Detection of the adulteration of fresh coconut water via NMR spectroscopy and chemometrics†

Paul I. C. Richardson,a Howbeer Muhamadali,a,b Yang Lei,a Alexander P. Golovanov,a,b David I. Ellis a,b* and Royston Goodacre a,b*

Here, we applied NMR spectroscopy in combination with chemometrics to quantify the adulteration of fresh coconut water, stretched with water–sugar mixtures. Coconut water was extracted from young Costa Rican coconuts and adulterated with concentrations of various sugar solutions. A total of 45 samples were analysed by 1D proton NMR spectroscopy and chemometrics. Results showed highly sensitive quantification, with a limit of detection of adulteration with sugars of 1.3% and a root-mean-squared error of prediction of 0.58%. Interestingly, we identified a regular drift in the chemical shift and a change in the lineshape of malic acid signals concomitant with increasing levels of adulteration. On further investigation, this was found to originate from changes in the concentration of divalent cations, such as magnesium, within the samples. It can be concluded that 1H NMR spectroscopy enables accurate quantification for the degree of adulteration in this product, with the added discovery finding that the shift and lineshape of the malic acid signal can be utilised as a potential diagnostic marker for partial substitution of fresh coconut water with extrinsic components such as sugar mixtures.

1. Introduction

Coconut water is a refreshing beverage obtained by extracting the liquid endosperm from immature (6–9 months) coconut fruits (Cocos nucifera). While all parts of the fruit and tree have been utilised throughout history within their native tropical regions for a multitude of purposes including building, horticultural substrate,1 or emergency ersatz intravenous drip,2,3 the high mineral content in coconut water along with its suggested health benefits have led to the drink being advertised as a natural alternative to isotonic sports drinks. These factors have contributed to a rapid expansion in the coconut water market, with sales in the UK reaching over £100 M in 2016;4 a 20-fold increase since 2012.

The sudden popularity of this product can lead to the risk of demand outstripping supply, as coconut palms require several years of maturation before they are able to bear fruit.5 Furthermore, with the vast majority of coconut water coming from only five countries worldwide (Brazil, Sri Lanka, Indonesia, Philippines and Thailand), these factors can result in coconut water being vulnerable to food fraud.6,7

The characterisation of coconut water has been undertaken using several different methods. High-performance liquid chromatography (HPLC) and liquid chromatography mass spectrometry (LC-MS) have been used to analyse multiple classes of phytohormones,8 and headspace solid phase micro-extraction gas chromatography (HS-SPME-GC) used to characterise the volatile organic profile. Nuclear magnetic resonance (NMR) spectroscopy combined with chemometric methods has also been applied to analyse and monitor the effect of different industrial processing methods.9 However, the novelty of coconut water and recent popularity has meant that the detection of fraud in this product has not as yet been studied in detail, with only one study published to date.10

Moreover, fraudulent activity involving coconut water products has already been observed.11–13 With an investigation in late 2017 by the National Food Crime Unit of the UK Food Standards Agency (FSA) showing that at least 14 major coconut water products contained undeclared sugars.14 These were detected using isotopic ratio mass spectrometry (IRMS)15 and site-specific natural isotope fractionation NMR (SNIF-NMR).16 While these are powerful and versatile method for the absolute detection of adulteration with sugars,17 they require lengthy preparation times (24 & 72 hours respectively), and are thus poorly suited as screening methods.

Here, we used 1H NMR and chemometric methods to detect product dilution masked with a normalised sugar profile sub-
stition. Fresh coconut water was adulterated at various concentrations using a buffered sugar solution emulating the natural concentrations of glucose, fructose, and sucrose present in our sample. In order to achieve our aim for a screening method, minimal sample processing was performed. Furthermore, we investigated a systematic drift observed for the signals assigned to malic acid with increasing levels of substitution with sugar solution, which revealed its origin in the changing metal coordination state of malate caused by this form of adulteration. This effect on malic acid signal shift and lineshape has the potential to be used as a diagnostic marker for tampering with coconut water products.

