**Characterization and Manipulation of a DGAT2 from the Diatom *Thalassiosira*
pseudonana: Improved TAG Accumulation without Detriment to Growth and Implications
for Chloroplast TAG Accumulation**

Kalpana Manandhar-Shrestha and Mark Hildebrand*

Marine Biology Research Division, Scripps Institution of Oceanography, University of
California, San Diego

* Corresponding author

Mail code 0202

9500 Gilman Dr.

La Jolla CA 92093-0202

858-822-0167

mhildebrand@ucsd.edu

Abstract

Improving the ability of microalgae cells to accumulate triacylglycerol (TAG) without a negative effect on growth or biomass accumulation is desirable for economically-viable biofuels production. Metabolic engineering is one method to increase TAG yields, provided that appropriate gene target(s) are identified. Microalgal lipid and TAG synthesis is incompletely characterized, but it is clear that it differs in fundamental ways from plants and other eukaryotes. We characterize a DGAT2 homolog from the diatom *Thalassiosira pseudonana*, and show that it shifts intracellular location from the chloroplast during exponential growth, to the endoplasmic reticulum in stationary phase when large cytoplasmic lipid droplets accumulate. This is the first direct localization of a chloroplast DGAT2, which confirms the ability of these organelles to accumulate TAG in evolutionarily-diverse classes of microalgae. The enzyme is also not directly localized to cytoplasmic lipid droplets, suggesting that TAG synthesis and lipid droplet formation are separate processes. Overexpression of the enzyme resulted in significantly increased TAG accumulation without a negative effect on growth or biomass accumulation. Overexpression lines had a distinct shift in fatty acid composition relative to wild-type. The results shed light on fundamental processes of algal lipid metabolism and a practical benefit of improved productivity in transgenic lines.

Keywords: diatom, *Thalassiosira pseudonana*, DGAT2, chloroplast TAG synthesis, metabolic engineering

1. Introduction

Many species of microalgae are excellent accumulators of triacylglycerol (TAG), which serves as a lipid and energy storage form for the organism. For human purposes, TAG is an excellent precursor to generate bio-based fuels. Even considering whole cell conversion processes such as hydrothermal liquefaction (HTL), the quality of the precursor material will improve the quality and value of the final product, therefore HTL will benefit from algae that accumulate higher lipid and TAG levels [1,2]. Improving lipid yields while maintaining high growth rate is predicted to provide the greatest benefit to reducing production costs [3]. In most types of algae, TAG does not accumulate appreciably unless cell growth is ceased by nutrient starvation or cell cycle arrest [4,5]. This is less than optimal in a production system, where overall productivity depends not only on the level of TAG in individual cells, but on the overall culture density and throughput, which can be limited by nutrient-based division arrest.

Efforts have been made to increase TAG levels or reduce TAG accumulation time in microalgae, either using mutagenesis/selection approaches[6-12], or through metabolic engineering of lipid biosynthesis or breakdown pathways[13,14]. Engineering efforts would be augmented by a better understanding of individual steps in algal lipid biosynthetic pathways, which are different than those in plants and mammals [15]. For example, TAG is usually associated with lipid droplet formation in the cytoplasm, but recent work indicates that *C. reinhardtii* can accumulate TAG in the chloroplast [16,17]. Intracellular compartmentation is likely to play an important role in microalgal TAG accumulation, particularly considering that complex compartmentation occurs in species derived from double endosymbiotic events, such as diatoms[18]. A better understanding of microalgal TAG synthesis processes could lead to informed manipulation approaches, hopefully without detriment to growth.

The enzymology of TAG synthesis is generally well understood. TAG synthesis in photosynthetic eukaryotic cells is carried out by multiple enzymatic reactions in different subcellular compartments. Fatty acids are synthesized in the chloroplast. For non-chloroplast TAG synthesis, fatty acids are exported from the chloroplast and acylated at different steps to produce TAG. Acetylation of glycerol-3-phosphate, catalyzed by the enzyme glycerol-3-phosphate-acyltransferase (GPAT) leads to the formation of lysophosphatidic acid (LPA). LPA can be then acetylated by lysophosphatidic acid acyltransferase (LPAAT) to form phosphatidic acid (PA). PA is the precursor for diacylglycerol (DAG) and anionic phospholipids. PA is dephosphorylated by PA phosphatase (PAP) to form diacylglycerol (DAG). DAG is an immediate precursor for TAG, which can utilize acyl-CoA-dependent and independent acylation metabolic pathways. Acyl-CoA-dependent acylation of DAG is catalyzed by DAG acyltransferases (DGATs), and acyl-CoA-independent acylation use phospholipids as acyl donors and DAG as an acyl acceptor catalyzed by phospholipid:DGAT. Three types of DGATs have been identified in eukaryotic cells. DGAT1 and DGAT2 are membrane proteins and play a direct role in synthesis of TAG in many eukaryotes, and DGAT3 is localized in cytosol and is not involved in oil production [19,20]. Although they catalyze the same reaction, the three types of DGAT belong to different gene families and share little sequence similarity [21-23].

There has been substantial interest recently in characterization of DGAT2s in diatoms. Xu et al. [24] functionally characterized a *T. pseudonana* DGAT2 (Thaps3_37867) in yeast, showing that it had DGAT2 activity, and altered the fatty acid profile resulting in increased C18:1, and incorporation of EPA and DHA into TAG. Gong et al. [25] characterized a *P. tricornutum* DGAT2 (Phatr2_49544) also in yeast, and showed that the transcript for this protein was induced prior to TAG accumulation in *P. tricornutum*. Niu et al. [13] characterized a distinct DGAT2

from *P. tricornutum* (Phatr2_49462), generating a transgenic line of *P. tricornutum* overexpressing the enzyme. They showed a slight decrease in biomass accumulation in the overexpression line relative to wild type, but because only one line was tested, it cannot be concluded whether that was due to overexpression or the genomic DNA insertion site. Nile red staining indicated 35% more TAG accumulation than wild type, but growth conditions were not reported. A substantial increase in long chain polyunsaturated fatty acids (LCPUFAs) was also reported in the transgenic line. Although this report suggests improvement in the transgenic line for TAG and LCPUFA levels with a detriment to maximum biomass, characterization of only one transgenic clone does not allow assessment of the reproducibility of the phenomenon.

