Monitoring of eicosapentaenoic acid (EPA) production in the microalgae *Nannochloropsis oceanica*

Marta Sá [1,2], Narcís Ferrer-Ledo [2], Rene Wijffels [2], João G. Crespo [1], Maria Barbosa [2,3], Claudia F. Galinha [*1]

1 LAQV-REQUIMTE, Chemistry Dept, FCT, Universidade Nova de Lisboa, 2829-516 Caparica, Portugal

2 Bioprocess Engineering, Wageningen University and Research, 6708PB Wageningen, Netherlands

3 Department of Biology, University of Bergen, PO Box 7803, 5006 Bergen, Norway

*corresponding author: cf.galinha@fct.unl.pt

Abstract

With the increase awareness for a healthier food regime and greener environmental processes, microalgae are being looked as a solution for a sustainable production of polyunsaturated fatty acids, such as omega-3 eicosapentaenoic acid (EPA). *Nannochloropsis oceanica* is an oleaginous microalga, well-known for the ability of EPA accumulation, although higher lipid productivities are still required to make the process competitive. Therefore, three cultivation parameters were tested in the present work (temperature, light cycles and nitrogen supply) in order to study the EPA profile in the polar and neutral fractions of the cells. In addition, an online monitoring tool based on a fluorescence spectroscopy technique was developed with the aim of increasing process knowledge at real time. The results of this work show that nitrogen depletion induces the highest variability in EPA accumulation in the neutral fraction (triacylglycerols). However, to increase the EPA content in the polar fraction a different strategy needs to be implemented, such as decreasing the cultivation temperature or the light available per cell. Chemometric models were developed through PCA (Principal Component Analysis) and PLS (Projection to Latent Structures), using only fluorescence spectra as inputs, enabling the monitoring of EPA in both fractions separately. High explained

variance was observed (above 85%) in both fractions, with R^2 above 0.81 and slopes above 0.93 for both validation and training data sets. Lower values of cross-validation and prediction errors were observed (between 0.29 and 0.49 % g/g_{DW}). The results obtained show that fluorescence spectroscopy is a powerful technique for online monitoring of non-fluorophore molecules, such as EPA, in complex process like microalgae cultivation.

Keywords: eicosapentaenoic acid (EPA); omega-3 fatty acid; Nannochloropsis oceanica; Fluorescence spectroscopy; Chemometric modelling.

1. Introduction

The importance of long-chain polyunsaturated fatty acids (PUFA), such as omega-3 (ω-3), has been extensively studied in the past years with the increased concern in the western world for a better and more equilibrated food regime. This class of lipids proved to have several pharmaceutical and nutraceuticals applications [1–3] and since they are essential for humans and most animals, and neither have the capacity to produce them, food and feed are considered the main vehicles for their supply [4]. Among the ω-3 fatty acids present in Nature, eicosapentaenoic acid (EPA, 20:5 ω-3), plays an essential role in long term health benefits of cardio and immune system [1–3,5].

The main source of EPA is fish and krill oil. To satisfy the human requirements, the World Health Organisation (WHO) advices a dietary intake of fish oil once or twice per week [6,7]. However, this solution is not sustainable in a long term since the supply of fish and krill oil is limited. As an alternative, marine photosynthetic organisms, like microalgae, are being regarded as a solution in aquaculture and terrestrial livestock feed as well as in human supplements, since they are the primary producers of PUFAs like EPA [2–4,8,9]. Microalgae can accumulate lipids in two distinct fractions, the polar lipids fraction (PL), mainly glycolipids and phospholipids; and the neutral lipids fraction (NL), in the form of triacylglycerols (TAG) [1,3,4]. There is some ongoing discussion about which of these two fractions is the best carrier of the EPA in food and feed, with some authors defending the TAG fraction [2] and others the PL fraction [10].

Several strategies are used to increase the lipid content in microalgae cells, namely nutrient limitation (nitrogen or phosphorous deprivation), high salinity, temperature, and high light intensity [1,3,4,11,12]. It is well known that under nitrogen limitation, TAG concentration increases [3]. For example, *Nannochloropsis* sp., a well-known oleaginous microalgae specie, can accumulate between 25 to 45% of total fatty acids under photoautotrophic conditions, among them the EPA [3–5]. It is also known that under different environmental conditions, the content and composition of fatty acids in PL and TAG fractions can vary substantially [1,5].

Using microalgae biomass has great advantages, such as the possibility to grow in non-arable land, using sea water and residual nutrients [2–4]. However, the production of lipids from microalgae

biomass still faces some challenges. Although *Nannochloropsis* genus is considered to be a model organism for lipid production [12], EPA content is still low, up to 4.3% on dry weight basis in *N. gaditana* [1,3,9] and between 2.7 and 5.2% in *N. oceanica* [3]. To make the process economically competitive, higher lipid productivities are required to decrease production costs [9] associated with high energy requirements for water management, and for lipid extraction of the biomass [2,4].

When aiming for the production of fatty acids, specially EPA, a real time and online monitoring of the culture would bring great advantages. Fatty acid analysis is known for being a laborious method, involving several steps: extraction, separation into different fractions, methylation and quantification. Different organic solvents are needed in this method, making the process non-green, and most of these steps are time consuming. Thus, the development of an online monitoring tool will enable to understand the effect of the process parameters in the product accumulation at real time, allowing the possibility to take important decisions in the moment, such as harvesting the culture when the maximum product content is achieved. To achieve this goal the development of a sensitive probe is of great importance for the overall economic efficiency of the microalgae production and biorefinery. Several spectroscopic techniques have been reported in the literature for the online monitoring and control of bioreactors, such as fluorescence spectroscopy, due to the possibility of tracking different metabolites simultaneously (substrates and products), and being a non-invasive and non-destructive tool. Fluorescence spectroscopy detects the presence of natural fluorophores of the media (extrinsic) and in the cells (intrinsic). The interaction between the two is rather complex, and for this reason, chemometric methods are often needed to deconvolute and find correlations between the concentration of substrates and products and the fluorescence captured [13–15].

In the present study*, N. oceanica* was cultivated in highly controlled lab-scale photobioreactors under different environmental conditions in order to better understand the effect of process condition on EPA accumulation and to develop a monitoring tool specific for this fatty acid. Different temperatures (15, 20, 25 and 30 ºC), light cycles (24 hours of light or a day/night cycle of 16 hours: 8 hours) and nitrogen supply (with or without) were tested. The EPA content was analysed in the PL and TAG fractions. To explore the possibility of using fluorescence spectroscopy as a monitoring tool for EPA production, excitation-emission matrices (EEMs) of the fluorescence spectra were acquired

during *N. oceanica* cultivation, and chemometric tools such as PCA (Principal Component Analysis) and PLS (Projection to Latent Structures) were used to develop prediction models.

2. Material and Methods

2.1. Strain, cultivation and pre-culture conditions

Nannochloropsis oceanica was provided by NECTON, S.A. (Olhão, Portugal). Pre-cultures of *N. oceanica* were incubated in 250 mL Erlenmeyer flasks with 100 mL of culture, at 25 ºC, in an orbital shaker (90 rpm) with an incident light of 100 $\mu mol.m^{-2}.s^{-1}$ and 0.2% CO_2. The cultures were kept under a day/night cycle light regime (16:8 hours) until further use. The cultivation media is constituted by natural sea water (Eastern Scheldt, the Netherlands) and filtered (0.2 µm) before use, supplemented with 10.7 mM of $NaNO_3$ and 0.535 mM of KH_2PO_4. A nutrient solution was added, Nutribloom from Phytobloom®, containing: 4 µM $ZnSO_4.7H_2O$, 2 µM $MnCl_2.4H_2O$, 0.2 µM $Na_2MoO_4.2H_2O$, 0.2 µM $CoCl_2.6H_2O$, 0.2 µM $CuSO_4.5H_2O$, 52.8 µM $Na_2EDTA.2H_2O$, 4 µM $MgSO_4.7H_2O$ and 40 µM $FeCl_3.6H_2O$. HEPES buffer (20 mM) was added and the pH was set to 7.8 with NaOH 5M. The medium was directly filtered sterilized by Sartobran® Capsule 0.2 µm into the Erlenmeyer flasks or the bioreactor vessel. Prior to the photobioreactor inoculation, inoculums were placed under day/night cycle or continuous light, according to the light regime of the experiment.

