


The Rametrix™ LITE Toolbox v1.0 for MATLAB®

Amanda K. Fisher¹ | William F. Carswell²  | Ahmad I.M. Athamneh³ |

Meaghan C. Sullivan² | John L. Robertson⁴ | David R. Bevan^{1,5} | Ryan S. Senger^{2,6} 

¹Genomics, Bioinformatics, and Computational Biology Interdisciplinary Program, Virginia Tech, Blacksburg, VA 24061, USA

²Department of Biological Systems Engineering, Virginia Tech, Blacksburg, VA 24061, USA

³Kindi Therapeutics and Drug Discovery, West Lafayette, IN 47906, USA

⁴Department of Biomedical Engineering and Mechanics, Virginia Tech, Blacksburg, VA 24061, USA

⁵Department of Biochemistry, Virginia Tech, Blacksburg, VA 24061, USA

⁶Department of Chemical Engineering, Virginia Tech, Blacksburg, VA 24061, USA

Correspondence

Ryan S. Senger, Department of Biological Systems Engineering, Virginia Tech, Blacksburg, VA 24061, USA.
Email: senger@vt.edu

Funding information

National Science Foundation, Grant/Award Number: NSF1254242

Abstract

To contribute to the growing interest in using Raman spectroscopy to analyze biological samples and provide chemometric analysis, we have developed a Raman Chemometrics (Rametrix™) Toolbox for use with MATLAB®. The LITE version of the Rametrix™ Toolbox is free to academic users through GitHub (<https://github.com/SengerLab/RametrixLITEToolbox>) and provides a graphical user interface for application of the following to Raman spectra: baseline correction with the Goldindex algorithm, vector or specific band normalization, principal component analysis (PCA), discriminant analysis of principal components (DAPC), identification of wavenumber loadings for PCA and DAPC, and calculation of total canonical distance. Raman spectroscopy and analysis with the Rametrix™ LITE Toolbox were applied to generate calibration curves, monitor enzymatic reactions, and track *Escherichia coli* culture growth. Results were quantitatively consistent with traditional methods of analysis. Additionally, the ability to distinguish urine specimens from healthy individuals and from patients receiving treatment for chronic kidney disease through peritoneal dialysis was demonstrated using PCA and DAPC of Raman spectra, suggesting future applications to detect or monitor progression of the disease. Overall, the Rametrix™ LITE Toolbox provides a streamlined application of PCA and DAPC chemometric techniques, and total canonical distance offers an additional quantitative measure to interpret Raman spectra of biological samples.

KEYWORDS

discriminate analysis of principal components, principal component analysis, Raman spectroscopy, total canonical distance

1 | INTRODUCTION

Raman spectroscopy is now used widely to monitor diverse biological applications, from contrasting lipids, nucleic acids, and collagen content of healthy versus cancer cells^[1] to identifying the chemical components of seeds, nuts, and oils.^[2] The strengths of Raman spectroscopy include short integration time (i.e., on the order of seconds), nondestructive measurement, simple sample preparation (i.e., no required dyes or chemical labels),

and highly detailed spectra.^[3–6] Its major challenges include interpretation of the resulting complex biological spectra, interference by fluorescence and cosmic events, and some instrument-to-instrument variability.^[7] Raman spectra of biological samples are often composed of many uncharacterized components in low and variable concentrations. Organic chemicals often have bands appearing at multiple wavenumbers, which can overlap with and enhance the intensity of those from other components.^[8] Interference from fluorescence by analyte or sample

impurities can impose a large background, and wavenumber bands may drift depending on the instrument and environment. For these reasons, assigning specific chemicals or molecular attributes to distinct wavenumber regions can be cumbersome, and such approaches have largely been supplemented by chemometric ones.^[9–11]

Chemometric approaches for interpreting complex biological spectra, such as the least squares regression-based technique biochemical component analysis^[12,13] and the more generally applicable technique of principal component analysis (PCA),^[9–11,14–16] allow comparative analysis of preprocessed spectra. PCA can be applied readily to any group of spectra, with or without prior knowledge of sample groups or classifications, and provides a visual representation of major similarities and differences among samples through clustering. PCA can be limited in its application to Raman spectra by its inability to provide assessment of groups or the relationships among groups of spectra.

Applying multivariate analysis of variance (MANOVA) to principal components arising from PCA is a more recent technique^[17] originally used to provide a faster and more robust alternative to Bayesian clustering algorithms when analyzing large datasets given by modern DNA sequencing technologies. Discriminant analysis of original variables focuses on between-group variability while neglecting within-group variation. Furthermore, the number of variables interpreted by MANOVA must be less than the number of samples given. In both Raman spectroscopy and sequencing data, MANOVA of Raman intensities at different wavenumbers or allele frequencies would require thousands of samples. Analyzing instead the principal components of the original variables discards any prior correlations and vastly reduces the number of variables provided to MANOVA while remaining information rich.

Discriminant analysis of principal components (DAPC) can be applied to a set of data classified into a priori groups based on experimental “factors” to provide a visual interpretation of the relationships among those groups. DAPC has been used successfully to interpret seasonal influenza hemagglutinin sequencing data, providing an obvious visual demonstration of a sudden change in allele frequency between seasons.^[17] It also has been used with Raman spectroscopy to characterize the phenotypic responses of *Escherichia coli* colonies exposed to various alcohol toxins.^[18,19] Unlike PCA alone, DAPC can interpret and use minor variances among data to arrange a priori groups into logical patterns, allowing correlation with independent variables.

When several principal components are used in DAPC, an equal number of “canonicals” (dimensions of DAPC) are formed. Dataset clustering in DAPC is often

visualized using a two- or three-dimensional plot, which is only representative of the first two or three canonicals.^[20–22] However, several more canonicals may exist and contain valuable information related to the separation of groups in the dataset. To capture all this multidimensional information into a single value for quantitative analysis, we have developed the concept of total canonical distance (TCD).^[23] The TCD is calculated using a standard distance formula across all canonical values between a reference group and an experimental group. The TCD calculation is shown in Equation 1, where j is the total number of canonicals generated from DAPC, C_i is the value of the i th canonical, and the subscripts *ref* and *exp* refer to reference and experimental groups, respectively.