Herein, we report the characterization of a distinct DGAT2 from *T. pseudonana*, Thaps3_25595. We include a detailed comparative phylogenetic analysis of the relatedness of DGAT2s from this species and other diatoms, determine the protein's intracellular location in relation to lipid droplet formation, and generate and characterize multiple overexpression transgenic lines. The localization data indicates a transition between intracellular compartments under different conditions, and distinguishes the location of the enzyme and location of lipid droplets. The localization results have implications related to chloroplast TAG synthesis in evolutionarily-diverse microalgae. We document improved TAG accumulation in all overexpression transgenic lines relative to wild type, with no detrimental effect on growth rate or biomass accumulation.

2. Materials and Methods

2.1. Culture conditions

Axenic cultures of the marine diatom *Thalassiosira pseudonana* were grown in artificial seawater (ASW) [26] at 18-20°C under continuous illumination at 150 μ einsteins. For silicon starvation, exponentially grown cultures were harvested and washed twice with silicon-free ASW and then incubated at 1.0×10^6 cells/ml in silicon-free ASW at 18-20°C in polycarbonate flasks. Cells were counted either using a hemocytometer or a Muse® Cell Analyzer (Millipore).

2.2. Construction of expression vectors

The intron-less coding region of a DGAT2 (Thaps3_25595) was PCR-amplified from genomic DNA of wild type *T. pseudonana* using Forward-5'-ATGGAAACAAAGCAAAAGAAGTCC-3' and Reverse-5'-CCAAACCTCAAGCTTTGCAACC-3' primers. The PCR product was cloned into a Gateway donor vector (pDONR 221 attB1-attB5) using BP clonase (Life Technologies) to create an entry clone. Similarly, TagRFP or eGFP entry clones were created using BP cloning into a Gateway donor vector (pDONR 221 attB5-attB3). To express the C-terminal tagged protein, DGAT2 and TagRFP or eGFP were cloned between the fucoxanthin chlorophyll a/c-binding gene (fcp) promoter and terminator in the destination vector pMHL_78 using LR clonase.

2.3. Transformation procedure and characterization of transgenic lines

Expression vectors encoding Thaps3_25595 TagRFP or eGFP were cotransformed with pMHL_9 encoding the selective marker nourseothricin acetyltransferase, NAT [27] into *T. pseudonana* cells using the Bio-Rad biolistic particle delivery System. NAT-resistant transgenic cells were selected either in liquid culture or on agar plates containing 100 μ g/ml

nourseothricin (clonNAT, Werner BioAgents, Germany). Cells expressing TagRFP or eGFP were then sorted from liquid culture by flow cytometry and subsequently plated, and colonies picked from agar plates were grown in liquid in a 24-well plate and screened by fluorescence microscopy. Insertion of DNA into transformed algal cells was further confirmed by PCR analysis and DNA sequencing.

2.4. DNA extraction

4-5 x 10⁶ cells were harvested and washed once with sterilized water. The supernatant was discarded and 25 µL TE was added to the pellet. Cells were then ruptured by 3 cycles of freezing on dry ice and thawing at 65° C followed by heating at 95° C for 5 min. Supernatants were collected after centrifugation at 12000 rpm and used for PCR.

2.5. Imaging flow cytometry

Cells were harvested at different time points, pelleted by centrifugation and the pellet was stored at -20° C until use. For analysis, pellets were thawed, resuspended in 0.45 M NaCl and stained with Nile red (1 µg ml⁻¹) for 10 min at room temperature in darkness. Nile red fluorescence as a proxy of TAG level was quantified by using an ImageStream X (Amnis Corp.) imaging flow cytometer. Cells were excited with a 405-nm laser at 10 mW, and brightfield and fluorescent images were collected for 20,000 events per sample. Data were collected at 40x magnification. Data was analyzed using Amnis IDEAS™ software for the ImageStream.

2.6. Staining of algal cells using BODIPY

Lipid droplets were visualized in BODIPY-stained cells using fluorescence microscopy. $1.0 - 2.0 \times 10^6$ cells were harvested, pelleted, and re-suspended in 0.45 M NaCl. To visualize neutral lipid droplets, cells were stained with $2.6 \mu\text{g ml}^{-1}$ BODIPY (4, 4-difluoro-3a, 4a diaza-s-indacene) (Invitrogen, Carlsbad CA, USA) incubating for 15-20 min at room temperature in the dark.

2.7. Fluorescence microscopy

Localization of Thaps3_25595:eGFP or TagRFP fusions in transgenic cells was carried out using the Zeiss Axio Observer.Z1 microscope with Plan-Apochromat 63x/1.40 Oil DIC M27 objective and filter set 38HE (Ex 470/40 nm, FT 495, Em 525/50 nm) for eGFP, filter set 46 (Ex 500/25 nm, FT 515 nm, Em 535/30 nm) for TagRFP and 05 filter set (Ex 395-440 nm, FT 460 nm, Em 470 nm LP) for chlorophyll. BODIPY fluorescence was observed using a Zeiss Axio Observer.Z1 microscope with filter set 46 (Ex 500/25 nm, FT 515 nm, Em 535/30 nm).

2.8. GC/MS analysis of fatty acid content

Exponentially grown cells were incubated in silicon-free medium for 24h. Cells were harvested at 0 and 24h washed with 0.4 M ammonium-formate and flushed with nitrogen gas to replace air in the tubes and stored at -80°C until used. Known amounts of lyophilized samples were placed in pre-combusted glass tubes, resuspended in 1 ml methanolic acid and 50 μl nonadecanoic acid methyl ester standard (C19:0, 5mg/ml in hexanes) and incubated at 65°C for 30 min. After the reaction, the tubes were allowed to cool to room temperature, and 1 ml of hexane was then added and vortexed. The upper hexane layer was collected into a GC-MS vial. FAME profiles were determined using Agilent 7890A [Agilent, US] gas chromatography system connected to a 5975C VL MSD quadrupole MS (EI). Samples were separated on a 60m DB23 Agilent GC-MS

column using helium as carrier gas and a gradient of 110 °C to 200 °C at 15 °C/min, followed by 20 minutes at 200 °C. The amount of each lipid class was calculated using retention times of known internal standards(C:19).

