Amino Acid Quantification in Bulk Soybeans by Transmission Raman Spectroscopy

ARTICLE *in* ANALYTICAL CHEMISTRY · OCTOBER 2013

Impact Factor: 5.64 · DOI: 10.1021/ac402284b · Source: PubMed

9 AUTHORS, INCLUDING:

Matthew Gelber

University of Illinois, Urbana-Champaign

4 PUBLICATIONS **26** CITATIONS

SEE PROFILE

Hossain Azam

Manhattan College

6 PUBLICATIONS **35** CITATIONS

SEE PROFILE

analytical chemistry

Article

pubs.acs.org/ac

Amino Acid Quantification in Bulk Soybeans by Transmission Raman Spectroscopy

Matthew V. Schulmerich,[†,‡,§] Matthew K. Gelber,[†,‡] Hossain M. Azam,[‡,§] Sandra K. Harrison,[∥] John McKinney,[∥] Dennis Thompson,[∥] Bridget Owen,[⊥,@] Linda S. Kull,[⊥,@] and Rohit Bhargava*[,†,‡,§,#]

[†]Beckman Institute for Advanced Science and Technology

[‡]Department of Bioengineering

[§]Micro and Nanotechnology Laboratory

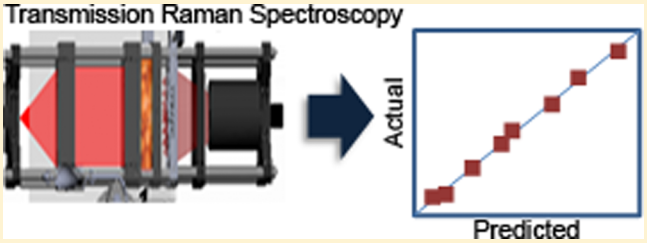
[∥]Illinois Crop Improvement Association

[⊥]National Soybean Research Laboratory

[@]College of Agriculture, Consumer, and Environmental Sciences

[#]Departments of Electrical and Computer Engineering, Mechanical Science and Engineering, Chemical and Biomolecular Engineering and Chemistry

ABSTRACT: Soybeans are a commodity crop of significant economic and nutritional interest. As an important source of protein, buyers of soybeans are interested in not only the total protein content but also in the specific amino acids that comprise the total protein content. Raman spectroscopy has the chemical specificity to measure the twenty common amino acids as pure substances. An unsolved challenge, however, is to quantify varying levels of amino acids mixed together and bound in soybeans at relatively low concentrations. Here we report the use of transmission Raman spectroscopy as a secondary analytical approach to nondestructively measure specific amino acids in intact soybeans. With the employment of a transmission-based Raman instrument, built specifically for nondestructive measurements from bulk soybeans, spectra were collected from twenty-four samples to develop a calibration model using a partial least-squares approach with a random-subset cross validation. The calibration model was validated on an independent set of twenty-five samples for oil, protein, and amino acid predictions. After Raman measurements, the samples were reduced to a fine powder and conventional wet chemistry methods were used for quantifying reference values of protein, oil, and 18 amino acids. We found that the greater the concentrations (% by weight component of interest), the better the calibration model and prediction capabilities. Of the 18 amino acids analyzed, 13 had R^2 values greater than 0.75 with a standard error of prediction c.a. 3−4% by weight. Serine, histidine, cystine, tryptophan, and methionine showed poor predictions (R^2 < 0.75), which were likely a result of the small sampling range and the low concentration of these components. It is clear from the correlation plots and root-mean-square error of prediction that Raman spectroscopy has sufficient chemical contrast to nondestructively quantify protein, oil, and specific amino acids in intact soybeans.



