uis.no

Dmitry Kechasov

Biochemical and molecular analysis of LC-PUFA biosynthesis in the microalga *Nannochloropsis*

Master thesis 2013

The thesis is submitted as a part of the
Master's degree study in Biological Chemistry
at the University of Stavanger

University of Stavanger

Faculty of Science and Technology

MASTER'S THESIS

Study program/ Specialization: Master program in Biological Chemistry	Spring semester, 2013 **Open** access
Writer: Dmitry Kechasov	…… (Writer's signature)
Faculty supervisor: Professor Sigrun Reumann External supervisor(s):	
Title of thesis: Biochemical and molecular analysis of LC-PUFA biosynthesis in the microalga *Nannochloropsis*	
Credits (ECTS): 60	
Key words: LC-PUFA, lipid staining, microalgae, *Nannochloropsis,* peroxisomes, cloning, transformation, contamination	Pages: 105 + enclosure: 5 Stavanger, 17.06.2013

Acknowledgement

I would like to express my gratitude to all those who gave me the possibility to complete this thesis.

First, and foremost, I would like to thank my supervisor, Prof. Sigrun Reumann for her friendly, encouraging and soft attitude and for support of all my ideas.

My thanks also go to Dr. T. Lingner, University of Göttingen (Germany), for providing the essential data for the peroxisomal protein prediction.

I would also like to take this opportunity to thank my friends and colleges from the Center for Organelle Research at the University of Stavanger, those who made a contribution to my thesis: Manish Budathoki, Kirsti Sørhagen, Aline Benichou, Gopal Chowdhary and others.

I would like to thank the University of Stavanger and whole Norway for giving me this amazing opportunity to be a student here.

Last but not least, I would like to give my special thanks to my wife, Liubov Fedotova, for her love, patience and comprehensive support.

Abstract

The present research project comprised both basic research and biotechnology and focused on investigating the role of peroxisomes in lipid metabolism in the lipid-rich microalgae, *Phaeodactylum* and *Nannochloropsis*, and on trying to optimize their productivity in synthesizing omega-3 fatty acids for Norwegian aquaculture by genetic engineering and technology-assisted breeding. The starting point of the present thesis was to develop the fundamental microalgal methodology in the Reumann group and to obtain important preliminary research data on peroxisome biology and fatty acid metabolism in order to establish the research group in this field and potentially qualify for research funding in the new BIOTEK 2021 program of the Norwegian Research Council. The objectives of this M. Sc. study were the following: a) establishment of basic culturing techniques and analytical methodology; b) prediction and analysis of the PTS1/2 proteome of soluble matrix proteins for *Nannochloropsis gaditana*; and c) cloning of full-length cDNAs or C-terminal exons of predicted PTS1 proteins in N. gaditana for subsequent subcellular localization and functional studies. After the fulfillment of extensive empirical and theoretical work, basic culturing techniques for Nannochloropsis were established, including those for analysis and optimization of growth conditions, subculturing and harvesting. The methodology for visualizing the accumulation of neutral lipids in lipid bodies by fluorescence microscopy upon nitrogen starvation was developed, and a protocol for DNA isolation and single-exon cloning was set-up. Among six *Nannochloropsis* species investigated, *N. gaditana* showed the most promising growth and productivity characteristics, followed by *N. oceanica*. The comprehensive analysis of the PTS1/2 proteome of soluble matrix proteins was carried out for *Nannochloropsis gaditana* and *N. oceanica*, combining various bioinformatic methods. The identified PTS1 proteins included not only conserved orthologs of known Arabidopsis PTS1 proteins, but also several novel genus- and species-specific proteins. To validate the predictions, two full-length cDNAs and one C-terminal exon of predicted PTS1 proteins in *N. gaditana* were cloned for subsequent subcellular localization. In addition nuclear transformation of *N. oceanica* was successfully attempted, as indicated by antibiotic resistance conferred by a given plasmid. In summary, the present Master project made major and highly significant contributions to the establishment of basic and applied microalgal research in the plant peroxisome group by Prof. Reumann at UiS and to National funding of the 10-Mio NOK BIOTEK2021 project “Microalgae 2021” in June 2013.

Table of Contents

1.Introduction

Algae represent a very polyphyletic group of organisms with chloroplasts. Such different organisms as blue-green bacterias (prokaryots), well-known unicellular model algae *Chlamidomonas* (Green Algae) or 45 meters-long giant kelp (Brown algae) (Varun, 2005). This group consists of several phylogenetic branches that merged and split a few times to form modern variety of species (McFadden, 2001). Algae live mostly in water, but they can be also found on a wet terrestrial objects as a bark of woods or even inside of lichens.

Algal organism can be formed by one single cell (unicellular) or by many cells (multicellular). At first sight, organisms which consist of one cell are simpler than multicellular organisms which are larger in size, but 'simple' does not mean 'weak', 'powerless'. Conversely, unicellular algae are more abundant in our world. They grow faster, evolve quicker and have enormous potentials in biological synthesis and they can be found in almost every ecological niche one can imagine. Usually unicellular algae are called as microalgae, while multicellular, visible by naked eye, are called as macroalgae.

The outstanding property of algae is their photosynthetic ability. Photosynthesis is an ability that distinguishes algae and plants from heterotrophic organisms. Algae can grow without any carbon source except carbon dioxide in air. They can synthesize compounds using the energy of sun which is an inexhaustible source of energy.

1.1.Microalgae

The scientific community has recently become much interested in algae. Humans face global energy problems as well as waste utilization ones and wish to solve these problems with new effective methods. Oil and gas are still cheap and available, but most of the oil reserves are concentrated on the territory of only a few countries. Fossil energy resources are not renewable and will be probably used up during the XXI century. Another acute problem is waste production and recycling. Waste recycling is energy-dependent and expensive. A third problem concerns the lack of food on the large territory of the world. All these challenges stimulate scientists to find organisms that can produce energy-rich products using a small amount of in-put resources in a limited period of time. On our planet we have only one large group of producers - photosynthetic organisms. The problem is that higher plants are very complex and specific organisms. It is therefore wiser to focus

on photosynthetic microorganisms when it comes to biotechnological production of high-energy and high-value bioproducts.

Microalgae are very promising candidates for future production of energy-rich substances, such as lipids and proteins. However, the development of algae cultivation is only beginning nowadays. There are major problems that have to be solved in order to get cost efficient results. Advantages and disadvantages of using microalgae are represented in Table 1.

Table 1. Advantages and disadvantages of using microalgae in biotechnology.

Advantages	Disadvantages
Fast growth rate	Very small size of the cells
No need of arable lands	Relatively low cell density
Sufficiency of CO2 as the only carbon source	High growth rates require the provision of concentrated CO2
Use of light as sole energy source	High growth rates require additional continuous illumination
Microalgae can accumulate lipids up to 50% wet weight	The fastest rates of lipids accumulation occurs under stress conditions
Marine algae can grow just in normal sea water	Energy expending extraction of lipids or other products

1.1.1.Evolution of algae

Algae represent a very large group of diverse organisms with unclear origin and relationships between taxa. The diversity of this group arises from their morphology which shows traces of complex evolutionary transformations.

In the beginning of the XX century the similarity between plastids and cyanobacteria was proposed for the first time by Schimper (1833). Later, the Russian biologist Mereschkowsky proposed the first endosymbiotic theory in the 1905 and made the first phylogenetic tree based on multiple endosymbiotic origins of plastids (both articles were written on german language, for details see McFadden, 2001). Figure 1 shows a phylogenetic tree of the origin and distribution of plastids between major groups that participated in the primary and secondary endosymbiotic processes. In the first stage the phagotrophic eukaryotic cell did not digest but retained the cyanobacteria-like phototrophic cell inside of its own cytosol. The new cell carried two genomes and was the result of the

primary endosymbiosis. From this organism three new clades evolved - these are plants/ green algae, red algae and glaucophytes. They differ in the presence of a number of metabolic pathways and properties of the photosynthetic apparatus. Several independent secondary endosymbioses occurred involving uptake of primary endosymbiotic cells by eukaryotic phagotrophs. According to one of the major theories, the heterokonts/ *Stramenopiles* evolved after the integration of red algal cell into a eukaryotic phagotroph (McFadden, 2001; Keeling, 2010). The diversity of plastids inside of the heterokonta group can be explained by several tertiary endosymbiosis after which, for example, the diatoms were born.

Novel techniques use genomic information to study evolutional diversity of organisms. The genomes of plastids are small in comparison to nuclear genomes, in addition to their different origin. This is one of the reasons why phylogenetic trees based on plastid genes alignment differs from those based on highly conserved genes, for example, ribosomal 18S RNA.

Two phylogenetic trees for Nannochloropsis can be found in Figures 2a and 2b, the first is based on the diversity of sequences of six highly conserved genes. Figure 2b shows relationship inside of Nannochloropsis genus based on ribosomal 18S RNA (Tyler et al., 2006; Radakovits et al., 2012).

The genomes of some *Stramenopiles* are available: *Nannochlorpsis gaditana, N. oceanica, Ectocarpus siliculosus, Phaeodactylum tricornutum, Talassiosira pseudonana, T.oceanica, Aureococcus anophagefferens, Phytophtora sp.* These species, representing algal- and fungal-like groups and were used in this research as ones, most closely related to *Nannochloropsis*.

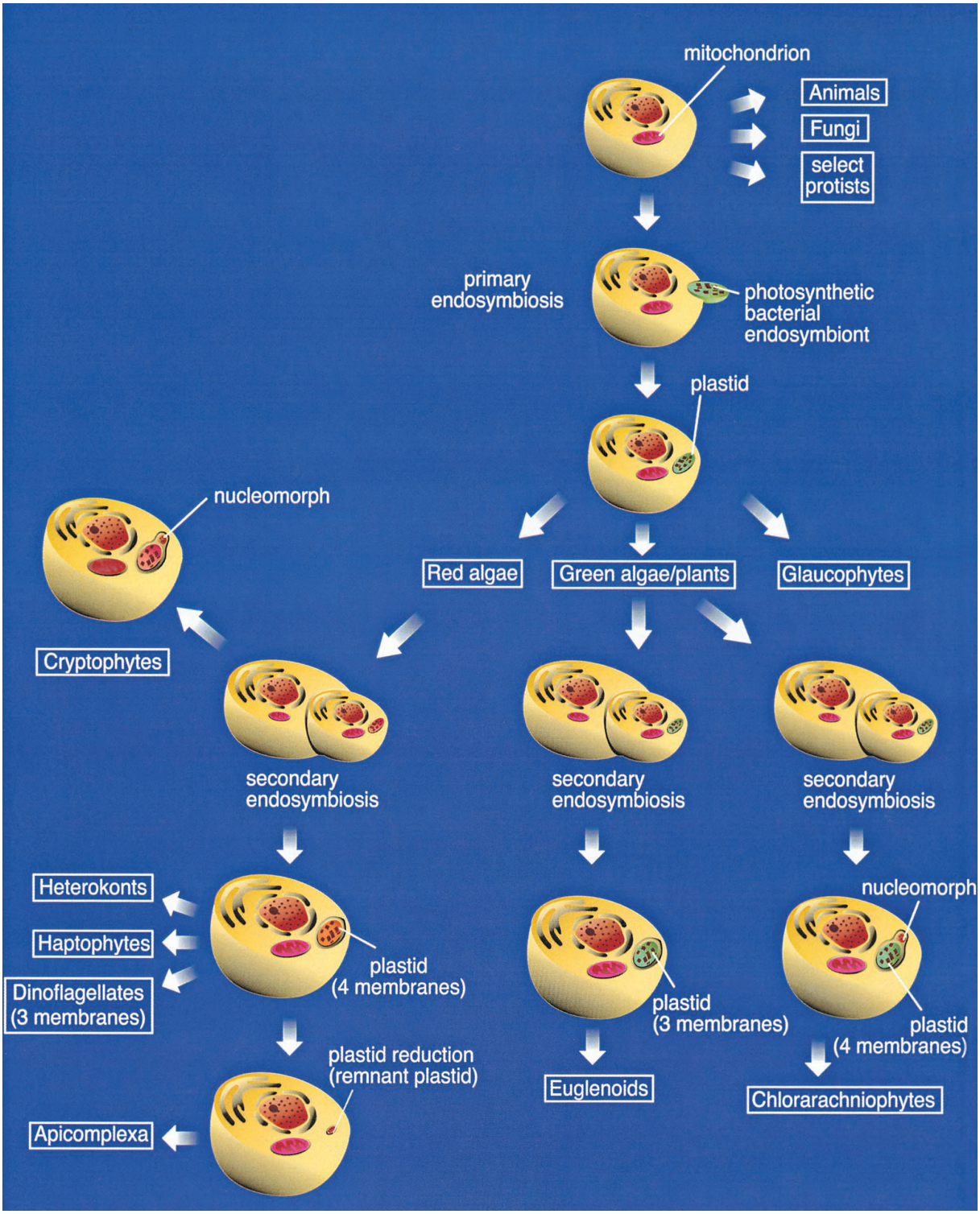


Figure 1. Hypotetic scheme reflecting evolution events involving primary and secondary endosymbiosis (McFadden, 2001). There appears to have been a single primary endosymbiosis in which a photosynthetic cyanobacterial-like prokaryote was engulfed and retained by a eukaryotic phagotroph. This event spawned three extant lineages: red algae, green algae/plants, and glaucophytes. These primary plastids are bounded by two envelope membranes. At least three secondary endosymbioses, in which a eukaryotic alga was engulfed and retained by eukaryotic phagotroph producing plastids with either three or four membranes, are known. At least one other secondary endosymbiosis, this time involving the engulfment of a red algal-like endosymbiont, occurred. This event led to the cryptophytes, which still retain a remnant of the red algal nucleus, namely the nucleomorph. Some interpretations suggest that this secondary endosymbiotic event was the origin of heterokonts, haptophytes, dinoflagellates, and apicomplexa plastids.

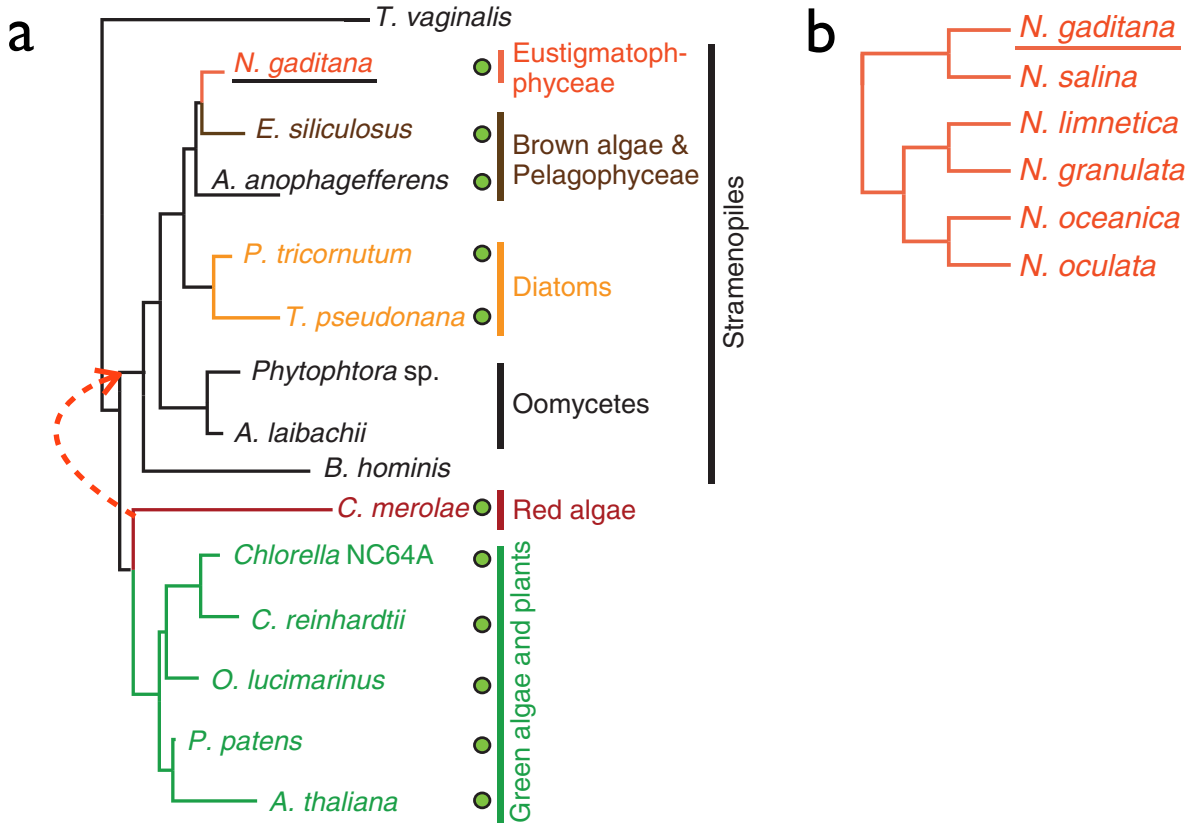


Figure 2. Phylogenetic analysis of the *N. gaditana* genome from Radakovits et al., 2012. (**a**) Schematic phylogenetic tree of *Stramenopiles* and photosynthetic algae. The dashed red line shows hypotetical secondary endosymbiotic event, when red algal cell was taken up by eukaryotic phagotroph resulting in the *stramenopile* origin (Tyler et al., 2006). Filled green circles on the right indicate photosynthetic species. (**b**) The tree indicates the relationship between different strains of *Nannochloropsis* based on 18S ribosomal RNA gene sequences.

1.2.Lipids

The term «total lipids» is defined as «the biochemical compounds not soluble in water, but soluble in organic solvents instead» from the Greenwell et al., 2010, P. 705. This is a very unspecified definition which includes molecules with different size and, more importantly, chemical properties. Lipids are traditionally divided into two classes: polar and neutral. Polar lipids are commonly represented by phospholipids (e.g. phosphatidylinositol, phosphatidylcholine, phosphatidylethanolamine) and glycolipids, which are combinations of (oligo-)saccharides and lipids (Figure 3). The presence of phosphate molecules and carbohydrates provides polar property of this class, representatives of which are more abundant in membranes. Neutral lipids are less polar due to the absence of high polar molecules and represented by the tri-, di- and monoglycerides, waxes and isoprenoid-type

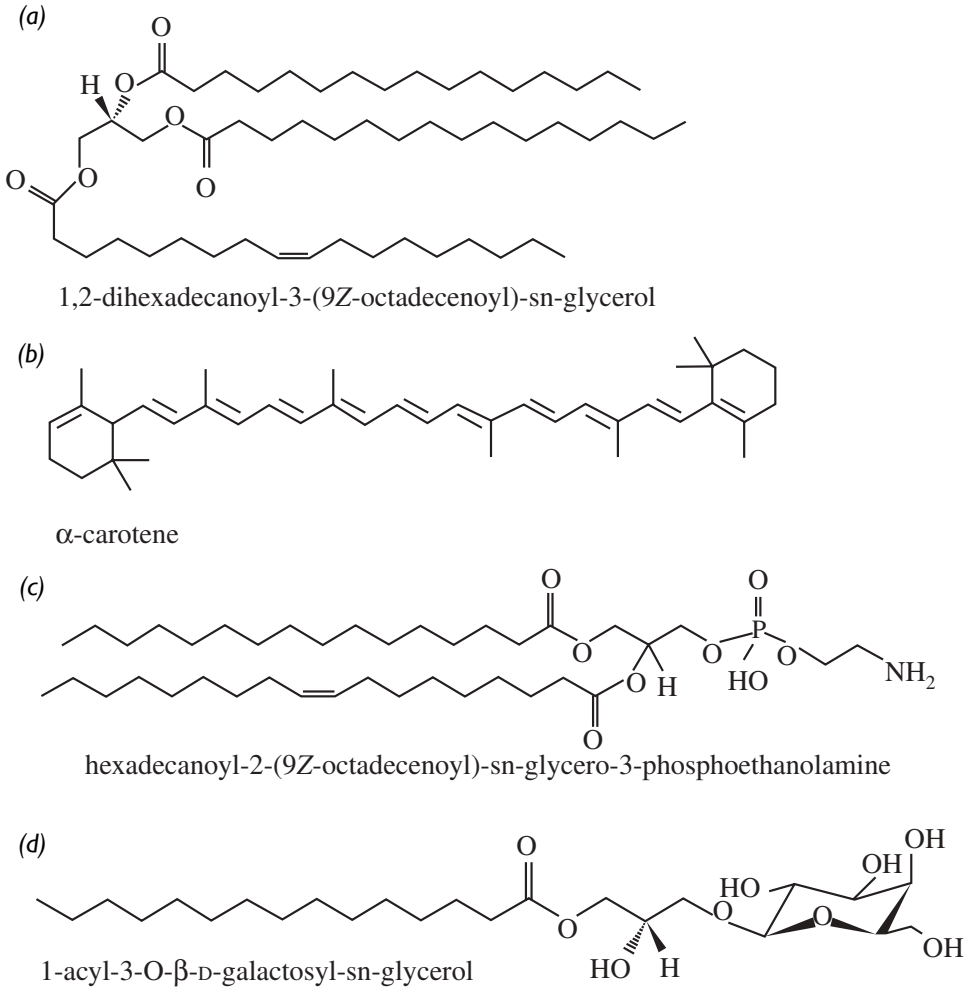


Figure 3. Overview of the chemical structures of the most common representatives from two lipid classes. Neutral lipids: triacylglycerides (a) and carotenoids (c); polar: phospholipids (b) and glycolipids (d) **(Greenwell et al., 2010)**.

The term «fat» is often applied to animal «solid» lipids. While the term «oil» is mostly used for plant and algal lipids. Algal lipids are liquid because of the high content of long chain fatty acids with high degree of unsaturation.

1.2.1.Long chain polyunsaturated fatty acids

Long-chain polyunsaturated fatty acids (LC-PUFAs) have a chain length of 20 and 22 carbons with two to six methylene-interrupted double bonds (not conjugated duble bonds, Δ3, Δ6, Δ9). In algae they perform energy-conserving function, stabilize membrane structure under stress conditions and protect against an overproduction of reactive oxygen species (Hoffmann et al., 2010). LC-PUFAs have some vital functions in humans being an important part of the membranes. Deficiency of LC-PUFAs can cause mental illness and even cardiac death (Khozin-Goldberg et al., 2011). PUFAs are subdivided in the two

groups by the number of carbons between the last double bond from the methyl end (ω) of the acyl chain: ω-3 and ω-6.

The capability of PUFA synthesis in animals is very limited. Some FA have to be ingested by the body with food. They are called essential PUFA. By contrast, plants and especially microorganisms (including algae) can often synthesis them.

The plant-derived human-essential PUFA linoleic acid (LA, 18:2 ω-6, this means 18 carbon fatty acids with two double bonds, the last double bond is located six carbons from the ω end of the carbon chain) and α-linolenic acid (ALA, 18:3 ω-3), are precursors for the LC-PUFA of the ω-6 group arachidonic acid (ARA, 20:4 ω-6) and of the ω-3 group eicosapentaenoic acid (EPA, 20:5 ω-3) and docosahexaenoic acid (DHA, 22:6 ω-3). ARA, EPA, DHA are not strictly essential, but the capability of human organism to their synthesis is very low and is not sufficient to provide necessary amount of these LC-PUFA for mental and cardiovascular health (Khozin-Goldberg et al., 2011).

The overall consumption of LC-PUFAs continuously increases both for aquaculture and for human food supplementation. The traditional methods of LC-PUFA enrichment such as extraction from fish and higher plants cannot provide efficient amounts of these fatty acids, which motivate the producers to find alternative sources of LC-PUFA production. One of the most promising candidates for the oil production is the microalga *Nannochloropsis*. It can accumulate high amount of lipids and produce very important EPA in high concentration. What is even better, the biosynthesis of EPA is accompanied by relatively high TAG biosynthesis, especially during stress conditions (Hoffmann et al., 2010). The EPA concentration can reach 35% of total fatty acids which concentration can reach 60-70% of dry weight of the cell (Khozin-Goldberg et al., 2011). It was shown that the DHA and EPA dissolved in TAG is a desirable form of dietary intake and digestion for human organism.

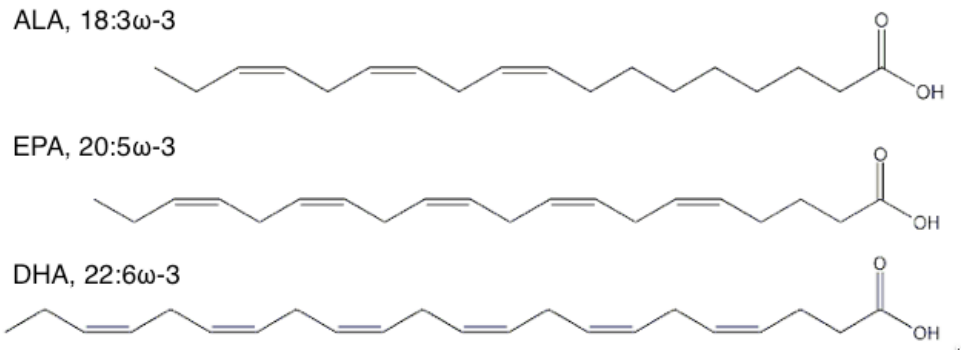


Figure 4. ω-3 Fatty Acids. From the top to bottom: α-linolenic acid (ALA) - the essential precursor of LC-PUFA, eicosapentaenoic (EPA) in the middle and docosahexaenoic (DHA) at the bottom, which are necessary for the normal functionality of human brain and cardiovascular system. Structures from www.LipidMAPS.org

The biosynthesis of fatty acids and particularly that of DHA from EPA involves the formation of polyunsaturated C_{24} intermediates, followed by a β-oxidation steps in the peroxisomes (Sprecher, 2000) Figure 5.

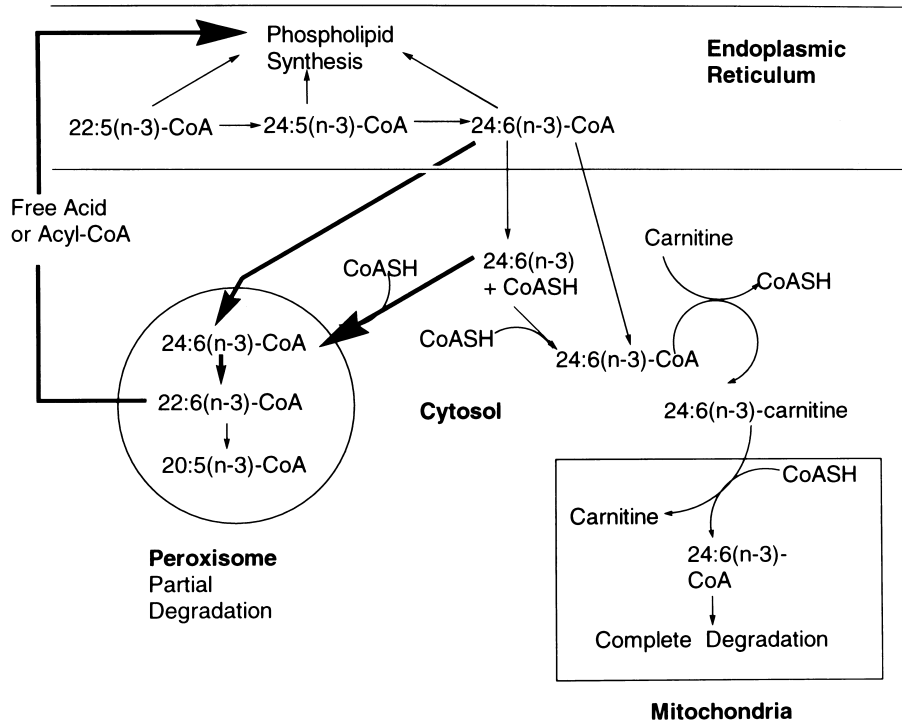


Figure 5. The intracellular movement of PUFA and their metabolism in the cell. The diagram shows that when PUFA are made in the endoplasmic reticulum, they may be used directly as substrates for phospholipid biosynthesis. The substrate and products of PUFA biosynthesis are the acyl-CoA derivatives, which can move directly to another subcellular compartment or be hydrolyzed to free acids by cytosolic acyl-CoA hydrolases. Biosynthesis of the highly unsaturated FA such as DHA from EPA in mammals involves the formation of polyunsaturated C24 intermediates by sequential elongation and desaturation, followed by a β-oxidation step of 24:6ω-3 to 22:6ω-3 in the peroxisomes **(Sprecher, 2000; Khozin-Goldberg et al., 2011).**

The peroxisomes enzymes play an important role in lipid biosynthesis and catabolism in multicellular eukaryotes, such as higher plants and mammals. However much work has to be done aimed at the analysis of the processes involved in peroxisomal part of the lipid metabolism in the most promising oil-producing organisms - microalgae.

1.3.Peroxisomes

Peroxisomes are very small (0.5 - 1 μm), ubiquitous eukaryotic organelles. They play some important roles in lipid metabolism, photorespiration and response to both abiotic

and biotic stresses in plants. The increasingly long list of all known metabolic pathways and functions consist of 14 members: ROS metabolism, fatty acids β-oxidation, photorespiration, purine catabolism, branched AA catabolism, jasmonic acid biosynthesis, auxin biosynthesis (IBA-to-IAA conversion), sulfur metabolism, polyamine catabolism, protein modification, pathogen defense, Co-factor metabolism, methylglyoxal detoxification, phylloquinone biosynthesis and pseudouridine catabolism (Reumann, 2011).

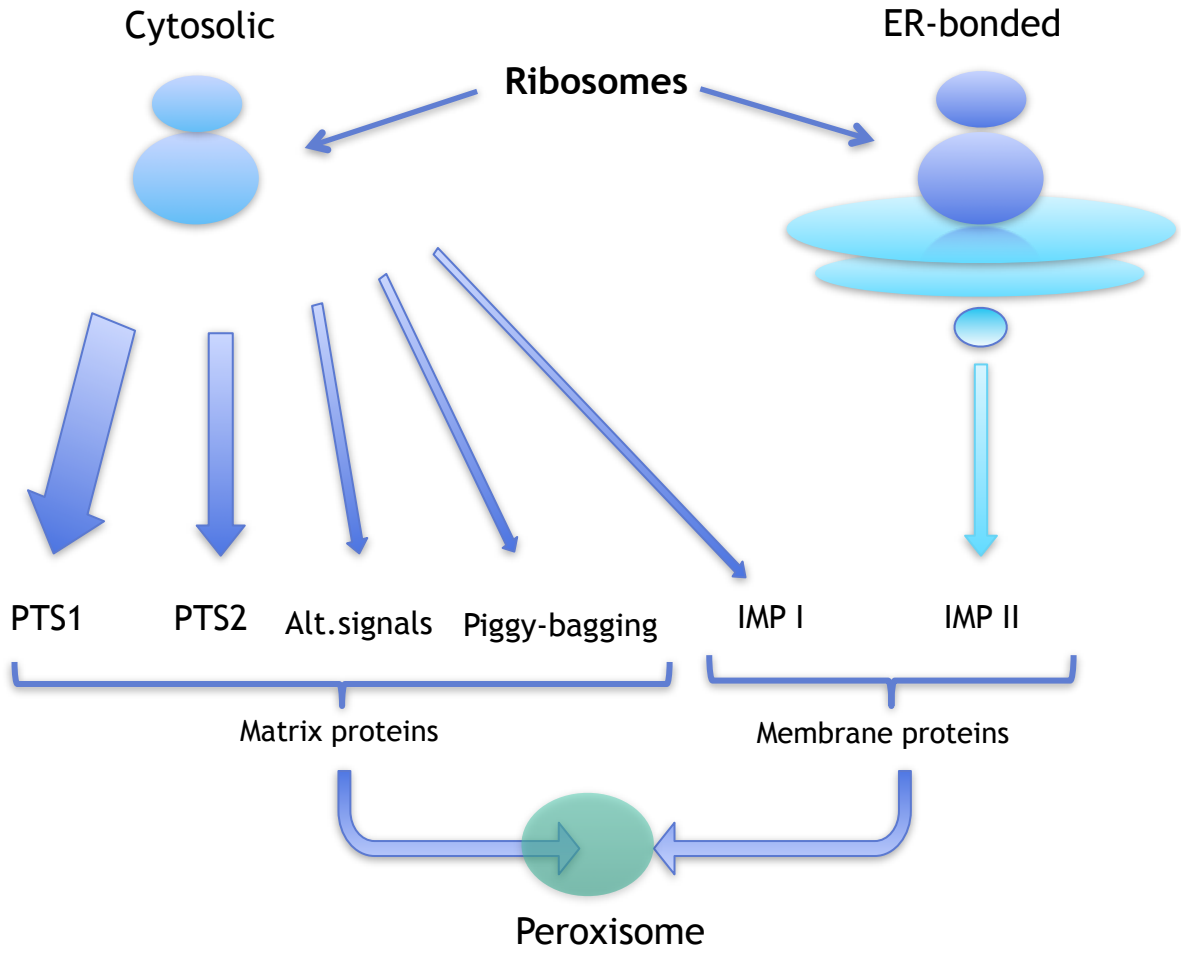


Figure 6. The most abundant way of transporting soluble proteins into peroxisomes is the PTS1 pathway. The peroxisomal targeting signal type 1 (PTS1) of the prototype SKL> is located at extreme C-terminal end of the protein. Generally, a combination of amino acids [S/A/C]-[K/R/H]-[L/M] fit this concept, but by far not completely (Reumann et al., 2007; Lanyon-Hogg et al., 2010). Another known targeting signal is the cleavable PTS2 nonapeptide of the prototype RLx5HL generally located in the 40 amino acid N-terminal domain. The following combination of amino acids was also considered as PTS2: [R/K]-[L/V/I]-x5-[H/Q]-[L/A] (where x denotes any amino acid) (Reumann et al., 2007; Lanyon-Hogg et al., 2010). This amino acid combination (not that of the PTS2, which is in progress, Lingner 2011 deals only with PTS1 prediction) represent a relatively constant composition and can be predicted by bioinformatical prediction algorithms with high accuracy (Lingner et al., 2011).

The most abundant way of transporting soluble proteins into peroxisomes is the PTS1 pathway. The peroxisomal targeting signal type 1 (PTS1) of the prototype SKL> is located

at extreme C-terminal end of the protein. Generally, a combination of amino acids [S/A/C]-[K/R/H]-[L/M] fit this concept, but by far not completely (Reumann et al., 2007; Lanyon-Hogg et al., 2010). Another known targeting signal is the cleavable PTS2 nonapeptide of the prototype RLx_5HL generally located in the 40 amino acid N-terminal domain. The following combination of amino acids was also considered as PTS2: [R/K]-[L/V/I]-x_5-[H/Q]-[L/A] (where x denotes any amino acid) (Reumann et al., 2007; Lanyon-Hogg et al., 2010). These amino acid combinations (PTS1 and PTS1) represent a relatively constant composition and can be predicted by bioinformatical prediction algorithms with high accuracy (Lingner et al., 2011). PTS1 prediction algorithms are already available, while PTS2 prediction alagorithms are in the progress.

In this study the proteins involved in biosynthesis of lipids in *Nannochloropsis* peroxisomes were investigated together with establishing basic growth techniques for microalgal cultivation. The control for cell proliferation and lipid accumulation are only two of them. Some of theoretical findings are presented in further sections.

1.4. Direct and indirect methods to determine microalgal cell concentrations

Microorganisms can be counted by several methods (Black, 2012):

- Turbidity measurements;
- Dry weight measurements;
- Microscopic counts using counting chambers;
- Measurements of fluorescence using a flow cytometer (Hyka et al., 2013).

