Research Article

Received: 7 April 2023 Revised: 30 August 2023 Accepted article published: 7 September 2023 Published online in Wiley Online Library:

(wileyonlinelibrary.com) DOI 10.1002/jctb.7512

Qualitative and quantitative evaluation of microalgal biomass using portable attenuated total reflectance-Fourier transform infrared spectroscopy and machine learning analytics

Mais Sweiss,[a] Sulaf Assi,[b*] Lina Barhoumi,[c] Dhiya Al-Jumeily,[d] Megan Watson,[b] Megan Wilson,[b] Tom Arnot[e] and Rod Scott[f†]

Abstract

BACKGROUND: Using microalgae for wastewater treatment offers an environmentally friendly method to produce microalgal biomass that can be used for many applications. However, the biochemical characteristics of microalgal biomass vary from species to species, from strain to strain, and between different growth stages within the same species/strain. This study utilized portable attenuated total reflectance-Fourier transform infrared (ATR-FTIR) spectroscopy to determine the composition of freeze-dried microalgal biomass corresponding to eight different locally isolated microalgae and a reference strain that were grown in wastewater and then harvested at the log and stationary phases, respectively.

RESULTS: The results showed that the portable ATR-FTIR spectroscopy offered a rapid, non-destructive, and accurate technique for monitoring changes in the biochemical composition of algal biomass at stationary and log phases, as well as quantifying their main constituents. For qualitative analysis of species, two machine learning analytics (MLAs; correlation in wavenumber space and principal component analysis) were able to differentiate between microalgae isolates in both their stationary and log phases. For quantification, univariate or multivariate regression offered accuracy in quantifying key microalgal constituents related to proteins, lipids, and carbohydrates. In this sense, multivariate methods showed more accuracy for quantifying carbohydrates, yet proteins and lipids were more accurately quantified with univariate regression. Based on quantification, the highest relative content of carbohydrates in the log phase was for Jordan-23 (Jo-23; *Desmodesmus sp.)*, whereas the highest content in the stationary phase was that for Jordan-29 (Jo-29; *Desmodesmus sp*). Regarding the relative lipid content in the log phase, Jo-23 had the highest lipid content, while the highest content in the stationary phase was for Jo-29.

CONCLUSION: ATR-FTIR spectroscopy offered a rapid and sustainable method for monitoring the microalgal biomass produced during wastewater treatment processes.

Keywords: attenuated total reflectance Fourier transform infrared; biomass; microalgae; spectroscopy; wastewater; principal component analysis

* *Correspondence to: S Assi, School Pharmacy and Biomolecular Sciences, Liverpool John Moores University, Liverpool, Byrom Street, L3 3AF, UK. E-mail: s.assi@ljmu.ac.uk*

† *Sadly, Professor Scott passed away in December 2018.*

a *Department of Biotechnology, Faculty of Agricultural Technology, Al-Balqa Applied University, Al-Salt, Jordan*

b *School Pharmacy and Biomolecular Sciences, Liverpool John Moores University, Liverpool, UK*

c *Department of Chemistry, Faculty of Science, Al-Balqa Applied University, Al-Salt, Jordan*

d *School of Computer Science and Mathematics, Liverpool John Moores University, Liverpool, UK*

e *Water Research & Innovation Centre, Department of Chemical Engineering, University of Bath, Bath, UK*

f *Department of Biology and Biochemistry, University of Bath, Bath, UK*

NOMENCLATURE

ATR-FTIR	Attenuated total reflectance-Fourier transform infrared
BSA	Bovine serum albumin
CCAP	collection of algae and protozoa
GTP	Glycerol tripalmitatete
Jo-2	Jordan_2 (*Desmodesmus sp.*)
Jo-4	Jordan_4 (*Desmodesmus sp.*)
Jo-12	Jordan_12 (*Coelastrella sp.*)
Jo-18	Jordan_18 (*Desmodesmus subspicatus* – Chodat – Hegewald and Schmidt.)
Jo-23	Jordan_23 (*Desmodesmus sp.*)
Jo-29	Jordan_29 (*Desmodesmus sp.*)
Jo-34	Jordan_34 (*Tetradesmus obliquus* –Turpin- Wynne)
Jo-40	Jordan-40 (*Chlorella sorokiniana Shihira & Krauss*)
L	Log phase
MLAs	Machine learning analytics
MSC	Multiplicative scatter correction
PCA	Principal component analysis
PLSR	Partial least square regression
PP	Polyphosphate
R. s	Reference strain (*Tetradesmus obliquus* CCAP 276/7 –Turpin- Wynne)
RMSE	Root mean square error
RMSEC	Root mean square error of calibration
RMSEP	Root mean square error of prediction
S	Stationary phase
(SD) FTIR	Second derivative FTIR
SPP	Sodium polyphosphate
TGA	Thermogravimetric analysis

INTRODUCTION

Microalgae are a diverse group of microorganisms that use carbon dioxide and light to synthesize their food and produce oxygen. They do not compete with plants for freshwater resources or arable land because they can grow in brackish water,[1-3] wastewater, and non-arable land.[4-8] These characteristics make them suitable for the production of biomass that can be used as bioenergy,[9-12] feed,[13,14] fertilizers,[15-17] food,[18,19] and pharmaceutical compound manufacturing.[20,21]

The biochemical composition of microalgal biomass (e.g. protein, lipid, carbohydrates, phenolic contents) varies between different species[22-24] and within the same species between different strains.[25] In addition, growth regulators, growth stage, and growth conditions (such as light intensity, temperature, and nutrient availability) affect the chemical properties of microalgal biomass.[26-29] Thus, the biochemical characteristics of the microalgal biomass determine its applications.

Since many factors affect the biochemical composition of green microalgae, it is important to find rapid, simple, and non-destructive methods to monitor microalgal biomass composition. Fourier transform infrared (FTIR) offers these advantages and does not need large amounts of dried biomass, making it feasible to use in monitoring the biochemical composition during experiments. Moreover, this technique does not require a multistep procedure to extract biochemical compounds, and it can simultaneously detect several biochemical compounds in the same spectrum.[30] In addition, the presence of a rigid cell wall for many algal species affects the efficiency of the biochemical compound extraction and the accuracy of the extraction outcomes.[31,32] A complete characterization of microalgal biomass through FTIR, thermogravimetric analysis (TGA), and an analysis of the percentage of the elemental composition of carbon, hydrogen, nitrogen, and sulfur (CHNS) has been conducted for four microalgal species for biofuel generation and nutrient removal.[33] The lipid productivity in two microalgae (*Chlamydomonas sp.* and *Monoraphidium contortum*) under various nutrient conditions was studied using second derivative (SD) FTIR. In this respect, SD spectral analysis showed a high production of lipids for both species under nutrient-deplete conditions.[34] FTIR was an efficient and reliable method to determine the carbohydrate and lipid contents of *Chlamydomonas reinhardtii* and *Scenedesmus subspicatus* grown in nitrogen-limiting conditions.[35] Furthermore, FTIR was applied in parallel with chemometrics for detecting grazers in microalgal cultures.[36] Since the invasion of predators can destroy the microalgal culture within a couple of days,[37] the combination of FTIR with two machine learning analytics (MLAs) can act as an early detection tool for real-time and early detection of algal grazers.[36] Moreover, this technique shows a specific chemical signature for each substance based on the functional groups present within said substance. As algal samples consist mainly of carbohydrates, proteins, and lipids that are strongly absorbed in the infrared region, FTIR serves as an ideal technique for their analysis.[38]

The combination of FTIR and chemometrics allows for mixtures of key constituents (carbohydrates, lipids, and proteins) to be quantified and for their uses to be assessed for industry sectors, such as bioenergy, food, green chemicals, cosmetics, and therapeutics. Furthermore, the application of MLAs, such as the correlation method, principal component analysis (PCA), and partial least square regression (PLSR), was undertaken. The correlation method and PCA allowed for the qualitative analysis of different microalgae species by revealing information about patterns between their spectra.[35,39,40] In addition, PLSR assisted in quantifying constituents (for example, carbohydrates, lipids, and proteins) in microalgae. As PLSR is a regression algorithm, it was crucial to have different concentrations of constituents in a PLSR model; different concentrations can be made by preparing different mixtures of the constituents and/or any other constituents of variable range.[41,42] This has been extensively applied to medicinal constituents.[43-45] Subsequently, this approach was adopted for microalgae constituents.

Jordan is one of the most water-scarce countries in the world.[46] Improving the quality of their treated wastewater is very important for safe reuse in agriculture and industry. The safe reuse of treated wastewater helps to reduce the existing stress on freshwater resources in the country.[47] This research is a second phase of a previous study, in which local green microalgae were isolated from Jordan, and their efficiency for wastewater treatment was assessed at a laboratory-scale based on their specific growth rate, specific nutrient removal rate from wastewater, and settleability by gravity. The removal of excess nutrients from the wastewater depends on how fast these nutrients can grow.[48] As an environmentally and economically friendly method to be applied in high-rate algal ponds (HRAP) in the future, the biomass of these microalgae is intended to be used as a by-product to reduce the cost of the wastewater treatment process. Since the produced microalgal biomass can be used as a slow-releasing fertilizer or as animal feed, it is worth noting that before applying the biomass, it should not have a high concentration of persisting organic pollutants or heavy metals.[49] In addition, the biomass can be converted into biofuel.[50]

Journal of Chemical Technology and Biotechnology published by John Wiley & Sons Ltd on behalf of Society of Chemical Industry (SCI).

Therefore, this research used portable ATR-FTIR spectroscopy with MLAs to determine microalgal constituents in log and stationary phases. The microalgae evaluated were obtained from eight different local green microalgae isolates from Jordan, in addition to a reference strain obtained from a culture collection known for its efficiency in removing nutrients from water.

MATERIALS AND METHODS

Materials

Eight different local green microalgal isolates from Jordan and a reference strain obtained from a culture collection were evaluated (Table 1). For carbohydrate, protein, and lipid references, the following chemicals were used as models: glucose, sucrose, bovine serum albumin (BSA; CAS number 9048–46-8), glycerol tripalmitate (GTP; CAS number 555–44-2), and sodium polyphosphate (SPP; CAS number 10361–03-2); all were purchased from Sigma-Aldrich.

Sample preparation

The isolated green microalgae were grown in triplicate in real municipal wastewater beside a reference species bought from the culture collection of algae and protozoa (CCAP): *T. obliquus* CCAP 276/7. The wastewater in this experiment was collected from a wastewater treatment plant in the UK and modified to become similar to real wastewater in Jordan, with the following characteristics: Phosphate (PO_4-P), 5.8 mg L^{-1}; Ammonium (NH_4-N), 20.15 mg L^{-1}; Nitrate (NO_3-N), 24.35 mg L^{-1}; the Total nitrogen was 45.2 mg L^{-1}. Then, the water was filtered through GF/C filter paper and sterilized using ozone gas for 15 min, followed by pumping filtered air for 20 min. The samples were grown at a temperature of 22 °C, with light intensity of 130–150 µmol·photons·m^{-2}·s^{-1}, light: dark cycle in hours of 16:8, and they were supplied with 2.5% CO_2 gas and seeded at *ca.* 1.0×10^6 cells mL^{-1}.

The growth of the samples was monitored by cells count using a Guava easyCyte™ flow cytometer (Millipore, Hayward, CA 94545, USA); when the samples were mid-log phase and in the stationary phase after 10 days of growth, 200 mL were collected and the cells were harvested by centrifugation at 3200 × g for 5 min. The algal pellets were washed twice with deionized water, then resuspended in deionized water, frozen as droplets in liquid nitrogen, and freeze-dried (Freezdryer Modylyo, Edwards, Britain; Vacuum pump RV8, Edwards, England).

For quantitative analysis, three different mixtures were prepared; binary, ternary, and quaternary mixtures containing the following constituents were prepared: BSA, glucose, GTP, and SPP (Table 2). In addition, two mixtures were made based on the standard addition of either BSA or glucose to algal samples.[42] Details of mixtures are listed below:

(1) Mixture 1: Binary mixture of glucose and SPP.
(2) Mixture 2: Ternary mixture of BSA, glucose, and SPP.
(3) Mixture 3: Quaternary mixture of BSA, glucose, GTP, and SPP.
(4) Mixture 4: Standard addition of BSA to freeze-dried algae powder, such as the algae of all eight species.
(5) Mixture 5: Standard addition of glucose to freeze-dried algae powder, such as the microalgae of all eight species.