2.2. Photobioreactor setup

All experiments were performed in a flat-panel airlift-loop photobioreactor, heat-sterilized, with a light path of 20.7 mm and working volume of 1.8 L (Labfors 5 Lux, Infors HT, Switzerland, 2010). The pH was controlled at 7.8 by CO_2 injection, and the culture is homogenised by filter sterilized air with a flow rate of 1L/min. The temperature was controlled by a water-jacket in direct contact with the cultivation vessel. The incident light was provided by 260 LED lamps (28V, 600 Watt) on the culture side of the photobioreactor, with warm white spectrum (450 – 620 nm), and the back side was covered to prevent interference of the ambient light. The incident light started at 200 $\mu mol.m^{-2}.s^{-1}$ and was increased to 636 $\mu mol.m^{-2}.s^{-1}$ when the back light reached 50 $\mu mol.m^{-2}.s^{-1}$. The light supplied

corresponds to the light experienced in a Dutch summer day, consisting of 16 hours of light supplied with a sinus function with solar noon at 1500 µmol.m^{-2}.s^{-1} [16].

2.3. Photobioreactor operation conditions

In this study eight experiments were performed, in batch mode, following a two-step approach, except the control experiment.

In the first step, bioreactors were inoculated at a cell concentration between 1.0 and 1.5 x10^7 cells.mL^{-1} and run under replete nitrogen conditions. This was followed by a second step with nitrogen starvation. Briefly, the bioreactors were emptied, the biomass was centrifuged (2500 rpm for 15 minutes) and washed with free-nitrogen medium, and the bioreactor was then refilled with the culture and free-nitrogen medium to prevent limitation of other nutrients, until a specific light supply rate of 1x10^{-13} µmol.cells^{-1}.s^{-1} (called N-starvation experiments).

The experiments are divided in two categories according to the light regime – continuous light (24 hours) or day/night (d/n) cycle (16:8 hours) (Figure 1). In the continuous light experiments, four temperatures were studied – 15, 20, 25 and 30 ºC, for a period of four days, starting at N-starvation (Batch names, respectively: "*15*", "*20*", "*25*" and "*30*"). The temperature was set in the beginning of the experiment and maintained during nitrogen depletion. In the d/n cycle experiments, three batches were performed to study the effect of N-starvation and sudden temperature decrease, versus a control batch, and followed for ten days. The first step of the cultivation was performed at 25 ºC, and in the second step the culture was either submitted to nitrogen-deplete medium (Batch name: "*N-starv*") or the temperature was decrease to 15 ºC (Batch name: "*25-15*"). In the control experiment, the bioreactor was inoculated and followed for ten days (Batch name: "*Control*").

2.4. Offline measurements

Biomass concentration was assessed by measuring dry weight (DW) and cell concentration. Dry weight was measured in triplicates, as described by Kliphuis *et al* [17], and 0.5 M ammonium formate was used to remove the salts from the culture. Cell concentration and cell size distribution was

measured in duplicates, using Isotone II diluent to dilute the samples, in a Multisizer II (Beckman Counter) using a 50 μm aperture tube.

Biomass volumetric production rate was calculated according to the following equation (Equation 1), where *DW(t)* and *DW(0)* correspond to the dry weight measured for day *t* and day *0*, respectively, and *t* is the time of the experiment (4 or 10 days):

$$Biomass\ volumetric\ porduction\ rate\ (g.L^{-1}.d^{-1})) = \frac{DW(t) - DW(0)}{t}$$

Lipid composition of *N. oceanica* was measured during the entire period of the cultivation, before and after the point of stress induction by nitrogen starvation or temperature decrease. Biomass samples were centrifuged, washed with 0.5 M ammonium formate, and stored at -20 ºC until lyophilisation. Lipids were extracted, separated into triacylglycerol (TAG) and polar (PL) fractions, and quantified as reported by Breuer *et al* [18] and Leon-Saiki *et al* [19]. Briefly, 10 mg of lyophilized biomass was disrupted with a beat beater and lipids were extracted with chloroform:methanol (1:1.25, v:v) containing the internal standards for TAG and PL fractions, 170 μg.mL^{-1} of tripentadecanoin (9:0) and 170 μg.mL^{-1} of 1,2-dipentadecanoyl-sn-glycero-3-[phosphor-rac-(1-glycerol)] (sodium salt) (15:0) respectively. TAG and PL fatty acids were separated by different elution solvents, hexane:diethylether (7:1, v:v) and methanol:acetone:hexane (2:2:1, v:v) respectively, in a SPE silica gel column (Sep-Pak Vac 6cc, Waters). Both fractions were methylated and quantified by gas chromatography (GC-FID). The results are expressed in g_{EPA}/g_{DW}.

Fluorescence spectroscopy excitation-emission matrices (EEMs) were measured in a Shimadzu RF-6000 spectrofluorophotometer with a cuvette. Each analysis takes around 5 minutes, where no cell sedimentation was observed. The spectra were acquired through an excitation wavelength range between 250 and 790 nm, with 5 nm steps, and emission wavelength range between 260 and 800 nm, also in 5 nm steps. Excitation and emission monochromator slit widths were 3 nm, with a scan speed of 12000 nm/min.

2.5.   Chemometric models development

The development of the chemometric models included three steps:

1) <u>Spectra pre-treatment</u>: water scatter peaks, like Rayleigh scatter of first and second order, can be a source of interference when aiming for a quantitative analysis using 2D fluorescence spectra. This wavelength-dependent scatter (peak emission ± 10 nm at each excitation wavelength) was removed with the use of an algorithm developed by Bahram *et al* [20], and replaced by interpolation of the surrounding data points. The program is available at www.models.kvl.dk.

2) <u>Principal Component Analysis</u>: through PCA is possible to compress and reduce the information in the EEMs, with minimal loss of information, by dividing the initial data into *n* linear combinations that have to follow two main rules, be uncorrelated and be ordered according to the explained variance they captured. The first principal component (PC1) will capture maximum variance in a certain direction (axis); then PC2 will be orthogonal to PC1 and will capture less variance; and so on. In total, 73 fluorescence spectra were compressed into 20 PCs, capturing more than 99% of the variance.

3) <u>Projection to Latent Structures modelling</u>: in order to correlate the fluorescence PCs (inputs) with the EPA concentration in TAG and PL fractions of *N. oceanica* (outputs), multivariate statistical modelling was used, namely PLS. In PLS modelling two sub data sets were created, one for training (calibration) the model, i.e. to build the function that better correlates the outputs with the inputs, and another for validation (prediction), to test the quality of the model to predict a new data set. Two validation approaches were studied, first using each batch at a time (batch-by-batch), and secondly using a random data set corresponding to 25 % of the total data. The cross validation (CV) of the models was performed with the remaining data, used for calibration (seven batches in the case of batch-by-batch validation, or random 75 % of all data). Shortly, several models were created using an independent set of data, selected by leave one out strategy (LOO) and repeatedly evaluating the errors of the models (RMSECV). Not all the PCs provided as inputs are required. Thus, to select the useful PCs for each model, an iterative stepwise elimination (ISE) was used [21]. To assess the quality of the models several parameters are evaluated, such as the variance captured (%), the root mean square error of cross-validation (RMSECV) and prediction (RMSEP), and the R^2 and slopes of the validation and training sets.

