Preface

This master thesis was performed during the spring of 2021 and was a continuation of the project thesis (TBT4500) in the fall of 2020. Some of the results presented in this report originates from this course. The master thesis was conducted at the Department of Biotechnology and Food Science (IBT) at the Norwegian University of Science and Technology (NTNU).

I am grateful to my supervisors professor Olav Vadstein and research scientist Tore Brembu for the opportunity to write this thesis and for the interesting experiment. Many thanks to Tore Brembu for all the help and guidance during the master and project thesis and for always being available for questions. Your advices have been to great help during the writing of both theses. Your knowledge and commitment is inspiring and I have learnt a lot from you during this time. Thank you to the ACMS group for the interesting presentations and papers presented during the meetings, and for the engaging discussions following the presentations. I want to especially thank PhD candidate Annika Messemer for all the training and help in the lab, as well as the help with the writing and interpreting the results. You have been a great support and an inspiration during this project.

Finally, I would like to thank all my friends and my family for the encouragement, support and love you have given me during these times, as well as my fellow students for creating a great environment during all the writing sessions we have shared. Lastly, many thanks to Kristin Bentzen for the support and guidance. The discussions we have had was to great help, and the time spent in the laboratory with you was a joy.

Trondheim, 2021-24-06

Christine Våge Sjevelås

Abstract

Diatoms is the most abundant and diversified subgroup of phytoplankton, and contributes to about one-fifth of the world's photosynthesis, creating more organic carbon annually than all terrestrial rain forests combined. They can be found in water wherever there is enough light and nutrients. The centric model diatom *Thalassiosira pseudonana* is built up of two halves, each consisting of a valve and its associated girdle bands. The frustule, made of amorphous, hydrated silica consists of layers of three-dimensional structures from nano- to micrometer scale, creating a complex, species-specific pattern of pores. Its morphology depends on different proteins and organic components, such as silaffins, silacidins, silicanins, and long chain polyamines. Because of their highly controlled nanopatterns that are produced cost-efficiently by biological self-assembly in large quantities, they are especially useful for nanotechnological applications. In order to be able to manipulate the cells to create certain structures that can be used for nanotechnological purposes, it is important that we understand the mechanism and the components involved in the synthesis of the frustule completely. Cultivation conditions and genetic manipulation can help optimize desired nanostructural parameters.

This report covers a three-part experiment involving the morphology of the frustule of the diatom *Thalassiosira pseudonana*. Firstly, I attempted to create single and double knock out *T. pseudonana* cell lines of *Tp23191* and *Tp6330* in group IV of the silicanin protein family. This experiment used gene editing with the CRISPR-Cas9 system to induce knock out mutations in target regions adjacent to protospacer adjacent motifs (PAM) sites. In total, six different target sites were chosen, two in each gene (Tp6330PAM1 and -2 and Tp23191PAM2 and -4) for the single knock out mutants, and two PAM sites that were identical in both genes (TP23191-6330PAM1 and -2) for the double knock out mutants. Screening was performed on replated cells that had previously given indications for mutations, but no mutants could be identified and isolated. The second experiment was a localization study, I attempted to tag two genes encoding proteins in the silicanin protein family, Tp23191 in group IV and Tp20931 in group III with the fluorescence protein mTurq. Several cloning strategies were attempted, but none proved to be successful. Lastly, the effects of cadmium and aluminium on the morphology of the frustule was studied. Image analysis of scanning electron microscopy (SEM) images of the frustule of *T. pseudonana* cells growing in ESAW medium with no added metal (control), 25 μg/L aluminium (Al_{25}) and 5 μg/L cadmium ($Cd_{5)}$) showed that the valve, fultoportulae, rimoportulae and cribrum pore diameter, and costae width was significantly different from the control in both samples, except cribrum pore diameter in Al_{25}, and fultoportulae and rimoportulae diameter in Cd_5.

Sammendrag

Kiselalger er den mest utbredte og diversifiserte undergruppen av fytoplankton. De bidrar til omtrent en femtedel av verdens fotosyntese, og produserer årlig mer karbon enn alle verdens regnskoger til sammen. De finnes overalt hvor det er nok lys og næringsstoffer. Den sylindriske modellorganismen *Thalassiosira pseudonana* består av to halvdeler som overlapper hverandre, og hver halvdel er bygd opp av en valve og tilhørende girdle bands. Celleveggen, som kalles en frustule, består av hydrert silica, og er bygd opp av lag med tredimensjonelle strukturer fra nano- til mikrometerskala. Disse strukturene danner et komplekst og artsspesifikt mønster av porer, hvis morfologi avhenger av ulike proteiner of organiske komponenter, slik som silaffiner, silacidiner, silicaniner og langkjededepolyaminer (long chain polyamines). Det komplekse mønsteret av porer produseres kostnadseffektivt ved biologisk selvmontering i store mengder, noe som gjør de spesielt nyttige for nanoteknologisk industri. For å kunne manipulere cellene til a danne visse strukturer som kan brukes til nanotegnologiske formål, er det viktig at vi forstår mekanismen og komponentene som er invovert i syntesen fullstendig. Man kan optimalisere ønskede nanostrukturelle paramtere ved å endre dyrkningsbetingelser eller ved genetisk manipulasjon.

Denne rapporten dekker et tredelt eksperiment som undersøker morfologien til *Thalassiosira pseudonana*s frustule. Først ble det forsøkt å danne enkle og doble knock out mutante *T. pseudonana* cellelinjer av *Tp23191* og *Tp6330* i gruppe IV i silicanin-proteinfamilien. Dette forsøket benyttet seg av geneditering med CRISPR-Cas9 systemet for å indusere knock out mutasjoner i disse genene. Seks ulike målsekvenser brukt, to i hvert gen (Tp6330PAM1 og -2 og Tp23191PAM2 og -4) for enkle knock out mutanter, og to målsekvenser som var identiske i begge genene (TP23191-6330PAM1 og -2) for doble knock out mutanter. Celler som hadde gitt indikasjoner på mutasjoner i tidligere arbeid hadde blitt replatet, og screening ble utført på disse koloniene. Ingen mutanter ble identifisert og isolert. Det andre forsøket omhandlet et lokaliseringsstudie, hvor to gener som koder for proteiner i silicanin-proteinfamilien, Tp23191 i gruppe IV og Tp20931 i gruppe III, ble forsøkt tagget med det fluoriscerende proteinet mTurq. Det ble brukt flere kloningsstrategier for å forsøke å sette sammen plasmidet som skulle bli brukt for å transformere *T. pseudonana*-celler med disse genene tagget med mTurq, men ingen var vellykkede. Til slutt ble morfologien til frustulen i *T. pseudonana* studert under påvirkning av kadmium og aluminium. Scanning electronmikroskopering (SEM) bilder av frustulen av *T. pseudonana*-celler som vokste uten påvirkning av metaller (kontroll), med 25 μg/L aluminium (Al_{25}) tilsatt, og 5μg/L kadmium (Cd_5) tilsatt ble analysert. Resultatene viste at diameter av frustulen, fultoportulae, rimoportulae, cribrumporer og bredden av costae var signifikante ulike fra kontrollprøvene i begge prøver, med unntak av diameteren til cribrumporene i Al_{25} og diameteren til fultoportulae og rimoportulae i Cd_5.

Abbrevations

AFIM	Ammonium fluoride insoluble matrix
Al	Aluminium
C	Control
Cas	CRISPR associated protein
Cd	Cadmium
DSB	Double strand break
Indel	Insertion/deletion
LCPA	Long chain polyamine
Mya	Million years ago
nat	N-acetyl transferase
Nou	Nourseothricin
PAM	Protospacer adjacent motif
SAP	Silicalemma associated protein
SDV	Silica deposition vesicle
SEM	Scanning electron microscopy
sgRNA	Single guide RNA
Sin	Silicanin
Std.dev	Standard deviation

Contents

Introduction

1.1 Background

The first recorded description we have of diatoms was made by an anonymous Englishman in 1702. When studying stalks of a water plant under a microscope he discovered what he described as rectangular oblongs. He speculated if the rectangular parallelograms could be a salt, but since they were always of the same size and did not alter their form when dried or re-exposed to warm water he decided instead that they were a plant [Anonymous, 1702]. In his confusion he caught the essence of diatoms; mineral utilizing plants. Although they were found fascinating and beautiful, they were deemed unimportant by scientists until the late 1800s, when Victor Hensen published his revolutionary thoughts on phytoplankton, of which diatoms are a part. He believed planktonic populations were a rapidly revolving link in the food chain, a way of thinking that was not accepted by his peers at the time, who were unable to conceive that the small could feed the large [Smetacek, 1999].

It has since been discovered that phytoplankton, the collective group of prokaryotes and eukaryotes that can photosyntesize, contribute to at least 50% of the world's primary production. The organic carbon that they produce is rapidly taken up by other organisms, and thus create the base of the food chain in the world's oceans, confirming Hensen's hypothesis [Armbrust, 2009]. The diatoms are the most abundant and diversified subgroup of phytoplankton with an estimated 100,000 different species [Mann and Vanormelingen, 2013] and contribute to about one fifth of the world's photosynthesis, creating more organic carbon annually than all of the terrestrial rainforests combined [Armbrust, 2009]. They can be found all over the world, wherever there is enough nutrients and light [Benoiston et al., 2017]. They are also important in the biochemichal recycling of nutritions such as nitrogen, and especially silicon. Every silicon atom entering the ocean is incorporated into the cell wall of a diatom on average 39 times before the cells eventually sink to the ocean floor Armbrust [2009]. This makes them important factors in carbon export, as the sinking cells are deposited on the ocean floor over geological time scales [Benoiston et al., 2017].

1.2 The Evolution of Diatoms

The earliest well-preserved diatom fossil is dated to about 190 million years ago (Mya), although molecular clock-based estimates suggest that the diatoms arose as early as 250 Mya. The emergence of diatoms, along with two other groups of eukaryotic phytoplankton,

initiated a decrease of atmospheric CO_2 concentrations as the heavier cells were more likely to sink to the sea floor, together with an increase of atmospheric O_2 concentrations as a result of their photosynthesis [Armbrust, 2009].

According to the secondary endosymbiotic hypothesis, diatoms evolved from two endosymbiotic events that happened sometime between 1200 to 700 Mya. The two events, illustrated in Fig 1.1, included the engulfment of (or invasion by) a red and a green algae into an eukaryotic heterotroph, which created the stramenopiles, the phylumn in which diatoms sit. The diatom gene sequences analyzed today suggests that the green algal endosymbiont preceeded that of the red one, as the genes apparently derived from green algae are very abundant within diatoms, whereas the genes derived from red algae are not. The green algae genes were incorporated into the nucleus, while the red algal endosymbiont gave rise to their chloroplast. This combination might have given the diatoms a selective advantage in ocean environments, and might be the cause for their dominance in the worlds oceans [Benoiston et al., 2017]. Their silicified cell wall have preserved the fossil record of diatoms remarkably well. These fossil records indicate that the centric diatom arose first, and were then followed by the pennate diatoms around 70 Mya [Bowler et al. 2009, Chacón-Baca et al. 2002]. The modern diatoms such as the radial centric *Coscinodiscus*, the multipolar centrics *Thalassiosirales*, and *Chaetocerotales*, and the araphid pennates *Thalassionema* and *Synedra* can be traced back to about 30 Mya, a result of a major shift in planktonic diatom diversification [Bowler et al., 2009].

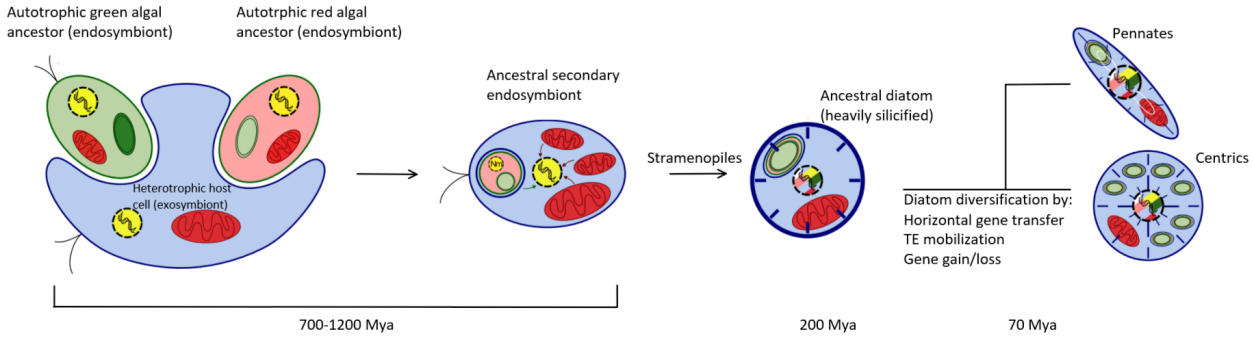

Figure 1.1: Schematic view of the evolution of diatoms according to the endosymbiotic hypothesis, which involved the engulfment of (or invasion by) a red and a green algae into an eukaryotic heterotroph. This created the ancestral diatoms, which diversified into the first pennate and centric diatoms about 70 Mya. The major events of the evolution is shown with the approximate dates (Mya = million years ago). Figure modified from Benoiston et al. [2017].

In addition to the two endosymbiotic events, there are evidence of gene transfer to diatoms from bacteria, representing as much as 5% of the total gene content in some species [Bowler et al., 2009]. Many of these genes are essential for diatoms and are shared amongst several

species, which indicates that they have ancient origins. The genome also contains diatom-specific transposable elements, which in turn may also have contributed to the rich diversity of species [Benoiston et al., 2017]. Surprisingly, the diatoms also have a complete urea cycle, which is usually only found in animals that consume complex nitrogen compounds and excrete nitrogenous waste products. They thus combine an animal-like ability to break down fat to generate chemical energy with a plant-like ability to generate metabolic intermediates from the breakdown. This combination allows them to survive long periods of darkness and resume cell division and growth when they return to the light [Armbrust, 2009].

1.3 The cell wall of Diatoms

The cell wall, called a frustule, is the diatoms most characteristic feature, essentially consisting of amorphous, hydrated silica ($SiO_2 \cdot nH_2O$) [Armbrust, 2009], and is in fact from where the diatoms get their name. The word diatom originates from the Greek word *diatomas*, meaning "cut in half", referring to their cell wall that is made up of two distinct halves. Each half is called a theca, where the slightly larger half, the epitheca, overlaps the smaller hypotheca, in a petridish like manner (Fig. 1.2) [Armbrust, 2009]. Diatoms are traditionally classified as either radial or pennate diatoms based on their symmetry, and range in size from microns to millimeters [Bowler et al., 2009]. The frustule provides the cell with rigidity, defense against predators, counterbalance towards turgor pressure, as well as providing photonic properties and giving the cell its shape. The frustule consists of layers of three-dimensional structures, ranging in size from nanometers to micrometers, creating a complex and highly species specific pattern of pores [Kröger and Poulsen 2008, Mock et al. 2008].

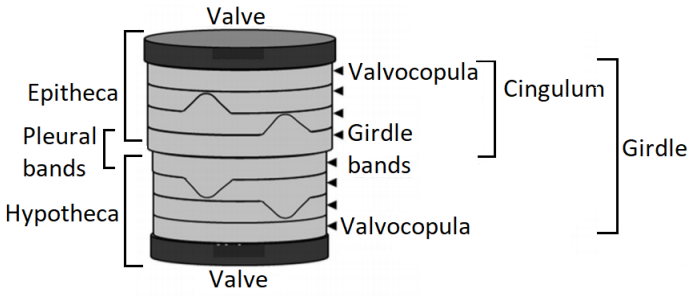

Figure 1.2: Schematic view of the frustule of a centric diatom. Each thecae consists of a valve and a cingulum. The larger half, the epitheca overlaps the smaller hypotheca in a petridish like manner in the section of the girdle called the pleural bands. Figure modified from Hildebrand et al. [2018].

Each theca consists of a valve and a cingulum. The two cingulums together form the girdle, which consists of several circular, overlapping structures called the girdle bands. The middle part of the girdle where the two thecae overlaps is called the pleural band, while the girdle band proximal to the valve is called valvocopula (Fig. 1.2) [Fattorini and Maier, 2021].

New valves can only be formed during cell division due to their rigidity, while girdle bands are synthesized during the interphase in a step wise manner to avoid gaps in the cell wall as the cell grows and the distance between the epitheca and hypotheca increases [Kröger and Poulsen, 2008]. The silica compounds of the frustule are synthesized in specialized organelles called silica deposition vesicles (SDV). Separate SDVs develop during cell division and interphase, where they synthesize the silica structures for the valves and girdle bands, respectively (Fig. 1.3) [Ehrlich and Witowski, 2015].The silica patterns are hypothesized to depend on both self-assembly processes and controlled silica polymerization with the precipitating silica further molded by cytoskeletal interactions. After the silica elements are synthesized, the entire organelle is exocytosed to deposit the elements on the cell surface [Mock et al., 2008]. Although it is not completely clear how this happens, the most likely scenario is that the proximal SDV membrane fuses with the plasma membrane while the distal SDV membrane is retrieved into the cytoplasm by a compensatory endocytosis [Kotzsch et al., 2017].

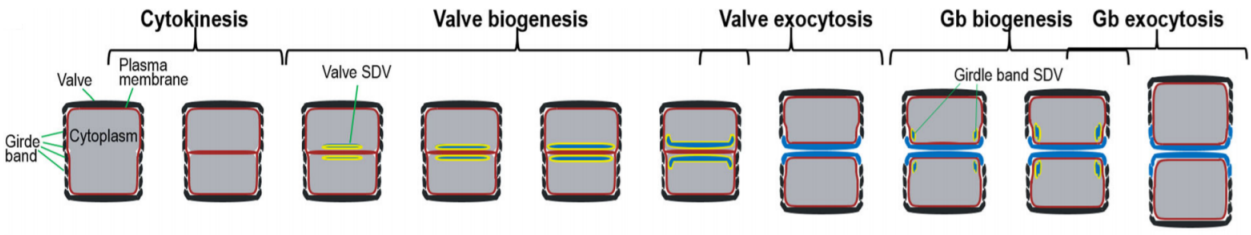

Figure 1.3: A schematic illustration showing how the frustule synthesis is coupled to the different stages of the diatom cell cycle. Valve synthesis occurs during cell division, while girdle bands are synthesized during interphase. Intracellular compartments, except for SDV, is not shown. Black and blue colors indicate mature and newly produced biosilica, respectively. Red and yellow colors depict the plasma and SDV membranes, respectively. Illustration obtained from Kotzsch et al. [2017].

Although many organisms are known to produce silica structures, none of them can compare to the diatoms' vast diversity of silica structures [Bedoshvili and Likhoshway, 2019]. The complexity and accuracy of the frustule's biosynthesis has spurred further research on diatoms, as understanding this process can inspire the development of novel syntheses for patterned inorganic materials with complex morphology and advanced properties. Diatomic structures are especially useful for nanotechnological applications, given their highly controlled nanopatters that are produced cost-efficiently by biological self-assembly in large quantities [Kröger and Poulsen, 2008]. In order to be able to manipulate the cells to create certain structures that can be used for nanotechnological purposes, it is important that we understand the mechanism and the components involved in the synthesis of the frustule completely.

1.3.1 SDV associated proteins

Several proteins and organic components have been identified or is suspected to be involved in the biosynthesis of the frustule. The proteins are generally divided into three classes, based on the biochemical approaches used to purify them and sequence characteristics (Fig. 1.4). 1) Soluble silica-associated proteins and long chain polyamines (LCPAs), which are isolated by isolation and detergent cleaning of the cell wall silica, followed by dissolution with ammonium fluoride. 2) The material that is not soluble using this method is termed ammonium fluoride insoluble matrix (AFIM), or the insoluble organic matrix. 3) Proteins that are associated with the SDV-membrane (the silicalemma). These proteins are characterized by having a single transmembrane domain, and have an intraluminal and a cytoplasmic portion [Hildebrand et al., 2018].

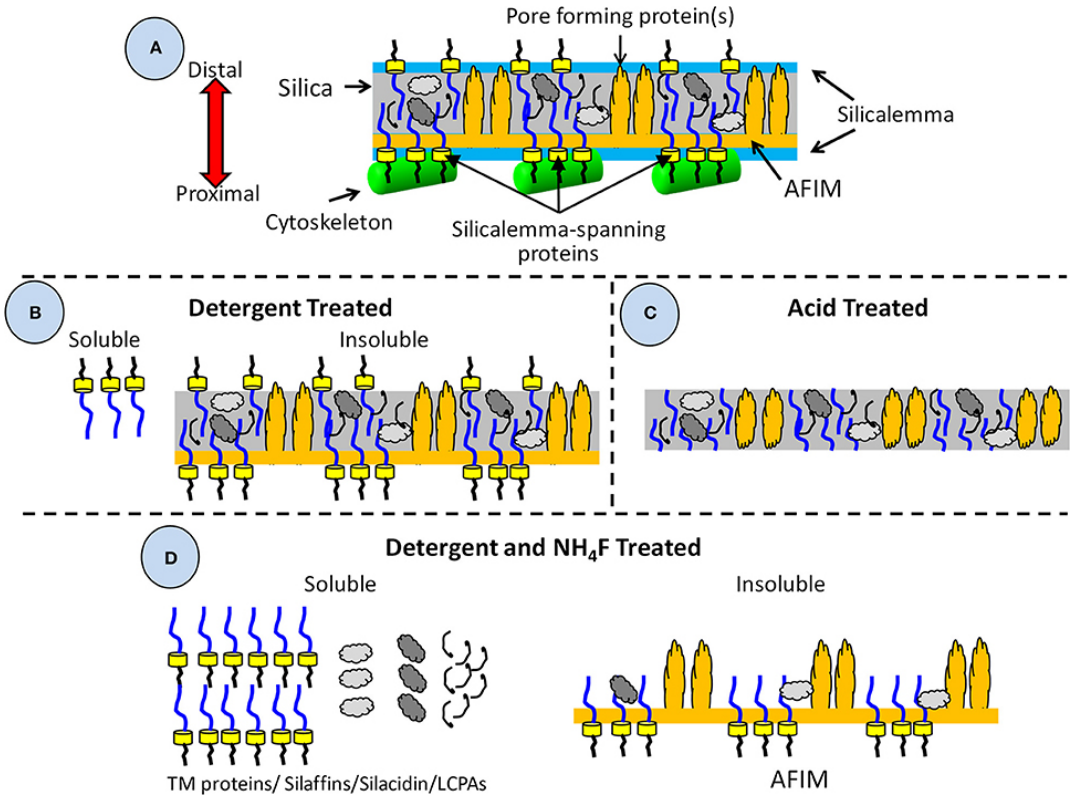

Figure 1.4: Diagram of the frustule and the solubility and isolation process of its associated components. A: all associated components, including those embedded in the silica, or associated with the silicalemma or AFIM, and the cytoskeleton. B: The frustule after detergent cleaning, resulting in the removing of the cytoskeleton and silicalemma, and proteins associated with the silicalemma. C: Acid treatment removes all organic material external to the silica. D: Components resulting from detergent cleaning and ammonium fluoride dissolution of the silica. This results in a soluble faction of silica-embedded and silicalemma components, and an insoluble faction, of the AFIM and any proteins that had become associated with it via crosslinks. Figure obtained from Hildebrand et al. [2018]

LCPAs are important for the silicification process, by rapidly catalyzing the polymerization

of silica. Altering the ratio of different size classes of LCPAs affected the silica morphology, generally by changing nanoparticle dimensions. *In vitro* experiments of silica polymerization demonstrated that the LCPAs required the presence of polyanions in the form of phosphate or pyrophosphate [Kröger et al., 2000]. This led to the discovery of the silacidin, which is rich in serine, and aspartic and glutamic acid. Silacidins might thus assist the LCPAs in silica polymerization [Wenzl et al., 2008]. The silaffins is another protein family belonging to the soluble silica-associated protein group. Purified silaffins rapidly catalyzed polymerization of silica, and the mixture of silaffins used affected the morphology of the silica [Kröger et al., 2002]. A study performed on five polypeptides from this family resulted in the division into two groups, regulatory silaffins, which polymerize silica dependent on the protein's concentration, and catalytic silaffins, which have a continuous stimulatory effect on silica formation [Poulsen and Kröger, 2004]. By searching for proteins with a similar amino acid composition as the silaffins, the cingulins was discovered. Characterization of the insoluble material of the frustule led to the discovery of these proteins being associated with "microrings" that precisely mimicked the structure and dimensions of the girdle bands. The microring material was then termed AFIM, and they also found plate-like insoluble matrix which was associated with the valve SDV [Kotzsch et al., 2016].

