


## The Chemical Aspects of Raman Spectroscopy: Statistical Structure-Spectrum Relationship in The Analyses of Bioflavonoids

Follow this and additional works at: <https://www.jfda-online.com/journal>

 Part of the [Food Science Commons](#), [Medicinal Chemistry and Pharmaceutics Commons](#), [Pharmacology Commons](#), and the [Toxicology Commons](#)



This work is licensed under a [Creative Commons Attribution-Noncommercial-No Derivative Works 4.0 License](#).

### Recommended Citation

Wang, Chih-Hsien; Huang, Chia-Chi; Chen, Wenlung; and Lai, Yen-Shi (2020) "The Chemical Aspects of Raman Spectroscopy: Statistical Structure-Spectrum Relationship in The Analyses of Bioflavonoids," *Journal of Food and Drug Analysis*: Vol. 28 : Iss. 2 , Article 5.

Available at: <https://doi.org/10.38212/2224-6614.1058>

This Original Article is brought to you for free and open access by Journal of Food and Drug Analysis. It has been accepted for inclusion in Journal of Food and Drug Analysis by an authorized editor of Journal of Food and Drug Analysis.

# The chemical aspects of Raman spectroscopy: Statistical structure-spectrum relationship in the analyses of bioflavonoids

Chih-Hsien Wang<sup>a</sup>, Chia-Chi Huang<sup>a,\*</sup>, Wenlung Chen<sup>a,\*\*</sup>, Yen-Shi Lai<sup>b</sup>

<sup>a</sup> Department of Applied Chemistry, National Chiayi University, No. 300, Syuefu Rd., Chiayi City 60004, Taiwan

<sup>b</sup> Colossus BioPharma Consultants Company, Ltd., NO 15, LN 98, Jhengde Road, Kaohsiung 80449, Taiwan

## Abstract

Raman spectroscopy has been accepted as a useful tool for the characterization of natural products. However, to identify a specific compound in a mixture sample of natural products using Raman spectra alone is highly challenging if not impossible. We demonstrated an effective solution to such issues using a method combining statistical Raman spectroscopy and Mass spectrometry. The method was validated with a successful application to the identification of the major anthocyanin components in a purple yam (*Dioscorea purpurea*) extract. Of particular interest is that statistical grouping of the bioflavonoid standards that formed the database of this study was found to correspond closely to the conventional chemical classification. An initial theory on the chemical aspects of Raman spectroscopy pertaining to the connectivity of Raman-active functional groups in bioflavonoids was developed based on the statistical correlation between chemical classification and Raman spectroscopy.

**Keywords:** FT-Raman spectroscopy, Anthocyanins, Catechins, Flavonoids, Statistical analysis

## 1. Introduction

*Dioscorea alata* L., *D. alata* L. var. *purpurea* (*Dioscorea purpurea* in short) is a species of yam featuring purple colored flesh characteristic of anthocyanins. It is a common food and a traditional medicine as well because of its functional constituents. A number of these constituents, including allantoin, crude anthocyanin,  $\beta$ -carotene lecithin, and dioscorin were investigated in one of our previous studies [1]. Except anthocyanin, these active compounds were identified unambiguously in this earlier work by Fourier transform Raman spectroscopy (FT-Raman).

FT-Raman with Nd: YAG lasers at 1064 nm, Michelson interferometers, and high-sensitivity GaAs detectors provided the advantages of high frequency precision, easy sample preparation, and

low fluorescence interference, and so has been widely applied in food and agriculture research [2–6]. Raman active molecules undergo a change in polarizability during molecular vibration leading to a set of Raman fingerprints that is useful for the characterization of phytochemicals [7,8]. With incompletely purified samples, however, the power of characterization can decrease drastically due to overlapping fingerprints.

Assemblies of phenolic fragments are ubiquitous in plant secondary metabolites such as bioflavonoids that include anthocyanins. Investigations of bioflavonoids using IR and Raman spectroscopies [9,10] have been conducted on purified standards and compilations of aromatic and hydroxy vibrational frequencies [11] were available. In our case, with omission of a tedious purification process, the composition of the principal *D. purpurea* anthocyanin fraction was so complex it was difficult to identify the

Received 7 November 2019; accepted 3 February 2020.  
Available online 27 June 2020

\* Corresponding author at:

\*\* Corresponding author at:

E-mail addresses: [chiachihuang@mail.ncyu.edu.tw](mailto:chiachihuang@mail.ncyu.edu.tw) (C.-C. Huang), [wlichen@mail.ncyu.edu.tw](mailto:wlichen@mail.ncyu.edu.tw) (W. Chen).

<https://doi.org/10.38212/2224-6614.1058>

2224-6614/Copyright © 2020 Food and Drug Administration, Taiwan. This is an open access article under the CC BY NC ND 4.0 license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

major anthocyanin based on this standardized information. This was a common issue in the study of many other plant materials and affected the throughput of Raman analysis of mixture in general. There was a practical need for methods that catalog the huge collection of standard spectral data in a way that mixture can be effectively analyzed by a transformed pattern in lieu of molecular fingerprinting.

A few studies have reorganized Raman spectral data with statistical methods such as hierarchical clustering and principal component analysis. The results provided correlations between spectral data and non-spectral properties such as color of flower petals in one example and species of citrus fruits in another [12,13]. The objective of the current study was to develop a simple spectral data processing method to aid the identification of anthocyanins in purple yam samples. Our method involved comparison of the FT-Raman spectrum of a real sample to a collection of standard bioflavonoids spectra. Specifically, the real sample was an HPLC fraction of purple yam extract. The comparison was made based on similarity distances between Raman spectra using a Pearson correlation algorithm [14]. Mass spectra of the real sample were subjected to structure analyses and used to validate the results of Pearson correlation. As bioflavonoids were chemically classified according to the connectivity pattern of their phenolic fragments, we further studied the correlation between structure classifications and Raman spectra of the bioflavonoid standards.

Rationalizations of the outcome in the form of structure-spectrum relationships were developed.

## 2. Materials and methods

Materials used in this study and the preparation of the real sample are detailed in the Supporting Information.

### 2.1. Spectral acquisition

FT-Raman spectra were obtained on a Bruker RFS-100 FT-spectrophotometer (Bruker Optik GmbH, Lubeck, Germany). Nd:YAG laser at 1064 nm (Coherent Lubeck GmbH, Lubeck, Germany) pumped by a diode laser was employed for excitation. Aiming of light on the samples were assisted by a HeNe laser beam cooperated with the excitation light. The output laser power was 100 mW. The resolution was at 4 cm<sup>-1</sup>, and the scattered radiation was collected at 180° with a set of reflectors. GaAs detector was cooled by liquid nitrogen, and Raman spectra were recorded over the Raman shift range 1–3500 cm<sup>-1</sup>. Each sample was scanned 500 times.

