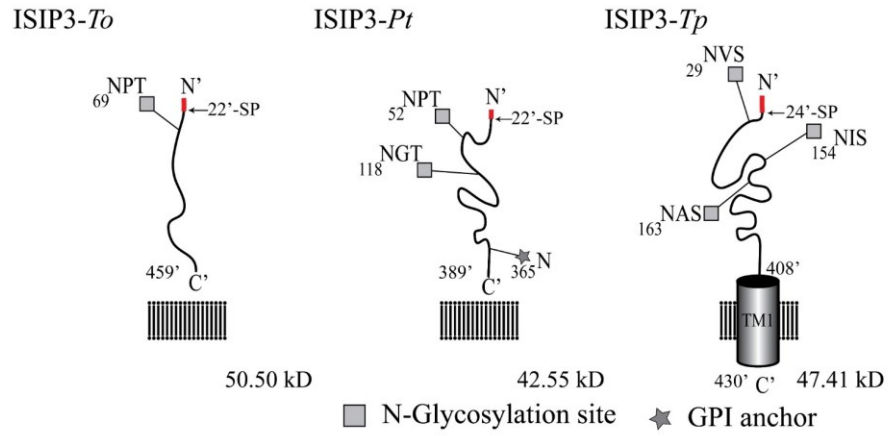


Figure 2.9: Overview of ISIP2 proteins concerning iron-binding sites and conserved domains.

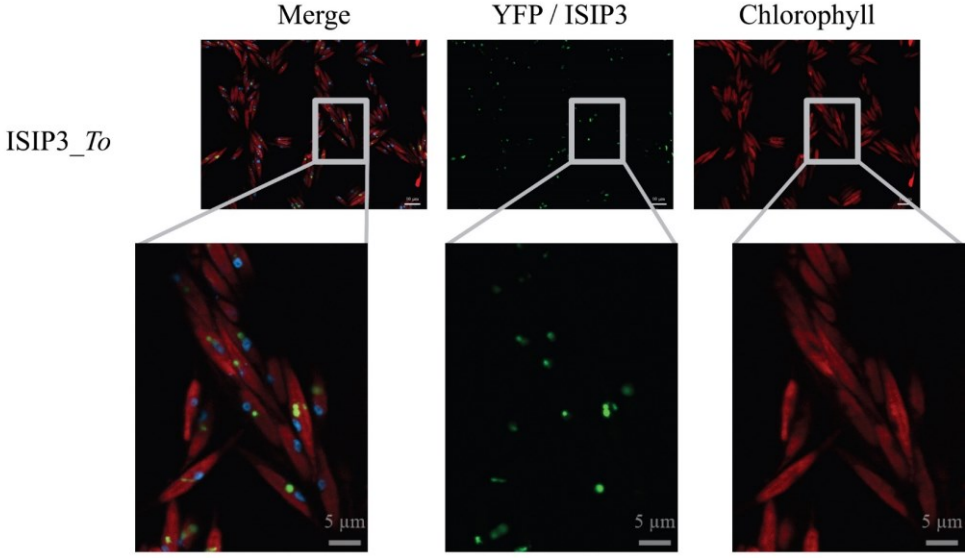
A) Full-length protein sequences were aligned (one-letter code) using MUSCLE in Mega7.0 (Kumar *et al.*, 2016). Conserved sites are based on McQuaid et al. (McQuaid *et al.*, 2018). D, Y, and H are iron-binding sites, whereas R is a CO_3^{2-} binding site. The positions are based on the *P. tricornutum* sequence. Consensus sequence and the coverage in % is shown below the alignment. B) Protein backbones are shown as black lines with an indication of conserved domains, presence of transmembrane domain (green circles), and signal peptides (purple hexagons). Transmembrane domains and signal peptides are predicted by TOPCONS (Shu *et al.*, 2015).

The third type of ISIPs is the group of ISIP3 proteins typified by a conserved domain of unknown function, DUF305, and an N-terminal signal peptide directing the protein to the secretory pathway (Figure 2.10). The conserved domain DUF305, present in all sequences, belongs to the ferritin superfamily and analysis via protein model finder, CPHmodels 3.2 server (Nielsen *et al.*, 2010), confirmed the presence of DUF305 in *G. theta*, *Cryptophyceae* and *Pelagophyceae* with high reliability even on a structural level (Figure 2.10). The presence of signal peptides and the lack of transmembrane domains suggests that most ISIP3 proteins are secreted or reside in intracellular vesicles. A heterologous transformation study in *P. tricornutum* showed that ISIP3_*To* confirms the location in distinct vesicles in close vicinity to chloroplasts (Figure 2.10). ISIP3 proteins from *E. huxleyi* and *G. theta* lack the signal peptide, and only three ISIP3 proteins have a transmembrane domain.

A)



B)



C)

	DUF305 CD	DUF305 CPH	TM	SP	Target
T. oceanica ISIP3	Y	N	0	Y	S
T. pseudonana ISIP3	Y	N	1	Y	S
P. tricornutum ISIP3	Y	N	0	Y	S
F. cylindrus ISIP3	Y	N	0	Y	S
Pelagophyceae ISIP3a	Y	N	0	Y	S
Pelagophyceae ISIP3b	Y	Y	0	Y	S
E. huxleyi ISIP3	Y	N	1	N	M
Pavlovales ISIP3	Y	N	0	Y	C
G. theta ISIP3	Y	Y	1	N	–
Cryptophyceae ISIP3	Y	Y	1	Y	S

Figure 2.10: Overview of structural information, cellular localization, and the presence of conserved domains in ISIP3 proteins.

A) Shown is the structure of three ISIP3 proteins found in diatoms. Indicated are N-glycosylations (Grey squares), GPI-anchor (asterix), transmembrane domains (TM), and signal peptides (SP). NetNGlyc 1.0 Server, TOPCONS, and SignalP 4.1 Server were used, respectively (Blom et al., 2004; Petersen et al., 2011; Chuang et al., 2012; Shu et al., 2015). B) Localization of ISIP3_*To*. ISIP3-YFP_*To* construct was transformed into *P. tricornutum*. ISIP3-YFP_*To* is shown in green while chloroplasts are red, and the DAPI stained nuclei blue. Scale bars are 5µm in the magnified image and 10 µm in others. C) Overview table showing results of TOPCONS (transmembrane domains and signal peptide), TargetP (subcellular localization), CD (conserved domain search, NCBI), and CPH-modeler -servers, respectively (Emanuelsson et al., 2000; Nielsen et al., 2010; Marchler-Bauer et al., 2017b). TM: Transmembrane domain; SP: Signal peptide; Target: S=secretory pathway, M=mitochondrion, C=chloroplast; DUF305 CD: the presence of domain DUF305 based on the conserved domain search on NCBI, DUF305 CPH: the presence of DUF305 in the structure of ISIP3 proven by CPH-modeler.

2.4 Discussion

The iron-homeostasis related proteins examined in this study revealed both structural and quantitative differences between the algal species analyzed here. These differences are unsurprising considering the timeline of evolutionary divergence going back 1 billion years for haptophytes and stramenopiles as the youngest separation (Yoon *et al.*, 2004). The investigation of the protein blocks involved in iron uptake by algae revealed four different iron acquisition strategies: siderophore-mediated uptake, reductive high-affinity uptake, phytotransferrin-mediated uptake of Fe^{3+}, and Fe^{2+} uptake via general divalent metal uptake proteins. The only species in our dataset that contains the protein building blocks for all four acquisition mechanisms is *F. cylindrus* (Figure 2.2), which possesses a large genome and can withstand very low iron concentration in the Southern Ocean (Mock *et al.*, 2017). Otherwise, the distribution of the different building blocks is not well defined and is unrelated to the taxonomic classification.

The general divalent metal uptake protein ZIP is present in all species analyzed here and shows high numbers compared to the other building blocks (Figure 2.2). The persistence and wide distribution of this protein family demonstrate the importance of ZIP proteins

for metal homeostasis. Considering that the divergence of green algae from the SAR supergroup and the haptophytes took place with the first endosymbiosis, 1.5 billion years ago (Yoon *et al.*, 2004), it is rather remarkable that both groups still possess this protein family in great abundance. The upregulation of ZIPs under low-iron conditions in field samples for *Thalassiosira* and *Chaetoceros* species was interpreted as a demonstration of their role in iron export from the vacuole indicating vacuolar iron storage in centric diatoms (Lampe *et al.*, 2018). This hypothesis is based on the intracellular localization of members of the ZIP family in *A. thaliana* (Milner *et al.*, 2012). ZIP proteins in *P. tricornutum* (Allen *et al.*, 2008; Smith *et al.*, 2016) and *O. tauri* (Lelandais *et al.*, 2016) are suggested to be permeases responsible for Fe^{2+} uptake, but involvement in iron export from vacuoles cannot be excluded here. Cell biology work on individual ZIP proteins is needed to gain knowledge about their specific functions. The conducted analysis does not show a negative correlation between ferritin and ZIP proteins, which would have been a good indication for the involvement of ZIP proteins in vacuolar iron storage for non-ferritin containing algae. The second protein building block involved in general divalent metal transport is the NRAMP protein family. Like ZIP proteins, NRAMP proteins may function in iron export from vacuolar storage in diatoms (Kustka & Allen, 2007; Lampe *et al.*, 2018) and *C. reinhardtii* (Urzica *et al.*, 2012), a hypothesis that is supported by the role of NRAMP proteins in iron export from vacuoles in *A. thaliana* (Bastow *et al.*, 2018). The NRAMP proteins analyzed show high similarity in terms of transmembrane domains and the location of the highly conserved iron-binding sites, making it difficult to differentiate them functionally.

Reductive high-affinity uptake systems, as described for *S. cerevisae* (Kosman, 2003), comprise another commonly observed acquisition mechanism (Figure 2.11). Although the wide distribution of the building blocks MCO and FRE among algal genomes indicates the potential for a wide distribution of the reductive high-affinity iron uptake system, the absence of an iron permease ortholog to FTR1 in most algal species suggests that the building blocks of this type of high-affinity system are not fully understood in marine algae, and indicates the use of alternative proteins for iron uptake as seen in *A. thaliana* (Ivanov & Bauer, 2017). Laboratory experiments with pure cultures showed that iron uptake in *T. oceanica* is dependent on copper, indicating a role for an MCO in iron uptake (Maldonado *et al.*, 2006). Iron reductase genes are upregulated in *T. oceanica* (Lommer *et al.*, 2012), *T. pseudonana* (Kustka & Allen, 2007), and *P. tricornutum* (Allen *et al.*, 2008), pointing towards a reductive high-affinity uptake system. Furthermore, iron reductase activity is increased under low iron conditions in *P. tricornutum* (Allen *et al.*, 2008) and *C. reinhardtii* (Lelandais *et al.*, 2016). All components involved in the classic reductive uptake system are present and upregulated under iron limitation in *C. reinhardtii* (Allen *et al.*, 2007). In contrast to *C. reinhardtii* and *T. pseudonan*a, *T. oceanica* and *P. tricornutum* do not possess an FTR1 like permease, which shows that there must be species-specific differences in this building block of the reductive iron uptake system. The reduction of iron is necessary to release iron from siderophores or other organic compounds. Therefore, reduction, whether intracellular or extracellular, is a crucial step in iron uptake in both algae and plants (Grillet & Schmidt, 2019). The building blocks of reductive high-affinity uptake overlap with the uptake or use of siderophore-bound iron based on the involvement of iron reductases in both strategies

(Maldonado & Price, 2001; Schröder *et al.*, 2003). The implementation of IRT1, belonging to the ZIP family, in the high-affinity uptake system in *A. thaliana* (Vert *et al.*, 2002) is another overlap of two iron acquisition strategies.

The overview of the building blocks of iron uptake systems (Figure 2.2) shows a broad distribution of iron uptake-associated proteins in this study. However, of the ISIP proteins, the phytotransferrin ISIP2 is the only one with a wide distribution, being present in all algal species studied except *P. multiseries*. The wide distribution of ISIP2 here agrees with previous findings in metatranscriptomic datasets (Marchetti *et al.*, 2012; Morrissey *et al.*, 2015). In general, ISIPs are up-regulated under low iron conditions, and ISIP1 and ISIP3 are solely expressed under iron limitation in *T. oceanica* (Lommer *et al.*, 2012). This upregulation was also verified in metatranscriptome analysis on environmental samples in low iron conditions (Marchetti *et al.*, 2012). ISIP proteins have been used as indicators for iron limitation in the ocean (Chappell *et al.*, 2015; Marchetti *et al.*, 2017) and were reported in areas of low-iron concentrations in the Tara Oceans (Caputi *et al.*, 2019).

ISIP2_*Pt* transports Fe^{3+} via endocytosis into the cell (McQuaid *et al*., 2018). The study conducted here shows substantial species-specific structural differences between the ISIP2 proteins in terms of conserved domains, the abundance of iron-binding sites, and the overall protein length. This group of proteins may fulfill more than just Fe^{3+} uptake as, for example, FEA proteins are shown to take up Fe^{2+} (Narayanan *et al.*, 2011) while being secreted in response to iron limitation (Allen *et al.*, 2007). Both characteristics stand in direct contrast to ISIP2a_*Pt,* but on the other hand, the induction of FEA1 upon high CO_2 concentrations (Allen *et al.*, 2007) and the abundance of all iron-binding sites

and the carbonate ion binding site in FEA2 show high similarities to ISIP2a_*Pt*. Nine members of the ISIP2 proteins reported here (Figure 2.9) have no transmembrane domain and would be either secreted similarly to FEA in *C. reinhardtii* or remain in vesicles inside the cell. Therefore, functional diversification of ISIP2 related proteins can be expected based on the described differences.

ISIP1_*Pt* is responsible for the internalization of hydroxamate-type siderophore-bound iron to the chloroplasts (Kazamia *et al.*, 2018). The theoretical random-encounters for one diatom cell are only 3.5 bacteria/day, assuming Brownian motion (Seymour *et al.*, 2017), suggesting that the siderophore uptake strategy would be more successful in a symbiotic relationship. The release of siderophores by bacteria has been suggested to increase the diffusion of iron and, therefore, to improve the overall availability of iron (Leventhal *et al.*, 2019). Some diatom–bacteria relationships have been described (Amin *et al.*, 2012; Behringer *et al.*, 2018; Sanchez *et al.*, 2018) with the area surrounding a phytoplankton cell, the phycosphere, attracting bacteria by the means of the host's exudates (Fouilland *et al.*, 2014; Seymour *et al.*, 2017). Both bacteria and diatoms possess and express attachment proteins that make this symbiotic relationship possible (Willis *et al.*, 2014; Guo *et al.*, 2017). A diatom-associated Marinobacter is an example of a symbiotic relationship where the host uses iron provided by the bacteria via vibrioferrin, a photolabile siderophore. (Amin *et al.*, 2012). While the use of iron by the diatom in this symbiotic relationship is dependent on the photoreduction of the siderophore, iron uptake via ISIP1 internalizes the siderophore directly (Kazamia *et al.*, 2018). Algae-associated bacterial communities are host species-specific, and more research is needed to demonstrate whether attached bacteria are capable of siderophore

production (Goecke *et al.*, 2013; Behringer *et al.*, 2018). Of note, the microbiome of *Fucus spiralis*, a brown macroalga, contains a high relative abundance of siderophore producing bacteria (Dogs *et al.*, 2017). Siderophore uptake in *P. tricornutum* is mediated by ISIP1, but only the internalization is inhibited in ISIP1 knock-down cells, not the binding of the siderophore on the cell surface, indicating the presence of a second protein that binds the siderophore (Kazamia *et al.*, 2018). *P. tricornutum*, *T. oceanica,* and *F. cylindrus* have ISIP1 and a ferrichrome-binding protein (FBP) as a possible siderophore binding protein (Allen *et al.*, 2008; Lommer *et al.*, 2012). Whether these proteins play a role in the transport of iron remains to be shown, but it is a surprise that the specificity of the siderophore differs between *P. tricornutum* and *T. oceanica*, with *T. oceanica* not able to use iron from hydroxamates (Kazamia *et al.*, 2018). This is in contrast to previous findings where *T. oceanica* used iron from a hydroxamate type siderophore (Maldonado & Price, 2001). The calculated Michaelis constant K_M of 5-7 nM for the siderophore uptake via ISIP1 in *P. tricornutum* indicates a fast and efficient uptake (Kazamia *et al.*, 2018). Similar studies on K_M values for other metal uptake proteins had half-saturation values of 76 nM for human hZIP4 (Antala *et al.*, 2015) and 28 nM for NRAMP1 in *A. thaliana* (Cailliatte *et al.*, 2010).

P130B is a homolog of ISIP1 (Figure 2.7) in *D. salina* and is part of a protein complex mediating iron uptake in combination with a transferrin and an MCO in an unknown way (Fig. 14) (Paz *et al.*, 2007). *T. oceanica* and *F. cylindrus* have all three components, whereas *P. tricornutum* is lacking the MCO. Although the function of ISIP1_*Pt* and ISIP2_*Pt* appears to be independent of one another (Kazamia *et al.*, 2018; McQuaid *et al.*, 2018), both proteins surround the chloroplast, and the uptake is described as

endocytic for both proteins (McQuaid *et al.*, 2018). Our localization study was limited to high iron conditions, and the localization of ISIP1_*To*, heterologously transformed in *P. tricornutum*, was on the cell surface similar to ISIP1_*Pt* (Figure 2.8). Overall, it is possible that ISIP1 and ISIP2 interact to fulfill iron transport, where ISIP1 acts as a non-specific enhancer of endocytosis-mediated iron transport under iron limitation.

Similar to ISIP3_*To*, the internalized siderophores in *P. tricornutum* are present in small vesicles close to the chloroplasts (Kazamia *et al.*, 2018), pointing towards a possible overlap or interaction of ISIP1-related uptake and ISIP3. ISIP3 could act as an iron storage protein based on the presence of DUF305, which belongs to the ferritin family. Additionally, both ISIP1 and ISIP3 show iron-mediated differential expression patterns greater than that of ISIP2, at least in *Pseudonitzschia granii* and *F. cylindrus* (Cohen *et al.*, 2018). Further studies are needed to investigate possible relationships between ISIP proteins, and new transformation systems for diatoms would be helpful to gain a better and more robust understanding of ISIP functions.

Iron storage is another critical intracellular mechanism based on the potential of free iron to cause cellular damage. Ferritin, found in all kingdoms, is widely used to store iron and has diverse functions in algae. The use of ferritin in the pennate diatom *Pseudo-nitzschia granii* has been postulated as an adaptation to survival in iron-limited regions, based on higher F_v/F_m and sustained growth under iron limitation compared to non-ferritin containing iron-limitation adapted diatoms such as *T. oceanica* (Marchetti *et al.*, 2009, 2017). The upregulation of ferritin under high-iron conditions in *P. granii* supports the role of ferritin in long-term storage (Cohen *et al.*, 2018). Furthermore, the maximal expression of ferritin in *O. tauri* at night suggests a diel regulation of ferritin, binding

excess iron in the night (Botebol *et al.*, 2015). Finally, *P. tricornutum* shows constant ferritin expression throughout the diel cycle indicating a general function in iron homeostasis (Smith *et al.*, 2016). Although first identified in pennate diatoms, ferritin has now also been found in centric diatoms, supporting multiple functions for ferritin in diatoms (Groussman *et al.*, 2015; Cohen *et al.*, 2018). Vacuolar storage of iron is another important iron storage mechanism which, in *A. thaliana,* involves the vacuolar iron transporter 1 (VIT1) for import (Kim *et al.*, 2006) and NRAMP proteins for iron export (Connorton *et al.*, 2017; Bastow *et al.*, 2018). Vacuolar storage may play an important role in algae that do not contain ferritin. Visualization of iron in *T. pseudonana* and *T. weissflogii* points towards vacuolar storage (Nuester *et al.*, 2012), and the upregulation of NRAMP proteins and ZIP proteins in *Thalassiosira* and *Chaetoceros* species under low iron conditions is suggested as a putative vacuolar iron export function (Lampe *et al.*, 2018).

The diversity in algal iron uptake strategies reflects the complex chemistry of iron in the ocean. Except for ISIP1, the different proteins involved in iron homeostasis all originated in bacteria (Debeljak *et al.*, 2019), and existing iron acquisition mechanisms present in algae appear to use different building blocks in different combinations. Furthermore, algae have adapted ISIPs to survive in extremely low iron concentrations. Figure 2.11 shows an overview of high-affinity uptake proteins in *A. thaliana*, *D. salina,* and *S. cerevisiae*. The different proteins found in these three species are also present in diatoms, with the addition of ISIPs indicating a high diversity of iron uptake proteins in diatoms, which might be one reason for their success over the last 100 million years. Research efforts must concentrate on specific proteins to solve these puzzles, and iron uptake

strategies might be organism-specific where the above-described building blocks overlap, and individual proteins, such as FRE, are involved in different uptake strategies.

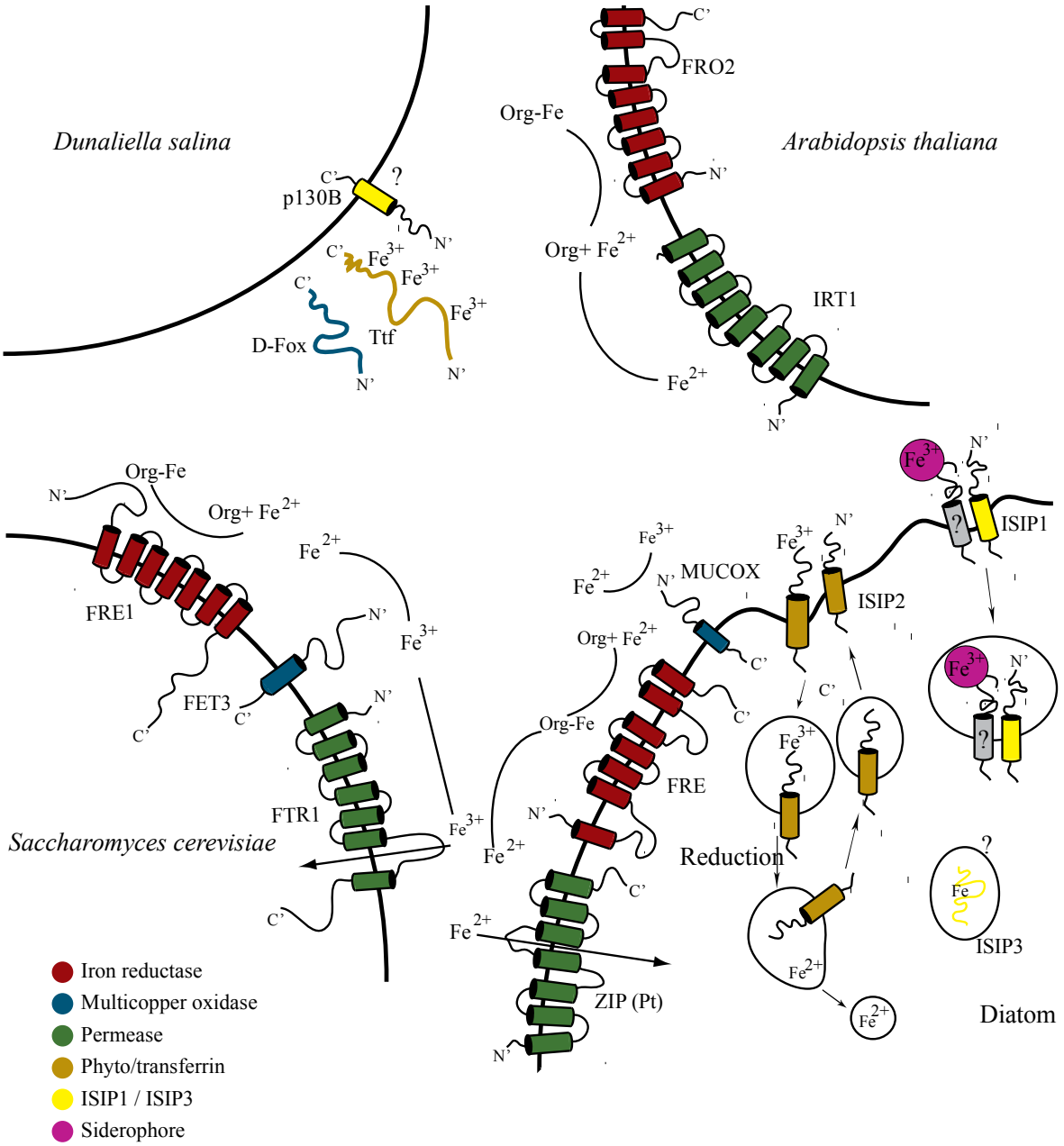


Figure 2.11: Iron uptake overview.

The overview represents four different iron uptake strategies in the form of high-affinity uptake strategies present in *D. salina*, *A. thaliana*, and *S. cerevisiae*. It also shows an overview of proteins involved in iron uptake in diatoms. TOPCONS was used for topology information (Tsirigos *et al.*, 2015). Topology of FRO2 is according to uniprot based on

published data. Colors are indicated and represent different types of proteins. The following abbreviations are use: Org: organic-ligand; FRO2: Iron reductase/oxidase; IRT1: Iron regulated transporter 1; ISIP: Iron starvation-induced protein; FRE: Iron reductase; ZIP: Zinc-regulated iron-regulated protein; FTR: Iron transporter (permease); Ttf: Triplicate transferrin; D-FOX: *Dunaliella*-Iron multicopper oxidase; MUCOX: Multicopper oxidase.

CHAPTER 3 SHORT-TERM RESPONSE TO IRON RESUPPLY IN IRON-LIMITED *THALASSIOSIRA OCEANICA*, AN OPEN OCEAN DIATOM, LEADS TO RAPID BUT DIFFERENTIAL DECAY OF IRON-RESPONSIVE TRANSCRIPTS

Joerg Behnke, Yun Chai, Hong Gu, Julie LaRoche

Contribution of authors:

Joerg Behnke:	Planning and conducting the experiment, sample preparation, data analysis, and drafting of the manuscript
Hong Gu, Yun Chai:	Analysis of transcript decay curves
Julie LaRoche:	Contributing to experimental design, planning and discussing the manuscript

This manuscript will be submitted to PLOS One.

3.0 Abstract

The iron hypothesis was proposed in 1989, and numerous studies have been conducted to characterize different aspects of the effect of iron on algae and how algae cope with fluctuating iron concentrations. Diatoms are one the most successful algae in low-iron ocean surface waters, and their dominance in iron-induced algae blooms is of great interest. Several studies have focused on transcriptomic differences in diatoms comparing low- and high-iron conditions. The study presented here analyzed the short-term temporal response following iron addition to an iron-starved open ocean diatom, *Thalassiosira oceanica*. The analysis included 54 transcripts and covered several cellular processes such as photosynthesis, N-linked glycosylation, iron transport, and transcription factors using the NanoString platform to generate a high-resolution time series. Nine transcripts were iron-responsive and showed an immediate response to the addition of iron. The downregulation of genes involved in iron uptake is the first and fastest response, followed by the expression of genes for iron-containing enzymes that have been replaced under iron-constraint. Simultaneously, transcripts for the replacing non-iron containing enzymes are decreasing. In a parallel experiment, a transcription inhibitor, actinomycin D, was used to explore underlying mechanisms of the decay of the iron-responsive transcripts, aiming to distinguish whether specific mRNAs are stabilized. The analysis of the decay curves revealed half-lives of 5-20 min for all transcripts except Mn-superoxide dismutase (Mn_SOD), plastocyanin (PETE), and ferredoxin (PETF). Iron starvation-induced protein 1 (ISIP1), fructose bisphosphate aldolase 4 (FBA4), flavodoxin 1 (FLDA1), and cellular repressor of EA1 regulated genes x2 (CREGx2) had significant differences in transcript decay between their iron-recovery samples and actD treated

samples. These differences suggest that regulatory mechanisms influenced their mRNA stability. The data gives insight into the short-term temporal response of *T. oceanica* on transcriptional and physiological levels. The Tara Oceans database showed a global distribution of the analyzed transcripts and the presence of iron-responsive genes in low iron areas. Overall, the data of this study is an essential step towards the understanding of diatom cell biology in the context of iron fertilization responses and will improve the modeling efforts of iron responses following sudden changes in iron concentrations.

3.1 Introduction

Iron is scarce in the ocean's surface water, and its complex chemistry results in multiple forms of iron with minor bioavailability. Iron precipitates in the oxidative contemporary ocean, but over 99% of the iron is bound to organic chelators (Gledhill & Buck, 2012; Lohan *et al.*, 2015; Tagliabue *et al.*, 2017). Since the early 1990s, it has been known that iron promotes algae growth (Martin, 1990), and it is now well established that 30-40% of the ocean surface waters are limited by iron (Westberry *et al.*, 2009). Several iron fertilization experiments have been conducted, resulting in diatom-dominated algal blooms (Landry *et al.*, 2000; de Baar *et al.*, 2005; Boyd *et al.*, 2007; Pollard *et al.*, 2009; Trick *et al.*, 2010). Diatoms are cosmopolitan unicellular algae that build an essential part of the global marine food-web and the carbon export in the ocean (Smetacek, 2018). Laboratory investigations of low-iron adapted diatoms revealed several adaptations including restructuring of the photosynthetic apparatus, replacing iron-containing proteins with alternative proteins, expressing a high-affinity iron uptake system, and modifying cellular parameters resulting in higher surface/volume ratios (La Roche *et al.*, 1996; Strzepek & Harrison, 2004; Peers & Price, 2006; Lommer *et al.*, 2012). Whereas

specific adaptations in open ocean diatoms such as the use of plastocyanin (PETE) or a decrease of the iron-bearing photosystem I (PSI) are permanent, most acclimations are temporary and can be reversed upon iron resupply (La Roche *et al.*, 1996; Strzepek & Harrison, 2004; Peers *et al.*, 2005). The switch from high- to low-iron conditions results in fundamental changes within the cell and can be detected in most metabolic processes (Lelandais *et al.*, 2016; Smith *et al.*, 2016). Iron limitation leads to an overall downregulation of photosynthetic genes (Allen *et al.*, 2008; Lommer *et al.*, 2012). In *T. oceanica,* the reduction of the number of chloroplasts, upregulation of endocytosis, and the upregulation of organic material degrading enzymes are discussed as a metabolic shift towards a mixotrophic lifestyle (Lommer *et al.*, 2012). Through these adaptations, open ocean diatoms can decrease their Fe:C ratio substantially (Marchetti *et al.*, 2006). Moreover, diatoms adapted well to the different forms of iron and evolved a variety of uptake strategies including a reductive iron uptake system, following the principles of a high affinity uptake system found in yeast (Maldonado & Price, 2001; Shaked *et al.*, 2005; Maldonado *et al.*, 2006), iron uptake via a phytotransferrin, the iron starvation-induced protein 2 (ISIP2) (McQuaid *et al.*, 2018), and an endocytosis-mediated siderophore uptake system via ISIP1 (Kazamia *et al.*, 2018).

Iron, as an essential trace metal for algae, can lead to radical oxygen production when intracellular concentrations are too high. The response to sudden changes in iron concentration surrounding the cell must, therefore, be fast to avoid excess uptake of iron. Temporal analysis of transcript counts responding to iron induction demonstrated a significant downregulation for flavodoxin (FLDA) and ISIP3 in *Thalassiosira oceanica* (Chappell *et al.*, 2015) and ferric reductases (FRE) in *Thalassiosira pseudonana* (Kustka

& Allen, 2007), but an exact temporal resolution of these changes remained unknown. The study here aimed to give insight into the short-term response to iron addition on a transcriptional and physiological level in *T. oceanica,* which is a globally distributed open ocean diatom adapted to chronic iron limitations. Fifty-four probes for mRNA transcripts that were previously described as differentially expressed under high-/low-iron concentrations were chosen for this analysis (Lommer *et al.*, 2012), and the transcript abundance was recorded throughout a 22 h period. The results of this study showed the downregulation of transcripts for iron uptake proteins as the first response, followed by the switch of transcripts for proteins that were replaced under iron limitation. Further, the study included the analysis of mRNA transcript stability and the tracking of physiological parameters to embed our data on transcript dynamics in a greater context.

Keywords: Diatom, iron hypothesis, targeted transcriptomic, iron response, F_v/F_m, mRNA half-life, ISIPs

3.2 Material and Methods

3.2.1 Artificial seawater and culturing

Axenic *Thalassiosira oceanica* (CCMP1005) was grown in ASW f/2 following a 14/10 h light/dark cycle at 22°C in trace metal clean polycarbonate (PC) bottles. ASW was prepared after Goldman et al. (Goldman *et al.*, 1978), which included the growth on nitrate (NO_3) as the only nitrogen source. The Aquil metal mix (Price *et al.*, 1988) was used with 100 µM EDTA final concentration and no addition of Ni^{2+}. Trace metal clean techniques were used for the preparation of high- and low-iron water, as well as all equipment used in this study. Acidified (pH=2) 10 mM solution of $FeCl_3$ without EDTA

was used for the addition of 10 µM final concentration of $FeCl_3$. Any modification or sampling was done in a trace metal clean environment. ASW was cleaned by gravity flow-through using a Sigma Fritted Luer Lock Chromatography Column (2.5cm x 23cm; Sigma-Aldrich Corporation, St. Louis, Missouri, USA) with 100 mg Chelex®100 (BioRad, Hercules, CA, USA) in a trace metal clean working space with a speed of 1-2 drops per second. All bottles were cleaned with Citronox (Alconox, White Plains, NY, USA) for 24 h, rinsed with MilliQ grade water, incubated in 3% hydrochloric acid (HCL) for 7 days and finally rinsed with MilliQ grade water before microwave sterilized. All materials used for trace metal clean water were rinsed with acidified MilliQ (pH=3).

3.2.2 Experimental design

Thalassiosira oceanica growth was kept in the exponential phase, and sampling was done in low to mid-exponential phase. Two separate experiments were performed, with the first experiment (long-term (LT)) covering 22 h, and the second experiment (short-term (ST)) covering a 6 h period. The treatments in the LT experiment included high-iron, low-iron, and iron-recovery. High-iron cultures were grown with an iron concentration of 10 µM, whereas no iron was added to the low iron cultures. The iron-recovery samples received 10 µM $FeCL_3$ after the initial measurement (T=0). The treatments analyzed in the ST experiment included high-iron, low-iron, and iron-recovery, as stated. Additionally, samples were treated with actinomycin D and iron (actd-Fe), only actinomycin D (actD), DMSO with iron (DMSO-Fe), and DMSO without iron (DMSO). The actD and DMSO treated samples were prepared in 500 ml PC bottles with a volume of 300 ml for the actD treatments and 200 ml for the DMSO treatments. The initial measurement was taken one hour after the start of the light period, and iron-addition

followed immediately after the initial sampling was completed. The following timepoints are based on the time-distance to the addition of iron. The polycarbonate bottles for the sampling of high- and low-iron samples were 4 L in the LT and 2 L in the ST experiment. For the sampling of the iron-recovery samples in the LT experiment two 4 L bottles were combined before the initial sampling. The sampling bottles in the LT experiment were inoculated with 2000 cells/ml, while the inoculum in the ST experiment was increased to 75000 cells/ml. The sampling times are shown in Figure 3.1.

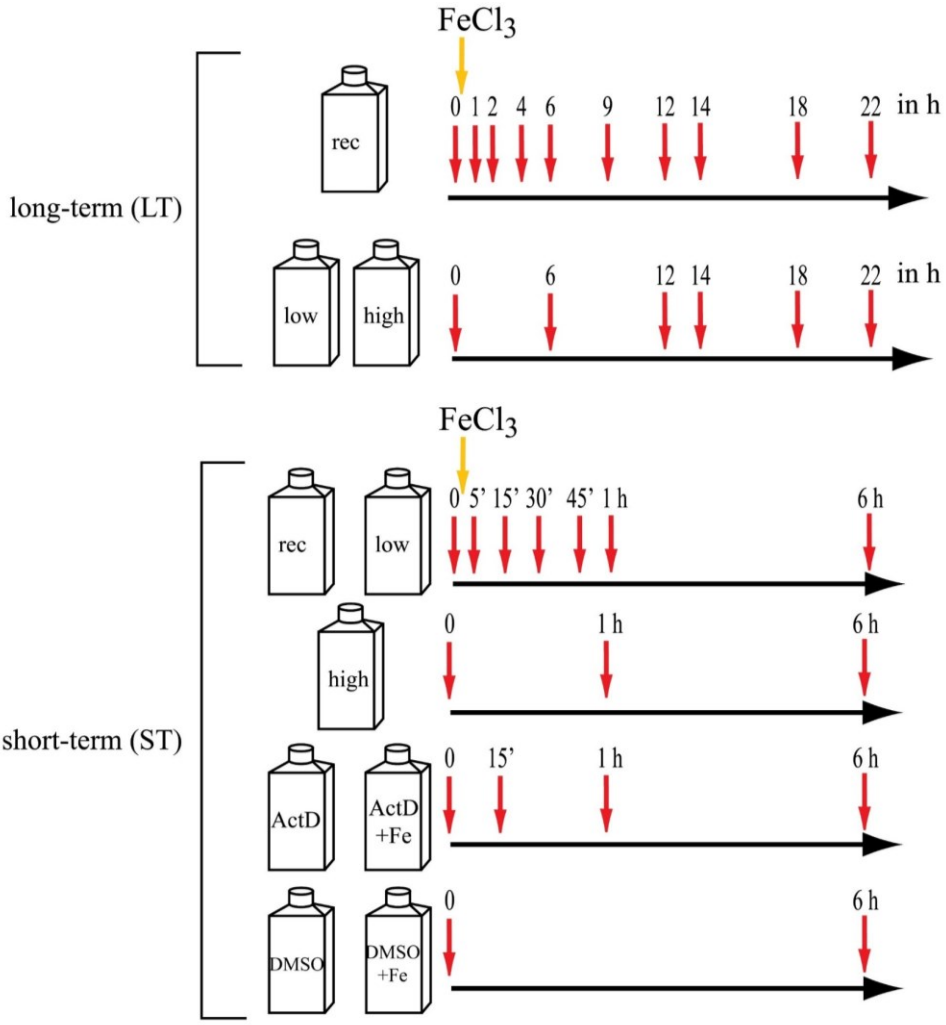


Figure 3.1: Experimental overview with sampling times.

Shown are the two different experiments that were conducted- the long-term (LT) experiment and the short-term (ST) experiment. The bottles indicate the different treatments that were analyzed in each experiment. Besides, the bottles are the sampling times in hours (h). The addition of iron is indicated with a yellow arrow. The iron-recovery samples (rec) were the only samples that received the addition of $FeCl_3$ as indicated.

In all experiments, the 1 h timepoint describes the timepoint at 1 h following the addition of $FeCl_3$. The LT experiment included triplicate experiments of each treatment (high, low, and recovery). The ST run included triplicates for each treatment (high, low, recovery, DMSO, DMSO-recovery) but only duplicate experiments for actD and actD-Fe. All replicates were taken from independent bottles.

3.2.3 Actinomycin D and DMSO treatment

Actinomycin D (actD) (Invitrogen, Carlsbad, CA, USA) treatments were done with a final concentration of 10 µg/ml actD in DMSO. DMSO concentration in the media was 0.1% final. A pre-test was run on triplicate cultures (data not shown) to check for the effect of actD. F_v/F_m measurements were used as a proxy for possible iron impurities and the effectiveness of actD. A parallel experiment with only DMSO (0.1% final concentration) was performed as a control in the pre-test as well as in the ST experiment itself.

3.2.4 Standard measurements and sampling of the culture

3.2.4.1 Flow cytometry

The Accuri C6 (Becton, Dickinson and Company (BD), Franklin Lakes, New Jersey, USA) was used to measure cell counts and to gain information about granularity using side-scatter (SSC), cell size using forward scatter (FSC), and chlorophyll content using fluorescence measurement (FL3, > 670 nm LP). The FL3 channel was used to gate the *T.*

oceanica population. All measurements were done on a 96well plate using 200 µl of culture. Samples were measured as duplicates in the ST experiment and as triplicates in the LT experiment.

3.2.4.2 Chlorophyll measurements

Ten ml culture was filtered onto glass fiber filters (GFF) and stored in scintillation vials at -20℃ until further processing. Chlorophyll was extracted in 90% acetone at -20℃ for 24 h. The measurement and the calculation was done using the Welschmeyer method (Welschmeyer, 1994) with a 10-005R fluorometer (Turner Designs, San Jose, CA, USA).

3.2.4.3 POC/PON measurements

Twenty ml of culture was used for the analysis of particulate organic carbon (POC) and particulate organic nitrogen (PON). The samples were filtered onto combusted GFF filters, which were dried for 24 h at 52°C following filtering. GFF filters were packed in tin cups and combusted at 1000C° to measure total organic C and N on an Elementar VarioMicro Cube Elemental Analyzer (Elementar, Langenselbold, Germany).

3.2.4.4 DUAL-PAM-100 measurements

A Waltz DUAL-PAM-100 (Heinz Walz GmbH, Effeltrich, Germany) was used to measure photophysiological parameters, including normalized variable fluorescence (F_v/F_m) for LT and ST experiment. Electron transfer rate (ETR), non-photochemical quenching (NPQ), and non-regulated heat dissipation (NO) were measured during the ST experiment using lightcurve measurements with 11 steps saturation pulse increasing in light intensity (0; 15;22;31;62;104;135;225;348;540;834 µE). All measurements were done following 10 min dark incubation with default settings for saturation pulse and measuring light (SP=10.000 µE for 600 ms, ML=128 µE) in the LT experiment.

Optimization of the light curve for F_v/F_m measurements led to an adjustment for the ST experiment resulting in a saturation pulse of level 1 (1000 µE for 600 ms) with a measuring light of level 10 (128 µE).

3.2.4.5 RNA extraction

RNA was extracted from a 150-200 ml sample with the Qiagen Plant RNeasy kit (Qiagen, Inc., Valencia, CA, USA) following manufactures protocol using 450 µl RLC buffer, 5 min incubation at 56℃ without sonification. LT samples were first filtered onto 2 µm PC filters, washed off, pelleted by centrifugation for 3 min at 5000 g, and flash-frozen in liquid nitrogen. ST experiment samples were filtered on 2 µm PC filters, immediately packed into 1.5 ml tubes and flash-frozen in liquid nitrogen. On-column DNA digestion using RNase free DNase by Qiagen (Qiagen, Inc., Valencia, CA, USA) was done to ensure the complete removal of DNA. The RNA was quantified with a Nanodrop (Thermo Fisher Scientific, Waltham, Massachusetts, USA) and stored at -80℃.

3.2.5 Transcript analysis using NanoString

The probes for the NanoString (NanoString Technologies ®, Seattle, Washington, USA) analysis were designed based on the available genome of *T. oceanica* (Lommer *et al.*, 2012)(Accession nr. AGNL01000000) and checked for cross binding activity (Supplemental Figure 2). Forty-four mRNAs were targeted in the LT experiment (Supplemental Table), and 24 (14 overlappings with LT experiment) in the ST experiment (Supplemental Table 3 and Supplemental Table 4).

Forty-eight samples of the LT experiment were analyzed for 44 mRNAs using 100 ng RNA of each sample in a first run (LT-Run) (Supplemental Figure 1). The target mRNA

sequence (100 bp) binds to two different probes within the NanoString assay. The reporter probe includes the fluorescently labeled part, and the capture probe includes biotin that binds to streptavidin beads throughout the automated NanoString procedure. The hybridization time was set at 20 h for the probes to bind the targeted transcripts in our samples. Otherwise, "nCounter® XT assay" from NanoString was followed.

In the second run, samples from LT and ST experiments were analyzed combined in two 96well plates (96well-Run) (Supplemental Table 5 and Supplemental Table 6). The "nCounter® PlexSet™ Reagents for Gene Expression User Manual" was followed. Each column on the 96well plate was pooled, after the 20 h hybridization incubation, to create one sample that is processed and analyzed. Two separate probes were used, Probe A and Probe B (Supplemental Table 3 and Supplemental Table 4). Probe A has a target-specific sequence, and a row specific fluorescently labeled reporter tag and Probe B has the universal capture tag. Based on high counts in the first LT-run, FLDA1, FBA3, FBA4, and, as high counts were expected, PETE were attenuated by the addition of inactive probes, aiming a 90% reduction of probe counts. A titration run was performed before the 96well plate run on a low-iron, high-iron, and an iron-recovery sample to ensure sufficient counts of targets. The titration run was used to calculate loading amounts of the different treatments, as well as the actual attenuation factor, which was used for the final analysis of the 96well plates (see below). Based on the titration run, 90 ng of RNA was used for high-iron samples (including iron-recovery samples that were sampled later than 30 min after the addition of iron), 70 ng for low-iron samples, and 80 ng for iron-recovery samples. All low iron samples and iron-recovery samples within 30 min after iron addition were attenuated for the above-mentioned genes.

3.2.6 Data analysis

3.2.6.1 The basic calculation of transcript counts

The absolute counts generated by the NanoString analyzer were downloaded and first processed in the NSolver software. Positive control count normalization was done by using the geomean of the Top 3 positive control probes that are included in the probesets provided by NanoString. Further, the samples were normalized towards the arithmetic mean over all samples of one house-keeping gene, the nuclear-import-exporter (THAOC_05312), which was done using Microsoft Excel.

The 96well-Run included a reference lane which was run in the first lane on both plates. The reference lane is used for in-plate probe calibration and contained the same sample in each well (Lane1 A-H). Here, a mix of one high-iron and one low-iron sample (1:1) was used to achieve sufficient counts for each target. The calibration of RNA amount loading differences and the correction for the attenuation was done using Excel. The housekeeping gene normalization for the actD samples was done independently, based on the degradation of all transcripts. Therefore, the arithmetic mean of the housekeeping gene counts of each timepoint was used to normalize each timepoint separately.

3.2.6.2 Decay curve analysis and validation

The data did not change linearly, and a broken-line model was used to calculate the decay curve. The slope of the curve before it became stable was used for the half-life calculations (changepoint) (Siegmund & Zhang, 1994).

Decay curve:

$$X_{np} = \beta_0 \times treatment_{np} + \beta_1 \times treatment_{np} \times \min(time, t_0)$$
$$+ \beta_2 \times treatment \times (time - t_0) + e$$
$$e \sim Normal(0, \sum^2)$$

X is the log-transformed data, and t_0 is the change point. t_0 is fixed for each treatment (calculation is shown below). β_0 is the unknown intercept for each treatment. β_1 is the unknown coefficient of the interaction between treatments before the change point, and β_2 is the unknown coefficients after the change point. Therefore, β_1 and β_2 are the decay rate vectors for the different treatments. The calculation of t_0 for each treatment was done using "lm.br" in R. The lm.br function fits the following model to each treatment's data:

$$y = alpha + B \times \min(time - t_0, 0) + Bp \times \max(time - t_0, 0) + e$$
$$e \sim Normal(0, \sigma^2)$$

Y is the log-transformed data for one treatment, while the parameters 'alpha', 'B', and 'Bp' are unknown.

Comparison of decay rates between treatments:

After β_1 was calculated, a likelihood ratio test was applied under the following hypothesis:

H_0 : The decay rates are the same between the two treatments.

(eg. β_1_actd = β_1_actd-Fe)

H_a : The decay rates are different between the two treatments.

(eg. β_1_actd ≠ β_1_actd-Fe)

The P-value of the test showed how strong the evidence was to reject that two treatments share the same decay rates. The decay constant was used to calculate the half-life by $t_{1/2}$= ln2/k with k being the decay constant.

3.2.7 Database searches and *in-silico* analysis

3.2.7.1 Visualization of transcript counts

The normalized transcript counts were analyzed and visualized in R (R Development Core Team, 2015) using ggplot2 (Wickham *et al.*, 2018). The visualization included the plotting of transcript counts over time to visualize changes. Fold changes between high-iron and low-iron samples were calculated, and two-tail Student's T-test (with unequal variance) was used for significance tests ($p < 0.05$).

3.2.7.2 Correlation matrix analysis

A correlation matrix analysis based on Spearman correlation was generated using log2 values of all transcript counts, excluding samples that were treated with actinomycin D, using the R-package ComplexHeatmap (Gu *et al.*, 2016), belonging to the Bioconductor package (Gentleman *et al.*, 2004). The final curation of the heat map was done with Adobe Illustrator.

3.2.7.3 In silico analysis

The *in-silico* analysis of putative transcription factors was performed using the Conserved Domain search tool from NCBI on all proteins from *T. oceanica* (Marchler-Bauer *et al.*, 2017). An NCBI blast analysis was used to confirm putative transcription factors. In general, *in-silico* analysis of protein localization was done with TargetP and

ChloroP (Emanuelsson *et al.*, 1999, 2000). The NetNGlyc 1.0 Server was used for the prediction of N-glycosylation sites in CREGX2 (Chuang *et al.*, 2012) and the SignalP 4.0 Server (Petersen *et al.*, 2011) was used for signal peptide prediction.

3.2.7.4 Tara Oceans database search

The Tara Oceans database (Villar *et al.*, 2018) was used to generate plots of the global distribution of all targeted transcripts used in the ST and LT experiment. Different e-values were used for each gene to find sequences with the highest similarity to *T. oceanica*. The e-values used for the search were between 10^{-70} and 10^{-250}. All targets were screened in the metagenome and the metatranscriptome dataset. The 0.8-5 µm and the 5-20 µm fraction from the surface and deep-chlorophyll max were analyzed.

3.3 Results

The effects of iron supply following prolonged iron-limitation in the open ocean diatom *T. oceanica* were monitored at the transcriptional and physiological levels. The NanoString platform was used to track a total of 54 transcripts in two different experiments. The first experiment covered 22 h (LT-experiment), and a second experiment was conducted over a 6 h period (ST-experiment). In both experiments, the initial sampling (timepoint=0) was done one hour after the light period started, and the addition of iron followed after the initial sampling. The LT experiment revealed adjustments of transcript counts in under one hour for some genes. Therefore, a temporal high-resolution time-course of the first hour was conducted on 24 transcripts in the ST experiment. Both experiments included high-iron samples, low-iron samples, and iron-recovery samples. Iron-recovery samples were grown equivalent to low-iron samples and received 10 µM $FeCl_3$ after the initial sampling. The analysis of transcript half-lives was

only conducted in the ST experiment using the transcription inhibitor, actinomycin D (actD). The use of dimethyl sulfoxide (DMSO) as a solvent for actD led to the analysis of DMSO treated low-iron and iron-recovery samples to reassure that the effects of actD are independent of DMSO.

3.3.1 Physiological response

Differences between low-iron and high-iron conditions were detected at the transcriptional and physiological levels. All physiological parameters shown in Figure 3.2 showed significant differences between high-iron and low-iron samples. Fluorescence (FL3), side scatter (SSC) and forward scatter (FSC) as a proxy for chlorophyll, cell granularity, and cell size, respectively (Figure 3.2), were tracked and are shown. High-iron cells revealed a stronger chlorophyll signal, higher granularity, and they were significantly larger. These parameters, measured with a flow cytometer, were further verified with POC/PON measurements and chlorophyll measurements. All three parameters showed significantly higher values for high-iron samples resulting in more chlorophyll/cell and higher carbon and nitrogen/cell. High-iron cells also grew faster and had significantly elevated F_v/F_m measurements (Figure 3.2). Only the C:N ratio remained equal between high-iron and low-iron samples.

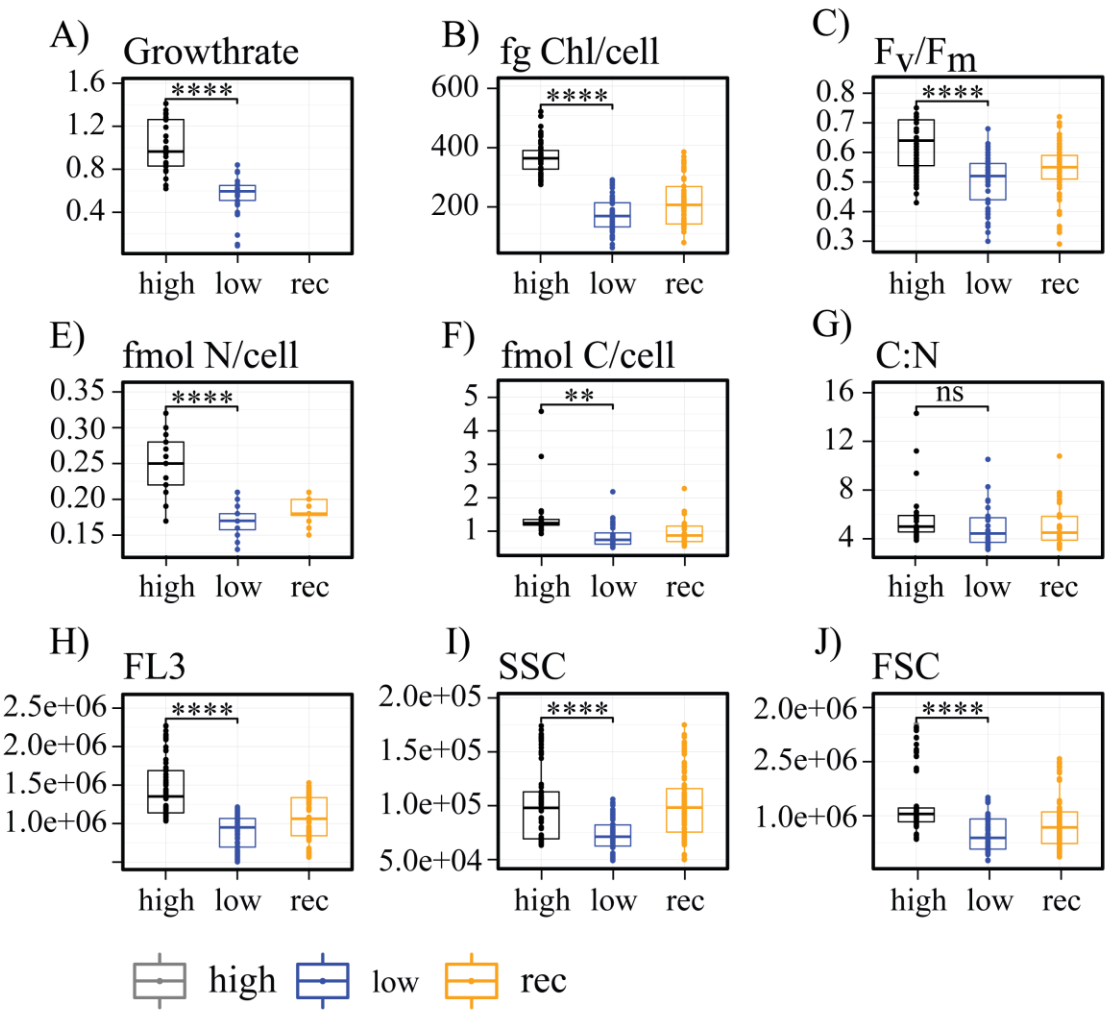


Figure 3.2: Statistical analysis of physiological parameters, comparing low-iron, high-iron, and iron-recovery samples.

Shown are parameters to characterize the state of the cell under low-iron, high-iron, and iron-recovery conditions. High-iron samples are black, low-iron samples are blue, and iron-recovery samples are shown in orange. Results from ST and the LT experiment are combined with every dot representing one sample, and a 2tail student's T-test was used for statistical analysis. A) The growth rate in μ (growth rate/day). The calculation of the growth rate is based on the final bottles in the LT experiment and on five consecutive bottles prior the final bottle in the ST experiment. B) Chlorophyll content per cell in fg/cell is based on chlorophyll measurements using a fluorometer. C) F_v/F_m measurements of all timepoints measured during LT and ST experiment D) Particulate organic nitrogen (PON) measurements were used to calculate N/cell in fmol/cell. E) Particulate organic carbon (POC) measurements were used to calculate C/cell in fmol/cell. F) C:N ratio for each measuring point of POC/PON. H) A flow cytometer was used to receive FL3 (LP > 670 nm) measurements as a proxy for chlorophyll. I) A flow cytometer was used for measurements of side scatter (SSC) as proxy for cell granularity. J) Flow cytometer

measurements were also used to gain information about forward scatter as a proxy for cell size. The following abbreviations were used: fluorescence (FL), side scatter (SSC), forward scatter (FSC), variable fluorescence (F_v), maximal fluorescence (F_m), nitrogen (N), carbon (C), chlorophyll (Chl). FL3, SSC, FSC, and F_v/F_m are in relative units. Chl, N, and C are in fg per cell.

The electron transport rate (ETR) had significantly higher values in the high-iron samples compared to low iron samples, and iron-recovery samples approached high-iron values throughout the 6 h of the ST experiment (Figure 3.3). The same trend was observed for photosystem efficiency Y(II) with significantly higher values in high-iron samples. Photochemical quenching Y(NPQ) showed higher values in low-iron samples, while non-regulated energy dissipation Y(NO) showed the opposite trend (Figure 3.3). The change of photo physiological parameters over time showed a significant change in F_v/F_m values. While the F_v/F_m values in the iron-recovery samples showed a significant upregulation after 6 h when compared to low-iron samples, the light curves of the iron-recovery samples also showed a clear distinction from the low-iron light curve after 6 h (Figure 3.3).

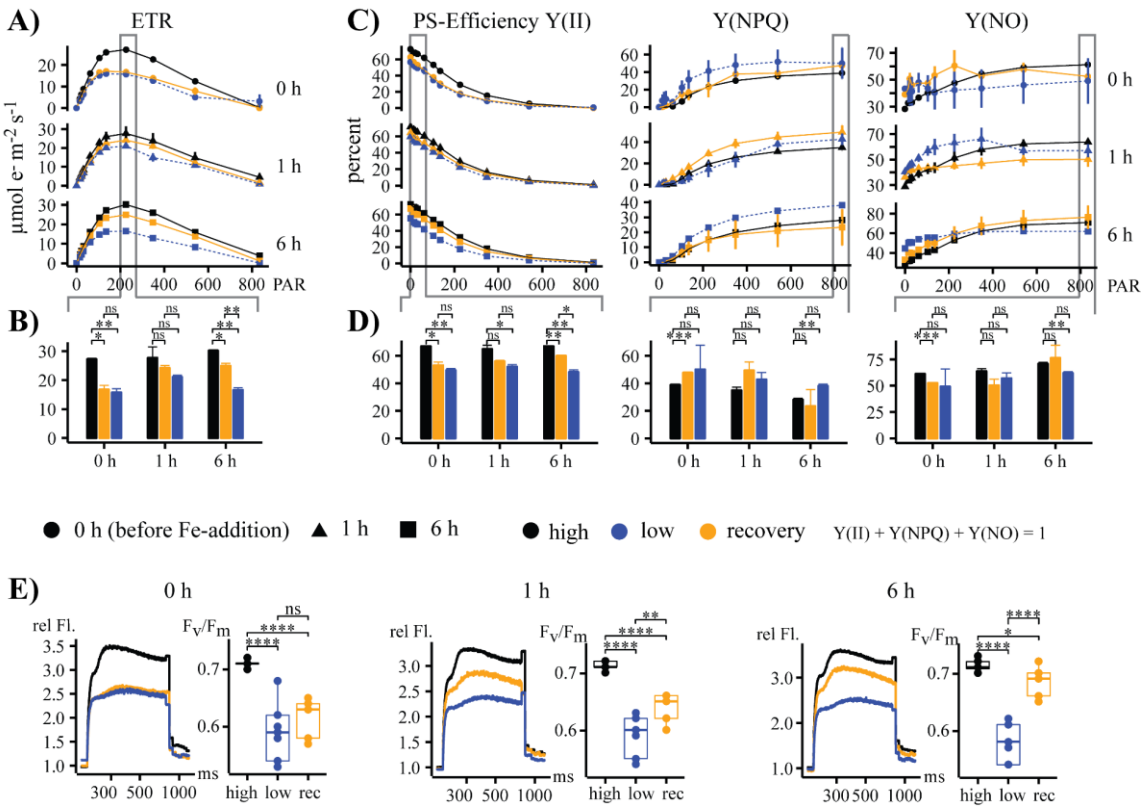


Figure 3.3: Changes in photo-physiological parameters over a period of 6 h during the short-term (ST) experiment.