All multivariate statistical analysis were performed using n-way toolbox for MATLAB - MathWorks®
[22].

3. Results and Discussion

In this work, the effect of several environmental conditions on the accumulation of EPA was studied in apolar (TAG) and polar (PL) fractions. The environmental conditions studied were light – continuous light (24 h) or day/night cycles (16 h : 8 h); temperature – low temperatures (15 and 20 ºC) and high temperatures (25 and 30 ºC); and nitrogen starvation. Samples were taken along the entire length of the experiments to measure fatty acid profile and obtain the EEMs.

To evaluate the effect of different stress factors on EPA accumulation, the fatty acid profile of the culture was compared in the beginning of the "stress phase", by nitrogen depletion or decrease of temperature, until the end of the batch, in a total of 4 days for 24 h light experiments, and 10 days for d/n cycle experiments. To monitor the evolution of the EPA during the experiments, models were performed with all the samples acquired through each batch, before and after the "stress phase". This strategy enables a higher data set to be used to calibrate and validate the models and also captures the variability expected in these experiments, from inoculation until EPA production.

3.1. Biomass concentration and cell size

In the experiments where light was provided during 24 h (Figure 2A), no differences were found in the biomass concentration between cultivation at 20, 25 and 30 ºC after 4 days of nitrogen depletion. The biomass volumetric productivities were on average 0.9 $g_{DW}.L^{-1}.d^{-1}$ for 20 and 30 ºC, and 1.0 $g_{DW}.L^{-1}.d^{-1}$ for 25 ºC. As expected, the biomass volumetric production rate was lower in the 15 ºC batch, 0.3 $g_{DW}.L^{-1}.d^{-1}$.

In the d/n cycle experiments (Figure 2B), the biomass concentration shows a similar profile in all three experiments for the first 4 days. However, nitrogen depletion seems to have a negative effect on the final biomass concentration than decreasing the temperature from 25 to 15 ºC. The biomass volumetric production rates were on average 0.5, 0.8 and 0.9 $g_{DW}.L^{-1}.d^{-1}$ for nitrogen starvation,

decrease in temperature and control experiment, respectively. In the decreased temperature and in the control experiments, no differences were found in the final biomass concentration.

The environmental conditions imposed to the culture also had an influence on the cell size (Table 1). For all experiments performed with 24 h of light and nitrogen depletion, it is possible to notice an increase in cell size. Nitrogen starvation is one of the strategies often used to increase the TAG concentration in microalgae cells, which is done in lipid bodies leading to an increased volume. Only in the experiment performed at 15 ºC (when the optimal growth temperature is around 25 ºC [23]) the final cell size was slightly lower than in the beginning of the starvation phase, indicating that the culture was probably already accumulating TAG when the nitrogen depletion was performed. This result can also explain why the biomass dry weight was higher in the beginning of the starvation phase of this experiment (d0) when compared with the other temperatures tested.

In the d/n cycle experiments, when the temperature was decreased (from 25 to 15 ºC), the cell size increased in the first two days (to 3.5 µm), but immediately after that it decreased to 3.00 and remained stable until the end of the batch. Cell concentration was constant on the first four days, only increasing after that (data not shown). This increase in the cell size can be explained by the accumulation of fatty acids inside lipid bodies, as an acclimation to the decrease in temperature, and a consequent decrease in the growth rate. The cells of the control batch without starvation or any other stress factor, increased size gradually throughout the experiment, accompanied by an increase in cell concentration. The process of increasing the amount of membrane surface was reported before as an adaption to low light conditions, as the one observed in the control batch, in order to increase the photosynthetic capacity to capture more light [9].

When comparing the two experiments with nitrogen depletion at 25 ºC and different light regimes, the same final cell size was achieved in the end of the experiments (d4 and d10), although the cell size increased faster for the culture exposed at 24 h of light as expected.

3.2.   EPA production

Nitrogen starvation, a well-known methodology to induce lipid accumulation in microalgae lipid bodies, was performed at four different temperatures, 15, 20, 25 and 30 ºC (Figure 3A).

For lower temperatures (15 and 20 ºC) it is possible to notice a higher content of EPA in the TAG fraction in the beginning of the starvation (d0) than for higher temperatures (25 and 30 ºC). In fact, the EPA content in TAG fraction is inversely proportional to the temperature, i.e., higher contents were observed in lower temperatures, and vice-versa. This might be due to the combined effect of temperature and light per cell, since at lower temperatures the biomass productivity decreases, which means that the culture will take longer to multiply, so more light will be available per cell. As a result, TAG content was already higher before starting the nitrogen starvation phase. Yet, when comparing the EPA accumulation in the TAG fraction during the four days of nitrogen starvation (final content of EPA compared to the initial), higher temperatures led to higher TAG productivities. For 30 ºC, the EPA accumulated in the TAG fraction was 0.017 g_{EPA}/g_{DW}, 0.012 g_{EPA}/g_{DW} for 25 ºC, 0.007 g_{EPA}/g_{DW} for 20 ºC and 0.005 g_{EPA}/g_{DW} for 15 ºC. This effect of increase long chain FA accumulation with high temperatures was also noticed by Ördög *et al.* in three *Chlorella* strains [24].

As studied before [9], nitrogen depletion leads to an increase of TAG content associated with a decrease in the PL fraction. One of the mechanisms proposed is the *de novo* synthesis of TAG by conversion of the membrane lipids, which consists mainly of PL. In this study it is possible to confirm that the increase observed in the TAG fraction was accompanied by a decrease in the PL fraction.

The decrease in the temperature was reported to increase the EPA content in the cell membranes, especially during growing conditions [1,25]. Lipid accumulation occurs during the day period, while cell division occurs at night [26,27]. Aiming to increase the EPA content in the PL fraction of *N. oceanica*, a second set of experiments were performed, keeping the culture under d/n cycle: nitrogen depletion (N-starv), temperature decrease (from 25 to 15 ºC) and a control batch, where the algae were allowed to grow without any imposed stressed (Figure 3B).

In the beginning of the experiments, N-starv and control batches started with a similar EPA content in the TAG fraction. Nevertheless, after ten days of nitrogen depletion media, the culture accumulated a large amount of EPA in the TAG fraction at the expense of the content in the PL fraction (Figure 3B, N-starv d0 and d10). These results confirm what was stated by other authors regarding the effect of nitrogen starvation [9,12,28]. The effect of lowering the temperature resulted in a slight increase of EPA in PL fraction, accompanied with a decrease in the TAG fraction (Figure 3B, 25-15º d0 and d10).

According to previous studies in *Nannochloropsis salina*, *Phaeodactylum tricornutum* and *Chlorella sp.* [3,29,30], temperature reduction increased the content of EPA and PUFA's, due to the need of increasing membrane fluidity. However, these studies point to previous research [31] where the FA were analysed in total lipid profile of the whole cell, and no distinction was done between TAG or PL fractions. The highest content of EPA in the microalgae membranes (PL fraction) was achieved in the control batch, where the temperature was maintained at 25 °C with d/n cycle (Figure 3B, d0 and d10). In this experiment, the EPA content in TAG fraction decreased after ten days from 0.005 g/g_{DW} to 0 g/g_{DW}.

Although there is no difference on the final dry weight between the control experiment and the decreased temperature (Figure 2), the cell size may be the contribution for the higher content of the EPA in PL fraction in the control batch. A higher cell concentration was achieved (data not shown), meaning that less light was available per cell. To adapt to this environmental condition, the microalgae increase the plastid membrane intending to increase the photosynthetic apparatus, and consequently the EPA content in this fraction increases [9]. And, as mentioned before, a cell size increase was noticed in this experiment (Table 1).

