L-Ascorbic Acid Prediction in Aqueous Solution Based on FTIR-ATR Terahertz Spectroscopy*  

Diding SUHANDY*1, Meinilwita YULIA*2, Yuichi OGAWA*2, Naoshi KONDO*2

Abstract

The feasibility of using attenuated total reflectance-terahertz (ATR-THz) spectroscopy for quantification of L-ascorbic acid (L-AA) in aqueous solutions was investigated. The spectra of 55 samples of L-AA solution ranging in concentration from 0-21% were acquired using Terahertz-based Fourier transform infrared (THz-FTIR) spectrometer. Spectral absorbance in the range 20 to 400 cm⁻¹ was used in this analysis. Full spectrum partial least squares (FS-PLS) regression was used to develop and validate the calibration model for determination of L-AA. The result showed that using the pre-processing of Savitzky-Golay first derivative spectra, prediction resulted in the lowest root mean square error of prediction (RMSEP) = 2.791%. A ratio of standard deviation to prediction error (RPD) value of 4.48 was obtained.

[Keywords] Terahertz wave, FS-PLS regression, RPD value, inter-molecular vibration, intra-molecular vibration

1 Introduction

L-ascorbic acid (L-AA), also known as vitamin C, is a water-soluble vitamin. It can be found widely in many plant materials, such as fruit and vegetables. Recently, many food products, such as juices and sports drinks, are fortified with L-AA. This is because L-AA is an essential nutrient and antioxidant, and thus plays an important role in health. Hence, there is increasing demand for vitamin C in both fresh and processed foods and drinks. Of the L-AA that is found in fresh food, such as fruits and vegetables, is however subject to degradation such process as heating, etc (Lee & Kader, 2000). For this reason, it is highly desirable to be able to quantify L-AA in foods, especially in juices or sports drinks, for quality control purposes.

Many papers have established non-spectroscopic methods for L-AA determination. These conventional methods include titrimetry, chemiluminescence, fluorometric, chromatographic and electrochemical methods (Arya et al., 1998; Arya et al., 2000). Each method though has its limitations. For example, while the titrimetry method using dichlorophenolidophenol as the titrant is rapid, the titrant itself is unstable and must be standardized before use. In the case of the chromatographic method, while it is accurate, it is expensive and time consuming.

Of the spectroscopic methods, Blanco et al., (1993) reports using near infrared (NIR) spectroscopy coupled with step-wise multiple linear regression (SMLR) and partial least squares (PLS) regression to determine L-AA in pharmaceutical products at concentrations of 16.67%, 22.88% and 40%. In this analysis relative standard error (RSE) was used to assess the quality of the calibration and validation model, both for the SMLR and PLS regression. The RSE for the SML regression was 0.59-1.82% (calibration), and 0.72-2.34% (validation); and 0.69-2.46% (calibration), 0.89-2.85% (validation) for PLS regression. Spectroscopic methods using NIR, Mid Infrared (MIR) and Raman spectroscopy for determination of L-AA in powder and liquid samples has been reported by Yang & Irudayaraj (2002). In these measurements, coefficients of determination (R²) for L-AA were 0.966-0.999, 0.973-0.980, and 0.941 for MIR, NIR and Raman spectroscopy, respectively. In the same report, determination of L-AA concentrations in the mid infrared region was conducted using Fourier transform infrared attenuated total reflectance (FTIR-ATR) spectroscopy in aqueous solutions. It was also demonstrated that using FTIR-ATR in the mid infrared region provided the best prediction of L-AA concentration, with R² = 0.999 and the lowest standard error of cross-validation (SECV) = 0.23%. In the far infrared region, an investigation of vibrational spectra of L-AA at low frequencies (below 100 cm⁻¹) using terahertz time domain spectroscopy (THz-TDS) was reported by Binghua et al., (2008). However, determination of L-AA concentrations

