International Food Research Journal 24(Suppl): S410-S415 (December 2017)
Journal homepage: *http://www.ifrj.upm.edu.my*

Algorithm for rapid identification of flavonoids classes

[1]Noh, C.H.C., [1*]Azmin, N.F.M., [1]Amid, A., [2]Asnawi, A.L.

[1]Department of Biotechnology Engineering, International Islamic University Malaysia 53100 Kuala Lumpur, Malaysia
[2]Department of Electrical and Computer Engineering, International Islamic University Malaysia 53100 Kuala Lumpur, Malaysia

Article history

Received: 17 June 2017
Received in revised form: 16 August 2017
Accepted: 18 September 2017

Keywords

Flavonoids
Fourier transform infrared spectroscopy
Principal component analysis
Fingerprint algorithm

Abstract

Bioactive compounds are one of the natural products used especially for medicinal, pharmaceutical and food application. Increasing research performed on the extraction, isolation and identification of bioactive compounds, however non to date has explored on the identification of flavonoids classes. Therefore, this study was focused on the development of algorithm for rapid identification of flavonoids classes which are flavanone, flavone and flavonol and also their derivatives. Fourier Transform Infrared (FTIR) spectroscopy coupled with multivariate statistical data analysis, which is Principal Component Analysis (PCA) was utilized. The results exhibited that few significant wavenumber range provides the identification and characterization of the flavonoids classes based on PCA algorithm. The study concluded that FTIR coupled with PCA analysis can be used as a molecular fingerprint for rapid identification of flavonoids.

Introduction

Flavonoid is one of the bioactive compounds that are widely distributed in fruits, plants, vegetables and microorganisms (Ignat *et al*., 2011; Park *et al*., 2012; Karpagasundari and Kulothungan 2014). For human consumption, flavonoid is important as they are shown to have high antioxidant activity, high radical scavenging, anti-inflammatory, anti-cancer (Kumar *et al*., 2014) and antiallergic.

Flavonoids have C6-C3-C6 general structure backbone which consists of two phenolic structures; ring A and ring B attached to heterocyclic ring C as in Figure 1. There are more than 7000 structures of flavonoids from different subgroups have been reported (Andersen and Markham, 2006). All flavonoids share a basic C6-C3-C6 phenyl-benzopyran backbone; however differ in their substituents (type, number and position) and in their insaturation (Pinheiro *et al*., 2012). Individual differences within each group are due to variation in number and arrangement of the hydroxyl groups and their extent of alkylation and/or glycosylation. Flavonoids aglycone can be divided into few subgroups which are flavanone, flavonol, flavone, flavanol, anthocyanidin and isoflavonoid (Knezevic *et al.*, 2012). For flavanone, flavonol and flavone, their variation within the groups is depending upon the presence of OH group at C-3, a saturated single bond between C-2 and C-3, and their

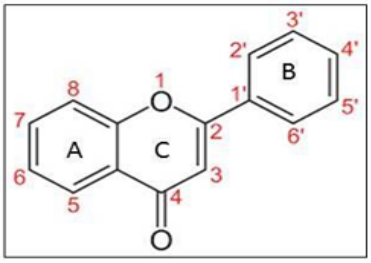

Figure 1. Backbone structure of flavonoids

conjugation and saturation behavior (Praisan *et al*., 2004; Noh *et al*., 2017).

Besides the extraction and isolation of the bioactive compound, as the multiple bioactive compound presences in compound, the characterization and identification of the bioactive compound become of interest due to complicated isolation procedure (Sasidharan *et al*., 2011).

Several analytical techniques have been developed for the identification of bioactive compound such as the application of High Performance Liquid Chromatography (HPLC) (Durum *et al.*, 2016), HPLC coupled with photodiode array and mass spectrometry detectors (Daqiu *et al*., 2015), nuclear magnetic resonance (NMR) spectroscopy and mass spectrometry (MS) evidences (Lingrong *et al*., 2014), Thin Layer Chromatography (TLC) plate (Samariya *et al*., 2013) and many more. However, some disadvantages of these method are less sensitive,

*Corresponding author.
Email: *norfadhillah@iium.edu.my*

expensive and the need for chemist to analyze has lead to the application of FTIR which is accurate, sensitive and inexpensive (Gok *et al*., 2015).