The experiments performed resulted in several scenarios in the accumulation of EPA by *N. oceanica*. Nitrogen depletion enables the accumulation of EPA in lipid bodies (TAG fraction) independently of the temperature, although higher temperatures led to higher accumulation. For nitrogen replete cultivation conditions, light played an essential role. Although the incident light was the same in all experiments, the biomass concentration of the control batch was higher, so the cells perceive lower light, and this led to an increase of EPA in PL fraction. Also, temperature decrease might slightly increase the EPA content in the PL fraction. When performing nitrogen depletion at 25 °C, a similar EPA final content in the biomass was reached for both 24h and d/n cycle batches, with the main difference being the time needed.

3.3. EPA monitoring

The experiments in this study led to several responses in *N. oceanica* biomass regarding cell concentration, size and physiological state (from non-stressed green cells, to yellowish when TAG synthesis is induced). This variability in the cells originates fluorescence EEMs with different characteristics. Changes in the media composition, such as nitrate concentration, were also reported in other microalgae cultivations to originate a different fluorescence profile [32]. Also, the variability of EPA content that can be found in the two fractions of the cell, apolar (TAG) and polar (PL), increases the complexity of monitoring such product. Depending on the commercial destination of the lipid enriched biomass, having information about the content and location of the EPA in the cell can be extremely useful namely for its recovery. Thus, a tool that can distinguish the content of EPA in both fractions of *N. oceanica* cells, simultaneously, was developed.

When aiming for industrialisation, a batch-by-batch validation approach seems logical, since the final model acquired with these experiments will then be validated with new data acquired from a new batch. However, the experiments of this work were thought to lead to a wider range of EPA concentrations, to be able to have broader range of scenarios. For that reason, some of the batches are not representative of a cultivation to produce EPA but were important to acquire concentration points in a lower or higher data ranges. Thus, a random data set was also used to create a general model, that would be more suitable for future use.

The prediction models are represented in a graphic (Figures 3 and 4 for the TAG fraction; Figures 5 and 6, for the PL fraction) where the observed values (y-axis) are plotted against the predicted values by the model (x-axis). Two parallel lines were added to each graphic corresponding to two times the standard deviation of all the experimental data acquired. Points represented outside these lines may be considered outlier observations.

Model parameters including variance explained (%), root mean square error of cross-validation (RMSECV) and prediction (RMSEP), validation and training R^2 and slopes, and the number of inputs selected, are shown in Tables 3 and 4 for TAG and PL fractions, respectively. The quality of a model can be described by a high explained variance, with coefficient of determination (R^2) and slope near to 1. Lower values of root mean square errors are preferable, and the values between the cross-

validation and prediction errors should be close, which means that both data sets (validation and training) are representative of all data variability.

3.3.1. Prediction models for EPA in TAG fraction

As mentioned before, the batches performed aimed to acquire a wide range of EPA concentration values. In Figure 4 it is possible to see that the experiments of d/n cycle (the validation data in the upper three graphics in Figure 4) show different data distribution. For instance, the decrease in temperature did not increase the EPA content in the TAG fraction. A similar distribution was observed in the control batch. For that reason, the EPA concentration range observed in these two batches is limited, meaning that when these batches are used for validation, the combination of R^2 and slope of the validation set is low (Table 2). Also, for both models, a high number of inputs is necessary (16 for both models) to explain the variance captured (around 91 % for both models). For the experiment N-starv, the validation set has four outliers (of a total of ten data points), experimental points represented above the line of two times the standard deviation of all data. This means that, without the data points of this batch to train the model, the prediction ability decreases (52.6 % of variance explained), together with a low R^2 (0.26) for the validation set and with the highest root mean square errors (0.58 and 0.69 % g/g$_{DW}$ for RMSECV and RMSEP, respectively).

The models obtained using the experiments performed at 24 h of light and nitrogen starvation as validation data set were more consistent (Figure 4, four lower graphics). Nitrogen depletion is known to induce the TAG accumulation in the cells, giving a broad range of concentrations through each experiment. The variance captured by these models was above 80 %, with RMSECV ranging from 0.32 to 0.41 % g/g$_{DW}$ and RMSEP between 0.26 and 0.31 % g/g$_{DW}$ (Table 2). The fact that the values of RMSECV and RMSEP are close reveals that the validation data set is representative of values of the training set.

For the model using as validation set the experiment performed at 20 ºC, a lower R^2 was found for both validation and training set (0.64 and 0.80, respectively). When using the batch performed at 30 ºC as validation set, the training R^2 was also lower (0.81). This can be explained by the fact that in both models only seven PCs were selected to build the model, and some outliers can be found in the

14

data distribution. Nevertheless, the remaining models, the R^2 of the validation were above 0.81, and above 0.89 for the training set, revealing the robustness of the fluorescence spectroscopy as a technique to monitor EPA concentrations in TAG fractions of *N. oceanica*.

A general model was built using 25 % of the total data as a validation set, randomly chosen, and the remaining 75 % as training set. As it can be seen in Figure 5, the validation set is widely spread through the entire concentration range of EPA found in the TAG fraction. Thirteen PCs, of twenty inputs given, were selected to explain 92.1 % of the variance captured by the model, which means that the fluorescence spectroscopy is a robust method for capturing the variability found in these experiments. The RMSEC and RMSEP were close (0.30 and 0.29, respectively) and both validation and training R^2 and slopes were high (all above 0.87).

The variability found in the quality of the models shows the importance of the calibration data set used. Not only it is important to provide as much data as possible, but also that data should be representative of the several scenarios that can be found. Overall, fluorescence spectroscopy proved to be a strong tool to monitor EPA concentration in the TAG fraction of *N. oceanica*. EPA is a fatty acid and therefore it is known for not being a natural fluorophore. Fluorescence spectroscopy was reported to be a scanning technique able to detect natural fluorophores, but also the interaction of those with the medium and other non-fluorophores components can provide information about the matrix constitution. For this reason, the monitoring of a non-fluorophore molecule like EPA, reveals the potential of this technique to monitor other lipid components.

3.3.2. Prediction models for EPA in PL fraction

In Figure 6 it is possible to see that the experiments of d/n cycle (the validation data in the upper three graphics) show different patterns. In the control batch, where the culture was allowed to grow at optimal conditions (25 ºC, d/n cycle), at a certain point the light became a limiting factor. It was shown in literature that decreasing light per cell increases EPA content, due to the increase of cell membranes [9]. This justifies the fact that EPA concentration was the highest in this batch. Decreasing the cultivation temperature was also reported to increase the unsaturated fatty acids in the PL fraction

of the algae, and it is possible to observe a high content of EPA in this experiment. However, and as reported for the models of EPA in TAG fraction, this means that using these data as validation set (meaning they are not used for calibration) is not a correct approach. Both models reach 89.5 % of variance explained, but there is a wide gap between RMSECV and RMSEP, and the combination between R^2 and slope of the validation set was not ideal (Table 3). When using the N-starv batch as validation, six of the ten points used for validation are considered outliers. Also, while observed values of EPA content varied between 1 and 6 % g/ g_{DW}, no prediction was performed above 4 % g/ g_{DW}. As described previously in the model of the EPA in the TAG fraction, the quality of the model decreases when these data points are not used to build the prediction model. Although the variance explained was 93.1 %, the difference between RMSECV and RMSEP is the highest observed in these models, with the lowest validation R^2 (0.01) and slope (0.21) (Table 3).

In the models obtained for the batches performed at 24h light and nitrogen starvation, a wider range of EPA concentration was observed (Figure 6, four lower graphics). Except for the batch at 20 °C, the variance captured was above 85 %, with RMSECV ranging from 0.41 to 0.46 % g/g_{DW} and RMSEP between 0.31 and 0.47 % g/g_{DW} (Table 3). Using as validation set the data of the batch performed at 20 °C, more outlier values were observed, and lower number of PCs were selected as inputs, resulting in a model with lower variance explained (67.4%), higher difference between RMSECV and RMSEP (0.6 and 0.29 % g/g_{DW}, respectively) and the lowest R^2 of the training data set (0.67) (Table 3).

