

# Identification of aquatically available carbon from algae through solution-state NMR of whole $^{13}\text{C}$ -labelled cells

Mohammad Akhter<sup>1,2</sup> · Rudraksha Dutta Majumdar<sup>1,2</sup> · Blythe Fortier-McGill<sup>1,2</sup> · Ronald Soong<sup>1,2</sup> · Yalda Liaghati-Mobarhan<sup>1,2</sup> · Myrna Simpson<sup>1,2</sup> · George Arhonditsis<sup>2</sup> · Sebastian Schmidt<sup>3</sup> · Hermann Heumann<sup>3</sup> · André J. Simpson<sup>1,2</sup>

Received: 9 February 2016 / Revised: 24 March 2016 / Accepted: 30 March 2016 / Published online: 13 April 2016  
© Springer-Verlag Berlin Heidelberg 2016

**Abstract** Green algae and cyanobacteria are primary producers with profound impact on food web functioning. Both represent key carbon sources and sinks in the aquatic environment, helping modulate the dissolved organic matter balance and representing a potential biofuel source. Underlying the impact of algae and cyanobacteria on an ecosystem level is their molecular composition. Herein, intact  $^{13}\text{C}$ -labelled whole cell suspensions of *Chlamydomonas reinhardtii*, *Chlorella vulgaris* and *Synechocystis* were studied using a variety of 1D and 2D  $^1\text{H}/^{13}\text{C}$  solution-state nuclear magnetic resonance (NMR) spectroscopic experiments. Solution-state NMR spectroscopy of whole cell suspensions is particularly relevant as it identifies species that are mobile (dissolved or dynamic gels), ‘aquatically available’ and directly contribute to the aquatic carbon pool upon lysis, death or become a readily available food source on consumption. In this study, a wide range of metabolites and structural components were identified within the whole cell suspensions. In addition, significant differences in the lipid/triacylglyceride (TAG) content of green algae and cyanobacteria were confirmed. Mobile species in algae are quite different from those in abundance

in ‘classic’ dissolved organic matter (DOM) indicating that if algae are major contributors to DOM, considerable selective preservation of minor components (e.g. sterols) or biotransformation would have to occur. Identifying the metabolites and dissolved components within algal cells by NMR permits future studies of carbon transfer between species and through the food chain, whilst providing a foundation to better understand the role of algae in the formation of DOM and the sequestration/transformation of carbon in aquatic environments.

**Keywords** Solution-state NMR · HMQC · Algae · Metabolomics · Dissolved organic matter

## Introduction

Green algae are a diverse group of eukaryotic organisms found in both freshwater and saltwater ecosystems. Blue-green algae (often termed cyanobacteria) are also prolifically found in water systems under certain nitrogen and phosphorus balances, outcompeting green algae at certain equilibria [1]. Both are autotrophic and carry out photosynthesis. As a result, both play a fundamental role in maintaining the energy balance in aquatic ecosystems as they belong to a lower trophic level and changes in their populations are magnified through the food web. In addition, green algae and cyanobacteria may play a role in the formation of dissolved organic matter (DOM), a topic of interest to a diversity of water researchers, as well as form a primary source for the production of biofuels.

Importantly, green algae and cyanobacteria have different molecular compositions and homeostatic set points, specifically concerning lipids and their metabolism [2]. This is of particular importance when cyanobacteria outcompete green algae in freshwater lakes, as this can have adverse effects on

**Electronic supplementary material** The online version of this article (doi:10.1007/s00216-016-9534-8) contains supplementary material, which is available to authorized users.

✉ André J. Simpson  
andre.simpson@utoronto.ca

<sup>1</sup> Environmental NMR Centre, University of Toronto Scarborough, 1265 Military Trail, Toronto, Ontario M1C 1A4, Canada

<sup>2</sup> Department of Physical and Environmental Sciences, University of Toronto Scarborough, 1265 Military Trail, Toronto, Ontario M1C 1A4, Canada

<sup>3</sup> Silantes GmbH, Gollierstr. 70 C, 80339 München, Germany

micro-crustaceans such as *Daphnia magna* (a keystone species) [3]. For example, exposure to cyanobacteria has been shown to increase the number of aborted eggs in *D. magna*, thus decreasing population size which directly impacts higher trophic levels [4]. Although the primary mechanisms by which these patterns are triggered is unclear, cyanobacteria are known to be of poor food quality as a result of mechanical interference, toxicity and (or) inadequate biochemical nutritional value (e.g. low content in omega-3 highly unsaturated fatty acids, such as docosahexaenoic and eicosapentaenoic acids, and sterols) for consumers. Consumers exclusively fed cyanobacteria typically exhibit near-starvation traits [5].

Furthermore, the amount and composition of DOM in water systems are of current interest, as DOM comprises a major percent of mobile organic carbon and plays a role in increasing the solubility of environmental contaminants [6]. This work will shed light on the metabolite profile of different algal species and the structural contributions that they can make to DOM when their cell walls are spliced. In particular, as low power solution-state nuclear magnetic resonance (NMR) techniques are employed (see ‘Conclusion’ section), soluble and mobile gel-like components (i.e. those readily dispersed into the aqueous phase) are emphasised within the whole cells. By understanding the potential flux of mobile metabolites and soluble precursors from algae into DOM, researchers may be able to better predict changes in DOM in freshwater in relation to the amount of algae or cyanobacteria present.

Further interest in algae stems from their application in the production of biofuels. Biofuels are synthesised from lipid oils in algae which may account for up to 50 % of the algae by weight in certain species [7]. Algae can be grown very rapidly in bioreactors and will likely become an important source for renewable energy in the future [7]. Current research on algae focuses on identifying the best (greatest yield of biodiesel per unit mass algae) species out of an estimated  $3 \times 10^5$  species, where a number of factors must be considered including the growth rate and the lipid oil content [8, 9]. Underlying these factors is the molecular composition of algae and the metabolomic changes that occur in response to altering environmental conditions. Lipids in cells may exist in different physical forms; however, it is the extractable lipid component that makes up the primary building block for the synthesis of biofuels and identifying this extractable lipid component is a challenge that researchers face in identifying the most viable species [10]. The first step in biofuel production is the extraction of lipids followed by esterification. Therefore, it is paramount that researchers identify the percent of lipids that are mobile within whole cells and thus readily available for simple extraction [10]. This work aims to identify and quantify a range of dissolved and mobile species within whole algae cells which includes the lipid fraction. The importance of algae for aquatic ecosystems, potential contribution to DOM formation and the production of biofuels drives a fundamental need to

better understand the metabolomic and structural profile of algae. This understanding will equip researchers with the knowledge necessary to precisely identify the impact of algal and cyanobacterial blooms at an ecosystem level as well as to develop genetically modified algae strains that will make bio-fuel production more economically viable [11].

The metabolite profile of different species of algae and cyanobacteria has been studied using various analytical techniques including gas chromatography mass spectrometry–time of flight–mass spectrometry (GC-MS TOF-MS) and high performance liquid chromatography inductively coupled plasma mass spectrometry (HPLC ICP-MS). Using GC-MS TOF-MS, the metabolomic changes in wild-type *Chlamydomonas reinhardtii* under nitrogen, phosphorus, sulphur, or iron-deprived conditions were examined [12]. Furthermore, HPLC ICP-MS was employed in identifying the speciation of arsenosugars in the brown alga *Fucus serratus* [13]. Such studies provide an excellent and sensitive targeted approach to evaluate specific metabolites in extracts but are often restricted to small molecular components that can be easily extracted and derivatized. Additionally, MS techniques are selective and reliant on MS libraries and thus unknown compounds may be missed [14]. In contrast, NMR requires little sample preparation and is a non-selective detector, with the main limitation being spectral overlap which can hamper assignment and quantification in complex matrices [15].

Some limited NMR studies have also been conducted on algae and cyanobacteria [16, 17]. Specifically, solution-state NMR was used to build a limited metabolomic profile of the algal species *Chlorophyta* and *Rhosophyta* with natural abundance  $^{13}\text{C}$  [16]. In addition, solid-state NMR has been used to study the light harvesting complexes (important for photosynthesis) in *C. reinhardtii* [17] and also to characterise the lipid and carbohydrate constituents of whole cells of marine (*Pavlova lutheri* and *Nannochloropsis oculata*) and freshwater (*C. reinhardtii*) microalgae which have different cell wall characteristics [18]. High resolution–magic angle spinning (HR-MAS)  $^1\text{H}$  and  $^{13}\text{C}$  NMR spectroscopy have also been shown to be useful for studying microalgal whole cells [19–21]. However, MAS tends to average anisotropic interactions and sharpens the signals from semi-solids and gels. Whilst this is often an advantage, if only dissolved metabolites are of interest, then it has been noted that peak resolution at the baseline tends to be poorer in the HR-MAS spectra compared to corresponding solution-state spectra, and chemical shifts can be slightly different [21]. To the best of our knowledge, a comparative study on the different species of  $^{13}\text{C}$ -enriched green algae and cyanobacteria whole cell suspensions using solution-state NMR including a diversity of 1D and 2D experiments has yet to be employed in the establishment of a more complete metabolic profile. Solution-state NMR provides high resolution molecular information concerning components in the sample that are dissolved or highly mobile in

their natural state within whole algal cells [22] and represents most common NMR technique available in the vast majority of NMR facilities. Importantly, most biologically relevant metabolites are found and operate in the cytoplasm of cells and are therefore visible through solution-state NMR. This makes solution-state NMR ideal for metabolomic studies. Furthermore, a thorough understanding of mobile metabolites will shed light on the contributions of algae to the structural components of DOM as these components will contribute directly to the aquatic environment on lysis.

In this study, the metabolomic profiles of two different strains of  $^{13}\text{C}$  isotopically labelled green algae and one cyanobacterium were analysed using solution-state NMR spectroscopy. Specifically, the strains *C. reinhardtii* (unicellular green algae with two characteristic flagella found in soils and fresh water), *Chlorella vulgaris* (unicellular green algae with no flagella found in fresh water) and *Synechocystis* (cyanobacteria found in fresh water) were examined. In a natural environment, the presence of flagella allows the former species to demonstrate phototactic behaviour and optimise their exposure to light for photosynthesis. The second species is considered a high quality food source for primary consumers with high protein content and other essential nutrients. The latter species is one of the most extensively studied types of cyanobacteria as it has the capacity to demonstrate both phototrophic growth by oxygenic photosynthesis in sunlight and heterotrophic growth by glycolysis and oxidative phosphorylation during dark periods. The evaluation of the solution-state NMR spectroscopic approach, in particular the ability to identify key metabolites and major differences between species, will facilitate future in vivo NMR-based metabolomic studies (for example feeding  $^{13}\text{C}$ -enriched algae to *D. magna* and following real-time metabolism). Additionally, understanding the physical state in which different metabolites are found will shed light on their ability to contribute to DOM as well as the potential extractability of lipid components for the production of biofuels.

## Materials and methods

### $^{13}\text{C}$ labelling of algae

The algae strains *C. reinhardtii*, *C. vulgaris* and the cyanobacteria *Synechocystis* were cultivated in a small scale closed loop system photobioreactor designed and built by Silantes GmbH. Each fermentation was carried out autotrophically and entirely with  $^{13}\text{CO}_2$  (enrichment 99%; purchased from Sigma-Aldrich, Isotec, St. Louis, MO). The environmental conditions for growth including media, temperature, light intensity and pH were optimised for maximum growth for each species. The biomass was then harvested and freeze dried for storage at  $-20\text{ }^\circ\text{C}$  prior to use in the NMR experiments,

which is the standard practice for microalgal whole cell samples [19, 20]. For further details on labelling methodology, refer to the Electronic Supplementary Material (ESM).

### Sample preparation

Uniformly labelled  $^{13}\text{C}$  algae strains *C. reinhardtii*, *C. vulgaris* and the cyanobacteria *Synechocystis* were studied using solution-state NMR spectroscopy. Twenty milligrams (dry weight) of each was suspended in 100 %  $\text{D}_2\text{O}$  (Cambridge Isotopes, Tewksbury, MA) and added to a high-precision 5-mm NMR tube (508-UP-7, Norell, Morganton, NC) for data collection.

### 1D NMR spectroscopy

The experiments were performed using a Bruker Avance III NMR spectrometer with a  $^1\text{H}$  Larmor frequency of 500.13 MHz equipped with a 5-mm four-channel ( $^1\text{H}$ ,  $^{13}\text{C}$ ,  $^{15}\text{N}$ ,  $^{19}\text{F}$ ) QXI NMR probe, fitted with an actively shielded z-gradient (Bruker BioSpin) at room temperature. A  $\text{D}_2\text{O}$  Lock was used for all experiments. Decoupling was used in all 1D and 2D experiments to remove  $^1\text{H}$ - $^{13}\text{C}$  coupling from the  $^{13}\text{C}$ -enriched sample. The decoupling schemes GARP-4 and WALTZ16 were used for the proton and carbon observed experiments, respectively. All NMR spectra were processed via TopSpin 3.1.

All  $^1\text{H}$  NMR spectra were recorded using the SPR-W5 WATERGATE water suppression sequence using a binomial delay of 125  $\mu\text{s}$  and GARP-4  $^{13}\text{C}$  decoupling during acquisition [23]. A  $90^\circ$  pulse was calibrated for each sample. A spectral width of 20 ppm was collected, and 512 scans were acquired with 16 k time domain points. The  $^1\text{H}$   $T_1$  was measured at 1 s using an inverse recovery sequence, and the recycle delay set five times this.  $^1\text{H}$  NMR spectra were processed using a zero filling factor of 2 and an exponential function corresponding to a line broadening of 0.3 Hz. For 1D  $^{13}\text{C}$  NMR spectra, a spectral width of 400 ppm was collected, 4 k scans were acquired with 16 k time domain points and a recycle delay was 5 s. Note that the carbon spectra were used for relative but not absolute quantification.  $^{13}\text{C}$  NMR spectra were processed using a zero filling factor of 2 and an exponential function corresponding to a line broadening of 5 Hz.

### Spectral editing

Diffusion edited (DE) proton and carbon spectra were produced using a bipolar pulse pair longitudinal encode-decode (BPLED) sequence with inverse gated decoupling [24]. Spectra were collected using encoding/decoding gradients of 2.5 ms at 55 gauss/cm and a diffusion time of 200 ms. DE  $^1\text{H}$  spectra and  $^{13}\text{C}$  spectra were processed using a zero filling factor of 2 and an exponential function corresponding to a line

broadening of 1.0 and 25 Hz, respectively. Inverse diffusion edited (IDE) and recovering relaxation losses from diffusion editing (RADE) were created via subtraction from the appropriate controls as previously described [25]. The IDE proton spectra and carbon spectra were processed using a zero filling factor of 2 and an exponential function corresponding to a line broadening of 1.0 and 5.0 Hz, respectively. The RADE proton spectra and carbon spectra were processed using a zero filling factor of 2 and an exponential function corresponding to a line broadening of 10.0 and 25.0 Hz, respectively. For spectral editing, the spectra were scaled until the spectrum being subtracted was nulled leaving a difference spectrum containing positive peaks [25]. The spectra were referenced to a D<sub>2</sub>O lock during acquisition and were then calibrated against a series of known compounds in the Bruker Biofluid Reference Compound Database (v 2.0.3).

## 2D NMR spectroscopy

2D (<sup>1</sup>H-<sup>13</sup>C) hetero-nuclear multiple-quantum correlation (HMQC) spectra were acquired using 2 k time domain points, 64 scans and 128 increments in the indirect dimension and a recycle delay of 1 s ( $1 \times T_1$ ). HMQC data were used for relative but not absolute quantification. HMQC spectra were processed with an exponential function corresponding to a line broadening of 15 Hz in F2 and using a q sine function in F1 shifted by  $\pi/2$  in both dimensions.

2D <sup>1</sup>H-<sup>1</sup>H total correlation spectroscopy (TOCSY) spectra were acquired in the phase-sensitive mode, using a mixing sequence with 300  $\mu$ s 40 KHz WURST-2 pulses within an X\_M16 mixing scheme [26]. Two thousand forty-eight data points (F2), 32 scans and a mixing time of 250 ms were used for each of the 128 increments in F1. F1 and F2 dimensions were processed using sine-squared functions with a  $\pi/2$  phase shift and a zero filling factor of 2.

## Compound identification

The spectra were calibrated against a series of known compounds in the Bruker Biofluid Reference Compound Database (v 2.0.3). Pattern matching of both 1D/2D spectra was performed using Analysis of MIXtures (AMIX, v 3.9.3, Bruker BioSpin) against the Bruker Biofluid Reference Compound Database (v 2.0.0 to v 2.0.3) using a procedure developed for complex mixtures [27]. Compounds with a greater than 80 % match (automated search) were selected for manual inspection. The chemical shifts of the identified compounds were compared with database values ( $r^2=0.99$ ,  $\sigma=0.01$ ) to confirm matching, and any compounds not meeting these requirements were removed. Major assignments were further checked for consistency against TOCSY data where possible and then further checked against the literature for consistency. Readers should note that error bars provided in the paper

represent the variance arising from processing and spectral analysis but due lack of biomass (and long NMR acquisition times) the samples could not be run in triplicate. Integration was performed using the ‘multi-integrate’ module within AMIX (v 3.9.3, Bruker BioSpin). The threshold was set at 1 % above the noise and ‘pattern files’ created for all metabolites by selecting the extremities of the peaks at the 1 % threshold. These files were then used to calculate integrals across the spectra in an automated fashion.