FTIR analysis

FTIR spectra were collected using the Bruker Alpha FTIR spectrometer equipped with ATR diamond holder (Fig. 1). Each spectrum

Table 2. Mixtures used in this study

MN	Nc	Constituents	Nv
MN1	2	Glucose SPP	11
MN2	3	BSA Glucose SPP	15
MN3	4	BSA Glucose GTP SPP	17
MN4	2	Algae species BSA	12
MN5	2	Algae species Glucose	12

Abbreviations: MN, mixture number; Nc, Number of constituents; Nv, number of vials. In all cases, the calibration range was 0–100% m/m of each constituent.

Table 1. Details of samples used in the study

Isolate	Log Phase (L; 3 biological replicates A, B & C)			Stationary Phase (S; 3 biological replicates A, B & C)		
Jordan_2 (Jo-2)	Jo-2_A/L	Jo-2_B/L	Jo-2_C/L	Jo-2_A/S	Jo-2_B/S	Jo-2_C/S
Jordan_4 (Jo-4)	Jo-4_A/L	Jo-4_B/L	Jo-4_C/L	Jo-4_A/S	Jo-4_B/S	Jo-4_C/S
Jordan_12 (Jo-12)	Jo-12_A/L	Jo-12_B/L	Jo-12_C/L	Jo-12_A/S	Jo-12_B/S	Jo-12_C/S
Jordan_18 (Jo-18)	Jo-18_A/L	Jo-18_B/L	Jo-18_C/L	Jo-18_A/S	Jo-18_B/S	Jo-18_C/S
Jordan_23 (Jo-23)	Jo-23_A/L	Jo-23_B/L	Jo-23_C/L	Jo-23_A/S	Jo-23_B/S	Jo-23_C/S
Jordan_29 (Jo-29)	Jo-29_A/L	Jo-29_B/L	Jo-29_C/L	Jo-29_A/S	Jo-29_B/S	Jo-29_C/S
Jordan_34 (Jo-34)	Jo-34_A/L	Jo-34_B/L	Jo-34_C/L	Jo-34_A/S	Jo-34_B/S	Jo-34_C/S
Jordan_4 0(Jo-40)	Jo-40_A/L	Jo-40_B/L	Jo-40_C/L	Jo-40_A/S	Jo-40_B/S	Jo-40_C/S
Reference Strain (R. s)	R.s_A/L	R.s_B/L	R.s_C/L	R.s_A/S	R.s_B/S	R.s_C/S

Note: Jordan_2, *Desmodesmus sp.*; Jordan_4, *Desmodesmus sp.*; Jordan_12, *Coelastrella sp.*; Jordan_18, *Desmodesmus subspicatus*; Jordan_23, *Desmodesmus sp.*; Jordan_29, *Desmodesmus sp.*; Jordan_34, *Tetradesmus obliquus*; Jordan_40, *Chlorella sorokiniana*; Reference strain, *Tetradesmus obliquus* CCAP 276/7.

Journal of Chemical Technology and Biotechnology published by John Wiley & Sons Ltd on behalf of Society of Chemical Industry (SCI).

Figure 1. The Bruker Alpha FTIR spectrometer equipped with ATR diamond holder.

was the sum of 16 scans over the wavenumber range 400–4000 cm^{-1} and with the optical resolution 4 cm^{-1}.

Algal samples, raw materials, and the mixture of raw materials were stored in glass vials prior to measurements. They were measured as received by placing a few milligrams of each sample on the ATR crystal sample interface. Three measurements were taken from each sample after being mixed for 2 min on a vortex mixer. Three spectra were taken for each measurement.

Qualitative and quantitative analysis

FTIR Spectra were exported in Matlab 2022a, where qualitative and quantitative analyses were applied. Prior to spectral interpretation, a baseline correction of spectra was made using multiplicative scatter correction (MSC). MSC corrected changes in the offset that resulted from different factors relating to sample properties and environmental conditions. This was achieved by the construction of a new spectrum that is a linear combination of the original spectrum (Eqn 1).[51]

$$y_{MSC,i} = \frac{(y_i - a)}{b} \tag{1}$$

Where,

yMSC,i is the corrected spectrum value,
i is the number,
yi is the original spectrum,
a is the intercept of the line,
b is the slope of the line.

Correlation method

The correlation method measured the momentum product between two or more spectra (Eqn 2).[51]

$$r_p = \frac{\sum (A_i - \overline{A})(B_i - \overline{B})}{\sqrt{\sum (A_i - \overline{A})^2 \sum (B_i - \overline{B})^2}} \tag{2}$$

Where,

r_p is the momentum product,
A is the first spectrum and $\overline{A}$ its mean,
B is the second spectrum and $\overline{B}$ its mean,
i is the corrected spectrum value.

In this respect, the correlation method measured the similarity of the spectra of the compared samples. The range of correlation is −1 to +1, where −1 indicates dissimilarity and + 1 is identical. The threshold adopted for the correlation coefficient (*r*) value was 0.95; hence, r values above 0.95 were considered a match and vice versa. Type I errors and type II errors were investigated for evaluation of the correlation method. A type I error was encountered when an algal sample gave an r value <0.95 against its own species. A type II error was met when an algal sample gave an r value >0.95 against a different species.[51]

Principal component analysis (PCA)

PCA is a multivariate classification technique that clusters samples according to variances among the samples. In this case, PCA reduces the dimension of the data into scores and loadings. Thus, PC1 has the highest variance, PC2 the second highest not related to the first, and so on. The scores show the distribution of the data in a multidimensional space, whereas the loadings show the important absorbance values that relate to significant constituents within the PCA model. In this respect, PCA was used to visualize the clustering of algal samples and constituents. The model was evaluated by considering type I errors and type II errors; in this model, a type I error was observed when a sample was not clustered with its species and a type II error was observed when a sample was clustered with different species.

The relationship between the spectra, scores, and loadings is described in Eqn 3.[52]

$$X = T \times P + Q \tag{3}$$

Where,

X is the original data matrix,
T represents the scores,
P represents the loadings,
Q represents the residuals.

Partial least square regression (PLSR)

PLSR was applied as a multivariate quantitative analytic considering the differences in the physicochemical properties in the mixtures.[53] The principle of PLSR was similar to that of PCA but applied to x and y variables; where x values were the absorbance values and y values represented the concentration values. In this respect, a PLSR model retrieved components in x values that were related to y values and rejected unrelated components. The components were retrieved based on finding factors that captured variance among the data such that one factor was added at a time in descending order. Hence, the first factor had the highest variance, the second had the second highest variance, and so on.[52]

$$X = T \times P + E \tag{4}$$

$$c = T \times q + f \tag{5}$$

Where,

X is the absorbance at different wavenumbers,
c is the concentration,
q is the loading vector,
T represents the scores,

Journal of Chemical Technology and Biotechnology published by John Wiley & Sons Ltd on behalf of Society of Chemical Industry (SCI).

P represents the loadings,
E and *f* represent the residuals.

Three parameters were considered when evaluating the PLSR model, including the regression correlation coefficient (r^2) value, and the root mean square error of calibration (RMSEC) and prediction (RMSEP) (Eqns 6 and 7). Thus, the higher the r^2 value, the more accurate the model; the lower the RMSEC and RMSEP values, the more precision the model possesses. Moreover, the closeness of RMSEC to RMSEP values indicated robustness of the models.[51]

$$RMSEC = \sqrt{\frac{\left(\sum_{(y-\bar{y})}{}^2\right.}{n-p-1}} \quad (6)$$

Where,

y is the % m/m of the calibration set,
is % m/m of each component,
n is the number of samples in the calibration set,
p is the number of factors used in the model.

$$RMSEP = \sqrt{\frac{\left(\sum_{(y-\bar{y})}{}^2\right.}{n-1}} \quad (7)$$

Where,

y is the % m/m of the prediction set,
is % m/m of each component,
n is the number of samples in the prediction set.

Univariate regression analysis

Quantitative analysis was conducted by univariate regression models that were developed based on the binary, ternary, and quaternary mixtures. The models were developed at key absorbances for carbohydrates, proteins, and lipids. For model evaluation, accuracy and precision were considered by measuring r^2 and root mean square error (RMSE) values. In this respect, the higher the r^2 and lower RMSE values, the more accurate and precise the model is. To reduce the peak absorption's fluctuation as a result of the inhomogeneous thickness of the sample, amide I (1724–1585 cm^{-1}) was selected as an internal reference peak for the evaluation of the carbohydrate and lipid contents relative to protein.

RESULTS AND DISCUSSION

Advantages of using portable infrared spectroscopy

Portable ATR-FTIR spectroscopy offered a rapid, simple, and sustainable approach when compared to laboratory-based techniques.[54] The ATR-FTIR spectrometer used (Fig. 1) could be

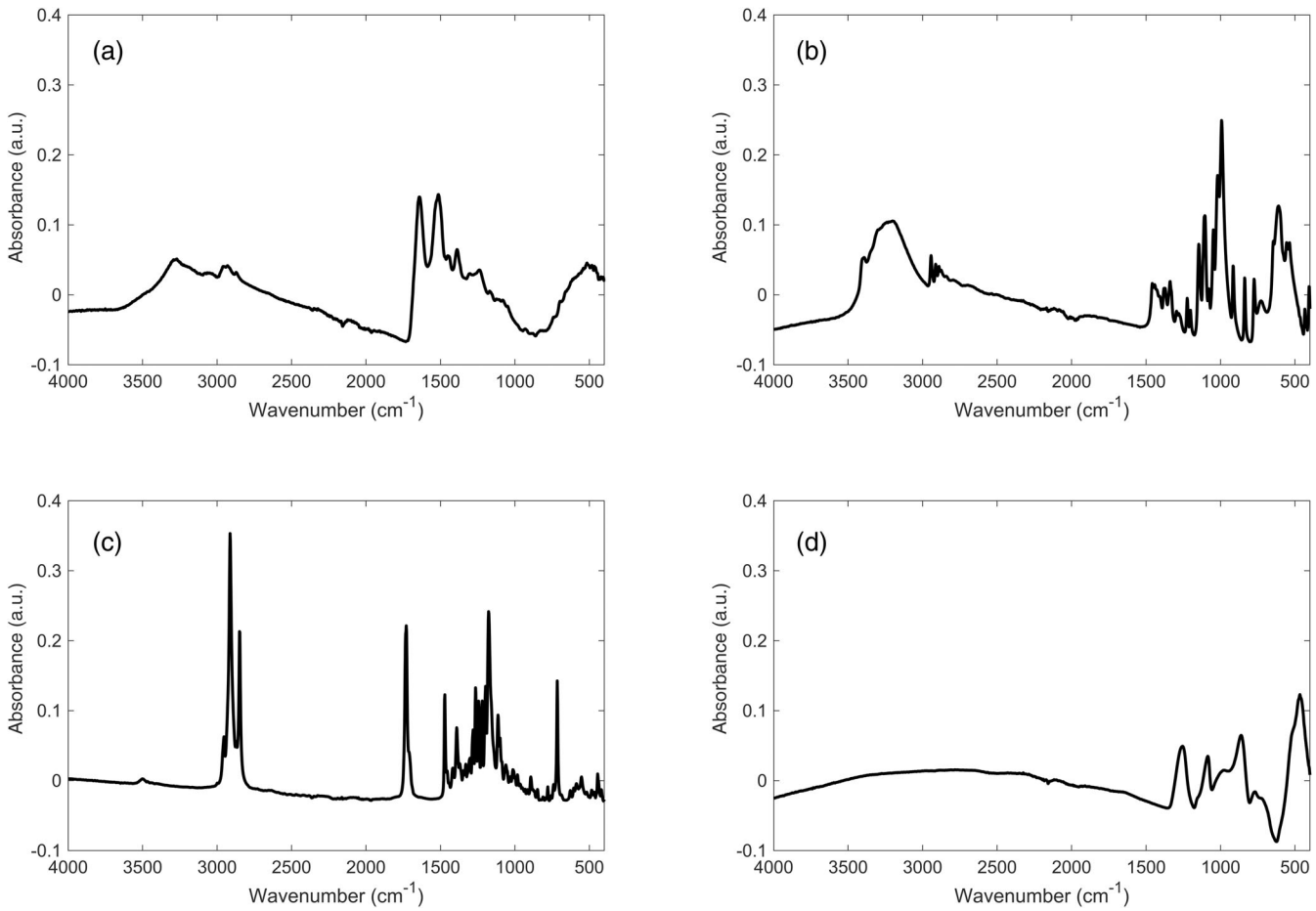


Figure 2. MSC-treated FTIR spectra of (a) BSA, (b) glucose, (c) GTP, and (d) SPP were measured using the Bruker Alpha FTIR spectrometer equipped with ATR diamond holder.

