Molecular biology of B vitamin metabolism genes and their regulation in *Phaeodactylum tricornutum*

Marcel Llavero Pasquina

Department of Plant Sciences

Girton College, University of Cambridge

This thesis is submitted for the degree of Doctor of Philosophy

September 2020

Declaration

This thesis is the result of my own work and includes nothing which is the outcome of work done in collaboration except as declared in the preface and specified in the text.

It is not substantially the same as any work that has already been submitted before for any degree or other qualification except as declared in the preface and specified in the text.

It does not exceed the prescribed word limit for the Biology Degree Committee.

Acknowledgments

First of all, I would like to specially thank Prof. Alison Smith, my supervisor, for giving me the opportunity to do this PhD. I sincerely appreciate all the support, mentorship, advice, feedback and trust you have given me with tremendous patience over these 4 years. It has been an honour to learn from you.

I owe much of what I have learned and achieved in my PhD to Dr Katrin Geisler and Dr Payam Mehrshahi. They both have been really patient in sharing all their expertise and knowledge with me and have provided advice and support in the key moments of my time in the lab.

I would also like to thank Lorraine Archer and Sue Aspinall for being two wonderful lab managers, enabling much of the research we do in the lab and always giving a really positive and contagious attitude to the work we do.

My experience and research during these years would not have been the same without all the collaboration and idea exchanges I have had with all members of the lab. Every single interaction has been special and has taught me an important lesson. Not to mention the really encouraging and fun atmosphere the group has built around our research project. Thank you very much to all of you Patrick Hickland, Dr Andy Sayer, Andre Holzer, Dr Gonzalo Mendoza-Ochoa, Dr Freddy Bunbury, Aleix Gorchs-Rovira, Sam Coffin, Dr Ian Bower, Stefan Grossfurthner, Shelby Newsad, Dom Absolon, Ellen Harrison, An Tran, Carrie Faessler, Dr Maria Huete-Ortega, Dr Matt Davey, Dr Pawel Mordaka, Dr Doris Gangl, Beata Sikora, Katie Sutherland and Monika Krolikowski.

It has also been an honour to have supervised three great undergraduate students: Jessie, Elena and Yaning. I have learned from them, as much as I hope they have learned from the lab, and they have certainly contributed to the results we have achieved.

I would also like to thank Prof. Jim Haseloff and Prof. Alex Webb, my second supervisor and GEC adviser, for providing critical feedback at key moments of the PhD that have helped me consider new research approaches.

Dr Amanda Hopes (University of East Anglia) gave me a few starting plasmids and advice on using CRISPR in diatoms which has been the basis of one of this thesis chapters and Dr Harry Jackson (University College London) shared reagents and instructions to perform western blots on algal samples.

Finally, I would like to most earnestly thank Blanca for all the support she has given me before and during these 4 years of PhD.

Summary

Thiamine (Vitamin B_1) in its diphosphate form is an essential cofactor for virtually all organisms. Thiamine biosynthesis is an expensive metabolic process involving suicidal and low-turnover enzymes, and many organisms have lost the ability to synthesise thiamine and depend on an exogenous source. Those prototrophs that synthesise thiamine, including species of bacteria, fungi, green algae and plants, have been shown to tightly regulate their thiamine-related genes in response to exogenous thiamine. In many cases this is via riboswitches, sequences in mRNA that fold into a tertiary structure (aptamer) able to specifically recognise a ligand and mediate a change of genetic expression in response. In diatoms, marine algae responsible for ~20 % of net primary productivity, *THIC*, which encodes one of the first enzymes in the thiamine biosynthetic pathway, has been predicted to have a thiamine pyrophosphate (TPP) riboswitch in the 3'UTR. Additionally, *THIC* has been previously observed to be downregulated by cobalamin (vitamin B_{12}) supplementation in diatoms reflecting the interconnectedness and co-regulation of different B vitamin metabolisms. The aim of this thesis is to investigate different aspects of the regulation of thiamine and cobalamin-related genes and the transport and metabolism of thiamine and cobalamin in diatoms.

Homology-based bioinformatic tools confirmed that all available diatom genomes contain homologues of *THIC* encoding the first enzyme in the pyrimidine branch of the thiamine biosynthesis pathway in bacteria, and/or NMT1, which catalyses the equivalent step in fungi. Additionally, it was found that many of these genes, and those coding for SSSP, a putative thiamine transporter, had predicted TPP aptamers in their 3'UTRs. A conserved polyadenylation site was found overlapping the diatom TPP aptamer sequences, which might be involved in a hypothetical mechanism of action. However, experiments using RT-qPCR and 3'-RACE showed that the *THIC* and *SSSP* genes in *Phaeodactylum tricornutum* did not respond to thiamine supplementation at the transcriptional or post-transcriptional level, even though thiamine is taken up by the cells. Furthermore, unlike in the green alga *Chlamydomonas reinhardtii*, which has experimentally characterised TPP riboswitches, the diatoms *P. tricornutum* and *Thalassiosira pseudonana* were insensitive to pyrithiamine, a thiamine antimetabolite that primarily inhibits growth by binding TPP riboswitches and downregulating thiamine biosynthesis genes. Reporter constructs confirmed that the *PtTHIC* regulatory sequences (promoter, 5'UTR, 3'UTR) could not regulate heterologous genes in response to thiamine supplementation. Nor did the *PtTHIC* aptamer respond in *C. reinhardtii* chimeric constructs. Finally, site-directed mutagenesis of the *PtTHIC* aptamer did not alter endogenous thiamine levels in *P. tricornutum*, in contrast to

equivalent mutations in other organisms with confirmed riboswitches. Together, these results demonstrate that the predicted TPP aptamer in the *PtTHIC* gene does not act as a riboswitch.

RT-qPCR experiments suggested that *PtTHIC* is downregulated by cobalamin supplementation similarly to cobalamin-independent methionine synthase (*PtMETE*). A motif-prediction algorithm revealed a conserved 14 bp motif in the promoters of *THIC, METE* and other cobalamin-downregulated genes in diatoms that could indicate the co-regulation of cobalamin and thiamine metabolism. Reporter construct experiments confirmed that the *PtTHIC* promoter could downregulate a reporter in response to cobalamin and that the *PtMETE* promoter with a mutation in the motif did not drive the expression of a reporter, demonstrating that the conserved motif is necessary for gene expression.

Finally, a CRISPR/Cas9 method was developed in collaboration with other members of the group to generate knock-out mutants of several genes in *P. tricornutum*. The *THIC* knock-out did not require thiamine for growth in f/2 minimal media, suggesting *P. tricornutum* can obtain thiamine or its pyrimidine moiety from an alternative source or pathway. A knock-out of *SSSP* demonstrated that this gene is required for thiamine uptake. The *METE* knock-out was auxotrophic for cobalamin and the Cobalamin Acquisition Protein 1 (*CBA1*) knock-out confirmed it is necessary for cobalamin uptake.

Taken together, the results in this thesis provide new insights into thiamine and cobalamin metabolism, transport and genetic regulation in diatoms. Unexpected results such as the unaltered growth of the *THIC* knock-out and the lack of function of the *P. tricornutum* predicted TPP riboswitches stand in contrast with existing knowledge of thiamine metabolism in other organisms. This raises important questions to understand the role of thiamine in the physiology, ecology and evolution of an algal group with global ecological relevance.

Table of Contents

Table of Figures

Table of Tables

List of Abbreviations

3’RACE	Rapid Amplification of cDNA 3’Ends
5-FC	5-Fluorocytosine
5-FOA	5-Fluoroorotic Acid
ABC Transporter	ATP-Binding Cassette Transporter
AIR	5-aminoimidazole ribotide
ALA	5-aminolaevulinic acid
APC	Amino acid-Polyamine-organoCation
bHLH	Basic Helix-Loop-Helix
bZIP	Basic Leucine Zipper
CCAP	Culture Collection of Algae and Protozoa
CDS	Coding Sequence
ChIP-Seq	Chromatin Immunoprecipitation - Sequencing
CRISPR	Clustered Regularly Interspaced Short Palindromic Repeats
DMB	Dimethybenzimidazole
DXP	1-deoxy-d-xylulose 5-phosphate
ECF Transporters	Energy-Coupling Factor Transporters
EMSA	Electrophoretic Mobility Shift Assay
EST	Expressed Sequence Tag
FAMP	N-formyl-4-amino-5-(aminomethyl)-2-methylpyrimidine
GA3P	Glyceraldehide-3-phosphate
GFP	Green Fluorescence Protein
GGPP	Geranylgeranyl Pyrophosphate
GOI	Gene of Interest
HDR	Homology Directed Repair
HET	5-(2-hydroxyethyl)-4-methylthiazole
HMM	Hidden Markov Model
HMP	4-amino-5-hydroxymethyl-2-methylpyrimidine
HPLC	High Performance Liquid Chromatography
HRM	High Resolution Melting
MFS	Major Facilitator Superfamily
MoClo	Modular Cloning
MSA	Multiple Sequence Alignment
NAD	Nicotinamide Adenine Dinucleotide
NCBI	National Center for Biotechnology Information
NHEJ	Non-Homologous End Joining
ORF	Open Reading Frame
PBG	Porphobilinogen
PCR	Polymerase Chain Reaction
PDB	Protein Data Bank
PLP	Pyridoxal-5-Phosphate
PVDF	Polyvinylidene Difluoride
RMSD	Root-Mean-Square Deviation
RNP	Ribonucleoprotein
RT-PCR	Reverse Transcription - Polymerase Chain Reaction
RT-qPCR	Reverse Transcription - Quantitative Polymerase Chain Reaction
SAH	S-adenosylhomocysteine
SAM	S-adenosylmethionine
SDS-PAGE	Sodium Dodecyl Sulfate–Polyacrylamide Gel Electrophoresis
sgRNA	Single Guide RNA

SSF	Sodium-solute Symporter Family
SVM	Support Vector Machine
TALEN	Transcription activator-like effector nucleases
TCA Cycle	Tricarboxylic Acid Cycle
THF	Tetrahydrofolate
TIDE	Track INDELs by DEcomposition
TMP	Thiamine Monophosphate
TPP	Thiamine Pyrophosphate
uORF	Upstream Open Reading Frame
UTR	Untranslated Region
Y1H	Yeast-1-Hybrid
YFP	Yellow Fluorescence Protein

Chapter 1. Introduction

1.1. B vitamins

Vitamins are an heterogenous group of small organic molecules characterised by being a dietary requirement in very low doses for animals in general and humans in particular. Vitamins are generally classified according to their solubility. Fat-soluble vitamins include retinol (vitamin A), cholecalciferol (vitamin D), tocopherol (vitamin E) and phytomenadione (vitamin K). Water-soluble vitamins include ascorbic acid (vitamin C), biotin (vitamin H) and the B vitamin group. There are eight B vitamins, thiamine (B_1), riboflavin (B_2), niacin (B_3), pantothenate (B_5) pyridoxal (B_6), biotin (B_7), folate (B_9) and cobalamin (B_{12}). Within each B vitamin type there can be more than one chemical form, generally called vitamers, that can meet the human physiological requirements, either because they can carry out the same biochemical function or because they can be metabolised into the active form of the vitamin (Bender, 2003).

Apart from their general vitamin characteristics, all B vitamins share three key features (Figure 1.1). B vitamins are used as cofactors in enzymes of key metabolic pathways like the Calvin Cycle, the Tricarboxylic Acid Cycle, fatty acid biosynthesis and one carbon metabolism (Godoy-Parejo *et al.,* 2020; Monteverde *et al.,* 2017). Most B vitamins are required by all living organisms, they all display unique biochemical properties, and except for biotin, their molecular structure and biosynthesis is related to nitrogenous bases. These three characteristics are interrelated and highlight the metabolic, physiological, ecological and evolutionary role of B vitamins. Additionally, different B vitamin biosynthetic pathways and the metabolic pathways dependent on B vitamins are intricately connected, with different instances of vitamins used as substrates or cofactors in other B vitamin biosynthesis pathways (Monteverde *et al.,* 2017). For instance, in many organisms cobalamin acts as a cofactor of methionine synthase, an enzyme in the C1 cycle necessary for the production of S-adenosylmethionine (SAM). SAM, in turn, is required by one of the enzymes in the thiamine biosynthetic pathway. Based on these observations, White (1976) suggested that B vitamins are the extant remnants of an ancient metabolism predating proteins, based on reactions catalysed by variant nucleic acids. At a later stage, and following a general acceptance of the RNA-world theory, it was proposed that the first ancient peptides would have bound to the catalytic nucleic acid to provide greater specificity to the reactions (Szathmáry, 1999). During evolution proteins would have substituted most of the catalytic nucleic acids, but many proteins would have retained B vitamins as a cofactor in the active site

of what now constitute modern enzymes. The key role of cofactors across living kingdoms is illustrated by the fact that more than 30 % of all enzymes found in the ExPASy database have been found to require a cofactor and 10 % require a B vitamin as a cofactor (Monteverde *et al.,* 2017; Mukhopadhyay *et al.,* 2019).

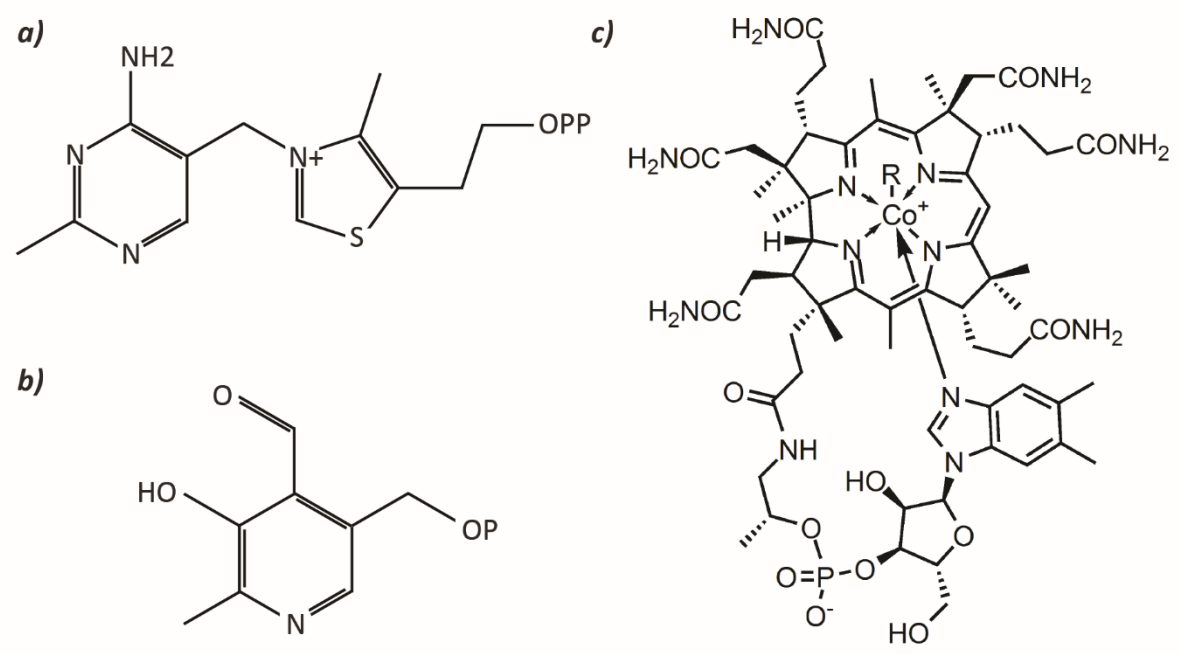


Figure 1.1. Molecular structure of TPP (a), pyridoxal (b) and cobalamin (c). *The R in the cobalamin molecule denotes the upper axial ligand position.*

In this thesis, thiamine (vitamin B1) biosynthesis regulation has been the main object of study, and it has become apparent that the metabolic relationship with cobalamin (vitamin B_{12}) influences the expression of thiamine-related genes. Therefore, to illustrate the biochemical uniqueness of B vitamins, their relation to nucleotide biosynthesis and the interconnection of their metabolic pathways, closer detail on the biochemistry, biosynthesis and transport of the two vitamins most studied in this thesis, thiamine and cobalamin, will be provided in the following subsections.

1.1.1 Thiamine

Biochemistry

Thiamine is formed by a pyrimidine and a thiazole moiety connected by a methylene bridge from position 5’ at the pyrimidine to the nitrogen in the thiazole ring, which as a consequence has a positive charge (Figure 1.1a). The pyrimidine ring has a methyl group at position 2’ and an amino group at position 4’. The thiazole has a methyl group at position 4 and the active cofactor is

pyrophosphorylated on the 2-hydroxyethyl group at position 5. In the mechanism of action of thiamine pyrophosphate (TPP) dependent enzymes, a tautomerization of the pyrimidine ring deprotonates the amino group at position 4, the resulting imino group then deprotonates the carbon at the thiazole position 2, which subsequently attacks the α-carbon in an α-ketoacid substrate. The bound substrate is decarboxylated to give the aldehyde product one carbon shorter. The sulphur atom of the thiazole ring plays a role in stabilizing some intermediate conformations of the molecule during decarboxylation (Malandrinos *et al.,* 2006). There have been previous attempts to simplify the molecule in the past, with only the demethylated pyrimidine being functional, but this is effective only in bacteria. Plants and fungi do not tolerate any reduction to the TPP molecule (Hanson *et al.,* 2018).

Its unique biochemistry makes TPP an essential cofactor in enzymes that participate in the most important central metabolism pathways. TPP-dependent enzymes include transketolase, involved in the pentose phosphate pathway and the Calvin cycle, pyruvate dehydrogenase, linking glycolysis with the tricarboxylic acid cycle (TCA cycle), oxoglutarate dehydrogenase, involved in the TCA cycle, pyruvate decarboxylase, involved in ethanol fermentation, and enzymes in both synthesis and degradation pathways of branched amino acids (Hanson *et al.,* 2018). All organisms are known to require thiamine for their survival, with the recently discovered exception of *Borrelia burgdorferi,* an obligate pathogen that causes lyme disease, which does not encode any thiamine-dependent enzymes and has a severely reduced carbon metabolism (Zhang *et al.,* 2016).

Biosynthesis

Most organisms, including plants and many fungi, bacteria and algae biosynthesise thiamine, while organisms such as animals and certain microorganisms require an exogenous supply of thiamine or its precursors. Thiamine is biosynthesised from two key intermediates, 4-amino-5-hydroxymethyl-2-methylpyrimidine (HMP) and 4-methyl-5-(2-hydroxyethyl)thiazole (HET). HMP is synthesised from 5-aminoimidazole ribonucleotide (AIR) in a SAM-dependent reaction by ThiC in bacteria and plants, whereas in fungi it is synthesised from pyridoxal-5-phosphate (PLP; vitamin B_6) in a suicide reaction that uses an internal histidine as substrate by THI5/NMT1 (Coquille *et al.,* 2012; Lai *et al.,* 2012). HET is synthesised from nicotinamide adenine dinucleotide (NAD; vitamin B_3), glycine and an internal cysteine residue by the suicide enzyme THI1/THI4 in plants, green algae, Cryptophytes and fungi (Chatterjee *et al.*, 2011; McRose *et al.,* 2014). Bacteria, Stramenopiles and Haptophytes use ThiS, ThiF, ThiG, ThiI, ThiO, DXS, and ThiG to synthesise HET from glycine, pyruvate, glyceraldehide-3-phosphate (GA3P) and cysteine (Figure 2; Jurgenson *et al.,* 2009; McRose *et al.,* 2014).

The thiamine biosynthesis pathway contains more than one suicide enzyme, and other enzymes such as *THIC* have a very low turn-over rate, leading to the prediction that thiamine metabolism can cost between 2 and 10 % of the maintenance energy in plants (Hanson *et al.,* 2018). Perhaps because it is a metabolically burdensome pathway, most organisms have evolved negative feedback loop regulations on the first enzymes of the pathway (more detail in the next subsection 1.1.2). The high metabolic cost of some of the earlier biosynthetic reactions might also explain why many microbes, including some algae, have a partial thiamine biosynthetic pathway and require an exogenous supply of one or both of HMP and HET for the biosynthesis of thiamine (Gutowska *et al.*, 2017). The loss of thiamine biosynthetic enzymes and the concomitant auxotrophy for thiamine or its precursors underpins the establishment and maintenance of important ecological relations in aquatic environments (Kazamia *et al.,* 2016).

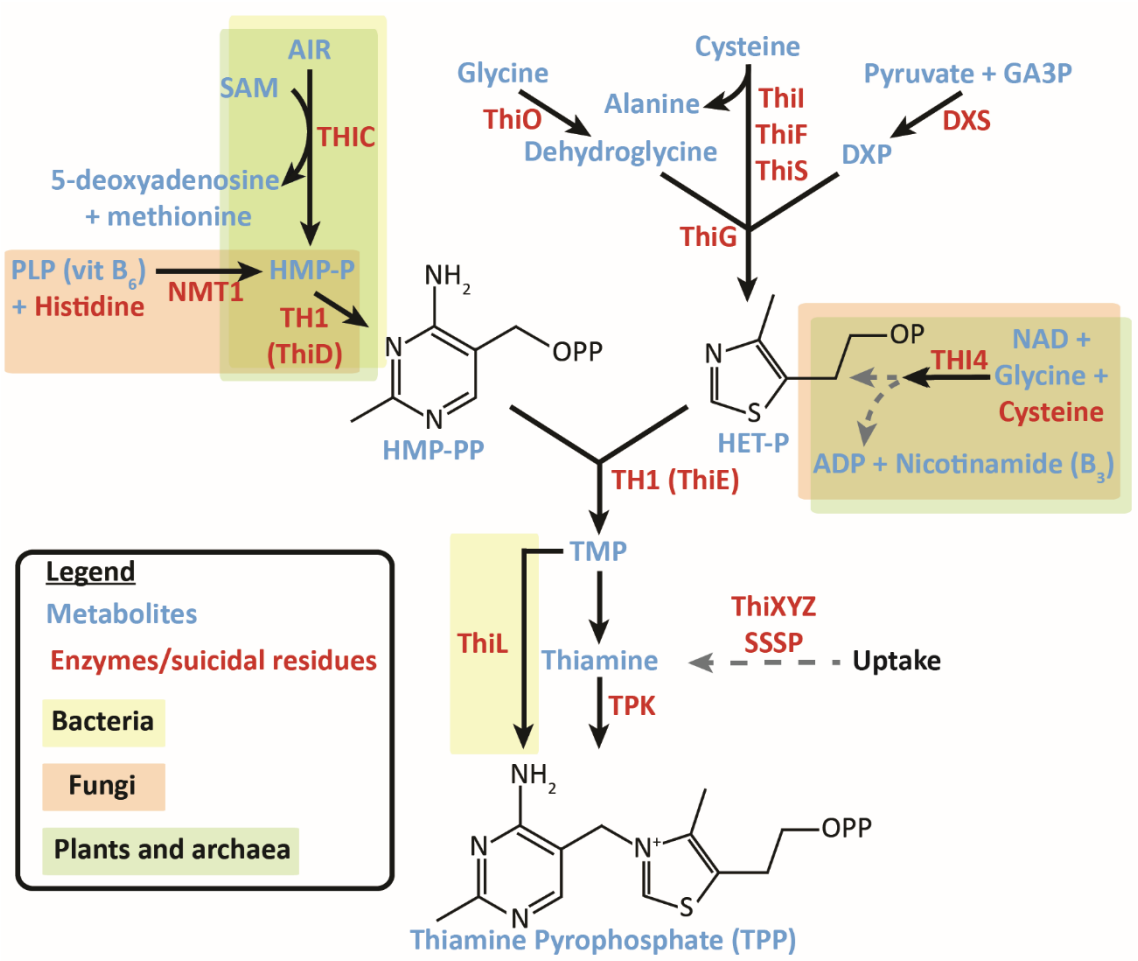


Figure 1.2. Thiamine biosynthesis pathway in different organism groups. *Bacteria, fungi, plants and archaea (in different colour shades) use different enzyme combinations to synthesise HMP-PP and HET-P. HMP-PP and HET-P are metabolised to thiamine by a common enzymatic step in all organisms (ThiE).*

Transport

Uptake of thiamine and its precursors is essential for auxotrophs, but it is also metabolically rewarding for prototrophs given the burden of *de novo* synthesis. Hence, all organisms have one of two known transport systems for thiamine and its precursors: an ATP-Binding Cassette (ABC) active transport system or a solute symport system. In bacteria, ABC transporters for thiamine, thiamine monophosphate (TMP), TPP, HMP and HET have been identified (Jurgenson *et al.,* 2009). In particular, the *thiBPQ* operon in *Salmonella typhimurium* has been determined to encode three proteins: the periplasmic component (ThiB), a transmembrane channel (ThiP), and an ATPase motor (ThiQ) of a Type I ABC transporter specific for thiamine, TMP and TPP (Webb *et al.,* 1998). ThiXYZ, another ABC transport system, is responsible for the uptake of the pyrimidine precursor N-formyl-4-amino-5-(aminomethyl)-2-methylpyrimidine (FAMP), and its expression is controlled by a TPP riboswitch in different bacterial species (Jaehme & Slotboom, 2015). A riboswitch is a sequence in mRNA that folds in a tertiary structure able to specifically bind a ligand and mediate a change in genetic expression in response (see section 1.2 for details). The substrate binding component of the system, ThiY, shares structural homology with THI5/NMT1 (Bale *et al.,* 2010). ABC transporters of the Energy-Coupling Factor (ECF) type have also been identified in *Lactococcus lactis,* where ThiF is the thiamine-specific component of the transport system and it is controlled by a TPP riboswitch. ThiUVWX is another ECF-type thiamine and HMP transporter in *B. subtilis*, and the operon in which it is encoded is also controlled by a TPP riboswitch (Devedjiev *et al.,* 2004; Schyns *et al.,* 2005). ThiW has been described as a thiazole-specific component of an ECF-type ABC transporter in multiple bacteria (Slotboom, 2014).

Thiamine and its precursors can also be imported by solute carrier symporters which are either members of the Amino acid-Polyamine-organoCation (APC) or the Major Facilitator (MFS) protein superfamilies. In the APC superfamily, *Methylobacillus flagellatus* ThiV, is member of the Sodium-solute Symporter Family (SSF), where a sodium cation is co-transported with the solute. ThiV is predicted to transport thiamine or its precursors, is controlled by a TPP riboswitch and is homologous to a pantothenate (vitamin B_5) transporter. In HMP-auxotrophic bacteria, CytX, in the Nucleobase Cation Symporter 1 family, has been hypothesised to transport HMP. In the MFS superfamily, ThiU and ThiT are hypothesised to be thiamine and HET transporters respectively (Jaehme & Slotboom, 2015). NiaP in the MFS family, which belongs to an homology group most of whose members specifically transport niacin (vitamin B_3), transports thiamine but not pyrimidine or thiazole precursors, and it is encoded in an operon with a predicted TPP riboswitch

(Jeanguenin *et al.,* 2012). The mechanism of action of the MFS transporters is not known, but it is possible that it involves proton or sodium symport.

In eukaryotes, most experimentally confirmed thiamine transporters rely on a chemiosmotic gradient to drive the uptake of the vitamin. In fungi, S*accharomyces cerevisiae* THI10 acts as a high-affinity transporter belonging to the MFS Family, and THI72 and NRT1 are two additional lower-affinity transporters (Donovan *et al.,* 2018; Enjo *et al.,* 1997). *Schizosaccharomyces pombe* THI9 has been shown to transport thiamine via a proton symport mechanism, contains a polyamine transporter domain and belongs to the Amino Acid/Choline Transporter family within the APC superfamily (Vogl *et al.,* 2008). *Candida parapsilosis* Dur31 has been shown to transport thiamine, contains an SSF domain and belongs to the solute carrier 5 transporter family within the APC superfamily. Dur31 and its homologue in *Neurospora crassa* are regulated by TPP riboswitches and Dur31 homologues in *S. cerevisiae* act as polyamine transporters (Donovan *et al.,* 2018).

In animals, *Homo sapiens* SLC19A2 and SLC19A3 are thiamine transporters that use a pH gradient to energise the transport via a proposed proton antiport system. They share high sequence conservation with HsSLC19A1, a folate (Vitamin B_9) transporter, and together belong to the Reduced Folate Carrier in the MFS superfamily (Ganapathy *et al.,* 2004). In plants, PUT3 acts as a thiamine and polyamine transporter in *Arabidopsis thaliana,* and it has been suggested to act as a proton symporter. PUT3 is a member of the amino acid permease family within the APC superfamily, like SpTHI9 in fungi. PUT3 is located in the phloem and it is thought to mediate directional transport from photosynthetic tissues where thiamine is mostly synthesised to sinks like roots (Martinis *et al.,* 2016).

Finally, thiamine transporters have been annotated, but not experimentally characterised, in all major algal groups. All algal lineages were found to have a predicted SSF family thiamine transporter under the annotation of SSSF, SSSP and SSSQ all of which contain a predicted TPP riboswitch in the untranslated regions of the genes (McRose *et al.,* 2014). SSSF, SSSP and SSSQ are orthologues and conserve homology to fungal transporters including *N. crassa* Dur31 and bacterial transporters. Prasinophytes also have a gene termed *ATS1* in the folate receptor family within the frizzled cysteine-rich domain-related superfamily, members of which family mediate folate transport in humans (Wibowo *et al.,* 2013). All transporter genes in *Micromonas* also showed the presence of predicted TPP riboswitches in the gene untranslated regions (UTRs) (McRose *et al.,* 2014).

1.1.2 Thiamine biosynthesis regulation

Numerous experimental reports have uncovered the widespread expression control of thiamine biosynthesis and thiamine transporter genes in all domains of life. The metabolic cost of *de novo* synthesis, and the low intracellular and environmental levels of the vitamin have contributed to a diversity of strategies that have evolved to regulate thiamine biosynthesis in a highly sensitive and tight manner in all kingdoms of life. In bacteria, thiamine biosynthesis and transport genes are generally co-located in operons and are regulated by TPP riboswitches in 5' untranslated regions (Winkler *et al.,* 2002). In archaea, riboswitches have been found in *ThiT1* and *ThiT2* in three species of the *Thermoplasma* genus*,* where they are found to directly sequester the Shine-Dalgarno sequence in the 5'UTR (Rodionov *et al.,* 2002).

In eukaryotes, TPP riboswitches can be found within introns in 5'UTRs, coding sequences and/or in 3' UTRs of thiamine-related genes (Figure 1.5). In plants, a TPP riboswitch has been found in the 3'UTR of *A. thaliana THIC* (Bocobza *et al.,* 2007; Wachter *et al.,* 2007). In green algae, riboswitches have been experimentally characterised in the 5'UTR of *Chlamydomonas reinhardtii THI4* and in the 6th intron of *THIC* (Croft *et al.,* 2007). In fungi, riboswitches were first described in three *N. crassa* genes, *NMT1, THI4* and *Dur31* (Cheah *et al.,* 2007). *NcDur31* is an homologue to *CpDur31* which has been described to have a TPP riboswitch in the first intron of the coding sequence (CDS) (Donovan *et al.,* 2018). The molecular mechanisms of these TPP riboswitches is explained in full in section 1.2.

Croft *et al.* (2007) predicted the presence of a TPP aptamer in the 3'UTR of diatoms *Phaeodactylum tricornutum* and *Thalassiosira pseudonana THIC*. More recently, McRose *et al.* (2014) took a bioinformatic approach based on sequence similarity and secondary structure conservation to predict the presence of TPP aptamers in eukaryotic photosynthetic lineages including Alveolata, Stramenopiles, Rhizaria, Haptophytes, Cryptophytes and Prasinophytes (see section 1.3 for algae taxonomy). They did not find any TPP aptamers in Rhodophytes and Glaucophytes, and to the best of my knowledge, TPP riboswitches have not been described in animals, Excavates and Amoebozoa either. The TPP aptamers were mostly found associated with *THIC* and *THI4*, the first enzymes in the thiamine biosynthesis pathway as well as to *SSSF, SSSP, SSSQ* and *ATS1* predicted transporters and exceptionally in *Guillardia theta THIM*. Close to half of the predicted TPP aptamers in algal lineages do not appear to regulate alternative splicing as they are found in single-exon genes. In the same study, using the Mamelliales *Micromonas sp. CCMP1545, SSSP* and *SSSF* that had riboswitches at both 5' and 3' UTRs, and *Micromonas sp. RCC299 ATS1* and *UNK1* that had riboswitches in the 3'UTR were shown to be downregulated by high thiamine levels. These are single exon genes that are not expected to involve alternative

splicing in their mechanism, although experimental confirmation at the molecular level is still lacking.

In *S. cerevisiae* no TPP riboswitches have been found to date. Instead, thiamine-related genes are tightly regulated by thiamine-specific transcription factors by which a thiamine sensor protein, Thi3p, binds transcription factors Thi2p and Pdc2p under low thiamine levels to enhance expression of thiamine-related gene promoters with a specific binding site for the transcription factors. Under increased levels of thiamine, TPP binds to Thi3p, which in consequence prevents it from binding the transcription factors (Maupin-Furlow, 2018). Pdc2p has also been described to regulate the expression of pyruvate dehydrogenase, linking the regulation of thiamine metabolism with the regulation of thiamine-dependent carbon metabolism (Nosaka, 2006). Thi3p and Thi2p have also been shown to be necessary to mediate downregulation of their own transcripts under thiamine supplementation (Nosaka *et al.,* 2005).

The vast majority of studied archaeal genomes have been shown to have a regulatory system based on a transcription factor similar to yeast. ThiR is the transcription factor with a DNA binding domain and a TPP binding domain orthologue to thiamine-phosphate synthase integrated in a single protein. ThiR is conserved throughout *Euryarchaeota* and *Crenarchaeota* and, although the DNA binding sites are different between different orders, they are consistently found in the promoter regions of the thiamine-related operons. Similar to *S. cerevisiae*, ThiR also contains its own binding sites in its promoter region in some species. The mechanism of action significantly differs from *S. cerevisiae* however, and in archaea ThiR seems to bind the promoter region of thiamine-related genes as a repressor under high-levels of thiamine; when thiamine levels decrease, the DNA binding affinity is reduced and the repression lifted to allow gene expression (Rodionov *et al.,* 2017; Hwang *et al.,* 2017). With systems in *Saccharomyces* and Archaea, we can conclude that there has been more than one independent example in evolution where thiamine regulation has bypassed the general RNA-based mechanism and has adopted a mechanism based on protein sensors and trans-acting transcription factors. On one hand, the transcription factor-based regulatory mechanism allows the regulation of gene expression at a transcriptional level instead of the post-transcriptional nature of riboswitch regulation, reducing the metabolic cost of transcription. On the other hand, the post-transcriptional regulation by riboswitches can respond faster to changes in environmental or physiological concentrations of the vitamin.

1.1.3 Cobalamin

Biosynthesis

Cobalamin is the vitamin with the most complex molecular structure. It comprises a tetrapyrrole-derived corrin ring connected with a dimethylbenzimidazole (DMB) nucleotide through a d-1-amino-2-propanol linker esterified to the nucleotide phosphate and with an amide link to the D pyrrole of the corrin ring. The four nitrogen atoms in the corrin ring together with a nitrogen from DMB coordinate a cobalt cation. Cobalt is also covalently bound to a variable upper axial ligand, which in the active forms of the vitamin in eukaryotes is a methyl or adenosyl group, forming methylcobalamin and adenosylcobalamin respectively (Figure 1.1c; Jägerstad & Arkbåge 2003). Pseudocobalamin, which has adenine instead of DMB as the lower axial ligand, is a variant synthesised by many bacteria, including cyanobacteria (Helliwell *et al.,* 2016).

Cobalamin biosynthesis represents one of the most complex biosynthetic pathways known in Nature, comprising around 30 different enzymatic steps, and is only present in prokaryotes (Raux *et al.,* 2000; Figure 1.3). There are three main parts in the cobalamin biosynthetic pathway, the first one is the synthesis of the corrin ring, the second is the synthesis of DMB nucleotide, and the third one is the condensation of the two moieties. DMB can be produced from reduced flavin mononucleotide (vitamin B_2) by BluB in *Sinorhizobium melliloti* in an oxygen-dependent manner (Taga *et al.,* 2007). DMB has also been described to be synthesised from AIR in an oxygen-independent manner by BzaABCDE in *Eubacterium limosnum* (Hazra *et al.,* 2015). AIR is the common substrate that leads to the biosynthesis of purines, thiamine and DMB. BzaA, BzaB and BzaF, a fusion of BzaA and BzaB, are homologous to THIC, the enzyme that catalyses the synthesis of the thiamine pyrimidine moiety from AIR (Figure 1.2; Mehta *et al.,* 2015).

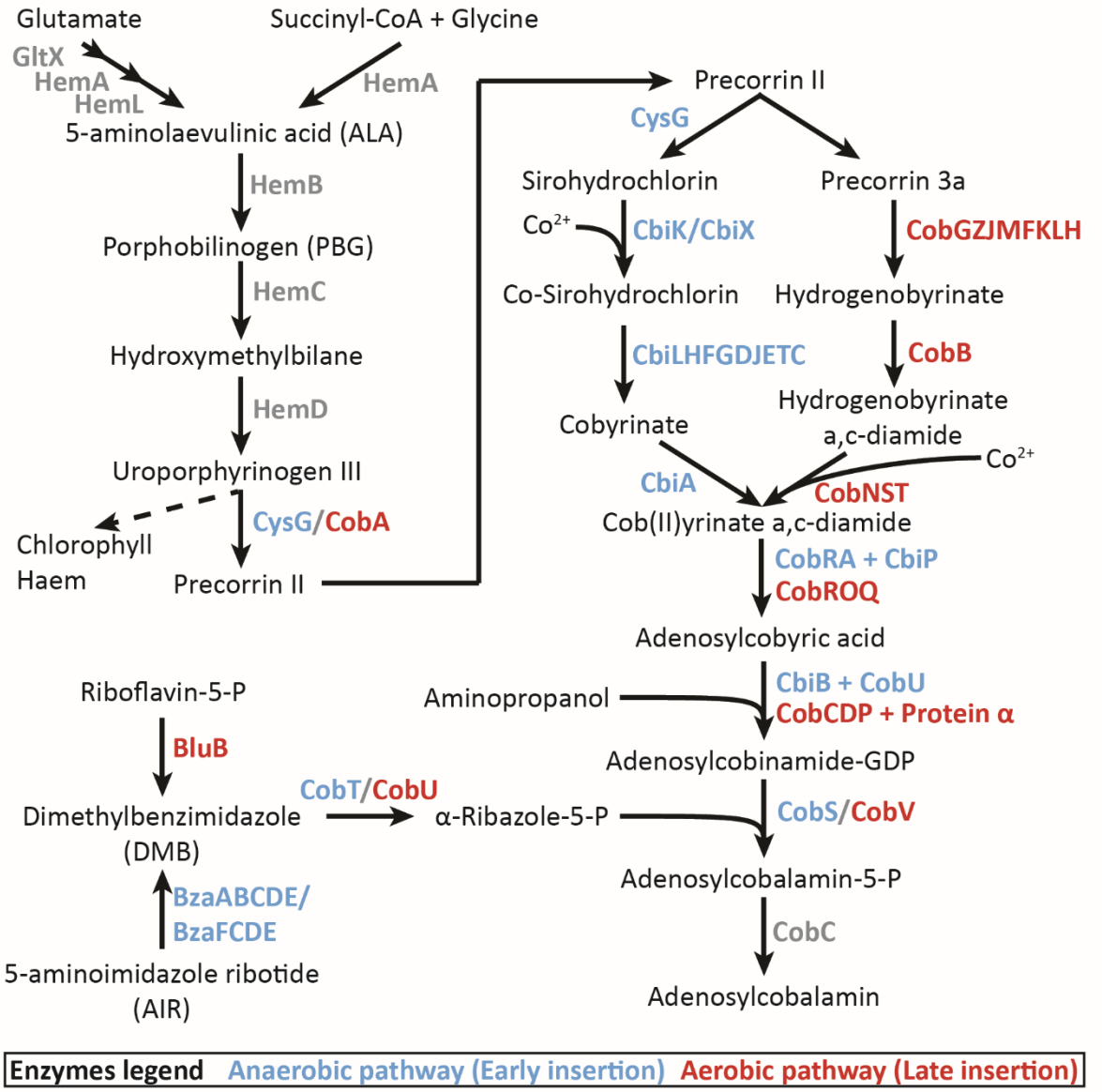


Figure 1.3. Cobalamin biosynthesis pathway. *Prokaryotes use either the anaerobic (blue) or aerobic (red) pathway to produce cobalamin de-novo. The first phase of the pathway is common and produces the tetrapyrrole precursor precorrin II, the second phase of the pathway is differentiated by whether the cobalt cation is inserted earlier (anaerobic pathway) or later (aerobic pathway) in the corrin ring to produce adenosylcobyric acid. This precursor is fused with DMB nucleotide as lower axial ligand via an aminopropanol linker to produce adenosylcobalamin. DMB can be produced from riboflavin (Vitamin B_2) by BluB, or from AIR by BzaABCDE in anaerobic conditions.*

The corrin ring is derived from the universal tetrapyrrole precursor, 5-aminolaevulinic acid (ALA), which is synthesised either from succinyl-CoA and glycine by ALA synthase, a reaction dependent on pyridoxal (vitamin B_6), or from glutamic acid sequentially by a dehydrogenase and an aminotransferase which also requires pyridoxal. Two ALA molecules are then condensed into porphobilinogen (PBG), the building block of a tetrapyrrole, by ALA dehydratase. PBG deaminase condenses four PBG molecules into a linear tetrapyrrole releasing four ammonia molecules. Uropor-phyrinogen III synthase then facilitates the cyclisation of the molecule, with inversion of ring D, to form the cyclic tetrapyrrole uropor-phyrinogen III, the intermediate from which chlorophyll, haem and cobalamin anabolism branch out. For cobalamin biosynthesis, the SAM-dependent uropor-phyrinogen III methylase catalyses the next step of the pathway.

From here onwards, there are two variants of the corrin synthesis pathway, one requires molecular oxygen as a substrate and the other does not. Both pathways have in common the sequential SAM-dependent methylation of the tetrapyrrole ring but differ in the step of addition of the cobalt cation, which is the first step of the oxygen-independent pathway, and almost the last step for the oxygen dependent pathway. The contraction of the pre-corrin ring between the A and D pyrroles is poorly understood in the oxygen-independent pathway, in the oxygen-dependent pathway it is catalysed by the action of an oxygen-dependent monooxygenase (CobG) and a subsequent methylase (CobJ). The amidation of the A and B pyrrole rings is not understood for the oxygen-independent pathway and is catalysed by CobB in the oxygen-dependent pathway just before the incorporation of the cobalt cation.

At this stage, the oxygen dependent and independent pathways rejoin, CobQ amidates all remaining carboxylic groups except for the one in the D pyrrole that is subsequently bound with d-1-amino-2-propanol with an amide bond in the first reaction of the pathway phase to link the corrin ring with the DMB nucleotide ligand. Finally, CobP catalyses the phosphoesterification of the hydroxyl group in the aminopropanol bridge with a guanine nucleotide monophosphate, and CobV replaces the guanine nucleobase for DMB. In general, the genes involved in corrin and DMB biosynthesis are clustered in 3-5 operons in bacterial genomes and are frequently controlled by cobalamin riboswitches (see section 1.2) that regulate translation initiation (Hazra *et al.,* 2015; Ravnum & Andersson, 2001).

Biochemistry

It has been suggested that the unique biochemistry of cobalamin was originally used to catalyse the creation of an oxidisable aldehyde via the deamination or dehydration of small substrates in fermentation processes (Roth *et al.,* 1996). The reactions would be of key importance for the

metabolism of fermenting microorganisms given that the aldehyde product could be further oxidised to obtain ATP or reduced to an alcohol that could be then excreted to regenerate reducing power carriers. In methanogens, cobalamin has an analogous physiological function and participates in the regeneration of reducing power by carrying a methyl group from acetate through four cobalamin-dependent methyl transferases until the release of molecular methane. The methyl group is covalently bound to the cobalt atom which when reduced to Co(I) is a super-nucleophile, thus giving unique biochemical properties to cobalamin (Banerjee & Ragsdale, 2003). A similar reaction is widely catalysed by cobalamin as a cofactor in many methyltransferases. Most significantly, many eukaryotes require cobalamin as a cofactor for methionine synthase (METH), a key step in the C1 cycle. Cobalamin-dependent methionine synthase catalyses the transfer of a methyl group from 5-methyltetrahydrofolate (vitamin B_9) to the homocysteine thiol group to produce methionine. Following the C1 cycle, methionine is S-adenosylated to form SAM, which is used as a methyl donor in a wide variety of methylation reactions. Important SAM-dependent reactions include the methylation of DNA and histones, as well as the synthesis of HMP from AIR catalysed by THIC in the first reaction of the thiamine biosynthesis pathway. As a result of the reaction, SAM is demethylated to S-adenosylhomocysteine which is further hydrolysed to homocysteine to replenish the C1 cycle. Finally, cobalamin also catalyses reductase dehalogenation reactions where it breaks the halogen-carbon bond, a reaction that is central to the bioremediation of sites polluted with toxic and recalcitrant halogenated compounds (Guo & Chen, 2018).

Plants do not require cobalamin as they contain a cobalamin-independent version of methionine synthase (METE) and lack its cobalamin-dependent form (METH). More than half of the 306 algal strains investigated in a survey of the literature were shown to require cobalamin (Croft *et al.* 2006), and in general cobalamin auxotrophy can be explained by a lack of METE in the genome. Cobalamin-independent algae generally have both METE and METH in their genomes. The physiological advantages of the METE cobalamin independence come with the downside that the enzyme is less efficient than METH, in *S. cerevisiae* it has been shown that MET6p (METE homolog) has a kinetic efficiency (Kcat/Km[Homocysteine]) of $2.57 \cdot 10^{-8}$ $s^{-1} \cdot M^{-1}$ (Suliman *et al.,* 2005) compared to $2.33 \cdot 10^{-5}$ $s^{-1} \cdot M^{-1}$ for *E.coli* METH (Banerjee *et al.,* 1990). Therefore, algal species with both methionine synthases have to optimally balance the expression of both genes, and it has been found that in *C. reinhardtii* and *P. tricornutum METE* is significantly downregulated in response to cobalamin supplementation (Bertrand *et al.,* 2012; Croft *et al.,* 2005; Helliwell *et al.,* 2016). Most algal species cannot use pseudocobalamin as a cofactor, but it has been shown that some algae auxotrophic for cobalamin have the capacity to refactor

pseudocobalamin into cobalamin and can thus grow with a supply of pseudocobalamin and DMB (Helliwell *et al.,* 2016).

Transport

The unique biochemical features of cobalamin, its costly and complex biosynthetic pathway only found in some prokaryotes, and its low abundance in the environment provide a strong incentive to develop highly efficient and specific uptake systems in organisms with cobalamin-dependent enzymes, even in those not strictly auxotrophs. The *E. coli* cobalamin transport system has been widely studied and structural and mechanistic studies have described the process in great detail. The first component of the transport system is BtuB, a TonB-dependent transporter in the outer membrane that actively transports cobalamin from the media to the periplasm (Chimento *et al.,* 2003). A three-component ABC active transport system, BtuCDF, uses ATP hydrolysis as motive force to translocate cobalamin from the periplasm to the cytosol. BtuF is the periplasmic component of the system that binds cobalamin with high affinity and directs the molecule to the BtuCD complex in the cell membrane. BtuC and BtuD form an heterotetramer where BtuC is the membrane spanning domain through which cobalamin goes through the membrane, and BtuD is located at the cytosolic side and is responsible to power and orchestrate cobalamin transport through a cycle of ATP binding, hydrolysis and release (Locher & Borth, 2004).

In mammals, free cobalamin in the small intestine is bound by Intrinsic Factor, which facilitates the endocytosis into intestinal cells mediated by Cubilin/AMN receptor. Intrinsic Factor is then degraded in the intestinal cells lysosomes and cobalamin is released into the blood stream associated with Transcobalamin. Cells uptake Transcobalamin from the blood stream through endocytosis mediated by the Transcobalamin Receptor. Transcobalamin is then degraded in the lysosome and cobalamin is transferred to the cytoplasm by ABCD4 which is proposed to be an ABC transport system similar to bacteria (Okamoto *et al.,* 2018; Xu *et al.,* 2019). Intrinsic Factor and Transcobalamin are proteins with high sensitivity for cobalamin with a DMB lower ligand and have a sub-picomolar affinity constant (K_d). It is hypothesised that these two proteins serve as a sieve to guarantee that only bioavailable cobalamin molecules are delivered to cells and they act as chaperones to prevent the highly-reactive cofactor from degradation or interaction with other molecules during its transport (Gherasim *et al.,* 2013).

In diatoms, Bertrand *et al.* (2012) identified a BtuF orthologue in *P. tricornutum* and *T. pseudonana* that was highly upregulated under cobalamin-depleted conditions, which they named Cobalamin Acquisition protein (CBA1). The protein was fluorescently tagged and

overexpressed in *P. tricornutum* to determine its location. It was found in the plasma membrane and the ER, and an increase in intracellular cobalamin was detected in the overexpressor mutant.

1.2. Riboswitches

Riboswitches are "structured noncoding RNA domains that selectively bind metabolites and control gene expression" (Breaker, 2012). The first endogenous riboswitch to be discovered inhibited *E. coli btuB* translation initiation upon cobalamin binding (Nahvi *et al.*, 2002), followed swiftly by the thiC and thiM riboswitches regulated by TPP (Winkler *et al.,* 2002). To date, up to 40 riboswitch structural classes binding to 24 different ligands, including vitamins, nucleotide derivatives, amino acids and other small molecules, have been molecularly characterised (McCown *et al.,* 2017) and new riboswitch classes continue to be discovered (Yu & Breaker, 2020). To conduct their function, riboswitches are composed of two functional modules: the aptamer and the expression platform. In short, the aptamer folds into a tertiary structure that upon binding of a specific ligand undergoes a conformational change. The expression platform translates the aptamer conformational change into regulation of genetic expression (Roth & Breaker, 2009). The following two subsections provide greater detail on the function and role of aptamers and expression platforms. The next subsection focuses on the aspects of the eukaryotic TPP riboswitch of special interest for this thesis concluding with a last subsection on the shared evolutionary origin proposed for vitamins and riboswitches.

1.2.1 Aptamers

Aptamers are short RNA sequences that have evolved the capacity to fold into complex structures formed by stems, loops and pseudo-knots. These RNA structures bind with remarkable sensitivity and specificity to their target ligands and are flexible enough to accommodate a dynamic equilibrium between different folding conformations. Aptamers are remarkably sensitive, with K_d values generally at the nanomolar range, and form very specific binding pockets to their ligands utilising the unique substituent groups in the RNA molecule to select their target ligand through electrostatic, steric and hydrogen bond interactions (Paselis & Serganov, 2014). In most cases metal ions are necessary to facilitate folding or mediate the binding of aptamers to their ligands (Serganov & Patel, 2012). Aptamers are folded co-transcriptionally, so the order in which the stems are transcribed is relevant to guide the folding. Some studies indicate that the folding of a riboswitch is influenced by the binding of a ligand co-transcriptionally (Gong *et al.,* 2017). The binding of a ligand can decrease the energy barrier to adopt a given conformation as well as stabilising a specific fold (Haller *et al.,* 2013).

Within their tight specificity, aptamers can allow binding flexibility to recognise derivatives and analogues of their canonical ligands. Binding different derivatives can be functionally relevant for riboswitches that sense a group of specific ligands (Roth & Breaker, 2009). It is the case of the TPP aptamer, which has been observed to specifically bind both TPP and HET-P in the *CrTHI4* riboswitch and responds to both TPP and its product HMP-PP in the *CrTHIC* riboswitch (Moulin *et al.,* 2013). This flexibility in the ligand specificity has also been proposed to have partly contributed to the emergence of new riboswitches within the same riboswitch family (Roth & Breaker, 2009). And aptamer flexibility has also been exploited to target riboswitches with unnatural analogues of their canonical ligands that interfere with the riboswitch function raising hopes of new anti-microbial treatment strategies (Blount & Breaker, 2006).

1.2.2 Expression platform

Sensitive and specific ligand binding is necessary but not sufficient for riboswitch function. Functional riboswitches also need to be able to change conformation upon ligand binding and modulate genetic expression in response to this conformational change. To create a functional riboswitch, aptamers are associated, usually with overlapping sequences, with expression platforms which mediate changes in gene expression. In general, riboswitches are positioned in untranslated regions of transcripts where they can interact with processes that guide gene expression from transcription through splicing and translation. Aptamer sequences are generally well conserved across a wide diversity of organisms since their function only involves binding of the ligand without any additional factors. In contrast, expression platforms need to interact with the genetic expression machinery of each species and therefore their sequences and modes of action are more variable (Roth & Breaker, 2009).

The processes necessary for genetic expression differ between prokaryotes and eukaryotes and as a result expression platforms employ completely different mechanisms in the different kingdoms of life. In bacteria, riboswitches are usually found in 5'UTRs where conformational changes lead to changes in ribosome accessibility to the ribosome binding site through the formation or disassociation of sequester or anti-sequester hairpins. It is also common that riboswitches regulate premature transcription termination through the formation and disassociation of terminator and anti-terminator stems. Other less common expression platforms involve self-cleavage or exposure to RNase degradation (Roth & Breaker, 2009). In archaea, riboswitches are found in 5'UTRs and they are predicted to directly sequester the ribosome binding site in their secondary structure upon thiamine recognition (Rodionov *et al.,* 2002). In eukaryotes, only TPP riboswitches have been found to date, so their mechanism of action is explained in the next subsection.

1.2.3 The eukaryotic TPP riboswitch

The TPP riboswitch is the most common of all riboswitches, and it is the only riboswitch class found in bacteria, archaea, and eukaryotes. The TPP aptamer sequence and structure are well conserved across the different kingdoms and X-ray crystallographic studies have demonstrated that it is formed by three main double-helixes radiating from a central junction (Edwards & Ferré-D'Amaré, 2006; Thore *et al.,* 2006). The first double-helix to be transcribed is the P2/3 which includes stems P2 and P3 and the J2/3 junction with some species also having a P3' stem branching from the J2/3 junction. The P4/5 double-helix is next transcribed and is formed by stems P4 and P5 and the J4/5 junction. Finally, the 3' end of the aptamer base-pairs with the first transcribed 5' end of the aptamer to form the third double-helix or P1 stem (Figure 1.4). The P3 stem is exceptionally variable in length and sequence beyond the first few bases and it is not required for TPP binding or intramolecular interactions, so it is proposed to function as a thermodynamic anchor to nucleate the folding of the rest of the P2/3 helix (Anthony *et al.,* 2012; Uhm *et al.,* 2018). The P1 and P2 stems have the lowest base-pairing energy and thus are the most dynamic domains of the riboswitch and overlap aptamer and expression platform in bacterial systems (Anthony *et al.,* 2012; Sudarsan *et al.,* 2005). The P4/5 stem has the second highest folding energy, but its sequence and length are well conserved, probably because the L5 loop at the end of P4/5 is responsible for interactions with the P2/3 helix in the closed conformation (Thore *et al.,* 2006; Anthony *et al.,* 2012).

The TPP molecule has contacts with both the P2/3 and P4/5 branches, the universally conserved CUGAGA motif in the J2/3 junction interacts with the pyrimidine substituents through hydrogen bonds, and the GCG sequence in the J4/5 chelates a Mg^{2+} cation that electrostatically interacts with the negative charges in the pyrophosphate (Figure 1.4). The bound TPP molecule brings together the two "hands" of the riboswitch, helixes P2/3 and P4/5, which close around the ligand through direct interactions between the L5 loop at the end of P4/5 and residues in the first base-pairs of the P3 stem. When P2/3 and P4/5 come together there is a conformational change involving residues in the "hinge" J2/4 junction that facilitates and stabilises the base-pairing of the two strands of the P1 stem (Anthony *et al.,* 2012; Thore *et al.,* 2006).

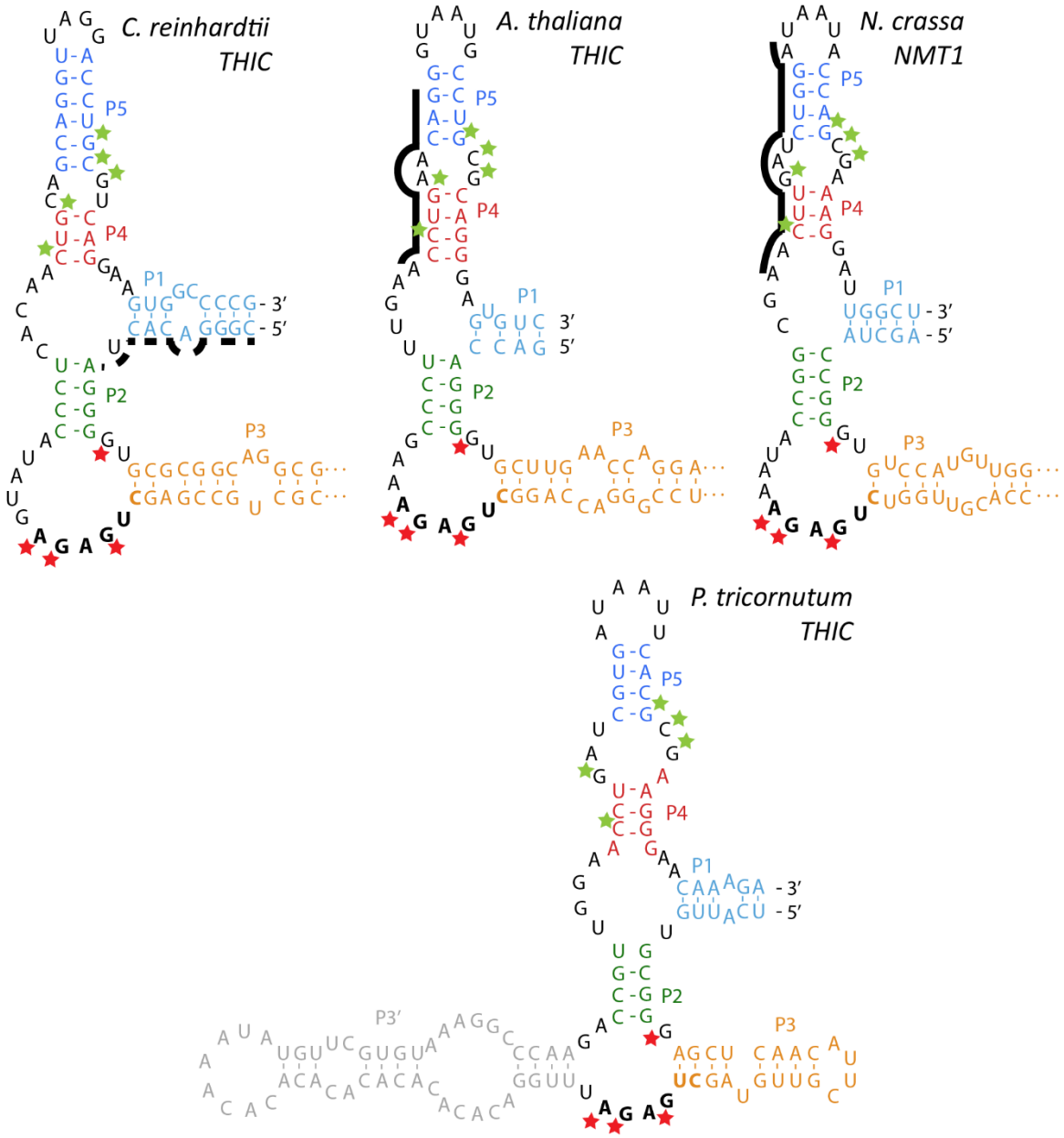


Figure 1.4. Secondary structure of eukaryotic TPP riboswitches. *The universally conserved CUGAGA motif is highlighted in bold. The conserved residues with direct interaction with the pyrimidine moiety are marked with a red star, the residues that interact with the pyrophosphate are marked with a green star. The sequence that has been shown to form a pseudo-knot blocking an alternative splice site are marked with a solid black line if experimentally validated and with a dotted line for the sequence proposed to base-pair with a splicing branching site in* CrTHIC.

The thiazole ring is not specifically bound by the aptamer, which allows the flexible recognition of pyrithiamine pyrophosphate, a TPP agonist, to activate the riboswitch (Thore *et al.,* 2008). Pyrithiamine is an analogue of thiamine with a pyridine ring in place of the thiazole moiety. It thus does not have the biochemical properties of the cofactor and cannot be converted to thiamine, so when pyrithiamine downregulates the expression of thiamine biosynthesis genes or it binds to apoenzymes in place of thiamine, it leads to cell death unless rescued with an alternative supply of thiamine (Sudarsan *et al.,* 2005). Mutations that inactivate the function of the TPP riboswitch have been shown to provide resistance to pyrithiamine in several species (Blount & Breaker, 2006; Moulin *et al.,* 2013; Sudarsan *et al.,* 2005).

Eukaryotic TPP riboswitches were predicted from bioinformatics by Sudarsan in 2003, but it was not until 2007 that their mechanism of action was characterised at the molecular level. That year, four independent studies found TPP riboswitches in different organisms: the fungus *Neurospora crassa* (Cheah *et al.*, 2007); *A. thaliana* (Bocobzca *et al.*, 2007; Wachter *et al.*, 2007); and *Chlamydomonas reinhardtii* (Croft *et al.*, 2007). All these riboswitches are associated with thiamine-related genes and their molecular mechanism is based on alternative splicing (Figure 1.5; Breaker, 2011). In the fungus *N. crassa,* TPP riboswitches are found in the *NMT1* and *THI4* 5'UTRs, in which the absence of TPP allows the sequestration of a potential splicing donor by specific base-pairing with the P4/5 stem (Figure 1.4). When the riboswitch is bound by TPP, this splicing donor is exposed, leading to an alternative splicing event that includes an upstream open reading frame (uORF) in the mRNA, which is translated preferentially over the main open reading frame (ORF) (Cheah *et al.,* 2007). Alternative splicing induced in the presence of TPP in the *A. thaliana THIC* riboswitch works in a similar manner, but in this case the alternative splicing event removes the canonical polyadenylation site, thus destabilising the transcript (Bocobzca *et al.,* 2007; Wachter *et al.,* 2007).

In the TPP riboswitch of the *C. reinhardtii THIC* 6th intron, the 5' strand of the P1 stem has been proposed to base-pair and sequester a potential branch site in the absence of TPP (Figure 1.4). Upon TPP binding, the P1 stem base-pairs and exposes the alternative branch site leading to a splicing event that extends the 6th exon in the mature transcript with an in-frame premature stop codon. The truncated protein is missing crucial catalytic regions from the C-terminus thus preventing enzyme activity (Croft *et al.,* 2007). It is not fully understood how the conformational change in the *CrTHI4* riboswitch mechanistically leads to an alternative splicing pattern, but it has been established that it leads to the inclusion of an uORF with analogous function to the *NcNMT1* riboswitch (Croft *et al.,* 2007; Nguyen *et al.,* 2016). Mutation of the uORF stop codon prevents repression of expression of a reporter gene controlled by the *CrTHI4* riboswitch,

indicating that the uORF translation is necessary for function of the riboswitch (Mehrshahi *et al.*, 2020).

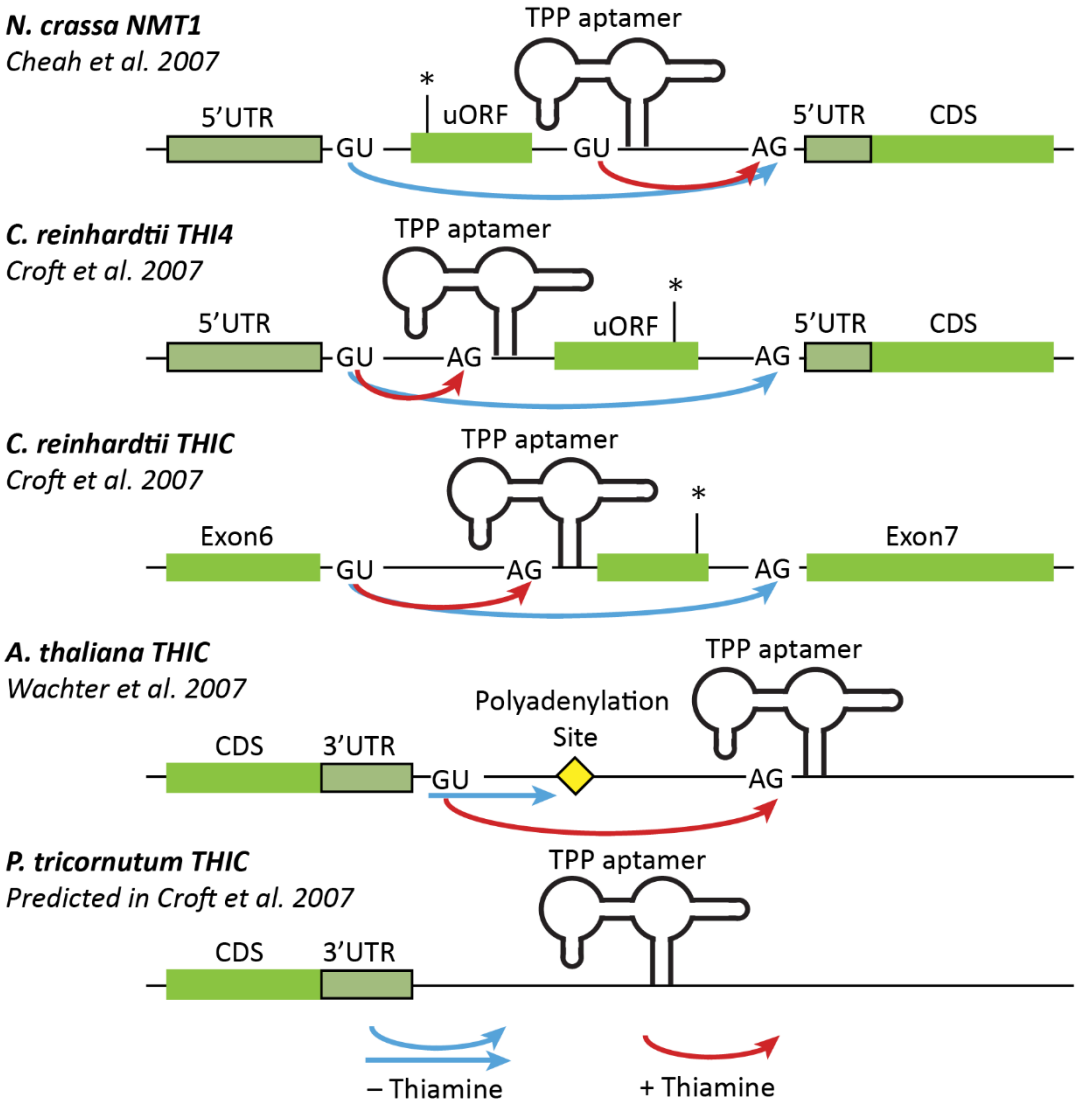


Figure 1.5. Alternative splicing events regulated by eukaryotic TPP riboswitches. *A TPP aptamer has been predicted at the 3'UTR of* P. tricornutum THIC *and its potential function has been studied in this thesis.*

1.2.4 Riboswitches and the RNA World

Despite their relatively recent discovery, riboswitches have been proposed to have originated in the prebiotic RNA world (Monteverde *et al.,* 2017). It is hypothesised that extant riboswitches could have evolved from ribozymes whose substrates or products are the ligands to modern day aptamers. A similar logic could explain why riboswitches that bind coenzymes today could be descendants of ribozymes that used these coenzymes to assist catalysis (Breaker, 2012). An alternative but not exclusive hypothesis is that aptamers were coupled with ribozymes in contiguous sequences so that the binding of a specific ligand to the aptamer regulated ribozyme activity through allosteric interactions (Breaker, 2011). This mechanism may have constituted the first primitive regulatory mechanism that conferred replicators in the RNA World the ability to respond to environmental cues.

The parallels between riboswitch and vitamin origins in the RNA World are not incidental. It is notable that the two most common riboswitch classes bind TPP and cobalamin, and there are also riboswitches that bind flavin mononucleotide (derived from vitamin B_2), tetrahydrofolate (THF, derived from vitamin B_9) and nicotinamide (NAD, derived from vitamin B_3) (Malkowski *et al.,* 2019). Other riboswitch classes bind to nucleotides and their derivatives like adenine, guanine, c-di-AMP, c-di-GMP, ZMP, Pre-Queuosine1 and SAM. Riboswitches are widely distributed across bacterial lineages, but the TPP riboswitch is uniquely found in bacteria, archaea and eukaryotes with a remarkably high conservation in aptamer sequence (McCown *et al.,* 2017). The wide distribution of riboswitches suggests a deep phylogenetic origin and the common binding of molecules that are thought to have been relevant in the RNA World supports the idea of a pre-DNA origin. So far, the TPP riboswitch is the only riboswitch family found in eukaryotes, it remains an open question why other riboswitch families have not been found in organisms other than bacteria (Breaker, 2011).

1.3. Microalgae

1.3.1 Taxonomy and evolution

The term microalgae refers to a loose grouping of all microscopic photosynthetic eukaryotes. Phylogenetically algae can be found in very divergent eukaryotic lineages and can broadly be classified based on whether their chloroplasts originated from a primary endosymbiosis of a cyanobacteria-like cell, or whether they originated from secondary endosymbiosis of other microalgal lineages (Figure 1.6). Primary endosymbionts are thought to originate from a single endosymbiosis event that gave rise to the Chlorophytes, commonly known as green algae, Rhodophytes, commonly known as red algae, and Glaucophytes. Photosynthetic pigment content is commonly used to differentiate between algal groups originating from primary endosymbiosis, Chlorophytes contain chlorophyll *a* and *b*, Rhodophytes and Glaucophytes contain chlorophyll *a* and phycobilins, and Glaucophytes uniquely have a layer of peptidoglycan between the inner and outer membranes of their cyanelles (Keeling, 2010).

The major groups of algae derived from secondary endosymbiosis are the Stramenopiles, Alveolates, Cryptophytes and Haptophytes (Figure 1.6). The chloroplasts in all these groups are generally regarded to come from a rhodophyte secondary endosymbiosis event and contain chlorophyll *a* and *c* (Archibald, 2012; McFadden, 2001). Other less abundant secondary endosymbionts are Euglenoids (within Discoba) and Chlorarachniophytes (within Rhizaria), which arose after the secondary endosymbiosis of a Chlorophyte (Figure 1.6; McFadden, 2001; Sibbald & Archibald, 2020).

P. tricornutum and *T. pseudonana,* the species studied in this thesis, are diatoms which belong to the Stramenopile lineage. Other economically-relevant members of the Stramenopile lineage are Oomycetes, which include plant parasites like *Phytophtora infestans,* which are not photosynthetic and there is no evidence they descend from a secondary endosymbiont (Sibbald & Archibald, 2020). Diatoms are single-celled organisms, although they may form chains, characterised by having a silica-frustule as their cell wall, reducing their dependence on carbohydrate polymers. Each frustule has two parts: the hypotheca and the epitheca, which is slightly larger than the former, separated by a girdle band that marks the region of cell division. There are two main groups of diatoms, pennate and centric, of which *P. tricornutum* and *T. pseudonana* are examples respectively. Diatoms in general, including *Thalassiosira* species, have been observed to undergo both asexual and sexual reproduction, in a life cycle dependent on the size of the theca, the silicate protective cell envelope, however *P. tricornutum* has not been observed to undergo sexual reproduction.

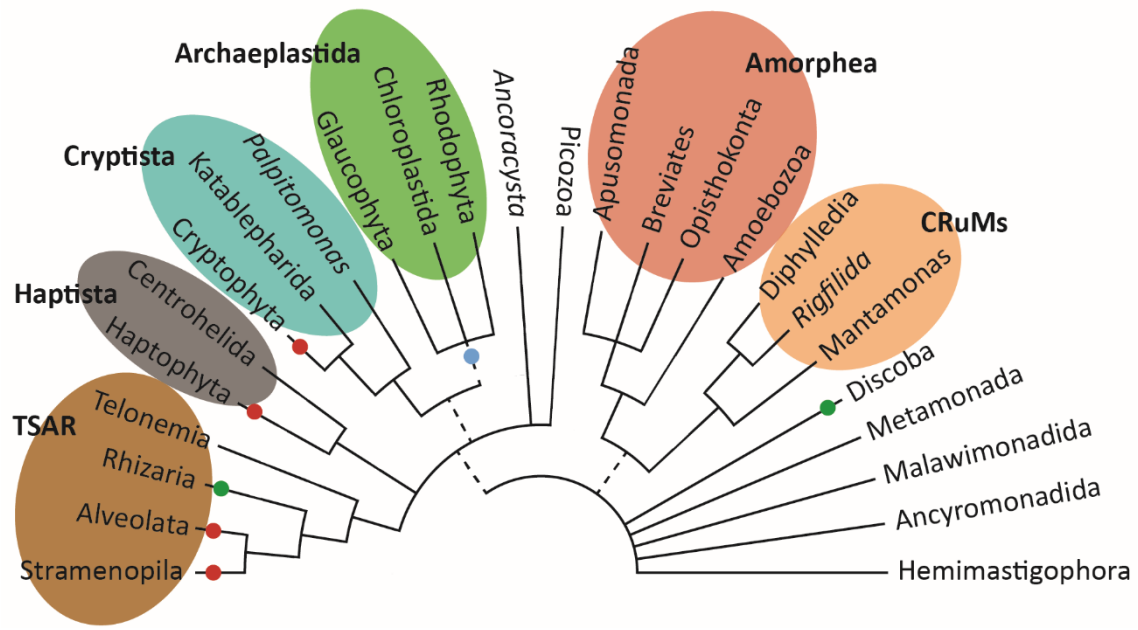


Figure 1.6. Eukaryotic tree of life including main endosymbiosis events. *Dashed lines represent uncertainties about the monophyly of the branch. The blue circle represents the suggested point of a cyanobacterial primary endosymbiosis. Red circles represent the suggested positions of rhodophyte secondary endosymbiosis. Green circles represent the suggested positions of chloroplastida secondary endosymbiosis. Secondary endosymbiosis events do not include all extant branches of each named taxon, eg. Oomycetes in the Stramenopila taxon are not descendant of a secondary endosymbiont and only some Euglenoids in the Discoba taxon are descendant from a secondary endosymbiont. Based on Burki et al. (2020) and Sibbald & Archibald (2020).*

1.3.2 Ecological relevance

The complexity of algal evolution has led to a wide diversity of species that play a crucial role in oceanic ecosystems at a planetary scale. It is predicted that diatoms are responsible for 20 % of global primary productivity, with another 15 % produced by other microalgae, 5 % by marine cyanobacteria and 60 % by land plants (Field *et al.,* 1998; Nelson *et al.,* 1995; Rousseux & Gregg, 2014). The large primary productivity contribution of microalgae is in part explained by events of explosive phytoplankton growth known as algal blooms in response to nutrient upwelling or favourable conditions, generally dominated by diatoms (Stramenopiles), coccolithophores (Haptophytes) and dinoflagellates (Alveolates). Often, species that dominate algal blooms produce toxins, including diatoms such as *Pseudo-nitzschia* species and thus understanding which factors limit and trigger unfettered algal blooming has attracted significant interest given the environmental and economic relevance of algal bloom effects on marine wildlife and human activities.

Many experiments have focused on the growth-limiting role of macronutrient depletion like nitrogen or phosphorus, but it has also been shown that B vitamins, only present at picomolar -

nanomolar concentrations in ocean environments (Sañudo-Wilhelmy *et al.,* 2012; Sañudo-Wilhelmy *et al.,* 2014; Suffridge *et al.,* 2017), can be an important limiting factor for algal blooms too (Croft *et al.*, 2005; Gobler *et al.*, 2007; Sañudo-Wilhelmy *et al.,* 2006). In a sample of 306 algal species whose growth in the laboratory had been assessed, more than half were cobalamin auxotrophs, 22 % were thiamine auxotrophs and 5 % were biotin auxotrophs (Croft *et al.,* 2006). Moreover, it was found that toxin-producing species show auxotrophy for B vitamins in an even higher proportion, with 26 out of 27 species being dependent on an exogenous source of cobalamin for growth (Tang *et al.,* 2010). Changes in B vitamin levels have indeed been shown to correlate with changes in the species composition of blooms in an example of species succession determined in part by B vitamin auxotrophies (Gobler *et al.*, 2007; Sañudo-Wilhelmy *et al.,* 2006). The species composition partly determined by B vitamin levels has also been suggested to influence the strength of the biological carbon pump that fixes carbon and sinks it to the ocean bed with implications for the global carbon cycle and climate change (Sañudo-Wilhelmy *et al.,* 2014).

Vitamins in the ocean can come as influx from land ecosystems or be generated by prototrophic organisms in the community. Research in our group has demonstrated that artificial co-cultures can be set up between the marine picoeukaryote *Ostreococcus tauri* and the rhodobacterium *Dinoroseobacter shibae* in which the microalga supplies the bacterium with vitamin B_3, B_7 and B_9, and the bacteria provides thiamine and cobalamin to the alga in exchange covering their vitamin auxotrophies reciprocally (Cooper *et al.,* 2019). In the natural environment, it is proposed that a complex network of mutualisms and symbiotic relationships between prototrophs and auxotrophs sustains a complex environmental community (Giovannoni, 2012; Kazamia *et al.,* 2016). Thiamine and cobalamin auxotrophies have arisen multiple times in different taxonomic groups suggesting that general patterns of environmental pressures lead to losing genes necessary for vitamin independence. For cobalamin it has been shown that the absence of cobalamin-independent methionine synthase, *METE*, in the genome correlates with cobalamin auxotrophy (Croft *et al.*, 2005; Helliwell *et al.*, 2011), and thiamine auxotrophy correlates with the absence of some of the genes for thiamine biosynthesis (Croft *et al.,* 2006; Bertrand & Allen, 2012). Since the thiamine biosynthesis pathway has two primary branches (Section 1.1.1) organisms can show auxotrophies for one of two main intermediates HMP and HET, for both, or for thiamine, depending on which genes are missing in the pathway (Gutowska *et al.,* 2017). So the picture of thiamine auxotrophy is complex and it is believed that the intermediates play an important role in supporting the growth of auxotrophs in the natural environment. The unique biochemistry of vitamins and their costly biosynthetic processes

(section 1) exert strong selection pressures that underpin the evolution of marine communities, vitamin auxotrophy and the regulation of vitamin-related genes (Giovannoni, 2012).

1.4. *Phaeodactylum tricornutum* genetic engineering tools

In parallel with the increasing interest in the ecological role and biotechnological applications of diatoms, *P. tricornutum* is increasingly used as a model organism. The omics datasets and molecular genetics toolbox to study and engineer *P. tricornutum* have been vastly expanded in recent years (Huang & Daboussi, 2017). In this section I will go through the different tools available to the community for the study and engineering of *P. tricornutum*.

1.4.1. Genome, epigenome, transcriptome and metabolome

The genome of *P. tricornutum* CCAP 1055/1 was sequenced in 2008 (Bowler *et al.,* 2008) with the third and most recent annotation version (Phatr3) being released in 2018 (Rastogi *et al.,* 2018). To assemble the latest version, Rastogi *et al.* screened 90 RNA-Seq datasets, which led to the annotation of 1489 new genes. The most recent piece of omics information added for the use of researchers working with *P. tricornutum* is the DNA methylome and histone modification epigenome (Veluchamy *et al.,* 2013; Veluchamy *et al.,* 2015). Transcriptomics datasets are available for different environmental conditions, but in most cases the studies focus on the effect of nitrogen limitation on central carbon and nitrogen metabolism and lipid accumulation (Levitan *et al.,* 2015b; McCarthy *et al.,* 2017; Remmers *et al.,* 2018). An Expressed Sequence Tag (EST) dataset taken in 16 different environmental conditions is easily accessible in the Ensembl portal for *P. tricornutum* and it is a useful tool to validate expression of annotated genes, map potential alternative splicing events and determine transcription termination sites (Maheswari *et al.,* 2009). The only transcriptomics study under conditions relevant to vitamin metabolism is an iron and cobalamin co-limitation study (Bertrand *et al.,* 2012), and there is an important knowledge gap in the genome-wide responses to different B vitamin limitations.

A metabolic model of *P. tricornutum* with user-friendly data visualization has been developed based on the genome sequence (Fabris *et al.,* 2012). Many metabolomics experiments provide information on the pools of different metabolites in *P. tricornutum* in different environmental conditions, generally with an especial interest in nitrogen depravation and lipid accumulation (Levitan *et al.,* 2015a; Popko *et al.,* 2016; Remmers *et al.,* 2018). But once again, the lack of studies on metabolic changes in response to changing vitamin availability or symbiotic conditions limits our knowledge of the role of vitamin limitation in diatoms and its implication for marine ecology.