FTIR spectrums however are difficult to interpret due to high dimensionality of the spectrum, unless for the expertise since the library spectrum are limited, and the software used to interpret the results somehow expensive (Smith, 1999). Henceforth, appropriate application of statistical analysis or chemometrics has provides accurate and rapid characterization of sample that contains multidimensional information (Royston *et al*., 2015).

Advantages offered by PCA that can analyze the high dimensional data have leads to the application of PCA for the analysis of spectrum data. PCA were shown to provide the pattern of the samples that indicates their similarities and differences based on the correlation factors (Kallithraka *et al.*, 2001).

Most often, studies performed on the identification of compound using few combination of chemometric techniques such as hierarchical cluster analysis (Seher *et al*., 2015), PCA, partial least square (PLS) and factor discriminant analysis (FDA) (Gouvinhas *et al.*, 2015). Taking PCA as an example, single applications of PCA into sample characterization sometimes were not sufficient to provide the pattern and validate the results. This is mainly cause by the improper choice of observation and less exploitation on the PCA interpretation. However, Bro and Smilde (2014) suggested that, critical understanding on the principles behind PCA would provide with significant information. Besides, based on Morrison (1990), the mathematical modeling coupled with the observable reality provides an algorithm for every statistical analysis.

Therefore, the objective of the studies is to develop an algorithm for flavonoids classes identification based on FTIR spectrum coupled with principal component analysis.

Materials and methods

The objective of the studies is to develop an algorithm for flavonoids classes identification. The materials used in the studies were divided into two which are the flavonoids subgroups samples (flavone, flavanone and flavonol) and the validation samples (quercetin, kaempferol, luteolin, apigenin, naringenin and ascorbic acid).

FTIR measurements

1. The dried powder of all samples was sent to Centre for Research and Instrumentation, Universiti Kebangsaan Malaysia and International Institute for Halal and Research Training (INHART) IIUM for FTIR analysis. The wavenumber range set for the FTIR is 4000-650 cm^{-1}. The FTIR was performed using the Attenuated Total Reflectance (ATR) where the solid samples were directly applied to the plate for measurements.

2. The subgroups of flavonoids including flavone, flavanone and flavonol. Meanwhile, the validation samples were selected according to the flavonoids subgroups derivatives. Luteolin and apigenin are derivatives of flavone, kaempferol and quercetin (flavonol), naringenin (flavanone) and ascorbic acid, the non flavonoids.

3. Utilizing the built in algorithm provided from software package of FTIR, the spectrum was pretreated using baseline correction to produce horizontal baseline shift and smoothing to reduce the noise.

4. The pretreated spectrums which comprise of numerical absorbance values and wavenumbers obtained were recorded in excel forms and treated as raw data that were further used for statistical data analysis on MATLAB R2013a 64 bit platform.

Development of algorithm

The development of algorithm provides key stages for flavonoids classification based on FTIR spectrum as summarized in Figure 2.

i)Pretreatment of data

1. Most often prior to performing the PCA, the data requires normalization and standardization (Bro and Smilde, 2014).

2. Normalization is usually used for the data that have different units to scales all the variable in range of [0,1] meanwhile; the standardization is performed to rescaled the variable to have a mean zero and a standard deviation of one.