Journal of Chemical Technology and Biotechnology published by John Wiley & Sons Ltd on behalf of Society of Chemical Industry (SCI).

easily carried to the field provided a power supply is available. This portability saves on the time and costs attributed to importing samples to a laboratory. The instrument weight is about 2 kg and it operates over variable temperature conditions. Moreover, the instrument has the optionality to incorporate in-built spectral libraries that offer identification on the spot. It is worth noting that an FTIR spectrum can be acquired in less than a minute. Subsequently, if instant measurements are combined with in-built identification, the characterization of samples can be made within minutes. This is advantageous when measuring a large number of samples off-site. Moreover, the ease of measurement of samples by using portable instruments allows them to be run by non-skilled personnel. It is the first research to evaluate the efficacy of using portable FTIR for algal biomass evaluation. This technique is important for monitoring the variations in dried algal biomass on the field for quality checks in the case of large-scale microalgae growing facilities (for example, wastewater treatment plants or microalgae producing facilities that have many ponds for growing microalgae). Drying the biomass is essential to minimize the degradation of the biomass and stabilize the biomass before processing it, so this step is essential for microalgae biomass production systems. In general, the biomass drying method varies in capital expenditures, operating expense investments, and depending on the end-product of the biomass; for example, in high-value products (pharmaceuticals), fast and energy-intensive methods could be applied (such as lyophilization, spray, fluidized bed).[55] On the other hand, if biofuel is the end-product, energy input must be minimized to make the process more economically viable and competitive with fossil fuels; for example, solar drying is very convenient for countries (like Jordan) that receive plenty of sunlight hours all year round. In addition, other factors should be considered based on the intended functional product when the drying method is chosen, such as the nature of the microalgal cell.[55]

Spectral quality

Raw materials

The spectral quality of the main constituents in algae were evaluated in relation to number of peaks, absorption, absorption intensity, and range. The absorption intensity varied between 0.1 and 0.6 absorbance units. SPP showed the lowest number of peaks ($n = 5$) and its absorption range was limited to the fingerprint region between 1200 and 400 cm^{-1} (Fig. 2).

The absorbance range of SPP was between 0.08 and 0.25 absorbance units. Likewise, BSA showed a low number of peaks ($n = 6$) and weak absorbance in the range of 0.06 and 0.14 absorbance units. On the other hand, glucose and GTP showed stronger absorbance when the number of peaks and intensities were compared. Glucose showed 24 peaks over the full wavenumber range and had absorbance intensities between 0.04 and 0.3 absorbance units. GTP was the strongest absorbing constituent and showed 24 peaks in the range of 0.02–0.49 absorbance units.

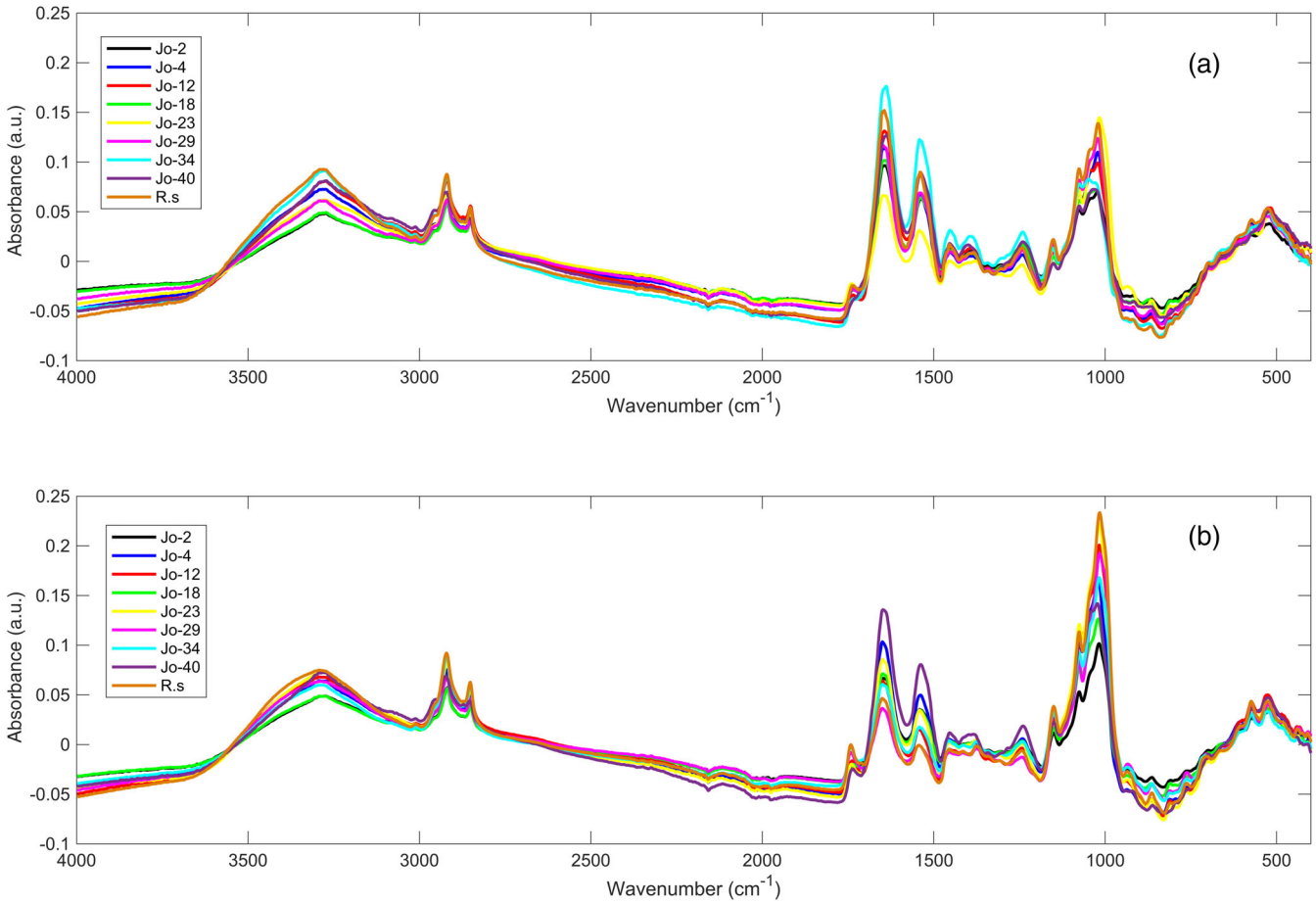


Figure 3. MSC-treated FTIR spectra of the (a) log phase and (b) stationary phase of microalgae samples measured using the Bruker Alpha FTIR spectrometer equipped with ATR diamond holder.

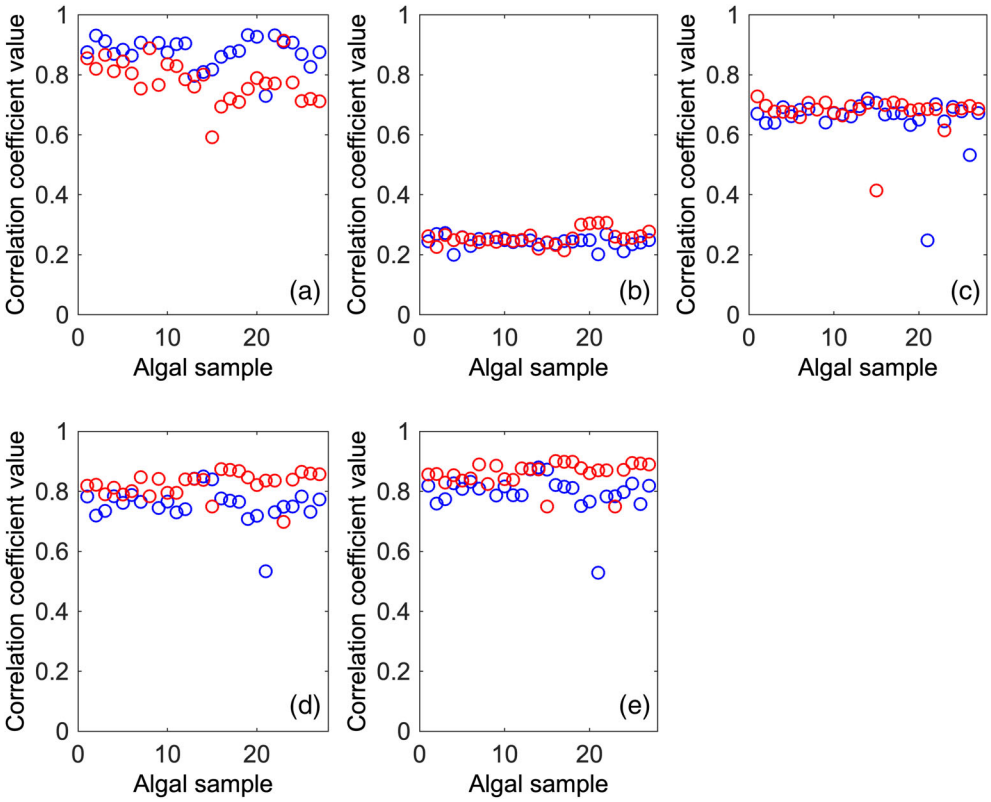


Figure 4. Correlation coefficient values of the MSC-treated FTIR spectra of microalgae samples in log (blue) and stationary (red) phases against (a) BSA, (b) glucose, (c) GTP, (d) SPP, and (e) sucrose. Measurements were conducted with the Bruker Alpha FTIR spectrometer equipped with ATR diamond holder.

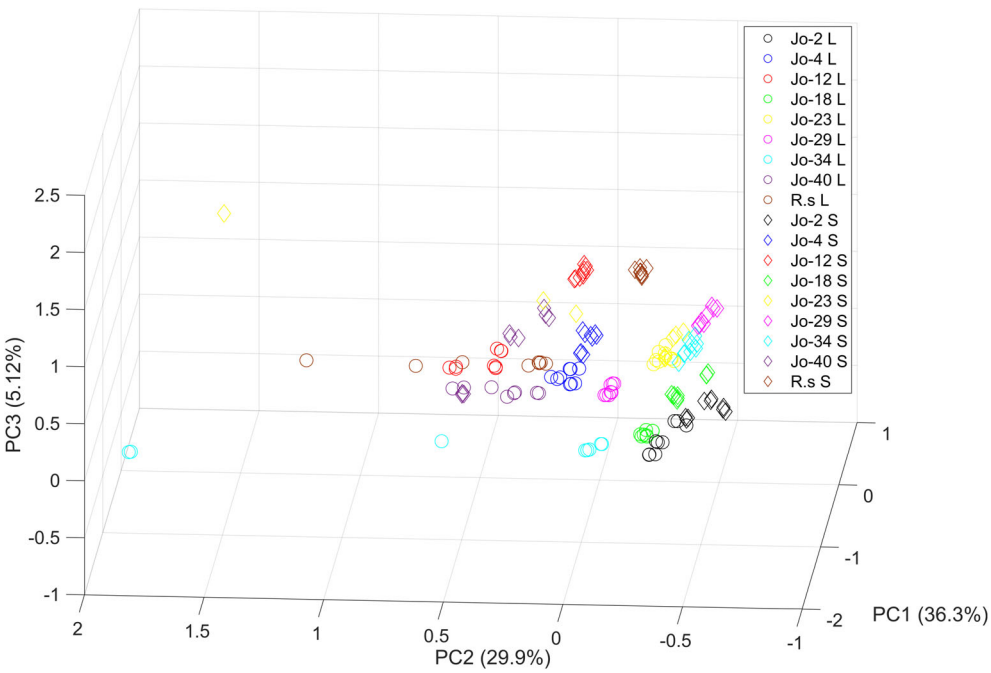


Figure 5. PCA score plot of PCA model 1 applied to the MSC-treated FTIR spectra of microalgae samples (stationary and logs) over the full wavenumber range.

Algal samples

When the algal biological replicates were compared, they showed FTIR activity corresponding mainly to protein and GTP. Key peaks corresponding to BSA and GTP were seen at 3284, 2919, 2851, 1744, 1638, 1542, and 1012 cm^{-1}. The aforementioned peaks corresponded to amine/NH (protein), methyl/methylene groups

Journal of Chemical Technology and Biotechnology published by John Wiley & Sons Ltd on behalf of Society of Chemical Industry (SCI).

(CH/=CH) (GTP), ester group (GTP), amide I (protein), amide II (protein), and ether (COC) (GTP) respectively.[56]

The changes in carbohydrate, lipid, and protein contents could be monitored by investigating of the IR spectra for the three biological replicates of the eight microalgal isolates and the reference strain in the log phase and stationary phase (after 10-day growth). The samples' IR spectra showed an absorbance range between 0.02 and 0.4 absorbance units over the full range (Fig. 3). The FTIR spectra showed about 20 peaks corresponding to proteins, lipids, and carbohydrates. The samples showed a decrease in protein content and an increase in carbohydrate content by three-fold. Nonetheless, the lipid content was constant in both log and stationary samples.