When using 25 % of random data as validation set (Figure 7), eleven of the twenty PCs were selected to explain 84.8 % of the variance and no outliers were observed. The quality of the model was high, with errors ranging between 0.37 and 0.49 % g/g_{DW} (RMSECV and RMSEP, respectively). Both R^2 of validation and training set were higher than 0.80 and slopes near 1.00.

As mentioned previously, a suitable calibration data set should be selected in order to capture the entire concentration range of EPA in the PL fraction, and thus, build a model representative of the different scenarios that can be encountered. The ability of the fluorescence spectroscopy to monitor EPA concentration in the cell membrane of *N. oceanica* is of high importance for the industrial application of this microalgae in the food and feed industry. It was described that the fatty acids

present in the microalgae membrane are more bioavailable for fish metabolism [10]. Having a tool able to determine EPA concentration at real time, with a fast and non-invasive methodology, allow the monitoring of the process as well as stirring it in order to have the desire amount of EPA in the biomass, improving the cultivation efficiency and thus the economic gain of the overall process.

4. Conclusions

For the first time, fluorescence spectroscopy coupled with chemometric tools was used to monitor EPA in microalgae biomass.

To enable the development of PLS prediction models, different environmental cultivation conditions were tested to induce different biological responses in the biomass profile and EPA accumulation. Stress factors, like nitrogen depletion and low light per cell, led to an increase in the cell size of *N. oceanica*. Nitrogen depletion led to an increase of EPA in the TAG fraction of the cells, while low light or decreased temperature lead to an increase of EPA in the PL fraction.

Statistical modelling combined with fluorescence spectroscopy is a step forward in the real monitoring of complex biological systems like microalgae cultivation. In this work, prediction models were developed that enable the monitoring of EPA content in both TAG and PL fractions of the cell. Two validation strategies were studied, batch-by-batch and random 25% of the total data, to access the importance of the batch conditions in the prediction capability of the method. The models developed demonstrate the potential of using fluorescence spectroscopy as a monitor tool of non-fluorophore molecules, such as EPA, revealing the potential of this technique to monitor lipid components. The use of such technology enables not only the possibility to monitor the cultivation system, but also of taking decisions at real time, empowering the optimisation of the cultivation of *N. oceanica* when aiming for EPA production.

Declaration of authors contributions

Statement of informed consent, human/animal rights

No conflicts, informed consent, or human or animal rights are applicable to this study.

References

[1]   M. Mitra, S.K. Patidar, S. Mishra, Integrated process of two stage cultivation of Nannochloropsis sp. for nutraceutically valuable eicosapentaenoic acid along with biodiesel, Bioresour. Technol. 193 (2015) 363–369. doi:10.1016/j.biortech.2015.06.033.

[2]   R.B. Draaisma, R.H. Wijffels, P. (Ellen) Slegers, L.B. Brentner, A. Roy, M.J. Barbosa, Food commodities from microalgae, Curr. Opin. Biotechnol. 24 (2013) 169–177. doi:10.1016/J.COPBIO.2012.09.012.

[3]   X.N. Ma, T.P. Chen, B. Yang, J. Liu, F. Chen, Lipid production from Nannochloropsis, Mar. Drugs. 14 (2016). doi:10.3390/md14040061.

[4]   Y. Li-Beisson, J.J. Thelen, E. Fedosejevs, J.L. Harwood, The lipid biochemistry of eukaryotic algae, Prog. Lipid Res. 74 (2019) 31–68. doi:10.1016/J.PLIPRES.2019.01.003.

[5]   E.T. Chua, P.M. Schenk, A biorefinery for Nannochloropsis: Induction, harvesting, and extraction of EPA-rich oil and high-value protein, Bioresour. Technol. 244 (2017) 1416–1424. doi:10.1016/J.BIORTECH.2017.05.124.

[6]   E.J. Baker, E.A. Miles, G.C. Burdge, P. Yaqoob, P.C. Calder, Metabolism and functional effects of plant-derived omega-3 fatty acids in humans, Prog. Lipid Res. 64 (2016) 30–56. doi:10.1016/j.plipres.2016.07.002.

[7]   World Health Organisation, Population nutrient intake goals for preventing diet-related chronic diseases, (n.d.).

[8]   A.R. Ribeiro, A. Gonçalves, M. Barbeiro, N. Bandarra, M.L. Nunes, M.L. Carvalho, J. Silva, J. Navalho, M.T. Dinis, T. Silva, J. Dias, Phaeodactylum tricornutum in finishing diets for gilthead seabream: effects on skin pigmentation, sensory properties and nutritional value, J. Appl. Phycol. 29 (2017) 1945–1956. doi:10.1007/s10811-017-1125-3.

[9]   J. Janssen, R. Wijffels, M. Barbosa, Lipid Production in Nannochloropsis gaditana during Nitrogen Starvation, Biology (Basel). 8 (2019) 5. doi:10.3390/biology8010005.

[10]   M.J. Caballero, M.S. Izquierdo, E. Kjørsvik, D. Montero, J. Socorro, A.J. Fernández, G. Rosenlund, Morphological aspects of intestinal cells from gilthead seabream (Sparus aurata) fed diets containing different lipid sources, Aquaculture. 225 (2003) 325–340. doi:10.1016/S0044-8486(03)00299-0.

[11]   I.A. Nascimento, S.S.I. Marques, I.T.D. Cabanelas, G.C. de Carvalho, M.A. Nascimento, C.O. de Souza, J.I. Druzian, J. Hussain, W. Liao, Microalgae Versus Land Crops as Feedstock for Biodiesel: Productivity, Quality, and Standard Compliance, Bioenergy Res. 7 (2014) 1002–1013. doi:10.1007/s12155-014-9440-x.

[12]   D. Simionato, M.A. Block, N. La Rocca, J. Jouhet, E. Maréchal, G. Finazzi, T. Morosinotto, The Response of Nannochloropsis gaditana to Nitrogen Starvation Includes De Novo Biosynthesis of Triacylglycerols, a Decrease of Chloroplast Galactolipids, and Reorganization of the Photosynthetic Apparatus, Eukaryot. Cell. 12 (2013) 665–676. doi:10.1128/ec.00363-12.

[13]   J.G. Henriques, S. Buziol, E. Stocker, A. Voogd, J.C. Menezes, Monitoring Mammalian Cell Cultivations for Monoclonal Antibody Production Using Near-Infrared Spectroscopy, in: T. Scheper, G. Rao (Eds.), Opt. Sens. Syst. Biotechnol., Springer, 2010: p. 262. doi:10.1007/BFb0103032.

[14]   S. Marose, C. Lindemann, R. Ulber, T. Scheper, Optical sensor systems for bioprocess monitoring, TIBTECH. 17 (1999) 30–34. doi:10.1002/elsc.201500014.

[15]   Joseph R. Lakowicz, Principles of fluorescence spectroscopy, 2006. doi:10.1007/978-0-387-46312-4.

[16]   I.T.D. Cabanelas, P.M. Slegers, H. Böpple, D.M.M. Kleinegris, R.H. Wijffels, M.J. Barbosa, Outdoor performance of Chlorococcum littorale at different locations, Algal Res. 27 (2017) 55–64. doi:10.1016/j.algal.2017.08.010.

[17]   A.M.J. Kliphuis, A.J. Klok, D.E. Martens, P.P. Lamers, M. Janssen, R.H. Wijffels, Metabolic modeling of Chlamydomonas reinhardtii: Energy requirements for photoautotrophic growth and maintenance, J. Appl. Phycol. 24 (2012) 253–266. doi:10.1007/s10811-011-9674-3.