FT-Raman spectra reported in this study were all-original and were not smoothed, normalized, or baseline corrected by data manipulation.

### 2.2. Data treatment

Spectral data, in the Raman shift range of 1–3500 cm<sup>-1</sup> at 2 wavenumber-unit intervals, amounting to a total of 1815 Raman shift variables, was processed and tabulated in Microsoft Excel. Statistical analyses were conducted in the IBM SPSS environment. Statistical score by Pearson correlation was used to measure the co-variance of Raman intensities within the space of 1815 dimensions of Raman shift. The initial result was then compared by means of similarity matrix, hierarchical clustering and multidimensional scaling. Statistical analyses of the structure indices and canonical correlation between spectral and structural data were conducted in a similar manner. For more detail about these methods and their applications, see Ref. [15]. Some result presentations such as conditional formatting and graphing were made with the aid of Microsoft Excel.

## 3. Result and discussion

### 3.1. FT-Raman database of flavonoid compounds

Twenty-eight bioflavonoids were chosen as the standards for this study and their FT-Raman spectra were acquired. These standards included 10 anthocyanins, 8 catechins (aka flavonols), and 10 flavonoids, with the last subgroup being loosely defined to allow for negative control. In Fig. 1 we show the basic skeleton and a representative molecule for each of these three subclasses. Table 1 gives the ID numbers and structure indices of the twenty-eight compounds. These indices were used as the basis of structure correlation later. The structures and FT-Raman spectra (printed over the range 350 cm<sup>-1</sup> to 1800 cm<sup>-1</sup>) of all 28 standards are included in the Supporting Information.

### 3.2. Partial purification of *Dioscorea purpurea* anthocyanins by HPLC and preparation of the real sample

Initially, fractionation of the yam extract was preceded by hydrolysis and de-glycosylation treatments in an effort to release the anthocyanidins for analyses. We have since realized that these pretreatments were detrimental to the preservation of the anthocyanin integrity. Therefore, the yam extract was fractionated by HPLC without any

pretreatments this time. The principal anthocyanin fractions included ten peaks, as shown in the chromatogram (Fig. S1, Supporting Information). Further analyses by Mass spectrometry showed that none of these peaks were homogeneous (Table S1, Supporting Information). The most populous fraction in the chromatogram, namely peak number 5, was used as the real sample in the FT-Raman study. It is difficult to identify an anthocyanin from the FT-Raman spectrum (Fig. S2, Supporting Information) of this sample. By referring to the closest literature, Raman peak assignments could be exhausted as follows:  $476\text{ cm}^{-1}$  ( $\tau\text{CC}$  vibration),  $540\text{ cm}^{-1}$  ( $\delta\text{CC}$  vibration),  $1184\text{ cm}^{-1}$  (OH stretching),  $1134$ ,  $1325$ ,  $1530$  and  $1598\text{ cm}^{-1}$  ( $\nu\text{CC}$  vibration), and  $1631\text{ cm}^{-1}$  ( $\nu\text{CC}$  and  $\nu\text{CO}$  vibrations) [16]. This, however, did little to help one reconstruct the structure of the compound because information about the connectivity of functional groups that differentiates various anthocyanins is absent.

### 3.3. Statistical treatments of Raman spectra

Peak intensities of the 28 standard Raman spectra as well as the real sample were tabulated across wavenumber ticks between  $1\text{ cm}^{-1}$  and  $3500\text{ cm}^{-1}$  for a total of 29 cases and 1815 variables (ticks). This table was imported to the SPSS software wherein similarity distances between all possible pairs of the 29 spectra were calculated against the 1815 variables based on Pearson product-moment. The results were shown in Fig. 2 as a proximity matrix where each cell contained a number proportional to the degree of pairwise similarity ranging from 0.00 to 1.00, with 1.00 indicating statistical identity. By

color-coding all the cells with a number greater than 0.60, three blocks were formed as outlined by the blue rectangles which coincided more or less with the preassignments in Table 1. Based on the simple formula of the catechins block (row/column 11–18) had the highest rate at 100%, followed by the anthocyanins (row/column 1–10) at 82%, and the flavonoids (row/column 19–28) at a borderline 52% that conformed well with the allowance for negative controls.

$$\% \text{ positive response} = 100 * (\text{number of red cells} / \text{number of all cells}),$$

In terms of individuals in each block, petunidin (No. 10) was highly similar to cyanidin (No. 2) and delphinidin (No. 3) but much less so to the remaining anthocyanins. Flavonone (No. 23) was notable for its similarity to the catechins as opposed to the other flavonoids. Other poor flavonoids performers included apigenin (No. 20), for its similarity to the anthocyanins, and flavone (No. 22) for essential dissociation from all the groups. Kaempferol (No. 25) and quercetin (No. 28) were both borderline between anthocyanins and flavonoids.

A remarkable outcome was that glycosylation was not a determining cause of dissimilarity so that malvidin (No. 4) was indistinguishable from malvin (No. 5), malvidin from oenin (No. 6), and pelargonidin (No. 7) from pelargonin (No. 8). Esterification by gallic acid was not determining either. Thus, epicatechin gallate (No. 13) was indistinguishable from epicatechin (No. 14), epigallocatechin gallate (No. 15) from epigallocatechin (No. 16) and galocatechin gallate (No. 17) from galocatechin (No. 18). This was fortunate for us since bypassing the hydrolysis/deglycosylation pretreatments during real sample preparation would be inconsequential for validating the statistical outcomes with Mass spectrometry as discussed later. Finally, the Raman spectrum of the real sample (No. 29) was found to associate with the anthocyanins only, highly similar to petunidin (No. 10) and cyanidin (No. 2).

We next applied a multiple dimensional scaling (MDS) to the set of 29 Raman spectra to confirm the results of the Pearson correlation. This task assigned each spectrum to a specific location in a conceptual two-dimensional space such that the distances between points in the space matched the given dissimilarities. As shown in Fig. 3 the MDS trended toward three groups, tightly with the catechins (VAR 11–18) and anthocyanins (VAR1-10) and loosely with the flavonoids (VAR19-28). Again, the real sample (VAR29) was shown to be distant from the catechins and the flavonoids and resembled

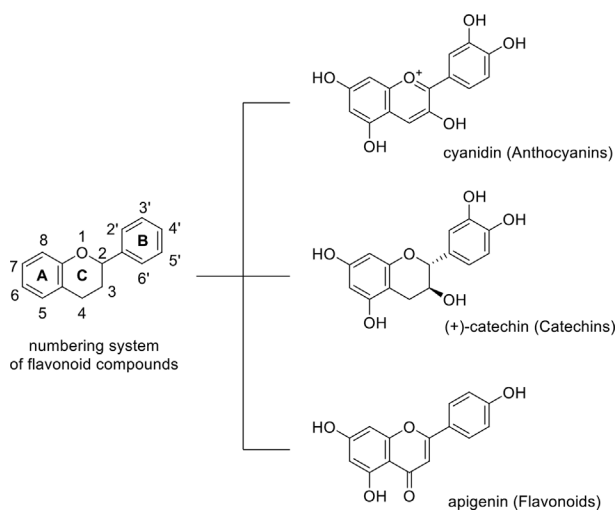


Fig. 1. Basic skeleton, conventional classification, and examples of flavonoids subclasses.



