Studying the targeting of frustule-associated proteins in the centric diatom *Thalassiosira pseudonana*

Dissertation

zur Erlangung des Grades eines

Doktor der Naturwissenschaften

(Dr. rer. nat.)

des Fachbereichs Biologie der Philipps-Universität Marburg

Vorgelegt von

Neri Fattorini

Aus Florenz, Italien

Marburg, 2021

Die vorliegende Dissertation wurde von Juli/2017 bis Juni/2021 am Fachbereich Biologie, Zellbiologie unter Leitung von Prof. Dr. Uwe G. Maier angefertigt.

Vom Fachbereich Biologie der Philipps-Universität Marburg (Hochschulkennziffer 1180) als Dissertation angenommen am _____

Erstgutachter:	Prof. Dr. Uwe G. Maier
Zweitgutachter:	Prof. Dr. Alfred Batschauer
Prof. Dr.	Prof. Dr. Andreas Brune
Prof. Dr.	Prof. Dr. Lars Voll

Tag der Disputation: _____

PUBLICATIONS

Parts of this work were included in:

Published

Fattorini, N., & Maier, U. G. (2021). Targeting of proteins to the cell wall of the diatom Thalassiosira pseudonana. *Discover Materials*, *1*(1), 1-10. (review article)

In preparation

Fattorini, N., & Maier, U. G. Studying the targeting of frustule-associated proteins in the centric diatom *Thalassiosira pseudonana*

“Se non esistessero le alghe
riusciresti a immaginarle?”

“If algae would not exist,
Could you imagine them?”

Bluvertigo
“Altre forme di vita”
Metallo non metallo (1998)

SUMMARY

LIST OF ABBREVIATIONS

bp: base pairs
$CaCO_3$: calcium carbonate
cDNA: complementary DNA
CDS: coding sequence
CinY1-4: Y-type cingulins (1-4)
CinW1-3: W-type cingulins (1-3)
RXL: Arginine-X-Leucine motif
eGFP: enhanced green fluorescent protein
fcp: Fucoxanthin chl a/c binding protein
DNA: deoxyribonucleic acid
DSi: dissolved silicon
ER: Endoplasmic reticulum
Fe_3O4: iron oxide
gDNA: genomic DNA
LCPA: Long-chain polyamines
MCS: multiple cloning site
NAT: nurseothrycin acetyl trasferase
NR: Nitrate reductase
G1: Gap 1 phase
S: Synthesis phase
G2: Gap 2 phase
M: Mitosis phase
PTM: Post-translational modifications
SP: Signal peptide
Si: Silicon
PAF: plastid autofluorescence
PPM: periplasmatic membrane
r.s.: restriction site
RNA: ribonucleic acid
SAP-1: Silicalemma-associated protein-1
SAP-3: Silicalemma associated protein-3
SDM: site-directed mutagenesis
SDV: Silica deposition vesicle
Sil1-4: Silaffin (1-4)
Sin1: Silicanin-1
SiO_2: silicon oxide
$SiOH_4$: metasilicic acid
SRP: signal recognition particle
TMD: trans-membrane domain
ZEO: zeocyn

LIST OF FIGURES

LIST OF TABLES

ABSTRACT

Diatoms are unicellular and photosynthetic organisms of highest ecological relevance. They are very well known because of their ability to exploit the dissolved silica available in the water to synthetize a modified cell wall, called the frustule. The frustule has a specific morphology for each species, and its formation process is complex and strictly intertwined with cell division, and therefore with the cell cycle. The diatom *T. pseudonana* has a strongly silicified frustule and is the first diatom species whose genome was sequenced, in 2004; since then, it has been the subject of many biomineralization studies and several proteins involved in the formation of the frustule, such as the silaffins and the cingulins, have been described, characterized, and localized *in vivo*. These proteins are believed to enter the secretory pathway at the ER, from which are then transported via vesicles to the Golgi, and from there to the silica deposition vesicle (SDV). When the synthesis of the frustule components inside the SDV is completed, its content is exocytosed outside of the cell. The only published study on the targeting of frustule proteins in *T. pseudonana* had identified short stretches of amino acids called pentalysine clusters (PLC) as potential targeting signal to the frustule of silaffins. Here, I have investigated the targeting of two cingulin representatives (CinW2 and CinY2) and one silaffins (Sil3) following different approaches: first, I have studied whether the presence of PLC-like regions in cingulins (called lysine-enriched regions (or LERs)) indicate that these regions play a similar role as it was for the PLC of silaffins; I have also studied if it was possible to identify a clear valve- or girdle band-targeting signal and, in that case, if this signal could be used to direct a valve-protein to the girdle or vice versa; second, since there is a remarkable difference in the gene expression between frustule-associated protein-encoding genes, I have investigated if the transcriptional regulation of the genes can affect the targeting of the encoded proteins; in parallel, I have obtained preliminary results regarding the isolation of the promoters of *sil3*, *cinW2*, *cinY2*. Additionally, I have identified unknown frustule proteins based on the presence of a PLC-like regions as well as a new discovered motif.

ZUSAMMENFASSUNG

Diatomeen sind einzellige, photosynthetische Organismen von höchster ökologischer Bedeutung. Sie sind bekannt für ihre Fähigkeit, die im Wasser gelöste Kieselsäure zu nutzen, um eine Zellwand, das so genannte Frustulum, zu synthetisieren. Das Frustulum hat für jede Art eine spezifische Morphologie, und ihre Synthese ist komplex und eng mit der Zellteilung und damit mit dem Zellzyklus verwoben. Die Diatomee *T. pseudonana* hat eine ausgeprägte silifizierte Zellwand und war die erste Kieselalgenart, deren Genom sequenziert wurde. Seit der Publikation der Sequenz im Jahre 2004 wurden zahlreiche Studien zur Biomineralisierung unternommen, und mehrere Proteine, die an der Bildung der Zellwand beteiligt sind, wie die Silaffine und Cinguline, wurden beschrieben, charakterisiert und *in vivo* lokalisiert. Es wird angenommen, dass diese Proteine in den sekretorischen Weg am ER eintreten, von wo aus sie dann mittels Vesikel zum Golgi und von dort zu einem Kompartiment, das „Silica Deposition Vesicle“ genannt wird (SDV), transportiert werden. Hier werden dann Zellwand-Komponenten synthetisiert und exozytiert. In der bisher einzigen publizierten Studie über das zelluläre Targeting von Frustulum-Proteinen in *T. pseudonana* wurden in Silaffinen kurze Abschnitte aus Aminosäuren, so-genannte Pentalysincluster (PLC), als potenzielle Targeting-Signale für das Frustulum identifiziert. In meinen Arbeiten habe ich das Targeting von zwei Cingulin-Proteinen (CinW2 und CinY2) und einem Silaffin (Sil3) mittels verschiedener Ansätze studiert. Zunächst wurde untersucht, ob PLC- ähnliche Regionen in Cingulinen (so genannte Lysine Enriched Regions oder LER) eine ähnliche Funktion haben wie die PLC von Silaffinen. Weiterhin war von Interesse, ob es möglich ist, „Valve“- oder „Girdle Band“-Targeting-Signale zu identifizieren, mit deren Hilfe ein „Valve“-lokalisiertes Protein in die „Girdle Bands“ dirigiert werden kann und vice versa. Da zwischen Genen, die für Frustulum-assoziierte Proteine kodieren, bemerkenswerte Unterschiede in der Genexpression bestehen, wurde weiterhin untersucht, ob die Transkriptionsregulation dieser Gene das Targeting der kodierten Proteine beeinflussen kann. Parallel hierzu wurden vorläufige Ergebnisse hinsichtlich der Isolierung der Promotoren von sil3, cinW2 und cinY2 erzielt. Darüber hinaus konnten, anhand von PLC ähnlichen Regionen sowie eines neue entdeckten Motivs, bisher unbekannte Zellwandproteine identifiziert werden.

1. INTRODUCTION

1.1 BIOMINERALIZATION

1.1.1 WHAT IS BIOMINERALIZATION?

Biomineralization is defined as the "*ability of living organisms to synthesize solid inorganic materials by using nutrient resources in their natural environment*" (Richthammer et al., 2011); the biominerals that are formed can be described as "*minerals formed from the remains of dead cells, or as the result of the metabolic activity of live cells"; these biominerals* "*encompass a wide range of chemistry and mineralogy, i.e., carbonates, silicates, sulfides, oxides*" (Sigel et al., 2008). Natural selection and evolution have driven the development of a large variety of adaptations, resulting in a multitude of different species – almost from every taxonomic group – capable of creating complex biomaterials, diverse in terms of their structure, chemical composition, morphology, and biological role (Kröger & Poulsen, 2008). At the present day, 70 different biomineral chemistries have been described, with the calcium carbonate-, calcium phosphate- and silica-based ones being the most abundant (Kröger & Brunner, 2014). Some examples are the bones and the teeth of animals, the Fe_3O_4 crystals of magneto-tactic bacteria, the complex cell walls of coccolitophores ($CaCO_3$-based) and radiolaria (SiO_2-based), and the calcareous or siliceous skeletal spicules of some sponges (Kröger & Poulsen, 2008), just to name a few.

In respect to biomineralization, one of the most thoroughly studied group are the diatoms. These organisms create a modified cell wall (called frustule) made of amorphous glass (SiO_2), finely decorated with extremely complex nano- and micro-scaled patterns of pores, ribs, and other silica structures. In the next section I will briefly describe the relationship between environmental silicon and diatoms.

1.1.2 SILICON BIOMINERALIZATION

Silicon (Si) is the seventh most abundant element in the Universe (P. J. Tréguer & De La Rocha, 2013) and the second most abundant element on Earth, making up 27.7% of earth's crust (Link1). Silicon does not occur uncombined, but mainly as silicates or in the oxidized form, as silicon dioxide (also called silica, or SiO_2). Virtually all dissolved silicon (DSi) in the oceans comes from the weathering of the Earth's crust, mainly thanks to four pathways: continental inputs (rivers and submarine water discharges), aeolian inputs, hydrothermal inputs, and dissolution of marine sediments (P. J. Tréguer & De La Rocha, 2013).

Biogenic silica is produced in the photic layer by siliceous organisms (radiolarians, silicoflagellates and diatoms) and it is in part recycled in the upper layer upon dissolution or entrapped in the seafloor as sediments where it continues to dissolve. In marine and freshwater environments, the predominant form of dissolved silicon is ortosilicic acid [$Si(OH)_4$], which is the chemical form of DSi that diatoms use to generate their cell walls (Sumper & Brunner, 2008). The average global oceanic concentration of ortosilicic acid is approximately 70 μM; polymerization does not occur when this chemical form of Si is in aqueous solution at concentrations not exceeding 2 mM; at higher concentrations, instead, ortosilicic acid polymerizes into colloidal particles (Sumper & Brunner, 2008).

Diatoms have a fundamental influence on the global silica cycle, as they are the group of marine organisms that contributes the most to the production of biogenic silica (P. J. Tréguer & De La Rocha, 2013): it has been calculated that once that an atom of Si has been delivered to the ocean, it passes through the biological silicon uptake/dissolution cycle 39 times, before being removed to the seabed (P. Tréguer et al., 1995). Diatoms master the ability of forming biominerals and, like the craftsmen from the past, they pass their secrets down from one generation to the others; in fact, their silica shells' shapes, their patterns of decoration, and every other morphological feature, are build following a species-specific design (Round et al., 2007). This characteristic allowed generations of taxonomists and diatoms' enthusiasts, to identify and classify the diatoms, based on their frustule's morphology.

1.2 DIATOMS

1.2.1 BIOLOGY, ECOLOGY AND EVOLUTION

Our world depends on phytoplankton: despite representing only the 0.2% of the global primary producers' biomass, phytoplankton is in fact responsible of 45% of the global net primary production (defined as the amount of photosynthetically fixed carbon available for the first heterotrophic level in an ecosystem) (Field et al., 1998).

Diatoms (class *Bacillariophyceae*) are unicellular photosynthetic *Stramenopiles* (Graham et al., 2016) and, together with dinoflagellates and coccolitophores, the most important phytoplanktonic group (Falkowski et al., 2004). They live in every fresh- and sea-water habitats, especially in nutrient-rich and upwelling zones, where they frequently are the dominant group. They are widely distributed all over the world and they can be pelagic or associated with the sediment (benthic diatoms) (Graham et al., 2016).

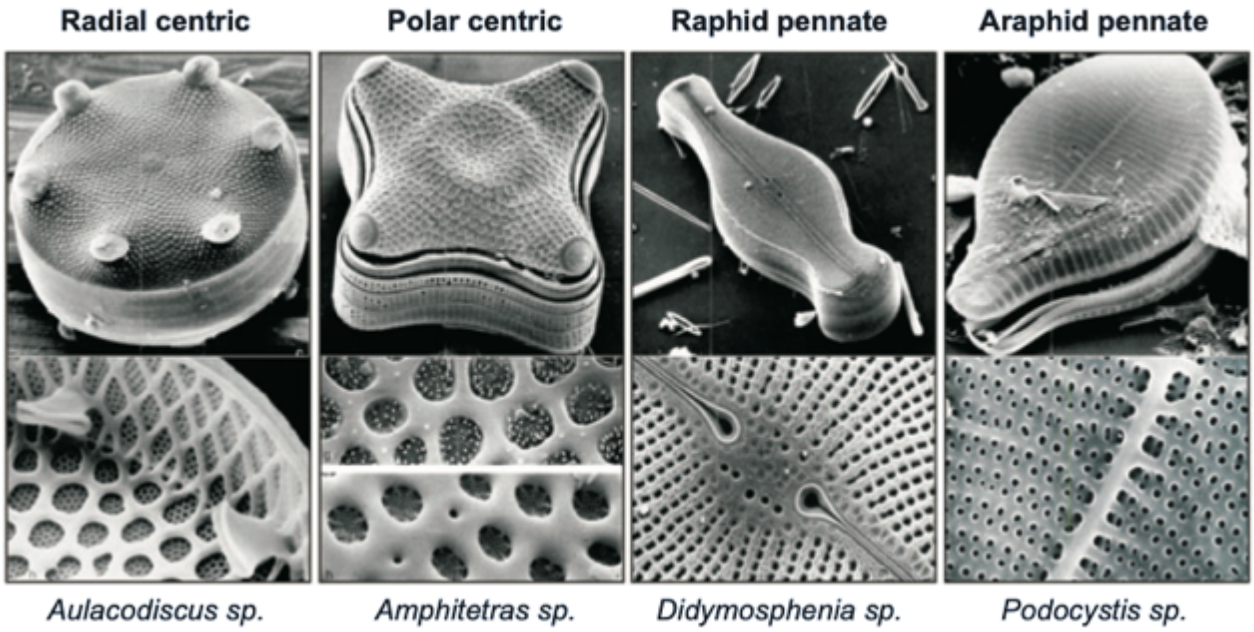


Figure 1 Some representative diatom species. The image shows examples of the diversity of shapes, symmetry, and patterns of decoration. The bottom row shows details of the valve shown in the upper row (from Kröger and Poulsen, 2008).

Diatoms are also the most diverse group of phytoplankton, estimated to be around 200000 extant species (Mann & Droop, 1996). When scientists describe them, one of the first things they usually highlight is the importance of these organisms as oxygen producers. They are, as a matter of fact, responsible for approximately 45% of total ocean's production (20% of global O_2 production), which is comparable to that of all rainforests together (Armbrust, 2009). But their ecological relevance on the global scale is not limited to the production of oxygen; as marine primary producers, in fact, diatoms support the global oceanic food chain (Armbrust, 2009; Benoiston et al., 2017).

The fluctuations in diatoms' communities have a huge impact on a global scale, influencing the carbon pump (*i.e.*, the amount of carbon that is sequestered into the ocean depths), the amount of CO_2 that is removed from the atmosphere and, ultimately, the global climate (P. Tréguer & Pondaven, 2000). The earliest diatom fossil is dated 190 million years ago (Early Jurassic), but they are thought to have arisen earlier, perhaps during the Triassic period, around 250 million years ago (Armbrust, 2009). During this period, dead diatom cells have sunk to the depths of the sea, accumulating on the seafloor, and creating overlaying layers of silica sediments, which we call *diatomite* or *diatomaceous earth* (Graham et al., 2016). The evolutionary history of diatoms is reflected in their cellular features (Benoiston et al., 2017; Maier et al., 2015).

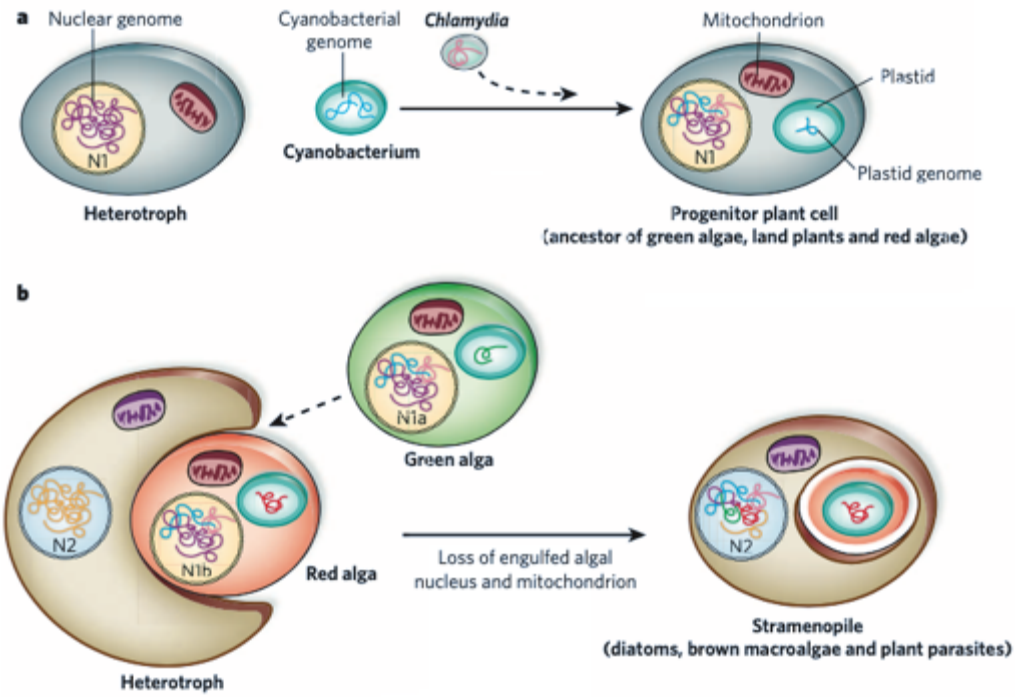


Figure 2 Origin of diatoms' plastids. Schematic representation the two endosymbiotic events that generated the diatoms. The evolution of the primary and secondary endosymbiont, the transfer of DNA from the symbiont to the host nucleus, and the loss of organelles and in the symbiont, have resulted in the Stramenopile-like type of cell (from Armbrust, 2009).

They have so-called *complex plastids*, chloroplasts surrounded by four membranes (Kroth, 2002; Maier et al., 2015); these four plastid membranes are an inheritance of the two endosymbiotic events that generated the *Chromalveolates* group (composed of heterokontophytes, haptophytes, cryptophytes, apicomplexa and peridinin-containing dinoflagellates) to which also diatoms belong (Maier et al., 2015). The first endo-symbiotic event took place around 1.5 billion years ago, when a cyanobacteria was engulfed (or invaded) by a heterotrophic eukaryote, creating the two membranes-bound plastid ancestor of land plants, red algae, and green algae (the innermost membrane comes from the cyanobacterium ancestor, whereas the outermost one from the first eukaryote host cell); consequently to this event, during the evolution of these ancestral endo-symbiotic organisms, many cyanobacterial genes have been transferred from the symbiont genome to the host genome (10% of plants nuclear genes are derived from cyanobacterial genome (Armbrust, 2009). A second endo-symbiotic even occurred approximately 500 million years later, when another heterotrophic eukaryote engulfed a red alga and reduced it to a complex plastid surrounded by four membranes, with the second outermost membrane (the periplastidal membrane, or PPM) deriving from the cytoplasmic membrane of the red alga, whereas the outermost membrane derives from a phagotrophic vacuole membrane

(Maier et al., 2015). As a result of this evolutionary process, the gene transfer from the red alga nuclear and plastid DNA to the host nucleus has shaped diatoms' genomes, creating a mix of plant-like and animal-like genes. For example, diatoms are equipped with a complete urea cycle (never observed in plants), and metabolic pathways used to generate energy from the breakdown of fat (like animals), but also to create chemical intermediates for other reactions (like plants); additionally, at least 170 red-algal genes have been identified in diatoms genomes, and some of them encode for plastid proteins (Armbrust, 2009). Their cellular features and metabolic adaptations make diatoms a fascinating group; in the last decades they gained a lot of attention because of their biotechnological potential to produce carbon neutral biofuels and other valuable compounds (see section 1.5).

One diatom feature has gained particular attention, their frustule (see **Figure 1**). The modified silica-made cell wall has been studied intensively, with the goal of exploiting diatom's ability to form these complex structures, and to generate bioinspired materials.

1.2.2 DIATOMS' FRUSTULE

1.2.2.1 MORPHOLOGY

Despite their outstanding diversity and complexity, diatom's cell walls share a common base structure (Ross et al., 1979; Round et al., 2007; see **Figure 3**): each frustule is composed of two halves, named *thecae*, that fit into each other like the two parts of a shoe box and the larger one (the *epitheca*) accommodates the smaller one (the *hypotheca*). Each theca is further subdivided in a *valve* and a *cingulum* (both can epi- or hypo-, depending on the theca to which they belong). Following the shoe box analogy, the valve would be the lid (or the bottom of the box), while the cingulum would be the sides of the lid (or of the bottom of the box); the epi- and hypocingulum together form the *girdle* of the cell and each cingulum is composed of a series of overlapping ring-like structures called *girdle bands*; the girdle band proximal to the valve is called *valvocopula*, whereas the terminal girdle band is called *pleural band* and corresponds to the region of the frustule where the epitheca overlaps with the hypotheca. Beyond this general base structure, what differentiates each frustule from the others is substantially the shape and the morphology of the valve (which allows to distinguish between *centrics* and *pennate* diatoms), its symmetry (*radial* or *polar* diatoms), and its pattern of decoration (the latter is normally more intricated in the valve respect to the girdle; see **Figure 1**).

Before the advent of molecular taxonomy, the morphology of the frustule has been used for the classification of diatoms down to the species level.

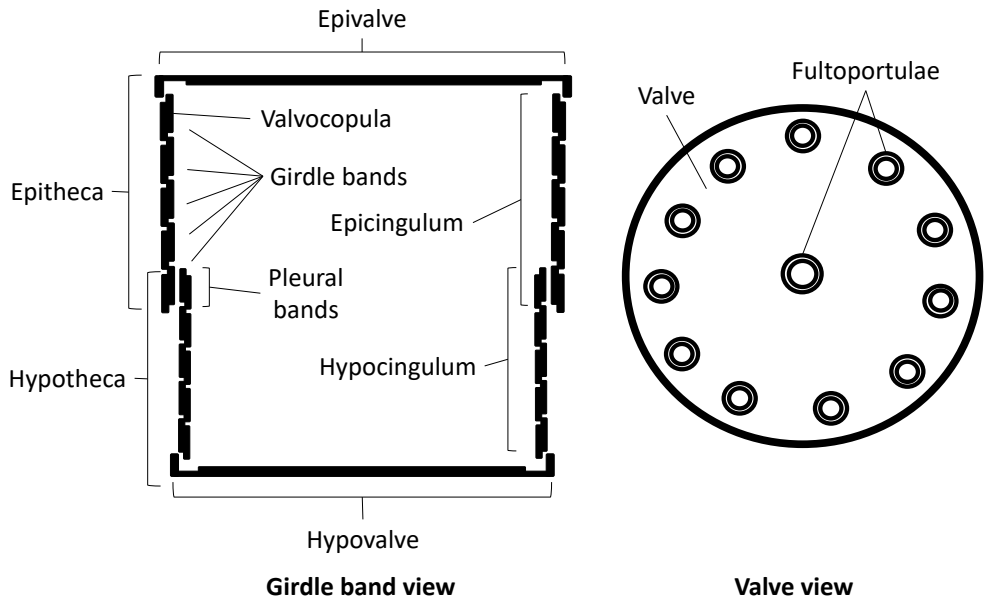


Figure 3 *T. pseudonana*'s frustule. This schematic rendering shows some of the features of the cell wall of diatoms. For details, see text.

1.2.2.2 MORPHOGENESIS

During vegetative growth diatom divide via mitotic divisions, but directly after cytokinesis the protoplasts of the newly formed daughter cells are still retained inside the mother cell's frustule and, before being able to separate, each cell must synthesize a new theca (**Figure 10**). Because of this constraint, each daughter cell inherits only one theca from the parental cell and synthesizes the other one *de novo* (Round et al., 2007); therefore, the inherited parental theca becomes the epitheca of the new cell, while the new theca becomes the hypotheca. After this step, the two daughter cells separate from each other, and each one synthesizes new girdle bands, thereby increasing its valve-to-valve distance and cell volume. Because of this peculiar division process, usually diatoms reduce their average population size with the proceeding of the generations, at least until they reach a certain threshold value that triggers the switch to sexual reproduction and the production of gametes (this size is usually one-third of the species' maximum size). The fusion of the gametes generates a zygote, the zygote generates an auxospore, and the auxospore generates an initial cell, whose size is the maximum size of the species; after the formation of the initial cell, the vegetative pattern of growth starts again (Graham et al., 2016). Despite this common diatom peculiarity, some extensively studied species (*Thalassiosira pseudonana*, *Phaeodactylum tricornutum and Cylindrotheca fusiformis)* show no size

reduction when grown in laboratory conditions, but the mechanism that they use to avoid shrinking is still unknown (Kröger & Poulsen, 2008). The formation of the new components of the frustule is a complex process, which is intertwined with the cell cycle and with cell division (Hildebrand et al., 2007), thus requiring an extremely fine regulation of the expression of the genes involved (as better detailed in section 1.4). This process represents a massive event for the cell, involving several organelles and structures, and a still not completely well understood intracellular trafficking pathway for the transport of the needed molecules (Poulsen et al., 2013; see also section 2.3).

One of the organelles involved in that is the silica deposition vesicle (SDV) (Drum & Pankratz, 1964); this organelle is provided with its own membrane (the silicalemma), and it is a cisternal compartment underlying the cytoplasmic membrane when fully expanded. The SDV is the organelle where the various components of the frustule are synthesized (Graham et al., 2016): inside and outside the SDV, an orchestrated team of molecular players (*e.g.*, soluble and membrane proteins, cytoskeleton, actin filaments) directs the nanopatterned precipitation of the silica and defines the moulding of the SDV, which will ultimately determine the shape, size, and pattern of decorations of the valves and the girdle bands (reviewed in Hildebrand et al., 2018). In the case of the valve, for example, the process starts with the formation of a base layer resembling the shape and the morphology of the valve; this layer is then thickened via silica precipitation, and the features of the valve (*i.e.*, ribs, costae, fultoportulae, cribrum pores) are formed (Heintze et al., 2020; Hildebrand et al., 2006). Once the synthesis of the biosilica is completed, the SDV content is exocytosed outside of the plasma membrane via a yet not completely understood mechanism (Pickett-Heaps, 1990). Recent data obtained by *in vivo* localization studies using the silicalemma protein Sin1 (Kotzsch et al., 2017) supports one of the previously proposed mechanism of exocytosis (reported in Pickett-Heaps, 1990), in which the distal silicalemma fuses with the plasma membrane at the centre of the SDV, and that the pulling of the membranes would trigger the exocytosis; the distal silicalemma would then become the new plasma membrane, whereas the old plasmalemma and the proximal silicalemma would be recycled (Pickett-Heaps, 1990).

The SDV shares some characteristics with the vacuole of higher plants: for example, it is an acidic compartment (Marty, 1999); additionally, a recent study described a V-type-H+-ATPase (VHA) as being responsible for the acidification of the SDV, depending on the ongoing cell wall synthesis step (and therefore on the cell cycle phase), VHA is translocated between the plasma membrane, the silicalemma, and the tonoplast

(*i.e.*, the vacuole-surrounding membrane; see also **Figure 9**); pharmacological inhibition of VHA prevented the acidification of the SDV, the synthesis of the frustule, the fusion of the SAP1-containing vesicles with the SDV (see 1.2.5.3), and it also affected the morphology of the frustule (Yee et al., 2020).

1.2.3 THE MODEL SPECIES *THALASSIOSIRA PSEUDONANA*

T. pseudonana is a centric diatom, belonging to the *Thalassiosirales* order; this species was the first diatom to have the genome sequenced (Armbrust et al., 2004) and since then it became a model species, especially in respect to the study of biomineralization in diatoms (Hildebrand et al., 2018; Hildebrand & Lerch, 2015; Poulsen & Kröger, 2004; Sumper & Brunner, 2008).

T. pseudonana has 11242 predicted genes in the nuclear genome, 144 plastid-encoded genes, and 40 mitochondrial-encoded genes (Armbrust et al., 2004). The molecular toolkit for this species has shortly followed the release of the genome (Poulsen et al., 2006), allowing researchers to exploit *T. pseudonana* for a large variety of transcriptomic-, proteomic, and lipidomic-based studies and experimental approaches, such as: the *in vivo* localization of proteins (Kotzsch et al., 2016, 2017; Poulsen et al., 2013; Samukawa et al., 2014; Scheffel et al., 2011; Tachibana et al., 2011; R. Tanaka et al., 2014; Tesson et al., 2017; Wustmann et al., 2020), the study of the intracellular trafficking of molecules (Kotzsch et al., 2017; Yee et al., 2020), the development of inducible transformation cassettes (Roshan P. Shrestha & Hildebrand, 2017), the study of diatoms' physiology and metabolism (Ashworth et al., 2013; Bender et al., 2012; Bromke et al., 2013; Clement et al., 2016; Dong et al., 2016; Dyhrman et al., 2012; Gruber et al., 2009; Hockin et al., 2012; Martin et al., 2011; Thamatrakoln et al., 2012; Yang & Gao, 2012; Yu et al., 2009), the study of their genetics and evolution (Alverson et al., 2011; Hopkinson et al., 2016; Montsant et al., 2007; Norden-Krichmar et al., 2011; Secq & Green, 2011), the study of their response to pollutants and toxic chemical (Carvalho, Bopp, et al., 2011; Carvalho, Burchardt, et al., 2011; Carvalho & Lettieri, 2011; Wu & Wang, 2011), and the study of their cellular organization and biochemical composition (Grouneva et al., 2011), just to name some examples.

In the last years, the molecular toolkit has also been enriched with the latest developments in molecular biology and genetic transformation (see for example: Görlich et al., 2019; Nawaly et al., 2020).

1.2.4 THE FRUSTULE OF *T. PSEUDONANA*

The cells of *T. pseudonana* are relatively small, the valve diameter is in fact approximately 3.0-4.2 µm, and the valve-to-valve distance is approximately 4.3-9.0 µm (Hildebrand et al., 2006). The frustule has a cylindrical shape (**Figure 4**), the valves are round and decorated with silica ribs (*costae*) radiating outwards, from a central silica structure (called *fultoportula*) towards the rim of the valve, where an additional series of fultoportulae is arranged in a circular fashion (Heintze et al., 2020; Hildebrand et al., 2006). The ribs are interconnected with transversal silica ridges, called *cross-connections,* creating trapezoidal areas named *areola pores* (Görlich et al., 2019) that usually enclose three or four *cribrum pores*. The girdle bands possess overlapping structures called *ligula*, located on opposite sides of the girdle on alternate bands (Hildebrand et al., 2006).

The intracellular organization shows a large central vacuole, squeezing most of the other cellular organelles to the periphery of the cell; the plastid is generally large, and often located close to the valves. The cell is tightly encased in the frustule, which is composed of the biosilica exocytosed from the SDV and other organic molecules associated with it. Several classes of *T. pseudonana*'s frustule biomolecules have been identified, described and characterized in the last 20 years, since the identification of the first frustule-associated proteins in this species (Poulsen & Kröger, 2004).

1.2.5 CELL WALL-ASSOCIATED BIOMOLECULES

Diatoms' biosilica is a complex organic–inorganic hybrid material composed not only of silica, but also proteins, carbohydrates, and other molecules. Some of these molecules are structural components of the biosilica (*e.g.*, silaffins, cingulins, silacidins, LCPA), some others instead, such as some rhamnose- and xylose-rich glycoproteins, are necessary to protect the frustule from dissolution (Armbrust et al., 2004).

It is possible to extract biosilica (see also section 5.1.4) from live cells by disruption of the cell's membranes and organelles, while maintaining intact the cell wall (Poulsen & Kröger, 2004); the frustule can then be dissolved using a mildly acidic solution of ammonium fluoride, resulting in one fraction containing the soluble material and another fraction containing the insoluble material (see for example Scheffel et al., 2011; reviewed in Hildebrand et al., 2018;).

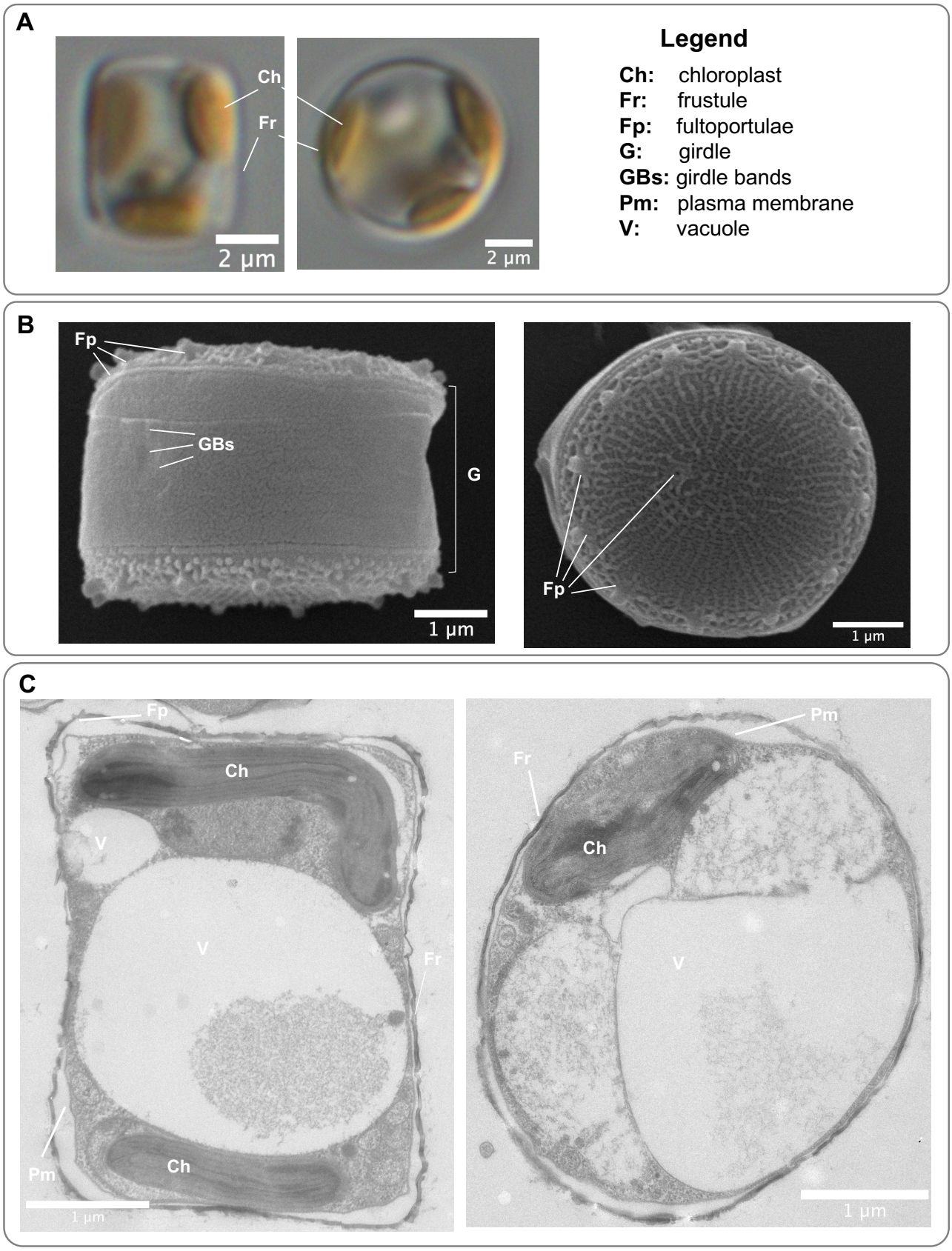


Figure 4 The *T. pseudonana*'s cell. Panel **A** shows images taken with an optical microscope; panel **B** shows scanning electron microscope (SEM) images of the frustule; panel **C** shows transmission electron microscope (TEM) images of cells' sections. Sample preparation and SEM imaging were performed by Dr. Karl-Heinz Rexer; sample preparation and TEM setup were performed by Dr. Thomas Heimerl.

This approach has several advantages: for example, to confirm that the signal is associated with the frustule, it can be used to compare the results of *in vivo* localization experiments with the fluorescent tagging of isolated biosilica, (Poulsen et al., 2013; Scheffel et al., 2011); another extremely useful application of this protocol is for the isolation of frustule-associated material and its characterization (for example novel frustule-associated proteins) and components (such as valves or girdle bands) (Kotzsch et al., 2016; Poulsen & Kröger, 2004; Wenzl et al., 2008). Some of the molecules that have been discovered in the last years using this strategy are described in the following sections.

1.2.5.1 SILAFFINS

The genome of *T. pseudonana encodes* for 4 silaffins, Sil1, Sil2, Sil3 and Sil4. The first silaffins identified – Sil1, Sil2 and Sil3 – were isolated from the frustule of *T. pseudonana* after dissolution of the biosilica with a mildly acidic solution (pH=4) of ammonium fluoride, followed by separation via ion exchange and gel permeation chromatography (Poulsen & Kröger, 2004); a fourth member (Sil4) has been identified later (Sumper & Brunner, 2008; Poulsen et al., 2013). The characterization of the silaffins revealed that they were acidic phosphoproteins, extensively post-translationally modified by glycosylation and sulfation (Poulsen & Kröger, 2004). All silaffins bear a SP for co-translational import into the ER (Poulsen et al., 2013; Poulsen & Kröger, 2004; see also **Figure 5** and section 7.3). From the ER, they are transported to the cell wall using a still unknown pathway (Fattorini & Maier, 2021) most likely involving a passage through the SDV, as these proteins are embedded in the biosilica matrix.

T. pseudonana's silaffins Sil1 and Sil2 share 91% of sequence similarity and exist in two isoforms each, denoted H and L (for "high" or "low", according to their apparent mass on an SDS-PAGE). The difference in size between the two isoforms is due to the alternative processing that the pro-peptides undergo during their maturation: after the SP, in fact, both Sil1 and Sil2 pro-peptides have an acidic N-terminal domain enriched in serines, prolines and threonines, followed by a strongly basic C-terminal domain in which is maintained the predominance of serines, but prolines and threonines are reduced. During the processing of the pro-peptides the N-terminal domain is cleaved off, resulting in the mature Sil1L and Sil2L isoforms; the higher mass isoforms of the proteins, Sil1H and Sil2H, are derived from the same precursor peptides of the -L isoforms but without the cleavage of the N-terminal domain.

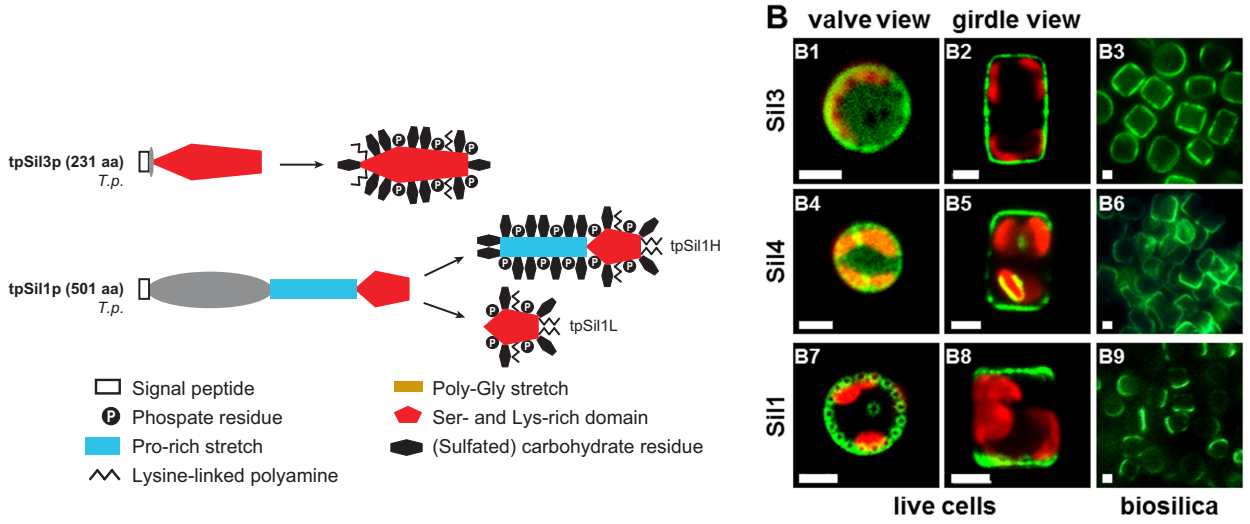


Figure 5 Silaffins. On the **left** are shown the schematic primary structures of Sil1 and Sil2 from *T. pseudonana*: grey ovals indicate portions of the proteins that are removed during their maturation (from Kröger and Poulsen, 2008). On the **right** are shown the results of *in vivo* localization experiments of eGFP-tagged silaffin fusion proteins, showing also the fluorescent tagging of the extracted biosilica (from Poulsen et al., 2013).

Also Sil3 undergoes post-translational processing, as it is synthetized as a precursor polypeptide having a 9 amino acids-long pro-peptide immediately following the SP, which is absent from the mature peptide.

Each protein or isoform has specific properties regarding its silica precipitation properties: when the Sil1/2H or Sil1/2L isoforms are mixed with long-chain polyamines (LCPA; see also section 1.2.5.5), the silica precipitation activity of the mixture is different, with the silica precipitation activity of lower molecular mass isoforms (Sil1/2L) increasing with increasing protein concentration, whereas the two higher molecular mass isoforms (Sil1/2H) show a regulatory effect on silica precipitation with increasing protein concentration. This effect might be explained by the different amino acid composition of the L and H isoforms: the acidic N-terminal domain of Sil1 and Sil2 has a high density of hydroxyamino acids that are most likely to be glycosylated or sulfated, than phosphorylated; it was previously demonstrated for another diatom species, *Cilindrotheca fusiformis*, that carbohydrate moieties inhibit silica precipitation whereas phosphate groups promote it (Poulsen et al., 2003); it is possible therefore that the higher density of phosphate groups, respect to sulfate or carbohydrate groups, of the basic C-terminal domain of the L-isoforms of Sil1 and Sil2, determines the silica precipitating activity of these isoforms. This could happen because of the formation of hydrogen bonds between the Si-OH groups of silicic acid and the hydroxyl groups of the carbohydrates moieties attached to the amino residues in the Sil1/2-H isoforms; these hydrogen bonds compete

with the hydrogen bonds formation between the same Si-OH groups and the amino groups of LCPA (Poulsen & Kröger, 2004; reviewed in Hildebrand et al., 2018 and Kröger & Poulsen, 2008;). Sil3 shows a similar behaviour to the high mass isoforms of Sil1 and Sil2: when Sil3 is mixed with LCPA silica precipitation is promoted, although only up to a 60-80 μM protein concentration; when the concentration is instead higher than this value, silica precipitation is inhibited.