Media with or without microorganisms differ in their turbidity or optical density, which can be detected by spectrophotometry. This method is very fast and technically simple, but the results can be affected by composition of pigments in cells, interaction of turbid culture with matter (scattering), etc. The method can only be used to indirectly determine cell numbers if a linear correlation between concentration of cells and optical density exists.

Dry weight measurements based on weighting of dehydrated cell biomass and calculation number of cells if the average weight of single dried cell is known. This method is quite time consuming and the results depend on cell size. The major advantage is that it can be used when dried algal biomass is needed for further chemical composition analysis or extraction of compounds.

Counting of cells using a counting chamber and a light microscope is a direct counting method. It means that in this method cells are counted directly, one by one in a defined volume. After making a calibration curve which reflects correlation between cell concentration and, for example, optical density, it is possible to calculate cell concentrations from optical density of a cell culture. The major disadvantage of this method is that it needs relatively long time to count the cells with sufficient accuracy (See Chapter 2.4).

Measuring number of cells using flow cytometry is probably the most accurate cells counting method. In the flow cytometer fluorescence of each single cell in culture can be detected. Using such instrument, the number of the cells can be investigated «in one click». The major disadvantages are very high prices for the flow cytometers and that natural or artificial fluorescent markers have to be present in the cell (Hyka et al., 2013).

1.1.Fluorescence spectroscopy

In order to look at the lipid composition of the cells a method to mark and visualize only the desired compounds is needed. There are some methods which can give an information about qualitative and quantitative composition of the cells. Some of them are based on the destruction of the cell and analysis of their composition: high-performance liquid chromatography (HPLC), thin-layer chromatography (TLC), gas chromatography (GC), etc. However cells die during these analysis, that is why these methods should be used when complete characterization is necessary (Brennan et al., 2012). Localization of compounds can be also inspected by fractionation of cells organelles or membranes and analyzing them separately. However it is time- and material-consuming technique. More advanced methods investigate localization of chemical compounds keeping cells alive.

A method that was used for lipid cell content analysis is fluorescence microscopy. Fluorescence is defined by emission of light by a certain molecule after absorption of light. The emitting light usually has longer wavelength than the exciting light due to energy loss. Fluorescence light microscopy uses different techniques compared to common light transmission microscopy. In the first case, the detected light is emitted by the desired object, which was excited by light having different wavelength, thus invisible for the detector. The specimen forms a very good contrasted image on the dark background (Lichtman et al., 2005). It is very different from common light microscopy, where the image

and its contrast appear as a result of interaction between visible light and the specimen matter.

Molecules which can excite fluorescent light are called fluorophores. Many organic molecules such as chlorophyll or other molecules which have conjugated π-orbitals in the complexes with conjugated double bonds. Nowadays, many fluorescent molecules are specifically designed to be used in fluorescence microscopy. They have different excitation and emission spectra, for instance, Green Fluorescence Protein (GFP) or Yellow Fluorescence Protein (YFP). These proteins can be fused to proteinaceous compounds of interest and emit light exactly from the place where they are located, thereby showing localization of analyzing compound (Shaner et al., 2005).

1.2.Lipids

Analysis of lipids in cells is usually performed by spectrophotometric or chromatographic methods. As was cited before, spectrophotometric methods are more rapid and can be done on living cells. Some specific molecules are used for this purpose.

Two fluorophores are nowadays used in algae lipid fluorescence microscopy - Nile Red and BODIPY 505/515. Both molecules are lipophilic and can penetrate cell structures stucking in lipids. They are not toxic and have linear correlation between concentration of die and total amount of lipids in algal cells (Cooper et al., 2010).

Microalgae typically have thick cell walls. To increase the penetration ability, microalgal cultures are treated with organic solvents: DMSO, glycerol or acetone. However high concentration of such compounds can be lethal for algae cells. BODIPY 505/515 penetrates Nannochloropsis cell walls easily, hence does not need high concentrated organic solvents in the final sample (Brennan et al., 2012).

Absorption and emission spectrums are very important, to precisely measure fluorescence of the dyes only. Nile Red can binds to phospholipids with resulting fluorescence maximum at 590 nm which is close to chlorophyll fluorescence 670 nm. BODIPY 505/515 does not have any fluorescence at 670 nm. These theoretical features demonstrate that BODIPY 505/515 is perfectly suitable for staining neutral lipids in the microalgae.

1.3.Bioinformatics

Informatics in biology helps to store, handle massive biological data (genomes, proteomes) and to visualize results. Bioinformatics cannot only handle biological data, but predict some process or phenomena. In the present study diverse bioinformatic tools were applied mostly to predict protein localization, validate protein orthology and improve gene and cDNA predictions (more detailed in Chapter 2.8)

For PTS1 predictions the PredPlantPTS1 peroxisomal prediction server was used (available at http://ppp.gobics.de/)(Lingner et al., 2011; Reumann et al., 2012). It represent results as position specific weight matrices (PWM) score which illustrate an information how last 14 amino acid at C-terminus end are correlated with know PTS1 peptides. The algorithm analyzes the 14 C-terminal amino acids and gives them a score according to their position and type based on collection of sequences of verified PTS1 peptides. It also considers the possible dependencies between amino acid residues. Such complex analysis of 14 amino acids at C-terminus end allows this prediction server to predict previously unidentified peroxisomal tripeptids. Final PWM score equal 1.000 should be considered as 100% probability to be peroxisomal, equal 0 - 0% probability to be peroxisomal with a threshold 0.412. Therefore, by using this prediction server we can quickly check proposed localization of any peptide.

In this project the actual protein structure from the corresponding nucleotide sequences will be analyzed and the prediction of the exon number will be made (See Chapter 2.8.2.1). For eukaryotic genes it is usual to have noncoding segments of nucleic acid that lie between coding regions and are called intervening sequences or introns. The introns are very important for self-splicing, alternative splicing mechanisms and gene shuffling. Exons are the coding parts of genes eventually expressed, usually by being translated into amino acid sequences (Campbell et al., 2008). If a gene has several exons, after transcription they are spliced together, without introns, forming mRNA that finally passes into the cytoplasm and can be translated (Black, 2012) (Figure 7).

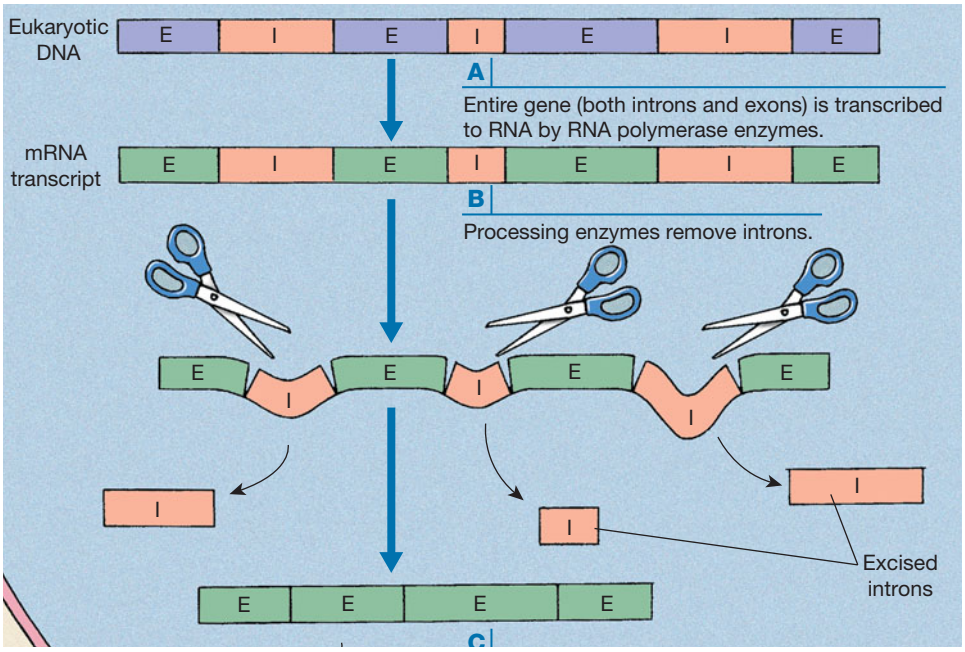


Figure 7. Forming the protein-coding structure of mRNA. Exons (E) and Introns (I), the mRNA which corresponded to the protein structure (C). Both exons and introns are transcribed into RNA, after that the introns are removed, while exons spliced together **(Black, 2012)**.

1.4.Genetic transformation of the algal cells

Transformation is a change in an organism's characteristics because of the transfer of genetic information (Black, 2012). It is a common way to introduce some parts of genome of one organism to another. It is mostly established for bacterial cells such as the «lab rabbit *Escherichia coli*», but can be applied to unicellular algae as well. Transformation is usually done by inserting the vector - a self-replicating gene carrier, for example a plasmid (Black, 2012). It can carry several genes such as antibiotic resistance genes, replicons, promoters, which help in the replication of the vector and effective gene translation. The target cell has to be able to uptake the vector. For this reason cells to be transformed have to be permeabilized, for instance by treatment with the electric power, low temperatures or chemicals which make their cell membranes porous or transparent for the vectors (Hanahan, 1983). Treated cells are called competent cells. The major techniques used in bacterial transformation is a cell bombardment by gold particles coated in vector DNA, a heat shock transformation and a electroporation (Dower et al., 1988).

Some protocols for efficient transformation of microalgae have been published recently: transformation of diatom *P. tricornutum* by microparticle bombardment (Apt et al., 1996) and the papers describing high efficiency transformation of *N. gaditana* (Radakovits et al., 2012) and *N. oceanica* (Vieler et al., 2012) and *N. sp* (Kilian et al., 2011) by

electroporation. Based on the analysis of these scientific papers electroporation method was chosen for transformation of the available Nannochloropsis species.

1.5.Thesis goals

In 2012 the research group by Prof. S. Reumann started a new research project on microalgae. The previous major research interest and expertise of the group was in peroxisome biogenesis and metabolism, applying different methodology including cell biology, protein chemistry including proteomics, plant molecular biology and bioinformatics. The new research project was intended to comprise both basic research and biotechnology and to focus on investigating of the role of peroxisomes in lipid metabolism in the lipid-rich microalgae, *Phaeodactylum* and *Nannochloropsis*, The long-term objective was to optimize their productivity in synthesizing omega-3 fatty acids for Norwegian aquaculture by genetic engineering and technology-assisted breeding. Both biosynthetic and catabolic fatty acid and lipid metabolic pathway have been linked to peroxisomes but any molecular details are hardly reported to date. The departure point of the present research was to establish the fundamental microalgal methodology in the Reumann group and to obtain important preliminary research data on peroxisome biology and fatty acid metabolism in order to establish the research group in this field and potentially qualify for the research funding from the new BIOTEK 2021 programme of the Norwegian Research Council.

The objectives of this M. Sc. study were the following:

1.The establishment of basic culturing techniques and analytical methodology in the Reumann research group:

a)Robust and reproducible culturing conditions and the methodology to monitor cell growth and proliferation in *Nannochloropsis* under standard growth conditions and nutrient starvation,

b)Fluorescence microscopy to monitor lipid accumulation in Nannochloropsis by fluorescent staining and microscopy,

c)Comparative analysis of microalgal growth and lipid productivity among different *Nannochloropsis* species, and

d)Investigation of bacterial, fungal and protozoan contamination analysis in non-axenic algal cultures and possibly partial purification;

2.Prediction of the peroxisomal proteome of soluble matrix proteins for *Nannochloropsis gaditana*:

a)Analysis of approx. 80 predicted PTS1 proteins of *N. gaditana* for PTS1 protein conservation in *Stramenopiles* and *Arabidopsis thaliana* by orthology and PTS1 protein analysis,

b)Manual verification and correction of published gene models of *N. gaditana* for predicted PTS1 proteins of highest interest,

c)Analysis of PTS1 protein conservation in *N. oceanica* taking advantage of publicly available contigs, and

d)Completion of the PTS1/2 proteome of *N. gaditana* by analysis of all known PTS1 and PTS2 proteins in *A. thaliana* for putative orthologs in *N. gaditana* and *N. oceanica* combined with PTS1 and PTS2 prediction analysis, and

e)Identification of approx. 10 PTS1 proteins in *N. gaditana* for experimental analyses. These proteins should be (i) predicted to be peroxisomal with high probability, (ii) be species- or genus-specific or conserved in higher plants and (iii) be known to be located in *A. thaliana* peroxisomes and orthologous to proteins reported to be involved in metabolism of unsaturated fatty acids or be of unknown intriguing and important predicted function.

3.Cloning of full-length cDNAs or C-terminal exons of predicted PTS1 proteins in *N. gaditana* for subsequent subcellular localization and functional studies:

a)Design of suitable pairs of cloning primers,

b)Isolation of genomic DNA and cloning of selected full-length cDNAs (of single exon genes) and C-terminal exons (of multiple exon genes) from genomic DNA into the transient storage vector pJET (Fermentas),

c)Sequence analysis of cloned DNA fragments

d)If time permits, subcloning of selected full-length cDNAs and C-terminal exons into the plant expression vector pCAT in the back of the reporter protein EYFP for experimental validation of predicted peroxisome targeting in a plant expression system.

In addition, the establishment of nuclear transformation of *N. oceanica* by pSELECT100 plasmid should been done.

2.Materials and methods

2.1.Organisms

Five Nannochloropsis species were used as axenic cultures: *N. gaditana* CCMP 526, *N. oculata* CCMP 525, *N. oceanica* CCMP 1779, N. salina CCMP 537, *N. granulata* CCMP 529. All these strains were purchased/obtained from the National Center for Marine Algae and Microbiota (NCMA, Bigelow, USA, www.ncma.bigelow.org/).

For contamination analysis non-axenic cultures of *N. gaditana* CCAP 849/5 were used . They were obtained from the Culture Collection of Algae and Protozoa (CCAP, Scotland, United Kingdom); *N. salina* NIVA-2/01, *N. oceanica* NIVA-2/03, *N. oculata* NIVA-3/04, *P. tricornutum* NIVA-BAC 2 was obtained from the Norwegian Institute for Water Research (NIVA), Oslo, Norway, (www.niva.no/).

2.2.Vectors

2.2.1.pCAT

The pCAT is a plant expression vector. Which is extensively used for subcellular localization experiments in the Reumann lab (Lingner et al., 2011). That one which was used in this study had inserted EYFP sequence without stop-codon inserted after the CamV 35S promoter. The known *A. thaliana* PTS1 protein DECR was located after the EYFP-coding gene. Such construct pCAT-EYFP-DECR express EYFP-DECR fused protein which will be targeted to the peroxisomes. In this study several putative peroxisomal proteins of *N. gaditana* will be inserted in to the vector in the place of DECR.

2.2.2.Blunt cloning vector pJET 1.2

The pJET1.2/blunt cloning vector was developed to be used in the CloneJET PCR Cloning Kit, Thermo Scientific, USA (www.thermoscientificbio.com/molecular-cloning/clonejet-pcr-cloning-kit/) (Figure 8).

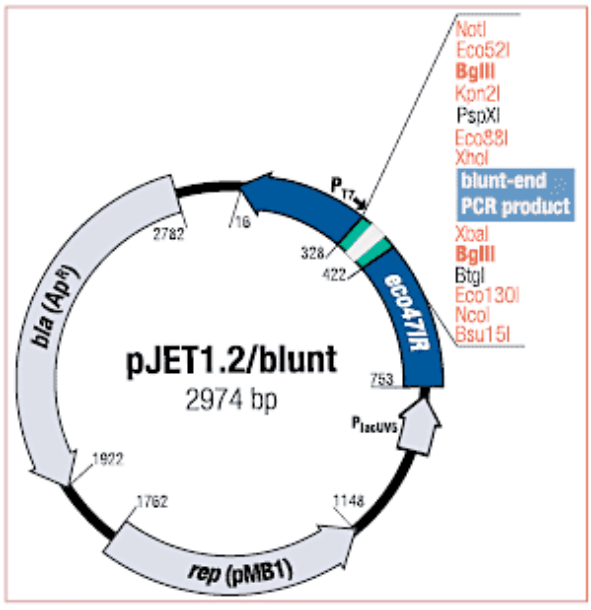


Figure 8. pJET1.2/blunt cloning vector. bla - (ApR)-ampicillin resistance gene, Plac-promoter, eco47IR-lethal gene with blunt ended multiple cloning site, rep-replicon from the pMBIp plasmid (www.thermoscientificbio.com/molecular-cloning/clonejet-pcr-cloning-kit/)

It is 2974 base pairs long and have the following elements: β-lactamase gene conferring resistance to ampicillin - bla (ApR), lethal gene eco47IR and its promoter Plac which are used for positive selection of recombinant plasmid, multiple cloning site with blunt DNA ends for ligation with insert and T7 RNA polymerase promoter for transcription of the cloned insert. Replication of insert occurs with the help of replicon (rep) from the pMBI plasmid.

If DNA construct has TA blunt terminal ends, it can be inserted into the pJET vector. Proof-reading DNA polymerases are well suitable for making blunt ends during PCR amplification of the construct. If the vector has linearized by itself without insert, the transformed cells die because the delineralization activates the lethal gene eco47IR.

2.2.3.Selection vector pSELECT100

pSelect100 is a vector designed for transformation of N. oceanica (Vieler et al., 2012). The vector was made based on the construct used for transformation of *C. reinhardtii*, pHyg3. It has two antibiotic resistance genes (Ampicillin and Hygromycin B), the original promoter of pHyg3 was replaced by LDSP promoter from *N. oceanica* CCMP1779 genomic DNA. In this project the vector was used to transform the *N. oceanica and N. gaditana* cells (See Chapter 2.9).

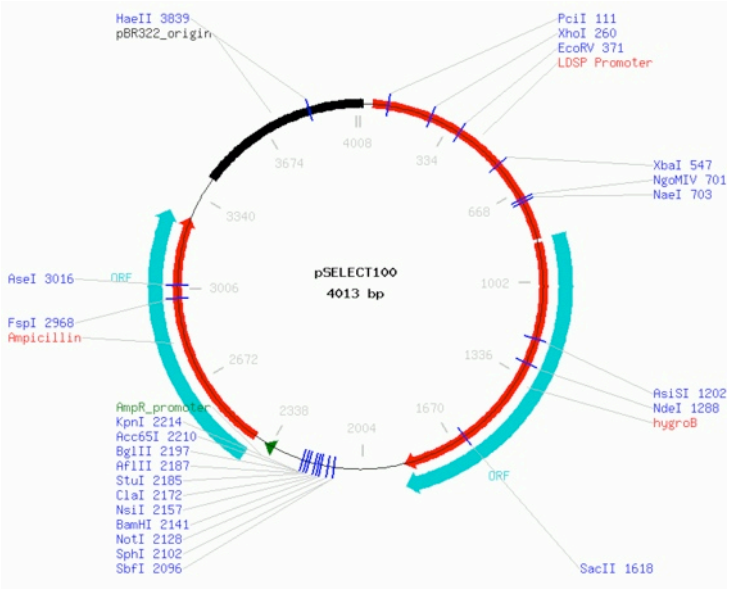


Figure 9. Gene map of selective vector pSELECT100 used for N. oceanica CCMP1779 genomic transformation (Vieler et al., 2012). The vector is 4013 base pairs long and contain native promoter LDSP, Amp and HygB resistance genes.

2.3.Basic growth conditions

The algae were grown in 250 ml Erlenmeyer flasks, using 100 ml of liquid culture in each. The flasks were closed by cotton plugs and covered by aluminum foil. Such construction helped to keep algae cultures sterile, but allowed air to go inside. Flasks with added medium, closed by cotton plugs and covered by aluminum foil were autoclaved and were stored in shelf board. Algae cultures were added into flasks inside of sterile box, to avoid contamination.

2.3.1.Medium

During study process growth conditions were changed according to optimization of growth rate, economy thoughts or other reasons. If not stated otherwise the standard growth medium was F/2 medium (Guillard et al., 1962; Guillard, 1975). The sea water stemmed from the North Sea and was collected near Hundvåg/Stavanger (Norway). The water had salinity 25,5 ‰, measured by portable salinity meter VWR CO310. Sea water was filter sterilized by 0.22 μm pore size filter using a vacuum pump to accelerate the process. The F/2 medium was prepared from a commercial 50x stock solution dissolved in sea water, or assembled from stock solutions.

According to a published protocol from the Culture Collection of Algae and Protozoa (CCAP) (http://www.ccap.ac.uk/media/recipes/F2.htm) with minor changes four stock solutions were made:

1. Stock solution «Trace elements (chelated), x1000», (1L): gram per liter

Na_2EDTA	4.16
$FeCl_3 \cdot 6H_2O$	3.15
$CuSO_4 \cdot 5H_2O$	0.01
$ZnSO_4 \cdot 7H_2O$	0.022
$CoCl_2 \cdot 6H_2O$	0.01
$MnCl_2 \cdot 4H_2O$	0.18
$Na_2MoO_4 \cdot 2H_2O$	0.006

2. Stock solution «Vitamin mix, x1000», (1L):

Cyanocobalamin (Vitamin B_{12})	0.0005
Thiamine HCl (Vitamin B_1)	0.1
Biotin (Vitamin B_7)	0.005

3. Stock solution «Nitrogen source, x1000», (1L):

$NaNO_3$	75

4. Stock solution «Phosphate source x1000)», (1L):

$NaH_2PO_4 \cdot 2H_2O$	5.65

To make 1 liter of F/2 medium, 1 ml of each solution were mixed and supplemented with natural sea water up to 1 liter. The pH was adjusted to 8.0 with 1 M NaOH or HCl. In order to make F/2 medium with silicate component, for example for diatoms, a stock solution of xx mM sodium metasilicate $Na_2SiO_3 \cdot 9H_2O$ (30 g should be dissolved in 1 liter) and 1 mL was added along with the others stock solutions during media preparation (from http://www.ccap.ac.uk/media/recipes/F2_Si.htm).

Ready medium was distributed over 10 Erlenmeyer flasks, using 100 ml in each. Sterilization was made by autoclaving. Sterile, closed flasks were stored in the lab shelfs together with other medium at the room temperature. Some sedimentation with slightly brown or white pellet was detected. Such sedimentation did not influence algae growth.

2.3.2.Temperature and light conditions

Two facilities were available at the Center of Organelle Research (CORE) at University of Stavanger in the period when this work was being performed.

On the one hand, a plant room with constant temperature of 23°C, day/night light cycle 18/6 hours respectively. Algae were grown in 250 ml Erlenmeyer flasks standing on an orbital shaker (1.2 speed setting) in a way that the light from the luminescent bulbs reached the cultures from the top.

Second facility, was introduced later, is plant growth chamber (Panasonic MLR-351). This chamber has precise programmable temperature and light control independently from surrounding conditions. The same 250 ml flasks on the shaker were transferred inside. Temperature were adjusted to 23°C, illuminance was about 13000 lux during 24 hours per day.

2.4.Algal growth analysis by cells counting

Microorganisms can be counted by several methods (Black, 2012), including direct (microscopic cell counts using counting chambers, flow cytometry) and indirect (turbidity and dry weight measurements).

Taking into account the available equipment, cell counting was done using counting chambers as a direct counting method and measuring optical density of cultures using the spectrophotometer as an indirect method.

2.4.1.Counting equipment

Cell counting was performed using a Bürker counting chamber. It has a design typical for counting chambers.

Counting chambers differ in marking of counting plates. Bürker variant has one large square, which is sometimes called counting grid (side length 3 mm) and several smaller squares represented at Figure 10. This counting grid has 9 squares, each side is surrounded by triple lines which have a length of 1 mm from the middle line of one triple set to the next middle line, area 1 mm^2. Each 1 mm^2 large square is subdivided into 16 small squares of identical size which are separated from each other by double lines, except those which share triple lines with large squares. Each side of small squares has a side length of 0.2 mm (area 0.04 mm^2).

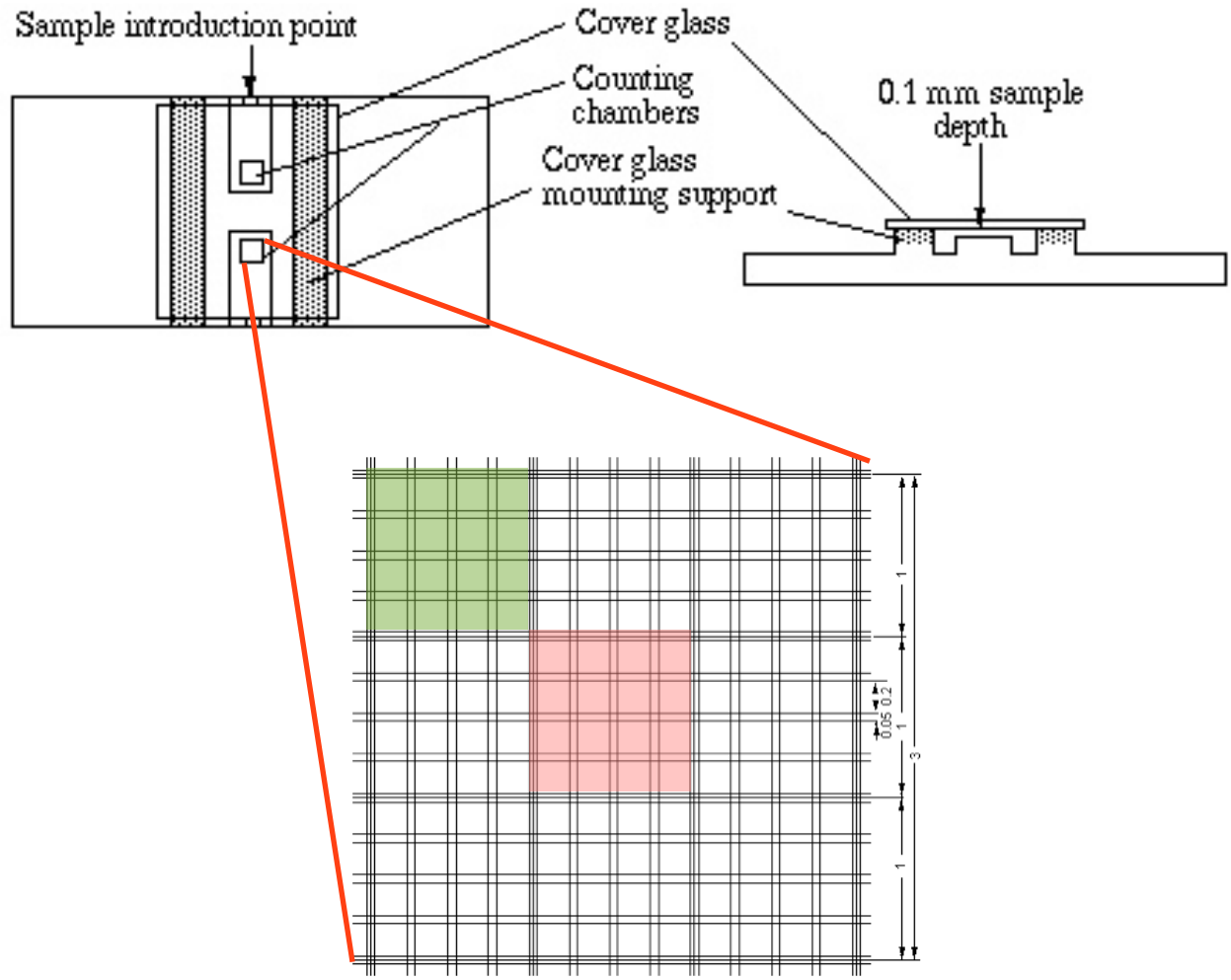


Figure 10. Construction of the Bürker counting chamber and counting grid. A depth of the counting chambers is 0.1 mm. The whole counting chamber is rounded by a red circle and a large 1 mm2 square is marked by green and red colors. Based on images of Experimental Biosciences Resources (www.ruf.rice.edu/~bioslabs/methods/microscopy/cellcounting.html) and Marienfeld Counting Chamber brochure (www.marienfeld-superior.com/index.php/counting-chambers.html).

2.4.2.Counting procedure

The method of determining cell concentration by cell counting is very simple and does not require expensive instruments or special working conditions. Though in order to get reliable results some rules need to be followed.

The total sample volume of a Bürker counting chamber is about 14 μL. This volume should be distributed equally between the two counting chambers with grids.

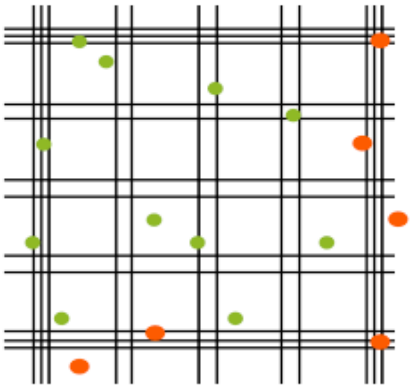


Figure 11. Regions in the large 1 mm^2 square. Red and green dots represent cells. Those cells which touch one of the right and bottom triple lines must not be counted and are marked by red color. Those ones which touch one of the left and upper triple lines have to be counted and are marked by green color.

All cells inside of large 1 mm^2 square have to be counted in such a way that none of them is missed or counted twice. Those cells which are located on one of the triple lines of the left and upper sides of large squares have to be counted, while those ones located on the right and down sides have to be skipped from counting (Figure 11). Thereby, exactly the number of cells in the 1 mm x 1 mm square is determined.

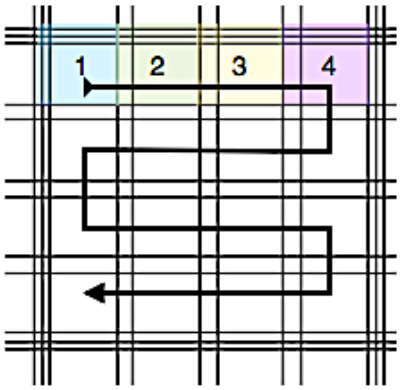


Figure 12. The proper way of counting inside of large square. Counting starts from upper left small square 1 and finishes at upper right square 4.

It is recommended to start counting from the left upper small square and then to move towards right upper corner, to write down number of counted cells, and to start the next series from the square which is located below the previous one (Figure 12). The direction has to be changed towards the left side of the large square. All cells that are located on the upper and left borders (double lines) of small squares should be added to the number of cells within double lines. Using this method one can avoid counting cells twice.

For cell counting two different samples were prepared, each sample was loaded into the counting chamber twice. Two large 1 mm^2 squares were counted for each loading. One from the top counting chamber and another one from the bottom counting chamber. To get the distribution of cells within chambers which is the closest to the real one, the location of counted squares was chosen in such a way, that the first counted square was at the top of the left corner (green region at Figure 10) of top counting chamber and the second counted square was in the middle position of the bottom counting chamber (red region at the Figure 10).

Thus, for each Bürker counting chamber two algal cell counts within the large 1 mm^2 squares were made. Since two algal aliquots of two Bürker cell loading replicates were made, eighth cell counts were done in total for each culture.

One large 1 mm^2 square has the following volume:

$$1\ mm * 1\ mm * 0.1\ mm = 0.1\ mm^3 = 0.1\ \mu l = 0.0001\ ml$$

Hence, the concentration of cells in one large square is:

Number of cells in one large square / 0.0001 ml, or №cells per ml * 10000

If the sample was diluted the cell concentration has to be multiplied by the dilution factor:

№cells in the large 1 mm^2 square * 10000 * dilution factor

Total cell concentration per mL calculates as the average of cell concentrations of all counted large squares for the culture.

2.5.Algal growth analysis by spectroscopic measurements

Direct microscopic cells counting take much time. However, if a correlation between cell number in a culture and optical properties of the same culture is known, it is much easier to measure optical density and calculate cell concentration from that.

To measure optical density every wavelength of visible light can be selected. Possible scattering of solvent and variation in pigments of algae should be considered. Three different wavelengths stated in the literature: 540 nm (Rocha et al., 2003), 625 nm (Hsueh et al., 2007), 680 nm (Gu et al., 2012) were chosen for measuring optical density of *Nannochloropsis* culture.

A generally accepted wavelength for measurements of *Nannochloropsis* optical density could not be found in the literature, so it was decided to get values for all of these wavelengths. Another reason for using only three different wavelengths was that the spectrophotometer which was available could measure exactly this amount of wavelengths as a maximum. To find a relation between cell concentration and optical density of culture measurements of these variables, the measurements were synchronized in time, i.e. they were performed within the same day. First the spectrophotometer SmartSpec Plus (BioRad, USA) had to be turned on, the optical density mode (button «OD») had to be chosen. Then, a number of wavelengths which would be measured (1-3) should have been typed on the keyboard.

To calibrate the instrument, 1 ml PS cuvette was filled with 1 ml F/2 media without algae, and placed into the cuvette holder. The lid was closed the button «Read blank» was pressed to start calibration, then the button «Continue» was pressed in order to jump to sample OD measurements.

1 ml of algae culture was transferred into each of three 1.5 Eppendorf tubes in the sterile conditions. The sample from the Eppendorf tube was transferred into the 1ml PS cuvette, and mixed by pipetting. The cuvette was placed into the cuvette holder, the lid was closed and button «Read sample» was pressed. The given values were notated. If the OD values were higher than 0.6, the sample was diluted with F/2 media. The procedure was repeated for two other samples.

The average OD value was calculated for each wavelength. For those samples which were diluted, the OD values were multiplied by the dilution factor and resulting value was

used to calculate the average. For each wavelength standard deviation was also calculated.

If it is necessary to make a calibration curve, then a new table should be created as shown in Table A in Appendix 1.

In a first row put results of cells counting, in the other put the results of OD measurements for the same samples (the time points can be added as well). A plot of such values as represented should be made as shown in a Figure 20.

2.6.Contamination analysis

A contamination can be removed by several methods: dilution, centrifugation, using antibiotics or detergents, plating, etc. Several methods were chosen and applied for *Nannochloropsis* algal cultures.

All procedures were done in the sterile conditions and using sterile instruments. For results see Chapter 3.1.3, Figure 21.

2.6.1.Centrifugation

Another method separates algae cells from bacterial ones using a combination of centrifugation and washing steps (Stein, 1980). This method cannot give an axenic culture, but can be used for the purification of the algal cultures.

The method was used for the purification of the *Nannochloropsis* cultures during harvesting and is described in Chapter 2.8.

2.6.2.Dilution method

1 ml of algal culture at mid-Log phase was transferred to the 1.5 Eppendorf tube, 2 μl of the culture were transferred into the 200 μl PCR tube, then 20 μl of sterile F/2 media were added to the tube and mixed by pipetting. The 2 μl of mixture were transferred into the new 200 μl PCR tube and the previous step was repeated. In total 4 dilutions were made, the sample was diluted in 10000 times. 10 μl of the final solution were put on the F/2 agar plate and distributed by using disposable loop or spatula.

2.6.3.Antibiotics

Treating a sample with antibiotics is used to slow down the growth rate of contaminators and/or kill them completely. Based on the literature (Jones et al., 1973; Bradley et al., 1988) the method of using antibiotic disks was chosen for this study. The following antibiotic disks were obtained from the MAST Group, UK: Amphotericin B 20 μg per disk, Imipenem 10 μg per disk, Neomycin 10 μg per disk, Nystatin 100 units per disk, Penicillin G 10 units per disk.