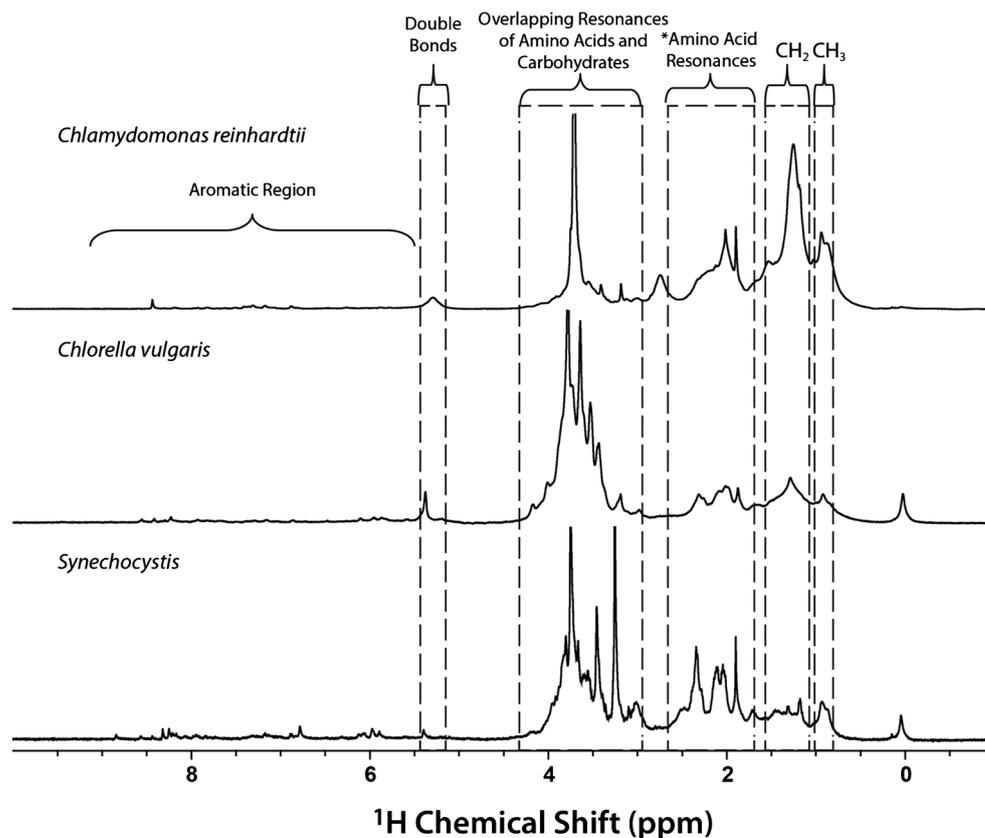
## Results and discussion

### 1D <sup>1</sup>H and <sup>13</sup>C NMR spectra of <sup>13</sup>C-enriched green algae and cyanobacteria

The 1D proton spectra of whole cell <sup>13</sup>C-enriched green algae and cyanobacteria suspended in D<sub>2</sub>O acquired using the SPR-W5 WATERGATE water suppression sequence are shown in Fig. 1. It is important to note that solution-state NMR only detects the soluble and dynamic components (true solids will not be observed using this technique). As such, throughout this paper, when for example the percentage difference of a component between the different species is compared, the comparison is specifically for the fraction of material that in the natural state (whole cells in water) is swollen/dissolved. This may not be the same as the percent fraction in total biomass. Due to its aquatic relevance, the swollen/dissolved fraction of algae most readily dispersed on lysis is the focus of this work. The swellability of each algae type varies resulting in differing signal strength from lipid, carbohydrate and protein components which may not be representative of the total amount of each component present. Therefore, in order to compare the differences more accurately, a relative swellability factor was calculated by integrating the total area in the <sup>13</sup>C spectra and normalising to *C. reinhardtii* (ESM Fig. S1). *C. reinhardtii* was determined to be the most swellable with *C. vulgaris* and *Synechocystis* exhibiting a swellability factor of 0.35 and 0.27, respectively, relative to *C. reinhardtii*. As a result, *C. reinhardtii* is ~3 times more swellable than the other two species thus providing more bio-available carbon (dissolved phase) on ingestion by grazers and to the aquatic environment on lysis.

The 1D <sup>1</sup>H NMR spectra show resonances that are consistent with signals for structural and metabolic components including lipids, carbohydrates and amino acids (Fig. 1). The lipids are most clearly defined by the -CH<sub>3</sub>, -CH<sub>2</sub> and double bond resonances and are most abundant in the *C. reinhardtii*, followed by *C. vulgaris* and with the least in the *Synechocystis*. Conversely, *C. vulgaris* and *Synechocystis* appear to contain more carbohydrates (3–4.2 ppm) than *C. reinhardtii* (note the tall single peak that dominates the *C. reinhardtii* spectrum in this region is not carbohydrate).

**Fig. 1** 1D  $^1\text{H}$  spectra of  $^{13}\text{C}$ -enriched *Chlamydomonas reinhardtii*, *Chlorella vulgaris* and *Synechocystis* cell suspensions in  $\text{D}_2\text{O}$  with select chemical shifts labelled (*Asterisk*: in *Chlamydomonas reinhardtii* where lipids dominate, this region contains strong contributions from lipid functionalities)



However, due to considerable overlap in the carbohydrate and amino acid regions, it was difficult to assign specific metabolites from the  $^1\text{H}$  NMR data alone. The aromatic region was expanded to show the presence of aromatic metabolites (Fig. 2). Due to less overlap and characteristic patterns from common metabolites, some specific assignments can be performed [28]. In particular, signals corresponding to the aromatic amino acids tyrosine and phenylalanine were observed along with aromatic nucleotide bases (6.0 ppm). The low signal-to-noise ( $S/N$ ) indicates these species are present at relatively low concentrations.

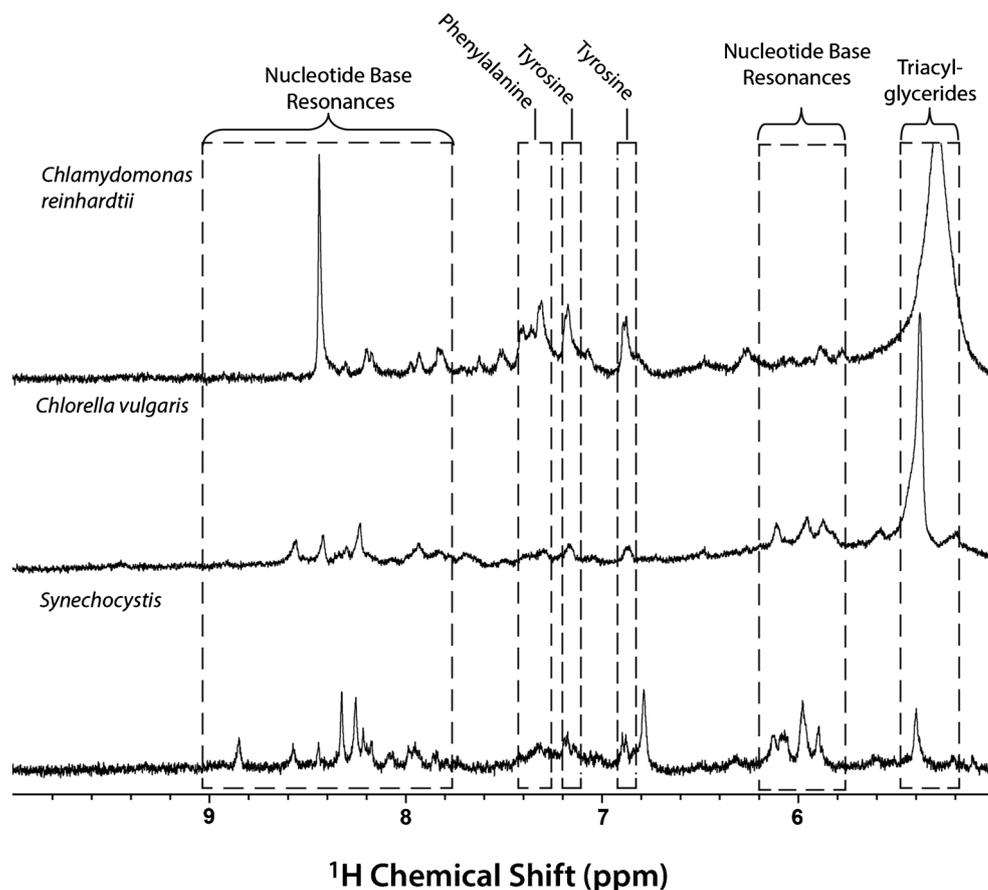
The 1D solution-state  $^{13}\text{C}$  NMR spectra of whole cell  $^{13}\text{C}$ -enriched green algae and cyanobacteria suspended in  $\text{D}_2\text{O}$  are shown in Fig. 3. The spectra show significant overlap; however, chemical shift regions corresponding to aliphatics (5–50 ppm),  $\alpha$ -protons in amino acids (50–60 ppm) and carbohydrates (60–80 ppm), anomeric carbons of carbohydrates (90–105 ppm), double bonds from triacylglycerides (TAG) (125–135 ppm) and carbonyl groups for proteins and lipids (170–190 ppm) were identified [28]. The characteristic double bond resonance ( $-\text{C}=\text{C}-$ , 125–135 ppm) indicates that the *C. reinhardtii* has a much higher TAG content than the *C. vulgaris* and *Synechocystis*. Carbohydrates are easily discerned in the carbon spectrum, which dominate in *C. vulgaris* with significantly less in *Synechocystis* and *C. reinhardtii*. Individual metabolites could not be identified

using the 1D carbon spectra alone due to spectral overlap and require additional information from 2D NMR.

## 2D HMQC ( $^1\text{H}$ - $^{13}\text{C}$ ) NMR

2D or 2D NMR spectroscopy is ideal for the identification of individual metabolites in complex heterogeneous samples as it increases spectral dispersion and provides coupling information allowing individual metabolites to be identified. HMQC was selected in this study due to its superior ability to suppress water in aqueous samples. Both HSQC and HMQC perform well, and their applications to environmental samples are discussed in the details in practical guides and reviews by Farooq et al. [29] and Simpson et al. [30]. Readers should note that although absolute quantification can be achieved in 2D NMR spectroscopy [31], only relative quantification was the goal here. Many metabolites including a range of amino acids and amino acid derivatives were confirmed by the HMQC NMR spectra (Fig. 4). In particular, the aromatic amino acids phenylalanine and tyrosine, which showed weak signals in  $^1\text{H}$  NMR, were clearly identifiable. A wide range of carbohydrates both monosaccharides and disaccharides were observable, although there still was considerable overlap in the carbohydrate region (Fig. 5). The anomeric carbon region provided further confirmation of carbohydrate metabolites. Furthermore, some amines, common organic acids and

**Fig. 2** 1D  $^1\text{H}$  spectra of the aromatic region (5 to 10 ppm) of  $^{13}\text{C}$ -enriched *Chlamydomonas reinhardtii*, *Chlorella vulgaris* and *Synechocystis* cell suspensions in  $\text{D}_2\text{O}$  with important metabolites labelled



nucleotide bases were also identified (Fig. 6). Metabolite assignments were cross-referenced with  $^1\text{H}$ - $^1\text{H}$  TOCSY data where possible (ESM Fig. S2) and confirmed with literature (ESM Table S1).

#### Detailed analysis of the metabolomic profile of *C. reinhardtii*, *C. vulgaris* and *Synechocystis*

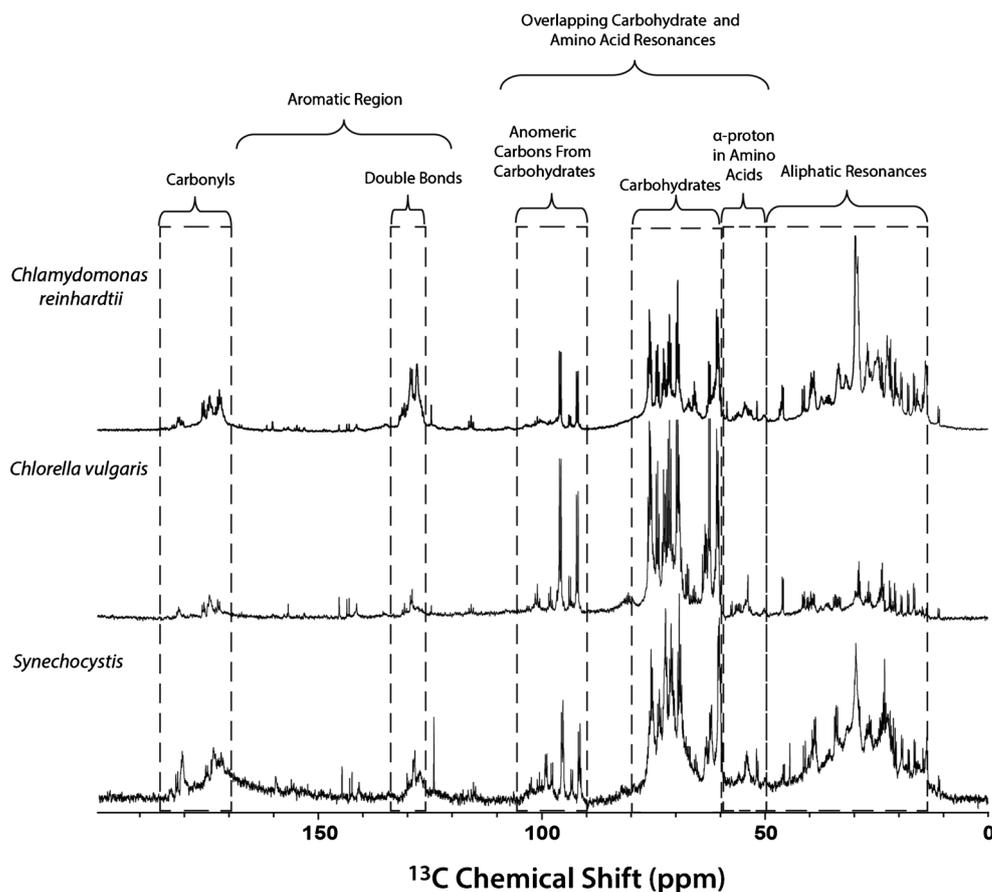
A total of 43 metabolites were identified in *C. reinhardtii*, 49 metabolites were identified in *C. vulgaris* (ESM Figs. S3–S5) and 48 metabolites were identified in *Synechocystis* (ESM Figs. S6–S8) whole cell suspensions in  $\text{D}_2\text{O}$  from HMQC spectra. The assignments correlated to expected chemical shifts with an  $r^2 \geq 0.99$ . However, due to spectral overlap, the assignment of any particular metabolite or lack thereof must be taken in light that a great range of organic residues are present in algae leading to a wide range of overlapping peaks. As such, assignments here will certainly bias the more intense and spectrally resolved metabolites that can be assigned with the most confidence. The NMR data are highly complex and likely contain a wealth of algae and species-specific information that unfortunately cannot be fully assigned. Nonetheless, it provides a single-technique approach for metabolite profiling, and was able to identify a wide arrange of metabolites, that otherwise may have required

multiple different analytical techniques for complete detection (see ESM, Table S1 and references therein). In the following section, a detailed overview of metabolites in each species and their potential impact in relation to the ecosystem, DOM and biofuel synthesis is presented.

