JOURNAL OF AGRICULTURAL AND FOOD CHEMISTRY

Metabolite Profiling Using ^{1}H NMR Spectroscopy for Quality Assessment of Green Tea, *Camellia sinensis* (L.)

GWÉNAËLLE LE GALL,* IAN J. COLQUHOUN, AND MARIANNE DEFERNEZ

Institute of Food Research, Norwich Research Park, Colney, Norwich NR4 7UA, United Kingdom

A set of 191 green teas from different countries was collected and analyzed by ^{1}H NMR. It was proposed to establish if the teas could be discriminated according to the country of origin or with respect to quality. Both principal component analysis (PCA) and cluster analysis were applied to the data. Some separation of Chinese and non-Chinese teas was observed. The present results did not allow allocation of samples to individual countries, but cluster analysis suggested that it might be possible with an augmented sample set. The PCA did show a separation between the Longjing type (highest quality Chinese tea) and most other Chinese teas and indicated some metabolites that could be responsible for the difference. Longjing teas showed higher levels of theanine, gallic acid, caffeine, epigallocatechin gallate, and epicatechin gallate and lower levels of epigallocatechin when compared with other teas. These compounds have been mentioned previously in connection with quality, but it was also shown that higher levels of theogallin (5-galloyl quinic acid), theobromine, 2-*O*-(β-L-arabinopyranosyl)-myo-inositol and some minor sugar-containing compounds were found in Longjing teas while higher levels of fatty acids and sucrose were found in the other teas. These new markers could prove to be useful for the authentication of bulk tea.

KEYWORDS: NMR; green; tea; metabolite; profiling; fingerprinting; composition; content; quality; authenticity; authentication; origin

INTRODUCTION

Tea is the most widely consumed drink in the world after water. There are three main types of tea: green (unfermented), oolong (semifermented), and black (fermented). The fermentation is an oxidation process, which produces theaflavins and thearubigins, two families of polyphenols present in black and oolong teas. Black tea is consumed worldwide while green and oolong teas are mainly consumed in Asia and Northern Africa.

The chemical composition of tea depends on several factors: genetic strain, climatic conditions, soil, growth altitude and horticultural practices, the plucking season, sorting (grading) of the leaves, the processing, storage, etc. Some parameters are more important than others. For example, the best green teas are usually plucked during the first flush in April or May (there are three main plucking times: spring, summer, and autumn). The quality of a tea is mainly assessed through its appearance (color, color intensity, and cloudiness), its flavor (astringency, bitterness, and sweetness), and its aroma (floral, sweet, grassy, etc.). However, the factors that determine quality differ somewhat between black and green teas. Theaflavins, thearubigins, catechins, and caffeine are known to be responsible for black tea quality (*1*, *2*), but green tea quality depends more on the content of amino acids, catechins, and caffeine (*3*−*5*).

Many characteristics are taken into account to judge the quality, and these are reflected in the price of a tea. Green teas of different origin have their own distinctive characters, but in general, high quality green teas are described as “delicate” or “sweet”. The highest quality tea from Japan called Gyokuro (the Ceremonial Tea “Matcha” is a pulverized version of Gyokuro) is grown under shade and is reported to contain a high amino acid but low catechin content (*4*, *6*). Sweetness is attributed to amino acids (*7*) especially theanine, which has the taste described as “umami” or “brothy”, while the catechins and caffeine contribute to the astringency (*1*). The appearance of the leaves and the color of the brew play a major role in the evaluation of green tea: young leaves and a clear brew with a pale green-yellowish tint are indicators of high quality. The quality of a green tea declines with signs of cloudiness and brown-reddish pigments. Last but not least, the aroma is also involved in the quality assessment of green tea.

The quality and price of a tea have been and continue to be established through the judgment of professional tea tasters, but there is a growing concern to assess the quality of teas by some form of measurement. Green tea has been increasingly analyzed during the last 10 years, more because of its health-related properties than for quality evaluation (although the claimed antioxidative, anticarcinogenic, and antitumorigenic properties as well as the ability to protect against cardiovascular disease are still to be demonstrated in vivo (*8*)). Most of the chromatographic techniques developed have been aimed at catechin quantification (*9*−*11*) or at quantification of catechins plus other compounds believed to be possible quality markers such as

* To whom correspondence should be addressed. Tel: +44 1603 255128. Fax +44 1603 507723. E-mail: gwenaelle.legall@bbsrc.ac.uk.