* Partly presented at the 126th JSAM Kansai Branch Annual Meeting at Wakayama Japan in September 2011 and at ASABE Annual International Meeting at Louisville Kentucky USA in August 7-11, 2011.
*1 JSAM Member, Corresponding author, Graduate School of Agriculture, Kyoto University, Sakyo ku, Kyoto 606-8502, Japan ; diding.suhandy@fs2.ecs.kyoto-u.ac.jp.
*2 JSAM Member, Graduate School of Agriculture, Kyoto University, Sakyo ku, Kyoto 606-8502, Japan
using FTIR-ATR in the far infrared region including low and high frequency has not been reported yet. The use of this region for spectroscopic analysis, which is now known as the terahertz (THz) region, has to date been little explored due to the lack of appropriate generating and detecting devices.

Recent developments in both emitting and more readily detecting methods of THz radiation have enabled terahertz spectroscopy to be undertaken. THz spectroscopy utilizes electromagnetic waves ranging from 0.1 to 10 THz. One THz corresponds to 33.33 cm\(^{-1}\), 4.14 meV, and has a wavelength of 300 \(\mu\)m (Beard et al., 2002; Pickwell and Wallace, 2006; Baxter and Guglietta, 2011; Ferguson and Zhang, 2002). This region, especially in the low frequencies (0.3–6.0 THz, or about 10–200 cm\(^{-1}\)) is rich in information about intra-molecular and inter-molecular vibrational modes in biological molecules (Laman et al., 2008). While NIR and mid infrared spectroscopy utilize electromagnetic radiation which correspond to stretching or bending motions of individual bonds in the molecule, in the far infrared or THz region electromagnetic radiation corresponding to motions of the entire structure is utilized (Laman et al., 2008). THz radiation has low photon energies (1 THz = 4.14 meV), one million times weaker than x-rays, and will not cause harmful photo ionization in biological tissues. This has advantages both for imaging biological materials, and in food analysis. In addition to being considered safe for most applications, THz radiation can penetrate through many commonly used non-polar dielectric materials such as paper, cloth, cardboard, textiles, plastics, wood, leather and ceramics with moderate attenuation. This allows THz spectroscopy to be used in non-invasive and non-destructive inspection systems. The range of potential applications for THz spectroscopy is expanding with the increased availability of many absorption spectra discovered recently in the THz region (e.g., fingerprint spectra). Those spectra are peculiar to specific chemicals, including vitamins, sugars, pharmaceuticals, and agricultural chemicals (Kawase et al., 2003). One study successfully applied a THz imaging technique for non-invasive detection of illicit drugs inside an envelope (Kawase et al., 2003). Other studies successfully acquired low vibrational spectra (below 200 cm\(^{-1}\) of wavenumber) of riboflavin (vitamin B\(_2\)) and related compounds, and spectral absorbance of several fatty acids and their analogues (Takahashi et al., 2005; Jiang et al., 2011).

Water, a strongly polar liquid, is highly absorptive in the THz region (Ogawa et al., 2009; querry et al., 1991). Therefore, it has not been recommended for use with the transmittance method of THz spectral probing in aqueous solutions. To overcome this limitation, the attenuated total reflectance (ATR) method is widely used for aqueous samples in the THz region, instead of using the transmittance method. There has however been no significant report on the application of FTIR-ATR-THz spectroscopy coupled with chemometric method for determination of L-AA in aqueous solutions. Therefore, the objective of this study was to demonstrate the potential use of FTIR-ATR-THz spectroscopy for quantification of L-AA in aqueous solutions. Specifically, a calibration model for L-AA using full spectrum partial least squares (FS-PLS) regression is developed. This calibration model can be used to predict the concentration of L-AA in aqueous solutions.