Soybeans are a commodity crop and a major food staple considered to be a "complete" source of protein capable of supplementing animal-based protein.[1,2] Soybeans are of significant economic and nutritional interest as the nutrition harvested makes it into the human food supply chain either directly by consuming soy-based products or indirectly when soy is used as a source for animal feed. As an important source of protein, buyers of soybeans for both human and animal consumption are interested not only in the total protein content but also in the specific amino acid that comprise the total protein content. Protein digestibility as measured by weight gain as a function of food intake is directly related to the ratio of amino acids present.[3] Essential amino acids have greater nutritional and economic value and are often supplemented in low quality animal feed to increase protein digestion.[4] As a result, the soybean industry is interested, especially in nondestructive, analytical techniques for quantifying specific amino acid content and amino acid ratios in bulk soybeans to gain the most value from the commodity.

One potential avenue is to use Raman spectroscopy, which has been shown to have sufficient chemical specificity to identify a pure sample of any given amino acid.[5,6] Of the twenty common amino acids, there are nine essential amino acids that the body is not able to synthesize and therefore must be received through diet or supplementation;[7−9] hence, technology for monitoring and recording the content of foodstuffs with these particular amino acids is especially crucial. While simultaneous measurements of different amino acid concen-

Received: July 24, 2013
Accepted: October 20, 2013
Published: October 21, 2013

   dx.doi.org/10.1021/ac402284b | *Anal. Chem.* 2013, 85, 11376−11381

trations are possible in a single spectrum with the chemical specificity of Raman spectroscopy, an enormous analytical challenge is to measure the concentration of the amino acids of interest in intact soybeans. Up to 18 amino acids present in bulk soybeans, with the attendant background from other molecules, makes this task especially challenging. Mature soybeans are comprised of ca. 36.5% total protein by weight.[10] This protein contains the amino acids detailed in Table 1 with

Table 1. Soybean Pallet of Amino Acid and Relative Concentrations

	nonessential	essential
glumatic acid	18.3%	
aspartic acid	11.9%	
leucine		7.7%
arginine	7.3%	[a]
lysine		6.3%
proline	5.5%	
serine	5.5%	
phenylalanine		4.9%
valine		4.7%
isoleucine		4.6%
alanine	4.5%	
glycine	4.4%	
threonine		4.1%
tyrosine	3.6%	[a]
histidine		2.6%
cystine	1.5%	[a]
tryptophan		1.4%
methionine		1.3%
% of total protein	62.5%	37.5%

[a]Sometimes considered essential in young animals to support growth.

essential amino acids consisting of ca. 37.5%. As an example, for 100 g of soybeans, the amino acid content ranges from methionine around 0.5 g to glutamic acid at around 6.7 g,[10] which are well within detection limits of modern spectroscopic instrumentation. While these ranges in amino acid concentrations are detectable with Raman spectroscopy, Raman spectroscopy is traditionally conducted in a backscattered configuration. This configuration collects light primarily from the surface of the sample and may not provide a true representative sampling of the amino acid content of the multiple beans.[11] Representative sampling of the soybeans is an important consideration and the traditional backscattered configuration may not be the best approach.[12] Recent work toward understanding and developing transmission Raman spectroscopy in light scattering materials offers an alternative approach to measurements.[13,14] Indeed, transmission configurations on light scattering samples like pharmaceutical tablets and soybeans are found to generate Raman signal from the sample interior, as opposed to the surface of the sample.[15−17] Raman spectroscopy offers an approach for optically homogenizing a soybean sample by taking advantage of the soybeans intrinsic light-scattering properties. Light scattering allows for the sampled volume to be larger than the spatial heterogeneity of the grain.[18]

We have recently demonstrated the capabilities of transmission Raman spectroscopy to nondestructively quantify the protein and oils content of individual soybeans.[18] Here we introduce an instrument designed to collect Raman spectrum from ca. 200 soybeans in a single acquisition (bulk measurements) using transmission measurements. Bulk measurements are necessary for quantifying low-abundance components, to better align with average measurements performed by destructive methods such as HPLC and to provide rapid assessments for average quality from samples in which average, not individual soybean composition, is important. Bulk measurements are also important to relate spectral measurements to gold-standard reference values provided by conventional destructive assays that are based on bulk sampling. Concordance of these measurements can not only help understand the analytical basis of our approach but also can form the basis for understanding the composition of single soybeans in single seed or imaging measurements. Here, the basic concept of optically homogenizing a soybean is simply extended to a bulk sample where the laser excitation and collection volumes are increased from a single bean to several beans. In this pilot study, we investigate the use of transmission Raman spectroscopy to noninvasively quantify protein, oil, and amino acid concentrations from bulk soybean samples.