1.4.2. Transformation and selection markers

The first genetic transformation of *P. tricornutum* was reported by Apt *et al.* in 1996 with a recorded maximum transformation frequency of 1 in 10^6 cells delivering a zeocin selection cassette by microparticle bombardment. An alternative method to transform the *P. tricornutum* nucleus became available with the development of electroporation delivery with a maximum reported transformation frequency of 1 in 2.8 x 10^5 cells (Zhang & Hu, 2014). Multi-pulse electroporation protocols have shown higher transformation efficiency at up to 1 in 4.5 x 10^5 cells (Miyahara *et al.,* 2013). In contrast to nuclear transformation, which is routinely used, there are sporadic reports of chloroplast transformation by microparticle bombardment (Materna *et al.,* 2009) and electroporation (Xie *et al.,* 2014), but these techniques have not been replicated widely by independent laboratories.

Recently, bacterial conjugation has been established to deliver non-integrative circular DNA fragments known as episomes to the *P. tricornutum* nucleus using a conjugating *E. coli* strain. The episomes are maintained at copy numbers equivalent to native chromosomes thanks to the *CEN-ARS-HIS* yeast sequence and it requires the *OriT* sequence (origin of transfer) to be mobilised from the *E. coli* donor (Karas *et al.,* 2015). The method was later improved to allow medium-throughput transformation in 24-well plates, and to trim the maintenance sequence to *CEN-ARS* (Diner *et al.,* 2016). A new conjugation method has been set up for high-throughput and automatable transformation in liquid in 96-well plates using *Sinorhizobium meliloti* as donor organism (Brumwell *et al.,* 2019). The advantages of conjugation over microparticle bombardment and electroporation include its higher transformation frequency of up to 1 in 4.0 × 10^4 cells, the avoidance of positional and deleterious effects derived from genome integration, and the capacity to remove the episome by removal of selection pressure. Nevertheless, although the episome can be maintained under selection pressure at a population level, it remains to be seen if the episome is retained and the genes expressed homogenously across the population.

Selection markers based on antibiotic resistance are used in *P. tricornutum* to maintain both randomly integrated or episome-based constructs. With the emergence of many new molecular engineering tools and the complexity of the genetic constructs built, more resistance markers have been required, and have been developed in recent years. The first selectable marker used was *Ble*, which confers resistance to zeocin and phleomycin (Apt *et al.,* 1996). Later on, the *NAT* cassette conferring resistance to nourseothricin and the *nptII* cassette conferring resistance to neomycin were reported (Zaslavskaia *et al.,* 2000). More recently, the *bsr* resistance cassette conferring resistance to blasticidin was reported to be functional in *P. tricornutum* (Buck *et al.,*

2018) and a UV-induced random mutation in phytoene desaturase (*PtPDS1*, *Phatr3_J35509*) was reported to lead to resistance to the herbicide norflurazon (Taparia *et al.,* 2019). With the development of the Clustered Regularly Interspaced Short Palindromic Repeats (CRISPR) technology in *P. tricornutum* it has become possible to engineer strains with resistance to other herbicides and knocking-out biosynthetic genes holds promise to develop auxotrophy complementation markers to vastly increase the gene stacking capacity in *P. tricornutum*. It has already been shown that CRISPR targeted mutation on the *PtUMPS* (*Phatr3_J11740*) and *PtAPT* (*Phatr3_J6834*) genes generate auxotrophic strains for uracil and adenine respectively (Serif *et al.,* 2018). Auxotrophic strains enhance the biocontainment of *P. tricornutum* mutants and ongoing efforts are already exploiting complementation markers (Slattery *et al.,* 2020).

1.4.3. Regulatory sequences, target peptides and construct assembly

Regulatory sequences

The first promoters to be used in *P. tricornutum* were *FCPA, FCPB, FCPC, FCPE* and *FCPF* which drive naturally high-level expression of fucoxanthin, chlorophyll *a*/*c*-binding proteins (Apt *et al.,* 1996; Falciatore *et al.,* 1999). The *FCP* promoters are constitutive but have higher expression under light conditions (Falciatore *et al.,* 1999) and were later shown to be much more active in exponential phase than in stationary phase (Erdene-Ochir *et al.,* 2016). The viral promoter, *CIP1*, was reported to show expression levels twice that of *FCPA* (Kadono *et al.,* 2015b). Other characterised endogenous constitutive promoters include *EF2* (Seo *et al.,* 2015), *V-ATPase C* (Watanabe *et al.,* 2018) and *GLNA* (Erdene-Ochir *et al.,* 2016) with reported maximum expression levels 2, 3 and 5-times higher than *FCPA/B* respectively, and with an expression profile not influenced by the diel cycle. Although it should be mentioned that the last study only assayed a single transformant for each construct randomly inserted in the genome, and so positional effects may explain part of the difference. New promoters will ideally need to test several independent transformants, and always be compared with other promoters in episome-based constructs or in targeted genomic integrations to avoid positional artefacts.

In order to control heterologous expression, inducible promoters have also been developed for *P. tricornutum*. The most widely used is the Nitrate Reductase promoter (P_{NR}) (Chu *et al.,* 2016; Niu *et al.,* 2012; Poulsen & Kröger, 2005), which has been shown to induce a 500-fold increase in Nitrate Reductase levels relative to photosynthetic rate when cells are stressed by nitrogen starvation (Levitan *et al.,* 2015b). The Ammonium Transporter promoter (P_{AMT}) has shown up to 25-fold increase in reporter protein levels three days after nitrogen starvation (Adler-Agnon *et al.,* 2018). The alkaline phosphatase promoter (P_{AP}) has been reported to induce an almost 100-fold increase in normalised fluorescence of a reporter construct 72h after phosphate starvation

(Lin *et al.,* 2017). The dynamic range and strength of the nitrate reductase and alkaline phosphatase promoters are excellent parameters for biotechnology application, but the wide metabolic impact and detriment to growth of nitrogen and phosphate starvation are important challenges to introduce these promoters in synthetic biology circuits. Moreover, the AP promoter is leaky (ie. expressed even under phosphorus replete conditions) when used to drive transgenes in our laboratory and thus has much smaller dynamic range (Patrick Hickland, personal communication). There is currently a need to develop orthologous inducible systems with sensitive tunability, high expression levels, high dynamic ranges, but without overarching effects on cell metabolism.

Besides their physiological interest, TPP riboswitches offer a new opportunity to develop modular genetic regulatory elements to use in synthetic biology designs for metabolic engineering. Riboswitches can be combined with promoters of varying strength to provide a breath of strength and controllable expression diversity. Riboswitches are very sensitive to their ligands with K_d values at the sub-nanomolar range which offers interesting savings when scaling up bioproduction systems. Finally, most conditional promoters used to date in diatoms and green algae either rely on light or nitrogen signals. Changes in light regime or nitrogen status have a dramatic impact on the whole metabolism of the cell bringing additional challenges to the metabolic engineering of bioproduction systems. TPP riboswitches would offer a competitive alternative whereby the regulation of the riboswitch would only have localised changes of expression in the thiamine biosynthesis pathway without altering or having a deleterious effect in other areas of cell metabolism. The modularity and orthogonality provided by this regulatory system are of special interest for synthetic biology applications and can help address some metabolic engineering challenges. For instance, Dr Payam Mehrshahi and Aleix Gorchs Rovira have already engineered a *C. reinhardtii* strain with a casbene synthase controlled by the *CrTHI4* riboswitch that conditionally produces casbene only in the absence of thiamine supplementation (Mehrshahi *et al.,* 2020).

Target peptides

The use of target peptides allows the targeting of heterologous proteins to specific compartments in the cell. Target peptides are useful in *P. tricornutum* to study the evolutionary implications of the chloroplast targeting mechanisms for the secondary-endosymbiosis acquisition of chloroplasts (Hempel *et al.,* 2010), to study particular metabolic pathways that happen in specific compartments, for instance HMP-P and HET-P biosynthesis being synthesised in the chloroplast, and to exploit specific pathways for the production of high-value compounds, particularly fatty acids and terpenoids in the chloroplast (Kadono *et al.,* 2015a; Radakovits *et al.,*

2011). Mitochondrial and chloroplast targeting peptides have been characterised using fluorescence proteins (Apt *et al.,* 2002), and a range of other target peptides directing heterologous proteins to different membranes and compartments of the cell were developed by Liu *et al.* (2016). Moreover, the secretion signal from the *HASP1* gene (encoding a protein of unknown function with the highest extracellular abundance in a *P. tricornutum* culture) was shown to direct heterologous proteins to the secretion pathway at high expression levels in combination with its native promoter (Erdene-Ochir *et al.,* 2019).

Construct assembly

Nowadays, some studies in *P. tricornutum* with a limited number of constructs continue to use classic cloning techniques based on two-step restriction-ligation with type II restriction enzymes using acceptor plasmids with multiple cloning sites (Buck *et al.,* 2018). Some groups use Gibson cloning to assemble constructs, but their studies generally include just a few plasmids (Diner *et al.,* 2016; Diner *et al.,* 2017). However, with the advent of more efficient, high-throughput and automatable transformation methods in the last years, the assembly of genetic constructs has become a bottleneck in the synthetic biology pipeline where multiple constructs need testing, and new, more high-throughput, assembly methods have been developed.

A group in Japan introduced Gateway® cloning in diatoms, a method that uses proprietary recombination sites and clonases to assemble constructs in a directional and modular way by recombination (Kadono *et al.,* 2015a; Watanabe *et al.,* 2018), but there has not been widespread adoption of the technique in diatom research. Instead, use of the Modular Cloning (MoClo) system has been developed leading to a library of interchangeable parts that can be flexibly and rapidly used to generate constructs. The MoClo system is based on the GoldenGate one-step restriction-ligation reaction facilitated by the use of type IIS restriction enzymes, which cut outside of their target site leaving a 4-nucleotide overhang with programmable sequence. A downside of the method is that internal restriction sites recognised by the IIS restriction enzymes need to be removed through targeted mutation or synthesis of the parts, commonly known as “domestication”. MoClo was designed to increase the DNA assembly throughput, combinatorial capacity, and the standardisation of parts using a common syntax to make it easier to share parts between different groups, and was widely adopted by the yeast and higher plant communities (Engler *et al.,* 2014; Patron *et al.,* 2015; D’Adamo *et al.,* 2019). In our group, we have assembled a library of MoClo parts for *P. tricornutum*, domesticating previously validated sequences of promoters, UTRs, transit peptides, antibiotic resistance markers, reporter genes, and episome parts (Geisler *et al.* in preparation; further discussed in chapter 4).

1.4.4. Gene silencing and genome editing

Both with the objective to study natural metabolic pathways and to increase production of lipids, gene silencing and genome editing tools have been developed in *P. tricornutum.* These tools have been recently optimised and offer the opportunity to exploit them for the study of vitamin metabolism in *P. tricornutum* in this thesis. Antisense RNA knock-downs were the first tool of its kind to be introduced in *P. tricornutum* (De Risso *et al.*, 2009), and despite not reducing expression to zero, it has been used to study the nitrogen metabolism of the cell (Allen *et al.,* 2011) and to enhance lipid production silencing Nitrate Reductase (Levitan *et al.,* 2015a). Transcription Activator-Like Effector Nucleases (TALEN) were used to knock-out UDP-glucose pyrophosphorylase, with the disruption to sugar metabolism leading to a 45-fold increase in triacylglycerol content (Daboussi *et al.,* 2014). A following study used TALEN-induced homologous recombination to insert a resistance cassette in the urease gene and disrupt its expression to study nitrogen metabolism (Weyman *et al.,* 2015).

TALEN has been quickly left obsolete by the adoption of CRISPR/Cas9 (Nymark *et al.,* 2016), which uses the Cas9 endonuclease from the adaptive immune system of *Streptococcus pyogenes* and a single guide RNA (sgRNA) artificially designed to guide Cas9 to cut at a specific genomic site. Off-target mutations and random integration of the Cas9 expression cassette are common unwanted effects of CRISPR methods. To overcome this bottleneck, several studies have included the Cas9 and sgRNA in episomes delivered by conjugation, which by removing the antibiotic selection pressure lead to the loss of the episome and the generation of non-transgenic knock-out lines (Russo *et al.,* 2018; Sharma *et al.,* 2018; Slattery *et al.,* 2018). The CRISPR/Cas9 technology is increasingly gaining traction with many researchers exploiting its novel potential to study the physiology and metabolism of *P. tricornutum* (Zhang *et al.,* 2019; Coale *et al.,* 2019; Helliwell *et al.,* 2019; Sharma *et al.,* 2020). An alternative solution is delivering Cas9 and sgRNA as a DNA-free ribonucleoprotein (RNP), using this approach it is possible to multiplex the generation of knock-outs with a report of a triple knock-out being generated with this method (Serif *et al.,* 2018). The RNP approach has the advantage of a significantly shorter exposure of the genome to Cas9, reducing the potential for off-target effects, but the reagents are more expensive (Further details in Chapter 5).

New CRISPR methods and variations are being developed continuously in *P. tricornutum* and a method has just been reported that allows targeted homologous recombination of an antibiotic-selection marker mediated by Cas9 double strand break (Moosburner *et al.,* 2020). In *P. tricornutum* non-homologous end joining (NHEJ) is several orders of magnitude more frequent than homologous recombination for DNA repair. Without CRISPR or TALEN double-strand break

induction, homologous recombination has been achieved in a *P. tricornutum* line with a knock-down of the ligase IV gene responsible for the last step of the NHEJ repair mechanism where targeted insertion by homologous recombination of an antibiotic selection cassette was observed without the need to induce a double-strand break, albeit at very low efficiency (Angstenberger *et al.,* 2018). The adoption of CRISPR-mediated homologous recombination allows for a rationally designed disruption of the target genes as well as a significantly easier screening system. Additionally, homologous recombination opens the possibility to solve the longstanding problem of varying positional effects of genome-inserted constructs by standardising insertion sites.

An unresolved drawback of the CRISPR/Cas9 technology is the generation of mosaic colonies in selection plates due to the Cas9-mediated mutagenesis being delayed a few generations (Huang & Daboussi*,* 2017). All studies to date have resolved the issue by re-streaking the primary selected colonies to individualise monoclonal secondary colonies. The process can take up to three weeks and significantly lengthens the knock-out production timeline. Addressing this challenge holds great potential to accelerate the adoption and utilisation of CRISPR technologies to understand and engineer diatoms.

1.5. Thesis aims

The overarching aim of this thesis is to investigate aspects of thiamine metabolism regulation in diatoms. In particular, the main aim of the thesis is to test whether the predicted riboswitch in *P. tricornutum THIC* 3'UTR mediates a genetic response to thiamine levels. To address these questions, I have used different experimental approaches, which are presented in three chapters.

In Chapter 3, I used bioinformatic methods to search for thiamine-related genes in diatom genomes and determine whether they are associated with predicted riboswitches. I also used structural bioinformatic tools to investigate the structural homologies of the predicted thiamine transporter *SSSP*. Motif-prediction tools were used to screen for potential regulatory sequences in the *PtTHIC* promoter and 3'UTR. Chapter 3 also includes experiments based on molecular biology techniques to investigate whether *P. tricornutum* takes up thiamine and whether thiamine or cobalamin supplementation lead to expression changes in thiamine-related genes.

In Chapter 4, genetic constructs were employed to further investigate whether the *P. tricornutum THIC* regulatory sequences can regulate the expression of heterologous genes in response to thiamine supplementation and whether *THIC* responds to intracellular thiamine levels. Reporter constructs were also used to determine whether a conserved motif present in *THIC* and other cobalamin downregulated genes in diatoms identified in Chapter 3 is necessary and sufficient for cobalamin regulation.

In Chapter 5, I developed a standardised CRISPR methodology to generate gene-specific knock-outs with a strategy based on sgRNA pairs and homologous recombination templates in collaboration with other members of our group. I then used this methodology to knock-out thiamine and cobalamin-related genes – *THIC, SSSP, METE* and *CBA1* – to investigate what are their physiological roles and inform the significance of their respective regulatory arrangements.

Chapter 2. Materials and Methods

2.1. Strains and culture conditions

2.1.1. Strains and culture conditions

Phaeodactylum tricornutum CCAP 1055/1 (*P. tricornutum*) was propagated from pre-existing cultures in the lab and routinely grown at 18°C and 30 µmol·m^{-2}·s^{-1} in a 16:8 hours day-night cycle in f/2 minus silica without vitamins (see Appendix 8.1.1 for recipe). *Thalassiosira pseudonana CCAP 1085/12* (*T. pseudonana*) was obtained from the Culture Collection of Algae and Protozoa (CCAP, Oban, UK) and grown in f/2 plus silica and cobalamin at the same temperature and light regime. *C. reinhardtii UVM4* (Neupert *et al.* 2009) was propagated from pre-existing cultures in the lab and routinely grown at 24°C and same light regime in TAP (see Appendix 8.1.2 for recipe). *Escherichia coli DH5α* (*E.coli*) was grown in LB medium (Formedium, Hunstanton, UK) and was used to clone the constructs that were later transformed into *P. tricornutum* and *C. reinhardtii.*

2.1.2. Vitamins and antibiotics

Zeocin (InvivoGen, San Diego, CA, USA) was used at 75 mg·l^{-1} to select transgenic *P. tricornutum* cells and at 10 mg·l^{-1} to select for *C. reinhardtii* transformants. Nourseothricin (Jena Bioscience, Jena, Germany) was used at 300 mg·l^{-1} to select transgenic *P. tricornutum* cells. Thiamine (Acros Organics, Geel, Belgium), pyrithiamine (Sigma-Aldrich, St. Louis, MO, USA), 4-amino-5-hydroxymethyl-2-methylpyrimidine (HMP) (Sigma-Aldrich), or 5-(2-hydroxyethyl)-4-methylthiazole (HET) (Sigma-Aldrich) were added to media at a final concentration of 10 µM – or an otherwise specified concentration – in respective supplementation experiments. Cobalamin (Sigma-Aldrich) was added to the media at a final concentration of 1 µg·l^{-1}, unless otherwise specified. Uracil (Sigma-Aldrich) was added at 30 mg·l^{-1} and 5-fluoroorotic acid (5-FOA) (Sigma-Aldrich) at 100 mg·l^{-1} to select for *UMPS* knock-outs in *P. tricornutum.*

2.1.3. Establishing axenic *P. tricornutum* cultures

Occasionally, *P. tricornutum* cultures were made axenic prior to phenotyping tests or transformations. *P. tricornutum* was subcultured in Nunc Flasks 1:50 up to 40 ml in f/2 supplemented with 100 mg·l^{-1} streptomycin (Melford, Chelsworth, UK), 10 mg·l^{-1} rifampicin (Duchefa, Haarlem, The Netherlands) and 100 mg·l^{-1} gentamycin (Melford). The culture was grown for 5 days at standard conditions and 100 µl were plated in Difco™ Marine Broth (Becton,

Dickinson & Co, Franklin Lakes, NJ, USA) with 1.5 % w/v agar (Melford) plates. If no colonies appeared in the Marine Broth plates after 2 days at 18°C, the culture was considered axenic. Controls plating xenic *P. tricornutum* cultures in Marine Broth plates with or without antibiotics were run in parallel to ensure the culture conditions supported bacterial growth.

2.2. Bioinformatics

2.2.1. *In-silico* structural characterisation

Transmembrane domains were predicted using the TMHMM Server Version 2.0 (Sonnhammer *et al.* 1998) giving the peptide sequence as input and leaving all parameters as default. Protein tertiary structures were predicted with PHYRE2 (Kelley *et al.,* 2015) giving the peptide sequence as input and leaving all parameters as default. Structure alignments were performed with the PDB StrucAlign software (Prlić *et al.,* 2010) run on a local Java console providing the output of the PHYRE2 predictions or experimentally determined structures for a pair of proteins in pdb format as input and leaving all parameters as default, both numeric and graphic outputs were obtained in the form of a calculated rmsd and an overlapped pdb file respectively.

2.2.2. Functional annotation by cell localisation prediction, GO Term and homology searches

SignalP3.0 (Bendtsen *et al.*, 2004) and ASAFind (Gruber *et al.*, 2015) were used to predict the subcellular location of various proteins using the default parameters in all cases. GO term searches against the third version of the *P. tricornutum* genome (Rastogi *et al.,* 2018) were performed in the EnsemblProtists genome browser. Sequence homology searches, BLAST, TBLASTN and BLASTP were done in the National Center for Biotechnology Information (NCBI) server with default parameters unless otherwise stated. The Multiple Sequence Alignment (MSA) for figure 3.2 was performed with Clustal W2 (Sievers *et al.,* 2011), and the output was used to build phylogenetic trees by Neighbour Joining using a BLOSUM62 scoring matrix in a local JalView applet (Waterhouse *et al.,* 2009). For figure 3.3, the MSA was performed with MUSCLE (Edgar, 2004), followed by the construction of a tree by Maximum-Likelihood method with 100 bootstrap iterations (Jones *et al.,* 1992) in the MEGA-X software (Kumar *et al.,* 2018).

2.2.3. Determining the splicing consensus and polyadenylation sites in *P. tricornutum*

Forty-eight introns were taken randomly from the different chromosomes of *P. tricornutum* genome. All intron annotations were supported by numerous EST reads in the Ensembl genome

database (Maheswari *et al.*, 2008). Consensus 5' and 3' splice sites logos were built using WebLogo (Crooks *et al.,* 2004). Branch sites were predicted with the Support Vector Machine – Branch Point (SVM-BP) online server described in Corvelo *et al.,* (2010) using *Homo sapiens* as selected organism. After discarding the predictions with *svm_scr* value lower than 1, the hit with the highest score for each of the 33 remaining introns was used to generate a consensus logo for the branch point. The consensus logos were validated against an independent set of introns. Polyadenylation signal site prediction on the *PtTHIC* 3'UTR was performed with the PASPA server described in Ji *et al.,* (2015) selecting *P. tricornutum* as target species and leaving all other values as default.

2.2.4. Identification of conserved motifs in cobalamin-downregulated genes

The 500 bp upstream of the start codon of the *THIC* gene for diatom species *P. tricornutum, F. cylindrus, P. multistrata, T. pseudonana* and *C. cryptica* were analysed with the MEME algorithm (Bailey *et al.,* 2009) with "Zero or One Occurrence per Sequence" and leaving other parameters as default. The 1000 bp fragments upstream of 32 genes downregulated by cobalamin in Bertrand *et al.* (2012) were selected and provided as input for the MEME algorithm with the "Any Number of Repetition" option and restricting the search to 25 bp long motifs.

2.2.5. TPP aptamer prediction in diatom genomes

The 26 bp long sequence spanning from the 3'P2 to the 3'P4 stems of the *PtTHIC, TpTHIC* and *FcTHIC* were used as queries to BLASTN other existing riboswitches in the genomes and transcriptomes of the other diatom species with Nucleotide-Nucleotide BLAST 2.6.0+ and "blastn-short" parameters. For all hits, the immediately upstream ORF was annotated through a Pfam search and a reciprocal TBLASTN search with *P. tricornutum*. The secondary structure for each predicted aptamer was annotated manually with the assistance of the RNAfold web server tool (Hofacker 2003).

2.3. Nucleic acids handling techniques

2.3.1. RNA isolation

RNA was extracted with the RNeasy Plant Mini Kit (Qiagen, Venlo, The Netherlands) from liquid nitrogen-frozen pellets from 20 ml cultures grown to early stationary phase, unless otherwise stated. Immediately after extraction, the RNA samples were treated with 1 U of TURBO DNase (Thermo Fisher, Waltham, MA, USA) for 30 minutes before cDNA synthesis.

2.3.2. cDNA synthesis

First strand cDNA was synthesised with SuperScriptIII reverse transcriptase (Thermo Fisher) primed with random hexamers. The reactions were conducted in a final volume of 10 µl following the supplier instructions for all other details. Control reactions leaving out the addition of reverse transcription were included to confirm there was no DNA contamination. cDNA was stored at -20°C and later used as template for Reverse Transcription-Polymerase Chain Reaction (RT-PCR) and Reverse Transcription-quantitative Polymerase Chain Reaction (RT-qPCR).

2.3.3. Semiquantitative RT-PCR

A 90 µl PHUSION (Thermo Fisher) Polymerase Chain Reaction (PCR) reaction mix with 1 µl cDNA template was set up for each primer pair and growth condition analysed following the specifications of the supplier. Primers were designed to amplify across introns giving bands of approximately 300 bp in unspliced templates and around 150 bp for templates that had undergone splicing (See Appendix 8.2.1). Each of the reaction mixes was aliquoted in 8 reactions of 10 µl, a PCR programme with 65°C annealing temperature and 20 seconds of extension time was run, and a 10 µl reaction tube was taken out every 2 cycles from 26 to 40 cycles for the target gene, and from 24 to 38 cycles for the Histone 4 housekeeping gene. The gene products were run in a 2 % agarose gel and the band intensity for each cycle was obtained with ImageJ. The band intensity was plotted against the cycle number for each of the reactions. The plot was linearised by taking the $\log_4$ of the band intensity plotted against the cycle number and a linear regression was fit to extract the amplification efficiency value as 4 raised to the slope of the linear model. A threshold value was set at a band intensity value that crossed all linear amplification curves at their exponential phase. A Ct value was calculated with the linear model for each of the genes and conditions, and the Ct values were then used to calculate the fold-change of the target gene between conditions using the DeltaDelta-Ct method (See section 2.3.5).

2.3.4. Analysis of Gene Expression by quantitative PCR

Reactions of 10 µl final volume were prepared for each primer pair (Appendix 8.2.1) and sample by mixing 5 µl of primer mastermix, containing 0.5 µM for each forward and reverse primer diluted in water, with 5 µl of template mastermix, prepared by mixing 2 µl of 1/8 diluted cDNA in 50 µl SybrGreen JumpStart Taq Buffer (Sigma-Aldrich). The reactions were run in 100-well disc in a RotorGene RT-qPCR thermocycler (Qiagen) for 40 cycles of 94°C for 20 seconds, 55°C for 20 seconds, and 72°C for 30 seconds. Three biological replicates and two technical replicates were run for each of the control and test conditions unless otherwise stated.

2.3.5. Calculations and statistics on RT-qPCR

The RotorGene RT-qPCR thermocycler software returned a list of Ct and amplification efficiency values for each of the samples analysed. First, all reactions with amplification efficiency below a cut-off quality value, between 1.4 and 1.6, were screened out. For each reaction, the efficiency value was raised to the negative value of the Ct power to obtain an RQ value for each sample (Equation 1). For each biological replicate, the RQ values of all the housekeeping (*H4, UBC* and *UBQ*, for *P. tricornutum*, and *ACTIN, EF2* and *RBCS,* for *T. pseudonana*) were geometrically averaged (Equation 2), and all sample RQ values were normalised by the housekeeping RQ average corresponding to their biological replicate (Equation 3). In this step, the standard error of the housekeeping RQ average was assumed to be negligible. Then, the average of all normalised RQ values for the control condition was set as calibrator for each gene (Equation 4). A relative expression value was obtained for each sample by dividing the normalised RQ of each sample by the calibrator RQ of their respective gene (Equation 5). A Welch double-sided T-test on the average relative expression was taken between the control and test conditions taking into account the deviation within biological and technical replicates. This RT-qPCR data analysis method is a modification of the DeltaDeltaCt method that takes into account individual amplification efficiencies for each sample (see Rao *et al.,* 2013 for similar RT-qPCR calculations).

$$Eq1.: RQ_i = Eff_i^{-Ct_i} \quad ; for\ each\ sample\ i$$

$$Eq2.: RQ_{Norm_j} = GeoMean\left[RQ_{HKG_j}\right] \quad ; for\ each\ biological\ replicate\ j$$

$$Eq3.: NRQ_i = \frac{RQ_i}{RQ_{Norm_j}} \quad ; for\ each\ sample\ i$$

$$Eq4.: Calibrator_k = Mean\left[NRQ_{Control_k}\right] \quad ; for\ each\ gene\ k$$

$$Eq5.: Relative\ Expression_i = \frac{NRQ_i}{Calibrator_k} \quad ; for\ each\ sample\ i$$

2.3.6. Rapid Amplification of cDNA 3'Ends (3'RACE)

First strand cDNA synthesis was performed with a polyT-VN primer with two anchor nucleotides at its 3' end and a universal adaptor in its 5'UTR (see Appendix 8.2.2 for primers) following the TVN method (Beilharz & Preiss, 2009). For TVN cDNA synthesis, the reaction components were kept as described in section 2.3.2, but the RNA-primer mix was incubated for 5 min at 42°C, followed by 1 h at 42°C and 1 h at 52°C after the addition of reverse transcription and buffer components. The cDNA was diluted 1/8 in nuclease free water and used as template for a first touch-down RT-PCR reaction primed with a forward high-specificity gene-specific primer (71°C annealing Tm) and a universal reverse primer annealing on the universal adaptor introduced with the TVN primer. Q5 High-Fidelity polymerase (New England Biolabs, Ipswitch, MA, USA) was employed with annealing temperatures of 72, 70 and 68°C for 5 cycles each before running 20 more cycles at 65°C. The extension step was kept at 72°C for 30 seconds throughout the 35 cycles. The PCR product of this first RT-PCR was then diluted 1/100 and used as a template for a semi-nested RT-PCR using a nested gene-specific forward primer and the universal reverse primer. For this reaction, Q5 polymerase was used again during 35 cycles using annealing temperature of 65°C and 30 seconds extension. RT-PCR products were run in a 2 % agarose gel and run at 130 mV for 25 minutes. Selected bands were cut, purified and sent for sequencing (Source Bioscience, Nottingham, UK).

2.3.7. Plasmids used and construct cloning

All constructs were cloned following the MoClo Golden Gate system (Engler *et al.* 2014). Level 0 parts were reused from existing constructs or were amplified from *P. tricornutum* genomic DNA using Q5 High Fidelity polymerase (New England Biolabs) (see Appendix 8.2.3 for primers). Level 1 constructs were assembled by BsaI (New England Biolabs) restriction-ligation of Level 0 parts. Level 2 constructs were assembled by BpiI (New England Biolabs) restriction-ligation of Level 1 constructs. All plasmids used in this thesis can be found in Appendix 8.3.

2.3.8. Genotyping and sequencing transformants

Generally, transformants were confirmed to include the expected construct before further phenotypic analysis. Colonies grown on selection plates were picked into selective liquid media in 96-well plates, a week later the transformants were subcultured 1:10 in selective media and after 7 days more 2 µl culture were mixed with 18 µl PHIRE Plant Direct PCR Kit (Thermo Fisher) dilution buffer and stored at -20°C for at least 30 min. One microlitre of the dilution mix was used as template for a PCR with the PHIRE Polymerase and construct-specific primers (see Appendix 8.2.4). The PCR products were run in an agarose gel, the bands were confirmed to have the expected size, and were further cut, purified with the Illustra GFX Gel Ban Purification

Kit (GE Helathcare, Chicago, IL, USA) and sent for Sanger Sequencing with construct-specific primers (Genewiz, South Plainfield, NJ, US; see Appendix 8.2.4).

2.3.9. Agarose Electrophoresis

Agarose (Sigma-Aldrich) was mixed with TAE buffer at a percentage between 0.5 and 2 % as indicated for each experiment, Midori Green (Nippon Genetics, Tokyo, Japan) or Ethidium Bromide (Sigma-Aldrich) at 1:20000 dilution was added, before casting the gel. The gel was run at 90 mV for 50 ml gels, 120 mV for 100 ml gels and at 150 mV for 150 ml gels for approximately 35 min.

2.4. Algae transformation techniques

2.4.1. Transformation of *C. reinhardtii* by electroporation

One microgram of linearised plasmid was electroporated into 2.5 $\cdot 10^7$ *C. reinhardtii UVM4* cells in early exponential phase using a Gene Pulser II electroporator (BioRad, Hercules, CA, USA) with a 0.4 cm cuvette at a capacitance of 25 mF and 800 $V \cdot cm^{-1}$, with no shunt resistor. Cells were recovered in a falcon tube with 10 ml TAP with 60 mM sucrose and incubated at 25ºC, 120 rpm, 2-3 $\mu mol \cdot m^{-2} \cdot s^{-1}$ overnight. The next day, cells were plated on TAP 1 % agar selection plates containing 5 $mg \cdot l^{-1}$ zeocin and incubated for 2 weeks after which colonies appeared.

2.4.2. Transformation of *P. tricornutum* by multi-pulse electroporation

Two and a half micrograms of linearised plasmid were electroporated into $1.25 \cdot 10^7$ *P. tricornutum* WT cells in early stationary phase using a NEPA21 Type II electroporator (Nepa Gene, Ichikawa, Japan) with a 0.2 cm cuvette. The poring pulse was set at 300 V, 5 ms length, 50 ms interval, 10 % decay rate, positive polarity and 3 to 6 pulses to achieve energies between 2.4 and 3.0 J. The transfer pulse was set at 20 V, 50 ms interval, 40 % decay rate, +/- polarity and 5 pulses. Cells were recovered in 5 ml f/2 plus silica for two hours in the dark before incubating them overnight at 18°C and 30 $\mu mol \cdot m^{-2} \cdot s^{-1}$ without shacking. The next day, cells were plated on f/2 minus silica 1 % agar selection plates containing 75 $mg \cdot l^{-1}$ zeocin and incubated for 3 weeks.

2.4.3. Transformation of *P. tricornutum* by microparticle bombardment

Twenty-five millilitres of an early exponential culture ($2\text{-}4 \cdot 10^6$ $cells \cdot ml^{-1}$) were spun down and resuspended in 250 µl of f/2 and $2.5 \cdot 10^7$ cells were plated at the centre of a f/2 + 1 % agar plate. Then, between 3 and 7.5 µg of plasmid at 1 $\mu g \cdot \mu l^{-1}$ were mixed with 1.5 mg of DNAdel™ 550 nm golden particles (Seashell Technology, La Jolla, CA, USA). Next, the mix was resuspended to 30 $mg \cdot ml^{-1}$ in 50 µl binding buffer and an equal volume of precipitation buffer was added. Gold

particles were then spun and washed with 100 % ethanol and resuspended in 30 µl of ethanol. Ten microlitres of gold particles were added to the centre of macrocarriers and were left to dry. The *P. tricornutum* plate was placed at the bottom of the biolistic device, with the microcarrier mounted on the top position and a 1350 psi rupture disk screwed into the gun. The DNA was bombarded with a pressure of 1550 psi. Cells were then scrapped with 800 µl of f/2 media and a spreader, and transferred to 4 f/2 + 1 % agar selection plates containing the required antibiotics. Colonies appeared after two to three weeks at 18°C and constant light.

2.4.4. Transformation of *P. tricornutum* by bacterial conjugation

P. tricornutum was transformed by conjugation following the protocol conceived by Karas *et al.* (2015). Briefly, an early exponential culture (2-4·10^6 cells·ml^{-1}) was spun down and concentrated to 10^8 cells·ml^{-1} of which 250 µl were plated on f/2 + 5 % LB + 1 % agar plates for every conjugation event. *P. tricornutum* cells were grown for 4 days before a 25 ml culture of *E. coli* carrying the pTA-MOB plasmid and the episome of interest was grown to 0.8-1.0 OD_{600nm}, spun down, resuspended in 250 µl, 200 µl of which were added to the conjugation plate and mixed with the *P. tricornutum* cells with a spreader. The conjugation plate was then incubated for 90 min at 30°C in the dark before moving to 18°C constant light for 2 days. Finally, cells were scraped off the conjugation plates exactly like in the microparticle bombardment procedure.

2.5. Phenotyping techniques

2.5.1. Fluorescence reporter, antibiotic resistance and vitamin auxotrophy assays

Triplicate cultures in replete and deplete conditions were grown in parallel for at least 7 days in 200 µl f/2 in 96-well plates. Every 24 hours, OD_{730nm} was measured as a proxy for growth, and Venus fluorescence was measured with excitation at 515 nm and emission at 550 nm with a CLARIOStar plate reader (BMG Labtech, Ortenberg, Germany).

2.5.2. Determination of intracellular thiamine quotas

Cell pellets were harvested 5 days post inoculation, washed three times with distilled water and wet weight of the final pellets was measured before flash freezing in liquid nitrogen and storing at -80°C. Pellets were treated with 250 µl 1 % (v/v) trichloroacetic acid (Sigma-Aldrich) and centrifuged at 10000 g for 10 minutes recovering the supernatant. TPP and thiamine were then derivatised to thiochrome pyrophosphate and thiochrome respectively by mixing 50 µl of the cell extract with 10 µl of freshly prepared 30 mM potassium ferricyanide (Sigma-Aldrich) in 15 % (w/v) sodium hydroxide (Thermo Fisher), 15 µl 1 M sodium hydroxide, and 25 µl methanol

(Sigma-Aldrich, HPLC-grade). The derivatisation mix was centrifuged at 4000 g for 10 minutes and 20 µl of the supernatant were injected for High Performance Liquid Chromatography (HPLC) analysis. An Accela HPLC setup (Thermo Scientific) was used with a C18 150 x 4,6 mm column (Phenomenex, Torrance, CA, USA). The fluid phase flowed at 1 ml·min^{-1} and was programmed as a gradient where at time 0 it was 5 % methanol, up to 47,5 % at 10 minutes, 100 % at 11 minutes, 100 % at 15 minutes, 5 % at 16 minutes and equilibration at 5 % methanol until 21 minutes. Thiochrome and thiochrome pyrophosphate fluorescence was measured using a Dionex UltiMate 3000 fluorescence detector (Thermo Scientific) with 375 nm excitation and 450 nm emission. The sensitivity of the fluorescence detector was set at 1 for the first 5 minutes of the HPLC programme and increased at 8 for the rest of the programme until the 21 minutes. The area of thiochrome pyrophosphate and thiochrome peaks was used to calculate the titre of each vitamer relative to their respective standard curves.

2.5.3. Cobalamin uptake assay

P. tricornutum cells at stationary phase were harvested and concentrated to 5·10^{6} cells·ml^{-1} and incubated in 1 ml f/2 media supplemented with 600 pg cobalamin in a microcentrifuge tube for one hour. A control with 1 ml f/2 and no cells spiked with 600 pg of cobalamin was included to control for cobalamin loss in the treatment of the samples. Cells were spun down to separate the cellular fraction (pellet) from the media fraction (supernatant). The cellular fraction was resuspended in 1 ml fresh f/2 media and each fraction was boiled for 10 minutes to release any intracellular cobalamin. The cobalamin content of each of the fractions was measured with a bioassay based on *Salmonella typhimurium AR3612 cysG metE*, auxotrophic for cobalamin (Raux *et al.,* 1996). For each fraction, 250 µl were diluted in 750 µl sterile water and 1 ml 2xM9 media plus 0,075 % of a washed *S. typhimurium* inoculum in a 0.9 % NaCl solution in a microtiter plate well. *S. typhimurium* cultures were incubated for 12 h at 37°C and 180 rpm, and OD_{595nm} was measured as a proxy for growth in a FLUOstar Omega plate reader (BMG Labtech). The concentration of cobalamin in each sample was estimated from the results of a standard curve of known cobalamin concentrations in f/2 media.

2.5.4. Luciferase assay

P. tricornutum transformants carrying the *Gaussia princeps* luciferase cassette were diluted 1:10 in 200 µl f/2 supplemented with 75 µg·ml^{-1} zeocin in 96-well plates. After 4 days of growth, their OD_{730nm} and luciferase activity were measured in a FLUOstar Omega plate reader (BMG Labtech). The assay was automated to add 100 ng of the luciferase substrate coelenterazine (Promega, Fitchburg, WI, USA) immediately before recording light emission every 0.2 seconds over 2 seconds for each transformant. The 10 luciferase activity data points were averaged. Those

clones having equal or higher luciferase activity than the lowest positive control were selected. The data for the selected clones was manually inspected to remove false positives which did not follow a logarithmic luciferase activity progression over the two seconds measurement.

2.5.5. Western Blots

Total cell extracts were obtained mixing pellets from 150 ml cultures grown to early stationary phase in X ml of 0.2 M sorbitol (Sigma-Aldrich), 1 % β-mercaptoethanol, and 0.8 M Tris-HCl pH 8.3 (Sigma-Aldrich), where X is equal to the culture OD_{750nm} before harvesting. Cell extracts were supplemented with 1 % Sodium Dodecyl Sulphate (Sigma-Aldrich) and boiled for 1 min. The samples were centrifuged at 16 krpm for 2 minutes and 15 µl were loaded in a 15 % Acrylamide Sodium Dodecyl Sulfate–Polyacrylamide Gel Electrophoresis (SDS-PAGE) gel. The electrophoresis was run at 150 mV for 90 minutes, the proteins were then transferred to a Polyvinylidene Difluoride (PVDF) membrane at constant 20 mA during 20 minutes in a semi-dry transfer cell (BioRad). The membrane was blocked in 0.5 % powder milk in TBS-T buffer at 4°C overnight, then incubated for 1 h with a rabbit-antiHA primary antibody (H6908, Sigma-Aldrich) in 2.5 % powder milk in TBS-T, washed 4 times with TBS-T, then incubated for 1 h with a goat-antirabbit secondary antibody conjugated with a Dy800 fluorophore (SA5-35571, Thermo Fisher) in 2.5 % powder milk in TBS-T. The membrane was finally washed 4 times in TBS-T and one last time in TBS before being imaged in a fluorescence scanner (Odyssey, Li-Cor, Lincoln, NE, USA).

2.5.6. Confocal imaging for intron testing constructs in *P. tricornutum*

For *P. tricornutum* transformants carrying the *pMLP2041*, *pMLP2042* and *pMLP2047* constructs, cells were grown for 2 days in f/2 and zeocin in 96-well plates before visualization in a confocal laser scanning microscope (TCS SP5, Leica Microsystems, Wetzlar, Germany). The images were taken with excitation from an Argon laser source at 514 nm and emission detected by PMT detectors between 518 nm and 560 nm for Venus and between 680 nm and 700 nm for chlorophyll.

2.5.7. Confocal imaging for *C. reinhardtii* chimeric riboswitch constructs

For *C. reinhardtii* transformants carrying the *pPM34* construct, cells were grown for 7 days in TAP media without supplementation, or in the presence of 10 µM thiamine or 10 µM HET, before visualization in a confocal laser scanning microscope (TCS SP8, Leica Microsystems). The images were taken in the sequential mode enabled by the Leica LAS software. The first image was acquired with excitation from a white light source at 476 nm and emission detection between 485 and 518 nm for the Green Fluorescence Protein (GFP) channel, and between 650 nm and

720 nm for the chlorophyll channel. The second image was acquired with excitation at 516 nm and emission detection between 526 and 592 nm for the Venus channel.

Chapter 3. Thiamine biosynthesis regulation in *P. tricornutum*

3.1 Introduction

The work in this chapter follows on the research done by numerous people of the research group that have extensively studied thiamine metabolism, its regulation, and the potential of TPP riboswitches for synthetic biology in *C. reinhardtii.* Croft *et al.* (2007) first reported the discovery of TPP riboswitches in *CrTHI4* 5'UTR and *CrTHIC* 6th intron, characterised their response to thiamine, and proposed a mechanism based on alternative splicing. Moulin *et al.* (2013) further characterised the role of the *C. reinhardtii* riboswitches in regulating the thiamine biosynthesis pathway, showing that *pyr1,* a strain with a mutation in *CrTHI4* aptamer, had increased levels of intracellular thiamine, and demonstrating that the *CrTHI4* riboswitch could regulate expression of a reporter construct. In their PhD theses, Balia Yusuf (2011) showed how the *pyr1* mutant was resistant to pyrithiamine, and Ginnie Nguyen (2014) built an aptamer platform based on the *CrTHI4* riboswitch where different aptamers could be combined with the expression platform in a modular fashion.

The work by Croft *et al.* (2007) was the first report of riboswitches in algae, in the same year the molecular mechanism of riboswitches in fungi and plants were also described (Cheah *et al.,* 2007; Wachter *et al.,* 2007). All eukaryotic riboswitches known to date are associated with thiamine metabolism genes, respond to TPP or intermediates of the thiamine biosynthesis pathway, and their mechanism of action is based on alternative splicing. The TPP binding domains and the 3D structure of riboswitches is highly conserved not only within eukaryotic organisms, but also in bacteria and archaea (Sudarsan *et al.,* 2003). The conservation of negative feedback regulation of thiamine biosynthesis across kingdoms and in most organisms known to synthesise thiamine *de novo* has been explained by the need to balance the metabolic cost of thiamine biosynthesis (Hanson *et al.,* 2018), with the essential requirement for thiamine (See section 1.1.1 for further detail).

Thiamine metabolism is also highly interconnected with other B vitamin metabolic pathways. For instance, Bertrand *et al.* (2012) showed that one of the most downregulated proteins by cobalamin in *P. tricornutum* is THIC*,* the enzyme responsible for the first reaction of the pyrimidine branch of the thiamine pathway that uses SAM as a substrate. SAM, in turn, is part of the C1 cycle where the cobalamin-dependent methionine synthase is a key enzyme. Folates

(vitamin B_9) are also closely related with the C1 cycle, with 5-methyltetrahydrofolate being the methyl donor to methionine synthase. PLP (vitamin B_6) is also related to folate and cobalamin metabolism as a cofactor of SHMT, an enzyme also downregulated by cobalamin supplementation in *P. tricornutum* (Bertrand *et al.,* 2012), and that produces 5,10-methylenetetrahydrofolate from tetrahydrofolate with a carbon being transferred from the side chain of serine. Finally, in fungi PLP has been shown to be the substrate of *THI5* and *NMT1*, suicidal enzymes that synthesise HMP-P, the pyrimidine moiety in the thiamine biosynthetic pathway (Coquille *et al.,* 2012).

In the early work in the lab by Croft *et al.* (2007) a TPP aptamer was predicted in the 3'UTR of the diatom *P. tricornutum* by secondary structure predictions, and the presence of the THI box sequence. More recently, McRose *et al.* (2014) used a pipeline mostly based on sequence conservation and secondary structure prediction to identify putative riboswitches in Haptophytes, Cryptophytes, Alveolata, Rhizaria, Mamelliophytes and numerous Stramenopiles (see section 1.3.1 for more information on algal lineages). The ecological relevance of these groups of photosynthetic eukaryotes cannot be overstated. The marine phototrophs are responsible for approximately 40 % of global primary production, and diatoms alone have been estimated to contribute 20 % of global primary production (Field *et al.,* 1998; Nelson *et al.,* 1995; Rousseux & Gregg, 2014). Previous studies in the laboratory have revealed that a significant proportion of marine algae require an exogenous source of B vitamins to support their growth, including 50 % of the species having a cobalamin requirement and 20 % having a thiamine requirement (Croft *et al.,* 2006). B vitamins have also been found to be involved in the formation of symbiosis and microbial ecology interplay in the ocean environment playing an important role in global biogeochemical cycles and in the appearance of algal blooms (Giovannoni, 2012).

The objectives of the work described in this chapter are to:

1. Identify candidate genes involved in thiamine biosynthesis and transport in *P. tricornutum*
2. Predict TPP aptamers in all available diatom genomes to study their conservation
3. Investigate whether *P. tricornutum* can take up thiamine and which transporter is responsible for it, if any
4. Investigate whether the *PtTHIC* riboswitch gives a genetic expression response to thiamine supplementation
5. Investigate the molecular mechanism by which cobalamin regulates *PtTHIC*

To address these questions, both bioinformatic and molecular biology methods have been used and their respective results are presented in two separated sections below.

3.2 Bioinformatics

3.2.1 Identification of the thiamine biosynthesis pathway in *P. tricornutum*

In order to set a metabolic framework to study the regulation of thiamine biosynthesis in *P. tricornutum,* the thiamine pathway with all the annotated gene IDs available was reconstructed. A combination of key word, GO term and homology searches was used to query the *P. tricornutum* genome and identify the genes involved in the thiamine biosynthesis pathway. Most of the genes found were already identified in Bertrand & Allen (2012), but the analysis revealed additional relevant candidates that may play an important role to understand thiamine biosynthesis regulation in diatoms. A graphical reconstruction of the pathway with all genes already described and newly identified can be found in figure 3.1.

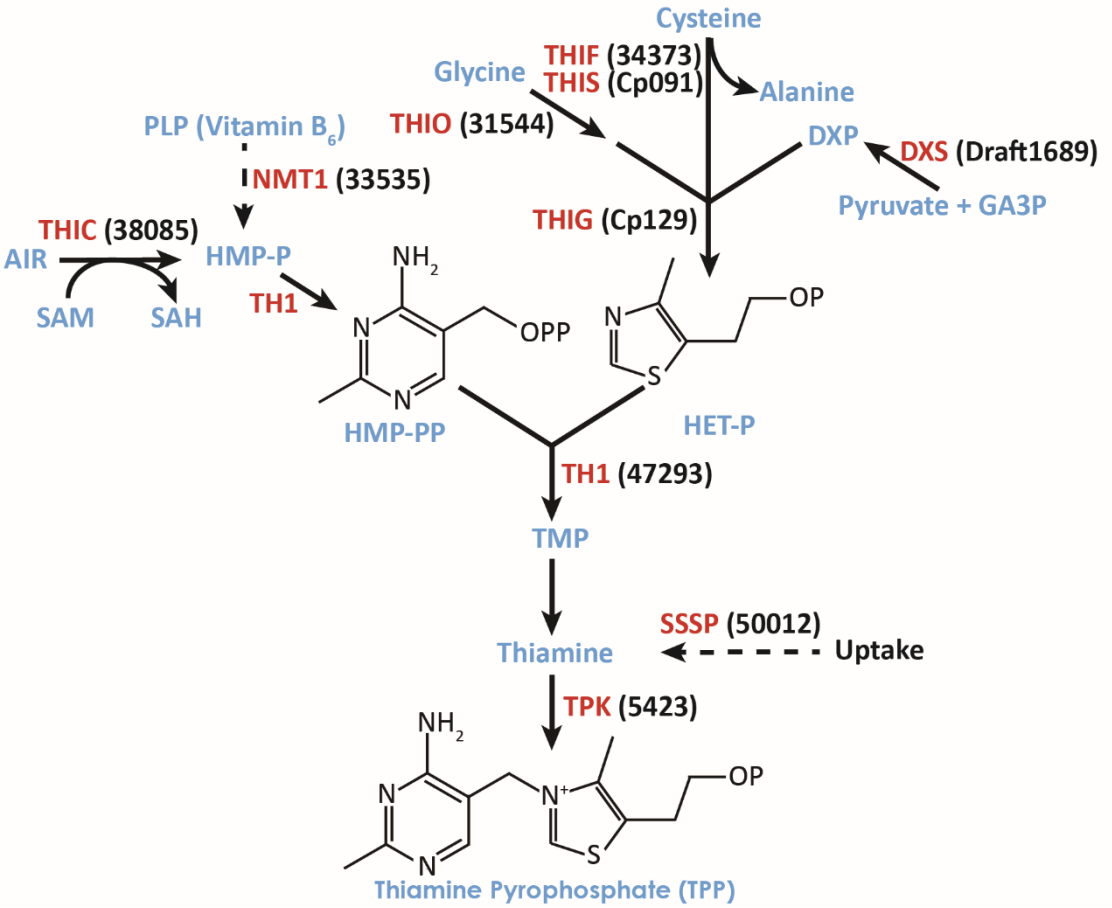


Figure 3.1* P. tricornutum *thiamine biosynthesis pathway. *Intermediates are represented in blue, genes are represented in red, and their corresponding gene ID is presented between brackets. The final product of the pathway, TPP, and the two products of the initial pathway branches, HMP-PP and HET-P, are represented with their molecular structure. Dashed arrows indicate hypothetical reactions without experimental support in* P. tricornutum*.*

The thiazole (HET-P) biosynthesis branch in *P. tricornutum* was found to be composed of three enzymes encoded in the nuclear genome (DXS, THIF and THIO) and two enzymes encoded in the chloroplast genome (THIS and THIG), suggesting that either the branch is partitioned between two cellular compartments or the nuclear encoded enzymes are targeted to the chloroplast. The last enzyme of the branch, THIG, is encoded in the chloroplast which suggests that HET-P is produced in the chloroplast and later exported to the cytoplasm to be used as substrate by TH1.

To test whether THIG substrates are themselves produced in the chloroplast or imported from another cellular component, the subcellular locations of DXS, THIF and THIO were tested. The online secondary-endosymbiont-specific subcellular targeting prediction software ASAFind returned with high confidence that DXS and THIO were targeted to the plastid. THIF was not predicted to have a signal peptide by SignalP3.0. THIS is the substrate of THIF and is encoded by the chloroplast genome, so it seems unlikely that THIF is not localised in the chloroplast. An explanation could lay in a wrong annotation of the coding sequence of the *THIF* gene in the current version of the genome, so 3 different truncations of the gene starting at the first 3 methionine codons were also tested with the same software as well as a 98-residue-long upstream ORF unannotated in the genome. All THIF sequence truncations tested were predicted to have no signal peptide by SignalP suggesting that either *THIF* was wrongly annotated by Bertrand & Allen (2012) or its subcellular location is not correctly predicted by SignalP.

All enzymes in the pyrimidine (HMP-PP) branch are encoded in the nuclear genome. In addition to *THIC* and *TH1* that were already described in Bertrand & Allen (2012), a search with the GO term for “thiamine biosynthetic process” returned two genes not previously reported. Firstly, the gene *Phatr3_EG00135* on chromosome 27 was also annotated as *THIC*, but showed no associated EST data and is unusually short with 126 amino acids compared to the *THIC* gene previously annotated that is 619 amino acids long (*Phatr3_J38085;* Bertrand & Allen, 2012), similar to the length of its homologue in *A. thaliana* (644 amino acids). *Phatr3_EG00135* has a partial hit to the PFAM domain for “phosphomethyl pyrimidine synthase THIC” and its peptide and nucleotide sequence is highly conserved with *Phatr3_J38085*. These results suggest that *Phatr3_EG00135* could have arisen as a partial gene duplication which lead to a dysfunctional gene, the expression of which was eventually lost.

Another gene found through GO-term search was *Phatr3_J33535* on chromosome 3, which contained associated EST reads confirming it is expressed. This gene is annotated as *NMT1,* which is the gene encoding for the suicidal HMP-P synthase and contains a TPP riboswitch in its 5’UTR in the fungi *N. crassa* (Cheah *et al.,* 2007). To provide a greater degree of confidence to

the *NMT1* annotation, a series of homology and structural prediction searches were carried out. In the first place, the Phatr3_J33535 peptide sequence was used as a query in a HMMER search with default parameters against the whole EMBL-EBI proteome repository. Hits included species in Stramenopiles *T. pseudonana* (GeneID *6708*), Cryptophytes *G. theta* (*165857*), Haptophytes *E. huxleyi* (*233314*) with an e-value lower than 1e-100 and in Chlorophytes *C. reinhardtii* (*07g332450*), *V. carteri* (*110035*), and Streptophytes *Physcomitrella patens* (*PHYPA_030848*) with e-values between 1e-5 and 1e-8. Out of the 453 hits in the Eukaryota lineage, 428 were under the fungi clade including *S. cerevisiae* THI5 and *N. crassa* NMT1. Both ScTHI5 and NcNMT1 have been experimentally shown to synthesise HMP from PLP in a suicide reaction that degrades a histidine in the active site of the enzyme (Coquille *et al.*, 2012; McColl *et al.*, 2003). The NcNMT1 peptide sequence was used as query in a HMMER search against the *P. tricornutum* proteome and a single hit corresponding to the candidate Phatr3_J33535 was found. In a search for protein family classification with the Phatr3_J33535 peptide sequence in NCBI returned three hits: NMT1 (e-value 4.82e-49), ThiY (e-value 4.81e-33) and TauA (e-value 5.94e-20). Both ThiY and TauA are annotated as ABC-type transporters, and ThiY has been proposed to be a structural homologue of NMT1 involved in the FAMP salvaging system in bacteria, once inside the cell FAMP can be metabolised into HMP-PP (Section 1.1.1; Bale *et al.,* 2010).

To elucidate whether the algal NMT1 candidates perform a biosynthetic function or instead they are misannotated in the genome and they code for a transporter similar to the ThiY family as suggested in Ginnie Nguyen Thesis (2014), a MSA including all algal NMT1 candidates, NcNMT1, ScTHI5, *Bacillus halodurans* ThiY, and *E. coli* SSUA transporter (as an outgroup) was carried out using ClustalW2 followed by the determination of a Neighbor-Joining phylogenetic tree (Figure 3.2). The phylogenetic tree clustered *P. tricornutum, Thalassiosira pseudonana* (Stramenopiles), *Guillardia theta* (Cryptophyte) and *Emiliania huxleyi* (Haptophyte) in a monophyletic group. The green algae and the diatom *Fistulifera solaris* NMT1 candidates were also monophyletic. The two clades of NMT1 candidates had equal distance to each other, BhThiY, and the ScTHI5 and NcNMT1 pair. It is interesting to note that FsNMT1 is the only gene encoding for a THI5/NMT1 protein shown to originate from an horizontal gene transfer event among 9 diatom genomes analysed (Vancaester *et al.,* 2020). The MSA showed that out of the known 9 active site residues for BhThiY and 15 active site residues for ScTHI5 (shadowed in blue and red respectively in figure 3.2; Bale *et al.,* 2010; Coquille *et al.,* 2012), green algae and *F. solaris* conserved 3 to 4 residues of the BhThiY active site and 1 to 2 residues of the ScTHI5 active site. The NMT1 candidates for *P. tricornutum, T. pseudonana, G. theta* and *E. huxleyi* conserved 4 to 5 residues of the BbThiY active site and 6 residues of the ScTHI5 active site. The different ratio of active site residues

conserved for the green algae or the secondary endosymbiont clades could indicate that the function of the NMT1 candidate could be different for the two groups. Unfortunately, the phylogenetic tree did not have enough resolution to resolve whether the algal THI5/NMT1 homologues are closer related to BhThiY and could have a transport function, or to ScTHI5 and could have a catalytic function.

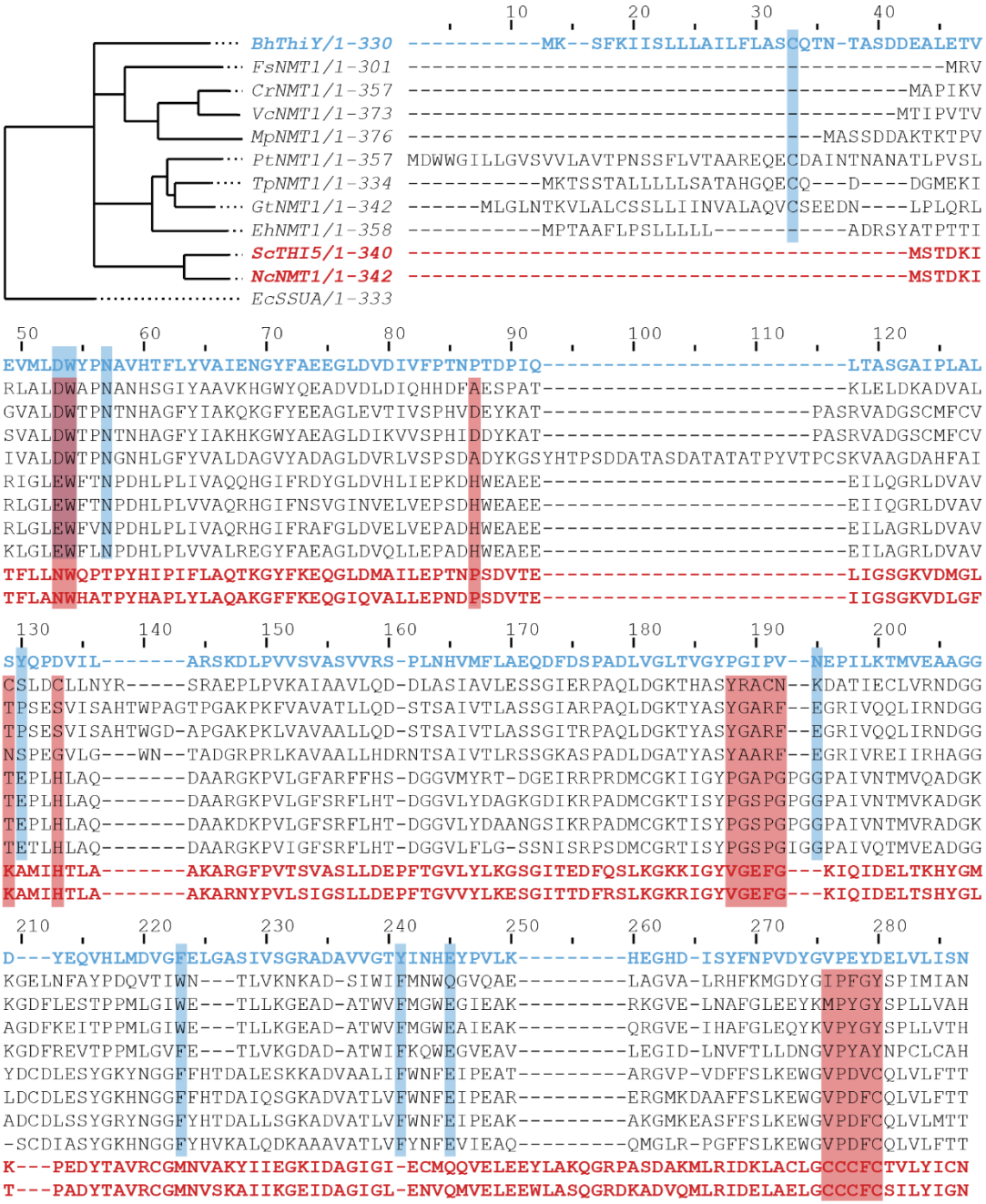


Figure 3.2. Phylogenetic tree and MSA for the NMT1/THI5-like algal candidates. *The first 287 positions of the MSA are shown for the NMT1 algal candidates, the bacterial thiamine transporter BhThiY (in blue letters) and the NcNMT1 and the ScTHI5 homologues (in red letters).The* E.coli *SSUA thiamine transporter was added as an outgroup for the tree and its sequence is not shown in the MSA. The tree was generated in JalView with a Blossum62 matrix and a Neighbour-Joining algorithm. The active site residues of BhThiY are highlighted in blue boxes and the active site residues for ScTHI5 are highlighted in red boxes to evaluate conservation of key residues in the NMT1 algal candidates. Bh:* Bacillus halodurans*; Fs:* Fistulifera solaris*; Cr:* Chlamyomonas reinhardtii*; Vc:* Volvox carteri*; Mp:* Micromonas pusilla*; Pt:* Phaeodactylum tricornutum*; Tp:* Thalassiosira pseudonana*; Gt:* Guillardia theta*; Eh:* Emiliania huxleyi*; Sc:* Saccharomyces cerevisiae*; Nc:* Neurospora crassa*; Ec:* Escherichia coli.

In order to get a phylogenetic tree with better resolution the sequences for several THI5/NMT1 and ThiY homologues from a previous phylogenetic study on the protein family (Bajor *et al.,* 2014) were used to generate an expanded MSA and phylogenetic tree with the algal NMT1 candidates. The protein sequences for all selected genes were aligned using a MUSCLE algorithm and the MSA was used to build a Maximum Likelihood phylogenetic tree with 100 bootstrap iterations. The analysis confirmed the green algae plus *F. solaris* and the secondary endosymbiont monophyletic branches and indicated with strong statistical support that the algal candidates are closer related to the bacterial FAMP transporter ThiY than the fungal HMP-P synthase THI5/NMT1. The green algae cluster is a sister branch with BhThiY with the secondary endosymbiont cluster being more distantly related to the bacterial homologue (Figure 3.3). Nevertheless, when the SignalP signal peptide prediction software was used on the PtNMT1 peptide sequence it did not predict the presence of a signal peptide, suggesting PtNMT1 resides in the cytoplasm which would not be compatible with a transport function.

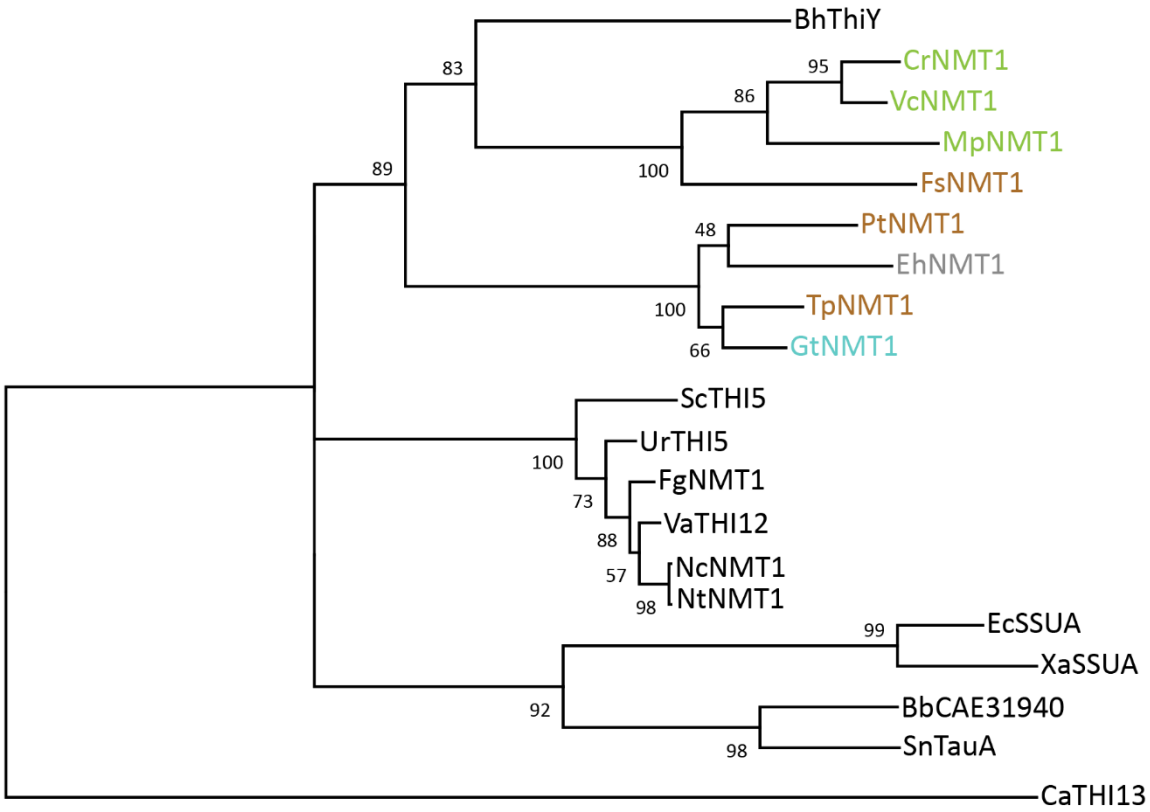


Figure 3.3. Maximum likelihood phylogenetic tree of algal NMT1 candidates and members of the protein family with ThiY and THI5/NMT1 annotations. *An MSA including 3 NMT1 candidates in green algae (green), 4 in diatoms (brown), one in haptophytes (grey) and another in cryptophytes (cyan) and known ThiY and THI5/NMT1 homologues in bacteria and fungi (Bajor et al., 2014) were aligned with MUSCLE. A Major Likelihood tree was built with 100 bootstrap replications and the bootstrap values are given for each branch. There is strong support to confirm that the algal candidates are more closely related to the bacterial ThiY transporter than to the fungal THI5/NMT1 HMP-P synthase.*

In order to see whether NMT1 homologues could co-occur with THIC in other eukaryotic organisms whose thiamine biosynthesis regulation has been extensively researched, the search was extended to plants with a blastp homology search with NcNMT1 as a query against the *A. thaliana* proteome. However, no significant hits were found. Reciprocally, the AtTHIC protein sequence was used as a query in a blastp homology search against the *N. crassa* proteome again finding no significant hits. Both homology searches against *P. tricornutum* proteome returned the expected PtNMT1 and PtTHIC hits. It seems then, that the co-occurrence of *THIC* and *NMT1* can be found in diverse algal groups (Chlorophytes, Stramenopiles, Haptophytes, Cryptophytes, Mamelliophytes) and even in Bryophytes but not in fungi or higher plants.

3.2.2 Functional characterisation of the *PtTHIC* 3'UTR sequence

Croft *et al.* (2007) predicted the presence of a TPP aptamer in the 3'UTR of *PtTHIC* and McRose *et al.* (2014) predicted the presence of a TPP aptamer in the 3'UTR of *PtSSSP*. In *A. thaliana, THIC* has a TPP aptamer in its 3'UTR that regulates gene expression in response to exogenous thiamine by modulating an alternative splicing event that splices out a polyadenylation site (Wachter *et al.,* 2007). In order to expand our knowledge on the potential molecular mechanisms occurring in the *PtTHIC* 3'UTR that could influence the regulation of the thiamine biosynthesis pathway, the *PtTHIC* 3'UTR sequence was analysed in more detail.

All eukaryotic TPP riboswitches known to date base their molecular mechanism on alternative splicing (Nguyen *et al.*, 2016). In order to investigate the potential molecular mechanism of the predicted riboswitches in *PtTHIC* and *PtSSSP* it was essential to know what the splicing consensus motifs in *P. tricornutum* are. With that objective, 48 random introns from all chromosomes of the *P. tricornutum* genome were collected after validating through EST reads that they were highly expressed and spliced. The 5' and 3' ends of the selected introns were aligned and the corresponding intron-exon junction sequence logos are shown in figure 3.4a. Then, a branch point prediction server based on a Support Vector Machine (SVM) learning algorithm was used to predict branching point sites for all the introns selected, the *H. sapiens* parameters were used as there were only mammalian training sets (Corvelo *et al.,* 2010). The best scoring branching point candidate for each of the 48 introns was selected and the 15 branching points with a score lower than 1 (the programme specifies not to consider those hits with score lower than 0) were removed. With the 33 remaining branching points the sequence logo was built and the distance to the 3' splicing site was plotted as a frequency distribution (Figure 3.4b). Most branching points are situated between 20 and 40 nucleotides from the 3' splicing site, and the sequence logos for all splicing motifs is highly conserved with plants and green algae examples (Szcześniak *et al.,* 2013). There are a few notable differences between *P. tricornutum* and *C. reinhardtii* splicing

sites, probably related to the differences in GC content. The polypyrimidine tract in the green alga is mainly cytosines whereas in plants and *P. tricornutum* it is mostly based on thymidines. The third position at the 5' end of the intron slightly differs between *C. reinhardtii*, which has a clearly overrepresented G in this position, whereas it is generally A in plants and *P. tricornutum*.

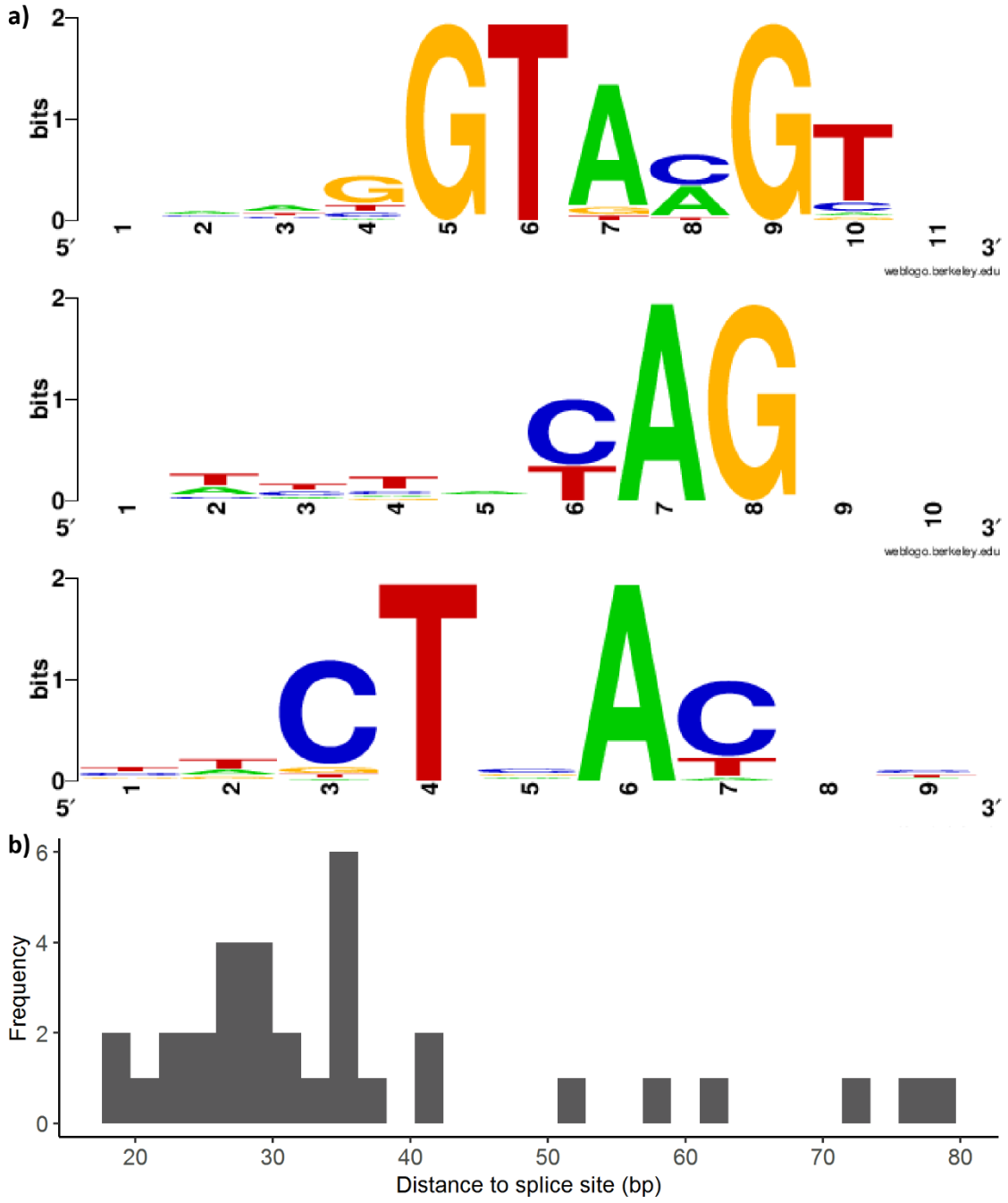


Figure 3.4 a) From top to bottom, sequence logos for the 5' and 3' splice site and the branching point in* P. tricornutum.** *The logos were built with the WebLogo online server with the alignment of the splicing sites found in 48 random introns in the* P. tricornutum *genome.* ***b) Distribution of distance between branching point and 3' splice site. *The branching site position was predicted for the different introns with the SVM-BP server. The distance distribution represents 33 out of the 48 introns that had an SVM score higher than 1.*

These data was then used, together with information from the *AtTHIC* alternative splicing model, to investigate the *PtTHIC* 3'UTR. First, splicing sites, branching sites and pyrimidine tracts matching the consensus were searched for in the *PtTHIC* 3'UTR. Two potential splicing donors (GU) were identified at 85 and 95 bp downstream from the stop codon and overlapping with the aptamer P1 and P2 stems 5' arms. There was only one potential branch site identified (CTGAT), 185 bp downstream of the stop codon, overlapping the P4-5 stem, but without a clear pyrimidine tract or 3' splice site following it. Finally, all the available EST in the Ensembl genome browser (Maheswari *et al.,* 2008) were analysed and no evidence of splicing in the *PtTHIC* 3'UTR was found (Figure 3.5). It is worth mentioning that, whilst the paper does not explicitly state the growth conditions, the norm for algal growth media is to include vitamins such as thiamine, and yet still no alternative splicing was observed.

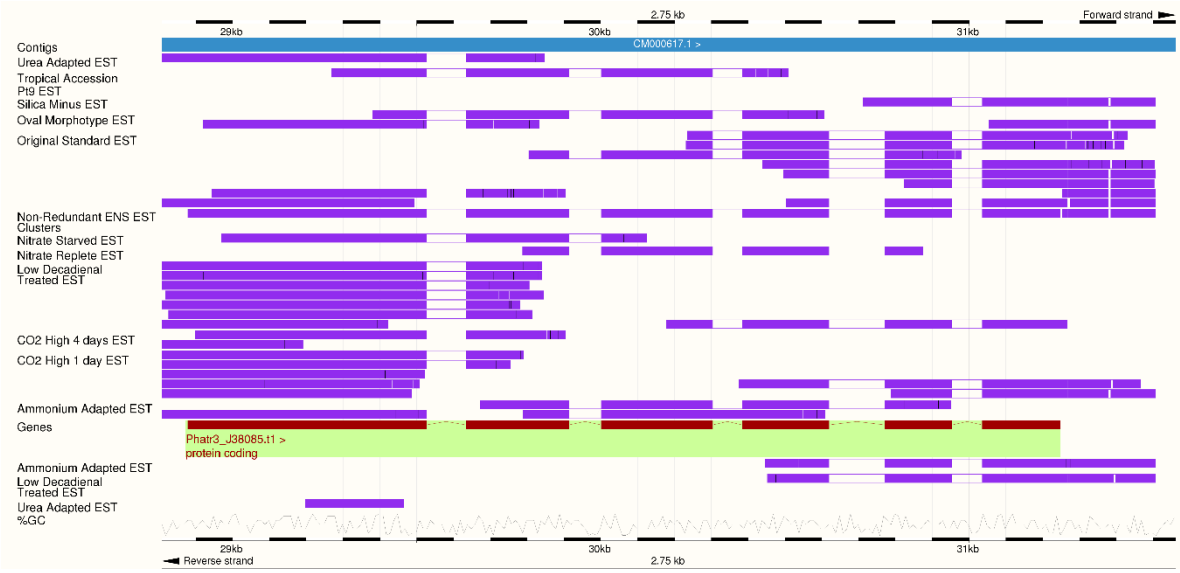


Figure 3.5. EST reads associated with* PtTHIC *(Phatr3_J38085). *The red boxes represent the 6 exons of* PtTHIC *with their 5 corresponding introns represented by lines. The EST reads are represented by the purple boxes above (sense) and below (antisense) the gene model. The 3'UTR is not annotated in the gene model but can be clearly identified with the majority of EST reads terminating 252 bp downstream of the stop codon.*

A second search was for the polyadenylation signal that marks the end of the 3'UTR, the termination of transcription and the start of the polyadenine tail with the PASPA software (Ji *et al.,* 2015). The most likely polyadenylation signal in *PtTHIC* was predicted to be the AACAAA sequence 212 bp downstream from the stop codon. This motif is very close to the consensus AAUAAA found in many eukaryotes (Tian & Manley, 2017) and overlaps with the 3' strand of the TPP aptamer P1 stem, which is precisely the domain with the lowest folding energy of the TPP

aptamer (Anthony *et al.,* 2012). The EST data found in the Ensembl genome browser was consistent with this prediction and most ESTs terminated 252 bp downstream the stop codon and just 34 bp downstream of the AACAAA motif (Figure 3.5).

In summary, with the evidence from the splicing and polyadenylation site analysis, it does not seem likely that *PtTHIC* is regulated via alternative splicing mediated by the TPP aptamer. Instead, given the overlap between the P1 stem and the predicted polyadenylation signal, it would be more likely that *PtTHIC* is regulated by the TPP aptamer via a mechanism of alternative polyadenylation.

3.2.3 *PtSSSP in-silico* structural characterisation

PtTHIC is not the only gene containing a predicted TPP riboswitch in *P. tricornutum*. McRose *et al.* (2014) predicted the presence of a TPP aptamer in the 3'UTR of *PtSSSP* and proposed the gene encoded for a thiamine or some of its intermediates transporter. Given the presence of a TPP riboswitch in its 3'UTR, and the role of thiamine uptake in the regulation of thiamine metabolism, *PtSSSP* was characterised bioinformatically to test whether its proposed annotation was correct.

First, the *PtSSSP* peptide sequence was run through the TMHMM transmembrane domain prediction software and 12 transmembrane domains were predicted, which gives an initial indication that *PtSSSP* is in principle located in the membrane and could act as a transporter. Furthermore, to find whether *PtSSSP* could share sequence similarity with experimentally confirmed thiamine transporters *A. thaliana PUT3* (Martinis *et al.*, 2016)*, H. sapiens SLC19A3* (Rajgopal *et al.*, 2001) or *S. cerevisiae THI7* (Singleton, 1997), a BLASTP search using the *PtSSSP* peptide sequence as query was conducted against *A. thaliana, S. cerevisiae* and *H. sapiens* proteomes. There were no significant results for *A. thaliana* or *H. sapiens*, but *S. cerevisiae* gave a single hit, *Dur3p,* annotated as a polyamine transporter. It has been experimentally shown that *AtPUT3* can transport both thiamine and polyamines (Martinis *et al.*, 2016). Then, the search was inverted and the *AtPUT3, HsSLC19A3* and *ScTHI7* peptide sequences were used as a query in a BLASTP search against *P. tricornutum.* There were no significant hits for *HsSLC19A3* or *ScTHI7*, and only two significant hits for *A. thaliana*: *Phatr3_J11160* and *Phatr3_J9233*, both with PFAM domains for polyamine transport, but only 10 and 6 predicted transmembrane domains respectively, while *AtPUT3* and *PtSSSP* have 12.