II Materials and Methods

1. Spectrometer for THz spectral acquisition

THz spectra of L-AA solutions were acquired using a Fourier-transform infrared (FTIR) based spectrometer (FARIS-1S, JASCO Corp., Japan) (Figure 1). This system is based on a FTIR spectrometer (range of measurement: 20-450 cm\(^{-1}\) or 0.6-13.5 THz) with a special light source made from a high pressure mercury lamp as a THz generating device. This lamp has low spectral intensity in the low frequency range (Hangyo et al., 2002). To solve this problem, a beam splitter made from silicon, which has high transmittance in the THz region, was used instead of one made from mylar. The THz spectral acquisition system was also equipped with an attenuated total reflectance (ATR) chamber. In this ATR system, a silicon prism which has high refractive index in the THz region was used as an internal reflection element (IRE). The spectrometer was also equipped with a room-temperature pyroelectric sensor made from deuterated L-alanine triglycine sulfate (DLTGS) as a detector. The software program Spectral Manager for Windows (JASCO Spectral Manager, JASCO Corp., Tokyo, Japan) was used to control the spectral acquisition process.

![Fig. 1 FARIS-1S spectrometer used for THz spectral acquisition (JASCO Corp., Japan)](image-url)
2. L-Ascorbic Acid (L-AA) Solutions
L-ascorbic acid (L-AA) powder (Wako Pure Chemical Industries, Ltd., Japan) was used to prepare L-AA solutions, by dissolving the powder in distilled water. The solutions were stirred well with a mixer (Tube Mixer TRIO HM-, AS ONE, Japan). In this study, 55 samples of L-AA solution were prepared. The concentration range of the L-AA solutions was 0-21% (mass/mass (w/w)). For FTIR-ATR-THz measurement, a 300μL sample of the L-AA solution was pipetted quickly onto the silicon prism surface using a micro pipette.

The samples were divided into two sample sets. A calibration and cross-validation sample set, consisting of 35 samples. This set was used for developing the calibration model and performing the cross-validation test. And a prediction sample set, consisting of 20 samples, was used for prediction purposes. Table 1 shows the two sample sets in detail, and the range of L-AA concentrations in the calibration and cross-validation set, which covers the range in the prediction set.

Table 1 Characteristic of calibration and cross-validation and prediction sample sets

<table>
<thead>
<tr>
<th>Items</th>
<th>Calibration and cross-validation sample set</th>
<th>Prediction sample set</th>
</tr>
</thead>
<tbody>
<tr>
<td>Number of samples</td>
<td>35</td>
<td>20</td>
</tr>
<tr>
<td>Minimum value*</td>
<td>0</td>
<td>3.21</td>
</tr>
<tr>
<td>Maximum value*</td>
<td>21.89</td>
<td>21.09</td>
</tr>
<tr>
<td>Mean*</td>
<td>10.13</td>
<td>12.61</td>
</tr>
<tr>
<td>Standard deviation*</td>
<td>6.69</td>
<td>6.81</td>
</tr>
<tr>
<td>Units</td>
<td>%</td>
<td>%</td>
</tr>
</tbody>
</table>

3. ATR THz spectra acquisition
FTIR-ATR-THz spectra of the L-AA solutions were acquired in the range 20-450 cm⁻¹, using a 16 cm⁻¹ of resolution. Each spectrum contains on average 200 scanning spectra. The reference for air was measured every 5 samples. The spectra intensity of samples and reference were obtained in single beam (SB) unit. During THz spectral measurement, the laboratory temperature and relative humidity were maintained at around 25°C and 70% respectively. The absorbance value of the sample was calculated using equation 1.

\[ A(\nu) = -\log_{10} \frac{S(\nu)}{R(\nu)} \]  

Where:
- \( A(\nu) \) is absorbance at wavenumber \( \nu \)
- \( S(\nu) \) is intensity of sample at wavenumber \( \nu \)
- \( R(\nu) \) is intensity of air reference at wavenumber \( \nu \)

The calculated absorbance value was corrected using the ATR correction function provided in the software (JASCO Spectral Manager, JASCO Corp., Tokyo, Japan). The corrected values were used for further analysis.