$$TCD = \sum_{i=1}^j \sqrt{(C_{ref,i} - C_{exp,i})^2}. \quad (1)$$

The Raman Chemometrics (Rametrix™) LITE Toolbox (v1.0) for MATLAB® was developed to quickly process Raman spectra and apply PCA, DAPC, and TCD calculations and visualize data clustering. The Rametrix™ LITE Toolbox is freely available to academic users through GitHub and provides an easy and graphical implementation of (a) spectral viewing/comparison and trimming, (b) baseline correction using the Goldindex algorithm,^[24] (c) vector normalization or normalization to a specific band intensity, (d) PCA, (e) DAPC, and (f) TCD for Raman spectral data. Here, we demonstrate that application of these tools to Raman spectra from biological systems allows both qualitative and quantitative analyses of system dynamics, comparable with traditional methods of analysis, such as UV/vis spectroscopy and enzymatic assays. In particular, we conducted the following experiments, arranged in order of increasing complexity, to demonstrate the methodology: (a) the construction of a calibration curve for 2-nitrophenol, (b) determination of glucose concentrations by enzymatic assay, (c) the construction of a calibration curve for bovine serum albumin (BSA) through monitoring a Bradford assay reaction, and (d) periodic monitoring of *E. coli* culture growth. Finally, we applied the Rametrix™ LITE Toolbox to Raman spectra of urine specimens from healthy individuals and patients being treated with peritoneal dialysis for chronic kidney disease (CKD). Analysis of urine samples by Raman and spectral processing with the Rametrix™ LITE Toolbox allowed a fast, noninvasive analysis that revealed critical molecular differences in the urine composition of healthy individuals and CKD patients. These studies demonstrate the wide variety of uses for Raman spectroscopy and chemometric analysis using the Rametrix™ LITE Toolbox.

2 | MATERIALS AND METHODS

2.1 | Rametrix™ LITE Toolbox modules and statistical methods

The Rametrix™ LITE Toolbox (v1.0) for MATLAB® is freely available for academic users and can be obtained from GitHub at the following address: <https://github.com/SengerLab/RametrixLITEToolbox>. The Rametrix™ LITE Toolbox provides user-friendly graphical user interface modules (arranged in tabular format) to conduct computational processing and statistical methods. It is also capable of generating and exporting MATLAB® compatible figures at each step of the analysis described below.

2.1.1 | Start module: Loading and exporting data

The Rametrix™ LITE Toolbox accepts raw spectra stored in SPC spectral file format (*.spc), space-delimited text file format (*.txt), or comma-separated value format (*.csv). These file formats accommodate one scan per file and include wavenumber data (Column 1) and Raman signal intensity data (Column 2). The Rametrix™ LITE Toolbox will load all spectral files stored in a folder specified by the user. The Rametrix™ LITE Toolbox can also save loaded data as a Microsoft Excel® formatted worksheet in a Windows® operating system or in comma-separated values format. Finally, the user can elect to save and load data in Raman data analysis format (*.rda). This format retains all data as well as all user-specified parameters. The Rametrix™ LITE Toolbox also allows the user to identify “factors” associated with each spectrum. Factors can be related to type of treatment, date of scan, and so forth and can be used as the basis for grouping spectra in later analyses (e.g., PCA or DAPC). Factors are identified in the filename of each spectral file and are separated by an underscore (“_”) in the filename. An unlimited number of factors are possible, but each file being loaded into the Rametrix™ LITE Toolbox must have the same number of factors specified.

2.1.2 | Explore module: Preprocessing spectra

The Explore module of the Rametrix™ LITE Toolbox is used for viewing (and overlaying) Raman spectra as well as performing the preprocessing steps of baselining and normalization. The user can select one or multiple spectra and view raw, raw accompanied by baseline, baselined, or normalized spectra in the viewing pane. The user can also specify if one or multiple spectra should be excluded from further analysis due to the presence of cosmic spikes or excessive noise. The user has the option of truncating spectra to any desired wavenumber range and applying baselining using the Goldindec algorithm, selected for its tolerance of high peaks and high peak ratios.^[24] The baseline polynomial order, the estimated peak ratio, and the smoothing window size can be specified as parameters for the Goldindec algorithm. Spectral normalization can be performed relative to a specified Raman band (chosen wavenumber) or by vector normalization according to Equation 2, where y is the Raman intensity at each wavenumber in the truncated range, p is the range of wavenumbers in each spectrum, and y' is the vector normalized intensity at each wavenumber in the truncated range.

$$y' = \frac{y}{\sqrt{\sum_{j=1}^p y_j^2}} \quad (2)$$

For the implementation examples presented, all collected raw Raman spectra were truncated to the biological fingerprint range^[7] of 600–1,800 cm^{-1} . The spectra were baselined by correcting for fluorescence and background drift with a smoothing window size of 5. The polynomial order and peak ratio were adjusted for each experiment and are recorded in Table 1.

2.1.3 | PCA module

The PCA module allows the user to perform PCA on the dataset, generating one less principal component than the number of spectra provided, and view principal

TABLE 1 Goldindec algorithm parameters and input parameters for DAPC in each experiment

Experiment	Baseline polynomial order	Estimated peak ratio	Dataset variability explained by PCs
2-Nitrophenol calibration curve	5	0.35	93.9%
Glucose assay	6	0.4	99.2%
BSA concentration curve	5	0.5	95.4%
<i>Escherichia coli</i> growth	3	0.5	95.2%
Rametrix™ urinalysis	9	0.6	98.1%

Note. DAPC = discriminant analysis of principal components; PC = principal component.

component scores (spectra plotted along principal component axes) in two or three dimensions. The user is also given the option to assign data labels on the resulting plot, so outliers can be identified easily. Axes can be adjusted to correspond to different principal components. PCA is performed in the Rametrix™ LITE Toolbox using the “pca” function from the Statistics and Machine Learning Toolbox™ in MATLAB®.

2.1.4 | PC Contributions module: Wavenumber contribution to principal components

The PC Contributions module allows the user to identify the total contribution of each principal component to the total dataset variance as well as the influence of each wavenumber to that principal component. Results for several principal components can be visualized simultaneously in a viewing pane, and these results can be used to identify specific Raman bands (that may correspond with specific molecules) that give rise to variance in the dataset. Specifically, these results reveal which bands lead to separations in PCA clustering. These bands can be linked to individual molecules through the use of spectral libraries.

In general, principal components are linear combinations of Raman intensities at all wavenumbers of all spectra loaded in the Explore module. The fractional contribution, w , of each wavenumber to each principal component is calculated by Equation 3, where z is the loading for each wavenumber (as calculated for a specific principal component by the “pca” function in MATLAB®) and p is the range of wavenumbers in each spectrum.

$$w = \frac{z^2}{\sum_{j=1}^p z_j^2} \quad (3)$$

This calculation is carried out separately for every wavenumber and principal component. It provides a graphical interpretation of how each wavenumber intensity contributes to the variance in the dataset, accounted for by each principal component.

2.1.5 | DAPC module

The DAPC module allows the user to perform DAPC with respect to any dataset “factor” specified. The user also has the option to perform DAPC using either a specified number of principal components or the required number of principal components to represent a specified variability (% total) of the dataset. The user can view results on a two- or three-dimensional canonical plot and assign

specific canonicals to axes. The formation of data clusters on the canonical plot is indicative of similarity (to be shown in Section 3). The DAPC analysis is performed using the “manova1” function from the MATLAB® Statistics and Machine Learning Toolbox™ with principal component scores from the PCA module.

