Characterisation of monogalactosyldiacylglycerol synthases in the model diatom Phaeodactylum tricornutum

Nolwenn Gueguen

Université Grenoble Alpes

THÈSE

Pour obtenir le grade de

DOCTEUR DE L'UNIVERSITÉ GRENOBLE ALPES

Spécialité : Biologie cellulaire

Arrêté ministériel : 25 mai 2016

Présentée par

Nolwenn GUEGUEN

Thèse dirigée par **Eric MARECHAL**

préparée au sein du **Laboratoire LPCV - Laboratoire de Physiologie Cellulaire Végétale**
dans **l'École Doctorale Chimie et Sciences du Vivant**

Caractérisation des monogalactosyldiacylglycérol synthases chez la diatomée modèle *Phaeodactylum tricornutum*

Characterisation of monogalactosyldiacylglycerol synthases in the model diatom *Phaeodactylum tricornutum*

Thèse soutenue publiquement le **13 mai 2022**,
devant le jury composé de :

Monsieur Eric MARECHAL
DIRECTEUR DE RECHERCHE, CNRS délégation Alpes, Directeur de thèse
Madame Yonghua LI-BEISSON
DIRECTRICE DE RECHERCHE, CEA centre de Cadarache, Rapporteure
Madame Corinne MERCIER
PROFESSEURE DES UNIVERSITES, Université Grenoble Alpes, Présidente
Madame Brigitte GONTERO
DIRECTRICE DE RECHERCHE, CNRS délégation Provence et Corse, Rapporteure
Monsieur Frédéric DOMERGUE
CHARGE DE RECHERCHE HDR, CNRS délégation Aquitaine, Examinateur
Monsieur Alberto AMATO
CHERCHEUR, CEA centre de Grenoble, Invité

Acknowledgments

My thesis at LPCV has allowed me to meet and exchange with many amazing people that I would like to thank for the production of this work. However, I cannot do so without also introducing the context in which I started my thesis. Indeed, this adventure did not start at LPCV 3.5 years ago, but 5.5 years ago when I began a thesis in another laboratory. For a number of reasons, I did not thrive neither as a person, nor as an apprentice researcher in this former thesis. This experience ended after almost 2 years, following the demission of my former thesis director, at which time I started to look for a new opportunity, which I found at LPCV in the LIPID team.

Therefore, I would like to address a huge thank-you to Éric Maréchal and Juliette Jouhet, who answered positively after taking notice of my situation, and offered me the great opportunity to join their team as a PhD student. This thesis will definitely stay as one of my best working experiences. I would like to personally thank Éric, through whom I received the funding to realise these extra three years of PhD, and later provided me with a 6 months extension to my contract following COVID-19 quarantine. Éric has been a wonderful mentor, who despite his responsibility has head of the laboratory has always been available for me. Éric has a passion for science that he is eager to communicate and led to many long and exciting talks in his company. He has been to me a truly open and empathetic person (although he may deny it!), gifted with a good (and sometimes scary!) sense of humour. I would like to thank him again, for always cheering me during my thesis, and wish that every PhD student could have such a great director!

I wish to thank Félix Cicéron, former post-doctoral student in the team, for he taught me everything about cultivating algae and performing CRISPR-Cas9 editing. I spent the first months of my thesis assisting him on what would become my PhD project, and he became one of the first friends I made in Grenoble. I also need to thank Alberto Amato and Juliette Salvaing, for all the complementary supervision they provided me on a weekly basis, their help in setting up new experiments and also for all the fun we had in our meetings.

I would like to thank again Juliette Jouhet, not only in her quality of team leader but also for all the help she provided me in lipid analysis. Grégory Si Larbi, for his help in the observation of my mutants by TEM, our scientific and non-scientific discussions, and through whom I discovered the insane world of intermittent fasting. Catherine Albrieux, who assisted me many times and in particular at the end of my PhD to finish in time several pending experiments. She too, has been not only a co-worker but also a friend with whom we had many laughs working. I also thank Valérie Gros, for teaching me how to extract and treat lipids with the most enjoyable protocols.

I would like to thank Dimitris Petroutsos, for his help in setting up photosynthesis experiment and interpretate their signification. Hanhua Hu and Yangmin Gong, for engineering and sending us *P. tricornutum* MGD-eGFP overexpressing strains. Giovanni Finazzi, for our discussion on MGDα particular localisation. Serena Flori, for our talks about the mysterious blob and how to recognize it. Lionel Imbert, for his guidance for protein production in cell free system. And Dylan Jabeguero, for assisting me in trying again and again to detect MGD activity at CERMAV. It was great to meet another PhD student working on an MGD and quoting the same papers!

Of course, I want to thank all the students that shared the office with me: Stéphanie, Adeline, Damien, Sébastien, Sarah, Baptiste and Pauline, who made of the office a great place to work and exchange. And in particular

Stephanie Bolik (Steph), the crazy person sitting on my left, who became a great friend and despite herself, knows me more than a lot of people. A personal thank to Damien Le Moigne too, who many times helped me take care of my algae when I was writing this manuscript.

A vast thank to every member of the incredible LIPID team that I did not mention, for their energy, the scientific discussions, the occasional cakes, and the football sessions with several of them. This would have been the perfect team if it were not for the broken leg curse that has hovered over us for the past few years!

A great thank to all the non-permanent people, for the organisation of scientific and non-scientific events, and the great solidarity between the students. A particular thank to Georgios and Mattia, for all the weekends spent playing boardgames at my place.

Special thanks to my friends from my former laboratory: 'Javi', Vincent, 'Jay', 'Joe', Heba, Stefan, and Remy, with whom we stick together. To Claude Maranges from INSA Toulouse, who besides having many responsibilities also acts as a godfather to all biochemical engineers coming out of the school, and always cheered me up when I was feeling down in my former thesis. And finally, to Etienne-Pascal Journet and Fayza Daboussi, who encouraged me to consider the LPCV when I was looking for an opportunity in the field of microalgae.

Finally, a big thank to my parents, sisters and brother, who supported me during these 5.5 years (and more). And to my fiancé, Mauro di Vito, who really tried hard to understand my rambling talks about algae, membranes, and lipids.

Table of contents

List of abbreviations

ACCase: acetyl-CoA carboxylase

ACK: acetate kinase

ACL: ATP-citrate lyase

ACP: acyl carrier protein

ACS: acetate synthase

ACSL: long-chain acyl-CoA synthases

ADS: stearoyl-CoA desaturase

ATS1: glycerol-3-phosphate acyltransferase 1

ATS2: 1-acylglycerol-3-phosphate-acyltransferase

BLAST: Basic Local Alignment Search Tool

BSA: bovine serum albumin

BTA: betaine lipid synthase

CalG1: Calicheamicin Glycostyltransferase 1

Cas9: CRISPR-associated protein 9

Catdom: catalytic domain

cDNA: complementary DNA

CDP: cytidine diphosphate

CDS: coding sequence

CIP: Calf Intestinal Phosphatase

CMC: critical micelle concentration

CMP: cytidine monophosphate

CoA: coenzyme A

CRIPSR: Clustered Regularly Interspaced Short Palindromic Repeats

Ct: cycle threshold

CTP: cytidine triphosphate

cTP: chloroplast transit peptide

CHAPS 3-[(3-cholamidopropyl)dimethylammonio]-1-propanesulfonate

DAG: diacylglycerol

DEL: deletion

DGAT: diacylglycerol acyltransferase

DGCC: 1,2-diacylglyceryl-3-O-carboxy-(hydroxymethyl)-choline

DGD: digalactosyldiacylglycerol synthase

DGDG: digalactosyldiacylglycerol

DGTA: diacylglycerylhydroxymethyltrimethyl-β-alanine

DGTS: diacylglyceryltrimethylhomoserine

DHA: docosahexaenoic acid

eGFP: enhanced Green Fluorescent Protein

ELO: elongase

ENR: enoyl-ACP reductase

EPA: Eicosapentaenoic acid

EpM: epiplastidial membrane

ER: endoplasmic reticulum

ERAD: Endoplasmic Reticulum Associated protein Degradation

ESAW: Enriched Seawater, Artificial Water

FA: Fatty acid

FAME: fatty acid methyl ester

FAS: fatty acid synthase

FAT: fatty acyl-ACP thioesterase

FCP: Fucoxanthin-chlorophyll a/c binding protein

FITC: fluorescein isothiocyanate

GC-FID: gas chromatography coupled to flame ionisation detection

GH: galactosyl hydrolase

GOI: gene of interest

GPAT: glycerol-3-phosphate-sn-1-acyl-CoA-acyltransferase

GT: glycosyltransferase

HAD: hydroxyacyl-ACP dehydratase

HexII: Hexagonal II

HPRT: hypoxanthine-guanine phosphoribosyltransferase

HRP: Horseradish Peroxidase

iEM: inner Envelope Membrane of primary plastids

iNE: inner Nuclear Membrane

INDEL: Insertion/Deletion

INS: insertion

KAR: 3-ketoacyl-ACP-reductase

KAS: ketoacyl-ACP synthase

kDa: kilo Dalton

KO: knockt-out

LACS: long-chain acyl-CoA synthases

LARP6C: La-related protein 6C

LAT1 : Lysolipid Acyltransferase 1

LC-MS/MS : Liquid chromatography coupled to tandem mass spectrometry

LD: lipid droplet

LHCF: Light Harvesting Complex containing Fucoxanthin

Lm: lamellar

LPA: lysophosphatidic acid

LPAT: lysophosphatidic acid acyltransferase

LPCAT: acyl-CoA:lysoPC acyltransferase

LB: Lysogeny Broth

MCMT: malonyl-CoA:ACP transferase

MDP: MGDG/DGDG/PG

MGD: monogalactosyldiacylglycerol synthase

MGDG: monogalactosyldiacylglycerol

MGlcD: monoglucosyldiacylglycerol synthase

MGlcDG: monoglucosyldiacylglycerol

mRNA: messenger RNA

MuRG: UDP-N-acetylglucosamine--N-acetylmuramyl-(pentapeptide) pyrophosphoryl-undecaprenol N-acetylglucosamine transferase

NPQ: non-photoquemical quenching

oEM: outer enveloppe membrane of primary plastid

oNE: outer Nuclear Membrane

OleI: Oleandomycin glycosyltransferase I

PA: phosphatidic acid

PAD: palmitoyl-ACP desaturase

PAM: protospacer adjacent motif

PAP: phosphatidic acid phosphatase

PAT: phosphateacetyltransferase

PC: phosphatidylcholine

Pchlide: protochlorophyllide

PCR: Polymerase Chain Reaction

PDAT: phospholipid:diacylglycerol acyltransferase

PDCT: phosphatidylcholine:diacylglycerol choline phosphotransferase

PDH: pyruvate dehydrogenase

PDK: pyruvate dehydrogenase kinase

PE: phosphatidylethanolamine

PG: phosphatidylglycerol

PGD1: plastid galactoglycerolipid degradation 1

PGP: phosphatidylglycerol phosphate

PGPP: phosphatidylglycerol phosphate phosphatase

PGPS: phosphatidylglycerol phosphate synthase

PI: phosphatidylinositol

PNPLA3: phospholipase domain-containing protein 3

PPC: periplastidial Compartment

PPM: periPlastidial Pembrane

PS: photosystem

PTM: post-translational modification

PUFA: polyunsaturated fatty acid

QC: qualified control

SAD: stearoyl-ACP desaturase

SELMA: 'symbiont-specific ERAD-like machinery'

sgRNA: single guide RNA

SQDG: sulfoquinovosyldiacylglycerol

STEM: scanning transmission electron microscopy

TAG: triacylglycerol

TE: thioesterase

TEM: transmission electron microscopy

TIC: translocon at the iEM

TOC: translocon at the oEM

UDP: uridine diphosphate

UDP-Gal: uridine diphosphate-galactose

VLC-PUFA: Very Long chain PolyUnsaturated Fatty Acid

WT: wild type

1. General introduction

Glycerolipids are major components of cell membranes. With the noticeable exception of archaea, less than a dozen of glycerolipid classes is sufficient to constitute the bulk of membranes of prokaryotes (bacteria, including cyanobacteria) and eukaryotes. Among them, monogalactosyldiacylglycerol (MGDG) is the major component of photosynthetic membranes (thylakoids), which makes it the most abundant membrane lipid on Earth. MGDG is essential for processes such as thylakoid membrane expansion and stabilisation of protein complexes that form the photosynthetic machinery. So far, the enzymes responsible for its synthesis have mostly been characterised in cyanobacteria and in land plants. Investigation on MGDG synthesis in eukaryotic algae has been limited to the detection and prediction of MGDG synthase genes in these organisms. Furthermore, in major phyla of photosynthetic eukaryotes, deriving from secondary endosymbiosis events, we still lack clear information about the precise localisation of MGDG in the three to four membranes delineating the complex plastids they harbour.

Diatoms are an important group of eukaryotic microalgae found in all aquatic ecosystems and responsible for 20 % of oxygen production on the planet. These unicellular organisms contain a secondary plastid surrounded by four membranes. In the model diatom *Phaeodactylum tricornutum*, three putative MGDG synthase isoforms have been identified based on sequence homology, and each isoform localises to a different subcellular compartment within the cell. The aim of this thesis was to determine their role in membrane lipid synthesis and potential influence on plastid biogenesis.

In order to characterise the three different MGD isoform in *P. tricornutum*, we first realised a phylogenetic analysis to trace back the appearance of each isoform in evolution. Additionally, further *in silico* analyses were performed to determine protein structure, targeting sequences, and active site localisation. To characterise the function of each isoform, we constructed lines overexpressing MGD proteins fused to a fluorescent protein, and we used the genome editing CRISPR-Cas9 technology to knock-out gene expression. Phenotypic analyses were conducted on mutant strains generated, such as lipid content analysis by GC-FID and LC-MS/MS, evaluation of membrane integrity by TEM, and photosynthetic efficiency measures under optimal growth condition. Lipid content was further monitored under nitrogen-deprivation condition.

Introduction Générale

Les glycérolipides sont des composants majeurs des membranes cellulaires. Avec l'exception notable des archaea, moins d'une douzaine de classes de glycérolipides sont suffisantes pour constituer l'essentiel des membranes de procaryotes (bactéries, dont cyanobactéries) et eucaryotes. Parmi elles, le monogalactosyldiacylglycérol (MGDG) est le composant majeur des membranes photosynthétiques (thylakoïdes), ce qui en fait le lipide membranaire le plus abondant sur Terre. Le MGDG est essentiel pour des processus tels que l'expansion des membranes des thylakoïdes et la stabilisation des complexes protéiques qui forment la machinerie photosynthétique. Jusqu'à présent les enzymes responsables de sa synthèse ont principalement été caractérisées chez les cyanobactéries et les plantes terrestres. Les recherches sur la synthèse de MGDG chez les algues eucaryotes ont été limitées à la détection et la prédiction de gènes codant pour des MGDG synthases (MGDs) chez ces organismes. En outre, dans les phylums majeurs d'eucaryotes photosynthétiques, dérivant d'évènements d'endosymbiose secondaire, il manque toujours des informations claires sur la localisation précise du MGDG dans les trois à quatre membranes délimitant les plastes complexes qu'ils contiennent.

Les diatomées sont un groupe important de microalgues eucaryotes rencontrées dans tous les écosystèmes aquatiques et responsables de 20 % de la production d'oxygène sur la planète. Ces organismes unicellulaires contiennent un plaste secondaire entouré de quatre membranes. Chez la diatomée modèle *Phaeodactylum tricornutum*, trois putatifs isoformes de MGDG synthase ont été identifiés par homologie de séquence, et chaque isoforme localise dans un compartiment subcellulaire différent dans la cellule. Le but de cette thèse fût de déterminer leur rôle dans la synthèse de lipides membranaires et leur potentielle influence sur la biogénèse du plaste.

Afin de caractériser les trois différents isoformes de MGD chez *P. tricornutum*, nous avons premièrement réalisé une analyse phylogénétique pour retracer l'apparition de chaque isoforme dans l'évolution. De plus, des analyses *in silico* approfondies ont été réalisées pour déterminer la structure protéique, les séquences d'adressage, et la localisation du site actif. Pour caractériser la fonction de chaque isoforme, nous avons construit des lignées surexprimant les protéines MGD fusionnées à une protéine fluorescente, et nous avons utilisé la technologie d'édition du génome CRISPR-Cas9 pour éliminer l'expression des gènes. Des analyses phénotypiques ont été conduites sur les souches mutantes générées, tel que l'analyse du contenu en lipides par GC-FID et LC-MS/MS, l'évaluation de l'intégrité des membranes par TEM, et des mesures de l'efficacité photosynthétique dans des conditions optimales de culture. Le contenu en lipides a en outre été suivi dans des conditions de carences d'azote.

2. Literature Review

This chapter regroups general knowledge about the diatom *Phaeodactylum tricornutum*, current knowledge about the synthesis of glycerolipids in plants, the role of monogalactosyldiacylglycerol in the cells and its synthesis, and the consequence of nutrient starvation on lipid homeostasis in plants. Parts of this chapter contain extracts from the review *'Lipid droplets in unicellular photosynthetic stramenopiles'* that I published last year in first author (Guéguen et al., 2021).

2.1. Diatoms and Phaeodactylum tricornutum

2.1.1. Diatoms, the most diverse group of microalgae

2.1.1.1. Ecological representation

Microalgae are photosynthetic unicellular organisms belonging to multiple prokaryotic and eukaryotic clades. They can be found in freshwater and seawater, but also in habitats such as soil, rock surface, deserts, hot springs, brackish water, and even ice or snow. The capacity to perform photosynthesis makes microalgae key primary producers in microbial nutritional networks. At equilibrium, microbial communities thus comprise a mixed biodiversity of microalgae and secondary and tertiary consumers. It happens that some disequilibrium can occur, for instance when a nutrient becomes abundant. In such disturbed environmental conditions, microalgae can grow at very high densities forming so-called "algal blooms", outcompeting other life forms.

Diatoms (Bacillariophyta) are a major monophyletic group of microalgae found in all aquatic ecosystems (de Vargas et al., 2015; Serôdio and Lavaud, 2021). The first representatives of this phylum, originating from the sub-reign of the Stramenopiles and the reign of the Chromoalveolata, have appeared 180 to 250 million years ago (Medlin et al., 2000; Sorhannus, 2007). They are believed to be one of the most abundant groups of phytoplankton, and thus major producers of the food chain (de Vargas et al., 2015). Estimates of the number of extant diatom species vary from several tens of thousands (20,000 species according to Guiry et al. 2012), up to 100,000 or even 200,000 species according to Kooistra et al (2007) and Mann and Droop (1996), respectively (Mann and Droop, 1996; Kooistra et al., 2007; Guiry, 2012). Known species are distributed in over 1,250 genera (Kooistra et al., 2007). Diatoms account for around 40 % of net primary production in oceans, which represents one fifth of all the photosynthesis on Earth, while only making up to 1 % of Earth's photosynthetic biomass (Field, 1998). Their high sinking rate makes diatoms critical in the export flux of carbon toward the ocean interior, where it can remain immobilized for centuries before entering the atmosphere again (Falkowski et al., 1998). Diatoms therefore play an important role in the carbon cycle. Diatoms also have an impact on the cycles of other important nutrients in oceans, such as the nitrogen and silicon cycles (Yool and Tyrrell, 2003; Bowler et al., 2010).

Diatoms are encased in porous silica shells called frustules. The name diatom (diá-tom-os, "cut in half") refers to these frustules being often symmetrical. Diatoms are classified based on their morphology between the centrics (radial symmetry) and pennates (bilateral symmetry). The pennate group radiated from the centric lineage about 150 million years ago (Falciatore et al., 2020) and is the most diverse one (Bowler et al., 2008). Pennate diatoms are further divided into the 'araphids' and the 'raphids', who have appeared 120 million years ago (Falciatore et al., 2020), following the emergence of a long longitudinal slit along the surface of the silica frustule, called a

'raphe'. Raphid pennates (Bacillariophyceae), the largest group, might have an additional ecological advantage with the gliding ability enabled by the raphe that allow them to move in response to changing environment (Underwood and Paterson, 2003).

Diatoms' ecological success has been attributed to a variety of competitive traits (Lima-Mendez et al., 2015). By contrast with other microalgae, the frustule provides a mechanical protection against predators (Hamm et al., 2003). Photosynthesis is optimized with the segregation of photosystems in subdomains to minimize physical contacts, thus improving light utilisation. Fast equilibration of electron carrier is then ensured by the interconnection of these subdomains through crosspoints between thylakoid layers (Flori et al., 2017). Furthermore, the proximity of the plastid with the mitochondria allows optimal energetic coupling between the two organelles (Bailleul et al., 2015; Flori et al., 2017). Vegetative diatom cells are diploid, and show an allele-specific expression conferring them phenotypic plasticity and contributing to their adaptability to unstable environments (Tirichine et al., 2017). Eventually, their genomes evolve fast, possibly due to the presence of transposable elements, known to be important in genome dynamics and species evolution (Maumus et al., 2009).

2.1.1.2. A complex cell architecture derived from multiple endosymbiosis events

In eukaryotes, photosynthesis acquisition always occurred following the engulfment of either a photosynthetic cyanobacterium (primary endosymbiosis) or of another photosynthetic eukaryote (secondary or tertiary endosymbiosis). In all cases, the engulfed organism gave rise to an organelle called a 'plastid', but the structure of this membrane structure shows a diversity reflecting the type of endosymbiosis at its origin.

Firstly, a unique initial acquisition of a cyanobacterium by an unknown ancestral heterotrophic eukaryote occurred around 1 to 1.5 billion years ago (Jensen and Leister, 2014; Maréchal, 2018) leading to the emergence of a 'primary plastid' – the so-called chloroplast (**Fig. 2.1**). Unicellular organisms possessing a primary plastid all derive from the same first endosymbiosis event. They are called Archaeplastida and radiated into four major lineages, the Glaucophyta, the Rhodophyta or red algae, the Chlorophyta or green algae, and the recently identified Prasinodermophyta lineage (Li et al., 2020). Archaeplastida also comprise the plant kingdom deriving from Chlorophyta, forming together with Prasinodermophyta the group of Viridiplantae.

A hypothetical '*Ménage à trois'* hypothesis suggests that conservation of the plastid inside the host was due to the co-infection of a Chlamidial pathogen together with the cyanobacterium, that would have helped establishing the initial flux of photosynthates to the host, providing some metabolic solutions to carbon storage in carbohydrates (Ball et al., 2013; Cenci et al., 2017). The inserted cyanobacteria thus stabilized 'became' a chloroplast following its genome reduction and loss of its capacity to live freely.

Secondary endosymbiosis refers to the acquisition of a complex or secondary plastid by eukaryote-to-eukaryote endosymbiosis. The engulfed organism is a photosynthetic eukaryote containing a primary plastid, i.e. an alga *stricto sensu* (**Fig. 2.1**). Secondary endosymbiosis is thought to have occurred several times during evolution, and has been performed by heterotroph eukaryotes, consequently becoming phototrophs (Reyes-Prieto et al., 2007). Unicellular photosynthetic organisms resulting from secondary endosymbiosis are sometimes incorrectly referred to as "algae", the term of "secondary endosymbionts" being possibly more appropriate. Higher order

endosymbiosis events also took place during evolution and involved more complex interactions, also leading to 'complex plastids' formation (Keeling, 2010, 2013; Oborník, 2019).

Diatoms stem from two endosymbiosis events, where the last endosymbiosis involved the engulfment of a microalga from the red lineage. However, the second endosymbiosis may imply different red algae and/or different heterotrophic cells (Keeling, 2013), as well as distinct patterns of horizontal gene transfers (Fan et al., 2020; Vancaester et al., 2020).

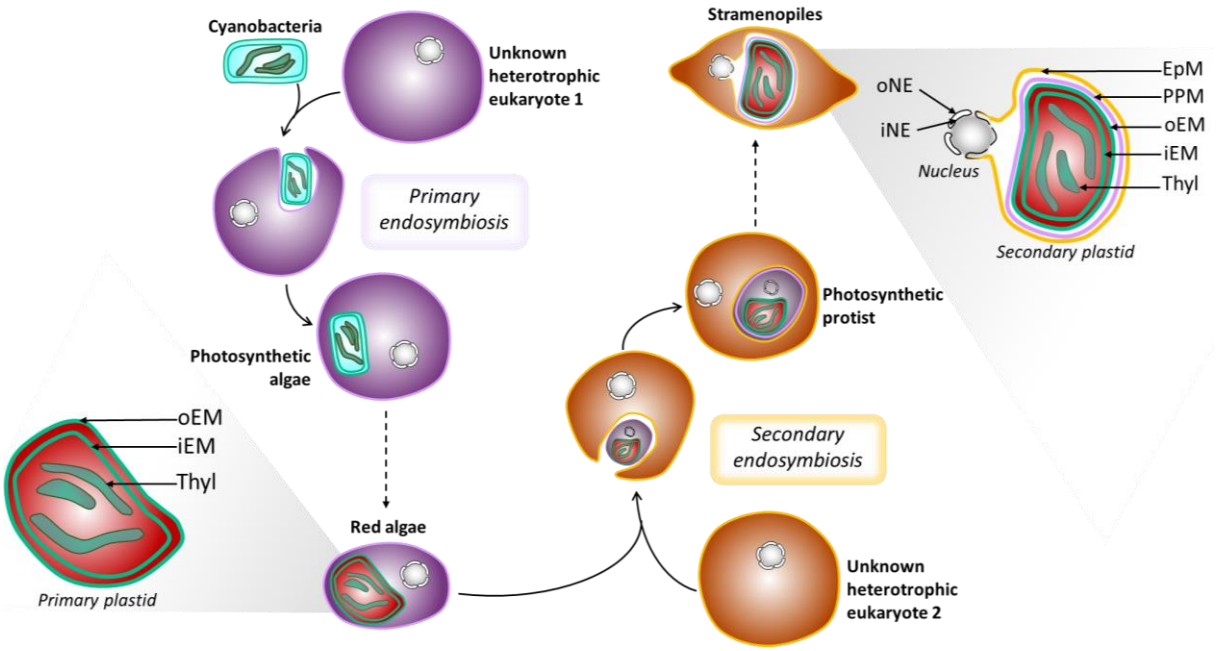


Figure 2.1: Schematic representation of plastid evolution - primary and secondary endosymbiosis, and organelle architecture. EpM, epiplastidial membrane; oEM and iEM, plastid outer and inner envelope membranes respectively; oNE and iNE, outer and inner nuclear envelope membranes respectively; PPM, periplastidial membrane; Thyl, thylakoids. Adapted from (Guéguen et al., 2021).

A feature shared by complex plastids is the presence of more than two membranes surrounding the plastid (**Fig. 2.1**). Some secondary plastids encountered in cryptophytes and chlorarachniophytes can even still contain a relic of the nucleus from the ancestral eukaryotic endosymbiont, called the nucleomorph (Curtis et al., 2012). Diatoms possess plastids bounded by four membranes and lack the nucleomorph.

At the beginning of my PhD project, the lipid composition of secondary plastid membranes was still unknown, and was often inferred based on evolutionary assumptions:

- The origin of the two innermost membranes of the plastid is not or little debated, and corresponds to the chloroplast envelope of the symbiont, called the outer and inner envelope membranes (oEM and iEM, respectively; **Fig. 2.1**).
- By contrast, different suppositions have been made about the origin of the additional bounding membranes. The outermost membrane, called the epiplastidial membrane (EpM, **Fig. 2.1**), could derive from the host phagocytic membrane. Underneath, the periplastidial membrane (PPM, **Fig. 2.1**) is usually considered to derive from the symbiont plasma membrane (Grosche et al., 2014). Alternatively it has

been hypothesized that the outer envelope membranes of red complex plastids could derive from the host endoplasmic reticulum (ER) (Gould et al., 2015).

- Interestingly, a vesicular network forming a 'blob-like' structure has been detected between the PPM and the oEM in *Phaeodactylum*, but its function is still elusive (Kilian and Kroth, 2004; Flori et al., 2016). This blob-like structure was proposed to be a relic of the endosymbiont endomembrane system (Cavalier-Smith, 2018).
- The EpM can also be found under the name of 'chloroplastic endoplasmic reticulum', as its peculiarity in diatoms is to be continuous on the one hand with the outer membrane of the nuclear envelope, forming a Nucleus-Plastid Continuum (NPC), and on the other hand with the cytosolic ER (Murakami and Hashimoto, 2009; Flori et al., 2016)(**Fig. 2.1**). In addition, membrane contact sites were detected between the inner membrane of the nuclear envelope and the PPM (Flori et al., 2016). Such a tight association between the plastid and the nucleus is hypothesized to facilitate the transport of proteins, RNAs, small molecules, and be involved in still-unknown functions.

Based on the suggested origin of the four membranes surrounding the secondary plastid, it is plausible but still not demonstrated that the two innermost membranes contain plastid-specific lipids such as monogalactosyldiacylglycerol (MGDG), digalactosyldiacylglycerol (DGDG) and sulfoquinovosyldiacylglycerol (SQDG) (Boudière et al., 2014), whereas the two outermost membranes may be related to the ER lipid composition, with lipids such as phosphatidylcholine and betaine lipids (Boudière et al., 2014; Dolch and Maréchal, 2015).

The events that led to the emergence of secondary plastids are critical milestones in the evolution of eukaryotes, and their impact on cell biology is still poorly understood. The tight structural and physical interactions of the secondary plastid with other organelles of the endomembrane system, such as the ER, the nucleus, or uncharacterized networks of vesicles, challenge our understanding of cellular processes usually assumed to be restricted to the ER.

2.1.1.3. Protein targeting in complex plastids

A major specificity of diatoms is therefore their complex cellular architecture. Presence of four membranes delineating the plastid is reflected into their organellar protein targeting system. Following endosymbiosis, endosymbiont genes were transferred to the host nucleus by lateral gene transfers. Nuclear-encoded proteins targeted to the plastid need to cross four membranes with distinct evolutive origins (**Fig. 2.2A**). To this end, precursors of proteins are first translated as preproteins presenting an N-terminal targeting sequence called a topogenic bipartite presequence. A bipartite presequence contains first a signal peptide that targets the protein to the ER/EpM, followed by a plastid transit peptide-like domain that then targets the protein to the plastid stroma (Kilian and Kroth, 2004) (**Fig. 2.2B**). Following import, each targeting peptide is cleaved, forming the mature protein. Swapping signal peptides between ER-targeted proteins and plastid-targeted proteins showed a functional conservation, and suggests that ER and plastid preproteins use the same pathway to enter the ER, though plastid preproteins need to be sorted further (Kilian and Kroth, 2004).

Some bipartite presequences are recognized *in silico* by the presence of a conserved motif at the cleavage site of the transit peptide (Kilian and Kroth, 2004; Gruber et al., 2007). This motif, namely ASAFAP, is missing in signal peptides of proteins putatively targeted to other destinations within the endomembrane system and thus may be specific to plastid-targeted preproteins (Kilian and Kroth, 2004) (**Fig. 2.2B**). Interestingly, presence of this motif following the signal peptide is sufficient for further sorting of the preprotein into the blob-like structure, located between the PPM and the oEM, in the absence of transit peptide (Kilian and Kroth, 2004; Flori et al., 2016). Within this motif, presence of an aromatic amino acid, generally phenylalanine (F) immediately after the cleavage site is crucial for the complete import into the plastid stroma, and is generally surrounded by two alanine residues (A) (Gruber et al., 2007) (**Fig. 2.2B**). The ASAFAP motif is thus used as a necessary sorting step for successful import to the plastid. A customized prediction tool now allows the identification of such motifs (ASAFind) (Gruber et al., 2015).

Intriguingly, although transfer mechanisms in the presence of a signal peptide and a transit peptide have been characterised in plants, the molecular determinant by which the ASAFAP motif allows the crossing of the PPM is still unclear. Two different models have been proposed, where the two middle membranes are either crossed *via* a vesicular shuttle system in a so-called 'vesicular shuttle model', or by proteins translocators and/or pores in a 'translocator model' (Kilian and Kroth, 2004) (**Fig. 2.2B**). In both models, a cotranslational transport across the EpM is postulated. The high conservation of an aromatic amino acid in position +1 of the cleavage site suggests that the transport/receptor system might be involved in its recognition, but this mechanism remains speculative (Gruber et al., 2007).

In plants, the N-terminal part of proteins targeted to the chloroplast is recognized by a protein translocator complex called the translocon at the iEM and oEM (TIC and TOC) mediating the import of the pre-protein to the chloroplast stroma (Schleiff et al., 2003; Nakai, 2015). In *P. tricornutum*, only components of the TIC apparatus were identified, suggesting that the transfer across the iEM could be mediated by a similar system (Gruber et al., 2007). TOC components seem to derive from bacterial Omp85 system (**Fig. 2.2B**).

In several organisms deriving from the engulfment of a red algae, including *P. tricornutum*, a plastid-located ER-associated degradation (ERAD)-like machinery has been identified (Hempel et al., 2009; Stork et al., 2012). In simple eukaryotic cells, the ERAD system is usually involved in the retrotranslation of misfolded proteins from the ER to the cytosol for proteasomal degradation. However, in complex plastids deriving from a red algae, the altered ERAD-related machinery was proposed to be involved in the regular transport of properly folded proteins out of the ER and into the PPC (Hempel et al., 2009; Stork et al., 2012). This plastid-located ERAD-like machinery was termed 'symbiont-specific ERAD-like machinery' (SELMA) (**Fig. 2.2B**). SELMA substrates presumably undergo ubiquitination. Consistently, two proteins involved in protein ubiquitination and de-ubiquitination have been identified in *P. tricornutum*, localised in the PPM and PPC respectively (Hempel et al., 2010).

Our understanding of secondary plastid biogenesis is currently limited to these basic data on protein import through the various limiting membranes, and at the beginning of my thesis, no machinery involved in the biogenesis of the lipid components was characterized.

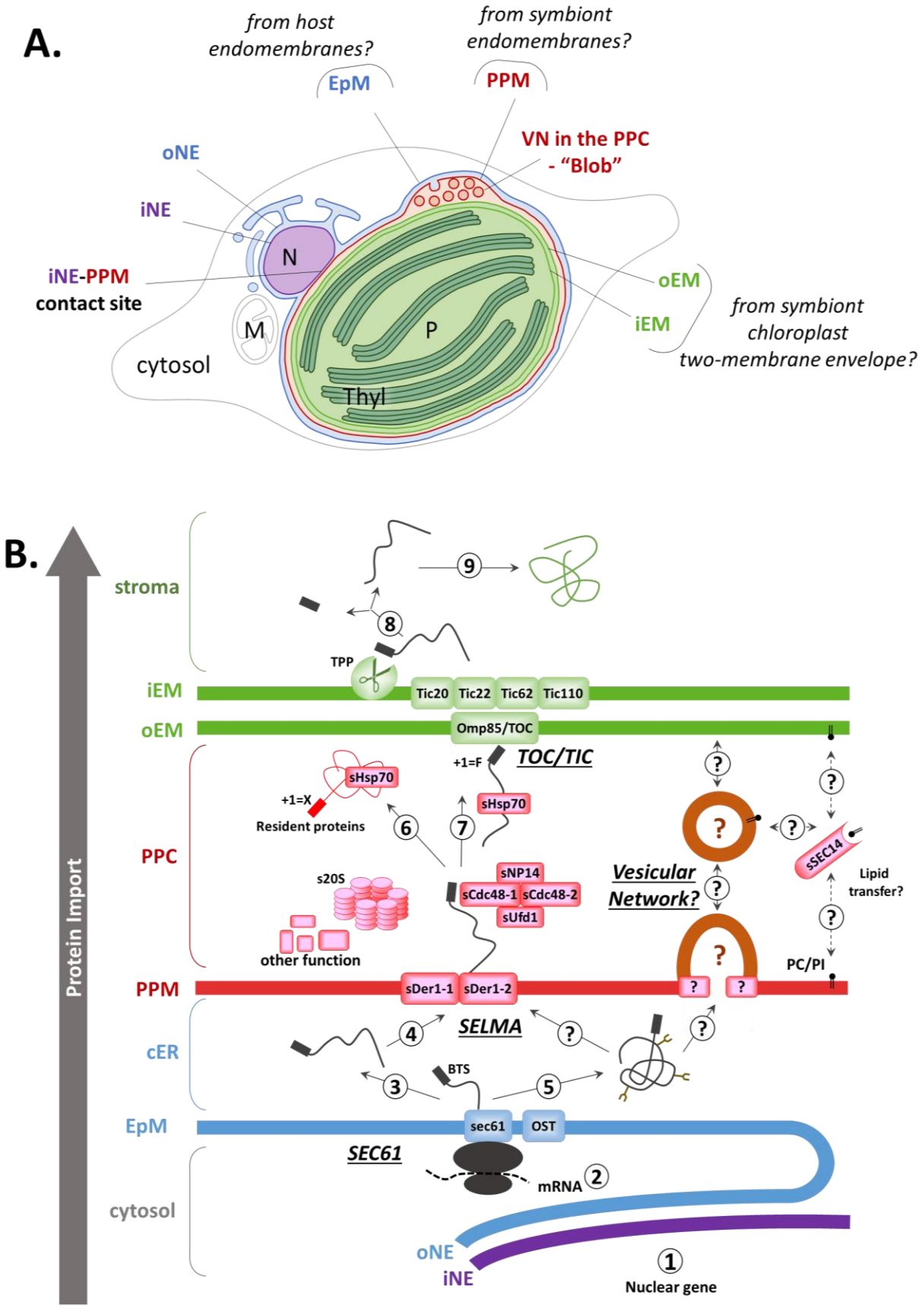

A.
from host endomembranes?
from symbiont endomembranes?
EpM
PPM
VN in the PPC - "Blob"
oNE
iNE
iNE-PPM contact site
N
M
P
Thyl
cytosol
oEM
iEM
from symbiont chloroplast two-membrane envelope?
B.
Protein Import
stroma
iEM
oEM
PPC
PPM
cER
EpM
cytosol
TPP
Tic20
Tic22
Tic62
Tic110
Omp85/TOC
TOC/TIC
+1=F
+1=X
Resident proteins
sHsp70
s20S
other function
sNP14
sCdc48-1
sCdc48-2
sUfd1
Vesicular Network?
sSEC14
Lipid transfer?
PC/PI
sDer1-1
sDer1-2
SELMA
BTS
sec61
OST
SEC61
mRNA
oNE
iNE
Nuclear gene

Figure 2.2: A. Chimeric organization of the secondary plastid in diatoms. A. The scheme shows a fusiform cell of Phaeodactylum. The plastid is limited by 4 membranes. The epiplastidic membrane (EpM), shown in blue, is continuous with the outer nuclear envelope membrane (oNE). The periplastidial membrane (PPM) is shown red. The outer and inner envelope membranes (oEM and iEM), shown in light green, are tightly apposed. The presence of a specific periplastidial compartment (PPC) is based on the detection of blob-like structures observed by confocal microscopy, in which protein precursors fused to GFP and crossing only the EpM and the PPM reside. The presence of a vesicular network (VN) in this PPC is addressed here. P, plastid; N, nucleus; M, mitochondrion. **B**. Protein import across the four plastid limiting membranes and *via* the PPC. Following mRNA transcription (1) and translation (2), plastid protein precursors harbour a bipartite topogenic signal (Bts). A sec61 complex operates very early by co-translational mediation of preproteins across the EpM (3) and release an unfolded protein precursor in the lumen of the chloroplast ER (EpM). This unfolded protein can be directed to the next membrane (4). Preproteins can also be N-glycosylated prior their transport through the PPM (5), probably by the action of an oligosaccharide transferase (OST). A translocon called the 'symbiont-specific ERAD-like machinery' (SELMA) is located in the PPM. In the SELMA, Derlin proteins, sDer1-1 and sDer1-2, interact together and with the Bts. Components of a symbiont ERAD machinery, i.e. sCdc48 ubiquitin-dependent AAA-ATPases, and their cofactors sUfd1 and sNP14, reside in the PPC. In the absence of an aromatic amino acid at position +1 of the Ctp (+1=X), proteins remain resident in the PPC. Presence of a phenylalanine or an aromatic residue (+1=F) determines the transport across the oEM and the iEM. Transport across the oEM and iEM involves components related to the chloroplast translocon, i.e. TOC and a TIC respectively. The TOC core component derives from a prokaryotic Omp85 sequence. Important TIC subunits well characterized in plant and alga chloroplasts are conserved in the iEM, i.e. Tic20, Tic22, Tic62 and Tic110. Cleavage of the Bts (8) by a transit peptide peptidase (TPP) releases mature proteins in the stroma (9). In this scheme the transfer of folded proteins is not deciphered. Possible routes *via* the VN in the PPC are shown.

2.1.2. *Phaeodactylum tricornutum,* a model for the study of diatoms

2.1.2.1. Introduction to *P. tricornutum*

P. tricornutum is a marine pennate diatom first described by Bohlin in 1897 and formerly known as *Nitzschia closterium* (Allen and Nelson, 1910; Bowler and Falciatore, 2019). To date, ten ecotypes have been characterised around the world (Bowler and Falciatore, 2019). It is a polymorphic diatom classified as a raphid pennate (**see Section 2.1.1.1**). It can present fusiform, triradiate, oval, cruciform or round morphotypes according to the strain and environmental conditions (**Fig.2.3A**) (Lewin et al., 1958; Martino et al., 2007; Martin-Jézéquel and Tesson, 2012). Under our laboratory optimal growth conditions, the studied *P. tricornutum* strain (CCAP1055/1) appeared mainly in the fusiform morphotype (**Fig.2.3B**), while oval forms arose in response to stressful conditions. Very rarely, triradiate cells are observed.

Although *P. tricornutum* is mostly a marine diatom, it includes brackish water strains, and two strains (CCAP1052/1B and CCAP1055/1) were reported to even grow in freshwater (Yongmanitchai and Ward, 1991; Listwan et al., 2018). *P. tricornutum* is a photoautotroph organism, meaning that it can grow with light as its sole source of energy. However, it is also capable of mixotrophy, feeding on either glucose, acetate, fructose or glycerol as external carbon sources (Villanova et al., 2017). Less silicified than other diatoms, *P. tricornutum* can be grown in the absence of Si, in which case it will not develop any silica frustule.

By contrast with some other diatoms, no substantial sexual reproduction evidence has been obtained. In laboratory conditions, *P. tricornutum* growth is achieved only through mitotic cell division. Briefly, *P. tricornutum* cytokinesis starts with the division of the plastid, followed by the formation of cleavage furrows at both longitudinal ends of the cell. Then the nucleus is split *via* the centripetal growth of the furrows, and division ends with the fusion of the furrows (**Fig.2.4**) (Tanaka et al., 2015a). One division is usually measured each day.

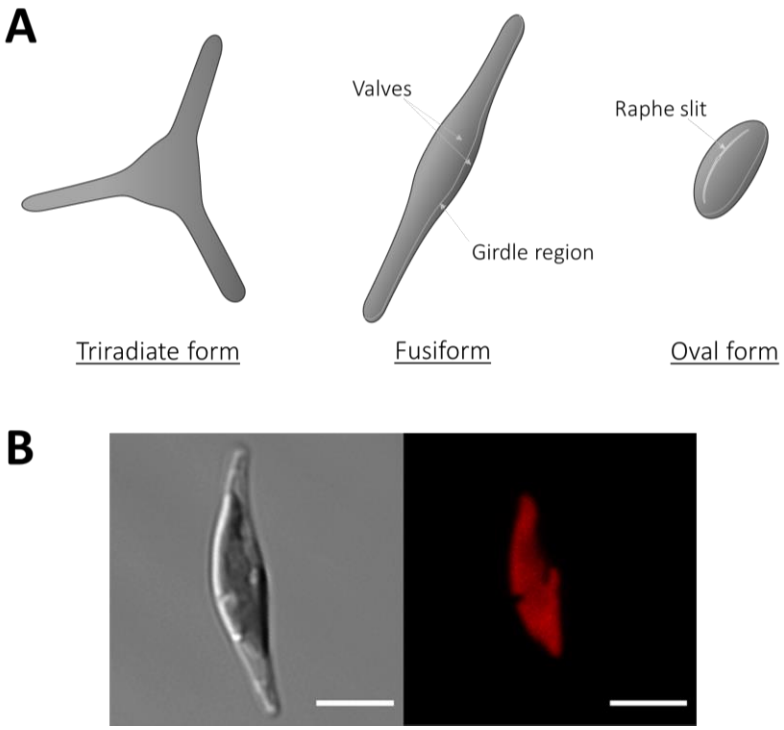


Figure 2.3: The three main morphotypes in *P. tricornutum.* **A**, Schematic drawing of the triradiate, fusiform and oval forms. Characteristics of frustules are indicated with arrows. The raphe structure is present only in oval form (Martino et al., 2007). **B**, Bright field and chlorophyll fluorescence pictures of a *P. tricornutum* cell cultivated in optimal growth conditions, in the fusiforme morphotype. Scale bar: 5 µm.

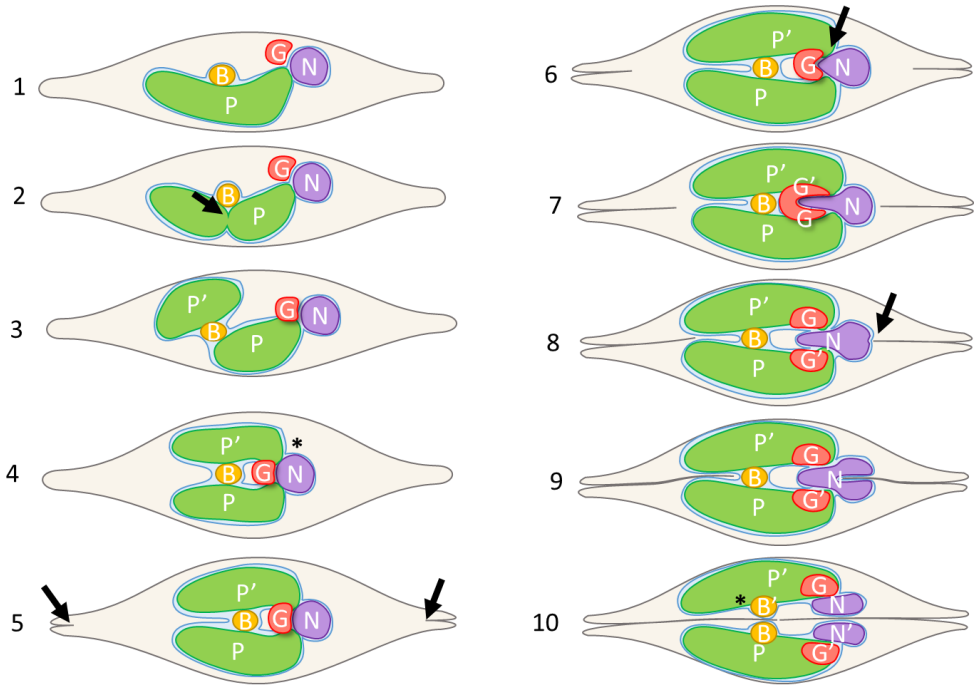


Figure 2.4: Cytokinesis steps in *P. tricornutum*. Schematic representation of the different steps of *P. tricornutum* cell division. **1**, cell before the beginning of the division. **2**, centripetal division of the plastid. **3**, rearrangement of the two plastids alongside each other. **4**, possible reattachment of the second plastid to the nucleus. **5**, Apparition of cleavage furrows at both longitudinal ends of the cell. **6** and **7**, cleavage of the Golgi apparatus by lengthening of the nucleus. **8** and **9**, cleavage of the nucleus by a growing furrow. **10**, fusion of the furrows and with possible cleavage of the blob by the furrows. Abbreviation: B, blob; G, Golgi apparatus; N, nucleus; P, plastid. The blue line corresponds to the most external membrane of the plastid and the nucleus. Adapted from (Tanaka et al., 2015a) and microscopy observations (**see sections 4.2 and 4.3.2.2**).

2.1.2.2. *Phaeodactylum* as a model organism

P. tricornutum is currently the most popular model organism for the study of diatoms. The complete genome of ecotype CCAP1055/1 has been sequenced and annotated (Scala et al., 2002; Montsant et al., 2005; Bowler et al., 2008). *P. tricornutum* has a genome size of 27.4 Mbp, and latest genome assembly was arranged into 33 chromosome-sized scaffolds (Bowler et al., 2008). The exact number of chromosomes is unknown, but there should be at least 29 chromosomes according to most recent genome examination (Filloramo et al., 2021). *P. tricornutum* genome comprises 12,233 coding genes, and 12 % of the genome are repetitive sequences mainly made of transposable elements (75%) (Rastogi et al., 2018).

P. tricornutum is a fast-growing species, easily cultivated in laboratory conditions and genetically transformable. Various transformation methods are available, such as biolistic, electroporation, and bacterial conjugation, and transformation can be done both on nuclear and plastid genomes (Apt et al., 1996; Xie et al., 2014; Karas et al., 2015; Nymark et al., 2016; Serif et al., 2018). Furthermore, the ability to grow without silica shell makes DNA introduction easier.

The pPhaT-1 plasmid is the most frequently used expression vector (Zaslavskaia et al., 2001). It contains a cassette conferring Zeocin resistance for primary selection, and a multicloning site flanked by the *P. tricornutum* fcpA promoter. The fcpA (LHCF1) and FcpB (LHCF2) promoters are the most frequently used, although they are light-dependent and inactive in the dark (Nymark et al., 2013). Other promoters with higher activity and no light phase dependence have been developed in the past years (Erdene-Ochir et al., 2016, 2019; Adler-Agnon et al., 2018; Watanabe et al., 2018).

2.1.3. Fundamental and applied research interest

Diatoms are models of choice for the study of secondary endosymbiosis, to better understand their evolution and metabolism (Keeling, 2010, 2013; Oborník, 2019). Our knowledge gap on the evolution, structure and biogenesis of this complex cell architecture, the unknown fundamental mechanisms, which cannot be addressed in more simple eukaryotic cells such as *Chlamydomonas*, *Arabidopsis*, yeast, *Drosophila* or humans, is immense. This is surprising that we know so little on such important phyla in the Tree of Life. The increasing research interest found in diatoms has led to a strikingly high number of genomic characterisations and transcriptomic studies (full sequences and annotations) (Armbrust, 2004; Oudot-Le Secq et al., 2007; Bowler et al., 2008; Lommer et al., 2010; Secq and Green, 2011; Vieler et al., 2012; Keeling et al., 2014; Di Dato et al., 2015; Tanaka et al., 2015b; Mock et al., 2017; Schwartz et al., 2018; Yang et al., 2018b; Osuna-Cruz et al., 2020).

Beside basic questions, microalgae have attracted the attention as potential cell factories for the production of high value compounds (**Fig. 2.5**). They are superior to plants in terms of biomass productivity, thrive in saline water, and can theoretically be grown without impact on non-arable land. Moreover, another advantage of microalgae is that they can feed on waste resources such as wastewater or CO_2 emitted by industry, allowing their cultivation in urbanized areas, as demonstrated for several diatoms (Soedarti et al., 2017; Hedayatkhah et al., 2018; Saranya and Ramachandra, 2020; Maréchal, 2021). Algae in general have been used for hundreds of

years for human purposes and are now exploited to produce agar, alginates, fertilizers, nutritional products and pigments. Our increasing understanding of algae thanks to the availability of sequenced genomes and transcriptomes allows us to consider their use for more biotechnological purposes (Kim et al., 2018; Lupette and Maréchal, 2018; Barkia et al., 2019; Ścieszka and Klewicka, 2019; Anto et al., 2020; Aziz et al., 2020; Chaudhary et al., 2020). Hereafter are detailed several industrial applications and prospects based on microalgae, with a focus on some *P. tricornutum* applications.

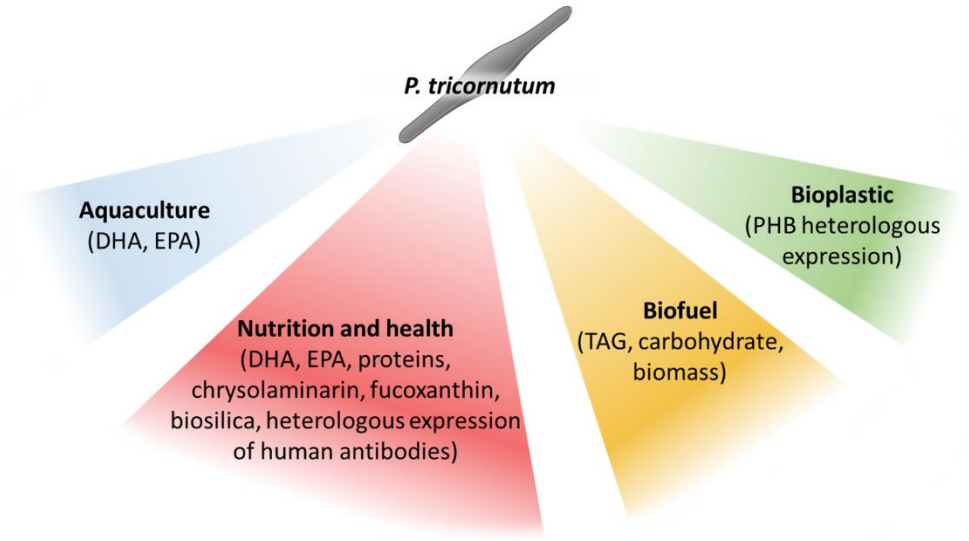


Figure 2.5: Possible industrial applications of *P. tricornutum*. *P. tricornutum* is rich in essential fatty acids, proteins and antioxidants which makes it attractive both as feed in aquaculture and as a nutraceutical product. Under certain conditions, it can accumulate oil in the form of TAG and carbohydrates, that can be used for biodiesel and bioethanol, respectively. Biomass can also be used for the production of biogas through methanisation. Genome edition of *P. tricornutum* has been achieved for the expression of non-native compound such as antibodies and PHB (Butler et al., 2020). In addition, diatom biosilica has attracted attention in the field of drug delivery (Uthappa et al., 2018). The feasibility of an integrated biorefinery for biofuels and high-value compounds was studied (Branco-Vieira et al., 2020).

Microalgae are primary producers of very long chain-fatty acids (VLC-PUFA) such as arachidonic acid (ARA; 20:4ω6), eicosapentaenoic acid (EPA; 20:5ω3), and docosahexaenoic acid (DHA; 22:6ω3). *P. tricornutum*, particularly rich in EPA, is already widely used in aquaculture as feed for molluscs and fish. As most animals, humans also require essential fatty acids (FA) of the ω3 and ω6 types. In modern 'Western diets', excessive consumption of ω6 fatty acids has led to a dietary ratio of ω6/ω3 of 15 instead of the recommended 3-4 (Cunnane et al., 2004; Hibbeln et al., 2006; Lands, 2014). Thus, there is a need to increase ω3 FAs consumption that fish oil cannot sustainably provide. Current research aims at increasing EPA and DHA in microalgae by controlling lipid metabolism, like in *P. tricornutum* (Mühlroth et al., 2013; Gao et al., 2017), which can also accumulate high DHA levels in modified strains (Hamilton et al., 2015).

Microalgae are also very interesting for their capacity to accumulate large amounts of energetically-dense oil in the form of triacylglycerol (TAG) that can be used for biofuel production, with diatoms being among the possible strains of interest to develop production strains (Ramachandra et al., 2009; Maréchal, 2021). Once harvested,

TAGs can be converted by a transesterification process into FA methyl esters (FAMEs) for biodiesel. Currently, the main unresolved issue with algal-derived biodiesel is its still too high cost compared to petroleum-based products. In order for microalgae to become more economically competitive, a better understanding of microalgae metabolic pathway is needed and has been the object of many studies aiming at increasing oil content. Methods to increase lipid production usually consist in (i) genetically engineer microalgal strains to either increase TAG biosynthesis or decrease lipid catabolism, and (ii) finding culture conditions triggering oil accumulation, like nutrient stresses or the use of chemical triggers (Widjaja et al., 2009; Jiang et al., 2012; Franz et al., 2013; Conte et al., 2018; Yang et al., 2018a). Schematically, the metabolic control of TAG accumulation can be achieved by "pushing" FAs production, "pulling" TAG biosynthesis and loading into lipid droplets (LDs) and "protecting" TAG from hydrolysis.

Most diatoms are able to accumulate chrysolaminarin (1,3-β-D-glucan) as carbon sink, a carbohydrate similar to starch (Myklestad and Granum, 2009; Caballero et al., 2016). Chrysolaminarin presents antioxidant properties and can be produced in extractable amounts with *P. tricornutum* (Xia et al., 2014; Gao et al., 2017).

Finally, most diatoms accumulate pigments such as chlorophylls *a* and *c*, beta-carotene, diatoxanthin, diadinoxanthin, and fucoxanthin, giving them their golden-brown colour (Bertrand, 2010). Such pigments have utility in the food, cosmetic and pharmaceutical industries (Henríquez et al., 2016). *P. tricornutum* fucoxanthin-containing products have been marketed (Yang et al., 2020).

More microalgae-derived products exist and will not be discussed further here.

2.2. Monogalactosyldiacylglycerol (MGDG), a key glycerolipid in photosynthetic organisms

2.2.1. The different glycerolipid classes in *P. tricornutum*

2.2.1.1. Membrane glycerolipids

Membrane bilayers are made of polar lipids, which harbour a 'polar head group' exposed to the surface, and a hydrophobic 'tail' inside the core of the bilayer. They can be of various chemical natures (e.g. glycerolipids, isoprenoids, sterols, sphingolipids, etc.). Considering glycerolipids conserved in bacteria (including cyanobacteria) and eukaryotes, the hydrophobic tail corresponds to two FAs of varying length esterified at positions *sn-1* and *sn-2* of a glycerol backbone, whereas the 'polar head group' linked at position *sn-3* varies according to the lipid class.

There are nine membrane glycerolipid classes in *P. tricornutum* (without considering lysolipids): five phospholipids, three glycolipids, and one betaine lipid (**Fig. 2.6**).

- The phospholipid classes comprise three anionic phospholipids, namely phosphatidic acid (PA), phosphatidylglycerol (PG), and phosphatidylinositol (PI), and two zwitterionic phospholipids, phosphatidylcholine (PC) and phosphatidylethanolamine (PE).

- The glycolipids classes comprise two galactolipids, monogalactosyldiacylglycerol (MGDG) and digalactosyldiacylglycerol (DGDG), and one anionic sulfolipid, sulfoquinovosyldiacylglycerol (SQDG).
- A betaine lipid found in *P. tricornutum* is diacylglyceryl hydroxymethyltrimethyl-β-alanine (DGTA).

In non-stressed conditions, the glycerolipid profile of *P. tricornutum* is dominated by MGDG, SQDG and PC, which together represent more than 75% of the total content (Abida et al., 2015).

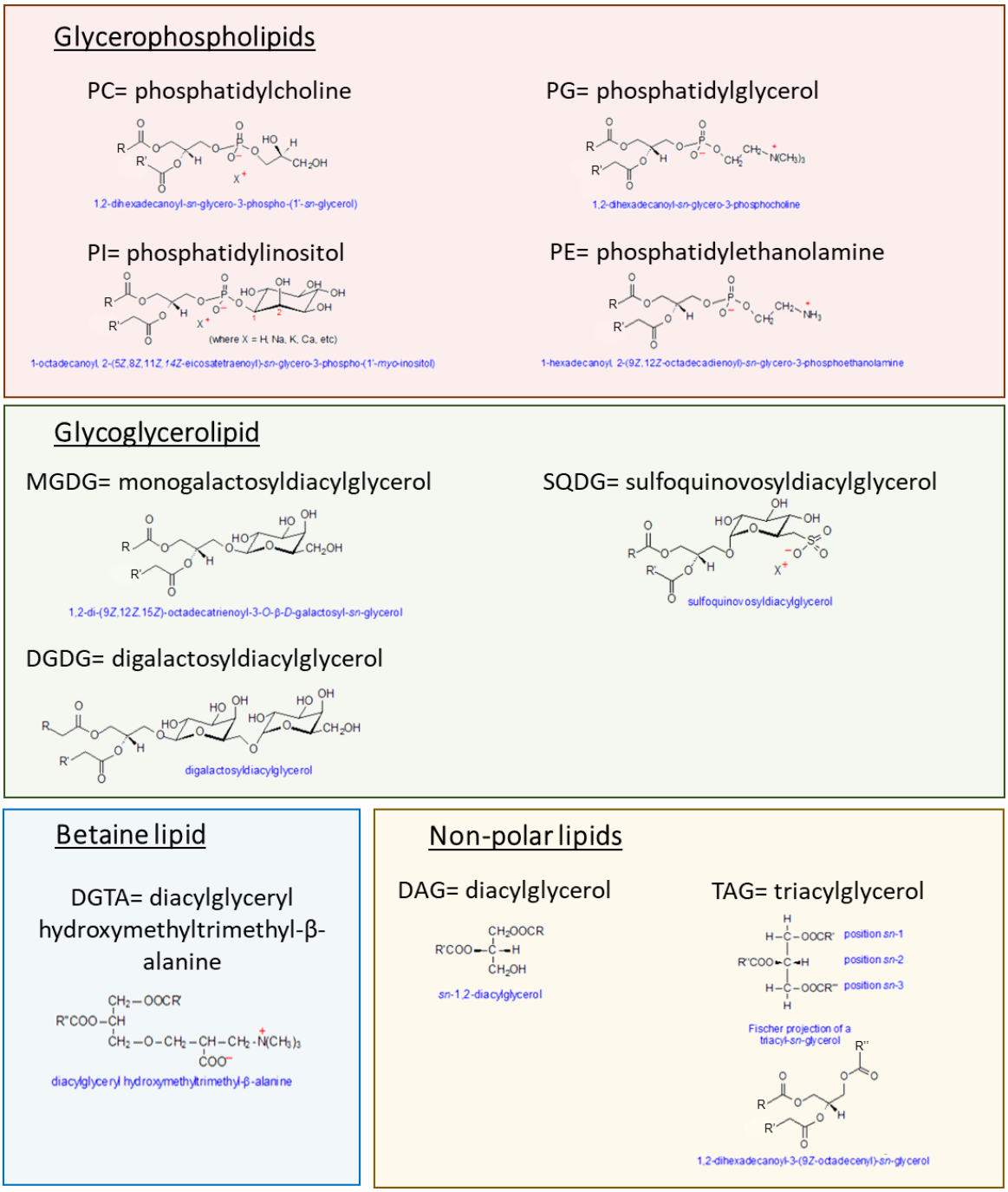


Figure 2.6: Membrane and storage glycerolipid classes in *P. tricornutum*. Lipid classes are grouped according to the nature of the polar head. MGDG, SQDG, PG and DGDG are the main component of photosynthetic membranes while PC and DGTA are the main lipids encountered in extraplastidial membranes.

SQDG, MGDG, DGDG and PG are conserved from cyanobacteria to 'primary' plastid membranes (Petroutsos et al., 2014; Guéguen and Maréchal, 2021), whereas PC, PE, PI and DGTA are usually found in the endomembrane

system of simple eukaryotes. PA and PG can be found both in the plastid and the endomembrane system. In *P. tricornutum*, the exact localisation of each lipid and of their corresponding synthetases in the plastid membranes is unknown.

Polar glycerolipids are not just the bulk of a water-impermeable barrier, in which membrane proteins can insert, and their diversity is not simply a marker of a sub-cellular compartment. In structural terms, depending on the equilibrium between polar head interactions and repulsive forces between the hydrophobic tails, each of the polar lipids can self-organize and form so-called 'phases', with either a positive curvature (e.g. micellar tubules, also known as 'hexagonal I', or HexI), no curvature ('lamellar', or Lm), or a negative curvature ('inverted hexagonal', hexagonal II, or HexII) (**Fig. 2.7**) (Jouhet, 2013; Guéguen and Maréchal, 2021). DGDG, PG, PC, SQDG and PI self-organise as Lm phase, corresponding to the classical bilayer. MGDG, PA and PE self-organise as inverted micelles (HexII phase). HexII structures can play roles in some enzyme activities and in membrane fusion and fission events (Jouhet, 2013). 'HexII lipids' can however be found in membrane bilayers without curvature, either because in their pure state they can form both HexII or Lm phases depending on temperature, hydration, etc., or because when they are mixed with Lm lipids, the lamellar phase is the more stable organisation.

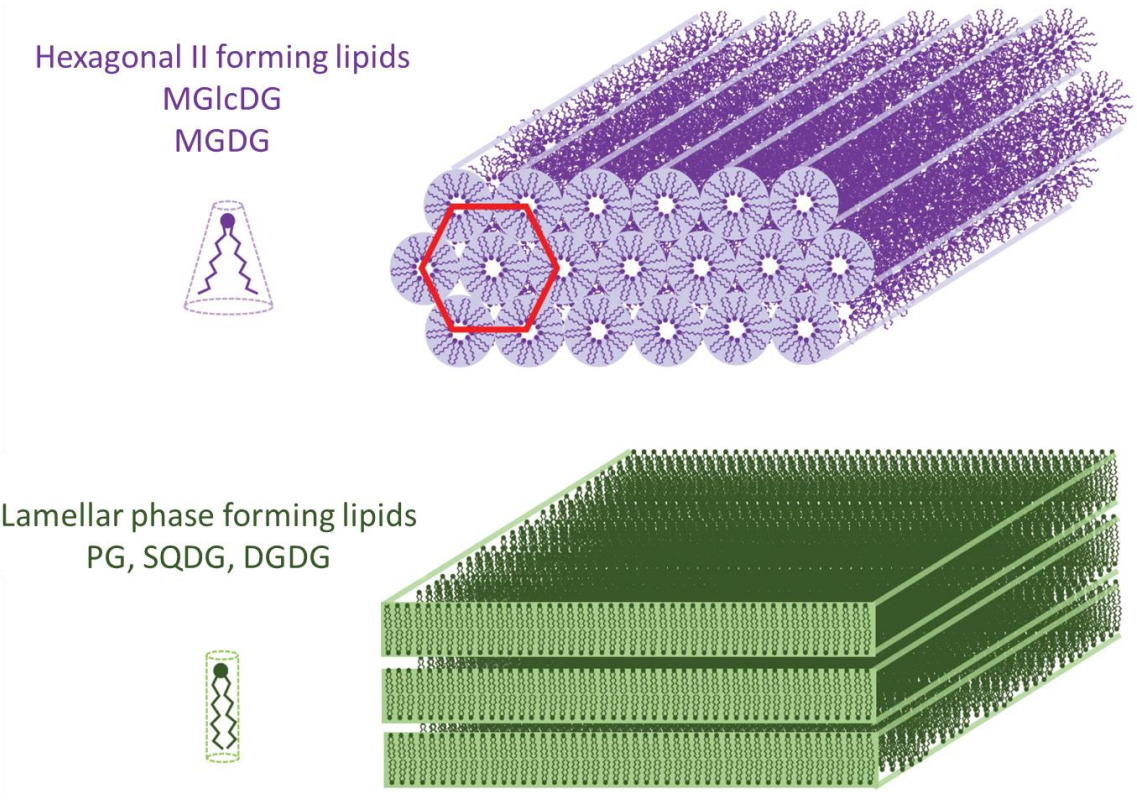


Figure 2.7: Hexagonal II and lamellar lipid phases. Lipids with a polar head (shown as a solid circle) that is relatively smaller than the volume occupied by their hydrophobic tail (acyl chains) have a molecular shape that fits within a truncated cone. They induce a negative curvature and favour self-assembly into inverted tubular micelles (hexagonal II or HexII phase, shown in purple). Arrays of tubular micelles are regular, with seven tubules having a hexagonal section, as shown in red. Lipids with a similar cross-sectional area for their polar head and their hydrophobic tail have a molecular shape that fits within a cylinder. They form bilayers (lamellar or Lm phase, shown in green). Some lipids can form both structures, based on some discrete structural differences, such as the number of double bonds in their acyl moiety, or on different physicochemical conditions

(temperature, hydration level). Lipid mixtures can adopt one conformation or the other, based on the relative proportions of HexII-forming and Lm-forming lipids in the mixture. Adapted from (Guéguen and Maréchal, 2021).

2.2.1.2. Triacylglycerol, a storage form

Triacylglycerols (TAGs) are non-polar lipid composed of three FAs esterified on a glycerol backbone. They are the most condensed form of carbon and energy storage, and are therefore known as a 'storage lipid'. Even though they are originally produced in a membrane bilayer, TAGs differ from polar lipids by their lack of lateral cohesion, and they self-organize as an isotropic phase, known as 'oil'. They are usually found in LDs located in the cytosol and/or in the plastid (plastoglobules). In stramenopiles, they can also probably be found in the ER and the EpM during LD biogenesis (Guéguen et al., 2021). In *P. tricornutum*, cytosolic LDs contain 99 % TAG and 1 % polar lipids (Lupette et al., 2019). The bulk of TAG contains saturated and monosaturated FAs, but EPA can also be found at position *sn-1* of several TAG species.

In standard 'unstressed' conditions, low levels of TAGs accumulate during the light period and are consumed in the dark (Chauton et al., 2013). High TAG production can be triggered by many abiotic stresses such as light and temperature exceeding tolerance thresholds (Alboresi et al., 2016), but the most studied stresses for the production of TAG are nutrient stresses like nitrogen and phosphate deprivations (Vieler et al., 2012; Simionato et al., 2013; Mayers et al., 2014; Abida et al., 2015; Mühlroth et al., 2017; Liang et al., 2019a, 2019b). Phosphate and nitrate limitations are supposed to induce an imbalance between carbon and the missing nutrient, with a carbon excess diverted to storage forms including TAG. Cells exposed to nitric oxide, hydrogen peroxide, and other oxidative reactive species, also respond by increasing TAG accumulation in LD (Dolch et al., 2017a; Conte et al., 2018). A plethora of chemicals proved efficient in triggering TAG accumulation in *P. tricornutum* (Conte et al., 2018).

TAG accumulation following a stress usually coincides with a cell cycle slowdown or arrest. The energy excess produced by photosynthesis that still operates during the first phases of a stress is stored and constitutes a pool of carbon and energy available upon stress release. When favourable growth conditions are restored, the energy released from LDs by TAG hydrolysis and FA beta-oxidation is used to reactivate cell cycle and photosynthesis activity.

In our laboratory, abiotic conditions inducing TAG accumulation are considered important for the study of enzymes synthesizing membrane glycerolipids, since in the case of a multigenic family, some of the isoforms may be specifically involved in the glycerolipid remodelling triggered in response to such stress.

Here, we can note that like most photosynthetic stramenopiles, *P. tricornutum* is also able to accumulate chrysolaminarin, another type of carbon storage similar to starch (Caballero et al., 2016). This polysaccharide is found in vacuoles (Schreiber et al., 2017), and like TAG, accumulates during the light period for later consumption in the dark. However, in contrast to the massive TAG accumulation occurring under nutrient stress, nitrogen deprivation does not seem to stimulate chrysolaminarin over-accumulation (Caballero et al., 2016).

2.2.1.3. Glycerolipid synthesis

Current knowledge about FAs and glycerolipid syntheses in algae is increasing greatly, mostly due to the interest in algal lipid production for industrial applications. However, the detailed mechanisms of these syntheses still need to be inferred by comparison with yeasts and plants where they have been better characterised.

2.2.1.3.1. Acetyl-CoA and malonyl-ACP, the original precursors

In eukaryotic species with a plastid, *de novo* FA synthesis occurs in the stroma. The building blocks necessary for FA synthesis are acetyl-Coenzyme A (acetyl-CoA) and malonyl-acyl carrier protein (malonyl-ACP). At least three different routes can lead to acetyl-CoA production. In plant and algae, most of the acetyl-CoA is produced from pyruvate by the plastid pyruvate dehydrogenase enzyme (PDH) (Mooney et al., 1999) that also generates CO_2 and NAPDH in the process (Johnston et al., 1997). Increase of PDH activity in *P. tricornutum* by the silencing of the pyruvate dehydrogenase kinase (PDK) leads to an increase in acetyl-CoA, that in turn increases neutral lipid content (Ma et al., 2014). Acetyl-CoA can also be produced from acetate by two ATP-dependent routes, either by an acetate synthase (ACS) or by the action of an acetate kinase (ACK) followed by a phosphate acetyltransferase (PAT). However, production of acetyl-CoA from acetate is not relevant for FA synthesis in plants as most of acetyl-CoA comes from PDH (Schwender et al., 2003). Finally, acetyl-CoA can be produced from citrate by an ATP-citrate lyase (ACL), although the involvement of ACL in FA synthesis in algae has not been demonstrated yet.

The committing step for FA synthesis is the generation of malonyl-CoA from acetyl-CoA by an acetyl-CoA carboxylase (ACCase) (**Fig. 2.8A**). ACCases are biotin-containing enzymes that perform carboxylation of acetyl-CoA in a two-step, ATP-dependent manner. In plants and algae, both homomeric and heteromeric forms of ACCases occur (Ohlrogge and Browse, 1995; Abu-Elheiga et al., 2000; Hu et al., 2008). By contrast, in algae derived from secondary endosymbiosis, only homomeric ACCases located in the plastid are present (Huerlimann and Heimann, 2013; Haq et al., 2017). The homomeric ACCase in *P. tricornutum* has been identified, and heterologous overexpression of *P. tricornutum* ACCase in *E. coli* was shown to increase a neutral lipid production, as detected by Nile red staining (Xie et al., 2013). Since *E. coli* major neutral lipid usually consists of an acetyl-CoA derived Poly(3-hydroxybutyrate) (PHB), it is not clear how the overexpression of an ACCase diverting the carbon flow to FAs (Page et al., 1994) used for membrane glycerolipids could lead to such phenotype. Malonyl-CoA is then converted to malonyl-ACP by a malonyl-CoA:ACP transferase (MCMT). In chromists, this enzyme (also known as FabD) has been characterised is several species, including *P. tricornutum* (Ryall et al., 2003).

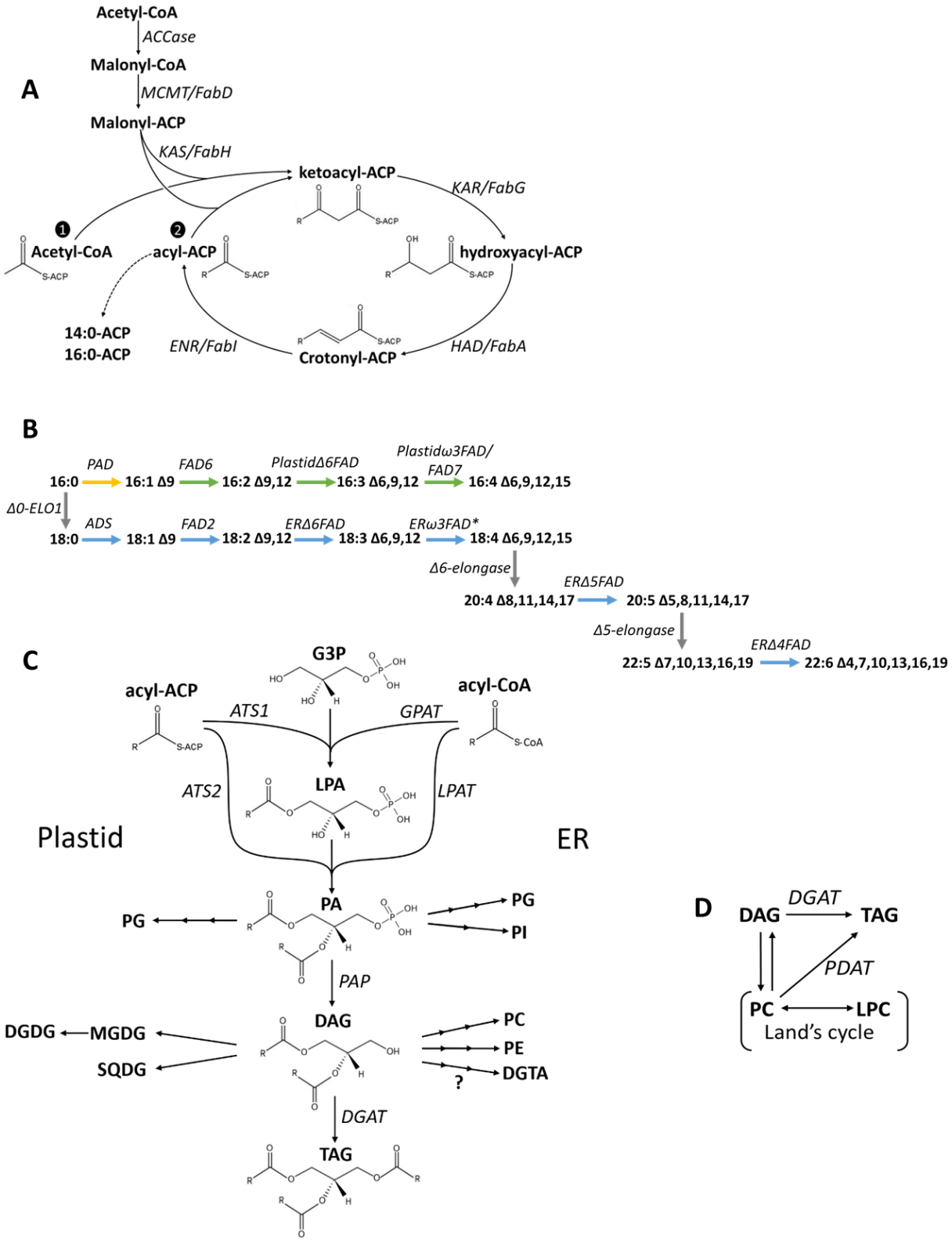


Figure 2.8 : Glycerolipid synthesis pathways. A, *De novo* fatty acid synthesis. **B**, Fatty acid elongation and desaturation in *P. tricornutum* (Dolch and Maréchal, 2015). Horizontal arrows represent desaturation steps, vertical arrows represent elongation steps. Arrow colour depends on the localisation of the step: plastid stroma (yellow), plastid membrane (green), cytoplasm (grey), or endoplasmic reticulum membrane (blue). (*) desaturase not identified. **C**, Glycerolipid synthesis. Steps corresponding to the Kennedy(/-like) pathway are detailed in the centre. Origin of the acyl moieties and localisation of the membrane glycerolipid syntheses are indicated on each side of the pathway. **D**, Representation of the different pathways leading to the synthesis of TAG through membrane lipid remodelling. PC is taken as an example.

2.2.1.3.2. De novo fatty acid synthesis

In plants and algae, *de novo* FA synthesis occurs in the plastid. FA synthesis starts with the condensation of an acetyl-CoA (Coenzyme A) and a malonyl-ACP (Acyl Carrier Protein) by a 3-ketoacyl-ACP synthase (KAS, also known as FabH), forming the four-carbon ketoacetyl-ACP and releasing one molecule of CO_2 from the malonyl-ACP (**Fig. 2.8A**). This reaction is followed by a reduction of the ketoacetyl-ACP into a 3-hydroxybutyryl-ACP catalysed by a 3-ketoacyl-ACP reductase (KAR or FabG) using NADPH as electron donor. Then it is dehydrated into a crotonyl-ACP by a hydroxyacyl-ACP dehydratase (HAD or FabA), and finally reduced into a butyryl-ACP by an enoyl-ACP reductase (ENR or FabI) that uses NADH or NADPH as electron donors. The butyryl-ACP, a four-carbon acyl-ACP, is then elongated by the same four enzymes with the addition of two carbons units at each iteration by the KAS, using malonyl-ACP as building block (Li-Beisson et al., 2013). Together, the KAS, KAR, HAD and ENR enzymes are called FA synthases (FAS) and are found either as a dissociated system (FAS of type II or FASII) or as a multienzymatic protein (FAS of type I or FASI) (Brown et al 2006).

Plants and diatoms have a FASII in the plastid stroma (Ryall et al., 2003). In flowering plants, *de novo* FA synthesis produces FAs with up to 16 and 18 carbons (Li-Beisson et al., 2013). However, it is still unknown if elongation in the plastid of diatoms reaches 18 carbons or only 16 carbons (Domergue et al., 2003; Abida et al., 2015; Dolch et al., 2017b). Based on the discovery of a Δ0-elongase in the cytosol of eustigmatophytes (Dolch et al., 2017b) and diatoms (Smith et al., 2021), elongating 16:0-CoA into 18:0-CoA, it is hypothesized in our laboratory that 18-carbon FAs produced in the plastid of stramenopiles is minor, if not absent.

In *A. thaliana*, three KAS isoforms with different specificities exist. KASIII performs the initial condensation of acetyl-CoA with malonyl-ACP, while KASI elongates acyl-ACP to a chain length of 16 carbons, and finally KASII perform the final elongation to reach 18 carbons (Pidkowich et al., 2007). All four enzymes of the FASII have been annotated in *P. tricornutum* genome, but only the ENR has been characterised in *P. tricornutum* and other chromists species (Ryall et al., 2003).

Freshly synthetised acyl-ACP is then either used directly for glycerolipids assembly, or converted to a free FA by an acyl-ACP thioesterase (FAT) before being transported by an unknown mechanism through the plastid envelope to the cytoplasm where further elongation can proceed.

2.2.1.3.3. Fatty acid desaturation and elongation

In *P. tricornutum*, *de novo* FA synthesis produces saturated acyls linked to ACP with 14 and 16 carbons (see above the discussion of the lack of 18-carbon FAs), that will be referred to as 14:0 and 16:0 in this manuscript. There are no desaturations occurring on 14:0. As many as four desaturations can be generated on the initial 16:0 (**Fig. 2.8B**). Desaturations of this FA occur in the plastid. Most of the FA desaturases in *P. tricornutum* were predicted *in silico*, some being characterized at the biochemical level (for review, Dolch and Maréchal, 2015). Desaturation of 16:0 to a $16{:}1^{\Delta 9}$ occurs on palmitoyl still linked to ACP, and is performed by an acyl-ACP Δ9-desaturase, called palmitoyl-ACP desaturase (PAD), in the stroma. Further desaturation of $16{:}1^{\Delta 9}$ is performed on specific glycerolipids in the plastid membranes by membrane-bound desaturases. The FAD6 catalysing the desaturation of $16{:}1^{\Delta 9}$ to $16{:}2^{\Delta 9,12}$ has been functionally characterized and works on acyl linked either to the *sn-1* or *sn-2*

positions of galactolipids (MGDG and DGDG) and SQDG. A PlastidΔ6FAD is then predicted to desaturate 16:2$^{\Delta 9,12}$ linked to MGDG at position *sn-2* into 16:3$^{\Delta 6,9,12}$. A final desaturation is predicted to be performed by the Plastidω3FAD/FAD7 on 16:3$^{\Delta 6,9,12}$ linked to MGDG at position *sn-2* to produce 16:4$^{\Delta 6,9,12,15}$. Although the PlastidΔ6FAD was not predicted to work on the *sn-1* position, presence of MGDG 16:3/16:3 (*sn-1*/*sn-2*) suggests that this enzyme might use acyl at the *sn-1* position as well. Production of 18:0 in diatom plastid is still unknown as only small amounts of FAs with 18 carbons can be detected in diatom plastids. The PAD found in *P. tricornutum* might work as a stearoyl-ACP desaturase (SAD) catalysing the desaturation of 18:0-ACP to 18:1$^{\Delta 9}$-ACP (Dolch and Maréchal, 2015), but recent characterisation shows that this desaturase works more as a PAD than a SAD (Smith et al., 2021). Alternatively, a cytosolic version stearoyl-CoA desaturase (ADS) has also been predicted, and desaturases performing further desaturation of 18:1$^{\Delta 9}$ have been characterised and localise in the ER (see below) (Dolch and Maréchal, 2015).