#### Detailed analysis of lipids

Green algae and cyanobacteria have distinct metabolite profiles which reflect their nutritional value within the food web. In particular, cyanobacteria are known to be deficient in lipids and TAG [32]. This is evident from  $^1\text{H}$  NMR spectra (Fig. 1) which show more intense signals for aliphatic metabolites ( $-\text{CH}_2$  and  $-\text{CH}_3$  moieties) in *C. reinhardtii* and *C. vulgaris* and weaker signal in *Synechocystis*. The 2D HMQC NMR spectra also confirm the difference in lipid content as the peak corresponding to lipids/TAG shows a decrease in signal strength in *Synechocystis* (ESM Fig. S8). Readers should note that although there is considerable overlap in the lipid peaks, the presence of TAG which makes a significant contribution to the total lipid content in algae was specifically identified based on the presence of two cross-peaks at  $\sim 5.2$  and  $\sim 69.0$  ppm and  $\sim 5.3$  and  $\sim 132.0$  ppm, arising from the glycerol backbone methine group and the double bonds in the long chain unsaturated fatty acids, respectively, both characteristic of TAG

**Fig. 3** 1D  $^{13}\text{C}$  spectra of  $^{13}\text{C}$ -enriched *Chlamydomonas reinhardtii*, *Chlorella vulgaris* and *Synechocystis* cell suspensions in  $\text{D}_2\text{O}$  with select chemical shifts labelled



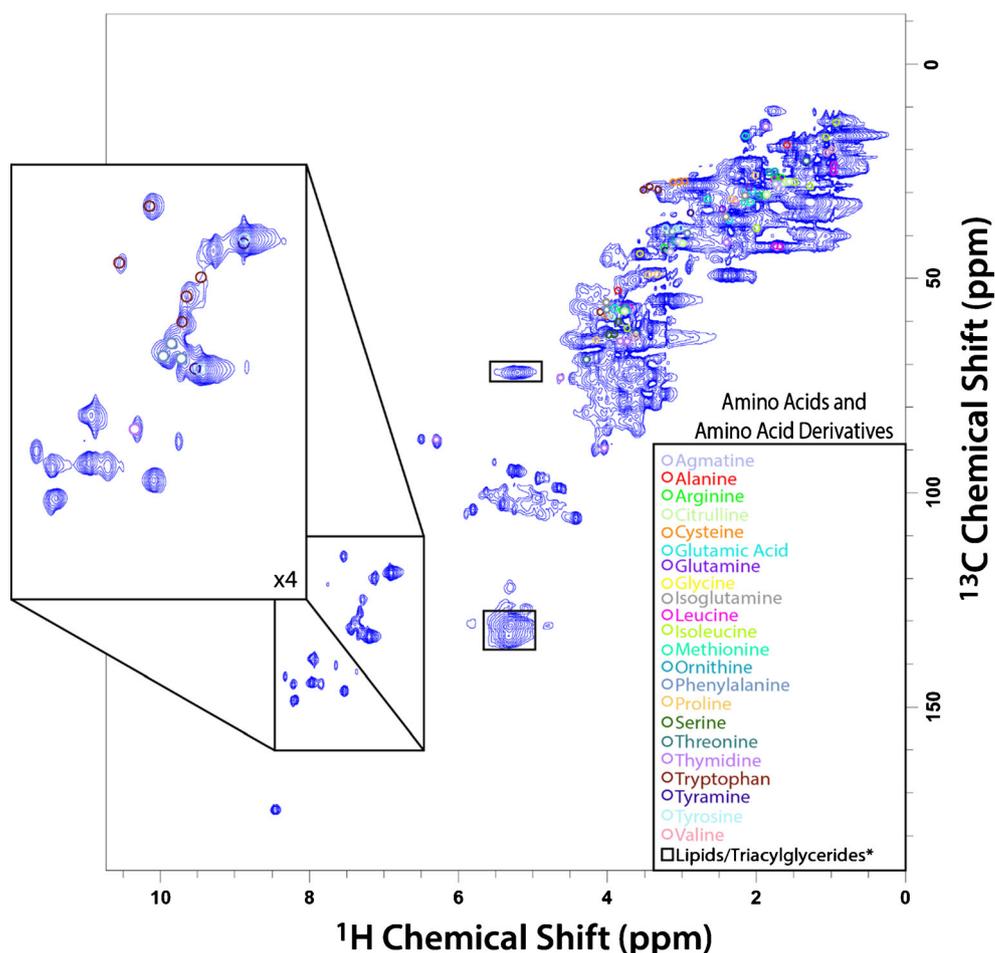
(see ESM Fig. S9 for NMR simulations of a 2D  $^1\text{H}$ - $^{13}\text{C}$  correlation spectrum of an example TAG molecule) [28]. Integration of HMQC NMR spectra shows that the amount of lipids in the swollen fraction of *C. reinhardtii* is greater than in *C. vulgaris* and *Synechocystis*. Readers should note that all quantifications of metabolites by integration of HMQC were scaled relative to spectral noise (rather than total spectral intensity) such that the different swellability of each algae is also captured in the quantification to provide a more direct and absolute comparison.

The concentration of lipids in *C. vulgaris* and *Synechocystis* was 92 and 97 % lower, respectively, relative to *C. reinhardtii* (ESM Figs. S10 and S11) based on  $^1\text{H}$  NMR data. Interestingly, whilst lipids (mainly TAG) are not detected in abundance in *Synechocystis* by  $^1\text{H}$  NMR detection, there is evidence for lipids in the sample from  $^{13}\text{C}$  NMR data (Fig. 3). Cellular components with very high rigidity can be considered to be in the solid phase, where strong  $^1\text{H}$ - $^1\text{H}$  homonuclear dipole couplings dominate leading to spectral broadening on the order of several kilohertz, which makes the observation of these signals via  $^1\text{H}$  detected solution-state NMR approaches extremely difficult. Meanwhile, the broadening effect is less for carbon as such the appearance of lipids in the  $^{13}\text{C}$  spectrum but not in the  $^1\text{H}$  spectrum suggests that they are more rigid, less dispersed in the water phase and as such less bioavailable

to filter feeders. It has been reported that *C. reinhardtii* generally contain a higher amount of TAG (~40–50 % dry weight) [33] compared to both *C. vulgaris* (~3 % dry weight) [34] and *Synechocystis* (negligible) [10]. Comparison of the TAG signals in both the 1D  $^1\text{H}$  spectra (Fig. 2) and the 2D HMQC (Fig. 4 and ESM Figs. S3 and S6) reveals that the detectable TAG content of *C. reinhardtii* is significantly higher than those of *C. vulgaris* and *Synechocystis*, which follows the literature trend.

A lack of ‘available’ lipids in the cyanobacterium *Synechocystis* supports previous work [35]. This is of interest to ecologists as algae and cyanobacteria form the primary food source for micro-crustaceans. Therefore, when cyanobacteria outcompete green algae in freshwater ecosystems, there is a decrease in the flux of essential lipids through the food web and a collapse of micro-crustacean populations. For example, cyanobacteria have been shown to have a direct species-specific toxic effect on zooplankton with a general decrease in survivorship of zooplankton [36]. Furthermore, there is a decrease in the number of eggs produced by *D. magna* when they are exposed to cyanobacteria due to its deficiency in polyunsaturated fatty acids [37]. Additionally, it has been shown that the lipid composition of silver carp is dependent on the fatty acid composition of zooplankton which forms an essential part of their diet [38]. Therefore, the impact of poor

**Fig. 4** 2D HMQC ( $^1\text{H}$ - $^{13}\text{C}$ ) spectra of a  $^{13}\text{C}$ -enriched *Chlamydomonas reinhardtii* cell suspension in  $\text{D}_2\text{O}$  with labelled amino acid and amino acid derivative metabolites. Asterisk: the highlighted regions for lipids/triacylglycerides represent the methine resonances of the glycerol backbone (top) and the double bond resonances in long chain unsaturated fatty acids (bottom)



nutrition (cyanobacteria) has potential to reach many trophic levels [38]. This study suggests there is a difference in availability of the lipids to the aqueous phase and implies not just the presence of specific lipids but also their physical form could play a role in their nutritional availability and warrants further study.

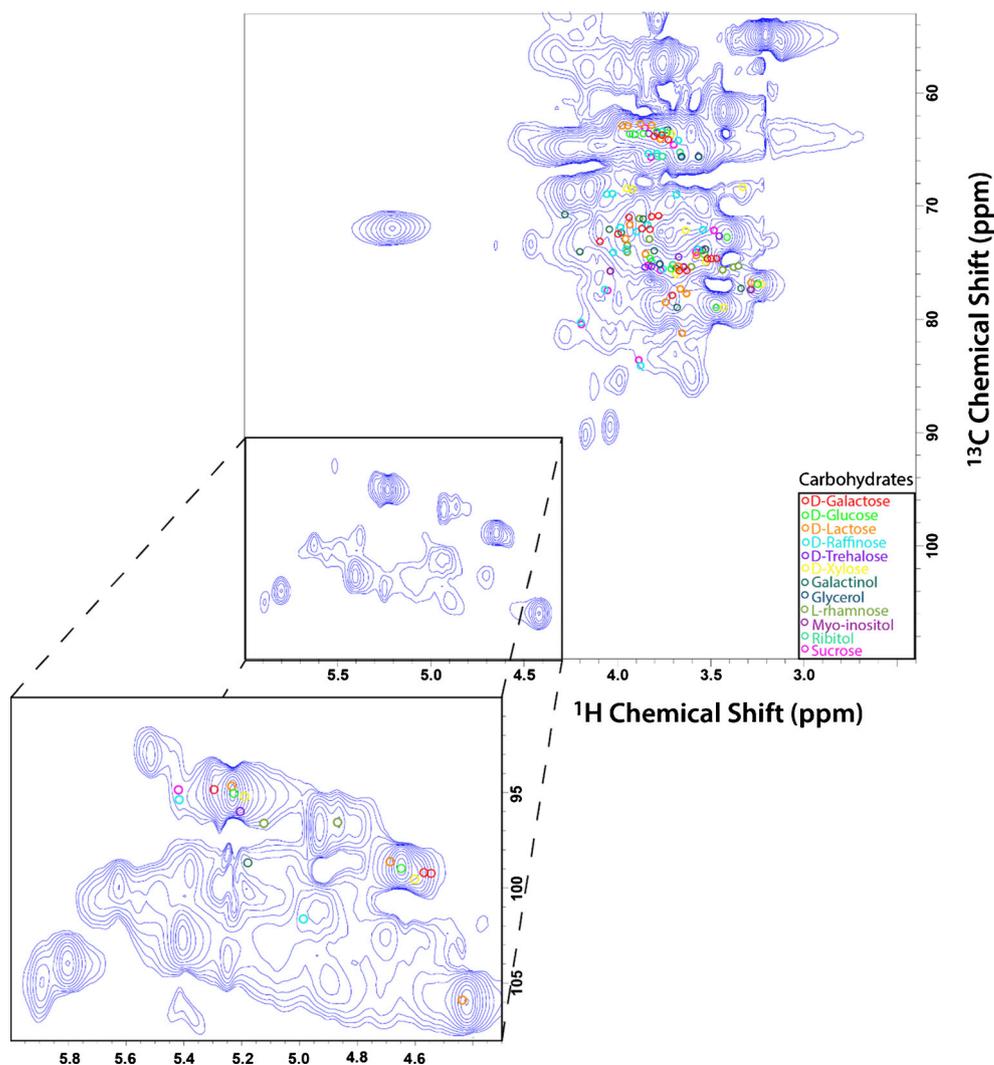
Furthermore, HMQC NMR data indicate that sterols and terpenoids are in low concentration in the algae species studied here, at least relative to the other components identified. This is interesting as sterols and terpenoids are known to be major components of DOM and resonate in a relatively unique spectral region making them difficult to miss by NMR [6, 39]. The low concentration of sterols and terpenoids suggests that algae and cyanobacteria may not be major direct contributors to structures abundant in classic DOM. However, it is not possible to completely rule out algae as a major input, given that the vast majority of species identified here, protein, lipids and small metabolites, are all excellent nutritional sources that will be quickly utilised by aquatic biota. In turn, many terpenoids are poor food sources that may accumulate; therefore, even small but continual inputs over time from algae could potentially have a significant impact on DOM composition.

Beyond the focus of this work, solution-state NMR may be an efficient tool in determining viable algae species for biofuel production and future research could easily focus on developing a database of algal species with a quantification of the amount of ‘aqueous available’ lipids in each species using standard NMR hardware. This is particularly important since super-heated water is one of the most economical and promising approaches for biofuel extraction from algae [40].

#### *Non-lipid metabolites*

Although there were obvious differences in the lipid content, there were many similarities in the carbohydrate and amino acid contents of each species. Using solution-state NMR, many key metabolites that carry out important biological functions in cells were identified. Whilst spectral overlap makes quantification of many metabolites challenging, it was possible to monitor the relative change in concentration of some of the more resolved metabolites between the species (ESM Figs. S10 and S11). Specifically, the aromatic amino acids tyrosine, tryptophan and phenylalanine, which are important for biological functioning in plant cells, were identified. The

**Fig. 5** 2D HMQC ( $^1\text{H}$ - $^{13}\text{C}$ ) spectrum of a  $^{13}\text{C}$ -enriched *Chlamydomonas reinhardtii* cell suspension in  $\text{D}_2\text{O}$  with labelled carbohydrate metabolites



low signals for these metabolites are indicative of lower concentrations. The percent difference of their concentrations in *C. vulgaris* and *Synechocystis* relative to *C. reinhardtii* was determined by integrating peak area (ESM Figs. S10 and S11) allowing for future real-time monitoring of these metabolites in complex natural samples. For example, monitoring tyrosine is important for the formation of radicals that are essential for the photo-synthetic oxygen-evolving system, a key component of photosynthesis [41]. Furthermore, tryptophan is a precursor to the plant hormone indoleacetic acid and is essential for maintaining physiological processes in plant cells [42]. Finally, phenylalanine is a precursor for the formation of phenylpropanoids which are essential for synthesis of lignin-like structural components such which is known to be present in some algae [43, 44].

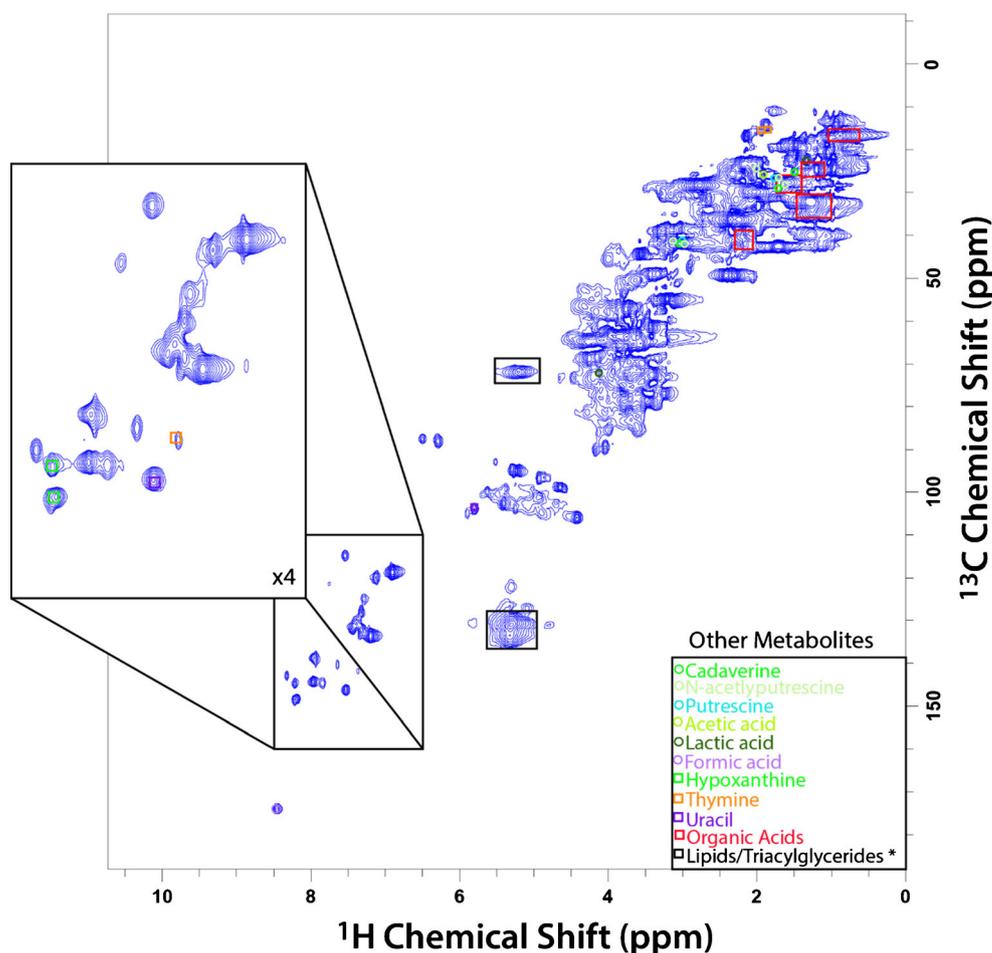
In addition, strong signals for proline, an amino acid that accumulates in plant cells during environmental stress, were observed in the HMQC NMR spectra. Therefore, this approach may be ideal for studying the change in the proline

signal, which is a biomarker for stress, in vivo [45, 46]. By integrating the total peak area, it was determined that the greatest amount of proline was found in *C. vulgaris*, 50 % more relative to *C. reinhardtii*, whereas the amount of proline in *Synechocystis* was lesser by ~70 % compared to *C. reinhardtii*. Additionally, several nucleotides including thymine and thymidine, which are involved in growth and development in plant cells demonstrating the potential application of this approach to metabolomics, were also identified [47].