MATERIALS AND METHODS

The source for measurements was 785 nm excitation light (Innovative Photonics Solutions, Monmouth Junction, NJ), launched from a 600 μm core optical fiber. The light was expanded and collimated to a 45 mm waist and then directed through a variable path length sampling chamber where the soybeans were held. A 12.5 mm path length was employed (sample chamber is shown in Figure 1). The light was collected

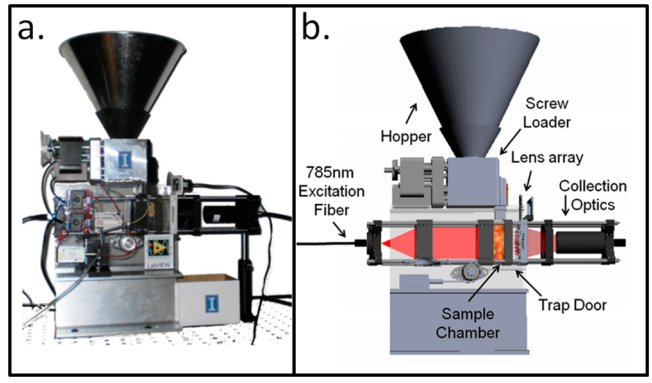


Figure 1. (a) Photograph and (b) schematic of the bulk soybean instrument.

from the opposite side of the chamber, providing all transmission geometry. The collection optics comprised a rotating lens-let array coupled with a telephoto lens pair utilized to focus the Raman-scattered light onto collection fibers (Fiber Tech Optica, Kitchener, Ontario). This optical design allows for both a large ~50 mm collection field of view along with an F-Number of 2 per lens. The lens-let array support structure was three-dimensionally printed to hold 12 lenses in 2 concentric circles. The inner circle was comprised of 3 lenses, and the outer circle was comprised of 9. The array was rotated by a servo-motor causing the 12 lenses to collect signal in 2 concentric circles during acquisitions. The signal collected by each point during the rotation averages over the duration of the acquisition time; therefore, precise sample positioning is not important because the regions of collection are actually moving during the measurement. The collection fibers consisted of a 7 × 7 array transposed into a 1 × 49 linear array, which entered

dx.doi.org/10.1021/ac402284b | Anal. Chem. 2013, 85, 11376−11381

the Raman spectrograph. The spectrograph (RXN-1, Kaiser Optical Systems Inc., Ann Arbor, MI) was optimized for 785 nm excitation and employed a back-illuminated, deep depletion CCD with antifringing (idus 420A BR-DD, Andor, Belfast). For all measurements, data in the spectral region 400−1800 cm^{-1} were acquired with 6−8 cm^{-1} spectral resolution. This instrumentation was controlled by a purpose-built application written in LabView (National Instuments, Austin, TX) to allow for automated sample loading and data collection.

Soybean Samples. Soybean samples anticipated to span a range of protein, oil, and amino acids were collected. The soybean batch selection was carefully performed to best represent the diversity of commercially available soybeans. Factors considered for representative population sampling included oil, protein and moisture concentrations, seed size, seed shape, year grown, seed coat sheen, hilum color, etc. Reference values for 18 amino acids as well as protein and oil content were measured and evaluated in this transmission Raman study. Protein concentrations of the 49 selected samples ranged from 38.41 to 50.71% dry basis and oil concentrations ranged from 16.36 to 22.41% dry basis. Ranges for amino acids are shown in Table 2. Before Raman measurements, each batch was manually inspected to ensure samples were debris free. Damaged or cracked soybeans were removed.