Production of very long chain polyunsaturated FAs (VLC-PUFA) occurs in the cytoplasm and the ER and requires the export of free 16:0 out of the plastid. 16:1 FAs are not used for generation of VLC-PUFA, which makes the PAD act as a key enzyme in (negatively) controlling VLC-PUFA level, as demonstrated by Smith et al (Smith et al., 2021). The FAT releasing free FA and ACP determines the quantity and type of FAs that are exported. In *P. tricornutum*, a FAT called PtTE has been characterized. PtTE shows no similarities to characterised plant and bacterial FAT (Gong et al., 2011), but overexpression in *P. tricornutum* leads to a strong increase in total FA amount with no impact on FA composition. Export of free 16:0 from the plastid to the cytosol requires an efficient transport mechanism through the four membranes surrounding the complex plastid of diatoms. Such transport mechanism has not been identified to date. Outside the plastid, free FAs are activated by esterification to Coenzyme A (CoA) by long-chain acyl-CoA synthases (LACS), and thus feed the cytosolic acyl-CoA pool involved in acyl-editing. In *P. tricornutum*, five LACS genes have been identified (Guo et al., 2014), and recently characterised using CRISPR-Cas9 editing (Hao et al., 2021). PtACSL2 and PtACSL4 were proposed to be involved in the activation of FAs exported form the plastid for glycerolipid assembly and acyl editing.

In *P. tricornutum* membrane glycerolipids, the most abundant FA is 20:5 (eicosapentaenoic acid, or EPA), and the longest VLC-PUFA is 22:6. Thus, several steps of elongation and desaturation are necessary for their formation from newly produced FAs (**Fig. 2.8B**). FA elongation occurs mostly in the ER (Leonard et al., 2004), and desaturations occur mostly on glycerolipids by membrane-bound desaturases, with the exception of plastidial PAD/SAD. As mentioned above, although it is still unknown if 18:0 are produced in diatom plastids, a cytosolic elongase Δ0-ELO1 generating 18:0 from 16:0 was identified in the eustigmatophyte *Nannochloropsis* (Dolch et al., 2017b; Liu et al., 2021), suggesting a similar mechanism exists in diatoms as well (Smith et al., 2021). Overexpression of Δ0-ELO1 was shown to "push" the flow of FAs toward the production of VLC-PUFA. Furthermore, two homologues of this Δ0-ELO1 were identified in *P. tricornutum* (Smith et al., 2021). The next step toward production of VLC-PUFA is the desaturation of 18:0 to 18:1$^{\Delta 9}$, and is predicted to be performed by ADS in the endomembrane system on acyl linked to CoA. All following desaturations have been either shown or predicted to be performed by ER-localised desaturases working on acyl linked to phospholipids, and potentially to betaine lipids as well (Dolch and Maréchal, 2015; Conte et al., 2018).

A characterised FAD2 reduces 18:1$^{\Delta 9}$ into 18:2$^{\Delta 9,12}$ in the ER. From there, different routes leading to 20:5 production are theoretically possible (Domergue et al., 2002). Pulse chase experiments with ^{14}[C] linoleic acid (Arao and Yamada, 1994) suggested that one route is mostly active. First, 18:2$^{\Delta 9,12}$ is reduced into 18:3$^{\Delta 6,9,12}$ and then to 18:4$^{\Delta 6,9,12,15}$ by the action of a characterised ERΔ6FAD and a non-identified desaturase, respectively. The non-identified desaturase was tentatively called ERω3FAD (Dolch and Maréchal, 2015). Then, 18:4$^{\Delta 6,9,12,15}$ is elongated to 20:4$^{\Delta 8,11,14,17}$ by a Δ6-elongase (Phatr3_J22274 and Phatr3_J20508) (Dolch et al., 2017a; Conte et al., 2018). Production of EPA is finally achieved by the characterised ERΔ5FAD.1 and the predicted ERΔ5FAD.2 catalysing the desaturation of 20:4$^{\Delta 8,11,14,17}$ into 20:5$^{\Delta 5,8,11,14,17}$. A last elongation is performed by a Δ5-elongase (Phatr3_J9255 and Phatr3_J34485), generating 22:5$^{\Delta 7,10,13,16,19}$, before the final desaturation into 22:6$^{\Delta 4,7,10,13,16,19}$ by ERΔ4FAD.

In plants, acyl editing occurs at position *sn-2* of PC (Griffiths et al., 1988, 1988; Bates et al., 2007, 2009; Tjellström et al., 2012). Integration of a newly synthesized FA on PC is achieved by the Land's cycle (Lands, 1960). Briefly, phospholipase cleavage of a FA from the *sn*-1 or *sn*-2 position of PC generates lysoPC and a free FA, and is followed by reacylation of lysoPC by an acyl-CoA:lysoPC (LPCAT) producing PC and free CoA. ^{14}C labelling experiments showed that in *Arabidopsis* the flux of newly synthetized FAs is mostly directed to acyl editing rather than glycerolipid synthesis, by their higher incorporation to the *sn*-2 position of PC than to their acylation to positions *sn*-1 and *sn*-2 of glycerol-3-phosphate (G3P) (Bates et al., 2007, 2009, 2012; Tjellström et al., 2012). High acyl-editing allows the non-accumulation of highly saturated FAs. The existence of a Land's cycle and the magnitude of this phenomenon in diatoms is unknown.

2.2.1.3.4. Assembly of glycerolipid precursors: PA and DAG

Glycerolipid *de novo* synthesis occurs both in the ER and the plastid. In the ER, the scaffolding starts with the esterification of an acyl-CoA at position *sn*-1 of G3P by a glycerol-3-phosphate-*sn*-1-acyl-CoA-acyltransferase (GPAT) leading to the formation of lysophosphatidic acid (LPA) (**Fig. 2.8C**). The addition of a second acyl-group at position *sn*-2 by a lysophosphatidic acid acyltransferase (LPAT) produces phosphatidic acid (PA). In the plastid, these esterification reactions use acyl-ACP instead of acyl-CoA and are performed by the glycerol-3-phosphate-acyl-ACP-acyltransferases ATS1 and ATS2 (Kunst et al., 1988; Yu et al., 2004). PA can be used to produce some of the phospholipid classes, or be dephosphorylated by a PA phosphatase (PAP) to produced diacylglycerol (DAG). In *P. tricornutum*, candidate GPAT, LPAT and PAP genes have been identified (Zulu et al., 2018). Among the GPAT candidates, two were proposed to code for the plastid-localised ATS1 and ATS2 (Abida et al., 2015). Overexpression of the suspected ATS2 in *P. tricornutum* leads to the accumulation of neutral lipid, presumably TAG (Niu et al., 2016). Interestingly, FA composition revealed a higher proportion of PUFA. In this article, the authors described this gene as coding for a plastid-localised GPAT rather than an ATS2 (or LPAT).

2.2.1.3.5. Synthesis of the different glycerolipid classes

DAG and PA are the universal precursors for all glycerolipid classes in eukaryotes. In the plastid, glycerolipids can contain FA species of ER origin on their DAG backbone, although the import mechanism of these FAs back to the plastid is still debated (**see section 2.2.1.4**). In *Arabidopsis*, the origin of a diacyl backbone is easily identified by two kinds of 'molecular signatures' defined by the FAs at position *sn*-2. The two pathways leading to their

synthesis are called “prokaryotic” or “eukaryotic”, when they occur in the plastid or the ER respectively. The use of these terms was recently contested (Sato and Awai, 2017). While such a distinction between topologically distinct pathways is quite evident in plants (Roughan and Slack, 1982), in most stramenopiles, a more thorough investigation is needed to depict the exact origin of the different building blocks of glycerolipids (Mühlroth et al., 2013). In the context of simple plastids, lipid classes exclusively synthesised in the plastid are SQDG, MGDG and DGDG, while those exclusively synthetised in the ER are PC, PE, and betaine lipids. PG on the other hand can be synthetised both in the plastid and in the ER. In *A. thaliana*, synthesis of plastid lipids occurs in the chloroplast envelope.

In *P. tricornutum*, at the beginning of my PhD thesis, the precise subcellular localisation of glycerolipids synthases was not known.

Among the biosynthetic steps considered as conserved in all plastids, PA can be directly used for PG synthesis in a three-step pathway. PA is converted to cytidine diphosphate-DAG (CDP-DAG) by a CDP-DAG synthase. Then a PG-phosphate (PGP) synthase (PGPS) uses G3P and CDP-DAG to produce PGP and cytidine monophosphate (CMP). Finally PG is formed from PGP by a PGP phosphatase (PGPP) (Andrews and Mudd, 1985). In *P. tricornutum*, a candidate gene was noted for PGPS (Zulu et al., 2018). PA not channelled to PG synthesis is dephosphorylated into DAG.

MGDG is synthesised by a MGDG synthase (MGD) catalysing the transfer of a galactosyl residue from a uridine diphosphate-galactose (UDP-Gal) to position *sn-3* of a DAG, forming a β-glycosidic linkage. DGDG is then synthesised by the transfer of a second galactosyl from UDP-Gal by a DGDG synthase (DGD), forming an α-glycosidic linkage between the two sugar residues (Kelly and Dörmann, 2004; Andersson and Dörmann, 2008).

In a similar way, SQDG is synthetised by the transfer of a sulfoquinovosyl from a UDP-sulfoquinovose donor to DAG (Benning, 2008). UDP-sulfoquinovose is assembled in the plastid stroma from sulfite and UDP-glucose.

In the ER, PG is synthetised as described above for the plastid. Similarly, PI is synthetised from PA by the formation of CDP-DAG, which is used together with *myo*-inositol as substrates for a PI synthase releasing PI and CMP (Klezovitch et al., 1993). PA not used for these two anionic phospholipids is converted into DAG.

DAG is the substrate for biosynthesis of the zwitterionic phospholipids PC and PE, and of DGTA. PC is synthetised from choline and DAG *via* the CDP-choline pathway involving serial reactions catalysed by a choline kinase, a CTP:phosphorylcholine cytidyltransferase and a CDP-choline:DAG cholinephosphotransferase. Similarly, PE is synthesised from ethanolamine and DAG *via* the CDP-ethanolamine pathway involving serial reactions catalysed by an ethanolamine kinase, a CTP:phosphorylethanolamine cytidyltransferase and a CDP-ethanolamine:DAG ethanolaminephosphotransferase.

DGTS is the sole betaine lipid for which a biosynthetic pathway is described. DGTS synthesis starts by the attachment of a C4 homo-serine moiety to a DAG, and finishes by the addition of three methyl groups to its amino group. These two reactions are performed in eukaryotes by a two-domain betaine lipid synthase A (BTA1). DGTA is believed to be synthesised from a DGTS precursor (Vogel and Eichenberger, 1989; Sato, 1992). DGTA

synthesis in *P. tricornutum* is still elusive as multiple enzymes could be involved. Currently, one protein was identified as a betaine-lipid synthase-like enzyme (BTA-like) (Conte et al., 2018).

DAG not only serves as a precursor for membrane glycerolipids, in particular MGDG, subject of my PhD thesis, but is also used for the biosynthesis of storage lipids in the form of TAG. It is therefore a key branching point in glycerolipid metabolism.

2.2.1.3.6. Synthesis of TAG: different DAG origins

TAG is commonly synthesised in the ER. TAG can be synthesised by a DAG acyl transferase (DGAT) that esterifies a third acyl-CoA at position *sn-3* of a DAG (**Fig. 2.8C**). This is the last step of the 'Kennedy pathway', also called the 'acyl-CoA-dependent pathway', which also includes the initial scaffolding of LPA, PA, and DAG. When TAG synthesis occurs in the plastid, acyl-ACP is used (Kennedy-like pathway) (Kennedy and Weiss, 1956). The last step of TAG formation can alternatively occur using FAs transferred from an existing membrane glycerolipid used as a donor. This pathway is then termed acyl-CoA independent (Dahlqvist et al., 2000).

P. tricornutum possesses several DGATs, namely, one PtDAGT1, five PtDGAT2, and a PtDGATX (Cui et al., 2018) or PtDGAT3 (Zhang et al., 2020), with a double wax ester synthase (WS) and DGAT function (Cui et al., 2013, 2018). All the PtDGATs were tested for activity either *in vitro* or by heterologous expression in the *Saccharomyces cerevisiae* quadruple mutant H1246.

- PtDGAT1 (GenBank accession MN061782) was recently reannotated and characterized (Zhang et al., 2020). Among the tested PtDGAT, PtDGAT1 is the most active and presented a slight preference toward mono-unsaturated medium chain FA like C16:1 over saturated medium chain ones (Zhang et al., 2020). A detectable, although reduced, activity was recorded on EPA (C20:5) too. PtDGAT1 seems to play a crucial role during nitrogen starvation, its expression being highly upregulated in a time-dependent manner at early and late starvation, i.e., when TAG cellular content was highly increased (Guihéneuf et al., 2011; Cui et al., 2018; Zhang et al., 2020). Overexpression of PtDGAT1 in *P. tricornutum* showed little effect on cell growth, but induced a twofold accumulation of lipids (Zhang et al., 2020). In nitrogen starvation conditions, PtDGAT1 overexpressing (OE) lines redirect carbon from carbohydrates and proteins toward TAG synthesis (Zhang et al., 2020).
- The PtDGAT2 family is represented by five protein coding genes in *P. tricornutum*, namely, PtDGAT2A (JX469835, Phatr3_49462), PtDGAT2B (JQ837823, Phatr3_49544) (Gong et al., 2013; Haslam et al., 2020; Zhang et al., 2020)], PtDGAT2C (JX469836, Phatr3_31662), PtDGAT2D (JX469837, Phatr3_43469), and PtDGAT2E (Phatr3_EG00369) (Chen and Smith, 2012). PtDGAT2B is the only one that rescued the phenotype of the H1246 mutant (Gong et al., 2013; Zhang et al., 2020), had an in vitro activity (Zhang et al., 2020), and presented a considerable overexpression in nitrogen starvation (Gong et al., 2013; Zhang et al., 2020). PtDGAT2B localizes at the cER, like PtDAGT1 (Zhang et al., 2020) and has been more intensively studied (Gong et al., 2013; Haslam et al., 2020; Zhang et al., 2020) compared with other PtDGAT2. It was demonstrated that PtDGAT2B expression varies along the growth curve in nitrogen replete condition with a peak at the end of the exponential phase (Gong et al., 2013), while in nitrogen

starvation, its expression stays invariably high (Gong et al., 2013; Zhang et al., 2020). Such pattern correlates with TAG accumulation. Consistently, PtDGAT2B-overexpressing lines accumulate TAGs (Haslam et al., 2020; Zhang et al., 2020). Analyses of TAG species in overexpressing lines suggest that PtDGAT2B may prefer C16 species and C16-containing DAG as substrates (Haslam et al., 2020). The role played by PtDGAT2A, PtDGAT2C, and PtDGAT2D in the Kennedy pathway remains uncertain because besides not complementing the phenotype of the H1246 mutant, neither showing any or very low activity *in vitro*, they are all poorly expressed in both nitrogen replete and deplete conditions (Gong et al., 2013; Zhang et al., 2020).

- In *P. tricornutum* genome, the locus Phatr3_J49708, encoding for a predicted protein was characterized by heterologous expression in the H1246 mutant and in vitro assay (Cui et al., 2013; Zhang et al., 2020). Phatr3_J49708 might be identified as PtDAGT3 by homology to *Acinetobacter calcoaceticus* DGAT3 (AE529086). PtDGAT3 is localized to the EpM (Zhang et al., 2020), and presents two domains, namely, a wax ester acyltransferase and acyl-CoA:diacylglycerol acyltransferase domains (Cui et al., 2013). Intriguingly, a few years later, the same predicted protein was identified as a novel dual-function PtWS/DGAT and named PtDGATX PtDGAT3 transcript is weakly expressed compared with PtDGAT1 and PtDGAT2B (Cui et al., 2018; Zhang et al., 2020), and along a growth curve, its expression peaks in the middle of the exponential phase. The overexpression of PtDGAT3 in alga did not impair growth nor photosynthesis efficiency, though it induced a TAG accumulation in both N-replete and N-deplete conditions (Cui et al., 2018; Zhang et al., 2020), which is coherent with the DGAT activity of the protein. Moreover, expression of PtDGAT3 in the H1246 mutant induced an accumulation, although to a lesser extent, of wax, as expected from the WS domain identified (Cui et al., 2013, 2018).

In diatoms, both FAs originating from the plastid and the ER seem to provide DAG for TAG biosynthesis (Radakovits et al., 2012; Vieler et al., 2012). A recent analysis in *Nannochloropsis oceanica* has focused on the four copies of LPATs, addressing their subcellular location and function in the synthesis of eukaryotic precursors, based on single and double knockout (KO) studies (Nobusawa et al., 2017). NoLPAT1 proved to be mainly involved in the transfer of 16:1 to the *sn*-2 position of LPA used for the synthesis of membrane glycerolipids, particularly PC and DGTS. This isoform does not influence TAG biosynthesis. By contrast, NoLPAT4 transfers 16:0 to the *sn*-2 position of LPA purely dedicated to TAG. Eventually, NoLPAT2 and NoLPAT3 are mainly involved in the transfer of 18:1 at the *sn*-2 position of precursors used for PC, PE, and of 16:0 at the *sn*-2 position of precursors for TAG (Nobusawa et al., 2017). Thus, in the early step of the Kennedy pathway, LPAT isoforms seem to control the fate of the produced PA, upstream membrane glycerolipid, and/or TAG pathways. Based on GFP-fusion analyses, NoLPAT1 and NoLPAT2 were likely localized at the ER, whereas NoLPAT3 and NoLPAT4 were located at the periphery of cytosolic LDs (Nobusawa et al., 2017). Based on this differential pattern, the role of NoLPAT2 in TAG formation would suggest that membranes related to the ER may be a platform for LD formation, at least in the early stages of LD biogenesis. The location of NoLPAT3 and NoLPAT4 suggests that they may be involved in the production of TAG loaded in more mature LDs. The location of NoLPAT3 further suggests that some part of the membrane lipid synthesis might occur at the vicinity of LDs. The role of LPATs is therefore likely to be critical in

the control of metabolic routes directed to membrane glycerolipids, TAG, or both. This role demonstrated in eustigmatophytes may be important in diatoms as well.

DAG used in the last step of the Kennedy pathway can stem from *de novo* synthesis as described here, but also from the recycling of membrane lipids (**Fig. 2.8D**) (Li-Beisson et al., 2010, 2019; Boudière et al., 2012; Petroutsos et al., 2014). Indeed, besides *de novo* synthesis via DGATs, TAGs may accumulate through glycerolipid recycling via the activity of a so-called phospholipid:diacylglycerol acyltransferase, or PDAT (Dahlqvist et al., 2000; Ma et al., 2016; Zulu et al., 2018; Falarz et al., 2020). The relative contributions of PDAT versus DGAT in plants is still a matter of debate (see (Woodfield et al., 2018)).

- DAG may derive from membrane glycerolipids, like PC, *via* the action of phospholipases C (PLC) or phosphatidylcholine:diacylglycerol choline transferases (PDAT). In *P. tricornutum*, genes coding for a putative PDAT have been predicted (Dolch et al., 2017b), but to our knowledge, not formally characterized. Their roles and possible involvement in the production of TAG substrates deserve investigation.
- It was also proposed that DGTA and MGDG may serve as a main FA sources for TAG formation under nitrogen formation in *P. tricornutum* (Popko et al., 2016). A number of studies in the green algae *Chlamydomonas reinhardtii* also concur to the important role of FA recycling from MGDG in TAG synthesis (Li et al., 2012b; Gu et al., 2021; Iwai et al., 2021).

The acyl-CoA-independent pathway consists in the transfer of a FA from the *sn*-2 position of a membrane lipid (usually PC, in non-photosynthetic organisms) to the *sn*-3 position of the glycerol backbone of DAG. This pathway was far less investigated than the acyl-CoA-dependent route in algae. Whereas in non-photosynthetic models, such as yeast, PDAT is located at the ER and using phospholipid as an acyl-donor, functional analysis of *Chlamydomonas reinhardtii* PDAT suggested a role in transferring FA groups from plastid membrane lipids as well (Yoon et al., 2012). By contrast with the vast number of DGATs, only one copy of PDAT is usually encountered in photosynthetic stramenopiles, such as *Nannochloropsis s.l.* (Dolch et al., 2017b; Nobusawa et al., 2017). On the basis of transcriptomic results, the PDAT pathway was suggested to be responsible for TAG accumulation in *Microchloropsis (formerly Nannochloropsis) gaditana* CCAP 849/5 alongside the DGAT-dependent synthesis (Mus et al., 2013; Yang et al., 2013, 2016; Alboresi et al., 2016). It seems that some physiological conditions may be in favour of a prominence of the DGAT- or the PDAT-dependent routes; however, these conditions have not yet been characterized and the level of possible redundancy/compensation is unknown. Thus, on the one hand, ^{13}C isotopic labelling results showed that at least in *M. gaditana* strain CCFM-01, most of the FAs in TAGs stemmed from *de novo* synthesis (Janssen et al., 2019a). On the other hand, in a knockdown (KD) line of an elongase converting 16:0-CoA into 18:0-CoA in the cytosol of *M. gaditana*, disturbing membrane glycerolipid composition, only the MgPDAT gene proved to be upregulated and likely responsible for the observed accumulation of TAG (Dolch et al., 2017b). The control of the acyl-CoA-independent pathway needs to be addressed in stramenopiles, as well as possible redundancy with the acyl-CoA dependent routes. Since PDAT allows a connection between TAG production and membrane lipid turnover, a role of the reorganization of cellular membranes needs also to be investigated.

In conclusion of this part dedicated to TAG synthesis, *P. tricornutum* shows a rather rich equipment to esterify acyl chains onto a DAG molecule deriving from the Kennedy pathway. Different DGAT proteins presumably exhibit different substrate specificities in order to cope with the wide array of FAs synthesized by *P. tricornutum*. Nevertheless, *in vitro* enzymatic activity tests on C16:0, C16:1, and C20:5 show no striking differences among the PtDGATs (Zhang et al., 2020). It is not excluded, though, that *in vitro* as well as *in vivo* assays tested to date were not appropriate to determine PtDGAT slight differences in substrate specificity. Moreover, the set of FAs tested *in vitro* was rather reduced. It is not possible to exclude a role of the inactive PtDGAT2s in very specific conditions. DGAT involved in peculiar yet to discover functions to cope with still undetermined conditions may be the reason for a huge genetic expansion in stramenopiles. Gene duplication in diatom is a relatively frequent event (Parks et al., 2018; Osuna-Cruz et al., 2020) and has been suggested as one of the reasons for the undeniable success of diatoms in the world oceans (Busseni et al., 2019).

2.2.1.4. Lipid trafficking

Glycerolipid synthesis described in the previous sections shows the required cooperation between the plastid and the ER in the production of membrane lipids, and implies that large intracellular fluxes of acyl chains occur between the two organelles. Furthermore, glycerolipid synthases are not only spatially separated between organelles, but sometimes asymmetrically located in the membranes (Bell et al., 1981; Benning, 2008). In *A. thaliana* for example, AtMGD1 is considered to be mostly localized in the leaflet facing the intermembrane space of the iEM (Xu et al., 2005), and AtDGD1 is localized in the leaflet facing the inter-membrane space of the oEM (Froehlich et al., 2001). Additionally, some glycerolipids are distributed asymmetrically between the two leaflets of a lipid bilayer. This is the case of the PC found in the oEM in plants, which localises to the lipid layer facing the cytosol (Dorne et al., 1985). Efficient lipid transport is thus needed to maintain membrane composition and structure, and the very existence of some membranes like the thylakoids.

Since in plants, the outer membrane of the plastid thus contains a few endomembrane lipids, such as PC (Dorne et al., 1985), it is possible that some lipids usually associated to the endomembrane are present in the outmost plastid membranes in *P. tricornutum* as well. Furthermore, when plant are subjected to a deprivation in phosphate, the plastid lipid DGDG can be relocalised to the endomembrane system as replacement for phospholipids (Jouhet et al., 2004). This implies that complete glycerolipid structures can be extracted from their host membranes and transported between organelles. Such mechanism has not been evidenced in *P. tricornutum* (Abida et al., 2015). In plants, DAG backbone of ER origin can be used for glycerolipid assembly inside plastids. The identity of the corresponding transported lipid providing this DAG moiety is still a matter of debate, as is the possibility of a similar mechanism in *P. tricornutum*.

To summarize, lipid transfer in cells comprises ‘acyl chain transports’ as well as ‘glycerolipid transports’, and requires mechanisms to move lipids between the two leaflets of a lipid bilayer, between membranes, and between organelles.

Several types of lipid transport mechanisms exist: transmembrane transport by leaflet-to-leaflet flipping and diffusion, transport between membranes with membrane contact sites and protein-mediated processes, and vesicular transport (Benning, 2008; Li et al., 2016).

The mechanisms involved in export and import of acyl chains from and to plastids are still uncertain.

- In plant and algae, the proposed mechanism for the export of newly synthetised acyl-ACP is that it is first converted to a free FA intermediate by an acyl-ACP thioesterase. Then, export through the plastid membranes can occur through facilitated diffusion across membranes (von Berlepsch et al., 2012), and eukaryotic FA transport system (Li et al., 2015). A recently discovered FA transporter in *A. thaliana* called FAX1 performs the transport of free FA across the iEM in the chloroplast (Li et al., 2015). A LACS located to the oEM would then rapidly convert the free FA into an acyl-CoA, usable by acyltransferases. The final transfer of the acyl-CoA to the ER might occur through acyl-CoA diffusion, binding to acyl-CoA-binding proteins (ACBPs) such as ER- and plastid-located LACSs (Li et al., 2016), by relocation through ATP-binding cassette (ABC) transporters (Kang et al., 2011; Kim et al., 2013), or acylation of lysoPC directly at the oEM (Li-Beisson et al., 2013). In *A. thaliana* for example, an ER-located ABC transporter called ABCA9 was shown to be involved in FA/acyl-CoA import to the ER (Kim et al., 2013).
- The import of cytosolic acyl-chains back to secondary plastids, in particular that of EPA, is called the "Omega pathway" and is discussed in section 2.2.1.5.

Passive equilibration of lipids between the two leaflets of a lipid bilayer can be achieved by ATP-independent flippases. However, the identity of these proteins remains largely unclear. In contrast, involved in membrane asymmetry in yeast and mammals are ATP-dependent flippases *sensu lato*, which includes ABC transporters and P-type ATPases. The use of energy through ATP hydrolysis allows these enzymes to move specific lipids against a concentration gradient. In *A. thaliana*, there is a family of P-type ATPases called aminophospholipid ATPases (ALAs) containing 11 members (ALA1 to ALA11). ALA1 and ALA10 were described as being involved in lipid membrane asymmetry (Gomès et al., 2000; Botella et al., 2016). In *Phaeodactylum*, one homolog of ALA (Phatr3_J52368), one of an associated ALIS (Phatr3_J45223) and one homolog of a scramblase (Phatr3_J55111) are detected in the genome, but have not been characterized yet.

Transport mechanisms moving glycerolipids through the plastid membranes are still largely unknown. In *A. thaliana*, an ABC transporter complex made of the TGD1, TGD2 and TGD3 proteins located in the iEM was identified (Xu et al., 2005; Benning, 2008, 2009). The TGD proteins were proposed to be involved in the transport of PA through the plastid membranes from the ER back to the stroma. However, no TGD homologues have been identified in *P. tricornutum* so far, suggesting that this mechanism might not apply to stramenopiles.

Another mechanism might be the transfer of lipids through membrane contact sites (MCS), that have been observed connecting all organelles (Levine and Loewen, 2006; Jouhet et al., 2007; Benning, 2008; Block and Jouhet, 2015; Michaud and Jouhet, 2019). Strong physical interaction between the ER and the plastid have been observed in plants (Andersson et al., 2007a, 2007b), and have been proposed to allow lipid transfer between these organelles (Block and Jouhet, 2015; Michaud and Jouhet, 2019). Furthermore, number of MCSs increases

under phosphate starvation, and MCSs were proposed to be involved in the relocalisation of DGDG to the mitochondria in stress conditions (Michaud and Jouhet, 2019). These mechanisms have not been addressed in diatom.

Finally, vesicular transport *via* budding and fusion of membrane vesicles plays an important role in the trafficking of certain membrane lipids between organelles (van Meer et al., 2008). Several studies suggest a role of vesicular trafficking in the biogenesis of thylakoids (Botté et al., 2011; Boudière et al., 2012; Bastien et al., 2016; Mareš et al., 2019; Gupta et al., 2021). However, in the absence of any evidence of membrane budding in cyanobacteria and chloroplasts, we have recently proposed an origin of thylakoids, at least in cyanobacteria but probably in primary and secondary plastids as well, through a non-vesicular process involving a rapid transition between HexII and Lm phases of MGDG, DGDG and SQDG mixtures (Guéguen and Maréchal, 2021).

In conclusion, glycerolipid biosynthesis in discrete subcellular compartments has led to many forms of lipid trafficking, and many of these mechanisms are still unknown or poorly described.

2.2.1.5. The 'omega pathway' in diatoms, a unique import of VLC-PUFAs from the cytosol to secondary plastids

EPA is the most abundant FA produced in *P. tricornutum* under non-stressed conditions. It is found in all membrane glycerolipids with the exception of PI (Abida et al., 2015). It is particularly enriched in plastid glycerolipids, where high quantities of EPA are found in MGDG. However, EPA is a VLC-PUFA produced in the ER. Thus, there must be a transport mechanism to "bring back" this FA to the plastid for its incorporation in plastid glycerolipids. To date, it is not known under which form the EPA makes this return journey to the plastid. It could be as a free FA, or esterified on a diacyl-moiety from an extra-plastidial glycerolipid (**Fig. 2.9**). This route was tentatively called the 'omega pathway' (Petroutsos et al., 2014).

In algae, several studies have shown that phospholipids, betaine lipids and TAG can provide EPA for the synthesis of MGDG (Haigh et al., 1996; Khozin-Goldberg et al., 2000; Khozin-Goldberg et al., 2002). The possible return of EPA on a recycled diacyl-moiety deriving from these lipids as seen in plants finds no evidence so far in *P. tricornutum*. Indeed, the TGD proteins identified in *A. thaliana* involved in the return of DAG backbones to the plastid have no homologues in *P. tricornutum* (Dolch et al., 2017b). Furthermore, the diacyl composition of MGDG does not fully reflects that of phospholipids and betaine lipids. Newly-formed EPA could be released from a phospholipid or betaine lipid into the cytosolic pool of acyl-CoA and transported to the plastid. Once in the plastid, EPA is assembled on G3P exclusively at position *sn-1* by ATS1 (Petroutsos et al., 2014), while a *de novo* synthetized acyl-ACP of 16 carbons is esterified at position *sn-2* by ATS2. Supporting this pathway, Hao et al recently identified in *P. tricornutum* a LACS enzyme localized to the EpM with high specificity for EPA, called ptACSL1 (Hao et al., 2021). Mutation of ptACSL1 resulted in decreased level of 20:5-containing species in MGDG, suggesting a role of this LACS in the transfer of EPA to the plastid.

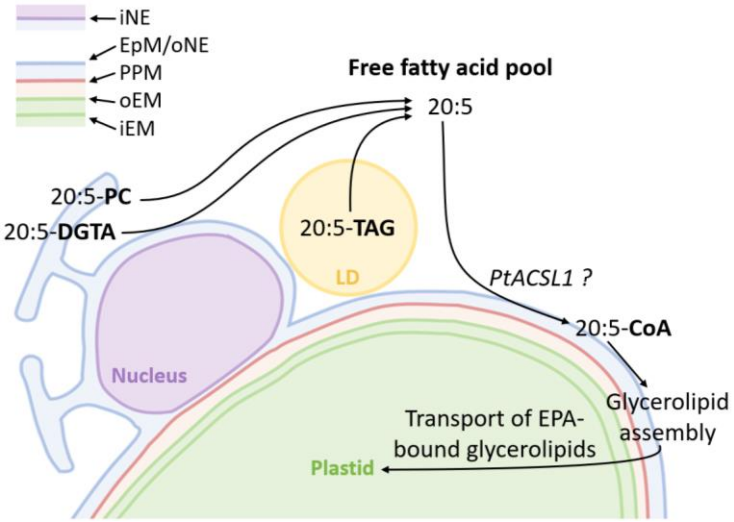


Figure 2.9: Possible scenarios for the Omega pathway in *P. tricornutum*, controlling the return of EPA to the plastid for incorporation in membrane glycerolipids. EPA bound to PC, DGTA or TAG is released to a cytosolic pool of free fatty acids. PtACSL1 would then specifically activate EPA at the EpM, facilitating its use in the assembly of glycerolipids at the EpM. EPA esterified on a glycerol backbone would then be transported inside the plastid, possibly through diffusion of lipids such as DAG and/or lysolipids (Jouhet et al., 2007). cER/oNE, chloroplastic Endoplasmic Reticulum/outer Nuclear Envelope membrane depending on its localisation around the plastid or the nucleus, respectively; DGTA, diacylglyceryl hydroxymethyltrimethyl-β-alanine; iEM, inner Envelope Membrane; iNE, inner Nuclear Envelope membrane; oEM, outer Envelope Membrane; oNE, outer Nuclear Envelope membrane; PC, Phosphatidylcholine; PPM, Periplastidial Membrane; PtACSL1, Long-chain acyl-Coenzyme A synthetase 1 in *P. tricornutum*.

2.2.1.6. Glycerolipid breakdown and β-oxidation

Glycerolipid degradation occurs in all living organism, releasing carbon and energy for the metabolism and vital cellular processes. Lipid degradation occurs during lipid remodelling when entering or leaving stressful conditions. These conditions can be mimicked by controlling nutrient availability. Degradation of glycerolipids can start by cleavage of either the polar head or an acyl chain by enzymes called lipases. Lipases releasing FAs from the glycerol backbone are called carboxylic ester hydrolases and are divided according to their substrate specificity (galactolipases A, phospholipases A, TAG lipases, DAG lipases and monoacylglycerol lipases). Lipases cleaving the head group of phospholipids are called phosphoric diester hydrolases, to which belong phospholipase C and D (Kong et al., 2018).

Several lipases have been identified in algae (Li et al., 2012a, 2012b; Yoon et al., 2012; Trentacoste et al., 2013; Barka et al., 2016; Siegler et al., 2017), mostly involved in TAG breakdown. In *P. tricornutum*, two TAG lipases were identified: one orthologue of SUGAR-DEPENDENT1 (Barka et al., 2016), the major TAG lipase in *A. thaliana* (Eastmond, 2006), as well as one putative patatin-like phospholipase domain-containing protein 3 (PNPLA3) (Wang et al., 2015). Very little is known about lipases involved in membrane glycerolipid degradation in algae. In *C. reinhardtii*, several membrane glycerolipids lipases were identified. First, three MGDG lipases A performing FA cleavage and producing lyso-MGDG were characterized: PGD1, CrGH and CrLAT1 (Li et al., 2012b; Gu et al., 2021; Iwai et al., 2021). PGD1 works preferentially on the *sn*-1 position (Li et al., 2012b), and putative orthologues of PGD1 are present in other algal genomes (Kong et al., 2018). CrLAT1 is a homologue of LPCAT involved in the recycling of MGDG, and thus is able to reacylate lyso-MGDG (Iwai et al., 2021). All three studies show that MGDG degradation is important for TAG accumulation in the green alga *C. reinhardtii*. Second, one PDAT protein in *C.*

reinhardtii is suspected to work as an acyl hydrolase on a broad range of glycerolipids (TAG, phospholipids, galactolipids).

Following their release from glycerolipids, FAs need to be activated by LACS enzymes before FA breakdown, known as β-oxidation. In *P. tricornutum*, it was recently proposed that PtACSL3 and PtACSL5 are involved in activation of FAs that are then channelled to β-oxidation in the mitochondria and the peroxisome respectively (Hao et al., 2021). B-oxidation consists in the serial cleavage of acetyl-CoA from an acyl-CoA. It involves oxidation of the β-carbon position of the acyl-CoA, hydration, dehydrogenation and thiolytic cleavage, and these four steps repeat until complete degradation of the acyl-CoA (C_n) into acetyl-CoAs (C_2). Enzymes involved in β-oxidation steps in algae have been reported (Kong et al., 2018; Li-Beisson et al., 2019).

2.2.2. MGDG synthases (MGDs)

MGDG is produced by the transfer of a galactosyl moiety from the sugar donor UDP-Gal to the DAG acceptor. This reaction is performed by MGDG synthases called MGDs. These enzymes have been characterised in various organisms, revealing their structures, functions, and binding properties.

2.2.2.1. From a two-step ancestral pathway to a one-step optimisation

MGDG is a hallmark of all photosynthetic membranes. Even *Gloeobacter*, which does not contain thylakoids but hosts its photosystem in its cytoplasmic membrane, contains MGDG (Raven and Sánchez-Baracaldo, 2021). Present-day *Gloeobacter* is believed to be close to the most ancient cyanobacteria lacking thylakoid membranes (Guéguen and Maréchal, 2021). Surprisingly, although MGD-containing plant and algae are all derived from cyanobacteria, both *Gloeobacter* and thylakoid-containing cyanobacteria lack MGDs. The galactolipid biosynthesis pathway in these photosynthetic prokaryotes involves a distinct set of enzymes than in plastids (**Fig. 2.10**).

In cyanobacteria, MGlcDG is first produced by a MGlcDG synthase (MgdA) (Awai et al., 2006; Apdila and Awai, 2017); then, the glucosyl polar head (in β conformation) is epimerized into a galactosyl residue, thus forming MGDG; eventually, DGDG is formed by the action of a digalactosyldiacylglycerol synthase (DgdA) (Awai et al., 2007; Sakurai et al., 2007; Apdila and Awai, 2017). The activity converting MGlcDG into MGDG in *Gloeobacter* was measured (Sato, 2015) but the corresponding enzyme has yet to be characterized molecularly. It has been tentatively called MgdX (Apdila and Awai, 2017). With a few exceptions, thylakoid-containing cyanobacterial clades which have radiated from a *Gloeobacter*-like ancestor contain another epimerase, called MgdE, shown to convert MGlcDG into MGDG (Awai et al., 2014).

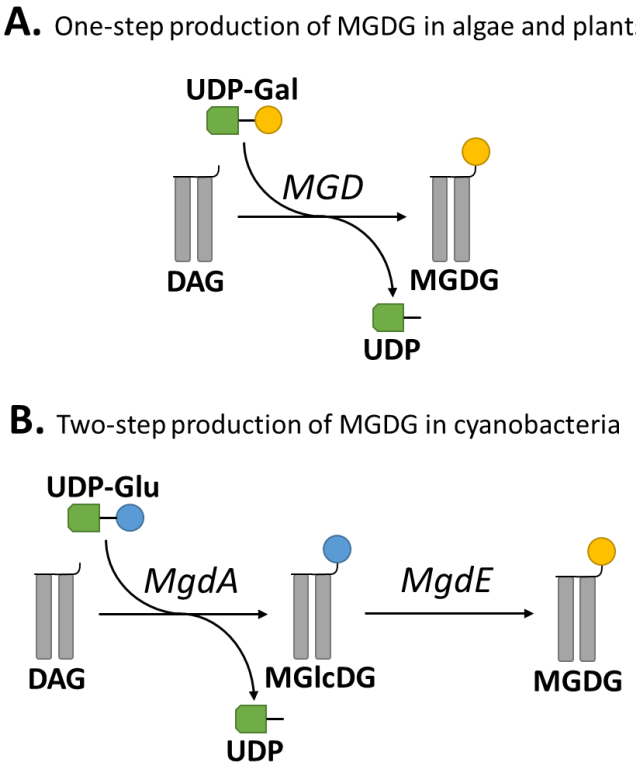


Figure 2.10: MGDG synthesis in plant and algae *versus* cyanobacteria. **A**, in plant and algae, MGDG is synthesised directly from DAG by an MGD. **B**, in cyanobacteria, DAG is first converted into a MGlcDG by an MgdA, then the glucose head group is converted to galactose by an epimerase (MgdE). Sugars are represented according to the official sugar symbol nomenclature: a yellow circle for galactose and a blue circle for glucose (Varki et al., 2015).

In plant and algal genomes, MGDs and DGDs are unrelated to their cyanobacterial counterparts (Mizusawa et al., 2009). No orthologs of MgdA have been identified (Awai et al., 2006), and only some red algae retained a cyanobacterial DgdA (Petroutsos et al., 2014). No cyanobacteria contain an ancestral type of MGD either, suggesting that MGDs in algae and plants were acquired by lateral gene transfer. Based on phylogenetic studies, MGDs are likely to have evolved from bacterial enzymes (Masuda et al., 2011; Yuzawa et al., 2012). In particular, an MGD identified in the bacteria *Chloroflexi* has a close relationship with eukaryotic MGDs (Yuzawa et al., 2012). It was proposed that this enzyme might be the ancestor MGD that would have been acquired by lateral gene transfer in parallel with the primary endosymbiosis. The simplification of the galactolipid biosynthetic pathway offered by the bacterial MGD compared to the MGlcDG synthase/epimerase system was then retained in evolution.

2.2.2.2. Classification and structure

MGDs, as well as MGlcDG synthases and DGDs, belong to the glycosyltransferase (GT) enzyme class. GTs are defined as enzymes catalysing the transfer of a sugar moiety from an activated donor molecule to an acceptor molecule. 114 families of GTs have been categorized according to the carbohydrate-active enzymes (CAZy; http://www.cazy.org) classification. MGDs belong to the GT28 family (Campbell et al., 1997; Coutinho et al., 2003; Levasseur et al., 2013). Interestingly, the cyanobacterial MGlcDG synthases contains motifs found in other GTs families, GT2 and GT21. On the other hand, both plant and cyanobacterial DGD belong to the GT4 family.

GTs can be functionally divided in two groups according to the stereochemistry of their products, which depends on whether the anomeric conformation of the sugar is 'retained' or 'inverted' (Sinnott, 1990). In the course of the enzymatic reaction performed by MGDs, the galactosyl group passes from an α configuration in UDP-Gal, to a β configuration in MGDG. Thus, MGDs are part of the 'inverting' GTs.

GTs can also be structurally divided into two main superfamilies, GT-A and GT-B, based on their three-dimensional folding (Breton et al., 2012). There is however no correlation between the stoichiometry of the reaction and the three-dimensional structure displayed by an enzyme (Breton et al., 2012). MGDs belong to the GT-B superfamily, thus presenting the conserved GT-B fold. The GT-B fold includes two Rossmann-type α/β/α domains of comparable size, named the N-domain and the C-domain, separated by a large cleft, and stabilized by two long C-terminal α-helices (**Fig. 2.11A**).

In angiosperms, MGD proteins are further sub-divided into two 'types', designated as type A and type B (Miège et al., 1999; Awai et al., 2001). Type A MGDs presents a cleavable N-terminal region of about 100 amino acids involved in the targeting of the protein to the chloroplast, whereas type B only have a short non-cleavable N-terminus of about 40 amino acids. For example, *A. thaliana* contains three enzymes of both types: AtMGD1 is a type A MGD whereas AtMGD2 and AtMGD3 are of type B (Awai et al., 2001). Comparison of *AtMGD1*, *AtMGD2* and *AtMGD3* genes showed that whereas *AtMGD1* and *AtMGD3* share a similar positional organisation of their eight exons, *AtMGD2* lacks two introns (Awai et al., 2001). *AtMGD1* and *AtMGD3* were suggested to have emerged from an early divergence, followed by deep modifications of their N-terminal part, and that *AtMGD2* derived from a gene duplication of an ancestral *AtMGD3* sequence. Phylogenetic studies showed that MGD proteins from different species cluster in two distinct group corresponding to type A and type B (Awai et al., 2001; Yuzawa et al., 2012). Further analysis of MGD enzymes phylogeny in Viridiplantae proposed that the divergence between type A and type B occurred after the emergence of Spermatophyta, but preceded that of Gymnosperm and Angiosperm, and was dated around 323 million years ago (Yuzawa et al., 2012).

To date, AtMGD1 is the best-characterised enzyme, and will be referred to as MGD1 for simplicity. Primary sequence of MGD1 contains 533 amino acids. It starts with the cleavable region typical of type A MGD (amino acid 1-106), followed by a short nucleotide sequence predicted to be disordered and not necessary for protein activity (amino acids 107-136) (Dubots et al., 2010). Most of the following sequence makes a large catalytic domain (amino acid 137-533) (Rocha et al., 2016). This catalytic domain (catdomMGD1, comprising the canonical N- and C-domains of a Rossmann fold) has been successfully expressed in *E. coli* and purified, allowing further characterisation of its biochemical and structure properties (Rocha et al., 2013). Notably, crystallisation of catdomMGD1 has been achieved in its apo-form and in complex with UDP (Rocha et al., 2016). Interestingly, the closest structural homologues of catdomMGD1 are bacterial GTs required for assembly of peptidoglycan (MuRG, family GT28), or synthesis of antibiotics (CalG1 and OleI, family GT1) (Hu et al., 2003; Bolam et al., 2007; Chang et al., 2011). A computational model was generated to predict catdomMGD1 structure in complex with UDP-Gal based on previous studies of MurG in complex with UDP-GlcNAc (Hu et al., 2003; Rocha et al., 2016).

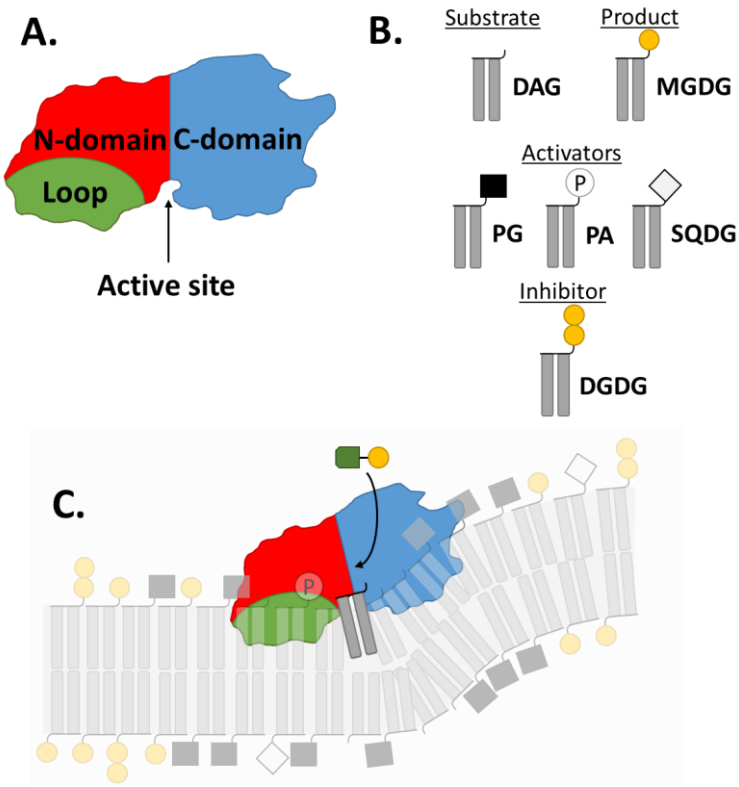


Figure 2.11: MGD1 structure and binding properties. A, molecular structure of MGD1: Two Rossmann folds corresponding to the N-domain (red) and C-domain (blue). The loop region is indicated in green inside the N-domain. The active site is located in the cleft between the two domains. **B**, list of glycerolipids inside the plastid membrane and their relation to MGD activity. **C**, model of MGD binding to the membrane. The MGD protein is localised in a curved region of the membrane with the N-domain at the bottom of the curve. The MGD protein binds to a PG-DAG cluster where DAG is surrounded by PG. A DAG is represented in the active site and the hydrophilic UDP-Gal donor is shown above the protein. A PA molecule is placed near the active site to activate enzyme activity. Adapted from (Makshakova et al., 2020).

In MGD1, the C-domain is involved in sugar donor binding. This domain is the most conserved among MGD1 homologues, reflecting conserved features in donor sugar accommodation (Rocha et al., 2016). MGD1 cannot use UDP-glucose as a donor, this sugar isomer resulting in a steric clash with a proline (Awai et al., 2006; Rocha et al., 2016). The N-domain however, involved in acceptor substrate binding, shows pronounced variations mirroring the diversity of acceptors substrates (Rocha et al., 2016). The catalytic domain of MGD1 displays a surprisingly large disordered region of about 50 amino acids residues inside the N-domain, conferring an elongated shape to the protein, an unusual feature among GT-B enzymes. The high flexibility of this region, designated as an internal “loop”, prevented its crystallisation. Without considering the loop, the N-domain (residues P140-V318 and E507-N524) shows the typical central parallel β-sheets (at least six) surrounded by α-helices (at least five). It is separated from the C-domain by a long loop (residues Y319-P333). The C-domain (residues K334-P506) is composed of six stranded β-sheets surrounded by eight α-helices, plus a β-hairpin protruding from the C-domain. The C-terminal part of the protein ends as an alpha helix in the N-domain for stabilisation, a signature of GT-B enzymes (Rocha et al., 2016).

Two catalytic residues have been identified, one in each domain. Residue His155 of MGD1, located in a glycine-rich loop in the N-domain, is involved in the binding of the acceptor substrate. Residue K430 in the C-domain binds the sugar donor. Importance of these residues for activity was demonstrated by point mutations (Botté et

al., 2005; Dubots et al., 2010; Rocha et al., 2016). The detailed mechanism of the enzymatic reaction is discussed in the following section.

A role for MGD1 loop has been investigated. A catdomMGD1 mutated form missing a large portion of the loop was shown to be inactive, although properly folded into a smaller protein (Rocha et al., 2016). The loop sequence presents an intriguingly high number of aromatic and basic residues, suggesting interactions with the acceptor substrate and/or the membrane. In particular, W182 was shown to play a role in DAG binding (Botté et al., 2005). Further evidence of the role of the loop in the binding of DAG but also in the binding of MGD1 activator PG is discussed hereafter.

2.2.2.1. Enzymatic reaction and activators

In two structural homologues of MGD1, OleI and CalG1, a catalytic His-Asp/Glu dyad has been described (Bolam et al., 2007; Chang et al., 2011). Hydrogen bonding between these positive and negative residues facilitates the deprotonation of the nucleophile OH group of the DAG acceptor (Rocha et al., 2016). This mechanism is however absent in MGD1 as no equivalent negative residue is present.

Interestingly, MGD enzymes need anionic lipids to be active, such as PA, PG and SQDG (**Fig. 2.11B**) (Coves et al., 1988; Maréchal et al., 1991, 1994; Dubots et al., 2010; Rocha et al., 2013; Nitenberg et al., 2020). MGD1 activation by PG has been the most studied to date, and key amino acids have been identified through mutational studies. Activation by SQDG is thought to proceed through the same mechanism as PG while PA however is believed to proceed through another mechanism. Indeed, PG and PA activation effects are synergic (Dubots et al., 2010). Furthermore, point mutations of P189 and W287 abolished activation by PG and SQDG, but not by PA, suggesting distinct binding sites (Dubots et al., 2010; Rocha et al., 2016; Nitenberg et al., 2020). Finally, PA activation exhibits an allosteric behaviour not observed with the two other lipid classes (Dubots et al., 2010).

PG is suggested to bind closely to the DAG binding site, and even interact with DAG, giving to this lipid a co-catalytic role (Makshakova et al., 2020). Residues in contact with PG are partly present in the globular portion of the N-domain, and partly in the loop. The second are also in contact with the DAG acceptor (Makshakova et al., 2020). Thus, a DAG and a PG molecules would colocalize at the active site of MGD1 (Nitenberg et al., 2020).

The role of PG in a potential PG-His catalytic dyad has been proposed as the catalytic His155 residue is suggested to bind PG (Makshakova et al., 2020; Nitenberg et al., 2020). This system would form an acid-base relay system similar to the His-Asp/Glu dyad and imply a catalytic role of anionic lipids. R156 and P189 are also expected to bind PG, stabilizing the MGD1-PG interaction, and help bring PG close to His155 (Rocha et al., 2016; Makshakova et al., 2020; Nitenberg et al., 2020). Thus, the catalytic reaction would involve binding of the anionic PG to MGD1 active site, creating an acid-base catalytic dyad with His155 facilitating deprotonation of the OH group of the DAG by the His base, while the basic residue K430 stabilizes the phosphates of the departing UDP.

Activation by PA is more puzzling. A hypothesis would be that PA, having a smaller head than PG and SQDG, would also perform a hydrogen bonding with His155 but in a different way than the two other lipid classes. It is also a possibility that PA has another regulatory binding site explaining its allosteric behaviour, not observed with PG and SQDG (Dubots et al., 2010; Makshakova et al., 2020; Nitenberg et al., 2020).

In addition, *in silico* structural modelling and simulations suggested that a change in the protein orientation with respect to the membrane might allow the switch between binding of the hydrophobic DAG and of the hydrophilic UDP-Gal (Nitenberg et al., 2020). Indeed, while the N-domain remains attached to the membrane, in simulation experiments, the C-domain can altern within 1 µs between an orientation away from the membrane, and an orientation that would bring it at the vicinity of the membrane. Only when close to the membrane can surface residues of the C-domain and catalytic H155 interact with PG. Anchoring of the N-domain in the membrane depends on the membrane lipid composition described in the following section.

2.2.2.2. Membrane binding properties

Arabidopsis MGD1 needs to be in position to reach its two substrates, one being hydrophobic, DAG, the other being water-soluble, UDP-Gal. It was shown that the protein binds to the surface of the iEM without spanning the entire bilayer (monotopic protein) (**Fig. 2.11C**) (Miège et al., 1999). A topology of MGD1 facing the intermembrane side of the iEM was initially proposed based on the apparent lack of UDP-Gal in the chloroplast stroma (Maréchal et al., 1995; Miège et al., 1999). However, an epimerase converting UDP-Glc to UDP-Gal in the stroma was detected in rice, and mutation of its homologue in *A. thaliana* leads to an embryonic defect (Li et al., 2011). Therefore, it cannot be ruled out that UDP-Gal could be available in the stroma and that MGD1 could face the stromal side of the iEM. In our laboratory, we even consider that both orientations, at the stromal and external sides of the iEM, could be possible.

Molecular dynamic simulations showed that the iEM composition likely promotes the formation of curved membranes, due to the high amount of lipid containing polyunsaturated tails (Makshakova et al., 2020). In these simulations, MGD1 locates on the slope of membrane wave with N-domain placed on the bottom of the curve. In this position, MGD1 active site is brought closely to the membrane surface. Contact between the protein and the membrane are thought to be made *via* interaction between hydrophobic residues of the N-domain and the lipids (Botté et al., 2005; Sarkis et al., 2014; Makshakova et al., 2020).

Several studies investigated the influence of lipid composition on MGD1 binding to membranes *in vitro* (Sarkis et al., 2014; Rocha et al., 2016; Makshakova et al., 2020; Nitenberg et al., 2020). Firstly, MGD1 appears as a very uncommon enzyme having a high affinity for its product MGDG, facilitating its binding to the MGDG-rich iEM and leading to the local accumulation of MGDG (Boudière et al., 2014; Sarkis et al., 2014). The cone-like shape of MGDG and its composition in unsaturated tails might induce curvatures of the membrane, creating a favoured binding site for MGD1. In contrast, MGD1 is excluded from membranes enriched in DGDG, revealing an unconventional retro-control by this down-product of MGDG, and emphasizing the importance of maintaining a proper MGDG/DGDG ratio in chloroplast membranes (Sarkis et al., 2014).

Interestingly, the MGD1 activators PA and PG were shown to exert a high attraction on MGD1 through electrostatic interactions that can counteract the negative effect of DGDG (Sarkis et al., 2014). His155 had a particularly important role in the binding to PG (Nitenberg et al., 2020).

The highest membrane binding affinity shown by MGD1 *in vitro* was obtained when introducing its DAG substrate. Rocha et al found that MGD1 binding rate to DAG monolayer was 10 times higher than with other

tested membranes (Rocha et al., 2016). Study of MGD1 in different membrane compositions ranked the strength of its interaction as follow: DAG ~ PG > MGDG ~ PA > MDP ~ DGDG, where 'MDP' corresponds to a mixed MGDG/DAG/PG composition mimicking that of the natural iEM (Rocha et al., 2016). Capture of the DAG substrate was suggested to be achieved by the loop. Indeed, the use of a catdomMGD1 variant missing a large portion of the loop showed that with this mutant, binding affinity for DAG was drastically dropped (Rocha et al., 2016).

Interestingly, it was hypothesized that PG contributes to the anchoring of MGD1 in specific sites where DAG accumulates.

2.2.2.3. Formation of DAG-PG cluster

In the iEM, DAG accounts for less than 1 mol% of the lipid composition. Conversion of DAG to MGDG is thus a rapid process, requiring an efficient recognition of DAG molecules by MGD1.

Computational simulations revealed that PG molecules tend to form nano-sized cluster embedding DAG molecules (Makshakova et al., 2020). This formation not only increases local concentration of DAG but presence of DAG and PG together allows the efficient binding of MGD1 to both its substrate and activator. Consistently, the simulations showed that MGD1 has a tendency to locate near these clusters (**Fig. 2.11C**) (Makshakova et al., 2020). Furthermore, presence of MGD1 itself triggers a higher accumulation of DAG molecules in the PG/DAG clusters at its vicinity (Makshakova et al., 2020; Nitenberg et al., 2020). In the simulations, MGD1 had a larger number of contacts with DAG molecules than with PG molecules. This phenomenon results from the interactions between the glycerol moiety of DAG with the surface residues of MGD1, and is facilitated by the lateral and transverse diffusion of DAG molecules through galactolipid membranes. It was thus hypothesize that PG would not only activate MGD1 but also bring it close to its substrate (Makshakova et al., 2020). Altogether, these data suggest the existence of spontaneous and/or induced nano-domain organisation of the iEM.

Interaction of PG and DAG with MGD1 is mostly performed by residues of the loop, suggesting that this region contributes to the anchoring of MGD1 to the membrane as well as capture of PG and DAG (Nitenberg et al., 2020).

2.2.2.4. Expression and location of MGDs in *Arabidopsis thaliana*

As mentioned in section 2.2.2.1, *A. thaliana* contains a multigenic family of MGDs, with MGD1 belonging to type A, whereas MGD2 and MGD3 belong to type B. Each MGD differ in terms of gene expression profile, subcellular localisation and substrate specificity.

MGD1 presents an N-terminal chloroplast transit peptide (cTP), allowing this nuclear-encoded protein to cross the oEM and localise to the iEM of chloroplasts (Miège et al., 1999; Awai et al., 2001). No predictable targeting sequences are visible on the short N-terminal portion of MGD2 and MGD3, which were shown to localise, at least, to the oEM (Awai et al., 2001). Consistently with this subcellular localization, MGD1 is more likely to be involved in the expansion of the iEM and the massive production of MGDG required for the building of thylakoids, whereas MGD2 and MGD3 might contribute more generally to the outer envelope membrane expansion. In addition, all three enzymes can use any type of DAG independently of their prokaryotic/eukaryotic origin, though

type B enzymes use preferentially DAG of eukaryotic origin, consistently with their more peripheral location (Awai et al., 2001).

Type A MGD1 is the most active enzyme. Its gene is abundantly expressed in green tissues of the whole plant at all developmental stages, except siliques and seeds (Awai et al., 2001). Type B *MGD2* and *MGD3* gene expression are mostly detected in non-photosynthetic tissues. *MGD2* shows a low and continuous expression throughout all developmental stages, in contrast to *MGD3* which is abundantly expressed at early developmental stages and at low levels at later stages (Awai et al., 2001). *MGD2* and *MGD3* are particularly expressed in inflorescences and roots, respectively. In these tissues, low levels of MGDG are present, consistent with the presence of small-sized plastids deprived of thylakoids. A study of MGD proteins localisation using β-glucuronidase (GUS) assays was consistent with gene expression (Kobayashi et al., 2004). This study further revealed that MGD1 only was found in etiolated cotyledons, that MGD2 and MGD3 also localise at leaf tips, and that all three enzymes were detected in pollen grains. In the pollen, only type B promoters were active after pollination.

Interestingly, *MGD2* expression increases upon pollen grains hydration and in pollen tubes, and the protein localises to the cytosol (Billey et al., 2020, 2021). Its mRNA expression level and distribution of the protein along the pollen tube is controlled by an RNA-binding protein called LARP6C. LARP6C may act as a translation repressor until the bound mRNA reaches the site where translation is required. At this site, LARP6C would switch to an activator of *MGD2* mRNA translation (Billey et al., 2020). Very importantly, the role of a cytosolic MGD2 in the pollen tube, in the absence of any plastid, is not understood. It was proposed that it could act as a cytosolic transporter of DAG from a donor compartment to a site of MGDG synthesis (Billey et al., 2021). The presence of MGD2 also coincides with DGDG found in the plasma membrane of the elongating pollen tubes, and may act in this unconventional extraplastidic galactolipid pathway (Botté et al., 2011).

The expression of type B enzymes is enhanced under phosphate deprivation, while *MGD1* expression remains almost constant (Awai et al., 2001; Kobayashi et al., 2004). Type B MGDs were particularly present in roots. Consistently, DGDG (produced from MGDG) accumulates greatly in the root under phosphate starvation (Härtel and Benning, 2000). Therefore, MGD2 and MGD3 are suspected to be involved in the feeding of MGDG species for the synthesis of DGDG under phosphate starvation. Furthermore, DGDG produced to replace the decreasing phospholipids in this condition is of eukaryotic structure, which is coherent with type B enzyme substrate preferences. It has been hypothesised that emergence of type B MGDs in Spermatophyta might have occurred to adapt to environmental changes (Petroutsos et al., 2014).

Interestingly, MGDs and DGDs have been proposed to operate in pairs (Petroutsos et al., 2014). *A. thaliana* contains two DGDG synthases, DGD1 and DGD2 (Dörmann et al., 1999). Expression of MGDs and DGDs in green versus non-green tissues and in response to environmental condition suggests that MGD1/DGD1 pair works for the production of galactolipid for thylakoids formation whereas (MGD2; MGD3)/DGD2 pairs work together for the production of galactolipids in non-green tissues and in response to environmental stresses (Awai et al., 2001; Froehlich et al., 2001; Kelly et al., 2003; Kobayashi et al., 2004).

2.2.2.5. MGDs in P. tricornutum

In *P. tricornutum*, three putative MGDs have been predicted previously (Phatr_14125, Phatr_54168 and Phatr_9619) (Petroutsos et al., 2014). In this manuscript, each of these MGDs will be referred to as follow:

- MGDα for Phatr_14125;
- MGDβ for Phatr_54168;
- MGDγ for Phatr_9619.

MGDα contains an unambiguous bipartite presequence suggesting that this protein localises to the plastid (**see Section 2.1.1.3**). MGDβ N-terminal sequence shares some features of bipartite presequence, and MGDγ does not contain any predictable targeting sequence. Therefore, these putative MGDs might localise in different plastid membranes or possibly outside of the plastid (Petroutsos et al., 2014).

A number of whole-genome microarray studies of *P. tricornutum* have been realised in different conditions, in which probes for *MGDα*, *MGDβ*, and *MGDγ* were present. From these, it appears that none of the putative MGD genes are differentially regulated in response to silica starved conditions (Sapriel et al., 2009), high light stress (Nymark et al., 2009), cadmium exposure (Brembu et al., 2011), exposure to dispersants (Hook and Osborn, 2012), re-illumination after dark treatment (Nymark et al., 2013), and exposure to different light radiations (Valle et al., 2014). A study of *P. tricornutum* acclimated to light/dark cycles showed that *MGDγ* expression increases during the light phase and is down regulated in the dark (Chauton et al., 2013). *MGDγ* expression might thus be related to processes occurring during the light period.

RNAseq analyses were also conducted on *P. tricornutum*. In an RNAseq performed by Dolch et al, *MGDα* expression was not detected, indicating that this putative MGD gene might have a very low basal expression level (Dolch et al., 2017a). In this study, no differential expression of *MGDβ* and *MGDγ* was detected following exposure to nitric oxide, although nitric oxid could posttranslationally inhibited MGD enzymatic activity. Another study on the effect of ethynylestradiol, a steroid hormonal disruptor, on TAG accumulation showed that none of the putative MGD genes was differentially regulated despite a slight decrease in MGDG content (Conte et al., 2018).

Interestingly, a study of phosphate shortage revealed the differential expression of all putative MGD genes in *P. tricornutum* (Yang et al., 2014). The expression of *MGDβ* was increased by 2.7-fold while *MGDα* and *MGDγ* were downregulated by 1.8 and 2-fold respectively. These results highly suggest that MGDβ may play an important role under phosphate limitation. Furthermore, the whole-cell proteome of *P. tricornutum* under nitrogen deprivation condition revealed a reduction of MGDα and MGDγ protein levels while MGDβ protein level remained stable (Lupette et al., 2019). This suggest that conservation of MGDβ activity might be important under both phosphate and nitrogen deprivation.

Eventually, in the raw data of two replicates over three of *P. tricornutum* LD proteome realised by Leyland *et al* (Leyland et al., 2020), the MGDγ enzyme was detected. In *P. tricornutum* LD proteome realised by Lupette *et al* (Lupette et al., 2019), this protein was not found. Presence of a MGD in the LD proteome is intriguing as MGDs

are usually plasitidal proteins (with the exception of AtMGD2 in the pollen tube). Detection of MGDγ following LDs purification could either mean that MGDγ actually plays a role in LDs, or that it is a contamination, potentially by the EpM.

2.2.3. MGDG, the most abundant membrane lipid on Earth

MGDG was first isolated in 1956 from wheat flour (*Triticum aestivum*) (Carter et al., 1956). This galactolipid is present in all photosynthetic membranes and therefore was recognised as a hallmark of oxygenic photosynthesis (Boudière et al., 2014), and the most abundant lipid class on Earth (Gounaris and Barber, 1983). This high level of conservation is believed to be connected to its crucial structural (**Fig. 2.12**) and functional roles in photosynthetic membranes.

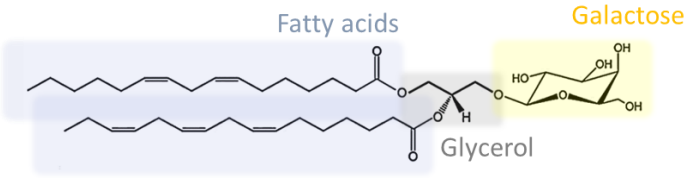


Figure 2.12: Biochemical structure of an MGDG molecule. MGDG is composed of two fatty acids (blue) esterified on a glycerol backbone (black) in position *sn-1* and *sn-2*, and possesses a galactose head group (yellow) in position *sn-3* of the glycerol backbone. In this representation, a MGDG 16:2-16:3 species is taken as an example.

2.2.3.1. Subcellular localization in primary plastids

Galactolipids are major components of photosynthetic membranes, and are only rarely encountered in other cell membranes. The proportion of MGDG and DGDG in each membrane in primary plastids is rather conserved from cyanobacteria to higher plants (Block et al., 1983; Mendiola-Morgenthaler et al., 1985; Boudière et al., 2014). The thylakoids and the iEM show a very similar composition characterized by a high proportion of MGDG. In these two membrane types, MGDG and DGDG account for around 50 and 30 mol% of membrane glycerolipids respectively (Block et al., 1983; Mendiola-Morgenthaler et al., 1985; Boudière et al., 2014). In contrast, the amount of DGDG in the oEM exceeds that of MGDG, with around 30 mol% of DGDG against 20 mol% of MGDG, and presents a high proportion of PC (30 mol%). Such differences between membranes presumably reflect their different functional properties.

In organisms deriving from secondary or tertiary endosymbiosis events, the composition of each plastid membrane is still unknown and must be inferred based on its evolutionary origin (**see section 2.1.1.2**). Thus, the iEM and oEM in the complex plastid of *P. tricornutum* may present a similar composition as that of the corresponding iEM and oEM in primary plastids.

2.2.3.2. Disruption of MGD activity in *A. thaliana*

Importance of MGDs has been investigated in *A. thaliana* by a functional genomic approach, *via* the study of knockdown and knockout mutants, inducible gene-suppression system, and by a chemical genetic approach, *via* treatments with a chemical inhibitor called galvestine-1 (Jarvis et al., 2000; Kobayashi et al., 2007; Botté et al., 2011; Boudière et al., 2012; Fujii et al., 2014).

MGD1 being the most important of the three MGDs in *A. thaliana*, the generation of a knockout line proved to be challenging. Indeed, total knockout of MGD1 results in embryo development arrest (Kobayashi et al., 2007). The first MGD1 mutant characterised was therefore a knockdown mutant, *mgd1-1* (Jarvis et al., 2000). In this mutant, a 75 % reduction was observed in both *MGD1* mRNA levels and MGD activity. This has led to a 42 % decrease in MGDG in the leaves compared to the WT, with an increased proportion of MGDG with eukaryotic signature. This is consistent with type B MGDs, compensating the partial loss of MGD1, and having a substrate preference for eukaryotic DAG substrates (Awai et al., 2001). The mutant has a yellow-green phenotype, correlated with a 50 % deficiency in total chlorophyll per plant and underdeveloped small round chloroplasts with fewer thylakoid membranes (Jarvis et al., 2000).

Years after the characterisation of *mgd1-1*, a knockout mutant *mgd1-2* was obtained (Kobayashi et al., 2007). Germination of *mgd1-2* seeds can only occur when grown heterotrophically on sucrose medium. The resulting albino seedlings lack galactolipids (MGDG content reduced by 98 %), and total impairment of photosynthetic ability prevents their growth in autotroph conditions. Such a loss of MGDG demonstrates that type B MGDs are not able to compensate for the complete loss of MGD1 function. *Mgd1-2* mutants cannot accumulate chlorophyll, and present severely underdeveloped or no internal membrane structures in the chloroplast. Total plant protein was generally similar in the mutant and the WT, however major synthetic proteins of thylakoids membranes were undetectable in the mutant. These results showed the crucial role of MGD1 in synthesising the bulk of MGDG. Interestingly, invaginations of the iEM were observed, either resulting from an artifact created by the inability of the plastid to form proper thylakoids, or suggesting that such invagination of the iEM could initiate thylakoid formation (Kobayashi et al., 2007). This latter hypothesis seems unlikely, since no iEM invagination is detected in WT chloroplasts, whereas such phenomenon should be frequently observed in normal conditions if thylakoids had to emerge from a budding of this membrane. We therefore suggest that this phenotype reflects an excess of bilayer forming lipids thus protruding, whereas Hexagonal II MGDG would rather promote a vesicle-independent formation of thylakoids (Guéguen and Maréchal, 2022). The dramatic decrease in DGDG in this mutant also demonstrates that MGD1 is predominantly responsible for the bulk of DGDG synthesis.

In order to characterise the role of MGD1 in seedlings without the drawback of working in heterotrophic conditions, an inducible gene-suppression system targeting *MGD1* was realised (Fujii et al., 2014). This study further highlights the importance of MGD1 in thylakoid membranes formation, especially during cotyledon greening. The role of galactolipids in photosynthetic electron transport and accumulation/maintenance of the photosystem core complexes was also suggested. Interestingly, MGD1 suppression downregulates genes responsible for photosynthesis and photorespiration and upregulates those involved in the glyoxylate cycle, suggesting a coordinated regulation of galactolipid synthesis and genes involved in those pathways.

The possible role of MGD1 in etioplasts (non-photosynthetic plastids developing in darkness) suggested by its gene expression localisation was surprising as *mgd1-1* does not seem to have an impact on etioplast architecture, nor does a *mgd2 mgd3* double mutant on MGDG composition in etiolated seedlings (Jarvis et al., 2000; Kobayashi et al., 2004, 2009). Recently however, using the inducible gene-suppression system, a 36 % loss in MGDG content was achieved in etiolated seedlings (with no impact of DGDG) that correlated with strong impairment of

prothylakoids formation, irregularly shaped prolamellar bodies, and altered etioplast shape (Fujii et al., 2017). Furthermore, formation of a chlorophyll intermediate, protochlorophyllide (Pchlide) was impaired, suggesting the possible role of MGDG in membrane-associated processes of the Pchlide biosynthesis pathway. Upon illumination, differentiation of etioplasts into chloroplasts and accumulation of chlorophyll was impaired (Fujii et al., 2019). The authors suggested that this phenotype may result from an altered MGDG/DGDG ratio.

Single and double mutants of *MGD2* and *MGD3* under optimal conditions have no impact on lipid composition in the shoot (Kobayashi et al., 2009). They are however important for DGDG synthesis in roots as a 41 % decrease in DGDG content is observed in the double mutant. Nonetheless, mutants do not show any impact on galactolipid FA composition in either shoot and root, and primary root growth was not impacted either. This further supports the idea that Type B enzymes are not essential for galactolipid synthesis and plant growth under optimal conditions (Kobayashi et al., 2009).

By contrast, the role of combined MGD2 and MGD3 reveals to be crucial for adaptation to low-phosphate conditions (Kobayashi et al., 2009). While knockout *mgd2-1* shows no or weak impact on the plant phenotype, knockout mutant *mgd3-1* has an important impact on the plant fitness under low-phosphate conditions, and phenotype was further aggravated in the double mutant *mgd2 mgd3*. Indeed, in the absence of both enzymes the production of DGDG under phosphate deprivation was significantly reduced, especially in roots where it was almost completely abolished. Although FA composition of galactolipids in a WT is altered under phosphate deprivation, no changes were observed in the *mgd2 mgd3* double mutant. Interestingly, the decrease in galactolipids in the shoot was compensated by a two-fold increase of PE and PC species but not of PG and PI. This suggests that MGDG produced by MGD2 and MGD3 in phosphate starvation feeds the synthesis of DGDG species destined to replace PC and PE. The inability to accumulate DGDG species during lipid remodelling triggered by phosphate deprivation led to reductions in fresh weight, root primary growth, and photosynthetic efficiency (particularly in cotyledons) in the *mgd2 mgd3* double mutant.

Based on a high-throughput chemical screen, a chemical inhibitor specific for the multigenic MGDs family in *A. thaliana* was developed (Botté et al., 2011). The molecule, called galvestine-1, competes with DAG binding on MGDs. As the generation of a triple *mgd1 mgd2 mgd3* mutant has not been reported to date, use of a chemical inhibitor serves as an alternative and complementary strategy to study the decreased activity all MGDs in the plant tissues. The use of galvestine-1 at the whole plant level was consistent with previous observations of MGD1 impairment effects such as reduced primary and secondary root growth, chlorosis, smaller chloroplast size and lower thylakoid content, invagination of the iEM, decrease of galactolipid content and increase of phospholipids proportion, and lethality after long-term treatment in autotrophic conditions.

Study of pollen tube growth in the presence of galvestine-1 revealed a negative impact on pollen tube elongation rate. This is surprising because in the pollen, only type B MGDs are present and double *mgd2 mgd3* mutant did not have an impact on pollen tube growth (Kobayashi et al., 2004, 2009). This study thus suggest a role of MGDG in the membrane expansion of growing pollen tubes (Botté et al., 2011). In the pollen tube, *MGD2* is the most abundant enzyme, and its expression is regulated by LARP6C (Billey et al., 2020). Interestingly, the use of galvestine-1 on *larp6c* mutant leads to a strong inhibition of pollen tube guidance (Billey et al., 2020). This

phenotype was not observed in WT pollen tubes treated with galvestine-1, and *larp6c* mutant pollen grown without the inhibitor has only a moderate inhibition of pollen guidance. These findings suggest that there may be a synergic effect between the loss of LARP6C and inhibition of MGD2 activity on pollen tube guidance.

In conclusion, although type B MGDs are not essential for the plant under optimal conditions, they play important roles in specific contexts such as phosphate deprivation and at specific developmental stages (Awai et al., 2001; Kobayashi et al., 2004, 2009; Billey et al., 2020). In addition, MGD2 function in pollen tube highlights that these enzymes play a role in the cytosol, a plastid-independent galactolipid pathway which needs to be unravelled.

2.2.3.3. Roles of galactolipids in the architecture of mature thylakoids

As mentioned in section 2.2.1.1, the highly unsaturated acyl chains of native MGDG give it a conical shape leading to its self-organisation in HexII phase. DGDG on the other hand forms a Lm phase due to its larger polar head group. Consequently, the distinct structural properties of galactolipids play an important role in the 3-dimensional architecture of thylakoids (Murphy, 1982; Simidjiev et al., 2000; Demé et al., 2014; Bastien et al., 2016). A remarkable feature of thylakoid membranes is their ability to form flattened cisternae that can pile up in stacked domains. It has been proposed that MGDG takes part in the stabilisation of the tightly concave thylakoid margin resulting from such structure, although no evidence has been provided since then (Murphy, 1982). However, the asymmetrical distribution of galactolipids in the thylakoid membrane, while not disproving, does not support this hypothesis as MGDG were found to be enriched in the stromal leaflet (>50%) whereas DGDG was most abundant in the luminal leaflet (>80%) (Rawyler et al., 1987).

In green algae and higher plants, thylakoids form appressed grana stacks as well as non-appressed stroma lamellae connecting the grana stacks. In the complex plastid of diatoms, thylakoids form loose stacks of mostly three thylakoids with membrane connections between the thylakoid layers (Bedoshvili et al., 2009). Reconstituted thylakoid lipid extracts was shown to self-organise *in vitro* as a regular stack of bilayers (Demé et al., 2014). This mixture can switch from HexII to Lm phases depending on hydration level and lipid composition, particularly the MGDG/DGDG ratio. The role of DGDG was important in membrane stacking *via* hydrogen bonds between polar heads of adjacent bilayers (Demé et al., 2014).

Organisation of thylakoids in large ordered lamellar structures also depends on proteins, in particular on Light-Harvesting Complex of photosystem II (LHCII) in green algae and higher plants. LHCII is responsible for capturing light energy and transfer it to photosystem II (PSII). In diatoms, this role is played by the fucoxanthin-chlorophyll proteins (FCPs) (Bhaya and Grossman, 1993). PSII localises to the appressed grana stacks in green algae and higher plants (Andersson and Anderson, 1980; Vallon et al., 1986). It was shown that the association of the non-bilayer forming lipid MGDG with chlorophyll-a/b light-harvesting antenna protein of LHCII *in vitro* leads to the formation of large ordered lamellar structures (Simidjiev et al., 2000). It is important to notice here that LHCII was able to reorganise the HexII phase of pure MGDG into Lm phase *in vitro*. It was proposed that MGDG may thus be responsible for the packing of thylakoid membrane bilayers as observed in grana (Lee, 2000). DGDG was also shown to bind to LHCII and be required for its stabilisation and macroorganisation (Nußberger et al., 1993; Krumova et al., 2010). Cyanobacteria do not contain stacked thylakoids due to their phycobilisome antenna being

larger than that of LHCII. Interestingly, thylakoids from cyanobacteria depleted from phycobilisome form stacked membranes, further indicating that galactolipid-rich membranes can still form stacked membranes in the absence of LHCII (Dobrikova et al., 2013). Therefore, both MGDG and DGDG have been suggested to lead to the stacking of thylakoid membranes through lipid-protein and lipid-lipid interactions, revealing that their chemical properties might be more relevant than their physical properties in this process.

2.2.3.4. The unresolved thylakoid biogenesis mechanism may involve HexII-forming MGDG

The process of thylakoid biogenesis has not been demonstrated yet. The common hypothesis is that thylakoids membranes are derived from the iEM due to their similar lipid composition (Boudière et al., 2014). Given the absence of the galactolipid synthases in the thylakoid membranes, a lipid transport mechanism must exist from the chloroplast envelope to the thylakoids. Bastien et al summarized in their review the possible mechanism of lipid export to the nascent thylakoids (Bastien et al., 2016). These scenarios include the budding of either spherical or flattened vesicles from the iEM, flattened invaginations, development of iEM-thylakoid bridges, and the recently reviewed generation of membranes *via* a non-vesicular process involving the transition of an HexII intermediate phase to Lm phase (Bastien et al., 2016; Guéguen and Maréchal, 2021).

Arguments taken in favour of a vesicular-based scenario include the observation of iEM invaginations in *A. thaliana* chloroplasts when MGD1 activity is affected (Kobayashi et al., 2007; Botté et al., 2011). These invaginations are interpreted as the budding of the iEM in the course of interrupted vesicle formations. However, as detailed above, this phenomenon could be an artifact resulting from the uncontrolled excess of Lm forming lipids in the iEM, which would better support a mechanism where phase transition is involved.

Vesicle budding and transfer to the thylakoids is thought to be mediated by a protein called Vesicle-Inducing Protein In Plastids1/Inner Membrane-associated protein of 30 kDa (VIPP1/IM30) (Kroll et al., 2001; Mechela et al., 2019). However, there has been so far no proof that VIPP1/IM30 actually guides vesicles to the thylakoid membranes. An ancestral form of IM30 exists in *Gloeobacter*, which lacks any thylakoid. Based on its high affinity for anionic lipids such as PG and SQDG (Hennig et al., 2015; Heidrich et al., 2016), it was proposed that VIPP1/IM30 could rather 'help' guiding anionic lipid during thylakoid biogenesis (Guéguen and Maréchal, 2021). Furthermore, scenarios involving vesicles as well as iEM-thylakoids bridges find little support as such structures are only rarely observed.

The hypothesis of a non-vesicular based scenario has emerged recently that would involve the ability of MGDG to form a HexII phase (Bastien et al., 2016). We recently proposed this mechanism as the possible origin of thylakoid in evolution (Guéguen and Maréchal, 2021). The hypothesis is that the iEM contains biosynthetic platforms allowing the accumulation of HexII forming lipids, most likely MGDG, toward the inside of the plastid. Given that MGD1 has a specificity for its product and is located in a different membrane than DGDs, local MGDG accumulation is highly probable. A 'trigger' would then be added to this system, to promote the transition of a HexII phase into a Lm one. It was shown for PE, another HexII lipid, that transition from one phase to another could be triggered by changes in the level of desaturation of its acyl chains, or by modulations in temperature or

hydration. In the case of MGDG, past studies have shown that HexII → Lm transition also occurred following changes in its desaturation level (Gounaris and Barber, 1983), but also in the presence of Lm forming lipids upon hydration (Demé et al., 2014). The trigger could therefore be a Lm thylakoid lipid such as SQDG. Furthermore, the transition HexII → Lm has a negative enthalpy (Gawrisch et al., 1992), meaning that it does not require external source of energy like ATP or proteins. Within nano- to microseconds, the HexII phase would turn into an unstable intermediary phase (with elongated structures called 'stalks') that subsequently would evolve into piled multilamellar membranes, the thylakoids. Consistently with this mechanism, a *Synechocystis* mutant forming MGlcDG instead of MGDG was viable and contained thylakoids, indicating that MGlcDG or MGDG as HexII lipids could be indistinctly usable for the formation of thylakoids.