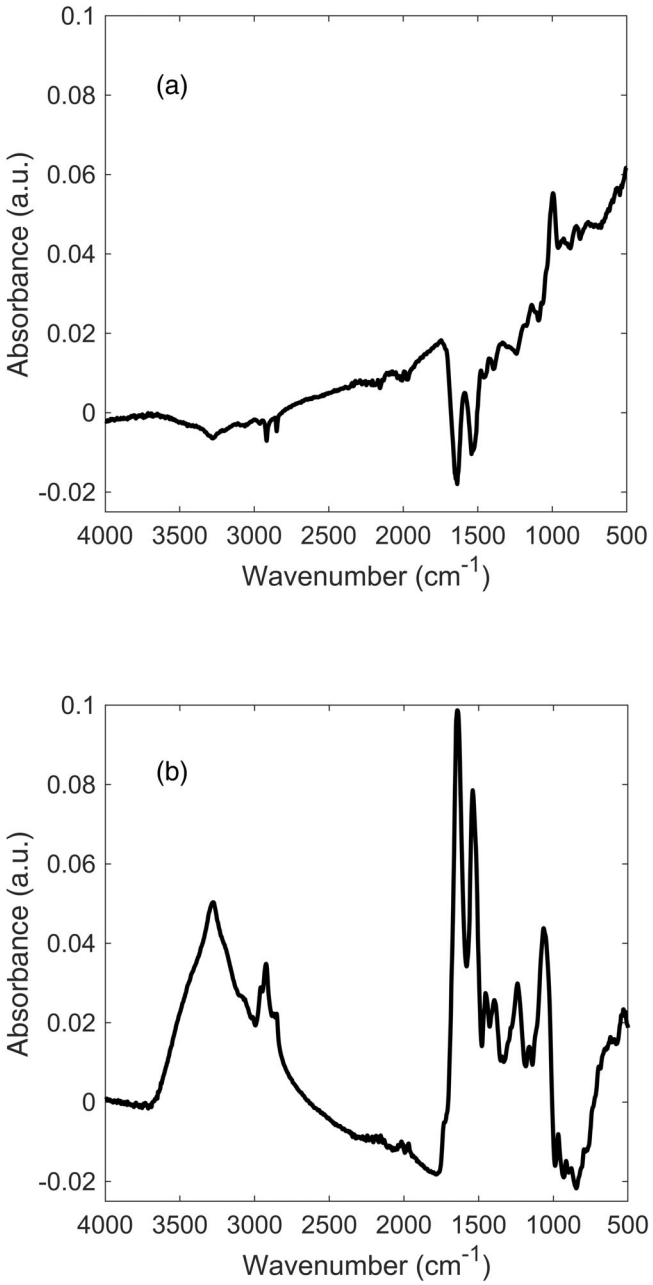


Figure 6. (a) PC1 and (b) PC2 loading plots of PCA model 1 that were applied to the MSC-treated FTIR spectra of microalgae samples (stationary and logs) over the full wavenumber range.

Qualitative analysis correlation in wavenumber space

The spectra of the algal samples were compared against the constituents' spectra and other spectra. The pattern of correlation was observed for samples that showed a decrease in BSA concentration in the stationary phase and an increase in lipids and sugars (Fig. 4). The r value ranges of stationary and log samples against BSA were in the range of 0.59–0.89 and 0.73–0.94, respectively. Moreover, the r values of glucose, GTP, SPP, and sucrose against stationary and log phases were in the range of 0.21–0.90 and 0.20–0.88, respectively.

Principal component analysis (PCA)

Two different PCA models were applied over two different ranges: the first (PCA model 1) being the full range and the second (PCA model 2) being the range of 900–1800 cm^{-1} that has the peaks of interest for proteins, sugars, and lipids.[57] PCA model 1 showed better discrimination between the log and stationary phases of each species; thus, no type II error was observed. However, type I errors were seen for both models. Hence, PCA model 1 showed type I errors for two species of the log phase (Jo-34 L and R.s L) and one species of the stationary phase (Jo-23 S) (Fig. 5).

This could be related to noise in the spectra of the specific samples where the majority of samples showed clear accuracy of qualification.[52] It is worth noting that the accuracy of identification was related to the protein, lipid, and carbohydrate constituents and this was confirmed by the PC loading plot of PCA model 1 (Fig. 6). Hence, both loading plots showed significant contribution of bands corresponding to the full algal spectra that showed key features of carbohydrates, lipids, and proteins. This, in turn, confirmed the explainability of the PCA model where the variances of the first two PCs corresponded to the full algal spectra.[58]

On the other hand, PCA model 2 showed type I errors for Jo-34 L at log phase and one species of the stationary phase (Jo-23 S). In addition, PCA model 2 showed type II errors for two species with overlapping scores (Fig. 7). These species were Jo-2 L (*Desmodesmus sp.*) and Jo-18 L (*Desmodesmus subspicatus)* and this could be related to similarities in the spectra between both species within the limited range of 900–1800 cm^{-1}. Thus, the PCA model applied over the full range offered more accuracy in discriminating samples than the limited range.

PCA model 2 did not show overlap between the stationary and log phases of each sample and this confirmed the model's ability to discriminate between both phases over the selected range. Explainability of the model was featured in the loading that showed key peaks to the constituents, where the highest peak was seen around 1000 cm^{-1}, thus contributing to COC group of GTP (Fig. 8).[56] Moreover, key contribution was seen around 1750 cm^{-1} that was related to ester group for lipids. This, in turn, confirmed that lipid and protein constituents contributed to differences among the different Jo samples.[56]

Quantitative analysis

Subsequently, univariate and multivariate regression models were constructed for the quantification of the carbohydrate, protein, and lipid contents in the samples.

Univariate analysis

A total of 14 univariate regression models were constructed at the key absorbances for the main constituents in algal samples. These included: bands at 1740 (ester), 2800 (methyl), 3000 (methylene)

Journal of Chemical Technology and Biotechnology published by John Wiley & Sons Ltd on behalf of Society of Chemical Industry (SCI).

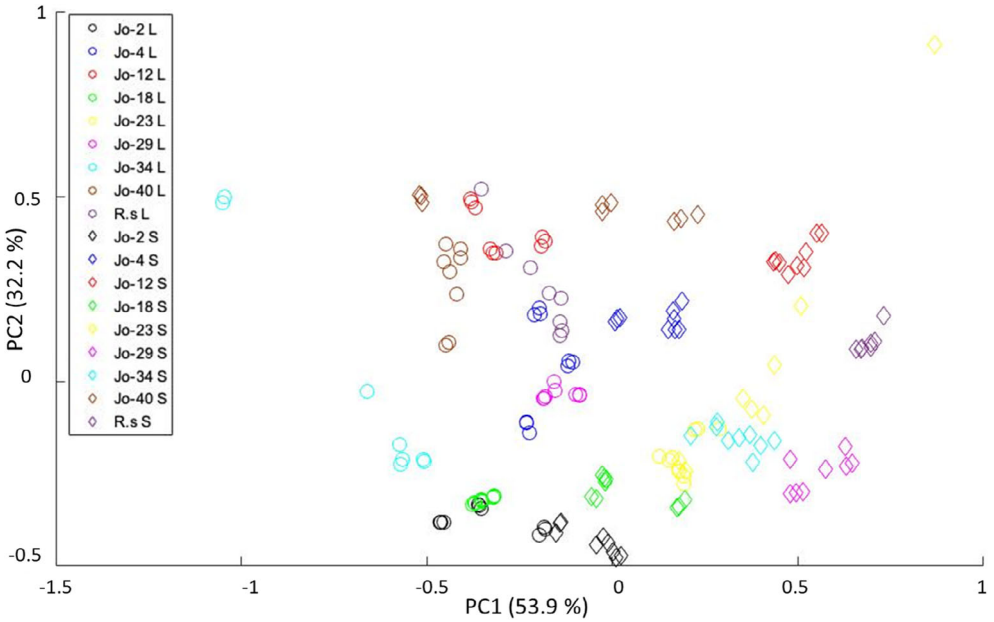


Figure 7. PCA scores plot of PCA model 1 applied to the MSC-treated FTIR spectra of microalgae samples (stationary and logs) over the wavenumber range of 900–1800 cm^{-1}.

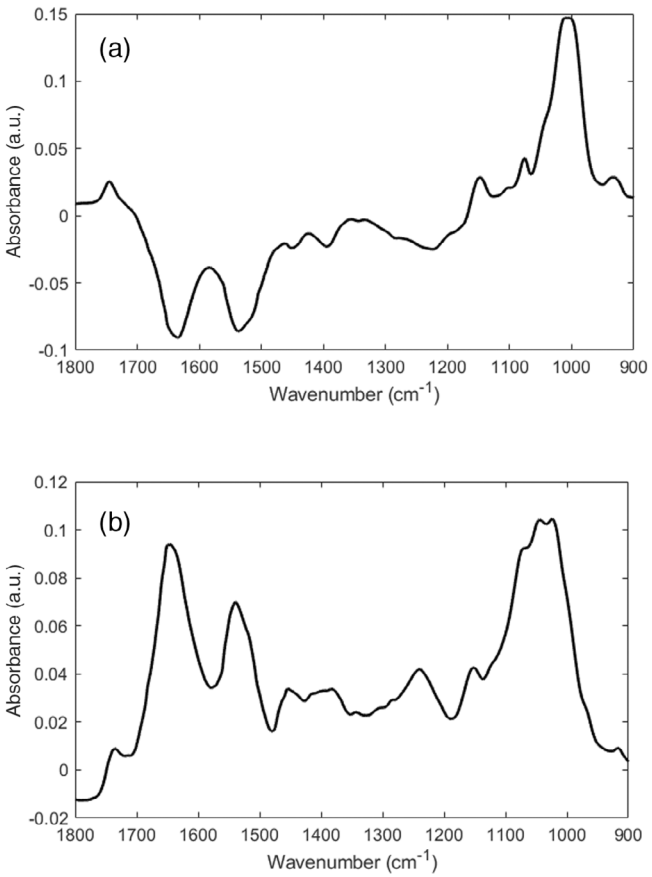


Figure 8. (a) PC1 and (b) PC2 loading plots of PCA model 1 that were applied to the MSC-treated FTIR spectra of microalgae samples (stationary and logs) over the wavenumber range of 900–1800 cm^{-1}.

cm^{-1} for lipids; bands at 1650 (amide I) and 1540 (amide II) cm^{-1} for proteins; and bands at 1050 cm^{-1} for carbohydrates. Depending on the linearity of the key bands present in each mixture, quantitative models were built based on carbohydrate, lipid, or protein quantifications (Table 3 and Appendix I Figs A1–A6).

For binary mixtures, only glucose bands showed linearity and, consequently, three models (M1 – M3) were constructed based on the absorbances at 3212, 1048, and 1016 cm^{-1} (Fig. 9). In all three cases, a calibration range of 10–90% m/m was used. All three models showed high accuracy of calibration with r^2 values of 0.96, 0.98, and 0.99, respectively. Nonetheless, there was variation in the precision, where M3 (at 1016 cm^{-1}) showed the highest precision, followed by M2 (at 1048 cm^{-1}) and M1 (at 3212 cm^{-1}). This was justified by the RMSEC values obtained for M3, M2, and M1 being 3.58, 4.13, and 5.94% m/m, respectively.

For ternary mixtures, models based on glucose and BSA were constructed. Glucose models (M4 and M5) were constructed in the range of 20–80% m/m at the absorbances of 993 and 1012 cm^{-1}, respectively. Both models showed similar high accuracy and precision, with r^2 and RMSEC values of 0.97 and 4% m/m, respectively. Likewise, high accuracy and precision were seen for BSA models (M6 and M7) constructed with the ternary mixture in the range of 10–75% m/m. In this respect, the r^2 and RMSEC values for M6 and M7 were around 0.99 and 2.5–2.99% m/m. This showed that the ternary mixtures had more accuracy and precision than the binary mixtures. High accuracy and precision in the latter case could be attributed to using more variables (three rather than two constituents in the mixture). However, including more constituents did not indicate higher accuracy and precision in all cases, and that depended on the constituent quantified and its IR activity. Hence, quaternary mixtures constructed with BSA, glucose, GTP, and SPP showed lower accuracy and precision for glucose quantification than BSA and GTP quantification. Glucose quantification at 993 and 1012 cm^{-1} gave r^2 values of 0.93 and 0.94 and RMSE values of 12.3 and 7.39% m/m, respectively. However, BSA quantification using models based on quaternary mixtures (at 1650 and 1515 cm^{-1}) showed high accuracy and precision with r^2 values of 0.97 and 0.98 and RMSE values of 4.63 and 5.46% m/m, respectively. GTP also

Journal of Chemical Technology and Biotechnology published by John Wiley & Sons Ltd on behalf of Society of Chemical Industry (SCI).