Nonetheless, although there is no sequence similarity evidence that *PtSSSP* could act as a thiamine transporter, the similarity to a polyamine transporter, the presence of a TPP aptamer in the 3'UTR and the same number of transmembrane domains as a known thiamine transporter

indicated that it could indeed be a thiamine transporter. Based on the match in transmembrane domains, it was proposed to use *in silico* structural alignment to test whether PtSSSP was structurally similar to experimentally confirmed thiamine transporters, and thus could in principle share function. To the best of my knowledge, there are no experimentally obtained 3D structures for eukaryotic thiamine transporters and therefore PHYRE2 was used to model the 3D structure of PtSSSP, AtPUT3, HsSLC19A3 and ScTHI7, as well as an unrelated transmembrane transporter in *P. tricornutum*, Phatr3_J22921, as a negative control. EcBtuF crystal structure known to have cobalamin uptake function and share structure similarity with PtCBA1 was retrieved from PDB (Borths *et al.,* 2002) and the structure of PtCBA1, known to be involved in cobalamin uptake (See section 1.1.2; Bertrand *et al.,* 2012), was modelled with PHYRE2 to have a technical control for the structural alignment. The predicted 3D structure of all these proteins was aligned pairwise with the StrucAlign structure alignment tool from PDB and the most relevant results are shown in table 3.1. The structure alignment tool quantified the deviation between two 3D structures with the Root-Mean-Square Deviation (RMSD) value which provides an average of the distance between atom positions in angstroms. The lower the RMSD value, the more closely the two structures align.

Table 3.1. Structural alignment of different thiamine transporter proteins. *The 3D structures of PtSSSP, and experimentally described thiamine transporters AtPUT3, ScTHI7 and HsSLC1943, as well as positive control PtCBA1 and negative control Phatr3_J22921 were modelled with PHYRE2. PtCBA1 structure was structurally aligned to EcBtuF as a positive control of a known structural conservation. The thiamine transporters were aligned pairwise with the putative transporter PtSSSP. And the predicted structure of a transmembrane protein not involved in transport (Phatr3_J22921) was aligned with AtPUT3 as a negative control. The alignments were performed with the StrucAlign tool and the results are ordered from lowest to highest RMSD (molecular deviation between the two structures). Identity refers to the percentage of residues with perfect match in a sequence alignment, similarity refers to the percentage of residues with either perfect match or residue substitutions with similar biochemical characteristics.*

Structure 1	Structure 2	Alignment length	Gaps	Identity	Similarity	RMSD (Å)
PtCBA1	EcBtuF	257	28 (11 %)	7.39 %	18.29 %	2.95
PtSSSP	AtPUT3	393	75 (19 %)	5.34 %	16.79 %	3.67
ScTHI7	AtPUT3	392	69 (18 %)	7.40 %	17.86 %	4.50
PtSSSP	ScTHI7	485	151 (31%)	5.15 %	15.67 %	4.71
PtSSSP	HsSLC19A3	341	199 (58 %)	3.23 %	10.26 %	6.57
Pt J22921	AtPUT3	426	282 (66 %)	2.58 %	7.28 %	7.70

The technical control alignment between PtCBA1 and EcBtuF showed an RMSD value which is relatively low consistent with their structural and functional similarity. On the other side of the spectrum, the alignment between Phatr3_J22921 and AtPUT3, two proteins that do not share any known structural or functional similarities was used as a negative control and has a significantly higher RMSD value. The best PtSSSP alignment was with AtPUT3, with an RMSD value close to that of the positive control and lower than between two known thiamine transporters (ScTHI7 and AtPUT3). When visually inspecting the overlaid structures of PtSSSP and AtPUT3 there is an evident close match in transmembrane helices overposition and a significant conservation of the central channel of the transporter can be seen from the top view of the structural alignment, providing further indication of a potential shared function (Figure 3.6). In the case of the negative control structural alignment visualization (not shown), much of the structure could not be visually aligned and some of the transmembrane helices crossed the membrane in a different orientation to those of AtPUT3.

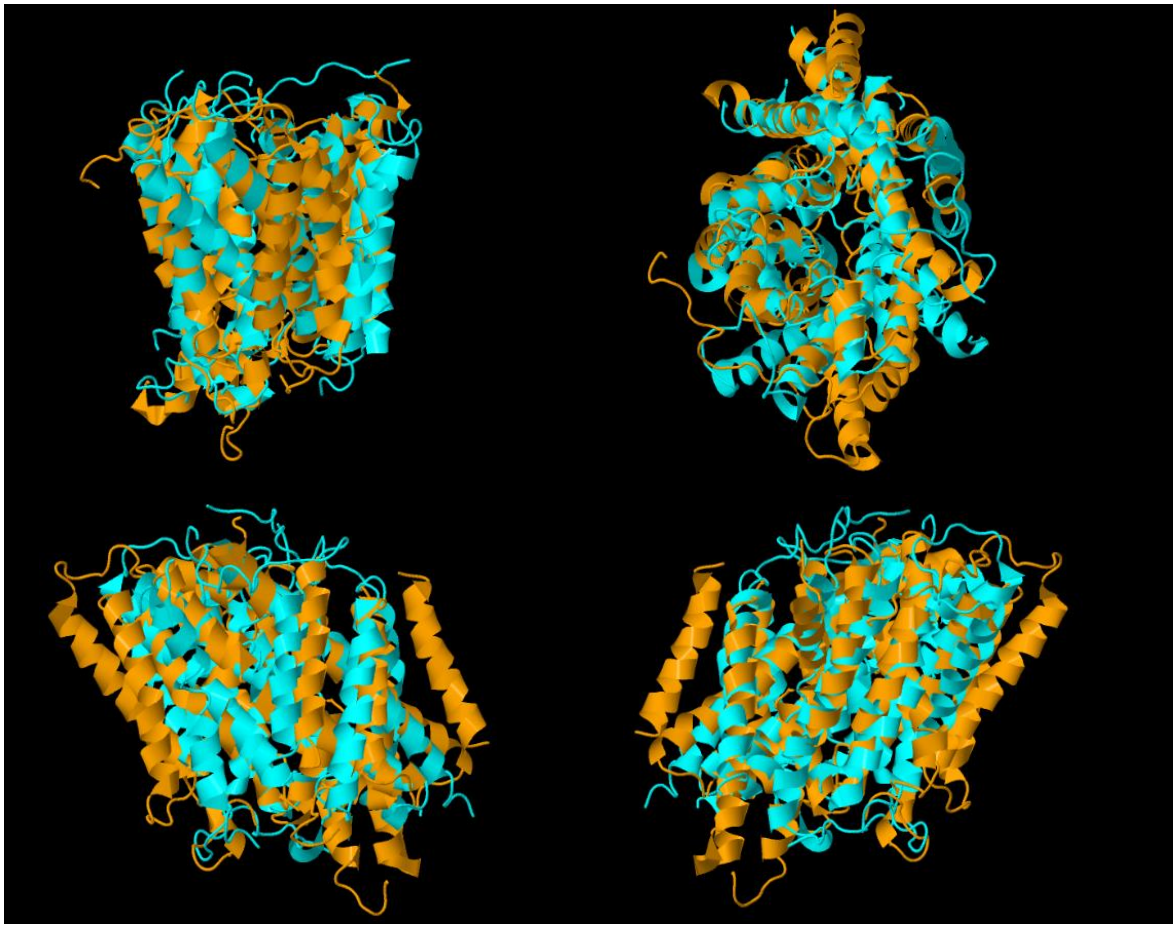


Figure 3.6. PtSSSP (blue) and AtPUT3 (orange) predicted 3D structures overlaid. *Four different views, front, top, right and left are shown in the figure. The images were generated with the Java applet with the structures predicted with PHYRE2.*

Whilst the sequence similarity information does not provide much information indicating a role in thiamine metabolism for *PtSSSP*, the structural similarity with known thiamine transporters and the presence of a predicted TPP aptamer in the 3'UTR of the gene are highly suggestive of it playing an important role in thiamine uptake. Later experiments were carried out to determine whether *P. tricornutum* can uptake thiamine (section 3.3.1), whether *PtSSSP* is exclusively responsible for it (section 5.6), and whether *PtSSSP* is regulated – either positively or negatively – by exogenous thiamine via its predicted TPP aptamer (section 3.3.2).

3.2.4 Identification of TPP aptamers in newly sequenced diatom genomes

In recent years numerous new diatom genomes have been published, including centric diatoms *Thalassiosira oceanica* (Lommer *et al.,* 2012), *Cyclotella cryptica* (Traller *et al.,* 2016), araphid pennate *Synedra acus* (Galachyants *et al,* 2015), and raphid pennates *Fistulifera solaris* (Tanaka *et al.,* 2015) and *Pseudo-nitzschia multistriata* (Basu *et al.,* 2017). These resources make it possible to search for riboswitches more widely in diatoms, and with an expanded diversity of sequence conservation data the opportunity to infer phylogenetic relationships and functionality arises.

The genomes of all diatoms available in the NCBI repository, including the above-mentioned species, were collected and with the help of Andre Holzer were assembled into a database that could then be BLAST-searched with various queries. First, the diatom genomes were queried with a BLAST search with parameters adjusted for short queries with the 26 bp long P4/5 stem from the *PtTHIC, T. pseudonana THIC* and *F. cylindrus THIC* aptamers. Five thousand basepairs upstream and downstream of every hit were extracted and immediate (<200 bp) upstream ORFs were annotated by pfam searches and TBLASTN searches against the NCBI nucleotide collection. Finally, the stems and conserved motifs of the predicted aptamers were annotated following the *PtTHIC* secondary structure template and with the help of the RNAfold Server (Hofacker, 2003). In parallel, *PtTHIC, PtNMT1* and *PtSSSP* peptide sequences were used in a TBLASTN search against the diatom database to find any other thiamine-related gene that was not captured with the aptamer query. The 5' and 3' UTR regions of these genes was then screened with the CTGAGA universally conserved ligand binding motif of the TPP aptamer to find potential riboswitches that had not been captured with the first search. With this pipeline all riboswitches already described in *P. tricornutum, T. pseudonana* and *F. cylindrus* and 11 new putative riboswitches could be identified, and the presence or absence of *THIC, NMT1* and *SSSP* in all available diatom genomes was determined (Figure 3.7a).

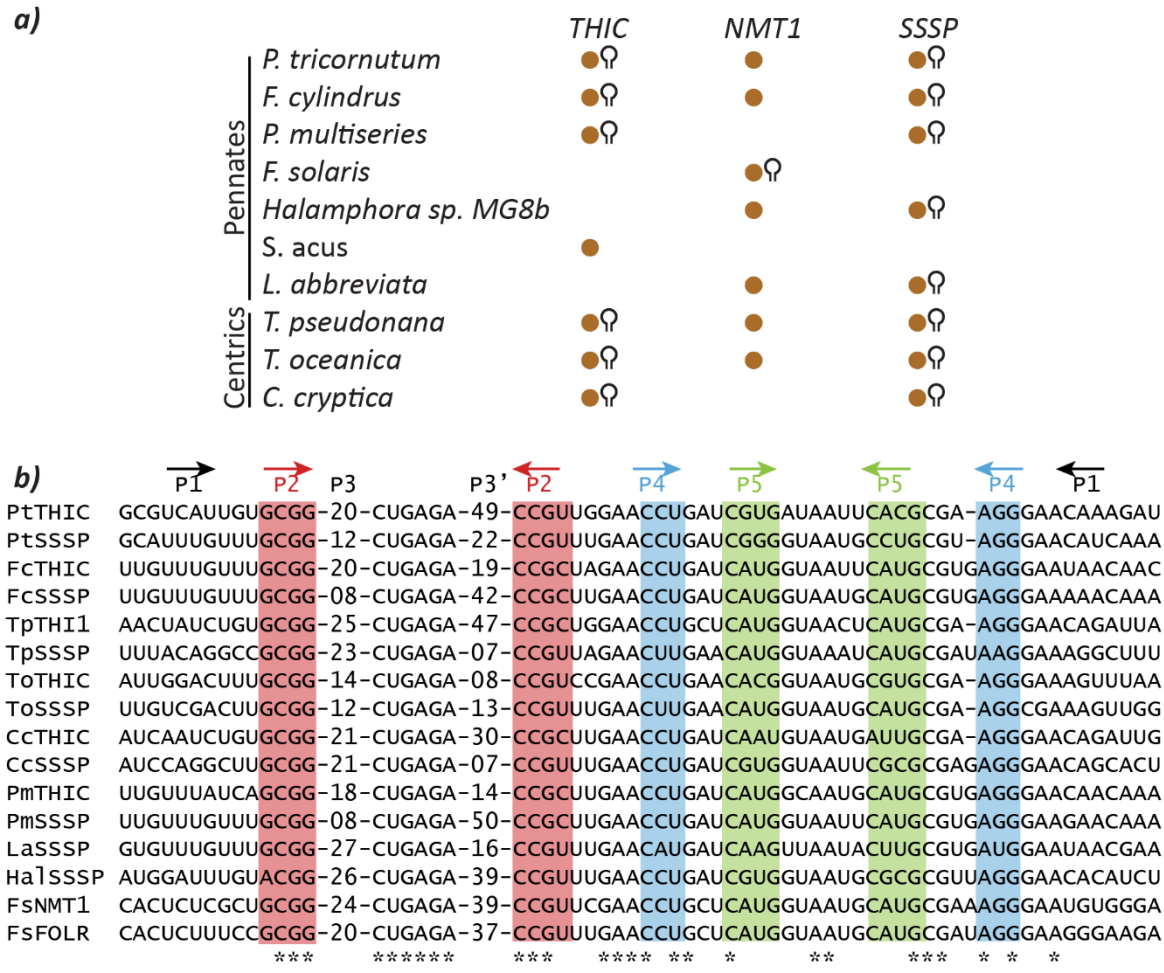


Figure 3.7. Identification of diatom predicted riboswitches in thiamine related genes. a) *Homology searches were used to identify the presence of THIC, NMT1 and SSSP in 9 available diatom genomes. A brown circle indicates presence of the gene in the genome, a stem next to the circle indicates the presence of a predicted TPP aptamer in the 3'UTR.* ***b)*** *MSA of all predicted TPP aptamers in diatoms. The shaded boxes mark conserved stems in the predicted secondary structure of the aptamers with its direction indicated by the arrows. P3 and P3' stems have variable lengths and sequences and have been summarised with its length in bp. P1 stems have variable lengths between all aptamers, and only the first 10 bp are shown to highlight their sequence conservation. Pt:* P. tricornutum*; Fc:* F. cylindrus*; Tp:* T. pseudonana*; To:* T. oceanica*; Cc:* C. cryptica*; Pm:* P. multistriata*; La:* L. abbreviata*; Hal:* Halamphora sp. MG8b*; Fs:* F. solaris*; Sr:* S. robusta

Four of the diatom species investigated encoded both a copy of *THIC* and *NMT1,* whilst the others had either THIC only (3 species) or NMT1 only (3 species). This evidence could support the hypothesis that both genes are active in the diatom lineage producing the same product, HMP-P, and that only one of the two genes would be sufficient to maintain thiamine biosynthesis levels. This hypothesis of course assumes that the observations are not biased by a potential lack of coverage in some species genomes that makes it impossible to find the genes. All the riboswitches found across the diatom lineage were in the 3'UTR of thiamine-related genes, mostly *THIC* and *SSSP,* and shared a very strong conservation indicating that the sequence reflects a defined and shared function related to thiamine metabolism. Exceptionally, riboswitches are also found in the 3'UTRs of *F. solaris NMT1* and *FOLR*, annotated as a folate receptor; this was also the case in *Micromonas pusilla* (McRose *et al.,* 2014). Only the *NMT1* gene in *F. solaris* was predicted to have a horizontal gene transfer origin among all the diatoms analysed, which could explain why it is the only species with a predicted riboswitch in its *NMT1* candidate (Vancaester *et al.,* 2020)

Turning to the predicted TPP aptamer sequences, all the putative diatom TPP riboswitches shared a very strong sequence conservation of stems P2, P4 and P5 as well as structurally conserving a P3' stem (Figure 3.7b). The 3' most strand of the predicted P1 stems were generally A-rich which could be indicative of a conservation of a polyadenylation site overlapping the P1 stem, supporting the hypothesis that the predicted aptamers work through a mechanism of alternative polyadenylation, as has already been discussed in detail for *PtTHIC* in section 3.2.2. Ideally, the 3'UTRs of all those diatom genes with a predicted TPP aptamer would have been analysed with the PASPA software (Ji *et al.,* 2015), but unfortunately the server is no longer available. The lengths and sequences of the predicted P3 and P3' stems were variable across different aptamers, lending support to the hypothesis that P3 and P3', being the two first stems to be fully transcribed and having the highest folding energy, act as a structural anchor that nucleates the folding of the rest of the aptamer (Anthony *et al.,* 2012).

As in previously experimentally investigated riboswitches in green algae, plants and fungi, the CUGAGA motif and overall secondary structure architecture was universally conserved in the putative diatom aptamers. The P4/P5 stem sequence was also well conserved between diatoms and green algae, plants, and fungal aptamers. However, whereas the P2 stem in green algae and plants has the AGGG sequence, which includes the alternative splicing acceptor (AG) used in the mechanism of action determined experimentally (Croft *et al.,* 2007, Wachter *et al.* 2007), the diatom predicted P2 stem has the sequence GCGG. Moreover, there was no other obvious AG splicing acceptor in the sequence of the aptamer conserved in diatoms, indicating that the

hypothetical molecular mechanism of the predicted riboswitches might differ from that in plants and green algae.

3.2.5 Identification of a conserved motif in diatom cobalamin-downregulated genes

Bertrand *et al.* (2012) studied to some extent the link between cobalamin and the C1 cycle with thiamine biosynthesis regulation. In particular, they found that the *PtTHIC* transcript and protein was downregulated by cobalamin supplementation. Research in our group, has determined that the *C. reinhardtii* and *P. tricornutum METE* promoters regulate gene expression in response to cobalamin (Helliwell *et al.,* 2014; Hickland *et al.,* in preparation). So, I wanted to investigate whether there is any conserved motif in the *THIC* promoter that could be linked to cobalamin regulation, and if so whether it is also present in *PtMETE.*

The diatom genomic dataset assembled in section 3.2.4 is of particular use when trying to unravel the molecular mechanisms that drive *THIC* downregulation in response to cobalamin. Working with the hypothesis that the cobalamin regulation on *THIC* was dependent on transcription factor regulation on the promoter region as for other cobalamin-regulated genes (Helliwell *et al.,* 2014; Hickland *et al.,* in preparation), the 500 bp immediately upstream of the *THIC* start codon in *P. tricornutum, F. cylindrus, P. multistrata, T. pseudonana, T. oceanica* and *C. cryptica* were assembled and run through the MEME software (Bailey *et al.,* 2009) with "zero or one occurrence per sequence" and other default parameters. The hit with the lowest e-value (6.7e+002) was 5'-KGAAKYACGTKCWKC-3' (where K stands for G or T, Y for C or T, and W for A or T following the standard IUPAC nomenclature) and was present in all 6 gene promoters in single copy. In order to validate the motif, the upstream region of a set of genes known to be downregulated by cobalamin from transcriptomic and proteomic data (Bertrand *et al.,* 2012) was searched with the consensus sequence. In parallel, a negative control test with a set of housekeeping genes not affected by cobalamin was also searched with the motif. The motif was found in 6 of the 7 upstream regions of the genes known to be downregulated by cobalamin, including 4 copies of the motif in *PtMETE* and genes in *T. pseudonana* and *F. cylindrus*. As expected, no copies of the motif could be found on any of the 8 negative controls (Table 3.2).

The motifs for the 6 original and 9 newly identified sequences were manually refined by removing the original first position (5' – **K**GAA…) given it contained very little information, and were used to build a sequence logo with the consensus sequence 5'- GAAGCACGTGCTTC -3' (Figure 3.8). It is notable that the consensus is a 14 bp palindrome, which is a widespread characteristic of transcription factor binding sites (Cox *et al.,* 2019). This important observation gives strength to the hypothesis that *THIC*, and all other cobalamin regulated genes, are

regulated by a unique transcription factor that recognises the motif in their promoter region. The palindromic nature of the motif suggests that the transcription factor could act as a homodimer, with each monomer binding each half of the motif in opposite directions. A Hidden Markov Model (HMM) was built with the 15 motif copies including the 20 bp regions flanking the central 14 bp palindrome by Dr Andrew Sayer. The HMM was used as a query to search against a database of 1000 bp fragments upstream of all known *P. tricornutum* ORFs assembled by Andre Holzer. Among the 15 top hits, 10 were for genes already known to be downregulated by cobalamin (*METE*, *CBA1* and *THIC*), and the other 5 hits corresponded to genes newly identified: *Phatr3_EG00541, Phatr3_J47571, Phatr3_J44048, Phatr3_J21868 and Phatr3_EG01741.* No annotation or associated pfam domains are known for *Phatr3_EG00541* or *Phatr3_J47571,* whilst *Phatr3_J44048* is annotated as a histidine phosphatase, with no known relation to cobalamin regulation. *Phatr3_J21868* is annotated as a GTP cyclohydrolase in the THF biosynthetic pathway with clear links to cobalamin given that 5-methyl-THF is the methyl donor to *METE* and this gene appears to be downregulated by cobalamin in the Bertrand *et al.* (2012) proteomic dataset. *Phatr3_EG01741* codes for a methyltransferase, an enzyme type which usually uses the C1 cycle intermediate SAM as a methyl donor. Different independent strands of evidence thus give support to suggest that the motif identified in diatom *THIC* genes is involved in the molecular mechanism by which cobalamin supplementation downregulates gene expression. In *C. reinhardtii* on the other hand, although *CrMETE* is also downregulated by cobalamin (Croft *et al.,* 2005), *CrTHIC* is not regulated by cobalamin according to various transcriptomic datasets obtained in the lab (Holzer *et al.,* in preparation; Sayer *et al.,* in preparation). The promoter region of *CrTHIC* was searched with the consensus sequence but no similar motif could be found, leading to the conclusion that the motif is not widely conserved across the eukaryotic tree of life.

Figure 3.8. Sequence logo for the conserved motif in diatom promoters downregulated by cobalamin. *The sequence logo was built with WebLogo using the sequences of the 6 diatom* THIC *promoter motifs and the 11 sequences of the motifs found in 8 other genes downregulated by cobalamin supplementation.*

With the database of 1000 bp fragments upstream of all ORFs in *P. tricornutum* genome now available, there is the opportunity to take an alternative approach to find a motif associated with cobalamin downregulation. The 1000 bp fragments upstream of the 32 genes downregulated by cobalamin in Bertrand *et al.* (2012) were selected and run on MEME with the "Any Number of Repetition" option and restricting the search to 25 bp long motifs. The second best hit with an e-value of 5e-10 followed the consensus 5' TMACWGTMA 3' and was found 39 times in 22 of the genes, including *PtMETE* and *PtTHIC.* This motif has less information than the first motif found but it is also a palindrome leading to the hypothesis that it could also be bound by a transcription factor dimer.

To investigate further the molecular mechanism by which the predicted sequences might downregulate genetic expression, the HMM of the motifs were used as a query against the JASPAR transcription factor binding site database selecting plants, fungi and protist groups (Fornes *et al.,* 2020). For the first 14 bp long motif (GAAKYACGTKCWKC), most of the higher-scoring hits showed the central conserved motif CACGTG, commonly known as E-box, and were bound by transcription factors dimers in the Myc-like family containing a basic Helix-Turn-Helix domain (bHLH). It is common for transcription factors to regulate the expression of their own gene in a feedback loop mechanism, and the presence of the binding motif in the promoter region of the gene is used as an indication to select candidate transcription factors. In Bertrand *et al.* (2012) dataset, there is a gene annotated as Myb-like transcription factor, *Phatr3_J32183*, that is upregulated by cobalamin supplementation and shows a motif in its upstream region which is mildly conserved with the consensus motif (5' – GTCGCACGTTCTTC – 3'). This evidence is not enough to narrow down a transcription factor counterpart to the cobalamin motif but provides a promising candidate to investigate experimentally. The HMM model for the second 9 bp long motif (TMACWGTMA) gave no significant hits and no similar sequences could be found in the upstream regions of the other diatom *THIC* genes.

Table 3.2. Summary of motifs found in diatom* THIC *genes and genes downregulated by cobalamin supplementation. *The position denotes the distance between the 3'end of the motif and the start codon of the gene. The genes which promoter region was used to find the motif with the MEME software are labelled as "Train", those that have been experimentally confirmed to be downregulated by cobalamin supplementation and have not been used in the MEME search are labelled "Test", and those not known to be downregulated by B_{12} are labelled "Negative Control". The 5' and 3' flanking sequences for the motif are included.*

Gene	Position		5' (21nt)	Motif (14nt)	3' (20nt)
PtTHIC	-284	Train	AACGGCAAAACGAATCTTTGG	GAATCACGTGCTTC	GAAGAATGCGAAAAGATCTC
FcTHIC	-186	Train	CTTTTTTCGCTTACTACAGTT	GCAGTACGTTCAGC	CAATCAGCGTGGGAGAAGGC
PmTHIC	-384	Train	TTTACTGAAACACGAGCAAAT	GAAGTACGTACTTC	AGAGAACGAAACTGATGGTT
TpTHIC	-235	Train	CTTCGAAACCACCTAGTTCCT	GAAGCACGTGCATC	ACGACATCCTCCACTTCGCA
CcTHIC	-274	Train	TAATGGTATCCATAAAATGAT	GTATCACGTTCGTC	GCTCGTTGAGGTTGATTAGA
ToTHIC	-144	Train	CGTCGGCGGTCTTCACATCGG	CAATCTCGCGCAGC	GACAATGCCGACATCACTCC
PtMETE1	-395	Test	TTGGAAAAAGAGAACATGTTC	AAATCACGTTCTTC	GCTAAGGTTGTTTCAAGAGC
PtMETE2	-133	Test	CGGGGAGCTCAATATCAGTTC	AAAGCACGTTCTTC	ATAGACCAGTGAATGAACTG
PtMETE3	-485	Test	AATTCGCGCCATCGGAACGGC	AAAACACGTGCTTC	CCGATATCTCCGGAATGGAA
PtMETE4	-262	Test	GCTGCCGTCCAAGGATATCAC	AAAGCACGTTCTAC	AGAAATGTAGTTTCAGAGGG
FkMETE	-273	Test	CGGTATGTCCGACCATATCTC	GAATCACGTTCCTC	GACATGTATCGATGGTAAAG
FcMETE	-404	Test	GCTTTTCGCCGCGCCATGTCA	GAATCACGTTCTTC	AAACTAGTTTTTTTAAATCT
PtCBA1	-263	Test	TCGATCGTGGATATTGAGCAC	GAAGCACGTGATTC	GCAATCCGGACGGGGCCGAG
PtSHMT1	-75	Test	TGGAACGTGCCAGCGAAAACG	TCTGCTCGTGCTTC	GCTCGGCCCATGGAGTAACG
Tp42612	-157	Test	CATTAGTGACAAAGTGCCAGA	CCATCACGTGCTAC	AACCACAACCGTCCGTGCAA
Tp22483		Test	Could not find		
PtUBQo		NegCtrl	Could not find		
PtH4		NegCtrl	Could not find		
PtGGPPS		NegCtrl	Could not find		
PtPSY		NegCtrl	Could not find		
PtPXD1		NegCtrl	Could not find		
PtUMPS		NegCtrl	Could not find		
PtPXD2		NegCtrl	Could not find		
PtMETH		NegCtrl	Could not find		

3.3 Molecular biology assays on WT strains

3.3.1 HPLC analysis of thiamine uptake

Having carried out a thorough bioinformatics study, there was enough information to consider testing experimentally the role and mechanism of the putative *PtTHIC* riboswitch. As a first step, in order to test whether exogenous thiamine is taken up by diatoms and thus exogenous thiamine supplementation can have an impact on the thiamine biosynthesis pathway regulation, an uptake assay by HPLC was performed. Thiamine content HPLC measurements have been previously performed in our research group to study the intracellular thiamine levels in different *C. reinhardtii* mutants and the riboswitch mediated response to thiamine supplementation confirms thiamine is taken up by this green alga (Moulin *et al.,* 2013).

Analysis of cultures grown at different thiamine concentrations showed that increasing levels of thiamine in the media contributed to significantly higher levels of intracellular thiamine both in *C. reinhardtii* and *P. tricornutum* (Figure 3.9). Under control conditions without any thiamine, the levels of intercellular thiamine were very low and increased significantly upon addition of extracellular thiamine. Intracellular TPP levels saw a 3-fold increase upon addition of 10 µM thiamine in the media, while higher levels of extracellular thiamine (100 and 1000 µM) increased TPP levels further in *C. reinhardtii*. Notably, *P. tricornutum* TPP levels in the absence of supplementation were similar to those in *C. reinhardtii* under 10 µM thiamine. *P. tricornutum* TPP levels increased 3-fold under 10 µM thiamine, indicating that exogenous thiamine is taken up and metabolised into TPP, and TPP levels further increased under 100 µM thiamine, but then saturated and slightly descended under 1000 µM.

Further evidence of thiamine uptake was provided by analysing the levels of thiamine in the media. After 6 days of culture, the media of a culture supplemented with 10 µM thiamine had no significant difference in thiamine levels compared to an unsupplemented culture control, while a control culture that was supplemented with 10 µM thiamine just before harvesting had 18 times higher thiamine concentration than the one cultured for 6 days. These results demonstrate that thiamine is taken up into the cell and is metabolised to TPP thus ruling out that the increase in thiamine levels is simply a result of adhesion to the cell surface.

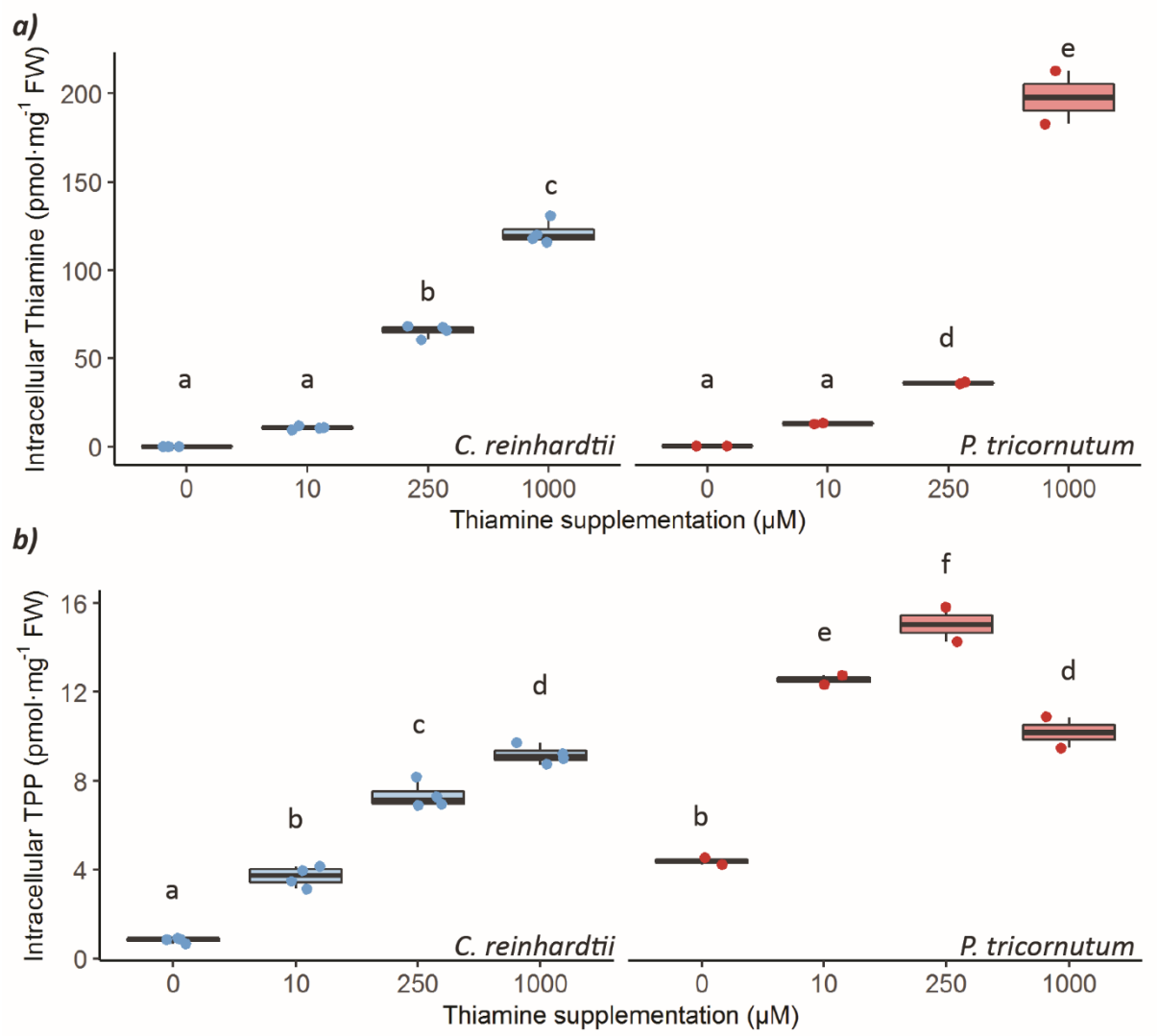


Figure 3.9. Thiamine (a) and TPP (b) intracellular levels in* C. reinhardtii *(Blue) and* P. tricornutum *(Red) under different levels of extracellular thiamine*. Cells were cultured under increasing concentrations of thiamine and the intracellular thiamine and TPP levels were quantified by HPLC and normalised by fresh weight. Two technical replicates were measured for* P. tricornutum *and two biological and two technical replicates for* C. reinhardtii*, each replicate is indicated by a single dot in the plot overlaid on a box plot summarising the data. Treatments with statistically undistinguishable measurements share the same letter according to a Tukey HSD test with 0.95 confidence level.*

3.3.2 Quantifying *THIC* transcript levels in response to vitamin treatments

A pilot semiquantitative RT-PCR experiment was used to test whether thiamine supplementation changed *PtTHIC* transcript levels as would be predicted by the presence of an aptamer in the 3'UTR similar to *AtTHIC* (Wachter *et al.* 2007). *P. tricornutum* 20 ml cultures were grown in f/2 to late exponential phase either without or with 10 µM thiamine, then RNA was extracted and used to make cDNA. The semiquantitative RT-PCR used primers targeting *PtTHIC* and the housekeeping *PtH4* encoding histone H4, the products were run in an agarose gel, and the intensity of the bands was plotted against the cycle number (Figure 3.10). It was calculated that under thiamine supplementation *PtTHIC* transcript levels decreased 13 % compared to the control. There were no biological repeats, so it was not possible to conduct a statistical test on this result, which nevertheless does not show a big difference between the conditions. In contrast there is almost complete repression of THIC in *C. reinhardtii* grown under the same conditions (Croft *et al.,* 2007).

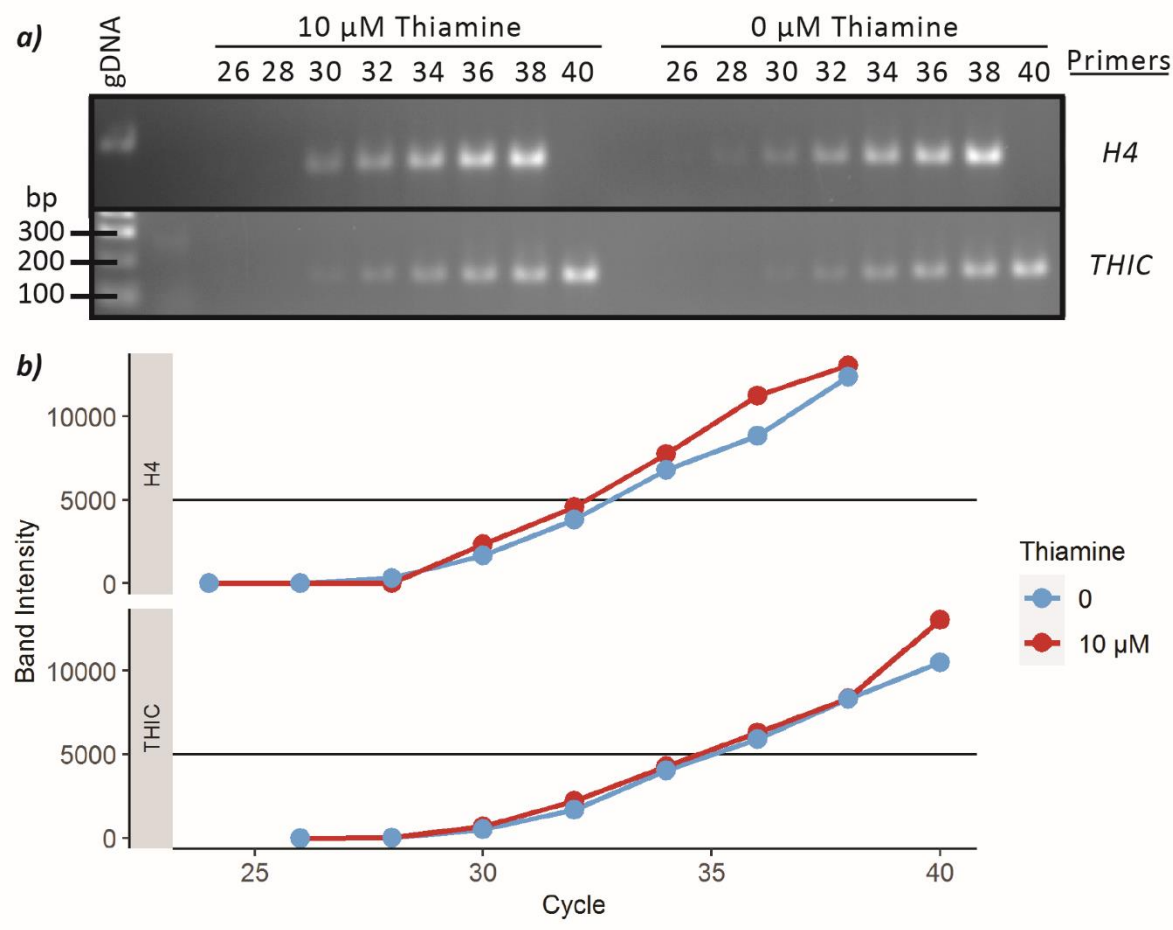


Figure 3.10. Semiquantitative RT-PCR on* PtTHIC *in response to thiamine supplementation. a) *A PHUSION PCR with primers targeting the CDS of* PtTHIC *and* PtHistone4 *(*H4*) was conducted on cDNA samples obtained from* P. tricornutum *cultures grown in the presence or absence of 10 µM thiamine. Ten microliter reactions for each of the indicated cycle numbers were carried out and run in a 0.5 % agarose gel.* ***b)*** *The intensity of each band was measured with ImageJ and plotted against its corresponding cycle number for each primer pair. A threshold intensity was set at 5000 units to cross the amplification curves at mid-exponential phase and set a corresponding Ct value for each primer pair and condition. The amplification efficiency was calculated by an exponential fit on the exponential phase of each primer pair. The Ct value and exponential efficiency were used to later calculate the fold-change in* PtTHIC *expression between the test and control condition.*

In order to have more statistical power, quantifiable data and the ability to screen more genes in more species and conditions, a Reverse Transcriptase-quantitative PCR (RT-qPCR) experiment was undertaken. *P. tricornutum* and *T. pseudonana* were grown in biological triplicate in 20 ml cultures with different supplementations. *P. tricornutum* was tested under 10 µM thiamine, 10 µM HMP and 1 $\mu g \cdot l^{-1}$ cobalamin supplementation independently against a no supplementation control. In *T. pseudonana* only the thiamine and control conditions were tested. Cells were harvested at late exponential phase, RNA was extracted and used to make cDNA. The cDNA was used as template in RT-qPCR reactions in technical duplicate with primers targeting thiamine metabolism genes *THIC, SSSP, NMT1,* and *THIG* for thiamine and HMP supplementations and *THIC, SSSP* and *METE* for the cobalamin conditions. Housekeeping genes Histone4 (*H4*), Ubiquitin Conjugating Enzyme (*UBC*) and Ubiquitin (*UBQ*) for *P. tricornutum* and Actin, Elongation Factor 1a (*EF1a*), and RuBisCO small subunit (*RBCS*) for *T. pseudonana* were used to normalise the abundance of each transcript using a modified DeltaDeltaCt method taking into account the individual amplification efficiency for each primer and sample combination (See section 2.3.5 for details; Rao *et al.,* 2013).

From the results in Figure 3.11, in response to thiamine supplementation, significant differences in transcript levels were observed only for *PtTHIG*, which showed a nearly 6-fold upregulation. However, this change was not observed in response to HMP supplementation (Figure 3.11a&b). Thiamine or HMP did not induce a change in transcript level for *THIC, SSSP* or *NMT1* in *P. tricornutum*. In *T. pseudonana*, thiamine supplementation caused a 2-fold statistically significant downregulation in *TpSSSP,* but *TpTHIC* and *TpNMT1* remained unchanged, and unlike *PtTHIG, TpTHIG* did not show any changes upon thiamine supplementation (Figure 3.11c). Upon cobalamin supplementation *PtMETE* was downregulated 135-fold and *PtTHIC* was downregulated 6-fold (not statistically significant), and no changes were observed for *PtSSSP* (Figure 3.11d). The absence of a significant change in *PtTHIC* levels in response to thiamine and HMP supplementation was consistent with the preliminary observations in the semiquantitative RT-PCR experiment and indicated that the predicted TPP riboswitch does not regulate expression at the transcript level under these conditions. Although the 6-fold downregulation of *PtTHIC* in response to cobalamin supplementation was not statistically significant given the low number of replicates, previous transcriptomic and proteomic data showed that *PtTHIC* is downregulated under cobalamin replete conditions (Bertrand *et al.,* 2012). The fact that both *PtTHIC* and *PtSSSP* have predicted TPP aptamers in their 3'UTR, but only *PtTHIC* appears to be downregulated by cobalamin would suggest that the predicted aptamer is not responsible for the cobalamin response, but probably the regulation is at the promoter region in agreement with the finding of a cobalamin-related motif in the *PtTHIC* promoter (section 3.2.5).

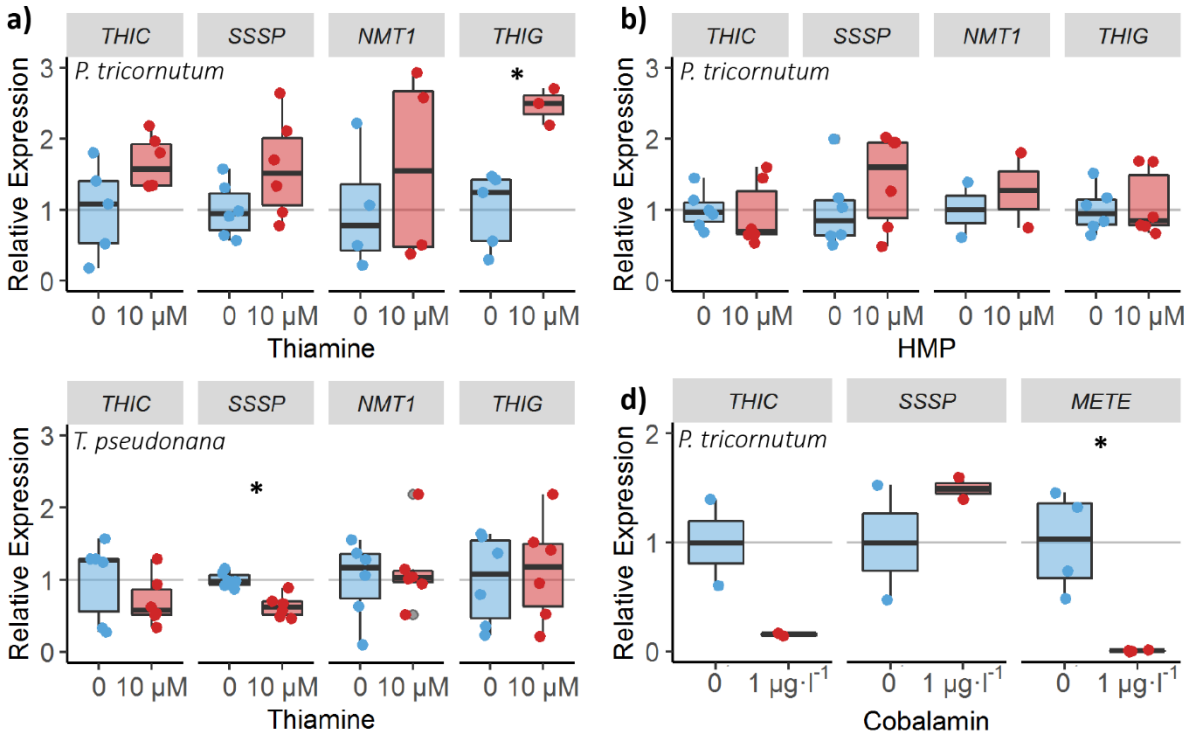


Figure 3.11. RT-qPCR on* P. tricornutum *and* T. pseudonana *thiamine related genes under different conditions. P. tricornutum *cultures were grown in the presence or absence of 10 µM thiamine* ***(a)****, 10 µM HMP* ***(b)****, 1 µg·l^{-1} cobalamin* ***(d)*** *or* T. pseudonana *cultures in the presence or absence of 10 µM thiamine* ***(c)****. All experiments are done in biological triplicates and technical duplicates with all genes being normalised to three house keeping genes, only those reactions finally used to calculate transcript levels (see methods, section 2.3.5) are shown as individual data points, box plots summarise the distribution of the sample. An asterisk incidcates where the relative expression was significantly different between test and control conditions under a two-sided Welch T-test with a 0.95 confidence interval.*

3.3.3 3'RACE

Semiquantitative RT-PCR and RT-qPCR experiments showed no significant changes in *PtTHIC* transcript levels upon thiamine or HMP supplementation (Figure 3.10 & 3.11). These results raise the question of whether the predicted TPP aptamer could regulate gene expression through a post transcriptional modification, including a polyadenylation event as proposed in section 3.2.2. Before the start of my thesis, preliminary 3'RACE experiments by Dr Mark Scaife in the lab showed that *PtTHIC* had a 253 bp 3'UTR in the absence of thiamine and it included an additional 588 bp 3'UTR isoform in the presence of thiamine. To test whether *PtTHIC* could have post-transcriptional modification in the 3'UTR, and to test whether the preliminary results by Dr Mark Scaife could be corroborated, a 3'RACE experiment was carried out on *PtTHIC* under thiamine or HMP supplementation compared to control conditions.

P. tricornutum was grown in biological triplicate in 20 ml cultures supplemented with 10 µM thiamine or 10 µM HMP, in parallel with a no supplement control. Cells were harvested at late exponential phase and RNA was extracted. cDNA was made following a modified TVN reverse transcription method by which a polyA primer with a 5' universal adaptor and a 3' TVN anchor (5' – TTGAAGACGTAGCGTAATGAACATAGTTTTTTTTTTTTVN – 3') primed the reverse transcription of all mature transcripts at the 5' end of the polyadenine tail (Lee *et al.,* 2014). To determine the length of the *PtTHIC* transcript, a PCR with a gene-specific forward primer annealing on *PtTHIC* 6th exon and a reverse primer annealing on the universal adaptor at the 3' end of the cDNA was performed. Given that the reverse primer annealed at an adaptor present in all transcripts, the first PCR was not specific enough. Hence, the PCR product was used as template in a second PCR with a nested forward primer specific to *PtTHIC* and a reverse primer on the adaptor to improve the assay specificity. The PCR product of the semi-nested PCR was run on an agarose gel to determine the length and number of *PtTHIC* 3'UTR isoforms. Some bands were cut and sent for sequencing to determine the exact position of the transcription termination site and the precise length of the 3'UTR.

The results show that there are no significant differences in the 3'UTR length of *PtTHIC* in response to thiamine or HMP supplementation (Figure 3.12b&c). The most intense and defined band for all the samples was a 347 bp fragment that was sequenced in at least one biological replicate for each condition and it precisely matched the transcription termination site that most EST reads share and that is only 34 bp downstream of the AACAAA motif overlapping the aptamer P1 stem hypothesised to play a role in the molecular mechanism in section 3.2.2. In addition, in all conditions a less intense, shorter band at 126 bp was observed, which terminated at an A-rich region of the *PtTHIC* 3'UTR (5' – AAAATAGA – 3') that could have produced a false positive amplification with the TVN method. At this point, the 3'RACE data obtained by Dr Mark Scaife previously in our research group and based on which the hypothesis of alternative polyadenylation was partly formulated, was re-examined. It was

found that the longer isoform of the 3'UTR matched an A-rich region 587 bp downstream the *PtTHIC* stop codon (5' – AGGAAAAGAACAAA – 3') and could have been a product of unspecific amplification on genomic DNA contamination in the RNA sample. In order to control the false positive amplification effect, a genomic DNA sample should have been added to the experiment at the TVN reverse transcription step, but this had not been done in the experiment run by Dr Mark Scaife nor in the experiment I ran. The genomic DNA controls in Figure 3.12 are added just as a template in the semi-nested PCR, and thus they do not control for unspecific amplification from the TVN primer. In the HMP 3'RACE on genomic DNA two bands could be seen, which could be explained by contamination from earlier PCR products.

For both thiamine and control conditions, but not in the HMP experiment, there was a higher band which the sequencing data confirmed was longer than the most intense 347 bp fragment. Unfortunately, only 37 bp beyond the transcription termination site could be sequenced and thus it is unknown where exactly the longer 3'UTR isoform terminates. There is no evidence in the EST datasets of *PtTHIC* 3'UTRs extending beyond the transcription termination site (Section 3.2.2), but it is possible that a few of the transcripts stochastically skip the canonical termination site and use an alternative one further downstream. In any case, this longer 3'UTR isoform appears in both treatment and control conditions and there does not seem to be observable differences in intensity in response to thiamine supplementation.

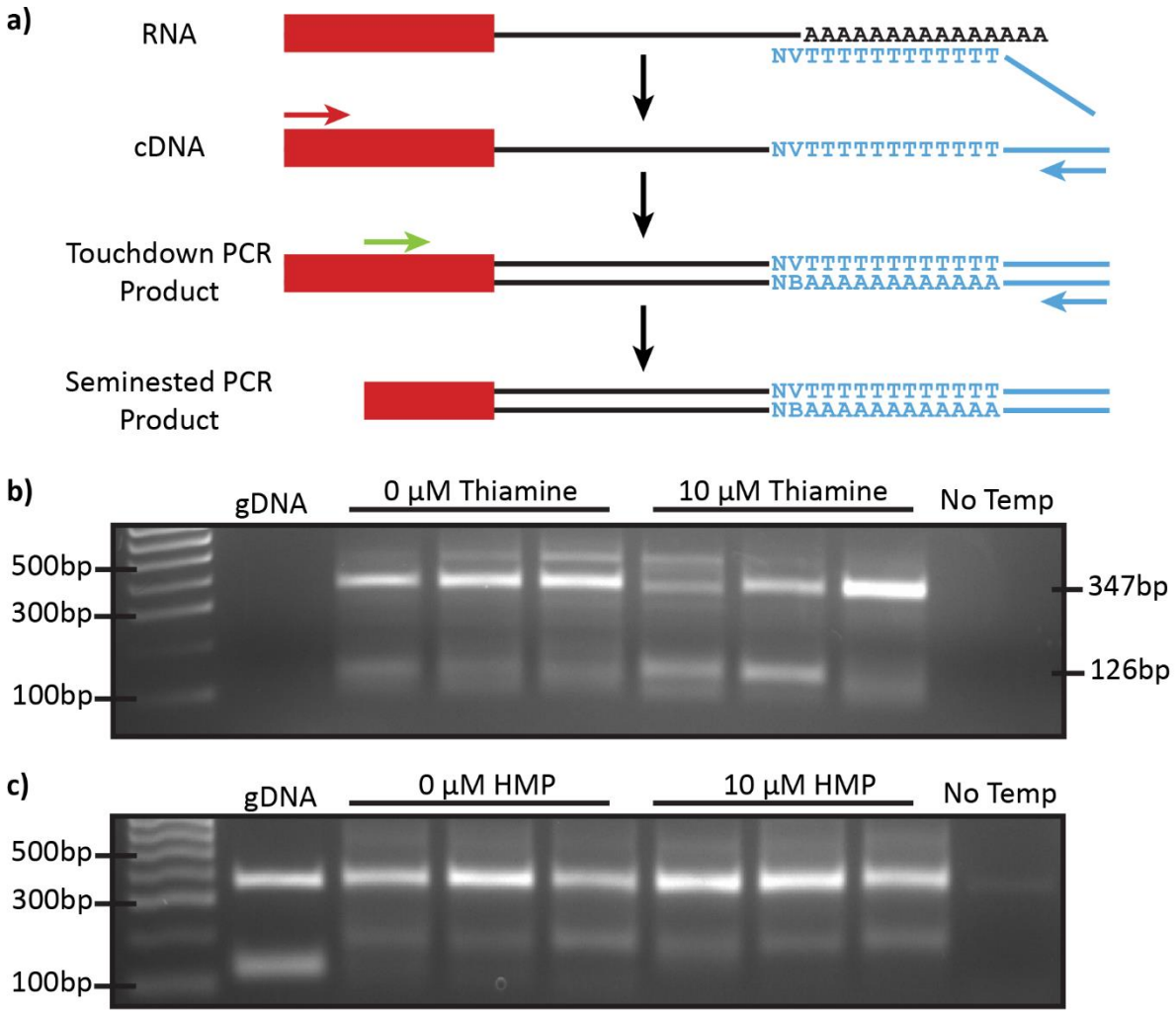


Figure 3.12. **PtTHIC** ***3'RACE RT-PCR products from*** **P. tricornutum** ***cells grown in the presence or absence of 10 µM thiamine (b) or HMP (c). a)*** *RNA isolated from* P. tricornutum *triplicate cultures in the presence or absence of 10 µM thiamine or HMP was retrotranscribed with a TVN primer binding the 5' start of the polyadenine tract of all messenger RNAs. A first touchdown PCR with a* PtTHIC *specific primer (red arrow) binding the CDS (red box) and a universal primer binding the adaptor included in the TVN primer (blue arrow) was followed by a semi-nested PCR with a second* PtTHIC *specific primer (green arrow) and the universal primer to increase the specificity of the 3'RACE. The products of this second PCR were run in a 0.5 % agarose gel.* ***b)*** *For the thiamine gel, a control with the second PCR reaction using genomic DNA (gDNA) as template was included.* ***c)*** *For the HMP gel, the genomic DNA template was included from the cDNA synthesis step.*

3.3.4 Pyrithiamine assays

Pyrithiamine is an antimetabolite, which has been shown to kill cells by competing with thiamine (Sudarsan *et al.,* 2005). In fact, whilst it can inhibit the activity of TPP-dependent enzymes essential for cell metabolism, the cytotoxic effect is manifested mainly at the level of gene expression. A functional TPP aptamer binds pyrithiamine and downregulates the expression of thiamine biosynthesis genes, in turn shutting down the production of thiamine. Supplementation with exogenous thiamine can rescue the effect of pyrithiamine (Woolley *et al.,* 1943). Thus, if pyrithiamine supplementation is lethal for an organism this suggests that its thiamine biosynthesis pathway is regulated by TPP riboswitches. Previous experiments in the lab have shown that *C. reinhardtii* does not survive pyrithiamine treatment unless it is supplemented with exogenous thiamine (Balia Yusof Thesis, 2011). *C. reinhardtii* can also be rescued with HET, since pyrithiamine degrades to release HMP which can be combined with HET to form thiamine even in the absence of functional pyrimidine and thiazole biosynthesis (Balia Yusof Thesis, 2011; Ginnie Nguyen Thesis, 2016). Accordingly, it was decided to test the effect of this antimetabolite on diatoms.

C. reinhardtii, P. tricornutum and *T. pseudonana* were grown in biological triplicate in 96-well plates under the presence or absence of 10 µM pyrithiamine and/or 10 µM thiamine, HET or HMP for 7 days recording OD_{730nm} as a proxy for growth every 24h. *C. reinhardtii* growth was disrupted by pyrithiamine, but normal growth could be rescued by thiamine and/or HET supplementation (Figure 3.13a). *P. tricornutum* and *T. pseudonana* were completely insensitive to the antimetabolite (Figure 3.13b&c). These results provide further evidence to support the hypothesis that the studied riboswitches in diatoms do not mediate a regulatory response to increased thiamine levels.

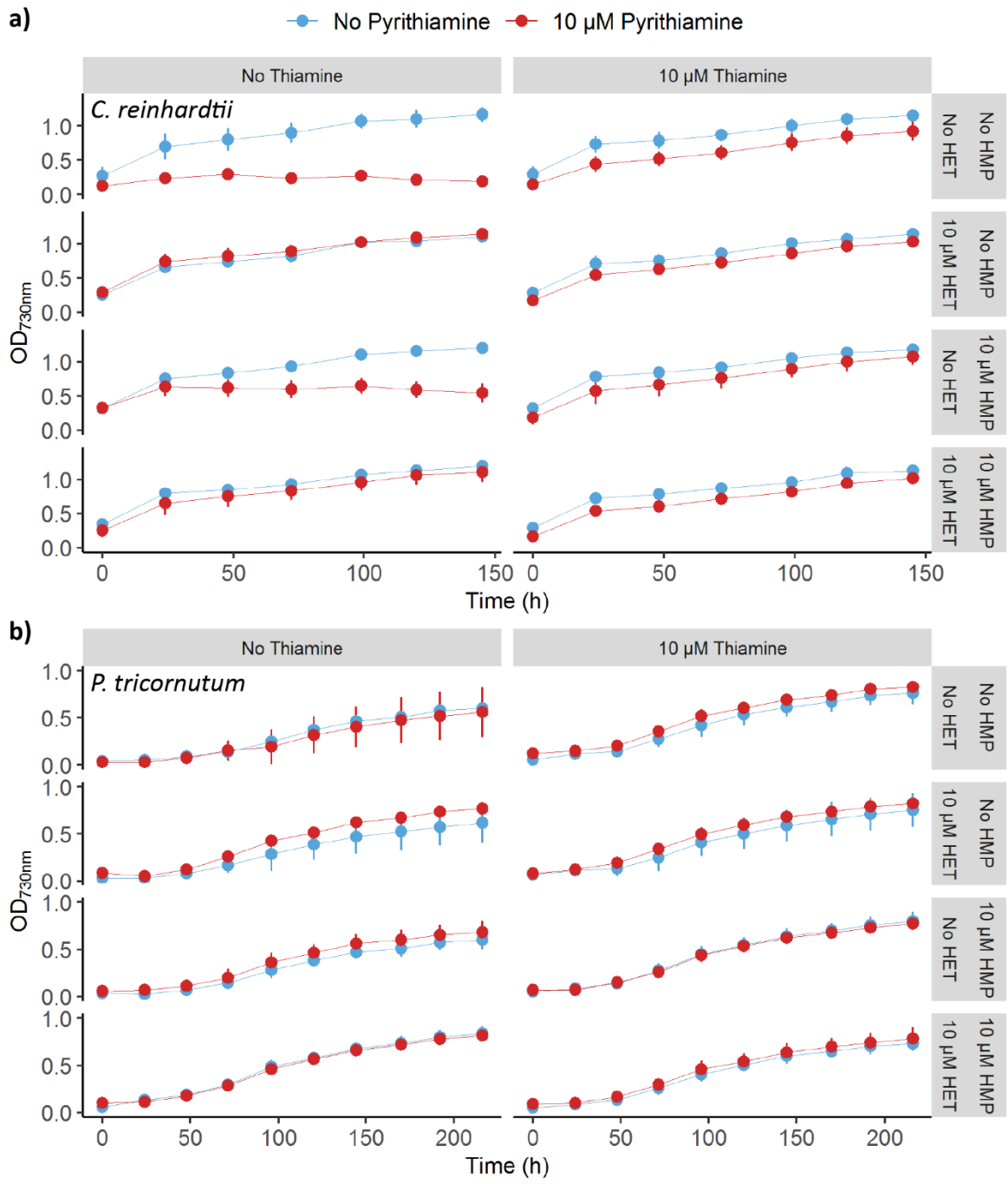


Figure 3.13. Pyrithiamine growth assays on* C. reinhardtii *(a),* P. tricornutum *(b) and* T. pseudonana *(c). *Triplicate cultures for each of the species was grown in the absence (blue) or presence (red) of 10 µM pyrithiamine, and in the presence (right) or absence (left) of 10 µM thiamine in 200 µl volume in 96-well plates. OD_{730nm} was measured with a plate reader every 24h to track growth.* ***Continues next page.***

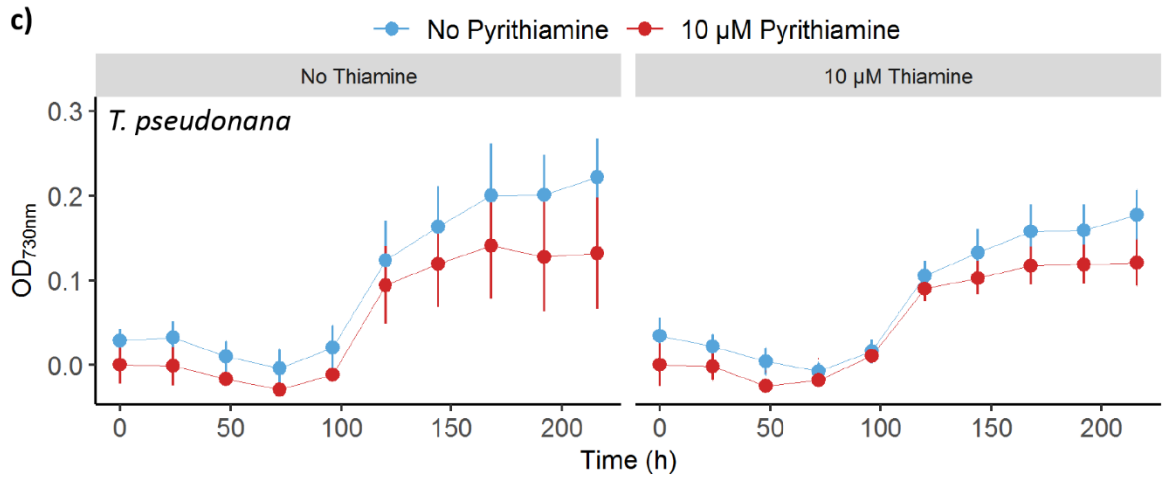


Figure 3.14. Continued

3.4 Discussion

The work in this chapter has been focused on unravelling the evolutionary prevalence of predicted TPP aptamers in the diatom lineages as well as understanding its potential function. The main results shown here challenge some assumptions in the literature such as the direct correlation between sequence similarity and function annotation.

Firstly, the thiamine biosynthesis pathway was investigated at a bioinformatic level in *P. tricornutum*. Most of the pathway was already unravelled and associated with the corresponding gene IDs in Bertrand & Allen (2012). Building on that knowledge, prediction of subcellular location indicated that all the enzymes in the thiazole branch are either expressed from or targeted to the chloroplast, except for THIF, which is encoded in the nuclear genome and not predicted to be targeted to the chloroplast. It seems unlikely that only one step of the pathway is catalysed in the cytosol, and therefore it remains to be seen whether SignalP fails to predict a target peptide in THIF, or the gene is wrongly annotated in the genome.

Turning to the pyrimidine branch of the pathway, identification of the THIC gene (Bertrand & Allen, 2012) led to the assumption that there was only one enzyme catalysing the production of HMP. But GO term and homology searches included in this chapter have confirmed the presence of *NMT1/THI5* homologues in Stramenopiles, Chlorophytes, Haptophytes, Cryptophytes and Mamelliophytes which had not been previously reported. The presence of a predicted TPP aptamer at the 3'UTR of the *F. solaris NMT1* gene provides additional evidence to indicate that these homologues have a relevant function in thiamine metabolism. In a phylogenetic tree, *F. solaris NMT1* candidate is monophyletic with the green algae *NMT1* candidates while the other diatom *NMT1* candidates cluster together with cryptophytes and haptophytes homologues. The clustering with green algae of the *F. solaris NMT1*

candidate and its unique predicted TPP aptamer can be explained by the recent discovery that the *F. solaris NMT1* candidate is the only member of the THI5/NMT1 family originated from horizontal gene transfer among 10 studied diatom genomes (Vancaester *et al.,* 2020) and notably the *NMT1* in the green alga *Micromonas RCC299* also has a TPP riboswitch in its 3'UTR (Nguyen *et al.,* 2016). Overall, the algal NMT1 candidates have higher sequence similarity with the bacterial ThiY*,* a periplasmic component of an ABC-transport system for an HMP salvage intermediate, than with yeast HMP synthases of the THI5/NMT1 family. Further experimental characterisation will be required to elucidate whether the THI5/NMT1 homologues in the different algae groups play a role in the transport system, like ThiY*,* or synthesise HMP from pyridoxal, like THI5/NMT1. This question is further discussed in Chapter 6.

From an evolutionary perspective, it is interesting to observe that fungi rely on THI5/NMT1 and lack THIC, while plants use THIC and do not have THI5/NMT1. Many algae have both THIC and the THI5/NMT1 homologue which raises the question of the evolutionary origin of both genes. In the case of diatoms in particular, a possibility is that one of the genes was inherited from the endosymbiont, believed to be an ancestor of the extant rhodophytes, and the other from the host of the secondary endosymbiosis event. A standard BLASTP search with PtTHIC against all rhodophyte sequences in NCBI gave hits with more than 65 % identity in 6 different species, while a BLASTP search with PtNMT1 found no significant hits. A possible explanation could be that the THI5/NMT1 homologue was provided by the host genome, while the *THIC* was incorporated from the endosymbiont in Stramenopiles, Cryptophytes and Haptophytes. This hypothesis would associate the HMP metabolism of the host cell with that present in current fungal species. It would be interesting to investigate further this question through comparative phylogenetics of *THIC* and *NMT1* genes in Rhodophyta, fungi and secondary endosymbionts with a rhodophyte-derived plastid. Turning to the thiazole branch of the biosynthetic pathway, Chlorophytes and Cryptophytes rely on the plant-like THI4 HET synthase, while Haptophytes and Stramenopiles rely on the bacterial-like THIG HET synthase (Figure 1.2; McRose *et al.,* 2020). In this case, the genes seem to follow the evolutionary patterns of the host genomes (see Figure 1.6 for the phylogenetic relationship of algal lineages).

Next, attention was turned to the regulation of the thiamine biosynthetic pathway, with a special interest in the TPP riboswitches previously predicted in the 3'UTR of *PtTHIC* and *PtSSSP* (Croft *et al.,* 2007; McRose *et al.,* 2012). All TPP riboswitches mechanistically described to date in eukaryotes are based on alternative splicing, but sequence analysis of the *PtTHIC* 3'UTR revealed no obvious splicing sites associated with the predicted TPP aptamer. Instead, a polyadenylation site overlapping the 3' arm of the P1 stem would be a good candidate to explain the molecular mechanism of the predicted aptamer. The polyadenylation site seems to be well conserved in the P1 stems of the riboswitches

found in diatoms in this work, which are all placed in 3'UTRs. Under this hypothetical mechanism, the riboswitch would mediate an alternative polyadenylation event in response to thiamine supplementation that would change the length of the 3'UTR. A longer 3'UTR could have additional regulatory domains that could lead to the repression or decay of the transcript leading to a downregulation of the gene. Similar alternative polyadenylation regulatory mechanisms have already been described in a wide variety of eukaryotes (Tian & Manley, 2017). However, the RT-qPCR experiments in the presence of exogenous thiamine or HMP showed there was no effective downregulation of *PtTHIC, PtNMT1* or *PtSSSP* and the 3'RACE experiments further showed that there were no changes in 3'UTR length in response to thiamine or HMP supplementation in *PtTHIC* and *PtSSSP*, so the hypothesis of a mechanism based on alternative polyadenylation regulation was in principle rejected.

It was possible that *P. tricornutum* did not take up thiamine, but HPLC analysis showed that intracellular levels of thiamine and TPP significantly increased upon thiamine supplementation. The increase in TPP unequivocally indicated that supplemented thiamine was taken up by *P. tricornutum* and pyrophosphorylated intracellularly. In Chapter 5, a *PtSSSP* knock-out is used to test whether SSSP is the transporter responsible for thiamine uptake using the HPLC-based uptake assay used here. In future experiments, the HPLC method could also be used to characterise the kinetics of thiamine import by studying a time-series at the minute and hour range of the thiamine uptake rate.

Once demonstrated that *P. tricornutum* could take up thiamine, the resistance to the pyrithiamine treatment provided further evidence to suggest the predicted aptamers did not regulate gene expression in response to thiamine supplementation. Previous experiments in *A. thaliana* and *C. reinhardtii* have shown that when TPP aptamers suffer loss-of-function mutations the organisms become insensitive to pyrithiamine (Bocobza *et al.,* 2013; Moulin *et al.,* 2013). Following this logic, the lack of response of the *PtTHIC* predicted riboswitch would prevent the downregulation of *THIC* by pyrithiamine keeping the flux through the pathway to sustain TPP levels.

Although the 6-fold downregulation of *PtTHIC* in the presence of cobalamin was not statistically significant due to the low number of replicates, previous studies showed that it is downregulated by cobalamin supplementation (Bertrand *et al.,* 2012). Given that *PtSSSP* also has a predicted TPP aptamer in its 3'UTR and it was not downregulated by cobalamin, it seems unlikely that the aptamer plays any role in cobalamin response of *PtTHIC*. Instead, the discovery of a conserved motif in the promoter region of *PtTHIC* might underpin the molecular mechanism regulating *PtTHIC* in response to cobalamin supplementation. The motif found in the *PtTHIC* promoter was also found in numerous genes known to be downregulated by cobalamin and related to the C1 cycle, including four copies of

the motif in the *PtMETE* promoter. The motif HMM could even be used for predictive searches in which we found the motif in the promoter regions of other genes related to the C1 cycle such as a GTP cyclohydrolase in the THF pathway. The motif is a 14 bp palindrome, which is a usual feature of binding sites for transcription factor that act as dimers. It would be interesting to find out which transcription factor binds the motif identified and test whether and how it mediates the response to cobalamin supplementation. The identification of both motif and transcription factor would define the explanation of the molecular mechanism shared in the regulation of cobalamin and thiamine metabolism. As explained in section 1.1, B vitamin metabolisms are intricately interrelated with one another and this is proposed to arise from their evolutionary origin in the RNA world. The discovery of a shared regulatory mechanism between thiamine and cobalamin related genes would be of special interest for our research group and would contribute to our understanding of the broader evolutionary and physiological relationships between B vitamin metabolisms. This question is also further discussed in Chapter 6.

Chapter 4. Studying vitamin metabolism regulation in *P. tricornutum* and *C. reinhardtii* with reporter constructs

4.1. Introduction

Semiquantitative RT-PCR, RT-qPCR and 3'RACE experiments presented in Chapter 3 have shown that the *PtTHIC* predicted TPP aptamer does not mediate gene regulation at a transcriptional or post-transcriptional level. These results contrast with the experimental evidence for thiamine biosynthesis regulation mediated by TPP aptamers in *A. thaliana, C. reinhardtii* and *N. crassa* (Cheah *et al.,* 2007; Croft *et al.,* 2007; Wachter *et al.,* 2007). Additionally, the *PtTHIC* unresponsiveness to thiamine supplementation is unexpected given the sequence conservation of its predicted TPP aptamer across the diatom lineage. As an alternative explanation for the results, the molecular methods used in Chapter 3 may have not been sufficient to detect a *PtTHIC* response to thiamine supplementation given that the regulation of the thiamine biosynthesis pathway may be only observed at a protein or metabolic level. Additional experimental approaches need to be employed to investigate whether thiamine biosynthesis may be regulated at the protein or metabolic level in *P. tricornutum*.

The reporter construct experiments carried out in our group to investigate thiamine regulation in *C. reinhardtii* give a precedent of how transgenic expression methods can be used to study genetic regulation at the protein and metabolic level. Reporter constructs were already used for the discovery and characterisation of TPP riboswitches in *C. reinhardtii* (Croft *et al.,* 2007). Reporter constructs were later used to determine the *CrTHI4* riboswitch response to different thiamine intermediates at the protein level (Moulin *et al.,* 2013). Recently, there has been a step change in our understanding of the *C. reinhardtii* riboswitches with the establishment by Dr Ginnie Nguyen of a riboswitch platform based on the *CrTHI4* 5'UTR in which different aptamers, including *CrTHIC* and *AtTHIC*, can be swapped-in in a modular way (Mehrshahi *et al.,* 2020). One of the experiments presented in this chapter contributed to this publication, which also includes the first example of the use of riboswitches in *C. reinhardtii* to regulate the expression of a terpene-producing heterologous pathway developed by Dr Payam Mehrshahi.

In Chapter 3, RT-qPCR experiments showed that *PtTHIC* is regulated by cobalamin supplementation in agreement with published transcriptomic and proteomic evidence (Bertrand *et al.,* 2012). In addition, the bioinformatic studies presented in section 3.2.6 revealed a conserved motif in diatom *THIC* promoter regions that is also present in most cobalamin-downregulated genes in the transcriptomic

and proteomic dataset, including four copies in *PtMETE*. The bioinformatic evidence is not sufficient to draw conclusions but allow us to hypothesise that there is a common molecular mechanism that downregulates expression of cobalamin-related genes in diatoms, including *THIC* and *METE*, based on a conserved motif in their promoter. This hypothesis remains to be tested and reporter constructs offer the opportunity to investigate cobalamin regulation as shown by numerous experiments already performed in our group.

Our interest for cobalamin-dependent genetic regulation arose from the realisation that more than 50 % of the algal species studied were cobalamin auxotrophs (Croft *et al.,* 2006). Cobalamin auxotrophy is related to the absence of the cobalamin-independent methionine synthase (METE) in the genome of auxotrophic species, and so must rely on the cobalamin-dependent version of methionine synthase (METH). It has been proposed that cobalamin is an important limiting factor for ecologically relevant algal auxotrophs such as the diatom *T. pseudonana* as cobalamin levels are really low or undetectable in large parts of the ocean (Sañudo-Wilhelmy *et al.,* 2012). Many algal species that have METE and thus are not cobalamin auxotrophs also contain METH, including both *C. reinhardtii* and *P. tricornutum* (Croft *et al.*, 2005; Helliwell *et al.,* 2011). Reporter constructs showed that the *CrMETE* promoter downregulates expression in response to cobalamin supplementation in *C. reinhardtii* (Helliwell *et al.,* 2014) and more recently, Patrick Hickland in our research group has extensively characterised the *PtMETE* promoter sequence showing it also downregulates expression in response to cobalamin supplementation. In this chapter I test the function of the cobalamin-responsive motif identified in *PtMETE* with a reporter construct experiment the results of which contribute to a manuscript in preparation on the *PtMETE* promoter by Patrick Hickland.

On the technical side, the experiments in this chapter have been possible thanks to the adoption of a MoClo system and the establishment of a reliable and reproducible pipeline for the transformation, screening and testing of reporter constructs in *P. tricornutum* and *C. reinhardtii* (Figure 4.1). The MoClo system is based on the Golden Gate cloning technique and uses Type-IIS restriction enzymes BpiI and BsaI to generate designer overhangs for scarless and directional ligation of several parts in a one-tube reaction (Weber *et al.,* 2011). In addition, MoClo sets a standardised syntax for the different constitutive parts of an expression cassette (Promoter, CDS, 3'UTR, etc) for the seamless permutation of parts and allows for the construct building blocks to be exchanged among the scientific community. MoClo also has different plasmid levels to give the system modularity and scalability to generate genetic constructs with multiple expression cassettes. The only minor downside to the system, is that new parts need to be domesticated to remove internal BpiI and BsaI sites (Engler *et al.,* 2014). The MoClo cloning standard was initially developed for plant systems (Patron *et al.*, 2015) but has been pioneered for algal genetic engineering by our laboratory who, together with 4 other groups,

generated the first MoClo kit for *Chlamydomonas reinhardtii* (Crozet *et al.*, 2018). A similar kit for *Phaeodactylum tricornutum* is in preparation and several of the parts developed for this chapter will be included in it (Geisler *et al.*, in preparation). These three aspects characteristic of the synthetic biology approach, standardisation, modularity and scalability, provide the means to generate large numbers of permuted constructs (for instance to test a single part such as a promoter) rapidly and easily and thus to address any biological question systematically.

P. tricornutum nuclear genome is conventionally transformed by microparticle bombardment and it takes around three weeks to get 10-100 colonies (Apt *et al.,* 1998). Members of the lab, in particular Dr Katrin Geisler, have optimised a reliable and reproducible *P. tricornutum* nuclear transformation method based on multi-pulse electroporation (Figure 4.1c; Miyahara *et al.,* 2013). With this method, *P. tricornutum* transformant colonies can be obtained in as little as two weeks and then be screened and analysed by different methods based on 96-well plates and plate readers. Reporter genes based on antibiotic resistance or fluorescent proteins give outputs that are easy and fast to track and can produce high-throughput data (Figure 4.1g). Similar methods have been available for *C. reinhardtii* for much longer and are also routinely used in our lab. The ability to rapidly assemble several constructs following the MoClo system and obtain hundreds of independent transformants for each of the constructs, in combination with the capacity to track reporter outputs with ease and high throughput have been instrumental to address questions related to thiamine and cobalamin genetic regulation in algae in this chapter.

In summary, this chapter describes experiments performed with the help of reporter constructs to investigate whether exogenous thiamine regulates *PtTHIC* at the protein and metabolic level. In addition, reporter constructs have also been used to attempt to introduce heterologous riboswitches from *C. reinhardtii* into *P. tricornutum* and vice versa. And finally, various reporter constructs have been used to investigate whether *PtTHIC* is regulated by cobalamin supplementation and what is the role of the predicted cobalamin-responsive motif in the *PtMETE* promoter. Before this thesis, there was no reporter construct experiments carried out to investigate the genetic regulation of *PtTHIC* and the function of its associated predicted TPP aptamer, the work in this chapter largely translates and expands the previous experiments performed in *C. reinhardtii* to the marine diatom.