Companion proteins of this process would have then evolved in the direction of helping, organizing, guiding, and orchestrating it and lead to the more sophisticated thylakoid structures observed today. VIPP1/IM30 for example could have evolved to guide anionic lipids toward the inside of the plastid based on its initial biophysical properties that makes it bind preferentially anionic lipids (Hennig et al., 2015; Guéguen and Maréchal, 2021). Furthermore, its ability to bind to PG suggest that it could locate to PG clusters close to MGD1, where MGDG accumulation would occur.

The fine tuning of the MGDG/DGDG ratio that regulates the transition from the HexII to Lm phase in lipid bilayers might thus be a key factor in thylakoid biogenesis.

2.2.3.5. Packing of large membrane proteins, including the photosynthetic machinery, by MGDG

Fine-tuning of the MGDG/DGDG ratio could also play a very important role for the function of mature thylakoids. Indeed, a low MGDG/DGDG ratio in plastids is generally associated with impaired photosynthesis (Boudière et al., 2014). *In vitro* and *in vivo* studies revealed that both HexII and Lm phases coexist in thylakoids membranes (Rawyler et al., 1987; Krumova et al., 2008; Demé et al., 2014). More specifically, the outer leaflet enriched in MGDG could be in a metastable state allowing the coexistence of both HexII and Lm phases, while in the inner leaflet the Lm phase would be more strongly promoted due to a lower MGDG/DGDG ratio (Rawyler et al., 1987). Certain hydrophobic proteins could be stabilised by the presence of HexII lipids by mean of inverted micelles (Murphy, 1982). The equilibrium between HexII and Lm phases was therefore suggested to regulate the lipid:protein ratio, and also influence the structural flexibility of the membrane (Garab et al., 2000). Furthermore, it has been proposed that the lateral packing pressure exerted by HexII lipids on membrane proteins influences their functional activity (Yeagle, 1989; Kruijff, 1997; Attard et al., 2000; Latowski et al., 2004).

Interestingly, main photosynthetic protein complexes are asymmetrically oriented between the two leaflets. Both photosystem I and II, the cytochrome b_6f and the ATP synthase contain lumenal and stromal domains playing distinct roles in the transport electron chain (Hankamer et al., 1997; Groth and Strotmann, 1999; Berry et al., 2000; Chitnis, 2001). In plants, most of these proteins are also known to show a strong lateral asymmetry between granal and stromal membranes (Andersson and Anderson, 1980; Anderson and Andersson, 1982; Pribil

et al., 2014). PSII in particular is mostly present in the appressed grana stacks. MGDG is precisely more prominent in the grana (Gounaris et al., 1983), supporting a role for MGDG in the packaging of this large protein complex.

The absence of MGDG results in the complete loss of PSII activity *in vivo* (Kobayashi et al., 2013). Consistently with this results, *in vitro* analysis showed that increasing level of MGDG enhance PSII activity, due to a better energy transfer between LHCII and PSII (Zhou et al., 2009). In addition, MGDG is crucial for ordered oligomerisation of LHCII (Schaller et al., 2011) and dimerization of PSII (Kansy et al., 2014). Stabilisation of LHCII is notably believed to be mediated through steric matching of the conical form of MGDG with the hourglass shape of trimeric LHCII (Seiwert et al., 2017). In diatoms, all thylakoid glycerolipids were found to bind at PSII-FCP dimer (Wang et al., 2020). More specifically, they were distributed at the interfaces between subunits suggesting their role in subunits interactions. In cyanobacteria, MGDG was shown to bind PSI as well, with a possible role in the formation of the reaction centre of PSI complex (Mizusawa and Wada, 2012).

Eventually, MGDG also plays a role in non-photochemical quenching (NPQ) mediated by the photo-protective xanthophyll cycle. In photosynthetic membranes, the xanthophyll cycle is a photo-protective mechanism consisting in the de-epoxidation of xanthophyll cycle pigments upon illumination by de-epoxidases enzymes. In higher plant and green algae, the initial epoxidized pigment is violaxanthin, while in diatoms it is diadinoxanthin (Goss and Jakob, 2010). Solubilisation of these pigments is essential for subsequent de-epoxidation. MGDG and PE are able to solubilise both violaxanthin and diadinoxanthin in an aqueous medium with their ability to form HexII phase, at much lower concentration that Lm-forming lipids (Goss et al., 2005). Furthermore, the HexII structure is essential for xanthophyll de-epoxidase activity and its binding to thylakoid membranes (Jahns et al., 2009). Although it is difficult to know where exactly such MGDG HexII phase would occur inside plastids, it seems functionally indispensable for de-epoxidation activity.

In conclusion, *via* its structure, MGDG plays an important role for the stabilisation of the photosynthetic machinery, its functioning and regulation.

2.2.3.6. Role of MGDG unsaturation level

Tuning of the MGDG/DGDG ratio might involve desaturases. Indeed, when unsaturated, some of the MGDG species cannot be used as substrates for DGDG production. This mechanism 'locking' a portion of MGDG in plastid membranes was demonstrated in plants and green algae (Boudière et al., 2012; Li et al., 2012c; Petroutsos et al., 2014). In the case of *P. tricornutum*, MGDG species always contain a 16 carbons (C16) FA at position *sn*-2 and either a C16 or EPA at position *sn-1* (Abida et al., 2015). MGDG species containing a C16 in *sn-1* are not used for DGDG if *sn-2* is bound to a 16:3 (16:x/16:3) and those with EPA in *sn-1* are not used if *sn*-2 is bound to a 16:4 (20:5/16:4) (Abida et al., 2015). Furthermore, DGDG contains only few amounts of 20:5/16:3 species although it is the main species in MGDG. The desaturation of MGDG in 16:3 and 16:4 could thus be a way to lock part of MGDG species and prevent their utilisation by DGDs in *P. tricornutum*. This mechanism explains why DGDG species are much more saturated than MGDG species.

Such a high degree of unsaturation for a membrane lipid is quite unusual, especially for a lipid closely associated with the photosynthetic machinery where oxidation of unsaturated lipids is favoured by exposure to light and

production of reactive oxygen species. Nevertheless, MGDG and DGDG high desaturation level may be strictly required to provide a fluid environment in which photosynthetic electron transfer can occur (Gounaris and Barber, 1983). Indeed, PSI and PSII are laterally separated with PSI localised to the stromal membranes and PSII to the granal membranes in plants (Pribil et al., 2014), while in the loosely stacked thylakoids of *P. tricornutum* the segregation is between the stromal-facing thylakoid membranes for PSI and the 'core' thylakoid membranes for PSII (Flori et al., 2017). Such a topological separation is needed to avoid energy spillover between the two complexes but requires an efficient long range electron transfer. This function is likely to be assured by plastoquinone, a hydrophobic electron carrier whose ability to diffuse through the membrane probably depends on the fluidity of the lipid matrix (Gounaris and Barber, 1983). Polyunsaturated lipids are precisely more flexible, providing the requested fluidity for quinone lateral navigation inside membranes.

Eventually, temperature fluctuations may lead to unwanted changes in membrane fluidity, resulting in transitions between HexII and Lm phases (Gawrisch et al., 1992; Rappolt et al., 2003). Regulation of membrane fluidity in response to changes in temperature is achieved by the control of lipid desaturation in plants and algae (Bandarra et al., 2008; Chen and Thelen, 2013; Fan and Evans, 2015; Bojko et al., 2017; Breton et al., 2020). An increase in the desaturation level of total FAs compensates for a decrease in temperature. In *P. tricornutum*, a shift of temperature from 25°C to 10°C leads to a 120% increase in PUFA yield (Jiang and Gao, 2004). For photosynthetic membranes, level of unsaturation of the major lipid MGDG must thus be tightly controlled. Consistently, changes in the degree of desaturation of MGDG species were observed under cold-stress response (Johnson and Williams, 1989; Sakamoto et al., 2006; Chen and Thelen, 2013; Breton et al., 2020).

2.2.3.7. Roles of MGDG in non-photosynthetic membranes

Taking all data presented above together, MGDG thus plays several roles in plastids for the correct structure and function of photosynthetic membranes. It is however noteworthy to mention that it is also present in non-green plastids of vascular plants and some diatoms (Anderson et al., 1978; Douce et al., 1987). These include proplastids, etioplasts, amyloplasts, leucoplasts, gerontoplasts, chromoplasts, proteinoplasts and elaioplasts (Douce et al., 1987; Ngernprasirtsiri et al., 1988; Charuvi et al., 2012; Kaundal et al., 2013; Suzuki et al., 2013). One type of non-green plastid however has been reported to not contain MGDG, the apicoplast in the Apicomplexa phylum (Botte et al., 2013). Loss of this class of lipid in Apicomplexa has been proposed to result from the very ancient loss of photosynthetic capacity of these organisms combined with simplification of the lipid metabolism (Botté and Maréchal, 2014). Although the role of MGDG in most non-green plastids has not been investigated, its presence indicates a physiological role that is not directly linked to photosynthesis.

MGDG synthesis has also been detected in the pollen grain and in pollen tubes, and revealed to be important for both pollen tube elongation and guidance (Kobayashi et al., 2004; Billey et al., 2020, 2021). Consistently, DGDG was detected in the plasma membrane of elongating pollen tubes (Botté et al., 2011). This demonstrates a novel role for a non-plastidic galactolipid pathway in the cytosol, possibly relocating DAG from a donor compartment to an acceptor membrane where MGDG synthesis occurs.

Eventually, MGDG also plays an indirect role in extraplastidic membranes during phosphate starvation, at least in plants. Although the amount of MGDG is unchanged, its production is increased in order to feed the synthesis

of DGDG that is then exported to other membranes as a replacement of phospholipids (Hartel et al., 2000), including the tonoplast, mitochondria and plasma membrane (Moellering and Benning, 2011).

To summarize, MGDG and more generally galactolipids play roles that are not solely restricted to building photosynthetic membranes, including possible role in non-plastidic compartments, and in response to some environmental stress conditions.

2.3. Glycerolipid remodelling under stress

2.3.1. From the study of the response to environmental parameters to the utilization of environmental stresses as model conditions to address specific biological processes

In nature as in laboratory conditions, microalgae can be subjected to changing parameters of the environment leading to abiotic (dim or exceeding light; low or high temperature; low or high salinity; nutrient shortage including carbon, nitrogen, phosphate, potassium and trace metals; exposure to toxic pollutants, heavy metals, endocrine disruptors, noxious gases, etc.) or biotic (viruses, pathogens, grazers, etc.) stresses. Microalgae responses to such conditions are being studied to better understand the impact on microalgae metabolism and adaptation strategies.

Such stresses can affect the grow rate and biochemical composition of cells. Biomass production, lipid content, lipid composition, starch content, protein content, carotenoid production, and photosynthesis efficiency are part of the main parameters monitored during these stresses as they influence cultivation strategy for industrial applications (Juneja et al., 2013). A selected stress can therefore become a model condition to better understand the role of a protein or a pathway. For example, the lipid synthesis enzymes MGD2 and MGD3 revealed specific role under phosphate deprivation (**see section 2.2.3.2**).

Some stresses are more studied than other. In particular, the TAG-inducing nitrogen and phosphate starvation conditions are part of the most popular stresses performed in laboratory.

2.3.2. Intense lipid remodelling under nitrogen and phosphate starvation

Stress-induced lipid remodelling corresponds to changes in the proportions of glycerolipid classes, in the FA contained in each class, and sometimes includes modification of subcellular location (Jaussaud et al., 2020).

Nitrogen deprivation in *P. tricornutum* leads to a drastic and rapid response, potentially due to the need of this element for protein synthesis. Interestingly, N-containing lipids such as PC, PE and DGTA are not affected by N deprivation and therefore do not appear to serve as N storage in the cell. Both MGDG and PG decrease and DGDG increase, while other lipid classes are stable (Abida et al., 2015) (**Fig. 2.13**). Such changes in galactolipids leads to a decrease in the MGDG/DGDG ratio at the whole cell level. This ratio being very important in thylakoid membranes, it would be particularly interesting to know in details how each plastid membrane is affected by the lipid variations. In any case, plastid size shrinkage is observed (Longworth et al., 2016). Furthermore, the decrease

of 20:5-containing MGDG species roughly corresponds to the increase of 20:5-containing TAG species, indicating that EPA released by MGDG degradation is reused for TAG synthesis.

Phosphate deprivation is an apparently less severe stress for *P. tricornutum* compared to a lack in nitrogen, possibly due to the presence of intracellular storage forms, nonetheless it also leads to an intense lipid remodelling. In general terms, phosphate deprivation triggers a breakdown of membrane phospholipids (Abida et al., 2015; Cañavate et al., 2017b, 2017a; Mühlroth et al., 2017; Liang et al., 2019a). Such decline points phospholipids as a likely substantial phosphate storage form.

- While in the non-plastidic membranes in plants, a PC-to-DGDG replacement appeared as a general phosphorus-saving process involving export of DGDG outside the plastid to reach the mitochondrion, vacuole or plasma membrane, no such mechanism could be evidenced in any heterokont model studied to date (Abida et al., 2015). Rather, the proportion of another class of non-phosphorus lipid, i.e. betaine lipids, increases (Abida et al., 2015; Cañavate et al., 2017b, 2017a; Mühlroth et al., 2017; Liang et al., 2019a; Meng et al., 2019): it was suggested that phospholipids were replaced by betaine lipids in membranes (Abida et al., 2015; Cañavate et al., 2017a; Mühlroth et al., 2017; Meng et al., 2019), as also suggested in other phytoplankton species (Van Mooy et al., 2009; Cañavate et al., 2017b). This specific PC-to-betaine lipid replacement has been proposed in several studies and is possibly restricted to the endomembrane system (Abida et al., 2015; Iwai et al., 2015; Meng et al., 2019).
- By the same token, in plastid membranes, phosphate deprivation triggers a striking increase in SQDG proportion in *P. tricornutum* (Abida et al., 2015; Cañavate et al., 2017b; Mühlroth et al., 2017). In *P. tricornutum* the decrease in PG is concomitant with this increase of SQDG (Abida et al., 2015; Cañavate et al., 2017b; Mühlroth et al., 2017), reflecting a specific PG-to-SQDG replacement (Abida et al., 2015; Iwai et al., 2015; Cañavate et al., 2017b). This replacement is reminiscent to that established in other organisms, such as *Arabidopsis* under phosphate deprivation (Nakamura, 2013) and has been reported in prokaryotic cyanobacteria and other eukaryotic phytoplankton exposed to a phosphate shortage as well (Van Mooy et al., 2009; Cañavate et al., 2017b).

Therefore, although DGDG increases in phosphate starvation, it was proposed that both DGTA and SQDG increases are enough to compensate for the loss of phospholipids (Abida et al., 2015). The increase of DGDG in phosphate and nitrogen limitations would rather serve to compensate for the loss of MGDG and protect photosynthetic membranes.

Degraded phospholipids can provide building blocks for TAG accumulation. FAs released from membrane glycerolipids may be consumed via the β-oxidation pathway in the mitochondrion. Alternatively, down-products can be recycled to form TAG. In particular, PC is supposed to be involved in acyl-CoA independent synthesis of TAG as inferred by the activation of the Land's cycle and the PDAT enzymes upon phosphate depletion (Mühlroth et al., 2017). Hypothetical flows of materials released from membrane lipid breakdown to TAG are summarized in **Fig. 2.13**.

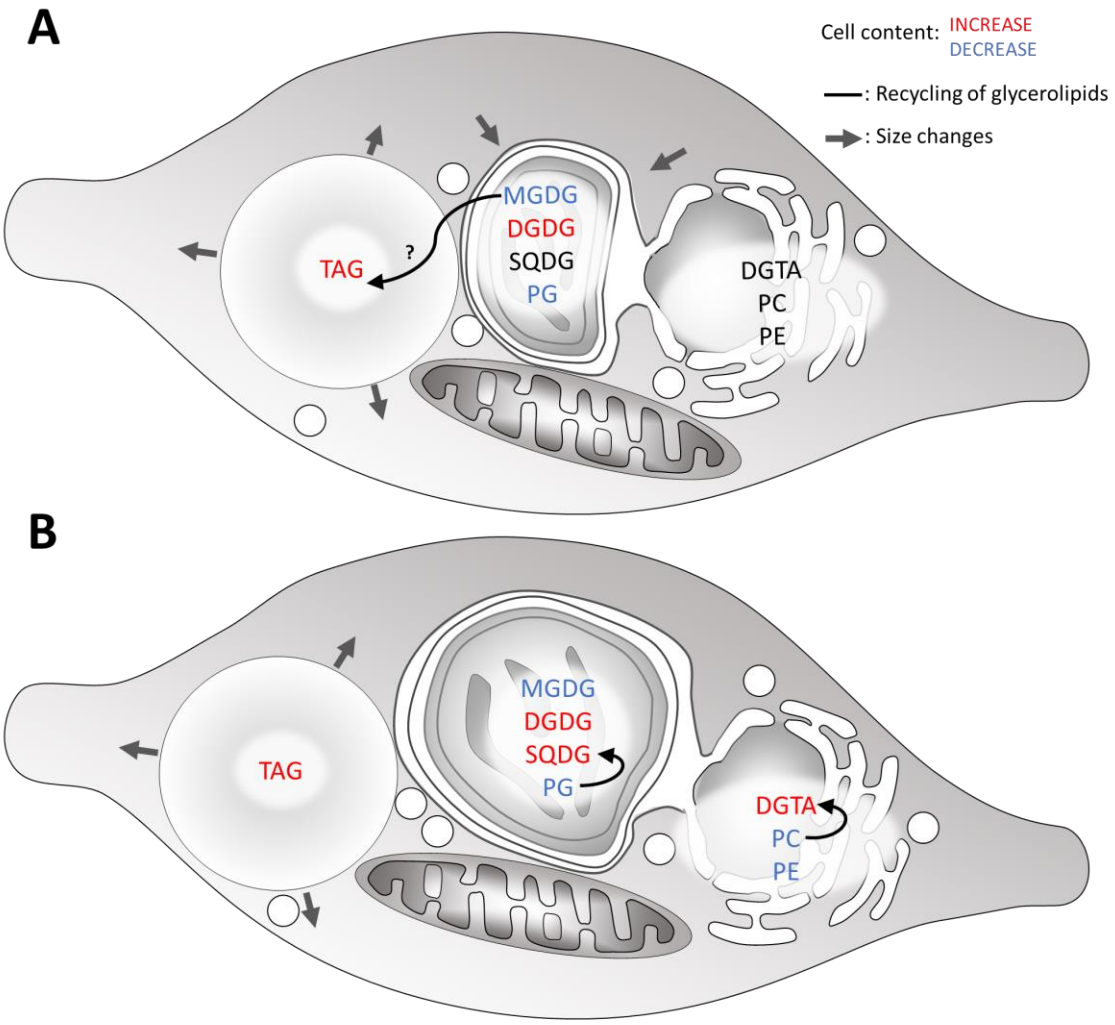


Figure 2.13: Hypothetical recycling of products of membrane glycerolipid breakdown for the formation of TAG in *P. tricornutum* subjected to a nitrogen or phosphate shortage. A. Lipid remodelling under nitrogen starvation. Most plastid glycerolipids are degraded, affecting the plastid size. B. Lipid remodelling under phosphate starvation. In both scenarios, fatty acids released from membrane lipids may be consumed via the β-oxidation pathway in the mitochondrion. Alternatively, down-products can be recycled to form TAG. Arrows show hypothetical conversions of membrane lipid down-products in the production of TAG, as well as in the increase in very-long-chained PUFA (e.g. 20:5) in LDs. Replacement by non-phosphorus lipids rescues the deleterious impact following phospholipids degradation. DGDG, digalactosyldiacylglycérol; EPA, eicosapentaenoic acid; MGDG, monogalactosyldiacylglycerol; PC, phosphatidylcholine; PG, phosphatidylglycerol; SQDG, sulfoquinovosyldiacylglycerol; TAG, triacylglycerol. Figure adapted from (Guéguen et al., 2021).

Thesis objective

At the beginning of my thesis, while most genes involved in membrane lipid biogenesis had been identified in diatoms *in silico* (Dolch et al., 2017b; Conte et al., 2018; Zulu et al., 2018), none were characterized. Biosynthesis of secondary plastids thus remained largely unknown. MGDG being the most important lipid in photosynthetic membranes, we decided to focus our attention on the characterisation of MGDs in *P. tricornutum*.

In *P. tricornutum*, three putative MGDs have been identified. Are all of these putative MGDs actually functional, and why would a cell be equipped with as many isoforms? Where are they localised, and do they contribute to the synthesis of distinct membranes? Do one of these MGDs act as the main enzyme under optimal condition while other come in support under stress conditions, as observed in Arabidopsis? Do they have distinct substrate specificity?

My thesis focuses on providing elements of response to all these questions. To this end, the main strategy consisted in the generation of knockout mutants of each putative MGDs in *P. tricornutum* followed by the analysis of their lipid profile, growth rate, membrane integrity and photosynthetic activity. Localisation of each enzyme in *P. tricornutum* was achieved using overexpression of GFP-fused MGDs. We also sought to demonstrate their enzyme activity using purified recombinant protein first *in vitro* and then *in vivo* in yeast. To further characterise these enzymes, lipid profile in the mutants was monitored under nutrient deprivation conditions.

3. Materials and Methods

3.1. Model organisms and culture conditions

3.1.1. P. tricornutum strains

P. tricornutum is a pennate diatom used as a study model for the following reasons: it can be cultivated in laboratory conditions in the absence of silica, its genome has been sequenced (Bowler et al., 2008), tools for genetic engineering (overexpression, editing, RNA silencing) are available (Daboussi et al., 2014; Nymark et al., 2016; Daboussi, 2017; Mann et al., 2017; Stukenberg et al., 2018), and its lipidome has been characterized (Abida et al., 2015).

P. tricornutum WT (Bohlin, 1896), ecotype Pt_1 (strain 8.6, CCAP 1055/3) was obtained from Pravosali-Guillard National Center for Marine Algae and Microbiota (NCMA).

MGD overexpressing *P. tricornutum* strains (MGDα-eGFP-A, MGDα-eGFP-B, MGDβ-eGFP-A, MGDβ-eGFP-B, MGDγ-eGFP-A, and MGDγ-eGFP-B) were kindly provided by Hanhua Hu (Laboratory of Algal Biology, Institute of Hydrobiology, Chinese Academy of Sciences, Wuhan 430072, China) and Yangmin Gong (Oil Crops Research Institute of Chinese Academy of Agricultural Sciences, Wuhan 430062, China). Overexpressing vectors were constructed as follows: full-length coding sequence of *MGDα*, *MGDβ* and *MGDγ* genes were amplified by PCR using cDNA derived from *P. tricornutum* as a template and MGDα-Fwd/Rev, MGDβ-Fwd/Rev and MGDγ-Fwd/Rev primers respectively (**Table 3.1**). The PCR products were cloned into pEASY-T1 simple (TransGen, Beijing) vector for DNA sequencing. Sequence-confirmed *MGDα*, *MGDβ* and *MGDγ* genes were excised from T-cloning vector with BamHI and XhoI in the case of *MGDα*, BamHI and SalI for *MGDβ*, and NheI and SalI for *MGDγ*. Each gene was inserted upstream a sequence coding for eGFP used as reporter gene in the multi cloning site of pPha-CG vector (**Fig. 3.1**) containing the restriction enzyme sites previously mentioned. pPha-CG vector was engineered by Yangmin Gong and results from the insertion of the eGFP inside the multiple cloning sites of the regular transformation vector pPha-T1 (GenBank AF219942) (Zaslavskaia et al., 2001). The pPha-CG vectors harbouring *MGDα-eGFP*, *MGDβ-eGFP* and *MGDγ-eGFP* were linearized with ScaI and introduced in *P. tricornutum* cells by electroporation as previously described (Zhang and Hu, 2014).

Primers for constructions in pPha-CG vector (Section 3.1.1)

Name	Sequence	Restriction enzyme
MGDα-Fwd	GGATCCATGGACAGACCAAAGGCGTC	BamHI
MGDα-Rev	CTCGAGTGATGTCCGTGTCAATTTAGC	XhoI
MGDβ-Fwd	GGATCCATGGTGTGGTCAAGGTCACC	BamHI
MGDβ-Rev	GTCGACCTTATTGTTGAGAAGCATCC	SalI
MGDγ-Fwd	GCTAGCATGGCTACAGGTTTCTCAACC	NheI
MGDγ-Rev	GTCGACCCGACTTGGTTCTTCGGCG	SalI

Primers for constructions in PIVEX 2.3d vector (Section 3.4.3.1)

Name	Sequence	Restriction enzyme
MGDα-Fwd-NcoI	TACATACCATGGGTTTGCGACCCACGTCA	NcoI
MGDα-Rev-SmaI	TTCAGACCCGGGTGATGTCCGTGTCAATTTAG	SmaI
MGDβ-Fwd-NdeI	GGAATTCCATATGAACGAGGCCGTTGCGAG	NdeI
MGDβ-Rev-SmaI	TTCAGACCCGGGCTTATTGTTGAGAAGCATCC	SmaI
MGDγ-Fwd-NdeI	GGAATTCCATATGGCTGGTTTAGAAATGGCG	NdeI
MGDγ-Rev-SacI	TTCAGAGAGCTCTCCGACTTGGTTCTTCGG	SacI

Primers for Quantitative Real Time PCR (Section 3.2.4)

Name	Sequence	Name	Sequence
168-qPCR1-Fwd	TCCTCATTCACTATCCGGGC	619-qPCR5-Fwd	CAGGACTCGGTTTCGTCACA
168-qPCR1-Rev	CTGCGGATGTGCTGATAGGT	619-qPCR5-Rev	AGCATGACGGGAAGAGACAC
168-qPCR2-Fwd	ACCTATCAGCACATCCGCAG	125-qPCR1-Fwd	GGCTTCGTCGAGAACATGGA
168-qPCR2-Rev	GCCCTAATTTTGGCTTCGCA	125-qPCR1-Rev	GCCGGGGAGATACGAAAAGA
168-qPCR3-Fwd	CTGTCGTGACGGATTTGGGA	125-qPCR1bis-Rev	GGCCGGGGAGATACGAAAAG
168-qPCR3-Rev	TCGTTTGGTAAGGCGTCGAA	125-qPCR2-Fwd	TCGTCGAAGAGGCTGGTTTT
168-qPCR4-Fwd	CGGGCTTTGGGGACTACAAT	125-qPCR2-Rev	CTAGCGCAGCATTTCGCATT
168-qPCR4-Rev	CCTTGTGCGCTTTTGCTCAT	125-qPCR3-Fwd	TCTTTTCGTATCTCCCCGGC
168-qPCR5-Fwd	GCGCATGGCTGAATACATGG	125-qPCR3-Rev	CAGCCACGAGCTCACAGTAT
168-qPCR5-Rev	AGTCAACGTTGCCAGCTTCT	125-qPCR4-Fwd	ACGGATATTGGCGGAACTGG
619-qPCR1-Fwd	AGCAGCTACCGCTCCTACA	125-qPCR4-Rev	CACTCCCGAGATCAGTGACG
619-qPCR1bis-Fwd	TGGATGAGCCTAATGTATCGGC	125-qPCR5-Fwd	CGGCAGCGTTCTCACCTAAT
619-qPCR1-Rev	GGTGTCGGACGAGAGGAAAA	125-qPCR5-Rev	TACTGCCGCTATCGAAACCG
619-qPCR2-Fwd	GGCGGCATCAAGTATACCGAA	TUBA-Fwd	CTGGGAGCTTTACTGCTTGGA
619-qPCR2-Rev	AAAGCTGGAACTGATTGGCG	TUBA-Rev	ATGGCTCGAGATCGACGTAAA
619-qPCR3-Fwd	ATTCAAGCCGACCTTCTCGG	RPS-Fwd	AATTCCTCGAAGTCAACCAGG
619-qPCR3-Rev	TCGCCACCTCCCATTACCA	RPS-Rev	GTGCAAGAGACCGGACATAC
619-qPCR4-Fwd	GTCTCTTCCCGTCATGCTTACA	HPRT-Fwd	AGCTTGGAAGTGTGTACTCTC
619-qPCR4-Rev	CCACATGCAAACTTCCTCGC	HPRT-Rev	TGAAGGTTACTCCTGTCGAAG

Primers for transformant screening (Section 3.2.7.2)

Name	Sequence	Name	Sequence
125-Fwd-138	GTGCCGAGGACGAATATGGT	168-Rev+359	GTGAGTACAGCACGAAAAGGT
125-Rev+210	AACGAGTGTACAGGCGACAG	619-Fwd-132	GCTCTCGTCGATGTATTGCC
125-Rev+1519	TCTAGGGCCTCTGAACGGTG	619-Rev+267	ACGAAACAAATAAAACCCGGAG
125-Fwd+205	TCGTTGTCATGTGCAAACTGT	619-Fwd+244	CCTCCGGGTTTTATTTGTTTCG
125-Rev+591	GGTTGCGCCGAAATGATAAA	619-Rev+713	AAACGCCCGGGAATTCGA
168-Fwd-131	CTACCCACTTCGGCATTGAT		

Table 3.1: Primers list. Primers used for the cloning of MGD genes in expression vectors, qPCR analysis of MGD expression, and PCR amplification and sequencing for transformant screening. Enzyme restriction sites are underlined when present.

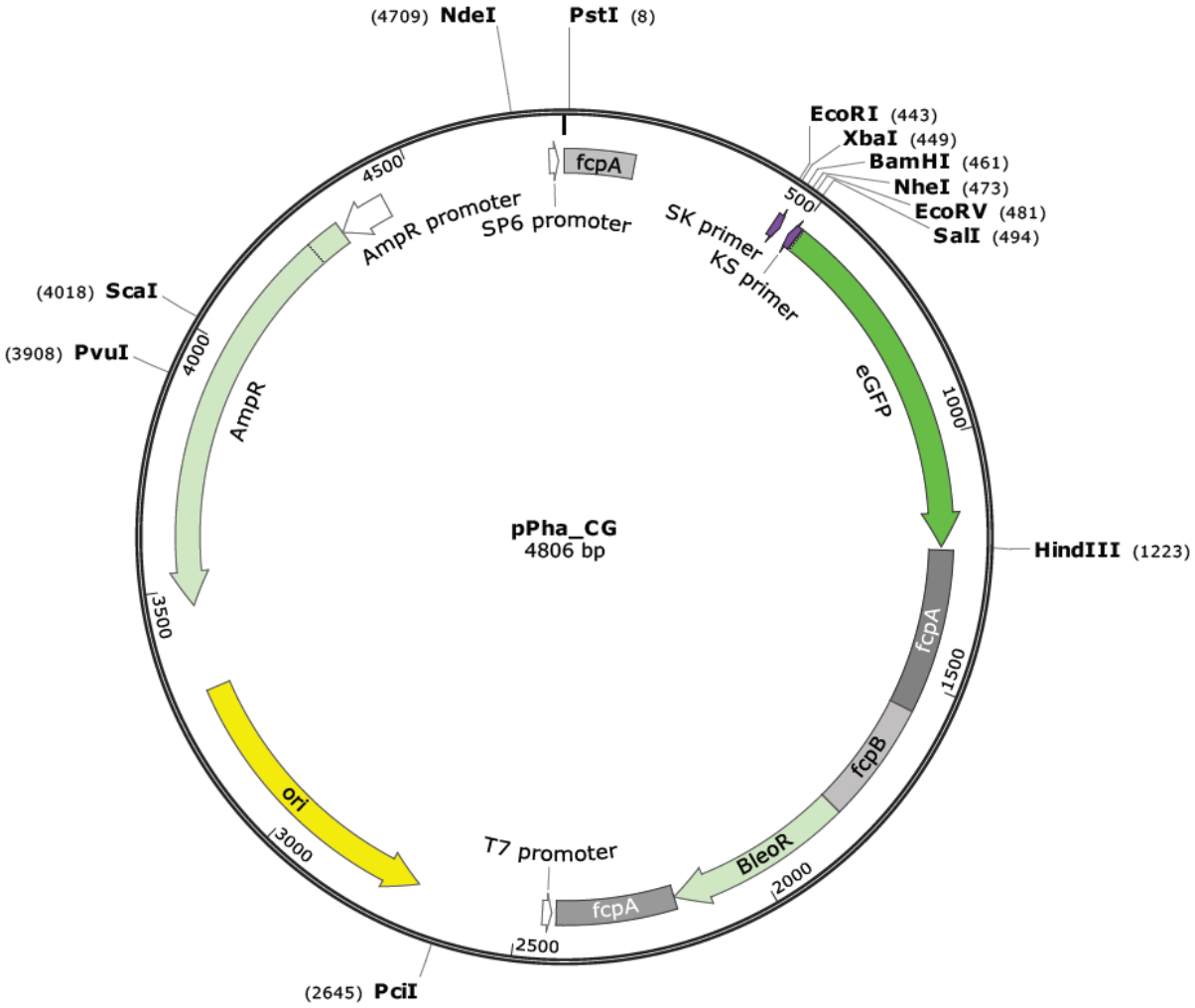


Figure 3.1: Map of the pPha_CG vector. Vector map was constructed with SnapGene software (from Insightful Science; available at snapgene.com) with functional sites and restriction enzyme sites. pPha_CG vector was used for the overexpression of MGD protein fused to eGFP in *P. tricornutum*. Abbreviation: AmpR, Ampicillin Resistance; BleoR, Zeocin Resistance; eGFP: enhanced Green Fluorescence Protein; fcpA, fucoxanthin chlorophyll a/c binding protein A promoter; fcpB, fucoxanthin chlorophyll a/c binding protein B promoter; ori, origin of replication.

3.1.2. Escherichia coli DH5α

Competent cells of *Escherichia coli* strain DH5α were obtained from Invitrogen (Invitrogen MAX Efficiency DH5α Competent Cells, catalogue number: 18258012). Genotype: F- Φ80lacZΔM15 Δ(lacZYA- argF) U169 recA1 endA1 hsdR17 (rk-, mark+) PhoA supE44 λ-thi-1 gyrA96 relA1. *E. coli* DH5α cells were stored in aliquots of 100 µL in 1.5-mL tubes at - 80 °C until use.

Bacteria were cultured in Lysogeny Broth (LB) medium, on solid medium (Agar 1% m/v) at 37°C or in 500-mL flasks containing 100 mL of liquid medium at 37°C and under gentle agitation (200 rpm). After transformation, bacteria were cultured on solid media in presence of Carbenicillin disodium salt (0.237 µM) for selection: only transformed bacteria containing the resistance cassette (AmpR) form colonies.

3.1.3. *P. tricornutum* growth conditions

3.1.3.1. ESAW medium

ESAW (Enriched Seawater, Artificial Water) medium is composed as follow: NaCl (326.7 mM), Na_2SO_4 (25 mM), KCl (8.03 mM), KBr (0.725), H_3BO_3 (0.372 mM), NaF (0.0657 mM), $MgCl_2$-$6H_2O$ (47.18 mM), $CaCl_2$-$2H_2O$ (9.134 mM), $SrCl_2$-$6H_2O$ (0.082 mM), Fe-EDTA (8.17 µM), Na_2EDTA-$2H_2O$ (8.3 µM), $ZnSO_4$-$7H_2O$ (0.254 µM), $CoCl_2$-$6H_2O$ (0.0672 µM), $MnCl_2$-$4H_2O$ (2.73 µM), Na_2MoO_4-$2H_2O$ (6.12 nM), Na_2SeO_3 (1 nM), $NiCl_2$-$6H_2O$ (6.27 nM), $CuSO_4$-$5H_2O$ (0.039 nM), vitamin B8 (4.09 nM), vitamin B12 (0.738 nM), vitamin B1 (0.593 µM), $NaNO_3$ (0.549 mM), and NaH_2PO_4-H_2O (0.022 mM). ESAW medium is prepared under sterile conditions from stock solutions sterilized either by autoclaving the solutions or by filtration through a 0.22-µm membrane filter.

Under optimal growth conditions, all cultures were conducted using enriched ESAW medium containing 10 times more $NaNO_3$ and NaH_2PO_4, corresponding to a concentration of 5.49 mM and 0.22 mM respectively. This medium was designated as ESAW 10N10P (Abida et al., 2015). For nitrogen deprivation experiments, ESAW 0N10P was used.

3.1.3.2. Culture on solid media

P. tricornutum cells were cultivated in an incubator (MLR-352-PE Climate Chamber) at 20°C under continuous light (50 µmol.m^{-2}.s^{-1}). Cultures of *P. tricornutum* cells on solid medium were performed with ESAW 10N10P complemented with Agar (1 % m/v). To guarantee axenic conditions, all media were supplemented with Carbenicillin disodium salt (0.237 µM) to which *P. tricornutum* is naturally resistant. For culture of mutant lines transformed with CRISPR-Cas9 vectors, Zeocin (Invitrogen; 0.07 µM) was added to the media. For overexpressing lines designed in our laboratory, the media were supplemented with Blasticidin (0.028 µM).

3.1.3.3. Culture in liquid media

Cells were transferred from solid to liquid media first in 25-mL, then in 50-mL, 100-mL, and finally 250-mL flasks containing 5 mL, 10 mL, 20 mL, and 50 mL of ESAW 10N10P media respectively. Flasks were placed in an incubator (INFORS HT Multitron Pro) at 20°C under gentle agitation (100 rpm), with a light cycle of 12h/12h (40 µmol.m^{-2}.s^{-1}). Cell concentration was kept between $0.5.10^6$ and 10.10^6 cell.mL^{-1}.

For nitrogen starvation, *P. tricornutum* cells were cultivated in triplicates in 50 mL of ESAW 10N10P in 250-mL flasks until a concentration of 3-4.10^6 cell.mL^{-1}. Cultures were transferred to 50-mL polycarbonate tubes and centrifuged at 1,500 g for 10 minutes. Supernatant was removed. Cells were washed with 50 mL of ESAW 0N10P medium, and centrifuged at 2,000 g for 10 minutes. Supernatant was removed. Cells from each triplicate were then transferred to 250-mL flaks with 50 mL of ESAW 0N10P at a concentration of 2.5-3.10^6 cell.mL^{-1}.

3.1.4. Cryopreservation

3.1.5. Harvest

Cultures were performed in triplicate in 250-mL flasks. Harvest was achieved when cells reached a cell density of 4-5.10^6 cell.mL^{-1}. For glycerolipid and RNA extractions, protein extraction, and DNA extraction, a volume containing 150.10^6, 200.10^6, and 100.10^6 cells was harvested respectively. Liquid cultures were centrifuged in 50-

mL polycarbonate tubes at 3,500 g for 10 minutes. Supernatant was either completely removed and tubes placed in liquid nitrogen for fast-freezing (protein extraction and RNA extraction experiments), or partially removed to be able to resuspend the cells, transfer them to 1.7 mL SafeSeal Microcentrifuge Tubes (27210, Sorenson) and centrifuge them at 100 g for 2 minutes. Supernatant was then removed before fast-freezing of the samples in liquid nitrogen (lipid extraction experiment). Cell pellets were stored at - 80°C.

3.2. Molecular biology

3.2.1. DNA extraction and purification

For extraction of genomic DNA from *P. tricornutum*, a cell pellet corresponding to 100.10^6 cells was resuspended in 500 µL of lysis buffer (250 mM Tris pH 8.2, 100 mM EDTA, 2% SDS, 350 mM NaCl) and transferred to a 1.5-mL tube. The sample was incubated 15 second at 60°C under agitation (450 rpm) using a thermobloc. 500 µL of cold phenol:chloroform:isoamyl alcohol (25:24:1) was added to the sample and homogenized gently by inverting the tube. The sample was centrifuged at 13,000 g at 4°C for 10 minutes to remove debris and form two clearly separated phases (an upper aqueous phase containing the DNA and a lower organic phase). About 300 µL of the upper aqueous phase was transferred to a new 1.5-mL tube. A second extraction was performed by addition of one volume of chloroform:isoamyl alcohol (24:1). The sample was gently mixed before centrifugation at 13,000 g at 4°C for 10 minutes. About 200 µL of the upper aqueous phase was transferred to a final 1.5-mL tube. For DNA precipitation, 30 µL of Na-Acetate 3 M pH 5 and three volumes of absolute ethanol were added, and the sample incubated at - 20°C for 20 minutes. The sample was centrifuged for 10 minutes at 13,000 g at 4°C to pellet the precipitated DNA and remove the supernatant. DNA pellet was washed with 700 µL of ethanol 70 % and centrifuged at 13,000 g at 4°C for 5 minutes. The supernatant was carefully removed, then the sample was left at room temperature for one hour to air dry the DNA pellet. DNA was resuspended in DNAse-free water. Concentration and purity of the DNA was evaluated using a NanoDrop2000 (ThermoFisher Scientific). Genomic DNA was considered pure for an A_{260}/A_{280} ratio of 1.8 and a A_{260}/A_{230} ratio between 1.8 and 2.2.

Plasmid extraction from *E. coli* was performed using the NucleoSpin Plasmid kit (MACHEREY-NAGEL) for 2-mL bacteria cultures and the NucleoBond Xtra Midi kit (MACHEREY-NAGEL) for 100-mL bacteria cultures according to manufacturer instructions.

DNA purification following PCR, enzyme digestion, or after gel migration was performed using the NucleoSpin Gel and PCR Clean-up kit (740609.250, MACHEREY-NAGEL).

3.2.2. RNA extraction and purification

For RNA extraction from *P. tricornutum* cells, 1.5 mL of TRI Reagent (mixture of guanidine thiocyanate and phenol in a monophasic solution; undisclosed proprietary composition, Sigma-Aldrich) was directly added to a frozen cell pellet corresponding to 150.10^6 cells (see section 3.1.5). The pellet was then resuspended and transferred to a 2-mL tube. The sample was first vigorously mixed using a vortex for 30 seconds and subsequently incubated 5 minutes at 60°C. This process was repeated twice for a thorough cell lysis. 300 µL of chloroform were added and

the mix was vigorously mixed. The sample was incubated at room temperature for 15 minutes before centrifugation at 11,000 g at 4°C for 15 minutes for phase separation. About 600 µL of the upper aqueous phase containing the RNA was transferred to a Phasemaker tube (ThermoFisher Scientific). A volume of 1-bromo-3-chloropropane (BCP) corresponding to one fifth of the transferred aqueous phase was added to the tube and mixed vigorously. The sampled was incubated 3 minutes at room temperature before centrifugation at 16,000 g at 4°C for 5 minutes. The Phasemaker tube contains a thick liquid polymer that separates and completely isolates the upper aqueous and lower organic phases, thus ensuring the absence of contamination in the upper phase containing the RNA. After centrifugation, 1 mL of the aqueous phase was transferred to a 2-mL tube. RNA was precipitated by the addition of one volume of isopropanol and gentle mixing. The sample was incubated for 30 minutes at room temperature. RNA was pelleted by centrifugation at 11,000 g at 4°C for 10 minutes and supernatant was removed. The RNA pellet was washed by the addition of 1 mL of ethanol 75 %, centrifugation at 7,500 g at 4 °C for 10 minutes, and removal of the supernatant. The sample was left to air dry for 30 minutes at room temperature. RNA was resuspended in 35 µL of RNA-free water.

In order to prevent RNA contamination with DNA, a first DNAse treatment was applied using the Invitrogen Ambion TURBO DNA-*free* kit following manufacturer's instructions. 0.1 volume of 10X turbo DNAse buffer and 1 µL of DNAse were added to the sample, then mixed gently. Sample was incubated at 37°C for 20 minutes in a thermobloc. 0.1 volume of DNAse inactivation reagent mix was added and mixed gently. Sample was incubated at room temperature for 5 minutes with regular mixing. Finally, the sample was centrifuged at 10,000 g for 1.5 minutes and supernatant transferred to a new tube.

Cleaning and purification of the sample was then achieved using the RNeasy MinElute Cleanup kit (Quiagen) following manufacturer's instructions. The volume was adjusted to 100 µL using RNAse DNAse-free water. 350 µL of lysis buffer (undisclosed proprietary composition) and 250 µL of absolute ethanol were added and mixed by pipetting. The sample was loaded on an RNeasy MinElute Spin Column and centrifuged at 10,000 g for 15 seconds. The flow-through was discarded and 500 µL of wash buffer (undisclosed proprietary composition) was added to the column and centrifuged at 10,000 g for 15 seconds. The flow-through was discarded, then 500 µL of ethanol 80% were added to the column and centrifuged at 1,000 g for 2 minutes. The flow-through was discarded and the column was centrifuged again at 14,000 g for 5 minutes to dry the column membrane. The column was then placed into a new collection tube. 14 µL of RNAse DNase-free water was added on the membrane and centrifuged at 14,000 g for 1 minute to eluate RNA.

RNA concentration and quality was assessed using a NanoDrop2000 (ThermoFisher Scientific) and based on a profil visualized after electrophoresis. RNA sample was considered pure for an A_{260}/A_{280} ratio above 1.8 and an A_{260}/A_{230} ratio of 1.8. Migration of *P. tricornutum* RNA on TAE 0.5X (0.5 mM EDTA disodium salt pH 8, 0.02 M Tris, 0.01 M acetic acid) agarose 2 % gel should show 2 to 4 disctinct bands. Genomic DNA and RNA degradation should not be visible.

RNA sample was stored at - 80°C until use.

3.2.3. Reverse transcription

Reverse transcription was achieved using the SuperScript IV VILO Master Mix with ezDNase Enzyme (ThermoFisher Scientific) following manufacturer's instructions. The first part of the method consists of a DNAse treatment. In a 0.2-mL tube, 500 ng of RNA was mixed with 1 µL of 10x ezDNAse buffer (undisclosed proprietary composition), 1 µL of ezDNAse enzyme, and adjusted to a volume of 10 µL with RNAse DNAse-free water. The mix was mixed by gentle agitation and incubated at 37°C for 2 minutes in a thermocycler (Bio-Rad T100 Thermal Cycler). The sample was then put quickly on ice before the reverse transcription part. 4 µL of superscript IV VILO Master Mix and 6 µL of RNAse DNAse-free water were added to the sample and mixed gently. The sample was incubated in a thermocycler at 25 °C for 10 minutes for primer annealing, 50 °C for 10 minutes for reverse transcription and 85 °C for 5 minutes for enzyme inactivation. Finally, the sample was diluted 5 times to correspond to a concentration of about 5 $ng.\mu L^{-1}$ of cDNA.

3.2.4. Quantitative Real Time PCR

3.2.4.1. Primer efficiency analysis

Quantitative polymerase chain reaction (qPCR) primers were designed using Primer-BLAST on NCBI on sequences XM_002181649.1 (*MGDα*), XM_002186319 (*MGDβ*), and XM_002176764.1 (*MGDγ*). qPCR primers were designed so that amplification products were between 90 and 150 nucleotides long, have a self-complementarity score below 7, and a self-3'-complementarity score below 3. Tested primers are listed in table 3.1.

qPCR primers were tested as previously described (Schmittgen and Livak, 2008). Briefly, qPCR primers were tested on *P. tricornutum* WT cDNA, with cDNA dilutions of 1:1, 1:10, 1:100, 1:1,000 and 1:10,000 in triplicates. Negative controls were done with DNAse RNAse-free water instead of cDNA. qPCR parameters are detailed in section 3.2.4.2. Efficiency was calculated as a function of the slope of the obtained cycle thresholds (Ct) and the logarithm base 10 (log_{10}) of corresponding cDNA dilutions. qPCR primer couples with an efficiency between 90 % and 110 % were considered appropriate for amplification.

3.2.4.2. Measure of *MGD* genes expression

To quantify *MGD* expression level in *P. tricornutum* WT and mutant lines, qPCR was performed after reverse transcription of extracted RNA of biological triplicates. One qPCR primer pair was used for each *MGD* gene. Selected pair were localised near the 3' borders of the genes: 125-qPCR2-Fwd/Rev for *MGDα*, 168-qPCR4-Fwd/Rev for *MGDβ*, and 619-qPCR5-Fwd/Rev for *MGDγ* (**Table 3.1**). Housekeeping genes *RPS* (40S Ribosomal Protein), and *HPRT* (hypoxanthine guanine phosphoribosyltransferase) were used as internal controls when analysing *MGD* expression in optimal growth conditions, while *RPS* and *TUBA* (Tubulin A) were used under nitrogen deprivation condition as *HPRT* was unstable under nitrogen starvation unlike *TUBA*.

qPCR were performed in hard-shell green shell/white wells 96-well PCR plates (Bio-Rad). Each reaction was achieved in technical triplicates. Power SYBR Green Master Mix (ThermoFisher Scientific) was used for the reaction: 5 µL of SYBR Green, 0.6 µL of each primer at 10 µM for a final concentration of 0.6 µM, 1.8 µL of RNAse

DNAse-free water, and 2 µL of cDNA at 5 ng.µL^{-1} for a final concentration of 1 ng.µL^{-1} were mixed in each well. qPCR plates were prepared on ice before filming of the plate and incubation in a thermocycler (Bio-Rad CFX Connect Real-Time System) using the Bio-Rad CFX Connect Real-Time System.

Incubation and plate reads (fluorophore analysis) were performed as follows: an amplification at 95 °C for 10 minutes (initial denaturation), followed by 40 cycles at 95 °C for 10 seconds (denaturation), 60 °C for 10 seconds (primer annealing), and 72 °C for 30 seconds (extension). After each cycle, SYBR Green fluorescence in the plate was measured to track DNA amplification in the samples. SYBR Green fluoresces when it binds to double stranded DNA. The cycle at which SYBR Green fluorescence was detected above the fluorescence detection threshold is called cycled threshold (Ct). The amplification was monitored by the melting curve. Melting curves were generated after the amplification cycles by first incubating the plate at 65 °C for 10 seconds, then SYBR Green fluorescence was measured every 0.5 °C between 65 °C and 95 °C. The amplicon in a sample starts denaturating at a certain temperature, known as melting temperature, depending on its length and on its G/C content. As the temperature rises, the SYBR Green fluorescence will decrease when the melting temperature corresponding to the given amplicon is reached. The derivative of the melting curve will show a peak at the melting temperature. Only one peak should be observed. If other peaks were visible for the same amplicon, it would mean that other products were amplified due to DNA contamination or unspecific primers for example.

Ct values inside technical triplicates were checked for reasonable standard deviations (below 0.2), and the Ct mean between technical replicates was then used to calculate relative gene expression values for each gene of interest (GOI) inside biological triplicates for each condition. Relative gene expression values for a GOI were calculated in each condition for each biological replicate as:

$E_R = 2^{-\Delta Ct_R}$, with $\Delta Ct_R = Ct_R - Ct_{\overline{ref}}$.

Where E_R is the relative gene expression calculated for one biological replicate, ΔCt_R is the difference between the Ct of the GOI in a given biological replicate and $Ct_{\overline{ref}}$, and $Ct_{\overline{ref}}$ is the mean of the three Ct of the biological triplicate for the reference gene. For graphical representation, we plotted the mean of the E_R of a GOI calculated in each biological triplicate for each condition. Standard deviation for each biological triplicate for each GOI in each condition was calculated as:

$$SD(\overline{E_R}) = SD(\overline{\Delta Ct_R}) * \left|\frac{\partial \overline{E_R}}{\partial \Delta Ct_R}\right| = SD(\overline{\Delta Ct_R}) * ln2 * 2^{-\overline{\Delta Ct_R}}$$

Where SD is the standard deviation, $\overline{E_R}$ is the mean of the E_R in a biological triplicate, and $\overline{\Delta Ct_R}$ is the mean of the ΔCt_R in a biological triplicate.

In the qPCR analysis comparing *MGD* gene expression in the mutants and the WT, the condition was the genetic background. In the qPCR analysis following *MGD* gene expression in the WT at different day of culture in either enriched medium or nitrogen-deprived medium, the condition was the day of cultivation and the medium.

3.2.5. Constructions for *P. tricornutum* transformation

3.2.5.1. CRISPR-Cas9 constructions

Single guide RNA (sgRNA) design was achieved using PhytoCrispex website (www.phytocrispex.biologie.ens.fr/CRISP-Ex/) with the reference genome of *P. tricornutum*, choosing NGG as Protospacer Adjacent Motif (PAM) sequence, and allowing CRISPR targets to start with any nucleotide. The PAM sequence is a specific sequence of nucleotide recognized by the nuclease. Here the nuclease is the Cas9 enzyme from *Streptococcus pyogenes*, which recognizes NGG as PAM sequence. If the sequence corresponding to the sgRNA in the genome is not directly followed by the PAM sequence, the Cas9 will not attach to its target. sgRNA were selected based on their proximity to START codons and active site residues. sgRNA forward and reverse sequences preceded with the short nucleotide sequences TCGA and AAAC respectively were obtained as oligonucleotides from ThermoFisher (www.thermofisher.com/fr). sgRNA from selected mutant lines are shown in table 3.2.

Target Gene	Name	Strand	Forward and reverse sequences	PAM sequence	Primers used for sequencing
MGDα	125a	+	TCGACAATTCAACAACCGTGCTCA AAACTGAGCACGGTTGTTGAATTG	**T**GG	125-Fwd-138 & 125-Rev+210
	125i	+	TCGACAGATACTCATGTCCGACAC AAACGTGTCGGACATGAGTATCTG	**T**GG	125-Fwd-138 & 125-Rev+1519
	125j	+	TCGACGACACTGGCGGGGGTCACA AAACTGTGACCCCCGCCAGTGTCG	**G**GG	125-Fwd+205 & 125-Rev+591
MGDβ	168b	+	TCGAGTCCCGAGGACGTATATGGT AAACACCATATACGTCCTCGGGAC	**T**GG	168-Fwd-131 & 168-Rev+359
	168c	-	TCGAGAAGTCGAAGTTGCCGATGG AAACCCATCGGCAACTTCGACTTC	**T**GG	168-Fwd-131 & 168-Rev+359
MGDγ	619a	+	TCGAAGGTTTCTCAACCCAGATAG AAACCTATCTGGGTTGAGAAACCT	**A**GG	619-Fwd-132 & 619-Rev+267
	619c	-	TCGAGGCTGCATTGACTTCAACGG AAACCCGTTGAAGTCAATGCAGCC	**A**GG	619-Fwd-132 & 619-Rev+267
	619h	+	TCGACTTTTCCTCTCGTCCGACAC AAACGTGTCGGACGAGAGGAAAAG	**C**GG	619-Fwd+244 & 619-Rev+713
	619i	+	TCGACTCTCGTCCGACACCGGTGG AAACCCACCGGTGTCGGACGAGAG	**A**GG	619-Fwd+244 & 619-Rev+713

Table 3.2: **Single guide RNAs list**. Single guide RNAs (sgRNA) used for the acquisition of *P. tricornutum* mutants with CRISPR-Cas9 technology. Orientation of the sgRNAs is given taking the target gene orientation as a reference. Forward and reverse sequences correspond to the oligonucleotide sequences ordered for sgRNA insert preparation. Short nucleotide sequences added for cloning purposes are underlined in the sequences (see section 3.2.5.1). The first possibly variable nucleotide of the associated PAM sequences is written in bold.

The sgRNA insert was prepared by annealing the corresponding pair of oligonucleotides, creating an insert with cohesive ends on both sides of the sgRNA sequence. 2.5 µL of each oligonucleotide at 100 µM were mixed with 45 µL of annealing buffer (1 mM EDTA, 50 mM NaCl, 10 mM Tris pH 7.6) for a final concentration of 5 µM, then heated at 95°C for 4 minutes, and cooled down at room temperature for 45 minutes for hybridization.

The pKSdiaCas9_sgRNA-zeo vector (**Fig. 3.2**) was kindly provided by Cécile Giustini (LPCV, Grenoble). This vector results from the insertion of the zeocin resistance cassette (BleoR) downstream of the FcpB promoter of the pKS diaCas9_sgRNA vector (Addgene, plasmid #74923). The pKSdiaCas9_sgRNA-zeo vector was digested with BSAI restriction enzyme (5'-GGTCTC(N_1)/(N_5)-3') (New England BioLabs), producing a linearized vector with cohesive

ends complementary to the prepared insert. Following digestion, BSAI enzyme recognition site is removed. 1 µL of prepared sgRNA insert and 100 ng of linearized vector were ligated using T4 DNA ligase (New England BioLabs).

The ligation product was transformed in *E. coli* DH5α (see section 3.2.6) for multiplication of the vector. Transformed bacteria were screened for insertion of the sgRNA in the vector by purifying the vector by a miniprep method and digesting the vector with BSAI restriction enzyme, using an empty pKSdiaCas9_sgRNA-zeo vector as a control. The empty vector still has a BSAI restriction enzyme site, while the vector containing the insert does not have it anymore. Thus, plasmid should not be linear after digestion if insert is present. Digestion with PSUI restriction enzyme was also performed to check digestion efficiency. PSUI enzyme cuts twice in the vector, regardless of the presence of the insert. Constructions validated by digestion were sent for sequencing (Biofidal) to ensure the proper insertion of the sgRNA in the vector.

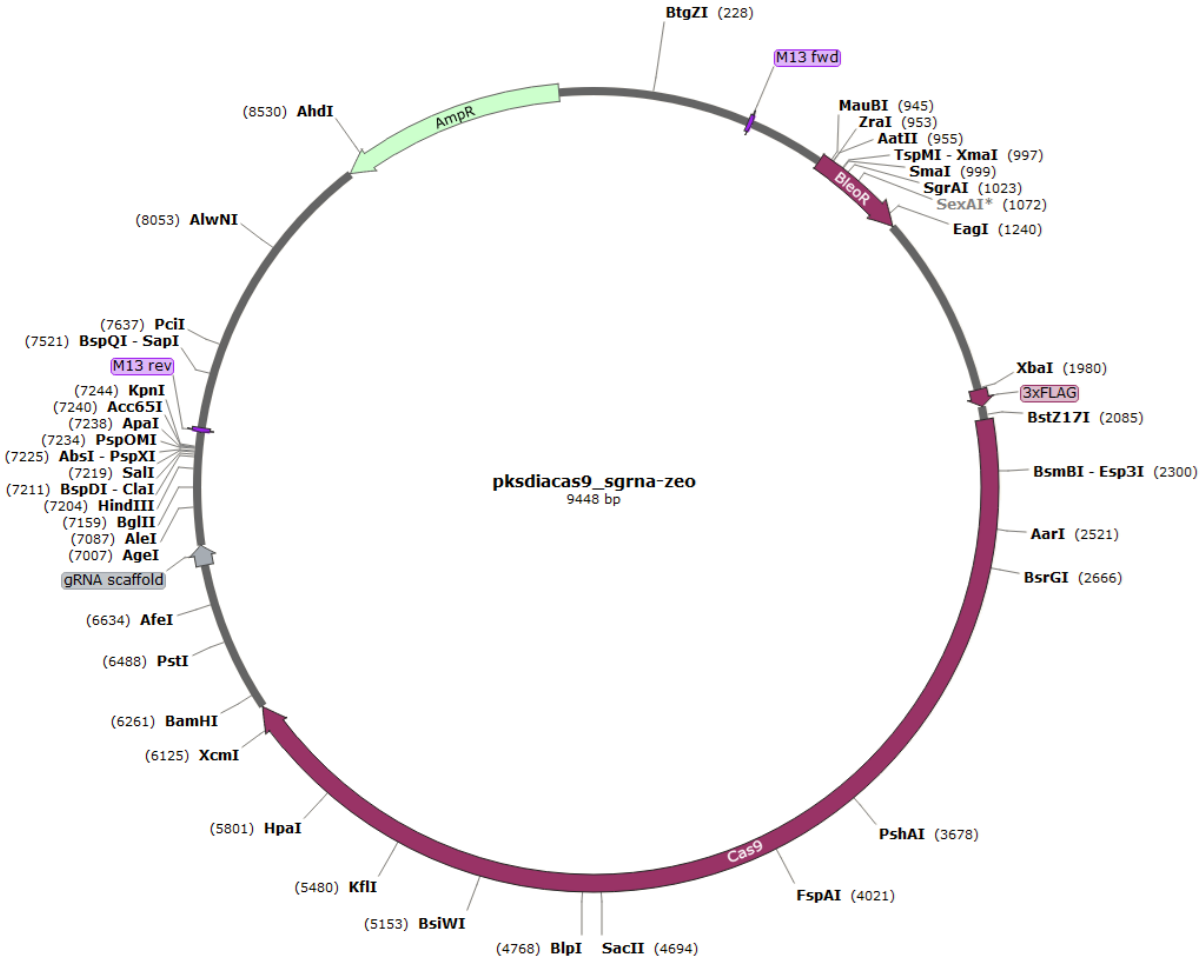


Figure 3.2: Map of the pKSDiaCas9-zeo vector. Vector map was constructed with SnapGene software (from Insightful Science; available at snapgene.com) with functional sites and restriction enzyme sites. PKSDiaCas9-zeo vector was used for the generation of *P. tricornutum* mutants using CRISPR-Cas9 technology. Abbreviation: 3xFLAG, three tandem FLAG-tag; AmpR, Ampicillin Resistance; BleoR, Zeocin Resistance; gRNA scaffold, single guide RNA scaffold.

3.2.5.2. Overexpression and complementation constructions

The overexpression vector pPha-T1_eGFP_Blasti was kindly provided by Cécile Giustini (LPCV, Grenoble) and results from the replacement of the zeocin resistance cassette with the blasticidin resistance cassette (bla) in the pPha-T1_eGFP vector (Gruber et al., 2007).

Each *MGD* gene sequence was optimised using Optimizer website (http://genomes.urv.es/OPTIMIZER/) and a *P. tricornutum* codon usage table kindly provided by Riccardo Aiese Cigliano (Sequentia, Barcelona). The STOP codon was not included in the final sequence as the protein was to be fused with eGFP and His tag. Optimised genes were further modified so that they can be used to complement mutant strains. Indeed, if the CRISPR machinery inside the mutant finds the PAM sequence associated to the target of its sgRNA in the optimised gene, it will mutate it. The first strategy was to modify the PAM sequences associated to the used sgRNAs by modification of at least one guanine without changing the encoded aminoacid. For some PAM sequences, such modification was not possible without modifying the aminoacid, and thus several changes in the corresponding sgRNA sequence were introduced to disturb the binding of the sgRNA, taking into account that the first nucleotides close to the NGG are the most important ones. Each time, the second most used codon was taken as a replacement for sequence modification if possible. EcoRI restriction site was added in 5' position and BamHI followed by BsrgI enzyme restriction sites were added in 3' positions for cloning purposes. Thus, presence of these enzyme restriction sites in the optimised genes was checked and modified if encountered. Each optimised *MGD* gene was synthesized by Invitrogen GeneArt Gene Synthesis (ThermoFisher Scientific) and delivered in the vector pMA-T.

Optimised gene sequences were checked by sequencing (Biofidal) before cloning in the pPha-T1_eGFP_Blasti vector. For constructs overexpressing an optimised MGD fused to eGFP and His tag, digestion was done using EcoRI and BamHI restriction enzymes. For constructs overexpressing an MGD fused to His tag only, EcoRI and BsrgI restriction enzymes were used. After digestion of pMA-T vectors containing the optimised *MGD* genes, 2% agarose gel electrophoresis and DNA purification on gel (see section 3.2.1) were performed to retrieve the inserts. pPha-T1_eGFP_Blasti vector digestion was done using Calf Intestinal Phosphatase (CIP) to prevent closing of the vector without the insert.

3.2.6. Heat-shock transformation of *E. coli*

For bacterial transformation, 2 µL of ligation product were added to a 100-µL aliquot of competent *E. coli* DH5α cells. After 30 minutes of incubation on ice, bacteria were heat-shocked in a water bath at 42°C for 40 seconds. Bacteria were then put back on ice for 2 minutes. To help bacteria recovery after transformation, 500 µL of liquid LB medium were added to the tube, followed by 1 hour of incubation at 37°C, 200 rpm. Cells were then plated on selection medium (**see section 3.1.2**).

3.2.7. Transformation of *P. tricornutum* and mutant selection

3.2.7.1. Biolistic transformation

P. tricornutum cells were maintained in exponential growth for a week (between 0.5 and 2.10^6 cell.mL^{-1}). Then 100.10^6 cells were collected by gentle centrifugation at 1,000 g for 10 minutes at room temperature, resuspended in 500 µL of ESAW 10N10P, and spread on a solid media with Carbenicillin. The plate was dried under a sterile hood for 30 minutes, then kept for 24 hours in an incubator (see section 3.1.3.2) before transformation. Vectors to transform were collected from overnight cultures of *E. coli* 100 mL containing the vector with NucleoBond Xtra Midi kit (MACHEREY-NAGEL) to obtain a concentration of 1 µg.µL^{-1}.

Vector preparation before transformation consists in the coating of the vector on microparticles. For each plate to transform, one 1.5-mL tube containing a 50-µL aliquot of 3 mg of tungsten beads M17 (Bio-Rad, Hercules, USA) in glycerol 50 % (v/v) was used. 4 to 5 µg of plasmid, 50 µL of $CaCl_2$ 2.5 M in absolute ethanol and 20 µL of spermidine 0.1 M (Sigma-Aldrich) were added to the tungsten beads. The mix was vigorously mixed using a vortex for 3 min at high speed then incubated at room temperature for 1 minute. Beads were pelleted by a spun down and the supernatant removed. 140 µL of cold ethanol 70 % were applied on the pellet without resuspending the beads, then removed. Next, 140 µL of cold absolute ethanol were applied on the pellet without resuspending the beads and removed. Finally, 50 µL of cold absolute ethanol were added on the pellet and the beads gently resuspended. Coated beads were kept agitated using a vortex at low speed until use.

Biolistic transformation was performed under a sterile laminar flow hood using a Bio-Rad Biolistic PDS-1000/He Particle Delivery System (Bio-Rad) according to manufacturer recommendations (Bio-Rad PDS-1000/He System User Manual). 1550 psi (per square inch) rupture discs were used. All materials for biolistic were sterilized with ethanol 70 % beforehand. The plate containing the cells to be transformed was placed at level 2 and 3 of the particle delivery system. For tungsten beads coated with CRISPR-Cas9 constructs, each prepared aliquot was spread on four macrocarriers, and all four macrocarriers were used on a given plate. In other words, tungsten beads from a single aliquot were delivered to a plate through four distinct bombardments. For tungsten beads coated with constructs for overexpression, the aliquot was entirely spread on two macrocarriers only. Rupture disk and macrocarrier were assembled inside the bombardment chamber according to manufacturer recommendations. Briefly, a partial vacuum pressure of 25 inches of mercury was created in the bombardment chamber of the biolistic system containing the cells to transform. Then, helium gas was sent to a gas acceleration tube, sealed by the rupture disk, above the bombardment chamber. Pressure inside the gas acceleration tube was monitored with a gauge. When the pressure inside the tube reached 1550 psi, the rupture disk burst, creating a helium shock wave inside the bombardment chamber. This causes the macrocarrier coated with the tungsten beads to be propelled downward at high velocity. The macrocarrier was halted after a short distance by a stopping grid, allowing only the tungsten beads to continue their trajectory through the cells to transform on the plate. The penetration of the tungsten beads, coated with the plasmid to transform, in the microalgae cells allowed the potential integration of the plasmid inside the microalga genome.

3.2.7.2. Transformants' selection, screening and purification method

Three days after transformation, transformed cells were transferred to a new ESAW solid medium with Carbenicillin and either Zeocin or Blasticidin for CRISPR-Cas9 vector and overexpression vector, respectively. Only cells having successfully integrated the plasmid with the resistance cassette form colonies. Colonies appeared after three to six weeks.

In the case of CRISPR-Cas9 vector transformation, colony PCR (Phire Plant Direct PCR Master Mix, F160, ThermoFisher Scientific) were conducted to analyse the mutation profile by sequencing (Biofidal). Primers were designed for each sgRNA to amplify a region of about 500 nucleotides surrounding the cutting site of the Cas9 protein (**Table 3.1**). In some rare cases, no amplification were obtained and primers amplifying a larger region were used to reveal a possible very large deletion. Three different kinds of mutation profiles were usually obtained: Either no edition happened (pure WT colony), several different editions (mosaic colony), or one edition only (pure mutant colony). For the analysis of mosaic colonies, TIDE (Tracking of Indels by Decomposition v3.3.0) and ICE (Inference of CRISPR Edits v2) online software were used for the analysis of the mutation profile composition. If a mosaic colony presented interesting insertions/deletions (INDELs), isolation of pure colonies was attempted. The equivalent of 200 cells from a mosaic colony were spread on a plate with Carbenicillin and Zeocin. Three to six weeks later, the same steps of PCR and sequencing were conducted. If pure mutant colonies were obtained, the screening process was stopped. Otherwise, another isolation attempt was performed.

3.3. Physiological analysis of *P. tricornutum*

3.3.1. Cell concentration determination and growth curve

Cell concentration was measured using a TECAN Infinite M1000 PRO as previously described (Conte et al., 2018). 300 µL per liquid culture were loaded on a 96-well flat bottom transparent plate (ThermoFisher Scientific). Absorbance was read at 730 nm and converted into cell concentration with the following linear regression:

$C = \frac{A_{730} - 0.03758}{0.01834}$ (Million cells per millilitre), where C is the concentration and A_{730} the absorbance.

For growth curve monitoring, cell cultures were conducted in triplicates in 100-mL flasks with 30 mL of liquid medium. On the first day, cells were diluted to a concentration of 1.10^6 cell.mL^{-1}. Cell concentration was measured every day at the same time for ten days.

3.3.2. Lipidomic analyses

3.3.2.1. Glycerolipid extraction

Glycerolipids were extracted according to the Folch method (Folch et al., 1957). Frozen cell pellets (**see section 3.1.5**) were lyophilised (CHRIST Alpha 2-4 LSCbasic) for a night before extraction. Lyophilised pellets were transferred to Corex glass tubes and grinded in liquid nitrogen. 4 mL of boiling absolute ethanol were quickly

added to a sample and the Corex tube transferred to a water bath at 85°C for 5 minutes for lipase inhibition. 2 mL of methanol and 8 mL of $ChCl_2$ were then added to the sample. Homogenisation of the sample was realised by Argon stirring for one minute. The Argon allows the removal of air from the sample and protects the lipids from oxidation. The sample was then capped with double aluminium cover and incubated at room temperature for one hour. Next, the sample was filtered through glass wool into a new Corex tube for removal of cell debris. 3 mL of $CHCl_2$/methanol (2:1; v/v) and 5 mL of NaCl 1 % were added to the sample to initiate biphase formation. The mix was homogenized by Argon stirring for 30 second. The sample was centrifuged for 10 minutes at 2,500 rpm at 4°C. Then, the lower organic phase containing the lipids was carefully transferred to a new Corex tube and dried under Argon flow. The lipids were recovered in 1.5 mL of $CHCl_2$, transferred to a glass haemolysis tube, dried again under Argon and stored at -20°C until use.

3.3.2.2. Methanolysis

One tenth of total glycerolipid extract was used for methanolysis. The lipids were solubilised with 1 mL of $CHCl_2$, and a 100-µL aliquot was transferred to a 10-mL crimp cap vial (Gerstel) using a Hamilton pipette. Methanolysis was performed by a MultiPurpose Sampler (MPS, Gerstel). Briefly, 5 µL of 15:0 FA (a FA with a 15-carbon chain length) at 1 $mg.mL^{-1}$ was added to the sample as internal standard for the following Gas Chromatography (GC) analysis. 3 mL of methanolysis buffer (methanol/sulphuric acid; 40:1 v/v) were added, vigorously mixed using a vortex, and incubated at 80°C for one hour. A transesterification reaction occurs between the glycerolipids, which contain FAs linked by ester bonds to a glycerol backbone, and the methanol in excess. This reaction yields FAs methyl esters (FAME) and glycerol backbones with different head group corresponding to the lipid classes methanolysed. This reaction is catalysed by the sulphuric acid, which allows the carbon of the reactant ester bond to be more nucleophile. Performing transesterification under acidic conditions also prevents reaction with the FAs of the sphingolipids linked by amide bonds. High temperature increases the rate of the reaction by facilitating hydrolysis of the reactant ester bond. Sample was cooled down at room temperature for 5 minutes before stopping the reaction with the addition of 3 mL of distilled water and 3 mL of hexane. Sample was vigorously mixed using a vortex for 5 seconds, then incubated at room temperature for 20 minutes. 1.8 mL of the upper phase containing the FAME was transferred to a 10-mL screw cap vial (MACHEREY-NAGEL). 2 mL of hexane were added to the sample, vigorously mixed using a vortex for 5 seconds, incubated 20 min at room temperature, and transferred to the same 10-mL screw cap vial. FAME were then dried under Argon flow and kept at -20°C.

3.3.2.3. Gas chromatography-Ion Flame Detection (GC-FID)

FAME were resuspended in 100 µL of hexane and analysed by Gas Chromatography (GC) coupled to Ion Flame Detection (FID). A GC-FID Perkin Elmer Clarus 580 equipped with a 30-m long cyanopropyl polysilphenesiloxane column with a diameter of 0.22 mm and a film thickness of 0.25 µm was used. Temperature increased from 130 to 180 °C, with nitrogen as vector gas. Identification of the FAME was achieved by comparison of their retention time with those of standard (Sigma-Aldrich). Surface peak method using the internal standard (15:0 FA) allowed the FAME species quantification and determination of the glycerolipid concentration in the initial sample.

3.3.2.4. Liquid Chromatograph Triple Quadrupole Mass Spectrometer (LC-MS/MS)

1 mL of $CHCl_2$ was used to resuspend total lipid extracts, and a volume containing 25 nmol of lipids was transferred to a glass haemolysis tube. Samples were dried under argon flow and kept at -20°C until use.

Glycerolipid dosage by high pressure liquid chromatography-tandem mass spectrometry was conducted as previously described (Dolch et al., 2017) by the LIPANG (Lipid Analysis in Grenoble) platform directed by Juliette Jouhet and Fabrice Rebeillé.

3.3.2.5. Lipid data analysis and statistical significance

Each lipid experiment was realised with three biological replicates and then repeated a second time. Data from both experiments were then treated in order to pool all replicates values. For analysis of FA profile, glycerolipid profile and molecular species profiles in the different lipids class, each value was first normalised as follow:

$V_N = \frac{V_i}{M_i} x\ M_x$, where V_N is the normalised value, V_i is the initial value, M_i is the median of the WT values in the associated experiment, and M_x is the median of all WT values from the two experiments.

Then, statistical differences were calculated by un unpaired multiple t test using the Sidak-Bonferroni method, without assuming consistent standard deviations, in GraphPad Prism software.

3.3.2.6. Nile red staining

TAG accumulation in lipid droplets was monitored by Nile Red (Sigma-Aldrich) fluorescent staining as previously described (Conte et al., 2018). Cell concentration in the liquid culture was determined by absorbance using a TECAN reader (see section 3.3.1). Then 160 µL of the culture was placed into a 96-wells flat bottom black plate (Greiner, 655086) and chlorophyll fluorescence determined with excitation and emission wavelengths of 440 and 680 nm respectively. Then, 40 µL of Nile Red fluorophore (9-diethylamino-5-benzo[α]phenoazinone; Sigma Aldrich) at 2.5 $\mu g.mL^{-1}$ in DMSO were added for lipid droplet staining. The plate was incubated in the dark at room temperature for 20 minutes before Nile Red measurement. Nile Red fluorescence was excited at 530 nm and read at 580 nm. Fluorescence was normalized to cell concentration ($Rfu.10^6\ cells^{-1}$).

3.3.3. Imaging

3.3.3.1. Epifluorescence microscopy

Cells were observed using an epifluorescence microscope (Zeiss AxioScope A1) equipped with a Zeiss AxioCam MRc. Images were captured using an oil-immersion x100 objective. Chloroplast autofluorescence was observed with a Fluorescein isothiocyanate filter (FITC: Excitation 488 nm / Emission 519 nm). Nile Red fluorescence in lipid droplets was also visualized with a FITC filter.

3.3.3.2. Transmission Electron Microscopy

Samples were prepared as previously described (Flori et al., 2018). Briefly, cells were fixed in 0.1 M phosphate buffer (PB) pH 7.4 containing 2.5 % (v/v) glutaraldehyde for 2 hours at room temperature, and then stored

overnight at 4 °C. Samples were then washed five times with 0.1 M PB pH 7.4. Samples were fixed in 0.1 M PB pH 7.4 containing 2 % osmium and 1.5 % ferricyanide potassium for 1 hour on ice. Samples were washed five times with 0.1 M PB pH 7.4 and then suspended in 0.1 M PB pH 7.4 containing 0.1 % tannic acid. Samples were incubated for 30 min in the dark at room temperature. Samples were washed five times with 0.1 M PB pH 7.4, dehydrated in ascending sequences of absolute ethanol, and infiltrated with ethanol/Epon resin mixture. Finally, the samples were embedded in Epon. Ultrathin sections (50–70 nm) were prepared with a diamond knife on a PowerTome ultramicrotome (RMC products, Tucson, AZ, USA) and collected on 200-µm nickel grids. Samples were visualized by scanning transmission electron microscopy (STEM) using a MERLIN microscope (Zeiss, Oberkochen, Germany) set up at 30 kV and 240 pA, or using a FEI tecnai OSIRIS microscope (Hillsboro, Oregon, USA) set up at 200 kV and ~300 pA.

3.3.3.3. Laser scanning confocal microscopy

Confocal microscopy images were obtained with a microscope Zeiss LSM880 equipped with a 63x/1.4 oil-immersed Plan-Apochromat objective, running Zen 2.3 SP1 acquisition software (Platform µLife, IRIG, LPCV). Chlorophyll autofluorescence and eGFP fluorescence were excited at 488 nm, and were detected at 600-730 nm and 500-531 nm, respectively. Laser intensity was set at 5 % for visualisation of chlorophyll, and at 2 %, 3 %, and 5 % for eGFP visualisation according to the eGFP signal intensity in each cell. 'Pseudo brightfield' images were acquired in parallel by differential interference contrast (DIC), using laser excitation at 488 nm at 0.6 % and a photomultiplier tube detector for transmitted light (T-PMT). Z-stacks containing consecutive images with a distance of 0.47 nm were obtained for each cell.

3.3.4. Photosynthetic activity determination

Fluorescence-based photosynthetic parameters were derived from imaging of chlorophyll fluorescence emission. Data were acquired with a pulse modulated amplitude fluorimeter, the MAXI version of *IMAGING-PAM M-Series* equipped with an IMAG-MAX/L LED-Array Illumination unit and an IMAG-K6 camera (Heinz Walz GmbH, Germany). A 200-µL aliquot of cells cultured in 100-mL flaks with 20 mL of ESAW medium at a concentration of 4-5.10^6 cells.mL^{-1} was deposited into a 96-wells flat bottom black plate (Greiner, 655086). Cells were dark acclimated for 15 minutes before measurements. Chlorophyll fluorescence was recorded using two different protocols as indicated. Both protocols start with measurements in the dark. In the first protocol, cells were subjected to a light intensity of 700 µmol photons.m^{-2}.s^{-1} during 9 minutes 30 seconds. Then, cells were left with a low light intensity of 20 µmol photons.m^{-2}.s^{-1} for 15 minutes to observe fluorescence relaxation. In the second protocol, cells were subjected to a 2-steps increase of light intensity from 55 to 335 µmol photons.m^{-2}.s^{-1}. Cells were allowed to reach steady state (time exposure of 15 minutes and then 10 minutes 30 seconds for each light intensity) before modification of the photon flux. Cells were then left with a light intensity of 20 µmol photons.m^{-2}.s^{-1} for 15 minutes for fluorescence relaxation measurements. Test parameters were optimised with the help of Dimitris Petroutsos (LPCV).

Effective photochemical quantum yield of PSII (Ψ_{II}) was calculated as ($F'm$-F_t)/$F'm$, and Non-Photochemical Quenching (NPQ) was calculated as ($F'm_0$-$F'm$)/$F'm$, where F_t is the steady-state fluorescence intensity

immediately prior to a saturating pulse of actinic light, while F'm and F'm_0 are the maximum fluorescence intensities in light- and dark-acclimated cells after a pulse, respectively (Maxwell and Johnson, 2000).

3.4. Protein analyses

3.4.1. Protein extraction

Frozen *Phaeodactylum* cell pellets in 50-mL polycarbonate tubes (see section 3.1.5) were lyophilised (CHRIST Alpha 2-4 LSCbasic) for a night before extraction to improve extraction yield. 350 µL of TANAKA buffer (sucrose 0.7 M, 0.5 M Tris, 30 mM HCl, 50 mM EDTA, 0.1 M KCl, 2 mM PMSF, 2 mM of β-mercaptoethanol, Antiprotease Roche 1x) were added on each frozen pellet. Cells were resuspended and transferred to a 2-mL tube. Another volume of TANAKA buffer (350 µL) was added to retrieve remaining cells and collect them in the same 2-mL tube. The cells were vigorously mixed using a vortex for 1 minute. 700 µL of phenol pH 7.5 were added to the mix and the sample mixed with a vortex for 1 minute. The sample was centrifuged at 20,000 g for 5 minutes at 4 °C. About 500 µL of the supernatant containing the proteins was transferred to a 1.5-mL tube. 1 mL of cold ammonium acetate 0.1 M in methanol was added. The sample was incubated at -20°C overnight for protein precipitation, then centrifuged at 20,000 g for 20 minutes at 4 °C and the supernatant was removed. The protein pellet was washed with 1 mL of cold ammonium acetate 0.1 M and centrifuged again at 20,000 g for 20 minutes at 4 °C. The supernatant was removed and this cleaning step was repeated twice. Finally, 1 mL of cold acetone 100 % was added to the pellet and directly removed. The sample was left to dry at room temperature for 30 minutes. 50 µL of extraction buffer (8 M urea, 1 mM EDTA, 50 mM Tris pH 6.8, 5 mM β-mercaptoethanol) were added to the pellet. The pellet was suspended with gentle tap and incubated for 2 hours before protein concentration measurement.

3.4.2. Protein assay

Protein concentration was determined using the Bradford method (Bradford, 1976). A 2-µL aliquot of protein sample was incubated with 1 mL of Bradford solution (Bio-Rad, Protein Assay Dye Reagent Concentrate; undisclosed proprietary composition) for 5 minutes at room temperature. Absorbance was measured at 595 nm with a spectrophotometer (Eppendorf Biophotometer D30). Absorbance in the sample was converted to protein concentration using a linear regression calculated from a protein scale prepared with standard quantities of Bovine Serum Albumin (BSA) set at 0, 2, 4, 6, 8, and 10 µg.