2.9. Bioinformatic analyses

The evolutionary relationships of members of the DGAT2 super family from different organisms were analyzed using the CLC main workbench 6.0 program[28]. BLASTp [29] was used to compare theThaps3_25595 sequence with DGAT2's from other organisms. SMART analysis [30]. TMHMM 2.0 [31] was used to map potential membrane-spanning segments. SignalP 3.0 or 4.1[32,33] were used to predict endoplasmic reticulum cleavage signals, and ChloroP 1.1 [34] was used to predict chloroplast targeting.

3. Results

3.1. Identification and characterization of a DGAT2 in *T. pseudonana*

By sequence analysis, we identified a possible DGAT2 (Thaps3_25595, Fig. S1) in the *Thalassiosira pseudonana* genome (<http://genome.jgi.doe.gov/Thaps3>). Multiple alignments with diverse eukaryotic phyla showed a conservation of important functional domains, indicating that the gene is indeed a DGAT2 enzyme (Fig. 1). BLASTp[29] searches showed that the amino acid sequence from *T. pseudonana* and from diverse organisms, including pathogens, mammals, microalgae and plants, had an LPLAT_MGAT-like domain at its C-terminus. The N terminal portion of Thaps3_25595 possessed an acyl-acceptor binding pocket which suggests that Thaps3_25595 has acyltransferase activity. Moreover, SMART analysis [30] showed a

204 PlsC(Phosphate acyltransferase involved in phospholipid biosynthesis) domain at residues 171-
205 277 (Fig. 1). Furthermore, a diacylglycerol O-acyltransferase (PLN02783) domain was identified
206 at residues 111-366. The sequence motif “YFP” plays a significant role in enzyme catalysis and is
207 highly conserved in plants, mammals and pathogens (Fig. 1), but is more variable in algal
208 DGAT2s [29]. A second highly conserved amino acid motif HPHG plays a significant role in
209 DGAT2 activity, and may be part of the active site [35]. The sequence “PH” is essential for high
210 levels of enzymatic DGAT2 activity [36,37]. The DGAT2 HPHG-motif is conserved in yeast,
211 fungi and algae, but it is equivalent to EPHS in plants. In addition, several conserved
212 hydrophobic amino acids which are important for DGAT2 activity were observed among all
213 DGAT2 enzymes.

214 A phylogenetic analysis of the DGAT2 amino acid sequences resulted in a generally well-
215 resolved tree and showed two main clades: the first clade comprised non-photosynthetic
216 pathogens and mammals, and the second clade comprised photosynthetic organisms, including
217 plants and algae (Fig. 2). Thaps3_25595 was most closely related to the centric diatom. *C.*
218 *cryptica* and followed by the pennate diatom *F. cylindrus*. Interestingly Thaps3_25595 was
219 found to be more closely related to the brown alga *E. siliculosus* than the diatom *P. tricornutum*.

220 Alignment of all DGAT2 sequences from three diatom genomes (Fig. S2) revealed
221 interesting relationships. Thaps3_25595 clades more closely with a gene from
222 *Ectocarpus siliculosus* than to any of the DGATs in *P. tricornutum*, suggestive of unique
223 characteristics for this enzyme.

224 Membrane protein topology analysis with TMHMM 2.0 [31] predicted three transmembrane
225 domains at amino acid residues 72-94, 99-121 and 163-185 (Fig. 1). Bioinformatic targeting
226 predictions were uninformative, SignalP 3.0 or 4.1 [32,33] did not predict an ER signal peptide,

ChloroP 1.1 [34] predicted chloroplast targeting, but since diatoms require a bipartite targeting sequence consisting of ER and chloroplast targeting sequences to enter the chloroplast [38], the prediction does not definitively indicate whether the protein might actually associate with the chloroplast.

3.2. Localization of Thaps3_25595 -eGFP or -TagRFP fusion proteins by fluorescence microscopy

To directly determine intracellular location, the full length of the Thaps3_25595 gene was PCR-amplified from genomic DNA, and green fluorescent protein (eGFP) or monomeric red fluorescent protein (TagRFP) was fused at the C-terminus. The construct was placed under the control of the fucoxanthin chlorophyll a/c binding protein (fcp) promoter and introduced into *T. pseudonana*. Transgenic clones were screened for antibiotic nourseothricin (nat1)-resistance conferred by a cotransformed vector expressing the nat1 gene under control of the ACCase promoter, and cells expressing eGFP or TagRFP were further enriched using a sorting flowcytometer. The sorted cells were plated on agar plates or grown in liquid ASW medium containing nourseothricin for a second round of sorting. Four independent transgenic clones were selected for further study.

Transgenic clones expressing the Thaps3_25595 -eGFP or -TagRFP were visualized by fluorescence microscopy. Thaps3_25595 protein was localized in the chloroplast during the exponential and early stationary phases of growth (Fig. 3). Interestingly, the localization was shifted to the endoplasmic reticulum in stationary phase (Fig. 3). It is possible that the subcellular localization of the protein could vary depending on its expression level, and that use of the constitutive fcp promoter could lead to mistargeting due to protein over-accumulation. To

address this, we constructed an expression vector using Thaps3_25595 fused with TagRFP under control of the native promoter. During the process, we obtained RNAseq data (Abbriano Burke, in preparation) showing that the Thaps3_25595 gene model was incorrect. Even though a contiguous open reading frame ran upstream to the sequence MQQF, the RNAseq data indicated that this methionine was not transcribed, and hence was not the authentic start codon. The authentic start codon (gene model g12083.t1) was the next downstream, METK (Fig. S1). Targeting predictions using this start codon were identical to Thaps3_25595. Use of the TagRFP-fused native construct revealed the same localization as Thaps3_25595, with chloroplast localization during exponential and early stationary phase growth, and ER localization during late stationary phase (Fig. 3). Based on this data, we assume the localization of Thaps3_25595 was either not affected by the additional amino acids, or the additional amino acids were not translated, and that the retargeting between the chloroplast and ER was authentic.

3.3. Comparison of growth and biomass accumulation in wild-type and transgenic lines

We monitored the growth and biomass accumulation of the wild type and Thaps3_25595 transgenic lines to evaluate the effect of recombinant protein expression. Cell numbers were counted using a MUSE cell counter for seven days. There was no difference in growth and biomass accumulation of wild type and transgenic clones in both exponential and stationary phases (Fig.4), indicating that expression of Thaps3_25595-eGFP had no detrimental effect.

