Brachiaria plantaginea as a Potential (New) Source of Shikimic Acid. Quantification by NIR and PLS Regression

Maria Cristina A. Costa1, Marcus B. Matallo2, Márcia M. C. Ferreira1, Sonia C. N. Queiroz1, Sydney Dionisio B. de Almeida2, Daniel A. de Siqueira Franco2

1 Theoretical and Applied Chemometrics Laboratory (LQTA), Institute of Chemistry, University of Campinas – Unicamp, Campinas – SP, Brazil
2 Instituto Biológico, jardim das Palmeiras, Campinas – SP, Brazil

Abstract

Shikimic acid is a natural organic compound produced in the biochemical pathways of eukaryotic organisms and is generally utilized as a starting material of the antiviral drug Oseltamivir (Tamiflu®). This study shows the results of shikimic acid accumulation in Brachiaria plantaginea, an abundant, grassy plant found in Brazil and other countries in Africa and America, after glyphosate spraying at three different doses. B. plantaginea plants harvested after 6 days of exposure to the herbicide showed that shikimic acid accumulation increased by 345%, on average, compared to unsprayed plants. The combination of near infrared spectroscopy and multivariate analysis using partial least square regression was applied for the quantification of shikimic acid in B. plantaginea. Spectra of 44 samples were obtained using the diffuse reflectance mode in the range of 4000 to 10 000 cm⁻¹ with 4 cm⁻¹ resolution. Different mathematical pretreatments were applied to the spectra. As preprocessing, the data were mean-centered. The calibration model with seven factors on twenty-nine samples (training set) exhibited a coefficient of determination = 0.9930 and a standard error of calibration = 84.05. For external validation (11 samples on the test set), the coefficient of determination = 0.9930 and the standard error of prediction = 154.91. The percent calibration error range was 1 to 10% for most of the samples and only two samples presented an error greater than 20%. For external validation, the mean prediction error was 10% and the range error ratio was 9.42, indicating that the model is qualified for screening calibration.

Key words
Brachiaria plantaginea · Poaceae · shikimic acid · NIR spectroscopy · PLS regression · chemometric analysis

Supporting information available online at http://www.thieme-connect.de/products

Recently, there is a growing concern about influenza A, especially subtypes H1N1 and H5N1 for their high transmission and mortality rates. Shikimic acid (3,4,5-trihydroxy-1-cyclohexene-1-carboxylic acid), Fig. 1, is a natural organic compound and an important intermediate in the biochemical pathways of plants and microorganisms. It is generally used as a starting material for industrial synthesis of the antiviral Oseltamivir (Tamiflu®, used against influenza A) and as a reactant in organic synthesis [1,2]. According to Chen et al. [3], shikimic acid and its derivatives possess several pharmacological effects, such as antithrombotic, anti-inflammatory, analgesic, antioxidant, anticancer, and antibacterial effects. However, shikimic acid is a scarce and expensive chemical, obtained mainly from the seeds of Illicium verum Hook.f. (Schisandraceae), a shrub native in China, and Illicium anisatum Gaertn., native to Japan [4]. So, there is an urgent need for the production of large amounts of shikimic acid from different sources [5]. Glyphosate is an herbicide widely used for controlling weeds around the world, whose mechanism of action is attributed to the activity inhibition of 5-enolpyruvyl-shikimic acid-3-phosphate (EPSP) synthase, an enzyme of the shikimic acid pathway, in a wide variety of plants with consequent shikimic acid accumulation. Glyphosate is the only commercially available herbicide that inhibits the enzyme EPSP, blocking the synthesis of aromatic amino acids [6]. The majority of the plant species are sensitive to glyphosate and when treated with this herbicide, accumulate high levels of shikimate [7,8]. Although glyphosate is currently the most important active ingredient for controlling weeds, low-dose applications offer the commercial use of glyphosate as a growth regulator [9]. A hormesis effect of glyphosate has been observed in various plants [10–14]. Studies on the increase of shikimic acid production in plants by treatment with low doses of glyphosate can be cited [15–18]. Recently, Franco et al. [19] investigated the possibility of using Brachiaria plantaginea (Link) Hitchc. (Poaceae) as another alternative source of shikimic acid. B. plantaginea is an annual grassy plant, common in Brazil and other countries in South America, Mexico, northeast and southeast USA, and Africa. In Brazil it is cultivated for forage, producing a rapid spring growth and allowing up to three cuts per cycle. The phenotypic characteristics and soil and climate requirements allow cultivation virtually throughout the national territory [20,21]. In this work, 44 samples of B. plantaginea were analyzed, and shikimic acid was quantified by high-performance liquid chromatography (HPLC) (Table 1). Although HPLC is versatile, quantitative, and precise, it can also be time-consuming and costly, requiring large quantities of expensive and toxic organic solvents [22,23]. Alternative methods based on near infrared spectroscopy (NIR) have been widely used in the determination and quantification of constituents of agricultural samples [22,24–27]. NIR spectroscopy is fast and nondestructive; however, the overtone and com-
The main objectives of this study were investigate the glyphosate and the maximum period of shikimate accumulation that would result in the greatest accumulation of shikimic acid, then, propose a fast and clean procedure for the quantification of shikimic acid in samples of *B. plantaginea* using NIR combined with PLS.