3.4.3. Recombinant protein production

3.4.3.1. *E. coli* heterologous expression vector constructions

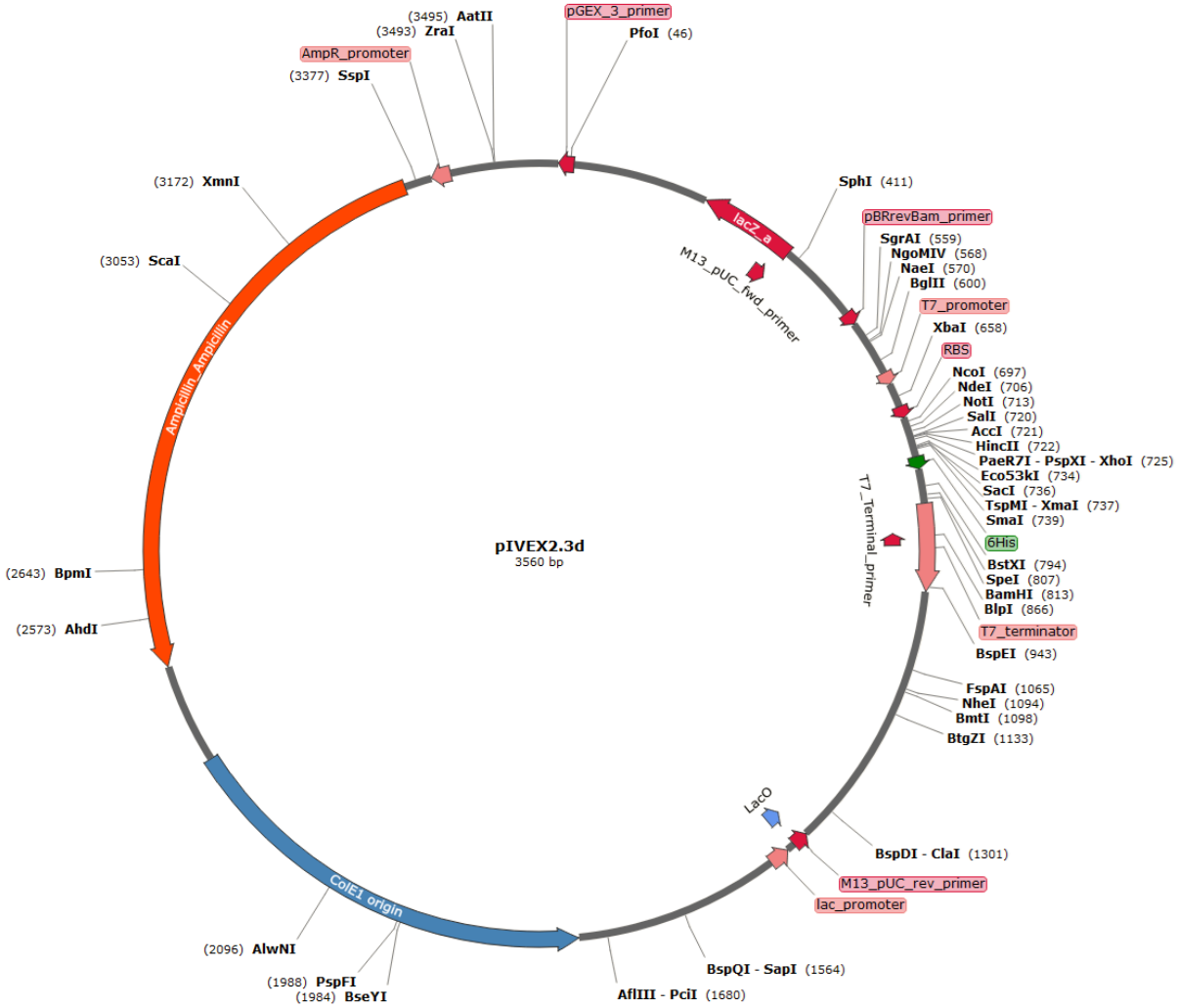


Figure 3.3: Map of the pIVEX2.3d expression vector. Vector map was constructed with SnapGene software (from Insightful Science; available at snapgene.com) with functional sites and restriction enzyme sites. PIVEX2.3d was used for the expression of MGD proteins in cell free system. Abbreviations: AmpR, Ampicillin Resistance; ColE1 origin, origin of replication; lacO, bacterial lac operator; lacZ_a, lacZ alpha gene; RBS, Ribosome Binding Site.

Expression vector for protein production pIVEX2.3d (**Fig. 3.3**) (biotechrabbit, BR1400701) was kindly provided by Lionel Imbert (IBS, Grenoble). *MGDαΔ126*, *MGDβΔ285* and *MGDγΔ189* genes were amplified by PCR (Phusion High-Fidelity DNA polymerase, F-530L, ThermoScientific) using *P. tricornutum* cDNA as a template, and MGDα-Fwd-NcoI with MGDα-Rev-SmaI, MGDβ-Fwd-NdeI with MGDβ- Rev-SmaI, and MGDγ-Fwd-NdeI with MGDγ- Rev-SacI as primers respectively (**Table 3.1**) (1307 bp, 1418 bp, and 1691 bp amplification products respectively). Primers for amplification included enzyme restriction sites for cloning. *MGDαΔ126*, *MGDβΔ285* and *MGDγΔ189* PCR products were digested with NcoI/SmaI, NdeI/SmaI and NdeI/SacI enzymes respectively and each cloned in pIVEX2.3d vector, generating pIVEX2.3d-MGDαΔ126, pIVEX2.3d-MGDβΔ285 and pIVEX2.3d-MGDγΔ189. These plasmids carry the MGD*αΔ126*, *MGDβΔ285* and *MGDγΔ189* genes followed by a non-cleavable C-terminal His_6-tag, generating the MGDαΔ126-His_6, MGDβΔ285-His_6 and MGDγΔ189-His_6 protein constructions. Plasmids were amplified in *E. coli* DH5α and harvested by a midiprep method to obtain a concentration of 1 $\mu g.\mu L^{-1}$ of plasmid.

3.4.3.2. Cell-free system protein production

Protein production was performed under the supervision of Lionel Imbert at the Cell Free expression platform of ISBG (Integrated Biology Structural Grenoble) at IBS (Institut de Biologie Structurale, Grenoble). Expression of each protein was assessed in different conditions before large-scale production. Expression with 12 mM, 14 mM and 16 mM of magnesium acetate ($MgO(Ac)_2$) were assessed. The use of different detergents for protein solubilisation was also evaluated using Brij 35 10 CMC (critical micelle concentration), Brij 35 50 CMC, Brij 58 10 CMC, Brij 58 50 CMC, DDM (Dodecyl-β-D-maltoside) 25 CMC and DDM 50 CMC. Following the first trials, successful expression of MGDβΔ95-His_6 and MGDγΔ62-His_6 proteins were obtained for a $MgO(Ac)_2$ concentration of 16 mM, and the use of Brij 35 50 CMC and Brij 58 10 CMC respectively.

Large-scale production was performed as previously described (Imbert et al., 2021) in GeBAflex dialysis device (Euromedex). A volume of 90 mL of feeding mix was prepared as follows: 0.8 mM of each CTP, GTP and UTP ribonucleotides (NU-1014, Jena Bioscience), HEPES 55 mM (Merck), ATP 1.2 mM, Folinic Acid 68 µM (Merck), cyclic AMP 0.64 mM (Merck), DTT 3.4 mM (Merck), spermidine 2 mM (Merck), NH_4OAc 27.5 mM (Merck), creatine phosphate 80 mM (Merck), 20 amino acids 1 mM each, potassium glutamate 208 mM (Merck), $Mg(OAc)_2$ 16 mM, and detergent at defined concentration. A volume of 9 mL of reaction mix was prepared as described for the feeding mix, with further addition of MRE600 tRNA 0.175 $mg.mL^{-1}$ (Roche), creatine kinase 250 $\mu g.mL^{-1}$ (Roche), homemade T7 RNA polymerase 1/100 (Imbert et al., 2021), homemade S30 extract 40% v:v (Imbert et al., 2021), and expression vector 10 $\mu g.mL^{-1}$. Three separate dialysis Gebaflex devices were loaded with 3 mL of reaction mix each, and placed inside a 50-mL cylinder containing 30 mL of the feeding mix and incubated overnight with stirring at 23°C before protein purification.

3.4.3.3. Protein purification

After protein production at the Cell Free expression platform, proteins were purified on a his-tagged protein purification resin. 1 mL of resin Ni Sepharose High Performance (GE17-5268-01, Merck) was loaded into a column and washed firstly with 20 mL of distilled water, then 25 mL of binding buffer (10 mM MOPS-KOH pH7.8, 6 mM CHAPS, 1 mM TCEP, and 50 mM KH_2PO_4/K_2HPO_4 pH 7.8). Samples were prepared by mixing 1 volume of protein mix with 1 volume of binding buffer concentrated twice (20 mM MOPS-KOH pH 7.8, 12 mM CHAPS, 2 mM TCEP, and 100 mM KH_2PO_4/K_2HPO_4 pH 7.8). Prepared mix was loaded on the column and the flowthrough was collected in a 15-mL polycarbonate tube. Then, weakly bound proteins were removed by the loading of 40 mL of binding buffer supplemented with imidazole 20 mM through the column. The flowthrough was collected in a 50-mL polycarbonate tube. The protein of interest was finally eluted with 10 mL of binding buffer supplemented with imidazole 300 mM. Elution was collected 1 mL at a time in separate 1.5-mL tubes. Protein purification was checked by gel electrophoresis (12-wells Bolt 4-12 % protein gel, Bis-Tris, ThermoFisher Scientific). Protein concentration was evaluated by Bradford assay (see section 3.4.2). Each 1 mL fraction after elution was divided in 250-µL aliquots in 5 % glycerol in 0.5-mL tubes, snap-freezed in liquid nitrogen and stored at -80°C.

3.4.4. MGDG synthase assay

MGDβΔ285-His_6 and MGDγΔ189-His_6 proteins activity was tested using the UDP-Glo Glycosyltransferase Assay (Promega). MGD enzymes interact with the membrane where their substrate is located. Thus, micelles were prepared to serve as platform to test their activity. In a glass tube, 10 µL of either dioleoylglycerol (18:1-18:1; Sigma Aldrich) or dipalmitoleylglycerol (16:1-16:1; Larodan) at 1 $mg.mL^{-1}$ in $CHCl_2$ and 4 µL of either phosphatidic acid (16:0-18:1; Avanti), phosphatidylglycerol (16:0-18:1; Avanti) or sulfoquinovosyldiacylglycerol (extracted from *P. tricornutum*) at 1 $mg.mL^{-1}$ in $CHCl_2$ were loaded, then dried under nitrogen flow. After evaporation, 12 µL of reaction buffer concentrated three times (75 mM Bis-Tris pH 6.5, 450 mM NaCl, 15 % glycerol, 3 mM TCEP and 18 mM CHAPS) and 17.5 µL of water were added to the glass tube. The glass tube was vigorously mixed using a vortex, then sonicated at 35 kHz for 10 minutes in a water bath filled with melting ice. After sonication, 2 µL of protein at 0.05 $mg.mL^{-1}$ (see section 3.4.3.3) were added to the glass tube. The reaction was started by adding 5 µL of UDP-Gal at 5 mM into the tubes (total reaction volume: 36.5 µL). Glass tubes were vigorously mixed using a vortex and then incubated at room temperature for 10 minutes. Enzymatic assays were performed in triplicate for each condition. Negative controls with either no DAG (substrate) or no UDP-Gal (donor) were also prepared. A positive control was done using recombinant AtMGD1Δ137 (produced in E. coli by Dylan Jabeguero from CERMAV, Grenoble) as a functional glycosyltransferase. Diluted UDP-Glo enzyme was prepared by mixing 1 µL of enzyme with 74 µL of enzyme dilution buffer. Then, UDP-detection reagent (UDR) was prepared by mixing 6 µL of diluted enzyme with 600 µL of nucleotide detection reagent. In a 96-wells flat bottom white lumitrac plate (Greiner bio-one), 25 µL of UDR were added per well and per reaction. After 10 minutes of glycosyltransferase reaction, 25 µL of the reaction mix were transferred from the glass tube to the plate to stop the reaction. The plate was incubated 1 hour at room temperature in the dark. During incubation, UDP produced by the glycosyltransferase reaction is converted to ATP, which is then converted into a luminescent signal that is proportional to the glycosyltransferase activity. The luminescence generated by the assay is highly stable. Luminescence in the plate was measured with a TECAN Spark 10M microplate reader after incubation.

3.4.5. Antibody production

Antibody production was performed in collaboration with the Agro-bio company (Stago group). For each MGD protein of *P. tricornutum*, three peptides were designed with the assistance of Agro-Bio experts to serve as antigens. For MGDα, the antigens were peptide 1, CLSRPSGNSGKKADM; peptide 2, DSNGATRERTGRKTPFC; and peptide 3, CGNNQEAKASLEKESWG. For MGDβ, antigens were peptide 1, CWRRRRSRRLQESLDR; peptide 2, RNENLRLSLEQRDWDC; and peptide 3, ERIRRLTKRRGGTEDC. For MGDγ, antigens were peptide 1, CPDSDHKRTADSVGEE; peptide 2, KWKELNDATHDQSTTC; and peptide 3, DRWNGALERNGISSSC. These peptides were produced by Agro-bio. Six rabbits were selected for antibody production and pre-immune rabbit serums were sampled a week before first antigen injection. All three antigens for each MGD protein were injected to two different rabbits at day 0, day 7, day 14, day 34, day 74, and day 104. Rabbits 00472 and 00478 were used for MGDα, rabbits 00585 and 00559 for MGDβ, and rabbits 00476 and 00492 for MGDγ. Samplings of rabbit serum were performed at day 28, day 42, day 84 and day 111. ELISA tests were achieved on pre-immune rabbit

serums and on rabbit serums at day 28, day 84 and day 111. All rabbit serums were also tested by Western blot analysis on *P. tricornutum* WT total proteins extracts (see section 3.4.6). Antibodies were purified on protein A from 20 mL of rabbit 585 exsanguination, and from 18 mL of day 84 samplings of rabbits 00472 and 00492. For MGDα, β and γ, 135.2 mg, 114 mg and 46.25 mg of antibodies were obtained, respectively.

3.4.6. Western-blot analysis

For the validation of produced MGD antibodies, samples containing 8 µg of *P. tricornutum* total proteins, 5 µL of Blot LDS Sample Buffer 4X (ThermoFisher Scientific), 2 µL of Sample Reducing Agent 10X (Life technologies), in a total volume of 20 µL adjusted with water were prepared and kept at -20°C until use. Prior electrophoresis, proteins were denatured by heating the samples at 95 °C for 5 minutes. 3 µL of Precision Plus Protein standard (Bio-Rad) were used as protein size markers. Denatured samples and size markers were subjected to electrophoresis on a 15-wells Bolt 4-12% Bis-Tris Plus gel (ThermoFisher Scientific) in a bath containing Blot MES SDS Running buffer (ThermoFisher Scientific). Migration occurred at 100 V for 10 minutes, then 120 V for 1 hour. Proteins were transferred to a 0.2-µm nitrocellulose blotting membrane (Amersham Protran) using a Mini Trans-blot cell (Bio-Rad) with transfer buffer (192 mM Glycine, 25 mM Tri-HCl pH 8.3, 20 % ethanol v/v). Transfer was performed at 90V, 35 mA for 1 hour 15 minutes. The membrane was washed in distilled water, and then placed in Red Ponceau preparation (0.2 % p/v Red Ponceau, 3 % v/v TCA) for 1 minute to visualize protein profile. The membrane was washed again in water and a picture was taken with a Camag TLC Vizualizer (Switzerland). Red Ponceau was removed with TBS buffer (Tris-HCl 20 mM pH 7.5, 150 mM NaCl). Membrane was blocked with TBS milk 5 % for 1 hour at room temperature under agitation. For antibody fixation, the membrane was incubated with TBS milk 5 % and a 1/1,000 dilution of the antibody at 4°C under agitation overnight. After incubation, the membrane was washed with TBS tween 0.1 % three times. Secondary antibody was fixed by incubating the membrane in TBS milk 5 % with a 1/10,000 dilution of HRP Goat IGG Anti Rabbit IGG (Interchim) at room temperature under agitation for 3 hours. The membrane was then washed in TBS Tween 0.1 %, and finally washed in TBS. For antibody revelation, Clarity Western ECL Substrate (Bio-Rad) was used. Imaging of bands reacting with the antibodies was achieved with an Image Quant 800 (Amersham·).

For visualisation of MGD-eGFP proteins in total protein extracts of MGD overexpressing *P. tricornutum* strains, we used the same protocol with minor modifications. Total proteins quantity in the samples was adjusted to 10 µg, and a 1/5,000 dilution of anti-GFP antibody HRP was used. Fixation of a secondary antibody is not necessary.

3.5. *In silico* analyses

3.5.1. Retrieval of *MGD* gene sequences

MGD genes in *P. tricornutum* were identified by sequence homology using a Basic Local Alignment Search Tool (BLAST). The BLASTP was used to search the protein database for *P. tricornutum* (taxid: 2850) using AtMGD1 (NP_194906, NCBI), AtMGD2 (NP_568394.2, NCBI), AtMGD3 (NP_565352.1, NCBI), soMGD1 (XP_021852153.1,

NCBI), NtMGD (NP_001312377.1, NCBI), CsMGD (NP_001292699.1, NCBI), and GmMGD (NP_001236930.1, NCBI) as protein queries. The identified locus tags of *P. tricornutum MGD* genes in NCBI are PHATRDRAFT_14125 for *MGD*α, PHATR_43938 for *MGDβ*, and PHATRDRAFT_9619 for *MGDγ*. All genes, cDNA and protein sequences for *P. tricornutum* MGDs were downloaded from EnsemblProtists. Gene identification numbers in EnsemblProtists are J14125 for *MGDα*, J54168 for *MGDβ*, and J9619 for *MGDγ*.

3.5.2. Targeting sequence prediction

Target sequence predictions were performed using several bioinformatics online servers. For each MGD in *P. tricornutum*, protein sequences were used as queries considering each possible START Methionine. For signal peptide prediction, SignalP-6.0 (https://services.healthtech.dtu.dk/service.php?SignalP) and HECTAR (https://webtools.sb-roscoff.fr/root?tool_id=abims_hectar) (Gschloessl et al., 2008; Teufel et al., 2022) were used. For chloroplast transit peptide prediction, WolF PSORT (https://wolfpsort.hgc.jp/) (Horton et al., 2007) and ChloroP - 1.1 (https://services.healthtech.dtu.dk/service.php?ChloroP-1.1) (Emanuelsson et al., 1999) were used.

3.5.3. Protein structure prediction

Each *P. tricornutum* MGD protein sequence was subjected to the Phyre^2 (Protein Homology/analogY Recognition Engine V 2.0) web portal for protein modelling, prediction and analysis (http://www.sbg.bio.ic.ac.uk/phyre2/html/page.cgi?id=index) (Kelley et al., 2015a). Protein models returned by the server were viewed with PyMOL software. AtMGD1 protein model (4X1T) was retrieved from UniProtKB server.

3.5.4. Phylogenetic analysis

All MGD protein sequences from other species used for phylogenetic analysis were retrieved from the National Centre for Biotechnology Information (NCBI) (**Table 3.3**) after a protein BLAST search using MGDα, MGDβ and MGDγ as protein queries. Sequences from different phyla were retained for phylogenetic analysis. The final dataset was manually curated in order to remove double sequences, misannotated sequences, etc. Sequences from the final dataset were individually inspected to validate the presence of typical conserved MGD domains including the active site. In the case of the allodiploid genome of *Fistulifera solaris*, we only kept one of each protein version: the proteins used in our tree are GAX23913.1, GAX09439.1, and GAX10983.1, which pair with GAX11826.1, GAX13030.1, and GAX24866.1, respectively.

Kingdom	Division or phylum	Class	Order	Family	Genus and Species	Protein idenfifier
Protozoa	Alveolata	Colpodellida	Chromerida	Vitrellaceae	*Vitrella brassicaformis*	CEM14063.1
						CEM34604.1
		Conoidasida (1)	Eugregarinorida	Gregarinidae	*Gregarina niphandrodes*	XP_011130274.5
Chromista	Stramenopiles	Bacillariophyceae (2)	Bacillariales	Bacillariaceae	*Pseudo-nitzschia multistriata*	VEU37191.1
						VEU39505.1
					Fragilariopsis cylindrus	OEU16730.1
						OEU06991.1
			Naviculales	Stauroneidaceae	*Fistulifera solaris*	GAX23913.1
						GAX10983.1
						GAX09439.1
				Phaeodactylaceae	*Phaeodactylum tricornutum*	XP_002186355
						XP_002176800
						XP_002181685
		Mediophyceae (2)	Thalassiosirales	Thalassiosiraceae	*Thalassiosira pseudonana*	XP_002295865
						XP_002293576.1
						XP_002294242
		Phaeophyceae	Ectocarpales	Ectocarpaceae	*Ectocarpus siliculosus*	CBJ28381.1
						CBJ28372.1
						CBN79326.1
		Pelagophyceae	Pelagomonadales	Pelagomonadaceae	*Aureococcus anophagefferens*	XP_009033523.1
						XP_009038839.1
						XP_009035780.1
		Eustigmatophyceae	Eustigmatales	Monodopsidaceae	*Microchloropsis gaditana/ Nannochloropsis gaditana*	XP_005855249.1
Plantae	Chlorophyta	Trebouxiophyceae	*incertae sedis*	Coccomyxaceae	*Coccomyxa subellipsoidea*	XP_005651388.1
			Chlorellales	Chlorellaceae	*Chlorella sorokiniana*	PRW45643.1
		Mamiellophyceae	Mamiellales	Bathycoccaceae	*Ostreococcus tauri*	OUS42062.1
		Chlorophyceae	Chlamydomonadales	Chlamydomonadaceae	*Chlamydomonas reinhardtii*	PNW74102.1
			Sphaeropleales	Selenastraceae	*Monoraphidium neglectum*	XP_013903204.1
					Raphidocelis subcapitata	GBF88428.1
	Magnoliophyta - Flowering plant	Magnoliopsida - Dicotyledons	Brassicales	Brassicaceae	*Arabidopsis thaliana*	NP_194906.1
						NP_565352.1
						NP_568394.2
			Rubiales	Rubiaceae	*Coffea arabica*	XP_027069738.1
						XP_027084801.1
			Caryophyllales	Chenopodiaceae	*Spinacia oleracea*	XP_021867203.1
						XP_021852153.1
			Nymphaeales	Nymphaeaceae	*Nymphaea colorata*	XP_031495579.1
						XP_031480803.1
			Amborellales	Amborellaceae	*Amborella trichopoda*	XP_006852865.1
						XP_006845407.1
		Liliopsida - Monocotyledons	Cyperales	Poaceae / Gramineae	*Brachypodium distachyon*	XP_010238179.1
						XP_003573675.2
						XP_003570331.1
					Oryza sativa	XP_015611851.1
						XP_015649135.1
						XP_015627712.1
	Bryophyta - Mosses	Bryopsida - True mosses	Funariales	Funariaceae	*Physcomitrium patens*	XP_024403181.1
						XP_024404076.1
	Lycopodiophyta - Lycopods	Lycopodiopsida	Selaginellales	Selaginellaceae	*Selaginella moellendorffii*	XP_024533920.1
	Rhodophyta	Cyanidiophyceae	Cyanidiales	Cyanidiaceae	*Cyanidioschyzon merolae*	XP_005536420.1
				Cyanidioschyzonaceae	*Cyanidiococcus yangmingshanensis*	KAF6002292.1
		Florideophyceae	Gracilariales	Gracilariaceae	*Gracilariopsis chorda*	PXF49423.1
						PXF42956.1
						PXF46603.1
Bacteria	Proteobacteria	Alphaproteobacteria	Hyphomicrobiales	Blastochloridaceae	*Blastochloris viridis*	WP_055036643.1

(1): From Apicomplexa clade
(2): From the diatom group

Table 3.3: Taxonomic classification of the MGD proteins used for phylogenetic analyses. Taxonomic classification according to Cavalier-Smith's system of classification (Turland et al., 2018; Burki et al., 2020). All selected proteins were predicted to be MGD proteins based on similarities with the MGD isoforms from *P. tricornutum*.

The alignment was performed with a customized pipeline in NGphylogeny.fr (Lemoine et al., 2019) using MUSCLE v3.8.1551 software. The ambiguously aligned regions were curated using the Block Mapping and Gathering with Entropy (BMGE v1.12_1) software implemented in NGphylogeny.fr using default settings. Preliminary phylogenetic trees were inferred using FastME with 2,500 bootstrap pseudoreplicates in order to rule out any aberrant results. MEGA X v10.0.5 software was fed with the aligned and curated data set. The best evolutionary model was evaluated and a Maximum Likelihood phylogenetic analysis was performed. In order to define the best evolutionary model, MEGA X was used to compare the 56 models implemented. The LG+G model was

chosen (lowest Bayesian Information Criterion (BIC) score). The phylogeny was inferred by Maximum Likelihood with 5,000 bootstrap pseudoreplicates. The tree with the highest log likelihood (-12414.63) was retained. Initial trees for the heuristic search were obtained by applying the BioNJ method (Gascuel, 1997) to a matrix of pairwise distances estimated using a JTT model. A discrete Gamma distribution was used to model evolutionary rate differences among sites (five categories (+G, parameter = 0.7081)). The best phylogenetic tree was retrieved in Newick format and visualized by iToL.

4. Results

4.1. In silico analysis of predicted MGD genes in *Phaeodactylum tricornutum*

Three MGDs have been predicted previously in *P. tricornutum* based on sequence similarity (Petroutsos et al., 2014). We re-evaluated this prediction using as queries the protein sequences of MGDs from Arabidopsis, spinach, tobacco, soybean and cucumber. We identified the same genes as previously reported, recorded in NCBI as *PHATRDRAFT_14125* (on chromosome 13), *PHATR_43938* (on chromosome 3), and *PHATRDRAFT_9619* (on chromosome 1), that we renamed *MGDα*, *MGDβ* and *MGDγ*, respectively.

Four to five peptides per MGD were detected in a whole-cell proteome analysis of *P. tricornutum* (Lupette et al., 2019) (**Supplemental Table 1**), validating that all three genes are expressed.

4.1.1. Reconstructed evolution of MGD genes through gene duplications and mutations

In order to investigate on the evolutionary origin of the MGD protein isoforms in *P. tricornutum*, an unrooted phylogenetic tree was generated using MGD protein sequences from different species retrieved from NCBI. A balanced number of species from different phyla were selected (**Table 3.3**). In particular, we selected species belonging to flowering plants (angiosperms or magnoliophyta), mosses (bryophyta), lycopods (lycopodiophyta), green algae (chlorophyta), red algae (rhodophyta), stramenopiles, alveolates, and a proteobacterium.

Fistulifera solaris is a diatom with an allodiploid genome structure. Therefore, its genes exist in two versions. We judged preferable for our tree to keep only one of each version as the differences between them were negligeable. For information, the proteins used in our tree are GAX23913.1, GAX09439.1, and GAX10983.1, which pair with GAX11826.1, GAX13030.1, and GAX24866.1 (not shown), respectively.

A distinct phylogenetic tree of MGD proteins had previously been generated, including *P. tricornutum* MGDs (Yuzawa et al., 2012). This tree was centred on the origin of MGD in photosynthetic membranes, and although it included many MGD homologues from prokaryotes absent in our tree, it contained fewer species from the green and red lineages. In both trees, MGDs from the stramenopiles and red algae are closely related, emphasising that stramenopiles derive from the engulfment of a red algae, from which they acquired an ancestral MGD.

The proteobacterium *Blastochloris viridis* is a photosynthetic purple alphaproteobacterium containing a very basic photosystem made only of PSII. The position of the MGD from this bacterium is quite different from that of the MGDs from the Chloroflexi phylum and the cyanobacteria (Yuzawa et al., 2012).This MGD protein clusters with the viridiplantae, possibly indicating that horizontal gene transfer might have occurred between ancestors of these organisms.

We first note that the phylogenetic tree we obtained is globally congruent with classification, organised in major clades corresponding to the different phyla (**Fig. 4.1**). Noteworthy, we see that the stramenopiles MGDs are polyphyletic. MGDα clusters with other diatoms and other stramenopiles (Eustigmatophyceae, Pelagophyceae, Phaeophyceae) putative MGD proteins, in a sister clade to the alveolates, whereas MGDβ and MGDγ form a

separate clade, sister to Pelagophyceae species. The Stramenopile clade containing MGDα clusters in a basal position to the alveolates, the green and the red lineages. Thus, MGDα, may have been vertically inherited from a MGD ancestor common to all lineages. MGDβ and MGDγ likely stemmed from gene duplication along the evolution at the base of diatoms. Consistently, Yuzawa *et al* concluded on the single origin of MGD in a common ancestor of green, blue and red lineages.

MGD proteins in the land plants also form a polyphyletic group (**Fig. 4.1**). In the angiosperms, selected species contain two to three MGD proteins. Only one MGD protein from each angiosperms species clusters together with type A AtMGD1, while the second and third MGDs cluster in a distinct clade with the type B AtMGD2 and AtMGD3 characterised in *A. thaliana* (Awai et al., 2001), as previously observed (Yuzawa et al., 2012). Gene duplication and subsequent functional differentiation thus occurred in the angiosperms, leading to the observation of two distinct clades containing type A and type B enzymes (Awai et al., 2001). Interestingly, MGD proteins from bryophyta and lycopodiophyta (which encode only one MGD) cluster together with type A proteins. This would lead to the conclusion that AtMGD1 may be the most ancient isoform in the land plants (Shaw et al., 2011).

A similar observation can be made within the stramenopiles. With the exception of *Microchloropsis gaditana* (previously known as *Nannochloropsis gaditana*), all selected stramenopiles contain two to three predicted MGD protein isoforms. The unique *M. gaditana* MGD clustered with MGDα, along with all the *Ectocarpus siliculosus* MGDs. All MGD proteins from the other species grouped within distinct clades containing MGDα, MGDβ and MGDγ, revealing a possible similar evolution of vertical inheritance of one MGD (homolog to MGDα) and subsequent duplication. The diatoms *Fistulifera solaris* and *Thalassiosira pseudonana* possess three predicted MGDs clustering separately with each MGDs of *P. tricornutum*. *Aureococcus anophagefferens* also possesses three MGDs, one clustered with MGDα but the other two clustered to a more basal position to the MGDβ/MGDγ clade. Interestingly, *Fragilariopsis cylindrus* seems to possess one MGDβ-like protein and one MGDγ-like protein, but no MGDα-like protein. In this species, a MGDα-like isoform was either undetected in the available genomic data, or lost.

In conclusion, all selected diatom species contain at least two MGD proteins, each grouping specifically with one of the *P. tricornutum* predicted MGDs.

We sought to reconstruct a possible evolutionary scenario. A first gene duplication of the ancestral MGD leading to MGDα, MGDβ and MGDγ could have led to a common ancestor to both MGDβ and MGDγ isoforms, which would later duplicate again. The first duplication is not shared by all stramenopiles, and seems to have occurred after the separation of the eustigmatophyceae and phaeophyceae from the other stramenopile classes. Indeed, the selected species for the eustigmatophyceae and phaeophyceae remain with a MGDα-like protein only. It cannot be ruled out however that these species could have lost the MGDβ-like and MGDγ-like isoforms. Syntheny and comparative genomics studies may shed light on the evolution of *MGD* genes in the stramenopiles, but this goes beyond the purposes of the present thesis.

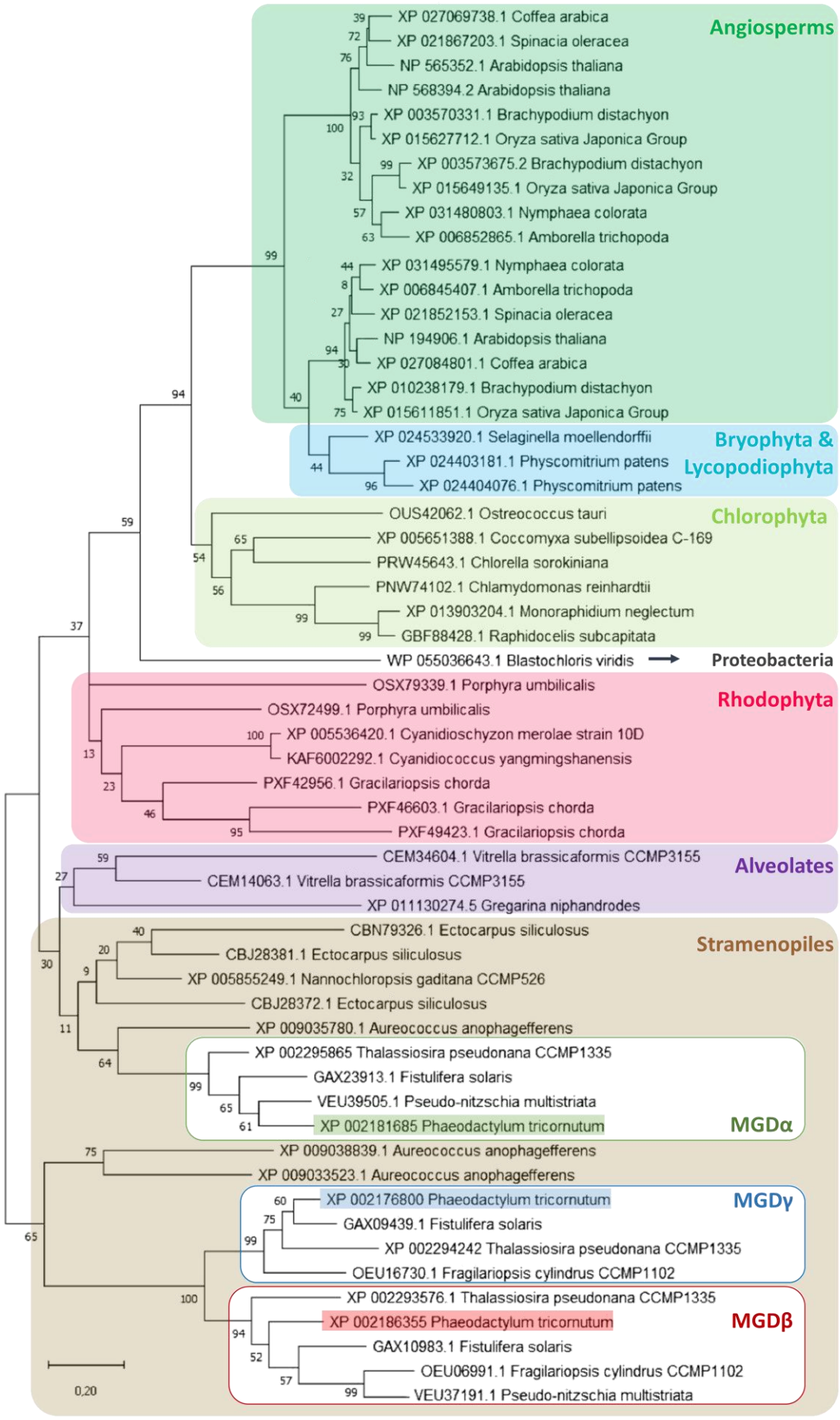

Angiosperms
XP 027069738.1 Coffea arabica
XP 021867203.1 Spinacia oleracea
NP 565352.1 Arabidopsis thaliana
NP 568394.2 Arabidopsis thaliana
XP 003570331.1 Brachypodium distachyon
XP 015627712.1 Oryza sativa Japonica Group
XP 003573675.2 Brachypodium distachyon
XP 015649135.1 Oryza sativa Japonica Group
XP 031480803.1 Nymphaea colorata
XP 006852865.1 Amborella trichopoda
XP 031495579.1 Nymphaea colorata
XP 006845407.1 Amborella trichopoda
XP 021852153.1 Spinacia oleracea
NP 194906.1 Arabidopsis thaliana
XP 027084801.1 Coffea arabica
XP 010238179.1 Brachypodium distachyon
XP 015611851.1 Oryza sativa Japonica Group
Bryophyta & Lycopodiophyta
XP 024533920.1 Selaginella moellendorffii
XP 024403181.1 Physcomitrium patens
XP 024404076.1 Physcomitrium patens
Chlorophyta
OUS42062.1 Ostreococcus tauri
XP 005651388.1 Coccomyxa subellipsoidea C-169
PRW45643.1 Chlorella sorokiniana
PNW74102.1 Chlamydomonas reinhardtii
XP 013903204.1 Monoraphidium neglectum
GBF88428.1 Raphidocelis subcapitata
WP 055036643.1 Blastochloris viridis
Proteobacteria
Rhodophyta
OSX79339.1 Porphyra umbilicalis
OSX72499.1 Porphyra umbilicalis
XP 005536420.1 Cyanidioschyzon merolae strain 10D
KAF6002292.1 Cyanidiococcus yangmingshanensis
PXF42956.1 Gracilariopsis chorda
PXF46603.1 Gracilariopsis chorda
PXF49423.1 Gracilariopsis chorda
Alveolates
CEM34604.1 Vitrella brassicaformis CCMP3155
CEM14063.1 Vitrella brassicaformis CCMP3155
XP 011130274.5 Gregarina niphandrodes
Stramenopiles
CBN79326.1 Ectocarpus siliculosus
CBJ28381.1 Ectocarpus siliculosus
XP 005855249.1 Nannochloropsis gaditana CCMP526
CBJ28372.1 Ectocarpus siliculosus
XP 009035780.1 Aureococcus anophagefferens
XP 002295865 Thalassiosira pseudonana CCMP1335
GAX23913.1 Fistulifera solaris
VEU39505.1 Pseudo-nitzschia multistriata
XP 002181685 Phaeodactylum tricornutum
MGDα
XP 009038839.1 Aureococcus anophagefferens
XP 009033523.1 Aureococcus anophagefferens
XP 002176800 Phaeodactylum tricornutum
GAX09439.1 Fistulifera solaris
XP 002294242 Thalassiosira pseudonana CCMP1335
OEU16730.1 Fragilariopsis cylindrus CCMP1102
MGDγ
XP 002293576.1 Thalassiosira pseudonana CCMP1335
XP 002186355 Phaeodactylum tricornutum
GAX10983.1 Fistulifera solaris
OEU06991.1 Fragilariopsis cylindrus CCMP1102
VEU37191.1 Pseudo-nitzschia multistriata
MGDβ
0,20

Figure 4.1: Phylogenetic analysis of the MGDs in photosynthetic organisms. Best evolutionary model obtained with MEGA X software, inferred by Maximum Likelihood with 5,000 bootstrap pseudoreplicates. The percentage of trees in which the associated taxa clustered together is shown next to the branches. The tree is drawn to scale, with branch lengths measured in the number of substitutions per site. Filled rectangles differentiate the main phyla, while delineated rectangles differentiate the clades in which each MGD from *P. tricornutum* (highlitghted) is found in the stramenopiles.

The *MGD* gene expansion in some Stramenopile lineages, taking into account the occurrence of organisms coding for only one MGD, like Eustigmatophytes, leads to the conclusion that in organisms bearing secondary plastids bounded by four membranes, only one MGD is sufficient for the cell requirements in MGDG.

Why do diatoms contain more than one MGD? Emergence of the MGDβ-like and MGDγ-like isoforms might have served a similar purpose as in *A. thaliana*, that is to cope to specific environmental stresses (Awai et al., 2001), or it might be linked to the complex cellular architecture of these secondary endosymbionts. To try answering this question, we focused our study on the model diatom *P. tricornutum*.

4.1.2. Three MGD gene sequences predicted in *P. tricornutum* genome: *MGDα*, *β* and *γ*

In order to study the *MGD* isoforms coding sequences (CDS) *in silico*, we retrieved the gene and protein sequences using the resources from NCBI and EnsemblProtists browser. EnsemblProtists was the most up to date for *P. tricornutum*. *MGDα* presents two possible CDS in EnsemblProtists: *PHATR*3*_J14125* (corresponding to the gene *PHATRDRAFT_14125* on NCBI) and *PHATR*3*_EG02525*. *PHATR*3*_EG02525* is the longest CDS (1,725 bp) and corresponds to a gene length of 1,930 bp (**Fig. 4.2**). *PHATR*3*_J14125* CDS (1,623 bp) begin with a start codon located 102 bp downstream of the predicted start codon for *PHATR*3*_EG02525*, with a corresponding gene length of 1828 bp. MGDα is predicted to contain three exons, with exon 1 being longer in the case of *PHATR*3*_EG02525* CDS (**Fig. 4.2**). In the NCBI database, the partial mRNA for this gene does not cover the entire coding sequence of either *PHATR*3*_J14125* or *PHATR*3*_EG02525*, but localisation of the introns is identical.

MGDβ is a 1,805-bp gene according to EnsemblProtists, with two exons. In NCBI, the partial mRNA for this gene starts 116 bp upstream the start codon predicted in EnsemblProtists, and localisation of the intron is consistent between genomic resources.

MGDγ is the longest of the three genes with 2,099 bp and contains two exons according to EnsemblProtists. However, on the NCBI database, the partial mRNA corresponding to this gene begin with a start codon located 440 bp downstream of the start codon predicted by Ensemble protist, downstream of the predicted intron. Furthermore, this partial mRNA does not contain a large nucleotide sequence of 327 bp in the middle of the transcript, suggesting the presence of a second intron, and at least two possible start codons (**Fig. 4.2**).

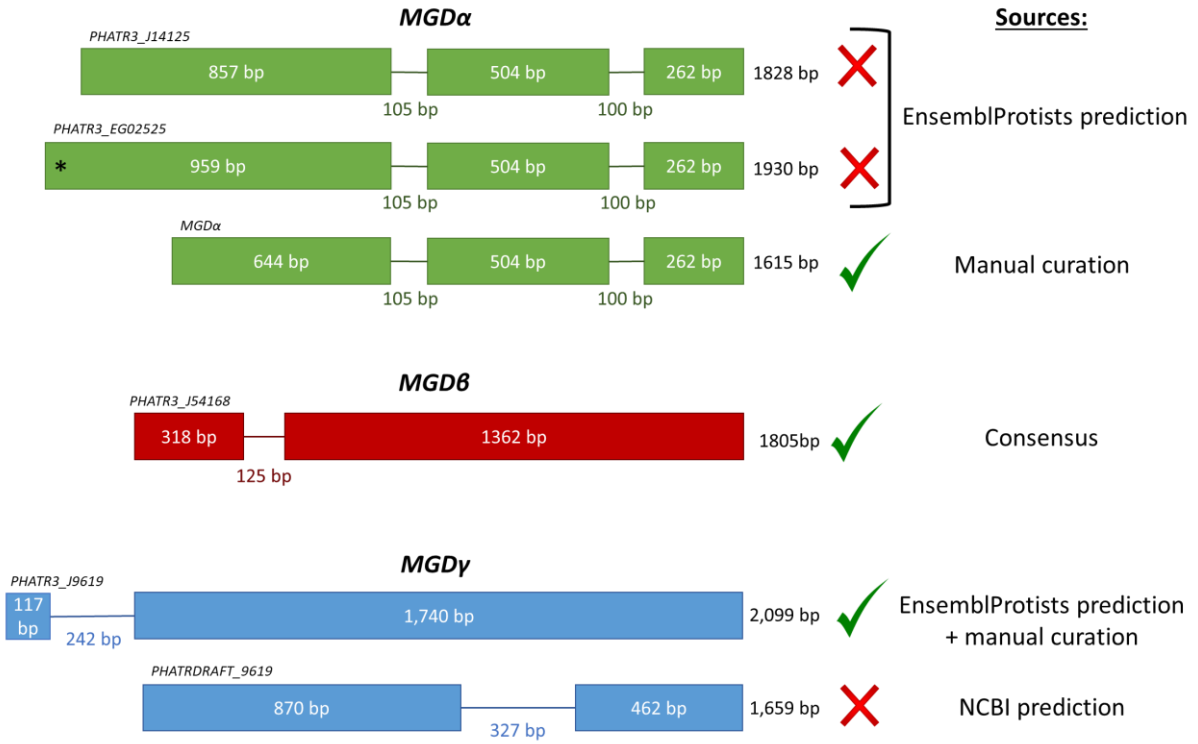


Figure 4.2: Gene structure analysis of *MGD* genes in *P. tricornutum*. Different gene structures are presented for each gene according to information extracted from NCBI and EnsemblProtists databases, and manual curation. Lines indicate introns and boxes indicate exons. Nucleotide length of each exon, intron and gene is indicated. Three gene structures are presented for *MGDα*, the first two structures (*PHATR3_J14125* and *PHATR3_EG02525*) derive from two possible CDSs retrieved from EnsemblProtists, while the third structure (*MGDα*) was generated following manual curation (**section 4.1.3.1.2**). The (*) indicate the localisation of the polymorphism mentioned in this section. For *MGDβ*, database information and personal results all pointed to the same gene structure. For *MGDγ*, two gene structures are presented as distinct CDSs were obtained from EnsemblProtists (*PHATR3_J9619*) and NBCI (*PHATRDRAFT_9619)*. Validation of one of these structures was obtained following cDNA sequencing.

Sequencing of *MGDα* and *MGDβ* cDNAs allowed us to validate the splicing of all predicted introns in the genes.

- For *MGDγ*, splicing of the first intron of 242 bp was confirmed. However, the nucleotide sequence corresponding to the second intron of 327 bp predicted with NCBI was present inside the cDNA sequence, revealing that this sequence was not an intron but part of the second exon. Furthermore, only one possible ATG codon is present upstream of the 242-bp intron, which corresponds to the start codon predicted by EnsemblProtists. Splicing of the intron thus validates that the gene begins with this first ATG codon.
- In the case of *MGDα*, cDNA included the start codon predicted for the CDS of *PHATR3_EG02525*. Surprisingly, a polymorphism was detected between the predicted start codons for *PHATR3_EG02525* and *PHATR3_J14125* (**Fig. 4.2**). The polymorphism originated from a region where there was either a repetition of three 'TC' (a thymine followed by a cytosine) or four 'TC' in the sequence. These two versions appear to be present at 50% each in the genome according to chromatogram analysis. The version containing four 'TC' is not in frame for protein translation. Therefore, it is unlikely that *MGDα* starts from the start codon predicted with *PHATR3_EG02525* CDS. A manual curation based on N-terminal addressing sequence, detailed below, allowed confirmation of the most likely start codon.

The length of the 5'-UTR was investigated by PCR amplification on cDNA. For *MGDα*, we could amplify the cDNA sequence until 591 bp before the ATG predicted for *PHATR3_J14125*. Therefore, the 5'-UTR region is at least ~600 bp-long. For *MGDβ*, the cDNA sequence could be amplified 219 bp upstream of the predicted start codon but not 460 bp upstream. Therefore, 5'-UTR measures between 220 and 460 bp. For *MGDγ*, the cDNA could be amplified 49 bp upstream of the predicted start codon of *PHATR3_J9619*, but not 247 bp upstream. Thus, this 5'-UTR is between 50 and 250 bp long. *MGDα* possesses the longest 5'-UTR, suggesting a possible control of this gene expression by multiple regulatory pathways.

Interestingly, MGDα and MGDγ show higher similarity to *A. thaliana* AtMGD1 than to AtMGD2 and AtMGD3, while MGDβ was equally similar to both AtMGD1 and AtMGD3 (**Fig. 4.3A**). MGDα was more similar to all AtMGDs, and in particular to AtMGD1, compared to MGDβ and MGDγ. Furthermore, MGDβ and MGDγ are slightly more similar to MGDα than to the AtMGDs (**Fig. 4.3**). This indicates than there seem to be no analogy between the three MGDs of *P. tricornutum* and the three MGDs of *A. thaliana*, as suggested by our phylogenetic analysis. MGDβ and MGDγ shared a high level of similarity (~50% identity) compared to that with MGDα (~35%) (**Fig. 4.3B**).

A.

	AtMGD1		AtMGD2		AtMGD3	
	Query cover	Per. Ident	Query cover	Per. Ident	Query cover	Per. Ident
MGDα	69%	40.00%	79%	36.43%	80%	37.25%
MGDβ	71%	32.38%	80%	30.67%	80%	32.89%
MGDγ	69%	35.09%	81%	33.74%	80%	32.83%

B.

	MGDα		MGDβ		MGDγ	
	Query cover	Per. Ident	Query cover	Per. Ident	Query cover	Per. Ident
MGDα			77%	35.27%	63%	33.80%
MGDβ	85%	35.02%			67%	53.4%
MGDγ	84%	37.93%	80%	49,89%		

Figure 4.3: Protein sequence similarities between MGDs from *P. tricornutum* and *A. thaliana*. Query cover percentage and percentage identity (Per. Ident) scores obtained for each *P. tricornutum* MGD (lines) using each *A. thaliana* MGD as queries (columns) (**A**), or using each MGD in *P. tricornutum* as queries (columns) (**B**), in a protein-protein Basic Local Alignment Search Tool on NCBI.

4.1.3. *Phaeodactylum* MGD protein structure analysis

4.1.3.1. Primary structure: identification of conserved domains and residues

AtMGD1 primary, secondary and tertiary structures have been well characterised following successful crystallisation of its catalytic domain (Rocha et al., 2016). Notably, catalytic residues have been structurally and functionally characterised as well as several residues binding to the substrate DAG, the sugar donor UDP-Gal, and the lipid activator PG. Figure 4.4 shows the sequence alignment of the three AtMGDs and the three *P. tricornutum* MGDs. The protein sequence of MGDα used for the alignment corresponds to the gene structure marked as validated in figure 4.2. Explanations for this choice will follow in section 4.1.3.1.2. Several small amino acid segments and isolated key residues show a high consensus level between these sequences.

4.1.3.1.1. Active site

Two catalytic residues have been identified in AtMGD1, one in each domain of the protein (Botté et al., 2005; Dubots et al., 2010; Rocha et al., 2016). The histidine (H) in position 159 of the alignment is involved in the deprotonation of the nucleophile hydroxyl (OH) group of the DAG substrate (**Fig. 4.4**). This residue is located in the first highly conserved region in proteins, which includes residues 148-170 of the alignment. The lysine (K) in position 576 of the alignment binds the sugar donor, stabilizing the phosphates of the departing UDP during the reaction. K576 is also found in a highly conserved region between proteins. The corresponding residues in the MGDs of *P. tricornutum* are easily recognized due to the high consensus between the sequences in these regions.

Presence of anionic lipids such as PA, PG and SQDG are mandatory for AtMGD1 activity. A PG-His catalytic dyad has even been proposed, that would form an acid/base relay system (Makshakova et al., 2020; Nitenberg et al., 2020). Such a catalytic dyad would facilitate the deprotonation of the acceptor substrate by H159.It was shown that PG develops interaction with the catalytic residue His159 and the neighbouring arginine R160, thus placing this lipid close to the active site (Makshakova et al., 2020). R160 is highly conserved in all sequences.

Other regions containing residues suspected to be involved in the binding of the substrate DAG and the activator PG in the N-domain of AtMGD1 were poorly conserved. However, the domain surrounding residue R576 and involved in UDP-Gal binding in the C-domain shows a much higher conservation level. Similar observations are often made with glycosyltransferases, and reflect the diversity of substrates by opposition to the similarities in donor sugar accommodation.

4.1.3.1.2. Prediction of protein localisation through identification of targeting sequences

Upstream of the first highly conserved domain that includes the catalytic His159, N-terminal sequences are highly variable. The N-terminal part of AtMGD1 contains a cTP allowing its localisation to the chloroplast. The N-terminal parts of varying length seen in *P. tricornutum* MGDs might thus correspond to targeting sequences as well, specific of this microorganism.

Protein targeting in complex plastid involves the succession of a signal peptide and a cTP, which is called a bipartite presequence. In MGDα, such a bipartite presequence could not be identified when considering the methionine start codon suggested by *Phatr3_J14125* CDS. However, a bipartite presequence was evident when considering the only other methionine located upstream of the first conserved region. Indeed, both a signal peptide and a cTP could be predicted using HECTAR server. We concluded that the protein sequence of MGDα most likely starts from the methionine from which a targeting sequence could be detected. Therefore, none of the CDSs predicted in EnsemblProtists start from the right start codon. Furthermore, this means that the gene 5'-UTR is actually at least 804 bp-long. This manual curation was used to reconstruct the gene structure shown in figure 4.2.

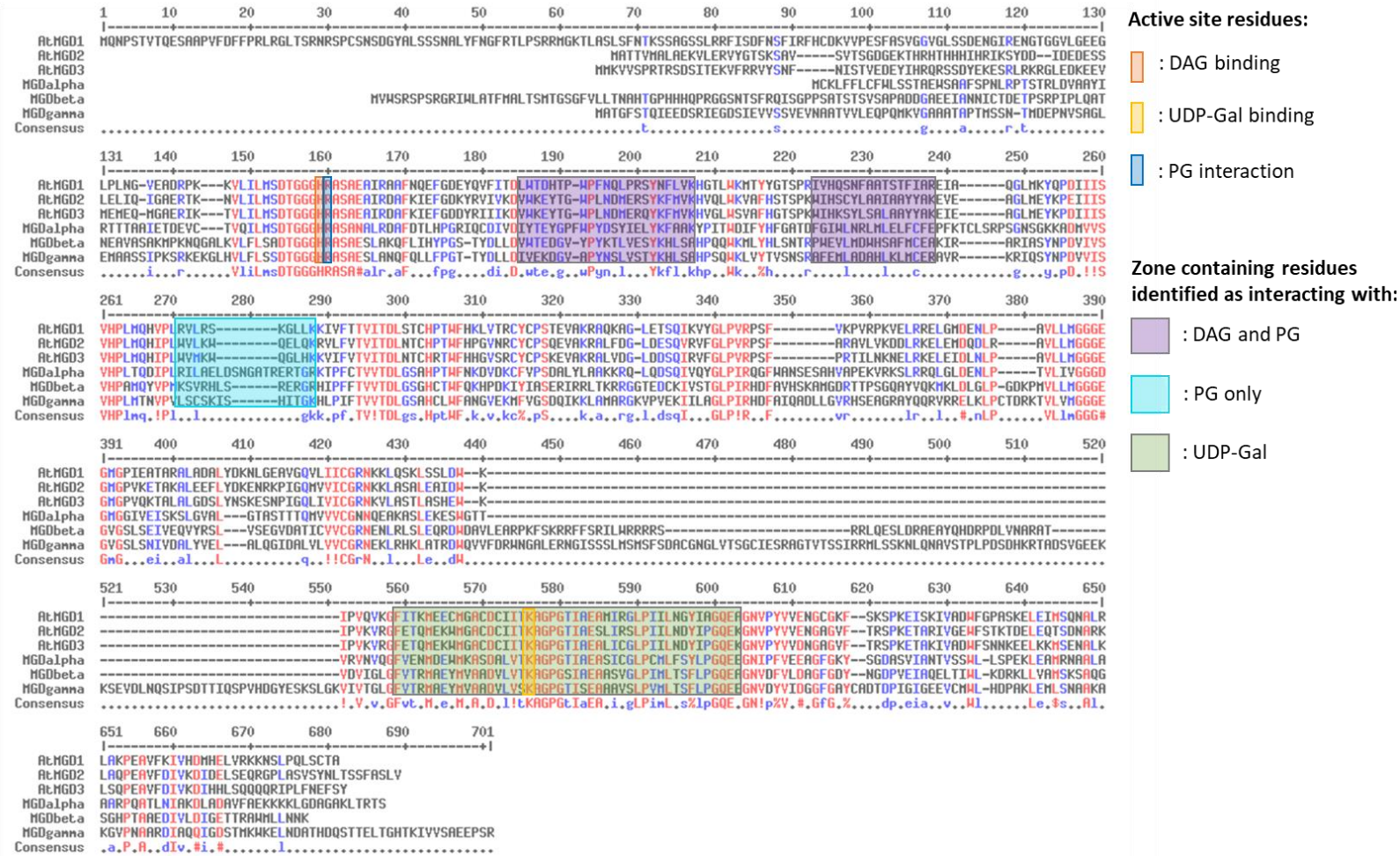


Figure 4.4: Multalin sequence alignment of AtMGDs and MGDs proteins. Comparison of the position of amino acids known to be involved in the interaction of AtMGD1 with its substrates and PG, with the amino acid sequences of *P. tricornutum* MGDs. Residue colours differentiate high (red) and medium (blue) conservation. The catalytic residues binding to the diacylglycerol and the UDP-Galactose are a histidine (orange square) in position 159 and a lysine (yellow square) in position 576 of the alignment, respectively. The residue interacting with PG and allowing a potential PG-His catalytic dyad is an arginine (dark blue square) in position 160 of the alignment. Zones containing a high number of residues in AtMGD1 interacting with DAG, PG and UDP-Galactose are indicated with purple squares when residues interact with both DAG and PG, with light blue squares when interacting with PG only, and with green squares when interacting with UDP-Galactose. Rocha et al., 2016, Makshakolva and Breton, 2020.

In MGDβ, four possible methionine start residues are located between the predicted methionine start of Phatr3_54168 and the first conserved region shown in the alignment. However, it is only when the protein sequence starts with the first methionine that we could obtain a low prediction (57.66 %) of a signal peptide with HECTAR. However, no cTP was predicted. Thus, we concluded that this first methionine was indeed the start methionine of MGDβ.

In MGDγ, we previously validated the first methionine as the methionine start based on the splicing of an intron placed before other possible methionine start. However, it should be noted that between the intron and the first conserved region of the protein, as many as four methionine residues are present in the long N-terminal part of MGDγ. No prediction of a bipartite presequence could be obtained using any of these methionine codons as a start. Thus, MGDγ protein does not appear to contain any predictable targeting sequence in its N-terminal part. Furthermore, it is highly probable that a protein starting from any of the four other methionine residues found after the intron may be functional and correctly localised.

4.1.3.2. Secondary and tertiary structures predictions

MGDs belong to the GT-B superfamily. The conserved GT-B fold consists of two Rossmann-type α/β/α domains of comparable size, named the N-domain and the C-domain, separated by a large cleft, and stabilized by two long C-terminal α-helices (Rocha et al., 2016). The N-domain in AtMGD1 correspond to positions 141-338 and 655-672, the C-domain to positions 362-654, and they are connected by a long loop corresponding to positions 339-361, in the alignment (**Fig. 4.4**). A large disordered region could not be crystalised in AtMGD1, corresponding to positions 186-235 (Rocha et al., 2016).

Secondary and tertiary structures of *P. tricornutum* MGDs were predicted using Phyre² (Protein Homology/analogY Recognition Engine V 2.0) web portal (Kelley et al., 2015b). All *P. tricornutum* MGD models retrieved from Phyre² correspond to the 3D model generated with the first hit, which in each case corresponds to the crystal structure of AtMGD1 in complex with UDP (template C4X1TA). The protein models obtained were compared to AtMGD1 structure (**Fig. 4.5**).

All *P. tricornutum* MGD protein structures show the same GT-B fold as in AtMGD1 crystal structure. All models start approximately from position 145 of the alignment (**Fig. 4.4**) and finishes around position 670.

The size of the N-domains is very similar (~180 amino acids) between MGDα and AtMGD1. However, the N-domains of MGDβ and MGDγ contain ~80 additional amino acids, respectively, upstream of the first conserved domain with MGDα and downstream of a signal peptide in the case of MGDβ. These additional amino acids share a low conservation level between the two proteins, and are absent from the models. In the models, the N-domain of MGDα appears distinct from the other due to the presence of a longer alpha helix protruding from the N-domain (**Fig. 4.5B**), which contains 18 amino acids against 10 in AtMGD1. All N-domains lack the portion corresponding to the disordered loop in AtMGD1 (~65 amino acids in *P. tricornutum* MGDs).

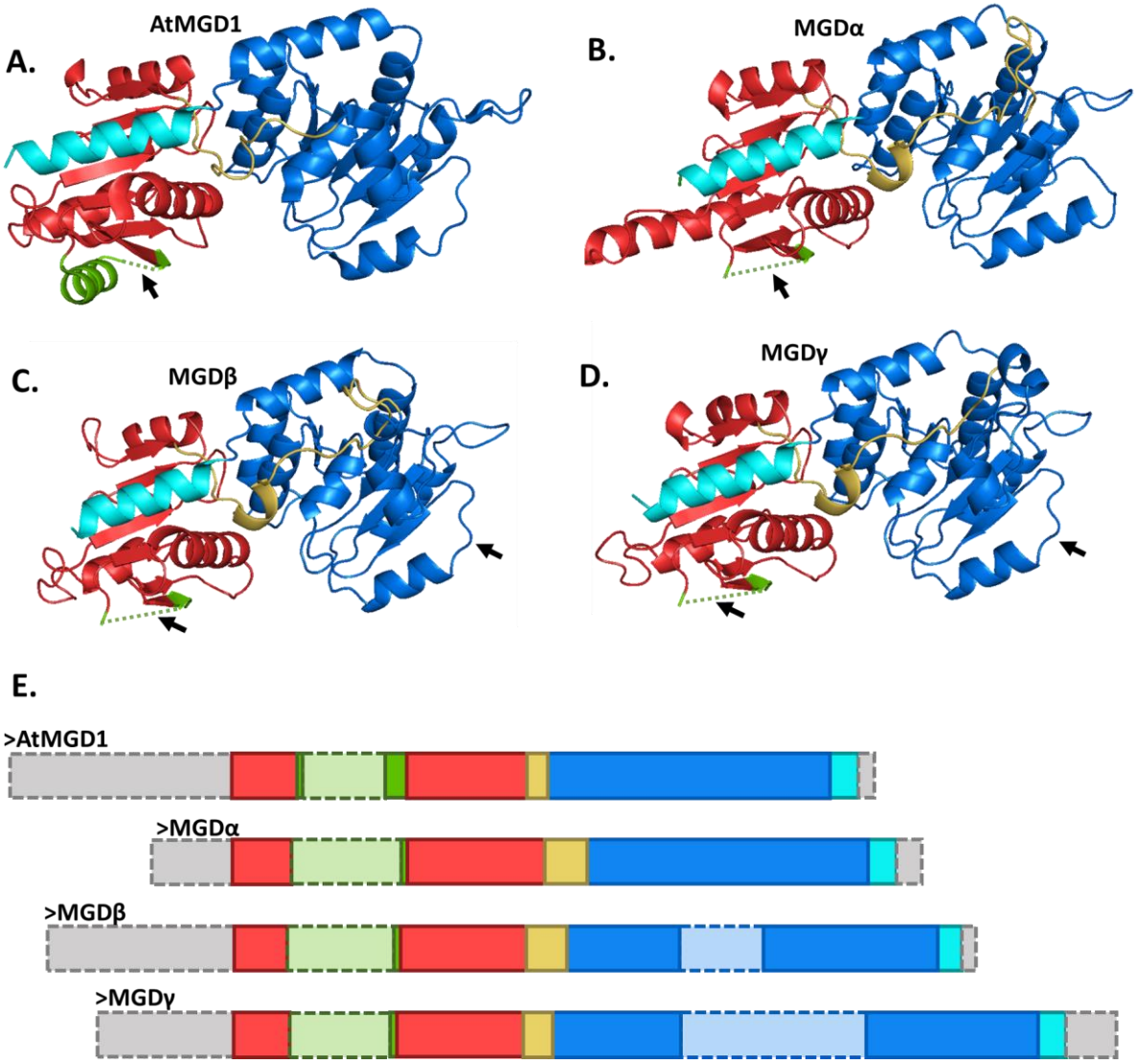


Figure 4.5: Protein models of *P. tricornutum* MGDs. Protein structures of AtMGD1 (A), MGDα (B), MGDβ (C) and MGDγ (D), and linear schematisations of protein domains (E). Models were viewed with PyMOL software using the cartoon representation in the same orientation. The N-domain is coloured in red except for the C-terminal helix coloured in cyan and the large disordered loop coloured in green. The C-domain is coloured in blue. Residues corresponding to the loop connecting both domains are coloured in yellow. Black arrows indicate the localisation of missing amino acids in the models. The domains that were absent from the models were represented with dashed outlines in the schematisation (E).

The C-domains are of similar size between AtMGD1 and MGDα (~170 amino acids). However, they are bigger in MGDβ (~223 amino acids) and MGDγ (~290 amino acids). The amino acid sequences responsible for the longer C-domains of MGDβ and MGDγ are absent from the models (**Fig. 4.5E**). Therefore, the predictions do not allow the modelling for substantial parts of the C-domains, which in the case of MGDγ represents a third of the C-domain.

We can also observe that the loop connecting both domains is longer in *P. tricornutum* MGDs, and is predicted to contain a small helix structure in the middle.

Given that C-domains show the highest conservation level in the amino acid sequence, the long additional stretches of residues present in MGDβ and MGDγ are very intriguing. If we compare MGDβ and MGDγ with MGDs

from *F. solaris*, the corresponding stretches are present in MGDβ-like and MGDγ-like proteins. When providing these sequences alone to Phyre², low confidence predictions of helix structures are returned. Alignment of these two sequences alone shows that the C-terminal part of the MGDγ extension is weakly similar to that of MGDβ. Furthermore, the hydrophobicity plots for these sequences suggests that they are hydrophilic, especially in the potentially common region (**Fig. 4.6**).

In conclusion, while the general structure of *P. tricornutum* MGDs is similar to that of AtMGD1, the size of the N- and C-domains of MGDβ and MGDγ is considerably larger. Interestingly, although the additional segments in MGDβ and MGDγ occur at the same locations in the proteins, they are only weakly similar.

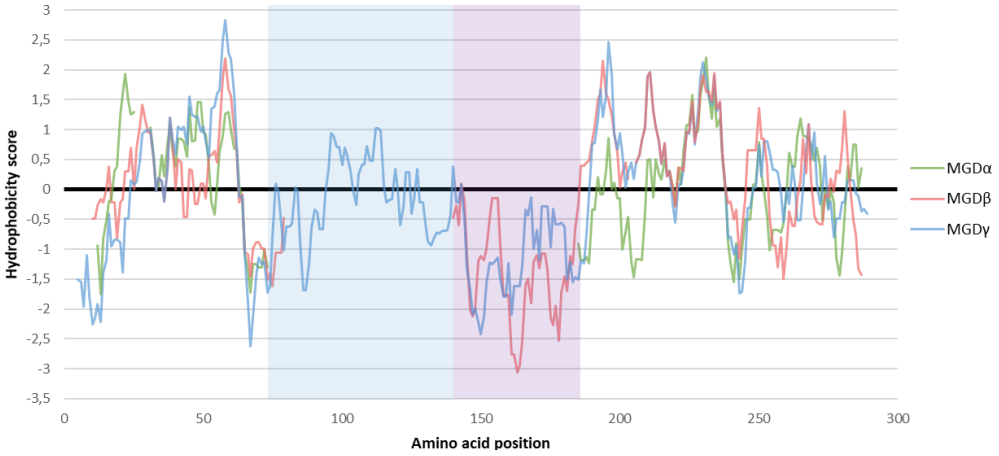


Figure 4.6: Hydrophobicity plot of the C-domain of *P. tricornutum* MGDs. Hydrophobicity plot was acquired with Expasy, using the hydrophobic scale of Kyte and Doolittle. The three plots were aligned so as to match protein sequence alignment.

4.2. Study of *Phaeodactylum* MGDα, MGDβ and MGDγ subcellular localisation by confocal imaging of eGFP-fused proteins

P. tricornutum strains overexpressing each *MGD* genes fused to eGFP were kindly provided by Hanhua Hu (Laboratory of Algal Biology, Institute of Hydrobiology, Chinese Academy of Sciences, Wuhan 430072, China) and Yangmin Gong (Oil Crops Research Institute of Chinese Academy of Agricultural Sciences, Wuhan 430062, China). For each MGD isoform, two overexpressing strains were provided: MGDα-eGFP-A and MGDα-eGFP-B, MGDβ-eGFP-A and MGDβ-eGFP-B, and MGDγ-eGFP-A and MGDγ-eGFP-B, overexpressing MGDα, MGDβ, and MGDγ, respectively.

Observation of these strains with confocal microscopy revealed distinct subcellular localisations of MGD isoforms in *P. tricornutum* (**Fig. 4.7**). We selected one transgenic strain for each MGD isoform and a representative observation for figure 4.7. In the range of detection used for observation of eGFP signal, chlorophyll auto fluorescence could also be observed as seen in the WT with a laser power of 5 %. However, eGFP signal clearly distinguished itself from chlorophyll auto fluorescence in the overexpressing strains.

With MGDα-eGFP, we obtained a clear intraplastidial localisation signal visible as spots in the plastid. These spots could correspond to a localisation in the thylakoids either in particular sub-domains or results from protein aggregation following overexpression. Alternatively, the spots could represent a localisation in the plastoglobules.

MGDβ-eGFP signal localises to the blob-like structure. The vesicular network inside the blob shows continuity with the PPM but not the oEM (Flori et al., 2016). Therefore, we suggest that MGDβ localisation might extend to the PPM.

MGDγ-eGFP signal appeared in the ER. This is intriguing since MGDs are usually found in the plastid, although AtMGD2 was recently found to localise to the cytosol of elongating pollen tubes in the absence of any plastid (Billey et al., 2021). Given that the ER is continuous with the EpM, MGDγ could probably also bind to the external membrane of the plastid. Furthermore, our collaborators Hanhua Hu and Yangmin Gong, using the same strains, had observed the MGDγ-eGFP in the cytosol. This could either be an artifact of the overexpressing system, or the sign that MGDγ is able to solubilise in the cytosol. Additionally, the extra amino acid sequence present in the C-domain of MGDγ is particularly rich in hydrophilic residues, potentially facilitating its solubilisation.

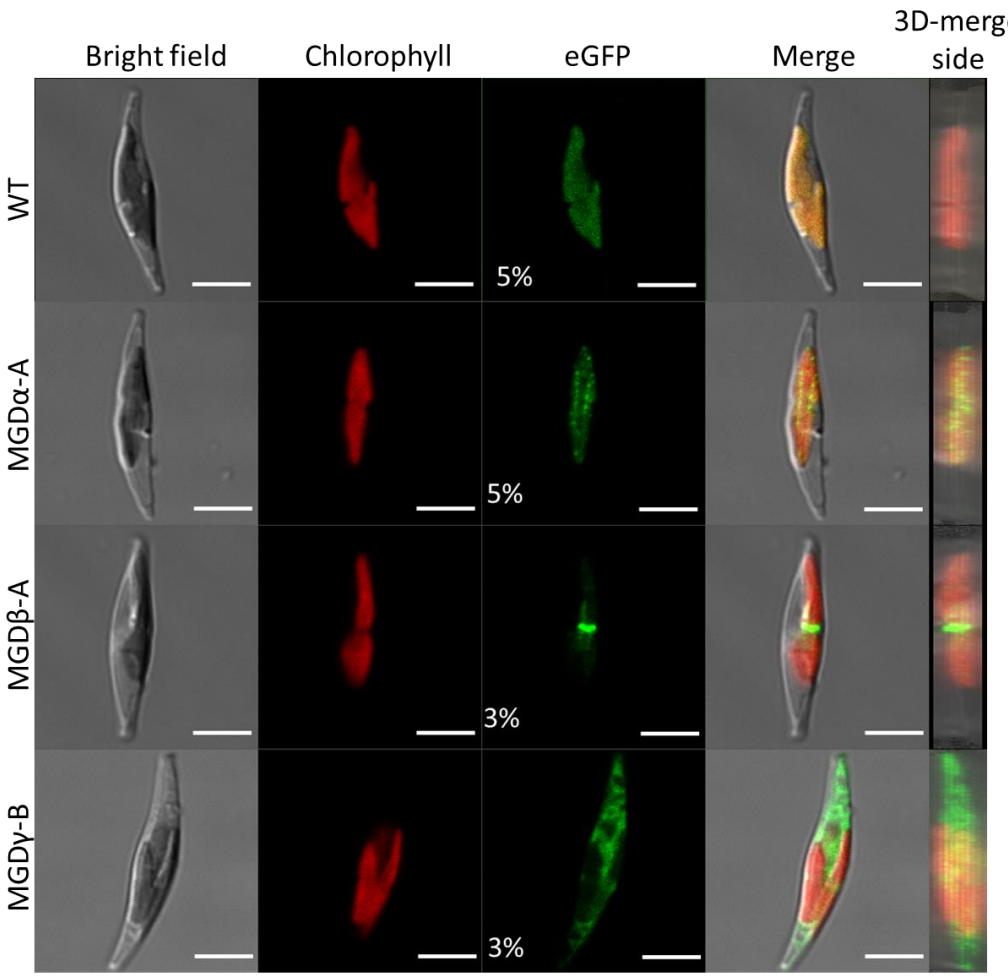


Figure 4.7: *In vivo* subcellular localisation of MGD isoforms in *P. tricornutum*. Detection of MGD-eGFP genes expression in transgenic *P. tricornutum* strains with wild type as a negative control. For each cell, bright field, chlorophyll fluorescence and eGFP signal are represented, with a merge of all images. eGFP excitation was realised at either 5 % or 3 % of laser power, as indicated in the pictures. In addition, we show the side of the cell after generation of a 3D-image from a Z-stack. An intraplastidial localisation was obtain when MGDα-eGFP is expressed. With MGDβ-eGFP, we recognize the eGFP signal in the blob-structure, indicating a localisation of MGDβ in the PPM. MGDγ-eGFP localises in the cytosol, suggesting that MGDγ might localise to both the cytosol and the EpM. Scale bar: 5 µm.

MGDβ corresponds therefore to a marker of the blob-like structure (**Fig. 4.8A**). Interestingly, the signal always appeared in the middle of plastids. The blob-like structure was most clearly observed in dividing plastids, which would be coherent with a hypothetical higher abundance of MGDβ level during division. In a plastid in division, the blob was localised at the interface between the two forming plastids. Then, the blob appeared as a link between the two generated plastid inside the dividing cell. This observation was supported with TEM images (**Fig. 4.8B and Fig.4.8C**). After separation of the daughter cells, blob structures appeared facing each other. These observations were integrated in the schematic representation drawn in figure 2.4. When MGDβ was observed in a plastid that is not in division, the signal was much more diffused. Thus, it seems that the blob-like structure plays an important role during cytokinesis, and that this structure is scattered in non-dividing cells.

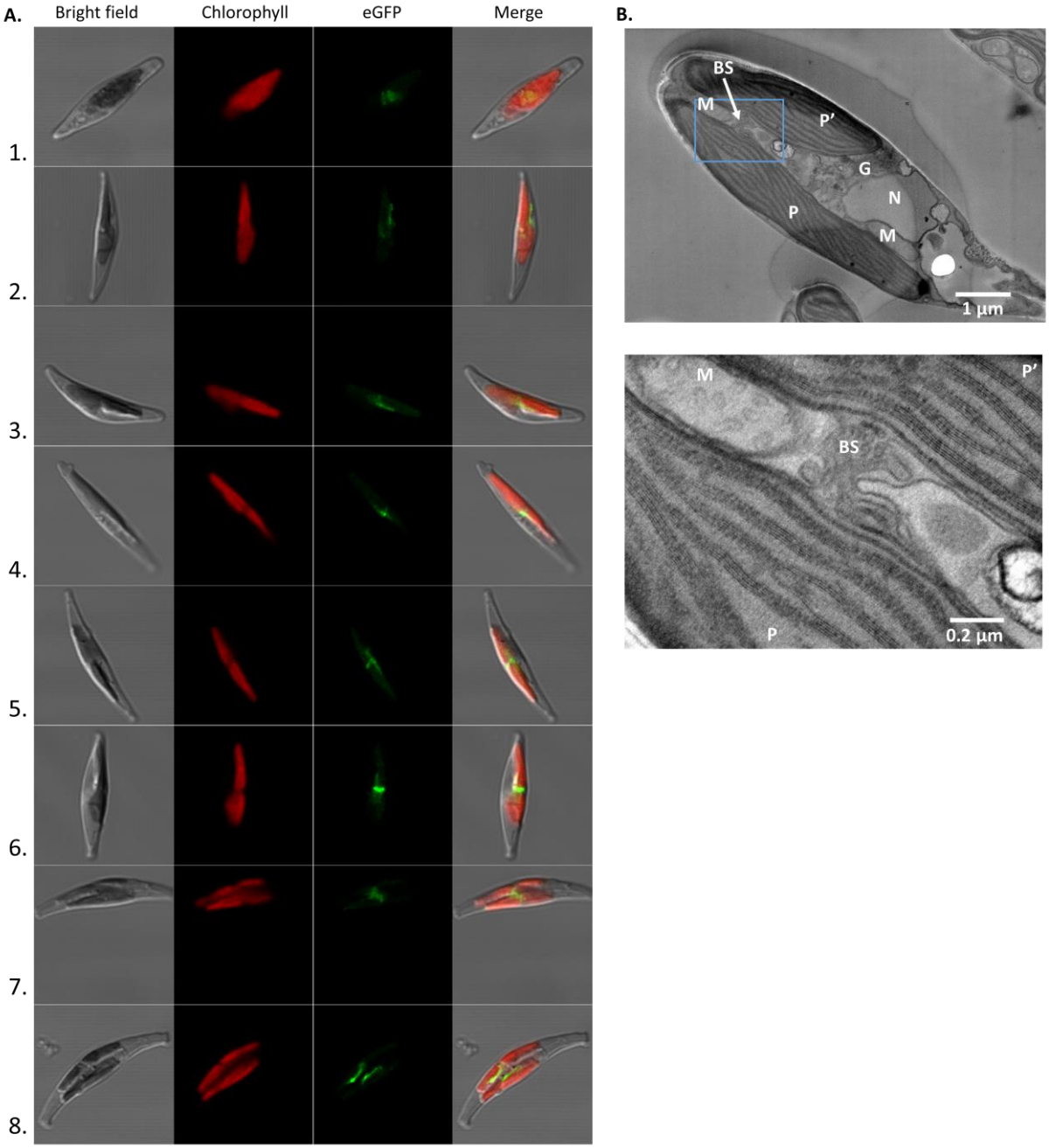


Figure 4.8: *In vivo* subcellular localisation of blob-like structure. Observation of blob-like structure localisation by confocal microscopy (A) and transmission electron microscopy (B). *P. tricornutum* transgenic strains overexpressing MGDβ-eGFP gene were used to monitor the appearance of blob-like structures in the cell (A). Cells 1 and 2 show a scattered eGFP signal, cells 3-6 show a blob-like structure at the central constriction of dividing plastids, cell 7 shows a blob-like structure extending from one plastid to the other inside a dividing cell, and cell 8 shows two blob-like structures facing each other on distinct plastids. In cell 8, it is not clear if the furrows have divided the cell completely or not. Observation of a wild type *P. tricornutum* cell containing two plastids following plastid division inside the cell (B). Cleavage furrows are visible at both end of the cell. A structure resembling to the blob-like structure was observed connecting the two plastids. BS, Blob-like structure; G, Golgi apparatus; M, mitochondrion; N, nucleus; P and P', plastids.

4.3. Generation of protein tools for the study of *P. tricornutum* MGD enzymes

At first, we sought to validate each MGD isoform activity, be able to detect the proteins level of each isoform in mutant lines, and potentially perform immunolocalisation experiments to complement recombinant protein localisation. To this end, we produced pure recombinant proteins as well as specific antibodies for each MGD isoform.

4.3.1. Recombinant protein production

Production of each MGD isoform was performed as described in the Material and Methods at the Cell Free expression platform of ISBG (Integrated Biology Structural Grenoble) at IBS (Institut de Biologie Structurale, Grenoble). The cell-free system was chosen for several reasons. First, it allows the overexpression of a protein without the drawbacks of fast protein degradation or cytotoxicity issues common in a cell system. Protein production can be done in a single tube, saving time and efforts. For our membrane proteins, detergent can be directly added into the expression system to prevent protein precipitation. Furthermore, it allowed us to perform a large-scale production (milligrams quantities) at low cost.

The design of each recombinant MGD isoform expression was based on the previous work on AtMGD1 (Botté et al., 2005; Dubots et al., 2010; Rocha et al., 2013). In AtMGD1, the primary sequence includes a canonical N-terminal chloroplast transit peptide followed by a short stretch of residues that represent the first 136 amino acids of the protein, prior to the catalytic domain (amino acids 137-533) (Botté et al., 2005; Dubots et al., 2010). Therefore, most studies expressed AtMGD1Δ137, a truncated version of the protein representing the catalytic domain, for enzyme activity tests and crystallisation studies (Botté et al., 2005; Dubots et al., 2010; Rocha et al., 2013).

Based on a protein alignment between AtMGD1Δ137 and each *P. tricornutum* MGD, we decided to express a similar His-tagged truncated versions of each MGD isoform: MGDαΔ42- His_6, MGDβΔ95-His_6 and MGDγΔ62-His_6. The truncated genes sequences were amplified from cDNA and cloned into an appropriate expression vector. It was previously demonstrated that *E. coli* was an appropriate host for the expression of AtMGD1Δ137 (Rocha et al., 2013), therefore, we used the biomolecular translation machinery from E. coli host background in our cell-free system.

To verify protein expression and ensure optimal expression conditions of the membrane proteins, different detergents and magnesium acetate concentrations were tested prior to large scale production. During these preliminary tests, MGDαΔ42-His_6 could not be produced, as it was undetected neither in a solubilised nor precipitated form. MGDβΔ95-His_6 and MGDγΔ62-His_6 could be produced, and addition of detergent coherently increased protein solubility.

Large scale production of MGDβΔ95-His_6 and MGDγΔ62-His_6 in 9-mL reaction volumes was successfully achieved at the Cell Free expression platform under the supervision of a qualified engineer, Lionel Imbert, in RNAse free conditions. Purifications of 3 mL and then 6 mL of the reaction mixes were realised using a his-tagged protein purification resin, under the guidance of Emmanuel Thevenon (team FLO_RE, LPCV, Grenoble) (**Fig. 4.9**). Protein

concentration was evaluated in the most concentrated 1-mL elution fractions (based on gel electrophoresis), by Bradford assay. Highest concentrations obtained for MGDβΔ95-His_6 and MGDγΔ62-His_6 were 0.306 $g.L^{-1}$ and 1.248 $g.L^{-1}$, respectively.

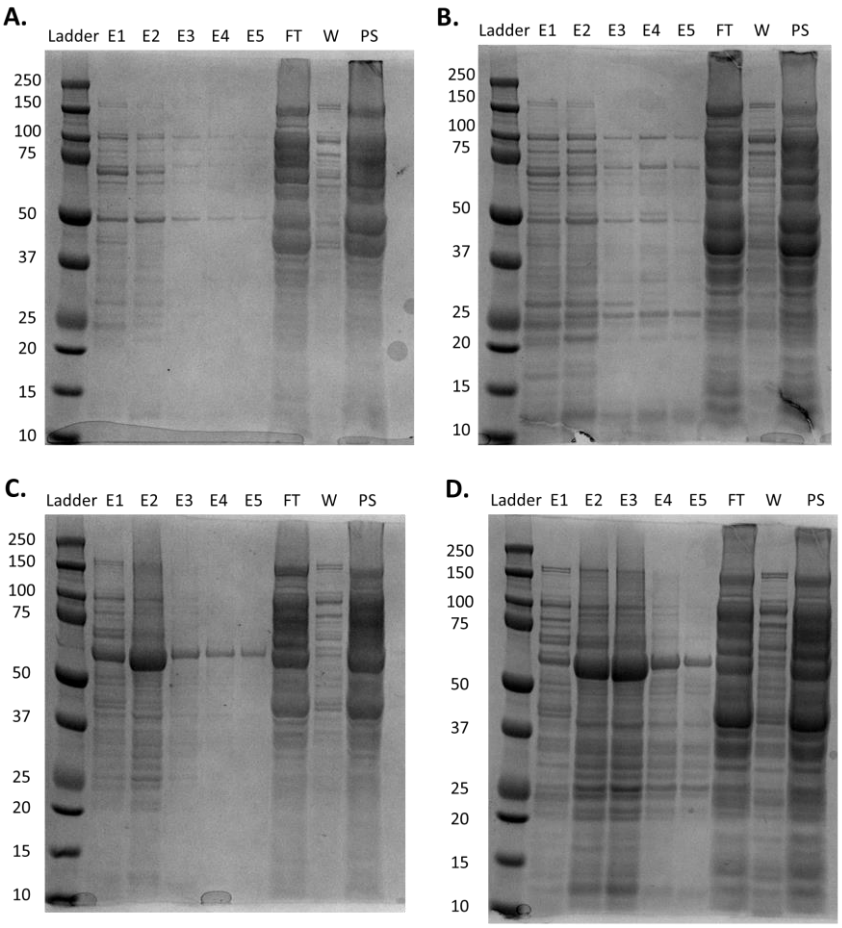


Figure 4.9: Purification of recombinant MGD proteins. Observation of each elution fraction obtained following protein purification of MGDβΔ95-His_6 (A, B) and MGDγΔ62-His_6 (C, D), by gel electrophoresis. Purification was first performed on a 3-mL aliquot (A, C), then a 6-mL aliquot (B, D). Expected size for MGDβΔ95-His_6 and MGDγΔ62-His_6 were 53 and 62 kDa, respectively. Matching bands were visible in the electrophoresis gels, as well as many proteins associated with the cell-free system. E: Elution; FT: Flow-through, W: Wash; PS: Prepared sample (before purification). Protein ladder is in kDa.