2. Methodology

Preparation of coconut water

Seven immature coconuts (6–9 months) originating from Costa Rica were obtained from a UK-based online retailer. The juice was extracted from each of these using a specialised Cocodrill® device obtained from the same retailer, centrifuged at 18 000g, for 10 min at 4 °C and pooled together to make a consistent stock solution. The pooled solution was then stored at −80 °C in 45 mL aliquots until required. Prior to use, the coconut water was thawed at room temperature and separated into 1 mL aliquots, which were heat-treated using a TechneDri-Cooler solution (0.0503 ppm containing the main peaks (these increments) each with a total volume of 2 mL. Each sample was spiked with 50 µL of a 4,4-dimethyl-4-silapentane-1-sulfonic acid (DSS) (Sigma-Aldrich) in D2O solution (Sigma-Aldrich) (98.17 mg mL−1) to provide chemical shift reference as well as a signal for radiofrequency lock. Finally, prior to analysis, samples were divided into three 0.6 mL aliquots. Excluding one sample at 75% adulteration that was rendered unusable due to human error, all prepared samples were analysed using 1H NMR spectroscopy, resulting in a total of 62 spectra for this experiment.

NMR analysis

Proton 1D spectra were acquired at 20 °C using a Bruker 800 MHz NMR AVANCE III spectrometer (Bruker, Coventry, UK) equipped with a 5 mm TCI cryoprobe with temperature control unit using standard pulse program “zgesp” from the Bruker library. A relaxation delay of 1.4 s and an acquisition time of 1.47 s were used with 16 scans, and exponential window function with 1 Hz line broadening was applied prior to Fourier transformation in Topspin 3.5 (Bruker). All NMR spectra were referenced to the DSS peak at 0.00 ppm.

Addition of Mg2+ ions to a sodium malate solution

A 11.88 mg mL−1 stock solution of malic acid (Sigma-Aldrich) was made, as well as a 0.5 M solution of MgCl2 (Fisher Scientific). Buffered malate solution, meant to approximate that found in coconut water, was made by combining 2896 µL of the malic acid stock solution, 824 µL 1 M KH2PO4 solution, 26 µL 1 M K2HPO4 solution, 563 µL 1 M KOH (Fisher Scientific) solution and 5711 µL H2O (3.40 mg mL−1 malate,18 pH: 5.32). The stock solution was then spiked with 25 µL DSS solution. An 800 µL aliquot of malate solution was added to an NMR tube and analysed using NMR as a control sample, then charged with 5 µL of the MgCl2 solution. It was analysed again, then sequentially charged with 10 µL aliquots of MgCl2 solution five times (end volume of MgCl2 added: 55 µL). This final solution was diluted in a malate-free buffer solution (stock: 824 µL 1 M KH2PO4 solution, 26 µL 1 M K2HPO4 solution, 9150 µL H2O) to 75, 50, 25, 12.5, and 6.25% concentration(s), and each of these was analysed using 1H NMR spectroscopy. A total of 12 spectra were obtained.

Data analysis

Fourier transformed NMR spectra were imported to Matlab and truncated to include only the spectral region between 5.6004 and −0.0503 ppm containing the main peaks (these comprised 13 277 data points per spectrum). Baseline correction was then performed using a de-trending algorithm, spectra were-scaled using the peak intensity of internal reference signal of DSS, and the spectral region containing water signal around 4.8 ppm removed. When required, the processed spectra were further divided into two spectral sections at 3.2 ppm. Principal components analysis (PCA) was employed to reduce the dimensionality within the spectral datasets into uncorrelated principal components (PCs) and describe relationships between samples. A linear multivariate predictive model, used for quantification, was also obtained using partial least squares regression (PLSR).19 The spectral dataset was divided into a training-set containing spectra for adulteration levels starting at 0% in 10% increments of increasing concentration (0%, 10%, 20%, 30%, up to 100%) and a test-set containing spectra for adulteration levels starting at 5% in 10%
increments of increasing concentration (e.g. 5%, 15%, 25%, 35%, up to 95%). Several predictive models were created using the training-set, with each of the models using an increasing number of latent variables (LVs). During calibration each model (in the training set only) was then tested using a $k$-fold cross-validation ($k = 20$) algorithm, and the optimal model was chosen by aiming to minimise the number of LVs along with the root-mean-squared error of cross-validation (RMSECV). Once chosen, the predictive model was evaluated by challenging the model with the test-set.