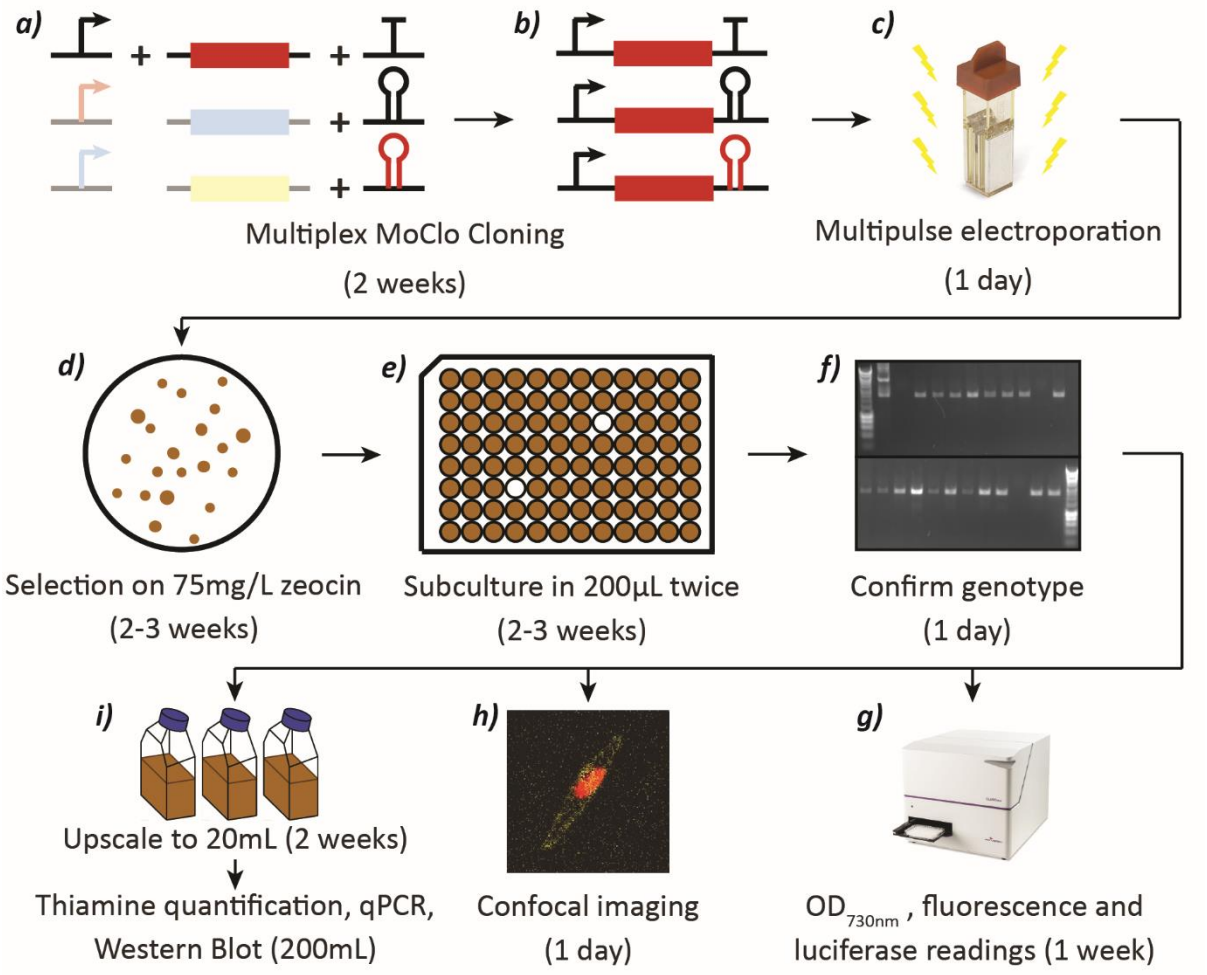


Figure 4.1. General experimental workflow for reporter constructs in* P. tricornutum.** ***a) *Modular genetic components were obtained from other group members or amplified and domesticated from* P. tricornutum *genome.* ***b)*** *A combination of the available parts was combined in permutation to generate a set of constructs using the MoClo system.* ***c)*** *Two and a half micrograms of each construct was linearised and electroporated into* $1.25 \cdot 10^7$ P. tricornutum CCAP 1055/1 *cells in triplicates.* ***d)*** *Hundreds of colonies were generally obtained after 2-3 weeks selection in 75* $mg \cdot l^{-1}$ *zeocin plates and were picked into 200 µl in 96-well plates.* ***e)*** *Cultures were subcultured twice in f/2 media with 75* $mg \cdot l^{-1}$ *zeocin to discard any clamps and ensure only resistant cells survived.* ***f)*** *Independent transformant were genotyped by PCR with the Phire Plant Direct PCR kit (Thermofisher) and a representative band was sent for sequencing to confirm the incorporation of the transgene.* ***g,h)*** OD_{730nm} *and fluorescence curves as well as confocal imaging could be started right away, and luciferase readings were taking 3 days after a 10 % subculture.* ***i)*** *RT-qPCR, thiamine quantification and Western Blots required more biomass, and selected independent transformants were upscaled to 20 ml or 200 ml before phenotyping. In total, it can take as little as 8 weeks from design to results to address any given question systematically. More detailed method descriptions can be found in Chapter 2.*

The objectives of this chapter are to:

1. Investigate whether the *PtTHIC* predicted riboswitch can control the expression of a transgene in response to thiamine in both *P. tricornutum* and *C. reinhardtii*
2. Investigate whether a mutation in the *PtTHIC* predicted riboswitch leads to deregulation of the thiamine biosynthesis pathway
3. Investigate whether *CrTHI4* and *CrTHIC* riboswitches can mediate a genetic expression change in response to thiamine supplementation in *P. tricornutum*
4. Investigate the molecular mechanism by which cobalamin regulates *PtTHIC* and *PtMETE*

4.2. *PtTHIC* 3'UTR reporter constructs

With the objective of determining the regulatory potential of *PtTHIC* 3'UTR at the protein level, a set of luciferase reporter constructs was designed (Figure 4.1a). The constructs contained the *PtFCPA* promoter-5'UTR driving the expression of the *Gaussia princeps* luciferase gene (*GpLuc*), and the *PtTHIC* 3'UTR in 3 different lengths: 217 bp including just until the end of the predicted aptamer secondary structure and truncating the AACAAA motif identified in section 3.2.4 (*pMLP2021*), 256 bp matching the 3'UTR observed in most EST reads and in the most common 3'RACE band in section 3.3.3 (*pMLP2022*), and 819 bp to make sure any relevant genetic context beyond the canonical transcription termination site, including any alternative polyadenylation sites, was included (*pMLP2023*). In addition, a construct with the *PtFCPC* 3'UTR instead of the *THIC* 3'UTR was added as a control (*pMLP2011*). The *PtFCPA* promoter-5'UTR was chosen because it has been previously characterised as a constitutive strong promoter (Apt *et al.*, 2016). It was expected that with a strong thiamine-independent promoter, one would be able to see a greater riboswitch dynamic range and maximise the signal-to-noise ratio. All constructs contained a zeocin selection cassette based on pAKS010, a plasmid previously assembled in the lab by Astrid Stubbusch, which contains the *PtFCPB* promoter, the zeocin resistance gene *Ble*, and the *PtFCPC* 3'UTR.

The constructs were introduced into *P. tricornutum CCAP 1055/1* by elecrotoporation followed by selection on 75 mg·l^{-1} zeocin. Colonies appeared after 2 weeks, and between 79 to 100 colonies were picked into 96-well plates for each of the constructs. After 4 weeks of subculturing, all independent transformants were then used to determine activity of the reporter gene in parallel with 6 biological replicates of a positive control strain that was developed by Dr Katrin Geisler which consisted of the *GpLuc* CDS controlled by the *PtFCPC* promoter and UTRs transformed into *P. tricornutum UTEX 646* background. Luciferase activity in cells grown for 4 days in 200 µl f/2 with 75 mg·l^{-1} zeocin and in the

absence of thiamine was analysed in a plate reader. The results showed that while most positive controls presented relatively high luciferase activity, very few transformants for the constructs tested showed luciferase activity higher than the lowest positive control (Figure 4.2b). Only 6 independent transformants carrying the control *PtFCPC* 3'UTR (*pMLP2011*), one independent transformant with the 256 bp long *PtTHIC* 3'UTR (*pMLP2022*) and another one with the longest version of the *PtTHIC* 3'UTR (*pMLP2023*) showed luciferase activity. The lack of transformants with measurable luciferase activity out of hundreds of transformants screened indicates that there may have been a fundamental problem with the construct design. It was therefore not possible to use these strains to test whether the *PtTHIC* 3'UTR responds to thiamine supplementation.

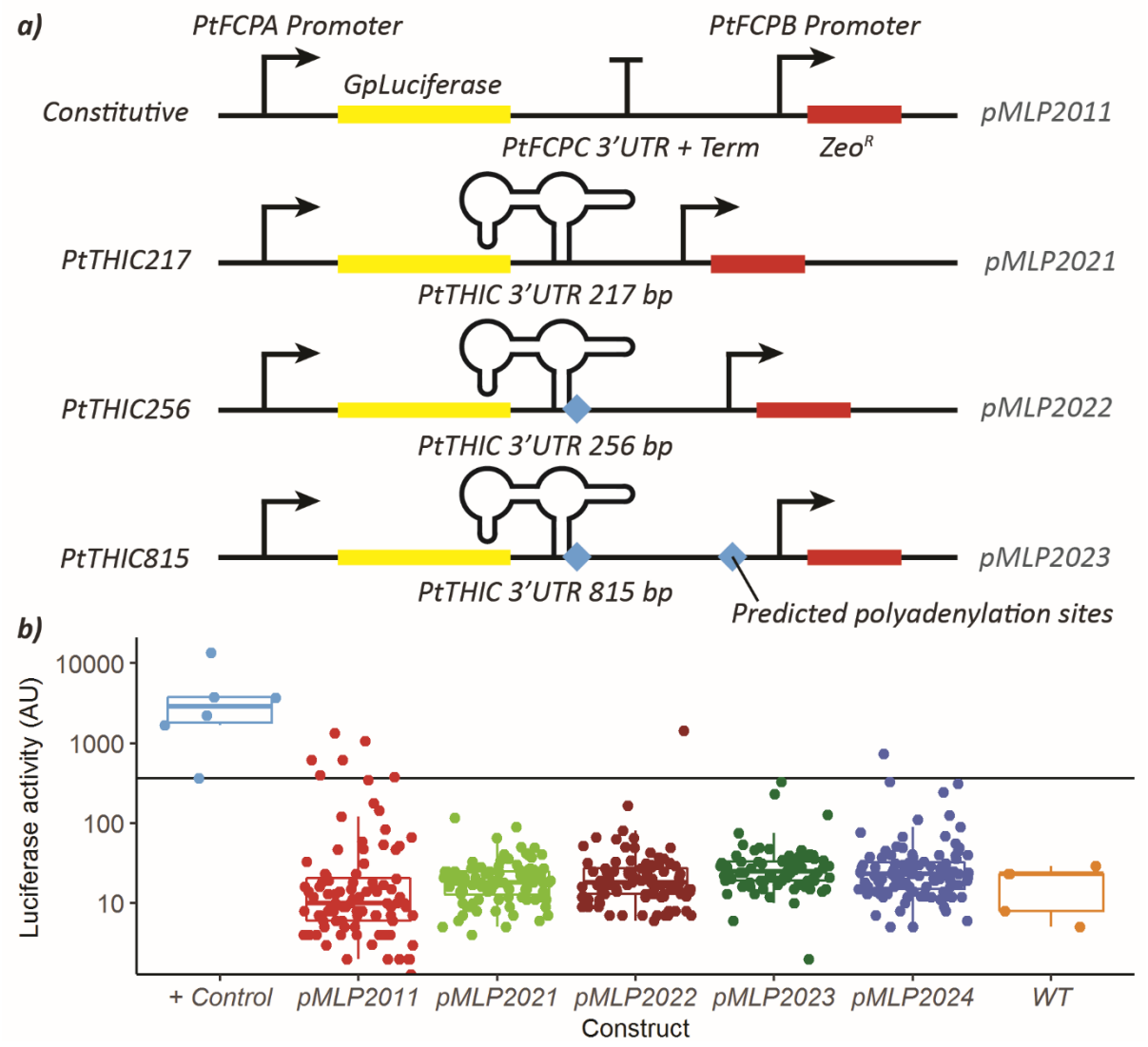


Figure 4.2 Luciferase assay on reporter constructs with a luciferase gene followed by different lengths of the* PtTHIC *3'UTR. *All recovered transformants from zeocin selection were grown in 200 µL in a 96-well plate for 4 days and tested in a luciferase assay in a plate reader. An independent positive control* P. tricornutum UTEX 646 *strain expressing luciferase provided by Dr Katrin Geisler and a negative control only containing a zeocin resistance cassette as well as a WT control were included in the experiment. Only those transformants that expressed higher levels of luciferase expression than the lowest value measure for a positive control (horizontal line) were considered to express luciferase.*

A second round of constructs was designed to attempt to overcome some of the problems of the luciferase reporter approach. This time the reporter construct had the *PtTHIC* promoter and 5'UTR as well as the 815 bp version of the *PtTHIC* 3'UTR controlling the expression of a *Ble-Venus* fusion (*pMLP2048*). A control construct with the *PtEF2* promoter and 5'UTR and the *PtFCPC* 3'UTR controlling the expression of *Ble-Venus* was cloned in parallel (*pMLP2047;* Figure 4.3a). It is known that 5' and 3'UTRs interact during the transcription process (Mazumder *et al.*, 2003), so it seemed sensible to have both *PtTHIC* UTRs in the reporter construct when assessing the regulatory effect at the protein level. Additionally, the zeocin resistance gene *Ble* was chosen as a reporter, rather than luciferase, so that in the presence of zeocin the addition of thiamine would lead to growth arrest only if the *PtTHIC* promoter or UTRs respond to thiamine supplementation. After electroporation and three weeks on zeocin selection plates, 29 and 445 independent transformants could be recovered for the control (*pMLP2047*) and *PtTHIC* (*pMLP2048*) constructs respectively, and since the same antibiotic resistance cassette is being used as the reporter, all these transformants are in principle suitable to test the function of the *PtTHIC* promoter and UTRs. Three independent transformants for each of the constructs and *P. tricornutum WT* were then grown in triplicate in 96-well plates under the presence or absence of 75 $mg \cdot l^{-1}$ zeocin and with or without 10 µM thiamine for 8 days, recording OD_{730nm} as a proxy for growth every 24 hours. The results showed that while the wild type did not grow under the presence of zeocin, all transformants for each construct grew well in the presence of zeocin, both in the presence and absence of thiamine (Figure 4.3b). These results provide evidence to suggest that neither the *PtTHIC* promoter nor its UTRs mediate a response to thiamine supplementation measurable in reporter levels.

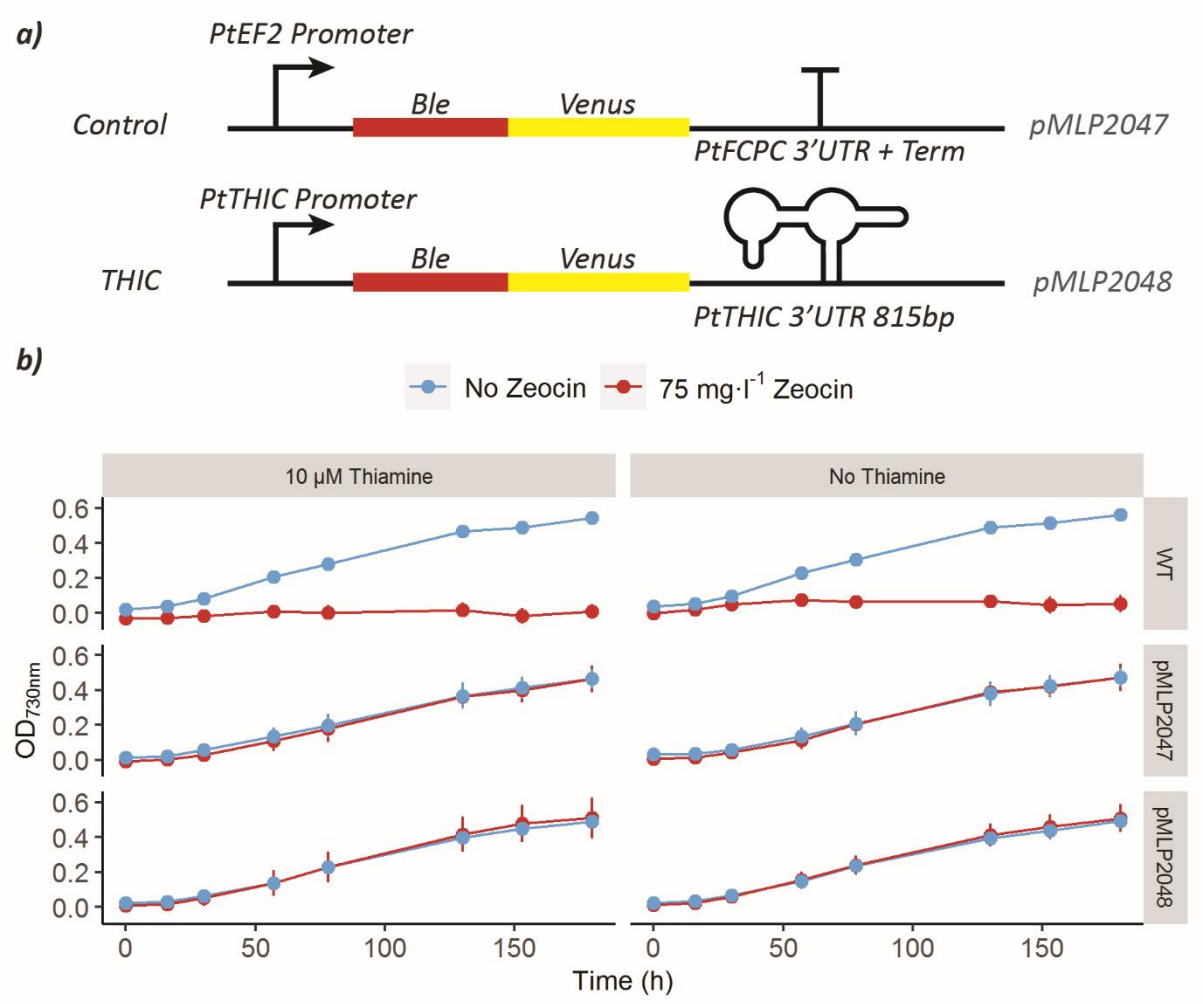


Figure 4.3. Zeocin sensitivity assays on* P. tricornutum *reporter strains to test the* PtTHIC *3'UTR response to thiamine in vivo. a) *Reporter constructs carrying a protein fusion between the Ble zeocin resistance gene and Venus fluorescent protein were flanked by the* PtTHIC *promoter and UTRs to test for their response to thiamine supplementation in comparison with a control were the* Ble-Venus *fusion was flanked by the constitutive promoter* PtEF2 *and terminator* PtFCPC*. **b)** Three independent transformants carrying the control or the test constructs, alongside three* P. tricornutum *WT biological triplicates were grown in triplicates in the presence (blue) or absence (red) of 75 mg·l^{-1} zeocin and the presence (left) or absence (right) of 10 µM thiamine in 96-well plates and their growth was recorded every 24h as OD_{730nm} with a plate reader.*

4.3. *PtTHIC* overexpressor strains with a mutated aptamer

After showing that the *PtTHIC* promoter and UTRs could not mediate a response at the protein level in a heterologous expression reporter two main questions arose. First, is it possible that the *PtTHIC* predicted aptamer requires sequences both in the 3'UTR and the CDS to regulate expression? Second, could a response to thiamine supplementation of the predicted *PtTHIC* aptamer be measured at a metabolic level? To address these questions a new set of constructs was designed inspired by studies in *C. reinhardtii* and *A. thaliana* riboswitch mutants. In this occasion, the *PtTHIC* promoter, CDS (including introns), and the 815 bp version of the 3'UTR were cloned including a C-terminal HA-Tag to quantify protein levels and distinguish the extra copy from the native copies of the gene (*pMLP2064*; Figure 4.4a). The HA-Tag was placed at the C-terminus despite its proximity with the TPP aptamer and at risk of interfering with the predicted aptamer potential molecular mechanism because PtTHIC has a predicted chloroplast targeting peptide and a tag at the N-terminus could interfere with subcellular localization processes. In parallel, a second construct was built with identical sequence except for a targeted mutation in the ligand binding motif of the predicted aptamer from CUGAGA to CUCUCU (*pMLP2065*). The mutation was equivalent to the *C. reinhardtii pyr1* mutant which is no longer able to respond to TPP (Croft *et al.*, 2007; Moulin *et al.,* 2013). If the native *PtTHIC* predicted aptamer was responsive to TPP, a mutant form that no longer bound TPP should be dominant over the two native copies of *PtTHIC* and would deregulate the thiamine biosynthetic pathway leading to increased levels of TPP. This was observed in *A. thaliana* plants overexpressing an extra copy of *AtTHIC* with a mutated riboswitch (Bocobza *et al.,* 2013).

The two constructs were electroporated into *P. tricornutum* and after three weeks of selection on zeocin, 94 independent transformants were picked into 96-well plates for each construct. Ten and eight independent transformants for the unmutated (*pMLP2064*) and mutated (*pMLP2065*) versions of the PtTHIC overexpression construct respectively were upscaled to 40 ml in f/2 together with four WT biological replicates. Twenty millilitres of each culture were harvested at late exponential phase and their thiamine and TPP intracellular content was measured by HPLC as described in Chapter 3. The results showed that thiamine and TPP levels for the transformants with the unmutated construct (*pMLP2064*) increased compared to the WT, although this was only statistically significant for TPP levels with a Tuckey HSD test with 0.05 confidence level. Transformants with the mutated construct (*pMLP2065*) did not show any observable difference in thiamine or TPP levels with the unmutated construct transformants (Figure 4.4b). These results show that the mutation in the predicted *PtTHIC* aptamer do not lead to increased intracellular levels of thiamine and TPP as would be expected if the aptamer mediated a response to TPP levels as previously observed in *C. reinhardtii* and *A. thaliana* equivalent mutants. The increase in thiamine and TPP levels compared to the WT cannot be explained

by the mutation and might be a consequence of PtTHIC higher abundance given that both unmutated (*pMLP2064*) and mutated (*pMLP2065*) constructs show very similar levels.

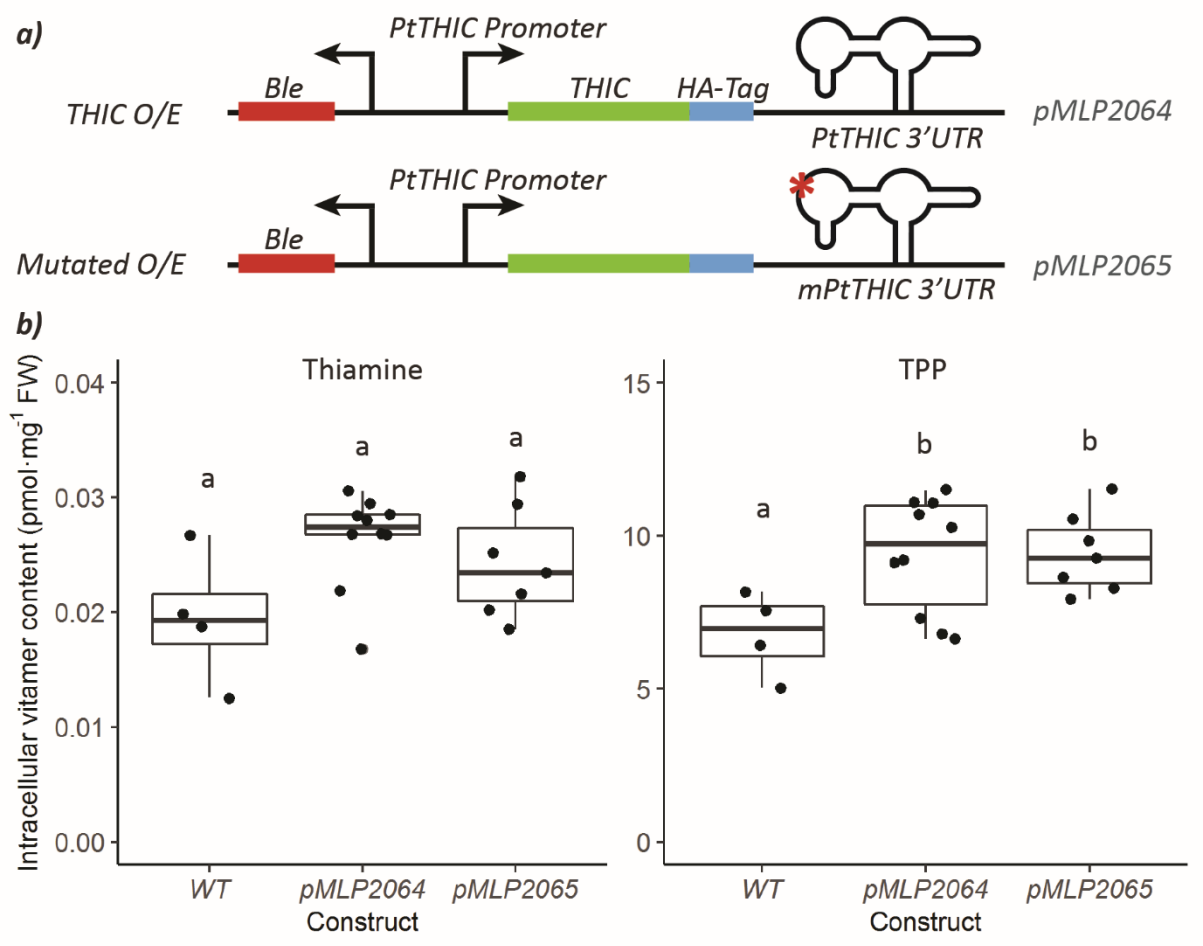


Figure 4.4. Intracellular thiamine and TPP levels in* P. tricornutum *overexpressing* THIC *with and without a mutated aptamer. a) *Constructs including an extra copy of* PtTHIC *including promoter, CDS with introns, and 815 bp downstream the STOP codon as 3'UTR. Constructs also included an HA-tag for western blotting, and in one of the constructs the thiamine-binding CUGAGA motif was mutated into CUCUCU.* ***b)*** *Four biological replicates for WT, 10 independent transformants for the* PtTHIC *overexpressor and 8 independent transformants for the mutated overexpressor were grown in f/2, harvested and their intracellular thiamine and TPP content was measured by HPLC and normalised by fresh weight. The results are shown as a boxplot and each measurement is shown as an individual data point. There are no statistical differences (Tukey HSD test with 0.95 confidence level) between the* PtTHIC *overexpressor and the mutated overexpressor, both overexpressor constructs show higher TPP levels than WT.*

The mild increase in thiamine and TPP levels by the overexpressing strains compared to WT suggests that the extra copy of *PtTHIC* is being translated and contributes to enhanced thiamine biosynthesis. To confirm that the *PtTHIC* extra copy was being expressed, and to test whether thiamine supplementation could impact protein levels in a riboswitch-dependent manner, a western blot was performed on total cell extracts from 200 ml cultures of a single independent transformant for the unmutated (*pMLP2064*) and mutated (*pMLP2065*) version of the overexpression construct grown in the presence or absence of 10 µM thiamine. A WT control grown in the absence of thiamine was also added to ensure the anti-HA antibodies did not cross-react with any native protein. The western blot results showed that the protein is being expressed in transformants of both constructs without any obvious difference between the transformants or between thiamine treatment and control conditions. Notably, the protein was detected at the expected size, but appeared as a double band (Figure 4.5). The double band might reflect a post-translational modification, including incorporation of the 4Fe-4S cluster, glycosylation, protease cleavage or chloroplast target peptide processing. With these results, it is possible to conclude that the targeted mutation in the predicted *PtTHIC* aptamer does not induce a deregulation of protein expression in agreement with the unchanged thiamine levels (Figure 4.4).

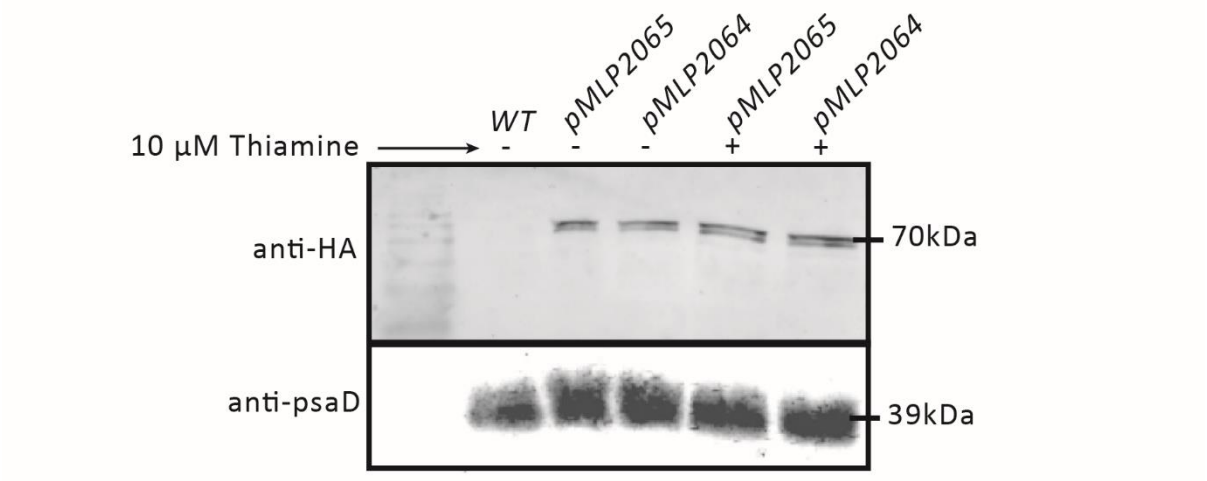


Figure 4.5 Western blot on* P. tricornutum *cells overexpressing* THIC *with a mutated (pMLP2065) or unmutated (pMLP2064) aptamer in the presence (+) or absence (-) of 10 µM thiamine*. A single independent transformant for each of the construct was grown in the presence of thiamine for 5 days, cells were harvested, lysed and run in an SDS-PAGE gel. The gel was then transferred to a PVDF membrane and incubated with anti-HA antibody and an anti-psaD loading control. A secondary anti-rabbit conjugated with a Dy800 fluorophore was incubated later, and observed using a fluorescence scanner shown in the image. PtTHIC has a predicted molecular weight of 70kDA corresponding to the band observed in the western.*

The overexpressor construct design also offered the possibility to test whether the *PtTHIC* downregulation in response to cobalamin supplementation observed by Bertrand *et al.* (2012) depends on the predicted *PtTHIC* aptamer or not. With this objective, three independent transformants for the mutated aptamer overexpression construct (*pMLP2065*), and six independent transformants for a newly assembled construct with the *PtFCPC* 3'UTR replacing the *PtTHIC* 3'UTR (*pMLP2066*) were grown in the presence or absence of 1 $\mu g \cdot l^{-1}$ cobalamin (Figure 4.6a). For each of the transformants and conditions, 5 ml were pelleted from 20 ml cultures at exponential phase and RNA was extracted, reverse-transcribed to cDNA, and an RT-PCR targeting the *PtTHIC* extra copy with the reverse primer on the HA-Tag was performed (see primers in Figure 4.6a). The results showed that for the mutated copy of *PtTHIC* (*pMLP2065*) the transcript levels were relatively higher than for the construct with the *PtFCPC* terminator (*pMLP2066*). For the mutated overexpressor (*pMLP2065*), we could see expression under both replete and deplete conditions, but the bands under cobalamin supplementation always showed lower intensity (Figure 4.6b). For the construct with the *PtFCPC* terminator (*pMLP2066*), we could only observe transcript in the cobalamin-deplete condition for 4 of the transformants (A, B, D and F), but yet again the addition of cobalamin seems to downregulate the expression of the *PtTHIC* extra copy below the limit of detection. Based on these results, it can be concluded that the *PtTHIC* cobalamin response depends on the promoter sequence and is independent of the predicted aptamer and indeed the entire *PtTHIC* 3'UTR.

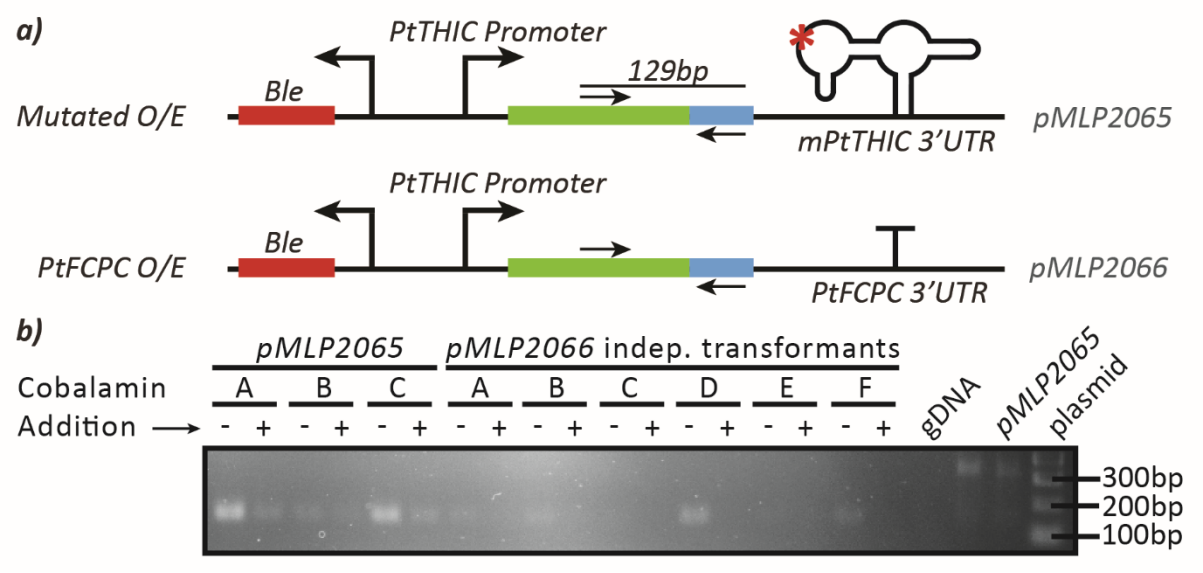


Figure 4.6. RT-PCR on* P. tricornutum *overexpressing* THIC *with a mutated aptamer or a* PtFCPC *3'UTR in the presence (+) or absence (-) of 1 µg·l^{-1} cobalamin. a) *Three independent transformants for a construct overexpressing* THIC *with a mutated aptamer (pMLP2065) and six independent transformants for a construct overexpressing* THIC *with a* PtFCPC *constitutive 3'UTR (pMLP2066) where grown for 5 days in f/2 with or without 1 µg·l^{-1} cobalamin.* ***b)*** *RNA was extracted, retrotranscribed, and an RT-PCR with primers targeting the overexpressed* THIC *gene, with a reverse primer base pairing with the HA-Taq was run for 35 cycles. The intensity of the bands in the 0,5 % agarose gels shows that cobalamin represses the expression of the reporter independently of the construct 3'UTR.*

4.4. *CrTHI4* and *CrT4_EPa50* co-regulation in *C. reinhardtii* cells

In parallel with the experiments investigating the predicted *PtTHIC* riboswitch, and after seeing convincing evidence showing it could not mediate a genetic expression response to thiamine supplementation, I turned to further investigate the native function, modularity and engineering of the TPP riboswitches in *C. reinhardtii*. This work follows on the modularisation of the aptamer and expression domains of the *CrTHI4* riboswitch which allowed the integration of heterologous aptamers in the *CrTHI4* riboswitch (Nguyen PhD Thesis, 2014). To modularise the *CrTHI4* riboswitch, Dr Ginnie Nguyen generated a modified *CrTHI4* riboswitch by moving the expression platform uORF out of the aptamer and into a region immediately downstream the aptamer which still conserved the native response to thiamine and HET (*CrTHI4_4N*). Once the expression platform was no longer co-located with the aptamer, it was possible to build chimeric riboswitches by replacing the *CrTHI4* aptamer for the *CrTHIC* (*CrTHIC_4N*) and the *AtTHIC* (*AtTHIC_4N*) aptamers which respectively respond to thiamine and HMP, and to thiamine only. Dr Ginnie Nguyen also used error-prone PCR on the native *CrTHI4* riboswitch to generate the *EPa50* mutant riboswitch, which responds to thiamine but not HET. These

novel riboswitches are the backbone of a paper recently published by our research group and in which the following experiment was included (Mehrshahi *et al.,* 2020).

The objective of this experiment was to test whether two engineered TPP riboswitches based with different affinity for different ligands could regulate simultaneously and with independent control two different transgenes. To address this question, Dr Payam Mehrshahi cloned and transformed into *C. reinhardtii UVM4* a construct with the *CrTHI4_4N* riboswitch controlling the expression of a *Ble-eGFP* fusion localised to the nucleus and the *EPa50* mutated riboswitch controlling the expression of a chloroplast targeted Venus fluorescent protein (*pPM34*; Figure 4.7a). Two weeks after electroporation colonies were picked into 96-well plates and subcultured for two weeks in 200 µl TAP. A few colonies were screened for Yellow Fluorescent Protein (YFP) and eGFP fluorescence with a Leica SP8 fluorescent confocal microscope and a representative transformant expressing both fluorescence reporters was identified. The selected transformant was then grown in TAP, TAP supplemented with 10 µM thiamine and TAP supplemented with 10 µM HET for 7 days and imaged with fluorescence confocal microscopy with channels detecting chlorophyll fluorescence, eGFP and Venus fluorescence. The results showed that the transformant had nuclear-localised eGFP fluorescence and chloroplast-targeted Venus fluorescence in TAP alone. Under thiamine supplementation eGFP fluorescence was downregulated below the detection limit, and Venus fluorescence significantly decreased. Under HET supplementation eGFP fluorescence was almost undetectable whilst Venus fluorescence did not show any obvious differences (Figure 4.7b). These results prove that two engineered riboswitches can independently regulate the expression of two transgenes responding to their respective ligands, opening the possibility to regulate increasingly complex synthetic biology circuits in *C. reinhardtii*.

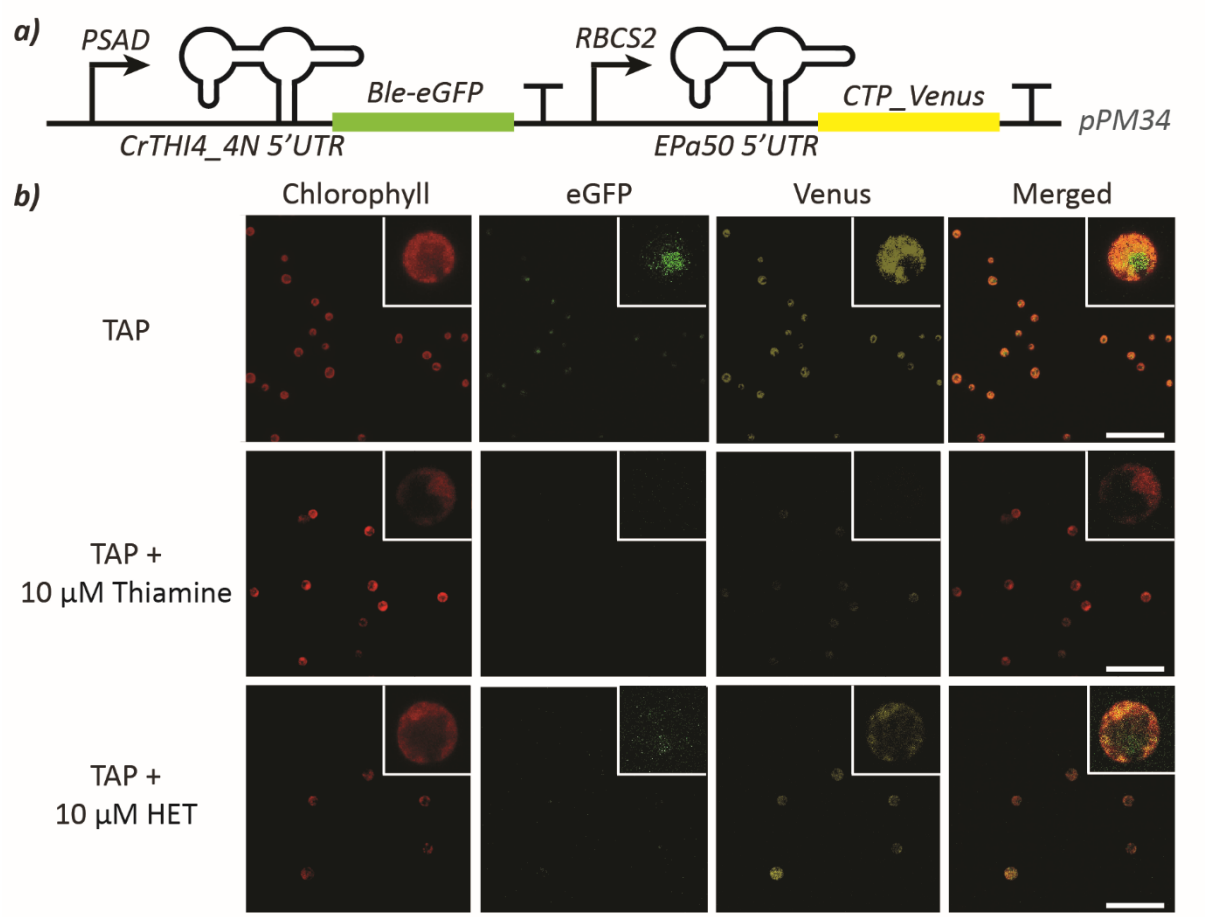


Figure 4.7. Confocal microscopy fluorescence images of a* C. reinhardtii *transformant simultaneously carrying fluorescent proteins regulated by* CrTHI4 *and* EPa50 *riboswitches. a) *A construct carrying a* Ble-eGFP *fusion controlled by the* CrTHI4 *5'UTR and a chloroplast-targeted Venus fluorescent protein controlled by the* EPa50 *riboswitch was cloned and transformed into* C. reinhardtii *by Dr Payam Mehrshahi.* ***b)*** *A representative independent transformant was cultured in a 96-well plate in TAP media, TAP plus 10 µM thiamine or TAP plus 10 µM HET. Cells were imaged in a confocal microscope with fluorescence channels detecting chlorophyll, eGFP and Venus. A general picture and a single cell close-up are shown. The white bar is 50 µm and the square insets are 13.5 µm wide.*

4.5. Testing the *CrTHI4* riboswitch in *P. tricornutum*

Since the native *PtTHIC* predicted riboswitch did not seem to mediate an effective response to thiamine supplementation, and with the objective of developing a modular TPP riboswitch that could control the expression of transgenes in synthetic biology designs in *P. tricornutum*, it was proposed to import the *CrTHI4* TPP riboswitch to *P. tricornutum.* To pursue this aim, two constructs were cloned and transformed into *P. tricornutum* which included a constitutive *PtFCPA* promoter followed by either the *CrTHI4_4N* (*pMLP2031*) or the *CrTHIC_4N* (*pMLP2032*) riboswitches presented in section 4.4 controlling the expression of a luciferase reporter with the *PtFCPC* terminator (Figure 4.8a). Three weeks after electroporation 42 and 63 colonies were picked respectively for *CrTHI4_4N (pMLP2031)* and *CrTHIC_4N (pMLP2032) constructs* and subcultured in 96-well plates in f/2 supplemented with 75 $mg \cdot l^{-1}$ zeocin. All transformants were tested for luciferase activity in parallel with the positive control provided by Dr Katrin Geisler (already used in section 4.2), a negative control with only a zeocin selection cassette and WT. None of the transformants carrying the *C. reinhardtii* riboswitches showed luciferase activity (Figure 4.8c). The results suggest that given the *C. reinhardtii* riboswitches were placed in the 5'UTR, a region that is involved in both transcription and translation initiation, the *C. reinhardtii* native sequences might not be recognised by the *P. tricornutum* transcription or translation initiation machinery preventing the expression of the reporter.

Next it was tested whether the lack of luciferase expression was explained at the transcript or the protein level. Nine independent transformants for the construct carrying the *CrTHI4_4N* riboswitch (*pMLP2031*) were grown in parallel with a culture of the luciferase positive control provided by Dr Katrin Geisler. DNA and RNA were extracted from 20 ml cultures grown to late exponential phase in f/2 and a genotyping PCR on the genomic DNA was performed in parallel with an RT-PCR on total cDNA. The results showed that all independent transformants with the *CrTHI4_4N* riboswitch had a luciferase band amplified from their genomic DNA, but no luciferase transcript could be detected. In contrast, a band could be observed from both cDNA and genomic DNA for the positive control (Figure 4.8c). These results lead to the conclusion that *CrTHI4_4N*, and possibly *CrTHIC_4N* by association, interfere with the *P. tricornutum* transcription machinery and do not allow expression of their associated genes.

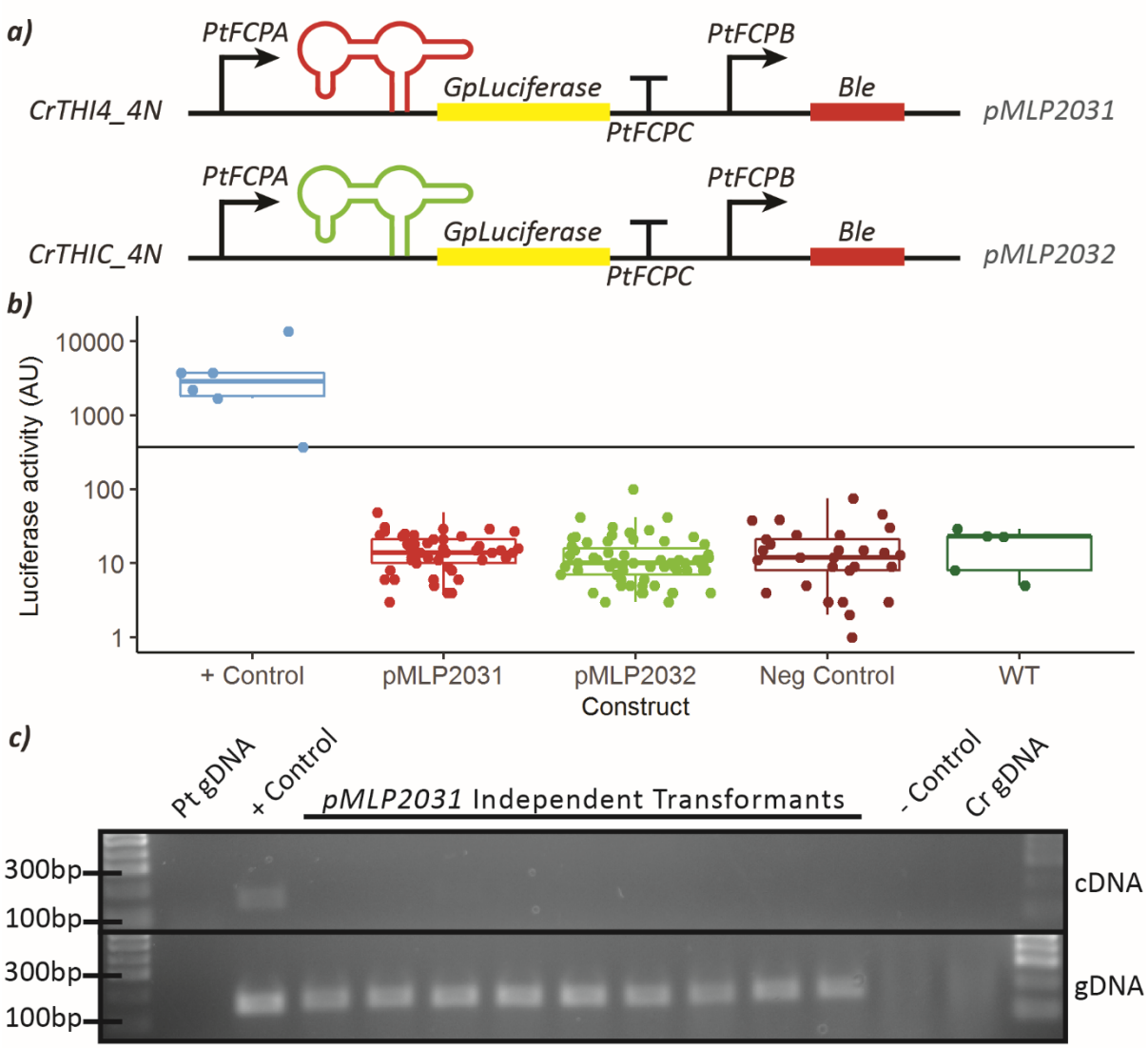


Figure 4.8. Luciferase assay and RT-PCR on* P. tricornutum *transformants carrying the* CrTHI4_4N *and* CrTHIC_4N *riboswitches. a) *Constructs carrying a luciferase reporter following either the* CrTHI4_4N *or the* CrTHIC_4N *riboswitch in the 5'UTR were transformed into* P. tricornutum. ***b)*** *All independent transformants selected on zeocin were phenotyped in the same luciferase experiment as in Figure 4.2.* ***c)*** *RNA from 9 independent constructs carrying the* CrTHI4 *riboswitch and a positive control was reverse-transcribed with random hexamers and tested for transcript presence in an RT-PCR on the cDNA with primers targeting the luciferase coding sequence. A genotyping control with the same PCR on the genomic DNA of the independent transformants was performed in parallel. Negative controls with no template, WT* P. tricornutum *and WT* C. reinhardtii *genomic DNA were also included.*

4.6. Testing the *CrTHIC* intron in *P. tricornutum*

Having seen that including *C. reinhardtii* functional riboswitches in the 5'UTR of expression cassettes interfered with transcription in *P. tricornutum,* but still with the objective of developing a functional riboswitch to control heterologous expression in *P. tricornutum*, the attention was turned to the *CrTHIC* riboswitch. *CrTHIC* has a riboswitch in its 6th intron that in response to thiamine supplementation induces alternative splicing and inclusion of an alternative exon with a premature stop codon (Croft *et al.,* 2007). To test the potential functionality of the *CrTHIC* riboswitch in a *P. tricornutum* expression cassette, an intron testing platform was developed by a protein fusion of the zeocin resistance cassette *Ble* and the *Venus* fluorescent protein. A variety of introns were then inserted in the +310 position of the *Venus* CDS, between two guanines, including the native *PtN78* intron (78 bp; *pMLP2041*), the widely used *CrRBCS2* intron that acts as an enhancer in *C. reinhardtii* (145 bp; *pMLP2042*) (Baier *et al.,* 2018), the whole *CrTHIC* 6th intron (1636 bp; *pMLP2043*), the first *CrTHIC* alternative intron (604 bp; *pMLP2044*), or a short version of the *CrTHIC* intron including only the 100 bp at the 5' and 3' end of the intron (200 bp; *pMLP2046*). The constructs were driven by the constitutive *PtEF2* promoter and the *PtFCPC* terminator and an intronless positive control was also included in the design (*pMLP2047;* Figure 4.9). Three weeks after transformation colonies appeared on zeocin selection plates, but not for all constructs. It was observed that while the intronless construct (*pMLP2047*) gave 29 colonies, and the controls with the *PtN78* (*pMLP2041*) and *CrRBCS2* (*pMLP2042*) introns gave 221 and 206 colonies respectively, no colonies appeared for any of the constructs bearing any of the versions of the *CrTHIC* intron (*pMLP2043-2046;* Table 4.1). The lack of colonies suggests that the expression of the *Ble-Venus* fusion is impaired by the presence of the *CrTHIC* intron. Given that the construct with the *CrRBCS2* intron successfully gave colonies, it is clear that the deleterious effect cannot be pinned down to all *C. reinhardtii* introns being incompatible with *P. tricornutum* but more precisely to a specific effect of the *CrTHIC* intron.

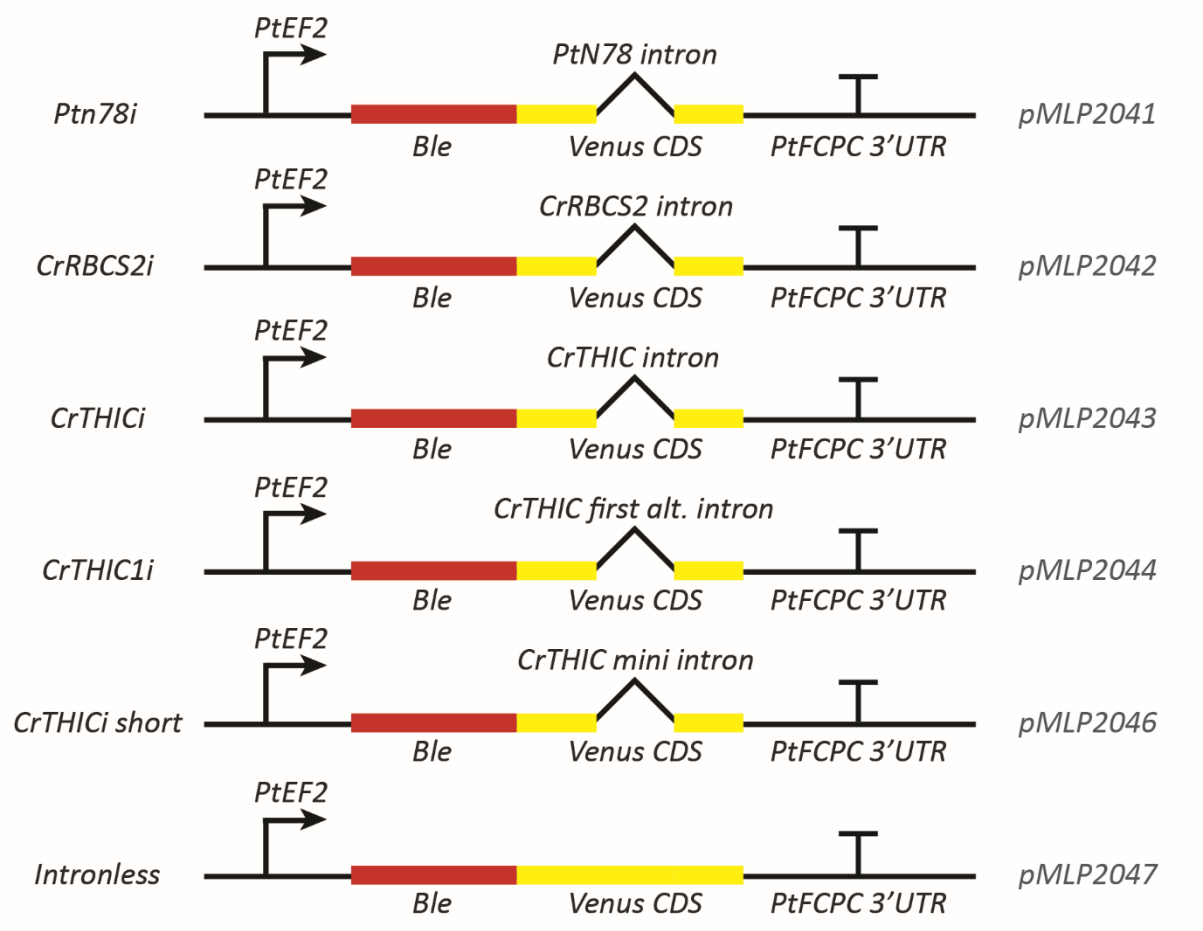


***Figure 4.9. Constructs used to test the* CrTHIC *riboswitch in* P. tricornutum.** *A splicing testing platform was set up by fusing the zeocin resistance gene (*Ble*) with a Venus fluorescent protein. Introns were inserted in a position known to correctly splice in the* Venus *CDS. A positive control with the native* PtN78 *intron was included, as well as a test with the widely used* CrRBCS2 intron*. Then, the full* CrTHIC *6th intron, including the riboswitch, the first alternative intron within the* CrTHIC *6th intron and a short version of the* CrTHIC *6th intron with only the 100 bp at the 5' and 3' most ends of the intron included. All constructs had a constitutive* PtEF2 *promoter and an* PtFCPC *terminator.*

***Table 4.1. Colonies recovered from the transformation with constructs to test* CrTHIC *intron function in* P. tricornutum.** *After electroporation cells were plated on zeocin selection plates and colonies counted 3 weeks later. Twelve colonies for each of the constructs with transformants were screened by confocal microscopy and the number of colonies expressing nuclear-targeted Venus counted. Fluorescence confocal images are shown in Figure 4.10.*

Construct	Intron	Intron length	Colonies on selection	Colonies screened by	Colonies expressing
pMLP2041	*PtN78*	78 bp	221	12	6
pMLP2042	*CrRBCS2*	145 bp	206	12	12
pMLP2043	*CrTHIC*	1636 bp	0	-	-
pMLP2044	*CrTHIC 1st*	604 bp	0	-	-
pMLP2046	*CrTHIC mini*	200 bp	0	-	-
pMLP2047	-	NA	29	12	2

Given that the *CrRBCS2* intron is used as an enhancer in *C. reinhardtii* constructs, it was significant that it could be incorporated and spliced in a *P. tricornutum* construct. To test whether the *CrRBCS2* intron also had an enhancer effect in *P. tricornutum* a dozen of the colonies for each of the *PtN78* (*pMLP2041*)*, CrRBCS2* (*pMLP2042*) and intronless (*pMLP2047*) constructs were grown to mid-exponential phase in f/2 in 96-well plates and screened for Venus fluorescence by confocal microscopy. The results showed that out of the 12 independent transformants tested for each construct, only 2 appeared to have observable Venus fluorescence for the intronless control (*pMLP2047*), and only 6 for the *PtN78* intron (*pMLP2041*). In contrast, all the colonies screened for *CrRBCS2* showed observable fluorescence (*pMLP2042*; Table 4.1). All constructs showed a nuclear localised Venus fluorescence as expected given the nuclear localisation target included in the *Ble* coding sequence, and there are no evident differences of fluorescence intensity between different constructs (Figure 4.10). These results would suggest that the heterologous intron could also have an enhancing effect in *P. tricornutum* and/or its splicing is more efficient than the controls.

To quantify the splicing efficiency of each of the *PtN78* and *CrRBCS2* introns, three independent transformants for each of the intronless (*pMLP2047*), *PtN78* (*pMLP2041*) and *CrRBCS2* (*pMLP2042*) constructs were grown to late exponential phase in 20 ml f/2 for 5 days. Cells were harvested, RNA extracted, reverse-transcribed, and a RT-qPCR with primers targeting the *Venus* CDS to measure total transcript levels and with primers targeting the exon-exon junction to measure the level of fully spliced transcripts, was carried out in collaboration with Nattaphorn Buayam (Figure 4.11a). The results showed no obvious difference in total transcript levels between the three constructs, and the splicing efficiency, calculated as the level of spliced *Venus* transcripts normalised by total *Venus* transcripts, is virtually 1 for each of the constructs (Figure 4.11b). These results give evidence that both the native

PtN78 and the *CrRBCS2* introns are fully spliced in the *P. tricornutum* heterologous expression cassette and the *CrRBCS2* intron does not have an enhancer effect in *P. tricornutum*.

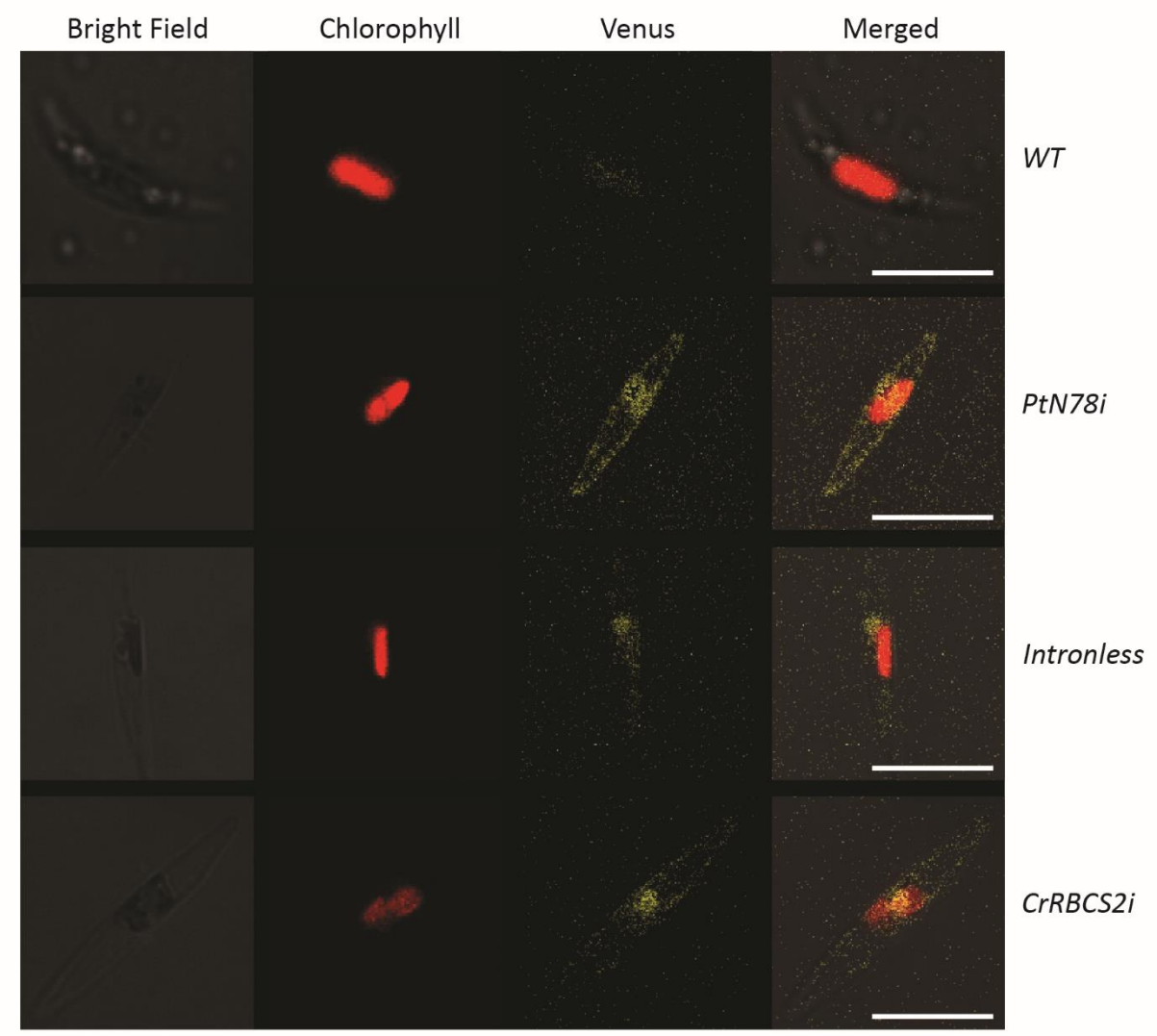


Figure 4.10. Confocal microscope fluorescent images of* P. tricornutum *carrying constructs testing introns. *Cells were grown for 3 days in f/2 without vitamins and 10 µl were visualised under a Leica SP5 fluorescence confocal microscope. A representative cell for each of the constructs that could be transformed (see Figure 4.9 and Table 4.1) is shown for each of the bright field, chlorophyll and Venus channels. The last column merges the three channels and the white bar corresponds to 10 µm.*

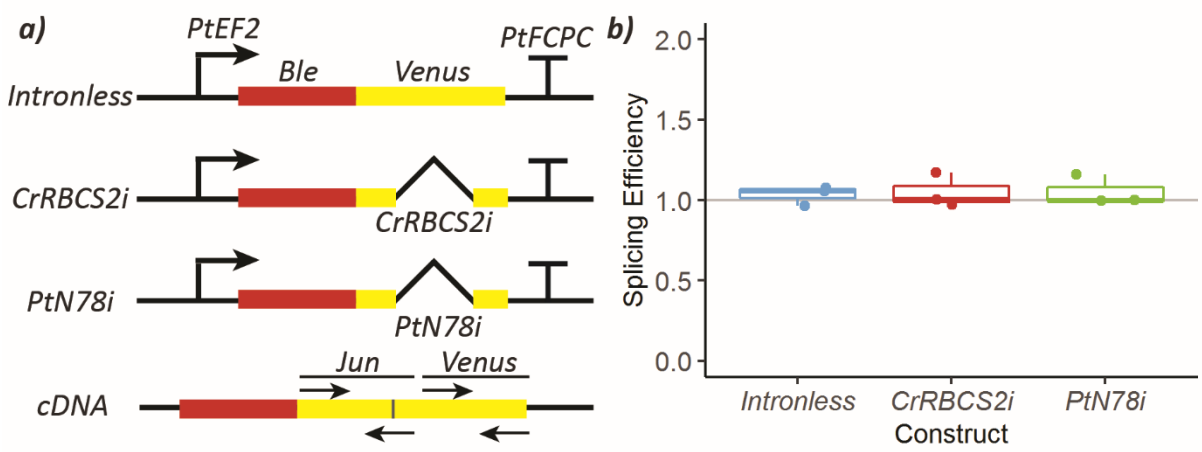


Figure 4.11. Splicing efficiency RT-qPCR measurement on* CrRBCS2 *intron in* P. tricornutum.** *Three independent transformants for an intronless control, a control with the native* PtN87 *intron and the construct with the heterologous* CrRBCS2 *intron were grown in f/2 media for 5 days, RNA was extracted and retrotranscribed.* ***a) *Primers targeting the* Venus *coding sequence (Venus) and a primer pair with the reverse primer annealing on the exon-exon junction (Jun) were used in technical duplicates in a RT-qPCR experiment.* ***b)*** *The splicing efficiency was calculated as the ratio between the relative level (ie. the amplification efficiency raised to the power of the Ct) of spliced transcripts (Jun primers) divided by the relative level of total transcript (Venus primers) and is plotted as boxplots for each construct. There are no statistical differences in means by T-tests between the intron controls and the intronless control.*

4.7. Chimeric *PtTHIC – CrTHIC* aptamers in the *CrTHI4* platform in *C. reinhardtii*

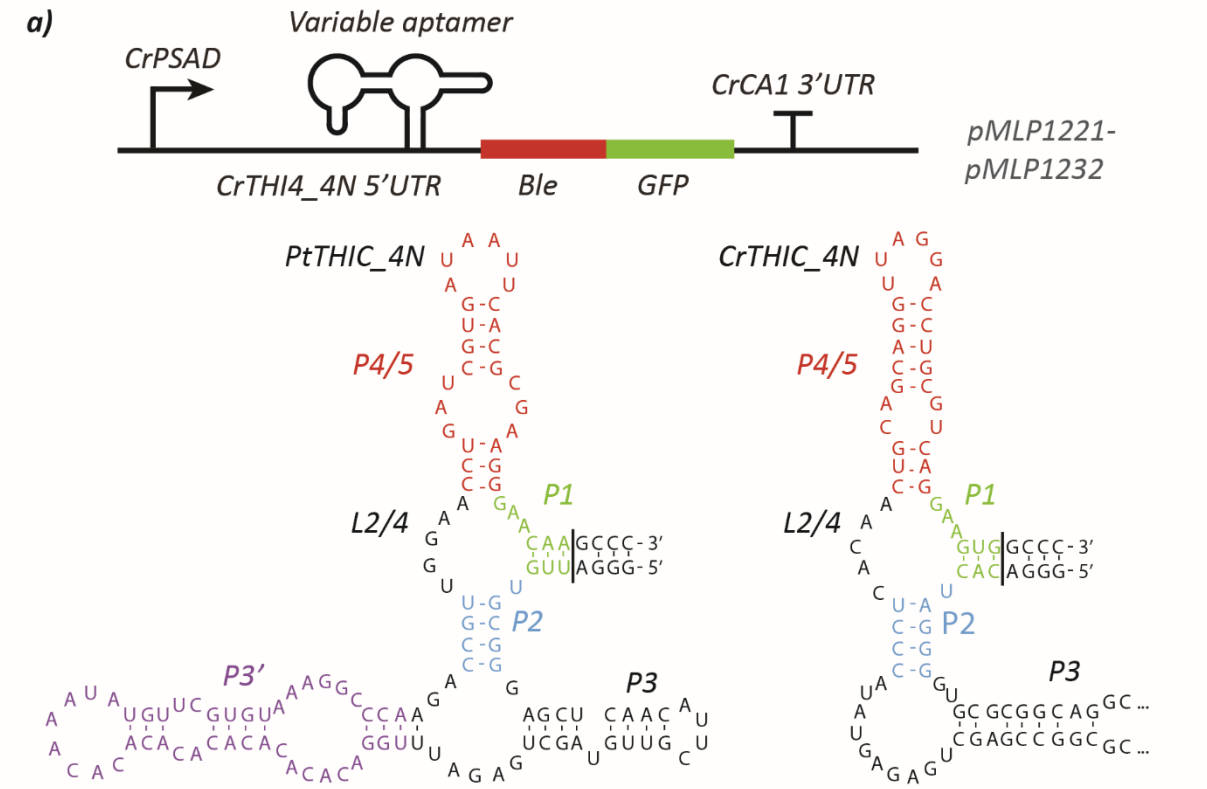
After observing that neither the native *PtTHIC* predicted riboswitch, nor the heterologous *CrTHI4_4N* and *CrTHIC* riboswitches could mediate a response to thiamine supplementation in *P. tricornutum*, it was proposed to investigate whether the function of the *PtTHIC* predicted aptamer could be rescued in *C. reinhardtii.* As detailed in section 4.4, the modular aptamer platform designed by Dr Ginnie Nguyen has already been shown to accommodate the *CrTHIC* aptamer (*CrTHIC_4N*) and the *AtTHIC* aptamer (*AtTHIC_4N*), therefore it was proposed that the functionality of the *PtTHIC* predicted aptamer could be tested in the *CrTHI4_4N* platform. Nevertheless, it is significant to highlight *CrTHIC* and *AtTHIC* aptamers share with *CrTHI4* an alternative splicing acceptor in their P2 stem key for the molecular mechanism of the riboswitch. As described in section 3.2.4, the *PtTHIC* predicted aptamer differs in key features with its counterparts in green algae and plants, including the absence of the splicing acceptor in the P2 stem and the presence of an additional P3' stem not found in *C. reinhardtii* or *A. thaliana* riboswitches. For this reason, apart from the basic design incorporating the native *PtTHIC* predicted aptamer into the platform (*pMLP1221*), a number of different chimeric aptamers combining structural domains of *PtTHIC* and *CrTHIC* aptamers were also added into the design (*pMLP1222-1231;* Figure 4.11b). The design also included a construct with all structural domains

matching the *CrTHIC* as a positive control (*pMLP1225*). All engineered aptamers were included in the *CrTHI4_4N* platform and cloned with a reporter *Ble-GFP* fusion, a *CrPSAD* promoter and a *CrCA1* terminator following the design used in section 4.4 (Figure 4.11a). *A priori*, with the *Ble* reporter the transformants with aptamers that had the ability to mediate a response to thiamine supplementation would see their growth impaired under the simultaneous presence of thiamine and zeocin.

In a first round, the constructs carrying the native *PtTHIC* aptamer (*pMLP1221*) or the chimeric aptamers with the P1, P2 or both from *CrTHIC* (*pMLP1222-1224*) were transformed into *C. reinhardtii UVM4.* Two weeks after electroporation between 30 and 59 colonies were picked into 96-well plates for each of the constructs. After two weeks of subculturing, five representative independent transformants were grown for 4 days in TAP supplemented with 10 $mg \cdot l^{-1}$ zeocin in a 96-well plate with or without 10 µM thiamine and OD_{730nm} was tracked in a plate reader as a proxy for growth. The results showed that all transformants could grow in the presence of zeocin with no evident growth differences between the thiamine-supplemented and the control condition for any of the constructs tested (Figure 4.11c). The results confirmed that, unlike the transformation of *CrTHI4_4N* and *CrTHIC_4N* riboswitches in *P. tricornutum* in sections 4.5 and 4.6, the *PtTHIC* predicted aptamer did not compromise expression in a *C. reinhardtii* heterologous expression cassette. The results further suggest that neither the *PtTHIC* predicted aptamer (*pMLP1221*) nor any of the P1/2 chimeras with the *CrTHIC* aptamer (*pMLP1222-1224*) are able to mediate a response to thiamine supplementation.

A second iteration of the experiment was set up to test whether differences in the P4/5 or the presence of the P3' in the *PtTHIC* predicted aptamer could prevent the aptamer from providing a response in the *CrTHI4* platform. Constructs with chimeric aptamers with P1/2 and/or P4/5 stems from *CrTHIC* and without P3' stems in all possible combinations (*pMLP1226-1231*) were cloned into the platform and transformed into *C. reinhardtii* in parallel with the full *CrTHIC* aptamer as a control (*pMLP1225*). Two weeks after transformation, between 15 and 50 colonies were picked into 96-well plates for each of the constructs. Five representative independent transformants for each of the constructs was grown for 9 days in the presence of 10 $mg \cdot l^{-1}$ zeocin in a 96-well plate with or without 10 µM thiamine and OD_{730nm} was tracked in a plate reader as a proxy for growth. The results showed that while the thiamine treatment clearly affects the growth of the positive control with the *CrTHIC* aptamer (*pMLP1225*), no growth detriment is observed for any of the chimeric aptamers (Figure 4.11c). Furthermore, it was observed that the positive control with the *CrTHIC* aptamer (*pMLP1225*) repressed expression to a certain extent compared to the *PtTHIC* aptamer even in the thiamine deplete condition. These results demonstrate that even when removing the P3' and swapping stems P1, P2, P4 and P5 for the *CrTHIC* aptamer, the predicted *PtTHIC* aptamer cannot mediate a response to thiamine supplementation in the *CrTHI4_4N* aptamer platform. Only the 5 bp in the L2/4 loop and

the P3 stem differ between the positive control (*pMLP1225*) and the chimeric aptamer resembling *CrTHIC* the most (*pMLP1229;* Figure 4.11b). One or both of the L2/4 or P3 domains must be necessary for the *CrTHIC* aptamer to give a response to thiamine supplementation. The downregulation of the heterologous reporter under the absence of thiamine for the *CrTHIC* aptamer has been observed in equivalent experiments by other members of the group and it is presumably due to the effects of intracellular TPP levels switching off a fraction of the transcripts (Mendoza-Ochoa *et al.,* in preparation).



b)

Plasmid	P1	P2	P3	P3'	L2/4	P4/5
pMLP1221	Pt	Pt	Pt	Pt	Pt	Pt
pMLP1222	Pt	Cr	Pt	Pt	Pt	Pt
pMLP1223	Cr	Pt	Pt	Pt	Pt	Pt
pMLP1224	Cr	Cr	Pt	Pt	Pt	Pt
pMLP1225	Cr	Cr	Cr	-	Cr	Cr
pMLP1226	Pt	Pt	Pt	-	Pt	Pt
pMLP1227	Cr	Cr	Pt	-	Pt	Pt
pMLP1228	Pt	Pt	Pt	-	Pt	Cr
pMLP1229	Cr	Cr	Pt	-	Pt	Cr
pMLP1230	Pt	Pt	Pt	Pt	Pt	Cr
pMLP1231	Cr	Cr	Pt	Pt	Pt	Cr

Figure 4.12.* PtTHIC – CrTHIC *chimeric aptamers tested in the* C. reinhardtii *aptamer platform.* a)** *A heterologous expression platform with a* CrPSAD *promoter, the zeocin resistance gene* ble *fused with* eGFP *and the* CrCA1 *terminator included the* CrTHI4 *5'UTR adapted in a way that the aptamer could be swapped by other aptamers.* **b)** *Chimeric aptamers were built combining different structural parts from* PtTHIC *(brown) and* CrTHIC *(green) aptamers in different combinations. pMLP1221 is the full* PtTHIC *aptamer, and pMLP1225 is the full* CrTHIC *aptamer. These chimeric riboswitches were cloned in the aptamer platform.* **c)** *Zeocin sensitivity assay on* C. reinhardtii *transformants carrying the 11 different chimeric aptamers in the aptamer platform. Cells were grown in the presence (red) or absence (blue) of 10 µM thiamine in the presence of 10 mg·l^{-1} zeocin and OD_{730nm} was measured every 24h in a plate reader.* ***Continues in next page.

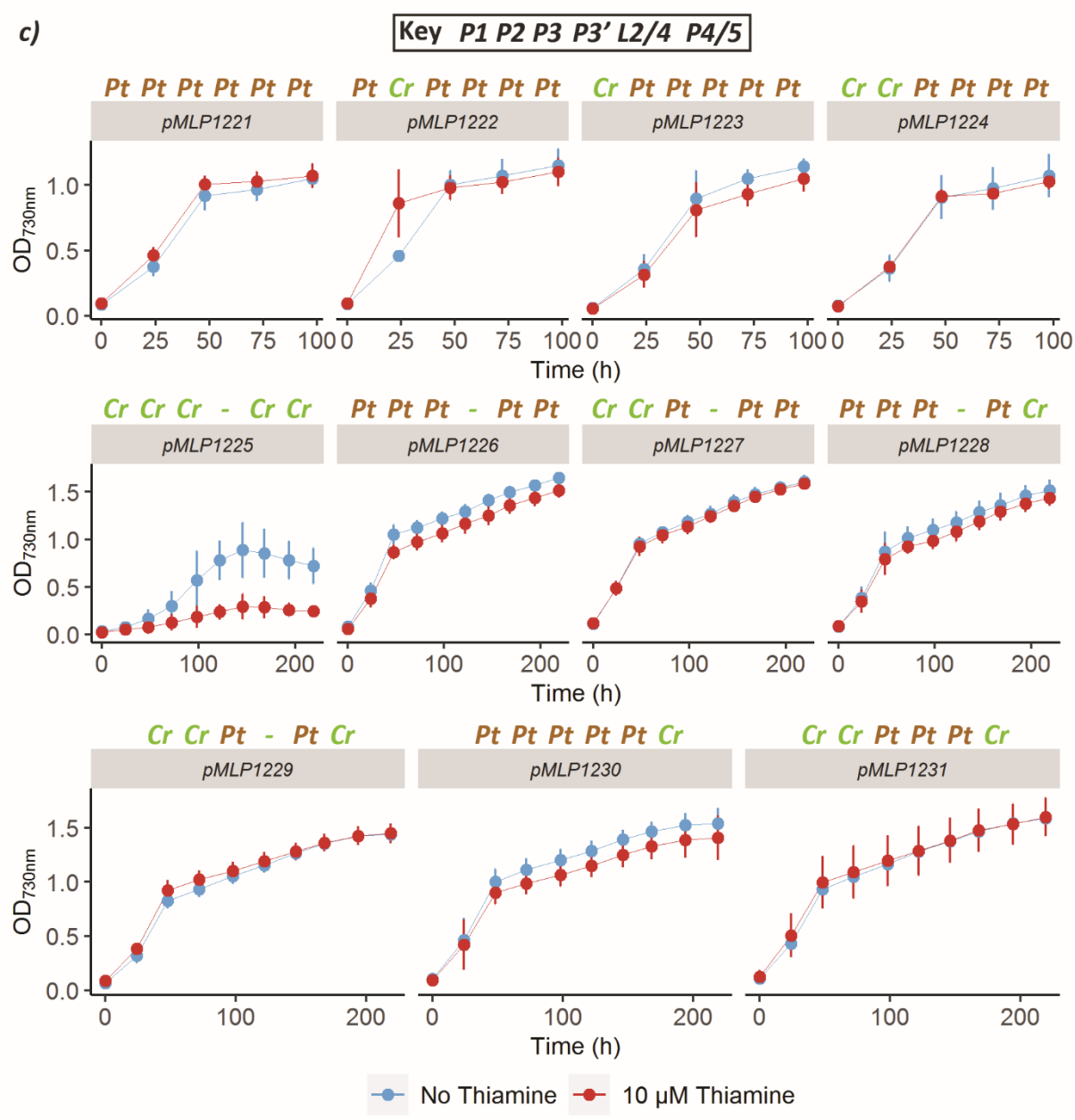


Figure 4.12. Continued.

4.8. Putative cobalamin-responsive motif is necessary for expression but not sufficient for regulation

The bioinformatic analysis in section 3.2.6 revealed the presence of a palindromic 14 bp motif upstream of genes known to be downregulated by cobalamin in diatoms, present in four copies upstream of *PtMETE*. This motif was also identified in the promoter region of all *THIC* genes in diatoms (Table 3.3). To experimentally test whether the motif mediated the promoter response to cobalamin supplementation I built a set of reporter constructs in collaboration with Shelby Newsad, another PhD student in the group, based on previous constructs designed by Patrick Hickland to study the *PtMETE* promoter. The 359 bp immediately upstream of the *PtMETE* start codon, which included two of the putative motifs and had previously been shown to drive expression in response to cobalamin by Patrick Hickland in our group, was used as a positive control promoter to drive the expression of a Venus reporter with a *PtFCPC* terminator (*pMLP2131*, Figure 4.12). A construct with a promoter with both 14 bp motifs sequences randomly scrambled was added to test whether the motifs were necessary for regulation (*pMLP2132*). Constructs with the *PtFCPA* constitutive promoter alone, with one or two added motifs were generated to test whether the motifs were sufficient to provide cobalamin responsiveness (*pMLP2133-2136*). All constructs were electroporated into *P. tricornutum* and 3 weeks later 14 to 37 independent colonies were picked into 96-well plates for each of the constructs. After two weeks of subculturing, all transformants were grown in the presence and absence of 1 $\mu g \cdot l^{-1}$ cobalamin in 200 µl and their Venus fluorescence normalised by OD_{730nm} was recorded for 7 days with a Clariostar plate reader.

