Authentication and Discrimination of Green Tea Samples Using UV-Visible, FTIR and HPLC Techniques Coupled with Chemometrics Analysis

Maram M. Aboulwafa [a†], Fadia S. Youssef[a†], Haidy A. Gad[a], Satyajit D. Sarker[b], Lutfun Nahar[b] Mohamed M. Al-Azizi[a], Mohamed L. Ashour[a*]

[a] Faculty of Pharmacy, Department of Pharmacognosy, Ain Shams University, 11566-Abbassia, Cairo, Egypt

[b] Medicinal Chemistry and Natural Products Research Group, School of Pharmacy and Biomolecular Sciences, Liverpool John Moores University, James Parsons Building, Byrom Street, Liverpool L3 3AF, United Kingdom

†These authors have contributed equally to this work

***Corresponding author**

Prof. Mohamed L. Ashour

Email: mohamed_ashour@pharm.asu.edu.eg; Tel.: +2010 68 222 354, Fax: +202 2 405 1107,

Highlights

Authentication and discrimination of green tea samples using UV-Visible, FTIR and HPLC techniques coupled with chemometrics analysis

- Quality of green tea from the biggest tea producing Asian countries was assessed.
- Simple UV-Visible, FTIR and HPLC and chemometrics were employed.
- PCA based on UV data successfully divided samples into informative clusters.
- SIMCA and PLS-DA models showed a good separation between South and the East tea
- Antioxidant activity of all samples were determined using DPPH· assay

ABSTRACT

Green tea is a popular beverage consumed worldwide. Its quality should be controlled adequately as the quality is influenced by several factors in addition to adulterations. This study aimed to develop a simple method for assessing the quality of green tea samples obtained from the South and the East Asian regions. The UV-Visible, FTIR and HPLC data from 38 samples were subjected to multivariate analyses using the unsupervised recognition techniques comprising Hierarchical Cluster Analysis (HCA) and Principal Component Analysis (PCA). The model for their authentication was constructed and validated by applying the supervised recognition techniques as Soft Independent Modeling of Class Analogy (SIMCA) and Partial Least Square Discriminant Analysis (PLS-DA). The percentages of caffeine in the identified samples were determined using a validated HPLC assay in addition to *in vitro* determination of their antioxidant activity using DPPH radical-scavenging capacity assay. HCA and PCA based on UV data successfully distributed the tested samples into informative clusters. However, that obtained from visible data could only differentiate samples with respect to their powdered condition. On the contrary, PCA from FTIR and HPLC data could hardly discriminate any of the samples. The models constructed using SIMCA and PLS-DA showed a good class separation between the South and the East Asian samples. The percentages of caffeine in the identified samples and the IC_{50} in DPPH assay are greatly diverse among all the tested samples. Thus, UV spectroscopy and chemometrics have provided a simple and quick tool for the quality control of commercial green tea samples.

KEYWORDS: *Camellia sinensis*; FTIR; HPLC; Multivariate data analysis; UV-Visible spectroscopy

1. Introduction

Camellia sinensis (L.) Kuntze of the family Theaceae is an evergreen shrub that produces the plant materials used in preparation of tea either from the leaves or the leaf buds [1]. Although the tea plant is native to Southeast Asia and cultivated in nearly about thirty countries, tea is considered as one of the most popular and safer beverages consumed worldwide [2]. Based upon FAO-Intergovernmental Group on Tea-22nd session held in May 2016, East Asia constitutes the largest tea producing area represented by China, which accounts for 38% of world tea production, followed by South Asian countries, India and Sri Lanka. Nowadays, drinking tea, in general, and green tea, in particular, is recognized to be crucial for its potential health benefits relying on a plethora of biological activities [3].

The best quality of green tea is attained when it is processed with fresh leaves collected in an early stage attributing to the variations in its sensory properties together with its active constituents exemplified by catechins and caffeine that, consequently, affects its price [4]. Moreover, many green tea consumers are greatly influenced by its geographical origin that is to a great extent affects its quality as well as its financial cost [5]. Thus, due to the massive elevation in green tea consumption all over the globe and owing to the ease of its adulteration, its quality should be tightly controlled and is felt obligatory worldwide [6].

