



Advanced mass balance characterization and fractionation of algal biomass composition

Stefanie Van Wychen¹ · Steven M. Rowland² · Kaitlin C. Lesco³ · Peter V. Shanta³ · Tao Dong¹ · Lieve M. L. Laurens³ 

Received: 11 February 2021 / Revised and accepted: 18 May 2021 / Published online: 1 July 2021
© The Author(s) 2021

Abstract

Opportunities associated with biomass production and bioproduct isolation from algae-derived feedstocks are plentiful and promising; however, there are challenges associated with realizing these applications. One of the most important, and often overlooked, challenges is the lack of availability of a strong foundation of compositional analysis methods validated on microalgal biomass. Currently, compositional analysis in algae is dominated by the use of interference-prone methods, a lack of full mass balance accounting, and the use of top-down approaches that bin all unaccounted-for mass into a single category, such as carbohydrates. We present here an approach based on a bottom-up algal biomass characterization aimed at moving towards, and highlighting the importance of, full and accurate mass closure to achieve the maximum economic potential from a sustainable and renewable feedstock. Algal biomass representing three genera, *Nannochloropsis*, *Scenedesmus*, and *Monoraphidium*, was subjected to a cell rupture and fractionation process, followed by detailed characterization of each fraction to determine the partitioning of measured and unknown components. The goal of this work is to identify where the missing components partition, and develop a strategy to close the mass balance or identify the unknowns, while utilizing a rigorous characterization approach for characterizing algal biomass. Although only 75–80% of the biomass was accounted for, the fractionation approach utilized here provides key insight into possible chemical components for future investigations.

Keywords Algae · Mass balance · Fractionation · Lipids · Carbohydrates · Protein

Introduction

The success of algae in the context of innovative agriculture and commodity production relies on the ability to accurately determine the composition of algal biomass for economic and process modeling. Accurate and robust analytical methods are paramount in determining the presence and quantification of compounds available for chosen industrial processes. Reduction in analytical uncertainty provides better data inputs for realistic cost targets and well-informed

strain selection for desired product portfolios. High-quality untargeted analyses also provide the potential for identification of high-value components, which may be utilized by a biorefinery approach for the valorization of algal biomass. Recent work has shown that products such as lipids and protein, which are predominantly used for biodiesel and animal feed, may actually produce high-value commodities such as therapeutics, bio-derived pigments, surfactants, and polymer precursors (Christaki et al. 2013; Pleissner et al. 2015; Hess et al. 2018; Sathasivam and Ki 2018; Sathasivam et al. 2019). Due to the difficult economics surrounding bioenergy production from algal biomass, it is more important than ever to include commodity chemicals in the algae production value chain in a biorefinery approach (Dong et al. 2016a; Laurens et al. 2017). However, transitioning from a bioenergy-focused biomass production strategy to include sources of higher-value bioproducts may only become feasible by understanding the complete composition of algae, which will aid with the discovery and isolation of novel products.

Compositional analysis of algae is known to be challenging because of the complexity of the often single-celled

✉ Lieve M. L. Laurens
Lieve.Laurens@nrel.gov

¹ Renewable Resources and Enabling Sciences Center, National Renewable Energy Laboratory, 15013 Denver West Parkway, Golden, CO, USA

² Catalytic Carbon Transformation and Scale-Up Center, National Renewable Energy Laboratory, 15013 Denver West Parkway, Golden, CO, USA

³ Biosciences Center, National Renewable Energy Laboratory, 15013 Denver West Parkway, Golden, CO, USA

biochemistry (as opposed to a dedicated storage compartment in plants) and the biological and biochemical diversity between different species. In addition, there are analytical challenges associated with the quantitative determination of each of the major organic constituents in algal biomass (*i.e.*, proteins, carbohydrates, and lipids). Nutritional testing methods used by commercial labs describe algae biomass using a “top-down” approach, as mandated in the Code of Federal Regulations Title 21, part 101 Food Labeling (Food and Drugs, 2020), which describes carbohydrates as 100%—ash—protein—lipids. The ash, proteins, and lipids are directly measured, but the carbohydrates are a calculation. Where this approach reaches 100% mass balance, it is done indirectly, absorbing any “unknown” components into the carbohydrate fraction and thus affecting downstream calculations used in economic assessment models. A more unambiguous assessment of the biomass uses direct measurements of each of the components, defined here as a bottom-up approach to compositional analysis and mass balance.

Directly measuring ash, carbohydrates, proteins, and lipids, as was introduced above as the bottom-up approach to mass balance, allows for accounting of close to 90% of the biomass composition in late harvest (*i.e.*, early and late stationary growth phase) algae, where the lipid and carbohydrate concentration are the highest, and the protein concentration is low (Table 1). This observation is consistent with published literature on shifts in biomass compositional profile as well as in mass balance closure (Laurens et al. 2014). However, using the same analytical methods on early harvest algae (*i.e.*, exponential growth phase), which is high in protein and low in lipids, accounts for approximately 70% or less of the biomass. Algae harvested in this early phase is of increasing interest due to the more favorable economics of shorter growth cycles. Therefore, developing an understanding of the unaccounted-for biomass becomes important to the continued development of bioproducts and biofuels. Table 1 shows the mass yields for the components described above, as well as the total mass balance, for early-, mid-, and late-harvest *Chlorella* sp. and *Scenedesmus* sp. samples.

In an effort to isolate and better understand the properties of the unaccounted-for mass in the high protein harvests, the

work presented here used a fractionation approach based on solubility in organic solvents, followed by a bottom-up approach to quantifying the major macromolecules in each fraction. The goal of this work was to understand which fraction is responsible for the unaccounted-for mass, in order to determine the possible composition of the unknown portion. Identification of the unknowns allows for planning and developing methods for their quantification, leading to a more complete mass balance for early harvest microalgae in future research.

Materials and methods

Biomass

Algal biomass from three strains was used for this work. *Nannochloropsis* sp. and *Scenedesmus acutus* (LRB-AP-0401), recently reclassified as *Tetrademus acutus* (Wynne and Hallan 2016), were grown under nutrient replete conditions at Arizona Center for Algae Technology and Innovation (AzCATI), at Arizona State University, in Mesa, Arizona, and generously provided by Dr. John McGowen. A third species, *Monoraphidium minutum* (26BAM), was grown at the University of Arizona, Tucson, at the algae testbed under nutrient replete conditions (Crowe et al. 2012; Huesemann et al. 2013). Previous analyses, shown in Table 1, were conducted on *Chlorella vulgaris* (LRB-1201) and *Scenedesmus acutus* (LRB-AP-0401) and are shown here to describe the difficulty in mass balance closure for early harvest/high protein algae. These samples were collected from flat-panel photobioreactor experiments, as three distinct harvest points in the growth cycle (early, mid, and late harvest, as shown in Table 1), specifically designed for nutrient depletion and compositional dynamics studies, as described previously in great detail (Laurens et al. 2014). Harvested biomass was lyophilized to preserve the physiological and biochemical integrity of the samples prior to long-term freezer storage. While we are aware that the process of lyophilization may alter the structure and accessibility of the cell wall, in collaborative settings, this process is

Table 1 Measured biomass composition (%DW) based on standard laboratory analytical procedures for two algae species and three harvest conditions, early, mid or late in the process of nutrient depletion. All data expressed as a percentage of the oven dried biomass

Species	Harvest	Ash	Carbohydrates	Protein	Lipids (FAME)	Mass balance
<i>Chlorella</i> sp.	Early	6.0	9.5	42.5	12.2	70.2
<i>Chlorella</i> sp.	Mid	2.7	42.7	13.1	22.4	80.9
<i>Chlorella</i> sp.	Late	2.5	30.9	10.7	40.5	84.6
<i>Scenedesmus</i> sp.	Early	7.0	22.6	39.5	8.6	77.7
<i>Scenedesmus</i> sp.	Mid	1.6	52.3	9.0	30.9	93.8
<i>Scenedesmus</i> sp.	Late	1.4	42.3	7.7	40.2	91.6

necessary and representative of the majority of algae samples treated for compositional analysis.

Cell disruption and extraction

Cell disruption and extraction experiments included accelerated solvent extraction (ASE), sonication with organic extraction, and bead-beating with organic extraction. Sonication and bead-beating were determined to yield comparable results, and bead-beating was utilized for the work below due to the ability to generate greater amounts of biomass in less time. The procedures for ASE extraction and sonication with extraction are located in the supplemental material. Bead-beating experiments were performed with a BioSpec Mini-Beadbeater (BioSpec Products, USA). Approximately 400 mg of lyophilized algae was added to polypropylene micro-vials, along with 10 stainless steel beads (2.4 mm). The samples were cooled on ice for 1 h prior to bead-beating to prevent excessive heating. Each vial was subject to 6 rounds of bead-beating in 1 min increments, for a total of 6 min of bead-beating, with approximately 10 min on ice between each round. The biomass was then transferred to Teflon centrifuge tubes, weighed for mass balance calculations, and rehydrated with 5 mL of water overnight at 4 °C. This procedure was conducted in triplicate for each algae strain. Each sample was extracted with the method described below.