[18]   G. Breuer, L. De Jaeger, V.P.G. Artus, D.E. Martens, J. Springer, R.B. Draaisma, G. Eggink, R.H. Wijffels, P.P. Lamers, Superior triacylglycerol (TAG) accumulation in starchless mutants of Scenedesmus obliquus: (II) evaluation of TAG yield and productivity in controlled photobioreactors, Biotechnol. Biofuels. 7 (2014) 1–11. doi:10.1186/1754-6834-7-70.

[19]   G.M. León-Saiki, I.M. Remmers, D.E. Martens, P.P. Lamers, R.H. Wijffels, D. van der Veen, The role of starch as transient energy buffer in synchronized microalgal growth in Acutodesmus obliquus, Algal Res. 25 (2017) 160–167. doi:10.1016/j.algal.2017.05.018.

[20]   M. Bahram, R. Bro, C. Stedmon, A. Afkhami, Handling of Rayleigh and Raman scatter for PARAFAC modeling of fluorescence data using interpolation, J. Chemom. (2006) 99–105. doi:10.1002/cem.978.

[21]   R. Boggia, M. Forina, P. Fossa, L. Mosti, Chemometric study and validation strategies in the structure-activity relationships of new cardiotonic agents, Quant. Struct. Relationships. 16 (1997) 201–213. doi:10.1002/qsar.19970160303.

[22]   C. a Andersson, R. Bro, The N -way Toolbox for MATLAB, Chemom. Intell. Labor. 52 (2000)

1–4.

[23] J.M. Sandnes, T. Källqvist, D. Wenner, H.R. Gislerød, Combined influence of light and temperature on growth rates of Nannochloropsis oceanica: Linking cellular responses to large-scale biomass production, J. Appl. Phycol. 17 (2005) 515–525. doi:10.1007/s10811-005-9002-x.

[24] V. Ördög, W.A. Stirk, P. Bálint, A.O. Aremu, A. Okem, C. Lovász, Z. Molnár, J. van Staden, Effect of temperature and nitrogen concentration on lipid productivity and fatty acid composition in three Chlorella strains, Algal Res. 16 (2016) 141–149. doi:10.1016/j.algal.2016.03.001.

[25] M.H.A. Michels, J. Camacho-Rodríguez, M.H. Vermuë, R.H. Wijffels, Effect of cooling in the night on the productivity and biochemical composition of Tetraselmis suecica, Algal Res. 6 (2014) 145–151. doi:10.1016/j.algal.2014.11.002.

[26] J. Fabregas, A. Maseda, A. Dominguez, M. Ferreira, A. Otero, Changes in the cell composition of the marine microalga, Nannochloropsis gbitana, during a light:dark cycle, Biotechnol. Lett. 24 (2002) 1699–1703.

[27] S.S. Nikaido, C.H. Johnson, Daily and Circadian Variation in Survival From Ultraviolet Radiation in Chlamydomonas reinhardtii, Photochem. Photobiol. 71 (2004) 758. doi:10.1562/0031-8655(2000)071<0758:dacvis>2.0.co;2.

[28] G. Benvenuti, P.P. Lamers, G. Breuer, R. Bosma, A. Cerar, R.H. Wijffels, M.J. Barbosa, Microalgal TAG production strategies: Why batch beats repeated-batch, Biotechnol. Biofuels. 9 (2016) 1–17. doi:10.1186/s13068-016-0475-4.

[29] H. Jiang, K. Gao, Effects of lowering temperature during culture on the production of polyunsaturated fatty acids in the marine diatom Phaeodactylum tricornutum (Bacillariophyceae), J. Phycol. 40 (2004) 651–654. doi:10.1111/j.1529-8817.2004.03112.x.

[30] Y. Wang, B. He, Z. Sun, Y.F. Chen, Chemically enhanced lipid production from microalgae under low sub-optimal temperature, Algal Res. 16 (2016) 20–27. doi:10.1016/j.algal.2016.02.022.

[31] T. Sakamoto, S. Higashi, H. Wada, N. Murata, D.A. Bryant, Low-temperature-induced desaturation of fatty acids and expression of desaturase genes in the cyanobacterium

Synechococcus sp. PCC 7002, FEMS Microbiol. Lett. 152 (1997) 313–320.

[32] M. Sá, J. Monte, C. Brazinha, C.F. Galinha, J.G. Crespo, Fluorescence coupled with chemometrics for simultaneous monitoring of cell concentration , cell viability and medium nitrate during production of carotenoid-rich Dunaliella salina, Algal Res. 44 (2019) 101720. doi:10.1016/j.algal.2019.101720.

Tables

Table 1: Cell size (µm) of *N. oceanica* at the first day of the "stress phase", and after 4 and 10 days. Experimental error inferior to 2.3% (calculated for duplicates of 23 samples)

Batch	Day 0	Day 4	Day 10
15	3.08	2.98	-
20	2.92	3.14	-
25	2.88	3.08	-
30	2.98	3.28	-
N-starv	2.85	2.85	3.04
25-15	3.10	3.30	3.00
Control	2.79	3.05	3.34

Table 2: Statistical parameters of the prediction models of eicosapentaenoic acid (EPA) content (% g/g$_{DW}$) in triacylglycerols (TAG) fraction of N. oceanica. Total number of observations used: 73

	Var	RMSECV	RMSEP	Validation		Training		Nr of
	(%)	*%g/g $_{DW}$*		R^2	Slope	R^2	Slope	**inputs**
N-starv (n_{val}=10)	93.1	0.29	1.55	0.01	0.21	0.93	1.00	16
25-15 (n_{val}=11)	89.5	0.42	0.21	0.78	0.53	0.89	1.00	14
Control (n_{val}=9)	89.5	0.46	0.21	0.49	0.69	0.89	1.00	17
15 (n_{val}=10)	85.9	0.46	0.31	0.68	1.23	0.86	1.00	11
20 (n_{val}=9)	67.4	0.62	0.29	0.79	0.92	0.67	1.00	5
25 (n_{val}=7)	89.9	0.42	0.47	0.74	0.99	0.90	1.00	14
30 (n_{val}=7)	89.6	0.41	0.31	0.93	0.77	0.90	1.00	15

Var – variance captured by the model;
RMSECV – root mean square error of cross-validation;
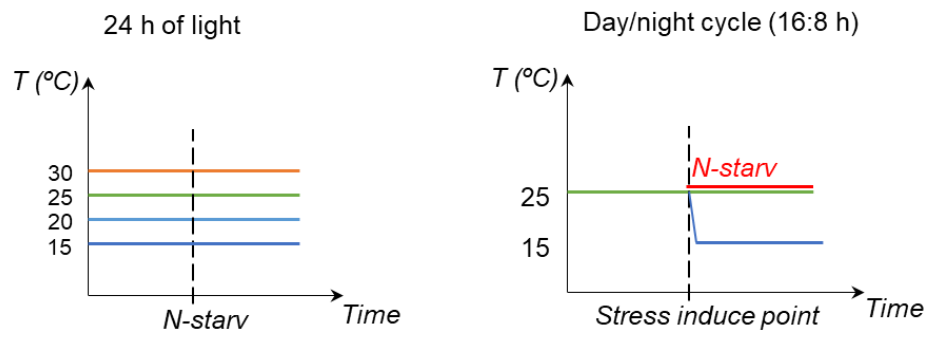RMSEP – root mean square error of prediction;
n_{val} – number of observations used for validation.