Chromatography fingerprinting coupled with chemometrics comprising PCA and HCA has recently been considered as one of the most frequent tool in the evaluation of chemical profiles of various botanicals [7]. In addition, HPLC-DAD-MS coupled with chemometrics was successfully adopted to differentiate between seven tea types experiencing different processing conditions [8]. UV-Visible and infrared spectroscopy coupled with chemometrics was previously reported to be an effective tool in the differentiation of food systems. It represented a successful discriminatory tool in the authentication of monofloral Yemeni Sidr honey [9] as well as different thyme and *Curcuma* samples [10, 11]. Despite some published studies on techniques to test the purity, quality as well as to authenticate and discriminate the green tea sourced from various geographical origins by different suppliers, nothing could be found in the literature regarding the use of a cost-effective, simple and robust technique in the assessment of the green tea quality as the use of UV-Vis and FTIR spectroscopy.

Hence, the forgoing study aimed to assess the possibility of coupling of various simple spectroscopic techniques, mainly, UV-Vis and FTIR spectroscopy with multivariate chemometrics analysis in order to establish a model for the authentication and discrimination of green tea supplied from different Asian regions and hopefully to prevent its adulteration worldwide. Moreover, the percentages of caffeine in the identified samples were determined using a validated HPLC assay in addition to *in vitro* determination of their antioxidant activity using DPPH radical-scavenging capacity assay.

2. Materials and Methods

2.1. Sample collection

A total of 38 dried green tea samples were purchased comprising 26 for model construction (11 samples from China, 11 from India and 4 from Sri Lanka). Additionally, 12 commercial samples from Chinese and Egyptian markets were used to validate the model and were assumed to be green tea. All samples were purchased between January and April 2015 and voucher specimens of the identified samples were kept at the Pharmacognosy Department, Faculty of Pharmacy, Ain Shams University. Names, codes as well as the geographical origin of the identified and the marketed samples are displayed in Supplementary Table 1S.

2.2. Sample preparation

Sample Preparation

From each sample, 2 g of coarsely powdered samples were packed in a tea bag material, labeled and macerated in 50 mL of HPLC grade methanol for 15 min to prepare the stock solution. For UV-Vis spectroscopic analysis, 1 mL was taken from the stock solution and diluted with methanol to 25 mL. Regarding HPLC, each sample was further diluted with HPLC grade methanol in a ratio of 1:1. However, for FTIR spectroscopy, a portion of each of the dried powdered sample was taken separately and ground using a mortar and a pestle, then the ground powder was mixed well with potassium bromide in a ratio of 1:30, respectively to make an intact transparent disc required for exposing it to IR radiations.

2.3. Ultraviolet and visible spectroscopy

Each of the prepared samples was analyzed by UV-Vis spectroscopy separately using a V-630 UV−Visible spectrophotometer (JASCO, Shimadzu, Japan) supplied with a quartz cell having an optical path of 1 cm and spectral resolution of 1 nm in the range 200−400 nm and 400−800 nm for UV and visible spectroscopy, respectively. Each sample was measured in triplicates.

2.4. Fourier Transform Infrared Spectroscopy

FTIR spectral analysis of the samples was carried out using Nicolet™ iS™ 10 FT-IR Spectrometer (Thermo Scientific™, Merck, Germany) covering IR radiation spectrum from 4000 to 400 cm_1 and measurements were performed in triplicates.

2.5. High Performance Liquid Chromatography (HPLC)

After sample dilution, 20 μL were auto-injected into Agilent 1200 series HPLC-DAD (Agilent Technologies, Santa Clara, CA, USA). All chromatographic separations were carried on a C18 reversed-phase column (150 mm, 4.6 mm, 5 μm) equipped with a quaternary pump. Mobile phases consisted of water (Solvent A) and methanol (Solvent B). A gradient elution system was adopted as follows: increasing concentration of solvent B by 10% every 3 min using a flow rate equals to 0.2 mL/min till reaching 40% followed by isocratic elution at 40% B for 15 min then returning back to gradient elution at 90% B till reaching 45 min as total run time. Detection was carried out at 278 nm and each sample was measured 3 times. Caffeine was quantitatively determined in each sample using HPLC *via* constructing of a calibration curve using standard caffeine solution and calculating the peak area of each sample.