Table 2. Summary of Calibration Results

	sampling range (% by weight)	latent variables	R^2	RMSECV	RMSEP
total protein	38.41−50.71	3	0.86	1.04	1.19
total oil	16.36−22.41	4	0.85	0.66	0.76
glutamic acid	6.33−8.77	4	0.83	0.25	0.29
asparticacid	4.20−5.74	5	0.81	0.20	0.21
leucine	2.90−3.79	4	0.84	0.09	0.10
arginine	2.67−4.48	8	0.86	0.19	0.19
lysine	2.50−3.19	6	0.85	0.07	0.08
proline	1.84−2.50	4	0.78	0.07	0.09
serine	1.63−2.24	3	0.62	0.10	0.11
phenylalanine	1.87−2.50	6	0.85	0.06	0.07
valine	1.86−2.44	4	0.75	0.07	0.07
isoleucine	1.74−2.30	8	0.83	0.07	0.07
alanine	1.60−2.08	6	0.80	0.05	0.06
glycine	1.62−2.12	6	0.85	0.05	0.05
threonine	1.40−1.86	4	0.76	0.05	0.05
tyrosine	1.36−1.84	5	0.79	0.05	0.06
histidine	0.98−1.32	4	0.66	0.05	0.06
cystine	0.50−0.78	3	0.42	0.06	0.06
tryptophan	0.34−0.62	3	0.19	0.04	0.07
methionine	0.52−0.71	3	0.51	0.03	0.03

Raman Measurements. The soybean samples were loaded into the hopper and dispensed using a feed screw mechanism into the sample chamber. The bulk sampling chamber was approximately 50 mL in volume, which can contain approximately 150−250 soybeans, depending on their size and shape. Acquisition time was integrated over 3 min, and then the sampling chamber was emptied and refilled. This was repeated for a total of three runs for each variety of soybean. Excitation light of 785 nm illuminated the soybeans inside the bulk chamber with a total laser power of ∼3.5 W. The light was diffuse and no damage was observed upon inspection of the samples after measurements. Twenty-four soybean varieties were used to develop a calibration model for predicting oil and protein content. The calibration model was then validated on an independent set of twenty-five soybean varieties for oil, protein, and amino acid predictions. All Raman data was processed in Matlab 2012b (The Mathworks, Nantucket, MA) and a partial least-squares (PLS) approach with a random-subset cross validation (PLS Toolbox 3.0, eigenvector Research Inc., www.eigenvector.com) was used for the calibration model development.

Wet Chemistry/Reference Method. After nondestructive Raman measurements, the samples were reduced to a fine powder and "destructive chemistry" methods were used for quantifying reference values of protein, oil, and amino acids. Crude oil and protein reference values were provided by the Illinois Crop Improvement Association. Samples higher than approximately 5% moisture were dried at 130 °C for 1 h to remove moisture that could interfere with extraction. Bulk soybeans were ground using a Retsch ZM200 mill equipped with a 12-tooth rotor and a 1.0 mm conidur screen. The mill was operated at 10000 rpm. Ground samples were placed into 2.0 mL microcentrifuge tubes (Fisher Scientific Catalogue no. 02-681-258) until analysis could begin. Oil concentration was determined using the AOCS Official Procedure Am 5-04 using an Ankom XT15 extractor (Ankom Technology, Macedon, NY) with a replicate sample size of approximately 1 g. Petroleum ether was used as the extraction solvent with an extraction time of 60 min. Protein concentrations were determined in accordance with AOCS Official Method Ba 4e-93 using a Thermo Finnigan Flash EA1112 Nitrogen/Protein analyzer (CE Elantech, Lakewood, NJ) and with a replicate sample size of approximately 40 mg.