The results showed that all constructs except the *PtMETE* promoter with scrambled motives (*pMLP2132*) exhibited detectable Venus fluorescence (Figure 4.14a). Only the control construct with the unmutated *PtMETE* promoter showed a clear downregulation of Venus fluorescence under cobalamin supplementation (*pMLP2131*) while all the constructs based on the *PtFCPA* promoter (*pMLP2133-2136*) showed similar expression patterns both in the presence and absence of cobalamin. In agreement with previous experiments, relative fluorescence values for both *PtMETE* (*pMLP2131*) and *PtFCPA* (*pMLP2133-2136*) reached their peak intensity at 3 days post-inoculation (Hickland *et al.,* in preparation). The large standard deviation of the measurement prevented us from finding statistically significant differences between constructs (Tuckey HSD test with 0.05 confidence level for normalised fluorescence at day 3 by construct excluding *pMLP2132*), but the average expression level of the *PtMETE* promoter under cobalamin deplete conditions (*pMLP2131*) appeared to be higher than the *PtFCPA* promoter (*pMLP2133-2136*) with equivalent results reported by Patrick Hickland. The standard deviation was due to the high variability between the independent transformants likely explained by positional effects of the random construct integration. The variation could be clearly

observed for the positive control constructs (*pMLP2131*), where four independent transformants showed at least twice the maximum fluorescence level compared to the other transformants, while one of the transformants showed no Venus fluorescence (Figure 4.13b). These results lead to the conclusion that the 14 bp motif found in the *PtMETE* promoter are necessary for gene expression, but not sufficient to confer cobalamin-responsive regulation when cloned upstream of the *PtFCPA.*

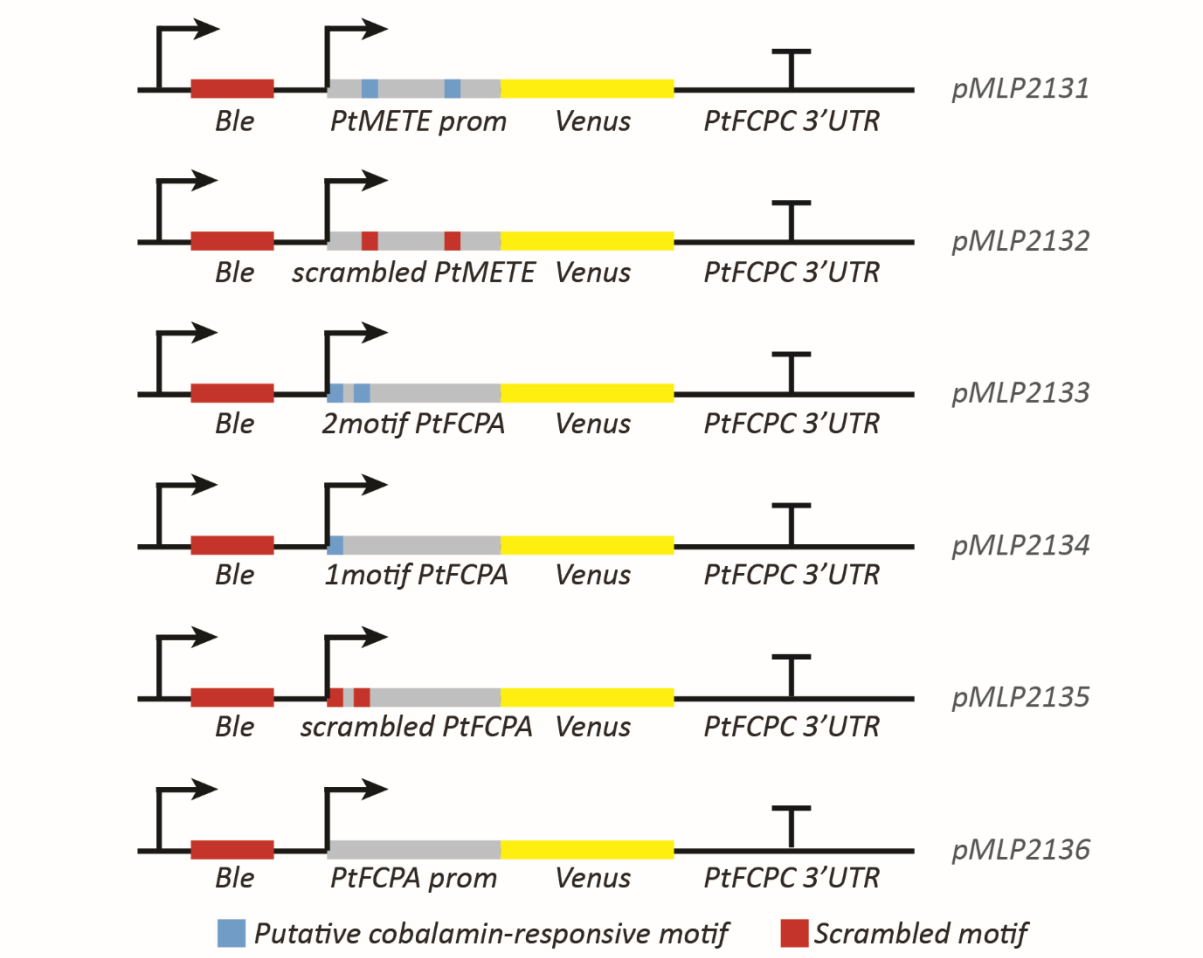


Figure 4.13 Construct design to test the putative cobalamin-responsive motif in* PtMETE *truncated promoter. *pMLP2131 contains the* PtMETE *promoter 326 bp truncation with no mutation, pMLP2132 is the same construct with targeted scrambling of the two 14 bp motifs. pMLP2136 is a negative control with the constitutive* PtFCPA *promoter, and pMLP2133-2135 contain the* PtFCPA *promoter with two, one or two scrambled motifs added to the 5'end of the promoter respectively. When two motifs are added, there is a 17 bp spacer between them.*

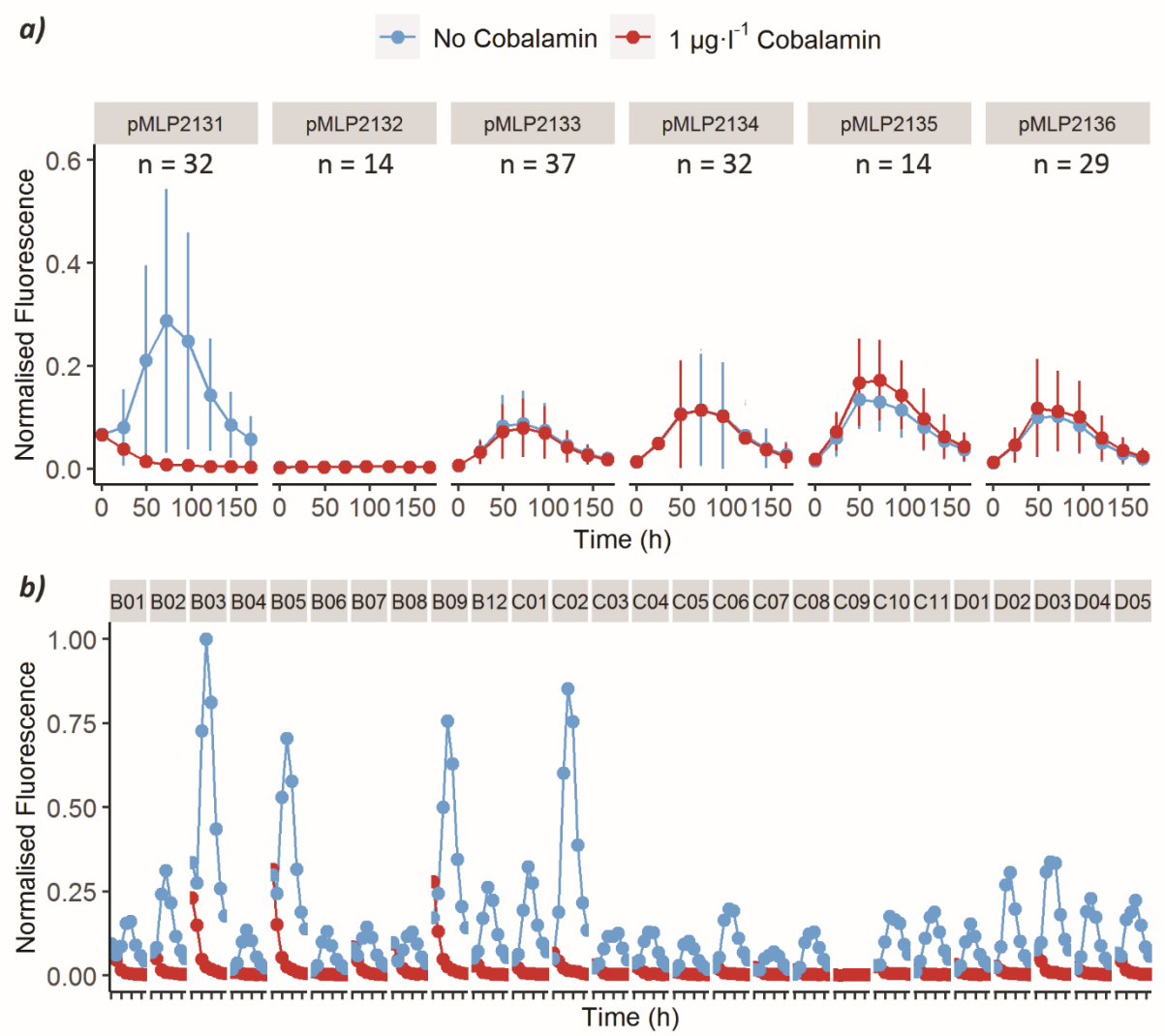


Figure 4.14. Time course of Venus fluorescence response to cobalamin supplementation in motif reporter constructs. a) *The Venus fluorescence of a varying number (n) of independent transformants for each of the reporter constructs in figure 4.13 was tracked for 7 days in the presence or absence of 1 mg·l^{-1} cobalamin in 96-well plates. Venus fluorescence is normalised to chlorophyll fluorescence for each individual measurement. Error bars represent the standard deviation of all independent transformants tested.* ***b)*** *Time course of Venus fluorescence response to cobalamin supplementation in pMLP2131 independent transformants. Shows the same data as in panel a) for each pMLP2131 individual transformant. The absolute fluorescence intensity is highly variable between independent transformants, but the response to cobalamin pattern is highly reproducible across all transformants.*

4.9. Discussion

This chapter has taken advantage of the use of reporter constructs to investigate in some detail the possible function of the predicted *PtTHIC* aptamer. The robust transformation, screening and testing pipeline set up in the lab which involved cloning constructs with multiple expression cassettes with the MoClo system in 2 weeks, electroporation and selection in 2-3 weeks, growth and stabilisation of up to hundreds of independent transformants in 96-well plate cultures in two more weeks, followed by screening and measuring reporter outputs with the Clariostar plate reader made it possible to address most biological questions with reporter constructs in as little as 8 weeks (Figure 4.1). The first set of constructs was based on the luciferase reporter. Unfortunately, this did not give enough independent transformants to test whether the *PtTHIC* 3'UTR could regulate transgene expression. It is unknown what led to the undetectable expression of luciferase, but perhaps the combination of a 5'UTR and a 3'UTR from unrelated genes has a negative impact on expression. The only construct to show more than one transformant with luciferase expression combined the *PtFCPA* 5'UTR with the *PtFCPC* 3'UTR, while the constructs that had a combination of *PtFCPA* 5'UTR and *PtTHIC* 3'UTR led at most to a single luciferase-expressing transformant (Figure 4.2). A similar trend can be observed in the *PtTHIC* overexpressing strains, when the *PtTHIC* 5'UTR and 3'UTR are combined (*pMLP2065*) the expression appears to be higher than the combination of the *PtTHIC* 5'UTR with the *PtFCPC* 3'UTR (*pMLP2065*) as detected by RT-PCR (Figure 4.5).

It is well understood that the contacts between 5' and 3' UTRs are important for translation and their interaction may be important for the regulation of translation initiation (Mazumder *et al.*, 2003). In previous experiments, we have shown that *C. reinhardtii* transformants with a *ble-eGFP* reporter controlled by the *CrTHI4* 5'UTR riboswitch cannot grow in the simultaneous presence of zeocin and thiamine given that thiamine effectively downregulates the expression of the antibiotic resistance reporter (Mehrshahi *et al.,* 2020). Therefore, in a second design iteration the *PtTHIC* promoter, 5' and 3' UTRs were cloned driving the expression of a *ble-Venus* reporter. There are both disadvantages and advantages of using *ble* over luciferase as a reporter. While luciferase provides a direct measurement reading of expression, and acts independently of the selection pressure, the *ble* reporter only allows for the measurement of OD_{730nm} as a proxy for zeocin resistance. On the other hand, the output of the *ble* reporter is easier and faster to measure and guarantees that all transformants selected on zeocin plates will be suitable to test for thiamine response. In addition, *ble* can be translationally fused to *Venus* to provide a direct quantification method that can be easily monitored in the Clariostar plate reader which can track growth and fluorescence intensity of hundreds of independent transformants over time with little effort. In contrast to the *CrTHI4* 5'UTR in *C. reinhardtii*, the *PtTHIC* promoter and UTRs do not regulate the expression of the *ble* reporter in response to thiamine supplementation in

P. tricornutum and the transformants do not show any growth anomaly compared to untreated controls (Figure 4.3).

All the Chapter 3 experiments performed on the wild type endogenous *PtTHIC* gene and the reporter constructs in this chapter consistently show that *PtTHIC* expression does not respond to thiamine supplementation. Nevertheless, a 10 µM thiamine supplementation does not reflect a realistic environmental scenario as levels measured in ocean samples range between 0 and 300 pM in ocean samples (Sañudo-Wilhelmy *et al.,* 2012) and thiamine requirements for marine algae auxotrophs can be met at picomolar concentrations (Gutowska *et al.,* 2017). Additionally, the likely function of the TPP riboswitch is to keep homeostasis of thiamine levels by regulating the biosynthesis pathway in response to its own product (Moulin *et al.,* 2013). With these two premises at hand, I turned my attention to mutagenesis experiments performed in various organisms: the *C. reinhardtii pyr1* mutant has a mutation in the *CrTHI4* riboswitch that leads to a deregulation of the biosynthesis pathway leading to increased intracellular thiamine levels and higher *CrTHI4* expression (Moulin *et al.,* 2013). Similarly, the engineered riboswitch mutant *B. subtilis 1A481* fails to regulate a reporter construct in response to thiamine supplementation and the *A. oryzae PtrA* mutant drastically reduces the affinity of the riboswitch to TPP (Sudarsan *et al.,* 2005). All three organisms are susceptible to the antimetabolite pyrithiamine, whereas *C. reinhardtii pyr1, B. subtilis 1A481* and *A. oryzae ptrA* mutants are resistant to the compound, implying that the antimetabolite target is the TPP riboswitch. In contrast with these examples, the transformants with an extra copy of *PtTHIC* with an equivalent mutated aptamer did not show the same phenotype in *P. tricornutum*. The transformants with a mutated *PtTHIC* aptamer did not accumulate more thiamine or TPP intracellularly and did not express higher PtTHIC protein levels compared to an unmutated control (Figures 4.4 and 4.5). These results corroborate the previous conclusions from Chapter 3 where *P. tricornutum* was shown not to respond to pyrithiamine, namely that the riboswitch does not respond to intracellular levels of thiamine.

Having confirmed that the native *PtTHIC* riboswitch is not responsive to thiamine supplementation, I attempted to import the *CrTHI4* and *CrTHIC* heterologous riboswitches from *C. reinhardtii* to establish an RNA-based regulatory part that could be used to regulate synthetic biology designs in *P. tricornutum*. Both *CrTHI4* and *CrTHIC* heterologous riboswitches failed to be expressed in *P. tricornutum* possibly due to species-specific differences in the basic transcription and splicing machinery (Figure 4.8; Table 4.1). These results highlight the fact that transferring eukaryotic riboswitches between relatively distant species is not trivial. In general, riboswitches are considered *cis*-acting regulatory elements, but they cannot be studied or understood outside of their biological context, as their expression platform mechanism generally requires interaction with the transcription, splicing and translation machinery. Transcription, splicing and translation processes are conserved in

eukaryote systems, but their functioning is finely tuned for each specific species making it challenging to transfer riboswitch elements interspecifically as demonstrated by the experiments presented in this chapter and in agreement with the design challenges to integrate aptamer domains with expression platforms to assemble new riboswitches with synthetic biology applications (Liang *et al.*, 2011).

Nevertheless, in attempting to import the *CrTHIC* intron to *P. tricornutum* I serendipitously discovered that the widely used *CrRBCS2* first intron (*CrRBCS2i*) from *C. reinhardtii* could be efficiently spliced in *P. tricornutum* (Figure 4.11). Constructs expressing enzymes for metabolic engineering are increasingly incorporating the *CrRBCS2i* in *C. reinhardtii* given its expression enhancement effect (Baier *et al.,* 2018; Lauersen *et al.,* 2018). Current efforts in the group are trying to reproduce some of the engineered metabolic pathways from *C. reinhardtii* in *P. tricornutum* and knowing that the *CrRBCS2i* can be efficiently spliced in *P. tricornutum* has allowed the shuttling of optimised parts encoding metabolic enzymes from one species to the other with ease. However, the enhancer effect of CrBBCS2i seen in *C. reinhardtii* is not carried over to *P. tricornutum*.

The generation of chimeric riboswitches incorporating domains from the *PtTHIC* and the *CrTHIC* aptamer in the *CrTHI4_4N* platform was designed to investigate which functional domains of the predicted *PtTHIC* aptamer were responsible for the lack of response to thiamine. However, experiments in *C. reinhardtii* did not allow us to discern which part of the *PtTHIC* riboswitch was responsible for the lack of thiamine response as none of the chimeric riboswitches could trigger a response to thiamine supplementation. Interestingly, even the chimeric riboswitch that contained just the 5 bp L2/4 loop and the P3 stem from *PtTHIC* did not show any response to thiamine (Figure 4.12). It logically follows that either the *CrTHIC* L2/4, its P3 stem, or both, are essential for the molecular mechanism of the *C. reinhardtii* riboswitch. The P3 stem is the first fully transcribed stem in the aptamer and has the highest folding energy of all stems, it has therefore been argued that it acts as a nucleation core for the folding of the rest of the aptamer (Anthony *et al.,* 2012) and it is not known to interact with the ligand or play a role in the interaction with the expression platform as seen in the tertiary structure of the *AtTHIC* aptamer (Thore *et al.,* 2006). In their discussion of the 3D structure, Thore *et al.* (2006) highlight that the J2/4 plays an important structural part in the closure of the "clamp" between the P2/3 and P4/5 helix and forms sequence-specific base-pairing interactions with residues in the P2 and P4 stem, especially through the G42 residue, which is interestingly conserved in *PtTHIC* but not in *CrTHIC.* Interestingly, the *AtTHIC* aptamer L2/4 only differs in one residue with the *PtTHIC* L2/4 (A to G; see Figure 1.4) but in contrast the *AtTHIC_4N* chimeric riboswitch has been shown to respond to thiamine in *C. reinhardtii* (Mehrshahi *et al.,* 2020). In conclusion, I hypothesise that it is the *PtTHIC* L2/4, and not the P3, which is interfering with the *CrTHIC* aptamer function. This hypothesis

could be tested by inverting the logic of the chimeric aptamer experiment, maintaining all the *CrTHIC* parts and only swapping in one domain at a time from *PtTHIC,* specifically J2/4 and P3. In sum, more experiments are needed to understand the mechanism of action of algal riboswitches, and techniques such as RNA probing and ligand-bound structural determination could help us gain important insights to explain the molecular basis for the lack of function of the *PtTHIC* predicted aptamer.

Finally, the cobalamin-response assays with the *PtTHIC* overexpressors showed that despite the mutation in the predicted aptamer *PtTHIC* was still downregulated by cobalamin confirming that the riboswitch was not involved in cobalamin sensing. The overexpressor construct with the *PtFCPC* 3'UTR in place of the *PtTHIC* 3'UTR further confirmed that cobalamin regulation was not based in the 3'UTR (Figure 4.6). This is in agreement with the findings of a conserved motif in *PtTHIC* that is found in the promoters of cobalamin-downregulated genes in diatoms (Section 3.2.6). The experiments with the *PtMETE* promoter with the scrambled motif experimentally confirm that the putative motif is necessary for expression, although the motif alone fails to confer cobalamin-responsiveness to the *PtFCPA* constitutive promoter (Figure 4.14). The discovery of the putative cobalamin-responsive motif has added new insights to the work led by Patrick Hickland in the group and is included in a manuscript addressing transcriptional cobalamin regulation in *P. tricornutum* (Hickland *et al.,* in preparation). Follow up experiments could use further permutations of reporter constructs to find which accessory sequences flanking the motif are required for an effective cobalamin response, and Yeast One hybrid (Y1H) and Electrophoretic Mobility Shift Assay (EMSA) techniques could be used to find the putative transcription factor counterpart to the motif.

In sum, the main results of this chapter confirm that *PtTHIC* is not regulated by thiamine supplementation or intracellular thiamine but is downregulated by cobalamin supplementation in agreement with the conclusions drawn from the experiments performed in wild type *P. tricornutum* and presented in Chapter 3. The results thus confirm that the conservation of the aptamer sequence across the diatom lineage alone does not logically lead to the annotation of a functional riboswitch. This is important to consider when applying bioinformatic approaches to predict the presence of functional riboswitches based on sequence and secondary structure conservation alone, given that the assumptions taken can lead to the missannotation of riboswitches. It is still unknown what are the underlying environmental conditions that facilitated the evolution of *THIC* expression unresponsive to thiamine supplementation in *P. tricornutum* given that thiamine biosynthesis is an expensive biosynthetic process, thiamine is only required at nanomolar concentrations and all eukaryotic organisms studied show strong feedback downregulation in thiamine biosynthetic genes (Hanson *et al.,* 2018; Wachter*,* 2010). On the molecular side of the evolutionary question, it will be interesting to discover which is the alternative functional basis that explains the high sequence conservation of the

predicted TPP aptamer in diatoms considering its function is not to mediate a genetic response to thiamine supplementation. These questions are relevant to address as they would inform a growing interest in the role that vitamin biosynthesis and auxotrophy play in the composition and dynamic of marine microbial communities with a key contribution to global biogeochemical cycles (Heal *et al.,* 2017; Kazamia *et al.,* 2016; Sañudo-Wilhelmy *et al.,* 2014).

Chapter 5. Developing CRISPR/Cas9 to study thiamine and cobalamin metabolism in *P. tricornutum*

5.1 Introduction

During the course of my PhD project genome editing techniques with CRISPR/Cas9 have been increasingly adopted and developed by the diatom research community. CRISPR/Cas9 relies on the action of the Cas9 endonuclease from *S. pyogenes* to generate a double strand break at a genomic position defined by sequence complementarity to a sgRNA. The double strand break can be naturally repaired by NHEJ generating small indels leading to frame-shift mutations, or by Homology Directed Repair (HDR) generating precise mutations based on a given homology repair template (Doudna & Charpentier, 2014; Figure 5.1b). The technical developments in *P. tricornutum* so far have been in three main categories: Cas9 delivery, type of mutation induced and mutant screening method.

For the Cas9 delivery system, there are three main methods that have been used so far. The first CRISPR/Cas9 genome editing report in *P. tricornutum* used microparticle bombardment to transform a construct encoding Cas9 and the sgRNA to be integrated and expressed from the genome (Nymark *et al.,* 2016). The sgRNA was designed to knock-out *SRP54,* a gene involved in the integration of proteins in the thylakoid membrane which gave a light-sensitivity phenotype easy to identify, and used a High Resolution Melting (HRM) PCR assay to detect edited genotypes (HRM is explained in greater detail below). The random integration in the genome of the Cas9 and sgRNA construct has potential deleterious effects, and additionally Cas9 is expected to be expressed continuously increasing the potential for off-target effects and interference with the phenotype of the edited lines. To solve these bottlenecks, a second approach was adopted in which the Cas9 and sgRNA expression cassettes were delivered in non-integrative episomes by bacterial conjugation (Figure 5.1a; Sharma *et al.,* 2018; Slattery *et al.,* 2018). This transformation approach was reported to be less efficient at giving genome-edited strains, but produced nearly 50 times more colonies to start with in comparison with microparticle bombardment delivery (Sharma *et al.,* 2018). The most significant improvement with the bacterial conjugation delivery is that the episome is not integrated in the genome and can be lost when the antibiotic selection is removed, leading to transgenic-free edited strains with limited off-target effects. Finally, a third method based on RNP delivery has been introduced to overcome the limitations of stable Cas9 expression and giving the opportunity to deliver several sgRNAs targeting

up to three genes in one microparticle bombardment shot (Serif *et al.,* 2018). The main drawback of this third method is the high cost of the reagents used.

The types of mutation introduced by CRISPR have also been constantly developing in *P. tricornutum* and influence the kind of genotyping screening techniques that can be used as well as the degree to which mutations can be rationally designed (Figure 5.1b). In the first CRISPR report in *P. tricornutum* Cas9 was used with a single sgRNA to create a single double strand break, which was repaired by NHEJ and lead to random small indels in the target site of the sgRNA (Nymark *et al.,* 2016). The method was improved by using the TevCas9 nuclease instead of the standard Cas9 (Slattery *et al.,* 2018). This modified nuclease is a fusion of Cas9 with the Tev nuclease and generates two double strand breaks that remove most of the target site sequence avoiding cycles of restriction-ligation and biasing the cleavage to more defined length deletions (Wolfs *et al.,* 2016). The next improvement came with the delivery of a pair of sgRNAs targeting a single gene of interest (GOI) at short distances (roughly 100 bp). This approach generally leads to defined short deletions that are easier to screen and give a more certain knock-out genotype when deleting a longer sequence of the gene, if not causing a frame-shift mutation (Hopes *et al.,* 2016; Serif *et al.,* 2018).

Most recently, the first peer-reviewed reports of CRISPR-mediated homologous recombination in diatoms have been published by which a plasmid encoding Cas9 and a sgRNA is co-transformed by microparticle bombardment with an homologous recombination template including a resistance cassette flanked by homology arms (Figure 5.1; Moosburner *et al.,* 2020). With this approach, CRISPR-Cas9 double strand breaks are used to induce HDR leading to the introduction of a selection cassette in the middle of the coding sequence of the GOI in a rationally designed and controlled manner. This approach not only facilitates genotyping by a simple PCR but most importantly opens a wide diversity of possibilities from introducing landing sites for routine and standardised genomic integration, to the precise editing of a GOI, to the introduction of novel regulatory sequences into existing genes for conditional expression of endogenous genes. There are preprint reports using a similar approach in *T. pseudonana* (Belshaw *et al.,* 2017 - preprint) in which I based my homologous recombination design in 2018, described here, seeking to pioneer the CRISPR-mediated homologous recombination technique in *P. tricornutum*.

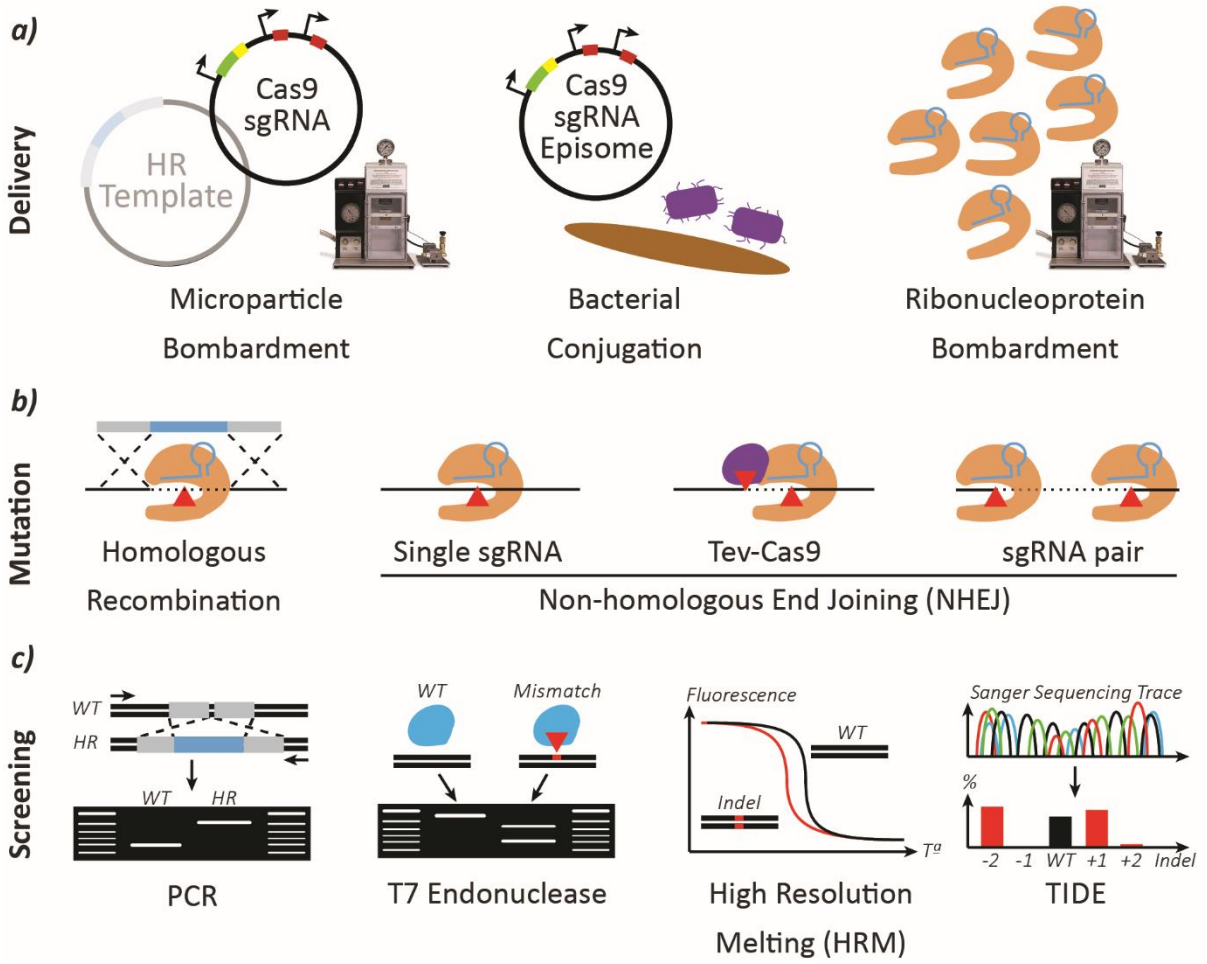


***Figure 5.1. Schematic representation of the different CRIPSR/Cas9 delivery, mutagenesis and screening methods used in* P. tricornutum.**

A further issue is that the screening of CRISPR mutants in *P. tricornutum* faces the challenge of colony mosaicism, by which a single colony transformed with Cas9 and sgRNAs will include a population of different genotypes, which is thought to be the result of Cas9 acting in successive early divisions of the transformed cell (Huang & Daboussi, 2017). To the best of my knowledge, all reports of CRISPR editing in *P. tricornutum* re-streak the genome-edited primary colonies to obtain monoclonal secondary colonies to overcome this challenge (Moosburner *et al.,* 2020; Nymark *et al.,* 2016; Serif *et al.,* 2018, Sharma *et al.,* 2018; Slattery *et al.,* 2018; Stuckenberg *et al.,* 2018). Finding an alternative solution to colony mosaicism could save up to 3 weeks in the process of generating genome editions in *P. tricornutum* and remains a technical problem of interest.

Finally, given that diatoms are diploid organisms, a further challenge is the need to identify clones with biallelic disruption, so it is essential to genotype both alleles independently to obtain a complete knock-out. With the objective of speeding up the screening for knock-out mutants, different

phenotyping and genotyping approaches have been employed (Figure 5.1c). For instance, target genes whose biallelic knock-out mutants can be selected positively by resistance to a certain molecule or negatively by auxotrophy for particular metabolites have been used to select strains with both alleles mutated. Some of the targets used include *PtUMPS* whose knock-out leads to resistance to 5-FOA and uracil auxotrophy, *PtAPT* whose knock-out leads to 2-fluoroadenine resistance and adenine auxotrophy, *PtPRA-PH* whose knock-out leads to histidine auxotrophy, or the urease gene which leads to failure to grow on urea (Serif *et al.,* 2018; Slattery *et al.,* 2018; Slattery *et al,* 2020). Early genotyping approaches to differentiate biallelic knock-out mutants from WT or heterozygous mutants include HRM assays for short indels undetectable by PCR (Nymark *et al.,* 2016; Sharma *et al.,* 2018), and T7 endonuclease assays to detect mismatches, short indels or point mutations not detectable by PCR (Moosburner *et al.,* 2020; Slattery *et al.,* 2018). However, with the adoption of sgRNA pairs to generate longer indels and homologous recombination designer mutagenesis to incorporate resistance cassettes at the target sites, it has become more straightforward to genotype mutants by standard PCR given that the mutations can be detected as length differences in agarose gels (Hopes *et al.,* 2016; Moosburner *et al.,* 2020; Serif *et al.,* 2018). Finally, the challenge of identifying multiple genotypes in the same colony and differentiating the two alleles in monoclonal colonies has been addressed by the use of the Track INDELs by DEcomposition (TIDE) algorithm, which deconvolutes compounded Sanger sequencing chromatograms into individual short indel mutations with an estimated relative abundance for each indel length (Brinkman *et al.,* 2014; Moosburner *et al.*, 2020; Serif *et al.,* 2018).

These many advancements in genome editing technologies and optimised CRISPR strategies in *P. tricornutum* open a range of opportunities both for basic physiological research and for metabolic engineering purposes. Specifically in our research group, I thought the application of the CRISPR/Cas9 genome editing techniques could help our research group further understand the physiological role of genes involved in thiamine and cobalamin metabolism, such as the genes whose regulation I have studied in Chapters 3 and 4. In addition, there are members of the group interested in developing *P. tricornutum* as a metabolic engineering platform to produce high-value terpenoids. For this purpose, genome editing offers the possibility to better understand terpene metabolism, and to knock-out endogenous pathways that compete with the introduced pathways. This is why in this chapter I have tried to knock-out a geranylgeranyl-pyrophosphate (GGPP) synthase (*GGPPS, Phatr3_J47271*) and a phytoene synthase (*PSY, Phatr3_EG02349*) involved in isoprenoid biosynthesis. *P. tricornutum* has two *GGPPS* encoded in its genome, one (*Phatr3_J19000*) is targeted to the chloroplast, is homologous to *GGPPS* found in bacteria and plants, and is thought to mostly contribute to the production of carotenoids and other pigments. The other (*Phatr3_J47271*), which I have targeted in this chapter, is

a putative cytosolic *GGPPS* that is homologous to *GGPPS* found in fungi and its function is not clear (Athanasakoglou & Kampranis, 2019; Fabris *et al.,* 2020).

Phytoene synthase (PSY) is targeted to the chloroplast and catalyses the first reaction of the carotenoid biosynthesis pathway. It is an interesting enzyme to understand physiologically given that diatoms contain specific and high-valued types of carotenoids, namely diatoxanthin, diadinoxanthin and fucoxanthin, which play an essential role in photoprotection and response to high-light stress (Athanasakoglou & Kampranis, 2019). From a metabolic engineering point of view, phytoene synthase is the main competitor for the GGPP pool, which is the required precursor to produce high-value diterpenoids. The challenge with mutating *GGPPS* and *PSY* is that they are likely to be essential for cell survival and the knock-outs would have a lethal phenotype. To overcome this bottleneck, in this chapter my aim was to utilise the CRISPR-mediated homologous recombination approach for precise and rationally designed genome editing to try to change the regulatory sequences of *GGPPS* and *PSY,* instead of fully knocking-out the genes. To do this I tried to replace their native promoter for the cobalamin-repressible *METE* promoter so that we can have a toggle for the expression of the genes that we can control easily.

The research presented in this chapter has been of interest to many people in the research group, and I have collaborated with Dr Katrin Geisler, Dr Andrew Sayer, Catherine Sutherland and Elena Bidash to perform some of the experiments presented here – I indicate their contributions as appropriate.

The objectives of this chapter are to:

1. Compare the efficiency and practicality of electroporation, microparticle bombardment and bacterial conjugation as delivery methods for a Cas9 construct
2. Establish a standard pipeline for the generation of CRISPR-mediated homologous recombination biallelic mutants of any desired gene in *P. tricornutum*
3. Generate knock-out mutants for thiamine-related genes *THIC* and *SSSP;* cobalamin-related genes *METE and CBA1*, and determine the phenotypic effects of the knockouts

5.2. Developing CRISPR/Cas9 on *PtUMPS* selectable gene

A proof-of-concept attempt to apply CRISPR/Cas9 in *P. tricornutum* was carried out in collaboration with Elena Bidash, an undergraduate student jointly supervised by Dr Katrin Geisler and me. We chose the *PtUMPS* gene as a target gene since the knock-out is easily phenotyped, can be positively selected by 5-FOA resistance, and negatively selected by uracil auxotrophy. Since electroporation is the preferred delivery method for genetic constructs in our group, this proof-of-concept experiment intended to compare the performance of electroporation with the techniques already reported in the literature, microparticle bombardment and bacterial conjugation. Finally, we were inspired by Prof. Thomas Mock research group (Hopes *et al.,* 2016; Belshaw *et al.,* 2017 – Preprint) to adopt in our methodology the use of sgRNA pairs and homologous recombination repair templates to have the ability to rationally design the mutations introduced in the genome and facilitate the screening of the transformants by simple PCR.

Our CRISPR methodology was based on the co-transformation of a first plasmid carrying a Cas9-YFP, a sgRNA pair and a zeocin resistance cassette, and a second plasmid carrying an homologous recombination template consisting of a nourseothricin cassette flanked by homology arms (Figure 5.2). The sgRNA target sites were designed to be a few hundred base pairs away from each other and have a high on-target and a low off-target score as predicted by the Broad Institute sgRNA designer and the RGEN Cas-Designer respectively (Bae *et al.,* 2014; Doench *et al.,* 2016; Hopes *et al.,* 2017). Two different sgRNA pairs, one targeting the first exon and a second targeting the second exon, were designed to maximise the chances of having an efficient sgRNA pair (*pMLP2071* and *pMLP2072* plasmids, respectively). In addition, a version of the constructs encoding Cas9-YFP and the sgRNA pairs was cloned with the *OriT* and *CEN-ARS-HIS* sequences to allow delivery by bacterial conjugation (*pMLP2073* and *pMLP2074*). The homologous recombination template included approximately 900 bp long homology regions upstream and downstream of the *PtUMPS* CDS flanking a nourseothricin resistance cassette (*NAT*) (*pMLP2075*).

Elena Bidash used the constructs encoding Cas9-YFP and a sgRNA pair (*pMLP2071* or *pMLP2072*) to transform $1.25 \cdot 10^7$ cells either individually or in co-transformation (2.5 µg for each plasmid) with the homologous recombination template (*pMLP2075*) by electroporation or microparticle bombardment. The constructs with conjugation parts (*pMLP2073* and *pMLP2074*) were transformed by bacterial conjugation to compare the efficiency of the three delivery methods. All single plasmid transformations were split in equal volumes after recovery and spread on two 75 $mg \cdot l^{-1}$ zeocin and 100 $mg \cdot l^{-1}$ uracil selection plates. All co-transformations were split into two equal volumes one spread on a selection plate containing 75 $mg \cdot l^{-1}$ zeocin and 100 $mg \cdot l^{-1}$ uracil and the other on a plate containing 300 $mg \cdot l^{-1}$ nourseothricin and 100 $mg \cdot l^{-1}$ uracil.

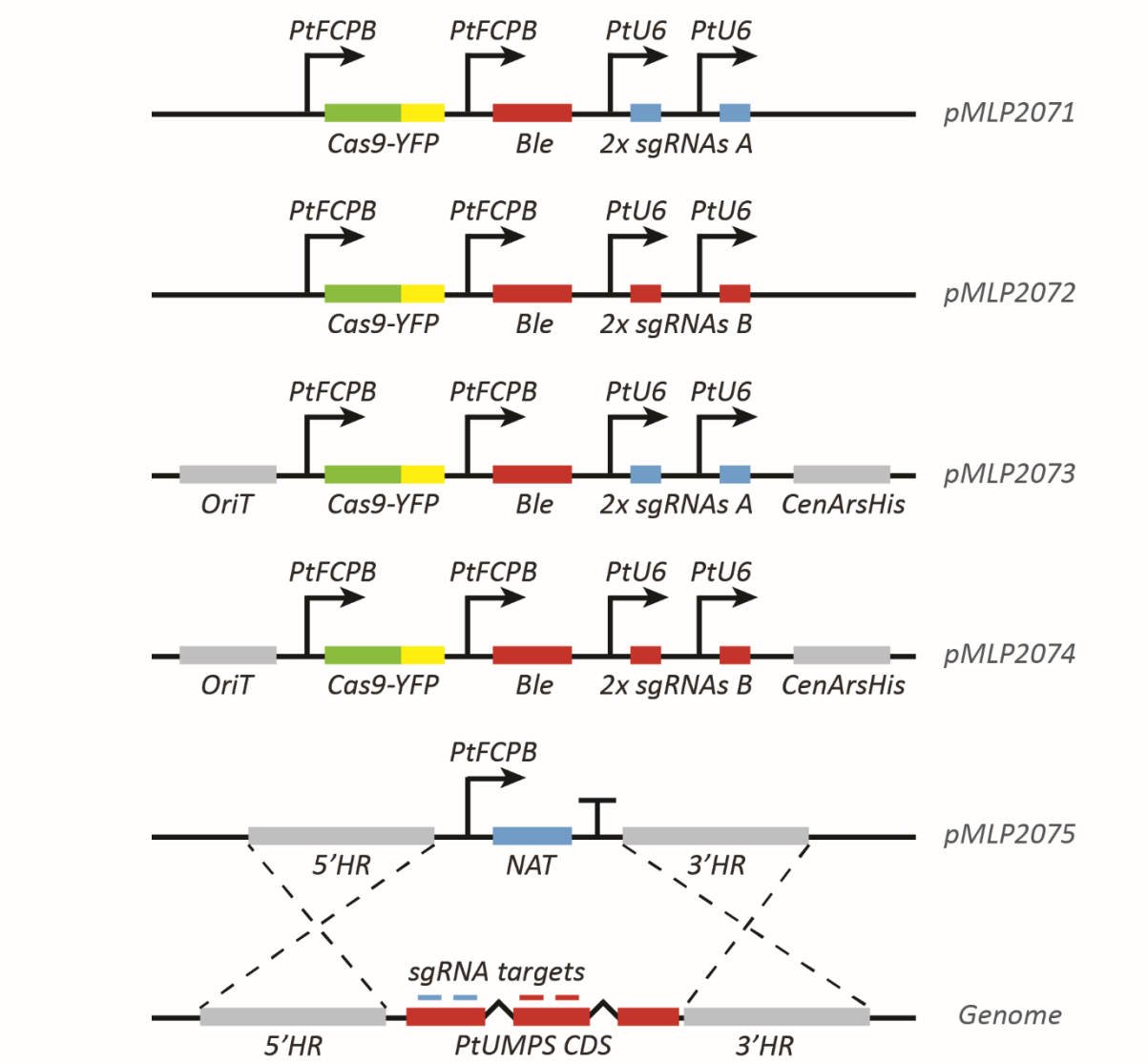


***Figure 5.2. Construct design to test the CRISPR/Cas9 homologous recombination approach on* PtUMPS.** *pMLP2071 and pMLP2072 constructs include* Cas9-YFP *and* Ble *(zeocin resistance gene) controlled by the* PtFCPB *promoter, in addition to a sgRNA pair controlled by the* PtU6 *promoter. pMLP2073 and pMLP2074 are equivalent to pMLP2071 and pMLP2072 but in addition contain the* OriT *and* CenArsHis *sequences allowing for transformation via bacterial conjugation. sgRNA pairs for pMLP02071 and pMLP2073 target the first exon of* PtUMPS*, and the sgRNA pair in pMLP2072 and pMLP2074 targets the second exon of* PtUMPS*. pMLP2075 is the homologous recombination template which includes roughly the 900 bp regions upstream and downstream of the* PtUMPS *CDS flanking a nourseothricin resistance cassette for selection and phenotyping.*

After electroporation, colonies appeared on zeocin selection plates after 12 days; in contrast it took 3 to 4 weeks for colonies to appear after microparticle bombardment. Slightly more than 40 colonies appeared on zeocin selection plates for electroporation with *pMLP2071* or co-transformation with the homologous recombination template (*pMLP2071* and *pMLP2075*) while 9-20 colonies per plate appeared for electroporation with *pMLP2072* or co-transformation with the homologous recombination template (*pMLP2072* + *pMLP2075*). Microparticle bombardment transformation gave fewer than 10 colonies per zeocin selection plate for transformation with *pMLP2071* or co-transformation with *pMLP2071* and *pMLP2075*, and only one colony per plate for transformation with *pMLP2072* and no colonies for co-transformation with *pMLP2072* and *pMLP2075.* Conjugation gave less than 15 colonies for *pMLP2074* and no colonies for *pMLP2073*. These initial results indicated that electroporation was the fastest and most efficient method to obtain transformants with our CRISPR/Cas9 construct design, although a higher number of colonies and replicates would be required to confirm this conclusion as statistically significant.

Next, we wanted to test whether the colonies selected on zeocin had been edited at the *UMPS* loci by testing whether they were resistant to 5-FOA. Colonies selected on zeocin were picked and subcultured in 96-well plates with 200 µl f/2 supplemented with 100 $mg \cdot l^{-1}$ uracil and 100 $mg \cdot l^{-1}$ 5-FOA. None of the more than 50 negative control colonies transformed with the homologous recombination template alone (*pMLP2075*) and initially selected on nourseothricin survived this treatment, while more than 60 % of colonies carrying the *pMLP2072* plasmid survived, and around 20 % of colonies transformed with plasmid *pMLP2071* survived both for electroporation and microparticle bombardment. Close to 40 % of the colonies conjugated with *pMLP2074* were resistant to 5-FOA. These results indicate that homologous recombination does not appear to happen spontaneously given that the homologous recombination template alone did not give any colonies resistant to 5-FOA. Furthermore, these results confirm that we were able to generate knock-outs with an expected phenotype in *P. tricornutum* with three different delivery methods. Notably, the percentage of 5-FOA resistant transformants differed significantly between sgRNA pairs, possibly indicating variability and unpredictability in effectiveness between sgRNAs. Given that there were no clear differences in percentage of 5-FOA resistant colonies between electroporation, microparticle bombardment and conjugation, and taking into account that electroporation gave more colonies per transformation in half the time, electroporation was chosen as the preferred delivery method to be used in all further experiments.

In parallel with phenotyping, we also wanted to genotype some transformants to confirm our CRISPR strategy had produced mutants by homologous recombination. Cells co-transformed by electroporation with the plasmid carrying the sgRNA pair targeting the first exon (*pMLP2071*) and the

homologous recombination template (*pMLP2075*) and initially selected on 300 mg·l^{-1} nourseothricin plates were picked and subcultured in the presence of 100 mg·l^{-1} uracil and 100 mg·l^{-1} 5-FOA. Fifteen of the colonies that were resistant to both nourseothricin and 5-FOA were taken forward to be genotyped as candidates to be homologous recombination mutants. A forward primer annealing at the genome upstream of the 5' homologous region and a reverse primer annealing at the nourseothricin resistance cassette were used to screen for homologous recombination events (Figure 5.3a). Eight out of the 15 colonies screened showed a positive band at the expected size (Figure 5.3b). The PCR confirmed that homologous recombination had been achieved with our CRISPR strategy, but the PCR setup did not allow us to discern whether the homologous recombination events were homozygous or heterozygous, nor whether the genotype was identical throughout the colony population or the colony presented mosaicism.

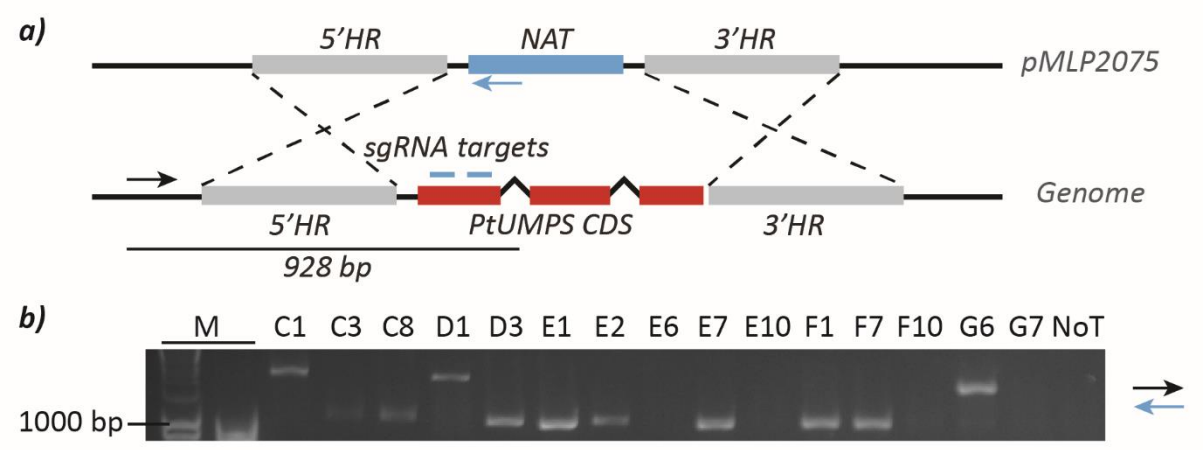


Figure 5.3. Preliminary genotyping PCR to screen for homologous recombination in primary colonies co-transformed with Cas9 and sgRNA primer pair (pMLP2071) and a homologous recombination template (pMLP2075). *Individual colonies were picked from nourseothricin selection plates 3 weeks after electroporation and subcultured in 100 mg·l^{-1} uracil and 100 mg·l^{-1} 5-FOA.* ***a)*** *Primers amplifying between the genomic region upstream the homologous recombination loci and the nourseothricin cassette were used to test whether individual colonies had incorporated the nourseothricin cassette by homologous recombination.* ***b)*** *Eight out of 15 colonies tested showed a band consistent with an homologous recombination mutation (C3, C8, D3, E1, E2, E7, F1 and F7).*

We therefore designed allele-specific primers to discern between the two alleles and determine whether the homologous recombination event was homozygous or heterozygous. The forward primer was designed to anneal at the genomic region upstream of the 5' homologous recombination region exactly matching an existing single nucleotide polymorphism (G/C) at its 3' end. The reverse primer was designed to anneal at the genome downstream of the 3' homologous recombination region so that the PCR would produce amplification bands of different lengths for WT or homologous-

recombined alleles (Figure 5.4a). Three of the selected colonies resistant to both 5-FOA and nourseothricin (D1, E1 and E7) were screened with this primer setup. The PCR results showed more than one band for each colony, colonies E1 and E7 showed a band at the expected homologous recombination size (2315 bp), and colonies D1 and E7 showed a band for the expected WT length (3267 bp). Colonies D1 and E1 showed bands of unexpected size that could be attributed to indel mutations. The amplification bands did not differ significantly between the two different allele-specific primers (Figure 5.4b,c). These PCR results were not sufficiently discriminating to differentiate between alleles, and it would be necessary to use a polymerase specifically commercialised for allele-specific PCR like the HiDi DNA polymerase (myPOLS, Germany; Serif *et al.,* 2018). Nevertheless, the multiple bands for each of the colonies indicate that the colonies were either heterozygous, mosaic, or both. Colonies D1 and E1, with more than 2 bands each, were likely to be mosaic.

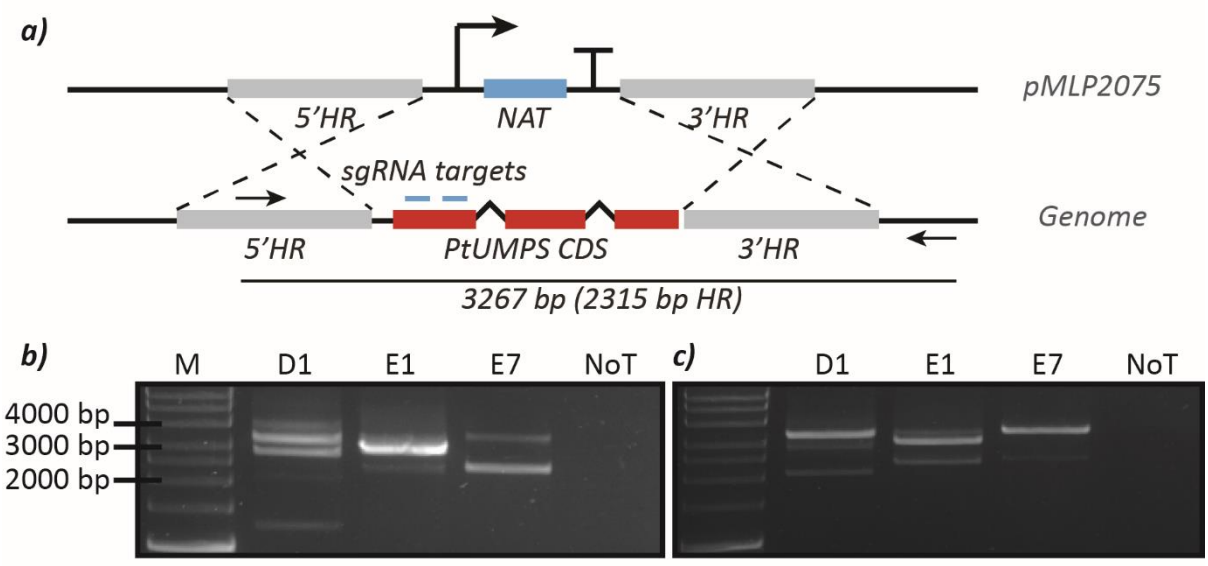


Figure 5.4. Preliminary genotyping of three primary colonies co-transformed with a Cas9 plus sgRNA pair construct (pMLP2071) and the homologous recombination template (pMLP2075). a) *The incorporation of the nourseothricin resistance cassette into the* PtUMPS *loci by homologous recombination is expected to generate a deletion of almost 1 kbp detectable with a PCR with primers flanking the recombination loci.* ***b, c)*** *PCR products for three independent transformants amplified with primers designed to bind in their 3' end to allele-specific polymorphisms for each of the two alleles in the* P. tricornutum CCAP 1055/1 *WT strain routinely used in our lab. Transformant E7 shows two clearly defined bands matching the expected sizes for the WT and the homologous recombination bands. Transformants D1 and E1 have a combination of several bands at the size expected for WT and other unexpected sizes. The variety of bands for each PCR reaction could reflect both an heterozygous genotype and a mosaic of different indel genotypes in the colonies picked after selection.*

With the objective to separate the mosaic colonies into monoclonal strains, 10 µl for three selected colonies resistant to both 5-FOA and nourseothricin (E1, E7 and F1) were streaked on zeocin selection plates. Twenty-four secondary colonies for each of the primary colonies were picked into 96-well plates with 200 µl f/2 containing 75 $mg \cdot l^{-1}$ zeocin and 100 $mg \cdot l^{-1}$ uracil and subcultured for 2 weeks. Four secondary colonies for each of the primary colonies were genotyped by PCR with four different primer pairs to test whether they had been successfully isolated as monoclonal colonies, and whether they had mono- or bi-allelic homologous recombination (Figure 5.5a). The first primer pair annealed immediately upstream and downstream of the homologous recombination loci and was expected to give a visibly lower band size for the homologous recombined genotype compared to the WT genotype. The second and third primer pairs were designed to amplify between the region upstream of the 5' homologous recombination arm and the nourseothricin cassette, and between the nourseothricin cassette and the region downstream of the 3' homologous recombination arm respectively, amplification was expected only for the homologous recombined genotype. The fourth primer pair amplified across the sgRNA pair target sites and should give a band of a smaller size if there was a deletion between the sgRNA pair target sites, no band for a biallelic homologous recombined genotype, or a band of expected size for WT. The results showed that the secondary colonies descended from the E1 and F1 primary colonies were heterozygous, with one allele homologously recombined, and the second containing a ~300 bp deletion. The secondary colonies descended from the E7 primary colony showed a homozygous genotype with both alleles mutated by homologous recombination (Figure 5.5b). Secondary colony B4, descended from E1, did not show a band for the primer pair amplifying across the sgRNA target sites, which could have been a technical error, whilst secondary colony B6 showed a WT band for the primer pair amplifying across the homologous recombination loci. This could possibly be explained by contamination in that particular reaction given that all other PCRs with this clone show a genotype consistent with a homologous recombined homozygous genotype. These results conclusively show that we had generated biallelic knock-out mutants by homologous recombination in *P. tricornutum* after a necessary step of restreaking a primary colony.

Next, we wanted to investigate whether the deletion identified in secondary colonies descended from E1 and E7 primary colonies was induced by Cas9 restriction or appeared spontaneously and was selected by 5-FOA resistance. To address this question a PCR band from one of the secondary colonies descended from E1 was excised from the gel, purified and sent for Sanger sequencing. The resulting chromatogram was aligned with the reference genome and revealed a 341 bp deletion, which led to a frame-shift mutation, spanning across the first intron and ending 9 bp downstream of the expected Cas9 cut site of one of the sgRNA target sites in the second exon (Figure 5.6). These results explain the

5-FOA resistance phenotype but were surprising given that the E1 colony was transformed with the sgRNA pair targeting the first exon (*pMLP2071*) and the deletion was associated with the sgRNA pair targeting the second exon (*pMLP2072*). The deletion could likely be explained as a mislabelling of the sample, or more unlikely by random mutation positively selected by 5-FOA resistance.

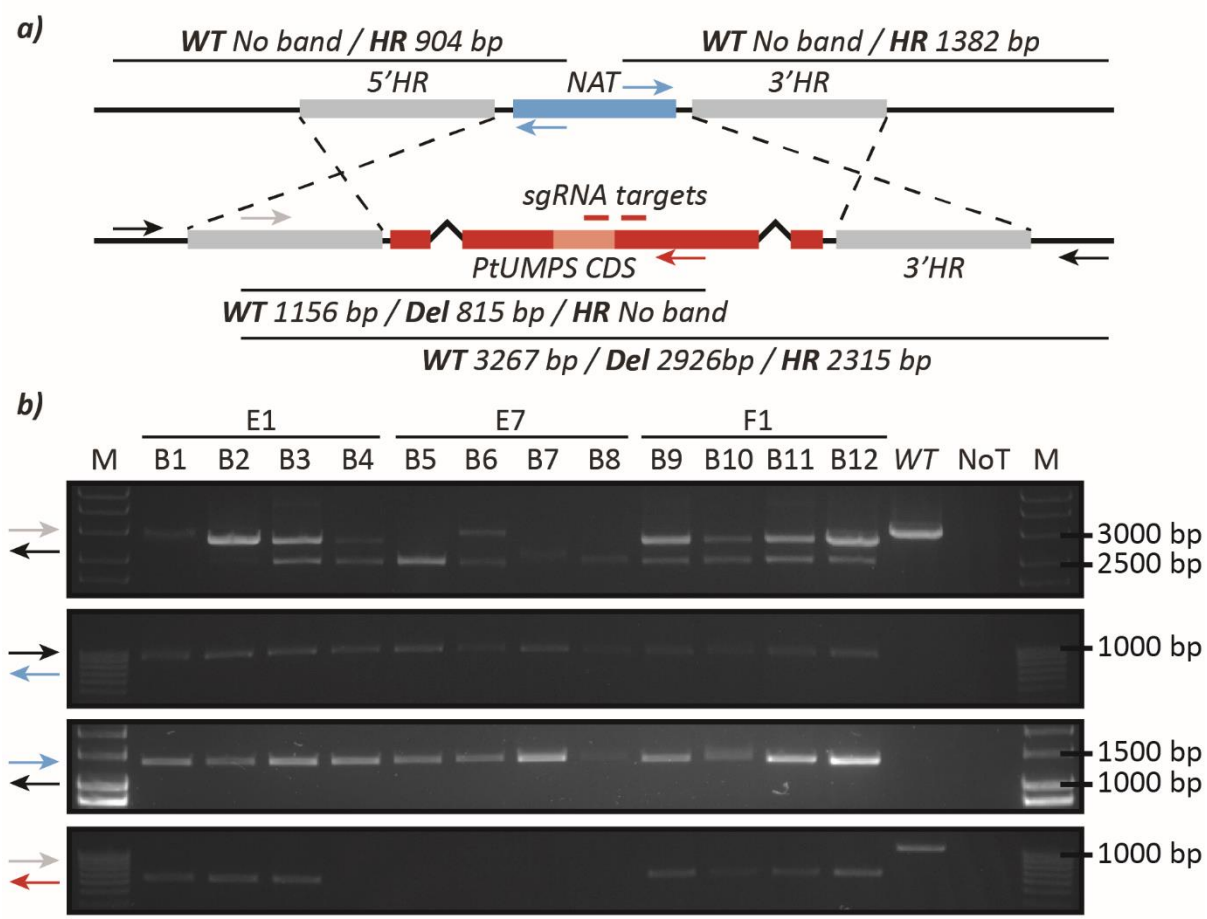


Figure 5.5. Genotyping PCRs on secondary colonies co-transformed with Cas9 plus sgRNA pair construct (pMLP2071) and the homologous recombination template (pMLP2075). a) *A combination of 4 primer pairs (colour coded arrows) was used to amplify distinctive DNA fragments (thin lines) with different expected sizes (numbers associated with lines) for wild type "WT", deletion "Del", and homologous recombination "HR" genotypes.* ***b)*** *Four secondary colonies (B1 to B12) for each of three primary colonies resistant to 5-FOA and nourseothricin (E1, E7 and F1) were genotyped with each of the four primer pairs (colour-coded arrows left of each agarose gel panel). Secondary colonies descendant from E1 and F1 generally show bands that match both an homologous recombination and a deletion genotype indicating they are heterozygous knock-out mutants. Secondary colonies descendant from E7 generally show bands matching an homologous recombination genotype with no bands consistent with WT or small deletion genotype indicating they are homozygous recombined knock-out mutants.*

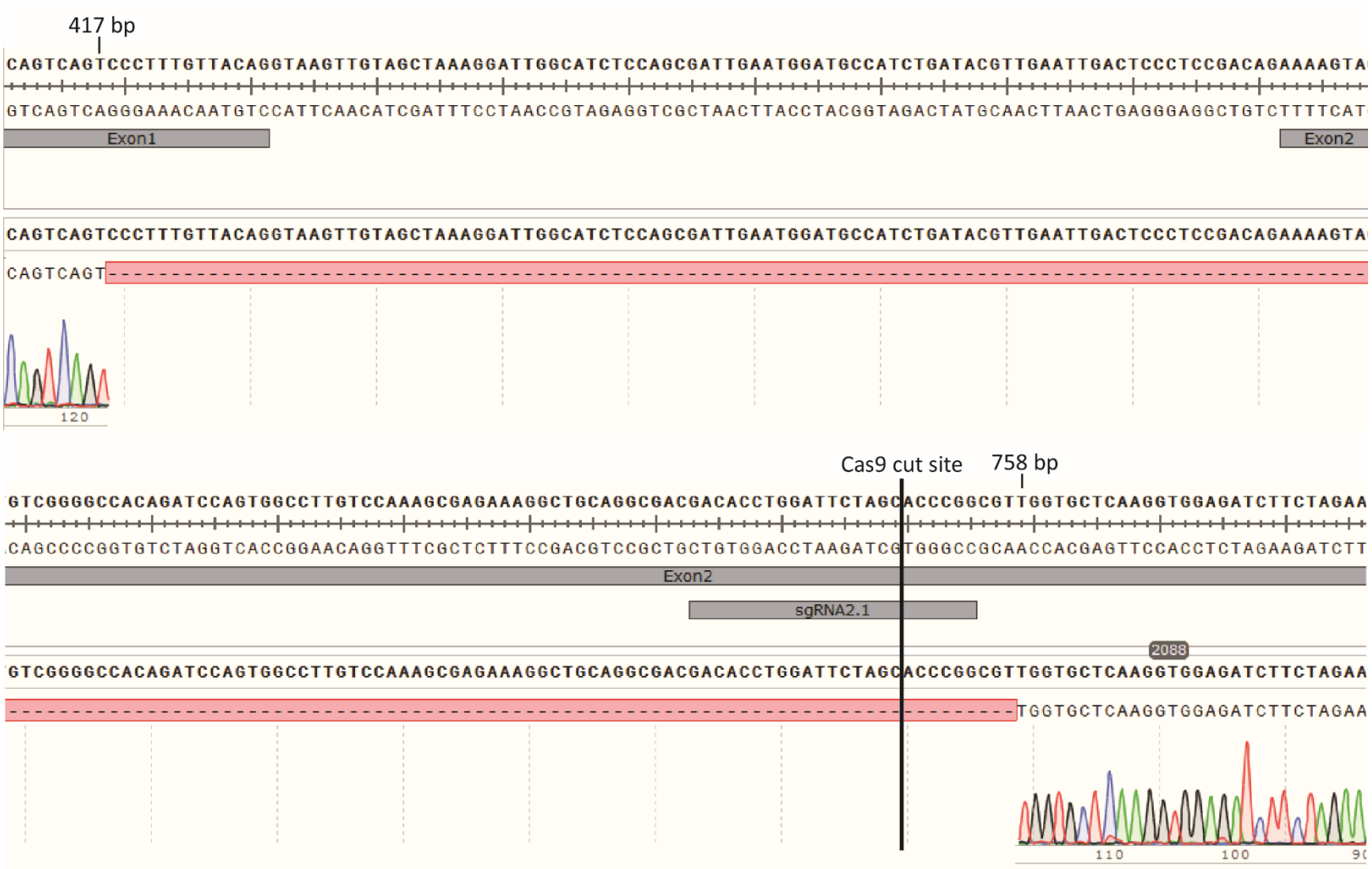


Figure 5.6. Sequence alignment between the Sanger chromatogram of* PtUMPS *knock-out mutant and the* P. tricornutum *reference genome. *The top sequence is the reference genome with annotation of* PtUMPS *exons, sgRNA target site and Cas9 cut site. The bottom sequence and chromatogram are for the* PtUMPS *knock-out mutant descendant of the E1 primary colony, which includes a 341 bp deletion between residues 417 bp and 757 bp downstream the* PtUMPS *start codon.*

5.3. Generating knock-outs for thiamine, cobalamin and terpene-related genes

With the experience gathered with the pilot test on *PtUMPS*, a transformation and screening pipeline was designed to generate *P. tricornutum* strains edited by homologous recombination at any given genomic position (Figure 5.7). We tested the pipeline to generate knock-out mutants for thiamine-related genes *THIC, NMT1* and *SSSP,* cobalamin-related genes *METE, METH* and *CBA1*, and terpene metabolism-related genes *GGPPS* and *PSY.* I was personally interested to investigate the phenotypes of the *THIC, NMT1, SSSP* knock-out mutants to further understand the physiological role of these thiamine-related genes, as it could give additional information to explain the physiological relevance of their genetic regulation. Many members of our research group are interested in questions around cobalamin metabolism and algae, and I chose to target *METE*, *METH* and *CBA1* as critical enzymes related to cobalamin in *P. tricornutum* (see section 1.1.3 for more detail on their relevance). Finally, some members of the research group are trying to engineer *P. tricornutum* to produce high value terpenoids. In our research group toolbox we already have a well-established capacity to overexpress multiple heterologous enzymes and control their expression level, but we cannot control the expression of endogenous enzymes that compete with the heterologous metabolic pathways. To better understand the role of terpene metabolism-related enzymes and to try and control the expression of these genes, I chose to also target *GGPPS* and *PSY.*

The first step in our standard methodology involved designing sgRNA pairs targeting sites a few hundred base pairs away from each other with high on-target scores and low off-target scores as predicted by the Broad Institute sgRNA designer (Doench *et al.,* 2016) and the RGEN Cas-Designer respectively (Bae *et al.,* 2014). At the same time, we designed 800 bp homology arms upstream and downstream the GOI CDS (Figure 5.7a). Inspired by the successful pilot on *PtUMPS* and with the objective of inducing an homologous recombination mutation in a targeted GOI, two plasmids were assembled for each GOI (Figure 5.7b; Figure 5.8): the first one encoded a Cas9-YFP fusion driven by the *PtFCPB* promoter, a zeocin resistance cassette and a sgRNA pair driven by the *PtU6* promoter (*pMLP2122-9* for our GOIs), and a second carried the homologous recombination template including a nourseothricin resistance cassette flanked by each GOI-specific homology arms (*pMLP2112-7* for each of the thiamine and cobalamin-related GOIs). Given that both *GGPPS* and *PSY* knock-out mutants could have a lethal phenotype, a different approach was tested for the terpene metabolism-related genes in which the homologous recombination template included an 800 bp homology region upstream of the GOI CDS followed by the nourseothricin resistance cassette and the *PtMETE* promoter driving the expression of the GOI CDS, which itself was designed to act as the 3' homology arm (*pMLP2118-9*). The constructs for *GGPPS* and *PSY* were intended to generate knock-in mutants in which the native promoter of the GOIs was replaced by the cobalamin-responsive *PtMETE* promoter.

In this way, it would be possible to downregulate the expression of the GOIs upon cobalamin supplementation (see section 4.8) avoiding the potential lethal phenotype of a full knock-out. An mTurquoise fluorescence reporter cassette was included downstream of the 3' homology region with the objective of being able to phenotypically discern whether the construct had been inserted randomly or by homologous recombination, in which case the fluorescent reporter cassette would not be included in the genome and no fluorescence would be detected. Dr Katrin Geisler cloned and later transformed these constructs.

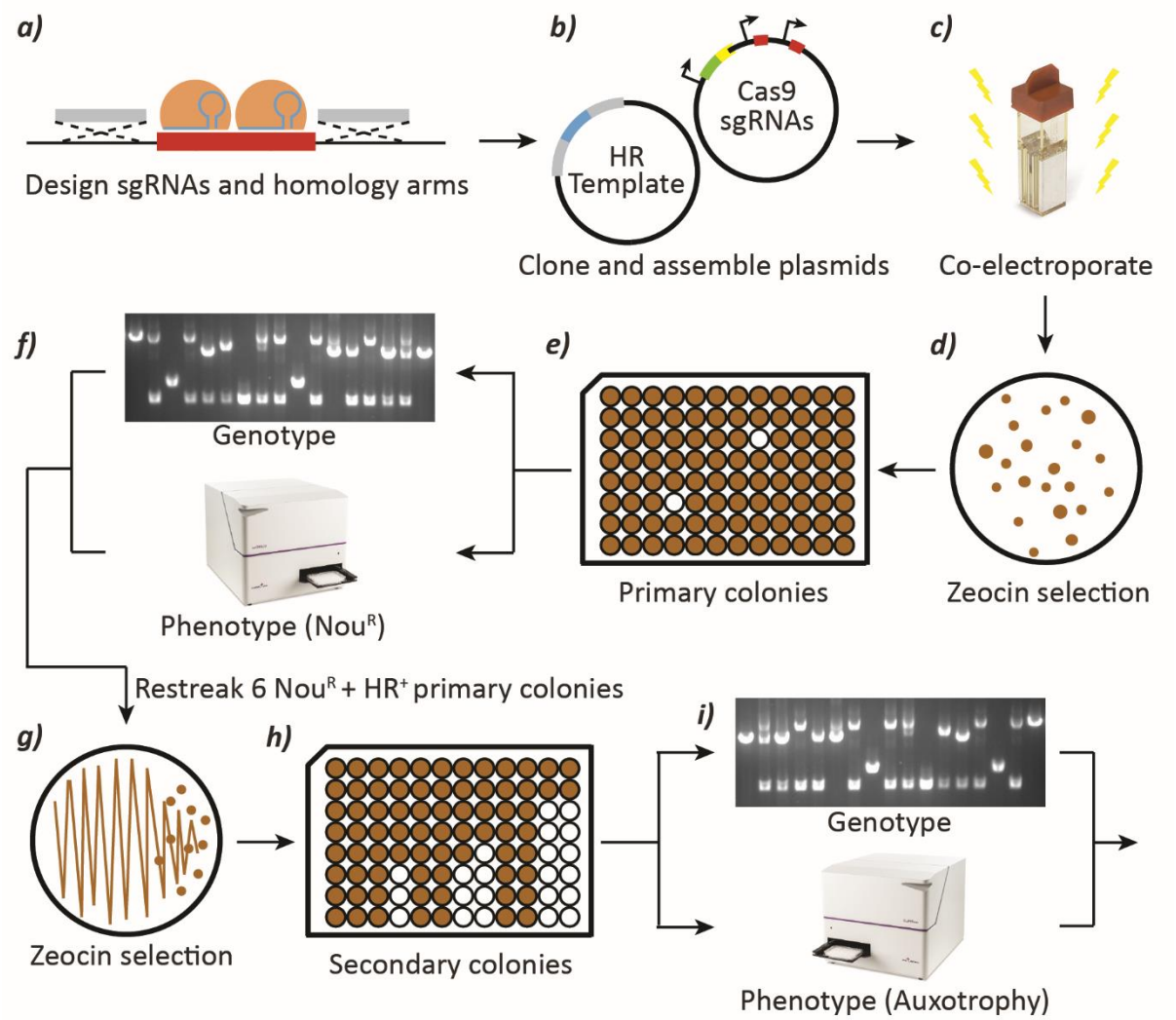
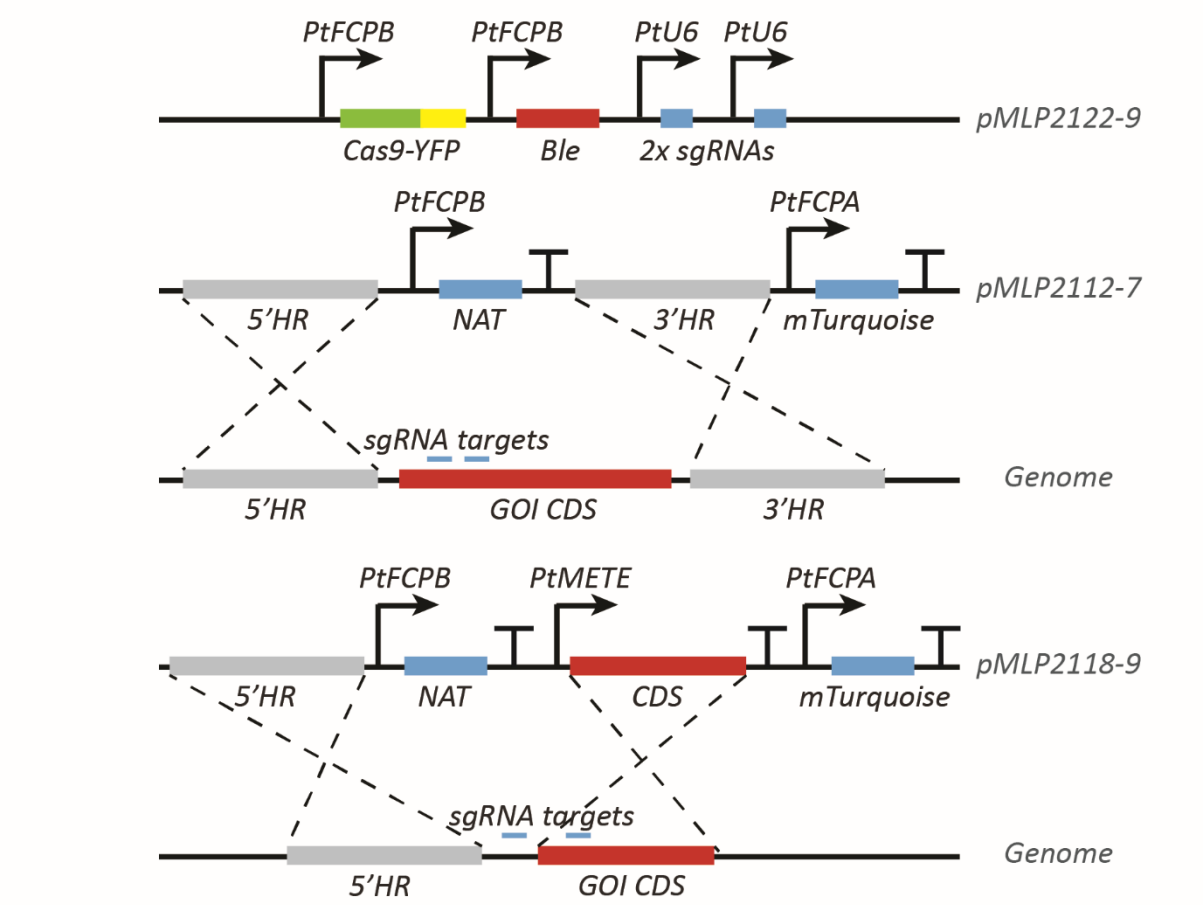


Figure 5.7. Screening pipeline to identify CRISPR/Cas9 knock-outs for thiamine, cobalamin and terpene-related genes. a) *The first step involves designing the sgRNA pair and homology arms for each target gene,* ***b)*** *followed by cloning the plasmids carrying* Cas9-YFP *and a sgRNA pair, and the homology repair template by Golden Gate.* ***c)*** P. tricornutum *is then co-electroporated with both plasmids and* ***d)*** *selected on 75 mg·l^{-1} zeocin plates for 3 weeks.* ***e)*** *Colonies are picked and subcultured in 96-well plates with 75 mg·l^{-1} zeocin for three weeks.* ***f)*** *Next, the primary colonies are genotyped by PCR and phenotyped with a growth assay on 300 mg·l^{-1} nourseothricin to identify primary colonies containing homologous recombined genotypes.* ***g)*** *Since primary colonies generally have a mosaic population including different genotypes, a selection of colonies resistant to nourseothricin and with promising genotype are re-streaked in 75 mg·l^{-1} zeocin plates to isolate monoclonal secondary colonies.* ***h)*** *After selection for three weeks, secondary colonies are picked and subcultured in 96-well plates for three more weeks.* ***i)*** *Finally, secondary colonies are genotyped and phenotyped again to select for monoclonal knock-out mutants that are taken forward for detailed phenotypic characterisation.*



GOI pMLP211X = 2: THIC ; 3: NMT1 ; 4: SSSP ; 5: METE ; 6: METH ; 7: CBA1 ; 8: GGPPS ; 9: PSY

Figure 5.8. Construct design to target 8 different* P. tricornutum *genes with CRISPR/Cas9-mediated homologous recombination. *The design follows the logic of the successful pilot to target* PtUMPS*, includes a first plasmid carrying the* Cas9-YFP *driven by the* PtFCPB *promoter and a pair of sgRNAs driven by the* PtU6 *promoter with a zeocin selection cassette (pMLP212X) and a second plasmid containing a nourseothricin resistance cassette flanked by homology arms of around 800 bp flanking the CDS of the target gene (pMLP211X). A different approach is proposed to target geranylgeranyl pyrophosphate synthase (*PtGGPPS*,* Phatr3_J47271*) and phytoene synthase (*PtPSY*,* Phatr3_EG02349*) genes, given that their knock-out would presumably lead to a lethal phenotype. For these two genes, the homologous recombination is designed not to replace the gene CDS for the nourseothricin resistance cassette, but instead swap the native promoter for the* PtMETE *promoter controlling the expression of the gene CDS (pMLP2118 and pMLP2119).* PtMETE *is downregulated by cobalamin supplementation, so the design is expected to produce conditional knock-outs only under the presence of cobalamin. The targeted genes with their respective numeric code (X) are: 2-*THIC*; 3-* NMT1*; 4-* SSSP*; 5-* METE*; 6-* METH*; 7-* CBA1*; 8-* GGPPS*; 9-* PSY*.*

Given that electroporation was the delivery method that showed the best outcomes in the pilot experiment (section 5.2), the constructs encoding Cas9 and the sgRNA pair (*pMLP212X*) and the homology repair template (*pMLP211X*) were co-transformed in the appropriate pairs (pMLP2122 with pMLP2112, etc) by electroporation into *P. tricornutum* for each individual GOI (Figure 5.7c). Transformants were then selected on 75 $mg \cdot l^{-1}$ zeocin plates with appropriate vitamin supplementation for expected auxotrophic phenotypes (10 µM thiamine for *THIC* and *NMT1* or 1 $\mu g \cdot l^{-1}$ cobalamin for *METE*) for 2-3 weeks (Figure 5.7d). Transformations for *THIC, SSSP, METE, METH* and *CBA1* gave more than 100 colonies, *NMT1* gave 65 colonies, *GGPPS* gave 27 colonies, and *PSY* did not give colonies (Table 5.1). The lack of colonies of the *PSY* mutant and the very few colonies for *GGPPS* could in part reflect a potential essential role of these genes, and the *NMT1* number of colonies could also indicate the function of the gene is physiologically important.

Colonies appearing on selection plates were picked and subcultured in 96-well plates in 200 µl f/2 containing 75 $mg \cdot l^{-1}$ zeocin and appropriate vitamins for 2 weeks (Figure 5.7e). In collaboration with Catherine Sutherland a fraction of the primary colonies obtained were genotyped to determine which colonies had incorporated the nourseothricin cassette by homologous recombination. In parallel, colonies were phenotyped by replica-plating into media containing 300 $mg \cdot l^{-1}$ nourseothricin to determine which were resistant to this antibiotic (Figure 5.7f). The genotyping was done with a 3-primer PCR in which a forward primer was designed to anneal at the genomic region upstream of the 5' homologous arm, one of the reverse primers was designed to anneal at the nourseothricin resistance cassette, and the second reverse primer was designed to anneal at the genomic region downstream of the 3' homologous arm (Figure 5.9a). With this PCR setup, a WT colony would give a single band of expected size (WT), a biallelic homologous recombined colony would give two bands one across the homologous recombination loci and shorter than the wild type (HRin) and a second of around 1000 bp across the homology recombination left (5') border (HRlb), a monoallelic homologous recombined colony or a mosaic colony would show all three bands. A sample gel image of the PCR products is shown in Figure 5.9b, and a summary of the genotyping results for all targeted genes is presented in Table 5.1.