In addition to these proteins, there is also evidence for the cytoskeleton being involved in forming diatom silica structures by providing a rigid framework for expansion and for positioning components of the SDV that lead to meso- and micro-scale structures. As these are assembled on the outside of the SDV, they need to transmit their assembly pattern across the silicalemma. It has been proposed that the silicalemma-spanning proteins could serve as intermediates that translate cytoskeletal assembly patterns into similar silica structures. The silicalemma associated protein (SAP) family has the predicted protein structure needed for this function, although the C-terminal portion, located at the cytosolic side of the silicalemma, lack any known cytoskeleton-interacting sequence [Tesson et al., 2017]. It is possible that they either contain a novel cytoskeleton-interacting sequence, that they interact indirectly with the cytoskeleton, or that they do not interact with it [Hildebrand et al., 2018].

The silicanin (Sin) protein family share the same overall structures as the SAPs. This protein was found to be associated with both the girdle bands and the valves, The Sin proteins are, like silacidin, highly negatively charged in the acidic environment inside the SDV, and was found to interact with LCPAs to catalyze the biosynthesis of silica [Kotzsch et al., 2017]. The first protein discovered in this family is called Silicanin-1 (Sin1, previously SiMat7), which is a predicted type 1 transmembrane protein, with a c-terminal cytosolic domain with no known cytoskeleton binding sites, preceded by a single transmembrane helix (Fig. 1.5). Like other transmembrane proteins spanning the silicalemma, they contain an N-terminal

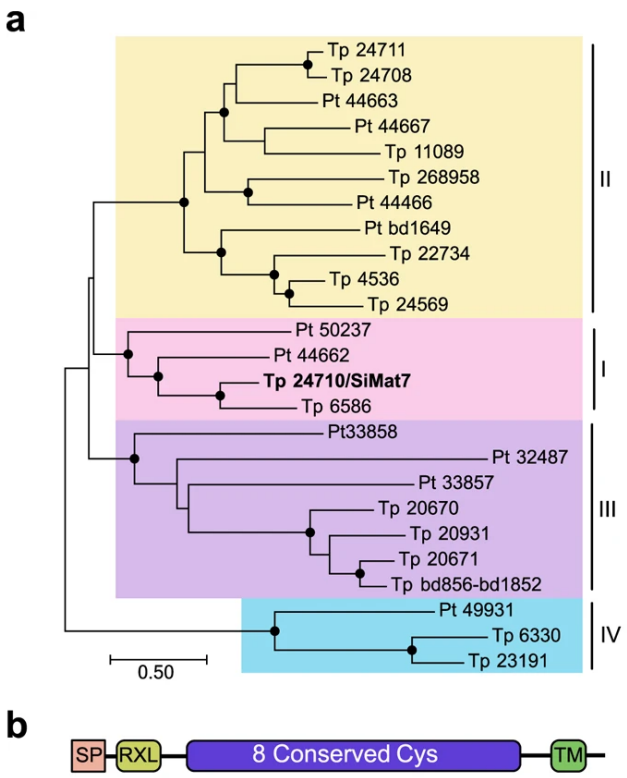

Figure 1.5: The silicanin protein family, predicted to be involved in frustule biosynthesis. a) Phylogenetic tree of silicanins. The proteins are divided into four groups (I-IV). The first discovered protein of this family, SiMat7 (later renamed Sin1) is shown in bold. The two proteins studied in the knock out experiment, Tp23191 and Tp6330, are located in group IV, and Tp20931 in the localization study is in group III. b) predicted domain structure of the silicanin proteins. The silicanin proteins consists of a signal peptide (SP) for cotranslational import into the endoplasmatic reticulum, a protease cleavage site (RXL), followed by a stretch with 8 conserved cysteins (Cys) and a transmembrane domain (TM). Figure obtained from Brembu et al. [2017].

signal peptide for co-translational import into the endoplasmatic reticulum. This suggests that the remaining part of Sin1 might be exposed to the extracellular space or the lumen of a secretory compartment. Following the signal peptide, the protein has a stretch of 30 amino acids ending with a protease cleavage site (RXL domain), which is typical for many diatom biosilica-associated proteins [Kotzsch et al., 2017]. A knockout experiment in *T. pseudonana* of the Sin1 gene performed by Görlich et al. [2019] showed that the mutant *T. pseudonana* cells lacking the Sin1 gene exhibited differences in valve morphology from wild type (see Section 1.6), while the girdle bands of the mutants had no observed differences. In these mutants, the cross connections between the ribs in the valve were mostly absent, and therefore also lacked the pattern of areola pores. The ribs were also found to be flatter than in the wild type. Due to these differences in valve morphology, the mutants had reduced mechanical strength and stiffness. From the results, it was therefore concluded that the Sin1-protein is important for biosilica production, but not essential.

1.3.2 Manipulating the frustule morphology by CRISPR-Cas9

The most fundamental way to achieve desired nanostructural parameters is by species or strain selection. Cultivation conditions and genetic manipulation can further optimize these properties. Extensive knowledge of the genome is required for genetic manipulation, which has so far only been achieved for a few diatom species [Su et al., 2018]. The most used method for genetic manipulation uses the clustured regularly interspaced palindromic repeats (CRISPR) system and the CRISPR associated (Cas) protein. This system originates from bacteria, where it functions as an adaptive immune system against viral infections. The system functions by integrating small sequences of viral or plasmid DNA/RNA into the genome of the bacteria. This sequence is then transcribed together with the CRISPR associated (Cas) protein, and serves as a single guide RNA (sgRNA) that base pairs with the complementary viral genome during a new infection [Doudna and Charpentier, 2014]. The Cas protein, which is an endonuclease, will then cut the foreign DNA or RNA (Fig. 1.6a). Some types of Cas proteins will create different types of cuts in DNA, while others cut RNA [Liu et al., 2020]. The sgRNA contains two critical features: a 20-nucletide sequence at the 5' end that base pairs with foreign DNA, and a double stranded structure at the 3' end that binds to the Cas protein. The 20 base pair structure at the 5' end determines the cut site. This sequence can be modified to create specific double strand breaks (DSBs) in any DNA sequence of interest, and can thus induce mutations in any organism. In order for the Cas protein to be able to make a DSB in the foreign DNA, the DNA targeting site needs to be located adjacent to a protospacer adjacent motif (PAM) [Doudna and Charpentier, 2014]. This sequence is important for initial DNA binding, and is a three base sequence of NGG, where N can be any base, although the system can also initiate base pairing with NAG, albeit at a lower efficiency [Hsu et al., 2014].

The first Cas protein discovered, the Cas9 protein, creates DSBs [Sander and Joung, 2014]. The cell's repair mechanism will try to repair this otherwise lethal damage to their DNA. The cell have two repair mechanisms to do so (Fig. 1.6b), homology-directed repair (HDR) or non-homologous end joining (NHEJ). HDR uses either the other allele or exogenous DNA as a template to repair DSB. NHEJ is a faster method, and usually the first reaction carried out after a DSB, although this method is often error-prone. The system works by creating single stranded overhangs to enable the two ends to be ligated together. In order for the two overhangs to match, the system might have to add or delete bases, and creates indel mutations of various lengths. This can create a frameshift mutation, which will cause a knock out of the gene [Lui et al., 2019].

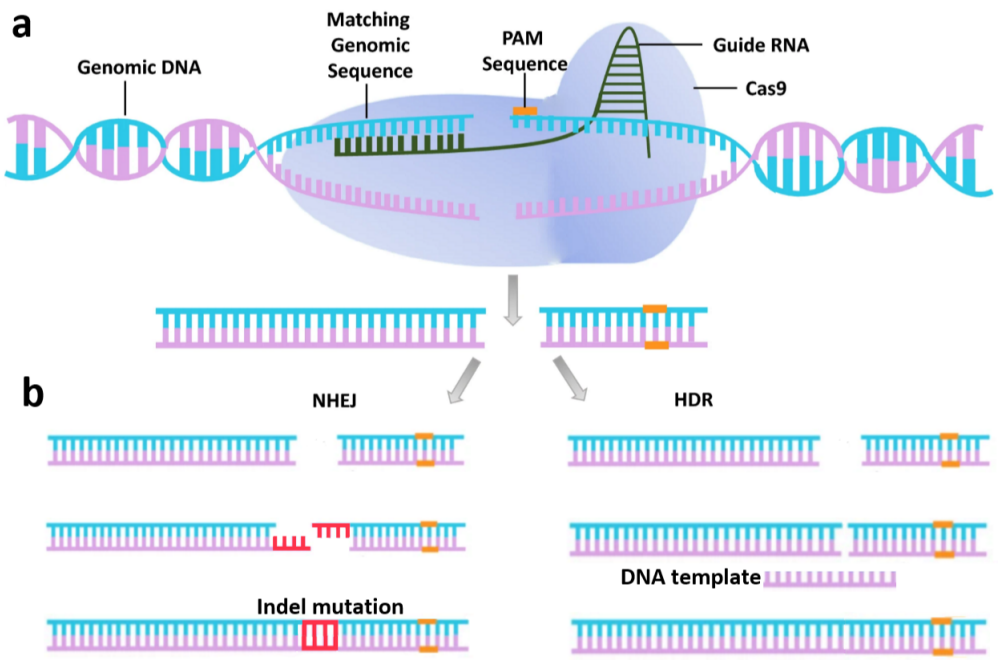

Figure 1.6: Schematic view of gene editing with CRISPR-Cas9. (a) The Cas9 associated sgRNA base pairs with the target DNA sequence, and the Cas9 protein makes a DSB in the DNA adjacent to the PAM sequence. (b) The cell's two repair mechanisms for repairing DSB in its DNA. Left: DNA is repaired by NHEJ, creating insertion or deletion (indel) mutations. Right: DNA is repaired by HDR, using the other allele as a template. No mutations occurs. Figure modified from [Gosh et al., 2019]

1.3.3 Manipulating the frustule morphology by cultivation conditions

Altering the cultivation conditions is a simpler way of optimizing the properties of the cell. The hierarchical pattern, shape and size of the frustule is sensitive to environmental changes like pH, salinity, inorganic nutrient dynamics, temperature, and light [Gensemer, 1990]. Diatom morphology and taxonomic composition have in fact been successfully used to monitor several disturbances in their environment, like eutrophication and acidification [Cattaneo et al., 2004]. Several studies have been conducted on the effect of metals on a range of diatom species [summarized in Su et al. [2018]]. The effects can vary from changes in cell length, to increases in the thickness of the valve, to morphological changes in the frustule or other components of the cell [Su et al., 2018]. Some of these metals are trace metals, others are toxic for diatoms, and since different species have different metal tolerance, pollution might alter the taxa of diatoms in the polluted environment. Metal pollution is often associated with acidification, which will in turn increase the bioavailability of metals, potentially making them more toxic [Gensemer, 1990]. Because of the ecological importance of diatoms, it is necessary to understand how they react to different pollutants.

1.4 *Thalassiosira pseudonana*

The centric diatom *Thalassiosira pseudonana* was the first diatom whose whole genome (34.5 Mb) was sequenced [Armbrust et al., 2004], and has become a model organism for biosilica composition and generation studies [Sumper and Brunner, 2008]. Its nuclear genome consists of 24 chromosome pairs and over 11 000 predicted proteins-coding genes. Over half of these cannot be assigned a function based on similarities to genes in other organisms, partly because diatoms have distinct features that cannot be understood by applying it to model systems [Armbrust et al., 2004].

The valve diameter is highly conserved at 3.8 μm, with most between 3.0 and 4.2 μm in diameter. The average length is affected by the salinity of the growth medium, but varies between 4.3 and 9.0 μm [Hildebrand et al., 2006]. Unlike most other diatoms species, *T. pseudonana* does not undergo sexual reproduction, and retains a consistent size over generations by an unknown mechanism [Fattorini and Maier, 2021]. While the girdle bands are mostly flat and have few distinct features, the valves have a complex hierarchical pattern of pores ranging from the micro to micrometer scale, shown in detail in Fig. 1.7 [Heintze et al., 2020]. At the center of the valves, a silica ring called an annulus is positioned, from which regularly spaced ribs, called costae, radiates. The costae remains mostly parallel to each other by branching, and therefore have a consistent average distance between them of 145 nm [Hildebrand et al., 2006]. In addition to the branching there are also silica bridges connecting neighboring costae in the central part of the valve. Two neighboring silica bridges and their interjacent costae constitutes the areola pore [Heintze et al., 2020]. Numerous nanopores (cribrum pores) with an average diameter of 18 nm are located all over the valve, covering about 4% of the surface. There is generally one (rarely two) larger pore offset from the center of the valve, called the fultoportula. Silica ridges are built upon the ribs distal to the cell center, and a various number of pores similar to the fultoportula, called the rimoportulae, are built upon this ridge [Hildebrand et al., 2006].

1.4.1 Synthesis of the valve of *Thalassiosira pseudonana* in silica deposition vesicles

A recent study by Heintze et al. [2020] have captured the details of valve formation in *T. pseudonana* inside the SDV (Fig. 1.8) by synchronizing a culture of *T. pseudonana* cells by two consecutive Si starvation-replenishment series, maximizing the proportion of cells bearing valve SDV. Its formation starts with the annulus. The costae emerge from the annulus and grow radially from there, branching soon after they emerge. Cribrum pores start to form already at early stages of costae growth, starting at the valve center and propagates radially

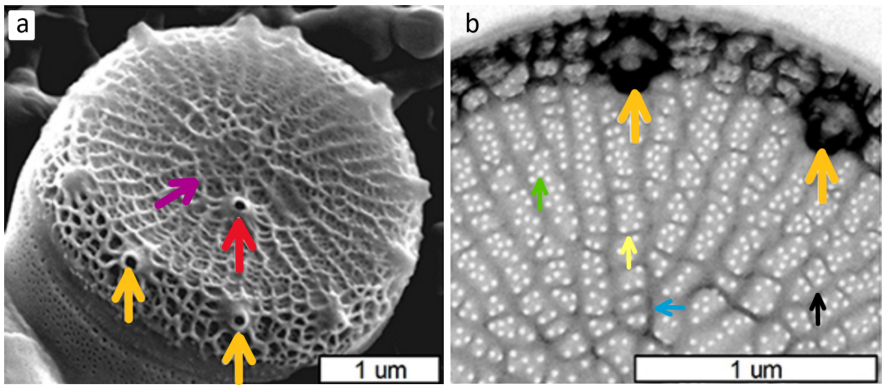

Figure 1.7: SEM images of *T. pseudonana* valve. The colored arrows indicates the valves biosilica structures. The colored arrows indicates the valves biosilica structures. Red: Fultoportula, purple: annulus, orange: rimoportulae, green: cribrum pore, yellow: silica bridge, blue: costae, black: areola pore. a) overview of the complete valve, b) close up image of the valve. Image modified from Hildebrand et al. [2018].

along the costae, while a thin silica layer develops around the pores and fill the entire space between the costae. Rimoportulae formation is highly synchronized and starts as the costae have almost reached its full length. They are formed from the ends of four (rarely three) costae that cease normal growth, and instead develops into a central tube with three satellite pores. The inter-costae space adjacent to the rimoportulae is completely silicified, lacking cribrum pores. Simultaneously as the formation of the rimoportulae begins, the costae in the central part of the valve becomes connected by silica bridges, and thus creates an areola pore. The formation of these bridges decrease towards the periphery, but becomes particularly prominent close to the valve margin. Costae not involved in rimoportulae formation continues to extend as the rimoportulae forms, but cease shortly after the formation of the rimoportulae is complete. The costae ends then merge into a patternless, non porous ring of silica, creating a continuous margin of the valve. Simultaneous as the formation of rimoportulae and areola pores, the whole valve increase in thickness. This process is called the z-expansion, and is the final step of valve formation in *T. pseudonana* [Heintze et al., 2020].

1.4.2 Aim of experiment

This report contains a three-part experiment to further investigate the frustule. Firstly, a knock-out experiment was performed. The aim of this experiment was to isolate and characterize knock out mutant *T. pseudonana* cell lines of the silicanin protein family genes *Tp23191* and *Tp6330*. *T. pseudonana* cells had previously been transformed with the pTpPUC3-Cas9 plasmid containing specific PAM sites in these genes. In total, six different target sites were chosen, two in each gene (Tp6330PAM1&2 and Tp23191PAM2&4), and two PAM sites that were identical in both genes (TP23191-6330PAM1&2) for the double knock out mutants. Some screening of these cells had been done in previous work, and this process was continued

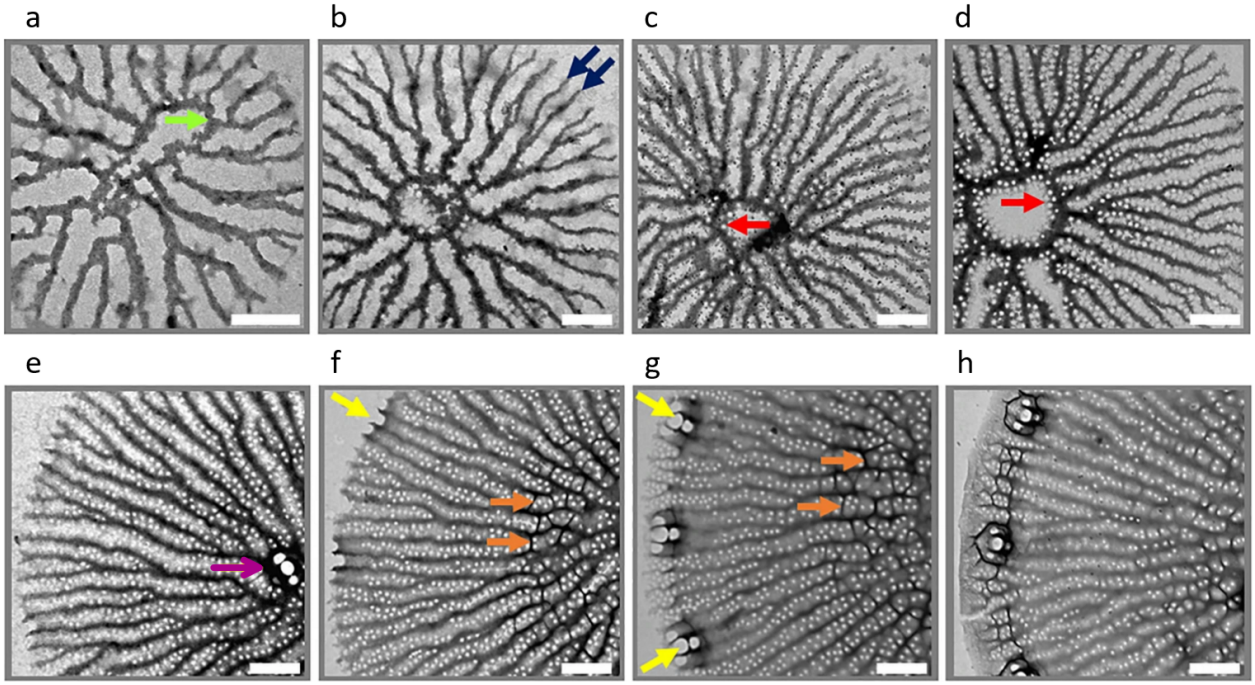

Figure 1.8: Valve development in *T. pseudonana*. Colored arrows marks different biosilica structures. Green: annulus, blue: costae, red: cribrum pore, purple: fultoportula, yellow: rimoportula, orange: areola pores. Scale bars: 1 μm. Image modified from Heintze et al. [2020].

in this experiment. Secondly, a localization experiment was performed on two silicanin proteins, Tp23191 in group IV and Tp20931 in group III. This was done by inserting these genes tagged with the fluorescence protein mTurq in the pTpPUC33 vector, and transforming this vector into *T. pseudonana* cells. Lastly, *T. pseudonana* cultures were grown in medium containing aluminium and cadmium at three different concentrations to study the effect of these metals on the morphology of the frustule. Image analysis was performed to calculate mean values of valve, fultoportulae, rimoportulae and cribrum pore diameter and costae width of *T. pseudonana* cells growing in ESAW medium with no added metal (control), supplemented with 25 μg/L aluminium or 5 μg/L cadmium.

Materials and Methods

2.1 Materials

All materials, medium, instruments etc. are listed in Appendix A.1.

2.2 Methods

2.2.1 Single and double knock out of *Tp23191* and *6330*

This experiment was a continuation of previous work I have done. *Thalassiosira pseudonana*-cells were transformed with the pTpPUC3-Cas9 vector (Fig. 2.2) containing the gene for Cas9 and its associated sgRNA with target sites located in *Tp23191* and *Tp6330*. The location of the target sites in each gene is shown in Fig. 2.1. Some clones whose results indicated mutations were replated. A detailed description of the cloning process, transformation and screening done during the previous work is attached in Appendix A.2.

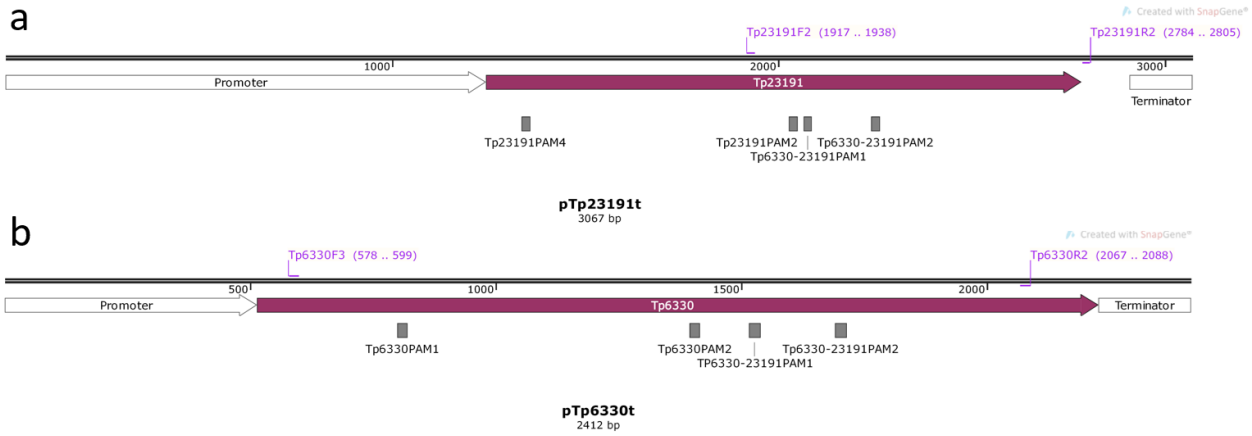

Figure 2.1: The two genes that were studied in this experiment, a) *Tp6330* and b) *Tp23191*. The maps show the gene in dark purple, with their promoter and terminator in white. The location for the different primers (listed in Tab. 2.1) that were used in this experiment in the present work is shown in purple. In the primer name, F represents forward primer and R reverse primer. The location for each target sequence in the Cas9 associated sgRNA in the pTpPUC3 vector is shown under each gene map in dark gray.