1. Using of the antibiotic disks on plated algal cultures. The 10 μl of algal culture were transferred to the F/2 agar plate and spread using spatula. Then, antibiotic disks were transferred to the agar surface and distributed equally. The colonies were analyzed after 2-3 weeks by microscopy.
2. 10 μl of liquid algal culture of each specie were placed into 2.0 ml Eppendorf tube together with 1 or two antibiotic disks, the solution was mixed and placed in the plant growing chamber for two weeks. After two weeks the solution was analyzed on the presence of contamination by microscopy.
3. The 2 μl of liquid culture from each of 2.0 ml Eppendorf tubes used in previous experiment were transferred into the first left hand well of the 96 well-plate. 200 μl of F/2 medium were added to the first well and mixed by pipetting. After that the 2 μl of the solution were transferred to the next well, 200 μl of F/2 media were added. The dilution was repeated eight times. The cultures were left for growing for 3 weeks and then analyzed by microscopy.

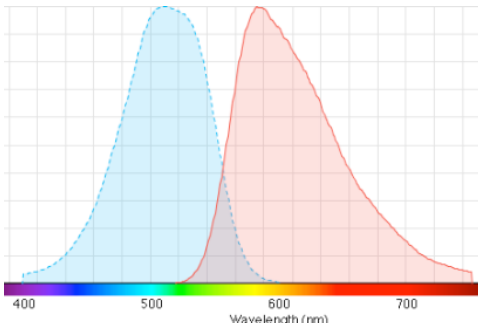
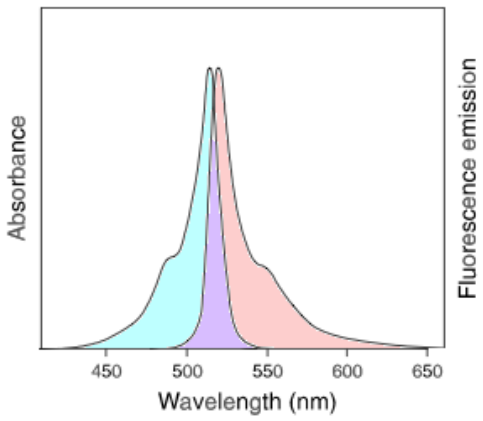
2.7.Neutral lipid staining by fluorescent dyes

Analysis of lipids in cells was performed by spectrophotometric method (Sea Chapter 3.2, Figure 22). The analysis of literature revealed that BODIPY 505/515 is better suitable for staining neutral lipids in the microalgae.

The fluorescence was analyzed using NIKON TE-2000U inverted fluorescence microscope equipped with Exfo X-cite 120 illumination system and followed filters for YFP (exciter HQ500/20, emitter S535/30), CFP (exciter S436/10, emitter S470/30) and for dSRed (exciter HQ560/40, emitter D630/60). The images were taking with Hamamatsu Orca ER 1394 cooled CCD camera connected with a PC.

For BODIPY 505/515 stained cells the settings used for YFP detection were chosen. For detection of autofluorescence of Chl a molecules, the RFP filters were used.

Table 2. Characteristics of major staining fluorescence dies for neutral lipids applying to algae. The table based on Invitrogen product information for the Nile Red (http://products.invitrogen.com/ivgn/product/N1142) and for the BODIPY 505/515 (http://products.invitrogen.com/ivgn/product/D3921).

	Nile Red	BODIPY 505/515
Molecular weight	318.369 g/mol	248.08 g/mol
Penetration of algal cell membranes	Good penetration ability when treated with relatively high temperatures or concentrations of DMSO, glycerol, acetone.	Easily at non toxic concentrations of glycerol or DMSO
Influence on living organism	Does not affect on organism, but high concentrations of treatment agents can be lethal for algal cells	
Excitation wave length	552 ± 3 nm	502 ± 3 nm
Emission wave length	Neutral lipids - Yellow-gold 636 ± 4 nm Phospholipids - Red about 590 nm	Neutral lipids - Green max @ 510 ± 4 nm
Chlorophyll interference @ 670 nm	Yes, for phospholipids	No
Absorption/ Emission Spectra		

2.7.1.Neutral lipid staining with BODIPY 505/515

(4,4-difluoro-1,3,5,7-tetramethyl-4-bora-3a,4a-diaza-s-indacene)

Stock and working solutions were made according to procedure which was first described by (Cooper et al., 2010) and improved by (Brennan et al., 2012).

First step: making 5 mM BODIPY in 100% DMSO solution.

1 ml of DMSO was added to the bottle containing 10 mg of BODIPY and mixed until all the crystals were dissolved. Then, the mixture was transferred into 10 ml measuring cylinder and 7 ml DMSO (100%) were added (to get 8 ml of DMSO in total). The cylinder was covered with parafilm and properly mixed. The solution was aliquoted in 15 1.5 ml Eppendorf tubes with 0.5 ml in each. Second step: making 0.5 mM BODIPY in 100% DMSO.

0.5 ml 5 mM BODIPY in 100 % DMSO was added to 4.5 ml 100% DMSO, the resulting solution was mixed and aliquoted in 9 Eppendorf tubes with 0.5 ml in each.

Third step: making 24 μM BODIPY in 100% DMSO.

48 μL of 0.5 mM BODIPY in 100% DMSO were added to 950 μL 100% DMSO in the new 1.5 ml Eppendorf tube, the mixture was aliquoted in 5 Eppendorf tubes with 0.2 ml in each.

In order to get **working solution** with a final concentration of 0.12 μg/mL BODIPY and 2% DMSO, 2 μL of 24 μM BODIPY in 100% DMSO were dissolved in 98 ml of algae culture. Microscopic observations can be started after 2 minutes of dye adding.

2.7.2.Nile Red staining dye for triacylglycerols

Composition and concentration of final solution for Nannochloropsis lipid staining were taken from (Doan et al., 2010). Because an acetone was used as a solvent for Nile Red, some precautions were made: glows were used, each new pipet tip was first dipped in acetone and afterwards used for the pipetting.

First, a **stock solution** of 2.5 mg/ml Nile Red in acetone was prepared.

1 ml of acetone was added to 25 mg of Nile Red, mixed until all the crystals were dissolved, the solution was transferred into 10 ml measuring cylinder. The previous step was repeated to transfer Nile Red completely. Then, 100% acetone was added to get 2.5 mg/mL solution (10 ml acetone in total), the cylinder was covered with parafilm and properly mixed. The solution was aliquotation in 10 Eppendorf tubes with 1 ml in each, 1 of

those tubes was aliquoted into 10 tubes - 100 μL in each, resulting in concentration 2.5 mg/mL Nile Red in 100% acetone in each tube.

Second step: 0.03 mg/mL Nile Red in 100% acetone (**Working solution**)

900 μL of 100% acetone were added into each tube from the previous step to get the concentration of 0.25 mg/mL Nile Red in 100% acetone. Then, 1 Eppendorf tube with 0.25 mg/ml Nile Red in 100% acetone was aliquoted into 10 tubes - 100 μL in each. Additional 733 μL of 100% acetone was added into each tube to get 0.833 μL of 0.03 mg/mL Nile Red in 100% acetone.

In total the following solutions were made: 9 tubes with 1 mL of 2.5 mg/mL concentration, 9 tubes with 1 mL of 0.25 mg/mL and 10 tubes with 0.833 μL of 0.03 mg/mL Nile Red in 100% acetone.

Working solution of Nile Red can be used directly for sample preparation. 0.5 mg/mL stock solution of glycerol is also needed for better penetration of the Nile Red through cell wall of Nannochloropsis (Doan et al., 2010).

Making stock solution 0.5 mg/mL glycerol in water:

5 g of glycerol were added into 10 ml measuring cylinder, the distilled water was added up to 10 ml. The solution was mixed (the cylinder was preliminary covered with parafilm). The mixture was distributed into 10 Eppendorf tubes - 1 ml 0.5 mg/ml glycerol in each.

According to the article (Doan et al., 2010) final concentration of Nile Red: 0.3 μg/ml, glycerol: 0.1 g/mL. To reach these concentrations 0.8 mL of algae culture were taken, 0.2 ml aqueous glycerol (with concentration - 0.5 g/mL) was added together with 10 μl Nile Red (concentration - 0.03 mg/mL in 100% acetone). Then, the solution was vortexed for about 1 minute and stored for minimum 4 minutes in the dark before it could be used.

2.8.Bioinformatics methods

Two species of microalgae were in the center of interest: *Nannochloropsis gaditana* strain CCMP527 and *Nannochloropsis oceanica* strain CCMP1779 . These species were chosen because their genomes were sequenced and published by Radakovits et al. (2012) and by Vieler et al. (2012) for *N. gaditana* and *N. oceanica* respectively. They had become available prior to the start of this thesis at GenBank of the National Center for

Biotechnology Information (NCBI) http://www.ncbi.nlm.nih.gov/genome/11691 and http://www.ncbi.nlm.nih.gov/genome/13215 for *N. gaditana* and *N. oceanica* respectively.

Gene and protein predictions are available for *N. gaditana*. They were retrieved and processed by Dr. T. Lingner (Institute of Microbiology and Genetics, University of Göttingen, Germany). Dr. T. Lingner created several lists with proteins («Plant», «Metazoa», «Combined») of *N. gaditana* which were predicted to be peroxisomal (For the full list of predicted proteins see Table B in Appendix 2). The proteomic data was taken from the Nannochloropsis Genome Project Solutions (NGPS) group which sequenced the *N. gaditana* genome (available at http://nannochloropsis.genomeprojectsolutions-databases.com/).

2.8.1.Orthology analysis

78 proteins with peroxisomal posterior probabilities above 0.500 (which corresponds to 0.412 of PWM score) were predicted using «Plant» prediction model. The underlying strategy was based on the idea that phylogenetically related organisms often contain orthologous genes, which are derived from a single ancestor gene as a result of speciation (Trachana et al., 2011). They usually encode proteins with similar function, structure and subcellular localization. Orthologous genes from related species have minor differences in nucleotide structure because of speciation, but mostly have similar or even identical parts of sequences which code major motifs or regions (high sequence similarity at the amino acid level, conserved folding, etc.). These parts determine the function ability of gene and have been usually inherited from a common ancestor. Therefore, genes which have similar structure elements might be homologues (orthologs) and encode proteins with almost identical functions and probably localization.

This was done by using BLAST search at http://blast.ncbi.nlm.nih.gov/Blast.cgi. Creators describe (Wheeler D, 2007) this powerful instrument as «The Basic Local Alignment Search Tool (BLAST) which finds regions of local similarity between sequences. The program compares nucleotide or protein sequences to sequence databases and calculates the statistical significance of matches». This tool allows scientists to compare nucleotide or amino acid sequence of interest to all sequences that have previously been deposited in GenBank, the world's largest sequence database. It is also possible to get information concerning function or allocation for protein. Restriction to the search parameters can be easily applied, for example, a search for similarities in genetic code can be applied only among defined group of organisms.

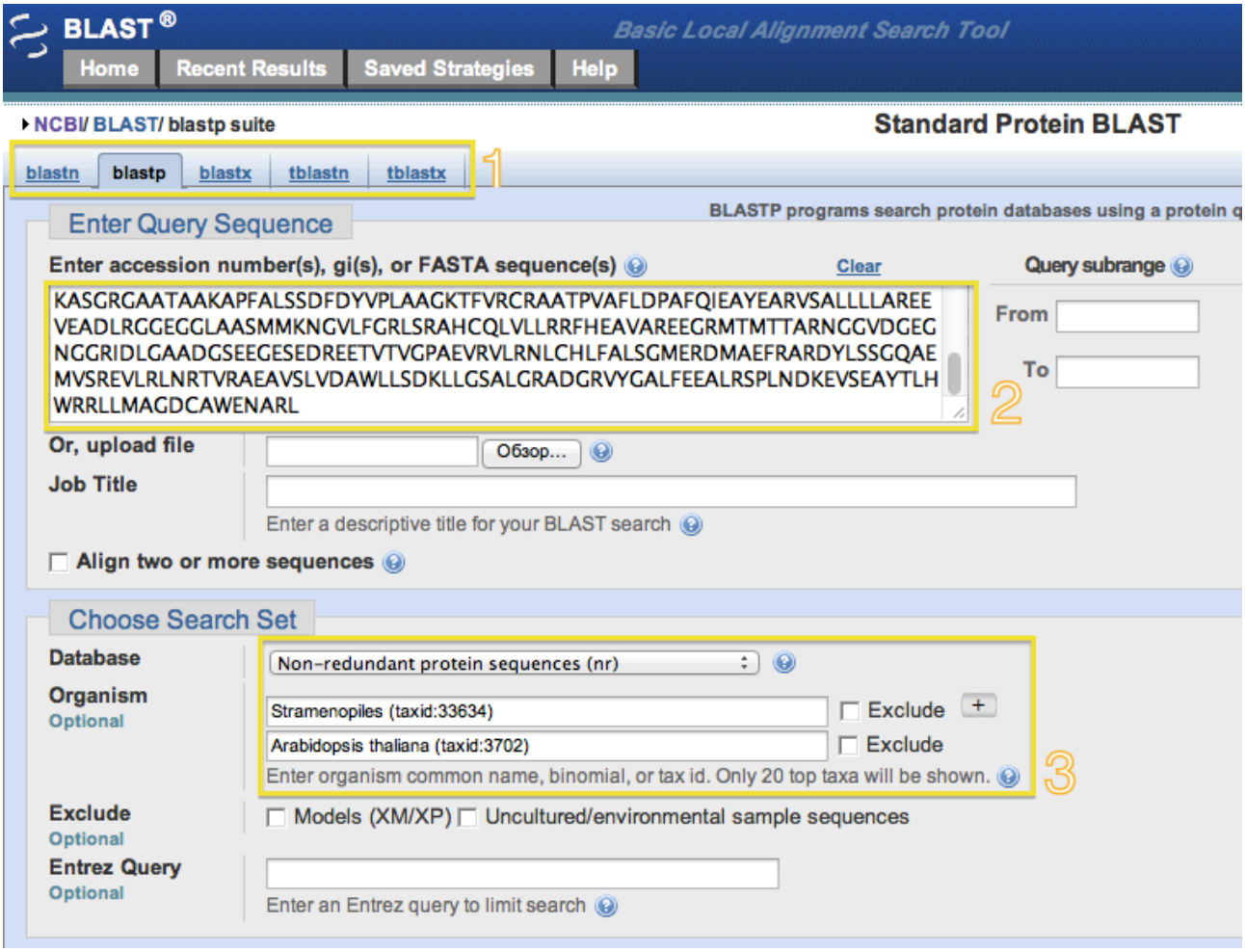


Figure 13. Print screen of protein blast search used orthology analysis with major elements. 1 - blast method, here blastp look for similarities in protein database using a protein query, 2 - Submission window for query sequence, suitable for pasting amino acid sequences or accession codes, here protein Nga03053, 3 - Chose of database, here «Non-redundant protein sequences (nr)» and restriction for search among submitted organisms, here *Stramenopiles* and *A. thaliana*.

The protein BLAST search was done for every amino acid sequence represented in the table provided by Dr. T. Lingner. Results were restricted to *Stramenopiles* and *Arabidopsis thaliana* Figure 4. *Stramenopiles* is a group of organisms having the closest organism similarity and origin with *Nannochloropsis* among algae. The *Arabidopsis* was chosen because of the most well-known plant organism (See section 1.1.1, Figure 2). Usually 4-5 proteins of high sequence identity (at least 30% at the amino acid level, E value from 1e-50 (the lowest similarity) to 1e-150 (the highest similarity)) from different species were picked up for comparison, such as, brown algae *Ectocarpus siliculosus* (Cock et al., 2010), three diatoms *Phaeodactylum tricornutum* (genome was published by Bowler, et al., in 2008) *Thalassiosira pseudonana* (Armbrust et al., 2004), *Thalassiosira oceanica* (Lommer et al., 2012) and well-known model plant *Arabidopsis thaliana* which genome was completed in 2000 by the *Arabidopsis* Genome Initiative. Every protein or gene

represented in the database has its own annotation page, where all known information about this sequence can be found, for example, length, location in genome (for genes), localization in the cell (for proteins), etc. This information was uploaded by the authorized users of the NCBI database, reviewed and revised periodically.

PredPlantPTS1 peroxisomal prediction server is available at http://ppp.gobics.de/ (Lingner et al., 2011; Reumann et al., 2012). It gives results as position specific weight matrices (PWM) score which gives an information on how last 14 amino acid at C-terminus end are correlated with know PTS1 peptides. The algorithm analyzes the 14 C-terminal amino acids and gives them a score according to their position and type based on the collection of sequences of verified PTS1 peptides. It also considers the possible decencies between amino acid residues. Such complex analysis of 14 amino acids at C-terminus end allows this prediction service to predict previously unidentified peroxisomal tripeptids. Final score of above 0.500 should be considered as probably peroxisomal, below - probably non-peroxisomal. The higher is the score, the better is the probability. Therefore using this prediction server it proposed localization of any peptide can be quickly checked.

There were three types of conclusions that could be made: high, moderate and low probability to be true peroxisomal protein (for the *N. gaditana* protein of interest). High probability was given to peptides which orthologs were mostly peroxisomal, moderate - to peptides which have some PTS1 proteins among orthologs, low probability was given to proteins which were very short, had annotations pointing to other compartment localization or which orthologs were not predicted/known as peroxisomal.

Even if not all the proteins were predicted to by peroxisomal based on orthology analysis, some of those which got a bad score can still be targeted to the peroxisomes. That is why all of them were further processed in the next step.

2.8.2.Orthology analysis verification by gene structure analysis

Genomes are represented by large collections of nucleotide sequences. Using bioinformatic algorithms it is possible to find among them the parts which encode genes. Programs analyze nucleotide sequence trying to recognize combination of nucleotides which encode transcription start signal (ATG), stop signal (TGA, TAA, TAG), borders of introns and splicing sites. These prediction algorithms (Augustus: http://bioinf.uni-greifswald.de/augustus/, GeneMark: http://exon.gatech.edu/) were tried to predict gene structures on another organisms - ascomycete - *Neurospora crassa* and oomycete -

Pythium ultimum. After the trial they were applied to the *N. gaditana* genome. The trials on the fungal model organisms can be the reason for the mistakes in gene predictions (Radakovits et al., 2012).

The verification of the protein structure predictions includes identification of the original nucleotide regions which were used for prediction and their actual borders. Investigation of actual structure of the proteins includes investigation of the number of exons in the gene. Therefore these two methods will be described together.

2.8.2.1.Investigation exons and introns in *N. gaditana* genes together with verifying the presence of PTS1

Most of the eukaryotic genes have exons and introns. The gene structure analysis was done using tBLASTn tool from NCBI. It was restricted to *N. gaditana* and *whole genome contigs* as a source of genomic data. (See Figure 13, where in the first section tblastn was used, in second section protein sequence was inserted and in the third section the *whole genome contigs* of *N. gaditana* genome were chosen). From the results of the search the most identical sequence was chosen for analysis. Because the prediction of structure of query protein was made based on the gene structure prediction, only the regions encoded exons are present in the resulted alignment and their identity have to be 100%. If the number of exons is more than one, it can be observed by the analysis of the alignment. If the first hit contains two matches in the same contig (Sequence ID) as shown in the Figure 14, (for the Nga00060) and these matches cover together 100% of the query sequence, it can be assumed that this protein contains two exons.

The PTS1 located at the extreme C-teminus, so the presence of stop codon immediately after the PTS1 tripeptide is important to identify. To check that two separated exons are really present in nucleotide sequence and that PTS1 is followed by the stop codon, the link «GeneBank» corresponding to the most interesting match had to be made active. Thereafter the new window corresponding to the selected region of the contig was automatically opened. The FASTA view mode was chosen, the selected region was extended by subtracting 500 bp from the start-point position and adding 500 bp to the end-point position in the «Change region shown» window. The shown sequence was translated using the ExPASy translation tool: (http://web.expasy.org/translate/). «Compact» output format was chosen in the translation options. The translation was done for three forward ORFs and three reverse ORFs. Our sequence was found in the first reverse ORF (blue frame in the Figure 15) by the copying the 14 amino acid terminal part of the aligned

sequence (position 65897 - 65911 in the Figure 15) and pasting into the search window of the internet browser (press Ctrl+F for Windows/Cmd+F for MacOs). The identical region was marked with yellow frame in Figure 15. These 14 amino acids were followed by the stop codon, therefore PTS1 prediction was correct. If the stop codon was not present directly after the PTS1, then the 14 amino acid terminated by the next following stop-codon were copied and verified in PredPlantPTS1.

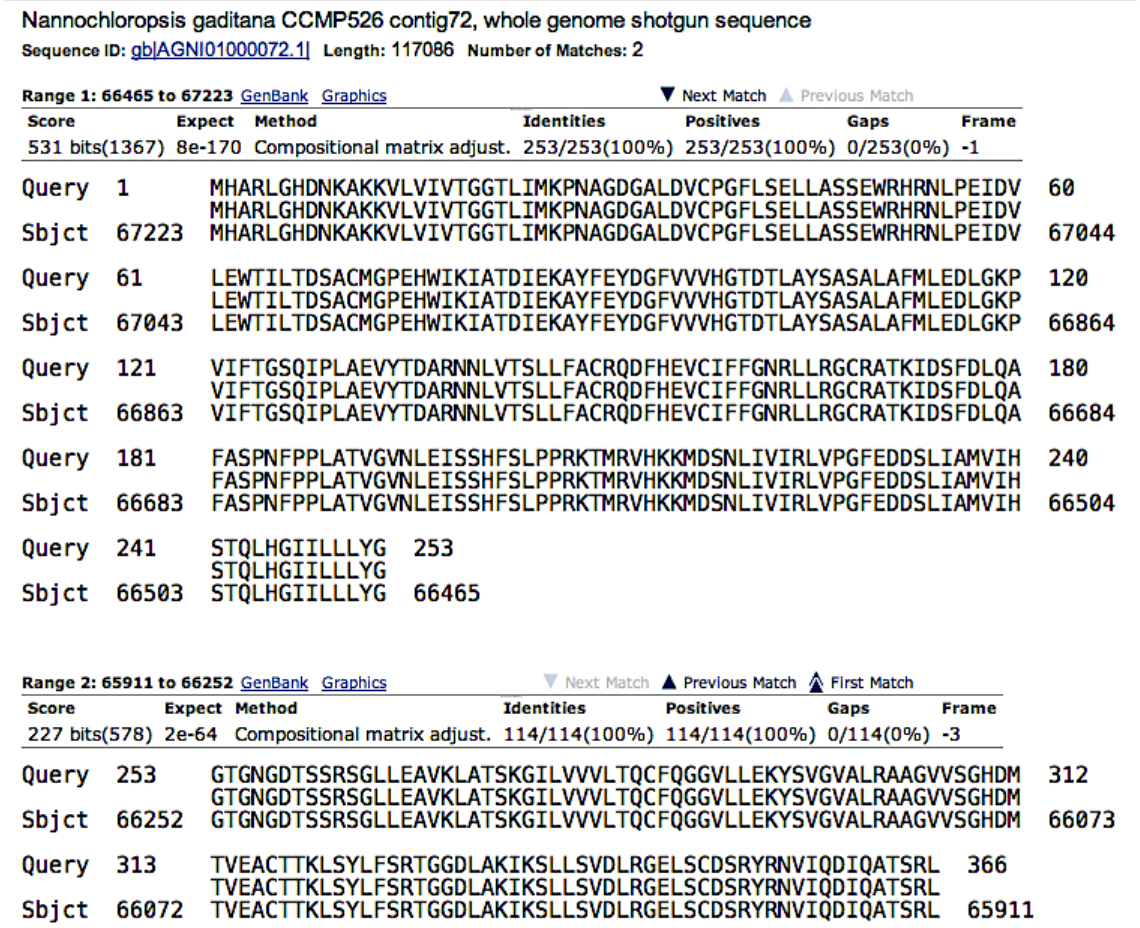


Figure 14. Print-screen of the tblatn search results of the predicted N. gaditana protein Nga00060 against whole genome contigs of Nannochloropsis species. Two alignment matches with different ORF in one contig cover 100% of the query protein. It was predicted that the corresponding gene has two exons.

The N-terminal region of the protein is not important for the PTS1 prediction, however for the successful peroxisomal localization results *in vivo* the complete sequence of at least one PTS1-carrying exon has to be estimated. For the single exon proteins Start codon sequence has to be verified. For two-exon protein Nga00060 the second exon was present in the same ORF (3'5' Frame 1»), however the first exon was translated using another ORF «3'5' Frame 2». The correct position of Start (M) and Stop (-) codons was checked by copy-pasting extreme N- and C-terminal regions from the protein alignment

represented in Figure 14. One exon gene cannot be translated by two different ORF, in such a way the hypothesis that the protein is coded by two-exon gene was verified.

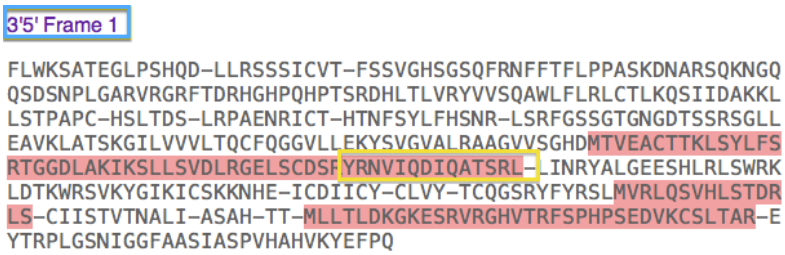


Figure 15. The result of translation of C-terminal exon using ExPASy translation tool. The ORF is marked by the blue frame. The last 14 amino acids of the C-terminal exon containing PTS1 marked by the yellow frame.

2.8.3.Prediction of PTS2 proteins in *N. gaditana and N. oceanica*

Proteins can be targeted to the peroxisomes by carrying C-terminal PTS1 tripeptide or by cleavable nonapeptide PTS2 which is located closer to the N-terminus. These are most abundant (but not the single) ways for soluble proteins to reach peroxisomes (see Chapter 1.3).

At the moment several peroxisomal protein databases are available. AraPerox 1.2 (http://www3.uis.no/araperoxv1/) was used in this study. It contains all known, experimentally verified proteins of *A. thaliana* with PTS2 represented (Reumann et al., 2004).

The proteins with verified PTS2 signals were blasted against nucleotide collection of *Nannochloropsis* genome to find homologs. BLAST parameters: tblastn, whole genome contigs, restriction to *Nannochloropsis*. The methodology is similar to verifying the presence of PTS1 during gene structure analysis described above (Chapter 2.8.2.1). However, for PTS2 prediction first exon had to be selected for the analysis and the region of the nucleotide sequence corresponding to this exon had to be extended by at least 1000 bp in the «Change region shown» window. Then, this nucleotide sequence was translated in ExPASy and the interested corresponding amino acid sequence was found in the one of six ORFs by «copy-pasting» as described above. The putative first exon length should be from the first start codon to the stop codon. The PTS2 nonapeptide should be located in the 40 amino acid region from the start codon and consist of represented amino acid combination [R/K]-[L/V/I]-x_5-[H/Q]-[L/A].

2.9.Molecular biology methods

Bioinformatical methods have to be proven by laboratory methods, because they have a lot of variabilities which can affect the result. One of such proving techniques is cloning of interesting genes, investigation of their sequences, with further subcloning and targeting analysis in a living cell. It is a very short description, the whole procedure is represented in Figure 16:

Peroxisomal protein prediction by bioinformatics

↓

Designing primers for 10 interesting genes

↓

Algal DNA isolation

↓

Analytical PCR (to verify correction of the primers for the 10 chosen genes and to find the best suitable Ta and elongation time)

↓

Analytical agarose gel electrophoresis (to check fragment size and PCR specificity)

↓

Analytical PCR (to verify correction of the primers for the 3 most interesting genes and to find the best suitable Ta and elongation time)

↓

Analytical agarose gel electrophoresis (to check fragment size and PCR specificity)

↓

Preparative PCR (to amplify 3 most interesting genes)

↓

Analytical agarose gel electrophoresis (to verify successful genes amplification)

↓

Preparative PCR (to isolate preparative PCR fragment)

↓

PCR fragment subcloning into pJET vector (TA cloning)

↓

E.coli transformation and pJET vector isolation

↓

Analytical PCR (to verify correction of the inserts for the 3 constructs and to find their orientation in the vector)

↓

Analytical agarose gel electrophoresis (to verify correct insert size and determine the DNA concentration)

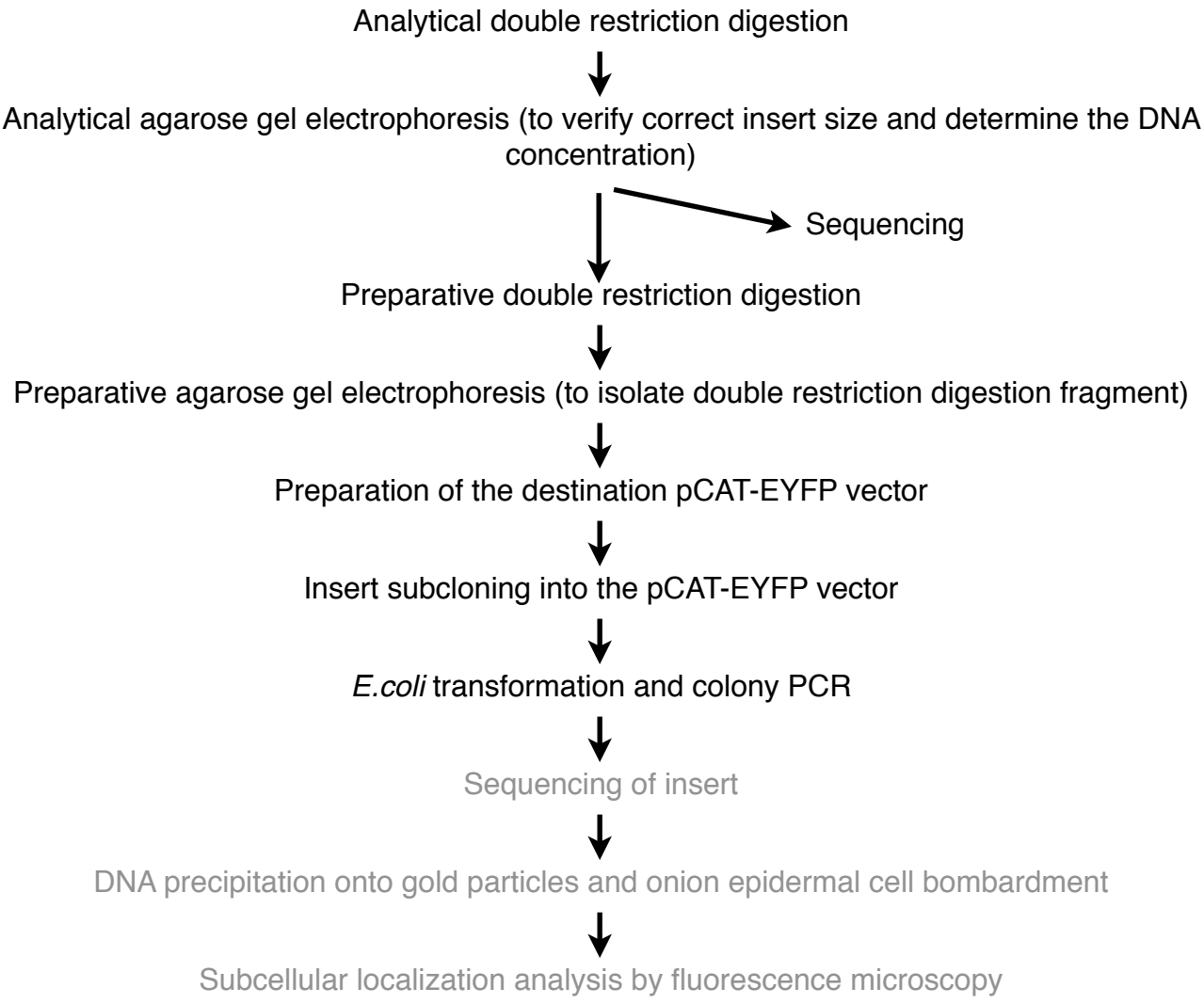


Figure 16. Summary of subcloning steps of PTS1 protein candidates in the back of EYFP using intermediate subcloning into pJET cloning vector. Three last steps were not performed due to the absence of transformed *E. coli* cells containing the designed constructs (See Chapter 3.4.5).

2.9.1.Algal cell harvesting

The protocol for harvesting the algal cell with intermediate partial purification was established for this project.

The algal culture was grown to reach maximum concentration, then it was distributed into the 50 ml Falcone tubes (for 100 ml of culture two tubes were used), the tubes were centrifuged in the Sorvall RC 5C Plus centrifuge using HS-4 rotor, 3500-4000 rcf 10-15 min at 4°C to get solid pellet. Then the supernatant was discarded and 50 ml of distilled water were added to the same tubes. The algal pellet was resuspended. The washing was repeated twice in order to get clear supernatant. After the last centrifugation the supernatant was removed and around 1 ml of water was added, the pelet was

resuspended and transferred into 1.5 ml (or 2.0 ml) Eppendorf tube. The tubes were stored in the freezer at -20°C.

A new protocol was established in order to get harvested cells ready for DNA extraction using DNA isolation kit (DNeasy Plant Mini Kit from Qiagen, Germany). 200 mg is a maximum recommended loading weight for the used kit, so it is necessary to make aliquots containing 200 mg wet weight of cells.

Preparation of harvested algal cells for the DNA extraction

The Eppendorf tubes containing algal cultures were taken out from the freezer and thawed on ice. Then the cultures were centrifuged at 10000 rcf for 5 min, the supernatant was removed, pellet was resuspended in about 200 μl of the nanopure water. The suspension was transferred into new 1.5 ml tube with known weight. Several suspensions were combined in this tube with known weight to get about 1.5 ml of total volume. When the tube was filled up, it was closed and centrifuged at 10000 rcf for 5 min. Supernatant was removed and the tube with algal paste was weighted and the weight of algal paste was calculated by subtracting tube weight from total weight. A small volume of nanopure water was added in order to dissolve the algae. The suspension was aliquoted between several new 1.5 ml tubes in such a way that 200 mg of algae in each tube were got. The tubes with algal suspensions were stored in the freezer at -20°C ready for the DNA extraction.

2.9.2.Extraction of algal DNA

The method of disruption of the frozen algae mass by mortar and pestle followed by additional disruption using manual homogenizer was adapted for algal DNA extraction.

Frozen harvested cells have to be thawed on ice. If the cells culture is dense, it has to be diluted, to be able to transfer all the algae mass. Liquid nitrogen from the storage tank has to be transferred into the special small container (the gloves and googles were used). The sterile pestle and mortar have to be cooled down by transferring some amount of liquid nitrogen to the mortar. When the nitrogen evaporated completely and pestle cooled down, algae solution using 200 μl pipetman should be very carefully transferred drop by drop. The drops immediately stick to the cold surface of the mortar and freeze. Then, the drops have to be squashed by applying powerful circular movements by the pestle (use gloves). To prevent movement of the mortar it has to be put onto stable pedestal. When all the drops are crushed in paste, it has to be transferred into special microcetrifuge tube with thick walls and mixed for some minutes using manual homogenizer with sterile tip.

When all the above steps were made, the disrupted cells were loaded into 1.5 ml tubes and DNeasy Plant Mini Kit from Qiagen (Germany) was used, where they melted down and handled according to the isolation kit instructions.