Methanol (5 mL) was added to the algae:water mixture and hand shaken for 1 min. The samples were then centrifuged at 9299 rcf for 20 min at 4 °C. The supernatant was carefully removed to avoid disruption of the algae pellet and collected in pre-weighed vials before being dried to about 1 mL volume at 25 °C under nitrogen. The samples were then subjected to four extractions as described below. Each extraction was performed by adding 6 mL of methanol and 6 mL of chloroform sequentially. The centrifuge tube was shaken by hand for 1 min after the addition of each solvent. Each sample was then allowed to sit for 10 min in organic solvent prior to centrifuging. All four methanol:chloroform extracts were combined into one vial and dried to approximately 1 mL volume at 25 °C under nitrogen. The residual biomass for each sample was grey/white in color after extraction, except for *Scenedesmus* which appeared to be grey with a faint green tint. The residual biomass was dried under vacuum at 40 °C overnight prior to obtaining a final weight. The 1-mL volume of each liquid fraction was diluted to a volume of 10 mL in a class A volumetric flask using the fraction's respective solvents. This procedure yielded three fractions (water:methanol, methanol:chloroform, and residual biomass) that were further studied to determine the chemical composition of each fraction. While all fractions' gravimetric recoveries are reported, it is necessary to note

that that absolute recovery yields may be different for different batches of materials, and also for fresh biomass.

Total solids and ash

Total solids and ash were determined on the whole and residual biomass as described below. Approximately 25 mg, exact weight recorded, of biomass was weighed into a pre-combusted, pre-weighed crucible and dried for 2 days under vacuum at 40 °C. A dry weight was then collected on the biomass prior to combustion using a ramping oven as is previously described (Van Wychen and Laurens 2015b). The water:methanol and methanol:chloroform extracts were analyzed for total solids and ash as follows: exactly 3 mL of extract was placed into a pre-combusted and pre-weighed crucible and slowly evaporated at 60 °C on a hot plate prior to being dried overnight at 40 °C under vacuum. A dry weight for the 3 mL was recorded, and the crucibles were combusted as described above for the solid samples.

Fatty acid methyl ester determination

Baseline content of fatty acid methyl esters (FAME) was determined on the whole biomass, as well as the fractions, using an in situ transesterification procedure (Van Wychen et al. 2015). Briefly, 7 to 10 mg of biomass, or the equivalent extracted fraction, was dried for 2 days at 40 °C under vacuum, and a dry weight was recorded. A known amount of tridecanoic acid (C13) methyl ester in hexane was added to each sample and standards as a recovery standard. Chloroform:methanol (2:1 v/v) was added to the samples in the amount of 0.2 mL, to facilitate cell wall penetration, before transesterification with 0.3 mL of HCl in methanol (2.1% v/v). Samples were then heated for 1 h at 85 °C on a digital hot block before 1 mL of n-hexane was added to extract FAMEs. FAMEs were analyzed by gas chromatography:flame ionization detection (GC:FID) on an Agilent 7890 N; DB-WAX-MS column with dimensions 30 m × 0.25 mm i.d. and 0.25 µm film thickness. Details of the temperature program, flow rates, and standards have been previously described (Laurens et al. 2012a; Van Wychen et al. 2015).

Lipid separation

The lipids from each sample were separated on a water-deactivated silica gel column. The silica gel was first dried in an oven overnight at 110 °C. Water was then added to the silica gel in a ratio of 6.7 mL of water to 20 g of silica gel. The water used was purified to 18.2 MΩ with a Milli-Q water purifier (MilliporeSigma, USA). A 10-mL SPE cartridge (Agilent TechnologiesUSA) was dry-packed with 2 g of water-deactivated silica gel, then conditioned

with 10 mL of hexane. Between 60 and 100 mg of the methanol:chloroform, extraction from each algae sample was dissolved in 1 mL of chloroform. The samples were pipetted into 1 mL of hexane in the headspace of the column and loaded onto the stationary phase with positive pressure. Five fractions were then eluted with the following: fraction 1) 5 mL of hexane:chloroform (1:1 v/v), fraction 2) 5 mL of chloroform (100%), fraction 3) 5 mL of chloroform (100%), fraction 4) 10 mL of chloroform:methanol (1:1 v/v), and fraction 5) 5 mL methanol (100%). Each fraction was dried and weighed for mass yield.

Carbohydrates as monomers

The carbohydrate content was determined for the whole biomass, as well as the three fractions generated from solvent extraction. Carbohydrate content was determined using a sulfuric acid hydrolysis, involving 2 steps, the first 1 h at 30 °C in 72% sulfuric acid (w/w) in a water bath and the second for 1 h at 121 °C in 4% (w/w) sulfuric acid in an autoclave, followed by quantification of resulting monomers using high performance anion-exchange chromatography (HPAEC) equipped with pulsed amperometric detection (PAD) (Van Wycken and Laurens 2015a). In brief, 25 mg of biomass or 10 mg of dried solvent extract were subjected to the two-stage sulfuric acid hydrolysis. Once hydrolyzed, samples were filtered, 0.2 µm nylon filters, and analyzed for monomeric carbohydrates (fucose, rhamnose, arabinose, galactose, glucose, xylose, and mannose) on a Dionex ICS-5000+, HPAEC-PAD system equipped with a PA-1 column (Dionex #035,391) and guard cartridge (Dionex #043,096). The column and detector were set to 35 °C and the following eluent regime was applied: 10 min at 1 mL min⁻¹ with 200 mM NaOH, and 30 min at 14 mM NaOH for equilibration purposes. Monomers were eluted with 14 mM NaOH for 20 min, and quantified by PAD using Waveform A as described before (Templeton et al. 2012).

N analysis

The N content of the whole and residual biomass was determined by combustion using an Elementar Vario EL cube (Langensfeld, Germany). In brief, approximately 5 mg of sample is combusted at 950 °C, and the resulting gas is transported in helium to the reduction and adsorption tubes. The intake pressure was set to 1200 psi. Detection was performed on a thermal conductivity detector. Nitrogen-to-protein conversion factors were calculated (Lourenço et al. 2004; Templeton and Laurens 2015). To determine N in extracts, a specific volume of extract was dried at 40 °C under vacuum, and a final weight was recorded prior to analysis as described above.

Amino acid analysis

The amino acid content and profiles were determined for the whole biomass, as well as the three fractions obtained from solvent fractionation. Approximately 2–5 mg biomass was hydrolyzed in 1 mL of 6 M HCl for 24 h at 110 °C in a digital hot block. After heating, 0.2 mL of the hydrolyzed sample was transferred to a separate vial and evaporated under a gentle flow of nitrogen. The dried hydrolysate was resuspended in 0.2 mL of 0.1 M HCl, and an internal standard of norvaline (10 µL) was added to the solution and mixed well using a pipette. The solution was transferred into a vial with an insert for HPLC analysis.

Amino acids were quantified by *o*-phthalaldehyde (OPA) and 9-fluorenylmethyl chloroformate (FOMC) derivatization method based on a previously reported method (Henderson et al. 2000). Briefly, 2.5 µL of borate buffer (0.4 M, Agilent 5061–3339) was mixed with 0.5 µL of sample, and then, 0.5 µL of OPA solution (Agilent 5061–3335) and 0.5 µL FOMC solution (Agilent 5061–3337) were added. The mixture was analyzed by HPLC (Agilent 1100) coupled with a DAD detector at 338 nm (for OPA) and 262 nm (for FOMC). The column was a Zorbax Eclipse-AAA column (3.5 µm × 4.6 mm × 250 mm), which was kept at 40 °C. The flow rate was 2 mL min⁻¹. Mobile phase A was 40 mM NaH₂PO₄, pH = 7.8, and B was acetonitrile:methanol:water (45:45:10, v/v/v). The gradient was 100% solvent A for the first 1.9 min; a linear decrease of solvent A from 100–43% from 1.9 to 18.1 min; a linear decrease of solvent A from 43 to 0% from 18.6 to 22.3 min; solvent A back to 100% at 23.2 min and stops at 26 min. Quantification of 19 amino acids, excluding tryptophan, was based on integration of individual peaks in the chromatograms and quantified using a 5-point calibration curve (10–1000 pmol L⁻¹). The individual amino acid concentrations were normalized based on the internal standard. The respective nitrogen-to-protein conversion factor was calculated for each of the fractions using previously described calculations (Lourenço et al. 2004; Templeton and Laurens 2015).

Nucleic acid analysis

A two-step nucleic acid procedure was adapted from the literature (Schneider 1945). In the first step, 50–100 mg of biomass was weighed into a 50 mL centrifuge tube. In accordance with the 1:3 ratio of DNA:RNA, 2 mg of DNA standard (Sigma D1626-250MG) and 6 mg of RNA standard (Sigma R6625-25G) were placed in a 50 mL centrifuge tube, and run with and without the addition of a protein standard spike (BSA, Sigma A7906-500G) to account for possible interferences. A reagent blank was included with all samples. Trichloroacetic acid (10 mL, 10%, v/v) was added to each tube; tubes were then vortexed and placed in an ice bath

for 10 min. After centrifugation at 8721 rcf for 20 min, the supernatant was decanted. Each precipitate was then washed twice with hot ethanol (90 °C). In the second step, 15 mL of 5% TCA was added to each sample and vortexed. The caps were tightly close and then twisted back a quarter turn in order to relieve pressure from heating. The samples were placed in a hot water bath at 90 °C for 25 min to hydrolyze the residual nucleic acid and vortexed halfway through incubation time. After 25 min, the samples were vortexed, cooled to room temperature, and then refrigerated at 4 °C before being centrifuged at 8721 rcf for 20 min at 0 °C. Finally, absorption measurements were taken at 260 and 280 nm on a Beckman Coulter DV 800 spectrophotometer, equipped with a six-position automated cell holder, and 10 mm path length, 1.0 mL matched quartz cuvettes (Fisher Scientific 50–823–023).