Table 1. Pre-categorization and structure indices of 28 flavonoid standards.

	No	Name	C2	C3	C4	C5	C6	C7	C8	C2'	C3'	C4'	C5'	C6'
anthocyanins	1	callistephin chloride	—	OGlu	H	OH	H	OH	H	H	OMe	OH	H	H
	2	cyanidin chloride	—	OH	H	OH	H	OH	H	H	OH	OH	H	H
	3	delphinidin chloride	—	OH	H	OH	H	OH	H	H	OH	OH	OH	H
	4	malvidin chloride	—	OH	H	OH	H	OH	H	H	OMe	OH	OMe	H
	5	malvin chloride	—	OGlu	H	OGlu	H	OH	H	H	OMe	OH	OMe	H
	6	oenin chloride	—	OGlu	H	OH	H	OH	H	H	OMe	OH	OMe	H
	7	pelargonidin chloride	—	OH	H	OH	H	OH	H	H	H	OH	H	H
	8	pelargonin chloride	—	OGlu	H	OGlu	H	OH	H	H	H	OH	H	H
	9	peonidin chloride	—	OH	H	OH	H	OH	H	H	OMe	OH	H	H
catechins	10	petunidin chloride	—	OH	H	OH	H	OH	H	H	OH	OH	OMe	H
	11	(+)-catechin	$\beta$ H	$\beta$ OH, $\alpha$ H	H,H	OH	H	OH	H	H	OH	OH	H	H
	12	(-)-catechin hydrate	$\alpha$ H	$\alpha$ OH, $\beta$ H	H,H	OH	H	OH	H	H	OH	OH	H	H
	13	(-)-epicatechin gallate	$\alpha$ H	$\beta$ OGal, $\alpha$ H	H,H	OH	H	OH	H	H	OH	OH	H	H
	14	(-)-epicatechin	$\beta$ H	$\beta$ OH, $\alpha$ H	H,H	OH	H	OH	H	H	OH	OH	H	H
	15	(-)-epigallocatechin gallate	$\beta$ H	$\alpha$ OGal, $\beta$ H	H,H	OH	H	OH	H	H	OH	OH	OH	H
	16	(-)-epigallocatechin	$\beta$ H	$\alpha$ OH, $\beta$ H	H,H	OH	H	OH	H	H	OH	OH	OH	H
	17	(-)-gallocatechin gallate	$\alpha$ H	$\alpha$ OGal, $\beta$ H	H,H	OH	H	OH	H	H	OH	OH	OH	H
flavonoids	18	(-)-gallocatechin	$\alpha$ H	$\alpha$ OH, $\beta$ H	H,H	OH	H	OH	H	H	OH	OH	OH	H
	19	3-hydroxyflavone	—	OH	OO	H	H	H	H	H	H	H	H	H
	20	apigenin	—	H	OO	OH	H	OH	H	H	H	OH	H	H
	21	daidzein	H	—	OO	H	H	OH	H	H	H	OH	H	H
	22	flavone	—	H	OO	H	H	H	H	H	H	H	H	H
	23	flavonone	$\alpha$ H	H,H	OO	H	H	H	H	H	H	H	H	H
	24	genistein	H	—	OO	OH	H	OH	H	H	H	OH	H	H
	25	kaempferol	—	OH	OO	OH	H	OH	H	H	H	OH	H	H
	26	luteolin	—	H	OO	OH	H	OH	H	H	OH	OH	H	H
	27	myricetin	—	OH	OO	OH	H	OH	H	H	OH	OH	OH	H
	28	quercetin	—	OH	OO	OH	H	OH	H	H	OH	OH	H	H
	29	real sample												