#### Spectral editing 1D spectra of $^{13}\text{C}$ -enriched *C. reinhardtii*

Algae and cyanobacteria cells are representative of complex natural samples; as such, a number of spectral editing techniques were employed to better study the metabolomic and structural profile due to spectral overlap. These include the acquisition of 1D DE (species with restricted diffusion, for example macromolecules, sorbed species), IDE (dissolved species) and RADE (most rigid molecules that can be detected

**Fig. 6** 2D HMQC ( $^1\text{H}$ - $^{13}\text{C}$ ) spectrum of a  $^{13}\text{C}$ -enriched *Chlamydomonas reinhardtii* cell suspension in  $\text{D}_2\text{O}$  with labelled amine, carboxylic acid, nucleotide base and acetate metabolites. Asterisk: the highlighted regions for lipids/triacylglycerides represent the methine resonances of the glycerol backbone (top) and the double bond resonances in long chain unsaturated fatty acids (bottom)



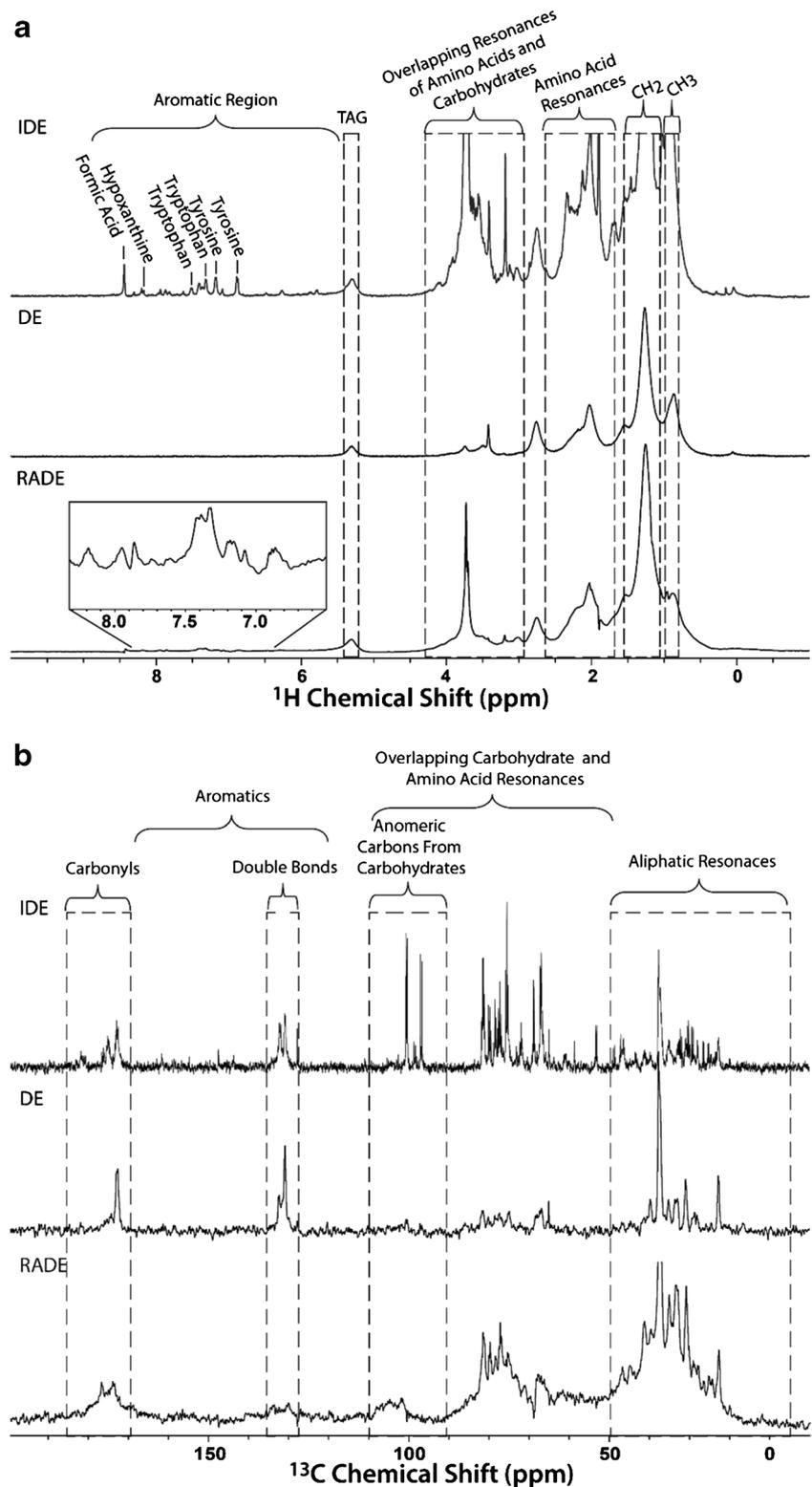
using the solution-state approaches employed here) spectra (Fig. 7a) [25]. The DE experiment involves the coding of the spatial position of signals at the start of the experiment and then decoding the position at the end using gradient pulses. The signal from molecules that diffuse, i.e. change their physical position, is not observed in the spectra resulting in a spectrum of large components or metabolites that show little to no diffusion. The  $^1\text{H}$  DE spectra of *C. reinhardtii* show a loss of signal in the carbohydrate region compared to the whole, unedited spectrum, whereas signals from lipids ( $-\text{CH}_2$  and  $-\text{CH}_3$  moieties) are retained (Fig. 7a) indicating that these later components have restricted diffusion consistent with lipid micelle/vesicles. The IDE spectra emphasise molecules with unrestricted diffusion, i.e. those truly in solution, and are created by the subtraction of the DE spectrum (contains larger components with restricted diffusion) from a reference spectrum (contains a profile of all components) [25]. The IDE spectrum of *C. reinhardtii* showed a relative increase in signal in the carbohydrate region compared to the whole spectrum and a decrease in signal in the lipid ( $-\text{CH}_2$  and  $-\text{CH}_3$  moieties) and amino acid regions suggesting that carbohydrates are dissolved in the cytoplasm and can move freely. Furthermore, a number of signals for mobile metabolites in solution such as

tyrosine, tryptophan, formic acid and thymidine were also observed.

Finally, a RADE spectrum accounts for large fast relaxing components that may be unobservable using traditional diffusion editing techniques with long diffusion delays [25]. The RADE spectrum therefore contains the components with very fast relaxation, for example more rigid structures such as cell membranes, structural proteins, etc. The RADE spectrum (Fig. 7a), similar to the DE spectrum, is dominated by lipids which are consistent with a range of membrane lipids and vesicles that would be present in algae cells. In contrast to the DE spectrum, the aromatic region of the RADE also shows signals from aromatic amino acids in proteins, which are abundant in algal cells.  $^1\text{H}$  spectral editing was also performed for *C. vulgaris* (ESM Fig. S12), and similar trends were observed with metabolites in the IDE spectrum, lipids in the DE spectrum and lipids and protein in the RADE spectrum. Interestingly, for *Synechocystis* (ESM Fig. S13), lipids were not seen in abundance in any of the  $^1\text{H}$  detected spectral editing approaches confirming that the lipids in *Synechocystis* are too rigid to be observed by  $^1\text{H}$  NMR.

Similar spectral editing approaches can also be applied to  $^{13}\text{C}$  NMR spectra. Specifically, IDE, DE and RADE  $^{13}\text{C}$

**Fig. 7** Comparison of 1D IDE, DE and RADE. **A**  $^1\text{H}$  spectra and **B**  $^{13}\text{C}$  spectra of a  $^{13}\text{C}$ -enriched *Chlamydomonas reinhardtii* and cell suspension in  $\text{D}_2\text{O}$  with select chemical shifts labelled



spectra for *C. reinhardtii* are displayed in Fig. 7b with regions of interest highlighted. In the IDE spectrum of *C. reinhardtii*, many sharp signals are visible in the amino acid and carbohydrate regions as these metabolites show unrestricted diffusion.

In addition, signals from the anomeric carbons in carbohydrates are clearly visible in the IDE spectrum. The carbon DE spectrum shows a decrease in the signal for carbohydrates which are mobile, whilst signals from lipids, which are more

restricted in their movement, are retained. Furthermore, the RADE carbon spectrum is dominated by lipids, particularly the fraction of lipids in cells that have a more rigid character. As well, the presence of aromatic amino acid signals in the RADE spectra is indicative of large protein molecules which exhibit restricted diffusion. The carbon IDE, DE and RADE spectra provide complimentary spectral information to the proton spectra. These diffusion edited NMR techniques are useful as they separate signals of the free dissolved metabolites from structural components (lipids and proteins), decrease overlap and provide some information as to the physical form of the components within the cells. Editing based on diffusivity has been previously applied to study complex mixtures such as blood plasma, tissues and urine [48–52]. However, the RADE experiment, which recovers fast relaxing components that otherwise may be overlooked by diffusion based editing, is relatively new [25]. When combined, inverse diffusion (metabolites), diffusion editing (restricted diffusion) and RADE (semi-solids) provide a convenient and discrete approach for the separation of major components in mixtures, including, as demonstrated here, application to whole cell suspensions.

## Conclusion

This study demonstrates that solution-state NMR can be used to study intact algae cells allowing the observation of major differences between species and identify key metabolites. Specifically, the difference in the lipid profile between green algae and cyanobacteria was confirmed, shedding light on the current understanding of the impact of cyanobacteria blooms on freshwater ecosystems. More importantly, a complete list of identifiable metabolites in each species was compiled, confirming and complimenting those previously reported. This work is important for future NMR-based metabolomic studies as algae and cyanobacteria are often the primary source for carbon labelling. For example, higher organisms such as mice can be  $^{13}\text{C}$ -labelled by feeding them fish that have eaten daphnia which have been grown on  $^{13}\text{C}$ -labelled algae [53]. In order to understand the metabolomic changes occurring in higher organisms, it is essential to understand the metabolite profile and flux from the primary source—algae and cyanobacteria—to higher organisms.

However, providing absolute quantification was challenging due to overlap in 1D spectra forcing identification and relative quantification from 2D NMR for specific metabolites. Whilst 2D is generally considered less quantitative than 1D NMR, improvements in the field and the extension of an electronic referencing method, Electronic REference To access In vivo Concentrations (ERETIC), have been shown to work for 2D NMR and once integrated into commercially available software will make absolute quantification for 2D NMR

possible in the future [54]. This will allow the monitoring of the metabolites using solution-state NMR to understand biological processes in vivo. Studies that follow the consumption of labelled algae through the food chain will also be possible and could be extremely informative as the fate and conversion of the label at the molecular level may be monitored, explaining how and why organisms depend on each other across trophic levels.

The study also suggests that algae and cyanobacteria are likely not major direct contributors to DOM composition as they contain low concentrations of sterols and terpenoids in the mobile state but still could be important sources after environmental fractionation and preservation. Conversely, algal inputs may be highly significant to protein-rich DOM that has been recently discovered in high-elevation lakes [55].

The ability to provide relative quantification as to the amount of lipids in each species may prove to be valuable to researchers trying to develop genetically modified algae strains with larger lipid fractions to increase the yield of biofuel per unit mass algae. Of the three strains of algae studied here, it is clear that *C. reinhardtii* has by far the highest proportion of aqueous available lipids, suggesting it as a potential strain for biofuel production. Even though valuable information can be gained using solution-state NMR, assignment of metabolites is challenging due to significant spectral overlap. Future studies using 3D NMR should provide the additional spectral dispersion required for further spectral assignments [56], presently not feasible due to spectral overlap.

In addition, solution-state NMR provides the ability to selectively study dissolved/mobile species. This can be an advantage where the goal of the study is to identify the aquatically available fraction as is the goal here but could be a drawback if all components in the cells (including true solids) need to be detected. The latter should be possible by the use of a novel technique called comprehensive multiphase NMR (CMP-NMR) introduced in 2012, which allows researchers to access molecular information from all three phases simultaneously [25]. This study demonstrated that lipids in *Synechocystis* were more ‘solid-like’ and less available to the aqueous phase than in the other strains studied. The CMP-NMR technique could be highly complementary providing the additional information on the semi-solid and solid phases needed to explain their availability.

In summary, this study provides a structural and metabolic overview as to the mobile/dissolved components in the natural state which are most likely to directly transfer into the aqueous phase on lysis. The study represents a key first step, permitting future studies of carbon transfer between species and through the food chain, as well as providing a foundation to better understand the role of algae in the formation of DOM and sequestration/transformation of carbon in the aquatic environment in general.

**Acknowledgments** A. J. S. would like to thank Mark Krembil and the Krembil Foundation for both inspiring and funding this research. A. J. S. would like to thank the Natural Sciences and Engineering Research Council of Canada (NSERC) for financially supporting this research.

#### Compliance with ethical standards

**Conflict of Interest** The authors declare no conflicts of interest.

## References

- Paerl HW, Xu H, McCarthy MJ, Zhu G, Qin B, Li Y, et al. Controlling harmful cyanobacterial blooms in a hyper-eutrophic lake (Lake Taihu, China): the need for a dual nutrient (N & P) management strategy. *Water Res.* 2011;45:1973–83.
- Piorreck M, Baasch K-H, Pohl P. Biomass production, total protein, chlorophylls, lipids and fatty acids of freshwater green and blue-green algae under different nitrogen regimes. *Phytochemistry.* 1984;23:207–16.
- Martin-Creuzburg D, Wacker A, von Elert E. Life history consequences of sterol availability in the aquatic keystone species *Daphnia*. *Oecologia.* 2005;144:362–72.
- Bednarska A, Slusarczyk M. Effect of non-toxic, filamentous cyanobacteria on egg abortion in *Daphnia* under various thermal conditions. *Hydrobiologia.* 2013;715:151–7.
- Perhar G, Arhonditsis GB, Brett MT. Modelling the role of highly unsaturated fatty acids in planktonic food web processes: a mechanistic approach. *Environ Rev.* 2012;20:155–72.
- Lam B, Baer A, Alaei M, Lefebvre B, Moser A, Williams A, et al. Major structural components in freshwater dissolved organic matter. *Environ Sci Technol.* 2007;41:8240–7.
- Demirbas A, Fatih DM. Importance of algae oil as a source of biodiesel. *Energy Convers Manag.* 2011;52:163–70.
- Griffiths MJ, Harrison STL. Lipid productivity as a key characteristic for choosing algal species for biodiesel production. *J Appl Phycol.* 2009;21:493–507.
- Scott SA, Davey MP, Dennis JS, Horst I, Howe CJ, Lea-Smith DJ, et al. Biodiesel from algae: challenges and prospects. *Curr Opin Biotechnol.* 2010;21:277–86.
- Wahlen BD, Willis RM, Seefeldt LC. Biodiesel production by simultaneous extraction and conversion of total lipids from microalgae, cyanobacteria, and wild mixed-cultures. *Bioresour Technol.* 2011;102:2724–30.
- Radakovits R, Jinkerson RE, Darzins A, Posewitz MC. Genetic engineering of algae for enhanced biofuel production. *Eukaryot Cell.* 2010;9:486–501.
- Bolling C, Fiehn O. Metabolite profiling of *Chlamydomonas reinhardtii* under nutrient deprivation. *Plant Physiol.* 2005;139:1995–2005.
- Madsen AD, Goessler W, Pedersen SN, Francesconi KA. Characterization of an algal extract by HPLC-ICP-MS and LC-electrospray MS for use in arsenosugar speciation studies. *J Anal At Spectrom.* 2000;15:657–62.
- Halket JM, Waterman D, Przyborowska AM, Patel RKP, Fraser PD, Bramley PM. Chemical derivatization and mass spectral libraries in metabolic profiling by GC/MS and LC/MS/MS. *J Exp Bot.* 2004;56:219–43.
- Aguilar JA, Cassani J, Delbianco M, Adams RW, Nilsson M, Morris GA. Minimising research bottlenecks by decluttering NMR spectra. *Chem Eur J.* 2015;21:6623–30.
- Gupta V, Thakur RS, Reddy CRK, Jha B. Central metabolic processes of marine macrophytic algae revealed from NMR based metabolome analysis. *RSC Adv.* 2013;3:7037–47.
- Pandit A, Morosinotto T, Reus M, Holzwarth AR, Bassi R, de Groot HJM. First solid-state NMR analysis of uniformly  $^{13}\text{C}$ -enriched major light-harvesting complexes from *Chlamydomonas reinhardtii* and identification of protein and cofactor spin clusters. *Biochim Biophys Acta Bioenerg.* 1807;2011:437–43.
- Arnold AA, Genard B, Zito F, Tremblay R, Warschawski DE, Marcotte I. Identification of lipid and saccharide constituents of whole microalgal cells by  $^{13}\text{C}$  solid-state NMR. *Biochim Biophys Acta.* 1848;2015:369–77.
- Chauton MS, Optun OI, Bathen TF, Volent Z, Gribbestad IS. HR MAS  $^1\text{H}$  NMR spectroscopy analysis of marine microalgal whole cells. *Mar Ecol Prog Ser.* 2003;256:57–62.
- Chauton MS, Røvik Størseth T, Johnsen G. High-resolution magic angle spinning  $^1\text{H}$  NMR analysis of whole cells of *Thalassiosira pseudonana* (Bacillariophyceae): broad range analysis of metabolic composition and nutritional value. *J Appl Phycol.* 2003;15:533–42.
- Chauton MS, Størseth TR, Krane J. High-resolution magic angle spinning NMR analysis of whole cells of *Chaetoceros muelleri* (Bacillariophyceae) and comparison with  $^{13}\text{C}$ -NMR and distortionless enhancement by polarization transfer  $^{13}\text{C}$ -NMR analysis of lipophilic extracts. *J Phycol.* 2004;40:611–8.
- Simpson AJ, Simpson MJ, Soong R. Nuclear magnetic resonance spectroscopy and its key role in environmental research. *Environ Sci Technol.* 2012;46:11488–96.
- Lam B, Simpson AJ. Direct  $^1\text{H}$  NMR spectroscopy of dissolved organic matter in natural waters. *Analyst.* 2008;133:263–9.
- Wu DH, Chen AD, Johnson CS. An improved diffusion-ordered spectroscopy experiment incorporating bipolar-gradient pulses. *J Magn Reson Ser A.* 1995;115:260–4.
- Courtier-Murias D, Farooq H, Masoom H, Botana A, Soong R, Longstaffe JG, et al. Comprehensive multiphase NMR spectroscopy: basic experimental approaches to differentiate phases in heterogeneous samples. *J Magn Reson.* 2012;217:61–76.
- Peti W, Griesinger C, Bermel W. Adiabatic TOCSY for C,C and H, H J-transfer. *J Biomol NMR.* 2000;18:199–205.
- Woods GC, Simpson MJ, Koerner PJ, Napoli A, Simpson AJ. HILIC-NMR: toward the identification of individual molecular components in dissolved organic matter. *Environ Sci Technol.* 2011;45:3880–6.
- Lam L, Soong R, Sutrisno A, de Visser R, Simpson MJ, Wheeler HL, et al. Comprehensive multiphase NMR spectroscopy of intact  $^{13}\text{C}$ -labeled seeds. *J Agric Food Chem.* 2014;62:107–15.
- Farooq H, Courtier-Murias D, Soong R, Bermel W, Kingery WM, Simpson AJ. HR-MAS NMR spectroscopy: a practical guide for natural samples. *Curr Org Chem.* 2013;17:3013–31.
- Simpson AJ, McNally DJ, Simpson MJ. NMR spectroscopy in environmental research: from molecular interactions to global processes. *Prog Nucl Magn Reson Spectrosc.* 2011;58:97–175.
- Giraudeau P. Quantitative 2D liquid-state NMR. *Magn Reson Chem.* 2014;52:259–72.
- Liu X, Sheng J, Curtiss R. Fatty acid production in genetically modified cyanobacteria. *Proc Natl Acad Sci.* 2011;108:6899–904.
- Hu Q, Sommerfeld M, Jarvis E, Ghirardi M, Posewitz M, Seibert M, et al. Microalgal triacylglycerols as feedstocks for biofuel production: perspectives and advances. *Plant J.* 2008;54:621–39.
- Stephenson AL, Dennis JS, Howe CJ, Scott SA, Smith AG. Influence of nitrogen-limitation regime on the production by *Chlorella vulgaris* of lipids for biodiesel feedstocks. *Biofuels.* 2010;1:47–58.
- Cai T, Ge X, Park SY, Li Y. Comparison of *Synechocystis* sp. PCC6803 and *Nannochloropsis salina* for lipid production using artificial seawater and nutrients from anaerobic digestion effluent. *Bioresour Technol.* 2013;144:255–60.