Measuring points of high-iron samples are shown in black, low-iron samples are in blue, and iron-recovery samples are in orange. 0 h, 1 h, and 6 h sampling are separated and represented as triangles, squares, and circles, respectively. A) The electron transfer rate (ETR) in µmol $e^{-}m^{-2}s^{-1}$ is shown. B) Measurements of the ETR_{max} is shown as a bar plot with student's t-test for statistical differences. C) Photosystem efficiency (PS-Efficiency Y(II)), non-photochemical quenching (Y(NPQ)), and non-regulated energy dissipation (Y(NO)) are shown as line-graphs. The data is in percent as all three parameters together result in 1. D) The bar graph of an early time-point is shown for Y(II) to demonstrated the difference in Y(II). Y(NPQ) and Y(NO) bar graphs are shown for the last time. P-values were generated based on a Student's t-test. E) The light-curves and the corresponding F_v/F_m values are presented with the light-curves as line-graph and the F_v/F_m measurement as boxplots. Student's t-test was used for statistical analysis.

Changes in chlorophyll/cell, as well as FL3/cell, SSC/cell, and FSC/cell over time, revealed clear trends for the iron-recovery samples. The ST experiment showed significant changes in the iron-recovery samples compared to the low-iron samples for FL3/cell, SSC/cell, and FSC/cell. When looking at the results of these parameters for the

LT experiment, the values of the iron-recovery samples reached almost high iron values (Figure 3.4).

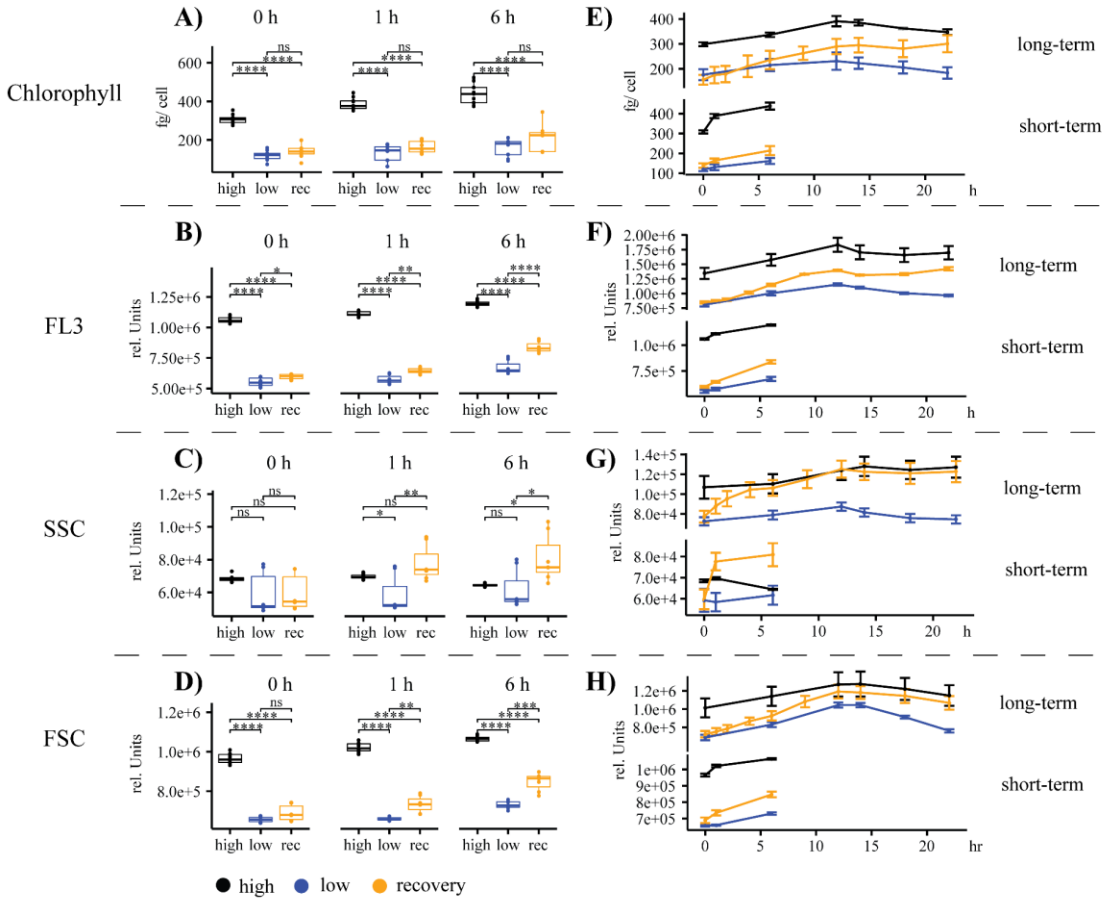


Figure 3.4: Changes in cellular chlorophyll content and cell properties over time.

Measurements of high-iron samples are shown in black, low-iron samples are shown in blue, and iron-recovery samples are shown in orange. Two tail student's t-test was used for statistical analysis of bar graphs in A)-D). The bar graphs represent six measurements of each timepoint (0 h, 1 h, 6 h) for the ST experiment. A) Chlorophyll content per cell in fg/cell B) FL3 values are shown in relative units C) Side-scatter (SSC) measurements are shown in relative units D) Forward-scatter (FSC) values shown in relative units. E)-F): The line-graphs show the trend of each parameter over the 22 h sampling period divided into long-term and short-term experiments. The standard error is indicated for each timepoint. E) Chlorophyll measurements are in fg/cell F) FL3 measurements are shown in relative units G) SSC measurements are shown in relative units H) FSC measurements are shown in relative units.

3.3.2 Changes in transcript abundance following iron resupply

The expression patterns were first analyzed in a correlation matrix based on Spearman correlation. The visualization in a heat map revealed nine iron-responsive genes, with seven genes that were upregulated under iron-limitation and two genes that were downregulated under iron-limitation. All nine genes showed a rapid change in transcript counts following iron resupply. The values for the iron reductase (FRE1), iron-starvation induced protein1 (ISIP1), ISIP3, cellular-repressor-of-EA1-regulated-genes x2 (CREGx2), fructose bisphosphate aldolase 3 (FBA3), FBA4, and flavodoxin (FLDA1) showed a sharp decline after iron resupply (cluster a in Figure 3.5), while FBA1 and ferredoxin (PETF) exhibited the opposite trend and showed a sharp increase in transcript numbers following the addition of iron (cluster b in Figure 3.5). The correlation of the samples was predominantly based on the iron status creating a strict division between high- and low-iron samples (Figure 3.5). For better visualization, a split region indicated as cluster 2 (Figure 3.5) was added. This cluster includes iron-recovery samples originated 15-60 min after iron supply. The adjustment of transcript counts for ISIP1, ISIP3, and FRE1 in the iron-recovery samples was completed before the adjustments of transcript counts in FBA3, FBA4, FLDA1, and CREGx2 reached their plateau. The heat map showed two clusters of samples (clusters 1 and 3 in Figure 3.5) from 7 pm and 9 pm, which is the switch from the light-period to the dark-phase (see also Supplemental Figure 4). Transcripts with count differences and a defined border in this cluster were FBA1, FBA3, PETF, and Mn_SOD. The DMSO treated samples, serving as a control for the actD treatment, were well distributed throughout different samples in the heatmap.

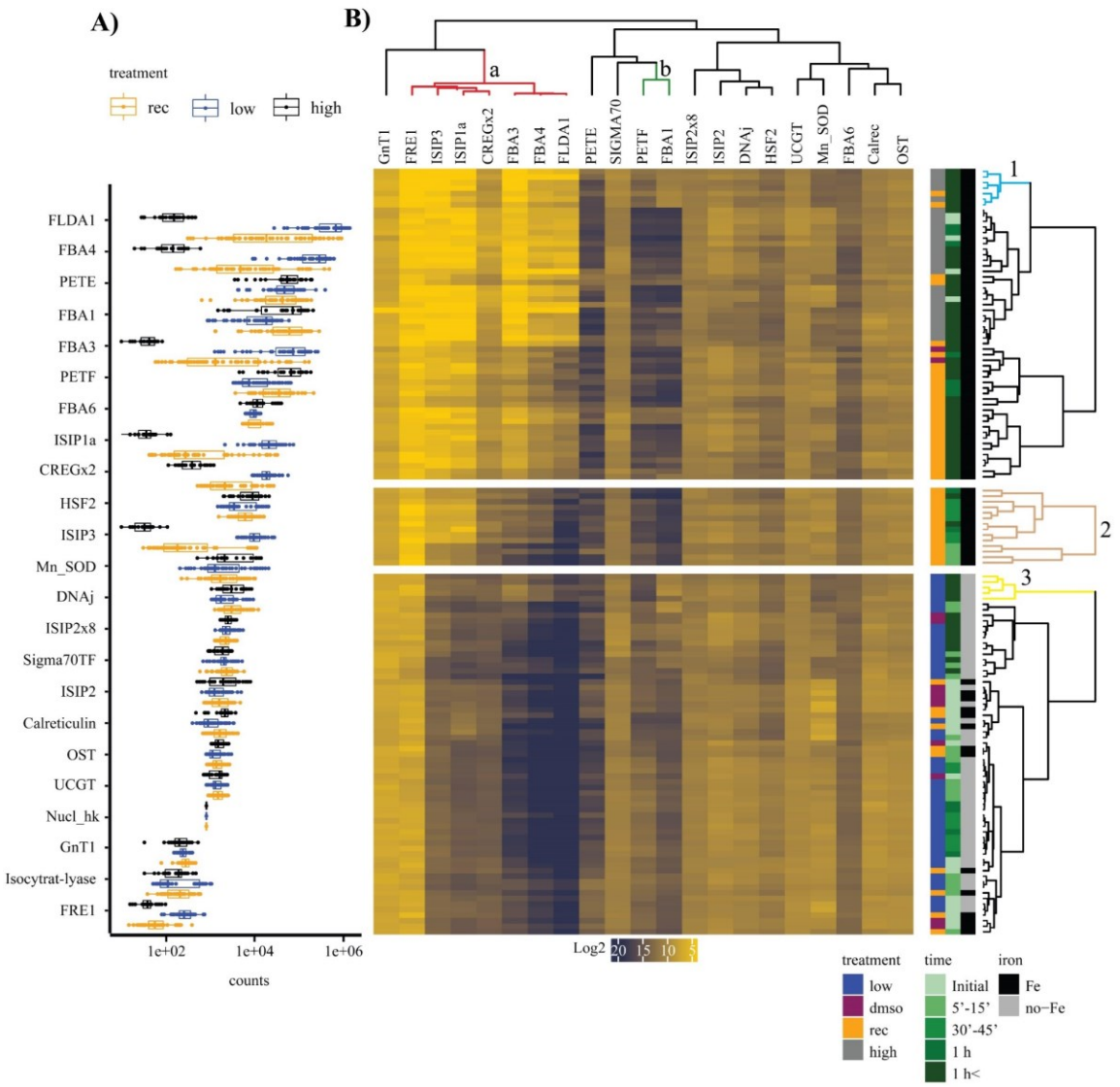


Figure 3.5: Correlation analysis of transcripts counts using Spearman correlation visualized in a heatmap.

A) Low (blue), high (black), and iron-recovery (orange) samples are plotted for each gene. The x-axis is plotted in log-scale, and the transcript counts range from less 100 to over 1 Mio. The genes are plotted in order of decreasing transcript abundance. B) Clustering based on Spearman correlation visualized in a Heatmap. Transcript counts are plotted as log2 values with yellow indicating low-iron counts and dark-blue indicating high-iron counts. Three colored bars are used for an indication of important parameters to differentiate samples. Low-iron samples are shown in blue, DMSO-treated samples are purple, iron-recovery samples are shown in orange, and high-iron samples are grey. Timepoints are separated in Initial, 5-15 min, 30-45 min, 1 h, and past 1 h in different shades of green with darker shades indicating later timepoints. Samples that received iron are indicated black, whereas samples without any iron are shown in grey. The colors in the dendrogram indicate

clusters with: a) including the downregulated genes following iron addition and b) the upregulated genes. Cluster 1 and 3 have samples from 7 pm and 9 pm, and the split region is shown as cluster 2 including iron-recovery samples.

The fold changes of all transcripts that were analyzed in the ST and the LT experiment were calculated. The nine iron-responsive genes, including ISIP1, ISIP3, FRE1, FBA3, FBA4, FBA6, PETF, FLDA, and CREGx2 showed significant differences between their high- and low-iron counts samples. The highest fold change (3279x) had FLDA1. FBA3 (2142) and FBA4 (1550) revealed high fold changes as well (Figure 3.6).

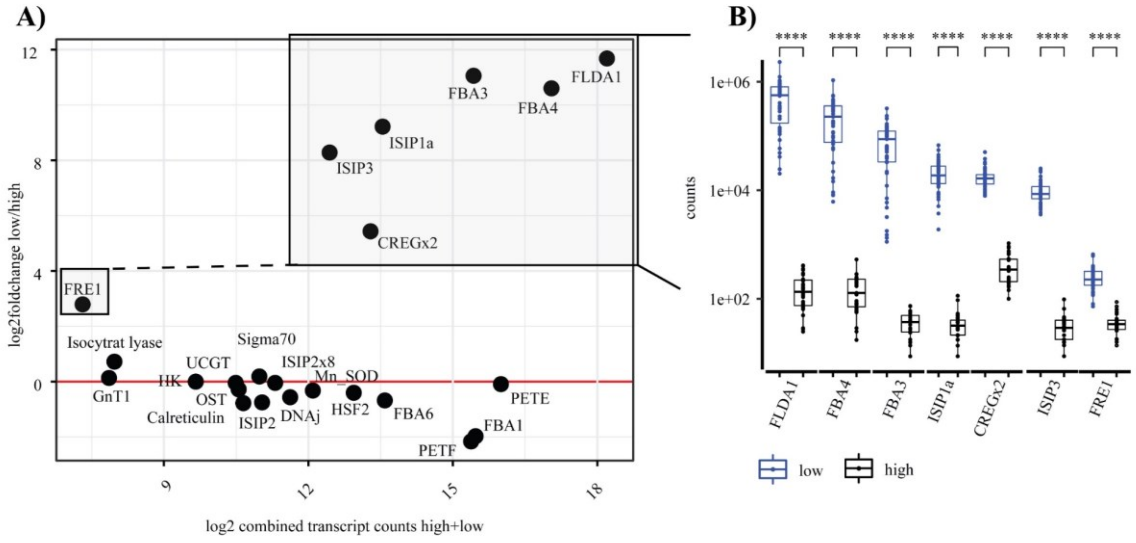


Figure 3.6: Overview of foldchanges between high- and low-iron samples and their significant statistical difference.

A) All genes analyzed in the ST and the LT experiment are shown with their mean counts across all high- and low-iron samples on the x-axes and their foldchange between high and low iron on the y-axes. Both values are transformed into log2 values for better comparison. B) The seven transcripts with the highest foldchanges are shown. Whisker plots representing high- and low-iron measurements are plotted with Student's t-test used for significance analysis between high- and low-iron transcript counts.

3.3.3 Transcript dynamics

Transcript counts were normalized against one house-keeping gene (nuclear import-export protein) and plotted on a log-scale (Figure 3.7). For better illustration, the genes

were divided into five different groups based on functional or spatial characteristics (Figure 3.7). All four ISIP proteins (ISIP1, ISIP2, ISIP2x8, ISIP3) were combined in one group as well as the fructose-bisphosphate aldolases (FBA1, FBA3, FBA4, and FBA6). The other three groups are transcripts for proteins involved in glycosylation (Calreticulin, oligosaccharyltransferase (OST), N-acetylglucosaminyltransferase 1 (GnT1), UDP-glucose:glycoprotein glucosyltransferase (UCGT)), transcripts involved in photosynthesis (flavodoxin (FLDA1), superoxide dismutase (Mn_SOD), plastocyanin (PETE), ferredoxin (PETF)), and a last group with 2 transcription factors (HSF2, SIGMA70), an iron reductase (FRE1) and CREGx2, a protein of unknown function.

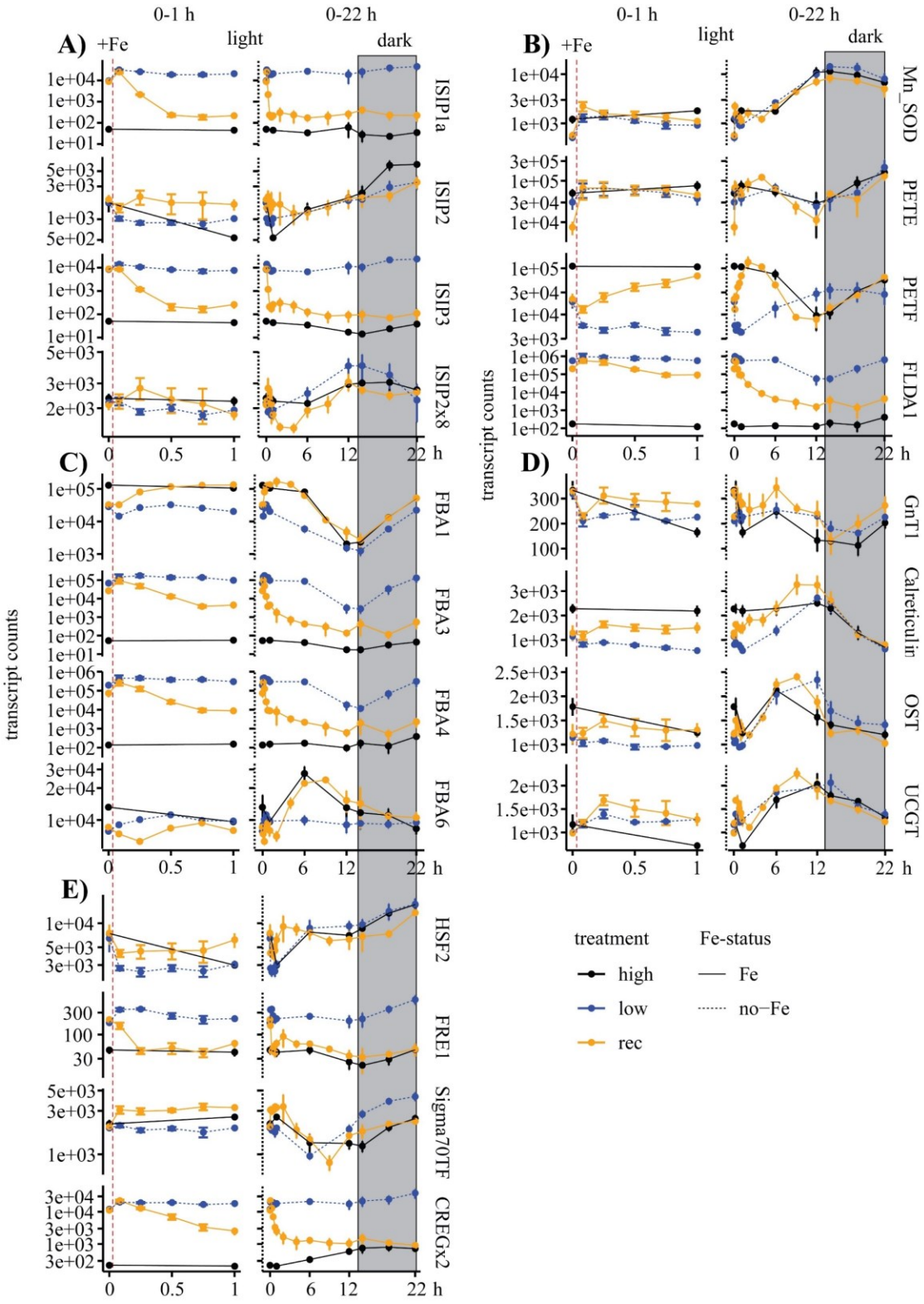


Figure 3.7: Overview of targeted transcript counts of high-iron, low-iron, and iron-recovery samples over the 6 h period and the 22 h period.

Transcript counts from the Ncounter analysis are on the y-axes as indicated, and time in hours is on the x-axes. The grey box indicates the dark phase. The first hour is shown on the left, and the addition of iron is indicated with a red dashed line. 0-22 h samples are shown on the right, and iron addition is not indicated based on the resolution. High-iron samples are in black, low-iron samples are shown in blue, and iron-recovery (rec) samples are shown in orange. Error bars indicate standard errors. Transcripts are grouped with A) iron starvation-induced proteins (ISIP), B) proteins involved in photosynthesis, C) Fructose-bisphosphate aldolases (FBA), D) proteins involved in the glycosylation pathway, E) 2 transcription factors, iron reductase (FRE1) and a protein of unknown function (CREGx2). The following abbreviations are used for transcripts: iron starvation-induced protein (ISIP1-3), fructose-bisphosphate aldolase (FBA1-4), heat-shock factor (HSF), iron-reductase (FRE1), transcription factor (TF), cellular repressor of EA1 activated genes (CREG), manganese superoxide dismutase (Mn_SOD), plastocyanin (PETE), ferredoxin (PETF), flavodoxin (FLDA), N-acetylglucosaminyltransferase (GnT1), oligosaccharyltransferase (OST), UDP-glucose glucosyltransferase (UCGT).

The main focus of this study was the temporal change of transcript counts following iron addition. Nine iron-responsive genes were identified in the Spearman correlation-based analysis and were further analyzed with respect to their temporal changes. When comparing the nine iron-responsive genes, FRE1 showed the fastest acclimation of transcript counts in the iron-recovery samples and reached high-iron levels after 15 min. ISIP1 and ISIP3 transcript counts were acclimated after 30 min upon iron addition. Transcript counts of PETF and FBA1 showed a fast acclimation as well, and FBA1 was acclimated after 30 min. PETF iron-recovery transcript counts reached high-iron levels after 60 min. FLDA1, FBA3, FBA4, and CREGx2 showed a fast decline, but the adjustment of transcript counts in iron-recovery samples to reach high-iron levels exceeded the first hour.

In contrast to ISIP1 and ISIP3, ISIP2 and ISIP2x8 did not exhibit a well-defined response to iron. ISIP2 counts steadily increased throughout the 22 h period independent of the treatment and reached its maximum at 5 am. ISIP2 high iron samples showed a minimum count at 8 am, and a maximum count at 1 am. ISIP2x8 revealed a diel cycle with

maximum transcript counts between 7 pm and 9 pm for all three conditions (Figure 3.7). ISIP1 and ISIP3 clustered with the iron-responsive genes in the heatmap, whereas ISIP2, and ISIP2x8 clustered with HSF2 and DNAj.

FBAs are essential enzymes in the Calvin cycle as well as in glycolysis and gluconeogenesis, converting C3 and C6 sugars. The expression patterns in this study differed based on the FBA type and their predicted intracellular localization. FBA1 and FBA6 belong to the class-II aldolases with a divalent metal as a cofactor, whereas FBA3 and FBA4 are part of the class-I aldolases, bearing a Schiff-base as a cofactor. Within the iron response, these proteins are part of the enzyme replacement strategy. FBA1 and FBA3 are targeted to the chloroplast based on *in-silico* analysis. While FBA1 was upregulated following the addition of iron, FBA3 showed a rapid downregulation. Both transcripts showed a minimum expression at 7 pm and 9 pm for low iron samples. FBA1 exhibited this trend in all three conditions. FBA4 and FBA6, putatively cytosolic proteins, showed differences in their transcript counts for all three treatments. FBA3 showed an almost identical pattern to FBA4 for all three treatments, resulting in a very high correlation (Figure 3.13). The iron-recovery samples of FBA6 followed the transcript counts of the high-iron samples, including a mid-day maximum that was absent in the low-iron samples (Figure 3.7).

The chloroplast targeted Mn_SOD showed a clear pattern for all three treatments with maximum counts between 9 pm and 1 am. There was no response following the addition of iron, and all treatments followed the same trend. Similarly, PETE counts showed no differences between the three treatments with a minimum expression at 7 pm, followed by an increase in the dark period. Flavodoxin counts had a strong response following the

addition of iron with a fast decrease. Low-iron counts of FLDA1 were similar to low-iron counts of FBA3, FBA1, and FBA4 with a minimum in the 7 pm and 9 pm samples (Figure 3.7).

Four transcripts were analyzed that are involved in N-linked glycosylation. These included OST, calreticulin, Gnt1, and UCGT. The only transcript showing a response to the addition of iron was calreticulin, with an upregulation of the iron-recovery samples during the light period when compared to the low iron samples (Figure 3.7).

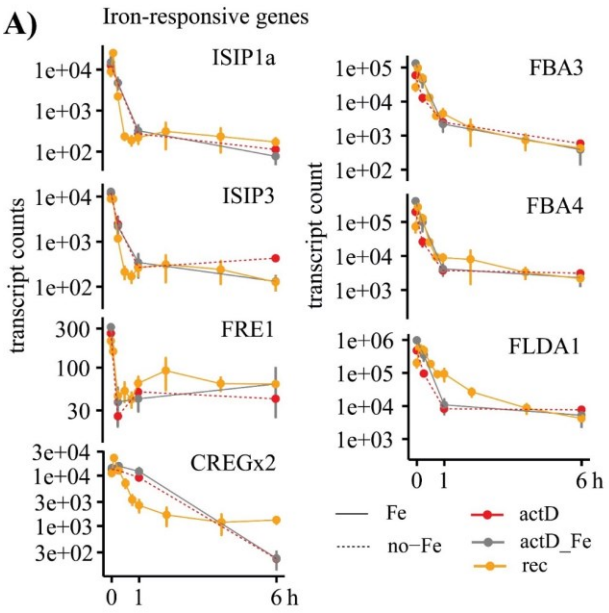
The last group contained two iron-responsive genes, FRE1 and CREGx2. Counts of both mRNAs showed a fast decrease after iron addition. The heat shock factor HSF2 showed first a decline for high-iron samples and iron-recovery samples while the counts for low iron samples increased. After six hours, all three treatments followed the same pattern, and a slight increase in counts was visible for all three treatments. The counts of the chloroplast targeted transcription factor SIGMA70 increased after the addition of iron, but the overall pattern appeared similar for all three treatments, with a minimum count around 1 pm and a subsequent steady increase until 5 am. Low-iron counts were elevated in the dark period compared to iron-recovery and high iron samples (Figure 3.7).

3.3.4 Actinomycin D decay curves

To answer the question of whether specific mRNAs are stabilized under iron limitation, the transcription inhibitor actinomycin D was used to investigate the half-life of the transcripts. Samples treated with actD showed a decrease for all targeted transcripts with some differences in respect to the resulting half-lives (Figure 3.8). The change point, defined as the point of time where the slope of the curve reaches its plateau, was also

calculated (Supplemental Table 7). There were seven iron-responsive transcripts (ISIP1a, ISIP3, FRE1, FBA3, FBA4, CREGx2, and FLDA1) that were upregulated under low-iron conditions and showed a rapid decline after the addition of iron (Figure 3.7). Four of these targets (ISIP1a, FLDA1, FBA4, and CREGx2) revealed significant differences in their half-lives, when comparing actD treated samples and iron-recovery samples.

The average half-life of all transcripts, excluding Mn_SOD, PETE, PETF, and CREGx2, was around 8 min. The half-lives of Mn_SOD, PETE, PETF, and CREGx2 were between 50 and 90 min (Figure 3.8 and Supplemental Figure 3). The actD experiment included the treatment of low-iron samples only with actD and with actD and iron in combination, but there were no significant differences for any targeted transcript between these two treatments.



B)

Gene	actD	actD_Fe	rec	in min
DNAj	11.06	16.12	-	
FBA1	7.82	8.87	-	
FBA2	7.43	7.49	-	
HSF2	7.95	10.77	-	
ISIP2	7.00	8.96	-	
ISIP2x8	7.40	7.08	-	
Mn_SOD	90.02*	92.42*	-	
PETE	67.19*	49.99*	-	
PETF	78.92*	93.25*	-	
Sigma70TF	11.85	15.18	-	
CREGx2	58.25*	56.13*	21.55	
FBA3	8.83	6.84	15.46	
FBA4	6.51	5.51	15.93	
FLDA1	8.45	7.39	43.96	
FRE1	5.42	4.77	7.86	
ISIP1a	10.14	10.19	5.38	
ISIP3	5.65	6.01	5.14	
average	8.12	8.86	16.47	

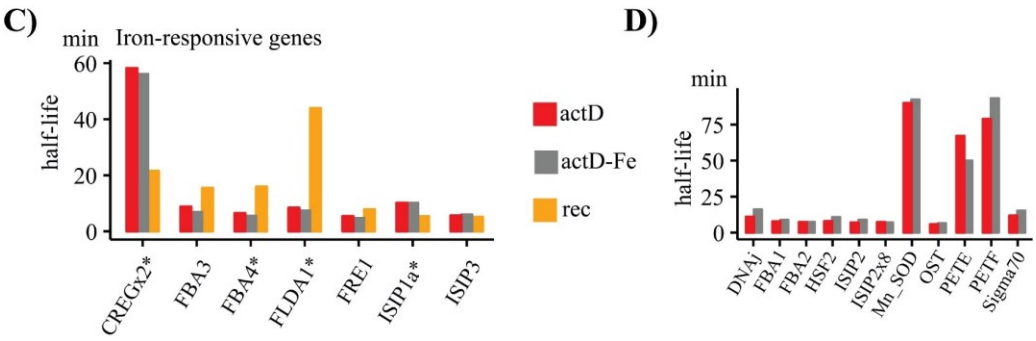


Figure 3.8: Overview of actinomycin D induced mRNA decay in comparison to iron-induced reduction of mRNA transcripts.

A) Changes in transcript count over time (6 h) in iron-recovery (orange), actD-treated cells (red), and actD-Fe cells (grey). "actD" samples are low-iron samples that received actinomycin D (red), "actD-Fe" are low-iron samples that received actinomycin D and iron (grey), and "rec" are low-iron samples that received only iron (orange). B) A table of half-lives for transcripts analyzed in the ST experiment is shown (in min). The half-lives were calculated based on the slope of the decay curve k (half-life= ln2/k) with the assumption that iron-responsive genes stop transcription after the addition of iron. C) Bar graphs of half-lives from the 7 iron-responsive genes that are rapidly downregulated are shown. Transcripts with an asterisk revealed significant differences between actD and iron-recovery samples. D) Bar graphs of half-lives for non iron-responsive genes are shown.

3.3.5 Tara Oceans analysis

The analysis of the global expression pattern of the targeted transcripts and the global distribution of *T. oceanica* was conducted by a blastp search on the Tara Oceans database targeting *T. oceanica* specific sequences. The Tara Oceans database taxonomic identifier showed *T. oceanica* specific sequences for Mn_SOD, FLDA1, and FBA4 (Figure 3.9). Other sequences were within the order of *Thalassiosirales,* and a blastp-search of sequences received from the Tara Oceans database for ISIP1 and ISIP2 returned *T. oceanica* as best-hits (data not shown). The low iron HNLC region in the east Pacific showed most genes present and expressed, and PETF showed the broadest distribution in abundance and expression (Figure 3.9).

Figure 3.9: Abundance analysis across the global ocean of targeted transcripts using the Tara Oceans database.

The Tara Oceans database was used to generate a proxy for the global distribution and expression of selected targeted transcripts. E-values of 10^{-50} to 10^{-250} were used to receive closely related genes to *T. oceanica*. The results are divided into the surface (SRF) and deep-chlorophyll max (DCM). The data is derived from the metagenome (MetaG) and the metatranscriptome (MetaT). Blue dots represent size fraction 0.8-5µm, and yellow dots show size-fraction 5-20 µm with the size of the dots being proportional to the percentage of total reads. The following abbreviations are used: iron starvation-induced protein (ISIP1-3), fructose-bisphosphate aldolase (FBA1-4), heat-shock factor (HSF), iron-reductase (FRE1), transcription factor (TF), cellular repressor of EA1 activated genes (CREG), manganese superoxide dismutase (Mn_SOD), plastocyanin (PETE), ferredoxin

(PETF), flavodoxin (FLDA), N-acetylglucosaminyltransferase (GnT1), oligosaccharyltransferase (OST), UDP-glucose glucosyltransferase (UCGT).

3.3.6 Transcription factor analysis

The *in-silico* search for transcription factors (Tf) in the *T. oceanica* genome resulted in 220 putative Tfs with heat shock factors (HSF) and Myb Tfs accounting for around 50% of the Tfs. A second large group included the bZIP proteins and Zn finger type Tfs. Based on previous results (Lommer *et al.*, 2012), differentially expressed transcription factors were analyzed in the NanoString experiment. The analysis showed HSF2 as the only gene with a significant difference between high- and low-iron samples. Sigma70 and bzip_20638 showed an increase after the addition of iron. Sigma70 was also included in the short-term analysis, and transcript counts increased after the addition of iron but only temporarily. The increase of Sigma70 and bzip_20638 after the first hour in the iron-recovery samples cannot be defined as iron-responsive as the control samples of high- and low-iron are missing for these timepoints. Most of the transcription factors showed a steady expression throughout the 22 h. Bzip_20638, Homeobox, HSF2, and SIGMA70 had a maximum at 5 am (Figure 3.10).

A)

	Phaeodactylum tricornutum	*Thalassiosira pseudonana*	*Thalassiosira oceanica*
HSF	70	94	49
Myb	34	37	54
bZIP	23	23	27
Zn_finger,C2H2-type	20	21	17
Zn_finger,CCCH-type	15	19	25
Zn-finger,TAZ-type	5	6	4
Zn_finger,GATA-type	0	0	0
bHLH	7	5	3
Fungal,TRF	1	4	4
CCAAT-binding	1	0	0
Homeobox	4	9	12
AP2-EREBP	3	12	8
MADS	0	0	0
Nin-like	0	0	0
TAF9	1	1	6
E2F-DP	5	3	1
ABI3/VP1	0	0	0
WRKY	0	0	0
Zn_finger,DOF type	0	0	0
CBF/NF-Y/archael	5	6	4
CXC/tesmin	5	5	0
Sigma-70	8	8	6
CSF	5	5	0
Total	212	258	220

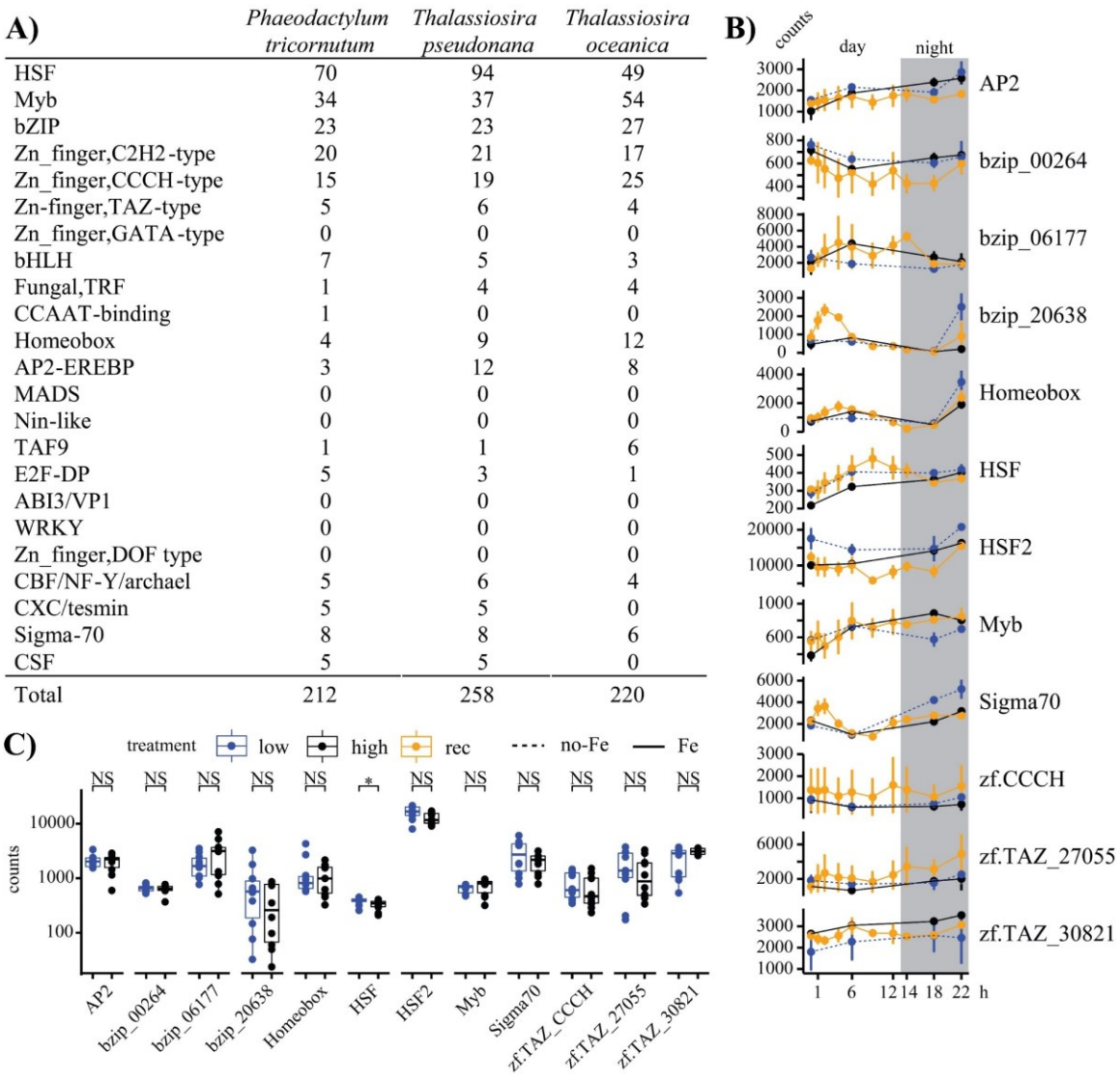


Figure 3.10: *In-silico* analysis of the *T. oceanica* proteome for transcription factors and transcript expression analysis in the LT experiment.

A) The table shows an overview of transcription factors in *P. tricornutum, T. pseudonana,* and *T. oceanica*. Tfs from *P. tricornutum, T. pseudonana* were taken from Rayko et al. (Rayko *et al.*, 2010). B) Expression patterns of the transcription factors under high-iron (grey) low-iron (blue), and iron-recovery (orange) are shown. The plotted data is taken from the long-term experiment run and spans over 22 h (x-axes). Transcript counts are on the y-axes. The grey box indicates the dark phase. Dashed lines indicate iron-free samples, while solid lines represent samples with iron. C) High- (grey) and low-iron (blue) samples are plotted in boxplots. Student's t-test was used for statistical analysis. Counts are on the y-axes, and the gene-names are indicated on the x-axes.

3.4 Discussion

The cellular response to iron limitation has been a focus of research efforts since the early 1990s after in-situ iron fertilization experiments demonstrated prompt algae growth upon iron induction resulting in diatom dominated blooms (Boyd *et al.*, 2007; Trick *et al.*, 2010). Laboratory studies on specific diatom species followed, demonstrating significant differences between high- and low-iron conditioned diatoms on transcriptional and physiological level (Peers *et al.*, 2005; Maldonado *et al.*, 2006; Kustka & Allen, 2007; Allen *et al.*, 2008; Lommer *et al.*, 2012; Smith *et al.*, 2016). The photosynthetic apparatus in *T. oceanica* is well adapted to chronic iron limitation. In more detail, the exchange of cytochrome c_6 by PETE, as well as the downregulation of photosystem I and the replacement of PETF with FLDA1 are critical adaptations, decreasing the iron-demand of the cell (La Roche *et al.*, 1993; Strzepek *et al.*, 2011; Lommer *et al.*, 2012). Iron-addition experiments in the north Pacific showed inconsistent responses of FLDA1 and PETE to the addition of iron, but the data confirmed a high transcription of PETE after iron addition (Marchetti *et al.*, 2012; Cohen *et al.*, 2017). While most studies looked at adaptive strategies, the study presented here focuses on the direct response to iron addition following iron limitation. The analysis of the response to iron addition was conducted in the open ocean diatom *T. oceanica,* which belongs to the worldwide distributed diatom genus *Thalassiosira* (Malviya *et al.*, 2016).

The results of the present study revealed a fast recovery of photo-physiological parameters, including ETR, PE (Y(II)), and F_v/F_m. All three parameters were almost completely recovered after 6 h, indicating an active remodeling of the photosynthetic apparatus (Figure 3.3). Field studies analyzing the transcriptional response of diatoms

from the genera *Pseudo-nitzschia* and *Thalassiosira* towards iron addition showed that both genera upregulate genes involved in the photosynthetic apparatus within 48 h (Cohen *et al.*, 2017). *T. oceanica* showed an overall decrease in chloroplast transcripts under iron-limiting conditions, and fg chlorophyll per cell was lower in *T. oceanica* (Lommer *et al.*, 2012) and *P. tricornutum* (Allen *et al.*, 2008) under iron limitation. In contrast, *T. pseudonana* showed continuous expression of PSI and PSII proteins following one week of acclimation to low-iron conditions (Nunn *et al.*, 2013). While the efficiency of photosystem II (F_v/F_m) improved after 1 h, the recovery of the chlorophyll content per cell (fg/cell) increased within 22 h but did not fully adjust to high-iron levels. Structural changes within the PSII complex, as observed in *Chaetoceros* (Petrou *et al.*, 2014), and an increase in PSII-D1 protein content in *T. oceanica* (Lommer *et al.*, 2012) will be contributing to the increase in F_v/F_m. Components of the photosystem were not recorded in the present study, but the fast recovery of the above-mentioned parameters suggests a fast remodeling of the photosynthetic apparatus in the proposed study as well. The slower increase in chlorophyll per cell compared to the recovery of F_v/F_m, ETR, and PE was most likely caused by the complexity of chlorophyll synthesis itself.

FLDA1 was described as an iron-regulated protein in algae in the mid-1990s and further suggested as an indicator for iron limitation (La Roche *et al.*, 1993, 1996; Geider & La Roche, 1994). FLDA1 expression is suggested as a marker for iron limitation (Chappell *et al.*, 2015), even though some diatom species express FLDA permanently, FLDA expression under iron-limitation can be controlled by manganese limitation (Wu *et al.*, 2019), and the short-term response to iron addition can result in an upregulation of FLDA1 (Pankowski & Mcminn, 2008; Marchetti *et al.*, 2012; Graff van Creveld *et al.*,

2016). In the study presented here, FLDA1 had the highest fold change between high- and low-iron samples, and it had the most transcript counts in the low-iron samples (Figure 3.6). Within the iron-responsive genes that were downregulated after iron addition, FLDA1 was the slowest, and it stayed above high-iron counts over the complete 22 h period (Figure 3.6). Upregulation of FLDA1 upon iron addition in *P. tricornutum* (Graff van Creveld *et al.*, 2016), as well as in field samples (Marchetti *et al.*, 2012), stands in contrast to the findings in the proposed study and points towards the existence of iron-responsive and iron-neutral types as proposed by Whitney *et al.* (Whitney *et al.*, 2011). FLDA1_*To* is part of the clade II which is the iron-responsive clade (Whitney *et al.*, 2011; Lommer *et al.*, 2012). Investigations into specific FLDAs are needed, and the use of FLDA as an iron indicator has to be handled with care, with species-specific probes being preferred. However, based on the rapid decline, the present study supports the use of FLDA_*To* as a useful measure of the iron status in the ocean. The upregulation of PETF was completed after 1 h (Figure 3.7), and not only the exchange of FLDA1 with PETF was the reason for the fast increase in PETF transcript counts, also the upregulation of ferredoxin dependent enzymes in the form of glutamate synthases and sulfite reductases as seen in field samples for the genera *Thalassiosira* could be contributing to the rapid upregulation (Cohen *et al.*, 2017). The broad range of functions could also be the reason that PETF had the highest abundance in the Tara Oceans database compared to the other transcripts of the ST experiment (Figure 3.9). On the other side, PETF had ten times lower transcript counts in our analysis compared to FLDA1, which might be a hint towards a lower efficiency of FLDA1 or an even broader range of functions for FLDA1 under iron limitation (Figure 3.7).

The half-life analysis with actD showed an average half-life of around 8 min with a few exceptions. One of the exceptions was PETF, which had an above-average half-life indicating a constant translation of PETF transcripts in the cell (Figure 3.8). The half-life of FLDA1 in the actD treated samples was 8 min, but a significantly higher half-life in the iron-recovery samples was observed. The slower decay of transcripts after the addition of iron compared to the immediate transcriptional stop caused by actD points towards a regulatory transcript that is controlling mRNA stability of FLDA1 (Figure 3.8). This regulatory transcript would not have been transcribed in the actD samples, which explains the similarity between the two actD treatments.

Radical oxygen species (ROS) are naturally occurring in the chloroplasts during photosynthesis, and superoxide dismutases (SOD) catalyze the dismutation of ROS (Peers & Price, 2004). The proposed study did not reveal any differences between the treatments regarding the expression pattern for Mn_SOD, and all three treatments followed the same trend with a maximum count in the dark phase (Figure 3.7). The similarity between the treatments was somewhat unexpected, as previous data showed no transcripts of Mn_SOD under low iron conditions (Lommer *et al.*, 2012). Considering that Mn_SOD responds to ROS production, the data in the study described here indicate that ROS production in *T. oceanica* is independent of the iron status. The diel pattern of Mn_SOD could be based on signaling activities between chloroplast and nucleus, controlled by H_2O_2 production in the chloroplast. Another reason could be a time delay between the transcription and translation of Mn_SOD. Similar to PETF, Mn_SOD had a longer half-life compared to the other transcripts, demonstrating a constant demand for the removal of ROS (Figure 3.8).

PETE replaces cytochrome c_6, and it is not widely distributed throughout diatoms, mostly present in low iron adapted diatoms (Groussman *et al.*, 2015). The proposed study did not show any apparent differences between the three treatments for PETE transcript counts. This, and a longer than average half-life confirm the permanent replacement of cytochrome c_6 by PETE, independent of iron status in *T. oceanica* (Peers & Price, 2006) (Figure 3.7 and Figure 3.8). The expression of PETE in the field showed a mixed response of PETE upon iron addition (Cohen *et al.*, 2017), but low transcript abundance of cytochrome c_6 in the field (Marchetti *et al.*, 2012) confirms the permanent use of PETE. While one laboratory study on *T. oceanica* reported downregulation of PETE transcripts under iron-limiting conditions (Lommer *et al.*, 2012), another study showed an upregulation of PETE proteins under iron limiting conditions (Hippmann *et al.,* in prep.). The concentration of iron, harvesting cells in different stages of their growth phase, as well as the amount of light, could influence the expression of PETE leading to these converse observations.

Adaptive strategies in diatoms for the survival in low-iron conditions include the upregulation of proteins that are involved in iron uptake, which is demonstrated by the upregulation of ISIP1, ISIP2, iron reductases, and multicopper oxidases (Allen *et al.*, 2008; Lommer *et al.*, 2012; Marchetti *et al.*, 2012; Morrissey & Bowler, 2012; Smith *et al.*, 2016; Cohen, Gong, *et al.*, 2018). ISIP1 and ISIP2 have recently been characterized as iron uptake proteins. ISIP1 is involved in endocytosis mediated uptake of iron-siderophore complexes that are delivered to the chloroplast (Kazamia *et al.*, 2018). ISIP2 is a general ferric ion uptake protein, belonging to the transferrin protein family based on the use of carbonate ions for functionality (McQuaid *et al.*, 2018).

The proposed study analyzed all four ISIPs (ISIP1, ISIP2, ISIP2x8, ISIP3) present in *T. oceanica* as well as the iron reductase FRE1, aiming to uncover the immediate response of iron uptake related proteins to iron supply. It was surprising that only ISIP1, ISIP3, and FRE1 responded to the addition of iron with a fast downregulation reaching high-iron counts after 30 min for both ISIPs and only 15 min for FRE1 (Figure 3.7 and Figure 3.8). The expression patterns of ISIP1 and ISIP3 showed a high correlation (R^2= 0.94) (Figure 3.11), which can be the result of a common regulation. The promoter motif that was previously described for ISIP1, FLDA1, and FBA3 (Lommer *et al.*, 2012), is absent in ISIP3. However, the proposed data indicate that other functional elements must be present, controlling these two genes in such a similar fashion. Localization studies on ISIP3_*To*, heterologously transformed into *P. tricornutum*, showed a localization similar to the localization of the internalized siderophores in the ISIP1_*Pt* study (Kazamia *et al.*, 2018). The localization and the similarity in their response to iron indicates a possible interaction between ISIP1 and ISIP3 or the involvement in the same pathway (Chapter 2). It is likely that ISIP1 and ISIP3 are involved in controlling iron homeostasis in the chloroplast. However, the exact role of ISIP3 and ISIP1 remains unclear. The classic high-affinity uptake system is a reductive uptake system with the involvement of an iron reductase (Terzulli & Kosman, 2010). The ferric-reductase activity increased in *P. tricornutum* under iron limitation, and reduction plays a role in capturing organically bound iron in *T. oceanica* (Maldonado & Price, 2001; Allen *et al.*, 2008). The data presented here confirmed the upregulation of the iron reductase FRE1 under low iron conditions (Figure 3.6), and FRE1 reached its baseline in the iron-recovery samples after

15 min (Figure 3.7). The overall low counts of FRE1 may indicate minor importance of the reductase activity in *T. oceanica* or high enzymatic efficiency.

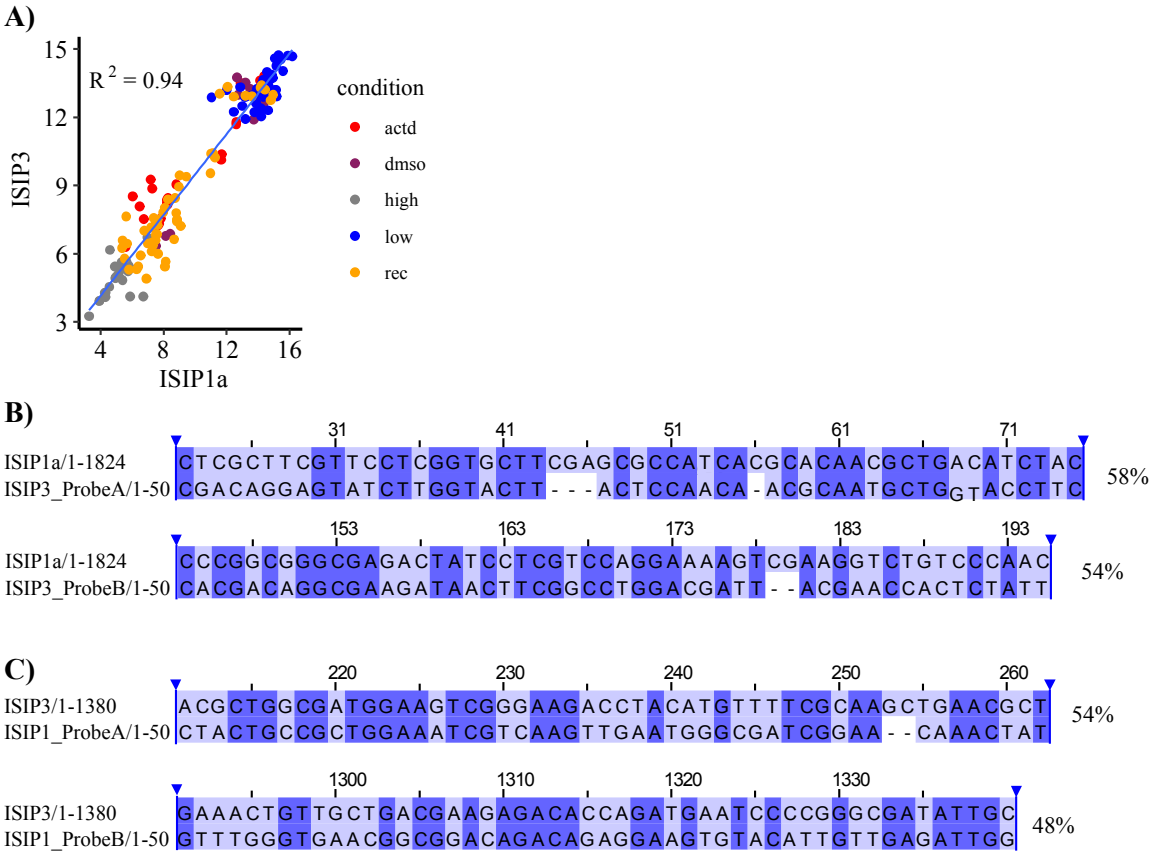


Figure 3.11: Correlation analysis of ISIP1/ISIP3 and NanoString probe analysis for putative cross binding.

A) All transcripts counts (log2 values) from ISIP1 and ISIP3 are plotted to show their correlation. ActD (red), DMSO (purple), high-iron (grey), low-iron (blue), iron-recovery (orange) B) Alignments of ISIP1 probe A and probe B onto full-length ISIP1 are shown. Dark blue indicates matching nucleotides. The percent identity is shown to the right of the alignments. A 75% identity is needed for possible false binding (Kane, 2000). C) Alignments of ISIP1 Probe A and B onto full-length ISIP3 is shown. Dark blue indicates identical nucleotides in both sequences with the full alignment percent identity shown to the right of each alignment.

The unknown protein cellular-repressor-of-EA1-regulated-genes x2 (CREGx2) was analyzed *in-silico* based on its strong upregulation (Figure 3.6) under low-iron conditions, the fast response to iron addition (Figure 3.12), and its unknown function.

The *in-silico* analysis led to the discovery of a complex I intermediate-associated protein 30 (CIA30) domain, a pyridoxamine 5'-phosphate oxidase (P5P) domain, and a histidine-rich site in CREGx2. Furthermore, a transmembrane domain and a signal peptide was identified, leading CREGx2 into the secretory pathway. A tree of the TOP100 hits of a blastp search showed that CREGx2_*To* clustered in a small clade differentiating CREGx2_*To* from other CREG proteins that contain only one of the conserved domains (Figure 3.12).

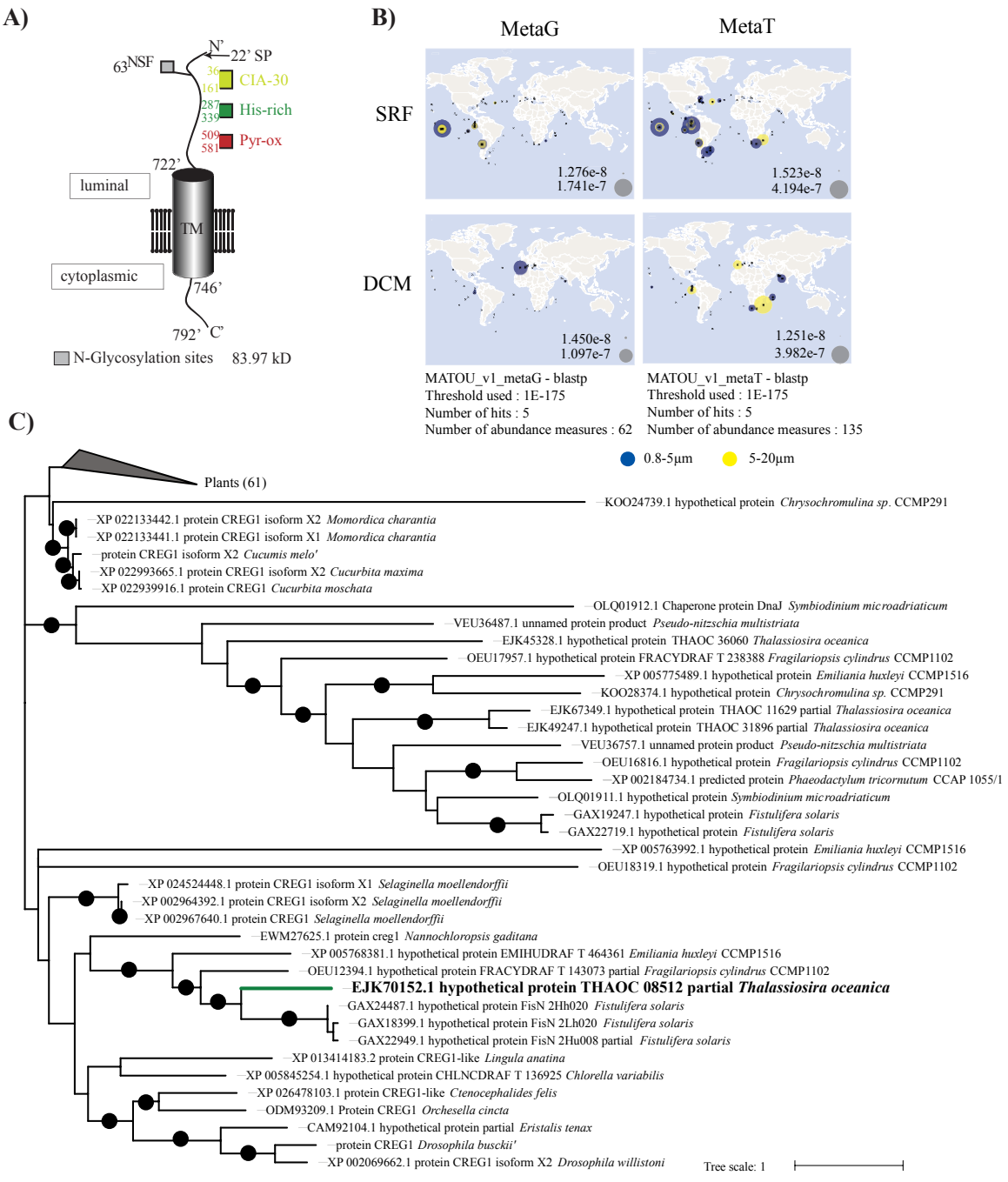


Figure 3.12: Analysis of CREGx2 for structural properties, transcript abundance in the Tara Oceans database, and the evolutionary distance to other CREG proteins.

A) The structure of CREGx2 is shown. Based on TMHMM (Emanuelson *et al.*, 2007) analysis the presence of a transmembrane domain (TM) is indicated as well as N-glycosylation sites predicted by NetNGlyco server (Chuang *et al.*, 2012). Two conserved domains were specified by NCBI domain search (Marchler-Bauer *et al.*, 2017) and are indicated in light-green and red. The third domain was identified by ProSite server analysis and is shown in green. The following abbreviations are used: signal peptide (SP), Complex

I intermediate-associated protein 30 (CIA-30), Pyrodoxamin-5'phosphate oxidase (Pyr-ox), histidine-rich region (His-rich). B) The Tara Oceans database was used for a global analysis of CREG genes with high similarity to *T. oceanica* as indicated under the plots. Results from the surface (SRF) and deep-chlorophyll max (DCM) samples are shown for 0.8-5 µm (blue dots) and 5-20 µm (yellow dots). The abundance is shown in percentage of total reads, and the circles reflect the percentage found in the metagenome (MetaG) and the metatranscriptome (MetaT) database. C) A blastp search on NCBI was used to generate a phylogenetic tree of the Top100 results. The *T. oceanica* sequence is bold.