Table 3: Statistical parameters of the prediction models of eicosapentaenoic acid (EPA) content (% g/g$_{DW}$) in the polar (PL) fraction of N. oceanica. Total number of observations used: 73

|	Var	RMSECV	RMSEP	Validation		Training		# Inputs
	(%)	*% g/g$_{DW}$*		R^2	Slope	R^2	Slope	
N-starv (n$_{val}$=10)	93.1	0.29	1.55	0.01	0.21	0.93	1.00	16
25-15 (n$_{val}$=11)	89.5	0.42	0.21	0.78	0.53	0.89	1.00	14
Control (n$_{val}$=9)	89.5	0.46	0.21	0.49	0.69	0.89	1.00	17
15 (n$_{val}$=10)	85.9	0.46	0.31	0.68	1.23	0.86	1.00	11
20 (n$_{val}$=9)	67.4	0.62	0.29	0.79	0.92	0.67	1.00	5
25 (n$_{val}$=7)	89.9	0.42	0.47	0.74	0.99	0.90	1.00	14
30 (n$_{val}$=7)	89.6	0.41	0.31	0.93	0.77	0.90	1.00	15

Var – variance captured by the model;
RMSECV – root mean square error of cross-validation;
RMSEP – root mean square error of prediction;
n$_{val}$ – number of observations used for validation.

Figure 1: Photobioreactor operating conditions; environmental conditions tested. N-starv: nitrogen depletion.
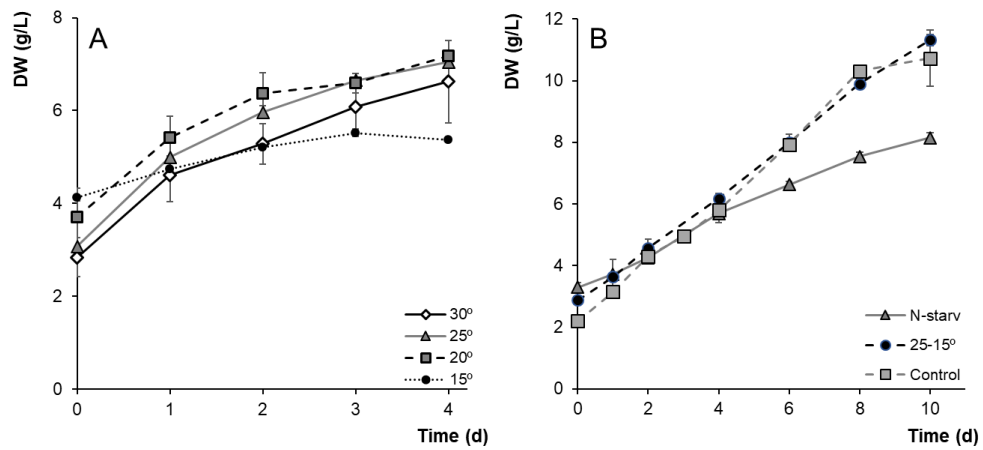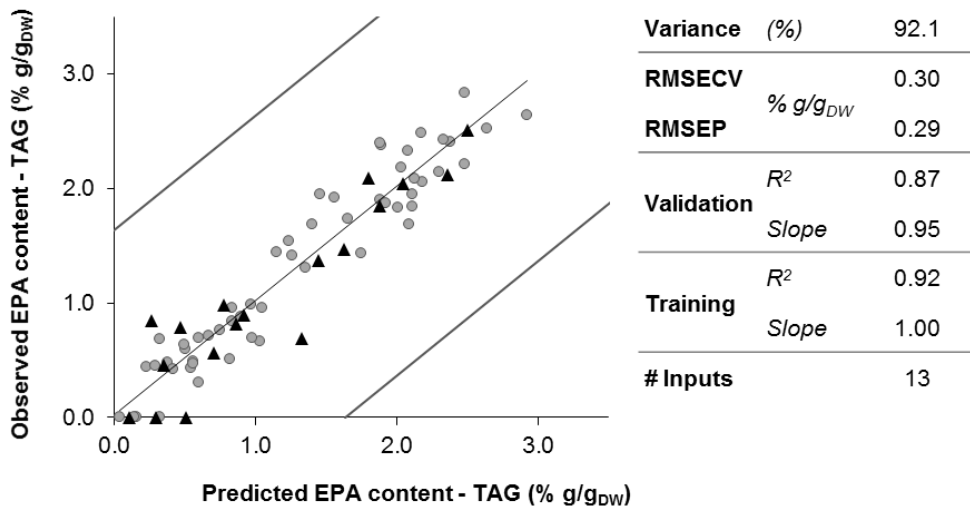
Figure 2: Average biomass concentration and respective standard deviation bars (n=3), expressed as dry weight, (DW in g/L), measured over time, from the moment of stress induction. A) For 24h of light and nitrogen starvation, comparison between different temperatures: 15 (circles), 20 (squares), 25 (triangles) and 30 ºC (diamonds); B) For day/night cycle (16/8), comparison between nitrogen starvation (N-starv, triangles), decrease of temperature from 25 to 15 ºC (25-15º, circles) and a control experiment (no starvation and no decrease of temperature, squares)
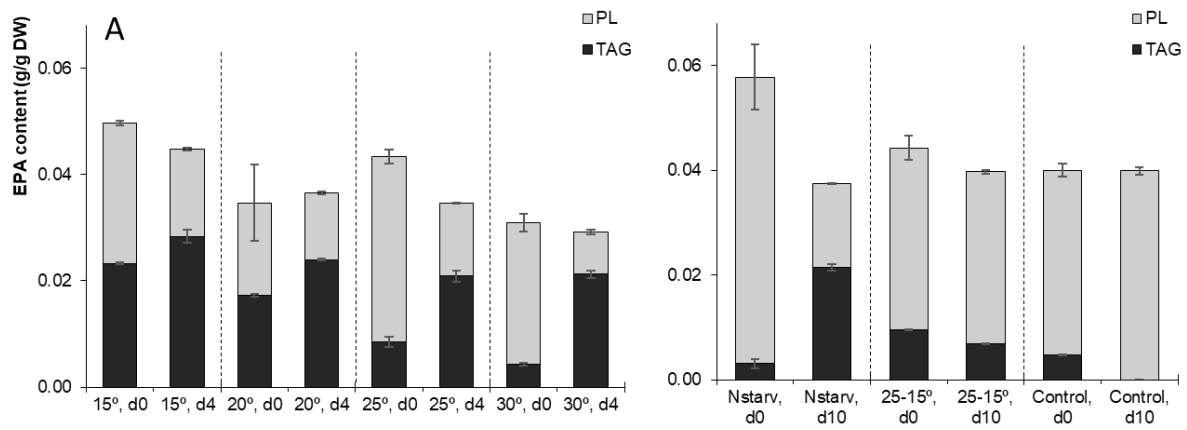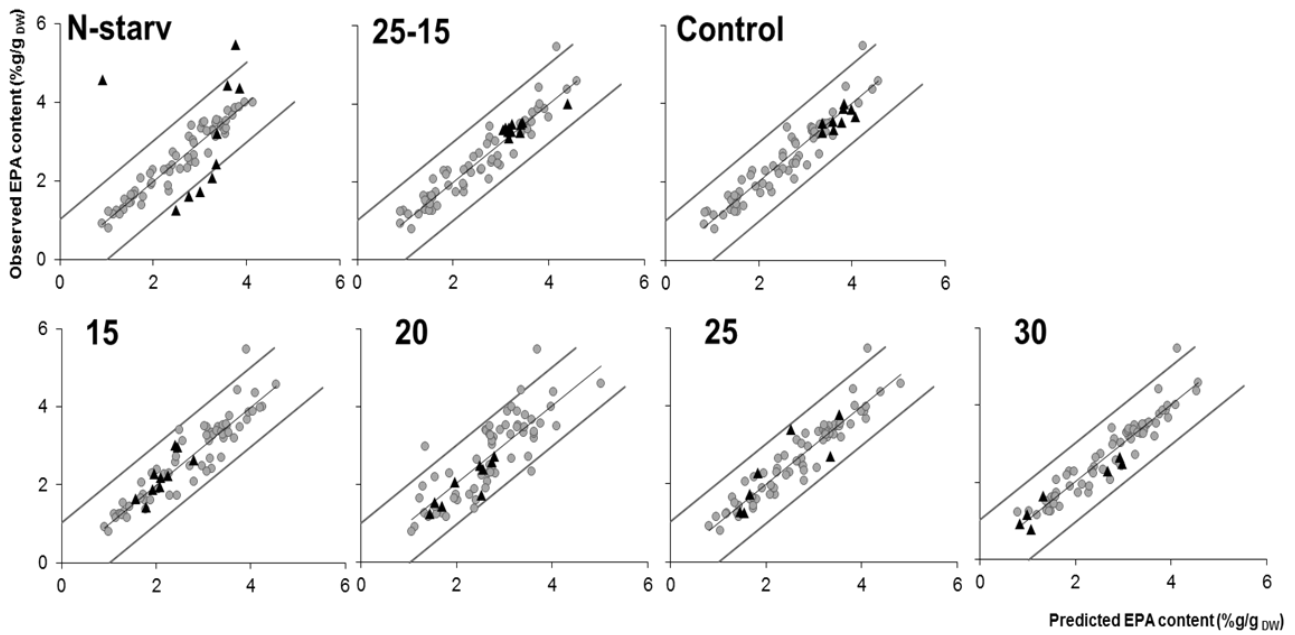
Figure 3: EPA (eicosapentaenoic acid) content expressed per dry weight of biomass (g EPA/g DW) in the triacylglycerols (TAG, dark grey) and polar lipids (PL, light grey) fractions. Error bars calculated as standard deviations (n=2). A) For 24h of light and nitrogen starvation, comparison between different temperatures: 15, 20, 25 and 30 °C. B) For day/night cycle (16/8), comparison between nitrogen starvation (N-starv), decrease of temperature from 25 to 15 °C (25-15) and a control experiment (no starvation and no decrease of temperature)
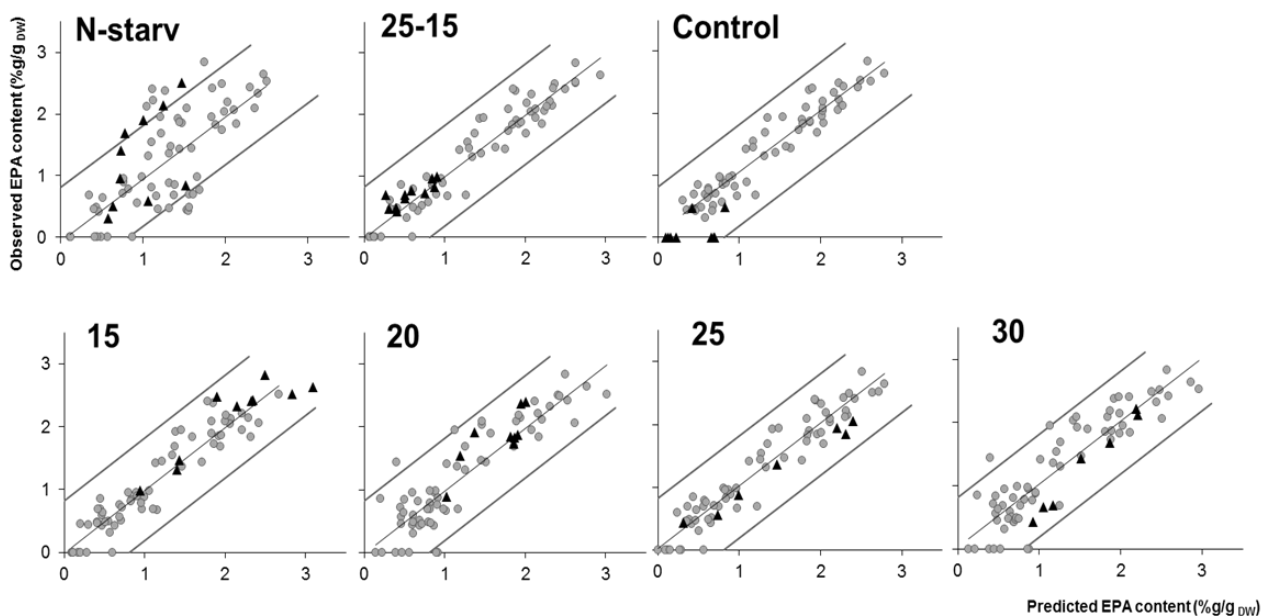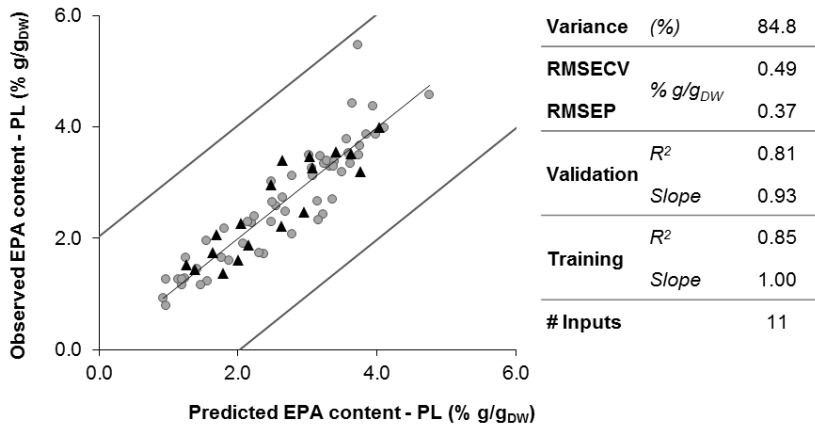
Figure 4: Prediction models of eicosapentaenoic acid (EPA) content in triacylglycerols (TAG) fraction of *N. oceanica*. Predicted values (x-axis) are plotted against observed values (y-axis). Training (●) and validation (▲) data are represented in percentage of grams of EPA per grams of biomass dry weight (% g/g$_{DW}$). Total number of observations used for each model 73