Sterol and phytol analysis

The sterol content of the whole biomass was determined as free sterols and phytol using a modification of the method described before (Ahmed et al. 2015). In brief, 10–12 mg of biomass was weighed out, and a known amount of an α -cholestane internal standard was added. The samples were then hydrolyzed with 0.5 mL methanolic KOH (10%, w/v) at 75 °C for 2 h. After hydrolysis, samples were cooled for at least 15 min before 0.2 mL of 0.9% NaCl (w/v) and 0.85 mL of n-hexane was added to the vial. The vial was vortexed and allowed to sit for at least 10 min. An aliquot of 0.7 mL of the hexane fraction was removed and evaporated in a separate vial under a gentle stream of nitrogen. Two more rounds of extraction were performed by the addition of 0.85 mL of hexane and subsequent removal and evaporation of 0.7 mL aliquots (into the same vial as the first extraction). If an emulsion formed during extraction, the vials were centrifuged at 930 rcf prior to removing the hexane layer. The dried extract was dissolved in 0.3 mL of chloroform, and 75 μ L was transferred to a new insert vial, to which 25 μ L of BSTFA (1% TMCS):pyridine (1:1, v/v) was added to derivatize the sterols for quantification on a GC-FID using the method described in the laboratory analytical procedure

for sterols available from NREL (Van Wychen and Laurens 2018).

Results

Baseline mass balance of whole algae biomass

The baseline compositional analysis of lyophilized whole biomass from three strains, representing high protein biomass harvested early in the growth phase, is shown in Table 2. The mass balance metric reflects the sum of the individually measured components in the biomass (*i.e.*, the sum of ash, carbohydrates, protein as amino acids, and lipids as FAME). Each component was measured as described above. Mass balances for the species reported in Table 2 are consistent with values from our previous work (Table 1) for high protein/early harvest samples, with approximately 25–35% of the biomass unaccounted for.

Application of cell disruption and fractionation method

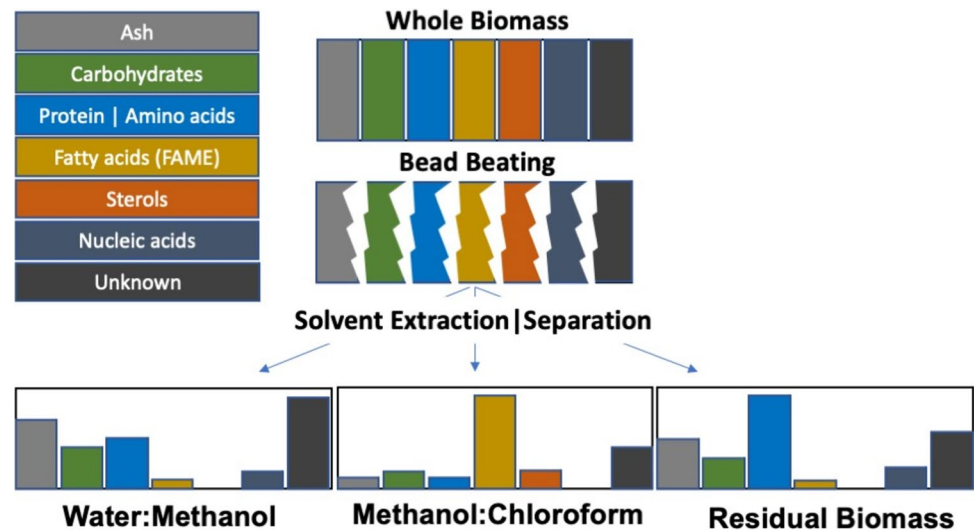
Three common cell disruption and extraction techniques were investigated to determine which yielded the best results. The three techniques used were high pressure (ASE extraction), sonication, and bead-beating, all followed by organic solvent extraction. Each technique was evaluated with respect to the mass recovery from extraction and completeness of extraction based on the residual FAME left in the residual solid biomass after extraction. The results from the comparisons, as well as the methods for ASE and sonication extractions, are found in the Supplemental Material. The gravimetric yields from the three extractions are shown in Table S1 and all showed mass balances near 100%. The mass balance of the fractions revealed greater yields in the residual fraction for ASE extraction when compared to sonication and bead-beating. The residual FAME analysis was conducted by applying the same *in situ* FAME procedure described above on the dried residual biomass and was used as a metric to determine the completeness of lipid extraction (Table S2). Although fatty acid methyl esters were detected

Table 2 Measured composition (baseline) of whole biomass (%DW) for 3 algae species

Species	Ash	Carbohydrates	Protein (amino acids)	Lipids (FAME)	Mass balance
<i>Monoraphidium</i> sp.	8.4 ± 0.0	20.0 ± 0.4	32.8 ± 0.7	9.3 ± 0.1	70.5 ± 1.1
<i>Nannochloropsis</i> sp.	17.6 ± 0.1	8.3 ± 0.4	31.6 ± 0.5	9.6 ± 0.1	67.1 ± 1.0
<i>Scenedesmus</i> sp.	17.8 ± 0.2	9.0 ± 0.3	35.7 ± 0.8	5.4 ± 0.0	67.9 ± 1.1

All data are expressed as a percentage of the oven dried biomass weight and are reported as an average of triplicate analyses (except for carbohydrates which was analyzed in duplicate) with associated standard deviation. Mass balance is calculated as the sum of the measured ash, carbohydrate, amino acid, and lipid (FAME) content and reported with the square root of the combined variances for each of the analyses

Fig. 1 Schematic of expected component fractionation after cell disruption, using bead beating as an example, and chosen analyses for comparing mass balances between whole biomass and the post-extraction fractions; the sum of the constituent composition of each of the fractions is being compared against the composition of the whole biomass



in the residual fractions for all extractions, the amount of FAME in the ASE extraction was considerably greater (15–31%) when compared to sonication (3–17%) and bead-beating (2–10%). For this reason, we chose bead-beating as the cell disruption method for the work discussed below.

Mass balance by fractionation

In an effort to reduce the biomass complexity and to understand the characteristics of the greater than 25% unknown fraction, we employed a sequential extraction approach. This approach generated three fractions based on solubility. In theory, the water:methanol fraction contains the most polar components, such as free salts, polar

metabolites, and small peptides. The methanol:chloroform fraction should contain primarily lipids and pigments, and the residual biomass that was not soluble in either solvent systems is likely comprised of ash, nucleic acids, protein, and carbohydrates. An example diagram of the component fractionation and respective fraction quantification is illustrated in Fig. 1. However, the fractions are not as discrete as described above, and each of the respective primary measured components and the unknown components are found in all of the fractions (Table 3). These results indicate that the unaccounted-for mass is likely comprised of more than one constituent, which was not entirely unexpected given the complexity of compounds in algal biomass. This methodology was designed based on

Table 3 Measured composition (% DW) of the whole biomass and respective water:methanol, methanol:chloroform, and residual fraction

	Ash	FAME	Carbohydrates	Protein (AA)	Nucleic acids ^a	Sterol ^b	Chlorophyll ^c	Unknown
<i>Monoraphidium</i>	8.37 ± 0.05	9.34 ± 0.07	20.03 ± 0.39	32.79 ± 0.75	4.24	0.71 ± 0.01	1.74 ± 0.03	22.78
H ₂ O:CH ₃ OH	2.84 ± 0.01	0.35 ± 0.01	1.86 ± 0.06	1.70 ± 0.04	-	-	-	4.93 ± 0.15
CH ₃ OH: CHCl ₃	0.58 ± 0.04	9.01 ± 0.26	1.39 ± 0.06	0.93 ± 0.16	-	0.71 ± 0.01	1.74 ± 0.03	6.30 ± 0.43
Residual	4.22 ± 0.37	0.18 ± 0.01	15.60 ± 0.35	28.77 ± 0.50	4.24	-	-	12.60 ± 1.31
<i>Nannochloropsis</i>	17.65 ± 0.07	9.60 ± 0.10	8.35 ± 0.44	31.56 ± 0.48	4.28	0.45 ± 0.01	1.65 ± 0.02	26.46
H ₂ O:CH ₃ OH	6.33 ± 0.03	0.19 ± 0.00	0.47 ± 0.01	3.53 ± 0.14	-	-	-	5.83 ± 0.16
CH ₃ OH: CHCl ₃	1.11 ± 0.04	9.29 ± 0.04	1.13 ± 0.04	0.84 ± 0.11	-	0.45 ± 0.01	1.65 ± 0.02	7.63 ± 0.16
Residual	10.03 ± 0.15	0.38 ± 0.00	6.08 ± 0.07	25.41 ± 0.51	4.28	-	-	13.21 ± 0.60
<i>Scenedesmus</i>	17.76 ± 0.25	5.45 ± 0.03	8.97 ± 0.28	35.68 ± 0.77	4.93	0.51 ± 0.03	2.17 ± 0.10	24.53
H ₂ O:CH ₃ OH	1.28 ± 0.02	0.28 ± 0.01	0.85 ± 0.03	2.92 ± 0.07	-	-	-	4.84 ± 0.06
CH ₃ OH: CHCl ₃	0.28 ± 0.04	3.87 ± 0.19	0.50 ± 0.02	1.13 ± 0.13	-	0.51 ± 0.03	2.17 ± 0.10	2.87 ± 0.23
Residual	14.57 ± 0.88	0.89 ± 0.02	7.17 ± 0.15	31.69 ± 1.01	4.93	-	-	16.82 ± 1.47

