Quantification of Sugar Mixtures with Near-Infrared Raman Spectroscopy and Multivariate Data Analysis

A Quantitative Analysis Laboratory Experiment

Liqun Wang
School of Chemistry and Biochemistry, Georgia Institute of Technology, Atlanta, GA 30332

Boris Mizaikoff and Christine Kranz*
School of Chemistry and Biochemistry, Georgia Institute of Technology, Atlanta, GA 30332
University of Ulm, Institute of Analytical and Bioanalytical Chemistry, Ulm, Germany; **christine.kranz@uni-ulm.de*

Chemometrics is a teaching focus gaining increasing importance in the undergraduate laboratory curriculum *(1–7)*. Multivariate statistical data evaluation techniques are particularly useful if simple linear regression methods are insufficient in analyzing complex and strongly interrelated data sets. For instance, multivariate data analysis can be applied if multiple constituents are contained in one sample resulting in convoluted analytical signatures (e.g., overlapping peaks in IR or Raman spectroscopy). Modern instrumental analysis enables the acquisition of enormous amounts of data, which requires smart data reduction or recognition of the data sets providing the most useful analytical information. Consequently, multivariate analysis methods are nowadays broadly implemented in academic and industrial environments for solving complex qualification or quantification problems.

Raman spectroscopy is considered among the most important nondestructive routine optical analysis tools providing molecule specific vibrational information. A main advantage over conventional IR spectroscopy is its utility for in situ measurements in aqueous systems, as water is considered largely transparent to Raman techniques. While both Raman and IR spectroscopy probe molecular vibrations, the mechanism of obtaining spectra is fundamentally different (i.e., direct absorption in IR versus scattering in Raman) along with different selection rules (i.e., IR-active versus Raman-active vibrations), thus frequently providing complementary spectroscopic information. At present, most undergraduate teaching laboratories are equipped with Fourier transform infrared (FT-IR) spectrometers, whereas Raman spectroscopy is still not considered among the standard spectroscopic teaching tools in the undergraduate chemistry curriculum. However, comprehensive undergraduate education in modern instrumental analysis should include the effective use of the inherent advantages of this powerful technique.

Similar to FT-IR spectra, Raman spectra of mixtures are often complex in nature with increasingly overlapping vibrational signatures as the compositional complexity of the sample increases. Multivariate data evaluation methods are capable of resolving qualification or quantification problems in complex sample matrices. Hence, it is essential if not inevitable to combine these spectroscopic techniques with appropriate data analysis techniques, as recently reported in this *Journal (2)* and demonstrated for evaluating Raman spectra of sugar mixtures *(8–11)*.

An experiment was developed and implemented for fourth-year undergraduate students majoring in chemistry. Students gain hands-on experience in Raman spectroscopy, and in a lecture accompanying this laboratory exercise students are provided with the theory of Raman spectroscopy. Furthermore, they are educated in the basic concepts, theory, and application of multivariate data analysis methods such as, for example, principal components analysis and regression (PCA/PCR) *(12)*. PCA/PCR calculations are performed in a two-step process: in the first step (PCA), principal components (eigenvectors) are extracted from the calibration spectra. The obtained scores are unique to each calibration spectrum and enable the reconstruction of each calibration spectrum from a common set of PCs and the corresponding set of scores associated with each spectrum. In the second step (PCR), the obtained scores are regressed back against the mass fractions that are known for each calibration sample. Consequently, for an unknown sample the mass fraction of the constituents is determined from the collected spectrum by first extracting the unique scores and subsequent regression against the matrix of known mass fractions. In addition, the students had access to interactive teaching software (Teach Me/Data analysis, Springer–Verlag) that comprehensively illustrates the underlying concepts of eigenvector-based multivariate techniques such as PCA/PCR.

The experiment is based on analyzing an unknown aqueous mixture of sugars containing fructose, galactose, and glucose. These analytes are nontoxic, environmentally safe, and low in cost. Adequate preparation of calibration standards for training the multivariate regression model is essential prior to performing this experiment *(13)* and has to satisfy the requirements of multivariate calibration techniques along with the demands for the experimental Raman spectroscopic measurement. The experiment is designed as a five-hour laboratory module. However, it can be adjusted to fit schedules with shorter (2–4 h) experimental modules.