3.4. Thaps3_25595 -TagRFP fusion protein expression

Using whole Affymetrix genome tiling microarrays, (Smith et al. submitted) we determined that mRNA levels for Thaps3_25595 become upregulated 8 hours after placing *T.*

pseudonana into silicon free medium, and remained upregulated to 24 hr (Fig. 5A). Triacylglycerol (TAG) levels noticeably increased between 8-12 hours (Fig. 5A). Thaps3_25595-TagRFP was expressed under an fcp promoter, which is generally constitutive during growth, but we show that it is also induced under Si limitation (Fig. 5B). We examined the level of expressed Thaps3_25595-TagRFP fusion protein, in cells grown under continuous light and silicon starvation, by monitoring TagRFP fluorescence intensities by imaging flow cytometry. The extent and timing of induction varied in the different transgenic lines, but Thaps3_25595 protein level in all the transgenic lines increased by 4 hours of silicon limitation, and remained consistent after 8h (Fig. 5C). The increase in expression level after transfer of culture to silicon starvation conditions could be attributed to the fcp promoter. Considering that the native protein should also be induced by 8 hour after Si limitation, the combination of native and fcp-driven expression should lead to overexpression of the total amount of Thaps3_25595 protein. Interestingly, the transgenic line Tg-1 expressed 1.5 fold higher than other three transgenic lines.

3.5. Expression of Thaps3_25595 increased TAG levels

Since DGAT2 is the final step of the TAG biosynthetic pathway, this enzyme may play an important role in the high accumulation of TAG under silicon starvation conditions [4,39]. We thus examined the ability of overexpression transgenic lines to accumulate TAG during silicon starvation. Cells were harvested at different time points (0, 4, 8, 12 and 24h), stained with the neutral lipid dye Nile-red, and relative TAG level was monitored by measuring Nile-red fluorescence by imaging flow cytometry. At 0h (unstressed condition), the level of TAG was similar in both the wild type and transgenic lines, with the exception of Tg-1 (Fig. 6). After 8h of silicon starvation, all of the transgenic lines accumulated a higher level of TAG than wild type.

At 24h of starvation, the transgenic clones accumulated a 1.52 to 1.95 fold higher level of TAG than the wild-type (Fig.6). Interestingly, Tg-1 accumulated a higher level of TAG both underunstarved and starved conditions, and also expressed a higher level of Thaps3_25595 protein than other transgenic clones (Fig. 5). The data are consistent with a higher level of TAG in transgenic clones resulting from higher levels of expression of Thaps3_25595.

3.6. Relative location of Thaps3_25595 protein and lipid droplets

We investigated the relative location of Thaps3_25595 and lipid droplet accumulation in *T. pseudonana*. Lipid droplets were visualized by fluorescence microscopy of transgenic clones after staining with the lipophilic-stain BODIPY. We observed lipid droplets in the cytoplasm in transgenic cells expressing Thaps3_25595 either in the chloroplast (Fig. 7A and B) or the ER (Fig. 7C). Therefore, in cells showing chloroplast localization of Thaps3_25595, there was no correlation with the location of lipid droplets (Fig. 7). In cells showing ER localization for Thaps3_25595, there was still not a strict correlation, but some protein appeared in the proximity of TAG (Fig. 7C). These results indicate that the location of DGAT2 does not correlate with the location of lipid droplet accumulation.

3.7. Fatty acid profile: increased PUFA content in TAG

Since the transgenic Thaps3_25595 expressing DGAT2 accumulated a higher level of TAG in silicon depleted medium (Fig. 6), we further investigated whether the fatty acid composition of total lipids was altered. Exponentially grown cells of two transgenic lines of Thaps3_25595 and wild type were harvested and washed twice with silicon-free ASW. Cells were then incubated in silicon-free ASW for 24h at 18-20°C under continuous illumination. Both wild type

and transgenic DGAT2-expressing cells were harvested at 0 and 24h and analyzed for fatty acid composition. As shown in Fig. 8A, saturated fatty acid C16:0 and C18:0 levels were lower than wild type after 24h of silicon starvation in transgenic Thaps3_25595 overexpressed cells. Unsaturated fatty acids 16:1 and 16:2 were not significantly different from wild type at 0 hours, but increased substantially compared with wild type at 24 hours (Fig. 8C). Substantial increases in C20:5(n3) (EPA) and C22:6 (n-3) (DHA) relative to wild type occurred in both exponential growth and silicon starvation.

4. Discussion

Sequence analysis indicated that Thaps3_25595 was a DGAT2, containing conserved amino acids characteristic of that enzyme (Figs. 1 and 2). Although DGAT2's have been characterized in this and other diatom species [13,24,25], Thaps3_25595 appears distinct, particularly because it clades preferentially with DGAT2 from *Ectocarpus siliculosus*, which is a filamentous brown alga, rather than the pennate diatom *Phaeodactylum tricornutum* DGAT2 (Fig. S2).

DGAT2's are typically localized in the ER membrane in other organisms, but this was not supported by targeting predictions on Thaps3_25595. Although chloroplast targeting was predicted, diatom chloroplasts are surrounded by the ER, and chloroplast localization also requires an ER signal [38]. Direct localization using eGFP or TagRFP fluorescent markers, under overexpression or native promoter control, showed targeting to either the ER or chloroplast, depending on culture stage. Chloroplast localization was prevalent during exponential and early stationary phase growth, and ER localization was prevalent during late stationary phase (Fig. 3). It is unclear how such dual targeting could occur. There is no evidence of alternative splicing or transcription start sites in our RNAseq data (Abbriano Burke,

unpublished). Thaps3_25595 has a sequence (ELKRIY) located 28 amino acids from the C-terminus that is highly similar to the yeast DGAT2 ER retrieval motif (ELKIVG), consistent with the idea that some ER targeting signal exists. If we assume that there is a cryptic ER signal sequence present, then the default targeting should be to the chloroplast, because of the presence of the chloroplast transit peptide. Related to the targeting question is the topology of Thaps3_25595. Topology studies in yeast and murine DGAT2 indicate ER (microsomal) localization, and suggest that both the N and C termini are positioned in the cytoplasm [37,40]. If that was the case with Thaps3_25595, then the chloroplast targeting signal would not be available for recognition, it has to be located in the ER lumen for that to occur. An interesting observation is that the first transmembrane segment overlaps with the chloroplast transit peptide sequence (Fig. 1), perhaps the transmembrane domain serves as an ER signal sequence (as occurs in some transporter proteins), and other factors influence the accessibility of the chloroplast transit peptide. This could suggest a post-translational regulatory mechanism involved in targeting. DGATs are known in other contexts to have differential targeting, in some organisms they are associated with the ER and also with lipid droplets that have no ER affiliation, although the proposed mechanism of relocation is not relevant to the diatom situation[41]. DGAT2 in mammalian cells has been shown to locate either in the ER or in mitochondrially-associated membranes that are specialized domains of the ER, and sequences responsible for the dual targeting have been identified [42].