On average, 57 % of the colonies tested showed resistance to nourseothricin with a range between 38 and 75 % for different target genes. The PCR showed that an average of 35 % of the colonies tested had PCR bands consistent with an homologous recombined genotype, showing one or both of the two expected bands for homologous recombination (HRin and/or HRlb). When also considering bands of unexpected size presumably attributed to CRISPR-induced indels (HRin and/or HRlb + other indels) the percentage increased to 50 % of the colonies tested. The other half of the colonies tested either showed a single band corresponding to the WT genotype or did not give a PCR product. There were

important differences between target genes, for example for *NMT1* only 6 % of the colonies showed bands consistent with an homologous recombination genotype, compared to 58 % of the colonies for *METE*. When comparing the genotyping information with the nourseothricin resistance results for each individual colony, we observed that on average 89 % of the colonies showing a band amplified across the 5' homology arm (HRlb) were also resistant to nourseothricin, with little differences between target genes. Conversely, 55 % of the nourseothricin resistant colonies showed a band amplified across the 5' homology arm (HRlb) and 40 % of the colonies missing the band amplified across the 5' homology arm (HRlb) were resistant to nourseothricin. These results show that nourseothricin resistance can be used to pre-select and enrich homologous recombination genotypes only losing 10 % of positive genotypes. Nonetheless, this enriched population should still be genotyped to confirm homologous recombination since a varying number of nourseothricin resistant colonies depending on the gene show no homologous recombination genotype, likely due to random integration of the nourseothricin resistance cassette. These ratios varied widely between target genes, so whilst 7 % of the *NMT1* colonies resistant to nourseothricin showed homologous recombination bands, the ratio was at least 40 % for all other target genes. These results demonstrate that the homologous recombination efficiency varies considerably between target genes, which could reflect differences in the physiological viabilities of the gene knock-outs, differences in the efficiency of the sgRNAs, or a combination of both.

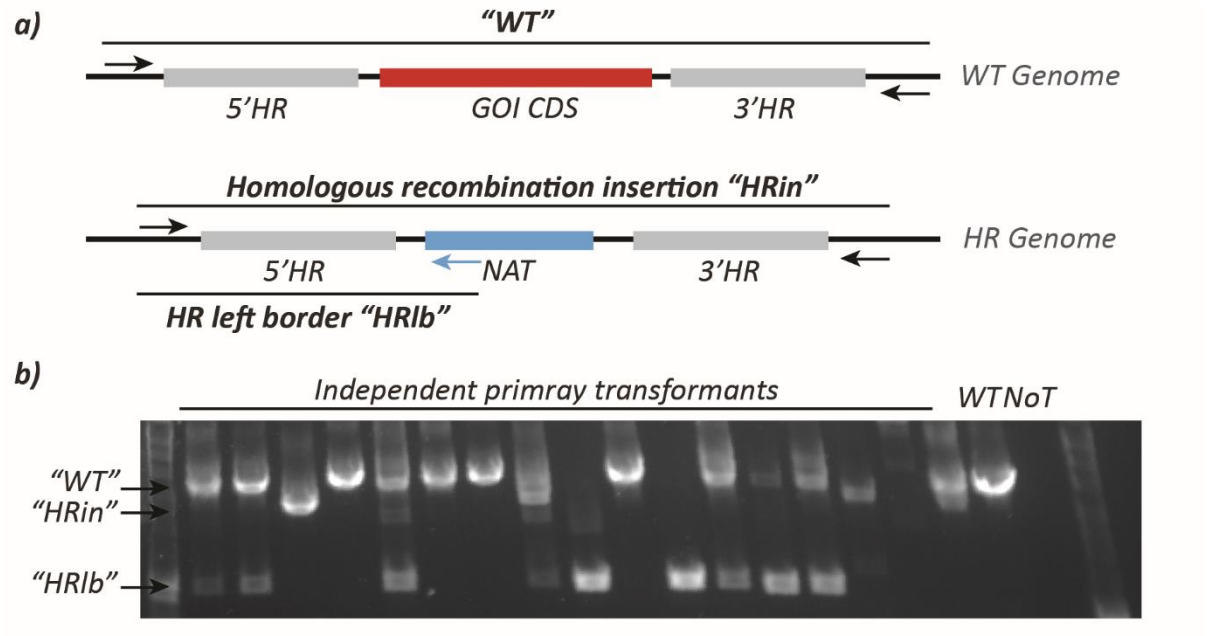


Figure 5.9 Preliminary genotyping of primary colonies by a three-primer PCR. a) *Primary colonies co-transformed with CRISPR/Cas9-mediated homologous recombination plasmids targeting* THIC, NMT1, SSSP, METE, METH, CBA1 *and* GGPPS *(pMLP211X and pMLP212X) were screened with a three-primer PCR approach that would generate a single band for colonies with wild type genotype at the loci ("WT"), would generate an shorter band of defined length for homologous recombination genotype ("HRin") as well as a specific band amplified between the region immediately upstream the recombination loci and the nourseothricin cassette ("HRlb"). A combination of the three bands would reflect an heterozygous or mosaic genotype and any other indel bands of unexpected sizes would represent uncontrolled indel mutation in the loci targeted.* ***b)*** *Example gel of a three-primer PCR showing transformants with WT, HRin and/or HRlb bands. The full summary of the results for all colonies analysed for each of the targeted genes can be found in Table 5.1.*

Table 5.1. Genotyping and phenotyping results for primary colonies of the 8 genes targeted by CRISPR/Cas9-meidated homologous recombination. *A subset of the primary colonies selected on zeocin were phenotyped for nourseothricin resistance (**Nou^R**) in growth assays in the presence of 300 mg·l^{-1} nourseothricin in 96-well plates and were genotyped for the presence of homologous recombination indel band and/or homologous recombination left border band (**HRin and/or HRlb**) with the three-primer PCR detailed in figure 5.9a. **Nou^R/HRlb** represents the percentage of colonies with a band across the homologous recombination left border which were also resistant to nourseothricin. **HRlb/Nou^R** represents the percentage of nourseothricin-resistant colonies which also had the homologous recombination-specific band. **Nou^R/No HRlb** represents the percentage of colonies without homologous recombination bands also resistant to nourseothricin.*

Gene	Colonies in zeocin selection plates	Transformants genotyped	Nou^R	HRin and/or HRlb	HRin and/or HRlb + other indels	$\frac{\text{Nou}^R}{\text{HRlb}}$	$\frac{\text{HRlb}}{\text{Nou}^R}$	$\frac{\text{Nou}^R}{\text{No HRlb}}$
THIC	286	144	61 %	26 %	42 %	95 %	41 %	49 %
NMT1	65	36	75 %	6 %	8 %	100 %	7 %	74 %
SSSP	142	55	45 %	38 %	38 %	90 %	76 %	18 %
METE	192	72	68 %	58 %	81 %	86 %	73 %	43 %
METH	180	72	57 %	46 %	72 %	88 %	71 %	31 %
CBA1	192	48	38 %	38 %	42 %	89 %	89 %	7 %
GGPPS	27	27	41 %	26 %	41 %	71 %	45 %	30 %
PSY	0	ND	ND	ND	ND	ND	ND	ND
Total	**1084**	**454**	**57 %**	**35 %**	**50 %**	**89 %**	**55 %**	**40 %**

Following the pipeline, primary colonies resistant to nourseothricin and with a homologous recombined genotype were restreaked on zeocin selection plates with appropriate vitamin supplementation to obtain monoclonal secondary colonies. For each target gene, 6 primary colonies showing a genotype compatible with homologous recombination were selected and 10 µl of culture were restreaked. After 2-3 weeks secondary colonies appeared and up to 16 independent colonies were picked into 96-well plates for each primary colony (Figure 5.7g). Secondary colonies were subcultured every week for two weeks and then genotyped using the same three-primer PCR used for primary colonies (Figure 5.9a) to determine how many colonies had a biallelic knock-out genotype (Figure 5.7i). The results are summarised in Table 5.2.

On average, 35 % of secondary colonies tested showed a WT genotype with only WT bands amplified, 33 % showed a monoallelic knock-out genotype with bands for both WT and homologous recombination, and 22 % showed a biallelic knock-out genotype with bands for homologous recombination only. 11 % of the colonies did not give a PCR product, and 3 % gave bands of unexpected sizes, possibly explained by indel mutations. However, the ratios differed highly between target genes: up to 78 % of the secondary colonies for *NMT1* showed a WT genotype, while only 11 % showed a monoallelic knock-out genotype and none appeared to have a biallelic knock-out genotype. *METH* had up to 81 % of the colonies showing a monoallelic knock-out genotype, while only 6 % had only WT bands, 13 % showed other indels, and again there were no biallelic knock-out mutants detected. *GGPPS* showed 47 % of the colonies with WT genotype and 43 % with a monoallelic knock-out genotype with no detected biallelic knock-outs either. *SSSP, CBA1* and *METE* all showed similar results with 27-42 % of colonies with a WT genotype, 28-58 % with a monoallelic knock-out genotype and 12-32 % with a biallelic knock-out genotype. *THIC* showed only 9 % of colonies with a WT genotype, 23 % with a monoallelic knock-out genotype, and up to 53 % of the colonies with a biallelic knock-out genotype. These results indicate that either differences in the sgRNA efficiency, or intrinsic differences in the genomic location or associated phenotypes of the genes could explain why 4 genes targeted gave biallelic knock-out genotypes while the other 3 did not.

For the four genes with biallelic knock-out genotypes, *SSSP*, *CBA1*, *METE* and *THIC*, at least ten individual secondary colonies with a combination of monoallelic and biallelic genotypes were taken forward for detailed genotypic characterisation. The target genes with only monoallelic knock-out genotypes were not brought forward for further investigation at this point since it was presumed that the deletion of a single copy of the gene would not lead to an observable phenotype.

Table 5.2. Summary of genotyping results for secondary colonies from each of the 7 genes targeted by CRISPR/Cas9. *The same three-primer PCR approach explained in figure 5.9a was used to screen monoclonal colonies obtained by restreaking the primary transformants. Band results consistent with a monoallelic knock-out genotype included the presence of a WT band in combination with the homologous recombination left border band (**WT & HRlb**), the homologous recombination indel band only (**WT & HRin**), or the three bands together (**WT & HRlb & HRin**). The sum of these three categories gave the total percentage of monoallelic knock-out genotypes (**Total monoallelic KO).** Band results consistent with a biallelic knock-out genotype included the presence of the homologous recombination left border band only (**HRlb only**), the homologous recombination indel band only (**HRin only**), or both together (**HRlb & HRin**). The sum of these three categories gave the total percentage of knock-out genotypes (**Total KO).***

				Monoallelic KO Genotypes				Biallelic KO Genotypes				
Gene	**Colonies screened**	**No bands**	**WT only**	**WT & HRlb**	**WT & HRin**	**WT & HRin & HRlb**	**Total Monoallelic KO**	**HRlb only**	**HRin only**	**HRlb & HRin**	**Total KO**	**Other indels**
THIC	88	11 %	9 %	11 %	2 %	10 %	23 %	35 %	6 %	13 %	**53 %**	2 %
NMT1	68	9 %	78 %	7 %	4 %	0 %	11 %	0 %	0 %	0 %	**0 %**	0 %
SSSP	52	0 %	27 %	48 %	6 %	4 %	58 %	0 %	2 %	10 %	**12 %**	4 %
METE	77	10 %	27 %	19 %	5 %	4 %	28 %	22 %	6 %	4 %	**32 %**	1 %
METH	16	0 %	6 %	75 %	0 %	6 %	81 %	0 %	0 %	0 %	**0 %**	13 %
CBA1	12	0 %	42 %	25 %	0 %	8 %	33 %	8 %	8 %	8 %	**25 %**	0 %
GGPPS	55	0 %	47 %	16 %	2 %	25 %	43 %	0 %	0 %	0 %	**0 %**	9 %
Total	**368**	**7 %**	**35 %**	**21 %**	**4 %**	**8 %**	33 %	**13 %**	**4 %**	**5 %**	**22 %**	**3 %**

5.4. *PtTHIC* knock-out does not show thiamine auxotrophy

The first gene knock-out to be analysed was *THIC*. For this gene, 18 secondary colonies, which showed either heterozygous or homozygous genotypes in the three-primer PCR screening, were genotyped with four different primer pairs to specifically determine their genotype. The primer design was similar to the one used for the selected secondary colonies in the pilot experiment with *UMPS* (Figure 5.5a). The first primer pair amplifies across the homologous recombination loci and should give bands of different sizes for WT and homologous recombination genotypes. The second primer pair spans the 5' homology arm and should give a band only in colonies with an homologous recombination genotype. Similarly, the third primer pair amplifies across the 3' homology arm and again should only give a band in colonies with homologous recombination genotype. The last primer pair amplifies across the sgRNA pair target sites and should give a band of expected size for a WT genotype, a shorter band for a deletion genotype, and no band for an homologous recombined genotype (Figure 5.10a).

The PCR results showed multiple different genotypes across the 18 screened colonies. Secondary colonies A3 and E4 clearly showed a biallelic homologous recombination genotype, with a shorter band for the PCR across the homologous recombination loci (1st primer pair) and no amplification with the primer pair amplifying across the sgRNA target sites (4th primer pair). Secondary colonies B3 and C3 showed a monoallelic homologous recombination genotype, with two bands corresponding respectively to the WT and the homologous recombined alleles for the PCR across the homologous recombination loci (1st primer pair). Secondary colonies, A5 and C5 showed a genotype which appears like the WT control, although there might be a short indel around the sgRNA target sites observed as a size increase in the PCR across the sgRNA target sites (Figure 5.10b and c).

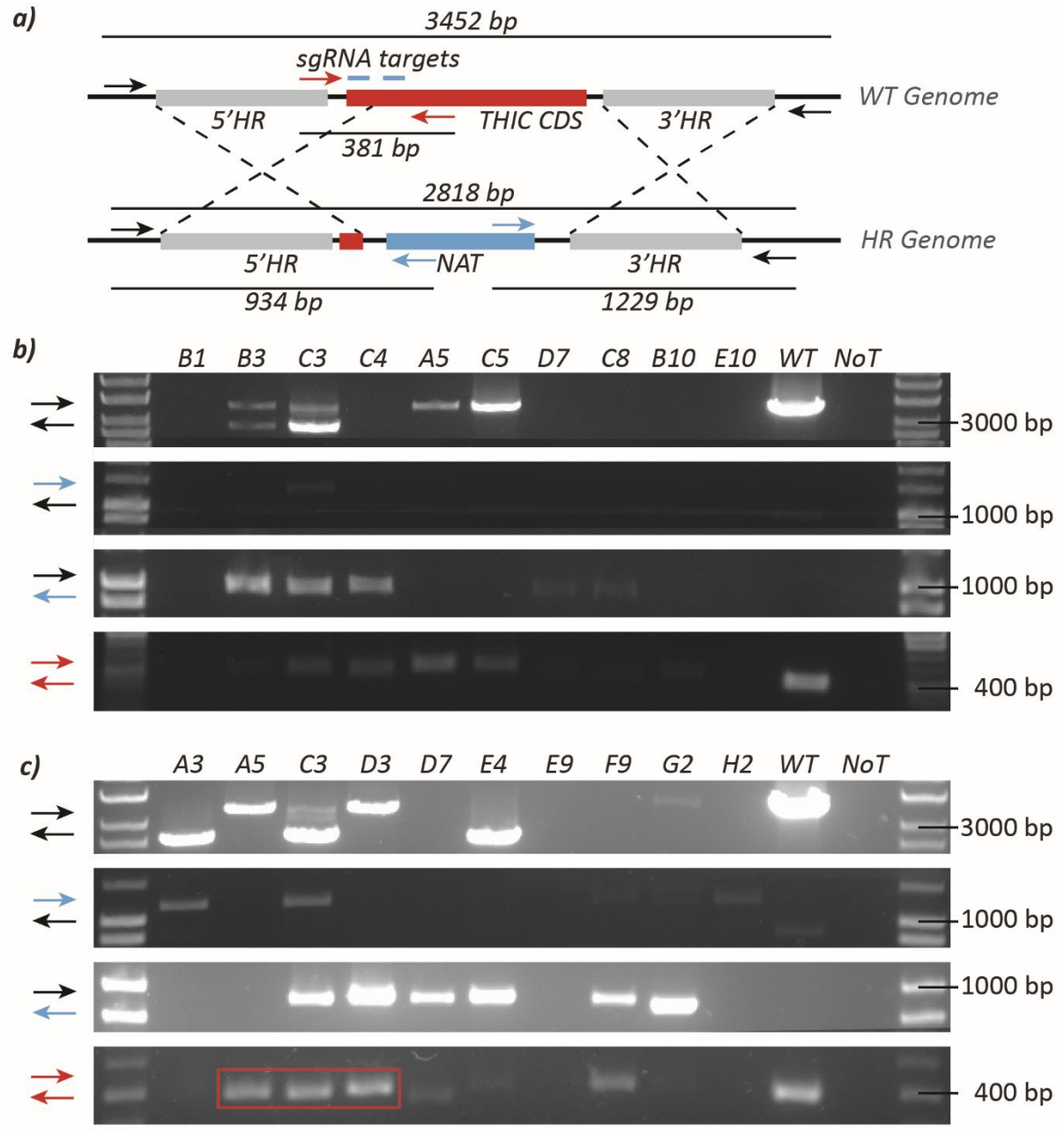


Figure 5.10. Genotyping of 18 selected secondary colonies for CRISPR/Cas9 targeting* THIC.** ***a) *Genotyping strategy for the* PtTHIC *loci with 4 primer pairs (colour coded) with the expected band sizes associated with corresponding bands (thin lines).* ***b)*** *and* ***c)*** *Two independent rounds of genotyping on 10 secondary colonies. The 4 panels correspond to the 4 primer pairs colour coded on the left and the top labels correspond to individual secondary colonies, wild type (WT) and no template control (NoT). A3 and E4 colonies show bands consistent with a biallelic homologous recombination genotype. B3 and C3 show heterozygous genotypes combining WT and homologous recombined bands. The* THIC *CDS bands for A5, C3 and D3 (red square) were cut, purified and sent for sequencing to screen for small indels undetectable by PCR. The sequencing results are shown in figure 5.11.*

To investigate whether there could be an indel mutation at the sgRNA target sites not obvious from the PCR results, PCR products for three independent colonies (A5, C3 and D3) were cut, purified and sent for sequencing (Figure 5.10c). The three PCR products were pooled together before sending to the sequencing service because the DNA yield for each individual sample was not sufficient to meet the sequencing requirements. The results showed that before reaching the sgRNA target site the Sanger chromatogram perfectly matched the reference genome, but after the Cas9 predicted cut site for the first sgRNA the chromatogram quality decreased sharply and showed overlapping peaks (Figure 5.11a). A chromatogram with overlapping peaks can be interpreted as the combination of different sequences in the same sample, which can be generated by small indel mutations. The TIDE algorithm has been specifically designed to deconvolute chromatograms from CRISPR-edited strains into a relative abundance of different indel length given a WT reference chromatogram (Brinkman *et al.,* 2014). The TIDE algorithm predicted that 15.3 % of the sequences present in our sample corresponded to a -1 deletion, 5.8 % to a -2 deletion, and 5.7 % were not mutated (Figure 5.11b). The results were not sufficient to discern which of the three secondary colonies (A5, C3 or D3) were mutated, and whether any of them had lost all functional copies of the gene. Nevertheless, these results suggest that small indels at the sgRNA target sites had been generated by our CRISPR approach in addition to inducing homologous recombination. Therefore, in addition to the known knock-outs by homologous recombination there could be other secondary colonies completely missing a functional copy of *THIC,* or indeed the other genes of interest.

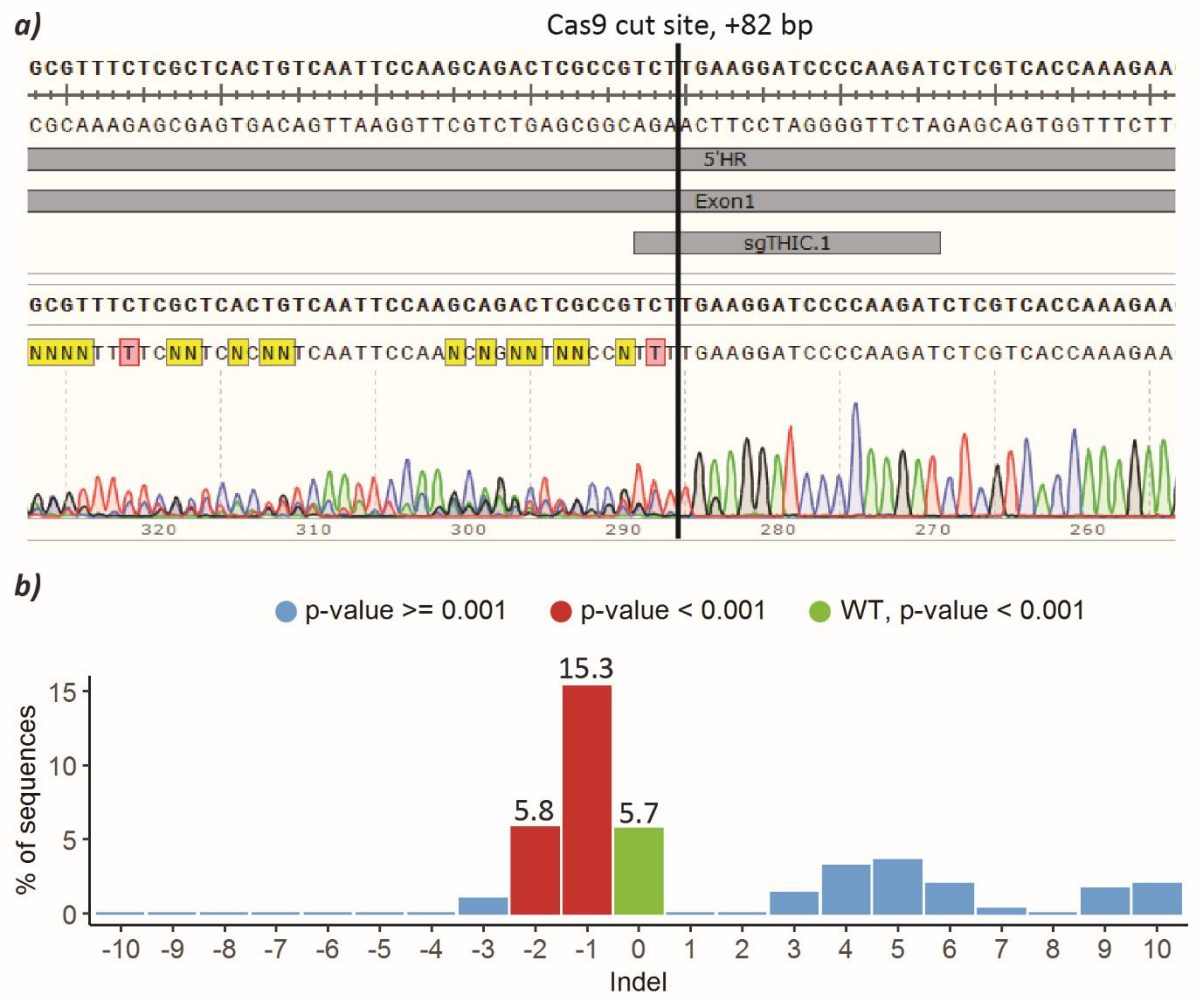


Figure 5.11. Sequencing results for the sgRNA target loci of secondary colonies A5, C3 and D3* THIC *CRISPR/Cas9 mutants. a) *The bands for the three secondary colonies were cut as shown in figure 5.10c, purified, pooled and sent for Sanger sequencing. The result shows that exactly at the expected Cas9 cut site, 82 bp downstream the start codon, the chromatogram loses accuracy and is formed by double-peaks.* ***b)*** *The TIDE algorithm to deconvolute Sanger sequencing double peak chromatogram in indel variants generated by CRISPR/Cas9 showed that the PCR product population included 15.3 % of reads with a 1 bp deletion followed by 5.8 % with a 2 bp deletion and 5.7 % with WT genotype. Since the DNA samples from the three colonies were pooled in a single sample, it is not possible to assign a genotype to the different strains, but the results confirm that short indel alleles not detected by PCR are present in one or all of the A5, C3 and D3 secondary clones.*

In parallel with the genotyping characterisation, the secondary colonies for *THIC* were tested for thiamine auxotrophy. Each of the secondary colonies was cultured in a 96-well plate in technical triplicate in f/2 medium in the presence or absence of 10 µM thiamine for 20 days, recording OD_{730nm} every 24 hours as a proxy for growth. In general, none of the secondary colonies tested showed any observable difference between treatments, including the secondary colony A3 which showed a clear biallelic homologous recombination genotype missing both copies of *THIC*. It would be interesting to phenotypically characterise other secondary colonies with a knock-out genotype, for instance E4, to confirm that the absence of *THIC* does not lead to thiamine auxotrophy, unfortunately the coronavirus pandemic lockdown prevented me from performing these experiments. If confirmed, these results would indicate that the absence of *THIC* does not prevent *P. tricornutum* from producing thiamine and might be explained by an alternative metabolic step that bypasses *THIC*.

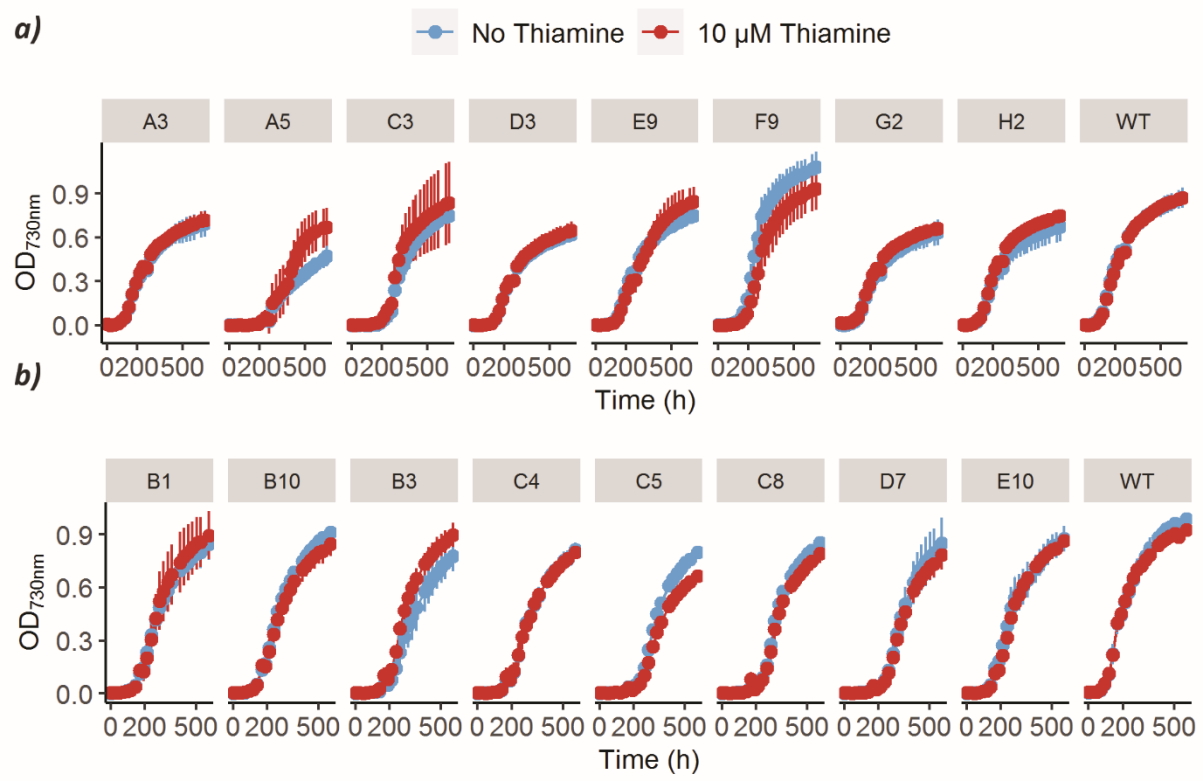


Figure 5.12. Growth characterization of* THIC *CRISPR/Cas9 mutants in the presence (10 µM) and absence of thiamine supplementation. a) *and* ***b)*** *Secondary colonies were subcultured in the presence and absence of thiamine supplementation in technical triplicates in 96-well plates and their growth tracked as* OD_{730nm} *for 20 days with a plate reader. None of the clones tested showed auxotrophy for thiamine, including the A3 secondary clone, which showed a clear knock-out genotype in figure 5.10c. This result shows that* THIC *is not necessary for thiamine biosynthesis in* P. tricornutum.

5.5. *PtMETE* knock-out requires cobalamin to grow

Ten secondary colonies for *METE* were genotypically and phenotypically characterised by PCR and a cobalamin auxotrophy test analogous to the experiments performed on the *THIC* mutants to determine whether the absence of *METE* leads to cobalamin auxotrophy in *P. tricornutum*. The PCR results showed multiple unspecific bands, probably explained by the repetitive nature of the *METE* coding sequence. Between the unspecific bands, secondary colonies C9 and H9 clearly showed a biallelic homologous recombined genotype. They both were missing the WT band and instead had the expected shorter bands for the PCR across the homologous recombination loci (1st primer pair), as well as bands for the PCRs amplifying across the homology arms (2nd and 3rd primer pairs), and showed no band for the primers amplifying across the sgRNA target sites on the gene CDS (4th primer pair). The secondary colony A3 seemed to have an heterozygous genotype given it showed both WT and homologous recombination insertion bands for the PCR across the homologous recombination loci (1st primer pair) and a WT band for the PCR across the sgRNA target sites (4th primer pair) (Figure 5.13b).

To test whether the absence of *METE* led to cobalamin auxotrophy, eight of the secondary colonies genotyped were grown in triplicate in f/2 medium both in the presence and absence of 1 $\mu g \cdot l^{-1}$ cobalamin in 96-well plates and their OD_{730nm} was measured as a proxy for growth in a plate reader for 20 days. Both C9 and H9, which had clear knock-out genotypes, could not grow in the absence of cobalamin. In addition, A9 also failed to grow without cobalamin supplementation. Secondary clone A3, which seemed to have a monoallelic homologous recombination genotype did not show any growth difference between conditions. Interestingly, B3 showed faster growth in the absence of cobalamin, and A11 showed a decreased growth rate in the absence of cobalamin, but given the high standard error of the latter, it is possible that there was a specific technical problem for one of the triplicate cultures (Figure 5.13c). These results indicate that the absence of *METE* leads to cobalamin auxotrophy in *P. tricornutum*, whereas a mono-allelic knock-out does not have a growth detriment in the absence of cobalamin.

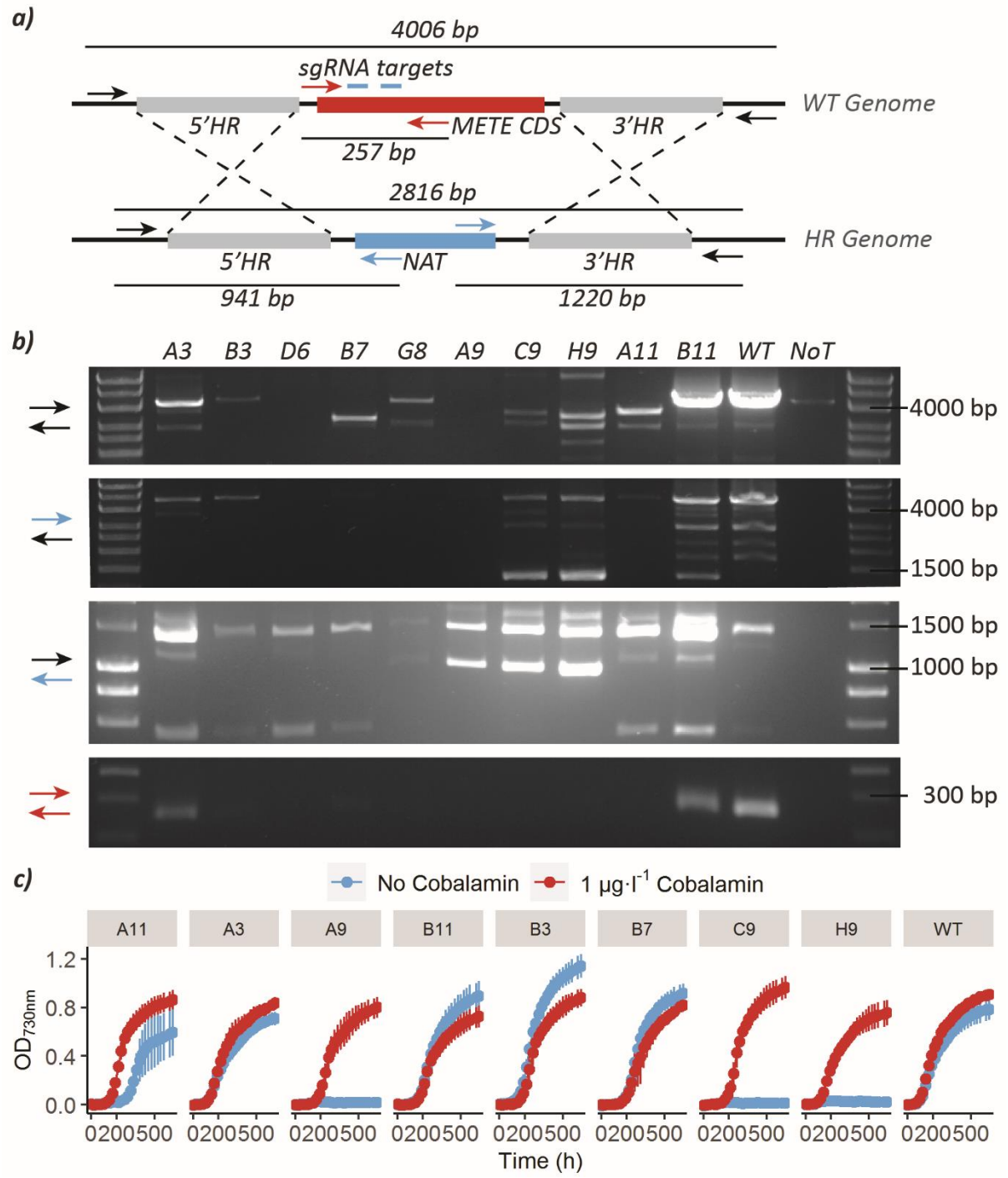


Figure 5.13. Genotyping and characterisation of CRISPR/Cas9* PtMETE *mutants. a) *Genotyping strategy with 4 primer pairs (colour coded) with expected band sizes indicated (associated with thin lines).* ***b)*** *PCR results for the four primer pairs in four different panels (colour coded on the left). The PCR for most primer pairs show many unspecific bands, probably explained by the repetitive nature of the* METE *gene. Colonies C9 and H9 show bands consistent with a biallelic homologous recombination knock-out genotype. A3 shows bands consistent with an heterozygous genotype combining WT and homologous recombination bands.* ***c)*** *Phenotypic characterization of 8 independent secondary colonies grown in triplicates with (1 µg·l^{-1}) or without cobalamin supplementation in technical triplicate in 96-well plates and growth tracked as OD_{730nm} for 20 days with a plate reader. Colonies A9, C9 and H9 show cobalamin auxotrophy consistent with their genotyping results.*

To investigate in more detail the auxotrophic phenotype, Dr Katrin Geisler grew the C9 biallelic knockout and the A3 monoallelic knock-out alongside a WT control in 20 ml triplicate cultures in an HT24 (Algenuity, Stewartby, UK) incubator with a starting cell-density of $5 \cdot 10^4$ cells·ml^{-1}. The three different samples were grown in the absence of cobalamin, in 100 ng·l^{-1} cobalamin and in 500 ng·l^{-1} cobalamin. The OD_{730nm} for each culture was recorded continuously by the bioreactor sensors as a proxy for growth. The growth of the monoallelic knock-out could not be differentiated from that of the WT control in any of the conditions. The C9 biallelic knock-out showed no growth in the absence of cobalamin but it could grow at the same rate with 100 or 500 ng·l^{-1} cobalamin, although the growth rate and carrying capacity were visibly lower compared to the WT and monoallelic knock-out (Figure 5.14). These results indicate that 100 ng·l^{-1} is sufficient to meet the cobalamin requirements of the *METE* knock-out, and it is likely that lower levels of cobalamin would still support its growth. The results also suggest that the absence of *METE* is detrimental to the growth of *P. tricornutum* even under cobalamin-replete conditions.

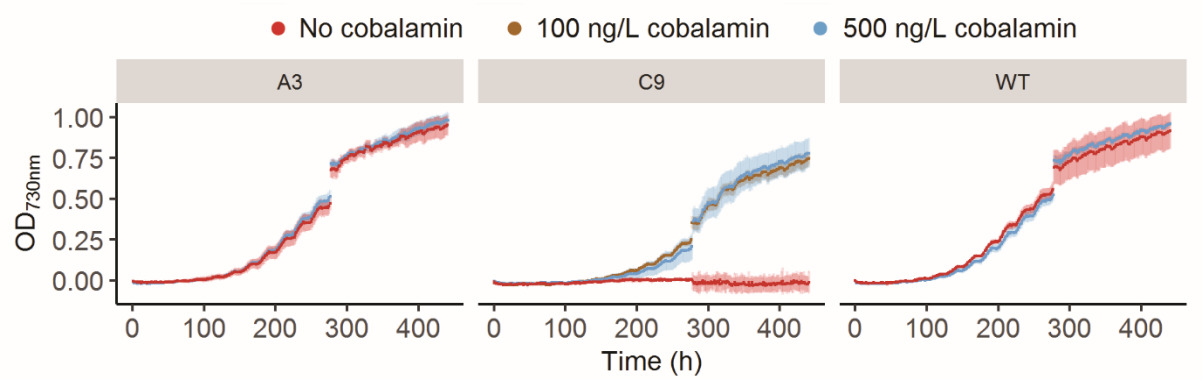


Figure 5.14 Cobalamin auxotrophy characterization of a monoallelic (A3) and a biallelic (C9)* METE *knock-out. *Mutants were grown in triplicates in 20 ml cultures in three different cobalamin conditions, deplete (blue), 100 ng·l^{-1} cobalamin (brown), 500 ng·l^{-1} cobalamin (red) with growth tracked continuously for 20 days as OD_{730nm} with an HT24 incubator. Due to technical limitations, the measurement had to be cut in two sequences. The biallelic knock-out (C9) shows cobalamin auxotrophy, consistent with its genotype and earlier characterization in 96-well plates, and 100 ng·l^{-1} cobalamin were sufficient to restore growth. The monoallelic knock-out (A3) did not show any observable differences with the WT. The biallelic knock-out (C9) grows slower than the monoallelic knock-out (A3) and WT even at 500 ng·l^{-1} cobalamin.*

5.6. Putative transporter PtSSSP is required for thiamine uptake

The secondary colonies for *SSSP* were genotyped and phenotyped to investigate whether a knock-out mutant for the gene leads to a loss of thiamine uptake phenotype. Ten independent secondary colonies were genotyped with a PCR design analogous to the *THIC* and *METE* knock-outs. The PCR results showed that the C3 secondary clone was the only clone with a biallelic knock-out genotype missing a WT band and having a band compatible with an homologous recombination insertion for the PCR across the homology recombination loci (1st primer pair) and no PCR amplification for the primer pair across the sgRNAs target sites (4th primer pair). D1 and A6 showed a WT band and a band compatible with an homologous recombination insertion for the PCR across the homologous recombination loci (1st primer pair) and a WT band for the PCR across the sgRNAs target sites (4th primer pair); D3 showed a WT band for both the PCR across the homologous recombination loci (1st primer pair) and the PCR across the sgRNAs target sites (4th primer pair) but also a band amplified across the 5' homology arm (2nd primer pair) (Figure 5.15b). These results are compatible with D1, D3 and A6 having a monoallelic homologous recombination genotype.

To test whether a *SSSP* knock-out genotype leads to a loss of thiamine uptake phenotype, the biallelic knock-out C3, the monoallelic knock-out D1 and WT were grown in the presence or absence of 10 µM thiamine in 20 ml cultures in biological duplicates. Five days post inoculation, cells were harvested, and duplicate total cell extracts from each culture were derivatised with the thiochrome method to detect intracellular thiamine content by HPLC (same method as sections 3.3.1 and 4.3). The results showed that whilst the WT has higher levels of intracellular thiamine in the replete condition compared to the deplete condition, there is no statistical difference in intracellular thiamine levels between conditions for the biallelic knock-out C3 (Tuckey HSD test, 0.95 confidence), despite the concentration of thiamine in the media being at least 4 orders of magnitude higher than in seawater (Sañudo-Wilhelmy *et al.,* 2012). The mono-allelic knock-out D1 showed higher levels under thiamine-replete conditions, but not as high as the levels reached by the WT (Figure 5.15c). These results indicate that the putative transporter SSSP is necessary for thiamine uptake in *P. tricornutum* and that a monoallelic knock-out leads to an intermediate phenotype with impaired uptake.

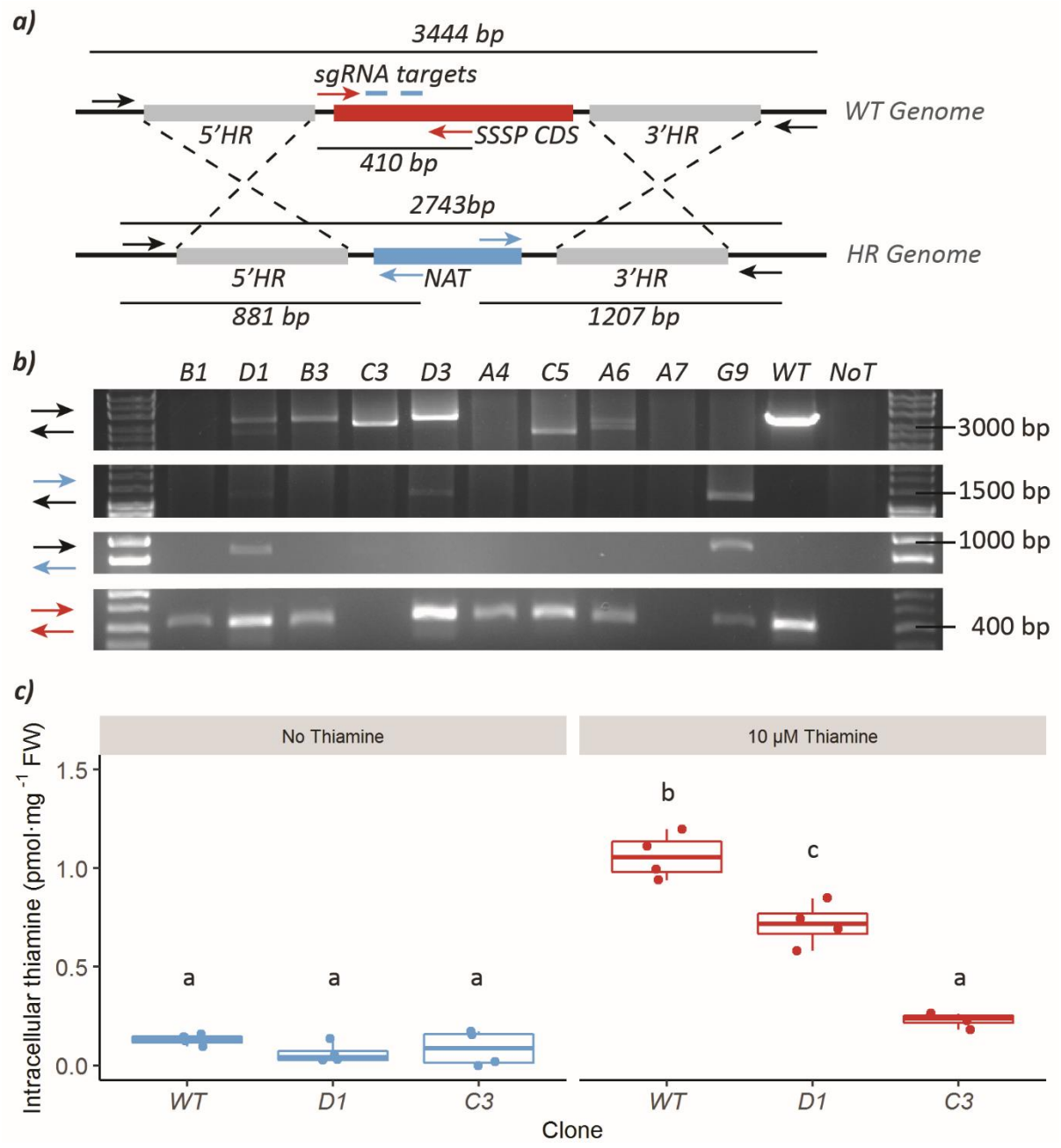


Figure 5.15. Genotyping and characterisation of CRISPR/Cas9* SSSP *mutants. a) *Genotyping strategy with 4 primer pairs (colour coded) with expected band sizes indicated (associated with thin lines).* ***b)*** *PCR results for the four primer pairs in four different panels (colour coded on the left). C3 secondary clone shows bands consistent with a biallelic homologous recombination insertion leading to loss of the* SSSP *CDS. D1, C5 and A6 show bands consistent with an heterozygous genotype combining WT and homologous recombination bands.* ***c)*** *Thiamine uptake assay by HPLC on WT, D1 and C3 secondary colonies. Two 20 ml biological replicates were grown in the presence (10 µM) or absence of thiamine supplementation, and cells were harvested at late exponential phase. A total extract from the cell pellet was derivatised with the thiochrome reaction and run in the HPLC for thiamine quantification in technical duplicate. A Tukey HSD test was performed to compare the mean thiamine content of the different clones and conditions. The results show that while WT uptakes thiamine, the monoallelic*

knock-out (D1) shows partial thiamine uptake, and the biallelic knock-out (C3) cannot take up thiamine.

5.7. Putative acquisition protein PtCBA1 is required for cobalamin uptake

To determine whether the *CBA1* knock-out leads to the loss of cobalamin uptake, ten secondary colonies for the transformants targeting *CBA1* were genotyped and phenotyped. The genotyping strategy was analogous to the one used for *THIC, METE* and *SSSP*. The PCR results showed that secondary clone B12 had a biallelic homologous recombination genotype with a band for the insertion expected size with the primers amplifying across the homologous recombination loci (1st primer pair) and no band for the PCR amplifying across the sgRNA target sites (4th primer pair). Secondary clone B11 showed a monoallelic homologous recombination genotype with bands for both WT and homologous recombination for the PCR across the homologous recombination loci (1st primer pair). Secondary clone E11 showed a biallelic knock-out mutation with a clear deletion band and no WT band for the PCR amplifying across the sgRNAs target sites (4th primer pair). Secondary clone A11 showed bands for WT and homologous recombination insertion for the PCR across the homologous recombination loci (1st primer pair) and a slightly shorter band than expected for a WT genotype in the PCR amplifying across the sgRNA target sites (4th primer pair) (Figure 5.16b).

To investigate whether a *CBA1* biallelic knock-out genotype leads to a loss of cobalamin uptake, and whether a monoallelic knock-out genotype leads to an intermediate phenotype, a cobalamin uptake assay was performed in collaboration with Dr Andrew Sayer. Four secondary clones, A11, B11, E11 and B12, were grown to late exponential phase in 20 ml cultures in biological quadruplicate. Cells were then concentrated to $5{\cdot}10^6$ cells$\cdot$ml^{-1} in 1 ml and incubated in f/2 supplemented with 600 pg cobalamin for 1h in a microcentrifuge tube. The tubes were then centrifuged to separate the cellular fraction (pellet) from the media fraction (supernatant). Each of the fractions was used to inoculate a bioassay based on the cobalamin-dependent growth of *S. typhimurium* (Raux *et al.,* 1996). The absolute level of cobalamin for each fraction was calculated based on a calibration curve for the growth of *S. typhimurium* under different known concentrations of cobalamin. The results showed that while WT contained virtually all the added cobalamin in the cellular fraction, the cobalamin levels for the biallelic knock-out (A11, E11 and B12) cellular fractions were indistinguishable from a negative control without cells (Tuckey HSD test, 0.95 confidence level). The monoallelic knock-out (B11) showed reduced cobalamin uptake (Figure 5.16c). These results indicate that putative acquisition protein CBA1 is essential for cobalamin uptake in *P. tricornutum* and that, like *SSSP*, a monoallelic genotype leads to an intermediate phenotype.

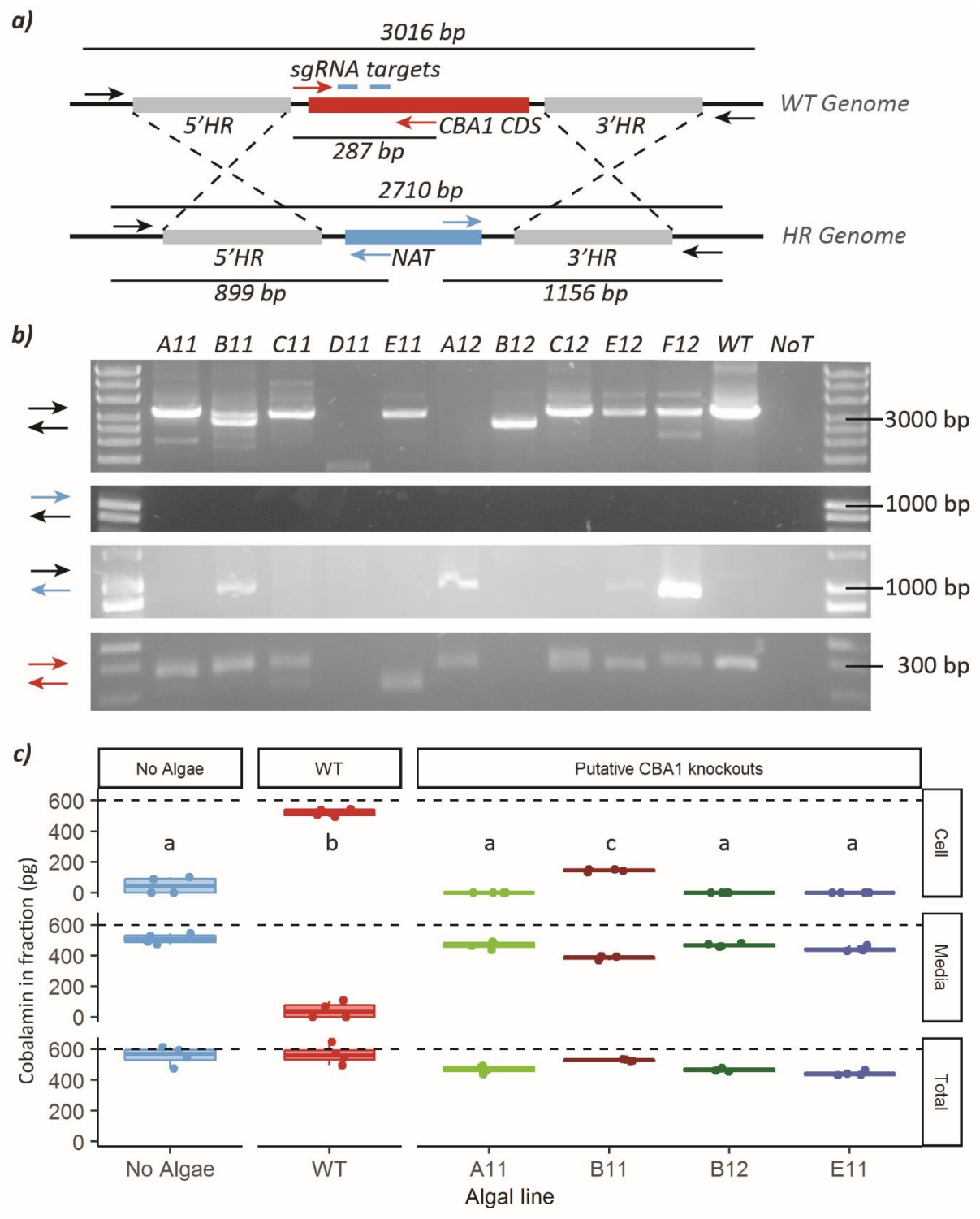


Figure 5.16. Genotyping and characterisation of CRISPR/Cas9* CBA1 *mutants. a) *Genotyping strategy with 4 primer pairs (colour coded) with expected band sizes indicated (associated with thin lines).* ***b)*** *PCR results for the four primer pairs in four different panels (colour coded on the left). B12 secondary colony shows band consistent with a biallelic homologous recombination genotype with exclusion of the* CBA1 *CDS. E11 shows bands consistent with a deletion at the sgRNA target site loci leaving no WT copy of the gene. B11 shows bands consistent with an heterozygous genotype combining WT and homologous recombination bands.* ***c)*** *Three biallelic knock-outs (A11, B12 and E11) and a monoallelic knock-out (B11) were grown in 20ml in four biological replicates to late exponential phase. The cultures where then harvested, concentrated to* $5 \cdot 10^6$ *cells·ml*$^{-1}$ *in 1 ml and incubated with f/2 plus 600 pg*

cobalamin for 1h in a microcentrifuge tube. After incubation, the cells fraction was recovered as a pellet and media fraction as supernatant after centrifugation. Vitamin B_{12} content in each fraction was determined by a bioassay based on the vitamin B_{12} auxotrophic growth of S. typhimurium *calibrated with known standards of cobalamin in a logistic model. A Tuckey HSD test was performed to compare the mean vitamin B_{12} content in the cellular fraction for each of the clones, WT and a negative control without algae. The results show that while WT uptakes cobalamin, the heterozygous knock-out (B11) shows partial uptake, and the biallelic knock-outs (A11, B12 and E11) cannot take up cobalamin.*

5.8. Discussion

The results presented in this chapter represent the first instance of CRISPR/Cas9 genome editing in *P. tricornutum* in our research group. As such they have opened many new possibilities which have been exploited here to investigate thiamine and cobalamin-related genes and will be useful to apply and use in future projects. The pilot test targeting *PtUMPS* and comparing multiple construct delivery methods is unprecedented in comparing side by side two existing Cas9 delivery methods, microparticle bombardment and bacterial conjugation, with a new method, electroporation. The results showed that electroporation gave the most colonies in the least time, and it gave a similar proportion of colonies resistant to 5-FOA compared with the two established methods. With the experiment on *PtUMPS* it was also realised that the primary colonies obtained showed genotypic mosaicism and that a round of restreaking and isolation of monoclonal secondary colonies was necessary to obtain knock-out lines.

With the pilot test targeting *PtUMPS* we also showed that CRISPR could mediate the incorporation of rationally designed expression cassettes in a given target region of the *P. tricornutum* genome by homologous recombination. The introduction of designed mutations by homologous recombination in *P. tricornutum* had only been achieved before using TALEN technology (McCarthy *et al.,* 2017; Weyman *et al.,* 2015). Since we obtained successful homologous recombination results, another research group published the successful introduction of a zeocin resistance cassette in the nitrate reductase gene by homologous recombination (Moosburner *et al.,* 2020). The introduction of whole expression cassettes by homologous recombination has also been reported in other Stramenopiles such as *N. oceanica* (Naduthodi *et al.*, 2019) and *T. pseudonana* (Belshaw *et al.,* 2017 - preprint). The homologous recombination design in this chapter uses the regions upstream and downstream of the CDS of the target gene as homology arms to swap the CDS of the target gene for the antibiotic resistance cassette, similarly to the previous examples in *N. oceanica* and *T. pseudonana*. In contrast, Moosburner *et al.* (2020) used the regions immediately upstream and downstream of the Cas9 cut site to introduce the antibiotic resistance cassette as an insertion disrupting the target gene CDS. Both designs allow ready genotyping of the mutants by PCR and have generated knock-out mutants with

intended phenotypes, but it would be worth comparing the two designs to test whether the proximity between homology regions and the proximity between these and the double-strand break has an impact on the homologous recombination efficiency.

Furthermore, with the experience of the pilot test on *PtUMPS* we were able to design an experimental pipeline for the generation of knock-out mutants of any gene by targeted replacement with a nourseothricin resistance cassette by homologous recombination. The pipeline was based on the co-electroporation of a first plasmid encoding Cas9 and a sgRNA pair targeting sites a few hundred basepairs away from each other on the target gene CDS with a second plasmid carrying a nourseothricin resistance cassette flanked by 800 bp long homology arms matching the regions upstream and downstream of the target gene CDS. The genotypic and phenotypic screen on the primary colonies revealed that almost 90 % of the colonies with homologous recombination bands were resistant to nourseothricin (Table 5.1). This high proportion provides a degree of confidence that selecting the colonies directly on nourseothricin after electroporation, or subjecting them to a post-selection screening on nourseothricin, would be an effective screening step, losing only 10 % of the potential mutants. By including a nourseothricin selection step before genotyping the primary colonies in our pipeline the proportion of colonies with an homologous recombination genotype, would increase from 35 % to 55 % according to the average experimental values, reducing the number of PCR reactions needed to detect enough primary candidates to be re-streaked. Nevertheless, the colonies would still need to be genotyped by PCR given that many nourseothricin resistant colonies did not show an homologous recombination genotype, likely explained by random integration of the nourseothricin resistance cassette in the genome.

The results in this chapter show that our CRISPR pipeline could generate biallelic knock-out mutants by homologous recombination for 4 out of the 8 targeted genes, *THIC, METE, SSSP* and *CBA1.* For *NMT1, METH* and *GGPPS* no secondary colonies showed a biallelic mutation in the target gene. Nevertheless, some of the colonies showed a monoallelic knock-out genotype with bands for both WT and homologous recombined genotypes. Given that the monoallelic knock-out genotypes for the *SSSP* and the *CBA1* gene led to intermediate phenotypes, it could be interesting to test the phenotype of the monoallelic knock-out mutants for *NMT1*, *METH,* and *GGPPS;* time did not permit this during my PhD. In the case of *NMT1,* it would be interesting to test whether the deletion of one of the alleles would lead to lower growth rates in the absence of thiamine, which would indicate that the gene is implicated in thiamine biosynthesis, like its homologues in *N. crassa* and *S. cerevisiae.* Or whether the monoallelic knock-out leads to a decreased capacity to salvage pyrimidine precursors in the thiamine biosynthetic pathway, which would indicate the gene conserves the function of its bacterial homologues. For now, we do not have an assay capable of measuring the pyrimidine salvage capacity

of *P. tricornutum*. The question of the potential function of PtNMT1 is further explored in Chapter 6. For *METH* monoallelic knock-out it would be interesting to test whether the loss of one of the alleles leads to growth anomalies in the presence or absence of cobalamin, and whether the transcript and metabolite levels of the C1 cycle are altered. Transcriptomics and targeted metabolomic studies on the C1 cycle in relation to cobalamin in *C. reinhardtii* have been of special interest for our research group in recent years, and we could apply similar techniques to investigate the phenotype of the *METH* and *METE* mutants alongside the *P. tricornutum WT* (Bunbury *et al.,* 2020; Holzer *et al.,* in preparation).

The monoallelic homologous recombined mutants generated for *GGPPS* are of particular interest given that they were not designed to knock-out the gene, but instead were engineered to swap the native promoter of the gene for the cobalamin-downregulated *METE* promoter. In this case it would be interesting to test whether cobalamin supplementation would lead to a controlled and predictable downregulation of *GGPPS* transcript and protein levels in the monoallelic knock-in. At a metabolic level, the controlled downregulation of GGPPS in the mutant could help us address some fundamental questions about terpene biosynthesis in *P. tricornutum*: Firstly, is there an available pool of GGPP in the cytosol? Is GGPP produced in the cytosol from GPP or FPP isoprenylation by a resident GGPP synthase, or do transport mechanisms channel GGPP from the chloroplast-targeted GGPPS (*Phatr3_19000*)? Which compounds, if any, are produced from GGPP in the cytosol? Could GGPP levels be boosted in the cytosol to increase production of diterpenoids in the cytosol by heterologous metabolic pathways? (Athanasakoglou & Kampranis, 2019). There are many labs around the world attempting to tap into the FPP and GGPP precursor pools in the chloroplast and cytosol in *C. reinhardtii* and *P. tricornutum* to produce novel high-value compounds with the introduction of heterologous terpene synthases and cytochrome P450s (eg D'Adamo *et al.,* 2019; Fabris *et al.,* 2020; Lauersen *et al.*, 2016; Mehrshahi *et al.,* 2020). To reveal new strategies to boost the pools of FPP and GGPP, more knowledge about the biosynthesis of these key precursors is needed, so engineering the native GGPP synthases to control their expression level with the *METE* promoter might be a promising avenue to funnel more precursors to the target high-value compounds. The monoallelic homologous recombined mutants for *GGPPS* developed in this chapter have the potential to help in both these directions, but since biallelic mutants would be more useful, more primary colonies could be screened, or the original constructs re-transformed.

For the transformants targeting *PSY* we were unable to obtain any colonies on selection plates, which could in part be explained if the *PSY* mutant had a lethal phenotype. In *C. reinhardtii*, no mutations in *PSY* (*Cre02.g095092*) are found in the CLiP insertional mutagenesis library, although this does have mutants with insertions in both the upstream and downstream gene in the same genomic region (Li

et al., 2019). This observation supports the idea that mutating *PSY* or even changing its promoter for the *METE* promoter could have a deleterious effect. Further attempts are required to confirm it is not possible to mutate *PSY* with our current methodology and with different sgRNAs, but if successful a *PSY* with expression controlled by cobalamin would be a great tool to repress an important pathway competing for a GGPP pool used as a precursor for the biosynthesis of high-value compounds. If we could demonstrate that a *GGPPS* monoallelic mutant obtained here or an hypothetical *PSY* mutant obtained in a second attempt can be regulated by the *METE* promoter, it would be the first time the regulatory sequence of an endogenous gene was altered in *P. tricornutum.* Such an advance would open the opportunity to use this method to specifically control the expression levels of any given gene, with application in both the study of the physiological role of genes and in the control of the expression strength and timing of endogenous metabolic pathways to enhance the production of target metabolic products. To test whether it is possible to swap an endogenous promoter for the *METE* promoter, the *UMPS* gene seems the best target for a pilot experiment, since the knock-out phenotype can be selected both positively and negatively, and we already have sgRNAs proven to mediate genome editing.

There are two main compatible explanations for why biallelic knock-outs could only be obtained for half of the targeted genes. Firstly, it could be explained by differences in the efficiency of the sgRNAs designed. Although the on and off target efficiency prediction tools are helpful in selecting promising sgRNA candidates, they do not account for all the possible variables and the sgRNA can fail to act *in vivo*. To overcome this challenge, sgRNAs can be tested *in vitro* in combination with Cas9 protein and a linear target DNA molecule amplified from the genome (Serif *et al.*, 2018). The samples are then run on an agarose gel and the on-target efficiency is calculated comparing the intensity of the bands for the template and cut DNA fragments (Serif *et al.,* 2018). We have demonstrated this process for *C. reinhardtii* genomic sequences using the CRISPR/LbCas12a method in our research group. Nevertheless, this process takes time, requires in vitro transcription and Cas9 production techniques, or the reagents can be bought at a significant cost. The alternative strategy used in this chapter was to include two sgRNAs for each target gene to increase the chance that at least one would work. However, in light of the results for genes that might be lethal, testing the sgRNA *in vitro* would seem to be an appropriate strategy before transformation to eliminate this possibility. With the knowledge that sgRNA work *in vitro*, failure to obtain biallelic homologously recombined mutants would point to the nature of the targeted gene. The second explanation for why biallelic mutants could not be obtained for *NMT1, METH* and *GGPPS* is that the knock-out phenotypes for the different target genes might lead to different degrees of viability. The phenotype of these genes cannot be anticipated fully, as the physiological effects of their knock-out could be multiple and unknown, and their function

might be essential for the survival of the cells. To overcome the challenge of phenotypic lethality, it is important to find appropriate culture conditions to select the mutants. We could attempt to select *METH* knock-outs with an exogenous source of methionine, and *GGPPS* knock-outs with an exogenous source of GGPP. Equally we could try to select for *PSY* knock-out with an exogenous source of phytoene.

Turning to the genes for which we could obtain biallelic knock-out mutants, the *THIC* knock-out revealed that *P. tricornutum* does not require a functional copy of THIC to grow, and thus it must have an alternative way to obtain thiamine. This result contradicts the established understanding of diatom thiamine metabolism, by which the pyrimidine moiety is considered to be synthesised solely by THIC (HMP-P synthase) (Bertrand & Allen, 2012). The other gene annotated as *THIC* in the *P. tricornutum* genome (Phatr3_EG00135; Section 3.2.1) seems unlikely to provide an alternative source of HMP-P given it does not have any associated EST reads and is only 124 amino acids long compared to other THIC homologues, including the PtTHIC knocked-out here, that are usually more than 600 amino acids long. A potential candidate to synthesise HMP in the absence of THIC is NMT1, which is homologous to HMP synthases in fungi and periplasmic components of ABC transporters for pyrimidine precursors in bacteria. The possible alternative sources of HMP are discussed in further detail in Chapter 6.

Research in our group has established a strong link between cobalamin auxotrophy and the absence of *METE* in algal genomes (Croft *et al.,* 2006; Helliwell *et al.,* 2011; Helliwell *et al.*, 2015) and this has been confirmed as more algal genomes have been sequenced (Bertrand & Allen, 2012; Helliwell, 2017). The cobalamin auxotrophy of the *METE* biallelic knock-outs confirms the prediction and offers an interesting phenotype to test a number of different hypotheses. Similar newly generated *C. reinhardtii METE* mutants show slower growth at low cobalamin concentrations or in co-culture with cobalamin -producing bacteria compared to *Lobomonas rostrata*, a naturally occurring cobalamin auxotroph closely related to *C. reinhardtii* (Bunbury *et al.,* 2020). It would be interesting to replicate the experiments with the *P. tricornutum METE* mutant and a naturally-occurring diatom missing *METE* like *Pseudo-nitzschia multiseries* or *T. pseudonana* to test whether the proposed adaptation of cobalamin auxotrophs to cobalamin deplete and co-culture conditions has evolutionary relevance in diatoms as well as in Chlorophytes. Finally, the cobalamin requirement of the *P. tricornutum METE* mutant could be exploited to increase the pressure for the diatom to enter into synthetic co-culture with cobalamin-producing bacteria. Co-cultures are a technique used to establish robust cultures of microalgae for biotechnological purposes (Lian *et al.,* 2018). Having the ability to co-culture a model diatom that has an extensive synthetic biology toolbox would give us the possibility to expand such investigation to a group of algae of key ecological relevance in the marine environment.

SSSP had already been proposed to be a transporter for thiamine (or one of its precursors) (McRose *et al.,* 2014). The results in this chapter confirm that SSSP is essential for thiamine uptake in *P. tricornutum* with experimental evidence. It would be interesting to test whether SSSP also transports thiamine precursors with uptake assays with WT and the *SSSP* knock-out and determining whether pyrimidine and thiazole precursors contribute to an increase in intracellular thiamine levels. Similarly, CBA1 had already been proposed as a cobalamin acquisition protein in an overexpression experiment (Bertrand *et al.,* 2012) and the results in this chapter confirm the role of this protein and are included in a manuscript (Sayer *et al.*, in preparation). It will be interesting to investigate the CBA1 mechanism of action, and whether it acts as a membrane receptor involved on receptor-mediated endocytosis in a mechanism similar to that described in humans (Setharam, 1999), as well as which other partners are necessary and interact with CBA1 and SSSP. Dr Gonzalo Mendoza-Ochoa and Shelby Newsad in our group are now designing experiments with molecular techniques to isolate and identify the proteins interacting with CBA1.

5.9. Technical challenges and opportunities

CRISPR genome-editing methods are fast evolving in *P. tricornutum* and are increasingly being adopted by the scientific community to study the physiology of the model diatom. However, there are still some challenges that have not yet been resolved, and also new opportunities to tap into with the adoption of the CRISPR/Cas9 technique in *P. tricornutum*. Here, I want to discuss in more detail the possibility of preventing colony mosaicism, creating homologous recombined mutants without random genomic integration, and exploiting CRISPR/Cas9 to generate new auxotrophic mutants with potential for synthetic biology and synthetic ecology applications.

Addressing colony mosaicism

To generate the knock-outs in this Chapter it was necessary to first genotype and phenotype primary colonies, and then re-streak the promising transformants to obtain monoclonal colonies, which then had to be genotyped and phenotyped again. This process is required given that Cas9 generates mutations in successive generations in each colony, giving rise to a variety of genotypes in a single colony on a selection plate, a phenomenon known as colony mosaicism (Huang & Daboussi, 2017). To the best of my knowledge all CRISPR/Cas9 reports in *P. tricornutum* have used the re-streaking approach to solve this challenge and obtain monoclonal colonies. However, re-streaking and re-screening the colonies is a lengthy process and remains the main bottleneck: avoiding colony mosaicism would save at least 3 weeks of work and streamline CRISPR protocols. To reduce the incidence of colony mosaicism three main strategies could be adopted in combination: speeding up

Cas9 editing, removing Cas9 after editing, and stalling cell cycle progression during editing (Mehravar *et al.,* 2019). In a study in bovine embryos, it was realised that to reduce CRISPR-derived genetic mosaicism it was critical that the editing event had to precede the replication of DNA (Lamas-Toranzo *et al.,* 2019). If extrapolated to *P. tricornutum*, arresting the cell cycle at G_1 phase during Cas9 delivery and editing could reduce or even eliminate the problem of cell mosaicism. In *P. tricornutum*, incubation in the dark during exponential phase leads to cell cycle arrest in G_1 phase (Brzezinski *et al.,* 1990) and 6-(cyclohexylmethoxy)-9H-purin-2-amine has also been used to arrest cell cycle at the G_1 phase by binding to two cyclin-dependent kinases, CDKA1 and CDKA (Kim *et al.,* 2017). Either treatment could be used to stall cells at the G_1 phase during Cas9 delivery and editing in cells containing a single copy of the genome. Cell cycle arrest with small-molecule inhibitors has already been used to enhance CRISPR/Cas9 HDR in human cells (Lin *et al.,* 2014).

The two delivery methods that allow for removing Cas9 after editing are RNP bombardment (Serif *et al.,* 2018) and episome conjugation (Slattery *et al.,* 2018). Episome delivery would in principle take longer to achieve edited strains than RNP delivery, since Cas9 and the sgRNAs have to be expressed and assembled. And it would take at least a generation to remove Cas9 by episome loss. In fact, episome segregation efficiency has been calculated to be 97 %, so this would be a rare event (Sharma *et al.*, 2018). Even if it were possible to repress Cas9 expression (eg. with the *PtMETE* promoter), Cas9 activity would take longer to decay than with RNP delivery since the protein could continue to be translated from mRNA until the transcript is degraded. Thus, in principle removing an episome-based Cas9 seems incompatible with arresting cell cycle. An alternative would be to let the episome-based Cas9 be expressed and act for 24 to 48h in liquid culture under positive selection before switching the cells to a media without selection for sufficient time to remove the episome, and only then plating in selection plates to isolate monoclonal edited strains. However, with high segregation efficiencies, it would be challenging to find colonies that have lost Cas9. To address this problem, a conditional negative selection marker could be included in the episome and cells be cultured in negative selection media before being plated to pre-select for the cells that have lost the episome (See figure 5.17 for a graphic representation of episome negative selection). To the best of my knowledge, the only negative selection markers used in *P. tricornutum* have been endogenous *UMPS* or *APT* (Serif *et al.,* 2018). However, it is only possible to use these markers in a heterologous construct in a *ΔUMPS* or *ΔAPT* background, which would likely have an altered metabolism (see Slattery *et al.,* 2020 for example). Therefore, the diatom synthetic biology community would greatly benefit from the adoption of orthologous negative selection markers. In the alveolate *Plasmodium falciparum* cytosine deaminase and thymidine kinase are used as negative selection markers (Duraisingh *et al.,* 2002). Cytosine deaminase catalyses the deamination of 5-fluorocytosine (5-FC) into 5-fluorouracil which is then

phosphoribosylated into the toxic compound 5-fluorouridine monophosphate, the same product of 5-FOA decarboxylation and phosphoribosylation by UMPS. Thymidine kinase phosphorylates the drug ganciclovir to produce the toxic ganciclovir triphosphate. If *P. tricornutum* is experimentally confirmed to be naturally resistant to 5-FC and ganciclovir, cytosine deaminase and thymidine kinase could be effectively adopted as negative selection markers to assist in selection of CRISPR mutants among many other applications. In the absence of suitable negative selection markers, an alternative solution would be to use fluorescent markers coupled with fluorescence activated cell sorting or microdroplet sorting, a method developed for *P. tricornutum* in part by our research group (Geisler *et al.,* in preparation).

In contrast to episomes, RNPs have been shown to act immediately after transformation, and degrade almost completely 24h after delivery in human cells (Kim *et al.,* 2014), making RNPs the best candidates to avoid colony mosaicism. RNPs could be used on cells arrested in G_1 phase and incubated for at least 24h in liquid culture under cell cycle arrest to allow for Cas9 to act and then degrade before plating on selection plates to isolate single colonies. However, the CRISPR/Cas9 RNP delivery method used in *P. tricornutum* is very inefficient (<0.4 % edited cells/transformed cells) and has only been able to generate knock-outs of genes with a selectable phenotype, like *UMPS*, or other genes in combination with a selectable gene (Serif *et al.,* 2018). Thus, we would have the problem of having all edited strains with an additional knock-out in *UMPS*, which is an important enzyme in the nucleotide-base metabolism and its mutation leads to uracil auxotrophy. In *N. oceanica* the co-electroporation of Cas RNPs with a dsDNA homologous repair template encoding a zeocin selection cassette increased editing efficiency up to 93 % of the clones selected on zeocin for FnCas12a (Naduthodi *et al.,* 2019). Our research group has adopted a method for *C. reinhardtii* (Bunbury *et al.,* 2020) based on CRISPR/LbCas12a RNP and a single-stranded oligodeoxynucleotide homology repair template co-electroporation that has reported an efficiency of up to 16 % editing (Ferenczi *et al.,* 2017). We are thus in a good position to adapt either the *N. oceanica* or the *C. reinhardtii* RNP electroporation-based methods for *P. tricornutum.* RNP delivery combined with HDR, which will be explored in greater detail in the next subsection, could address the low efficiency of RNP editing in *P. tricornutum* and holds promise to solve the challenge of colony mosaicism.

Homology directed repair without random genome integration

HDR offers a range of opportunities for genome editing through the precise and rational modification of targeted genomic sequences. In this chapter, HDR has been used to introduce an antibiotic resistance cassette to facilitate phenotypic and genotypic screenings, and to switch the promoter of a native gene for a tuneable promoter, although it still needs to be tested whether the gene can be regulated by the new promoter. In addition, HDR can be used to introduce fluorescent tags or epitopes

for the localisation and quantification of native genes (Kuang *et al.,* 2017; Roberts *et al.,* 2017), and has been proposed for introducing landing sites based on bacterial site specific recombinases to offer a standardised insertion point for the expression of heterologous constructs avoiding the undesirable positional effects of random integration (Kroth *et al.,* 2018). However, HDR has only been achieved in *P. tricornutum* with methods that require the random integration of Cas9 or TALEN nucleases (McCarthy *et al.,* 2017; Moosburner *et al.,* 2020, Weyman *et al.,* 2015; this chapter) leading to potential disruption of endogenous genes and increased potential of off-target effects due to the continued expression of the nucleases (Kroth *et al.,* 2018). Thus, the next step in the adoption of genome editing by homologous directed repair in *P. tricornutum* would be to achieve precise editing events without random genomic integration.

RNP bombardment is a DNA-free Cas9 delivery method which allows insertion-less gene editing in *P. tricornutum* (Serif *et al.,* 2018). It should be tested whether co-transforming RNPs with an homology repair template containing an antibiotic resistance cassette could lead to homologous directed repair in *P. tricornutum* (Figure 5.17a). This approach has been shown to work for *N. oceanica* with different Cas variants (Cas9, FnCas12a, AsCas12a and LbCas12a) with up to 93 % of antibiotic resistant colonies having an homology directed mutation in the target loci (Naduthodi *et al.,* 2019). It is possible that in this approach some of the homology repair templates could be integrated randomly in the genome or be inserted in the target site via NHEJ (Figure 5.17b). It would be then necessary to test by Southern blot and PCR, or equivalent techniques, whether the repair template has been incorporated only at the target site by HDR. The screening process could be facilitated by including a negative selection marker outside of the homology arms in the homologous recombination template, those cells that have incorporated the repair template by random insertion will not be able to survive since they will encode the negative selection marker (Figure 5.17c).

Episomes have been shown not to integrate in the *P. tricornutum* genome (George *et al.,* 2020), so bacterial conjugation could also be used to generate CRISPR/HDR mutations without integration. In *P. falciparum* an episome containing Cas9 and an episome containing the sgRNA expression cassette and the homology recombination template were co-transformed to integrate a GFP reporter cassette in a precise genomic location without random integration (Mogollon *et al.,* 2016). It has already been shown that *P. tricornutum* can contain two episomes with different selection markers in a single cell (Slattery *et al.,* 2018), and the low copy number of episomes does not seem to prevent homologous recombination from an episome, based on the success of the *P. falciparum* experiment. The *P. falciparum* approach could be adopted for *P. tricornutum* with two sequential episome conjugations to generate rationally-designed HDR mutations without positional effects (Figure 5.17a). The advantage of having two separate episomes is that a background strain containing an episome

encoding Cas9 could be conjugated subsequently with the homology template and sgRNA episome for each specific target (Figure 5.17a). Alternatively, given that size does not seem to be limiting for episomes (up to 33 kbp episomes have been conjugated into *P.* tricornutum; Slattery *et al.,* 2018), all necessary components for CRISPR/HDR could also be cloned in a single episome with the GoldenGate based cloning methods in our lab without additional steps (Figure 5.17a).

A priori, the main challenge with episome-based homologous recombination would seem to be that the replaced genomic sequence would be recombined between the homologous arms into the episome allowing for successive rounds of editing and HDR (Figure 5.17b). To avoid continuous cycles of editing and HDR, the homology repair template could be designed to split the sgRNA target site and avoid re-editing, although this would restrict the range of possible mutations. An alternative solution would be to remove the episome containing the replaced genomic sequence and/or Cas9 after homologous directed repair. If the cells were only selected for the antibiotic resistance marker between the homology arms (ie. nourseothricin for *NAT*) this could be achieved naturally by segregation at a really low efficiency given that the recombined episome would not have the resistance cassette anymore. The process could be sped up with the introduction of a negative selection marker (eg. cytosine deaminase) outside of the homology recombination arms. Yeast cytosine deaminase (*yfcu*) has already been added in homology repair templates outside of the homology recombination arms to select for those colonies with only HDR integration in *P. falciparum* (Mogollon *et al.,* 2016). Under positive-negative selection (eg. 5-fluorocytosine and nourseothricin), only those cells that have included the antibiotic resistance marker by HDR and successively lost the episome would survive (Figure 5.17c). In the case of transforming Cas9 with the sgRNAs and homology repair template in a single episome, positive-negative selection would also remove Cas9 and could have the co-benefit of minimising colony mosaicism.