Experiment

Preparation of Standard Solutions

D-(−)-Fructose, D-(+)-galactose, and D-(+)-glucose are ACS reagent grade and were purchased from Sigma–Aldrich (St. Louis, MO). Fourteen solutions with different sugar compositions were prepared by mixing three different sugars in a deionized water matrix (Table 1). Mass fractions of fructose and galactose range from 0 to 20% (m/m); the mass fraction of glucose ranges from 0 to 30% (m/m) in the standard samples. A major effort in preparing adequate calibration standards for PCR entails establishing a sufficiently large set of samples, while minimizing collinearity of the mass fraction variations within the calibration set. Hence, a mass fraction matrix (Table 1) was constructed ensuring minimum correlation among the mass

fraction variations of different sugars within the calibration set utilizing a *minimum correlation table* (Table 2), which was derived from a table of randomized trials *(14, 15)*. A minimum correlation table is a generic approach to developing multivariate calibration schemes ensuring that confounding variations among components are minimized, thereby maximizing orthogonality for eigenvector-based data evaluation schemes *(15)*. Practically, the mass fraction range of each component in the experiment was uniformly divided into 14 levels, as shown in Table 2. The minimum correlation table provides the level that should be used for each constituent during preparation of each standard. To verify the utility of the calibration set, standards 4 and 11 were set aside as quasi-unknowns for testing the predictive capability of the established PCA/PCR models derived from the remaining standard mixtures. PCA/PCR requires a minimum number of calibration samples, which is determined by the classical design rule of at least 2^3 standards for three analytes at two mass fraction levels *(16)*, while minimizing the correlation among the standards. During development of this experiment, it was experimentally determined that 12 standard mixtures were sufficient for establishing sufficiently robust PCR calibration models if there are no other interfering factors such as, for example, impurities of the samples or—in this specific case—isomerization of sugars. The calibration solutions were prepared in sample vials and were then heated to 35 °C using a sand bath for dissolving all sugars. Prior to Raman analysis, the samples were equilibrated to room temperature.

Raman Analysis of Sugar Solutions

Students recorded FT-Raman spectra utilizing a FT-IR/Raman spectrometer (FRA106/S, Bruker Optics, Billerica, MA) equipped with a Nd:YAG excitation laser source emitting at 1064 nm and an indium–gallium–arsenide (InGaAs) detector. In principle, the experiments can be conducted with any Raman system providing the capability to record spectra of liquid-phase samples with sufficient signal-to-noise ratio and spectral resolution. However, the obtained predictive errors may vary from the data reported herein. The measurements were performed in a 180° backscattering mode. The excitation laser power was set at 400 mW, and the excitation laser was focused onto a quartz cuvette (volume: 1.5 mL) with an optical path length of 5 mm. The cuvette was equipped with a reflecting mirror layer at the rear for enhanced efficiency. Spectra were collected at a spectral resolution of 2 cm^{-1} in the frequency range of 1600–200 cm^{-1}; each spectrum represented an average of 32 scans.

Table 1. Mass Fraction of Standard Solutions

Standard	Fructose (%)	Galactose (%)	Glucose (%)	Distilled Water (%)
1	0	2.86	17.14	80.00
2	1.43	17.14	30.00	51.43
3	2.86	14.29	10.71	72.14
4	4.29	5.71	19.29	70.71
5	5.71	20.00	0	84.29
6	7.14	4.29	25.71	62.86
7	8.57	15.71	8.57	67.14
8	11.43	12.86	6.43	69.29
9	12.86	1.43	23.57	62.14
10	14.29	8.57	15.00	62.14
11	15.71	11.43	27.86	45.00
12	17.14	7.14	4.29	71.43
13	18.57	18.57	12.86	50.00
14	20.00	0	2.14	77.86

Table 2. Minimum Correlation Table for the 14 Standards

Standard	Fructose	Galactose	Glucose
1	1	3	9
2	2	12	14
3	3	10	6
4	4	5	10
5	5	14	1
6	6	4	12
7	7	11	5
8	8	9	4
9	9	2	11
10	10	7	8
11	11	8	13
12	12	6	3
13	13	13	7
14	14	1	2

Multivariate Data Analysis Using Raman Spectra

PCA/PCR calibration models were established using the PLS_Toolbox_3.5 software (Eigenvector Research Inc., Wenatchee, WA). The students recorded spectra from each of the twelve standards excluding standards 4 and 11, which were set aside as quasi-unknown samples for testing the performance of the established predictive model. Spectra of the 12 standard solutions were mean-centered prior to extracting the principal components *(17)*. During this procedure, the mean spectrum (i.e., average spectrum) was calculated from all calibration spectra and then subtracted from each calibration spectrum. Thereby, variations common to all calibration samples were removed. Consequently, the remaining variations among the spectra appeared enhanced prior to extracting the remaining directions of largest variance within the data set represented by the principal components.