Table 3. Details of univariate regression models

N	Mixture name	Substance quantified	Calibration range (% m/m)	Wavenumber (cm^{-1})	r^2	RMSE (% m/m)
M1	Binary	Glucose	10–90	3212	0.9563	5.94
M2	Binary	Glucose	10–90	1048	0.9776	4.13
M3	Binary	Glucose	10–90	1016	0.9848	3.58
M4	Tertiary	Glucose	20–80	993	0.9701	4.16
M5	Tertiary	Glucose	20–80	1012	0.9704	4.44
M6	Tertiary	BSA	10–75	1644	0.9899	2.96
M7	Tertiary	BSA	10–75	1515	0.9905	2.54
M8	Quaternary	Glucose	5–80	993	0.9259	12.3
M9	Quaternary	Glucose	5–80	1020	0.944	7.39
M10	Quaternary	BSA	5–80	1650	0.9712	4.63
M11	Quaternary	BSA	5–80	1515	0.9737	5.46
M12	Quaternary	GTP	5–80	2913	0.9781	4.4
M13	Quaternary	GTP	5–80	2849	0.9688	5.18
M14	Quaternary	GTP	5–80	1730	0.9885	3.24

Abbreviations: N, model number; M, model; r^2, regression correlation coefficient; %m/m, percentage mass per mass.

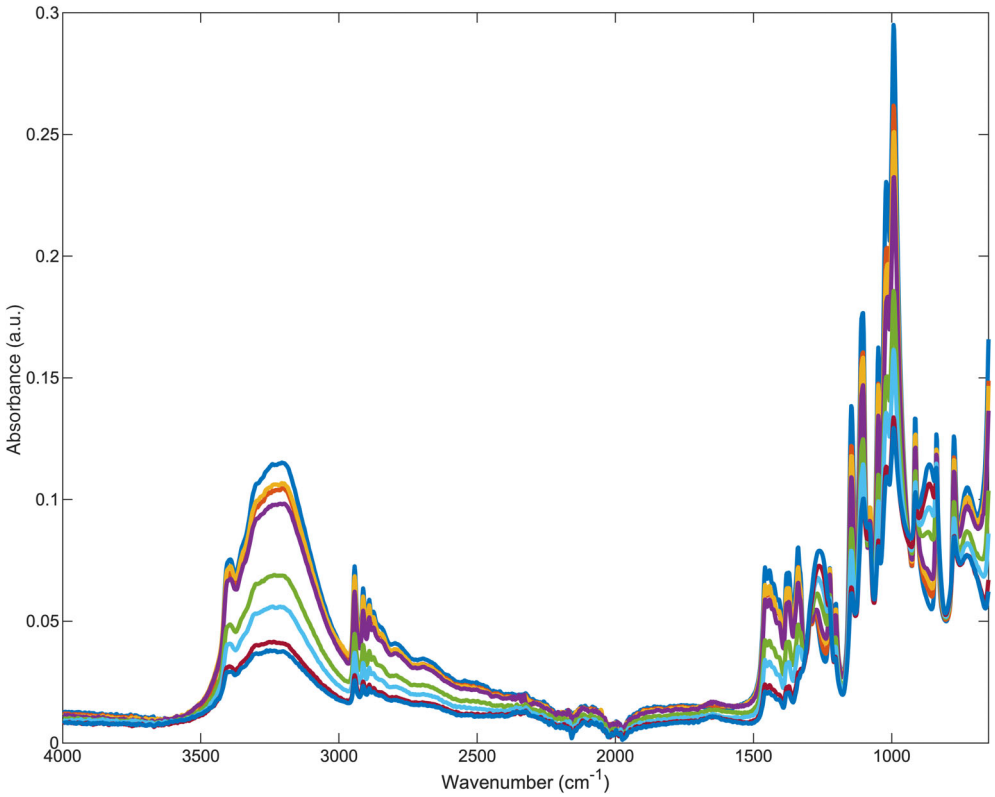


Figure 9. MSC-treated FTIR spectra of the binary mixture measured using the Bruker Alpha FTIR spectrometer equipped with ATR diamond holder.

showed accurate and precise quantitative models at 2913, 2849, and 1730 cm^{-1}. The aforementioned three models yielded r^2 values of 0.98, 0.97, and 0.99, and RMSE values of 4.40, 5.18, and 3.24% m/m, respectively.

Multivariate analysis

Multivariate regression models based on PLSR were constructed for quantification of glucose and SPP (Table 4).

PLSR models based on the binary mixture (MN1) of glucose and SPP showed accurate quantifications that were dependent on the concentration range and spectral range of the quantified analyte in the mixture. Out of the 15 PLSR models applied to MN1, the most accurate models for quantifying glucose and SPP were obtained for the concentration range of 10–100% m/m of each glucose and SPP. Thus, for glucose quantification, the PLSRM6 that was applied over the wavenumber range of 1533.84–418.13 cm^{-1} showed r^2C/r^2V values and RMSEC/RMSEP values of 0.9613/0.9761

Journal of Chemical Technology and Biotechnology published by John Wiley & Sons Ltd on behalf of Society of Chemical Industry (SCI).

Table 4. PLSR models based on binary and ternary mixtures

N	Range		Analyte	Analyte (% m/m)	F	C:V	r^2C	RMSEC	r^2V	RMSEP
PLSRM1	Full range	Binary	Glucose	0–100	3	18:9	0.9567	5.38	0.9859	4.13
PLSRM2	Full range	Binary	Glucose	0–90	1	20:10	0.8789	9.99	0.9239	9.98
PLSRM3	Full range	Binary	Glucose	10–100	3	20:10	0.9586	5.84	0.9647	5.44
PLSRM4	Full range	Binary	Glucose	10–90	3	20:10	0.9567	5.38	0.9929	4.13
PLSRM5	3544. 9-2965.7	Binary	Glucose	10–100	2	20:10	0.9506	6.38	0.9166	9.4
PLSRM6	1533.84–418.13	Binary	Glucose	10–100	3	20:10	0.9613	5.65	0.9761	4.48
PLSRM7	3544. 9-2965.7, 1533.84–418.13	Binary	Glucose	10–100	3	20:10	0.9569	5.96	0.9641	5.48
PLSRM8	Full range	Binary	SPP	0–100	6	26:13	0.944	8.09	0.8606	13.9
PLSRM9	Full range	Binary	SPP	0–90	5	24:12	0.9413	5.38	0.9754	4.94
PLSRM10	Full range	Binary	SPP	10–100	4	20:10	0.9686	5.09	0.9883	3.94
PLSRM11	Full range	Binary	SPP	10–90	3	18:9	0.9566	5.38	0.9858	4.13
PLSRM12	3012. 6-2782.1	Binary	SPP	10–100	3	20:10	0.9419	6.93	0.9639	8.27
PLSRM13	1786.755–1672.533	Binary	SPP	10–100	2	20:10	0.9115	8.55	0.9819	7.18
PLSRM14	1493.042–409.9746	Binary	SPP	10–100	4	20:10	0.9719	4.79	0.995	3.37
PLSRM15	3012. 6-2782.1, 1786.755–1672.533, 1493.042–409.9746	Binary	SPP	10–100	4	20:10	0.9747	4.57	0.9954	3.28
PLSRM16	Full range	Tertiary	Glucose	0–100	6	34:17	0.9641	7.04	0.9181	9.77
PLSRM17	Full range	Tertiary	Glucose	0–90	8	32:16	0.9736	5.69	0.9204	8.67
PLSRM18	Full range	Tertiary	Glucose	10–100	6	26:13	0.9704	6.53	0.9319	8.37
PLSRM19	Full range	Tertiary	Glucose	5–90	6	24:12	0.9631	6.27	0.9134	8.44
PLSRM20	932–1224	Tertiary	Glucose	5–100	2	9:27	0.9889	3.93	0.7963	11.88
PLSRM21	932–1224	Tertiary	Glucose	0–100	10	34:17	0.9808	5.35	0.9657	6.32

Note: F represents the number of components.

and 5.65/4.48% m/m, respectively. Likewise, the PLSRM7 that was applied over the wavenumber ranges of 3544.9–2965.7 and 1533.84–418.13 cm^{-1} showed r^2C/r^2V values and RMSEC/RMSEP values of 0.9613/0.9761 and 5.65/4.48% m/m, respectively. Similarly, the PLSRM14 and PLSRM15 that were applied for SPP quantification over the ranges of 1493.042–409.9746 and 3012.6–2782.1, 1786.755–1672.533 and 1493.042–409.9746 cm^{-1} showed r^2C/r^2V values of 0.9719/0.995 and 0.9747/0.9954, respectively. The latter two models showed RMSEC/RMSEP values of 4.79/3.37% m/m and 4.57/3.28% m/m, respectively. Hence, PLSR models showed high accuracy and precision when based on binary mixtures, especially in the range of 10–100% m/m. Models based on the ternary mixture (MN2), which included PLSRM16-PLSRM21, showed accurate quantifications for glucose with r^2C/r^2V values in the ranges of 0.9631–0.9889 and 0.7963–0.9657, respectively, and RMSEC/RMSEP values in the ranges of 3.93–7.04/6.32–9.77% m/m, respectively. Thus, out of these latter models, PLSRM21 showed the closest RMSEP/RMSEC values and r^2C/r^2V values and confirmed the reproducibility of the model. On the other hand, quantification of BSA and SPP based on ternary and quaternary mixtures did not yield accurate or precise results. Moreover, quaternary models did not yield accurate nor precise results for glucose quantification. This, in turn, could be related to the low number of samples compared to the variables assessed and that often results in overfitting of the PLSR models.[59]

Relative carbohydrate, lipid, and polyphosphate contents

Relative carbohydrate content was determined by the absorption area ratio of the carbohydrates (1200–950 cm^{-1}) band to the amide I band (referred to as carbohydrates: amide I). Relative lipids content was determined by the absorption area ratio of the lipids band (1740 cm^{-1}) to the amide I band (referred to as lipids: amide I). Relative polyphosphate (PP) content was determined by the absorption area ratio of the band 1240 cm^{-1} to the amide I band (referred to as polyphosphate: amide I).

One way ANOVA test was performed to compare the means of the carbohydrate: amide I ratio and lipid: amide I ratio at the log and stationary phases, followed by Tukey *post Hoc* test at $P = 0.05$ to pinpoint any differences when using IBM SPSS statistics 21 (Fig. 10). The aim was to find out which sample had the highest relative carbohydrates content or relative lipids content with a significant difference, along with the other samples in the same phase. The results showed that Jo-23 has the highest mean of carbohydrates: amide I ratio, with a significant difference among the other isolates in the log phase. While in the stationary phase, Jo-29 had the highest mean of carbohydrates: amide I ratio, with significant differences from those of Jo-2, Jo-4, Jo-18, Jo-34, and Jo-40 (Fig. 10(a)). The high carbohydrate content in the biomass suggests that it can be used as a fermentation feedstock for bioethanol production.[60] In comparison with lignocellulosic plants, the microalgae biomass has a lower content of hemicellulose and lignin.[61]

For the lipids: amide I ratio at the log phase, Jo-23 had the highest mean of lipid: amide I ratio with a significant difference among the other isolates. While in the stationary phase, there was an increase in the relative lipids content (except for Jo-23), but there was no significant difference in the increase between the samples. High lipid content in the biomass suggests that biomass can be used for biodiesel production.[62]

The samples were not normally distributed for the relative polyphosphate content, so the median was compared using Kruskal-Wallis Test to see if there was a significant difference at

Journal of Chemical Technology and Biotechnology published by John Wiley & Sons Ltd on behalf of Society of Chemical Industry (SCI).