The experimental results presented in Table 1 show a higher accumulation of shikimic acid after 6 days of glyphosate application at the concentration of 36.0 g·ha$^{-1}$ of glyphosate ha$^{-1}$. The rates of glyphosate and the maximum period of shikimate accumulation are consistent with the literature [29–32]. The high standard deviation observed in Table 1 probably resulted from the intrinsic level of shikimic acid in plants associated with external factors. Chen et al. [33], studying the shikimic acid contents in conifer needles from different regions in China, observed great differences within the same or different species, even in the same region, the shikimic acid content of samples had distinct variances.

Shikimic acid is produced as an intermediate metabolite in the biochemical pathways of plants, and for the same dose of herbicide and same harvest time, variations in the amount of shikimic acid produced by the individuals studied were observed, suggesting differential sensitivity to glyphosate due to a higher demand for shikimate pathway products. A greater sensitivity to glyphosate, and consequently a greater shikimate accumulation, was detected by Pline et al. [34] in reproductive tissue over vegetative tissue. Furthermore, drifts of glyphosate applications are common and consequently the results are also subject to variations, which may lead to acid concentration values different from those expected. Shikimate accumulation in corn decreased from 349% at 0 m to 93% at 15.8 m, and shikimate levels were unaffected beyond 25.6 m downwind from a single aerial application of 866 g a.e·ha$^{-1}$ of glyphosate (Reddy et al.) [35].

The greatest accumulation of shikimic acid was observed in *B. plantaginea* sprayed with glyphosate at 36.0 g a.e·ha$^{-1}$. The peak concentration occurred within 6 days of exposure to the herbicide and was 345% higher compared to unsprayed plants. These results confirm the potential of this plant as a source of shikimic acid.

Before building the regression model, principal component analysis was performed in the mean-centered spectral data (X matrix) to check for trends or clusters among samples and the presence of outliers. The first two PCs explain about seventy percent of the total variance. Analyzing the two-dimensional PC1 vs. PC2 scores plot based on the mean-centered data (Fig. 2), one can observe that sample 46 is outside the 95% confidence ellipse, but its spectrum does not show atypical features. A significant overlap with a tendency towards discrimination of plants harvested at 6, 9, and 12 days after treatment with glyphosate along PC1 can be seen in this plot. The scores tend to become more positive in the first principal component with increasing time of exposure to glyphosate. When the samples with an exposition time of 3 days are included in the PC analysis, they occupy an intermediate region in the PC1 vs. PC2 scores plot, but tend to be discriminated mainly by PC2 (Fig. 1). Supporting Information).