4. Calibration model using full spectrum PLS (FS-PLS) regression
To extract the relationship between THz spectra and the concentration of L-AA, a number of calibration models were developed using FS-PLS regression with different pre-processed spectra, including smoothing (moving average, median filter, Gaussian filter and Savitzky-Golay smoothing), and derivation (Savitzky-Golay first and second derivative). The internal validation method, known as full cross-validation, was conducted for each calibration model. FS-PLS regression is a very powerful and useful technique for quantitative studies in analytical science, including spectroscopy. In this method, FS-PLS regression utilized both information from all wavenumbers of the THz spectra and the L-AA concentration data, and new variables called latent variables (PLS factors). In the development of all the calibration models, 20 PLS factors were set up as a maximum. Pre-processing of the spectra and the development of a calibration model using FS-PLS regression was performed by using the commercial multivariate analysis software The Unscrambler® version 9.8 (CAMO, Oslo, Norway).

5. Evaluation of the calibration model and prediction
Performance of the calibration model and cross-validation was evaluated using the following statistical parameters; coefficient of determination between predicted and measured L-AA concentration (R²), root mean square error of calibration (RMSEC), root mean square error of cross-validation (RMSECV), and ratio of standard deviation to prediction error (RPD). Other important parameters are the range error ratio (RER) or ratio of range of the original data (maximum–minimum value) to standard error of cross-validation (SECV). A good calibration model should have high R², low RMSEC and RMSECV, but also a small difference between the two (Camps & Christen, 2009). The quality of validation results were then evaluated using RPD and RER values. This is a qualitative measure for the assessment of the validation results. The smaller the error of prediction is, compared to the variance of the reference values, the larger the RPD value is, and therefore the better the model (Carles Collell et al., 2010; Conzen, 2006). An RPD between 1.5 and 2 means that the calibration model can discriminate between low and high values of the response variable; a value between 2 and 2.5 indicates that coarse quantitative predictions are possible, and a value between 2.5 and 3 or above correspond to good and excellent prediction accuracy, respectively. It is recommended that the RER should be 10 or
higher (Williams & Sobering, 1996; Conzen, 2006). Another important parameter is the number of PLS factors used to explain the model. This number also indicates the complexity of the model and is based on the minimum RMSECV. The performance of the final model was evaluated according to the root mean square error of prediction (RMSEP) and the coefficient determination ($R^2$) in the prediction set.

III Results and Discussion

1. Typical spectra of L-AA solution in THz region

Figure 2 shows the typical spectra of L-AA in solution and water in the THz region (20-400 cm$^{-1}$). Spectra data above 401 cm$^{-1}$ was not used because high noise was identified in the wavenumber range of 401-450 cm$^{-1}$. Both spectra are very similar in shape. Water is the main component in our L-AA solution. The absorbance peak of water spectrum in the THz region observed in this study is consistent with a previous report (Ogawa et al., 2009). The peak was located at 154.3 cm$^{-1}$ of wavenumber or at 4.6 THz. The spectrum of water is higher than that of L-AA in solution which indicates the influence of the L-AA molecule. The Terahertz region is rich in information about intra-molecular and inter-molecular stretching, and vibration of organic compounds (Laman et al., 2008). The presence of the L-AA molecules in water results in intra-molecular and inter-molecular stretching vibration mode associated with organic bonding between L-AA and water molecules (L-AA–water, water–water and L-AA–L-AA) and results in a weaker absorbance of a L-AA solution compared to that of pure water (Figure 2).

2. Calibration and cross-validation result of L-AA concentration determination

Using the FS-PLS regression method, calibration and cross-validation were performed on the original and several pre-processed spectra (Table 2). The developed calibration model was further validated by full cross-validation using the “leave one out” technique. In this technique, one of the calibration samples is removed. Subsequently, the removed sample was predicted with the revised model using the residual samples, and the procedure was repeated until each sample was excluded once from the model (Gurdeniz et al., 2007). Seen in Table 2, all of the calibration models had a very high $R^2 = 0.981-0.996$. The RMSEC values of L-AA varied from 0.396% to 0.909% (w/w). The RMSECV values varied from 1.430% to 2.598% (w/w). These values were very similar compared to a previous result of L-AA determination in powder and liquid samples. Yang & Iruudayaraj (2002) obtained SECV values 0.23%, 1.35% and 2.47% for FTIR-ATR mid infrared spectroscopy, NIR spectroscopy and Raman spectroscopy, respectively. This result shows that FTIR-ATR-THz spectroscopy has the potential for quantification of L-AA in aqueous solution. RPD and RER values were also satisfactory. Most of the developed calibration models had an RPD of more than 3 and an RER of more than 10, except for the Savitzky-Golay second derivative spectra.