Reported analytical values are an average and standard deviation for the triplicate fractionation extractions performed on the bead-beaten material for each species. Nucleic acids were determined on a single whole biomass sample for each species. All values are expressed on a whole biomass dry weight basis. ^aNucleic acids were only quantified for the whole sample (due to limited fraction material). For this purpose, we assumed that all nucleic acids remained in the residual fraction. ^bSterols were only quantified for the whole sample (due to limited fraction material). Based on the hydrophobic nature of sterols, we assumed that all sterols were in the nonpolar fraction. ^cChlorophyll was only quantified for the whole sample (due to limited fraction material). Based on the fraction colors, we assumed that all chlorophyll was in the nonpolar fraction

component solubility, with the goal of determining if one fraction accounted for the bulk of the unknown material. However, insight into the solubility of unknown components may provide some indication as to the composition of these unaccounted-for compound classes. Table 3 shows the totals from all analyses, for each species, subdivided into solubility fractions. The measurements made on each fraction were ash, FAME, carbohydrates, and protein as amino acids, as well as the percentage of each fraction that was unaccounted for in the analysis. Table 4 shows the sum of the biomass that was accounted for by the above analyses, as well as the gravimetric results that correspond to each fraction. The sum of the fraction's measured components does not equal that of the whole sample because the value for the whole sample was directly measured and was not the sum of the fractions. The nitrogen content was also measured in each fraction to determine the non-protein nitrogen content (Table 5). The nucleic acids, sterols, and chlorophyll were all measured for the whole biomass only (due to limited fraction mass); however, we have assumed that nucleic acids will all be found in the residual biomass in the form of DNA and that the chlorophyll and sterols would end up in the methanol:chloroform fraction. In the case of chlorophyll, due to visual validation (i.e., the residual biomass was grey for *Monoraphidium* and *Nannochloropsis* and was mostly grey with a slight green tint for *Scenedesmus*), we believe this to be a good assumption.

Table 4 Sum of the measured components (% DW) from Table 3 for the whole biomass and water:methanol, methanol:chloroform, and residual fractions and total fraction weights as fractions of the whole biomass

	Measured component sum ^a	Gravimetric yield as fraction of whole biomass
<i>Monoraphidium</i>	77.22 ± 1.14	
H ₂ O:CH ₃ OH	6.75 ± 0.09	11.68 ± 0.19
CH ₃ OH: CHCl ₃	14.36 ± 0.07	20.66 ± 0.43
Residual	53.01 ± 0.89	65.50 ± 0.42
<i>Nannochloropsis</i>	73.54 ± 1.06	
H ₂ O:CH ₃ OH	10.52 ± 0.13	16.34 ± 0.16
CH ₃ OH: CHCl ₃	14.47 ± 0.13	22.10 ± 0.29
Residual	46.18 ± 0.60	59.39 ± 0.31
<i>Scenedesmus</i>	75.47 ± 1.21	
H ₂ O:CH ₃ OH	5.33 ± 0.07	10.17 ± 0.13
CH ₃ OH: CHCl ₃	8.46 ± 0.07	11.33 ± 0.30
Residual	59.25 ± 1.53	76.07 ± 0.09

^aMeasured component sum represents the sum of the components measured in the fractions, expressed as a % of the whole biomass

Table 5 Total nitrogen content, expressed on a whole biomass basis (% DW), calculated N-to-protein conversion factors, and non-protein nitrogen (%DW) for the whole biomass and each fraction

	Total N	Nitrogen-to-protein factor	Non-protein N
<i>Monoraphidium</i>			
Whole algae	7.20	4.6 ± 0.1	25.77 ± 1.77
H ₂ O:MeOH	0.42 ± 0.01	4.1 ± 0.1	33.37 ± 1.86
Chl:MeOH	0.29 ± 0.02	3.2 ± 0.4	49.08 ± 6.05
Residual	5.98 ± 0.03	4.8 ± 0.1	21.51 ± 1.34
<i>Nannochloropsis</i>			
Whole algae	6.68	4.7 ± 0.1	23.95 ± 1.27
H ₂ O:MeOH	0.66 ± 0.01	5.3 ± 0.3	20.66 ± 3.55
Chl:MeOH	0.32 ± 0.01	2.6 ± 0.2	60.15 ± 3.78
Residual	5.36 ± 0.01	4.7 ± 0.1	22.49 ± 1.8
<i>Scenedesmus</i>			
Whole algae	8.64	4.1 ± 0.1	32.43 ± 1.44
H ₂ O:MeOH	0.88 ± 0.00	3.3 ± 0.1	42.13 ± 1.23
Chl:MeOH	0.32 ± 0.04	3.5 ± 0.1	44.05 ± 0.7
Residual	7.3 ± 0.26	4.3 ± 0.3	29.15 ± 4.52

Nitrogen content (N) is measured by combustion and expressed on a dry weight (DW) basis, whereas non-protein nitrogen is expressed as the difference between the measured amino acid content in each fraction relative to the total nitrogen contribution in that fraction

Carbohydrate composition across fractions

Monosaccharide composition as a function of the solubility fraction gives us some understanding into the carbohydrates present. Figure 2 shows the distribution of monosaccharides, determined by acid hydrolysis with chromatographic analysis, for each species and fraction where each panel is normalized to total 100%. We observe that the methanol:chloroform fraction primarily yields galactose, likely due to the high concentration of mono- and digalactosyl lipids extracted. We also observe a significant amount of glucose, mannose, and, in the case of *Scenedesmus*, ribose.

Amino acids, nucleic acids, and non-proteinaceous nitrogen

For each fraction, as well as the whole biomass, the protein content was quantified as the sum of the individual acid hydrolyzed amino acids and the total nitrogen content was determined by combustion. The protein content as amino acids and the N by combustion data for the whole biomass and fractions are presented in Table 3. The majority of the protein from the whole biomass was observed in the residual fraction (80–88%), and only a small fraction was detected in the water:methanol and methanol:chloroform fractions. We used the individual amino acid concentrations, and their respective N contents, to determine the N associated with

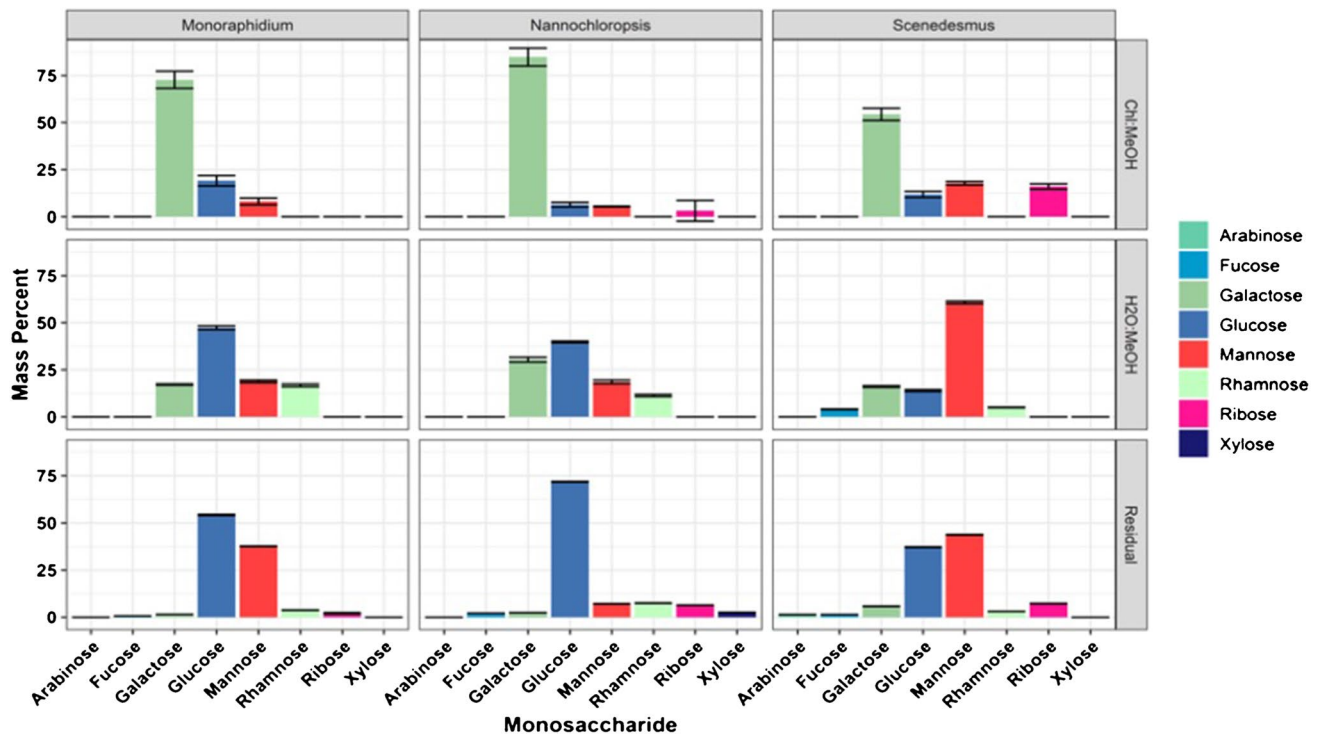


Fig. 2 Distribution of carbohydrate monomer composition within each solubility fraction. Rows correspond to the individual fractions, and columns correspond to algae species. For each subplot, the sum of all bars is normalized to 100%