CREGx2 was first described in *P. tricornutum* (Allen *et al.*, 2008) as an iron-responsive gene, upregulated under iron limitation, and later in *T. oceanica* (Lommer *et al.*, 2012). CREG proteins are suggested to be involved in cell proliferation (Allen *et al.*, 2008; Lommer *et al.*, 2012). The His-rich region present in CREGx2_*To* is present in ZIP proteins and functionally relevant for divalent metal transport (Ivanov & Bauer, 2017). A phylogenetic analysis of the Top100 sequences of a blastp search showed CREGx2_*To* as the only protein with all three domains and the only protein with a His-rich region (Figure 3.12). The data of the proposed study led to the hypothesis that CREG2 is involved in iron homeostasis. The strong upregulation under iron limitation, the immediate response to iron addition, and the histidine-rich region point towards a function as an iron uptake protein, but regulatory functions cannot be excluded at this point. The half-life analysis using actD revealed a significant difference between the decay of mRNAs after transcription stop through actD and the decay following iron addition. Opposite to the finding for FLDA1, the half-life in the iron-recovery experiment was significantly higher when compared to the actD samples. This difference indicates a regulatory transcript that influences the mRNA stability of CREGx2. The analysis of the Tara Oceans base demonstrated a high abundance and high expression of CREGx2 in the HNLC region of the equatorial Pacific, demonstrating the expression of CREGx2 under low-iron concentration in the environment.

The enzyme replacement strategy is an important contributor to lower the iron demand in the cell. The switch between PETE and FLDA1, as discussed above, is an example within the photosynthetic apparatus. The FBA enzymes are another example where enzymes are expressed according to their cofactor, and four FBA proteins, FBA1, 3, 4, and 6, were analyzed. FBA1 and FBA3 are putatively targeted to the chloroplasts, while FBA4 and FBA6 are putatively located in the cytosol. The localization of FBA proteins in pyrenoids has only been confirmed in *P. tricornutum* (Allen *et al.*, 2012), but FBA1_*To* and FBA3_*To* have been discussed to be localized in the pyrenoids (Lommer *et al.*, 2012). Localization studies are needed to show subcellular localization of each FBA. While FBA1 and FBA6 belong to class-II aldolases and require a divalent metal in their active site, FBA3 and FBA4 belong to the class-I aldolases using a Schiff-base as a cofactor. Aldolases are critical enzymes in glycolysis, gluconeogenesis, and the Calvin-Benson cycle catalyzing the conversion of fructose-1,6- bisphosphate (FBP) into dihydroxyacetone phosphate (DHAP) and glyceraldehyde-3-phosphate (GAP) (Allen *et al.*, 2012). The two class-II FBA proteins, FBA3 and FBA4, showed, together with FLDA1, the highest fold change between high- and low-iron samples. All three proteins had high transcript counts in the low-iron samples (Figure 3.6), and transcript-counts for both FBA proteins decreased instantly after iron addition (Figure 3.7). The expression data of FBA3 and FBA4 correlated very strongly (R^2>0.97) (Figure 3.13), but cross-hybridization can be excluded based on the sequence alignments that indicate identities of under 75% (Kane, 2000) (Figure 3.13). The strong correlation was somewhat unexpected, considering their differences in terms of the predicted localization, the resulting putative function, and a complete phylogenetic separation (Allen *et al.*, 2012). An *in-silico* search

for regulatory elements upstream of FBA3 and FBA4 did not result in any common putative promoter sequences, and the previously characterized iron-responsive promoter is only present in FBA3 (Lommer *et al.*, 2012; Yoshinaga *et al.*, 2014). The similarity is also visible in the actD treated samples, where FBA3 and FBA4 had very similar half-lives (Figure 3.8). Both transcripts had longer half-lives in the iron-recovery samples than in the actD samples, which could result from the transcription of a regulatory element that influences mRNA decay similar to FLDA1. The low-iron samples from FBA3 and FBA4, as well as all samples from FBA1 followed a diel cycle similar to enzymes of the Calvin-Benson in *P. tricornutum* (Smith *et al.*, 2016). The adjustment of FBA1 transcript counts following iron induction took 30 min, and iron-recovery samples showed a similar pattern to high-iron samples after that. The low fold change (log2= -1.97) of FBA1 between low- and high-iron samples, as well as the expression of FBA1 in the low-iron samples itself, shows that FBA1 was still active under iron-limiting conditions and was not fully exchanged by its putative replacement partner FBA3, which could indicate a lower enzymatic efficiency of FBA3. The expression of FBA4 also follows the CB cycle, as shown for *P. tricornutum* (Smith *et al.,* 2016), but putative localization of FBA4 in the cytosol indicates the involvement in a different metabolic pathway such as gluconeogenesis or glycolysis. The FBA4_*Pt* protein has a high sequence similarity (93% coverage, 62% ID, e-value: 2e-126) to FBA4_*To* but it showed an opposite expression pattern with a maximum count at the beginning of the dark phase correlating with other cytosolic proteins that are part of the glycolysis pathway (Smith *et al.*, 2016). Furthermore, FBA4_*Pt* was unresponsive to the iron status of the media, which is different towards FBA4_*To* (Allen *et al.*, 2012). Both FBA4s have a bacterial origin in

common, but based on the described differences, they are most likely involved in different metabolic processes. The iron-response of FBA6 and FBA4 is opposite, and while FBA6 transcript counts increased, FBA4 counts decreased rapidly within the first hour after iron-addition. Whether FBA6 and FBA4 are interchangeable needs to be shown, but the expression patterns are very different, pointing towards different functions in the cell.

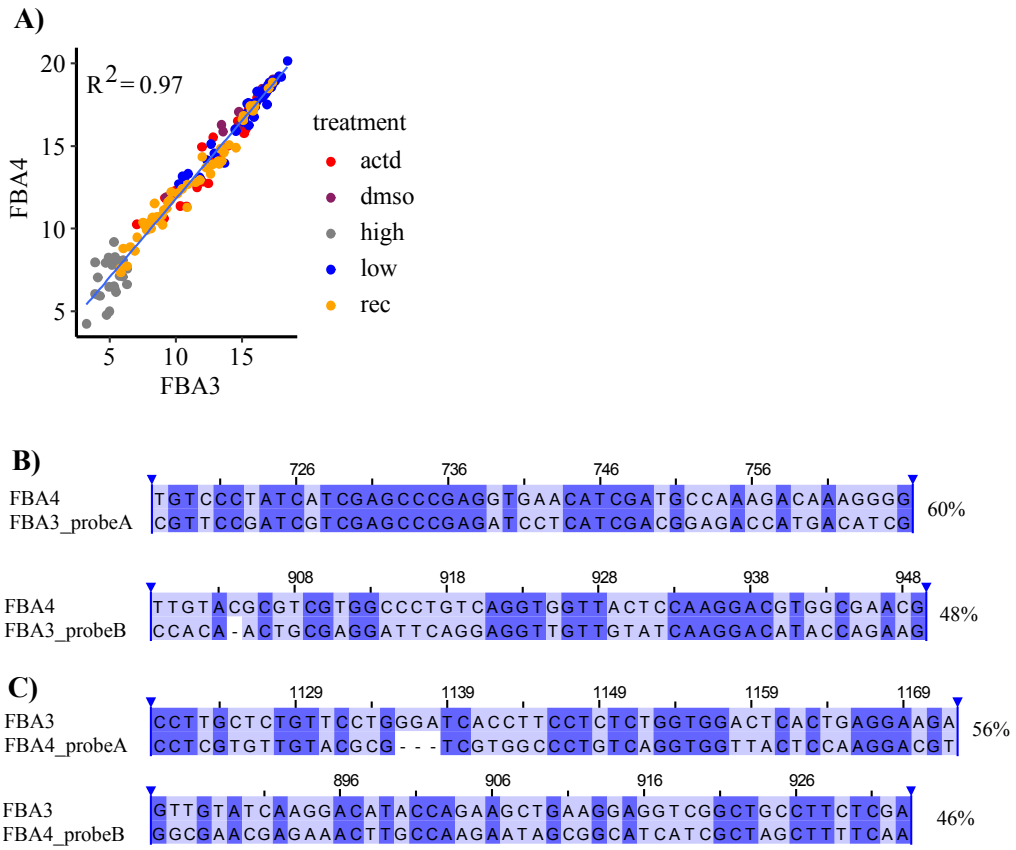


Figure 3.13: Correlation analysis of FBA3/FBA4 and NanoString probe alignment for possible cross binding activity.

A) Correlation analysis of all sample points for FBA3 and FBA4. actD (red), DMSO (purple), high-iron (grey), low-iron (blue), iron-recovery (orange) B) Alignments of FBA3 probe A and probe B onto full-length FBA4 are shown. Dark blue indicates matching nucleotides. The percent identity is shown to the right of the alignments. 75% identity is the threshold for possible false binding (Kane *et al.*, 2000). C) Alignments of FBA4 Probe A and B onto full-length FBA3 is shown. Dark blue indicates matching nucleotides and the percent identity of the alignment is shown to the right.

The results of the *in-silico* analysis of the *T. oceanica* genome for transcription factors showed that specific transcription factors (Tf) count had similar quantities in comparison to *P. tricornutum* and *Thalassiosira pseudonana* (*T. pseudonana*) (Figure 3.10) (Rayko *et al*., 2010). The sum of Tfs and the distribution of different types of Tfs correlates well with the other two diatom species. CSF and CXC/tesmin are two transcription factors that were not identified in *T. oceanica* that are present in *P. tricornutum* and *T. pseudonana*. The long-term experiment included the analysis of 12 transcription factors, but only one transcription factor from the heat shock factor family showed a significant difference between high and low iron samples (Figure 3.10).

SIGMA70 is one of three transcription factors that is upregulated under iron limitation in *P. tricornutum* (Smith *et al.*, 2016). SIGMA70_*Pt* is chloroplast targeted and is thought to control the chloroplast genome (Smith *et al.*, 2016). One of the SIGMA70 Tfs in *T. oceanica* was upregulated under iron limitation in the LT experiment, but this upregulation only appeared in the dark phase. The immediate increase in transcript counts for the iron-recovery samples seems contrary to the overall downregulation. Overall, SIGMA70_*To* is targeted to the chloroplast, and changes in transcript counts were expected following iron induction. Overall, there were no significant changes between high- and low-iron samples, and the response to iron addition was therefore not very well defined.

In conclusion, seven genes showed a rapid decrease in transcript counts and two genes a rapid increase demonstrating a quick alteration of the cell's metabolism following iron induction, reversing adaptive strategies like chlorosis, enzyme replacements, and the upregulation of high-affinity iron-uptake proteins (Figure 3.14).

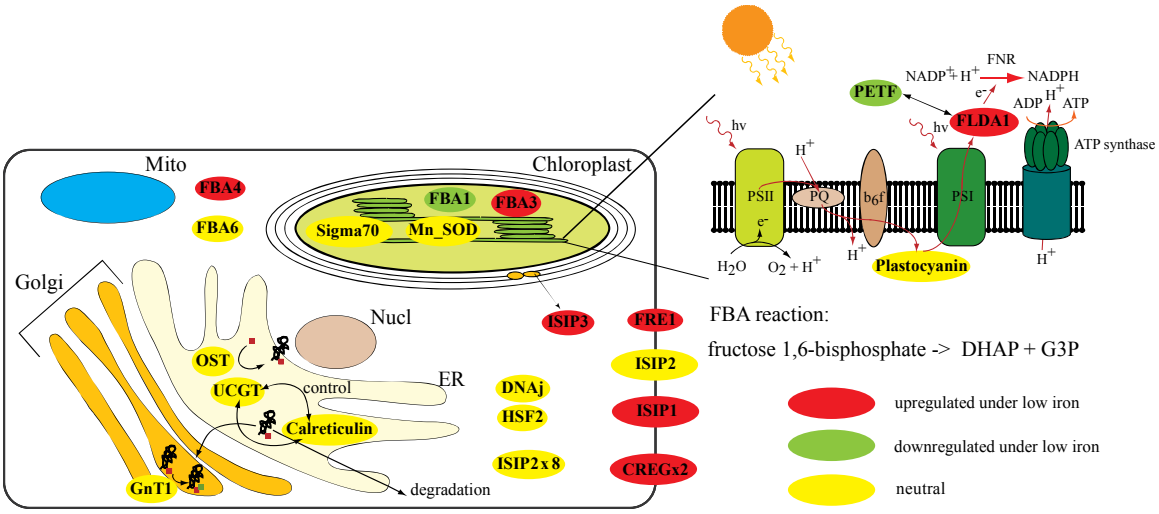


Figure 3.14: Cellular overview of the targeted transcripts and their localization in the cell.

The scheme of a *T. oceanica* cell and its compartments are shown: chloroplast (light-green), mitochondrion (Mito/blue), endoplasmic reticulum (ER/beige), Golgi apparatus (Golgi/orange), and the nucleus (Nucl/brown). The position of the gene name indicates the localization with respect to published data (ISIP1, ISIP2, ISIP3, FBA1, and FBA3) or putative location based on a prediction by TargetP (Emanuelson *et al.*, 2000) and ChloroP (Emanuelson *et al.*, 2000). The regulatory effect of iron is shown by the color surrounding the gene. Red circles show genes that are upregulated under iron limitation, green circles represent genes that are downregulated under iron limitation, and yellow circles are genes with no difference between the treatments.

The high-resolution temporal data demonstrate a stepwise iron-recovery process that started with the downregulation of ISIP1, ISIP3 as well as the iron reductase FRE1 that all adjusted to high-iron counts in under 30min. Based on localization studies on ISIP3 and a similar expression like ISIP1, ISIP3 might be involved in iron uptake/homeostasis in the chloroplast. It was shown previously that ISIP1 is involved in iron uptake towards the chloroplast, and therefore an interaction of ISIP1 and ISIP3 is possible. It appears that the first step of the iron response is the stop of excess uptake of iron into the cell and foremost into the chloroplast protecting the cell from ROS development. The second step of the iron-recovery response is the upregulation of PETF and FBA1 transcripts. These proteins are critical enzymes in the chloroplast and might improve the efficiency of the

chloroplast to produce glucose. This improvement can also be seen in the fast recovery of photo-physiological parameters that goes along the reinstating of replaced proteins. At the same time, but with slower speed, the non-iron containing enzymes FBA3, FBA4, and FLDA1 are downregulated. The analysis with actD revealed CREGx2, Mn_SOD, PETE, and PETF with a longer half-life and a possible stabilization of their mRNAs. Interestingly, the transcript half-life of CREGx2 in the iron-recovery samples is significantly shorter than in the actD samples. A secondary transcript product must control the half-life within the iron response. The opposite was observed for FLDA1 with a significantly longer half-life in the iron-recovery samples. The search for *T. oceanica* specific transcripts in the Tara Oceans database revealed a global distribution, demonstrating the global relevance of *T. oceanica* in the ocean (Figure 3.9). The proposed study describes the iron response of an open ocean diatom in a temporal high-resolution fashion. This knowledge will have implications on modeling efforts of anthropogenic or natural occurring iron induction events such as volcanic eruption (Rogan *et al.*, 2016), dust storms (Ratten *et al.*, 2015), or fertilizing icebergs (Duprat *et al.*, 2016).

CHAPTER 4 IDENTIFICATION OF N-LINKED GLYCOSYLATION SITES AND TARGETED TRANSCRIPTOMICS REVEAL DIURNAL EXPRESSION OF MEMBERS OF THE N-LINKED GLYCOSYLATION PATHWAY AND A HIGH PERCENTAGE OF NXT-TYPE GLYCOSYLATION SITES IN THE OPEN OCEAN DIATOM THALASSIOSIRA OCEANICA

Joerg Behnke, Alejandro Cohen, Julie LaRoche

Contribution of authors:

Joerg Behnke:	Planning and conducting the experiment, sample preparation and data analysis, drafting of the manuscript
Alejandro Cohen:	LC-MS/MS analysis of N-linked glycosylated peptides
Julie LaRoche:	Contributing to experimental design, planning and discussing the manuscript

This manuscript will be submitted to Frontiers in Plant Science.

4.0 Abstract

N-linked glycosylation is a posttranslational modification relevant for protein folding, protein function, cell-cell communication, and numerous other aspects. The process of N-linked glycosylation in algae is only marginally investigated, and further knowledge is needed to understand the cell biology of algae and the evolution of posttranslational N-linked glycosylation. The proposed study investigated the N-linked glycosylation pathway in the open ocean diatom *Thalassiosira oceanica*, demonstrating that all enzymes necessary for N-linked glycosylation are present. *T. oceanica's* ability to survive in very low iron concentrations led to the analysis of expression patterns under high- and low-iron concentrations for calreticulin, oligosaccharyltransferase (OST), N-acetylglucosaminyltransferase (GnT1), and UDP-glucose glucosyltransferase (UCGT). Employing targeted transcriptomics, these analyzed transcripts revealed a diurnal pattern, mostly independent of the iron status. The identification of 118 N-linked glycosylated peptides was based on the solid-phase extraction of N-linked glycosylated peptides (SPEG) method, analyzing samples retrieved from high- and low-iron conditions. The identified peptides had 86% NXT-type motifs, with X being any amino acids except proline. The presence of N-linked glycosylation sites in the iron starvation-induced protein 1a (ISIP1a) confirmed its predicted topology, contributing to the biochemical characterization of ISIP proteins in diatoms. An ocean-wide analysis using the Tara Oceans database showed a global abundance of calreticulin, OST, and UCGT, underlining the importance of glycosylation in microalgae.

4.1 Introduction

Post-translational modifications of proteins are often essential for proper protein function with N-linked glycosylation widely present in numerous proteins. N-linked glycosylation is primarily used for protein folding control in the endoplasmic reticulum with further functions based on the maturation of glycan structures in the Golgi apparatus. Whereas putative functions of N-linked glycosylation in algae are unknown, functions in plants are marginally established, and complete blockage of N-linked glycosylation leads to embryonic death demonstrating the need of N-linked glycosylation in plants (Lerouxel *et al.*, 2005; Nagashima *et al.*, 2018). The maturation of N-linked glycan structures differs based on specific enzymes in the Golgi apparatus, and while vertebrates show a high diversity of glycan structures, many of these enzymes present in vertebrates have not been identified in plants and microalgae (Baïet *et al.*, 2011; Strasser, 2016). The diversity of glycan structures in vertebrates is reflected by their diverse functions ranging from cell-cell communication or involvement in auto-immune diseases to inflammatory reactions (Moremen *et al.*, 2012). Based on these functions, it is not surprising that N-linked glycosylation is an important component for the functionality of biopharmaceuticals, where microalgae are seen as an attractive alternative for recombinant protein production because they are eukaryotic, photo-autotroph and fast-growing (Yusibov *et al.*, 2016). The first production of a functional monoclonal antibody in *Phaeodactylum tricornutum* was an important step towards a recombinant protein expression systems in diatoms for pharmaceutical usage (Hempel *et al.*, 2011; Hempel & Maier, 2012; Vanier *et al.*, 2015, 2018). *P. tricornutum* is capable of N-linked glycosylation resulting mainly in high-mannose glycans (Baïet *et al.*, 2011). The pathway

leading to matured N-linked glycosylation structures is a multi-step process that starts in the endoplasmic reticulum (ER) and continues throughout the Golgi apparatus. It starts with the formation of a lipid-linked oligosaccharide (LLO) by well-conserved asparagine-linked enzymes (ALG). An oligosaccharyltransferase (OST) attaches the LLO to an asparagine that is part of an NXS/T motif, with x being any amino acid except proline. The attached glycan is then trimmed in the ER before N-linked glycosylated proteins enter a folding control in the ER. The exact mechanism of the folding control is unknown, but the chaperones calreticulin and calnexin are involved in the folding process, and UDP-glucose glucosyltransferase (UCGT) glycosylates misfolded proteins, forcing a new folding control. Correctly folded proteins enter the Golgi apparatus, while misfolded proteins are eventually exported out of the ER and degraded in the proteasome. Once the glycosylated protein reaches the Golgi-apparatus, the glycan is first trimmed and then either modified by N-acetylglucosaminyltransferase I (GnT1) for further glycan maturation by a variety of enzymes (Lannoo & Damme, 2015) or enzymes alter the glycan structures in a GnT1-independent fashion. The proposed study investigated the difference of N-linked glycosylation in high- and low-iron concentration, aiming to identify the functional aspects of N-linked glycosylation and gain novel insight into the adaption of diatoms to chronic iron limitation.

Microalgae are gaining popularity for industrial purposes, and the successful expression of a monoclonal antibody showed their potential for recombinant protein production (Vanier *et al.*, 2014). The understanding of algal cell-biology is an urgent need to sustain and broaden industrial microalgae usage. The proposed study aimed to characterize glycosylation sites in the open ocean diatom *Thalassiosira oceanica*, reassuring

underlying mechanisms of glycosylation are similar to plants or mammals. Four critical proteins involved in the N-linked glycosylation pathway were analyzed in terms of differential expression under high- and low-iron conditions, diurnal expression, and their respective half-lives. The expression data was complemented with a blast search in the Tara Oceans database, analyzing the global abundance of these genes. Furthermore, the study was used to provide information about the topology of specific proteins as N-linked glycosylation is luminal. Overall, the findings give insight into a new field of microalgal cell biology.

4.2 Material and Methods

4.2.1 *In-silico* identification of enzymes involved in N-linked glycosylation

The *in-silico* search was performed on the *T. oceanica* proteome and the MMETSP database (Keeling *et al.*, 2014), using previously identified sequences from *Phaeodactylum tricornutum* (Baïet *et al.*, 2011). Further analysis of enzymes involved in glycan maturation in the Golgi apparatus included protein sequences retrieved from *Arabidopsis thaliana*, *Homo sapiens*, *Mus musculus,* and *Drosophila Melanogaster*. Identified proteins in *T. oceanica* were confirmed by a blast search against NCBI's non-redundant (nr) database (Supplemental Table 12).

4.2.2 Artificial seawater and culturing conditions

The exact protocol of the targeted transcriptome experiment is described in Chapter 3. Briefly, axenic *Thalassiosira oceanica* (CCMP1005) cultures were grown in ASW f/2 with a 14/10 light/dark cycle at 22°C and 100 µE light. ASW f/2 was prepared after Goldman et al. (Goldman *et al.*, 1978), using the Aquil metal mix (Price *et al.*, 1988) with

100 µM EDTA, 10 µM $FeCl_3$, without Ni^{2+}. Trace metal clean techniques were used at all times, and all equipment was trace metal clean. An Accuri C6 (Becton, Dickinson and Company (BD), Franklin Lakes, New Jersey, U.S.) was used to monitor cell growth. The iron status of the cell was assessed by measuring normalized variable fluorescence (F_v/F_m) with a Waltz Dual-PAM-100 (Heinz Walz GmbH, Effeltrich, Germany).

The cell culture for the SPEG-analysis was done separately with ASW f/2 following the recipe from Goldman et al. (Goldman *et al.*, 1978), using high purity salts (BioUltra). Batch cultures were regularly assessed for iron-limitation by measurements of normalized variable fluorescence (F_v/F_m). *T. oceanica* cultures grown in high-iron media were grown in the presence of 10 µM Fe, 10 µM EDTA, and 880 µM of $^{15}NO_3$ as the only nitrogen source.

4.2.3 Experimental setup

The detailed experimental protocol of the targeted transcriptome analysis is described in chapter 3. Briefly, *T. oceanica* growth was kept in the exponential phase, and sampling was done in low- to mid-exponential phase. The experiment included high-iron, low-iron, and iron-recovery treatments. High-iron cultures were grown in the presence of 10 µM $FeCL_3$, and iron-recovery samples received 10 µM $FeCL_3$ after the initial measurement (T=0). Additionally, samples were treated with actinomycin D and iron (actd-Fe), solely actinomycin D (actD), as well as DMSO with iron (DMSO-Fe), and DMSO without iron (DMSO). The initial sampling was done one hour after the start of the light period, and iron-addition followed the initial sampling. In all experiments, the 1 h timepoint describes the timepoint at 1 h following the addition of $FeCl_3$.

4.2.4 Actinomycin D and DMSO treatment

A final concentration of 10 µg/ml actD in DMSO was used for the actinomycin D treatment. The final DMSO concentration in the culture media was 0.1%. A pre-test was run on triplicate cultures to check for the effect of actD. F_v/F_m measurements were used as a proxy for possible Fe impurities and effectiveness of actD (data not shown). In parallel, samples with only DMSO (0.1% final concentration) were analyzed as a control in the pre-test as well as in the 6 h experiment.

4.2.5 RNA extraction and NanoString analysis

The exact protocol can be found in chapter 3. In brief, RNA concentrations were measured with a Nanodrop. Ninety ng of RNA was loaded for high-iron samples (including iron-recovery samples sampled later than 30 min after iron-addition), 70 ng for low-iron samples, and 80 ng for iron-recovery samples. Data was first processed in NSolver, and the house-keeping gene normalization was done in Excel using the nuclear-import-export protein transcript counts (THAOC_05312).

4.2.6 Half-life determination

The detailed protocol is provided in chapter 3. In short, the analysis of the decay rate was based on a broken-line model (Siegmund & Zhang, 1994). Furthermore, the decay rate calculation needed the calculation of the changepoint, which describes the point of the curve where it reaches its plateau. The changepoint was determined using the function "lm.br" in R, and the resulting slope (k) was used to calculate the half-lives ($t_{1/2}=\ln 2/k$).

4.2.7 Solid-phase extraction of N-linked glycopeptides

Sample volumes of 850 ml for low-iron cultures and 500 ml for high-iron samples were filtered onto 2 µm polycarbonate filters, rinsed off, and subsequently pelleted. A protocol by Tian et al (Tian *et al*., 2007) was followed for the capturing of the N-linked glycosylation motifs with minor modifications. Proteins were extracted in 6 M urea buffer with 6 min sonication. The bicinchoninic acid assay (BCA) was used to measure the protein concentration of the samples, and 500 µg of each sample was combined for further processing. The combined lysate was reduced for 60 min at 60°C in a final concentration of 10 mM DTT. It followed an alkylation step with a final concentration of 12 mM iodoacetamide for 30 min at room temperature (RT). The urea concentration was diluted under 2 M by the addition of 100 mM (pH8) potassium phosphate buffer, and twenty µg trypsin was added for overnight digestion. The sample was acidified (pH<3) with formic acid and washed in HLB cartridges. Columns were conditioned with 1x1 ml 50% acetonitrile (ACN) with 0.1% trifluoroacetic acid (TFA) and 2x1 ml 0.1% TFA. The sample was loaded, and the column was washed 5x with 0.1% TFA. Two elution steps followed with 600 µl 50% ACN – 0.1% TFA. The peptides were oxidized with a final concentration of 10 mM sodium periodate for 1h at 4 °C in the dark, followed by a dilution step to reach an ACN percentage of under 5% and acidified with formic acid (pH<3). A final column-wash followed as described above. Hydrazine beads (75 µl) were washed with 1ml water, and oxidized peptides were added and left overnight at RT. Beads were washed 3x with 1 ml water and 3x with 100 mM. The washed beads were transferred to a new tube and resuspended in 75 µl ammonium bicarbonate buffer. Three µl PNGase F was added to the beads and incubated for 3 h at 37 °C. Beads were spun

down and washed twice with the ammonium bicarbonate buffer. Washes and supernatant were combined and washed in HLB cartridge, as described above. Peptides were dried in a SpeedVac and stored at -20 °C for further processing.

4.2.8 Peptide identification

The samples were dried to a pellet in a vacuum centrifuge and subsequently resuspended in 20 µL of a 3% ACN, 0.5% formic acid solution. The samples were transferred to a 300 µL HPLC vial and subject to analysis by LC-MS/MS on a VelosPRO orbitrap mass spectrometer (ThermoFisher Scientific, Waltham, Massachusetts, USA) equipped with an UltiMate 3000 Nano-LC system (ThermoFisher Scientific, Waltham, Massachusetts, USA). Chromatographic separation of the digests was performed on PicoFRIT C18 self-packed 75 mm x 60 cm capillary column (New Objective, Woburn, Massachusetts, USA) at a flow rate of 300 nl/min. MS and MS/MS data were acquired using a data-dependent acquisition method in which a full scan was obtained at a resolution of 30,000, followed by ten consecutive MS/MS spectra in both higher-energy collisional dissociation (HCD) and collision-induced dissociation (CID) mode (normalized collision energy 36%). Internal calibration was performed using the ion signal of polysiloxane at m/z 445.120025 as a lock mass. Raw MS data were analyzed using Proteome Discoverer 2.2 (ThermoFisher Scientific, Waltham, Massachusetts, USA). Peak lists were searched against the *T. oceanica* protein (txid159749) database as well as the cRAP database of common contaminants (Global Proteome Machine Organization). Cysteine carbamidomethylation and all nitrogen atoms labeled as ^{15}N were set as a fixed modification, while asparagine to aspartic deamidation (N to D to account for PNGase F hydrolysis), methionine (Met) oxidation, N-terminal Met loss, and phosphorylation on

serine, threonine, and tyrosine were included as variable modifications. A mass accuracy tolerance of 5 ppm was used for precursor ions, while 0.02 Da for HCD fragmentation or 0.6 Da for CID fragmentation was used for productions. Percolator was used to determine confident peptide identifications using a 0.1% false discovery rate (FDR). Quantification of the peptides was performed by exporting the ProteomeDiscoverer's .msf to Skyline. Precursor ion quantification was done Skyline's MS1 filtering mode.

4.2.9 Tara Oceans database analysis

The Tara Oceans database was used to screen for the global abundance of OST, calreticulin, UCGT, and GnT1. The protein sequences were used in a blastp search against the Metagenome database with an e-value threshold of 10^{-50}. Surface water samples and deep chlorophyll max samples were analyzed. The data for the size fractions of 0.8-5 μm and 5-20 μm were analyzed and plotted.

4.2.10 *In-silico* analysis of N-linked glycosylated protein

N-linked glycosylated proteins were analyzed for the presence of transmembrane domains with TMHMM Server v. 2.0 (Krogh *et al*., 2001). Conserved domains were analyzed with the NCBI conserved domain search using the default settings (Marchler-Bauer *et al.*, 2017). N-linked glycosylation sites were confirmed by prediction with NetNGlyc 1.0 Server (Chuang *et al.*, 2012). Subcellular localization of the proteins were analyzed with TargetP 1.1 Server (Emanuelsson *et al.*, 2000) and the presence of a signal peptide with SignalP 4.1 Server (Petersen *et al.*, 2011). Chloroplast localization predicted by TargetP 1.1 Server were verified with ChloroP 1.1 Server (Emanuelsson *et al.*, 2000).

Besides a blast search for best hits the KEGG database (Kanehisa & Goto, 2000) was used for functional analysis.

4.3 Results

4.3.1 Identification of enzymes involved in N-linked glycosylation

The first step of N-linked glycosylation is the formation of a lipid-linked oligosaccharide (LLO) by asparagine-linked glycosylation enzymes (ALG). This process takes place at the membrane of the endoplasmic reticulum and involves several ALG enzymes. The *in-silico* search in *T. oceanica* resulted in 10 ALG enzymes and showed that only ALG10, responsible for the transfer of the last glucose onto the LLO, is missing (Figure 4.1). Throughout the formation of the LLO, a flippase turns the glycan structure from the outside of the endoplasmic reticulum to the inside. The finished glycan structure gets attached by OST to a specific amino acid motif that is usually NXT or NXS, with X being any amino acid except proline. (Figure 4.2).

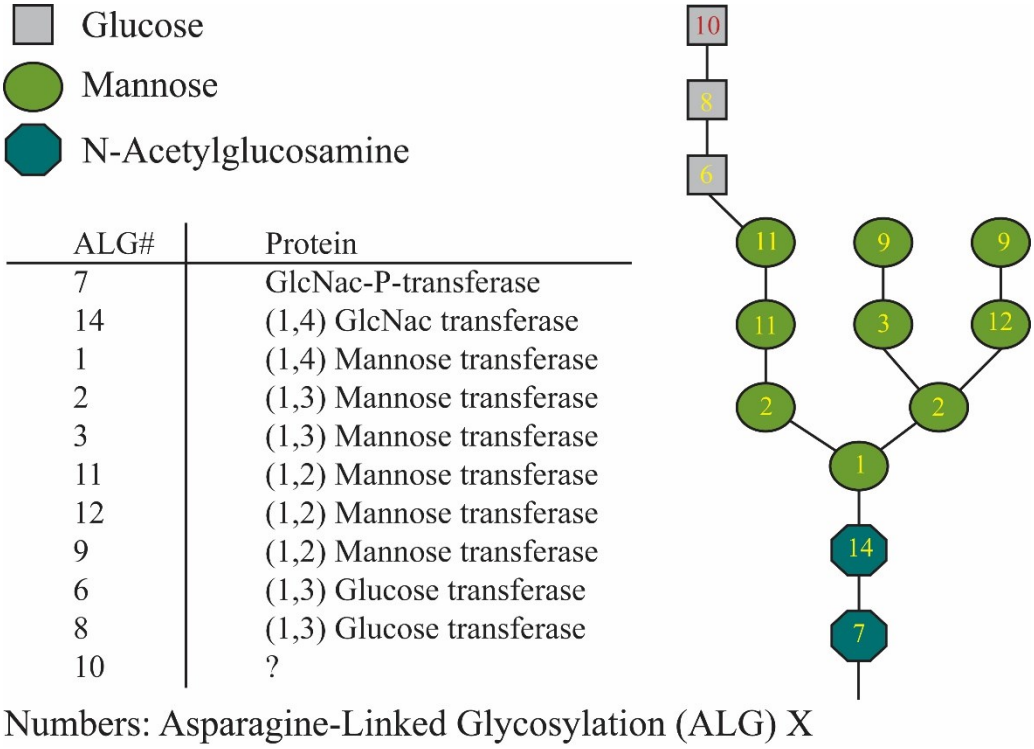


Figure 4.1: Overview of asparagine-linked glycosylation (ALG) enzymes and the lipid-linked glycan structure.

The figure shows the structure of a lipid-linked oligosaccharide (LLO) and the corresponding enzymes responsible for the attachment of the individual residues. The THAOC numbers in the table refer to identified enzymes in the *T. oceanica* proteome. The glycan structure shows the arrangement of the different sugar residues. Grey Squares are glucose residues, green spheres are mannose residues, and blue octagons are N-acetylglucosamines. The numbers in the structure indicate the corresponding ALG number. Numbers in red indicate missing enzymes.

Once the glycan is attached to the protein in the ER, glucose residues are removed by a glucosidase (GCII), and a chaperone-mediated protein-folding control within the ER follows. Calreticulin is a putative chaperone identified in the *T. oceanica* proteome, that is putatively involved in the folding control. The exact mechanism of the folding control is unknown, but UCGT reglycosylates incorrectly folded proteins, and only correctly folded proteins enter the Golgi apparatus. Here, the glycan structure gets trimmed by a mannosidase (α-ManI) and can be further modified by GnT1, leading to GnT1-specific

complex glycan structures (Figure 4.2). The high diversity of glycan structures within N-linked glycosylation is based on a wide variety of enzymes located in the Golgi apparatus. These enzymes are responsible for the attachment of various sugar residues to the glycan structure, often following the modification by GnT1. In addition to GnT1, *T. oceanica* possesses a fucosyltransferase ((1,3)-FucT) and a galactosylase (β1,4-GalT) that can alter the glycan structure in the Golgi apparatus.

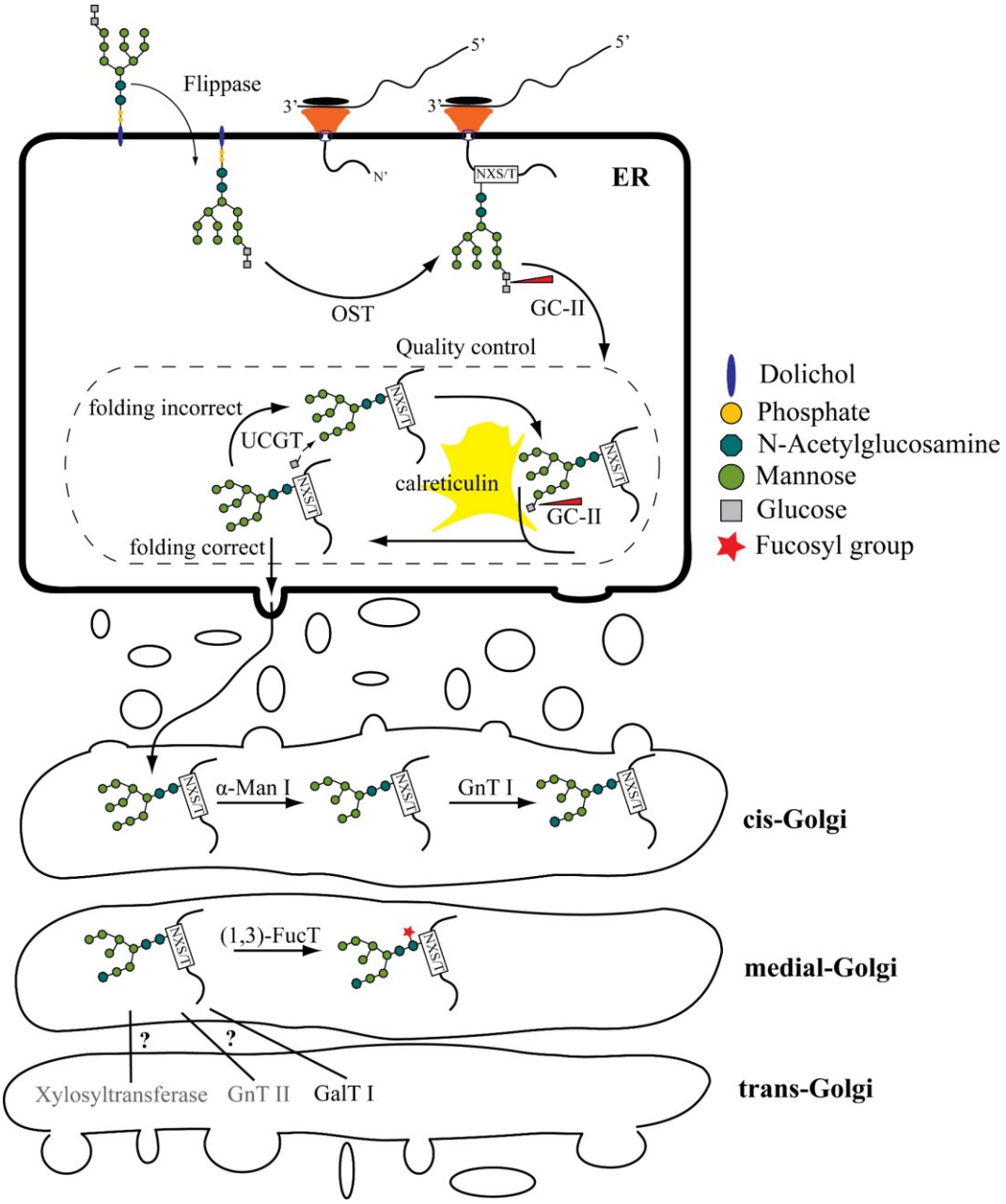


Figure 4.2: Simplified scheme of the N-linked glycosylation pathway in the endoplasmic reticulum and the Golgi apparatus.

The N-linked glycosylation pathway is shown in a simplified scheme. The mRNA is being attached and translated at the membrane of the endoplasmic reticulum, and the glycan is attached by the oligosaccharyltransferase (OST) and trimmed by α-glucosidase II (GC-II). The protein enters the folding control (dashed box). Calreticulin and UDP-glucose:glycoprotein glucosyltransferase (UCGT) are the main component of the system. Correctly folded proteins are further transported to the Golgi apparatus, and N-linked glycan structures are further processed by specific enzymes. The following abbreviations are used: alpha-mannosidase (α-Man), N-acetylglucosaminyltransferase (GnTI), 1,3

fucosyltransferase ((1,3)-FucT), galactosyltransferase (GalTI). In grey are two enzymes as examples of missing proteins that are involved in glycan maturation in other organisms. Components of the glycan structure are indicated with blue being dolichol, yellow is phosphate, grey squares are glucose residues, green spheres are mannose residues, and blue octagons are N-acetylglucosamines.

Table 1: Overview of proteins involved N-linked glycosylation.

The table gives a comparison of proteins in *P. tricornutum* (Baiet et al. 2011) and the corresponding proteins in *T. oceanica* that are involved in N-linked glycosylation. The following abbreviations are used: transmembrane domain (TM), *N*-acetylglucosaminyltransferase (GlcNAcT), mannosyltransferase (ManT), glucosyltransferase (GlcT), dolichol-phosphate –glucosyltransferase (P-Dol GlcT), dolichol-phosphate mannosyltransferase (P-Dol ManT), glucosidase II (Glc II), mannosidase I (Man I), *N*-acetylglucosaminyltransferase I (GnTI), mannosidase II (Man II), UDP-glucose:glycoproteinglucosyltransferase (UCGT).

Predicted Protein function	THAOC Nr	Length To	Length Pt	TM To	TM Pt	PFAM To	Pfam Pt	Protein Nr._Pt
(1,2)-ManT ALG11	THAOC_09775	673	433	2	4	PF15924, PF00534	PF00534	54621
(1,2)-ManT ALG12	THAOC_28372	361	581	6	6	PF03901, PF00153	PF03901	44425
(1,2)-ManT ALG9	THAOC_13160	637	556	7	10	PF03901	PF03901	44574
(1,3)-FucT	THAOC_00736	1273	481	0	1	PF00852	PF00852	54599
(1,3)-FucT	THAOC_14764	281	770	0	1	PF00852	PF00852	46109
(1,3)-FucT	?		718		1		PF00852	46110
(1,3)-GlcT ALG6	THAOC_34646	826	532	9	11	PF03155	PF03155	44117
(1,3)-GlcT ALG8	THAOC_03361	499	436	11	9	PF03155	PF03155	44905
(1,3)-ManT ALG2	THAOC_32241	572	503	0	1	PF15924, PF00534	PF00534	22554
(1,3)-ManT ALG3	THAOC_17486	571	414	8	9	PF05208	PF05208	10976
(1,4)-GlcNAcT ALG13	?		170		0		PF04101	9427
(1,4)-GlcNAcT ALG14	THAOC_25680	215	180	1	2	PF02245, PF08660	PF08660	14444
(1,4)-ManT ALG1	THAOC_33337	507	448	0	2	PF13439,PF13692	PF00534	14002
beta-1,4-galactosyltransferase 4	THAOC_18453	344	361	0	0	PF13733, PF02709	PF13733, PF02709	XP_002180427.1
calreticulin	THAOC_17167	399	421	0	0	PF00262	PF00262	41172
Flippase	MMETSP_09271-36046	564	451	8	6	PF04506	PF04506	XP_002177395.1
GlcII,subunit	THAOC_03472	797	712	0	0	PF12999, PF13015	PF01055	50836
GlcII,subunit	THAOC_03472	797	803	0	0	PF12999, PF13015	PF07915	54169

GlcNAc-P-transferase ALG7	THAOC_33316	415	440	6	9	PF00953	PF00953	9724
GnTI	THAOC_02312	529	444	1	1	PF03071	PF03071	54844
ManI	THAOC_08479	832	666	0	1	PF01532	PF01532	1815
ManII	?		1498		1		PF01074, PF09261, PF07748	52248
OST (STT3 subunit)	THAOC_04081	972	911	11	10	PF02516	PF02516	55197
OST (STT3 subunit)	THAOC_24768	1065	894	5	10	PF02516	PF02516	55198
P-Dol Man T (DPM1)	THAOC_00809	290	236	0	0	PF00535	PF00535	19705
P-DolGlcT ALG5	THAOC_13573	627	348	0	1	PF00535	PF00535	45980
UCGT	THAOC_35806	1737	499	0	0	PF18402, PF06427, PF18404	PF06427	54787

4.3.2 Targeted transcriptome analysis and half-life determination

The presence and expression of OST, calreticulin, UCGT, and GnT1 as key-enzymes in the N-linked glycosylation pathway was analyzed in a targeted transcriptome experiment, employing the NanoString technology. The conducted experiment included the analysis of transcript abundance under iron-replete and iron-deprived conditions, showing only temporary differences. The diurnal pattern for OST, calreticulin, and UCGT showed a maximum expression towards the end of the light phase and a steady downregulation during the night. Calreticulin, as well as OST, showed higher transcript counts for high-iron samples in the first 6 h, resulting in an overall significant upregulation under iron deprived conditions (Figure 4.3). The analysis of transcript abundance included the use of actinomycin D, a transcription inhibitor. The inhibition of gene transcription allowed the calculation of the respective half-lives, revealing half-lives of 4-8 min for all four transcripts. The analysis of the global abundance of the targeted transcripts in the Tara

Oceans dataset exhibited an overall higher abundance of calreticulin, OST and UCGT compared to GnT1 (Figure 4.5).

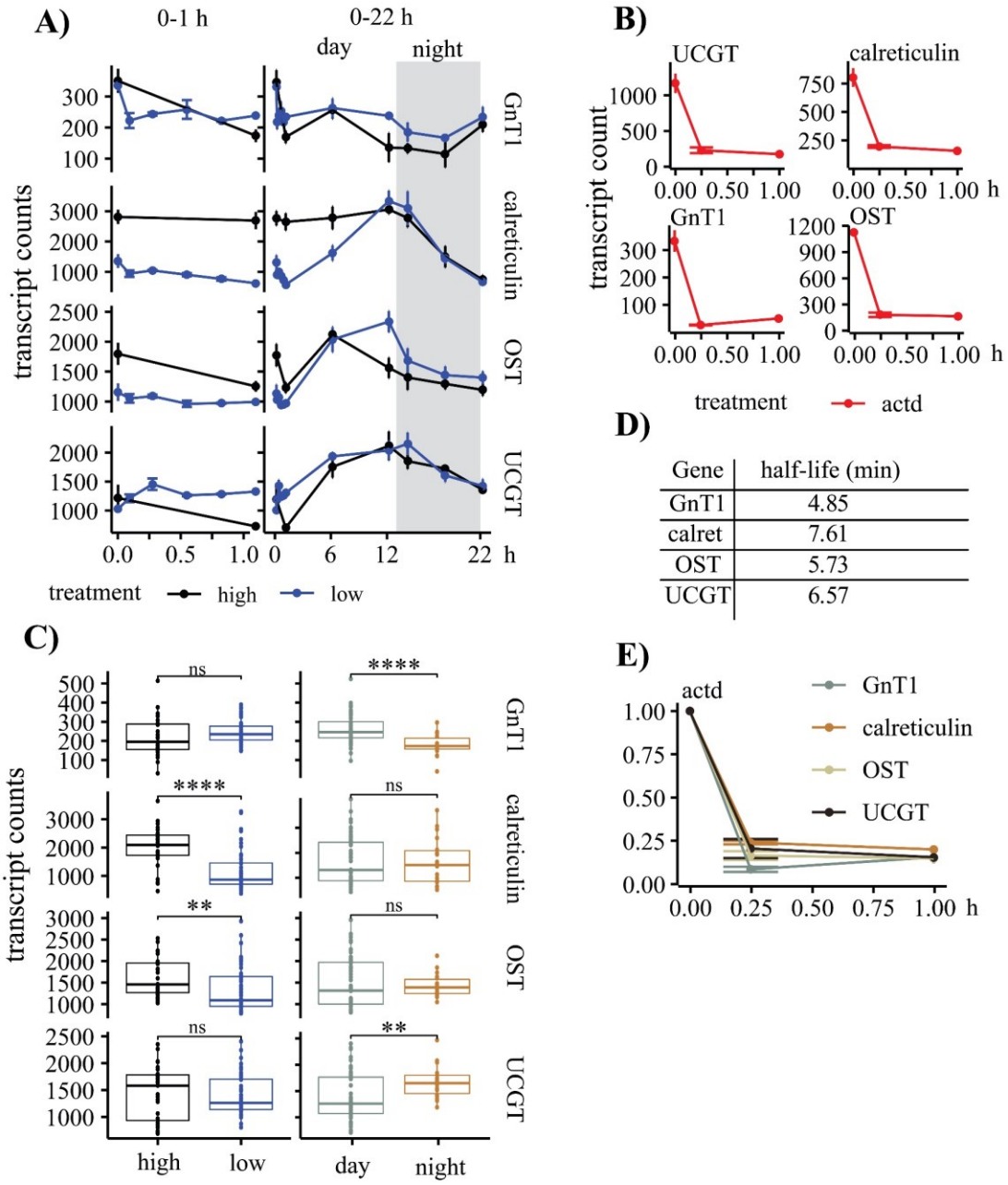


Figure 4.3: Analysis of transcript dynamics and half-lives of selected enzymes involved in the N-linked glycosylation pathway.

A) Transcript dynamics of N-acetylglucosaminyltransferase I (GnT1), calreticulin, oligosaccharyltransferase (OST), and UDP-glucose glucosyltransferase (UCGT). High-iron samples are shown in black, and low-iron samples are in blue. The left panel shows

transcript counts of the first hour, and the right side shows transcript counts over the 22 h period. The dark period is highlighted in grey. B) Transcript counts of the first hour after the samples when treated with actinomycin D (actD). C) Boxplots for each target analyzing statistical variance between high/low (left) and night/day (right) samples. Student's t-test was used as significance test. D) The normalized samples are shown below, with the highest equal to 1. GnT1 is grey, and calreticulin is orange, OST bright brown, and UCGT is dark brown. E) The table shows the half-life of the mRNA based on the slope in min.

4.3.3 Solid-phase extraction of N-linked glycopeptides

The analysis of putative glycosylation sites was done following a protocol from Tian et al (Tian *et al.*, 2007). Proteins were digested with trypsin, and glycosylated peptides were immobilized on hydrazine beads. Following purification, the peptides were released via PNGase F digest and identified via LC-MS/MS. The use of PNGase F allows only the analysis of non-1,3-fucosylated proteins. The resulting 118 identified peptides came from 115 proteins (Figure 4.4 and Supplemental Table 10). Besides the nitrate reductase, iron-starvation induced protein1a (ISIP1a), and a ferrichrome-binding protein (FBP), all other proteins were of unknown function, and putative functional characteristics were identified through blast and KEGG database searches. The analysis of the proteins through the KEGG database resulted in putative functions for 23 proteins. The identified functions were distributed throughout various cellular pathways reaching from glycan biosynthesis, genetic information processing to amino acid metabolism (Supplemental Table 11). Most of the motifs (102) occurred with an NXT type, motif, and only 18 peptides had an NXS motif, with x being any amino acid except proline. Within the 120 motifs, 16 putative bacterial glycosylation motifs of the D/E-X_1-N-X_2-S/T type were identified, and, interestingly, 32 of all motifs had a valine located either in the motif or directly in front or following the motif (Figure 4.4). Only one of the proteins with a bacterial glycosylation motif was previously identified with a bacterial origin, acquired

via lateral gene transfer from a cyanobacteria (Lommer *et al.*, 2012). Overall, 75 of the identified proteins had a predicted transmembrane domain or a signal peptide, and 33 proteins were predicted to go into the secretory pathway. Thirty proteins were undefined in terms of their location and would be considered cytosolic proteins. Previously identified N-linked glycosylated peptides of *C. reinhardtii* were used for comparison. Here, 90 N-linked glycosylated proteins were identified with 45 proteins predicted into the secretory pathway, 15 protein targeted to the chloroplast, 12 proteins into the mitochondrion, and 18 proteins were undefined. Both analyses were conducted with iron-replete and iron deprived cells. The low iron samples of the proposed study on *T. oceanica* were grown in presence of ^{15}N nitrate as sole N-source. This approach led to the detection of 52 peptides in both conditions, 38 in high-iron and 29 peptides in low-iron samples (Figure 4.4). The study on *C. reinhardtii* revealed 21 peptides in the low iron samples and only 13 peptides were identified solely in the low-iron samples (Mathieu-Rivet *et al.*, 2013).

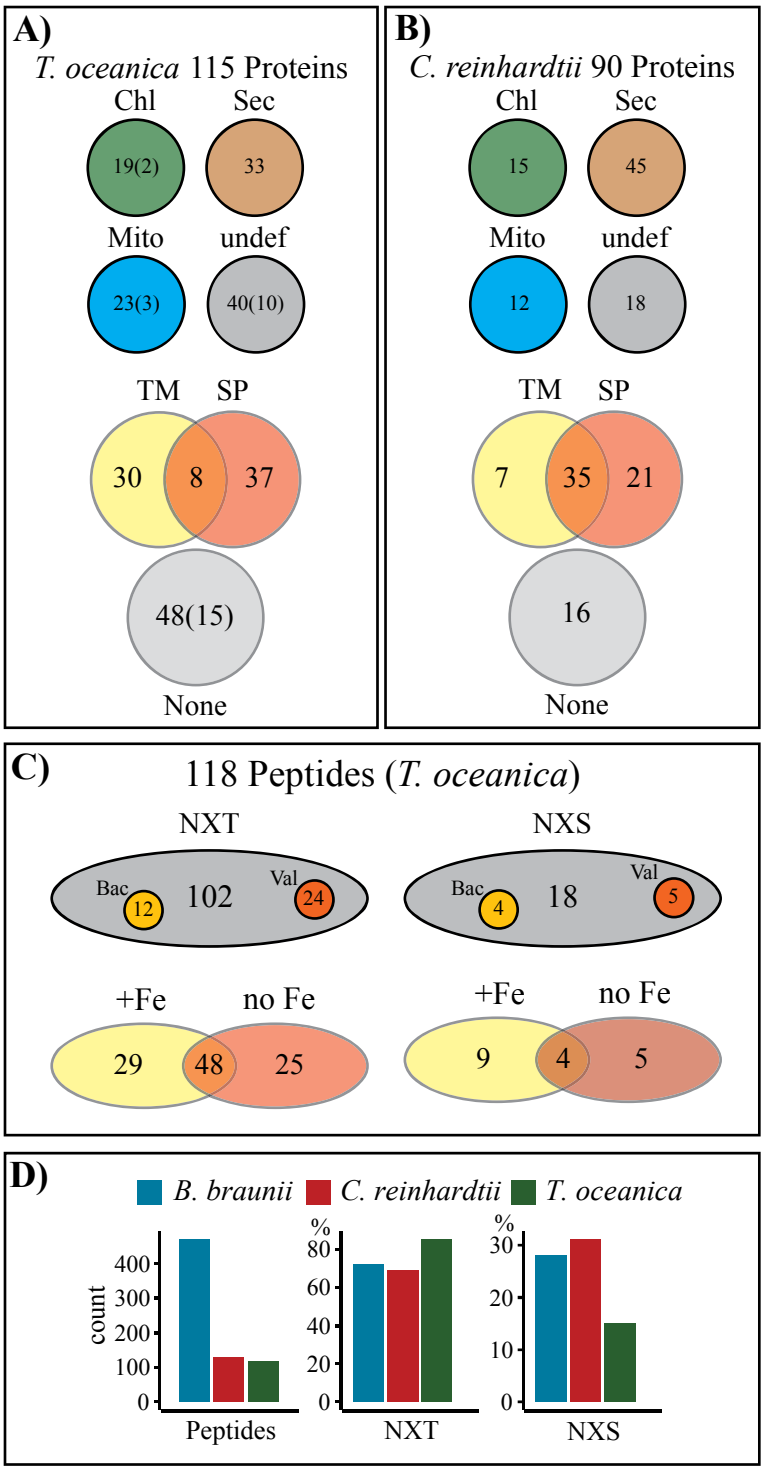


Figure 4.4: Overview of N-linked glycosylated proteins and peptides.

A) Overview of the 115 identified proteins in *T. oceanica*. The proteins are grouped into their respective predicted cellular localization (chloroplast (Chl); secretory pathway (Sec); mitochondria (Mito); undefined (undef)), predicted with TargetP 1.1 server (Emanuelsson

et al., 2005). Furthermore, the proteins are grouped by the presence of transmembrane domains (TM) predicted with TMHMM Server 2.0 (Krogh *et al*., 2001), and signal peptides (SP) predicted by SignalP 4.0 Server (Emanuelsson *et al.*, 2005). The amount of incomplete sequences is shown in brackets. B) Overview of glycosylated proteins identified in *C. reinhardtii* (Mathieu-Rivet *et al.*, 2013). The 90 proteins are grouped similar to A), based on the same server analysis for the respective predictions. C) The 118 peptides identified in *T. oceanica* are shown in terms of the motif type (NXT or NXS) with one letter code for the amino acid and x being any amino acid except proline. Small circles within the grey circles show the number of motifs that could be putatively bacterial (orange) or that have a valine (red). The numbers of motifs found in high-iron treatment (yellow), low-iron treatment (brown), or both treatments (overlapping area) are shown. D) Comparison with peptides found in *B. braunii* (Schulze *et al.*, 2017) and *C. reinhardtii* (Mathieu-Rivet *et al.*, 2013). Shown are the total motifs found in each study and the percent of NXT and NXS motifs.

4.4 Discussion

The focus of the proposed study was the N-linked glycosylation pathway in the context of iron-replete and iron-limited conditions, as well as N-linked glycosylation motifs and the identification of members of the pathway in general, supporting the understanding of N-linked glycosylation in diatoms. Furthermore, this work contributes to research on microalgae as a promising alternative to bacteria, mammalian cell lines, or yeast for the production of recombinant proteins based on their cheap and easy growth as well as their eukaryotic character (Yan *et al.*, 2016; Taunt *et al.*, 2018). Posttranslational modifications ,such as N-linked glycosylation, are an important aspect in the production of recombinant proteins as it is often essential for the proper function of the produced proteins (Moremen *et al.*, 2012). The understanding of N-linked glycosylation in microalgae is still in its infancy and research has been done on a few microalgae including *C. reinhardtii* (Mamedov & Yusibov, 2011; Mathieu-Rivet *et al.*, 2017)*, P. tricornutum* (Baïet *et al.*, 2011)*, Porphyridium sp* (Levy-Ontman *et al.*, 2011, 2014)*, B. braunii* (Schulze *et al.*, 2017) and *Chlorella vulgaris* (Mócsai *et al.*, 2019). These studies focused mainly on

glycan structures, indicating a high diversity of glycan structures within N-linked glycosylated proteins.

All enzymes needed for the production of the N-linked glycan precursor were identified in *T. oceanica*. Two enzymes, ALG10 and GC-I, were not identified, confirming previous data that showed ALG10 absent in diatoms and GC-I not being present in *P. tricornutum* (Baïet *et al.*, 2011; Levy-Ontman *et al.*, 2014). Interestingly, ALG10 is responsible for the attachment of the outer glucose residue, and GC-I removes this glucose residue (Figure 4.1). Following the attachment of the glycan in the ER, a protein folding control is mediated by chaperone proteins. Calreticulin was identified as a chaperone protein most likely involved in the folding control in *T. oceanica,* as well as UCGT, which attaches a glucose residue to misfolded proteins (Cherepanova *et al.*, 2016). Enzymes involved in glycan maturation throughout the Golgi apparatus were also identified (Table 1).