3. In this project, since the data are in the same unit (absorbance), only the standardization was performed. Considering x, the two dimensional matrix representation of (IxJ) where I rows, the observation and J columns, the variable involved. By default, PCA algorithm used in MATLAB standardized the raw data of x.

ii) Matrix representation

1. Prior to performing PCA, the three dimensional data matrix, X which comprises of spectrums was built (Azmin, 2013).

2. Figure 3 shows the configuration of the three dimensional array X in which the data set is Xk, an IxJ matrix, where K is the region number, with I rows (i = 1, … , I) is the sample types which

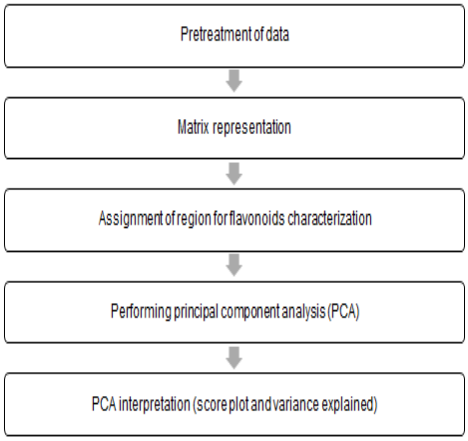

Figure 2. Development of algorithm utilizing PCA for flavonoids classification

represent observation and J columns (j = 1, … , J) is the wavenumber range which represent variables .

iii) Assignment of regions for flavonoids characterization

1. Few significant regions have been assigned as in Table 1 for the identification of flavonoids based on their structural differences.
2. The assigned region will provide the variables for the PCA algorithm.

iv) Statistical data analysis (Principal Component Analysis)

1. PCA algorithm was performed on the data matrix, X based on the assigned region.
2. Principal Component Analysis statistical tool in MATLAB software was used to correlate the wavenumber range (regions) with the absorbance values of each data points of the sample (peak intensity), which provides the variables and sample used for principal components derivation hence provides the sample characterization.
3. By default, algorithm used in MATLAB R2013a to calculate the PCA is based on Singular Value Decomposition (SVD).
4. The information provided by the PCA algorithms were further utilized for interpretation and visualization.

PCA interpretation and validation of model

i) Score and variance explained

1. The score plot were generated to visualize the differences and similarities of the flavonoids of interest at selected region for their characterization.
2. The score stored in the MATLAB workspace contains the actual principal components that used to visualize the pattern of the data.
3. The combinations of principal component were chose upon since it will affect the pattern of the data.

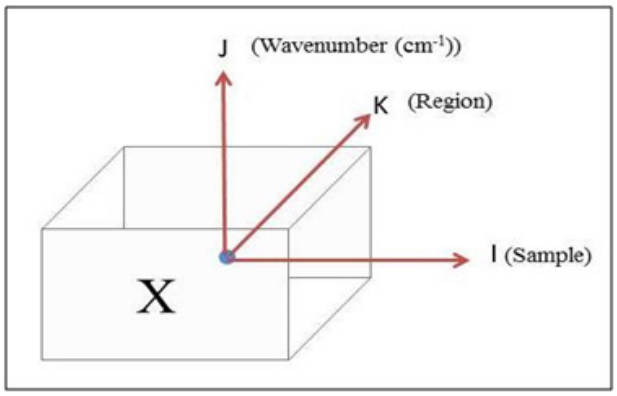

Figure 3. Data matrix representation

Results

The FTIR data obtained from the flavonoids and validation sample has shown to produce unique spectrum since they consist of different structure. Regardless of the structure of the rest of the molecule, the functional groups tend to absorb infrared radiation in the same wavenumber range (Smith, 1999). However, the overall structure will cause the wavenumber shift under the wavenumber range which is of interest of the study.

Figure 4 shows the PCA result of flavonoids on fingerprint region. In practice, the first 4 PCs were considered to represent the pattern of spectrum which is significant to their classification (Gok *et al.*, 2015). The different combinations of PCs were performed to obtain the score plot that provides clear discrimination of the sample. Based on Figure 4, clear discrimination was observed between the flavonoids classes and their derivatives by the combination of the first and fourth principal component. This describes 62.9% of the total variance for PC1 and 6.5% for PC4. The quercetin and kaempferol were positioned together with the flavonol; the luteolin and apigenin with the flavone and naringenin with the flavanone. Meanwhile, the ascorbic acid which is not a flavonoid was not in any groups of flavonoids. Focusing on the flavonoids structure of the heterocyclic ring C as in Figure 1, the flavonoids subgroups somehow exhibit similarities under the PCA with their derivatives.