4.3.2. Antibodies production

Antibody production was performed in collaboration with the Agro-bio company (Stago group) as described in the materials and methods. Antibodies efficiency were tested by Western blot on total protein extracts of *P. tricornutum* WT and MGD mutant strains, and on the purified recombinant proteins (**Fig. 4.10**).

Overall, the quantity of protein extract used was insufficient to detect native MGD expression in the WT, and many unspecific bands were detected (**Fig. 4.10**). For MGDα antibody, a band at around 48.9 kDa was expected. However, we detected an unspecific band at around 50 kDa in both the WT and the mutant strains. Therefore, although we could not verify if the antibody could detect MGDα, presence of this unspecific band makes this antibody useless for the detection of MGDα in total protein extract of *P. tricornutum*.

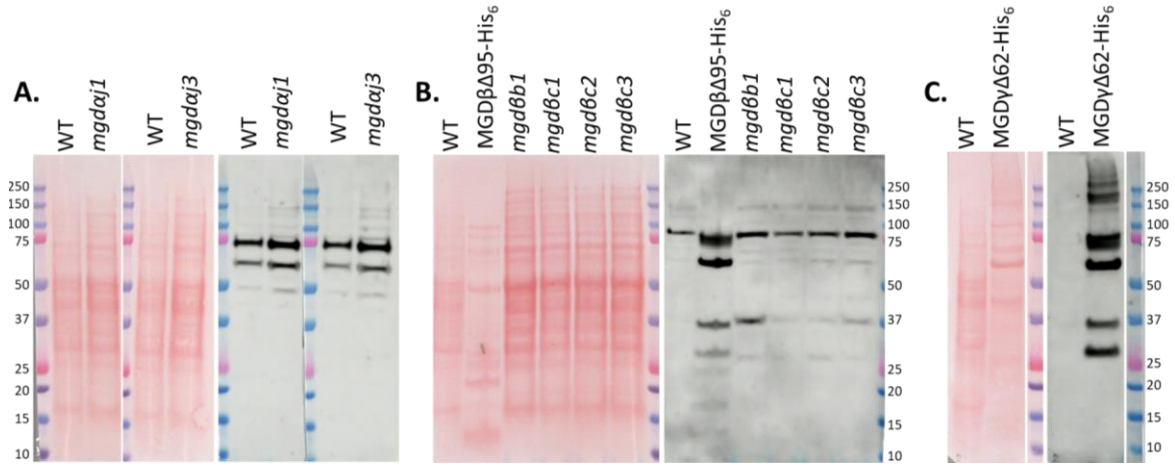


Figure 4.10: Western blot analysis of antibody efficiency. Antibodies for MGDα (A), MGDβ (B) and MGDγ (C) were tested on 8 µg of total protein extracts of the WT and a selection of MGD mutant strains, as well as on the produced recombinant proteins. Red ponceau images and antibody revelation images are presented.

For MGDβ antibody, we expected a band at around 53 kDa in the recombinant protein. However, no such band was detected. In the WT and the *mgdβc2* mutant only, we expected a band at around 58.9 kDa. However, we did not detect a band at this size in the WT, and a weak band could be detected in all MGDβ knockout mutants and *mgdβc2*. Therefore, we concluded that the antibody was not specific of MGDβ.

For MGDγ however, the expected band size was detected for MGDγΔ62-His_6 (62 kDa). In the WT we should have seen a band at 66.5 kDa, but it could not be visible due to the weak expression of the protein. Therefore, MGDγ antibody was successfully obtained.

4.3.3. Activity test

In order to validate the MGDG synthase activity of the MGD isoforms, we performed a UDP-Glo Glycosyltransferase Assay using MGDβΔ95-His_6 and MGDγΔ62-His_6 (see Materials and Methods). The experiment consisted in loading the enzyme on prepared micelles made of the DAG substrate and an anionic lipid for MGD activation. Then, UDP-Galactose was added to the reaction, allowing the MGDG synthase to use both DAG and UDP-Galactose to produced MGDG and the by-product UDP. The reaction was then stopped, and the UDP production, proportional to the glycosyltransferase activity, was measured (see Materials and Methods). AtMGD1 catalytic domain AtMGD1Δ137 was used as a positive control.

In our first experiment, DAG 18:1-18:1 and PG 16:0-18:1 were used, and the MGDs were incubated for 10 minutes with their substrates. No glycosyltransferase activity could be detected with MGDβΔ95-His_6 and MGDγΔ62-His_6 in these conditions. We repeated this experiment a second time, increasing the incubation time from 10 minutes to 1 hour with similar results. Then, we changed the substrate and used the DAG 16:1-16:1, which is one of the major substrates of MGDs isoforms in *P. tricornutum* (seen later in this chapter). We also used SQDG extracted from *P. tricornutum* as anionic lipid activator instead of PG. Again, no activity was detected with our recombinant proteins while the control with AtMGD1Δ137 was positive, indicating that this enzyme can be activated with SQDG from *P. tricornutum* (data not shown).

In conclusion, we could not detect MGD activity using MGDβΔ95-His_6 and MGDγΔ62-His_6 with the UDP-Glo Glycosyltransferase Assay. Either some parameters of this test still need optimisation for *P. tricornutum* system, more appropriate DAG molecular species should be used as substrate, or the recombinant proteins were unfortunately not enough stable. Indeed, MGDβΔ95-His_6 and MGDγΔ62-His_6 could be misfolded and incapable of enzyme activity, or have precipitated following a long storage. Future experiments with other expression system and optimised conditions could therefore be considered.

4.4. Functional study of *Phaeodactylum* MGDs based on a functional genomic strategy

4.4.1. Generation of MGD knock-out and overexpressing lines

4.4.1.1. Methodology for the selection of interesting *Phaeodactylum* lines

4.4.1.1.1. CRISPR-Cas9 mutants

Our main strategy to functionally characterise MGD isoforms in *P. tricornutum* was to generate knockout lines using CRISPR-Cas9 technology. To this end, we initially designed single guide RNAs (sgRNA) directing the Cas9 to the predicted start codons of each *MGD* genes, and obtain mutations inducing frameshift at the beginning of the gene sequences. However, this first design was done before the study of the correct start codons. Therefore, the first sgRNAs targeted the predicted start codon of *Phatr3_J14125*, *Phatr3_J9619*, and *Phatr3_J54168* CDSs present in EnsemblProtists.

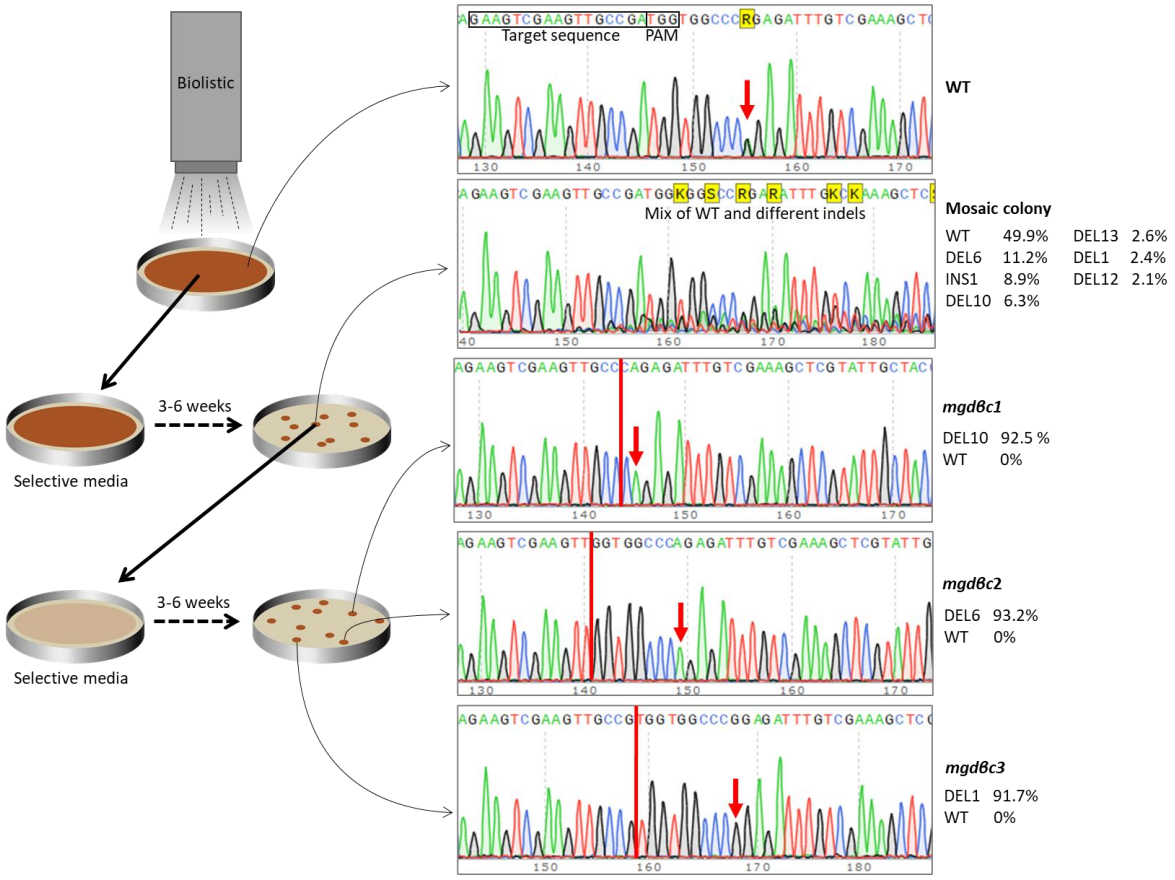


Figure 4.11: Methodology for the obtention of pure mutant strains. Cells were transferred to a selective media three days after transformation. Only transformant cells survive and form colonies. Colonies were visible after three to six weeks. PCR on colony was performed: Around 500 bp surrounding the sequence targeted by the sgRNA were amplified and sequenced. Chromatograms obtained during the isolation of MGDβ mutants edited with the sgRNA '168c' are presented as an example. Here, chromatograms were read by TIDE to assess the composition of the mutation profiles. The first transformant cells usually led to the formation of mosaic colonies. The equivalent of 200 cells from a mosaic colony containing interesting indels was spread on a new selective medium in order to separate cells and obtain colonies from single mutant cells. A colony was considered pure if more than 90% of its mutation profile was represented by one indel. Vertical red lines indicate the position of the deletions in the example. Red arrows point at the location of a single nucleotide polymorphism present in the WT and lost in the isolated mutants.

For MGDα, the mutations obtained were located 200 bp before the real start codon, and the mutant strains did not have an altered lipid profile. For MGDγ, the mutant strains obtained had a normal lipid profile as well. We therefore considered that MGDγ translation could 'leak' from another ATG in phase located downstream of the start codon. Following comparison of MGD isoforms with AtMGD1, we designed new sgRNAs for MGDα and MGDγ targeting the catalytic histidine residue.

CRISPR constructs were transformed in *P. tricornutum* cells using biolistic transformation. The wild type cells to transform were spread on plates a day before the transformation. The CRISPR constructs coated on tungsten beads were bombarded directly on the plates. Cells were allowed to rest for three days after transformation before transfer to selective media. We usually observed the appearance of colonies of transformed cells after three to six weeks on selective media.

Not all colonies are expected to contain mutations. Indeed, biolistic transformation is known to possibly cause some fragmentation and incomplete incorporation of the Cas9 gene and sgRNA, preventing CRISPR mutagenesis (Jacobs et al., 2015).

Our objective was to study mutant strains containing a pure mutation profile. However, transformed colonies usually contain a variety of insertion/deletions (indels) at the target site (**Fig. 4.11**). These colonies are called 'mosaic' colonies. This phenomenon is due to the fact that a transformed cell might divide several times before an indel is generated at the target site. Therefore, several cells in a colony can end up with different independent indels that will be present in different proportions in a colony.

Our strategy was to analyse the mutation profile of each colony by sequencing to identify promising mutant strains (**Fig. 4.11**). Tools such as ICE (Inference of CRISPR Edits; https://ice.synthego.com/) and TIDE (Tracking of Indels by Decomposition; https://tide.nki.nl/) can identify the mutation and their corresponding levels from the chromatogram of a sequenced mosaic colony. Then we spread around 200 cells of a selected colony onto a new selective medium in order to obtain new colonies originating from a single mutant cell. The new colonies were analysed to verify the mutation profile. This screening process was repeated until colonies with a pure mutation profile were obtained, and this process could last from one to three months.

P. tricornutum cells are diploid. Therefore, we could expect to only isolate colonies with two types of indels, resulting from the separate edition of each allele (biallelic heterozygous mutations). However, this was only rarely observed and most of our isolated mutant strains had a homozygous mutation (**Fig. 4.11**). Indeed, it has been observed that loss of heterozygosity, over a region that can range from 20 kb to an entire chromosome, can occur following CRISPR editing (Russo et al., 2018). In the chromatograms used as example in figure 4.11, an SNP present in the WT sequence is lost in the mutants. This base that was either a guanine or an alanine became an alanine in mutants *mgdβc1* and *mgdβc2*, and a guanine in the mutant *mgdβc3*. This further indicate that loss of heterozygosity occurred in these strains. Such an important change in the genome must be kept in mind when comparing mutant strains against the WT.

Usually, deletions (DEL) ranging from 1 to 20 bp were obtained, while insertions (INS) were limited to 1 to 4 bp. However, very large deletions were detected several times, such as a deletion of 280 residues (DEL280), a DEL468

and even a DEL1116. Two large insertions were also detected. An INS145, corresponding to the insertion of a sequence originating from *E. coli* genome, and a INS169, corresponding to a fragment of the CRISPR vector. Presence of a fragment of *E. coli* genome is probably due to the fact that CRISPR vector was harvested from an *E. coli* culture for vector amplification. The insertion of a vector fragment probably results from vector fragmentation following biolistic transformation.

4.4.1.1.2. Overexpression of optimised MGDs

We wanted to generate *MGD* overexpressing constructions in order to obtain our own *MGD* overexpressing strains, localise the proteins with an eGFP fusion, and complement our knockout mutant strains. The overexpression vector pPha-T1_eGFP_Blasti was kindly provided by Cécile Giustini (LPCV, Grenoble). Overexpression is driven by the fcpA promoter, and position of restriction enzyme sites allow the insertion of a gene with or without fusion with eGFP.

In the knockout strains, the CRISPR-Cas9 machinery is still present and would target the *MGD* vector if we attempt to complement the mutant strains without modifying the transgene. Therefore, we decided to have the gene sequences synthetized with specific point mutations to disable the recognition of the sgRNA complementary sequence. We took the advantage of synthesizing the genes to optimise the sequence using a codon usage table for *P. tricornutum*, kindly provided by Riccardo Aiese Cigliano (Sequentia, Barcelona).

In order to realise the insertion of the modified *MGD* sequences in the overexpressing vector using restriction enzymes, we also verified that no enzyme restriction sites were present in the genes, and modified the nucleotide sequence where sites were located.

After optimisation of the *MGD* sequences, we thus replaced some of the optimised codons with the next most used codons for the corresponding amino acids in order to prevent the binding of sgRNAs and restriction enzymes without altering the protein sequence.

We successfully obtained constructs for the overexpression of each optimised *MGD* isoforms, either fused to eGFP and His tag (*MGDopt-eGFP-His6*), or only to the His tag (*MGDopt-His6*). The *MGDopt-eGFP-His6* constructs were generated for protein localisation studies only, while the *MGDopt-His6* constructs were intended for the generation of overexpressing strains and the complementation of knockout lines.

Successful acquisition of transformants was achieved for *MGDαopt-His6*, and *MGDβopt-His6* but not for *MGDγopt-His6*. Unfortunately, due to the sanitary situation in 2020-2021, we did not have time to verify the expression level of the transgenes and study the lipid profiles. We did not have enough time either to attempt mutant complementation and acquisition of *MGDopt-eGFP-His6* overexpressing strains.

4.4.1.2. Knock-out and overexpressing lines

4.4.1.2.1. Mutation profile of selected knock-out lines

The use of CRISPR-Cas9 editing coupled with biolistic transformation may lead to adverse side-effects of selected lines. Indeed, biolistic transformation leads to the random integration of the vector DNA in the algae genome. Furthermore, constitutive expression of the Cas9 may lead to off-target mutations. Therefore, for each *MGD*

isoform, we tried to select several mutant strains obtained from different sgRNAs (distinct potential off-targets), and not originating from the same initial transformant colony (distinct vector insertion in the genome) (Table 3.2).

For *MGDα*, we selected one mutant obtained with the sgRNA 125i (*mgdαi1*), and three independent mutants obtained with the sgRNA 125j (*mgdαj1*, *mgdαj2* and *mgdαj3*) (**Fig. 4.12**). Interestingly, the mutants obtained with sgRNA 125j had a pure mutation profile at the first screening step and therefore do not derive from an initial mosaic colony. *Mgdαj1* and *mgdαj2* mutants had the exact same deletion, and later appeared to have extremely similar phenotypes. It is therefore highly probable that they derive from the division of the initial transformant cell prior to transfer to selective media.

For *MGDβ*, we selected one mutant obtained with the sgRNA 168b (*mgdβb1*), and three mutants obtained with the sgRNA 168c (*mgdβc1*, *mgdβc2* and *mgdβc3*) (**Fig. 4.12**). The mutants obtained with sgRNA 168c correspond to the example shown in figure 4.11. They derive from the same initial transformed cell, and thus harbour the same vector insertion. However, *mgdβc2* mutant has a DEL6, which does not generate a frameshift in the sequence (**Fig. 4.12**). This DEL6 leads to the deletion of two amino acids in the sequence (in position 94 and 95 of the alignment in figure 4.4) located downstream of the predicted signal peptide and upstream of the first conserved region between proteins. Therefore, this deletion is expected to be a silent mutation and this mutant was used to control the potential effect of an insertional mutation in *mgdβc1* and *mgdβc3* mutants.

For *MGDγ*, we selected two mutants obtained with the sgRNA 619h (*mgdγh1* and *mgdγh2*) and one obtained with the sgRNA 619i (*mgdγi1*) (**Fig. 4.12**). For this gene, obtention of mutant colonies was particularly laborious as the mutants had a growth phenotype on solid medium greatly affecting the formation of colonies. Only one mosaic colony was obtained after transformations with 619h and 619i sgRNAs. Isolation of pure mutant strains with the classical method consisting of spreading around 200 cells on a plate never lead to the growth of a single colony. By simply spreading the cells on solid medium using a loop, we managed to isolate three pure colonies from the mosaic colony obtained with 619h. For the mosaic colony obtained with 619i, the mutant profile revealed that there was no trace of WT sequence, and all indels lead to either a frameshift in the sequence or the deletion of crucial amino acids in the active site region (**Fig. 4.12**). As we were unsuccessful at isolating pure mutants from this colony, we decided to use this strain without further purification.

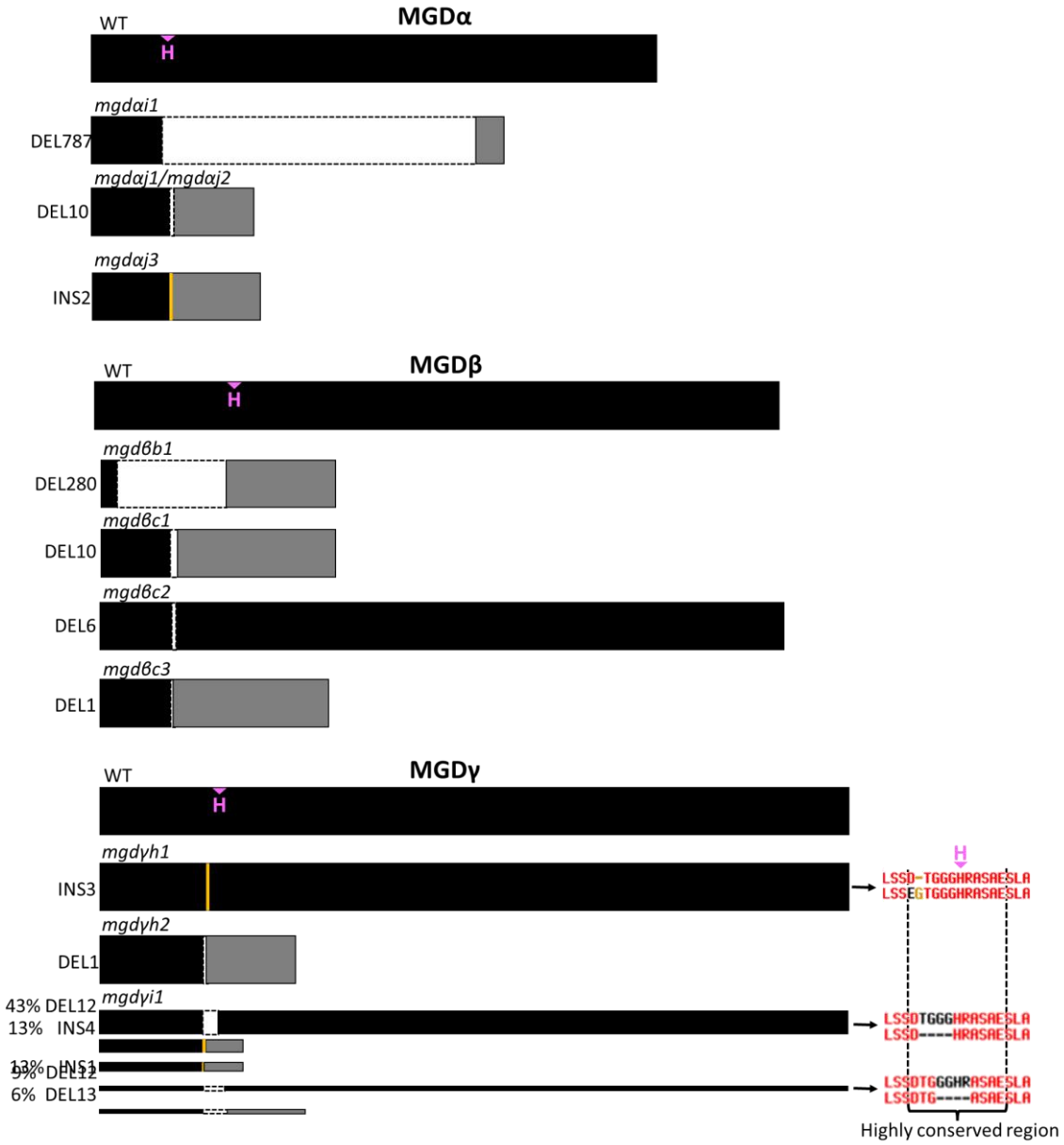


Figure 4.12: Presentation of selected mutant for phenotyping analysis. Schematisation of the impact of the indels on the protein sequences in the mutants. The relative position of the catalytic histidine is indicated in pink in the wild type protein sequences. Nature of the indels is indicated next to the mutant protein sequence in base pair unit. The colours of the rectangles indicate if the sequence is in phase (black) or not (grey) with the wild type protein sequence. White rectangles correspond to portion of the sequence that are absent due to a deletion. Yellow marks indicate the position of an insertion. In the case of *mgdγi1* mixed mutant, the five most abundant indels were represented. *Mgdγh1* and *mgdγi1* strains contain indels that do not induce a frameshift. For these, alignments of the wild type and mutant sequences show the impact of the indels on a highly conserved region of the protein.

4.4.1.2.2. Protein expression in overexpressing lines

MGD-eGFP protein expression in the strains provided by Hanhua Hu was verified by detection of an eGFP signal by confocal microscopy (**see section 4.2**). In order to further measure the relative protein overexpression levels, we performed a Western blot analysis on total protein extract (**Fig 4.11**). The protein weight expected for each MGD protein prior cleavage of targeting sequences are 51.25 kDa for MGDα, 61.41 kDa for MGDβ, and 66.49 kDa

for MGDγ. Given that the eGFP protein weights 26.9 kDa, we expect to detect the chimeric proteins around 78 kDa for MGDα-eGFP, 89 kDa for MGDβ-eGFP, and 93 kDa for MGDγ-eGFP.

Western blot analysis revealed a band at 70 kDa for MGDα-eGFP and 80 kDa for MGDβ-eGFP, which is around 8 kDa less than predicted (**Fig 4.11**). These results are coherent with the cleavage of the bipartite presequences in these proteins. Two bands were detected for MGDγ-eGFP, at 90 kDa and 100 kDa. The 100 kDa band seems stronger than the 90 kDa band. This is surprising given that no alternate splicing was detected by cDNA sequencing of the transcript. Therefore, either the translation can start from two different methionine start codon, or the protein can present distinct post-translational modifications (PTM). MGDγ does not contain any predicted bipartite presequence, therefore no cleavage of the protein is expected at its N-terminal and the molecular weight is expected to be close to the prediction. Therefore, the band at 90 kDa might correspond to the protein starting from the predicted start methionine without PTM. Another start methionine upstream of the predicted start methionine is not possible due to the presence of a stop codon. Therefore, the addition of 10 kDa to the protein might be linked to PTMs such as glycosylations. Such PTM could influence protein solubility and serve as a control mechanism for instance for the localization of the protein to the EpM and to the cytosol.

MGDα-eGFP expression levels are similar in both overexpressor strains (**Fig 4.11**). MGDβ-eGFP seems slightly more expressed in the MGDβ-eGFP-A strain than the MGDβ-eGFP-B strain. MGDγ-eGFP is much more expressed in the MGDγ-eGFP-B strain than the MGDγ-eGFP-A strain, which is consistent with previous confocal microscopy observations (data not shown). These differences in expression level were taken into account when analysing the lipid profile.

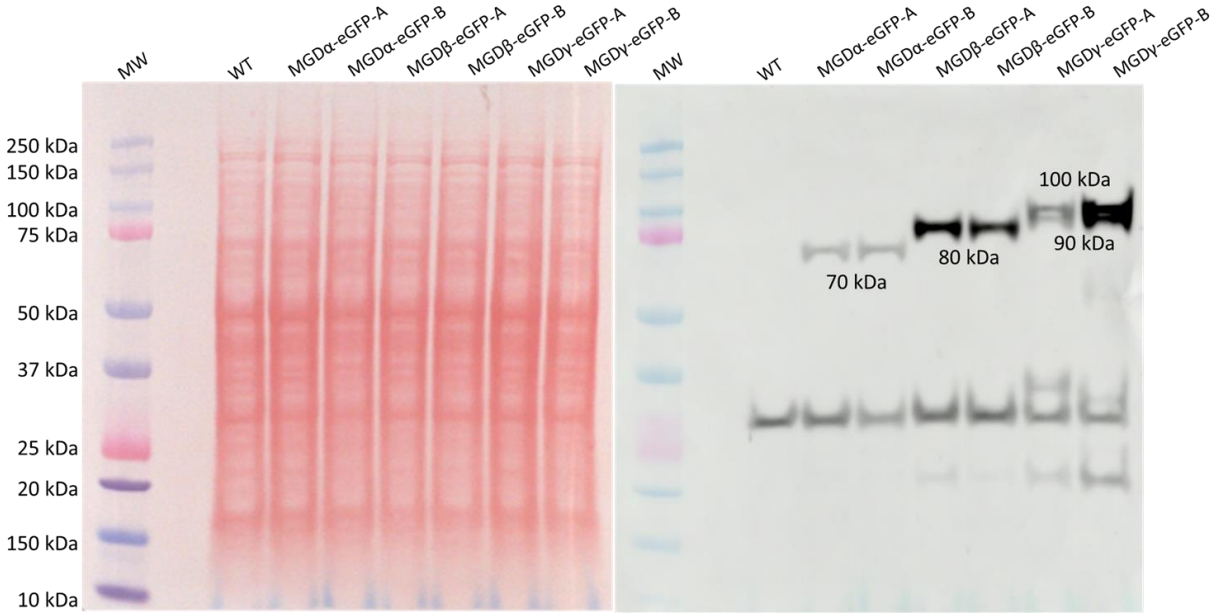


Figure 4.13: Visualisation of MGD-eGFP proteins overexpression by Western Blot. Red ponceau and western blot analyses of total protein extract from *MGD-eGFP* overexpressing *P. tricornutum* strains using wild type (WT) as a control. Quantity of total protein extract per well: 10 µg. Membrane revelation was performed with antibody anti GFP-HR at 1/5000.

4.4.2. Phenotypic analysis

4.4.2.1. Impact of MGD knock-out and overexpression on *Phaeodactylum* fitness under optimal growth conditions

A first key observation on the *MGD* mutant strains is that, contrary to what happens in *A. thaliana*, none of the *MGD* isoform knockouts are lethal. The growth of each mutant strains was monitored (**Fig. 4.14**). Most of the knockout mutant strains had a slightly lower growth than the WT strain. Of all the mutants, the decrease in growth rate was the most important in the *mgdγh1* and *mgdγh2* strains (**Fig. 4.14 E and F**). This decrease was not as pronounced in the *mgdγi1* strain, and therefore the severity of the phenotype in the *mgdγh1* and *mgdγh2* strains might be linked to insertional mutation or off-target mutations. On the contrary, overexpression of MGDα and MGDγ but not MGDβ led to a slight increase in growth (**Fig. 4.14 G**).

The impact of *MGD* knockouts on the growth curves showed that none of the MGD isoforms seems to play a major role in the cell, which potentially reveal a compensation effect between the proteins.

Mutant fitness was also evaluated by measuring the photosynthetic efficiency (Ψ_{II}) and non-photoquemical quenching (NPQ) during a stress (**Fig. 4.15**). The amount of light energy absorbed by a cell that can be used for photosynthesis depends on the amount of photosynthetic chain transporters. When exposed to a light intensity higher that what a cell can use for photosynthesis, the excessive energy is dissipated into heat by NPQ. This photoprotection mechanism is itself limited by the amount of photoprotection effectors, and saturation of both photosynthesis and NPQ leads to oxidative stress. As MGDG plays an important role in the stabilisation and function of the photosystem, as well as in the xanthophyll cycle, we sought whether MGD knockouts had any effect on photosynthesis and induction of photoprotection.

Cells were exposed to two different kinds of stress. A high light stress consisted in exposing dark-adapted cells to high light (700 µE) (**Fig. 4.15 A-C**), while a moderate light stress consisted in exposing dark-adapted cells to a 2-steps increase in light intensity (55 then 335 µE) (**Fig. 4.15 D-F**). In both cases, Ψ_{II} recovery and NPQ relaxation was then monitored at low light intensity (20 µE). Low light was used for recovery because it was shown that NPQ relaxation is very slow in the dark in diatoms, compared to plants or green algae (Ruban et al., 2004; Goss et al., 2006).

As expected, Ψ_{II} diminished compared to the dark-adapted state during illumination, in accordance with the light intensities, while NPQ was induced. In low light, Ψ_{II} recovered while NPQ relaxed, until reaching levels adapted to low light intensity.

Knockout of MGDα, which is located inside the plastid, only had a small effect on photoprotection. During illumination at high and moderate light intensities, there was no visible effect on Ψ_{II} but we could observe a slightly higher NPQ in the mutants (reaching ~1.73) compared to the WT (reaching ~1.63) (**Fig. 4.15 A, D**). After illumination at 700 µE, Ψ_{II} recovered similarly in the mutants and the WT, except for *mgdαj1* (stabilising at ~0.49 against ~0.54 in the WT). NPQ relaxation however was slower in all mutants following high illumination (**Fig. 4.15 A**).

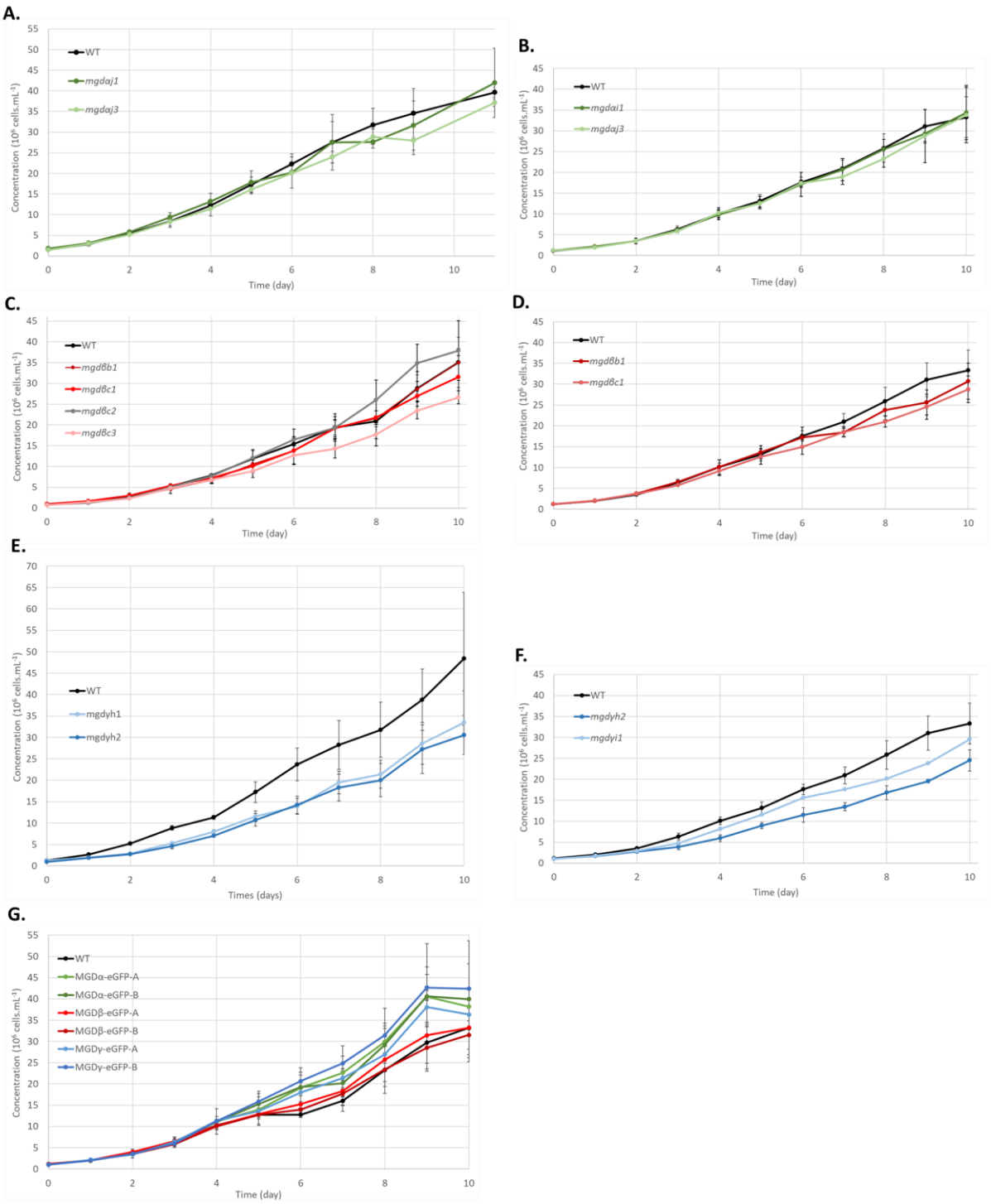


Figure 4.14: Growth curves of the MGD mutant and overexpressing strains compared to the WT. Cell concentration in the mutant strains cultured in parallel with the WT was monitored in independent experiments for *MGDα* (A, B), *MGDβ* (C, D) and *MGDγ* (E, F). MGD overexpressing strains were cultured with a WT in parallel in a single experiment (G). Graphs B, D and F were obtained during the same experiment. The algae cells were cultured at 20 °C in ESAW 10N10P medium, and the cell concentration (10^6 cells.mL^{-1}) in 100-mL flasks was measured every day for 10 to 11 days at the same hour using a TECAN Infinite M1000 PRO. Data are the average of three independent experiments. Values are means and standard errors.

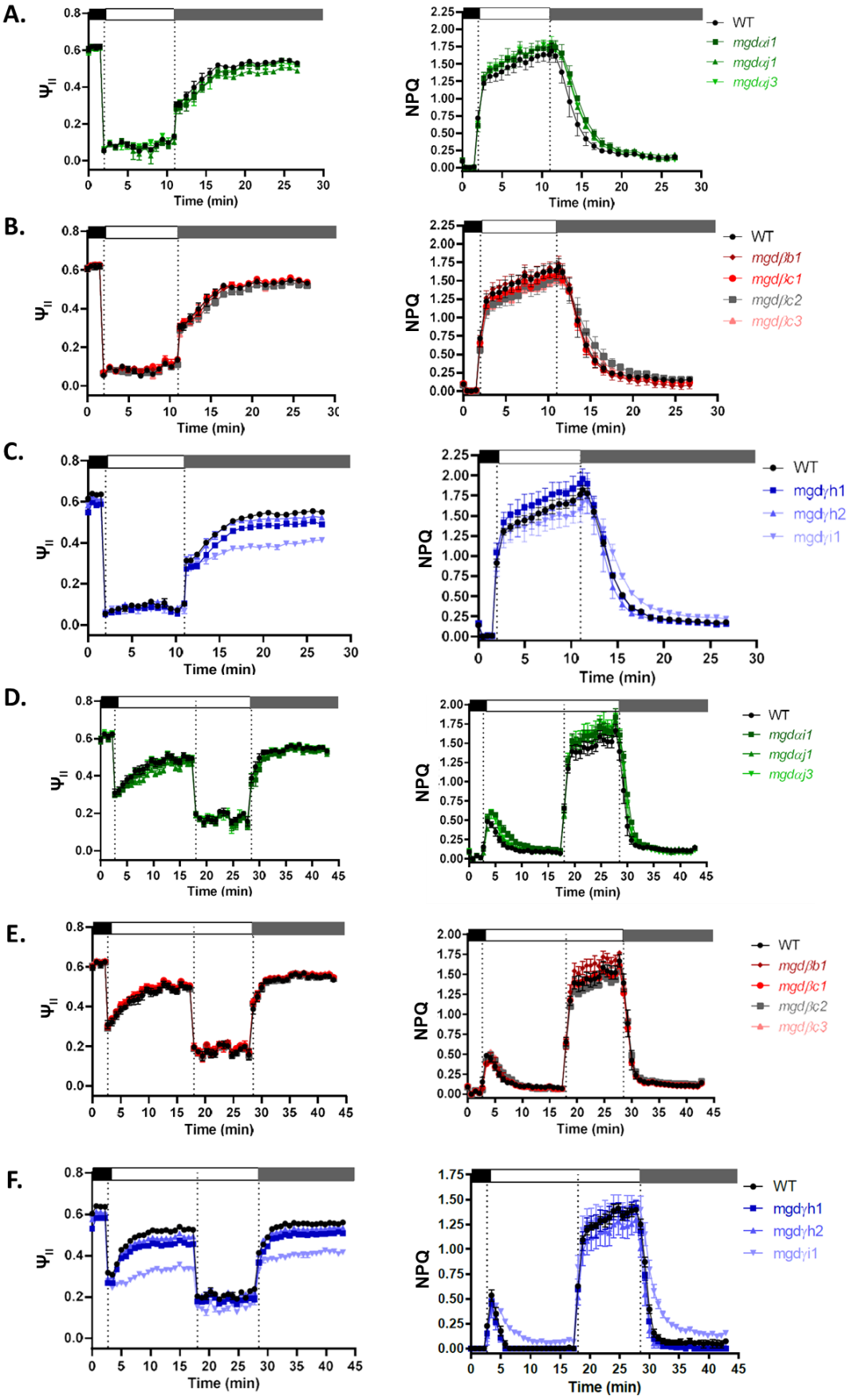

A.
B.
C.
D.
E.
F.
ΨII
NPQ
Time (min)
WT
mgdαi1
mgdαj1
mgdαj3
mgdβb1
mgdβc1
mgdβc2
mgdβc3
mgdγh1
mgdγh2
mgdγi1

Figure 4.15: Photosynthetic efficiency and NPQ induction under high and moderated stresses. *MGDα* (A, D), *MGDβ* (B, E) and *MGDγ* (C, F) mutant strains were exposed to high (700 µE; A-C) and moderate (55 µE then 335 µE; D-F) light intensities, followed by relaxation in low light (20 µE). Chlorophyll fluorescence was measured to calculate Ψ_{II} and NPQ.

Knockout of MGDβ did not lead to any observable changes in the photosynthetic efficiency and NPQ (**Fig. 4.15 B, E**). Only the mutant *mgdβb1* had a slightly higher NPQ level (~1.69) during the second step of the moderate stress compared to the WT (~1.57).

Knockout of MGDγ did not lead to any common phenotype on NPQ in the mutants. Nevertheless, all mutants seemed affected in the recovery of Ψ_{II}, especially *mgdγi1* strain (~0.40 against ~0.55 in the WT).

The inability to fully recover Ψ_{II} is a signature of photoinhibition. A weak photoinhibition was thus observed in MGDα and MGDγ mutant strains, while photoprotection was only slightly impacted in the MGDα mutant strains.

Overall, the changes in photosynthetic efficiency and NPQ in the mutants were very subtle and might result from a mild or indirect stress due to the absence of an MGD. Therefore, the overexpressors phenotype was not investigated further.

4.4.2.2. Cell morphology and membrane structure in studied lines

Cell morphology in several mutants was examined by epifluorescence microscopy (**Fig. 4.16**). All mutant cells appeared fusiform with a normal sized plastid compared to the WT. No changes in cell size were observed either.

In the MGD overexpressing strains, there was no changes in cell morphotype either, but chlorophyll fluorescence seemed higher in MGDα and MGDγ overexpressing lines, possibly due to the fluorescence of the eGFP in the plastid and the EpM/cytosol, respectively.

Potential impact of MGD knockouts on the plastid membranes was checked by scanning transmission electron microscopy (STEM) (**Fig. 4.17**). A selection of mutants was observed: *mgdαi1* and *mgdαj3* strains, *mgdβc2* and *mgdβc3* strains, and *mgdγh2* and *mgdγi1* strains. For MGDβ mutant strains, *mgdβc2* was used as a reference for a possible phenotype. Only one representative image is shown for each isoform.

No clear modifications were observed in the cell membrane structure in any MGD mutants (**Fig. 4.17**). For MGDβ mutant, we wanted ideally to observe a blob-like structure, however, it is difficult to obtain clear images of the blob as it requires to have cut the cell right inside the blob during sample preparation. Figure 4.15 D-G, as well as 4.8 C show possible blob-like structures.

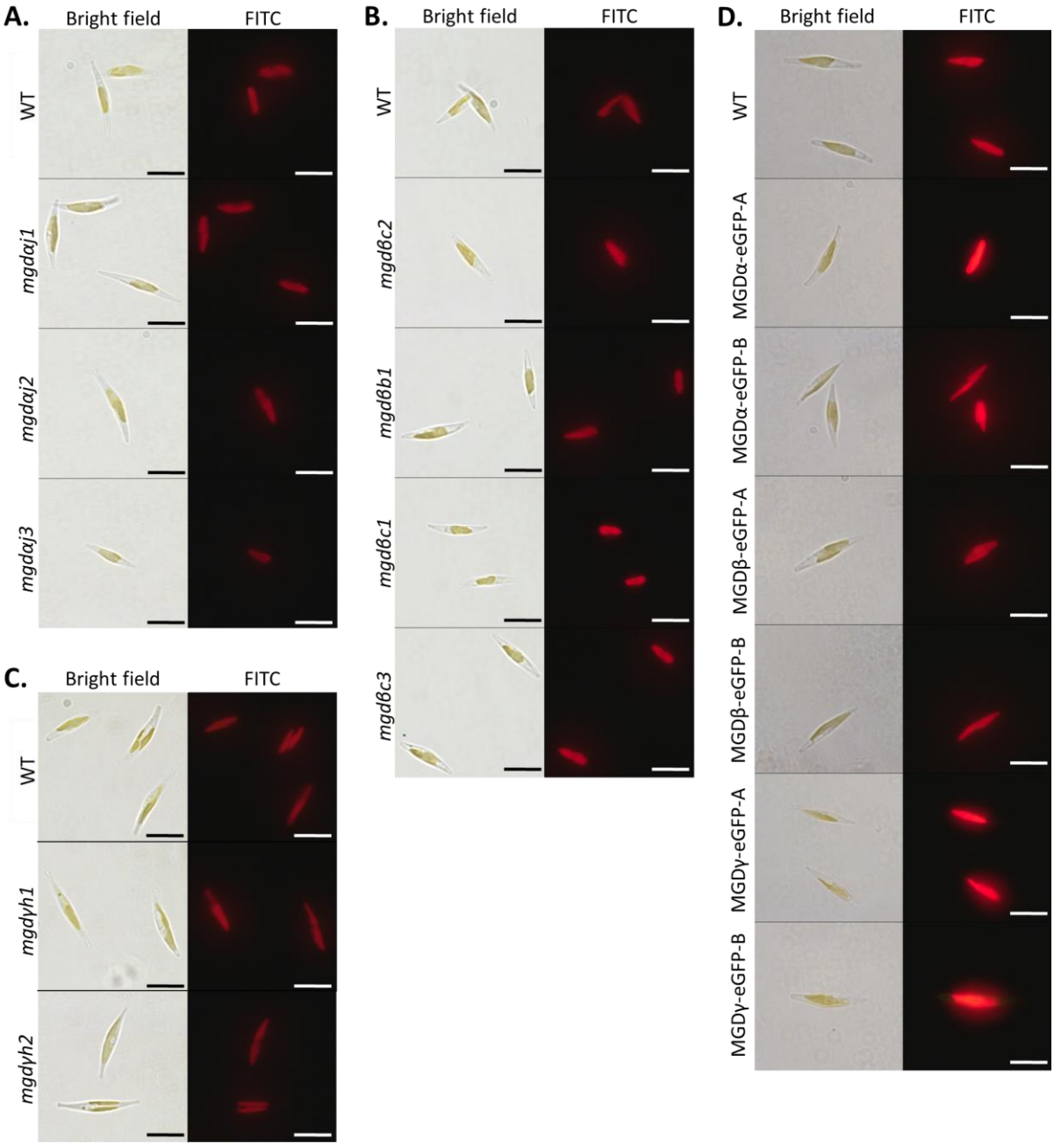


Figure 4.16: Observation of cell morphology by epifluorescence microscopy. Cell cultures were observed with an epifluorescence microscope. Mutant strains for MGDα (A), MGDβ (B) and MGDγ (C) were cultured separately with a wild type (WT), while overexpressing strains MGDα-eGFP-A and -B, MGDβ-eGFP-A and -B, and MGDγ-eGFP-A and -B were all cultured in parallel, with the WT (D). Bright field images and chlorophyll autofluorescence images were taken with an oil-immersed objective 100x. Chlorophyll fluorescence was observed with an FITC filter. Scale bar: 10 µm.

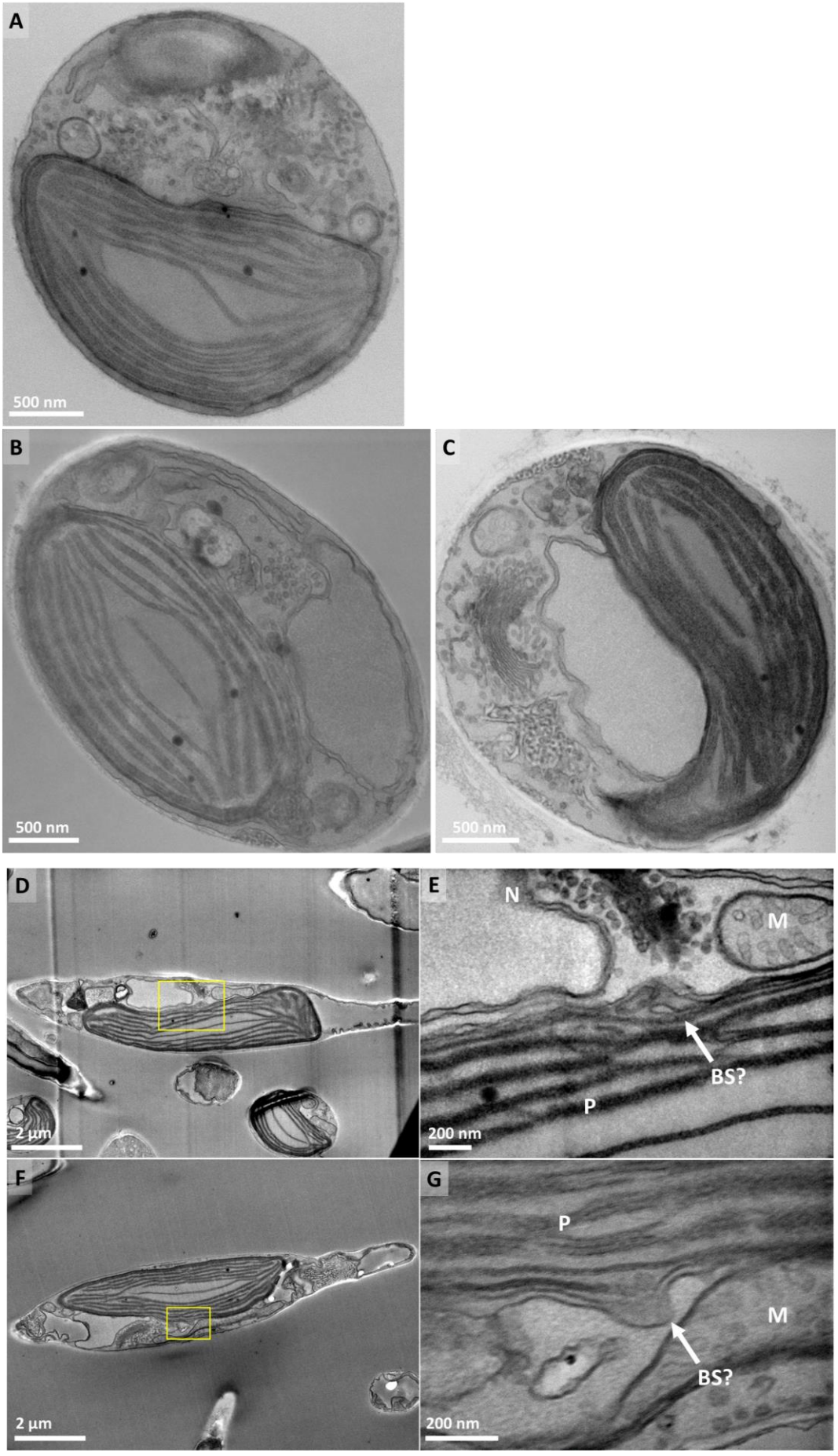

A
500 nm
B
500 nm
C
500 nm
D
2 µm
E
N
M
BS?
P
200 nm
F
2 µm
G
P
M
BS?
200 nm

Figure 4.17: Observation of membrane integrity by scanning transmission electron microscopy (STEM) in MGD mutants compared to the WT. *Mgdαj3* (B) and *mgdγh2* (C) were both cultured in parallel with a WT (A). In a separate experiment, *mgdβc3* mutant (F, G) was observed with a WT (D, E) cultivated in parallel. Images E and G are a magnification of images D and F, respectively.

4.4.2.3. Changes in the glycerolipidome following alteration of MGD expression

We cultured the MGD knockout strains in parallel with the WT in a 10N10P medium and carefully extracted the lipids as described in the 'Materials and Methods'. An aliquot of the total lipid extract was methanolized to produce FA methyl esters (FAMEs). The FAMEs were analysed by gas chromatography coupled with flame ionisation detection (FID) to obtain the global FA profile in each lipid extract. Based on the lipid quantification obtained by GC-FID, a 5-nmol aliquot of total lipid extract was analysed by high pressure liquid chromatography-tandem mass spectrometry using qualified controls (QC) as previously described (Jouhet et al., 2017). This method allows the approximate quantification of each lipid class, and each lipid species inside the lipid classes.

4.4.2.3.1. Total fatty acids and lipid classes profiles are mainly unaffected

The total FA distributions obtained in our experiments are similar to those already reported in the literature in the same conditions (Abida et al., 2015). *P. tricornutum* is rich in eicosapentaenoic acid (20:5) (about 25-30 % of total FAs) and C16 molecular species (mainly 16:0, 16:1, 16:3), while poor in C18 FA.

Overall, the total amount of FA in all MGD mutants was similar to the WT analysed in parallel (between 6.64 to 7.10 nmol.10^{-6} cells on average) (**Fig. 4.18A**). In MGDα mutant strains, there was a slight increase in the proportion of 16:1 FA (19.7 % on average in the mutants against 17.0 % in the WT) acid while 16:3 and 16:4 tended to decrease, in comparison with the WT. A decrease in 18:2 was also observed (1.13 % on average in the mutants against 1.69 % in the WT). In MGDβ, no change in FA distribution could be observed. MGDγ mutants tended to have slightly more 16:0 and less 14:0 species. Significant changes in FA proportions were mainly limited to one or two mutants, nevertheless there seemed to be fewer amounts of 18:0 in all mutants (2.11 % on average in the mutants against 2.87 % in the WT).

The distribution of the different lipid classes was analysed by LC-MS/MS. The obtained glycerolipid profile is consistent with previous observations (**Fig. 4.18B**) (Abida et al., 2015). The profile is dominated by MGDG, PC and SQDG, which combined accounted for about 70 % of total glycerolipid content.

The glycerolipid profile was mainly similar between all MGD mutants and the WT. The distribution of the different lipid classes indicated that the mutants had no impact on galactolipids proportions.

- MGDα mutants showed lower levels of SQDG (16.0 % in the mutants on average against 19.6 % in the WT) and a tendency to contain more PC (17.8 % in the mutants on average against 14.10 % in the WT). The non-polar lipid DAG is present in minor amounts in 10N10P conditions, so variations of this lipid class are very subtle. However, it seems that a lower proportion of DAG was detected in MGDα mutant strains (1.28 % on average) compared to the WT (1.54 %).

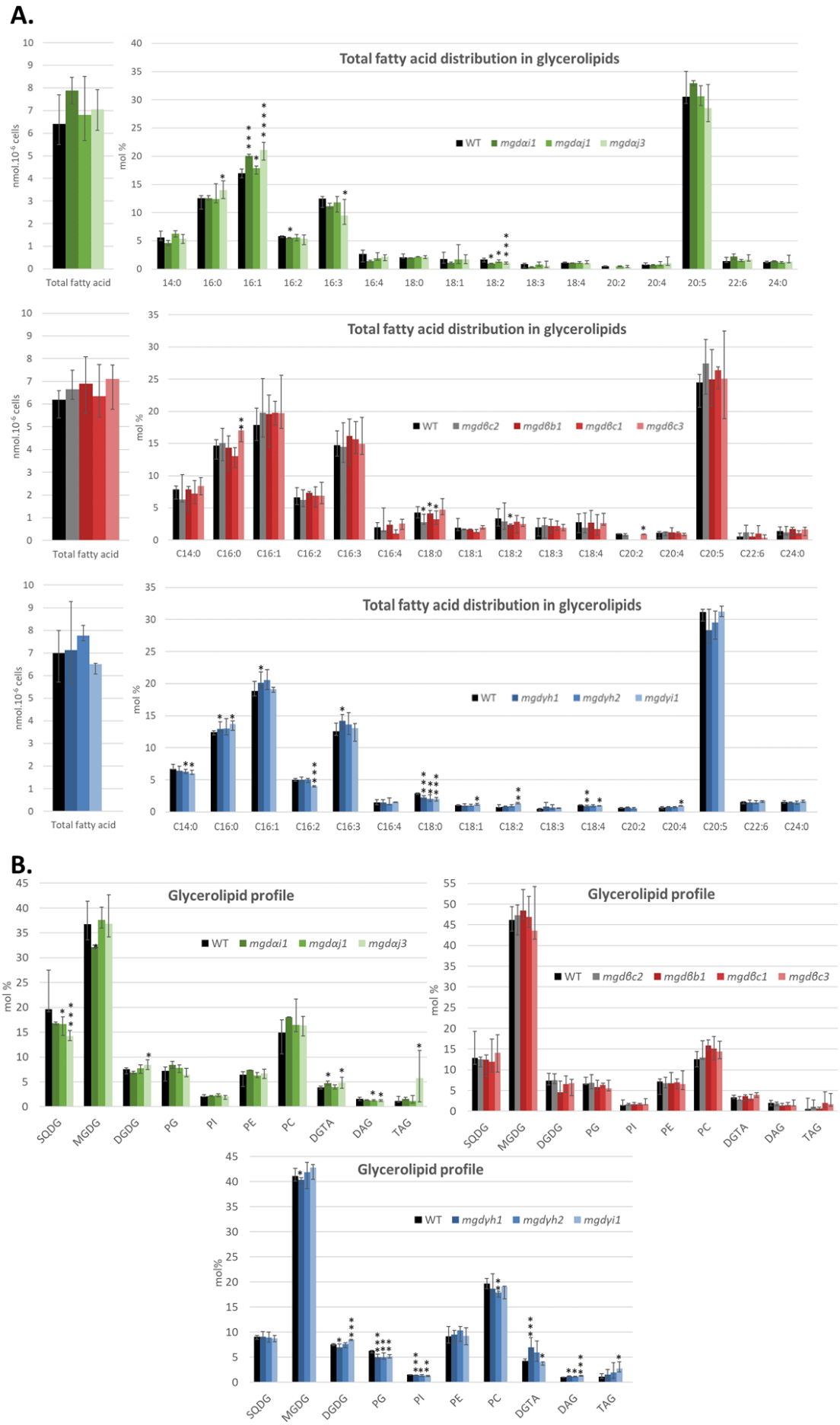

A.
Total fatty acid distribution in glycerolipids
nmol.10⁻⁶ cells
mol %
Total fatty acid
WT mgdαi1 mgdαj1 mgdαj3
14:0 16:0 16:1 16:2 16:3 16:4 18:0 18:1 18:2 18:3 18:4 20:2 20:4 20:5 22:6 24:0
Total fatty acid distribution in glycerolipids
WT mgdβc2 mgdβb1 mgdβc1 mgdβc3
Total fatty acid distribution in glycerolipids
WT mgdyh1 mgdyh2 mgdyi1
B.
Glycerolipid profile
WT mgdαi1 mgdαj1 mgdαj3
SQDG MGDG DGDG PG PI PE PC DGTA DAG TAG
Glycerolipid profile
WT mgdβc2 mgdβb1 mgdβc1 mgdβc3
Glycerolipid profile
WT mgdyh1 mgdyh2 mgdyi1
mol%

Figure 4.18: Quantitative analysis of FA and glycerolipid content in MGD mutants. Lipids from *P. tricornutum* MGD mutant strains and the WT grown in 10N10P medium were extracted, and separated as described in 'Materials and Methods'. **A**, total FA content (given in nmol.10^{-6} cells) and global FA profile (given in molar percentage) in total glycerolipid extracts. **B**, molar profile of the different glycerolipid classes in total glycerolipid extracts. The *mgdβc2* mutant contains a silent mutation and was used as a control for MGDβ knockout mutants. Each result is the median of six biological replicates ± min and max values. Significant differences are indicated with one asterisk for 0.05 > p-value > 0.01, two asterisks for 0.01 > p-value > 0.001, and three asterisks for 0.001 > p-value, and were calculated by an unpaired multiple t test using GrahPad Prism software.

- In MGDβ mutants, the proportion of PC was higher in the knockout mutants (14.10 % on average) compared to both controls (12.7 % on average), which is coherent with a decrease in PC in the MGDβ overexpressing strains (Fig. 4.19B).
- All MGDγ mutants had less PG (5.1 % on average against 6.3 % in the WT) and PI (1.40 % on average against 1.53 % in the WT), and more DAG (1.23 % against 1.05 % in the WT). Consistently, DAG tended to decrease in MGDγ overexpressing strains. The *mdgγh1* and *mgdγh2* strains tended to have higher proportion of DGTA (6.45 % on average) than in the WT (4.33 %), consistent with a slight decrease in DGTA in the MGDγ-eGFP-B strain (**Fig. 4.19B**). The *mgdγi1* on the contrary tended to have about the same proportion of DGTA (4.01 %). With the exception of DGTA which level might be slightly altered, lipid class distribution was well conserved between mutants.

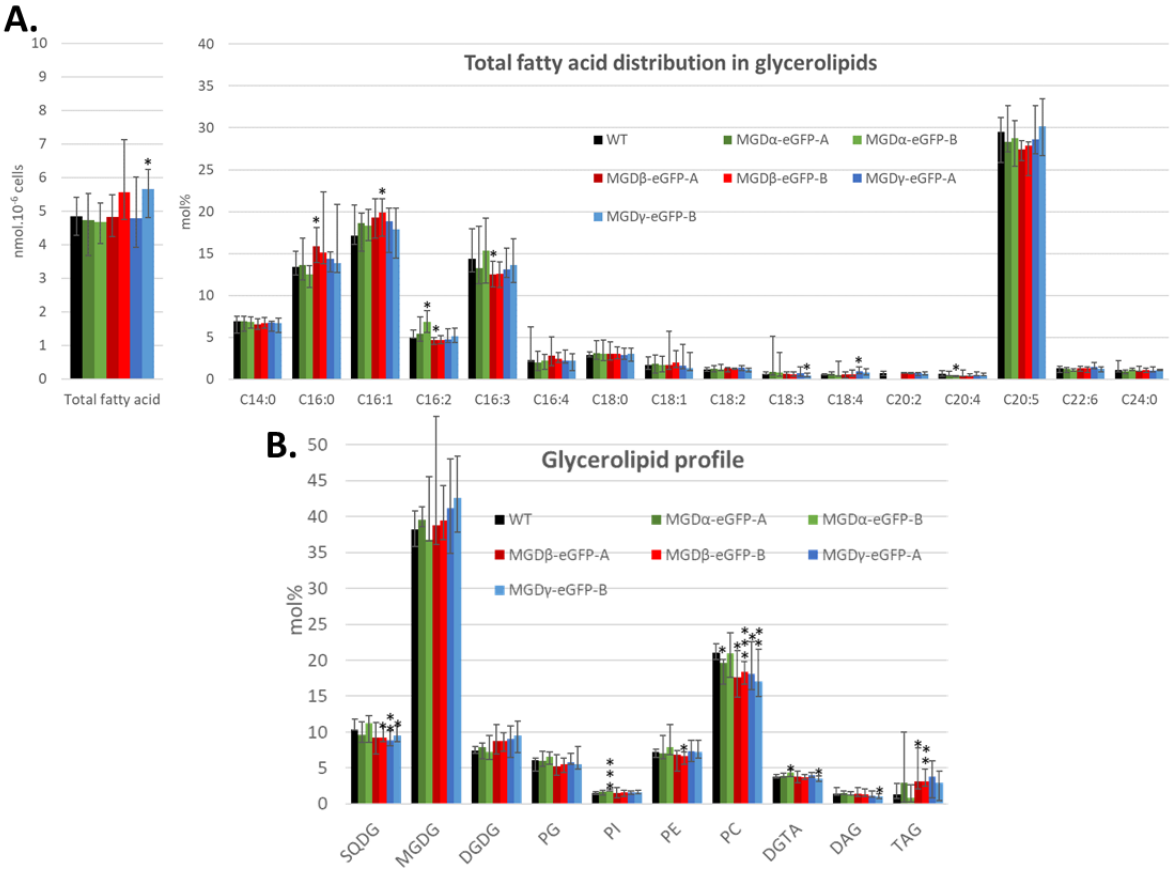


Figure 4.19: Quantitative analysis of FA and glycerolipid content in MGD overexpressing lines. Lipids from *P. tricornutum* MGD overexpressing lines and the WT grown in 10N10P medium were extracted, and separated as described in 'Materials and Methods'. **A**, total FA content (given in nmol.10^{-6} cells) and global FA profile (given in molar percentage) in total glycerolipid extracts. **B**, molar profile of the different glycerolipid classes in total glycerolipid extracts. Each result is the median of six biological replicates ± min and max values. Significant differences are indicated with one asterisk for 0.05 > p-value > 0.01, two asterisks for 0.01 > p-value > 0.001, and three asterisks for 0.001 > p-value, and were calculated by an unpaired multiple t test using GrahPad Prism software.

In MGD overexpressing lines, total FA amounts were similar to the WT (**Fig. 4.19**). Only one overexpressor, MGDγ-eGFP-B, had a weak increase (p-value = 0.016), which was not observed in the other MGDγ overexpressing line. In terms of FA distribution, MGDα overexpressors tended to have slightly more 16:1 and 16:2 FAs. MGDβ overexpressors had a small increase in 16:0 and 16:1, and a small decrease in 16:2 and 16:3. They also tended to have less 20:5. MGDγ overexpressors were very similar to the WT and might have a tendency for more 16:1 and 18:4.

Glycerolipid profiles were quite similar between all MGD overexpressors and the WT. However, MGDβ and MGDγ overexpressing lines showed a decrease in PC proportion accompanied by a small increase in TAG. These lines also tended to have less SQDG.

4.4.2.3.2. Lipid species composition reveals an important impact of MGD mutation on acyl fluxes in lipid homeostasis

In order to better understand the impact of MGDs' knockouts on the lipid profile, we looked in detail at the variation of each lipid molecular species in each lipid class. The FA positioning for each glycerolipid was previously characterised in *P. tricornutum* (Abida et al., 2015; Jouhet et al., 2017).

Our observations highlighted distinct roles of MGDs in the fluxes of diacyl molecular species, based on the nature of acyl groups at positions *sn-1* and *sn-2* of glycerol backbone. We considered that molecular species decreasing in a mutant line were determined by the mutated enzyme, and we highlighted consistencies in overexpressing lines by a corresponding increase. Our second principle was to assume that an increase of a molecular species proportion in a mutated line might be determined by one or both isoforms acting in a compensation mechanism.

In MGDα mutants, all molecular species containing a C16 in position *sn-1* decreased in MGDG (**Fig. 4.20**), with the notable exceptions of 16:1-16:1, 16:3- 16:3, and 16:4-16:3. Consistently, MGDG 16:1-16:2, 16:1-16:3 and 16:2-16:3 species increased in one of the MGDα overexpressing strain, MGDα-eGFP-B (**Fig. 4.26**).

On the contrary, the proportion of MGDG molecular species containing a 20:5 FA increased, except the 20:5-16:4. This is coherent with a decrease in MGDG 20:5-16:3 in MGDα-eGFP-B. The proportion of all molecular species containing a 14:0 FA decreased in MGDG (**Fig. 4.20**).

DGDG molecular species profile did not mirrored that of MGDG. An increase in species containing a 16:1 in position *sn-1* was observed while the species containing a 20:5 decreased, with the exception of 20:5-16:0 and 20:5-16:4 molecular species. In MGDα-eGFP-B, DGDG 16:1-16:0 and 16:1-16:1 tended to decrease, while 20:5-16:2 species was increased.

In SQDG, the species containing a C16 FA in position *sn-1* tended to increase, while species containing a 20:5 FA tended to decrease. SQDG 14:0-16:0 and 14:0-16:1 species decreased.

In PG, the 16:1-16:0 molecular species decreased in proportion and the 16:1-16:1 species increased.

Many molecular species were impacted in non-plastidial glycerolipids of MGDα mutants (**Fig. 4.21**). All 14:0 containing species decreased, as seen in plastidial glycerolipids. Consistently, these species increased in the non-plastidial lipids of MGDα-eGFP-B strain. All C18-containing species decreased, with the exception of 18:2-18:1 and 18:2-18:2 in PC, and 20:5-18:4 in PE. All species containing a C16 in positions *sn-1* and *sn-2* increased in PC and DGTA. However, 16:0-16:0 and 16:0-16:1 molecular species

decreased in DAG while 16:1-16:1 increased. Species containing a 20:5 in position *sn-1* and a C16 in position *sn-2* increased or tended to increase in PC, DGTA and PE. The 20:5-22:6 molecular species increased in PC, while it decreased in MGDα-eGFP-B.

For the analysis of the impact of MGDβ mutants on glycerolipids, we will only mention the significant differences that marked MGDβ knockout lines without similarly affecting *mgdβc2* mutant.

In the MGDβ knockout mutants, MGDG molecular species containing a 20:5 in position *sn-1* decreased (**Fig. 4.22**), with the exception of 20:5-16:0 and 20:5-16:4. MGDG 16:1-16:0 and MGDG 16:1-16:1 species decreased, while other MGDG species containing a C16 FA in positions *sn-1* and *sn-2* and two or more unsaturations on a C16 increased. The opposite was observed in the MGDβ overexpressing strains.

The molecular species 14:0-16:0 and 14:0-16:1 globally decreased in all plastidial glycerolipids (**Fig. 4.22**). In particular, the MGDG 14:0-16:1 was decreased by half in the knockouts. In contrast, the PG 14:0-16:2 and MGDG 14:0-16:3 tended to increase. In the MGDβ overexpressing strains, MGDG 14:0-16:3 was slightly decreased, as were the other 14:0-containing MGDG species (**Fig. 4.26**).

DGDG molecular species where little impacted in the mutants compared to MGDG species. 14:0-16:1 tended to decrease and 16:1-16:1 strongly decreased, consistently with the impact on MGDG, while 16:1-16:2 strongly decreased as well. 20:5-containing DGDG species were almost not affected, with only a slight increase in 20:5-16:0 compared to the controls.

Except for the species already discussed in SQDG and PG, few significant changes were observed in comparison to both controls in the mutants.

Few changes were observed in non-plastidial glycerolipids (**Fig. 4.23**). In PC, DGTA and PE, 16:1-16:2, 16:1-16:3, and 20:5-16:2 species either increased or tended to increase in the mutants while 20:5-18:1, 20:5-18:2 and 20:5-18:3 species tended to decrease. PE 20:5-20:5 molecular species increased, while it decreased in the MGDβ overexpressing strains. No impact was observed on DAG.

In MGDγ mutant lines, levels of 16:0-16:0, 16:1-16:0 and 16:1-16:1 were strongly decreased in galactolipids (**Fig. 4.24**), while species with the same composition were stable in SQDG and PG. In MGDγ overexpressing lines, these same galactolipid species increased, particularly in MGDG.

The proportion of all 14:0 FA-containing molecular species in plastidial glycerolipids were decreased compared to the WT (**Fig. 4.24**). In particular, 14:0-16:1 species in galactolipids were reduced to half the level of the WT, while MGDG 14:0-16:1 increased in the MGDγ overexpressors. For several molecular species, the *mgdγh1* and *mgdγh2* mutants behaved differently compared to *mgdγi1* mutant. Therefore, we cannot conclude on the impact of MGDγ mutation of the proportion of 16:1-16:2, 16:1-16:3, 16:2-16:3 and 16:3-16:3 species. However, 16:1-16:3, 16:2-16:3 and 16:3-16:3 species clearly decreased in the overexpressors. As for the 20:5-containing MGDG species, no changes were observed except a strong decrease in 20:5-16:4, that tended to increase in the overexpressing lines. In DGDG and SQDG, 20:5-containing species tended to increase with the exception of DGDG 20:5-16:4. PG molecular species distribution was globally unimpacted by MGDγ mutations.

In non-plastidial membrane glycerolipids of MGDγ mutants, many changes in species proportions were observed (**Fig. 4.25**). 16:0-16:1, 16:1-16:0, 16:1-16:1, 16:1-18:1, and 16:1-18:2 molecular species in PC and DGTA increased, while DGTA 16:1-18:3 and PC 16:1-18:4 decreased. Again, the distinct behaviour of the mutants did not allow us to conclude on a change in several 20:5-containing species. Nevertheless, there was a clear decrease in 20:5-20:4 and 20:5-20:5 molecular species of PC, DGTA and PE. Compared to MGDα and MGDβ mutations, 14:0-containing species were not impacted in MGDγ mutants. In DAG, changes were limited to 16:0-16:1 and 20:5-16:0, which tended to increase and decrease, respectively.

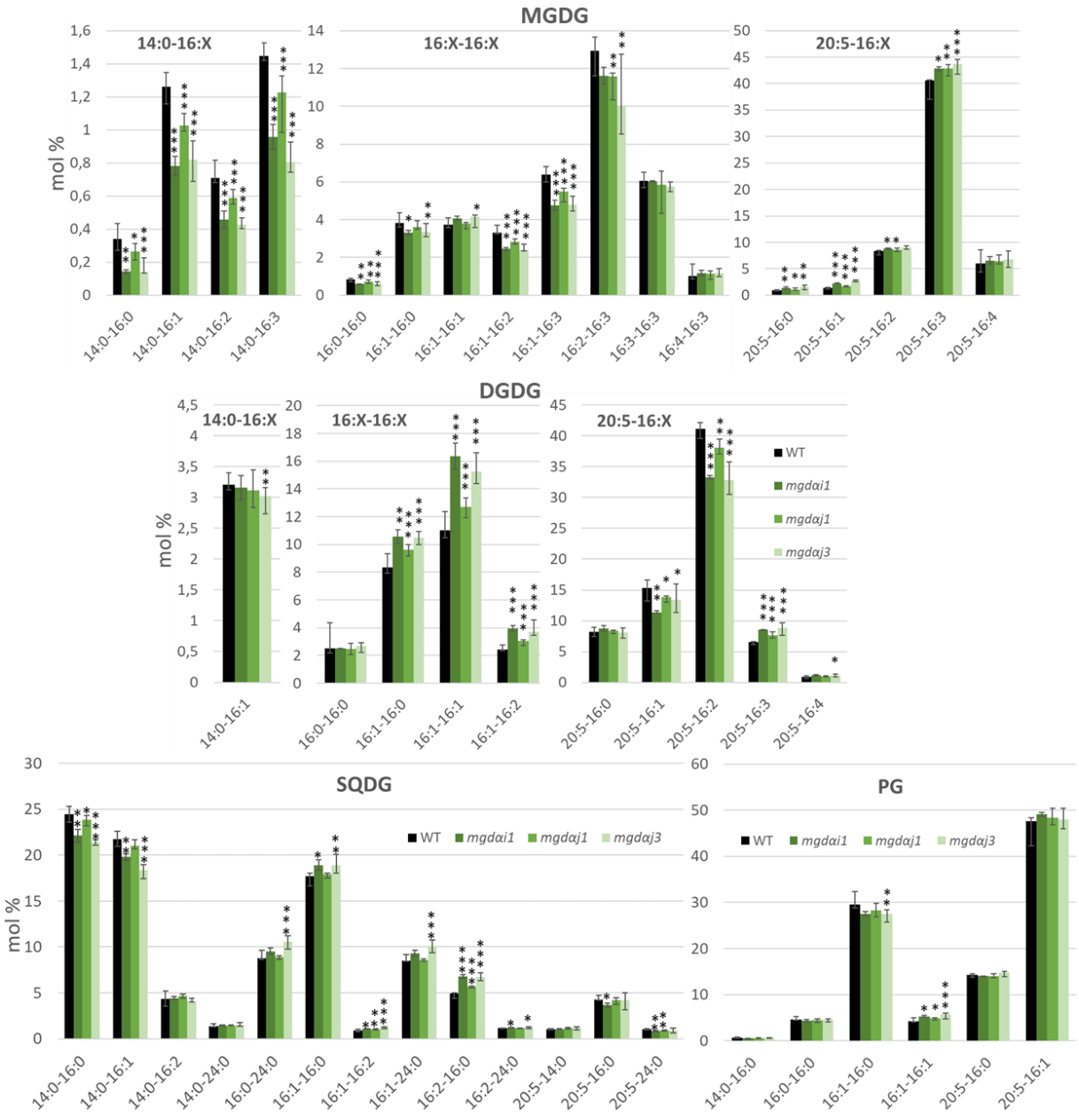


Figure 4.20: Impact of MGDα mutations on the molecular species constituting the main plastidial glycerolipids in *P. tricornutum*. Lipid molecular species in MGDG, DGDG, SQDG and PG were analysed by LC-MS/MS. Each result is the median of six biological replicates ± min and max values. Significant differences are shown by one asterisk for 0.05 > p-value > 0.01, two asterisks for 0.01 > p-value > 0.001, and three asterisks for 0.001 > p-value, and were calculated by an unpaired multiple t test using GrahPad Prism software.

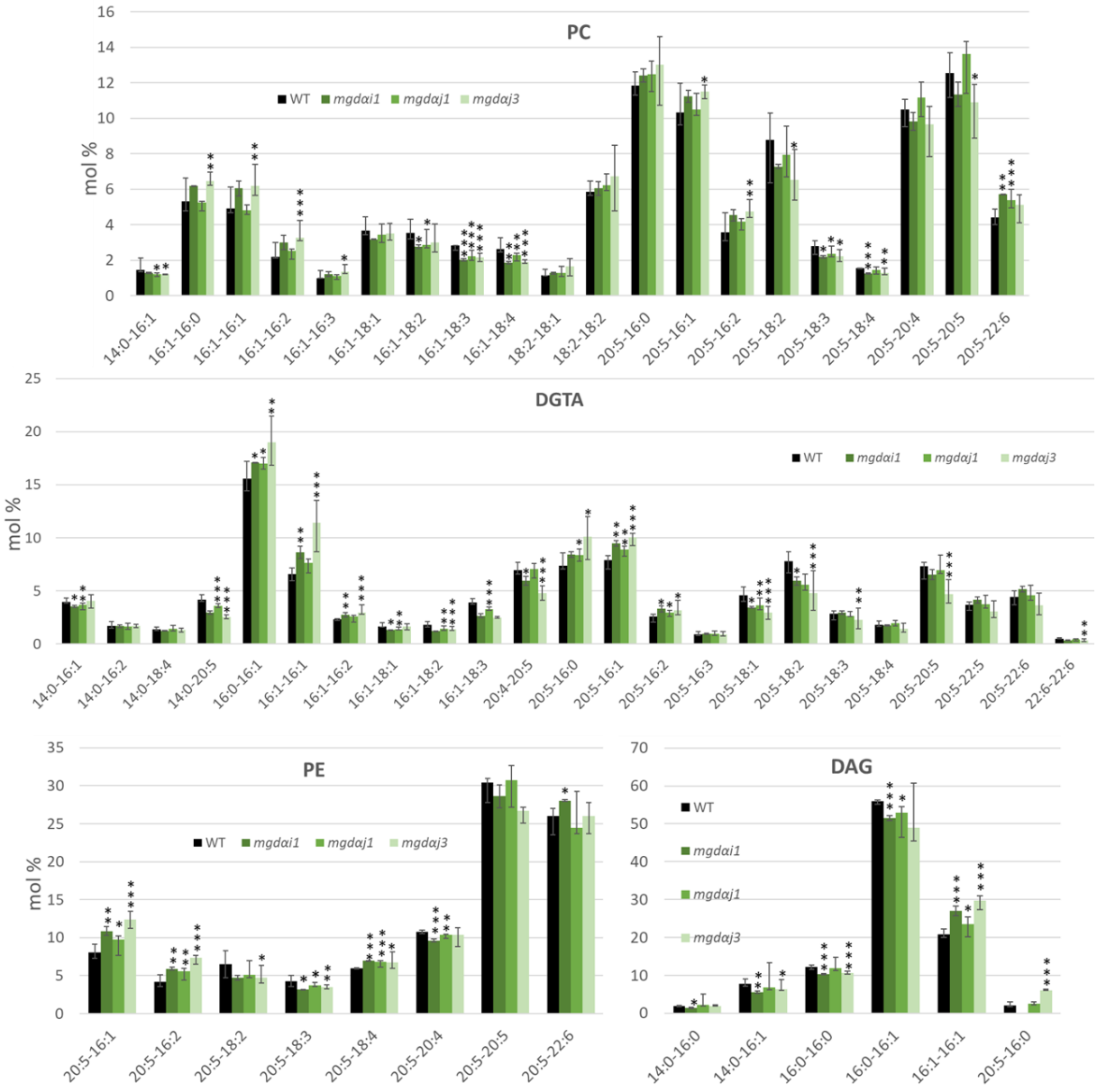


Figure 4.21: Impact of MGDα mutations on the molecular species constituting non-plastidial membrane glycerolipids in *P. tricornutum*. Lipid molecular species in PC, DGTA, PE, and DAG were analysed by LC-MS/MS. Each result is the median of six biological replicates ± min and max values. Significant differences are shown by one asterisk for 0.05 > p-value > 0.01, two asterisks for 0.01 > p-value > 0.001, and three asterisks for 0.001 > p-value, and were calculated by an unpaired multiple t test using GrahPad Prism software.