Also the morphology of the precipitated silica is quite different between each protein-LCPA mixture tested: when Sil1/2L isoforms are mixed with LCPA, they drive the formation of silica spheres at all protein concentrations tested, with the average diameter increasing from 230 nm (at 10 μM) to 2.3 μm (at 100 μM); mixtures of LCPA with Sil1/2H generated irregular, porous silica sheet of non-uniform shapes; mixtures of LCPA with Sil3 formed two distinct kind of structures, namely plates of densely packed, extremely small silica particles, and large polydisperse silica spheres with diameter ranging from 0.9 μm to 4.2 μm (Poulsen & Kröger, 2004). Since the silaffins alone cannot precipitate silica particles *in vivo* without LCPA, and that the presence of multivalent inorganic ions (like phosphate) in LCPA mixtures was proven to drive the formation of silica spheres (Kröger et al., 2000), it has been hypothesized that the role of silaffins is merely to offer abundance of negative charges for the LCPA, and that silaffins alone cannot precipitate silica because they lack the LCPA modification of lysine residues that are instead present in *C. fusiformis* (Poulsen & Kröger, 2004).

In the protein Sil3 from *T. pseudonana*'s, 30 out of 33 total lysines are embedded in a K-(A/S/Q)-X-K tetrapeptide. Sumper et al., 2007 have studied the pattern of lysine-associated post-translational modifications with di-methyl moieties, diazaoctanyl moiteies, or quaternary ammonium groups and, based on their results, they described a pattern of modification that follows specific rules, depending on: i) the position of the lysine in the tetrapeptide; ii) if the tetrapeptide motif is clustered with other motifs,; iii) if a single lysine is located no more than two amino acid away from a motif. Given the specificity of this modification code, the presence of several K-(A/S/Q)-X-K motifs and the properties of Sil3, it has been hypothesised that this modification scheme determines the overall silica precipitation properties of Sil3 by influencing its self-assembly and the interactions with the other frustule-forming biomolecules (Sumper et al., 2007), and also that the targeting of Sil3 is somehow depending on the type of modification that characterizes the lysines (Poulsen et al., 2013). These KxxK motifs are typical of silaffins but are present also in other proteins, like the cingulins (Scheffel et al., 2011), suggesting that they might play a

role both in influencing the silica precipitation properties of the proteins and in their targeting (see section 1.3).

The *in vivo* localization of silaffins by eGFP-tagging and confocal imaging (Poulsen et al., 2013; Scheffel et al., 2011; see **Figure 5**) confirmed that these proteins are associated with the frustule and indicated specific sub-localization for each protein: Sil1 was localized in the area of the valve surrounding the fultoportulae; Sil4 was localized homogenously in the valve and in some gridle bands; Sil3 was localized in the valve (leaving gaps in the fluorescent signal resembling the fultoportulae pattern) and in the area of the girdle (with some difference depending on the promoter used for the expression of the gene, see section 1.4.3).

An interesting characteristic of silaffins is the fact that these proteins share no sequence homology between diatom species (Poulsen and Kröger, 2004); it appears therefore, that their amino acid composition and post-translational modifications, rather than the amino acid sequence itself, are important for the functioning of this proteins. As detailed in the next section, this assumption allowed to use these criteria to search for new frustule-proteins.

1.2.5.2 CINGULINS

Following a bioinformatics screening of *T. pseudonana*'s proteome a family of six cingulin proteins has been identified (Scheffel et al., 2011; an additional one was identified from Kotzsch et al., 2016), based on the presence of a SP and an acidic domain (at least 100 amino acid-long) containing ≥18% serines and ≥10% lysines. These six proteins were grouped, according to their amino acid composition, in two types: the W-type cingulins (CinW1-3) are enriched in tryptophan residues, while the Y-type cingulins (CinY1-4) are enriched in tyrosine residues (**Figure 6**).

In vivo localization experiments identified these proteins as associated with the cingulum, and the proteins were therefore named cingulins (Scheffel et al., 2011). After dissolution of the frustule, these proteins are found in the ammonium fluoride insoluble material (AFIM). Scanning electron microscope (SEM) analysis of the AFIM fraction obtained by cingulin-eGFP expressing transformants demonstrated the presence of organic, nanopatterned microrings, resembling the morphology of the non-porous parts of the girdle; these microrings were capable of precipitating silica particles with or without LCPA (Scheffel et al., 2011). Based on these results, it has been hypothesized (Scheffel et al., 2011) that the microrings are part of a continuous organic layer embedded in the

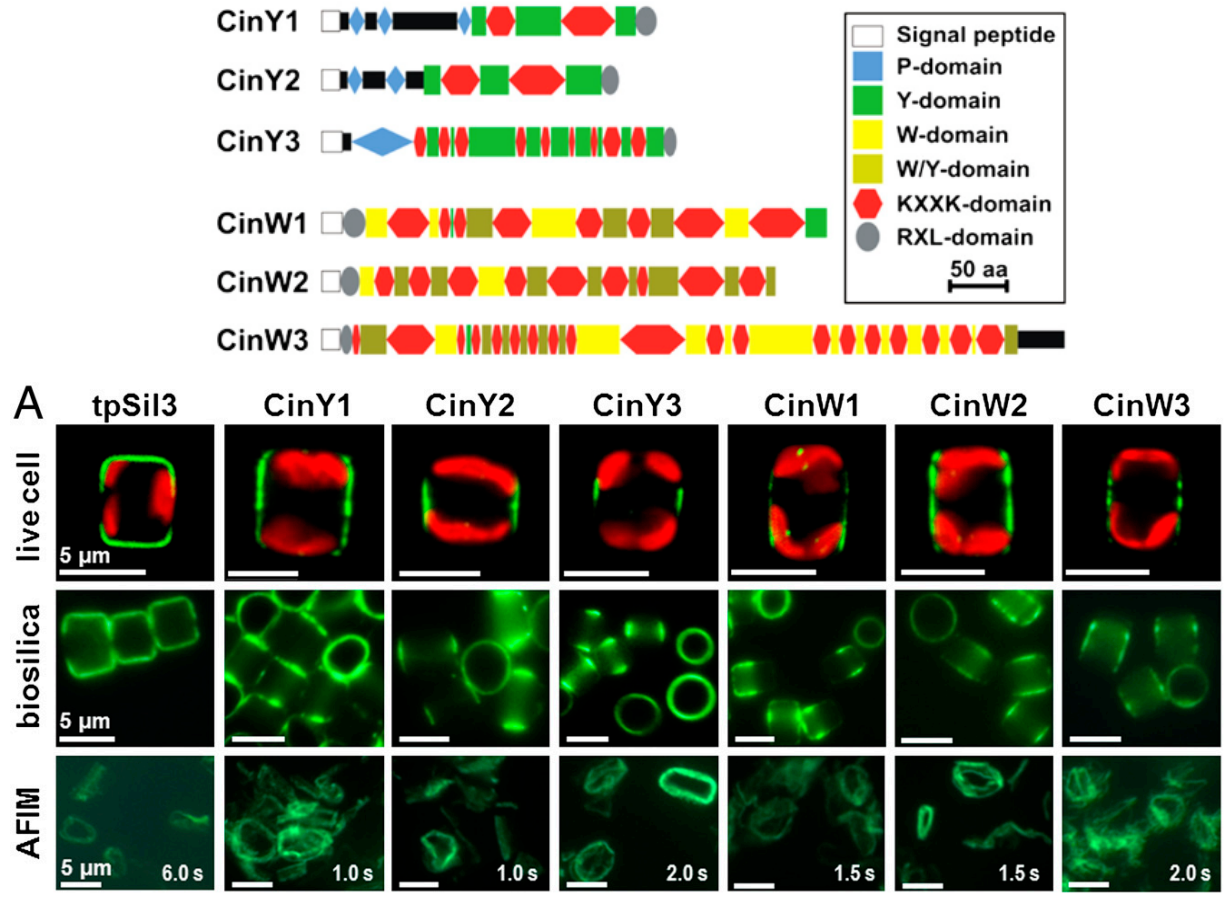


Figure 6 Cingulins. The **upper** panel shows the schematic primary structure of cingulins. The **bottom** panel shows the results of *in vivo* localization experiments of the eGFP-tagged cingulin fusion proteins and of the fluorescent tagging of both the extracted biosilica and the ammonium fluoride insoluble material (AFIM) (from Scheffel et al., 2011).

biosilica of the girdle bands (Hildebrand et al., 2009) and that cingulins are one of the components of an organic layer that functions as template for the synthesis of the girdle bands' non-porous biosilica.

Subsequent work on *T. pseudonana*'s biosilica extracts extended our knowledge of the frustule's chemical composition even further (Kotzsch et al., 2016), as the authors identified a fourth member of the Y-type cingulin group, a valve-associated insoluble organic matrix (nanopatterned microplates), and several other frustule-associated proteins (see following sections). They also found Sil1 to be present in the insoluble organic fraction (small amount of Sil3 proteins were also detected in the insoluble fraction in a previous work; Scheffel et al., 2011).

1.2.5.3 SDV MEMBRANE PROTEINS

Some of the proteins related with the frustule's formation process that have been identified in *T. pseudonana* are silicalemma membrane proteins. For example, the protein silicanin-1

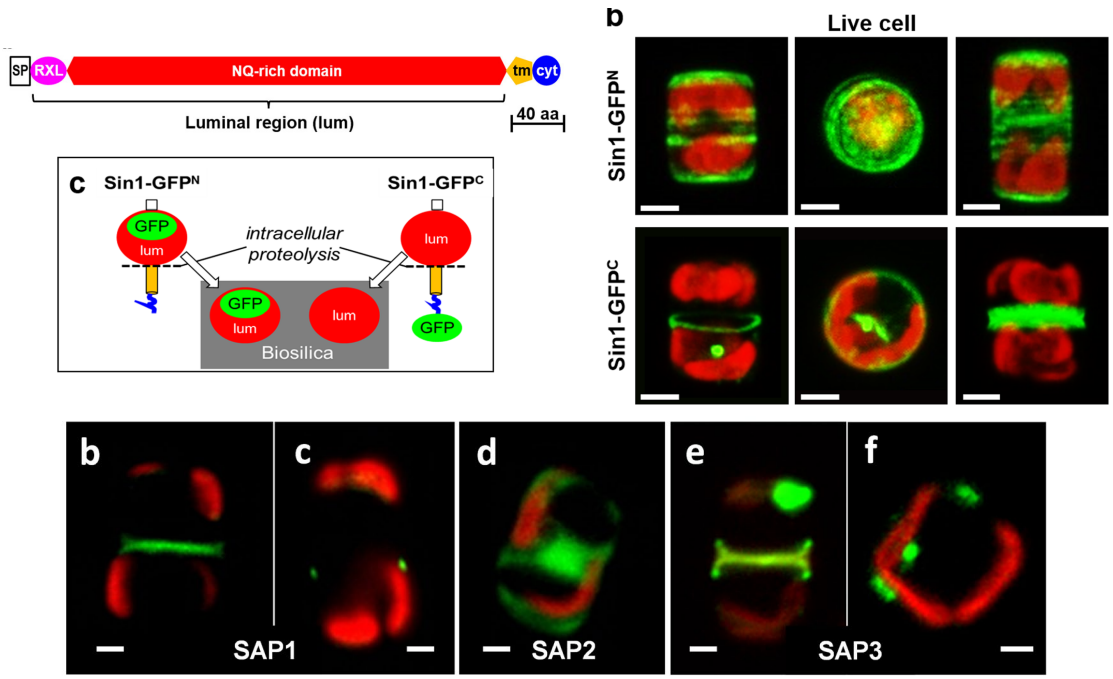


Figure 7 SDV membrane proteins. **Top left** panel shows a schematic representation of the Sin-1 primary sequence and of the hypothesized proteolytic processing in the Sin1-eGFP fusion proteins (Kotzsch et al., 2017); **top right** panel shows *in vivo* localizations of the Sin1-eGFP fusion proteins (upper row: eGFP at the N-ter; lower row: eGFP at the C-ter) (Kotzsch et al., 2017); **bottom panel** shows *in vivo* localization of the SAP proteins (Tesson et al., 2017).

(Sin1) was identified from *T. pseudonana*'s frustule (Kotzsch et al., 2016) and later on (Kotzsch et al., 2017) further characterized as a highly conserved diatom silicalemma protein; a second silicanin protein (Sin2) was found via sequence homology.

Sin1 has a SP immediately followed by a luminal domain; the luminal domain is then followed by a TMD, and after the TMD there is the predicted cytosolic domain. It has been proposed that the faith of Sin1 is split between two populations of protein (Kotzsch et al., 2017): one population gets the TMD and cytosolic domains cleaved off, whereas the other population keeps the transmembrane domain; additionally, the luminal portion of the protein might be involved in the biosilica synthesis machinery, as indirectly suggested by the fact that this protein is highly conserved among many diatom species (Kotzsch et al., 2017). Additional experiments demonstrated an influence of Sin1 on the strength of the frustule (Görlich et al., 2019).

Another class of silicalemma-proteins was identified in the same year (Tesson et al., 2017): one protein (silicalemma-associated protein-1, or SAP1) was identified starting from a dataset containing genes that were upregulated in a similar way as the *sil3* gene during a synchronization experiments (Shrestha et al., 2012); two other proteins, SAP2 and SAP3, were found by sequence homology. Similarly to Sin1, these proteins had a SP followed by a luminal domain, separated from a cytosolic domain by a TMD.

Knockdown experiments on SAP1 and SAP3 were shown to be involved in the morphogenesis of the frustule (Tesson et al., 2017), and when fusing the eGFP to the luminal portion of SAP3, as also demonstrated for Sin1, the fluorescent signal remains trapped in the biosilica (Tesson et al., 2017), indicating strong evidence of an involvement of these proteins in the biosilica formation taking place inside the SDV.

1.2.5.4 SILACIDINS

Further analysis of the soluble fraction of the dissolved frustule from *T. pseudonana* (Wenzl et al., 2008) identified three small similar peptides (named silacidin A, B and C), all encoded by the same silacidin gene (see also Sumper & Brunner, 2008). The three peptides, in fact, derive by the same precursor peptide having repeated domains interspaced by RRL motifs that are absent from the mature peptides; the RXL motifs are often target sites for proteolytical cleavage in frustule-associated proteins (Poulsen et al., 2013; Scheffel et al., 2011). The pro-peptide processing generates 4 silacidin A, one silacidin B and one silacidin C peptides.

Regarding their post-translational modification, silacidins share similarities to other frustule-associated proteins: these proteins, in fact, have at least 60% of serine residues phosphorylated and are therefore extremely acidic proteins (Wenzl et al., 2008). Silica precipitation experiments demonstrated that silacidins, when combined with LCPA, are extremely efficient in catalysing the formation of silica nanoparticles, suggesting that these proteins' role might be to provide the polyanionic component required from the LCPA to direct silica precipitation (Wenzl et al., 2008); additionally, when the environmental availability of Si is decreased (*i.e.,* when *T. pseudonana* cells are Si-starved) an increase in the amount of silacidin protein in the frustule was detected, suggesting that in case the Si is depleted from the environment an increased efficiency in silica processing is required (Richthammer et al., 2011).

1.2.5.5 LONG-CHAIN POLYAMINES (LCPA)

Long-chain polyamines (LCPA), like silaffins, are embedded in the biosilica, and can be extracted only upon dissolution of the frustule (Poulsen & Kröger, 2004; Sumper et al., 2005). These molecules are common in diatoms, although the chain length and degree of modification differ from species to species (Kröger & Poulsen, 2008). LCPA are composed of a propylenediamine, putrescine, or spermidine basis molecule to which linear oligopropyleneimine chains are attached (**Figure 8**).

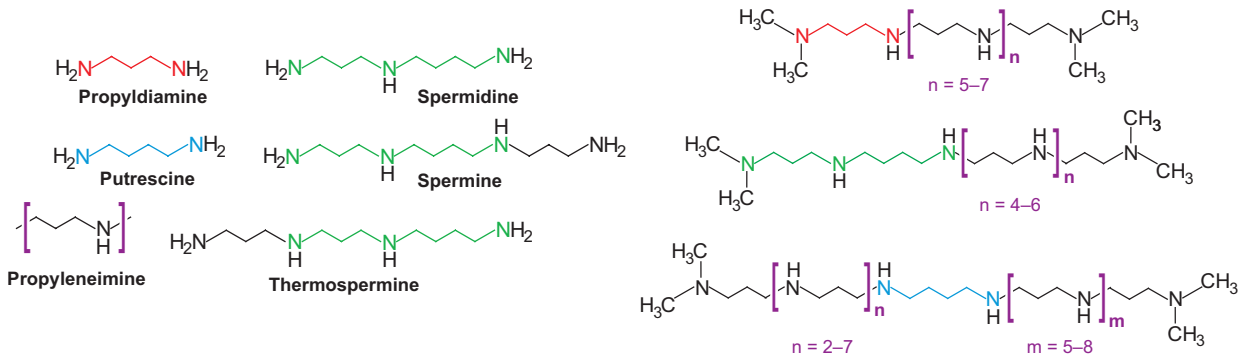


Figure 8 Long-chain polyamines (LCPA). On the **left** the chemical structures of natural polyamines; on the right the chemical structures of LCPA from diatoms (from Kröger and Poulsen, 2008).

In *T. pseudonana*, these molecules interact with the negatively charged silaffins, creating large supramolecular assemblies; this interaction creates a microscopic phase separation that, in presence of dissolved Si, precipitates microparticles of silica. Interestingly, the interaction between LCPA and the three silaffins tested (Sil3, Sil1/2H, and Sil1/2L) indicates different properties of each silaffins in respect to precipitating silica, with Sil3 and Sil1/2H showing regulatory activity at higher concentrations (Poulsen & Kröger, 2004). The existence of species-specific differences in polyamine structures, even among diatoms within the same genus, supports the idea of a specific role of polyamines in creating species-specific silica nanostructures (Sumper et al., 2008).

1.2.5.6 Chitin

Chitin is a very common molecule on our planet; as a matter of fact, it is the world's second most abundant carbohydrate after cellulose (the first most abundant in the sea), and it is present also in some diatom genus (Durkin et al., 2009). In *T. pseudonana* it has a dual function: it is synthesized inside the cell and then extruded in long fibres through the fultoportulae to ensure buoyancy in the water column (Chiriboga et al., 2020). In addition to that, a chitin network-like structure resembling the shape and the size of the silica cell wall could be isolated after dissolution of *T. pseudonana*'s frustule with ammonium fluoride (Brunner et al., 2009); it was proposed that this structure could function as scaffold for the deposition of the silica during the formation of the frustule, or for strengthen and mechanically stabilize the cell wall.

1.3 INTRACELLULAR TRAFFICKING AND PROTEIN TRANSPORT

As eukaryotic organisms, diatoms have a highly complex cellular organization and compartmentalization: they possess several membrane-bound organelles that are specialized for different purposes and cellular needs, creating distinct cell compartments and thus allowing the separation between specific molecular (*e.g.,* the nucleus, the endoplasmic reticulum, or the Golgi Apparatus) or chemical environments (for example, lysosomes or the peroxisomes). Diatoms are also photosynthetic organisms and are therefore equipped with one or more photosynthetic organelles (the chloroplasts) that, thanks to their evolution, are surrounded by four membranes, with the outermost one being continuous with the ER membrane (as described in section 1.2.1). The additional presence of a silica-forming specialized organelle (the SDV), with its own set of biomolecules that need to be transported there, further increases the complexity of diatoms' cell, its compartmentalization, and its transport machinery. Therefore, diatoms need specific targeting routes for soluble or membrane proteins that must be transported to their organelles.

1.3.1 TRANSPORT ROUTES

In a typical eukaryotic cell, a large percentage of the newly synthesized proteins are either destined to some cell organelle (*e.g.,* the nucleus, the ER, the Golgi apparatus, the peroxisome, the chloroplast), integrated into the membranes system (*e.g.,* receptors or transporters), or secreted (*e.g.,* hormones or neurotransmitters). These molecules need a precise and efficient sorting mechanism, allowing them to reach their destination without mistargeting.

Two different mechanisms are known for the sorting of proteins: signal-based targeting and vesicle-based trafficking (Lodish et al., 2016). The second process involves what is also known as secretory pathway: soluble proteins taking this pathway are co-translationally imported into the ER, where they are properly folded and often processed via post-translational modification. The proteins are then sorted out by means of vesicles budding from the ER and then transported to other intermediate or final cell locations (for example, the Golgi apparatus is the intermediate step after the ER for vesicles needed to be transported to the lysosome or to the plasma membrane). Despite some differences exist, protein sorting generally relies on the presence of targeting sequences (*e.g.*, short stretches of amino acids or mannose-6-phosphate) that designate the protein's destination by ensuring the specific cargo-receptor interaction, therefore avoiding mistargeting.

For many proteins taking the secretory pathway, and finally acting as soluble proteins, a first targeting signal is represented by the *signal peptide* (SP), a short stretch of amino acid, usually (but not always) located at the very N-terminal of the nascent protein. This sequence is recognized by the *signal recognition particle* (SRP), that binds to the SP of the nascent protein and to the ribosome; the complex is then translocated to the ER membrane, where it interacts with the *SRP receptor* (an ER integral membrane protein); the protein is then forced through a translocon (the *Sec61* complex) and enters the ER lumen, where a signal peptidase removes the SP. The proteins can now be folded into their native state and be post-translationally modified (*e.g., N*-glycosylation). The proteins are then sorted out to the *cis*-Golgi in transport vesicles that bud from the ER membrane system; this movement is defined as *anterograde*, because it goes from the ER to the Golgi; *retrograde* vesicle trafficking, instead, carries cargo proteins needed in the ER back to this organelle, where they can be recycled (in case of ER-resident proteins) or degraded (in case of misfolded proteins). According to this model, the vesicles carrying the cargo proteins that bud from the ER then fuse with each other, forming the *cis*-Golgi cisterna. Thanks to a process called *cisternal maturation*, retrograde vesicles carry the needed Golgi-resident proteins back from mature cisterna to the early cisterna. During this process, the *cis*-Golgi cisterna moves from the *cis*-position (close to the ER) to the *trans*-position (farther from the ER), by first becoming the *medial*-, and then the *trans*-Golgi cisterna. From the *trans*-Golgi, the proteins are then packed into other budding vesicles that are then transported to their destination (for example, the lysosome or the plasma membrane).

The recognition of the cargo proteins, the formation of the budding vesicles, and their fusion with the target membranes, despite differences between each type of vesicle (*i.e.,* anterograde, retrograde, or those budding from the plasma membrane and the *trans*-Golgi, directed to the endosomes), share some general characteristics (Lodish et al., 2016): proteins belonging to the GTPase superfamily of switch proteins (*e.g.*, *ARF* for COPI vesicles, *Sar1* for COPII vesicles) are responsible for the initiation of the vesicle coat formation and disassembly, for the recruitment of the desired cargo proteins, and for the budding of the vesicles. *Rab* proteins are a second group of GTPase; they substantially represent the “delivery tag” of each vesicle; in fact, they designate the vesicle’s destination by means of interaction with so-called *Rab effectors*, that are generally located on the target membrane (in some cases they directly interact with motor proteins), and they also allow docking of the vesicle to the target membrane.

The close interaction between the vesicle and the target membranes (and their fusion) is ensured by v-type (vesicle) and t-type (target) SNAREs proteins. The v-SNAREs are exposed on the vesicle surface after disassembly of the vesicle coat; after Rab-mediated docking of the vesicle, the v-SNAREs are free to interact with the cognate SNAREs on the target membrane.

In diatoms, one of the most studied protein targeting and trafficking mechanism is certainly the one underlying the transport of nuclear-encoded proteins to the plastid (reviewed in Maier et al., 2015), but also other examples exist, such as the peroxisomal targeting studies in *P. tricornutum* (Gonzalez et al., 2011). Additionally, several marker proteins for specific cell compartments are known (see for example Liu et al., 2016), allowing for co-localization studies. Little is known, instead, about the intracellular trafficking of frustule-associated proteins, or about the SDV-targeting. As mentioned, silaffins' and silacidins' precursor peptides undergo extensive processing (*e.g.*, proteolytic cleavage) and post-translational modification (such as phosphorylation, glycosylation, and sulfation), whereas cingulins possess several KxxK and RXL motifs, which are candidate target sites of lysine modifications and proteolytic cleavage, respectively (Poulsen & Kröger, 2004; Scheffel et al., 2011; Wenzl et al., 2008). All these proteins are derived from precursor polypeptides bearing a SP for co-translational import into the ER, and an ER-resident kinase that phosphorylates silaffins (*in vitro*) has been identified and localized by eGFP-tagging (V. Sheppard et al., 2010); *in vivo* localization experiments of the SP from Sil3, and membrane fractions isolation of Sil3-derived constructs showed it to be present in ER and Golgi membrane fractions (Poulsen et al., 2013). Based on these data, a model has been proposed, in which proteins needed for the biosynthesis of the frustule (silaffins, silacidins, cingulins) are synthesized at the ER, transported to the Golgi Apparatus and from there sorted out via cargo-receptor interactions-mediated sorting and transferred to the SDV in transport vesicles, whereas non-silica forming proteins associated with the frustule (*e.g.,* frustulins) are targeted to the plasma membrane directly from the ER (Poulsen et al., 2013).

Another pathway is known, at least in plants, where vesicles budding from the ER are directly targeted to the vacuole, bypassing the Golgi apparatus (Viotti, 2014); given the similarities between the SDV and the vacuole, it is possible that a similar pathway exists as well. Moreover, a recent study showed that a V-type H^+-ATPase is not only responsible for the acidification of the SDV, and involved in the vesicular trafficking to this organelle,

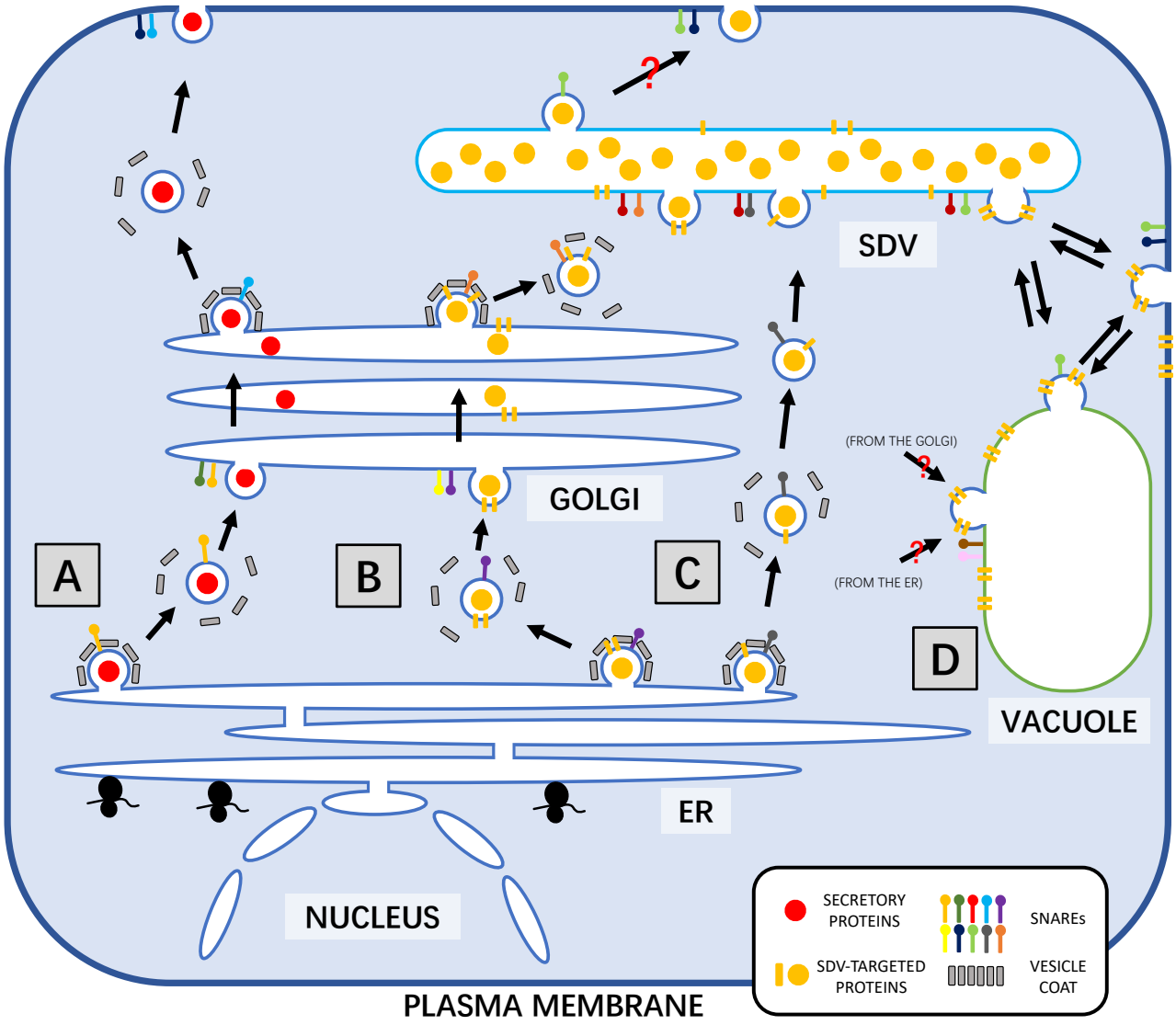


Figure 9 Intracellular trafficking model for *T. pseudonana*. The figure shows four possible pathways for the delivery of cell-wall associated proteins to their final cellular localization (description in the text). For simplicity, all organelles and cell structures are omitted, except for the nucleus, the cytoplasm membrane, the endoplasmic reticulum (ER), the Golgi apparatus, the silica deposition vesicle (SDV) and the vacuole. The inlet at the bottom right of the figure specifies the cargos, vesicle coat proteins and SNAREs (from Fattorini and Maier, 2021).

but also that this protein is dynamically translocated between the silicalemma, the plasma membrane and the tonoplast (*i.e.*, the vacuolar membrane) (Yee et al., 2020).

Using this new data, a new extended model has been hypothesized to describe the protein sorting and transport of silica-forming biomolecules (see **Figure 6**; Fattorini & Maier, 2021): via route **A**, vesicles containing proteins that are not directly involved in the formation of biosilica (for example, frustulins) are targeted from the ER directly to the plasma membrane, and are then exocytosed; via route **B**, **C**, or **D** vesicles containing cargo proteins needed for the formation of the frustule are transported to the SDV. Via route **B**, the cargo proteins are transported from the ER to the Golgi Apparatus, and from there to the SDV via transport vesicles; via route **C**, vesicles budding from the ER are

directly transported to the SDV; via route **D**, vesicles coming from the ER or the Golgi are transported to the vacuole, and from there to the SDV. Anyway, it is also possible that some of the frustule-forming proteins do not transit through the SDV and get embedded in the biosilica via some other pathway.

At the present time, except for the membrane proteins Sin1, VHA, SAP1 and SAP3 (Kotzsch et al., 2017; Tesson et al., 2017; Yee et al., 2020), there is no published data demonstrating an SDV-transit for other proteins, such as the silaffins, cingulins or silacidins. In any case, vesicle-mediated cell sorting and intracellular trafficking to the SDV would require the presence of specific receptors (in the ER, or both in the ER and the Golgi), SNAREs, and Rab proteins; in this respect, the isolation of SDV fractions and the characterization of the related proteomic composition, together with *in vivo* localization and co-localization experiments of selected candidate proteins, would greatly enrich our knowledge on the protein sorting and vesicular targeting to the SDV

1.3.2 TARGETING MOTIFS IN FRUSTULE'S PROTEINS

As mentioned in the previous section, protein sorting requires the presence of internal targeting motifs to ensure the specificity of the cargo protein intracellular transport.

Regarding the targeting sequences in frustule-associated proteins from *T. pseudonana*, the only data available come from a study of 2013 (Poulsen et al., 2013): researchers have created eGFP-tagged truncations of the Sil3 amino acid sequence and found out that some portions of the protein still retained silica-targeting properties, whereas other portions showed no, or very reduced silica-targeting. The analysis of these portions identified short amino acid stretches in which five non-consecutive lysines were interspaced by other amino acids, and it highlighted how the presence of these motifs was associated with efficient frustule-targeting; the researchers identified two of these motifs in Sil3, two in Sil4, one in Sil1 and one in Sil2 (see **Table 1**), and they named them pentalysine clusters (PLC). *In vivo* localization of one PLC isolated from Sil3 (PLC3ct) and one from Sil1 (PLC1) revealed that both constructs retained silica-targeting activity, although with remarkable differences (PLC1 had a much more significant targeting activity respect to PLC3ct). Interestingly, the *in vivo* localization experiment of an artificial PLC (PLCart), where lysines were interspaced only by serine residues, indicated a strong targeting activity. Those portions of Sil3 that did not possess any PLC at all, when localized *in vivo*, showed instead very reduced silica targeting, respect to the portions bearing one PLC.

Therefore, the authors hypothesized that the PLC were involved in the efficient targeting of Sil3, but that also other motifs or factors were responsible for the frustule-targeting of the native Sil3.

Table 1 Amino acid sequence of the pentalysine clusters identified the silaffins (from Poulsen et al., 2013).

Motif name	Motif sequence	protein	position
PLC3nt	KAAKLFKP--KASK	Sil3	62–73
PLC3ct	KAAKIFKG--KSGK	Sil3	189–200
PLC1	KATKTLKVS-KSGK	Sil1	430–442
PLC2	KVSKSGKSS-KSSK	Sil2	420–432
PLC4a	KAAKLFKP--KSSK	Sil4	47–58
PLC4b	KATKVFKSSGKSGK	Sil4	213–226
PLCart	KSSKSSKSS-KSSK	-	-

Silaffins might not be the only proteins having PLC-like motifs as targeting sequences. Cingulins are characterized by the presence of several lysine residues, many of which are contained in clusters of KxxK motifs (see also section 1.2.5.2), suggesting perhaps that also in these proteins these motifs might be involved in the targeting.

1.4 CELL CYCLE-REGULATED EXPRESSION

As mentioned, the synthesis of the frustule is correlated with the progression through the cell cycle (**Figure 10**); consequently, the expression of the genes encoding for proteins or enzymes directly involved in the frustule's biosynthesis must be carefully coordinated. Luckily, a very useful approach has been developed, that allows to enrich the algal cultures for cells undergoing the same biological process (see next section). The interpretation of the available transcriptomic data, together with comparing results from *in vivo* localization experiments, provide additional clues on the process (Fattorini & Maier, 2021; Hildebrand et al., 2018).

1.4.1 CULTURE SYNCHRONIZATION

In an algal culture growing during the exponential phase of the growth curve, and under continuous light, the algal population is composed of cells from different stages of the cell cycle, because their growth is not coordinated. Enriching the culture for synchronous-growing cells would provide a higher percentage of cells undergoing a certain biologic

event and, if a certain biological process can be precisely associated with a specific phase of the cell cycle, synchronous growth is then a powerful tool for the study of that process. The synchronization protocol for *T. pseudonana* was developed in 2007 (Hildebrand et al., 2007) and it consists in a silicon depletion-based synchronization of the cell growth (Hildebrand et al., 2007): cells are first grown in normal (Si-repleted) medium, and then transferred in Si-free medium; the absence of Si triggers a stop in the growth of the cells,

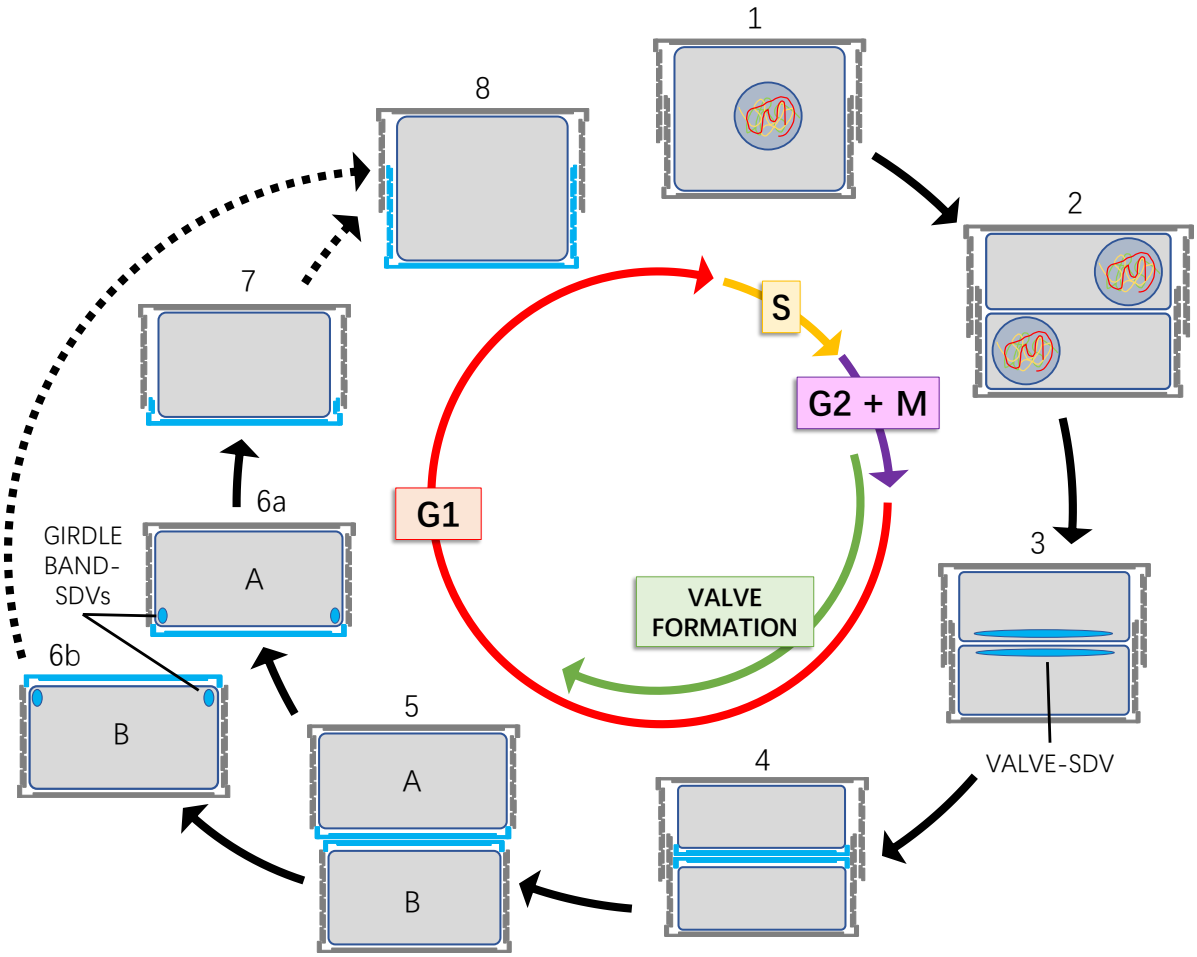


Figure 10 Schematic representation of *T. pseudonana*'s cell cycle and frustule formation. The relationship between valve formation and the cell cycle phases are taken from Schmid, 1994. The steps are: (**1**) the mother cell starts a new cell cycle. (**2**) Following DNA replication (S phase) the cell enters the G2+M phase; after cytokinesis (end of M phase) the two daughter cells' protoplasts have formed but are still retained inside the parental frustule. (**3**) In each protoplast, a valve-SDV is formed and the synthesis of the building blocks for valve generation starts within it. (**4**) When the process is completed, the content of the SDVs is exocytosed. (**5**) The daughter cells are not encased in the parental frustule any longer and they separate; cell A inherits the parental epitheca and it is therefore larger than cell B, which instead inherits the hypotheca. (**6a**, **b**) Analogously to what happens for valve synthesis, the building blocks for girdle bands formation are synthesized inside girdle band-SDV. (**7**) When the process is completed, the content of the girdle band-SDV is exocytosed; (**8**) by stepwise synthesis of new girdle bands (only the formation of the first girdle band is depicted), the cells increase the girdle length and their cell volume. By a yet unknown mechanism, the smaller cell restores the original size, and the average size of the population does not decrease with the mitotic divisions. When the cells have reached the maximum number of girdle bands they can enter a new round of the cell cycle. For simplicity, no organelle is shown except for the plasma membrane, the nucleus (in 1 and 2), and the SDV. The newly synthesized biosilica is in cyan; the parental frustule is in grey; the cytoplasm membrane, the nuclear membrane, and the SDV membrane are depicted in dark blue. The relative lengths of the cell cycle phases are shown in the center.

which arrests in the late G1 phase. Upon Si re-addition the cell cycle is resumed and a variable percentage of cells (see also Heintze et al., 2020) concludes the G1 phase and proceeds synchronously through the cell cycle. This approach can be particularly useful for microscopic observations, for the enrichment of cell-cycle specific structures in the culture (*e.g.*, valve- or girdle bands-forming SDVs), for organelle isolation, for the study of metabolic diel transitions, or for gene expression studies, just to name a few examples (Ashworth et al., 2013; Brembu et al., 2017; Görlich et al., 2019; Heintze et al., 2020; Hildebrand et al., 2007, 2009; Mock et al., 2008; Poulsen et al., 2013; Roshan Prakash Shrestha et al., 2012; Yee et al., 2020).

1.4.2 Expression of genes encoding cell wall-associated proteins

The expression of genes encoding for proteins involved in frustule synthesis might be specifically cell cycle-regulated; if so, expression patterns should be expected to be associated with the cell cycle phases in a synchronized culture. The synchronous growth protocol has been exploited for investigating transcript level changes (Brembu et al., 2017; Mock et al., 2008; Shrestha et al., 2012) and, following the data available from Shrestha et al., 2012, interpretations on the relationship between the expression of genes (plotted as fold changes in transcript level versus hours after Si re-addition) and the localization of the encoded proteins have been proposed (Fattorini & Maier, 2021; Hildebrand et al., 2018). These interpretations assume that the gene expression could be reflected in the protein levels as well.