Discussion

In PCA analysis, the correlation factors provide the pattern of the samples. The sample that is close to each other indicates their similarities meanwhile far

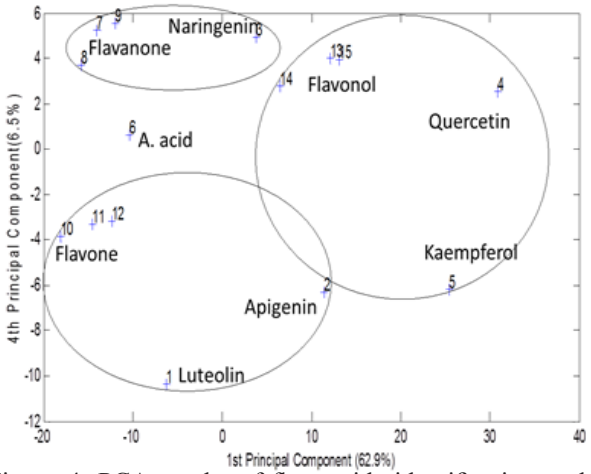

Figure 4. PCA results of flavonoids identification under the fingerprint region (1300 – 900 cm^{-1}).

to each other indicate their differences. This pattern of analysis provides clear discrimination between samples as a tool of identification.

There are various method used to illustrate the high dimensional data into two or three dimensional that best represent the data (Jollife, 2004). In PCA analysis, understanding on the interpretation of the score plot will provides significant information. For the PCA, it is important to identify how much principal component to retain and it will depend upon the purpose of analysis. Most often, the score plot was illustrated based on first principal component versus second principle component which carries the largest total variance (usually more or less 80% explained variance) that regarded as "good" presentation of data (Jollife, 2004). However, the principal component that carries less variance sometimes contains the important information that must be considered as shown in Figure 4.

Six significant regions have been assigned in Table 1 for their identification based on its structural difference. The first five regions were referred as group frequency regions; meanwhile the sixth region indicates the fingerprint region. Figure 4 shows the results of one of the PCA analysis of flavonoids spectrum under the fingerprint region.

The identification of flavone, flavanone and flavonol are focusing on the heterocyclic ring C that represents the major structural difference between the flavonoids. Focusing on the heterocyclic ring C, the flavonoids derivatives are expected to sit together with the flavonoids subgroups hence provide the validation of algorithm. The score plot were illustrated by few attempts of different combinations of principle component since sometimes, the principle component that carries less variance contains important information that provides distinct feature of flavonoids sample.

Table 1. Assigned region for flavonoids subgroups characterization

Region	Classification	Wavenumber range (cm^{-1})	Spectral characteristics	References
1	Quick diagnose for the presence or absence of the intense group	3500-3200	O-H	Smith, 1999
2		3200-2800	C-H	Smith, 1999
3		1800-1600	C=O	Smith, 1999
4		1900-1500	C=C	Silverstein and Webster, 1996
5		1300-800	C-O	Silverstein and Webster, 1996
6		1300-900	Fingerprint region	Silverstein and Webster, 1996

Based on Table 1, 3500-3200 cm^{-1} (region 1) has been identified for the presence of hydroxyl group in flavonoids, 3200-2800 cm^{-1} (region 2) for the presence of C-H groups, 1800-1600 cm^{-1} (region 3) for the C=O group, 1900-1500 cm^{-1} (region 4) for the C=C groups, 1300-800 cm^{-1} (region 5) for the C-O group and 1300-900 cm^{-1} (region 6) for the fingerprint region.