To investigate the relationship between the chemical shift of malic acid signals and adulteration, maxima locations for the strongest peaks in the signal at 2.7 ppm were found using the Matlab “findpeaks” command. These data were then imported into Microsoft Excel and a least-squares linear regression algorithm was used to acquire a line of best fit. Due to the lack of malic acid in the 100% adulteration samples, these spectra were not included in these calculations.

3. Results and discussion

Although the full spectral range measured during experiments was 12 to $-3$ ppm, the vast majority of strong signals were contained within 5.4 to 0 ppm. To minimise the effect of noise and simplify further calculations, spectra were therefore truncated to an effective spectral range of 5.6 to $-0.05$ ppm. Furthermore, the high concentrations of sugar in coconut water relative to its other constituents led to these signals being orders of magnitude more intense and subsequently overpowering statistical models. To remedy this and obtain a clear understanding of the effect of variance within other coconut water constituents, spectra were further divided into two subsections: the first containing the signals associated with sugars (5.6 to 3.2 ppm) and the second containing signals associated with other coconut water constituents (3.2 to $-0.05$ ppm). Both of these sections were then analysed with PCA (Fig. 1).

PCA scores plot for the coconut water constituents region (Fig. 1A) showed very little variance within adulteration levels – indicating that replicate measurements were highly reproducible. While PC 1 shows a clear and constant trend from positive to negative scores with increasing adulteration, PC 2 shows a “boomerang effect”, with the scores increasing until approximately 50% adulteration and a gradual decrease to return to their initial negative scores as the level of adulteration increases.

Loadings in PC 1 (Fig. S1B†) closely resemble the NMR spectra (Fig. S1A†), indicating that the trend is an overall decrease in intensity, with the only peaks not present being those ascribed to the internal DSS standard at 0.00, 0.63, 1.75, 2.91, and 3.14 ppm. This trend is expected, as virtually all the signals in this region are ascribed to coconut water, and increasing the level of adulteration will consequently decrease the intensity of those peaks. The loadings of PC 2 (Fig. S1C†) however, show the main cause for this boomerang effect to be due to the two signals at $\sim 2.70$ and $\sim 2.45$ ppm assigned to malic acid.20 Along with a decrease in intensity, a decrease in

Fig. 1  PCA scores plots for adulteration of coconut water with our mixed sugar solution. To obtain a meaningful representation of the variance within different regions of the spectral range, spectra were truncated into two sections prior to analysis: a sugar region (5.00–3.20 ppm) containing higher intensity peaks originating from glucose, sucrose and fructose; and a coconut region (3.20 to $-0.05$ ppm) containing signals originating from other components of coconut water. A: PCA scores plot for the coconut region, obtained with two PCs and 89.7% total explained variance (EV). B: PCA scores plot for the sugar region, obtained with two PCs and 88.4% TEV. The inset colour bars represent the percentage level of adulteration.
chemical shift occurs with increasing levels of adulteration as well as signal sharpening (Fig. 4). For both signals, this is reflected in the loadings by negative weighting at the more de-shielded part of the NMR spectra in samples with lower adulteration levels, and a positive weighting with the effect of decreasing intensity of the signals.