36. DeMott WR, Zhang Q-X, Carmichael WW. Effects of toxic cyanobacteria and purified toxins on the survival and feeding of a copepod and three species of *Daphnia*. *Limnol Oceanogr*. 1991;36:1346–57.
37. Wacker A, Martin-Creuzburg D. Allocation of essential lipids in *Daphnia magna* during exposure to poor food quality. *Funct Ecol*. 2007;21:738–47.
38. Domaizon I, Desvillettes C, Debroas D, Bourdier G. Influence of zooplankton and phytoplankton on the fatty acid composition of digesta and tissue lipids of silver carp: mesocosm experiment. *J Fish Biol*. 2000;57:417–32.
39. Woods GC, Simpson MJ, Simpson AJ. Oxidized sterols as a significant component of dissolved organic matter: evidence from 2D HPLC in combination with 2D and 3D NMR spectroscopy. *Water Res*. 2012;46:3398–408.
40. Savage PE, Levine R, Pinnarat T. Method of producing biodiesel from a wet biomass. US Patent. 2014;8673028:B2.
41. Barry BA, Babcock GT. Tyrosine radicals are involved in the photosynthetic oxygen-evolving system. *Proc Natl Acad Sci*. 1987;84:7099–103.
42. Bartel B. Auxin biosynthesis. *Annu Rev Plant Physiol Plant Mol Biol*. 1997;48:51–66.
43. Delwiche CF, Graham LE, Thomson N. Lignin-like compounds and *Sporopollenin coleochaete*, an algal model for land plant ancestry. *Science*. 1989;245:399–401.
44. Ritter H, Schulz GE. Structural basis for the entrance into the phenylpropanoid metabolism catalyzed by phenylalanine ammonia-lyase. *Plant Cell Online*. 2004;16:3426–36.
45. Siripornadulsil S, Traina S, Verma DPS, Sayre RT. Molecular mechanisms of proline-mediated tolerance to toxic heavy metals in transgenic microalgae. *Plant Cell Online*. 2002;14:2837–47.
46. Wu J, Chang S, Chou T. Intracellular proline accumulation in some algae exposed to copper and cadmium. *Bot Bull Acad Sin*. 1995;36:89–93.
47. Stasolla C, Katahira R, Thorpe TA, Ashihara H. Purine and pyrimidine nucleotide metabolism in higher plants. *J Plant Physiol*. 2003;160:1271–95.
48. Tang H, Wang Y, Nicholson JK, Lindon JC. Use of relaxation-edited one-dimensional and two dimensional nuclear magnetic resonance spectroscopy to improve detection of small metabolites in blood plasma. *Anal Biochem*. 2004;325:260–72.
49. Wang Y, Bollard ME, Keun H, Antti H, Beckonert O, Ebbels TM, et al. Spectral editing and pattern recognition methods applied to high-resolution magic-angle spinning <sup>1</sup>H nuclear magnetic resonance spectroscopy of liver tissues. *Anal Biochem*. 2003;323:26–32.
50. Beckonert O, Keun HC, Ebbels TMD, Bundy J, Holmes E, Lindon JC, et al. Metabolic profiling, metabolomic and metabonomic procedures for NMR spectroscopy of urine, plasma, serum and tissue extracts. *Nat Protoc*. 2007;2:2692–703.
51. Smith LM, Maher AD, Cloarec O, Rantalainen M, Tang H, Elliott P, et al. Statistical correlation and projection methods for improved information recovery from diffusion-edited NMR spectra of biological samples. *Anal Chem*. 2007;79:5682–9.
52. Liu M, Nicholson JK, Lindon JC. High-resolution diffusion and relaxation edited one- and two-dimensional <sup>1</sup>H NMR spectroscopy of biological fluids. *Anal Chem*. 1996;68:3370–6.
53. Krüger M, Heumann H. Isotopic labeling of higher organisms. US Patent. 2014;20140328760:A1.
54. Michel N, Akoka S. The application of the ERETIC method to 2D-NMR. *J Magn Reson*. 2004;168:118–23.
55. Goldberg SJ, Ball GI, Allen BC, Schladow SG, Simpson AJ, Masoom H, et al. Refractory dissolved organic nitrogen accumulation in high-elevation lakes. *Nat Commun*. 2015;6:6347.
56. Simpson AJ, Kingery WL, Hatcher PG. The identification of plant derived structures in humic materials using three-dimensional NMR spectroscopy. *Environ Sci Technol*. 2003;37:337–42.

**Analytical and Bioanalytical Chemistry**

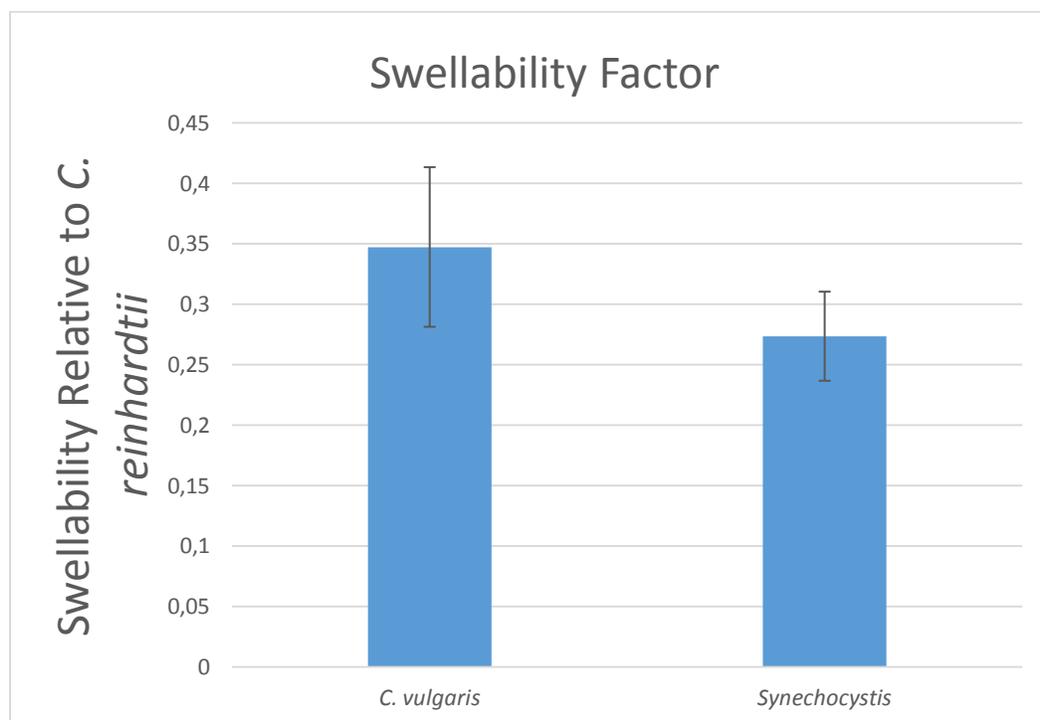
**Electronic Supplementary Material**

**Identification of aquatically available carbon from algae through solution-state NMR of whole  $^{13}\text{C}$ -labelled cells**

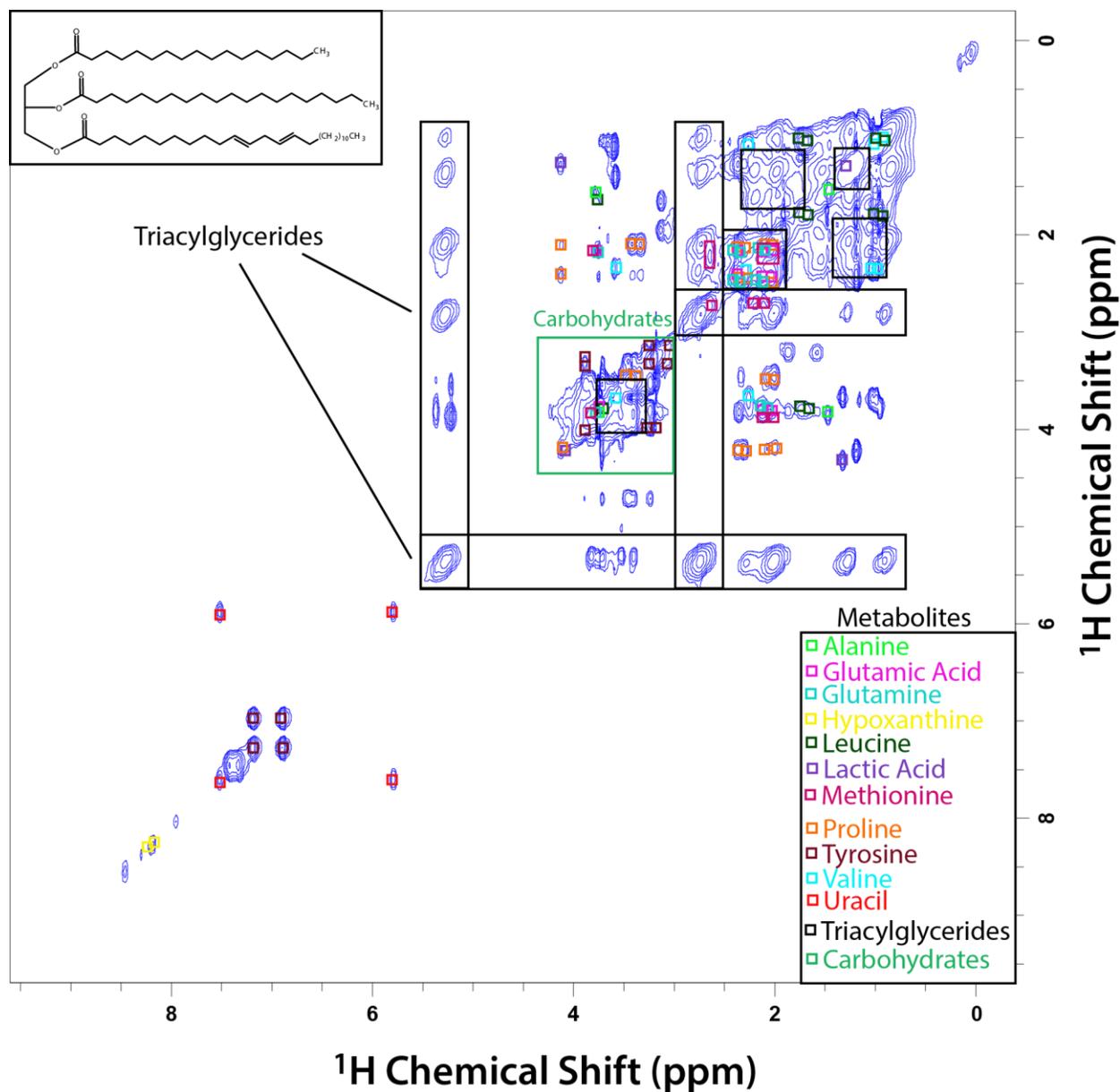
Mohammad Akhter, Rudraksha Dutta Majumdar, Blythe Fortier-McGill, Ronald Soong, Yalda Liaghati Mobarhan, Myrna Simpson, George Arhonditsis, Sebastian Schmidt, Hermann Heumann, André J. Simpson

## **<sup>13</sup>C Labeling Methodology**

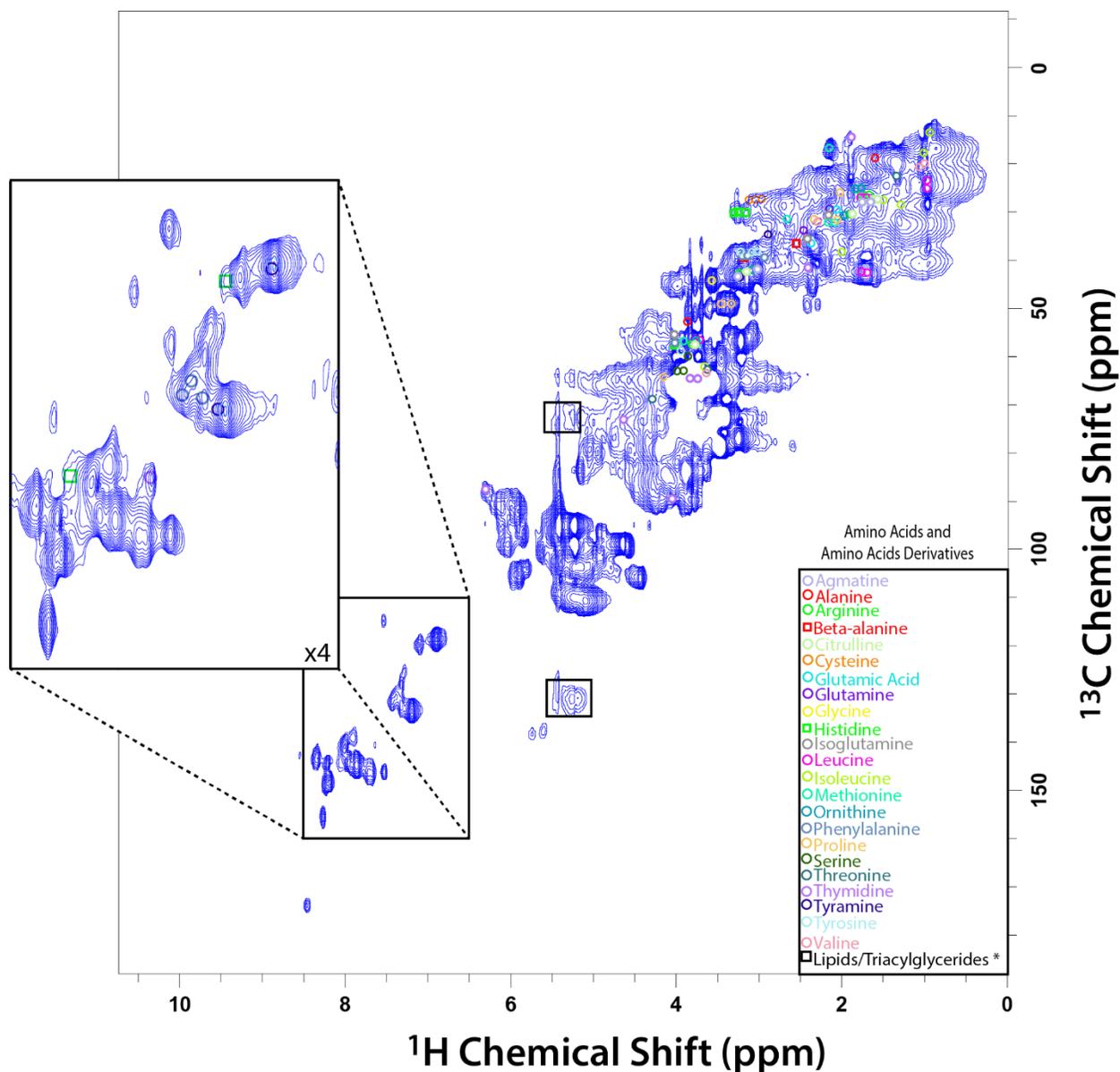
<sup>13</sup>C-labeling of the algae (wt, SAG culture collection Goettingen, Germany) was achieved by growing the cells in <sup>13</sup>CO<sub>2</sub> (enrichment 99%; purchased from Sigma Aldrich Isotec, St Louis, Missouri) for 150 hours in a photobioreactor (PBR) constructed by Silantes GmbH (Munich Germany). The PBR is built as a closed system avoiding loss of the stable isotope labeled <sup>13</sup>CO<sub>2</sub>. It is an airlift driven external loop tubular fermenter having 20L operating volume, permitting circulation of the media volume with 1 vvm. The algae were cultivated in regular TP media (Tris-Phosphate w/o Acetate + <sup>15</sup>NH<sub>4</sub>Cl), as previously described (Gorman and Levine, 1965), using the following parameters: pH 7.0 - 7.4 (pH-probe InPro<sup>®</sup>3253I/Sg, Mettler Toledo), temperature 30°C (InPro<sup>®</sup>3253I), light intensity 1300 μmol/(m<sup>2</sup>·s) (quantum detector LI-250A, Li-Cor Biosciences GmbH). A computer-controlled gas management system (Labview 10.1, National Instruments) was incorporated to keep the <sup>13</sup>CO<sub>2</sub> content (pCO<sub>2</sub>-probe InPro<sup>®</sup>5000, Mettler Toledo) at a concentration of 2%. The nitrogen carrier gas was allowed to vary between 65% and 85% corresponding to the oxygen content which increased during autotrophic growth of algae between 15% and 30%. If the oxygen content reached a value of 30% (pO<sub>2</sub>-probe Visiferm Do Arc., Hamilton) an N<sub>2</sub> purging step was introduced in order to reduce the O<sub>2</sub> content to the initial value of 15%. To avoid loss of <sup>13</sup>CO<sub>2</sub> during N<sub>2</sub> purging, the <sup>13</sup>CO<sub>2</sub> addition was stopped at the end of a growth cycle until a value of 0,5% was reached by metabolic depletion. If the oxygen reached the initial value of 15% due to N<sub>2</sub>-purging, the computer controlled valves (Valve 221606, Buerkert) were closed and another growth cycle was resumed. The <sup>13</sup>C-content of the algae biomass was determined by analyzing the enrichment in “C<sub>18</sub>” fatty acids of isolated algae biomass by GCMS (Thermo Quest Polaris Q MS / Trace GC2000, Thermo Fisher). Isotopic enrichment of 98-99% <sup>13</sup>C was achieved in the algae biomass.