The aim of this experiment in this work was to isolate and characterise stable knock out mutant *T. pseudonana* cells. Primer pairs were tested to find the most effective primer pair and their annealing temperature for amplification of *Tp6330* with the least off target product. It was decided that Tp6330F3 and Tp6330R2, with annealing at 72°C was the best pair, and

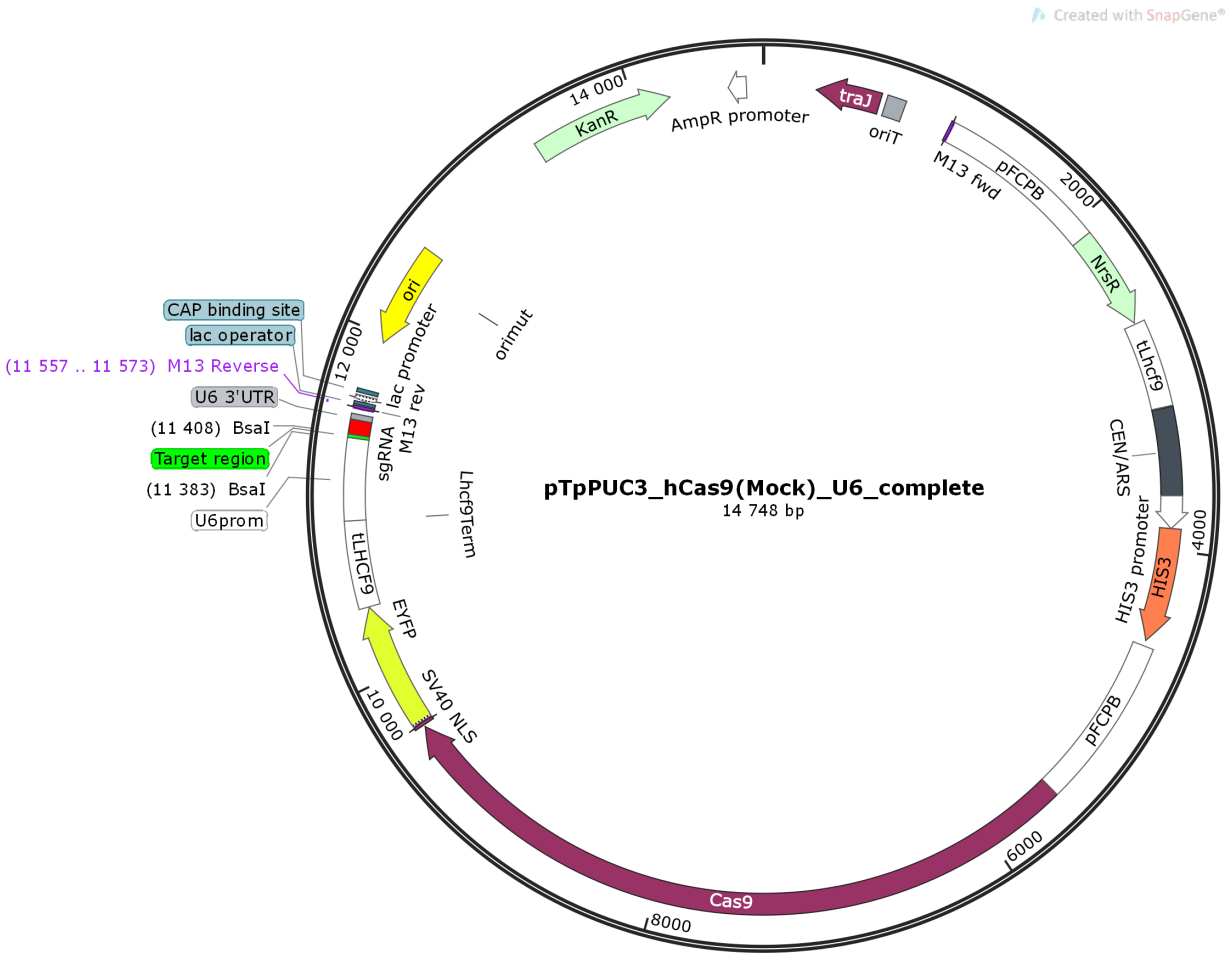

Figure 2.2: The pTpPUC3 vector (14.748 bp) used for transforming the gene for Cas9 (purple) and its associated sgRNA (red) with the specific target regions into *T. pseudonana* cells. The plasmid contains the genes for two antibiotics resistances (both in mint green) for selection of clones, kanamycin (*E. coli*) and nourseothricin (*T. pseudonana*). The Cas9 gene is tagged with a yellow fluorescence protein (EYFP, yellow) and a nuclear localization sequence (NLS, purple).

thus these primers were used for all PCR amplifications during the screening process. Primers for *Tp23191* had previously been tested by my supervisor, and the most optimal pair was found to be Tp23191F2 and Tp23191R2 at 65°C. The localization of these primers are shown in Fig. 2.1, and their sequence is listed in Tab. 2.1.

Table 2.1: Primers and their sequence used for PCR amplification of the genes Tp6330 and Tp23191.

Primer	Sequence	Annealing temp. (°C)
Tp6330F3	GTACCTTTCCACAAAGCCATCA	72
Tp6330R2	TCATTGATGGTGTTCCATGGCT	
Tp23191F2	CTTGCCTACGAATGCTACCAAG	65
Tp23191R2	CAACACAGGGAGGTCAAACTCT	

2.2.1.1 Screening for knock out *T. pseudonana* cells

Replated clones were picked and resuspended in the PCR reaction solution described for Phusion High-Fidelity DNA polymerase. The genes of interest from clones and wild type were amplified by PCR, following the procedure for this polymerase using the primers listed in Tab. 2.1. For clones transformed with the pTpPUC3 vector containing target sites in both genes (double knock out transformants), the same clone was used to amplify both genes by picking the same clone twice. The PCR products of the clones were compared to the wild type genes by gel electrophoresis analysis (1% agarose). For samples with a successful amplification, the PCR product was cleaned with ExS pure enzymatic kit, by mixing the enzyme (2 μL) with the PCR product (5 μL) and incubating it at 37°C for 4 min before inactivating the enzyme at 90°C for 1 min. Primer (2.5 μL) and water (to a total of 10 μL) was added to this product, which was then submitted for Sanger sequencing. The sequencing results were compared to the wild type gene previously sequenced in the lab, using the Snapgene (v5.2.4) align tool.

2.2.2 Localization study

The aim of this experiment was to tag two proteins in the silicanin protein family, Tp23191 in group IV and Tp20931 in group III with the fluorescence protein mTurq to observe their locations *in vivo*. A representative map of the vector, pTpPUC-mTurq, used for transforming the genes tagged with mTurq is shown in Figure 2.3. This map shows the *Tp23191* tagged with the fluorescence protein mTurq inserted in the plasmid. *mTurq* was inserted downstream of the start of the gene because of the proteolytic removal of the signal peptide. This was done by amplifying the fragments needed, and inserting them into the linearized pTpPUC3 vector by different cloning techniques. The primers used for amplification of the fragments are shown in Tab. 2.2. The sequence of these primers, and other primers used for e.g. colony screening is attached in Appendix A.3

2.2.2.1 Constructing vector used for conjugation with mTurq-tagged genes

The DNA used as a template for amplifying the gene, promoter and terminator fragments was extracted from wild type *T. pseudonana* cells by spinning down *T. pseudonana* culture in f/2+Si medium (10 μL) at 17,000g for 1 min. The supernatant was removed and lysis buffer (20 μL) was added. The solution was kept on ice for 15 min, before it was incubated in a thermo shaker at 95°C for 10 min.

The fragments were amplified by PCR with Phusion High-Fidelity DNA polymerase following the protocol for this enzyme. The primers used for amplification and their annealing tem-

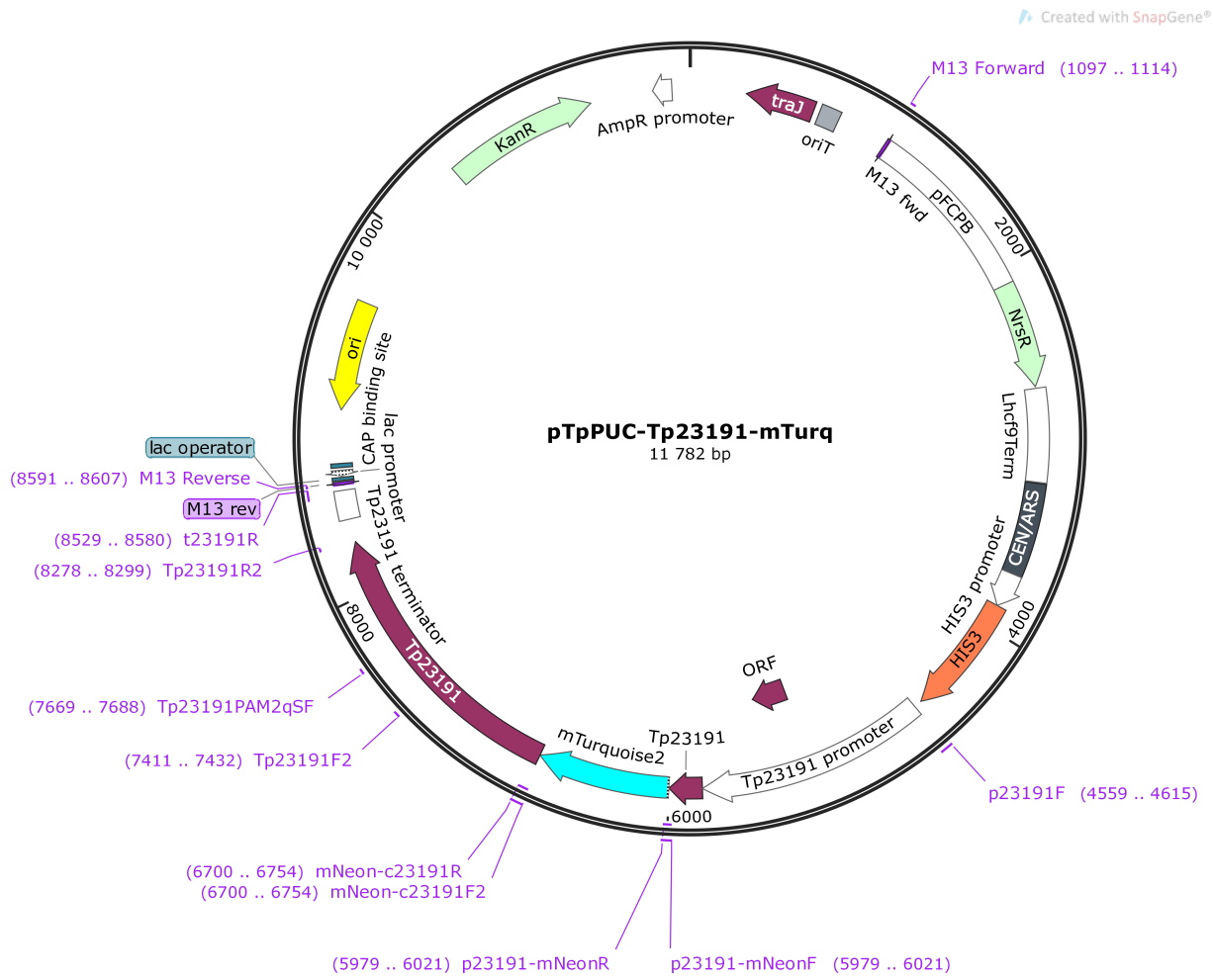

Figure 2.3: Map of the plasmid used for transforming *T. pseudonana* cells with the genes tagged with the fluorescence protein mTurq. The gene for mTurq is inserted downstream of the start of the gene because of the proteolytic removal of the signal peptide located in the beginning. The gene's promoter and terminator is shown in white, the gene is in purple, and mTurq is in cyan. The plasmid contains two antibiotic resistance genes, *KanR* and *NrsR*, both in mint green, used for selection of transformed *E. coli* and *T. pseudonana* cells, respectively. The primers used for amplification, colony screening and Sanger sequencing is included.

perature is shown in Tab 2.2. Amplification was confirmed by analyzing the PCR product by gel electrophoresis. The vector was cut with PstI-HF at 37°C for 30 min or 2 hours by adding cut-smart buffer (5 μL) and PstI-HF (0.5 μL) to the plasmid (10 μL) and diluting with MQ-water (to 50 μL). The enzyme was inactivated by incubating the reaction at 65°C for 15 min. To assemble the vector, Gibson assembly, nested Gibson, CPEC and two step Gibson was performed.

Table 2.2: Primers and their annealing temperature for amplifying fragments used in assembly for cloning pTpPUC3-mTurq vectors. The primers are engineered to create an overhang between the amplified fragment and the fragment/vector adjacent to it. The last column shows which fragment is created, and which DNA template was used in the PCR.

Forward primer	Reverse primer	Annealing temp. (°C)	Amplified fragment (DNA template)	Length (bp)
P20931F	P20931mNeonR	60.0	Tp20931 Promotor (*T. pseudonana)*	1424
P20931mNeonF	mNeonc20931R2	67.2	mTurq-Tp20931 (pBKS-mTurq)	717
mNeonc20931F2	t20931	69,6	Tp20931 Gene + terminator (*T. pseudonana)*	1881
P23191F	P23191mNeonR2	60.0	Tp23191 Promotor (*T. pseudonana)*	660
P23191mNeonF	mNeonc23191R	67.2	mTurq-Tp23191 (pBKS-mTurq)	717
mNeonc23191F	t23191R	69.6	Tp23191 Gene + terminator (*T. pseudonana)*	1889

Gibson assembly

The Gibson reaction was prepared by adding the different fragments and vector to the Gibson Master Mix (10 μL) and diluting with ddH_2O (to 20 μL). This reaction was repeated 3 times, using different concentrations of fragments and backbone, and different incubation times. An example of the combination of fragments is shown in Tab. 2.3. The reaction was incubated at 50°C for 60 min or 2h.

Table 2.3: An example of the combination of fragments used for Gibson assembly, showing the concentration of each fragments and the amount added to the reaction.

Fragment	Concentration (ng/μL)	Amount (μL)
Linearized pTpPUC3	298.4	0.3
Tp23191 Promoter	116	0.5
mTurq	78.0	0.75
Tp23191 Terminator	23.6	2.75
Linearized pTpPUC3	175.5	1
Tp20931 Promoter	61.8	0.75
mTurq	68.8	0.75
Tp20931 Terminator	22	2.5

Nested Gibson

If the Gibson assembly is unsuccessful, a nested Gibson can be performed using the failed Gibson product. This procedure uses a polymerase (here, Phusion High-Fidelity DNA polymerase was used) to amplify two or more fragments that have been ligated together during the Gibson assembly. The reaction was set up according to the protocol for this polymerase. Primers (Tab. 2.2) was selected based on the gel electrophoresis performed on the Gibson assembly product. The product was checked by gel electrophoresis, and a new Gibson reaction was performed using this product combined with the remaining fragments.

Two-step Gibson

The fragments were split into two tubes, with two or three adjacent fragments in one tube,

and the rest of the fragments and vector in the other. The reaction set up is the same as for the Gibson assembly described above. The reactions were incubated separately for either 15 or 30 min before they were combined and further incubated for 45 min or 1.5 hours, respectively. This procedure was repeated with different combinations and concentrations of fragments. An example is given in Tab. 2.4.

Table 2.4: An example of a two step Gibson reaction, showing which fragments are incubated together before combining them. The reactions were combined after 15/45 min and incubated further for 45 min/1.5h.

Fragment	Amount tube 1 (μL)	Amount tube 2 (μL)
Vector	0.3	-
Promotor Tp20931	0.72	-
mTurq Tp20931	-	0.72
Terminator Tp20931	-	2.5

Circular polymerase extension cloning

Circular polymerase extension cloning (CPEC) uses a polymerase for extension, where the overlaps between each fragments work as primers. CPEC was performed using Phusion High-Fidelity DNA polymerase according to protocol. This procedure was repeated, using different concentrations of the vector and fragments, and different annealing temperatures. An example of the reaction set up is shown in Tab. 2.5.

Table 2.5: An example of the reaction set-up for CPEC using Phusion High-Fidelity DNA polymerase. The PCR reaction was run according to protocol for this enzyme.

Compound	Amount (μL)
Vector	1
Inserts	3·1.33
Phusion buffer	5
dNTPs	1
Phusion polymerase	0.5
ddH_2O	To a total of 25

2.2.2.2 Heat shock transformation and colony screening

After each cloning procedure, the product was analyzed by gel electrophoresis (1% agarose) and heat shock transformed into competent NEB stable *E. coli* cells by the following method. Plasmid (0.5 μL) was added to competent *E. coli* cells (50μL). The solution was incubated on ice for 30 min and then heat shocked at 42°C for 45s. The solution was put on ice for 2 min before LB-medium (1 mL) was added and incubated at 30°C for 1h with shaking (250 rpm). The culture was plated out on LB-plates with kanamycin (100 μg/ml), and incubated at 30°C overnight, or until clones appeared. Clones were picked and plated out

on a new LB-plate (colony plate). The leftover cells on the pipette tip used for picking were mixed with the solution described in the protocol for RedTaq DNA polymerase Master Mix, and a PCR was performed according to protocol. The product was analyzed by gel electrophoresis. Colonies from the colony screening plate was picked and resuspended in LB-medium (5 mL) and incubated at 30°C overnight. A Miniprep was conducted on these cultures using Miniprep GeneJet kit according to protocol, except warm ddH_2O (70μL) was used instead of the elution buffer. To the product (7.5 μL), primer (2.5 μL) was added and this was submitted for Sanger sequencing. The results were compared to the SnapGene file of the plasmid to check for correct insertion of all fragments.

2.2.3 The effects of aluminium and cadmium on the frustule

The aim of this experiment was to observe the effects of two metals, aluminium (Al) and cadmium (Cd) at three concentrations each, on the frustule of *Thalassiosira pseudonana*. For Al, these concentrations were 5 μg/L, 15 μg/L and 25 μg/L. For Cd, these were 0.15 μg/L, 2.5 μg/L and 5 μg/L. A control with no added metal was included, and three replicates of each treatment were made. The effects were studied by scanning electron microscopy (SEM) and by measuring photosystem II efficiency.

2.2.3.1 Cultivation conditions

All samples were incubated at 18°C in a 16h:8h light-dark cycle with 175 μmol m^{-1} s-1 illumination. *T. pseudonana* cells were made axonic by spinning down f/2+Si-cell culture (5 mL) at 3000rpm for 10 min, and inoculating them in f/2+Si medium (12.5 ml) with antibiotics (Tab. 2.6) two consecutive times. ESAW medium with no added metal was inoculated with the axonic *T. pseudonana* culture, and the cells were adapted to the new media for 1 week. The ESAW medium with supplemented metal were prepared by adding sterile filtrated stock solutions of the metal to ESAW medium (15 mL), and they were inoculated (30 000 cells/mL) with ESAW *T. pseudonana* culture.

Table 2.6: Antibiotics mixture for making axonic *T. pseudonana* culture.

Antibiotic	Concentration (μg/mL)
Steptomycin	50
Gentamycin	67
Ciprofloxacin	20
Chloramphenicol	2.2
Ampicillin	100

2.2.3.2 Scanning electron microscopy

T. pseudonana culture in ESAW medium was spun down at 3000 rpm for 10 min and the supernatant was removed. The cells were washed three times with MQ water (10 mL) to remove salt, centrifuged at 4500 rpm for 10 min between each washing, and stored at 60°C overnight to dry. SDS-EDTA solution (1 mL, 50 g/L SDS in 100mM titriplex III) was added to the dried cells, and the solution was mixed by pipetting up and down several times. After being left still for 20 min it was centrifuged at 4500 rpm for 10 min. The SDS-EDTA washing procedure was repeated 6 times, after which the frustules were washed three times with MQ water as described above. The frustules were stored in ethanol (96%) at -20°C.

The fixation, drying, mounting and coating of the samples in preparation for SEM was performed by my supervisor (A. messemer). The SEM pictures were analyzed using ImageJ (v153.C). The diameter of the valve, fultoportulae, rimoportulae, cribrum pores and the width of costae was measured. Statistical analyses (T-test) was calculated using Microsoft Excel (v2105).

2.2.3.3 Measuring efficiency of photosystem II

Quantum yield is a measure of the efficiency of photosystem II. In dark adapted samples this is equivalent to Fv/Fm, where Fv is maximal variable fluorescence and Fm is maximal fluorescence intensity. Culture (1mL) was transferred to a cuvette and placed in the aquapen for three minutes for the samples to be dark adapted before the quantum yield was measured.

Results

3.1 Single and double knock out of the genes Tp23191 and Tp6330

This project is a continuation of previous work, which is summarized in the following paragraph. The aim of the experiment was to create knock out mutant *Thalassiosira pseudonana* lines with a single or double knock out in the genes Tp23191 and Tp6330 encoding two members of group IV of the silicanin protein family. This was performed by transforming *T. pseudonana* cells with the pTpPUC3-Cas9 plasmid with the gene-specific PAM target sites by conjugation with DH10β *E. coli*. Six different target sequences were used: two target sites for each gene in the single knock out experiment, and two target sequences that were identical in both genes for the double knock out experiment. Clones from the initial conjugation plate were picked and incubated in liquid f/2+Si-medium. At least 24 clones from each conjugation plate was picked, while for some conjugations, 36 or 48 clones were picked. Screening was performed by extracting DNA from the clones transformed with the pTpPUC3-Cas9 plasmid, and amplifying the genes of interest by PCR and comparing the product with the same gene amplified from a wild type cell by gel electrophoresis analysis. High resolution melting (HRM) analysis was also performed to look for small indel mutations that would not be visible from screening with PCR and gel electrophoresis analysis. Additionally, flow cytometry was also performed to look for changes in cell size or other morphological changes (results in Appendix A.5). The results indicated mutations only in the Tp6330 gene in clones that were transformed with the pTpPUC3-Cas9 vector containing the target site Tp6330PAM2 or Tp23191-6330PAM2. Some of these clones were therefore replated, to attempt to obtain a single and stable mutant cell line. A detailed description of this work is attached in Appendix A.2.

In the present work, the screening process was continued on the replated cells by amplifying the genes by PCR and comparing them to the same gene amplified from wild type *T. pseudonana* cells. Since results of gene amplification by PCR in the previous work showed several bands when ran on an agarose gel, even for the wild type, different primer pairs were tested to find the pair with the least unspecific amplification. Fig. 3.1 shows that the product using the primer pair Tp6330F3 and Tp6330R2 at 72°C had only one fragments when analyzed by gel electrophoresis. Therefore, this pair was used during the screening process onward.

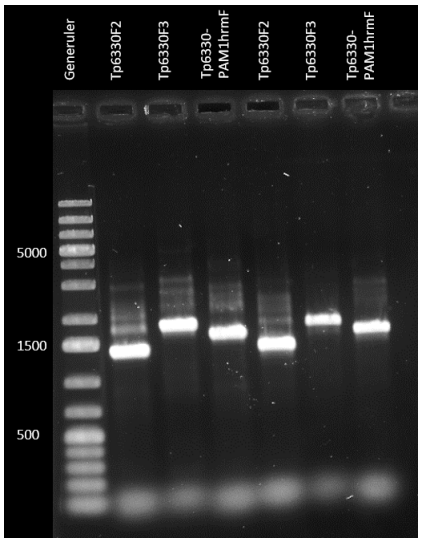

Figure 3.1: Testing of primers to find the optimal primer pair for screening of the Tp6330 gene in *T. pseudonana* clones. All tests used Tp6330R2 as reverse primer, while the forward primer used is shown in the label of each sample. The primer pair Tp6330F3 and Tp6330R2 at 72°C had the least unspecific amplification. The numbers on the left indicate fragment sizes in the generuler (1kb plus)

The results from the screening for clones edited in Tp6330PAM2 is shown in Fig. 3.2, and the results from clones attempted gene edited in Tp6330-23191PAM2 is shown in Fig. 3.3. Some of the old results are included for comparison. In the figures, the clones are numbered. Clones with only one number, e.g. C1, are from the initial conjugation plate, and was screened in previous work (a and b, both figures). The clones with two numbers were screened in this work (Fig. 3.2c and Fig. 3.3c and d). The first number represents the clone from the original conjugation plate which was then replated, and the second number represents the clone picked from this new plate.

In Fig. 3.2a, all samples contains a mix of fragments of different lengths, including the wild type. As mentioned, this is most likely due to the primers creating off target product. However, for some samples (C2, C3, C4, C6, C7, C8, C10, C11 and C12), some fragments which was not amplified from the wild type gene can be seen. This is an indication of mutations being present in the culture. Additionally, a mix of fragments of different length often indicates that the culture is mosaic. However, as the PCR product amplified from the wild type gene also contains different fragments, the results are somewhat ambiguous. Flow cytometry was performed on these clones around the same time as this screening, and the results showed that some of the clones had a shift in forward scattering A and H, indicated enlarged cells compared to the wild type.

Fig. 3.2b shows the same clones screened again about a month later. These results do not immediately indicate mutations at the same level as the previous screening results. Addi-

tionally, the wild type sample do not show the same amount of fragments as in the first screening. However, in this gel electrophoresis analysis, the gel was stained with GelRed after the electrophoresis was complete, instead of being mixed into the loading dye as in the first screening. This might have caused the already low-signal fragments to be too weak to be observed. The flow cytometry performed in previous work on these clones in between the two primary screenings did show some differences in the forward scattering H and A compared to the wild type, which is an indication of a phenotypic effect caused by mutations. The clones were replated around the same time as the second screening (Fig. 3.2) based on the results from both flow cytometry and gene amplification.