2.1.6 | Canonical Contributions module: Wavenumber contributions to canonicals

Similar to the PC Contributions module, the Canonical Contributions module applies Equation 3 to DAPC results to determine which wavenumbers give rise to separations seen in the canonical plot(s). Specifically, the fractional contribution of each principal component to each canonical is calculated (Equation 3) and multiplied by the original matrix of wavenumber contributions for principal components. The resulting matrix contains the fractional contribution of each wavenumber to each canonical. This analysis helps determine which wavenumber(s) drive the cluster separations in DAPC. These wavenumber(s) can then be attributed to specific molecules through the use of spectral libraries, such as the following reference for biological tissues.^[25] This module allows the user to check that the intended molecule(s) are giving rise to cluster separations in DAPC; otherwise, the chemistry of the system may be changing in an unanticipated way. When the mechanisms involved are unknown, the Canonical Contributions module can be used to discover how a system is changing.

2.1.7 | TCD module

TCD is applied in the TCD module of the Rametrix™ LITE Toolbox to further quantify DAPC results. A reference factor is selected, and the distance across all canonicals used in DAPC is calculated between that reference and every other factor group. The correlation between the TCD and independent variable values is then calculated, as well as a linear least squares equation of a fitted line through the data.

2.1.8 | Implementation

For the examples presented in this article, the Goldindex baseline algorithm parameters and the number of principal components used for DAPC analysis are given in Table 1. Each application of DAPC was provided with “factor” labeling of each spectrum based on the experimental design.

2.2 | Experimental methods

2.2.1 | Instrumentation

All Raman spectra were collected using an Agiltron (Woburn, MA) PeakSeeker PRO-785 Raman spectrometer utilizing a 100-mW, 785-nm laser with spot size 0.1–0.2 mm. Spectra spanned the 200- to 2,000-cm⁻¹ wavenumber range with resolution of 8 cm⁻¹. An integration time of 15 s was used on all samples, which were prepared for Raman measurement in 2-ml screw thread glass autosampler vials with 10-mm screw thread caps (Thermo Fisher; Waltham, MA).

All spectrophotometric measurements were taken with a Spectronic Genesis 10 Bio spectrophotometer (Thermo Fisher). Samples were prepared for spectrophotometric measurement in Plastibrand 1.5-ml disposable cuvettes 12.5 × 12.5 × 45 mm (Thermo Fisher).

2.2.2 | Calibration curves

2-Nitrophenol

2-Nitrophenol (Sigma-Aldrich; St. Louis, MO) was prepared in deionized (DI) water to final concentrations of 20 mM, 10 mM, 5 mM, 1,000 μM, 500 μM, and 100 μM. Two 1-ml samples of the 20-mM concentration and four 1-ml samples of all other concentrations were measured with the spectrophotometer at 420 nm and with the Raman spectrometer, using DI water used for spectrophotometer blank. The 20-mM concentrations were out of absorbance range of the spectrophotometer. A dark subtract was performed before each round of Raman measurements to remove noise caused by charge accumulation on the charge coupled device detector.

Glucose assay

Three samples each of 0, 4, 8, 12, and 16 μl of 1-mg/ml glucose in 200 μl of 1% benzoic acid standard were mixed with 400 μl of Sigma Glucose (GO) assay reagent (Sigma-Aldrich) containing glucose oxidase and peroxidase. The samples were incubated in a 37 °C water bath for 30 min before 400 μl of 12 N sulfuric acid was added to each sample to quench the enzymatic reaction. Samples were measured by Raman spectroscopy and by the spectrophotometer at 340 nm. The 0-mg/ml samples contained the assay reagent and were used as the spectrophotometer blank and a dark subtract before each round of Raman measurements to remove noise caused by charge accumulation on the charge coupled device detector.

Protein content

Three samples each of 0-, 50-, 100-, 250-, 500-, 750-, and 1,000-μg/ml purified BSA (New England Biolabs; Ipswich,

MA) were prepared in DI water and measured with Raman spectroscopy after dark subtract using a DI water blank. The samples, except for the 50-μg/ml sample (out of range concentration), were then evaluated by traditional Bradford assay^[26] using Coomassie Plus™ Protein Assay Reagent (Thermo Fisher) with spectrophotometric measurement at 595 nm. The 0-μg/ml BSA sample served as the spectrophotometer blank.

Bacterial growth

Chemically competent *E. coli* 10-β cells (New England Biolabs) were grown in 2-ml lysogeny broth overnight at 37 °C and 200 RPM. Four 2.5-ml samples were made with 250 μl of the culture and 9.75 ml of fresh lysogeny broth. A 1-ml aliquot of each sample was measured with the Raman spectrometer and on an Eppendorf BioPhotometer plus for OD₆₀₀ immediately after introduction of fresh media and every hour thereafter for 5 hr, while incubating the cultures at 37 °C and 200 RPM in between measurements.

RametriX™ urinalysis

Institutional review board-approved protocols were in place at Virginia Tech for this study. Urine specimens were collected from 20 healthy volunteers on the Virginia Tech campus and 31 patients undergoing peritoneal dialysis treatment for management of CKD from Fresenius Kidney Care Crystal Springs (Roanoke, VA). Specimens were frozen and stored at -30 °C for no longer than 4 weeks prior to analysis. Specimens were thawed and warmed to 37 °C, transferred to glass sample vials, and scanned by Raman with a 100-mW 785-nm laser with a 10-s integration time. Each sample was scanned 10 times.

3 | RESULTS

3.1 | The RametriX™ LITE Toolbox

The RametriX™ LITE Toolbox for MATLAB® was developed for streamlined applications of Goldindec algorithm baselining, normalization, PCA, DAPC, and TCD to sets of Raman spectra. The toolbox is organized into the following seven tabs. (a) Start: loading and saving files, (b) Explore: exploring and preprocessing spectra, (c) PCA: performing PCA on spectra, (d) PC Contributions: examining wavenumber contributions to principal components, (e) DAPC: performing DAPC, (f) Canonical Contributions: examining wavenumber contributions to canonicals, and (g) TCD: calculating TCD between experimental groups and a reference group. A full examination of the functionalities provided by the RametriX™ LITE Toolbox and recommendations for sample collection and preprocessing are available in Appendix S1.

3.2 | Experimental implementations

Modules of the Rametrix™ LITE Toolbox (including PCA, DAPC, and TCD) were applied to Raman spectra collected from common wet lab experiments designed to demonstrate the versatility and applicability of the approach. Results were compared with traditional methods of analysis for determining chemical concentration, protein content, and cellular density in liquid culture. A final experiment, referred to as Rametrix™ urinalysis, demonstrates that Raman spectroscopy with analysis by PCA can be used to detect differences in urine samples from healthy individuals and patients with CKD.