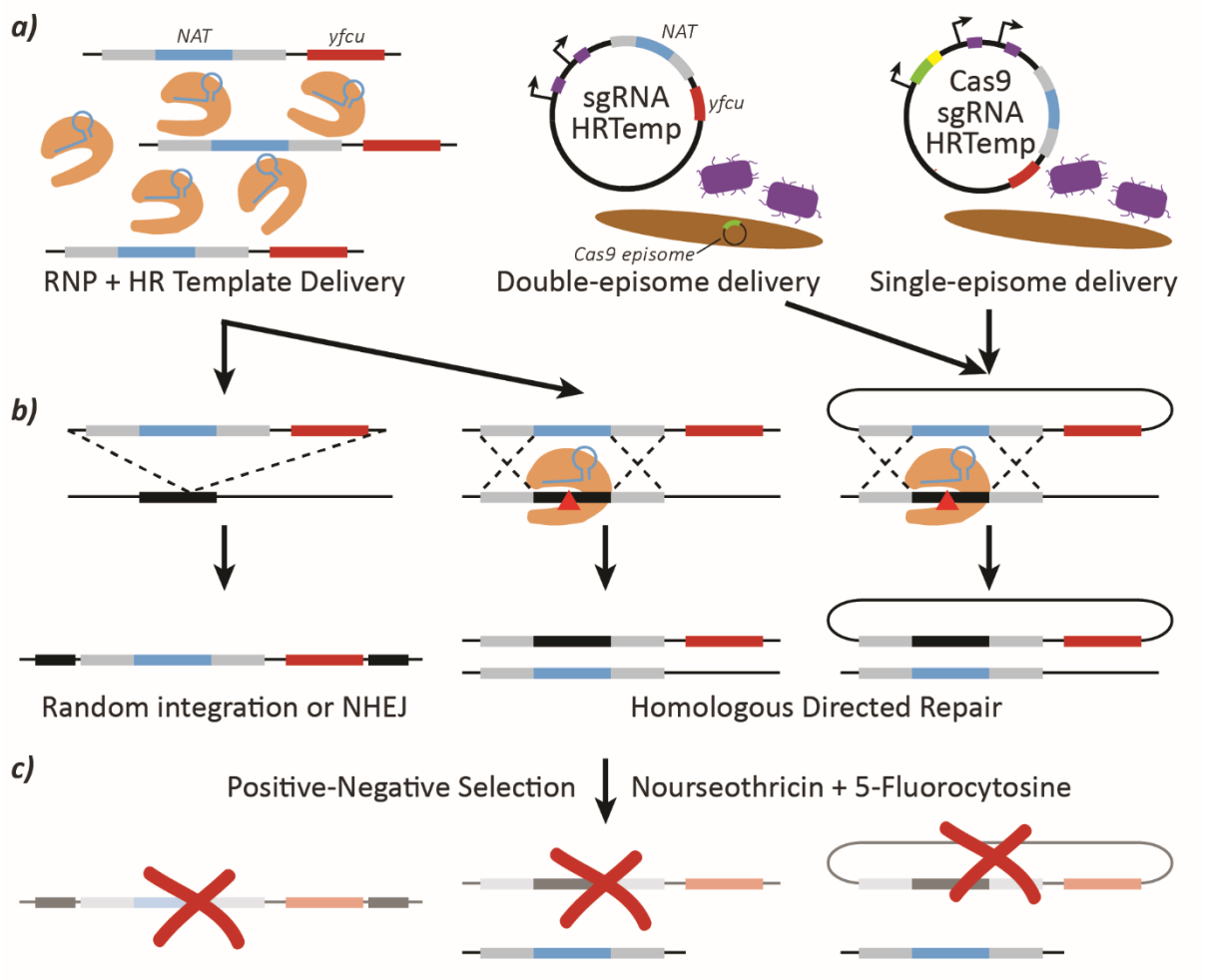


Figure 5.17. Proposed strategies to achieve precise mutations by CRISPR-mediated homologous directed repair without random genomic integration. a) *Three different delivery methods could be used: (i) RNP and homologous repair template (HRTemp) co-delivery via electroporation or bombardment, (ii) delivery of an episome containing sgRNAs cassettes and homology repair template into a strain carrying a* Cas9 *cassette encoded in another episome, (iii) delivery of an episome containing* Cas9*, sgRNAs and the homology repair template into WT* P. tricornutum*.* ***b)*** *RNP delivery could lead to the desired HDR, or unintended random integration or NHEJ of the homologous repair template. In principle, episome-based HDR would lead to the inclusion of the target genomic sequence in the recombined episome which could lead to iterative cycles of editing and HDR.* ***c)*** *The inclusion of a positive selection marker conferring nourseothricin resistance (*NAT*, blue) between the homology recombination arms, and a negative selection marker conferring sensitivity to 5-fluorocytosine (*yfcu*, red) outside the homologous recombination arms would allow to use positive-negative selection with nourseothricin and 5-fluorocytosine to remove those cells that have randomly integrated the homology repair template and that have retained the episome recombination product.*

Vitamin auxotrophic mutants to expand the marker gene toolbox

The *METE* knock-out developed by CRISPR in this chapter has shown it is possible to generate vitamin auxotrophic mutants for *P. tricornutum.* CRISPR, preferably through non-integrative methods, could be used to generate a library of *P. tricornutum* auxotrophs for the different B vitamins by knocking-out a single gene involved in their respective biosynthetic pathway. Our research group has extensive expertise in the topic of vitamin metabolism in algae, and so we are in a good position to identify the most amenable gene targets (eg. *TH1* for thiamine (B_1) auxotrophy, *PDX1* for pyridoxine (B_6) auxotrophy) and to culture and analyse the auxotrophic mutants. A library of *P. tricornutum* B vitamin auxotrophic mutants could have promising applications for synthetic biology and synthetic ecology.

There is a limited availability of selection markers for *P. tricornutum*. Zeocin and nourseothricin resistance markers are the most commonly used (Apt *et al.,* 1996; Zaslavskaia *et al.,* 2000), and only recently have a resistance gene for blasticidin (Buck *et al.,* 2018) and a selection system based on a mutation to the endogenous phytoene desaturase conferring resistance to the herbicide norflurazon (Taparia *et al.,* 2019) been added to the toolbox of selection markers. With the increasing complexity of synthetic biology and metabolic engineering in *P. tricornutum* more marker genes are needed for gene stacking of various constructs. B vitamin auxotrophy could be exploited to expand the range of selection markers available in *P. tricornutum.* The genes missing in the auxotrophic mutants could be included in heterologous genetic constructs to positively select the synthetic construct by complementation of an auxotrophic background strain selected in vitamin-deplete plates. Uracil and tryptophan auxotrophic markers on knock-out strains have been recently developed in *P. tricornutum* in a similar way to the amino acid auxotrophy markers commonly used in *S. cerevisiae* (Slattery *et al.,* 2020). In contrast to amino acid or nucleotide auxotrophic markers, vitamin auxotrophy markers would require much lower concentrations of exogenous vitamin to be maintained, and the effects of vitamin supplementation on gene expression and metabolism would in principle be less widespread than amino acid supplementation. The use of auxotrophic markers also offers an additional level of biocontainment by avoiding the use and potential environmental release of antibiotic resistance cassettes. In yeast genetic engineering, auxotrophic markers are used to positively select and maintain autonomous plasmids and also for targeted insertions in the genome where the constructs are inserted by homologous recombination in the auxotrophic marker loci to restore the function of the marker gene and incorporate the heterologous construct (Tuite, 1992). Similarly, vitamin auxotrophy markers could be used in *P. tricornutum* to select for episomes, and as positive markers for targeted insertion of constructs in the auxotrophy marker loci via CRISPR/HDR.

Vitamin auxotrophy could also have applications for synthetic ecology, which seeks to rationally design artificial microbial communities with specific functions using individual species as building

blocks (Dunham, 2007). Synthetic ecology can construct model systems to study the dynamics and functions of species interactions in the natural environment (Vranken *et al.,* 2019) and holds promise to boost the productivity of algae bioproduction (Kazamya *et al.,* 2012). Interactions between algae and bacteria have been shown to be mediated by the interchange of vitamins between different species in co-culture (Cooper *et al.,* 2019). *P. tricornutum* has an extensive genetic engineering toolbox empowering it for the production of high value compounds from heterologous biosynthetic pathways, but naturally does not require any vitamin for growth. *P. tricornutum* engineered B vitamin auxotrophs might promote the establishment of interdependent mutualisms with microbial partners able to meet the B vitamin requirement and reliant on photosynthate from *P. tricornutum*. A *P. tricornutum* auxotrophic library could allow to choose the specific vitamin requirements to match the biosynthetic capacity of a chosen bacterial target, enabling logic and modular design of synthetic co-cultures. Double auxotrophic mutants could be generated to facilitate increasingly complex co-cultures with more than two species.

Chapter 6. Summary of work and future perspectives

Contrary to established knowledge for species of bacteria, fungi and plants, the evidence gathered in this thesis shows that the diatom *P. tricornutum* does not require THIC for thiamine biosynthesis and demonstrates that the predicted TPP aptamers in diatoms do not respond to thiamine supplementation despite their high sequence conservation. Additionally, the results presented here confirm that PtTHIC is downregulated by cobalamin supplementation and suggest that a conserved motif identified upstream of cobalamin-downregulated genes in diatoms is involved in cobalamin-downregulation. Finally, the development of a CRISPR/Cas9 method with homologous directed repair has enabled the generation of various *P. tricornutum* knock-outs, which confirmed that SSSP is a thiamine transporter, CBA1 is a cobalamin transporter and METE is required for cobalamin-independent growth, demonstrating that the CRISPR technique can greatly accelerate physiological research in diatoms. Here, I wish to comment on the evolutionary, ecological and physiological relevance of some of the new findings in this thesis and propose potential experiments to address questions opened by these results.

The thiamine biosynthetic pathway: is there an alternative to THIC?

The biallelic *PtTHIC* knock-out generated by CRISPR/Cas9 was able to grow in thiamine-deplete media (Figure 5.12), and therefore *P. tricornutum* must be able to source HMP or thiamine without THIC. This observation contradicts much of our knowledge on thiamine biosynthesis, and as such should be validated by testing more independent biallelic THIC knock-outs, and testing there is no unnoticed THIC expression by RT-PCR. If confirmed, how *P. tricornutum* obtains thiamine in the absence of THIC would be a relevant question given that virtually all organisms need thiamine. The knock-outs were diluted 1:10 twice in media without thiamine and the growth of the culture was then tracked for 20 days, and although thiamine or its precursors might be present in the de-ionised water or sea salts used to make up the media (Carini *et al.,* 2014), it seems unlikely that thiamine can accumulate in the cell in sufficient levels to survive this long. To test whether the f/2 media could contain some uncontrolled thiamine or pyrimidine precursors it could be used with or without thiamine supplementation to grow algae with known auxotrophy for HMP, like *O. tauri* (Cooper *et al.,* 2018) or *E. huxleyi* (Bertrand & Allen, 2012). Alternatively, an isotope-labelling experiment supplying $^{15}NO_3$ in the media followed by HPLC-MS determination of the thiamine isotope present in *P. tricornutum*

would elucidate whether thiamine or its precursors are present in the water and/or sea salts, or they are synthesised during the culture. Another possible source of thiamine could be bacterial contamination, which seems unlikely given that *P. tricornutum* cultures were treated with antibiotics before transformation and always handled following sterile techniques. This hypothesis could be tested by measuring the thiamine present in the media fraction of the PtSSSP knock-out with the HPLC thiamine quantification assay, which has a limit of detection of 50 nM. The only alternative left is that thiamine could be synthesised by *P. tricornutum* independently of THIC.

Besides *THIC,* the *P. tricornutum* genome has another annotated HMP-synthase homologous to the fungal NMT1 (Figure 3.1). Based on overall sequence identity, and the conservation of key residues in the binding pocket, it has been suggested that NMT1 and ThiY, the periplasmic component of a FAMP ABC-transporter found in bacteria, share a common ancestor (Bajor *et al.,* 2014; Bale *et al.,* 2010). Since ThiY and NMT1 both have a conserved periplasmic binding protein (PBP) fold, it seems plausible that ThiY first evolved from a PBP ancestor and was then repurposed as a HMP-P synthase in an ancestor common to all fungi. The putative refunctionalisation of ThiY into the resource-intensive and single-turnover HMP-synthase would have faced negative selection pressure and would have required a strong positive selection likely conditioned by HMP auxotrophy in a thiamine-deplete environment. The *P. tricornutum* NMT1 candidate is phylogenetically closer to ThiY (Figure 3.3), so it is not clear whether PtNMT1 could have an HMP-synthase function that could sustain thiamine biosynthesis in the PtTHIC knock-out. Other algal species like *Aureococcus anophagefferens* (Stramenopile) and *E. huxleyi* (Haptophyte) lack *THIC* but contain an *NMT1* candidate (Dom Absolon, personal communication) and are thiamine auxotrophs (Tang *et al.,* 2010; Bertrand & Allen, 2012), lending support to a functional annotation of the algal NMT1 candidates as transporters for pyrimidine precursors rather than HMP-P synthases. At a molecular level, ThiY is tethered to the cell membrane in gram-positive bacteria by the acylation of a conserved cysteine residue (C19 in *B. halodurans* ThiY) which also appears to be conserved in *P. tricornutum, T. pseudonana* and *G. theta* (Figure 3.2; Bale *et al.,* 2010). However, PtNMT1 has more active site residues conserved with the fungi NMT1 homologues than the bacterial ThiY and does not have a predicted signal peptide for the secretory pathway indicating it is probably localised in the cytoplasm. All in all, with the bioinformatic evidence gathered here it is not possible to draw a firm conclusion on whether PtNMT1 has a transport or a catalytic function and this should be discerned experimentally.

Experimentally determining the function of the algal NMT1 candidates would clarify our understanding of thiamine metabolism in organisms belonging to several eukaryotic supergroups with significant ecological significance. To address this question, a *P. tricornutum* CRISPR/Cas9 knock-out for NMT1 could be tested for its capacity to grow in the absence of thiamine, or uptake pyrimidine

precursors. Unfortunately, the CRISPR/Cas9 attempt in this thesis did not yield an NMT1 biallelic knock-out but a fresh attempt with new sgRNAs could in principle generate the mutant of interest. A complementary experiment would be to replicate complementation assays in the RWY16 yeast strain, which misses all genes coding for HMP-P synthases, to test whether PtNMT1 could rescue its growth in the absence of thiamine (Coquille *et al.,* 2012). Or similarly, a *B. halodurans* or *B. subtilis* mutant missing ThiY could be complemented to test whether PtNMT1 could rescue the FAMP transport function. Finally, since *S. cerevisiae* PLP synthases SNZ2 and SNZ3 are downregulated by thiamine supplementation (Paxhia & Downs, 2019), it would also be interesting to test whether the *P. tricornutum* PLP synthase (*PtPDX1, Phatr3_J29885*) is affected by thiamine supplementation and whether the *PtNMT1* and *PtTHIC* transcript levels are altered under PLP-replete conditions.

If PtNMT1 was confirmed to be involved in the synthesis of HMP-P, it would lead to a previously unconsidered scenario where two HMP-P synthases, THIC and NMT1, coexist in the same organism. This new metabolic arrangement would have interesting ecological and regulatory implications. So far, our understanding is that in plants, algae and bacteria, HMP-P is synthesised from AIR by THIC and in fungi HMP-P is synthesised from PLP by NMT1 (Figure 1.2). Both enzymes are known to have a significant metabolic cost: NMT1 is a suicide enzyme that uses a histidine in its active site as substrate (Coquille *et al.,* 2009) and THIC is inactivated after a few turnovers (Palmer & Downs, 2013) consuming the 0.2 % of the maintenance energy in plants as calculated by Hanson *et al.* (2018). However, it is likely that one of the two enzymes has a slightly higher metabolic cost, and in evolutionary terms one could hypothesise that the organisms that have lost the less efficient enzyme would have been favourably selected. To test this hypothetic scenario, one could transform *S. cerevisiae*, which naturally synthesises HMP-P via THI5, with *A. thaliana* THIC and in an artificial evolution experiment test whether over hundreds of generations deleterious mutations appear on the native *THI5* or in the heterologous *THIC*. The experiment could also be carried out similarly in *E. coli* expressing heterologous THI5 from *S. cerevisiae.* However, the coexistence of both enzymes could also confer the organism with more metabolic flexibility, which would be a positive selected trait in environments with rapid changes in nutrient availability, where one of the enzymes can synthesise HMP-P when the precursors for the other enzyme are scarce. In this latter situation, it would be important that the two HMP-P synthases are well regulated to respond to the levels of precursors and cofactors, similarly to the METE/METH pair for methionine biosynthesis. An example of regulation depending on substrate availability can be found in the downregulation of PtTHIC by cobalamin supplementation. Cobalamin deprivation has been suggested to reduce the levels of SAM in diatoms, which is the methyl-donor in the reaction catalysed by THIC, explaining the downregulation of the gene in *P. tricornutum* (Bertrand *et al.,* 2012).

Alternatively, if PtNMT1 was confirmed to be involved in the transport of pyrimidine precursors, it would expand our understanding of thiamine precursor salvage in algae and would help elucidate the role this process plays in shaping marine microbial metabolism and community composition. NMT1 candidates are found in Chlorophytes, Cryptophytes, Haptophytes, Mamelliales, Rhizaria, Stramenopiles and Streptophytes (Figure 3.1; McRose *et al.,* 2014). In ecological and evolutionary terms, it is still an open question how higher plants in land environments could benefit from having transporters involved in thiamine precursor salvage. In contrast, it seems that the salvage of thiamine precursors confers a clear selective advantage for free-living organisms in aquatic environments. In particular, free thiamine has been shown to degrade into FAMP when exposed to UV light (Carini *et al.,* 2014; Gutowska *et al.,* 2017), a process which is specifically relevant to planktonic environments. The salvage of free FAMP could support organisms auxotrophic for the pyrimidine moiety and minimise the metabolic burden of prototrophs. The characterisation of a FAMP transporter conserved in all major algal lineages would expand our understanding on the role of thiamine precursor salvage in the ecological dynamics of marine microbial communities.

TPP aptamers in diatoms: high conservation without its canonical function?

Quantitative RT-PCR and 3'RACE experiments in this thesis showed that *PtTHIC* and *PtSSSP* are not regulated at the transcript level by thiamine supplementation (Figure 3.11; Figure 3.12). Additional experiments with reporter constructs confirmed that the *PtTHIC* promoter and 3'UTR cannot regulate the expression of a transgene (Figure 4.3), and that the putative *PtTHIC* riboswitch does not regulate expression to maintain homeostasis of thiamine levels (Figure 4.4). Similarly, *PtSSSP* does not seem to be downregulated by thiamine supplementation at the protein level, given that thiamine uptake did not stop when really high levels of thiamine (10 µM) were taken up by *P. tricornutum* (Figure 3.9). The lack of function of the putative TPP aptamer does not seem to be restricted to *P. tricornutum* as it can also be observed in at least one other diatom. Both *T. pseudonana* and *P. tricornutum* appeared to be resistant to pyrithiamine (Figure 3.13) and *TpTHIC* levels are not affected by thiamine supplementation (Figure 3.11).

From an evolutionary perspective, there is a conceptual tension between the lack of function of the diatom predicted TPP riboswitches and their strong sequence conservation (Figure 3.7). Strong sequence conservation is generally associated with a conserved function, and in the case of sequences in untranslated regions this becomes more pronounced, as mutations could easily appear if there was no selection pressure to conserve the sequence. Thus, why do the predicted TPP riboswitches in diatoms conserve the sequence if they have lost their expected function? In the light of the long

evolutionary history of riboswitches, dating back to the RNA World more than 3.5 billion years ago (Monteverde *et al.*, 2007), the possibility that the diatom TPP riboswitch sequence is still conserved because it has have only recently lost its function seems remote. Given the evidence gathered in this thesis, the earliest point at which this hypothetical loss of function could have arisen is at the last common ancestor between *P. tricornutum* and *T. pseudonana* some 100 million years ago (Armburst, 2009), in which time mutations would have had sufficient time to erode the sequence conservation of a locus without function. Consequently, the conundrum posed by the sequence conservation of the diatom TPP aptamers and the lack of thiamine response seems to have another explanation, and although an answer cannot be provided conclusively with the evidence gathered in this thesis, I propose that it might be related with the unexpected behaviour of the thiamine biosynthesis pathway.

A first hypothesis is that the sequence conservation despite the lack of genetic response might be explained by an unknown source of thiamine in f/2 media that keeps the TPP riboswitches constantly repressed and masks any possible response of the thiamine-related genes to thiamine supplementation. However, the transcript levels of the thiamine-related genes investigated by RT-qPCR do not seem to be at a significant low level and in fact, cobalamin supplementation can drive transcript levels to a significantly lower level (Figure 3.11). Additionally, the *PtTHIC* promoter can drive the expression of a Ble-Venus reporter to sufficient levels to confer zeocin resistance and to fluorescence levels comparable with the *PtEF2* promoter (Figure 4.3; Figure 4.10). It could also be suggested that an unknown source of thiamine could mask the pyrithiamine effect. However, in *C. reinhardtii* thiamine levels need to be a minimum of 10 % of the pyrithiamine concentration in order to rescue growth (Balia Yusof Thesis, 2011). That would represent having at least 1 µM thiamine of unknown origin in our media, which seems highly improbable given the low environmental concentrations of the vitamin. Nevertheless, in *C. reinhardtii*, HET supplementation alone can rescue growth in the presence of pyrithiamine (Figure 3.13). It could then be argued that since the thiazole branch in *P. tricornutum* is not associated with TPP riboswitches, the constant supply of HET together with the degradation of pyrithiamine to HMP can rescue *P. tricornutum* growth and mask the effect of pyrithiamine on the TPP riboswitch. As suggested in the previous section, the hypothesis of an unknown thiamine source could be tested by performing a comparative thiamine quantification assay in the media fraction of *P. tricornutum* WT and the *SSSP* knock-out.

A second hypothesis, based on the idea that *P. tricornutum* might have an alternative pathway to synthesise thiamine that does not involve THIC, is that the THIC-independent production of thiamine would overtake the production from THIC under laboratory conditions, keeping the predicted riboswitch activated and thus masking the possible effect of thiamine supplementation. Under this hypothesis, the predicted riboswitch would not be involved in controlling gene expression in response

to environmental levels of the vitamin in agreement with the experimental evidence gathered in this thesis. Instead, the riboswitch function would be limited to repress the expression of THIC in response to internal thiamine levels, to switch off the production of HMP from THIC when the alternative pathway can deliver sufficient thiamine. In this case, rather than a full loss-of-function, the predicted riboswitch would have not completely lost its expected function, providing sufficient selective advantage to explain the high sequence conservation among the diatom TPP aptamers. However, this hypothesis seems to be incompatible with the results of the mutated *THIC* overexpressor, which did not show increased thiamine levels compared to an unmutated control (Figure 4.4). Equivalent mutations in *B. subtilis THIC* and *C. reinhardtii THI4* riboswitches lead to significant increases in intracellular thiamine concentration (Schyns *et al.,* 2005; Moulin *et al.,* 2013 respectively). Although a deleterious mutation in the predicted TPP aptamer should in principle be dominant over the native alleles, the deregulated mutant in *C. reinhardtii* has mutations in the native gene. With the CRISPR/Cas9-mediated HDR technique, it is now possible to generate designer mutations in the native *PtTHIC* loci, and the deregulation mutagenesis experiment could be repeated targeting the native copy of *THIC* to validate the results of the experiment presented in this thesis. If the PtNMT1 knockout could be generated via CRISPR/Cas9 as mentioned above it would also offer the possibility to test whether *THIC* would then be downregulated by thiamine supplementation.

In the light of the evidence gathered here, it becomes apparent that bioinformatic approaches based on sequence conservation alone should not be solely relied on to annotate riboswitches. This is important because most annotated riboswitches have no experimental validation and are annotated based on sequence conservation and secondary structure prediction (McRose *et al.,* 2014; Moldovan *et al.,* 2018; Yadav *et al.,* 2015). It is important that the riboswitches are experimentally validated before conclusively assigning a functional annotation, especially in eukaryotic lineages outside of Chloroplastida and Opisthokonts (Figure 1.6), the only phyla where eukaryotic TPP riboswitch characterisations have been carried out to date. The results in these two eukaryotic groups should not be generalised to the rest of eukaryotic lineages, especially considering that all experimentally determined eukaryotic riboswitches have been described in organisms from land environments, whereas many relevant eukaryotic species are found in oceanic environments, with significantly differences in vitamin environmental concentrations (Gobler *et al.,* 2007). In the case of TPP riboswitch prediction in algal lineages, McRose *et al.* (2014) experimentally validated the function of the riboswitches found in two *Micromonas* species (Prasinophytes). Interestingly, *Micromonas* has no thiamine biosynthesis genes, and its riboswitches are located in the 3'UTRs (and sometimes in both UTRs) of thiamine transporter genes with no evidence of alternative splicing (Worden *et al.,* 2009). It would be interesting to test whether *Micromonas* is resistant to pyrithiamine and to investigate if the

molecular mechanism of its riboswitches could be based on alternative polyadenylation, as originally hypothesised for the putative diatom riboswitches (Section 3.2.4). And given that the *Micromonas* riboswitch mechanism of action does not seem to rely on alternative splicing, it would be interesting to try and use the it to regulate heterologous genes in *P. tricornutum*, something that could not be achieved with alternative-splicing based riboswitches from *C. reinhardtii* (Figure 4.8; Table 4.1).

Leaving evolutionary considerations aside, the results in this thesis show that thiamine supplementation does not influence the expression of any of the thiamine-related genes investigated. In fact, in all the experiments comparing thiamine replete with thiamine deplete conditions we could not observe any genetic or physiological changes despite the active uptake of the vitamin. It would be interesting to test whether the presence of thiamine has an effect on any *P. tricornutum* gene through a transcriptomic analysis, which could also serve to validate the results of this thesis. In parallel, it would also be important to test whether exogenous thiamine induces any changes at the physiological level, including photosynthesis and respiration rates, or the metabolism of hydrocarbons and lipids. Considering we could not observe any changes in gene expression upon thiamine supplementation, *P. tricornutum* might be a good chassis to incorporate orthologous thiamine-regulatory mechanisms from other organisms without interfering with the expression of native genes. For instance, the *S. cerevisiae* thiamine response mechanism, involving a sensor protein (Thi3p) and a transcription factor (Thi2p) (Nosaka *et al.,* 2005) could be imported to *P. tricornutum* to control heterologous genes. This system would meet two general synthetic biology principles: it would be orthogonal, having minimal interaction with native processes, and would confer a strong repression level, with reported fold-changes in expression of up to 1000 for some yeast thiamine-controlled promoters (Hohmann & Meacock, 1998).

Why and how is *PtTHIC* downregulated by cobalamin?

Despite not responding to thiamine supplementation, PtTHIC has been shown to be downregulated by cobalamin supplementation in previous proteomic experiments (Bertrand *et al.,* 2012). The regulatory link between *THIC* and cobalamin is not totally surprising as it is known that different B vitamin metabolisms are interrelated and often the biosynthesis of one B vitamin requires the use of another as a cofactor (Monteverde *et al.,* 2007). In fact, the results in this thesis are consistent with a previous transcriptomics and proteomics analysis (Bertrand *et al.,* 2012), in which the authors discussed that the downregulation of *THIC* in response to cobalamin might be linked to the use of SAM as a substrate of THIC. Interestingly, SAM levels in *C. reinhardtii*, which also has both METE and METH, show no observable differences between cultures grown with or without 1 $\mu g \cdot l^{-1}$ cobalamin for 30

hours (Bunbury *et al.,* 2020). This might explain why THIC is not downregulated by cobalamin in this species (Dr Andrew Sayer*,* personal communication). In contrast, SAM levels are significantly increased under replete (200 pM) compared to control (1 pM) cobalamin concentrations in *T. pseudonana,* a diatom with only METH (Heal *et al.,* 2017). It would be interesting to investigate whether *P. tricornutum* SAM levels in replete and deplete cobalamin conditions resemble more the *C. reinhardtii* or the *T. pseudonana* response to cobalamin supplementation. We could also investigate SAM concentrations as well as *PtTHIC* transcript levels under replete and deplete cobalamin conditions in the *P. tricornutum METE* mutant generated and presented in Chapter 5. The results of these experiments would test whether the reasoning for *PtTHIC* downregulation in response to cobalamin proposed by Bertrand *et al.* (2012) reflects measurable changes in SAM levels and would provide more insight into the physiological effects of *PtTHIC* upregulation under cobalamin depletion.

An alternative and more speculative hypothesis is that cobalamin regulation can be explained by assigning a DMB (cobalamin lower axial ligand) synthesis function to THIC. It has been shown that THIC is homologous to BzaF, the first enzyme in the DMB biosynthetic pathway in some anaerobic bacteria. Both THIC and BzaF contain a Fe-S cluster, share a common substrate (AIR) and a similar mechanism of action involving SAM as a substrate (Mehta *et al.,* 2015). Additionally, it has been shown that genes involved in the anaerobic biosynthesis of DMB are downregulated by cobalamin riboswitches in bacteria (Hazra *et al.,* 2015). One could then speculate that THIC is actually involved in DMB biosynthesis, which would be consistent with the lack of regulation of *THIC* by thiamine supplementation and the growth of the THIC knock-out in the absence of thiamine. Recent genomic evidence shows that the *P. tricornutum* genome contains the necessary genes to remodel corrinoids into cobalamin with DMB as lower axial ligand (CobC, CobS, CobT; *Phatr3_J16445*) (Vancaester *et al.,* 2020). *P. lutheri* and a *C. reinhardtii METE* mutants have shown they can remodel pseudocobalamin into cobalamin with a given exogenous DMB source and other cobalamin-auxotrophic algal species have shown they can grow on pseudocobalamin without DMB supplementation (Helliwell *et al.,* 2016). Having the ability to remodel corrinoids only seems to be positively selectable in an environment with sufficient levels of DMB or in the case of DMB prototrophy. DMB concentrations in seawater collected in the western English Channel are too low to allow corrinoid remodelling in algae (Helliwell *et al.,* 2016), so the hypothesis that *P. tricornutum* might be able to synthesise DMB is not entirely discardable. Nevertheless, three reasons highly discourage this hypothesis: to the best of my knowledge there are no known examples of DMB biosynthesis in eukaryotes, in principle BzaA and BzaF act under anaerobic conditions (Hazra *et al.,* 2015), and PtTHIC conserves the key active site residues in AtTHIC, and not those of the BzaF homolog, identified by Mehta *et al.* (2015). A

complementation assay with PtTHIC in a *E. coli* THIC mutant replicating the experiment with AtTHIC (Kong *et al.,* 2008) would elucidate whether PtTHIC can act as an HMP-P synthase.

Whether *P. tricornutum* is able to assemble DMB into a corrinoid ring to form cobalamin, and whether it can refactor corrinoids into cobalamin without exogenous DMB, could be experimentally tested with growth assays with the PtMETE knock-out. If growth of the METE knock-out could be rescued by pseudocobalamin or the corrinoid moiety alone, it would demonstrate that *P. tricornutum* can synthesise DMB *de novo* and utilise it to synthesise cobalamin when provided with an exogenous corrinoid ring. This ability would constitute a trait positively selected in competitive oceanic environments with two distinct pools of cobalamin and pseudocobalamin (Heal *et al.,* 2017) and would have important implications for the ecological cycling of corrinoids (Helliwell *et al.,* 2016). A bioassay based on a *S. meliloti bluB* mutant or liquid chromatography – mass spectrometry (LC-MS) methods (Crofts *et al.*, 2014) could be used to detect the presence of DMB in the *P. tricornutum* WT and THIC knock-out strains, and would ultimately validate whether THIC is responsible for DMB biosynthesis. These experiments would be of ecological interest given that the role of pseudocobalamin-refactoring species is only beginning to be understood to play an important role in the dynamics of microbial communities with cobalamin auxotrophs and low levels of the vitamin.

Turning to the mechanistic characterisation of THIC cobalamin downregulation, reporter constructs in this thesis showed that the *PtTHIC* promoter and 5'UTR regions were sufficient to confer cobalamin responsiveness (Figure 4.6). In addition, bioinformatic prediction of conserved motifs in the promoter regions of six diatom *THIC* genes revealed a conserved 14 bp palindromic sequence that might constitute the binding site of a transcription factor (Figure 3.8). The same conserved motif is also conserved in other cobalamin-downregulated genes in diatoms, including four copies of the motif in the highly expressed and tightly regulated *PtMETE* promoter (Table 3.2). Previous truncation experiments on the *PtMETE* promoter showed that a longer version of the promoter with four copies of the motif drives the expression of a reporter gene several times higher than a *PtMETE* truncation with only two motifs, and again several times higher than a *PtMETE* truncation with only one motif (Hickland *et al.,* in preparation). The additive effect that the motif has on expression level and the observation that mutating the motifs in the *PtMETE* promoter abolishes the expression of the reporter (Figure 4.14) lend support to the hypothesis that cobalamin regulation is mediated by a transcriptional activator binding the motifs and activating the transcription machinery in the absence of cobalamin. A mechanism with an analogous repressible activator logic has been extensively described for the PHO4 transcription activator in *S. cerevisiae* (Komeili & O'Shea, 2000). PHO4 activates the expression of phosphate starvation-related genes but it is inhibited by the presence of phosphate in a phosphorylation-mediated mechanism that exports it to the cytoplasm (Komeili & O'Shea, 1999).

Repressible activators, of which PHO4 is a clear example, seem to be especially suited to regulate responses to nutrient-deplete conditions. There are also instances of repressible activators in sulphur (Maruyama-Nakashita *et al.,* 2006; Tao & Marzluf, 1998), glucose (Carlson, 1999) and thiamine depletion (Maupin-Furlow, 2018). It follows, that a repressible activator might be used in diatoms to regulate the response to cobalamin depletion. The mechanisms by which the putative transcriptional activator is inactivated in response to cobalamin supplementation might involve the prevention of dimerization by post-translational modification (Pireyre & Burow, 2015), the relocation to the cytoplasm (Merkle, 2001), or the tagging for degradation (Desterro *et al.,* 2000). Some of these processes might be influenced by a sensor protein responding to the presence of cobalamin, or a C1 cycle metabolite like methionine or SAM, or even responding to changes in the SAM-to-SAH ratio, which indicates the methylation potential in the cell (Ye *et al.,* 2017).

PHO4 is a bHLH domain-containing transcription factor whose binding site core sequence is CACGTG (Shimizu *et al.,* 1997). The CACGTG motif is also known as G-box and is recognised by bHLH and basic Lucin Zipper (bZIP) DNA binding domains (Ezer *et al.,* 2017). Coincidentally, the cobalamin-downregulated motif in *P. tricornutum* has a highly conserved G-box at its core (Figure 3.8). In the *P. tricornutum* genome there are only 7 genes encoding a bHLH domain and 23 genes encoding a bZIP domain (Rayko *et al.*, 2010), of which 3 are involved in circadian rhythm regulation (Annunziata *et al.,* 2019), 3 in light acclimation and cell cycle progression (Costa *et al.,* 2013; Huysman *et al.,* 2013), one in tricarboxylic acid cycle regulation (Matthijs *et al.,* 2017), and one in CO_2 response regulation (Ohno *et al.,* 2012). Given the distribution of the conserved motif in cobalamin-downregulated promoters (Table 3.2), it seems unlikely that one of the already characterised transcription factors binds the motif, and instead I would propose that one of the remaining 21 uncharacterised bHLH and bZIP transcription factors would be responsible for cobalamin-responsive gene regulation. There is evidence that the genomic context flanking the G-Box provides specificity for transcription factors to specifically bind their target sequence (Ezer *et al.,* 2017) and both sequence and DNA topology seem to be deterministic for binding specificity (Gordân *et al.,* 2013). The conservation of flanking GAAG-CTTC sequences around the G-Box in the conserved motif in diatom cobalamin-downregulated promoters (Figure 3.8) would then indicate that it is bound by a specific transcription factor rather than by any transcription factor with bHLH or bZIP domains.

From an evolutionary perspective, it is interesting to note that the promoter region of the human Transcobalamin II gene contains a G-Box bound by USF1/USF2, bHLH-containing transcription factors (Li & Seetharam, 1998). In *S. cerevisiae,* the *METE* homologue (*MET6*) is regulated by CBF1-MET4-MET28, a bHLH-bZIP heteromeric transcription factor involved in the regulation of sulphurated amino acids in response to sulphur deprivation (Kuras *et al.,* 1997; Thomas & Surdin-Kerjan, 1997). *ScMET6*

is mildly downregulated by methionine and SAM supplementation and has two CBF1-MET4-MET28 binding sites (TCACGTG) which are very similar to the predicted motif in the cobalamin-downregulated genes in diatoms (Figure 3.8; Thomas & Surdin-Kerjan, 1997). This observation invites speculation that a hypothetical cobalamin regulation by a bZIP/bHLH transcription factor might have an evolutionary relationship with the sulphur starvation regulatory mechanism and might respond indirectly to cobalamin supplementation by sensing an increase in methionine or SAM levels.

Y1H or bacterial 1 hybrid (B1H) experiments could be performed to elucidate which proteins in a library of all proteins, or a library of the 21 uncharacterised bHLH and bZIP transcription factors, bind the conserved motif. Libraries of all *P. tricornutum* proteins have already been developed and used in Y1H to identify other transcription factors and could be reused to investigate the cobalamin-related motif (Huysman *et al.,* 2013; Matthijs *et al.,* 2016; Matthijs *et al.,* 2017). The challenge with the Y1H technique is that bHLH and bZIP transcription factors bind to their binding sites as dimers, so if the transcription factor recognising the cobalamin motif binds as an heterodimer the Y1H assay would not be able to identify the candidate, as only one peptide is tested at a time. Instead, a DNA pull-down assay on a *P. tricornutum* protein extract followed by SDS-PAGE and MS identification of the purified protein(s) could be used (Tran *et al.,* 2015). Unlike Y1H, DNA pull-down methods are not limited to testing transcription factors in isolation and can identify more than one protein binding cooperatively to the DNA probe. For the DNA-pull down assay, a negative control in the presence of cobalamin could differentiate between constitutive binding proteins, and proteins only associated with the promoter under cobalamin-deplete conditions. If a transcription factor is identified through the Y1H or pull-down assay, an EMSA could be performed to validate the specific binding and to investigate in more detail the precise sequence requirements for the protein-DNA interaction (Ohno *et al.,* 2012). With only 21 candidates, it could even be conceivable to express all the transcription factors in *E. coli* and perform EMSA assays for each of them to identify the transcription factor(s) selectively binding the conserved motif in cobalamin-downregulated promoters (Ohno *et al.,* 2012). If a transcription factor is identified, its function could be further studied with overexpressing and knock-out strains to investigate the phenotypic consequences (Slattery *et al.,* 2014), and with Chromatin Immunoprecipitation-Sequencing (ChIP-Seq) analysis to identify all genomic locations interacting with the transcription factor to identify which genes are regulated by cobalamin regulation (Park, 2009).

Besides the identification of the transcription factor, identifying the minimal DNA sequence to confer cobalamin-dependent expression would be interesting to understand the regulatory process and to use the cobalamin-repressible module in synthetic biology constructs. From the reporter constructs used in this thesis, it becomes apparent that the conserved 14 bp motif is necessary for expression, but not sufficient for regulation (Figure 4.14). A truncated *PtMETE* promoter including just 242 bp

upstream of the ATG and 149 bp upstream of the transcription start site, and only including the most proximal of the four motifs, was shown to regulate the expression of a reporter in response to cobalamin (Hickland *et al.,* in preparation). Within the 242 bp truncation there needs to be a minimal regulatory sequence including the 14 bp motif which confers cobalamin responsiveness to the *PtMETE* promoter. Given that *PtMETE* is a repressible promoter, a directed evolution assay on *P. tricornutum* transformed with a zeocin resistance marker expressed from the 242 bp truncation of the *PtMETE* promoter under combined zeocin and cobalamin supplementation could be contemplated. Only those cells with spontaneous mutations in the promoter that prevented cobalamin downregulation but allowed gene expression would survive. Sequencing the mutants that survived the selection would indicate which regions are specifically important for cobalamin-downregulation. This method would fail if cobalamin regulation and expression depend on the same or overlapping motifs, which is what a repressible activator model would suggest. It would also not give the expected result if the cells mutate the transcription factor rather than the DNA binding site to release the regulation, although this is less favourable for the cell as all the other genes regulated by the transcription factor would be affected. Other reporter construct approaches including progressive truncations, and targeted mutagenesis of the promoter could also help elucidate the minimal regulatory sequence. And once the transcription factor is identified, EMSA assays could also be used to test different mutated sequences to understand which sequences are necessary for binding.

Concluding remarks

The unique biochemical properties of B vitamins make them essential cofactors for a wide variety of metabolic reactions and have ensured their continued use since the RNA World in all forms of life. It is generally accepted that the earliest aptamer-based genetic regulatory mechanisms arose to regulate the metabolism of B vitamins in the RNA World and are still conserved nowadays across the tree of life. Despite this, the results in this thesis contradict most of our initial preconceptions on riboswitches and thiamine biosynthesis, partly due to the extrapolation of the knowledge of the characterised examples of these systems in select eukaryotic lineages to other eukaryotic groups. In particular, the conundrum posed by the lack of thiamine response of the diatom TPP riboswitches despite their strong sequence conservation and the lack of knowledge on the function of the NMT1 homologue in several algal lineages show that there are important gaps in our understanding of the evolution, regulation and environmental relevance of thiamine metabolism in the marine environment. Addressing the questions opened by the results of this thesis requires further experimental focus on the largely unstudied B vitamin metabolism of several algal lineages and will further illuminate the importance of B vitamins in the population dynamics and nutrient cycling of

marine ecosystems. Finally, the investigation of the cobalamin regulation of *THIC* highlights the interconnection of B vitamin metabolisms and paves the way for the elucidation of the molecular mechanism underpinning the diatom cobalamin-depletion response, and the CRISPR/Cas9 method developed here will accelerate further physiological research and metabolic engineering in diatoms.

7. References

Adler-Agnon, Z., Leu, S., Zarka, A., Boussiba, S. & Khozin-Goldberg, I. (2018). Novel promoters for constitutive and inducible expression of transgenes in the diatom *Phaeodactylum tricornutum* under varied nitrate availability. *Journal of Applied Phycology*, 30(5), 2763-2772.

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. *Nature*, 473(7346), 203-207.

Angstenberger, M., Krischer, J., Aktaş, O. & Büchel, C. (2018). Knock-down of a ligIV homologue enables DNA integration via homologous recombination in the marine diatom *Phaeodactylum tricornutum*. *ACS synthetic biology*, 8(1), 57-69.

Annunziata, R., Ritter, A., Fortunato, A. E., Manzotti, A., Cheminant-Navarro, S., Agier, N., Huysman, M.J., Winge, P., Bones, A.M., Bouget, F.Y. & Lagomarsino, M. C. (2019). bHLH-PAS protein RITMO1 regulates diel biological rhythms in the marine diatom *Phaeodactylum tricornutum*. *Proceedings of the National Academy of Sciences*, 116(26), 13137-13142.

Anthony, P. C., Perez, C. F., García-García, C. & Block, S. M. (2012). Folding energy landscape of the thiamine pyrophosphate riboswitch aptamer. *Proceedings of the National Academy of Sciences*, 109(5), 1485-1489.

Apt, K. E., Grossman, A. R. & Kroth-Pancic, P. G. (1996). Stable nuclear transformation of the diatom *Phaeodactylum tricornutum*. *Molecular and General Genetics*, 252(5), 572-579.

Apt, K. E., Zaslavkaia, L., Lippmeier, J. C., Lang, M., Kilian, O., Wetherbee, R., Grossman, A.R. & Kroth, P. G. (2002). In vivo characterization of diatom multipartite plastid targeting signals. *Journal of Cell Science*, 115(21), 4061-4069.

Archibald, J. M. (2012). The Evolution of Algae by Secondary and Tertiary Endosymbiosis. In Gwenaël Piganeau (Ed.), *Genomic Insights into the Biology of Algae*, 87–118. Academic Press.

Armbrust, E. V. (2009). The life of diatoms in the world's oceans. *Nature*, 459(7244), 185-192.

Athanasakoglou, A. & Kampranis, S. C. (2019). Diatom isoprenoids: Advances and biotechnological potential. *Biotechnology advances*, 37(8), 107417.

Bae, S., Park, J. & Kim, J. S. (2014). Cas-OFFinder: a fast and versatile algorithm that searches for potential off-target sites of Cas9 RNA-guided endonucleases. *Bioinformatics*, 30(10), 1473-1475.

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

Bailey, T. L., Boden, M., Buske, F. A., Frith, M., Grant, C. E., Clementi, L., Ren, J., Li, W.W. & Noble, W. S. (2009). MEME SUITE: tools for motif discovery and searching. *Nucleic acids research*, 37(suppl_2), W202-W208.

Bajor, J., Tkaczuk, K. L., Chruszcz, M., Chapman, H., Kagan, O., Savchenko, A. & Minor, W. (2014). The crystal structure of pyrimidine/thiamin biosynthesis precursor-like domain-containing protein CAE31940 from proteobacterium *Bordetella bronchiseptica RB50*, and evolutionary insight into the NMT1/THI5 family. *Journal of structural and functional genomics*, 15(2), 73-81.

Bale, S., Rajashankar, K. R., Perry, K., Begley, T. P. & Ealick, S. E. (2010). HMP binding protein ThiY and HMP-P synthase THI5 are structural homologues. *Biochemistry*, 49(41), 8929-8936.

Banerjee, R., Frasca, V., Ballou, D. P., & Matthews, R. G. (1990). Participation of cob (I) alamin in the reaction catalyzed by methionine synthase from Escherichia coli: a steady-state and rapid reaction kinetic analysis. *Biochemistry*, *29*(50), 11101-11109.

Banerjee, R. & Ragsdale, S. W. (2003). The many faces of vitamin B12: catalysis by cobalamin-dependent Enzymes1. *Annual review of biochemistry*, 72.

Basu, S., Patil, S., Mapleson, D., Russo, M. T., Vitale, L., Fevola, C., Maumus, F., Casotti, R., Mock, T., Caccamo, M. & Montresor, M. (2017). Finding a partner in the ocean: molecular and evolutionary bases of the response to sexual cues in a planktonic diatom. *New Phytologist*, 215(1), 140-156.

Belshaw, N., Grouneva, I., Aram, L., Gal, A., Hopes, A. & Mock, T. (2017). Efficient CRISPR/Cas-mediated homologous recombination in the model diatom *Thalassiosira pseudonana*. *BioRxiv*, 215582 - PREPRINT.

Bender, D. A. (2003). *Nutritional biochemistry of the vitamins.* Cambridge University Press.

Bendtsen, J. D., Nielsen, H., von Heijne, G. & Brunak, S. (2004). Improved prediction of signal peptides: SignalP 3.0. *Journal of molecular biology*, 340(4), 783-795.

Bertrand, E. M. & Allen, A. E. (2012). Influence of vitamin B auxotrophy on nitrogen metabolism in eukaryotic phytoplankton. *Frontiers in microbiology*, 3, 375.

Bertrand, E. M., Allen, A. E., Dupont, C. L., Norden-Krichmar, T. M., Bai, J., Valas, R. E. & Saito, M. A. (2012). Influence of cobalamin scarcity on diatom molecular physiology and identification of a

cobalamin acquisition protein. *Proceedings of the National Academy of Sciences*, 109(26), E1762-E1771.

Blount, K. F. & Breaker, R. R. (2006). Riboswitches as antibacterial drug targets. *Nature biotechnology*, 24(12), 1558-1564.

Bocobza, S. E., Malitsky, S., Araújo, W. L., Nunes-Nesi, A., Meir, S., Shapira, M., Fernie, A.R. & Aharoni, A. (2013). Orchestration of thiamin biosynthesis and central metabolism by combined action of the thiamin pyrophosphate riboswitch and the circadian clock in *Arabidopsis*. *The Plant Cell*, 25(1), 288-307.

Bocobza, S., Adato, A., Mandel, T., Shapira, M., Nudler, E. & Aharoni, A. (2007). Riboswitch-dependent gene regulation and its evolution in the plant kingdom. *Genes & development*, 21(22), 2874-2879.

Borths, E. L., Locher, K. P., Lee, A. T. & Rees, D. C. (2002). The structure of *Escherichia coli* BtuF and binding to its cognate ATP binding cassette transporter. *Proceedings of the National Academy of Sciences*, 99(26), 16642-16647.

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., 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., Oudot–Le Secq, M., Napoli, C., Obornik, M., Parker, M.S., Petit, J., 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(7219), 239.

Breaker, R. R. (2011). Prospects for riboswitch discovery and analysis. *Molecular cell*, 43(6), 867-879.

Breaker, R. R. (2012). Riboswitches and the RNA world. *Cold Spring Harbor perspectives in biology*, 4(2), a003566.

Brinkman, E. K., Chen, T., Amendola, M. & van Steensel, B. (2014). Easy quantitative assessment of genome editing by sequence trace decomposition. *Nucleic acids research*, 42(22), e168-e168.

Brumwell, S. L., MacLeod, M. R., Huang, T., Cochrane, R. R., Meaney, R. S., Zamani, M., Matysiakiewicz, O., Dan, K.N., Janakirama, P., Edgell, D.R. & Charles, T. C. (2019). Designer *Sinorhizobium meliloti* strains and multi-functional vectors enable direct inter-kingdom DNA transfer. *PloS one*, 14(6).

Brzezinski, M. A., Olson, R. J. & Chisholm, S. W. (1990). Silicon availability and cell-cycle progression in marine diatoms. *Marine ecology progress series*, 83-96.

Buck, J. M., Bártulos, C. R., Gruber, A. & Kroth, P. G. (2018). Blasticidin-S deaminase, a new selection marker for genetic transformation of the diatom *Phaeodactylum tricornutum*. *PeerJ*, 6, e5884.

Bunbury, F., Helliwell, K. E., Mehrshahi, P., Davey, M. P., Salmon, D. L., Holzer, A., Smirnoff, N. & Smith, A. G. (2020). Responses of a Newly Evolved Auxotroph of *Chlamydomonas* to B12 Deprivation. *Plant Physiology*, 183(1), 167-178.

Burki, F., Roger, A. J., Brown, M. W. & Simpson, A. G. (2020). The new tree of eukaryotes. *Trends in ecology & evolution*, 35(1), 43-55.

Carlson, M. (1999). Glucose repression in yeast. *Current opinion in microbiology*, 2(2), 202-207.

Chatterjee, A., Abeydeera, N. D., Bale, S., Pai, P. J., Dorrestein, P. C., Russell, D. H., Ealick, S.E. & Begley, T. P. (2011). *Saccharomyces cerevisiae* THI4p is a suicide thiamine thiazole synthase. *Nature*, 478(7370), 542-546.

Cheah, M. T., Wachter, A., Sudarsan, N. & Breaker, R. R. (2007). Control of alternative RNA splicing and gene expression by eukaryotic riboswitches. *Nature*, 447(7143), 497.

Chimento, D. P., Mohanty, A. K., Kadner, R. J. & Wiener, M. C. (2003). Substrate-induced transmembrane signaling in the cobalamin transporter BtuB. *Nature Structural & Molecular Biology*, 10(5), 394-401.

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

Coale, T. H., Moosburner, M., Horák, A., Oborník, M., Barbeau, K. A. & Allen, A. E. (2019). Reduction-dependent siderophore assimilation in a model pennate diatom. *Proceedings of the National Academy of Sciences*, 116(47), 23609-23617.

Cooper, M. B., Kazamia, E., Helliwell, K. E., Kudahl, U. J., Sayer, A., Wheeler, G. L. & Smith, A. G. (2019). Cross-exchange of B-vitamins underpins a mutualistic interaction between *Ostreococcus tauri* and *Dinoroseobacter shibae*. *The ISME journal*, 13(2), 334-345.

Coquille, S., Roux, C., Fitzpatrick, T. B. & Thore, S. (2012). The Last Piece in the Vitamin B1 Biosynthesis Puzzle structural and functional insight into yeast 4-amino-5-hydroxymethyl-2-methylpyrimidine phosphate (HMP-P) synthase. *Journal of Biological Chemistry*, 287(50), 42333-42343.

Corvelo, A., Hallegger, M., Smith, C. W. & Eyras, E. (2010). Genome-wide association between branch point properties and alternative splicing. *PLoS computational biology*, 6(11), e1001016.

Costa, B. S., 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(9), e74451.

Cox, J. S., Moncja, K., Mckinnes, M. & Van Dyke, M. W. (2019). Identification and Characterization of Preferred DNA-Binding Sites for the *Thermus thermophilus HB8* Transcriptional Regulator TTHA0973. *International journal of molecular Sciences*, 20(13), 3336.

Croft, M. T., Lawrence, A. D., Raux-Deery, E., Warren, M. J. & Smith, A. G. (2005). Algae acquire vitamin B12 through a symbiotic relationship with bacteria. *Nature*, 438(7064), 90.

Croft, M. T., Moulin, M., Webb, M. E. & Smith, A. G. (2007). Thiamine biosynthesis in algae is regulated by riboswitches. *Proceedings of the National Academy of Sciences*, 104(52), 20770-20775.

Croft, M. T., Warren, M. J. & Smith, A. G. (2006). Algae need their vitamins. *Eukaryotic cell*, 5(8), 1175-1183.

Crofts, T. S., Men, Y., Alvarez-Cohen, L. & Taga, M. E. (2014). A bioassay for the detection of benzimidazoles reveals their presence in a range of environmental samples. *Frontiers in microbiology*, 5, 592.

Crooks, G. E., Hon, G., Chandonia, J. M. & Brenner, S. E. (2004). WebLogo: a sequence logo generator. *Genome research*, 14(6), 1188-1190.

Crozet, P., Navarro, F. J., Willmund, F., Mehrshahi, P., Bakowski, K., Lauersen, K. J., Pérez-Pérez, M., Auroy, P., Gorchs Rovira, A., Sauret-Gueto, S., Niemeyer, J., Spaniol, B., Theis, J., Trösch, R., Westrich, L., Vavitsas, K., Baier, T., Hübner, W., de Carpentier, F., Cassarini, M., Danon, A., Henri, J., Marchand, C.H., de Mia, M., Sarkissian, K., Baulcombe, D.C., Peltier, G., Crespo, J., Kruse, O., Jensen, P., Schroda, M., Smith, A.G. & Lemaire, S.D. (2018). Birth of a photosynthetic chassis: A MoClo toolkit enabling synthetic biology in the microalga *Chlamydomonas reinhardtii*. *ACS synthetic biology*, 7(9), 2074-2086.

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. *Nature communications*, 5(1), 1-7.

D'Adamo, S., Schiano di Visconte, G., Lowe, G., Szaub-Newton, J., Beacham, T., Landels, A., Allen, M.J., Spicer, A. & Matthijs, M. (2019). Engineering the unicellular alga *Phaeodactylum tricornutum* for high-value plant triterpenoid production. *Plant biotechnology journal*, 17(1), 75-87.

De Riso, V., Raniello, R., Maumus, F., Rogato, A., Bowler, C. & Falciatore, A. (2009). Gene silencing in the marine diatom *Phaeodactylum tricornutum*. *Nucleic acids research*, 37(14), e96-e96.

Denslow, S. A., Rueschhoff, E. E. & Daub, M. E. (2007). Regulation of the *Arabidopsis thaliana* vitamin B6 biosynthesis genes by abiotic stress. *Plant Physiology and Biochemistry*, 45(2), 152-161.

Desterro, J. M. P., Rodriguez, M. S. & Hay, R. T. (2000). Regulation of transcription factors by protein degradation. *Cellular and Molecular Life Sciences*, 57(8-9), 1207-1219.

Devedjiev, Y., Surendranath, Y., Derewenda, U., Gabrys, A., Cooper, D. R., Zhang, R. G., Lezondra, L., Joachimiak, A. & Derewenda, Z. S. (2004). The structure and ligand binding properties of the *B. subtilis* YkoF gene product, a member of a novel family of thiamin/HMP-binding proteins. *Journal of molecular biology*, 343(2), 395-406.

Diner, R. E., Bielinski, V. A., Dupont, C. L., Allen, A. E. & Weyman, P. D. (2016). Refinement of the diatom episome maintenance sequence and improvement of conjugation-based DNA delivery methods. *Frontiers in Bioengineering and Biotechnology*, 4, 65.

Diner, R. E., Noddings, C. M., Lian, N. C., Kang, A. K., McQuaid, J. B., Jablanovic, J., Espinoza, J.L., Nguyen, N.A., Anzelmatti Jr., M.A., Jansson, J., Bielinski, V.A., Karas, B.J., Dupont, C.L., Allen, A.E. & Weyman, P.D. (2017). Diatom centromeres suggest a mechanism for nuclear DNA acquisition. *Proceedings of the National Academy of Sciences*, 114(29), E6015-E6024.

Doench, J. G., Fusi, N., Sullender, M., Hegde, M., Vaimberg, E. W., Donovan, K. F., Smith, I., Tothova, Z., Wilen, C., Orchard, R., Virgin, H.W., Listgarten, J. & Root, D.E. (2016). Optimized sgRNA design to maximize activity and minimize off-target effects of CRISPR-Cas9. *Nature biotechnology*, 34(2), 184-191.

Donovan, P. D., Holland, L. M., Lombardi, L., Coughlan, A. Y., Higgins, D. G., Wolfe, K. H. & Butler, G. (2018). TPP riboswitch-dependent regulation of an ancient thiamin transporter in *Candida*. *PLoS genetics*, 14(5), e1007429.

Doudna, J. A. & Charpentier, E. (2014). The new frontier of genome engineering with CRISPR-Cas9. *Science*, 346(6213).

Dunham, M. J. (2007). Synthetic ecology: a model system for cooperation. *Proceedings of the National Academy of Sciences*, 104(6), 1741-1742.

Duraisingh, M. T., Triglia, T. & Cowman, A. F. (2002). Negative selection of *Plasmodium falciparum* reveals targeted gene deletion by double crossover recombination. *International journal for parasitology*, 32(1), 81-89.

Edgar, R. C. (2004). MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic acids research*, 32(5), 1792-1797.

Edwards, T. E. & Ferré-D'Amaré, A. R. (2006). Crystal structures of the thi-box riboswitch bound to thiamine pyrophosphate analogs reveal adaptive RNA-small molecule recognition. *Structure*, 14(9), 1459-1468.

Engler, C., Youles, M., Gruetzner, R., Ehnert, T. M., Werner, S., Jones, J. D., Patron, N.J. & Marillonnet, S. (2014). A golden gate modular cloning toolbox for plants. *ACS synthetic biology*, 3(11), 839-843.

Enjo, F., Nosaka, K., Ogata, M., Iwashima, A. & Nishimura, H. (1997). Isolation and characterization of a thiamin transport gene, THI10, from *Saccharomyces cerevisiae*. *Journal of Biological Chemistry*, 272(31), 19165-19170.

Erdene-Ochir, E., Shin, B. K., Huda, M. N., Kim, D. H., Lee, E. H., Song, D. G., Kim, Y.M., Kim, S.M. & Pan, C. H. (2016). Cloning of a novel endogenous promoter for foreign gene expression in *Phaeodactylum tricornutum*. *Applied Biological Chemistry*, 59(6), 861-867.

Erdene-Ochir, E., Shin, B. K., Kwon, B., Jung, C. & Pan, C. H. (2019). Identification and characterisation of the novel endogenous promoter HASP1 and its signal peptide from *Phaeodactylum tricornutum*. *Scientific reports*, 9(1), 1-10.

Fabris, M., George, J., Kuzhiumparambil, U., Lawson, C. A., Jaramillo-Madrid, A. C., Abbriano, R. M., Vickers, C.E. & Ralph, P. (2020). Extrachromosomal genetic engineering of the marine diatom *Phaeodactylum tricornutum* enables the heterologous production of monoterpenoids. *ACS synthetic biology*, 9(3), 598-612.

Fabris, M., Matthijs, M., Rombauts, S., Vyverman, W., Goossens, A. & Baart, G. J. (2012). The metabolic blueprint of *Phaeodactylum tricornutum* reveals a eukaryotic Entner–Doudoroff glycolytic pathway. *The Plant Journal*, 70(6), 1004-1014.

Falciatore, A., Casotti, R., Leblanc, C., Abrescia, C. & Bowler, C. (1999). Transformation of nonselectable reporter genes in marine diatoms. *Marine Biotechnology*, 1(3), 239-251.

Ferenczi, A., Pyott, D. E., Xipnitou, A. & Molnar, A. (2017). Efficient targeted DNA editing and replacement in *Chlamydomonas reinhardtii* using Cpf1 ribonucleoproteins and single-stranded DNA. *Proceedings of the National Academy of Sciences*, 114(51), 13567-13572.

Field, C. B., Behrenfeld, M. J., Randerson, J. T. & Falkowski, P. (1998). Primary production of the biosphere: integrating terrestrial and oceanic components. *Science*, 281(5374), 237-240.

Fornes, O., Castro-Mondragon, J. A., Khan, A., Van der Lee, R., Zhang, X., Richmond, P. A., Modi, B.P., Correard, S., Gheorghe, M., Baranašić, D., Santana-Garcia, W., Tan, G., Chèneby, J., Ballester, B., Parcy, F., Sandelin, A., Lenhard, B., Wasserman, W.W. & Mathelier, A. (2020). JASPAR (2020): update of the open-access database of transcription factor binding profiles. *Nucleic acids research*, 48(D1), D87-D92.

Galachyants, Y. P., Zakharova, Y. R., Petrova, D. P., Morozov, A. A., Sidorov, I. A., Marchenkov, A. M., Logacheva, M.D., Markelov, M.L., Khabudaev, K.V, Likhoshway, Y.V. & Grachev, M.A. (2015). Sequencing of the complete genome of an araphid pennate diatom *Synedra acus subsp. radians* from Lake Baikal. *Doklady Biochemistry and Biophysics*, 461(1), 84-88.

Ganapathy, V., Smith, S. B. & Prasad, P. D. (2004). SLC19: the folate/thiamine transporter family. *Pflügers Archiv*, 447(5), 641-646.

George, J., Kahlke, T., Abbriano, R. M., Kuzhiumparambil, U., Ralph, P. J. & Fabris, M. (2020). Metabolic engineering strategies in diatoms reveal unique phenotypes and genetic configurations with implications for algal genetics and synthetic biology. *Frontiers in Bioengineering and Biotechnology*, 8, 513.

Gherasim, C., Lofgren, M. & Banerjee, R. (2013). Navigating the B12 road: assimilation, delivery, and disorders of cobalamin. *Journal of Biological Chemistry*, 288(19), 13186-13193.

Giovannoni, S. J. (2012). Vitamins in the sea. *Proceedings of the National Academy of Sciences*, 109(35), 13888-13889.

Gobler, C. J., Norman, C., Panzeca, C., Taylor, G. T. & Sañudo-Wilhelmy, S. A. (2007). Effect of B-vitamins (B1, B12) and inorganic nutrients on algal bloom dynamics in a coastal ecosystem. *Aquatic Microbial Ecology*, 49(2), 181-194.

Godoy-Parejo, C., Deng, C., Zhang, Y., Liu, W., & Chen, G. (2020). Roles of vitamins in stem cells. *Cellular and Molecular Life Sciences*, *77*(9), 1771-1791.

Gong, S., Wang, Y., Wang, Z. & Zhang, W. (2017). Co-transcriptional folding and regulation mechanisms of riboswitches. *Molecules*, 22(7), 1169.

Gordân, R., Shen, N., Dror, I., Zhou, T., Horton, J., Rohs, R. & Bulyk, M. L. (2013). Genomic regions flanking E-box binding sites influence DNA binding specificity of bHLH transcription factors through DNA shape. *Cell reports*, 3(4), 1093-1104.

Gorman, D. S. & Levine, R. P. (1965). Cytochrome f and plastocyanin: their sequence in the photosynthetic electron transport chain of *Chlamydomonas reinhardtii*. *Proceedings of the National Academy of Sciences*, 54(6), 1665-1669.

Gruber, A., Rocap, G., Kroth, P. G., Armbrust, E. V. & Mock, T. (2015). Plastid proteome prediction for diatoms and other algae with secondary plastids of the red lineage. *The Plant Journal*, 81(3), 519-528.

Guillard, R. R. (1975). Culture of phytoplankton for feeding marine invertebrates. In Smith, W.L. & Chanley, M.H. (Eds.), *Culture of marine invertebrate animals*, 29-60. Plenum Press.

Guo, M. & Chen, Y. (2018). Coenzyme cobalamin: biosynthesis, overproduction and its application in dehalogenation—a review. *Reviews in Environmental Science and Bio/Technology*, 17(2), 259-284.

Gutowska, M. A., Shome, B., Sudek, S., McRose, D. L., Hamilton, M., Giovannoni, S. J., Begley, T. P. & Worden, A. Z. (2017). Globally important haptophyte algae use exogenous pyrimidine compounds more efficiently than thiamin. *MBio*, 8(5), e01459-17.

Haller, A., Altman, R. B., Soulière, M. F., Blanchard, S. C. & Micura, R. (2013). Folding and ligand recognition of the TPP riboswitch aptamer at single-molecule resolution. *Proceedings of the National Academy of Sciences*, 110(11), 4188-4193.

Hanson, A. D., Amthor, J. S., Sun, J., Niehaus, T. D., Gregory III, J. F., Bruner, S. D. & Ding, Y. (2018). Redesigning thiamin synthesis: Prospects and potential payoffs. *Plant Science*, 273, 92-99.

Hazra, A. B., Han, A. W., Mehta, A. P., Mok, K. C., Osadchiy, V., Begley, T. P. & Taga, M. E. (2015). Anaerobic biosynthesis of the lower ligand of vitamin B12. *Proceedings of the National Academy of Sciences*, 112(34), 10792-10797.

Heal, K. R., Qin, W., Ribalet, F., Bertagnolli, A. D., Coyote-Maestas, W., Hmelo, L. R., Moffett, J. W., Devol, A. H., Armbrust, E. V., Stahl, D. A. & Ingalls, A. E. (2017). Two distinct pools of B12 analogs reveal community interdependencies in the ocean. *Proceedings of the National Academy of Sciences*, 114(2), 364-369.

Helliwell, K. E., Chrachri, A., Koester, J. A., Wharam, S., Verret, F., Taylor, A. R., Wheeler, G. L. & Brownlee, C. (2019). Alternative mechanisms for fast Na+/Ca2+ signaling in eukaryotes via a novel class of single-domain voltage-gated channels. *Current Biology*, 29(9), 1503-1511.

Helliwell, K. E., Lawrence, A. D., Holzer, A., Kudahl, U. J., Sasso, S., Kräutler, B., Scanlan, D. J., Warren, M. J. & Smith, A. G. (2016). Cyanobacteria and eukaryotic algae use different chemical variants of vitamin B12. *Current Biology*, 26(8), 999-1008.

Helliwell, K. E., Scaife, M. A., Sasso, S., Araujo, A. P. U., Purton, S. & Smith, A. G. (2014). Unraveling vitamin B12-responsive gene regulation in algae. *Plant Physiology*, 165(1), 388-397.

Hempel, F., Felsner, G. & Maier, U. G. (2010). New mechanistic insights into pre-protein transport across the second outermost plastid membrane of diatoms. *Molecular microbiology*, 76(3), 793-801.

Hofacker, I. L. (2003). Vienna RNA secondary structure server. *Nucleic acids research*, 31(13), 3429-3431.

Hohmann, S. & Meacock, P. A. (1998). Thiamin metabolism and thiamin diphosphate-dependent enzymes in the yeast *Saccharomyces cerevisiae*: genetic regulation. *Biochimica et Biophysica Acta (BBA)-Protein Structure and Molecular Enzymology*, 1385(2), 201-219.

Hopes, A., Nekrasov, V., Belshaw, N., Grouneva, I., Kamoun, S. & Mock, T. (2017). Genome Editing in Diatoms Using CRISPR-Cas to Induce Precise Bi-allelic Deletions. *Bio-protocol*, 7, 23.

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), 49.

Huang, W. & Daboussi, F. (2017). Genetic and metabolic engineering in diatoms. *Philosophical Transactions of the Royal Society B: Biological Science*s, 372(1728), (2016)0411.

Huysman, M. J., Fortunato, A. E., Matthijs, M., Costa, B. S., Vanderhaeghen, R., Van den Daele, H., Sachse, M., Inzé, D., Bowler, C., Kroth, P. G., Wilhelm, C., Falciatore, A., Vyverman, W. & De Veylder, L. (2013). AUREOCHROME1a-mediated induction of the diatom-specific cyclin dsCYC2 controls the onset of cell division in diatoms (*Phaeodactylum tricornutum*). *The Plant Cell*, 25(1), 215-228.

Hwang, S., Cordova, B., Abdo, M., Pfeiffer, F. & Maupin-Furlow, J. A. (2017). ThiN as a versatile domain of transcriptional repressors and catalytic enzymes of thiamine biosynthesis. *Journal of bacteriology*, 199(7), e00810-16.

Jaehme, M. & Slotboom, D. J. (2015). Diversity of membrane transport proteins for vitamins in bacteria and archaea. *Biochimica et Biophysica Acta (BBA)-General Subjects*, 1850(3), 565-576.

Jägerstad, M. & Arkbåge, K. (2003). Cobalamins| properties and determination. In Caballero, B., Trugo, L. & Finglas, P. (Eds.), *Encyclopedia of Food Sciences and Nutrition* (Second Edition), 1419-1427. Elsevier Science.

Jeanguenin, L., Lara-Núñez, A., Rodionov, D. A., Osterman, A. L., Komarova, N. Y., Rentsch, D., Gregory III, J. F. & Hanson, A. D. (2012). Comparative genomics and functional analysis of the NiaP family uncover nicotinate transporters from bacteria, plants, and mammals. *Functional & integrative genomics*, 12(1), 25-34.

Ji, G., Li, L., Li, Q. Q., Wu, X., Fu, J., Chen, G. & Wu, X. (2015). PASPA: a web server for mRNA poly (A) site predictions in plants and algae. *Bioinformatics*, 31(10), 1671-1673.

Jones, D. T., Taylor, W. R. & Thornton, J. M. (1992). The rapid generation of mutation data matrices from protein sequences. *Bioinformatics*, 8(3), 275-282.

Jurgenson, C. T., Begley, T. P. & Ealick, S. E. (2009). The structural and biochemical foundations of thiamin biosynthesis. *Annual review of biochemistry*, 78, 569-603.

Kadono, T., Kira, N., Suzuki, K., Iwata, O., Ohama, T., Okada, S., Nishimura, T., Akakabe, M., Tsuda, M. & Adachi, M. (2015a). Effect of an introduced phytoene synthase gene expression on carotenoid biosynthesis in the marine diatom *Phaeodactylum tricornutum*. *Marine drugs*, 13(8), 5334-5357.

Kadono, T., Miyagawa-Yamaguchi, A., Kira, N., Tomaru, Y., Okami, T., Yoshimatsu, T., Hou, L., Ohama, T., Fukunaga, K., Okauchi, M., Yamaguchi, H., Ohnishi, K., Falciatore, A. & Adachi, M. (2015b). Characterization of marine diatom-infecting virus promoters in the model diatom *Phaeodactylum tricornutum*. *Scientific reports*, 5, 18708.

Karas, B. J., Diner, R. E., Lefebvre, S. C., McQuaid, J., Phillips, A. P., Noddings, C. M., Brunson, J. K., Valas, R. E., Deerinck, T. J., Jablanovic, J., Gillard, J. T. F., Beeri, K., Ellisman, M. H., Glass, J. I., Hutchison III, 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. *Nature communications*, 6, 6925.

Kazamia, E., Helliwell, K. E., Purton, S. & Smith, A. G. (2016). How mutualisms arise in phytoplankton communities: building eco-evolutionary principles for aquatic microbes. *Ecology letters*, 19(7), 810-822.

Keeling, P. J. (2010). The endosymbiotic origin, diversification and fate of plastids. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 365(1541), 729-748.

Kelley, L. A., Mezulis, S., Yates, C. M., Wass, M. N. & Sternberg, M. J. (2015). The Phyre2 web portal for protein modeling, prediction and analysis. *Nature protocols*, 10(6), 845.

Kim, J., Brown, C. M., Kim, M. K., Burrows, E. H., Bach, S., Lun, D. S. & Falkowski, P. G. (2017). Effect of cell cycle arrest on intermediate metabolism in the marine diatom *Phaeodactylum tricornutum*. *Proceedings of the National Academy of Sciences*, 114(38), E8007-E8016.

Kim, S., Kim, D., Cho, S. W., Kim, J. & Kim, J. S. (2014). Highly efficient RNA-guided genome editing in human cells via delivery of purified Cas9 ribonucleoproteins. *Genome research*, 24(6), 1012-1019.

Komeili, A. & O'Shea, E. K. (2000). Nuclear transport and transcription. *Current opinion in cell biology*, 12(3), 355-360.

Komeili, A. & O'Shea, E. K. (1999). Roles of phosphorylation sites in regulating activity of the transcription factor Pho4. *Science*, 284(5416), 977-980.

Kong, D., Zhu, Y., Wu, H., Cheng, X., Liang, H., & Ling, H. Q. (2008). AtTHIC, a gene involved in thiamine biosynthesis in Arabidopsis thaliana. *Cell research*, *18*(5), 566-576.

Kroth, P. G., Bones, A. M., Daboussi, F., Ferrante, M. I., Jaubert, M., Kolot, M., Nymark, M., Río Bártulos, C., Ritter, A., Russo, M. T., Serif, M., Winge, P. & Falciatore, A. (2018). Genome editing in diatoms: achievements and goals. *Plant cell reports*, 37(10), 1401-1408.

Kuang, D., Qiao, J., Li, Z., Wang, W., Xia, H., Jiang, L., Dai, J., Fang, Q. & Dai, X. (2017). Tagging to endogenous genes of *Plasmodium falciparum* using CRISPR/Cas9. *Parasites & vectors*, 10(1), 1-8.

Kumar S., Stecher G., Li M., Knyaz C., and Tamura K. (2018). MEGA X: Molecular Evolutionary Genetics Analysis across computing platforms. *Molecular Biology and Evolution,* 35, 1547-1549.

Kuras, L., Barbey, R. & Thomas, D. (1997). Assembly of a bZIP–bHLH transcription activation complex: formation of the yeast Cbf1–Met4–Met28 complex is regulated through Met28 stimulation of Cbf1 DNA binding. *The EMBO journal*, 16(9), 2441-2451.

Lai, R. Y., Huang, S., Fenwick, M. K., Hazra, A., Zhang, Y., Rajashankar, K., Philmus, B., Kinsland, C., Sanders, J. M., Ealick, S. E. & Begley, T. P. (2012). Thiamin pyrimidine biosynthesis in *Candida albicans*: a remarkable reaction between histidine and pyridoxal phosphate. *Journal of the American Chemical Society*, 134(22), 9157-9159.

Lamas-Toranzo, I., Galiano-Cogolludo, B., Cornudella-Ardiaca, F., Cobos-Figueroa, J., Ousinde, O. & Bermejo-Álvarez, P. (2019). Strategies to reduce genetic mosaicism following CRISPR-mediated genome edition in bovine embryos. *Scientific reports*, 9(1), 1-8.

Lauersen, K. J., Baier, T., Wichmann, J., Wördenweber, R., Mussgnug, J. H., Hübner, W., Huser, T. & Kruse, O. (2016). Efficient phototrophic production of a high-value sesquiterpenoid from the eukaryotic microalga *Chlamydomonas reinhardtii*. *Metabolic engineering*, 38, 331-343.

Lauersen, K. J., Wichmann, J., Baier, T., Kampranis, S. C., Pateraki, I., Møller, B. L. & Kruse, O. (2018). Phototrophic production of heterologous diterpenoids and a hydroxy-functionalized derivative from *Chlamydomonas reinhardtii*. *Metabolic engineering*, 49, 116-127.

Lee, M. C., Jänicke, A. & Beilharz, T. H. (2014). Using Klenow-mediated extension to measure poly (A)-tail length and position in the transcriptome. In Rorbach, J. & Bobrowicz, A. J. (Eds.), *Polyadenylation*, 25-42. Humana Press.

Levitan, O., Dinamarca, J., Zelzion, E., Gorbunov, M. Y. & Falkowski, P. G. (2015a). An RNA interference knock-down of nitrate reductase enhances lipid biosynthesis in the diatom *Phaeodactylum tricornutum*. *The Plant Journal*, 84(5), 963-973.

Levitan, O., Dinamarca, J., Zelzion, E., Lun, D. S., Guerra, L. T., Kim, M. K., Kim, J., Van Mooy, B. A. S., Bhattacharya, D. & Falkowski, P. G. (2015b). Remodeling of intermediate metabolism in the diatom *Phaeodactylum tricornutum* under nitrogen stress. *Proceedings of the National Academy of Sciences*, 112(2), 412-417.

Li, N. & Seetharam, B. (1998). A 69-Base Pair Fragment Derived from Human Transcobalamin II Promoter Is Sufficient for High Bidirectional Activity in the Absence of a TATA Box and an Initiator Element in Transfected Cells Role of an E Box in Transcriptional Activity. *Journal of Biological Chemistry*, 273(43), 28170-28177.

Li, X., Patena, W., Fauser, F., Jinkerson, R. E., Saroussi, S., Meyer, M. T., Ivanova, N., Robertson, J. M., Yue, R., Zhang, R., Vilarrasa-Blasi, J., Wittkopp, T. M., Ramundo, S., Blum, S. R., Goh, A., Laudon, M., Srikumar, T., Lefebvre, P. A., Grossman, A. R. & Jonikas, M. C. (2019). A genome-wide algal mutant library and functional screen identifies genes required for eukaryotic photosynthesis. *Nature genetics*, 51(4), 627-635.

Lian, J., Wijffels, R. H., Smidt, H. & Sipkema, D. (2018). The effect of the algal microbiome on industrial production of microalgae. *Microbial biotechnology*, 11(5), 806-818.

Liang, J. C., Bloom, R. J. & Smolke, C. D. (2011). Engineering biological systems with synthetic RNA molecules. *Molecular cell*, 43(6), 915-926.

Lin, H. Y., Yen, S. C., Kuo, P. C., Chung, C. Y., Yeh, K. L., Huang, C. H., Chang, J. & Lin, H. J. (2017). Alkaline phosphatase promoter as an efficient driving element for exogenic recombinant in the marine diatom *Phaeodactylum tricornutum*. *Algal research*, 23, 58-65.

Lin, S., Staahl, B. T., Alla, R. K. & Doudna, J. A. (2014). Enhanced homology-directed human genome engineering by controlled timing of CRISPR/Cas9 delivery. *eLife*, 3, e04766.

Liu, X., Hempel, F., Stork, S., Bolte, K., Moog, D., Heimerl, T., Maier, U. G. & Zauner, S. (2016). Addressing various compartments of the diatom model organism *Phaeodactylum tricornutum* via sub-cellular marker proteins. *Algal research*, 20, 249-257.

Locher, K. P. & Borths, E. (2004). ABC transporter architecture and mechanism: implications from the crystal structures of BtuCD and BtuF. *FEBS letters*, 564(3), 264-268.

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 biology*, 13(7), R66.

Maheswari, U., Mock, T., Armbrust, E. V. & Bowler, C. (2009). Update of the Diatom EST Database: a new tool for digital transcriptomics. *Nucleic acids research*, 37(suppl_1), D1001-D1005.

Malandrinos, G., Louloudi, M. & Hadjiliadis, N. (2006). Thiamine models and perspectives on the mechanism of action of thiamine-dependent enzymes. *Chemical Society Reviews*, 35(8), 684-692.

Malkowski, S. N., Spencer, T. C. & Breaker, R. R. (2019). Evidence that the nadA motif is a bacterial riboswitch for the ubiquitous enzyme cofactor NAD+. *RNA*, 25(12), 1616-1627.

Martinis, J., Gas-Pascual, E., Szydlowski, N., Crèvecoeur, M., Gisler, A., Bürkle, L. & Fitzpatrick, T. B. (2016). Long-distance transport of thiamine (vitamin B1) is concomitant with that of polyamines. *Plant Physiology*, 171(1), 542-553.

Maruyama-Nakashita, A., Nakamura, Y., Tohge, T., Saito, K. & Takahashi, H. (2006). *Arabidopsis* SLIM1 is a central transcriptional regulator of plant sulphur response and metabolism. *The Plant Cell*, 18(11), 3235-3251.

Materna, A. C., Sturm, S., Kroth, P. G. & Lavaud, J. (2009). First induced plastid genome mutations in an alga with secondary plastids: psba mutations in the diatom *Phaeodactylum tricornutum* (bacillariophyceae) reveal consequences on the regulation of photosynthesis. *Journal of phycology*, 45(4), 838-846.

Matthijs, M., Fabris, M., Broos, S., Vyverman, W. & Goossens, A. (2016). Profiling of the early nitrogen stress response in the diatom *Phaeodactylum tricornutum* reveals a novel family of RING-domain transcription factors. *Plant Physiology*, 170(1), 489-498.

Matthijs, M., Fabris, M., Obata, T., Foubert, I., Franco-Zorrilla, J. M., Solano, R., Fernie, A. R., Vyverman, W. & Goossens, A. (2017). The transcription factor bZIP14 regulates the TCA cycle in the diatom *Phaeodactylum tricornutum*. *The EMBO journal*, 36(11), 1559-1576.

Maupin-Furlow, J. A. (2018). Vitamin B1 (Thiamine) Metabolism and Regulation in Archaea. In LeBlanc, J. G. & Savoy De Giori, G. (Eds.), *B Group Vitamins-Current Uses and Perspectives*. IntechOpen. http://dx.doi.org/10.5772/intechopen.77170

Mazumder, B., Seshadri, V. & Fox, P. L. (2003). Translational control by the 3'-UTR: the ends specify the means. *Trends in biochemical Sciences*, 28(2), 91-98.