400 µl of Buffer AP1 and 4 µl of RNase A were added. The solution was vortexed and incubated for 10 min at 65°C. The tubes should be inverted 2–3 times during incubation. Then, 130 µl Buffer P3 were added, mixed and incubated for 5 min on ice. After that the lysate was centrifuged for 5 min at the top speed. The lysate was pipetted into a QIAshredder spin column and placed in a 2 ml collection tube. Next centrifugation was done at 20,000 x g for 2 min. The flow-through was transferred into a new tube without disturbing the pellet if present and 1.5 volumes of Buffer AW1 were added. The solution was mixed by pipetting.

650 µl of the mixture were transferred into a DNeasy Mini spin column placed in a 2 ml collection tube. Centrifugation was done at ≥ 6000 x g (≥ 8000 rpm) for 1 minute. The flow-through was discarded. The last step was repeated with the remaining sample. The spin column was placed into a new 2 ml collection tube. 500 µl Buffer AW2 were added, and centrifugation was performed at ≥6000 x g for 1 minute. The flow-through was discarded.

After that, another 500 µl Buffer AW2 were added and the mixture was centrifuged at 20,000 x g for 2 min. The spin column should be carefully removed from the collection tube, so that the column does not meet the flow-through.

In the last step, the spin column was placed to a new 1.5 ml or 2 ml tube, 50 µl nanopure water was added for elution. After the incubation for 5 min at room temperature (15–25°C), the solution was centrifuged for 1 min at ≥6000 x g. The elution could be done twice into the same tube or to a new one. The DNA concentration was measured at Nanodrop spectrometer.

2.9.3.Determination of DNA concentration in solution

The Nanodrop 2000 spectrometer (Thermo scientific, USA) was used for determination of genomic or plasmid DNA concentration in microliters of solvent.

Before the use of instrument the lenses must be cleaned by adding 2 *µ*l of nanopure water on them and wiping them using special lens tissue. To calibrate the instrument 1 *µ*l of nanopure water was placed on the instrument and measurements were done by clicking corresponding button in the computer interface. When the calibration was done, 1 *µ*l of the

sample was pipetted on the lens, the sample was given a name in the corresponding window of the computer program and the measurements were performed. After each measurement the lenses of the spectrometer have to be wiped using lens tissue. The concentration values were printed using connected printer.

2.9.4.Polymerase chain reaction (PCR)

PCR is a very powerful and useful technique nowadays. In this technique nucleic acid regions can be amplified in vitro by a thermostable DNA polymerase. As a result one gets high amount of copies of the region of interest. They can be used in the gel electrophoresis or directly in a transformation of a vector (Campbell et al., 2008).

PCR, as all other methods in natural science, has two types of analysis - two modes of working process: analytical mode and preparative one. Typical aim of choosing analytical mode is to get evidence of something or to test working parameters. It is different for the preparative mode, which is used when the high quality/quantity results are needed. Hence, it is wise to use more material (more time and money) for the second mode if one`s aim is to use the results in subsequent steps. If the results of experiment are not supposed to be used as a material for the next experiments, the analytical PCR should be chosen. The master mix for analytic PCR can be found in Table 3.

Table 3.Typical analytical PCR master mix:

Component	Volume (μL)	Final concentration
Sterile nanopure water	13.5	
10x Taq buffer	2	1x
25 mM MgCl2	2	2,5 mM
10 mM dNTPs	0.5	0,25 mM
10 μM forward primer	0.5	0,25 μM
10 μM reverse primer	0.5	0,25 μM
Template DNA	0.5	
Taq polymerase (1U/μl)	0.5	
Total volume	20	

Composition of PCR mix used in preparative PCR is the same as in analytical PCR, with the only exception that 0,5 μL more water is used, based on the assumption that the

volume of the sample is close to zero. The master mix for preparative PCR can be found in Table 4.

Table 4. Typical preparative PCR master mix.

Component	Volume (μL)	Final concentration
Sterile nanopure water	34.5	
10x Taq buffer	5	1x
25 mM MgCl2	5	2,5 mM
10 mM dNTPs	1	0,25 mM
10 μM forward primer	1	0,25 μM
10 μM reverse primer	1	0,25 μM
Template DNA	1	
Taq polymerase	0.5	
Total volume	50	

For analytical PCR a homemade Taq DNA polymerase was used. It is relatively cheap and therefore can be effectively used in large scale qualitative analysis. For preparative PCR more advanced, effective and expensive Phusion High-Fidelity DNA Polymerase (Thermo Scientific, USA) was used. PCR elongation time is adapted according to the expected length of the PCR product, about 1 minute per 1 kb. The annealing temperature (T_a) for each primer was calculated by the formula:

$$T_m = 69.3°C + 41\%GC - 650/n,$$

$$T_a = \ - 3°C,$$

Where Tm is melting temperature, %GC is a ration of the bases guanine (G) and cytosine (C) to the total number of bases (n) of the primer. If the primers had different Tm, the lowest value, was taking for calculation of Ta. The two types of annealing temperatures were used for the PCR programme: for cycles from the first to the tenth lower Ta was used, than for cycles from the eleventh to the thirty-fifth. It gave more unspecific products at the first stage which were further amplified during the next one. This technique is usually used for «difficult» amplicons. The typical programs for analytical and preparative PCR can be found in Table 5.

Table 5. Typical PCR programs

Step	Cycle	Preparative PCR		Analytical PCR	
		Temp. (C)	Time	Temp. (C)	Time
Initial denaturation	1	96	2 min	96	5min
Denaturation	1-10	96	30 sec	96	45 sec
Annealing	1-10	Ta	30 sec	Ta	45 sec
Elongation	1-10	72	30 sec - 4 min	72	30 sec - 4 min
Denaturation	11-35	96	30 sec	96	45 sec
Annealing	11-35	Ta	30 sec	Ta	45 sec
Elongation	11-35	72	30 sec - 4 min	72	30 sec - 4 min
Final elongation	36	72	10 min	72	10 min
Cooling	36	12		12	

Colony PCR was used to identify transformants that contain the desired plasmid. It is usually done by growing distinct clones of transformants on agar LB medium. A portion of a colony was touched with fine pipette tip in order to get a small bit of cells culture and, then added to the master mix and used in PCR solution.

2.9.5.Gel electrophoresis

Agarose gel electrophoresis was used to separate DNA fragments according to their mobility under electric field in the gel. It was used to separate DNA sequences from 380 bp to the vectors which are more than 4000 bp in size. The preparative and analytical methods are also used in this technique. The difference between them was in amount of loading sample. The analytical agarose gel electrophoresis was used to check fragment size, PCR specificity and parameters of PCR program. Preparative agarose gel electrophoresis was used to separate preparative PCR products and double restriction digestion fragments.

An agarose powder dissolved in 1x TAE buffer was used for the preparation of the common 1% agarose gel. Lower concentration of agarose was used for separation of large DNA fragments to increase separation speed. The agaroze dissolved in 1xTAE buffer was heated up in microwave oven and heated up to about 60°C before pouring into the gel casting apparatus with a comb in it. After solidifying the comb was removed and the gel

plate together with the gel was transferred in to a gel running apparatus containing 1xTAE buffer. The level of the buffer should be several mm higher, than the gel to cover it.

To make a 1000 ml of 50xTAE buffer 242 g of 2Mtris-base was added to the 100 ml of 0.5 M EDTA (pH 8.0), than 57.1 ml of acetic acid (pH 8.3) was added and the sterile nanonpure water was added to get the total volume 1000 ml. The 50x TAE buffer was diluted to get working concentration 1x.

To make 1% (w/v) agarose gel, powdered agarose 1% (w/v) was melted in 1x TAE buffer (e.g. 0.5 g/50 ml) and then casted into a plate with the comb(s) for well-making.

Into a first well a mixture of DNA size marker (1 kb Gene Ruler, 0.5 ng/μl) and GelRed was loaded to compare. The size marker was used to compare the intensity and position of the sample bands. In the next wells the sample mixture was loaded. The sample mixture consists of the sample (1-3 µl), GelRed fluorescent dye (1 µl), 6x DNA loading dye (2 µl). The GelRed binds to the DNA and was used to visualize the position of the DNA in the gel under UV light. The loading dye was used to see the form of the moving DNA. The speed of moving of the DNA in the gel depends on the amount of the DNA loaded into the well (directly proportional). The gel was electrophoresed at 60 - 90V for 40 - 90 minutes respectively. The pictures of the gel were done under UV light.

2.9.6.Purification of DNA from the agarose gel electrophoresis and PCR reactions

The purification of DNA from agarose gel electrophoresis and PCR reactions was done using PCR Purification kit and Gel Extraction purification kit respectively. Both kits were obtained from Fermentas, Germany.

To extract DNA from the agarose gel the band corresponding to the interesting fragment was cut out and placed into the 1.5 ml tube with known weight, the total weight was measured and a weight of the gel slice was calculated. The Binding buffer was added in a proportion 1:1 (volume of buffer:weight of gel) and then mixed and incubated in water bath at 55°C for 10 min.

To purify the PCR mixture the Binding buffer in proportion of 1:1 (v:v) was added to the PCR mixture.

After adding the Binding buffer to the DNA mixture from PCR and gel the procedure is the same for both kits and is represented below.

The solution was transferred into the GenJET purification column and centrifuged at 12000 rpm for one minute. The flow-through was discarded and 700 µl of Wash buffer were added into the column. The mixture was centrifuged at 12000 rpm for one minute, the flow through was discarded,the empty column was centrifuged in order to remove the rest of Wash buffer. The column was placed into a new collection 1.5 ml tube and 35 µl of sterile nanopure water was added. After the centrifuging at 12000 rpm for one minute, the flow through containing DNA was collected. The concentration was measured using the method described before (Chapter 2.8.3). The DNA was stored at -20°C.

2.9.7.Restriction digestion

The restriction digestion was done both in analytical and preparative ways. The analytical one was performed to check the probability of digestion, while preparative one was used for subcloning. The mixture for double digest was made according to Table 6, where Buffer is the most suitable buffer for the pair of restriction nucleases (first RE and second RE) from Table 7. The mixture was incubated at 37°C for 3 h (analytical) or overnight (preparative).

Table 6.Components of analytical restriction double digest

Component	Amount/Volume (µL)	Final concentration
Sterile nanopure water	up to 10 µl	
plasmid DNA	1 µg	0.1 g/l
10x Buffer	1 µl	1x
10 U/µl first RE	0.5 µl	5 U
10 U/µl second RE	0.5 µl	5 U

The Single digest by BamHI was done for the linearization of pSelect100 plasmid used for the *Nannochloropsis* transformation. For the double digest of *N. gaditana* proteins Nga00060 and Nga00170 SacII and NotI were used in G buffer; to cut Nga05502 NotI and XbaI were used in O buffer. The amount of RE used was correlated with their activity in a chosen buffer according to the table 7. For example, for combination NotI and XbaI in G buffer, the NotI used in 2-3 times more concentrated than XbaI. All the enzymes used produce sticky ends.

Table 7. Restriction enzymes and their activity in different buffers

DNA sequence	RE	Recommended buffer	Restriction enzyme activity, %					
			B	G	O	R	Tango1X	Tango2X
Nga00060	SacII	B	100	50-100	(0-20)	(0-20)	50-100	(0-20)
	NotI	O	(0-20)	20-50	100	20-50	(0-20)	20-50
Nga00170	SacII	B	101	50-100	(0-20)	(0-20)	50-101	(0-20)
	NotI	O	(0-20)	20-50	101	20-51	(0-20)	20-51
Nga05502	NotI	O	(0-20)	20-50	100	20-50	(0-20)	20-50
	XbaI	Tango	50-100	50-100	20-50	(0-20)	100	50-100
pSelect100	BamHI	Unique	20-50	100	20-50	50-100	100	50-100

2.9.7.1.Ligation of inserts into pJET vector

Two different types of ligations were used. TA-ligation of the constructs into the pJET blunt cloning vector (See Chapter 2.2.2) and ligations of constructs into the pCAT-EYFP plant expression vector (See Chapter 2.2.1). In both cases the T4 DNA Ligase enzyme was used to ligate ends of insert to the ends of the vector. In cloning to the pJET vector the blunt end DNA strands were used (no overhang), while for the cloning into pCAT vector the DNA strands of the vector and insert had small overhangs which are identical if the same RE were used for cleavage. The amount of the insert added to the reaction mixture was calculated using the followed formula:

ng of insert = 3 * ng of vector * size of insert / by size of the vector

The insert in the TA-ligation was a PCR product containing TA blunt ends generated by proofreading DNA polymerase. The linearized pJET cloning vector contained similar TA blunt ends. The components represented in Table 7 were added in to 1.5 ml tube. The mixture was vortexed and incubated for 5 minutes at room temperature. After that it was used for the transformation of *E. coli* cells (Chapter 2.8.8).

Table 8.Components used for the cloning blunt-end PCR products into pJET vector

Component	Volume
2X Reaction Buffer	10 µl
Non-purified PCR product *or* purified PCR product/other blunt-end DNA fragment	1 µl 0.15 pmol ends
pJET1.2/blunt Cloning Vector (50 ng/µl)	1 µl (0.05 pmol ends)
Water, nuclease-free	up to 19 µl
T4 DNA Ligase	1 µl
Total volume	20 µl

2.9.7.2.Ligation of inserts into pCAT-EYFP vector

For the ligation of the inserts into the pCAT-EYFP the construct pCAT-EYFP-DECR was used. The cDNAs released from the pJET cloning vector by restriction endonuclease cleavage (See Chapter 2.9.7) were inserted in the back of EYFP instead of DECR which was cleaved by the same restriction nucleases as ones used to cut cDNAs from the pJET vector. The mixture was made in the 0.2 ml tube which was placed into the incubator at 16°C overnight.

Table 9. Components used for the cloning of sticky end PCR products into pCAT vector

Component	Amount/Volume (µL)
Sterile nanopure water	up to 10 µl
pCAT vector	100 ng
cDNA insert	calculated
10x T4 DNA ligase buffer	1 µl
T4 DNA ligase	1 µl

2.9.8.Transformation of *E. coli* JM109

The 1.5 ml tube containing competent cell *E. coli* JM109 was taken from the freezer and thawed on ice. The 1µl - 10 µl containing 1 - 10 ng DNA for transformation was added to the competent cells and mixed carefully by pipetting. Then, the tube was incubated for 30 minutes on ice. After that the hit-shock was applied to the cells by transferring the tube into the water bath with the temperature of 42°C for 75 seconds. After that the tube was immediately placed back on ice for two minutes. Then, 600 µl of LB media were added to

the cell mixture and the tube was place into the incubator with shaking for one hour at 37°C. Later, the cells were plated on 1% ampicillin-containing LB agar plates. The plates were incubated at 37°C overnight.

2.9.9.Plasmid DNA isolation

GeneJet plasmid miniprep kit (from Fermetas, Germany) was used to isolate plasmid DNA from *E. coli*. Single bacteria colony was transferred from the LB ampicillin containing plate to the 10 ml (15 ml for low copy plasmid) of liquid LB ampicillin by sterile wooden stick. The culture was incubated overnight at 37°C with shaking. The cells were pelleted by centrifugation at 6000 rpm for 2 minutes. The supernatant was removed and the bacterial pellet was resuspended by adding 250 µl of the resuspension solution and vortexing. Then, 250 µl of Lysis solution were added. The tube was inverted four-six times. After that 350 µl of Neutralization solution were added and mixed by inverting the tube four-six times. The tube was centrifuged at the maximum speed (20000 rpm) for 5 minutes. The supernatant was transferred into a GeneJET spin column which was placed in a collection tube and centrifuged at maximum speed for one minute. The flow-through was discarded and the 500 µl of Wash solution were added. The mixture was centrifuged at the maximum speed for one minute. The flow-through was discarded and the washing step repeated with 500 µl of Wash solution. To completely remove the Wash solution from the column the sample was centrifuged at the maximum speed for one minute. Then the spin column was transferred into a new sterile Eppendorf tube and 35 µl of the nanopure water was added. After two minutes of incubation at room temperature the tube with a column was centrifuged at the maximum speed for one minute. The flow-through containing plasmid DNA was collected and a concentration was measured using the Nanodrop 2000 (for details see Chapter 2.8.3). The isolated plasmid DNA was stored at -20°C.

2.10.Genetic transformation of the *Nannochloropsis* algae by electroporation

The transformation is a transfer of genetic information of one organism into another one with the following change of properties of target organism.

The genetic transformation of *Nannochloropsis* by electroporation has been recently established and published (Kilian et al., 2011; Radakovits et al., 2012; Vieler et al., 2012). The pSelects100 vector was designed for the transformation of *N. oceanica* and obtained from C. Benning, Department of Biochemistry and Molecular Biology, Michigan State University, USA. The availability of the vector and of the transformation protocol

determined the decision to establish genetic transformation primary for the *N. oceanica* by electroporation.

For the transformation the following species were chosen: N. oceanica CCMP 1779 and N. gaditana CCMP 526. The vector is pSELECT100.

In the transformation of Nannochloropsis oceanic ccmp1779 the following materials were used:

•pSELECT100 selection vector linearized by BamHI restriction enzyme (Chapter 2.9.7).

•Salmon Sperm DNA (must be heat denatured at 95°C for 1 min)

•375 mM Sorbitol (Mw 182.17 g/l), filter sterilized (0.24 μm) and cooled down on ice;

•Recovery tubes (1 per sample): sterile and prelabeled 15 ml Falcon tubes with 5 ml ½ salinity f/2 autoclaved medium;

•Selection Plates (1 per sample): 150x10 mm ½ salinity f/2 agar plates containing 50 μg/ml Hygromycin B;

•Sterile electroporation cuvettes with 2 mm gap for the MicroPulser Electroporator, BioRad, USA (www.bio-rad.com/) (Figure 17);

•Early log phase Nannochloropsis cells (1.0x108 cells per transformation reaction);

•Top Agar: ½ salinity f/2 culture medium with 1% microbiological agar.

The cells were grown in normal f/2 media, the cell concentration was estimated by optical density measurements at the 540 nm wavelength. The cells have to be at the early stage of logarithmic growth phase when their cell walls are not very thick. The cell cultures were transferred into the Falcon tubes (5 ml in each) which were put inside of rubber adapters for the centrifuge rotor and centrifuged for 15 minutes at 4000g and 4°C. The supernatant was removed and the cells were resuspended in 15 ml of 375mM Sorbitol to remove salts. The washing was done two-three times. Finally, the pellet was resuspended in 0.2 ml 375 mM Sorbitol for each tube.

3-5 μg of linearized vector pSELECT100 (Chapter 2.9.7) were added together with 40 μg of Salmon Sperm DNA. The mixture was incubated on ice for 10 min. The volume of the vector DNA used per reaction depends on the volume of DNA that is added to the cells and should not exceed 10% of the total volume. After 10 minutes the cultures were transferred into the 2 mm cuvettes.

The electroporation was performed on the MicroPulser with the following parameters: Voltage: 2.2 kV, time of pulse - 10 sec, Maximum capacitance 10 μF. Actual volts and time constant can be read after each electroporation. If the sample has a low resistance, the arc can occur. It means that no significant electric flow went through the sample. The arc occurs in the medium containing relatively high salt concentration, that is why the volume of the added vector solution should be less than 10% of the total volume. In order to avoid the arc production, the sample has to be twice diluted with Sorbitol or the voltage has to be decreased (down to 1.6kV).

Immediately after the electroporation, the cells were transferred from the cuvette to the prepared culture tube with 5 mL sterile 1/2 salinity F/2 media, then these Falcone tubes were transferred to the plant growing chamber (Panasonic MLR-351), and cells were left to recover for 48 hours at slow shaking (1.2 speed setting) and light 13000lux, temperature 22°C.

After 48 hours the top agar was prepared (melted in microwave and kept at 42°C in a preheated water bath) and only after that 5 ml of agar were transferred into each culture tube containing 5 ml cell suspension. The solution was mixed by pipetting up and down several times (or inverting the tube 3-4 times) and poured on 150 mm diameter selection plates containing 50mg/ml Hygromycin B. The plates were stored in the same plant growth chamber and sealed with parafilm only after 5 days of storing.

After two weeks single colonies should become visible. For screening the colonies have to be picked, resuspended in 200 μl of water, spread on a new selection plate to get single colony. The colonies should grow for a few days to became visible and then screened on presence of insert using PCR, Western Blot, etc.

Figure 17. MicroPulser electroporation apparatus and a cuvette from BioRad, USA (www.bio-rad.com/).

3.Results

The present project was focused on the establishment of microalgal research. Several basic culturing techniques for *Nannochloropsis* were established, such as, cell counting and spectroscopic measurements, harvesting and genomic DNA isolation. The analysis of growth curve of three microalgal species was done.

The experiments for purification of contaminated cultures, such as, sample dilution, antibiotic treatment and plating, were performed.

The methodology of neutral lipid staining by fluorescence microscopy was established and four *Nannochloropsis* species were analyzed by this technique.

The prediction and analysis of PTS1/2 proteome of soluble matrix proteins for *N. gaditana* were made. Two full-length cDNAs and one C-terminal exons of predicted PTS1 proteins of *N. gaditana* were cloned into pJET vector. In addition the successful transformation *of N. oceanica* was performed. The transformation of *N. gaditana* was unsuccessful.

3.1.Comparative growth analysis of different Nannochloropsis species

Monitoring growth and proliferation of microalgae is fundamental to all attempts in studying and optimizing growth rates and lipid productivity. Several methods for quantify microalgal growth have been reported in the literature, and the most important ones have been established in the Reumann group in the present M.Sc. thesis to monitor growth and proliferation of microalgae (see above, Chapter 2.4, 2.5, 2.6).

3.1.1.Choice of the optimum wavelength for spectroscopic measurements

Several methods for cell counting are available. They can be classified into two categories: those that are based on the direct cell counting (based on using microscopy and based on using flow cytometry) and indirect methods. Indirect methods of investigation of the amount of cells are based on the correlation between some properties of cells culture and cell number such as dry weight and optical density measurements.

The optical density of algal cultures was initially measured at three different wavelengths, namely 540, 625 and 680 nm. These three wavelengths were chosen because their application had been previously reported for *Nannochloropsis* given (Rocha et al., 2003; Hsueh et al., 2007; Gu et al., 2012). If the absorbance values were higher than 0.6, the

samples were diluted in order to avoid scattering in oversaturated media. The comparative OD measurements were done for three representative species, namely available *N. salina* (strain NIVA-2/01), *N. oceanica* (strain NIVA-2/03) and *P. tricornutum* (NIVA-BAC 2). The experiment was carried out over 36 days with periodic OD measurements every 2-3 days (Fig. 18, 19). A 100-ml culture (containing nitrogen source - $NaNO_3$) was inoculated with a few ml of a logarithmic algal culture and grown under standard conditions (18/6 of light/ dark cycle, shaking at 1.2 speed, Chapter 2.3.2).

For all measured microalgal cultures it was revealed that the OD at 540 nm is higher than both at 625 and 680 nm which have very similar values. The increase in OD over culturing time was very close to linear for all the wavelengths. Figure 18 represents a plot of absorption values of N. salina versus culturing days when measurements were performed from the moment of culturing up to day 36 at three different wavelengths. This figure is representative for the data obtained for the different microalgae.

Interestingly, in this long-term culturing experiment, a change in pigmentation was observed in the *Nannochloropsis* cultures. In the early growth phase, i.e., up to an OD (540 nm) of approx. OD = 1 the cultures were green, indicating a high chlorophyll content. Later, the cultures became yellowish, and after one month of subculturing the cultures were totally yellow. These observations correlate with the facts found in literature that the carotenoids/chlorophyll ration should increase during propagation of *Nannochloropsis* species (Solovchenko et al., 2011) Similar changes were observed for *P. tricornutum*, except for the fact that this culture was initially brown.

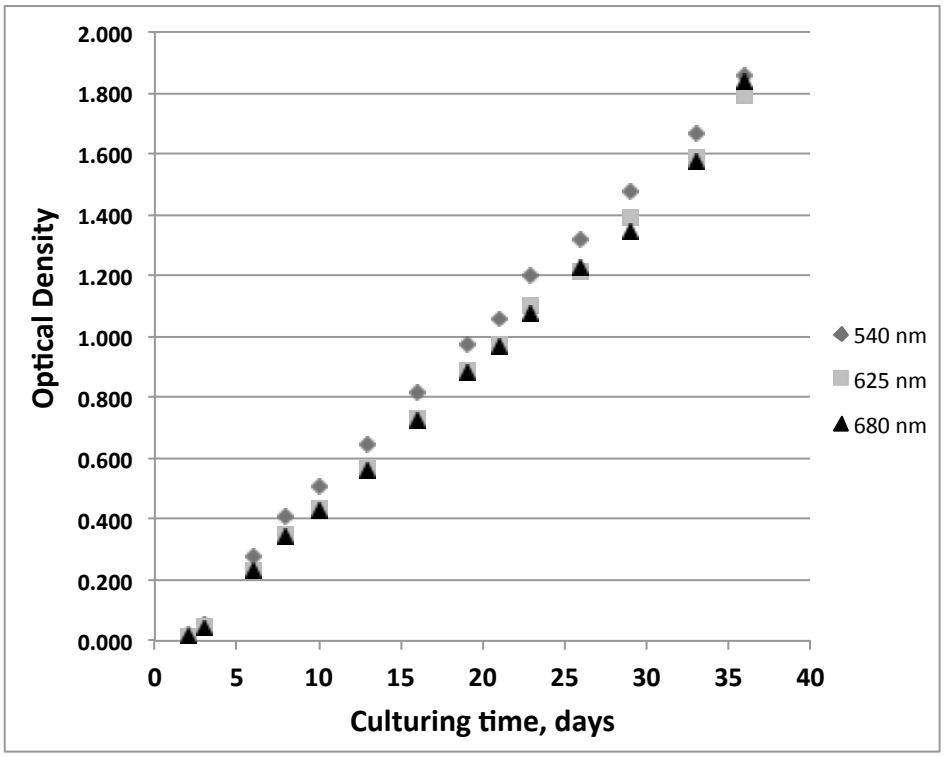


Figure 18. Observation of optical density at 540, 625 and 680 nm of N. salina NIVA-2/01 cultures in F/2 medium during 36 days. The measurements were performed three times for every time point. When OD values overgrew 0.6 the samples were diluted in proportion 1:2 and measured values were multiplied by the dilution factor 2. The absorption at 540 nm was the highest, while absorption values at 625 and 680 nm were slightly lower. Such difference can be explained by pigment composition of the cells.

It was concluded from these data that OD (540 nm) measurements are a quick and relatively reliable methods to monitor growth and proliferation of the microalgae of interest and that the chosen wavelength influenced the OD measurements only marginally. Nevertheless, OD at 540 nm measurements might be affected by parameters such as changes in pigmentation and cell size and do not give very accurate data under varying culturing conditions and in comparison between microalgae.

Figure 19 illustrates the changes in optical density of the three above-mentioned algal cultures, as shown representatively for the wavelength of 540 nm. The choice of using preferably 540 nm wavelength was made to avoid influence of probable chlorophyl a degradation during culturing, which can effect on the absorption at 625/680 nm.

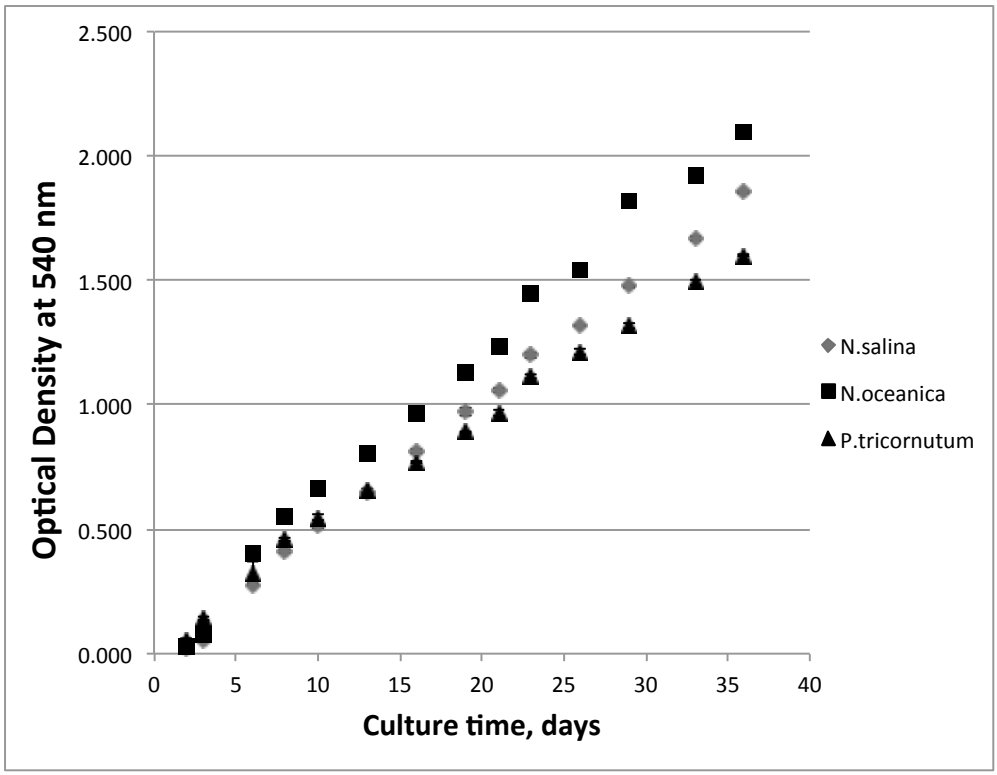


Figure 19. Comparative growth analysis of three microalgae (*N. salina*, *N. oceanica* and *P. tricornutum*) in F/2 medium over 36 days by measurement of the optical density at 540 nm. The measurements were performed in triplicates for each time point (See Chapter 2.4.3). When OD(540)>0.6, the samples were diluted to OD<0.6 and measured values were multiplied by the dilution factor. The culture of *N. salina* had the highest absorption from the third day. The OD values of *P. tricornutum* were smaller relative to the other microalgae, resulting in the lowest slope. The OD values for *N. oceanica* lied in the middle range between those of *N. salina* and *P. tricornutum*.

All cultures had similar and very low OD values in the first five days after subculturing. Later and until the end of measurements the *N. oceanica* culture showed the highest values and the steepest slope, indicating the highest growth rate. *N. gaditana* showed similar OD (540 nm) values as *P. tricornutum* up to the thirteenth day and then the OD values were in the middle range between *N. salina* and *P. tricornutum* (Figure 19). The OD values for P. tricornutum values after the thirteenth day were the lowest in comparison to the other cultures.

The standard deviation of the three algal aliquots of the same culture was calculated for all measurements and was very low (approx. 1.5% for undiluted samples and around 2.5% for diluted algae samples, see Fig. 19). A correlation between standard deviation values and wavelength were not found, indicating that the accuracy of the measurements made at all the wavelengths was equal.

3.1.2.Cell counting

As was written above direct and indirect methods of estimation of cell concentration are available (Chapter 3.1.1). Direct methods are more accurate, but time consuming, while indirect ones are generally fast, but susceptible to the different factors, such as, scattering of the turbid medium or presence of bacterial contaminators.

The indirect methods for microalgal growth analysis, appropriate calibration curves have to be made. Considering the availability of instruments, direct counting using the counting chambers (Chapter 2.4.2) and indirect cell growth analysis by OD measurement (Chapter 2.4.3) were chosen as our working methods. The objective of this thesis was to establish both a highly accurate and a «quick and easy» method to monitor the growth and proliferation of microalgae.

The methodology of direct cell number analysis is provided in the Material and Methods (Chapter 2.4.2). Where the cells loaded into the Bürker counting chamber were counted using microscopic observation. The number of algal cells counted in 0.1 μl was multiplied by x10000 and the dilution factor in order to obtain the cell concentration per ml of culture. The direct cell counting of microalgal cultures was done periodically with time intervals of three to seven days during the long-term growth experiment of 36 days int total started from the day of subculturing. The algae species used for this analysis were the same as used for the spectroscopic analysis of growth: *N. oceanica*, *N. salina* and *P. tricornutum* (Chapter 3.1.1 and Figures 18 and 19). For each culture two samples were collected and four repeats of counting were performed for each sample (one repeat was equal to counting of one 0.1 mm^2 square, thus eight counts per algal culture). The counts were averaged, and the standard deviation indicated the accuracy of the measurements. Following measurements were done every 2-4 days (Figure 20).

For *Nannochloropsis* species the cell concentrations were around $1.2x10^5 \pm 2.0x10^4$ cells/ml and $4.2x10^5 \pm 6.0x10^4$ cells/ml for the second and third day, respectively. For *P. tricornutum* the first measurment was made on the third day, cell number was 820000±24000 ($0,82*10^6$)per ml. Next measurement was made on the tenth day. By direct cell counting, the highest cell concentrations were detected for *N. oceanica* (e.g. $3.1x10^7 \pm 2.4x10^6$ cell/ml of day 19, Fig. 20), moderate concentrations for *N. salina* ($1.8x10^7 \pm 2.4x10^6$ cells/ml of day 19, Fig. 20) and the lowest concentrations for *P. tricornutum* ($8.96x10^6 \pm 1.65x10^6$ cells/ml of day 19, Fig. 20). Considering very similar inocula used for all three cultures, the data

indicated that *N. oceanica* grew best, followed by *N. salina* and *P. tricornutum*. During culturing time *N. oceanica* continuously had the highest cell concentration with maximum $3.1x10^7 \pm 2.4x10^6$ cell/ml detected on the twenty-first day. A few measurements (at the thirteenth and nineteenth days) showed an unexpected decrease in cell growth which might have been caused by a technical error, but these results do not have sufficient level of significance and should not be considered as totally correct.

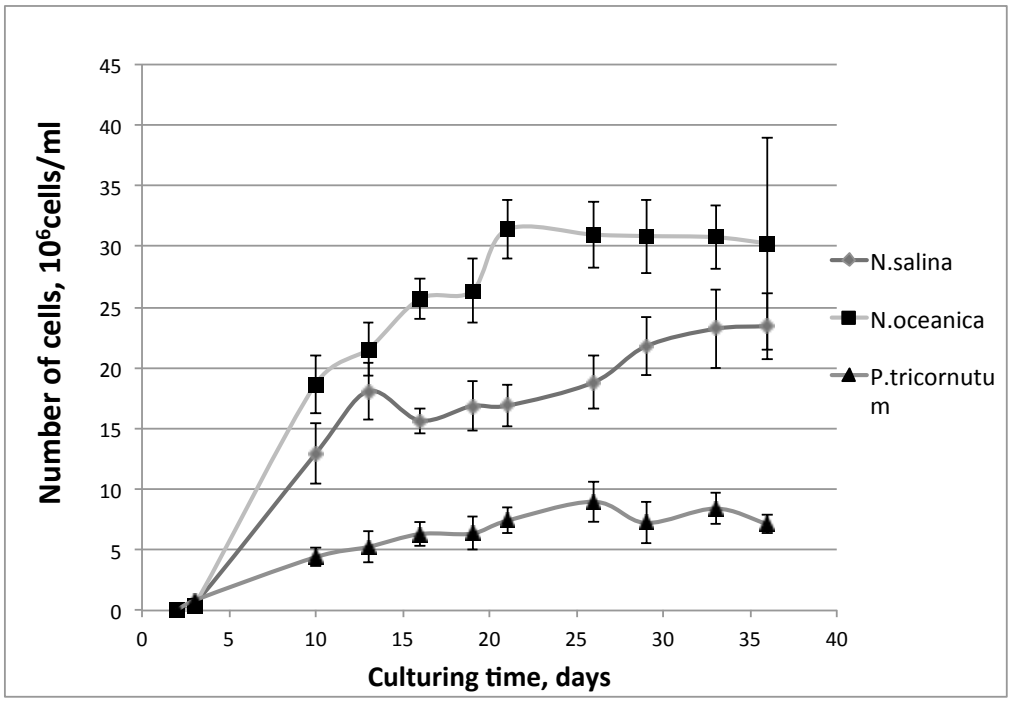


Figure 20. Direct determination of cell concentrations of three microalgal cultures. Number of cells of *N. salina*, *N. oceanica*, *P. tricornutum* were counted using the Bürker counting chamber during 36 days of culturing. On the Y-axis the cell concentrations are expressed in millions per ml. Time intervals between counts vary from 3 to 7 days. The error bars show standard deviation values.