The targeted transcriptome approach was used to verify the presence and expression of some of the key-enzymes in the N-linked glycosylation pathway. These enzymes included the catalytic domain Staurosporine and temperature-sensitive 3 (STT3) as a subdomain of OST (Mohorko *et al.*, 2011), calreticulin, UCGT, and GnT1 as a key enzyme for complex glycan maturation in the Golgi apparatus. The gene expression patterns of OST, UCGT, and calreticulin were similar to an increase throughout the day, which could relate to an increased expression of secreted proteins and transmembrane transporters. In *P. tricornutum*, similar expression patterns were shown for proteins of the cytosolic glycolysis and the citric cycle (Smith *et al.*, 2016). Both pathways could have N-linked glycosylated members, or the upregulation of these pathways increases the

demand for uptake proteins to supply macro and micronutrients and therefore increasing the flow-through of proteins in the ER.

The transcription pattern of GnT1 showed a mid-day maximum and mid-night minimum with only minor changes throughout the 22 h period. Most conserved sites of GnT1, as discussed for *P. tricornutum* (Baïet *et al.*, 2011), are present in *T. oceanica* (Supplemental Figure 6). Transcript counts of GnT1 were less when compared to the other three transcripts, indicating a possible minor activity of GnT1 in *T. oceanica*. Whereas a knockout of GnT1 in mice was embryonic lethal (Ioffe & Stanley, 1994), the importance of GnT1 in plants was only visible under stress conditions (Strasser *et al.*, 2005). It is not known if GnT1 activity is important for microalgae, and GnT1-dependent glycan structures that were identified in *P. tricornutum* (Baïet et al., 2011) were also found in the non-GnT1 harboring *C. reinhardtii* (Mathieu-Rivet et al., 2013), questioning the activity of GnT1 in *P. tricornutum*. *In-silico* analysis showed the presence of GnT1 in numerous algae (Vanier et al., 2014), but GnT1-dependant glycan structures were so far only identified in the green algae *Botryococcus braunii* (Schulze et al., 2017). Analysis of qPCR data verified the activity of GnT1 in *P. tricornutum* (Baïet et al., 2011), and the targeted transcriptome analysis in the study presented here showed GnT1 gene expression in *T. oceanica*, although in low abundance. Whether GnT1 is active in *T. oceanica* and leads to GnT1-dependent glycan structures still needs to be shown. The search for gene abundance in the Tara Oceans dataset showed a lower abundance of GnT1 genes compared to OST, UCGT, and calreticulin (Figure 4.5), supporting the hypothesis of a restricted abundance of GnT1 in microalgae. The targeted transcriptome analysis in the proposed study revealed diurnal patterns that were overlaying an effect of iron, but it

showed that the identified genes are actually expressed, and the similarity in expression supports their functional connection. The half-lives of all four mRNAs were very similar ranging from 4.85 min for GnT1 to 7.61 min for calreticulin (Figure 4.3).

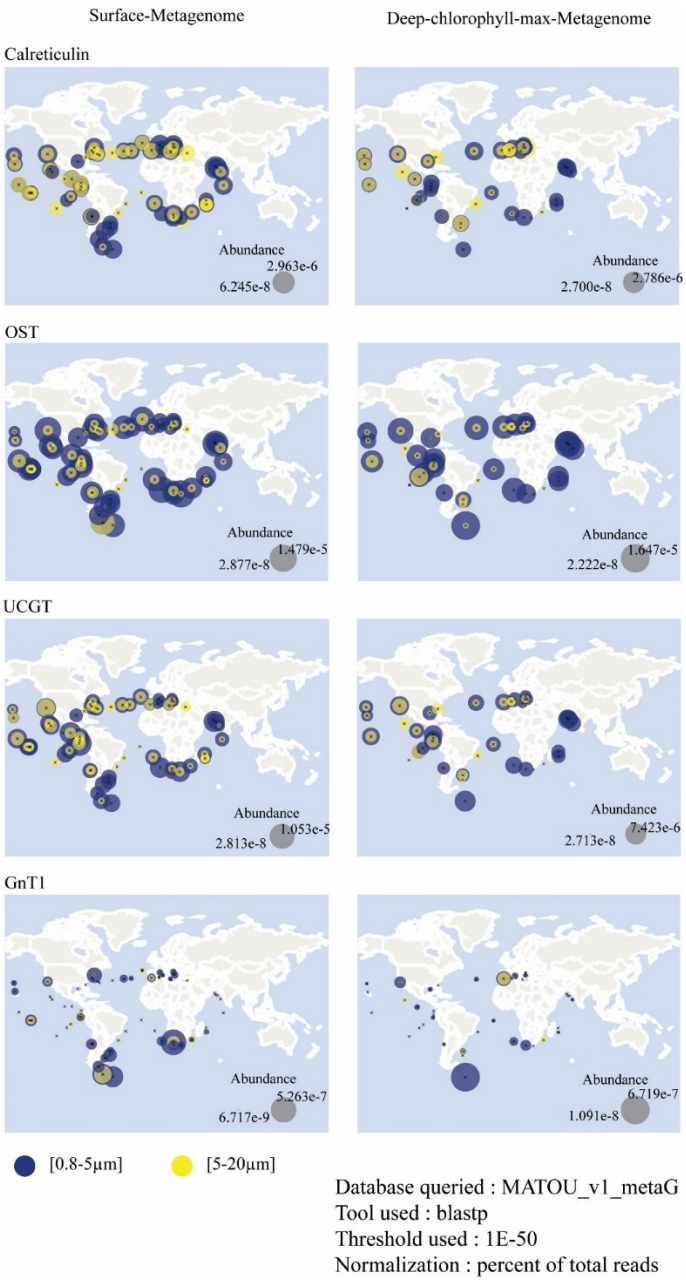


Figure 4.5: Global distribution of members of the N-linked glycosylation pathway.

The global distribution of calreticulin, oligosaccharyltransferase (OST), UDP-glucose glucosyltransferase (UCGT), and N-acetylglucosaminyltransferase (GnT1) was analyzed

using a blastp (e-value threshold of 1E-50) search in the Tara Oceans Metagenome database (Villar et al. 2018). Two size fractions are shown, 0.8-5 µm (blue) and 5-20 µm (yellow). The size of the circle represents the percentage of total reads. The left panel shows surface samples, and the right panel shows deep-chlorophyll max samples.

The analysis of the SPEG experiment resulted in 118 peptides from 115 proteins (Figure 4.4). The majority of motifs that were identified were from NXT-type (102) motif, and only 18 peptides belong to the NXS-type. These proportions are very similar for *B. braunii* and *C. reinhardtii* (Figure 4.4). Further, 16 motifs had a putatively bacterial N-linked glycosylation motif, following D/E-X_1-N-X_2-S/T-type (Kowarik *et al.*, 2006), but no bacterial OST was identified in *T. oceanica in-silico*. All motifs except one were predicted by NetNGlyc 1.0 Server with high confidence (Blom *et al.*, 2004). Interestingly, 30 peptides had a valine either within the motif itself or following the motif, which could be of interest for future identification of glycosylation motifs in diatoms. Valine is part of N-linked glycosylation sites in insects where a NXV motif has been reported (Scheys *et al.*, 2018).

The analysis of cell extract from iron-replete and iron-deprived conditions resulted in the identification of glycosylation sites in important proteins involved in the low-iron adaptation. Within this dataset, two N-linked glycosylation sites in ISIP1a were confirmed. ISIP1a is involved in endocytosis-mediated iron uptake via siderophores in *P. tricornutum* (Kazamia *et al.*, 2018). The direct function is unknown as it does not bind the siderophore, but initiates the endocytosis of the iron-siderophore complex (Kazamia *et al.*, 2018). The predicted conformation shows an extracellular N-terminal domain, which was confirmed through the detection of glycosylation sites in the N-terminal domain in the study presented here. Two different putative ferrichrome-binding proteins (FBP) (THAOC_08758 and THAOC_28875) were identified that might play an

important role in the low-iron adaptation. The identification of N-linked glycosylation peptides under iron-deprived conditions demonstrates the presence of FBP proteins in *T. oceanica* under iron limitation. Previous transcriptomic data supports its sole expression under iron-deprived conditions (Lommer *et al.*, 2012). Both transporters have a signal peptide leading them to the endoplasmic reticulum, but only THAOC_28875 possesses a transmembrane domain, anchoring it into a membrane. This protein is a promising candidate for the binding of the iron-siderophore complex that is being internalized via ISIP1a. Furthermore, a putative ferredoxin dependent nitrite-reductase was identified. This peptide was upregulated under high-iron conditions, which confirms published data that demonstrated an upregulation of ferredoxin-dependent enzymes in the genera *Thalasssiosira* after the addition of iron in environmental samples (Cohen *et al*., 2017). Besides the ferredoxin-dependent nitrite reductase, a peptide of a carbonic anhydrase was identified in high- and low-iron samples, showing an upregulation under low-iron conditions, and confirming previous result on N-linked glycosylation in carbonic anhydrases (Lehtimäki *et al.*, 2015). The UCGT enzyme was also identified and showed a slight upregulation under high-iron conditions.

The identified peptides were analyzed with TargetP in terms of their subcellular location (Emanuelsson *et al.*, 2000). Incomplete protein sequences must be ignored, as the signaling peptide for subcellular sorting is located in the N-terminal region of the protein. It was expected that most proteins would be targeted towards the secretory pathway, but only 33 proteins were positive for the secretory pathway, while 17 and 20 proteins were targeted towards the chloroplast or the mitochondrion, respectively (Figure 4.4). Glycosylated proteins are described for both compartments, and it is defined as an

alternative route for some plastid proteins to go through the ER (Kung *et al.*, 2009; Hempel *et al.*, 2013). N-linked glycosylation of chloroplast targeted proteins such as the carbonic anhydrase (Lehtimäki *et al.*, 2015) and a plastidial pyrophosphatase (Nanjo *et al.*, 2006) has been shown. A study on N-glycosylated proteins in *C. reinhardtii* showed 15 out of 90 proteins targeted to the chloroplast, and 12 proteins are targeted to the mitochondrion. In *C. reinhardtii*, 50% of the proteins were predicted to go into the secretory pathway (Figure 4.4). Both datasets show that N-linked glycosylation might play an important role for proteins in different cellular compartments, demonstrating a complex function of N-linked glycosylation in microalgae.

In conclusion, the study presented here gives insight into N-linked glycosylation in *T. oceanica,* identifying enzymes involved in the glycosylation process, showing expression patterns of key-enzymes under high- and low-iron conditions, and revealing N-linked glycosylated peptides. In general, the analysis of N-linked glycan structures identified in microalgae shows a high diversity based on species-specific glycan structures. It is still very early to speculate, but besides the importance of N-linked glycosylation for the folding control of proteins of different compartments, N-linked glycosylation in microalgae could also play a role in intra/inter-species signaling and communication in terms of algae-algae or algae-bacteria relationships (Amin *et al.*, 2012). The importance of N-linked glycosylation for cell-cell communication in multicellular eukaryotes and the species-specific glycan structures found in microalgae point towards the role of N-linked glycosylation that goes beyond the folding control. Further genomic sequencing of algal species, analysis of glycan structures, and functional characterization of specific proteins

involved in N-linked glycosylation are needed to gain insight into the evolution and function of N-linked glycosylation in the ocean.

CHAPTER 5 SUCCESSFUL GENETIC TRANSFORMATION VIA MICROPROJECTILE BOMBARDMENT OF THE LOW-IRON ADAPTED OPEN OCEAN DIATOM THALASSIOSIRA OCEANICA LEADS TO ANTIBIOTIC RESISTANCE AND EXPRESSION OF THE GENE-OF-INTEREST

Joerg Behnke, Julie LaRoche

Contribution of authors:

Joerg Behnke:	Planning and conducting experiments, troubleshooting, drafting of the manuscript
Julie LaRoche:	Troubleshooting, discussing the manuscript

5.0 Abstract

Transgenic techniques are well established for various model organisms leading to advanced knowledge of biochemical processes and often used for industrial and commercial use. The rising interest in diatom research, based on their biogeochemical importance and their diverse industrial use, led to the establishment of several transformation systems for specific diatom species. Most research on biochemistry and cell biology in diatoms involving genetic modification have been done with *Phaeodactylum tricornutum,* a coastal diatom species with a high growth rate, and high tolerance towards numerous stressors. The high diversity of diatoms in terms of habitat adaptations, as well as high genomic variability between species, justifies the development of further transformation systems. This study develops a transformation system for the open ocean diatom *Thalassiosira oceanica*. The first step was to design a pDEST vector with *T. oceanica*-specific regulatory elements. This vector was successfully used via particle bombardment, verified by the expression of the resistance gene, leading to growth in nourseothricin-containing media. This work is an essential foundation for the successful transformation of a globally and environmentally relevant open ocean diatom.

5.1 Introduction

Diatoms are one of the major groups of microalgae. Their biogeochemical importance in the global ocean (Assmy *et al.*, 2013) impacts the marine food web (Tréguer *et al.*, 2018) and the global carbon cycle (Smetacek, 2018), demonstrating the need for a detailed understanding of diatom cell biology. Moreover, microalgae are gaining popularity for

industrial purposes, including the production of food supplements, biofuel, and biopharmaceuticals (Vanier *et al.*, 2018).

The first algal transformation system was developed for *Chlamydomonas reinhardtii* in the mid-1980s (Rochaix & Van Dillewijn, 1982; Kindle *et al.*, 1989). One decade later, a transformation system for diatoms followed (Dunahay & Jarvis, 1995), which was the starting point for the development of numerous transformation systems of different diatom species. Research on diatoms using genetic transformation has been focused on *P. tricornutum* and *Thalassiosira pseudonana*, contributing significantly towards the understanding of fundamental aspects of diatom cell biology (Allen *et al.*, 2012; Samukawa *et al.*, 2014; Kazamia *et al.*, 2018; McQuaid *et al.*, 2018). For *P. tricornutum* and *T. pseudonana,* transformation assays were used to silence genes (De Riso *et al.*, 2009) and popular systems such as the TALEN system (Daboussi *et al.*, 2014) and the CRISPR/Cas9 system (Hopes *et al.*, 2016; Nymark *et al.*, 2016; Stukenberg *et al.*, 2018) have been successfully applied. These techniques allow precise editing of DNA and the knock-out of targeted genes. However, the first step of any gene-editing technique is the successful delivery of DNA into the cell, and as such, microparticle bombardment is the most widely used technique to transform diatoms (Doron *et al.*, 2016; Huang & Daboussi, 2017). More recently, bacterial transformation (Karas *et al.*, 2015) and electroporation (Niu *et al.*, 2012) have been developed for *P. tricornutum* with improvements in transformation efficiency and the capability of introducing long pieces of DNA, so-called episomes.

P. tricornutum and *T. pseudonana* are well-established model organisms, but both species are restricted to coastal areas. The Tara Oceans database, covering coastal and open

ocean surface water samples globally (Carradec *et al.*, 2018), did not detect *P. tricornutum* in the dataset, which resembles the minor environmental relevance of *P. tricornutum*. Not only does the high diversity of diatoms require further transformation systems, but also physiological and genetic differences between coastal and open ocean diatom species (Strzepek & Harrison, 2004; Strzepek *et al.*, 2011) demonstrate that the catalog of model organisms in diatom research needs to be expanded.

The open ocean diatom *Thalassiosira oceanica* is well adapted to open ocean environments and can withstand chronic iron limitation (Lommer *et al.*, 2012). *T. oceanica* sequences have been recovered from global metagenomics surveys (Chappell *et al*., 2015), and it belongs to the globally distributed genus *Thalassiosira* (Malviya *et al.*, 2016). Expression patterns of specific genes in *T. oceanica* have been used to assess the iron status in the north-east Pacific (Chappell *et al.*, 2015), demonstrating the environmental importance of this species. *T. oceanica* is also found in the Tara Oceans dataset, supporting its global distribution and importance (Chapter 3). A successful transformation system for *T. oceanica* is therefore of great interest, and it would allow the verification of previous results from *P. tricornutum* and *T. pseudonana*. This study presents the development of a transformation system for *T. oceanica*, including the design of a pDEST vector with *T. oceanica* native promoters and terminators, and the successful delivery of the vector into *T. oceanica* using microparticle bombardment. The analysis of the first transformants showed that this study is a good foundation for the further development of a transformation system in *T. oceanica*.

5.2 Material and Methods

5.2.1 Cell culturing

Axenic *T. oceanica* cells were grown in artificial seawater (ASW f/2) (Goldman *et al.*, 1978) and kept at 20°C with 100 µE light and a 14/10 h light/dark cycle. The cells were grown in batch culture, and cell counts were monitored with a flow cytometer as described in chapter 3. For plating, cells were filtered onto 2 µm polycarbonate (PC) filters and further concentrated by centrifugation at 3000 g for 3 min. The resulting pellet was resuspended in a small volume of ASW to achieve 200 µl with $1x10^8$ cells per plate.

5.2.2 RNA extraction and cDNA preparation

DNA and RNA from *T. oceanica* were extracted using the Qiagen AllPrep DNA/RNA kit (Qiagen, Inc., Valencia, CA, USA) following the manufacture's protocol, using 450 µl RLT buffer for lysis with previous sonification for cell disruption. The use of RNase free DNase (Qiagen, Inc., Valencia, CA, USA) via on-column digest served to ensure the removal of DNA contamination. cDNA was generated with Superscript III (Invitrogen, Carlsbad, CA, USA) following manufactures protocol with oligo(dt) primers and maximum RNA volume.

5.2.3 Vector design

The pDEST vector was received from Andrew Allen (Scripps Institution, LaJolla, CA, USA) and *T. oceanica* specific regulatory elements, including promoters and terminators, were introduced into the vector via Gibson Assembly® cloning technique (New England Biolabs, Ipswich, Massachusetts, USA) (Supplemental Table 15). The NEBuilder Assembly Tool® (New England Biolabs, Ipswich, Massachusetts, USA) was used to

generate Gibson primers for specific reactions. The assembly was divided into three separate reactions to obtain a better assembly efficiency. The scheme is shown in Figure 5.1, and primer sequences are listed in Supplemental Table 15. The first step included the exchange of promoter and terminator of the resistance gene. In the second step, the resistance gene was replaced, and the nourseothricin acetyltransferase (NAT) was introduced into the vector. The third and final step included the introduction of promoter and terminator sequences controlling the transcription of the gene cassette. After each step, the vector was sequenced to verify the sequence and direction of the introduced DNA. The gene of interest was introduced into the vector using the Gateway® system (Invitrogen, Carlsbad, CA, USA). Gene of interests were first amplified via PCR from cDNA using the Phusion® High-fidelity polymerase (Invitrogen, Carlsbad, CA, USA) and a second PCR step was used to add a 5' CACC overhang upstream of the gene to modify the gene for introduction into the pENTR™ Directional TOPO® vector using the pENTR™ Directional TOPO® Cloning Kit (Figure 5.1).

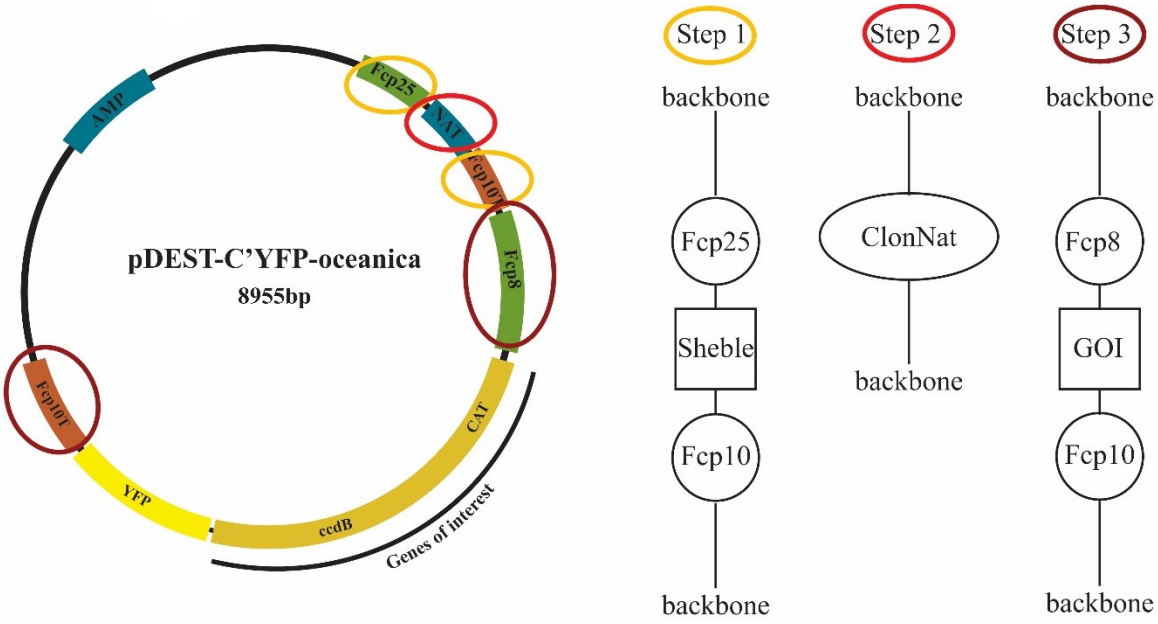


Figure 5.1: Gibson Assembly® cloning scheme for the introduction of *T. oceanica* specific promoter and terminator sequences into the pDEST vector, and the exchange of the resistance gene.

The Gibson Assembly® cloning system was used to introduce *T. oceanica* specific promoters and terminators into the pDEST vector. The reaction was separated into three steps. Step1 was the exchange of promoter and terminator of the resistance gene. The second step was the exchange of the resistance gene itself, and the third step was the exchange of promoter and terminator of the gene of interest. The following abbreviations are used: ampicillin (AMP), fucoxanthin-chlorophyll binding protein (FCP), nourseothricin N-acetyl transferase (NAT), chloramphenicol acetyltransferase gene (CAT), control of cell death B (ccdB), yellow fluorescence protein (YFP).

5.2.4 Transformation of *T. oceanica* cells using a biolistic transformation approach

Cells were grown to mid-late exponential phase, and $1x10^8$ cells were concentrated via filtering (2 µm PC filters) with subsequent centrifugation (3 min at 2000 g) and placed in the middle of an ASW plate (1.2% Agarose). After 24 h, cells were bombarded using 0.7 µm (M10) tungsten beads (Bio-Rad, Hercules, CA, USA), coated with up to 20 mg of vector DNA with a pressure of 900 psi using a PDS-1000/He system (Bio-Rad, Hercules, CA, USA). Cells were transferred immediately into ASW f/2 media without antibiotics and kept in the media for 48 h with regular light/dark cycle. After the recovery time, the cells were transferred to selective ASW f/2 with 50 µg/ml nourseothricin (Figure 5.2). Antibiotic-resistant cultures were plated (50 µg/ml nourseothricin) using the pour-plating technique to achieve individual colonies (Figure 5.2).

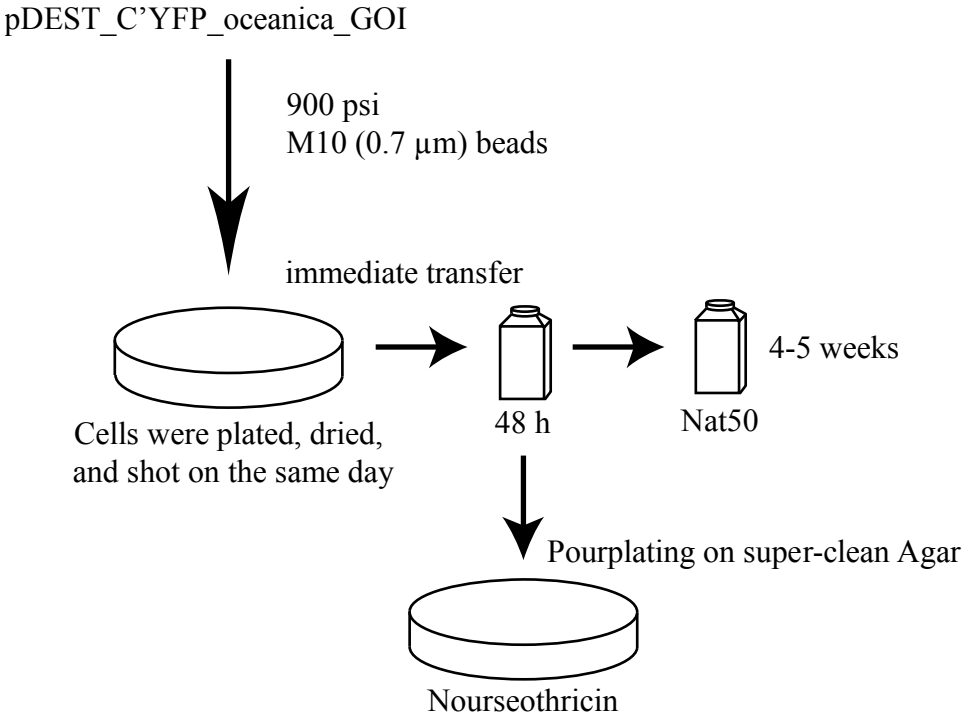


Figure 5.2: Transformation scheme for *T. oceanica*.

Cells were transformed via microparticle bombardment. First, the cells were plated, dried, and transformed on the same day. After the cells were bombarded, they immediately got transferred into media without antibiotics for a 48 h recovery time. The cells were then transferred into selective media or used for pour plating.

5.2.5 Pour-plating

Difco™ Agar from Becton Dickinson (BD, Franklin Lakes, New Jersey, USA) was further cleaned before use. One hundred g of agar was washed by stirring with 3 L of MilliQ for 30 min. The agar was filtered onto Whatman F4 filter paper in a Buechner funnel. This procedure was repeated until the filtrate was clear. The agar was then washed with 3 L of 95% ethanol, followed by a final 3 L wash with analytical grade acetone. The agar was then dried at 50°C in glass baking dishes for 2–3 days and stored tightly sealed at room temperature. For pour-plating, ASW, vitamins, macro-nutrients, and the metal mix were autoclaved separately and kept at 50°C. Cleaned agar was

autoclaved in Milli-Q water with a final concentration of 0.32% agar in the plate. Reagents were mixed in a 50 ml tube, and the antibiotic was added in various concentrations. The cells were added once the ASW in the tube reached 32°C and plated immediately.

5.2.6 Antibiotic resistance test of *T. oceanica* cells

Several antibiotic tests were done. First tests with zeocin and phleomycin showed no tolerance of *T. oceanica* at any concentration tested (data not shown). The first test using nourseothricin was done on suspension cultures with concentrations ranging from 1-100 µg/ml. The initial inoculum was $4x10^5$ cells/ml, and the viability of the cells was checked after 2 weeks visually based on the color and via microscopy. The second antibiotic testing was done for the pour-plating experiment. Two different cell densities were used ($1x10^6$ and $1x10^7$ cells per plate), and nourseothricin concentrations ranged from 0.01–10 µg/ml. All conditions were tested as triplicates. The cells were kept in 24 h low-light (25 µE) and assessed after four weeks.

5.2.7 PCR analysis of transformants

The Phusion® High-Fidelity DNA Polymerase (New England Biolabs, Ipswich, Massachusetts, USA) was used for PCR reactions, and primer-specific annealing temperatures were calculated using the NebTm calculator (Polymerase (New England Biolabs, Ipswich, Massachusetts, USA). The general PCR reaction was based on the manufacturer's protocol.

5.3 Results

5.3.1 Pour-plating and antibiotic testing of *T. oceanica* cells

The establishment of a transformation system is dependent on antibiotic resistance to select individual clones with the inserted DNA-plasmid. The first tests with zeocin and phleomycin showed no tolerance of *T. oceanica* cells. The nourseothricin system was used instead, and *T. oceanica* cells grew in suspension cultures with concentrations of 10 µg/ml nourseothricin (Figure 5.3).

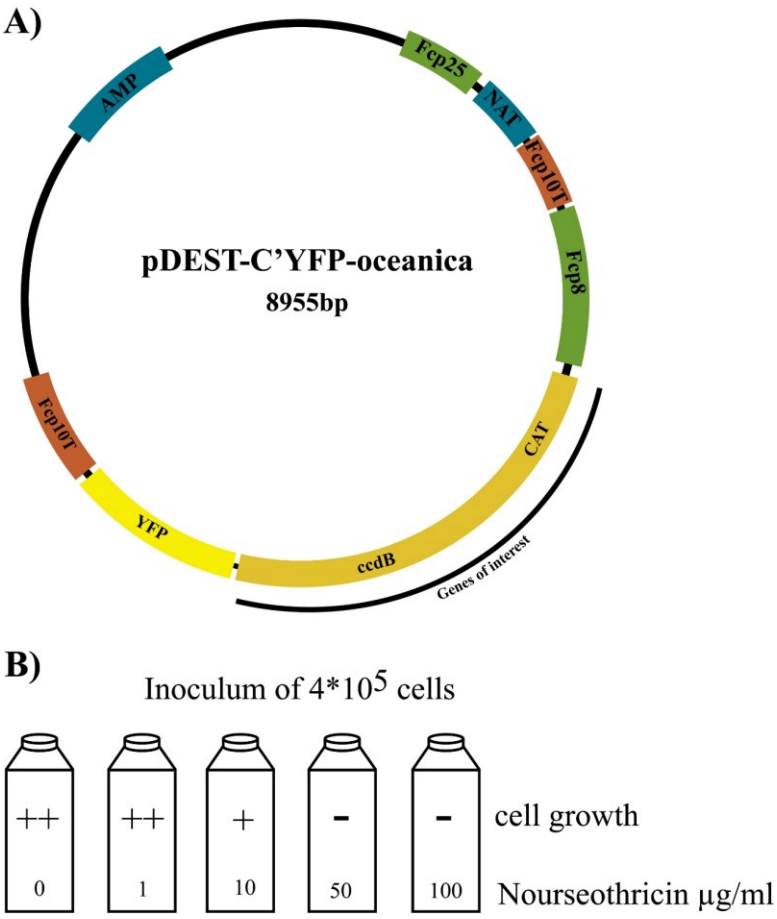


Figure 5.3: Vector map of pDEST-C'YFP-oceanica vector and antibiotic testing.

A) The pdest-C'YFP-oceanica vector was equipped with *T. oceanica* specific promoters and terminators. The following abbreviations are used: ampicillin (AMP), fucoxanthin-chlorophyll binding protein (FCP), nourseothricin N-acetyl transferase (NAT),

chloramphenicol acetyltransferase gene (CAT), control of cell death B (ccdB), yellow fluorescence protein (YFP). B) Antibiotic testing of wildtype *T. oceania* cells towards varying concentrations of Nourseothricin in µg/ml. The media was inoculated with $4x10^5$ cells/ml, cell growth was monitored, and results were taken after 2 weeks.

The first successful transformants were achieved by microparticle bombardment, 48 h recovery time in non-selective media, and a final transfer into selective media using 50 µg/ml nourseothricin (Figure 5.2). Following the successful transformation in suspension cultures, the establishment of the pour-plating technique was the next step, aiming to receive single colonies. The pour plating technique with a previously transformed suspension culture resulted in single colonies, and PCR analysis showed the expression of the resistance gene in two of the three clones analyzed and the presence of the gene-of-interest in one clone. Results on DNA showed the same pattern with one more positive clone for the resistance gene (Figure 5.4).

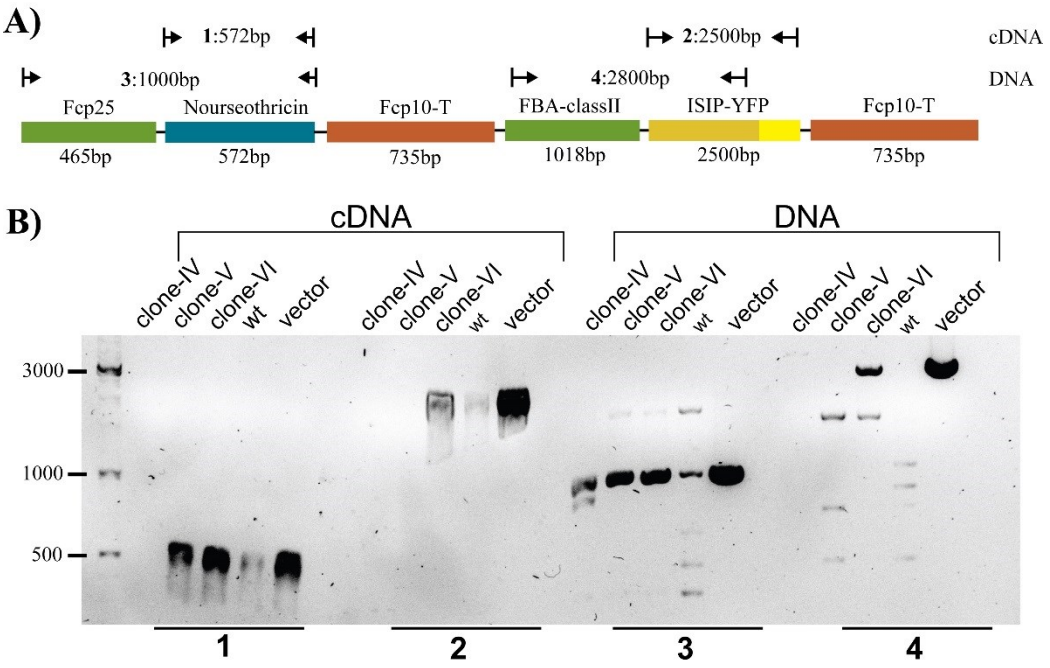


Figure 5.4: PCR analysis of transformants.

Previously antibiotic-resistant suspension cultures were used for pour-plating, and three clones (IV, V, VI) were picked and analyzed. A) Shown is a part of the vector indicating

the amplified areas as 1, 2, 3 and 4. The length of each fragment is under the respective block. B) DNA and cDNA were analyzed. Untransformed DNA/RNA was used as a negative control, while the respective vector served as a positive control. The blocks indicated in A) are also indicated under the gel as 1,2,3, and 4.

The next step was an antibiotic sensitivity test of *T. oceanica* in pour plates towards nourseothricin as well as a cell density test (Figure 5.5). These tests were done with *T. oceanica* cells that were transformed with empty beads. The first test aimed to detect the right plating density, and plates with a cell density 10^5-10^7 cells/plate were efficient to produce colonies (data not shown). The second test aimed to find the antibiotic threshold of *T. oceanica,* which was between 1-5 µg/ml in the pour plating technique. Here, a cell density of 10^7 resulted in a higher plating efficiency compared to 10^6 (Figure 5.5). Unfortunately, the following transformations with vector DNA and subsequent pour-plating did not result in any colonies.

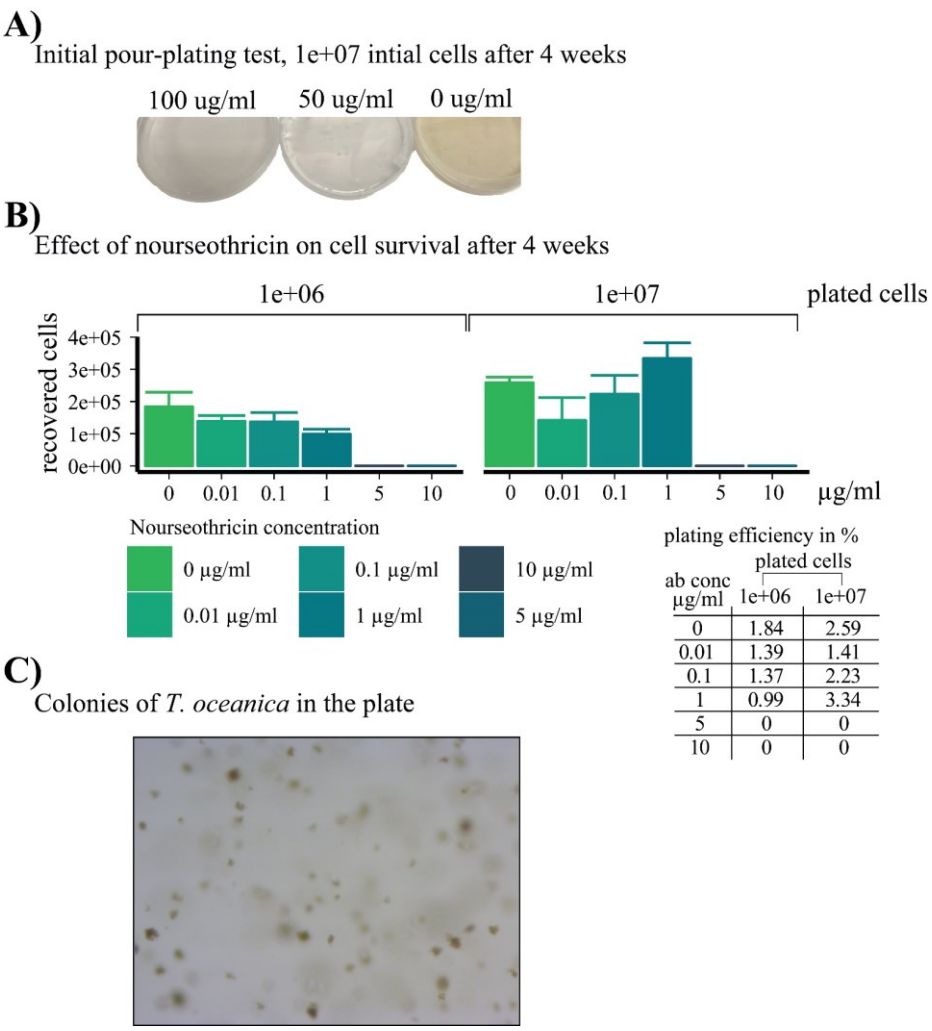


Figure 5.5: Antibiotic testing with different concentrations of nourseothricin using the pour plating technique.

A) An initial test was run on *T. oceanica* cultures to test pour-plating abilities and antibiotic resistance. The cells were kept under regular growth conditions (14/10 light/dark with 100 µE) for four weeks. B) Five different concentrations ranging from 0.01-10 µg/ml of nourseothricin were in combination with two different cell concentrations (1e+06 and 1e+07) to assess the antibiotic resistance threshold. The cells were kept in continuous low light with 25 µE, and cells were counted after four weeks. The plating efficiency is shown in percentage. C) Shown are *T. oceanica* colonies on a plate from the antibiotic testing.

5.3.2 Vector design of pDEST-C'YFP-oceanica

The pDEST-C'-YFP vector (Siaut *et al.*, 2007), previously used for transformation in *P. tricornutum*, was modified for the use with *T. oceanica*, including the exchange of

promoters and terminators (Figure 5.3). The downstream sequence of the fucoxanthin-chlorophyll a-c binding protein 10 (Fcp10) gene was used as a terminator. The promoter of the resistance gene is an upstream regulatory element of Fcp25, and several promoter elements were tried for the gene-of-interest. The promoters for the gene-of-interest included upstream elements of Fcp8, Fcp25, fructose-bisphosphate aldolase (FBA) class-II, nitrate reductase, heat shock protein 70 (HSP70), and the F2A system from the foot-mouth disease virus was tried as an alternative, putatively resulting in co-expression of resistance gene and gene-of-interest (Figure 5.6). The FBA-class-II promoter revealed the best results and was used in combination with several genes for further exploration.

5.3.3 Analysis of cDNA/DNA of antibiotic-resistant transformants

The pDEST-C'YFP-oceanica vector was used for transformation of *T. oceanica* using a biolistic gun (Figure 5.2). The first plating attempt was not successful, and transformants were therefore recovered from suspension cultures using 50 µg/ml nourseothricin (Figure 5.1). Throughout first transformations, a pressure of 900 psi in combination with a bead size of 0.7 µm resulted in positive antibiotic-resistant *T. ocenica* cells, grown in suspension cultures. Higher pressure and larger bead size did not result in positive transformants (data not shown). The first analysis of transformants on DNA/cDNA level showed negative results for the gene of interest (data not shown), which led to the exchange of the Fcp8 promoter with several other promoters (Figure 5.6). Subsequent transformants demonstrated that the FBA-class-II promoter resulted in the most antibiotic-resistant cultures. Several transformants were successfully screened for the presence of nourseothricin acetyltransferase on DNA as well as on cDNA level (Figure

5.6). Furthermore, the ISIP3-YFP construct was amplified on cDNA level and promoter-ISIP3 construct on DNA level (Figure 5.6).

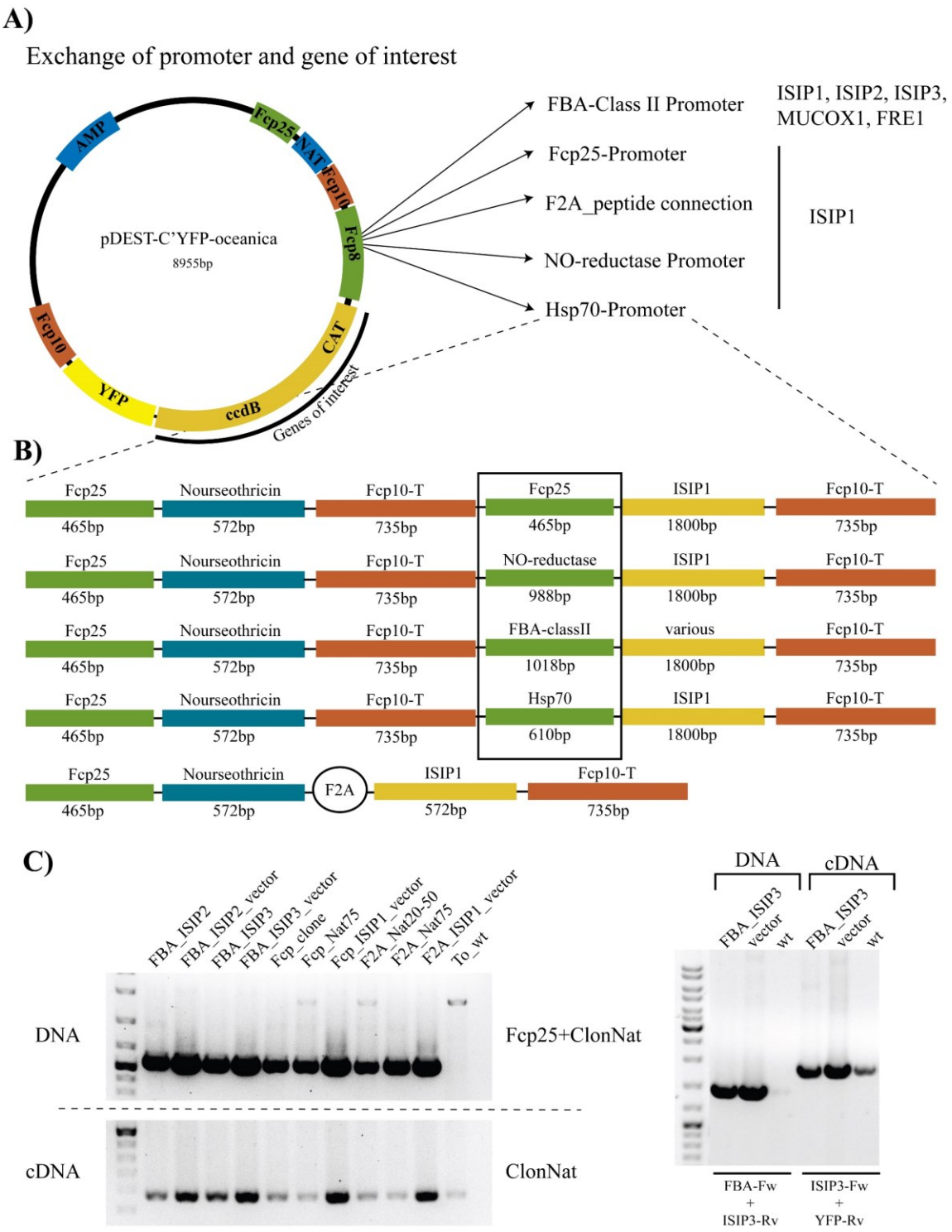


Figure 5.6: Overview of different promoters used in the pDEST-C'YFP-oceanica vector and PCR analysis of antibiotic-resistant cultures.

A) The promoter of the gene of interest was exchanged to test several different promoters. Following the first positive results with an FBA-class-II promoter, several genes were introduced as gene-of-interest with the FBA-class-II promoter. The following abbreviations are used: ampicillin (AMP), fucoxanthin-chlorophyll binding protein (FCP), nourseothricin N-acetyl transferase (NAT), chloramphenicol acetyltransferase gene (CAT), control of cell deathB (ccdB), yellow fluorescence protein (YFP). B) The resulting constructs are shown with the different promoters for the gene-of-interest in the black box. The names of the blocks are provided above, and the respective length is indicated below each block. C) Antibiotic-resistant transformants were analyzed via PCR on DNA and cDNA level for the presence of ClonNat as well as the gene-of-interest. The complete promoter-gene region was amplified for the DNA analysis. Wildtype (wt) DNA/cDNA was used as a negative control, and the respective vector served as a positive control.

5.4 Discussion

The increasing interest in the cell biology of diatoms over the last decades led to the development of numerous transformation systems for different diatom species (Doron *et al.*, 2016; Huang & Daboussi, 2017; Johansson *et al.*, 2019b). Industrial usage of diatoms for various applications and compounds (Radakovits *et al.*, 2010; Tabatabaei *et al.*, 2011; Delalat *et al.*, 2015), and the implication on the global biogeochemical cycle (Boyd *et al.*, 2007; Armbrust, 2009; Assmy *et al.*, 2013; Boyd, 2013; Villar *et al.*, 2018) are the main reasons for the interest in molecular techniques aimed at genetic manipulation in microalgae. Over the last years, transgenic work led to the functional characterization of novel proteins (Allen *et al.*, 2012; Samukawa *et al.*, 2014; Morrissey *et al.*, 2015; Kazamia *et al.*, 2018; McQuaid *et al.*, 2018). The environmental relevance and the ability to withstand chronic iron-limitation (Maldonado & Price, 2001; Peers & Price, 2006; Lommer *et al.*, 2012) are reasons for the development of a transformation system for *T. oceanica*. Adaptations to iron limitation are of great interest as iron impacts algae growth in over 30% of the ocean (Tagliabue *et al.*, 2017). Moreover, *T. oceanica* is a globally distributed open ocean species (Chappell *et al.*, 2015; Carradec *et al.*, 2018) with specific

characteristics, distinguishing *T. oceanica* from coastal species (Strzepek & Harrison, 2004; Peers *et al.*, 2005).

The design of the vector was the first step of the transformation system and included the exchange of promoters and terminators in the pDEST-C'YFP vector (Siaut *et al.*, 2007). One of the first transformation systems that has been established in diatoms was on *P. tricornutum.* The significant advantage of *P. tricornutum* is the absence of a silica shell, which creates a harsh barrier in other diatom species when conducting biolistic transformations. This barrier was one of the concerns regarding the transformation of *T. oceanica*. Therefore, it was not surprising that a smaller bead size, compared to *P. tricornutum,* was successful. However, it was surprising that lower pressure was needed. In *Fistulifera sp*, the pressure was irrelevant and no impact on the transformation efficiency was detected (Muto *et al.*, 2013).

Other advantages of *P. tricornutum* include the ability to grow on plates for several months and to withstand a wide range of antibiotics (Apt *et al.*, 1996). The two antibiotics, zeocin and phleomycin, that are used as a selective marker for *P. tricornutum* were too strong for the use in *T. oceanica*. This led to the development of a transformation system with the nourseothricin system, which was successfully used for *T. pseudonana* (Poulsen *et al.*, 2006). Therefore, the pDEST-C'YFP vector was modified according to the use of nourseothricin and the nourseothricin acetyltransferase was incorporated as the resistance gene (Figure 5.1). Promoter and terminator elements of Fcps were successfully used in numerous transformation systems in diatoms, including *P.tricornutum*, *T. pseudonana*, *Thalassiosira weisflogii*, *Chaetoceros gracilis*, and *Fistulifiera sp* (Huang & Daboussi, 2017). The incomplete assembly of the *T. oceanica*

genome with over 50,000 scaffolds limited the choice of promoters and terminators, but suitable downstream elements of Fcp10 as a terminator and upstream sequences of Fcp25 as a promoter for the resistance gene were identified and used. Initially, upstream elements of Fcp8 were used for the gene of interest, but the lack of successful expression of the gene-of-interest and the lack of YFP fluorescence led to the use of alternative promoter regions. Here, the promoter region of the nitrate reductase as previously used in *P. tricornutum* (Schellenberger Costa *et al.*, 2013) and *T. pseudonana* (Poulsen *et al.*, 2006) was used. Additionally, Fcp25, which was used for the resistance gene, heat-shock protein 70 (Hsp70) as used in *Chlamydomonas reinhardtii* (Doron *et al.*, 2016), and a fructose-bisphosphate aldolase class-II promoter region based on its constant expression (Lommer *et al.*, 2012) were used. The FBA-class-II promoter resulted in the most positive transformants. The F2A peptide from the foot-mouth virus (Minskaia *et al.*, 2013) was used for co-expression of the resistance gene and the gene-of-interest, avoiding the need for two promoters, but no antibiotic-resistant transformants were detected (Figure 5.6).

Antibiotic-resistant transformants were retained with the biolistic approach in suspension cultures without plating, which has been established for *Pseudo-nitzschia multistrata* and *P. aryensis* (Sabatino *et al.*, 2015). DNA delivery via a biolistic gene gun is still the most widely used technique (Apt *et al.*, 1996; Doron *et al.*, 2016; Huang & Daboussi, 2017) but promising results have also been shown with electroporation (Niu *et al.*, 2012; Muñoz *et al.*, 2018), multi-pulse electroporation (Miyahara *et al.*, 2013), and bacterial conjugation (Karas *et al.*, 2015).

In the study presented here, suspension cultures containing antibiotic-resistant *T. oceanica* cells were tested via PCR analysis, demonstrating the expression of the resistance gene (Figure 5.4 and Figure 5.6). Some transformants revealed the amplification of the gene-of-interest with the promoter on the DNA level, as well as the expression of the gene-of-interest YFP construct on cDNA level (Figure 5.4 and Figure 5.6). However, there was no detection of YFP fluorescence under the microscope or in the flow cytometer. A western blot analysis showed bands with the size of YFP, but no bands were detected for the gene-of-interest YFP construct (Supplemental Figure 7). The same issue was described for the centric diatom *Chaetoceros sp.* (Miyagawa-Yamaguchi *et al.*, 2011), which was resolved by the use of Azami-green, an alternative fluorescent protein used to replace GFP (Ifuku *et al.*, 2015). To what extent the use of Azami-green would solve this issue in *T. oceanica* needs to be tested. The disadvantage of using keeping the cells in suspension to retain antibiotic-resistant cells is the inability to select individual clones. This led to the establishment of the pour-plating technique for *T. oceanica*, using extra clean agar and testing of antibiotic resistance towards nourseothricin in the plate (Figure 5.5). Initial results showed that transformed *T. oceanica* cells can grow in plates without antibiotics, but no transformed cells were retained in antibiotic-containing plates.

Overall, this work is an initial step for the development of a successful transformation system in *T. oceanica*. The next step would be to improve the transformation efficiency by testing the parameters used for the biolistic approach, and by the use of alternative DNA delivery techniques. The establishment of bacterial transformation, as well as the use of multi-pulse electroporation, are two techniques that should be tested. Both

approaches should explore using suspension cultures, as well as pour-plates with varying antibiotic concentrations. The use of suspension cultures would give a good indication if the alternative transformation techniques worked with the overall goal to use the pour-plating technique to receive individual colonies. The recovery time between transformation and plating and the light conditions during the recovery process can be a factor for the efficiency of the pour-plating technique and need to be optimized. In more detail, the length of the recovery time, comparing 12 h, 24 h, 36 h, and 48 h, and different light conditions, including varying light levels as well as different variations of the light/dark cycle, should be tested. The salt concentration of the pour-plates and the growth conditions in terms of light levels and light/dark cycles are also parameters that need to be tested further. The optimization of the transformation efficiency and the pour plating conditions could lead to the detection of YFP-positive transformants, but it should be considered to exchange YFP with Azami green to test whether YFP is useable in *T. oceanica*. Here, the exchange of YFP with shorter protein tags, such as HA-Tags, His-Tag or FLAG-tags, would bring further insight into the use of different tags in *T. oceanica*. Research on diatoms needs more useful and reliable transformation systems to cover more diatom species as diatoms have significant genetic differences on species level (Allen *et al.*, 2008; Lommer *et al.*, 2012). Cell biology research on two species, *P. tricornutum* and *T. pseudonana,* is a good start but not sufficient to describe the cell biology of an estimated 100,000 species (Armbrust, 2009).

CHAPTER 6 DISCUSSION

6.0 General Discussion

Thalassiosira oceanica is a globally distributed open ocean diatom which can withstand very low concentrations of iron (Sunda *et al.*, 1991; Malviya *et al.*, 2016). Iron concentrations and iron bioavailability are a constraint in the open ocean and algae have evolved a variety of mechanisms to cope with low iron concentration (Allen *et al.*, 2008; Lommer *et al.*, 2012; Cohen *et al.*, 2017; Luxem *et al.*, 2017). Research efforts on low-iron adapted diatom species revealed the presence of three groups of proteins, coined iron starvation-induced proteins (ISIPs), which are highly upregulated under iron-starvation (Allen *et al.*, 2008; Lommer *et al.*, 2012). *T. oceanica* possesses five ISIP proteins, ISIP1a and ISIP1b, ISIP2, ISIP2x8, and ISIP3. ISIP1b has not been detected at the transcriptional level and is likely derived from a gene duplication event with a loss of function. All other ISIPs are significantly upregulated under iron limitation in *T. oceanica* (Lommer *et al.*, 2012). The targeted transcriptome study using NanoString technology revealed conflicting results on both ISIP2 proteins, as they were not significantly upregulated under low-iron conditions in the study presented here (Chapter 2). Furthermore, there was no response of ISIP2 to the addition of iron. The expression of ISIP2 in *P. tricornutum* is depending on carbonate ions and not only on the concentration of iron (McQuaid *et al.*, 2018), which was not addressed in the targeted transcript analysis performed in this thesis. ISIP2 transcript counts in *Pseudo-nitzschia granii* were only upregulated by 1.4 fold in iron-limited samples (Cohen *et al.*, 2018), and ISIP2a was even induced upon iron addition in a field study (Graff van Creveld *et al.*, 2016). These diverse responses to iron addition show that the mechanisms controlling ISIP2 expression

are not fully understood and that the expression might be species-specific. The analysis of structural properties of ISIP2 proteins across 15 algal species (Chapter 2) revealed a high variation in terms of the presence of transmembrane (TM) domains and conserved domains (NCBI Conserved domain search), supporting the hypothesis of a species-specific function or regulation.

The localization study of YFP-tagged ISIP1a and ISIP3 from *T. oceanica* in *P. tricornutum* (Chapter 2) showed ISIP1a_*To* at the cell surface similar to ISIP1_*Pt* under high iron concentrations (Kazamia *et al.*, 2018). ISIP1_*Pt* is involved iron-siderophore uptake via endocytosis, with an unknown protein binding the iron-siderophore complex and ISIP1 causing the internalization (Kazamia *et al.*, 2018). The sequences of ISIP1a_*To* and ISIP1_*Pt* have high sequence similarity (95% Coverage, 38.98% Identity, e-value: $4x10^{-125}$), and both proteins show a similar topology with an endocytosis motif in their short C-terminal domain (Chapter 2). The topology of ISIP1a_*To* was confirmed by the presence of two glycosylation sites in its extracellular N-terminal domain. Considering the similarities between ISIP1_*Pt* and ISIP1a_*To*, the latter may also be involved in siderophore uptake (Chapter 2). Ferrichrome-binding proteins (FBP) are present in *P. tricornutum* and *T. oceanica* and might be the iron-siderophore binding protein. These proteins also share a high sequence similarity (Coverage: 53%, e-value: $8x10^{-26}$, ID: 30.61%, THAOC_08758: PHAEO_XP_002185795.1). The identification of two putative FBP proteins (THAOC_28875 and THAOC_08758) in *T. oceanica* in the N-linked glycosylation study verifies the translation and presence of FBP proteins in *T. oceanica* (Chapter 4). Both FBP proteins were detected in the low-iron sample, indicating the possibility of iron-siderophore binding in the presence of ISIP1a. Previously identified

binding sites in ferrichrome-binding protein FhuA (Ferguson *et al.*, 1998) and FhuD (Clarke *et al.*, 1998) from *E. coli* were not identified in the FBP proteins from *T. oceanica* or *P. tricornutum* (data not shown).

The targeted transcriptome analysis revealed a strong correlation between ISIP1 and ISIP3 ($R^2 = 0.94$), suggesting common regulatory mechanisms that control the expression (Chapter 3). An alignment of the ISIP1/ISIP3 probes on full-length ISIP3/ISIP1 gene sequences confirmed the specificity of the probes to each target and showed no indication of possible cross-detection based on identities of under 60% in the alignments (Chapter 3), which is below the threshold of 75% for cross-reactivity of 50 bp long oligomers (Kane *et al.*, 2000). The localization study showed ISIP3_*To* in vesicular structures (Chapter 2) that could be periplastidal compartments (Kilian & Kroth, 2005), and more interestingly, these vesicles looked very similar to the structures the internalized siderophores in *P. tricornutum* were located in (Kazamia *et al.*, 2018). The exact function of ISIP3_*To* is unknown, but the presence of a domain of unknown function (DUF305), belonging to the ferritin family, and the localization close to the chloroplasts favors an iron-storage/homeostasis function for the chloroplast (Chapter 2). ISIP3_*To* expression was elevated in the dark in low-iron conditions (Supplemental Figure 5) similar to previous studies on ISIP3 (Chappell *et al.*, 2015; Smith *et al.*, 2016) and the diel cycle of ferritin in *O. tauri* (Botebol *et al.*, 2015). ISIP1 also showed an upregulation at night, but Fv/Fm values remained constant, indicating no increase in iron-stress, which could have caused the upregulation of ISIP1 and ISIP3. Taken together, the similarity of the localization of ISIP3_*To* with the localization of the siderophores, the presence of a domain belonging to the ferritin family, and the high correlation of ISIP1_*To* and

ISIP3_*To* expression, led to the hypothesis that ISIP1 and ISIP3 are involved in the same mechanisms- iron uptake/homeostasis in the chloroplast (Figure 6.1). The iron-responsive promoter sequence in upstream elements of ISIP1, FLDA1, and FBA3 that was identified in *T. oceanica*, *P. tricornutum*, and *F. cylindrus* (Lommer *et al.*, 2012) and later confirmed through mutagenesis studies (Yoshinaga *et al.*, 2014), is absent in ISIP3_*To*, contradicting the hypothesis, but the strong correlation shows that there have to be other regulatory factors.

6.1 Conclusions

The second chapter showed that ISIP proteins are structurally different from ZIP and NRAMP proteins. Differences within the group of ISIP2 proteins in terms of structural properties and the abundance of conserved domains suggests a further diversification of this group. The localization study and the *in-silico* analysis of ISIP3 led to the conclusion that ISIP3 acts as an iron storage protein. Overall, this study shows the diversity of iron acquisition strategies in algae and that some protein families are involved in different iron-acquisition strategies.