Considering the cingulins, despite all of them have a girdle band localization (Scheffel et al., 2011), the gene expression pattern of the Y-type genes is different respect to that of the W-type ones (**Figure 11.B**): the Y-type show a peak in the expression at the end of the G1 phase (between 2–4 hours, when the cell is supposed to synthesize the last girdle bands (Hildebrand et al., 2007)), whereas the W-type cingulins show no (*cinW1, cinW3*) or very little (*cinW2*, between 7–8 hour) cell cycle-associated regulation.

In respect to the transcriptional regulation of the silaffin genes (**Figure 11.A**), a similar picture appears: *sil4* shows only a very slight transcriptional level change in the different cell cycle phases (Hildebrand et al., 2018), whereas *sil3* had a maximum in the phase of valve generation together with a minor peak in the time frame of girdle band synthesis (Frigeri et al., 2006; Hildebrand et al., 2018; Shrestha et al., 2012); *sil1* showed a maximum during the first 2–4 h after silicon replenishment, followed by downregulation (Frigeri et al., 2006).

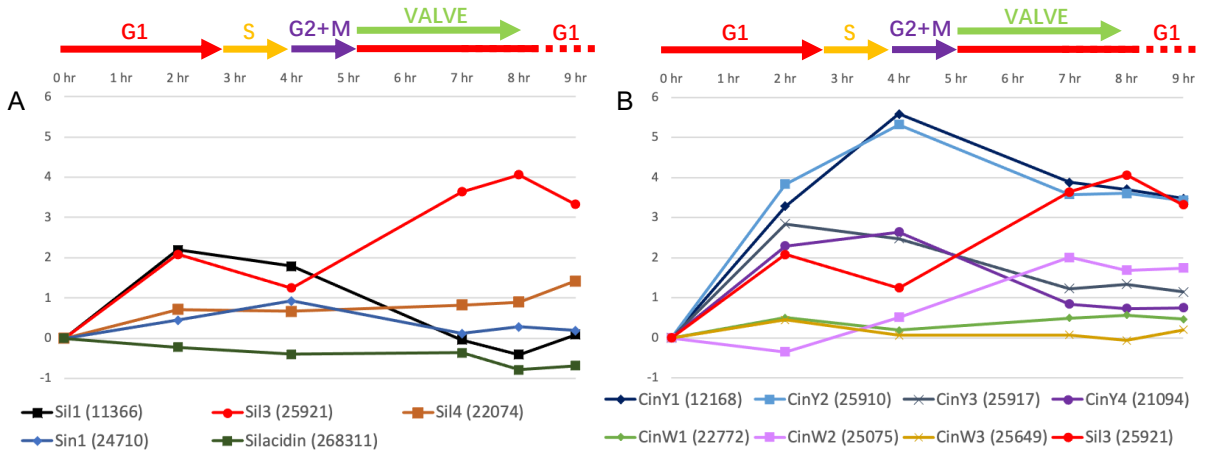


Figure 11 Expression patterns of selected genes encoding for cell wall-related proteins. The raw data regarding the expression pattern of shown genes (the JGI ID of each gene is shown next to the name) are taken from a transcriptomics experiment with synchronized cells (Shrestha et al., 2012). Plot **A** shows the expression pattern of the *sil1*, *sil3*, *sil4, silicanin-1* and *silacidin* genes; plot **B** shows the expression pattern of the W-type and Y-type *cingulin* genes. On the x-axis are reported the hours after Si-replenishment, and on the y-axis is indicated the log2 fold changes of the mRNA expression at indicated time points after Si has been added (t = 0 h). The expression pattern of the Sil3 gene is shown in both plots to approximately indicate the time of valve synthesis; additionally, the cell cycle phases are reported as in **Figure 10**, and shown above each plot.

1.4.3 Role of the promoters in the targeting

By analyzing the expression data on both cingulins and silaffins, some observations can be made: for example, all the studied cingulins (with the exception of CinY4; Kotzsch et al., 2016) were localized *in vivo* using the NR promoter and terminator to regulate the expression of their genes (Scheffel et al., 2011); this promoter, under the experimental conditions used (*i.e.,* with NO_3^- as source of nitrogen) is constitutively expressed. So, despite being expressed in phases in which a valve-SDV only or even no SDV is present, the eGFP-tagged cingulin proteins localized exclusively in the girdle bands. Additionally, the *in vivo* localization of the proteins CinW2 and CinY2 using both their native and the NR promoters and terminators was investigated: for CinW2 no substantial difference was observable between the two conditions; for CinY2, when the gene was expressed using the native promoter and terminator the signal was located in the pleural bands only, whereas when using the NR promoter it was located in all parts of the girdle (Kotzsch et al., 2016).

For the silaffins, a similar picture emerges when comparing the *in vivo* localization of Sil3 obtained when the expression of the *sil3* gene was regulated by the native promoter (Scheffel et al., 2011) or by a constitutively expressed one (Poulsen et al., 2007, 2013). In the former case, the signal was localized in all areas of the frustule except for the pleural bands (see **Figure 6**); in the latter case, instead, all the frustule was labelled, also the

pleural bands (see **Figure 5**). These results suggest that the timing of the expression represent an additional level of regulation, influencing the targeting of some of the studied proteins. Therefore, it appears extremely important the use of the native promoter-gene-terminator cassette when studying recombinant genes for the *in vivo* localization of proteins, or at least to always compare the localization obtained using constitutively active promoters with the one obtained with the native expression.

1.5 BIOTECHNOLOGICAL APPLICATION OF DIATOMS

As mentioned in section 1.2.1, diatom's ecological relevance and their biogeochemical importance on a global scale are undoubtful. Not surprisingly, these organisms have also proven to be extremely reliable source for bio-derived compounds and applications. Diatoms, in fact, have certain characteristics that make them ideal organisms for biotechnological strategies: they are relatively easy to cultivate, as they require only nutrients, CO_2, water, and light to grow; they quickly grow exponentially, by dividing roughly once per day; additionally, they can synthesize bioproducts by using the atmospheric CO_2, therefore representing a renewable source of valuables (for a review, see Hempel & Maier, 2016).

To date, several diatom genomes have been sequenced (Armbrust et al., 2004; Basu et al., 2017; Bowler et al., 2008; Lommer et al., 2012; Oliver et al., 2021; Osuna-Cruz et al., 2020; T. Tanaka et al., 2015; Traller et al., 2016), and despite they represent only a small sampling of the total diatom biodiversity, the data obtained by the genome sequencing of these few species have provided invaluable insights into diatoms' adaptations, evolution, and metabolism. This kind of resource significantly boosted diatom-based biotechnological research, already accounting for a number of effective or potential applications (reviewed in Sharma et al., 2021). Some examples are given by the production of bioplastic (Hempel, Bozarth, et al., 2011) and the synthesis of recombinant proteins (Hempel, Lau, et al., 2011).

The applications derived from the engineering of the frustule are of particular interest for material scientists, since what biomineralizing organisms such as diatoms do is a sort of dream for them: they can produce inorganic-organic hybrid structures in very little time, with a complexity level down to the nanometre scale, at room (or even lower) temperature, and with high-fidelity results. The advent of strategies like the LiDSI (*li*ve *d*iatom *s*ilica *i*mmobilization; Kröger & Brunner, 2014) allowed the genetic engineering of diatom biosilica, for example to create frustules coated with functional enzymes, where the

specific valve- or girdle band-localization of the inserted enzyme can be chosen based on the fusion protein created (Kumari et al., 2020; Poulsen et al., 2007; V. C. Sheppard et al., 2012); this approach has the advantage of ensuring that enzymes are less prone to degradation by heat or solvents, or by proteolytic enzymes. Moreover, diatom frustules can be used as templates to create inorganic replicas made of several materials (graphene, gold, silver, palladium), so that the materials' chemical and optical properties can be exploited synergically with the frustule-derived nanopattern of decorations (see for example Pan et al., 2014; reviewed briefly in Kröger & Brunner, 2014). Other potential applications have been proposed, such as: biosensors or microelectromechanical systems (see Kröger & Poulsen, 2008), or as drug delivery systems (reviewed in Delasoie & Zobi, 2019).

2. AIM OF THE PROJECT

The aim of my project was to investigate the targeting of proteins in the diatom *Thalassiosira pseudonana*. Given the peculiar morphology of diatom cell wall and its complex morphogenesis, we expect that a finely regulated mechanism of protein sorting and transport exist in diatoms and in *Thalassiosira pseudonana*.

The available literature suggests that an intertwined relation between signal sequence-based targeting, genomic transcriptional regulation, and post-translational processing of the proteins of interest is responsible for the trafficking of frustule-associated proteins to the cell wall. On the other hand, it is still unknown whether a specific valve-or girdle band-targeting signal exist. So, the objectives of my project relate to the following questions: can we identify specific targeting signal also in cingulins? Can these signals be specific for the girdle, whereas other for the valve? In case they do, can these signals be used to re-direct a valve- protein to the girdle, or vice vera?

To answer these questions, I have decided to study the role of PLC-like regions in cingulins and to test if their functioning is analogous to that demonstrated in Sil3. Moreover, I intended to demonstrate that the presence of PLC-like regions (or other potential targeting motifs) could be used to identify novel frustule-associated proteins. In parallel to that, I intended to experimentally test if the transcriptional regulation of the gene expression could result in an influence on the frustule-targeting, and in case to which extent. As a subproject, I intended to isolate the promoters of the *sil3*, *cinW2*, and *cinY2* genes.

My study will help understanding the mechanisms through which diatom cell-wall proteins are targeted to specific sub-compartments of the frustule (*i.e.,* the valve or the girdle bands)

3. RESULTS

3.1 TRANSFORMATION VECTORS USED

As a result of the cloning strategies used, I have worked with two "families" of plasmids: the ones deriving from the pTpNAT-MCS parental vector and the ones deriving from the pTpINV-NAT parental vector (which is itself a derivation of the first; see also section 7.4). The difference between the two groups is given by the position of the NAT cassette respect to the MCS; in the first group the NAT cassette is located immediately downstream respect to the MCS, whereas in the second group the NAT cassette is located immediately upstream respect to the MCS. The extension of the various fragment composing the vectors remains unaltered between the two groups. In the Supplementary section 7.4 are shown both the plasmid maps and the features of the various vectors.

3.2 LYSINES IN CINW2 AND CINY2

The analysis of the amino acid composition of CinW2, CinY2 and Sil3 indicates that lysines are the third more abundant amino acid in the three proteins (see **Table 2**). In CinW2, CinW1, and CinW3 all lysine residues except for the second amino acid of the SP are contained into motifs where lysines are interspaced by other amino acids; these short motifs (KxK, KxxK, KxxxK, or KxxxxK, where X = G or S), which for simplicity will be referred to as KxxK motifs, are often clustered in PLC-like regions. The structure of this KxxK-rich regions from CinY2 and CinW2 is similar to the one of PLC, but it's not identical; additionally, when the PLC were identified, it was not investigated whether the presence of five lysines was a necessary feature for the silica targeting or not (Poulsen et al., 2013). Because of these reasons, I have selected a few regions of the CinW2 and CinY2 peptide sequence having KxxK-motifs of at least four lysines and I have referred to them as lysine-enriched regions (LERs). The sequence of the corresponding motifs in the studied proteins are shown in the protein sequences of section 7.2, and in **Figure 13**, **Figure 16** and **Figure 18**.

3.3 TARGETING STUDIES BY *IN VIVO* LOCALIZATIONS

3.3.1 SIL3, CINW2 AND CINY2

In the parental transformation vector that I have used for the cloning of my constructs, the expression of the *nat* gene is regulated by the *fcp* promoter and terminator regions. Under the experimental conditions applied, the NAT cassette ensured a sufficient screening

capability against those clones which had integrated the vector bearing this cassette; this indicated that the choice of reducing the extension of these regulatory regions did not affect the efficacy of the antibiotic resistance cassette. Furthermore, the extension of the genes' native promoter and terminator regions has been reduced also in the three vectors pTpSil3, pTpW2 and pTpY2 (to approximately 800 bp for the promoter and 300 bp for the terminator).

Table 2 Amino acid composition of the proteins Sil3, CinW2 and CinY2

AMINO ACID COMPOSITION OF CinW2, CinY2 AND Sil3									
CinW2				CinY2			Sil3		
Amino acid		Counts	Percent.	Amino acid		Counts	Percent.	Amino acid		Counts	Percent.
Ser	S	126	32.9%	Gly	G	58	23.4%	Ala	A	40	17.3%
Gly	G	87	22.7%	Ser	S	37	14.9%	Ser	S	40	17.3%
Lys	**K**	**41**	**10.7%**	**Lys**	**K**	**22**	**8.9%**	**Lys**	**K**	**34**	**14.7%**
Asp	D	34	8.9%	Pro	P	19	7.7%	Glu	E	19	8.2%
Trp	W	19	5.0%	Thr	T	19	7.7%	Met	M	19	8.2%
Glu	E	13	3.4%	Tyr	Y	19	7.7%	Gly	G	13	5.6%
Tyr	Y	13	3.4%	Asp	D	11	4.4%	Asp	D	12	5.2%
His	H	8	2.1%	Leu	L	8	3.2%	Leu	L	9	3.9%
Val	V	7	1.8%	Ala	A	7	2.8%	Thr	T	9	3.9%
Ala	A	6	1.6%	Arg	R	7	2.8%	Val	V	8	3.5%
Thr	T	6	1.6%	Asn	N	7	2.8%	His	H	7	3.0%
Leu	L	5	1.3%	Ile	I	7	2.8%	Pro	P	7	3.0%
Asn	N	4	1.0%	Glu	E	6	2.4%	Ile	I	5	2.2%
Pro	P	4	1.0%	Val	V	6	2.4%	Arg	R	3	1.3%
Gln	Q	3	0.8%	Met	M	4	1.6%	Phe	F	3	1.3%
Ile	I	3	0.8%	Gln	Q	3	1.2%	Gln	Q	2	0.9%
Arg	R	2	0.5%	His	H	3	1.2%	Cys	C	1	0.4%
Met	M	1	0.3%	Phe	F	3	1.2%	Asn	N	0	0.0%
Phe	F	1	0.3%	Cys	C	2	0.8%	Trp	W	0	0.0%
Cys	C	0	0.0%	Trp	W	0	0.0%	Tyr	Y	0	0.0%

To confirm that this reduction did not affect the expression of the corresponding DNA constructs, and to have a native localization to use for the comparison during my experiments, the *in vivo* location of the three native proteins (*i.e.*, without any sequence modification) was confirmed: the Sil3-eGFP fusion protein was localized in the valve (except the fultoportulae) and in the first girdle bands, but no labelling was localized in the pleural bands (**Figure 12**.**A1-3**). The CinW2-eGFP fusion protein was localized exclusively in the girdle, from the valvocopula to the central area of the girdle (where the two thecas overlap), labelling several girdle bands with different apparent intensity, which were differentially labelled because of the morphology of *Thalassiosira pseudonana*'s cell wall (**Figure 12.B1-3**). The fusion protein CinY2-eGFP was also localized merely in the girdle area, although the fluorescent signal labelled only the area of the pleural bands (**Figure 12.C1-3**). In some image, it might appear as if this localization is not central respect to the

frustule and it is instead shifted towards one pole of the cell; this can be explained by keeping in mind that the two opposite valves of a diatom cell move away from each other during growth, therefore when a diatom in an early stage of growth is imaged, the newly formed hypotheca is still shorter than the epitheca. As expected, for all the three proteins, the localization obtained was identical to the previously published localization (Scheffel et al., 2011; Kotzsch et al., 2016).

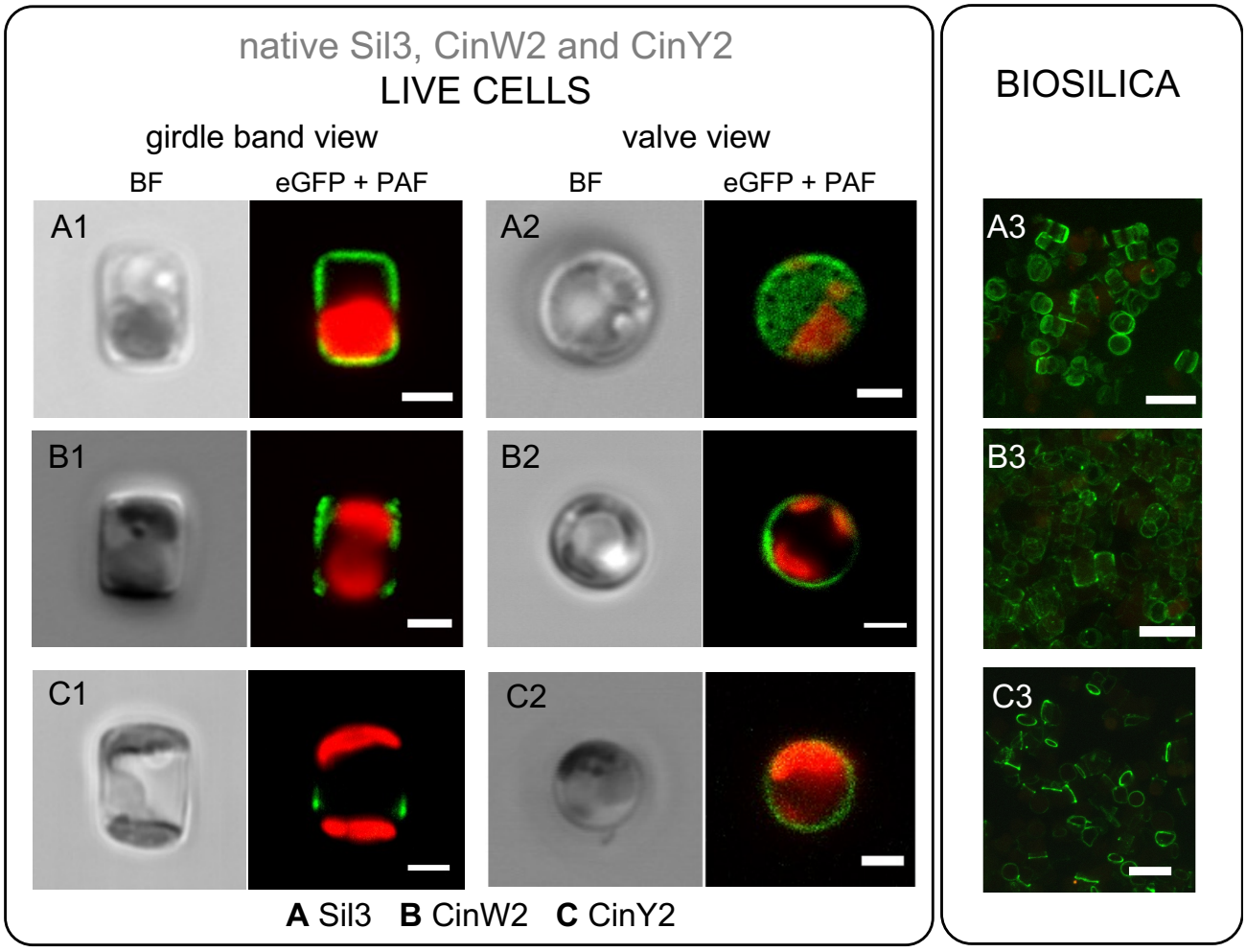


Figure 12 *In vivo* localization of the native Sil3, CinW2, and CinY2 as eGFP-fusion proteins. The figure shows both bright field images (BF) together with merged green (eGFP) and red (plastid autofluorescence, PAF) channels, of representative *T. pseudonana* cells. A1, B1, and C1 show transapical sections of the cells in girdle band view; A2, B2, and C2 show transversal sections in valve view (A2 shows a valve surface view image); A3, B3 and C3 show instead fluorescent images of extracted biosilica. Scale bars are 2 µm for the *in vivo* localization images and 10 µm for the biosilica images.

3.3.2 CinW2-derived constructs

To identify putative targeting motifs in CinW2, I have studied the *in vivo* localization by eGFP tagging of constructs derived from the *cinW2* gene. These constructs encoded modified versions (truncations, deletions, mutations) of the CinW2 protein (see **Figure 13**

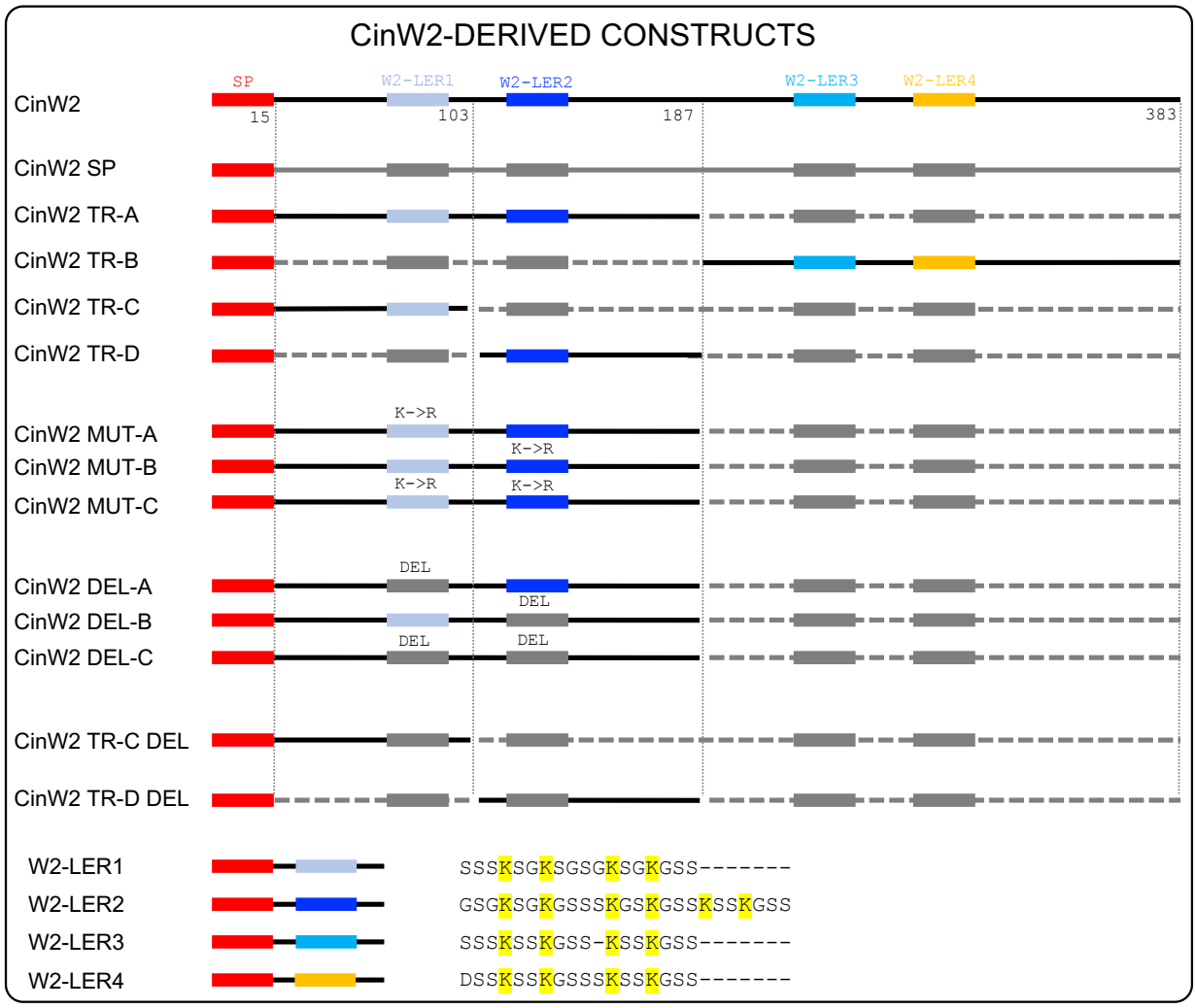


Figure 13 CinW2-derived constructs, graphical depiction. The first construct is the full-length CinW2 protein; the color code is indicated on top of each feature; the number of the amino acid residue corresponding to the end of a certain region (and to the end of the peptide) is indicated below the graphic. The regions depicted in grey are the deleted ones. For the isolated LERs the amino acid sequence is shown. Size of the constructs not in scale.

for a graphical summary). All the CinW2-derived constructs have been generated using the plasmid pTpCinW2 as parental plasmid.

3.3.2.1 CinW2 Sequence truncations

The construct CinW2 SP encodes the first 15 amino acid residues, that are predicted to be the SP; this construct was localized in an intracellular sub-compartment (**Figure 14.A**), possibly associated with the plastid.

The truncation CinW2 TR-A included the predicted SP (amino acids from 1 to 15) and the amino acids from 16 to 187, representing therefore the 172 aa-long N-terminal half of the mature CinW2. The localization of the corresponding eGFP-fusion protein fluorescent signal (**Figure 14.B**) resulted very similar to that obtained by the full-length CinW2-eGFP fusion protein, although with one main difference respect to the latter

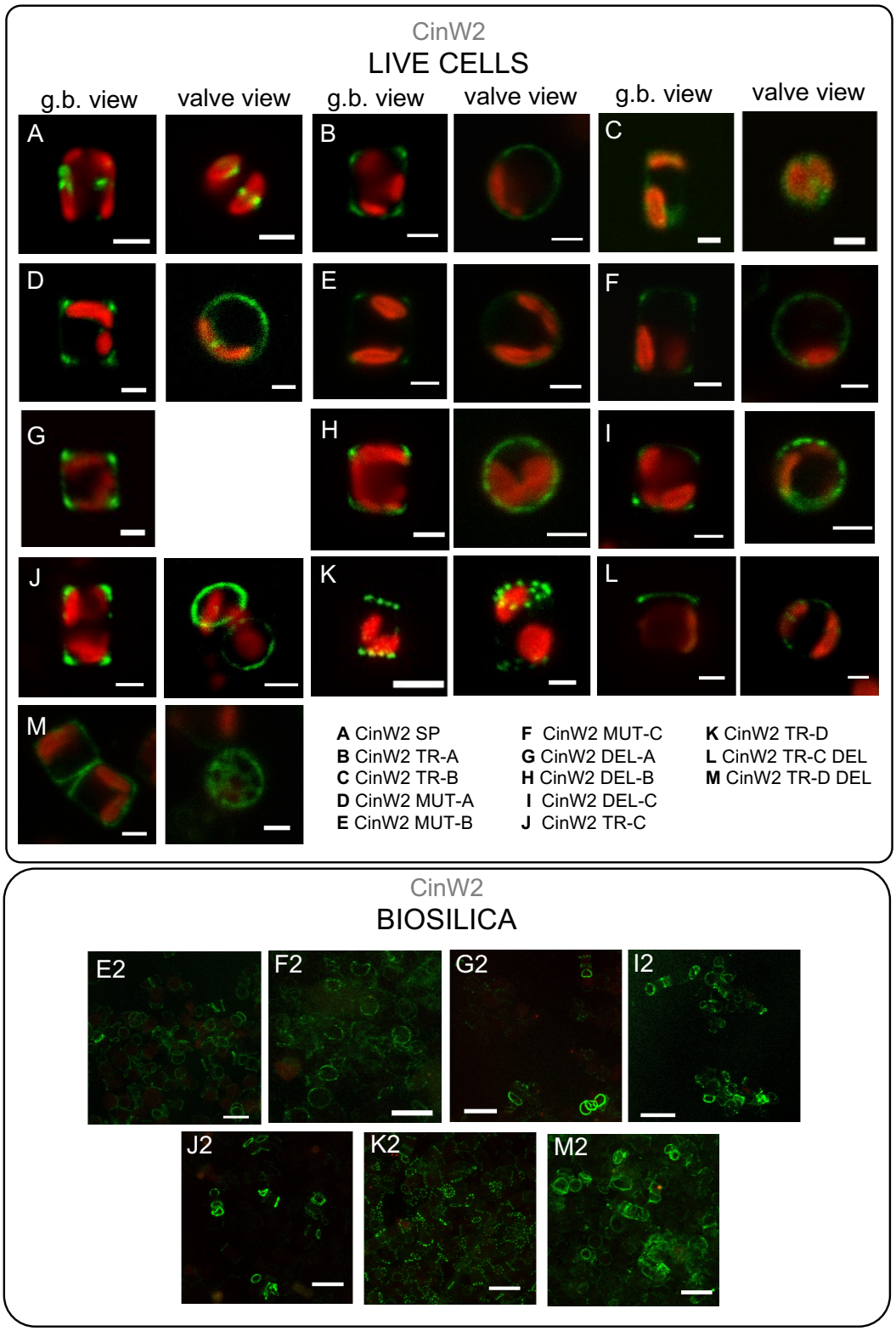


Figure 14 *In vivo* localizations of CinW2-derived constructs. Confocal images of live cells (representative for each construct) from *in vivo* localization experiments using the truncated, mutated, or deleted CinW2-derived constructs. The upper panel shows the merge of the eGFP (green) and PAF (red) channels. For each construct, the first column shows transapical sections of the cells (in girdle band view), whereas the second column shows transversal sections of the cells (in valve view). For the constructs in G, J and K there is no valve view image, but for J and K maximum projections of tilted cells are shown, highlighting the fluorescent signal from the valve; for the construct in M, the second image shows the valve surface and not the transversal section. The bottom panel shows confocal images prepared merging the eGFP (green) and the PAF (red) channels of images of extracted biosilica from some of the clones shown in the upper panel. Scale bars are 2 µm for the *in vivo* images and 10 µm for the biosilica images.

location: the signal was mostly coming from the region of the frustule comprised between the valve and the first girdle bands (corresponding to the rim of the valve and the valvocopula, the proximal girdle band). The full-length CinW2 protein (**Figure 12**.**B1**) is instead located not only in the abovementioned area but also all over the girdle; this latter location appears to be completely abolished in the CinW2 TR-A construct. The CinW2 TR-B truncation included the predicted SP (amino acids from 1 to 15) and the amino acids from 188 to 383, representing therefore the 196 aa-long C-terminal half of the mature CinW2. The resulting *in vivo* localization of the construct (**Figure 14.C**) showed no frustule specificity; instead, a weak signal appears to be in the cytosol, suggesting that perhaps this truncation cannot target the eGFP to the cell wall and that it has most likely lost the specific targeting sequence.

3.3.2.2 CinW2 MUTATIONS AND DELETIONS

According to these first preliminary results, I hypothesized that some targeting motif was present in the portion of the sequence comprised between amino acids 16 to 187. The CinW2 protein contains multiple LERs (see **Figure 13** and section 7.3.1) and two of them, W2-LER1 and W2-LER2, are included in the CinW2 TR-A construct. As mentioned in the introduction, the LERs have certain striking structural similarity to the PLCs, which are involved in the efficient targeting of the Sil3 protein; I have therefore decided to test whether W2-LER1 and W2-LER2 has some function in the targeting mechanism of CinW2 TR-A.

To do so, I conducted two distinct experiments: in the first experiment I have mutated all the lysine residues in the motifs into arginine residues (this approach was previously used to demonstrate that arginine cannot functionally substitute lysine to achieve efficient silica targeting in a Sil3-derived cell wall-targeting fragment; Poulsen et al., 2013). Three new CinW2 TR-A-deriving constructs were generated, in which W2-LER1 (construct CinW2 MUT-A), W2-LER2 (CinW2 MUT-B), or both LERs (CinW2 MUT-C) were mutated (see **Figure 13**). As shown in **Figure 14.D-F**, none of the three constructs resulted in a mistargeted location; on the contrary, the final cell wall location appears to be identical between them, and to the CinW2 TR-A construct.

In the second experiment, I generated three new constructs, where W2-LER1 was deleted (CinW2 DEL-A), W2-LER2 was deleted (CinW2 DEL-B), or both were deleted (CinW2 DEL-C). Also in this case (**Figure 14**.**G-I**), the resulting localizations showed no loss of specific targeting in the corresponding constructs, with the eGFP signal coming

from the very same localization as for the CinW2 TR-A and the mutated constructs. For the constructs CinW2 MUT-B, CinW2 MUT-C, CinW2 DEL-A and CinW2 DEL-C, the biosilica extracts confirmed that the corresponding eGFP-tagged fusion proteins were embedded in the silica (**Figure 14**, images **E2, F2, G2, I2**).

I made two additional sets of truncations of the CinW2 protein: the first set directly derived from construct CinW2 TR-A and generated the construct CinW2 TR-C, comprising amino acids from 1 to 103 (therefore also including W2-LER1), and the construct CinW2 TR-D comprising the SP plus amino acids from 104 to 187 (therefore including W2-LER2). The localization of the construct CinW2 TR-C (**Figure 14.J)** was substantially identical to the localization of CinW2 TR-A and its mutated or deleted derived constructs: the fluorescent labelling was located only in the valvocopula, highlighting therefore two ring-like regions at the rims of the valves, without any fluorescence coming from the remaining areas of the valves or the girdle. The localization of the CinW2 TR-D, instead, resulted in a completely different labelling (**Figure 14.K**): the only area of the frustule showing fluorescent tagging was the valve-located circular pattern resembling the organization of the fultoportulae.

The second set of truncations was created from CinW2 DEL-C and the truncation position was the same as the previous one (between amino acid residues 103-104). Since the two constructs generated (CinW2 TR-C DEL and CinW2 TR-D DEL) were derived from CinW2 DEL-C (in which W2-LER1 and W2-LER2 had been deleted), they were lacking LER1 and LER2, respectively. Construct CinW2 TR-C DEL (**Figure 14.L**) showed the same valvocopula-located labelling as previously observed for other constructs, albeit the signal was not only very faint and not homogenously distributed in that area, but it was also apparently absent from the opposite pole of the cell; additionally, intracellular sub-compartments showed minor amounts of fluorescence. Construct CinW2 TR-D DEL (**Figure 14.M**) was instead localized mostly in the valve's area, but a minor fraction of the fluorescent signal was also present in the girdle and inside the cell; interestingly, the fultoportulae appear to be excluded from the valve labelling.

3.3.2.3 CinW2 isolated LERs

To test if the isolated LERs from CinW2 can target the eGFP to the frustule, four new constructs were generated, containing the CinW2 SP and then each individual LER; the LERs have been isolated together with the three amino acids flanking the first and the last lysine (see **Figure 13**).

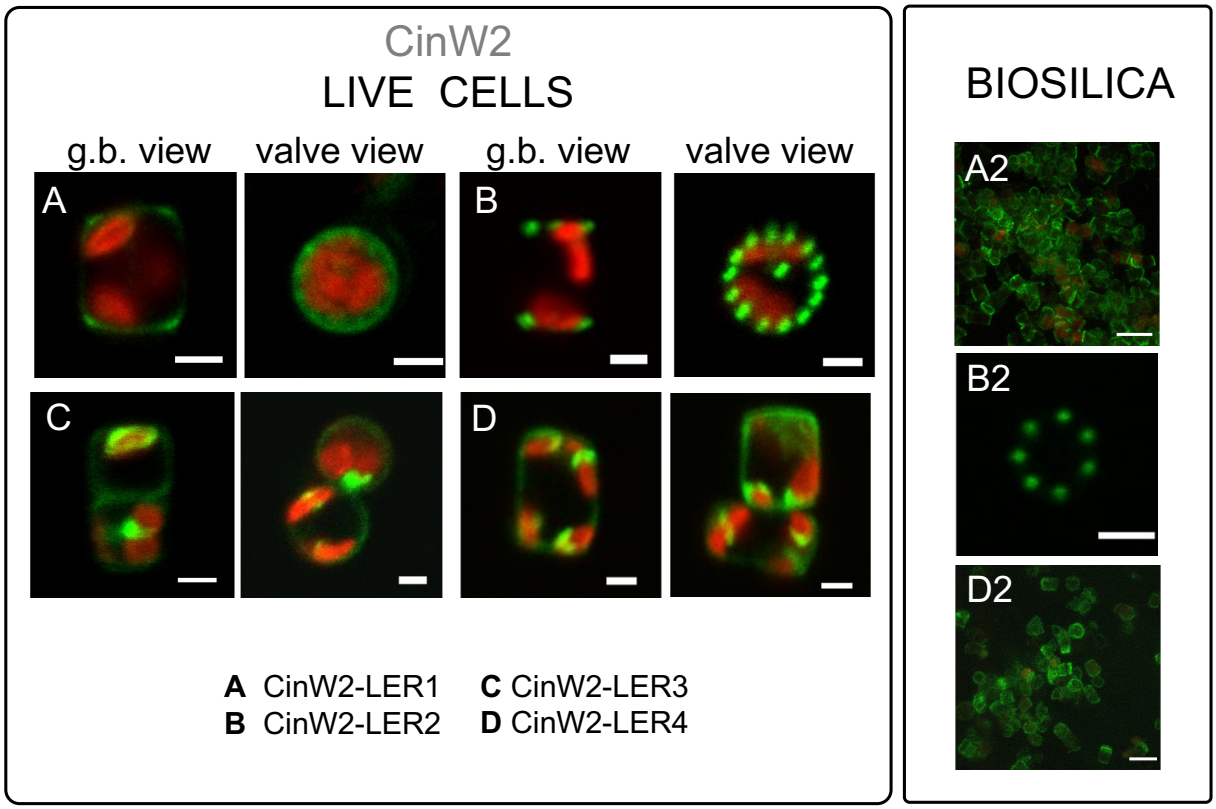


Figure 15 *In vivo* localizations of the CinW2-LERs constructs. Confocal images of live cells (representative of each construct) from *in vivo* localization experiments of the isolated CinW2-derived LERs. The panel on the left shows the merge of the eGFP (green) and PAF (red) channels. For each construct (except for the construct in D), the first column shows transapical sections of the cells (in girdle band view), whereas the second column shows transversal sections of the cells (in valve view); the second image from A shows the surface of the valve. The panel on the right shows confocal images prepared merging the eGFP (green) and the PAF (red) channels of images of extracted biosilica from some of the clones shown in the panel on the left. Scale bars are 2 µm for the *in vivo* images (and image B2) and 10 µm for the biosilica images.

Construct W2-LER1 (**Figure 15.A**) gave unclear results because despite the labelling was certainly pronounced in the valve (especially in the fultoportulae and in the rim of the valve) some cells show fluorescent labelling also in the girdle, as confirmed by the fluorescent labelling in the extracted biosilica (**Figure 15.A2**). The localization of the W2-LER2 construct (**Figure 15.B**), resulted instead in the eGFP signal tagging an area of the valve that was clearly resembling the shape of the fultoportulae pattern that decorates the valve, without labelling at all any part of the girdle; the labelling of the extracted biosilica confirmed this localization (**Figure 15.B2**). The localization of construct W2-LER3 (**Figure 15.C**) resulted in a faint but apparently homogenous frustule labelling, but conspicuous amounts of eGFP signal could be detected in some intracellular compartment closely associated with the plastid (possibly the periplastid compartment, PPC); unfortunately, no biosilica extraction image is available for this construct. The localization of construct W2-LER4 (**Figure 15.D**) resulted in a location widely distributed all over the frustule too and, also in this case, a conspicuous part of the eGFP signal was trapped in what appears to

be the same PPC location; biosilica extracts imaging confirmed that at least a certain percentage of the total fluorescent signal is still retained from the frustule (**Figure 15.D2**). For both W2-LER3 and W2-LER4 the ratio between frustule labelling and intracellular labelling based on the apparent intensity of fluorescence is shifted towards the intracellular location.

3.3.3 CinY2-DERIVED CONSTRUCTS

Analogously to what was done for CinW2, I have generated some modified CinY2-derived constructs (**Figure 16**) and I have localized the encoded protein fused to eGFP, to investigate the role of the Y2-LERs and to identify potential targeting sequences.

3.3.3.1 CinY2 SEQUENCE TRUNCATIONS

The construct CinY2 SP is analogous to the construct CinW2 SP, as it composed of the first 15 amino acids of the CinY2 amino acid sequence, representing the predicted SP of the protein; as for the CinW2 SP construct, the eGFP signal from CinY2 SP (**Figure 17.A**) is localized exclusively in an intracellular compartment which appears to be closely associated with the plastid. Similarly to what was done for CinW2, I created two truncated fragments of the CinY2 full-length protein: CinY2 TR-A and CinY2 TR-B.

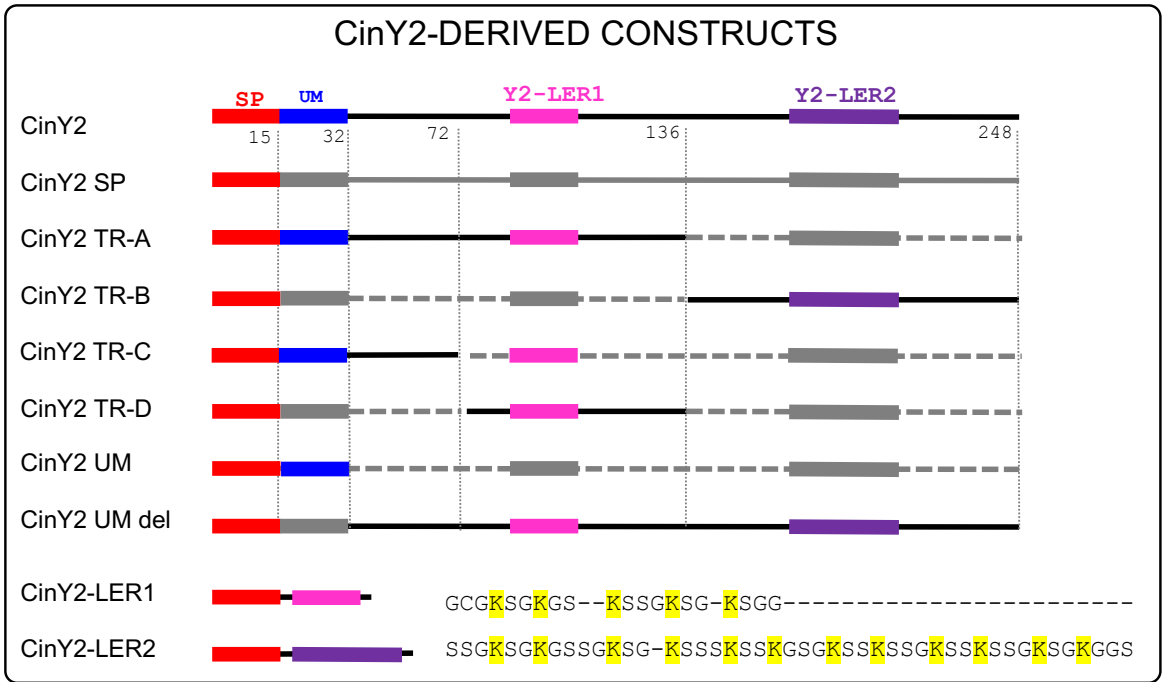


Figure 16 CinY2-derived constructs; graphical depiction of the amino acid sequences. The first is the full-length CinY2 protein (the color code is indicated on top of each feature). The numbers indicate the last amino acid residue in a truncation. The regions depicted in grey are the ones that have been deleted. For the isolated LERs, of the amino acid sequence is also shown. Size of the constructs is not in scale.