2.5.1. Validation of HPLC method

The adopted HPLC method was validated for accuracy, linearity, precision and recovery following the International Conference on Harmonization (ICH) guidelines.

2.5.1.1. Calibration curve

Standard solution of caffeine was prepared and diluted to suitable concentrations for constructing the calibration curve. Various concentrations of caffeine in the range of 10-100 μg/mL were used in which each is measured three times. The calibration curve was drawn by plotting the peak

areas as a function of the used caffeine concentration.

2.5.1.2. Limit of detection and quantification

The limits of detection (LOD) and quantification (LOQ) were determined under the chromatographic conditions. They were determined *via* calculating the signal-to-noise ratio of the standard compound after the injection of a series of solution until the S/N ratio 3 for LOD and 10 for LOQ [10].

2.5.1.3. **Precision and accuracy**

Measurement of intra- and inter-day variability was achieved to detect repeatability of the method adopting the method previously described by Gad et.al [10]. The relative standard derivation (R.S.D.) was calculated by the formula:

RSD (%) = (S.D./mean)×100%

S.D: is the standard deviation

The recoveries were determined by the formula:

Recovery (%) = (amount found/original amount ×100%).

2.6. In vitro antioxidant evaluation using diphenyl picryl hydrazyl radicle scavenging capacity assay (DPPH·)

Equal volumes of sample solutions containing 0.001–10 mg/mL of the tested samples and 0.2 mM methanol solution of DPPH• were mixed together. After incubation in the dark for 30 min at room temperature, the absorbance was then measured against a blank at λ_{max} = 520 nm using Awareness Technology ChroMate® Microplate Reader (Florida, USA) and compared to DPPH• control after background subtraction. The percent inhibition was calculated from three different experiments using the following equation [12].

Inhibition (%) = [Ac – As/Ac] x 100,

Where, Ac: absorbance of control; As: absorbance of sample.

2.7. Chemometric analysis

The UV-Visible and IR spectral data as well as HPLC data were subjected separately to various chemometrics techniques namely HCA and PCA as unsupervised pattern recognition techniques.

HCA was performed using Hierarchical Clustering Explorer 3.5 (Human computer interaction laboratory, University of Maryland, College Park, MD, USA). However, PCA was performed employing Unscrambler® 9.7 (CAMO SA, Oslo, Norway). All spectral data matrices were subjected to mean centering preprocessing by applying a default option in the Unscrambler®9.7 software before analysis. However, spectral data obtained from FTIR was additionally subjected to standard normal variate "SNV" method that removes scatter effects from spectra by centering and scaling each individual spectrum. SIMCA as supervised recognition method was also applied for better sorting and discrimination among samples. However, for partial least square discriminant analysis (PLS-DA), each sample in the calibration set was given an arbitrary number describing if the sample belongs to a particular country or not (1 = Eastern Asia and 2 = Southern Asia samples). The training and validation sets were as that used in the SIMCA model and were chosen randomly. A green tea sample is considered from East Asia or South Asia if its value lies between 0.5 -1.5 and 1.5-2.5 respectively. Both SIMCA and PLS-DA were applied by Unscrambler® 9.7 software [9, 13, 14].

3. Results and discussion

The methanol extracts of the green tea samples showed UV absorption bands in the UV region between 200 and 400 nm. This partly relied upon the richness of green tea with polyphenolic compounds particularly catechins, flavonoids and phenolic acids. These phytoconstituents are characterized by the presence of various chromophores and conjugated systems that act as UVabsorbing systems. (-)-Epigallocatechin-3-gallate (EGCG), (-)-epigallocatechin (EGC), (-)-epicatechin-3-gallate (ECG) and (-)-epicatechin (EC) represent the four major catechins that predominate in green tea. EGCG and ECG showed absorbance at spectral range between 246 to 325 nm with λ max at 276 and 279.2 nm in methanol, respectively. Moreover, kaempferol, myricetin and quercetin are highly prevalent in the green tea with absorbance maximum at 367 nm. Additionally, phenolic acids exemplified by gallic, chlorogenic and caffeic acids that mainly exist in green tea showed absorbance maxima at 273 nm for the first and 330 nm for the latter [11, 15].