PCA scores plot for the sugar signal section (Fig. 1B) also shows a clear trend related to adulteration, this time along a composite axis combining PC 1 and PC 2, allowing for a clear separation between different levels of adulteration. Unlike the coconut water constituents section, however, the sugar signals section shows significant variance within classes, indicating that, predictably, the detection is not as obvious. Nonetheless, the ability to detect even small discrepancies readily in the sugar profile using NMR is promising. Interestingly, a further examination of the loadings plot allows for the discrimination of signals for each sugar and provides further details concerning the imperfections in our masking of dilution. Fig. 2 provides an enhanced view of the signals associated with glucose, fructose and sucrose (4.3–3.2 ppm). Although the overlaid spectra (Fig. 2A) show a high signal density in this area, the loadings plot of PC 1 (Fig. 2B) combined with the NMR spectrum of each sugar allows for the assignment of peaks and thus a description of the imperfections in masking. For example, our mixed solution was found to contain slightly less glucose and fructose, while containing more sucrose.

Moreover, utilising this information could potentially allow for the differentiation between natural variation in the sugar profiles of coconut water samples and the intentional addition of sugars.

Quantitative analysis

In addition to exploratory analysis, a quantitative predictive model was created using partial least-squares regression (PLSR). PLSR was performed on both sugar signals and coconut water constituent subsections, along with the full effective spectral range. Plots of each model are presented in Fig. 3, while a summary of the overall statistics of PLS in terms of linearity, error and detection limits are presented as Table 1.

It is immediately clear upon observing the resulting plots (Fig. 3) that the combination of NMR and PLSR allow for the creation of a powerful predictive model. Both individual subsections, along with the full effective range, showed high accuracy within both the test and cross-validation sets even at low levels of adulteration and despite the naturally occurring sugar profile being emulated within the adulterant.

The strength of this analytical model is further confirmed by the various statistical results (Table 1). Based on the $R^2$ and $Q^2$ values for the cross-validation and test sets consistently reaching $>0.99$, these data demonstrated a very good fit to the linear predictive models. Additionally, the RMSE on the test sets are all within 1%, indicating that the models can quanti-
tatively predict the level of adulteration within 5%. Two models are presented for the coconut water region, one using 3 LVs and one using 5 LVs. While the optimal model can generally be discerned from the plot comparing the RMSECV to the number of latent variables used in each model (data not shown), both could be viewed as acceptable without the requirement for supplementary testing. It is also interesting to compare the prediction error of the whole spectral range to that of the individual regions; while one would intuitively expect to see either an average of the two or, similarly to the PCA scores, a peak intensity-based weighting leading to equivalent results to the sugar section, the combined sections seem to create a more powerful model than its individual components (with the provision that the 3 LV model is correct for the coconut water section). A possible reason for this may be a higher number of peaks that can be used to validate the model’s predictions, thus lowering random uncertainty. While increasing the number of data points used in the spectra also often increases the contribution of background noise, the signal-to-noise ratio in these spectra is so high that this drawback is not significant. Furthermore, the limits of detection (LODs), which were calculated using a comparison of the predicted and real values on the test set with a false positive/negative rate of 0.05, were found to be within 2%, reaching 0.6% for the 5LV coconut constituent region model. Given the results of each model, we recommend the use of the entire dataset (5.6–0 ppm) rather than individual parts to ensure that all factors relevant to adulteration are taken into account.

Previously, we presented Raman spectroscopy for the first time as an effective screening method for the detection of adulteration in coconut water, as it is an inexpensive, portable and very fast method, requiring virtually no sample preparation.10 Using this vibrational spectroscopic method, we showed it was possible to detect adulteration by dilution and its masking with single sugars down to 1.9%. However, due to its reliance on discrepancies in the sugar profile to adulteration, the ability of Raman spectroscopy to detect fraudulent dilution of coconut water was hampered by normalising the individual concentrations of glucose, fructose, and sucrose throughout the adulteration. The use of NMR spectroscopy, though more time and resource intensive, would detect virtually all cases of intentional substitution while retaining the advantage of minimal sample preparation, as very low levels of adulteration would likely not be commercially worth the risk inherent to the practice.