10.1021/jf034828r CCC: $27.50
Published on Web 01/27/2004

caffeine, gallic acid, and theanine (*3*, *5*, *12*). There has been little systematic analysis of other types of compound in relation to green tea quality. A few articles have described characterization of the aroma compounds of high quality teas (*13*−*15*), mainly by gas chromatography. Another paper reports the use of chemical data as various as fiber and amino acid contents (*16*). Near-infrared spectroscopy (NIR) has also been used (*17*−*19*), but the technique is mainly aimed at replacing time-consuming analyses such as the Kjeldahl and Folin−Ciocalteu methods used for quantifying the total nitrogen and the total polyphenol contents. To the best of our knowledge, high-resolution NMR has never been used for quality evaluation of tea.

The present study shows that ^{1}H NMR can be used to analyze simultaneously the catechins, the amino, organic, phenolic, and fatty acids, and the sugars from a single green tea extract. As well as the "usual" markers (catechins, caffeine, or theanine), compounds such as theogallin (5-galloyl quinic acid) or 2-*O*-(β-L-arabinopyranosyl)-myo-inositol are also detected. The latter compounds have not often been analyzed but have been described as potentially playing a role in quality (*20*, *21*). Theogallin has been quantified a few times by chromatography (*21*−*23*) but was only once considered as a quality marker by the authors (*21*). As for 2-*O*-(β-L-arabinopyranosyl)-myo-inositol, this sugar was first determined to be present in tea by Sakata et al. (*20*) and has never been analyzed for quality purposes.

The database of green teas collected for this study is large: just under 200 green teas from China, Japan, Vietnam, India, Indonesia, and Bangladesh were gathered over 2 years. It is first established whether the teas can be discriminated according to the country of origin by chemometric analysis of their ^{1}H NMR spectra. This is followed by an analysis targeted at determining which metabolites are associated with the most highly regarded and priced Chinese green teas, the Longjing type.

MATERIALS AND METHODS

Materials. One hundred ninety-one green teas were available for this study, of which 17 were jasmine teas (scented). About half of them (including all the non-Chinese teas) were supplied by an industrial partner, and the rest were purchased from the Tea Research Institute, Hangzhou (China). The database was composed of 168 green teas originating from China, seven from Japan, six from Indonesia, five from India, three from Vietnam, and two from Bangladesh. Only the seven Japanese teas had undergone a steaming process. These consisted of one Gyokuro superfine (finest quality of Japanese tea), one Bancha (made of coarse/old leaves), and five Sencha of different grades (Sencha is one of the types most commonly drunk in Japan). The rest were roasted. Thirty-eight Chinese teas were Longjing, also called "Dragon well" teas. They were graded from superfine to quaternary (superfine, first, second, third, and quaternary grade). These were the most expensive teas along with the Gyokuro. They were grown mainly in Zhejiang county (east coast of China). Only a few Longjing teas came from other areas such as Shandong (northeast China closer to Beijing) or Chongqing district in Sichuan (west China near the mountains). In general, a high proportion of the 168 Chinese teas came from Zhejiang county where the most intensive cultivation in the whole country is carried out. Other Chinese teas were from nearby counties such as Fujian, Anhui, Hubei, or Hunnan or more remote counties such as Guizhou, Sichuan, and Guangdong. A variety of styles with different grades is also represented in the database: Longjing (made of the two youngest leaves and buds), gunpowder (leaves of different ages rolled to look like gun bullets), and imperial (made of young to medium leaves in a twisted style) such as Monkey king or Chunmee.

Methods. *Extraction.* A 0.096 ± 0.005 g amount of tea (finely ground using a coffee grinder) was stirred with 1200 μL of extraction solvent consisting of 70% methanol-d_4 and 30% buffer solution (30 mM Na_2HPO_4 and 5 mM sodium 3-(trimethylsilyl)propionate-d_4 as a reference in D_2O) at 70 °C for 10 min, allowed to cool, and centrifuged at 10 000 rpm for 10 min (Jouan A14 centrifuge). Each NMR sample consisted of 750 μL of the supernatant, which was stored at −18 °C until required for analysis.

NMR Spectroscopy. ^{1}H NMR spectra were recorded at 27 °C with a 400 MHz JEOL GX spectrometer fitted with an autosampler. Methanol-d_4 was used as the internal lock. Each spectrum consisted of 104 scans of 8192 complex data points with a spectral width of 5000 Hz, an acquisition time of 1.64 s, and a recycle delay of 2 s per scan. The pulse angle was 50°. A presaturation sequence was used to suppress the residual water signal with low power selective irradiation at the water frequency during the recycle delay. Spectra were Fourier transformed with 1 Hz line broadening, phased, and baseline corrected using the JEOL (Delta) software. Spectra were converted to Felix 2000 software format and saved as ASCII files. Spectra were further transferred to a personal computer for data analysis.