The third chapter revealed a detailed response of iron-limited *T. oceanica* cells to the addition of iron at the transcriptional and physiological levels. The main conclusion is that the adjustment of transcript copies starts within 5 min after the addition of iron. In more detail, the downregulation of transcripts for proteins involved in high-affinity iron-uptake takes place first, a response that is likely linked to the protection of the cells against an excess uptake of iron. The switch of transcripts for iron-containing proteins that were replaced by non-iron containing proteins under iron-limitation follows, improving enzyme efficiency. Further conclusions include that the improvement in

photosystem efficiency is faster than the increase in chlorophyll fg/cell. The half-life analysis revealed a constant expression of MnSOD, PETE, and PETF, which shows a constant expression of these genes, including the permanent replacement of cytochrome c_6 with PETE. The study also indicated that the downregulation of CREGx2, a protein of unknown function, and FLDA1 transcripts upon iron-addition is regulated by other transcripts.

An interesting aspect of the targeted transcriptome data is the interpretation of differential gene expression at different timepoints. More specific, PETF appears downregulated under iron-limitation in the morning and upregulated around the switch from the light phase to the dark phase. This observation should be put into consideration in any study that analyzes differential gene expression.

The fourth chapter provided insight into glycosylated peptides under iron-replete and iron-limiting conditions. One of the conclusions is that NXT-type motifs are more often found for N-linked glycosylation than NXS-type motifs. Furthermore, the expression of members of the N-linked glycosylation pathway showed a diurnal pattern with only minor differences between high- and low-iron conditions, demonstrating that N-linked glycosylation, as well as protein folding, are not affected by the iron-status. The presence of N-linked glycosylation sites in ISIP1a confirmed the predicted topology. The identification of two ferrichrome-binding proteins (FBP) in low-iron samples confirms the presence and expression of FBP proteins and suggests an involvement of FBP proteins in ISIP1-mediated siderophore uptake.

The last chapter was the development of a transformation system for *T. oceanica*. The design of a *T. oceanica*-specific transformation vector was completed successfully. The

main conclusions are that *T. oceanica* can be used for genetic transformation and that *T. oceanica* grows in plates when using the pour-plating technique. The main objective of the project to localize YFP-tagged proteins in the cell was not achieved.

6.2 Overall conclusions

The main conclusions are that *T. oceanica* responds immediately to the addition of iron, downregulating two of its ISIP proteins (ISIP1a and ISIP3) and an iron reductase first, and therefore inhibiting an excess uptake of iron. Secondly, ISIP1 and ISIP3 might both be involved in iron uptake/homeostasis into the chloroplast, and both proteins show high structural similarities between different algal species, while ISIP2 proteins possibly act more species-specific based on their high structural variance. Further, the data confirms the predicted topology of ISIP1 and shows the presence of ferrichrome-binding proteins under low-iron conditions, indicating a putative involvement in the ISIP1-mediated siderophore uptake.

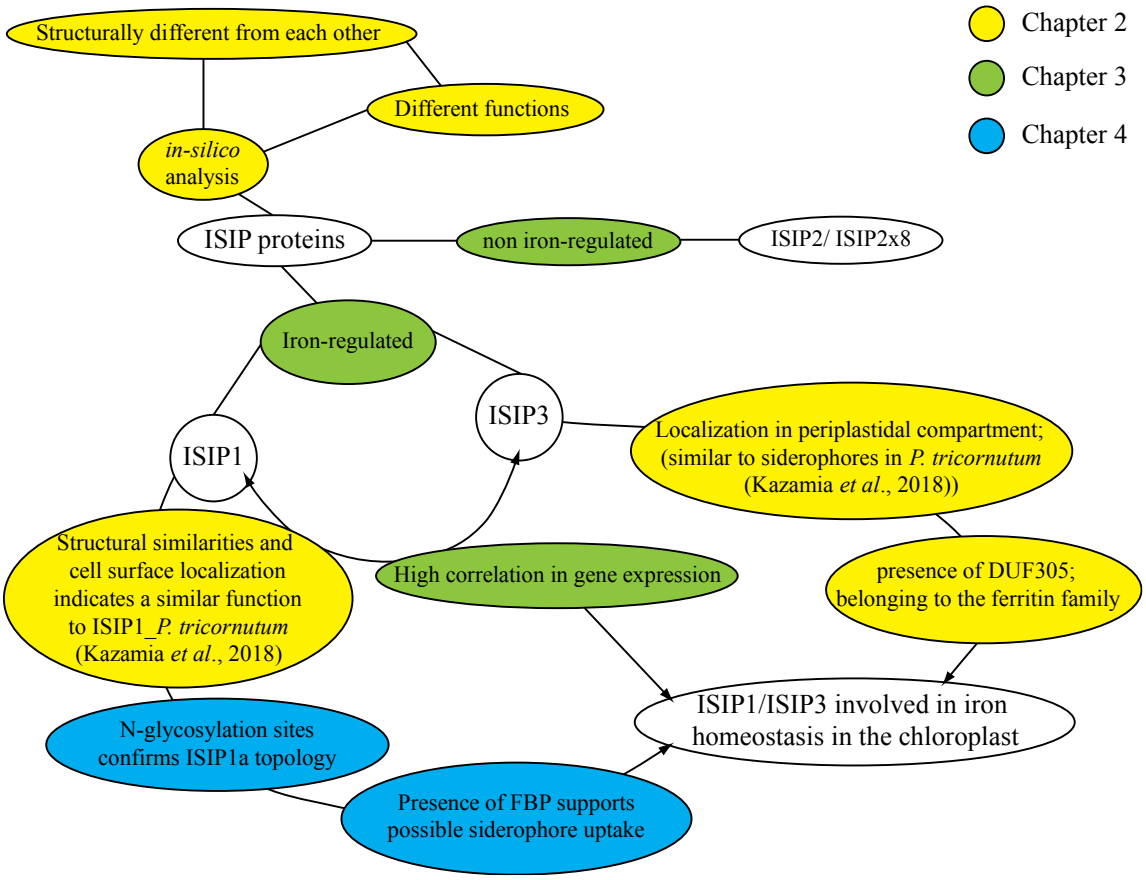


Figure 6.1: Flowchart of results related to ISIP proteins throughout the different chapters.

The chapters are indicated by color, with chapter 2 in yellow, chapter 3 in green, and chapter 4 in blue. The main findings of each chapter regarding the ISIP proteins are presented and connected through lines. The following abbreviations are used: iron starvation-induced protein (ISIP), ferrichrome-binding protein (FBP).

6.3 Future directions

Chapter 2

The analysis of iron uptake related proteins revealed novel aspects of widely distributed proteins such as ZIP and NRAMP proteins, as well as more recently discovered proteins, the ISIP proteins. The analysis revealed results on all three groups of ISIP proteins but did not cover ISIP proteins with domain duplication. The group of ISIP2 proteins with domain duplication needs to be explored in more depth. The proposed C-terminal and N-

terminal domains need to be better defined, and localization studies on members of this subgroup of ISIP2 proteins would be of great interest. Localization studies are often the start for functional experiments, and the question of whether specific NRAMP proteins are involved in iron export from vacuolar stored iron could be partially answered with a localization study. Systematic localization studies of all members of the ZIP and NRAMP family in *T. oceanica* and *P. tricornutum* would be of great interest. The integration of the analyzed proteins in a phylogenetic background could help to understand evolutionary relationships and functional diversification. Of particular interest would be the analysis of ISIP1a_*To* transformed into *P. tricornutum* to test if ISIP1a_*To* provides *P. tricornutum* with the ability to grow in the presence of a catecholate siderophore instead of a hydroxamate siderophore, testing the results by Kazamia et al. (Kazamia *et al.*, 2018)

Chapter 3

The targeted transcriptome study revealed a very rapid, immediate response of *T. oceanica* to the addition of iron. A similar study would be desirable for the analysis of a whole transcriptome using short time intervals within the first hour and hourly measurements after. The whole transcriptome study should focus on four different aspects. The first focus should be on the transcripts of iron-uptake related proteins including ZIP and NRAMP. The second focus should be the recovery of the photosystem as the improvement in photosystem efficiency was faster than the recovery of the chlorophyll content fg/cell, and the third focus should be mitochondrial genes, as well as genes for carbon compound degrading enzymes to analyze the suggested metabolic shift further (Lommer *et al.*, 2012). The last aspect should be a detailed analysis of the

identified transcription factors. The use of the transcription inhibitor actinomycin D revealed mRNA half-lives, which leads to the question of half-lives of corresponding proteins. Therefore, comparing transcriptomic data with proteomic data would not only bring insight into protein half-lives but also show the correlation between transcriptome and proteome data. Staining of cellular compartments in addition to electron microscopy of *T. oceanica* under varying iron concentrations could reveal aspects of cellular adjustments in terms of the number of endosomes, higher degradation of material in lysosomes, and the reduced chloroplast size.

Chapter 4

The next step of this project should be the analysis of glycan structures and the comparison of these structures with other microalgae. The identified members of the N-linked glycosylation pathway (OST, GnT1, Calreticulin, or UCGT) should be analyzed in a localization study. The silencing of OST, GnT1, Calreticulin, or UCGT individually through siRNA, RNAi or CRISPR/Cas9 could bring first hints towards the functional relevance of N-linked glycosylation in diatoms. This approach would not be possible in *T. oceanica* but in *P. tricornutum*. The analysis of N-linked glycopeptides of different cellular compartments could result in higher coverage and verify the predicted localization of identified N-linked glycosylated proteins.

Chapter 5

The transformation system was partially successful, and two main issues need to be addressed. The first issue is the visualization of YFP tagged proteins, and the second issue is the plating efficiency to generate single colonies of transformed cells. Initial

western blot analysis did not show bands for YFP-tagged proteins. The next step would be to use smaller tags, HA-tag for example, to analyze whether YFP is just not suitable as a tag in *T. oceanica* and is degraded or if the whole protein is not being translated correctly, with a potential issue in the vector. It also should be clarified through a Southern Blot analysis, whether the vector-DNA is located in the nucleus. The pour plating technique of *T. oceanica* cells was successful for untransformed cells or transformed cells that were first grown in suspension. There were no single colonies achieved when using transformed cells in antibiotic-containing plates, even with relative low concentration of Nourseothricin of only 5 µg/ml, compared to 50 µg/ml when successfully growing transformed cells in selective media. The use of microparticle bombardment was successful for generating antibiotic transformants in suspension cultures, but the transformation efficiency might be too low to generate colonies with the pour plating technique. Therefore, bacterial transformation and electroporation should be tested for the pour plating. Further, the salt concentration of the plates should be varied to analyze if that effects the plating efficiency of transformed cells in antibiotic-containing plates.

REFERENCES

Aguirre, J.D., Clark, H.M., McIlvin, M., Vazquez, C., Palmere, S.L., Grab, D.J., Seshu, J., Hart, P.J., Saito, M., & Culotta, V.C. (2013) A manganese-rich environment supports superoxide dismutase activity in a lyme disease pathogen, *Borrelia burgdorferi*. *J. Biol. Chem.*, **288**:8468–8478.

Allen, A.E., Dupont, C.L., Oborník, M., Horák, A., Nunes-Nesi, A., McCrow, J.P., Zheng, H., Johnson, D.A., Hu, H., Fernie, A.R., & Bowler, C. (2011) Evolution and metabolic significance of the urea cycle in photosynthetic diatoms. TL - 473. *Nature*, **473 VN**-:203–207.

Allen, A.E., Laroche, J., Maheswari, U., Lommer, M., Schauer, N., Lopez, P.J., Finazzi, G., Fernie, A.R., & Bowler, C. (2008) Whole-cell response of the pennate diatom *Phaeodactylum tricornutum* to iron starvation. *Proc. Natl. Acad. Sci. U. S. A.*, **105**:10438–10443.

Allen, A.E., Moustafa, A., Montsant, A., Eckert, A., Kroth, P.G., & Bowler, C. (2012) Evolution and functional diversification of fructose bisphosphate aldolase genes in photosynthetic marine diatoms. *Mol. Biol. Evol.*, **29**:367–379.

Allen, M.D., Del Campo, J.A., Kropat, J., & Merchant, S.S. (2007) FEA1, FEA2, and FRE1, encoding two homologous secreted proteins and a candidate ferrireductase, are expressed coordinately with FOX1 and FTR1 in iron-deficient *Chlamydomonas reinhardtii*. *Eukaryot. Cell*, **6**:1841–1852.

Alonso, J.M., Hirayama, T., Roman, G., Nourizadeh, S., & Ecker, J.R. (1999) EIN2, a bifunctional transducer of ethylene and stress responses in *Arabidopsis*. *Science (80-.).*, **284**:2148–2152.

Amin, S.A., Parker, M.S., & Armbrust, E. V. (2012) Interactions between Diatoms and Bacteria. *Microbiol. Mol. Biol. Rev.*, **76**:667–684.

Antala, S., Ovchinnikov, S., Kamisetty, H., Baker, D., & Dempski, R.E. (2015) Computation and functional studies provide a model for the structure of the zinc transporter hZIP4. *J. Biol. Chem.*, **290**:17796–17805.

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

Archibald, F. (1983) *Lactobacillus plantarum* , an organism not requiring iron. *FEMS Microbiol. Lett.*, **19**:29–32.

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

Armbrust, E.V., Lopez, P.J., La Roche, J., De Martino, A., Porcel, B.M., Bowler, C., Martin-Jézéquel, V., Cadoret, J., Heijde, M., Schnitzler Parker, M., Maheswari, U., Mock, T., Weissenbach, J., Vardi, A., Katinka, M., Jabbari, K., Petit, J., Allen, A.E., Kaas, R., & Meichenin, A. (2010) Digital expression profiling of novel diatom transcripts provides insight into their biological functions. *Genome Biol.*, **11**:R85.

Assmy, P., Jansen, S., Fuchs, N., Kragefsky, S., Latasa, M., Steigenberger, S., Herndl, G.J., Webb, A., Breitbarth, E., Berg, G.M., Bathmann, U., Arrieta, J.M., Henjes, J., Klaas, C., Schuller, S.E., Peeken, I., Strass, V.H., Montresor, M., Friedrichs, L., Scharek, R., Rottgers, R., Smetacek, V., Wolf-Gladrow, D., & Cisewski, B. (2013) Thick-shelled, grazer-protected diatoms decouple ocean carbon and silicon cycles in the iron-limited Antarctic Circumpolar Current. *Proc. Natl. Acad. Sci.*, **110**:20633–20638.

Bai, L., Qiao, M., Zheng, R., Deng, C., Mei, S., & Chen, W. (2016) Phylogenomic analysis of transferrin family from animals and plants. *Comp. Biochem. Physiol. Part D Genomics Proteomics*, **17**:1–8.

Baïet, B., Burel, C., Saint-Jean, B., Louvet, R., Menu-Bouaouiche, L., Kiefer-Meyer, M.C., Mathieu-Rivet, E., Lefebvre, T., Castel, H., Carlier, A., Cadoret, J.P., 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.

Bastow, E.L., Garcia de la Torre, V.S., Maclean, A.E., Green, R.T., Merlot, S., Thomine, S., & Balk, J. (2018) Vacuolar iron stores gated by NRAMP3 and NRAMP4 are the primary source of iron in germinating seeds. *Plant Physiol.*, **177**:1267–1276.

Behringer, G., Ochsenkühn, M.A., Fei, C., Fanning, J., Koester, J.A., & Amin, S.A. (2018) Bacterial communities of diatoms display strong conservation across strains and time. *Front. Microbiol.*, **9**:1–15.

Benoiston, A., Ibarbalz, F.M., Bittner, L., Guidi, L., Jahn, O., Dutkiewicz, S., Bowler, C., & Bowler, C. (2017) The evolution of diatoms and their biogeochemical functions.

Bertrand, E.M., McCrow, J.P., Moustafa, A., Zheng, H., McQuaid, J.B., Delmont, T.O., Post, A.F., Sipler, R.E., Spackeen, J.L., Xu, K., Bronk, D.A., Hutchins, D.A., & Allen, A.E. (2015) Phytoplankton–bacterial interactions mediate micronutrient colimitation at the coastal Antarctic sea ice edge. *Proc. Natl. Acad. Sci.*, **112**:9938–9943.

Blom, N., Sicheritz-Pontén, T., Gupta, R., Gammeltoft, S., & Brunak, S. (2004) Prediction of post-translational glycosylation and phosphorylation of proteins from the amino acid sequence. *Proteomics*, **4**:1633–1649.

Botebol, H., Lelandais, G., Six, C., Lesuisse, E., Meng, A., Bittner, L., Lecrom, S., Sutak, R., Lozano, J.C., Schatt, P., Vergé, V., Blain, S., & Bouget, F.Y. (2017) Acclimation of a low iron adapted *Ostreococcus strain* to iron limitation through cell biomass lowering. *Sci. Rep.*, **7**:1–11.

Botebol, H., Lesuisse, E., Šuták, R., Six, C., Lozano, J.C., Schatt, P., Vergé, V., Kirilovsky, A., Morrissey, J., Léger, T., Camadro, J.M., Gueneugues, A., Bowler, C., Blain, S., & Bouget, F.Y. (2015) Central role for ferritin in the day/night regulation of iron homeostasis in marine phytoplankton. *Proc. Natl. Acad. Sci. U. S. A.*, **112**:14652–14657.

Bowler, C., Allen, A.E., Badger, J.H., Grimwood, J., Jabbari, K., Kuo, A., Maheswari,

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

Boyd, P.W. (2013) Diatom traits regulate Southern Ocean silica leakage. *Proc. Natl. Acad. Sci.*, **110**:20358–20359.

Boyd, P.W., Jickells, T., Law, C.S., Blain, S., Boyle, E. a, Buesseler, K.O., Coale, K.H., Cullen, J.J., de Baar, H.J.W., Follows, M., Harvey, M., Lancelot, C., Levasseur, M., Owens, N.P.J., Pollard, R., Rivkin, R.B., Sarmiento, J., Schoemann, V., Smetacek, V., Takeda, S., Tsuda, a, Turner, S., & Watson, a J. (2007) Mesoscale iron enrichment experiments 1993-2005: synthesis and future directions. *Science*, **315**:612–617.

Bozzi, A.T., Bane, L.B., Weihofen, W.A., McCabe, A.L., Singharoy, A., Chipot, C.J., Schulten, K., & Gaudet, R. (2016) Conserved methionine dictates substrate preference in Nramp-family divalent metal transporters. *Proc. Natl. Acad. Sci.*, **113**:10310–10315.

Butler, A. & Theisen, R.M. (2010) Iron(III)-siderophore coordination chemistry: Reactivity of marine siderophores. *Coord. Chem. Rev.*, **254**:288–296.

Cailliatte, R., Mari, S., Briat, J.F., Curie, C., & Schikora, A. (2010) High-Affinity Manganese Uptake by the Metal Transporter NRAMP1 Is Essential for *Arabidopsis* Growth in Low Manganese Conditions. *Plant Cell*, **22**:904–917.

Caputi, L., Carradec, Q., Eveillard, D., Kirilovsky, A., Pelletier, E., Pierella Karlusich, J.J., Rocha Jimenez Vieira, F., Villar, E., Chaffron, S., Malviya, S., Scalco, E., Acinas, S.G., Alberti, A., Aury, J.-M., Benoiston, A.-S., Bertrand, A., Biard, T., Bittner, L., Boccara, M., Brum, J.R., Brunet, C., Busseni, G., Carratalà, A., Claustre, H., Coelho, L.P., Colin, S., D'Aniello, S., Da Silva, C., Del Core, M., Doré, H., Gasparini, S., Kokoszka, F., Jamet, J.-L., Lejeusne, C., Lepoivre, C., Lescot, M., Lima-Mendez, G., Lombard, F., Lukeš, J., Maillet, N., Madoui, M.-A., Martinez, E., Mazzocchi, M.G., Néou, M.B., Paz-Yepes, J., Poulain, J., Ramondenc, S., Romagnan, J.-B., Roux, S., Salvagio Manta, D., Sanges, R., Speich, S., Sprovieri, M., Sunagawa, S., Taillander, V., Tanaka, A., Tirichine, L., Trottier, C., Uitz, J., Veluchamy, A., Veselá, J., Vincent, F., Yau, S., Kandels-Lewis, S., Searson, S., Dimier, C., Picheral, M., Bork, P., Boss, E., de Vargas, C., Follows, M.J., Grimsley,

N., Guidi, L., Hingamp, P., Karsenti, E., Sordino, P., Stemmann, L., Sullivan, M.B., Tagliabue, A., Zingone, A., Garczarek, L., D'Ortenzio, F., Testor, P., Not, F., Ribera d'Alcalà, M., Wincker, P., Bowler, C., Iudicone, D., Acinas, S.G., Bork, P., Boss, E., de Vargas, C., Follows, M.J., Gorsky, G., Grimsley, N., Hingamp, P., Jaillon, O., Kandels-Lewis, S., Karp-Boss, L., Karsenti, E., Krzic, U., Not, F., Ogata, H., Pesant, S., Raes, J., Reynaud, E.G., Sardet, C., Sieracki, M., Speich, S., Stemmann, L., Sullivan, M.B., Sunagawa, S., Velayoudon, D., & Weissenbach, J. (2019) Community-Level Responses to Iron Availability in Open Ocean Planktonic Ecosystems. *Global Biogeochem. Cycles*, 391–419.

Carradec, Q., Pelletier, E., Da Silva, C., Alberti, A., Seeleuthner, Y., Blanc-Mathieu, R., Lima-Mendez, G., Rocha, F., Tirichine, L., Labadie, K., Kirilovsky, A., Bertrand, A., Engelen, S., Madoui, M.A., Méheust, R., Poulain, J., Romac, S., Richter, D.J., Yoshikawa, G., Dimier, C., Kandels-Lewis, S., Picheral, M., Searson, S., Acinas, S.G., Boss, E., Follows, M., Gorsky, G., Grimsley, N., Karp-Boss, L., Krzic, U., Pesant, S., Reynaud, E.G., Sardet, C., Sieracki, M., Speich, S., Stemmann, L., Velayoudon, D., Weissenbach, J., Jaillon, O., Aury, J.M., Karsenti, E., Sullivan, M.B., Sunagawa, S., Bork, P., Not, F., Hingamp, P., Raes, J., Guidi, L., Ogata, H., De Vargas, C., Iudicone, D., Bowler, C., & Wincker, P. (2018) A global ocean atlas of eukaryotic genes. *Nat. Commun.*, **9**.

Castaings, L., Caquot, A., Loubet, S., & Curie, C. (2016) The high-affinity metal Transporters NRAMP1 and IRT1 Team up to Take up Iron under Sufficient Metal Provision. *Sci. Rep.*, **6**:1–11.

Chappell, P.D., Whitney, L.P., Wallace, J.R., Darer, A.I., Jean-Charles, S., & Jenkins, B.D. (2015) Genetic indicators of iron limitation in wild populations of *Thalassiosira oceanica* from the northeast Pacific Ocean. *ISME J.*, **9**:592–602.

Cherepanova, N., Shrimal, S., & Gilmore, R. (2016) N-linked glycosylation and homeostasis of the endoplasmic reticulum. *Curr. Opin. Cell Biol.*, **41**:57–65.

Chuang, G.Y., Boyington, J.C., Gordon Joyce, M., Zhu, J., Nabel, G.J., Kwong, P.D., & Georgiev, I. (2012) Computational prediction of N-linked glycosylation incorporating structural properties and patterns. *Bioinformatics*, **28**:2249–2255.

Clarke, T.E., Ku, S.Y., Dougan, D.R., Vogel, H.J., Tari, L.W., Ferguson, A.D., Hofmann, E., Coulton, J.W., Diederichs, K., & Welte, W. (1998) The structure of the ferric siderophore binding protein FhuD complexed with gallichrome. *Nat. Struct. Biol.*, **7**:2215–2220.

Cohen, A., Nelson, H., & Nelson, N. (2000) The family of SMF metal ion transporters in yeast cells. *J. Biol. Chem.*, **275**:33388–33394.

Cohen, N.R., Gong, W., Moran, D.M., McIlvin, M.R., Saito, M.A., & Marchetti, A. (2018) Transcriptomic and proteomic responses of the oceanic diatom *Pseudo-nitzschia granii* to iron limitation. *Environ. Microbiol.*, **20**:3109–3126.

Cohen, N.R., Mann, E.L., Stemple, B., Moreno, C.M., Rauschenberg, S., Jacquot, J.E., Sunda, W.G., Twining, B.S., & Marchetti, A. (2018) Iron storage capacities and

associated ferritin gene expression among marine diatoms. *Limnol. Oceanogr.*, **63**:1677–1691.

Cohen, N.R., Sunda, W.G., McNair, H., Twining, B.S., Bruland, K.W., Marchetti, A., Brzezinski, M.A., Maldonado, M.T., Thamatrakoln, K., Ellis, K.A., Bargu, S., Kuzminov, F.I., Lampe, R.H., & Till, C.P. (2017) Diatom Transcriptional and Physiological Responses to Changes in Iron Bioavailability across Ocean Provinces. *Front. Mar. Sci.*, **4**:360.

Connorton, J.M., Balk, J., & Rodríguez-Celma, J. (2017) Iron homeostasis in plants-a brief overview. *Metallomics*, **9**:813–823.

Courville, P., Chaloupka, R., Veyrier, F., & Cellier, M.F.M. (2004) Determination of Transmembrane Topology of the *Escherichia coli* Natural Resistance-associated Macrophage Protein (Nramp) Ortholog. *J. Biol. Chem.*, **279**:3318–3326.

Curie, C., Panaviene, Z., Loulergue, C., Dellaporta, S.L., Briat, J.F., & Walker, E.L. (2001) Maize *yellow stripe1* encodes a membrane protein directly involved in Fe(III) uptake. *Nature*, **409**:346–349.

Curtis, B.A., Tanifuji, G., Maruyama, S., Gile, G.H., Hopkins, J.F., Eveleigh, R.J.M., Nakayama, T., Malik, S.B., Onodera, N.T., Slamovits, C.H., Spencer, D.F., Lane, C.E., Gray, M.W., Archibald, J.M., Burki, F., Hirakawa, Y., Reyes-Prieto, A., Keeling, P.J., Fast, N.M., Green, B.R., Grisdale, C.J., Gruber, A., Kroth, P.G., Irimia, M., Arias, M.C., Ball, S.G., Kuo, A., Schmutz, J., Grimwood, J., Lindquist, E., Lucas, S., Salamov, A., Grigoriev, I. V., Rensing, S.A., Symeonidi, A., Elias, M., Herman, E.K., Klute, M.J., Dacks, J.B., Oborník, M., Kořený, L., Durnford, D.G., Neilson, J.A.D., Armbrust, E.V., Rocap, G., Aves, S.J., Liu, Y., Beiko, R.G., Coutinho, P., Henrissat, B., Hempel, F., Maier, U.G., Zauner, S., Häppner, M.P., Ishida, K.I., Shirato, S., Suzuki, S., Kim, E., Richards, T.A., Mc Rose, D., Worden, A.Z., Mock, T., Poole, A.M., Pritham, E.J., Roy, S.W., Schaack, S., Bell, C., Bharti, A.K., Crow, J.A., Kramer, R., & Mc Fadden, G.I. (2012) Algal genomes reveal evolutionary mosaicism and the fate of nucleomorphs. *Nature*, **492**:59–65.

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

de Baar, H.J.W., Boyd, P.W., Coale, K.H., Landry, M.R., Tsuda, A., Assmy, P., Bakker, D.C.E., Bozec, Y., Barber, R.T., Brzezinski, M.A., Buesseler, K.O., Boyé, M., Croot, P.L., Gervais, F., Gorbunov, M.Y., Harrison, P.J., Hiscock, W.T., Laan, P., Lancelot, C., Law, C.S., Levasseur, M., Marchetti, A., Millero, F.J., Nishioka, J., Nojiri, Y., van Oijen, T., Riebesell, U., Rijkenberg, M.J.A., Saito, H., Takeda, S., Timmermans, K.R., Veldhuis, M.J.W., Waite, A.M., & Wong, C.S. (2005) Synthesis of iron fertilization experiments: From the iron age in the age of enlightenment. *J. Geophys. Res. C Ocean.*, **110**:1–24.

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.

Debeljak, P., Toulza, E., Beier, S., Blain, S., & Obernosterer, I. (2019) Microbial iron metabolism as revealed by gene expression profiles in contrasted Southern Ocean regimes. *Environ. Microbiol.*, **21**:2360–2374.

Delalat, B., Sheppard, V.C., Rasi Ghaemi, S., Rao, S., Prestidge, C.A., McPhee, G., Rogers, M.L., Donoghue, J.F., Pillay, V., Johns, T.G., Kröger, N., & Voelcker, N.H. (2015) Targeted drug delivery using genetically engineered diatom biosilica. *Nat. Commun.*, **6**:8791.

DiDonato, R.J., Roberts, L.A., Sanderson, T., Eisley, R.B., & Walker, E.L. (2004) *Arabidopsis Yellow Stripe-Like2 (YSL2)*: A metal-regulated gene encoding a plasma membrane transporter of nicotianamine-metal complexes. *Plant J.*, **39**:403–414.

Ding, J., Ji, C., & Cai, H. (2017) Membrane Transporter Families of Metal Microelements make Plants Grow Better and Healthier **7**:74–100.

Dogs, M., Daniel, R., Bergen, N., Wemheuer, B., Brinkhoff, T., Wolter, L., & Simon, M. (2017) *Rhodobacteraceae* on the marine brown alga *Fucus spiralis* are abundant and show physiological adaptation to an epiphytic lifestyle. *Syst. Appl. Microbiol.*, **40**:370–382.

Doron, L., Segal, N., & Shapira, M. (2016) Transgene Expression in Microalgae—From Tools to Applications. *Front. Plant Sci.*, **7**:1–24.

Dunahay, T.G. & Jarvis, E.E. (1995) Genetic transformation of the diatoms *Cyclotella Cryptica* and *Navicula Saprophila*. *J. Phycol.*, **31**:1004–1012.

Duprat, L.P.A.M., Bigg, G.R., & Wilton, D.J. (2016) Enhanced Southern Ocean marine productivity due to fertilization by giant icebergs. *Nat. Geosci.*, **9**:219–221.

E.W Juergen, Polle, a, b Kerrie Barry, c John Cushman, d Jeremy Schmutz, c, E., Duc Tran, a Leyla T. Hathwaik, d Won C. Yim, d Jerry Jenkins, E., Zaid McKie-Krisberg, a Simon Prochnik, C., * Erika Lindquist, c Rhyan B. Dockter, C., Catherine Adam, c Henrik Molina, f Jakob Bunkenborg, g EonSeon Jin, H., & Mark Buchheim, i J.M. (2017) Draft Nuclear Genome Sequence of the Halophilic and Beta-Carotene-Accumulating Green Alga *Dunaliella salina* Strain CCAP 19/18. *Genome Announc.*, **5**:e01105-17.

Emanuelsson, O., Nielsen, H., Brunak, S., & Von Heijne, G. (2000) Predicting subcellular localization of proteins based on their N-terminal amino acid sequence. *J. Mol. Biol.*, **300**:1005–1016.

Emanuelsson, O., Nielsen, H., & Heijne, G. Von (1999) ChloroP, a neural network-based method for predicting chloroplast transit peptides and their cleavage sites. *Protein Sci.*, **8**:978–984.

Falkowski, P., Scholes, R.J., Boyle, E., Canadell, J., Canfield, D., Elser, J., Gruber, N., Hibbard, K., Hogberg, P., Linder, S., Mackenzie, F.T., Moore, B., Pedersen, T., Rosental, Y., Seitzinger, S., Smetacek, V., & Steffen, W. (2000) The global carbon cycle: A test of our knowledge of earth as a system. *Science (80-.).*,.

Fang, H.-M. & Wang, Y. (2002) Characterization of iron-binding motifs in *Candida albicans* high-affinity iron permease CaFtr1p by site-directed mutagenesis. *Cell*, **647**:641–647.

Ferguson, A.D., Hofmann, E., Coulton, J.W., Diederichs, K., & Welte, W. (1998) Siderophore-mediated iron transport: Crystal structure of FhuA with bound lipopolysaccharide. *Science (80-.).*, **282**:2215–2220.

Fisher, M., Gokhman, I., Pick, U., & Zamir, A. (1997) A structurally novel transferrin-like protein accumulates in the plasma membrane of the unicellular green alga *Dunaliella salina* grown in high salinities. *J. Biol. Chem.*, **272**:1565–1570.

Fouilland, E., Tolosa, I., Bonnet, D., Bouvier, C., Bouvier, T., Bouvy, M., Got, P., Le Floc'h, E., Mostajir, B., Roques, C., Sempéré, R., Sime-Ngando, T., & Vidussi, F. (2014) Bacterial carbon dependence on freshly produced phytoplankton exudates under different nutrient availability and grazing pressure conditions in coastal marine waters. *FEMS Microbiol. Ecol.*, **87**:757–769.

Fuqua, B.K., Vulpe, C.D., & Anderson, G.J. (2012) Intestinal iron absorption. *J. Trace Elem. Med. Biol.*, **26**:115–119.

Geider, R.J. & La Roche, J. (1994) The role of iron in phytoplankton photosynthesis, and the potential for iron-limitation of primary productivity in the sea. *Photosynth. Res.*, **39**:275–301.

Gentleman, R.C., Carey, V.J., Bates, D.M., Bolstad, B., Dettling, M., Dudoit, S., Ellis, B., Gautier, L., Ge, Y., Gentry, J., Hornik, K., Hothorn, T., Huber, W., Iacus, S., Irizarry, R., Leisch, F., Li, C., Maechler, M., Rossini, A.J., Sawitzki, G., Smith, C., Smyth, G., Tierney, L., Yang, J.Y.H., & Zhang, J. (2004) Bioconductor: open software development for computational biology and bioinformatics. *Genome Biol.*,.

Gledhill, M. & Buck, K.N. (2012) The organic complexation of iron in the marine environment: A review. *Front. Microbiol.*, **3**:69.

Goecke, F., Thiel, V., Wiese, J., Labes, A., & Imhoff, J.F. (2013) Algae as an important environment for bacteria – phylogenetic relationships among new bacterial species isolated from algae. *Phycologia*, **52**:14–24.

Goldman, J.C., Mccarthy, J.J., & Goldman, J.C. (1978) Steady state growth and ammonium uptake of a marine diatom. *Limnol. Oceanogr.*, **23**:695–703.

Graff van Creveld, S., Rosenwasser, S., Levin, Y., & Vardi, A. (2016) Chronic iron limitation confers transient resistance to oxidative stress in marine diatoms. *Plant Physiol.*, **172**:pp.00840.2016.

Grass, G., Taudte, N., Nies, D.H., Maguire, M.E., Rensing, C., Franke, S., & Kucharski, L.M. (2005) The Metal Permease ZupT from *Escherichia coli* Is a Transporter with a Broad Substrate Spectrum. *J. Bacteriol.*, **187**:1604–1611.

Grillet, L. & Schmidt, W. (2019) Iron acquisition strategies in land plants: not so different after all. *New Phytol.*, 11–18.

Große, C., Scherer, J., Koch, D., Otto, M., Taudte, N., & Grass, G. (2006) A new ferrous iron-uptake transporter, EfeU (YcdN), from *Escherichia coli*. *Mol. Microbiol.*, **62**:120–131.

Groussman, R.D., Parker, M.S., & Armbrust, E.V. (2015) Diversity and evolutionary history of iron metabolism genes in diatoms. *PLoS One*, **10**:e0129081.

Gu, Z., Eils, R., & Schlesner, M. (2016) Complex heatmaps reveal patterns and correlations in multidimensional genomic data. *Bioinformatics*, **32**:2847–2849.

Guerinot, M. Lou (2000) The ZIP family of metal transporters. *Biochim. Biophys. Acta - Biomembr.*, **1465**:190–198.

Guo, S., Campbell, R.L., Yashunsky, V., Stevens, C.A., Smith, S.P., Allingham, J.S., Braslavsky, I., Escobedo, C., Davies, P.L., Vance, T.D.R., Voets, I.K., Olijve, L.L.C., Bar-Dolev, M., Langelaan, D.N., Yazdi, S.R., & Graham, L.A. (2017) Structure of a 1.5-MDa adhesin that binds its Antarctic bacterium to diatoms and ice. *Sci. Adv.*, **3**:e1701440.

Harrington, J.M. & Crumbliss, A.L. (2009) The redox hypothesis in siderophore-mediated iron uptake. *BioMetals*, **22**:679–689.

Hassler, C.S., Schoemann, V., Boye, M., Tagliabue, A., Rozmarynowycz, M., & McKay, R.M.L. (2012) Iron bioavailability in the Southern Ocean. *Oceanogr. Mar. Biol.*, **50**:1–64.

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

Hempel, F., Lau, J., Klingl, A., & Maier, U.G. (2011) Algae as protein factories: Expression of a human antibody and the respective antigen in the diatom phaeodactylum tricornutum. *PLoS One*, **6**.

Hempel, F. & Maier, U.G. (2012) An engineered diatom acting like a plasma cell secreting human IgG antibodies with high efficiency. *Microb. Cell Fact.*, **11**:2–7.

Hider, R.C. & Kong, X. (2010) Chemistry and biology of siderophores. *Nat. Prod. Rep.*, **27**:637–657.

Hippmann, A.A., Schuback, N., Moon, K.M., McCrow, J.P., Allen, A.E., Foster, L.J., Green, B.R., & Maldonado, M.T. (2017) Contrasting effects of copper limitation on the photosynthetic apparatus in two strains of the open ocean diatom *Thalassiosira oceanica*. *PLoS One*, **12**:1–34.

Hoffmann, L.J., Breitbarth, E., Boyd, P.W., & Hunter, K.A. (2012) Influence of ocean warming and acidification on trace metal biogeochemistry. *Mar. Ecol. Prog. Ser.*, **470**:191–205.

Hopes, A., Nekrasov, V., Kamoun, S., & Mock, T. (2016) Editing of the urease gene by CRISPR-Cas in the diatom *Thalassiosira pseudonana*. *Plant Methods*, **12**:1–12.

Huang, W. & Daboussi, F. (2017) Genetic and metabolic engineering in diatoms. *Philos. Trans. R. Soc. B Biol. Sci.*, **372**.

Ifuku, K., Yan, D., Miyahara, M., Inoue-Kashino, N., Yamamoto, Y.Y., & Kashino, Y. (2015) A stable and efficient nuclear transformation system for the diatom *Chaetoceros gracilis*. *Photosynth. Res.*, **123**:203–211.

Ilbert, M. & Bonnefoy, V. (2013) Insight into the evolution of the iron oxidation pathways. *Biochim. Biophys. Acta - Bioenerg.*, **1827**:161–175.

Ioffe, E. & Stanley, P. (1994) Mice lacking N-acetylglucosaminyltransferase I activity die at mid-gestation, revealing an essential role for complex or hybrid N-linked carbohydrates. *Proc. Natl. Acad. Sci. U. S. A.*, **91**:728–732.

Ivanov, R. & Bauer, P. (2017) Sequence and coexpression analysis of iron-regulated ZIP transporter genes reveals crossing points between iron acquisition strategies in green algae and land plants. *Plant Soil*, **418**:61–73.

Johansson, O.N., Töpel, M., Pinder, M.I.M., Kourtchenko, O., Blomberg, A., Godhe, A., & Clarke, A.K. (2019a) *Skeletonema marinoi* as a new genetic model for marine chain-forming diatoms. *Sci. Rep.*, **9**:5391.

Johansson, O.N., Töpel, M., Pinder, M.I.M., Kourtchenko, O., Blomberg, A., Godhe, A., & Clarke, A.K. (2019b) *Skeletonema marinoi* as a new genetic model for marine chain-forming diatoms. *Sci. Rep.*, **9**:5391.

John H. Martin & Fitzwater, S. (1988) Iron deficiency limits phytoplankton growth in the north-east Pacific subarctic. *Nature*, **331**:343–345.

Kane, M.D. (2000) Assessment of the sensitivity and specificity of oligonucleotide (50mer) microarrays. *Nucleic Acids Res.*, **28**:4552–4557.

Kanehisa, M. & Goto, S. (2000) KEGG: Kyoto Encyclopedia of Genes and Genomes. *Nucleic Acids Res.*, **28**:27–30.

Karas, B.J., Diner, R.E., Lefebvre, S.C., McQuaid, J., Phillips, A.P.R., Noddings, C.M., Brunson, J.K., Valas, R.E., Deerinck, T.J., Jablanovic, J., Gillard, J.T.F., Beeri, K., Ellisman, M.H., Glass, J.I., Hutchison, C.A., Smith, H.O., Venter, J.C., Allen, A.E., Dupont, C.L., & Weyman, P.D. (2015) Designer diatom episomes delivered by bacterial conjugation. *Nat. Commun.*, **6**:1–10.

Kazamia, E., Sutak, R., Paz-Yepes, J., Dorrell, R.G., Vieira, F.R.J., Mach, J., Morrissey, J., Leon, S., Lam, F., Pelletier, E., Camadro, J.M., Bowler, C., & Lesuisse, E. (2018) Endocytosis-mediated siderophore uptake as a strategy for Fe acquisition in diatoms. *Sci. Adv.*, **4**:eaar4536.

Keeling, P.J., Burki, F., Wilcox, H.M., Allam, B., Allen, E.E., Amaral-Zettler, L.A., Armbrust, E.V., Archibald, J.M., Bharti, A.K., Bell, C.J., Beszteri, B., Bidle, K.D., Cameron, C.T., Campbell, L., Caron, D.A., Cattolico, R.A., Collier, J.L., Coyne, K., Davy, S.K., Deschamps, P., Dyhrman, S.T., Edvardsen, B., Gates, R.D., Gobler, C.J., Greenwood, S.J., Guida, S.M., Jacobi, J.L., Jakobsen, K.S., James, E.R., Jenkins, B., John, U., Johnson, M.D., Juhl, A.R., Kamp, A., Katz, L.A., Kiene, R.,

Kudryavtsev, A., Leander, B.S., Lin, S., Lovejoy, C., Lynn, D., Marchetti, A., McManus, G., Nedelcu, A.M., Menden-Deuer, S., Miceli, C., Mock, T., Montresor, M., Moran, M.A., Murray, S., Nadathur, G., Nagai, S., Ngam, P.B., Palenik, B., Pawlowski, J., Petroni, G., Piganeau, G., Posewitz, M.C., Rengefors, K., Romano, G., Rumpho, M.E., Rynearson, T., Schilling, K.B., Schroeder, D.C., Simpson, A.G.B., Slamovits, C.H., Smith, D.R., Smith, G.J., Smith, S.R., Sosik, H.M., Stief, P., Theriot, E., Twary, S.N., Umale, P.E., Vaulot, D., Wawrik, B., Wheeler, G.L., Wilson, W.H., Xu, Y., Zingone, A., & Worden, A.Z. (2014) The Marine Microbial Eukaryote Transcriptome Sequencing Project (MMETSP): Illuminating the Functional Diversity of Eukaryotic Life in the Oceans through Transcriptome Sequencing. *PLoS Biol.*, **12**:e1001889.

Kenshi, K. (2001) Biogeochemistry of iron in seawater. *Chikyu.Ac.Jp*, 93–102.

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

Kim, S.A. & Guerinot, M. Lou (2007) Mining iron: Iron uptake and transport in plants. *FEBS Lett.*, **581**:2273–2280.

Kim, S.A., Punshon, T., Lanzirotti, A., Li, A., Alonso, J.M., Ecker, J.R., Kaplan, J., & Guerinot, M. Lou (2006) Localization of iron in Arabidopsis seed requires the vacuolar membrane transporter VIT1. *Science (80-.).*, **314**:1295–1298.

Kindle, K.L., Schnell, R.A., Fermindez, E., & Lefebvre, P.A. (1989) Stable Nuclear Transformation of *Chlamydomonas* Using the *Chlamydomonas* Gene for Nitrate Reductase. *Cell*, **109**:2589–2601.

Koay, M.S., Antonkine, M.L., Gärtner, W., & Lubitz, W. (2008) Modelling Low-Potential [Fe 4 S 4] Clusters in Proteins **5**:1571–1587.

Kosman, D.J. (2003) Molecular mechanisms of iron uptake in fungi. *Mol. Microbiol.*,.

Kowarik, M., Young, N.M., Numao, S., Schulz, B.L., Hug, I., Callewaert, N., Mills, D.C., Watson, D.C., Hernandez, M., Kelly, J.F., Wacker, M., & Aebi, M. (2006) Definition of the bacterial N-glycosylation site consensus sequence. *EMBO J.*, **25**:1957–1966.

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

Kumar, S., Stecher, G., & Tamura, K. (2016) MEGA7: Molecular Evolutionary Genetics Analysis Version 7.0 for Bigger Datasets. *Mol. Biol. Evol.*, **33**:1870–1874.

Kung, L.A., Tao, S.C., Qian, J., Smith, M.G., Snyder, M., & Zhu, H. (2009) Global analysis of the glycoproteome in *Saccharomyces cerevisiae* reveals new roles for protein glycosylation in eukaryotes. *Mol. Syst. Biol.*, **5**:1–14.

Kustka, A.B. & Allen, A.E. (2007) SEQUENCE ANALYSIS AND TRANSCRIPTIONAL REGULATION OF IRON ACQUISITION GENES IN TWO MARINE DIATOMS. *J. Phycol.*, **729**:715–729.

Kwok, E.Y., Severance, S., & Kosman, D.J. (2006) Evidence for Iron Channeling in the Fet3p − Ftr1p High-Affinity Iron Uptake Complex in the Yeast Plasma Membrane. *Biochemistry*, **45**:6317–6327.

La Roche, J., Boyd, P.W., McKay, R.M.L., & Geider, R.J. (1996) Flavodoxin as an in situ marker for iron stress in phytoplankton. *Nature*, **382**:802–805.

La Roche, J., Geider, R.J., Graziano, L.M., Murray, H., & Lewis, K. (1993) Induction of specific proteins in eukaryotic algae grown under iron-, phosphorus-, or nitrogen-deficient conditions. *J. Phycol.*, **29**:767–777.

La Roche, J., Murray, H., Orellana, M., & Newton, J. (1995) FLAVODOXIN EXPRESSION AS AN INDICATOR OF IRON LIMITATION IN MARINE DIATOMS. *J. Phycol.*, **31**:520–530.

Laird, K.R., Kingsbury, M. V., & Cumming, B.F. (2010) Diatom habitats, species diversity and water-depth inference models across surface-sediment transects in Worth Lake, northwest Ontario, Canada. *J. Paleolimnol.*, **44**:1009–1024.

Lampe, R.H., Mann, E.L., Cohen, N.R., Till, C.P., Thamatrakoln, K., Brzezinski, M.A., Bruland, K.W., Twining, B.S., & Marchetti, A. (2018) Different iron storage strategies among bloom-forming diatoms. *Proc. Natl. Acad. Sci.*, **115**:E12275–E12284.

Landry, M.R., Constantinou, J., Latasa, M., Brown, S.L., Bidigare, R.R., & Ondrusek, M.E. (2000) Biological response to iron fertilization in the eastern equatorial Pacific (IronEx II). III. Dynamics of phytoplankton growth and microzooplankton grazing. *Mar. Ecol. Prog. Ser.*, **201**:57–72.

Lannoo, N. & Damme, E.J.M. Van (2015) Plant Science Review / N -glycans : The making of a varied toolbox. *Plant Sci.*, **239**:67–83.

Lau, C.K.Y., Krewulak, K.D., & Vogel, H.J. (2016) Bacterial ferrous iron transport: The Feo system. *FEMS Microbiol. Rev.*, **40**:273–298.

Lehtimäki, N., Koskela, M.M., & Mulo, P. (2015) Posttranslational Modifications of Chloroplast Proteins: An Emerging Field. *Plant Physiol.*, **168**:768–775.

Lelandais, G., Scheiber, I., Paz-Yepes, J., Lozano, J.C., Botebol, H., Pilátová, J., Žárský, V., Léger, T., Blaiseau, P.L., Bowler, C., Bouget, F.Y., Camadro, J.M., Sutak, R., & Lesuisse, E. (2016) *Ostreococcus tauri* is a new model green alga for studying iron metabolism in eukaryotic phytoplankton. *BMC Genomics*, **17**:319.

Lerouxel, O., Mouille, G., Andème-Onzighi, C., Bruyant, M.P., Séveno, M., Loutelier-Bourhis, C., Driouich, A., Höfte, H., & Lerouge, P. (2005) Mutants in DEFECTIVE GLYCOSYLATION, an *Arabidopsis* homolog of an oligosaccharyltransferase complex subunit, show protein underglycosylation and defects in cell differentiation and growth. *Plant J.*, **42**:455–468.

Letunic, I. & Bork, P. (2016) Interactive tree of life (iTOL) v3: an online tool for the display and annotation of phylogenetic and other trees. *Nucleic Acids Res.*,

44:W242–W245.

Leventhal, G.E., Ackermann, M., & Schiessl, K.T. (2019) Why microbes secrete molecules to modify their environment: the case of iron-chelating siderophores. *J. R. Soc. Interface*, **16**:20180674.

Levy-Ontman, O., Arad, S., Harvey, D.J., Parsons, T.B., Fairbanks, A., & Tekoah, Y. (2011) Unique N-glycan moieties of the 66-kDa cell wall glycoprotein from the red microalga *Porphyridium sp. J. Biol. Chem.*, **286**:21340–21352.

Levy-Ontman, O., Fisher, M., Shotland, Y., Weinstein, Y., Tekoah, Y., & Arad, S.M. (2014) Genes involved in the endoplasmic reticulum N-Glycosylation pathway of the red microalga Porphyridium sp.: A bioinformatic study. *Int. J. Mol. Sci.*, **15**:2305–2326.

Lohan, M.C., Buck, K.N., & Sander, S.G. (2015) Organic ligands — A key control on trace metal biogeochemistry in the oceans. *Mar. Chem.*, **173**:1–2.

Lommer, M., Specht, M., Roy, A.-S., Kraemer, L., Andreson, R., Gutowska, M.A., Wolf, J., Bergner, S. V, Schilhabel, M.B., Klostermeier, U.C., Beiko, R.G., 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.

Luxem, K.E., Ellwood, M.J., & Strzepek, R.F. (2017) Intraspecific variability in *Phaeocystis antarctica's* response to iron and light stress. *PLoS One*, **12**:1–15.

Lyons, T.W., Reinhard, C.T., & Planavsky, N.J. (2014) The rise of oxygen in Earth's early ocean and atmosphere. *Nature*, **506**:307–315.

Maldonado, M.T., Allen, A.E., Chong, J.S., Lin, K., Leus, D., Karpenko, N., & Harris, S.L. (2006) Copper-dependent iron transport in coastal and oceanic diatoms. *Limnol. Oceanogr.*, **51**:1729–1743.

Maldonado, M.T. & Price, N.M. (1999) Utilization of iron bound to strong organic ligands by plankton communities in the subarctic Pacific Ocean. *Deep. Res. Part II Top. Stud. Oceanogr.*, **46**:2447–2473.

Maldonado, M.T. & Price, N.M. (2001) Reduction and transport of organically bound iron by *Thalassiosira oceanica* (Bacillariophyceae). *J. Phycol.*, **37**:298–309.

Malviya, S., Scalco, E., Audic, S., Vincent, F., Veluchamy, A., Poulain, J., Wincker, P., Iudicone, D., de Vargas, C., Bittner, L., Zingone, A., & Bowler, C. (2016) Insights into global diatom distribution and diversity in the world's ocean. *Proc. Natl. Acad. Sci. U. S. A.*, **113**:E1516-25.

Mamedov, T. & Yusibov, V. (2011) Green algae *Chlamydomonas reinhardtii* possess endogenous sialylated N-glycans. *FEBS Open Bio*, **1**:15–22.

Marchetti, A. & Maldonado, M.T. (2016) Iron. In *The Physiology of Microalgae*. pp. 233–279.

Marchetti, A., Maldonado, M.T., Lane, E.S., & Harrison, P.J. (2006) Iron requirements of

the pennate diatom *Pseudo-nitzschia*: Comparison of oceanic (high-nitrate, low-chlorophyll waters) and coastal species. *Limnol. Oceanogr.*, **51**:2092–2101.

Marchetti, A., Moreno, C.M., Cohen, N.R., Oleinikov, I., deLong, K., Twining, B.S., Armbrust, E.V., & Lampe, R.H. (2017) Development of a molecular-based index for assessing iron status in bloom-forming pennate diatoms. *J. Phycol.*, **53**:820–832.

Marchetti, A., Parker, M.S., Moccia, L.P., Lin, E.O., Arrieta, A.L., Ribalet, F., Murphy, M.E.P., Maldonado, M.T., & Armbrust, E.V. (2009) Ferritin is used for iron storage in bloom-forming marine pennate diatoms. *Nature*, **457**:467–470.

Marchetti, A., Schruth, D.M., Durkin, C.A., Parker, M.S., Kodner, R.B., Berthiaume, C.T., Morales, R., Allen, A.E., & Armbrusta, E.V. (2012) Comparative metatranscriptomics identifies molecular bases for the physiological responses of phytoplankton to varying iron availability. *Proc. Natl. Acad. Sci. U. S. A.*, **109**:E317–E325.

Marchler-Bauer, A., Bo, Y., Han, L., He, J., Lanczycki, C.J., Lu, S., Chitsaz, F., Derbyshire, M.K., Geer, R.C., Gonzales, N.R., Gwadz, M., Hurwitz, D.I., Lu, F., Marchler, G.H., Song, J.S., Thanki, N., Wang, Z., Yamashita, R.A., Zhang, D., Zheng, C., Geer, L.Y., & Bryant, S.H. (2017) CDD/SPARCLE: Functional classification of proteins via subfamily domain architectures. *Nucleic Acids Res.*, **45**:D200–D203.

Martin, J.H. (1990) Glacial-interglacial CO2 change: The Iron Hypothesis. *Paleoceanography*, **5**:1–13.

Martin, J.H., Coale, K.H., Johnson, K.S., Fitzwater, S.E., Gordon, R.M., Tanner, S.J., Hunter, C.N., Elrod, V.A., Nowicki, J.L., Coley, T.L., Barber, R.T., Lindley, S., Watson, A.J., Van Scoy, K., Law, C.S., Liddicoat, M.I., Ling, R., Stanton, T., Stockel, J., Collins, C., Anderson, A., Bidigare, R., Ondrusek, M., Latasa, M., Millero, F.J., Lee, K., Yao, W., Zhang, J.Z., Friederich, G., Sakamoto, C., Chavez, F., Buck, K., Kolber, Z., Greene, R., Falkowski, P., Chisholm, S.W., Hoge, F., Swift, R., Yungel, J., Turner, S., Nightingale, P., Hatton, A., Liss, P., & Tindale, N.W. (1994) Testing the iron hypothesis in ecosystems of the equatorial Pacific Ocean. *Nature*, **371**:123–129.

Masuda, H., Shimochi, E., Hamada, T., Senoura, T., Kobayashi, T., Aung, M.S., Ishimaru, Y., Ogo, Y., Nakanishi, H., & Nishizawa, N.K. (2017) A new transgenic rice line exhibiting enhanced ferric iron reduction and phytosiderophore production confers tolerance to low iron availability in calcareous soil. *PLoS One*, **12**:1–21.

Mathieu-Rivet, E., Lerouge, P., Vanier, G., Plasson, C., Lucas, P.-L., Remy-Jouet, I., Tchi-Song, P.C., Vanier, J., Richard, V., Walet-Balieu, M.-L., Bernard, S., Driouich, A., Loutelier-Bourhis, C., Afonso, C., & Bardor, M. (2017) Heterologous expression of the N-acetylglucosaminyltransferase I dictates a reinvestigation of the N-glycosylation pathway in *Chlamydomonas reinhardtii*. *Sci. Rep.*, **7**:1–12.

Mathieu-Rivet, E., Scholz, M., Arias, C., Dardelle, F., Schulze, S., Le Mauff, F., Teo, G., Hochmal, A.K., Blanco-Rivero, A., Loutelier-Bourhis, C., Kiefer-Meyer, M.C.,

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.

McQuaid, J.B., Kustka, A.B., Oborník, M., Horák, A., McCrow, J.P., Karas, B.J., Zheng, H., Kindeberg, T., Andersson, A.J., Barbeau, K.A., & Allen, A.E. (2018) Carbonate-sensitive phytotransferrin controls high-affinity iron uptake in diatoms. *Nature*, **555**:534–537.

Milner, M.J., Seamon, J., Craft, E., & Kochian, L. V. (2012) Transport properties of members of the ZIP family in plants and their role in Zn and Mn homeostasis. *J. Exp. Bot.*, **64**:396–381.

Minskaia, E., Nicholson, J., & Ryan, M.D. (2013) Optimisation of the foot-and-mouth disease virus 2A co-expression system for biomedical applications. *BMC Biotechnol.*, **13**:67.

Miyagawa-Yamaguchi, A., Okami, T., Kira, N., Yamaguchi, H., Ohnishi, K., & Adachi, M. (2011) Stable nuclear transformation of the diatom *Chaetoceros sp. Phycol. Res.*, **59**:113–119.

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.

Mock, T., Otillar, R.P., Strauss, J., McMullan, M., Paajanen, P., Schmutz, J., Salamov, A., Sanges, R., Toseland, A., Ward, B.J., Allen, A.E., Dupont, C.L., Frickenhaus, S., Maumus, F., Veluchamy, A., Wu, T., Barry, K.W., Falciatore, A., Ferrante, M.I., Fortunato, A.E., Glöckner, G., Gruber, A., Hipkin, R., Janech, M.G., Kroth, P.G., Leese, F., Lindquist, E.A., Lyon, B.R., Martin, J., Mayer, C., Parker, M., Quesneville, H., Raymond, J.A., Uhlig, C., Valas, R.E., Valentin, K.U., Worden, A.Z., Armbrust, E.V., Clark, M.D., Bowler, C., Green, B.R., Moulton, V., Van Oosterhout, C., & Grigoriev, I. V. (2017) Evolutionary genomics of the cold-Adapted diatom *Fragilariopsis cylindrus*. *Nature*, **541**:536–540.

Mócsai, R., Figl, R., Troschl, C., Svehla, E., Windwarder, M., Thader, A., Strasser, R., & Altmann, F. (2019) N-glycans of the microalga *Chlorella vulgaris* are of the oligomannosidic type but highly methylated. *Sci. Rep.*, **9**:2–9.

Mohorko, E., Glockshuber, R., & Aebi, M. (2011) Oligosaccharyltransferase: The central enzyme of N-linked protein glycosylation. *J. Inherit. Metab. Dis.*, **34**:869–878.

Moore, C.M., Mills, M.M., Arrigo, K.R., Berman-Frank, I., Bopp, L., Boyd, P.W., Galbraith, E.D., Geider, R.J., Guieu, C., Jaccard, S.L., Jickells, T.D., La Roche, J., Lenton, T.M., Mahowald, N.M., Marañón, E., Marinov, I., Moore, J.K., Nakatsuka, T., Oschlies, A., Saito, M.A., Thingstad, T.F., Tsuda, A., & Ulloa, O. (2013) Processes and patterns of oceanic nutrient limitation. *Nat. Geosci.*,.

Morel, F.M.M. & Price, N.M. (2003) The biogeochemical cycles of trace metals in the oceans. *Science (80-.).*, **300**:944–947.

Moremen, K.W., Tiemeyer, M., & Nairn, A. V. (2012) Vertebrate protein glycosylation: Diversity, synthesis and function. *Nat. Rev. Mol. Cell Biol.*, **13**:448–462.

Morrissey, J. & Bowler, C. (2012) Iron utilization in marine cyanobacteria and eukaryotic algae. *Front. Microbiol.*, **3**:1–13.