3.2.1 | 2-Nitrophenol calibration curve

To demonstrate the applicability of the Rametrix™ LITE Toolbox in MATLAB®, a simple 2-nitrophenol calibration curve was generated using (a) Raman measurements with PCA, DAPC, and TCD and (b) traditional spectroscopy using absorbance measurements at 420 nm. Four of the analysis methodologies included in the Rametrix™ LITE Toolbox are demonstrated in Figure 1: (a) baselining and normalization of raw Raman spectra (Figure 1a), (b)

PC wavenumber loadings (Figure 1b), (c) DAPC (Figure 1c), and TCD (Figure 1d). PCA results were also generated but are not shown in Figure 1. With PCA, samples aligned along principal component 1 (PC1) in order of ascending concentration. The PC1 accounted for 79.2% of the dataset variance, and PC2 accounted for 5.2%. PCA of Raman spectra was able to also incorporate measurements of 20 mM of 2-nitrophenol, which was beyond the scope of the spectrophotometer. To verify that the largest source of variability in the first principal component was the differing concentrations of 2-nitrophenol, the wavenumber loadings were evaluated by the PC Contribution module (Figure 1b) and compared with the expected wavenumber bands for 2-nitrophenol. The wavenumbers that contributed most to the first principal component were 821, 1032, 1142, 1,254 cm^{-1} , and the 1,285- to 1,375- cm^{-1} range. The 821- cm^{-1} band is commonly indicative of NO_2 in-plane angle bending, and the remaining bands indicate CC, CO, and NO stretching,^[27] all of which are characteristic of 2-nitrophenol structure. DAPC results (Figure 1c) provide less within-sample variability than observed with PCA (not shown) and align in order of ascending concentration along Canonical 1. Here, the calibration along Canonical 1 is nonlinear,

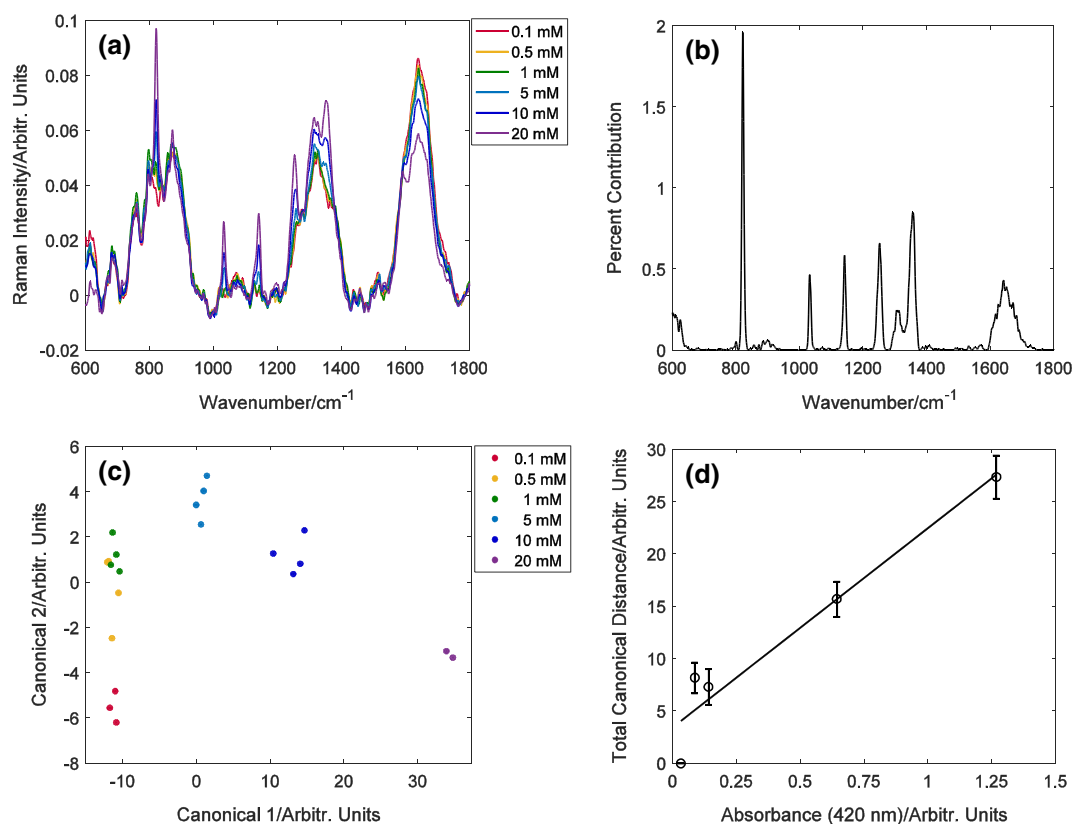


FIGURE 1 2-Nitrophenol calibration curves by absorbance and Raman spectroscopy. (a) Averaged Raman spectra, baselined with the Goldinddec algorithm and vector normalized. (b) Wavenumber loadings for the first principal component. (c) Discriminant analysis of principal components of Raman spectra. (d) Total canonical distance of each factor group plotted against each sample absorbance at 420 nm [Colour figure can be viewed at wileyonlinelibrary.com]

and considerable influence of Canonical 2 is observed at low concentrations. To quantify the sample concentration differences using DAPC, the TCD was calculated and plotted against spectrophotometric absorbance data (Figure 1d). A correlation coefficient (R^2) value of 0.94 was calculated (excluding the 20-mM samples), indicating high correlation between TCD of Raman spectra and spectrophotometric absorbance. Data are given in Table S1.

3.2.2 | Enzymatic assay for glucose

Next, an enzyme-based commercial glucose assay was run, and results were obtained by traditional absorbance measurements at 340 nm and Raman spectroscopy. This implementation is more complex than the previous example because enzymatic reactions were involved to measure glucose concentrations. The calibration curve based on absorbance (340 nm) is given in Figure 2a, and an R^2 value of 0.97 was obtained between glucose concentration and absorbance at 340 nm. PCA and DAPC results of the Raman approach are given in Figure 2c,d. Data clusters were arranged in ascending order along PC1 and PC2 and Canonical 1. The multivariate analysis technique employed by DAPC transformed the loosely grouped

PCA scores of the data into linearly arranged cohesive units, with little variance in Canonical 2 compared with the overall range of Canonical 1. Application of DAPC to Raman spectra allows recovery of logical progression from least to greatest concentration, which is not observable in any single peak location in the complex baselined and vector normalized Raman spectra shown in Figure 2d. Calculating the TCD (not shown) between DAPC groups and correlating with absorbance produced an R^2 value of 0.99 (excluding the 0- $\mu\text{g}/\text{mL}$ samples). Data are given in Table S2.