The association of Thaps3_25595 with chloroplasts during exponential and early stationary phase growth suggests that TAG formation occurs in the chloroplast in diatoms. Chloroplast TAG formation has been reported in *C. reinhardtii*[16]. In fact, both chloroplast and cytoplasmically located TAG is derived from DAG synthesized in the chloroplast [16]. Over

50% of cellular lipid can be associated with the chloroplast [43]. Such a large demand may require the establishment of a TAG storage pool in the chloroplast to reduce short-term energetic and metabolic demands for making FAs for membranes during the period of chloroplast replication. This pool apparently does not aggregate into substantial lipid droplets under normal conditions; chloroplast TAG has been identified in small plastoglobules[16,44,45]. Similarly, in *T. pseudonana* or other diatoms, we have never observed BODIPY or Nile-red stainable material in the chloroplast, and freeze fracture images have not revealed sizable droplets in that location (Roth, Goodenough, Tesson, Hildebrand, unpublished observations). In *C. reinhardtii*, sizable chloroplast-localized lipid droplets do accumulate, but only in *sta6*starchless mutants at the time when starch normally accumulates [17], suggesting that the inability to store carbon as carbohydrate exposes a pre-existing pathway to accumulate TAG. None of the *C. reinhardtii* DGATs has predicted chloroplast targeting, prompting a hypothesis that chloroplast TAG synthesis is associated with the chloroplast envelope or that another set of enzymes may be involved in chloroplast TAG synthesis[15]. Correct predictive targeting is also an issue with Thaps3_25595. Similar to the *T. pseudonana* enzyme, all four of the *C. reinhardtii* DGATs have membrane spanning domains near the N-terminus (unpublished observation). Perhaps the enrichment of hydrophobic amino acids in this region obscures targeting predictive capabilities. To date, no direct localization of the *C. reinhardtii* DGATs has been reported. Our direct localization (Fig. 3) shows that Thaps3_25595 is present throughout the chloroplast and not just in the envelope.

Taken together, the data suggest that chloroplast TAG synthesis occurs in highly evolutionarily diverse classes of microalgae. It is unclear how much TAG can be accumulated in the chloroplast under normal growth conditions. The absence of large droplets does not

necessarily indicate that little is accumulated. Given the membrane-rich environment inside the chloroplast, perhaps TAG can associate with the substantial hydrophobic environment therein without being driven to accumulate in droplets. The data also suggest that unknown mechanisms may be involved to override the normal targeting processes in the DGAT enzymes targeted to the chloroplast. In *C. reinhardtii*, DGATs have no predicted chloroplast targeting, whereas in *T. pseudonana*, they do, but lack a clear ER signal prediction, which would be a requirement for chloroplast targeting. Also in *T. pseudonana* is the shift in localization between chloroplast and ER, which cannot be ascribed to a currently known mechanism.

Diatoms accumulate large lipid droplets in the cytoplasm resulting from silicon starvation [4,14,39,46]. The transition in location of Thaps3_25595 to the ER in stationary phase is consistent with the enzyme being positioned more closely to the site of major lipid droplet formation. Under silicon starvation, the lack of growth and requirement to replicate chloroplasts may obviate the need for chloroplast TAG synthesis. We show that Thaps3_25595 does not strictly co-localize with forming lipid droplets (Fig. 7), which does occur in some other organisms [41].

Overexpression of Thaps3_25595 resulted in significant increases in TAG during silicon starvation (Fig. 6). A simple explanation is that the increase is due to facilitated carbon flux through the last step in the TAG synthesis pathway. Particular steps in FA and TAG synthesis may be especially sensitive to manipulation, for example, work in the Aquatic Species Program in which the enzyme involved in the first committed step to FA synthesis, acetyl CoA-carboxylase (ACCase) was overexpressed, resulted in increased mRNA and enzymatic activity, but no increase in TAG [47].

Overexpression of a DGAT2 in *P. tricornutum* resulted in a slight decrease in biomass accumulation, and an apparent 1.35-fold increase in TAG, as evaluated by Nile red staining [13]. Unfortunately, only a single transgenic line was characterized in that paper, which doesn't allow an evaluation of the robustness of the data. In our experiments, we characterized 4 independent lines, and all showed improved TAG accumulation, with no detriment to growth or biomass accumulation (Figs. 4 and 6). There was an apparent positive correlation between expression level and TAG accumulation ability in transgenic line Tg-1 (Fig. 5 and 6).

DGATs play vital roles in controlling the quantitative [48] and qualitative [49-51] flux of fatty acids into storage TAGs. Interesting changes occurred in FA composition as a result of Thaps3_25595 overexpression (Fig. 6), summarized by a general decrease in C16:0 and C18:0, and increase in DHA and EPA. Niu et al.[13] over expressed a type 2 DGAT in *P. tricornutum*, and showed a substantial increase in EPA (DHA was not detectable), however that enzyme (Phatr2_49462) clades distantly from Thaps3_25595 (Fig. S2). In *T. pseudonana*, we have demonstrated an increase in DHA and EPA resulting from silicon-limitation [52], when TAG accumulates. Perhaps any trigger that stimulates TAG accumulation, such as Thaps3_25595 overexpression, shifts the FA composition similarly. In addition to higher TAG levels, the improvement in EPA and DHA increases the value of the lipid product.

Previous work in which three DGATs in *C. reinhardtii* were overexpressed did not result in increased TAG, or a shift in FA composition [53]. The authors concluded that further insights into lipid biosynthesis pathways were necessary to generate improvements in lipid yield. We agree that further insights will be beneficial, however our results demonstrate that simple manipulations can be productive, and that generalizations based on examination of one algal species are not necessarily universal. Although the situation may be more complicated in *C.*

433 *reinhardtii*, we note that overexpression in that study was not demonstrated under conditions of
434 lipid accumulation [53], so there may be another explanation for the lack of response.