Multivariate Analysis. *A. Principal Component Analysis (PCA).* The application of the technique to NMR data and the interpretation of PC loadings have already been described in previous work (*24*).

B. Hierarchical Cluster Analysis (25, 26). Cluster analysis techniques represent the intrinsic structure of the data without making a priori assumptions about the origin of the samples. Most of the techniques are based on the measurements of distances or correlation between objects. Clusters are formed hierarchically on the basis of proximity of objects or clusters of objects. In this work, we used an agglomerative clustering procedure based on standardized Euclidean distances and the nearest neighbor (single-link) method. The inputs for the cluster analysis were the PC scores of the samples (obtained by correlation method PCA). The outputs of the cluster analysis are presented as a dendrogram (tree-shaped diagram) that pictures the similarities and dissimilarities between the samples.

Univariate Analysis. Results of multivariate analyses can be difficult to interpret in terms of specific compounds so we have applied analysis of variance (ANOVA) (*27*) to selected NMR signals to determine whether there are significant differences between mean concentrations of individual compounds in the "Longjing" and "other tea" groups.

Software. PCA was carried out in Matlab, version 6.1 (The MathWorks Inc, Natick, MA). For this, 5140 points were extracted from the original NMR spectrum in Felix ASCII format (originally 8192 points) using a PASCAL program written in-house. Parts of the spectrum that do not contain any signals were excluded (the region between points numbered 1701 and 6840 was kept for PCA). An in-house Matlab macro was used to achieve alignment of peaks within a chosen region across a series of spectra (see Results and Discussion section). *F* values and box plots were calculated using the Matlab macro "anova1" (Statistics toolbox). The cluster analysis was performed in Matlab using an interactive function called gcluster.m from the PLS Toolbox version 2.0 (Eigenvector Research, Manson, WA).

RESULTS AND DISCUSSION

Signal Assignments. Figure 1A shows the ^{1}H NMR spectrum of a superfine grade Longjing green tea selected as a typical example for detailed signal assignments. The procedure adopted for the signal assignment has already been described previously (*28*). **Table 1** summarizes the chemical shift information available for green tea from the two-dimensional (2D) spectra and the reference standards.

About 30 compounds were identified in the one-dimensional (1D) and 2D spectra, and more than 50 signals or groups of signals were indexed overall. Apart from ubiquitous compounds such as amino and fatty acids and common sugars such as sucrose and glucose that are assigned in the literature compilations (*29*), the signals of phenolics, flavonoids (flavan-3-ols or catechins, flavonols), xanthines, and minor sugars can be observed. Theanine, unique to tea, is the predominant amino acid present (*7*). There is a variety of catechins present in the green tea used to assign the proton signals. Catechins account