In PCA score configurations, flavonoids classes were observed to posses different characteristic at different regions. Hence, this indicates the variation of molecular structure was based on the region classification. For flavonol, clear discrimination had shown at region 1, 2,3 and 4. The presence of hydroxyl group at carbon 3 of heterocyclic ring C provides the distinct characteristics of structure arrangement as compared to flavanone and flavone. For flavone, clear discrimination had shown at region 3 and 4 as compared to flavanone. Meanwhile, for flavanone, clear discrimination shown at region 5 as compared to flavonol and flavone. This is due to the molecule attraction behavior of the single bond at C2 and C3 of flavanone as compared to flavone and flavonol.

Region 1 until 5 however, do not encompass the information to discriminate the flavonoids classes and its derivatives, hence utilization of the fingerprint region. The fingerprint region, between 1300-900 cm^{-1} (Figure 4) is utilized and has successfully discriminated the flavonoids classes and its derivatives. It is valuable due to its attributes of which it encompasses all combination bands (Silverstein and Webster, 1997).

The main contribution in this study is, the selected region based on PCA will provides significant identification of flavonoids based on its structure arrangement. Henceforth, the isolation procedure for example, will be less time consuming

and identification of flavonoids presence in fruits and plants can be optimized.

Conclusion

In this study, the identification of flavonoids classes was performed using PCA and FTIR spectroscopy. Clear discrimination of the flavonoids classes and their derivatives has shown in wavenumber range of 3500-3200 cm^{-1}, 3200-2800 cm^{-1}, 1800-1600 cm^{-1} and 1900-1500 cm^{-1} for flavonol, 1800-1600 cm^{-1} and 1900-1500 cm^{-1} for flavone and 1300-800 cm^{-1} for flavanone. At 1300-900 cm^{-1}, the fingerprint region, all flavonoids and their derivatives were clearly discriminate. The combination of principal component, considering the small variance provides the important information that should not be ignore. The similarities and difference in FTIR spectrum region based on PCA algorithm has successfully extract the distinct feature of the flavanone, flavone and flavonol in FTIR spectrum. It is recommended that the proposed methodology performed on the other bioactive compound for rapid and economical identification of bioactive compound.

Conflict of Interest

The authors declare no conflict of interest.

Acknowledgments

Authors acknowledge to International Islamic University of Malaysia "IIUM" and Ministry of Higher Education "MOHE" for the research grant "RAGS14-050-0113".

References

Andersen, O.M. and Markham, K.R. 2006. Flavonoids-Chemistry, biochemistry and application. Boca Raton: CRC Press.

Azmin, N.F.M. 2013. Fingerprinting of complex bioprocess data. New Castle, United Kingdom: Newcastle University, PhD thesis.

Bro, R. and Smilde, A.K. 2014. Tutorial review: Principal component analysis. Royal Society of Chemistry 6: 2812–2831.

Daqiu, Z., Wenhui, T., Zhaojun, H. and Jun, T. 2015. Identification of flavonoids and expression of flavonoid biosynthetic genes in two coloured tree peony flowers. Biochemical and Biophysical Research Communications: 450-456.

Durum, S., Allam, R.N.A.M. and Shazely, A.S.E. 2016. Identification of phenolic acids and flavonoids compounds from some varieties *Triticum Aestivum* and *Triticum*. Middle East Journal of Applied Sciences 6(3): 521-530.

Gok, S., Severcan, M., Goormaghtigh, E., Kandemir, I. and Severcan, F. 2015. Differentiation of Anatolian honey samples from different botanical origins by ATR-FTIR spectroscopy using multivariate analysis. Food Chemistry 1(170): 234–240.

Gouvinhas, I., Almeido, J.M.M.M.D., Carvalho, T. and Barros, A.I.R.N.A. 2015. Discrimination and characterisation of extra virgin olive oils from three cultivars in different maturation stages using Fourier transform infrared spectroscopy in tandem with chemometrics. Food Chemistry 174: 226-232.

Ignat, I., Volf, I. and Popa, V. 2011. A critical review of methods for characterisation of polyphenolic compounds in fruits and vegetables. Food Chemistry 126(4): 1821-1835.