Considering the measurements before thirteenth day of culturing *N. oceanica had* the steepest growth slope, *N. salina* had moderate steepness of the slope and *P. tricornutum* had the most gentle one.

N. salina grew most rapidly and reached intermediate maximum on the 23rd day with concentration $1.8x10^7 \pm 2.4x10^6$ then, cell number became lower and grew with lower intensity than before reaching the highest value on the last day of experiment - 36th. At that moment a cell number was $2.4x10^7 \pm 2.7x10^6$ cells in 1 ml.

P. tricornutum started to grow with the highest initial cell number 820000±24000 of cells per ml on the third day, after that it increased gently and reached its maximum on the

twenty-sixth day with the number of cells of $8.96x10^6 \pm 1.65x10^6$ in 1 ml. In the period from day 26 to 36 the cell concentration slightly decreased. The last measurements showed $7.1x10^6 \pm 7.6x10^5$ cells/ml. Standard deviation values were the lowest for *P. tricornutum* cultures and equally higher for both *Nannochloropsis* cultures. While relative standard deviation was almost equal for *Nannochloropsis* species (12.6%), it was higher for *Phaeodactylum* (16.4%).

3.1.3.Contamination analysis and purification of cultures

After obtaining *Nannochloropsis* and *Phaeodactylum* cultures from the NIVA, Norway (See Chapter 2.1) it was found by microscopy that the cultures had contaminators. Later the originally axenic cultures from CCMP were also contaminated during the growth. It was decided to find a method to remove the contamination.

Contamination is a very serious problem in microbiology and culturing of untransformed microorganisms in particular. Unwanted microorganisms often grow more rapidly than a major strain and can affect the accuracy of measurements or growth of the strain. In order to get more precise results, it is recommended to work with axenic cultures. Dougherty in 1959 wrote that «Axenic cultures, usually referred to as “pure cultures”...» and «axenic cultivation is the rearing of one or more individuals of a single species on a nonliving medium». In some cases, like DNA sequencing or physiological experiments this is necessary to obtain the axenic cultures.

A contamination can be removed by plenty of methods:

The use of detergents can remove bacteria sticking to algal cell membranes (McDaniel et al., 1962) or a combination of using detergents and centrifugation can give better results (Stein, 1980). Water samples can be diluted multiple times to get uni-cellular cultures. The dilution method works when the targeted species predominate in the sample. The method of plating works only with algae which grow on agar, from where the single colonies can be isolated. However, according to previous studies (Sensen et al., 1993), (Jones et al., 1973), a combination of several methods should be used in order to get appropriate results. Another method, such as treating a sample with antibiotics in order to slow down the growth rate of contaminators and/or kill them completely, is probably the most effective.

That is why several methods were tested (See Chapter 2.6).

The centrifugation as a purifying step was not used as a single method of getting axenic cultures, however it was used during cell harvesting and between the subculturing of old cultures into new media (See Chapter 2.9.1).

The multiple dilution method using PCR 0.2 ml tubes in combination with plating was not efficient. In both groups of plates where the cultures were spread by using the disposable loop or the spatula, the evidence of contamination was found (the pale halo around the algae colonies) (See Figure 21.4). However it was found that the use of spatula can be better applied for the less concentrated samples, because a few colonies were better separated. While the use of the disposable loop can effectively separate the concentrated sample distributing the colonies on the whole surface of agar plate. The single colonies were separated in both cases, however the microscopy analysis using x100-150 magnification showed the presence of non-algal cells (See Figure 21.5 and 21.6).

Based on the literature, concerning antibiotic treatment of algae (Jones et al., 1973; Bradley et al., 1988), the rapid method of using antibiotic disks was chosen for this study. The following antibiotic disks were obtained from the MAST Group, UK: Amphotericin B 20 μg per disk, Imipenem 10 μg per disk, Neomycin 10 μg per disk, Nystatin 100 units per disk, Penicillin G 10 units per disk. Amphotericin and Nystanin were used for removing probable fungal contaminators, Imipenem, Penicillin G were used for removing bacteria and Neomicin was used against broad range of contaminators.

The antibiotic disks were placed on the surface of agar plates with growing algal colonies. It was supposed that a region free of contamination should appear around antibiotics disks. However, even after 1 month of growing, the colonies of contaminators were found alongside with algal colonies. The only exception was relatively low density of algal and contaminators colonies around the disk of Amphotericin B (AMB in Figure 21.2), but such situation was unique and has not been observed later. At the same time the presence of the antibiotic disks increased the lag-time, before the colonies were observed and slightly decreased their growth, without visible effect on the algal colonies, which grew even on the disk surface.

The next experiment involved subculturing of algae in the liquid media containing antibiotic disk. The algal cultures grew in the 1.5 ml Eppendorf tubes, each tube contained dipped antibiotic disk. After two weeks the cultures were checked on the presence of contamination by microscopy. The autofluorescence of Chlorophyll was also detected in order to control its possible degradation under antibiotic stress. The best result was shown

in the tube containing the Imipenem disk. Only a few contaminants were observed, while the pure culture and the other tubes contained a lot of contaminants. The chlorophyll degradation has not been observed. The next check was done after the month of culturing, it showed that the tube with Imipenem disk had lower number of contaminants, while tubes with Nystatine and Neomycin showed larger contamination than the control. The last check was done at the forty-fifth day of subculturing, and showed that the concentration of contaminants became higher and was equal among all the tubes. The degradation of Chlorophyll was detected in the tube contained imipenem disk.

At the same time another experiment was started. After one month of culturing in liquid media with antibiotic disks, the cells were placed into the wells of 96 well-plate and diluted (See Chapter 2.6.3).

The microscopic measurements after one month of subculturing showed that a very few number of contaminants was found in the last wells with diluted Neomicin, Nystatine and Imipenem cultures, while the control was saturated by contaminants. No changes in chlorophyll concentration were found.

The most promising antibiotics were Imipenem, Nystatin, Neomycin and Amphotericin B. However, the concentration of antibiotics was not enough to remove all contaminators. The disks decreased the growth rate of contaminators. Higher concentration has to be used in combination with dilution and plating methods for the efficient purification of *Nannochloropsis* cultures.

The nature of contaminators has not been investigated in this study. According to the positive results of the use of antibiotics against bacteria and fungi, it can be concluded that both groups were present.

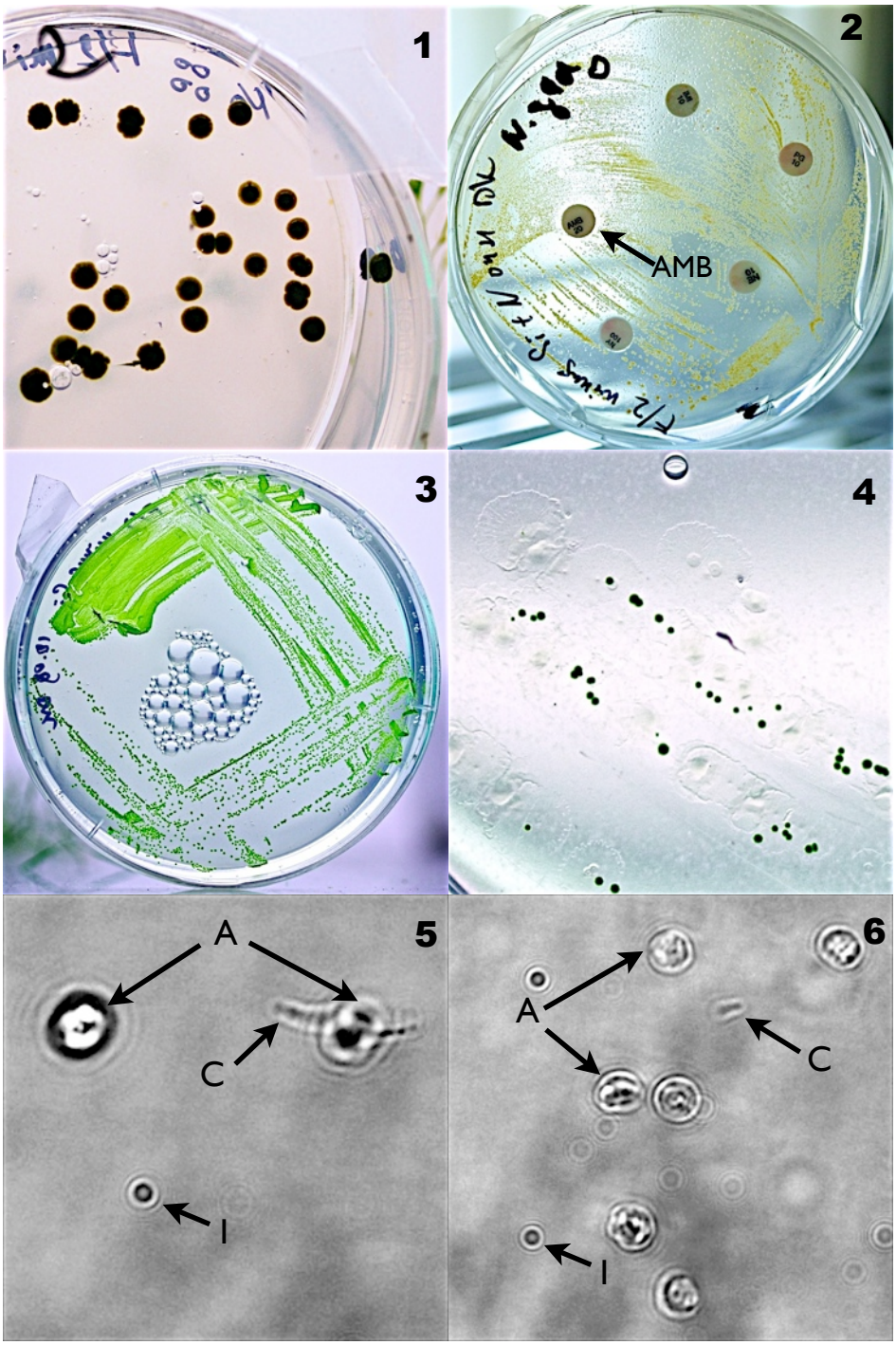


Figure 21. The results of contamination analysis and appliance of different purification methods for *N. gaditana* cultures. Figures 21.1 - 21.4 show results of plating the algal sample on F/2 agar plates. The cells were spread on the agar surface using spatula (21.1 and 21.2) or the disposable loop (21.3 and 21.4). The well separated colonies represented in the figure 21.1, were contaminated which was detected by microscopy. The Amphotericin B with probable contamination-free zone shown in the figure 21.2 (AMB). The colonies of contaminators growing together with algal one are shown in the figure 21.4. The figures 21.5 and 21.6 represented bright-field microscopic images made at x150 (figure 21.5) and x100 (figure 21.6) magnification, where A - algae cells, C - moving (bacterial) contaminators, I - probably inorganic contamination (dirt). If the objects like «C» (in figures 21.5 and 21.6) were present in a sample, it was considered as contaminated.

3.2.Lipid content

3.2.1.Neutral lipid staining by BODIPY 505/515

To visualize the presence and location of neutral lipids in *Nannochloropsis* cells, the fluorescent staining dye BODIPY 505/515 was preferred over the more traditional dye Nile Red because the former had been recently reported to show improved performance in microalgal lipid analyses (for comparison of the BODIPY and NileRed see Chapter 2.7 and Table 2). For instance, BODIPY 505/515 penetrates algal cell walls easily upon cell pretreatment with DMSO and binds to neutral lipids inside of the cells (Brennan et al., 2012). By fluorescence microscopy the molecule can be excited at one defined wavelength (for example at the maximum absorption 505 nm) and the emitted fluorescent light is detected at longer wavelength (max emission at 515 nm). The location and intensity of the source of the fluorescent light gives the information about the location and the amount of neutral lipids in algal cells.

For staining of neutral lipids 2 *μ*L of BODIPY 505/515 prepared in 100% DMSO were added to 98 ml of *Nannochloropsis* cell culture to get final concentration of 0.12 *μ*g/mL of BODIPY 505/515 and 2% of DMSO (See Chapter 2.5.3). The detection of fluorescence was performed on Nikon TE2000U inverted fluorescence microscope. The cell at mid-Log phase were used for analysis grown under standard and nitrogen depletion conditions..

To detect fluorescence of BODIPY 505/515 settings for the YFP detection were used. To detect autofluorescence of Chl a molecules, the dSRed filters were used, (see Chapter 2.7). The intensity of excited light was unstable and relatively low and quickly declined. Therefore, the required exposure time was usually more than 10 sec and sometimes even up to 30 sec to compensate the low light intensity. Hence, prerequisite of cell imaging at such long exposure times was that the microalgae did not move, which made imaging challenging.

3.2.2.Lipid accumulation under normal and nitrogen depletion stress conditions

Under standard growth conditions *Nannochloropsis* species usually contain up to 30-40% total lipids per dry biomass (Rodolfi et al., 2009; Gu et al., 2012). Applying stress conditions can increase lipid concentration up to 60-70% of dry mass (Hoffmann et al., 2010; Kilian et al., 2011). Stress factors like low nitrogen concentration in growing medium

or low light and temperature conditions trigger protective mechanisms in microalgae which enhance the production of lipids. *Nannochloropsis* algae produce mainly neutral lipids and store them in lipid bodies. It also produce very important EPA (omega-3 LC-PUFA) in high concentration (up to 35% of total fatty acids), accompanied by relatively high TAG biosynthesis (Khozin-Goldberg et al., 2011). The combination of this two components are recommended for human dietary intake (Hoffmann et al., 2010) (Capter 1.2.1).

In the present thesis lipid production in *Nannochloropsis* cells was induced by subculturing equal volumes of the algae cultures in the mid-log growth phase in new media with and without nitrogen source ($NaNO_3$) (For detailed medium composition see Chapter 2.3.1). After two days of growth under the identical environmental conditions 98 ml of each culture were subjected to BODIPY staining and microscopic analysis. The fluorescent structures were located close to the cells border, generally low in number, i.e. of only one single structure (most likely lipid body, Figures 22B, 22D, 22F, 22I) or a pair (Figures 22A, 22C, 22E) per algal cell.

After two days of microalgal grown under stress conditions the algae contained a significantly higher content of BODIPY-stained subcellular structures (lipids bodies), as indicated by both stronger fluorescence and larger structures, in relation to cell size, and these structures were more abundant among cell population compared to the cells grown under normal conditions. This result was obtained for species analyzed. The intensity of fluorescence from the *N. salina* cells grown in +N medium are shown in Figure (22D) was lower than for the cells of *N. salina* grown in -N medium (22E). Similar results were obtained for the cell of *N. oceanica* grown at standard conditions (+N) (see figure 22F), and nitrogen depleted medium (-N) in figure (22G). The fluorescence intensity detected from the *N. oculata* cells grown in normal medium was lower (figure 22H), than from the cells grown in absence of nitrogen source, one of which is represented in figure (22I). However, it remained difficult to conclude in which species the fluorescence of BODIPY 505/515 was more intense or abundant, because time constraints did not permit in-depth semi-quantitative analyses. However, preliminary data indicated that *N. gaditana* showed the maximum TAG accumulation in lipid bodies. As for the cells grown in nitrogen containing medium (+N) (figure 22A) and for the cells grown in the nitrogen depletion medium (-N) which are represented in figures (22B and 22C).

Changes in cell size or shape were not detected between cultures grown in the standard and the nitrogen depletion conditions. The average size of the cells was 2 μm, and the cell shape varied from oval in *N. salina* to round in the other species.

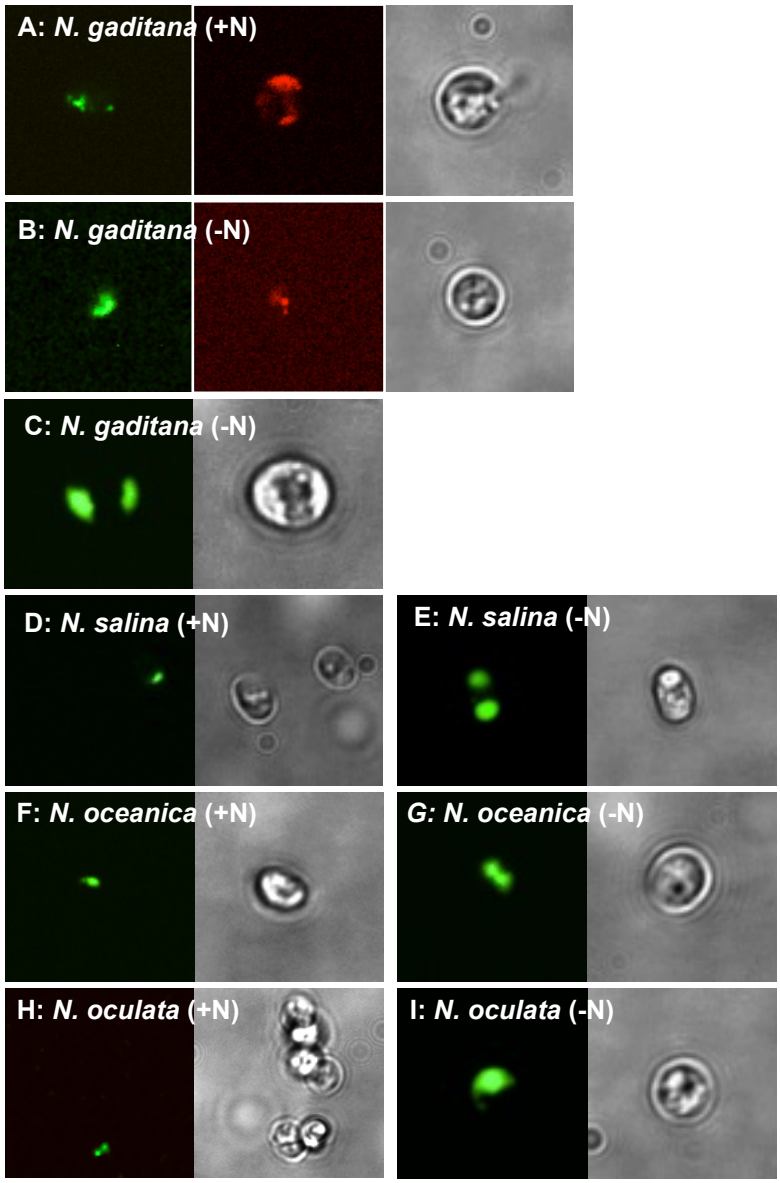


Figure 22. Fluorescence microscopy images of Nannochloropsis cells stained by BODIPY 505/515. The red color shows autofluorescence of chlorophyll a, the green color indicates fluorescence of BODIPY 505/515 and the bright field images were taken alongside with fluorescent ones. All pictures were taken by a Nikon TE2000 fluorescence microscope with x100 magnification, except, picture 22A, where the magnification was x150. The cells of *N. gaditana* which grew in medium with nitrogen source (+N) represented (22A) and without nitrogen containing compounds (-N) (22B and 22C). *N. salina* cells grew in +N and -N media are shown (22D and 22E), respectively. *N. oceanica* cells which were grown in +N and -N media are shown (22F and 22G), respectively, *N. oculata* cells grown in +N and -N media are represented at figures (22H, and 22I), respectively. The intensity of the fluorescence was higher in the cells which were grown in nitrogen depletion media.

3.3.Bioinformatic analyses of the predicted peroxisomal proteome of *Nannochloropsis*

Bioinformatics is a very powerful integration of computer based algorithms of calculation and visualization of biological data. The time when biology and chemistry were only descriptive studies are over. Nowadays, biology as all other scientific studies is no longer «homogeneous», but consists of some other branches and has become highly interdisciplinary, which is advantageous and needed in order to get, process and represent important information (Krawetz et al., 2003; Baxevanis et al., 2004). For example, it is simply impossible to work on massive genetic data without informatics tools.

In the present study bioinformatics was used to analyze proteomic and genomic information of *Nannochloropsis.* At time point when performing the present research, only two genomes of *Nannochloropsis* species were freely and conveniently available at Genbank of NCBI. These were *N. gaditana* CCMP527 and *N. oceanica* CCMP1779. Unfortunately, information about predicted protein sequences was available only for *N. gaditana* CCMP527, and no literature concerning peroxisomal proteins was available for *Nannochloropsis* in general. The significance of investigations of the peroxisomal proteome for the understanding and optimization of lipid metabolism, several steps of which occur in the peroxisomes, has been outlined in the introduction (see 1.2.1, 1.3).

The complete work-flow summarizing all the steps of prediction and analyzing of PTS1/2 proteome of N. gaditana is shown in Figure 23.

Fig. A

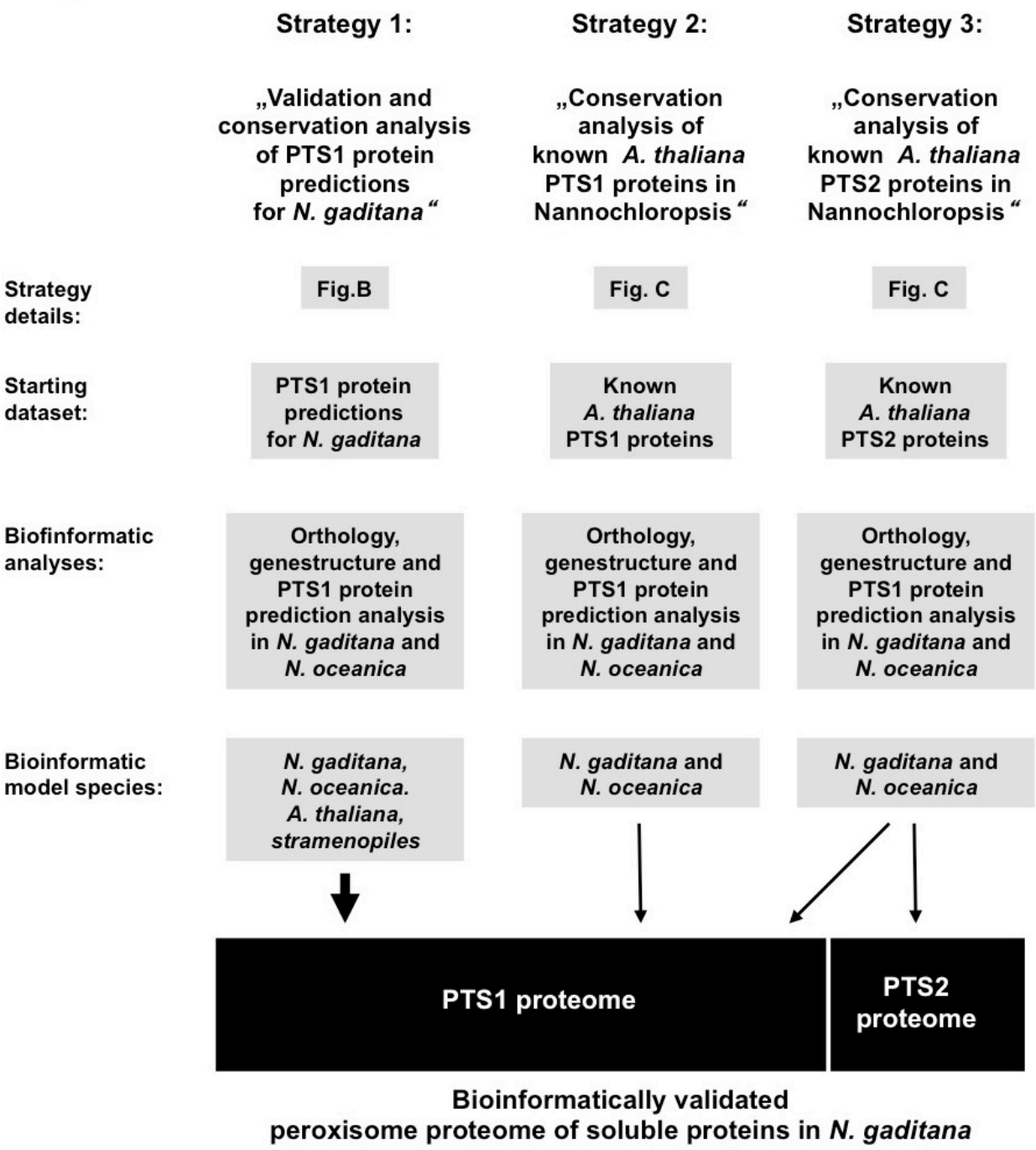
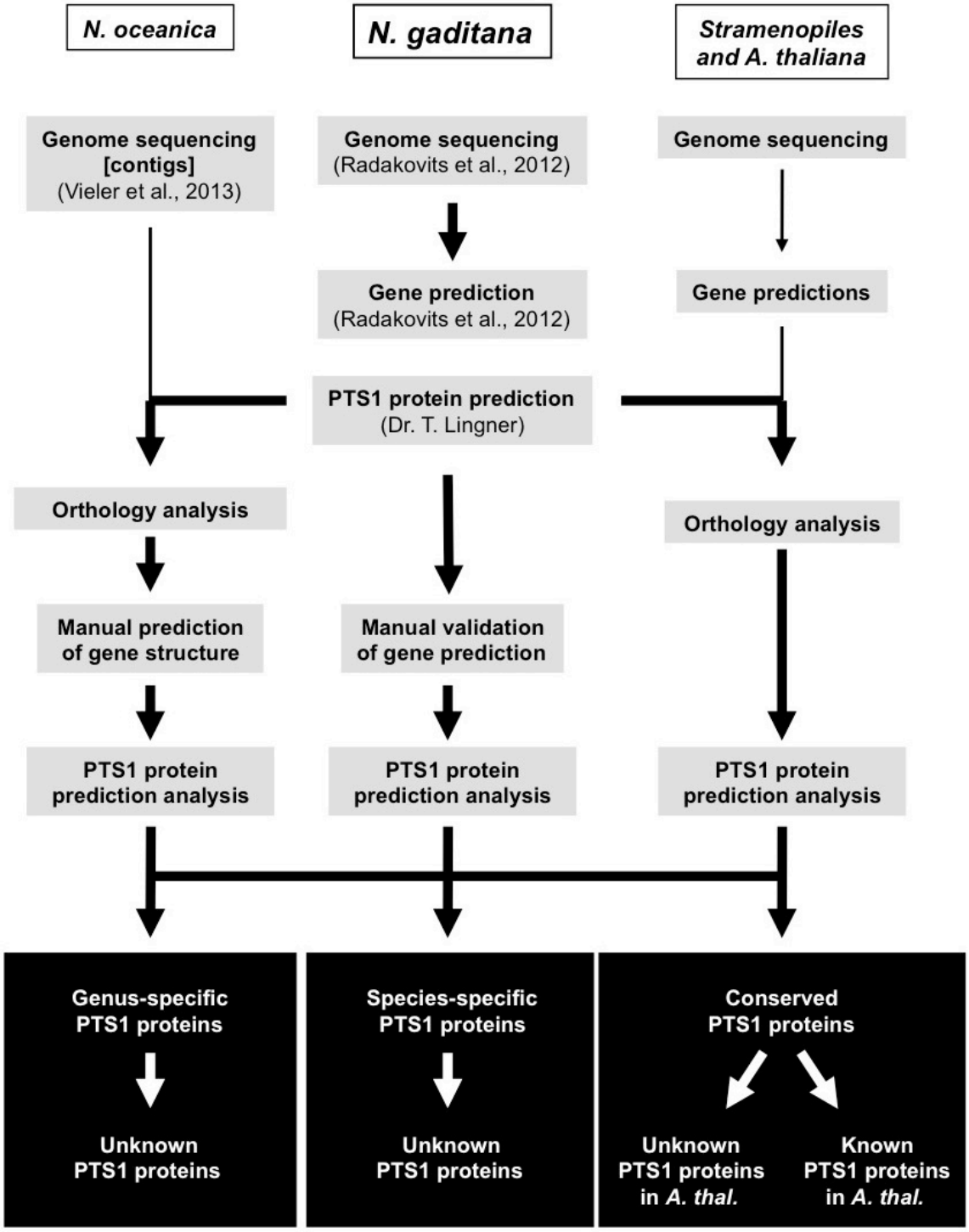

Strategy 1:
„Validation and conservation analysis of PTS1 protein predictions for N. gaditana“
Strategy 2:
„Conservation analysis of known A. thaliana PTS1 proteins in Nannochloropsis“
Strategy 3:
„Conservation analysis of known A. thaliana PTS2 proteins in Nannochloropsis“
Strategy details:
Fig.B
Fig. C
Fig. C
Starting dataset:
PTS1 protein predictions for N. gaditana
Known A. thaliana PTS1 proteins
Known A. thaliana PTS2 proteins
Biofinformatic analyses:
Orthology, genestructure and PTS1 protein prediction analysis in N. gaditana and N. oceanica
Orthology, genestructure and PTS1 protein prediction analysis in N. gaditana and N. oceanica
Orthology, genestructure and PTS1 protein prediction analysis in N. gaditana and N. oceanica
Bioinformatic model species:
N. gaditana, N. oceanica. A. thaliana, stramenopiles
N. gaditana and N. oceanica
N. gaditana and N. oceanica
PTS1 proteome
PTS2 proteome
Bioinformatically validated peroxisome proteome of soluble proteins in N. gaditana

Fig. B



Bioinformatically validated PTS1 proteome of *N. gaditana*

Fig. C

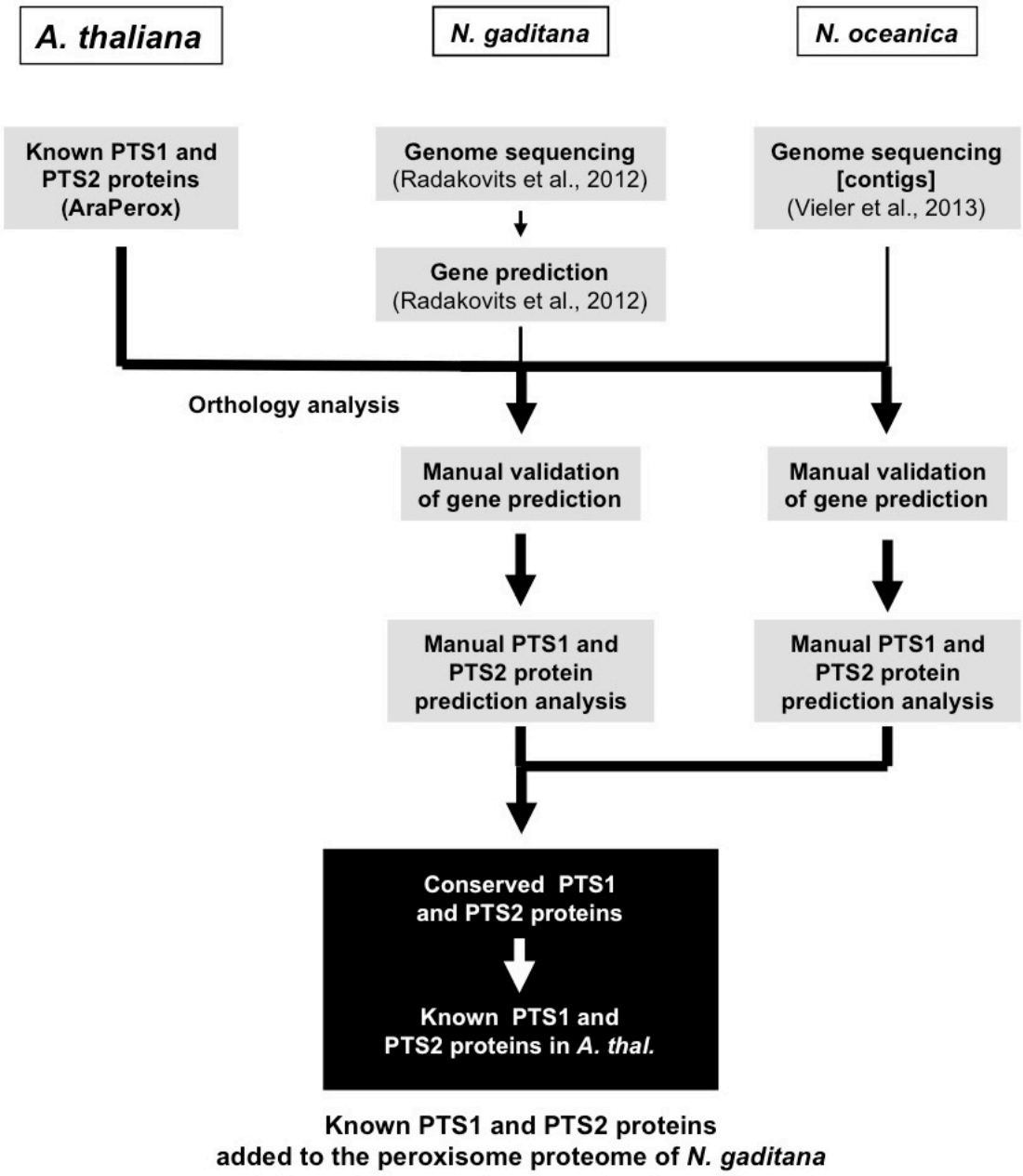


Figure 23. Work-flow of the identification peroxisomal proteome of *N. gaditana* and *N. oceanica*. In the figure 2323 summarizes all major steps of the PTS1/2 proteins predictions. There three strategies were used (figure 23A). In a strategy 1 (23A and 23B) the predictions of Dr. T. Lingner were used and verified by gene structure analysis, orthology analysis among proteins of *A. thaliana* and *Stramenopiles* (Results in Chapters 3.3.3 and 3.3.4). Additional PTS1 and PTS2 orthologues proteins were found by direct blasting of known *A. thaliana* proteins (strategies 2 and 3, figure 23A and 23C, results in Chapter 3.3.5). *Nannochloropsis* is a genus of the *Stramenopiles/ Heteroconta* phylum (See Chapter 1.1.1). In total 42 putative PTS1 *N. gaditana* proteins were found, among them 30 conserved, 12 genus-specific. The presence of two PTS2 proteins orthologues to *A. thaliana* PTS2 proteins was predicted.

3.3.1.Evaluation of three different PTS1 protein prediction models for *Nannochloropsis gaditana*

All predicted proteins of *N. gaditana* from the NGPS database (http://nannochloropsis.genomeprojectsolutions-databases.com/) were subjected to PTS1 protein prediction to investigate whether they contained a predicted PTS1 in the C-terminal end. Each protein was given a PWM prediction score and a posterior probability ranging from 0 to 1.000. They express the probability for protein to be targeted to peroxisomes. The posterior probabilities were ranged from 0 to 1.000, where 0 - 0% probability to have a PTS1, 1.000 means 100% probability to contain a PTS1, with 0.5 as a threshold (which is equal to 0.412 of PWM score) (Lingner et al., 2011).

Predicted PTS1 proteins were arranged into a hierarchical list sorted by PTS1 posterior protein prediction probability in such a way that proteins with the highest score were located at the top of the list and those with the lowest at the bottom. The lowest cutoff value was associated with the posterior prediction probability threshold equal to 0.5.