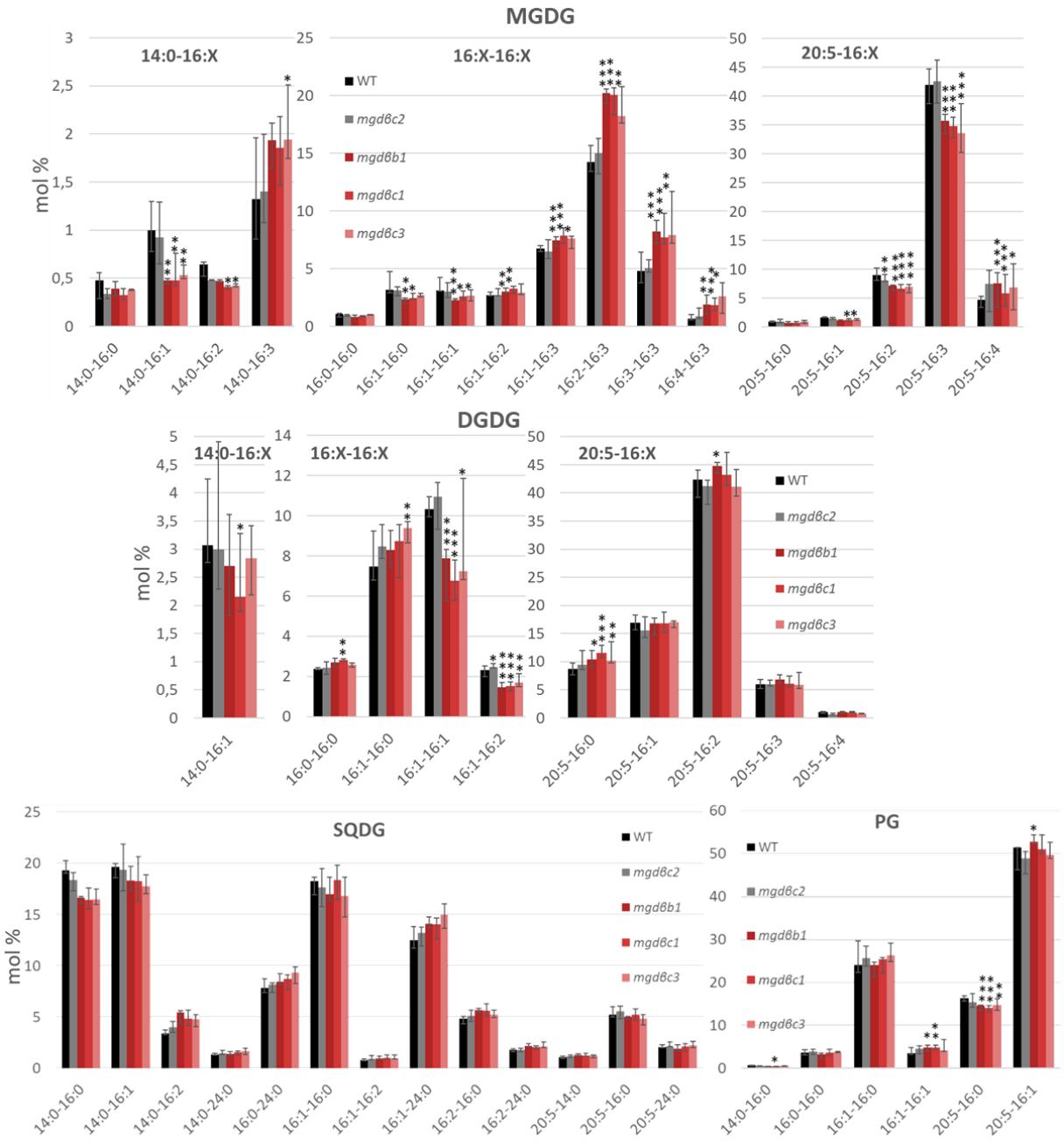


Figure 4.22: Impact of MGDβ mutations on the molecular species constituting the main plastidial glycerolipids in *P. tricornutum*. Lipid molecular species in MGDG, DGDG, SQDG and PG were analysed by LC-MS/MS. The *mgdβc2* mutant contains a silent mutation and was used as a control. Each result is the median of six biological replicates ± min and max values. Significant differences with the WT are shown by one asterisk for 0.05 > p-value > 0.01, two asterisks for 0.01 > p-value > 0.001, and three asterisks for 0.001 > p-value, and were calculated by an unpaired multiple t test using GrahPad Prism software.

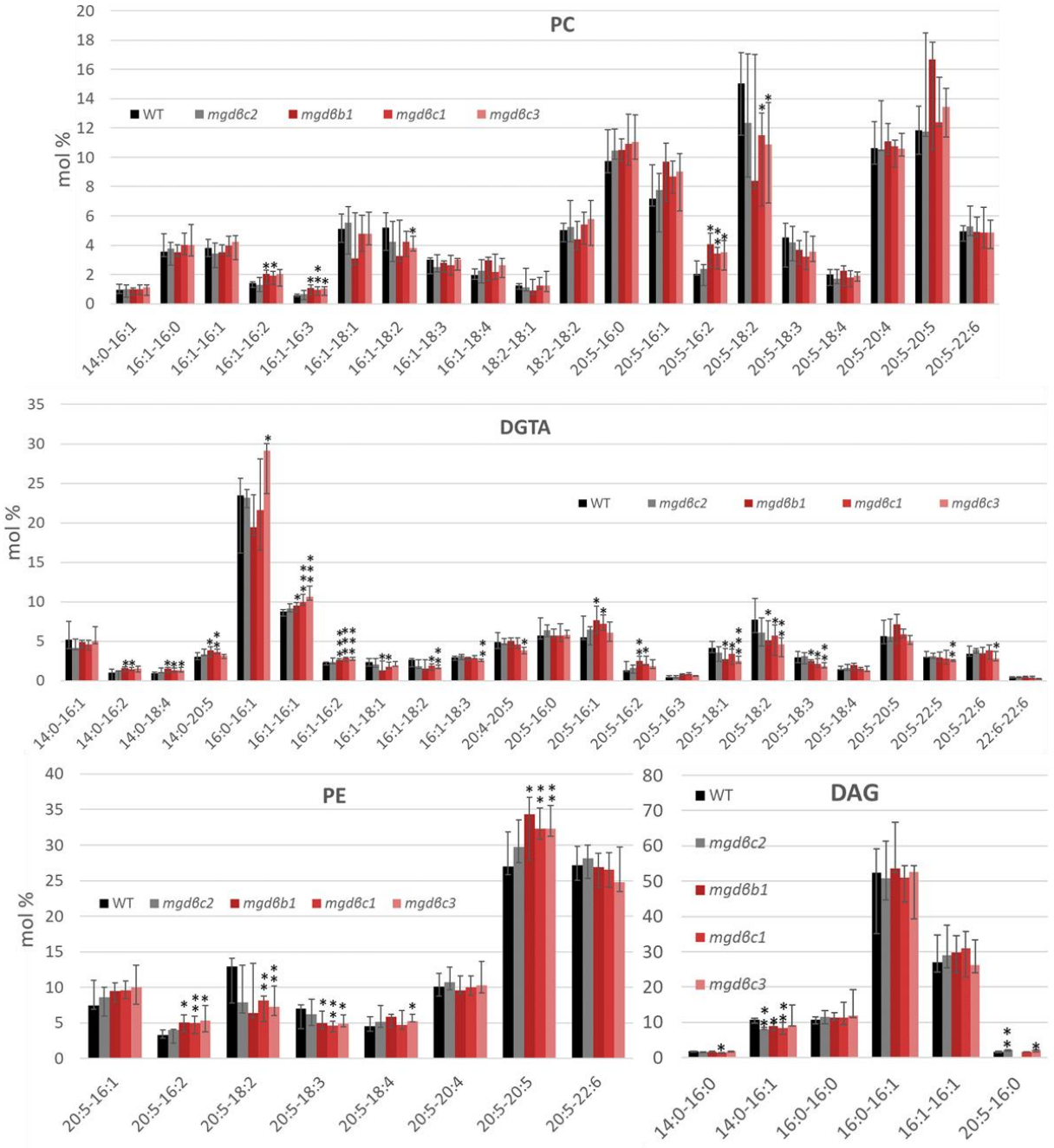


Figure 4.23: Impact of MGDβ mutations on the molecular species constituting non-plastidial membrane glycerolipids in *P. tricornutum*. Lipid molecular species in PC, DGTA, PE, and DAG were analysed by LC-MS/MS. The *mgdβc2* mutant contains a silent mutation and was used as a control. Each result is the median of six biological replicates ± min and max values. Significant differences with the WT are shown by one asterisk for 0.05 > p-value > 0.01, two asterisks for 0.01 > p-value > 0.001, and three asterisks for 0.001 > p-value, and were calculated by an unpaired multiple t test using GrahPad Prism software.

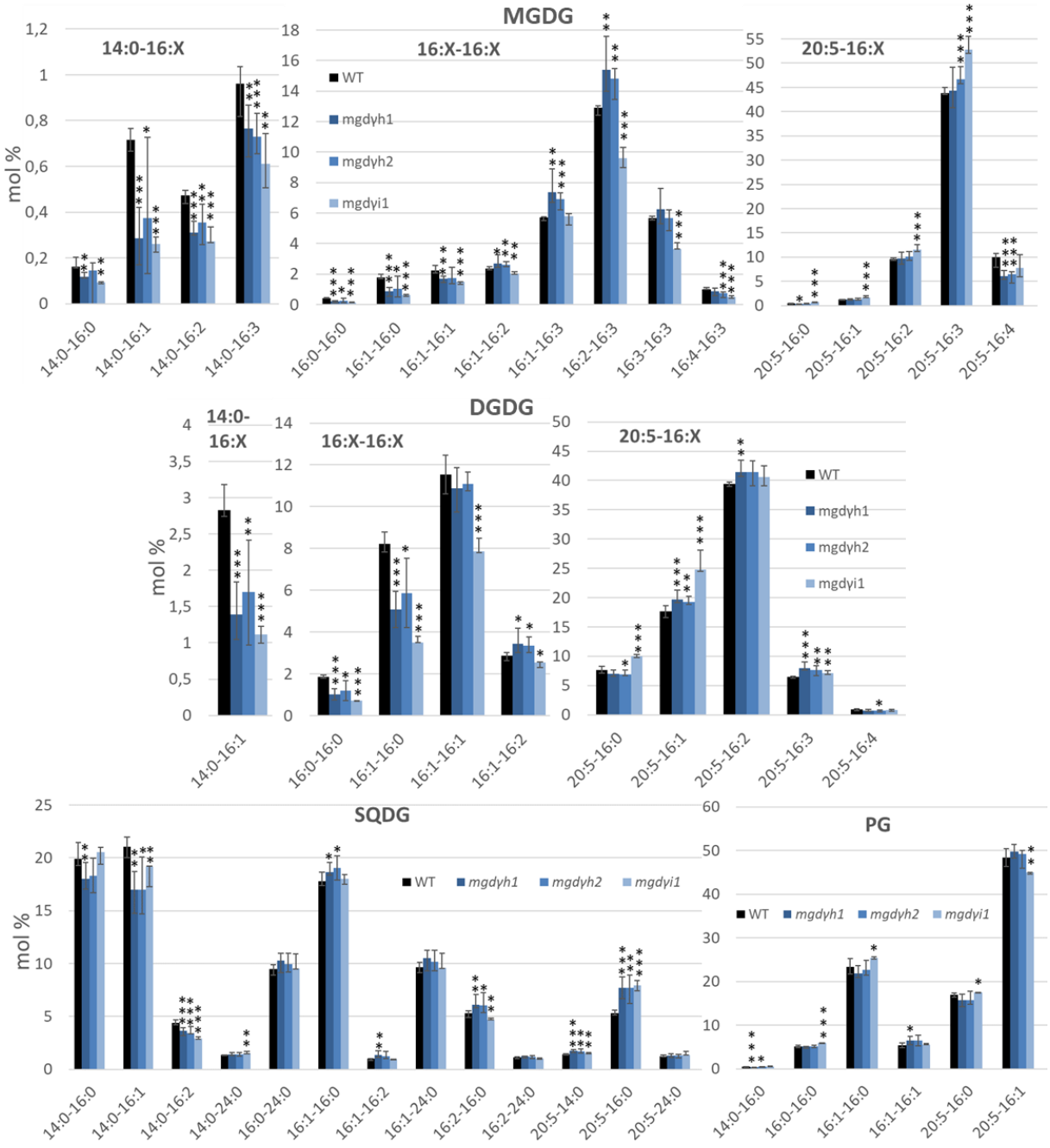


Figure 4.24: Impact of MGDγ mutations on the molecular species constituting the main plastidial glycerolipids in *P. tricornutum*. Lipid molecular species in MGDG, DGDG, SQDG and PG were analysed by LC-MS/MS. Each result is the median of six biological replicates ± min and max values. Significant differences are shown by one asterisk for 0.05 > p-value > 0.01, two asterisks for 0.01 > p-value > 0.001, and three asterisks for 0.001 > p-value, and were calculated by an unpaired multiple t test using GrahPad Prism software.

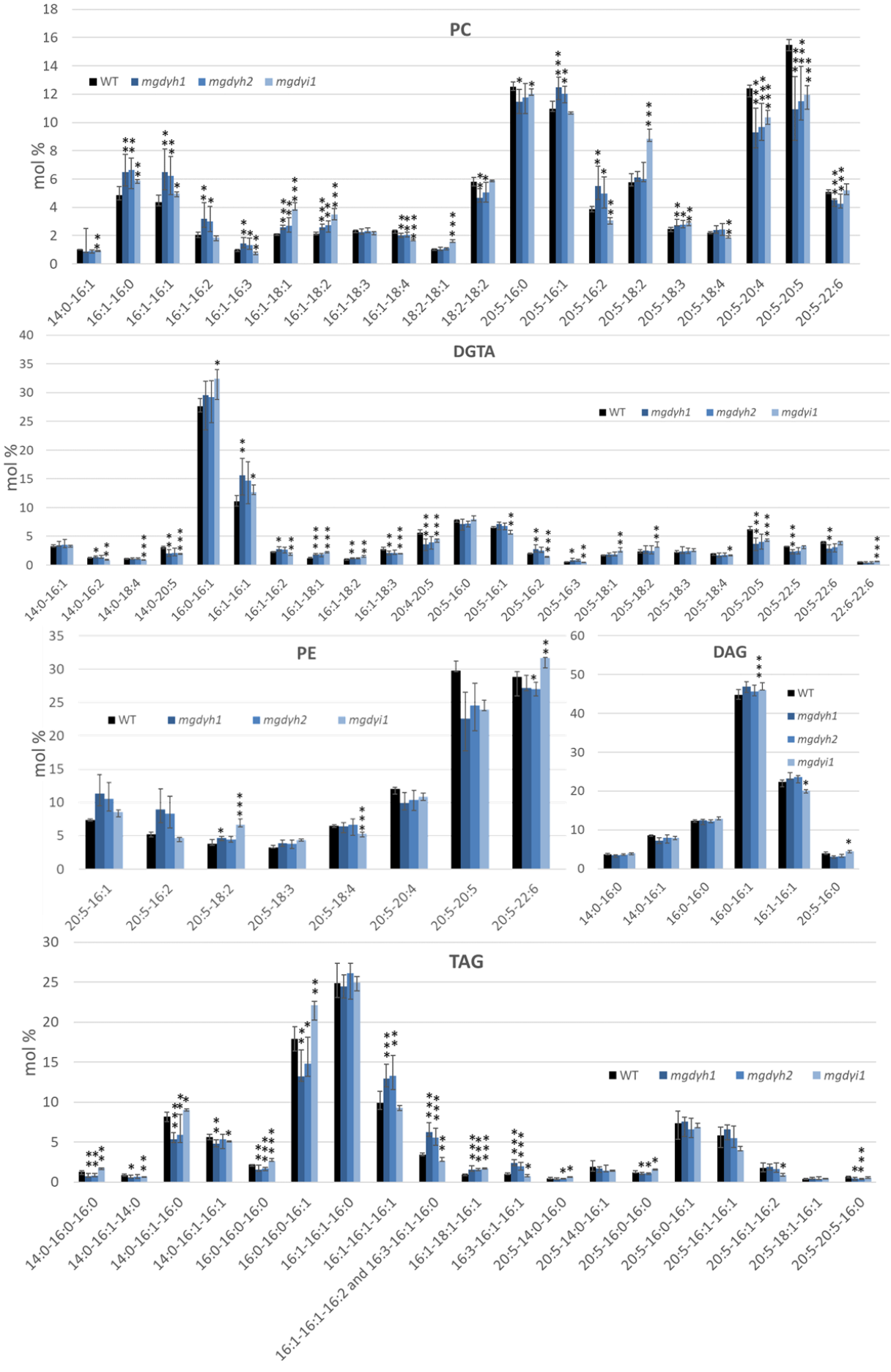

PC
WT
mgdyh1
mgdyh2
mgdyi1
mol %
14:0-16:1
16:1-16:0
16:1-16:1
16:1-16:2
16:1-16:3
16:1-18:1
16:1-18:2
16:1-18:3
16:1-18:4
18:2-18:1
18:2-18:2
20:5-16:0
20:5-16:1
20:5-16:2
20:5-18:2
20:5-18:3
20:5-18:4
20:5-20:4
20:5-20:5
20:5-22:6
DGTA
14:0-16:1
14:0-16:2
14:0-18:4
14:0-20:5
16:0-16:1
16:1-16:1
16:1-16:2
16:1-18:1
16:1-18:2
16:1-18:3
20:4-20:5
20:5-16:0
20:5-16:1
20:5-16:2
20:5-16:3
20:5-18:1
20:5-18:2
20:5-18:3
20:5-18:4
20:5-20:5
20:5-22:5
20:5-22:6
22:6-22:6
PE
20:5-16:1
20:5-16:2
20:5-18:2
20:5-18:3
20:5-18:4
20:5-20:4
20:5-20:5
20:5-22:6
DAG
14:0-16:0
14:0-16:1
16:0-16:0
16:0-16:1
16:1-16:1
20:5-16:0
TAG
14:0-16:0-16:0
14:0-16:1-14:0
14:0-16:1-16:0
14:0-16:1-16:1
16:0-16:0-16:0
16:0-16:0-16:1
16:1-16:1-16:0
16:1-16:1-16:1
16:1-16:1-16:2 and 16:3-16:1-16:0
16:1-18:1-16:1
16:3-16:1-16:1
20:5-14:0-16:0
20:5-14:0-16:1
20:5-16:0-16:0
20:5-16:0-16:1
20:5-16:1-16:1
20:5-16:1-16:2
20:5-18:1-16:1
20:5-20:5-16:0

Figure 4.25: Impact of MGDγ mutations on the molecular species constituting non-plastidial membrane glycerolipids in *P. tricornutum*. Lipid molecular species in PC, DGTA, PE, DAG and TAG were analysed by LC-MS/MS. Each result is the median of six biological replicates ± min and max values. Significant differences are shown by one asterisk for 0.05 > p-value > 0.01, two asterisks for 0.01 > p-value > 0.001, and three asterisks for 0.001 > p-value, and were calculated by an unpaired multiple t test using GrahPad Prism software.

In MGD overexpressing lines, MGDG species were particularly marked by MGDβ and MGDγ overexpressions. In particular, 16:1-16:0 proportion was increased while 16:2-16:3 and 16:3-16:3 were decreased.

There was also an increase in 14:0-16:1 and 16:0-16:0, and a decrease in 16:1-16:3 in MGDγ overexpressors.

MGDβ overexpressing lines tended to have less 14:0-16:2 and 14:0-16:3.

For MGDα, the two overexpressing lines behaved slightly differently. MGDG species distribution in MGDα-eGFP-A did not differ significantly from the WT, while MGDα-eGFP-B had less 16:0-16:0 and 20:5-16:3, and more 16:1-16:2, 16:1-16:3 and 16:1-16:3.

DGDG composition was impacted in all overexpressing lines:

Again, the two MGDα overexpressors behaved with some slight differences for several lipid species, but both appeared to contain less DGDG 14:0-16:1.

In MGDβ overexpressors, there were a decrease in 14:0-16:1, 20:5-16:2 and 20:5-16:3 DGDG species, and an increase in 16:1-16:0 and 20:5-16:1.

Following MGDγ overexpression, there were an increase in 16:1-16:0 and a decrease in 20:5-16:3. The MGDγ-eGFP-B line also contain slightly less 16:1-16:1 and more 16:0-16:0 and 20:5-16:0.

The molecular species distribution of other plastidial glycerolipid were also mostly impacted in MGDβ and MGDγ overexpressing lines, while MGDα overexpression had little effect on the distributions.

In both SQDG and PG, the 16:1-16:0 species were increased in the two MGDβ overexpressing lines. MGDβ overexpression also led to a decrease in SQDG 14:0-16:1, PG 20:5-16:0 and PG 20:5-16:1, a slight increase in SQDG 20:5-16:0, and a strong increase in PG 16:0-16:0.

MGDγ overexpressing lines tended to have less SQDG 14:0-16:1, PG 20:5-16:0 and PG 20:5-16:1, and more SQDG 16:1-16:2, PG 16:0-16:0 and PG 16:1-16:0.

The SQDG and PG molecular species distribution was not impacted in MGDα-eGFP-A line. The second MGDα overexpressor had small differences in proportions compared to the WT. In the SQDG species, MGDα-eGFP-B had less 14:0-16:0, 16:1-16:0, 20:5-14:0 and 20:5-24:0, while more 14:0-16:2, 16:1-16:2, and 16:2-16:0. In the PG species, it showed more 16:1-16:0, 16:1-16:1, and less 20:5-16:1.

Non-plastidial glycerolipid were mostly impacted by MGDα-eGFP-B line, although MGDα-eGFP-A had again little effect on the distributions. MGDβ and MGDγ overexpressions led to small and similar variations.

Following MGDβ and MGDγ overexpressions, there was a decrease of all C14-containing species in non-plastidial lipids (**Fig. 4.27**), except the 14:0-16:0 in DAG.

In the MGDβ overexpressors, there was a decrease of 16:1-18:3, 20:4-20:5, and 20:5-20:5, and an increase in 16:0-16:1, 16:1-18:1, 20:5-22:5, and 20:5-22:6 in all non-plastidial glycerolipids containing these species. Additionally, there were an increase of 20:5-16:0 in PC but not in DAG, an increase in PC 18:2-18:1 and 18:2-18:2, and a decrease in PC 16:1-18:4 and 20:5-18:4.

In the MGDγ overexpressing lines, there was an increase in 16:1-18:1 and 20:5-22:6, and a decrease in 20:5-16:1 in all non-plastidial glycerolipids containing these species. The DGTA 20:5-22:5 increased while the PE 20:5-22:5 decreased. There was an increase in 18:2-18:1 and 18:2-18:2 in PC, and an increase in 16:0-16:0 and 16:0-16:1 in DAG.

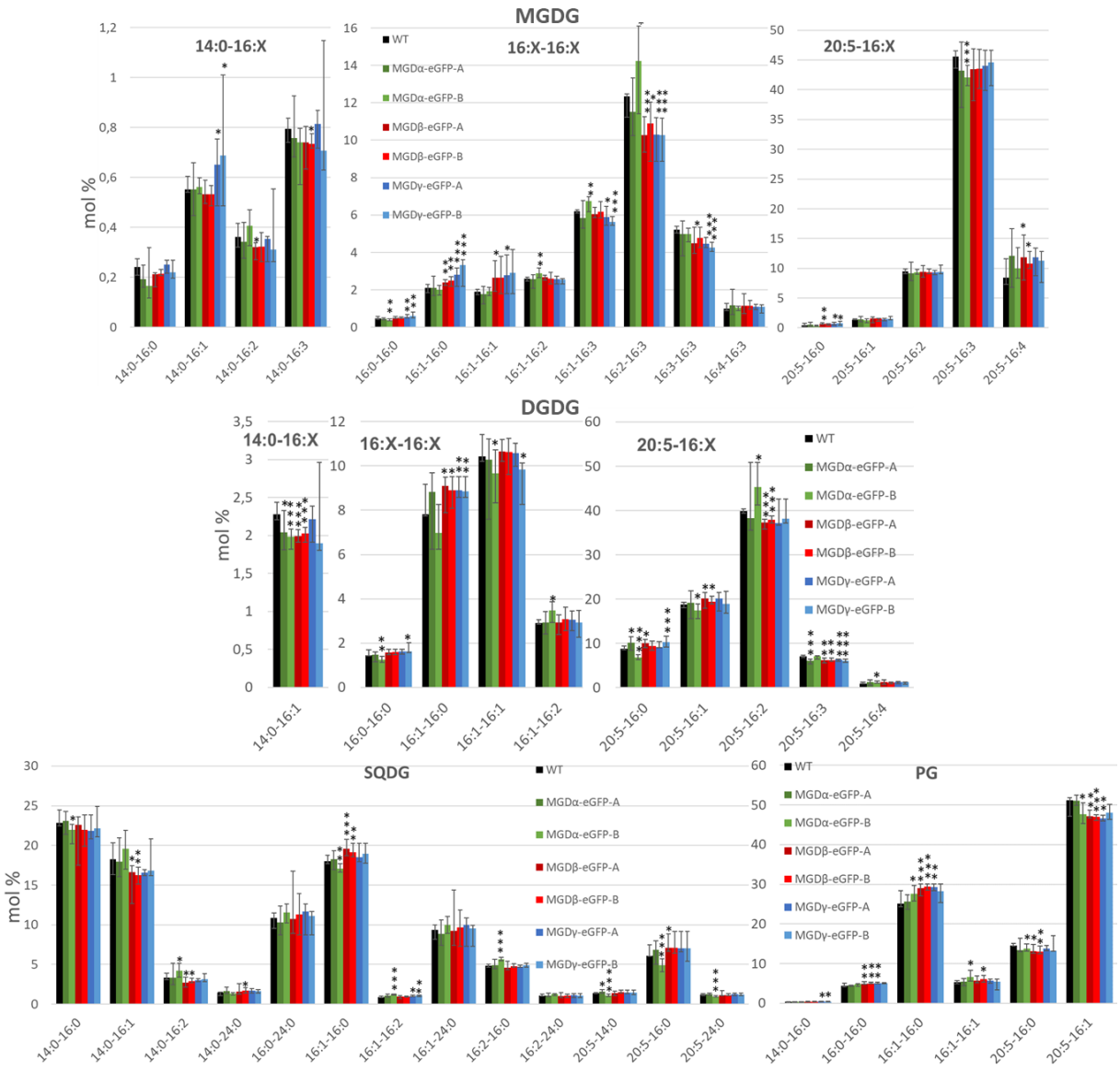


Figure 4.26: Impact of MGDs overexpression on the molecular species constituting the main plastidial glycerolipids in *P. tricornutum*. Lipid molecular species in MGDG, DGDG, SQDG and PG were analysed by LC-MS/MS. Each result is the median of six biological replicates ± min and max values. Significant differences are shown by one asterisk for 0.05 > p-value > 0.01, two asterisks for 0.01 > p-value > 0.001, and three asterisks for 0.001 > p-value, and were calculated by an unpaired multiple t test using GrahPad Prism software.

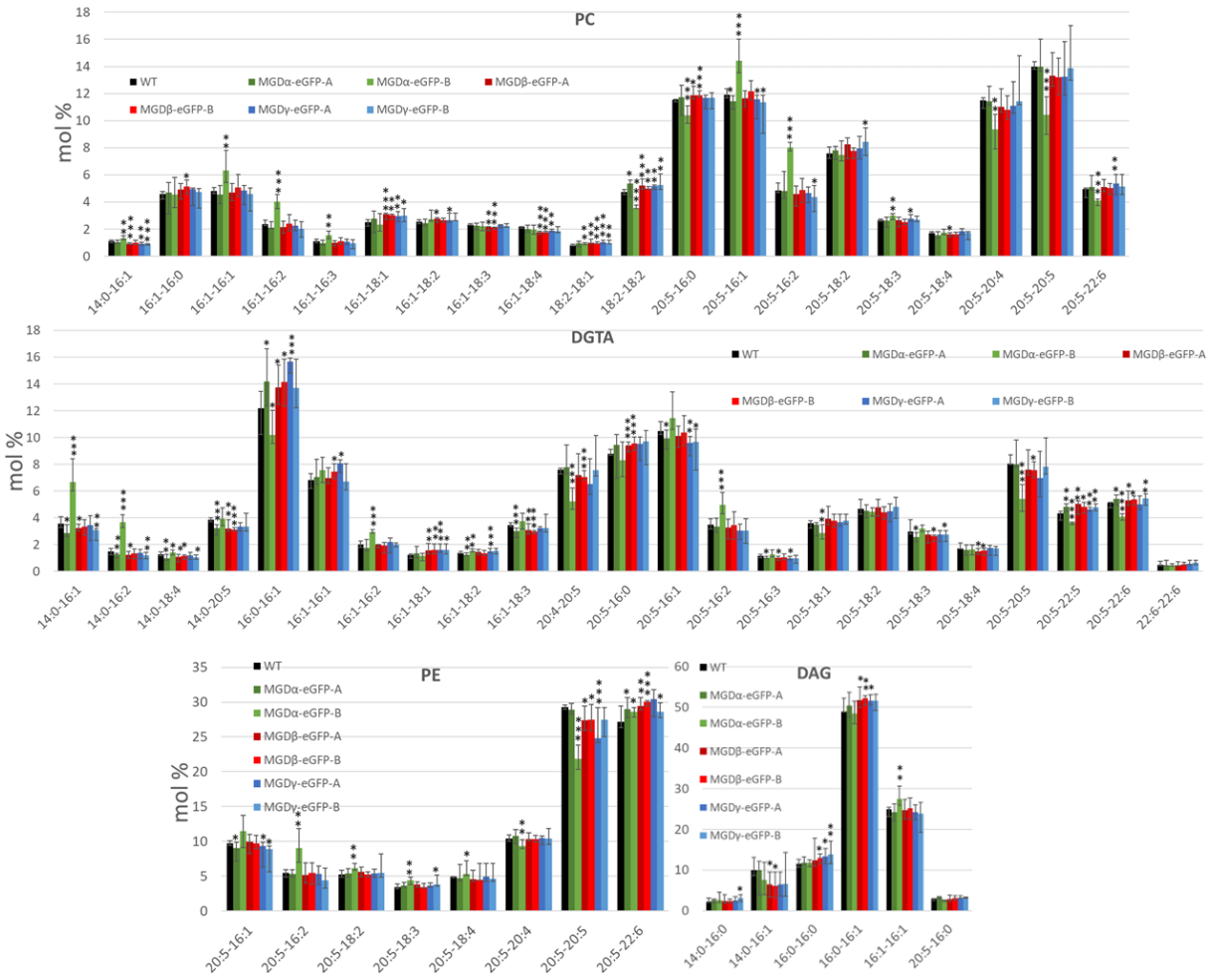


Figure 4.27: Impact of MGD overexpression on the molecular species constituting non-plastidial membrane glycerolipids in *P. tricornutum*. Lipid molecular species in PC, DGTA, PE, and DAG were analysed by LC-MS/MS. Each result is the median of six biological replicates ± min and max values. Significant differences are shown by one asterisk for 0.05 > p-value > 0.01, two asterisks for 0.01 > p-value > 0.001, and three asterisks for 0.001 > p-value, and were calculated by an unpaired multiple t test using GrahPad Prism software.

4.4.2.4. No compensation at the transcriptomic level following the loss of an MGD

Given the moderate changes in phenotype at the level of lipid class relative proportions and the low impact on growth and photosynthesis in the knockout mutants, qPCR analyses were performed on the MGD mutants and a WT cultured in parallel to check for a possible compensation effect at the transcriptomic level between *MGD* genes.

A first observation was that in the WT, *MGDβ* and *MGDγ* were expressed at similar levels, and that MGDα was around 4 times more expressed that *MGDβ* and *MGDγ* (**Fig. 4.28**). In the culture of MGDβ mutants in parallel with a WT, a much lower MGDα expression was observed, with standard deviations being relatively high. Therefore, there might have been a problem either of culture condition or during RNA and cDNA obtention in this sample.

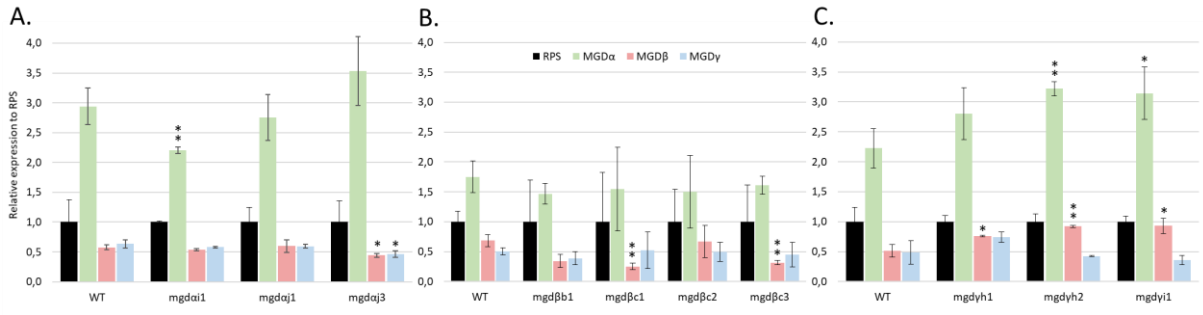


Figure 4.28: *MGDs* relative gene expression in the mutant lines compared to the WT. Each graph represents the relative gene expression of each *MGD* gene in the mutant lines of *MGDα* (A), *MGDβ* (B), and *MGDγ* (C), compared to a WT grown in parallel, normalised to *RPS* housekeeping gene expression. Each result is the average of three biological replicates ± standard error. Significant differences are shown by one asterisk for 0.05 ≥ p-value > 0.01, two asterisks for 0.01 > p-value > 0.001, and three asterisks for 0.001 > p-value, and were calculated by an unpaired multiple t test using GrahPad Prism software.

No compensation effect could be detected in the mutants of *MGDα* and *MGDβ* (**Fig. 4.28**). In MGDγ mutants, an increase of *MGDα* and *MGDβ* gene expression by 1.37 and 1.69-fold on average, respectively, was detected. This indicates that there might be a compensation mechanism when MGDγ is absent.

Another phenotype was the decrease in *MGDβ* gene expression in *MGDβ* knockout mutants by 2.25-fold on average (p-value for *mgdβb1* was 0.05028), which was not observed in the control mutant *mgdβc2* (**Fig. 4.28**). The decreased level of mRNA of the mutated gene could be linked either to the recognition of the mutated RNA as aberrant RNA that would then be degraded by the cellular machinery, or to a positive regulation loop of MGDβ on itself.

In conclusion, MGDα appeared to be the main MGD isoform under optimal condition in the WT. Only *MGDγ* knockout seemed to trigger a moderate compensation effect at the transcriptional level of the two other MGD. If a compensational effect exists in the absence of MGDα and MGDβ, it might be at the protein level and involve mechanisms such as protein degradation system and protein relocalisation.

4.4.3. Nutrient starvation

The diversification of the MGD isoforms in diatoms might play a role for environmental stress adaptation. As nitrogen deprivation is a common stress model applied to microalgae, and that a decrease in MGDα and MGDγ proteins has been detected previously following a removal of nitrogen in the culture medium (Lupette et al., 2019), we used this condition to test the effect of each MGD mutation on *P. tricornutum* phenotype.

In the whole-cell proteome analysis performed previously in our team by Josselin Lupette, protein levels of MGDα and MGDγ were found to be reduced to 6.1 % and 26.3 %, respectively, compared to their levels under optimal growth condition. Protein levels of MGDβ showed a slight increase to 127.0% of its usual level. The experiment had been realised with the same medium composition as in our experiment, however algae were cultured in a volume of 3 L and were harvested after seven days. In our experimental context, we cultivated *P. tricornutum* cells in 50-mL volume, and harvested the cell earlier, after 5 days of deprivation. As cells tend to grow faster in a small volume, both conditions might still be relatively comparable.

4.4.3.1. MGDs' expression following nitrogen limitation

We verified if nitrogen limitation led to a decrease in mRNA expression of *MGDα* and *MGDγ*. We measured by qPCR the expression of each *MGD* genes at different days of cultivation in 1N10P medium compared to cultivation in 10N10P medium.

Note that for this experiment, the 1N10P condition was used instead of a more severe stress in 0N10P medium as mRNAs tend to degrade rapidly in 0N10P, greatly lowering the quality of qPCR analysis. The cDNA samples were kindly provided by Elodie Armanet, intern in our team at the time of the experiment.

The housekeeping genes RPS, HPRT and TUBA were initially used as references for qPCR analyses, but only the RPS gene was found relatively stable and was subsequently selected for gene expression level normalization.

At day 1, the expression of all *MGD* genes was highly increase in 1N10P medium (data not shown) compared to the 10N10P medium. However, after only one day, the effect of nitrogen limitation in 1N10P should not have appeared yet. The relative expressions of each MGD compared to RPS were also very different to analyses performed in rich media. Standard deviations for all data were rather high. Therefore, RNA quality might be poor and unfit for comparative analysis, even in the 10N10P condition.

In conclusion, given the quality of the cDNA samples, any moderate change in MGD genes expression during the first 6 days of nitrogen limitation may have been missed. No drastic change of MGD genes expression was detected, but this experiment should be repeated.

4.4.3.2. Effect of nitrogen deprivation on growth and non-polar lipid accumulation

A selection of two knockout mutants per MGD were used for the analysis of the impact of nitrogen deprivation: *mgdαi1*, *mgdαj3*, *mgdβb1*, *mgdβc1*, *mgdγh2*, and *mgdγi1*. Cells grown in nutrient-replete medium (enriched ESAW 10N10P) were shifted to nitrogen-deprived medium (0N10P) at a starting cell concentration of 3 to 3.5 10^6 cells.mL^{-1}.

Growth rate was monitored at day 0, 3, 4 and 5 of nitrogen deprivation (**Fig. 4.29**). Nitrogen deprivation condition did not lead to complete growth arrest during the course of our experiment as previously observed (Abida et al., 2015), but cells divided more slowly. Cultures kept growing between day 3 and day 5, with WT cells reaching a concentration of around 9.10^6 cells.mL^{-1} at day 5, which is about half of the concentration that would be expected in 10N10P condition. Growth rate did not look affected in MGDα and MGDβ mutants (**Fig. 4.29**), while growth was lower in both MGDγ mutants. This observation is similar to that made in 10N10P condition (**Fig. 4.14 A-D**).

The TAG accumulation in response to nitrogen deprivation was monitored by Nile Red staining at days 3, 4, and 5 of deprivation. However, the amounts of TAG reached at day 5 were too high to be reliably measured by Nile Red fluorescence reaching a saturation level. We observed that TAG accumulation was significantly faster in MGDβ mutants compared with the WT (**Fig. 4.30**). This could be related to MGDβ mutants being more stressed than the WT, or to TAG synthesis pathways being impacted by the loss of MGDβ.

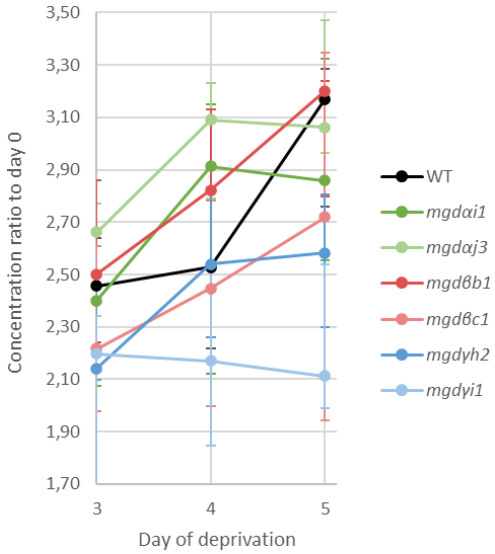


Figure 4.29: *P. tricornutum* growth during nitrogen limitation. Cell concentration in the mutant strains cultured with a WT in parallel was monitored in two separate experiments. The algae were cultured at 20 °C in ESAW 0N10P medium, and the cell concentration in 250-mL flasks was measured after 0, 3, 4 and 5 days at the same hour using a TECAN Infinite M1000 PRO. Cell concentration increase compared to concentration at day 0 were plotted. Data are the median of six biological replicates. Error bars represent minimal and maximal values.

Cells stained with Nile Red at day 5 of nitrogen deprivation were observed by epifluorescence microscopy (**Fig. 4.31**). All strains were fusiform and showed large lipid droplets. Plastid size had shrank compared to cells grown in 10N10P condition (**Fig. 4.16**), consistent with the senescence of plastid membranes previously observed during nitrogen starvation (Abida et al., 2015). During the observation of the *mgdγi1* mutant, a lipid droplet was seen 'popping' out of a cell (**Fig. 4.31**, second set of images for *mgdγi1*) and a few lipid droplets were detected in the medium, possibly reflecting a higher fragility of *mgdγi1* mutant cells under these conditions. During observation, no striking difference in cell, plastid and lipid droplet size could be noticed between the WT and the mutant lines.

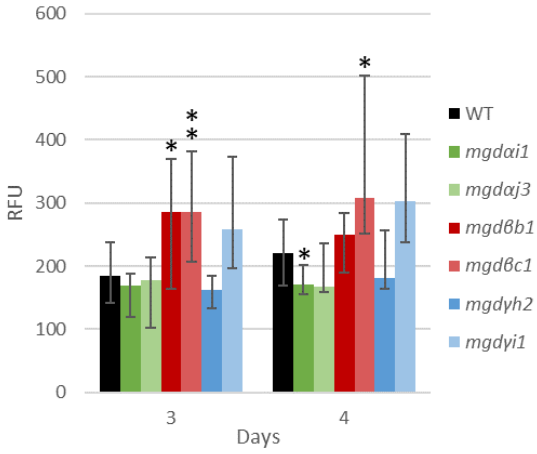


Figure 4.30: Accumulation of non-polar lipids during nitrogen limitation. *P. tricornutum* WT and mutant strains were cultured in ESAW 0N10P medium. Non-polar lipid accumulation was measured after 3, and 4 days by Nile Red staining using a TECAN Infinite M1000 PRO, and expressed as fluorescence intensity normalized by cell number (RFU, relative fluorescence units). Data are the median of three biological replicates. Error bars represent minimal and maximal values. Significant differences are shown by one asterisk for 0.05 > p-value > 0.01, and two asterisks for 0.01 > p-value > 0.001, and were calculated by an unpaired multiple t test using GrahPad Prism software.

4.4.3.3. Impact of Nitrogen shortage on the fatty acid and glycerolipid content of MGD mutants

After 5 days of nitrogen deprivation, cells were harvested for glycerolipid analysis. TAG content reached 4.10 nmol.10^{-6} cells after 5 days of nitrogen deprivation, which represented around 60% of total glycerolipid content (**Fig. 4.32B**). As a result, total FA profile became closer to the composition of TAG molecular species (**Fig. 4.32A**). TAG molecular species are rich in 16:0 and 16:1 FA, which became the dominant species in the FA profile.

When considering membrane glycerolipid only, the main differences in the glycerolipid profile under nitrogen starvation compared to our nutrient replete condition were an increase in DGDG and DAG proportions, and a decrease in PG and PE (**Fig. 4.32B**). However, our two conditions cannot be further compared given that cultivation time and starting cell concentration were distinct.

Although Nile Red staining indicated that TAG might have accumulated faster in the MGDβ mutants, TAG content was rather similar to the WT at day 5. In MGDγ mutant lines however, TAG accumulation tended to be lower.

Consistently with a very low MGDα protein level in the WT under nitrogen deprivation, no changes could be detected in the proportions of membrane glycerolipids in MGDα mutants compared to the WT.

For MGDβ mutants, a slight increase in PE and PG, and a very small decrease in PI were observed.

The changes of membrane glycerolipid distribution in MGDγ mutants were more striking. There were a lower proportion of DGDG (by 22.5% on average) and DAG (by 9.9 % on average) proportions, and an increase in PC (by 27.0 % on average) and DGTA (by 56.6 % on average) compared to the WT. The decrease in DGDG proportion compared to the WT was roughly equivalent to the increase in PC. In terms of nmol.10^{-6} cells, the decrease of DGDG seen in mol% of total membrane glycerolipid was not significant. Therefore, reasoning in mol% or in nmol.10^{-6} cells about DGDG species resulted in the same observations (**Fig. 4.33 & Fig. 4.35**).

MGDG acyl composition was only slightly impacted in MGDα knockout mutants (**Fig. 4.33**). This highlights a dispensable role of MGDα in nitrogen shortage.

A decrease of 14:0-16:3 and an increase in 16:3-16:3 and 16:4-16:3 species were observed. Consistently, 16:3-16:3 and 16:4-16:3 species were the only stable species containing a C16 in position sn-1 in optimal growth condition while the other decreased, except for 16:1-16:1 species which had shown a slight increase (**Fig. 4.20**). In DGDG, there were an increase in 16:1-16:1 and a decrease in 20:5-16:0 and 20:5-16:1.

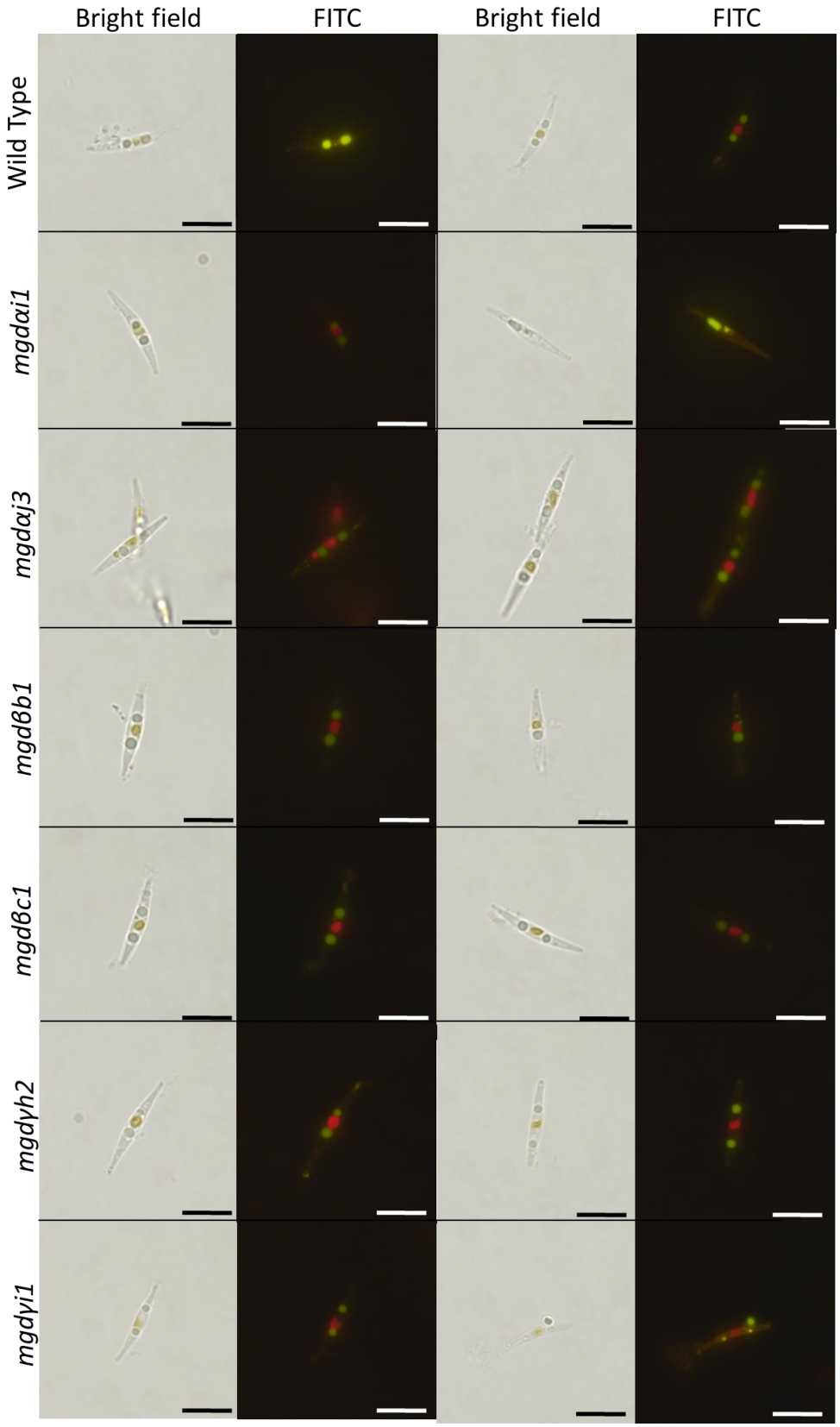


Figure 4.31: Observation of cell morphology and TAG accumulation by epifluorescence microscopy. *P. tricornutum* WT and mutant strains were observed with an epifluorescence microscope after 5 days of culture in ESAW 0N10P medium. Two sets of images were taken per strain. Images were taken with an oil-immersed objective 100x. Chlorophyll autofluorescence and Nile Red staining were observed using a FITC filter. Scale bar: 10 µm.

By contrast, MGDG molecular species composition was severely affected in MGDβ and MGDγ mutants (**Fig. 4.33**). This highlights a likely role for those two isoforms upon nitrogen stress.

In MGDβ mutant lines, a strong decrease in 14:0-16:2 and 14:0-16:3 was observed. The decrease was such that these species were detected only in one replicate in the mutants, and not at all in the case of 14:0-16:2 in *mgdβb1*. Although MGDG 14:0-16:1 seemed stable, a decrease in DGDG 14:0-16:1 was observed. MGDG species containing a C16 FA in position *sn-1* and a maximum of one unsaturation on either of its two FAs were increased. However, DGDG 16:1-16:1 was decreased. Then, MGDG 16:1-16:2, 16:1-16:3 and 16:2-16:3 species were strongly decreased while 16:3-16:3 and 16:4-16:3 were strongly increased. All 20:5-containing MGDG species were decreased, except the 20:5-16:4 that strongly increased. In DGDG, only 20:5-16:2 and 20:5-16:4 species showed a significant increase, while other 20:5-containing species were stable.

Changes in galactolipid species distribution were the most spectacular in MGDγ mutants (**Fig. 4.33**). MGDG and DGDG species containing a 16:0 or a 16:1 in position *sn-2* and either a 14:0 or a C16 in position *sn-1* were severely decreased. On the contrary, species with a 16:2 or 16:3 in position *sn-2* were increased. 20:5-containing galactolipid species were mainly increased, except the 20:5-16:4 in *mgdγi*.

In other plastidial lipids, the impact of MGDα mutation was moderate (**Fig. 4.33**), while MGDβ and MGDγ mostly impacted PG in an opposite way.

Very small but highly significant increases in SQDG 16:2-16:0 and 16:2-24:0 were detected in MGDα mutants, while there was a small decrease in SQDG 20:5-14:0.

In MGDβ knockout lines, there were a decrease in the proportion of 14:0-16:0, 14:0-16:1, 20:5-16:0 and 20:5-16:1 in both SQDG and PG. Increased species were the 16:1-16:0 in both SQDG and PG, the 16:1-16:2, 16:2-16:0 and 16:2-24:0 in SQDG, and the 16:0-16:0 and 16:1-16:1 in PG.

MGDγ mutation had no impact on SQDG. However, the proportion of PG species containing a C16 FA in position *sn-1* decreased significantly while 20:5-containing species increased. It seems that the impacted PG species were affected similarly as the galactolipid species with the same FA composition.

Non-plastidial glycerolipid were mostly unimpacted by MGDα mutation (**Fig. 4.34**). Both MGDβ and MGDγ mutations influenced non-plastidial molecular species distributions, with a particularly strong impact of MGDγ mutation on a couple of species in DGTA and TAG.

In PC, there were a decrease in the proportion of 16:1-18:1 in MGDα mutants, and an increase in 18:2-18:2 and 20:5-16:0. In DGTA, there was a decrease in 16:1-16:2 and 22:6-22:6 species, and an increase in 20:4-20:5 and 20:5-20:5. No changes were detected in PE and DAG.

In MGDβ mutants, PC 16:1-18:1,16:1-18:2 and 20:5-18:2 species were increased while PC 18:2-18:2 and 20:5-16:0 decreased (**Fig. 4.34**). In DGTA, 16:0-16:1, 16:1-16:1 and 16:1-18:1 increased while 20:4-20:5, 20:5-22:6 and 22:6-22:6 decreased. In PE, only 20:5-20:4 tended to decrease. In DAG, there was a small decrease in 14:0-16:1 compensated with a slight increase in 16:0-16:1.

For MGDγ mutants, it should first be reminded that both PC and DGTA content were largely increased in these lines (**Fig. 4.34B**). As a consequence, all PC and DGTA species content were increased compared to the WT (**Fig. 4.35**), with the exception of DGTA 22:6-22:6, which was decreased in *mgdγh2* and never detected in any *mgdγi1* replicate. However, the increase in PC and DGTA was not accompanied with a conserved lipid species distribution in either lipid class.

In both PC and DGTA, there was an increase in the proportion of 14:0-16:1, 16:1-16:0 or 16:0-16:1, and 16:1-16:1 species. Interestingly, the proportion of these species were decreased in galactolipids. The proportion of 16:1-16:2 was increased in PC, but only increased in *mgdγh2* in DGTA.

In PC, 20:5-16:0 tended to decrease while DGTA 20:5-16:0 increased. Species with the composition 20:5-16:1 were increased in PC and DGTA in both mutants, and in PE for *mgdγh2* line only. The 20:5-20:5 species of PC, DGTA and PE were all strongly decreased in proportion.

In PC only, 16:1-18:2, 16:1-18:3 and 20:5-16:2 increased in proportion.

In DGTA, 14:0-20:5, 20:4-20:5, 20:5-18:1, 20:5-18:2, 20:5-22:5 and 22:6-22:6 species were decreased.

Overall, it seemed that more 20:5-containing species were decreased in proportion than increased, which is consistent with a higher proportion of 20:5-containing species in galactolipids.

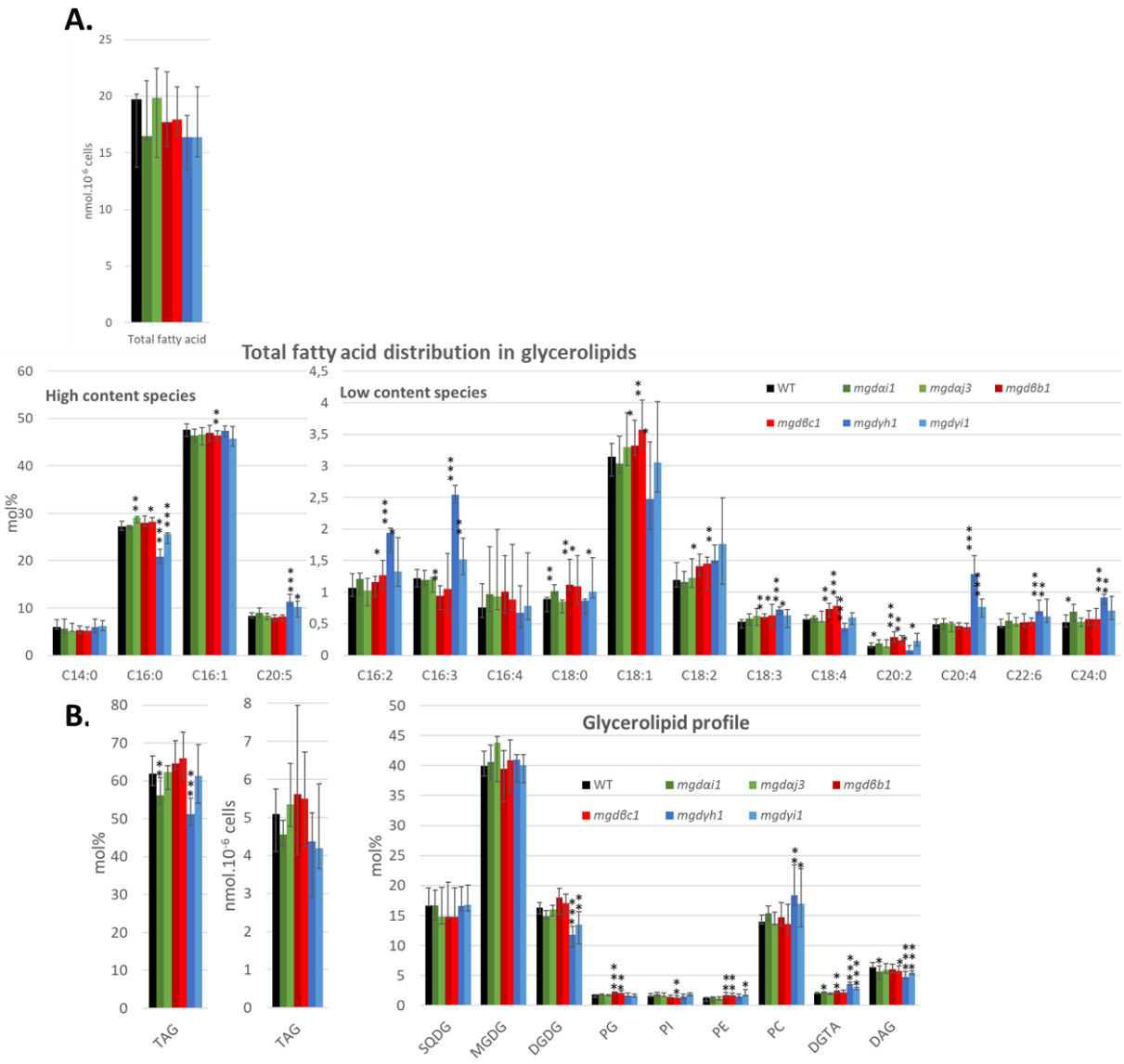


Figure 4.32: Quantitative analysis of FA and glycerolipid content in MGD mutant lines. Lipids from *P. tricornutum* MGD mutant lines and the WT grown in 0N10P medium were extracted, and separated as described in 'Materials and Methods'. **A**, total FA content (given in nmol.10^{-6} cells) and global FA profile (given in molar percentage) in total glycerolipid extracts. **B**, TAG content (in nmol.10^{-6} cells and molar percent of total glycerolipids), and glycerolipid profile (in molar percent without including TAG). Each result is the median of six biological replicates ± min and max values. Significant differences are indicated with one asterisk for 0.05 > p-value > 0.01, two asterisks for 0.01 > p-value > 0.001, and three asterisks for 0.001 > p-value, and were calculated by an unpaired multiple t test using GrahPad Prism software.

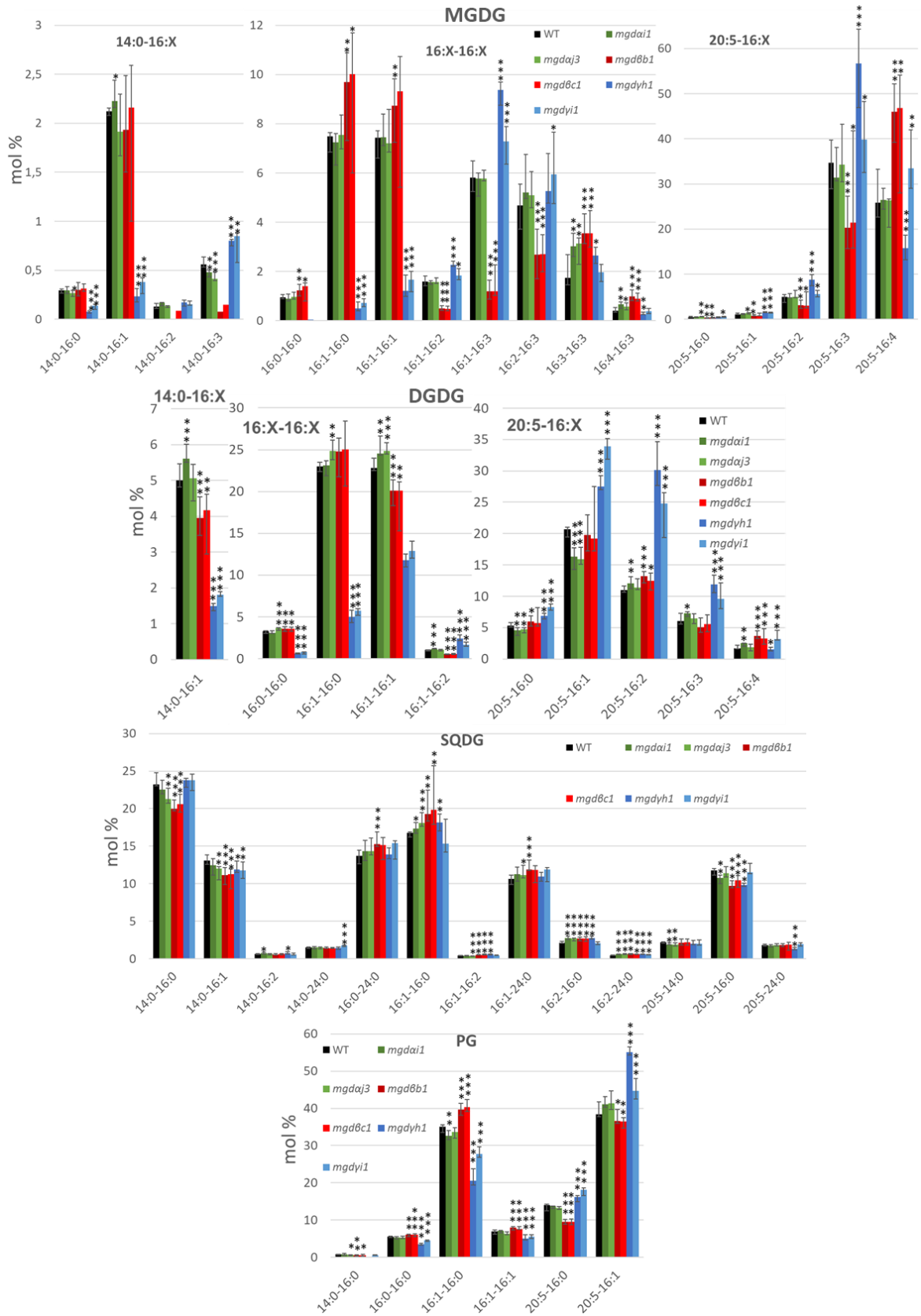


Figure 4.33: Impact of MGD mutations on the molecular species constituting the main plastidial glycerolipids in nitrogen-deprived *P. tricornutum* cells. Lipid molecular species in MGDG, DGDG, SQDG and PG were analysed by LC-MS/MS. Each result is the median of six biological replicates ± min and max values. Significant differences are shown by one asterisk for 0.05 > p-value > 0.01, two asterisks for 0.01 > p-value > 0.001, and three asterisks for 0.001 > p-value, and were calculated by an unpaired multiple t test using GrahPad Prism software.

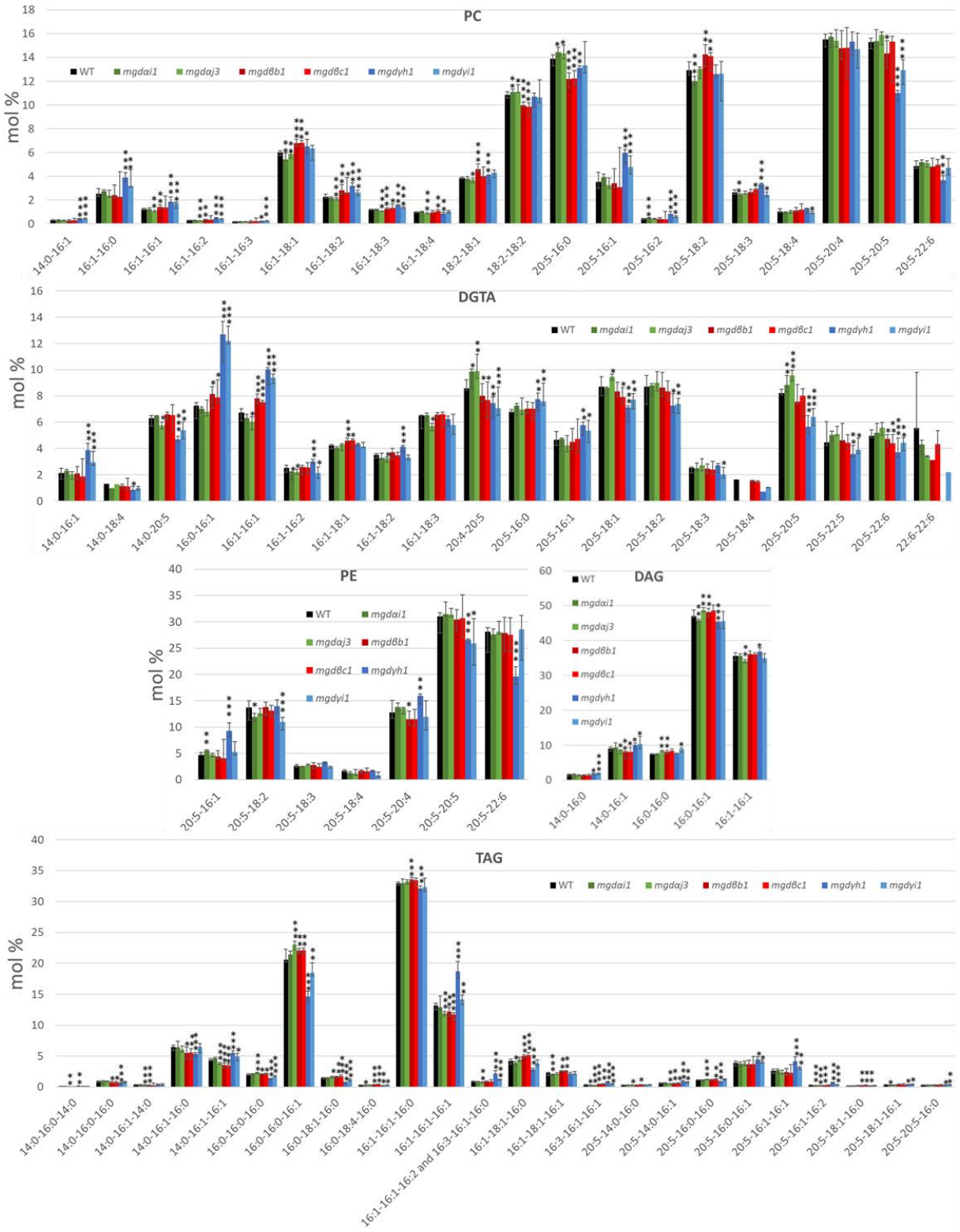


Figure 4.34: Impact of MGD mutations on the molecular species constituting non-plastidial membrane glycerolipids in nitrogen-deprived *P. tricornutum* cells. Lipid molecular species in PC, DGTA, PE, DAG and TAG were analysed by LC-MS/MS. Each result is the median of six biological replicates ± min and max values. Significant differences are shown by one asterisk for 0.05 > p-value > 0.01, two asterisks for 0.01 > p-value > 0.001, and three asterisks for 0.001 > p-value, and were calculated by an unpaired multiple t test using GrahPad Prism software.

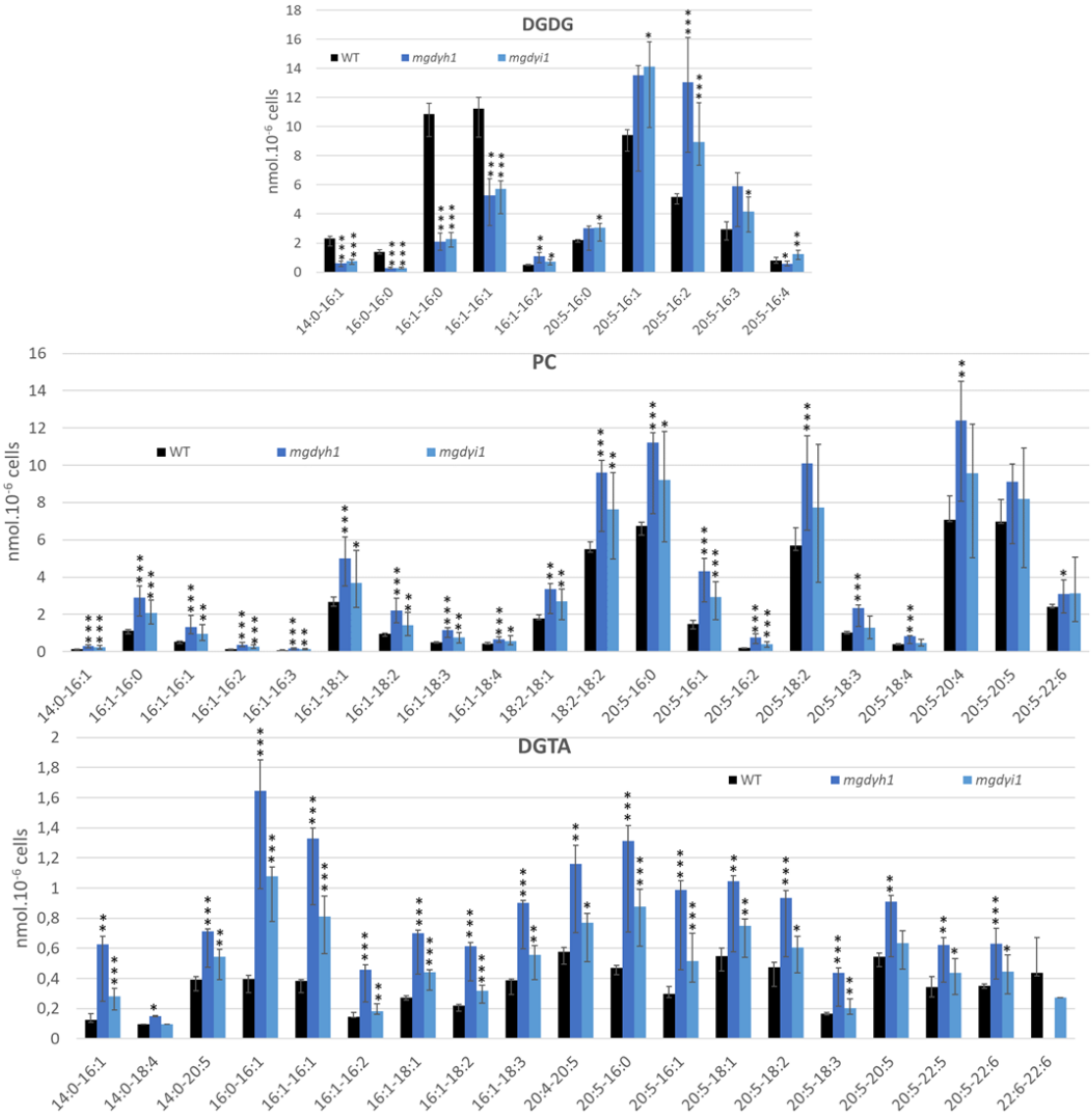


Figure 4.35: Impact of MGDγ mutations on the molecular species constituting DGDG, PC and DGTA in nitrogen-deprived *P. tricornutum* cells. Lipid molecular species in DGDG, PC, and DGTA were analysed by LC-MS/MS. Each result is the median of six biological replicates ± min and max values. Significant differences are shown by one asterisk for 0.05 > p-value > 0.01, two asterisks for 0.01 > p-value > 0.001, and three asterisks for 0.001 > p-value, and were calculated by an unpaired multiple t test using GrahPad Prism software.

Altogether, the impact of MGDα mutation during nitrogen deprivation was very weak, which is consistent with a very low MGDα protein level in the WT. Considering that MGDα is nearly inactive upon nitrogen starvation, the mutations of either MGDβ or MGDγ led to a stronger phenotype than in enriched ESAW medium as there was mainly only one active MGD left in the cell after each mutation. In these conditions, MGDγ seemed to play a very important role on galactolipids with a low unsaturation level, and MGDγ mutation tended to have a more important effect on non-plastidial lipids than MGDβ mutants. The increase in DGTA 16:0-16:1 and 16:1-16:1 in MGDγ mutants was particularly striking, and highlights a role for MGDγ outside the plastid. MGDβ looked important to keep high level of polyunsaturated galactolipids, which therefore did not impact so much DGDG composition, and seemed to be more important with plastidial glycerolipids than MGDγ. Interestingly, MGDβ and γ mutations had opposite effects on PG, which is a lipid class encountered in both plastidial and extra-plastidial membranes.

5. General Discussion and Conclusion

5.1. *Phaeodactylum* contains three confirmed and non-redundant MGD enzymes

Phylogenetic analysis of MGD proteins in multiple phyla revealed that MGDs in diatoms represent a polyphyletic group, similar to what is observed in the Viridiplantae between an ancestral type A and a type B MGDs having emerged in Angiosperms (**Fig. 4.1**). In diatoms, three major types of MGD were identified, originating from gene duplications and specialisation. An ancestral MGD might have been vertically inherited in the stramenopile group following the engulfment of a red algae at the origin of the MGDα isoform. Then, a first gene duplication occurred, potentially following the separation of the eustigmatophyceae and phaeophyceae from the other stramenopiles. This first gene duplication would have led to the emergence of a type of MGD distinct from MGDα, at the origin of the β and γ isoforms. A second gene duplication potentially followed the separation of the pelagophyceae from the diatoms. Among the diatoms selected in our phylogenetic analysis, three out of five species contain the three types of MGD, including *P. tricornutum*, while the two others only contain two types of MGD. There is likely a selection pressure in diatoms for the conservation of two to three different types of MGDs. However, in other species close phylogenetically to diatoms such as *M. gaditana*, only one MGD was detected. Both diatoms and *M. gaditana* originate from two endosymbiosis events, and contain four membranes surrounding the plastid. Therefore, presence of two to three MGDs in diatoms cannot be solely justified by their complex membrane architecture.

We focused on the three MGDs present in the *P. tricornutum* model. *In silico* analysis and manual curation allowed us to confirm the correct start methionine of each preprotein (before cleavage of targeting peptides) (**Fig. 4.2**), and revealed that MGDα was not correctly annotated in available online resources. Indeed, it was clear following the search of targeting peptides in the N-terminal part of the protein that MGDα sequence begins at a methionine located downstream of the two previously suggested start methionines. We also identified highly conserved amino acids between *P. tricornutum* and *A. thaliana* MGDs, including the two catalytic residues previously characterised in AtMGD1 (Botté et al., 2005; Dubots et al., 2010; Rocha et al., 2016). The high similarities between the active sites of *P. tricornutum* MGDs and AtMGD1 strongly suggest that these proteins are functional MGDs, which was confirmed with the lipid profile analysis of MGD mutants (**see section 4.4.2.3**). Analyses of predicted protein models suggested that *P. tricornutum* MGDs present the classical GT-B fold, and that the proteins regions are globally similar with AtMGD1 (**Fig. 4.5**). The main difference between the proteins was the size of the N and C-domains of MGDβ and MGDγ that both contain additional residue extensions. In the C-domain, MGDβ and MGDγ contain 50 and 110 additional amino acids, respectively. These two regions appeared to be essentially hydrophilic, and could therefore influence protein solubility (**Fig. 4.6**). In the N-domain, MGDβ and MGDγ contain 82 and 81 additional amino acids, respectively, upstream of the first conserved amino acids with MGDα and downstream of a signal peptide in the case of MGDβ. Therefore, MGDβ and MGDγ mature proteins are relatively bigger than MGDα. The role of these large extensions is unknown.

MGDα is predicted to contain a clear bipartite presequence suggesting its localisation inside the plastid, which was confirmed by protein subcellular localisation using fluorescent protein fusions (**Fig. 4.7**). Its cleavage site,

identified using SignalP-6.0 server, harbours a canonical ASAFAP motif: 'SAAFSP' (Kilian and Kroth, 2004). The eGFP signal observed for the recombinant MGDα-eGFP protein supports its localisation to the thylakoid membranes. However, the spot pattern observed is quite unusual and could mean that the protein localises to subdomains in the thylakoids, or that overexpression led to the formation of protein aggregates inside the stroma of the plastid. In MGDβ, a weak prediction of a signal peptide was obtained, for which the cleavage site was weakly similar to the ASAFAP motif: 'GSGFVL'. Quite consistently, protein localisation showed that MGDβ localises to the blob-like structure between the second and the third membrane surrounding the plastid (**Fig. 4.7**). Therefore, MGDβ successfully crosses the EpM and PPM membranes but is not transported through the final two inner membranes of the plastid, which correspond to the chloroplast envelope membranes in primary plastids. Failure to cross the oEM and iEM might be linked to a tuning of the ASAFAP motif. Additionally, MGDβ localisation could extend to the PPM, as the vesicular network present in the blob-like structure was seen as emerging from the PPM (Flori et al., 2016). Finally, no targeting sequence could be predicted in MGDγ, which is consistent with an MGDγ-eGFP signal outside of the plastid, in the ER/EpM and possibly the cytosol (**Fig. 4.7**). A localisation to the cytosol is surprising given that MGD proteins are classically localised on a membrane, and an intense cytosolic labelling might result from the overexpression system. Indeed, MGDγ could be localised to the ER and EpM, and be solubilised by the eGFP fusion and accumulate in the cytoplasm following overexpression. This result is also coherent with the detection of MGDγ in two replicates of the LD proteome of Leyland *et al* (Leyland et al., 2020), following potential contamination by the EpM. We do not exclude that MGDγ is partitioned between the ER, EpM and the cytosol. AtMGD2 was recently shown to localise to the cytosol of elongating pollen tubes (Billey et al., 2021). Thus, MGDγ could also be localised to the cytosol, suggesting a possible gain of function of this protein. It could possibly dock on plastidial or even non-plastidial membranes to perform MGDG synthesis activity. Besides, the additional protein regions present in its N- and C-domains might also be linked to a gain of function. Localisation of MGDγ to the ER raises other questions, such as the possible presence of MGDG in the ER, and the binding of this enzyme to the membrane. Indeed, it is known that MGD enzymes have an affinity for their product MGDG. If there is no MGDG in the ER, that would mean that MGDγ has distinct binding properties than other MGDs, or that the presence of PE in the ER, which is also a HexII lipid, allows the binding.

Therefore, each MGD protein in *P. tricornutum* localises to a distinct subcellular compartment. MGDα, which is the closest to the ancestral MGD, was localised inside the thylakoids. Given that other plant MGDs are generally localised to the chloroplast envelope, it shows that this MGD 'migrated' toward the inside of the plastid. In contrast, MGDβ and MGDγ are localised in membranes that do not derive from the chloroplast envelope. Therefore, for these two MGDs, a relocalisation outside of the endosymbiont plastid took place.

Noteworthy, we detected one MGD with a conserved active site in a basal apicomplexa, *Gregarina niphantrodes*. Apicomplexa lost their ability to perform photosynthesis, but most of them remained with a vestigial plastid called the apicoplast. In endoparasitic *Toxoplasma gondii* and *Plasmodium falciparum*, which contain an apicoplast, MGDG synthesis could not be demonstrated and no MGD could be predicted (Maréchal et al., 2002). In basal apicomplexa, gregarines constitute a clade of mesoparasites which has evolved independently. *G. niphantrodes* does not seem to contain an apicoplast nor a plastid genome (Toso and Omoto, 2007). Looking

more closely at the MGD gene of *G. niphantrodes*, we observed that there was no targeting peptide and that it contains large additional regions compared to other MGD, which structurally brings it closer to MGDγ and MGDβ. Therefore, this apicomplexan contains another example of an MGD found outside of a plastid, possibly in the cytosol. Future work is needed to characterise this uncommon non-plastidial MGD in gregarines.

Presence of an MGD outside the plastid strongly suggests that MGDG in diatoms (and in basal apicomplexa) can be found in extraplastidial membranes, like the endomembrane system. The EpM and PPM lipid compositions, previously expected to be mainly related to the ER lipid composition, may additionally contain galactolipids. These findings imply that MGDG probably plays new and undiscovered roles in non-plastidial membranes in these organisms.

It is generally accepted that MGDG is mostly localised in the thylakoid membranes and acts as a key building block for photosynthetic membranes. Indeed, for a plastid with a diameter of 1 µm and a surface of 3.14 µm², the thylakoid membranes can reach 50-70 µm² depending on the thylakoids membranes arrangement (Guéguen and Maréchal, 2021). Furthermore, in primary plastids, MGDG represent 50 % of total glycerolipids in the thylakoids and iEM (Block et al., 1983; Mendiola-Morgenthaler et al., 1985; Boudière et al., 2014), while the oEM only contain 20% of MGDG. If we consider that the iEM and oEM have the same surface, then more than 90 % of MGDG is localised in the thylakoids. However, in *P. tricornutum*, the plastid is surrounded by four membranes, and the localisation of the MGDs suggest that all four membranes contain MGDG. Besides, extra-plastidial membranes might also contain MGDG. Therefore, the MGDG species distribution profile in *P. tricornutum* may not exactly reflect the MGDG species distribution in thylakoids membranes.

Observation of MGDβ-eGFP in the blob-like structure between the PPM and oEM also gave insightful information about the positioning of this structure, particularly during plastid division (**Fig. 4.8**). The MGDβ-overexpressing system might not be ideal to draw conclusion about the normal behaviour of the blob, but similar positioning of this structure was also observed when eGFP protein was fused to a signal peptide and an incomplete transit peptide (Kilian and Kroth, 2004). Our observations of MGDβ-eGFP fluorescence by confocal microscopy and of the WT by TEM (**Fig. 4.8**) suggest that the vesicular network might take a blob-like structure mostly during cytokinesis, from the moment the plastid starts dividing until the whole cell has completed division. When the plastid does not show any sign of division, the MGDβ-eGFP fluorescence signal was more diffuse, either suggesting that the vesicular network itself loses the blob structure and become more scattered, or that MGDβ protein relocalises to the vicinity of the blob. During plastid division, the blob-like structure was located at the central constriction of the dividing plastid. Therefore, MGDβ could be particularly important during this step for the provision of MGDG to the growing and dividing plastid membranes. Following division, the two generated plastids seem to remain attached by the blob-like structure, meaning that the PPM and EpM are not fully divided, probably until the cleavage furrow comes through the blob (**Fig. 2.4**). This is particularly important as it would mean that the oNE-EpM continuum is never broken during plastid division, a point that was not addressed by Tanaka *et al* in their characterisation of plastid division (Tanaka et al., 2015a).

The distinct localisations of MGDs in *P. tricornutum*, are a first indication that these enzymes are not redundant. Furthermore, qPCR analysis of the relative expression of each isoform in the *P. tricornutum* WT strain showed

that MGDα seems to be the most expressed isoform under our optimal laboratory conditions compared to MGDβ and MGDγ, which relative expressions were around 4 times lower (**Fig. 4.28**). This could mean that MGDα is the main isoform under optimal conditions, consistent with its higher similarity to the MGDs from other phylogenetic lineages. Nonetheless, MGDα knochout mutants were only weakly impacted by the absence of this isoform. Furthermore, the relative protein level of MGDα was shown to drop to 6 % under nitrogen deprivation (Lupette et al., 2019). This is coherent with the decrease of thylakoid membranes observed in this stressful condition (Abida et al., 2015), and reveals that MGDβ and MGDγ might be sufficient to sustain the cell in nitrogen starvation. Therefore, we decided to address the question of what could be the respective roles of each isoform in optimal and stressful conditions. To this end, we generated CRISPR-Cas9 mutants of each MGD, and used MGD-eGFP overexpressing strains.

5.2. Substrate preferences of each isoform: a reflection on steric restriction or enzyme subcellular localisation?

Our methodology to obtain CRISPR-Cas9 mutants allowed us to select mostly mutants with a pure mutation profile, with the exception of *mgdγi1* that stayed with a mosaic mutation profile without further purification (**Fig. 4.12**). For phenotypic analyses of mutated MGDs, a minimum of two knockout mutants each obtained with a different single guide RNA were used. Two overexpressing strains were used for each MGD isoform. Western blot analysis of MGD-eGFP expression in the overexpressors confirmed the overexpression of each MGDs (**Fig. 4.13**), and revealed distinct protein levels in the two overexpressors of MGDβ and MGDγ. Expression levels of MGDα-eGFP were relatively low compared to those of MGDβ-eGFP and MGDγ-eGFP. This might be linked to the fact that *MGDα* gene was cloned into the overexpressing vector with extra 213 bp from its 5'-UTR, which might contain regulatory regions moderating the overexpression level.

Surprisingly, none of the MGD isoforms are mandatory for cell survival as knockout mutants could be successfully obtained for each MGD. Besides, no clear impact of knockout mutants on cell growth, cell morphology, membrane integrity, nor photosynthetic activity could be detected (**Fig. 4.14**, **4.15**, and **4.16**).