Proximity Matrix																													
Correlation between Vectors of Values																													
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29
1	1.00	0.71	0.69	0.89	0.73	0.86	0.92	0.93	0.94	0.59	0.62	0.45	0.47	0.39	0.45	0.35	0.45	0.54	0.32	0.84	0.31	0.31	0.27	0.39	0.74	0.67	0.49	0.69	0.33
2	0.71	1.00	0.86	0.69	0.49	0.85	0.64	0.78	0.70	0.93	0.34	0.09	0.26	0.09	0.23	0.18	0.25	0.31	0.25	0.64	0.19	0.20	0.04	0.23	0.58	0.51	0.39	0.55	0.60
3	0.69	0.86	1.00	0.60	0.64	0.69	0.68	0.66	0.63	0.90	0.30	0.09	0.35	0.10	0.27	0.16	0.32	0.30	0.35	0.59	0.25	0.30	0.07	0.33	0.67	0.63	0.57	0.69	0.57
4	0.89	0.69	0.60	1.00	0.62	0.81	0.77	0.93	0.92	0.51	0.50	0.39	0.29	0.25	0.29	0.20	0.30	0.41	0.19	0.90	0.19	0.21	0.16	0.24	0.58	0.55	0.29	0.51	0.34
5	0.73	0.49	0.64	0.62	1.00	0.48	0.71	0.56	0.69	0.55	0.37	0.29	0.34	0.24	0.33	0.29	0.32	0.37	0.36	0.53	0.24	0.25	0.18	0.34	0.54	0.59	0.51	0.64	0.24
6	0.86	0.85	0.69	0.81	0.48	1.00	0.69	0.95	0.85	0.66	0.58	0.31	0.35	0.30	0.35	0.25	0.36	0.49	0.17	0.76	0.19	0.17	0.15	0.26	0.59	0.51	0.30	0.51	0.48
7	0.92	0.64	0.68	0.77	0.71	0.69	1.00	0.78	0.84	0.59	0.48	0.34	0.43	0.33	0.40	0.28	0.39	0.44	0.31	0.72	0.28	0.33	0.25	0.36	0.69	0.61	0.54	0.68	0.33
8	0.93	0.78	0.66	0.93	0.56	0.95	0.78	1.00	0.93	0.58	0.62	0.42	0.40	0.34	0.39	0.28	0.40	0.52	0.21	0.87	0.23	0.23	0.21	0.29	0.64	0.57	0.33	0.55	0.41
9	0.94	0.70	0.63	0.92	0.69	0.85	0.84	0.93	1.00	0.55	0.56	0.41	0.38	0.32	0.36	0.28	0.36	0.47	0.24	0.85	0.24	0.24	0.22	0.31	0.65	0.58	0.40	0.61	0.34
10	0.59	0.93	0.90	0.51	0.55	0.66	0.59	0.58	0.55	1.00	0.21	-0.02	0.23	0.01	0.18	0.14	0.20	0.22	0.31	0.47	0.20	0.22	0.00	0.25	0.57	0.51	0.49	0.61	0.62
11	0.62	0.34	0.30	0.50	0.37	0.58	0.48	0.62	0.56	0.21	1.00	0.91	0.69	0.80	0.68	0.73	0.70	0.79	0.32	0.54	0.49	0.40	0.63	0.58	0.50	0.54	0.33	0.45	0.10
12	0.45	0.09	0.09	0.39	0.29	0.31	0.34	0.42	0.41	-0.02	0.91	1.00	0.60	0.75	0.63	0.77	0.63	0.70	0.27	0.44	0.44	0.38	0.66	0.53	0.36	0.43	0.23	0.31	-0.08
13	0.47	0.26	0.35	0.29	0.34	0.35	0.43	0.40	0.38	0.23	0.69	0.60	1.00	0.68	0.71	0.60	0.79	0.63	0.52	0.39	0.64	0.49	0.60	0.67	0.56	0.59	0.58	0.60	0.09
14	0.39	0.09	0.10	0.25	0.24	0.30	0.33	0.34	0.32	0.01	0.80	0.75	0.68	1.00	0.66	0.71	0.65	0.71	0.32	0.31	0.53	0.36	0.67	0.58	0.32	0.40	0.28	0.35	-0.03
15	0.45	0.23	0.27	0.29	0.33	0.35	0.40	0.39	0.36	0.18	0.68	0.63	0.71	0.66	1.00	0.69	0.71	0.70	0.42	0.37	0.53	0.48	0.72	0.52	0.52	0.53	0.42	0.48	0.06
16	0.35	0.18	0.16	0.20	0.29	0.25	0.28	0.28	0.28	0.14	0.73	0.77	0.60	0.71	0.69	1.00	0.63	0.72	0.31	0.26	0.49	0.42	0.69	0.56	0.32	0.39	0.30	0.34	-0.03
17	0.45	0.25	0.32	0.30	0.32	0.36	0.39	0.40	0.36	0.20	0.70	0.63	0.79	0.65	0.71	0.63	1.00	0.72	0.53	0.38	0.67	0.53	0.53	0.63	0.48	0.54	0.56	0.55	0.07
18	0.54	0.31	0.30	0.41	0.37	0.49	0.44	0.52	0.47	0.22	0.79	0.70	0.63	0.71	0.70	0.72	0.72	1.00	0.39	0.45	0.59	0.44	0.61	0.62	0.44	0.50	0.41	0.49	0.09
19	0.32	0.25	0.35	0.19	0.36	0.17	0.31	0.21	0.24	0.31	0.32	0.27	0.52	0.32	0.42	0.31	0.53	0.39	1.00	0.33	0.72	0.66	0.31	0.61	0.59	0.71	0.68	0.65	0.11
20	0.84	0.64	0.59	0.90	0.53	0.76	0.72	0.87	0.85	0.47	0.54	0.44	0.39	0.31	0.37	0.26	0.38	0.45	0.33	1.00	0.27	0.33	0.25	0.34	0.68	0.67	0.34	0.56	0.29
21	0.31	0.19	0.25	0.19	0.24	0.19	0.28	0.23	0.24	0.20	0.49	0.44	0.64	0.53	0.53	0.49	0.67	0.59	0.72	0.27	1.00	0.61	0.45	0.74	0.45	0.60	0.63	0.58	0.08
22	0.31	0.20	0.30	0.21	0.25	0.17	0.33	0.23	0.24	0.22	0.40	0.38	0.49	0.36	0.48	0.42	0.53	0.44	0.66	0.33	0.61	1.00	0.42	0.53	0.49	0.61	0.53	0.49	0.08
23	0.27	0.04	0.07	0.16	0.18	0.15	0.25	0.21	0.22	0.00	0.63	0.66	0.60	0.67	0.72	0.69	0.53	0.61	0.31	0.25	0.45	0.42	1.00	0.50	0.33	0.38	0.20	0.26	-0.05
24	0.39	0.23	0.33	0.24	0.34	0.26	0.36	0.29	0.31	0.25	0.58	0.53	0.67	0.58	0.52	0.56	0.63	0.62	0.61	0.34	0.74	0.53	0.50	1.00	0.48	0.62	0.57	0.60	0.08
25	0.74	0.58	0.67	0.58	0.54	0.59	0.69	0.64	0.65	0.57	0.50	0.36	0.56	0.32	0.52	0.32	0.48	0.44	0.59	0.68	0.45	0.49	0.33	0.48	1.00	0.81	0.67	0.81	0.28
26	0.67	0.51	0.63	0.55	0.59	0.51	0.61	0.57	0.58	0.51	0.54	0.43	0.59	0.40	0.53	0.39	0.54	0.50	0.71	0.67	0.60	0.61	0.38	0.62	0.81	1.00	0.64	0.76	0.23
27	0.49	0.39	0.57	0.29	0.51	0.30	0.54	0.33	0.40	0.49	0.33	0.23	0.58	0.28	0.42	0.30	0.56	0.41	0.68	0.34	0.63	0.53	0.20	0.57	0.67	0.64	1.00	0.87	0.20
28	0.69	0.55	0.69	0.51	0.64	0.51	0.68	0.55	0.61	0.61	0.45	0.31	0.60	0.35	0.48	0.34	0.55	0.49	0.65	0.56	0.58	0.49	0.26	0.60	0.81	0.76	0.87	1.00	0.29
29	0.33	0.60	0.57	0.34	0.24	0.48	0.33	0.41	0.34	0.62	0.10	-0.08	0.09	-0.03	0.06	-0.03	0.07	0.09	0.11	0.29	0.08	0.08	-0.05	0.08	0.28	0.23	0.20	0.29	1.00

This is a similarity matrix

Fig. 2. Spectral proximity matrix of the bioflavonoid standards and real sample.

cyanidin (VAR2) and oenin (VAR6) better than petunidin (VAR10) as measured by the concentric purple dashed circle.

Hierarchical clustering of the 29 spectra resulted in the dendrogram in Fig. 4. Three groups were formed at the heterogeneity level around 15. The flavonoids were the most dissociated, with apigenin, daidzein, flavonone and genistein being misplaced with the anthocyanins and the catechins. The real sample was clustered with the anthocyanins on a secondary level, being closest to cyanidin, delphinidin, and petunidin.

In summary, all three statistical models agreed in two terms. There were clear associations within the catechins and the anthocyanins, but less so with the flavonoids as it was by design of negative control. The real sample was recognized as an anthocyanin and, specifically, an analog of cyanidin. For independent support, a cross reference was made to a structure analysis of the real sample's MS spectrum. The results (peak 5, Table S1, Supporting Information) suggested that the anthocyanins in this sample were cyanidin glucoside or pyranoside esters of ferulic acid or sinapic acid. Since glycosylation/esterification would not prevent the statistical model from recognizing anthocyanidins by the Raman spectra of anthocyanins or esters, the statistical models have concluded highly consistently with MS in this case.