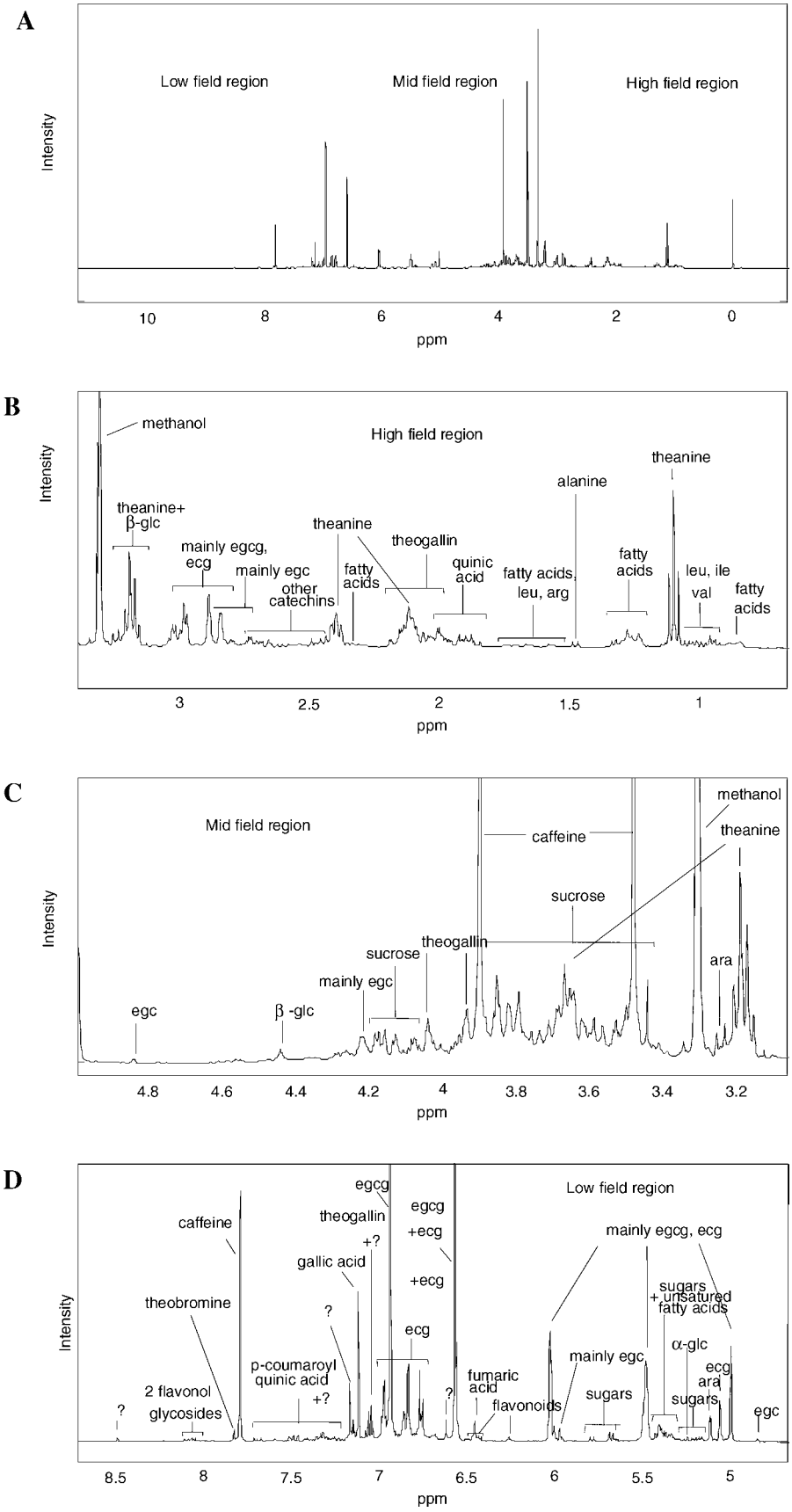


Figure 1. Details of ^{1}H NMR spectrum of a high grade Longjing green tea extract. Key: leu, leucine; ile, isoleucine; val, valine; arg, arginine; glc, glucose; and ara, 2-*O*-(*β*-L-Arabinopyranosyl)-myo-inositol.

for 30% of solids (wt/wt) in a typical tea beverage (*30*). Standards of available catechins were run for this study in the extraction solvent to help the catechin assignment. The chemical shifts were in good agreement with those previously reported for catechins in deuterated acetone (*31*). Five unidentified catechins can be detected along with the five predominant ones,

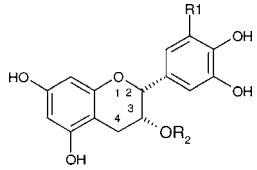
Table 1. ^{1}H Chemical Shifts of Compounds from 1D and 2D Spectra of a High Grade Longjing Green Tea

compound	chemical shifts (ppm)[a]
linolenic acid	0.96, 1.28, 1.30, 1.58, 2.07, 2.35, 2.80, 5.37
other fatty acids	0.88, 1.28, 1.60, 2.24, 2.36, 5.32
leucine	0.98, 1.78, 1.67, 3.66
isoleucine	1.04, 1.98
valine	1.06, 2.31
theanine	1.12, 2.15, 2.45, 3.22, 3.70
lactic acid[b]	1.34, 4.00
threonine[b]	1.35, 4.22
alanine	1.51, 3.71
arginine	1.73, 3.27
lysine	1.73, 3.03
quinic acid	1.90, 1.92, 2.03, 2.04, 3.98, 4.05, 3.55
γ-amino butyric acid	1.92, 2.34, 3.03
theogallin	2.02, 2.15, 2.20, 3.85, 4.21, 5.42, 7.08
glutamic acid[b]	2.03, 2.45, 3.70
unknown catechin 1	2.82, 2.97, 3.94
unknown catechin 2	2.48, 2.77, 4.31
C	2.54, 2.84, 4.07, 4.63
unknown catechin 3	2.64, 2.84, 3.84, 4.22
EGC	2.74, 2.89, 4.25, 4.87, 6.58
EC	2.74, 2.89, 4.26, 4.81
unknown catechin 4	2.89, 3.00, 5.03
EGCG	2.89, 3.03, 5.03, 5.50, 6.03, 6.59, 6.95
ECG	2.89, 3.03, 5.09, 5.50, 6.03, 6.78, 6.88, 6.99
unknown catechin 5	3.10, 3.34, 3.91
caffeine	3.33, 3.50, 3.92, 7.80
β-glucose	4.58, 3.21
2-*O*-(β-L-arabinopyranosyl)-myo-inositol	5.14, 3.26, 3.61, 3.68, 3.89, 3.97, 4.18
α-glucose	5.20, 3.50
sugar 1	5.36
sugar 2	5.38
sugar 3	5.42
sucrose	5.42, 3.43, 3.53, 3.76, 3.84, 3.80, 4.05, 4.19
sugar?	5.64
sugar 4	5.67, 3.73, 4.02,[b] 3.85, 4.14, 4.88, 5.21[b]
sugar 5	5.79, 3.76, 3.97, 5.26[b]
hydroxycinnamic compound 1[c]	6.34, 7.58
fumaric acid[b]	6.48
hydroxycinnamic compound 2[c]	6.43, 7.66
p-coumaroyl quinic acid[b]	6.46, 6.88, 7.51, 7.72
?[d]	6.83, 7.18
?[d]	6.83, 7.18, 7.63
kaempferol glycoside	6.98, 8.11
quercetin glycoside	6.98, 7.63, 8.09
gallic acid[b]	7.14
threobromine	7.84