protein. The % N, associated with the amino acids and the total nitrogen determined by combustion, allowed us to calculate the nitrogen-to-protein conversion factors for the whole biomass and each fraction (Table 5). The range of calculated conversion factors illustrate that the whole biomass and each fraction contain different ratios of protein and non-proteinaceous nitrogen. For the whole biomass, the averaged factor is 4.47 ± 0.31 across species, which is consistent with the averaged factor reported in the literature (Lourenço et al. 2004). The variation shown between the different species and fractions supports our approach of using amino acid data to calculate the respective protein content. As with any calculation of such factors, there are assumptions and we applied the factor calculation methodology previously reported based on the N associated with the free amino acid concentration related to the total measured nitrogen content (Lourenço et al. 2004; Templeton and Laurens 2015). As a percentage of the whole biomass, the non-protein-nitrogen content calculated using the data in Table 5 is the greatest in the residual fraction where the majority of protein also resides (5–7%, expressed on a whole biomass basis). It was observed that 21–29% of the nitrogen in the residues (based on the residual fraction weight) were from non-protein sources, likely attributed to the nucleic acids almost exclusively confined to the residual fractions. The N concentrations measured in the water:methanol

and methanol:chloroform fractions were considerably lower (0.3–0.9%, expressed on a whole biomass basis), and between 20 and 60% of that nitrogen was attributed to sources other than protein.

Lipid analysis

The lipid quantification method utilized here is based on a previously published *in situ* FAME method (Van Wychen et al. 2015). This type of analysis only accounts for the acyl portion of a lipid. However, in early growth phase (high-protein) algae, there are a wide range of polar lipids that make up the majority of the lipids present. The FAME content for the whole biomass and each fraction is shown in Table 3, and the gravimetric yields for each fraction as a % of the whole biomass are shown in Table 4. Of the whole biomass, 11–22% is extracted in the methanol:chloroform solvent extraction, and approximately half of that mass is accounted for as FAME for *Monoraphidium* sp. and *Nannochloropsis* sp., whereas for the *Scenedesmus* sp., only 35% of the methanol:chloroform extract is explained as FAME.

To understand the distribution of polar and non-polar lipids, a chromatographic separation was utilized to separate the lipids into 5 fractions. Figure 3 shows the mass distribution from gravimetric analysis for each chromatographic fraction. For each sample, the mass recovery is the

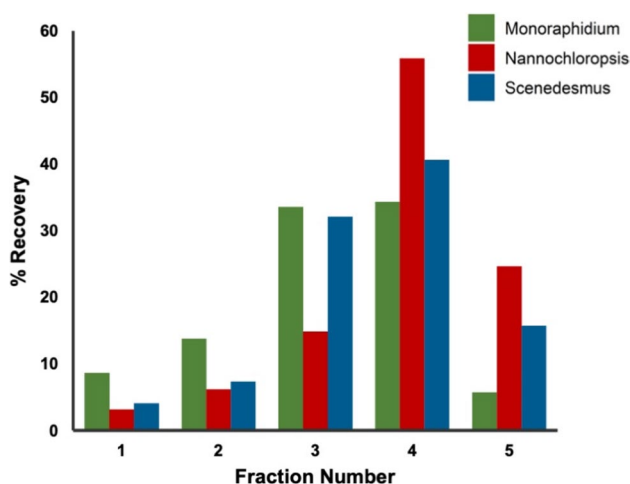


Fig. 3 Mass distribution of the five methanol:chloroform solubility fractions collected from silica gel separation. Each bar represents the percentage of material that eluted in a given fraction based on the initial sample loading

percentage of the original sample mass that was loaded onto the silica gel column and therefore represents the composition of the methanol:chloroform fraction. The fractions were analyzed by ultrahigh-resolution mass spectrometry to determine the lipid classes present in each fraction. The first and second fractions contained non-polar lipids (i.e., wax esters and carotenoids in fraction 1 and TAG lipids in fraction 2). Fraction 3 was comprised primarily of chlorophyll, mono- and diacylglycerols, and phospholipids, whereas fractions 4 and 5 were predominantly glycolipids and phosphosphingolipids (data not shown). We see that only between 9 and 22% of the lipids extracted in the methanol:chloroform fraction are found in the non-polar fractions. The highest concentration of lipids is observed in the polar fractions, which indicates that determination of lipid concentration by FAME analysis underestimates the total mass of lipids in the sample.

Discussion

Importance of biochemical context for compositional analysis

There are caveats to most of the analytical methods utilized for microalgae, some of which are the result of the biochemical shifts in the composition of algae due to changes in environmental conditions. It has been shown that strain selection, growth phase, and cultivation parameters, such as light and available nutrients, lead to vastly different biomass macromolecular makeups (Becker 2007; Laurens et al. 2012b, 2014; Finkel et al. 2016). A comprehensive review of

published composition data in algae resulted in a collection of 1500 data points for a range of algae species (Finkel et al. 2016). The data were tabulated, and median composition values were determined to be 32.2% protein, 17.3% lipid, 15.0% carbohydrate, 17.3% ash, 5.7% RNA, 1.1% chlorophyll-*a*, and 1.0% DNA, reported as percent dry weight for the active growth phase. This study also revealed that variability based on taxonomy was observed in the microalgae composition data. The analysis of data derived from the literature has great potential and value for understanding trends for microalgae composition; however, several works have illustrated the variability in biomass composition (Gatenby et al. 2003; Finkel et al. 2016; Bernaerts et al. 2018).

Caveats and limitations to analytical methodologies

Quantitative determination of protein, a major biochemical constituent in microalgae, is particularly challenging due to the fact that protein constituents are typically present as free amino acids, peptides, proteins, and protein complexes (with sugars and/or lipids) and, depending on their biochemical function, range from highly hydrophobic to highly hydrophilic (Gatenby et al. 2003; Huo et al. 2011; Safi et al. 2012; Bleakley and Hayes 2017). Protein is commonly calculated from measured elemental nitrogen content determined by combustion or Kjeldahl procedures, where the nitrogen from these methods is converted to protein content using a conversion factor. Although these measurements are both robust and accurate with respect to nitrogen content, published research illustrates that nitrogen-to-protein conversion factors are variable, sometimes up to approximately twofold, between species and growth conditions (Lourenço et al. 2004; Templeton and Laurens 2015). Typically, protein quantification based on amino acid content yields more accurate results compared with the use of a traditional nitrogen-to-protein conversion of 6.25, known to not be appropriate for microalgae due to the presence of non-proteinaceous nitrogen. A more appropriate factor of 4.78 or 4.97 has been published in the literature (Lourenço et al. 2004; Templeton and Laurens 2015). This work utilized an acid hydrolysis to measure the content of amino acids present. This approach, coupled with measuring the total nitrogen for each fraction, allows us to understand the amount of nitrogen that corresponds to protein for each solubility fraction. In the fractionation approach used here, the protein and non-protein-derived nitrogen ratios used for the whole biomass may not apply to each of the solubility fractions; therefore, a constant nitrogen-to-protein conversion factor is not applicable. Table 5 shows that nitrogen-to-protein factors for the fractions range from 2.6 to 5.3.