Under optimal culture condition, weak differences could be observed in the glycerolipidome of all mutant strains. The level of these changes indicates that in the mutants, the remaining unmutated MGDs were able to compensate for the absence of enzyme activity of the mutated isoform. Given that a compensation mechanism at the transcriptional level could only be detected in MGDγ mutants by qPCR, other compensation types might take place when MGDα and MGDβ are knocked out. Control of MGD protein stability, enzymatic regulation (i.e. activation of remaining isoforms *via* metabolic regulatory loops activated by mutation), MGDs relocalisation or MGDG transport through the membranes could maybe re-establish the required MGDG level at the different subcellular compartments.

In nitrogen deprivation condition, MGDα knockout mutants behaved similarly to the WT. This was expected as MGDα protein level in the WT was shown to be strongly downregulated, and nearly abolished in this condition.

Consistently, the effects of MGDβ and MGDγ mutations on the glycerolipidome were much stronger in low nitrogen where mainly only one MGD isoform could compensate for the absence of the other.

Combined observations of MGDα, MGDβ and MGDγ mutation effects in optimal and nutrient deplete conditions allowed us to identify the main DAG substrates for each isoform and their products fate (Fig. 5.1).

Under optimal condition, MGDα mutations most strongly affected MGDG species containing a C16 FA with two or more unsaturations, except those containing a C20:5 in *sn-1* position. Consistently, these species are highly reduced in nitrogen deprivation, which might either reflect the specific degradation of these species in nitrogen deprivation, and/or the absence of MGDα protein for their production. In the total FA distribution in MGDα mutants, a small increase in 16:1 was observed, while 16:2, 16:3, and 16:4 FAs tended to decrease. Given that C16 desaturations are performed by desaturases acting mostly on MGDG (Dolch and Maréchal, 2015), and that the 16:1 FA can be produced prior to its integration on a glycerol backbone (Dolch and Maréchal, 2015), we propose that MGDα could mainly use DAG with the following compositions: 14:0-16:0, 14:0-16:1, 16:0-16:0, 16:1-16:0 and 16:1-16:1 (which correspond to the main DAG species detected by LC-MS), that would stem from the *de novo* production of FA (**Fig. 5.1**). Once the corresponding MGDG products are formed, they would barely accumulate as plastid-localised desaturases would quickly use them to produce MGDG species with higher unsaturation levels (Dolch and Maréchal, 2015), and provide the cell with 16:2, 16:3, and 16:4 FAs (**Table 5.1**).

Dolch *et al* described as 'chloroplast membranes' the localisation of the desaturases working on C16 FA (Dolch and Maréchal, 2015). This localisation was attributed because a bipartite presequence was detected in these enzymes. Thus, it means that the preproteins for these enzymes can cross the two innermost membranes of the plastid, and could be localised in the oEM, iEM, and/or the thylakoids. A quick desaturation of MGDα products would suggest that desaturases localisation include the thylakoids membranes.

The interpretation of MGDα substrates identity and products fate would be consistent with MGDα being localised close to *de novo* FA synthesis as well as to the plastid desaturases.

As for the impact of MGDα mutation on DGDG, two explanations are possible. Indeed, we observed a decrease in DGDG species containing a 20:5 in position *sn-1*, and an increase in DGDG species containing a 16:1 in position *sn-1*. This does not reflect the impact seen on MGDG species. Therefore, either:

- (i) MGDα also operates massively on DAG 20:5-16:0 and 20:5-16:1 but the subsequent products do not accumulate as they are quickly used for DGDG synthesis, or
- (ii) in the absence of MGDα, the DAG species that were not used by MGDα are available for the other MGDs, that will produce the MGDG 16:0-16:0, 16:0-16:1 and 16:1-16:1. These MGDG species would then need to be transported to the thylakoids. However, prior to their arrival to the thylakoids, these species might be available for DGD enzymes that would be present in the plastid membranes where MGDβ and MGDγ are localised, and therefore these DGD enzymes would receive a higher proportion of MGDG with two C16 FA (**Fig. 5.1**).

In one case, MGDα would be important for the synthesis of 20:5-containing DGDG species through its production of 20:5-containing MGDG species, and a possible association with a DGD enzyme specific for the production of

DGDG species with a 20:5. In the other case, MGDα would not directly impact DGDG composition as it would not produce 20:5-containing MGDG species, and its other MGDG products would not be used either for DGDG synthesis because they quickly acquire high unsaturation levels that are not used by DGDs. Rather, the impact on the DGDG species would be due to a higher production of DGDG with two C16 FAs. This hypothesis relies on the assumption that important DGD enzymes would be localised in some of the four plastid surrounding membranes. MGDG species with two C16 FAs would be synthesised in the membranes where these DGDs are localised, increasing their availability compared to 20:5-containing MGDG species, that would also be present in these membranes.

Of the two MGDα overexpressing lines, only one showed a strong phenotype on galactolipid species. Given that in both overexpressors, similar MGDα-eGFP expression levels were detected, we cannot really argue in favour of one phenotype. Nevertheless, MGDα-eGFP-B line showed more MGDG species containing a C16 with two or more unsaturations, except those containing a 20:5. This is coherent with the suggested role of MGDα.

Under optimal conditions, MGDβ mutation strongly decreased the accumulation of 20:5-containing MGDG species, except MGDG 20:5-16:4. Interestingly, MGDG 14:0-16:1, 16:1-16:0 and 16:1-16:1 species were also decreased. This firstly suggests that under normal conditions, MGDβ would mostly use DAG 20:5-16:0 and 20:5-16:1, and that the corresponding MGDG products would be quickly desaturated, leading to the accumulation of 20:5-containing MGDG species with a higher unsaturation level of the C16 at position *sn-2* (**Table 5.1 and Fig. 5.1**).

MGDβ is localised in the blob-like structure and potentially the PPM, and its direct products do not seem to accumulate a lot in the cell compared to the more unsaturated form of these species. Therefore, MGDβ 20:5-containing products must be quickly available to the desaturases. This means that either desaturases are present in the oEM and iEM and that the 20:5-containing MGDG species are quickly transported to these two membranes, or that the transport of these species is fast enough to reach thylakoid-localised desaturases without accumulating (**Fig. 5.1**). In either case, the 20:5-containing species are unambiguously transported to the thylakoid membranes where most of the MGDG is usually found, as these represent more than 50 % of total MGDG. The role of MGDβ in the production of 20:5-containing MGDG species is also consistent with its subcellular localisation. Indeed, the 20:5 FA is produced outside of the plastid following elongation and desaturation steps (**Fig. 5.1**). For this FA to 'come back' to the plastid, it must cross the four different plastid envelope membranes, including the PPM where MGDβ is localised. Finally, given the very low level level of DAG containing a 20:5 FA in the cell, the fact that this DAG is the principal substrate of MGDβ highly suggest a form of concentrating mechanism of this lipid in the blob.

On the other hand, the decrease of MGDG 14:0-16:1, 16:1-16:0 and 16:1-16:1 species suggests that these three species might also be direct products of MGDβ, but that most of these MGDβ products would not be made available for the desaturases. Indeed, when we look at the MGDG species distribution profile in a WT, we see that MGDG 14:0-16:1, 16:1-16:0 and 16:1-16:1 species are able to accumulate. This would not be coherent if all of these species were supposed to be rapidly converted by desaturases into other MGDG species as soon as they are produced. Therefore, these species might accumulate in membranes where desaturases are not present, *e.g.*

potentially in the PPM and EpM. The noticeable decrease of MGDG 14:0-16:1n 16:1-16:1 and 16:1-16:1 species in MGDβ knock-out mutants therefore means that MGDβ participates to the accumulation of these species. Consistently, MGDG 16:1-16:0 and 16:1-16:1 proportions are increased in the MGDβ overexpressing lines, while the MGDG 16:2-16:3 and 16:3-16:3 decrease.

Under normal conditions, only DGDG 16:1-16:1 and 16:1-16:2 proportions significantly decreased in MGDβ knockouts, while DGDG 14:0-16:1 tended to decrease. This would be consistent with the role of MGDβ in the production of MGDG 14:0-16:1 and MGDG 16:1-16:1, in membranes where they are made available to DGDs. The decrease in DGDG 16:1-16:2 although MGDβ is not impacting MGDG 16:1-16:2 can also be explained. Indeed, the desaturation of 16:1 to 16:2 in position *sn-2* of DGDG can be performed by FAD6 (Dolch and Maréchal, 2015). Therefore, the decrease in DGDG 16:1-16:2 may originate from the decrease in DGDG 16:1-16:1 from which it could be produced. These results highlight the possible association of MGDβ with a DGD enzyme specific of the production of DGDG 16:1-16:1. The stable proportion of 20:5-containing DGDG species might be due to compensation effects by MGDα and MGDγ. In line with this hypothesis, 20:5-containing MGDG are still produced, and could be used in priority to maintain 20:5-containing DGDG levels, which would explain why MGDG 20:5-16:4 was not decreased as this species is barely used for DGDG synthesis (species not always detected by LC-MS). Furthermore, it was suggested above that DGDG with a 20:5 could be produced by a DGD associated with MGDα.

Under nitrogen deprivation, MGDβ mutation led to a decrease in 20:5-containing MGDG species, except, again, for the MGDG 20:5-16:4. This is consistent with the role of MGDβ in the production of 20:5-containing MGDG species, and the priority given to the production of 20:5-containing DGDG species. We also observed a decrease of MGDG 14:0-16:2, 14:0-16:3, 16:1-16:2, 16:1-16:3, and 16:2-16:3. These species need to be produced by desaturation of MGDs products. This might mean that in nitrogen deprivation, a large proportion of the MGDG 14:0-16:1, 16:1-16:0 and 16:1-16:1 produced by MGDβ are transported toward the plastid desaturases. It is possible then that part of MGDG 14:0-16:1, 16:1-16:0 and 16:1-16:1 species produced by MGDβ under optimal conditions are also transported to the desaturases, but that this effect is negligible on the total amount of highly unsaturated MGDG with a C16 in position *sn-1*. However, under nitrogen deprivation, the very low level of MGDα may increase the availability of DAG species 14:0-16:1, 16:1-16:0 and 16:1-16:1, and because MGDα will not lead to the production of highly unsaturated MGDG with a C16 at position *sn-1*, MGDβ contribution to their production is much increased.

Under nitrogen deprivation, the impact of MGDβ on DGDG species is again a decrease in 14:0-16:1, 16:1-16:1, and 16:1-16:2, similar to what happens under normal conditions. Which also means that MGDγ could largely compensate for the production of DGDG species.

Under optimal conditions, MGDγ mutation led to a strong decrease in all 14:0-containing MGDG species and in MGDG 16:0-16:0, 16:1-16:0, 16:1-16:1 and 20:5-16:4. Therefore, the main DAG substrates for MGDγ under normal conditions might be 14:0-16:0, 14:0-16:1, 16:0-16:0, 16:0-16:1 and 16:1-16:1. Given that MGDγ is located on the outermost membrane of the plastid, its MGDG products are not directly available for plastid desaturases (**Table 5.1**). Coherently, MGDγ mutation did not have an impact on MGDG species containing a C16 with two or more unsaturations, except for 14:0-16:2 and 14:0-16:3 species. The strong impact on MGDG 14:0-16:0, 14:0-

16:1, 16:0-16:0, 16:1-16:0, 16:1-16:1 species suggest that MGDγ is responsible for their production and accumulation (**Table 5.1 and Fig. 5.1**). Under nitrogen deprivation condition, MGDγ also led to a very strong decrease of the same molecular species, which further confirm its role in the production and accumulation of these species. In nitrogen starvation, MGDG 14:0-16:2 and 14:0-16:3 species were not impacted by MGDγ mutation, suggesting that their decrease under optimal conditions might not be directly linked to MGDγ mutation. Coherently, MGDG 14:0-16:1, 16:0-16:0, 16:1-16:0 and 16:1-16:1 species were increased in MGDγ overexpressing lines, while MGDG 16:1-16:3, 16:2-16:3 and 16:3-16:3 were decreased. The products of MGDγ are therefore able to accumulate in membranes where plastid desaturases are not found. Such membranes can be the EpM and PPM, but could also comprise extra-plastidial membranes.

MGDγ main products contain a very low unsaturation level (from 0 to 2). In these MGDG species, the volume occupied by the acyl chain is rather small, which gives these molecules a shape that better fits a cylinder than a cone. Therefore, these species naturally tend to form a lamellar (Lm) phase. In other words, the MGDG species produced by MGDγ may have distinct physical properties that the MGDG species produced by the activities of MGDα and MGDβ.

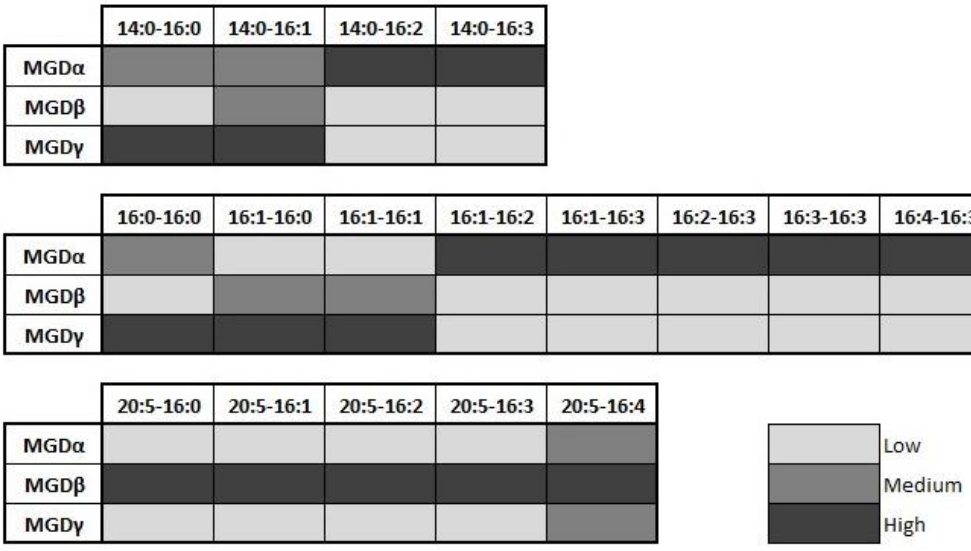


Table 5.1: Contribution of each MGD to the accumulation of the different MGDG species. The level of contribution of each MGD to the accumulation of each MGDG species is indicated with three different intensities of grey. This is not to be confounded with the level of use of each possible DAG substrate.

Consistently with its impact on MGDG species, MGDγ led to a decrease of DGDG 14:0-16:1, 16:0-16:0, 16:1-16:0, and 16:1-16:1, and to a proportional increase in 20:5-containing DGDG species. These decreases were moderate in optimal conditions, and very strong in nitrogen deprivation condition. Therefore, we can speculate that a particular DGD enzyme might be highly dependent on the products of MGDγ for the production of DGDG species with a 14:0 or a C16 in position *sn-1*. This DGD would be localized in or close to the EpM. Nevertheless, it must be noted that in nitrogen deprivation condition, in the absence of MGDβ following its mutation and most probably of MGDα too due to its low protein level, MGDγ was able to produce 20:5-containing MGDG species, although these species did not reach the same proportions as in the WT, and lead to the production of 20:5-containing DGDG species. Therefore, MGDγ can also use DAG 20:5-16:0 and 20:5-16:1 for the production of

MGDG species (**Fig. 5.1D**). Nevertheless, it seems that MGDγ support to 20:5-containing MGDG species is very low when other MGDs are active.

Overall, it seems that all MGDs in *P. tricornutum* are able to use the same DAG substrates. However, they seem to show substrate apparent preferences. These preferences might result from actual enzyme specificities determined by steric features around the binding an active sites, or reflect the DAG substrate availability in the distinct subcellular compartments where MGDs are localised (**Fig. 5.1**). Indeed, MGDα localisation to the thylakoid membranes makes it closer to DAG substrate newly produced following *de novo* FA synthesis. However, it is found far from the site of production of 20:5 FA. Consistently, MGDα seems to mainly use 14:0-16:0, 14:0-16:1, 16:0-16:0, 16:0-16:1 and 16:1-16:1 DAG. MGDβ is localised in the vesicular network found as a blob-like structure between the PPM and the oEM, and maybe extend to all the PPM. This vesicular network has been discussed as a possible way to transport plastid-targeted proteins through the PPM (**Fig. 2.2B**). Here it seems that DAG 20:5-16:0 and 20:5-16:1 species show a high availability where MGDβ localises, as they appear to be its main substrates. Therefore, the vesicular network may be involved in the transport of 20:5-containing lipid species inside the plastid, as part of the so-called Ω-pathway. Interestingly MGDγ localisation in the ER/EpM and cytosol suggests that it may dock to the cytosolic surface of the EpM and to other non-plastidial membranes, where 20:5-containing DAG could be found. However, it seems that MGDγ is more exposed to the same DAG as MGDα, and barely uses 20:5-containing DAG in a system where every MGD is expressed (in the WT under optimal growth condition). This is particularly intriguing, and suggests that the 20:5 FA may use a particular route that avoids the conversion of a large part of DAG 20:5-16:0 and 20:5-16:1 in the EpM by MGDγ.

We observed that the production of 20:5-containing MGDG is possible under nitrogen condition in the absence of both MGDα and MGDβ, which means that MGDγ is still able to produce these species. Therefore, production of 20:5-containing MGDG species might also take place in the EpM where these products would have to be able to relocate to the next plastid membranes rapidly to reach the plastidial desaturases (**Fig. 5.1C**). Alternatively, this production might occur in the PPM if relocalisation of MGDγ takes place in the absence of MGDβ.

Interestingly, the proportions of most of the C18-containing lipid species in non-plastidial glycerolipids were decreased in MGDα mutants. However, PC and DGTA proportions tended to increase. Therefore, the C18-containing species might be relatively low compared to the main PC and DGTA species which accumulated in slightly higher amounts than in the WT. Interestingly, the DGTA 16:0-16:1 and 16:1-16:1 species were increased in MGDα mutants, which might indicate that DAG composed of a combination of 16:0 and 16:1 FAs were slightly more available for synthesis of non-plastidial lipids, which would explain their very slight increase in proportion. DAG proportion however was decreased in the glycerolipid profile, and the DAG 16:1-16:1 species was found to be increased while all other DAG species decreased.

In MGDβ mutants, DAG 20:5-16:0 proportion was not particularly increased. However, there seemed to be an increase in 20:5-20:5 species in PC, DGTA and PE. In *P. tricornutum*, the formation of 20:5 is suspected to occur on both phospholipids and betaine lipids (Dolch and Maréchal, 2015). The increase of 20:5-20:5 in PC and DGTA could therefore reflect the accumulation of 20:5 formed on these species due to its lack of integration in MGDG.

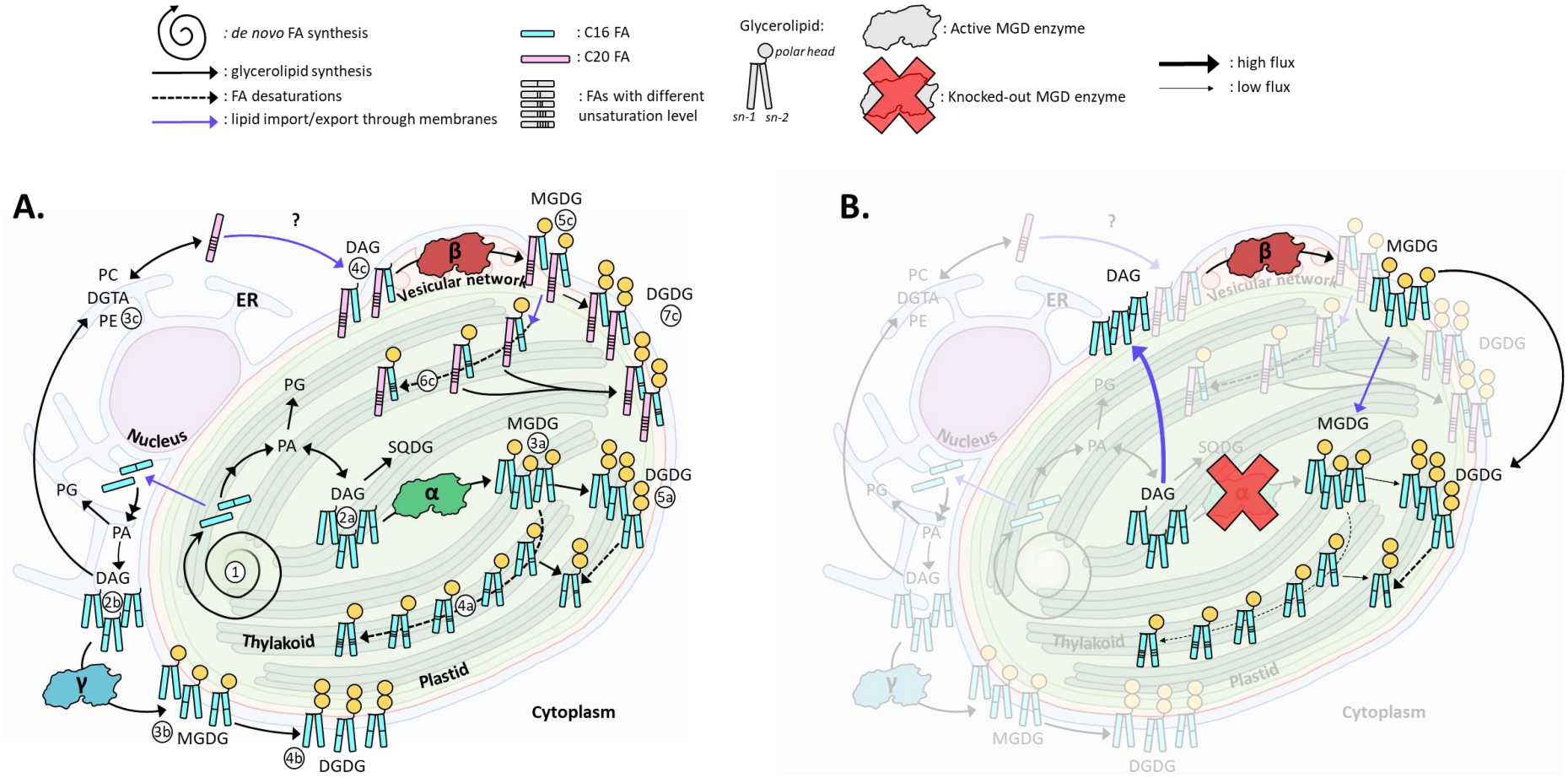
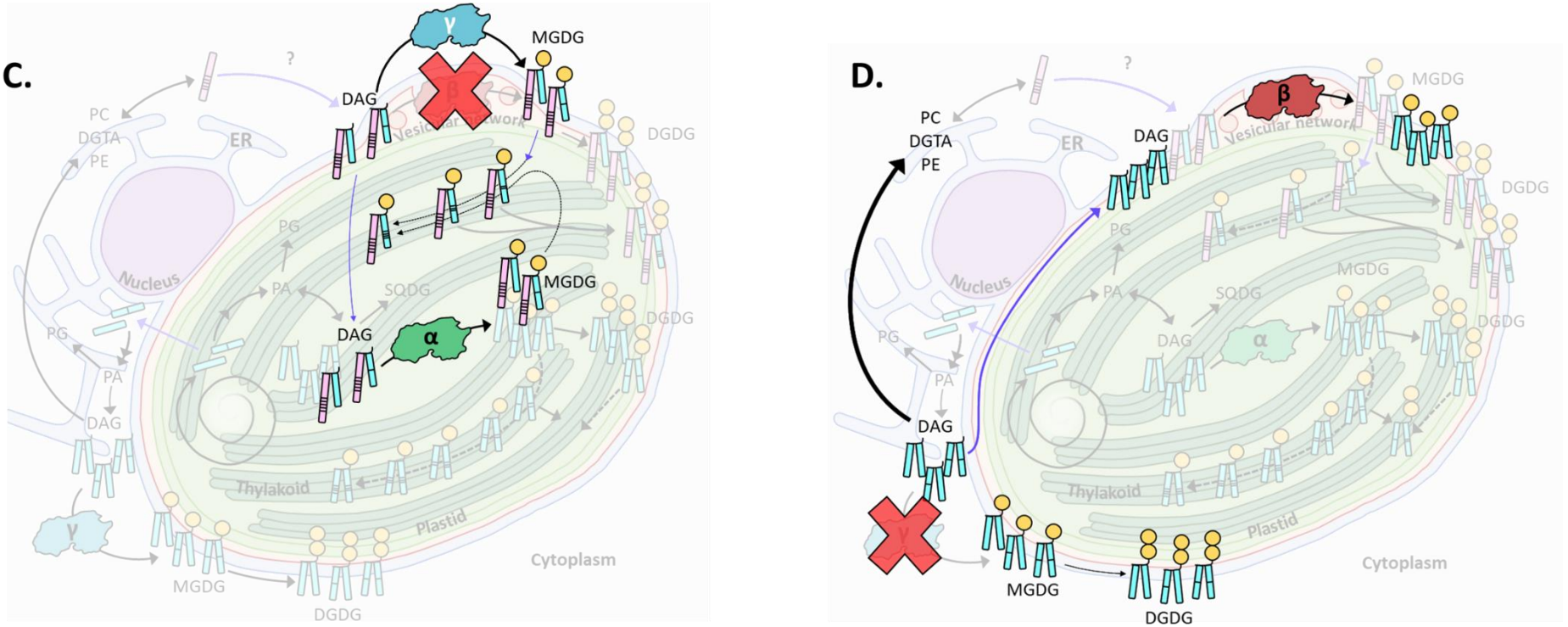


Figure 5.1: Interpretation of lipid fluxes in *P. tricornutum* WT and MGD mutant cells. Proposed pathways for lipid synthesis and transport in a *P. tricornutum* cell, reconstructed based on the analysis of MGD overexpressing and mutated lines. DAG, MGDG and DGDG compositions are detailed, showing the main species consumed/produced by each MGD in the different subcellular compartments. For simplification, species with a 14:0 FA are not represented. **A**, Schematisation of lipid fluxes in a WT. FA are *de novo* synthesised in the plastid (1). Then, FA are either directly assembled on a glycerol backbone for the production of plastidial glycerolipids, or transported through the plastid envelope membranes to the cytosol. In the plastid, DAG species 16:0-16:0, 16:0-16:1 and 16:1-16:1 (2a) are used by MGDα to produce MGDG species (3a) that will rapidly be desaturated (4a) and/or used for DGDG synthesis (5a). FA exported to the cytosol are assembled on a glycerol backbone for the production of extraplastidial glycerolipids. DAG species 16:0-16:0, 16:0-16:1 and 16:1-16:1 (2b) are used by MGDγ to produce MGDG species (3b) at the EpM



and/or extraplastidial membranes. These species can accumulate and/or be used for DGDG synthesis (4b). FA elongation in the cytosol and desaturation on both phospholipids and betaine lipids (3c) leads to the formation of 20:5 FA. 20:5 FA is transported 'back' to the plastid through an unknown mechanism. In the PPM, DAG 20:5-16:0 and 20:16:1 (4c) is used by MGDβ. MGDβ direct products (5c) are transported to the plastid where they are rapidly desaturated (6c) and/or used for DGDG synthesis (7c). **B**, In the absence of MGDα, the unused DAG substrates are possibly transported through the plastid envelope membranes and become mostly available to MGDβ. The subsequent products are likely to be partly transported back to the thylakoids where they are rapidly desaturated, and partly used for the synthesis of DGDG. **C**, In the absence of MGDβ, the unused DAG substrates are possibly transported through the plastid envelope membranes and become available to both MGDα and MGDγ. The flux of DGDG production from the 20:5-containing MGDG species is unchanged. **D**, In the absence of MGDγ, the unused DAG substrates are likely to be partly used for the synthesis of extraplastidial lipids, and partly transported through the envelope membranes where they become mostly available to MGDβ. DAG, diacylglycerol; DGDG, digalactosyldiacylglycerol; DGTA, diacylglycerylhydroxymethyltrimethyl-β-alanine; MGDG, monogalactosyldiacylglycerol; PE, phosphatidylethanolamine, PC, phosphatidylcholine.

MGDγ mutation had a strong impact on non-plastidial lipid composition. In particular, there was an increase of PC and DGTA species containing a combination of 16:1 and 16:0 FAs on their glycerol backbone, both under optimal and nitrogen deprived conditions. MGDγ being located on the outside of the plastid, there is a very high probability that this enzyme locates near to the sites of PC and DGTA synthesis. Therefore, it is consistent that the DAG species that were no used by MGDγ for the production of MGDG may be used as substrates for the synthesis of PC and DGTA, and potentially PE (although we could not detect these species in PE) (**Fig. 5.1D**). It is reasonable to imagine that the 16:0 FA amounts available for elongation into 18:0 and for following desaturation in phospholipids and DGTA increased. Consistently, the amounts of PC, PE and DGTA containing an 18:1 or an 18:2 FA increased. It seems however that the desaturases and the elongase responsible for the desaturations of 18:2 to 18:3 and 18:3 to 18:4, as well as the elongation of 18:4 to 20:4 were not fast enough to respond to the increase of their substrate. This might be why an imbalance appeared with a decrease in the proportions of PC, PE and DGTA species containing in *sn-2* position an 18:4, 20:4, 20:5 or 22:6 FA. This does not mean however that less 20:5 species were produced, and there was actually a slight increase in the proportion of 20:5-containing MGDG species, and a strong increase in SQDG 20:5-16:0.

Under nitrogen deprivation, MGDβ seemed to accumulate TAG faster than in the WT at day 3 of deprivation. However, its TAG species distribution measured at day 5 was only slightly impacted by the mutation. Therefore, it is unclear why MGDβ mutants would lead to such phenotype. In MGDγ, the total TAG amount in the cell was lower than in the WT. Interestingly, MGDγ mutation led to strong differences in the proportion of several TAG species. In particular, there was a decrease in 16:0-16:0-16:0 and 16:0-16:0-16:1 species, and an increase in 16:1-16:1-16:1. There were also small increases and decreases of TAG species consistent with corresponding decreases and increases observed with the MGDγ overexpressing lines. It shows that MGDG species produced by MGDγ maybe be used for the synthesis of TAG species, as already suspected in *C. reinhardtii* (Li et al., 2012c; Gu et al., 2021; Iwai et al., 2021).

In conclusion, we have shown that each MGD leads to the production of different proportions of MGDG species, and that their use of DAG and the fate of produced MGDG species depend on their subcellular localisation (**Fig. 5.1**). We also observed that under optimal condition, the absence of an MGD can be largely compensated by the activity of the two other MGD isoforms, which might suggest that either MGD enzymes are able to relocalise, or more probably that MGDG species are able to travel fast across different membranes, through a mechanism that still need to be elucidated. Under nitrogen deprivation, MGDα protein level is decreased, which is consistent with a decrease in thylakoids membranes induced in this condition. Under this stressful condition, we see that the roles of MGDβ and MGDγ are important for lipid homeostasis, especially that of MGDγ. Therefore, as in *A. thaliana*, it seems that the emergence of additional MGDs in diatoms is linked with the improvement of adaptation to stressful conditions.

6. Perspectives

This work allowed a first characterisation of each MGD in *P. tricornutum* under optimal and nitrogen deprivation conditions. Many questions have emerged following our discoveries. First, it seems clear that none of the MGD in *P. tricornutum* were crucial for cell survival under optimal and nitrogen-deprived conditions, especially because of compensation effects that seemed to be implemented in the MGD single mutants. Therefore, phenotyping of double MGD mutants would permit to refine our interpretation of each MGD substrates and products, and possibly highlight the actual importance of MGDβ and MGDγ under nitrogen deprivation. Other stress conditions should also be tested, such as phosphate deprivation, high light stress, and salt stress.

Although presence of a conserved active site and the impact of MGD mutants on MGDG composition demonstrate that each MGD isoform must have an MGDG synthase activity, the actual detection of MGDG synthase activity would add more weight to this study. Given that detection of glycosyltransferase activity failed using a UDP-Glo glycosyltransferase assay, other tests could be performed. For example, the expression of recombinant *P. tricornutum* MGD isoforms in yeast followed by the detection of the presence or absence of MGDG in transformed yeast (as yeast do not naturally produce MGDG).

The production of recombinant MGD proteins should be tried again. Addition of micelles with a composition similar to plastid membranes to the cell-free system instead of detergent might facilitate MGDα production. Furthermore, MGDγ and MGDβ may need to be produced again, as their N-terminal extension of around 80 amino acids was removed from the produced recombinant proteins, in spite of its potential importance for activity. Alternatively, heterologous expression in bacteria or yeast should be evaluated, with appropriate protein tags for straightforward purification.

Using Langmuir monolayers, the binding properties of MGDγ could be investigated using an ER-like lipid composition, and test wether PE and MGDG are interchangeable.

The produced MGDγ antibody could be used for immunolocalisation assay. This would not only complement our observation of MGDγ-eGFP in the ER of *P. tricornutum*, but also reveal if this isoform can also localise to the EpM and other extra-plastidial membranes.

Eventually, a role for MGD outside of the plastid should definitely be investigated in MGDγ-like containing diatoms and in other related phyla.

7. Annexes

Supplemental Table 1: *P. tricornutum* MGDs information retrieved from the whole cell proteome analysis performed by Josselin Lupette in 10N10P and 0N10P condition. Analyses were performed as described in (Lupette et al., 2019). The log.iBAQ values are the log2 of the Maxquant intensities in 10N10P (p) and 0N10P (m). The fold change (FC) indicate the change in protein level in 0N10P compared to 10N10P.

Protein ID	Gene Symbol	Description	Uniprot Accession	Chr Contig (phatr3)	JGI corresponding ID (phatr2)	MW (kDa)	Seq. length
Phatr3_J14125.p1	MGDα	Monogalactosyldiacylglycerol synthase α	B7G3X5	13	jgi|Phatr2|14125|e_gw1.13.68.1	58,993	540
Phatr3_J54168.p1	MGDβ	Monogalactosyldiacylglycerol synthase β	B5Y4U3	3	jgi|Phatr2|54168|estExt_Phatr1_ua_kg.C_chr_30020	61,806	559
Phatr3_J9619.p1	MGDγ	Monogalactosyldiacylglycerol synthase γ	B7FQ04	1	jgi|Phatr2|9619|e_gw1.1.264.1	66,48	618

Protein ID	Number of peptides	Sequence coverage (%)	Score (Maxquant)	Log.iBAQ.m1	Log.iBAQ.m2	Log.iBAQ.m3	Log.iBAQ.p1	Log.iBAQ.p2	Log.iBAQ.p3	P_Value	logFC	FC
Phatr3_J14125.p1	5	11,1	49,668	13,9670	16,4245	14,2270	19,2089	19,3636	18,1659	1,10E-03	-4,0399	0,061
Phatr3_J54168.p1	7	15,9	18,493	20,0538	20,1923	19,4084	19,4053	19,5128	19,7017	3,83E-01	0,3449	1,270
Phatr3_J9619.p1	7	14,1	11,333	17,1974	16,4054	17,3026	19,0765	18,6593	18,9487	2,01E-03	-1,9263	0,263

8. References

Abida, H., Dolch, L.-J., Meï, C., Villanova, V., Conte, M., Block, M. A., et al. (2015). Membrane Glycerolipid Remodeling Triggered by Nitrogen and Phosphorus Starvation in *Phaeodactylum tricornutum*. *Plant Physiology* 167, 118–136. doi: 10.1104/pp.114.252395.

Abu-Elheiga, L., Brinkley, W. R., Zhong, L., Chirala, S. S., Woldegiorgis, G., and Wakil, S. J. (2000). The subcellular localization of acetyl-CoA carboxylase 2. *Proc Natl Acad Sci U S A* 97, 1444–1449. doi: 10.1073/pnas.97.4.1444.

Adler-Agnon, Z., Leu, S., Zarka, A., Boussiba, S., and Khozin-Goldberg, I. (2018). Novel promoters for constitutive and inducible expression of transgenes in the diatom Phaeodactylum tricornutum under varied nitrate availability. *J Appl Phycol* 30, 2763–2772. doi: 10.1007/s10811-017-1335-8.

Alboresi, A., Perin, G., Vitulo, N., Diretto, G., Block, M. A., Jouhet, J., et al. (2016). Light Remodels Lipid Biosynthesis in Nannochloropsis gaditana by Modulating Carbon Partitioning Between Organelles. *Plant Physiol.*, pp.00599.2016. doi: 10.1104/pp.16.00599.

Allen, E. J., and Nelson, E. W. (1910). On the Artificial Culture of Marine Plankton Organisms. *J. Mar. Biol. Ass.* 8, 421–474. doi: 10.1017/S0025315400073690.

Anderson, J. M., and Andersson, B. (1982). The architecture of photosynthetic membranes: lateral and transverse organization. *Trends in Biochemical Sciences* 7, 288–292. doi: 10.1016/0968-0004(82)90014-7.

Anderson, R., Livermore, B. P., Kates, M., and Volcani, B. E. (1978). The lipid composition of the non-photosynthetic diatom nitzschia alba. *Biochimica et Biophysica Acta (BBA) - Lipids and Lipid Metabolism* 528, 77–88. doi: 10.1016/0005-2760(78)90054-1.

Andersson, B., and Anderson, J. M. (1980). Lateral heterogeneity in the distribution of chlorophyll-protein complexes of the thylakoid membranes of spinach chloroplasts. *Biochimica et Biophysica Acta (BBA) - Bioenergetics* 593, 427–440. doi: 10.1016/0005-2728(80)90078-X.

Andersson, M. X., and Dörmann, P. (2008). “Chloroplast Membrane Lipid Biosynthesis and Transport,” in *The Chloroplast: Interactions With the Environment* Plant Cell Monographs., eds. A. S. Sandelius and H. Aronsson (Berlin, Heidelberg: Springer Berlin Heidelberg), 125–159. doi: 10.1007/978-3-540-68696-5_4.

Andersson, M. X., Goksör, M., and Sandelius, A. S. (2007a). Membrane Contact Sites: Physical Attachment Between Chloroplasts and Endoplasmic Reticulum Revealed by Optical Manipulation. *Plant Signaling & Behavior* 2, 185–187. doi: 10.4161/psb.2.3.3973.

Andersson, M. X., Goksör, M., and Sandelius, A. S. (2007b). Optical Manipulation Reveals Strong Attracting Forces at Membrane Contact Sites between Endoplasmic Reticulum and Chloroplasts. *Journal of Biological Chemistry* 282, 1170–1174. doi: 10.1074/jbc.M608124200.

Andrews, J., and Mudd, J. B. (1985). Phosphatidylglycerol Synthesis in Pea Chloroplasts: Pathway and Localization. *Plant Physiol.* 79, 259–265. doi: 10.1104/pp.79.1.259.

Anto, S., Mukherjee, S. S., Muthappa, R., Mathimani, T., Deviram, G., Kumar, S. S., et al. (2020). Algae as green energy reserve: Technological outlook on biofuel production. *Chemosphere* 242, 125079. doi: 10.1016/j.chemosphere.2019.125079.

Apdila, E. T., and Awai, K. (2017). Configuration of the sugar head of glycolipids in thylakoid membranes. *Genes Genet. Syst.* 92, 235–242. doi: 10.1266/ggs.17-00047.

Apt, K. E., Grossman, A. R., and Kroth-Pancic, P. G. (1996). Stable nuclear transformation of the diatomPhaeodactylum tricornutum. *Molec. Gen. Genet.* 252, 572–579. doi: 10.1007/BF02172403.

Arao, T., and Yamada, M. (1994). Biosynthesis of polyunsaturated fatty acids in the marine diatom, Phaeodactylum tricornutum. *Phytochemistry* 35, 1177–1181. doi: 10.1016/S0031-9422(00)94817-9.

Armbrust, E. V. (2004). The Genome of the Diatom Thalassiosira Pseudonana: Ecology, Evolution, and Metabolism. *Science* 306, 79–86. doi: 10.1126/science.1101156.

Attard, G. S., Templer, R. H., Smith, W. S., Hunt, A. N., and Jackowski, S. (2000). Modulation of CTP:phosphocholine cytidylyltransferase by membrane curvature elastic stress. *Proceedings of the National Academy of Sciences* 97, 9032–9036. doi: 10.1073/pnas.160260697.

Awai, K., Kakimoto, T., Awai, C., Kaneko, T., Nakamura, Y., Takamiya, K., et al. (2006). Comparative Genomic Analysis Revealed a Gene for Monoglucosyldiacylglycerol Synthase, an Enzyme for Photosynthetic Membrane Lipid Synthesis in Cyanobacteria. *Plant Physiology* 141, 1120–1127. doi: 10.1104/pp.106.082859.

Awai, K., Marechal, E., Block, M. A., Brun, D., Masuda, T., Shimada, H., et al. (2001). Two types of MGDG synthase genes, found widely in both 16:3 and 18:3 plants, differentially mediate galactolipid syntheses in photosynthetic and nonphotosynthetic tissues in Arabidopsis thaliana. *Proceedings of the National Academy of Sciences* 98, 10960–10965. doi: 10.1073/pnas.181331498.

Awai, K., Ohta, H., and Sato, N. (2014). Oxygenic photosynthesis without galactolipids. *PNAS* 111, 13571–13575. doi: 10.1073/pnas.1403708111.

Awai, K., Watanabe, H., Benning, C., and Nishida, I. (2007). Digalactosyldiacylglycerol is Required for Better Photosynthetic Growth of Synechocystis sp. PCC6803 Under Phosphate Limitation. *Plant and Cell Physiology* 48, 1517–1523. doi: 10.1093/pcp/pcm134.

Aziz, E., Batool, R., Khan, M. U., Rauf, A., Akhtar, W., Heydari, M., et al. (2020). An overview on red algae bioactive compounds and their pharmaceutical applications. *Journal of Complementary and Integrative Medicine* 0. doi: 10.1515/jcim-2019-0203.

Bailleul, B., Berne, N., Murik, O., Petroutsos, D., Prihoda, J., Tanaka, A., et al. (2015). Energetic coupling between plastids and mitochondria drives CO2 assimilation in diatoms. *Nature* 524, 366–369. doi: 10.1038/nature14599.

Ball, S. G., Subtil, A., Bhattacharya, D., Moustafa, A., Weber, A. P. M., Gehre, L., et al. (2013). Metabolic Effectors Secreted by Bacterial Pathogens: Essential Facilitators of Plastid Endosymbiosis? *The Plant Cell* 25, 7–21. doi: 10.1105/tpc.112.101329.

Bandarra, N. M., Gokpinar, S., Gama Pereira, T., Nunes, M. L., Gouveia, L., Monterio, M., et al. (2008). Effect of Temperature on Growth and Biochemical Composition (Sterols, α-tocopherol, Carotenoids, Fatty Acid Profiles) of the Microalga, Isochrysis galbana. *Israeli Journal of Aquaculture - Bamidgeh*. doi: 10.46989/001c.20492.

Barka, F., Angstenberger, M., Ahrendt, T., Lorenzen, W., Bode, H. B., and Büchel, C. (2016). Identification of a triacylglycerol lipase in the diatom Phaeodactylum tricornutum. *Biochimica et Biophysica Acta (BBA) - Molecular and Cell Biology of Lipids* 1861, 239–248. doi: 10.1016/j.bbalip.2015.12.023.

Barkia, I., Saari, N., and Manning, S. R. (2019). Microalgae for High-Value Products Towards Human Health and Nutrition. *Marine Drugs* 17, 304. doi: 10.3390/md17050304.

Bastien, O., Botella, C., Chevalier, F., Block, M. A., Jouhet, J., Breton, C., et al. (2016). "New Insights on Thylakoid Biogenesis in Plant Cells," in *International Review of Cell and Molecular Biology* (Elsevier), 1–30. doi: 10.1016/bs.ircmb.2015.12.001.

Bates, P. D., Durrett, T. P., Ohlrogge, J. B., and Pollard, M. (2009). Analysis of Acyl Fluxes through Multiple Pathways of Triacylglycerol Synthesis in Developing Soybean Embryos. *Plant Physiology* 150, 55–72. doi: 10.1104/pp.109.137737.

Bates, P. D., Fatihi, A., Snapp, A. R., Carlsson, A. S., Browse, J., and Lu, C. (2012). Acyl Editing and Headgroup Exchange Are the Major Mechanisms That Direct Polyunsaturated Fatty Acid Flux into Triacylglycerols. *Plant Physiology* 160, 1530–1539. doi: 10.1104/pp.112.204438.

Bates, P. D., Ohlrogge, J. B., and Pollard, M. (2007). Incorporation of Newly Synthesized Fatty Acids into Cytosolic Glycerolipids in Pea Leaves Occurs via Acyl Editing. *Journal of Biological Chemistry* 282, 31206–31216. doi: 10.1074/jbc.M705447200.

Bedoshvili, Ye. D., Popkova, T. P., and Likhoshway, Ye. V. (2009). Chloroplast structure of diatoms of different classes. *Cell Tiss. Biol.* 3, 297–310. doi: 10.1134/S1990519X09030122.

Bell, R. M., Ballas, L. M., and Coleman, R. A. (1981). Lipid topogenesis. *J Lipid Res* 22, 391–403.

Benning, C. (2008). A role for lipid trafficking in chloroplast biogenesis. *Progress in Lipid Research* 47, 381–389. doi: 10.1016/j.plipres.2008.04.001.

Benning, C. (2009). Mechanisms of Lipid Transport Involved in Organelle Biogenesis in Plant Cells. *Annu. Rev. Cell Dev. Biol.* 25, 71–91. doi: 10.1146/annurev.cellbio.042308.113414.

Berry, E. A., Guergova-Kuras, M., Huang, L., and Crofts, A. R. (2000). Structure and Function of Cytochrome *bc* Complexes. *Annu. Rev. Biochem.* 69, 1005–1075. doi: 10.1146/annurev.biochem.69.1.1005.

Bertrand, M. (2010). Carotenoid biosynthesis in diatoms. *Photosynth Res* 106, 89–102. doi: 10.1007/s11120-010-9589-x.

Bhaya, D., and Grossman, A. R. (1993). Characterization of gene clusters encoding the fucoxanthin chlorophyll proteins of the diatom Phaeodactylum tricornutum. *Nucleic Acids Res* 21, 4458–4466. doi: 10.1093/nar/21.19.4458.

Billey, E., Hafidh, S., Cruz-Gallardo, I., Litholdo, C. G., Jean, V., Carpentier, M.-C., et al. (2020). LARP6C regulates selective mRNA translation to promote pollen tube guidance in *Arabidopsis thaliana*. doi: 10.1101/2020.11.27.401307.

Billey, E., Hafidh, S., Cruz-Gallardo, I., Litholdo, C. G., Jean, V., Carpentier, M.-C., et al. (2021). LARP6C orchestrates posttranscriptional reprogramming of gene expression during hydration to promote pollen tube guidance. *The Plant Cell* 33, 2637–2661. doi: 10.1093/plcell/koab131.

Block, M. A., Dorne, A. J., Joyard, J., and Douce, R. (1983). Preparation and characterization of membrane fractions enriched in outer and inner envelope membranes from spinach chloroplasts. II. Biochemical characterization. *J Biol Chem* 258, 13281–13286.

Block, M. A., and Jouhet, J. (2015). Lipid trafficking at endoplasmic reticulum–chloroplast membrane contact sites. *Current Opinion in Cell Biology* 35, 21–29. doi: 10.1016/j.ceb.2015.03.004.

Bojko, M., Olchawa-Pajor, M., Chyc, M., Goss, R., Schaller-Laudel, S., and Latowski, D. (2017). Acclimatization of Thalassiosira pseudonana Photosynthetic Membranes to Environmental Temperature Changes. in doi: 10.11159/icepr17.120.

Bolam, D. N., Roberts, S., Proctor, M. R., Turkenburg, J. P., Dodson, E. J., Martinez-Fleites, C., et al. (2007). The crystal structure of two macrolide glycosyltransferases provides a blueprint for host cell antibiotic immunity. *Proceedings of the National Academy of Sciences* 104, 5336–5341. doi: 10.1073/pnas.0607897104.

Botella, C., Sautron, E., Boudiere, L., Michaud, M., Dubots, E., Yamaryo-Botté, Y., et al. (2016). ALA10, a Phospholipid Flippase, Controls FAD2/FAD3 Desaturation of Phosphatidylcholine in the ER and Affects Chloroplast Lipid Composition in *Arabidopsis thaliana*. *Plant Physiol.* 170, 1300–1314. doi: 10.1104/pp.15.01557.

Botté, C., Jeanneau, C., Snajdrova, L., Bastien, O., Imberty, A., Breton, C., et al. (2005). Molecular Modeling and Site-directed Mutagenesis of Plant Chloroplast Monogalactosyldiacylglycerol Synthase Reveal Critical Residues for Activity. *Journal of Biological Chemistry* 280, 34691–34701. doi: 10.1074/jbc.M505622200.

Botté, C. Y., Deligny, M., Roccia, A., Bonneau, A.-L., Saïdani, N., Hardré, H., et al. (2011). Chemical inhibitors of monogalactosyldiacylglycerol synthases in Arabidopsis thaliana. *Nat Chem Biol* 7, 834–842. doi: 10.1038/nchembio.658.

Botté, C. Y., and Maréchal, E. (2014). Plastids with or without galactoglycerolipids. *Trends in Plant Science* 19, 71–78. doi: 10.1016/j.tplants.2013.10.004.

Botte, C. Y., Yamaryo-Botte, Y., Rupasinghe, T. W. T., Mullin, K. A., MacRae, J. I., Spurck, T. P., et al. (2013). Atypical lipid composition in the purified relict plastid (apicoplast) of malaria parasites. *Proceedings of the National Academy of Sciences* 110, 7506–7511. doi: 10.1073/pnas.1301251110.

Boudière, L., Botté, C. Y., Saidani, N., Lajoie, M., Marion, J., Bréhélin, L., et al. (2012). Galvestine-1, a novel chemical probe for the study of the glycerolipid homeostasis system in plant cells. *Molecular BioSystems* 8, 2023. doi: 10.1039/c2mb25067e.

Boudière, L., Michaud, M., Petroutsos, D., Rébeillé, F., Falconet, D., Bastien, O., et al. (2014). Glycerolipids in photosynthesis: Composition, synthesis and trafficking. *Biochimica et Biophysica Acta (BBA) - Bioenergetics* 1837, 470–480. doi: 10.1016/j.bbabio.2013.09.007.

Bowler, C., Allen, A. E., Badger, J. H., Grimwood, J., Jabbari, K., Kuo, A., et al. (2008). The Phaeodactylum genome reveals the evolutionary history of diatom genomes. *Nature* 456, 239–244. doi: 10.1038/nature07410.

Bowler, C., and Falciatore, A. (2019). Phaeodactylum tricornutum. *Trends in Genetics* 35, 706–707. doi: 10.1016/j.tig.2019.05.007.

Bowler, C., Vardi, A., and Allen, A. E. (2010). Oceanographic and Biogeochemical Insights from Diatom Genomes. *Annu. Rev. Mar. Sci.* 2, 333–365. doi: 10.1146/annurev-marine-120308-081051.

Bradford, M. M. (1976). A rapid and sensitive method for the quantitation of microgram quantities of protein utilizing the principle of protein-dye binding. *Analytical Biochemistry* 72, 248–254. doi: 10.1016/0003-2697(76)90527-3.

Branco-Vieira, M., San Martin, S., Agurto, C., Freitas, M. A. V., Martins, A. A., Mata, T. M., et al. (2020). Biotechnological potential of Phaeodactylum tricornutum for biorefinery processes. *Fuel* 268, 117357. doi: 10.1016/j.fuel.2020.117357.

Brembu, T., Jørstad, M., Winge, P., Valle, K. C., and Bones, A. M. (2011). Genome-Wide Profiling of Responses to Cadmium in the Diatom *Phaeodactylum tricornutum*. *Environ. Sci. Technol.* 45, 7640–7647. doi: 10.1021/es2002259.

Breton, C., Fournel-Gigleux, S., and Palcic, M. M. (2012). Recent structures, evolution and mechanisms of glycosyltransferases. *Current Opinion in Structural Biology* 22, 540–549. doi: 10.1016/j.sbi.2012.06.007.

Breton, S., Jouhet, J., Guyet, U., Gros, V., Pittera, J., Demory, D., et al. (2020). Unveiling membrane thermoregulation strategies in marine picocyanobacteria. *New Phytol* 225, 2396–2410. doi: 10.1111/nph.16239.

Burki, F., Roger, A. J., Brown, M. W., and Simpson, A. G. B. (2020). The New Tree of Eukaryotes. *Trends in Ecology & Evolution* 35, 43–55. doi: 10.1016/j.tree.2019.08.008.

Busseni, G., Rocha Jimenez Vieira, F., Amato, A., Pelletier, E., Pierella Karlusich, J. J., Ferrante, M. I., et al. (2019). Meta-Omics Reveals Genetic Flexibility of Diatom Nitrogen Transporters in Response to Environmental Changes. *Molecular Biology and Evolution* 36, 2522–2535. doi: 10.1093/molbev/msz157.

Butler, T., Kapoore, R. V., and Vaidyanathan, S. (2020). Phaeodactylum tricornutum: A Diatom Cell Factory. *Trends in Biotechnology* 38, 606–622. doi: 10.1016/j.tibtech.2019.12.023.

Caballero, M. A., Jallet, D., Shi, L., Rithner, C., Zhang, Y., and Peers, G. (2016). Quantification of chrysolaminarin from the model diatom Phaeodactylum tricornutum. *Algal Research* 20, 180–188. doi: 10.1016/j.algal.2016.10.008.

Campbell, J. A., Davies, G. J., Bulone, V., and Henrissat, B. (1997). A classification of nucleotide-diphospho-sugar glycosyltransferases based on amino acid sequence similarities. *Biochem J* 326 (Pt 3), 929–939. doi: 10.1042/bj3260929u.

Cañavate, J. P., Armada, I., and Hachero-Cruzado, I. (2017a). Aspects of phosphorus physiology associated with phosphate-induced polar lipid remodelling in marine microalgae. *Journal of Plant Physiology* 214, 28–38. doi: 10.1016/j.jplph.2017.03.019.

Cañavate, J. P., Armada, I., and Hachero-Cruzado, I. (2017b). Interspecific variability in phosphorus-induced lipid remodelling among marine eukaryotic phytoplankton. *New Phytol* 213, 700–713. doi: 10.1111/nph.14179.

Carter, H. E., McCluer, R. H., and Slifer, E. D. (1956). Lipids of Wheat Flour. I. Characterization of Galactosylglycerol Components [1]. *J. Am. Chem. Soc.* 78, 3735–3738. doi: 10.1021/ja01596a051.

Cavalier-Smith, T. (2018). Kingdom Chromista and its eight phyla: a new synthesis emphasising periplastid protein targeting, cytoskeletal and periplastid evolution, and ancient divergences. *Protoplasma* 255, 297–357. doi: 10.1007/s00709-017-1147-3.

Cenci, U., Bhattacharya, D., Weber, A. P. M., Colleoni, C., Subtil, A., and Ball, S. G. (2017). Biotic Host–Pathogen Interactions As Major Drivers of Plastid Endosymbiosis. *Trends in Plant Science* 22, 316–328. doi: 10.1016/j.tplants.2016.12.007.

Chang, A., Singh, S., Phillips, G. N., and Thorson, J. S. (2011). Glycosyltransferase structural biology and its role in the design of catalysts for glycosylation. *Current Opinion in Biotechnology* 22, 800–808. doi: 10.1016/j.copbio.2011.04.013.

Charuvi, D., Kiss, V., Nevo, R., Shimoni, E., Adam, Z., and Reich, Z. (2012). Gain and Loss of Photosynthetic Membranes during Plastid Differentiation in the Shoot Apex of *Arabidopsis*. *Plant Cell* 24, 1143–1157. doi: 10.1105/tpc.111.094458.

Chaudhary, R., Nawaz, K., Khan, A. K., Hano, C., Abbasi, B. H., and Anjum, S. (2020). An Overview of the Algae-Mediated Biosynthesis of Nanoparticles and Their Biomedical Applications. *Biomolecules* 10, 1498. doi: 10.3390/biom10111498.

Chauton, M. S., Winge, P., Brembu, T., Vadstein, O., and Bones, A. M. (2013). Gene Regulation of Carbon Fixation, Storage, and Utilization in the Diatom Phaeodactylum tricornutum Acclimated to Light/Dark Cycles. *PLANT PHYSIOLOGY* 161, 1034–1048. doi: 10.1104/pp.112.206177.

Chen, J. E., and Smith, A. G. (2012). A look at diacylglycerol acyltransferases (DGATs) in algae. *Journal of Biotechnology* 162, 28–39. doi: 10.1016/j.jbiotec.2012.05.009.

Chen, M., and Thelen, J. J. (2013). *ACYL-LIPID DESATURASE2* Is Required for Chilling and Freezing Tolerance in *Arabidopsis*. *The Plant Cell* 25, 1430–1444. doi: 10.1105/tpc.113.111179.

Chitnis, P. R. (2001). PHOTOSYSTEM I: Function and Physiology. *Annu. Rev. Plant. Physiol. Plant. Mol. Biol.* 52, 593–626. doi: 10.1146/annurev.arplant.52.1.593.

Conte, M., Lupette, J., Seddiki, K., Meï, C., Dolch, L.-J., Gros, V., et al. (2018). Screening for Biologically Annotated Drugs That Trigger Triacylglycerol Accumulation in the Diatom *Phaeodactylum*. *Plant Physiology* 177, 532–552. doi: 10.1104/pp.17.01804.

Coutinho, P. M., Deleury, E., Davies, G. J., and Henrissat, B. (2003). An evolving hierarchical family classification for glycosyltransferases. *J Mol Biol* 328, 307–317. doi: 10.1016/s0022-2836(03)00307-3.

Coves, J., Joyard, J., and Douce, R. (1988). Lipid requirement and kinetic studies of solubilized UDP-galactose:diacylglycerol galactosyltransferase activity from spinach chloroplast envelope membranes. *Proceedings of the National Academy of Sciences* 85, 4966–4970. doi: 10.1073/pnas.85.14.4966.

Cui, Y., Zhao, J., Wang, Y., Qin, S., and Lu, Y. (2018). Characterization and engineering of a dual-function diacylglycerol acyltransferase in the oleaginous marine diatom Phaeodactylum tricornutum. *Biotechnol Biofuels* 11, 32. doi: 10.1186/s13068-018-1029-8.

Cui, Y., Zheng, G., Li, X., Lin, H., Jiang, P., and Qin, S. (2013). Cloning and characterization of a novel diacylglycerol acyltransferase from the diatom Phaeodactylum tricornutum. *J Appl Phycol* 25, 1509–1512. doi: 10.1007/s10811-013-9991-9.

Cunnane, S., Drevon, C. A., and Harris, W. S. (2004). Recommendations for intakes of polyunsaturated fatty acids in healthy adults. *ISSFAL News* 11, 12–25.

Curtis, B. A., Tanifuji, G., Burki, F., Gruber, A., Irimia, M., Maruyama, S., et al. (2012). Algal genomes reveal evolutionary mosaicism and the fate of nucleomorphs. *Nature* 492, 59–65. doi: 10.1038/nature11681.

Daboussi, F. (2017). Advances in editing microalgae genomes. *Perspectives in Phycology* 4, 17–23. doi: 10.1127/pip/2017/0071.

Daboussi, F., Leduc, S., Maréchal, A., Dubois, G., Guyot, V., Perez-Michaut, C., et al. (2014). Genome engineering empowers the diatom Phaeodactylum tricornutum for biotechnology. *Nature Communications* 5. doi: 10.1038/ncomms4831.

Dahlqvist, A., Stahl, U., Lenman, M., Banas, A., Lee, M., Sandager, L., et al. (2000). Phospholipid:diacylglycerol acyltransferase: An enzyme that catalyzes the acyl-CoA-independent formation of triacylglycerol in yeast and plants. *Proceedings of the National Academy of Sciences* 97, 6487–6492. doi: 10.1073/pnas.120067297.

de Vargas, C., Audic, S., Henry, N., Decelle, J., Mahe, F., Logares, R., et al. (2015). Eukaryotic plankton diversity in the sunlit ocean. *Science* 348, 1261605–1261605. doi: 10.1126/science.1261605.

Demé, B., Cataye, C., Block, M. A., Maréchal, E., and Jouhet, J. (2014). Contribution of galactoglycerolipids to the 3-dimensional architecture of thylakoids. *FASEB j.* 28, 3373–3383. doi: 10.1096/fj.13-247395.

Di Dato, V., Musacchia, F., Petrosino, G., Patil, S., Montresor, M., Sanges, R., et al. (2015). Transcriptome sequencing of three Pseudo-nitzschia species reveals comparable gene sets and the presence of Nitric Oxide Synthase genes in diatoms. *Sci Rep* 5, 12329. doi: 10.1038/srep12329.

Dobrikova, A. G., Domonkos, I., Sözer, Ö., Laczkó-Dobos, H., Kis, M., Párducz, Á., et al. (2013). Effect of partial or complete elimination of light-harvesting complexes on the surface electric properties and the functions of cyanobacterial photosynthetic membranes. *Physiol Plantarum* 147, 248–260. doi: 10.1111/j.1399-3054.2012.01648.x.

Dolch, L.-J., Lupette, J., Tourcier, G., Bedhomme, M., Collin, S., Magneschi, L., et al. (2017a). Nitric Oxide Mediates Nitrite-Sensing and Acclimation and Triggers a Remodeling of Lipids. *Plant Physiology* 175, 1407–1423. doi: 10.1104/pp.17.01042.

Dolch, L.-J., and Maréchal, E. (2015). Inventory of Fatty Acid Desaturases in the Pennate Diatom Phaeodactylum tricornutum. *Marine Drugs* 13, 1317–1339. doi: 10.3390/md13031317.

Dolch, L.-J., Rak, C., Perin, G., Tourcier, G., Broughton, R., Leterrier, M., et al. (2017b). A Palmitic Acid Elongase Affects Eicosapentaenoic Acid and Plastidial Monogalactosyldiacylglycerol Levels in Nannochloropsis. *Plant Physiol.* 173, 742–759. doi: 10.1104/pp.16.01420.

Domergue, F., Lerchl, J., Zähringer, U., and Heinz, E. (2002). Cloning and functional characterization of *Phaeodactylum tricornutum* front-end desaturases involved in eicosapentaenoic acid biosynthesis: Δ5- and Δ6-fatty acid desaturases from diatom. *European Journal of Biochemistry* 269, 4105–4113. doi: 10.1046/j.1432-1033.2002.03104.x.

Domergue, F., Spiekermann, P., Lerchl, J., Beckmann, C., Kilian, O., Kroth, P. G., et al. (2003). New Insight into *Phaeodactylum tricornutum* Fatty Acid Metabolism. Cloning and Functional Characterization of Plastidial and Microsomal Δ12-Fatty Acid Desaturases. *Plant Physiology* 131, 1648–1660. doi: 10.1104/pp.102.018317.

Dörmann, P., Balbo, I., and Benning, C. (1999). *Arabidopsis* Galactolipid Biosynthesis and Lipid Trafficking Mediated by DGD1. *Science* 284, 2181–2184. doi: 10.1126/science.284.5423.2181.

Dorne, A. J., Joyard, J., Block, M. A., and Douce, R. (1985). Localization of phosphatidylcholine in outer envelope membrane of spinach chloroplasts. *J Cell Biol* 100, 1690–1697. doi: 10.1083/jcb.100.5.1690.

Douce, R., Alban, C., Bligny, R., Block, M. A., Covès, J., Dorne, A.-J., et al. (1987). "Lipid Distribution and Synthesis Within the Plant Cell," in *The Metabolism, Structure, and Function of Plant Lipids*, eds. P. K. Stumpf, J. B. Mudd, and W. D. Nes (Boston, MA: Springer New York), 255–263. doi: 10.1007/978-1-4684-5263-1_46.

Dubots, E., Audry, M., Yamaryo, Y., Bastien, O., Ohta, H., Breton, C., et al. (2010). Activation of the Chloroplast Monogalactosyldiacylglycerol Synthase MGD1 by Phosphatidic Acid and Phosphatidylglycerol. *Journal of Biological Chemistry* 285, 6003–6011. doi: 10.1074/jbc.M109.071928.

Eastmond, P. J. (2006). *SUGAR-DEPENDENT1* Encodes a Patatin Domain Triacylglycerol Lipase That Initiates Storage Oil Breakdown in Germinating *Arabidopsis* Seeds. *Plant Cell* 18, 665–675. doi: 10.1105/tpc.105.040543.

Emanuelsson, O., Nielsen, H., and Heijne, G. V. (1999). ChloroP, a neural network-based method for predicting chloroplast transit peptides and their cleavage sites. *Protein Sci.* 8, 978–984. doi: 10.1110/ps.8.5.978.

Erdene-Ochir, E., Shin, B.-K., Huda, M. N., Kim, D. H., Lee, E. H., Song, D.-G., et al. (2016). Cloning of a novel endogenous promoter for foreign gene expression in Phaeodactylum tricornutum. *Appl Biol Chem* 59, 861–867. doi: 10.1007/s13765-016-0235-y.

Erdene-Ochir, E., Shin, B.-K., Kwon, B., Jung, C., and Pan, C.-H. (2019). Identification and characterisation of the novel endogenous promoter HASP1 and its signal peptide from Phaeodactylum tricornutum. *Sci Rep* 9, 9941. doi: 10.1038/s41598-019-45786-9.

Falarz, L. J., Xu, Y., Caldo, K. M. P., Garroway, C. J., Singer, S. D., and Chen, G. (2020). Characterization of the diversification of phospholipid:diacylglycerol acyltransferases in the green lineage. *Plant J* 103, 2025–2038. doi: 10.1111/tpj.14880.

Falciatore, A., Jaubert, M., Bouly, J.-P., Bailleul, B., and Mock, T. (2020). Diatom Molecular Research Comes of Age: Model Species for Studying Phytoplankton Biology and Diversity. *Plant Cell* 32, 547–572. doi: 10.1105/tpc.19.00158.

Falkowski, P. G., Barber, R. T., and Smetacek, V. (1998). Biogeochemical Controls and Feedbacks on Ocean Primary Production. *Science* 281, 200–206. doi: 10.1126/science.281.5374.200.

Fan, W., and Evans, R. M. (2015). Turning Up the Heat on Membrane Fluidity. *Cell* 161, 962–963. doi: 10.1016/j.cell.2015.04.046.

Fan, X., Qiu, H., Han, W., Wang, Y., Xu, D., Zhang, X., et al. (2020). Phytoplankton pangenome reveals extensive prokaryotic horizontal gene transfer of diverse functions. *Sci. Adv.* 6, eaba0111. doi: 10.1126/sciadv.aba0111.

Field, C. B. (1998). Primary Production of the Biosphere: Integrating Terrestrial and Oceanic Components. *Science* 281, 237–240. doi: 10.1126/science.281.5374.237.

Filloramo, G. V., Curtis, B. A., Blanche, E., and Archibald, J. M. (2021). Re-examination of two diatom reference genomes using long-read sequencing. *BMC Genomics* 22, 379. doi: 10.1186/s12864-021-07666-3.

Flori, S., Jouneau, P.-H., Bailleul, B., Gallet, B., Estrozi, L. F., Moriscot, C., et al. (2017). Plastid thylakoid architecture optimizes photosynthesis in diatoms. *Nature Communications* 8, 15885. doi: 10.1038/ncomms15885.

Flori, S., Jouneau, P.-H., Finazzi, G., Maréchal, E., and Falconet, D. (2016). Ultrastructure of the Periplastidial Compartment of the Diatom Phaeodactylum tricornutum. *Protist* 167, 254–267. doi: 10.1016/j.protis.2016.04.001.

Flori, S., Jouneau, P.-H., Gallet, B., Estrozi, L. F., Moriscot, C., Schoehn, G., et al. (2018). “Imaging Plastids in 2D and 3D: Confocal and Electron Microscopy,” in *Plastids* Methods in Molecular Biology., ed. E. Maréchal (New York, NY: Springer US), 113–122. doi: 10.1007/978-1-4939-8654-5_7.

Folch, J., Lees, M., and Stanley, G. H. S. (1957). A simple method for the isolation and purification of total lipides from animal tissues. *Journal of Biological Chemistry* 226, 497–509. doi: 10.1016/S0021-9258(18)64849-5.

Franz, A. K., Danielewicz, M. A., Wong, D. M., Anderson, L. A., and Boothe, J. R. (2013). Phenotypic screening with oleaginous microalgae reveals modulators of lipid productivity. *ACS Chem Biol* 8, 1053–1062. doi: 10.1021/cb300573r.

Froehlich, J. E., Benning, C., and Dörmann, P. (2001). The Digalactosyldiacylglycerol (DGDG) Synthase DGD1 Is Inserted into the Outer Envelope Membrane of Chloroplasts in a Manner Independent of the General Import Pathway and Does Not Depend on Direct Interaction with Monogalactosyldiacylglycerol Synthase for DGDG Biosynthesis. *Journal of Biological Chemistry* 276, 31806–31812. doi: 10.1074/jbc.M104652200.

Fujii, S., Kobayashi, K., Nagata, N., Masuda, T., and Wada, H. (2017). Monogalactosyldiacylglycerol Facilitates Synthesis of Photoactive Protochlorophyllide in Etioplasts. *Plant Physiol* 174, 2183–2198. doi: 10.1104/pp.17.00304.

Fujii, S., Kobayashi, K., Nakamura, Y., and Wada, H. (2014). Inducible Knockdown of *MONOGALACTOSYLDIACYLGLYCEROL SYNTHASE1* Reveals Roles of Galactolipids in Organelle

Differentiation in Arabidopsis Cotyledons. *Plant Physiology* 166, 1436–1449. doi: 10.1104/pp.114.250050.

Fujii, S., Nagata, N., Masuda, T., Wada, H., and Kobayashi, K. (2019). Galactolipids Are Essential for Internal Membrane Transformation during Etioplast-to-Chloroplast Differentiation. *Plant and Cell Physiology* 60, 1224–1238. doi: 10.1093/pcp/pcz041.

Gao, B., Chen, A., Zhang, W., Li, A., and Zhang, C. (2017). Co-production of lipids, eicosapentaenoic acid, fucoxanthin, and chrysolaminarin by Phaeodactylum tricornutum cultured in a flat-plate photobioreactor under varying nitrogen conditions. *J. Ocean Univ. China* 16, 916–924. doi: 10.1007/s11802-017-3174-2.

Garab, G., Lohner, K., Laggner, P., and Farkas, T. (2000). Self-regulation of the lipid content of membranes by non-bilayer lipids: a hypothesis. *Trends in Plant Science* 5, 489–494. doi: 10.1016/S1360-1385(00)01767-2.

Gascuel, O. (1997). BIONJ: an improved version of the NJ algorithm based on a simple model of sequence data. *Molecular Biology and Evolution* 14, 685–695. doi: 10.1093/oxfordjournals.molbev.a025808.

Gawrisch, K., Parsegian, V. A., Hajduk, D. A., Tate, M. W., Gruner, S. M., Fuller, N. L., et al. (1992). Energetics of a hexagonal-lamellar-hexagonal-phase transition sequence in dioleoylphosphatidylethanolamine membranes. *Biochemistry* 31, 2856–2864. doi: 10.1021/bi00126a003.

Gomès, E., Jakobsen, M. K., Axelsen, K. B., Geisler, M., and Palmgren, M. G. (2000). Chilling tolerance in Arabidopsis involves ALA1, a member of a new family of putative aminophospholipid translocases. *Plant Cell* 12, 2441–2454.

Gong, Y., Guo, X., Wan, X., Liang, Z., and Jiang, M. (2011). Characterization of a novel thioesterase (PtTE) from Phaeodactylum tricornutum. *J. Basic Microbiol.* 51, 666–672. doi: 10.1002/jobm.201000520.

Gong, Y., Zhang, J., Guo, X., Wan, X., Liang, Z., Hu, C. J., et al. (2013). Identification and characterization of PtDGAT2B, an acyltransferase of the DGAT2 acyl-Coenzyme A: Diacylglycerol acyltransferase family in the diatom *Phaeodactylum tricornutum*. *FEBS Letters* 587, 481–487. doi: 10.1016/j.febslet.2013.01.015.

Goss, R., Ann Pinto, E., Wilhelm, C., and Richter, M. (2006). The importance of a highly active and ΔpH-regulated diatoxanthin epoxidase for the regulation of the PS II antenna function in diadinoxanthin cycle containing algae. *Journal of Plant Physiology* 163, 1008–1021. doi: 10.1016/j.jplph.2005.09.008.

Goss, R., and Jakob, T. (2010). Regulation and function of xanthophyll cycle-dependent photoprotection in algae. *Photosynth Res* 106, 103–122. doi: 10.1007/s11120-010-9536-x.

Goss, R., Lohr, M., Latowski, D., Grzyb, J., Vieler, A., Wilhelm, C., et al. (2005). Role of Hexagonal Structure-Forming Lipids in Diadinoxanthin and Violaxanthin Solubilization and De-Epoxidation. *Biochemistry* 44, 4028–4036. doi: 10.1021/bi047464k.

Gould, S. B., Maier, U.-G., and Martin, W. F. (2015). Protein Import and the Origin of Red Complex Plastids. *Current Biology* 25, R515–R521. doi: 10.1016/j.cub.2015.04.033.

Gounaris, K., and Barber, J. (1983). Monogalactosyldiacylglycerol: The most abundant polar lipid in nature. *Trends in Biochemical Sciences* 8, 378–381. doi: 10.1016/0968-0004(83)90366-3.

Gounaris, K., Sundby, C., Andersson, B., and Barber, J. (1983). Lateral heterogeneity of polar lipids in the thylakoid membranes of spinach chloroplasts. *FEBS Letters* 156, 170–174. doi: 10.1016/0014-5793(83)80271-3.

Griffiths, G., Stymne, S., and Stobart, A. K. (1988). The utilisation of fatty-acid substrates in triacylglycerol biosynthesis by tissue-slices of developing safflower (Carthamus tinctorius L.) and sunflower (Helianthus annuus L.) cotyledons. *Planta* 173, 309–316. doi: 10.1007/BF00401017.

Grosche, C., Hempel, F., Bolte, K., Zauner, S., and Maier, U. G. (2014). The periplastidal compartment: a naturally minimized eukaryotic cytoplasm. *Current Opinion in Microbiology* 22, 88–93. doi: 10.1016/j.mib.2014.09.017.

Groth, G., and Strotmann, H. (1999). New results about structure, function and regulation of the chloroplast ATP synthase ($CF_0 CF_1$). *Physiologia Plantarum* 106, 142–148. doi: 10.1034/j.1399-3054.1999.106120.x.

Gruber, A., Rocap, G., Kroth, P. G., Armbrust, E. V., and Mock, T. (2015). Plastid proteome prediction for diatoms and other algae with secondary plastids of the red lineage. *Plant J* 81, 519–528. doi: 10.1111/tpj.12734.

Gruber, A., Vugrinec, S., Hempel, F., Gould, S. B., Maier, U.-G., and Kroth, P. G. (2007). Protein targeting into complex diatom plastids: functional characterisation of a specific targeting motif. *Plant Molecular Biology* 64, 519–530. doi: 10.1007/s11103-007-9171-x.

Gschloessl, B., Guermeur, Y., and Cock, J. M. (2008). HECTAR: A method to predict subcellular targeting in heterokonts. *BMC Bioinformatics* 9, 393. doi: 10.1186/1471-2105-9-393.

Gu, X., Cao, L., Wu, X., Li, Y., Hu, Q., and Han, D. (2021). A Lipid Bodies-Associated Galactosyl Hydrolase Is Involved in Triacylglycerol Biosynthesis and Galactolipid Turnover in the Unicellular Green Alga Chlamydomonas reinhardtii. *Plants* 10, 675. doi: 10.3390/plants10040675.

Guéguen, N., Le Moigne, D., Amato, A., Salvaing, J., and Maréchal, E. (2021). Lipid Droplets in Unicellular Photosynthetic Stramenopiles. *Front. Plant Sci.* 12, 639276. doi: 10.3389/fpls.2021.639276.

Guéguen, N., and Maréchal, E. (2021). Origin of cyanobacterial thylakoids via a non-vesicular glycolipid phase transition and their impact on the Great Oxygenation Event. *Journal of Experimental Botany*, erab429. doi: 10.1093/jxb/erab429.

Guihéneuf, F., Leu, S., Zarka, A., Khozin-Goldberg, I., Khalilov, I., and Boussiba, S. (2011). Cloning and molecular characterization of a novel acyl-CoA:diacylglycerol acyltransferase 1-like gene (PtDGAT1) from the diatom Phaeodactylum tricornutum: P. tricornutum DGAT1 cloning and characterization. *FEBS Journal* 278, 3651–3666. doi: 10.1111/j.1742-4658.2011.08284.x.

Guiry, M. D. (2012). How many species of algae are there? *Journal of Phycology* 48, 1057–1063. doi: 10.1111/j.1529-8817.2012.01222.x.

Guo, X., Jiang, M., Wan, X., Hu, C., and Gong, Y. (2014). Identification and biochemical characterization of five long-chain acyl-coenzyme A synthetases from the diatom Phaeodactylum tricornutum. *Plant Physiology and Biochemistry* 74, 33–41. doi: 10.1016/j.plaphy.2013.10.036.

Gupta, T. K., Klumpe, S., Gries, K., Heinz, S., Wietrzynski, W., Ohnishi, N., et al. (2021). Structural basis for VIPP1 oligomerization and maintenance of thylakoid membrane integrity. *Cell* 184, 3643-3659.e23. doi: 10.1016/j.cell.2021.05.011.

Haigh, W. G., Yoder, T. F., Ericson, L., Pratum, T., and Winget, R. R. (1996). The characterisation and cyclic production of a highly unsaturated homoserine lipid in Chlorella minutissima. *Biochimica et Biophysica Acta (BBA) - Lipids and Lipid Metabolism* 1299, 183–190. doi: 10.1016/0005-2760(95)00205-7.

Hamilton, M. L., Warwick, J., Terry, A., Allen, M. J., Napier, J. A., and Sayanova, O. (2015). Towards the Industrial Production of Omega-3 Long Chain Polyunsaturated Fatty Acids from a Genetically Modified Diatom Phaeodactylum tricornutum. *PLoS ONE* 10, e0144054. doi: 10.1371/journal.pone.0144054.

Hamm, C. E., Merkel, R., Springer, O., Jurkojc, P., Maier, C., Prechtel, K., et al. (2003). Architecture and material properties of diatom shells provide effective mechanical protection. *Nature* 421, 841–843. doi: 10.1038/nature01416.

Hankamer, B., Barber, J., and Boekema, E. J. (1997). Structure and membrane organization of Photosystem II in green plants. *Annu. Rev. Plant. Physiol. Plant. Mol. Biol.* 48, 641–671. doi: 10.1146/annurev.arplant.48.1.641.

Hao, X., Chen, W., Amato, A., Jouhet, J., Maréchal, E., Moog, D., et al. (2021). Multiplex CRISPR/Cas9 editing of the long-chain acyl-CoA synthetase family in the diatom *Phaeodactylum tricornutum* reveals that mitochondrial ptACSL3 is involved in the synthesis of storage lipids. *New Phytologist*, nph.17911. doi: 10.1111/nph.17911.

Haq, S., Bachvaroff, T., and Place, A. (2017). Characterization of Acetyl-CoA Carboxylases in the Basal Dinoflagellate Amphidinium carterae. *Marine Drugs* 15, 149. doi: 10.3390/md15060149.

Härtel, H., and Benning, C. (2000). Can digalactosyldiacylglycerol substitute for phosphatidylcholine upon phosphate deprivation in leaves and roots of Arabidopsis? *Biochem Soc Trans* 28, 729–732.

Hartel, H., Dormann, P., and Benning, C. (2000). DGD1-independent biosynthesis of extraplastidic galactolipids after phosphate deprivation in Arabidopsis. *Proceedings of the National Academy of Sciences* 97, 10649–10654. doi: 10.1073/pnas.180320497.

Haslam, R. P., Hamilton, M. L., Economou, C. K., Smith, R., Hassall, K. L., Napier, J. A., et al. (2020). Overexpression of an endogenous type 2 diacylglycerol acyltransferase in the marine diatom Phaeodactylum tricornutum enhances lipid production and omega-3 long-chain polyunsaturated fatty acid content. *Biotechnol Biofuels* 13, 87. doi: 10.1186/s13068-020-01726-8.

Hedayatkhah, A., Cretoiu, M. S., Emtiazi, G., Stal, L. J., and Bolhuis, H. (2018). Bioremediation of chromium contaminated water by diatoms with concomitant lipid accumulation for biofuel production. *Journal of Environmental Management* 227, 313–320. doi: 10.1016/j.jenvman.2018.09.011.

Heidrich, J., Wulf, V., Hennig, R., Saur, M., Markl, J., Sönnichsen, C., et al. (2016). Organization into Higher Ordered Ring Structures Counteracts Membrane Binding of IM30, a Protein Associated with Inner Membranes in Chloroplasts and Cyanobacteria. *Journal of Biological Chemistry* 291, 14954–14962. doi: 10.1074/jbc.M116.722686.

Hempel, F., Bullmann, L., Lau, J., Zauner, S., and Maier, U. G. (2009). ERAD-derived preprotein transport across the second outermost plastid membrane of diatoms. *Mol Biol Evol* 26, 1781–1790. doi: 10.1093/molbev/msp079.

Hempel, F., Felsner, G., and Maier, U. G. (2010). New mechanistic insights into pre-protein transport across the second outermost plastid membrane of diatoms: Ubiquitination and de-ubiquitination in a minimized cytosol. *Molecular Microbiology* 76, 793–801. doi: 10.1111/j.1365-2958.2010.07142.x.

Hennig, R., Heidrich, J., Saur, M., Schmüser, L., Roeters, S. J., Hellmann, N., et al. (2015). IM30 triggers membrane fusion in cyanobacteria and chloroplasts. *Nat Commun* 6, 7018. doi: 10.1038/ncomms8018.

Henríquez, V., Escobar, C., Galarza, J., and Gimpel, J. (2016). "Carotenoids in Microalgae," in *Carotenoids in Nature* Subcellular Biochemistry., ed. C. Stange (Cham: Springer International Publishing), 219–237. doi: 10.1007/978-3-319-39126-7_8.

Hibbeln, J. R., Nieminen, L. R., Blasbalg, T. L., Riggs, J. A., and Lands, W. E. (2006). Healthy intakes of n−3 and n−6 fatty acids: estimations considering worldwide diversity. *The American Journal of Clinical Nutrition* 83, 1483S-1493S. doi: 10.1093/ajcn/83.6.1483S.

Hook, S. E., and Osborn, H. L. (2012). Comparison of toxicity and transcriptomic profiles in a diatom exposed to oil, dispersants, dispersed oil. *Aquatic Toxicology* 124–125, 139–151. doi: 10.1016/j.aquatox.2012.08.005.

Horton, P., Park, K.-J., Obayashi, T., Fujita, N., Harada, H., Adams-Collier, C. J., et al. (2007). WoLF PSORT: protein localization predictor. *Nucleic Acids Research* 35, W585–W587. doi: 10.1093/nar/gkm259.