McCarthy, J. K., Smith, S. R., McCrow, J. P., Tan, M., Zheng, H., Beeri, K., Roth, R., Lichtle, C., Goodenough, U., Bowler, C. P., Dupont, C. L. & Allen, A. E. (2017). Nitrate reductase knockout uncouples nitrate transport from nitrate assimilation and drives repartitioning of carbon flux in a model pennate diatom. *The Plant Cell*, 29(8), 2047-2070.

McColl, D., Valencia, C. A. & Vierula, P. J. (2003). Characterization and expression of the Neurospora crassa nmt-1 gene. *Current genetics*, 44(4), 216-223.

McCown, P. J., Corbino, K. A., Stav, S., Sherlock, M. E. & Breaker, R. R. (2017). Riboswitch diversity and distribution. *RNA*, 23(7), 995-1011.

McFadden, G. I. (2001). Primary and secondary endosymbiosis and the origin of plastids. *Journal of phycology*, 37(6), 951-959.

McRose, D., Guo, J., Monier, A., Sudek, S., Wilken, S., Yan, S., Mock, T., Archibald, J. M., Begley, T. P., Reyes-Prieto, A. & Worden, A. Z. (2014). Alternatives to vitamin B 1 uptake revealed with discovery of riboswitches in multiple marine eukaryotic lineages. *The ISME journal*, 8(12), 2517-2529.

Mehravar, M., Shirazi, A., Nazari, M. & Banan, M. (2019). Mosaicism in CRISPR/Cas9-mediated genome editing. *Developmental biology*, 445(2), 156-162.

Mehrshahi, P., Nguyen, G. T. D., Gorchs Rovira, A., Sayer, A., Llavero-Pasquina, M., Lim Huei Sin, M., Medcalf, E. J., Mendoza-Ochoa, G. I., Scaife, M. A. & Smith, A. G. (2020). Development of Novel Riboswitches for Synthetic Biology in the Green Alga *Chlamydomonas*. *ACS synthetic biology*, 9(6), 1406–1417.

Mehta, A. P., Abdelwahed, S. H., Fenwick, M. K., Hazra, A. B., Taga, M. E., Zhang, Y., Ealick, S. E. & Begley, T. P. (2015). Anaerobic 5-hydroxybenzimidazole formation from aminoimidazole ribotide: An unanticipated intersection of thiamin and vitamin B12 biosynthesis. *Journal of the American Chemical Society*, 137(33), 10444-10447.

Merkle, T. (2001). Nuclear import and export of proteins in plants: a tool for the regulation of signalling. *Planta*, 213(4), 499-517.

Miyahara, M., Aoi, M., Inoue-Kashino, N., Kashino, Y. & Ifuku, K. (2013). Highly efficient transformation of the diatom *Phaeodactylum tricornutum* by multi-pulse electroporation. *BioScience, biotechnology, and biochemistry*, 77(4), 874-876.

Mogollon, C. M., Van Pul, F. J., Imai, T., Ramesar, J., Chevalley-Maurel, S., De Roo, G. M., Veld, S. A. J., Kroeze, H., Franke-Fayard, B. M. D., Janse, C. J. & Khan, S. M. (2016). Rapid generation of marker-free P. falciparum fluorescent reporter lines using modified CRISPR/Cas9 constructs and selection protocol. *PloS one*, 11(12), e0168362.

Moldovan, M. A., Petrova, S. A. & Gelfand, M. S. (2018). Comparative genomic analysis of fungal TPP-riboswitches. *Fungal Genetics and Biology*, 114, 34-41.

Monteverde, D. R., Gómez-Consarnau, L., Suffridge, C. & Sañudo-Wilhelmy, S. A. (2017). Life's utilization of B vitamins on early Earth. *Geobiology*, 15(1), 3-18.

Moosburner, M. A., Gholami, P., McCarthy, J. K., Tan, M., Bielinski, V. A. & Allen, A. E. (2020). Multiplexed knockouts in the model diatom *Phaeodactylum* by episomal delivery of a selectable cas9. *Frontiers in microbiology*, 11.

Moulin, M., Nguyen, G. T., Scaife, M. A., Smith, A. G. & Fitzpatrick, T. B. (2013). Analysis of *Chlamydomonas* thiamin metabolism in vivo reveals riboswitch plasticity. *Proceedings of the National Academy of Sciences*, 110(36), 14622-14627.

Mukhopadhyay, A., Borkakoti, N., Pravda, L., Tyzack, J. D., Thornton, J. M. & Velankar, S. (2019). Finding enzyme cofactors in Protein Data Bank. *Bioinformatics*, 35(18), 3510-3511.

Naduthodi, M. I. S., Mohanraju, P., Südfeld, C., D'Adamo, S., Barbosa, M. J. & Van Der Oost, J. (2019). CRISPR–Cas ribonucleoprotein mediated homology-directed repair for efficient targeted genome editing in microalgae *Nannochloropsis oceanica IMET1*. *Biotechnology for biofuels*, 12(1), 1-11.

Nahvi, A., Sudarsan, N., Ebert, M. S., Zou, X., Brown, K. L. & Breaker, R. R. (2002). Genetic control by a metabolite binding mRNA. *Chemistry & biology*, 9(9), 1043-1049.

Nelson, D. M., Tréguer, P., Brzezinski, M. A., Leynaert, A. & Quéguiner, B. (1995). Production and dissolution of biogenic silica in the ocean: revised global estimates, comparison with regional data and relationship to biogenic sedimentation. *Global Biogeochemical Cycles*, 9(3), 359-372.

Nguyen, G. T., Scaife, M. A., Helliwell, K. E. & Smith, A. G. (2016). Role of riboswitches in gene regulation and their potential for algal biotechnology. *Journal of phycology*, 52(3), 320-328.

Nguyen, G.T.D.T. (2014). *Thiamine pyrophosphate riboswitches in Chlamydomonas reinhardtii* (Doctoral Dissertation, University of Cambridge).

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(6), 1-3.

Nosaka, K. (2006). Recent progress in understanding thiamin biosynthesis and its genetic regulation in *Saccharomyces cerevisiae*. *Applied microbiology and biotechnology*, 72(1), 30-40.

Nosaka, K., Onozuka, M., Konno, H., Kawasaki, Y., Nishimura, H., Sano, M. & Akaji, K. (2005). Genetic regulation mediated by thiamin pyrophosphate-binding motif in *Saccharomyces cerevisiae*. *Molecular microbiology*, 58(2), 467-479.

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

Ohno, N., Inoue, T., Yamashiki, R., Nakajima, K., Kitahara, Y., Ishibashi, M. & Matsuda, Y. (2012). CO2-cAMP-responsive cis-elements targeted by a transcription factor with CREB/ATF-like basic zipper domain in the marine diatom *Phaeodactylum tricornutum*. *Plant Physiology*, 158(1), 499-513.

Okamoto, T., Kawaguchi, K., Watanabe, S., Agustina, R., Ikejima, T., Ikeda, K., Nakano, M., Morita, M. & Imanaka, T. (2018). Characterization of human ATP-binding cassette protein subfamily D reconstituted into proteoliposomes. *Biochemical and biophysical research communications*, 496(4), 1122-1127.

Palmer, L. D. & Downs, D. M. (2013). The thiamine biosynthetic enzyme ThiC catalyses multiple turnovers and is inhibited by S-adenosylmethionine (AdoMet) metabolites. *Journal of Biological Chemistry*, *288*(42), 30693-30699.

Park, P. J. (2009). ChIP–seq: advantages and challenges of a maturing technology. *Nature reviews genetics*, 10(10), 669-680.

Patron, N. J., Orzaez, D., Marillonnet, S., Warzecha, H., Matthewman, C., Youles, M., Raitskin, O., Leveau, A., Farré, G., Rogers, C., Smith, A. G., Hibberd, J., Webb, A. A. R., Locke, J., Schornack, S., Ajioka, J., Baulcombe, D. C., Zipfel, C., Kamoun, S., Jones, J. D. G., Kuhn, H., Robatzek, S., Van Esse, H. P., Sanders, D., Oldroyd, G., Martin, C., Field, R., O'Connor, S., Fox, S., Wulff, B., Miller, B., Breakspear, A., Radhakrishnan, G., Delaux, P., Loqué, D., Granell, A., Tissier, A., Shih, P., Brutnell, T. P., Quick, W. P., Rischer, H., Fraser, P. D., Aharoni, A., Raines, C., South, P. F., Ané, J., Hamberger, B. R., Langdale, J., Stougaard, J., Bouwmeester, H., Udvardi, M., Murray, J. A. H., Ntoukakis, V., Schäfer, P., Denby, K.,

Edwards, K. J., Osbourn, A. & Haseloff, J. (2015). Standards for plant synthetic biology: a common syntax for exchange of DNA parts. *New Phytologist*, 208(1), 13-19.

Paxhia, M. D. & Downs, D. M. (2019). SNZ3 encodes a PLP synthase involved in thiamine synthesis in *Saccharomyces cerevisiae*. *G3: Genes, Genomes, Genetics*, 9(2), 335-344.

Pireyre, M. & Burow, M. (2015). Regulation of MYB and bHLH transcription factors: a glance at the protein level. *Molecular plant*, 8(3), 378-388.

Popko, J., Herrfurth, C., Feussner, K., Ischebeck, T., Iven, T., Haslam, R., Hamilton, M., Sayanova, O., Napier, J., Khozin-Goldberg, I. & Feussner, I. (2016). Metabolome analysis reveals betaine lipids as major source for triglyceride formation, and the accumulation of sedoheptulose during nitrogen-starvation of *Phaeodactylum tricornutum*. *PloS one*, 11(10).

Poulsen, N. & Kröger, N. (2005). A new molecular tool for transgenic diatoms: control of mRNA and protein biosynthesis by an inducible promoter–terminator cassette. *The FEBS journal*, 272(13), 3413-3423.

Prlić, A., Bliven, S., Rose, P. W., Bluhm, W. F., Bizon, C., Godzik, A. & Bourne, P. E. (2010). Pre-calculated protein structure alignments at the RCSB PDB website. *Bioinformatics*, 26(23), 2983-2985.

Radakovits, R., Eduafo, P. M. & Posewitz, M. C. (2011). Genetic engineering of fatty acid chain length in *Phaeodactylum tricornutum*. *Metabolic engineering*, 13(1), 89-95.

Rajgopal, A., Edmondnson, A., Goldman, I. D. & Zhao, R. (2001). SLC19A3 encodes a second thiamine transporter ThTr2. *Biochimica et Biophysica Acta (BBA)-Molecular Basis of Disease*, 1537(3), 175-178.

Rao, X., Huang, X., Zhou, Z. & Lin, X. (2013). An improvement of the 2ˆ(–delta delta CT) method for quantitative real-time polymerase chain reaction data analysis. *Biostatistics, bioinformatics and biomathematics*, 3(3), 71.

Rastogi, A., Maheswari, U., Dorrell, R. G., Vieira, F. R. J., Maumus, F., Kustka, A., McCarthy, J., Allen, A. E., Kersey, P., Bowler, C. & Tirichine, L. (2018). Integrative analysis of large scale transcriptome data draws a comprehensive landscape of *Phaeodactylum tricornutum* genome and evolutionary origin of diatoms. *Scientific reports*, 8(1), 1-14.

Raux, E., Lanois, A., Levillayer, F., Warren, M. J., Brody, E., Rambach, A. & Thermes, C. (1996). *Salmonella typhimurium* cobalamin (vitamin B12) biosynthetic genes: functional studies in *S. typhimurium* and *Escherichia coli*. *Journal of bacteriology*, 178(3), 753-767.

Raux, E., Schubert, H. L. & Warren, M. J. (2000). Biosynthesis of cobalamin (vitamin B12): a bacterial conundrum. *Cellular and Molecular Life Science*s, 57(13-14), 1880-1893.

Ravnum, S. & Andersson, D. I. (2001). An adenosyl–cobalamin (coenzyme-B12)-repressed translational enhancer in the cob mRNA of *Salmonella typhimurium*. *Molecular microbiology*, 39(6), 1585-1594.

Rayko, E., Maumus, F., Maheswari, U., Jabbari, K. & Bowler, C. (2010). Transcription factor families inferred from genome sequences of photosynthetic stramenopiles. *New Phytologist*, 188(1), 52-66.

Remmers, I. M., D'Adamo, S., Martens, D. E., de Vos, R. C., Mumm, R., America, A. H., Cordewener, J. H. G., Bakker, L. V., Peters, S. A., Wijffels, R. H. & Lamers, P. P. (2018). Orchestration of transcriptome, proteome and metabolome in the diatom *Phaeodactylum tricornutum* during nitrogen limitation. *Algal research*, 35, 33-49.

Roberts, B., Haupt, A., Tucker, A., Grancharova, T., Arakaki, J., Fuqua, M. A., Nelson, A., Hookway, C., Ludmann, S. A., Mueller, I. A., Yang, R., Horwitz, R., Rafelski, S. M. & Gunawardane, R. N. (2017). Systematic gene tagging using CRISPR/Cas9 in human stem cells to illuminate cell organization. *Molecular biology of the cell*, 28(21), 2854-2874.

Rodionov, D. A., Leyn, S. A., Li, X. & Rodionova, I. A. (2017). A novel transcriptional regulator related to thiamine phosphate synthase controls thiamine metabolism genes in archaea. *Journal of bacteriology*, 199(4), e00743-16.

Rodionov, D. A., Vitreschak, A. G., Mironov, A. A. & Gelfand, M. S. (2002). Comparative genomics of thiamin biosynthesis in procaryotes New genes and regulatory mechanisms. *Journal of Biological Chemistry*, 277(50), 48949-48959.

Roth, A. & Breaker, R. R. (2009). The structural and functional diversity of metabolite-binding riboswitches. *Annual review of biochemistry*, 78, 305-334.

Roth, J. R., Lawrence, J. G. & Bobik, T. A. (1996). Cobalamin (coenzyme B12): synthesis and biological significance. *Annual review of microbiology*, 50(1), 137-181.

Russo, M. T., Cigliano, R. A., Sanseverino, W. & Ferrante, M. I. (2018). Assessment of genomic changes in a CRISPR/Cas9 *Phaeodactylum tricornutum* mutant through whole genome resequencing. *PeerJ*, 6, e5507.

Sañudo-Wilhelmy, S. A., Cutter, L. S., Durazo, R., Smail, E. A., Gómez-Consarnau, L., Webb, E. A., Prokopenko, M. G., Berelson, W. M. & Karl, D. M. (2012). Multiple B-vitamin depletion in large areas of the coastal ocean. *Proceedings of the National Academy of Sciences*, 109(35), 14041-14045.

Sañudo-Wilhelmy, S. A., Gobler, C. J., Okbamichael, M. & Taylor, G. T. (2006). Regulation of phytoplankton dynamics by vitamin B12. *Geophysical research letters*, 33(4), L04604.

Sañudo-Wilhelmy, S. A., Gómez-Consarnau, L., Suffridge, C. & Webb, E. A. (2014). The role of B vitamins in marine biogeochemistry. *Annual review of marine Science*, 6, 339-367.

Schyns, G., Potot, S., Geng, Y., Barbosa, T. M., Henriques, A. & Perkins, J. B. (2005). Isolation and characterization of new thiamine-deregulated mutants of *Bacillus subtilis*. *Journal of bacteriology*, 187(23), 8127-8136.

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

Serganov, A. & Patel, D. J. (2012). Metabolite recognition principles and molecular mechanisms underlying riboswitch function. *Annual review of biophysics*, 41, 343-370.

Serif, M., Dubois, G., Finoux, A. L., Teste, M. A., Jallet, D. & Daboussi, F. (2018). One-step generation of multiple gene knock-outs in the diatom *Phaeodactylum tricornutum* by DNA-free genome editing. *Nature communications*, 9(1), 1-10.

Sharma, A. K., Nymark, M., Sparstad, T., Bones, A. M. & Winge, P. (2018). Transgene-free genome editing in marine algae by bacterial conjugation–comparison with biolistic CRISPR/Cas9 transformation. *Scientific reports*, 8(1), 1-11.

Sharma, A.K., Mühlroth, A., Jouhet, J., Maréchal, E., Alipanah, L., Kissen, R., Brembu, T., Bones, A. M. & Winge, P. (2020). The Myb-like transcription factor phosphorus starvation response (PtPSR) controls conditional P acquisition and remodelling in marine microalgae. *New Phytologist*, 225(6), 2380-2395.

Shimizu, T., Toumoto, A., Ihara, K., Shimizu, M., Kyogoku, Y., Ogawa, N., Oshima, Y. & Hakoshima, T. (1997). Crystal structure of PHO4 bHLH domain–DNA complex: flanking base recognition. *The EMBO journal*, 16(15), 4689-4697.

Sibbald, S. J. & Archibald, J. M. (2020). Genomic insights into plastid evolution. *Genome Biology and Evolution,* 12(7), 978–990.

Sievers, F., Wilm, A., Dineen, D., Gibson, T. J., Karplus, K., Li, W., Lopez, R., McWilliam, H., Remmert, M., Söding, J., Thompson, J. D. & Higgins, D. G. (2011). Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. *Molecular systems biology*, 7(1), 539.

Singleton, C. K. (1997). Identification and characterization of the thiamine transporter gene of *Saccharomyces cerevisiae*. *Gene*, 199(1-2), 111-121.

Slattery, S. S., Diamond, A., Wang, H., Therrien, J. A., Lant, J. T., Jazey, T., Lee, K., Klassen, Z., Desgagné-Penix, I., Karas, B. J. & Edgell, D. R. (2018). An expanded plasmid-based genetic toolbox enables Cas9 genome editing and stable maintenance of synthetic pathways in *Phaeodactylum tricornutum*. *ACS synthetic biology*, 7(2), 328-338.

Slattery, S. S., Wang, H., Giguere, D. J., Kocsis, C., Urquhart, B. L., Karas, B. J. & Edgell, D. R. (2020). Plasmid-based complementation of large deletions in *Phaeodactylum tricornutum* biosynthetic genes generated by Cas9 editing. *Scientific reports*, 10(1), 1-12.

Slotboom, D. J. (2014). Structural and mechanistic insights into prokaryotic energy-coupling factor transporters. *Nature Reviews Microbiology*, 12(2), 79-87.

Sonnhammer, E. L., Von Heijne, G. & Krogh, A. (1998). A hidden Markov model for predicting transmembrane helices in protein sequences. In Glasgow, J., Littlejohn, T., Major, F., Lathrop, R., Sankoff, D. & Sensen, C. (Eds.), *Proceedings of the Sixth International Conference on Intelligent Systems for Molecular Biology*, 175-182. AAAI Press.

Sudarsan, N., Barrick, J. E. & Breaker, R. R. (2003). Metabolite-binding RNA domains are present in the genes of eukaryotes. *RNA*, 9(6), 644-647.

Sudarsan, N., Cohen-Chalamish, S., Nakamura, S., Emilsson, G. M. & Breaker, R. R. (2005). Thiamine pyrophosphate riboswitches are targets for the antimicrobial compound pyrithiamine. *Chemistry & biology*, 12(12), 1325-1335.

Sudarsan, N., Wickiser, J. K., Nakamura, S., Ebert, M. S. & Breaker, R. R. (2003). An mRNA structure in bacteria that controls gene expression by binding lysine. *Genes & development*, 17(21), 2688-2697.

Suffridge, C., Cutter, L. & Sañudo-Wilhelmy, S. A. (2017). A new analytical method for direct measurement of particulate and dissolved B-vitamins and their congeners in seawater. *Frontiers in Marine Science*, 4, 11.

Suliman, H. S., Sawyer, G. M., Appling, D. R., & Robertus, J. D. (2005). Purification and properties of cobalamin-independent methionine synthase from *Candida albicans* and *Saccharomyces cerevisiae*. *Archives of Biochemistry and Biophysics*, *441*(1), 56-63.

Szathmáry, E. (1999). The origin of the genetic code: amino acids as cofactors in an RNA world. *Trends in Genetics*, 15(6), 223-229.

Szcześniak, M. W., Kabza, M., Pokrzywa, R., Gudyś, A. & Makałowska, I. (2013). ERISdb: a database of plant splice sites and splicing signals. *Plant and Cell Physiology*, 54(2), e10.

Taga, M. E., Larsen, N. A., Howard-Jones, A. R., Walsh, C. T. & Walker, G. C. (2007). BluB cannibalizes flavin to form the lower ligand of vitamin B 12. *Nature*, 446(7134), 449-453.

Tanaka, T., Maeda, Y., Veluchamy, A., Tanaka, M., Abida, H., Maréchal, E., Bowler, C., Muto, M., Sunaga, Y., Tanaka, M., Yoshino, T., Taniguchi, T., Fukuda, Y., Nemoto, M., Matsumoto, M., Wong, P. S., Aburatani, S. & Fujibuchi, W. (2015). Oil accumulation by the oleaginous diatom *Fistulifera solaris* as revealed by the genome and transcriptome. *The Plant Cell*, 27(1), 162-176.

Tang, Y. Z., Koch, F. & Gobler, C. J. (2010). Most harmful algal bloom species are vitamin B1 and B12 auxotrophs. *Proceedings of the National Academy of Sciences*, 107(48), 20756-20761.

Tao, Y. & Marzluf, G. A. (1998). Synthesis and differential turnover of the CYS3 regulatory protein of Neurospora crassa are subject to sulphur control. *Journal of bacteriology*, 180(3), 478-482.

Taparia, Y., Zarka, A., Leu, S., Zarivach, R., Boussiba, S. & Khozin-Goldberg, I. (2019). A novel endogenous selection marker for the diatom *Phaeodactylum tricornutum* based on a unique mutation in phytoene desaturase 1. *Scientific reports*, 9(1), 1-12.

Thomas, D. & Surdin-Kerjan, Y. (1997). Metabolism of sulphur amino acids in *Saccharomyces cerevisiae*. *Microbiology and Molecular Biology Reviews*, 61(4), 503-532.

Thore, S., Frick, C. & Ban, N. (2008). Structural basis of thiamine pyrophosphate analogues binding to the eukaryotic riboswitch. *Journal of the American Chemical Society*, 130(26), 8116-8117.

Thore, S., Leibundgut, M. & Ban, N. (2006). Structure of the eukaryotic thiamine pyrophosphate riboswitch with its regulatory ligand. *Science*, 312(5777), 1208-1211.

Tian, B. & Manley, J. L. (2017). Alternative polyadenylation of mRNA precursors. *Nature reviews Molecular cell biology*, 18(1), 18.

Traller, J. C., Cokus, S. J., Lopez, D. A., Gaidarenko, O., Smith, S. R., McCrow, J. P., Gallaher, S. D., Podell, S., Thompson, M., Cook, O., Morselli, M., Jaroszewicz, A., Allen, E. E., Allen, A. E., Merchant, S. S., Pellegrini, M. & Hildebrand, M. (2016). Genome and methylome of the oleaginous diatom *Cyclotella cryptica* reveal genetic flexibility toward a high lipid phenotype. *Biotechnology for biofuels*, 9(1), 258.

Tran, D. H., Shishido, Y., Chung, S. P., Trinh, H. T. T., Yorita, K., Sakai, T. & Fukui, K. (2015). Identification of DNA-binding proteins that interact with the 5'-flanking region of the human d-amino acid oxidase gene by pull-down assay coupled with two-dimensional gel electrophoresis and mass spectrometry. *Journal of pharmaceutical and biomedical analysis*, 116, 94-100.

Tuite, M. F. (1992). Strategies for the genetic manipulation of *Saccharomyces cerevisiae*. *Critical reviews in biotechnology*, 12(1-2), 157-188.

Uhm, H., Kang, W., Ha, K. S., Kang, C. & Hohng, S. (2018). Single-molecule FRET studies on the cotranscriptional folding of a thiamine pyrophosphate riboswitch. *Proceedings of the National Academy of Sciences*, 115(2), 331-336.

Vancaester, E., Depuydt, T., Osuna-Cruz, C. M. & Vandepoele, K. (2020). Comprehensive and functional analysis of horizontal gene transfer events in diatoms. *Molecular Biology and Evolution*.

Veluchamy, A., Lin, X., Maumus, F., Rivarola, M., Bhavsar, J., Creasy, T., O'Brien, K., Sengamalay, N. A., Tallon, L. J., Smith, A. D., Rayko, E., Ahmed, I., Le Crom, S., Farrant, G. K., Sgro, J. Y., Olson, S. A., Splinter Bondurant, S., Allen, A. E., Rabinowicz, P. D., Sussman, M. R., Bowler, C. & Tirichine, L. (2013). Insights into the role of DNA methylation in diatoms by genome-wide profiling in *Phaeodactylum tricornutum*. *Nature communications*, 4(1), 1-10.

Veluchamy, A., Rastogi, A., Lin, X., Lombard, B., Murik, O., Thomas, Y., Dingli, F., Rivarola, M., Ott, S., Liu, X., Sun, Y., Rabinowicz, P. D., McCarthy, J., Allen, A. E., Loew, D., Bowler, C. & Tirichine, L. (2015). An integrative analysis of post-translational histone modifications in the marine diatom *Phaeodactylum tricornutum*. *Genome biology*, 16(1), 102.

Vogl, C., Klein, C. M., Batke, A. F., Schweingruber, M. E. & Stolz, J. (2008). Characterization of Thi9, a novel thiamine (Vitamin B1) transporter from *Schizosaccharomyces pombe*. *Journal of Biological Chemistry*, 283(12), 7379-7389.

Wachter, A. (2010). Riboswitch-mediated control of gene expression in eukaryotes. *RNA biology*, 7(1), 67-76.

Wachter, A., Tunc-Ozdemir, M., Grove, B. C., Green, P. J., Shintani, D. K. & Breaker, R. R. (2007). Riboswitch control of gene expression in plants by splicing and alternative 3' end processing of mRNAs. *The Plant Cell*, 19(11), 3437-3450.

Watanabe, Y., Kadono, T., Kira, N., Suzuki, K., Iwata, O., Ohnishi, K., Yamaguchi, H. & Adachi, M. (2018). Development of endogenous promoters that drive high-level expression of introduced genes in the model diatom *Phaeodactylum tricornutum*. *Marine genomics*, 42, 41-48.

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

Webb, E., Claas, K. & Downs, D. (1998). thi BPQ Encodes an ABC Transporter Required for Transport of Thiamine and Thiamine Pyrophosphate in *Salmonella typhimurium*. *Journal of Biological Chemistry*, 273(15), 8946-8950.

Weber, E., Engler, C., Gruetzner, R., Werner, S. & Marillonnet, S. (2011). A modular cloning system for standardized assembly of multigene constructs. *PloS one*, 6(2).

Weyman, P. D., Beeri, K., Lefebvre, S. C., Rivera, J., McCarthy, J. K., Heuberger, A. L., Peers, G., Allen, A. E. & Dupont, C. L. (2015). Inactivation of *Phaeodactylum tricornutum* urease gene using transcription activator-like effector nuclease-based targeted mutagenesis. *Plant biotechnology journal*, 13(4), 460-470.

White, H. B. (1976). Coenzymes as fossils of an earlier metabolic state. *Journal of Molecular Evolution*, 7(2), 101-104.

Wibowo, A. S., Singh, M., Reeder, K. M., Carter, J. J., Kovach, A. R., Meng, W., Ratnam, M., Zhang, F. & Dann III, C. E. (2013). Structures of human folate receptors reveal biological trafficking states and diversity in folate and antifolate recognition. *Proceedings of the National Academy of Sciences*, 110(38), 15180-15188.

Winkler, W., Nahvi, A. & Breaker, R. R. (2002). Thiamine derivatives bind messenger RNAs directly to regulate bacterial gene expression. *Nature*, 419(6910), 952-956.

Wolfs, J. M., Hamilton, T. A., Lant, J. T., Laforet, M., Zhang, J., Salemi, L. M., Gloor, G. B., Schild-Poulter, C., & Edgell, D. R. (2016). Biasing genome-editing events toward precise length deletions with an RNA-guided TevCas9 dual nuclease. *Proceedings of the National Academy of Sciences*, 113(52), 14988-14993.

Woolley, D. W. & White, A. G. C. (1943). Selective reversible inhibition of microbial growth with pyrithiamine. *The Journal of experimental medicine*, 78(6), 489-497.

Worden, A. Z., Lee, J. H., Mock, T., Rouzé, P., Simmons, M. P., Aerts, A. L., Allen, A. E., Cuvelier, M. L., Derelle, E., Everett, M. V., Foulon, E., Grimwood, J., Gundlach, H. , Henrissat, B., Napoli, C., McDonald, S. M., Parker, M. S., Rombauts, S., Salamov, A., Von Dassow, P., Badger, J. H., Coutinho, P. M., Demir, E., Dubchak, I., Gentemann, C., Eikrem, W., Gready, J. E., John, U., Lanier, W., Lindquist, E. A., Lucas, S., Mayer, K. F. X., Moreau, H., Not, F., Otillar, R., Panaud, O., Pangilinan, J., Paulsen, I., Piegu, B., Poliakov, A., Robbens, S., Schmutz, J., Toulza, E., Wyss, T., Zelensky, A., Zhou, K., Armbrust, E. V., Bhattacharya, D., Goodenough, U. W., Van de Peer, Y. & Grigoriev, I. V. (2009). Green evolution and

dynamic adaptations revealed by genomes of the marine picoeukaryotes *Micromonas*. *Science*, *324*(5924), 268-272.

Xie, W. H., Zhu, C. C., Zhang, N. S., Li, D. W., Yang, W. D., Liu, J. S., Sathishkumar, R. & Li, H. Y. (2014). Construction of novel chloroplast expression vector and development of an efficient transformation system for the diatom *Phaeodactylum tricornutum*. *Marine Biotechnology*, 16(5), 538-546.

Xu, D., Feng, Z., Hou, W. T., Jiang, Y. L., Wang, L., Sun, L., Zhou, C. Z. & Chen, Y. (2019). Cryo-EM structure of human lysosomal cobalamin exporter ABCD4. *Cell research*, 29(12), 1039-1041.

Yadav, S., Swati, D. & Chandrasekharan, H. (2015). Thiamine pyrophosphate riboswitch in some representative plant species: a bioinformatics study. *Journal of Computational Biology*, 22(1), 1-9.

Yu, D. & Breaker, R. R. (2020). A bacterial riboswitch class senses xanthine and uric acid to regulate genes associated with purine oxidation. *RNA*, rna-075218.

Yusof, Z. N. B. (2011). *Regulation of thiamine (vitamin B1) biosynthesis in Chlamydomonas reinhardtii* (Doctoral Dissertation, University of Cambridge).

Zaslavskaia, L. A., Lippmeier, J. C., Kroth, P. G., Grossman, A. R. & Apt, K. E. (2000). Transformation of the diatom *Phaeodactylum tricornutum* (Bacillariophyceae) with a variety of selectable marker and reporter genes. *Journal of phycology*, 36(2), 379-386.

Zhang, C. & Hu, H. (2014). High-efficiency nuclear transformation of the diatom *Phaeodactylum tricornutum* by electroporation. *Marine genomics*, 16, 63-66.

Zhang, K., Bian, J., Deng, Y., Smith, A., Nunez, R. E., Li, M. B., Pal, U., Yu, A. M., Qiu, W., Ealick, S. E. & Li, C. (2016). Lyme disease spirochaete *Borrelia burgdorferi* does not require thiamin. *Nature microbiology*, 2(1), 1-6.

Zhang, P., Burel, C., Plasson, C., Kiefer-Meyer, M. C., Ovide, C., Gügi, B., Wan, C., Teo, G., Mak, A., Song, Z. & Driouich, A. (2019). Characterization of a GDP-Fucose Transporter and a Fucosyltransferase Involved in the Fucosylation of Glycoproteins in the Diatom *Phaeodactylum tricornutum*. *Frontiers in plant science*, *10*, 610.

8. Appendix

8.1. Media receipes

8.1.1. f/2

F/2 media was prepared according to Guillard (1975) by adding the following components to 1L of de-ionised water before filter sterilising:

Component	Mass or Volume	Stock Solution Concentration	Final Concentration
Sea Salts (Sigma)	30g	N/A	30g·l^{-1}
$NaNO_3$ (Fisher Chemical)	1ml	75g·l^{-1}	75 mg·l^{-1}
$NaH_2PO_4 \cdot H_2O$ (Fisher Scientific)	1ml	5g·l^{-1}	5 mg·l^{-1}
Trace Metal Solution	1ml	N/A	See table below
$Na_2SiO_3 \cdot 9\ H_2O$	1ml (only for f/2+Si)	30g·l^{-1}	30 mg·l^{-1}

pH was adjusted to 8.0

To prepare the Trace Metal Solution the following components were added to 250ml of de-ionised water before filter sterilising and keeping at 4ºC:

Component	Mass or Volume	Stock Solution Concentration	Final Concentration
$FeCl_3 \cdot 6\ H_2O$ (VWR)	0.7875g	N/A	3.15 mg·l^{-1}
$Na_2EDTA \cdot 2\ H_2O$ (Fisher Chemical)	1.0900g	N/A	4.36 mg·l^{-1}
$CuSO_4 \cdot 5\ H_2O$ (Sigma)	250 µl	9.8g·l^{-1}	9.8µg·l^{-1}
$Na_2MoO_4 \cdot 2\ H_2O$ (Sigma)	250 µl	6.3g·l^{-1}	6.3µg·l^{-1}
$ZnSO_4 \cdot 7\ H_2O$ (Sigma)	250 µl	22.0g·l^{-1}	22µg·l^{-1}
$CoCl_2 \cdot 6\ H_2O$ (Sigma)	250 µl	10.0g·l^{-1}	10µg·l^{-1}
$MnCl_2$ (Sigma)	250 µl	114.5g·l^{-1}	114.5µg·l^{-1}

8.1.2. TAP

TAP media was prepared according to Gorman & Levine (1965) by adding the following to 1L of de-ionised water before autoclaving:

Component	Mass or Volume	Stock Solution Concentration	Final Concentration
TRIS Base (Melford)	2.42g	N/A	2.42g·l^{-1}
Phosphate Buffer II	1ml	288g·l^{-1} K_2HPO_4 (Fisher Chemical) 144g·l^{-1} KH_2PO_4 (Fisher Chemical)	288 mg·l^{-1} 144 mg·l^{-1}
Solution A	10ml	1.5g·l^{-1} NH_4Cl (Sigma) 4g·l^{-1} $MgSO_4 \cdot 7\ H_2O$ (Fisher Chemical) 2g·l^{-1} $CaCl_2 \cdot 2\ H_2O$ (Fisher Chemical)	15 mg·l^{-1} 40 mg·l^{-1} 20 mg·l^{-1}
$Na_2EDTA \cdot 2\ H_2O$ (Fisher Chemical)	1ml	25mM	25 µM
$(NH_4)_6Mo_7O_{24} \cdot 4H_2O$ (MP Biomed.)	1ml	28.5 µM	28.5nM
CuEDTA	1ml	2mM $CuCl_2 \cdot 2\ H_2O$ (Sigma) 2mM $Na_2EDTA \cdot 2\ H_2O$(Fisher Chem.)	2 µM 2 µM
$ZnSO_4 \cdot 7\ H_2O$ (Sigma)	1ml	2.5mM	2.5 µM
$MnCl_2$ (Sigma)	1ml	6mM	6 µM
FeEDTA	1ml	20mM $FeCl_3 \cdot 6H_2O$ (VWR) 22mM $Na_2EDTA \cdot 2H_2O$(Fisher Chem.) 22mM Na_2CO_3 (Fisher Chemical)	20 µM 22 µM 22 µM
Glacial Acetic Acid (Fisher Chem.)	1ml	N/A	1 % (v/v)

pH was adjusted to 7.0

Solid media plates were prepared by mixing 1:1 a 3 % w/v agar (Melford) solution with f/2, or adding 1.5 % w/v agar (Formedium) to TAP.

8.2. Primers

8.2.1. RT-qPCR primers

#	Section	Primer	Sequence
1	3.3.2	H4.fwd	AGGTCCTTCGCGACAATATC
2	3.3.2	H4.rv	ACGGAATCACGAATGACGTT
3	3.3.2	UBC.fwd	TGCTTCGAAGGCAGCTTAAT
4	3.3.2	UBC.rv	ATCGTGCGGAAATACCAGAC
5	3.3.2	UBQ.fwd	GGCGGTCGATCTTAAACGTA
6	3.3.2	UBQ.rv	ATCGGCTCTTTCGTTCTTCA
7	3.3.2	THIC.fwd	CGCCATTATTTGCTCCAAC
8	3.3.2	THIC.rv	CACAGCATGCTCCATTGC
9	3.3.2	NMT1.fwd	TTTGCATTTGGCACAAGACG
10	3.3.2	NMT1.rv	TCCGGGATAGCCAATGATC
11	3.3.2	SSSP.fwd	GTCATGGTGTTTGACGTTTCC
12	3.3.2	SSSP.rv	GACGCTTTCATTCTGATTCAGC
13	3.3.2	THIG.fwd	AGGATGTGCGACTGTTATGC
14	3.3.2	THIG.rv	GCTTGCGTAGCTTCACTTG
15	3.3.2	METE.fwd	GGAAAGGATCGCTGAGTGAAG
16	3.3.2	METE.rv	CCAAATTCTCGAAGCGATGAGG
17	3.3.2	TpActin.fwd	ACGTGACCTCACGGACTACC
18	3.3.2	TpActin.rv	CAGTAAGGTTGGGCTGGAAA
19	3.3.2	TpEF1a.fwd	GTATCGGCACTGTCCCTGTT
20	3.3.2	TpEF1a.rv	ATCTCCTCGCCGAATATCCT

21	3.3.2	TpRBCS.fwd	TAAAGGCTGGGCTATGAACG
22	3.3.2	TpRBCS.rv	AACCTGGCTCATTAGCTGGA
23	3.3.2	TpTHIC.fwd	GAGAAGGTTGATGGCATTGC
24	3.3.2	TpTHIC.rv	TCCATTTGGCATGGATAGATC
25	3.3.2	TpNMT1.fwd	GCGTTTGGATGTAGCAGTGAC
26	3.3.2	TpNMT1.rv	TATCTGCTGGTCGCTTGATG
27	3.3.2	TpSSSP.fwd	AGGTAAAGAAGGAGGCTGGATC
28	3.3.2	TpSSSP.rv	CCAAGAGAGGTTAGGATTAGCACC
29	3.3.2	TpTHIG.fwd	AATACAGCTGGATCACAAACAGC
30	3.3.2	TpTHIG.rv	AGCTTTTAATGTACCTATTGGATCAGG
31	4.3	THIC3U.fwd	CAACAAGCTCTACGTGGAAGATG
32	4.3	HATag.rv	TAGTCGGGGACGTCGTAGG
33	4.6	Junction.fwd	GAAGCTGATCTGCACCACC
34	4.6	Junction.rv	CGGGTCTTGTAGTTGCCG
35	4.6	Venus.fwd	AGCGCACCATCTTCTTCAAG
36	4.6	Venus.rv	CGTCCTCCTTGAAGTCGATG

8.2.2. 3'RACE primers

#	Section	Primer	Sequence
37	3.3.3	TVN.rv	TTGAAGACGT**AGCG**TAATGAACATAGTTTTTTTTTTTTVN
38	3.3.3	UAdp.rv	TTGAAGACGT**AGCG**TAATGAACATAG
39	3.3.3	THIC71.fwd	GAGCTTTACAAGGAACTGGGCAACAAG
40	3.3.3	THIC3U.fwd	CAACAAGCTCTACGTGGAAGATG

8.2.3. Cloning primers

#	Section	Primer	Sequence

41	4.2; 4.3	THIC3UTR.fwd	AAGAAGACTT**GCTT**GAAATTGGAAAGAATAGTGGTGG
42	4.2	THIC217.rv	AAGAAGACTT**AGCG**TTGTTCCCTTCGCGTG
43	4.2	THIC256.rv	CCGAAGACGT**AGCG**GTAAACCAAGTATTATTATTATTGTCTACCG
44	4.2; 4.3	THIC815.rv	AAGAAGACTT**AGCG**TTCCGGTAGTGCGGG
45	4.2; 4.3	THICP5U.fwd	AAGAAGACTT**GGAG**TTGGTCGACTTCACTCAAAGG
46	4.2; 4.3	THICP5U.rv	AAGAAGACTT**CATT**TGCTGCAACTGTAACCGC
47	4.3	THICCDS.fwd	TTGAAGACAT**AATG**GCTGGGTGCCATTC
48	4.3	THICm1.rv	TTGAAGACTG**TCCT**CAATCTGAAAGCGTAAGTGG
49	4.3	THICm1.fwd	TTGAAGACAG**AGGA**CGAGGTGGAAAAGTTGC
50	4.3	THICm2.rv	TTGAAGACTA**CCTC**GGCAATCCCGTCC
51	4.3	THICm2.fwd	TTGAAGACTA**GAGG**ACTTGACGTGGGAATGTTTC
52	4.3	THICm3.rv	TTGAAGACAT**GTAG**ACCAAGATGCTCCTTTGG
53	4.3	THICm3.fwd	TTGAAGACAT**CTAC**CCAACCGCGACG
54	4.3	THICCDS.rv	TTGAAGACAT**AAGC**CTAGGACGCTAGATCTTTCAAAGG
55	4.3	THICCDSTag.rv	TTGAAGACAT**CGAA**CTGGACGCTAGATCTTTCAAAGG
56	4.3	THICAptm.fwd	TTGAAGACAT**ACAT**TCGTTGTAGCTCTCTTTTGGACACAC
57	4.3	THICAptm.rv	TTGAAGACAT**ATGT**TGAGCTCCCGCAC
58	4.5; 4.8	FCPAP.fwd	AAGAAGACTT**GGAG**GGGCTGCAGGAC
59	4.5	FCPAP.rv	AAGAAGACTT**AGTA**TCGAAACGGCAGACAAATTTG
60	4.5	CrTHI45U.fwd	TTGAAGACAA**TACT**GAGCTGTCGCATAGATCG
61	4.5	CrTHI45U.rv	TTGAAGACAA**CATT**GACTTTGGTGAAGAAGG
62	4.6	BleCDS.fwd	AAGAAGACTT**AATG**GCCAAGTTGACCTCC
63	4.6	VenusIn.rv	AAGAAGACTT**ACCG**TCGTCCTTGAAGAAGATG
64	4.6	VenusIn.fwd	AAGAAGACGT**AGGC**AACTACAAGACCCGC

65	4.6	VenusCDS.rv	TTGAAGACTT**AAGC**TTACTTGTACAGCTCGTCCAT
66	4.6	N78i.fwd	TTGAAGACTT**CGGT**GCGTCTCTCAATCG
67	4.6	N78i.rv	CGTCGAAGACTT**GCCT**ACAGTAAAAAGAGATAATCGTTG
68	4.6	CrRBCS2i.fwd	AAGAAGACTT**CGGT**GAGTCGACGAGCAAGCC
69	4.6	CrRBCS2i.rv	AAGAAGACTT**GCCT**GCAAATGGAAACGGC
70	4.6	CrTHICi.fwd	AAGAAGACTT**CGGT**AAGTGCCGTACAGCTGC
71	4.6	CrTHICi.rv	AAGAAGACTT**GCCT**GGGGTGGAGAGC
72	4.6	CrTHIC1i.rv	AAGAAGACTT**GCCT**AGTGTCCCGAGAGTCCG
73	4.6	CrTHICmini.rv	AAGAAGACTT**CGCA**ACAGCAATGATGCAT
74	4.6	CrTHICmini.fwd	TTGAAGACAA**TGCG**GCTGATGCTGCCTGCAAGTG
75	4.7	TC4N.fwd	AAGAAGACTT**TACT**GAGCTGTCGCATAGATCG
76	4.7	TC4N.rv	AAGAAGACTT**CATT**GACTTTGGTGAAGAAGGTTC
77	4.7	PtP1/2.rv	AAGAAGACTT**CAAT**CCCGAGGCCGTGC
78	4.7	CrP1/P2.rv	AAGAAGACAC**CCCT**AGTGTCCCGAGG
79	4.7	CrP2.rv	AAGAAGACAC**CCCT**ACAATCCCGAGGC
80	4.7	CrP1up.rv	AAGAAGACTT**CCGC**AGTGTCCCGAGGCCGT
81	4.7	PtP1.fwd	AAGAAGACTT**ACAA**GCCCCGGTGCCTG
82	4.7	CrP1.fwd	AAGAAGACTT**GGGA**AGTGGCCCCGG
83	4.7	PtP1.rv	AAGAAGACTT**TTGT**TCCCTTCGCGTGA
84	4.7	PtP1/2.fwd	AAGAAGACTT**ATTG**TGCGGGAGCTCAAC
85	4.7	CrP2.fwd	AAGAAGACTT**AGGG**GAGCTCAACATTCGTTGTAG
86	4.7	CrP1.rv	AAGAAGACTT**TCCC**TTCGCGTGAATTATC
87	4.7	CrP1up.fwd	AAGAAGACTT**GCGG**GAGCTCAACATTC
88	4.7	CrP2up.rv	AAGAAGACTT**TCCA**AGGGTCTTGGGCCTTTAC

89	4.7	CrP2up.fwd	AAGAAGACTT**TGGA**ACCTGATCGTGATAATTCA
90	4.7	PtTHICP3a.rv	AAGAAGACTT**AATC**TCAGCTACAACGAATGTTGAG
91	4.7	PtTHICP3a.fwd	AAGAAGACTT**GATT**AGACCGTTGGAACCTGATCG
92	4.7	PtTHICP3a2.fwd	AAGAAGACTT**GATT**AGACCCTTGGAACCTGATCG
93	4.7	CrTHICP4/5.fwd	AAGAAGACTT**GATT**AGACCGTTGGAACTGCAGCAGGTTAG GACCTGCGTCAGAAGTCTTCTT
94	4.7	CrTHICP4/5.rv	AAGAAGACTT**CTGA**CGCAGGTCCTAACCTGCTGCAGTTCC AACGGTCTAATCAAGTCTTCTT
95	4.7	CrTHICP4/5a.fwd	AAGAAGACTT**GATT**AGACCCTTGGAACTGCAGCAGGTTAG GACCTGCGTCAGAAGTCTTCTT
96	4.7	CrTHICP4/5a.rv	AAGAAGACTT**CTGA**CGCAGGTCCTAACCTGCTGCAGTTCC AAGGGTCTAATCAAGTCTTCTT
97	4.7	P4/5Pt.fwd	AAGAAGACTT**TCAG**GAACAAGCCCCGGTG
98	4.7	P4/5Cr.fwd	AAGAAGACTT**TCAG**GAAGTGGCCCCGGT
99	4.7	CrP4.rv	AAGAAGACTT**CTGA**CGCAGGTCCTAACCTGCTGCAGTTCC AAGGGTCTTGGGC
100	4.7	PtP4.rv	AAGAAGACTT**CTGA**CGCAGGTCCTAACCTGCTGCAGTTCC AACGGTCTTGGGC
101	4.8	METEm1.fwd	AAGAAGACTT**GGAG**ACTTCCGTTTGCTGAATGAAG
102	4.8	METEm1.rv	AAGAAGACTT**TGAT**ATCCTTGGACGGCAG
103	4.8	METEm2.fwd	AAGAAGACTT**ATCA**ATCTCACATACGGCAAGAAATGTAGT TTCAGAGGGCTGAG
104	4.8	METEm2.rv	AAGAAGACTT**GATA**TTGAGCTCCCCGAAATG
105	4.8	METEm3.fwd	AAGAAGACTT**TATC**AGTTATTCTACAGCGTCACATAGACCA GTGAATGAACTGTAATGC
106	4.8	METEm3.rv	AAGAAGACTT**CATT**TGTTGATCGAGATAGCTTTTTGG
107	4.8	FCPAm.rv	AAGAAGACTT**CATT**TCGAAACGGCAGACAAATTTGTG

108	4.8	FCPAm2.fwd	AAGAAGACTT**GGAG**CAAAGCACGTTCTACGATCCGACAAC CTTGTCCAAAGCACGTTCTTCGGGCTGCAGGACGCAATG
109	4.8	FCPAm.fwd	AAGAAGACTT**GGAG**CAAAGCACGTTCTTCGGGCTGCAGG ACGCAATG
110	4.8	FCPA2scr.fwd	AAGAAGACTT**GGAG**ATCTCACATACGGCAGATCCGACAAC CTTGTCATTCTACAGCGTCACGGGCTGCAGGACGCAATG
111	5.2	sgUMPS1.1.fwd	AAGGTCTCA**CGAG**GCACAAGAGAGAGTTGACTGGTTTTAG AGCTAGAAATAGCAAG
112	5.2	sgUMPS1.2.fwd	AAGGTCTCA**CGAG**GAGAAAGAGCTGTTTGCAGAGTTTTAG AGCTAGAAATAGCAAG
113	5.2	sgUMPS2.1.fwd	AAGGTCTCA**CGAG**GACACCTGGATTCTAGCACCGTTTTAG AGCTAGAAATAGCAAG
114	5.2	sgUMPS2.2.fwd	AAGGTCTCAT**CGAG**GTTGCTGATACCCCTAGACACGTTTTA GAGCTAGAAATAGCAAG
115	5.2; 5.3	sgRNA.rv	TGGTCTCA**AGC**GTAATGCCAACTTTGTACAAG
116	5.2	5'UMPS_HR.fwd	AAGGTCTCA**GGAG**GCAAGGGCATCAAAGTGC
117	5.2	5'UMPS_HR.rv	AAGGTCTCA**AGCG**TTTGATCCGGGAATCTTGTG
118	5.2	3'UMPS_HR.fwd	AAGGTCTCA**GGAG**ATACTCTGGCGAGGTTTTCCAC
119	5.2	3'UMPS_HR.rv	AAGGTCTCA**AGCG**TGCCACTTCCAGGTGGAC
120	5.3	sgTHIC.1.fwd	AAGGTCTCA**CGAG**GATCTTGGGGATCCTTCAAGAGTTTTA GAGCTAGAAATAGCAAG
121	5.3	sgTHIC.2.fwd	AAGGTCTCA**CGAG**GTGGATTTGTACGATACCTCGGTTTTA GAGCTAGAAATAGCAAG
122	5.3	sgNMT1.1.fwd	AAGGTCTCA**CGAG**GTTTCCCAAGATGGATTGGTGGTTTTA GAGCTAGAAATAGCAAG
123	5.3	sgNMT1.2.fwd	AAGGTCTCA**CGAG**GCTCCTCCTTCCTCGTCACGGGTTTTAG AGCTAGAAATAGCAAG

124	5.3	sgSSSP.1.fwd	AAGGTCTCA**CGAG**GTGACGGATAGGTTTCAGGGGTTTTAGAGCTAGAAATAGCAAG
125	5.3	sgSSSP.2.fwd	AAGGTCTCA**CGAG**GCATTATGAGAAAGACCATGGGTTTTAGAGCTAGAAATAGCAAG
126	5.3	sgMETE.1.fwd	AAGGTCTCA**CGAG**GCCAAATTCTCGAAGCGATGGTTTTAGAGCTAGAAATAGCAAG
127	5.3	sgMETE.2.fwd	AAGGTCTCA**CGAG**GCGTGTCGAATGGATAATATGGTTTTAGAGCTAGAAATAGCAAG
128	5.3	sgMETH.1.fwd	AAGGTCTCA**CGAG**GCGTAGTCAAACTCGACCCGGGTTTTAGAGCTAGAAATAGCAAG
129	5.3	sgMETH.2.fwd	AAGGTCTCA**CGAG**GAATGTCGTTCCGCTGCGCGGGTTTTAGAGCTAGAAATAGCAAG
130	5.3	sgCBA1.1.fwd	AAGGTCTCA**CGAG**GACCTACCTCCTCTACCAGTGGTTTTAGAGCTAGAAATAGCAAG
131	5.3	sgCBA1.2.fwd	AAGGTCTCA**CGAG**GCGTTTGCGAGAGACCCACGTGTTTTAGAGCTAGAAATAGCAAG
132	5.3	sgGGPPS.1.fwd	AAGGTCTCA**CGAG**GACCTGTTTGAAACAGGACAGTTTTAGAGCTAGAAATAGCAAG
133	5.3	sgGGPPS.2.fwd	AAGGTCTCA**CGAG**GTCCCGGGAAAGGATGTACGGTTTTAGAGCTAGAAATAGCAAG
134	5.3	sgPSY.1.fwd	AAGGTCTCA**CGAG**GAGGTGCGAAGAGACGAGACGGTTTTAGAGCTAGAAATAGCAAG
135	5.3	sgPSY.2.fwd	AAGGTCTCA**CGAG**GCAACAAGATGCATTTCTCGGGTTTTAGAGCTAGAAATAGCAAG
136	5.3	5'THIC.HR.fwd	AAGGTCTCA**GGAG**TGGTCGACTTCACTCAAAG
137	5.3	5'THIC.HR.rv	AAGGTCTCA**AGCG**CCTTCTTGGGATCAACG
138	5.3	3'THIC.HR.fwd	AAGGTCTCA**GGAG**AGGTAAGCGCAAAGGTC
139	5.3	3'THIC.HR.rv	AAGGTCTCA**AGCG**GTGTTTCCATCGAGAGTC

140	5.3	5'NMT1.HR.fwd	AAGGTCTCA**GGAG**AAAACGGGAAGTCGAG
141	5.3	5'NMT1.HR.rv	AAGGTCTCA**AGCG**ACTGTGAGGTAGAAAGAAAGG
142	5.3	3'NMT1.HR.fwd	AAGGTCTCA**GGAG**TGAAGTTGCCTGGTAGACTG
143	5.3	3'NMT1.HR.rv	AAGGTCTCA**AGCG**CCGAATACCGTAAGAAGG
144	5.3	5'SSSP.HR.fwd	AAGGTCTCA**GGAG**TATCGTGGACTCGGTCC
145	5.3	5'SSSP.HR.rv	AAGGTCTCA**AGCG**ACTGATCAGGAGAATATTGC
146	5.3	3'SSSP.HR.fwd	AAGGTCTCA**GGAG**AAGATCCGGAATCCTACG
147	5.3	3'SSSP.HR.rv	AAGGTCTCA**AGCG**TGCTTCGGTTCGAATG
148	5.3	5'METE.HR.fwd	AAGGTCTCA**GGAG**CCGAATTACTCCGAATAACTG
149	5.3	5'METE.HR.rv	AAGGTCTCA**AGCG**TGATAAGATCCACTTCACTCAG
150	5.3	3'METE.HR.fwd	AAGGTCTCA**GGAG**CCAGCAATTATGTAGACG
151	5.3	3'METE.HR.rv	AAGGTCTCA**AGCG**ATTTGGTGAGCTCCTTG
152	5.3	5'METH.HR.fwd	AAGGTCTCA**GGAG**CCAACTATAGGATTCTTTGCTC
153	5.3	5'METH.HR.rv	AAGGTCTCA**AGCG**ACGTTCGGGTTGTATG
154	5.3	3'METH.HR.fwd	AAGGTCTCA**GGAG**ACGAGATGGTTGATGTTG
155	5.3	3'METH.HR.rv	AAGGTCTCA**AGCG**AACTTCTTCACTAGAGCG
156	5.3	5'CBA1.HR.fwd	AAGGTCTCA**GGAG**TCGACCACCACTTATTCC
157	5.3	5'CBA1.HR.rv	AAGGTCTCA**AGCG**TTCGGAGTATCCTGATGG
158	5.3	3'CBA1.HR.fwd	AAGGTCTCA**GGAG**GACTACTCAGTCTTATACACGTATTG
159	5.3	3'CBA1.HR.rv	AAGGTCTCA**AGCG**ACACAGAAATGATGCCTC
160	5.3	5'GGPPS.HR.fwd	AAGGTCTCA**GGAG**TCCTTGTTCTCGATCTGAG
161	5.3	5'GGPPS.HR.rv	AAGGTCTCA**AGCG**ATGATTGACAGAAGCTC
162	5.3	5'PSY.HR.fwd	AAGGTCTCA**GGAG**CAAACTTTACGAAAGGC
163	5.3	5'PSY.HR.rv	AAGGTCTCA**AGCG**AGAACTTCCGCGACTTG

164	5.3	GGPPSCDS.fwd	AAGGTCTCA**AATG**TCTTCTCAGGAGACAAAGAAAG
165	5.3	GGPPSCDS.rv	AAGGTCTCA**AAGC**TTATGCTTCGTCAATCTCTACC
166	5.3	PSYCDS.fwd	AAGGTCTCA**AATG**AAAGTTTCGACAAAGCTC
167	5.3	PSYCDS.rv	AAGGTCTCA**AAGC**TCATACTTGATCCAATTGGAC
168	5.3	L1.CFP.fwd	AAGGTCTCT**GGAG**GGGCTGCAG
169	5.3	L1.CFP.rv	AAGGTCTCA**AGCG**ACTAACACACCTGG

8.2.4. Genotyping and sequencing primers

#	Section	Primer	Sequence
170	Na	L0seq.fwd	AATACGCAAACCGCCTCTC
171	Na	L0seq.rv	AAATAGGCGTATCACGAGG
172	Na	L1seq.fwd	ATGCACATACAAATGGACG
173	Na	L1seq.rv	TTAATAACACATTGCGGACG
174	4.5	Luc.fwd	CCCATGGAGCAGTTCATC
175	4.5	Luc.rv	CCTTGATCTTGTCCACCTG
176	Chap. 5	NAT.rv	AGTGAACACGACGCTGAAGG
177	Chap. 5	NAT.fwd	GAGGTCACCAACGTCAACG
178	5.2	UMPSout.fwd	TTAGAAGACCCGGTCGAAGG
179	5.2	UMPSout.rv	GGTCGAGGCCGTAAGAATAGG
180	5.2	UMPS.SNP.G.fwd	TTTCTCTGTCACTGTCAAGGTGAG
181	5.2	UMPS.SNP.C.fwd	TTTCTCTGTCACTGTCAAGGTGAC
182	5.2	gUMPSin.fwd	CTCTACGACAAATTCACAAGATTCC
183	5.2	gUMPSin.rv	GCTTTCTGAATCCTCTCCTGC
184	5.3; 5.4	gTHIC.fwd	TGTAGGGCATCTCTGGATGG

185	5.3; 5.4	gTHIC.rv	CGTGGGAGTCTCGTTTCG
186	5.3	gNMT1.fwd	GTTGTTTCCACAATGACGTGC
187	5.3	gNMT1.rv	AATCACGGTATTGGAAACAGAGC
188	5.3; 5.6	gSSSP.fwd	AACCGACAGCATTCGTATGG
189	5.3; 5.6	gSSSP.rv	TTGATGTGCCTCCTCGAGAC
190	5.3; 5.5	gMETE.fwd	TATCCGCGAACGACCTACAG
191	5.3; 5.5	gMETE.rv	GAACAGCGCGACTTTTTGG
192	5.3	gMETH.fwd	TTGTACAGGCGTGCAAGTTG
193	5.3	gMETH.rv	GAAACGATCCGCCAAAGC
194	5.3; 5.7	gCBA1.fwd	GTTTCCCCCAAGCCTTTG
195	5.3; 5.7	gCBA1.rv	CAGCAAGGACGCTATTCAGG
196	5.3	gGGPPS.fwd	ACGGCCTGCTTTATTGTGG
197	5.3	gGGPPS.rv	AGCTACCCGAAGATGGTTGG
198	5.4	gTHICin.fwd	CCATGGTTAACGCCTTTCGTAG
199	5.4	gTHICin.rv	CGTCCGTCCATTCCTGG
200	5.5	gMETEin.fwd	CACAGTGGGTGATATGTACCTCTACG
201	5.5	gMETEin.rv	CATTTTTTCATGCCTATTCACATTAGC
202	5.6	gSSSPin.fwd	CGGTATTGGTATTACTATTGCGATTG
203	5.6	gSSSPin.rv	CGTAGGATTCCGGATCTTTGG
204	5.7	gCBA1in.fwd	CGCTCTTCTCCCAAGGATG
205	5.7	gCBAin.rv	GATCTCGTCCAAGAAGCAAGG

8.3. Plasmids

8.3.1. Level 0

Plasmid	Description	5’ Border	3’ Border	Storage

pMLP0004	PtFCPA promoter	**GGAG**	TACT	Box1_A1
pMLP0101	PtFCPA promoter	**GGAG**	**AATG**	Box1_A2
pMLP0106	PtTHIC promoter	**GGAG**	**AATG**	Box1_A3
pMLP0107	PtEF2 promoter	**GGAG**	**AATG**	-
pMLP0111	PtMETE359 promoter	**GGAG**	**AATG**	Glycerol Stock, -80ºC, labelled pMLP0211
pMLP0112	PtMETE359scram prom	**GGAG**	**AATG**	Glycerol Stock, -80ºC, labelled pMLP0212
pMLP0113	FCPA_2METEmotifs prom	**GGAG**	**AATG**	Glycerol Stock, -80ºC, labelled pMLP0213
pMLP0114	FCPA_1METEmotif prom	**GGAG**	**AATG**	Glycerol Stock, -80ºC, labelled pMLP0214
pMLP0115	FCPA_2METEscrammotifs prom	**GGAG**	**AATG**	Glycerol Stock, -80ºC, labelled pMLP0215
pMLP0201	CrTHI4 5'UTR	TACT	**AATG**	Box1_A4
pMLP0202	CrTC4N 5'UTR	TACT	**AATG**	Box1_A5
pMLP0203	PtTC4N 5'UTR	TACT	**AATG**	Box1_A6
pMLP0204	PtTC4N_CrP2 5'UTR	TACT	**AATG**	Box1_A7
pMLP0205	PtTC4N_CrP1 5'UTR	TACT	**AATG**	Box1_A8
pMLP0206	PtTC4N_CrP1/2 5'UTR	TACT	**AATG**	Box1_A9
pMLP0207	CrTC4N 5'UTR	TACT	**AATG**	Box1_A10
pMLP0208	PtTC4N_NoP3a 5'UTR	TACT	**AATG**	Box1_B1
pMLP0209	PtTC4N_NoP3a_CrP1/2 5'UTR	TACT	**AATG**	Box1_B2
pMLP0210	PtTC4N_NoP3a_CrP4/5 5' UTR	TACT	**AATG**	Box1_B3
pMLP0211	PtTC4N_NoP3a_CrP1/2_CrP4/5	TACT	**AATG**	Box1_B4
pMLP0212	PtTC4N_CrP4/5 5'UTR	TACT	**AATG**	Box1_B5

pMLP0213	PtTC4N_CrP1/2_CrP4/5 5'UTR	TACT	**AATG**	Box1_B6
pMLP0401	Luc CDS	**AATG**	**GCTT**	Box1_B7
pMLP0403	Venus CDS	**AATG**	**GCTT**	Box1_B8
pMLP0405	BleVenus_PtN78i CDS	**AATG**	**GCTT**	Box1_B9
pMLP0406	BleVenus_CrRBCS2i	**AATG**	**GCTT**	Box1_B10
pMLP0407	BleVenus_CrTHICi	**AATG**	**GCTT**	Box1_C1
pMLP0408	BleVenus_CrTHIC1i	**AATG**	**GCTT**	Box1_C2
pMLP0409	BleVenus_CrTHIC2i	**AATG**	**GCTT**	Box1_C3
pMLP0410	BleVenus_CrTHICimini	**AATG**	**GCTT**	Box1_C4
pMLP0411	BleVenus CDS	**AATG**	**GCTT**	Box1_C5
pMLP0412	PtTHIC CDS	**AATG**	**GCTT**	Glycerol Stock, -80ºC, labelled pJD_A
pMPL0413	PtTHIC CDS	**AATG**	TTCG	Glycerol Stock, -80ºC, labelled pJD_B
pMLP0414	mTurquoise CDS	**AATG**	**GCTT**	-
pMLP0503	PtTHIC217 3UTR	**GCTT**	GGTA	Box1_C6
pMLP0504	PtTHIC256 3UTR	**GCTT**	GGTA	Box1_C7
pMLP0505	PtTHIC815 3UTR	**GCTT**	GGTA	Box1_C8
pMLP0601	PtFCPC Term	**GCTT**	CGCT	Box1_C9
pMLP0606	PtTHIC815 Term	**GCTT**	CGCT	Box1_C10
pMLP0607	PtTHiC815m Term	**GCTT**	CGCT	-
pMLP0701	PtFCPC Term	GGTA	CGCT	Box1_D2
pKG141	Ble-eGFP CDS	**AATG**	**GCTT**	Katrin Geisler
pKG270	HA-Tag	TTCG	**GCTT**	Katrin Geisler
pPM11	CrCA1 terminator	**GCTT**	CGCT	Payam Mehrshahi

pPM19	CrPSAD promoter	**GGAG**	**AATG**	Payam Mehrshahi

8.3.2. Level 1

Plasmid	Position	Description	Includes L0s	Storage
pMLP1101	1	5'hrUMPS	-	Box1_D3
pMLP1102	1	5'hrTHIC	-	Box1_D4
pMLP1103	1	5'hrNMT1	-	Box1_D5
pMLP1104	1	5'hrSSSP	-	Box1_D6
pMLP1105	1	5'hrMETE	-	Box1_D7
pMLP1106	1	5'hrMETH	-	Box1_D8
pMLP1107	1	5'hrCBA1	-	Box1_D9
pMLP1108	1	5'hrGGPPS	-	Box1_D10
pMLP1109	1	5'hrPSY	-	Box1_D1
pMLP1111	1	FCPA_luc_FCPC	pMLP0101, pMLP0401, pMLP0601	Box1_E1
pMLP1116	1	FCPA_luc_THIC217_FCPC	pMLP0101, pMLP0401, pMLP0503, pMLP0701	Box1_E2
pMLP1117	1	FCPA_luc_THIC256_FCPC	pMLP0101, pMLP0401, pMLP0504, pMLP0701	Box1_E3
pMLP1119	1	FCPA_luc_THIC815_FCPC	pMLP0101, pMLP0401, pMLP0506, pMLP0701	Box1_E4
pMLP1211	2	FCPA_CrTHI4_luc_FCPC	pMLP0004, pMLP0201, pMLP0401, pMLP0601	Box1_E5
pMLP1212	2	FCPA_CrTC4N_luc_FCPC	pMLP0004, pMLP0202, pMLP0401, pMLP0601	Box1_E6

pMLP1213	2	EF2_BleVenus_PtN78i_FCPC	pMLP0107, pMLP0601	pMLP0405,	Glycerol Stock, -80ºC,
pMLP1214	2	EF2_BleVenus_CrRBCS2i_FCPC	pMLP0107, pMLP0601	pMLP0406,	Glycerol Stock, -80ºC,
pMLP1215	2	EF2_BleVenus_CrTHICi_FCPC	pMLP0107, pMLP0601	pMLP0407,	Glycerol Stock, -80ºC,
pMLP1216	2	EF2_BleVenus_CrTHIC1i_FCPC	pMLP0107, pMLP0601	pMLP0408,	Glycerol Stock, -80ºC,
pMLP1217	2	EF2_BleVenus_CrTHIC2i_FCPC	pMLP0107, pMLP0601	pMLP0409,	Glycerol Stock, -80ºC,
pMLP1218	2	EF2_BleVenus_CrTHICimini _FCPC	pMLP0107, pMLP0601	pMLP0410,	Glycerol Stock, -80ºC,
pMLP1219	2	EF2_BleVenus_FCPC	pMLP0107, pMLP0601	pMLP0411,	Glycerol Stock, -80ºC,
pMLP1220	2	THIC_BleVenus_THIC815	pMLP0106, pMLP0606	pMLP0411,	Glycerol Stock, -80ºC,
pMLP1221	**2**	CRPSAD_PtTC4N_BleGFP_CA1	pPM19, pKG141, pPM11	pMLP0203, pKG270,	Box1_E7
pMLP1222	**2**	CRPSAD_PtTC4N_CrP2 _BleGFP_CA1	pPM19, pKG141, pPM11	pMLP0204, pKG270,	Box1_E8
pMLP1223	**2**	CRPSAD_PtTC4N_CrP1 BleGFP_CA1	pPM19, pKG141, pPM11	pMLP0205, pKG270,	Box1_E9
pMLP1224	**2**	CRPSAD_PtTC4N_CrP1/2 BleGFP_CA1	pPM19, pKG141, pPM11	pMLP0206, pKG270,	Box1_E10

pMLP1225	**2**	CRPSAD_CrTC4N_BleGFP_CA1	pPM19, pKG141, pPM11	pMLP0207, pKG270,	Box1_F1
pMLP1226	**2**	CRPSAD_PtTC4N_NoP3a BleGFP_CA1	pPM19, pKG141, pPM11	pMLP0208, pKG270,	Box1_F2
pMLP1227	**2**	CRPSAD_PtTC4N_NoP3a _CrP1/2_BleGFP_CA1	pPM19, pKG141, pPM11	pMLP0209, pKG270,	Box1_F3
pMLP1228	**2**	CRPSAD_PtTC4N_NoP3a _CrP4/5_BleGFP_CA1	pPM19, pKG141, pPM11	pMLP0210, pKG270,	Box1_F4
pMLP1229	**2**	CRPSAD_PtTC4N_NoP3a_CrP1/ 2 _CrP4/5_BleGFP_CA1	pPM19, pKG141, pPM11	pMLP0211, pKG270,	Box1_F5
pMLP1230	**2**	CRPSAD_PtTC4N_CrP4/5 _BleGFP_CA1	pPM19, pKG141, pPM11	pMLP0212, pKG270,	Box1_F6
pMLP1231	**2**	CRPSAD_PtTC4N_CrP1/2 _CrP4/5_BleGFP_CA1	pPM19, pKG141, pPM11	pMLP0213, pKG270,	Box1_F7
pMLP1241	**2**	THIC_THIC_THIC	pMLP0106, pMLP0606	pMLP0412,	Box1_F8
pMLP1242	**2**	THIC_THIC_THICm	pMLP0106, pMLP0607	pMLP0412,	Box1_F9
pMLP1243	**2**	THIC_THIC_FCPC	pMLP0106, pMLP0601	pMLP0412,	Box1_F10
pMLP1244	**2**	THIC_THIC_HaTag_THIC	pMLP0106, pKG270, pMLP0606	pMLP0413,	Box1_G1

pMLP1245	**2**	THIC_THIC_HaTag_THICm	pMLP0106, pMLP0413, pKG270, pMLP0607	Box1_G2
pMLP1246	**2**	THIC_THIC_HaTag_FCPC	pMLP0106, pMLP0413, pKG270, pMLP0601	Box1_G3
pMLP1301	**3**	3'hrUMPS	-	Box1_G4
pMLP1302	**3**	3'hrTHIC	-	Box1_G5
pMLP1303	**3**	3'hrNMT1	-	Box1_G6
pMLP1304	**3**	3'hrSSSP	-	Box1_G7
pMLP1305	**3**	3'hrMETE	-	Box1_G8
pMLP1306	**3**	3'hrMETH	-	Box1_G9
pMLP1307	**3**	3'hrCBA1	-	Box1_G10
pMLP1308	**3**	3'hrGGPPS	-	Box1_H1
pMLP1309	**3**	3'hrPSY	-	Box1_H2
pMLP1311	**3**	PtMETE359_Venus62_FCPC	pMLP0111, pMLP0403, pMLP0601	Box1_H3
pMLP1312	**3**	PtMETE359scram_Venus62_ FCPC	pMLP0112, pMLP0403, pMLP0601	Box1_H4
pMLP1313	**3**	FCPA_2METEmotifs_Venus62_ FCPC	pMLP0113, pMLP0403, pMLP0601	Box1_H5
pMLP1314	**3**	FCPA_1METEmotif_Venus62_ FCPC	pMLP0114, pMLP0403, pMLP0601	Box1_H6
pMLP1315	**3**	FCPA_2METEscrammotifs_ Venus62_FCPC	pMLP0115, pMLP0403, pMLP0601	Box1_H7
pMLP1316	**3**	FCPA_Venus62_FCPC	pMLP0101, pMLP0403, pMLP0601	Box1_H8
pMLP1321	**3**	5'hrUMPS	-	Box1_H9

pMLP1401	**4**	FCPA_CFP_FCPC	pMLP0101, pMLP0414, pMLP0601	Box1_H10
pMLP1411	**4**	U6_sgUMPS1.1	-	Box1_I1
pMLP1412	**4**	U6_sgUMPS2.1	-	Box1_I2
pMLP1501	**5**	U6_sgUMPS_1	-	Box1_I3
pMLP1502	**5**	U6_sgTHIC_1	-	Box1_I4
pMLP1503	**5**	U6_sgNMT1_1	-	Box1_I5
pMLP1504	**5**	U6_sgSSSP_1	-	Box1_I6
pMLP1505	**5**	U6_sgMETE_1	-	Box1_I7
pMLP1506	**5**	U6_sgMETH_1	-	Box1_I8
pMLP1507	**5**	U6_sgCBA1_1	-	Box1_I9
pMLP1508	**5**	U6_sgGGPPS_1	-	Box1_I10
pMLP1509	**5**	U6_sgPSY_1	-	Box1_J1
pMLP1511	**5**	U6_sgUMPS1.2	-	Box1_J2
pMLP1512	**5**	U6_sgUMPS2.2	-	Box1_J3
pMLP1521	**5**	3'hrUMPS	-	Box1_J4
pMLP1601	**6**	U6_sgUMPS_2	-	Box1_J5
pMLP1602	**6**	U6_sgTHIC_2	-	Box1_J6
pMLP1603	**6**	U6_sgNMT1_2	-	Box1_J7
pMLP1604	**6**	U6_sgSSSP_2	-	Box1_J8
pMLP1605	**6**	U6_sgMETE_2	-	Box1_J9
pMLP1606	**6**	U6_sgMETH_2	-	Box1_J10
pMLP1607	**6**	U6_sgCBA1_2	-	Box2_A1
pMLP1608	**6**	U6_sgGGPPS_2	-	Box2_A2
pMLP1609	**6**	U6_sgPSY_2	-	Box2_A3

pAKS010	**2**	FCPB_Ble_FCPC	-	Katrin Geisler
pKG018	**1**	OriT episomal part	-	Katrin Geisler
pKG023	**6**	Cen-Ars-His episomal part	-	Katrin Geisler
pKG025	**3**	FCPB_Ble_FCPC	-	Katrin Geisler
pKG425	**2**	FCPB_NAT_FCPC	-	Katrin Geisler
pKG234	**4**	FCPB_NAT_FCPC	-	Katrin Geisler
Cas9	**2**	FCPB_Cas9-Venus_FCPC	-	Box2_A4
sgRNA template			-	Box2_A5
PtU6 promoter			-	Box2_A6

8.3.3. Level 2

Plasmid	Description	Includes L1s	Storage
pMLP2011	FCPA_luc_FCPC	pMLP1111, pAKS010	Glycerol Stock, -80ºC,
pMLP2021	FCPA_luc_THIC217	pMLP1116, pAKS010	Glycerol Stock, -80ºC,
pMLP2022	FCPA_luc_THIC256	pMLP1117, pAKS010	Glycerol Stock, -80ºC,
pMLP2023	FCPA_luc_THIC815	pMLP1119, pAKS010	Glycerol Stock, -80ºC,
pMLP2031	FCPA_CrTHI4_luc_FCPC	pMLP1211, pKG025	Box2_A7
pMLP2032	FCPA_CrTC4N_luc_FCPC	pMLP1212, pKG025	Box2_A8
pMLP2041	EF2_BleVenus_CrRBCS2i_FCPC	pMLP1214, pKG025	Glycerol Stock, -80ºC,
pMLP2042	EF2_BleVenus_CrTHICi_FCPC	pMLP1215, pKG025	Glycerol Stock, -80ºC,

pMLP2043	EF2_BleVenus_CrTHIC1i_FCPC	pMLP1216, pKG025	Glycerol Stock, -80ºC,
pMLP2044	EF2_BleVenus_CrTHIC2i_FCPC	pMLP1217, pKG025	Glycerol Stock, -80ºC,
pMLP2045	EF2_BleVenus_CrTHICimini_FCPC	pMLP1218, pKG025	Glycerol Stock, -80ºC,
pMLP2046	EF2_BleVenurs_PtN78i_FCPC	pMLP1213, pKG025	Glycerol Stock, -80ºC,
pMLP2047	EF2_BleVenus_FCPC	pMLP1219, pKG025	Glycerol Stock, -80ºC,
pMLP2048	THIC_BleVenus_THIC815	pMLP1220, pKG025	Glycerol Stock, -80ºC,
pMLP2061	THIC_THIC_THIC	pMLP1241, pKG018, pKG023, pKG025	Box2_A9
pMLP2062	THIC_THIC_THICm	pMLP1242, pKG018, pKG023, pKG025	Box2_A10
pMLP2063	THIC_THIC_FCPC	pMLP1243, pKG018, pKG023, pKG025	Box2_B1
pMLP2064	THIC_THIC_HaTag_THIC	pMLP1244, pKG018, pKG023, pKG025	Box2_B2
pMLP2065	THIC_THIC_HaTag_THICm	pMLP1245, pKG018, pKG023, pKG025	Box2_B3
pMLP2066	THIC_THIC_HaTag_FCPC	pMLP1246, pKG018, pKG023, pKG025	Box2_B4
pMLP2071	Cas9_sgUMPS1	Cas9, pKG025, pMLP1411, pMLP1511	Box2_B5
pMLP2072	Cas9_sgUMPS2	Cas9, pKG025, pMLP1412, pMLP1512	Box2_B6

pMLP2073	Conj_Cas9_sgUMPS1	Cas9, pKG018, pKG023, pKG025, pMLP1411, pMLP1511	Box2_B7
pMLP2074	Conj_Cas9_sgUMPS2	Cas9, pKG018, pKG023, pKG025, pMLP1412, pMLP1512	Box2_B8
pMLP2075	hrUMPS_NAT	pMLP1321, pMLP1521, pKG234	Box2_B9
pMLP2111	Cas9_sgUMPS	Cas9, pKG025, pMLP1501, pMLP1601	Katrin Geisler
pMLP2112	Cas9_sgTHIC	Cas9, pKG025, pMLP1502, pMLP1602	Katrin Geisler
pMLP2113	Cas9_sgNMT1	Cas9, pKG025, pMLP1503, pMLP1603	Katrin Geisler
pMLP2114	Cas9_sgSSSP	Cas9, pKG025, pMLP1504, pMLP1604	Katrin Geisler
pMLP2115	Cas9_sgMETE	Cas9, pKG025, pMLP1505, pMLP1605	Katrin Geisler
pMLP2116	Cas9_sgMETH	Cas9, pKG025, pMLP1506, pMLP1606	Katrin Geisler
pMLP2117	Cas9_sgCBA1	Cas9, pKG025, pMLP1507, pMLP1607	Katrin Geisler
pMLP2118	Cas9_sgGGPPS	Cas9, pKG025, pMLP1508, pMLP1608	Katrin Geisler
pMLP2119	Cas9_sgPSY	Cas9, pKG025, pMLP1509, pMLP1609	Katrin Geisler
pMLP2121	hrUMPS_NAT	pMLP1101, pMLP1301, pMLP1401, pKG425	Katrin Geisler
pMLP2122	hrTHIC_NAT	pMLP1102, pMLP1302, pMLP1401, pKG425	Katrin Geisler

pMLP2123	hrNMT1_NAT	pMLP1103, pMLP1401, pKG425	pMLP1303,	Katrin Geisler
pMLP2124	hrSSSP_NAT	pMLP1104, pMLP1401, pKG425	pMLP1304,	Katrin Geisler
pMLP2125	hrMETE_NAT	pMLP1105, pMLP1401, pKG425	pMLP1305,	Katrin Geisler
pMLP2126	hrMETH_NAT	pMLP1106, pMLP1401, pKG425	pMLP1306,	Katrin Geisler
pMLP2127	hrCBA1_NAT	pMLP1107, pMLP1401, pKG425	pMLP1307,	Katrin Geisler
pMLP2128	hrGGPPS_NAT	pMLP1108, pMLP1401, pKG425	pMLP1308,	Katrin Geisler
pMLP2129	hrPSY_NAT	pMLP1109, pMLP1401, pKG425	pMLP1309,	Katrin Geisler
pMLP2131	PtMETE359_Venus62_FCPC	pMLP1311, pAKS010		Box2_C2
pMLP2132	PtMETE359scram_Venus62_FCPC	pMLP1312, pAKS010		Box2_C3
pMLP2133	FCPA_2METEmotifs_Venus62_FCPC	pMLP1313, pAKS010		Box2_C4
pMLP2134	FCPA_1METEmotif_Venus62_FCPC	pMLP1314, pAKS010		Box2_C5
pMLP2135	FCPA_2METEscrammotifs_Venus62_FCPC	pMLP1315, pAKS010		Box2_B10
pMLP2136	FCPA_Venus62_FCPC	pMLP1316, pAKS010		Box2_C1