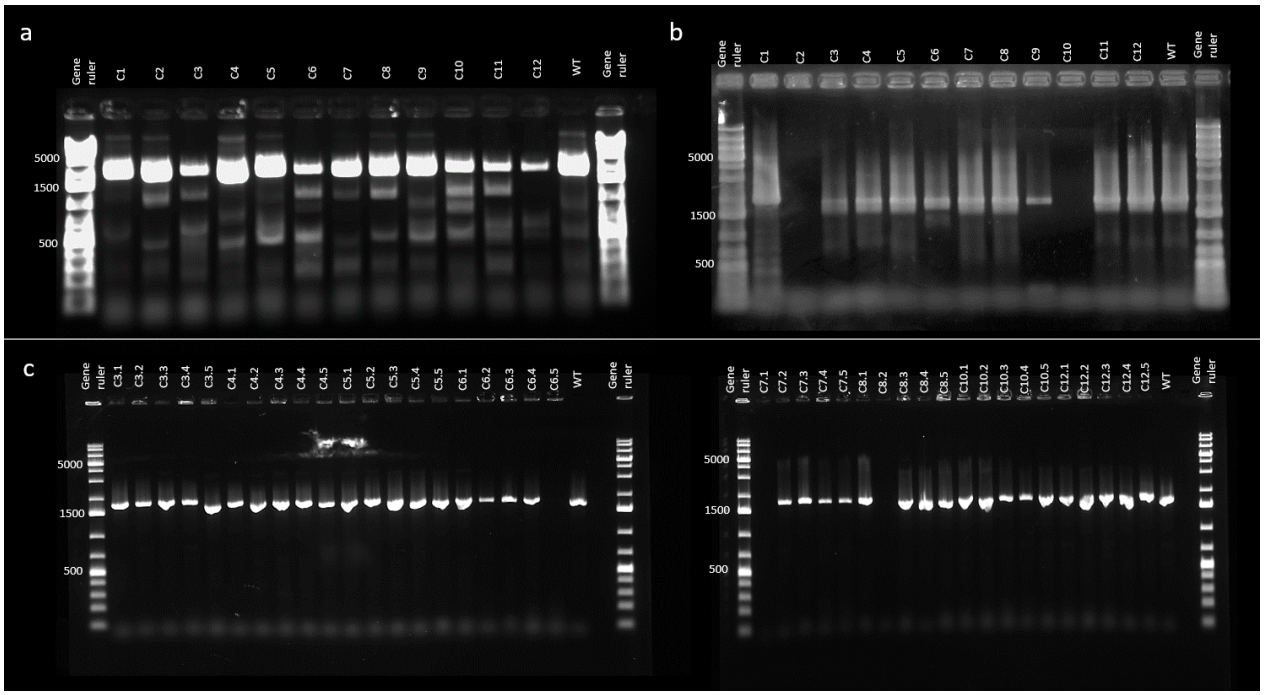

Figure 3.2: Gel electrophoresis analysis of PCR amplified *Tp6330* from *T. pseudonana* clones transformed with the pTpPUC3-Cas9 vector containing the target site Tp6330PAM2. a) The first screening of the primary clones from the original conjugation plate, using Tp6330F1 and Tp6330R2 (Expected length =1700 bp). b) Screening of the same clones as in a) about a month later, using the same primers. c) Screening of replated clones, using Tp6330F3 and Tp6330R2 (Expected length =1500 bp). The numbers on the left for each individual image indicate fragment sizes in the generuler (1kb plus).

The screening of the replated cells are shown in Fig. 3.2c. The primer pair Tp6330F3 and Tp6330R2 was used, which had previously been shown to not produce any off target product (Fig. 3.1). Some of the PCR product from the clones (e.g. C3.5, C4.2 and C12.5) seem to be of a different length than the gene amplified from the wild type, although no large differences can be seen. As gel electrophoresis is not an exact method for screening for small indel mutations, these results are inconclusive.

Fig. 3.3 compares the new and previous results of PCR amplifications of *T. pseudonana*

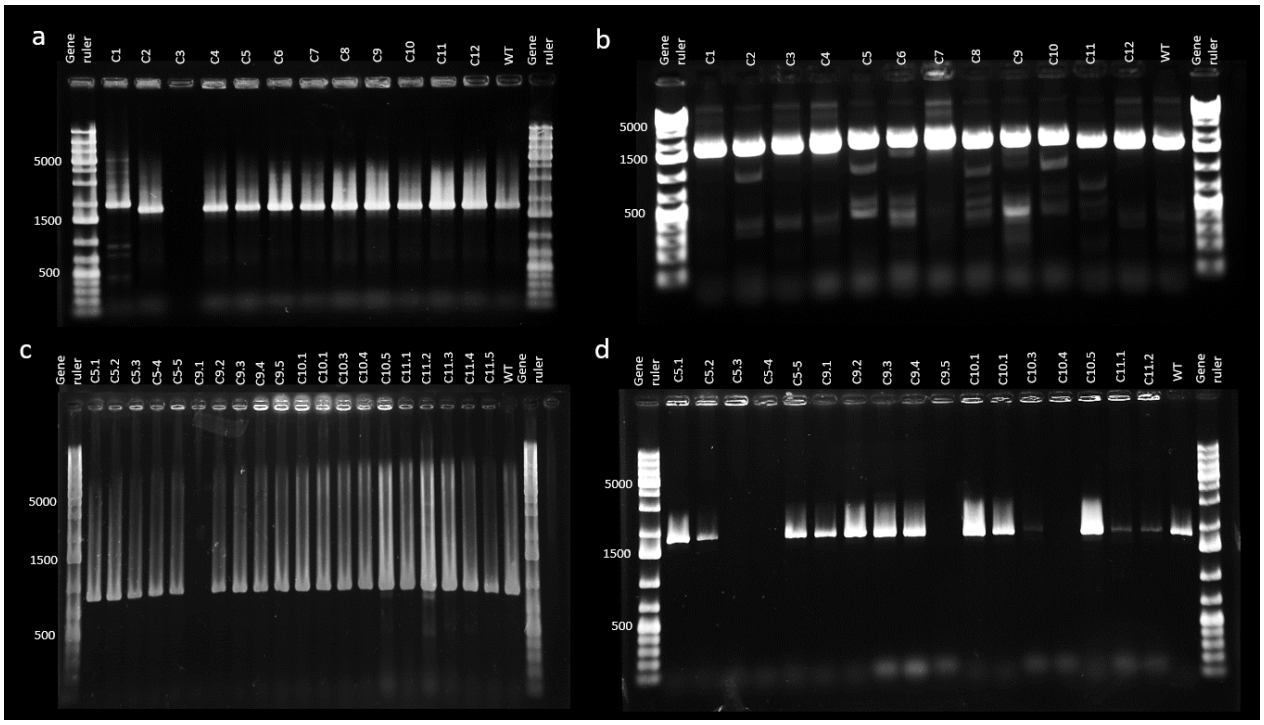

Figure 3.3: Gel electrophoresis of amplified genes from *T. pseduonana* clones transformed with the pTpPUC3 vector containing the target site Tp6330-23191PAM2. Both genes were amplified from the same clones. a) PCR amplification of *Tp23191* from primary clones from the original conjugation plate, using Tp23191F1 and Tp23191R2 (Expected length: ∼1600 bp). The wrong wild type sample was used for comparison, as this PCR product amplified *Tp6330* instead of *Tp23191*. b) PCR amplification of *Tp6330* from the same clones as in a), using Tp6330F1 and Tp6330R2 (Expected length: ∼1700 bp). c) PCR amplification of *Tp23191* from replated cells, using Tp23191F2 and Tp23191R2 (Expected length: ∼1900 bp). d) PCR amplification of *Tp6330* from the same replated cells as in c), using Tp6330F3 and Tp6330R2 (Expected length: ∼1500bp). The numbers on the left for each individual image indicate fragment sizes in the generuler (1kb plus).

clones transformed with the pTpPUC3 vector containing the target site Tp6330-23191PAM2. Fig. 3.3a and c show the PCR product where *Tp23191* was amplified, while Fig. 3.3b and d show the PCR product of *Tp6330* amplified. The two genes are amplified from the same clone, using primers specific for each gene. The wrong wild type was used in Fig. 3.3a for comparison. However, as Cas9 induced mutations creates random indel mutations of various lengths, the results can still be interpreted by comparing the product amplified from the clones with each other. The PCR product from all samples have the same length, and they also contains only one fragment. The results of amplified *Tp6330* from the same clones (Fig. 3.3b), show some differences in the PCR product amplified from the clones C2, C5, C6, C8, C9, C10 and C11 compared to the wild type gene. Based on these results, combined with the flow cytometry results, some clones were selected for replating. The results from the screening of these clones are shown in Fig. 3.3c and d. The amplified *Tp23191* of these clones do show some differences in length, although this is most likely due to the gel being uneven, as the samples follows a smooth curve (Fig. 3.3c). C10.5 and C11.2 contains two

additional fragments that can not be seen in the other samples or the wild type. Both of these fragments are shorter than the wild type fragment, which may indicate a deletion mutation. No clear differences can be seen in the amplified *Tp6330* for the replated clones (Fig. 3.3d).

All samples that had a successful PCR-amplification were submitted for Sanger Sequencing, and the results were compared to the wild type gene, which had been previously sequenced in this lab. These results verified that there were no mutations.

3.2 Localization study

The aim of this part of the experiment was to tag two genes of the silicanin protein family, *Tp23191* in group IV and *Tp20931* in group III, with a fluorescence marker to identify the intracellular location of the gene product *in vivo*. This was done by PCR amplification of the fragments (promoter, mTurq, and the gene and terminator). These fragments and the pTpPUC3 vector were then assembled, and the vector was heat shock transformed into DH5α-*E. coli*. Colony screening was then performed to look for clones containing the plasmid with all fragments correctly inserted.

A representative image of the gel electrophoresis analysis performed on the PCR product after amplification of the different fragments is shown in Figure 3.4. Different annealing temperatures were used to find the optimal, included in the label of each sample. These fragments were used to clone the vectors, using different cloning techniques. Gibson, Nested Gibson, two-step Gibson and CPEC were performed.

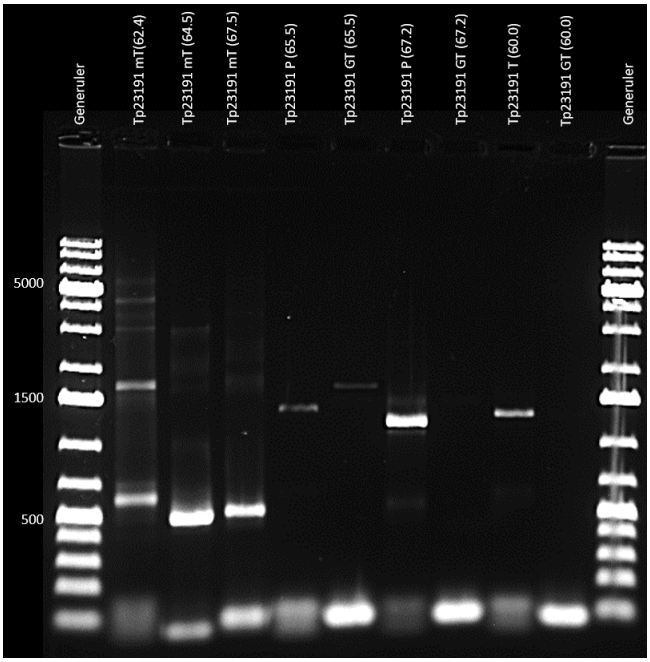

Figure 3.4: Gel electrophoresis of the fragments amplified by PCR for creating the vector used for fluorescence tagging of the silicanin gene Tp23191. Different annealing temperatures were used, indicated in the label name. mT: mTurq fragment, P: promoter fragment, GT: gene and terminator fragment. The numbers on the left for each individual image indicate fragment sizes in the generuler (1kb plus).

Results from all cloning procedures are shown in Fig. 3.5. The vector with all of the fragments inserted should be ∼11 kBp for both vectors. A weak signal around this length can be seen for two samples in Fig. 3.5b and one sample in Fig. 3.5d (samples are marked with an

asterisk). Even though no signal at the desired length was observed for most of the samples, the product was still heat shock transformed into competent *E. coli*.

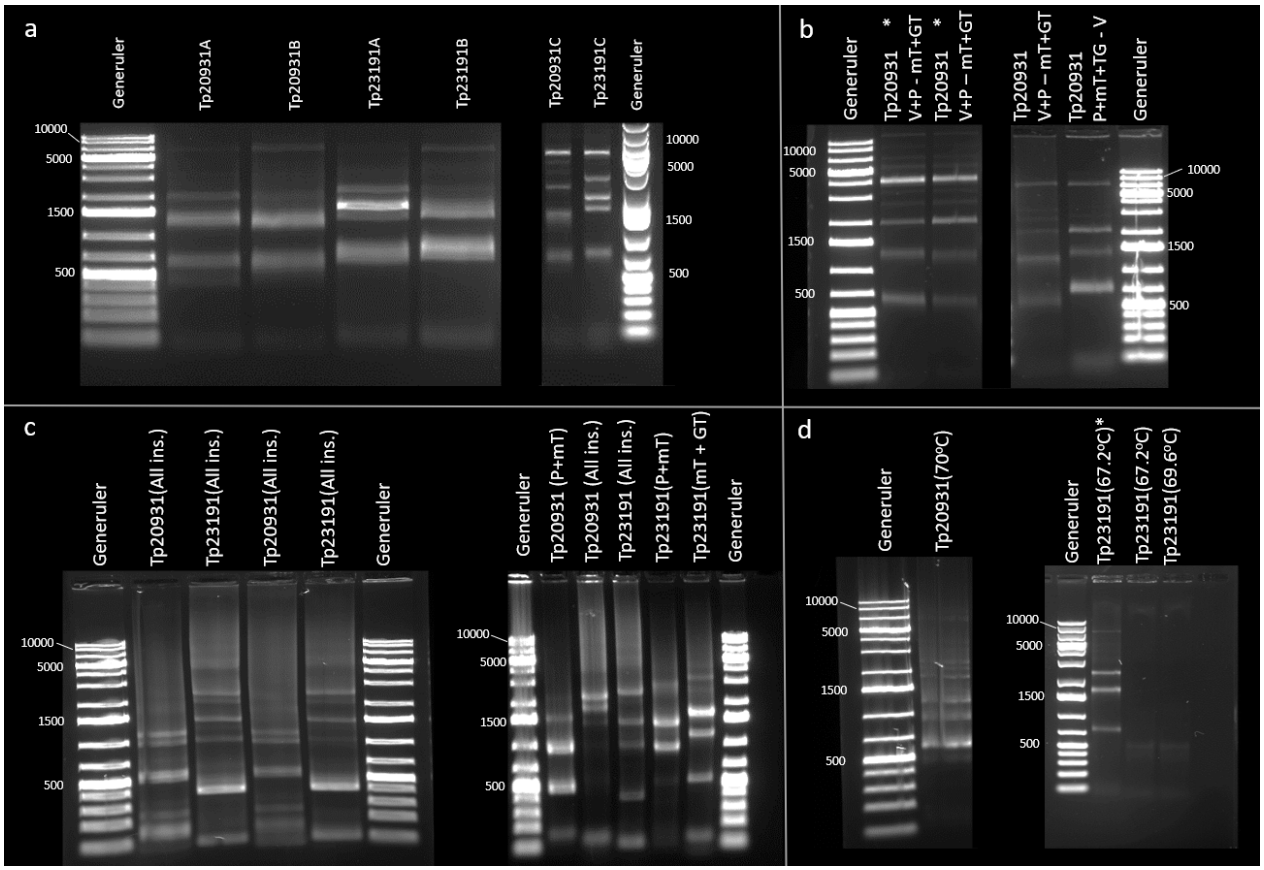

Figure 3.5: The results from the different cloning procedures performed to produce the two vectors for fluorescence tagging of the two genes Tp20931 and Tp23191. The length of each fragment is listen in Tab: 2.2. The linearized vector is 7.7 Kbp. For Tp20931-mTurq, the vector with all inserts equals 11.0 Kbp, and for Tp23191 it equals 11.8 KBP. P=Promoter, mT=mTurq, GT=Gene and terminator, V=vector, All ins=All inserts. The numbers on the left for each individual image indicate fragment sizes in the generuler (1kb plus). a) Gibson assembly. Different concentrations of the fragments were used in the different reactions. The two reactions marked with A uses the concentrations listed in Tab. 2.3. b) Two-step Gibson. Different combinations of fragments before combining the reactions were used in the different reactions, included in the label. The fragments incubated together before combining the reactions are included in the label, and the two reactions are separated by a hyphen. Two samples had a weak signal at the expected length (11kb) for the vector with all fragments inserted, marked with an asterisk. c) Nested Gibson. Primers were used to amplify some fragments that had been ligated together in a failed Gibson assembly. The primers used were the same as those used to amplify the fragment (Tab. 2.2), combining the forward primer of the first fragment and the reverse primer of the last fragment. Which fragments were amplified is included in the label of each sample. d) CPEC. Different annealing temperatures were used. One sample had a weak signal at the expected length (11kb) of the vector with all fragments inserted, marked with an asterisk.

Colony screening was performed on most of these cells, and an example of the results from this is shown in Fig. 3.6. Several bands can be seen, which implies that the primers might be creating non-specific product. However, for most samples in Fig. 3.6a and for two samples

in Fig. 3.6b a band can be seen at the expected length for correct insertion. The primers used in this colony screening, p23191F and p23191mNeonR (Fig. 3.6a), and p20931F and mNeonc20931R2 (Fig. 3.6b), amplify the *Tp23191* promoter fragment alone and *Tp20931* promoter and *mTurq* fragments combined, respectively. The plasmid was extracted from heat shock transformed *E. coli* cells whether or not colony screening was performed on them, and sent in for Sanger sequencing. None of the samples showed a correct insertion of all fragments.

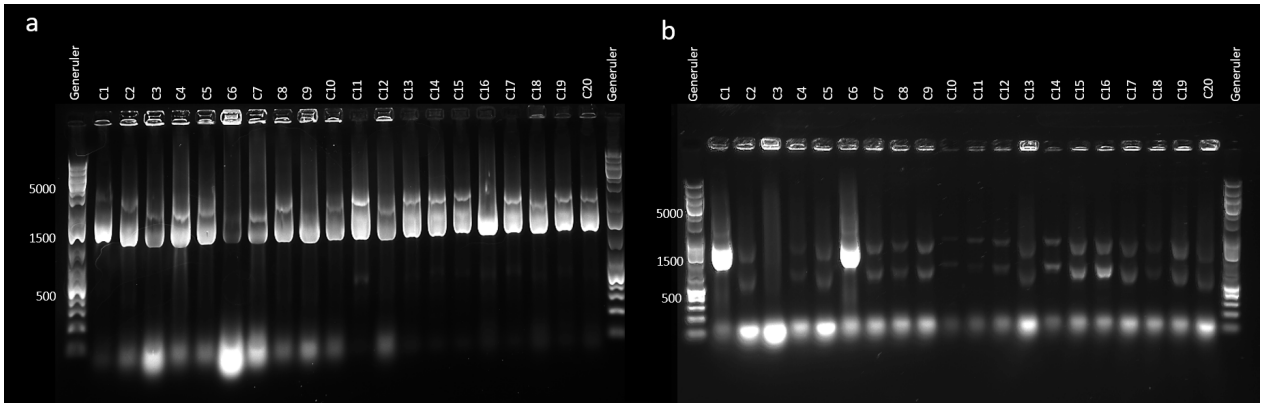

Figure 3.6: Gel electrophoresis of some of the colony screenings performed on competent *E. coli*- cells transformed with Gibson assembly product. a) Tp23191-mTurq assembly. The primers p23191F and p23191mNeonR was used. Expected length = 1500bp. b) Tp20931-mTurq assembly. The primers p20931F and mNeonc20931R2 was used. Expected length = 1400bp. The numbers on the left for each individual image indicate fragment sizes in the generuler (1kb plus).

3.3 The effects of Aluminium and Cadmium on the frustule

3.3.1 Photosystem II efficiency

The mean quantum yields for the control and for each concentration of Al and Cd are shown in Fig. 3.7. A student T-test was performed (α=0.05) on these samples compared to the control. No significant differences was found. The data behind this graph and the p-values (T.test) is attached in Appendix A.6.

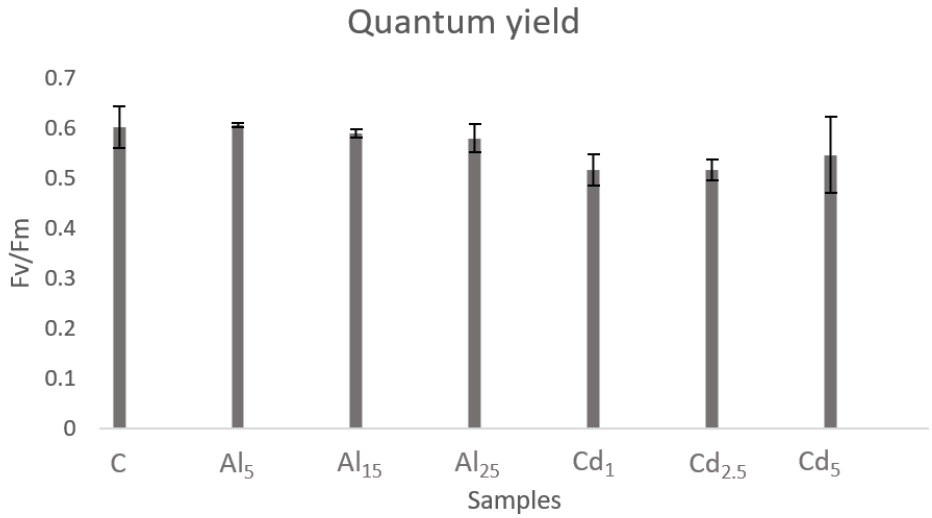

Figure 3.7: The mean quantum yield (=Fv/Fm) for the control and all concentrations of Al and Cd. Standard deviation is included. No significant differences was found using T-test (α=0.05).

3.3.2 Scanning electron microscopy

It was originally planned to take SEM images of all samples, but because of time restrictions only the three replica of the control (C), Al_{25} and Cd_5 samples was chosen. In total, five SEM images of C, Al_{25} and Cd_5 was used for calculating the valve, fultoportulae, rimoportulae and cribrum pore diameter, and costae width. A representative selection of these images are shown in Fig. 3.8. The rest of these images and an example of how the forementioned biosilica structures was measured are attached in Appendix A.7. The mean value of these measurements is given in Tab. 3.1. All parameters were significantly different in both treatments, except for cribrum pore diameter in Al_{25}, and fultoportulae and rimoportulae diameter in Cd_5. Both metals increased the diameter of the valve and the width of the costae, while fultuportulae (for Al_{25}), rimoportulae (for Al_{25}) and cribrum pore (for Cd_5) decreased in both treatments.