The dendrogram (Fig. 3) shows three major groups with a similar value around 0.6. These clusters, G1, G2 and G3, are mainly composed of samples harvested at 12, 6, and 3 days, respectively, after treatment with glyphosate, but independent of the applied dose. In conclusion, exploratory data analysis showed some tendency to discriminate samples with respect to the harvest time, but not with the glyphosate dose applied. The PLS model was built using the shikimic acid concentration values determined by HPLC and the spectral data. The final model was obtained after removing four outliers (samples 29, 33, 34,
and 44) from the data set. After the removal of the outliers, the PLS model was built with 29 samples in the training set and the remaining 11 comprising the test set. The parameters for model evaluation are displayed in Table 2. Among all pretreatments applied to each mean spectrum, the smoothing and first derivative with windows of 7 and 13 points, respectively, based on a Savitzky-Golay polynomial filter [36], resulted in the best model. Leave-one-out cross-validation was applied during the PLS modeling in order to determine the number of factors in the model based on predictive ability. The optimum number of factors is indicated by the local minimum in the plot of standard error of cross-validation (SECV) vs. the number of factors (Fig. 2 S, Supporting Information). Table 2 shows the coefficients of determination for calibration and validation ($R^2_{\text{Cal}}$ and $R^2_{\text{Val}}$), SECV, and the standard error of calibration (SEC) for the final model with seven factors. The regression vector (Fig. 3 S, Supporting Information) shows the positive and negative coefficients in the model. We can observe the signals corresponding to the combination bands of lignin (C–H stretching) and the peaks in the region of the combination bands of O–H and C=O stretches of carboxylic acids [37,38].

![Dendrogram of the spectral data of 44 B. plantaginea samples. G1, G2, and G3 are the groups consisting mainly of samples harvested after 12, 6, and 3 days, respectively, of treatment with glyphosate herbicide. G1 and G3 showed a good classification of samples, while G2 is smaller and was formed by a mixture of samples harvested after 6 and 9 days. Samples 3, 21, and 33 in G1 were incorrectly classified, having been harvested with 3, 9, and 9 days, respectively. G3 contains eleven samples, whereas only two (32 and 41) were incorrectly classified. The remaining samples formed smaller groups with only two or three components, or are mixed in larger clusters.](image)

**Table 2** Parameters for model evaluation and validation* of the best PLS model obtained.

<table>
<thead>
<tr>
<th>Matrix size</th>
<th>Pretreatment</th>
<th>Factors</th>
<th>Outliers</th>
<th>$R^2$</th>
<th>SEC</th>
<th>SECV</th>
<th>SEP</th>
<th>RSD</th>
<th>RER</th>
</tr>
</thead>
<tbody>
<tr>
<td>44</td>
<td>Smoothing, first deriv.</td>
<td>7</td>
<td>4</td>
<td>0.99</td>
<td>0.84</td>
<td>0.93</td>
<td>84.05</td>
<td>349.53</td>
<td>154.91</td>
</tr>
</tbody>
</table>

* Original 44 samples, 4 outliers, 29 in training, and 11 in validation sets

and 44) from the data set. After the removal of the outliers, the PLS model was built with 29 samples in the training set and the remaining 11 comprising the test set. The parameters for model evaluation are displayed in Table 2. Among all pretreatments applied to each mean spectrum, the smoothing and first derivative with windows of 7 and 13 points, respectively, based on a Savitzky-Golay polynomial filter [36], resulted in the best model. Leave-one-out cross-validation was applied during the PLS modeling in order to determine the number of factors in the model based on predictive ability. The optimum number of factors is indicated by the local minimum in the plot of standard error of cross-validation (SECV) vs. the number of factors (Fig. 2 S, Supporting Information). Table 2 shows the coefficients of determination for calibration and validation ($R^2_{\text{Cal}}$ and $R^2_{\text{Val}}$), SECV, and the standard error of calibration (SEC) for the final model with seven factors. The regression vector (Fig. 3 S, Supporting Information) shows the positive and negative coefficients in the model. We can observe the signals corresponding to the combination bands of lignin (C–H stretching) and the peaks in the region of the combination bands of O–H and C=O stretches of carboxylic acids [37,38].