### Table 1

PLSR results for the optimal models of the sugar region, coconut region, and the combined spectral range. LVs represents the number of latent variables used in each model, $R^2$ and $Q^2$ represent the goodness of the model’s fit to the dataset, while trn, CV and tst denote results pertaining to the training, cross-validation and test set, respectively. The limit of detection (LOD) is also presented. For the coconut (CW) region, the results of two models with different number of LVs used in model construction are presented as both seemed equally viable.

<table>
<thead>
<tr>
<th>Method</th>
<th>LVs</th>
<th>$R^2$</th>
<th>$Q^2$ (CV)</th>
<th>$Q^2$ (tst)</th>
<th>RMSE (trn, %)</th>
<th>RMSE (CV, %)</th>
<th>RMSE (tst, %)</th>
<th>LOD (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sugar region</td>
<td>3</td>
<td>0.9992</td>
<td>0.9991</td>
<td>0.9991</td>
<td>0.8863</td>
<td>0.7810</td>
<td>0.8416</td>
<td>1.9836</td>
</tr>
<tr>
<td>CW region</td>
<td>3</td>
<td>0.9994</td>
<td>0.9889</td>
<td>0.9991</td>
<td>0.7631</td>
<td>0.8426</td>
<td>0.8515</td>
<td>1.6790</td>
</tr>
<tr>
<td>CW region</td>
<td>5</td>
<td>0.9999</td>
<td>0.9995</td>
<td>0.9999</td>
<td>0.2993</td>
<td>0.5341</td>
<td>0.2974</td>
<td>0.6513</td>
</tr>
<tr>
<td>Full range</td>
<td>4</td>
<td>0.9998</td>
<td>0.9995</td>
<td>0.9996</td>
<td>0.4948</td>
<td>0.5889</td>
<td>0.5767</td>
<td>1.3418</td>
</tr>
</tbody>
</table>

**Fig. 3** PLSR results plots comparing the predicted actual concentrations for the coconut region (A), sugar region (B) and combined spectral range (C). For Fig. 3A, the model with 3 LVs was used.
chemical shifts of two visible peaks in the 2.7 ppm range depicted a linear drift (Fig. 4D & E).

Three hypotheses were postulated and tested to explain the potential source(s) of this behaviour: (i) slight variations in pH; (ii) the variation of concentration of malic acid itself; and (iii) variation of divalent cation concentrations leading to differences in malate metal coordination state. The pH hypothesis was tested by altering the pH of coconut water aliquots using 1 M solutions of NaOH and HCl (Fig. S3†), while the concentration hypothesis was tested by diluting a buffered 3.4 mg mL\(^{-1}\) solution of sodium malate, emulating the natural concentrations in coconut water (Fig. S4†). Neither of these led

---

**Fig. 4** A–C: \(^1\)H NMR spectra for the adulteration of coconut water with our mixed sugar solution ranging from 0% (blue) to 100% (red) adulteration, zoomed in on each peak assigned to malate. D–E: chemical shift of malic acid peaks plotted against level of adulteration (0% being pure coconut water, 100% being pure sugar solution) for the malate signal at ~2.7 ppm, along with regression results.

---

**Fig. 5** A: Overlaid \(^1\)H NMR spectra depicting the effect on the ~2.7 ppm malate signal multiplet of the incremental addition of aqueous Mg\(^{2+}\) (MgCl\(_2\)) to a buffered 3.44 mg mL\(^{-1}\) solution of sodium malate ranging from 0 mM Mg\(^{2+}\) (blue) to 31.43 mM Mg\(^{2+}\) (red). B: Plot comparing the drift in chemical shift of each peak in the malate signal multiplet relative to its corresponded initial position prior to addition of Mg\(^{2+}\) ions.
to any significant drift in the malate peaks and were therefore ruled out. Further details on these experiments can be found in the ESL.†

To test the divalent cation coordination hypothesis, separate aliquots of buffered sodium malate solution were spiked with increasing concentrations of zinc (in the form of ZnCl₂) and magnesium (in the form of MgCl₂), two metal cations naturally present in coconut water, and the drift in chemical shift of the malate signal at ∼2.7 ppm examined. All analysis was performed on this signal as it was the most sharply defined and contained a consistent double doublet; however, a similar drift was present in all three malate signals.