Concentrations of amino acids were determined at the University of Missouri Agricultural Experiment Station Chemical Laboratories (www.aescl.missouri.edu/MethRefs.html), Columbia, MO using standard methods (Amino acids: AOAC Official Method 982.30 E(a,b,c), chp. 45.3.05, 2006.); C-PER and DC-PER models were used together to measure the amino acids content using the "Amino acid Analyzer", which was capable of measuring individual amino acids at concentrations as low as 20 nM. Three types of hydrolysates were used for all types of amino acids. Acid hydrolysis was employed to determine all amino acids except methionine, cystine and/or cysteine, and tryptophan. Performic acid oxidation followed by acid hydrolysis was used to determine methionine (Met) and cystine/cysteine (Cys) and alkaline hydrolysis is used to determine tryptophan (Tryp). Standard amino acid solutions were used to calibrate the analyzer at least every 24 h. Each amino acid peak utilized in the final calculation had ≥85% resolution.

Signal Processing. All processing was performed in Matlab2012b (The Mathworks). The instrument's output was a frame from the charged coupled device (CCD) that consisted of 1024 × 255 pixels. The dispersive grating in the spectrograph projected the wavelength axis along the width of the chip (1024 pixels), while the chip height (255 pixels) contained spectra projected from each of the collection fibers. Cosmic rays were removed manually. A CCD frame acquired without a sample present and processed in the same manner was then subtracted from the sample frame to correct for the CCD's dark current. A "pin-cushion" and rotation correction was applied to the sample frame to correct for slit image curvature and a slight rotation of the CCD relative to the spectrograph's slit.[19] The wavelength-dependent response of the CCD was corrected by focusing the collection optics onto a NIST traceable white light source

emitted from a calibration accessory (HCA, Kaiser Optical Systems Inc., Ann Arbor, MI) and dividing the sample frame by the CCD's white light response. The neon channel of the same calibration accessory was then used to collect the atomic emission spectrum of neon and convert the wavelength axis of the CCD from pixels to wavenumbers. The Raman shift was then calculated with the Raman spectra collected from a standard Teflon sample. The neon lamp was used to calibrate the wavelength dispersion on the spectrograph, and the Teflon was used to calculate the laser wavelength. The soybean data was reduced to a single spectrum per variety by employing singular value decomposition (SVD) on each frame and then reconstructing the spectrum with only the first eigenvector. Baseline correction was then accomplished by fitting the background signal with a fifth-order polynomial and then subtracting that background from the spectra. Each spectrum was then normalized by area and then mean centered. This processing resulted in a single Raman spectrum per soybean variety, which was used as an input for PLS model development. Raman measurements from each soybean variety were collected in triplicate to yield a total of 147 spectra (49 × 3). The resulting Raman spectra, before mean centering, are shown in Figure 2a, and the average and one standard deviation of this data set is shown in Figure 2b. Figure 2 clearly delineates differences in the Raman spectra between soybean varieties.

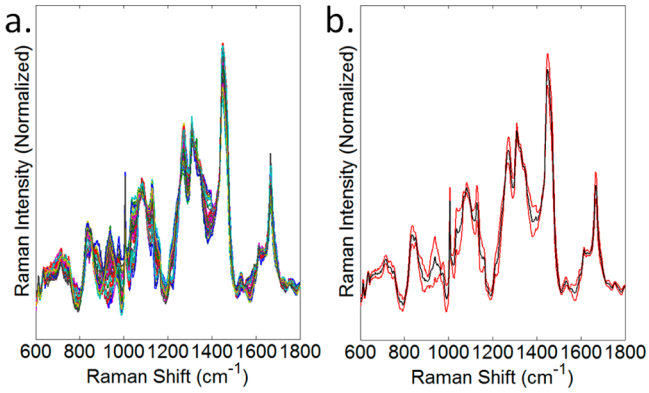


Figure 2. (a) Bulk Raman spectra of 49 different soybean varieties, (b) average (black) and 1 standard deviation (red).