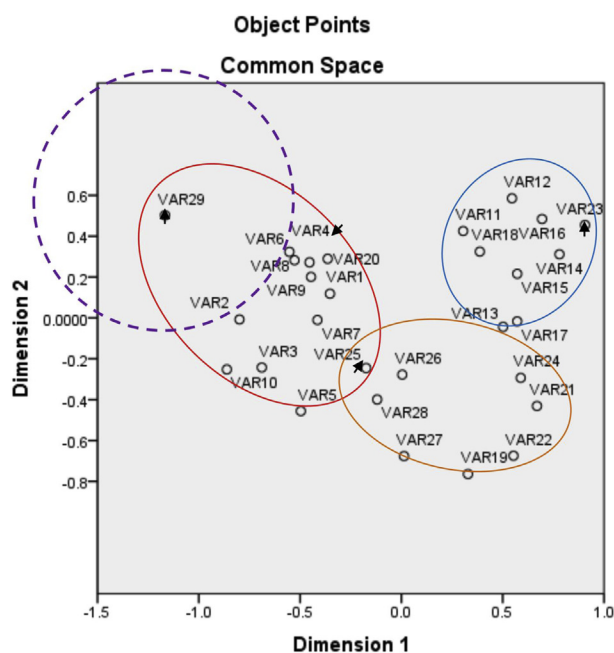


Fig. 3. Multiple dimensional scaling of the bioflavonoid standards and real sample.

### 3.4. The chemical aspects of Raman spectroscopy

Though broadly applied to the study of chemicals, a particularly underdeveloped application of Raman spectroscopy was in the elucidation of chemical structures. We have learned from the current study that statistical modeling of Raman data cataloged bioflavonoids in a way similar to structure-based grouping. To gain more insight into the intersection of Raman spectrum and chemical structure, we conducted a statistical analysis of the structure of the 28 bioflavonoid standards. The structure indices in Table 1 were recoded to scale values as shown in Table S2 (Supporting Information). With the constant variables C6, C8, C2', and C6' in Table 1 excluded, distances between all possible pairs of the 28 structures were calculated by Pearson product-moment and the resultant proximity matrix is shown in Fig. 5. Application of the 60%-or-better-similarity rule cataloged the structures into anthocyanins, catechins, and flavonoids at 88%, 100%, and 62% responsive rates. By this result, it would seem that natural products chemists had run their own statistical analyses in mind when the conventional classification of bioflavonoids was devised decades ago. As there appeared to be a common place between Figs. 2 and 5, we were curious whether there was a dependency between the two seemingly independent datasets.

Assignment of values to the structure indices in the recoding process was random. Therefore, they were standardized to Z scores. The Z scores of the structure variables C2, C3, C4, C5, C7, C3', C4', and C5' were then correlated to the intensity of the 1815 spectrum variables using the canonical correlation algorithm provided by SPSS. The task was set to tell how the index variables will predict the spectrum variables and the results are presented in Fig. 6, which showed the association of individual structure indices and Raman shifts with a significance level of  $p < 0.005$  (two-tailed). For discussion purpose, one exception was allowed for C4, which had a significance of  $0.005 < p < 0.007$ . Fig. 6 shows that the structure influence on Raman spectra was long range and involved electronic resonance as a key mechanism.

Molecular orbital hybridizations at C2 and C3 are key for the conventional classification of flavonoids, i.e.  $sp^3$  in catechins and  $sp^2$  in anthocyanins/flavonoids. Anthocyanin/flavonoid C2=C3 bonds are conjugated to both ring A and ring B and are in place to influence the vibrations of the phenols by resonance with the oxygen lone pairs. Indeed, C2 and C3 were found to correlate with benzene ring stretching ( $1550\text{--}1630\text{ cm}^{-1}$ ). C3 might also affect