3.2.3 | Protein concentration

Calibration curves for BSA were obtained (a) by standard Bradford assays with absorbance measurements at 595 nm and (b) through Raman spectroscopic analysis and calculation of the TCD. The purpose was to demonstrate that the Raman methodology can distinguish differing concentrations of large, chemically complex molecules (i.e., proteins) as well as it does for small molecules (e.g., 2-nitrophenol). The calibration curve resulting from the standard Bradford assay and absorbance measurements at 595 nm is shown in Figure 3a. This analysis produced an R^2 of 0.99 between concentration and

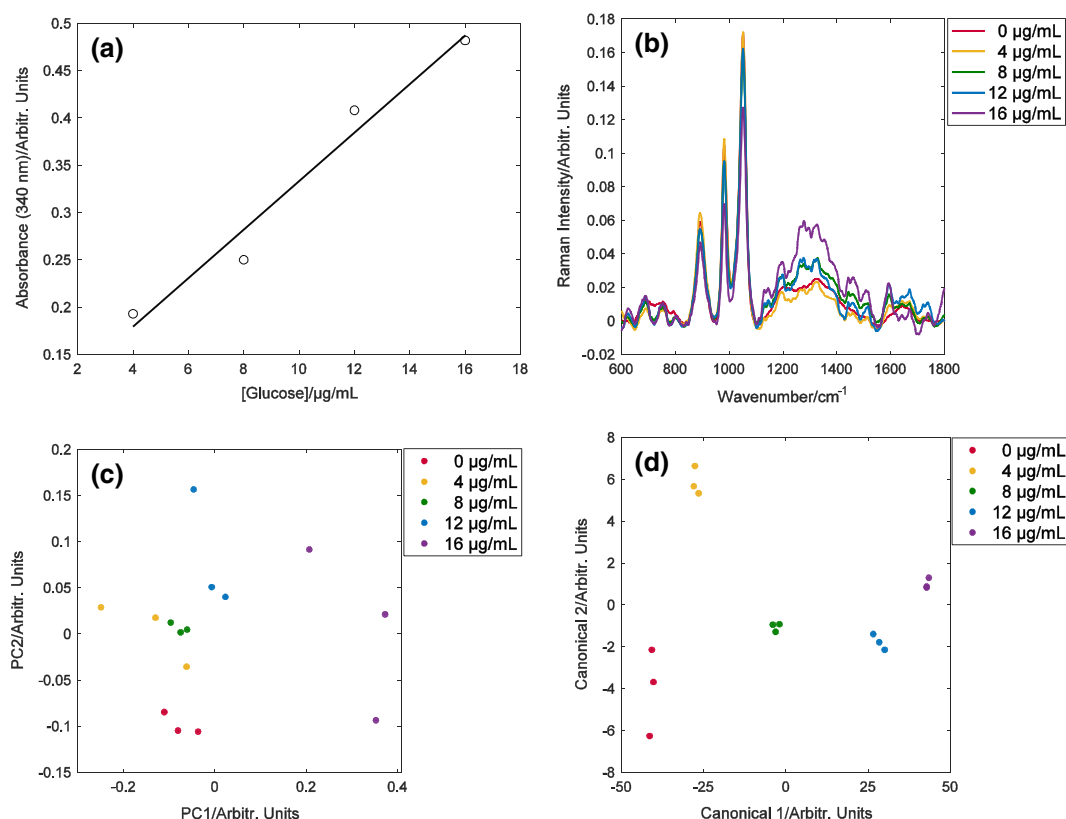


FIGURE 2 Glucose assay measured by absorbance and Raman spectroscopy. (a) Averaged glucose absorbances at 340 nm. (b) Averaged Raman spectra, baselined with Goldindex algorithm and vector normalized. (c) Principal component analysis of Raman spectra. (d) Discriminant analysis of principal components of Raman spectra [Colour figure can be viewed at wileyonlinelibrary.com]

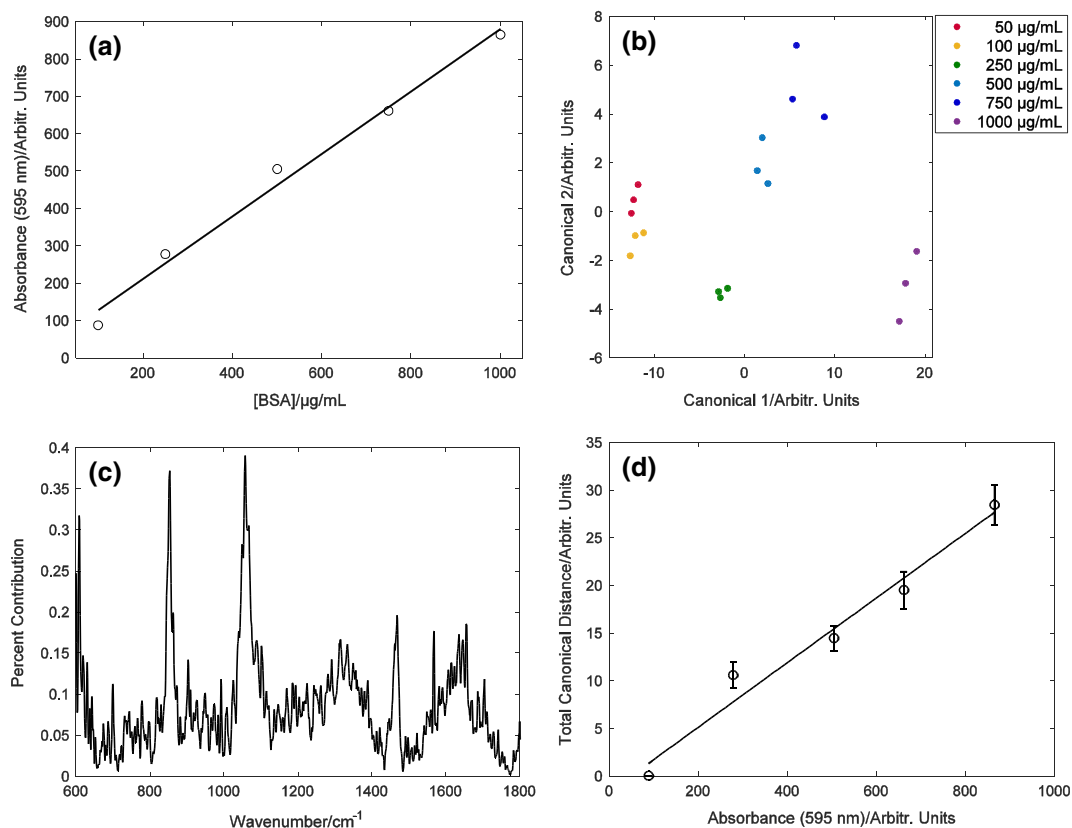


FIGURE 3 Bradford assay of bovine serum albumin (BSA) measured by absorbance and Raman spectroscopy. (a) Averaged BSA with Bradford reagent absorbance measurements at 595 nm with standard deviations. (b) Discriminant analysis of principal components of Raman spectra. (c) Wavenumber loadings of Canonical 1. (d) Total canonical distance of each factor group plotted against each sample absorbance at 595 nm [Colour figure can be viewed at wileyonlinelibrary.com]

absorbance. DAPC results of the Raman spectra taken before addition of Bradford reagent are shown in Figure 3b. The wavenumber loadings for the first canonical are given in Figure 3c and show four prominent bands around 600, 850, 1,060, and 1,470 cm⁻¹. These are most likely representative of disulfide bonds, tyrosine, C–N bonds, and CH₂ and CH₃ angle bending,^[28,29] all of which are common features of proteins, including BSA. TCD results were plotted against average absorbance of samples treated with Bradford reagent in Figure 3d. This analysis revealed a correlation R^2 value of 0.97 (excluding the 50-μg/ml samples). The raw TCD and absorbance data are given in Table S3.