RESULTS AND DISCUSSION

The Raman calibration results for bulk protein, oil, and amino acids are summarized in Table 2. The number of latent variables used to obtain these results was chosen to minimize the difference between the root-mean-square error in the calibration-validation set (n = 24) and the validation set (n = 25). The R^2 values reported in Table 2 were calculated based on the combined predictions of the calibration-validation set and the validation set. As expected, it is clear from these results that the greater the concentrations (% by weight) of the component of interest, the better the calibration model and prediction capabilities. This is apparent in the prediction models coefficient of determination (R^2), which describes the linearity of the transfer function relative to the validation data. In general, the closer the R^2 value is to 1.0, the better the model. However, a model with an R^2 value greater than 0.75 is generally recognized to have predictive capabilities. Of the 18 amino acids, 13 had R^2 values greater than 0.75. Serine, histidine, cystine, tryptophan, and methionine showed poor predictions (R^2 < 0.75), which are likely a result of the small sampling range and the low concentration of these components by weight.

Both calibration and validation correlation plots for protein and oil are depicted in Figure 3. The Raman predictions based

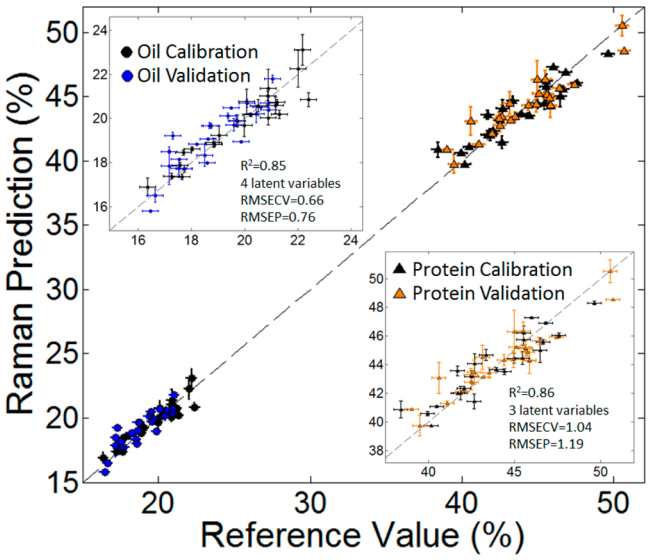


Figure 3. Calibration-validation and validation correlation plot for protein and oil.

on the calibration set are depicted in black, and the validation sets are presented in color. The error of predictions for protein is 1.19% and 0.76% for oil. This is slightly higher than expected as the reference method has a standard deviation of 0.4% for protein and 0.3% for oil. The larger-than-expected prediction error is likely the result of a relatively small calibration set.[20] As with all secondary analytical approaches, the larger the number and range of the calibration set, the greater the prediction capabilities.

Both calibration and validation correlation plots for the nonessential amino acids are shown in Figure 4. The Raman

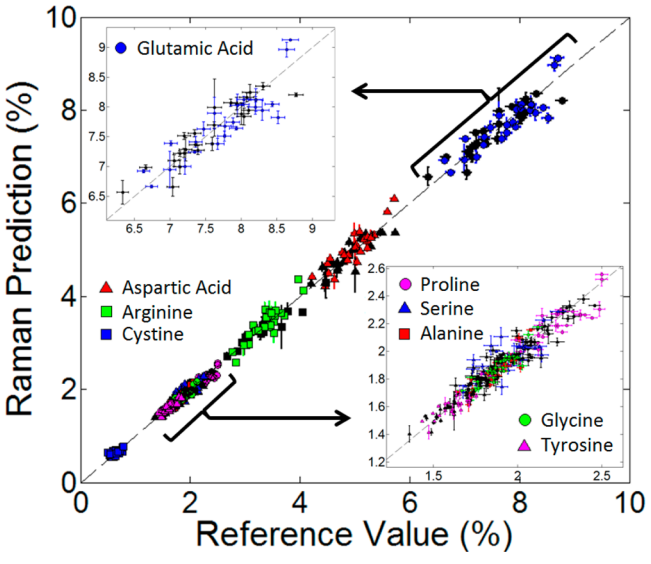


Figure 4. Calibration-validation and validation correlation plot for nonessential amino acids.