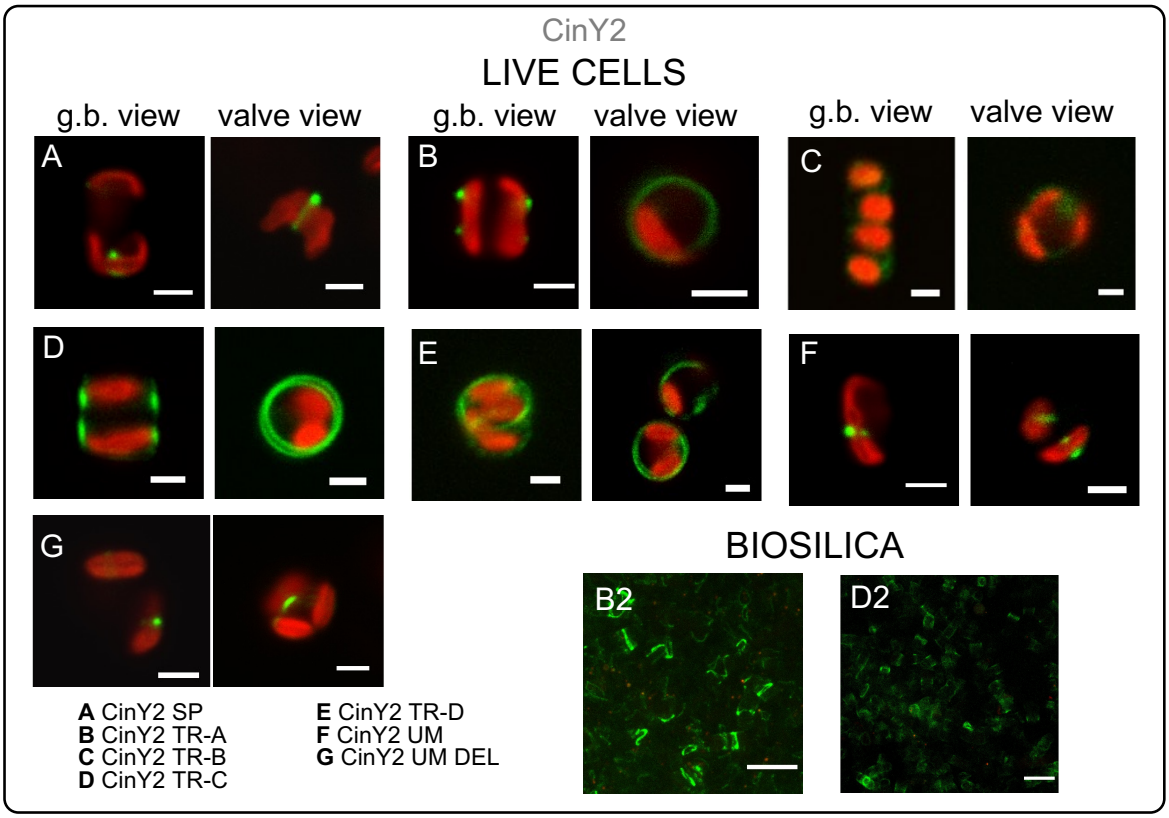


Figure 17 *In vivo* localizations of CinY2-derived constructs. The upper part of the figure shows the merge of the eGFP (green) and PAF (red) channels of confocal images of live cells (representative of each construct); the bottom part of the figure shows confocal images of extracted biosilica from two of the clones shown in the upper part. For each construct, the first column shows transapical sections of the cells (in girdle band view), whereas the second column shows transversal sections of the cells (in valve view). For the construct in E there is no girdle band view image, but a maximum projection of a tilted cell. Scale bars are 2 µm for the *in vivo* images and 10 µm for the biosilica images.

The construct CinY2 TR-A included the predicted SP and the amino acids from 16 to 136; the resulting peptide therefore represents the 121 aa-long N-terminal portion of the predicted CinY2 mature protein. The *in vivo* localization of this construct differed respect to the CinY2 native one: two distinct areas of the girdle were labelled, apparently corresponding to the same girdle band but on the two different thecas, thus resulting in a symmetrical labelling (**Figure 17.B**). The CinY2 TR-B construct included the predicted SP and the amino acids from 137 to 248, therefore representing the 111 aa-long C-terminal portion of the predicted native CinY2 protein. This construct did not result in any frustule-associated eGFP fluorescent signal (**Figure 17C**), indicating that the C-terminal half of CinY2 could not target the eGFP to the frustule (similarly to the CinW2 TR-B construct).

According to the results obtained, I hypothesized that a frustule-targeting motif was still present in CinY2 TR-A but absent in the CinY2 TR-B. Hence, I decided to generate two additional truncations of the CinY2 TR-A construct: the first construct, CinY2 TR-C, included amino acids from 1 to 72; the second truncation, CinY2 TR-D, included instead the predicted SP plus amino acids from 73 to 136. In both cases, the encoded eGFP-

fusion proteins were targeted to the frustule, even though with different specific sub-localizations: for CinY2 TR-C a dual location was observed, in which the fluorescent signal symmetrically labelled those girdle bands which are located in the central part of the cingulum of each theca, and that are proximal to the pleural bands (**Figure 17.D**). For CinY2 TR-D, another dual – although asymmetrical – localization was observed (**Figure 17.E**); the eGFP signal labelled a ring-like structure closely associated with the valve (most likely the valvocopula) and, on the other pole of the cell, a ring-forming pattern of pore-like structures (most likely the fultoportulae).

3.3.3.2 CinY2 isolated LERs

Similarly to what was done for CinW2, to study the targeting activity of LERs in CinY2 I created two constructs containing the isolated Y2-LER1 and Y2-LER2. The transformation efficiency (intended as the ratio between the number of eGFP-positive clones and the total number of colonies on selection plates after transformation) was very low for both constructs, and despite the repeated experiments, I was able to observe eGFP signal labelling some parts of the frustule only in few clones, and these clones showed very weak signals. This signal appeared to be associated with the valve and perhaps even with the fultoportulae but, given the abovementioned factors, I consider these as preliminary results, that would require further experiments before being discussed.

3.3.4 Sil3-derived constructs

To extend the investigation already carried on Sil3 from Poulsen et al., 2013, I have generated Sil3-derived constructs, for *in vivo* localization experiments (**Figure 18**). The construct Sil3 SP is composed of the first 17 amino acid residues, representing the predicted SP of Sil3; the localization of this construct was previously confirmed to be in the ER by Poulsen et al., 2013 (the expression of the gene was regulated by the light-induced *fcp* gene's regulatory regions. For my experiments, instead, the native promoter and terminator were used but – not unexpectedly – the fluorescent signal of construct Sil3 SP was localized in an intracellular compartment associated with the plastid, most likely the cER (**Figure 19.A**), similarly to what was obtained by Poulsen et al., 2013.

According to Poulsen et al., 2013, the two PLC present in the Sil3 sequence (PLC3nt and PLC3ct) are important for the efficient silica-targeting of the protein, but they did not localize the full length protein when the two PLC have been deleted. So, I have localized a construct (Sil3 DEL), derived from the full-length Sil3 protein, in which only the

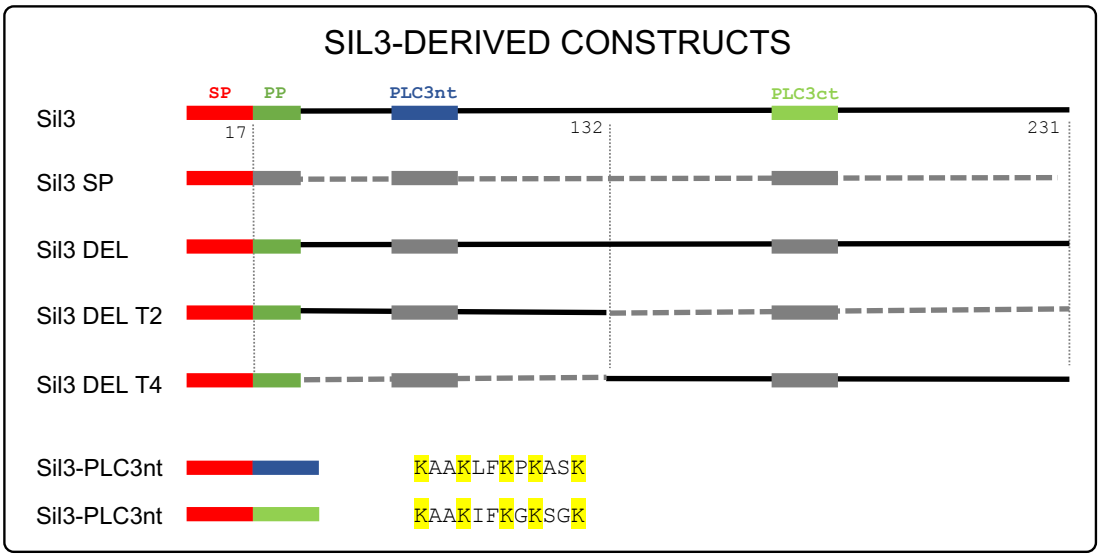


Figure 18. Sil3-derived constructs. Graphical depiction of the amino acid sequences. The first is the full length Sil3 protein (the color code is indicated on top of each feature); PP is pro-peptide, the PLCs are those described in the text. The number indicates the last amino acid residue in a truncation. The regions depicted in grey are the ones that have been deleted. For the isolated PLCs, the amino acid sequence is also shown. Size of the constructs is not in scale.

two abovementioned PLC were absent. As previously described, the full-length Sil3-eGFP fusion protein labels all parts of the frustule except for the region corresponding to the overlap of the two thecae (the pleural bands) (**Figure 12.A**). In the case of the Sil3 DEL construct, the resulting fluorescent signal mostly comes from the frustule, even though a conspicuous amount of the signal can be detected inside the cell (**Figure 19.B**); the frustule labelling is analogous to the one obtained from full-length Sil3-eGFP protein, albeit some of the parts of the frustule which are labelled in the latter case seem to be free of any fluorescence, as if there were gaps in the signal.

I created two additional Sil3 DEL-derived constructs by generating a truncation between amino acids 132 and 133. The two fragments created are identical to the T2 and T4 fragments localized from Poulsen et al., 2013 (which are substantially the N-terminal and C-terminal halves of the protein, respectively), but since they derived from Sil3 DEL no PLC is present in the fragments. The two new constructs I created have been named Sil3 DEL T2 and Sil3 DEL T4, and they represent the N-terminal and C-terminal portions of the Sil3 DEL protein, respectively. The localization of these constructs indicated that both of them still retained frustule targeting activity: the construct Sil3 DEL T2 (**Figure 19.C**) showed a higher ratio between intracellular fluorescence and frustule fluorescence compared to Sil3 DEL, but generally the frustule localization appeared similar (with some parts of the frustule not labelled); an analogous localization, but with a much weaker signal, was obtained by the construct Sil3 DEL T4 (**Figure 19.D**).

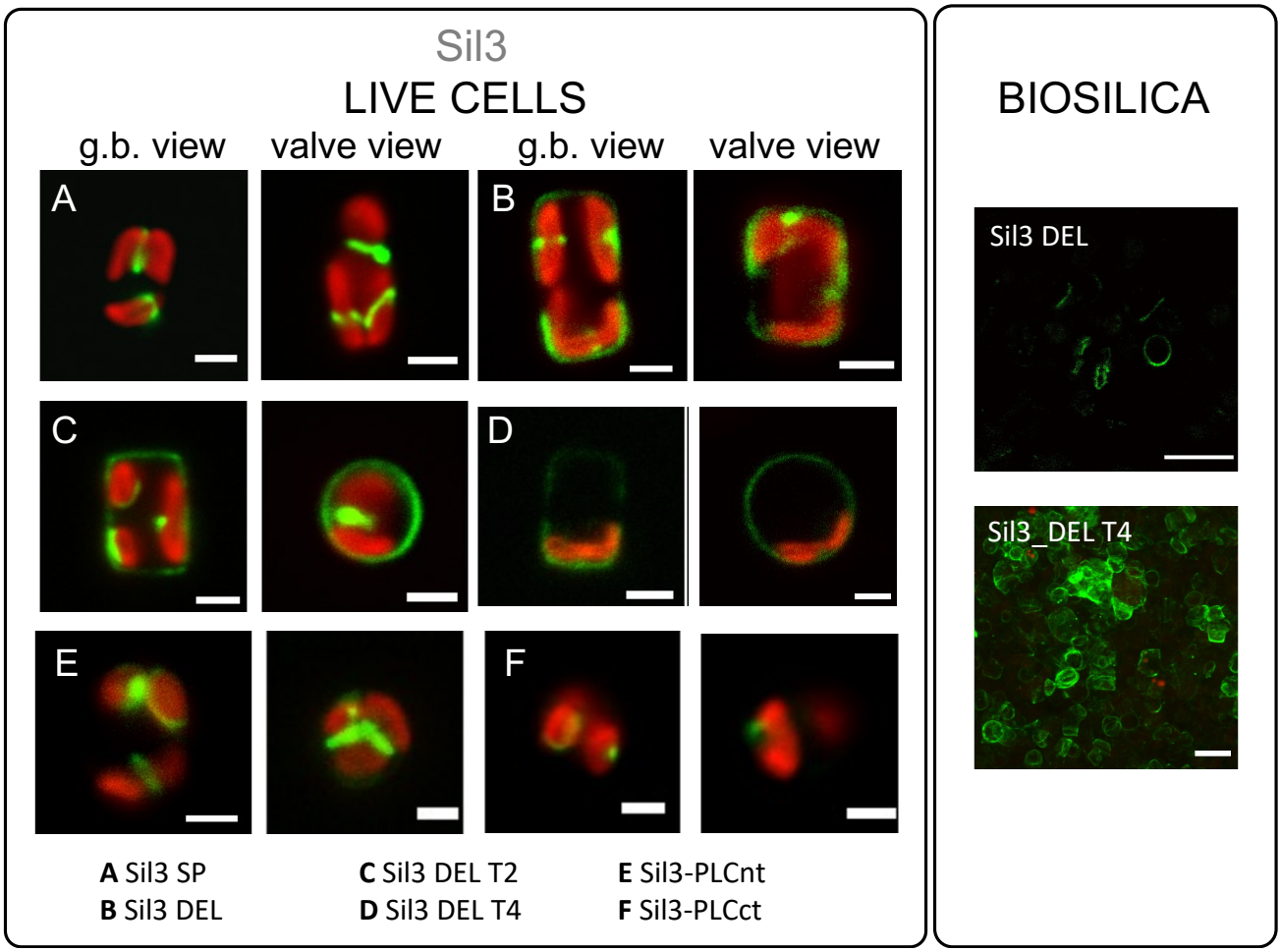


Figure 19 *In vivo* localizations of Sil3-derived constructs. Confocal images of live cells (representative of each construct) from *in vivo* localization experiments of the Sil3-derived constructs. The panel on the left shows the merge of the eGFP (green) and PAF (red) channels, whereas the panel on the right shows the merge of the same channels with images of extracted biosilica from the clones shown in the panel on the left. For each construct, the first column shows transapical sections of the cells (in girdle band view), whereas the second column shows transversal sections of the cells (in valve view). For the constructs in A and B there is no valve view image, but only two girdle band view images. Scale bars are 2 µm for the *in vivo* images and 10 µm for the biosilica images.

The silica targeting of the isolated PLC3ct from Sil3 was also investigated by Poulsen et al., 2013, indicating that this motif is capable of directing the eGFP to the frustule, although with very low efficiency; the expression of the gene was regulated using a light-induced promoter and therefore the peptide was overexpressed under the conditions used. To test if the use of the native promoter could affect the targeting of PLC3ct, and to verify if also PLC3nt retained the targeting activity, I have generated two Sil3-derived constructs, one for each PLC, in which the motifs were flanked by the SP at the N-terminal and by the eGFP at the C-terminal. For both constructs, the results of the localization experiments (**Figure 19**.**E**,**F**) showed fluorescent signal coming from internal cell compartments, apparently in close association with the chloroplast; this location was similar to what observed for Sil3 SP (see **Figure 19.A)**, and also for CinW2 SP (**Figure 14.A**) and CinY2 SP (**Figure 15.A**). Unfortunately, no confocal images of biosilica extracts are available.

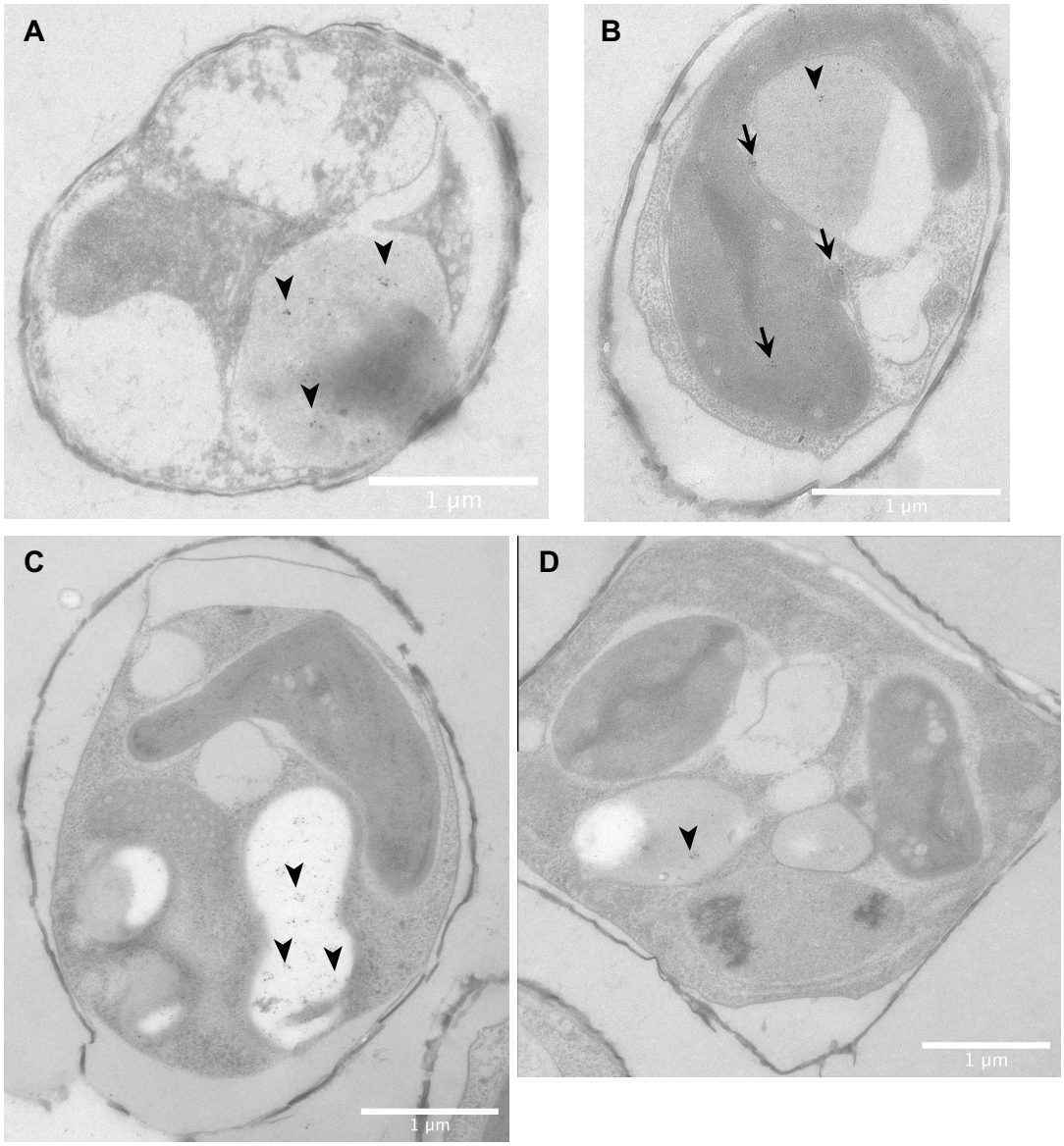


Figure 20 TEM images of immunolabeling experiments with Sil3 PLC3ct (A, B) and Sil3 PLC3nt (C, D). Arrows mark the plastid localization, whereas arrowheads mark the vacuole localization. Sample preparation and imaging performed by Dr. Thomas Heimerl.

Immunolabeling experiments have been performed to investigate the intracellular localization of the Sil3 PLC3nt-eGFP and Sil3 PLC3ct-eGFP fusion proteins (**Figure 20**): although the obtained labelling is poor, the results show the immunolabeled peptides as located inside the vacuole and, for Sil3 PLC3ct, also in some sub-compartment of the plastid; in this case, the gold particles labelled a region of the cell where the vacuole and the chloroplast appear to be in contact (**Figure 20.B**).

Given the structure of PLC and LERs, in which several KxxK motif are arranged together, forming a larger motif enriched in lysines, I have investigated also if a KxxK motif alone could function as targeting sequence; to do so, I have created an artificial construct in which the KSGK tetrapeptide was inserted between the SP and the eGFP of the pTpSil3

vector, and I have localized it (**Figure 22**). This experiment resulted in the eGFP signal stuck inside the cell, in a location closely associated with the plastid.

3.3.5 SEARCHING FOR NEW TARGETING MOTIFS

My results with CinW2, CinY2, and Sil3, together with the previously published data, indicate that PLC and LERs are not the only motifs involved in the targeting of the studied proteins, but they can – once isolated – target the proteins to the frustule. Thus, I have decided to search for new frustule-targeting motifs among the already known biosilica-associated proteins, and to exploit these motifs (and the PLC/LERs) to find yet unknown proteins that are targeted to the biosilica.

3.3.5.1 UNKNOWN MOTIF FROM CINY2

The analysis of the amino acid sequence of CinY1, CinY2 and CinY3 (see section 7.3.2 for the CinY2 amino acid sequence) allowed me to identify a 17 amino acid-long region which is conserved among the three proteins, not only for its amino acid composition and sequence but also for its position (the motif is located after the SP); I named this motif "*unknown motif*" (or UM). To investigate if the UM contributes to the frustule-targeting of the proteins to which it belongs, I have chosen CinY2 as candidate for further experiments.

In the first experiment, I have deleted all the genomic sequence downstream of the UM (see **Figure 16**), therefore leaving only the portion of the DNA construct encoding the SP and the UM (the resulting construct was name CinY2 UM). In the second experiment, I have deleted from vector pTpCinY2 only the genomic sequences encoding for the UM, leaving the remaining DNA coding sequence unaltered (this construct was named CinY2 UM DEL; see **Figure 16**). The resulting transformants cell lines generated using the abovementioned constructs showed no region of the frustule labelled; instead, the eGFP signal was localized exclusively inside the cell. The localization of both constructs appeared to be similar, if not identical, to the one obtained by the CinY2 only SP construct (**Figure 17**), although it is difficult to clearly identify the cellular sub-compartment only by analysing the eGFP channel and the plastid autofluorescence channel.

Considering these first results and given the presence of the UM in at least two other frustule-associated proteins (CinY1 and CinY3), I hypothesized that the UM could be a targeting motif and that it could be common in other frustule proteins too; thus, I made a

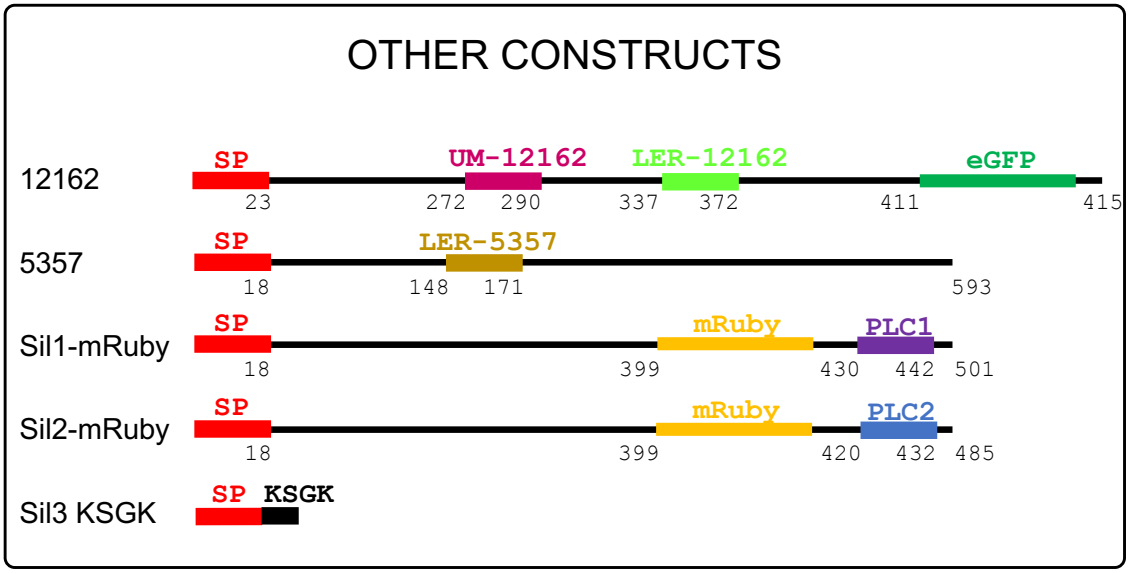


Figure 21 Graphical depiction of the primary sequences of the additional constructs studied. The position of the fluorescent tag is shown only for those constructs in which its position is internal to the main amino acid sequence (otherwise, the position is at the C-ter). The number indicates the last amino acid residue in a region. Size of the constructs is not in scale.

BLAST search of the UM against the *T. pseudonana*'s protein database, retrieving two hits corresponding to protein 12162 and protein 12163 (IDs are as in the JGI Browser database). According to the *in silico* data, both proteins have a SP for co-translational import into the ER and no predicted TMDs; protein 12162 is 415 amino acids long, whereas protein 12163 is 1155 amino acid long (see sections 7.3.4). Interestingly, the two proteins present one (12162) and two (12163) regions enriched in KxxK motifs. The position of the UM in the sequence is different between the two proteins: in protein 12162 the motif is located internally, close to the C-terminal; in the predicted protein 12163, instead, the motif is located immediately after the SP (as it was for the abovementioned cingulins CinY1, CinY2 and CinY3).

In both proteins there are RXL motifs (one in 12162 and three in 12163); these motifs are frequent in frustule-associated proteins and are considered potential target sites of proteolytic cleavage. When planning the insertion of the eGFP, I have considered this factor: in 12162 the fluorescent protein has been inserted immediately upstream of the RML tripeptide (after residue 410), whereas in 12163 it has been inserted before the first RVL tripeptide (after residue 562). In this way, the eGFP protein would be located downstream to the UM and to the KxxK motifs-enriched regions but upstream respect to the RXL motif in both constructs; so if cleavage occurred at any RXL it would not separate

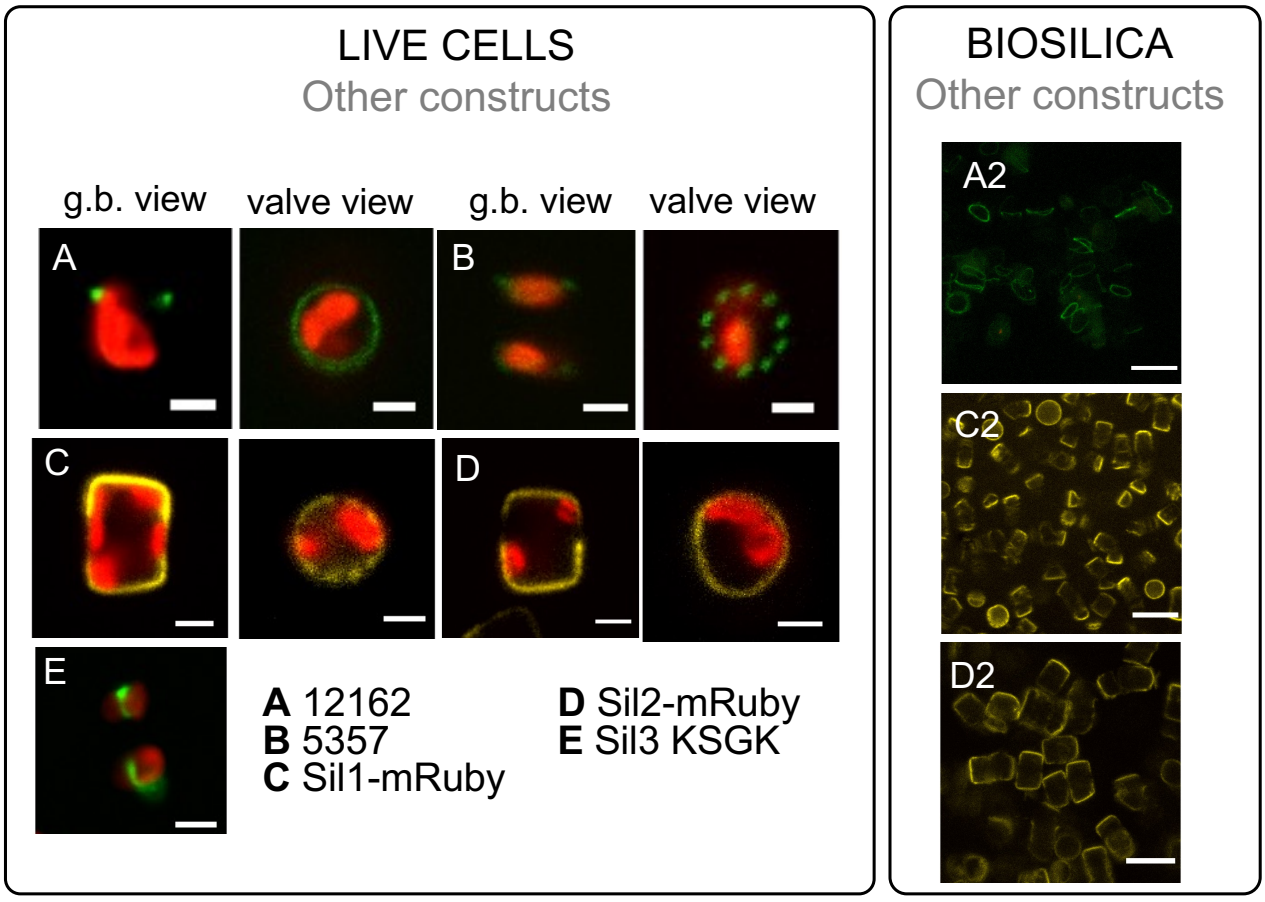


Figure 22 *In vivo* localizations of 12162, 5357, Sil1, Sil2, and KSGK. The panel on the left shows the merge of the eGFP (green) and PAF (red) channels of confocal images of live cells (representative of each construct) from *in vivo* localization experiments. For each construct, the first column shows transapical sections of the cells (in girdle band view), whereas the second column shows transversal sections of the cells (in valve view) The panel on the right shows the same channels' merge of confocal images of extracted biosilica from some of the clones shown in the panel on the left. For the construct E, there is no valve view image; the valve images of B and C show the surface of the valve. Scale bar is 2 µm

the region of interest from the fluorescent tag. The transformation vectors generated were named pTp12162-GFP and 12163-GFP.

The transformation with the construct pTp12162-eGFP was successful, although the efficiency was very low; in fact, I managed to obtain only one clone showing eGFP-tagging of the frustule (**Figure 22**). The fluorescent signal appears to be localized on a circular band close to only one valve region and not both, as it would be expected given the symmetrical morphogenesis of diatom's frustule. I hypothesize that this protein has a similar localization to that of CinY2 and that this frustule localization is observed because the imaged cells might have just synthesized the new valve, and therefore the hypocingulum was still in the early stage of development (resulting therefore in an overlap of the fluorescent signal coming from both thecas). I have, in fact, observed also cells showing an eGFP signal that was in a central position respect to the girdle (no image available).

The biosilica extraction images confirmed that the signal was localized in the frustule, specifically in only one circular region of the girdle (most likely one, or few girdle bands). This protein is therefore a new frustule protein.

Unfortunately, I was never able to obtain GFP-positive transformants using the pTp12163-GFP construct; analysis by colony PCR, indicated that none of the NAT resistant clones integrated the 12163-eGFP genomic region. Despite that, given its characteristics, I still consider the 12163 predicted protein to be a potential frustule-protein candidate, but obviously in need of experimental validation.

3.3.5.2 SCREENING OF THE SILAFFIN-LIKE RESPONSE GENES (SLRG) DATABASE

I screened the SLRG dataset (see section 1.4.2) for candidate proteins based on: i) the presence of a SP for co-translational import into the ER; ii) the absence of TMDs; iii) the presence of KxxK-motifs. I have therefore obtained five candidates: 22220, 8219, 23861, 22630 and 5357 (IDs are reported as in the JGI database).

The gene 22220 was discarded during the cloning because the corresponding protein was localized as a fultoportulae-protein by another research group (Christophe Heintze, personal communication). Genes 23861, 8219 and 22630 were cloned into the vector pTpNAT-MCS together with their upstream and downstream regions (800 bp and 300 bp, respectively) and with the eGFP at the C-terminal, but for all three genes the transformation experiments never resulted in any eGFP positive clone. Gene 5357 was also cloned together with its upstream and downstream regions and with the eGFP at the C-terminal (**Figure 21**), and the transformant cell lines expressed a fusion protein that localized in the valve, more precisely tagging the pattern of distribution of the fultoportulae (**Figure 22.B**); also this protein is therefore a frustule-associated protein. Unfortunately, also for this construct no images of extracted biosilica are available.

3.4 PROMOTER STUDIES AND TARGETING

To investigate whether the timing of the expression of the *sil3*, *cinW2* and *cinY2* genes could influence the valve- or girdle band-targeting of the encoded proteins I have performed the following experiments.

3.4.1 PROMOTERS EXCHANGE

The different frustule localizations that we have seen in the known cell wall-associated proteins could be obtained by a combination of targeting sequences internal to the

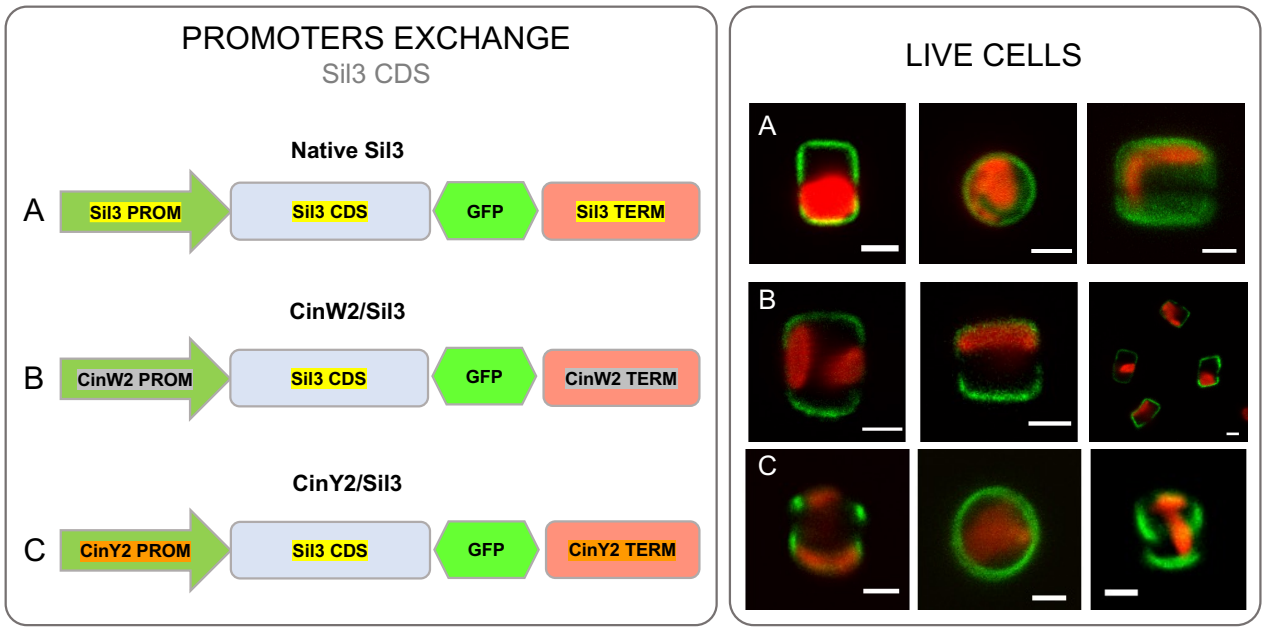


Figure 23 Exchange of promoters for the Sil3 CDS. In the panel on the left are shown graphical representations of the constructs used for transformation; in the panel on the right, there are confocal images from representative live cells, created by merging the eGFP channel (green) with the PAF channel (red). The third image in A and C are Z-stack projections. Scale bar is 2 µm.

peptides and appropriate timing of the expression of their encoding genes; swapping promoters between genes encoding for proteins having a different frustule localization would result in an alteration of their expression, that could eventually cause the protein's mistargeting. The exchange of the promoters, from the technical point of view, is an exchange of the coding sequences of the genes rather than their promoters and terminators. The 6 new plasmids generated with the permutation of their genomic regions of interest have been named by specifying both the regulatory regions and the CDS, in this order; for example, the constructs Sil3/CinW2 and Sil3/CinY2 are those in which the Sil3 promoter and terminator regulate the expression of the CinW2 and CinY2 genes, respectively. The vectors have been used for biolistic transformation and the resulting transformants screened at the confocal microscope for eGFP detection. Some representative images of the transformants are shown in **Figure 23**, **Figure 24** and **Figure 25**.

The eGFP labelling of the CinW2/Sil3 (**Figure 23.B**) construct appears to be very similar if not identical to the labelling of the native Sil3 protein (**Figure 23.A**); the fluorescent signal is localized both in the valve and in the girdle area, but not in the pleural bands (or perhaps neither in the girdle bands proximal to the pleural bands). The fluorescent signal of the construct CinY2/Sil3, instead, is absent from the valve plate but it is localized in the valvocopulae and in one additional girdle band (**Figure 23.C**); the overall

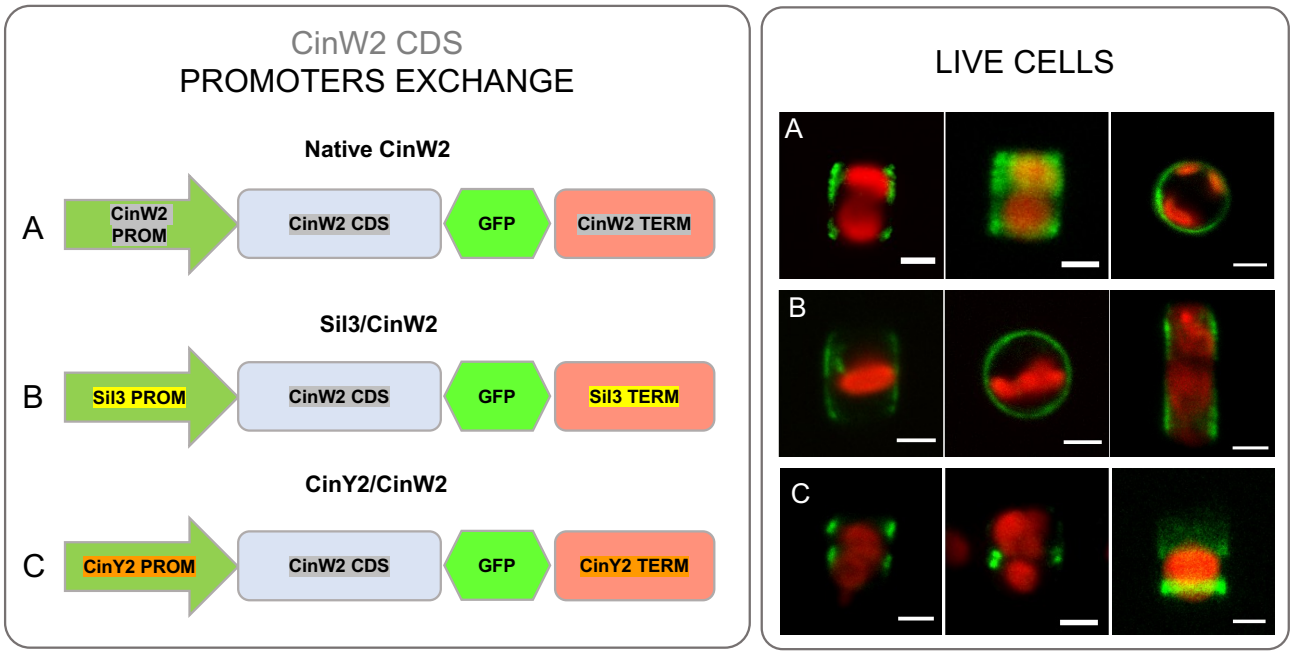


Figure 24 Exchange of promoters for the CinW2 CDS. In the panel on the left are shown graphical representations of the constructs used for transformation; in the panel on the right, there are confocal images from representative live cells, created by merging the eGFP channel (green) with the PAF channel (red). The second image in A, and the third image in C are Z-stack projections. Scale bar is 2 µm.

appearance of the fluorescent tagging appears to be somehow asymmetrical, with both valvocopulae and only one girdle band tagged.

The construct Sil3/CinW2 (**Figure 24.B**) also resulted in an eGFP tagging that was substantially identical to the one obtained by the native CinW2 (**Figure 24.A**); a minor fraction of the signal could be detected inside the cell, but nevertheless the multiple girdle bands labelling was clear. For construct CinY2/CinW2, instead, there were observable differences respect to the native CinY2 protein's localization: a dual, asymmetrical fluorescent tagging marked the area of the girdle, without any valve labelling; there was a thick fluorescent ring (apparently a single girdle band, although it was not quite clear) on one theca, and a less pronounced but somehow diffuse fluorescent signal on the other theca (**Figure 24.C**).