![Plot of reference vs. predicted values for (*) calibration and (•) external validation of shikimic acid in B. plantaginea sample models with seven factors.](image)

*Original 44 samples, 4 outliers, 29 in training, and 11 in validation sets*

The regression vector (Fig. 3 S, Supporting Information) shows the positive and negative coefficients in the model. We can observe the signals corresponding to the combination bands of lignin (C–H stretching) and the peaks in the region of the combination bands of O–H and C=O stretches of carboxylic acids [37,38].

**Fig. 4** Plot of reference vs. predicted values for (*) calibration and (•) external validation of shikimic acid in B. plantaginea sample models with seven factors.

The reason for the good fit in the training set is that the model was built with the best pretreatments of each mean spectrum, which was done before the removal of outliers. After the removal of the outliers, the PLS model was built with 29 samples in the training set and the remaining 11 comprising the test set. The parameters for model evaluation are displayed in Table 2. Among all pretreatments applied to each mean spectrum, the smoothing and first derivative with windows of 7 and 13 points, respectively, based on a Savitzky-Golay polynomial filter [36], resulted in the best model. Leave-one-out cross-validation was applied during the PLS modeling in order to determine the number of factors in the model based on predictive ability. The optimum number of factors is indicated by the local minimum in the plot of standard error of cross-validation (SECV) vs. the number of factors (Fig. 2 S, Supporting Information). Table 2 shows the coefficients of determination for calibration and validation ($R^2_{\text{Cal}}$ and $R^2_{\text{Val}}$), SECV, and the standard error of calibration (SEC) for the final model with seven factors. The regression vector (Fig. 3 S, Supporting Information) shows the positive and negative coefficients in the model. We can observe the signals corresponding to the combination bands of lignin (C–H stretching) and the peaks in the region of the combination bands of O–H and C=O stretches of carboxylic acids [37,38].

**Fig. 4** Plot of reference vs. predicted values for (*) calibration and (•) external validation of shikimic acid in B. plantaginea sample models with seven factors.

The reason for the good fit in the training set is that the model was built with the best pretreatments of each mean spectrum, which was done before the removal of outliers. After the removal of the outliers, the PLS model was built with 29 samples in the training set and the remaining 11 comprising the test set. The parameters for model evaluation are displayed in Table 2. Among all pretreatments applied to each mean spectrum, the smoothing and first derivative with windows of 7 and 13 points, respectively, based on a Savitzky-Golay polynomial filter [36], resulted in the best model. Leave-one-out cross-validation was applied during the PLS modeling in order to determine the number of factors in the model based on predictive ability. The optimum number of factors is indicated by the local minimum in the plot of standard error of cross-validation (SECV) vs. the number of factors (Fig. 2 S, Supporting Information). Table 2 shows the coefficients of determination for calibration and validation ($R^2_{\text{Cal}}$ and $R^2_{\text{Val}}$), SECV, and the standard error of calibration (SEC) for the final model with seven factors. The regression vector (Fig. 3 S, Supporting Information) shows the positive and negative coefficients in the model. We can observe the signals corresponding to the combination bands of lignin (C–H stretching) and the peaks in the region of the combination bands of O–H and C=O stretches of carboxylic acids [37,38].

**Fig. 4** Plot of reference vs. predicted values for (*) calibration and (•) external validation of shikimic acid in B. plantaginea sample models with seven factors.

The reason for the good fit in the training set is that the model was built with the best pretreatments of each mean spectrum, which was done before the removal of outliers. After the removal of the outliers, the PLS model was built with 29 samples in the training set and the remaining 11 comprising the test set. The parameters for model evaluation are displayed in Table 2. Among all pretreatments applied to each mean spectrum, the smoothing and first derivative with windows of 7 and 13 points, respectively, based on a Savitzky-Golay polynomial filter [36], resulted in the best model. Leave-one-out cross-validation was applied during the PLS modeling in order to determine the number of factors in the model based on predictive ability. The optimum number of factors is indicated by the local minimum in the plot of standard error of cross-validation (SECV) vs. the number of factors (Fig. 2 S, Supporting Information). Table 2 shows the coefficients of determination for calibration and validation ($R^2_{\text{Cal}}$ and $R^2_{\text{Val}}$), SECV, and the standard error of calibration (SEC) for the final model with seven factors. The regression vector (Fig. 3 S, Supporting Information) shows the positive and negative coefficients in the model. We can observe the signals corresponding to the combination bands of lignin (C–H stretching) and the peaks in the region of the combination bands of O–H and C=O stretches of carboxylic acids [37,38].