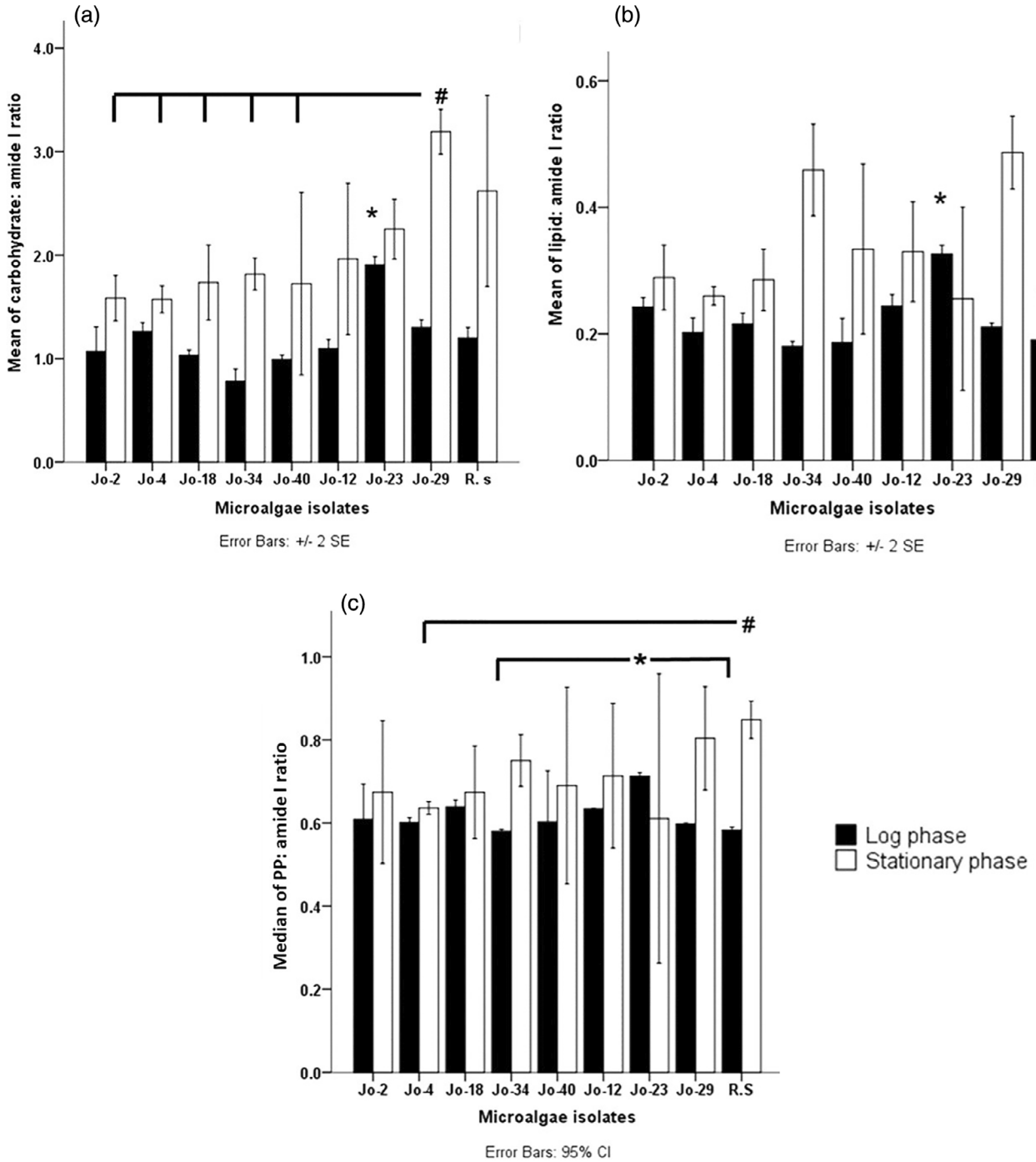


Figure 10. Statistical analysis for the mean and the median, and One way ANOVA test to compare the mean of (a) the carbohydrate: amide I ratio at the log phase and the stationary phase; and (b) the lipid: amide I ratio at the log phase and the stationary phase, followed by Tukey *post Hoc* test at $P = 0.05$, error bars represent the standard error (±SE) and n = 3. (c) Kruskal-Wallis Test at $P = 0.05$ for the PP: amide I at log phase and stationary phase. Asterix (*) denotes the difference at the log phase, the number sign (#) denotes the difference at the stationary phase.

$P = 0.05$; the difference among the samples was pinpointed using multiple comparisons (stepwise step-down) by IBM SPSS statistics 21. In the log phase, Jo-23 had the highest PP: amide I ratio and it was significantly different from Jo-34 and the reference strain. In the stationary phase, the reference strain had the highest PP: amide I ratio and was only significantly different from Jo-4.

In this research, there is special interest in the relative polyphosphate content because phosphorus is an essential element that plants need as a nutrient to grow and one that also is used in varied industries. Since phosphate rock is non-renewable, it is crucial to find ways of recycling the element. Therefore, microalgae biomass can be used to recover and reuse the phosphorus from wastewater and prevent its adverse effects on the water and

Journal of Chemical Technology and Biotechnology published by John Wiley & Sons Ltd on behalf of Society of Chemical Industry (SCI).

environment;[63] in addition, the microalgal biomass can be used as a slow-releasing fertilizer.[16]

CONCLUSIONS

The findings of the study showed the feasibility of portable ATR-FTIR spectroscopy for the qualitative and quantitative determination of algal samples. When used in portable form, the technique was shown to be rapid, non-destructive, and specific to determining algal samples and their constituents. The method showed high accuracy and precision when using both qualitative and quantitative approaches. Moreover, when FTIR was combined with PCA, clustering was seen among samples, depending on the growth phase for each species.

Microalgae are applied for wastewater treatment in the log phase to ensure their active growth and efficient nutrient removal. As was mentioned earlier, the microalgal biomass characteristics vary depending on the environmental conditions and water characteristics. Portable FTIR offers a rapid method for evaluating and monitoring the main characteristics of biomass when applying microalgae for wastewater treatment. This will help in determining the fate of the biomass.

ACKNOWLEDGEMENTS

The authors would like to thank the University of Bath (UK) for offering the equipment, consumables, and algal growing facility, as well as the Al-Balqa Applied University (Jordan) and the Ministry of Higher Education and Scientific Research (Jordan) for providing the PhD studentship. ESystem Engineering Society (eSES) is acknowledged for support with machine learning analytics. The authors would also like to pay tribute to the late Professor Rod Scott, who passed away before publishing this work, for his major contributions in developing and supporting the idea of this research.

CONFLICT OF INTEREST

All the authors declared that there are no conflicts of interest in publishing this article.

REFERENCES

1 von Alvensleben N, Magnusson M and Heimann K, Salinity tolerance of four freshwater microalgal species and the effects of salinity and nutrient limitation on biochemical profiles. *J Appl Phycol* **28**:861–876 (2016). https://doi.org/10.1007/s10811-015-0666-6.
2 Arora N, Kumari P, Kumar A, Gangwar R, Gulati K, Pruthi PA *et al*., Delineating the molecular responses of a halotolerant microalga using integrated omics approach to identify genetic engineering targets for enhanced TAG production. *Biotechnol Biofuels* **12**:2 (2019). https://doi.org/10.1186/s13068-018-1343-1.
3 Kim M, Oh HJ, Nguyen K and Jin E, Identification and characterization of *Dunaliella salina* OH214 strain newly isolated from a saltpan in Korea. *Algae* **37**:317–329 (2022). https://doi.org/10.4490/algae.2022.37.9.13.
4 Chi Z, Zheng Y, Jiang A and Chen S, Lipid production by culturing oleaginous yeast and algae with food waste and municipal wastewater in an integrated process. *Appl Biochem Biotechnol* **165**:442–453 (2011). https://doi.org/10.1007/s12010-011-9263-6.
5 Daneshvar E, Zarrinmehr MJ, Koutra E, Kornaros M, Farhadian O and Bhatnagar A, Sequential cultivation of microalgae in raw and recycled dairy wastewater: microalgal growth, wastewater treatment and biochemical composition. *Bioresour Technol* **273**:556–564 (2019). https://doi.org/10.1016/j.biortech.2018.11.059.
6 Li K, Liu Q, Fang F, Luo R, Lu Q, Zhou W *et al*., Microalgae-based wastewater treatment for nutrients recovery: a review. *Bioresour Technol* **291**:121934 (2019). https://doi.org/10.1016/j.biortech.2019.121934.
7 García-Galán MJ, Arashiro L, Santos LHMLM, Insa S, Rodríguez-Mozaz S, Barceló D *et al*., Fate of priority pharmaceuticals and their main metabolites and transformation products in microalgae-based wastewater treatment systems. *J Hazard Mater* **390**:121771 (2020). https://doi.org/10.1016/j.jhazmat.2019.121771.
8 Sutherland DL, Park J, Heubeck S, Ralph PJ and Craggs RJ, Size matters – microalgae production and nutrient removal in wastewater treatment high rate algal ponds of three different sizes. *Algal Res* **45**:101734 (2020). https://doi.org/10.1016/j.algal.2019.101734.
9 Zhou W, Li Y, Min M, Hu B, Chen P and Ruan R, Local bioprospecting for high-lipid producing microalgal strains to be grown on concentrated municipal wastewater for biofuel production. *Bioresour Technol* **102**:6909–6919 (2011). https://doi.org/10.1016/j.biortech.2011.04.038.
10 Isdepsky A and Borowitzka MA, In-pond strain selection of euryhaline *Tetraselmis sp*. strains for reliable long-term outdoor culture as potential sources of biofuel and other products. *J Appl Phycol* **31**:3359–3370 (2019). https://doi.org/10.1007/s10811-019-01873-y.
11 Ganesan R, Manigandan S, Samuel MS, Shanmuganathan R, Brindhadevi K, Lan Chi NT *et al*., A review on prospective production of biofuel from microalgae. *Biotechnol Rep (Amst)* **27**:e00509 (2020). https://doi.org/10.1016/j.btre.2020.e00509.
12 Siddiki SYA, Mofijur M, Kumar PS, Ahmed SF, Inayat A, Kusumo F *et al*., Microalgae biomass as a sustainable source for biofuel, biochemical and biobased value-added products: an integrated biorefinery concept. *Fuel* **307**:121782 (2022). https://doi.org/10.1016/j.fuel.2021.121782.
13 Enzing C, Ploeg M, Barbosa M and Sijtsma L, Microalgae-based products for the food and feed sector: an outlook for Europe, JRC scientific and policy reports, in *Institute for Prospective Technological Studies, Joint Research Centre, European Commission, Luxembourg*, ed. by Vigani M, Parisi C and Cerezo ER. European Commission, Luxembourg, pp. 7 (2014). https://doi.org/10.2791/33391.
14 Yaakob Z, Ali E, Zainal A, Mohamad M and Takriff MS, An overview: biomolecules from microalgae for animal feed and aquaculture. *J Biol Res (Thessalon)* **21**:6 (2014). https://doi.org/10.1186/2241-5793-21-6.
15 Uysal O, Uysal FO and Ekinci K, Evaluation of microalgae as microbial fertilizer. *Eur J Sustain Dev* **4**:77–82 (2015). https://doi.org/10.14207/ejsd.2015.v4n2p77.
16 Coppens J, Grunert O, Van Den Hende S, Vanhoutte I, Boon N, Haesaert G *et al*., The use of microalgae as a high-value organic slow-release fertilizer results in tomatoes with increased carotenoid and sugar levels. *J Appl Phycol* **28**:2367–2377 (2016). https://doi.org/10.1007/s10811-015-0775-2.
17 Dineshkumar R, Kumaravel R, Gopalsamy J, Sikder MNA and Sampathkumar P, Microalgae as bio-fertilizers for Rice growth and seed yield productivity. *Waste Biomass Valorization* **9**:793–800 (2018). https://doi.org/10.1007/s12649-017-9873-5.
18 Santhakumaran P, Ayyappan SM and Ray JG, Nutraceutical applications of twenty-five species of rapid-growing green-microalgae as indicated by their antibacterial, antioxidant and mineral content. *Algal Res* **47**:101878 (2020). https://doi.org/10.1016/j.algal.2020.101878.
19 Torres-Tiji Y, Fields FJ and Mayfield SP, Microalgae as a future food source. *Biotechnol Adv* **41**:107536 (2020). https://doi.org/10.1016/j.biotechadv.2020.107536.
20 Cuellar-Bermudez SP, Aguilar-Hernandez I, Cardenas-Chavez DL, Ornelas-Soto N, Romero-Ogawa MA and Parra-Saldivar R, Extraction and purification of high-value metabolites from microalgae: essential lipids, astaxanthin and phycobiliproteins. *J Microbial Biotechnol* **8**:190–209 (2015). https://doi.org/10.1111/1751-7915.12167.
21 Mehariya S, Goswami RK, Karthikeysan OP and Verma P, Microalgae for high-value products: a way towards green nutraceutical and pharmaceutical compounds. *Chemosphere* **280**:130553 (2021). https://doi.org/10.1016/j.chemosphere.2021.130553.
22 Laurens LML, Van Wychen S, McAllister JP, Arrowsmith S, Dempster TA, McGowen J *et al*., Strain, biochemistry, and cultivation-dependent measurement variability of algal biomass composition. *Anal Biochem* **452**:86–95 (2014). https://doi.org/10.1016/j.ab.2014.02.0092.
23 Rosenberg JN, Kobayashi N, Barnes A, Noel EA, Betenbaugh MJ and Oyler GA, Comparative analyses of three *chlorella* species in response to light and sugar reveal distinctive lipid accumulation pat-