dx.doi.org/10.1021/ac402284b | Anal. Chem. 2013, 85, 11376−11381

predictions based on the calibration set are depicted in black, and the validation sets are presented in color. Of the nonessential amino acids, glutamic acid has the highest concentration of ∼6−9% of the soybean's weight. The Raman prediction has the capability of predicting the glutamic acid concentration within a root-mean-square error of prediction of 0.3%. This is a standard error [100 (RMSEP/mean % by weight)] of 3−4% by weight. Tyrosine has a much lower range of 1.36−1.84% of the soybeans weight. The Raman base calibration has a root-mean-square error of prediction of 0.06%, which is also a standard error of 3−4%. The Raman calibration approach was not able to predict cystine concentrations at this level, likely because of the low concentrations and small range in sampling. The range of cystine concentration was between 0.50 and 0.78% of the soybeans weight; however, the R^2 value is 0.42, indicating a poor calibration model. Hence, we are confident that the approach can work well for predicting concentrations of more abundant species but likely needs a careful understanding of calibration, sources of variance, instrument design, and optimization before low abundance or trace species can be confidently quantified using this approach. This study focused on an initial evaluation and did not seek to further optimize all parameters given the limited range in concentrations in our calibration set.

Both calibration and validation correlation plots for the essential amino acids are shown in Figure 5, where predictions

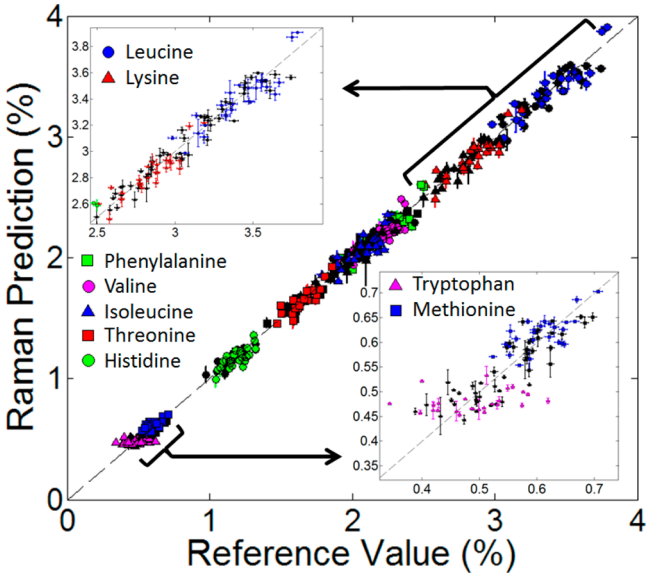


Figure 5. Calibration-validation and validation correlation plot for essential amino acids.

based on the calibration set are depicted in black and the validation sets are presented in color. Similar to the nonessential amino acids, the components with a higher weight percent have better Raman-based predictions. Leucine has a range of 2.90−3.79% by weight. The Raman based calibration has a root-mean-square error of prediction of 0.10%, which is a standard error of 3−4%. Similarly, lysine with a range between 2.50 and 3.19% by weight has a root-mean-square error of prediction of 0.08%. This is a standard error of 2−3%. The Raman-based calibration begins to perform poorly at concentrations below 1.5% by weight, as can be seen with correlation plots for histidine, tryptophan, and methionine.

CONCLUSIONS

In this pilot study, we ran small-scale calibration models for 18 amino acids using 24 soybean varieties as a calibration set and 25 soybean varieties as a validation set. It is clear from the correlation plots that Raman spectroscopy has sufficient chemical contrast to nondestructively quantify protein, oil, and specific amino acids. The calibration models with R^2 values greater than 0.75 showed standard errors of predictions c.a. 3−4% by weight. However, the accuracy of the Raman predictions is different depending on the specific component being calibrated. It is plausible that using a significantly larger training set would allow more accurate quantitation of the lower concentration amino acid components. While these results are encouraging, several interesting directions are also apparent. Optimization of the developed instrument for predicting low-abundance species and determination of ultimate analytical limits remains an endeavor, studies to directly compare NIR spectroscopic predictions to Raman spectroscopic predictions would also likely help practitioners understand the relative merits of each technology and provide optimal solutions.