Transformation experiments using Sil3/CinY2 (**Figure 25.B**) and CinW2/CinY2 (**Figure 25.C**) have been more challenging than the previous ones, in fact the screening for eGFP-positive clones among all the transformants obtained was often unsuccessful and required multiple experiments. Additionally, the few clones I managed to identify as positive during the confocal microscope screening had only low amounts of fluorescent signal, mostly located in intracellular sub-compartments which appear to be associated with the plastid (for both constructs) and with the cytosol (for construct CinW2/CinY2); a

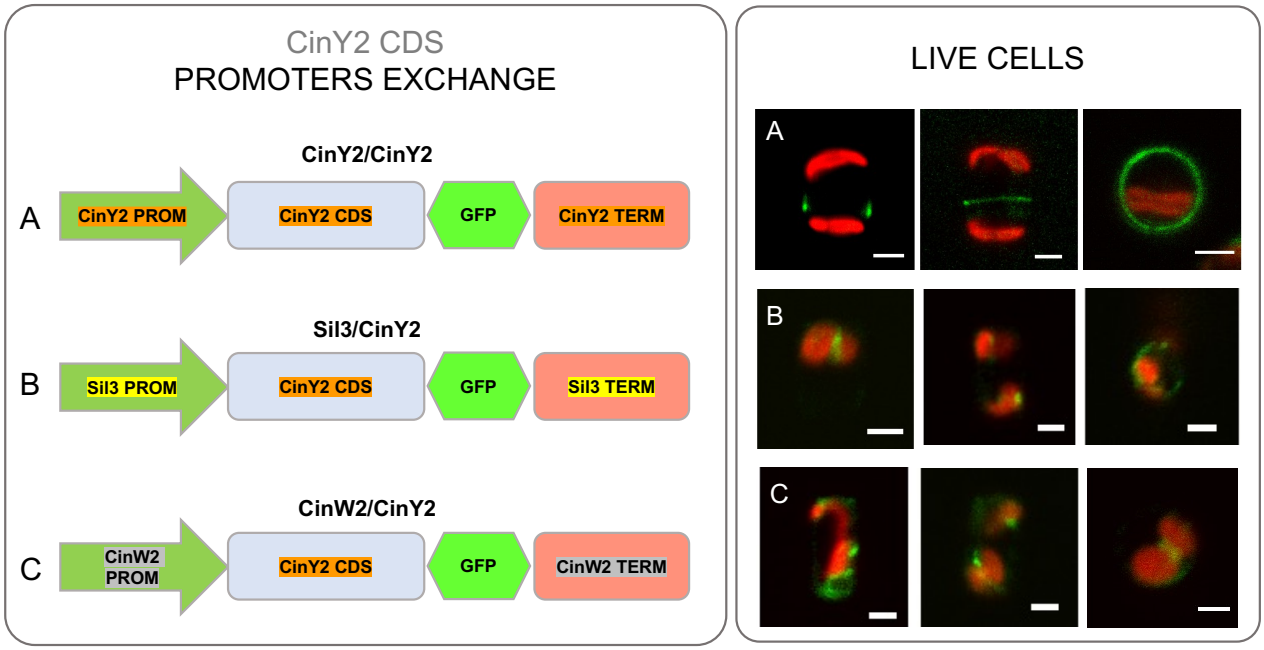


Figure 25 Exchange of promoters for the CinY2 CDS. In the panel on the left are shown graphical representations of the constructs used for transformation; in the panel on the right, there are confocal images from representative live cells, created by merging the eGFP channel (green) with the PAF channel (red). The second image in A is a Z-stack projection. Scale bar is 2 µm.

minor fraction was also found in the frustule, as ring-shaped girdle band-like structures (for Sil3/CinY2), although is not clear the exact position of this ring-shaped area.

3.4.2 PROMOTERS ISOLATION

As a sub-project, I intended to isolate the cis-acting regions specifically involved in regulating the expression of the *sil3*, *cinY2* and *cinW2* genes. To achieve that, I have decided to follow an approach based on the *in vivo* localization of the eGFP upon regulating the expression of its gene by stepwise truncated versions of each one of the sil3, *cinW2* and *cinY2* promoters and terminators (**Figure 27**). After the deletion of the CDS of the frustule-associated gene from the corresponding vector, and following the transformation, the expression of the *egfp* gene would be driven by the gene's specific regulatory regions, but all specific intracellular signal sequences internal to the protein (*e.g.*, the SP, or any targeting motif) are lost; the transcripts derived from these constructs are translated in the cytosol and, consequently, the expected localization of the eGFP in these experimental conditions is cytosolic. Obtaining no more cytosolic localization or a strongly reduced intensity of fluorescence might indicate loss of motifs responsible for the transcriptional regulation, that could be then precisely identified and studied.

To reduce as much as possible any undesired regulation of the expression from the backbone of the plasmid, I have used vectors derived from vector pTpINV-NAT, in which

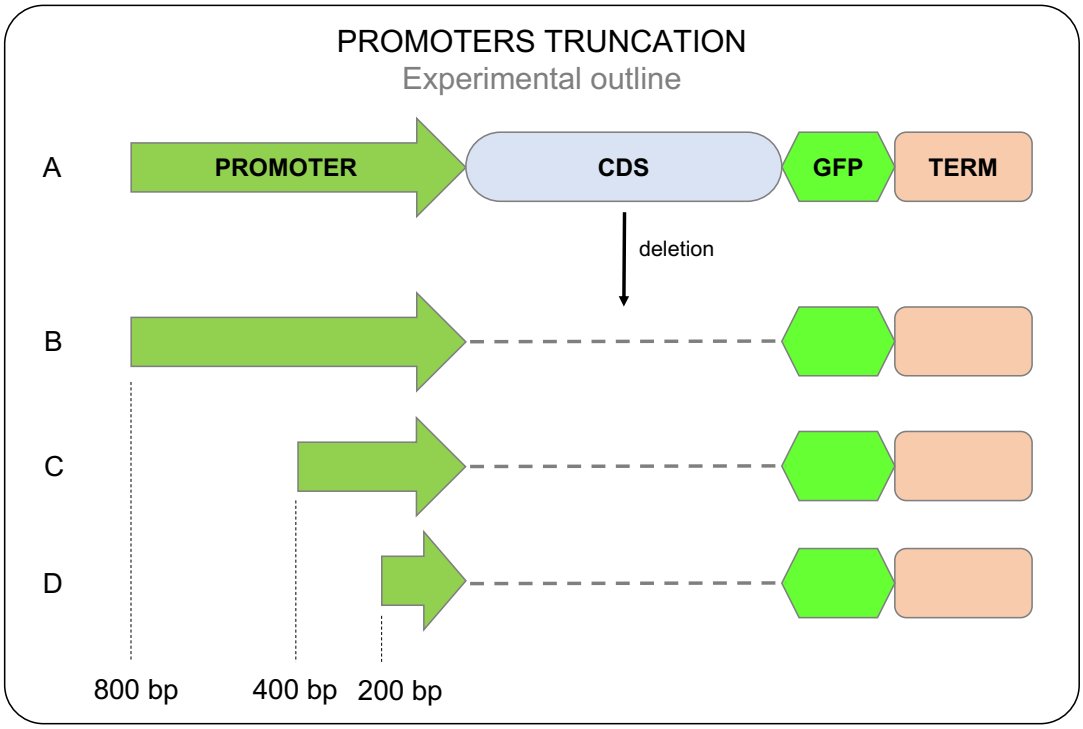


Figure 27 Graphical summary of the promoters' truncation strategy. **A**) General structure of the native gene cassette, where the promoter (800 bp) is located upstream to the CDS and the eGFP is located between the CDS and the terminator; this structure is the same for the three genes studied **B**) After deletion of the CDS, the promoter is immediately upstream to the eGFP. **C**) First truncation (400 bp). **D**) Second truncation (200 bp).

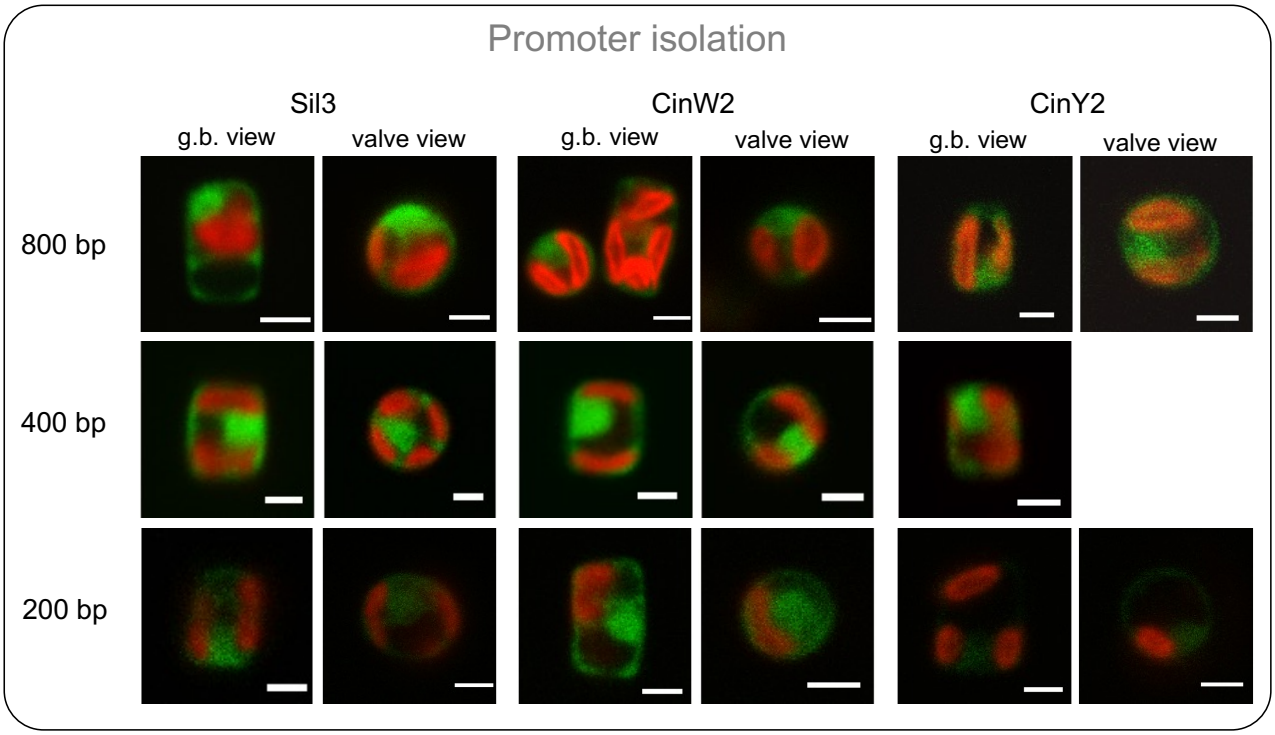


Figure 27 *In vivo* localization of the constructs derived from the truncations of the promoters. For each gene and for each truncation (except for CinY2 400 bp), images of cells in both girdle band view and valve view are shown. Scale bars are 2 µm.

the NAT cassette has been shifted from its original position (downstream respect to the terminator region) to a new position in the region immediately upstream of the promoter. In this way the *fcp* terminator is immediately upstream to the start of the promoter of the gene of interest. The localization of the GFP upon transformation with the “800 bp” constructs was, as expected, in the cytosol for all the three genes (**Figure 27**). The constructs derived from the second truncations, (“prom-400”) also resulted in transformants expressing the GFP in the cytosol for all the three genes; the same was the case of the “prom-200”, even though in this case the intensity of the fluorescent signal was much fainter than what was observed for the other two sets of constructs, especially in the case of CinY2.

3.4.3 NATIVE SIL1 AND SIL2 LOCALIZATION

The genomic sequences of the predicted *sil1/2* gene were amplified as described in the Materials and Methods section, including the predicted native promoter and terminator regions; the cloning of the PCR product into the pJet shuttle vector and the following mini preparations extraction gave a mixed population of vectors: among 6 mini preparations sequenced, 4 of them matched the published Sil1 sequence, whereas the remaining two mini preparations matched the available Sil2 sequence, according to the pattern of differences existing between the Sil1 and Sil2 protein (as from the available published *sil1* transcript sequence; Poulsen & Kröger, 2004). Both genomic sequences were inserted into the pTpNAT-MCS vector and the fluorescent tag mRuby (Marter et al., 2020) was inserted after amino acid 399 in both construct; this position is the same position as that of the eGFP in the previously published localization experiments in Sil1 (Poulsen et al., 2013). The two vectors generated have been named pTpSil1-mRuby and pTpSil2-mRuby.

Not surprisingly, the *in vivo* localization obtained by transformation with both constructs was identical (**Figure 22**); as already discussed elsewhere (Poulsen & Kröger, 2004), the two proteins have 91% of amino acid sequence identity and it was therefore expectable to obtain the same frustule localization for both. Unexpectedly, this localization was not exactly corresponding to the one previously published (Poulsen et al., 2013) in which the Sil1 protein was localized merely in the valve (specifically around the fultoportulae) and apparently also the first girdle band (the valvocopula), when the gene’s expression is regulated by the ligh-induced *fcp* promoter. In my experiment, instead, both Sil1 and Sil2 proteins are localized in the valve and in the girdle area, leaving untagged only the central area of the girdle, in a very similar way as the Sil3 protein. This was the first documented use of the fluorescent tag mRuby in *Thalassiosira pseudonana*.

4. DISCUSSION

With my work I have tried to answer to the question whether the lysine-enriched regions (LERs) in cingulins have a similar function to the pentalysine clusters (PLC) of silaffins, and if it was possible to use a frustule-targeting sequence to re-direct a valve protein to the girdle band or vice versa. Additionally, I have studied if a different transcriptional regulation of the genes could cause a significative alteration of the targeting in the encoded frustule-proteins and, in parallel, I have partially isolated *cis*-acting components of the non-coding genomic upstream regions of *sil3*, *cinW2*, and *cinY2*. Last, I have used the available published data and the results of my experiments to identify novel frustule-proteins.

4.1 TARGETING SEQUENCES

Given the complexity of the cellular and molecular mechanisms involved in the biosynthesis of the frustule, its peculiar morphogenesis (creating distinct compartments, like valves and girdle bands), and the dependence on the cell cycle, when interpreting the results of the *in vivo* localization experiments (sections 3.3.2, 3.3.3, 3.3.4, and 3.3.5) I have distinguished between three different types of modification (**Figure 28**): **type-I**) no apparent mistargeting of the mutated protein in respect to the native one was detected (**Figure 28.C**); this scenario would imply that the mutated/deleted motif(s) or region(s) have no influence in the targeting of the constructs, or that other factors can completely overcome the modification; **type-II**) silica targeting was achieved, but the specific girdle band- or valve-localization was somehow different in the mutated protein respect to the native one (**Figure 28.D**); this scenario would imply that despite the changes in the protein sequence the intracellular trafficking and protein sorting is still achieved; the observed differences may arise by loss of specificity during the sorting (perhaps caused by the loss of just one out of more targeting motifs), or by loss of silica precipitation properties (caused by mutation or deletion of functional domains); **type-III**) silica targeting was completely abolished (**Figure 28**.**E-F**); in this scenario, the targeting to the frustule is compromised because of complete loss of the motif(s) needed for the correct intracellular trafficking. My results will be discussed following this interpretative guideline.

The picture that emerged from the targeting experiments in CinW2 and CinY2 is that for both CinW2 and CinY2, the deletion of the N-terminal portion of the protein determined a complete abolishment of the frustule targeting (a type-III modification), indicating that some targeting sequence was missing; the deletion of the C-terminal half,

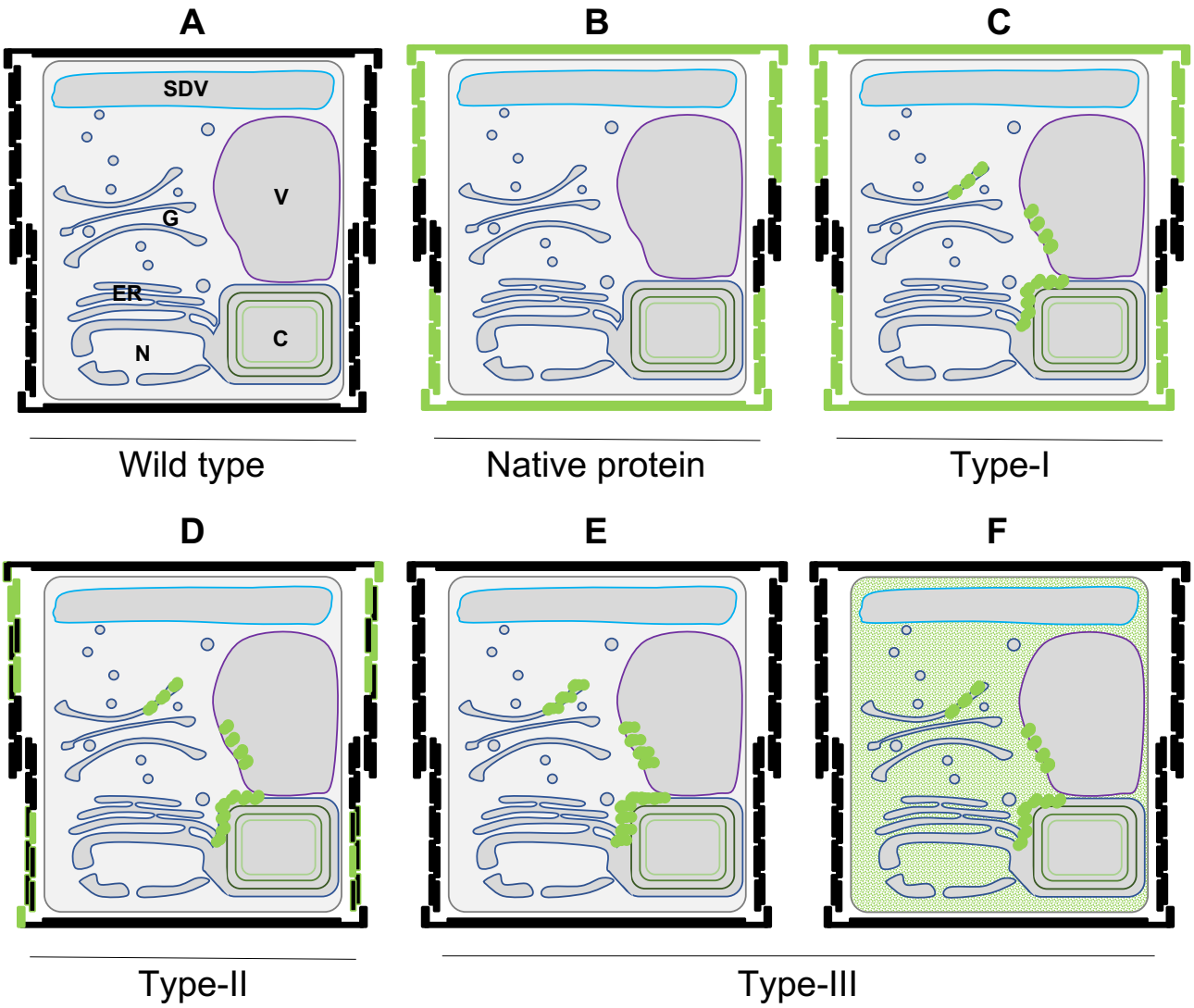


Figure 28 Schematic representation of the three different types of cell or frustule labelling obtained. The graphics show theorical scenarios resulting from different modifications of a hypothetical frustule protein; as an example, in its native, full-length form this hypothetical protein has a Sil3-like localization. **A**) wild type cells showing no eGFP labelling; **B**) hypothetical frustule labelling of the eGFP-fused protein; **C**) example of type-I modification; **D**) example of type-II modification; **E**) and **F**) examples of type-III modification. In C, D, E and F, random intracellular localizations are shown, to indicate the possibility of retained fluorescent signal inside the cell. For simplicity, the only organelles or cell components shown are: the nucleus (N), the endoplasmic reticulum (ER), the Golgi apparatus (G), the vacuole (V), the chloroplast (C), the silica deposition vesicle (SDV), the plasma membrane (in dark grey), and the frustule (in black).

instead, apparently left intact a portion of the native protein that retained the sequence(s) necessary for frustule-targeting, as the *in vivo* localization indicated that the truncation is a type-II modification. In both CinW2 and CinY2, each half of the proteins contains at least one LER (two in each CinW2's half and one in each CinY2's half), so, the loss of silica targeting in constructs CinW2 TR-B and CinY2 TR-B cannot be tracked back to the absence of LER, but it is possible that this result was due to different targeting properties of each LERs.

The experiments where lysines have been exchanged with arginines (MUT constructs), or were the LERs were deleted (DEL constructs) showed no changes in the localization of the constructs respect to that of CinW2 TR-A (type-I modification); the results of the MUT experiments differed from what was shown for the Sil3-derived construct T7 (Poulsen et al., 2013), where the lysine-to-arginine affected the silica targeting, although in this case the modified peptide was much shorter than the three MUT constructs (only 22 amino acids long respect to 187), therefore I cannot exclude that other internal portions of the truncation were involved. Possible explanations might be: i) the two tested W2-LERs are not involved in the targeting; ii) the mutations did indeed have an effect, but other factors overcame this effect and ensured that silica targeting was still achieved.

On the other hand, the localization experiments using isolated LERs indicated that these motifs can indeed target the eGFP to the frustule, but with remarkable differences between each LERs, and respect to the native CinW2 localization: W2-LER1 and W2-LER2 have clear targeting properties, as all fluorescent signal was apparently located in the frustule and not inside the cell. It is particularly interesting to observe that for both LERs the fluorescent signal was localized in (or at least, also in) the valve, more specifically in the fultoportulae, where the native CinW2 was never localized. This means that two LERs isolated from the girdle-protein CinW2 can function as valve-targeting sequences. Both W2-LER3 and W2-LER4 showed weak but homogeneous frustule labelling, whereas a conspicuous amount of fluorescence was stuck inside the cell, suggesting perhaps that the silica targeting is for some reason less efficient respect to the other two LERs from CinW2. It might sound reasonable to associate this result with the unsuccessful frustule-targeting of the CinW2 TR-B construct (that contains W2-LER3 and W2-LER4), but I am tempted to consider the results regarding the deletion of the LERs as not backing up this hypothesis, and instead as pointing towards a scenario in which other factors are responsible for the targeting. Another possibility could be that the deletion of the N-terminal half deprived the protein of some sort of transit peptide located in the close proximity of the SP, similarly to what is known regarding the transport of nuclear-encoded plastid proteins in diatoms (Maier et al., 2015). I exclude this possibility, given the following observations: when analysing the results of the *in vivo* localization of the CinW2 TR-C, CinW2 TR-D, CinW2 TR-C DEL and CinW2 TR-D DEL constructs, it can be observed that CinW2 TR-C gives a type-I modification respect to CinW2 TR-A (from which it derives); this suggests that the removal of the amino acid residues from 104 to 187 had no visible effect. Instead, construct CinW2 TR-C DEL (created by deleting W2-LER1 from construct

CinW2 TR-C) showed a type-II result: the biosilica was labelled, although not homogeneously and with low efficiency (as indicated by the faint signal and the additional intracellular location); this suggests that W2-LER1 is not necessary for silica targeting, but its deletion has a strong influence on the targeting efficiency and on the overall distribution of the fluorescent signal. The construct CinW2 TR-D resulted in a fultoportulae localization (type-II modification), which was identical to the localization obtained by the isolated W2-LER2, as if the LER was indeed the signal responsible for the frustule localization of the construct. When deleting this LER from the construct CinW2 TR-D (resulting in construct CinW2 TR-D DEL) a significant different frustule localization respect to the one obtained by CinW2 TR-D was obtained, with the valve plate and some of the girdle tagged, but the fultoportulae untagged (again, a type-II modification). These results indicate that despite lacking the first N-terminal 103 amino acid residues, both CinW2 TR-D and CinW2 TR-D DEL are still targeted to the frustule; so, the need of a transit-like sequence at the N-terminal of the protein is unlikely in CinW2.

Constructs CinW2 DEL-C, CinW2 TR-C DEL, and CinW2 TR-D DEL have no LER, instead they have three, two, and one KSGK motifs, respectively. I have questioned if the targeting activity of certain PLC or LERs could be due to the presence of single KxxK motifs, rather than clusters of them. My attempt to use a single KSGK motif fused to the SP of Sil3 to target the frustule was unsuccessful, as all fluorescent signal was retained inside the cell. It seems therefore that there are three possible explanations: i) the short motif cannot function as a targeting sequence; ii) the KSGK cannot exert the targeting function when expressed as a short tetrapeptide motif only; in this case, a better approach instead of using a single KxxK motif, would be to delete such short motifs from a longer peptide that still retains the frustule-targeting properties but has no PLC or LER, like for example the three constructs CinW2 DEL-C, CinW2 TR-C DEL, and CinW2 TR-D DEL; iii) this motif is a targeting sequence but is involved only in early or late steps of the trafficking; in this case, TEM immunolabeling and co-localization studies would help understanding the exact location of the eGFP-tagged peptide.

Similarly to what observed for CinW2, also in CinY2 the series of truncations that I have made did not give a clear identification of the sequences responsible for the targeting. In addition to CinY2 TR-A and CinY2 TR-B discussed above, both the constructs CinY2 TR-C (no LERs) and CinY2 TR-D (containing CinY2-LER1) were localized in the frustule (both type-II modifications). Thus, as it was in CinW2, it appears that this LER might be not the dominant targeting sequence (although the targeting activity of LERs

isolated from CinY2 is still to be tested). Moreover, despite the deletion of the N-terminal half of the protein in the construct CinY2 TR-B abolished the targeting (type-III modification), the results obtained by the localization of construct CinY2 TR-D suggests that no transit-like sequences are present in the N-terminal portion of the protein immediately following the SP, as I have hypothesized for CinW2.

Interestingly, the so-called Unknown Motif (UM) appears to have some influence in the targeting process, although the results are a bit puzzling. The deletion of only this motif from the native CinY2 sequence (CinY2 UM DEL), in fact, did not allow frustule-targeting in the mutated peptide, suggesting perhaps that the motif could be involved in the process. The results of the localization of CinY2 UM, instead, where the UM was the only portion of the protein left after the SP, might be explained by the UM being a motif involved in the first steps of the intracellular protein trafficking. After comparing the results from CinY2 TR-C, CinY2 TR-D, CinY2 UM and CinY2 UM DEL, I hypothesize that a multi-level targeting mechanism exist for the CinY2 protein, involving the UM, and some unknown sequences located between amino acids 32-72, and 73-136. Additional, although indirect, evidences of a potential involvement of the UM motif in the protein targeting to the frustule came from the identification of the protein 12162. This protein, in fact, was identified as a new frustule protein thanks to the presence of a UM-like motif, indirectly supporting the hypothesis that the UM is indeed involved in the frustule-targeting of the proteins, although further experiments would be required. It would be useful, for example, to test if the deletion of this motif from the full-length protein 12162 determines the same mistargeting encountered with CinY2 UM DEL; given the position of the UM in the primary sequences of CinY1 and CinY3, and its amino acid composition, it is possible that in CinY1 and CinY3 it has a similar function as the one suggested in CinY2. Similarly, protein 5357 was identified thanks to the presence in its primary sequence of a LER region.

The deletion of both PLC from the Sil3 full length protein showed that the modification applied affected more the overall frustule distribution of the eGFP-tagged protein than its frustule-targeting, since both the valve and the girdle bands were tagged (a type-II modification). These results somehow confirm what was already proposed by Poulsen and co-workers, that pentalysine clusters are not the only motifs responsible for the correct frustule-targeting of Sil3. The fact that the Sil3 DEL construct was at least partially transported to the frustule indicates that the putative cell wall targeting motif(s) were still present. Additionally, the localization of the Sil3 DEL T2 and Sil3 DEL T4 constructs suggested that these putative targeting sequences were contained in both the

Sil3 DEL T2 and Sil3 DEL T4 protein fragments (which have 5 and 7 KxxK motifs respectively). The gaps in the frustule-associated fluorescent labelling of the construct Sil3 DEL could be determined by an inefficient silica precipitation activity of the modified protein, caused by the deletion of the two PLC; post-translational modification of lysines and serines is relevant in ensuring the silica precipitation properties of Sil3 (Kröger & Poulsen, 2008) and therefore the deletion of both PLC could deprive the protein of some of those KxxK domain which are needed for the biosilicification process. In silaffins, the targeting function of PLC has been previously investigated (Poulsen et al., 2013); different PLC showed different silica targeting properties, with PLC1 from Sil1 and the artificial PLCart showing a widespread, homogenous tagging on the frustule; instead, with PLC3ct from Sil3, although biosilica extracts showed frustule labeling, a conspicuous amount of eGFP remained stuck inside the cell, indicating poor silica targeting; all these constructs were localized upon overexpression using the light-induced *fcp* promoter. In my experiments, always using the native promoter and terminator, I have localized the very same PLC3ct and the other PLC from Sil3 (PLC3nt); it is unclear to me why no frustule labelling occurred, although – as mentioned – no biosilica was extracted and I cannot therefore completely exclude that a certain (perhaps small) percentage of the total eGFP-fused protein have tagged the frustule. In any case, the immunolabeling and TEM observation identified both constructs in the vacuole of the cell, and PLC3ct also in the chloroplast. These experiments would need confirmation from a higher number of clones, but if backed up by additional localizations it would indicate a vacuolar transit for the two isolated PLC from Sil3.

Why do such relevant differences in the frustule-targeting of the isolated LERs from CinW2 exist? When checking the amino acid composition of the LERs from CinW2 it is immediately evident how much W2-LER3 and W2-LER4 are similar. In W2-LER4, in fact, except for an additional serine residue immediately before the fourth lysine, the only other difference between the two constructs is given by an aspartic acid substituting the first amino acid (a serine) of the construct. The differences between W2-LER1 and W2-LER2 are more significant, the latter is 7 amino acids longer than the former and there are three additional single residues changes.

It has been demonstrated that the post-translational modification pattern of lysines in Sil3 is a regulated by a specific set of rules (see section 1.2.5.1 and Sumper et al., 2007), depending on the clustering of lysines into KxxK motifs and on the amino acid composition in the close proximity of the lysine studied. Poulsen and co-workers proposed

that the overall frustule targeting mechanism in *T. pseudonana* could rely on the post-translational modification of lysines and serines, rather than on a specific peptide sequence (Poulsen et al., 2013); they proposed also that the mechanism regulating the trafficking of the SDV-targeted protein could be dependent on protein aggregation, as the silaffins are known to create supramolecular assemblies given their zwitterionic nature. My results from the experiments with CinW2 and Sil3 are – at least in theory – compatible with the first hypothesis. If a similar set or rules exists in cingulins as well, and if single-amino acid modifications can influence the protein sorting and trafficking of the related protein fragment, then it is possible that the presence of one or more KxxK motifs is the reason why some of the studied Sil3- and CinW2-derived modified fragments are still targeted to the frustule; these constructs in fact possess at least one KxxK motif and these motifs could be the target site of specific post-translational moieties addition that function as targeting signals. My hypothesis would also explain the differences in the frustule-targeting properties of the isolated PLC from CinW2: the structure of the KxxK motifs, the amino acids flanking the motifs, and the distance from other similar motifs could influence the pattern of modification, and therefore the silica precipitation properties and/or the targeting activity of the motif.

4.2 PROMOTERS STUDIES

As indicated by the available expression data (Shrestha et al., 2012; **Figure 11**), not all the genes encoding for frustule-associated proteins of *T. pseudonana* are regulated in a similar way, despite their involvement in the same process (*i.e.*, the formation of the cell wall). On the contrary, there are remarkable differences between some of the known genes involved in the bio-silicification process, both regarding the timing of the expression and the relative abundance of the transcripts. The synthesis of the frustule is a complex and massive biological process, most likely involving several cell compartments and organelles (Fattorini and Maier, 2021). It appears logic, therefore, that proteins necessary for the formation of the valve would be synthesized in the necessary amount and in the proper moment of the cell cycle; same is true for the proteins needed only for the girdle bands. Proteins which are instead needed ubiquitously in respect to the bio-silicification process, would be synthesized somehow constantly or, at least, for as long as the frustule-synthesizing machinery is at work.

Available data regarding the comparison between the *in vivo* localization of proteins using the native promoter or a non-native, constitutively expressed, one, has shown that

both for Sil3 (Poulsen et al., 2013; Poulsen et al., 2007; Scheffel et al., 2011) and for CinY2 (Kotzsch et al., 2016) the frustule localization obtained is altered when using a non-native promoter respect to native one (reviewed in Fattorini and Maier, 2021).

If a certain protein depends on internal targeting sequences for the intracellular trafficking, then observing a change in the frustule-localization of a protein after modifying the transcriptional regulation of its gene indicates that these unknown targeting sequences are functioning to transport the protein to the frustule, but that the pathway taken it's not the correct one, or that the cellular machinery underlying the transport works slightly differently. This is for example the case of the Sil3 protein: when the gene is expressed using the CinY2 promoter, it is not transported to the valve but only to the girdle; the same protein, when using the CinW2 promoter, is instead transported to the frustule in the same way as it is for the native Sil3 protein, as if *sil3* and *cinW2* promoters could be functionally exchanged. An explanation could be given by the fact that the *cinw2* gene is not regulated (Shrestha et al. 2012; Fattorini and Maier, 2021) and its constitutive expression during the cell cycle ensures that the CinW2 protein is available at any time it is needed; on the contrary, the *ciny2* gene is strongly regulated. Similarly, the CinW2 protein is transported to its correct frustule location when using the *sil3* promoter instead of the native one; on the contrary, when the *cinw2* gene is expressed upon transcriptional regulation of the *ciny2* gene, the CinW2 is localized in a restricted area of the girdle, different than the native one. The preliminary results on the localization of the CinW2 and Sil3 proteins using the *cinY2* regulatory regions, showed very poor silica targeting and conspicuous amounts of intracellular fluorescent signal.

It appears that the exchange between the *sil3* and the *cinw2* regulatory regions is almost irrelevant in respect to causing modifications in the native frustule localization of the proteins. Instead, in both the cases when the *sil3* or *cinW2* genes' expression is regulated by the *ciny2* promoters, the silica targeting is altered. One hypothesis could explain these observations: when diatoms cells grow, they increase the valve-to-valve distance by synthesizing new girdle bands, but the total number of girdle bands in each cell is variable (Hildebrand et al., 2006); in any case, before mitosis and before the synthesis of the new valves, the formation of new girdle bands is stopped; thus, to proceed through the cell cycle and start a new mitotic division there must be some internal cue that triggers the switch from growth to division. The *in vivo* localization of CinY2 always showed that the protein is located in the last girdle bands (the pleural bands), no matter the length of the cell and therefore the total number of girdle bands; so it might be possible that the

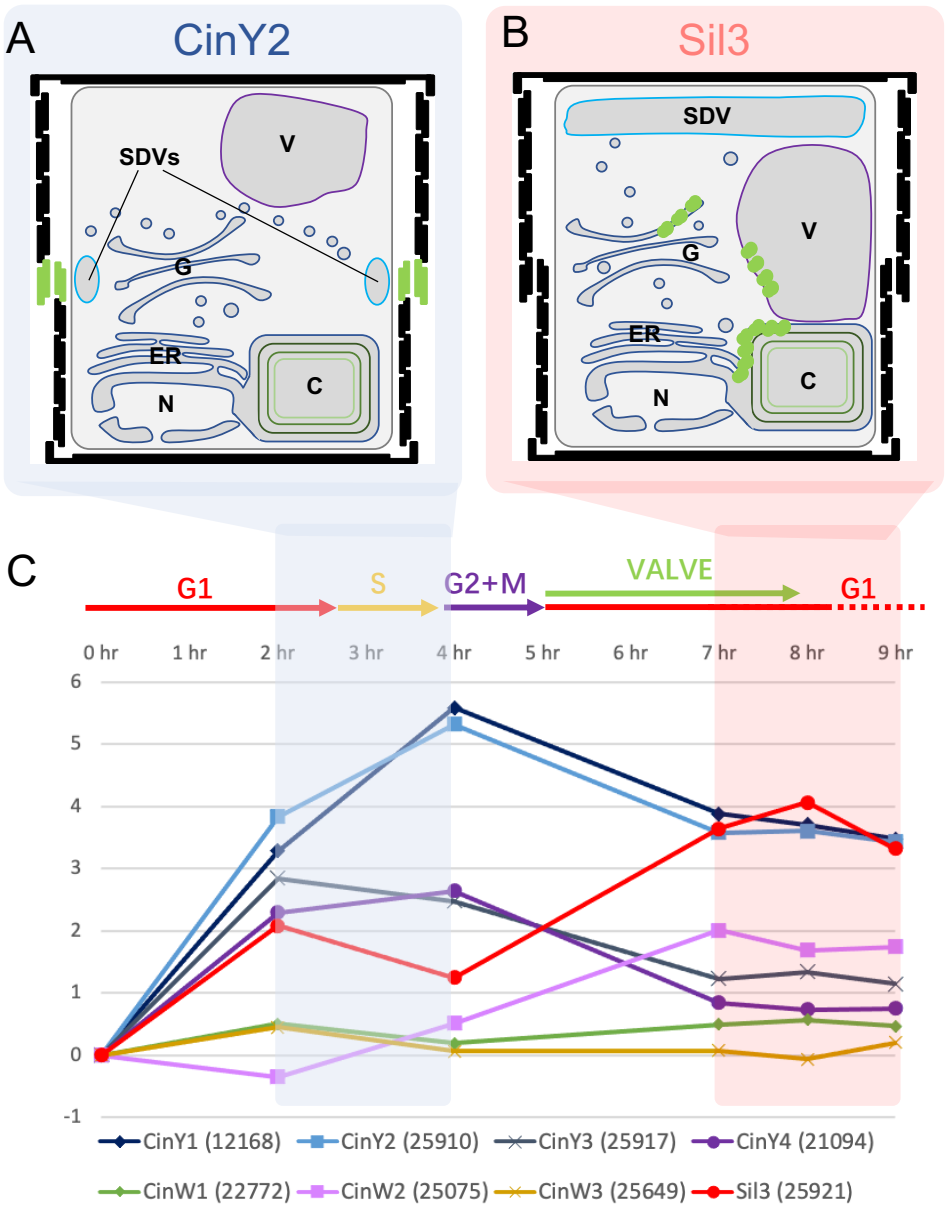


Figure 29 Promoter exchange and targeting in CinY2. This figure shows the hypothetical relationship between the expression of the *cinY2* gene and the targeting of the corresponding protein. Figure **A** shows the frustule localization and cell organization when the *cinY2* gene is expressed using its native promoter and terminator; figure **B** shows instead the frustule localization and cell organization when the *cinY2* gene is expressed using the *sil3* promoter and terminator (the eGFP localization is just an example of an intracellular mistargeting). The plot in **C** is identical to the one shown in **Figure 11**; the colored areas highlight the peaks in the expression of the *cinY2* and *sil3* genes (in light blue and light red, respectively). The organelles are labelled as in **Figure 28**.

synthesis, the transport, and the silica embedding of the CinY2 protein is in some way correlated with some cell signal triggering the abovementioned switch from growth to cell division.

In this scenario, CinY2 could exploit a specific intracellular trafficking machinery that is unavailable when the *cinY2* gene was expressed using the *cinW2* or *sil3* regulatory regions (see **Figure 29**). As a result of that, as the gene is expressed at the wrong time (**Figure 29.B**), the protein is not able to find its way to the frustule and eventually gets

degraded (this would also explain why the transformants screening was so difficult). This speculative scenario might be compatible with a sort of check-point role for the protein CinY2. If true, it would be interesting to test if the absence of Sil3 and Sil1/2 from the pleural bands, where CinY2 instead is localized, is related with it.

In respect to the study of the promoters, my preliminary work on the isolation of the *cis*-acting region of the *Sil3, cinY2 and cinW2* genes, will be useful for the further steps of this approach, such as determining the differential expression of the *egfp* gene upon regulation of the expression from the truncated versions of the upstream region. Evaluating the level of expression only by analysing the amount of fluorescence can be biased by differences in the amount of protein synthesized that can – for example – depend on the number of genomic integration events following transformation.

4.3 SIL1 AND SIL2

My results suggest two possibilities regarding the origin of the two almost identical proteins Sil1 and Sil2: i) they are synthesized as a result of the expression of two alleles of the same gene, resulting in two distinct, functional proteins (the two native proteins have been identified upon their isolation from alive cells-extracted frustules; Poulsen & Kröger, 2004); ii) alternatively, a gene duplication event occurred involving a genomic region of unknown extension but including the one that I have managed to amplify. Both scenarios are compatible with the results obtained by a single PCR amplification using the primer pair that I have designed on the available *in silico* data.

It is difficult to hypothesize why *in vivo* localization of Sil1 obtained during my experiments is different respect to the one previously published. The only other difference rather than the use of native promoter/terminator regions instead of the light-induced *fcp* promoter/terminator, is that the previous experiments used a mix of genomic and cDNA sequences to generate the transformation vectors (Poulsen et al., 2013). I am tempted to say that the use of a constitutively expressed promoter should result in a less restricted protein distribution on the frustule, rather than a more restricted (if no other factors are changed).

4.4 CONCLUSIONS

To summarize, the transport of frustule proteins and their intracellular trafficking and targeting is a complex process most likely regulated by a combination of factors, including the timing of their expression, internal signal sequences and post-translational amino acid

modifications. What is not yet completely clear is exactly what the targeting sequences are. Despite no clear, unique girdle band-, valve-, or neither simply a general SDV-targeting motif was identified, I still managed to obtain insights into the targeting of the studied proteins. Lysine-enriched regions in CinW2 show a similar behavior as the PLC in Sil3, with different silica targeting properties depending on their structure; but additionally, they appear to be sufficient but not necessary targeting motifs. The changes in the transcriptional regulation of the genes have a substantial effect only on certain proteins, depending on the promoters that have been used for the permutation.

4.5 FURTHER STEPS

To further investigate the topic, several approaches could be taken. For example, in respect to the search and study of frustule-targeting sequences, the targeting properties of PLC and LERs could be used to test whether these sequences can be added at the N-terminal or C-terminal of full length proteins (for example Sil3 or Sil1) to change their specific targeting; otherwise, another approach could be to delete the remaining KxxK motifs from some of the truncated versions of CinW2 in order to understand the minimum amount of KxxK motifs needed to ensure the targeting (if they are indeed needed). For the study about the effects of transcriptional regulation on the targeting, instead, other protein pairs could be chosen to have their promoters exchanged, such as proteins targeted only to the valve with proteins targeted only to the girdle. For the investigation regarding the intracellular trafficking, a screening of the protein database of *T. pseudonana* for v-SNARE and t-SNARE homologs could result in a set of candidates for co-localization studies using SDV-membrane proteins (such as SAP3 or Sin1), useful for searching transport vesicles targeted to the SDV; otherwise, another approach could be to use compounds known to inhibit specific steps of the vesicle trafficking (*e.g.*, vesicles fusion, vesicle coat assembly, cytoskeleton fibers-based transportation) and then isolating organelle fractions from clones expressing recombinant frustule proteins.

5. MATERIALS AND METHODS

5.1 ALGAL CULTURES

5.1.1 *T. PSEUDONANA'S* STRAIN USED

The diatom strain used in this work is the Thalassiosira pseudonana (Hustedt) Hasle et Heimdal strain CCMP 1135. This strain was isolated for the first time in 1958 from Moriches Bay (Long Island, New York) and was kept in continuous culture since then at the Centre for Culture of Marine Phytoplankton (Link 2). This strain is the same strain that was used for sequencing the genome of the species in 2004 (Armbrust et al., 2004; Link 3).

5.1.2 CULTURING AND MAINTENANCE

To determine the cell density of the cultures, cell counting was performed using a Thoma counting chamber (**Figure 30**; Link 4). Approximately 8 µL of a thoroughly mixed algal culture were used for the counting; first, a coverslide was positioned on top of the two counting grids of the chamber, then the culture aliquot was pipetted, from the side of the coverslide, on top of one grid; after a couple of minutes, the counting was performed according to the chamber's specifications (see legend of **Figure 30** for details).