**Fig. 4** Plot of reference vs. predicted values for (*) calibration and (•) external validation of shikimic acid in B. plantaginea sample models with seven factors.
The difficulty in getting a better model can be explained in part by the wide range of shikimic acid concentrations (333.90 to 3592.45 µg × g$^{-1}$).

Despite the difficulty in establishing a PLS model for this data set, the results obtained for both calibration and prediction sets were satisfactory, demonstrating that NIR spectroscopy associated with PLS regression is a possible alternative to quantify shikimic acid in *B. plantaginea*. The model was considered validated, as shown by the results found for the external set of samples. The coefficient of determination was above 0.90 and the RER value was close to 10.0, indicating that the proposed model is acceptable for quality control.

**Material and Methods**

**Samples:** The trial was conducted in a greenhouse. Seeds of *B. plantaginea* were planted in vessels on loamy soil classified as purple Eutrustox under a completely randomized design with three replications. Prior to planting, fertilization was performed with 260 kg × ha$^{-1}$ of 00–20–20 (N–P–K) fertilizer, and the herbicide glyphosate (Roundup Original) was sprayed at reduced doses of 36.0, 3.6, and 0.36 g acid equivalent (a.e.) of glyphosate ha$^{-1}$, which are very low compared to the minimum dose suggested as herbicide by Rodrigues and Almeida (360 g a.e. × ha$^{-1}$) [40].

Treatments were applied at the beginning of the reproductive stage of *B. plantaginea*, 71 days after sowing, using a CO$_2$ backpack sprayer equipped with four nozzles spaced at 0.5 m and 110.03 tips (Magno) at a constant pressure of 35 lb × in$^{-2}$ adjusted for a spray volume of 100 L·ha$^{-1}$. At the time of spraying, wind force, according to the Beaufort scale, varied between 4.0 and 5.0 km × h$^{-1}$; air temperature was 24°C and relative air humidity was 65%.

The determination of shikimic acid in *B. plantaginea* leaves followed the procedure described by Matallo et. al. [41]. At 3, 6, 9, and 12 days after treatments, the leaves were collected, always in the morning, dried in a forced air circulation oven at 60°C for 48 h, and then crushed at 25 000 rpm (IKA A11 Basic), frozen at temperatures ≤10 °C, stored in polystyrene boxes, sealed, and transported to the laboratory. The extraction of shikimic acid was performed in replicates of 400 mg of dry matter, mixed with 10 mL of water, acidified to pH 2 with phosphoric acid (Merck), heated in a microwave oven (Panasonic NN-S62B) at 100 W for 10 seconds at a temperature of 49.8°C ± 2.8°C, allowed to cool for 5 min and then was filtered through Whatman N. 1 filter paper and a 0.22-µm membrane filter (Millex-GV, Millipore) for quantification. Shikimic acid was quantified by HPLC (Fig. 5A) using a Shimadzu LC 2010 A chromatographer (isocratic mode) with a diode array detector at a wavelength of 212 nm, injection volume of 20 µL, and a Phenomenex Gemini C18 110 Å (250.0 mm × 4.0 mm; 5 µm particle size) LC column. The mobile
phase used was ultrapure water acidified solution with phosphoric acid and methanol (95:5) at a flow rate of 1.0 mL x min\(^{-1}\). It was verified that the chromatograms were reproducible and so each replicate was analyzed only once by HPLC. The results were taken on a dry weight basis in order to eliminate the variability of the water content of the plants. Data were subjected to analysis of variance by the F-test (\(\text{Table 1}\)).