AUTHOR INFORMATION

Corresponding Author

*E-mail: rxb@uiuc.edu.

Notes

The authors declare no competing financial interest.

ACKNOWLEDGMENTS

We gratefully acknowledge support for this work through the United Soybean Board Grant code 9282.

REFERENCES

(1) Michelfelder, A. J. *Am. Fam. Physician* **2009**, *79*, 43−47.
(2) Velasquez, M. T.; Bhathena, S. J. *Int. J. Med. Sci.* **2007**, *4*, 72−82.
(3) Fuller, M. F.; Garlick, P. J. *Annu. Rev. Nutr.* **1994**, *14*, 217−241.
(4) Friedman, M. *J. Agric. Food Chem.* **1996**, *44*, 6−29.
(5) Jenkins, A. L.; Larsen, R. A.; Williams, T. B. *Spectrochim. Acta, Part A* **2005**, *61*, 1585−1594.
(6) De Gelder, J.; De Gussem, K.; Vandenabeele, P.; Moens, L. *J. Raman Spectrosc.* **2007**, *38*, 1133−1147.
(7) WHO/FAO/UNU, *E.C. Protein and Amino Acid Requirements in Human Nutrition*; Technical Report Series: 935; World Health Organization: Geneva, 2007.
(8) Pencharz, P. B.; Ball, R. O. *Annu. Rev. Nutr.* **2003**, *23*, 101−116.
(9) Lyman, C. M.; Kuiken, K. A.; Hale, F. *J. Agric. Food Chem.* **1956**, *4*, 1008−1013.
(10) Nutrient Data Laboratory, A. *USDA Choline Study*; United States Department of Agriculture: Washington, D.C., 2003.
(11) Matousek, P. *Chem. Soc. Rev.* **2007**, *36*, 1292−1304.
(12) Qiao, Y.; Van Kempen, T. A. T. G. *J. Anim. Sci.* **2004**, *82*, 2596−2600.
(13) Shin, K.; Chung, H.; Kwak, C. *Analyst* **2012**, *137*, 3690−3696.
(14) Hwang, J.; Kang, S.; Lee, K.; Chung, H. *Talanta* **2012**, *101*, 488−494.
(15) Everall, N.; Priestnall, I.; Dallin, P.; Andrews, J.; Lewis, I.; Davis, K.; Owen, H.; George, M. W. *Appl. Spectrosc.* **2010**, *64*, 476−484.
(16) Everall, N.; Matousek, P.; Macleod, N.; Ronayne, K. L.; Clark, I. P. *Appl. Spectrosc.* **2010**, *64*, 52−60.
(17) Matousek, P.; Everall, N.; Littlejohn, D.; Nordon, A.; Bloomfield, M. *Appl. Spectrosc.* **2011**, *65*, 724−733.
(18) Schulmerich, M. V.; Walsh, M. J.; Gelber, M. K.; Kong, R.; Kole, M. R.; Harrison, S. K.; McKinney, J.; Thompson, D.; Kull, L. S.; Bhargava, R. *J. Agric. Food Chem.* **2012**, *60*, 8097−8102.
(19) Esmonde-White, F. W. L.; Esmonde-White, K. A.; Morris, M. D. *Appl. Spectrosc.* **2011**, *65*, 85−98.

(20) Kwak, J. T.; Reddy, R.; Sinha, S.; Bhargava, R. *Anal. Chem.* **2012**, *84*, 1063−1069.

dx.doi.org/10.1021/ac402284b | *Anal. Chem.* 2013, 85, 11376−11381