Hazards

There are no hazards involved in this experiment. All compounds used are nontoxic. Furthermore, the combined FT-IR/Raman instrument is protected by an interlock from accidentally switching on the laser while the sample compartment is still open and vice versa.

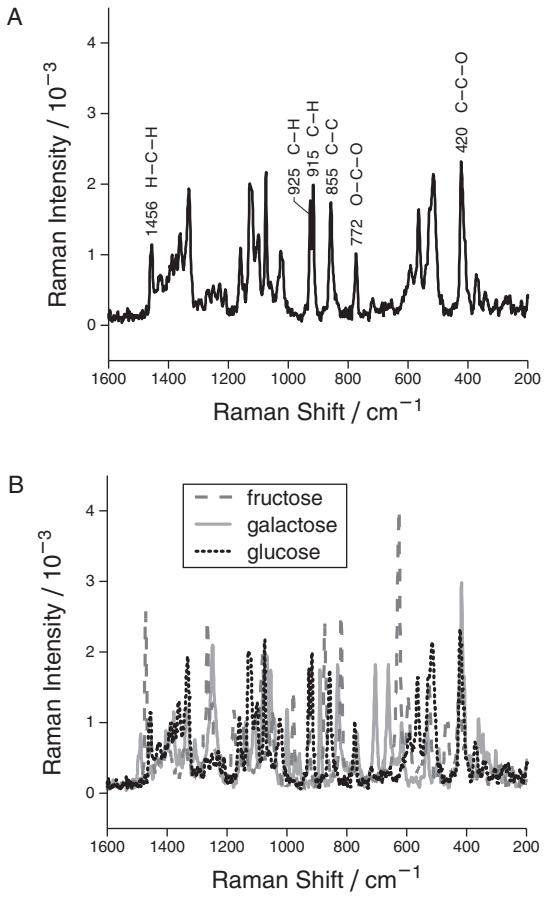

Figure 1. (A) Raman spectrum of solid glucose with selected peak assignment. (B) Overlay of Raman spectra of solid sugars: fructose, galactose, and glucose.

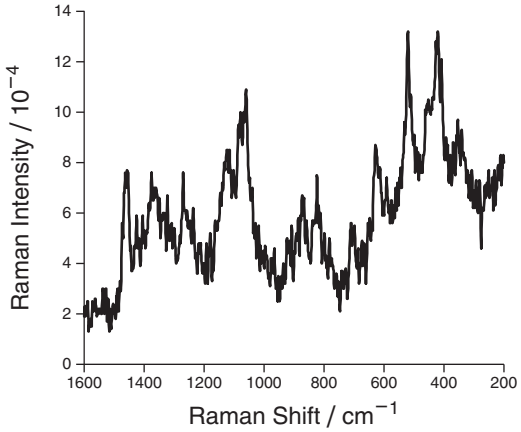

Figure 2. Raman spectrum of sugar mixture in aqueous solution (fructose:galactose:glucose:H_2O, 15.71:11.43:27.86:45.00, m/m/m/m).

Results

The Raman spectrum of pure solid glucose with peak labeling for selected major vibrations *(18, 19)* is shown in Figure 1A. A detailed peak assignment for all features is beyond the scope of this study, and the reader is referred to suitable references *(18–21)* for a more detailed peak assignment of the sugars studied in this experiment. An overlay of the Raman spectra for the pure sugars—glucose, galactose, and fructose—is shown in Figure 1B. A typical Raman spectrum for an aqueous solution of the sugar mixture (fructose:galactose:glucose:H_2O, 15.71:11.43:27.86:45.00, m/m/m/m) is shown in Figure 2. From these spectra it is immediately evident that significant overlap of the characteristic bands renders the quantification of each constituent impossible utilizing univariate statistics.

However, the developed multivariate PCA/PCR models exhibit the anticipated linear correlation between the predicted and the actual mass fraction values of each constituent contained in the standard mixtures. The developed model provides excellent results for fructose with a goodness of the fit $r^2 = 0.992$ and a predictive accuracy with a root mean square error of calibration (RMSEC) of ~0.57. The model for galactose has a r^2 value of 0.985 and a RMSEC of ~0.75; whereas the model for glucose is characterized by an r^2 value of 0.992 and a RMSEC of ~0.84. The superior fit for the fructose model derives from the fact that the spectrum of fructose (see Figure 1B) reveals much stronger Raman signatures in contrast to the spectra of galactose and glucose at the same measurement conditions. The obtained predicted mass fraction values along with the actual values for the quasi-unknown samples 4 and 11 are provided in Table 3. Given the weaker signals for glucose and galactose, relatively higher predictive errors are imminent. Furthermore, it is conceivable that the manual sample preparation and isomerization of the sugars add to the error.