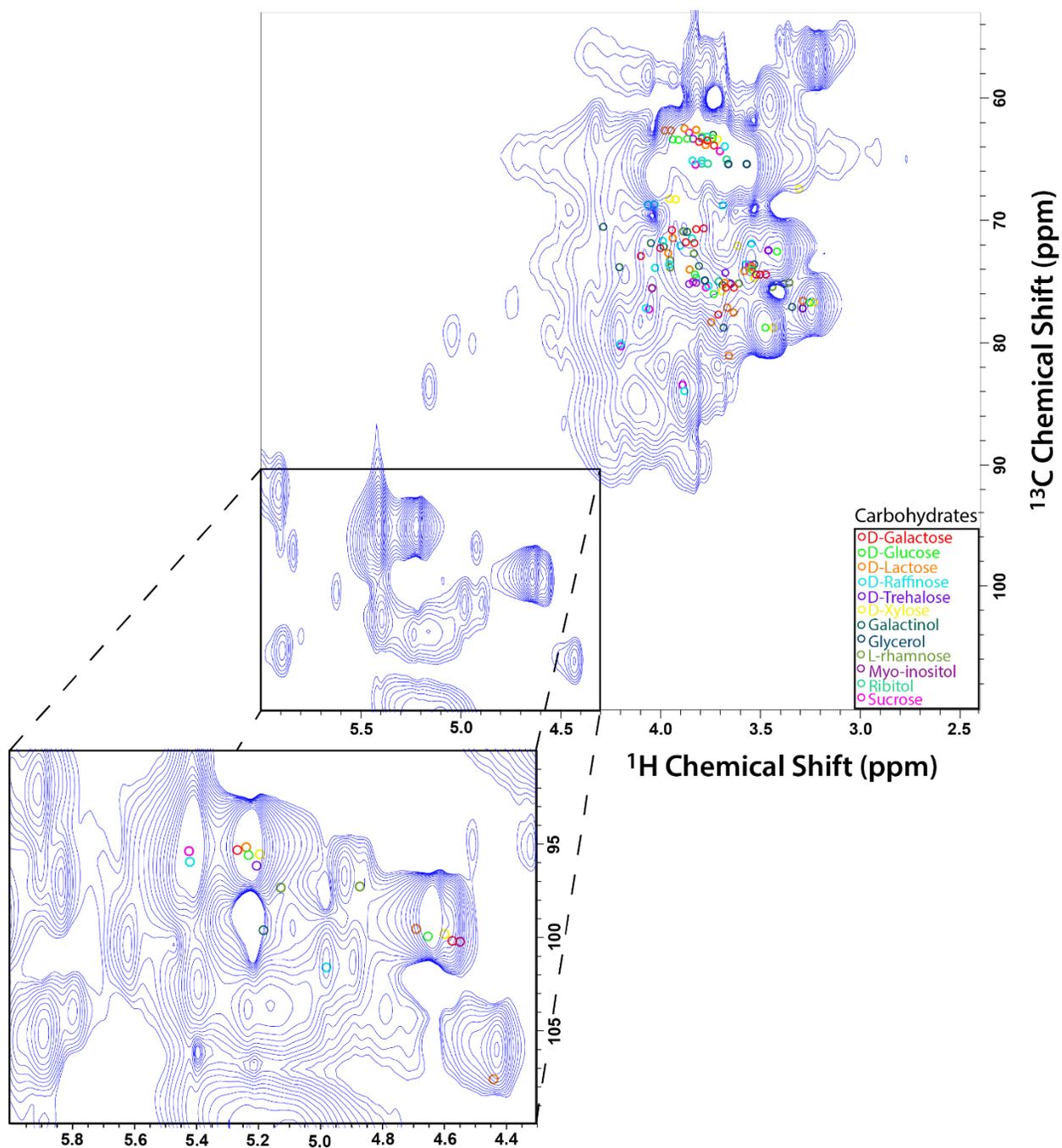
**Fig. S1** Swellability of *Chlorella vulgaris* and *Synechocystis* relative to *Chlamydomonas reinhardtii* as determined by integration of 1D  $^{13}\text{C}$  spectra



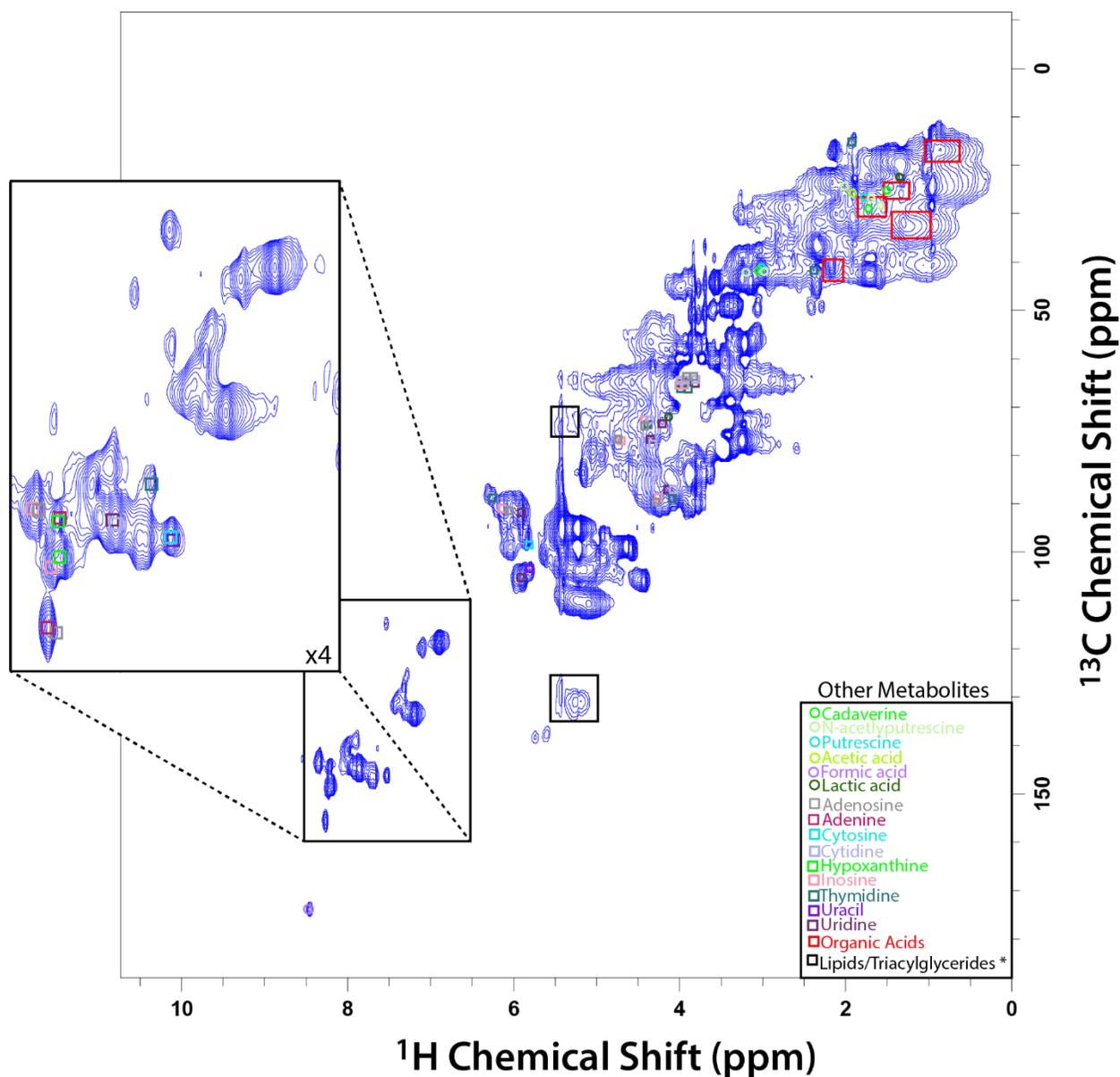
**Fig. S2** 2D TOCSY ( $^1\text{H}$ - $^1\text{H}$ ) spectrum of a  $^{13}\text{C}$ -enriched *Chlamydomonas reinhardtii* cell suspension in  $\text{D}_2\text{O}$  with metabolites and general regions for triacylglycerides and carbohydrates identified. An example triacylglyceride molecule is also displayed



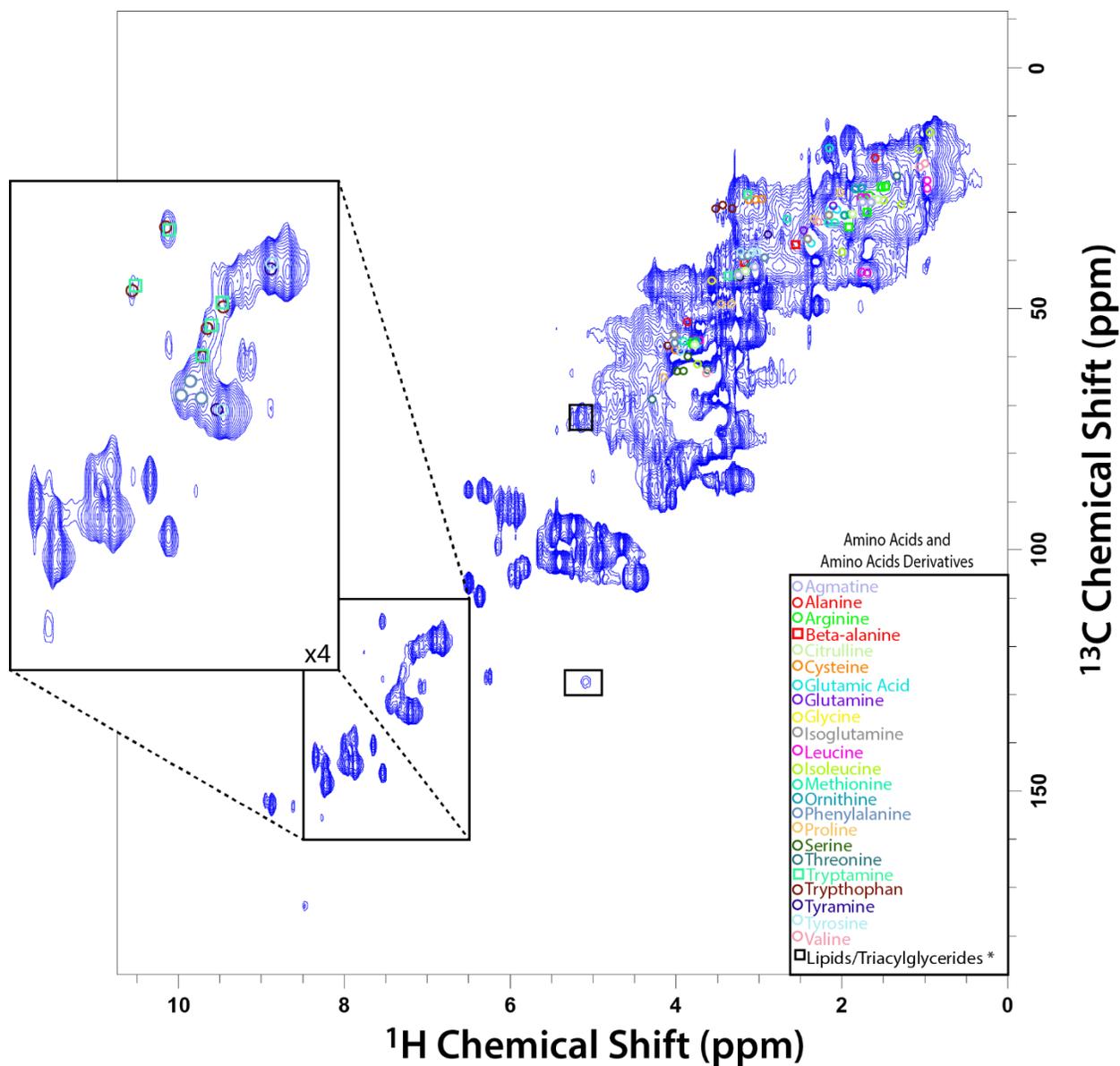
**Fig. S3** 2D HMQC ( $^1\text{H}$ - $^{13}\text{C}$ ) spectrum of a  $^{13}\text{C}$ -enriched *Chlorella vulgaris* cell suspension in  $\text{D}_2\text{O}$  with labelled amino acid and amino acid derivative metabolites. \*The highlighted regions for Lipids/Triacylglycerides represent the methine resonances of the glycerol backbone (top), and the double bond resonances in long chain unsaturated fatty acids (bottom)



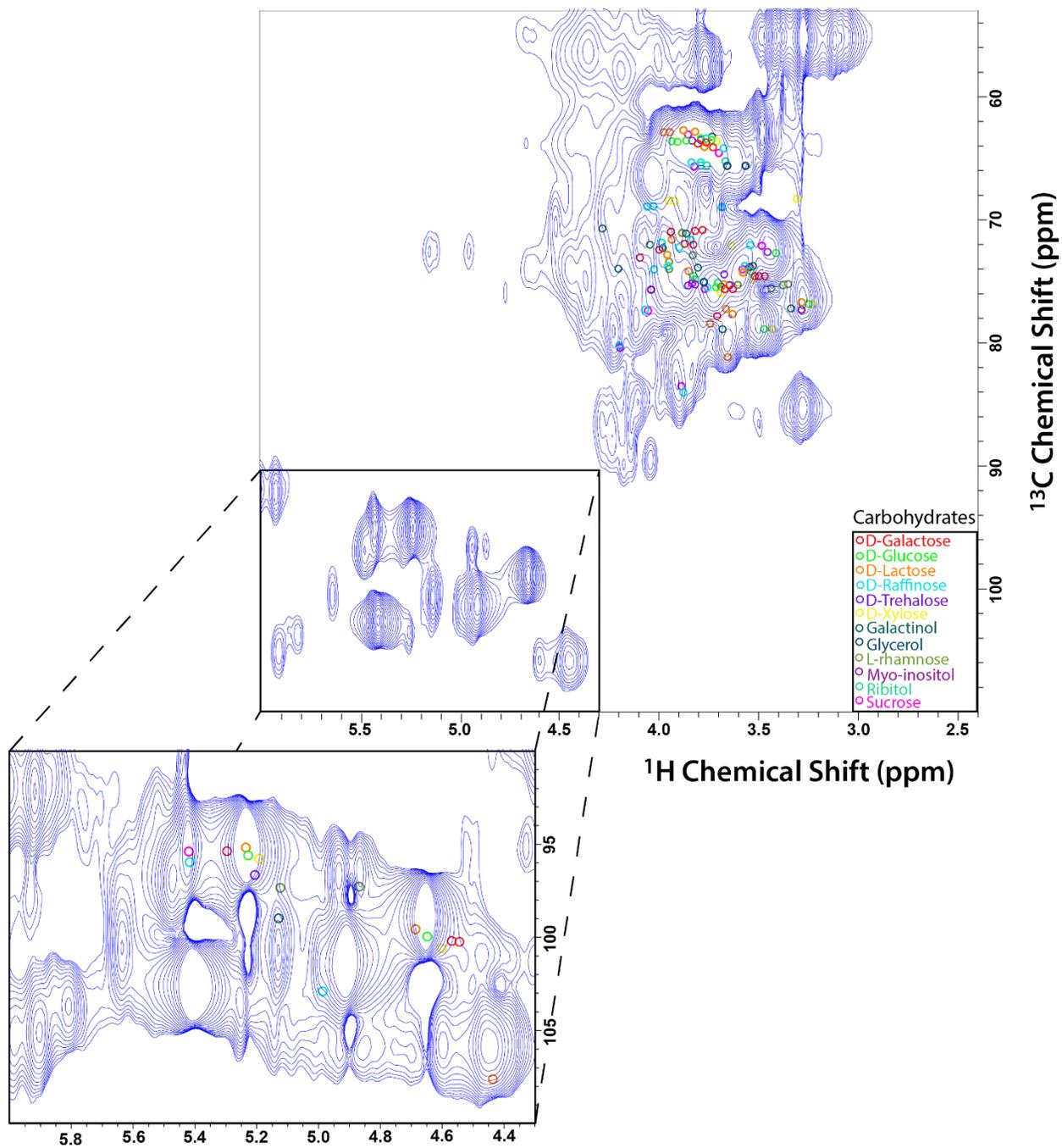
**Fig. S4** 2D HMQC ( $^1\text{H}$ - $^{13}\text{C}$ ) spectrum of a  $^{13}\text{C}$ -enriched *Chlorella vulgaris* cell suspension in  $\text{D}_2\text{O}$  with labelled carbohydrate metabolites