For the purpose of determining the most accurate protein content, we utilized an amino acid analysis to determine protein content in the samples and the fractions, rather than

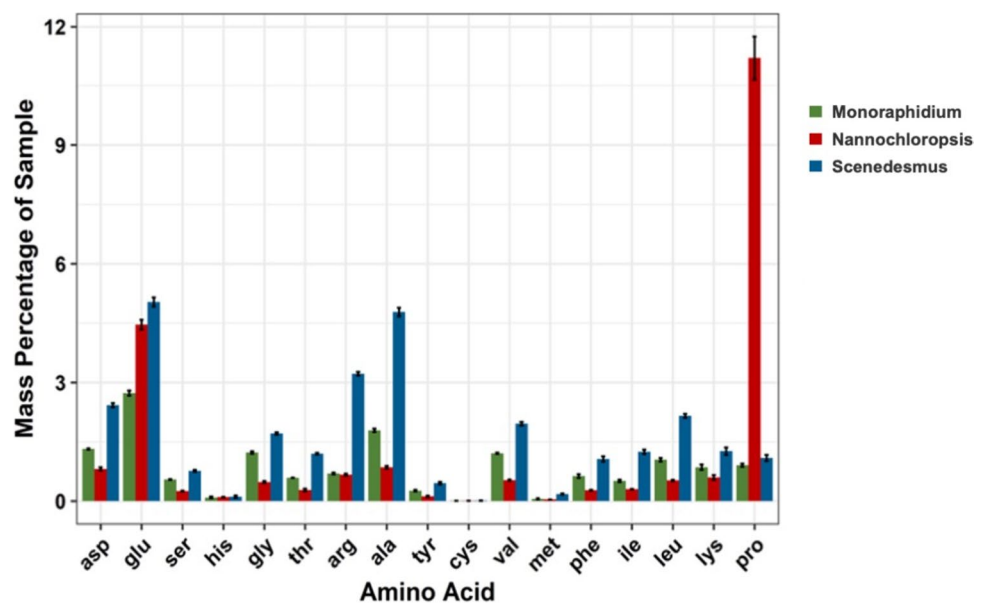
the more traditional calculation of protein from nitrogen content using a conversion factor. Direct quantification of protein by amino acid analysis provides increased detail in the chemical composition of the algae, by providing the amino acid distribution in both the whole algae and the fractions and, in general, is a more accurate reflection of protein content than using a nitrogen conversion factor. The methanol:chloroform fraction, as well as the residual biomass, yielded similar amino acid profiles. However, the water:methanol fraction revealed that *Nannochloropsis* contains a significantly greater concentration of proline compared to the other two species studied here (Fig. 4). It has been well documented that proline accumulation is linked to stress conditions within plants and algae (Vanlerberghe and Brown 1987; Siripornadulsil et al. 2002; Hayat et al. 2012). Here, we utilize a technique that provides a more accurate quantification of protein, and also gain some insight into the physiological conditions of the *Nannochloropsis* sample that would otherwise be lost from protein determination by nitrogen content.

By comparing the N content measured by combustion for each fraction to the total N calculated with the measured amino acid content, we are able to determine the distribution of N-containing compounds in each of the fractions, separated into protein and non-protein nitrogen (NPN) species. Nitrogen from each amino acid was quantified based on the % N for each amino acid, e.g., for proline nitrogen is 12.3% of the structure of the total amino acid molecular weight. Some of the potential sources of non-protein nitrogen include nucleic acids, chlorophyll, and nitrogen-containing lipids, metabolites, and inorganic N species (including remnants of the nutrients in the cultivation media), which present themselves across all three of the fractions studied

here. We see in Table 5 that the percentage of NPN changes drastically between fractions and accounts for 45–60% of the nitrogen in the methanol:chloroform fraction, as opposed to only 21–29% of the residual fraction. For this reason, it is necessary to use direct protein analysis in the form of amino acid analysis for samples that have been fractionated, either by extraction or perhaps hydrolysis, rather than a nitrogen-to-protein conversion factor.

Carbohydrates in algae are present as storage or structural polymers, and occasionally as free sugars or osmoprotectants in some salt-water species (Rosell and Srivastava 1984; Reboloso-Fuentes et al. 2001; Scholz et al. 2014). The monomeric units consist of amino sugars, sugar alcohols, uronic acids, and neutral sugars, forming a myriad of polymeric sugar structures and complexes with lipids and proteins. Carbohydrates can be quantified spectrophotometrically or through chemical or enzymatic hydrolysis and chromatographic detection of the monomeric components. (Rao and Pattabiraman 1989; Masuko et al. 2005; Templeton et al. 2012; Van Wychen et al. 2017) However, spectrophotometric methods are prone to interferences and are often non-specific regarding the composition of carbohydrates, introducing potential inaccuracies due to using a single sugar as a calibration standard (Rahman and Richards 1987; Rao and Pattabiraman 1989). In some instances, as per nutritional labeling regulations, carbohydrates are calculated as the difference between the sum of protein, ash, and crude fat and an assumed mass closure of 100% (Lane et al. 2021). This approach not only introduces skewed data that assumes all other components were accounted for but also incorporates error from all other measurements into the final carbohydrate calculation. Compounded errors such as those observed from carbohydrates calculated by difference

Fig. 4 The distribution of amino acids present in the water:methanol fraction. Many of these species are likely present initially as free amino acid metabolites



may lead to significant overestimations of bioproduct potential which could prove problematic when considering the associated biomass value and corresponding economics are often heavily based on the compositional profile. The direct determination of sugar content, as reported here, yields a significantly lower carbohydrate content when compared to the calculation of carbohydrates by difference. This indicates that the calculation of carbohydrates by difference may have negative effects on processes such as the prediction of bioethanol yields during fermentation.

The distribution of monosaccharides as a function of solubility also provides some insight into the composition of carbohydrates. Figure 2 shows that glucose and mannose are the primary monomers in the residual fraction, whereas the water:methanol fraction shows appreciable concentrations of galactose and rhamnose. This may indicate that glucose and mannose are preferentially incorporated into macromolecular structures within the microalgae. The monosaccharide distribution for the methanol:chloroform fraction is composed primarily of galactose, likely due to the presence of mono- and digalactosyl lipids. We also observe a significant amount of glucose, mannose, and, in the case of *Scenedesmus*, ribose. These sugars are not known to exist in algal lipids and may be present due to co-extraction of carbohydrates or lipid/carbohydrate complexes. Although these components are present in low abundance, this does highlight the fact that quantitation of a compound class, such as lipids, by gravimetric analysis of a solubility fraction is not ideal due to the potential of co-extracted species of different biochemical families.

Lipids are a chemically diverse group of components, consisting of polar and non-polar molecules which include glycolipids, phospholipids, sphingolipids, triacylglycerols, and pigments, and their contributions to single celled organisms, such as microalgae, are critical for physiological functions and cellular integrity (Fahy et al. 2005; Borowitzka 2013; Wu et al. 2014; Yao et al. 2015; Dong et al. 2016b). Lipid detection is generally less prone to interferences when compared to carbohydrates; however, the accurate quantification of lipids has its own nuances. Determining intact lipids based on lipid classes in algae can prove challenging on a number of fronts, including incomplete extraction, poor separation of lipid types, and a lack of quantitative lipid standards that cover the wide range of lipid classes prohibiting the respective quantitative determination. Therefore, lipid content is typically determined either by acid or base hydrolysis followed by solvent extraction of intact lipids and then a gravimetric weight, by solvent extraction alone with gravimetric analysis (Bligh and Dyer 1959), or by *in situ* transesterification to fatty acid methyl esters (FAME) (Laurens et al. 2012a; Ryckebosch et al. 2013, 2014; Dong et al. 2015). These methods can yield very different results, especially if the gravimetric analysis co-extracts components

such as metabolites and non-polar peptides. The data presented here suggests that small amounts of both carbohydrates and proteins/peptides are in fact present as co-extractants in both the water:methanol and methanol:chloroform fractions. For food analysis, it is common to use an acid or base hydrolysis method followed by extraction of the fatty acids with organic solvent and gravimetric measurement. It is worth pointing out that the acid/base treatment of lipids prior to extraction cleaves the polar head groups, which are unaccounted for in the overall lipid mass percentage and FAME analysis may also neglect portions of non-polar lipids such as wax esters, hydrophobic metabolites, and pigments (e.g., carotenoids) typically found in the unsaponifiable fraction of the lipids, and large, polar, headgroups, on, e.g., phosphor- and sphingolipids. This point is relevant to the current work as it is known that early growth phase (high-protein) algae possesses a wide range of polar lipids that make up the majority of the lipids present. However, FAME analysis was chosen here as a conservative, and least ambiguous measure of the lipid fraction of the biomass, as well as most likely to not overestimate the amount of lipid present. Additionally, some of the lipid constituents not accounted for as FAME are captured with the measurement of carbohydrates, chlorophyll, ash, and potentially protein (as in the case of lipoproteins). The carbohydrate measured in the methanol:chloroform fraction is likely due to glycolipid head groups, e.g. galactose in galactolipids, and we assume that at least a portion of the ash is comprised of phosphate from phospholipids, as we have observed full recovery of the mass of polyphosphates after ashing (NREL unpublished data). Unsaponifiables on the methanol:chloroform fractions were determined following previously documented procedures (Ahmed et al. 2015) and across species, were remarkably consistent, between 32.8 and 34.6% of the fraction (data not shown). While the unsaponifiables may potentially describe the majority of the missing mass in the methanol:chloroform fraction, the gravimetric determination combined with the molecular complexity of the unsaps could potentially lead to double counting, so we chose to not include this component in the mass balance accounting. Our decision is further supported by results from the literature which indicated 70% of the unsaponifiables from *Nannochloropsis* were unknown and the risks of counting the non-hydrolyzable biopolymer algaenan as part of the unsaponifiables fraction may be misleading (Wang and Wang 2011). Despite accounting for some lipid headgroups indirectly, there still remains a large portion of the methanol:chloroform fraction that is unaccounted for and will be the focus of ongoing detailed lipidomics mass spectrometry work in our laboratory.