435 We previously reported [14] on the downregulation of a TAG lipase leading to substantially
436 increased TAG levels in transgenic lines. This demonstrated that affecting the cell's ability to
437 utilize TAG could increase TAG yields. The current report demonstrates that increasing the
438 cell's ability to make TAG also increases TAG yields. The fact that distinct types of
439 manipulations lead to improved TAG productivity is encouraging for metabolic engineering
440 approaches in general. Specific metabolic steps may influence other cellular processes in various
441 ways, by having multiple options for manipulation, possible detrimental effects may be avoided,
442 and the best manipulation may be identified. Identifying multiple steps also enables coordinate
443 manipulation of multiple genes, "trait stacking", which could synergistically increase yields. The
444 first two engineering approaches we have reported on were both successful in terms of increasing
445 TAG yield without a detriment to growth. This suggests that choosing gene targets based on
446 fundamental metabolic principles is a valid initial approach for attempts to improve productivity,
447 and that a possible metabolic drain of increased TAG can be managed by the cell.

456 **Acknowledgements**

457 We appreciate the help of Sarah R. Smith in processing and analysis of the microarray data, and
458 Raffaella Abbriano Burke for processing of the RNAseq data. The research was supported by Air
459 Force Office of Scientific Research (AFOSR) grants FA9550-08-1-0178 and FA9550-08-1-
460 0178, US Department of Energy grants DE-EE0001222 and DEEE0003373, National Science
461 Foundation grant CBET-0903712, California Energy Commission's 'California Initiative for
462 Large Molecule Sustainable Fuels', agreement number: 500-10-039.

References

1. P. Biller, A.B. Ross, Potential yields and properties of oil from the hydrothermal liquefaction of microalgae with different biochemical content, *Bioresource Tech.* 102 (2001) 215-225.
2. S. Sawayama, T. Minowa, S.Y. Yokoyama, Possibility of renewable energy production and CO₂ mitigation by thermochemical liquefaction of microalgae, *Biomass Bioenergy* 17 (1999) 33-39.
3. R. Davis, A. Aden, P.T. Pienkos, Techno-economic analysis of autotrophic microalgae for fuel production, *Appl. Energy* 88 (2011) 3524-3531.
4. P. Roessler, Effects of silicon deficiency on lipid composition and metabolism in the diatom *Cyclotella cryptica*, *J. Phycology* 24 (1988) 394-400.
5. Q. Hu, M. Sommerfeld, E. Jarvis, M. Ghirardi, Microalgal triacylglycerols as feedstocks for biofuel production: perspectives and advances, *Plant J.* 54 (2008) 621-639.
6. Y. Li, D. Ham, G. Hu, M. Sommerfeld, Q. Hu, Inhibition of starch synthesis results in overproduction of lipids in *Chlamydomonas reinhardtii*, *Biotechnol. Bioeng.* 107 (2010) 258-268.

- 486
- 487 7. V. H. Work, R. Radakovits, R.E. Jinkerson, J.E. Meuser, L.G. Elliott,
- 488 D.J.Vinyard, L.M.L. Laurens, G.C. Dismukes, M.C. Posewitz, Increased lipid
- 489 accumulation in the *Chlamydomonas reinhardtii* sta7-10 starchless isoamylase
- 490 mutant and increased carbohydrate synthesis in complemented strains, Eukaryot.
- 491 Cell 9 (2010) 1251–1261.
- 492
- 493 8. T.Y. Doan, J.P. Obbard, Enhanced intracellular lipid in *Nannochloropsis* sp. via
- 494 random mutagenesis and flow cytometric cell sorting. Algal Res. 1 (2012) 17–21.
- 495
- 496 9. K. Manandhar-Shrestha, M. Hildebrand, Development of flow cytometric
- 497 procedures for the efficient isolation of improved lipid accumulation mutants in a
- 498 *Chlorella* sp. microalga, J. Appl. Phycol. 25 (2013) 1643–1651.
- 499
- 500 10. B. Xie, D. Stessman, J.H. Hart, H.L. Dong, Y.J. Wang, D.A. Wright, B.J. Nikolau,
- 501 M.H. Spalding, L.J. Halverson, High-throughput fluorescence-activated cell
- 502 sorting for lipid hyperaccumulating *Chlamydomonas reinhardtii* mutants, Plant
- 503 Biotech. J. 12 (2014) 872–882.
- 504
- 505 11. L. de Jaeger, R.E.M. Verbeek, R.B. Draaisma, D.E. Martens, J. Springer, G.
- 506 Eggink, R.H. Wijffels, Superior triacylglycerol (TAG) accumulation in starchless
- 507 mutants of *Scenedesmus obliquus*: (I) mutant generation and characterization.
- 508 Biotech. Biofuels 7 (2014): 69.

12. M. Terashima, E.S. Freeman, R.E. Jinkerson, M.C. Jonikas, A fluorescence-activated cell sorting-based strategy for rapid isolation of high-lipid *Chlamydomonas* mutants, *Plant J.* 81 (2015) 147-159.
13. Y.F. Niu, M.H. Zhang, D.W. Li, W.D. Yang, J.S. Liu, W.B. Bai, H.Y. Li, Improvement of neutral lipid and polyunsaturated fatty acid biosynthesis by overexpressing a type 2 diacylglycerol acyltransferase in marine diatom *Phaeodactylum tricornutum*, *Mar Drugs* 11(2013) 4558-4569.
14. E.M. Trentacoste, R. Shrestha, S.R. Smith, C. Glé, A.C. Hartmann, M. Hildebrand, W.H. Gerwick, Metabolic engineering of lipid catabolism increases microalgal lipid accumulation without compromising growth, *Proc. Natl. Acad. Sci. USA.* 110 (2013) 19748-19753.
15. B. Liu, C. Benning, Lipid metabolism in microalgae distinguishes itself, *Curr. Opin. Biotech.* 24 (2013) 300–309.
16. J. Fan, C. Andre, C. Xu, A chloroplast pathway for the de novo biosynthesis of triacylglycerol in *Chlamydomonas reinhardtii*. *FEBS Lett.* 585 (2011) 1985–1991.
17. C. Goodson, R. Roth, Z.T. Wang, U. Goodenough, Structural correlates of cytoplasmic and chloroplast lipid body synthesis in *Chlamydomonas*

532 *reinhardtii* and stimulation of lipid body production with acetate boost, Euk. Cell
533 10 (2011) 1592–1606.