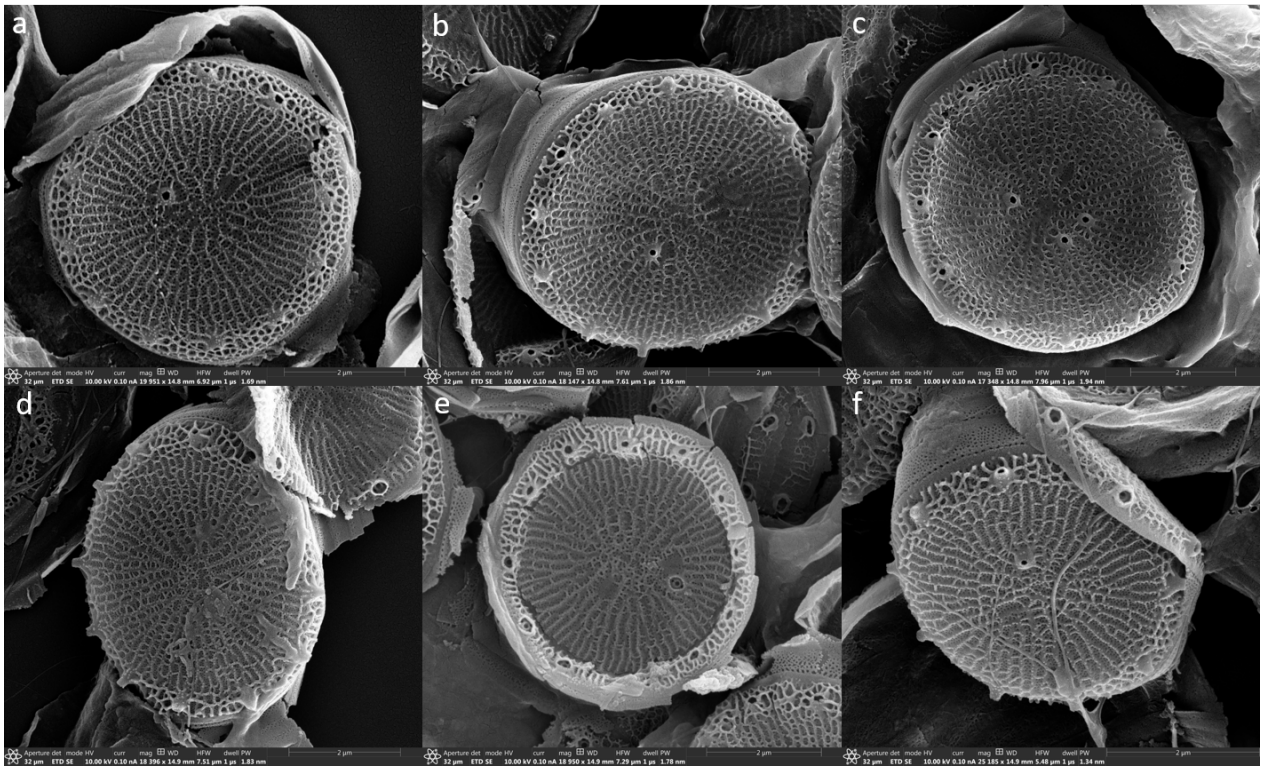

Figure 3.8: Comparison of biosilica morphology of the frustule affected by metals. a) and d) Control sample. Scale bar = 2 μm. b) and e) Al_{25} sample. Scale bar = 2 μm. c) and e) Cd_5 sample. Scale bar = 2 μm, respectively.

Biosilica structure	C (μm ± std.dev)	Al_{25} (μm ± std.dev) ; p-value	Cd_5 (μm ± std.dev) ; p-value
Valve diameter	3.8± 0.3	4.4± 0.4 ; 1.72E-4	4.3± 0.7 ; 0.0150
Costae width	0.05±0.01	0.065± 0.01 ; 5.19E-10	0.07± 0.01 ; 7.20E-14
Fultoportulae diameter	0.16± 0.04	0.1± 0.03 ; 6.24E-4	0.12± 0.06 ; 0.0512
Rimoportulae diameter	0.13± 0.05	0.10± 0.03 ; 7.61E-3	0.12± 0.05 ; 0.95
Cribrum pore diameter	0.026± 0.005	0.026± 0.005 ; 0.58	0.024± 0.005 ; 1.78E-3

Table 3.1: The mean diameter or width of biosilica structures in the control (C), 25 μg/L Aluminium (Al_{25}) and 5 μg/L cadmium (Cd_5) samples and the standard deviation for each measurement. A T-test (α=0.05) was performed for the values of Al_{25} and Cd_5 compared to the control. All measurement are statistically different except for cribrum pore diameter for Al_{25}, rimoportulae diameter for Cd_5 and fultoportulae diameter for Cd_5.

Discussion

This report presents a three-part experiment studying the morphology of the frustule and the components involved in its synthesis. Firstly, single and double knock out *Thalassiosira pseudonana* cell lines of *Tp23191* and *Tp6330* in group IV of the silicanin protein family were attempted to be identified and isolated. In the second part, two proteins of the silicanin protein family, Tp23191 in group IV and Tp20931 in group III was attempted to be tagged with the fluorescence protein mTurq to observe their subcellular location *in vivo*. Lastly, the effects of aluminium and cadmium on the biosynthesis of the valve was studied by image analysis of SEM images of *T. pseudonana* cells growing in ESAW medium supplemented by 25 μg/L aluminium or 5 μg/L cadmium.

4.1 Single and double knock out of Tp23191 and Tp6330

The results from the previous work indicated that there were mutations in the *T. pseudonana* cells transformed with the pTpPUC3Cas9 plasmid containing the target site Tp6330PAM2 or Tp6330-23191PAM2. For the double knock out transformant, the results only indicated mutations in the Tp6330 gene. However, after closer examination by re-screening the primary cells at a later time and screening the replated clones, the transformants did not show any mutations. For the samples that did indicate mutations (Fig. 3.2a and Fig. 3.3b), several bands can be seen. This may be due in some part by the primer creating nonspecific product, since the wild type also contains several bands. However, fragments that could not be amplified from the wild type DNA is also observed in the clones. A mix of bands of two different lengths usually implies an heterozygous mutant, with either one wild type allele and one mutant allele, or two different mutations on each allele [Bhattacharya and Van Meir, 2019]. As these clones were picked from the plate and incubated in liquid medium at an early stage, and was incubated under the influence of antibiotics for some time, the Cas9 protein is most likely continuously transcribed and will keep cutting the genome until the plasmid is lost due to removal of antibiotic pressure. As the cells will also be continuously dividing, there will most likely be a mix of cells with different mutations, which corresponds with the results in the Sin1-knock out experiment by [Görlich et al., 2019], where primary clones were found to be mosaic. Some of the bands that is not replicated from the wild type sample may therefore be caused by mutants being present in the culture.

Interestingly, the screening of the same clones at a later time (Fig. 3.2b), about 1 month after the first screening), did not show the same mix of bands of different lengths. The reason for this is not clear. One possible explanation is that the mutants grew slower than the wild

type, and since *T. pseudonana* cultures needs to be fed weekly by transferring an aliquot to fresh medium, it is possible that the mutants were diluted to a point where they could no longer be detected. Although Görlich et al. [2019] reported that Sin1 knock out mutants did not exhibit a significantly different growth rate compared to the wild type, cell growth have been shown to vary depending on the cell size even within a species, with larger cells growing more slowly [Martin-Jézéquel et al., 2003]. The flow cytometry performed on clones in the previous work indicated that the transformants were larger than the wild type cells, which may have caused them to grow slower. This was not confirmed by calculating the growth rate, as no stable mutant was isolated.

Additionally, the wild type sample did not contain several fragments in the second screening in the previous work (Fig. 3.2b) like it did in the first one (Fig. 3.2a), even though the same primers and PCR program was used. It is unknown why these results were so different, as the wild type sample used was the same in both screening. One explenation could be that this is caused by the gel red being added after the gel was run, instead of being mixed into the loading dye as in the first screening. The signal for these other fragments might therefore be too weak to be observed clearly. It is therefore possible that the clones as well contain some other fragments in addition to the strong signal from the wild type, although too weak to be seen.

The results from the screening of the replated cells (Fig. 3.2c) do show some, although small, differences in length between some clones and the wild type. The results from the Sanger sequencing did however not show any mutations in the clones that were screened. These differences might therefore be caused by the samples not running evenly, or other errors which caused false positive results. Previous screening performed by both gel electrophoresis and flow cytometry indicated mutants being present, although ambiguously. However, the wild type fragment exhibited a much stronger signal in the gel electrophoresis for the clone samples than the fragments that could not be amplified in the wild type sample. One possibility is that there are mutations present, but in very small concentrations, and that it was not detected as not enough clones were screened from the replicated cells.

The gel electrophoresis analysis performed on the amplified *Tp23191* from clones transformed with the pTpPUC3-Cas9 vector containing the target site Tp6330-23191PAM2 did not indicate any mutations. The wrong wild type sample was used in Fig. 3.3a, so it is not possible to draw any conclusions by comparing the clone samples with the wild type. However, since Cas9 induce random indel mutations, one would expect mutants to be different from each other, and thus the results can be interpreted by comparing the different clones to each other. All PCR amplifications of the clones created fragments of the same length. Additionally, all samples contain only one fragment instead of the mix of fragments of different lengths as

can be seen in Fig. 3.2a and Fig. 3.3b. As these clones were not replated, they would most likely contain several fragments if there had been mutants present. The replated cells did show some differences is PCR product length, but this is probably caused by the gel being uneven, as these fragments follows a smooth curve. The Sanger sequencing results confirms that there were no mutations in these clones. However, in two clones (C10.5 and C11.2), two additional fragments can be seen. This is unexpected, as replated cells should be single-celled colonies. It is possible to see two bands in single-celled colonies if they are heterozygous, but three bands is not expected if the colonies originates from a single cell. Additionally, these cells had not been under antibiotic pressure for some time and the pTpPUC3-Cas9 plasmid should therefore not be expressed, meaning Cas9 is most likely not present to induce new mutations. As the Sanger sequencing confirmed there were no mutations in these cells, it is possible that these fragments are caused by contamination, or some unspecific amplification that was more effectively amplified in these samples.

Conversely, in some PCR amplifications of *Tp6330* in the primary clones Fig. (3.3b), fragments that could not be amplified in the wild type could be seen. The target sites are identical in both genes, and should therefore theoretically have the same efficiency in both genes. This can indicate that *Tp23191* is an essential gene. However, if a gene attempted to be knocked out is essential, one can still detect small in-frame mutations in transformants [You et al., 2020]. Since no such mutations were observed, this hypothesis can not be confirmed. If this gene is not essential, it is possible that there were no mutations in either gene, and the fragments that can only be seen in the clones and not the wild type is caused by the signal of these fragments in the wild type is too weak to be seen. This could then imply that the indications of mutations observed in *Tp6330* for clones transformed with both vectors discussed (target sites in Tp6330PAM2 and Tp23191-6330PAM2) were coincidental.

No mutants could be identified and isolated in this experiment. Several other students performed similar experiments using the same vector, although with different target sites, and none proved to be successful. This strongly indicates that the vector that was used was not efficient. Studies performed in another project investigating the expression of different genes in the vector showed that cells growing under pressure of nourseothricin (Nou) had a very low expression of the nourseothricin N-acetyl-transferase (nat) gene. These results indicate that the cells might have a different mechanism for resisting the antibiotic, or that *nat* is a highly effective antibiotic resistance gene. This could result in only some clones of a culture cultivating the plasmid, while most clones lost it [Bentzen, 2021]. This might explain why no mutants could be observed. As mentioned above, the wild type signal was much stronger in the clones that did indicate mutations. As aliquotes of these cultures were transferred to fresh medium weekly, and after a while to medium with no antibiotics,

the possible mutants might be diluted to a point where they could no longer be detected. Although Nou is a commonly used selection marker for gene manipulation in *T. pseudonana* cells and is proved to be efficient [see e.g. Karas et al. 2015], antibiotic sensitivity have been found to be dependent on cell density [Miyagawa-Yamaguchi et al. 2011], which may also have impacted the concentration of wild type cells in the cultures.

In this experiment, and the similar experiments performed by my co-students, genetic manipulation in *T. pseudonana* by bacterial conjugation proved to be highly ineffective. Biolistic shooting is an optional method for gene manipulation, and is a well implemented method. This method will, however, permanently integrate the plasmid into the cell's genome. The cells will therefore exhibit a continuous expression of Cas9, which will continue to cut the host genome until the target sequence is no longer recognizable. It may therefore be challenging to obtain stable mutants. Additionally, the permanent expression of Cas9 may increase the chance of off-target editing. This method is also requires expensive equipment to be performed. By using biolistic shooting, one will also experience large insertions at the DSB, caused by partial insertion of the vector [Kroth et al., 2018], and may make mutants easier to detect during screening. Although biolistic shooting has been found to be more efficient, bacterial conjugation is still the preferable method because of the negative effects caused by permanent integration of the episome in the host genome when using biolistic shooting.

4.2 Localization study

I attempted to clone the plasmid that was going to be used as a vector to transform *Tp23191* and *Tp20931* tagged with the fluorescence protein mTurq into *T. pseudonana* cells using several different cloning methods, and each method was performed several times with some modifications in incubation time or annealing temperature. None of these methods resulted in obtaining a plasmid with all fragments correctly inserted. For some of the earlier results, e.g. Fig. 3.5a, right, a strong signal could be seen at $\sim$ 7000bp, which is the length of the plasmid. The vector was therefore tested on a gel, and was found to not be completely digested. A new restriction digestion reaction was therefore performed with a longer incubation time, and when examined by gel electrohoresis no undigested plasmid could be observed. This might explain the false positive clones appearing on the selection plate from the first cloning procedure, and why no clones contained the plasmid with the fragments correctly inserted.

A very faint signal at the expected length of the plasmid with all inserts can be seen for the samples in Fig. 3.5b (sample 1, 2 and 4) and Fig. 3.5d (sample 2). This signal was unfortunately not strong enough to be seen before the pictures were digitally altered to increase the light and contrast. It is therefore possible that the cloning was somewhat successful using

these methods (two-step Gibson and CPEC), although not enough clones were screened after this product was heat shock transformed into competent *E. coli* cells.

As two-step Gibson and CPEC was the only methods that indicated the correct insertion of all of the fragment, these methods may be the most effective for cloning a plasmid with several inserts (4 fragments to be assembled in total, including the vector). The reason why this cloning showed to be problematic may be because there are several ends which needs to be ligated together in the correct manner. Therefore, the chance of this happening between all four fragments in the same plasmid might therefore be quite small. The two-step Gibson method may be more effective as this method incubated two reactions separately for some time. Some of the fragments in these reaction might therefore be ligated together before the reactions are combined and all fragments are present in the same reaction, which means that after combining these reactions, there is only two fragments that needs to be ligated together, increasing the chance of correct insertion of all fragments. CPEC works in a different manner. This method uses a polymerase, and the overlaps between each fragment works as a primer for the extension caused by the polymerase in each cycle. Therefore, for each cycle there will be produced more and more product of some fragments ligated together, and unused fragments which have yet not been ligated to another fragment may function as a primer in the following cycles, increasing the chance of all fragments being combined, until the full plasmid is cloned. It is possible that the cloning would be more successful by first ligating the fragments together either by Gibson assembly or a modified CPEC (non-circular??) before attempting to insert them into the vector. This way, the product can be analyzed before insertion to ensure all fragments are ligated together. If gel electrophoresis analysis show that several products is produced during the ligation, the desired product can be extracted by cutting out the fragment from the gel. This way, we can assure that all inserts into the backbone contains all fragments.

4.3 The effects of aluminium and cadmium on the frustule

Scanning electron microscopy images were taken of *T. pseudoana* cells growing in ESAW medium supplemented with 25 μg/L aluminium or 5 μg/L cadmium. The images were used to calculate mean diameter of the valve, fultoportulae, rimoportulae, cribrum pores and the width of the costae. All values were significantly different, except for cribrum pore diameter for Al_{25}, and rimoportulae and fultoportulae diameter for Cd_5. Both metals increased the diameter of the valve and the width of the costae, while fultoportulae, rimoportulae and cribrum pore (for Cd_5) decreased in both samples.

Soleimani et al. [2020] found that Al^{3+} ions increased the silica uptake for the diatom *Pinnularia sp.* while also decreasing the growth rate, causing a higher content of inorganic content per cell. This led to a decrease in the small pores of the frustule. Although the cribrum pore diameter was the same for Al_{25} as the control, the fultoportulae and rimoportulae decreased significantly. However, a higher content of inorganic content per cell is likely to affect small pores as well, unless these are more strictly controlled. The diameter ranged from 17-44 nm in the control, and from 18-36 nm in the aluminium samples, which is more than two fold for both samples. It is possible that the angle the images were taken at or the contrast/light settings may have impacted the appearance of the pores. As the mean diameter of the pores in the control sample is larger than what is found in the literature (18 nm, Hildebrand et al. [2006]), it is hard to say how accurate these measurement was. See Appendix A.7 for an example of how the different parameters were measured.

Both metals increased the mean diameter of the valve, which is oppositely of what is found in the diatom species *Asterionella formosa* and *Asterionella ralfsii* [Cattaneo et al. 2004, Gensemer 1990]. For this parameter, the control sample was consistent with the findings in the literature (3.8μm, Hildebrand et al. [2006]). It is unknown why the diameter was affected differently in other species. Although these diatoms are greatly different in shape and morphology diatom compared to *T. pseudonana* as these are pennate diatoms, one main difference is that cell diameter decreases over generations as new hypotheca are synthesized within the parent cell restricted by the diameter of the mature valves, which induce auxosporulation and sexual reproduction, in contrast to *T. pseudonana* cells, which maintain its size during division. Although this process is not well understood, it is hypothesized that silacidins and other regulators are used to keep diameter low, especially under stressful conditions such as nutrient limitations. Gensemer et al. [1993] found that at pH 6 with 20 μmol/L Al, the V_m (maximum Si uptake rates) was more than twice as high as samples with no Al. It is possible that aluminium therefore somehow have affected the size regulators in the cells, and therefore caused the increase in the cell.

Heredia et al. [2012] studied the morphological effects of cadmium in the diatom species *Nitzschia palea*. They found that the pores became smaller at increasing cadmium concentrations. This is the same effect observed in this experiment, were fultoportulae and cribrum pore diameter were significantly different in Cd_5. Unlike alluminium, Heredia et al. did not find evidence for cadmium being incorporated into the biosilica of the frustule. Instead, the structural changes seemed to depend on growth (for macroscopic features), and the content of biomolecules in the biosilica. Additionally, the costae became more silicified in this diatom species, which is also the same effect as observed in this experiment. Both Heredia et al. [2012] and Radić et al. [2020] observed no differences in size or overall shape of the frustule.

The silica uptake and growth rates were supposed to be measured, but this had to be omitted because we had no access to a working flow cytometer at the time, and counting the cells manually was too time consuming. It would be interesting to see if *T. pseudonana* cells are affected similarly to the other diatoms mentioned above. If aluminium increases the inorganic content of the valve, this would explain the decrease in fultoportulae and rimoportulae diameter. The growth rate could also indicate toxicity of these two metals. The *T. pseudonana* cells were viable at all concentrations for both metals, and the quantum yield was not significantly different for either metal. However, as the quantum yield was indeed lower for the cadmium samples, it is possible that too few samples were measured for them to be significant.

Conclusion

The aim of the single and double knock out experiment was to isolate and identify stable mutant *Thalassiosira pseudonana* cell lines with no functional *Tp23191* and *Tp6330* gene product. Some preliminary results from screening for mutants by amplification of the genes to observe genotypic differences and flow cytometry to observe phenotypic differences indicated mutations. These clones were replated and screened again by amplification of the genes. No knock out mutants were obtained. As this was the problem in other studies using the same plasmid as a backbone to transformed *T. pseudonana* cells with Cas9 and its associated sgRNA with target sequences in several other genes, it is hypothesized that the pTpPUC3 vector was not efficient. It is possible that the low efficiency is caused by *nat* being so effective in resisting the antibiotic pressure, which results in only a few clones of a culture cultivating the plasmid, while most clones lose it.

A localization study was performed by attempting to tag the silicanin proteins Tp20931 and Tp23191 with the fluorescence protein to identify their subcellular location *in vivo*. Four different cloning methods were performed, but none proved to be successful. It is possible that the amount of fragments needed to be ligated together is the reason for the difficulties in cloning the plasmids. The cloning may therefore be more successful if the inserts are ligated together before attempting to insert them into the backbone. This can be done by either a Gibson assembly or a modified CPEC reaction with only the fragments added to the reaction. The product can then be analyzed by gel electrophoresis, and if different products are produced, the desired product can be extracted from the gel and used in a new Gibson assembly, containing only the vector and one insert consisting of all fragments combined.

The effect of cadmium and aluminium on the morphology of the frustule of *T. pseudonana* cells were studied by image analysis of SEM images. Valve diameter and costae width increased significantly in both treatments, while fultoportulae and rimoportulae decreased significantly in Al_{25} but were insignificantly different in Cd_5, and cribrum pore diameter decreased significantly in Cd_5. The decrease in the different pores in Al_{25} is most likely caused by an increase in inorganic content in the valve, while this was most likely caused by the content of biomolecules in the biosilica in Cd_5 treatments. The observed increase in valve diameter in both treatments is unusual compared to the literature, and the cause for this is not known.

Bibliography

Anonymous (1702). Two letters from a gentleman in the country, relating to Mr. Leeuwenhoek's letter in transaction, no 283. *Phil. Trans. R. Soc. Lond. B*, 23:1494–1501.

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

Armbrust, E. V., Berges, J. A., Bowler, C., Green, B. R., et al. (2004). The genome of the diatom *Thalassiosira pseudonana*: Ecology, evolution, and metabolism. *Science*, 306:79–86.

Bedoshvili, Y. D. and Likhoshway, Y. V. (2019). Cellular mechanisms of diatom valve morphogenesis. In *Diatoms: Fundamentals and applications*, chapter 5, pages 99–114. John Wiley & Sons, Ltd.

Benoiston, A. S., Ibarbalz, F. M., Bittner, L., Guidi, L., et al. (2017). The evolution of diatoms and their biogeochemical functions. *Phil. Trans. R. Soc. B*, 372(20160397).

Bentzen, K. L. L. (2021). Functional studies of two genes encoding closely related group ii silicanins in the diatom *Thalassiosira pseudonana* - msc thesis. *Norwegian University of Technology*.

Berges, J. A., Franklin, D. J., and Harrison, P. J. (2001). Evolution of an artificial seawater medium: Improvements in enriched seawater, artificial water over the last two decades. *J. Phycol.*, 37:1138–1145.

Bhattacharya, B. and Van Meir, E. G. (2019). A simple genotyping method to detect small CRISPR-Cas9 induced indels by agarose gel electrophoresis. *Sci. Rep.*, 9(4437).

Bowler, C., Vardi, A., and Allen, A. E. (2009). Oceanographic and biogeochemical insights from diatom genomes. *Annu. Rev. Mar. Sci.*, 2:333–365.

Brembu, T., Chauton, M. S., Winge, P., Bones, A. M., et al. (2017). Dynamic responses to silicon in *Thalasiossira pseudonana* - identification, characterisation and classification of signature genes and their corresponding protein motifs. *Sci. Rep.*, 7(4865).

Cattaneo, A., Couillard, Y., Wunsam, S., and Courcelles, M. (2004). Diatom taxonomic and morphological changes as indicators of metal pollution and recovery in Lac Dufault (Québec, Canada). *J. Paleolimnol.*, 32:163–175.

Chacón-Baca, E., Beraldi-Campesi, H., Cevallos-Ferriz, S. R. S., Knoll, A. H., and Golubic, S. (2002). 70 Ma nonmarine diatoms from northern Mexico. *Geology*, 30:279–281.

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

Ehrlich, H. and Witowski, A. (2015). Biomineralization in diatoms: The organic templates. In *Evolution of lightweight structures*, chapter 3, pages 39–47. Springer Dordrecht Heidelberg New York London.

Fattorini, N. and Maier, Q. G. (2021). Targeting of proteins to the cell wall of the diatom *Thalassiosira Pseudonana*. *Discov. Mater.*, 1(5).

Gensemer, R. W. (1990). Role of aluminum and growth rate on changes in cell size and silica content of silica-limited populations of *Asterionella ralfsii* var. *Americana* (Bacillariophyceae). *J. Phycol*, 26:250–258.

Gensemer, R. W., Smith, R. E. H., and Duthie, H. C. (1993). Comparative effects of pH and aluminium on silica-limited growth and nutrient uptake in *Asterionella ralfsii* var. *Americana* (bacillariophyceae). *J. Phycol.*, 29:36–44.

Gosh, D., Venkataramani, P., Nandi, S., and Bhattecharjee, S. (2019). CRISPR–Cas9 a boon or bane: the bumpy road ahead to cancer therapeutics. *Cancer Cell Int*, 19(12).

Görlich, S., Pawolski, D., Zlotnikow, I., and Kröger, N. (2019). Control of biosilica morphology and mechanical performance by the conserved diatom gene *Silicanin-1*. *Commun. Biol.*, 2(245).

Heintze, C., Formanek, P., Haupstein, J., Rellinghaus, B., et al. (2020). An intimate view into the silica deposition vesicles of diatoms. *BMC mat.*, 2(11).

Heredia, A., Figueira, E., Rodrigues, C. T., Rodríguez-Galván, A., et al. (2012). Cd^{2+} affects the growth, hierarchical structure and peptide composition of the biosilica of the freshwater diatom *Nitzschia palea* (Kützing) W. Smith. *Phycol. Research*, 60:229–240.

Hildebrand, M., Lerch, S., and Shrestha, R. (2018). Understanding diatom cell wall silicification - moving forward. *Front. Mar. Sci.*, 5(125).

Hildebrand, M., York, E., Kelz, J. I., Davis, A. K., et al. (2006). Nanoscale control of silica morphology and three-dimensional structure during diatom cell wall formation. *J. Mat, Res.*, 21:2689–2698.