UV spectroscopic data region between 200 and 400 nm were subjected to PCA using cross validation method after mean centering of the data. PCA score plot (Fig. 1a) successfully

discriminated the tested samples into two informative clusters with PC1 and PC2 computed for 75% and 19% of the data variance, respectively. The first cluster comprised the majority of the East Asian samples (from China), while the second one contained the entire South Asian samples (from India & Sri Lanka). Although CD, CX, CM and CL are Chinese samples, they were clustered within the cluster of Sri Lanka, this might be attributed to the high quality of Dongting Biluochun, Xinyang Maojian, Huangshan Maofeng and Lushan Yunwu green teas that approaches that of the Sri Lankan tea. Ceylon green tea exhibited an extremely high quality owing to its great nutritional and biological value and it is more likely to be adulterated [16]. To challenge PCA model, different marketed samples were tested, the resulted PCA score plot (Fig. 1b) explained 94% of total variance in the data where PC1 and PC2 accounted for 73% and 21% respectively. The scatter plot showed that almost all the marketed samples were grouped within the South Asian cluster.

Regarding the visible region, PCA was also performed on the average replicates of reading absorbencies in the region between 400 and 800 nm. It was obvious from the score plot (Fig. 1S) that no significant segregation was observed among different samples although PC1 and PC2 accounting for 98 and 1% of the data variance, respectively. Moreover, hierarchical cluster analysis was applied to classify the samples based on the similarities and differences among their UV spectral data**.** Besides, average group linkage method using Pearson's correlation was applied to build hierarchical agglomerative dendrogram using UV spectral data at a similarity level of 0.5. The clustering dendrogram (Fig. 2a) revealed that all samples were divided into three main clusters; Clusters A & B include the majority of South Asian samples in addition to CD, CX, CM and CL whereas cluster C represents most of the East Asian samples and this further ascertains the results obtained from PCA. Upon addition of the marketed samples, two discriminative clusters are observed in which cluster A represents the majority of the identified East and South Asian samples; however, cluster B comprises all of the commercial samples (Fig. 2b).

The results obtained from the unsupervised pattern recognition techniques (HCA and PCA) clarified the differentiation of the samples into discriminate clusters in accordance with their geographical origin, which reflect the notable variation among the green tea samples

composition. As a result, a more accurate segregation is demanded to be achieved between the green tea samples from different Asian regions as well as the marketed samples. Consequently, SIMCA and PLS-DA (Partial Least Square Discriminant Analysis) as supervised pattern recognition techniques were adopted to further ascertain the results obtained from the unsupervised techniques.

Soft Independent Model of Class Analogy **(**SIMCA) technique effectively handles a number of classes and attributes each sample to its respective class [9]. The classification using SIMCA process comprises two main steps, which are the training and the validation ones. In the former the individual models of the data classes are performed, and in the latter, new samples (not used in the training stage) are classified within the established class models to evaluate the model's efficacy [9]. Two classes corresponding to the East Asian and the South Asian samples were developed using independent PCA. Cooman's plots were used for the validation of the SIMCA model using the different models of green tea constructed for the different regions of Asia where a sample belonging to a particular class should fall within the membership limit. The validation samples were all within the limit of each membership, indicating the superior predictability and sensitivity of the model constructed from the East and South Asian ones (Fig. 3a). The developed SIMCA model showed interclass distances of about 32, and all variables exhibited a discrimination power more than 2 and a modeling power nearly 0.85, which clarified the discriminative potential of the developed model to discriminate the spectral signals of the East Asian samples with respect to those from South Asia, with 95% confidence limits. SIMCA results confirmed that of PCA regarding the clustering of Chinese samples (CD, CX, CM and CL) with the South Asian samples. For the above reasons, the constructed SIMCA model was employed in the classification of marketed samples (Fig. 3b). All marketed samples were nearer to the South Asian model relative to the East Asian one.

PLS-DA was also carried out to enhance the separation between the green tea samples obtained from different localities. PLS-DA calibration model was developed using full cross validation. The score plot of the first and the second latent variables for the calibration set in (Fig. 2S) showed that a good class separation was attained between the East and the South Asian samples, where two distinct clusters were observed. The South Asian samples were clearly

clustered on the right quadrant, while those from East Asia appeared occupying the left side of the plot. On validation with test set samples, the PLS-DA model showed 100 % correct classification for all the South Asian samples as all the test set were assigned to its correct class as seen in Table 1. Authentication of the marketed green tea samples was additionally performed using the developed PLS-DA model. The results illustrated in Table 1 further supported that obtained from PCA as well as SIMCA models and revealed the resemblance of all the commercial samples to South Asian ones and thus considered as South Asian origin. Results of PLS-DA were in accordance with that obtained from SIMCA regarding the segregation of both East and South Asian samples and commercial ones.