534
535 18. Hildebrand, M, Abbriano, RM, Polle, JEW, Traller, JC, Trentacoste, EM, Smith,
536 SR, Davis, AK. 2013. Metabolic and cellular organization in evolutionarily
537 diverse microalgae as related to biofuels production. Curr. Opin. Chem. Biol.
538 17:506-514

539
540 19. H. Cao, J.M. Shockey, K.T. Klasson, D.C. Chapital, C.B. Mason, B.E. Scheffler,
541 Developmental regulation of diacylglycerol acyltransferase family gene
542 expression in tung tree, PLOS ONE 8 (2013) e76946.

543
544 20. M. Luisa Hernandez, L. Whitehead, Z.S. He, V. Gazda, A. Gilday, E.
545 Kozhevnikova, F.E. Vaistij, T.R. Larson, I.A. Graham, A cytosolic
546 acyltransferase contributes to triacylglycerol synthesis in sucrose-rescued
547 *Arabidopsis* seed oil catabolism mutants, Plant Physiol. 160 (2012) 215-225.

548
549 21. S. Cases, S.J. Smith, Y.W. Zheng, H.M. Myers, S.R. Lear, et al. Identification of
550 a gene encoding an acyl CoA :diacylglycerol acyltransferase, a key enzyme in
551 triacylglycerol synthesis, Proc. Natl. Acad. Sci. USA. 95 (1998) 13018-13023.

552
553 22. K.D. Lardizabal, J.T. Mai, N.W. Wagner, A. Wyrick, T. Voelker, D.J. Hawkins,
554 DGAT2 is a new diacylglycerol acyltransferase gene family - Purification,

cloning, and expression in insect cells of two polypeptides from
Mortierellaramanniana with diacylglycerol acyltransferase activity, J. Biol.
Chem. 276 (2001) 38862-38869.

23. S. Saha, B. Enugutti, S. Rajakumari, R. Rajasekharan, Cytosolic triacylglycerol
biosynthetic pathway in oilseeds. Molecular cloning and expression of peanut
cytosolic diacylglycerol acyltransferase, Plant Physiol. 141 (2006) 1533-1543.

24. J. Xu, M. Kazachkov, Y. Jia, Z. Zheng, J. Zou, Expression of a type 2
diacylglycerol acyltransferase from *Thalassiosira pseudonana* in yeast leads to
incorporation of docosahexaenoic acid beta-oxidation intermediates into
triacylglycerol. FEBS J. 280 (2013) 6162-6172.

25. Y. Gong, J. Zhang, G. Junping; X.J. Guo, X. Wan, Z. Liang, C.J. Hu, M.L. Jiang,
Identification and characterization of PtDGAT2B, an acyltransferase of the
DGAT2 acyl-Coenzyme A: Diacylglycerol acyltransferase family in the diatom
Phaeodactylum tricornutum, FEBS Lett. 587 (2013) 481-487.

26. W.M. Darley, B.E. Volcani, Role of silicon in diatom metabolism. A silicon
requirement for deoxyribonucleic acid synthesis in the diatom *Cylindrotheca*
fusiformis Reimann and Lewin, Exp. Cell Res. 58 (1969) 334-342.

27. N. Poulsen, P.M. Chesley, N. Kroger, Molecular genetic manipulation of the diatom *Thalassiosira pseudonana* (Bacillariophyceae), J. Phycol. 42 (2006) 1059-1065.
28. CLC main workbench 6.0 program, <http://www.clcbio.com/products/clc-main-workbench/>.
29. S.F. Altschul, W. Gish, W. Miller, E.W. Myers, D.J. Lipman, Basic local alignment search tool. J MolBiol 215 (1990) 403-410.
30. I. Letunic, T. Doerks, P. Bork. SMART: recent updates, new developments and status in 2015, Nuc Acids Res 43 (2015) D257-D260.
31. TMHMM Server, <http://www.cbs.dtu.dk/services/TMHMM/>
32. J.D. Bendtsen, H. Nielsen, G.vonHeijne, S.Brunak, Improved prediction of signal peptides: SignalP 3.0, J. Mol. Biol. 340 (2004) 783-795.
33. T.N. Petersen, S. Brunak, G. von Heijne, H. Nielsen, SignalP 4.0: discriminating signal peptides from transmembrane regions, Nature Meth. 8 (2011) 785-786.

598 34. O. Emanuelsson, H. Nielsen, G. von Heijne, ChloroP, a neural network-based
599 method for predicting chloroplast transit peptides and their cleavage sites, Protein
600 Sci. 8 (1999) 978-984.

601 |

602 35. J.E. Chen, A.G. Smith, A look at diacylglycerol acyltransferases (DGATs) in
603 algae. J. Biotech. 162 (2012) 28-39.

604

605 36. J.F. Lockwood, J. Cao, P. Burn, Y. Shi, Human intestinal monoacylglycerol
606 acyltransferase: differential features in tissue expression and activity. American
607 journal of physiology, Endocrin. Metabol. 285 (2003) E927-937.

608

609 37. Q. Liu, R.M. Siloto, C.L. Snyder, R.J. Weselake, Functional and topological
610 analysis of yeast acyl-CoA:diacylglycerol acyltransferase 2, an endoplasmic
611 reticulum enzyme essential for triacylglycerol biosynthesis. J. Biol. Chem. 286
612 (2011) 13115-13126.

613

614 38. M. Lang, K.E. Apt, P.G. Kroth, Protein transport into "Complex" diatom plastids
615 utilizes two different targeting signals, J. Biol. Chem. 273 (1999) 30973-30978.

616

617 39. E.T. Yu, F.J. Zendejas, P.D. Lane, S. Gaucher, B.A. Simmons, T.W. Lane,
618 Triacylglycerol accumulation and profiling in the model diatoms *Thalassiosira*
619 *pseudonana* and *Phaeodactylum tricornutum* (Bacillariophyceae) during
620 starvation, J Appl. Phycol. 21 (2009) 669-681.