3.2.4 | Monitoring microbial culture growth

Another biological application to test the robustness of the Raman methodology and Rametrix™ LITE Toolbox was monitoring *E. coli* liquid culture growth over time with both OD₆₀₀ and Raman spectroscopy. A typical growth curve was observed in the *E. coli* samples over the first 5 hr of incubation following inoculation when

measured by OD₆₀₀ (Figure 4a). PCA of Raman spectra (Figure 4b) demonstrated that the spectra are extremely variable and difficult to categorize before applying DAPC (Figure 4c), whereupon logical progression by time was observed, decreasing along Canonical 1. A large gap along Canonical 2 between the first and second hours of growth may be representative of the shift from lag to exponential growth phases. TCD results were plotted against OD₆₀₀ in Figure 4d, and an R^2 value of 0.92 was obtained for the correlation. The TCD and absorbance values are given in Table S4. To determine if the growth rate could still be ascertained by TCD, both TCD and OD₆₀₀ growth curves were normalized and used to calculate the growth rate. These data are also presented in Table S4, and very good agreement was observed whether OD₆₀₀ or TCD was used to calculate growth rate.

3.2.5 | Rametrix™ urinalysis

Raman spectroscopy and processing with the Rametrix™ LITE Toolbox in MATLAB® can replicate the qualitative analysis returned by a wide range of standard analytical techniques, but it can also provide qualitative information

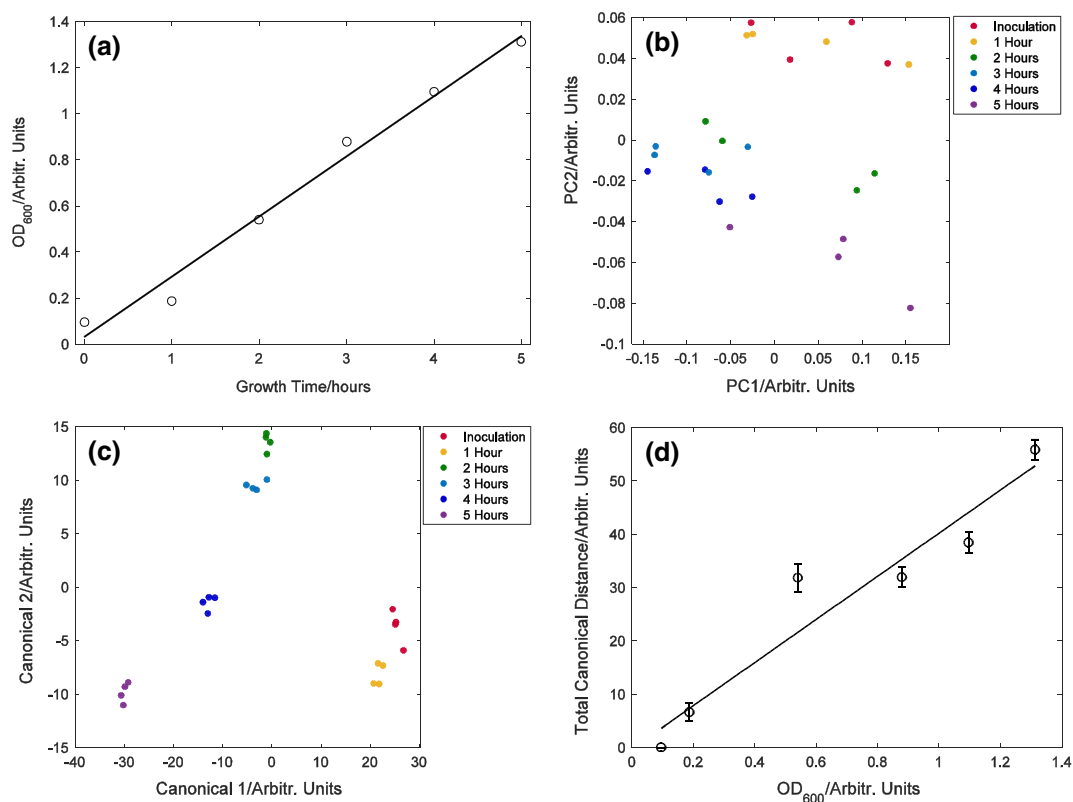


FIGURE 4 *Escherichia coli* liquid culture growth over time measured by OD₆₀₀ and Raman spectroscopy. (a) Averaged OD₆₀₀ with standard deviations. (b) Principal component analysis of Raman spectra. (c) Discriminant analysis of principal components of Raman spectra. (d) Total canonical distance of each factor group plotted against each sample OD₆₀₀ [Colour figure can be viewed at wileyonlinelibrary.com]

for complex samples that would otherwise require metabolomics analysis. To demonstrate this capability, urine specimens from healthy individuals and from CKD patients undergoing peritoneal dialysis treatments were analyzed by Raman spectroscopy and the spectra processed using the Rametrix™ LITE Toolbox. Spectra were first baselined and normalized (Figure 5a) and then compared by PCA (Figure 5b) and DAPC (Figure 5c). The separation of healthy individuals and CKD patients is obvious in PCA (Figure 5b), without the benefit of defined groups, but this separation became more defined by DAPC (Figure 5c). Urine represents a particularly complex matrix, with over 2,500 known chemical components,^[30] and the Raman spectra are accordingly complex. The greatest distinguishing sources of variability in the spectra between the healthy individuals and CKD patients were identified in the wavenumber loadings of Canonical 1 from DAPC (Figure 5d). The urea band at 1,003 cm⁻¹ is the major distinguishing feature, but the presence of many other bands in the wavenumber loadings indicate that there are many more features from which to extract patient health data. These have the potential to be previously unknown biomarkers of disease. Further detailed analysis could be conducted utilizing factors such as

disease state, age, sample collection time, and other sample qualifiers to track disease progression and give valuable insight into overall patient health.