Table 2 Calibration and validation result for L-AA concentration determination

<table>
<thead>
<tr>
<th>Pre-Processing</th>
<th>PLS factors</th>
<th>$R^2_{cal}$</th>
<th>RMSEC</th>
<th>RMSECV</th>
<th>RPD</th>
<th>RER</th>
<th>$R^2_{pred}$</th>
<th>RMSEP</th>
</tr>
</thead>
<tbody>
<tr>
<td>Original</td>
<td>12</td>
<td>0.994</td>
<td>0.507</td>
<td>1.513</td>
<td>4.36</td>
<td>14.26</td>
<td>0.884</td>
<td>3.627</td>
</tr>
<tr>
<td>MAS$^a$</td>
<td>12</td>
<td>0.993</td>
<td>0.542</td>
<td>1.517</td>
<td>4.35</td>
<td>14.22</td>
<td>0.897</td>
<td>3.659</td>
</tr>
<tr>
<td>MFS$^b$</td>
<td>12</td>
<td>0.994</td>
<td>0.493</td>
<td>1.430</td>
<td>4.61</td>
<td>15.08</td>
<td>0.899</td>
<td>3.777</td>
</tr>
<tr>
<td>GFS$^c$</td>
<td>15</td>
<td>0.996</td>
<td>0.396</td>
<td>1.475</td>
<td>4.48</td>
<td>14.66</td>
<td>0.912</td>
<td>4.322</td>
</tr>
<tr>
<td>SGS$^d$</td>
<td>11</td>
<td>0.993</td>
<td>0.546</td>
<td>1.518</td>
<td>4.34</td>
<td>14.21</td>
<td>0.906</td>
<td>3.615</td>
</tr>
<tr>
<td>SGF$^e$</td>
<td>10</td>
<td>0.991</td>
<td>0.621</td>
<td>1.474</td>
<td>4.48</td>
<td>14.64</td>
<td>0.916</td>
<td>2.791</td>
</tr>
<tr>
<td>SGSD$^f$</td>
<td>12</td>
<td>0.981</td>
<td>0.909</td>
<td>2.598</td>
<td>2.54</td>
<td>8.31</td>
<td>0.889</td>
<td>3.714</td>
</tr>
</tbody>
</table>

$^a$Moving average smoothing with number of segment:3
$^b$Median filter smoothing with number of segments:7
$^c$Gaussian filter smoothing with number of segment:5
$^d$Savitzky-Golay smoothing with number of segments:3, polynomial order:2
$^e$Savitzky-Golay first derivative with number of segments:11, polynomial order:2
$^f$Savitzky-Golay second derivative with number of segments:13, polynomial order:2

3. Prediction of L-AA concentration

To be valid as an analytical method, the capability of using the developed calibration models for prediction of L-AA concentrations of independent samples was confirmed. For this purpose, 20 samples of L-AA solution with different concentrations were used. The RMSEP obtained varied from 2.791% to 4.322%. This result was inferior to that obtained in a previous report of L-AA determination (Yang & Iruudayaraj, 2002). That study used NIR, Mid Infrared and Raman spectroscopy for L-AA determination, and reported an overall error of prediction 0.2-3.0%. As seen in Table 2, the calibration model using Savitzky-Golay first derivative
spectra resulted in the lowest RMSEP (RMSEP = 2.791%) and with the highest coefficient of determination (R²pred = 0.916). Hence, this calibration model was selected as the best calibration model for predicting L-AA concentration in aqueous solutions. A scatter plot between actual and predicted values for calibration and validation using Savitzky-Golay first derivative spectra is depicted in Fig. 3.