Hu, Q., Sommerfeld, M., Jarvis, E., Ghirardi, M., Posewitz, M., Seibert, M., et al. (2008). Microalgal triacylglycerols as feedstocks for biofuel production: perspectives and advances. *Plant J* 54, 621–639. doi: 10.1111/j.1365-313X.2008.03492.x.

Hu, Y., Chen, L., Ha, S., Gross, B., Falcone, B., Walker, D., et al. (2003). Crystal structure of the MurG:UDP-GlcNAc complex reveals common structural principles of a superfamily of glycosyltransferases. *Proceedings of the National Academy of Sciences* 100, 845–849. doi: 10.1073/pnas.0235749100.

Huerlimann, R., and Heimann, K. (2013). Comprehensive guide to acetyl-carboxylases in algae. *Critical Reviews in Biotechnology* 33, 49–65. doi: 10.3109/07388551.2012.668671.

Imbert, L., Lenoir-Capello, R., Crublet, E., Vallet, A., Awad, R., Ayala, I., et al. (2021). "In Vitro Production of Perdeuterated Proteins in H2O for Biomolecular NMR Studies," in *Structural Genomics* Methods in Molecular Biology., eds. Y. W. Chen and C.-P. B. Yiu (New York, NY: Springer US), 127–149. doi: 10.1007/978-1-0716-0892-0_8.

Iwai, M., Hori, K., Sasaki-Sekimoto, Y., Shimojima, M., and Ohta, H. (2015). Manipulation of oil synthesis in Nannochloropsis strain NIES-2145 with a phosphorus starvation–inducible promoter from Chlamydomonas reinhardtii. *Front. Microbiol.* 6. doi: 10.3389/fmicb.2015.00912.

Iwai, M., Yamada-Oshima, Y., Asami, K., Kanamori, T., Yuasa, H., Shimojima, M., et al. (2021). Recycling of the major thylakoid lipid MGDG and its role in lipid homeostasis in *Chlamydomonas reinhardtii*. *Plant Physiology* 187, 1341–1356. doi: 10.1093/plphys/kiab340.

Jacobs, T. B., LaFayette, P. R., Schmitz, R. J., and Parrott, W. A. (2015). Targeted genome modifications in soybean with CRISPR/Cas9. *BMC Biotechnol* 15, 16. doi: 10.1186/s12896-015-0131-2.

Jahns, P., Latowski, D., and Strzalka, K. (2009). Mechanism and regulation of the violaxanthin cycle: The role of antenna proteins and membrane lipids. *Biochimica et Biophysica Acta (BBA) - Bioenergetics* 1787, 3–14. doi: 10.1016/j.bbabio.2008.09.013.

Jarvis, P., Dormann, P., Peto, C. A., Lutes, J., Benning, C., and Chory, J. (2000). Galactolipid deficiency and abnormal chloroplast development in the Arabidopsis MGD synthase 1 mutant. *Proceedings of the National Academy of Sciences* 97, 8175–8179. doi: 10.1073/pnas.100132197.

Jaussaud, A., Lupette, J., Salvaing, J., Jouhet, J., Bastien, O., Gromova, M., et al. (2020). Stepwise Biogenesis of Subpopulations of Lipid Droplets in Nitrogen Starved Phaeodactylum tricornutum Cells. *Front Plant Sci* 11. doi: 10.3389/fpls.2020.00048.

Jensen, P. E., and Leister, D. (2014). Chloroplast evolution, structure and functions. *F1000Prime Rep* 6. doi: 10.12703/P6-40.

Jiang, H., and Gao, K. (2004). EFFECTS OF LOWERING TEMPERATURE DURING CULTURE ON THE PRODUCTION OF POLYUNSATURATED FATTY ACIDS IN THE MARINE DIATOM PHAEODACTYLUM TRICORNUTUM (BACILLARIOPHYCEAE)1: PUFAs in PHAEODACTYLUM TRICORNUTUM. *Journal of Phycology* 40, 651–654. doi: 10.1111/j.1529-8817.2004.03112.x.

Jiang, Y., Yoshida, T., and Quigg, A. (2012). Photosynthetic performance, lipid production and biomass composition in response to nitrogen limitation in marine microalgae. *Plant Physiology and Biochemistry* 54, 70–77. doi: 10.1016/j.plaphy.2012.02.012.

Johnson, G., and Williams, J. P. (1989). Effect of Growth Temperature on the Biosynthesis of Chloroplastic Galactosyldiacylglycerol Molecular Species in *Brassica napus* Leaves. *Plant Physiol.* 91, 924–929. doi: 10.1104/pp.91.3.924.

Johnston, M. L., Luethy, M. H., Miernyk, J. A., and Randall, D. D. (1997). Cloning and molecular analyses of the Arabidopsis thaliana plastid pyruvate dehydrogenase subunits1The nucleotide sequence data in this article have been entered into the GenBank database under the accession numbers U80185 and U80186.1. *Biochimica et Biophysica Acta (BBA) - Bioenergetics* 1321, 200–206. doi: 10.1016/S0005-2728(97)00059-5.

Jouhet, J. (2013). Importance of the hexagonal lipid phase in biological membrane organization. *Front Plant Sci* 4, 494. doi: 10.3389/fpls.2013.00494.

Jouhet, J., Lupette, J., Clerc, O., Magneschi, L., Bedhomme, M., Collin, S., et al. (2017). LC-MS/MS versus TLC plus GC methods: Consistency of glycerolipid and fatty acid profiles in microalgae and higher plant cells and effect of a nitrogen starvation. *PLOS ONE* 12, e0182423. doi: 10.1371/journal.pone.0182423.

Jouhet, J., Maréchal, E., Baldan, B., Bligny, R., Joyard, J., and Block, M. A. (2004). Phosphate deprivation induces transfer of DGDG galactolipid from chloroplast to mitochondria. *Journal of Cell Biology* 167, 863–874. doi: 10.1083/jcb.200407022.

Jouhet, J., Maréchal, E., and Block, M. A. (2007). Glycerolipid transfer for the building of membranes in plant cells. *Progress in Lipid Research* 46, 37–55. doi: 10.1016/j.plipres.2006.06.002.

Juneja, A., Ceballos, R., and Murthy, G. (2013). Effects of Environmental Factors and Nutrient Availability on the Biochemical Composition of Algae for Biofuels Production: A Review. *Energies* 6, 4607–4638. doi: 10.3390/en6094607.

Kang, J., Park, J., Choi, H., Burla, B., Kretzschmar, T., Lee, Y., et al. (2011). Plant ABC Transporters. *The Arabidopsis Book* 9, e0153. doi: 10.1199/tab.0153.

Kansy, M., Wilhelm, C., and Goss, R. (2014). Influence of thylakoid membrane lipids on the structure and function of the plant photosystem II core complex. *Planta* 240, 781–796. doi: 10.1007/s00425-014-2130-2.

Karas, B. J., Diner, R. E., Lefebvre, S. C., McQuaid, J., Phillips, A. P. R., Noddings, C. M., et al. (2015). Designer diatom episomes delivered by bacterial conjugation. *Nat Commun* 6, 6925. doi: 10.1038/ncomms7925.

Kaundal, R., Sahu, S. S., Verma, R., and Weirick, T. (2013). Identification and characterization of plastid-type proteins from sequence-attributed features using machine learning. *BMC Bioinformatics* 14, S7. doi: 10.1186/1471-2105-14-S14-S7.

Keeling, P. J. (2010). The endosymbiotic origin, diversification and fate of plastids. *Phil. Trans. R. Soc. B* 365, 729–748. doi: 10.1098/rstb.2009.0103.

Keeling, P. J. (2013). The Number, Speed, and Impact of Plastid Endosymbioses in Eukaryotic Evolution. *Annu. Rev. Plant Biol.* 64, 583–607. doi: 10.1146/annurev-arplant-050312-120144.

Keeling, P. J., Burki, F., Wilcox, H. M., Allam, B., Allen, E. E., Amaral-Zettler, L. A., et al. (2014). The Marine Microbial Eukaryote Transcriptome Sequencing Project (MMETSP): Illuminating the Functional Diversity of Eukaryotic Life in the Oceans through Transcriptome Sequencing. *PLoS Biol* 12, e1001889. doi: 10.1371/journal.pbio.1001889.

Kelley, L. A., Mezulis, S., Yates, C. M., Wass, M. N., and Sternberg, M. J. E. (2015a). The Phyre2 web portal for protein modeling, prediction and analysis. *Nat Protoc* 10, 845–858. doi: 10.1038/nprot.2015.053.

Kelley, L. A., Mezulis, S., Yates, C. M., Wass, M. N., and Sternberg, M. J. E. (2015b). The Phyre2 web portal for protein modeling, prediction and analysis. *Nat Protoc* 10, 845–858. doi: 10.1038/nprot.2015.053.

Kelly, A. A., and Dörmann, P. (2004). Green light for galactolipid trafficking. *Current Opinion in Plant Biology* 7, 262–269. doi: 10.1016/j.pbi.2004.03.009.

Kelly, A. A., Froehlich, J. E., and Dörmann, P. (2003). Disruption of the Two Digalactosyldiacylglycerol Synthase Genes *DGD1* and *DGD2* in Arabidopsis Reveals the Existence of an Additional Enzyme of Galactolipid Synthesis. *Plant Cell* 15, 2694–2706. doi: 10.1105/tpc.016675.

Kennedy, E. P., and Weiss, S. B. (1956). The Function Of Cytidine Coenzymes In The Biosynthesis Of Phospholipides. *Journal of Biological Chemistry* 222, 193–214. doi: 10.1016/S0021-9258(19)50785-2.

Khozin-Goldberg, I., Didi-Cohen, S., Shayakhmetova, I., and Cohen, Z. (2002). Biosynthesis Of Eicosapentaenoic Acid (EPA) In The Freshwater Eustigmatophyte <I>Monodus Subterraneus</I> (Eustigmatophyceae) ¹. *Journal of Phycology* 38, 745–756. doi: 10.1046/j.1529-8817.2002.02006.x.

Khozin-Goldberg, I., Yu, H. Z., Adlerstein, D., Didi-Cohen, S., Heimer, Y. M., and Cohen, Z. (2000). Triacylglycerols of the red microalga Porphyridium cruentum can contribute to the biosynthesis of eukaryotic galactolipids. *Lipids* 35, 881–889. doi: 10.1007/S11745-000-0597-8.

Kilian, O., and Kroth, P. G. (2004). Identification and characterization of a new conserved motif within the presequence of proteins targeted into complex diatom plastids: Protein targeting into diatom plastids. *The Plant Journal* 41, 175–183. doi: 10.1111/j.1365-313X.2004.02294.x.

Kim, J., Lee, J.-E., Kim, K., and Kang, N. (2018). Beneficial Effects of Marine Algae-Derived Carbohydrates for Skin Health. *Marine Drugs* 16, 459. doi: 10.3390/md16110459.

Kim, S., Yamaoka, Y., Ono, H., Kim, H., Shim, D., Maeshima, M., et al. (2013). AtABCA9 transporter supplies fatty acids for lipid synthesis to the endoplasmic reticulum. *Proceedings of the National Academy of Sciences* 110, 773–778. doi: 10.1073/pnas.1214159110.

Klezovitch, O., Brandenburger, Y., Geindre, M., and Deshusses, J. (1993). Characterization of reactions catalysed by yeast phosphatidylinositol synthase. *FEBS Letters* 320, 256–260. doi: 10.1016/0014-5793(93)80598-O.

Kobayashi, K., Awai, K., Nakamura, M., Nagatani, A., Masuda, T., and Ohta, H. (2009). Type-B monogalactosyldiacylglycerol synthases are involved in phosphate starvation-induced lipid remodeling, and are crucial for low-phosphate adaptation. *The Plant Journal* 57, 322–331. doi: 10.1111/j.1365-313X.2008.03692.x.

Kobayashi, K., Awai, K., Takamiya, K., and Ohta, H. (2004). Arabidopsis Type B Monogalactosyldiacylglycerol Synthase Genes Are Expressed during Pollen Tube Growth and Induced by Phosphate Starvation. *Plant Physiology* 134, 640–648. doi: 10.1104/pp.103.032656.

Kobayashi, K., Kondo, M., Fukuda, H., Nishimura, M., and Ohta, H. (2007). Galactolipid synthesis in chloroplast inner envelope is essential for proper thylakoid biogenesis, photosynthesis, and embryogenesis. *Proceedings of the National Academy of Sciences* 104, 17216–17221. doi: 10.1073/pnas.0704680104.

Kobayashi, K., Narise, T., Sonoike, K., Hashimoto, H., Sato, N., Kondo, M., et al. (2013). Role of galactolipid biosynthesis in coordinated development of photosynthetic complexes and thylakoid membranes during chloroplast biogenesis in Arabidopsis. *Plant J* 73, 250–261. doi: 10.1111/tpj.12028.

Kong, F., Romero, I. T., Warakanont, J., and Li-Beisson, Y. (2018). Lipid catabolism in microalgae. *New Phytol* 218, 1340–1348. doi: 10.1111/nph.15047.

Kooistra, W. H. C. F., Gersonde, R., Medlin, L. K., and Mann, D. G. (2007). "The Origin and Evolution of the Diatoms: Their Adaptation to a Planktonic Existence," in *Evolution of Primary Producers in the Sea* (Elsevier), 207–249. doi: 10.1016/B978-012370518-1/50012-6.

Kroll, D., Meierhoff, K., Bechtold, N., Kinoshita, M., Westphal, S., Vothknecht, U. C., et al. (2001). VIPP1, a nuclear gene of Arabidopsis thaliana essential for thylakoid membrane formation. *Proceedings of the National Academy of Sciences* 98, 4238–4242. doi: 10.1073/pnas.061500998.

Kruijff, B. de (1997). Lipid polymorphism and biomembrane function. *Current Opinion in Chemical Biology* 1, 564–569. doi: 10.1016/S1367-5931(97)80053-1.

Krumova, S. B., Dijkema, C., de Waard, P., Van As, H., Garab, G., and van Amerongen, H. (2008). Phase behavior of phosphatidylglycerol in spinach thylakoid membranes as revealed by 31P-NMR. *Biochimica et Biophysica Acta (BBA) - Biomembranes* 1778, 997–1003. doi: 10.1016/j.bbamem.2008.01.004.

Krumova, S. B., Laptenok, S. P., Kovács, L., Tóth, T., van Hoek, A., Garab, G., et al. (2010). Digalactosyl-diacylglycerol-deficiency lowers the thermal stability of thylakoid membranes. *Photosynth Res* 105, 229–242. doi: 10.1007/s11120-010-9581-5.

Kunst, L., Browse, J., and Somerville, C. (1988). Altered regulation of lipid biosynthesis in a mutant of Arabidopsis deficient in chloroplast glycerol-3-phosphate acyltransferase activity. *Proceedings of the National Academy of Sciences* 85, 4143–4147. doi: 10.1073/pnas.85.12.4143.

Lands, B. (2014). Historical perspectives on the impact of n-3 and n-6 nutrients on health. *Progress in Lipid Research* 55, 17–29. doi: 10.1016/j.plipres.2014.04.002.

Lands, W. E. (1960). Metabolism of glycerolipids. 2. The enzymatic acylation of lysolecithin. *J Biol Chem* 235, 2233–2237.

Latowski, D., Åkerlund, H.-E., and Strzałka, K. (2004). Violaxanthin De-Epoxidase, the Xanthophyll Cycle Enzyme, Requires Lipid Inverted Hexagonal Structures for Its Activity. *Biochemistry* 43, 4417–4420. doi: 10.1021/bi049652g.

Lee, A. G. (2000). Membrane lipids: It's only a phase. *Current Biology* 10, R377–R380. doi: 10.1016/S0960-9822(00)00477-2.

Lemoine, F., Correia, D., Lefort, V., Doppelt-Azeroual, O., Mareuil, F., Cohen-Boulakia, S., et al. (2019). NGPhylogeny.fr: new generation phylogenetic services for non-specialists. *Nucleic Acids Research* 47, W260–W265. doi: 10.1093/nar/gkz303.

Leonard, A. E., Pereira, S. L., Sprecher, H., and Huang, Y.-S. (2004). Elongation of long-chain fatty acids. *Progress in Lipid Research* 43, 36–54. doi: 10.1016/S0163-7827(03)00040-7.

Levasseur, A., Drula, E., Lombard, V., Coutinho, P. M., and Henrissat, B. (2013). Expansion of the enzymatic repertoire of the CAZy database to integrate auxiliary redox enzymes. *Biotechnol Biofuels* 6, 41. doi: 10.1186/1754-6834-6-41.

Levine, T., and Loewen, C. (2006). Inter-organelle membrane contact sites: through a glass, darkly. *Current Opinion in Cell Biology* 18, 371–378. doi: 10.1016/j.ceb.2006.06.011.

Lewin, J. C., Lewin, R. A., and Philpott, D. E. (1958). Observations on Phaeodactylum tricornutum. *Journal of General Microbiology* 18, 418–426. doi: 10.1099/00221287-18-2-418.

Leyland, B., Zarka, A., Didi-Cohen, S., Boussiba, S., and Khozin-Goldberg, I. (2020). High Resolution Proteome of Lipid Droplets Isolated from the Pennate Diatom *Phaeodactylum tricornutum* (*Bacillariophyceae*) Strain pt4 provides mechanistic insights into complex intracellular coordination during nitrogen deprivation. *J. Phycol.*, jpy.13063. doi: 10.1111/jpy.13063.

Li, C., Wang, Y., Liu, L., Hu, Y., Zhang, F., Mergen, S., et al. (2011). A Rice Plastidial Nucleotide Sugar Epimerase Is Involved in Galactolipid Biosynthesis and Improves Photosynthetic Efficiency. *PLoS Genet* 7, e1002196. doi: 10.1371/journal.pgen.1002196.

Li, L., Wang, S., Wang, H., Sahu, S. K., Marin, B., Li, H., et al. (2020). The genome of Prasinoderma coloniale unveils the existence of a third phylum within green plants. *Nature Ecology & Evolution*. doi: 10.1038/s41559-020-1221-7.

Li, N., Gügel, I. L., Giavalisco, P., Zeisler, V., Schreiber, L., Soll, J., et al. (2015). FAX1, a Novel Membrane Protein Mediating Plastid Fatty Acid Export. *PLoS Biol* 13. doi: 10.1371/journal.pbio.1002053.

Li, N., Xu, C., Li-Beisson, Y., and Philippar, K. (2016). Fatty Acid and Lipid Transport in Plant Cells. *Trends in Plant Science* 21, 145–158. doi: 10.1016/j.tplants.2015.10.011.

Li, X., Benning, C., and Kuo, M.-H. (2012a). Rapid Triacylglycerol Turnover in Chlamydomonas reinhardtii Requires a Lipase with Broad Substrate Specificity. *Eukaryot Cell* 11, 1451–1462. doi: 10.1128/EC.00268-12.

Li, X., Moellering, E. R., Liu, B., Johnny, C., Fedewa, M., Sears, B. B., et al. (2012b). A Galactoglycerolipid Lipase Is Required for Triacylglycerol Accumulation and Survival Following Nitrogen Deprivation in *Chlamydomonas reinhardtii*. *Plant Cell* 24, 4670–4686. doi: 10.1105/tpc.112.105106.

Li, X., Moellering, E. R., Liu, B., Johnny, C., Fedewa, M., Sears, B. B., et al. (2012c). A Galactoglycerolipid Lipase Is Required for Triacylglycerol Accumulation and Survival Following Nitrogen Deprivation in *Chlamydomonas reinhardtii*. *Plant Cell* 24, 4670–4686. doi: 10.1105/tpc.112.105106.

Liang, Iqbal, Wen, Tong, and Liu (2019a). Phosphorus-Induced Lipid Class Alteration Revealed by Lipidomic and Transcriptomic Profiling in Oleaginous Microalga Nannochloropsis sp. PJ12. *Marine Drugs* 17, 519. doi: 10.3390/md17090519.

Liang, J., Wen, F., and Liu, J. (2019b). Transcriptomic and lipidomic analysis of an EPA-containing Nannochloropsis sp. PJ12 in response to nitrogen deprivation. *Sci Rep* 9, 4540. doi: 10.1038/s41598-019-41169-2.

Li-Beisson, Y., Shorrosh, B., Beisson, F., Andersson, M. X., Arondel, V., Bates, P. D., et al. (2010). Acyl-Lipid Metabolism. *Arabidopsis Book* 8. doi: 10.1199/tab.0133.

Li-Beisson, Y., Shorrosh, B., Beisson, F., Andersson, M. X., Arondel, V., Bates, P. D., et al. (2013). Acyl-Lipid Metabolism. *arbo.j* 2013. doi: 10.1199/tab.0161.

Li-Beisson, Y., Thelen, J. J., Fedosejevs, E., and Harwood, J. L. (2019). The lipid biochemistry of eukaryotic algae. *Progress in Lipid Research* 74, 31–68. doi: 10.1016/j.plipres.2019.01.003.

Lima-Mendez, G., Faust, K., Henry, N., Decelle, J., Colin, S., Carcillo, F., et al. (2015). Determinants of community structure in the global plankton interactome. *Science* 348, 1262073. doi: 10.1126/science.1262073.

Listwan, S., Tokarek, W., Kleszcz, K., Chowaniec, M., Porębska, Z., Krawczyk, K., et al. (2018). Phaeodactylum Tricornutum as a Potential Phytoremediator of Sea and Fresh Waters. in doi: 10.11159/icepr18.147.

Liu, J., Liu, M., Pan, Y., Shi, Y., and Hu, H. (2021). Metabolic engineering of the oleaginous alga Nannochloropsis for enriching eicosapentaenoic acid in triacylglycerol by combined pulling and pushing strategies. *Metabolic Engineering* 69, 163–174. doi: 10.1016/j.ymben.2021.11.015.

Lommer, M., Roy, A.-S., Schilhabel, M., Schreiber, S., Rosenstiel, P., and LaRoche, J. (2010). Recent transfer of an iron-regulated gene from the plastid to the nuclear genome in an oceanic diatom adapted to chronic iron limitation. *BMC Genomics* 11, 718. doi: 10.1186/1471-2164-11-718.

Longworth, J., Wu, D., Huete-Ortega, M., Wright, P. C., and Vaidyanathan, S. (2016). Proteome response of Phaeodactylum tricornutum , during lipid accumulation induced by nitrogen depletion. *Algal Research* 18, 213–224. doi: 10.1016/j.algal.2016.06.015.

Lupette, J., Jaussaud, A., Seddiki, K., Morabito, C., Brugière, S., Schaller, H., et al. (2019). The architecture of lipid droplets in the diatom Phaeodactylum tricornutum. *Algal Research* 38, 101415. doi: 10.1016/j.algal.2019.101415.

Lupette, J., and Maréchal, E. (2018). "Phytoplankton Glycerolipids: Challenging but Promising Prospects from Biomedicine to Green Chemistry and Biofuels," in *Blue Biotechnology* (Weinheim, Germany: Wiley-VCH Verlag GmbH & Co. KGaA), 191–215. doi: 10.1002/9783527801718.ch6.

Ma, X.-N., Chen, T.-P., Yang, B., Liu, J., and Chen, F. (2016). Lipid Production from Nannochloropsis. *Marine Drugs* 14, 61. doi: 10.3390/md14040061.

Ma, Y.-H., Wang, X., Niu, Y.-F., Yang, Z.-K., Zhang, M.-H., Wang, Z.-M., et al. (2014). Antisense knockdown of pyruvate dehydrogenase kinase promotes the neutral lipid accumulation in the diatom Phaeodactylum tricornutum. *Microb Cell Fact* 13, 100. doi: 10.1186/s12934-014-0100-9.

Makshakova, O., Breton, C., and Perez, S. (2020). Unraveling the complex enzymatic machinery making a key galactolipid in chloroplast membrane: a multiscale computer simulation. *Sci Rep* 10, 13514. doi: 10.1038/s41598-020-70425-z.

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

Mann, M., Serif, M., Jakob, T., Kroth, P. G., and Wilhelm, C. (2017). PtAUREO1a and PtAUREO1b knockout mutants of the diatom Phaeodactylum tricornutum are blocked in photoacclimation to blue light. *Journal of Plant Physiology* 217, 44–48. doi: 10.1016/j.jplph.2017.05.020.

Maréchal, E. (2018). Primary Endosymbiosis: Emergence of the Primary Chloroplast and the Chromatophore, Two Independent Events. *Methods Mol Biol* 1829, 3–16. doi: 10.1007/978-1-4939-8654-5_1.

Maréchal, E. (2021). Carburants à base d'algues oléagineuses Principes, filières, verrous. *Techniques de l'Ingenieur* pp.IN186 v2.

Maréchal, E., Azzouz, N., Santos de Macedo, C., Block, M. A., Feagin, J. E., Schwarz, R. T., et al. (2002). Synthesis of Chloroplast Galactolipids in Apicomplexan Parasites. *Eukaryot Cell* 1, 653–656. doi: 10.1128/EC.1.4.653-656.2002.

Maréchal, E., Block, M. A., Joyard, J., and Douce, R. (1991). Purification of UDP-galactose: 1,2-diacylglycerol galactosyltransferase from spinach chloroplast envelopes membranes. *C.R. Acad. Sci* 313, 521–528.

Maréchal, E., Block, M. A., Joyard, J., and Douce, R. (1994). Kinetic properties of monogalactosyldiacylglycerol synthase from spinach chloroplast envelope membranes. *J Biol Chem* 269, 5788–5798.

Maréchal, E., Miège, C., Block, M. A., Douce, R., and Joyard, J. (1995). The Catalytic Site of Monogalactosyldiacylglycerol Synthase from Spinach Chloroplast Envelope Membranes. *Journal of Biological Chemistry* 270, 5714–5722. doi: 10.1074/jbc.270.11.5714.

Mareš, J., Strunecký, O., Bučinská, L., and Wiedermannová, J. (2019). Evolutionary Patterns of Thylakoid Architecture in Cyanobacteria. *Front. Microbiol.* 10, 277. doi: 10.3389/fmicb.2019.00277.

Martin-Jézéquel, V., and Tesson, B. (2012). "Phaeodactylum tricornutum polymorphism : an overview," in *Advances in Algal Cell Biology*, eds. K. Heimann and C. Katsaros (Berlin, Boston: DE GRUYTER). doi: 10.1515/9783110229615.43.

Martino, A. D., Meichenin, A., Shi, J., Pan, K., and Bowler, C. (2007). Genetic and phenotypic characterization of *Phaeodactylum tricornutum* (Bacillariophyceae) accessions [1]. *Journal of Phycology* 43, 992–1009. doi: 10.1111/j.1529-8817.2007.00384.x.

Masuda, S., Harada, J., Yokono, M., Yuzawa, Y., Shimojima, M., Murofushi, K., et al. (2011). A Monogalactosyldiacylglycerol Synthase Found in the Green Sulfur Bacterium *Chlorobaculum tepidum* Reveals Important Roles for Galactolipids in Photosynthesis. *Plant Cell* 23, 2644–2658. doi: 10.1105/tpc.111.085357.

Maumus, F., Allen, A. E., Mhiri, C., Hu, H., Jabbari, K., Vardi, A., et al. (2009). Potential impact of stress activated retrotransposons on genome evolution in a marine diatom. *BMC Genomics* 10, 624. doi: 10.1186/1471-2164-10-624.

Maxwell, K., and Johnson, G. N. (2000). Chlorophyll fluorescence—a practical guide. *Journal of Experimental Botany* 51, 659–668. doi: 10.1093/jexbot/51.345.659.

Mayers, J. J., Flynn, K. J., and Shields, R. J. (2014). Influence of the N:P supply ratio on biomass productivity and time-resolved changes in elemental and bulk biochemical composition of Nannochloropsis sp. *Bioresource Technology* 169, 588–595. doi: 10.1016/j.biortech.2014.07.048.

Mechela, A., Schwenkert, S., and Soll, J. (2019). A brief history of thylakoid biogenesis. *Open Biol.* 9, 180237. doi: 10.1098/rsob.180237.

Medlin, L. K., Kooistra, W. H. C. F., and Schmid, A.-M. M. (2000). A review of the evolution of the diatoms - a total approach using molecules, morphology and geology. *The origin and early evolution of the diatoms: fossil, molecular and biogeographical approaches.*, 13–35.

Mendiola-Morgenthaler, L., Eichenberger, W., and Boschetti, A. (1985). Isolation of chloroplast envelopes from Chlamydomonas. Lipid and polypeptide composition. *Plant Science* 41, 97–104. doi: 10.1016/0168-9452(85)90109-8.

Meng, Y., Cao, X., Yang, M., Liu, J., Yao, C., and Xue, S. (2019). Glycerolipid remodeling triggered by phosphorous starvation and recovery in Nannochloropsis oceanica. *Algal Research* 39, 101451. doi: 10.1016/j.algal.2019.101451.

Michaud, M., and Jouhet, J. (2019). Lipid Trafficking at Membrane Contact Sites During Plant Development and Stress Response. *Front. Plant Sci.* 10, 2. doi: 10.3389/fpls.2019.00002.

Miège, C., Maréchal, E., Shimojima, M., Awai, K., Block, M. A., Ohta, H., et al. (1999). Biochemical and topological properties of type A MGDG synthase, a spinach chloroplast envelope enzyme catalyzing the synthesis of both prokaryotic and eukaryotic MGDG: MGDG synthase A of spinach chloroplast envelope membranes. *European Journal of Biochemistry* 265, 990–1001. doi: 10.1046/j.1432-1327.1999.00801.x.

Mizusawa, N., Sakurai, I., Sato, N., and Wada, H. (2009). Lack of digalactosyldiacylglycerol increases the sensitivity of *Synechocystis* sp. PCC 6803 to high light stress. *FEBS Letters* 583, 718–722. doi: 10.1016/j.febslet.2009.01.021.

Mizusawa, N., and Wada, H. (2012). The role of lipids in photosystem II. *Biochimica et Biophysica Acta (BBA) - Bioenergetics* 1817, 194–208. doi: 10.1016/j.bbabio.2011.04.008.

Mock, T., Otillar, R. P., Strauss, J., McMullan, M., Paajanen, P., Schmutz, J., et al. (2017). Evolutionary genomics of the cold-adapted diatom Fragilariopsis cylindrus. *Nature* 541, 536–540. doi: 10.1038/nature20803.

Moellering, E. R., and Benning, C. (2011). Galactoglycerolipid metabolism under stress: a time for remodeling. *Trends in Plant Science* 16, 98–107. doi: 10.1016/j.tplants.2010.11.004.

Montsant, A., Jabbari, K., Maheswari, U., and Bowler, C. (2005). Comparative genomics of the pennate diatom Phaeodactylum tricornutum. *Plant Physiol* 137, 500–513. doi: 10.1104/pp.104.052829.

Mooney, B. P., Miernyk, J. A., and Randall, D. D. (1999). Cloning and Characterization of the Dihydrolipoamide *S* -Acetyltransferase Subunit of the Plastid Pyruvate Dehydrogenase Complex (E2) from Arabidopsis1. *Plant Physiology* 120, 443–452. doi: 10.1104/pp.120.2.443.

Mühlroth, A., Li, K., Røkke, G., Winge, P., Olsen, Y., Hohmann-Marriott, M., et al. (2013). Pathways of Lipid Metabolism in Marine Algae, Co-Expression Network, Bottlenecks and Candidate Genes for Enhanced Production of EPA and DHA in Species of Chromista. *Marine Drugs* 11, 4662–4697. doi: 10.3390/md11114662.

Mühlroth, A., Winge, P., El Assimi, A., Jouhet, J., Maréchal, E., Hohmann-Marriott, M. F., et al. (2017). Mechanisms of Phosphorus Acquisition and Lipid Class Remodeling under P Limitation in a Marine Microalga. *Plant Physiol.* 175, 1543–1559. doi: 10.1104/pp.17.00621.

Murakami, R., and Hashimoto, H. (2009). Unusual Nuclear Division in Nannochloropsis oculata (Eustigmatophyceae, Heterokonta) which May Ensure Faithful Transmission of Secondary Plastids. *Protist* 160, 41–49. doi: 10.1016/j.protis.2008.09.002.

Murphy, D. J. (1982). The importance of non-planar bilayer regions in photosynthetic membranes and their stabilisation by galactolipids. *FEBS Letters* 150, 19–26. doi: 10.1016/0014-5793(82)81297-0.

Mus, F., Toussaint, J.-P., Cooksey, K. E., Fields, M. W., Gerlach, R., Peyton, B. M., et al. (2013). Physiological and molecular analysis of carbon source supplementation and pH stress-induced lipid accumulation in the marine diatom Phaeodactylum tricornutum. *Appl Microbiol Biotechnol* 97, 3625–3642. doi: 10.1007/s00253-013-4747-7.

Myklestad, S. M., and Granum, E. (2009). "Biology of (1,3)-β-Glucans and Related Glucans in Protozoans and Chromistans," in *Chemistry, Biochemistry, and Biology of 1-3 Beta Glucans and Related Polysaccharides* (Elsevier), 353–385. doi: 10.1016/B978-0-12-373971-1.00010-8.

Nakai, M. (2015). The TIC complex uncovered: The alternative view on the molecular mechanism of protein translocation across the inner envelope membrane of chloroplasts. *Biochimica et Biophysica Acta (BBA) - Bioenergetics* 1847, 957–967. doi: 10.1016/j.bbabio.2015.02.011.

Nakamura, Y. (2013). Phosphate starvation and membrane lipid remodeling in seed plants. *Progress in Lipid Research* 52, 43–50. doi: 10.1016/j.plipres.2012.07.002.

Ngernprasirtsiri, J., Harinasut, P., Macherel, D., Strzalka, K., Takabe, T., Akazawa, T., et al. (1988). Isolation and Characterization of the Amyloplast Envelope-Membrane from Cultured White-Wild Cells of Sycamore (*Acer pseudoplatanus* L.). *Plant Physiol.* 87, 371–378. doi: 10.1104/pp.87.2.371.

Nitenberg, M., Makshakova, O., Rocha, J., Perez, S., Maréchal, E., Block, M. A., et al. (2020). Mechanism of activation of plant monogalactosyldiacylglycerol synthase 1 (MGD1) by phosphatidylglycerol. *Glycobiology* 30, 396–406. doi: 10.1093/glycob/cwz106.

Niu, Y.-F., Wang, X., Hu, D.-X., Balamurugan, S., Li, D.-W., Yang, W.-D., et al. (2016). Molecular characterization of a glycerol-3-phosphate acyltransferase reveals key features essential for triacylglycerol production in Phaeodactylum tricornutum. *Biotechnol Biofuels* 9, 60. doi: 10.1186/s13068-016-0478-1.

Nobusawa, T., Hori, K., Mori, H., Kurokawa, K., and Ohta, H. (2017). Differently localized lysophosphatidic acid acyltransferases crucial for triacylglycerol biosynthesis in the oleaginous alga *Nannochloropsis*. *Plant J* 90, 547–559. doi: 10.1111/tpj.13512.

Nußberger, S., Dörr, K., Wang, D. N., and Kühlbrandt, W. (1993). Lipid-protein Interactions in Crystals of Plant Light-harvesting Complex. *Journal of Molecular Biology* 234, 347–356. doi: 10.1006/jmbi.1993.1591.

Nymark, M., Sharma, A. K., Sparstad, T., Bones, A. M., and Winge, P. (2016). A CRISPR/Cas9 system adapted for gene editing in marine algae. *Scientific Reports* 6. doi: 10.1038/srep24951.

Nymark, M., Valle, K. C., Brembu, T., Hancke, K., Winge, P., Andresen, K., et al. (2009). An Integrated Analysis of Molecular Acclimation to High Light in the Marine Diatom Phaeodactylum tricornutum. *PLoS ONE* 4, e7743. doi: 10.1371/journal.pone.0007743.

Nymark, M., Valle, K. C., Hancke, K., Winge, P., Andresen, K., Johnsen, G., et al. (2013). Molecular and Photosynthetic Responses to Prolonged Darkness and Subsequent Acclimation to Re-Illumination in the Diatom Phaeodactylum tricornutum. *PLoS ONE* 8, e58722. doi: 10.1371/journal.pone.0058722.

Oborník (2019). Endosymbiotic Evolution of Algae, Secondary Heterotrophy and Parasitism. *Biomolecules* 9, 266. doi: 10.3390/biom9070266.

Ohlrogge, J., and Browse, J. (1995). Lipid Biosynthesis. *The Plant Cell* 7, 957. doi: 10.2307/3870050.

Osuna-Cruz, C. M., Bilcke, G., Vancaester, E., De Decker, S., Bones, A. M., Winge, P., et al. (2020). The Seminavis robusta genome provides insights into the evolutionary adaptations of benthic diatoms. *Nat Commun* 11, 3320. doi: 10.1038/s41467-020-17191-8.

Oudot-Le Secq, M.-P., Grimwood, J., Shapiro, H., Armbrust, E. V., Bowler, C., and Green, B. R. (2007). Chloroplast genomes of the diatoms Phaeodactylum tricornutum and Thalassiosira pseudonana: comparison with other plastid genomes of the red lineage. *Mol Genet Genomics* 277, 427–439. doi: 10.1007/s00438-006-0199-4.

Page, R. A., Okada, S., and Harwood, J. L. (1994). Acetyl-CoA carboxylase exerts strong flux control over lipid synthesis in plants. *Biochimica et Biophysica Acta (BBA) - Lipids and Lipid Metabolism* 1210, 369–372. doi: 10.1016/0005-2760(94)90242-9.

Parks, M. B., Nakov, T., Ruck, E. C., Wickett, N. J., and Alverson, A. J. (2018). Phylogenomics reveals an extensive history of genome duplication in diatoms (Bacillariophyta). *Am J Bot* 105, 330–347. doi: 10.1002/ajb2.1056.

Petroutsos, D., Amiar, S., Abida, H., Dolch, L.-J., Bastien, O., Rébeillé, F., et al. (2014). Evolution of galactoglycerolipid biosynthetic pathways – From cyanobacteria to primary plastids and from primary to secondary plastids. *Progress in Lipid Research* 54, 68–85. doi: 10.1016/j.plipres.2014.02.001.

Pidkowich, M. S., Nguyen, H. T., Heilmann, I., Ischebeck, T., and Shanklin, J. (2007). Modulating seed -ketoacyl-acyl carrier protein synthase II level converts the composition of a temperate seed oil to that of a palm-like tropical oil. *Proceedings of the National Academy of Sciences* 104, 4742–4747. doi: 10.1073/pnas.0611141104.

Popko, J., Herrfurth, C., Feussner, K., Ischebeck, T., Iven, T., Haslam, R., et al. (2016). Metabolome Analysis Reveals Betaine Lipids as Major Source for Triglyceride Formation, and the Accumulation of Sedoheptulose during Nitrogen-Starvation of Phaeodactylum tricornutum. *PLoS ONE* 11, e0164673. doi: 10.1371/journal.pone.0164673.

Pribil, M., Labs, M., and Leister, D. (2014). Structure and dynamics of thylakoids in land plants. *Journal of Experimental Botany* 65, 1955–1972. doi: 10.1093/jxb/eru090.

Radakovits, R., Jinkerson, R. E., Fuerstenberg, S. I., Tae, H., Settlage, R. E., Boore, J. L., et al. (2012). Draft genome sequence and genetic transformation of the oleaginous alga Nannochloropsis gaditana. *Nat Commun* 3, 686. doi: 10.1038/ncomms1688.

Ramachandra, T. V., Mahapatra, D. M., B, K., and Gordon, R. (2009). Milking Diatoms for Sustainable Energy: Biochemical Engineering versus Gasoline-Secreting Diatom Solar Panels. *Ind. Eng. Chem. Res.* 48, 8769–8788. doi: 10.1021/ie900044j.

Rappolt, M., Hickel, A., Bringezu, F., and Lohner, K. (2003). Mechanism of the Lamellar/Inverse Hexagonal Phase Transition Examined by High Resolution X-Ray Diffraction. *Biophysical Journal* 84, 3111–3122. doi: 10.1016/S0006-3495(03)70036-8.

Rastogi, A., Maheswari, U., Dorrell, R. G., Vieira, F. R. J., Maumus, F., Kustka, A., et al. (2018). Integrative analysis of large scale transcriptome data draws a comprehensive landscape of Phaeodactylum tricornutum genome and evolutionary origin of diatoms. *Sci Rep* 8, 4834. doi: 10.1038/s41598-018-23106-x.

Raven, J. A., and Sánchez-Baracaldo, P. (2021). *Gloeobacter* and the implications of a freshwater origin of Cyanobacteria. *Phycologia* 60, 402–418. doi: 10.1080/00318884.2021.1881729.

Rawyler, A., Unitt, M. D., Giroud, C., Davies, H., Mayor, J. P., Harwood, J. L., et al. (1987). The transmembrane distribution of galactolipids in chloroplast thylakoids is universal in a wide variety of temperate climate plants. *Photosynth Res* 11, 3–13. doi: 10.1007/BF00117669.

Reyes-Prieto, A., Weber, A. P. M., and Bhattacharya, D. (2007). The Origin and Establishment of the Plastid in Algae and Plants. *Annu. Rev. Genet.* 41, 147–168. doi: 10.1146/annurev.genet.41.110306.130134.

Rocha, J., Audry, M., Pesce, G., Chazalet, V., Block, M. A., Maréchal, E., et al. (2013). Revisiting the expression and purification of MGD1, the major galactolipid synthase in Arabidopsis to establish a novel standard for biochemical and structural studies. *Biochimie* 95, 700–708. doi: 10.1016/j.biochi.2012.11.011.

Rocha, J., Sarkis, J., Thomas, A., Pitou, L., Radzimanowski, J., Audry, M., et al. (2016). Structural insights and membrane binding properties of MGD1, the major galactolipid synthase in plants. *The Plant Journal* 85, 622–633. doi: 10.1111/tpj.13129.

Roughan, P. G., and Slack, C. R. (1982). Cellular Organization of Glycerolipid Metabolism. *Annu. Rev. Plant. Physiol.* 33, 97–132. doi: 10.1146/annurev.pp.33.060182.000525.

Ruban, A., Lavaud, J., Rousseau, B., Guglielmi, G., Horton, P., and Etienne, A.-L. (2004). The super-excess energy dissipation in diatom algae: comparative analysis with higher plants. *Photosynth Res* 82, 165–175. doi: 10.1007/s11120-004-1456-1.

Russo, M. T., Aiese Cigliano, R., Sanseverino, W., and Ferrante, M. I. (2018). Assessment of genomic changes in a CRISPR/Cas9 *Phaeodactylum tricornutum* mutant through whole genome resequencing. *PeerJ* 6, e5507. doi: 10.7717/peerj.5507.

Ryall, K., Harper, J. T., and Keeling, P. J. (2003). Plastid-derived Type II fatty acid biosynthetic enzymes in chromists. *Gene* 313, 139–148. doi: 10.1016/S0378-1119(03)00671-1.

Sakamoto, T., Higashi, S., Wada, H., Murata, N., and Bryant, D. A. (2006). Low-temperature-induced desaturation of fatty acids and expression of desaturase genes in the cyanobacterium Synechococcus sp. PCC 7002. *FEMS Microbiology Letters* 152, 313–320. doi: 10.1111/j.1574-6968.1997.tb10445.x.

Sakurai, I., Mizusawa, N., Wada, H., and Sato, N. (2007). Digalactosyldiacylglycerol Is Required for Stabilization of the Oxygen-Evolving Complex in Photosystem II. *Plant Physiology* 145, 1361–1370. doi: 10.1104/pp.107.106781.

Sapriel, G., Quinet, M., Heijde, M., Jourdren, L., Tanty, V., Luo, G., et al. (2009). Genome-Wide Transcriptome Analyses of Silicon Metabolism in Phaeodactylum tricornutum Reveal the Multilevel Regulation of Silicic Acid Transporters. *PLoS ONE* 4, e7458. doi: 10.1371/journal.pone.0007458.

Saranya, G., and Ramachandra, T. V. (2020). Biofuel and bioremediation prospects of estuarine diatoms. *IJRET* 11, 248. doi: 10.1504/IJRET.2020.10033986.

Sarkis, J., Rocha, J., Maniti, O., Jouhet, J., Vié, V., Block, M. A., et al. (2014). The influence of lipids on MGD1 membrane binding highlights novel mechanisms for galactolipid biosynthesis regulation in chloroplasts. *FASEB j.* 28, 3114–3123. doi: 10.1096/fj.14-250415.

Sato, N. (1992). Betaine lipids. *Bot Mag Tokyo* 105, 185–197. doi: 10.1007/BF02489414.

Sato, N. (2015). Is Monoglucosyldiacylglycerol a Precursor to Monogalactosyldiacylglycerol in All Cyanobacteria? *Plant Cell Physiol* 56, 1890–1899. doi: 10.1093/pcp/pcv116.

Sato, N., and Awai, K. (2017). "Prokaryotic Pathway" Is Not Prokaryotic: Noncyanobacterial Origin of the Chloroplast Lipid Biosynthetic Pathway Revealed by Comprehensive Phylogenomic Analysis. *Genome Biol Evol* 9, 3162–3178. doi: 10.1093/gbe/evx238.

Scala, S., Carels, N., Falciatore, A., Chiusano, M. L., and Bowler, C. (2002). Genome properties of the diatom Phaeodactylum tricornutum. *Plant Physiol* 129, 993–1002. doi: 10.1104/pp.010713.

Schaller, S., Latowski, D., Jemioła-Rzemińska, M., Dawood, A., Wilhelm, C., Strzałka, K., et al. (2011). Regulation of LHCII aggregation by different thylakoid membrane lipids. *Biochimica et Biophysica Acta (BBA) - Bioenergetics* 1807, 326–335. doi: 10.1016/j.bbabio.2010.12.017.

Schleiff, E., Soll, J., Küchler, M., Kühlbrandt, W., and Harrer, R. (2003). Characterization of the translocon of the outer envelope of chloroplasts. *Journal of Cell Biology* 160, 541–551. doi: 10.1083/jcb.200210060.

Schmittgen, T. D., and Livak, K. J. (2008). Analyzing real-time PCR data by the comparative CT method. *Nat Protoc* 3, 1101–1108. doi: 10.1038/nprot.2008.73.

Schreiber, V., Dersch, J., Puzik, K., Bäcker, O., Liu, X., Stork, S., et al. (2017). The Central Vacuole of the Diatom Phaeodactylum tricornutum : Identification of New Vacuolar Membrane Proteins and of a Functional Di-leucine-based Targeting Motif. *Protist* 168, 271–282. doi: 10.1016/j.protis.2017.03.001.

Schwartz, A. S., Brown, R., Ajjawi, I., McCarren, J., Atilla, S., Bauman, N., et al. (2018). Complete Genome Sequence of the Model Oleaginous Alga *Nannochloropsis gaditana* CCMP1894. *Genome Announc* 6, e01448-17, /ga/6/7/e01448-17.atom. doi: 10.1128/genomeA.01448-17.

Schwender, J., Ohlrogge, J. B., and Shachar-Hill, Y. (2003). A Flux Model of Glycolysis and the Oxidative Pentosephosphate Pathway in Developing Brassica napus Embryos. *Journal of Biological Chemistry* 278, 29442–29453. doi: 10.1074/jbc.M303432200.

Ścieszka, S., and Klewicka, E. (2019). Algae in food: a general review. *Critical Reviews in Food Science and Nutrition* 59, 3538–3547. doi: 10.1080/10408398.2018.1496319.

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

Seiwert, D., Witt, H., Janshoff, A., and Paulsen, H. (2017). The non-bilayer lipid MGDG stabilizes the major light-harvesting complex (LHCII) against unfolding. *Sci Rep* 7, 5158. doi: 10.1038/s41598-017-05328-7.

Serif, M., Dubois, G., Finoux, A.-L., Teste, M.-A., Jallet, D., and Daboussi, F. (2018). One-step generation of multiple gene knock-outs in the diatom Phaeodactylum tricornutum by DNA-free genome editing. *Nat Commun* 9, 3924. doi: 10.1038/s41467-018-06378-9.

Serôdio, J., and Lavaud, J. (2021). "Diatoms and Their Ecological Importance," in *Life Below Water* Encyclopedia of the UN Sustainable Development Goals., eds. W. Leal Filho, A. M. Azul, L. Brandli, A. Lange Salvia, and T. Wall (Cham: Springer International Publishing), 1–9. doi: 10.1007/978-3-319-71064-8_12-1.

Shaw, A. J., Szövényi, P., and Shaw, B. (2011). Bryophyte diversity and evolution: Windows into the early evolution of land plants. *American Journal of Botany* 98, 352–369. doi: 10.3732/ajb.1000316.

Siegler, H., Valerius, O., Ischebeck, T., Popko, J., Tourasse, N. J., Vallon, O., et al. (2017). Analysis of the lipid body proteome of the oleaginous alga Lobosphaera incisa. *BMC Plant Biol* 17, 98. doi: 10.1186/s12870-017-1042-2.

Simidjiev, I., Stoylova, S., Amenitsch, H., Javorfi, T., Mustardy, L., Laggner, P., et al. (2000). Self-assembly of large, ordered lamellae from non-bilayer lipids and integral membrane proteins in vitro. *Proceedings of the National Academy of Sciences* 97, 1473–1476. doi: 10.1073/pnas.97.4.1473.

Simionato, D., Block, M. A., La Rocca, N., Jouhet, J., Maréchal, E., Finazzi, G., et al. (2013). The Response of Nannochloropsis gaditana to Nitrogen Starvation Includes *De Novo* Biosynthesis of Triacylglycerols, a Decrease of Chloroplast Galactolipids, and Reorganization of the Photosynthetic Apparatus. *Eukaryotic Cell* 12, 665–676. doi: 10.1128/EC.00363-12.

Sinnott, M. L. (1990). Catalytic mechanism of enzymic glycosyl transfer. *Chem. Rev.* 90, 1171–1202. doi: 10.1021/cr00105a006.

Smith, R., Jouhet, J., Gandini, C., Nekrasov, V., Marechal, E., Napier, J. A., et al. (2021). Plastidial acyl carrier protein Δ9-desaturase modulates eicosapentaenoic acid biosynthesis and triacylglycerol accumulation in *Phaeodactylum tricornutum*. *Plant J* 106, 1247–1259. doi: 10.1111/tpj.15231.

Soedarti, T., S., T., H., S., and Kuncoro, E. P. (2017). Bioremediation of Mercury (II) Contaminated Seawater Using the Diatom Skeletonema costatum. *KLS* 3, 62. doi: 10.18502/kls.v3i6.1115.

Sorhannus, U. (2007). A nuclear-encoded small-subunit ribosomal RNA timescale for diatom evolution. *Marine Micropaleontology* 65, 1–12. doi: 10.1016/j.marmicro.2007.05.002.

Stork, S., Moog, D., Przyborski, J. M., Wilhelmi, I., Zauner, S., and 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. *Eukaryot Cell* 11, 1472–1481. doi: 10.1128/EC.00183-12.

Stukenberg, D., Zauner, S., Dell'Aquila, G., and Maier, U. G. (2018). Optimizing CRISPR/Cas9 for the Diatom Phaeodactylum tricornutum. *Front. Plant Sci.* 9, 740. doi: 10.3389/fpls.2018.00740.

Suzuki, T., Tsunekawa, S., Koizuka, C., Yamamoto, K., Imamura, J., Nakamura, K., et al. (2013). Development and disintegration of tapetum-specific lipid-accumulating organelles, elaioplasts and tapetosomes, in Arabidopsis thaliana and Brassica napus. *Plant Science* 207, 25–36. doi: 10.1016/j.plantsci.2013.02.008.

Tanaka, A., De Martino, A., Amato, A., Montsant, A., Mathieu, B., Rostaing, P., et al. (2015a). Ultrastructure and Membrane Traffic During Cell Division in the Marine Pennate Diatom Phaeodactylum tricornutum. *Protist* 166, 506–521. doi: 10.1016/j.protis.2015.07.005.

Tanaka, T., Maeda, Y., Veluchamy, A., Tanaka, M., Abida, H., Maréchal, E., et al. (2015b). Oil Accumulation by the Oleaginous Diatom *Fistulifera solaris* as Revealed by the Genome and Transcriptome. *Plant Cell* 27, 162–176. doi: 10.1105/tpc.114.135194.

Teufel, F., Almagro Armenteros, J. J., Johansen, A. R., Gíslason, M. H., Pihl, S. I., Tsirigos, K. D., et al. (2022). SignalP 6.0 predicts all five types of signal peptides using protein language models. *Nat Biotechnol*. doi: 10.1038/s41587-021-01156-3.

Tirichine, L., Rastogi, A., and Bowler, C. (2017). Recent progress in diatom genomics and epigenomics. *Current Opinion in Plant Biology* 36, 46–55. doi: 10.1016/j.pbi.2017.02.001.

Tjellström, H., Yang, Z., Allen, D. K., and Ohlrogge, J. B. (2012). Rapid Kinetic Labeling of Arabidopsis Cell Suspension Cultures: Implications for Models of Lipid Export from Plastids. *Plant Physiology* 158, 601–611. doi: 10.1104/pp.111.186122.

Toso, M. A., and Omoto, C. K. (2007). Gregarina niphandrodes may Lack Both a Plastid Genome and Organelle. *J Eukaryotic Microbiology* 54, 66–72. doi: 10.1111/j.1550-7408.2006.00229.x.

Trentacoste, E. M., Shrestha, R. P., Smith, S. R., Gle, C., Hartmann, A. C., Hildebrand, M., et al. (2013). Metabolic engineering of lipid catabolism increases microalgal lipid accumulation without compromising growth. *Proceedings of the National Academy of Sciences* 110, 19748–19753. doi: 10.1073/pnas.1309299110.

Turland, N., Wiersema, J., Barrie, F., Greuter, W., Hawksworth, D., Herendeen, P., et al. eds. (2018). *International Code of Nomenclature for algae, fungi, and plants*. Koeltz Botanical Books doi: 10.12705/Code.2018.

Underwood, G. J. C., and Paterson, D. M. (2003). "The importance of extracellular carbohydrate productionby marine epipelic diatoms," in *Advances in Botanical Research* (Elsevier), 183–240. doi: 10.1016/S0065-2296(05)40005-1.

Uthappa, U. T., Brahmkhatri, V., Sriram, G., Jung, H.-Y., Yu, J., Kurkuri, N., et al. (2018). Nature engineered diatom biosilica as drug delivery systems. *Journal of Controlled Release* 281, 70–83. doi: 10.1016/j.jconrel.2018.05.013.

Valle, K. C., Nymark, M., Aamot, I., Hancke, K., Winge, P., Andresen, K., et al. (2014). System Responses to Equal Doses of Photosynthetically Usable Radiation of Blue, Green, and Red Light in the Marine Diatom Phaeodactylum tricornutum. *PLoS ONE* 9, e114211. doi: 10.1371/journal.pone.0114211.

Vallon, O., Wollman, F. A., and Olive, J. (1986). Lateral distribution of the main protein complexes of the photosynthetic apparatus in Chlamydomonas reinhardtii and in spinach: an immunocytochemical study using intact thylakoid membranes and a PS II enriched membrane preparation. *Photobiochemistry and photobiophysics* 12, 203–220.

van Meer, G., Voelker, D. R., and Feigenson, G. W. (2008). Membrane lipids: where they are and how they behave. *Nat Rev Mol Cell Biol* 9, 112–124. doi: 10.1038/nrm2330.

Van Mooy, B. A. S., Fredricks, H. F., Pedler, B. E., Dyhrman, S. T., Karl, D. M., Koblížek, M., et al. (2009). Phytoplankton in the ocean use non-phosphorus lipids in response to phosphorus scarcity. *Nature* 458, 69–72. doi: 10.1038/nature07659.

Vancaester, E., Depuydt, T., Osuna-Cruz, C. M., and Vandepoele, K. (2020). Comprehensive and Functional Analysis of Horizontal Gene Transfer Events in Diatoms. *Molecular Biology and Evolution*, msaa182. doi: 10.1093/molbev/msaa182.

Varki, A., Cummings, R. D., Aebi, M., Packer, N. H., Seeberger, P. H., Esko, J. D., et al. (2015). Symbol Nomenclature for Graphical Representations of Glycans. *Glycobiology* 25, 1323–1324. doi: 10.1093/glycob/cwv091.

Vieler, A., Wu, G., Tsai, C.-H., Bullard, B., Cornish, A. J., Harvey, C., et al. (2012). Genome, Functional Gene Annotation, and Nuclear Transformation of the Heterokont Oleaginous Alga Nannochloropsis oceanica CCMP1779. *PLoS Genet* 8, e1003064. doi: 10.1371/journal.pgen.1003064.

Villanova, V., Fortunato, A. E., Singh, D., Bo, D. D., Conte, M., Obata, T., et al. (2017). Investigating mixotrophic metabolism in the model diatom *Phaeodactylum tricornutum*. *Phil. Trans. R. Soc. B* 372, 20160404. doi: 10.1098/rstb.2016.0404.

Vogel, G., and Eichenberger, W. (1989). "A New Betaine Lipid from Ochromonas Danica : Diacylglyceryl-O-2-(Hydroxymethyl) (N,N,N-Trimethyl)-ß-Alanine (DGTA)," in *Biological Role of Plant Lipids*, eds. P. A.

Biacs, K. Gruiz, and T. Kremmer (Boston, MA: Springer US), 167–168. doi: 10.1007/978-1-4684-1303-8_42.

von Berlepsch, S., Kunz, H.-H., Brodesser, S., Fink, P., Marin, K., Flügge, U.-I., et al. (2012). The Acyl-Acyl Carrier Protein Synthetase from *Synechocystis* sp. PCC 6803 Mediates Fatty Acid Import. *Plant Physiology* 159, 606–617. doi: 10.1104/pp.112.195263.

Wang, W., Zhao, S., Pi, X., Kuang, T., Sui, S., and Shen, J. (2020). Structural features of the diatom photosystem II–light-harvesting antenna complex. *FEBS J* 287, 2191–2200. doi: 10.1111/febs.15183.

Wang, X., Liu, Y.-H., Hu, D.-X., Balamurugan, S., Lu, Y., Yang, W.-D., et al. (2015). Identification of a putative patatin-like phospholipase domain-containing protein 3 (PNPLA3) ortholog involved in lipid metabolism in microalga Phaeodactylum tricornutum. *Algal Research* 12, 274–279. doi: 10.1016/j.algal.2015.09.005.

Watanabe, Y., Kadono, T., Kira, N., Suzuki, K., Iwata, O., Ohnishi, K., et al. (2018). Development of endogenous promoters that drive high-level expression of introduced genes in the model diatom Phaeodactylum tricornutum. *Marine Genomics* 42, 41–48. doi: 10.1016/j.margen.2018.06.003.

Widjaja, A., Chien, C.-C., and Ju, Y.-H. (2009). Study of increasing lipid production from fresh water microalgae Chlorella vulgaris. *Journal of the Taiwan Institute of Chemical Engineers* 40, 13–20. doi: 10.1016/j.jtice.2008.07.007.

Woodfield, H. K., Cazenave-Gassiot, A., Haslam, R. P., Guschina, I. A., Wenk, M. R., and Harwood, J. L. (2018). Using lipidomics to reveal details of lipid accumulation in developing seeds from oilseed rape (Brassica napus L.). *Biochimica et Biophysica Acta (BBA) - Molecular and Cell Biology of Lipids* 1863, 339–348. doi: 10.1016/j.bbalip.2017.12.010.

Xia, S., Gao, B., Li, A., Xiong, J., Ao, Z., and Zhang, C. (2014). Preliminary Characterization, Antioxidant Properties and Production of Chrysolaminarin from Marine Diatom Odontella aurita. *Marine Drugs* 12, 4883–4897. doi: 10.3390/md12094883.

Xie, W.-H., Pang, F., Niu, Y.-F., Zhang, M.-H., Yang, W.-D., Liu, J.-S., et al. (2013). Functional characterization of an ACCase subunit from the diatom *Phaeodactylum tricornutum* expressed in *Escherichia coli*: Characterization of Diatom PtACC-1 in *E. coli*. *Biotechnology and Applied Biochemistry* 60, 330–335. doi: 10.1002/bab.1091.

Xie, W.-H., Zhu, C.-C., Zhang, N.-S., Li, D.-W., Yang, W.-D., Liu, J.-S., et al. (2014). Construction of Novel Chloroplast Expression Vector and Development of an Efficient Transformation System for the Diatom Phaeodactylum tricornutum. *Marine Biotechnology* 16, 538–546. doi: 10.1007/s10126-014-9570-3.

Xu, C., Fan, J., Froehlich, J. E., Awai, K., and Benning, C. (2005). Mutation of the TGD1 Chloroplast Envelope Protein Affects Phosphatidate Metabolism in *Arabidopsis*. *The Plant Cell* 17, 3094–3110. doi: 10.1105/tpc.105.035592.

Yang, L., Chen, J., Qin, S., Zeng, M., Jiang, Y., Hu, L., et al. (2018a). Growth and lipid accumulation by different nutrients in the microalga Chlamydomonas reinhardtii. *Biotechnol Biofuels* 11, 40. doi: 10.1186/s13068-018-1041-z.

Yang, M., Lin, X., Liu, X., Zhang, J., and Ge, F. (2018b). Genome Annotation of a Model Diatom Phaeodactylum tricornutum Using an Integrated Proteogenomic Pipeline. *Molecular Plant* 11, 1292–1307. doi: 10.1016/j.molp.2018.08.005.

Yang, R., Wei, D., and Xie, J. (2020). Diatoms as cell factories for high-value products: chrysolaminarin, eicosapentaenoic acid, and fucoxanthin. *Critical Reviews in Biotechnology* 40, 993–1009. doi: 10.1080/07388551.2020.1805402.

Yang, Z., Zheng, J., Niu, Y., Yang, W., Liu, J., and Li, H. (2014). Systems-level analysis of the metabolic responses of the diatom *P haeodactylum tricornutum* to phosphorus stress. *Environ Microbiol* 16, 1793–1807. doi: 10.1111/1462-2920.12411.

Yang, Z.-K., Niu, Y.-F., Ma, Y.-H., Xue, J., Zhang, M.-H., Yang, W.-D., et al. (2013). Molecular and cellular mechanisms of neutral lipid accumulation in diatom following nitrogen deprivation. *Biotechnol Biofuels* 6, 67. doi: 10.1186/1754-6834-6-67.

Yeagle, P. L. (1989). Lipid regulation of cell membrane structure and function. *FASEB J* 3, 1833–1842.

Yongmanitchai, W., and Ward, O. P. (1991). Growth of and omega-3 fatty acid production by Phaeodactylum tricornutum under different culture conditions. *Appl Environ Microbiol* 57, 419–425. doi: 10.1128/aem.57.2.419-425.1991.

Yool, A., and Tyrrell, T. (2003). Role of diatoms in regulating the ocean's silicon cycle. *Global Biogeochem. Cycles* 17, n/a-n/a. doi: 10.1029/2002GB002018.

Yoon, K., Han, D., Li, Y., Sommerfeld, M., and Hu, Q. (2012). Phospholipid:diacylglycerol acyltransferase is a multifunctional enzyme involved in membrane lipid turnover and degradation while synthesizing triacylglycerol in the unicellular green microalga Chlamydomonas reinhardtii. *Plant Cell* 24, 3708–3724. doi: 10.1105/tpc.112.100701.

Yu, B., Wakao, S., Fan, J., and Benning, C. (2004). Loss of Plastidic Lysophosphatidic Acid Acyltransferase Causes Embryo-Lethality in Arabidopsis. *Plant and Cell Physiology* 45, 503–510. doi: 10.1093/pcp/pch064.

Yuzawa, Y., Nishihara, H., Haraguchi, T., Masuda, S., Shimojima, M., Shimoyama, A., et al. (2012). Phylogeny of Galactolipid Synthase Homologs Together with their Enzymatic Analyses Revealed a Possible Origin and Divergence Time for Photosynthetic Membrane Biogenesis. *DNA Research* 19, 91–102. doi: 10.1093/dnares/dsr044.

Zaslavskaia, L. A., Lippmeier, J. C., Kroth, P. G., Grossman, A. R., and Apt, K. E. (2001). Transformation of the diatom Phaeodactylum tricornutum (Bacillariophyceae) with a variety of selectable marker and reporter genes. *Journal of Phycology* 36, 379–386. doi: 10.1046/j.1529-8817.2000.99164.x.

Zhang, C., and Hu, H. (2014). High-efficiency nuclear transformation of the diatom Phaeodactylum tricornutum by electroporation. *Marine Genomics* 16, 63–66. doi: 10.1016/j.margen.2013.10.003.

Zhang, Y., Pan, Y., Ding, W., Hu, H., and Liu, J. (2020). Lipid production is more than doubled by manipulating a diacylglycerol acyltransferase in algae. *GCB Bioenergy*, gcbb.12771. doi: 10.1111/gcbb.12771.

Zhou, F., Liu, S., Hu, Z., Kuang, T., Paulsen, H., and Yang, C. (2009). Effect of monogalactosyldiacylglycerol on the interaction between photosystem II core complex and its antenna complexes in liposomes of thylakoid lipids. *Photosynth Res* 99, 185–193. doi: 10.1007/s11120-008-9388-9.

Zulu, N. N., Zienkiewicz, K., Vollheyde, K., and Feussner, I. (2018). Current trends to comprehend lipid metabolism in diatoms. *Progress in Lipid Research* 70, 1–16. doi: 10.1016/j.plipres.2018.03.001.

List of scientific publications

Scientific reviews:

Guéguen, N., Le Moigne, D., Amato, A., Salvaing, J., and Maréchal, E. (2021). Lipid Droplets in Unicellular Photosynthetic Stramenopiles**. Frontiers in Plant Science**. 12, 639276. doi:10.3389/fpls.2021.639276.

Guéguen, N., and Maréchal, E. (2021). Origin of cyanobacterial thylakoids via a non-vesicular glycolipid phase transition and their impact on the Great Oxygenation Event. **Journal of Experimental Botany**, erab429. doi:10.1093/jxb/erab429.

Le Moigne, D., **Guéguen, N.**, and Salvaing, J. (2022). Lipid droplets in plants: More than a simple fat storage; *in* **Advances in Botanical Research** (Elsevier), 191–223. doi:10.1016/bs.abr.2021.07.004.

Article in preparation:

Guéguen, N., Cicéron, F., Gros, V., Si Larbi, G., Jouhet, J., Hu, H., Gong, Y., Salvaing, J., Amato, A., and Maréchal, E. Characterisation of the three monogalactosyldiacylglycerol synthase isoforms in *P. tricornutum*. *In preparation.*

List of scientific communications

Oral communications:

Guéguen, N., Cicéron, F., Gros, V., Si Larbi, G., Jouhet, J., Hu, H., Gong, Y., Salvaing, J., Amato, A., and Maréchal, E. Characterisation of the Monogalactosyldiacylglycerol Synthases in *Phaeodactylum tricornutum*. Invited speaker at LIPME (UMR 2594) internal seminar. 5th of April 2022, **Castanet-Tolosan (France)**.

Guéguen, N., Cicéron, F., Gros, V., Si Larbi, G., Jouhet, J., Hu, H., Gong, Y., Salvaing, J., Amato, A., and Maréchal, E. Characterisation of the Monogalactosyldiacylglycerol Synthases in *Phaeodactylum tricornutum*. ISF-Young, organisers: Uppsala Universitet (Sweden) and Université Grenoble Alpes (France), supervised by the Royal Society of Chemistry. 26th-27th of July 2021, **online conference**.

Guéguen, N., Cicéron, F., Gros, V., Si Larbi, G., Jouhet, J., Hu, H., Gong, Y., Salvaing, J., Amato, A., and Maréchal, E. Characterisation of the Monogalactosyldiacylglycerol synthases in Phaeodactylum tricornutum. 21st scientific day of IMBL. 8th of July 2021, **Montpellier (France)**

Guéguen, N., Cicéron, F., Gros, V., Si Larbi, G., Jouhet, J., Hu, H., Gong, Y., Salvaing, J., Amato, A., and Maréchal, E. Characterisation of the Monogalactosyldiacylglycerol Synthases in *Phaeodactylum tricornutum*. I-BE-C. 16th-18th of February 2021, **web congress**.

Guéguen, N., Cicéron, F., Gros, V., Si Larbi, G., Jouhet, J., Hu, H., Gong, Y., Salvaing, J., Amato, A., and Maréchal, E. Characterisation of the Monogalactosyldiacylglycerol Synthases in *Phaeodactylum tricornutum*. Lipids in the Ocean, organisers: CESAM & LAQV & University of Aveiro (Portugal), LEMAR & IUEM & Université de Bretagne (France). 18th of November 2020, **online conference**.

Guéguen N., Cicéron F., Falconet D., Jouhet J., Salvaing J., Amato A. and Maréchal E. Characterisation of glycerolipid synthases in *Phaeodactylum tricornutum*. Alg'in Provence, European Workshop. 1st-2nd of October 2019, **Arles (France)**.

Poster communications:

Guéguen, N., Cicéron, F., Gros, V., Si Larbi, G., Jouhet, J., Hu, H., Gong, Y., Salvaing, J., Amato, A., and Maréchal, E. Characterisation of Monogalactosyldiacylglycerol synthases in the model diatom *Phaeodactylum tricornutum* using CRISPR-Cas9 technology. Gerli 16th international lipidomics meeting. 26th-29th of September 2021, **Bordeaux (France)**.

Guéguen, N., Cicéron, F., Gros, V., Si Larbi, G., Jouhet, J., Hu, H., Gong, Y., Salvaing, J., Amato, A., and Maréchal, E. Characterisation of the Monogalactosyldiacylglycerol Synthases in *Phaeodactylum tricornutum*. ISF-Young, Organisers: Uppsala Universitet (Sweden) and Université Grenoble Alpes (France), supervised by the Royal Society of Chemistry. 26th-27th of July 2021, **online conference**.

Guéguen, N., Cicéron, F., Gros, V., Si Larbi, G., Jouhet, J., Hu, H., Gong, Y., Salvaing, J., Amato, A., and Maréchal, E. Characterisation of the Monogalactosyldiacylglycerol Synthases in *Phaeodactylum tricornutum*. Molecular Life of Diatoms 6. 12th-14th of July 2021, **online conference**.

Guéguen N., Cicéron F., Falconet D., Jouhet J., Salvaing J., Amato A. and Maréchal E. Characterisation of glycerolipid synthases in *Phaeodactylum tricornutum*. Glyco@Days. 4th-5th of November 2019, **Autrans (France)**.

Guéguen N., Cicéron F., Falconet D., Jouhet J., Salvaing J., Amato A. and Maréchal E. Characterisation of MonoGalactosylDiacylGlycerol synthases in the model diatom Phaeodactylum tricornutum using CRISPR-Cas9 technology. 19th scientific day of IMBL. 21st of May 2019, **Montpellier (France)**.

Invited seminars:

Bolik S., **Guéguen N**. A focus on microalgae: from ecological to physiological approaches. Plymouth Marine Laboratory. 24th of July 2019, **Plymouth (United Kingdom)**.

Organisation of conferences:

ISF-Young. Organisers: Uppsala Universitet (Sweden) and Université Grenoble Alpes (France), supervised by the Royal Society of Chemistry. Conference Committee members: **Guéguen, N.** (Co-chair), Redman H.-J. (Co-chair), Kaeffer, N., Le Moigne, D., Lorenzi, M., McCarthy, B. 26th-27th of July 2021, **Webinar**.

Participation to the tasks of common interest in the laboratory

- Co-responsible for the 'biomolecular room': Security formation of new room users, monitoring of the level of consumables in the room, evacuation of the biochemical waste.
- Authorisation to manage the autoclave: Co-responsible for launchings and openings of the autoclave (sterile material, sterile medium, waste).

Formations during the PhD

Category: researcher activity and job practice

☛ THURSDAY OF SECURITY: Biological risks (20th of December 2018) AMPHI 3 Stendhal - 1180 avenue Centrale

3.5 hours

Category: Scientific formation

☛ Marine Biodiversity: Expertise in Flora (105 hours of formation) Roscoff Marine Station

40 hours

Category: cross disciplinary formation

☛ Papirus 2020 session n°3 (9th of September 2020 – 6th of November 2020) Philippe Lenoir, platform of e-formation INRAE

35 hours

Category: International, Languages

☛ Create your CV in English - 1 day (29th of May 2017) PROGRESS - 2 impasse Michel Labrousse - La Maison des lois - 31100 Toulouse

7 hours

Category: Thesis methodology - literature search and publication

☛ Research bibliometric and evaluation- 1 day (19th of October 2017) Federal University of Toulouse Midi-Pyrénées - Maison de la Recherche et de la Valorisation (MRV) 118 route de Narbonne

6 hours

☛ Manage literature resources with Zotero: initiation - 1 day (13th of December 2017) Federal University of Toulouse Midi-Pyrénées - Maison de la Recherche et de la Valorisation (MRV) 118 route de Narbonne

6 hours

Category: Career pursuit

☛ Communication in public - 2 days (2nd of May 2017 – 3rd of May 2017) PROGRESS - 2 impasse Michel Labrousse - La Maison des lois - 31100 Toulouse

13 hours

☛ Meeting direction - 1 day (23rd of November 2017) PROGRESS - 2 impasse Michel Labrousse - La Maison des lois - 31100 Toulouse

7 hours

☛ Mobility/international career preparation- 1 day (6th of June 2018) Federal University of Toulouse Midi-Pyrénées - Maison de la Recherche et de la Valorisation (MRV) 118 route de Narbonne

7 hours

☛ How to present oneself to a recruitment consultant - 1 day (11th of June 2018) PROGRESS

7 hours

Category: Pedagogic practice for the University

☛ Understanding of what is learning, for efficient teaching - 1 day (18th of January 2018) Federal University of Toulouse Midi Pyrénées- Maison de la Recherche et de la Valorisation (MRV) 118 route de Narbonne

6 hours

☛ Body and space control - 1 day (23rd of June 2017) Federal University of Toulouse Midi Pyrénées - Maison de la Recherche et de la Valorisation (MRV) 118 route de Narbonne

6 hours

Category: Scientific

☛ Project management for thesis - 1 day (24th of November 2017) PROGRESS - 2 impasse Michel Labrousse - La Maison des lois - 31100 Toulouse

7 hours

☛ Scientific ethic and integrity - 1 day (16th of February 2018) Federal University of Toulouse Midi Pyrénées - Maison de la Recherche et de la Valorisation (MRV) 118 route de Narbonne

7 hours

☛ GlobalHealth - Biodiversity, Ecosystems Health, Human Healt - Agreenium - LYON

18 hours

☛ Farming concerns for tomorrow's society - Agreenium - Rennes (Agrocampus Ouest)

30 hours

Total participations: 205.5 hours / 16 modules

Total Credits: 35

Teaching

- 32 hours of practical work of chemistry to 1st year engineer students at INSA Toulouse
- Supervision of a master 1 student for seven weeks

CV

Nolwenn GUÉGUEN

+33 6 04 15 44 07

gueguen.nolwenn4@gmail.com

LinkedIn : nolwenn-guéguen

ENGINEER-PhD STUDENT

Passionate about research in plant lipid metabolism

PROFESSIONAL EXPERIENCE

PhD student in biology

CEA - LPCV, Grenoble, France | Nov 2018 – May 2022

Characterisation of the monogalactosyldiacylglycerol synthases in the microalgae *Phaeodactylum tricornutum*

Research engineer

INRAE - LIPME, Castanet-Tolosan, France | Oct 2016 – Sept 2018

Role of the transcription factor MYB30 in seed development and protection in *Arabidopsis thaliana*

Researcher intern

INRAE - LIPME, Castanet-Tolosan, France | Jan 2016 – May 2016

Identification and characterisation of new protein partners of MIEL1 in pollen in *Arabidopsis thaliana*

Researcher intern

Gembloux Agro-Bio Tech, Belgium | Aug 2015 – Sept 2015

Study of antibiotics expression in bacterial strains valuable as biopesticides

EDUCATION

Doctorate in molecular biology

University Grenoble Alpes, Grenoble | Nov 2018 – May 2022

Engineer in Biochemical Engineering

INSA Toulouse| 2011 – 2016

Master2 Research Plant Biosciences

University Paul Sabatier| 2015 – 2016

Baccalaureate S, SVT - with distinction

Lycée Renée Cassin, Montfort sur Meu | 2011

SCIENTIFIC AND TECHNICAL SKILLS

- Molecular Biology
- Genetic Engineering
- Microalgae cultivation, initiation to PBR
- Microalgal metabolism and physiology
- DNA, RNA, lipid, and protein extraction
- GC-FID
- Lipid metabolism
- CRISPR-Cas9, biolistic transformation

Software:

- Image J
- PYMOL
- Matlab
- R studio
- MegaX
- Photoshop
- Microsoft Office

LANGUAGES

French:
Mother tongue

English:
Fluent (read, written, spoken)
TOEIC: 920/990

Spanish:
Fluent (read, written, spoken)

Portuguese and Italian:
Elementary

INTERPERSONAL SKILLS

- **Oral communication in English**
  - Presentation at international conferences
- **Written communication in English**
  - Writing of scientific articles
  - Posters
- **Teaching**
  - Supervising of M1 intern
  - Practical work of Chemistry at INSA Toulouse
- **Project management**
- **Team work**
- **Adaptable**
- **Autonomous**
- **Organised**
- **Creative**

HOBBIES

Karate (black belt), **gym**

Painting, digital painting, pyrography

Accordion

Abstract

Monogalactosyldiacylglycerol (MGDG) is a membrane lipid class found abundantly in photosynthetic membranes. MGDG is known to play roles in thylakoid biogenesis, and in stabilisation and function of the photosystems. Despite its importance, few information is available on MGDG synthesis in complex plastids found in some eukaryotic algae. In particular, diatoms contain a secondary plastid surrounded by four membranes in which the precise glycerolipid composition is still unknown. In this thesis, we focused on the three MGDG synthase (MGD) isoforms detected by sequence homology in the model diatom *Phaeodactylum tricornutum*.

Phylogenetic analysis supports that the existence of the three types of MGD identified in *P. tricornutum* seems a conserved feature in diatoms. GFP-fusion localisation by confocal microscopy showed that each isoform (numbered α, β, and γ) is localised to a distinct subcellular compartment. MGDα is located inside the plastid at the vicinity of thylakoids. MGDβ is in the periplastidial vesicular network (known also as the "blob") and the periplastidial membrane. MGDγ is an intriguing cytosol-localised isoform. We also detected an MGDγ-like sequence in genomic data from basal apicomplexans of the Gregarina phylum, missing the usual vestigial plastid of this clade. Together, these results indicate that in these complex organisms, a MGD has been relocalised to extraplastidial compartments of the cell during evolution.

The successful CRISPR-Cas9 edition of targeted MGD genes in *P. tricornutum* was achieved. Several knock-out mutants were generated for each isoform. Although MGDα is the most expressed isoform and localize at the vicinity of photosynthetic membranes, seeming therefore to act as the main isoform in the cell, MGDα knockout did not lead to any strong alteration in cell growth, photosynthetic activity and membrane integrity. Neither did MGDβ and MGDγ knockout lines. Such phenotype suggest that highly efficient compensation mechanisms exist under optimal growth conditions. In wild type cells cultivated in low nitrogen conditions, MGDα and MGDγ protein levels drop. Cultivation of the MGD mutant strains revealed that MGDβ and MGDγ may play an important role under such abiotic stress. Such specific role should be investigated in the future.

Variation of the fatty acid profiles was observed in different classes of membrane glycerolipids. The impact on MGDG suggests that each isoform has distinct substrate preferences, which might be linked to substrate availability in their respective subcellular compartment. Mutation of the MGDγ isoform led to a strong alteration of the composition of several extraplastidial lipid classes, compared to mutations of MGDα and MGDβ isoforms, suggesting a critical role in lipid homeostasis at the whole cell level. Future prospects on MGDG synthesis, localisation and function are proposed.

Résumé

Le monogalactosyldiacylglycérol (MGDG) est une classe de lipide rencontrée abondamment dans les membranes photosynthétiques. Le MGDG est connu pour jouer des rôles dans la biogénèse des thylakoïdes, et dans la stabilisation et la fonction des photosystèmes. Malgré son importance, peu d'informations sont disponibles sur la synthèse du MGDG dans des plastes complexes rencontrés dans certaines algues eucaryotes. En particulier, les diatomées contiennent un plaste secondaire entouré par quatre membranes dans lesquelles la composition précise en glycérolipides est encore inconnue. Dans cette thèse, nous nous sommes concentrés sur les trois isoformes de MGDG synthase (MGD) détectées par homologie de séquence chez la diatomée modèle *Phaeodactylum tricornutum*.

L'analyse phylogénétique soutient que l'existence de trois types de MGD identifiés chez *P. tricornutum* semble être un trait conservé chez les diatomées. La localisation de fusions GFP par microscopie confocale a montré que chaque isoforme (numéroté α, β et γ) est localiseé dans un compartiment subcellulaire différent. MGDα est localisée à l'intérieur du plaste au voisinage des thylakoïdes. MGDβ se trouve dans le réseau vésiculaire périplastidial (aussi connu en tant que « blob ») et dans la membrane périplastidiale. MGDγ est une isoformee intriguante localisée dans le cytosol. Nous avons également détecté une séquence similaire à MGDγ dans des données génomiques d'un clade basal d'apicomplexe du phylum Grégarine, où le plaste vestigial habituel de ce groupe est absent. Ensemble, ces résultats indiquent que dans ces organismes complexes, une MGD a été relocalisée vers des compartiments extraplastidiaux de la cellule au cours de l'évolution.

L'édition des gènes de MGDs cibles par CRISPR-Cas9 chez *P. tricornutum* a été réalisée avec succès. Plusieurs mutants « knock-out » ont été générés pour chaque isoforme. Bien que MGDα soit l'isoforme la plus exprimée dans la cellule, la suppression de l'expression de MGDα n'a pas conduit à de fortes altérations de la croissance cellulaire, de l'activité photosynthétique, ou de l'intégrité des membranes. De même, l'absence d'expression de MGDβ et MGDγ n'a conduit à aucun effet à l'échelle cellulaire. Un tel phénotype suggère que des mécanismes de compensation très efficaces existent en conditions optimales de culture. Chez la souche sauvage cultivée en carence d'azote, le niveau protéique de MGDα et MGDγ chute. La culture de mutants de MGDs a révélé que MGDβ et MGDγ pourraient jouer un rôle important sous de telles conditions de stress. Ce rôle spécifique devrait être étudié dans le futur.

La variation des profils d'acides gras a été observée chez les différentes classes de glycérolipides membranaires. L'impact sur le MGDG suggère que chaque isoforme présente des préférences de substrat différentes, qui pourraient être liées à la disponibilité du substrat dans leur compartiment cellulaire respectif. La mutation de l'isoforme MGDγ a mené à de fortes altérations de la composition de plusieurs classes de lipides extraplastidiaux, comparée à celles des mutations des isoformes MGDα et MGDβ, suggérant un rôle critique dans l'homéostasies des lipides à l'échelle de la cellule entière. Des perspectives futures sur la synthèse, localisation et fonction du MGDG sont proposées.

Keywords/mots-clefs: CRISPR-Cas9 ; Monogalactosyldiacylglycerol ; glycerolipids ; diatoms ; *Phaeodactylum tricornutum.*