Jolliffe, I. T. 2002. Principal component analysis. New York: Springer.

Kallithraka, S., Arvanitoyannis, I. S., Kefalas, P. and El-zajouli, A. 2001. Instrumental and sensory analysis of Greek wine; implementation of principal component analysis (PCA) for classification according to geographical origin. Food Chemistry 73: 501-514.

Karpagasundari, C. and Kulothungan, S. 2014. Analysis of Bioactive Compounds in Physalis minima leaves using GC MS, HPLC, UV-VIS and FTIR techniques. Journal of Pharmacognosy and Phytochemistry 3(4): 196-201.

Knezevic S.V., Blazekovic, B., Stefan, M.B. and Babac M. 2012. Plant polyphenols as antioxidants influencing the human health. In Rao, D.V. (Ed.), Phytochemicals as Nutraceuticals- Global Approach to Their Role in Nutrition and Health, p. 155-180. Europe: InTech.

Kumar, S.K., Suresh, M., Kumar, S.A. and Kalaiselvi, P. 2014. Bioactive compounds, radical scavenging, antioxidant properties and FTIR spectroscopy study of Morinda Citifolia fruit extracts. International Journal of Current Microbiology and Applied Science 3(2): 28-42.

Lingrong, W., Dan, W., Yueming, J.K.P.N., Sen, L., Guoxiang, J., Jirui, H., Mouming, Z., Wei, L. and Bao, Y. 2014. Identification of flavonoids in litchi (*Litchi chinensis* Sonn.) leaf and evaluation of anticancer activities. Journal of Functional Food 555-563.

Morrison, D.F. 1990. Multivariate statistical methods. New York: McGraw-Hill.

Noh, C.H.C., Azmin, N.F.M. and Amid, A. 2017. Principal Component Analysis Application on Flavonoids Characterization. Advances in Science, Technology and Engineering Systems Journal 2(3): 435-440.

Park, Y.S., Cvikrová, M., Martincová, O., Ham, K.S., Kang, S.G., Park, Y.K., Namiesnik, J., Rombolà, A.D., Jastrzebski, Z. and Gorinstein, S. 2015. *In vitro* antioxidative and binding properties of phenolics in traditional, citrus and exotic fruits. Food Research International 74: 37–47.

Pinheiro, P.F. and Justino, G. C. 2012. Structural analysis of flavonoids and related compound: A review of spectroscopic applications. In Rao, D.V. (Ed.), Phytochemicals-A Global Perspective of Their Role

in Nutritrion and Health, p. 33-56. Europe: InTech.
Praisan, J.K., Wang, C.C. and Barnes, S. 2004. Mass spectrometric methods for the determination of flavonoids in biological samples. Free Radical Biology and Medicine 37(9): 1324-1350.
Royston, G., Emma, V.Y., James, K.H. and Ian, M.S. 2003. Chemometric discrimination of unfractionated plant extracts analyzed by electrospray mass spectrometry. Phytochemistry 859–863.
Samariya, K. and Sarin, R. 2013. Identification of flavonoids from *Cyperus Rotundus* Linn. *in vivo* and *in vitro*. Journal of Drug Delivery and Therapeutics 3(2): 109-113.
Sasidharan, S., Chen, Y., Saravanan, D., Sundram, K.M. and Yoga, L. 2011. Extraction, isolation and characterization of bioactive compounds from plants' extracts. African Journal of Traditional, Complementary and Alternative Medicine 8(1): 1-10.
Seher, G., Mete, S., Erik, G., Irfan, K. and Feride, S. 2015. Differentiation of Anatolian honey samples from different botanical origins by ATR-FTIR spectroscopy using multivariate analysis. Food Chemistry 170(2015): 234–240.
Silvertein, R.M. and Webster, F.X. 1997. Spectrometric identification of organic compounds. United States: John Wiley and Sons, Inc.
Smith, B. C. 1999. Infrared spectral interpretation: A systematic approach. United States: CRC Press.