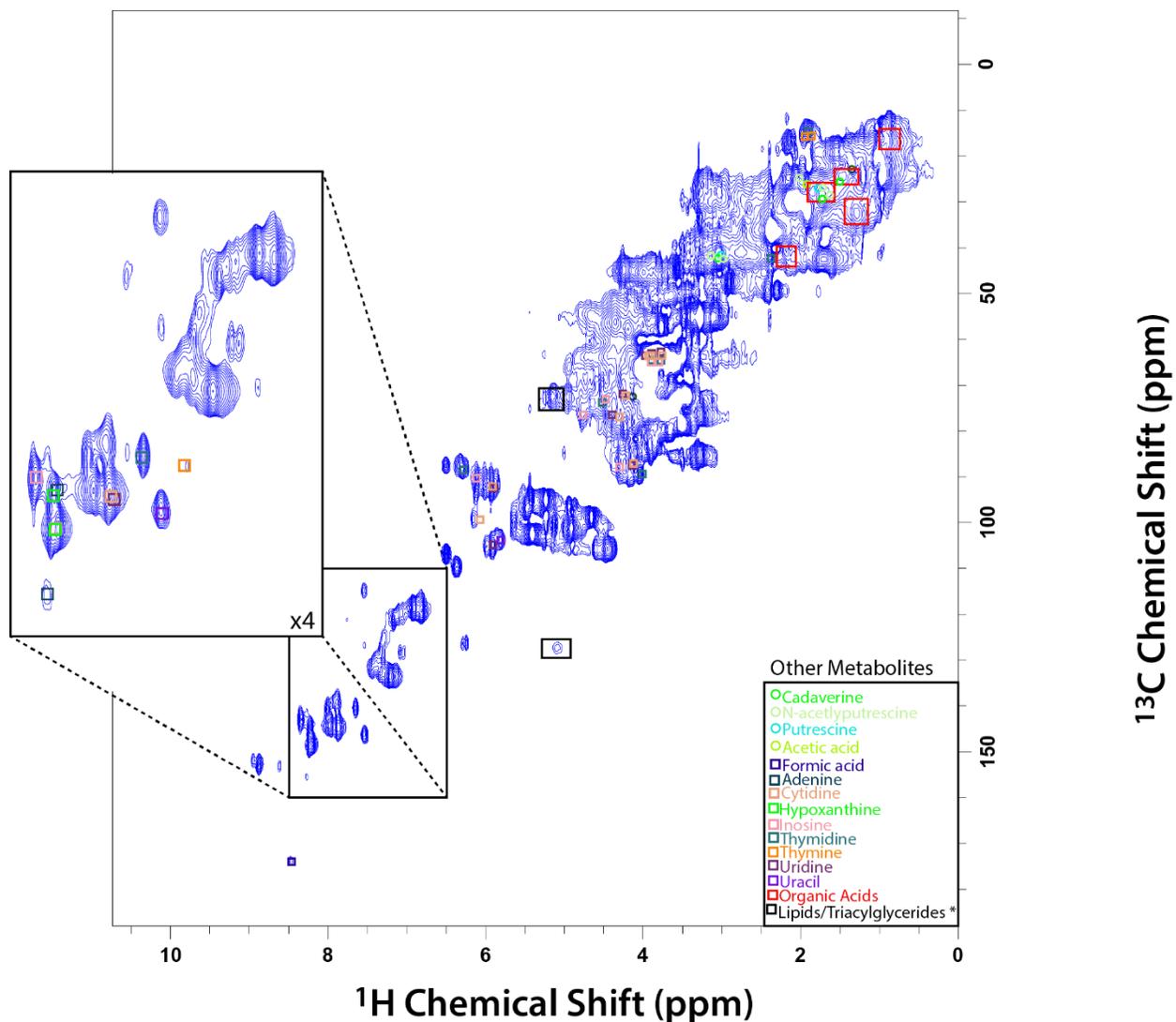
**Fig. S5** 2D HMQC ( $^1\text{H}$ - $^{13}\text{C}$ ) spectrum of a  $^{13}\text{C}$ -enriched *Chlorella vulgaris* cell suspension in  $\text{D}_2\text{O}$  with labelled amine, carboxylic acid, nucleotide base and acetate metabolites. \*The highlighted regions for Lipids/Triacylglycerides represent the methine resonances of the glycerol backbone (top), and the double bond resonances in long chain unsaturated fatty acids (bottom)



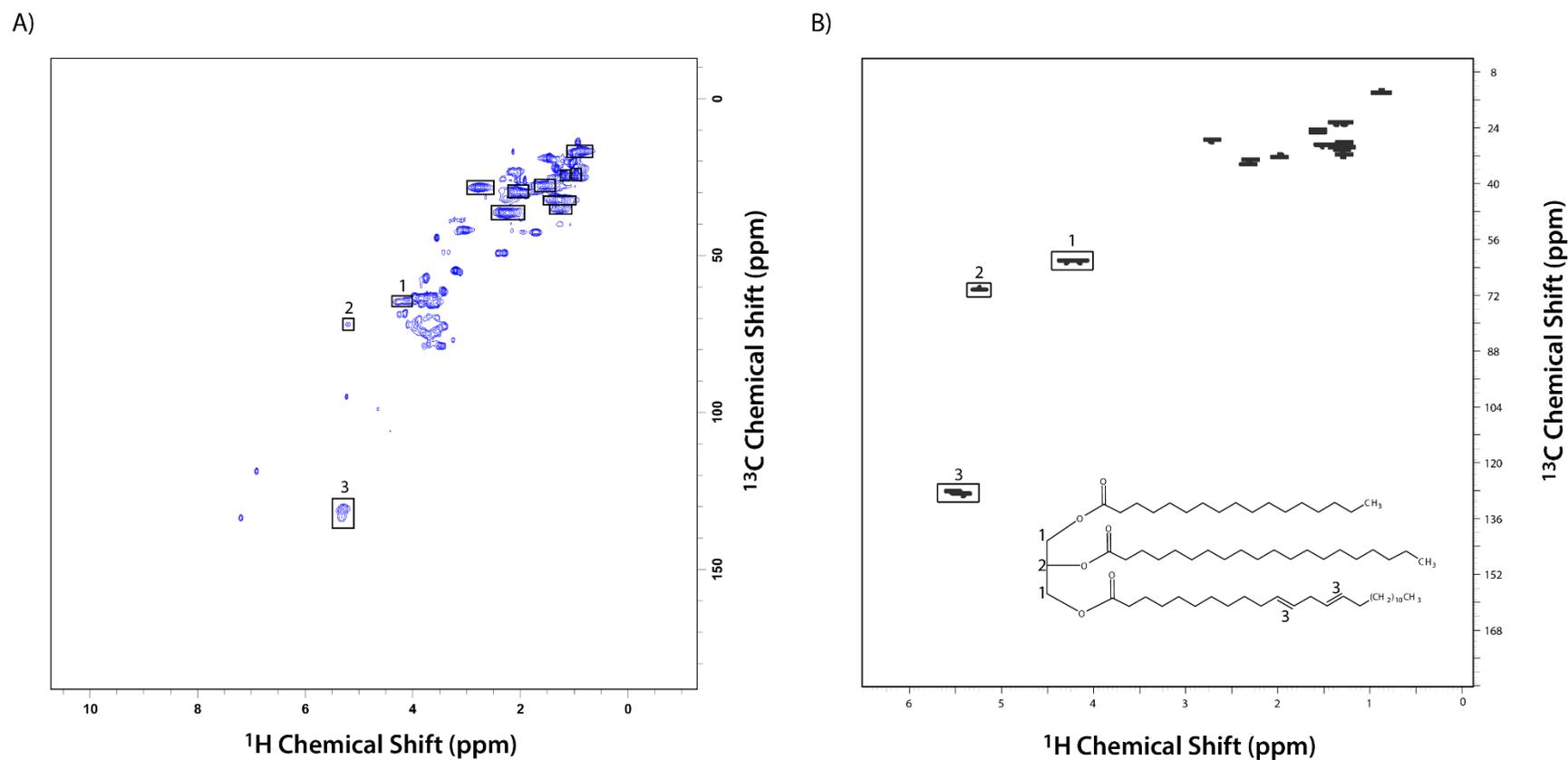
**Fig. S6** 2D HMQC ( $^1\text{H}$ - $^{13}\text{C}$ ) spectrum of a  $^{13}\text{C}$ -enriched *Synechocystis* suspension in  $\text{D}_2\text{O}$  with labelled amino acid and amino acid derivative metabolites. \*The highlighted regions for Lipids/Triacylglycerides represent the methine resonances of the glycerol backbone (top), and the double bond resonances in long chain unsaturated fatty acids (bottom)



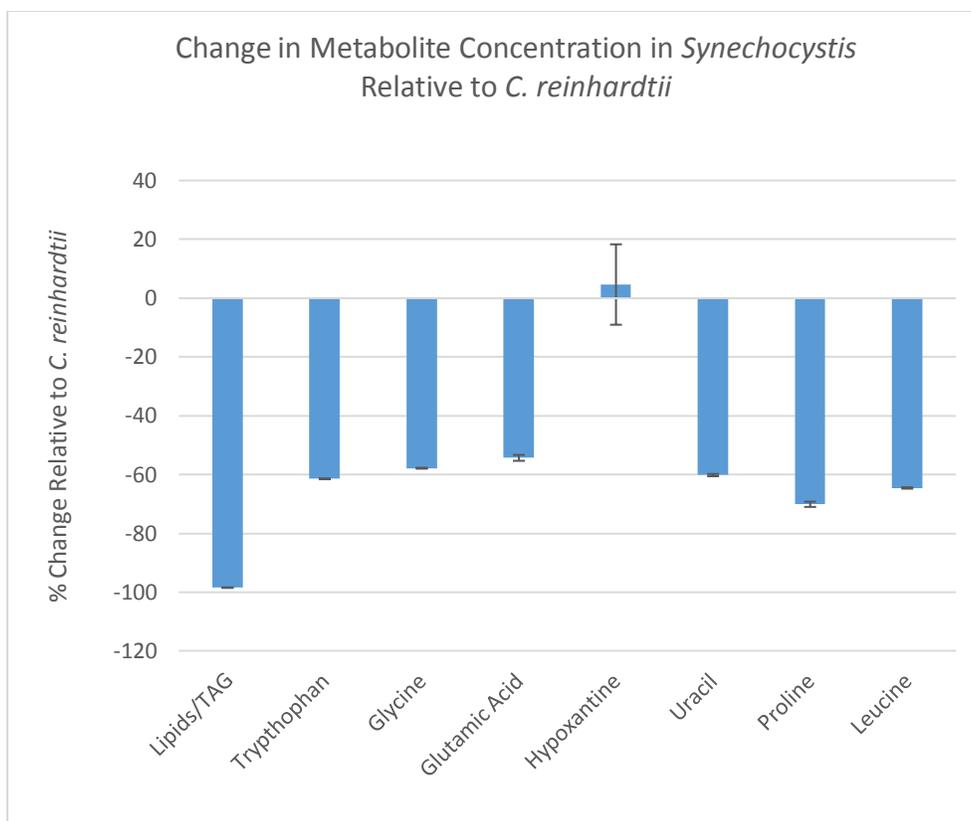
**Fig. S7** 2D HMQC ( $^1\text{H}$ - $^{13}\text{C}$ ) spectrum of a  $^{13}\text{C}$ -enriched *Synechocystis* cell suspension in  $\text{D}_2\text{O}$  with labelled carbohydrate metabolites



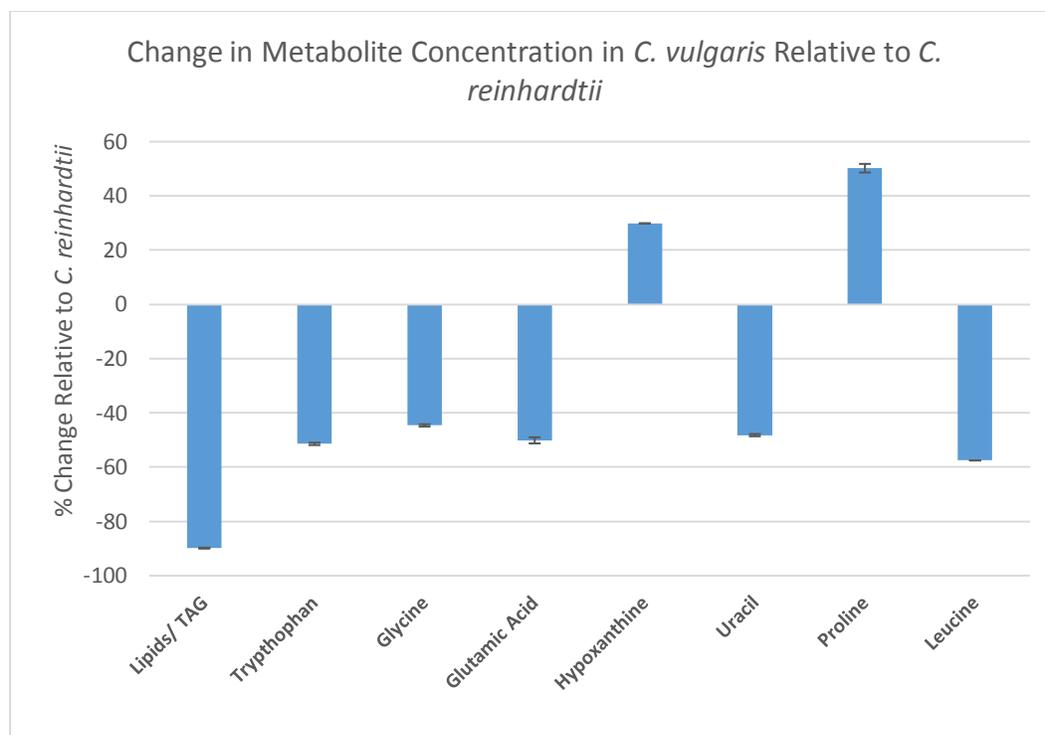
**Fig. S8** 2D HMQC ( $^1\text{H}$ - $^{13}\text{C}$ ) spectrum of a  $^{13}\text{C}$ -enriched *Synechocystis* cell suspension in  $\text{D}_2\text{O}$  with labelled amine, carboxylic acid, nucleotide base and acetate metabolites. \*The highlighted regions for Lipids/Triacylglycerides represent the methine resonances of the glycerol backbone (top), and the double bond resonances in long chain unsaturated fatty acids (bottom)