Morrissey, J., Sutak, R., Paz-Yepes, J., Tanaka, A., Moustafa, A., Veluchamy, A., Thomas, Y., Botebol, H., Bouget, F.Y., McQuaid, J.B., Tirichine, L., Allen, A.E., Lesuisse, E., & Bowler, C. (2015) A novel protein, ubiquitous in marine phytoplankton, concentrates iron at the cell surface and facilitates uptake. *Curr. Biol.*, **25**:364–371.

Muñoz, C.F., de Jaeger, L., Sturme, M.H.J., Lip, K.Y.F., Olijslager, J.W.J., Springer, J., Wolbert, E.J.H., Martens, D.E., Eggink, G., Weusthuis, R.A., & Wijffels, R.H. (2018) Improved DNA/protein delivery in microalgae – A simple and reliable method for the prediction of optimal electroporation settings. *Algal Res.*, **33**:448–455.

Muto, M., Fukuda, Y., Nemoto, M., Yoshino, T., Matsunaga, T., & Tanaka, T. (2013) Establishment of a Genetic Transformation System for the Marine Pennate Diatom *Fistulifera sp.* Strain JPCC DA0580-A High Triglyceride Producer. *Mar. Biotechnol.*, **15**:48–55.

Nagashima, Y., von Schaewen, A., & Koiwa, H. (2018) Function of N-glycosylation in plants. *Plant Sci.*, **274**:70–79.

Nanjo, Y., Oka, H., Ikarashi, N., Kaneko, K., Kitajima, A., Mitsui, T., Munoz, F.J., Rodriguez-Lopez, M., Baroja-Fernandez, E., & Pozueta-Romero, J. (2006) Rice Plastidial N-Glycosylated Nucleotide Pyrophosphatase/Phosphodiesterase Is Transported from the ER-Golgi to the Chloroplast through the Secretory Pathway. *Plant Cell Online*, **18**:2582–2592.

Narayanan, N.N., Ihemere, U., Chiu, W., Siritunga, D., Rajamani, S., Singh, S., Oda, S., & Sayre, R.T. (2011) The iron assimilatory protein, FEA1, from Chlamydomonas reinhardtii facilitates iron-specific metal uptake in yeast and plants. *Front. Plant Sci.*, **2**:1–13.

Naumann, B., Busch, A., Allmer, J., Ostendorf, E., Zeller, M., Kirchhoff, H., & Hippler, M. (2007) Comparative quantitative proteomics to investigate the remodeling of bioenergetic pathways under iron deficiency in *Chlamydomonas reinhardtii*. *Proteomics*, **7**:3964–3979.

Nevo, Y. & Nelson, N. (2006) The NRAMP family of metal-ion transporters. *Biochim. Biophys. Acta - Mol. Cell Res.*,.

Nielsen, M., Lundegaard, C., Lund, O., & Petersen, T.N. (2010) CPHmodels-3.0-remote homology modeling using structure-guided sequence profiles. *Nucleic Acids Res.*, **38**:W576–W581.

Niu, Y.F., Yang, Z.K., Zhang, M.H., Zhu, C.C., Yang, W.D., Liu, J.S., & Li, H.Y. (2012) Transformation of diatom *Phaeodactylum tricornutum* by electroporation and

establishment of inducible selection marker. *Biotechniques*, **52**:1–3.

Nuester, J., Vogt, S., & Twining, B.S. (2012) Localization of iron within centric diatoms of the genus Thalassiosira. *J. Phycol.*, **48**:626–634.

Nunn, B.L., Faux, J.F., Hippmann, A.A., Maldonado, M.T., Harvey, H.R., Goodlett, D.R., Boyd, P.W., & Strzepek, R.F. (2013) Diatom Proteomics Reveals Unique Acclimation Strategies to Mitigate Fe Limitation. *PLoS One*, **8**:e75653.

Nymark, M., Sharma, A.K., Sparstad, T., Bones, A.M., & Winge, P. (2016) A CRISPR/Cas9 system adapted for gene editing in marine algae. *Sci. Rep.*, **6**:6–11.

Palenik, B., Grimwood, J., Aerts, A., Rouze, P., Salamov, A., Putnam, N., Dupont, C., Jorgensen, R., Derelle, E., Rombauts, S., Zhou, K., Otillar, R., Merchant, S.S., Podell, S., Gaasterland, T., Napoli, C., Gendler, K., Manuell, A., Tai, V., Vallon, O., Piganeau, G., Jancek, S., Heijde, M., Jabbari, K., Bowler, C., Lohr, M., Robbens, S., Werner, G., Dubchak, I., Pazour, G.J., Ren, Q., Paulsen, I., Delwiche, C., Schmutz, J., Rokhsar, D., Van de Peer, Y., Moreau, H., & Grigoriev, I. V. (2007) The tiny eukaryote *Ostreococcus* provides genomic insights into the paradox of plankton speciation. *Proc. Natl. Acad. Sci.*, **104**:7705–7710.

Pankowski, A. & Mcminn, A. (2008) Iron availability regulates growth, photosynthesis, and production of ferredoxin and flavodoxin in Antarctic sea ice diatoms. *Aquat. Biol.*, **4**:273–288.

Pausch, F., Bischof, K., & Trimborn, S. (2019) Iron and manganese co-limit growth of the Southern Ocean diatom Chaetoceros debilis 1–16.

Paz, Y., Katz, A., & Pick, U. (2007) A multicopper ferroxidase involved in iron binding to transferrins in *Dunaliella salina* plasma membranes. *J. Biol. Chem.*, **282**:8658–8666.

Peers, G. & Price, N.M. (2004) A role for manganese in superoxide dismutases and growth of iron-deficient diatoms. *Limnol. Oceanogr.*, **49**:1774–1783.

Peers, G. & Price, N.M. (2006) Copper-containing plastocyanin used for electron transport by an oceanic diatom. *Nature*, **441**:341–344.

Peers, G., Quesnel, S.-A., & Price, N.M. (2005) Copper requirements for iron acquisition and growth of coastal and oceanic diatoms. *Limnol. Oceanogr.*, **50**:1149–1158.

Petersen, T.N., Brunak, S., von Heijne, G., & Nielsen, H. (2011) SignalP 4.0: discriminating signal peptides from transmembrane regions. *Nat. Methods*, **8**:785–786.

Petrou, K., Trimborn, S., Rost, B., Ralph, P.J., & Hassler, C.S. (2014) The impact of iron limitation on the physiology of the Antarctic diatom Chaetoceros simplex. *Mar. Biol.*, **161**:925–937.

Pfaffen, S., Abdulqadir, R., Le Brun, N.E., & Murphy, M.E.P. (2013) Mechanism of ferrous iron binding and oxidation by ferritin from a pennate diatom. *J. Biol. Chem.*, **288**:14917–14925.

Pfaffen, S., Bradley, J.M., Abdulqadir, R., Firme, M.R., Moore, G.R., Brun, N.E.L., & Murphy, M.E.P. (2015) A diatom ferritin optimized for iron oxidation but not iron storage. *J. Biol. Chem.*, **290**:28416–28427.

Pollard, R.T., Salter, I., Sanders, R.J., Lucas, M.I., Moore, C.M., Mills, R.A., Statham, P.J., Allen, J.T., Baker, A.R., Bakker, D.C.E., Charette, M.A., Fielding, S., Fones, G.R., French, M., Hickman, A.E., Holland, R.J., Hughes, J.A., Jickells, T.D., Lampitt, R.S., Morris, P.J., Nédélec, F.H., Nielsdóttir, M., Planquette, H., Popova, E.E., Poulton, A.J., Read, J.F., Seeyave, S., Smith, T., Stinchcombe, M., Taylor, S., Thomalla, S., Venables, H.J., Williamson, R., & Zubkov, M. V. (2009) Southern Ocean deep-water carbon export enhanced by natural iron fertilization. *Nature*, **457**:577–580.

Poulsen, N., Chesley, P.M., & Kröger, N. (2006) Molecular genetic manipulation of the diatom *Thalassiosira pseudonana* (Bacillariophyceae). *J. Phycol.*, **42**:1059–1065.

Price, N.M., Harrison, G., Hering, J., Hudson, R., Nirel, P., Palenik, B., & Morel, F.M. (1988) Preparation and chemistry of the artificial algal culture medium aquil. *Biol. Oceanogr.*, **6**:443–461.

R Development Core Team (2015) R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL http://www.R-project.org/. *R Found. Stat. Comput. Vienna, Austria.*,.

Radakovits, R., Jinkerson, R.E., Darzins, A., & Posewitz, M.C. (2010) Genetic Engineering of Algae for Enhanced Biofuel Production. *Eukaryot. Cell*, **9**:486–501.

Ratten, J.M., LaRoche, J., Desai, D.K., Shelley, R.U., Landing, W.M., Boyle, E., Cutter, G.A., & Langlois, R.J. (2015) Sources of iron and phosphate affect the distribution of diazotrophs in the North Atlantic. *Deep. Res. Part II Top. Stud. Oceanogr.*, **116**:332–341.

Raven, J.A., Evans, M.C.W., & Korb, R.E. (1999) The role of trace metals in photosynthetic electron transport in O2-evolving organisms. *Photosynth. Res.*, **60**:111–150.

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

Read, B.A., Kegel, J., Klute, M.J., Kuo, A., Lefebvre, S.C., Maumus, F., Mayer, C., Miller, J., Monier, A., Salamov, A., Young, J., Aguilar, M., Claverie, J.M., Frickenhaus, S., Gonzalez, K., Herman, E.K., Lin, Y.C., Napier, J., Ogata, H., Sarno, A.F., Shmutz, J., Schroeder, D., De Vargas, C., Verret, F., Von Dassow, P., Valentin, K., Van De Peer, Y., Wheeler, G., Dacks, J.B., Delwiche, C.F., Dyhrman, S.T., Glöckner, G., John, U., Richards, T., Worden, A.Z., Zhang, X., Grigoriev, I. V., Allen, A.E., Bidle, K., Borodovsky, M., Bowler, C., Brownlee, C., Mark Cock, J., Elias, M., Gladyshev, V.N., Groth, M., Guda, C., Hadaegh, A., Iglesias-Rodriguez, M.D., Jenkins, J., Jones, B.M., Lawson, T., Leese, F., Lindquist, E., Lobanov, A., Lomsadze, A., Malik, S.B., Marsh, M.E., MacKinder, L., Mock, T.,

Mueller-Roeber, B., Pagarete, A., Parker, M., Probert, I., Quesneville, H., Raines, C., Rensing, S.A., Riaño-Pachón, D.M., Richier, S., Rokitta, S., Shiraiwa, Y., Soanes, D.M., Van Der Giezen, M., Wahlund, T.M., Williams, B., Wilson, W., Wolfe, G., & Wurch, L.L. (2013) Pan genome of the phytoplankton *Emiliania* underpins its global distribution. *Nature*, **499**:209–213.

Rochaix, J.D. & Van Dillewijn, J. (1982) Transformation of the green alga *Chlamydomonas reinhardii* with yeast DNA. *Nature*, **296**:70–72.

Rogan, N., Achterberg, E.P., Le Moigne, F.A.C., Marsay, C.M., Tagliabue, A., & Williams, R.G. (2016) Volcanic ash as an oceanic iron source and sink. *Geophys. Res. Lett.*, **43**:2732–2740.

Sabatino, V., Russo, M.T., Patil, S., D'Ippolito, G., Fontana, A., & Ferrante, M.I. (2015) Establishment of Genetic Transformation in the Sexually Reproducing Diatoms *Pseudo-nitzschia multistriata* and *Pseudo-nitzschia arenysensis* and Inheritance of the Transgene. *Mar. Biotechnol.*, **17**:452–462.

Sakurai, T. & Kataoka, K. (2007) Structure and function of type I copper in multicopper oxidases. *Cell. Mol. Life Sci.*, **64**:2642–2656.

Samperio-Ramos, G., Santana Casiano, J.M., & González Dávila, M. (2016) Effect of ocean warming and acidification on the Fe(II) oxidation rate in oligotrophic and eutrophic natural waters. *Biogeochemistry*, **128**:19–34.

Samukawa, M., Shen, C., Hopkinson, B.M., & Matsuda, Y. (2014) Localization of putative carbonic anhydrases in the marine diatom, *Thalassiosira pseudonana*. *Photosynth. Res.*, **121**:235–249.

Sanchez, N., Brown, E.A., Olsen, Y., Vadstein, O., Iriarte, J.L., Gonzalez, H.E., & Ardelan, M. Van (2018) Effect of Siderophore on Iron Availability in a Diatom and a Dinoflagellate Species: Contrasting Response in Associated Bacteria. *Front. Mar. Sci.*, **5**:1–15.

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

Scheys, F., Van Damme, E.J.M., De Schutter, K., Staes, A., Gevaert, K., & Smagghe, G. (2018) Evolutionarily conserved and species-specific glycoproteins in the N-glycoproteomes of diverse insect species. *Insect Biochem. Mol. Biol.*, **100**:22–29.

Schröder, I., Johnson, E., & De Vries, S. (2003) Microbial ferric iron reductases. *FEMS Microbiol. Rev.*, **27**:427–447.

Schulze, S., Urzica, E., Reijnders, M.J.M.F., van de Geest, H., Warris, S., Bakker, L. V., Fufezan, C., Martins dos Santos, V.A.P., Schaap, P.J., Peters, S.A., & Hippler, M. (2017) Identification of methylated GnTI-dependent N-glycans in *Botryococcus brauni*. *New Phytol.*, **215**:1361–1369.

Severance, S., Chakraborty, S., & Kosman, D.J. (2004) The Ftr1p iron permease in the yeast plasma membrane: orientation, topology and structure-function relationships.

Biochem. J., **380**:487–496.

Seymour, J.R., Amin, S.A., Raina, J.B., & Stocker, R. (2017) Zooming in on the phycosphere: The ecological interface for phytoplankton-bacteria relationships. *Nat. Microbiol.*, **2**:17065.

Shaked, Y., Kustka, A.B., & Morel, F.M.M. (2005) A general moel for iron acquasition by eukaryotic phytoplankton. *Limnol. Oceanogr.*, **50**:872–882.

Shi, D., Xu, Y., Hopkinson, B.M., & Morel, F.M.M. (2010) Effect of ocean acidification on iron availability to marine phytoplankton. *Science*, **327**:676–679.

Shu, N., Elofsson, A., Tsirigos, K.D., Peters, C., & Käll, L. (2015) The TOPCONS web server for consensus prediction of membrane protein topology and signal peptides. *Nucleic Acids Res.*, **43**:W401–W407.

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.

Siegmund, D.O.. & Zhang, H. (1994) Confidence Regions in Broken Line Regression. *Lect. Notes-Monograph Ser.*, **23**:292–316.

Smetacek, V. (2018) Seeing is Believing: Diatoms and the Ocean Carbon Cycle Revisited. *Protist*, **169**:791–802.

Smetacek, V., Klaas, C., Strass, V.H., Assmy, P., Montresor, M., Cisewski, B., Savoye, N., Webb, A., D'Ovidio, F., Arrieta, J.M., Bathmann, U., Bellerby, R., Berg, G.M., Croot, P., Gonzalez, S., Henjes, J., Herndl, G.J., Hoffmann, L.J., Leach, H., Losch, M., Mills, M.M., Neill, C., Peeken, I., Röttgers, R., Sachs, O., Sauter, E., Schmidt, M.M., Schwarz, J., Terbrüggen, A., & Wolf-Gladrow, D. (2012) Deep carbon export from a Southern Ocean iron-fertilized diatom bloom. *Nature*, **487**:313–319.

Smith, S.R., Gillard, J.T.F., Kustka, A.B., McCrow, J.P., Badger, J.H., Zheng, H., New, A.M., Dupont, C.L., Obata, T., Fernie, A.R., & Allen, A.E. (2016) Transcriptional Orchestration of the Global Cellular Response of a Model Pennate Diatom to Diel Light Cycling under Iron Limitation. *PLoS Genet.*, **12**:1–39.

Soria-Dengg, S. & Horstmann, U. (1995) Ferrioxamines B and E as iron sources for the marine diatom Phaeodactylum tricornutum. *Mar. Ecol. Prog. Ser.*, **127**:269–277.

Strasser, R. (2016) Plant protein glycosylation. *Glycobiology*, **26**:926–939.

Strasser, R., Stadlmann, J., Glössl, J., Svoboda, B., Altmann, F., & Mach, L. (2005) Molecular basis of N -acetylglucosaminyltransferase I deficiency in *Arabidopsis thaliana* plants lacking complex N-glycans. *Biochem. J.*, **387**:385–391.

Strzepek, R.F. & Harrison, P.J. (2004) Photosynthetic architecture differs in coastal and oceanic diatoms. *Nature*, **431**:689–692.

Strzepek, R.F., Maldonado, M.T., Hunter, K.A., Frew, R.D., & Boyd, P.W. (2011) Adaptive strategies by Southern Ocean phytoplankton to lessen iron limitation:

Uptake of organically complexed iron and reduced cellular iron requirements. *Limnol. Oceanogr.*, **56**:1983–2002.

Stukenberg, D., Zauner, S., Dell'Aquila, G., & Maier, U.G. (2018) Optimizing CRISPR/cas9 for the diatom *Phaeodactylum tricornutum*. *Front. Plant Sci.*, **9**:740.

Sunda, W.G. & Huntsman, S.A. (1995) Iron uptake and growth limitation in oceanic and coastal phytoplankton. *Mar. Chem.*, **50**:189–206.

Sunda, W.G., Swift, D.G., & Huntsman, S.A. (1991) Low iron requirement for growth in oceanic phytoplankton. *Nature*, **351**:56–57.

Tabatabaei, M., Tohidfar, M., Jouzani, G.S., Safarnejad, M., & Pazouki, M. (2011) Biodiesel production from genetically engineered microalgae: Future of bioenergy in Iran. *Renew. Sustain. Energy Rev.*, **15**:1918–1927.

Tagliabue, A., Bowie, A.R., Boyd, P.W., Buck, K.N., Johnson, K.S., & Saito, M.A. (2017) The integral role of iron in ocean biogeochemistry. *Nature*, **543**:51–59.

Taunt, H.N., Stoffels, L., & Purton, S. (2018) Green biologics: The algal chloroplast as a platform for making biopharmaceuticals. *Bioengineered*, **9**:48–54.

Terzulli, A. & Kosman, D.J. (2010) Analysis of the high-affinity iron uptake system at the *Chlamydomonas reinhardtii* plasma membrane. *Eukaryot. Cell*, **9**:815–826.

Thomine, S. & Ji, S. (2015) Plant Metal Transporters with Homology to Proteins of the NRAMP Family Genomic Analysis of the NRAMP Family in Plant Species 1–9.

Thomine, S. & Lanquar, V. (2010) Iron Transport and Signaling in Plants. *Pumps Plant Signaling, Signal. Commun. Plants, eds Geisler M, Venema K (Springer, Berlin)*, 99–131.

Thomine, S., Lelièvre, F., Debarbieux, E., Schroeder, J.I., & Barbier-Brygoo, H. (2003) AtNRAMP3, a multispecific vacuolar metal transporter involved in plant responses to iron deficiency. *Plant J.*, **34**:685–695.

Tian, Y., Zhou, Y., Elliot, S., Aebersold, R., Zhang, H. (2007) Solid-phase Extraction of N-linked Glycopeptides. *Nat Protoc.*, **2**:334–339.

Tréguer, P., Bowler, C., Moriceau, B., Dutkiewicz, S., Gehlen, M., Aumont, O., Bittner, L., Dugdale, R., Finkel, Z., Iudicone, D., Jahn, O., Guidi, L., Lasbleiz, M., Leblanc, K., Levy, M., & Pondaven, P. (2018) Influence of diatom diversity on the ocean biological carbon pump. *Nat. Geosci.*, **11**:27–37.

Trick, C.G., Bill, B.D., Cochlan, W.P., Wells, M.L., Trainer, V.L., & Pickell, L.D. (2010) Iron enrichment stimulates toxic diatom production in high-nitrate, low-chlorophyll areas. *Proc. Natl. Acad. Sci.*, **107**:5887–5892.

Tripathy, S.C. & Jena, B. (2019) Iron-Stimulated Phytoplankton Blooms in the Southern Ocean: a Brief Review. *Remote Sens. Earth Syst. Sci.*,.

Tsang, H.-F., Xue, V.W., Koh, S.-P., Chiu, Y.-M., Ng, L.P.-W., & Wong, S.-C.C. (2017) NanoString, a novel digital color-coded barcode technology: current and future

applications in molecular diagnostics. *Expert Rev. Mol. Diagn.*, **17**:95–103.

Turnsek, J. & Dupont, C. (2017) Pour plating of Thalassiosira pseudonana (Tp) [WWW Document]. URL https://www.researchgate.net/publication/322255671_Pour_plating_of_Thalassiosira_pseudonana_Tp

Urzica, E.I., Casero, D., Yamasaki, H., Hsieh, S.I., Adler, L.N., Karpowicz, S.J., Blaby-Haas, C.E., Clarke, S.G., Loo, J.A., Pellegrini, M., & Merchanta, S.S. (2012) Systems and Trans-system level analysis identifies conserved iron deficiency responses in the plant lineage. *Plant Cell*, **24**:3921–3948.

Vanier, G., Hempel, F., Chan, P., Rodamer, M., Vaudry, D., Maier, U.G., Lerouge, P., & Bardor, M. (2015) Biochemical characterization of human anti-hepatitis b monoclonal antibody produced in the microalgae *Phaeodactylum tricornutum*. *PLoS One*, **10**:1–19.

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

Vanier, G., Stelter, S., Vanier, J., Hempel, F., Maier, U.G., Lerouge, P., Ma, J., & Bardor, M. (2018) Alga-Made Anti-Hepatitis B Antibody Binds to Human Fcγ Receptors. *Biotechnol. J.*, **13**:1700496.

Vert, G., Grotz, N., Dédaldéchamp, F., Gaymard, F., Guerinot, M. Lou, Briat, J.-F., & Curie, C. (2002) IRT1, an Arabidopsis transporter essential for iron uptake from the soil and for plant growth. *Plant Cell*, **14**:1223–1233.

Villar, E., Vannier, T., Vernette, C., Lescot, M., Cuenca, M., Alexandre, A., Bachelerie, P., Rosnet, T., Pelletier, E., Sunagawa, S., & Hingamp, P. (2018) The Ocean Gene Atlas: Exploring the biogeography of plankton genes online. *Nucleic Acids Res.*, **46**:W289–W295.

Waterhouse, A.M., Procter, J.B., Martin, D.M.A., Clamp, M., & Barton, G.J. (2009) Jalview Version 2-A multiple sequence alignment editor and analysis workbench. *Bioinformatics*, **25**:1189–1191.

Weinberg, E.D. (1989) Cellular Regulation of Iron Assimilation. *Q. Rev. Biol.*, **64**:261–290.

Welschmeyer, N.A. (1994) Fluorometric analysis of chlorophyll a in the presence of chlorophyll b and pheopigments. *Limnol. Oceanogr.*, **39**:1985–1992.

Westberry, T.K., Boss, E.S., O'Malley, R.T., Siegel, D.A., Wiggert, J.D., McClain, C.R., Feldman, G.C., Doney, S.C., Moore, J.K., Dall'Olmo, G., Milligan, A.J., & Mahowald, N. (2009) Satellite-detected fluorescence reveals global physiology of ocean phytoplankton . *Biogeosciences*, **6**:779–794.

Whitney, L.A.P., Lins, J.J., Hughes, M.P., Wells, M.L., Dreux Chappell, P., & Jenkins, B.D. (2011) Characterization of putative iron responsive genes as species-specific indicators of iron stress in thalassiosiroid diatoms. *Front. Microbiol.*, **2**:1–14.

Whittaker, S., Bidle, K.D., Kustka, A.B., & Falkowski, P.G. (2011) Quantification of nitrogenase in *Trichodesmium* IMS 101: Implications for iron limitation of nitrogen fixation intheocean. *Environ. Microbiol. Rep.*, **3**:54–58.

Wickham, H., Chang, W., Henry, L., Pedersen, T.L., Takahash, K., Wilke, C., & Woo, C. (2018) ggplot2: Create Elegant Data Visualisations Using the Grammar of Graphics. *R Packag. version 3.1.0*,.

Willis, A., Bowler, C., Eason-Hubbard, M., Wetherbee, R., Hodson, O., & Maheswari, U. (2014) Adhesion molecules from the diatom Phaeodactylum tricornutum (Bacillariophyceae): genomic identification by amino-acid profiling and in vivo analysis . *J. Phycol.*, **50**:837–849.

Wu, J. (2001) Soluble and Colloidal Iron in the Oligotrophic North Atlantic and North Pacific. *Science (80-.).*, **293**:847–849.

Wu, M., McCain, J.S.P., Rowland, E., Middag, R., Sandgren, M., Allen, A.E., & Bertrand, E.M. (2019) Manganese and iron deficiency in Southern Ocean Phaeocystis antarctica populations revealed through taxon-specific protein indicators. *Nat. Commun.*, **10**:3582.

Yan, N., Fan, C., Chen, Y., & Hu, Z. (2016) The potential for microalgae as bioreactors to produce pharmaceuticals. *Int. J. Mol. Sci.*, **17**:1–24.

Yoon, H.S., Hackett, J.D., Ciniglia, C., Pinto, G., & Bhattacharya, D. (2004) A Molecular Timeline for the Origin of Photosynthetic Eukaryotes. *Mol. Biol. Evol.*, **21**:809–818.

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.

Yusibov, V., Kushnir, N., & Streatfield, S.J. (2016) Antibody Production in Plants and Green Algae. *Annu. Rev. Plant Biol.*, **67**:669–701.

APPENDIX A: SUPPLEMENTAL TABLES

Supplemental Table 1: Full-length nucleotide sequences for all targeted transcripts.

This list includes the identifiers of the targeted transcriptome experiment. The probes were designed based on these sequences.

>Adhesin 1_THAOC_03112
>AP2_THAOC_08614_TF
>ATP-ADPtranslocase_THAOC_10028
>bzip_THAOC_00246_TF
>bzip_THAOC_06177_TF
>bzip_THAOC_20638_TF
>Calreticulin_THAOC_17167
>carbamoyl-phosphate-synthase_Urea-cycle genes _THAOC_01996
>CB_cell-surface-bacterial-protein-of-unknown-function_THAOC_33887
>CHIT_cell-surface-chitinase_THAOC_33534
>Cre-like recombinase_THAOC_12920
>CREGx2_THAOC_8512
>cytochromec-oxidase-subunitI_gi|347462428|gb|JP296424.1
>DNAj _house-keeping_Thaoc_13951
>DupDom_THAOC_19266
>FBA1_THAOC_00388
>FBA3_THAOC_02112
>FBA4_THAOC_24978
>FBA6_THAOC_24977
>FLDA1_THAOC_31152
>FRE1_THAOC_06605
>Glutamine synthase_gb|AGNL01046868.1
>GnT1_THAOC_02312
>Homeobox_THAOC_04964
>HSF_THAOC_16896_TF
>HSF2_THAOC_05090
>Hsp20a_THAOC_12165
>Hsp20g_Thaoc_21083
>Isocitratlyase_THAOC_35650
>Isocitratlyase_THAOC_35651
>ISIP1a_THAOC30955
>ISIP2_THAOC_21195
>ISIP2x8_THAOC_36747
>ISIP3_THAOC_34758
>Lipidtransporter1_THAOC_13660
>MUCOX2 THAOC_26314

>M-Protease_THAOC_07439
>Mitochondrial-glutamine-synthetase_classI_THAOC_33749
>Mn_SOD_THAOC_02860
>Myb_THAOC_35238_TF
>NADH-dehydrogenase-subunit7_ gi|347462926|gb|JP296922.1
>NRPS2_bacterial-non-ribosomal-peptidesynthetase_THAOC_26229
>Nuclear_import-export receptor_THAOC_05312-HK
>Nuclear-reverse-transcriptase_THAOC_00575
>OST_THAOC_24768
>PETF_Ferredoxin_THAOC_25559
>PETE_Plastocyanine_THAOC_06915
>RNA-binding-protein_THAOC_29934
>Sigma70_THAOC_16512_TF
>UCGT_THAOC_35806
>Urease_T.oceanica_ACB99416.1
>zf-CCCH_THAOC_16092_TF
>zf-TAZ_THAOC_27055_TF

Supplemental Table 2: List of target sequences for NanoString probes the LT-Run.

This table provides the 100 bp long target sequences that were used for the design of the NanoString probes.

THAOC Nr	Name	Sequence (100 bp)
JP296424	Cytochrome C oxidase	TATGGGCCCACCATATGTTTACTGTTGGATTAGATATAGATACAA GAGCTTACTTTACTGCGGCTACTATGATTATCGCGATCCCAACAG GAATTAAAAT
JP296922	NADH dehydrogenase	GGTTGCCACGCAATGGATGTCGGTGCTATGACACCTTTTTTGTGA GCTTTTGAAGAACGAGAAAAGCTAATGGAATTTTATGAACGAGT TTCAGGTGCGC
THAOC_00246	bzip TF	GCGACAATGAGCTGTACACAAAGGCTGCCGACGGTGCTGTTCAT GTATTTTTATCTTGCTTCATGGGCTTGTGTCAGAAGATGCCATCA AGTCCAGCGAT
THAOC_00575	nuclear reverse transcriptase	ATGTGGAATGAAGCAGTCTGAACTTGATCCCTGTCTGTTCATAGG CCCAAAAGTGATTTGCATTGTCTACGTTGACGACCTGCTATTTTG GGCGCGAGAC
THAOC_01996	carbamoyl-phosphate synthase Urea	AGTTCCATCCCGAGTACAGGAACGGACCAACTGATACCGAGTTC ATGTTTGACACTTTCCTCGCTGCCTGTAAGGATCCCAAGGCTCCA ATCCAGTTCCC
THAOC_02112	FBA3	CGTTCCGATCGTCGAGCCCGAGATCCTCATCGACGGAGACCATG ACATCGCCACAACTGCGAGGATTCAGGAGGTTGTTGTATCAAGG ACATACCAGAAG
THAOC_03112	Adhesin 1	CGACCTTTTCGACGAGGATCAATACAGAATGGTGCTTCGTACTA AGGGCGGGCTTTACGCCGGATTCAATCGCGCCAAAGGAGTCAAT GCGGGTGTCAAG
THAOC_04964	Homeobox TF	CTCACCATCCGATCATCCGTACCAGAGGAGCGAACCGTACTGGC CACCTTCAGGGCTCCCATTGAGTACACGGTCCCGTATGACATGA ACGACCGCAAGA
THAOC_05090	HSF2	GCCCGACGGGCTTCTACGGGATCTCTTCAACATCTCAAGCCGATC GTCATGAAGTTGAATGCGTCGCAAGGTGTATCTCCAACAGAAGA ACTCAACATAT

THAOC_05312	nuclear import-export receptor	CCTCATGGGTTATAGACAACGGCAGCTACGACGAAAAGCGAATG AAGATTCGTTCGATTGTGCAGCCGCTACTCTCTGAACTTGCGAAG ATGATCGTCAG
THAOC_06177	bzip TF	TCGTTACGAGCAAATGGAAGAACGAATGTCTCAAATGCAGGAAC GAATCGATACACTGACAAGGTCCTTGAGCGGTCGGGCCTTCGCA TCGTCGATTGAA
THAOC_06605	FRE1	TCCTGACAAGTCTCTGGAAGGCACGTGACAATGTCAACGCTGTC ACTATTTTGCATGCTCATGCAGTAGAGTCTTGGTACAAATCACTC TTCCAAAGGGG
THAOC_07439	M-Protease	TCAAAGGCATTATCCTTGGCCTCGTTAGCCGCTACTGCAATCGTT GGACTGGTATGCCTACTGTTAATAGAGTTCAACTGTGGTATTGTA GCAGGAGCCA
THAOC_08614	AP2 TF	GTGAGCATAACGCCGCCACTGCGTACAAAGTGGTCCACGATCTA TGTTCAACGTACGAGGAGTTTCCTTCTAGTGTTCAAGTCGAGCAG AATGTCAAAGC
THAOC_10028	ATP-ADP translocase	AATCGAAATCTCAGCTTCGACTCAAGCGACTTGCATTCGTAGGGT TCGGGGGCACCCTCGGCGGTATTATAGGAAGTGTCATCGCATCA TTTGCAGCGGA
THAOC_12165	HSP20a house-keping	GCGATAGTTCAGGGACATGCACCCGAGTTGACGACGACGGAGTC AAAACAAAAGAAGGCCACGGCTCCTCCAAAGGCCAACGAGATG GAAGATGCCAAGG
THAOC_12920	CRE	GTACGGCACGGTACTTCAGTCCGTCCTCATAAGGGCGTACAACTT GCGAATTACGTATCCCGATCGTGACATTTGTATTCACGCCAACGA CGTGGCGGGA
THAOC_13660	Lipid transporter 1 mito	CAAGTCTACCTTTGCGAATATGCTGTGCTGCGTGCAAAACCCAAC TGGCGGAGACATATTGGTAAATGGGAACTCGGTGACAACAGATC AATACACCGTG
THAOC_13951	DNAj protein house-keeping	ACGACCCCGAGGCCGCCGAGAAGTTCCAGAAGCTGGGACAGGC GTACCAGGTCCTGTCGAACGAGTCGAGCCGTGCCAACTACGACA AGAACGGCAAGCC
THAOC_16092	zf-CCCH TF	AGCAAGTACGAGAAGGCTCTAACTGCATTGGAAGAGAGTAGGA GGAAGCCCTACAACAGAAATGATCGAAATGACACCAAAGGTGC GTCAGGTGATCGAG

THAOC_16512	Sigma70 TF	GAGACCACCGACGTACCAGGAGTGGGCTGCATCAATCGATCCAG AAATGACGGTCATACAGCTCAAACGTCAGATCCGACGCAGTCTC CGGGCCAAGGCA
THAOC_16896	HSF TF	GGGAGCGGACGCGTTCCTGGATCGCATCGAGCAGTCGTCGGGCC AGCTCAACGATTTTCAAATGGGCAGCCTGCCCGAGGGTCTCGCG GATATGGCAGTT
THAOC_19266	DupDom	ATGCTTCTGCATTGGACGATCTCAGGGCCGAGTTGAAAGCCGAT GCGTTCGCCAATCTTGAATCATCACTGACAGTGCACCGCCAAAA GCACCAAGACGA
THAOC_20638	bzip TF	TCCGACTCTTTCGTATCCAGCAGCACCCTCGGCAAACAATTCACA GCTAGCGCAGAATTCAAATAAGGTGAAGAGGGAATCAGGTCCG AGCCTCCAATCG
THAOC_21083	Hsp20g house-keeping	CCGAGGTGGATACTGGCAAGATGACCGCGTCATTCTCGGATGGC ACACTCACAGTTGTCGCTCCTATGGTTGAGCCGACAGCTGGACTT GAGCAGGATCA
THAOC_21195	ISIP2	CGATGGTGTCGTCAGTTTGGAATTCGGGGACTACCAATACGCTTA CAACCCACTCGAGGGCAACAAGAATGGACGAACGCTCGCGACTA TTTCACAAATG
THAOC_24978	FBA4	CCTCGTGTTGTACGCGTCGTGGCCCTGTCAGGTGGTTACTCCAAG GACGTGGCGAACGAGAAACTTGCCAAGAATAGCGGCATCATCGC TAGCTTTTCAA
THAOC_26229	NRPS2	TGTCCAATGTTGCAAGCATCGACTCGGACTTCTTCGATATTGGCG GTGACAGTCTCGGCTCGATGAAATTAGCCCGAGCGATCCAGAAA GATCTCGGTGT
THAOC_26314	MUCOX2	TTACACCGTTTGGAACCGAGGCGGCGATCAGTACGGTGGATTTC CCTTGACGATATTTACAGAATTGCACCACGACAAAATCTGGAAT AAGAAGAGGCCG
THAOC_27055	zf-TAZ TF	CTTAATACCGAAATGAATAGCGAGCCACCGAGGAAGAAACCCA GTATTCTTTGCGGTGACATGACGGCCACTCCCTCTATGTCTATCA AGTCTGGTGTTT
THAOC_29934	cytosolic-nuclear zinc finger protein	TAAGATTCAACGCGATGACAGAAGAACGGATCTGCCGGGATCTA CTTCGAGAGAGATCAAGGTTGAGAAGATCTCCAATGGGTTTGGG TTTGGGCACTGT

THAOC_30821	zf-TAZ TF	CAACGGGGCAGTGTCGTCTTCACCGGACGAGGTAAATTCATCAA CAGTTGAACGTGGCGGAAAGTTTGCTGCCATGGAAAAGAAGCTG AAAAAGAACAGT
THAOC_30995	ISIP1a	CTACTGCCGCTGGAAATCGTCAAGTTGAATGGGCGATCGGAACA AACTATGTTTGGGTGAACGGCGGACAGACAGAGGAAGTGTACAT TGTTGAGATTGG
THAOC_31152	FLDA1	GAGTGCCGGCCTTTTCTACTCGACCCAGACTGGAAACACTGAAA CTGTCGCTGGATATTTGGCCGACGCTACGGGCCTCGAAATGAAG GACATCGGAGAC
THAOC_33534	CHIT cell surface chitinase	CTCTAACTGCAGGATGTCTCGACGGCACCACGCCTCCTCCACCTT CGCCGCCAAACACAACGTCCAACCCAACAAGCGGTGTCACCGCT CAGCCGACCAG
THAOC_33749	glutamine synthase	CGAGGTAGTTTTGGGTTACGAGAATGCCCTTCAAATAGCGGACA ACATCTTTTACGCAAAGGAAACAATCTCCGCCGTTGCAAAGCAG CATGGGATGAAG
THAOC_33887	CB unknown function	TGGATGTCGCAGTCACTCAGCTGCGTGTCCGCTCTTATTCCAGCG GTTCGTGCACTTTTCGTATCTACCACAAAGCCGGGGATTTTGTTG GCTCCGAGAC
THAOC_34758	ISIP3	CGACAGGAGTATCTTGGTACTTACTCCAACAACGCAATGCTGGT ACCTTCCACGACAGGCGAAGATAACTTCGGCCTGGACGATTACG AACCACTCTATT
THAOC_35238	Myb TF	GCAAGGAGGTGCATATCCAAAAGATATGGTCGTATGAAAAGGTC TCCTGTTGAAGAGACAGCTTCTGAACTGTCTTGCAGCTTCAGTAT TGATTCAGAGC
THAOC_35650	Isocytrat lyase	TCCAGTTGCTCGCAAGTTCAGCCAAGGTGTTAAGGCTGTATTTCC ACACCAAATGCTTGCGTACAACCTTAGGTGCGTGAAGATTGTGT GCTGCTTTCAT
THAOC_35651	mito isocytrat lyase	AATGCGGTCACATGGGCGGAAAAGTGCTTGTATCTACGCAGGAG CATGTGGACAGATTGATCGCTGCTCGTTTGGCTGCCGATGTACTG GGCGTTGAATT
THAOC_36747	ISIP2x8	GAGGAAATGTGCGCCAACTTTCAATCTTGCAATGAAGACGGTAA ATCTCAAGTGAATGTGAAATTGATCGAGGAGTTCTTGAACGGAC AGAACAAACTCG

THAOC_8512	CREGx2	CCACGGGGCGTATCTTCTTCTATATGATGGGCTCGCATCATCTCC ACAAGTCAACTCTCACAGTAAGCCAGGCGAGCGTGAATCCTGAC CTCTTCTCCGT
urease.1	urease.1	CGAATCCAATCTGACCACACACATTACGGAGACGAATGCAAATT CGGCGGCGGAAAGAGTTTACGCGAGGGAATGGGCCAGATGACTT CGGTTTCCTCTG

Supplemental Table 3: List of NanoString Probes A for 96well Run.

This table shows the sequences of the probe A used in the ST experiments. The sequences includes 50 nt of the target sequence and the sequence that attaches to the row specific fluorophore.

THAOC Nr.	Name	Sequence
THAOC_03152	Hypothetical transporter protein	AATTGGATATTGGGTGGCGTAGTTGGGTTGAGCGTACTTG GAAGCGAGGTCCTCAAGACCTAAGCGACAGCGTGACCTTG TTTCA
THAOC_35650	Isocytrat lyase	GGTGTGGAAATACAGCCTTAACACCTTGGCTGAACTTGCG AGCAACTGGACATCCTCTTCTTTTCTTGGTGTTGAGAAGAT GCTC
THAOC_02112	FBA3	CGATGTCATGGTCTCCGTCGATGAGGATCTCGGGCTCGAC GATCGGAACGCACAATTCTGCGGGTTAGCAGGAAGGTTAG GGAAC
THAOC_06605	FRE1	AATAGTGACAGCGTTGACATTGTCACGTGCCTTCCAGAGA CTTGTCAGGACTGTTGAGATTATTGAGCTTCATCATGACCA GAAG
THAOC_13951	DNAj	CTGGTACGCCTGTCCCAGCTTCTGGAACTTCTCGGCAAAGA CGCCTATCTTCCAGTTTGATCGGGAAACT

THAOC_16512	Sigma70 TF	TCATTTCTGGATCGATTGATGCAGCCCACTCCTGGTACGTC GGTGCGAACCTAACTCCTCGCTACATTCCTATTGTTTTC
THAOC_21195	ISIP2	GGTTGTAAGCGTATTGGTAGTCCCCGAATTCCAAACTGAC GACACCATCGCCAATTTGGTTTTACTCCCCTCGATTATGCG GAGT
THAOC_24978	FBA4	ACGTCCTTGGAGTAACCACCTGACAGGGCCACGACGCGTA CAACACGAGGCTTTCGGGTTATATCTATCATTTACTTGACA CCCT
THAOC_30995	ISIP1a	ATAGTTTGTTCCGATCGCCCATTCAACTTGACGATTTCCAG CGGCAGTAGCAACAGCCACTTTTTTTCCAAATTTTGCAAGA GCC
THAOC_31152	FLDA1	CGACAGTTTCAGTGTTTCCAGTCTGGGTCGAGTAGAAAAG GCCGGCCACCGTGTGGACGGCAACTCAGAGATAACGCATA T
THAOC_34758	ISIP3	GAAGGTACCAGCATTGCGTTGTTGGAGTAAGTACCAAGAT ACTCCTGTCGCCTGGAGTTTATGTATTGCCAACGAGTTTGT CTTT
THAOC_36747	ISIP2x8	TGAGATTTACCGTCTTCATTGCAAGATTGAAAGTTGGCGCA CATTTCCTCCAGATAAGGTTGTTATTGTGGAGGATGTTACT ACA
THAOC_8512	CREGx2	CTTGTGGAGATGATGCGAGCCCATCATATAGAAGAAGATA CGCCCCGTGGCTTCCTTCCTGTGTTCCAGCTACAAACTTAG AAAC
THAOC_05090	HSF2	ATGACGATCGGCTTGAGATGTTGAAGAGATCCCGTAGAAG CCCGTCGGGCCATAAAATTGGTTTTGCCTTTCAGCAATTCA ACTT

THAOC_05312	nuclear import-export_receptor_HK	AATCTTCATTCGCTTTTCGTCGTAGCTGCCGTTGTCTATAACCCATGAGGCTGGTCAAGACTTGCATGAGGACCCGCAAATTCCT
THAOC_25559	PETF; Ferredoxin	GGCGGCATCGACGATGAAGACGTCATCGGCGCACTCGAAGGTGGCTTTCGTTGGGACGCTTGAAGCGCAAGTAGAAAAC
THAOC_00388	FBA1	AACGTGCTTCGCTCCCGCAATGGTTCCGGCGATGGCCAGCAGACCTGCAATATCAAAGTTATAAGCGCGT
THAOC_24977	FBA6	GCGGCCTCCATGACGGTGTTGCACACCGACGAATGCCTGCCAATGCACTCGATCTTGTCATTTTTTTGCG
THAOC_06915	PETE; Plastocyanine	CGGAGTCGGCTCCCATCTTAACCTCAACAGTCTGGCAAACTGGAGAGAGAAGTGAAGACGATTTAACCCA
THAOC_02860	Mn_SOD	CCTGCGCTTCAAAGGCCGCCTTCATCTCGTCAAAAGACCCGAATGACTTGCGATTGCTGCATTCCGCTCAACGCTTGAGGAAGTA
THAOC_17167	Calreticulin	CATGACGCCGTAGGGCGTGTCGCCGCCGAACTTTGCTGAGGCTGTTAAAGCTGTAGCAACTCTTCCACGA
THAOC_24768	OST	AACACCAACGTCATCGCAAACACGGCCACGCTCTCATTATCATACCCTCCCTAGGACGCAAATCACTTGAAGAAGTGAAAGCGAG
THAOC_35806	UCGT	CCGCAGTTTCTTGTCCTCTTCTCGAATTACACTGAGCAACTGGAATACATCCACGCGATGACGTTCGTCAAGAGTCGCATAATCT
THAOC_02312	GnT1	GCCAACTGTCGAAAGATCTCCATGCTGTTGTAGGTTAATCGAACATGGCTCATTTGGAATGATGTGTACTGGGAATAAGACGACG

Supplemental Table 4: List of NanoString probe B for 96well plates.

The sequences listed in this table include 50 bp of the target sequence and a short universal sequence for the capture probe (CGAAAGCCATGACCTCCGATCACTC).

THAOC Nr	Name	Sequence
THAOC_03152	Hypothetical transporter protein	CGAAAGCCATGACCTCCGATCACTCGAAGATAGCCGACTTGG TTTACGCGCACTGCGGATGAGTACACATCTGGG
THAOC_35650	Isocitrate lyase	CGAAAGCCATGACCTCCGATCACTCATGAAAGCAGCACACAA TCTTCACGCACCTAAGGTTGTACGCAAGCATTT
THAOC_02112	FBA3	CGAAAGCCATGACCTCCGATCACTCCTTCTGGTATGTCCTTGA TACAACAACCTCCTGAATCCTCGCAGTTGTGG
THAOC_06605	FRE1	CGAAAGCCATGACCTCCGATCACTCCCCCTTTGGAAGAGTGA TTTGTACCAAGACTCTACTGCATGAGCATGCAA
THAOC_13951	DNAj	CGAAAGCCATGACCTCCGATCACTCCGTTCTTGTCGTAGTTGG CACGGCTCGACTCGTTCGACAGGAC
THAOC_16512	Sigma70 TF	CGAAAGCCATGACCTCCGATCACTCTGGCCCGGAGACTGCGT CGGATCTGACGTTTGAGCTGTATGACCG
THAOC_21195	ISIP2	CGAAAGCCATGACCTCCGATCACTCCATTTGTGAAATAGTCG CGAGCGTTCGTCCATTCTTGTTGCCCTCGAGTG
THAOC_24978	FBA4	CGAAAGCCATGACCTCCGATCACTCTTGAAAAGCTAGCGATG ATGCCGCTATTCTTGGCAAGTTTCTCGTTCGCC
THAOC_30995	ISIP1a	CGAAAGCCATGACCTCCGATCACTCCCAATCTCAACAATGTA CACTTCCTCTGTCTGTCCGCCGTTCACCCAAAC
THAOC_31152	FLDA1	CGAAAGCCATGACCTCCGATCACTCCGATGTCCTTCATTTCGA GGCCCGTAGCGTCGGCCAAATATCCAG

THAOC_34758	ISIP3	CGAAAGCCATGACCTCCGATCACTCAATAGAGTGGTTCGTAA TCGTCCAGGCCGAAGTTATCTTCGCCTGTCGTG
THAOC_36747	ISIP2x8	CGAAAGCCATGACCTCCGATCACTCCGAGTTTGTTCTGTCCGT TCAAGAACTCCTCGATCAATTTCACATTCACT
THAOC_8512	CREGx2	CGAAAGCCATGACCTCCGATCACTCGAAGAGGTCAGGATTCA CGCTCGCCTGGCTTACTGTGAGAGTTGA
THAOC_05090	HSF2	CGAAAGCCATGACCTCCGATCACTCATATGTTGAGTTCTTCTG TTGGAGATACACCTTGCGACGCATTCAACTTC
THAOC_05312	nuclear import-export_receptor_HK	CGAAAGCCATGACCTCCGATCACTCCTGACGATCATCTTCGCA AGTTCAGAGAGTAGCGGCTGCACAATCGAACG
THAOC_25559	PETF; Ferredoxin	CGAAAGCCATGACCTCCGATCACTCCGGCACGAGTAAGGAAG GTCAACTCCCTCCTCCTC
THAOC_00388	FBA1	CGAAAGCCATGACCTCCGATCACTCTGGTCGGTGTGGAGAAC AACAGGGACACCGTAAAGCTTGGCAACCTCGTG
THAOC_24977	FBA6	CGAAAGCCATGACCTCCGATCACTCCGTTCGACGCCTGGATG ATCACCGGCGAGTTGAACTTGGCG
THAOC_06915	PETE; Plastocyanine	CGAAAGCCATGACCTCCGATCACTCCCTTGCAGACAGTGACC TTGGCAGGCTCGAAGACAAGAAGTC
THAOC_02860	Mn_SOD	CGAAAGCCATGACCTCCGATCACTCCATACCCAACCACTTCC AAACACTGCTCCGGGCGCAG
THAOC_17167	Calreticulin	CGAAAGCCATGACCTCCGATCACTCAAAATAACGTGCGTGCG CTTGTTGGATGATCCGCAGATGTCGGGACCGAA
THAOC_24768	OST	CGAAAGCCATGACCTCCGATCACTCCCCCTTTTCCGTCACCTC GCAACGATCGAGTCCACAAATAG

THAOC_35806	UCGT	CGAAAGCCATGACCTCCGATCACTCAATTTCAGCACTGGTCTC ACCCGGCTCTCTAACTC
THAOC_02312	GnT1	CGAAAGCCATGACCTCCGATCACTCTAGGTACGCCAGCCTTCT CATCTTCCGTAATGTCAAACTGTGAG

Supplemental Table 5: Overview of Plate1 of the 96well plate run for the targeted transcriptome analysis.

Attenuated samples are shown in bold. The first lane is a reference lane, loaded with the same sample mix (high:low; 1:1). The sample name consists of the type: Low (no iron added), Rec (Iron added after initial measurement), High (grown with 10 µM $FeCL_3$), ActD (addition of actinomycin D after the initial measurement), ActD+Fe (addition of actD and iron after the initial measurement), DMSO (addition of DMSO after initial measurement), DMSO+Fe (addition of DMSO and iron after the initial measurement). This is followed by the time (0h for the initial measurement) taken after the addition of iron in min (‘)/ hours (h) and a unique experiment identifier.

	1	2	3	4	5	6	7	8	9	10	11	12
A	**Ref**	High_0h_ 108A	**Low _60'_706**	Rec_6h_706	ActD+Fe 6h_706	High_60'_ 108B	**Low _6h_2206**	**ActD _0h_2206**	**DMSO _0h_2206**	High_6h_1 08C	**Rec _0h_1607**	**ActD _15'_1607**
B	**Ref**	High_60'_ 108A	**Low _6h_706**	**ActD _0h_706**	**DMSO _0h_706**	High_6h_1 08B	**Rec _0h_2206**	**ActD _15'_2206**	**DMSO _6h_2206**	**Low _0h_1607**	**Rec _5'_1607**	**ActD _60'_1607**
C	**Ref**	***Low _45'_706***	**Rec _0h_706**	**ActD _15'_706**	**DMSO _6h_706**	**Low _0h_2206**	**Rec _5'_2206**	**ActD _60'_2206**	**DMSO+Fe _0h_2206**	**Low _5'_1607**	**Rec _15'_1607**	**ActD _6h_1607**
D	**Ref**	**Low _0h_706**	**Rec _5'_706**	**ActD _60'_706**	**DMSO+Fe _0h_706**	**Low _5'_2206**	**Rec _15'_2206**	**ActD _6h_2206**	DMSO+Fe 6h_2206	**Low _15'_1607**	Rec _30'_1607	**ActD+Fe _0h_1607**
E	**Ref**	**Low _5'_706**	**Rec _15'_706**	**ActD _6h_706**	DMSO+Fe _6h_706	**Low _15'_2206**	Rec _30'_2206	**ActD+Fe _0h_2206**	ActD _60'_2206-20ng	**Low _30'_1607**	Rec _45'_1607	**ActD+Fe _15'_1607**
F	**Ref**	**Low _15'_706**	Rec _30'_706	**ActD+Fe _0h_706**	ActD _60'_706-20ng	**Low _30'_2206**	Rec _45'_2206	**ActD+Fe _15'_2206**	ActD _60'_2206-40ng	**Low _45'_1607**	Rec _60'_1607	ActD+Fe _60'_1607
G	**Ref**	**Low _30'_706**	Rec _45'_706	**ActD+Fe _15'_706**	ActD _60'_706-40ng	**Low _45'_2206**	Rec _60'_2206	ActD+Fe _60'_2206	ActD _6h_2206_40ng	**Low _60'_1607**	Rec _6h_1607	ActD+Fe _6h_1607
H	**Ref**	*High_6h_ 108A*	Rec _60'_706	ActD+Fe _60'_706	ActD _6h_706_40ng	**Low _60'_2206**	Rec _6h_2206	ActD+Fe _6h_2206	High_0h_108B	**Low _6h_1607**	**ActD _0h_1607**	High_0h_1 08C

Supplemental Table 6: Overview of the targeted transcriptome experiment plate2 of the 96well run.

Attenuated samples are shown in bold. The first lane is a reference lane, loaded with the same sample mix (high:low; 1:1). The sample name consists of the type: Low (no iron added), Rec (Iron added after initial measurement), High (grown with 10 µM

$FeCL_3$), ActD (addition of actinomycin D after the initial measurement), ActD+Fe (addition of actD and iron after the initial measurement), DMSO (addition of DMSO after initial measurement), DMSO+Fe (addition of DMSO and iron after the initial measurement). This is followed by the time (0h for the initial measurement) taken after the addition of iron in min (‘)/ hours (h) and finally an experiment identifier.

	1	2	3	4	5	6	7	8	9	10	11	12
A	**Ref**	**DMSO _0h_1607**	High_0h_27012	**low_12h_16075**	rec_6h_9032	**Low_0h_2206**	**low_6h_30054**	rec_18h_30054	**Low_60'_2206**	**low_0h_11032**	rec_2h_26075_	**DMSO_0h_2206**
B	**Ref**	**DMSO _6h_1607**	High_0h_27012_	**low_14h_16075**	rec_9h_9032	High_0h_29013	**low_12h_30054**	rec_6h_30054	Low_60'_706_20ng	**low_6h_11032**	rec_4h_26075	**DMSO+Fe_0h_1607**
C	**Ref**	**DMSO+Fe _0h_1607**	High_12h_27021	**low_18h_16075**	rec_12h_9032	High_6h_2901	**low_14h_30054**	rec_9h_30054	High_0h_10084	**low_12h_11032**	rec_6h_26075	**Low_30'_706**
D	**Ref**	DMSO+Fe _6h_1607	High_14h_27012	**low_22h_16075**	rec_14h_9032	High_12h_29013	**low_18h_30054**	rec_12h_30054	High_6h_10084	**low_14h_11032**	rec_9h_26075	Low_0h_1607_
E	**Ref**	**Rec_5'_706**	High_18h_27012	**rec_0h_9032**	rec_18h_9032	High_14h_29013	**low_22h_30054**	rec_14h_30054	High_12h_10084_	**Low_18h_11032_**	rec_12h_26075_	low_5'_1607
F	**Ref**	ActD _60'_1607	High_22h_27012	rec_1h_9032	rec_22h_9032	High_18h_29013	**rec_0h_30054**	rec_18h_30054	High_14h_10084	**low_22h_11032**	rec_14h_26075	**low_5'_1607-30ng**
G	**Ref**	**Rec_15'_1607**	**low_0h_16075**	rec_2h_9032	**ActD_0h_1607**	High_22h_29013	rec_1h_30054	rec_22h_30054	High_18h_10084_	**rec_0h_26075**	**rec_18h_26075**	High_6h_27012
H	**Reference**	High_60'_108C	**low_6h_16075**	rec_4h_9032	**Low_45'_1607**	**low_0h_30054**	rec_2h_30054	Low_0h_2206-40ng	High_22h_10084	rec_1h_26075	Rec_22h_26075	Low_30_706-30ng

Supplemental Table 7: Changepoints and half-lives in min.

Based on the transcripts counts following the addition of iron (rec) or the addition of actinomycin D (actD) or actinomycin D and iron (actDFe) a changepoint was determined. The changepoint is the point of time when the slope reaches its plateau. The half-life was calculated based on the slope of the curve before reaching the changepoint used as k in $t_{1/2}$= ln2/k.

Gene	Changepoint actD	actdFe	rec	Half-life actd	actdfe	rec
ISIP1a	55.44	55.44	34.92	10.14	10.19	5.38
ISIP2	20.94	20.94	256.2	7.00	8.96	-1808.21
ISIP3	32.04	32.04	30.48	5.65	6.01	5.14
FBA1	16.86	16.86	882	7.82	8.87	194.34
FBA3	40.5	40.5	97.2	8.83	6.84	15.46
GnT1	15	15	1038	4.86	4.24	2599.30
Mn_SOD	60	60	4.98	90.02	92.42	-3.44
HSF2	16.32	16.32	120	7.95	10.77	-1299.65
FRE1	15	15	15	5.42	4.77	7.86
PETE	36.78	36.78	5.46	67.19	49.99	-1.56
DNAj _hk	17.4	17.4	1080	11.06	16.12	-967.18
Sigma70TF	15	15	540	11.85	15.18	349.49
Calreticulin	16.92	16.92	654	7.62	7.80	-424.38
OST	15.06	15.06	453	5.74	6.55	-447.19
FBA6	18.6	18.6	394.2	7.43	7.49	-229.77
FBA4	36.72	36.72	85.2	6.51	5.51	15.93
PETF	15.96	15.96	50.52	78.92	93.25	-25.99
FLDA1	49.92	49.92	280.2	8.45	7.39	43.96

UCGT	16.86	16.86	512.4	6.57	6.70	-562.01
ISIP2x8	16.44	16.44	106.2	7.40	7.08	-2970.63
CREGx2	29.82	29.82	81	58.25	56.13	21.55

Supplemental Table 8: Overview of calculated changepoints and the p-value for targeted transcripts.

The table shows the changepoints of targeted transcripts for three different treatments and the significance of the slope difference between and after the changepoint ($p < 0.01$).