Journal of Chemical Technology and Biotechnology published by John Wiley & Sons Ltd on behalf of Society of Chemical Industry (SCI).

terns in the microalga *C. Sorokiniana*. *PloS One* **9**:e92460 (2014). https://doi.org/10.1371/journal.pone.0092460.
24 Tibbetts SM, Milley JE and Lall SP, Chemical composition and nutritional properties of freshwater and marine microalgal biomass cultured in photobioreactors. *J Appl Phycol* **27**:1109–1119 (2015). https://doi.org/10.1007/s10811-014-0428-x.
25 Kobayashi N, Noel EA, Barnes A, Watson A, Rosenberg JN, Erickson G *et al*., Characterization of three *Chlorella sorokiniana* strains in anaerobic digested effluent from cattle manure. *Bioresour Technol* **150**: 377–386 (2013). https://doi.org/10.1016/j.biortech.2013.10.032.
26 Juneja A, Ceballos R and Murthy G, Effects of environmental factors and nutrient availability on the biochemical composition of algae for biofuels production: a review. *Energies* **6**:4607–4638 (2013). https://doi.org/10.3390/en6094607.
27 George B, Pancha I, Desai C, Chokshi K, Paliwal C, Ghosh T *et al*., Effects of different media composition, light intensity and photoperiod on morphology and physiology of freshwater microalgae *Ankistrodesmus falcatus* – a potential strain for bio-fuel production. *Bioresour Technol* **171**:367–374 (2014). https://doi.org/10.1016/j.biortech.2014.08.086.
28 Esakkimuthu S, Krishnamurthy V, Wang S, Hu X, Swaminathan K and Abomohra AE, Application of p-coumaric acid for extraordinary lipid production in *Tetradesmus obliquus*: a sustainable approach towards enhanced biodiesel production. *Renew Energy* **157**:368–376 (2020). https://doi.org/10.1016/j.renene.2020.05.005.
29 Ahmad SFK, Lee KT and Vadivelu VM, Emerging trends of microalgae bio-granulation research in wastewater treatment: a bibliometric analysis from 2011 to 2023. *Biocatal Agric Biotechnol* **102684**: 102684 (2023). https://doi.org/10.1016/j.bcab.2023.102684.
30 Mayers JJ, Flynn KJ and Shields RJ, Rapid determination of bulk microalgal biochemical composition by Fourier-transform infrared spectroscopy. *Bioresour Technol* **148**:215–220 (2013). https://doi.org/10.1016/j.biortech.2013.08.133.
31 Kim D-Y, Vijayan D, Praveenkumar R, Han J-I, Lee K, Park J-Y *et al*., Cell-wall disruption and lipid/astaxanthin extraction from microalgae: *chlorella* and *Haematococcus*. *Bioresour Technol* **199**:300–310 (2016). https://doi.org/10.1016/j.biortech.2015.08.107.
32 Ferro L, Gojkovic Z, Gorzsás A and Funk C, Statistical methods for rapid quantification of proteins, lipids, and carbohydrates in nordic microalgal species using ATR–FTIR spectroscopy. *Molecules* **24**:3237 (2019). https://doi.org/10.3390/molecules24183237.
33 Arif M, Li Y, El-Dalatony MM, Zhang C, Li X and Salama E-S, A complete characterization of microalgal biomass through FTIR/TGA/CHNS analysis: an approach for biofuel generation and nutrients removal. *Renew Energy* **163**:1973–1982 (2021). https://doi.org/10.1016/j.renene.2020.10.066.
34 Grace CE, Mary MB, Vaidyanathan S and Srisudha S, Response to nutrient variation on lipid productivity in green microalgae captured using second derivative FTIR and Raman spectroscopy. *Spectrochim Acta A Mol Biomol Spectrosc* **270**:120830 (2022). https://doi.org/10.1016/j.saa.2021.120830.
35 Dean AP, Sigee DC, Estrada B and Pittman JK, Using FTIR spectroscopy for rapid determination of lipid accumulation in response to nitrogen limitation in freshwater microalgae. *Bioresour Technol* **101**: 4499–4507 (2010). https://doi.org/10.1016/j.biortech.2010.01.065.
36 Deore P, Beardall J, Palacios YM, Noronha S and Heraud P, FTIR combined with chemometric tools — a potential approach for early screening of grazers in microalgal cultures. *J Appl Phycol* **33**:2709–2722 (2021). https://doi.org/10.1007/s10811-021-02543-8.
37 Wang H, Zhang W, Chen L, Wang J and Liu T, The contamination and control of biological pollutants in mass cultivation of microalgae. *Bioresour Technol* **128**:745–750 (2013). https://doi.org/10.1016/j.biortech.2012.10.158.
38 Wagner H, Liu Z, Langner U, Stehfest K and Wilhelm C, The use of FTIR spectroscopy to assess quantitative changes in the biochemical composition of microalgae. *J Biophotonics* **3**:557–566 (2010). https://doi.org/10.1002/jbio.201000019.
39 Jebsen C, Norici A, Wagner H, Palmucci M, Giordano M and Wilhelm C, FTIR spectra of algal species can be used as physiological fingerprints to assess their actual growth potential. *Physiol Plant* **146**: 427–438 (2012). https://doi.org/10.1111/j.1399-3054.2012.01636.x.
40 Bajhaiya AK, Dean AP, Driver T, Trivedi DK, Rattray NJW, Allwood JW *et al*., High-throughput metabolic screening of microalgae genetic variation in response to nutrient limitation. *Metabolomics* **12**:9 (2016). https://doi.org/10.1007/s11306-015-0878-4.
41 Laurens LML and Wolfrum EJ, Feasibility of spectroscopic characterization of algal lipids: chemometric correlation of NIR and FTIR spectra with exogenous lipids in algal biomass. *Bioenergy Res* **4**:22–35 (2011). https://doi.org/10.1007/s12155-010-9098-y.
42 Assi S, Khan I, Edwards A, Osselton D and Al-Obaidi H, On-spot quantification of modafinil in generic medicines purchased from the internet using handheld Fourier transform-infrared, near-infrared and Raman spectroscopy. *J Anal Sci Technol* **11**:35 (2020). https://doi.org/10.1186/s40543-020-00229-3.
43 Kim S, Kano M, Nakagawa H and Hasebe S, Estimation of active pharmaceutical ingredients content using locally weighted partial least squares and statistical wavelength selection. *Int J Pharm* **421**:269–274 (2011). https://doi.org/10.1016/j.ijpharm.2011.10.007.
44 Kvalheim OM, Chan H, Benzie IFF, Szeto Y, Tzang AH, Mok DK *et al*., Chromatographic profiling and multivariate analysis for screening and quantifying the contributions from individual components to the bioactive signature in natural products. *Chemom Intel Lab Syst* **107**:98–105 (2011). https://doi.org/10.1016/j.chemolab.2011.02.002.
45 He R, Ma T, Gong M, Xie K, Wang Z and Li J, The correlation between pharmacological activity and contents of eight constituents of Glycyrrhiza uralensis Fisch. *Heliyon* **9**:e14570 (2023). https://doi.org/10.1016/j.heliyon.2023.e14570.
46 Ministry of Water and Irrigation – Jordan, *Climate Change Policy for a Resilient Water Sector*. Amman, p. 1 (2016).
47 Ministry of Water and Irrigation – Jordan, *Water Substitution and Reuse Policy*. Amman, pp. 1–4 (2016).
48 Sweiss MA, *Microalgae for Wastewater Treatment and Biomass Production from Bioprospecting to Biotechnology*. Ph.D. dissertation. University of Bath, UK, pp. 73–143 (2017).
49 Delrue F, Álvarez-Díaz PD, Fon-Sing S, Fleury G and Sassi J-F, The environmental biorefinery: using microalgae to remediate wastewater, a win-win paradigm. *Energies* **9**:1–19 (2016). https://doi.org/10.3390/en9030132.
50 Jones CS and Mayfield SP, Algae biofuels: versatility for the future of bioenergy. *Curr Opin Biotechnol* **23**:346–351 (2012). https://doi.org/10.1016/j.copbio.2011.10.013.
51 Jee RD, Near-infrared spectroscopy, in *Clarke's Analysis of Drugs and Poisons*, 3rd edn, ed. by Moffat AC, Osselton MD, Widdop B and Watts J. Pharmaceutical Press, London, pp. 346–357 (2004).
52 Brereton RG, *Chemometrics: Data Analysis for the Laboratory and Chemical Plant*. John Wiley & Sons, West Sussex, pp. 183–317 (2003).
53 Burns DA and Ciurczak EW, *Handbook of near-Infrared Analysis*, 3rd edn. CRC press, Florida, p. 594 (2007).
54 Assi S, Moorey P, Kouris N, Kneller P and Osselton D, Identification of counterfeit tobacco using Fourier transform infrared spectroscopy, in *Meeting of the London Toxicology Group*. London Toxicology Group, London (2014).
55 de Carvalho JC, Magalhães AI, de Melo Pereira GV, Medeiros ABP, Sydney EB *et al*., Microalgal biomass pretreatment for integrated processing into biofuels, food, and feed. *Bioresour Technol* **300**: 122719 (2020). https://doi.org/10.1016/j.biortech.2019.122719.
56 Stuart BH, Organic molecules, in *Infrared Spectroscopy: Fundamentals and Applications*. John Wiley & Sons Ltd, West Sussex, UK, pp. 71–93 (2004).
57 Giordano M, Kansiz M, Heraud P, Beardall J, Wood B and McNaughton D, Fourier transform infrared spectroscopy as a novel tool to investigate changes in intracellular macromolecular pools in the marine microalga Chaetoceros muellerii (Bacillariophyceae). *J Phycol* **37**:271–279 (2001). https://doi.org/10.1046/j.1529-8817.2001.037002271.x.
58 Fyvie M, McCall JA and Christie LA, Towards explainable metaheuristics: PCA for trajectory mining in evolutionary algorithms, in *Artificial Intelligence XXXVIII: 41st SGAI International Conference on Artificial Intelligenceproc. SGAI-AI 2021*, ed. by Bramer M and Ellis R. Springer International Publishing, Cambridge, UK, pp. 89–102 (2021). https://doi.org/10.1007/978-3-030-91100-3_7.
59 Andersen CM and Bro R, Variable selection in regression—a tutorial. *J Chemometr* **24**:728–737 (2010). https://doi.org/10.1002/cem.1360.
60 Baker A, Ceasar SA, Palmer AJ, Paterson JB, Qi W, Muench SP *et al*., Replace, reuse, recycle: improving the sustainable use of phosphorus by plants. *J Exp Bot* **66**:3523–3540 (2015). https://doi.org/10.1093/jxb/erv210.
61 Harun R, Danquah MK and Forde GM, Microalgal biomass as a fermentation feedstock for bioethanol production. *J Chem Technol Biotechnol* **85**:199–203 (2009). https://doi.org/10.1002/jctb.2287.

62 Lakshmikandan M, Murugesan AG, Wang S, Abomohra AE-F, Jovita PA and Kiruthiga S, Sustainable biomass production under CO2 conditions and effective wet microalgae lipid extraction for biodiesel production. *J Clean Prod* **247**:119398 (2020). https://doi.org/10.1016/j.jclepro.2019.119398.

63 Sanchez Rizza L, Sanz Smachetti ME, Do Nascimento M, Salerno GL and Curatti L, Bioprospecting for native microalgae as an alternative source of sugars for the production of bioethanol. *Algal Res* **22**:140–147 (2017). https://doi.org/10.1016/j.algal.2016.12.021.

APPENDIX

Nominal versus predicted of models based on mixtures MN1-3.