The *T. pseudonana* cell lines have been maintained using liquid growth medium or agar plates, depending on the intended use: cells for immediate use (*e.g.*, DNA/RNA/biosilica extraction, transformation, microscope imaging) have been grown on NEPC medium (see section 5.1.3 for details, and **Table 3** for the corresponding recipe), whereas cells for long- and mid-term maintenance have been kept on NEPC-based Bacto agar 1.5% (w/v) plates, with the agar plates incubated in the growth chamber having the agar layer at the bottom). Periodically, to reduce bacterial contamination, the cultures have been treated with a mix of ampicillin and streptomycin (500 µg/mL and 50 µg/mL, respectively).

The agar plates have been prepared as follows: since the Bacto agar is apparently sensitive to the chemical components of the NEPC medium, to make 1 L of NEPC-agar solution, I have prepared 500 mL of two times concentrated NEPC medium and 500 mL of two-times concentrated agar-dH_2O solution; after autoclaving, the two solutions were mixed together, and the obtained one-time concentrated mixture of NEPC medium, water and agar, was kept stirring until the solution was cooled down; then I have added the

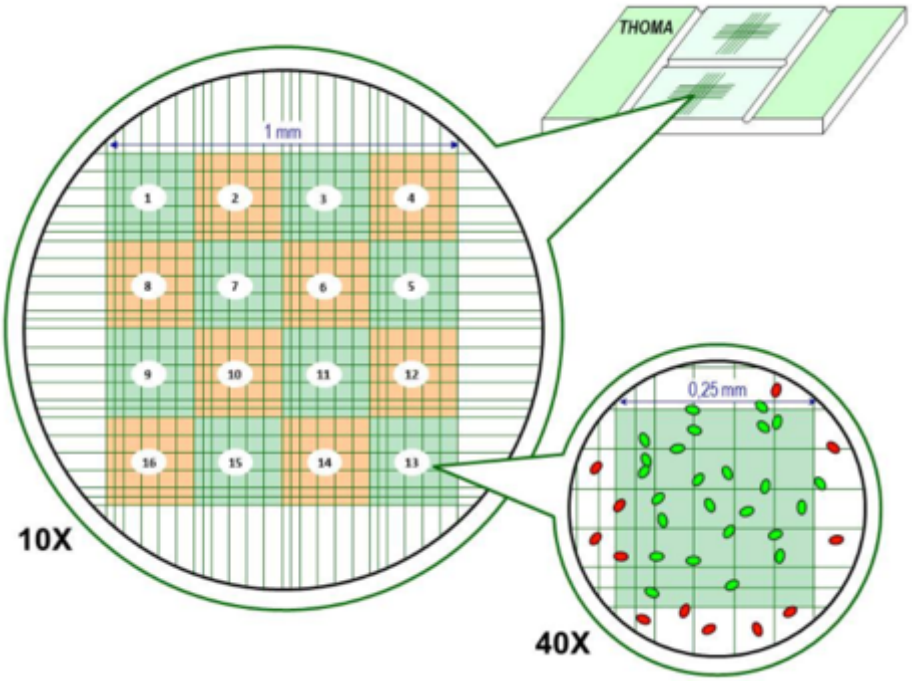


Figure 30 Schematic representation of the Thoma counting chamber. The large circle shows a magnified view of one of the two counting grids; in the small circle is shown one of the 16 squares used for counting the cells located on the grid, The total number of cells in all 16 squares is then multiplied by 10,000 and the resulting number represent the estimated number of cells per mL of culture.

supplements and, if needed, the antibiotic; the two medium-agar solution could then be poured.

Transformants cell lines have been supplemented with the antibiotic nurseothricin (NAT or clonNAT) as selection marker, at the final concentration of 150 µg/mL (Link 5). The light regime consisted in 14 hours of light and 10 hours of dark, at an irradiance of approximately 80-90 µE m^{-2} s^{-1} and temperature of 18 °C.

Selected clones were chosen for additional long-term storage upon cryopreservation. Approximately 50 mL of exponentially growing culture (roughly $1.5x10^6$ cell/mL) were centrifugated at 3200 g for 12 min; 4.4 mL of the surnatant were used to gently resuspend the pellet; 600 µL of DMSO were added to the solution; 1.2 mL of the solution were transferred into two screw-cap cryo-vials for each clone; the tubes were then transferred into a Corning Coolcell freezing box filled with isopropanol and pre-cooled at 4 °C; the box was transferred for three hours at -80 °C, then transferred into storage boxes and kept at -80 °C.

5.1.3 MEDIUM PREPARATION

The medium used for cultivation was prepared following a slightly modified version of the North East Pacific Culture Collection (NEPCC) original medium recipe which is available

from the website of the Canadian Center for the Culture of Microorganisms (Link 6). The exact components, their quantities, and the steps of the medium preparation are listed in Table 3.

Table 3 Recipes of the growth media used.

NEPC MEDIUM RECIPE (1 L)		
Salts		
NaCl		20.8 g
$MgCl \cdot 6H_2O$		9.6 g
Na_2SO_4		3.5 g
Dissolve in 900 mL of dH_2O and add:		
1 M $CaCl_2$		9 mL
3 M KCl		2.6 mL
Enrichment stocks		
ES1	$NaNO_3$ (47 g/L)	1 mL
ES2	$SrCl_2 \cdot 6H_2O$ (21 g/L)	1 mL
ES3	$Na_2SiO_3 \cdot 9H_2O$ (30 g/L)	1 mL
ES5*	$Na_2EDTA \cdot 2H_2O$ (1.86 g/L) $ZnCl_2$ (32.7 mg/L) $CoCl_2 \cdot 6H_2O$ (20.2 mg/L) $Na_2MoO_4 \cdot 2H2O$ (126 mg/L) $MnCl_2 \cdot 4H_2O$ (475 mg/L)	1 mL
ES6	H_3BO_3 (25 g/L)	1 mL
ES7*	$Na_2EDTA \cdot 2H_2O$ (3.72 g/L) $FeCl_3 \cdot 6H_2O$ (1.76 g/L)	1 mL
ES8	$Na_2SeO_3 \cdot 5H_2O$ (47 g/L)	0.1 mL
Adjust to pH = 8.0-8.2, autoclave and then add:		
Supplements**		
S1	NaF (3 g/L) KBr (85 g/L)	1 mL
S2***	Na_2- Glycerophosphate$\cdot H_2O$ (6 g/L)	1 mL
S3	Thiamin·HCl (200 mg/L) Vitamin B12 (4 mg/L) Biotin (2 mg/L)	1 mL

NEPC-agar plates 1.5 (w/v) (1 L)	
Prepare:	
2x NEPC medium	500 mL
2x agar-dH_2O (15 g/L)	500 mL
Autoclave, mix and cool down	
Add supplements/antibiotics	
Pour on plates and let gelify	
LB medium (1 L)	
Tryptone	10 g
Yeast extract	5 g
NaCl	10 g

* Prepare in a separate solution and adjust to pH = 6
** Sterilize solutions by filtration through 0.22 µm filter;
*** Dissolve in a saturated solution of $NaHCO_3$).

5.1.4 BIOSILICA EXTRACTION

To isolate intact frustules of *T. pseudonana*, I followed a slightly modified protocol respect to published one (see for example Poulsen et al., 2013): roughly 4 mL of an exponentially growing culture (approximately $1.5 \cdot 10^6$ cell/mL) were centrifugated at 3200 g for 10 min; the surnatant was discarded and the pellet was resuspended with an extraction buffer composed of 100 mM EDTA (pH= 8) and 1 mM PMSF; after the pellet was completely resuspended, SDS was added to the resuspension at the final concentration of 2% w/v; the resuspension was then kept at 55 °C for 1 hour; after this step, the resuspension was centrifugated at 3200 g for 10 min; the surnatant was discarded and the pellet (containing extracted cell walls and degraded membranes and intracellular content) was extracted

again until it was colourless (a greenish colour of the pellet indicates that the extraction is not completed, and that there is still intracellular components). After the extraction, the pellet was washed three times with 1 mL of distilled water, once with 1 mL acetone 80% and again three times with 1 mL of distilled water. The final pellet was resuspended in 20 µL and it was ready for the observation with the microscope.

5.1.5 BACTERIAL CELL CULTURES AND TRANSFORMATION

The *Escherichia coli* TOP10 competent cells aliquots were previously stored at -80 ℃ in LB medium (see **Table 3**) with 50% of Glycerin; before use, each 100 µL aliquot was placed on ice until completely thawed. Depending on the intended purpose (*e.g.*, for mini or for midi preparations), a DNA mixture or bacterial cryo-stock inoculum were added to the thawed aliquots.

For plasmid mini preparations, the 50 µL of DNA-bacterial solutions were mixed and placed in a thermal bath at 42 °C for 42 sec; immediately after this step, the aliquots were placed on ice for approximately 5-10 min; the aliquots were then plated on LB-based agar plates (1.5% w/v, supplemented with 50 µg/mL ampicillin); the plates were let dry shortly and incubated at 37 °C for approximately 14-18 hours; on the following day, colonies of transformants bacteria were picked using toothpicks (generally 6-10 colonies/construct, depending on the total number of colonies on the plate), and transferred into sterile glass tubes containing 3 mL of LB medium supplemented with µg/mL ampicillin; the tubes were then incubated for 14-18 hours shaking at approximately 150 rpm. The bacterial cultures were then processed as described in section 5.2.6.4.

For midi preparations, two methods of inoculation were followed: for the first method, a very small amount of plasmid DNA (approximately less than 100 ng) was added to the 50 µL thawed aliquot of bacterial cells, and the bacterial aliquot was transformed and processed as previously described (with the difference that only one colony was then taken for the inoculation). For the second method instead, an already existing LB-glycerin bacterial stock was removed from the -80 °C and, immediately after, it was picked using a sterile pipette tip, that was then used for the inoculation. In both cases, the inoculum was transferred into a sterile Erlenmeyer flask containing 50 mL of LB medium supplemented with 50 µg/mL of ampicillin; the flasks were then incubated as previously described. The midi extraction procedure is described in section 5.2.6.4.

5.2 MOLECULAR BIOLOGY

5.2.1 DNA AND RNA EXTRACTION

For the DNA this protocol has been followed: approximately 100 mL of exponentially growing cells (~1.5 million cells/mL) were harvested at 10000 g for 12 min at 4 °C; then the surnatant was discarded, and 400 mg of sterile silica beads were added to the pellet, together with 500 µL of TE buffer and 500 µL of phenol; the mixture was vortexed for 85 sec at 30 Hz three times, placing the sample on ice between the vortex steps; the mixture was then centrifugated at 20000 g, and the aqueous phase was transferred into a new 1.5 mL Eppendorf tube; then 50 µL of 3 M sodium acetate (pH = 5), 1 mL of 96% ethanol and 2 µL glycogen were added to the solution, that was incubated overnight at -20 °C; the overnight samples were then centrifugated at 20000 g for 30 min at 4 °C; the surnatant was discarded and the whitish pellet at the bottom of the tube washed with 1 mL of 70% ethanol (20000 g for 5 min at 4 °C); the surnatant was discarded and the pellet dried in a vacuum-driven exicator; the pellet was then resuspended with 50 µL of warm, sterile, distilled water and kept on an heating block at 55 °C for 30 min; the DNA concentration of the sample was then measured and the DNA mixture aliquoted into aliquots having ~100 ng of DNA/µL, ready to be use as templates for the PCR reactions; the DNA was then stored at -20 °C.

This protocol has been instead followed for the RNA extraction: all equipment (bench, centrifuges, gloves, pipettes and tubes) have been thoroughly wiped with RNAse Zap to avoid degradation of the RNA from RNAse enzymes; 150 mL of fresh culture were harvested at 10000 g for 12 min at 21 °C; the surnatant was discarded and 1 mL of Trizol reagent was added to the pellet, together with approximately 400 mg of silica beads; the mixture was shaken at 1400 rpm on a heating block for 10 min at 60 °C and then vortexed three times for 30 sec; the mixture was then centrifugated at 10,000 g for 5 min at 4 °C and the surnatant transferred into a new 2 mL Eppendorf tube; 200 µL of chloroform were added to the mixture, the tube was agitated vigorously and the samples were then placed at room temperature for 15 min, then the samples were centrifugated at 12000 g for 15 min at 4 °C (this step with chloroform was done two times); 500 µL of aqueous surnatant were transferred into a new tube, an equal amount of isopropanol was added and the samples were kept at room temperature for 10 min; the samples were then centrifugated at 12,000 g for 10 min at 4 °C and the surnatant discarded; 1 mL of 70% ethanol was added and the samples to wash the pellet; the samples were centrifugated at 7,500 g for 10 min at 4 °C; the surnatant was then discarded and the pellet dried in a vacuum-driven

exicator until complete evaporation of the ethanol; the RNA pellet was then resuspended with 30 µL of sterile, DEPC water; the samples were kept at 55 °C for 30 min to better dissolve. The RNA concentration in the solution was measured and then the samples were then stored at -80 °C.

5.2.2 cDNA SYNTHESIS

Before synthesizing the cDNA from the RNA extracted, a DNAseI treatment (Thermo Fischer Scientific) was performed to remove any residual genomic DNA contamination from the RNA solution. A small aliquot of the RNA extract containing approximately 1 µg of RNA was mixed with 1 unit of the DNAseI enzyme, with 1 µL of 10x $MgCl_2$ buffer and sufficient sterile DEPC-water to reach 10 µL of total volume; the mixture was kept at 37 °C for 1 hour and then stopped by adding 1 µL of a 50 mM solution of EDTA and by keeping the mixture at 65 °C for 10 min; 2-4 µL of the DNA-free RNA extract were loaded on a agarose gel to check for absence of contamination. Approximately 500 ng of the DNA-free RNA were used to synthesize the cDNA, using the Fast Gene Scriptase II cDNA synthesis kit (Nippon Genetics) and following the manufacturer's instructions.

5.2.3 POLYMERASE CHAIN REACTION (PCR)

Different sets of thermocycling conditions have been used, depending on the purpose of the reaction and on the characteristics of the primer pairs used (**Table 5**).

The PCRs needed for of the molecular cloning or for any mutation, insertion or deletion, have been made using the NEB 2x Q5 site-directed mutagenesis (SDM) kit, which is provided with a 2x concentrated Q5 polymerase master mix. Instead, for the colony PCRs I have used the 2x super master mix from Bio Tools, or the 2x One*Taq* master mix from NEB (see section 5.2.6.7). The PCR mixes has been prepared as indicated in **Table 4**. For SDM reactions, I have generally prepared a 10 µL PCR mix.

When cloning a gene of interest from gDNA (and for this purpose I have generally always used traditional cloning techniques), I have usually scaled the reaction mix volume up to 25 µL, to have a higher DNA yield after gel extraction and purification. For the Gibson Assembly reaction, I have generally scaled the reaction mix up to 50 µL, to obtain an even higher concentration of DNA in the eluate respect to that needed for the traditional cloning technique, given that the total volume of the Gibson ligation mix is only 5 µL

Table 4 Composition of the PCR mixes prepared for the experiments

MASTER MIX PREPARATION			
	Q5 PCR MIX (NEB)	2x SMM (Bio Tools)	OneTaq 2X master mix (NEB)
Sterile dH2O	Up to 10 µL	Up to 20 µL	Up to
primer fw (5 µM)	1 µL	0.5 µL	0.4 µL
primer rv (5 µM)	1 µL	0.5 µL	0.4 µL
DNA template	5 – 20 ng	70 – 100 ng	70 – 100 ng
Enzyme mix	5 µL	10 µL	5 µL
TOTAL	10 µL	20 µL	10 µL

The thermocycling conditions of the PCR reactions I have performed are indicated in **Table 5**. Depending on the template, or on the primer pair used, the settings might have been adjusted within the indicated range of values.

Table 5 Thermocycling conditions of the PCR performed for the experiments

THERMOCYCLING SETTINGS						
	2x Q5 HF (NEB)		2x SMM (Bio Tools)		2x (NEB)	
	Temp	T	Temp	T	Temp	T
Initial denaturation	98 °C	30 sec	94 °C	5 min	95 °C	30 sec
Denaturation	98 °C	10-30 sec	94 °C	20 sec	95 °C	15 – 30 sec
Annealing	variable	30 sec	variable	30 sec	variable	15 – 60 sec
Extension	72 °C	20-30 sec/Kb	72 °C	30 sec/Kb	68 °C	1 min/Kb
Number of cycles	25 - 40		35			
Final extension	72 °C	2 min	72 °C	5 min	68 °C	5 min

5.2.4 AGAROSE GEL ELECTROPHORESIS AND EXTRACTION

To check the results of a PCR amplification, of an enzymatic digestion, or of an RNA/DNA extraction, the nucleic acid solution obtained has been analysed by means of agarose gel separation with an electrophoretic apparatus.

The agarose gel was prepared using 150 mL of 1x TBE buffer (see **Table 6**); in general, the agarose concentration in the gel can be chosen depending on the size of the fragments to separate during the run, but a 1% w/v agarose gel was usually used for my purposes (thus, 1.5 g agarose/150 mL buffer); 4.5 µL of fluorescent dye Roti-Safe gel stain (Carl Roth) were added to the solution for the staining of the nucleic acids.

The nucleic acid mixture to be analysed was prepared by mixing 2 µL of 6x DNA loading buffer (see **Table 6**) with 10 µL of nucleic acid/distilled water solution (when the nucleic acid concentration is high, then only 1-2 µL are sufficient for the analysis, and the rest can be added as water). This mix can be scaled up to higher volumes; this was frequently the case when preparing a Gibson Assembly reaction, for which higher yield of DNA are useful and therefore larger wells are needed; I have obtained better results during the extraction procedure when I have used single, larger wells to accommodate the

50 µL PCR product for the electrophoretic run, respect to when I have used multiple, smaller wells. The gel separation was conducted using TBE as running buffer and applying a voltage of ~140 V. The size of the separated bands was compared to the fragments of a specific marker; depending on the size of the expected fragments, I have used the ladder 1 kb or the ladder 100 (see **Figure 31**).

Table 6 Concentration of the components in the TBE buffer and in the 6x loading buffer

ELECTROPHORESIS BUFFERS			
TBE buffer		6X loading buffer	
Tris-HCl (pH = 8.8)	1 M	Urea	4 M
Boric Acid	0.83 M	EDTA	50 nM
EDTA	10 nM	Saccharose	50 w/v
		Bromophenol Blue	0.1 % w/v
		Xylene Cyanole	0.1 % w/v

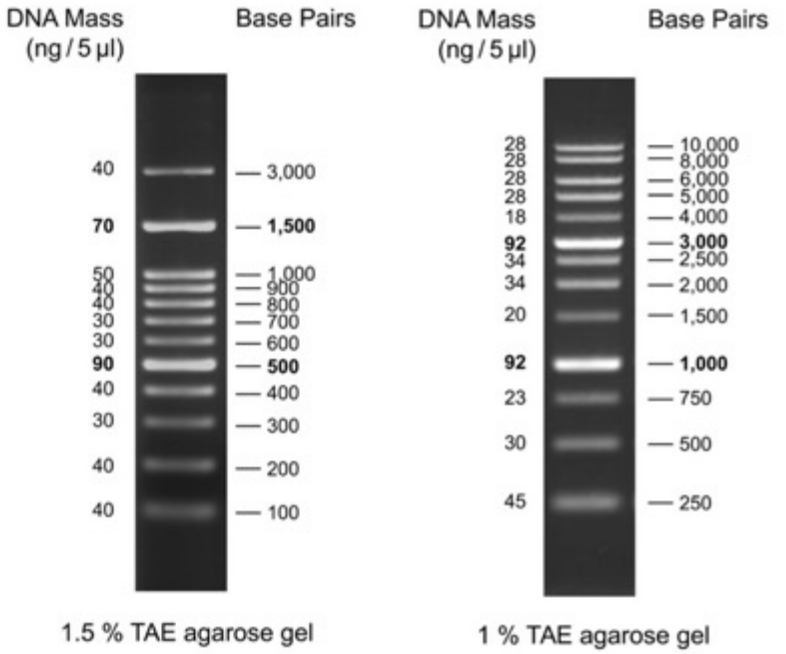


Figure 31 The images show the two gene markers used to compare the size of the fragments that have been separated on the agarose gel. On the left is shown the ladder 100bp and on the right is shown the ladder 1kb.

To extract and purify the PCR products separated on the agarose gel, I have used the Zymoclean Gel DNA Recovery Kit (Zymo Research), following the manufacturer's instructions. The final elution step has been done using 12 µL of warm (75 °C); sterile dH_2O.

5.2.5 DIGESTION AND LIGATION

To cut out the fragment(s) of interest from DNA molecule(s) (*i.e.*, vector or PCR product), I have used restriction endonucleases; the enzymes have been chosen according to the

restriction sites that were inserted via the primer pair at the 5'- and 3'-ends of the PCR product during the amplification step. An enzymatic digestion using that specific pair of restriction enzymes has been setup for both the donor and recipient molecules, generating therefore fragments bearing compatible ends that could be used for the ligation. The digestion mix has been prepared as indicated in **Table 7**; the buffer used and the corresponding volume, together with the reaction time and temperatures, have been chosen depending on the enzymes used, according to the manufacturer's recommendations.

The ligation between the digested fragments have been performed by maintaining a 1:1, 1:3 or 3:1 ratio between the number of vector's ends and insert's ends, depending on their relative size. To setup the ligation reaction, I have first calculated the molecular weight of each fragment and then determined the picomoles contained in 1 µL of DNA solution. Then I have used the calculated values and the established vector:insert ratio to determine the volume needed for the ligation reaction.

The reaction mix has been prepared by mixing the calculated amounts of DNA solution, with 1 µL of T4 ligase buffer, 0.5 µL of the enzyme DNA T4 ligase, and bringing the volume up to 10 µL with sterile distilled water. To insert the PCR amplified products into the pJET1.2 shuttle vector, the ligation reaction has been setup by eluting the PCR product with 6.5 µL of sterile dH_2O, to which 0.5 µL of DNA T4 ligase, 0.5 µL of pJET1.2 vector (50 ng/µL) and 7.5 µL of 2x reaction buffer, have been added. The ligation reactions have been conducted at room temperature.

Table 7 Components of the mix used to perform a restriction enzyme digestion reaction using two different restriction enzymes

DOUBLE DIGESTION MIX	
Components	Volumes
Sterile dH2O	Up to 10 µL
Enzyme 1	0.2 µL (~2 U)
Enzyme 2	0.2 µL (~2 U)
Buffer	*variable*
DNA	*variable* (500 – 1000 ng)
TOTAL	10 µL

5.2.6 MOLECULAR CLONING

For the creation of the transformation vectors needed for my experiments, I have used three different strategies: sticky ends ligation-based traditional cloning, Gibson Assembly (GA), and site-directed mutagenesis (SDM). All primers pairs cited in this section are listed in **Table 9** and **Table 10** of the supplementary information.

5.2.6.1 TRADITIONAL CLONING

As general guideline for this approach, the purified PCR products of interest have been inserted into the shuttle vector pJET1.2 via blunt-end ligation (the NEB Q5 polymerase generates blunt ends), following the manufacturer's instructions. The ligation mixture was then used to transform *E. coli* competent cells. The cells were then processed as described in section 5.1.5. The pJET1.2 plasmids validated by sequencing and the recipient plasmid were then digested with the appropriate restriction enzyme, to generate the desired final vector (see also section 5.2.5).

This type of cloning strategy (*i.e.,* sticky ends ligation) has been used to generate the parental plasmid pTpNAT-MCS, with e procedure that I will briefly describe here: this plasmid was designed to have a multiple cloning site (MCS) and the NAT cassette, made of the *nurseothricin acetyl transferase* (NAT) gene (that confers resistance to the antibiotic nurseothricin) flanked by the *fcp* (*fucoxanthin chlorophyll a/c binding potein*, a light-induced gene) promoter and terminator (Poulsen et al., 2006). This plasmid was created as follows: using as templates the vector pPhaNR, that contained the desired MCS (Stork et al., 2012), and the vector 937, that contained the desired NAT cassette (Kotzsch et al., 2016), I have amplified the MCS from pPhaNR with the primers pPhaT1MCSfw and pPhaT1MCSrv (that insert a NdeI and AccIII r.s., respectively), and the NAT cassette from vector 937 using the primers Fcp9promFw and Fcp9termRv2 (that insert an AccIII and XhoI r.s., respectively); with these primers, the extension of the promoter and terminator regions were reduced to 500 bp and 400 bp, respectively. The vector pPhaNR and the amplified MCS have been digested using the restriction enzymes NdeI and AccIII, creating therefore two fragments bearing compatible sticky-ends. The two fragments have been ligated together, creating an intermediate vector having the desired MCS, but also the undesired zeocyn resistance cassette from pPhaNR. The intermediate vector and the amplified NAT cassette have therefore been digested with the restriction enzymes AccII and XhoI, generating two fragments bearing compatible sticky-ends. The ligation of the two fragments generated the vector pTpNAT-MCS. This vector has been used as parental plasmid for all the other constructs, as indicated in **Table 9**.

5.2.6.2 SITE-DIRECTED MUTAGENESIS (SDM)

The site-directed mutagenesis (SDM) technique has been used to generate all mutations, truncations, deletions, insertions, and all pTpSil3-, pTpW2- and pTpY2-derived constructs,

including the isolated LERs or PLC (in this case the constructs were generated via two sequential deletions); the primers used are indicated in **Table 9**).

As a general description: the PCR products of interest were checked on an agarose gel, and the desired band extracted and purified (see section 5.2.4); the eluate was then quantified at the spectrophotometer, and approximately 5–15 ng of purified PCR product were added to the mix of **Table 8**, to perform the reaction using the KLD enzyme mix (Kinase, Ligase, DpnI): this reaction generates the final, circular plasmid, and degrades any remaining molecule of the template used for the PCR. The mix was kept at room temperature for 5 min, and then used to transform 50 µL of *E. coli* competent cells (as described in section 5.1.5.

Table 8 Components of the KLD mix from the NEB site-directed mutagenesis kit.

KLD MIX	
Components	Volume
dH20	Up to 5 µL
KLD buffer	0.5 µL
KLD enzyme mix	0.5 µL
PCR product	5 – 20 ng
Total	5 µL

5.2.6.3 GIBSON ASSEMBLY

This cloning strategy has been used to generate the plasmids: pTpCinY2, pTpINV-NAT, pTpINV-Sil3, pTpINV-CinW2, pTpINV-CinY2 and all 6 vectors belonging to the promoter exchange experiment (see section 3.4.1), which is briefly described here: the first step was deleting the CDS from each vector, obtaining therefore a construct in which the START codon of the GFP was positioned immediately downstream to the last nucleotide before the START codon of the native gene (which means the last nucleotide of the upstream region of the native gene); the resulting vectors have been named pTpINV-Sil3 prom-GFP, pTPINV-CinW2 prom-GFP and pTpINV-CinY2 prom-GFP.

As a general description of the method: the purified PCR products of interest were quantified, and the ligation mix prepared using volumes according to the values calculated with a Gibson Assembly excel sheet available online (Link 7); then, the protocol was concluded following the manufacturer's instructions. The ligation mixture was then used to transform E. coli competent cells (as described in section 5.1.5).

5.2.6.4 PLASMID PREPARATIONS EXTRACTION

Plasmid mini preparations have been processed according to the alkaline lysis protocol, as follows: 1.5 mL of the grown overnight *E. coli* cultures (see section 5.1.5) were transferred into 1.5 mL Eppendorf tubes; the tubes were then centrifugated at 10000g for 2 min at 4 °C; the surnatant discarded and the pellet thoroughly dried from residual medium; 200 µL of buffer P1 were then added to the pellet, that was then resuspended; 200 µL of buffer P2 was added to the solution; the tubes were closed and gently mix upon inversion; after no long than 1 min, 200 µL of buffer P3 and 40 µL of chloroform were added; the tube was thoroughly mixed and placed on ice for 5 min; then the tubes were centrifugated at 20000g for 5 minutes at 4 °C; the aqueous layer containing the nucleic acids was then transferred into a new 1.5 mL tube and 400 µL of isopropanol were added; the solution was mixed and centrifugated at 20000g for 20 min at 4 °C; after this step, a whitish pellet, containing the plasmid DNA, appears at the bottom of the tube; the isopropanol was then discarded and 300 µL of 70% ethanol were added to the tube; the tubes were then centrifugated at 20000g for 10 min at 4 °C; the ethanol was then discarded and the tubes placed to dry in a vacuum-driven exsiccation apparatus until the ethanol was completely evaporated; then the plasmid DNA was resuspended in 40-50 µL of warm, sterile, distilled water and placed on heating block at 55 °C for 30-60 min to completely dissolve the pellet. The concentration of the plasmid DNA preparation was then measured using the Nanodrop spectrophotometer. After measuring the concentration, in order to identify the positive clones, a digestion reaction using restriction enzymes was setup: in case of an insertion of a fragment inside the vector, the restriction enzymes were chosen according to the restriction sites inserted in the PCR product via the primer pair used; in case of deletions, insertions or mutations, the restriction enzymes were chosen to cut the vector into fragments of specific length, allowing the identification of positive clones by electrophoresis gel separation of the resulting digested fragments (see section 5.2.4). The selected candidate clones were further validated by sequencing (see next section).

To obtain high purity of plasmid preparations for the biolistic transformation experiments, plasmid midi preparations have been processed starting from 50 mL cultures (see section 5.1.5); before proceeding with the extraction, ~750 µL of *E. coli* culture was transferred into a properly labelled 2 mL Eppendorf tube with ~750 µL of glycerine; the tube was then vortexed and flash-frozen in liquid nitrogen; then the tubes were transferred into a box and stored at -80 °C. The extraction was performed by using the Nucleobond Xtra midi extraction kit from Macherey-Nagel, following the manufacturer's instructions.

The pellet obtained was then processed in the same way as for the mini preparations, without performing any confirmation digestion at the end.

5.2.6.5 SEQUENCING

After every ligation, deletion, insertion, or mutation performed, but also in general to confirm the nucleotide sequence of a certain DNA fragment, the samples' nucleotide sequences have been obtained and compared to the predicted *in silico* sequences. To do so, we have used the sequencing service of the Macrogen company. Normally, one or two clones from each construct were chosen for sequencing, and the samples were prepared according to the company's requirements: approximately 200-300 ng of plasmid DNA were mixed with 5 µL of a 5 µM solution of the chosen sequencing primer; then the volume was adjusted to 10 µL with sterile distilled water; the sample tube was appropriately labelled and then sent to the company.

5.2.6.6 BIOLISTIC TRANSFORMATION

The *biolistic* (from *biol*ogical ball*istic*) particle delivery-based method of transformation consists in the bombardment of biological tissues with DNA-coated tungsten microparticles. For each construct to be used for transformation, 300 millions of target cells are needed. I have always considered a 30-40% of cell loss during centrifugation (Hildebrand et al., 2007), therefore I have normally centrifugated around 420 million cell per construct. Liquid cultures of *T. pseudonana* in the exponential phase of growth (~$1.5x10^6$ cells/mL) were harvested by centrifugation (3200g for 12 min at 21 °C); the surnatant was discarded and the algal pellet resuspended using fresh, sterile NEPC medium (approximately 900 µL were used for each construct); the algal suspension was then plated onto NEPC-agar plates (approximately 300 µL per plate) into a circular area of approximately 5 cm of diameter; the plates were then let dry under the sterile bench; to prepare the microparticles, ready-to-use aliquots of M10 tungsten microparticles (0.7 µm of diameter) contained in 1.5 mL Eppendorf tubes were thawed on ice; 5 µg of plasmid DNA were then added to the microparticles and the mixture was shorty vortexed; then 50 µL of $CaCl_2$ were added to the tube and the mixture was shortly vortexed; then 20 µL of 0.1 M spermidin were added and the mixture was vortexed for three minutes; after this step, the mixture was briefly centrifugated for 10 sec to precipitate the microparticles; the surnatant was carefully discarded and the microparticles were washed with 250 µL of RNA-grade 96% ethanol and vortexed for 1 min; then the mixture was briefly centrifugated

again to precipitate the microparticles; the surnatant was again discarded and the microparticles resuspended in 50 µL of 96% ethanol; after this preparation the microparticles are stable for 1 hour and can be immediately used for the transformation experiment.

The particle delivery device (*gene gun*), its component and the bench where the device was located were thoroughly wiped with ethanol; the microparticles were loaded on the microcarrier disk (three microcarrier disks per construct): each microparticles aliquot was split into three 15 µL fractions (one for each microcarrier disk), and loaded on the disks; when the ethanol was completely evaporated from the microparticle mixture, a small metal grid (the stopping screen) was flamed for sterilization and placed on the support, immediately below the microcarrier disk; once sealed by a metal ring, this component was inserted into the gene gun; before shooting, a rupture disk conceived to break at 1550 psi was inserted at the end of the gas connector, where the helium gas enters the gene gun chamber; the agar plate with the cell was positioned on the second shelf from the top and the vacuum was created inside the chamber (~27 psi); the helium gas was then pumped inside the connector and, once reached the rupture disk breaking pressure, the air flow projected the microparticles against the cells on the plate; immediately after the bombardment, the cells were gently scraped using a sterile spatula and resuspended into 100 mL of fresh, sterile NEPC medium. The cells were then kept for 24 hours under continuous light and at 21 °C; the following day they were harvested (3200 g for 12 min at 21 °C) and plated (~8 million cells per plate) on 10 NEPC-agar plates supplemented with 150 µg/mL of NAT. The plates were kept under continuous light for two weeks; after that, single colonies normally had appeared that could be transferred and observed.

5.2.6.7 COLONY PCR

To verify that the whole fragment of interest contained in the transformation vector had been integrated, all the clones that showed eGFP expression under the confocal microscope had been tested and validated via colony PCR (except for the construct CinY2 TR-C). The DNA used as template was obtained as follows: a small amount of cell culture growing on plate was directly picked with a sterile inoculation loop and then resuspended in ~40 µL of sterile, distilled water; the suspension was then boiled at 96 °C for approximately 10 min (at least until the colour of the suspension turned green); the suspension was then centrifugated at 20000 g for 5 min at 4 °C and 2-3 µL were used as template for the PCR reaction indicated in **Table 4**. The primers I have normally used are

indicated in **Table 10**; the forward primers were usually binding in the backbone of the vector, a few bases upstream respect to the MCS of the vector, whereas the reverse primers were binding downstream respect to the MCS. The PCR program used is indicated in **Table 5**.

5.2.7 MICROSCOPY

5.2.7.1 OPTICAL

A Carl Zeiss optical microscope equipped with a 40x objective was used for the cell counting and for general check of the cells' health.

5.2.7.2 CONFOCAL

For the observation of live cells and simultaneous detection of the fluorescent tags and plastid fluorescence, approximately 15 µL of cell suspension were transferred on a microscope slide, covered with a cover slip, and then a drop of Leica Immersion Oil (standard and type "F") was placed on top of the cover slip. An upright Leica DM 6000 B confocal microscope was used for the observation; this device was equipped with a HCX PL APO 63x/1.40 oil PH3 objective, a 100 mW Argon laser which could excite at 458, 476, 488, 496 and 514 nm and a DPSS that can excite at 561 nm. To excite the eGFP and trigger plastid autofluorescence, the 488 nm excitation wavelength was used, whereas two different emission detection channels were setup, one for the eGFP (500-520 nm) and one for the plastid autofluorescence (620-720 nm). To excite mRuby, the 561 nm excitation wavelength was used, whereas the emission detection was in the range between 580-605 nm (Marter et al., 2020). All the acquired images were exported using the Leica LAS AF software, and edited using the ImageJ software (Schneider et al., 2012).

6. BIBLIOGRAPHY

6.1 SCIENTIFIC PUBLICATIONS

Alverson, A. J., Beszteri, B., Julius, M. L., & Theriot, E. C. (2011). The model marine diatom Thalassiosira pseudonana likely descended from a freshwater ancestor in the genus Cyclotella. *BMC Evolutionary Biology*, *11*(1), 125. https://doi.org/10.1186/1471-2148-11-125

Armbrust, E. V. (2009). The life of diatoms in the world's oceans. *Nature*, *459*(7244), 185–192. https://doi.org/10.1038/nature08057

Armbrust, E. V., Berges, J. A., Bowler, C., Green, B. R., Martinez, D., Putnam, N. H., Zhou, S., Allen, A. E., Apt, K. E., & Bechner, M. (2004). The genome of the diatom Thalassiosira pseudonana: ecology, evolution, and metabolism. *Science*, *306*(5693), 79–86.

Ashworth, J., Coesel, S., Lee, A., Armbrust, E. V., Orellana, M. V., & Baliga, N. S. (2013). Genome-wide diel growth state transitions in the diatom Thalassiosira pseudonana. *Proceedings of the National Academy of Sciences of the United States of America*, *110*(18), 7518–7523. https://doi.org/10.1073/pnas.1300962110

Basu, S., Patil, S., Mapleson, D., Russo, M. T., Vitale, L., Fevola, C., Maumus, F., Casotti, R., Mock, T., Caccamo, M., Montresor, M., Sanges, R., & Ferrante, M. I. (2017). Finding a partner in the ocean: molecular and evolutionary bases of the response to sexual cues in a planktonic diatom. *New Phytologist*, *215*(1), 140–156. https://doi.org/10.1111/nph.14557

Bender, S. J., Parker, M. S., & Armbrust, E. V. (2012). Coupled Effects of Light and Nitrogen Source on the Urea Cycle and Nitrogen Metabolism over a Diel Cycle in the Marine Diatom Thalassiosira pseudonana. *Protist*, *163*(2), 232–251. https://doi.org/10.1016/j.protis.2011.07.008

Benoiston, A. S., Ibarbalz, F. M., Bittner, L., Guidi, L., Jahn, O., Dutkiewicz, S., & Bowler, C. (2017). The evolution of diatoms and their biogeochemical functions. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *372*(1728). https://doi.org/10.1098/rstb.2016.0397

Bowler, C., Allen, A. E., Badger, J. H., Grimwood, J., Jabbari, K., Kuo, A., Maheswari, U., Martens, C., Maumus, F., Otillar, R. P., Rayko, E., Salamov, A., Vandepoele, K., Beszteri, B., Gruber, A., Heijde, M., Katinka, M., Mock, T., Valentin, K., … Grigoriev, I. V. (2008). The Phaeodactylum genome reveals the evolutionary history of diatom genomes. *Nature*, *456*(7219), 239–244. https://doi.org/10.1038/nature07410

Brembu, T., Chauton, M. S., Winge, P., Bones, A. M., & Vadstein, O. (2017). Dynamic responses to silicon in Thalasiossira pseudonana - Identification, characterisation and classification of signature genes and their corresponding protein motifs. *Scientific Reports*, *7*(1), 4865. https://doi.org/10.1038/s41598-017-04921-0

Bromke, M. A., Giavalisco, P., Willmitzer, L., & Hesse, H. (2013). Metabolic Analysis of Adaptation to Short-Term Changes in Culture Conditions of the Marine Diatom Thalassiosira pseudonana. *PLoS ONE*, *8*(6), 1–11. https://doi.org/10.1371/journal.pone.0067340

Brunner, E., Richthammer, P., Ehrlich, H., Paasch, S., Simon, P., Ueberlein, S., & van Pée, K.-H. (2009). Chitin-Based Organic Networks: An Integral Part of Cell Wall Biosilica in the Diatom Thalassiosira pseudonana. *Angewandte Chemie International Edition*, *48*(51), 9724–9727. https://doi.org/10.1002/anie.200905028

Carvalho, R. N., Bopp, S. K., & Lettieri, T. (2011). Transcriptomics responses in marine diatom Thalassiosira pseudonana exposed to the polycyclic aromatic hydrocarbon benzo[a]pyrene. *PLoS ONE*, *6*(11). https://doi.org/10.1371/journal.pone.0026985

Carvalho, R. N., Burchardt, A. D., Sena, F., Mariani, G., Mueller, A., Bopp, S. K., Umlauf, G., & Lettieri, T. (2011). Gene biomarkers in diatom Thalassiosira pseudonana exposed to polycyclic aromatic hydrocarbons from contaminated marine surface sediments. *Aquatic Toxicology*, *101*(1), 244–253. https://doi.org/10.1016/j.aquatox.2010.10.004

Carvalho, R. N., & Lettieri, T. (2011). Proteomic analysis of the marine diatom Thalassiosira pseudonana upon exposure to benzo(a)pyrene. *BMC Genomics*, *12*(1), 159. https://doi.org/10.1186/1471-2164-12-159

Chiriboga, O. G., LeDuff, P., & Rorrer, G. L. (2020). Extracellular Chitin Nanofibers from Marine Diatoms. *Encyclopedia of Marine Biotechnology*, *13*, 1083–1092. https://doi.org/10.1002/9781119143802.ch43

Clement, R., Dimnet, L., Maberly, S. C., & Gontero, B. (2016). The nature of the CO2-concentrating mechanisms in a marine diatom, Thalassiosira pseudonana. *New Phytologist*, *209*(4), 1417–1427. https://doi.org/10.1111/nph.13728

Delasoie, J., & Zobi, F. (2019). Natural diatom biosilica as microshuttles in drug delivery systems. *Pharmaceutics*, *11*(10), 1–21. https://doi.org/10.3390/pharmaceutics11100537

Dong, H. P., Dong, Y. L., Cui, L., Balamurugan, S., Gao, J., Lu, S. H., & Jiang, T. (2016). High light stress triggers distinct proteomic responses in the marine diatom Thalassiosira pseudonana. *BMC Genomics*,