Gene	CP-actd	CP-actd in min	p-value	CP-actdFe	CP-actdFe in min	p-value	CP-rec	CP-rec_hrs	p-value
CREGx2	0.497	29.82	0.018	0.497	29.82	0.018	1.35	81	9.82E-15
DNAj_hk	0.29	17.4	3.86E-05	0.29	17.4	3.86E-05	18	1080	1
FBA1	0.281	16.86	6.79E-08	0.281	16.86	6.79E-08	14.7	882	1.06E-05
FBA3	0.675	40.5	4.66E-07	0.31	18.6	6.65E-07	6.57	394.2	6.15E-12
FBA4	0.612	36.72	1.54E-09	0.675	40.5	4.66E-07	1.62	97.2	1.45E-10
FBA6	0.31	18.6	6.65E-07	0.612	36.72	1.54E-09	1.42	85.2	5.43E-08
FLDA1	0.832	49.92	3.45E-09	0.832	49.92	3.45E-09	4.67	280.2	8.41E-09
FRE1	0.25	15	3.81E-05	0.25	15	3.81E-05	0.25	15	4.73E-07
HSF2	0.272	16.32	5.64E-06	0.272	16.32	5.64E-06	2	120	0.719
ISIP1a	0.924	55.44	2.14E-11	0.924	55.44	2.14E-11	0.582	34.92	4.95E-16
ISIP2	0.349	20.94	3.29E-06	0.349	20.94	3.29E-06	4.27	256.2	0.013
ISIP2x8	0.274	16.44	4.82E-07	0.274	16.44	4.82E-07	1.77	106.2	0.007
ISIP3	0.534	32.04	3.21E-09	0.534	32.04	3.21E-09	0.508	30.48	1.83E-19
Mn_SOD	1	60	0.108	1	60	1.08E-01	0.083	4.98	4.39E-04
PETE	0.613	36.78	0.092	0.613	36.78	0.092	0.091	5.46	3.65E-05
PETF	0.266	15.96	0.01	0.266	15.96	0.01	0.842	50.52	9.46E-04
Sigma70TF	0.25	15	7.28E-07	0.25	15	7.28E-07	9	540	1.91E-07

Supplemental Table 9: Overview of the calculated slopes for actD, actDFe, and iron-recovery samples with their respective p-values.

The table shows the slopes for the different decay curves and the p-values of the different slopes. The values are significantly different when $p<0.01$.

Gene	slope-actD	slope-actDFe	slope-rec	p-value actd-acdtFe	p-value recovery-actD	p-value recovery-actDFe
CREGx2	-0.714	-0.741	-1.93	0.693	2.20E-16	2.20E-16
DNAj_hk	-3.76	-2.58	0.043	0.537	0.003	0.069
FBA1	-5.32	-4.69	-0.214	0.823	0.006	0.034
FBA6	-5.6	-5.55	0.181	0.974	2.07E-10	5.63E-08
FBA3	-4.71	-6.08	-2.69	0.413	5.30E-02	0.012
FBA4	-6.39	-7.55	-2.61	0.565	2.55E-03	2.35E-03
FLDA1	-4.92	-5.63	-0.946	0.582	4.48E-07	6.71E-06
FRE1	-7.67	-8.71	-5.29	0.648	0.173	0.068
HSF2	-5.23	-3.86	0.032	0.549	5.47E-04	0.022
ISIP1a	-4.1	-4.08	-7.73	0.982	1.29E-07	8.61E-06
ISIP2	-5.94	-4.64	0.023	0.431	1.59E-08	2.21E-04
ISIP2x8	-5.62	-5.87	0.014	0.834	1.33E-12	4.22E-11
ISIP3	-7.36	-6.92	-8.09	0.711	0.369	0.246
Mn_SOD	-0.462	-0.45	12.1	0.907	4.98E-05	5.07E-05
PETE	-0.619	-0.832	26.65	0.215	1.12E-08	8.63E-09
PETF	-0.527	-0.446	1.6	0.475	7.05E-12	9.31E-11
Sigma70TF	-3.51	-2.74	-0.119	0.524	3.50E-05	0.004

Supplemental Table 10: Overview of the identified N-linked glycosylated peptides/proteins in terms of the presence of transmembrane domains, signal peptides, and their subcellular localization.

The table provides information on all identified N-Glycosylated proteins and peptides. THAOC numbers highlighted in orange are incomplete sequences. NXS type motifs are highlighted in blue, and motifs containing a valine are highlighted in yellow.

TMHMM Server v. 2.0 (Krogh *et al*., 2001) was used for transmembrane analysis. N-linked glycosylation sites were confirmed by prediction with NetNGlyc 1.0 Server (Chuang *et al.*, 2012). Subcellular localization of the proteins was analyzed with TargetP 1.1 Server (Emanuelsson *et al.*, 2000) and ChloroP 1.1 Server (Emanuelsson *et al.*, 2000), and the presence of a signal peptide with SignalP 4.1 Server (Petersen *et al.*, 2011).

THAOC_Nr	Peptides(bacterial-motif)	+Fe/-Fe	predicted_motif	SignalP	TMHMM	TargetP
THAOC_01082	IENGTLTDTDEFIR	+Fe	NGTL	N	o	M
THAOC_01861	IDVDNSNFTK	+Fe	NFTK	N	i48-70o	_
THAOC_02303	GLFTPPPIYNETSEATR	+Fe	NETS	N	o	_
THAOC_02593	NEDGNLSYTFIPSNTTLTAK	+Fe	NLSY	N	i506-525o	M
THAOC_03029	GVYEVVANVTGR	+Fe	NVTG	N	o	_
THAOC_06669	LNSTLVSNSQGTK	+Fe	NSTL	N	o	_
THAOC_07762	NLTVFVDPIDGTR	+Fe	NLTV	N	o	_
THAOC_09459	NGTSTLAYLANEK	+Fe	NGTS	N	o	_
THAOC_09965	VLNATYPK	+Fe	NATY	Y	o	M
THAOC_10672	YVNITETEEDLK	+Fe	NITE	N	o	S
THAOC_10830	TGNSTTTFR	+Fe	NSTT	N	o	C
THAOC_15509	NVSETLVGIDLSPK	+Fe	NVSE	Y	o	S
THAOC_15574	ENATAVPK	+Fe	NATA	N	i35-57o	C
THAOC_15644	ALNNGTVDLIAGALR	+Fe	NGTV	N	o	C
THAOC_15735	DSTYELSDGSNK	+Fe	NKSF	N	i96-118o	_
THAOC_15783	VLEAGLNASR	+Fe	NASR	N	o	_
THAOC_16186	IPQDSYENCTR	+Fe	NCTR	N	i104-126o	_
THAOC_16348	GEYACPFAGAEEGNVTIGPR	+Fe	NVTI	N	o	_
THAOC_16678	NLTVFVDPIDGTR	+Fe	NLTV	N	o	_
THAOC_16927	VGLNASDER	+Fe	NASD	N	o	S
THAOC_16944	SSVFFGNETR	+Fe	NETR	N	o	_
THAOC_17196	NLELDKNCSSLAGR	+Fe	NCSS	N	o	M
THAOC_17197	YVNITETEEDLK	+Fe	NITE	Y	o	M

THAOC_18661	AGNNTGVLGR	+Fe	*NNTG*	N	o	M
THAOC_19290	NSTYELSDGSNK	+Fe	NSTY	N	o	_
THAOC_20249	HPADNFTR	+Fe	NFTR	N	i29-46o	M
THAOC_20416	SKNITTSPLR	+Fe	NITT	Y	o	S
THAOC_21105	YGANVTGITLSR	+Fe	NVTG	N	o15-37i58-80o	S
THAOC_25916	DSAADHFNVSGR	+Fe	NVSG	N	o	_
THAOC_28805	VYNVTR	+Fe	NVTR	N	o324-346i568-590o974-996i1003-1022o1032-1054i	C
THAOC_28865	ESNEAISATDEKNVTR	+Fe	NVTR	N	o	C
THAOC_31319	ETNQTLEEER	+Fe	NQTL	N	o	_
THAOC_33645	LFQYINNTVR	+Fe	NNTV	Y	i7-29o	S
THAOC_34498	FVNITR	+Fe	NITR	N	o	M
THAOC_35708	GSNDTLVQAYVVGSR	+Fe	NDTL	Y	i5-27o	S
THAOC_35708	TGVDEITNQTK	+Fe	NQTK			
THAOC_36060	ITFSDISNETLGTPR	+Fe	NETL	N	o218-240i	_
THAOC_36156	FINSVLVNASR	+Fe	NASR	N	o	_
THAOC_36968	ETSNSSDLIDAIDGK	+Fe	*NSSD*	N	o509-531i619-641o699-721i734-756o	C
THAOC_37130	SKNITTSPLR	+Fe	NITT	N	o	_
THAOC_37151	SLNETNGAVTK	+Fe	NETN	N	i7-26o	M
THAOC_00016	NQTSFQSALDSVR	+Fe and -Fe	NQTS	Y	o	S
THAOC_00282	ENVTTQVEIR	+Fe and -Fe	NVTT	N	o	_
THAOC_01040	ESIDSNGTDVTR	+Fe and -Fe	NGTD	N	o	_
THAOC_01215	LEEDFFSNATALK	+Fe and -Fe	NATA	Y	o	S
THAOC_01294	LANNTLFR	+Fe and -Fe	NNTL	Y	o	C
THAOC_01622	DAANSTVSDDFLR	+Fe and -Fe	NSTV	Y	o	C
THAOC_01697	VTGIGANHSEDFQILK	+Fe and -Fe	NHSE	Y	o	S
THAOC_02237	LGGDVNLTLIAER	+Fe and -Fe	NLTL	N	o	C

THAOC_02407	NAAAQDVTFTEFFQNR	+Fe and -Fe	NRSE	N	i7-29o	S
THAOC_02444	NVTGELQR	+Fe and -Fe	NVTG	Y	o	C
THAOC_02645	NLTVTGEYVLK	+Fe and -Fe	NLTV	Y	o222-241i253-270o285-302i432-454o469-491i	S
THAOC_03578	GNVTSASQLVNVLVANR	+Fe and -Fe	NVTS	N	o	_
THAOC_04813	VNATIPGGLANVR	+Fe and -Fe	NATI	N	i161-183o	_
THAOC_08721	NSTTGLISNEANWQR	+Fe and -Fe	NSTT	N	o	_
THAOC_08721	AENYTLIDESK	+Fe and -Fe	NYTL			
THAOC_08875	NDTLTEDDRLER	+Fe and -Fe	NDTL	Y	o	S
THAOC_09665	STDDLANLTFDLPVDPR	+Fe and -Fe	NLTF	N	o	_
THAOC_10687	ENVTTQVEIR	+Fe and -Fe	NVTT	N	o	C
THAOC_11178	GNTTDELLTNLR	+Fe and -Fe	NTTD	N	o	C
THAOC_12170	NLTFSATKPFDGSVEDPDGIPTR	+Fe and -Fe	NLTF	Y	o	M
THAOC_12670	FTEVQKNETVAEK	+Fe and -Fe	*NETV*	N	o	_
THAOC_13195	VSGGLSAGDNETK	+Fe and -Fe	NETK	N	i12-34o	S
THAOC_13263	SNTTFLNDTIEETIAR	+Fe and -Fe	NDTI	N	i16-38o	_
THAOC_13263	SNTTFLNDTIEETIAR	+Fe and -Fe	NTTF			
THAOC_13263	STNDTIDLIR	+Fe and -Fe	NDTI			
THAOC_13611	AGNVTQDKLDEALDYAK	+Fe and -Fe	NVTQ	N	o	M
THAOC_17042	FGGLSYLNTTSLADDGAHR	+Fe and -Fe	NTTS	N	i7-26o	S
THAOC_19641	CDSVGPQYVNVTK	+Fe and -Fe	NVTK	N	o	M
THAOC_19641	IENSTELPDFVSEYYK	+Fe and -Fe	NSTE			
THAOC_19688	ATFQHILPLNSTGLAK	+Fe and -Fe	NSTG	Y	o311-333i	S
THAOC_20624	NELNLTHPK	+Fe and -Fe	NLTH	N	o	_
THAOC_20766	KIALDNDLCVNMEVAPQAGANVTK	+Fe and -Fe	NVTK	N	i21-43o	_

THAOC_21564	LANETNLDAPETTSK	+Fe and -Fe	NETN	Y	o	S
THAOC_23462	LNATLSDEK	+Fe and -Fe	NATL	N	o	C
THAOC_24488	YSLNFTNSTEGAGDAR	+Fe and -Fe	NFTN	N	o	_
THAOC_24488	YSLNFTNSTEGAGDAR	+Fe and -Fe	NSTE			
THAOC_24857	RPQVQPDNTTK	+Fe and -Fe	NTTK	Y	o	_
THAOC_25360	YSFSAVTANGTQTK	+Fe and -Fe	NGTQ	Y	o	S
THAOC_28112	ATFQHILPLNSTGLAK	+Fe and -Fe	NSTG	Y	o	M
THAOC_30958	YVFANYTDSPALR	+Fe and -Fe	NYTD	N	i21-43o	M
THAOC_31443	VLGLPAAVVAGGNFTR	+Fe and -Fe	NFTR	Y	o	M
THAOC_32088	NFNVTYLNKPAAPR	+Fe and -Fe	NVTY	Y	o	C
THAOC_32289	ITNIIDFAAVNGTEPLLR	+Fe and -Fe	NGTE	Y	o317-339i	S
THAOC_32289	LLFTTAADATFSSNSTR	+Fe and -Fe	NSTR			
THAOC_32605	DISYERNDTAR	+Fe and -Fe	*NDTR*	N	i131-153o	_
THAOC_34298	WNSTLESQVIK	+Fe and -Fe	NSTL	N	o	_
THAOC_34531	YGNATFVQDGLTLK	+Fe and -Fe	NATF	N	o	_
THAOC_34531	LLNQSLTDGGFK	+Fe and -Fe	NQSL			
THAOC_34839	VQANEFLNNSK	+Fe and -Fe	NNSK	Y	o	S
THAOC_34986	KIALDNDLCVNMEVAPQAGANVTK	+Fe and -Fe	NVTK	N	o	C
THAOC_35212	GLNSTTSDGVTPK	+Fe and -Fe	NSTT	Y	o	S
THAOC_35477	DRLNTTLVDGK	+Fe and -Fe	NTTL	N	i89-111o	_
THAOC_35806	SVSQFNASDVK	+Fe and -Fe	NASD	N	o	_
THAOC_35840	EEVAVNETR	+Fe and -Fe	NETR	N	o301-323i330-352o362-384i412-434o444-466i525-547o825-847i854-876o924-946i953-975o1292-1314i1321-	_

					1343o1353-1375i1403-1425o1435-1457i1516-1538o1816-1838i1845-1867o1877-1899i1912-1934o1944-1966i	
THAOC_37387	ENVTTQVEIR	+Fe and -Fe	NVTT	N	o	_
THAOC_37792	NGTVADSLIQPLAAK	+Fe and -Fe	NGTV	N	i93-115o	C
THAOC_03372	SVNHNLNAAMTNLTHR	-Fe	NLTH	N	i96-113o128-150i552-574o	_
THAOC_05598	NQNETDSLYFENR	-Fe	NETD	N	i63-80o	M
THAOC_07125	YEDIVPIFHNR	-Fe	NRSV	N	o	_
THAOC_08463	AANLTVQSTVDEDEEER	-Fe	NLTV	N	o4-26i422-444o	S
THAOC_08758	NMVNTSQIDFIVTIDNGNDAQMK	-Fe	NTSQ	Y	o	S
THAOC_08758	VYNSSAGENR	-Fe	NSSA			
THAOC_08899	SPLPANETDSELRGEVDPALMTR	-Fe	NETD	Y	o	S
THAOC_10893	YCEASTPHSNFTCYGLAK	-Fe	NFTC	Y	i7-26o	S
THAOC_12275	DFFNITSSDNSSESMK	-Fe	NSSE	Y	o	C
THAOC_12809	VIGYGVDDATSGSVDKLNSTEIVEK	-Fe	NSTE	N	o685-704i744-766o821-843i873-895o910-932i	_
THAOC_13367	ELNEMNRLIEK	-Fe	NRLI	N	o	M
THAOC_13367	LDVSFYAPNGTLLR	-Fe	NGTL			
THAOC_14224	VGEAYYQENNGTVVVNK	-Fe	NGTV	Y	o258-280i	S
THAOC_15714	LYDLNITNLR	-Fe	NITN	N	i13-35o	M
THAOC_15991	SPEPYMNMTMK	-Fe	NMTM	Y	o	S

THAOC_17301	IASAMPNATVTILEK	-Fe	NATV	N	o	_
THAOC_20624	MADSIENYTCADESLNTTEAKR	-Fe	NTTE			
THAOC_22915	MADSIENYTCADESLNTTEAK	-Fe	NYTC	Y	o	S
THAOC_28875	AGDYAPGSGNFLANETK	-Fe	NETK	Y	o	S
THAOC_30375	NQNETDSLYFENR	-Fe	NETD	N	i62-79o	M
THAOC_30692	VNNTETLTASVSDPDIVSR	-Fe	NNTE	N	o	M
THAOC_30995	SEDSNGEVDVFSTNSLSSNHELVNFNATGSR	-Fe	NATG	Y	o533-555i	S
THAOC_30995	GSGSPVLVTTGPSASNEMQNGTCSSGCK	-Fe	NGTC			
THAOC_31818	LEDGAVCETDSIDTNVTDVR	-Fe	NVTD	Y	o	S
THAOC_32920	LQGSYVQNNTLR	-Fe	NNTL	N	o122-139i	M
THAOC_35081	VNNTETLTASVSDPDIVSR	-Fe	NNTE	N	o	M
THAOC_35093	VNFNITMMDLR	-Fe	NITM	N	o567-589i	S
THAOC_35148	MAVNTTQLENFR	-Fe	NTTQ	Y	o	C
THAOC_35840	MMNDNNSSIEELLR	-Fe	NSSI			
THAOC_37707	FTDADGVTPLMNAAENGTAAVLK	-Fe	NGTA	Y	o	S
THAOC_37707	GMLNVTEALLEK	-Fe	NVTE			
THAOC_37891	LANVTLDELDAVEMIGGAMR	-Fe	NVTL	N	o	M

Supplemental Table 11: Overview of functional characterization of full-length sequences of the identified N-linked glycosylated proteins.

The protein number is shown on the left (THAOC). "Log2 +Fe/-Fe" shows the quantitative analysis using Skyline for all peptides that were identified in high- and low-iron samples. Conserved domain search (CDD) on NCBI (Marchler-Bauer *et al.*, 2017)was used, in combination with a general blast search against NCBI's nr database. Additionally, the KEGG database was searched.

THAOC	+Fe or -Fe	log2 foldchange Fe/+Fe	CDD_best-hit	blast_best-hit-genename (bitscore - evalue - Ident)	KEGG_analysis
THAOC_01082	+Fe		DOT1	OLQ11018.1 Ankyrin-1 [Symbiodinium microadriaticum] 64.3 5e-07 27%	
THAOC_01861	+Fe			XP_028606906.1 serine/threonine-protein kinase ULK4 isoform X2... 40.8 7.8 32%	
THAOC_02303	+Fe		pro_imino_pep_1	PZQ34147.1 prolyl aminopeptidase [Pseudomonas oleovorans] 148 3e-37 35%	
THAOC_02593	+Fe		PLN00113	RAK00555.1 Leucine-rich repeat (LRR) protein [Larkinella arbor... 189 5e-45 34%	
THAOC_03029	+Fe		MscS	OEU16058.1 MS_channel-domain-containing protein [Fragilariopsi... 94.0 3e-18 34%	
THAOC_06669	+Fe		PT	GBG29637.1 Integrin alpha-D [Hondaea fermentalgiana] 51.2 0.064 27%	
THAOC_07762	*+Fe*		IPPase	KOO26960.1 3'(2'),5'-bisphosphate nucleotidase 1-like protein ... 102 3e-23 48%	
THAOC_09459	+Fe		DAP2	GAX15253.1 acylaminoacyl-peptidase [Fistulifera solaris] 610 0.0 45%	APEH; acylaminoacyl-peptidase [EC:3.4.19.1]
THAOC_09965	+Fe		PHA03247	XP_024247912.1 E3 ubiquitin-protein ligase rnf213-alpha-like [... 102 3e-18 50%	

THAOC_10672	+Fe		Peptidase_C1A	OEU18519.1 Peptidase_C1-domain-containing protein [Fragilariop... 211 5e-64 51%	CTSL; cathepsin L [EC:3.4.22.15]
THAOC_10830	+Fe		COG1233	OEU13913.1 carotenoid isomerase [Fragilariopsis cylindrus CCMP... 479 4e-155 47%	
THAOC_15509	+Fe		COG4976	OEU09404.1 S-adenosyl-L-methionine-dependent methyltransferase [Fragilariopsis cylindrus CCMP1102]... 257 5e-76 36%	
THAOC_15574	+Fe			WP_090006886.1 DNRLRE domain-containing protein [Lentzea viola... 81.3 4e-12 32%	
THAOC_15644	+Fe		PBP2_BztA	KPQ18091.1 ABC-type L-amino acid uptake system substrate-bindi[Rhodobacteraceae bacterium HLUCCO18]... 77.8 2e-11 26%	
THAOC_15735	+Fe		Ricin_B_lectin	XP_026285426.1 polypeptide N-acetylgalactosaminyltransferase 5... 53.9 0.005 27%	
THAOC_15783	+Fe		DnaJ	RMF19234.1 molecular chaperone DnaJ [Candidatus Dadabacteria b... 68.9 3e-09 49%	
THAOC_16186	+Fe		Beta_propel	OEU15308.1 hemopexin domain-containing protein [Fragilariopsis... 213 1e-53 28%	
THAOC_16348	+Fe			WP_020568116.1 T9SS type A sorting domain-containing protein [... 55.1 0.001 40%	
THAOC_16678	+Fe		IPPase	OEU17810.1 carbohydrate phosphatase [Fragilariopsis cylindrus ... 289 4e-88 43%	cysQ; 3'(2'), 5'-bisphosphate nucleotidase [EC:3.1.3.7]
THAOC_16927	+Fe		Laminin_G_3	XP_022838756.1 Concanavalin A-like lectin/glucanases superfami... 113 2e-20 35%	

THAOC_16944	+Fe			HAB15428.1 TPA: dehydrogenase [Verrucomicrobiales bacterium] 68.9 2e-07 42%	
THAOC_17196	+Fe		Smc	XP_021174756.1 ankyrin repeat domain-containing protein 26 [Fu... 48.1 0.23 25%	
THAOC_17197	+Fe		Peptidase_C1A	XP_002296462.1 probable papain cyctein protease [Thalassiosira... 389 2e-130 55%	
THAOC_18661	+Fe		COG4102		
THAOC_19290	+Fe		Ricin_B_lectin	TFK97697.1 ricin B lectin domain-containing protein [Pterula g... 50.4 0.003 30%	
THAOC_20249	+Fe		pp-GalNAc-T	PXF39435.1 Polypeptide N-acetylgalactosaminyltransferase 10 [G... 444 2e-139 47%	GALNT; polypeptide N-acetylgalactosaminyltransferase [EC:2.4.1.41]
THAOC_20416	+Fe			WP_081909449.1 alpha/beta fold hydrolase [Aureispira sp. CCB-QB1] 41.6 4.9 32%	
THAOC_21105	+Fe		Cfa	RLB44392.1 class I SAM-dependent methyltransferase [Deltaprote... 296 9e-92 43%	E2.1.1.317; sphingolipid C9-methyltransferase [EC:2.1.1.317]
THAOC_25916	+Fe		Herpes_BLLF1	XP_002292690.1 predicted protein [Thalassiosira pseudonana CCM... 965 0.0 33%	
THAOC_28805	+Fe		Chitin_synth_2	ACL00587.1 chitin synthase 1 [Thalassiosira rotula] 866 0.0 51%	CHS1; chitin synthase [EC:2.4.1.16]
THAOC_28865	+Fe			XP_002290655.1 predicted protein [Thalassiosira pseudonana CCM... 533 2e-174 62%	
THAOC_31319	+Fe		Smc	XP_022340696.1 CAP-Gly domain-containing linker protein 1-like... 44.7 2.0 28%	
THAOC_33645	+Fe		GTP_HflX	GAX27929.1 GTPase [Fistulifera solaris] 658 0.0 64%	hflX; GTPase

THAOC_34498	+Fe		SpoT	XP_002289042.1 GTP pyrophosphokinase-like protein [Thalassiosi... 794 0.0 83%	relA; GTP pyrophosphokinase [EC:2.7.6.5]
THAOC_35708a	+Fe		DLH	XP_013721012.1 carboxymethylenebutenolidase homolog [Brassica ... 73.2 5e-10 29%	
THAOC_35708b	+Fe		DLH	XP_013721012.1 carboxymethylenebutenolidase homolog [Brassica ... 73.2 5e-10 29%	
THAOC_36060	+Fe		CIA30	OLQ01912.1 Chaperone protein DnaJ [Symbiodinium microadriaticum] 131 3e-30 39%	
THAOC_36156	+Fe		CotH	WP_119147344.1 spore coat protein CotH [Cohnella sp. K2E09-144] 77.8 7e-11 25%	
THAOC_36968	+Fe		S2P-M50_like_2	WP_072620443.1 site-2 protease family protein [Spirulina major] 107 1e-20 36%	
THAOC_37130	+Fe			WP_081909449.1 alpha/beta fold hydrolase [Aureispira sp. CCB-QB1] 42.0 2.1 29%	
THAOC_37151	+Fe		AarF	OEU20769.1 ABC1-domain-containing protein [Fragilariopsis cyli... 874 0.0 59%	
THAOC_00016	+Fe and -Fe	-3.01	PLN02431	XP_002289265.1 nitrite reductase-ferredoxin dependent [Thalass... 990 0.0 86%	nirA; ferredoxin-nitrite reductase [EC:1.7.7.1]
THAOC_00282	+Fe and -Fe	-0.51		XP_001747551.1 hypothetical protein [Monosiga brevicollis MX1] 91.7 2e-14 31%	
THAOC_01040	+Fe and -Fe	-0.93	RluA	XP_002293927.1 psedouridylate synthase [Thalassiosira pseudona... 435 3e-151 74%	rluA; tRNA pseudouridine32 synthase / 23S rRNA pseudouridine746 synthase [EC:5.4.99.28 5.4.99.29]
THAOC_01215	+Fe and -Fe	-1.09	2OG-FeII_Oxy_3	KOO34593.1 proline hydroxylase [Chrysochromulina sp. CCMP291] 96.3 5e-17 30%	

THAOC_01294	+Fe and -Fe	-1.21	DUF179	WP_099438613.1 YqgE/AlgH family protein [Pedobacter ginsengisoli] 52.8 1e-04 29%	
THAOC_01622	+Fe and -Fe	-0.84	SpoVK	OEU19899.1 AAA-domain-containing protein [Fragilariopsis cylin... 353 3e-112 55%	NSF; vesicle-fusing ATPase [EC:3.6.4.6]
THAOC_01697	+Fe and -Fe	-1.99	P4Hc	GAX23806.1 prolyl 4-hydroxylase [Fistulifera solaris] 329 2e-107 60%	P4HA; prolyl 4-hydroxylase [EC:1.14.11.2]
THAOC_02237	+Fe and -Fe	-0.91	SpoVK	OEU16619.1 AAA-domain-containing protein [Fragilariopsis cylin... 531 3e-170 85%	
THAOC_02407	+Fe and -Fe	-0.40	SPFH_prohibitin	MAR20860.1 peptidase [Flavobacteriales bacterium] 55.5 4e-05 27%	
THAOC_02444	+Fe and -Fe	-0.85	Smc	XP_028518291.1 putative leucine-rich repeat-containing protein... 52.8 0.010 24%	
THAOC_02645	+Fe and -Fe	-0.40	Lung_7-TM_R	OEU09634.1 transmembrane receptor [Fragilariopsis cylindrus CC... 290 8e-90 42%	
THAOC_03578	+Fe and -Fe	-0.74	SOUL	OLP93311.1 Heme-binding-like protein, chloroplastic [Symbiodin... 132 3e-30 39%	
THAOC_04813	+Fe and -Fe	0.53	CYCc		
THAOC_08721a	+Fe and -Fe	-0.50	Glyco_hydro_36N	PKN91130.1 cellobiose phosphorylase [Chloroflexi bacterium HGW... 782 0.0 40%	
THAOC_08721b	+Fe and -Fe	8.61	Glyco_hydro_36N	PKN91130.1 cellobiose phosphorylase [Chloroflexi bacterium HGW... 782 0.0 40%	
THAOC_08875	+Fe and -Fe	-1.35	GstA	OEU21517.1 glutathione S-transferase [Fragilariopsis cylindrus... 397 5e-130 45%	GST; glutathione S-transferase [EC:2.5.1.18]

THAOC_09665	+Fe and -Fe	0.09	Phytochelatin	GAX12778.1 hypothetical protein FisN_15Hh257 [Fistulifera sola... 256 1e-75 46%	
THAOC_10687	+Fe and -Fe	-0.68		XP_004991002.1 hemagglutinin/amebocyte aggregation factor [Sal... 162 1e-35 27%	
THAOC_11178	+Fe and -Fe	-2.07	Tic110	XP_002185478.1 translocator of the inner chloroplast envelope Pt... 851 0.0 43%	
THAOC_12170	+Fe and -Fe	-0.67	CRAL_TRIO	OEU21376.1 CRAL/TRIO domain-containing protein [Fragilariopsis... 234 8e-70 41%	
THAOC_12670	+Fe and -Fe	-0.82	AarF		
THAOC_13195	+Fe and -Fe	-0.99			
THAOC_13263a (2 motifs)	+Fe and -Fe	-0.04	Nucleotid_trans		
THAOC_13263a (2 motifs)	+Fe and -Fe	-0.04	Nucleotid_trans		
THAOC_13263b	+Fe and -Fe	0.38	Nucleotid_trans		
THAOC_13611	+Fe and -Fe	-1.35	S2P-M50_like_2	PXF48806.1 putative zinc metalloprotease EGY2, chloroplastic (Gracilariopsis chorda) [... 103 6e-19 31%	
THAOC_17042	+Fe and -Fe	-1.56	UMP_CMP_kin_fam	GAX17423.1 UMP-CMP kinase [Fistulifera solaris] 267 5e-82 39%	CMPK1; UMP-CMP kinase [EC:2.7.4.14]
THAOC_19641a	+Fe and -Fe	undefined	CuRO_1_Tth-MCO_like		
THAOC_19641b	+Fe and -Fe	-0.75	CuRO_1_Tth-MCO_like		
THAOC_19688	+Fe and -Fe	0.45		XP_011395472.1 putative plastid-lipid-associated protein 8, ch... 44.7 0.20 30%	
THAOC_20624	+Fe and -Fe	-1.94	PHA03307	SFF45954.1 Myxococcus cysteine-rich repeat-containing protein ... 114 1e-23 41%	

THAOC_20766	+Fe and -Fe	5.65	CdCA1		
THAOC_21564	+Fe and -Fe	-1.57	MDN1		
THAOC_23462	+Fe and -Fe	-0.77	PLN02517		E2.3.1.158; phospholipid:diacylglycerol acyltransferase [EC:2.3.1.158]
THAOC_24488	+Fe and -Fe	-0.92		EWM23412.1 signal peptide peptidase-aspartyl protease family ... 59.7 2e-06 43%	
THAOC_24488	+Fe and -Fe				EWM23412.1 signal peptide peptidase-aspartyl protease family ... 59.7 2e-06 43%
THAOC_24857	+Fe and -Fe	1.91	FAS1		
THAOC_25360	+Fe and -Fe	-0.28	MSA-2c		
THAOC_28112	+Fe and -Fe	0.45			
THAOC_30958	+Fe and -Fe	1.29		KUF96176.1 UV excision repair protein [Phytophthora nicotianae] 108 5e-21 23%	SGT1; peptidyl serine alpha-galactosyltransferase [EC:2.4.1.-]
THAOC_31443	+Fe and -Fe	-0.08		XP_011395472.1 putative plastid-lipid-associated protein 8, ch... 65.9 4e-08 30%	
THAOC_32088	+Fe and -Fe	-0.87	MhpC	KOO34546.1 hydrolase or acyltransferase of alpha beta superfam... 138 9e-34 37%	
THAOC_32289a	+Fe and -Fe	-2.04		OLQ04553.1 Nuclear cap-binding protein subunit 2-A [Symbiodini... 186 9e-49 41%	
THAOC_32289b	+Fe and -Fe	-0.48		OLQ04553.1 Nuclear cap-binding protein subunit 2-A [Symbiodini... 186 9e-49 41%	
THAOC_32605	+Fe and -Fe	-0.84	PRK00811	OEU21075.1 S-adenosylmethionine decarboxylase [Fragilariopsis ... 133 2e-33 78%	

THAOC_34298	+Fe and -Fe	0.15	COG1916	OUS48755.1 SpoU rRNA methylase family protein [Ostreococcus ta... 114 1e-23 29%	
THAOC_34531a	+Fe and -Fe	-1.06			
THAOC_34531b	+Fe and -Fe	0.06			
THAOC_34839a	+Fe and -Fe	0.06	DUF3754		
THAOC_34986b	+Fe and -Fe	3.91	CdCA1		
THAOC_35212	+Fe and -Fe	-0.35	Peptidases_S8_Tripeptidyl_Aminopeptidase_II	XP_006030111.1 tripeptidyl-peptidase 2 isoform X2 [Alligator s... 566 2e-175 34%	TPP2; tripeptidyl-peptidase II [EC:3.4.14.10]
THAOC_35477	+Fe and -Fe	-0.16	Sulfotransfer_1		
THAOC_35806	+Fe and -Fe	-1.33	GT8_HUGT1_C_like		HUGT; UDP-glucose:glycoprotein glucosyltransferase [EC:2.4.1.-]
THAOC_35840	+Fe and -Fe	-0.23	Patched	sterol regulatory element-binding protein	
THAOC_37387	+Fe and -Fe	-0.68	DUF285	WP_083930705.1 BspA family leucine-rich repeat surface protein[Eudoraea adriatica]... 107 7e-21 35%	
THAOC_37792	+Fe and -Fe	-0.23	COG3349	XP_002186389.1 PDS-like 1, phytoene desaturase-like protein, phytoene dehydrogenase-like protein [Phaeodactylum tricornutum CCAP 1055/1] 764 0.0 66%	
THAOC_03372	-Fe		COPIIcoated_ERV	EWM21113.1 Thioredoxin domain protein [Nannochloropsis gaditana] 449 3e-150 47%	
THAOC_05598	-Fe			EWM21517.1 gluconolactonase [Nannochloropsis gaditana] 96.7 1e-17 28%	

THAOC_07125	-Fe		Aldo_ket_red	HCC57794.1 TPA: oxidoreductase [Bryobacterales bacterium] 190 1e-53 40%	
THAOC_08463	-Fe		DltE	WP_101616912.1 SDR family oxidoreductase [Bifidobacterium marg... 103 9e-21 33%	
THAOC_08758a	-Fe			OEU15614.1 ferrichrome ABC transporter-like protein [Fragilari... 74.7 3e-10 27%	
THAOC_08758b	-Fe			OEU15614.1 ferrichrome ABC transporter-like protein [Fragilari... 74.7 3e-10 27%	
THAOC_08899	-Fe		COG4338	OLQ09640.1 Ubiquinone/menaquinone biosynthesis C-methyltransfe... 118 3e-24 60%	
THAOC_10893	-Fe				
THAOC_12275	-Fe		PRK11778	OEU15743.1 Peptidase_S49-domain-containing protein [Fragilario... 374 8e-121 60%	
THAOC_12809	-Fe		S2P-M50_like_2	XP_005705620.1 peptidase, M50 family protein [Galdieria sulphu... 121 1e-24 37%	
THAOC_13367a	-Fe		RluA		rluA; tRNA pseudouridine32 synthase / 23S rRNA pseudouridine746 synthase [EC:5.4.99.28 5.4.99.29]
THAOC_13367b	-Fe		RluA		rluA; tRNA pseudouridine32 synthase / 23S rRNA pseudouridine746 synthase [EC:5.4.99.28 5.4.99.29]
THAOC_14224	-Fe			WP_071106749.1 PEP-CTERM sorting domain-containing protein [Moorea producens]... 136 3e-35 41%	
THAOC_15714	-Fe		UDP_GE_SDE_e	XP_002179157.1 nad-dependent epimerase/dehydratase [Phaeodacty... 440 8e-150 61%	E5.1.3.6; UDP-glucuronate 4-epimerase [EC:5.1.3.6]

THAOC_15991	-Fe		P4Hc	XP_002293760.1 prolyl 4-hydroxylase alpha subunit [Thalassiosi... 253 9e-79 67%	P4HA; prolyl 4-hydroxylase [EC:1.14.11.2]
THAOC_17301	-Fe		crtI_fam	XP_002186335.1 phytoene desaturase. zeta-carotene desaturase. ... 536 2e-179 51%	AL1; phytoene desaturase (3,4-didehydrolycopene-forming) [EC:1.3.99.30]
THAOC_20624	-Fe		PHA03307	SFF45954.1 Myxococcus cysteine-rich repeat-containing protein ... 114 1e-23 41%	
THAOC_22915	-Fe		P4Hc	GAX11170.1 prolyl 4-hydroxylase [Fistulifera solaris] 311 9e-98 38%	
THAOC_28875	-Fe				
THAOC_30375	-Fe			WP_068777435.1 undecaprenyl/decaprenyl-phosphate alpha-N-acety... 40.4 1.5 38%	
THAOC_30692	-Fe		NADH_oxidase	XP_005853895.1 NADH dehydrogenase [Nannochloropsis gaditana CC... 128 1e-30 30%	
THAOC_30995a	-Fe			OEU14956.1 iron starvation induced protein [Fragilariopsis cyl... 459 1e-151 41%	
THAOC_30995b	-Fe			OEU14956.1 iron starvation induced protein [Fragilariopsis cyl... 459 1e-151 41%	
THAOC_31818	-Fe				
THAOC_32920	-Fe		DUF2360		
THAOC_35081	-Fe		NADH_oxidase	XP_005853895.1 NADH dehydrogenase [Nannochloropsis gaditana CC... 126 3e-30 30%	
THAOC_35093	-Fe		COPIIcoated_ERV	EWM21113.1 Thioredoxin domain protein [Nannochloropsis gaditana] 373 6e-120 40%	
THAOC_35148	-Fe		CIA30	KOO21365.1 NmrA-like family protein [Chrysochromulina sp. CCMP... 112 3e-24 32%	
THAOC_35840	-Fe		Patched	sterol regulatory element-binding protein	

THAOC_37707a	-Fe		ANK	GAX17142.1 ankyrin [Fistulifera solaris] 1053 0.0 60%	
THAOC_37707b	-Fe		ANK		
THAOC_37891	-Fe		HYOU1-like_NBD	OEU12263.1 HSP70-domain-containing protein [Fragilariopsis cyl... 465 1e-144 48%	HYOU1; hypoxia up-regulated 1

Supplemental Table 12: Overview of blast analysis with *T. oceanica* sequences that were previously identified as putative enzymes involved in N-linked glycosylation pathway to verify their functions.

Proteins previously identified in *P. tricornutum* were used to blast in the *T. oceanica* database. The resulting hits were used for a blastp search against the non-redundant database to confirm putative function.

Pt protein number	Function	THAOC Nr	*To* blastp hit	unique identifier	ID in %	e-value
46110	(1,3)-FucT	00736, partial	Alpha-(1,3)-fucosyltransferase 11 [Daphnia magna]	1022767837	24.92	2.18E-10
19705	P-Dol_ManT_(DPM1)	00809	dolichyl-phosphate mannosyltransferase [Phaeodactylum tricornutum CCAP 1055/1]	219116887	66.67	7.86E-106
54844	GnT_I	02312	GNT-I-domain-containing protein [Fragilariopsis cylindrus CCMP1102]	1072228931	42.28	6.17E-90
44117	(1,3)-GlcT_ALG_8	03361	alpha-1,3-glucosyltransferase [Fistulifera solaris]	1210515835	35.86	1.19E-75
54169	GlcII,subunit3	03472	Glucosidase II beta-subunit [Operophtera brumata]	914568173	35	5.95E-21
55197	OST_(STT3_subunit)	04081	STT3 subunit-like protein [Phaeodactylum tricornutum CCAP 1055/1]	219130538	55.64	0
54169	GlcII,subunit5	08367, partial	PREDICTED: glucosidase 2 subunit beta [Pseudopodoces humilis]	929511375	28.83	2.19E-05
54621	(1,2)-ManT_ALG_11	09775	alpha-1,2-mannosyltransferase ALG11 [Rhizophagus irregularis DAOM 197198w]	595482990	39.74	1.25E-119
44574	(1,2)-ManT_ALG_9	13160	alpha-1,2-mannosyltransferase [Fistulifera solaris]	1210525995	38.41	4.29E-107
45980	P-Dol_GlcT_ALG_5	13573, partial	dolichyl-phosphate beta-glucosyltransferase [Nannochloropsis gaditana]	585102551	45.52	2.36E-63

46110	(1,3)-FucT	14764	fucosyltransferase, partial [Thalassiosira pseudonana CCMP1335]	224004060	59.66	2.94E-70
41172	Calreticulin	17167	calreticulin [Phaeodactylum tricornutum CCAP 1055/1]	219129933	71.78	0
10976	(1,3)-ManT_ALG_3	17486	alpha-1,3-mannosyltransferase [Fistulifera solaris]	1210512764	46.82	4.63E-109
55197	OST_(STT3_subunit)	24768	STT3 subunit-like protein [Phaeodactylum tricornutum CCAP 1055/1]	219130538	45.4	0
14444	(1,4)-GlcNAcT_ALG_14	25680	UDP-N-acetylglucosamine transferase subunit ALG14 [Fragilariopsis cylindrus CCMP1102]	1072229432	37.36	5.80E-08
44425	(1,2)-ManT_ALG_12	28372, partial	alpha-1,6-mannosyltransferase [Fistulifera solaris]	1210526352	41.34	8.54E-72
22554	(1,3)-ManT_ALG_2	32241	mannosyltransferase, partial [Phaeodactylum tricornutum CCAP 1055/1]	219125608	37.58	1.83E-84
9724	GlcNAc-P-transferase_ALG_7	33316	predicted protein, partial [Phaeodactylum tricornutum CCAP 1055/1]	219111951	63.14	1.38E-171
14002	(1,4)-ManT_ALG_1	33337	beta-1,4 mannosyltransferase, partial [Thalassiosira pseudonana CCMP1335]	224006207	58.25	0
44117	(1,3)-GlcT_ALG_6	34646	dolichyl pyrophosphate Glc1Man9GlcNAc2 alpha-1,3-glucosyltransferase [Phaeodactylum tricornutum CCAP 1055/1]	219113735	43.09	2.96E-125
54787	UDP-glucose:glycoprotein_glucosyltransferase	35806, partial	UDP-glucose:glycoprotein glucosyltransferase [Fistulifera solaris]	1210512199	40.33	0
XP_002177395.1	Flippase	MMETSP_09271-36046	oligosaccharide translocation protein RFT1 [Fistulifera solaris]	GAX22268.1	41.48	3.00E-129
XP_002180427.1	beta-1,4-galactosyltransferase 4	18453	galactosyl transferase [Phaeodactylum tricornutum CCAP 1055/1]	XP_002180427.1	56.47	3.00E-141

Supplemental Table 13: Overview of changepoints in actD, actDFe, and iron-recovery samples for targeted transcript involved in the N-linked glycosylation pathway.

The table shows the changepoints of targeted transcripts for three different treatments and the significance of the slope difference between and after the changepoint ($p < 0.01$).

Gene	CP-actd	CP-actd in min	p-value	CP-actdFe	CP-actdFe in min	p-value	CP-rec	CP-rec_hrs	p-value
calreticulin	0.282	16.92	4.61E-09	0.282	16.92	4.61E-09	10.9	654	1.29E-10
GnT1	0.25	15	7.45E-06	0.25	15	7.45E-06	17.3	1038	0.183
OST	0.251	15.06	6.23E-07	0.251	15.06	6.23E-07	7.55	453	6.99E-09
UCGT	0.281	16.86	8.53E-07	0.281	16.86	8.53E-07	8.54	512.4	2.38E-07

Supplemental Table 14: Overview of slope in the actD, actDFe, and iron-recovery samples for the targeted transcripts that are involved in the N-linked glycosylation pathway.

The table shows the slopes of targeted transcripts for actD, actdFe, and iron-recovery treatments and including the significance p-value between the different slopes ($p < 0.01$).

Gene	slope-actD	slope-actDFe	slope-rec	p-value actd-acdtFe	p-value recovery-actD	p-value recovery-actDFe
calreticulin	-5.46	-5.33	0.098	0.918	5.14E-11	1.47E-08
GnT1	-8.56	-9.8	-0.016	0.471	1.93E-13	1.49E-14
OST	-7.25	-6.35	0.093	0.434	2.20E-16	3.24E-14
UCGT	-6.33	-6.21	0.074	0.913	2.20E-16	2.38E-15

Supplemental Table 15: Primer List for the design of the pDESTC'YFP-oceanica vector used for transforming *T. oceanica.*

Primer name	Sequence 5'>3'	Comment
R1Backbone_rev	TTAGGTACCAGCTTTTGTTC	Gatewat pDEST vector design primer_reaction1
R1Fcp25_Prom_fwd	gaacaaaagctggtacctaaTTCGTCCGTATGAACGCC	Gatewat pDEST vector design primer_reaction1
R1Fcp25_Prom_rev	ataccgtcgaGGTTCAAAATGAGTCTCTTGTAACTG	Gatewat pDEST vector design primer_reaction1
R1SheBle_fwd	attttgaaccTCGACGGTATCGATAATATTCTAGCTGAG	Gatewat pDEST vector design primer_reaction1
R1SheBle_rev	tgtgttggcaCCGCATGGGCCCAGATCT	Gatewat pDEST vector design primer_reaction1
R1Fcp10_Term_fwd	gcccatgcggTGCCAACACATGTAATTTTTTC	Gatewat pDEST vector design primer_reaction1
R1Fcp10_Term_rev	caccatgaataggcgagattCTTCATTTTTGCTGCAGTC	Gatewat pDEST vector design primer_reaction1
R1Backbone_fwd	AATCTCGCCTATTCATGG	Gatewat pDEST vector design primer_reaction1
R2fragment_rev	GGGTACCCTCAGCTAGAATATTATCGATACC	Gatewat pDEST vector design primer_reaction2
R2ClonNat_fwd	tattctagctgagggtacccATGACCACTCTTGACGACACG	Gatewat pDEST vector design primer_reaction2
R2ClonNat_rev	gaccggcggtgttggtcggcgtcCTAGGGGCAGGGCATGCT	Gatewat pDEST vector design primer_reaction2
R2fragment_fwd	GACGCCGACCAACACCGC	Gatewat pDEST vector design primer_reaction2
R3backbone_rev	CTTGTATACACCATGAATAGGC	Gatewat pDEST vector design primer_reaction3
R3Fcp8_Prom_fwd	ctattcatggtgtatacaagAGGAGAATACATCGAGAAGATAC	Gatewat pDEST vector design primer_reaction3
R3Fcp8_Prom_rev	agagcggccgGGTTCAATTGTTTGGTGTG	Gatewat pDEST vector design primer_reaction3
R3cassette_fwd	caattgaaccCGGCCGCTCTAGAACTAG	Gatewat pDEST vector design primer_reaction3
R3cassette_rev	tgtgttggcaCTCAGAATTCTTACTTGTACAGC	Gatewat pDEST vector design primer_reaction3
R3Fcp10_Term_fwd	gaattctgagTGCCAACACATGTAATTTTTTC	Gatewat pDEST vector design primer_reaction3
R3Fcp10_Term_rev	aaagggaacaaaagctggtaCTTCATTTTTGCTGCAGTC	Gatewat pDEST vector design primer_reaction3
R3backbone_fwd	TACCAGCTTTTGTTCCCTTTAG	Gatewat pDEST vector design primer_reaction3
YFPFw	GGCGAGGAGCTGTTCA	YFP-verficiation primer
YFPRv	GTACAGCTCGTCCATGC	YFP-verficiation primer
ISIP1acDNARv	CTAAGCTACCTGCTTTGAGCC	gene-specific primer for cDNA amplification
ISIP1acDNAFw (72C Phusion)	ATGAAGTTCCCGGTCGCCCTC	gene-specific primer for cDNA amplification

ISIP3cDNARv (58.5C Phusion)	TCACAAGAACAAGAAGGTCTT	gene-specific primer for cDNA amplification
ISIP3cDNAFw (62C Phusion)	ATGTACCTTGTCACTGCCTTG	gene-specific primer for cDNA amplification
Fre1cDNARv	TTACATTTCAAACGCCTCAGG	gene-specific primer for cDNA amplification
Fre1cDNAFw	ATGGGGCTGCTCGGCGACGTC	gene-specific primer for cDNA amplification
Fre1cDNA15bFw	GACGTCCTCCGGTGTCTC	gene-specific primer for cDNA amplification
Fre1cDNA2256bRv	CCTCAGGATAAAGGGATATACG	gene-specific primer for cDNA amplification
Mucox2cDNARv	TCACACAAACAGCGACATCGA	gene-specific primer for cDNA amplification
Mucox2cDNAFw	ATGTTAGCGCTTCCTGCGATC	gene-specific primer for cDNA amplification
ISIP1cDNAGateentryC' GFPFw	CACCATGAAGTTCCCGGTC	gene-specific primer to generate construct for pEntry D-Topo vector
ISIP1cDNAGateentryC' GFPRv	AGCTACCTGCTTTGAGCC	gene-specific primer to generate construct for pEntry D-Topo vector
ISIP3cDNAGateentryC' GFPFw	CACCATGTACCTTGTCACTGCC	gene-specific primer to generate construct for pEntry D-Topo vector
ISIP3cDNAGateentryC' GFPRv	CAAGAACAAGAAGGTCTTTCTAGT	gene-specific primer to generate construct for pEntry D-Topo vector
Fre1cDNAGateentryC' GFPFw	CACCATGGGGCTGCTCGGC	gene-specific primer to generate construct for pEntry D-Topo vector
Fre1cDNAGateentryC' GFPRv	CATTTCAAACGCCTCAGG	gene-specific primer to generate construct for pEntry D-Topo vector
Mucox2cDNAGateentry C'GFPFw	CACCATGTTAGCGCTTCCT	gene-specific primer to generate construct for pEntry D-Topo vector
Mucox2cDNAGateentry C'GFPRv	CACAAACAGCGACATCGA	gene-specific primer to generate construct for pEntry D-Topo vector
ClonNatFw	ATGACCACTCTTGACGACACG	Amplification primer
ClonNatRv	CTAGGGGCAGGGCATG	Amplification primer
ClonNat_Fw	ATGACCACTCTTGACGACACG	Amplification primer pair without StopCodon
ClonNat_Rv	GGGGCAGGGCATGCTCAT	Amplification primer pair without StopCodon
Fcp865CFw	GAAGATACTCCTCGTTGGACCC	Amplification primer, promoter exchange in pDEST vector

Fcp2564CFw	GAACGCCGAGAAAATTAGTGAC	Amplification primer, promoter exchange in pDEST vector
Fcp3ProcheckFw	GAAAAGTCCATCATCCGCAC	Amplification primer, promoter exchange in pDEST vector
Fcp3ProcheckRv	AGGAGTCCAAGAGGATCGTA	Amplification primer, promoter exchange in pDEST vector
Hsp70ProcheckFw	GACCGCATTGTTCTTGTTGA	Amplification primer, promoter exchange in pDEST vector
Hsp70ProcheckRv	ATTGTGAATTGCCTCTTGCC	Amplification primer, promoter exchange in pDEST vector
NOredProcheckFw	ATCTATTCTGTCGGCGTTGT	Amplification primer, promoter exchange in pDEST vector
NOred ProcheckRv	TCGACGGCTTTATGAGAGTC	Amplification primer, promoter exchange in pDEST vector
FBAProcheckFw	GACTGAGCCACCAGGAAC	Amplification primer, promoter exchange in pDEST vector
FBA ProcheckRv ’	ACATGCTCAAGGCACTGTT3	Amplification primer, promoter exchange in pDEST vector
Fcp25_fwd	attcatggtgtatacaagTTCGTCCGTATGAACGCC	Gatewat pDEST vector design primer for promoter exchange
Fcp25_rev	gttctagagcggccgGGTTCAAAATGAGTCTCTTGTAACTG	Gatewat pDEST vector design primer for promoter exchange
NO_ref_fwd	attcatggtgtatacaagATCGAATCTATTCTGTCGG	Gatewat pDEST vector design primer for promoter exchange
NO_ref_rev	gttctagagcggccgGGTTGTCTGTCTTCAATGC	Gatewat pDEST vector design primer for promoter exchange
Hsp70_fwd	attcatggtgtatacaagGACCGCATTGTTCTTGTTG	Gatewat pDEST vector design primer for promoter exchange
Hsp70_rev	gttctagagcggccgTGTTTGTTGTGTTATTATCGTTC	Gatewat pDEST vector design primer for promoter exchange
FBAclass-II_fwd	attcatggtgtatacaagGACTGAGCCACCAGGAAC	Gatewat pDEST vector design primer for promoter exchange
FBAclass-II_rev	gttctagagcggccgGGTTGGGTAGATTACAATCAAAC	Gatewat pDEST vector design primer for promoter exchange
ClonNatF2a30fusion_Rv	CTTTACAGGTGCCACAATTTTCTGCTTGTGGGGGCAGGGCATGCTCAT	Gatewat pDEST vector design primer for F2A construct

ClonNatF2a60fusion_Rv	CAACTTGAGAAGGTCAAAGTTCAGTGTTTGCTTTACAGGTGCCACAAT	Gatewat pDEST vector design primer for F2A construct
ClonNatF2a90fusion_Rv	GGGCCCAGGATTGGACTCAACGTCTCCAGCCAACTTGAGAAGGTCAAA	Gatewat pDEST vector design primer for F2A construct
F2A_Backbone_rev	GGGTACCCTCAGCTAGAATATTATCG	Gatewat pDEST vector design primer for F2A construct
F2A_Backbone_fwd	ATGGTGAGCAAGGGCGAG	Gatewat pDEST vector design primer for F2A construct
ClonNatF2A_fwd	tagctgagggtacccATGACCACTCTTGACGACACG	Gatewat pDEST vector design primer for promoter exchange
ClonNatF2A_rev	gaacttcatGGGCCCAGGATTGGACTC	Gatewat pDEST vector design primer for promoter exchange
ISIP1aF2A_fwd	ctgggcccATGAAGTTCCCGGTCGCC	Gatewat pDEST vector design primer for promoter exchange
ISIP1aF2A_rev	gcccttgctcaccatAGCTACCTGCTTTGAGCCTAATG	Gatewat pDEST vector design primer for promoter exchange

Supplemental Table 16: Nucleotide sequences of vector parts and full-length sequence of the pDEST-C'YFP-oceanica

>segment 1 :

TACCAGCTTTTGTTCCCTTTAGTGAGGGTTAATTTCGAGCTTGGCGTAATCATGGTCATA
GCTGTTTCCTGTGTGAAATTGTTATCCGCTCACAATTCCACACAACATACGAGCCGGAAG
CATAAAGTGTAAAGCCTGGGGTGCCTAATGAGTGAGCTAACTCACATTAATTGCGTTGCG
CTCACTGCCCGCTTTCCAGTCGGGAAACCTGTCGTGCCAGCTGCATTAATGAATCGGCCA
ACGCGCGGGGAGAGGCGGTTTGCGTATTGGGCGCTCTTCCGCTTCCTCGCTCACTGACTC
GCTGCGCTCGGTCGTTCGGCTGCGGCGAGCGGTATCAGCTCACTCAAAGGCGGTAATACG
GTTATCCACAGAATCAGGGGATAACGCAGGAAAGAACATGTGAGCAAAAGGCCAGCAAAA
GGCCAGGAACCGTAAAAAGGCCGCGTTGCTGGCGTTTTTCCATAGGCTCCGCCCCCCTGA
CGAGCATCACAAAAATCGACGCTCAAGTCAGAGGTGGCGAAACCCGACAGGACTATAAAG
ATACCAGGCGTTTCCCCCTGGAAGCTCCCTCGTGCGCTCTCCTGTTCCGACCCTGCCGCT
TACCGGATACCTGTCCGCCTTTCTCCCTTCGGGAAGCGTGGCGCTTTCTCATAGCTCACG
CTGTAGGTATCTCAGTTCGGTGTAGGTCGTTCGCTCCAAGCTGGGCTGTGTGCACGAACC
CCCCGTTCAGCCCGACCGCTGCGCCTTATCCGGTAACTATCGTCTTGAGTCCAACCCGGT
AAGACACGACTTATCGCCACTGGCAGCAGCCACTGGTAACAGGATTAGCAGAGCGAGGTA
TGTAGGCGGTGCTACAGAGTTCTTGAAGTGGTGGCCTAACTACGGCTACACTAGAAGGAC
AGTATTTGGTATCTGCGCTCTGCTGAAGCCAGTTACCTTCGGAAAAAGAGTTGGTAGCTC
TTGATCCGGCAAACAAACCACCGCTGGTAGCGGTGGTTTTTTTGTTTGCAAGCAGCAGAT
TACGCGCAGAAAAAAAGGATCTCAAGAAGATCCTTTGATCTTTTCTACGGGGTCTGACGC
TCAGTGGAACGAAAACTCACGTTAAGGGATTTTGGTCATGAGATTATCAAAAAGGATCTT
CACCTAGATCCTTTTAAATTAAAAATGAAGTTTTAAATCAATCTAAAGTATATATGAGTA
AACTTGGTCTGACAGTTACCAATGCTTAATCAGTGAGGCACCTATCTCAGCGATCTGTCT
ATTTCGTTCATCCATAGTTGCCTGACTCCCCGTCGTGTAGATAACTACGATACGGGAGGG
CTTACCATCTGGCCCCAGTGCTGCAATGATACCGCGAGACCCACGCTCACCGGCTCCAGA
TTTATCAGCAATAAACCAGCCAGCCGGAAGGGCCGAGCGCAGAAGTGGTCCTGCAACTTT
ATCCGCCTCCATCCAGTCTATTAATTGTTGCCGGGAAGCTAGAGTAAGTAGTTCGCCAGT
TAATAGTTTGCGCAACGTTGTTGCCATTGCTACAGGCATCGTGGTGTCACGCTCGTCGTT
TGGTATGGCTTCATTCAGCTCCGGTTCCCAACGATCAAGGCGAGTTACATGATCCCCCAT
GTTGTGCAAAAAAGCGGTTAGCTCCTTCGGTCCTCCGATCGTTGTCAGAAGTAAGTTGGC
CGCAGTGTTATCACTCATGGTTATGGCAGCACTGCATAATTCTCTTACTGTCATGCCATC
CGTAAGATGCTTTTCTGTGACTGGTGAGTACTCAACCAAGTCATTCTGAGAATAGTGTAT
GCGGCGACCGAGTTGCTCTTGCCCGGCGTCAATACGGGATAATACCGCGCCACATAGCAG
AACTTTAAAAGTGCTCATCATTGGAAAACGTTCTTCGGGGCGAAAACTCTCAAGGATCTT
ACCGCTGTTGAGATCCAGTTCGATGTAACCCACTCGTGCACCCAACTGATCTTCAGCATC
TTTTACTTTCACCAGCGTTTCTGGGTGAGCAAAAACAGGAAGGCAAAATGCCGCAAAAAA
GGGAATAAGGGCGACACGGAAATGTTGAATACTCATACTCTTCCTTTTTCAATATTATTG
AAGCATTTATCAGGGTTATTGTCTCATGAGCGGATACATATTTGAATGTATTTAGAAAAA
TAAACAAATAGGGGTTCCGCGCACATTTCCCCGAAAAGTGCCACCTAAATTGTAAGCGTT
AATATTTTGTTAAAATTCGCGTTAAATTTTTGTTAAATCAGCTCATTTTTTAACCAATAG