4 | DISCUSSION

To provide a streamlined computational pipeline for processing Raman spectra and performing multivariate statistical analyses, such as PCA, DAPC, and TCD, we have developed and provide the Rametrix™ LITE Toolbox for MATLAB®, free to academic users. Here, we have demonstrated the use of the Rametrix™ LITE Toolbox and have shown how it can provide qualitative (i.e., separation of groups through DAPC) and quantitative analyses (i.e., recreating calibration curves with TCD). The quantitative analysis approach showed good agreement with standard analytical techniques involving spectrophotometric absorbance measurements. Although many of the examples given in this article are simple in nature, such as generating calibration and growth curves, the use of TCD to do this is a new concept. It is valuable because a Raman spectrum captures a snapshot of the chemical composition of a sample, which can then be deconvoluted to reveal

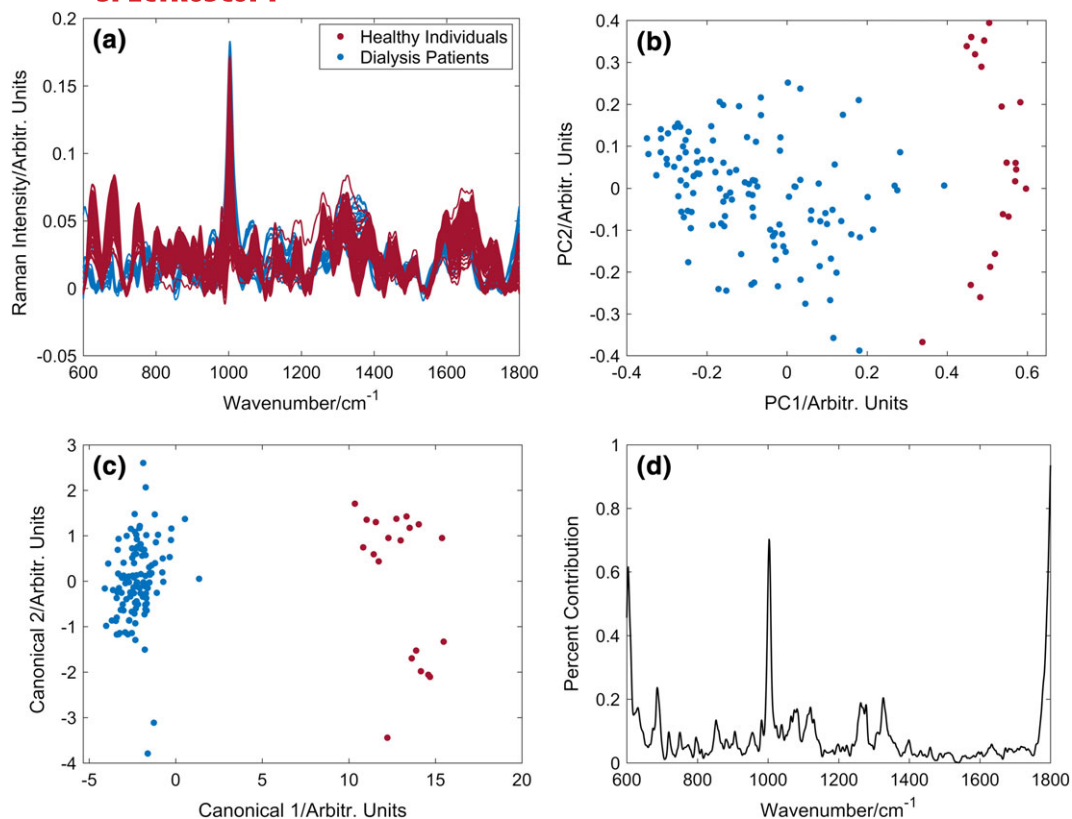


FIGURE 5 Rametrix™ urinalysis of healthy individuals and dialysis patients. (a) Averaged Raman spectra of different time points per individual, baselined with the Goldindex algorithm and vector normalized. (b) Principal component analysis of Raman spectra. (c) Discriminant analysis of principal components of Raman spectra. (d) Wavenumber loadings of Canonical 1 [Colour figure can be viewed at wileyonlinelibrary.com]

further information. For example, here, we have shown that Raman spectroscopy and calculation of TCD can generate a reliable growth curve for *E. coli*. Though OD₆₀₀ measurements are easily obtained, additional information is contained in the Raman spectra that can contribute to more robust analysis. We have shown previously that properties such as membrane fatty acid configuration, membrane fluidity, and amino acids content can be extracted from Raman spectra of *E. coli* cells growing under stressed conditions.^[18] Additionally, analysis of *E. coli* cultures by DAPC can distinguish among the *E. coli* phenotypes that arise from exposure to nonlethal doses of alcohols and different classes of antibiotics.^[19,22] Such changes in cellular phenotype arise from genetically driven stress responses and are accompanied by changes in chemical composition (i.e., metabolomics and lipidomics). With current technologies, the characterization of these chemical changes requires cell deconstruction and several types of analysis, many involving mass spectrometry. Although we may not yet be able to deconvolute the complicated Raman signal to decipher the chemical composition of a complex biological sample, this signal can be used along with multivariate statistics to

determine the similarities and differences of experimental groups. The observed differences may then be traced back to individual Raman bands, which can be identified using spectral libraries. The Rametrix™ LITE Toolbox enables this type of analysis through inclusion of the PC Contributions and Canonical Contributions modules. Although this was not demonstrated here for the case of *E. coli* culture growth, it was demonstrated for the calibration curves of 2-nitrophenol and BSA. For the case of *E. coli* culture growth, how would one know if observed changes were due to growth, as opposed to changes in pH or by-product accumulation? The PC and Canonical Contribution modules can answer this question. Notably, bands associated with major biological macromolecules (e.g., protein, fatty acids, and nucleic acids) should dominate the PC and Canonical Contributions modules if observed changes are due to culture growth. Raman bands for these molecules can be found in spectral libraries, including this one.^[25] Although results were not shown in detail here, we confirm this was the case for *E. coli* culture growth.

Along these lines, the Rametrix™ LITE Toolbox can be used to discover why (or how) a system changes when the mechanism(s) are unknown. This was true for the

example of CKD presented here. Nephrologists have located several useful clinical biomarkers of CKD to enable early diagnosis. However, it is likely that additional useful biomarkers exist that could provide additional information to clinicians. One or more of these may be responsible for the uncharacterized bands in Figure 5d. Further investigation into these, through the use of spectral libraries and/or mass spectrometry, may yield a novel biomarker. This illustrates one way that Raman spectroscopy and the Rametrix™ LITE Toolbox can be used for discovery.

Furthermore, Raman spectroscopy is an immensely flexible analytical technique that has now become inexpensive, reliable, and portable. The past decade has seen the development of high-quality handheld Raman devices, and the instrument used in the research presented here has the footprint of a laptop computer, which is in stark contrast to the expensive and cumbersome Raman instruments that were required of biological research just a short time ago. Raman spectroscopy also accommodates near real-time measurement of diverse samples collected in many different fields of research, without the need for extensive training, complex sample preparation, chemical labeling, or use of hazardous materials. Thus, the future of Raman-based analytics in biological research is expected to continue to grow, and the Rametrix™ LITE Toolbox for MATLAB® will provide a user-friendly means of processing spectra and performing multivariate statistical analysis to generate qualitative and quantitative results. Though almost all the replicated analytical techniques used here to demonstrate the application of Raman spectroscopy and the Rametrix™ LITE Toolbox were of a biological nature, the methodologies available in the Rametrix™ LITE Toolbox can be applied to any source of spectra, or nearly any data that can be provided in matrix format. We hope that by providing a centralized toolbox in a widely used mathematics platform, the use of Raman spectroscopy in evaluation of complex biological samples will continue to grow. Evaluation with PCA and DAPC distills the complex measurement of the chemical spectrum in a sample into visually digestible scatter plots for analysis, without sacrificing depth of information, and TCD has proven to provide a quantitative element to this analysis.