Concerning FTIR spectroscopy**,** the average reading of three replicates of each sample versus 3800 variables representing the peaks intensity in the region between 4000 and 400 cm^{-1}, were recorded. These data were then subjected to PCA using standard normal variate "SNV" method. The resulting score plot (Fig. 3S) revealed that FTIR spectral data did not offer a strong discriminatory tool for the authentication and discrimination of green tea from the East and the South Asian countries as all the samples either in the same form or in different ones are scattered randomly in the plot although PC1 and PC2 accounting for 79 and 12% of the data variance, respectively.

For High Performance Liquid Chromatography (HPLC), the average area percentages of three replicates of each sample versus 64 variables representing the retention time of the predominating compounds in the Asian green tea samples were recorded. HPLC did not provide an effective tool for the differentiation of green tea samples obtained from the East and the South Asian regions as most of the samples are scattered in a diverse manner in the plot (Fig. 4S). However, HPLC chromatograms offered an excellent tool for the effective profiling of most of the chemical constituents prevailing in the green tea samples that undoubtedly discriminated them from other non-green tea samples (Fig. 5S).

Moreover, the percentages of caffeine in all the tested samples were determined by a validated HPLC method and are illustrated in Table 2. Caffeine calibration curve showed a good linearity over the examined range with $R^2 > 0.984$. In addition, the LOD and LOQ were found to

be 11.46 and 34.73 μg/mL, respectively. Meanwhile the overall recovery was found to be 95.59 ± 2.34% and the intra-day and inter-day variability (%) were 0.227 and 4.453, respectively. Moreover, the antioxidant activity of all the samples was determined *via* the assessment of the IC_{50} using DPPH radical-scavenging capacity assay. Results are illustrated in Table 3. This was carried out in an effort to correlate the antioxidant activity to the concentration of caffeine in the samples but unfortunately, there was no correlation that clearly existed between them. Undoubtedly, this could be due to the existence of many other antioxidant phytochemicals predominating in the green tea that synergistically explain the antioxidant behavior of the samples.

4. Conclusion

It could be concluded that authentication and discrimination of green tea samples sourced from the South and the East Asian regions could be effectively achieved using UV spectroscopic data in combination with chemometrics. This undoubtedly could offer a credible simple model for the quality control of herbal drugs based upon their predominant active constituents.

Conflict of interest

The authors declared no conflict of interest

References

[1] V.S.P. Chaturvedula, I. Prakash, The aroma, taste, color and bioactive constituents of tea, J. Med. Plant Res. 5 (2011) 2110-2124.

[2] H. Mukhtar, N. Ahmad, Tea polyphenols: prevention of cancer and optimizing health, The Am. J. Clin. Nutr. 71 (2000) 1698s-1702s.

[3] A. Samali, A. Rukaiyatu, K. Mustapha, Qualitative and quantitative evaluation of some herbal teas commonly consumed in Nigeria, Afr. J. Pharm. Pharmacol. 6 (2012) 384-388.

[4] Q. Chen, Z. Guo, J. Zhao, Identification of green tea's (*Camellia sinensis* L.) quality level according to measurement of main catechins and caffeine contents by HPLC and support vector classification pattern recognition, J. Pharm. Biomed. Anal. 48 (2008) 1321-1325.