The presence of divalent cations led to a significant increase in chemical shift. The addition of MgCl₂ solution to the sodium malate solution led to the malate signal to shift to more positive values, which is in agreement with the effect observed with increasing concentrations of coconut water. Furthermore, based on the linear interpolation of the fit, the amplitude of the drift fits with that found in coconut water. The concentrations of the two main divalent cation constituents in coconut water, calcium and magnesium, add up to a total of ∼17 mM: similar addition of MgCl₂ would lead to an approximate shift of 0.017 ppm, based on the results presented in Fig. 5B, which matches the total magnitude of 0.017–0.022 ppm drift observed as a result of adulteration (Fig. S3†). The agreement between the shifts in adulteration experiments and separate malate titration with divalent metal ions provides strong evidence to defend our hypothesis that the signal shift is due to changed metal coordination state of malate. Furthermore, increasing the concentration of Zn²⁺ (Fig. S5†) to 10 mM led to a drift of 0.05 ppm in the chemical shift; although the concentration of zinc cations in coconut water is <0.1 mM, meaning the absolute shift as a result of the presence of Zn cations is likely not very significant, the presence of this effect provides further evidence for these cations being the source of the drift.

A final confirmation of our hypothesis was acquired through the dilution of the final MgCl₂-spiked sodium malate solution with buffer solution (Fig. 6). Although the change in chemical shift was less linear than that observed in the spiking experiments, the chemical shift did return close to its original position prior to spiking with divalent cations (Fig. 5A). Interestingly, while the source of the signal drift has now been accounted for, the relative signal broadening of malate peaks when the higher percentage of coconut water is present in the sample is currently unclear but can likely be explained by further malate binding to other yet unknown natural ingredients of coconut water. Importantly, malate signals serve as a sensitive reporter for the relative concentration of this genuine product in adulterated samples.

4. Conclusion

Coconut water is a refreshing and nutritious beverage which has gained huge popularity in recent years. This sudden increase in popularity and sales, however, has greatly increased the vulnerability and potential risk of illicit behaviours such as economically motivated adulteration (food fraud). Further, this relatively recent and sudden increase in popularity and sales of this product could be said to result in there not currently being a sufficient range and application of testing methods available to food regulatory bodies, which instead rely on IRMS and SNIF-NMR. Here, we proposed a combination...
of $^1$H NMR spectroscopy and chemometrics to create a powerful model which, importantly, requires virtually no sample preparation. Using this approach, we were able to detect substitution of coconut water with a sugar solution emulating the natural concentrations of glucose, fructose and sucrose (this drink’s major soluble solids) at levels as low as 1.3%, while being able to quantify it to within 0.6% error. Additionally, we investigated a regular linear drift observed in the chemical shift of malate peaks as the percentage of adulteration increased. While changes in pH and concentration of malate were not found to be significant, changes in the concentration of $^{2+}$ cations was found to cause a drift in our malate solution that is in complete agreement with our findings in coconut water spectra. Chemical shift being an absolute measurement, this property could be utilised as a diagnostic marker$^{21}$ of adulteration, indicative of the intentional substitution, or stretching, of coconut water with foreign, cheaper liquids.

**Conflicts of interest**

There are no conflicts of interest to declare.

**Acknowledgements**

DIE and RG thank the UK ESRC and UK FSA for funding (Food fraud: a supply network integrated systems analysis (grant number ES/M003183/1)).

**References**