17(1), 1–14. https://doi.org/10.1186/s12864-016-3335-5
Drum, R. W., & Pankratz, H. S. (1964). Post mitotic fine structure of Gomphonema parvulum. *Journal of Ultrastructure Research*, *10*(3–4), 217–223.
Durkin, C. A., Mock, T., & Armbrust, E. V. (2009). Chitin in Diatoms and Its Association with the Cell Wall. *Eukaryotic Cell*, *8*(7), 1038–1050. https://doi.org/10.1128/EC.00079-09
Dyhrman, S. T., Jenkins, B. D., Rynearson, T. A., Saito, M. A., Mercier, M. L., Alexander, H., Whitney, L. A. P., Drzewianowski, A., Bulygin, V. V., Bertrand, E. M., Wu, Z., Benitez-Nelson, C., & Heithoff, A. (2012). The transcriptome and proteome of the diatom thalassiosira pseudonana reveal a diverse phosphorus stress response. *PLoS ONE*, *7*(3). https://doi.org/10.1371/journal.pone.0033768
Falkowski, P. G., Katz, M. E., Knoll, A. H., Quigg, A., Raven, J. A., Schofield, O., & Taylor, F. J. R. (2004). The evolution of modern eukaryotic phytoplankton. *Science*, *305*(5682), 354–360.
Fattorini, N., & Maier, U. G. (2021). Targeting of proteins to the cell wall of the diatom Thalassiosira pseudonana. *Discover Materials*, *1*(1). https://doi.org/10.1007/s43939-021-00005-z
Field, C. B., Behrenfeld, M. J., Randerson, J. T., & Falkowski, P. (1998). Primary production of the biosphere: integrating terrestrial and oceanic components. *Science*, *281*(5374), 237–240.
Frigeri, L. G., Radabaugh, T. R., Haynes, P. A., & Hildebrand, M. (2006). Identification of proteins from a cell wall fraction of the diatom Thalassiosira pseudonana: insights into silica structure formation. *Molecular & Cellular Proteomics*, *5*(1), 182–193.
Gonzalez, N. H., Felsner, G., Schramm, F. D., Klingl, A., Maier, U. G., & Bolte, K. (2011). A single peroxisomal targeting signal mediates matrix protein import in diatoms. *PLoS ONE*, *6*(9). https://doi.org/10.1371/journal.pone.0025316
Görlich, S., Pawolski, D., Zlotnikov, I., & Kröger, N. (2019). Control of biosilica morphology and mechanical performance by the conserved diatom gene Silicanin-1. *Communications Biology*, *2*(1). https://doi.org/10.1038/s42003-019-0436-0
Graham, L. E., Graham, J. M., Cook, M. E., & Wilcox, L. W. (2016). *Algae*. http://docstore.library.uvic.ca/Algae_3rd_ed_2016.pdf
Grouneva, I., Rokka, A., & Aro, E. M. (2011). The thylakoid membrane proteome of two marine diatoms outlines both diatom-specific and species-specific features of the photosynthetic machinery. *Journal of Proteome Research*, *10*(12), 5338–5353. https://doi.org/10.1021/pr200600f
Gruber, A., Weber, T., Bártulos, C. R., Vugrinec, S., & Kroth, P. G. (2009). Intracellular distribution of the reductive and oxidative pentose phosphate pathways in two diatoms. *Journal of Basic Microbiology*, *49*(1), 58–72. https://doi.org/10.1002/jobm.200800339
Heintze, C., Formanek, P., Pohl, D., Hauptstein, J., Rellinghaus, B., & Kröger, N. (2020). An intimate view into the silica deposition vesicles of diatoms. *BMC Materials*, *2*(1), 1–15.
Hempel, F., Bozarth, A. S., Lindenkamp, N., Klingl, A., Zauner, S., Linne, U., Steinbüchel, A., & Maier, U. G. (2011). Microalgae as bioreactors for bioplastic production. *Microbial Cell Factories*, *10*(1), 81. https://doi.org/10.1186/1475-2859-10-81
Hempel, F., Lau, J., Klingl, A., & Maier, U. G. (2011). Algae as protein factories: Expression of a human antibody and the respective antigen in the diatom phaeodactylum tricornutum. *PLoS ONE*, *6*(12). https://doi.org/10.1371/journal.pone.0028424
Hempel, F., & Maier, U. G. (2016). Microalgae as solar-powered protein factories. *Advances in Experimental Medicine and Biology*, *896*, 241–262. https://doi.org/10.1007/978-3-319-27216-0_16
Hildebrand, M., Frigeri, L. G., & Davis, A. K. (2007). SYNCHRONIZED GROWTH OF THALASSIOSIRA PSEUDONANA (BACILLARIOPHYCEAE) PROVIDES NOVEL INSIGHTS INTO CELL-WALL SYNTHESIS PROCESSES IN RELATION TO THE CELL CYCLE ^1. *Journal of Phycology*, *43*(4), 730–740. https://doi.org/10.1111/j.1529-8817.2007.00361.x
Hildebrand, M., Kim, S., Shi, D., Scott, K., & Subramaniam, S. (2009). 3D imaging of diatoms with ion-abrasion scanning electron microscopy. *Journal of Structural Biology*, *166*(3), 316–328.
Hildebrand, M., & Lerch, S. J. L. (2015). Diatom silica biomineralization: Parallel development of approaches and understanding. *Seminars in Cell & Developmental Biology*, *46*, 27–35. https://doi.org/10.1016/j.semcdb.2015.06.007
Hildebrand, M., Lerch, S. J. L., & Shrestha, R. P. (2018). Understanding Diatom Cell Wall Silicification---Moving Forward. *Frontiers in Marine Science*, *5*. https://doi.org/10.3389/fmars.2018.00125
Hildebrand, M., York, E., Kelz, J. I., Davis, A. K., Frigeri, L. G., Allison, D. P., & Doktycz, M. J. (2006). Nanoscale control of silica morphology and three-dimensional structure during diatom cell wall formation. *Journal of Materials Research*, *21*(10), 2689–2698. https://doi.org/10.1557/jmr.2006.0333
Hockin, N. L., Mock, T., Mulholland, F., Kopriva, S., & Malin, G. (2012). The response of diatom central carbon metabolism to nitrogen starvation is different from that of green algae and higher plants. *Plant Physiology*, *158*(1), 299–312. https://doi.org/10.1104/pp.111.184333
Hopkinson, B. M., Dupont, C. L., & Matsuda, Y. (2016). The physiology and genetics of CO2 concentrating mechanisms in model diatoms. *Current Opinion in Plant Biology*, *31*, 51–57. https://doi.org/10.1016/j.pbi.2016.03.013

Kotzsch, A., Gröger, P., Pawolski, D., Bomans, P. H. H., Sommerdijk, Nico A. J. M., Schlierf, M., & Kröger, N. (2017). Silicanin-1 is a conserved diatom membrane protein involved in silica biomineralization. *BMC Biology*, *15*(1). https://doi.org/10.1186/s12915-017-0400-8

Kotzsch, A., Pawolski, D., Milentyev, A., Shevchenko, A., Scheffel, A., Poulsen, N., Shevchenko, A., & Kröger, N. (2016). Biochemical Composition and Assembly of Biosilica-associated Insoluble Organic Matrices from the Diatom Thalassiosira pseudonana. *Journal of Biological Chemistry*, *291*(10), 4982–4997. https://doi.org/10.1074/jbc.M115.706440

Kröger, N., & Brunner, E. (2014). Complex-shaped microbial biominerals for nanotechnology. *Wiley Interdisciplinary Reviews: Nanomedicine and Nanobiotechnology*, *6*(6), 615–627. https://doi.org/10.1002/wnan.1284

Kröger, N., Deutzmann, R., Bergsdorf, C., & Sumper, M. (2000). Species-specific polyamines from diatoms control silica morphology. *Proceedings of the National Academy of Sciences of the United States of America*, *97*(26), 14133–14138. https://doi.org/10.1073/pnas.260496497

Kröger, N., & Poulsen, N. (2008). Diatoms---From Cell Wall Biogenesis to Nanotechnology. *Annual Review of Genetics*, *42*(1), 83–107. https://doi.org/10.1146/annurev.genet.41.110306.130109

Kroth, P. G. (2002). Protein transport into secondary plastids and the evolution of primary and secondary plastids. *International Review of Cytology*, *221*, 191–255. https://doi.org/10.1016/S0074-7696(02)21013-X

Kumari, E., Görlich, S., Poulsen, N., & Kröger, N. (2020). Genetically Programmed Regioselective Immobilization of Enzymes in Biosilica Microparticles. *Advanced Functional Materials*, *30*(25). https://doi.org/10.1002/adfm.202000442

Liu, X., Hempel, F., Stork, S., Bolte, K., Moog, D., Heimerl, T., Maier, U. G., & Zauner, S. (2016). Addressing various compartments of the diatom model organism Phaeodactylum tricornutum via sub-cellular marker proteins. *Algal Research*, *20*, 249–257. https://doi.org/10.1016/j.algal.2016.10.018

Lommer, M., Specht, M., Roy, A. S., Kraemer, L., Andreson, R., Gutowska, M. A., Wolf, J., Bergner, S. V., Schilhabel, M. B., Klostermeier, U. C., Beiko, R. G., Rosenstiel, P., Hippler, M., & LaRoche, J. (2012). Genome and low-iron response of an oceanic diatom adapted to chronic iron limitation. *Genome Biology*, *13*(7), R66. https://doi.org/10.1186/gb-2012-13-7-r66

Maier, U. G., Zauner, S., & Hempel, F. (2015). Protein import into complex plastids: Cellular organization of higher complexity. *European Journal of Cell Biology*, *94*(7–9), 340–348. https://doi.org/10.1016/j.ejcb.2015.05.008

Mann, D. G., & Droop, S. J. M. (1996). Biodiversity, biogeography and conservation of diatoms. *Hydrobiologia*, *336*(1–3), 19–32. https://doi.org/10.1007/BF00010816

Marter, P., Schmidt, S., Kiontke, S., & Moog, D. (2020). Optimized mRuby3 is a Suitable Fluorescent Protein for in vivo Co-localization Studies with GFP in the Diatom Phaeodactylum tricornutum. *Protist*, *171*(1), 125715. https://doi.org/10.1016/j.protis.2020.125715

Martin, P., Van Mooy, B. A. S., Heithoff, A., & Dyhrman, S. T. (2011). Phosphorus supply drives rapid turnover of membrane phospholipids in the diatom Thalassiosira pseudonana. *ISME Journal*, *5*(6), 1057–1060. https://doi.org/10.1038/ismej.2010.192

Marty, F. (1999). Plant vacuoles. *The Plant Cell*, *11*(4), 587–599.

Mock, T., Samanta, M. P., Iverson, V., Berthiaume, C., Robison, M., Holtermann, K., Durkin, C., BonDurant, S. S., Richmond, K., Rodesch, M., Kallas, T., Huttlin, E. L., Cerrina, F., Sussman, M. R., & Armbrust, E. V. (2008). Whole-genome expression profiling of the marine diatom Thalassiosira pseudonana identifies genes involved in silicon bioprocesses. *Proceedings of the National Academy of Sciences of the United States of America*, *105*(5), 1579–1584. https://doi.org/10.1073/pnas.0707946105

Montsant, A., Allen, A. E., Coesel, S., Martino, A. De, Falciatore, A., Mangogna, M., Siaut, M., Heijde, M., Jabbari, K., Maheswari, U., Rayko, E., Vardi, A., Apt, K. E., Berges, J. A., Chiovitti, A., Davis, A. K., Thamatrakoln, K., Hadi, M. Z., Lane, T. W., … Bowler, C. (2007). Identification and comparative genomic analysis of signaling and regulatory components in the diatom Thalassiosira pseudonana. *Journal of Phycology*, *43*(3), 585–604. https://doi.org/10.1111/j.1529-8817.2007.00342.x

Nawaly, H., Tsuji, Y., & Matsuda, Y. (2020). Rapid and precise genome editing in a marine diatom, Thalassiosira pseudonana by Cas9 nickase (D10A). *Algal Research*, *47*(August 2019), 101855. https://doi.org/10.1016/j.algal.2020.101855

Norden-Krichmar, T. M., Allen, A. E., Gaasterland, T., & Hildebrand, M. (2011). Characterization of the small RNA transcriptome of the diatom, Thalassiosira pseudonana. *PLoS ONE*, *6*(8). https://doi.org/10.1371/journal.pone.0022870

Oliver, A., Podell, S., Pinowska, A., Traller, J. C., Smith, S. R., McClure, R., Beliaev, A., Bohutskyi, P., Hill, E. A., Rabines, A., Zheng, H., Allen, L. Z., Kuo, A., Grigoriev, I. V., Allen, A. E., Hazlebeck, D., & Allen, E. E. (2021). Diploid genomic architecture of Nitzschia inconspicua, an elite biomass production diatom. *Scientific Reports*, *11*(1), 1–14. https://doi.org/10.1038/s41598-021-95106-3

Osuna-Cruz, C. M., Bilcke, G., Vancaester, E., De Decker, S., Bones, A. M., Winge, P., Poulsen, N., Bulankova, P., Verhelst, B., Audoor, S., Belisova, D., Pargana, A., Russo, M., Stock, F., Cirri, E.,

Brembu, T., Pohnert, G., Piganeau, G., Ferrante, M. I., … Vandepoele, K. (2020). The Seminavis robusta genome provides insights into the evolutionary adaptations of benthic diatoms. *Nature Communications*, *11*(1). https://doi.org/10.1038/s41467-020-17191-8

Pan, Z., Lerch, S. J. L., Xu, L., Li, X., Chuang, Y. J., Howe, J. Y., Mahurin, S. M., Dai, S., & Hildebrand, M. (2014). Electronically transparent graphene replicas of diatoms: A new technique for the investigation of frustule morphology. *Scientific Reports*, *4*, 1–6. https://doi.org/10.1038/srep06117

Pickett-Heaps, J. (1990). The cell biology of diatom valve formation. *Progress in Phycological Research*, *7*, 1–101.

Poulsen, N., Berne, C., Spain, J., & Kröger, N. (2007). Silica Immobilization of an Enzyme through Genetic Engineering of the DiatomThalassiosira pseudonana. *Angewandte Chemie International Edition*, *46*(11), 1843–1846. https://doi.org/10.1002/anie.200603928

Poulsen, N., Chesley, P. M., & Kröger, N. (2006). MOLECULAR GENETIC MANIPULATION OF THE DIATOM THALASSIOSIRA PSEUDONANA (BACILLARIOPHYCEAE). *Journal of Phycology*, *42*(5), 1059–1065. https://doi.org/10.1111/j.1529-8817.2006.00269.x

Poulsen, N., & Kröger, N. (2004). Silica Morphogenesis by Alternative Processing of Silaffins in the Diatom Thalassiosira pseudonana. *Journal of Biological Chemistry*, *279*(41), 42993–42999. https://doi.org/10.1074/jbc.M407734200

Poulsen, N., Scheffel, A., Sheppard, V. C., Chesley, P. M., & Kröger, N. (2013). Pentalysine Clusters Mediate Silica Targeting of Silaffins in Thalassiosira pseudonana. *Journal of Biological Chemistry*, *288*(28), 20100–20109. https://doi.org/10.1074/jbc.M113.469379

Poulsen, N., Sumper, M., & Kröger, N. (2003). Biosilica formation in diatoms: characterization of native silaffin-2 and its role in silica morphogenesis. *Proceedings of the National Academy of Sciences*, *100*(21), 12075–12080.

Richthammer, P., Börmel, M., Brunner, E., & van Pée, K.-H. (2011). Biomineralization in diatoms: The role of silacidins. *ChemBioChem*, *12*(9), 1362–1366.

Ross, R., Cox, E. J., Karayeva, N. I., Mann, D. G., Paddock, T. B. B., Simonsen, R., & Sims, P. A. (1979). *An amended terminology for the siliceous components of the diatom cell*.

Round, F. E., Crawford, R. M., & Mann, D. G. (2007). *Diatoms: Biology and Morphology of the Genera*.

Samukawa, M., Shen, C., Hopkinson, B. M., & Matsuda, Y. (2014). Localization of putative carbonic anhydrases in the marine diatom, Thalassiosira pseudonana. *Photosynthesis Research*, *121*(2–3), 235–249. https://doi.org/10.1007/s11120-014-9967-x

Scheffel, A., Poulsen, N., Shian, S., & Kröger, N. (2011). Nanopatterned protein microrings from a diatom that direct silica morphogenesis. *Proceedings of the National Academy of Sciences*, *108*(8), 3175–3180.

Schmid, A.-M. M. (1994). Aspects of morphogenesis and function of diatom cell walls with implications for taxonomy. In *The protistan cell surface* (pp. 43–60). Springer.

Schneider, C. A., Rasband, W. S., & Eliceiri, K. W. (2012). NIH Image to ImageJ: 25 years of image analysis. *Nature Methods*, *9*(7), 671–675. https://doi.org/10.1038/nmeth.2089

Secq, M. P. O. Le, & Green, B. R. (2011). Complex repeat structures and novel features in the mitochondrial genomes of the diatoms Phaeodactylum tricornutum and Thalassiosira pseudonana. *Gene*, *476*(1–2), 20–26. https://doi.org/10.1016/j.gene.2011.02.001

Sharma, N., Simon, D. P., Diaz-Garza, A. M., Fantino, E., Messaabi, A., Meddeb-Mouelhi, F., Germain, H., & Desgagné-Penix, I. (2021). Diatoms Biotechnology: Various Industrial Applications for a Greener Tomorrow. *Frontiers in Marine Science*, *8*(February). https://doi.org/10.3389/fmars.2021.636613

Sheppard, V. C., Scheffel, A., Poulsen, N., & Kröger, N. (2012). Live diatom silica immobilization of multimeric and redox-active enzymes. *Applied and Environmental Microbiology*, *78*(1), 211–218. https://doi.org/10.1128/AEM.06698-11

Sheppard, V., Poulsen, N., & Kröger, N. (2010). Characterization of an Endoplasmic Reticulum-associated Silaffin Kinase from the Diatom Thalassiosira pseudonana. *Journal of Biological Chemistry*, *285*(2), 1166–1176. https://doi.org/10.1074/jbc.M109.039529

Shrestha, Roshan P., & Hildebrand, M. (2017). Development of a silicon limitation inducible expression system for recombinant protein production in the centric diatoms Thalassiosira pseudonana and Cyclotella cryptica. *Microbial Cell Factories*, *16*(1), 1–14. https://doi.org/10.1186/s12934-017-0760-3

Shrestha, Roshan Prakash, Tesson, B., Norden-Krichmar, T., Federowicz, S., Hildebrand, M., & Allen, A. E. (2012). Whole transcriptome analysis of the silicon response of the diatom Thalassiosira pseudonana. *BMC Genomics*, *13*(1), 499. https://doi.org/10.1186/1471-2164-13-499

Sigel, A., Sigel, H., & Sigel, R. K. O. (2008). Biomineralization. In A. Sigel, H. Sigel, & R. K. O. Sigel (Eds.), *Biomineralization: From Nature to Application* (Vol. 4). John Wiley & Sons, Ltd. https://doi.org/10.1002/9780470986325

Stork, S., Moog, D., Przyborski, J. M., Wilhelmi, I., Zauner, S., & Maier, U. G. (2012). Distribution of the SELMA translocon in secondary plastids of red algal origin and predicted uncoupling of ubiquitin-dependent translocation from degradation. *Eukaryotic Cell*, *11*(12), 1472–1481.

https://doi.org/10.1128/EC.00183-12
Sumper, M., & Brunner, E. (2008). Silica Biomineralisation in Diatoms: The Model Organism Thalassiosira pseudonana. *ChemBioChem*, *9*(8), 1187–1194. https://doi.org/10.1002/cbic.200700764
Sumper, M., Brunner, E., & Lehmann, G. (2005). Biomineralization in diatoms: Characterization of novel polyamines associated with silica. *FEBS Letters*, *579*(17), 3765–3769. https://doi.org/10.1016/j.febslet.2005.06.001
Sumper, M., Hett, R., Lehmann, G., & Wenzl, S. (2007). A Code for Lysine Modifications of a Silica Biomineralizing Silaffin Protein. *Angewandte Chemie International Edition*, *46*(44), 8405–8408. https://doi.org/10.1002/anie.200702413
Tachibana, M., Allen, A. E., Kikutani, S., Endo, Y., Bowler, C., & Matsuda, Y. (2011). Localization of putative carbonic anhydrases in two marine diatoms, Phaeodactylum tricornutum and Thalassiosira pseudonana. *Photosynthesis Research*, *109*(1–3), 205–221. https://doi.org/10.1007/s11120-011-9634-4
Tanaka, R., Kikutani, S., Mahardika, A., & Matsuda, Y. (2014). Localization of enzymes relating to C4 organic acid metabolisms in the marine diatom, Thalassiosira pseudonana. *Photosynthesis Research*, *121*(2–3), 251–263. https://doi.org/10.1007/s11120-014-9968-9
Tanaka, T., Maeda, Y., Veluchamy, A., Tanaka, M., Abida, H., Maréchal, E., Bowler, C., Muto, M., Sunaga, Y., Tanaka, M., Yoshino, T., Taniguchi, T., Fukuda, Y., Nemoto, M., Matsumoto, M., Wong, P. S., Aburatani, S., & Fujibuchi, W. (2015). Oil accumulation by the oleaginous diatom Fistulifera solaris as revealed by the genome and transcriptome. *Plant Cell*, *27*(1), 162–176. https://doi.org/10.1105/tpc.114.135194
Tesson, B., Lerch, S. J. L., & Hildebrand, M. (2017). Characterization of a New Protein Family Associated With the Silica Deposition Vesicle Membrane Enables Genetic Manipulation of Diatom Silica. *Scientific Reports*, *7*(1). https://doi.org/10.1038/s41598-017-13613-8
Thamatrakoln, K., Korenovska, O., Niheu, A. K., & Bidle, K. D. (2012). Whole-genome expression analysis reveals a role for death-related genes in stress acclimation of the diatom Thalassiosira pseudonana. *Environmental Microbiology*, *14*(1), 67–81. https://doi.org/10.1111/j.1462-2920.2011.02468.x
Traller, J. C., Cokus, S. J., Lopez, D. A., Gaidarenko, O., Smith, S. R., McCrow, J. P., Gallaher, S. D., Podell, S., Thompson, M., Cook, O., Morselli, M., Jaroszewicz, A., Allen, E. E., Allen, A. E., Merchant, S. S., Pellegrini, M., & Hildebrand, M. (2016). Genome and methylome of the oleaginous diatom Cyclotella cryptica reveal genetic flexibility toward a high lipid phenotype. *Biotechnology for Biofuels*, *9*(1), 1–20. https://doi.org/10.1186/s13068-016-0670-3
Tréguer, P. J., & De La Rocha, C. L. (2013). The world ocean silica cycle. *Annual Review of Marine Science*, *5*, 477–501. https://doi.org/10.1146/annurev-marine-121211-172346
Tréguer, P., Nelson, D. M., Van Bennekom, A. J., Demaster, D. J., Leynaert, A., & Quéguiner, B. (1995). The silica balance in the world ocean: A reestimate. *Science*, *268*(5209), 375–379. https://doi.org/10.1126/science.268.5209.375
Tréguer, P., & Pondaven, P. (2000). Silica control of carbon dioxide. *Nature*, *406*(6794), 358–359. https://doi.org/10.1038/35019236
Viotti, C. (2014). ER and vacuoles: never been closer. *Frontiers in Plant Science*, *5*. https://doi.org/10.3389/fpls.2014.00020
Wenzl, S., Hett, R., Richthammer, P., & Sumper, M. (2008). Silacidins: Highly Acidic Phosphopeptides from Diatom Shells Assist in Silica Precipitation In Vitro. *Angewandte Chemie International Edition*, *47*(9), 1729–1732. https://doi.org/10.1002/anie.200704994
Wu, Y., & Wang, W. X. (2011). Accumulation, subcellular distribution and toxicity of inorganic mercury and methylmercury in marine phytoplankton. *Environmental Pollution*, *159*(10), 3097–3105. https://doi.org/10.1016/j.envpol.2011.04.012
Wustmann, M., Poulsen, N., Kröger, N., & van Pée, K.-H. (2020). Chitin synthase localization in the diatom Thalassiosira pseudonana. *BMC Materials*, *2*(1), 1–7. https://doi.org/10.1186/s42833-020-00016-9
Yang, G., & Gao, K. (2012). Physiological responses of the marine diatom Thalassiosira pseudonana to increased pCO 2 and seawater acidity. *Marine Environmental Research*, *79*, 142–151. https://doi.org/10.1016/j.marenvres.2012.06.002
Yee, D. P., Hildebrand, M., & Tresguerres, M. (2020). Dynamic subcellular translocation of V--type H+--ATPase is essential for biomineralization of the diatom silica cell wall. *New Phytologist*, *225*(6), 2411–2422.
Yu, E. T., Zendejas, F. J., Lane, P. D., Gaucher, S., Simmons, B. A., & Lane, T. W. (2009). Triacylglycerol accumulation and profiling in the model diatoms Thalassiosira pseudonana and Phaeodactylum tricornutum (Baccilariophyceae) during starvation. *Journal of Applied Phycology*, *21*(6), 669–681. https://doi.org/10.1007/s10811-008-9400-y

6.2 INTERNET LINKS

Link1https://www.rsc.org/periodic-table/element/14/silicon#:~:text=Silicon%20makes%20up%2027.7%25%20of,(silica)%20and%20as%20silicates.

Link 2 https://ncma.bigelow.org/CCMP1335?quantity=1&product-type=9

Link 3 https://mycocosm.jgi.doe.gov/Thaps3/Thaps3.home.html

Link 4 http://insilico.ehu.eus/counting_chamber/thoma.php

Link 5 https://www.jenabioscience.com/images/741d0cd7d0/NTC-Flyer.pdf

Link 6 http://cccm.botany.ubc.ca/resources/marine-media-receipes/

Link 7 https://barricklab.org/twiki/bin/view/Lab/ProtocolsGibsonCloning

7. SUPPLEMENTARY INFORMATION

7.1 LIST OF CONSTRUCTS (WITH PRIMERS)

Table 9 List of constructs and primers used to generate them. In the first column is indicated the name of the construct created (in bold is shown the working label, whereas the label used in the thesis is shown in normal font); in the second column is indicated the template used for each PCR reaction; in the third column is indicated the name of the primer used for the PCR reaction; in the last column is indicated the purpose of the reaction. The background color of the cells indicates the cloning strategy used: **grey** is for traditional cloning (sticky-ends ligation), **light blue** is for site-directed mutagenesis, **yellow** is for Gibson Assembly. Because of a change in the experimental planning, the cloning of constructs W2-LER2-4 and Y2-LER2 was obtained by first inserting a glycine spacer (made of 6 glycines) between the portion of interest and the eGFP; the spacer has been then removed.

CONSTRUCT	TEMPLATE	PRIMERS	USE
pTpNAT-MCS	937	Fcp9promFw	Amplification of NAT cassette
		Fcp9termRv2	
	pPhaNR	pPhaT1MCSfw	Amplification of MCS
		pPhaT1MCSrv	
pTpSil3	396 pTpNAT-MCS	Sil3EcoRVwholeFw	Insertion of Sil3 gene cassette
		Sil3AccI300DwRv	
pTpW2	936 pTpNAT-MCS	CinW2XbaI800upFw	Insertion of CinW2 gene cassette
		CinW2AccIWholeRv	
pTpY2	937	(Gib)CinY2insert_fw	Insertion of CinY2 gene cassette
		(Gib)CinY2insert_rv	
	pTpNAT-MCS	(Gib)Y2insertion_pTpBB_fw	
		(Gib)Y2insertion_pTpBB_rv	
pTp5357-GFP	gDNA pTpNAT-MCS	5357_SacI800up_fw	Insertion of 5357 gene cassette in pTpNAT-MCS
		5357_XbaI300dw_rv	
	pTp5357	pTp5357_gib_fw	Insertion of eGFP
		pTp5357_gib_rv	
	pTpSil3 (for the eGFP)	CinW2 N-ter_fw	
		eGFP_rv	
pTp12162-GFP	gDNA	pTp12162_800up(gib)_fw	Insertion of 12162 gene cassette into pTpNAT-MCS
		pTp12162_300dw_rv(gib)CORRECT	
	pTpNAT-MCS	fcp9_p_fw(no r.s.)	
		pTpInvNAT_bb_rv	
	pTpSil3 (for the eGFP)	CinW2 N-ter_fw	Insertion of eGFP into pTp12162
		eGFP_rv(no stop) for 12162(gib)	
	pTp12162	fcp9_fw_12162(gib)	
		pTp12162_300up(gib)_rv	
pTpSil1-mRuby	gDNA pTpNAT-MCS	Sil1_KpnI800up_fw	Amplification from gDNA
		Sil1_SalI-300dw_rv	
	(from previous step)	(gib)Sil1/2_mRubyINS_fw	Insertion of eGFP
		(gib)Sil1/2_mRubyINS_rv	
pTpSil2-mRuby	gDNA pTpNAT-MCS	Sil1_KpnI800up_fw	Amplification from gDNA
		Sil1_SalI-300dw_rv	
	(from previous step)	(gib)Sil1/2_mRubyINS_fw	Insertion of eGFP
		(gib)Sil1/2_mRubyINS_rv	
pTpINV-NAT	pTpNAT-MCS	pTpInvNAT_bb_fw	Deletion of NAT cassette
		pPhaT1MCSrv	
	pTpNAT-MCS	(gib)NATcassette_fw	Insertion of NAT (amplification of the NAT fragment)
		(gib)NATcassette_rv(vector)	
	(from previous step)	pTpInvNAT_bb_rv	Insertion of NAT (amplification of the backbone fragment)
		pPhaT1MCSfw	
pTpINV-Sil3 prom-GFP	pTpSil3 prom-GFP	pTpInvNAT_bb_fw	Deletion of NAT cassette
		Sil3AccI300DwRv	
	pTpSil3 prom-GFP	(gib)NATcassette_fw	Insertion of NAT (amplification of the NAT fragment)
		(gib)NATcassette_rv(Sil3)	
	(from first step)	pTpInvNAT_bb_rv	Insertion of NAT (amplification of the backbone fragment)
		Sil3_800up_fw	
pTpINV-Sil3 prom-400	pTpINV-Sil3 prom-GFP	Sil3_promoter400_fw	Sil3 400 bp promoter truncation
		Fcp9TermRv2	
pTpINV-Sil3 prom-200	pTpINV-Sil3 prom-	Sil3_promoter200_fw	Sil3 200 bp promoter truncation

	GFP	Fcp9TermRv2	
pTpINV-CinW2 prom-GFP	pTpCinW2 prom-GFP	pTpInvNAT_bb_fw	Deletion of NAT cassette
		CinW2AccIWholeRv	
	pTpCinW2 prom-GFP	(gib)NATcassette_fw	Insertion of NAT (amplification of the NAT fragment)
		(gib)NATcassette_rv(W2)	
	(from first step)	pTpInvNAT_bb_rv	Insertion of NAT (amplification of the backbone fragment)
		W2_800up_fw	
pTpINV-CinW2 prom-400	pTpINV-CinW2 prom-GFP	CinW2_promoter400_fw	CinW2 400 bp promoter truncation
		Fcp9TermRv2	
pTpINV-CinW2 prom-200	pTpINV-CinW2 prom-GFP	CinW2_promoter200_fw	CinW2 200 bp promoter truncation
		Fcp9TermRv2	
pTpINV-CinY2 prom-GFP	pTpCinY2 prom-GFP	pTpInvNAT_bb_fw	Deletion of NAT cassette
		CinY2AccIII300DwRV	
	pTpCinY2 prom-GFP	(gib)NATcassette_fw	Insertion of NAT (amplification of the NAT fragment)
		(gib)NATcassette_rv(Y2)	
	(from first step)	pTpInvNAT_bb_rv	Insertion of NAT (amplification of the backbone fragment)
		Y2_800up_fw	
pTpINV-CinY2 prom-400	pTpINV-CinY2 prom-GFP	CinY2_promoter400_fw	CinY2 400 bp promoter truncation
pTpINV-CinY2 prom-200	pTpINV-CinY2 prom-GFP	CinY2_promoter200_fw	CinY2 200 bp promoter truncation
Sil3(BB)-CinW2(CDS)	pTpSil3	(new)D_Sil3_TermFw	Amplification of the Sil3 BB
		D_Sil3_PromRv	
	pTpCinW2	D_CinW2_GeneFw(Sil3_PromFw)	Amplification of the CinW2 CDS
		(new)D_CinW2_GeneRv(Sil3_TermRv)	
Sil3(BB)-CinY2(CDS)	pTpSil3	(new)D_Sil3_TermFw	Amplification of the Sil3 BB
		D_Sil3_PromRv	
	pTpCinY2	(new)D_CinY2_GeneFw(Sil3_PromFw)	Amplification of the CinY2 CDS
		(new)D_CinY2_GeneRv(Sil3_TermRv)	
CinW2(BB)-Sil3(CDS)	pTpCinW2	(new)D_CinW2_TermFw	Amplification of the CinW2 BB
		D_CinW2_PromRv	
	pTpSil3	D_Sil3_GeneFw(CinW2_PromFw)	Amplification of the Sil3 CDS
		(new)D_Sil3_GeneRv(CinW2_TermRv)	
CinW2(BB)-CinY2(CDS)	pTpCinW2	(new)D_CinW2_TermFw	Amplification of the CinW2 BB
		D_CinW2_PromRv	
	pTpCinY2	(new)D_CinY2_GeneFw(CinW2_PromFw)	Amplification of the CinY2 CDS
		(new)D_CinY2_GeneRv(CinW2_TermRv)	
CinY2(BB)-Sil3(CDS)	pTpCinY2	D_CinY2_TermFw	Amplification of the CinY2 BB
		(new)D_CinY2_PromRv	
	pTpSil3	(new)D_Sil3_GeneFw(CinY2_PromFw)	Amplification of the Sil3 CDS
		(new)D_Sil3_GeneRv(CinY2_TermRv)	
CinY2(BB)-CinW2(CDS)	pTpCinY2	D_CinY2_TermFw	Amplification of the CinY2 BB
		(new)D_CinY2_PromRv	
	pTpCinW2	(new)D_CinW2_GeneRv(CinY2_TermRv)	Amplification of the CinW2 CDS
		(new)D_CinW2_GeneFw(CinY2_PromFw)	
CinW2 N-ter (CinW2 TR-A)	pTpCinW2	CinW2Trunc_N-ter_fw	Truncation
		CinW2Trunc_N-ter_rv	
CinW2 C-ter (CinW2 TR-B)	pTpCinW2	CinW2Trunc_C-ter_fw	Truncation
		CinW2Trunc_C-ter_rv	
CinY2 N-ter (CinY2 TR-A)	pTpCinY2	CinY2Trunc_N-ter_fw	Truncation
		CinY2Trunc_N-ter_rv	
CinY2 C-ter (CinY2 TR-B)	pTpCinY2	CinY2Trunc_C-ter_fw	Truncation
		CinY2Trunc_C-ter_rv	
CinW2 N-ter PLC1 mut (CinW2 MUT-A)	pTpCinW2	W2_PLC1(k-to-r)_fw	Lysine to arginine mutation of W2-LER1
		W2_PLC1(k-to-r)_rv	
CinW2 N-ter PLC2 mut (CinW2 MUT-B)	pTpCinW2	CinW2PLC2(K-to-R)fw	Lysine to arginine mutation of W2-LER2
		CinW2PLC(K-to-R)rv	
CinW2 N-ter PLC1/PLC2 mut (CinW2 MUT-C)	CinW2 N-ter PLC1 mut	CinW2PLC2(K-to-R)fw	Lysine to arginine mutation of both LERs in CinW2
		CinW2PLC(K-to-R)rv	
CinW2 N-ter PLC1 del (CinW2 DEL-A)	pTpCinW2	W2_PLC1del_fw2	Deletion of W2-LER1
		W2_PLC1del_rv2	
CinW2 N-ter PLC2 del (CinW2 DEL-B)	pTpCinW2	W2_PLC2del_fw	Deletion of W2-LER2
		W2_PLC2del_rv	
CinW2 N-ter PLC1/PLC2 del (CinW2 DEL-C)	CinW2 N-ter PLC1 del	W2_PLC2del_fw	Deletion of both LERs from CinW2
		W2_PLC2del_rv	
W2 PLC1 (W2-LER1)	pTpW2	CinW2_PLC1_fw	Deletion of the region comprised between the SP and the LER1
		CinW2Trunc_C-ter_rv	
	(from previous step)	CinW2Trunc_N-ter_fw	Deletion of the region comprised between the LER1 and the eGFP
		CinW2_PLC1_rv	
W2 PLC2 (W2-LER2)	pTpW2	W2_SP+GS+GFP_fw	Insertion of a glycine spacer between the LER2 and the eGFP
		W2_SP+GS+GFP_rv	

	(from previous step)	W2_PLC2_fw2 W2_PLC2_rv2	Insertion of LER2
	(from previous step)	CinW2Trunc_N-ter_fw W2_PLC2_rv	Deletion of the glycine spacer
W2 PLC3 (W2-LER3)	pTpW2	W2_SP+GS+GFP_fw W2_SP+GS+GFP_rv	Insertion of a glycine spacer between the LER3 and the eGFP
	(from previous step)	W2_PLC3_fw2 W2_PLC3_rv2	Insertion of LER3
	(from previous step)	CinW2Trunc_N-ter_fw W2_PLC3_rv	Deletion of the glycine spacer
W2 PLC4 (W2-LER4)	pTpW2	W2_SP+GS+GFP_fw W2_SP+GS+GFP_rv	Insertion of a glycine spacer between the PLC and the eGFP
	(from previous step)	W2_PLC4_fw2 W2_PLC4_rv2	Insertion of LER4
	(from previous step)	CinW2Trunc_N-ter_fw W2_PLC4_rv	Deletion of the glycine spacer
Y2 PLC1 (W2-LER4)	pTpY2	Y2_PLC1_fw CinW2Trunc_N-ter_fw	Deletion of the region comprised between the SP and the LER1
	(from previous step)	CinY2Trunc_C-ter_rv Y2_PLC1_rv	Deletion of the region comprised between the LER1 and the eGFP
Y2 PLC2 (W2-LER4)	pTpY2	W2_SP+GS+GFP_fw Y2_SP+GS+GFP_rv	Insertion of a glycine spacer between the LER2 and the eGFP
	(from previous step)	Y2_PLC2_fw2 Y2_PLC2_rv2	Insertion of LER2
	(from previous step)	CinW2Trunc_N-ter_fw Y2_PLC2_rv	Deletion of the glycine spacer
Sil3 PLC1 (Sil3-PLCnt)	pTpSil3	Sil3_PLC1_no GS_fw Sil3_PLC1_no GS_rv	Insertion of PLCnt
Sil3 PLC2 (Sil3-PLCnt)	pTpSil3	Sil3_PLC2_no GS_fw Sil3_PLC2_no GS_rv	Insertion of PLCct
CinW2 N-ter 1-103 (CinW2 TR-C)	pTpCinW2 N-ter	CinW2Trunc_N-ter_fw CinW2_Trunc_N-ter(1-103)rv	Truncation
CinW2 N-ter PLC1/PLC2 del 1-103 (CinW2 TR-C DEL)	pTpCinW2 N-ter PLC1/PLC2 del	CinW2Trunc_N-ter_fw CinW2_Trunc_N-ter(1-103)rv	Truncation
CinW2 N-ter 104-187 (CinW2 TR-C)	pTpCinW2 N-ter	CinW2_Trunc_N-ter(104-187)fw CinW2Trunc_C-ter_rv	Truncation
CinW2 N-ter PLC1/PLC2 del 104-187 (CinW2 TR-D DEL)	pTpCinW2 N-ter PLC1/PLC2 del	W2-N-ter-PLC1/2del(104-187)_fw CinW2Trunc_C-ter_rv	Truncation
CinY2 1-72 (CinW2 TR-C)	pTpCinY2 N-ter	CinY2Trunc_N-ter_fw CinY2_Trunc_N-ter(1-72)rv	Truncation
CinY2 73-136 (CinW2 TR-D)	pTpCinY2 N-ter	CinY2_Trunc_N-ter(73-136)fw CinY2Trunc_C-ter_rv	Truncation
CinY2 17 aa after SP (CinW2 UM)	pTpCinY2	CinY2Trunc_N-ter_fw CinY2_17aa_afterSPrv	Truncation
CinY2 17 aa after SP del (CinW2 UM DEL)	pTpCinY2	Y2_17aa del_fw CinY2Trunc_C-ter_rv	Truncation
Sil3 PLC1/PLC2 del (Sil3 DEL)	pTpSil3	Sil3_PLC1del_fw Sil3_PLC1del_rv	Deletion of both PLC
	pTpSil3 PLC1 del	Sil3_PLC2del_fw Sil3_PLC2del_rv	
Sil3 PLC1/PLC2 del T2 (Sil3 DEL T2)	pTpSil3 PLC1/PLC2 del	CinW2Trunc_N-ter_fw Sil3_T2_rv	Truncation
Sil3 PLC1/PLC2 del T4 (Sil3 DEL T4)	pTpSil3 PCL1/PLC2 del	Sil3_T4_fw Sil3_SP_rv	Truncation
Sil3 only SP (Sil3 SP)	pTpSil3	CinW2Trunc_N-ter_fw Sil3_SP_rv	Truncation
CinW2 only SP (CinW2 SP)	pTpCinW2	CinW2Trunc_N-ter_fw CinW2Trunc_C-ter_rv	Truncation
Y2 only SP (CinY2 SP)	pTpCinY2	CinW2Trunc_N-ter_fw CinY2Trunc_C-ter_rv	Truncation
Sil3 KSGK	pTpSil3	KSGK_fw Sil3_SP_rv	Truncation

7.2 LIST OF PRIMERS' SEQUENCES

Table 10 List of primers in alphabetical order with corresponding 5'-3' nucleotide sequence