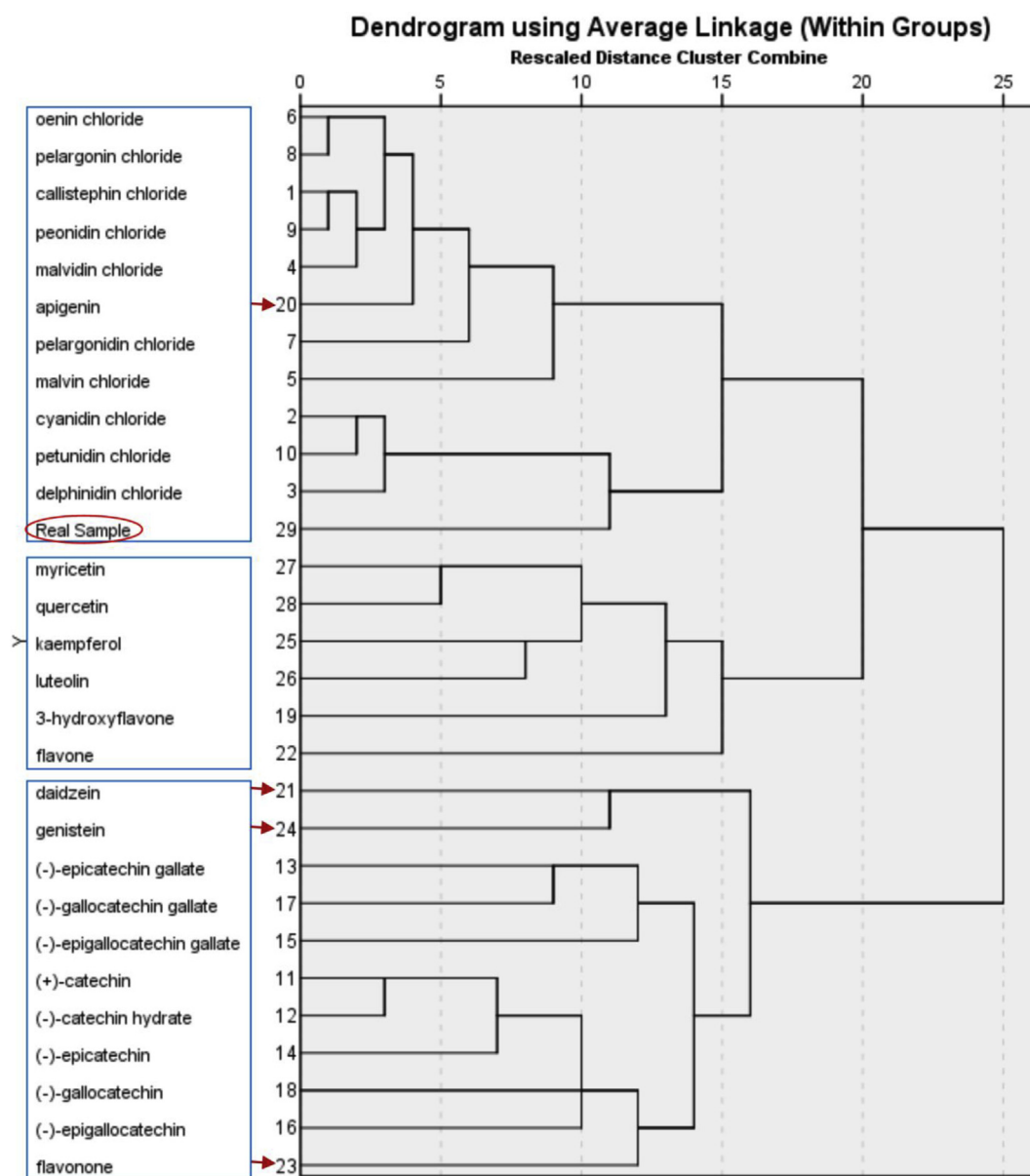


Fig. 4. Hierarchical clustering of the bioflavonoid standards and real sample.

the phenolic COH bending in the range of  $1300\text{--}1360\text{ cm}^{-1}$  [17]. C4 was significant in anthocyanins where it could resonate with the oxygen lone pairs at C3 and couple with glucose vibrations at  $2960\text{--}2900\text{ cm}^{-1}$  and  $560\text{--}500\text{ cm}^{-1}$  [18]. The lower p value of C4 was consistent with the observation in Fig. 2 that glycosylation did not lessen the similarity between anthocyanin and anthocyanidin. Hydroxy substituents on C5, C7, and C4' could contribute their lone pairs to an extended  $\pi$ -system involving ring C and affect the vibrations of the benzene rings as well as the phenol groups. C3' and C5' are *meta* to the rest of the molecules in all three subclasses of

compounds and, being deprived of chance of remote couplings via resonance, have found no significance in Fig. 6.

Retrospectively, the above scheme explains why petunidin (No. 10), cyanidin (No. 2), and delphinidin (No. 3) were not differentiated by FT-Raman despite that they were structurally differentiated at C5'. Flavonone (No. 23) was included in the catechins flock in Fig. 2 because they all have saturated C2–C3 bonds. The enhanced weight of C2–C3 in the absence of resonance was presumably also the main driver behind the high responsive rate within the catechins themselves.



	Correlation between Vectors of Values																											
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28
1	1.00	0.98	0.81	0.77	0.74	0.76	0.72	0.72	0.99	0.73	0.67	0.43	0.50	0.67	0.34	0.38	0.29	0.32	-0.13	0.31	0.18	-0.27	-0.19	0.35	0.40	0.55	0.53	0.65
2	0.98	1.00	0.83	0.76	0.71	0.74	0.83	0.81	0.99	0.75	0.58	0.33	0.42	0.58	0.21	0.26	0.18	0.22	-0.05	0.31	0.23	-0.27	-0.17	0.40	0.49	0.49	0.54	0.67
3	0.81	0.83	1.00	0.99	0.96	0.97	0.69	0.67	0.83	0.99	0.50	0.30	0.37	0.50	0.39	0.44	0.34	0.38	-0.14	0.18	0.11	-0.34	-0.18	0.26	0.32	0.32	0.65	0.48
4	0.77	0.76	0.99	1.00	0.98	0.99	0.56	0.55	0.77	0.99	0.49	0.31	0.38	0.49	0.43	0.49	0.39	0.43	-0.17	0.10	0.05	-0.34	-0.17	0.17	0.22	0.29	0.65	0.42
5	0.74	0.71	0.96	0.98	1.00	0.99	0.48	0.49	0.73	0.97	0.50	0.34	0.39	0.50	0.50	0.55	0.44	0.47	-0.22	0.09	0.07	-0.33	-0.19	0.13	0.15	0.31	0.62	0.38
6	0.76	0.74	0.97	0.99	0.99	1.00	0.52	0.52	0.75	0.98	0.54	0.38	0.42	0.54	0.53	0.57	0.47	0.50	-0.23	0.13	0.04	-0.34	-0.19	0.17	0.19	0.34	0.63	0.40
7	0.72	0.83	0.69	0.56	0.48	0.52	1.00	0.98	0.75	0.62	0.37	0.14	0.23	0.37	0.04	0.08	0.01	0.05	0.05	0.46	0.35	-0.21	-0.17	0.55	0.67	0.33	0.44	0.54
8	0.72	0.81	0.67	0.55	0.49	0.52	0.98	1.00	0.73	0.60	0.40	0.18	0.25	0.40	0.10	0.14	0.06	0.08	0.00	0.50	0.41	-0.20	-0.21	0.55	0.65	0.37	0.43	0.55
9	0.99	0.99	0.83	0.77	0.73	0.75	0.75	0.73	1.00	0.75	0.61	0.36	0.44	0.61	0.24	0.29	0.21	0.24	-0.06	0.26	0.19	-0.28	-0.17	0.34	0.42	0.50	0.54	0.66
10	0.73	0.75	0.99	0.99	0.97	0.98	0.62	0.60	0.75	1.00	0.46	0.27	0.34	0.46	0.41	0.46	0.37	0.41	-0.15	0.14	0.08	-0.34	-0.17	0.21	0.27	0.27	0.65	0.41
11	0.67	0.58	0.50	0.49	0.50	0.54	0.37	0.40	0.61	0.46	1.00	0.96	0.97	1.00	0.83	0.85	0.86	0.87	-0.33	0.46	0.34	-0.19	0.33	0.53	0.31	0.65	0.46	0.51
12	0.43	0.33	0.30	0.31	0.34	0.38	0.14	0.18	0.36	0.27	0.96	1.00	0.99	0.96	0.89	0.90	0.94	0.94	-0.41	0.39	0.27	-0.17	0.41	0.44	0.15	0.55	0.31	0.33
13	0.50	0.42	0.37	0.38	0.39	0.42	0.23	0.25	0.44	0.34	0.97	0.99	1.00	0.97	0.84	0.86	0.90	0.92	-0.35	0.37	0.31	-0.21	0.46	0.48	0.22	0.54	0.37	0.40
14	0.67	0.58	0.50	0.49	0.50	0.54	0.37	0.40	0.61	0.46	1.00	0.96	0.97	1.00	0.83	0.85	0.86	0.87	-0.33	0.46	0.34	-0.19	0.33	0.53	0.31	0.65	0.46	0.51
15	0.34	0.21	0.39	0.43	0.50	0.53	0.04	0.10	0.24	0.41	0.83	0.89	0.84	0.83	1.00	1.00	0.97	0.96	-0.50	0.35	0.15	-0.15	0.21	0.30	0.03	0.51	0.35	0.18
16	0.38	0.26	0.44	0.49	0.55	0.57	0.08	0.14	0.29	0.46	0.85	0.90	0.86	0.85	1.00	1.00	0.98	0.97	-0.48	0.35	0.17	-0.17	0.23	0.32	0.06	0.51	0.40	0.22
17	0.29	0.18	0.34	0.39	0.44	0.47	0.01	0.06	0.21	0.37	0.86	0.94	0.90	0.86	0.97	0.98	1.00	1.00	-0.47	0.30	0.18	-0.19	0.37	0.32	0.02	0.44	0.33	0.17
18	0.32	0.22	0.38	0.43	0.47	0.50	0.05	0.08	0.24	0.41	0.87	0.94	0.92	0.87	0.96	0.97	1.00	1.00	-0.45	0.29	0.19	-0.20	0.39	0.34	0.05	0.43	0.36	0.20
19	-0.13	-0.05	-0.14	-0.17	-0.22	-0.23	0.05	0.00	-0.06	-0.15	-0.33	-0.41	-0.35	-0.33	-0.50	-0.48	-0.47	-0.45	1.00	0.41	0.59	0.82	0.48	0.49	0.64	0.32	0.51	0.57
20	0.31	0.31	0.18	0.10	0.09	0.13	0.46	0.50	0.26	0.14	0.46	0.39	0.37	0.46	0.35	0.35	0.30	0.29	0.41	1.00	0.79	0.61	0.35	0.93	0.87	0.87	0.66	0.76
21	0.18	0.23	0.11	0.05	0.07	0.04	0.35	0.41	0.19	0.08	0.34	0.27	0.31	0.34	0.15	0.17	0.18	0.19	0.59	0.79	1.00	0.60	0.60	0.86	0.83	0.68	0.66	0.73
22	-0.27	-0.27	-0.34	-0.34	-0.33	-0.34	-0.21	-0.20	-0.28	-0.34	-0.19	-0.17	-0.21	-0.19	-0.15	-0.17	-0.19	-0.20	0.82	0.61	0.60	1.00	0.46	0.52	0.54	0.54	0.45	0.49
23	-0.19	-0.17	-0.18	-0.17	-0.19	-0.19	-0.17	-0.21	-0.17	-0.17	0.33	0.41	0.46	0.33	0.21	0.23	0.37	0.39	0.48	0.35	0.60	0.46	1.00	0.58	0.41	0.34	0.42	0.41
24	0.35	0.40	0.26	0.17	0.13	0.17	0.55	0.55	0.34	0.21	0.53	0.44	0.48	0.53	0.30	0.32	0.32	0.34	0.49	0.93	0.86	0.52	0.58	1.00	0.94	0.80	0.73	0.83
25	0.40	0.49	0.32	0.22	0.15	0.19	0.67	0.65	0.42	0.27	0.31	0.15	0.22	0.31	0.03	0.06	0.02	0.05	0.64	0.87	0.83	0.54	0.41	0.94	1.00	0.71	0.75	0.87
26	0.55	0.49	0.32	0.29	0.31	0.34	0.33	0.37	0.50	0.27	0.65	0.55	0.54	0.65	0.51	0.51	0.44	0.43	0.32	0.87	0.68	0.54	0.34	0.80	0.71	1.00	0.75	0.87
27	0.53	0.54	0.65	0.65	0.62	0.63	0.44	0.43	0.54	0.65	0.46	0.31	0.37	0.46	0.35	0.40	0.33	0.36	0.51	0.66	0.66	0.45	0.42	0.73	0.75	0.75	1.00	0.86
28	0.65	0.67	0.48	0.42	0.38	0.40	0.55	0.54	0.66	0.41	0.51	0.33	0.40	0.51	0.18	0.22	0.17	0.20	0.57	0.76	0.73	0.49	0.41	0.83	0.87	0.87	0.86	1.00
This is a similarity matrix																												