In total three lists of predicted proteins were made, depending on which prediction model was applied. The «Plant» PWM model is identical to that published previously (Lingner et al., 2011), while the «Metazoa» and the «Combined» model had not been published (T. Lingner and S. Reumann, unpubl. data). The «Plant» PTS1 model that had been developed first (Lingner et al., 2011) had been built by using the homology search of *Arabidopsis* PTS1 proteins in plant proteins and ESTs. The «Metazoa» model was based on a model that had been made analogously to the plant model using the homology search of human PTS1 proteins in ESTs and metazoan proteins, while the «Combined» model reflected a model that had been built using (positive and negative) sequences from animals, fungi and plants. The original model development and predictions to *N. gaditana* and organization of the results were made by Dr. T. Lingner (Institute of Microbiology and Genetics, University of Göttingen, Germany).

As a part of this thesis a strategy was developed to compare and combine the three PTS1 protein prediction models («Plant», «Metazoa» and «Combined») for *Nannochloropsis*. First, the «Plant» PTS1 prediction model was used as a standard, based on the assumption that this model would give the largest number of PTS1 proteins and with highest prediction accuracy. Next, each Nannochloropsis protein predicted to possess a PTS1 by the «Plant» model prediction was searched in the other lists e.g. «Metazoan» and «Combined», for similar PTS1 protein prediction. Finally, all proteins predicted

consistently by in the «Plant» model and simultaneously one or both of the others were grouped together and their number determined (Figure 24).

Seventy-eight *N. gaditana* proteins were predicted as PTS1 proteins above the threshold for the «Plant» model. Thirty-one (out of 51) and 63 (out of 89) predicted PTS1 proteins from the «Metazoa» and the «Combined» model, respectively, overlapped with the «Plant» model (and thirty-one with all three models), thereby further increasing the probability that these predicted PTS1 proteins are indeed true peroxisomal proteins in *Nannochloropsis gaditana*. It was concluded that «Metazoa» model is the least suitable for *Nannochloropsis* predictions and that the «Plant» model predicts many more PTS1 proteins (For the full list of proteins predicted by the different models see Table B in Appendix 2).

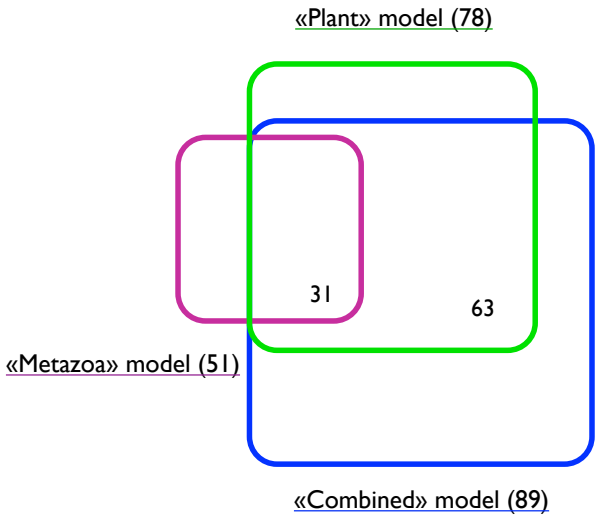


Figure 24. Number of peroxisomal proteins predicted by the «Plant», and found in the «Metazoa» and the «Combined» models. Numbers inside of the figures represent amount of the proteins predicted by «Plant» model and found in the «Combined» model (63 of 89) and in the «Metazoa» model (31 of 51). The total number of predicted proteins using different models is 78, 89 and 51 for the «Plant», «Combined» and «Metazoa», respectively

3.3.2. Prediction and bioinformatic validation of the PTS1 proteome of Nannochloropsis

The prediction of PTS1 *N. gaditana* proteins by the «Plant» prediction model were verified by orthology analysis among proteins of several species of *Stramenopiles* and *A. thaliana.* The N. gaditana proteins were blasted against proteins of *Stramenopiles* and *A. thaliana,* the presence of PTS1 in the C-terminus end of the found proteins was predicted using PredPlantPTS1 server (for method See Chapter 2.5.2). Those proteins which had putative PTS1 orthologs among *Stramenopiles* and *A. thaliana* were considered as conserved among *Stramenopiles* and *A. thaliana* with high probability, those which had only one PTS1 ortholog were annotated as conserved peroxisomal proteins among *Stramenopiles* and *A. thaliana* with moderate probability. The proteins without PTS1 orthologs among *Stramenopiles* and *A. thaliana* were described as peroxisomal proteins with low probability

to be conserved among *Stramenopiles* and *A. thaliana* (detailed methodology used is shown in Figure 23).

Table 10 shows representative examples of three categories on the overall probability of *Nannochloropsis gaditana* proteins to be peroxisome-targeted by the PTS1 pathway based on the results of comparison of *N. gaditana* protein information with ortholog information.

For instance, for the protein Nga03053 annotated as peroxisomal acyl coenzyme A oxidase 1/5 (ACX 1/5) and predicted to be peroxisome-targeted by the PTS1 tripeptide ARL, four putative orthologs were found for *Stramenopiles* and *A. thaliana* in the protein database (Genbank at NCBI) . All four proteins had similar annotations (which predict their functions), were of similar size (about 700±50 amino acid), and shared significant sequence identity at the amino acid level (≥34%). These four putative orthologs were analyzed for the presence of PTS1 domains using the PredPlantPTS1 prediction servicer (http://ppp.gobics.de/) (Lingner et al., 2011; Reumann et al., 2012). All four putative orthologs were shown to contain predicted PTS1 domain, thereby increasing overall the probability significantly that the *N. gaditana* protein Nga03053 was indeed correctly predicted as a PTS1 protein.

Another protein Nga30150 was annotated as «mitochondrial carrier domain containing protein» for which three putative orthologs were found, which were annotated with similar keywords such as «mitochondrial» and «carrier», had similar sizes (about 300±15 amino acid), varying degrees of sequence identity (31% for *A. thaliana* and more than 60% for *Stramenopiles*). However, none of these proteins contained a predicted PTS1 domain, as indicated by very negative PWM values. Therefore, it was overall concluded that the protein Nga30150 had overall a low probability to be peroxisomal (see Table 6). The proteins that had only one putative ortholog with predicted PTS1, were considered as proteins with moderate peroxisomal localization probability (Table 10, protein L-asparaginase Nga00060, SRL>).

Table 10. Representative examples of orthology and PTS1 prediction analysis of putative *N. gaditana* PTS1 proteins. Blastp search was used to find orthologous proteins for the *N. gaditana* protein query. Sequence name, predicted GO-term and size of the *N. gaditana* protein are presented in the first three columns. Information about organism «Org.», in which homologs were found, their annotation, size, identity, presence of PTS1 predicted by PredPlantPTS1 with PWM score of prediction (values from 0 to 1, where 0 - not peroxisomal, 1 means 100% peroxisomal, treshold 0.412), C-terminal tripeptide. The conclusion about probability of being peroxisomal for the *N. gaditana* proteins was done based on total similarity information of ortholgs to the query and is

shown in the last column. Organisms acronyms: A*rabidopsis thaliana (*At*)*, *Aureococcus anophagefferens (Aa)*, *Phaeodatylum tricornutum (Pt)*, *Ectocarpus siliculosus (Es)*, *Thalassiosira pseudonana (Ts)*, *Thalassiosira oceanica (To)*.

***Nannochloropsis* proteins**			**Homologs to *Nannochloropsis* protein query**							**Conclusion**
Seq.Name	GO term	Size,aa	Org.	Annotations	Size,aa	ID (%)	C-trip.	PTS1	PWM	
Nga03053	**peroxisomal acyl-coenzyme a oxidase 1**	**751**	At	peroxisomal acyl-CoA oxidase	664	36	**AKL**	yes	0.80	High probability of being a true peroxisomal protein
			Pt	precursor of acyl-coa oxidase	706	36	**SRL**	yes	0.86	
			Tp	acyl-coa oxidase	685	35	DPL	no	-0.55	
			To	acyl-coenzyme A oxidase	761	34	**SRL**	yes	0.64	
Nga00060	**L-asparaginase**	**366**	Aa	put. Asparaginase	456	49	VRA>	no	-0.57	Moderate probability of being a true peroxisomal protein
			Es	Asparaginase	446	49	**SKL>**	**yes**	**0.84**	
			Tp	L-Asparaginase I	355	48	HYY>	no	-0.98	
			Pt	(none)	332	48	PQE>	no	-0.88	
			At	no put. orthologs						
Nga30150	**mitochondrial carrier protein**	**315**	Es	Mitochondrial carrier protein	320	65	GVL	no	-0.41	Low probability of being a true peroxisomal protein
			Tp	Mitochondrial carrier protein	292	61	PKK	no	-0.14	
			At	S-adenosylmethionine carrier	345	31	HNA	no	-1.11	

Applying these orthology and PTS1 protein analyses systematically to the predicted PTS1 protein of N. gaditana, out of 78 PTS1 predicted proteins by «Plant» prediction model 28 proteins of *N. gaditana* were predicted to be peroxisomal with high (23 proteins) or moderate probability (5 proteins) based on the orthology analysis. Among these proteins with high overall PTS1 protein prediction probability, two proteins were new/unknown proteins of very intriguing predicted function: an alpha/beta-hydrolase domain-containing protein (Nga02873, SRL>) and an NAD^+ kinase/Prenylcysteine methylesterase (Nga01349,SRL>). The proteins with moderate probability were the following: L-asparaginase Nga00060, SRL>, peroxiredoxine Nga01942, SRL>, Nga03561, Glucose-6-phosphate-1-dehydrogenase Nga00833, AKI>, Photosystem II oxigen-envolving protein Nga05405, PKL>, ATP phosphoribosyltransferase Nga03561, SRM> (see Table 13).

The genome of *N. oceanica* was recently been sequenced (Vieler et al., 2012) and whole genome contigs were available at GeneBank of NCBI. However, no gene or protein predictions were publicly available. Therefore, a manual method to find orthologs of *N. gaditana* predicted proteins in *N. oceanica* was established. *Nannochloropsis oceanica* CCMP 1779 whole genome contigs collection was blasted on presence of orthologs to *N. gaditana* proteins. In total 78 *N. gaditana* proteins from the results of «Plant» PTS1 prediction model were blasted against N. oceanica CCMP 1779 whole genome contigs, using tblastn (NCBI). Last 14 amino acids were analyzed in PredPlantPTS1 on the presence of PTS1. 27 PTS1 proteins were conserved among *Nannochloropsis*, *Stramenopiles* and *A. thaliana* species (See Table 13).

3.3.3.Verification of predictions

Verification of putative peroxisomal of *N. gaditana* proteins orthologous to the PTS1 proteins of *A. thaliana* and conserved among *Stramenopiles* was supplemented by a gene structure analysis. Such analysis was made for all the predicted 28 PTS1 proteins which got positive results to be peroxisomal from the orthology search. However, the remaining 50 proteins were analyzed as well in order to avoid the possibility of losing probable organism-specific proteins, which did not have peroxisomal orthologs outside of the genus *Nannochloropsis*.

The methodology included blasting *N. gaditana* proteins as query against the translated nucleotide collection (whole genome contigs) of *N. gaditana* and checking for the presence of a PTS1 at the C-terminal end of the manually translated sequence. This was performed by tblastn search for each protein (see Chapter 2.8.2).

All twenty-eight proteins orthologous to *Stramenopiles* and *A. thaliana* PTS1 proteins were verified to contain PTS1s using gene structure analysis. Twenty-six orthologs of these proteins were predicted to have a PTS1 in *N. oceanica*. The C-terminus of «protein #6» (PVQ>) was not found in the genomic data. Possibly the protein has a PTS1 tripeptide in the end, because the sequence identity with the *N. gaditana* homolog was very good. N. oceanica's «Protein #8»(RRL>) showed moderate identity with the *N. gaditana* homolog and did not carry a PTS1 at the C-terminus. The length of the *N. gaditana* protein was quite short, perhaps the exon structure or length of this protein even in *N. gaditana* was wrongly predicted.

Analyzing 50 proteins from the group with low probability to be peroxisomal among orthologs, twelve *N. gaditana* proteins (nine for putative *N. oceanica* orthologs) that were assigned prediction scores above threshold were verified to have PTS1. Four of them had only one exon (seven for *N. oceanica*). Two predicted *N. gaditana* proteins completely lacked any annotation (Nga00291, SRL> and Nga20149, CRM>).

Two additional proteins for both species of *Nannochloropsis* were identified. Protein 41 (ARL>) by direct blasting the known PTS1 multifunctional protein AIM1 from *A. thaliana* against whole genome contigs of *Nannochloropsis* species. It was manually translated and PTS1 predicted using the PredPlantPTS1. The protein 42 (AHL>) of *N. gaditana* orthologous to the known *A. thaliana* PTS2 protein Aspartate aminotransferase (ASP3) was found during PTS2 protein prediction (see Chapter 3.3.4), and verification was done in

the same way as for the protein #41. In *N. oceanica* genome both proteins are present, but the protein #42 (PHL>) was not predicted to carry PTS1 by PredPlantPTS1, while Protein 41 has the same ARL tripeptide in the N-terminal end.

In total 40 proteins from *N. gaditana* were verified to have a PTS1 by the gene structure analysis method. This number constitutes 51% of the total proteins predicted by the «Plant» prediction model. 35 putative peroxisomal orthologs to N. gaditana PTS1 proteins were found in the *N. oceanica* genome.

3.3.4.Investigation exons and introns in *N. gaditana* genes of interest for subsequent gene/C-terminal exon cloning

Most of the eukaryotic genes have exons and introns.

The final protein sequence does not have any introns and is functionally active only when it is entirely complete. Using predicted protein sequence as a template in the collection of nucleotide sequences it was possible to find those parts of genes which were considered as exons and were present in protein. It can be investigated by looking at the results of blasting. If there are several parts of identity present, and they have different open reading frames, it can already be assumed that this analyzed protein has several exons. Nucleotide sequence which corresponds to the parts of identity was looked at, taking all nucleotides which are between N- and C-terminus of corresponded protein plus roughly 500 bp at each side, was translated, using Internet-based translation tool: ExPASy (http://web.expasy.org/translate/). In the results of translation different open reading frames (orf) were looked at and it was tried to find the regions identical to given protein sequence (See Chapter 2.8.2.1).

13 predicted proteins of *N. gaditana* were encoded by single exon genes, which is 32% of total amount of predicted proteins. 14 predicted proteins of *N. oceanica* were encoded by single exon gene and they represent 40% of all the predicted proteins. Two species share nine orthologues proteins encoded by single exon genes. However, the structure of 61% of the peroxisomal proteins predicted by the «Plant» model was wrongly predicted. This is a reason why not all of the 78 proteins were verified to carry PTS1.

Possible correlations between the predicted protein sequence and the manually translated one from the related nucleotide sequence are represented in Figure 25.

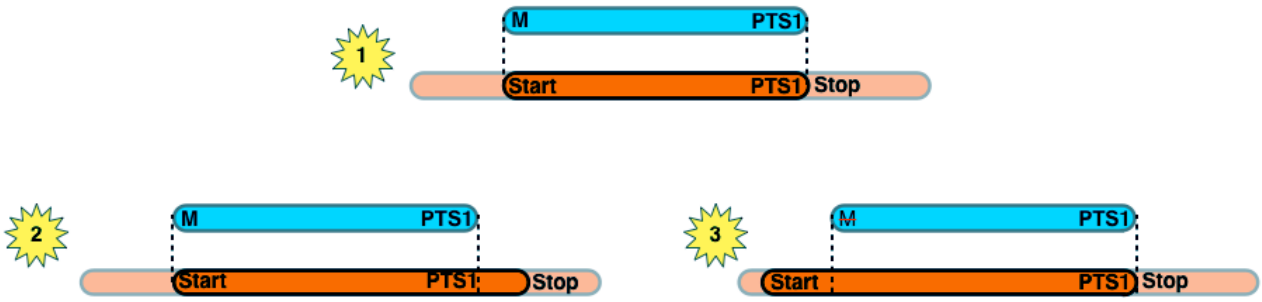


Figure 25. Correlation between predicted protein sequences and manually translated nucleotide sequence. Blue boxes - predicted amino acid protein sequence, saturated orange boxes - nucleotide sequence of the gene related to the protein, pale orange - non coding regions of nucleotide sequence, Start - start codon, M-methionine, M - any amino acid, PTS1 - peroxisomal targeting tripeptide type 1, Stop - stop codon.

Three possible cases represented in Figure 25 can be explained:

1. Predicted protein sequence is completely identical to manually translated nucleotide sequence in size.
2. Predicted protein sequence is mostly identical to the manually translated nucleotide sequence in size, excepting that the stop signal is located not directly after PTS1 (protein is actually longer at C-terminus).
3. Predicted protein sequence is mostly identical to the manually translated nucleotide sequence in size, excepting that the start signal of from the gene is not present in predicted protein sequence (protein is actually shorter at N-terminus).

3.3.5.PTS2 prediction

In total 14 known *A. thaliana* proteins from AraPerox v1.2 (http://www3.uis.no/araperoxv1/) were analyzed using the newly established method described in Chapter 2.8.3. According to it, the *A. thaliana* protein sequences were blasted against *Nannochloropsis* whole genome contigs, then putative N-terminal end was estimated and investigated on the presence of followed amino acid combination [R/K]-[L/V/I]-x_5-[H/Q]-[L/A] (Lanyon-Hogg et al., 2010) of PTS2. In additional all proteins were analyzed on the presence of PTS1 in the C-terminal end as described in the Chapter 2.8.2.1.

For all analyzed *A. thaliana* proteins were found putative homologs (Table 11). However seven proteins of *N. gaditana* (six - for *N. oceanica*) lacked any PTS. Four homologs

lacked PTS2 in *N. gaditana*, but had PTS1. They are orthologs to ASP - protein family, ACX - protein family and CSY - protein family and LACS6 protein. For *N. oceanica* three orthologs to the same *A. thaliana* proteins were found, except ASP3, the which ortolog did not contain predicted PTS1.

Table 11. Comparison of known PTS2 proteins from *A. thaliana* with predicted peroxisomal proteins from *N. gaditana* and *N. oceanica*. The *A. thaliana* proteins were used as query sequences in tblastn search. There are tree types of accession codes for *N. gaditana*: $Ngax_5$ - for predicted proteins from NgGP, $AGNI/AEUMx_8$ - for whole genome contigs at GeneBank of NCBI. Presence of PTS1 for homologs predicted by PredPlantPTS1, PTS2 by manual search, both signals are shown in the table, if present.

A. thaliana		*N. gaditana*				*N. oceanica*			
acc.number	Acronym	acc.code	Signal	PTS1	PTS2	Acc.code	Signal	PTS1	PTS2
AT5G65110	ACX2	Nga04370, Nga03053	ARL	yes	no	AEUM01001941	ARL	yes	no
AT1G06290.1	ACX3		ARL	yes	no		ARL	yes	no
AT1G06310.1	ACX6		ARL	yes	no		ARL	yes	no
AT3G58740.1	CSY1	Nga05548	ARL	yes	no	AEUM01002396	ARL	yes	no
AT3G58750.1	CSY2		ARL	yes	no		ARL	yes	no
AT2G42790.1	CSY3		ARL	yes	no		ARL	yes	no
AT5G11520.1	ASP3	AGNI01000127.1	AHL	yes	no			no	no
AT3G05970.1	LACS6	Nga00919	ARL	yes	no	AEUM01005249	SRL	yes	no
XP_002865685.1	KAT5/PKT1/2	Nga01710, Nga06820	RL_X5_HL	no	yes	AEUM01004300	RL_X5_HL	no	yes
AT2G33150.1	PKT3			no	yes			no	yes
AT1G04710.1	PKT4			no	yes			no	yes
AT2G22780.1	PMDH1			no	no	AEUM01003686	RL_X5_HA	no	yes
AT5G09660	PMDH2			no	no			no	yes
AT5G58220	TTL	AGNI01005008	RL_X5_HL	no	yes	AEUM01005008	RL_X5_HL	no	yes
AT1G50510.1	INDA			no	no			no	no
AT1G60550.1	ECHID			no	no			no	no
AT5G48545.1	histidine triad			no	no			no	no
AT3G56490.1	HIT3			no	no			no	no
AT1G06460.1	ACD32.1			no	no			no	no

Three proteins of *N. gaditana* orthologues to two protein families were predicted to have PTS2: Nga01710, Nga06820 - similar to PKT/KAT-protein family (both had <RLx_5HL>) and AGNI01005008 (<RLx_5HL>) which is similar to the second domain of TTL-protein (acc. number AT5G58220). The proteins Nga01710, Nga06820 are paralogs of the same protein. In the same time thee proteins of *N. oceanica* orthologues to three protein families were predicted to have PTS2: AEUM01004300 (<RLx_5HL>) - orthologues to PKT/KAT-protein family, AEUM01005008 (<RLx_5HL>) orthologues to the second domain of TTL-protein (acc. number AT5G58220) and AEUM01003686 which is putative ortholog to the PMDH-protein family (PMDH1- AT2G22780 and PMDH2 - AT5G09660).

In total three conserved PTS2 proteins orthologues to the PKT/KAT protein family and TTL-protein from *A. thaliana* were found in *Nannochloropsis* genus.

This data shows that PTS2 pathway is present in *Nannochloropsis* microalgae.

These results were compared with the data available for red, green algae, higher plants, diatoms and yeast based on the study of A. Shinozaki et al. (2009). The data for *Nannochloropsis* genus represented in Table 12. was predicted in this study and was not verified by localization studies. The data for the other species was collected by the authors of the article. If the presence of the protein was verified, then the type of PTS together with the corresponding amino acid peptide was represented in the table. In case of not peroxisomal localization of homologues protein, the «-» was used to indicate this.

Table 12. Peroxisomal targeting signal in algae, plant, and yeast (based on (Shinozaki et al., 2009)). Comparison of some known peroxisomal proteins from Green algae, Red algae, Diatoms, Plants, Yeast and predicted *Nannochloropsis* genus proteins. If protein is present, the signal and its type is present, if not «-», if no data available - «no data» is written. *Nannochloropsis* proteins were predicted in this study.

Peroxisomal enzymes	Organism					
	Green algae	Plant	*Nannochloropsis*	Red algae	Diatoms	Yeast
Malate synthase (MS)	SRM (PTS1-like)	SRL (PTS1)	SRL (PTS1)	–	AKL (PTS1-like)	SKL (PTS1)
Isocitrate lyase (ICL)	–	SRM (PTS1)	SRL (PTS1)	No data	–	–
Hydroxypyruvate reductase (HPR)	–	SKL (PTS1)	–	SRL (PTS1-like)	–	no data
Malate dehydrogenase (MDH)	RI_X5_HL (PTS2-like)	RI_X5_HL (PTS2)	RL_X5_HA (PTS2)	No data	–	SKL (PTS1)
Citrate synthase (CS)	–	RL_X5_HL (PTS2)	ARL (PTS1)	SRL (PTS1-like)	–	SKL (PTS1)
Thiolase (THIO)	–	RQ_X5_HL (PTS2)	RL_X5_HL (PTS2)	–	–	RL_X5_HL (PTS2)
Glycolate oxidase (GO)	CKL (PTS-1like)	PRM (PTS1)	SRL, SKI (PTS1)	SKL (PTS-1like)	SRL (PTS1-like)	no data

Malate dehydrogenase PMDH is present as PTS2 protein in Green algae, Higher plants and Nannochloropsis, The thiolase PKT/KAT is present as PTS2 only in Higher plants, yeast and Nannochloropsis. This data shows that the PTS2 pathway have the highest similarity with the one in Higher plants. At the same time glycolate oxidase seams to be very conserved among these groups.

Table 13 represents all predicted peroxisomal proteins for two species *N. gaditana* and *N. oceanica*. First group of proteins was taken from the results of prediction of orthologs to the *A. thaliana* PTS1 proteins (Chapter 3.3.3), next group of proteins was predicted by verification of gene structure predictions (Chapter 3.3.4) and two proteins are represented in in the third group. The protein 41 (AGNI010009, ARL>) was predicted by direct blasting of *Arabidopsis* AIM1 protein (At4g29010) sequence against *Nannochloropsis* whole genome contigs (tblastn). The peptide №42 was found during the PTS2 prediction (Chapter 3.3.4). *N. gaditana* proteins which protein sequences are available at NGPS (http://nannochloropsis.genomeprojectsolutions-databases.com/) have a name in format $Ngax_5$, while the proteins which were predicted by gene structure analyzation have the names related to the accession code of the GeneBank at NCBI, where they can be found. The annotation was obtained from Dr. T. Lingner which made GO-term search among

othologs and was additionally verified and completed by the integration of annotations from the *Stramenopiles* orthologs.

The proteins in the table are sorted by decreasing PTS1 probability predicted by PredPlantPTS1. The proteins marked with **bold** font were selected for the followed localization studies (See Chapter 3.4). *N. gaditana* proteins marked with *italic* font were probably wrongly predicted or they represent species-specific proteins, because of insufficient annotation and moderate/low identity with *N. oceanica* homologs.

Table 13. Predicted PTS1 proteins of *N. gaditana* and *N. oceanica*. Putative *N. gaditana* PTS1 proteins were used as query to identify orthologues genes in *N. oceanica* by tBLASTn. The N- and C-terminal ends of the *N. oceanica* genes were estimated by manual translation and sequence computing by dual sequence alignment with *N. gaditana* ortholog. PTS1 prediction analysis was performed for the manually predicted *N. oceanica* genes using PredPlantPTS1. For each protein of *N. gaditana* sequence name from NGP (http://nannochloropsis.genomeprojectsolutions-databases.com/) are available, except, independently found proteins №41 and №42. These proteins and *N. oceanica* proteins have only accession code for the genomic contigs where they where found. Annotations were obtained from machine GO-term search, missing annotations were completed manually. Last 3 amino acid of the C-terminal end of predicted protein present as well. Proteins marked in with **bold** were selected for *in vivo* validation of predictions. Proteins marked in with *italic* probably wrongly predicted (see text for details) For decoding of abbreviations see Appendix Table D.

	Nannochloropsis gaditana CCMP 526			*Nannochloropsis oceanica* CCMP 1779		
	Seq. Name	**Last 3 aa**	**Annotation**	**acc.number**	**Last 3 a**	**Peroxisoma**
Proteins conserved among *Stramenopiles* and *A. thaliana*						
1	**Nga05502**	**SRL**	**DECR/SDRb**	gb|AEUM01003195.1	SKL	yes
2	Nga04682	AKL	GR1	gb|AEUM01000955.1	AKL	yes
3	**Nga02873**	**SRL**	**α/β-hydrolase domain-containing protein**	gb|AEUM01002165	ARL	yes
4	Nga04370	ARL	ACX1/5	gb|AEUM01004509.1	ARL	yes
5	**Nga00060**	**SRL**	**ASP**	gb|AEUM01002974	AKL	yes
6	Nga06726	SRL	MS	gb|AEUM01001952	PVQ	no*
7	Nga02565	SKL	ECH2	gb|AEUM01003526	SKL	yes
8	Nga01349	SRL	NAD+ kinase/Prenylcysteine methylesterase	gb|AEUM01002529	RRL	no
9	Nga00919	ARL	LACS7	gb|AEUM01005249	SRL	yes
10	Nga01393	SRL	GOX	gb|AEUM01003813	SKL	yes
11	Nga00886	SRL	ICL	gb|AEUM01004791	SRL	yes
12	Nga01291	SRL	IBR3	gb|AEUM01002061	SRL	yes
13	**Nga00224**	**ARL**	**GGT1/2**	gb|AEUM01002696	ARL	yes
14	**Nga02369**	**SKL**	**DCI**	gb|AEUM01002521	SKL	yes
15	Nga02787	ARL	SCP2	gb|AEUM01001347	ARM	yes
16	Nga03561	SRM	ATP-PRT2	gb|AEUM01002601	SRM	yes
17	Nga01942	SRL	PRX5	gb|AEUM01000142	SRL	yes
18	Nga01381	SKL	NS	gb|AEUM01004442	SRL	yes
19	Nga05548	ARL	CSY1/2/3	gb|AEUM01002396	ARL	yes
20	Nga02546	SRL	TTL	gb|AEUM01003526	SRL	yes
21	**Nga00082**	**SKI**	**GOX2**	gb|AEUM01001952	SKI	yes
22	Nga04020	PRL	SDRa or NADH	gb|AEUM01000477/g	SKL	yes
23	Nga06551	AKL	AAE5	gb|AEUM01002605	AKL	yes
24	Nga06128	ARL	PTER/DECR/SDRa	gb|AEUM01004610	AKL	yes
25	Nga02557	AKL	BADH2	gb|AEUM01003526	AKL	yes
26	Nga00833	AKI	G6PD	gb|AEUM01001967	AKI	yes
27	Nga05405	PKL	Photosystem ii oxygen evolving complex	gb|AEUM01002967	PKL	yes
28	**Nga03053**	**ARL**	**ACX1/5**	gb|AEUM01001941	ARL	yes
***Nannochloropsis* genus specific proteins**						
29	**Nga01077**	**SKL**	**carbohydrate esterase/prenylcysteine me**	gb|AEUM01004771	ARL	yes
30	**Nga00170**	**SKL**	**peroxiredoxin-like PRX/TioTRX**	gb|AEUM01002948	SKL	yes
31	Nga30059	SRL	GLYI8	gb|AEUM01001580	SRL	yes
32	Nga00291	SRL	---*NA*---	gb|AEUM01003836	SKL	yes
33	Nga04192	ARL	α/β-hydrolase domain-containing protein	gb|AEUM01004824	AKL	yes
34	**Nga00077**	**PKL**	**CM-PDT-PDH**	gb|AEUM01001952	PKL	yes
35	Nga06021	SSL	gamma-butyrobetaine dioxygenase	gb|AEUM01002684	ASL	yes
36	Nga03919	SKI	YCII-related domain protein	gb|AEUM01003173	SKL	yes
37	*Nga40012*	*SLL*	*50s ribosomal protein*	gb|AEUM01000225	SIL	yes
38	*Nga01170*	*ALL*	*MtN3-like protein*	gb|AEUM01001403	ALL	no
39	*Nga05974*	*DRL*	*serine carboxypeptidase like protein*	gb|AEUM01002908	GRL	no
40	*Nga20149*	*CRM*	---*NA*---	gb|AEUM01002985	CRI	no
Manually predicted conserved proteins						
41	AGNI010009	ARL	AIM1	gb|AEUM01004762.1	ARL	yes
42	AGNI010001	AHL	ASP	gb|AEUM01004243	PHL	no

3.3.6.Metabolic pathways analysis of predicted peroxisomal *Nannochloropsis* proteins

It is well known that peroxisomes play a very important role in several metabolic processes, such as ROS metabolism, fatty acid beta-oxidation, etc. (Chapter 1.3).

45 peroxisomal proteins of *N. gaditana* were analyzed during PTS1 and PTS2 predictions (the work flow is represented in figure 23, the results in Table 7 and 9). Based on the annotation of the orthologues *A. thaliana* proteins (more information present) the proteins were classified according to the metabolic pathway groups.

I. Lipid biosynthesis:

1. Fatty acid oxidation: LACS (Nga00919, ARL>), ACX-family (Nga04370, ARL>, Nga 03053, ARL>), AIM1/EHHADH (Protein 41 «AGNI01009», ARL>), PKT/KAT (Nga01710 Nga06820, RLx_5HL), SCP2 (Nga02787, ARL);

2. Unsaturated fatty acid biosynthesis: DECR (Nga05502, SRL>, Nga06128, ARL>), DCI (Nga02369, SKL>), ECH2 (Nga02565, SKL), IBR3 (Nga01291, SRL), SDRa/b (Nga06128, ARL>, Nga04020, PRL>), AAE5 (Nga06551, AKL>);

II. ROS metabolism: GR (Nga04682, AKL>), PRX (Nga01941, SRL>, Nga00170, SKL>)

III. Glyoxylate cycle: CS (Nga05548,ARL>), ICL (Nga00886, SRL>), MS (Nga06726, SRL>), MDH (Protein 42 *N. oceanica* «AEUM010003686», RLx_5HA)

IV. Photorespiration: GOX (Nga01393, SRL>, Nga00082, SKI>), GGT (Nga00224, ARL>)

V. Hexose-monophosphate metabolism: G6PD (Nga00833, AKI>)

VI. Phylloquinone biosynthesis: NS (Nga01381)

VII. Polyamine metabolism: BADH (Nga02557, AKL>)

VIII. Urate metabolism: TTL (Nga02546, SRL>, «AGNI01005008», RLx_5HL)

IX. Protein/aminoacid metabolism: ASP (Nga00060, SRL> and Protein 42 «AGNI010001», AHL>)

For decoding of abbreviation see Table D in Appendix. The function of some proteins was not identified, however 14 proteins, which is almost half of the amount of proteins with identified functions and 17% of the total number of predicted proteins, were involved into lipid metabolism.

3.4.Cloning of the predicted peroxisomal *Nannochloropsis gaditana* proteins.

Work with algal cell differs from one with another organisms which were used by the Reumann group. Cell harvesting is not an exception. The new protocol was therefore developed (Chapter 2.8.1). Using this protocol *Nannochloropsis* cells were partially purified and harvested.

For the further genes amplification algal DNA was needed. The DNA isolation from Nannochloropsis microalgae had not been established in the Reumann lab before the present work was started. That is why some important experiments have been made.

The established method of algal DNA extraction was based on the extraction of DNA from the plant cells, where plant tissues were disrupted in a mortar using pestle. Additional precursor steps: liquid nitrogen treatment and use of the mobile homogenizer improved DNA concentration (Chapter 2.8.2). The harvested and aliquoted *Nannochloropsis* cells were used in the genomic DNA isolation. However due to thick cell walls algae cells were disrupted before the use of isolation kits. Three strategies were used: 1) Disruption by manual homogenizer; 2) Disruption of frozen algae mass by mortar and pestle and 3) disruption of frozen algae mass by mortar and pestle followed by additional disruption using manual homogenizer.

Experiments showed that the third option usually gives the best concentration, while the first one was the worst (data not shown). The second strategy gave the result close to the best one. Thus, using mortar and pestle we can efficiently disrupt frozen algae cells.

3.4.1.Amplification of cDNAs by PCR

To explore the presence of the 10 predicted proteins of *N. gaditana* analytical PCR was carried out using different annealing temperatures (T_a) and elongation times. The most complete results were obtained at T_a 55°C for the 1-10 repeats and 60°C for the 11-25 repeats and the elongation time of 150 seconds. The genes were amplified using these parameters and were loaded in to the 1% agarose gel for the electrophoretic separation alongside with the 1 kb of DNA ladder (See Materials and Methods in Chapter 2.9.4 and

2.9.5). Visualization of the separation results shows that most of the proteins were amplified Figure 26. At these parameters 9 out of 13 constructs were amplified correctly NgP1.1 (full length protein Nga00060, 1333bp), NgP1.2 (last exon of Nga00060, 349 bp), NgP3 (Nga00082, 1152 bp), NgP4 (Nga000170, 402 bp), NgP6.1 (Nga01077, 1237 bp), NgP6.2 (last exon of Nga01077, 616 bp), NgP7 (Nga02369, 759 bp), NgP8 (Nga2873, 1482 bp), NgP10 (Nga05502, 548 bp). Four constructs were not amplified: NgP2.1 (Nga00077, 2235 bp), NgP2.2 (C-term. exon of Nga00077, 2030 bp), NgP5 (Nga00224, 1488 bp), NgP9 (Nga03053, 2289 bp) (The corresponding primers are represented in Table C in Appendix).