Mass balance by fractionation

It is commonly shown that conventional direct analytical techniques do not allow for closing the mass balance for microalgae, specifically in the early growth phase. Using common analytical techniques, across a wide range of species, only 70–80% of the biomass is accounted for (Bernaerts et al. 2018). By utilizing a fractionation approach, we aimed to identify the fraction where mass closure could not be completed and thereby gain information into the composition of the unidentified material. When considering the sum of all fractions, there was not just one fraction in which there was unidentified material, but in fact, the unidentified material was spread across all of the solubility fractions. This outcome indicates there are likely several types or classes of analytes that are not being measured. Expressed on a whole biomass basis, there was ~5% unknown in the water:methanol fraction, 3–6% unknown in the methanol:chloroform fraction, and 12–17% unknown in the residual biomass. Although this approach did not result in an increase in the percentage of biomass identified, it does reveal several hypotheses about the composition of unknown compounds and provides useful insight into the direction of future analyses. These points are discussed in greater detail below.

Potential composition of unidentified components

There remains value in understanding the solubility characteristics of the unidentified portions of algal biomass. For instance, we know that analytes present in the water:methanol fraction will be polar/hydrophilic compounds. Since inorganic salts, carbohydrates, and amino acids were accounted for in this fraction, it is likely that the remaining unidentified compounds are comprised of a range of small polar metabolites. Metabolomics experiments have been previously conducted on algae species and reveal a broad range of potential metabolites (Kind et al. 2012; Werner et al. 2019), with some possibly making a substantial contribution to the soluble fractions created and discussed above. However, very little work has been conducted to quantify these metabolites in microalgae. Likewise, the methanol:chloroform fraction may contain compounds with a range in polarities. It is likely that, in addition to lipids, this fraction may also include some hydrophobic metabolites, such as organic vitamins, that are not captured via the macromolecular analyses performed in this work. We also did not account for carotenoids in this work, which may account for a portion of the unidentified components (Reboloso-Fuentes et al. 2001; Borowitzka 2013; Ryckebosch et al. 2014; Huang et al. 2018). The residual biomass contained the greatest portion of the unidentified components, ranging from 12 to 17% on a whole biomass basis (Table 3). It is unclear what the composition of this unknown fraction may

be. It is possible that the acid hydrolysis steps used to determine carbohydrate and protein quantities do not adequately break down all of the components present or that there is an interaction between the two groups. There may also be other components such as algaenan that are resistant to hydrolysis and may contribute to a portion of the residual fraction, up to 8% of the whole biomass (Allard and Templier 2000; Kodner et al. 2009; Scholz et al. 2014). Unfortunately, we do not yet have enough information to form a hypothesis as to the composition of the unknowns in the residual fraction, and future work will focus on understanding the composition of this material and include a quantitative recovery of the algaenan biopolymer.

Conclusion

The composition of algae is variable, both as a function of nutrients and environmental factors. Using analytical methodologies to account for protein, carbohydrates, lipids, and ash, greater than 90% of the biomass components could be identified and accounted for in high carbohydrate and high lipid algal biomass. However, in samples harvested from early growth phase cultures, only ~75% of components were accounted for using the same techniques. A fractionation approach was used to separate biomass components by solubility, in the attempt to isolate and therefore identify the characteristics of the unaccounted-for component. While this technique is not recommended for general compositional analysis of the biomass, it is a practical approach used here to reduce the complexity of the biomass for future research on the identities and subsequent development of quantification methods for the unknown components. After analysis of each fraction, it was determined that all fractions contained significant amounts of the unaccounted-for mass. The solubility properties of the water:methanol and methanol:chloroform fractions indicate that the approximately 7–13% of unidentified components in these fractions are potentially from metabolites and lipids that were not accounted for by ash, FAME, carbohydrates, and amino acid analysis. The residual fraction contained the greatest amount of unidentified material ranging from 12 to 17% on a whole biomass basis. The unidentified mass in the residual fraction may prove to be the most challenging to identify, as all we know about it currently is that it is insoluble in polar and non-polar solvents, or at the very least, difficult to extract. Future work will focus on the detection and identification of unknowns in all fractions, but specifically the residual fraction, with a final goal of quantification of all constituents for mass balance closure. Full characterization of early growth stage algae is critical to the valorization of algae components that may ultimately improve the economics of algae production.

Supplementary Information The online version contains supplementary material available at <https://doi.org/10.1007/s10811-021-02508-x>.

Acknowledgements This work was authored by the National Renewable Energy Laboratory. The US Government retains and the publisher, by accepting the article for publication, acknowledges that the US Government retains a nonexclusive, paid-up, irrevocable, worldwide license to publish or reproduce the published form of this work, or allow others to do so, for the US Government purposes.

Funding This work is funded by the US Department of Energy under Contract No. DE-AC36-08GO28308 with the National Renewable Energy Laboratory, as part of the DOE Office of Energy Efficiency and Renewable Energy, Bioenergy Technologies Office, WBS 1.2.3.001.

Data availability All data generated or analyzed during this study are included in this published article as supplementary information files. Supporting calculation background and information can be obtained from the corresponding author on reasonable request.

Declarations

Disclaimer The views expressed in the article do not necessarily represent the views of the DOE or the US Government.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

References

- Ahmed F, Zhou W, Schenk PM (2015) *Pavlova lutheri* is a high-level producer of phytosterols. *Algal Res* 10:210–217
- Allard B, Templier J (2000) Comparison of neutral lipid profile of various trilaminar outer cell wall (TLS)-containing microalgae with emphasis on algaenan occurrence. *Phytochemistry* 54:369–380
- Becker EW (2007) Micro-algae as a source of protein. *Biotechnol Adv* 25:207–210
- Bernaerts TMM, Gheysen L, Kyomugasho C, Jamsazzadeh Kermani Z, Vandionant S, Foubert I, Hendrickx ME, Van Loey AM (2018) Comparison of microalgal biomasses as functional food ingredients: focus on the composition of cell wall related polysaccharides. *Algal Res* 32:150–161
- Bleakley S, Hayes M (2017) Algal proteins: extraction, application, and challenges concerning production. *Foods* 6:33
- Bligh EG, Dyer WJ (1959) A rapid method of total lipid extraction and purification. *Can J Biochem Physiol* 37:911–917
- Borowitzka MA (2013) High-value products from microalgae-their development and commercialisation. *J Appl Phycol* 25:743–756
- Christaki E, Bonos E, Giannenas I, Florou-Paneria P (2013) Functional properties of carotenoids originating from algae. *J Sci Food Agric* 93:5–11
- Crowe B, Attalah S, Agrawal S, Waller P, Ryan R, Van Wagenen J, Chavis A, Kyndt J, Kacira M, Ogden KL, Huesemann M (2012) A comparison of *Nannochloropsis salina* growth performance in two outdoor pond designs: conventional raceways versus the ARID pond with superior temperature management. *Int J Chem Eng* 2012:920608
- Dong T, Knoshaug EP, Davis R, Laurens LML, Van Wychen S, Pienkos PT, Nagle N (2016) Combined algal processing: a novel integrated biorefinery process to produce algal biofuels and bioproducts. *Algal Res* 19:316–323
- Dong T, Knoshaug EP, Pienkos PT, Laurens LML (2016) Lipid recovery from wet oleaginous microbial biomass for biofuel production: a critical review. *Appl Energy* 177:879–895
- Dong T, Yu L, Gao D, Yu X, Miao C, Zheng Y, Lian J, Li T, Chen S (2015) Direct quantification of fatty acids in wet microalgal and yeast biomass via a rapid in situ fatty acid methyl ester derivatization approach. *Appl Microbiol Biotechnol* 99:10237–10247
- Fahy E, Subramaniam S, Brown HA, Glass CK, Merrill AH Jr, Murphy RC, Raetz CR, Russell DW, Seyama Y, Shaw W, Shimizu T, Spener F, van Meer G, VanNieuwenhze MS, White SH, Witztum JL, Dennis EA (2005) A comprehensive classification system for lipids. *J Lipid Res* 46:839–861
- Finkel ZV, Follows MJ, Liefer JD, Brown CM, Benner I, Irwin AJ (2016) Phylogenetic diversity in the macromolecular composition of microalgae. *PLoS ONE* 11:e0155977
- Food and Drugs, 21, Code of Federal Regulation, Title 21, volume 1, part 101, Food Labeling (2020)
- Gatenby CM, Orcutt DM, Kreeger DA, Parker BC, Jones VA, Neves RJ (2003) Biochemical composition of three algal species proposed as food for captive freshwater mussels. *J Appl Phycol* 15:1–11
- Hayat S, Hayat Q, Alyemeni MN, Wani AS, Pichtel J, Ahmad A (2012) Role of proline under changing environments: a review. *Plant Signal Behav* 7:1456–1466
- Henderson JW, Ricker RD, Bidlingmeyer B A, Woodward C (2000) Rapid, accurate, sensitive, and reproducible HPLC analysis of amino acids, Agilent Application Note 5980–1193E, <https://www.agilent.com/cs/library/chromatograms/59801193.pdf>
- Hess SK, Lepetit B, Kroth PG, Mecking S (2018) Production of chemicals from microalgae lipids – status and perspectives. *Eur J Lipid Sci Technol* 120:1–26
- Huang W, Lin Y, He M, Gong Y, Huang J (2018) Induced high-yield production of zeaxanthin, lutein, and β -carotene by a mutant of *Chlorella zofingiensis*. *J Agric Food Chem* 66:891–897
- Huesemann MH, Van Wagenen J, Miller T, Chavis A, Hobbs S, Crowe B (2013) A screening model to predict microalgal biomass growth in photobioreactors and raceway ponds. *Biotechnol Bioeng* 110:1583–1594
- Huo YX, Cho KM, Rivera JGL, Monte E, Shen CR, Yan Y, Liao JC (2011) Conversion of proteins into biofuels by engineering nitrogen flux. *Nat Biotechnol* 29:346–351
- Kind T, Meissen JK, Yang D, Nocito F, Vaiya A, Chen Y-C, Vandergheynst JS, Fiehn O (2012) Qualitative analysis of algal secretions with multiple mass spectrometric platforms. *J Chromatogr A* 1244:139–147
- Kodner RB, Summons RE, Knoll AH (2009) Phylogenetic investigation of the aliphatic, non-hydrolyzable biopolymer algaenan, with a focus on green algae. *Org Geochem* 40:854–862
- Lane M, Van Wychen S, Politis A, Laurens LML (2021) A data-driven comparison of commercially available testing methods for algae characterization. *Algal Res* 53:102134
- Laurens L, Quinn M, Van Wychen S, Templeton DW, Wolfrum EJ (2012) Accurate and reliable quantification of total microalgal fuel potential as fatty acid methyl esters by in situ transesterification. *Anal Bioanal Chem* 403:167–178
- Laurens LML, Dempster TA, Jones HDT, Wolfrum EJ, Van Wychen S, McAllister JSP, Rencenberger M, Parchert KJ, Gloe LM (2012)