Primer name	5'-3' sequence
(Gib)CinY2insert_fw	GGTACCCGGGGATCCTCTAGAGGTGGTGGTGTTGGCTGCTAGC
(Gib)CinY2insert_rv	GCTTGCATGCCTGCAGGTCGACCTAACAAATACTAATTCGCTGTTGC
(gib)NATcassette_fw	GCAGATTGTACTGAGAGTGCACGCAACGAATATTTCC
(gib)NATcassette_rv(Sil3)	CATCAGTATCCTGCCGTAATGGATCTAGGGG
(gib)NATcassette_rv(vector)	CGAATTCCATATGTAATGGATCTAGGGGATGC
(gib)NATcassette_rv(W2)	CCATTAGTTAGGTCTGTGCAATAATGGATCTAGGGGATGC
(gib)NATcassette_rv(Y2)	GCTAGCAGCCAACACCACCACCTAATGGATCTAGGGGATGC
(gib)Sil1/2_mRubyINS_fw	GAATGGACGAATTGTACAAGGAAGACCACTACTTTTTTGGCAAG
(gib)Sil1/2_mRubyINS_rv	TCTTCTCCCTTGGAGACCATGGTGGGCGATATAGTAGGCATTTC
(Gib)Y2insertion_pTpBB_fw	GTCGACCTGCAGGCATGCAAGC
(Gib)Y2insertion_pTpBB_rv	TCTAGAGGATCCCCGGGTACC
(new)D_CinW2_GeneFw(CinY2_PromFw)	ATCATACAGCAAATCGACAATATATACAATGAAGCTCGCTCTCTTCCTAACGATCC
(new)D_CinW2_GeneRv(CinY2_TermRv)	CGGCATGGACGAGCTGTACAAGTAAGTGTTGATGTCTCCTTCAAGTGACACC
(new)D_CinW2_GeneRv(Sil3_TermRv)	CGGCATGGACGAGCTGTACAAGTAAGTTCATCATCTTCATATCGTATGAAGTGG
(new)D_CinW2_TermFw	TAAATAACCACAACTATCTACC
(new)D_CinY2_GeneFw(CinW2_PromFw)	CTCAACGATAAACGAACAAAAGAAACCATGAAGTTAATCATCGCCCTCAGCGC
(new)D_CinY2_GeneFw(Sil3_PromFw)	CCTATCAACACATCACCATATCAAAAATGAAGTTAATCATCGCCCTCAGCGC
(new)D_CinY2_GeneRv(CinW2_TermRv)	GGTGCTTCTCACAACAACAGGATATAATAAATAACCACAACTATCTACC
(new)D_CinY2_GeneRv(Sil3_TermRv)	GGTGCTTCTCACAACAACAGGATATAAGTTCATCATCTTCATATCGTATGAAGTGG
(new)D_CinY2_PromRv	ATCATACAGCAAATCGACAATATATACA
(new)D_Sil3_GeneFw(CinY2_PromFw)	ATCATACAGCAAATCGACAATATATACAATGAAGACTTCTGCCATTGTATTGC
(new)D_Sil3_GeneRv(CinW2_TermRv)	GGACGAGCTGTACAAGTAATAAATAACCACAACTATCTACC
(new)D_Sil3_GeneRv(CinY2_TermRv)	GGACGAGCTGTACAAGTAAGTGTTGATGTCTCCTTCAAGTGACACC
(new)D_Sil3_TermFw	GTTCATCATCTTCATATCGTATGAAGTGG
5357_SacI800up_fw	GAGCTCCCTTTGGATTCAACG
5357_XbaI300dw_rv	TCTAGACTTTCTGGAAGTTTGACG
CinW2_PLC1_fw	AAATCTGGCAAGTCCGGATCGG
CinW2_PLC1_rv	CTTGCCGGACTTGCCCGATCC
CinW2_promoter200_fw	ACTCCGAACGGACAGCACAGTTCTACG
CinW2_promoter400_fw	ATGAAAATCCTGTGAGAAGCCTTCCG
CinW2_Trunc_N-ter(1-103)rv	GGAGTCATCTTCAATGTACACC
CinW2_Trunc_N-ter(104-187)fw	AGCGATGGCAGTGGTAAGAGTGG
CinW2AccIWholeRv	GTCGACCTACATCAAATGAGATGAAG
CinW2AccIWholeRv	GTCGACCTACATCAAATGAGATGAAG
CinW2PLC(K-to-R)rv	GGAAGACGATCCCCTGCCACTCCTACCACTGCCATCGCTGGA
CinW2PLC2(K-to-R)fw	AGGGGAAGCAGGGGGAGTAGCAGGAGTAGTAGGGGAAGCAGTAGTGATG
CinW2Trunc_C-ter_fw	GATGATTCCTCTTCCTCCAAGTCAAGCAAGGG
CinW2Trunc_C-ter_rv	GGCGGCAATGAGGGTGGGGATCG
CinW2Trunc_N-ter_fw	ATGGTGAGCAAGGGCGAGGAGC
CinW2Trunc_N-ter_fw	ATGGTGAGCAAGGGCGAGGAGC
CinW2Trunc_N-ter_rv	ATCGTCAGATCCCCAGTGCGAGG
CinW2XbaI800upFw	TCTAGATTGCACAGACCTAACTAATGGTGTC
CinY2_17aa_afterSPrv	GGGGGTAGGACGGCCTGGGAATGG
CinY2_promoter200_fw	GTCGGAATGGACGAGACGGCGAGGAGGG
CinY2_promoter400_fw	ACTCCGAACGGACAGCACAGTTCTACG
CinY2_Trunc_N-ter(1-72)rv	CAAGGTAGGAAGAGGAGTAGGC
CinY2_Trunc_N-ter(73-136)fw	GGGGGGGTTCCTACGACGAAGATGC
CinY2AccIII300DwRV	TCCGGACTAACAAATACTAATTCGCTGTTGC
CinY2Trunc_C-ter_fw	GACAACTACGCCGACGACTACACC
CinY2Trunc_C-ter_rv	TGCCGAAGCGAGGGTGATTGC
CinY2Trunc_C-ter_rv	TGCCGAAGCGAGGGTGATTGC
CinY2Trunc_N-ter_fw	ATGGTGAGCAAGGGCGAGGAGC
CinY2Trunc_N-ter_fw	ATGGTGAGCAAGGGCGAGGAGC
CinY2Trunc_N-ter_rv	TCCGTATCCGTATCCTGTTGTTAGATGTTGGAGG
D_CinW2_GeneFw(Sil3_PromFw)	CCTATCAACACATCACCATATCAAAAATGAAGCTCGCTCTCTTCCTAACGATCC
D_CinW2_PromRv	CTCAACGATAAACGAACAAAAGAAACC
D_CinY2_TermFw	GTGTTGATGTCTCCTTCAAGTGACACC
D_Sil3_GeneFw(CinW2_PromFw)	CTCAACGATAAACGAACAAAAGAAACCATGAAGACTTCTGCCATTGTATTGC
D_Sil3_PromRv	CCTATCAACACATCACCATATCAAAA
eGFP_rv	TTACTTGTACAGCTCGTCCATGCCG
eGFP_rv	TTACTTGTACAGCTCGTCCATGCCG
eGFP_rv(no stop) for 12162(gib)	CTACTGACGCAACATCCTCTTGTACAGCTCGTCCATGCCG
eGFP_rv(no stop) for 12162(gib)	CTACTGACGCAACATCCTCTTGTACAGCTCGTCCATGCCG
eGFP(no stop)_rv	CTTGTACAGCTCGTCCATGCCG
fcp9_fw_12162(gib)	AGGATGTTGCGTCAGTAGATTAGTTATGCGCCGCTCTTTCTCG

fcp9_up_fw(no r.s.)	GCAACGAATATTTCCAGAAGG
Fcp9promFw	TCCGGAGCAACGAATATTTCCAGAAGGAG
Fcp9termRv2	CTCGAGTAATGGATCTAGGGGATGCT
KSGK_fw	AAGTCTGGAAAGATGGTGAGCAAGGGCGAGGAGC
mRuby_fw	ATGGTCTCCAAGGGAGAAGAGC
mRuby_rv	TTACTTGTACAATTCGTCCATTCCTCC
mRuby(no stop)_rv	CTTGTACAATTCGTCCATTCCTCC
pPhaT1MCSfw	CATATGGAATTCGATATCATCGACTA
pPhaT1MCSfw	CATATGGAATTCGATATCATCGACTA
pPhaT1MCSrv	TCCGGAAAGCTTGCATGCCTGCAGG
pPhaT1MCSrv	TCCGGAAAGCTTGCATGCCTGCAGG
pTp12162_300dw_rv(gib)CORRECT	TGGAAATATTCGTTGCCAATACCAGTCGTTTTATATGG
pTp12162_300up(gib)_rv	CGCCCTTGCTCACCATCCTTCCCTTTCCCGACTTTCC
pTp12162_800up(gib)_fw	AGATTGTACTGAGAGTGCACTTTCGTACTGTATGAGCGTCG
pTp5357_gib_fw	TGGACGAGCTGTACAAGTAAAGGGTGCAAATACTGATTCATGAATTG
pTp5357_gib_fw	TGGACGAGCTGTACAAGTAAAGGGTGCAAATACTGATTCATGAATTG
pTp5357_gib_rv	TCCTCGCCCTTGCTCACCATTGGCACTGCTGCCAAAGTC
pTp5357_gib_rv	TCCTCGCCCTTGCTCACCATTGGCACTGCTGCCAAAGTC
pTpInvNAT_bb_fw	GCCGGTCTCCCTATAGTGAGTCG
pTpInvNAT_bb_rv	GTGCACTCTCAGTACAATCTGC
pTpInvNAT_bb_rv	GTGCACTCTCAGTACAATCTGC
Sil1_KpnI800up_fw	GGTACCCGACTGAAGGACATAGCAGC
Sil1_SalI-300dw_rv	GTCGACGTAGTATGATAGCATAAGTGC
Sil3_800up_fw	CGGCAGGATACTGATGATAGTTATTATTAGG
Sil3_PLC1_no GS_fw	AAGCCCAAAGCAAGCAAGATGGTGAGCAAGGGCGAGGAGC
Sil3_PLC1_no GS_rv	GAAAAGCTTGGCTGCCTTGGCAGCAGTGGTGGCGAGAACG
Sil3_PLC1del_fw	GCTGGTTCCATGCCTGATGAGG
Sil3_PLC1del_rv	TGCAGGGCCAGCAACATCC
Sil3_PLC2_no GS_fw	AAGGGAAAGAGTGGGAAGATGGTGAGCAAGGGCGAGGAGC
Sil3_PLC2_no GS_rv	GAAGATCTTGGCAGCCTTGGCAGCAGTGGTGGCGAGAACG
Sil3_PLC2del_fw	TCCGGGAGTCTCTCCATGC
Sil3_PLC2del_rv	CGAAGAGGGCATGCTAGACTCC
Sil3_promoter200_fw	GCCGGCGCCAATTTCGTGCTTGACTGAACG
Sil3_promoter400_fw	GGAACAAAAATGTCCAGCTATCGTCTGCTCCG
Sil3_SP_rv	GGCAGCAGTGGTGGCGAGAACG
Sil3_T2_rv	GACGTCCTCGGCCTCGGCCTTG
Sil3_T4_fw	ACTGAGATGTCCATGGCAAAGG
Sil3AccI300DwRv	GTCGACCCTCACTGATCCATACGATG
Sil3AccI300DwRv	GTCGACCCTCACTGATCCATACGATG
Sil3EcoRVwholeFw	GATATCCGGCAGGATACTGATGATAG
W2_800up_fw	TTGCACAGACCTAACTAATGG
W2_PLC1(k-to-r)_fw	TCGGGCAGGTCCGGCAGGGGAAGTAGTGG
W2_PLC1(k-to-r)_rv	TCCGGACCTGCCAGATCTGGAACTGG
W2_PLC1del_fw2	GGAAGTAGTGGACCACACGGC
W2_PLC1del_rv2	GGAACTGGAGTAGTCTCCTCCCCATC
W2_PLC2_fw2	AAGCAAGGGGAGTAGCAAGAGTAGTAAGGGAAGCAGTGGTGGCGGAGGGATG GTGAGC
W2_PLC2_rv	ACTGCTTCCCTTACTACTCTTGCTACTCC
W2_PLC2_rv2	CCCTTGGAAGACGATCCCTTGCCACTCTTACCACTGCCGGCGGCAATGAGGGT GGGGATCG
W2_PLC2del_fw	GGAAGCAGTAGTGATGATAGTACGG
W2_PLC2del_rv	ACCACTGCCATCGCTGGAG
W2_PLC3_fw2	CGTCCAAGTCAAGCAAGGGATCGTCGGGTGGCGGAGGGATGGTGAGC
W2_PLC3_rv	CGACGATCCCTTGCTTGACTTGG
W2_PLC3_rv2	ATCCCTTGGAAGACTTGCTGCTTGAGGCGGCAATGAGGGTGGGGATCG
W2_PLC4_fw2	TCGTCCAAGTCATCCAAGGGTTCATCCGGTGGCGGAGGGATGGTGAGC
W2_PLC4_rv	GGATGAACCCTTGGATGACTTGG
W2_PLC4_rv2	TGATCCCTTGCTTGACTTGGATGAGTCGGCGGCAATGAGGGTGGGGATCG
W2_SP+GS+GFP_fw	GGAGGGATGGTGAGCAAGGGCGAGG
W2_SP+GS+GFP_rv	GCCACCGGCGGCAATGAGGGTGGGG
W2-N-ter-PLC1/2del(104-187)_fw	AGCGATGGCAGTGGTGGAAGC
Y2_17aa del_fw	AACCCCACAATGGTGAATACCATCG
Y2_800up_fw	GGTGGTGGTGTTGGCTGCTAGC
Y2_PLC1_fw	GGGTGTGGAAAGAGTGGTAAGG
Y2_PLC1_rv	TCCACCACTCTTGCCACTCTTTCC
Y2_PLC2_fw2	TCTGGCAAGTCGTCCAAAAGCTCCGGCAAGTCGTCCAAGAGCTCTGGCAAGAG TGGAAAGGGAGGAAGTGGTGGCGGAGGGATGGTGAGC
Y2_PLC2_rv	ACTTCCTCCCTTTCCACTCTTGC
Y2_PLC2_rv2	TCCCTTGCTCGACTTGCTTGAGGACTTGCCGCTTTTGCCAGAAGATCCCTTGC CGGACTTGCCTGAGGATGCCGAAGCGAGGGTGATTGC
Y2_SP+GS+GFP_rv	GCCACCTGCCGAAGCGAGGGTGATTGC
SEQUENCING PRIMERS	
pJet1.2 fw	GCACAAGTGTTAAAGCAGTT

pJet1.2 rv	CTCTCAAGATTTTCAGGCTGTAT
pPha fw	GGCGTATCACGAGGCCCTTTCG
pPha_5'vorNde	GCTTAACTATGCGGCATCAG
pTpNAT-MCS200dw	TGCAGTGTGGAGGTGAATGATCGG
pTpNAT/MCSseqRV	GTTCAACTCTCTAAACTTCTCC

7.3 PROTEIN SEQUENCES

7.3.1 CINW2 CONSTRUCTS

>CinW2 full length (383 aa)
MKLALFLTIPTLIAAQQSSVRGVATTSSRQLDEWGDDAWGSSDSGSSGKSGKSGGSASSGDGWETDGWGGDYSSSKSGKS
GSGKSGKGSSGPHGHWVYIEDDSSDGSGKSGKGSSSKGSKGSSKSSKGSSSDDSTDDSWDGGWGGHGGWNGDNSGKSGKG
SYGSGKSGKGSSYPSSHWGPSHWGSDDDDSSSSKSSKGSSESSSKSSKGSSDSSSKSSKGSSSSEDEGHWEWEGGYGSGK
SGKGSYSGSSGKSGKSGSGDSWVGDYGSSGKSGKGSYGGDSWGGNYNGWGGHYDVDVDDDDSSSSKSSKGSSKSSKGSSE
DSSKSSKGSSSKSSKGSSSSEDEGHWVWEGSYGSGKSGKGSYSGSSGKSGKSGSGDEGWYSGW

>CinW2 SP (15 aa)
MKLALFLTIPTLIAA

>CinW2 TR-A (187 aa)
MKLALFLTIPTLIAAQQSSVRGVATTSSRQLDEWGDDAWGSSDSGSSGKSGKSGGSASSGDGWETDGWGGDYSSSKSGKS
GSGKSGKGSSGPHGHWVYIEDDSSDGSGKSGKGSSSKGSKGSSKSSKGSSSDDSTDDSWDGGWGGHGGWNGDNSGKSGKG
SYGSGKSGKGSSYPSSHWGPSHWGSDD

>CinW2 TR-B (15/196 aa)
MKLALFLTIPTLIAADDSSSSKSSKGSSESSSKSSKGSSDSSSKSSKGSSSSEDEGHWEWEGGYGSGKSGKGSYSGSSGK
SGKSGSGDSWVGDYGSSGKSGKGSYGGDSWGGNYNGWGGHYDVDVDDDDSSSSKSSKGSSKSSKGSSEDSSKSSKGSSSK
SSKGSSSSEDEGHWVWEGSYGSGKSGKGSYSGSSGKSGKSGSGDEGWYSGW

>CinW2 TR-C (103 aa)
MKLALFLTIPTLIAAQQSSVRGVATTSSRQLDEWGDDAWGSSDSGSSGKSGKSGGSASSGDGWETDGWGGDYSSSKSGKS
GSGKSGKGSSGPHGHWVYIEDDS

>CinW2 TR-D (15/84 aa)
MKLALFLTIPTLIAASDGSGKSGKGSSSKGSKGSSKSSKGSSSDDSTDDSWDGGWGGHGGWNGDNSGKSGKGSYGSGKSG
KGSSYPSSHWGPSHWGSDD

>CinW2 MUT-A (187 aa)
MKLALFLTIPTLIAAQQSSVRGVATTSSRQLDEWGDDAWGSSDSGSSGKSGKSGGSASSGDGWETDGWGGDYSSSRSGRS
GSGRSGRGSSGPHGHWVYIEDDSSDGSGKSGKGSSSKGSKGSSKSSKGSSSDDSTDDSWDGGWGGHGGWNGDNSGKSGKG
SYGSGKSGKGSSYPSSHWGPSHWGSDD

>CinW2 MUT-B (187 aa)
MKLALFLTIPTLIAAQQSSVRGVATTSSRQLDEWGDDAWGSSDSGSSGKSGKSGGSASSGDGWETDGWGGDYSSSKSGKS
GSGKSGKGSSGPHGHWVYIEDDSSDGSGRSGRGSSSRGSRGSSRSSRGSSSDDSTDDSWDGGWGGHGGWNGDNSGKSGKG
SYGSGKSGKGSSYPSSHWGPSHWGSDD

>CinW2 MUT-C (187 aa)
MKLALFLTIPTLIAAQQSSVRGVATTSSRQLDEWGDDAWGSSDSGSSGKSGKSGGSASSGDGWETDGWGGDYSSSRSGRS
GSGRSGRGSSGPHGHWVYIEDDSSDGSGRSGRGSSSRGSRGSSRSSRGSSSDDSTDDSWDGGWGGHGGWNGDNSGKSGKG
SYGSGKSGKGSSYPSSHWGPSHWGSDD

>CinW2 DEL-A (175 aa)
MKLALFLTIPTLIAAQQSSVRGVATTSSRQLDEWGDDAWGSSDSGSSGKSGKSGGSASSGDGWETDGWGGDYSSSGSSGP
HGHWVYIEDDSSDGSGKSGKGSSSKGSKGSSKSSKGSSSDDSTDDSWDGGWGGHGGWNGDNSGKSGKGSYGSGKSGKGSS
YPSSHWGPSHWGSDD

>CinW2 DEL-B (168 aa)
MKLALFLTIPTLIAAQQSSVRGVATTSSRQLDEWGDDAWGSSDSGSSGKSGKSGGSASSGDGWETDGWGGDYSSSKSGKS
GSGKSGKGSSGPHGHWVYIEDDSSDGSGGSSSDDSTDDSWDGGWGGHGGWNGDNSGKSGKGSYGSGKSGKGSSYPSSHWG
PSHWGSDD

>CinW2 DEL-C (156 aa)
MKLALFLTIPTLIAAQQSSVRGVATTSSRQLDEWGDDAWGSSDSGSSGKSGKSGGSASSGDGWETDGWGGDYSSSGSSGP
HGHWVYIEDDSSDGSGGSSSDDSTDDSWDGGWGGHGGWNGDNSGKSGKGSYGSGKSGKGSSYPSSHWGPSHWGSDD

>CinW2 TR-C DEL (91 aa)
MKLALFLTIPTLIAAQQSSVRGVATTSSRQLDEWGDDAWGSSDSGSSGKSGKSGGSASSGDGWETDGWGGDYSSSGSSGP
HGHWVYIEDDS

>CinW2 TR-D DEL (15/65)
MKLALFLTIPTLIAASDGSGGSSSDDSTDDSWDGGWGGHGGWNGDNSGKSGKGSYGSGKSGKGSSYPSSHWGPSHWGSDD

>W2-LER1
MKLALFLTIPTLIAASSSKSGKSGSGKSGKGSS

>W2-LER2
MKLALFLTIPTLIAAGSGKSGKGSSSKGSKGSSKSSKGSS

>W2-LER3
MKLALFLTIPTLIAASSSKSSKGSSKSSKGSS

>W2-LER4
MKLALFLTIPTLIAADSSKSSKGSSSKSSKGSS

7.3.2 CinY2

>CinY2 full length (248 aa)
MKLIIALSAITLASAGTNKTLPPTPFPGRPTPNPTMVNTIGTPGPSFIVTEQTPAPTPGDVLTPQPTPLPTLGGVPTTKM
PTEMSYGYGYGDYGIVDCFGKSGKSGSGCGKSGKGSKSSGKSGKSGGGGGGGYGYGDNYADDYTPSTDDYEYGYGHGGSS
GKSGKGSSGKSGKSSSKSSKGSGKSSKSSGKSSKSSGKSGKGGSRDDGHGYGGYGGYEGYGGYEGYQYGGDEYVRRNRRL
GASHNNRI

>CinY2 SP (15 aa)
MKLIIALSAITLASA

>CinY2 TR-A (136 aa)
MKLIIALSAITLASAGTNKTLPPTPFPGRPTPNPTMVNTIGTPGPSFIVTEQTPAPTPGDVLTPQPTPLPTLGGVPTTKM
PTEMSYGYGYGDYGIVDCFGKSGKSGSGCGKSGKGSKSSGKSGKSGGGGGGGYGYG

>CinY2 TR-B (15/112 aa)
MKLIIALSAITLASADNYADDYTPSTDDYEYGYGHGGSSGKSGKGSSGKSGKSSSKSSKGSGKSSKSSGKSSKSSGKSGK
GGSRDDGHGYGGYGGYEGYGGYEGYQYGGDEYVRRNRRLGASHNNRI

>CinY2 TR-C (72 aa)
MKLIIALSAITLASAGTNKTLPPTPFPGRPTPNPTMVNTIGTPGPSFIVTEQTPAPTPGDVLTPQPTPLPTL

>CinY2 TR-D (15/64 aa)
MKLIIALSAITLASAGGVPTTKMPTEMSYGYGYGDYGIVDCFGKSGKSGSGCGKSGKGSKSSGKSGKSGGGGGGGYGYG

>CinY2 UM (15/17 aa)
MKLIIALSAITLASAGTNKTLPPTPFPGRPTP

>CinY2 UM DEL (15/231 aa)
MKLIIALSAITLASANPTMVNTIGTPGPSFIVTEQTPAPTPGDVLTPQPTPLPTLGGVPTTKMPTEMSYGYGYGDYGIVD
CFGKSGKSGSGCGKSGKGSKSSGKSGKSGGGGGGGYGYGDNYADDYTPSTDDYEYGYGHGGSSGKSGKGSSGKSGKSSSK
SSKGSGKSSKSSGKSSKSSGKSGKGGSRDDGHGYGGYGGYEGYGGYEGYQYGGDEYVRRNRRLGASHNNRI

>Y2-LER1 (15/20 aa)
MKLIIALSAITLASAGCGKSGKGSKSSGKSGKSGG

>Y2-LER2 (15/46)
MKLIIALSAITLASASSGKSGKGSSGKSGKSSSKSSKGSGKSSKSSGKSSKSSGKSGKGGS

7.3.3 SIL3

>Sil3 full length (231 aa)
MKTSAIALLAVLATTAATEPRRLRTLEGHGGDHSISMSMHSSKAEKQAIEAAVEEDVAGPAKAAKLFKPKASKAGSMPDE
AGAKSAKMSMDTKSGKSEDAAAVDAKASKESHMSISGDMSMAKSHKAEAEDVTEMSMAKAGKDEASTEDMCMPFAKSDKE
MSVKSKQGKTEMSVADAKASKESSMPSSKAAKIFKGKSGKSGSLSMLKSEKASSAHSLSMPKAEKVHSMSA

>Sil3 SP (17 aa)
MKTSAIALLAVLATTAA

>Sil3 DEL (204)
MKTSAIALLAVLATTAATEPRRLRTLEGHGGDHSISMSMHSSKAEKQAIEAAVEEDVAGPAAGSMPDEAGAKSAKMSMDT
KSGKSEDAAAVDAKASKESHMSISGDMSMAKSHKAEAEDVTEMSMAKAGKDEASTEDMCMPFAKSDKEMSVKSKQGKTEM
SVADAKASKESSMPSSSGSLSMLKSEKASSAHSLSMPKAEKVHSMSA

>Sil3 DEL T2 (132 aa)
MKTSAIALLAVLATTAATEPRRLRTLEGHGGDHSISMSMHSSKAEKQAIEAAVEEDVAGPAAGSMPDEAGAKSAKMSMDT
KSGKSEDAAAVDAKASKESHMSISGDMSMAKSHKAEAEDV

>Sil3 DEL T4 (17/99 aa)
MKTSAIALLAVLATTAATEMSMAKAGKDEASTEDMCMPFAKSDKEMSVKSKQGKTEMSVADAKASKESSMPSSSGSLSML
KSEKASSAHSLSMPKAEKVHSMSA

>Sil3 PLC3nt (17/12 aa)
MKTSAIALLAVLATTAAKAAKLFKPKASK

>Sil3 PLC3ct (17/12 aa)
MKTSAIALLAVLATTAAKAAKIFKGKSGK

>Sil3 KSGK (17/4 aa)
MKTSAIALLAVLATTAAKSGK

7.3.4 OTHER PROTEINS

>Sil1 (501 aa)
MKVTTSIITLLFASCGAADVQRVLEDVTEPAVTTPAATPAPITPEPATPAPTICEGRNFYYDEETRKCSNEATGGIYGTL
IDCCVAISGSVSCPYVDICNTLQPSPSPETNEPSAKPITAAPISSAPVSAAPVTSAPVAAPVETTSMTGPTTIVASIVST
NAPSLTNAPSSSLEAVVTRIPVETTNTASPTTTAASIVSTNAPSSSPEAVVTPRPTFRPSPEGTESNTSPASIASDVMFG
PPKTSTPTSTPTSSSHPSSSEPTLSPSVSKEPTGYPTSSPSHSPTKSPSKSPSSSPTTSPSASPTETPTETPTESPTESP
TESPTLSPTESPTLSPTESPSLSPTLSTTWSPTGYPTLAPSPSPSISSAPSVSSAPSSPPSISSAPSVSSAPSKNFGFLP
GLTEMPTISPTEDHYFFGKSHKSHKSHKSKATKTLKVSKSGKSAKSSKSSGRRPLFGVSQLSEGIAVGYAKSSGRSSQQA
VGSWMPVAAACILGALSFFLN

>Sil2 (485 aa)
MKVTTSIITLLFASCGAADVQRVLEDVTEPAVTTPAATPAPITPEPATPAPTICEGRNFYRDDDTGKCSNEATGGIYGTL
IECCVAISGSDSCPYVDICNTLQPSPSPETNEPSAKPITAAPISSAPVSAAPVTSAPVAAPVETTSMTGPTTIVASIVST
NAPSSTNAPSSSLEAVVTRIPVETTNTASPTTTAASIVSTNAPSSSPEAVVTPRPTFRPSPKGTESNTFPASIASDVMFD
PARSEPTFTPTSSSQPSSSEPTLSPSVSKEPTRYPTSSPSHSPTKSPSKSPSSSPTTSPSASPTETPTETPTESPTELPT
LSPTEFPSLSPTLSPTWSPTGYPTLAPSPSPSISSAPSVSSAPSSSPSISSAPSVSSAPSKNFGFLPGRNEMPTISPTED
HYFFGKSHKSHKSKATKTLKVSKSGKSSKSSKSSGSRPLFGVSQLSEGIAAGYAKSSGRSSQQAVGSWMPVAAACILGAL
SFFLN

Comparison between the PLC from Sil1 (PLC1) and Sil2 (PLC2)

PLC1 KSHKSHKSHKSKATKTLKVSKSGKSAKSSK
PLC2 KSHKSHK---SKATKTLKVSKSGKSSKSSK

>12162(415 aa)

MKHHAILLTALLALSLTTPSVEAADSNPTKKNIKAVKKLLRKKIRKQRKKNMKEDAATNAAATTTTSTAAAVVSKNNTID
GNKPTWSPTMWPTYWPTYSPTNEKMSEETGESIPVSAPVTKQPTPKPSKAPTPSPSKAALVQSQVEIPVANPDSAQTNLN
NLVFISVLVNDVPAMGQSLDVFAIASNAANGNCAVSLNIQQVVYTPNDGFVGFDECTYEACDGMEMCDTAIVTIEVLKEE
NTPVPTFANTRVNPTPFPSQDVITFEPTIASGTNKTLTPTVTPLRPTPFPTENTPAPTFIVDTPTTRLPTGYTDEKYYTD
DYGYSMPYLSHIDLSSKSGKGGKGSKSGSKSAKSSSKGSKSSGKSGKSGSWKGDMHGDYGDDDGYTFDDGYAHDDGWYGG
EAGGKSGKGRRMLRQ

>12163 (1155 aa)

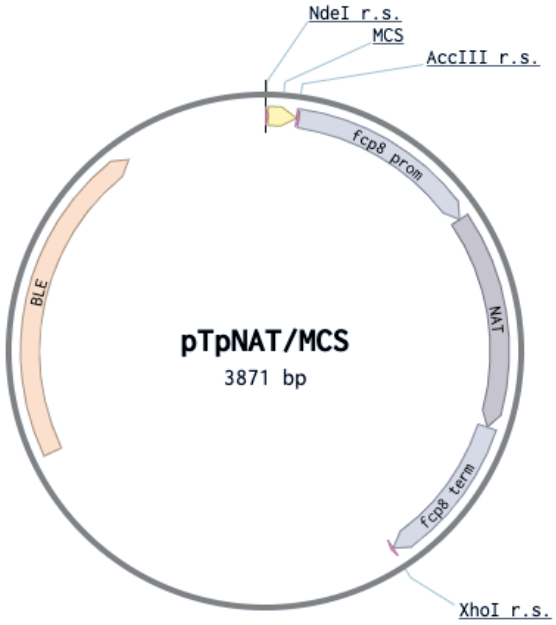
MISLLLVSTTFALVSAGTNKTLPPTVTPGRPTQTPTESPPTLFPTFVPTSGYNADKPYYGGSSMSYSYWDTLERISLGKS
AKAKGSKGTSKSGKAIQHMVWGDDGFDTYEKYDDNYVSEDDYHYGGGGSWHASSKSGKAKSGKAGSKSSKGSKSDKSGKV
SSKSGKHSSTMSRSSSKSGKSNGGSGHYNKPIVTMPYYYGENNWSGYGGNSKIEQHHPLQIPPNTQDPDYFHPSAVGGNA
YKQGKDGAAVIVSKLFSKSMDEPLELRANFFFVNNVLQEYEIPGDRFSEPYDQVVTIKVWGAGGGGCGGGRGEETEPSSR
LDEETGIVYLSEGFGGGYVTASFKLPIGETLQITIGGGGNTEGSQSDSLGGSGGYNGGFPGKDDGKSGGGGGGGMSTVEW
NGTILVAAYGGDGGGNTTYCSAYGGRGGSLRGQTKDAGGYVNTNLTLVDNDVKSGAQCPSEPLAFHLSHDSASFTWNAGS
NQHSTSKEFYVHKYIVRLSTAELERDGEETNFCSHYDVHEHILSDPTRSNDNATTSMSNLHPQTSYCLQVEAFSVEGLSL
GKRVLPFTTRPAPVNEWIKMDVRQTVPSDISANSVNDEQSWCEYSSIRPTGRRGHSMAVINDQAYIFGGATVKCGCSYDS
FFNETRCSRKNVYSDELWQFDTISSRFSLLHDMSSIDETTPRGREQHSATEMPDGSLVVIGGMTSFNDDLTIDQESNVLG
DVWRLPNPHHTSSFVFSGIGDEARESQAMELTSGKISKHRLYVSSEEDVEMCIADIQVKISLDHDCPKDIHYISLIGPDN
SSGINAHESRYRETKLYISSIGSQKNDCHASTLNLVFSDKAYDSILSYSDIPTAGEYRPANSLMSMFDSWPVDGEWTVGI
AQSQSSHLEEHRGRVLDWELRVKGKPCHAKANWQQLSRQSNFPPRRMHTAVAIGNSIFVSGGFSDHHLDDLWRFDYGPNT
WTELNASASGKPLPMNGQTALLGQFGLLVYGGIQPLGTKEHLADLWLLDVFENEWSKVPIARNHSMPRGRYLSTIGMCKS
HDVDRMNHVVIVFGGDGGLLHNSYQDSYGFMSNAFFQDVWMLSLGGISGSMKERHREEHCDWRLSLNSTAQREWNASCGW
SESNNSSPDECSWDEILVMAWCKEQYQSFYMEYAH

>5357(593 aa)

MRVLCLLAFLQVTSFAVSNEVDSELPSRKHLLRNTRDPQQRSLQQKEYRLVKPKDRRVVRKERNSKLAREDDGERTEVIP
ELERASNGGTFEFEVDDEESVNVVVTVNVETTATEEEEPSEDVDEMCPEWEIVYVKQIQHQHHHHASKSGKGGKSSKSTK
SHKSKAGKSVKEVPVKKCKTTKPTSKPAPVTPMPIGGATPIPANITPDPTDVPPTSTTSSPTKAGDNITSNPTIDSIETS
APVASAPPNPVPTDVPVLAPTMPPMSGPLPTLPPAEATSSPTVSVPQPTDVPISVAPTMPPMSGPLPTLPPAEATSSPTV
SVPQPTDVPISVAPTMPPMSALTSAPVATGSVSPTVNNQTSASPTVNNETIVTDAPSAAPSTTTATDAPSVSAAPSTGEK
QPTESPTYMPSYIPTKAPVAGIVPSTAPTTGSGLTTLSWDFEDQQFPIDPWTTGGDGVWAIDDTNADGGSVYSLKSPDLE
DDTNPSPQLSNATLTLSDTFAGGVMRMRVLASVLPPRDIFVVYIDGESAAQLVDVQEFTDVVLGVGPGAHVIDFSYQYNF
FGVDPLPPSPPSRIGSVWIDNVSIETLAAVP

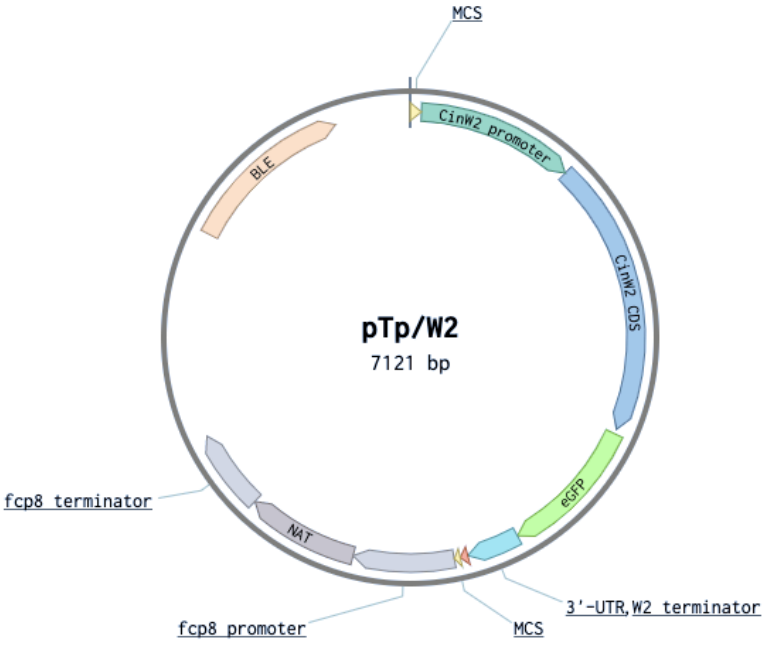
7.4 PLASMID MAPS AND CONSTRUCTS

7.4.1 PTPNAT-MCS



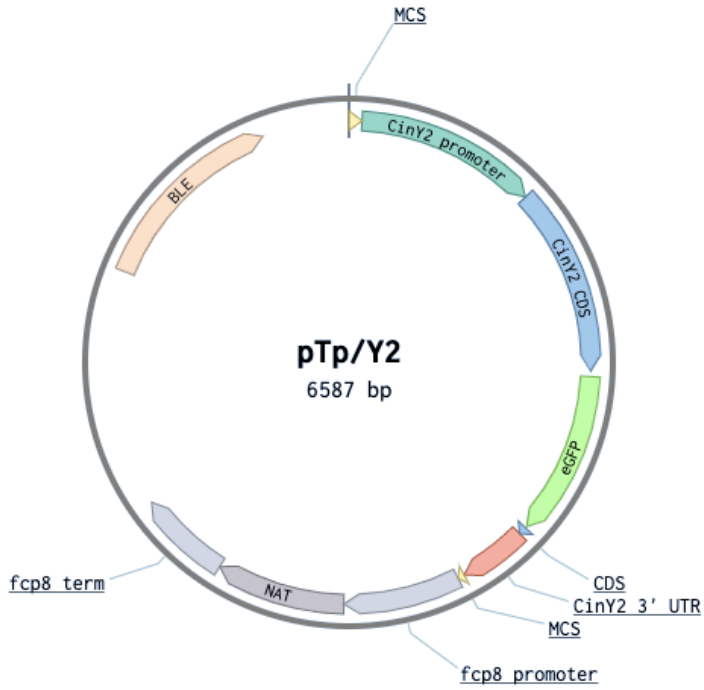
feature	position	Length
MCS	7-81	75 bp
fcp8 promoter	88-600	513 bp
nat CDS	601-1169	569 bp
fcp8 terminator	1170-1578	409 bp
ble	2629-3488	860 bp

7.4.2 PTPCINW2



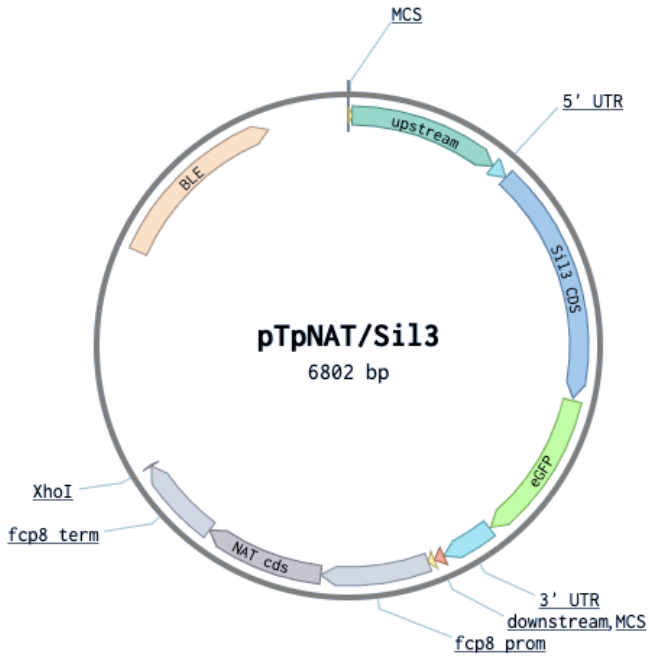
feature	position	Length
MCS	1-57	57 bp
CinW2 promoter	58-857	800 bp
CinW2 CDS	858-2255	1398 bp
eGFP	2280-2999	720 bp
CinW2 3'-UTR	3000-3268	269 bp
CinW2 terminator	3269-s3307	39 bp
MCS	3308-3337	30 bp
fcp8 promoter	3338-3850	513 bp
nat CDS	3851-4419	569 bp
fcp8 terminator	4420-4828	409 bp
ble	5879-6739	861 bp

7.4.3 pTpCinY2



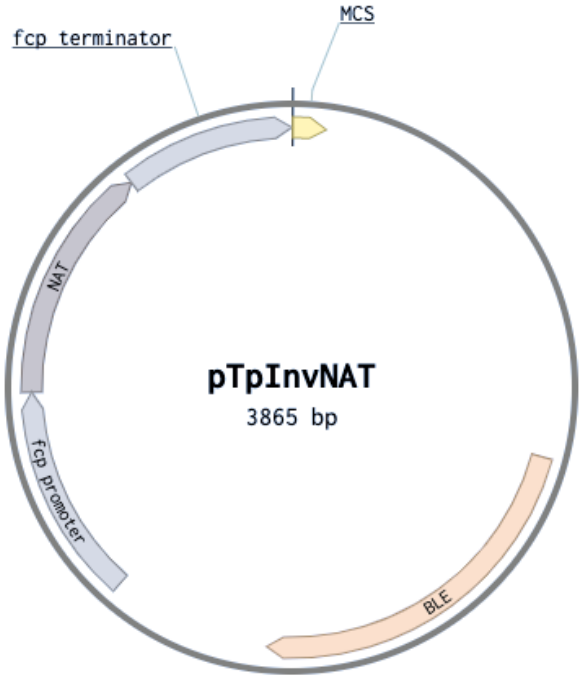
feature	position	Length
MCS	1-57	57 bp
CinY2 promoter	58-857	800 bp
CinY2 CDS	858-2255	1684 bp
eGFP	1709-2428	720 bp
CinY2 3'-UTR	2473-2773	301 bp
MCS	2774-2803	30 bp
fcp8 promoter	2804-3316	513 bp
nat CDS	3317-3885	569 bp
fcp8 terminator	3886-4294	409 bp
ble	5345-6205	861 bp

7.4.4 pTpSil3



feature	position	Length
MCS	1-57	57 bp
CinY2 promoter	58-857	800 bp
CinY2 CDS	858-2255	1684 bp
eGFP	1709-2428	720 bp
CinY2 3'-UTR	2473-2773	301 bp
MCS	2774-2803	30 bp
fcp8 promoter	2804-3316	513 bp
nat CDS	3317-3885	569 bp
fcp8 terminator	3886-4294	409 bp
ble	5345-6205	861 bp

7.4.5 pTpINV-NAT



feature	position	Length
fcp8 promoter	1-513	513 bp
NAT	514-1082	569 bp
fcp8 terminator	1083-1491	409 bp
ble	2623-3482	861 bp

ACKNOWLEDGMENTS

First, I want to thank Prof. Dr. Uwe Maier, not only for having entrusted me with his project, but also for letting me work independently and because he has always supported me, no matter if small or bigger issues have appeared on the path.

I would like to thank the members of my thesis evaluation committee: Prof. Dr. Alfred Batschauer, Prof. Dr. Andreas Brune, and Prof. Dr. Lars Voll for their availability and for their experienced review on my thesis work.

I would like to thank Dr. Stefan Zauner for all the time he spent answering the silliest questions about basically every aspect of molecular biology. I would like to thank Dr. Daniel Moog, Dr. Franziska Hempel, and Dr. Thomas Heimerl, for being always there to help me with their vast knowledge and smart suggestions or criticisms (and, Thomas, also for letting me play with the TEM); I would like to thank Elmar Utesch, Angela Zimmer and Marion Schön, for being always so nice and available, and for making everybody feel a little bit more at home.

I would like to thank my PhD colleagues Jana, Vicki and Stephie, you all helped me staying focused on the target and never mind about the problems on the road. A special thanks goes to Gianluca, I cannot imagine my PhD without you sitting at your desk next to mine. I also would like to thank all the bachelors, master students, and trainees that have worked in our lab, because they made me perceive once again the nice atmosphere of the student life, full of jokes, pranks and fun.

I would like to also thank my parents and my in-laws, for all the useful help they provided, allowing me to have more time for writing and for analysing the data. Finally, I want to thank my wife Giulia and my two little monsters Lorenzo and Jana, for allowing me see everything from a better perspective.

Erklärung

Hiermit versichere ich, dass ich meine Dissertation

“Studying the targeting of frustule-associated proteins in the centric diatom *Thalassiosira pseudonana”*

Selbstständig, ohne unerlaubte Hilfe angefertigt und mich dabei keiner anderen als den von mir ausdrücklich bezeichneten Quellen bedient habe.
Die Dissertation wurde in der jetzigen oder einer ähnlichen Form noch bei keiner anderen Hochschule eingereicht und hat noch keinen sonstigen Prüfungszwecken gedient.

Marburg, den 30.12.2021 Neri Fattorini