In general, matrix interferences are a relevant factor possibly affecting the predictive accuracy. If interferences within unknown samples are also present within the calibration samples, their effects are considered while establishing the calibration model. However, if interferences within the unknown sample are not present or not fully represented within the calibration data set, additional error is introduced into the predictive model, which has to be carefully considered when designing such experiments.

Increasing the number of averaged Raman spectra for each sample and increasing the number of standards for establishing the multivariate model are viable strategies for reducing the predictive error. However, these experimental settings would exceed the practically useful duration for educational laboratory experiments.

Acknowledgments

This work was supported by the National Science Foundation within the CCLI (A & I) program (grant #02-095). The authors would like to thank Bruce Thompson for valuable discussions during designing the standard sample set for PCA/PCR analysis.

Table 3. Comparison of Student Data for the Predicted and Actual Mass Fractions of Two Quasi-Unknowns

Sample	Fructose (%)		Galactose (%)		Glucose (%)	
	Predicted	Actual	Predicted	Actual	Predicted	Actual
4	4.43	4.29	6.11	5.71	16.99	19.29
11	15.03	15.71	13.34	11.43	27.26	27.86
Average Relative Error	3.8		11.5		7.1	

Literature Cited

1. Chau, F. T.; Chung, W. H. *J. Chem. Educ.* **1995,** *72,* A84–A85.
2. Ribone, M. E.; Pagani, A. P.; Olivieri, A. C.; Goicoechea, H. C. *J. Chem. Educ.* **2000,** *77,* 1330–1333.
3. Oeberg, T. *J. Chem. Educ.* **2006,** *83,* 1178–1181.
4. Msimanga, H. Z.; Elkins, P.; Tata, S. K.; Smith, D. R. *J. Chem. Educ.* **2005,** *82,* 415–424.
5. Cazar, R. A. *J. Chem. Educ.* **2003,** *80,* 1026–1029.
6. Delaney, M. F.; Warren, F. V., Jr. *J. Chem. Educ.* **1981,** *58,* 646–651.
7. Cartwright, H. *J. Chem. Educ.* **1986,** *63,* 984–987.
8. Batsoulis, A. N.; Siatis, N. G.; Kimbaris, A. C.; Alissandrakis, E. K.; Pappas, C. S.; Tarantilis, P. A.; Harizanis, P. C.; Polissiou, M. G. *J. Agric. Food Chem.* **2005,** *53,* 207–210.
9. Mrozek, M. F.; Weaver, M. J. *Anal. Chem.* **2002,** *74,* 4069–4075.
10. Mrozek, M. F.; Zhang, D.; Ben-Amotz, D. *Carbohydr. Res.* **2004,** *339,* 141–145.
11. Arboleda, P. H.; Loppnow, G. R. *Anal. Chem.* **2000,** *72,* 2093–2098.
12. Beebe, K. R.; Pell, R. J.; Seasholtz, M. B. *Chemometrics: A Practical Guide,* Wiley: New York, 1998.
13. Miller, J. C.; Miller, J. N. *Statistics and Chemometrics for Analytical Chemistry;* Prentice Hall: London, 2000.
14. E. I. du Pond de Nemours and Company. *Strategy of Experimentation,* 2nd ed.; Appl. Technol. Div.: Wilmington, DE, 1988.
15. Thompson, B. T.; Mizaikoff, B. *Applied Spectrosc.* **2006,** *60,* 672–678.
16. Montgomery, D. C. *Design and Analysis of Experiments,* 4th ed.; John Wiley & Sons: New York, 1996; Chapter 7.
17. Kramer, R. *Chemometric Techniques for Quantitative Analysis;* Marcel Dekker: New York, 1998; p 173.
18. Söderholm, S.; Roos, Y. H.; Meinander, N.; Hotokka, M. *J. Raman Spectrosc.* **1999,** *30,* 1009–1018.
19. Mathlouthi, M.; Luu, D. V. *Carbohydr. Res.* **1980,** *81,* 203–212.
20. Sekkal, M.; Legrand, P.; Vergoten, G.; Dauchez, M. *Spectrochim. Acta, Part A* **1992,** *48A,* 959–973.
21. Wells, H. A., Jr.; Atalla, R. H. *J. Mol. Struct.* **1990,** *224,* 385–424.

Supporting *JCE Online* Material

http://www.jce.divched.org/Journal/Issues/2009/Nov/abs1322.html

Abstract and keywords

Full text (PDF) with links to cited *JCE* articles

Supplement

- Notes for instructors
- Step-by-step instructions for students