Hsu, P. D., Lander, E. S., and Zhang, F. (2014). Development and Applications of CRISPR-Cas9 for Genome Engineering. *Cell*, 157:1262–1278.

Karas, B. J., Diner, R. E., Lefebvre, S. C., McQuaid, J., et al. (2015). Designer diatom episomes delivered by bacterial conjugation. *Nat. Commun.*, 6(6925).

Kotzsch, A., Gröger, P., Pawolski, D., Bomans, P. H. H., et al. (2017). Silicanin-1 is a conserved diatom membrane protein involved in silica biomineralization. *BMC Biology*, 15(65).

Kotzsch, A., Pawolski, D., Milentyev, A., Poulsen, N., et al. (2016). Biochemical composition and assembly of biosilica-associated insoluble organic matrices from the diatom *Thalassiosira pseudonana*. *J. Biol. Chem*, 291:4982–4997.

Kroth, P. G., Bones, A. M., Daboussi, F., Ferrante, M. I., et al. (2018). Genome editing in diatoms: achievements and goals. *Plant. Cell. Rep.*, 10:1401–1408.

Kröger, N., Deutzmann, R., Bergsdorf, C., and Sumper, M. (2000). Species-specific polyamines from diatoms control silica morphology. *Proc. Natl. Acad. Sci. USA*, 97:14133–14138.

Kröger, N., Lorenz, S., Brunner, E., and Sumper, M. (2002). Self-assembly of highly phosphorylated silaffins and their function in biosilica morphogenesis. *Science*, 298:584–586.

Kröger, N. and Poulsen, N. (2008). Diatoms—From cell wall biogenesis to nanotechnology. *Annu. Rev. Genet*, 42:83–107.

Liu, Z., Dong, H., Cui, Y., Cong, L., et al. (2020). Application of different types of CRISPR/Cas-based systems in bacteria. *Microb. Cell. Fact.*, 19(172).

Lui, M., Rehman, S., Tang, X., Gu, K., et al. (2019). Methodologies for Improving HDR Efficiency. *Front. Genet.*, 7(691).

Mann, D. G. and Vanormelingen, P. (2013). An inordinate fondness? The number, distributions, and origins of diatom species. *Eukaryot. Microbiol*, 60:414–420.

Martin-Jézéquel, V., Hildebrand, M., and Brzezinski, M. A. (2003). Silicon metabolism in diatoms: Implications for growth. *J. Phycol*, 36:821–840.

Miyagawa-Yamaguchi, A., Okami, T., Kira, N., Yamaguchi, H., et al. (2011). Stable nuclear transformation of the diatom chaetoceros sp. *Phycol. Research*, 59:113–119.

Mock, T., Samanta, M. P., Iverson, V., Berthiaume, C., et al. (2008). Whole-genome expression profiling of the marine diatom *Thalassiosira pseudonana* identifies genes involved in silicon bioprocesses. *PNAS*, 105:1579–1584.

Naito, Y., Hino, K., Bono, H., and Ui-Tei, K. (2015). CRISPRdirect: software for designing CRISPR/Cas guide RNA with reduced off-target sites. *Bioinformatics*, 31:1120–1123.

Poulsen, N. and Kröger, N. (2004). Silica morphogenesis by alternative processing of silaffins in the diatom *Thalassiosira pseudonana*. *J. Biol. Chem*, 279:42993–42999.

Radić, T. M., Čačković, A., Penezić, A., Dautović, J., et al. (2020). Physiological and morphological response of marine diatom *Cylindrotheca closterium* (bacillariophyceae) exposed to cadmium. *Europ. J. Phycol.*, 56:24–36.

Sander, J. D. and Joung, J. K. (2014). CRISPR-Cas systems for editing, regulating and targeting genomes. *Nat. Biotechnol.*, 32:347–355.

Smetacek, V. (1999). Revolution in the ocean. *Nature*, 401(647).

Soleimani, M., Rutten, L., Maddala, S. P., Wu, H., and other (2020). Modifying the thickness, pore size, and composition of diatom frustule in *Pinnularia sp.* with Al^{3+} ions. *Sci. Rep.*, 10(19498).

Su, Y., Lundholm, N., and Ellegaard, M. (2018). Effects of abiotic factors on the nanostructure of diatom frustules - ranges and variability. *Appl. Microbiol. Biotechnol.*, 102:5889–5899.

Sumper, M. and Brunner, E. (2008). Silica biomineralisation in diatoms: The model organism *Thalassiosira pseudonana*. *ChemBioChem*, 9:1187–1194.

Tesson, B., Lerch, S. J. L., and Hildebrand, M. (2017). Characterization of a new protein family associated with the silica deposition vesicle membrane enables genetic manipulation of diatom silica. *Sci. Rep.*, 7(13457).

Wenzl, S., Hett, R., Richthammer, R., and Sumper, M. (2008). Silacidins: Highly acidic phosphopeptides from diatom shells assist in silica precipitation *In Vitro*. *Angew. Chem. Int. Ed.*, 47:1729–1732.

You, Y., Ramachandra, S. G., and Jin, T. (2020). A CRISPR-based method for testing the essentiality of a gene. *Sci. Rep.*, 10(14779).

Appendix

A.1 Materials

Liquid Luria-Bertani medium (LB-medium)

	Amount (g/L dH_2O)	Manufacturer
Trypton	10	VWR Life Science
Yeast extract	5	Oxoid
NaCl	5	Sigma Life Science
Bactoagar (only for agar plates)	15	Oxoid

f/2 + Si-medium

Stocks		Amount per liter
(1)	**Trace elements (chelated)**	
	Na_2EDTA	4.36g
	$FeCl_2 \cdot 6H_2O$	3.15g
	$CuSO_4 \cdot 5H_2O$	0.01g
	$ZnSO_4 \cdot 7H_2O$	0.022g
	$CoCl_2 \cdot 6H_2O$	0.01g
	$MnCl_2 \cdot 4H_2O$	0.18g
	$Na_2MO_2 \cdot 2H_2O$	0.006g
(2)	**Vitamin mix**	
	cyanocobalamin (Vitamin B_{12})	0.0005g
	Thiamine HCl (Vitamin B_1)	0.1g
	Biotin	0.0005g
(3)	**Sodium metasilicate**	
	$Na_2SIO_3 \cdot 9H_2O$	30.0g
Medium		per litre
	$NaNO_3$	0.075g
	$NaH_2PO_4 \cdot 2H_2O$	0.00565g
	Trace elements stock solution (1)	1.0 mL
	Vitamin mix stock solution (2)	1.0 mL
	Sodium metasilicate stock solution (3)	1.0 mL
	Agar (only for agar plates)	15g/L

ESAW [Berges et al., 2001]

Component	Stock Solution (g/L)	Quantity
Salt solution I: Anhydrous salts		600mL
NaCl		21.194g
Na_2SO_4		3.550g
KCl		0.599g
$NaHCO_3$		0.174g
KBr		0.0863
H_3BO_3		0.0230
NaF		0.0028g
Salt Solution II: Hydrated salts		300mL
$MgCl_2{\cdot}6H_2O$		9.592
$CaCl_2{\cdot}2H_2O$		1.344g
$SrCl_2{\cdot}6H_2O$		0.0218g
Major nutrients		
$NaNO_3$	46.67	1mL
$NaH_2PO_4{\cdot}H_2O$	3.094	1mL
$Na_2SiO_3{\cdot}9H_2O$	15	2mL
Iron-EDTA stock solution		1mL
$Na_2EDTA{\cdot}2H_2O$		2.44g
$FeCl_3{\cdot}6H_2O$	1.77	1mL
Trace metals II Stock solution		1mL
$Na_2EDTA{\cdot}2H_2O$		3.09g
$ZnSO_3{\cdot}7H_2O$		0.073g
$CoSO_4{\cdot}7H_2O$		0.016g
$MnSO_4{\cdot}4H_2O$		0.54
$Na_2MoO_4{\cdot}2H_2O$	1.48	1mL
Na_2SeO_g	0.173	1mL
$NiCl_2{\cdot}6H_2O$	1.49	1mL
Vitamin stock solution		1mL
Thiamin·HCl (vit. B_1		0.1g
Biotin (vit. H)	1.0	1mL
Cyanocobalamin (vit. B_{12}	2	1mL

SOC-medium

	Amount
Trypton	20g/L
Glucose	3.6 g/L
$MnCl_2{\cdot}$ $2H_2O$	5.08 g/L
KCl	2.5 mM

SDS-EDTA

	Concentration
Sodium dodecyl sulfate	50g/L
Titriplex III	100mM

***Thalassiosira pseudonana* lysis buffer**

	Concentration
Tripton x-100	10%
Tris-HCl (pH8)	20 mM
EDTA	10mM

1X TAE buffer

	Amount
Tris-base	4.84 g
Glacial acetic acid	1.142 mL
EDTA	$1.2 \cdot 10^{-3}$M

Enzymes

Enzyme	Manufacturer
T4 DNA Ligase	New England BioLabs (ref: M0202S)
BSA-I HF v2	New England BioLabs (ref: R3733S)
PstI-HF	New England BioLabs (ref: R3140S)

Antibiotics stock

Antibiotics	Manufacturer
Kanamycin	panReac Applichem
Gentamycin	Gibco
Nourseothricin	Jena Biosciences
Steptomycin	Sigma Aldrich
Ciprofloxacin	Sigma Aldrich
Chloramphenicol	Sigma Aldrich
Ampicillin	Sigma Aldrich

Microorganisms

Microorganism	Strain(s)
E.coli	DH5α
	DH10β
	NEB Stable
Thalassiosira Pseudonana	CCMP 1335

Vectors

Vector	Manufacturer
pTpPUC3	AddGene (Plasmid #62864)
pNCS-mNeonGreen	Allele biotechnology
mNeonGreen-mTurqoise	Addgene (Plasmid #98886)

Kits

Kit	Manufacturer (Product number)
Phusion hot start II High-fidelity PCR DNA polymerase	Thermo Scientific
Phusion Hot Start II High-Fidelity PCR master mix	Thermo Scientific
RedTaq 2X Master mix	VWR Life Science
Light cycler 480 High Resolution Melting Master	Rocher (Version 07)
ExS-pure enzymatic PCR purification kit	NimaGen
GeneJET plasmid Miniprep Kit	Thermo Scientific
Q5 High-Fidelity DNA Polymerase	New England BioLabs
Gibson Master Mix	New England Biolabs

Various compounds

Compound	Manufacturer
T4 DNA Ligase buffer (10x buffer, with 10 mM ATP)	New England Biolabs
Cut Smart buffer	New England BioLabs
GelRed nucleic acid stain 10,000x	Biotium
6x DNA Loading Dye	Thermo Scientific
dNTPs	VWR Life Science
SeaKem LE agarose	Lanza

Instruments

Instrument	Manufacturer
Mikrocentrifuge	VWR Himac, CT15E
Multifuge	Thermo Scientific Heraeus multifuge, x1R Centrifuge
Flow Cytometer	BD ACCURI C6 flow cytometer and Biosciences Novocyte
Thermo-Shaker	Grant bio, version: V.4GW
Gelelectrophoresis	Consort EV1450
Gel tray	Thermo Scientific OWL easycast B2/B1/B1A
PCR	Biorad t100 Thermo Cycler
Incubator with shaking	Multitron 2, INFORS AG
Qubit 4 Fluorometer	Invitrogen
Electron microscopy	Teneo SEM, Thermo Scentific
Nanodrop	NanoDrop^{TM} 1000 Spectrophotometer, Thermo Scientific
AquaPen	Ap110/C, ICT international

A.2 Methods - Project thesis

The aim of these experiments was to perform single knock outs of the *Thalassiosira pseudonana* genes Tp23191 and Tp6330 and a double knockout of both of these genes, and to tag two genes (Tp23191 and Tp20931) with a fluorescence marker. Both of these experiments were done by modifying a vector and transforming it into *T. pseudonana* by conjugation.

A.2.1 Protocols

Heat shock-transformation

Plasmids can be taken up by competent *E.coli*-cells by heat shock transformation. The reaction set-up is shown in Table A.1. The solution was incubated on ice for 30 min and then at 42°C with shaking (250 rpm) for 45s. It was then incubated on ice for 2 min before LB-medium (1 mL) was added. After incubating for 1h at 37°C, culture (100 μL) was plated out on LB-plates with kanamycin (100 μg/mL). The plates were incubated at 37°C overnight.

Table A.1: Components for heat chock transformation

	Amount
Competent DH5α- or DH10β-*E. coli*	50 μL
Plasmid	0.5 μL

Colony PCR with RedTaq polymerase

Clones were picked and plated out on a new LB-plate (colony plate). The leftover cells on the pipette tip used for picking were mixed with the solution described in Table A.2 by pipetting up and down a few times. A PCR was performed according to Table A.3.

Table A.2: PCR-reaction set-up for colony screening with RedTaq DNA polymerase master mix.

Compound	Volume/reaction
RedTaq polymerase master mix	12.5 μL
Reverse primer	0.5 μL
Forward primer	0.5 μL
ddH_2O	To a total of 25 μL

Table A.3: Colony PCR program.

Inital denaturation	95°C	5 min	
Denaturation	95°C	30s	
Annealing	60°C	30s	repeat 34 times
Extention	72°C	20s	
Final extention	72°C	5 min	

Extraction of nuclear DNA from *Thalassiosira pseudonana*

Thalassiosira pseudonana culture in f/2+Si-medium (100 μL) was spun down at 17,000g for 1 min. The supernatant was removed and lysis buffer (20 μL) was added. The solution was kept on ice for 15 min, before it was placed in a Thermo shaker at 95°C for 10 min.

A.2.2 Single and double knock-out of Tp23191 and Tp6330

A list of primers used throughout this experiment is shown in Table A.4, with their sequence and application. The locations of each primer is shown in Figure A.1. The vector used to transform the genes for Cas9 in the knock-out experiment is shown in Figure A.2. This vector contains the genes for kanamycin resistance and Cas9, tagged with a yellow fluorescent protein (EYFP) and a nuclear localization sequence, and its associated sgRNA. The two BSAI cutting sites flanking the target sequence in the sgRNA made it possible to modify the sgRNA to make it specific for chosen a target sequence. The target sequences were chosen (by Tore Brembu) using CRISPRdirect [Naito et al. [2015]] and checked with another program made by Associate professor Per Winge that recognizes PAMs. The possible sequences were carefully checked to decrease the risk of off-target effects or low efficiency. The target regions of the sgRNA that eventually were chosen are shown in Table A.5, and their location within the genes are shown in Figure A.1. The double knock out experiment was possible since these two genes contains some sequences that are identical to each other. This made it possible to find a target sequence adjacent to a PAM motif that was located in the sequences these two genes share.

Table A.4: List of primers showing their sequence and application used throughout these experiments. F represents for forward primers and R reverse primers. The location of these primers are shown in Figure A.1.

Primers	Sequence	Application
Tp6330F1	GAGTACAAGCAGGTTCCGACCA	Amplifaction of
Tp6330R2	TCATTGATGGTGTTCCATGGCT	*Thalassiosira pesudonana*
Tp23191F1	TTGATCAGCAGTGACCTCAACA	gene by PCR
Tp23191R2	CAACACAGGGAGGTCAAACTCT	
Tp23191PAM1qSF	GCAGGTTTCTGAGTTCATCA	
Tp23191PAM1qSR	TCCTCGTCAACAAAGACAGC	
Tp23191PAM4qSF	ACCGCCAAAACCATCAAACTCG	HRM primers
Tp23191PAM4qSR	CGAGGAGAAGAGGTGGTTCCTG	for Tp23191
Tp23191PAM2qSF	AACCCCTACGCTAATGGTGA	
Tp23191PAM2qSR	TGTCCGTAGGCAACGTACTC	
Tp6330PAM1qSF	GAGCAAGTTTCTGAGTGGAT	
Tp6330PAM1qSR	TGCTATAGATGGAGCAATCC	
Tp6330PAM2qSF	CGGCGAAGAGATCAACTACC	
Tp6330PAM2qSR	CTGAAGGGGCTCTTCGTACT	HRM primers
Tp6330PAM1hrmF	GATGTCAAGACCACCAGTGAGG	for Tp6330
Tp6330PAM1hrmR	CGTCGTAGTAACAGTCGTTTCCT	
Tp6330PAM2hrmF	GGCCATCGACTGTGACCAATG	
Tp6330PAM2hrmR	AGATCCAGATGCCTCACACTCAG	
M13-Rev	GTCATAGCTGTTTCCTG	Amplifying target
Forward primer in Table A.5	See Table A.5	sequence in pTpPUC3

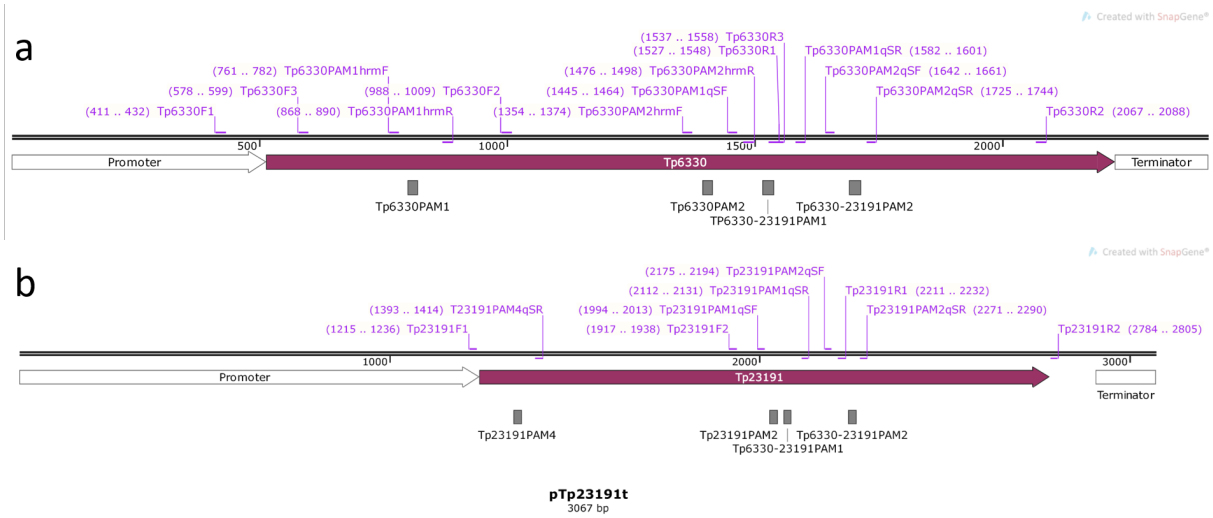

Figure A.1: Gene maps of the two *Thalassiosira pseudonana* genes that were analyzed in this experiment, **(a)**: Tp6330 and **(b)**: Tp23191. The maps show the gene in dark purple, with their promoter and terminator in white. The location for the different primers (listed in Table A.4) that were used throughout this experiment is shown. In the primer name, F represents forward primer, R reverse primer, and hrm and qS primers used for high resolution melting analysis. The location for each target sequence in the Cas9 associated sgRNA in the pTpPUC3 vector is shown under each map.

Table A.5: The primers used for oligo annealing to create the target site of the sgRNA associated with Cas9.

Target gene	Forward primer	Reverse primer
Tp6330PAM1	ATTGTCTAATTGAGTACGCTCAGGCA	AAACTGCCTGAGCGTACTCAATTAGA
Tp6330PAM2	ATTGTCACCCTCGTCGGCCTGGTCTA	AAACTAGACCAGGCCGACGAGGGTGA
TP23191PAM2	ATTGTGTGAAGCCTCAGGGTCTCAGA	AAACTCTGAGACCCTGAGGCTTCACA
TP23191PAM4	ATTGTCGGTGGCGGGCTCCCCGCGAA	AAACTTCGCGGGGAGCCCGCCACCGA
TP6330-23191PAM1	ATTGTTGCAGATAGCTCCGTAGTAGA	AAACTCTACTACGGAGCTATCTGCAA
TP6330-23191PAM2	ATTGTACAGCTCATGGTATCACTGAA	AAACTTCAGTGATACCATGAGCTGTA

A.2.2.1 Construction of pTpPUC3 vector used for conjugation in *Thalassiosira pseudonana*

Preparing pTpPUC3-vector by restriction digestion

The pTpPUC3 vector was retransformed into DH5α-*E. coli* by heat shock transformation according to the protocol described in Section A.2.1. Clones from the plate were picked and incubated in LB-medium (5 mL) at 37°C overnight. Miniprep was performed using the overnight culture (1 mL) according to Thermo Scientific GeneJet plasmid Miniprep protocol to extract the plasmid.

The vector was cut with BsaI-HF v2 by mixing plasmid DNA (1 μg), 10x Cutsmart buffer (5 μL), BsaI-HF v2 (1 μL) and ddH_2O (to a total of 50 μL), and incubated on ice for 15 min. This linear vector would later be ligated with the specific oligo sequences containing the desired target sequences.

Preparing oligo fragments with the specific target sequences by oligo annealing

The oligo primers were annealed by mixing forward and reverse primers of a pair (see Table A.5; 1.4 μL of each, 10 μM), ligase buffer (5 μL) and ddH_2O (end volume 50 μL). The solution was annealed in a Thermo cycler starting at 85°C for 10 min, then the temperature was lowered by 1°C every 1.5 min 60 times.

Ligation of plasmid and oligo fragments

The oligo fragments and the linear vector was ligated with T4 DNA ligase according to the New England BioLabs T4 DNA ligase protocol. After the ligation was complete, the enzyme was inactivated at 65°C for 10 min.

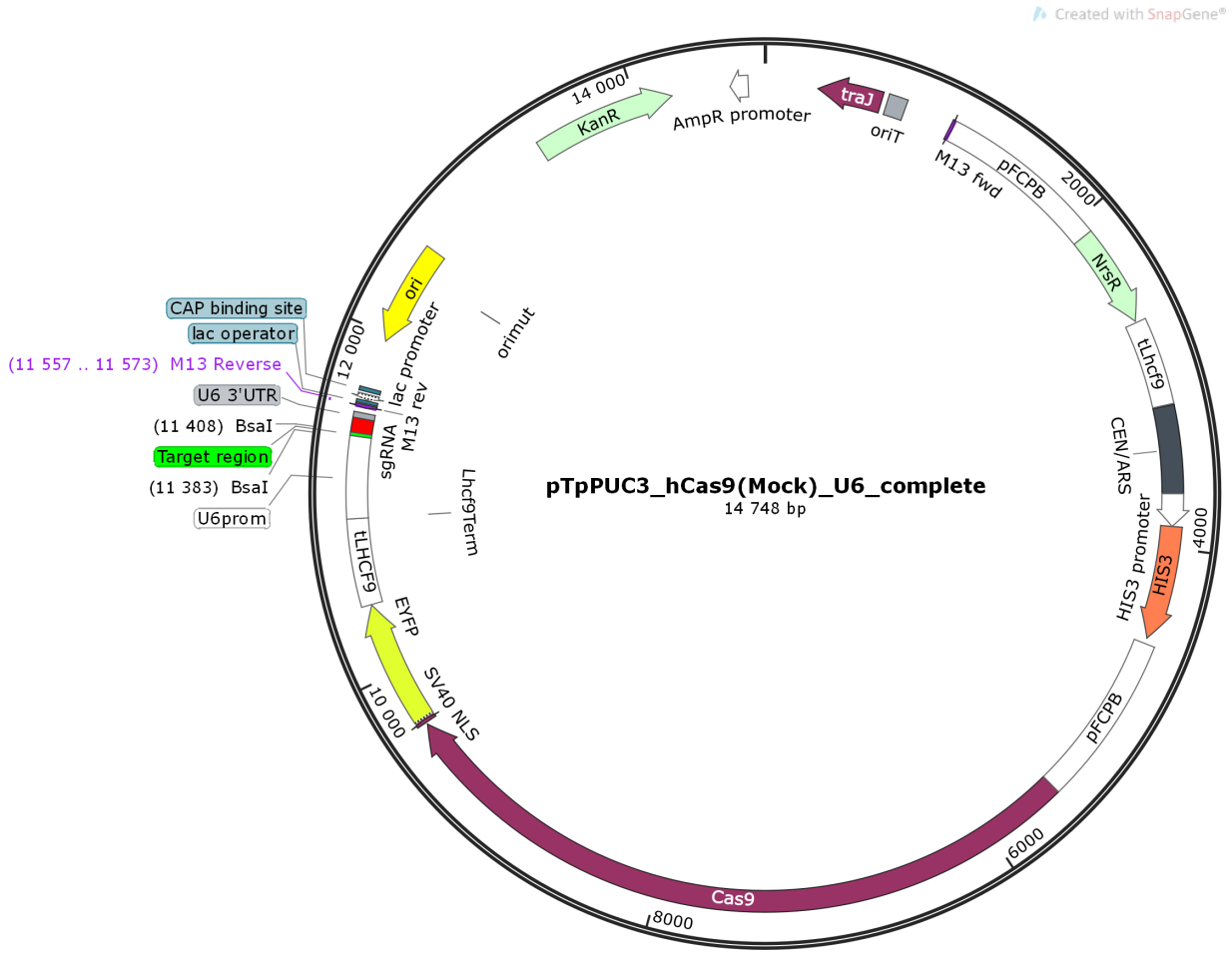

Figure A.2: Plasmid map of the vector, pTpPUC3 (14,748 bp), used for transforming the genes for Cas9 into *T. pseudonana*. The vector contains the genes for kanamycin and nourseothricin resistance (both in mint green) and for Cas9, tagged with a yellow fluorescence protein (EYFP) and a nuclear localization sequence (SV40 NLS), and its associated sgRNA (red) with the target sequence (green). The target sequence is flanked by two BsaI cutting sites. The location of the M13-rev primer that was used in for this plasmid is shown in purple. The forward primers (see Table A.5) binds to the target region.