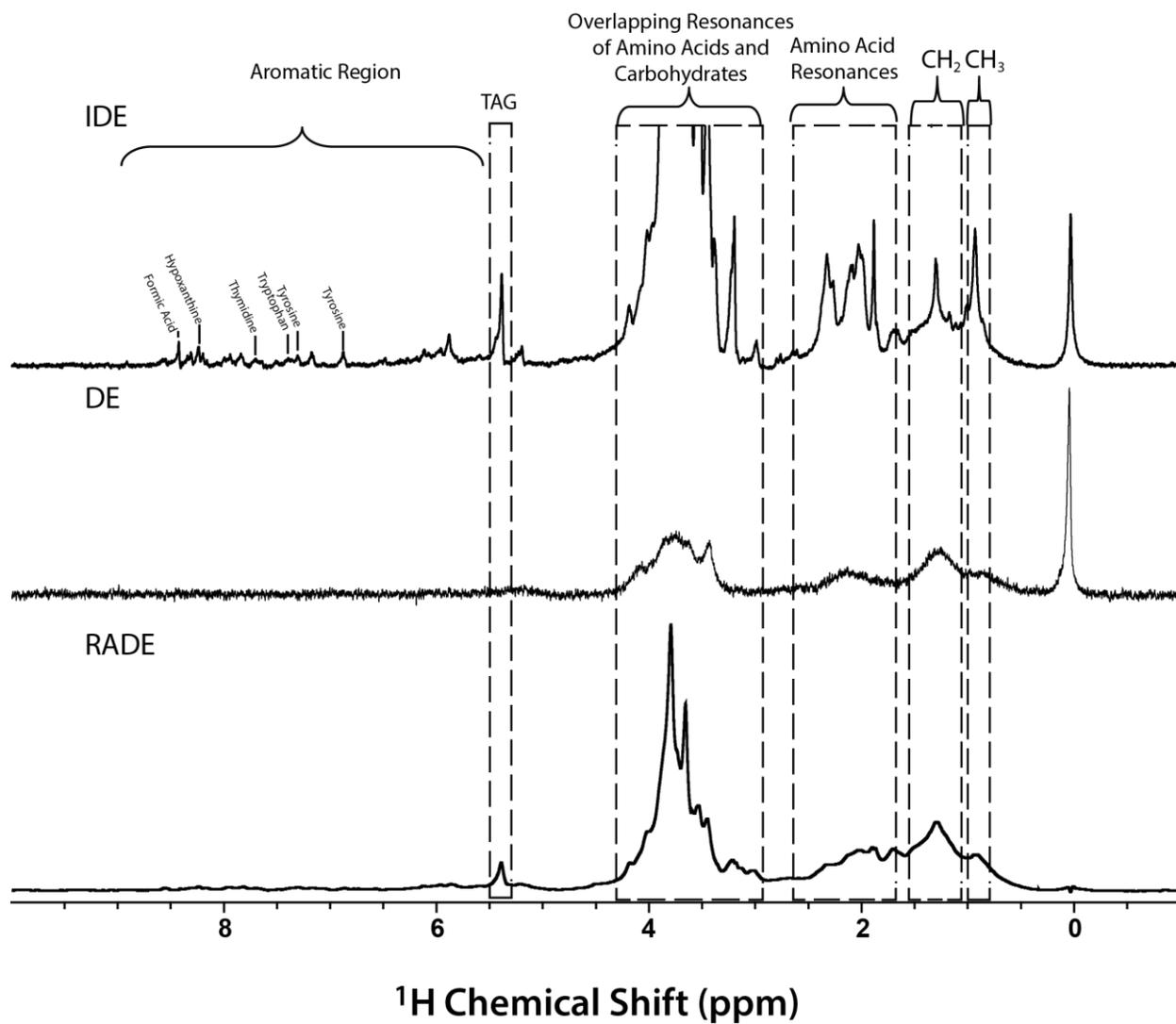
**Fig. S9** A) 2D HMQC ( $^1\text{H}$ - $^{13}\text{C}$ ) spectrum of a  $^{13}\text{C}$ -enriched *Chlamydomonas reinhardtii* cell suspension in  $\text{D}_2\text{O}$  with triacylglyceride peaks labelled. B) A 2D HMQC ( $^1\text{H}$ - $^{13}\text{C}$ ) simulated spectra (simulated spectra was produced using ACD software) of an example triacylglyceride molecule with resonances unique to triacylglycerides denoted by numbers



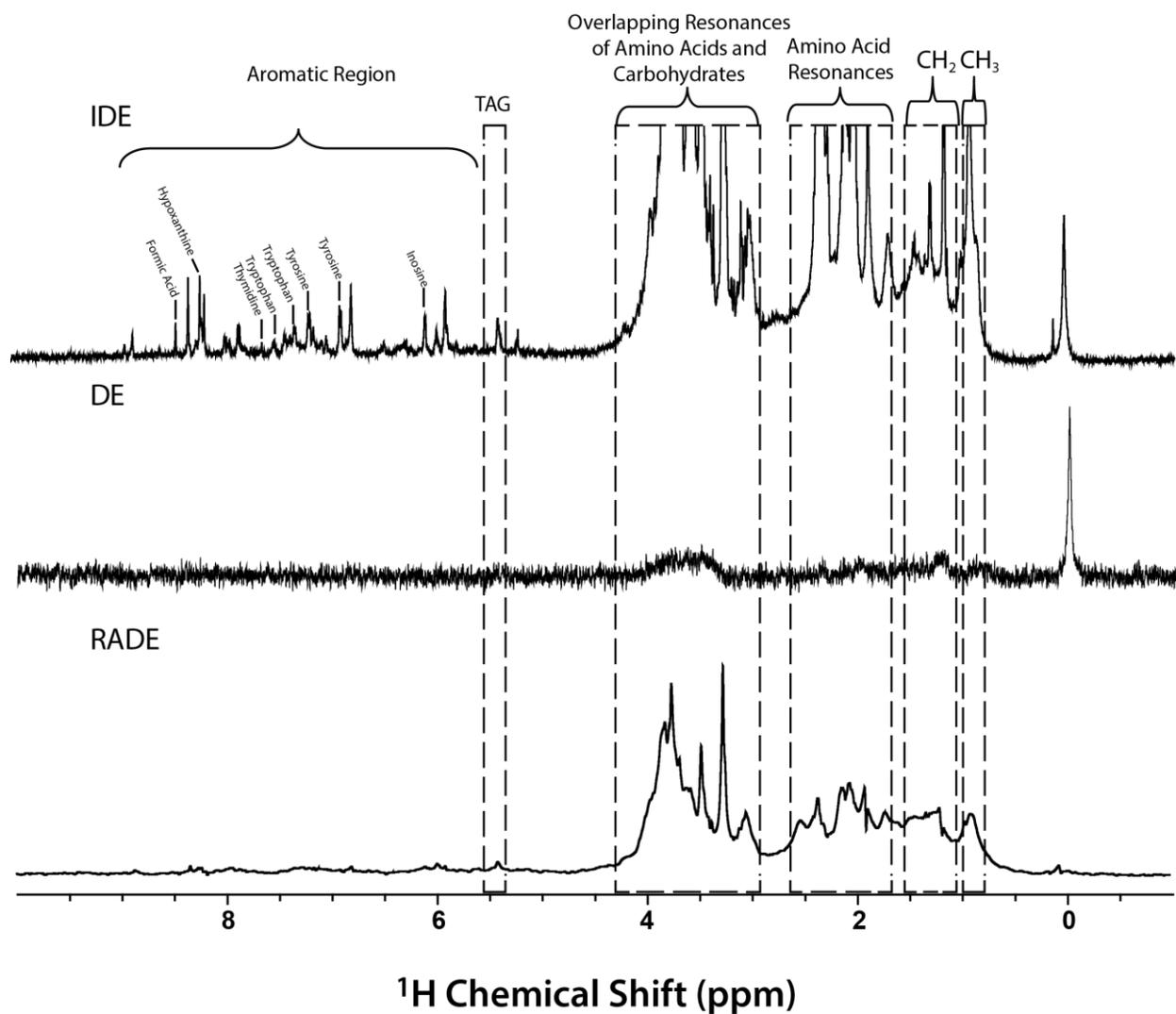
**Fig. S10** Change in concentration (% change) of select identifiable metabolites in  $^{13}\text{C}$ -enriched *Chlorella vulgaris* relative to  $^{13}\text{C}$ -enriched *Chlamydomonas reinhardtii* as determined by integration of peak area



**Fig. S11** Change in concentration (% change) of select identifiable metabolites in  $^{13}\text{C}$ -enriched *Synechocystis* relative to  $^{13}\text{C}$ -enriched *Chlamydomonas reinhardtii* as determined by integration of peak area



**Fig. S12** Comparison of 1D IDE, DE and RADE  $^1\text{H}$  spectra of a  $^{13}\text{C}$ -enriched *Chlorella vulgaris*, cell suspension in  $\text{D}_2\text{O}$  with select chemical shifts labelled



**Fig. S13** Comparison of 1D IDE, DE and RADE  $^1\text{H}$  spectra of a  $^{13}\text{C}$ -enriched *Synechocystis*, cell suspension in  $\text{D}_2\text{O}$  with select chemical shifts labelled

**Table S1** List of all metabolite assignments for  $^{13}\text{C}$ -labelled *Chlamydomonas reinhardtii*, *Chlorella vulgaris* and *Synechocystis* with literature references for confirmation

<i>Chlamydomonas reinhardtii</i>	<i>Chlorella vulgaris</i>	<i>Synechocystis</i>	Literature Reference
<b>Amino Acids and Amino Acid Derivatives</b>			
Agmatine	Agmatine	Agmatine	(Hamana and Matsuzaki 1985)
Alanine	Alanine	Alanine	(Eisenhut et al. 2008, Krall et al. 2009, Lee and Fiehn 2008, May et al. 2008)
Arginine	Arginine	Arginine	(Eisenhut et al. 2008, Krall et al. 2009, May et al. 2008)
	Beta-alanine	Beta-alanine	(Eisenhut et al. 2008, Krall et al. 2009, Lee and Fiehn 2008, May et al. 2008)
Citrulline	Citrulline	Citrulline	(Bölling and Fiehn 2005, Eisenhut et al. 2008, May et al. 2008)
Cysteine	Cysteine	Cysteine	(Bölling and Fiehn 2005, May et al. 2008, Yoshikawa et al. 2013)
Glutamic acid	Glutamic acid	Glutamic acid	(Bölling and Fiehn 2005,

			Eisenhut et al. 2008, Lee and Fiehn 2008)
Glutamine	Glutamine	Glutamine	(Bölling and Fiehn 2005, Eisenhut et al. 2008, Krall et al. 2009)
Glycine	Glycine	Glycine	(Bölling and Fiehn 2005, Eisenhut et al. 2008, Lee and Fiehn 2008)
-----	Histidine	-----	(Bölling and Fiehn 2005, May et al. 2008, Yoshikawa et al. 2013)
Isoglutamine	Isoglutamine	Isoglutamine	(Belmont & Miller, 1965)
Leucine	Leucine	Leucine	(Bölling and Fiehn 2005, Eisenhut et al. 2008, Krall et al. 2009)
Isoleucine	Isoleucine	Isoleucine	(Bölling and Fiehn 2005, Krall et al. 2009, Lee and Fiehn 2008)put
Methionine	Methionine	Methionine	(Bölling and Fiehn 2005,

			Eisenhut et al. 2008, Lee and Fiehn 2008)
Ornithine	Ornithine	Ornithine	(Bölling and Fiehn 2005, Eisenhut et al. 2008, Krall et al. 2009)
Phenylalanine	Phenylalanine	Phenylalanine	(Bölling and Fiehn 2005, Eisenhut et al. 2008, Krall et al. 2009, Lee and Fiehn 2008)
Proline	Proline	Proline	(Bölling and Fiehn 2005, Eisenhut et al. 2008, Krall et al. 2009, Lee and Fiehn 2008)
Serine	Serine	Serine	(Bölling and Fiehn 2005, Eisenhut et al. 2008, Krall et al. 2009, Lee and Fiehn 2008)
Threonine	Threonine	Threonine	(Bölling and Fiehn 2005, Eisenhut et al. 2008, Krall et al. 2009, Lee

			and Fiehn 2008)
-----	-----	Tryptamine	(Byeon et al. 2013)
Tryptophan	Tryptophan	Tryptophan	(Bölling and Fiehn 2005, Eisenhut et al. 2008)
Tyramine	Tyramine	Tyramine	(Yoshikawa et al. 2013)
Tyrosine	Tyrosine	Tyrosine	(Bölling and Fiehn 2005, Eisenhut et al. 2008, Krall et al. 2009, Lee and Fiehn 2008)
Valine	Valine	Valine	(Bölling and Fiehn 2005, Eisenhut et al. 2008, Krall et al. 2009)
<b>Carbohydrates</b>			
D-Galactose	D-Galactose	D-Galactose	(Eisenhut et al. 2008, Lee and Fiehn 2008)
D-Glucose	D-Glucose	D-Glucose	(Eisenhut et al. 2008, Krall et al. 2009, May et al. 2008)
D-Lactose	D-Lactose	D-Lactose	(Krall et al. 2009)
D-Raffinose	D-Raffinose	D-Raffinose	(Salerno and Pontis 1989)
D-Trehalose	D-Trehalose	D-Trehalose	(Bölling and Fiehn

			2005, Krall et al. 2009, Lee and Fiehn 2008)
D-Xylose	D-Xylose	D-Xylose	(Eisenhut et al. 2008, May et al. 2008)
Galactinol	Galactinol	Galactinol	(Bölling and Fiehn 2005, May et al. 2008)
Glycerol	Glycerol	Glycerol	(Bölling and Fiehn 2005, Lee and Fiehn 2008)
L-Rhamnose	L-Rhamnose	L-Rhamnose	(May et al. 2008)
Myo-inositol	Myo-inositol	Myo-inositol	(Bölling and Fiehn 2005, Krall et al. 2009, Lee and Fiehn 2008)
Ribitol	Ribitol	Ribitol	(Maruo et al. 1965)
Sucrose	Sucrose	Sucrose	(Eisenhut et al. 2008, Lee and Fiehn 2008, Maruo et al. 1965, May et al. 2008)
<b>Other Metabolites</b>			
Cadaverine	Cadaverine	Cadaverine	(Hamana and Matsuzaki 1985)
N-acetylputrecine	N-acetylputrecine	-----	(May et al. 2008)
Putrescine	Putrescine	Putrescine	(Bölling

			and Fiehn 2005, Lee and Fiehn 2008)
Acetic Acid	Acetic Acid	Acetic Acid	(Ross et al. 2010)
Formic acid	Formic Acid	Formic Acid	(Ross et al. 2010)
Lactic Acid	Lactic Acid	Lactic Acid	(May et al. 2008)
-----	Adenine	Adenine	(Eisenhut et al. 2008, Krall et al. 2009, Lee and Fiehn 2008, May et al. 2008)
-----	Adenosine	-----	(Eisenhut et al. 2008, Lee and Fiehn 2008)
-----	Cytidine	Cytidine	(Yoshikawa et al. 2013)
Cytosine	Cytosine	-----	(Yoshikawa et al. 2013)
Hypoxanthine	Hypoxanthine	Hypoxanthine	(Yoshikawa et al. 2013)
-----	Inosine	Inosine	(Yoshikawa et al. 2013)
Thymine	-----	Thymine	(Eisenhut et al. 2008, Glaser et al. 1973)
-----	Thymidine	Thymidine	(Glaser et al. 1973)
Uracil	Uracil	Uracil	(Bölling and Fiehn 2005, Eisenhut et al. 2008)
-----	Uridine	Uridine	(May et al. 2008)

## References

- Belmont, L. and Miller, J. (1965) The utilization of glutamine by algae. *Journal of Experimental Botany* 16(2), 318-324.
- Bölling, C. and Fiehn, O. (2005) Metabolite profiling of *Chlamydomonas reinhardtii* under nutrient deprivation. *Plant physiology* 139(4), 1995-2005.
- Byeon, Y., Lee, K., Park, Y.I., Park, S. and Back, K. (2013) Molecular cloning and functional analysis of serotonin N- acetyltransferase from the cyanobacterium *Synechocystis* sp. PCC 6803. *Journal of pineal research* 55(4), 371-376.
- Eisenhut, M., Huege, J., Schwarz, D., Bauwe, H., Kopka, J. and Hagemann, M. (2008) Metabolome phenotyping of inorganic carbon limitation in cells of the wild type and photorespiratory mutants of the cyanobacterium *Synechocystis* sp. strain PCC 6803. *Plant physiology* 148(4), 2109-2120.
- Glaser, V., Al-Nuri, M., Groshev, V. and Shestakov, S. (1973) The labelling of nucleic acids by radioactive precursors in the blue-green algae *Anacystis nidulans* and *Synechocystis aquatilis* Sanv. *Archiv für Mikrobiologie* 92(3), 217-226.
- Gorman, D.S., Levine, R. P. (1965) Cytochrome f and plastocyanin: their sequence in the photosynthetic electron transport chain of *Chlamydomonas reinhardtii*. *Proc. Natl. Acad. Sci. USA*, 54(6), 1665–1669.
- Hamana, K. and Matsuzaki, S. (1985) Distribution of polyamines in prokaryotes, algae, plants and fungi. *Polyamines: basic and clinical aspects* (Eds K. Imanoh, F. Suzuki, O. Suzuki, U. Bachrach), VNU Science Press, Netherlands, 105-112.
- Krall, L., Huege, J., Catchpole, G., Steinhauser, D. and Willmitzer, L. (2009) Assessment of sampling strategies for gas chromatography–mass spectrometry (GC–MS) based metabolomics of cyanobacteria. *Journal of Chromatography B* 877(27), 2952-2960.
- Lee, D.Y. and Fiehn, O. (2008) High quality metabolomic data for *Chlamydomonas reinhardtii*. *Plant methods* 4(1), 7.
- Maruo, B., Hattori, T. and Takahashi, H. (1965) Excretion of ribitol and sucrose by green algae into the culture medium. *Agricultural and Biological Chemistry* 29(12), 1084-1089.
- May, P., Wienkoop, S., Kempa, S., Usadel, B., Christian, N., Rupprecht, J., Weiss, J., Recuenco-Munoz, L., Ebenhöf, O. and Weckwerth, W. (2008) Metabolomics-and proteomics-assisted genome annotation and analysis of the draft metabolic network of *Chlamydomonas reinhardtii*. *Genetics* 179(1), 157-166.

Ross, A., Biller, P., Kubacki, M., Li, H., Lea-Langton, A. and Jones, J. (2010) Hydrothermal processing of microalgae using alkali and organic acids. *Fuel* 89(9), 2234-2243.

Salerno, G.L. and Pontis, H.G. (1989) Raffinose synthesis in *Chlorella vulgaris* cultures after a cold shock. *Plant physiology* 89(2), 648-651.

Yoshikawa, K., Hirasawa, T., Ogawa, K., Hidaka, Y., Nakajima, T., Furusawa, C. and Shimizu, H. (2013) Integrated transcriptomic and metabolomic analysis of the central metabolism of *Synechocystis* sp. PCC 6803 under different trophic conditions. *Biotechnology journal* 8(5), 571-580.