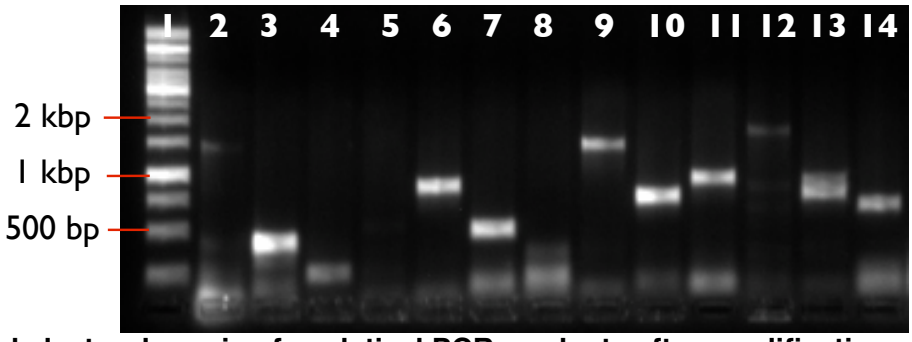


Figure 26. Gel electrophoresis of analytical PCR products after amplification.

Gel electrophoresis was carried out in a 1% agarose gel. 1 kb GeneRuler DNA ladder (lane 1). NgP1.1 (lane 2), NgP1.2 (lane 3), NgP2.1 (lane 4), NgP2.2 (lane5), NgP3 (lane 6), NgP4 (lane 7), NgP5 (lane 8), NgP6.1 (lane 9), NgP6.2 (lane 10), NgP7 (lane 11), NgP8 (lane 12), NgP9 (lane 13), NgP10 (lane 14). The Ta was 55°C-60°C, elongation time 150 seconds.

Three of the most interesting constructs were chosen for further analysis: NgP1.2, NgP4 and NgP10. The experiment aimed at finding precise parameters, most suitable for PCR of these constructs was made. The results of this experiment are represented in the Figure 27: 3 different Ta: 50°C, Ta 55°C and Ta 50°C for the 1-10 cycles with Ta 55°C for the 11-35 cycles (lanes 2-4, 5-7, and 8-10 respectively in the Figure 27), elongation time 90 seconds. All the constructs were amplified with correct size at all three Ta programs. The Ta 55°C was chosen for preparative PCR analysis for the same constructs.

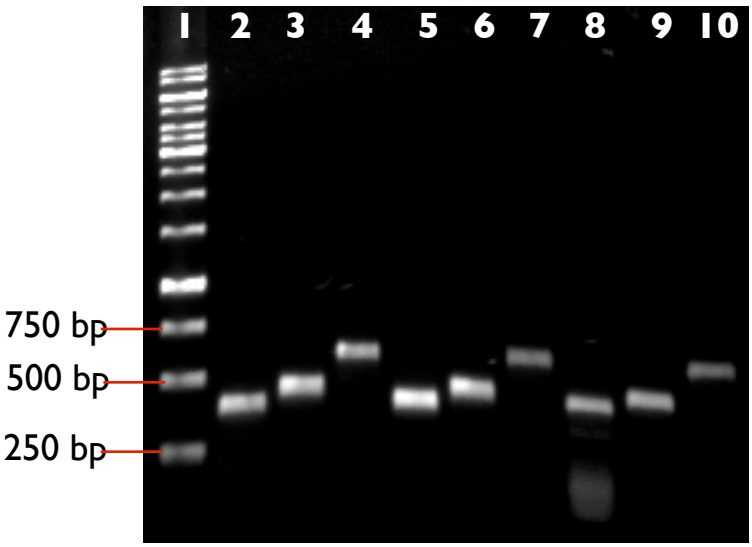


Figure 27. Gel electrophoresis of three analytical PCR products at different Ta.
Gel electrophoresis was carried out in a 1% agarose gel. 1 kb GeneRuler DNA ladder (lane1). NgP1.2 (lane2), NgP4 (lane3), NgP10 (lane4), Ta 50°C; NgP1.2 (lane5), NgP4 (lane6), NgP10 (lane7), Ta 55°C; NgP1.2 (lane8), NgP4 (lane9), NgP10 (lane10), Ta 50°C-55°C. The PCR amplification products for the NgP1.2 was between 250 bp and 500 bp, NgP4 was closer to 500 bp and NgP10 between 500 bp and 750 bp. Which are very similar to the calculated size NgP1.2 - 349 bp, NgP4 - 402 bp, NgP10 - 548 bp. The results shows that the amplification occurred at the correct regions in all three temperatures.

Preparative PCR was carried out using Phusion High-Fidelity DNA Polymerase at Ta 55°C. The result of preparative PCR showed similar PCR products of correct size without unspecific PCR products (Figure 28).

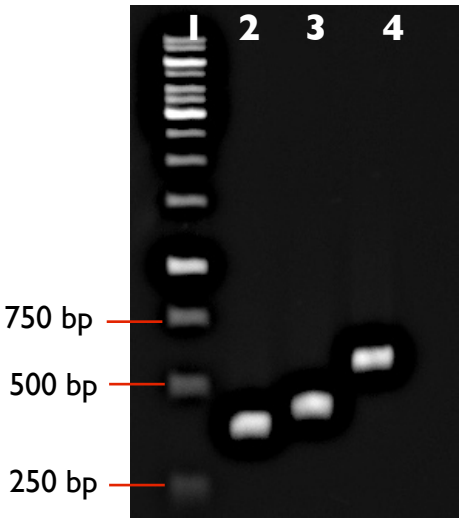


Figure 28. Gel electrophoresis of three preparative PCR products.
Gel electrophoresis was carried out in a 1% agarose gel. 1 kb GeneRuler DNA ladder (lane1). NgP1.2 (lane2), NgP4 (lane3), NgP10 (lane4), Ta 55°C,elongation time 90 sec. The calculated size of products NgP1.2 - 349 bp, NgP4 - 402 bp, NgP10 - 548 bp is similar to the given results.

The PCR product was purified using PCR Purification kit (Fermentas). The concentration of purified DNA was measured on the spectrophotometer (NanoDrop 2000). The constructs were used for subcloning into the pJET cloning vector.

3.4.2.Cloning cDNAs into the pJET cloning vector

The constructs were cloned in to pJET 1.2 cloning vector using CloneJet kit (See Chapter). The constructs were inserted into the multiple cloning cite in the middle of the lethal gene eco47IR. The clones were checked on the presence of the vector by resistance to the ampicillin on the LB agar plates. In order to amplify the number of constructs inserted into

the pJet vector, the competent cell *E. coli* JM 109 was transformed by the plasmids. The colonies were subcultured in the liquid LB medium and plasmid DNA was extracted using GeneJet plasmid miniprep kit from Fermentas, Germany. For the analysis of the correct insert and orientation of the constructs the analytical PCR was performed.

3.4.3.Analysis of the insert

In order to find the orientation of the construct in the vector a pair of vector and gene specific primers were used. In Figure 29. The results of the PCR are represented. The lanes from 2 - 12 represent the constructs in reverse orientation, where NgP1.2 (lanes 2, 8 and 10), NgP4c5 (lane 7); the lanes 13-19 represent the constructs in forward orientation, where NgP4c3 (lanes 15 and 18), NgP10 (lanes 13 and 19). So all the constructs of NgP1.2 taken from two different colonies (c2 and c5) and the vector taken from the colony №3 of NgP4 had the reverse orientation; all the constructs of NgP10 taken from two different colonies (c2 and c4) and the vector taken from the colony №5 of NgP4 had the forward orientation. The lanes 4, 9 and 14 did not have any vectors. The bonds around 2 and 3 kbp were probably the vector DNA in supercoiled and nicked circular forms respectively. The position of bonds related to the constructs of NgP1.2, NgP4 and NgP10 are slightly different (they covered less distance) from the results of native cDNA amplification (Figure 29) because PCR products, made with the use of one of the vector specific primers, make the constructs longer up to around 70 bp.

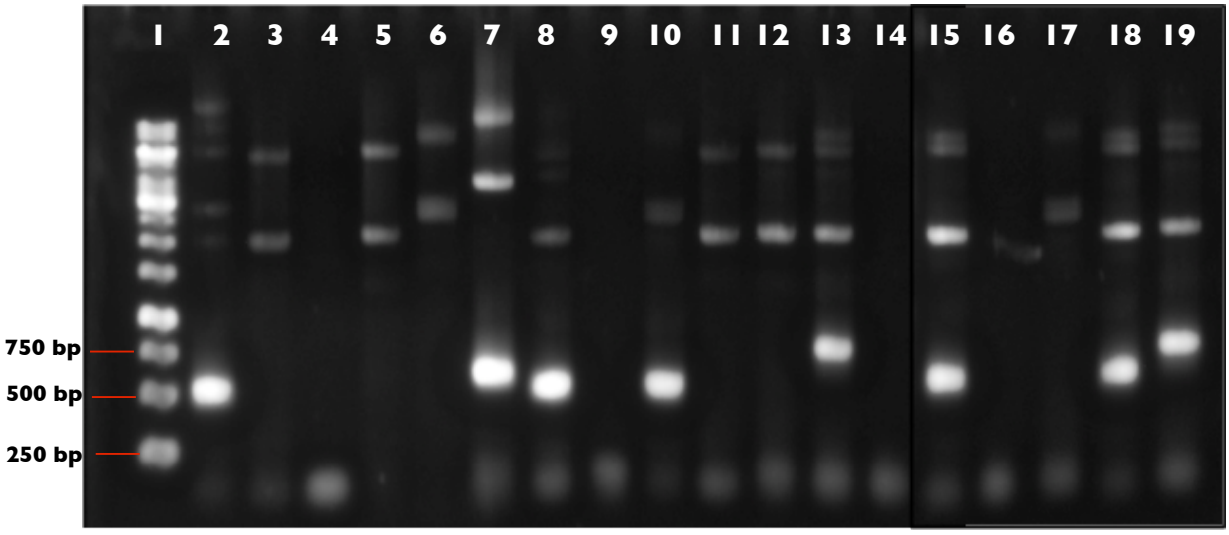


Figure 29. Gel electrophoresis of the insert orientation analysis.

Gel electrophoresis was carried out in a 1% agarose gel. 1 kb GeneRuler DNA ladder (lane1). Proposed reverse orientation: NgP1.2 (**lane 2**), NgP4c3 (lane 3), NgP4c5 (~~lane 4~~), NgP10c4 (lane 5), NgP10c2 (lane 6), NgP4c5 (**lane 7**), NgP1.2c2 (**lane 8**), NgP10c2 (~~lane 9~~), NgP1.2c5 (**lane 10**), NgP4c3 (lane 11), NgP10c4 (lane 12); proposed forward orientation: NgP10c4 (**lane 13**), NgP10c2 (lane 14), NgP4c3 (**lane 15**), NgP10v2 (~~lane 16~~), NgP1.2c5 (lane 17), NgP4c3 (**lane 18**), NgP10c4

(**lane 19**). The lanes marked in bold showed positive results of orientation, the crossed lanes showed absence of vector DNA the others non marked lanes showed negative results of proposed orientation. The PCR amplification products for the NgP1.2 was around 500 bp, NgP4 was a bit higher than 500 bp and NgP10 was in between 500 bp and 750 bp. Which are very similar to the calculated size NgP1.2 - 420 bp, NgP4 - 480 bp, NgP10 - 520 bp.

Based on the results of analytical PCR, the constructs NgP1.2c2, NgP4c5 and NgP10c4 were chosen for the preparative PCR. After the preparative PCR the chosen constructs were sent purified and were sent for sequencing. However the results of sequencing were not available

3.4.4.Cloning cDNAs into the pCAT-EYFP vector

After the investigation of the orientation of *N. gaditana* protein constructs in the cloning vector pJET the following constructs were double digested to get sticky ends. For double digest of the constructs NgP1.2 and NgP4 SacII and NotI restriction endonucleases were used, for the NgP10 NotI and XbaI restriction endonucleases were used Figure 30. The cDNAs NgP1.2, NgP4 and NgP10 were cut out from pJET vector (lanes 2-3, 4-5 and 6-7 respectively). The destination vector pCAT-EYFP-DECR was also double digested with the same pairs of restriction nucleases. In Figure lanes 8-11 corresponding to the double digest made by SacII and NotI, while 12-13 were made with NotI and XbaI. The digested pCAT vector released DECR insert (1000 bp). The length of the other inserts is hard to define accurately, because the ladder proteins covered longer distance than the other samples. However, the NgP1.2 (350 bp) and NgP10 (550 bp) and DECR (1000 bp) can be assumed to be digested correctly. At the same time the product of pJET-NgP4 looked like incorrectly digested. The experiment was repeated and the correct product of pJET-NgP4 (400 bp) digestion was obtained (data not shown).

The correct products of digestion should have been purified from the gel and used for the ligation into the opened vectors with corresponding sticky ends, however the extraction of the DNA from the gels using two different kits did not give useful concentrations (< 10 ng/ μl). The ligation was done with this low concentration, following by transformation of *E.coli* cells, but the living colonies were not obtained. The repeating experiments were not performed due to the time constrains.

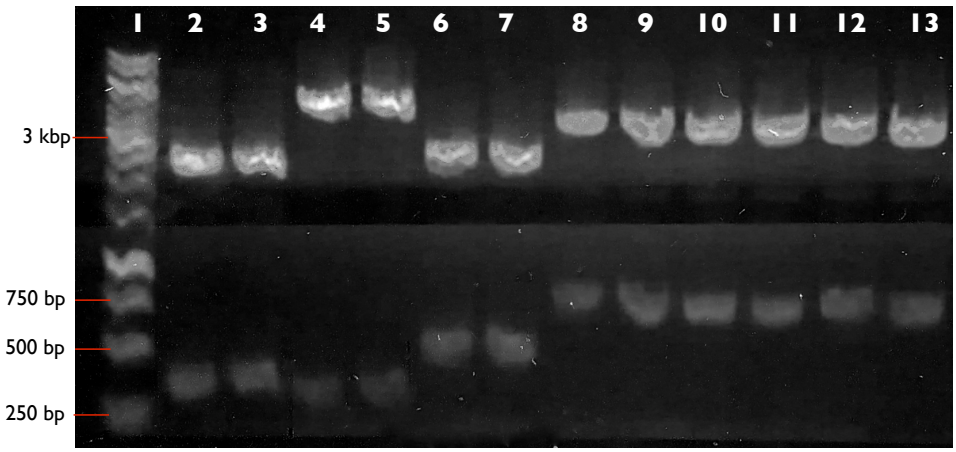


Figure 30. Gel electrophoresis of the double digest products.

Gel electrophoresis was carried out in a 1% agarose gel. 1 kb GeneRuler DNA ladder (lane 1). Product of digest NgP1.2 by SacII and NotI (lanes 2, 3); Product of digest NgP4 by SacII and NotI (lanes 4, 5); Product of digest NgP10 by SacI and XbaI (lanes 6, 7); Product of digest pCAT-EYFP-DECR by SacII and NotI (lanes 8 - 11); Product of digest pCAT-EYFP-DECR by SacI and XbaI (lanes 12, 13). The lower bonds corresponds to the inserts, the upper bonds corresponds to the double digested vectors. Calculated NgP1.2 insert size - 350 bp, NgP4 - 400 bp, NgP10 - 550 bp, pJET - 3 kbp, pCAT-EYFP - 3.8 kbp. The products were correct, except the NgP4, which size did not correspond to the one predicted.

3.5.Genomic transformation of the *Nannochloropsis* algae by electroporation

The genomic transformation of *Nannochloropsis oceanica* by electroporation has been recently established and published (Vieler et al., 2012). The pSelects100 vector and the transformation protocol were designed for the *N. oceanica* and obtained from C. Benning, Department of Biochemistry and Molecular Biology, Michigan State University, USA.

The transformation was done for *N. oceanica* CCMP 1779 and *N. gaditana* CCMP 526. The pSELECT100 plasmid was dissolved, dropped and dried on the sheet of Watman paper and sent to us by the C. Benning. The region corresponding to the droplet was cut out and placed in to the 1.5 Eppendorf tube and filled by 0.5 ml 0.5 M Tris-HCl buffer, mixed and stored in the refrigerator at -20°C. The *E. coli* competent cells JM 109 were transformed by this vector, positive colonies were selected by resistance to Ampicillin (Chapter 2.9.8). The positive colonies were subcultured in liquid media (10 ml in each tube) and the plasmid was extracted using GeneJet plasmid miniprep kit. The DNA concentration was measured on the NanoDrop2000 spectrometer. A common concentration for 10 ml of liquid culture was 150 ng/μl. The vector was linearized overnight by BamHI restriction enzyme in the specific buffer (Chapter 2.9.7). The analytical gel electrophoresis was done to check the restriction product (Figure 31). A half of total

amount of the linearized vector was purified using PCR Purification kit (Fermentas). The other part was used unpurified to avoid lost of DNA during purification).

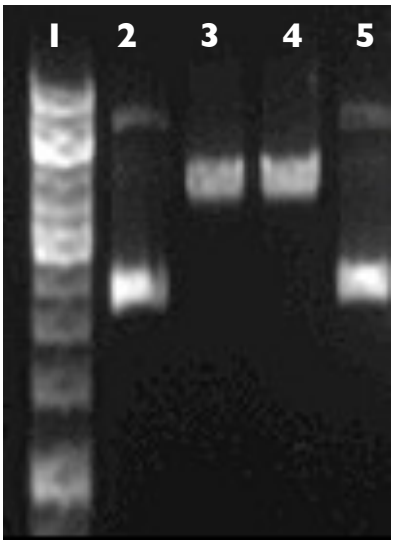


Figure 31. The results of restriction digest of pSELECT100 by BamHI restriction enzyme.

1kb DNA ladder (lane 1),1 μl of the native pSELECT100 (lanes 2 and 5), 1 μl of the digested pSELECT100 vector by BamHI (lanes 3 and 4). The native pSELECT100 plasmid exists in two forms: nicked circular and supercoiled; the digested vector is represented only by linearized form which electrophoretic mobility is lower than supercoiled and higher than nicked circular.

It was not possible to apply recommended 2.2 kV to the samples, containing vector DNA (See Chapter). Even if the linearized vector was purified, the salts were still present in the final mixture making the resistance of the sample lower than the working value 500-600 Ohms. In the samples with low resistance the «arc» message on the display often lighted up meaning that the arc prevention system and quenching system have been made activ and the pulse has been terminated. Unfortunately the MicroPulser does not have a function of measuring resistance of the sample before the pulse. The pulse has to be done to see the actual parameters of the electric flow. In case the sample had a lower resistance it was diluted twice by non-polar solvent Sorbitol with the concentration of 0.375M. Such addition diluted a sample and the dissolved salts, it increased resistance and in some cases helped to get a pulse at 2.2kV. However in the majority of the cases it was not effective and the only solution which helped was decreasing the voltage setting, generally up to 1.6 kV (which is 28% lower than recommended). On the other hand this change in settings decreased the time of the pulse as well, down to 1.6 seconds (which is 60% lower than maximum value 4.0 sec. and only 6.5% of recommended value - 25 sec).

The number of pulses which were done was different from one proposed in the original protocol to cover five fold difference in capacitance ability between our instrument and one described by the authors of the transformation protocol. (Vieler et al., 2012). Pulse number varied from 5 to 10 for *N. oceanica* at 1.6 kV and 1.6 sec. pulse duration. For *N. gaditana* 7 pulses were made with the same characteristics and 4 pulses with 1.5 kV during 3.6 sec. were made for *N. gaditana* negative control.

One week after the transformation the top agar contained plenty of very small and pale-green colonies both in samples and negative controls. However after two weeks these colonies almost disappeared and only few dark-green colonies larger in size (about 1-2 mm in diameter) left on the top agar surface (See Figure 32). All 4 plates of *N. oceanica* containing vector DNA got colonies and one of the two negative controls got a few as well. *N. gaditana* plates do not contain any dark green colonies both in a sample and negative control. The colour of the top agar containing small pale initial colonies is yellow. Probably the cells were not transformed and died because of Hygromycin B sensitivity. There can be several explanations of the presence of the growing colonies in the control plates that will be further discussed in the Chapter 4.5.

At the moment of writing of this thesis the colonies are still growing on the plates and further screening by PCR, Wester Blotting, etc., has to be done to confirm the presence of the inserted DNA inside of the algal cells.

The positive results, stated above, confirm that the efficient genetic transformation of promising oil-producing microalgae *N. oceanica* was successfully established.

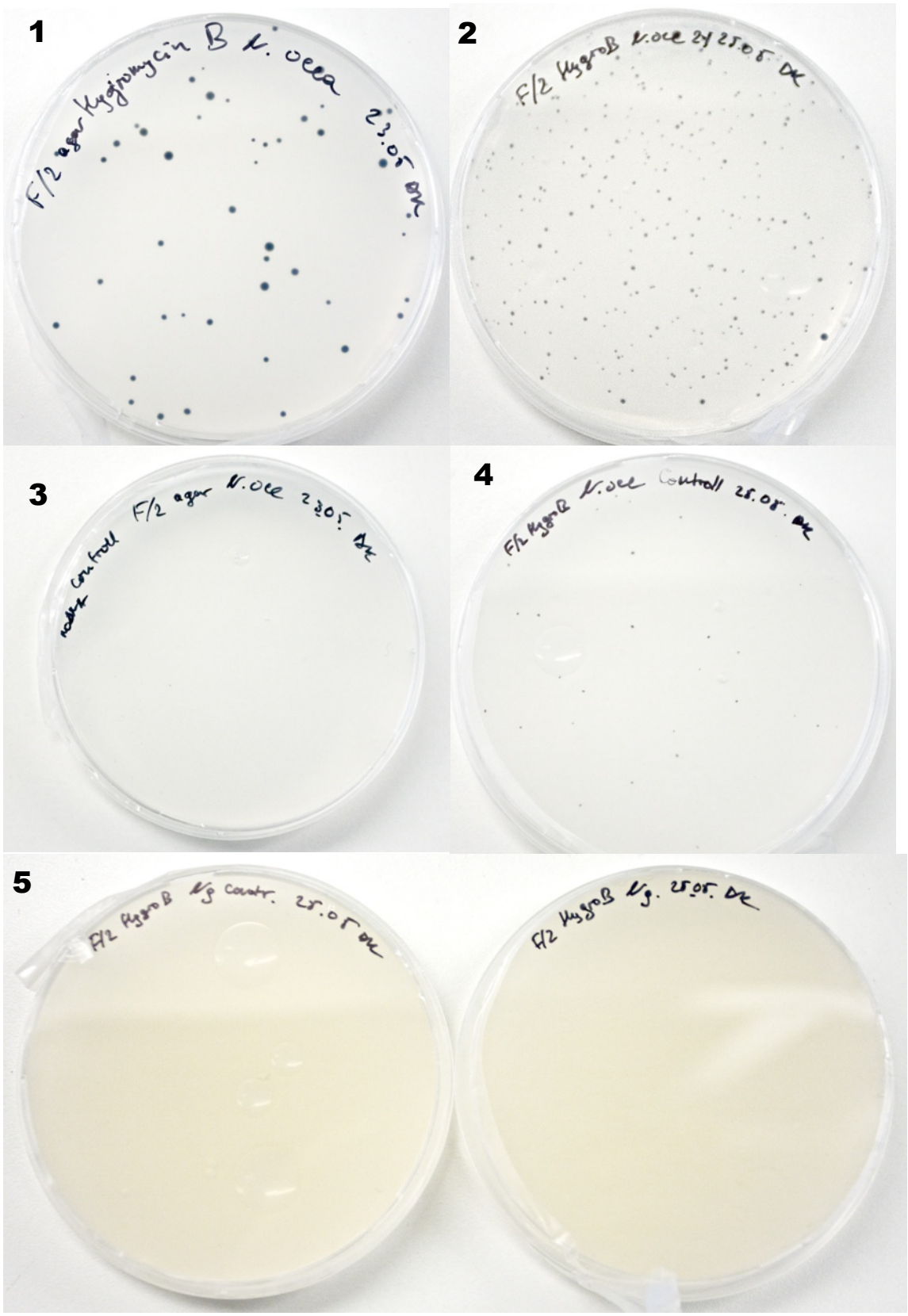


Figure 32. The results of the transformation of *N. oceanica* CCMP 1779 and *N. gaditana* CCMP 526 by electroporation using pSELECT100 selection vector. *N. oceanica* colonies of positive transformants (32.1 and 32.2); *N. oceanica* negative controls - no DNA inserted (32.3 and 32.4), where *N. oceanica* colonies are present in the figure 32.4; negative results of transformation *N. gaditana* both in the samples with (left plate) or without - negative control (right plate) DNA.

4.Discussion

4.1.Growth analysis

4.1.1.Choice of optimum wave length for spectroscopic measurements

The optical density of algal cultures (*N. salina* (strain NIVA-2/01), *N. oceanica* (strain NIVA-2/03) and *P. tricornutum* (NIVA-BAC 2)) was initially measured at three different wavelengths, namely 540, 625 and 680 nm based on published data for *Nannochloropsis* (Rocha et al., 2003; Hsueh et al., 2007; Gu et al., 2012). The experiment was done in 36 days with periodic OD measurements every 2-3 days (Fig. 18, 19).

The highest OD (540 nm) values compared to OD (625 nm) and OD (680 nm) (Fig. 18) can be explained by the hypothesis that in such species the amount of carotenoids was higher than the amount of chlorophyll a (Chl a), which is the only chlorophyll molecule present in these microalgae. Chl a has the maximum absorption at 670 nm (Wellburn, 1994). The OD (625 nm) and OD (680 nm) values were very similar and close to the absorption maximum of Chl a. Compared to colorless microorganisms such as *E. coli* and *S. cerevisiae*, whose OD routinely measured at 600 and 650 nm, respectively (Chen et al., 1998; Sezonov et al., 2007), the OD values of microalgae are increased due to chl a and carotenoid absorption, depending on the chosen wavelength and the absorption maxima of the pigments (Wellburn, 1994).

The change of the color of *Nannochloropsis* algae cultures during long time experiments (more tan 1 month) can be explained by the destruction of chlorophyll a molecules in the cell. Decreasing level of chlorophyll a makes red-yellow colored pigments - carotenoids more visible. The cells become yellowish. The reason for this phenomenon is unknown, but probably it happened because of the increasing number of dying cell during the long growing time. The most probable explanation of such massive death is that the algal cell cultures reached their maximum density and used up some vital compounds from the medium. We could suppose that such critical compound is the nitrogen. In the experiments of growing microalgae in the medium without nitrogen the algae become yellowish much faster (during one week). This hypothesis is supported by Solovchenko et al. (2011), who wrote about the increasing proportion of carotenoids/chlorophyll in *Nannochloropsis* algae during culturing time in stress conditions.

The analysis of standard deviation values shows that all the results of measurements at these wavelengths are very similar (1.5% of measurements values for undiluted samples and around 2.5% for the 1:2 diluted) and it is hard to find the best wavelength by this comparison. The higher deviation in the diluted samples can be explained by the loss of some sample volume during the transportation to the cuvette, irregularity of dilution and insufficient mixing.

No experiments aimed at analyzing such correlation have been undertaken, but the results of measurements of algae optical density show that the proportion of carotenoids increases with aging of the algal culture.

Measuring optical density of the culture with known cell concentration by spectrophotometer requires additional knowledge in spectroscopy. Spectroscopes are very precise instruments, that is why the results are affected by the quality of samples. Major factors in this respect are: 1) Homogeneity of sample concentration. The same as for sample preparation for direct microscopic measurements using counting chambers. 2) Optical density of the sample. Too diluted sample can give a value equal to statistical error, too dense samples scatter and reflect light from the source, resulting in higher than actual cell concentration. 2) Constant environment. Calibration curve made by plotting optical density of the culture versus cell concentration can be applied only for the investigation of cell concentration of the same organism which was grown up in the same conditions. Otherwise, the quality of cell number prediction can be low.

4.1.2.Cell counting

The direct cell counting using microscopic observation of cells on the counting chambers was chosen as primary method of investigation of cell concentration of *Nannochloropsis* algae. The counting was done for the same species as stated in the Chapter 4.1.2 during the same time interval. The results are represented in Figure 20. The highest cell concentrations were detected for *N. oceanica* (e.g. $3.1x10^7 \pm 2.4x10^6$ cell/ml), moderate concentrations for *N. salina* ($1.8x10^7 \pm 2.4x10^6$ cells/ml) and the lowest concentrations for *P. tricornutum* ($8.96x10^6 \pm 1.65x10^6$ cells/ml). The relative standard deviation was similar for *Nannochloropsis* species (12.6%), it was higher for *Phaeodactylum* (16.4%). Probable explanation of higher relative standard deviation values for *Phaeodactylum,* than for *Nannochloropsis* species can be that this algae can have different shapes of the cells from

circular shape to star-like one, while needle-like shape of the cell is the more abundant. Probably some circular cells were not counted.

The normal logaripmic growth was expected for all the cultures, however at the middle of the experiment the decrease of growth was found for the *Nannochloropsis* species (Figure 20). It can be explained by error in counting or by influence of some unknown stress-factors which were the reason of the cell death.

A quality of cell counting using counting chambers is influenced by several factors:

1)Homogeneity of sample concentration. The difference in concentration between samples should be as low as possible (a culture should be mixed properly). 2) The volume counted. The larger is the sample, the lower is the error. 3) Concentration of cells. The number of cells in the large square should be in a range of 50-400, larger cell concentration becomes more advantageous, lower concentration increases the error. 4) Loaded volume. Recommended volume of Bürker counting chamber is 14 μl. Lower volume affects cell distribution, large - increases sample thickness and some cells are out of focal plane as a result. 4) Human factor. This is probably the most important factor. Omitting cell by mistake or counting undesired objects (bacterials, dirt) can affect the results. It seems that the most important in this respect is to follow the counting rules, to have excellent knowledge in the microscopy and to be accurate when performing the counting.

4.1.3.Contamination analysis and culture purification

There are several methods of purifying the samples and obtaining axenic cultures available, such as centrifugation, sonication, using detergents, antibiotics, dilution (Stein, 1980). However it is very difficult to get axenic culture using only one method. It happens because of inhomogeneity of microalgae population, where every strain has its own antibiotic resistance. Another reason is that bacteria can make a symbiotic relationship with algae. Moreover, contaminators heterogeneity can be a problem - antibiotics generally kill bacteria but not eukaryotic contaminators or kill only specific groups of bacteria (e.g. ampicillin). It was reported in the literature that many algal species are sensitive to antibiotics against fungi. At the same time some algae are resistant to some groups of antibiotics and it is possible that the methods of removing bacterials do not work for fungi (Jones et al., 1973). That is why the combination of the methods has to be applied (Stein, 1980; Sensen et al., 1993).

The experiment from which the best results were obtained consisted of a combination of two techniques: using antibiotics for the liquid cultures with followed multiple dilution on 96 well plate (See Chapter 3.1.3). The addition of the purification step (centrifugation) before subculturing into the media with antibiotic and plating the multiple diluted cultures on the agar plates can strongly increase the probability of getting axenic cultures. However, this is a very time consuming combination of methods.

Probably, the method of plating the algae on agar containing antibiotic can be sufficient as well. Therefore further studies aimed at determining the antibiotic resistivity and ability of *Nannochloropsis* algal species to grow on agar have to be undertaken.

4.2. Lipid content

4.2.1.Neutral lipid staining by BODIPY 505/515

The lipid staining of oil producing microalgae *Nannochloropsis* was performed using BODIPY 505/515 neutral lipid staining dye dissolved in DMSO according to the previously described method (Cooper et al., 2010; Brennan et al., 2012). The fluorescence emitted by this dye was observed on Nikon TE2000U epifluorescent microscope. The excitation was performed at 440-480 nm wavelength, emission was detected at 535-550 nm which is similar to the setting used for GPF-observation. BODIPY 505/515 fluorescence was detected inside of the cells in compact spots, probably lipid bodies.

Because of the irregular light intensity from the light source and the use of setting, which is not totally compatible with the use of staining dye, either excitation or emission maximum, the quality of images was generally low. In order to improve the quality, exposure time was increased up to 10-30 seconds, resulting in impossibility of taking pictures of moving cells. Chlorophyll autofluorescence was detected only for the *N. gaditana* cells in the first experiment. Probably, due to the fact that at that time the light source was better. This fail was the reason for the impossibility to make semiquantitative analysis of the lipid content.

Several suggestions to improve lipid content analysis by the BODIPY 505/515 can be named. First, increased stability and quality of light source is necessary for the detection of fluorescence. Second, using the wavelengths closer to maximum for the dye (excitation at 505 nm, emission at 515 nm) seams to improve the detection of the emitting light. Probably, confocal light microscope equipped with different lasers and filters could be more suitable instrument for such analysis. Third, fixation of algal cells by adding into a

sample some substances which can increase its viscosity without affecting microscopic measurements, for example glycerol, may solve the problem of moving cells.

The conclusion that can be drawn is that the improvements of technical properties of the microscope have to be done in order to increase the quality of images. The quality of staining can be considered as sufficient and the staining procedure can be used without modifications in future experiments.

4.2.2.Lipid accumulation at normal and nitrogen depletion stress conditions

Numerous studies show that microalgae accumulate lipid under stress conditions (Rodolfi et al., 2009; Hoffmann et al., 2010; Solovchenko et al., 2011). Nitrogen depletion, low or high temperatures, extreme light conditions can induce protection mechanisms in algae cells, however these mechanisms are not well known in microalgae research because of the large diversity between species which causes problems for the application of the methods of analysis used for known model organisms to the other ones.

It was shown that microalgae genius Nannochloropsis produce in high amounts mostly neutral lipids in the form of triacylglicerols TGA and store them in the lipid bodies in cytoplasm. Nitrogen limited conditions push algal cell to produce lipids, which level can reach 60-70% of dry weight (Rodolfi et al., 2009). At the same time growth rate and protein content decreases, which can be explained in the following way: the fact that algal cells inhibit their processes leads to growth and division and makes a storage of energy reach compounds - lipids. Another reason for the activation of biosynthesis of fatty acids in the algal cells that could be named is the protection of photosynthetic apparatus, which can generate excessive amount of electrons under nitrate limitation. Such electrons could produce undesirable and harmful reactive oxygen species (ROS). Fatty acid biosynthesis requires relatively high number of electrons which leads to the decreasing level of «free» electrons in electron transport chain and lower production of ROS (Greenwell et al., 2010; Hoffmann et al., 2010).

Probably this is the reason why some parts of lipid biosynthesis occur in peroxisomes - organelles with well known important role in ROS metabolism.

In these study four different species of oil producing microalgae *Nannochloropsis* were used for the visualization of neutral lipids by fluorescence dye BODOPY 505/515 growing in normal and nitrogen depleted conditions. In Figure 22 (Capter 3.2.2) brightfield and

fluorescent images made for all these species (see annotation) are represented. It can be assumed that in all the represented algae lipid content increased, especially in the cells of *N. gaditana*, but at the same time the culture became yellow which can be explained by degradation of ch a and/or by cell death. (Chapter 4.1.1) However it is hard to say anything accurately due to the impossibility of the quantitative comparison of the given results. At the same time it should be mentioned that the method of staining was successfully established, first positive results are obtained and the way for future studies is open.