Figure 4 shows the scatter plot of the prediction result. By a 95% confidence paired t-test, there were no significant differences between the actual L-AA concentrations measured using the reference method and that predicted by FTIR-ATR-THz spectroscopy. This result shows that a calibration model for spectroscopy-based determination of L-AA concentration using FTIR-ATR-THz spectroscopy is possible. Prediction results show that FTIR-ATR-THz spectroscopy successfully predicted the concentration of L-AA in solution and has the potential for L-AA determination in complex systems, such as fruit juices and other aqueous systems.

4. Evaluating the structure of the calibration model
In order to clarify the behavior of the calibration model, the regression coefficient was plotted against the wavenumber (Figure 5). Both radiations at higher and lower frequencies contribute significantly to the calibration model for L-AA determination. However, it was clear that the regression coefficient in the lower frequencies is a little superior to that in higher frequencies. Thus, it confirms the important role of inter-molecular bonding force of L-AA in lower frequencies (Binghua et al., 2008). In the lower frequencies, the specific wavenumber, which has significant regression coefficient (regression coefficient is more than ±3000), includes wavenumbers of 53.99 cm⁻¹ (1.62 THz), 65.57 cm⁻¹ (1.97 THz), 84.85 cm⁻¹ (2.55 THz), 119.56 cm⁻¹ (3.59 THz), 142.71 cm⁻¹ (4.28 THz) and in the higher frequencies at 227.56 cm⁻¹ (6.83 THz), 320.13 cm⁻¹ (9.60 THz) and 362.55 cm⁻¹ (10.88 THz). These important wavenumbers may correspond with L-AA determination in aqueous systems. In a study of solid L-ascorbic acid measurement using THz-Time Domain Spectroscopy (TDS) in low frequencies it was reported that the inter-molecular vibration of solid L-ascorbic acid at 2.05 THz and 2.34 THz (Binghua et al., 2008). In another study using the transmittance method, spectra of solid L-AA (5% in concentration) in the range of 0-200 cm⁻¹ was successfully acquired. The transmittance data was transformed into absorbance data. Some peaks of L-AA absorbance in spectra of solid samples were found at 65.57 cm⁻¹ and 119.56 cm⁻¹ of
wavenumber (RIKEN Terahertz Database). However, since L-AA in solid and liquid samples forms different spectra in the THz region, further investigation is needed to clarify this finding.

Fig. 5 Regression coefficient plot calculated for L-AA determination (Calculations were conducted by using Savitzky-Golay first derivative spectra of wavenumber 20–400 cm\(^{-1}\) using FS-PLS regression.)

IV Summary and Conclusions

In summary, quantitative measurement of L-AA concentration in aqueous solutions using FTIR-ATR in terahertz (THz) region is possible. Our calibration and validation model results show a high correlation between terahertz spectra of L-AA solution and L-AA concentration. The calibration model using the calculated Savitzky-Golay first derivative was found to give the best prediction of L-AA concentrations in aqueous solutions. A prediction error of 2.791% was obtained and it seemed to be applicable for accurate determination of L-AA concentrations in aqueous solutions. The overall results indicate that the concentration of L-AA in aqueous solutions can be accurately determined by FTIR-ATR-THz spectroscopy coupled with a FS-PLS regression calibration model. Further research will focus on realizing the potential of using FTIR-ATR-THz spectroscopy for the determination of L-AA concentrations in complex systems, such as fruit juices and other aqueous systems.

Acknowledgement

This work was supported in part by KAKENHI (Grant-in-Aid for challenging Exploratory Research No. 23658209). We also are grateful to many of our academic colleagues for many discussions in this field.

References


RIKEN Terahertz Database: http://thzdb.org


(Received : 19. January. 2012, Accepted : 15. May. 2012)