This is a similarity matrix

Fig. 5. Structural proximity matrix of the bioflavonoid standards.

The structure-spectrum relationship would also have predicted the placement of apigenin (No. 20) and kaempferol (No. 25) with the anthocyanins because they all have  $sp^2$ -hybridized C2–C3 as well as a C4' OH group. The significant structural presence or absence of hydroxy groups at C3' and C5' was statistically inconsequential because C3' and C5' were not an effector. The effect of C4 was not universal, which might be an addition reason why its p value was high in the canonical correlation. Thus, monoglycosylation, though a massive structure modification, did not warrant a decisive spectral differentiation of malvidin (No.4) from malvin (No. 5), neither from oenin (No. 6). The same went to pelargonidin (No. 7) and pelargonin (No. 8). The effect of C4 was more significant between the bis-glucoside malvin and mono-glucoside oenin as shown in Fig. 2. Bis-glucosidation would pose higher conformational restraint to and elevate the C4-mediated coupling with the sugar vibrations.

#### 4. Conclusions

Similarity between spectra was a drawback in the analysis of flavonoids mixture by Raman spectroscopy. We have established a method of processing FT-Raman spectral data to turn this around. In this method, Pearson correlation, multiple dimensional scaling, and hierarchical clustering based on similarity measures to classify 28 bioflavonoids standards as anthocyanins, catechins, and flavonoids. The model was applied successfully to identify the major anthocyanin components in a purple yam extract as cyanidin derivatives. It was hoped that this would set a precedent for resolving similar issues in the studies of other plant materials in general; and, with this example in order, the FT-Raman data of the 28 bioflavonoid standards can be repeatedly used by us or an interested colleague to analyze another bioflavonoid-containing real sample. Statistical classifications of the bioflavonoid standards based on their Raman spectra and