[5] Q. Chen, J. Zhao, H. Lin, Study on discrimination of Roast green tea (*Camellia sinensis* L.) according to geographical origin by FT-NIR spectroscopy and supervised pattern

recognition, Acta Mol. Biomol. Spectrosc. 72 (2009) 845-850.
[6] L. Nitin Seetohul, M. Islam, W.T. O'Hare, Z. Ali, Discrimination of teas based on total luminescence spectroscopy and pattern recognition, J. Sci. Food Agr. 86 (2006) 2092-2098.
[7] T. Yi, Q. Chen, X. He, S. So, Y. Lo, L. Fan, J. Xu, Y. Tang, J. Zhang, Z. Zhao, Chemical quantification and antioxidant assay of four active components in *Ficus hirtaroot* using UPLC-PAD-MS fingerprinting combined with cluster analysis, Chem. Cent. J. 7 (2013) 115, doi.org/10.1186/1752-153X-7-115.
[8] T. Yi, L. Zhu, W.-L. Peng, X.-C. He, H.-L. Chen, J. Li, T. Yu, Z.-T. Liang, Z.-Z. Zhao, H.-B. Chen, Comparison of ten major constituents in seven types of processed tea using HPLCDAD-MS followed by principal component and hierarchical cluster analysis, LWT-Food Sci. Technol. 62 (2015) 194-201.
[9] A.-R.A. Roshan, H.A. Gad, S.H. El-Ahmady, M.S. Khanbash, M.I. Abou-Shoer, M.M. Al-Azizi, Authentication of monofloral yemeni sidr honey using ultraviolet spectroscopy and chemometric analysis, J. Agr. Food Chem. 61 (2013) 7722-7729.
[10] H.A. Gad, A. Bouzabata, Application of chemometrics in quality control of Turmeric (*Curcuma longa*) based on Ultra-violet, Fourier transform-infrared and 1H-NMR spectroscopy, Food Chem. 237 (2017) 857-864.
[11] H.A. Gad, S.H. El- Ahmady, M.I. Abou- Shoer, M.M. Al- Azizi, A modern approach to the authentication and quality assessment of thyme using UV spectroscopy and chemometric analysis, Phytochem. Anal. 24 (2013) 520-526.
[12] F. S. Youssef, M. Ashour, Sobeh M, H. El-Beshbishy, A. Singab, M. Wink, *Eremophila maculata*- Isolation of a rare naturally-occurring lignan glycoside and the hepatoprotective activity of the leaf extract, Phytomedicine. 23 (2016) 1484-1493.
[13] M. Sobeh, M.S. Braun, S. Krstin, F.S. Youssef, M.L. Ashour, M. Wink, Chemical profiling of the essential oils of *Syzygium aqueum*, *Syzygium samarangense* and *Eugenia uniflora* and their discrimination using chemometric analysis, Chem. Biodivers. 13(2016) 1537–1550.
[14] M.L. Ashour, F.S. Youssef, H.A. Gad, M. Wink, Inhibition of cytochrome P450 (CYP3A4) activity by extracts from 57 plants used in traditional chinese medicine (TCM) , Pharmacog. Mag. 13 (2017) 300 -308.
[15] D.P. Rivelli, V.V.d. Silva, C.D. Ropke, D.V. Miranda, R.L. Almeida, T.C.H. Sawada,

S.B.d.M. Barros, Simultaneous determination of chlorogenic acid, caffeic acid and caffeine in hydroalcoholic and aqueous extracts of *Ilex paraguariensis* by HPLC and correlation with antioxidant capacity of the extracts by DPPH· reduction, Braz..J. Pharmacog. 43 (2007) 215-222.

[16] G. Ma, Y. , Zhang, J. Zhang , G. Wang, L., Chen, M., Zhang, T. Liu , X. Liu, C. Lu, Determining the geographical origin of Chinese green tea by linear discriminant analysis of trace metals and rare earth elements: taking Dongting Biluochun as an example, Food Cont. 59 (2016) 714-720.

Fig. 1. PCA score plot using the UV data matrix (200–400nm) of methanol extracts of **(a)** the identified green tea samples, **(b)** identified and the marketed green tea samples

Fig. 2. A clustering dendrogram using the UV data matrix (200–400nm) of methanol extracts of **(a)** the identified green tea samples, **(b)** identified and the marketed green tea samples

Fig. 3. (a) Cooman's plot for the validation of the SIMCA model using the East and South Asian models, **(b)** Authentication of marketed green tea samples using SIMCA model

Table 1

Table 2

Composition of caffeine (mg/g) dry weight in different green tea samples estimated using HPLC[a]

Table 3

In vitro antioxidant activity of the tested green tea samples using (DPPH·) scavenging capacity assay *via* the determination of their IC_{50} (mg/mL)