GCCGAAATCGGCAAAATCCCTTATAAATCAAAAGAATAGACCGAGATAGGGTTGAGTGTT
GTTCCAGTTTGGAACAAGAGTCCACTATTAAAGAACGTGGACTCCAACGTCAAAGGGCGA
AAAACCGTCTATCAGGGCGATGGCCCACTACGTGAACCATCACCCTAATCAAGTTTTTTG
GGGTCGAGGTGCCGTAAAGCACTAAATCGGAACCCTAAAGGGAGCCCCCGATTTAGAGCT
TGACGGGGAAAGCCGGCGAACGTGGCGAGAAAGGAAGGGAAGAAAGCGAAAGGAGCGGGC
GCTAGGGCGCTGGCAAGTGTAGCGGTCACGCTGCGCGTAACCACCACACCCGCCGCGCTT
AATGCGCCGCTACAGGGCGCGTCCCATTCGCCATTCAGGCTGCGCAACTGTTGGGAAGGG
CGATCGGTGCGGGCCTCTTCGCTATTACGCCAGCTGGCGAAAGGGGGATGTGCTGCAAGG
CGATTAAGTTGGGTAACGCCAGGGTTTTCCCAGTCACGACGTTGTAAAACGACGGCCAGT
GAATTGTAATACGACTCACTAACCATGATTACGCCAAGCTCGAAATTAACCCTCACTAAA
GGGAACAAAAGCTGGTACCTAA

>segment2(Fcp25Promoter)

TTCGTCCGTATGAA
CGCCGAGAAAATTAGTGACACCCTTGCCTTCGCTTCGCCAGCTTACGCTTGACTTCGCTC
GCGATCGCCCAGCATGCTTGGCAGTGGGGGGCGCCTTAATCTCTCGGCCCCCTGGCCACT
ATTGTTATGGTTGCGAGAAATCGATGAATGAACCTCACAGCGTCGCACAGGCGTATGGAG
CGGCGCGCGGCGCGCGGCCGGCGGCGCATACGATAAGCCTAAGCACCTGGATCCACTTCA
ACTTCAGCCGTGAGTAAGAATTTGCACGAAATGAACGGGATTGTGCATGAAAAGGTAACA
GCATGCTCCCTTTGCATGGTAAAGAAACCAAAGGTGCGTTTTCTTCAGATTTTGATTTCC
AATATTTGCATCATTTTTAGTACAACGAGCAAAGCCGAAGCCGCATCGGCTATTAGACGA
ACACAGTTACAAGAGACTCATTTTGAACC

>nourseothricin acetyltransferase

ATGACCACTCTTGACGACACGGCTTACCGGTACCGCACCAGTGTCCCGGGGGACGCCGAG
GCCATCGAGGCACTGGATGGGTCCTTCACCACCGACACCGTCTTCCGCGTCACCGCCACC
GGGGACGGCTTCACCCTGCGGGAGGTGCCGGTGGACCCGCCCCTGACCAAGGTGTTCCCC
GACGACGAATCGGACGACGAATCGGACGACGGGGAGGACGGCGACCCGGACTCCCGGACG
TTCGTCGCGTACGGGGACGACGGCGACCTGGCGGGCTTCGTGGTCATCTCGTACTCGGCG
TGGAACCGCCGGCTGACCGTCGAGGACATCGAGGTCGCCCCGGAGCACCGGGGGCACGGG
GTCGGGCGCGCGTTGATGGGGCTCGCGACGGAGTTCGCCGGCGAGCGGGGCGCCGGGCAC
CTCTGGCTGGAGGTCACCAACGTCAACGCACCGGCGATCCACGCGTACCGGCGGATGGGG
TTCACCCTCTGCGGCCTGGACACCGCCCTGTACGACGGCACCGCCTCGGACGGCGAGCGG
CAGGCGCTCTACATGAGCATGCCCTGCCCCTAG

>segment4(Fcp10Terminator):

TGCCAACACATGTAATTTTTTTCAATTCCGAAAAGTCCATCATCCGCACACCGCACAATTG
CACCGCGAGTGAGCGCCGAAACGATGGACGTAAGAGATTACGTCATTAGTGAGTAGCGTT
TAGCGTTTATTGATAAGCCGGCAGCGGGCACATCGCCAACTTACGGCGTTGAGCTGAGTG
GAGGCCTTGCCGGTCTGGGCAGGGGCGAAGGCGGCGGCGGAGCCGATAAGGGCAAGGGCG
GCAACAGACTTCATCATGGTTGGTTTGTTACGATGGTGCAAAAATGGGGTCTACAGGCAG
CAGGCGGACTCTTCGAAACCAGCTCGCCTCTCTTGCGTGAGGGTCAACATTACATTGGTA
CTCCCATTCATTATTAATTACTCAACTCTACCAAAATATATTTTCCATTGGTTCATTCTC
TATGAATATACCTAGGTAAAGTAACATTATGGAAAAAATGCGACGCTAAATTACTAATGC
TACCCCTTCATAGTTGAACGTATCTTAGCAACGTGTGACGGGCGCGGAGGCAAGCAGTAT
CGATATCGTAGCCCGATCTTTGGCCTTCCTACCTACTTGGTTGTTCTCACATTTCAAACC
TTTGGAACAATTACTGTGAGTTAGCCGGAGGGGAGCAATCGACTCGGAGACTGCTACGAA

CCGGTTTGATCTCCATGTCTTGTGTCATGTGATAACACTCACGCTTGTTCTTCGACTGCA
GCAAAAATGAAGTCTGTTGCCGC
>Fcp8 Promoter:
CACAGTCAGGAATAACACTAGCTCGTCTTCATAGGGCG-(linker)-AGGAGAATACATCG
AGAAGATACTCCTCGTTGGACCATGTTGAGCCGTCTCTTTCCTCCGATACGAACCCGGTG
CTCTAGGAGGACGAGAAGACGAAAAGGGAATCAAAAGGCATCTGCGGTTAGGCGGGCATC
AAAGAATGCACGGACCAAGCCGCCTGTATGTAACGCCGAACGAAGAAAACAGTTGGCTAG
AGAGTGAATCTATCAGCGTTCGGCGTTACGTTGTGTTAGGTTACAATGCTTCTGTACGTT
TGCAAAATGAAAGAGAACGTGGCAGAAAGGAAACGAAAATGGAAGGTGACGATTGACGTG
GAAAATGGCTCCTTCTGGCCATGAACAGCCTTTGTGATAAAAGTAGGTACGGCTACAGAC
CGTGCCCAAACAAAGCGTCATTAGAAGGCGTGTCACAAAGAACACCGCAAGCAAGCAGAA
CCCACGATGTCTGAAGAGTGACGATCCGTCGGTGAATAAAAGAGAACGTTCGTGTCGATC
TTATTCGGCCGTAAAACGAAAGCAAGCTGTTGATTAACGGCGCGGCGCGACGGAGCATGG
CCGTGAGAATGATATGGAGAAAAGAAGATCACATTGCCTTCCTTGAGAGGAGAATGAAAA
CACTACACTTCAGGAGGATGAAGTCTCGCCAAGTTGATCAAAGTGATGCACCGCCTGATG
AATATTATTAAGTTATTACTTTCGAAACATAATAGAGGATAAGAATTAATGAGCGGGTGG
TTGTCGACGGTTCTATCATCCTTGCATTGCATCTCGAGTACGATGGGATACGAGTCGACA
CTTTTGCCAGCTTGGTAGTTAGTTAGGAAGCAGACCAGCTGCCGCGTCACGAGTTTACGT
TTGGGTCCGAAGGAAGTTCGTGAAATTCGGGGGCTCCCCCCTCGGGAAGTGGTGAAATTG
AGGTGTATCTTCTTCCAAACCAAACACGCGACGCACATTCTGTCTACTCTCGTCACACAC
CAAACAATTGAACC
>NO_promoter
ATCGAATCTATTCTGTCGGCGTTGTCGGTGGTATGAGTATGAGGTGGTATCTGGTGGTAT
GTATGTCGTGGCAGGAGTCGGTTACACTCCCTCGTCCCGAGTTCGAAATCGCGCGCACGA
CGATTGCAGCTCTGGTTGGCTTCTGGCTCTTCTGGTTTGGTCGCGCAGTTTCTGCCCGGA
TATCTTTTCGATCTTGTACAATATCGTCCGAGCGAGCTGTGCTATCTTCTACAAGTCGTG
TCACTTGCTACATATACCTATGCTTGTAGTATTGACTGTGTCCTTGTTGACTTGCAACAA
GGACCAAGGGTGGATGCAGTTAAAAAAGTTTGTGTTGCGGTGCTGTACCGTGCTATTGTA
GAGCTAGCTACTAGTAATAGTAGGGTAGCTAATAAGATACCTCTTCTAGATCTTAATCCC
GCGGTTAGTTCTCAGCTCCATCGCTCTTCGATAGCTCAAATTACACGTCATCAACTGCCT
GCTGGATGAAACGAAGCCCCGAAGCAAAACACGACTCCATATCCCGCTGTTTCGTGCCTA
ATCATCGTTTGTTTCTCATTGTTCTCTCACGTACACACTGGCGCATGCTGCTGCTAATCG
TGATAATAAGACGCGGCTTCTGCCTTCTACCGTAGGTACTGGAAGTTTCGATTGGGAGGG
TGCTACTTCCGGCGGCAACGAGTTGCGCTCCGTCGGTGTTTTTAGCAGCAGCGAGCGCGG
GCATGCGGCAGCACAGTAGCGGGATCTCGCATCATCGGGAGCCGCCCGCGGCGGCGGAGA
GTACGATAGCTAGGCCGAGAGCAGGATGCGATGATGTGGCCTCCTTCGACTCCTTCAGCT
TCTTTATGGTGCTCGAATCTATTATCTGCTGGCTGACGACGCGACGACGGTAAGGGCGGG
CGGATCAGCCAGGCAGACGGATCAACTCTCCCAGCTGGTACGATCCTTCGCGGAGCCGAG
TGTACGAAGGACTTGCTGCGACCATCACACTTTCCATTCATTCGACAGAAACAACGTCCT
CGTCCAATCTTGACGAGCTGGGGAGGAACACTTACCAACTCACCCACTCGAGACATTCTT
TACGAGGAAAACACTGATAGATCCAATCGCACGCCGAGAGCGAGAGCACATCATCGAAGG
AGTGAAGCATTGAAGACAGACAACC
>FBA_promoter

GACTGAGCCACCAGGAACCCCGGCCGGAAATGTGACTCGATACGCTGCCTTGGGCTCACT
CGATTAGAGCAATTTATTTGTTTCTCTTGACTCTCCCTCTTTACTTCTTTGCTCTCGAGA
ACCTTGCTCTCGAGAACCAGCGGGCAGCACAACAGAGGAAATGAACCTCTGAGCACAATA
GTAAGTACCAGTCAACCATTCCTAACAATAACATTGGCACATTGGCATCATGATCATCAT
GCAACACGTCAAAGATGAGACAGGAGCAATGAGTGATGCGATCAAGCATTATTGACATAT
AATTATAACAATATAATGATCATGCAACACGTGACGGGAGTGTCCCTATTTTTTCCTGCC
GTGACACTCGCCCTTCACTGTATTGCTTGCTGTCTGGTAGTGCATCTGCTCACACCTACT
ACCACCTACCAAGATCTTTTGGCAGGGGATGACTGGTATGCCATAGTTTGTGTGTCCCCG
CCTCTTCCCTAACTAAACGAACTAACTTATACGGAGATGATTTCGATTTAATCCTCGTGT
GTCCTTCGTTTCAGCTTGTAAGAGACGGATTTCTCCGCTTCCAACGGTACTGATAAACTG
CTGTAAACTGACAATGTGCTGTGTGCACCCGGTCAGGGTCTAGGGACCGGGCGGCACTGC
TCCCAGCGCTGATACATACTAGGCATAGTATGTACAGATAAGCTGGGCGATGATCATGGT
GACTGGCATCTTACATAGGTACATGACATGATGAGCTACAGCAGTGTCCTGCTACAGGGG
CTGGGCCCCACACACGGCATGGAAACAACAACGTAGGGGGCGCACGTGTACGTGGGCGGT
GACGCTGACGTTTGTACTTGACTGCACACAGCTAGCTACTTTATTGGCATTGGCACAGCG
CCAGGTGGGTGGTTTTGTGGGGCAGGGACGGCGACGCGACGGCGACAGCAGTTGTCGGCG
CTTCGGAAGCCTTTGAGATCACATTCTTCGTCGACGTTTGATTGTAATCTACCCAACC
>Hsp70_promoter
GACCGCATTGTTCTTGTTGACGGCAAGGTGGTGAGTAAAACTGAAGACCTACTGAAAGAC
ACCAACAGGGACCGTATCATTGGTATAACCAGAGGTTAGGAATAAGCATCTAATACCAAG
GTTGATGATCACCCTCTCTTTGGACTGTCTCTTTCCCTCTCCTTGCGCGGGCAGCGCGCG
GGCAGCACAAAGAGCCCGACCTCTGGGGCCCTCCCTTCATTCCAGCCCCCCCTTTCGCCC
TCTCCCAAGGACCCCTCTCCCCCCCCTCGGCCCTGGGGGTAGGGTGTATCTCCACGTGGA
TTCCCTCCATGGGAGTATGTGTTTTCGTTCACATCCGATCAAGCGCTCACAAAACACAAC
GTGAAAACACTATATTGATTCTGTGCCCCCCCCAATAGCCTCACTAAAACAAATCAACCG
GTTGGTATTATTTGCCACCGTGGAAATTTCTTGTGCGTGGTTTAGACGTTGTTAGACCGG
CCGCCAACAACGACACAACGACCCATGACCACCCACCGCACGTTCATCTCGCATAATCTA
CACCAACCTTTGCAGTTGAGTGTTGTATTCCAACTCTATCTCTTTGAGAACGATAATAAC
ACAACAAACA
>gene cassette :
CGGCCGCTCTAGAACTAGTGGATCCCCCGAGATATCACAAGTTTGTACAAAAAAGCTGAA
CGAGAAACGTAAAATGATATAAATATCAATATATTAAATTAGATTTTGCATAAAAAACAG
ACTACATAATACTGTAAAACACAACATATCCAGTCACTATGGCGGCCGCATTAGGCACCC
CAGGCTTTACACTTTATGCTTCCGGCTCGTATAATGTGTGGATTTTGAGTTAGGATCCGT
CGAGATTTTCAGGAGCTAAGGAAGCTAAAATGGAGAAAAAAATCACTGGATATACCACCG
TTGATATATCCCAATGGCATCGTAAAGAACATTTTGAGGCATTTCAGTCAGTTGCTCAAT
GTACCTATAACCAGACCGTTCAGCTGGATATTACGGCCTTTTTAAAGACCGTAAAGAAAA
ATAAGCACAAGTTTTATCCGGCCTTTATTCACATTCTTGCCCGCCTGATGAATGCTCATC
CGGAATTCCGTATGGCAATGAAAGACGGTGAGCTGGTGATATGGGATAGTGTTCACCCTT
GTTACACCGTTTTCCATGAGCAAACTGAAACGTTTTCATCGCTCTGGAGTGAATACCACG
ACGATTTCCGGCAGTTTCTACACATATATTCGCAAGATGTGGCGTGTTACGGTGAAAACC
TGGCCTATTTCCCTAAAGGGTTTATTGAGAATATGTTTTTCGTCTCAGCCAATCCCTGGG
TGAGTTTCACCAGTTTTGATTTAAACGTGGCCAATATGGACAACTTCTTCGCCCCCGTTT

TCACCATGGGCAAATATTATACGCAAGGCGACAAGGTGCTGATGCCGCTGGCGATTCAGG
TTCATCATGCCGTTTGTGATGGCTTCCATGTCGGCAGAATGCTTAATGAATTACAACAGT
ACTGCGATGAGTGGCAGGGCGGGGCGTAAACGCGTGGATCCGGCTTACTAAAAGCCAGAT
AACAGTATGCGTATTTGCGCGCTGATTTTTGCGGTATAAGAATATATACTGATATGTATA
CCCGAAGTATGTCAAAAAGAGGTATGCTATGAAGCAGCGTATTACAGTGACAGTTGACAG
CGACAGCTATCAGTTGCTCAAGGCATATATGATGTCAATATCTCCGGTCTGGTAAGCACA
ACCATGCAGAATGAAGCCCGTCGTCTGCGTGCCGAACGCTGGAAAGCGGAAAATCAGGAA
GGGATGGCTGAGGTCGCCCGGTTTATTGAAATGAACGGCTCTTTTGCTGATGAGAACAGG
GACTGGTGAAATGCAGTTTAAGGTTTACACCTATAAAAGAGAGAGCCGTTATCGTCTGTT
TGTGGATGTTCAGAGTGACATTATTGACACGCCCGGGCGACGGATGGTGATCCCCCTGGC
CAGTGCCCGCCTGCTGTCAGACAAAGTCTCCCGTGAGCTTTACCCGGTGGTGCATGTCGG
GGATGAAAGCTGGCGCATGATGACCACCGATATGGCCAGTGTGCCGGTCTCCGTTATCGG
GGAAGAAGTGGCTGATCTCAGTCACCGCGAAAATGACATCAAAAACGCCATTAACCTGAT
GTTCTGGGGAATATAAATGTCAGGCTCCCTTATACACAGCCAGTCTGCAGGTCGACCATA
GTGACTGGATATGTTGTGTTTTACAGTATTATGTAGTCTGTTTTTTATGCAAAATCTAAT
TTAATATATTGATATTTATATCATTTTACGTTTCTCGTTCAGCTTTCTTGTACAAAGTGG
TGATATCCATGGTGAGCAAGGGCGAGGAGCTGTTCACCGGGGTGGTGCCCATCCTGGTCG
AGCTGGACGGCGACGTAAACGGCCACAAGTTCAGCGTGTCCGGCGAGGGCGAGGGCGATG
CCACCTACGGCAAGCTGACCCTGAAGTTCATCTGCACCACCGGCAAGCTGCCCGTGCCCT
GGCCCACCCTCGTGACCACCTTCGGCTACGGCCTGCAGTGCTTCGCCCGCTACCCCGACC
ACATGAAGCAGCACGACTTCTTCAAGTCCGCCATGCCCGAAGGCTACGTCCAGGAGCGCA
CCATCTTCTTCAAGGACGACGGCAACTACAAGACCCGCGCCGAGGTGAAGTTCGAGGGCG
ACACCCTGGTGAACCGCATCGAGCTGAAGGGCATCGACTTCAAGGAGGACGGCAACATCC
TGGGGCACAAGCTGGAGTACAACTACAACAGCCACAACGTCTATATCATGGCCGACAAGC
AGAAGAACGGCATCAAGGTGAACTTCAAGATCCGCCACAACATCGAGGACGGCAGCGTGC
AGCTCGCCGACCACTACCAGCAGAACACCCCCATCGGCGACGGCCCCGTGCTGCTGCCCG
ACAACCACTACCTGAGCTACCAGTCCGCCCTGAGCAAAGACCCCAACGAGAAGCGCGATC
ACATGGTCCTGCTGGAGTTCGTGACCGCCGCCGGGATCACTCTCGGCATGGACGAGCTGT
ACAAGGAATTCTGAG

>Fcp10 Terminator

GTTTCCGGACTTGTCAATATCTCCACCGAAGTGAACTCCAGCACGCGTGATAAGGTTTCC
GACGAATGCGAGCTGCGAAACACGTCCGTGCTTAAGCTCGACATATCGCAGACGGTCGAA
AGTCTCCTGGTCGGCATTGTCGAGGAAGCCAAGAGGGTCCCAGAAGCCAAGAGGAGCCTG
TGCACCGAGCTCAGACTCGAACTATAGTGCCAACACATGTAATTTTTTTCAATTCCGAAA
AGTCCATCATCCGCACACCGCACAATTGCACCGCGAGTGAGCGCCGAAACGATGGACGTA
AGAGATTACGTCATTAGTGAGTAGCGTTTAGCGTTTATTGATAAGCCGGCAGCGGGCACA
TCGCCAACTTACGGCGTTGAGCTGAGTGGAGGCCTTGCCGGTCTGGGCAGGGGCGAAGGC
GGCGGCGGAGCCGATAAGGGCAAGGGCGGCAACAGACTTCATCATGGTTGGTTTGTTACG
ATGGTGCAAAAATGGGGTCTACAGGCAGCAGGCGGACTCTTCGAAACCAGCTCGCCTCTC
TTGCGTGAGGGTCAACATTACATTGGTACTCCCATTCATTATTAATTACTCAACTCTACC
AAAATATATTTTCCATTGGTTCATTCTCTATGAATATACCTAGGTAAAGTAACATTATGG
AAAAAATGCGACGCTAAATTACTAATGCTACCCCTTCATAGTTGAACGTATCTTAGCAAC
GTGTGACGGGCGCGGAGGCAAGCAGTATCGATATCGTAGCCCGATCTTTGGCCTTCCTAC

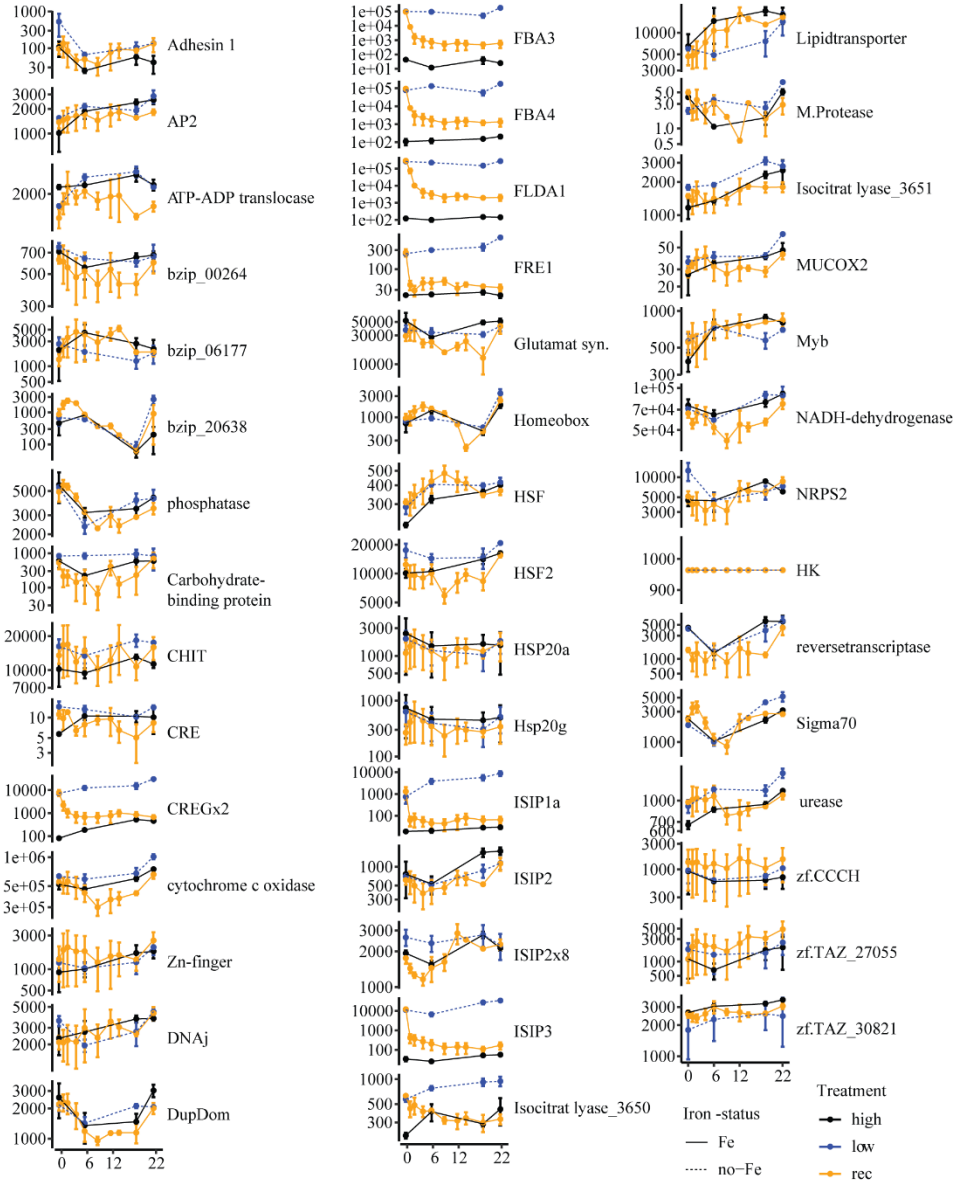
CTACTTGGTTGTTCTCACATTTCAAACCTTTGGAACAATTACTGTGAGTTAGCCGGAGGG
GAGCAATCGACTCGGAGACTGCTACGAACCGGTTTGATCTCCATGTCTTGTGTCATGTGA
TAACACTCACGCTTGTTCTTCGACTGCAGCAAAAATGAAGTCTGTTGCCGC
>Full-length sequence of pdestc'YFP-oceanica
TATCAGGGCGATGGCCCACTACGTGAACCATCACCCTAATCAAGTTTTTTGGGGTCGAGG
TGCCGTAAAGCACTAAATCGGAACCCTAAAGGGAGCCCCCGATTTAGAGCTTGACGGGGA
AAGCCGGCGAACGTGGCGAGAAAGGAAGGGAAGAAAGCGAAAGGAGCGGGCGCTAGGGCG
CTGGCAAGTGTAGCGGTCACGCTGCGCGTAACCACCACACCCGCCGCGCTTAATGCGCCG
CTACAGGGCGCGTCCCATTCGCCATTCAGGCTGCGCAACTGTTGGGAAGGGCGATCGGTG
CGGGCCTCTTCGCTATTACGCCAGCTGGCGAAAGGGGGATGTGCTGCAAGGCGATTAAGT
TGGGTAACGCCAGGGTTTTCCCAGTCACGACGTTGTAAAACGACGGCCAGTGAATTGTAA
TACGACTCAC[TTCGTCCGTATGAACGCCGAGAAAATTAGTGACACCCTTGCCTTCGCTT
CGCCAGCTTACGCTTGACTTCGCTCGCGATCGCCCAGCATGCTTGGCAGTGGGGGGCGCC
TTAATCTCTCGGCCCCCTGGCCACTATTGTTATGGTTGCGAGAAATCGATGAATGAACCT
CACAGCGTCGCACAGGCGTATGGAGCGGCGCGGCGCGCGGCCGGCGGCGCATACGATA
AGCCTAAGCACCTGGATCCACTTCAACTTCAGCCGTGAGTAAGAATTTGCACGAAATGAA
CGGGATTGTGCATGAAAAGGTAACAGCATGCTCCCTTTGCATGGTAAAGAAACCAAAGGT
GCGTTTTCTTCAGATTTTGATTTCCAATATTTGCATCATTTTTAGTACAACGAGCAAAGC
CGAAGCCGCATCGGCTATTAGACGAACACAGTTACAAGAGACTCATTTTGAACCTCGACG
GTATCGATAATATTCTAGCTGAGGGTACCCATGACCACTCTTGACGACACGGCTTACCGG
TACCGCACCAGTGTCCCGGGGGACGCCGAGGCCATCGAGGCACTGGATGGGTCCTTCACC
ACCGACACCGTCTTCCGCGTCACCGCCACCGGGGACGGCTTCACCCTGCGGGAGGTGCCG
GTGGACCCGCCCCTGACCAAGGTGTTCCCCGACGACGAATCGGACGACGAATCGGACGAC
GGGGAGGACGGCGACCCGGACTCCCGGACGTTCGTCGCGTACGGGGACGACGGCGACCTG
GCGGGCTTCGTGGTCATCTCGTACTCGGCGTGGAACCGCCGGCTGACCGTCGAGGACATC
GAGGTCGCCCCGGAGCACCGGGGGCACGGGGTCGGGCGCGCGTTGATGGGGCTCGCGACG
GAGTTCGCCGGCGAGCGGGGCGCCGGGCACCTCTGGCTGGAGGTCACCAACGTCAACGCA
CCGGCGATCCACGCGTACCGGCGGATGGGGTTCACCCTCTGCGGCCTGGACACCGCCCTG
TACGACGGCACCGCCTCGGACGGCGAGCGGCAGGCGCTCTACATGAGCATGCCCTGCCCC
TAGCCGACGCCGACCAACACCGCCGGTCCGACGCGGCCCGACGGGTCCGAGGCCTCGGAG
ATCTGGGCCCATGCGGTGCCAACACATGTAATTTTTTTCAATTCCGAAAAGTCCATCATC
CGCACACCGCACAATTGCACCGCGAGTGAGCGCCGAAACGATGGACGTAAGAGATTACGT
CATTAGTGAGTAGCGTTTAGCGTTTATTGATAAGCCGGCAGCGGGCACATCGCCAACTTA
CGGCGTTGAGCTGAGTGGAGGCCTTGCCGGTCTGGGCAGGGGCGAAGGCGGCGGCGGAGC
CGATAAGGGCAAGGGCGGCAACAGACTTCATCATGGTTGGTTTGTTACGATGGTGCAAAA
ATGGGGTCTACAGGCAGCAGGCGGACTCTTCGAAACCAGCTCGCCTCTCTTGCGTGAGGG
TCAACATTACATTGGTACTCCCATTCATTATTAATTACTCAACTCTACCAAAATATATTT
TCCATTGGTTCATTCTCTATGAATATACCTAGGTAAAGTAACATTATGGAAAAAATGCGA
CGCTAAATTACTAATGCTACCCCTTCATAGTTGAACGTATCTTAGCAACGTGTGACGGGC
GCGGAGGCAAGCAGTATCGATATCGTAGCCCGATCTTTGGCCTTCCTACCTACTTGGTTG
TTCTCACATTTCAAACCTTTGGAACAATTACTGTGAGTTAGCCGGAGGGGAGCAATCGAC
TCGGAGACTGCTACGAACCGGTTTGATCTCCATGTCTTGTGTCATGTGATAACACTCACG
CTTGTTCTTCGACTGCAGCAAAAATGAAGTCTGTTGCCGCAATCTCGCCTATTCATGGTG

TATACAAGAGGAGAATACATCGAGAAGATACTCCTCGTTGGACCATGTTGAGCCGTCTCT
TTCCTCCGATACGAACCCGGTGCTCTAGGAGGACGAGAAGACGAAAAGGGAATCAAAAGG
CATCTGCGGTTAGGCGGGCATCAAAGAATGCACGGACCAAGCCGCCTGTATGTAACGCCG
AACGAAGAAAACAGTTGGCTAGAGAGTGAATCTATCAGCGTTCGGCGTTACGTTGTGTTA
GGTTACAATGCTTCTGTACGTTTGCAAAATGAAAGAGAACGTGGCAGAAAGGAAACGAAA
ATGGAAGGTGACGATTGACGTGGAAAATGGCTCCTTCTGGCCATGAACAGCCTTTGTGAT
AAAAGTAGGTACGGCTACAGACCGTGCCCAAACAAAGCGTCATTAGAAGGCGTGTCACAA
AGAACACCGCAAGCAAGCAGAACCCACGATGTCTGAAGAGTGACGATCCGTCGGTGAATA
AAAGAGAACGTTCGTGTCGATCTTATTCGGCCGTAAAACGAAAGCAAGCTGTTGATTAAC
GGCGCGGCGCGACGGAGCATGGCCGTGAGAATGATATGGAGAAAAGAAGATCACATTGCC
TTCCTTGAGAGGAGAATGAAAACACTACACTTCAGGAGGATGAAGTCTCGCCAAGTTGAT
CAAAGTGATGCACCGCCTGATGAATATTATTAAGTTATTACTTTCGAAACATAATAGAGG
ATAAGAATTAATGAGCGGGTGGTTGTCGACGGTTCTATCATCCTTGCATTGCATCTCGAG
TACGATGGGATACGAGTCGACACTTTTGCCAGCTTGGTAGTTAGTTAGGAAGCAGACCAG
CTGCCGCGTCACGAGTTTACGTTTGGGTCCGAAGGAAGTTCGTGAAATTCGGGGGCTCCC
CCCTCGGGAAGTGGTGAAATTGAGGTGTATCTTCTTCCAAACCAAACACGCGACGCACAT
TCTGTCTACTCTCGTCACACACCAAACAATTGAACCCGGCCGCTCTAGAACTAGTGGATC
CCCCGAGATATCACAAGTTTGTACAAAAAAGCTGAACGAGAAACGTAAAATGATATAAAT
ATCAATATATTAAATTAGATTTTGCATAAAAAACAGACTACATAATACTGTAAAACACAA
CATATCCAGTCACTATGGCGGCCGCATTAGGCACCCCAGGCTTTACACTTTATGCTTCCG
GCTCGTATAATGTGTGGATTTTGAGTTAGGATCCGTCGAGATTTTCAGGAGCTAAGGAAG
CTAAAATGGAGAAAAAAATCACTGGATATACCACCGTTGATATATCCCAATGGCATCGTA
AAGAACATTTTGAGGCATTTCAGTCAGTTGCTCAATGTACCTATAACCAGACCGTTCAGC
TGGATATTACGGCCTTTTTAAAGACCGTAAAGAAAAATAAGCACAAGTTTTATCCGGCCT
TTATTCACATTCTTGCCCGCCTGATGAATGCTCATCCGGAATTCCGTATGGCAATGAAAG
ACGGTGAGCTGGTGATATGGGATAGTGTTCACCCTTGTTACACCGTTTTCCATGAGCAAA
CTGAAACGTTTTCATCGCTCTGGAGTGAATACCACGACGATTTCCGGCAGTTTCTACACA
TATATTCGCAAGATGTGGCGTGTTACGGTGAAAACCTGGCCTATTTCCCTAAAGGGTTTA
TTGAGAATATGTTTTTCGTCTCAGCCAATCCCTGGGTGAGTTTCACCAGTTTTGATTTAA
ACGTGGCCAATATGGACAACTTCTTCGCCCCCGTTTTCACCATGGGCAAATATTATACGC
AAGGCGACAAGGTGCTGATGCCGCTGGCGATTCAGGTTCATCATGCCGTTTGTGATGGCT
TCCATGTCGGCAGAATGCTTAATGAATTACAACAGTACTGCGATGAGTGGCAGGGCGGGG
CGTAAACGCGTGGATCCGGCTTACTAAAAGCCAGATAACAGTATGCGTATTTGCGCGCTG
ATTTTTGCGGTATAAGAATATATACTGATATGTATACCCGAAGTATGTCAAAAAGAGGTA
TGCTATGAAGCAGCGTATTACAGTGACAGTTGACAGCGACAGCTATCAGTTGCTCAAGGC
ATATATGATGTCAATATCTCCGGTCTGGTAAGCACAACCATGCAGAATGAAGCCCGTCGT
CTGCGTGCCGAACGCTGGAAAGCGGAAAATCAGGAAGGGATGGCTGAGGTCGCCCGGTTT
ATTGAAATGAACGGCTCTTTTGCTGATGAGAACAGGGACTGGTGAAATGCAGTTTAAGGT
TTACACCTATAAAAGAGAGAGCCGTTATCGTCTGTTTGTGGATGTTCAGAGTGACATTAT
TGACACGCCCGGGCGACGGATGGTGATCCCCCTGGCCAGTGCCCGCCTGCTGTCAGACAA
AGTCTCCCGTGAGCTTTACCCGGTGGTGCATGTCGGGGATGAAAGCTGGCGCATGATGAC
CACCGATATGGCCAGTGTGCCGGTCTCCGTTATCGGGGAAGAAGTGGCTGATCTCAGTCA
CCGCGAAAATGACATCAAAAACGCCATTAACCTGATGTTCTGGGGAATATAAATGTCAGG

CTCCCTTATACACAGCCAGTCTGCAGGTCGACCATAGTGACTGGATATGTTGTGTTTTAC
AGTATTATGTAGTCTGTTTTTTATGCAAAATCTAATTTAATATATTGATATTTATATCAT
TTTACGTTTCTCGTTCAGCTTTCTTGTACAAAGTGGTGATATCCATGGTGAGCAAGGGCG
AGGAGCTGTTCACCGGGGTGGTGCCCATCCTGGTCGAGCTGGACGGCGACGTAAACGGCC
ACAAGTTCAGCGTGTCCGGCGAGGGCGAGGGCGATGCCACCTACGGCAAGCTGACCCTGA
AGTTCATCTGCACCACCGGCAAGCTGCCCGTGCCCTGGCCCACCCTCGTGACCACCTTCG
GCTACGGCCTGCAGTGCTTCGCCCGCTACCCCGACCACATGAAGCAGCACGACTTCTTCA
AGTCCGCCATGCCCGAAGGCTACGTCCAGGAGCGCACCATCTTCTTCAAGGACGACGGCA
ACTACAAGACCCGCGCCGAGGTGAAGTTCGAGGGCGACACCCTGGTGAACCGCATCGAGC
TGAAGGGCATCGACTTCAAGGAGGACGGCAACATCCTGGGGCACAAGCTGGAGTACAACT
ACAACAGCCACAACGTCTATATCATGGCCGACAAGCAGAAGAACGGCATCAAGGTGAACT
TCAAGATCCGCCACAACATCGAGGACGGCAGCGTGCAGCTCGCCGACCACTACCAGCAGA
ACACCCCCATCGGCGACGGCCCCGTGCTGCTGCCCGACAACCACTACCTGAGCTACCAGT
CCGCCCTGAGCAAAGACCCCAACGAGAAGCGCGATCACATGGTCCTGCTGGAGTTCGTGA
CCGCCGCCGGGATCACTCTCGGCATGGACGAGCTGTACAAGTAAGAATTCTGAGTGCCAA
CACATGTAATTTTTTTCAATTCCGAAAAGTCCATCATCCGCACACCGCACAATTGCACCG
CGAGTGAGCGCCGAAACGATGGACGTAAGAGATTACGTCATTAGTGAGTAGCGTTTAGCG
TTTATTGATAAGCCGGCAGCGGGCACATCGCCAACTTACGGCGTTGAGCTGAGTGGAGGC
CTTGCCGGTCTGGGCAGGGGCGAAGGCGGCGGCGGAGCCGATAAGGGCAAGGGCGGCAAC
AGACTTCATCATGGTTGGTTTGTTACGATGGTGCAAAAATGGGGTCTACAGGCAGCAGGC
GGACTCTTCGAAACCAGCTCGCCTCTCTTGCGTGAGGGTCAACATTACATTGGTACTCCC
ATTCATTATTAATTACTCAACTCTACCAAAATATATTTTCCATTGGTTCATTCTCTATGA
ATATACCTAGGTAAAGTAACATTATGGAAAAAATGCGACGCTAAATTACTAATGCTACCC
CTTCATAGTTGAACGTATCTTAGCAACGTGTGACGGGCGCGGAGGCAAGCAGTATCGATA
TCGTAGCCCGATCTTTGGCCTTCCTACCTACTTGGTTGTTCTCACATTTCAAACCTTTGG
AACAATTACTGTGAGTTAGCCGGAGGGGAGCAATCGACTCGGAGACTGCTACGAACCGGT
TTGATCTCCATGTCTTGTGTCATGTGATAACACTCACGCTTGTTCTTCGACTGCAGCAAA
AATGAAGTACCAGCTTTTGTTCCCTTTAGTGAGGGTTAATTTCGAGCTTGGCGTAATCAT
GGTCATAGCTGTTTCCTGTGTGAAATTGTTATCCGCTCACAATTCCACACAACATACGAG
CCGGAAGCATAAAGTGTAAAGCCTGGGGTGCCTAATGAGTGAGCTAACTCACATTAATTG
CGTTGCGCTCACTGCCCGCTTTCCAGTCGGGAAACCTGTCGTGCCAGCTGCATTAATGAA
TCGGCCAACGCGCGGGGAGAGGCGGTTTGCGTATTGGGCGCTCTTCCGCTTCCTCGCTCA
CTGACTCGCTGCGCTCGGTCGTTCGGCTGCGGCGAGCGGTATCAGCTCACTCAAAGGCGG
TAATACGGTTATCCACAGAATCAGGGGATAACGCAGGAAAGAACATGTGAGCAAAAGGCC
AGCAAAAGGCCAGGAACCGTAAAAAGGCCGCGTTGCTGGCGTTTTTCCATAGGCTCCGCC
CCCCTGACGAGCATCACAAAAATCGACGCTCAAGTCAGAGGTGGCGAAACCCGACAGGAC
TATAAAGATACCAGGCGTTTCCCCCTGGAAGCTCCCTCGTGCGCTCTCCTGTTCCGACCC
TGCCGCTTACCGGATACCTGTCCGCCTTTCTCCCTTCGGGAAGCGTGGCGCTTTCTCATA
GCTCACGCTGTAGGTATCTCAGTTCGGTGTAGGTCGTTCGCTCCAAGCTGGGCTGTGTGC
ACGAACCCCCCGTTCAGCCCGACCGCTGCGCCTTATCCGGTAACTATCGTCTTGAGTCCA
ACCCGGTAAGACACGACTTATCGCCACTGGCAGCAGCCACTGGTAACAGGATTAGCAGAG
CGAGGTATGTAGGCGGTGCTACAGAGTTCTTGAAGTGGTGGCCTAACTACGGCTACACTA
GAAGGACAGTATTTGGTATCTGCGCTCTGCTGAAGCCAGTTACCTTCGGAAAAAGAGTTG

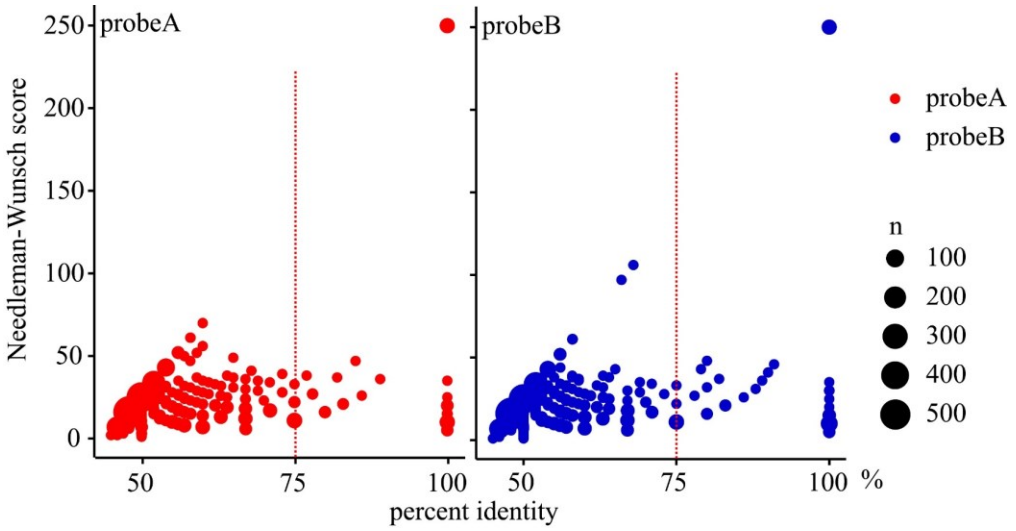
GTAGCTCTTGATCCGGCAAACAAACCACCGCTGGTAGCGGTGGTTTTTTTGTTTGCAAGC
AGCAGATTACGCGCAGAAAAAAAGGATCTCAAGAAGATCCTTTGATCTTTTCTACGGGGT
CTGACGCTCAGTGGAACGAAAACTCACGTTAAGGGATTTTGGTCATGAGATTATCAAAAA
GGATCTTCACCTAGATCCTTTTAAATTAAAAATGAAGTTTTAAATCAATCTAAAGTATAT
ATGAGTAAACTTGGTCTGACAGTTACCAATGCTTAATCAGTGAGGCACCTATCTCAGCGA
TCTGTCTATTTCGTTCATCCATAGTTGCCTGACTCCCCGTCGTGTAGATAACTACGATAC
GGGAGGGCTTACCATCTGGCCCCAGTGCTGCAATGATACCGCGAGACCCACGCTCACCGG
CTCCAGATTTATCAGCAATAAACCAGCCAGCCGGAAGGGCCGAGCGCAGAAGTGGTCCTG
CAACTTTATCCGCCTCCATCCAGTCTATTAATTGTTGCCGGGAAGCTAGAGTAAGTAGTT
CGCCAGTTAATAGTTTGCGCAACGTTGTTGCCATTGCTACAGGCATCGTGGTGTCACGCT
CGTCGTTTGGTATGGCTTCATTCAGCTCCGGTTCCCAACGATCAAGGCGAGTTACATGAT
CCCCCATGTTGTGCAAAAAAGCGGTTAGCTCCTTCGGTCCTCCGATCGTTGTCAGAAGTA
AGTTGGCCGCAGTGTTATCACTCATGGTTATGGCAGCACTGCATAATTCTCTTACTGTCA
TGCCATCCGTAAGATGCTTTTCTGTGACTGGTGAGTACTCAACCAAGTCATTCTGAGAAT
AGTGTATGCGGCGACCGAGTTGCTCTTGCCCGGCGTCAATACGGGATAATACCGCGCCAC
ATAGCAGAACTTTAAAAGTGCTCATCATTGGAAAACGTTCTTCGGGGCGAAAACTCTCAA
GGATCTTACCGCTGTTGAGATCCAGTTCGATGTAACCCACTCGTGCACCCAACTGATCTT
CAGCATCTTTTACTTTCACCAGCGTTTCTGGGTGAGCAAAAACAGGAAGGCAAAATGCCG
CAAAAAAGGGAATAAGGGCGACACGGAAATGTTGAATACTCATACTCTTCCTTTTTCAAT
ATTATTGAAGCATTTATCAGGGTTATTGTCTCATGAGCGGATACATATTTGAATGTATTT
AGAAAAATAAACAAATAGGGGTTCCGCGCACATTTCCCCGAAAAGTGCCACCTAAATTGT
AAGCGTTAATATTTTGTTAAAATTCGCGTTAAATTTTTGTTAAATCAGCTCATTTTTTAA
CCAATAGGCCGAAATCGGCAAAATCCCTTATAAATCAAAAGAATAGACCGAGATAGGGTT
GAGTGTTGTTCCAGTTTGGAACAAGAGTCCACTATTAAAGAACGTGGACTCCAACGTCAA
AGGGCGAAAAACCGTC

APPENDIX B: SUPPLEMENTAL FIGURES



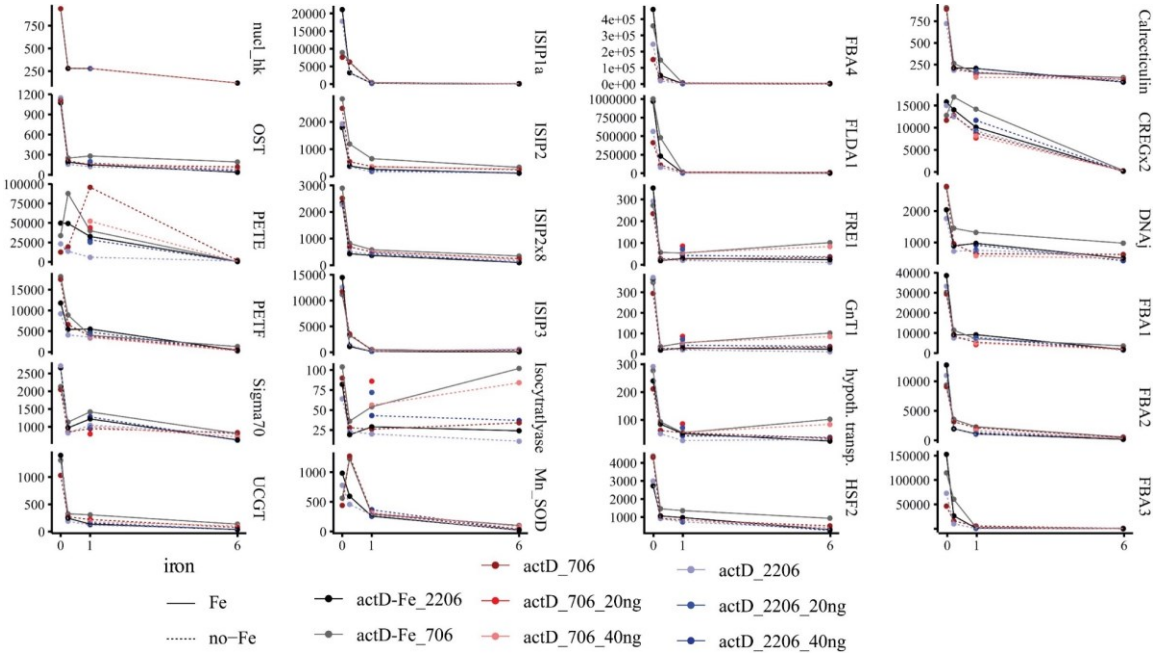
Supplemental Figure 1: Transcript counts of the first LT run over the period of 22 h.

Shown are the results from the first NanoString run of 44 targeted transcripts. The x-axes are in hours, and three treatments are shown. High-iron samples are black, low-iron samples are blue, and iron-recovery samples are orange. The addition of 10 µM $FeCl_3$ was done after the initial measurement at timepoint 0. Twelve of these transcripts were further analyzed in the ST experiment. The y-axes are on log scale.



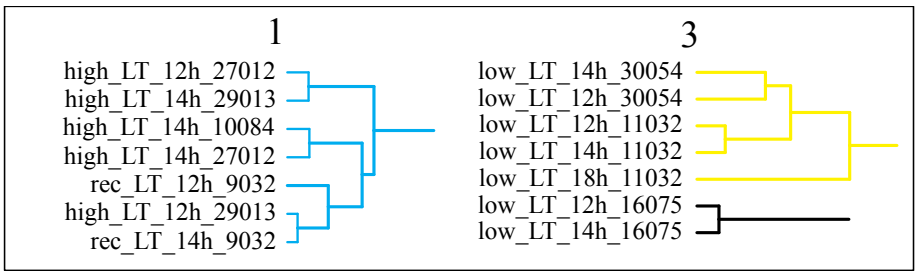
Supplemental Figure 2: Overview of percent identity of all probe sequences aligned to each gene analyzed in this study.

All target sequences were divided into their respective probeA and probeB resulting in 50bp length. Each probe was then aligned with all targeted genes that were analyzed in this study. The alignment was done using the Needleman-Wunsch algorithm with a gapopen penalty of 99 and a gapextend penalty of 10 (max), aiming to align the probes without gaps. X-axes is percent identity, and y-axes are Needleman-Wunsch score. The size of each circle represents the number of alignments included in this circle. The exact hit between probe and its target gene has 100% identity and a Needleman-Wunsch score of 250. The sequences higher than the 75% identity threshold show very low Needleman-Wunsch scores, which indicated partial alignment of the probes.



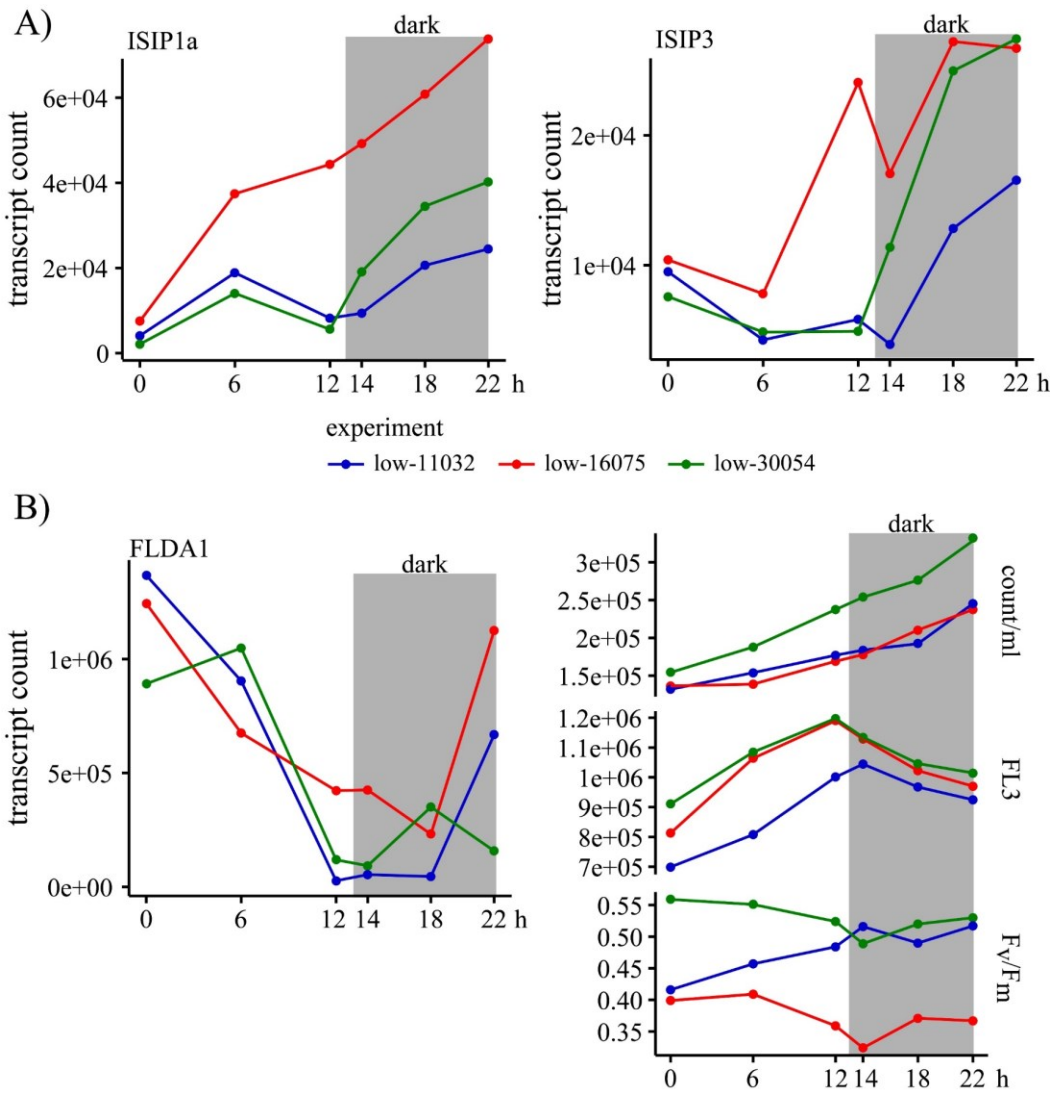
Supplemental Figure 3: Overview of transcript counts during the ST experiment treated with actinomycin D.

All 24 targeted transcripts that were used in the ST and LT experiment are shown. All measuring points of actD and actD-Fe are shown including samples that were measured as duplicates.



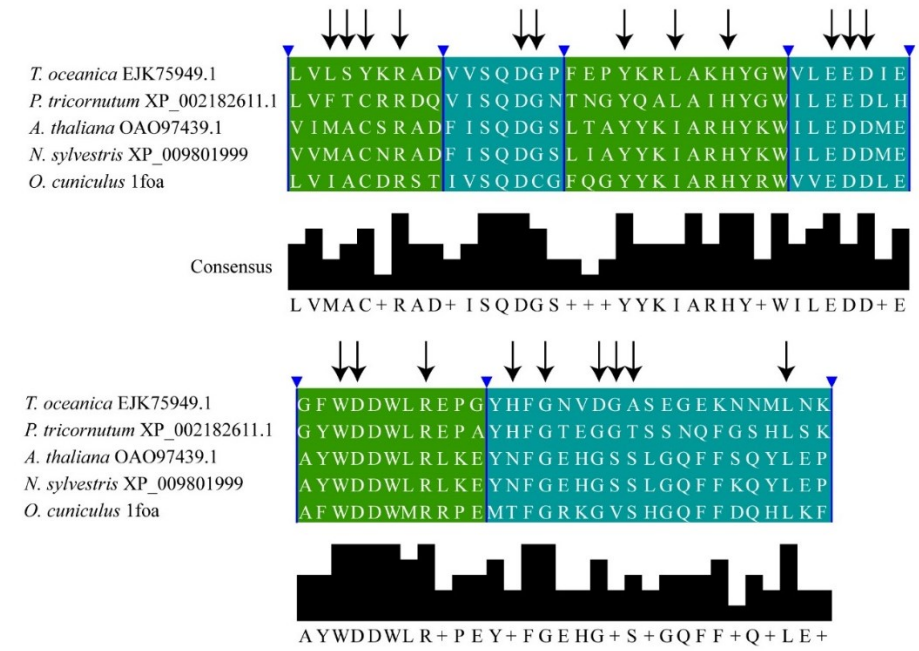
Supplemental Figure 4: Heatmap clusters 1 and 3 with sample-specific names.

Cluster 1 and 3 as outlined in the text and indicated in Figure 3.5 are shown with sample-specific names. Name is divided into: treatment (high/low), experiment-type (LT-long-term experiment) point of time after iron addition)_experiment-specific identifier. Twelve h and 14 h are 7 PM and 9 PM, respectively.



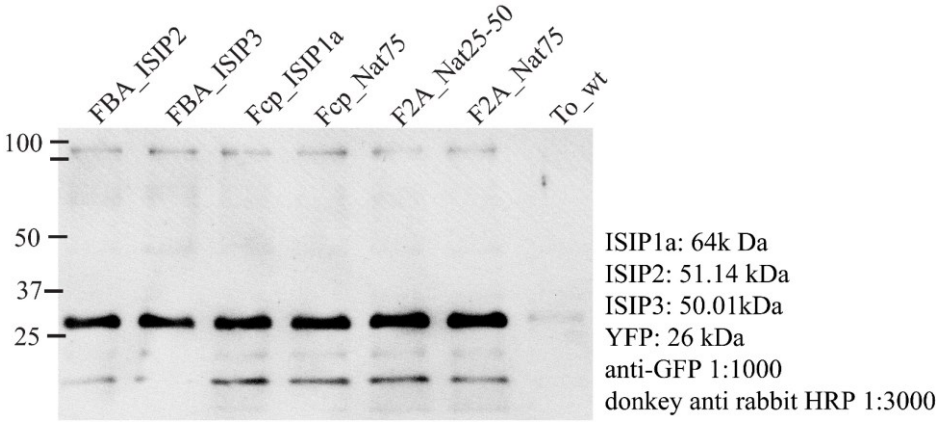
Supplemental Figure 5: ISIP1a and ISIP3 transcript counts in low-iron samples.

A) Transcript counts of ISIP1a, and ISIP3 in low-iron samples over the 22 h period. Each line represents one experiment: low-11032(blue), low-16075 (red), low-30054(green). The dark phase is indicated by a grey box. B) FLDA1, cell count, FL3 values, and F_v/F_m measurements are shown to reassure that the iron stress did not increase over time.



Supplemental Figure 6: Sequence alignment of GnT1 protein sequences.

The *T. ocanica* GnT1 protein sequence is on top and previously identified amino acids that are relevant for substrate binding are marked with an arrow (Baiet *et al.*, 2010). The sequence was aligned with *Phaeodactylum tricornutum* (*P. tricornutum*), *Arabidopsis thaliana* (*A. thaliana*), *Nicotiana sylvestris* (*N. sylvestris*), and *Oryctolagus cuniculus* (*O. cunniculus*).



Supplemental Figure 7: Western blot analysis of antibiotic-resistant *T. oceanica* cultures.

Shown is the detection of proteins with a green fluorescent protein (GFP) antibody on a nitrocellulose membrane. Antibiotic-resistant *T. oceanica* cultures were harvested and proteins were extracted. Each lane is loaded with 20 µg of protein lysate. The transformants were retained using suspension cultures without pour plating.