The ligated vector was heat shock-transformed into DH5α-*E. coli* and clones were screened with colony PCR according to the protocols described in Section A.2.1. The primers used for the colony PCR were M13-rev primer and forward primer listed in Table A.5. The colony PCR products were analyzed by gel electrophoresis. For the samples that showed a signal on the gel, the corresponding clones on the colony plate were picked and the pipette tip was released into a tube with LB-medium (5 mL) with kanamycin (100 μg/mL). The culture was incubated at 37°C overnight, and a miniprep was contucted on these cultures according to Thermo Scientific GeneJet plasmid miniprep. The extracted plasmids were sent in for Sanger sequencing, using M13-rev as a primer, to check if the oligo fragments were correctly inserted. The plasmids that were found to be correct were heat shock transformed to competent DH10β-*E. coli*, as this *E. coli* strain contains the conjugational plasmid pTA-Mob, which carries the genes for gentamycin resistance and enables them to transform plasmids by conjugation. Clones were picked and resuspended in liquid LB-medium (5 mL) and incubated overnight at 37°C.

A.2.2.2 Conjugation

LB-medium (150 mL) with kanamycin (100 μg/mL) and gentamycin (20 μg/mL) was inoculated with transformed DH10β-*E.coli* and incubated with shaking (220 rpm) at 37°C until OD reached 0.3. The cells were then spun down for 10 min at 3000g and resuspended in SOC media (800 μL).

T. pseudonana culture was spun down at 4000g at 10°C for 5 min. Most of the supernatant was removed. The pellet was resuspended in the remaining media and cell density was measured using BD ACCURI C6 flowcytometer. $4.85{\cdot}10^5$ *T. pseudonana cells* were mixed with *E. coli* cells (200 μL) by pipetting up and down a few times. Cells were plated on f/2+Si plates, and incubated in the dark at 30°C for 90min. The plates were moved to 18°C in the light and incubated for 4h. f/2+Si medium (1 mL) was added to the plates and the cells were scraped off. Scraped cells (200 μL/800 μL) were transferred to f/2+Si plates with nourseothricin (100 μg/mL) and incubated at 18°C in the light until colonies appeared.

A.2.2.3 Screening

For all following PCR set-ups, the extension step was increased by 1 min to ensure the reaction was complete even for gene mutations with large insertions.

Clones were picked and transferred to tissue culture plates containing f/2+Si-medium (1 mL) with nourseothricin (100 μg/L) for cultivation. DNA was extracted according to the procedure described in Section A.2.1.

A PCR was conducted to amplify the genes of interest to look for possible mutations by the following method. Extracted DNA (1 μL) was added to Phusion Hot start HFII polymerase Master mix according to the Thermo Scientific protocol for this Master mix, using 1 μL of each primer (see Table A.4, primers for amplification of gene) to a total volume of 20 μL. The product was analyzed by gel electrophoresis to look for successful amplification and indications of possible mutations.

High Resultion Melting (HRM)

Extracted DNA (5 μL) of clones and wild type were used to amplify the genes of interest by PCR with Phusion hot start II High-fidelity PCR DNA polymerase and primers listed in Table A.4, following the Thermo Scientific protocol for this enzyme. The PCR product was diluted (1:4,000,000) with ddH_2O. The qPCR was set up according to Life science LightCycler 480 High Resolution Melting Master, using 2.4 μL $MgCl_2$ and diluted PCR product (5 μL) as DNA template. The primers that were used are shown in Table A.4. For clones that were different from the wild type, 10^{-1} or 10^{-5} diluted culture (100 μL), depending on the density of the culture, was plated out on f/2+Si plates for further screening.

Flow Cytometry

Liquid culture of clones and wild type was transferred to flow cytometry tubes and analyzed using Novocyte flow cytometer to look for differences in size in clones compared to the wild type. The clones that diverged from the wild type, 10^{-1} diluted culture (100 μL) was plated out on f/2+Si plates for further screening.

Colony screening of replated clones

The replated clones were further screened by colony PCR using Phusion High fidelity hot start DNA polymerase, and the primers for amplification of genes listed in Table A.4 by the following method. The clones were picked and replated, and the leftover cells on the pipette tip were used to extract DNA by immersing the pipette tip in lysis buffer (20 μL) and pipetting up an down a few times. The rest of the DNA extraction was conducted according to the protocol described in Section A.2.1, and the PCR was set up according to the protocol for this enzyme. A gel electrophoresis was performed to check for product. The PCR products that showed a signal on the gel were cleaned using Exs-pure enzymatic PCR purification kit, by adding iX-Pure reagent (2 μL) to the PCR-product (5 μL), and incubating the solution at 37°C for 4 min, The enzyme was then inactivated at 95°C for 1 min. To this solution, primer (forward primer used for amplification of gene, see Table A.4; 2.5 μL) and water (end volume 10 μL) was added, and the samples were sent in for Sanger sequencing to identify possible frame shift mutations.

A.2.3 Tagging Tp23191 and Tp20931 with mNeon/mTurq

The aim of this part of the experiment was to tag two genes of the silicanin protein family, Tp23191 in group I and Tp20931 in group III with a fluorescence marker to see their location *in vivo*. A representative map of the vector (pTpPUC-mNeon) used for transforming the genes tagged with a fluorescence marker (mNeon/mTurq) is shown in Figure A.3. This map shows the inserted gene (Tp23191) tagged with mNeon. mNeon was inserted downstream of the start of the gene because of the proteolytic removal of the signal peptide.

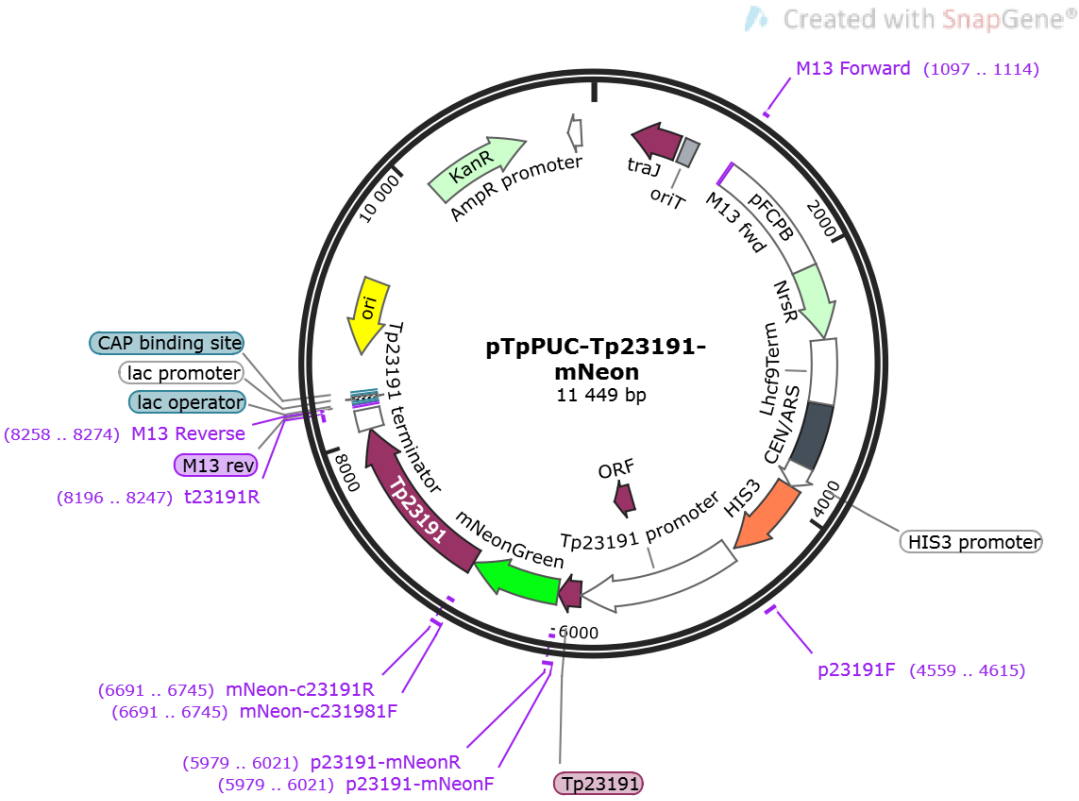

Figure A.3: Plasmid map of pTpPUC-Tp23191-mNeon. The vector contains the genes of interest (Tp23191 is shown in purple), with the promoter and terminator (white) tagged with mNeon (green), and the gene for kanamycin resistance (mint green).

A.2.3.1 Constructing vector used for conjugation with mTurq-tagged genes

Cutting pTpPUC3

pTpPUC-mNeon (1 μL) was cut with PstI-HF (1 μL) in a solution (end volume 50 μL) containing CutSmart buffer (5 μL). 70 ng vector was used for the assembly.

Amplifying fragments for assembly

Wild type *T. pseudonana* DNA was extracted according to the protocol described in Section

A.2.1. A PCR was performed to amplify fragments used for the assembly. To find the correct annealing temperature, a PCR with several annealing temperatures was performed. The annealing temperature that were found to be optimal are shown in Table A.6, together with a description of which fragments were created, and the PCR program is shown in Table A.7. For the mNeon gene, the plasmid pBKS-mNeon was used as DNA-template, and extracted DNA from wild-type *T. pseudonana* cells were used as template DNA for the promoter, and gene and terminator fragments. The PCR products were analyzed by gel electrophoresis to check if the fragments had been amplified. The concentration for each PCR product were measured using Qubit assay, and 70 ng of each was used in the assembly.

Table A.6: Primers and their annealing temperature for amplifying fragments used in assembly for creating pTpPUC3-mNeon vectors. The last column shows which fragment is created, and which DNA template was used in the PCR.

Forward primer	Reverse primer	annealing temperature (°C)	Amplified fragment (DNA template)
P20931F	P20931mNeonR	60.0	Tp20931 Promotor (*T. pseudonana)*
P20931mNeonF	mNeonc20931R2	67.2	mTurq (pNCS-mNeon plasmid)
mNeonc20931F2	t20931	69,6	Tp20931 Gene + terminator (*T. pseudonana)*
P23191F	P23191mNeonR2	60.0	Tp23191 Promotor (*T. pseudonana)*
P23191mNeonF	mNeonc23191R	67.2	mTurq (pNCS-mNeon plasmid)
mNeonc23191F	t23191R	69.6	Tp23191 Gene + terminator (*T. pseudonana)*

Table A.7: PCR-program for amplifying fragments used in assembly for creating pTpPUC3-mNeon vectors.

	temperature (°C)	time	repeats
Hot start	98	∞	
Initial denaturation	98	30s	
Denaturation	98	10s	
Annealing	see table A.6	15s	34x
Extention	72	15-30s/kb	
final extention	72	5 min	

Circular polymerase extension cloning (CPEC)

Assembly of the PstI cut vector and fragments were performed following the protocol for CPEC by Quan and Tian (2011). The products were ran on a 1% agarosegel. The PCR product for the pTpPUC3-Tp23191-mNeon assembly was then heat shock-transformed into DH5α-*E. coli*. This was not done for pTpPUC3-Tp20931-mNeon as no product was observed by gel electrophoresis for this assembly. The heat shock transformed DH5α-*E. coli* were plated out, and colony screening was performed on these clones according to the protocol described in Section A.2.1, using M13-rev primer and forward primer for Tp23191 for amplification of gene by PCR, see Table A.4.

The colony PCR of DH5α-*E. coli* cells that were heat shock transformed with the Tp23191-

mNeon vector gave no result. This experiment was attempted again, using mTurq instead of mNeon. mNeon was replaced with mTurq because a different student had created clones where the proteins Tp21058 and Tp23225 of the ankyrin protein family was tagged with mNeon. It is suspected that these proteins interact in the cell wall. By tagging the silicanins genes with mTurq and transforming them into the clones with mNeon tagged ankyrins, this could be investigated by looking for fluorescence resonance energy transfer. Since the mTurq and mNeon genes has an identical start and end, the same primers, annealing temperature and procedure as described above was used, except a different plasmid, which had been previously modified from mNeon-mturquoise to only contain the mTurq gene, was used.

A.3 Primers

Table A.8: Primers and their sequence. The localizations can be seen in Figs. A.4-A.7

Primer	Sequence
M13rev	AGCGGATAACAATTTCACACAGG
M13For	CCCAGTCACGACGTTGTAAAACG
Tp23191F1	TTGATCAGCAGTGACCTCAACA
Tp23191PAM4qsR	CGAGGAGAAGAGGTGGTTCCTG
Tp23191F2	CTTGCCTACGAATGCTACCAAG
Tp23191PAM1qsF	GCAGGTTTCTGAGTTCATCA
Tp23191PAM1qsR	TCCTCGTCAACAAAGACAGC
Tp23191qsF	AACCCCTACGCTAATGGTGA
Tp23191R1	TGATGTAGTTCTCGGCGTAAGT
Tp23191PAM2qsR	TGTCCGTAGGCAACGTACTC
Tp23191R2	CAACACAGGGAGGTCAAACTCT
p23191F	GACCTCTAGAGTCGACCTGCATGAAAGAGTATGCAAGATTTTTCTATTCCAAAAGAG
p23191-mNeonF	CACCTCTTCTCCTCGTAAACTTGAGATGGTGAGCAAGGGCGAG
p23191-mNeonR	CTCGCCCTTGCTCACCATCTCAAGTTTACGAGGAGAAGAGGTG
mNeon-c23191F	GGCATGGACGAGCTGTACAAGGAAGAAGAAATAGATGGATCCTACTCATTGAAGT
mNeon-c23191R	ACTTCAATGAGTAGGATCCATCTATTTCTTCTTCCTTGTACAGCTCGTCCATGCC
t23191R	CCAAGCTTGCATGCCTGCAAAGTAGCAGAGATAAAGGTACTTCAATTATCAG
Tp6330F1	GAGTACAAGCAGGTTCCGACCA
Tp6330F3	GTACCTTTCCACAAAGCCATCA
Tp6330PAM1hrmF	GATGTCAAGACCACCAGTGAGG
Tp6330PAMhrmR	CGTCGTAGTAACAGTCGTTTCCT
Tp6330F2	ACAATTCGAGGAGCAGTGTACT
Tp6330PAM2hrmF	GGCCATCGACTGTGACCAATG
Tp6330qSF	GAGCAAGTTTCTGAGTGGAT
Tp6330PAM2hrmR	AGATCCAGATGCCTCACACTCAG
Tp6330R1	CCATAAGGACTGCAGATAGCTC
Tp6330R3	AACTCCATCTCCATAAGGACTG
Tp6330PAM1qSR	TGCTATAGATGGAGCAATCC
Tp6330PAM2gSF	CGGCGAAGAGATCAACTACC
Tp6330PAM2qSR	CTGAAGGGGCTCTTCGTACT
Tp6330R2	TCATTGATGGTGTTCCATGGCT
p20931F	TTGACCTCTAGAGTCGACCTGCACTAAATGACATTGTATTGTTGCACCGTAGA
p20931-mNeonF	AGAGAGGACGCAAGCTAAGCATGGTGAGCAAGGGCGAG
p20931-mNeonR	CTCGCCCTTGCTCACCATGCTTAGCTTGCGTCCTCTCT
mNeon-c20931F2	GGCATGGACGAGCTGTACAAGAACATGAAGGAGTTTGAGGCTGAG
mNeon-c20931R2	CTCAGCCTCAAACTCCTTCATGTTCTTGTACAGCTCGTCCATGCC
t20931R	CAAGCTTGCATGCCTGCACAGTCAGAAATTAAGTTTGTTTAGTAACAATTTTATTCTATACAA

A.4 Gene maps and vectors

The gene maps of the two genes studied in the knock out experiment, *Tp23191* and *Tp6330* with all primers are shown in Figs. A.4 and A.5. The two vectors used for transforming *Tp20931* and *Tp23191* tagged with *mTurq* in the localization study is shown in Figs. A.6 and A.7, with all primers used for amplification of fragments, colony screening and Sanger sequencing included. See Tab. A.4 for a list of primers and their sequence used in all experiments.

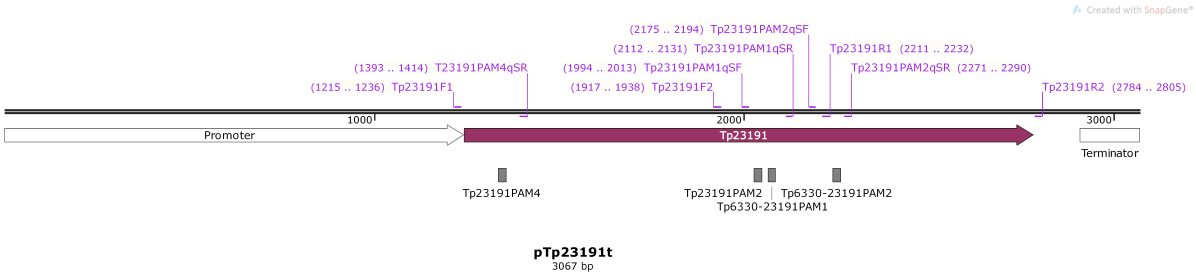

Figure A.4: Gene map of *Tp23191*, showing the gene in purple and the promoter and terminator in white. All primers located in the gene are included.

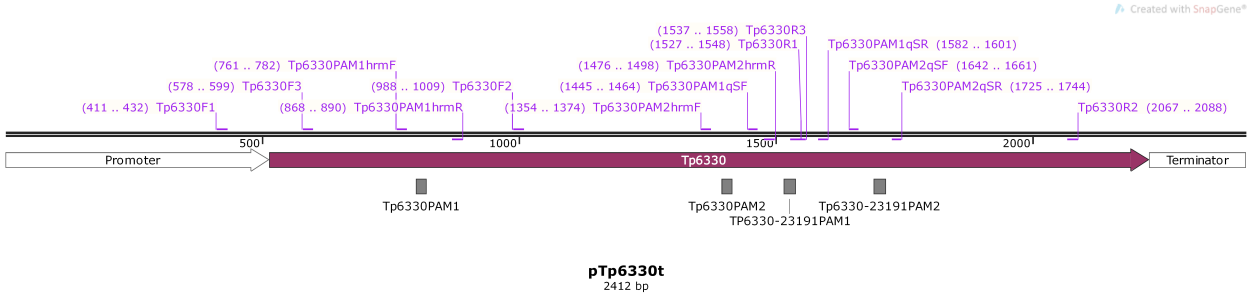

Figure A.5: Gene map of *Tp6330*, showing the gene in purple and the promoter and terminator in white. All primers located in the gene are included.

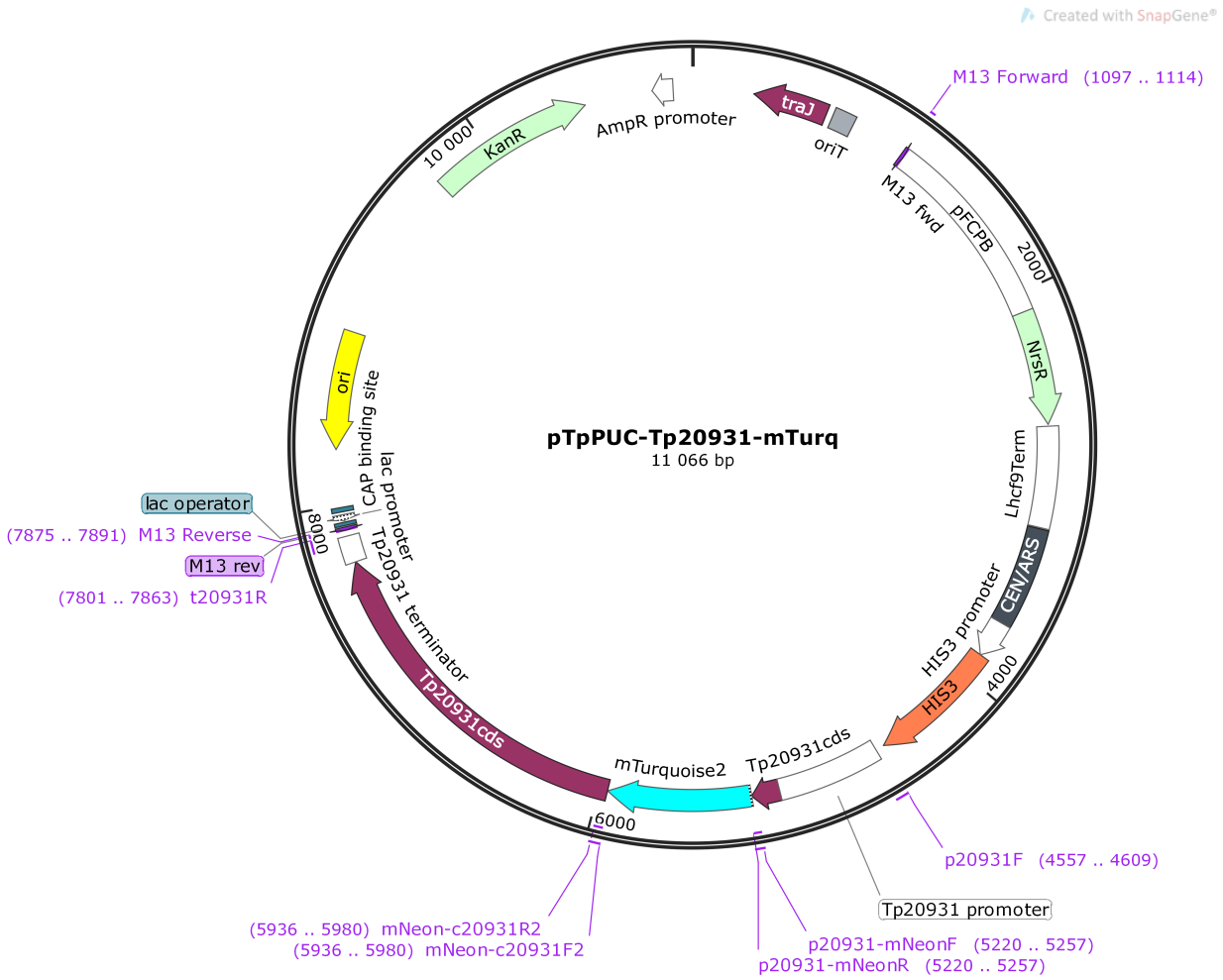

Figure A.6: The vector used for transforming *Tp20931* tagged with *mTurq* for the localization study. All primers used for amplification of fragments, colony screening and Sanger sequencing is included.