4.3.Bioinformatic analyses of Nannochloropsis proteins

Nannochloropsis gaditana and *N. oceanica* genomes have been recently sequenced. For *N. gaditana* gene and protein predictions were made. Dr. T. Lingner made the predictions of peroxisomal localization for all the available proteins using the prediction servicer PredPlantPTS1. The result of this prediction was the list of 78 proteins with posterior probability to have PTS1 above 0.5.

4.3.1. Prediction and bioinformatic validation of the PTS1 proteome of Nannochloropsis

The first step of validation of the T. Lingner predictions included orthology analysis of all 78 predicted proteins. The orthologs for predicted proteins were searched among *stramenopiles* and *A. thaliana* proteins. If at least one of the homologues proteins had peroxisomal targeting, this protein was considered as probably correctly predicted. Peroxisomal targeting of the proteins was identified by the examination of its annotation and manual prediction using PredPlantPTS1.

It can be concluded that this method was efficient for the validation of the peroxisomal targeting only for the proteins which have orthologs among analyzed organisms. This group of proteins has probably conserved function and localization which reflect their importance. However the rest 50 proteins were not considered as non-peroxisomal. Some proteins can change its structure during evolution. That is why it was decided to check those 50 proteins for carrying the PTS1 tripeptide manually. It was done by blasting amino acid sequence of *N. gaditana* protein against the whole genome contigs of the same organism.

The results showed that the predicted amino acid sequence is not always 100% identical to the related nucleotide sequence. In some cases protein aa sequence didn’t start from

the methionine (which is the first aa of all the proteins), in the others predicted C terminus didn't have the stop codon in related nucleotide sequence. In another words the borders of aa sequence differ from the border of the nucleotide sequence from which this protein sequence was predicted.

Such situation cannot be explained by alternative splicing, because, firstly, it was detected even in single exon genes, secondly, the missing parts of nucleotide sequence were located in the same exons as the closest correctly translated aa regions of the protein.

The probable explanation could be that the authors of protein prediction used not totally corrected prediction algorithms. Possible reason could be that they didn't use the phylogenetically closest organism as a model for predictions. Differences in gene structure and low number of regions of similarity interrupted correct identification of terminal regions.

After the investigation of actual C-terminal ends of the proteins and prediction of presence of the PTS1 tripeptide there, only 12 proteins remained in the list of predicted peroxisomal proteins. Considering that these proteins didn't have orthologs among sequenced genomes of stramenopiles and A.thaliana, they were labeled «organism specific». They are probably unique and are present only in *Nannochloropsis.* It means that they have previously unknown function in peroxisomes or their function and structure are very different from the other organisms.

Another 28 proteins from orthology analysis were analyzed, and the presence of PTS1 in C-terminus end was confirmed for all of them.

Two more PTS1 proteins not from the list provided by Dr. T. Lingner were found later. The protein 41 was found by blasting known peroxisomal multifunctional protein of *A.thaliana* (AIM1) against nucleotide collection of *N. gaditana* with following manual translation in to amino acid sequence using web-based translation tool: ExPASy (http://web.expasy.org/translate/). The presence of PTS1 signal ARL was confirmed by the prediction servicer PredPlantPTS1. Protein 42 was found during the analysis of blast search results for the *A.thaliana* PTS2 protein aspartate aminotransferase ASP against *Nannochloropsis* whole genome contigs. The procedure of verification that is similar to the one described above was undertaken as well.

30 proteins were predicted to be conserved PTS1 proteins among *Stramenopiles* and *A. thaliana.* While 12 were annotated as probably genus-specific. Among them 4 are probably have wrongly predicted structure (See Chapter 3.3.3, Table 13)

Five PTS1 proteins predicted for N. gaditana were not found in N. oceanica, while one of PTS2 proteins of N. oceanica was not predicted for the N. gaditana (See tables 11 and 13 in Chapter 3.3). These proteins which are found only in one organism inside of the genus *Nannochloropsis* can be name as species-specific.

4.3.2.Prediction of PTS2 proteins

In total 14 known *A. thaliana* proteins from AraPerox v1.2 (http://www3.uis.no/araperoxv1/) were analyzed using the newly established method described in Chapter 2.8.3. The *A. thaliana* PTS2 protein sequences were blasted against *Nannochloropsis* whole genome contigs, putative N-terminal ends were estimated and the presence of PTS2 was manually checked. The presence of PTS1 in the C-terminal end was also checked (Chapter 3.3.5. Table 11).

Seven proteins of *N. gaditana* (six - for *N. oceanica*) lacked any PTS. Probably they are not peroxisomal in *Nannochloropsis* species. Four *N. gaditana* proteins orthologues to the A. thaliana PTS2 proteins exchanged PTS2 on PTS1. They are orthologs to ASP - protein family, ACX - protein family and CSY - protein family and LACS6 protein. In *N. oceanica* ortholog to ASP3 protein did not contained any PTS, while orthologs to ACX, CSY and LACS6 had PTS1. Such «exchange» can be explained by hypotesis that PTS1 gives better localization probability than PTS2, however the real reason and evolutionary driving force remain unknown.

Three proteins containing PTS2 in *N. gaditana* were predicted. They are ortholgs to PKT/ KAT-protein family and TTL ptotein of *A. thaliana.* The same proteins were predicted for *N. oceanica* including one more ortholog to PMDH-protein family. These proteins have conserved structure and keeping PTS2, because they can perform their function only in the present form.

4.4.Genetic transformation of the Nannochloropsis algae by electroporation

The genetic transformation of *Nannochloropsis oceanica* by electroporation using pSelect100 vector with Hygromycin B and Ampicillin resistance genes was done according to the protocol described by Vieler et al.(2012).

The transformation was originally designed for *N. oceanica* CCMP 1779 (Vieler et al., 2012) but in the other protocols for the *Nannochloropsis* transformation such as *N. gaditana* CCMP 526 (Radakovits et al., 2012) and *N. sp* by (Kilian et al., 2011) similar electroporetic parameters were used (See Chapter 2.10).

The parameters of pulse which were used by authors are the followed: 2.2 kV for 25 sec. However our instrument had in five times lower capacitance and the maximum pulse duration was four sec. It was decided to increase number of pulses to compensate short time of the pulse. The number of pulses during the experiments varied from 5 to 10 for *N.oceanica* sample containing DNA and negative control at 1.6 kV and 1.6 sec. pulse duration. For *N. gaditana* sample with vector DNA seven pulses were made with the same characteristics, while four pulses with 1.5 kV during 3.6 sec. were made for *N. gaditana* negative control. The time between pulses was about 30 seconds. The maximum number of pulses depended on heating of the solution and foam formation together with the fast pelting of the cells. The correlated between number of pulses and efficiency of transformation was not found. More experiments have to be done.

The impossibility of the application 2.2 kV to the samples, containing vector DNA can be explained by arc formation because of the low resistance of the solution, containing relatively high salt concentration. The MicroPulser cannot measure the resistance of the sample before the pulse. The decrease of the pulse duration when the time was also decreased has to be explored. At the same time the better methods for removing the salts from the solution of the cells and DNA have to be developed.

After around 5 days plates were covered by parafilm according to the protocol (Vieler et al., 2012). The explanation of this ventilation step is that the evaporation of water from the 0.5% top agar is desiring, this will make this layer thinner and bring colonies closer to the agar containing antibiotic. Diffusion of Hygromycin at the border of two agars is much better, meaning that the concentration is higher, resulting in the fact that the cells without resistance gene to the antibiotic die.

One week after the transformation small initial colonies of microalgae appeared, but after 2 weeks most of them disappeared and only several became easily visible (For results see Chapter 3.5).

The positive colonies were found in all the 4 plates of *N. oceanica* containing vector DNA and one of two negative controls. *N. gaditana* plates did not contain any visible colonies.

Most probably the cells were not transformed and died because of Hygromycin B sensitivity. Another possible reason is that the cells died under transformation, but this is less probable. The presence of the growing colonies in the control plates can be explained in the following ways: 1) the individual cells have resistivity to Hygromycin B; 2) the colonies located in the plate or region of the plate where there is antibiotic has low concentration. The conclusions are the following1) further experiments for checking sensitivity of *N. oceanica* and *N. gaditana* to Hygromycin B should be made, 2) longer time period of growing can probably help top agar to evaporate and bring algae colonies closer to the bottom agar. In addition new repeating experiments have to be made with a screening on the PCR or other methods to verify the presence of the Vector DNA.

5.Conclusion and future perspectives

The present Master project made significant contributions to the establishment of microalgal research in the plant peroxisome group by Prof. Reumann at UiS. Several basic culturing techniques for *Nannochloropsis* were established, including those for analysis and optimization of growth conditions (cell counting and spectroscopic measurements) and for subculturing, harvesting and genomic DNA isolation.

The experiments for purification of contaminated cultures showed that combination of methods, such as sample dilution, antibiotic treatment and plating, gives the best results, even if axenic cultures were not obtained in this study.

The methodology of visualizing the accumulation of neutral lipids in lipid bodies by fluorescence microscopy was established. The largest amount of lipids was detected in the *N. gaditana* cells under normal and nitrogen starvation conditions.

The prediction and analysis of the exon/intron structure of PTS1/2 proteome of soluble matrix proteins for *N. gaditana* were made. 30 conserved putative orthologs of known *Arabidopsis* PTS1 proteins and 12 genus-specific proteins were found. In total 42 PTS1 proteins were predicted, 10 of them were chosen for subcloning, two full-length cDNAs (Nga00170 and Nga05502) and one C-terminal exons (Nga00060) of predicted PTS1 proteins of *N. gaditana* were cloned into pJET vector. However the cloning into the pCAT-EYFP vector was not successful due to the insufficient extraction of the DNA from the agarose gel.

Two conserved putative orthologs of known *Arabidopsis* PTS2 proteins were identified. However, further experiments have to be done to verify subcellular localization of predicted peroxisomal proteins and presence of PTS2 pathway.

In addition the successful transformation *of N. oceanica* was performed. The transformation of *N. gaditana* was unsuccessful.

The combination of the best growth and productivity characteristics among six *Nannochloropsis* species, investigated together with the availability of sequenced genome and the prediction of 14 peroxisomal proteins involved in lipid metabolism, makes *N. gaditana* the most promising candidate for production of LC-PUFA in the large scale.

6.References

Apt, K. E., P. G. Kroth-Pancic, et al. (1996). "Stable nuclear transformation of the diatom Phaeodactylum tricornutum." Mol Gen Genet **252**(5): 572-579.

Arabidopsis Genome, I. (2000). "Analysis of the genome sequence of the flowering plant Arabidopsis thaliana." Nature **408**(6814): 796-815.

Armbrust, E. V., J. A. Berges, et al. (2004). "The genome of the diatom Thalassiosira pseudonana: ecology, evolution, and metabolism." Science **306**(5693): 79-86.

Baxevanis, A. D. and B. F. Ouellette (2004). Bioinformatics: a practical guide to the analysis of genes and proteins, Wiley-interscience.

Black, J. G. (2012). Microbiology: Principles and Explorations, Wiley.

Bowler, C., A. E. Allen, et al. (2008). "The Phaeodactylum genome reveals the evolutionary history of diatom genomes." Nature **456**(7219): 239-244.

Bradley, P., D. Cheney, et al. (1988). "One step antibiotic disk method for obtaining axenic cultures of multicellular marine algae." Plant Cell, Tissue and Organ Culture **12**(1): 55-60.

Brennan, L., A. Blanco Fernandez, et al. (2012). "Enhancement of BODIPY505/515 lipid fluorescence method for applications in biofuel-directed microalgae production." J Microbiol Methods **90**(2): 137-143.

Campbell, N. A. and J. B. Reece (2008). Biology, Pearson, Benjamin Cummings.

Chen, J., A. Brevet, et al. (1998). "Control of 5',5'-dinucleoside triphosphate catabolism by APH1, a Saccharomyces cerevisiae analog of human FHIT." J Bacteriol **180**(9): 2345-2349.

Cock, J. M., L. Sterck, et al. (2010). "The Ectocarpus genome and the independent evolution of multicellularity in brown algae." Nature **465**(7298): 617-621.

Cooper, M. S., W. R. Hardin, et al. (2010). "Visualizing "green oil" in live algal cells." J Biosci Bioeng **109**(2): 198-201.

Doan, T.-T. Y. and J. P. Obbard (2010). "Improved Nile Red staining of Nannochloropsis sp." Journal of Applied Phycology **23**(5): 895-901.

Dougherty, E. C. (1959). "INTRODUCTION TO AXENIC CULTURE OF INVERTEBRATE METAZOA: A GOAL*." Annals of the New York Academy of Sciences **77**(2): 27-54.

Dower, W. J., J. F. Miller, et al. (1988). "High efficiency transformation of E. coli by high voltage electroporation." Nucleic Acids Res **16**(13): 6127-6145.

Greenwell, H. C., L. M. Laurens, et al. (2010). "Placing microalgae on the biofuels priority list: a review of the technological challenges." J R Soc Interface **7**(46): 703-726.

Gu, N., Q. Lin, et al. (2012). "Effect of salinity on growth, biochemical composition, and lipid productivity of Nannochloropsis oculataCS 179." Engineering in Life Sciences **12**(6): 631-637.

Guillard, R. L. (1975). Culture of Phytoplankton for Feeding Marine Invertebrates. Culture of Marine Invertebrate Animals. W. Smith and M. Chanley, Springer US**:** 29-60.

Guillard, R. R. and J. H. Ryther (1962). "Studies of marine planktonic diatoms. I. Cyclotella nana Hustedt, and Detonula confervacea (cleve) Gran." Can J Microbiol **8**: 229-239.

Hanahan, D. (1983). "Studies on transformation of Escherichia coli with plasmids." J Mol Biol **166**(4): 557-580.

Hsueh, H. T., H. Chu, et al. (2007). "A batch study on the bio-fixation of carbon dioxide in the absorbed solution from a chemical wet scrubber by hot spring and marine algae." Chemosphere **66**(5): 878-886.

Hyka, P., S. Lickova, et al. (2013). "Flow cytometry for the development of biotechnological processes with microalgae." Biotechnol Adv **31**(1): 2-16.

Jones, A. K., M. E. Rhodes, et al. (1973). "The use of antibiotics to obtain axenic cultures of algae." British Phycological Journal **8**(2): 185-196.

Khozin-Goldberg, I., U. Iskandarov, et al. (2011). "LC-PUFA from photosynthetic microalgae: occurrence, biosynthesis, and prospects in biotechnology." Applied microbiology and biotechnology **91**(4): 905-915.
Kilian, O., C. S. Benemann, et al. (2011). "High-efficiency homologous recombination in the oil-producing alga Nannochloropsis sp." Proc Natl Acad Sci U S A **108**(52): 21265-21269.
Krawetz, S. A. and D. D. Womble (2003). Introduction to bioinformatics: a theoretical and practical approach, Humana PressInc.
Lanyon-Hogg, T., S. L. Warriner, et al. (2010). "Getting a camel through the eye of a needle: the import of folded proteins by peroxisomes." Biol Cell **102**(4): 245-263.
Lichtman, J. W. and J.-A. Conchello (2005). "Fluorescence microscopy." Nature Methods **2**(12): 910-919.
Lingner, T., A. R. Kataya, et al. (2011). "Identification of novel plant peroxisomal targeting signals by a combination of machine learning methods and in vivo subcellular targeting analyses." The Plant Cell **23**(4): 1556-1572.
Lommer, M., M. Specht, et al. (2012). "Genome and low-iron response of an oceanic diatom adapted to chronic iron limitation." Genome Biol **13**(7): R66.
McDaniel, H. R., J. B. Middlebrook, et al. (1962). "Isolation of Pure Cultures of Algae from Contaminated Cultures." Appl Microbiol **10**(3): 223.
McFadden, G. I. (2001). "PRIMARY AND SECONDARY ENDOSYMBIOSIS AND THE ORIGIN OF PLASTIDS." Journal of Phycology **37**(6): 951-959.
Radakovits, R., R. E. Jinkerson, et al. (2012). "Draft genome sequence and genetic transformation of the oleaginous alga Nannochloropis gaditana." Nat Commun **3**: 686.
Reumann, S. (2011). "Toward a definition of the complete proteome of plant peroxisomes: Where experimental proteomics must be complemented by bioinformatics." Proteomics **11**(9): 1764-1779.
Reumann, S., D. Buchwald, et al. (2012). "PredPlantPTS1: A Web Server for the Prediction of Plant Peroxisomal Proteins." Front Plant Sci **3**: 194.
Reumann, S., C. Ma, et al. (2004). "AraPerox. A database of putative Arabidopsis proteins from plant peroxisomes." Plant Physiol **136**(1): 2587-2608.
Rocha, J. M., J. E. Garcia, et al. (2003). "Growth aspects of the marine microalga Nannochloropsis gaditana." Biomol Eng **20**(4-6): 237-242.
Rodolfi, L., G. Chini Zittelli, et al. (2009). "Microalgae for oil: Strain selection, induction of lipid synthesis and outdoor mass cultivation in a low-cost photobioreactor." Biotechnology and Bioengineering **102**(1): 100-112.
Sensen, C. W., K. Heimann, et al. (1993). "The production of clonal and axenic cultures of microalgae using fluorescence-activated cell sorting." European Journal of Phycology **28**(2): 93-97.
Sezonov, G., D. Joseleau-Petit, et al. (2007). "Escherichia coli physiology in Luria-Bertani broth." J Bacteriol **189**(23): 8746-8749.
Shaner, N. C., P. A. Steinbach, et al. (2005). "A guide to choosing fluorescent proteins." Nat Methods **2**(12): 905-909.
Shinozaki, A., N. Sato, et al. (2009). "Peroxisomal targeting signals in green algae." Protoplasma **235**(1-4): 57-66.
Solovchenko, A., I. Khozin-Goldberg, et al. (2011). "Stress-induced changes in optical properties, pigment and fatty acid content of Nannochloropsis sp.: implications for non-destructive assay of total fatty acids." Mar Biotechnol (NY) **13**(3): 527-535.
Sprecher, H. (2000). "Metabolism of highly unsaturated n-3 and n-6 fatty acids." Biochim Biophys Acta **1486**(2-3): 219-231.
Stein, J. R. (1980). Handbook of phycological methods: culture methods and growth measurements, Cambridge University Press.

Trachana, K., T. A. Larsson, et al. (2011). "Orthology prediction methods: a quality assessment using curated protein families." Bioessays **33**(10): 769-780.
Tyler, B. M., S. Tripathy, et al. (2006). "Phytophthora genome sequences uncover evolutionary origins and mechanisms of pathogenesis." Science **313**(5791): 1261-1266.
Varun, S. (2005). Dictionary of Botany, Isha Books.
Vieler, A., G. Wu, et al. (2012). "Genome, functional gene annotation, and nuclear transformation of the heterokont oleaginous alga Nannochloropsis oceanica CCMP1779." PLoS Genet **8**(11): e1003064.
Wellburn, A. R. (1994). "The Spectral Determination of Chlorophylls a and b, as well as Total Carotenoids, Using Various Solvents with Spectrophotometers of Different Resolution." Journal of Plant Physiology **144**(3): 307-313.
Wheeler D, B. M. (2007). BLAST QuickStart: Example-Driven Web-Based BLAST Tutorial. Totowa (NJ), Humana Press.

7.Abbreviations

aa	amino acid
amp	ampicillin
bp	base pair
cDNA	complementary
DNA	deoxyribonucleic acid
dNTP	deoxynucleoside triphosphates
E. coli	Escherichia coli
ER	endoplasmic reticulum
EYFP	enhanced yellow fluorescent protein
ROS	reactive oxigen spesies
kbp	kilobase pair
LB medium	Luria-Bertani medium
PCR	polymerase chain reaction
PTS	protein targeting signal
Ta	annealing temperature
Taq	*Thermus aquaticus*
Tm	melting temperature
UV	ultraviolet
LC-PUFA	long chain polyunsaturated fatty acids
FA	fatty acids
v/v	volume to volume
RED	red fluorescent protein
IBA	Indole-3-butyric acid
IAA	Indole-3-acetic acid
ALA	α-linolenic acid
ARA	arachidonic acid
EPA	eicosapentaenoic acid
DHA	docosahexaenoic acid
Ch a	Chlorophyll a

8.Appendix

Table A. OD at 540 nm versus number of cells/ml for *N. salina*, *N. oceanica*, and *P. tricornutum*.

	N. salina		***N. oceanica***		**P. *tricornutum***	
Days	OD	Conc. (cells/ml)	OD	Conc. (cells/ml)	OD	Conc. (cells/ml)
2	0.021	1.20E+05	0.031	1.25E+05	0.050	0.00E+00
3	0.053	4.87E+05	0.080	4.13E+05	0.141	8.21E+05
10	0.510	1.29E+07	0.662	1.87E+07	0.548	4.40E+06
13	0.646	1.81E+07	0.809	2.15E+07	0.655	5.23E+06
16	0.812	1.57E+07	0.968	2.57E+07	0.769	6.30E+06
19	0.972	1.69E+07	1.131	2.63E+07	0.890	6.34E+06
21	1.058	1.69E+07	1.230	3.14E+07	0.968	7.39E+06
26	1.201	1.88E+07	1.443	3.10E+07	1.111	8.96E+06
29	1.319	2.18E+07	1.543	3.09E+07	1.211	7.22E+06
33	1.475	2.32E+07	1.820	3.08E+07	1.319	8.40E+06
36	1.667	2.34E+07	1.920	3.03E+07	1.493	7.13E+06

Table B. The *N. gaditana* proteins predicted by three prediction models.

Proteins predicted by all three models marked in green, those predicted only by the Plant model are red.

	Plant		Plant model Metazoa		Combined			
hit	**score**	**prob.**	**score**	**prob.**	**score**	**prob.**	**motif**	**sequence**
1	1.16	1.000	1.46	0.99	1.46	0.99	SANAETISRPSSRL	Nga05502.01 gene
2	1.04	1.000	1.37	0.99	1.51	1.00	GLSGKKGEKARAKL	Nga04682 maker-N
3	1.01	1.000	1.28	0.98	1.41	0.99	GRGRTVGGMPLARL	Nga06148 maker-r
4	0.98	1.000	1.16	0.90	1.38	0.99	EMKREGNKTLLSRL	Nga02873 augustu
5	0.97	1.000	1.08	0.75	1.32	0.98	PQWKVEEQTERARL	Nga04370.1 maker
6	0.96	1.000	1.25	0.96	1.55	1.00	YRNVIQDIQATSRL	Nga00060 maker-r
7	0.96	1.000	1.01	0.56	1.27	0.97	YRYIVDTTGALSRL	Nga06726 maker-r
8	0.96	1.000	1.50	0.997	1.47	0.995	KNSPVVQEALQSKL	Nga02565 maker-r
9	0.92	1.000	-	-	1.10	0.875	PPTHRDAPQRDSRL	Nga01349 maker-N
10	0.92	1.000	1.18	0.93	1.38	0.99	YATRDLHQGSKARL	Nga00919 maker-N
11	0.91	1.000	-	-	1.22	0.96	REGGRGSYIRPSRL	Nga01393 maker-N
12	0.91	1.000	1.76	1.00	1.60	1.00	AEKVGIGNTLKSKL	Nga01077 augustu
13	0.88	1.000	-	-	1.23	0.96	WSQGIWIRITRSRI	Nga05065 augustu
14	0.87	1.000	1.13	0.855	1.49	0.996	SGAGAARAGAISRL	Nga00886.01 make
15	0.87	1.000	1.19	0.933	1.43	0.993	HAPLEEVYSAASKL	Nga00170 genema
16	0.87	1.000	1.10	0.807	1.47	0.995	VVGMGGPASPQSRL	Nga30059 Transcri
17	0.85	1.000	1.05	0.677	1.40	0.991	SGGEFAKENGQSRL	Nga01291 augustu
18	0.83	1.000	-	-	1.32	0.981	KGSTGRGTGARARL	Nga00224.01 augu
19	0.83	1.000	1.54	0.998	1.39	0.990	QGAMLKETPTFSKL	Nga02369 genema
20	0.81	1.000	1.25	0.964	1.32	0.982	AGDVRIPHPRTSRL	Nga00291 maker-r
21	0.80	1.000	1.22	0.950	1.20	0.949	SSNPTVGDGDLSKL	Nga20601 Transcri
22	0.80	1.000	1.07	0.728	1.22	0.957	LGPILTAARPTARL	Nga02787 maker-r
23	0.78	0.999	1.24	0.960	1.41	0.992	HEGKEALTDTRARL	Nga04192 maker-c
24	0.78	0.999	-	-	1.26	0.970	LVPLPSPVPLIARL	Nga20954 Transcri
25	0.77	0.999	-	-	1.26	0.970	ATDILLFPISNSRM	Nga03561 maker-N
26	0.76	0.999	1.25	0.964	1.30	0.979	HIKFSRAENVLSRL	Nga01942 maker-r
27	0.75	0.999	1.44	0.995	1.39	0.990	KRKPAFRALSTSKL	Nga01381 maker-r
28	0.74	0.999	-	-	1.09	0.858	RLSPVVKPTAPPKL	Nga00077 augustu
29	0.72	0.998	1.09	0.795	1.41	0.991	EPERLLPSTTRARL	Nga05548 genema
30	0.69	0.997	-	-	1.03	0.767	SILDASSVSPPSSL	Nga20299 Transcri
31	0.69	0.997	-	-	1.41	0.992	TFKFPTRVQNLSRL	Nga01184 maker-N
32	0.67	0.997	1.10	0.813	1.15	0.917	LLTFLGTREDSSRL	Nga02546 augustu
33	0.67	0.996	-	-	0.96	0.621	RSMVTHVVSYQSKI	Nga00082 augustu
34	0.66	0.996	-	-	1.32	0.983	VLVVDGGILVKPRL	Nga04020.01 make
35	0.64	0.993	1.04	0.661	1.26	0.970	MGGKVKGSAPRSSL	Nga06021 maker-N
36	0.63	0.992	-	-	1.24	0.962	WQVRQWTVVAGSKI	Nga03919.01 make
37	0.63	0.991	1.30	0.980	1.17	0.933	AKGRIPACNEEAKL	Nga20777 Transcri
38	0.62	0.991	-	-	-	-	DWVALAYHICYRRL	Nga30623 Transcri
39	0.62	0.990	-	-	0.92	0.525	IARGVYIILEKLKL	Nga02192 maker-N
40	0.60	0.986	-	-	0.97	0.656	EAEEEEARRLRSSM	Nga30885 Transcri
41	0.58	0.980	-	-	-	-	GWKGWHPKTCSQRL	Nga00389.01 make
42	0.58	0.980	-	-	-	-	STMGPSIELDYSLL	Nga40012 rpl1: nt_
43	0.58	0.979	-	-	1.04	0.790	FSSCVRSSKRLSRI	Nga20070 Transcri
44	0.57	0.976	1.34	0.987	1.34	0.985	KAEYQKIVSGKAKL	Nga06551 maker-r
45	0.57	0.976	-	-	1.12	0.894	VAQDLEEKLKTPKL	Nga30826 Transcri
46	0.57	0.975	1.08	0.773	1.30	0.979	LEELMAVEAGAARL	Nga01960 maker-r

47	0.57	0.973	1.28	0.975	1.24	0.964	LLPVYGELPMKARL	Nga06128 maker-r
48	0.55	0.965	1.32	0.983	1.29	0.976	VTENALGWYNVAKL	Nga02557 maker-r
49	0.55	0.962	-	-	-	-	VAGAVSMETLPTRM	Nga20364 Transcri
50	0.55	0.962	-	-	1.38	0.989	YSGRYKWKNPAAKI	Nga00833 maker-c
51	0.55	0.961	1.29	0.978	1.30	0.978	TMEEMYERAEFAKL	Nga20492 Transcri
52	0.55	0.959	-	-	1.00	0.705	RAWKGNLVFESSKI	Nga06068.1 makeı
53	0.52	0.929	-	-	1.04	0.794	DLLKGVIGSFLPKL	Nga05405 maker-r
54	0.52	0.928	-	-	1.05	0.811	PALQAYGQALKASL	Nga04041 genema
55	0.52	0.924	-	-	-	-	HRVNREIERLRRRM	Nga02433 maker-r
56	0.51	0.920	-	-	0.94	0.572	WKLIDDQLKSKAHM	Nga30150 Transcri
57	0.50	0.900	-	-	-	-	RCLGPPSPSGPALL	Nga21015 Transcri
58	0.50	0.893	-	-	-	-	VVEDRTIIRRHYRL	Nga02862 maker-r
59	0.50	0.891	-	-	-	-	SVSWYFFRRHLSYL	Nga00181 maker-r
60	0.50	0.890	-	-	0.96	0.627	ASAAVEGGKMVAQL	Nga00380 maker-r
61	0.49	0.866	-	-	1.00	0.710	NRMKNLGRRQNSEL	Nga20710 Transcri
62	0.48	0.837	-	-	0.96	0.635	FEPQHLGYLHPSSL	Nga30662 Transcri
63	0.47	0.819	1.20	0.937	0.92	0.525	GRGDQSLTMELSSL	Nga05036.01 augu
64	0.46	0.744	-	-	1.10	0.875	DVGLSKAEVLKTRL	Nga06259 maker-N
65	0.45	0.737	-	-	0.98	0.667	LLMAGDCAWENARL	Nga03053 augustu
66	0.45	0.726	-	-	1.11	0.880	GTNTAPGKVIKIRM	Nga04183.1 makeı
67	0.44	0.680	-	-	0.95	0.605	FAGDGGKEMQAALL	Nga01170 maker-n
68	0.44	0.661	-	-	-	-	EAFFLFSEYLYDRL	Nga05974 maker-r
69	0.44	0.659	-	-	-	-	LSIVEACRSSRCRM	Nga20149 Transcri
70	0.44	0.639	-	-	1.01	0.742	TEIELYTGREAHKL	Nga04594 augustu
71	0.42	0.578	-	-	-	-	TWPKWPAQSVNLRL	Nga21079.1 Transc
72	0.42	0.559	-	-	-	-	HHPHPSLRGLPAQL	Nga04067 maker-r
73	0.42	0.553	-	-	0.97	0.652	VKSTSHSFPSVAKI	Nga20109.1 Transc
74	0.42	0.538	-	-	-	-	WDEPRAAQKKRNRM	Nga05243 maker-r
75	0.42	0.537	-	-	0.94	0.566	VDSDDFYLGRRNRL	Nga02847 maker-r
76	0.42	0.533	-	-	1.14	0.912	RGAALNAVIIAEKL	Nga00719 maker-r
77	0.42	0.529	-	-	-	-	GDVTQNYNPGYSAL	Nga04136.01 make
78	0.41	0.508	-	-	-	-	VEGAVHGSMRVATM	Nga20160 Transcri
Total	78.000	78.000	31.000	31.000	63.000	63.000		

Table C. Primers designed for *N. gaditana* predicted PTS1 proteins.
These primers were used for amplifying 10 *N. gaditana* proteins by PCR.

Acronym	Name	Size, bp	Primer name	Primer sequence	Tm	RE
NgP1.1	Nga_00060	1333	DK_1f_NgP1	aag gcg gcc gct ATG CAT GCT CGA TTA GGC	72.2	NotI
			DK_3r_NgP1	tatatctagacCTATAGCTTGGGGGGCGCTGT	69.5	SacII
NgP1.2	Nga_00060	349	DK_2f_NgP1	aag gcg gcc gct GGT ACT GGC AAT GGG GAT ACG	75.7	NotI
			DK_3r_NgP1	tatatctagacCTATAGCTTGGGGGGCGCTGT	69.5	SacII
NgP2.1	Nga_00077	2325	DK_4f_NgP2	aag gcg gcc gct ATG GGC ACC CTC CCG CTC	77.7	NotI
			DK_6r_NgP2	tatatctagacCTATAGCTTGGGGGGCGCTGT	69.5	XbaI
NgP2.2	Nga_00077	2030	DK_5f_NgP2	aag gcg gcc gct GCG GTC GTG CCG ATT GAA AAC	75.7	NotI
			DK_6r_NgP2	tatatctagacCTATAGCTTGGGGGGCGCTGT	69.5	XbaI
NgP3	Nga_00082	1152	DK_7f_NgP3	aag tct aga gta ATG GGT GCA GCG GAA GAG	68.1	XbaI
			DK_8r_NgP3	tatatctagacCTAGATCTTGGATTGGTACGA	64.4	XbaI
NgP4	Nga_00170	402	DK_9f_NgP4	aag gcg gcc gct ATG GAT TTG AGC ACA CAA	69.5	NotI
			DK_10r_NgP4	tataccgcggtTTACAACTTAGAGGCGGCGCT	70.8	SacII
NgP5	Nga_00224	1488	DK_11f_NgP5	aag gcg gcc gct ATG GCA GCA CAA CCC CGT	75.0	NotI
			DK_12r_NgP5	tatatctagacCTACAGTCGTGCGCGCGCACC	72.1	XbaI
NgP6.1	Nga_01077	1237	DK_13f_NgP6	aag Aga gct cta ATG GTA GTC CCT AAG CTT	65.4	SacI
			DK_15r_NgP6	tatatctagacTCACAATTTGGACTTCAGCGT	64.4	XbaI
NgP6.2	Nga_01077	616	DK_14f_NgP6	aag gcg gcc gct GTG GAC GAC TTA CAG CAT TGC	74.5	NotI
			DK_15r_NgP6	tatatctagacTCACAATTTGGACTTCAGCGT	64.4	XbaI
NgP7	Nga_02369	759	DK_16f_NgP7	aag gcg gcc gct ATG AAT AGT AAG TTT TGG	66.8	NotI
			DK_17r_NgP7	tatatctagatTCACAACTTCGAGAACGTCGG	65.6	XbaI
NgP8	Nga_02873	1482	DK_18f_NgP8	aag gcg gcc gct ATG ATC TCC GAC GAG GTG	73.6	NotI
			DK_19r_NgP8	tatatctagacCTATAAGCGCGATAACAAGGT	64.4	XbaI
NgP9	Nga_03053	2289	DK_20f_NgP9	aag gcg gcc gct ATG GTC AAC GAG AAT CCT	70.9	NotI
			DK_21r_NgP9	tatatctagacCTACAGTCGCGCGTTTTCCCA	68.2	XbaI
NgP10	Nga_05502	548	DK_22f_NgP10	aag gcg gcc gct ATG AGT CGC GCC GCT TTT	73.6	NotI
			DK_23r_NgP10	tatatctagatTCACAGTCGACTCGACGGTCG	68.2	XbaI

Table D. The list of acronyms for the proteins identified in this study

LACS	long chain acyl-CoA synthetases
ACX	acyl-coenzyme A oxidase
AIM	*A. thaliana* peroxisomal multifunctional protein
EHHADH	enoyl hydratase/3-hydroxyacyl CoA dehydrogenase
PKT/KAT	peroxisomal 3-ketoacyl thiolase
SCP	sterol carrier protein
DECR	2,4-dienoyl-CoA reductase
DCI	dodecenoyl-CoA isomerase
ECH	enoyl-CoA hydratase
IBR	in-between-ring domain containing protein
SDR	short-chain dehydrogenase/reductase
AAE	acyl-activating enzyme
GR	glutathione reductase
PRX	peroxyredoxine
CS	citrate synthase
ICL	isocitrate lyase
MS	malate synthase
MDH	NAD malate dehydrogenase
GOX	glycolate oxidase
GGT	gamma-glutamyltranspeptidase
G6PD	glucose-6-phosphate dehydrogenase
NS	naphthoate synthase
BADH	betaine aldehyde dehydrogenase
TTL	transthyretin-like protein
ASP	aspartate aminotransferase