- Algal biomass constituent analysis: method uncertainties and investigation of the underlying measuring chemistries. *Anal Chem* 84:1879–1887
- Laurens LML, Markham J, Templeton DW, Christensen ED, Van Wychen S, Vadelius EW, Chen-Glasser M, Dong T, Davis R, Pienkos PT (2017) Development of algae biorefinery concepts for biofuels and bioproducts; a perspective on process-compatible products and their impact on cost-reduction. *Energy Environ Sci* 10:1716–1738
- Laurens LML, Olstad JL, Templeton DW (2018) Total protein content determination of microalgal biomass by elemental nitrogen analysis and a dedicated nitrogen-to-protein conversion factor. In: Spilling K (ed) *Biofuel from algae*. Humana Press, NY, pp 233–242
- Laurens LML, Van Wychen S, McAllister JP, Arrowsmith S, Dempster TA, McGowen J, Pienkos PT (2014) Strain, biochemistry, and cultivation-dependent measurement variability of algal biomass composition. *Anal Biochem* 452:86–95
- Lourenço SO, Barbarino E, Lavín PL, Lanfer Marquez UM, Aida E (2004) Distribution of intracellular nitrogen in marine microalgae: calculation of new nitrogen-to-protein conversion factors. *Eur J Phycol* 39:17–32
- Masuko T, Minami A, Iwasaki N, Machima N, Nishimura S, Lee YC (2005) Carbohydrate analysis by a phenol-sulfuric acid method in microplate format. *Anal Biochem* 339:69–72
- Pleissner D, Lau KY, Zhang C, Lin CSK (2015) Plasticizer and surfactant formation from food-waste- and algal biomass-derived lipids. *Chemosuschem* 8:1686–1691
- Rahman MD, Richards GN (1987) Interference by flavonoids in the phenol-sulfuric acid analysis of carbohydrates. *Carbohydr Res* 112–115
- Rao P, Pattabiraman TN (1989) Reevaluation of the phenol-sulfuric acid reaction for the estimation of hexoses and pentoses. *Anal Biochem* 181:18–22
- Reboloso-Fuentes MM, Navarro-Pérez A, García-Camacho F, Ramos-Miras JJ, Guil-Guerrero JL (2001) Biomass nutrient profiles of the microalga *Nannochloropsis*. *J Agric Food Chem* 49:2966–2972
- Rosell KG, Srivastava LM (1984) Seasonal variation in the chemical constituents of the brown algae *Macrocystis integrifolia* and *Nereocystis luetkeana*. *Can J Bot* 62:2229–2236
- Ryckebosch E, Bermúdez SPC, Termote-Verhalle R, Bruneel C, Muylaert K, Parra-Saldivar R, Foubert I (2013) Influence of extraction solvent system on the extractability of lipid components from the biomass of *Nannochloropsis gaditana*. *J Appl Phycol* 26:1501–1510
- Ryckebosch E, Bruneel C, Termote-Verhalle R, Muylaert K, Foubert I (2014) Influence of extraction solvent system on extractability of lipid components from different microalgae species. *Algal Res* 3:36–43
- Safi C, Charton M, Pignolet O, Silvestre F, Vaca-García C, Pontalier P-Y (2012) Influence of microalgae cell wall characteristics on protein extractability and determination of nitrogen-to-protein conversion factors. *J Appl Phycol* 25:523–529
- Sathasivam R, Ki JS (2018) A review of the biological activities of microalgal carotenoids and their potential use in healthcare and cosmetic industries. *Mar Drugs* 16:26
- Sathasivam R, Radhakrishnan R, Hashem A, Abd Allah EF (2019) Microalgae metabolites: a rich source for food and medicine. *Saudi J Biol Sci* 26:709–722
- Schneider WC (1945) Phosphorus compounds in animal tissues: I. Extraction and estimation of desoxypentose nucleic acid and of pentose nucleic acid. *J Biol Chem* 161:293–303
- Scholz MJ, Weiss TL, Jinkerson RE, Jing J, Roth R, Goodenough U, Posewitz MC, Gerken HG (2014) Ultrastructure and composition of the *Nannochloropsis gaditana* cell wall. *Eukaryot Cell* 13:1450–1464
- Siripornadulsil S, Traina S, Verma DPS, Sayre RT (2002) Molecular mechanisms of proline-mediated tolerance to toxic heavy metals in transgenic microalgae. *Plant Cell* 14:2837–2847
- Templeton DW, Quinn M, Van Wychen S, Hyman D, Laurens LML (2012) Separation and quantification of microalgal carbohydrates. *J Chromatogr A* 1270:225–234
- Templeton DW, Laurens LML (2015) Nitrogen-to-protein conversion factors revisited for applications of microalgal biomass conversion to food, feed and fuel. *Algal Res* 11:359–367
- Van Wychen S, Laurens LML (2015a) Determination of total carbohydrates in algal biomass - Laboratory Analytical Procedure (LAP). Golden, CO, National Renewable Energy Laboratory. NREL/TP-5100-60957. <https://www.nrel.gov/docs/fy16osti/60957.pdf>
- Van Wychen S, Laurens LML (2018) Determination of total sterols in microalgae by acid hydrolysis and extraction: Laboratory Analytical Procedure (LAP), Golden, CO, National Renewable Energy Laboratory. NREL/TP-5100-72990. <https://www.nrel.gov/docs/fy19osti/72990.pdf>
- Van Wychen S, Laurens LML (2015b) Determination of total solids and ash in algal biomass: Laboratory Analytical Procedure (LAP), Golden, CO, National Renewable Energy Laboratory. NREL/TP-5100-60956. <https://www.nrel.gov/docs/fy16osti/60956.pdf>
- Van Wychen S, Long W, Black SK, Laurens LML (2017) MBTH: a novel approach to rapid, spectrophotometric quantitation of total algal carbohydrates. *Anal Biochem* 518:90–93
- Van Wychen S, Ramirez K, Laurens LML (2015) Determination of total lipids as Fatty Acid Methyl Esters (FAME) by in situ transesterification: Laboratory Analytical Procedure (LAP). Golden, CO, National Renewable Energy Laboratory. NREL/TP-5100-60958 <https://www.nrel.gov/docs/fy16osti/60958.pdf>
- Vanlerberghe GC, Brown LM (1987) Proline overproduction in cells of the green alga *Nannochloris bacillaris* resistant to azetidine-2-carboxylic acid. *Plant Cell Environ* 10:251–257
- Wang G, Wang T (2011) Characterization of lipid components in two microalgae for biofuel application. *J Am Oil Chem Soc* 89:135–143
- Werner A, Broeckling CD, Prasad A, Peebles CAM (2019) A comprehensive time-course metabolite profiling of the model cyanobacterium *Synechocystis* sp. PCC 6803 under diurnal light:dark cycles. *Plant J* 99:379–388
- Wu S, Zhang B, Huang A, Huan L, He L, Lin A, Niu J, Wang G (2014) Detection of intracellular neutral lipid content in the marine microalgae *Prorocentrum micans* and *Phaeodactylum tricorutum* using Nile red and BODIPY 505/515. *J Appl Phycol* 26:1659–1668
- Wynne MJ, Hallan JK (2016) Reinstatement of *Tetradismus* G. M. Smith (Sphaeropleales, Chlorophyta). *Feddes Repert* 126:83–86
- Yao L, Gerde JA, Lee S-L, Wang T, Harrata KA (2015) Microalgae lipid characterization. *J Agric Food Chem* 63:1773–1787

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.