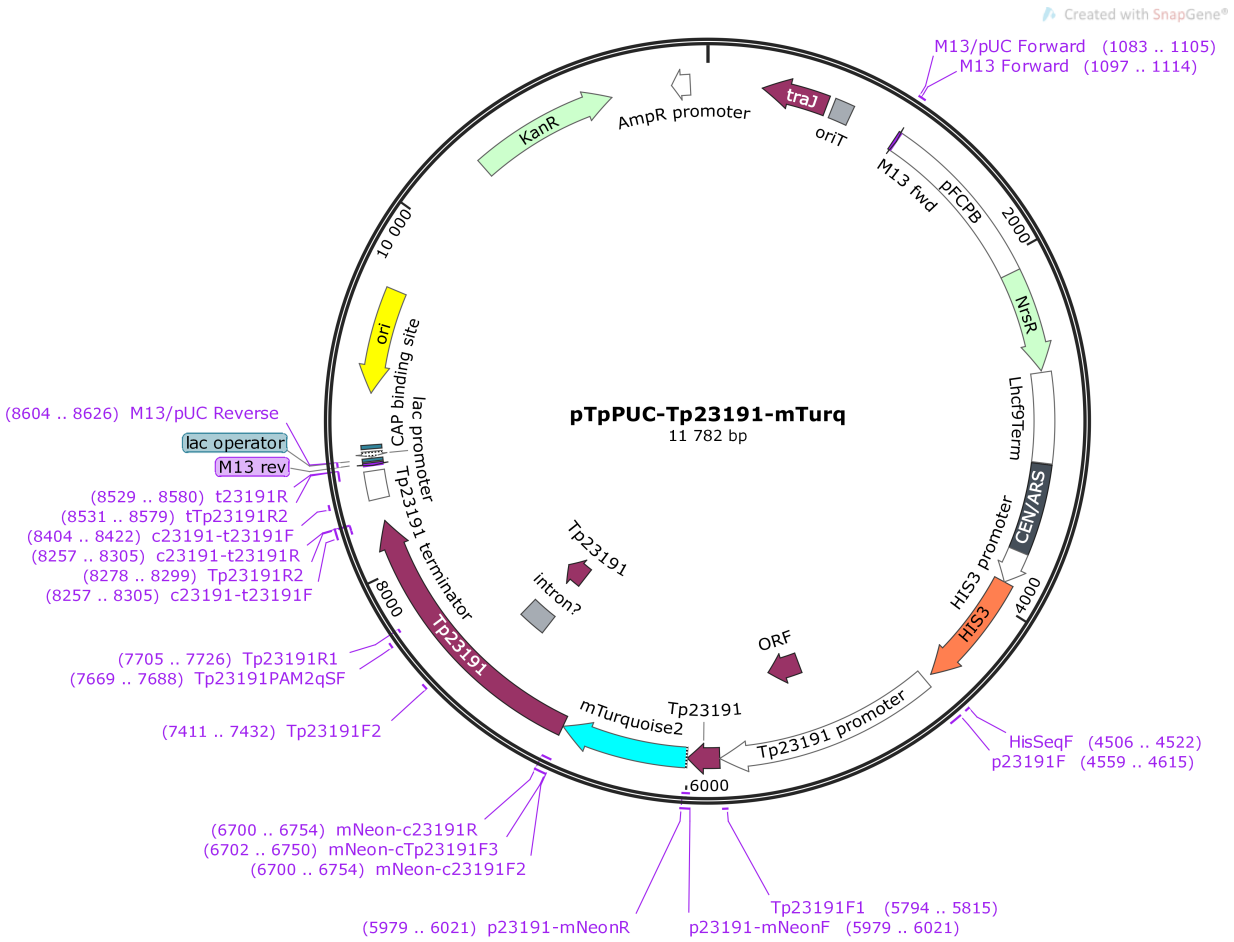

Figure A.7: The vector used for transforming *Tp23191* tagged with *mTurq* for the localization study. All primers used for amplification of fragments, colony screening and Sanger sequencing is included.

A.5 Flow cytometry results

Graphs for flow cytometry performed on *Thalassiosira pseudonana* clones transformed with the pTpPUC3 vector containing the target sequence Tp6330PAM2 is shown in Figures A.8 and A.9. Graph 1: number of cells. The red bar shows the M1 gate, indicating living cells. Graph 2: Forward scattering (FSC) for clones (red) compared to wild type (black). Graph 3: Percentage of cell count for cells at different FSC-H values for clones (blue) compared to wild type (gray)

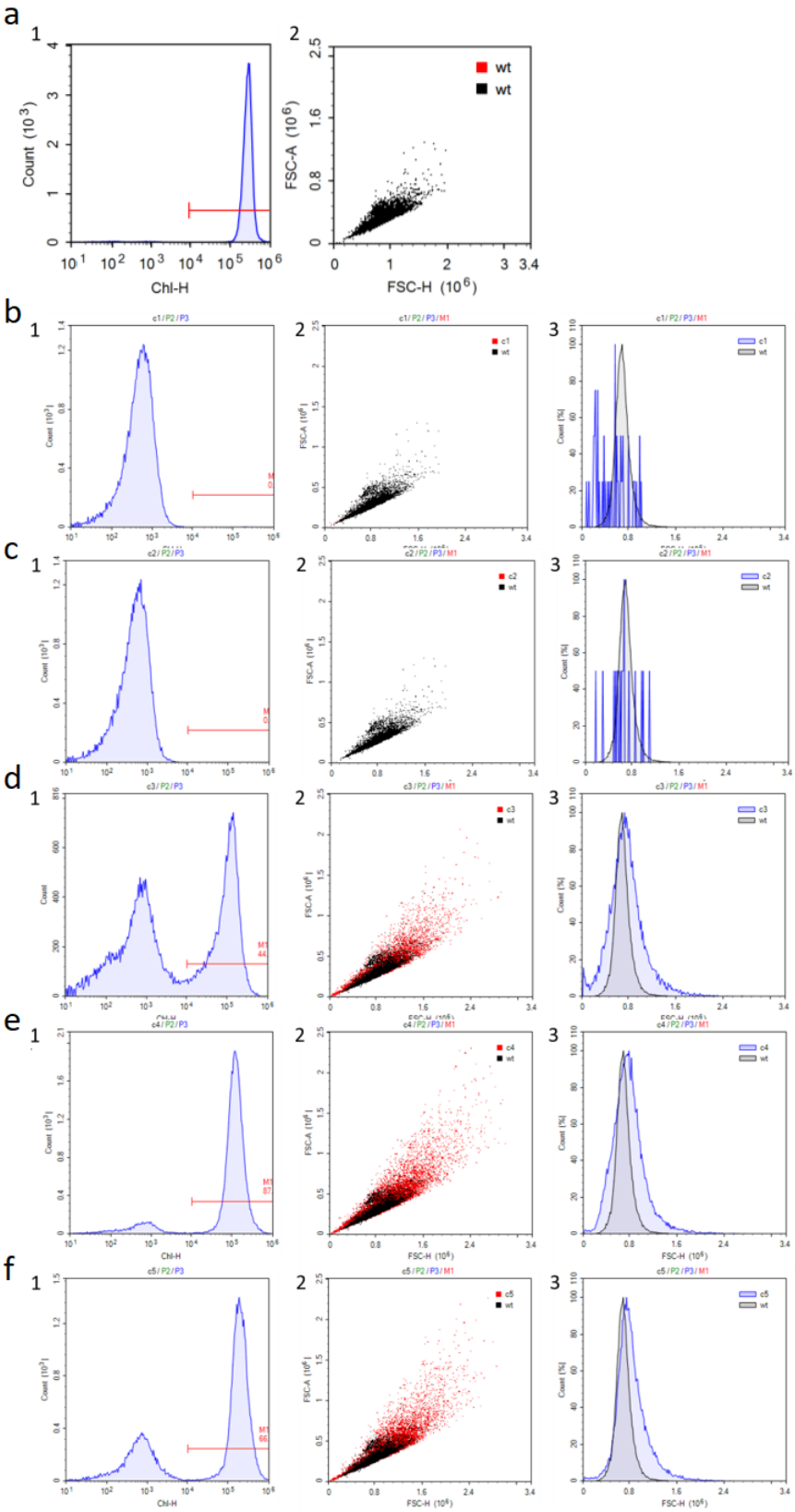

Figure A.8: Flow cytometry data for **(a)** Wild type, **(b)**C1, **(c)** C2, **(d)**C3, **(e)**C4 and **(f)**C5

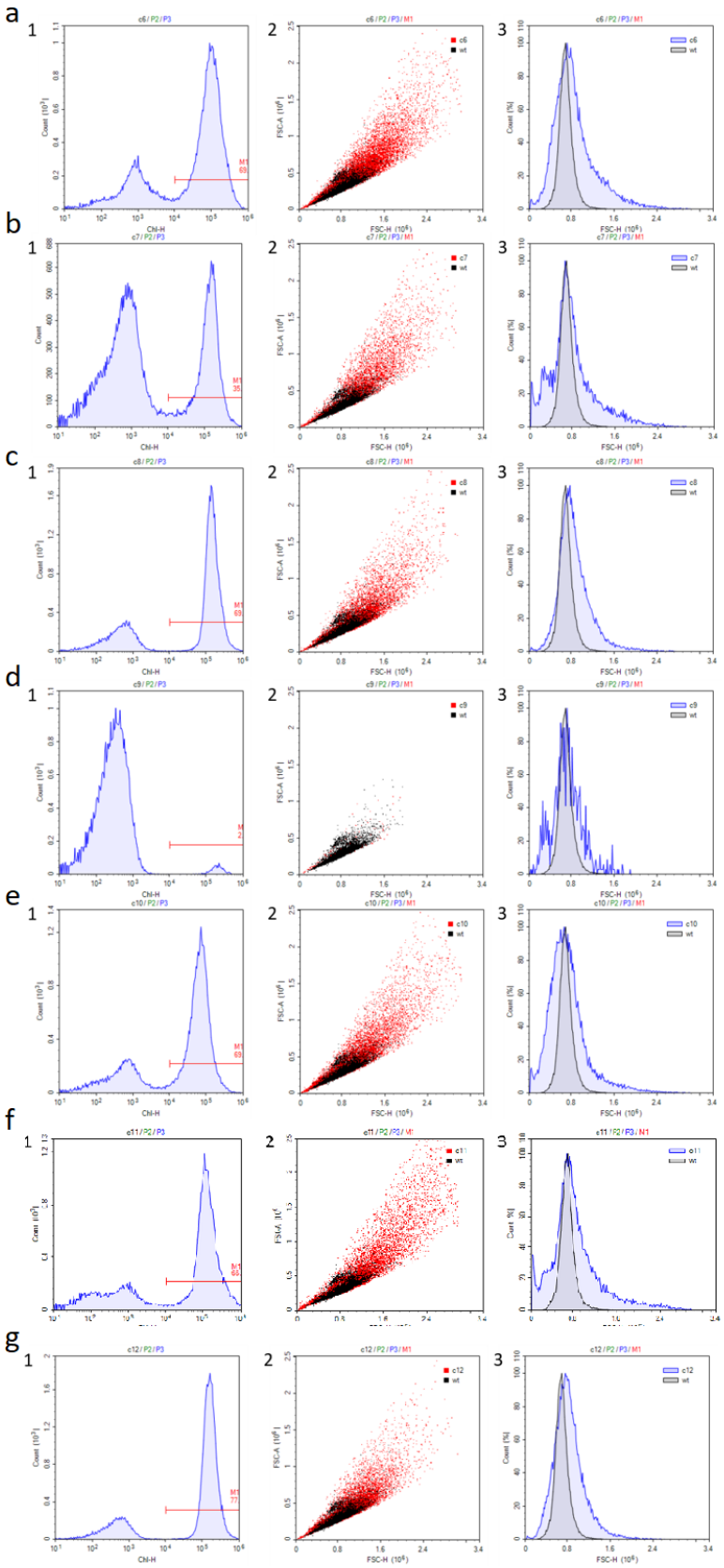

Figure A.9: Flow cytometry data for **(a)** C6, **(b)**C7, **(c)** C8, **(d)**C9, **(e)**C10, **(f)**C11 and **(g)**C12

A.6 Quantum yield data

Table A.9: Quantum yield of each replicate, and the mean with standard deviation for each sample.

Sample	Qyantum Yield (Fv/Fm)	Mean (± Std Dev)
C_1	0.66	
C_2	0.59	0.60± 0.04
C_3	0.56	
$Al_{5,1}$	0.61	
$Al_{5,2}$	0.61	0.607±0.005
$Al_{5,3}$	0.60	
$Al_{15,1}$	0.58	
$Al_{15,2}$	0.60	0.590±0.008
$Al_{15,3}$	0.59	
$Al_{25,1}$	0.60	
$Al_{25,2}$	0.60	0.58±0.03
$Al_{25,3}$	0.54	
$Cd_{1,1}$	0.49	
$Cd_{1,2}$	0.56	0.52±0.03
$Cd_{1,3}$	0.50	
$Cd_{2.5,1}$	0.52	
$Cd_{2.5,2}$	0.54	0.52±0.02
$Cd_{2.5,3}$	0.49	
$Cd_{5,1}$	0.61	
$Cd_{5,2}$	0.59	0.55±0.08
$Cd_{5,3}$	0.44	

A.7 SEM images and measurments of biosilica structures

An example of how the different parameters were measured is shown in Fig. A.10. The valve in many images were taken at an angle and not directly above. Because of this, the diameter of each valve was measured up to four times at different locations, each time from the inner silica ring around the rim, through the center at the valve (Fig. A.10a). The rimoportulae was always measured tangent to the circumference of the valve as the angle may affect the visible diameter perpendicular to the circumference.

All images used for measuring the different biosilica structures are shown in Figs. A.11-A.13.

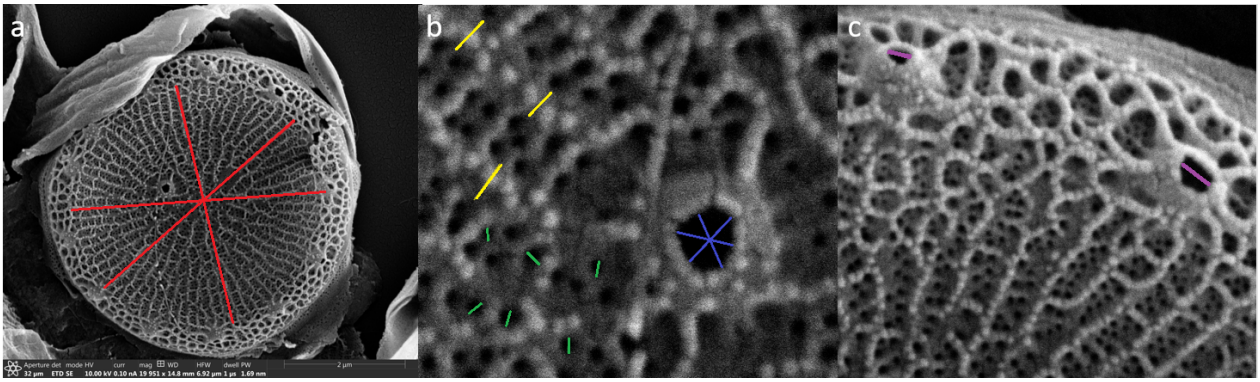

Figure A.10: An example of how the biosilica structures were measured. a) The valve diameter (red) was measured up to four times at different angles, from the inner ring of the ring and across the center. b) Fultuportulae diameter (blue) was measured at least twice at different angles. The costae width (yellow) was measured between silica bridges or branching, tangent to the length. Cribrum pore diameter (green) was measured once each. c) Rimoportulae diameter (purple) was always measured once each, tangent to the valve circumference.

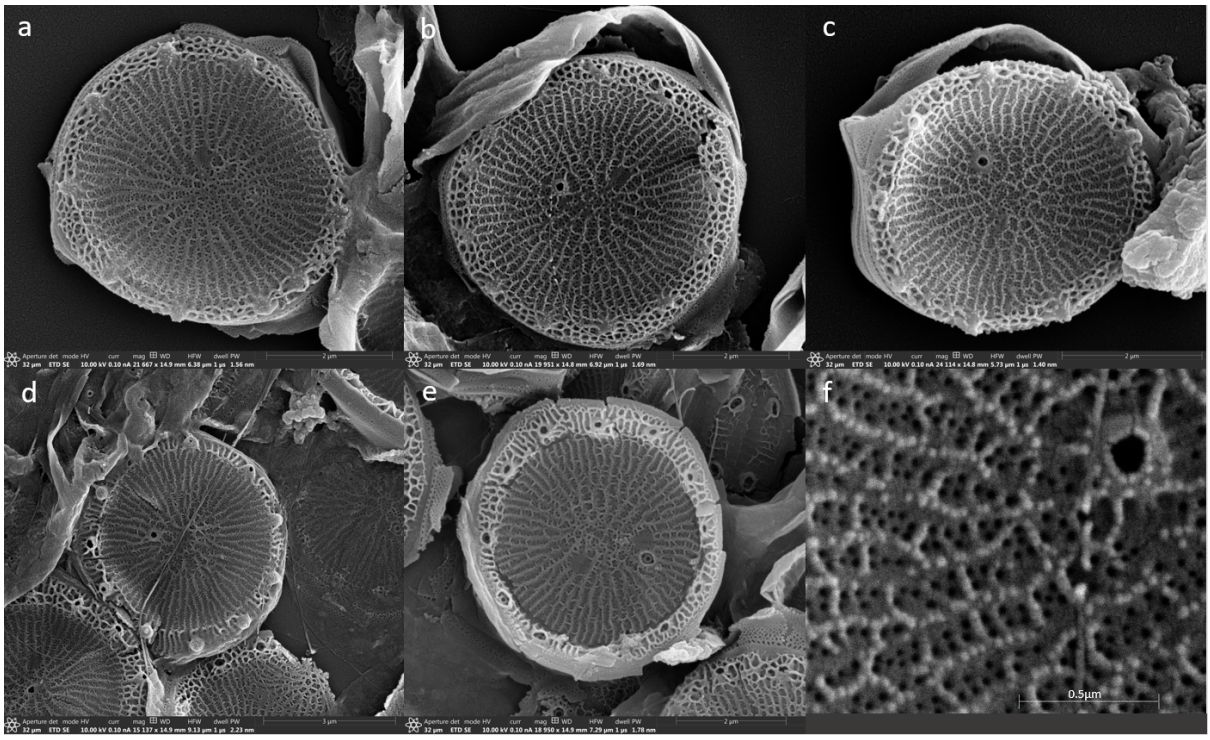

Figure A.11: SEM images of control samples. a)-e) Full size SEM image. Scale bar in a), b), c) and e) = 2 μm, in d) = 3 μm. f) Close up image of the valve in b), showing the fultuportulae, costae and cribrum pores. Scale bar = 0.5 μm.

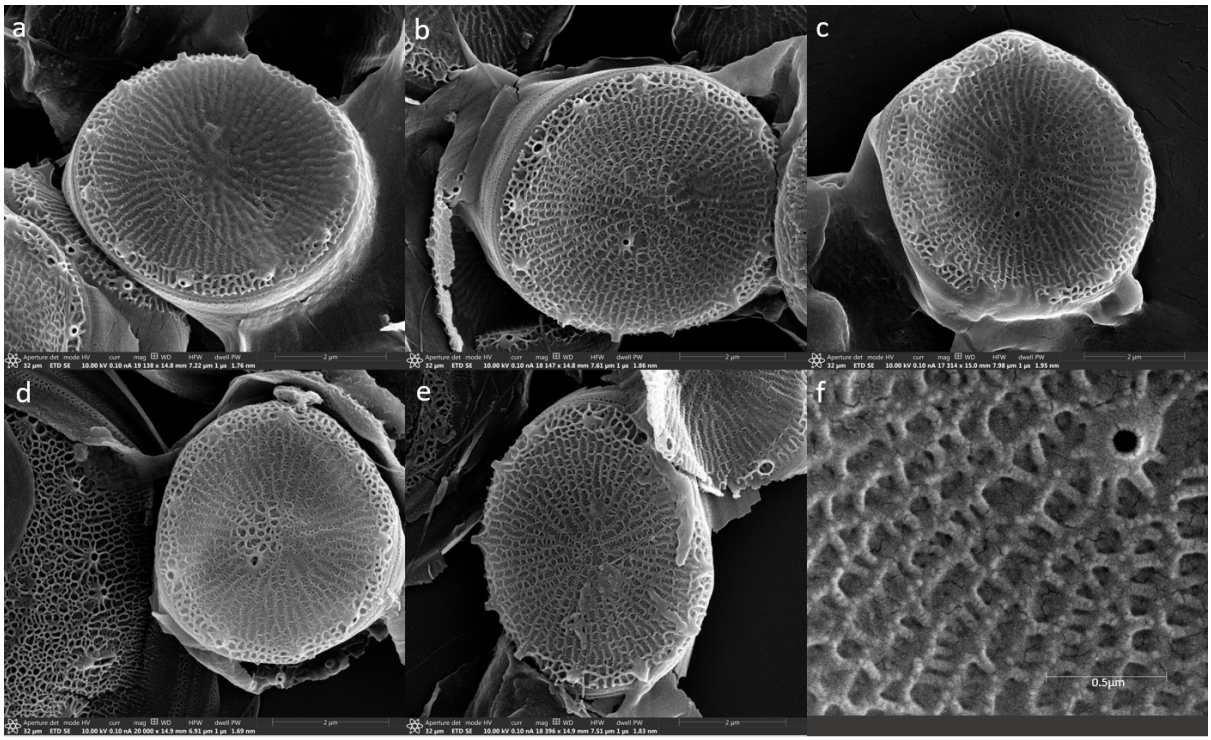

Figure A.12: SEM images of Al_{25} samples. a)-e) Full size SEM image. Scale bar = 2 μm. f) Close up image of the valve in b), showing the fultuportulae, costae and cribrum pores. Scale bar = 0.5 μm.

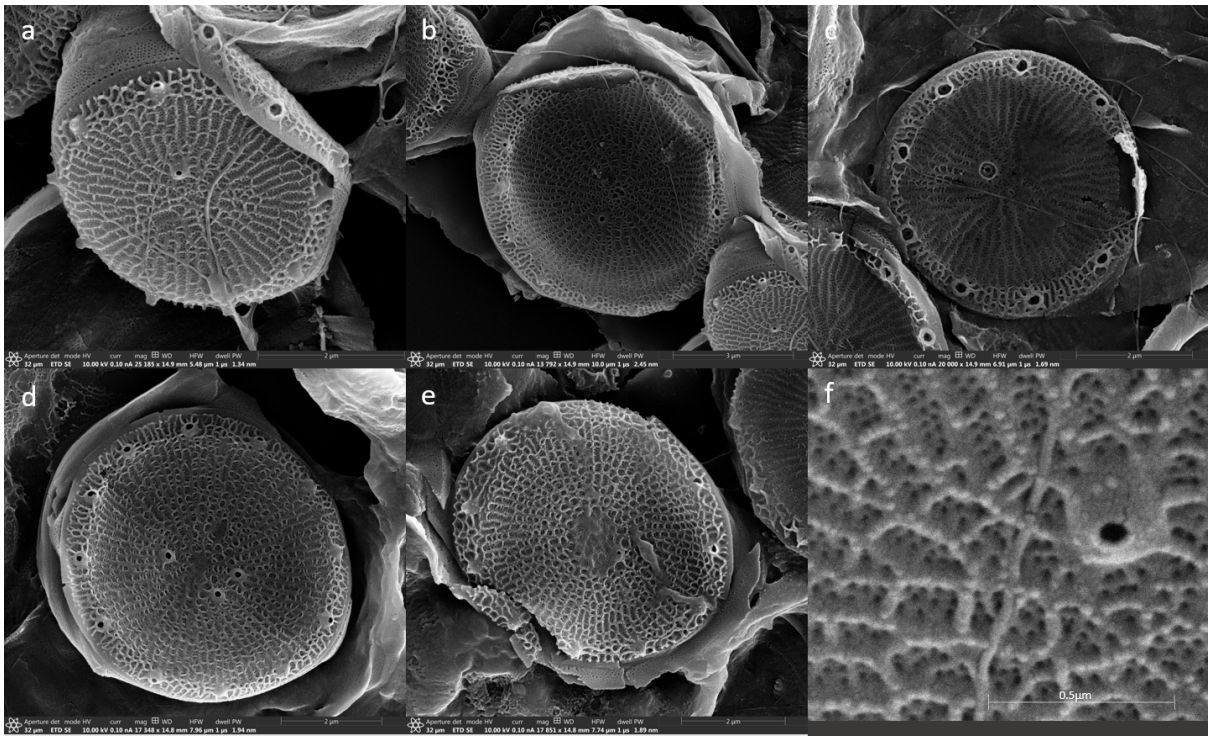

Figure A.13: SEM images of Cd_5 samples. a)-e) Full size SEM image. Scale bar in a), c), d)and e) = 2 μm, in b) = 3 μm. f) Close up image of the valve in b), showing the fultuportulae, costae and cribrum pores. Scale bar = 0.5 μm