Near infrared spectroscopy data analysis: The diffuse reflectance spectra in the near infrared region were acquired using an Antares II FT-NIR spectrometer (Thermo Fisher Scientific) in the 4500 to 10000 cm\(^{-1}\) range. The spectra were generated by averaging 16 successive scans with 4 cm\(^{-1}\) nominal resolution (yielding 3112 wavenumbers). The signals were obtained in the reflectance mode (%R) and transformed into absorbance by using a log transformation. \(\text{Fig. 5B}\) shows a generic spectrum. Three spectra were recorded for each sample, and the average spectrum was used for data analysis.

The mean spectra were organized in a matrix format \(X (44, 3112)\) where each row corresponds to a sample and the columns correspond to the absorbance (log 1/R) values. Different mathematical transformations were tested; the best results were obtained by applying smoothing and the first derivative. The vector \(y\) of the concentrations was correlated with spectral information through the PLS regression method on the mean-centered data using Pirouette® 4.5 software. Leave-one-out cross-validation was used to determine the number of factors (latent variables) in the calibration model. The presence of outliers was investigated by analyzing the plot of leverage vs. studentized residuals and by principal component analysis. After removing the outliers, the data set was randomly split into two subsets: the training (calibration) set containing 29 samples and the test set with the remaining 11. The final regression model was evaluated by analyzing the values of \(R^2, \text{SEC, SECV, and the relative error (RE)}\) as shown in \(\text{Table 2}\). Supporting Information.

Supporting Information

The relative error values for the complete set of 40 samples, the equations for the calculation of the validation parameters, and PC1 vs. PC2 scores plot of the mean-centered NIR spectra of the samples are available as Supporting Information.

Acknowledgements

\(\text{MMCF acknowledges the Brazilian Governmental Agency CNPq for financial support. M.B.M. thanks CNPq for funding the project 431927/2013–9.}\)

Conflict of Interest

\(\text{The authors declare no conflict of interest.}\)

References


7. Burke IC, Thomas WE, Wendy A, Pline-Srni DA. The presence of outliers was investigated by analyzing the plot of leverage vs. studentized residuals and by principal component analysis. After removing the outliers, the data set was randomly split into two subsets: the training (calibration) set containing 29 samples and the test set with the remaining 11. The final regression model was evaluated by analyzing the values of \(R^2, \text{SEC, SECV, and the relative error (RE)}\) as shown in \(\text{Table 2}\). Supporting Information.

Supporting Information

The relative error values for the complete set of 40 samples, the equations for the calculation of the validation parameters, and PC1 vs. PC2 scores plot of the mean-centered NIR spectra of the samples are available as Supporting Information.

Acknowledgements

\(\text{MMCF acknowledges the Brazilian Governmental Agency CNPq for financial support. M.B.M. thanks CNPq for funding the project 431927/2013–9.}\)

Conflict of Interest

\(\text{The authors declare no conflict of interest.}\)

References


7. Burke IC, Thomas WE, Wendy A, Pline-Srni DA. The presence of outliers was investigated by analyzing the plot of leverage vs. studentized residuals and by principal component analysis. After removing the outliers, the data set was randomly split into two subsets: the training (calibration) set containing 29 samples and the test set with the remaining 11. The final regression model was evaluated by analyzing the values of \(R^2, \text{SEC, SECV, and the relative error (RE)}\) as shown in \(\text{Table 2}\). Supporting Information.

Supporting Information

The relative error values for the complete set of 40 samples, the equations for the calculation of the validation parameters, and PC1 vs. PC2 scores plot of the mean-centered NIR spectra of the samples are available as Supporting Information.

Acknowledgements

\(\text{MMCF acknowledges the Brazilian Governmental Agency CNPq for financial support. M.B.M. thanks CNPq for funding the project 431927/2013–9.}\)

Conflict of Interest

\(\text{The authors declare no conflict of interest.}\)

References

26 Parreira TF, Ferreira MMC, Sales HJS, Almeida WB. Quantitative determination of epoxidized soybean oil using near-infrared spectroscopy and multivariate calibration. Appl Spectrosc 2002; 56: 1607–1614