Variance	(%)	92.1
RMSECV	% g/g$_{DW}$	0.30
RMSEP		0.29
Validation	R^2	0.87
	Slope	0.95
Training	R^2	0.92
	Slope	1.00
# Inputs		13

Figure 5: Prediction model of eicosapentaenoic acid (EPA) content in triacylglycerols (TAG) fraction (total number of observations 73) using 75 % of the data for training (●) and 25 % for validation (▲), represented in percentage of grams of EPA per grams of dry weight (% g/g$_{DW}$). Model performance parameters: variance captured (Variance); root mean square error of cross-validation (RMSECV); root mean square error of prediction (RMSEP); coefficients of determination (R2) and slopes of linear regression between observed and predicted data obtained respectively for the training and validation data sets; total number of inputs used by the model (# inputs)

Figure 6: Prediction models of eicosapentaenoic acid (EPA) content in polar lipids (PL) fraction of *N. oceanica*. Predicted values (x-axis) are plotted against observed values (y-axis). Training (●) and validation (▲) data are represented in percentage of grams of EPA per grams of biomass dry weight (% g/g$_{DW}$). Total number of observations used for each model 73

Figure 7: Prediction model of eicosapentaenoic acid (EPA) content in polar lipids (PL) fraction (total number of observations 73) using 75 % of the data for training (●) and 25 % for validation (▲), represented in percentage of grams of EPA per grams of dry weight (% g/g_DW). Model performance parameters: variance captured (Variance); root mean square error of cross-validation (RMSECV); root mean square error of prediction (RMSEP); coefficients of determination (R2) and slopes of linear regression between observed and predicted data obtained respectively for the training and validation data sets; total number of inputs used by the model (# inputs)