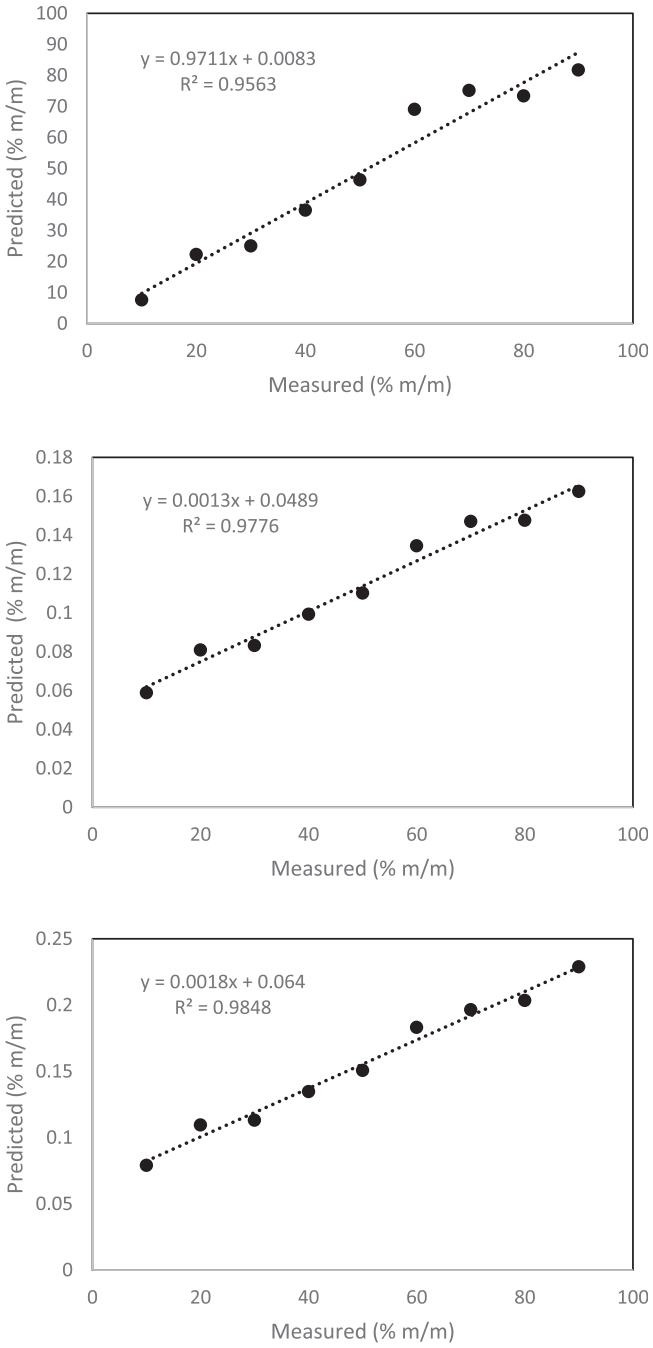


Figure A1. Univariate regression of binary mixture (glucose quantification) at 3212, 1048, and 1012 cm^{-1} (top to bottom).

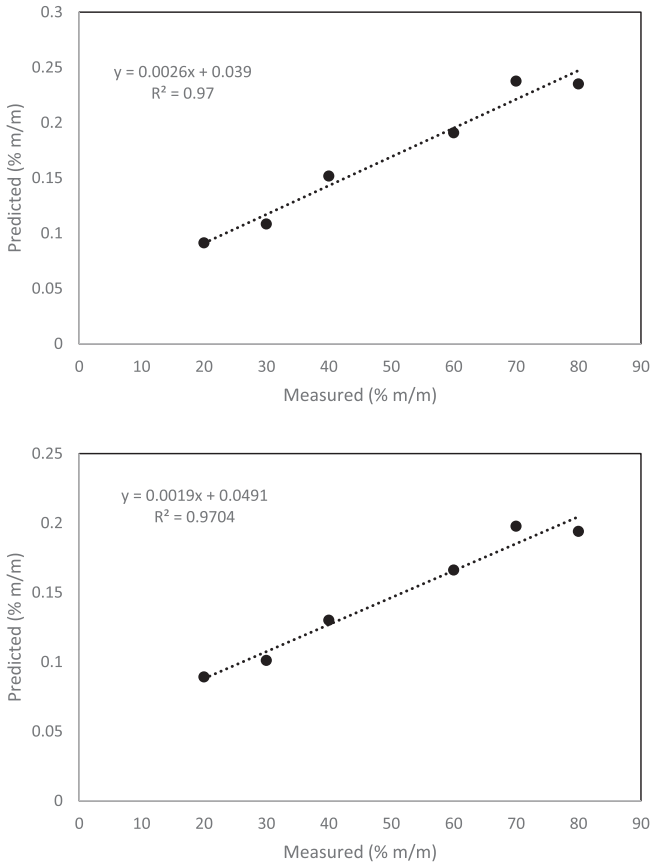


Figure A2. Univariate regression of tertiary mixture (glucose quantification) at 993 and 1012 cm^{-1} (top to bottom).

Journal of Chemical Technology and Biotechnology published by John Wiley & Sons Ltd on behalf of Society of Chemical Industry (SCI).

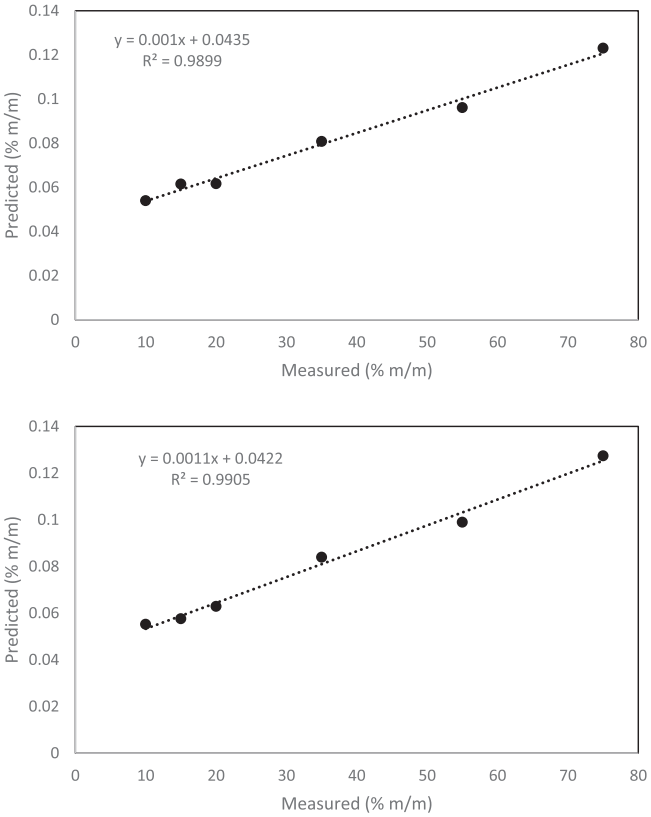


Figure A3. Univariate regression of tertiary mixture (BSA quantification) at 1644 and 1515 cm^{-1} (top to bottom).

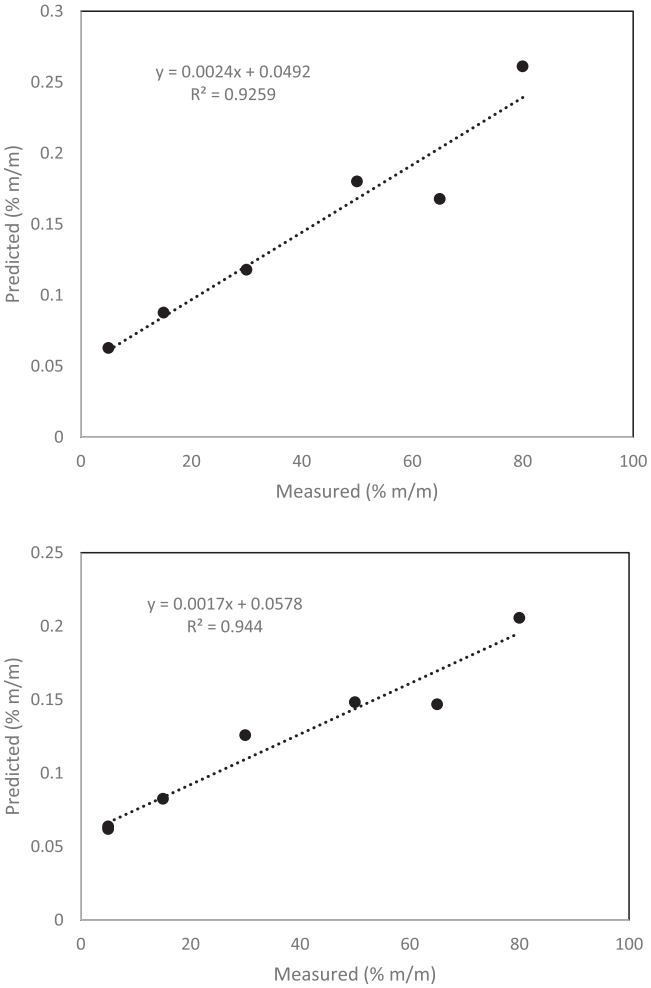
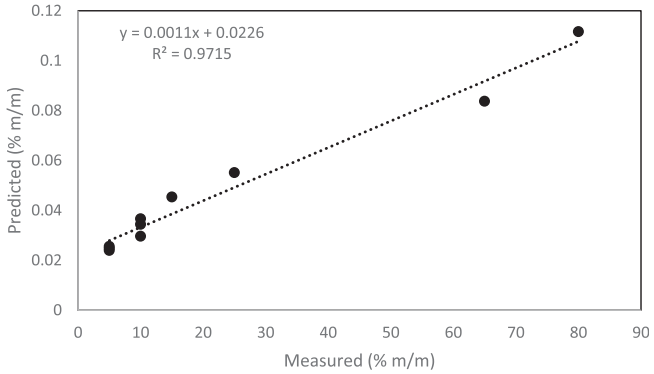


Figure A4. Univariate regression of quaternary mixture (glucose quantification) at 993 and 1020 cm^{-1} (top to bottom).

Journal of Chemical Technology and Biotechnology published by John Wiley & Sons Ltd on behalf of Society of Chemical Industry (SCI).



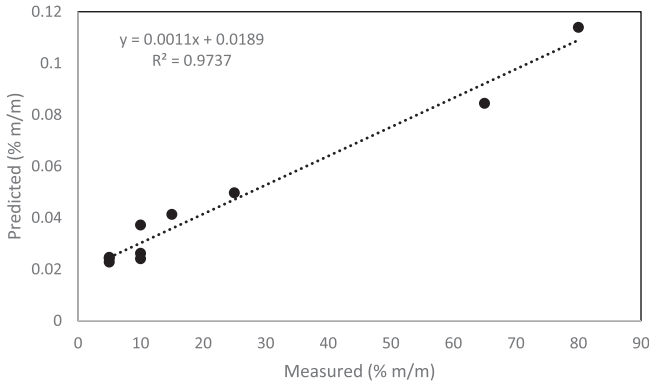
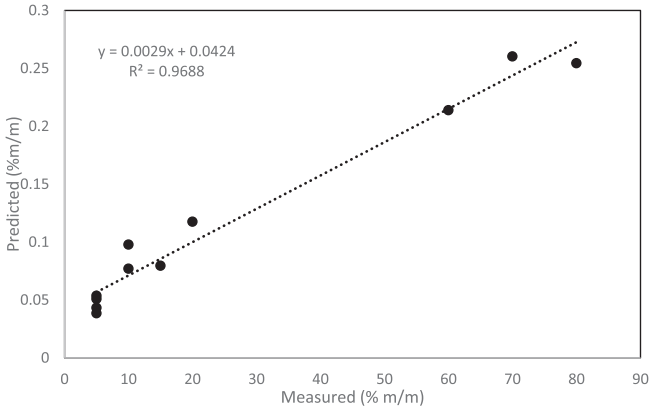


Figure A5. Univariate regression of quaternary mixture (BSA quantification) at 1650 and 1515 cm^{-1} (top to bottom).



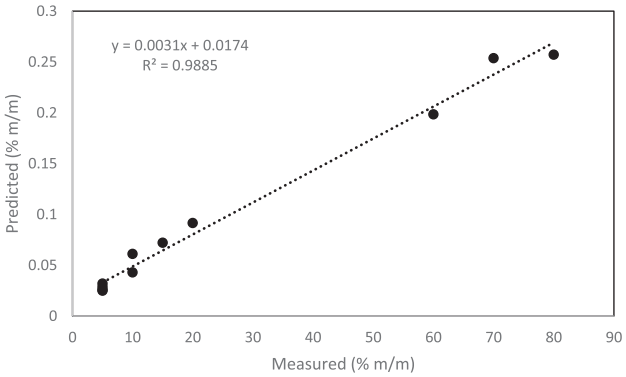


Figure A6. Univariate regression of quaternary mixture (GTP quantification) at 2849 and 1730 cm^{-1} (top to bottom).

Journal of Chemical Technology and Biotechnology published by John Wiley & Sons Ltd on behalf of Society of Chemical Industry (SCI).