Although the Rametrix™ LITE Toolbox is freely available to academic users through GitHub, we are also constructing a Rametrix™ PRO Toolbox version. Although the capabilities of the LITE version have been highlighted in this article, the PRO version will be made available through license agreement and contains additional functions for sample classification and predictions. This is useful when dealing with samples of “unknown” origin or classification.

ACKNOWLEDGEMENTS

Funding for this research was obtained from the National Science Foundation, Award NSF1254242. R. S. S. and J. L. R. are cofounders and owners of DiallySensors, Inc., which intends to commercialize Rametrix™ technologies.

ORCID

William F. Carswell  <http://orcid.org/0000-0002-5112-6565>

Ryan S. Senger  <http://orcid.org/0000-0002-2450-6693>

REFERENCES

- [1] K. Kong, C. Kendall, N. Stone, I. Notingher, *Adv. Drug Delivery Rev.* **2015**, *89*, 121.
- [2] C. E. da Silva, P. Vandennebeele, H. G. Edwards, L. F. de Oliveira, *Anal. Bioanal. Chem.* **2008**, *392*, 1489.
- [3] A. F. Chrimes, K. Khoshmanesh, P. R. Stoddart, A. Mitchell, K. Kalantar-Zadeh, *Chem. Soc. Rev.* **2013**, *42*, 5880.
- [4] D. Drescher, J. Kneipp, *Chem. Soc. Rev.* **2012**, *41*, 5780.
- [5] P. Matousek, *Chem. Soc. Rev.* **2007**, *36*, 1292.
- [6] A. Molckovsky, L. M. Song, M. G. Shim, N. E. Marcon, B. C. Wilson, *Gastrointest. Endosc.* **2003**, *57*, 396.
- [7] T. Bocklitz, A. Walter, K. Hartmann, P. Rösch, J. Popp, *Anal. Chim. Acta* **2011**, *704*, 47.
- [8] T. Ishiyama, V. V. Sokolov, A. Morita, *J. Chem. Phys.* **2011**, *134*, 024509.
- [9] Y. H. Ong, M. Lim, Q. Liu, *Opt. Express* **2012**, *20*, 22158.
- [10] H. Shinzawa, K. Awa, W. Kanematsu, Y. Ozaki, *J. Raman Spectrosc.* **2009**, *40*, 1720.
- [11] J. Guicheteau, L. Argue, D. Emge, A. Hyre, M. Jacobson, S. Christesen, *Appl. Spectrosc.* **2008**, *62*, 267.
- [12] C. L. Zavaleta, B. R. Smith, I. Walton, W. Doering, G. Davis, B. Shojaei, M. J. Natan, S. S. Gambhir, *Proc. Natl. Acad. Sci. U. S. A.* **2009**, *106*, 13511.
- [13] J. R. Mourant, K. W. Short, S. Carpenter, N. Kunapareddy, L. Coburn, T. M. Powers, J. P. Freyer, *J. Biomed. Opt.* **2005**, *10*, 031106.
- [14] A. G. Shen, J. Peng, Q. H. Zhao, L. Su, X. H. Wang, J. M. Hu, J. Yang, *Laser Physics Lett.* **2012**, *9*, 322.
- [15] A. G. Ryder, *J. Forensic Sci.* **2002**, *47*, 275.
- [16] R. Y. Sato-Berru, E. V. Mejia-Uriarte, C. Frausto-Reyes, M. Villagran-Muniz, H. Murrieta, J. M. Saniger, *Spectrochim. Acta A Mol. Biomol. Spectrosc.* **2007**, *66*, 557.
- [17] T. Jombart, S. Devillard, F. Balloux, *BMC Genet.* **2010**, *11*, 94.
- [18] T. N. K. Zu, A. I. M. Athamneh, R. S. Wallace, E. Collakova, R. S. Senger, *J. Bacteriol.* **2014**, *196*, 3983.
- [19] T. N. K. Zu, A. I. M. Athamneh, R. S. Senger, *Fermentation* **2015**, *2*, 3.
- [20] M. Singh, Y. Tong, K. Webster, E. Cesewski, A. P. Haring, S. Laheri, B. Carswell, T. J. O'Brien, C. H. Aardema, R. S. Senger, J. L. Robertson, B. N. Johnson, *Lab Chip* **2017**, *17*, 2561.

- [21] B. G. Freedman, T. N. Zu, R. S. Wallace, R. S. Senger, *Biotechnol. J.* **2016**, *11*, 877.
- [22] A. I. Athamneh, R. A. Alajlouni, R. S. Wallace, M. N. Seleem, R. S. Senger, *Antimicrob. Agents Chemother.* **2014**, *58*, 1302.
- [23] R. S. Senger, C. D. DeLaTorre, W. F. Carswell, K. Webster, M. Sullivan, J. Gong, P. Du, J. L. Robertson, *J. Raman Spectrosc.* **2018**. Submitted
- [24] J. Liu, J. Sun, X. Huang, G. Li, B. Liu, *Appl. Spectrosc.* **2015**, *69*, 834.
- [25] Z. Movasaghi, S. Rehman, I. U. Rehman, *Appl. Spectrosc. Rev.* **2007**, *42*, 493.
- [26] M. M. Bradford, *Anal. Biochem.* **1976**, *72*, 248.
- [27] A. Kovacs, V. Izvekoy, G. Keresztury, G. Pongor, *Chem. Phys.* **1998**, *238*, 231.
- [28] B. A. Bolton, J. R. Scherer, *J. Phys. Chem.* **1989**, *93*, 7635.
- [29] V. J. C. Lin, J. L. Koenig, *Biopolymers* **1976**, *15*, 203.
- [30] S. Bouatra, F. Aziat, R. Mandal, A. C. Guo, M. R. Wilson, C. Knox, T. C. Bjorndahl, R. Krishnamurthy, F. Saleem, P. Liu, Z. T. Dame, J. Poelzer, J. Huynh, F. S. Yallou, N. Psychogios, E. Dong, R. Bogumil, C. Roehring, D. S. Wishart, *PLoS One* **2013**, *8*.

SUPPORTING INFORMATION

Additional Supporting Information may be found online in the supporting information tab for this article.

How to cite this article: Fisher AK, Carswell WF, Athamneh AIM, et al. The Rametrix™ LITE Toolbox v1.0 for MATLAB®. *J Raman Spectrosc.* 2018;49:885–896. <https://doi.org/10.1002/jrs.5348>