621
622
623
624
625
626
627
628
629
630
631
632
633
634
635
636
637
638
639
640
641
642
643

40. S.J. Stone, M.C. Levin, R.V. Farese Jr. , Membrane topology and identification of key functional amino acid residues of murine acyl-CoA:diacylglycerol acyltransferase-2, J. Biol. Chem. 281 (2006) 40273-40282.

41. F. Wilfling, H. Wang, J.T. Haas, N. Krahmer, T.J. Gould, et al., Triacylglycerol synthesis enzymes mediate lipid droplet growth by relocating from the ER to lipid droplets, Dev. Cell 24 (2013) 384-399.

42. S.J. Stone, M.C. Levin, P. Zhou, J. Han, T.C. Walther, R.V. Farese Jr. The endoplasmic reticulum enzyme DGAT2 is found in mitochondria-associated membranes and has a mitochondrial targeting signal that promotes its association with mitochondria. J. Biol. Chem. 284: (2009) 5352-5361.

43. D.R. Janero, R. Barnett, Isolation and characterization of an ether-linked homoserine lipid from the thylakoid membrane of *Chlamydomonas reinhardtii* 137+, J. Lipid Res. 23: (1982) 307-316.

44. J.R. Austin II, E. Frost, P.-A. Vidi, F. Kessler, L.A. Staehelin, Plastoglobules are lipoprotein subcompartments of the chloroplast that are permanently coupled to thylakoid membranes and contain biosynthetic enzymes, Plant Cell 18 (2006) 1693–1703.

- 667 51. R.Z. Li, K.S. Yu, T. Hatanaka, D.F. Hildebrand, *Vernonia* DGATs increase
668 accumulation of epoxy fatty acids in oil, *Plant Biotechnol. J.* 8 (2010) 184-195.
669
- 670 52. O. Cook, M. Hildebrand, Enhancing LC-PUFA production in *Thalassiosira*
671 *pseudonana* by overexpressing the endogenous fatty acid elongase genes. *J. Appl.*
672 *Phycol.* (2015) DOI 10.1007/s10811-015-0617-2
673
- 674 53. M. La Russa, C. Bogen, A. Uhmeyer, A. Doebbe, E. Filippone, O. Kruse, J.H.
675 Mussnug, Functional analysis of three type-2 DGAT homologue genes for
676 triacylglycerol production in the green microalga *Chlamydomonas reinhardtii*. *J.*
677 *Biotechnol.* 162 (2012) 13-20.
678
679

Figure Legends

Fig. 1. Domain structure of Thaps3_25595 and identification of key functional amino acid residues. Upper, amino acid sequence of Thaps3_25595 and position of domains. Lower, key functional amino acids in DGAT2's from diverse organisms. Proteins used in the analysis were available from the NCBI database (*Thalassiosira pseudonana* XP_002294564.1; *Fragilariopsis cylindrus* XP_001019739.1; *Phaeodactylum tricornutum* XP_002184163.1; *Ectocarpus siliculosus* CBN79671.1; *Cyanidioschyzon merolae* XP_005534894.1; *Galdieria sulphuraria* XP_005708616.1; *Ostreococcus tauri* XP_003083539.1; *Chlamydomonas reinhardtii* AFB73929.1; *Arabidopsis lyrata* XP_002876092.1; *Zea mays* NP_001150174.1; *Glycine max* XP_006599271.1; *Ricinus communis* XP_002528531.1; *Vernicia fordii* ABC94473.1; *Schizosaccharomyces pombe* XP_001713160.1; *Aspergillus fumigatus* XP_749138.1; *Paracoccidioides* sp. XP_002796975.1; *Rhodospiridium toruloides* CDR39395.1; *Macrophomina phaseolina* EKG14568.1; *Homo sapiens* NP_835470.1; *Ornithorhynchus anatinus* XP_001518899.1; *Gallus gallus* XP_003641802).

Fig. 2. Phylogenetic tree showing relationships among the DGAT2 protein sequences from algae, plant, mammals and pathogens. The tree was generated using the CLC workbench 6 software using the same amino acid sequences as in Fig. 1. Bootstrap values are shown in percentages at nodes.

Fig. 3. Subcellular localization of Thaps3_25595. Exponential (growth stage) and early stationary phase transgenic lines were examined. Thaps3_25595 protein was fused with GFP, the native sequence was fused with TagRFP, and visualized with a fluorescence microscope.

Fig. 4. Growth of wild type and transgenic cells. Wild-type and transgenic Thaps3_25595-overexpressing cells were inoculated at 0.1×10^6 cells $\cdot$ mL⁻¹. Cells were grown at 18°C under continuous light in artificial sea water. Growth was monitored by counting cell number. Values are means $\pm$ SE ($n = 4$).

Fig. 5. Expression of Thaps3_25595 protein and mRNA in *T. pseudonana*. A) Transcript profile of Thaps3_25595 in wild type cells under silicon starvation (Smith et al., submitted) and the relative TAG content over time. B) Transcript profile of fucoxanthin chlorophyll a/c-binding mRNA (Thaps3_268127) in *T. pseudonana* cells under silicon starvation (Smith et al., submitted). C) Expression of Thaps3_25595 TagRFP protein in transgenic clones, monitored by TagRFP fluorescence intensities. TagRFP fluorescence intensities were monitored in an imaging flow cytometer. Values are means $\pm$ SE ($n = 5$).

Fig. 6. TAG accumulation in wild type and transgenic Thaps3_25595 overexpression cells. Exponentially grown wild type and transgenic cells were placed into silicon-free medium. Cell samples at indicated time points were isolated and stained with Nile red to monitor TAG level using imaging flow cytometry. Values are means $\pm$ SE ($n = 5$).

Fig. 7. Subcellular localization of lipid droplets and Thaps3_25595 protein. Cells were stained with BODIPY and lipid droplets were visualized under a fluorescence microscope. A and B) exponentially growing cells. C) stationary phase cells.

Fig. 8. Ratio of fatty acid constituents in overexpressing Thaps3_25595 transgenic lines of *T. pseudonana* at 0 hr and 24 hr of silicon starvation relative to wild-type. Two transgenic lines were examined.

Figureg

Figure 1

Figure 2

Figure 3

Figure 4

Figure 5

Figure 6

Figure 7

Figure 8