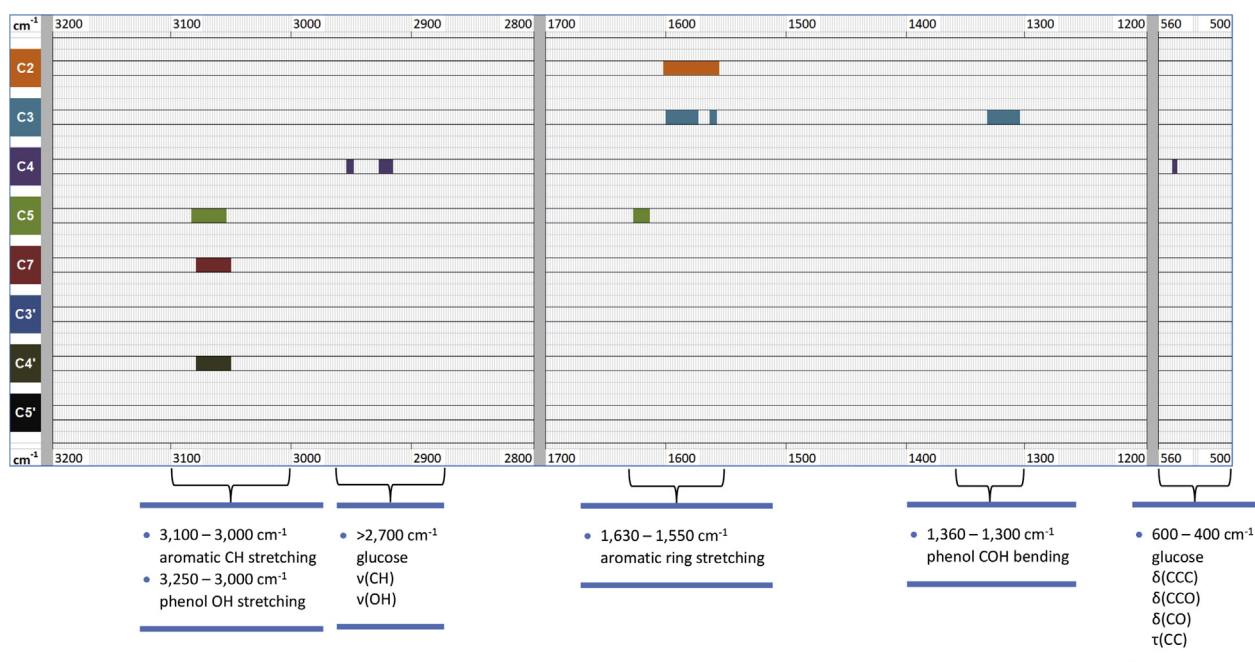


Fig. 6. A presentation of Structure-Spectrum Relationship of the bioflavonoid standards (for the individual structure indices see Table 1).

structure indices were both consistent with the conventional classification. Statistical correlation between the structure indices and Raman spectra was established using a canonical correlation method. The result was highly selective and could be explained using a long-range coupling theory. By this we have developed a structure-spectrum relationship for bioflavonoids which can be summarize as follows: (1) Hybridization status at C2 and C3 had a determinant predicting effect on the Raman spectra. *Vice versa*, by examining the aromatic ring stretching and phenolic COH bending alone, attribution to a subclass of C2–C3 hybridizations can often be made. (2) C4 is only important when it is  $sp^2$  hybridized and where conformational difference exists to affect its participation in resonance with neighboring groups. (3) Bioflavonoids different only in the substitution at C3' and C5' may not be differentiated by Raman spectroscopy if no other rules apply.

### Compliance with ethics requirements

This article does not contain any studies with human or animal subjects.

### Funding sources

Financial support from the Ministry of Science and Technology, Taiwan, under the project numbers MOST 107-2113-M-415-002 and MOST 108-2113-M-415-006 was highly appreciated.

### Declaration of Competing Interest

The authors declare that they have no conflicts of interest.

### Appendix A. Supplementary data

Supplementary data to this article can be found online at <https://doi.org/10.38212/2224-6614.1058>.

### References

- [1] Wang C-H, Chen M-J, Tseng C-Y, Wei K-M, Chen W. Structural characterization of functional compositions isolated from *Dioscorea purpurea* (cultivar of ming-chien) by Raman spectroscopy. *Nat Prod Chem Res* 2013;1:1–8.
- [2] Batsoulis AN, Siatas NG, Kimbaris AC, Alissandrakis EK, Pappas CS, Tarantilis PA, et al. FT-Raman spectroscopic simultaneous determination of fructose and glucose in honey. *J Agric Food Chem* 2005;53:207–10.
- [3] Agarwal UP, Reiner RR, Ralph SA. Estimation of cellulose crystallinity of lignocelluloses using near-IR FT-Raman spectroscopy and comparison of the Raman and segal-WAXS methods. *J Agric Food Chem* 2013;61:103–13.
- [4] Li Y-S, Church JS. Raman spectroscopy in the analysis of food and pharmaceutical nanomaterials. *J Food Drug Anal* 2014;22:29–48.
- [5] Persic M, Mikulic-Petkovsek M, Halbwirth H, Solar A, Veberic R, Slatnar A. Red walnut: characterization of the phenolic profiles, activities and gene expression of selected enzymes related to the phenylpropanoid pathway in pellicle during walnut development. *J Agric Food Chem* 2018;66:2742–8.
- [6] Yazgan Karacaglar NN, Bulat T, Boyaci IH, Topcu A. Raman spectroscopy coupled with chemometric methods for the discrimination of foreign fats and oils in cream and yogurt. *J Food Drug Anal* 2019;27:101–10.

- [7] Liao Y-H, Wang C-H, Tseng C-Y, Chen H-L, Lin L-L, Chen W. Compositional and conformational analysis of yam proteins by near infrared fourier transform Raman spectroscopy. *J Agric Food Chem* 2004;52:8190–6.
- [8] Hsu B-L, Weng Y-M, Liao Y-H, Chen W. Structural investigation of edible zein films/coatings and directly determining their thickness by FT-Raman spectroscopy. *J Agric Food Chem* 2005;53:5089–95.
- [9] Harnly JM, Lin L, Bhagwat S. Profiling methods for the determination of phenolic compounds in foods and dietary supplements. *Anal Bioanal Chem* 2007;389:47–61.
- [10] McGoverin CM, Weeranantanaphan J, Downey G, Manley M. Review: the application of near infrared spectroscopy to the measurement of bioactive compounds in food commodities. *J Near Infrared Spectrosc* 2010;18: 87–111.
- [11] Pompeu DR, Larondelle Y, Rogez H, Abbas O, Pierna JAF, Baeten V. Characterization and discrimination of phenolic compounds using Fourier transform Raman spectroscopy and chemometric tools. *BASE* 2018;22:13–28.
- [12] Gamsjaeger S, Baranska M, Schulz H, Heiselmayer P, Musso M. Discrimination of carotenoid and flavonoid content in petals of pansy cultivars (*Viola x wittrockiana*) by FT-Raman spectroscopy. *J Raman Spectrosc* 2011;42:1240–7.
- [13] Feng X, Zhang Q, Zhu Z. Rapid classification of citrus fruits based on Raman spectroscopy and pattern recognition techniques. *Food Sci Technol Res* 2013;19:1077–84.
- [14] Falk R, Well AD. Many faces of the correlation coefficient. *J Stat Educ* 1997;5:1–14.
- [15] Lee JK. Statistical bioinformatics: for biomedical and life science researchers. John Wiley & Sons; 2011.
- [16] Zaffino C, Russo B, Bruni S. Surface-enhanced Raman scattering (SERS) study of anthocyanidins. *Spectrochim Acta Mol Biomol Spectrosc* 2015;149:41–7.
- [17] Lin-Vien D, Colthup NB, Fateley WG, Grasselli JG. The handbook of infrared and Raman characteristic frequencies of organic molecules. Elsevier; 1991.
- [18] Söderholm S, Roos YH, Meinander N, Hotokka M. Raman spectra of fructose and glucose in the amorphous and crystalline states. *J Raman Spectrosc* 1999;30:1009–18.