[a] Spectra referenced to methanol = 3.3 ppm. [b] Provisional assignment. [c] Either chlorogenic (5-caffeoyl quinic acid), neo-chlorogenic (3-caffeoyl quinic acid), *p*-coumaroyl quinic acid or caffeic acid. [d] Could come from chlorogenic, neo-chlorogenic, caffeic acids, or any compound containing an aromatic ring with an OH substitution in position 3′ and 4′ (e.g., quercetin).

which are (−)-epigallo-catechin-3-gallate (EGCG), (−)-epicatechin-3-gallate (ECG), (−)-epigallo-catechin (EGC), (−)-epicatechin (EC), and (+)-catechin (C). This family of compounds can easily be identified by the characteristic signals arising from H-3 (3.80−5.20 ppm) and H-4 (2.50−3.10 ppm) of the heterocyclic ring (see **Scheme 1**), which are cross-correlated in COSY spectra. At least two other unidentified catechins were observed in the ^{1}H NMR spectra of extracts of other roasted teas (data not shown). These unknown catechins could be (−)-gallocatechin-3-gallate (GCG), (−)-gallocatechin (GC), (−)-catechin-3-gallate (CG), (−)-epigallo-catechin-3-(3″-O-methyl)-gallate, (−)-epigallocatechin-3,5-digallate, (−)-epicatechin-3,5-digallate, or epiafzelechin (*32*). GCG, GC, and CG are often seen as heat-converted catechins (*32*−*34*). Caffeine is the major

Scheme 1. Chemical Structure of Some Catechins[a]



[a] EGCG, R_1 = OH, R_2 = gallate; ECG, R_1 = H, R_2 = gallate; EGC, R_1 = OH, R_2 = H; and EC, R_1 = H, R_2 = H.

xanthine observed in the spectrum. Theobromine is also detected (7.84 ppm), but the theophylline level is too low to allow its detection. Minor signals in the low field region were assigned to kaempferol and quercetin glycosides (flavonols) along with signals of gallic acid, theogallin, some hydroxycinnamic derivatives (caffeoyl quinic acids?), and probably *p*-coumaroyl quinic acid (**Figure 1D**). In the high field region, signals of theogallin are seen as well as those of quinic acid (**Figure 1B**). Finally, minor signals attributed to sugars are seen in the midfield region of the Longjing green tea spectrum (**Figure 1D**).

Preparation of NMR Data for Chemometrics. Because caffeine peak positions as well as other peaks (including catechins) may be displaced from one sample to another, a program was written to allow the local alignment of the peaks within a data matrix. This program is a Matlab macro written in-house based on the principles of partial linear fit (PLF) (*35*). Briefly, the method consists of selecting portions of the spectrum, comparing them to a reference spectrum (mean spectrum), and shifting them left or right so that the difference between the particular region and the corresponding region in the reference spectrum is minimized (*35*). Additionally, the macro used is an iterative program that recalculates the mean spectrum at each pass and then repeats the shifting procedure so that the difference is further minimized. In relatively polar solvents, as used here, caffeine molecules associate both with each other and with polyphenols to form a range of oligomeric species in which the proton chemical shifts are different from those of the single molecules. The observed chemical shifts are average values that depend on the oligomeric species present, and because the nature and proportion of these are concentration-dependent, the effective shifts vary from sample to sample. Small differences of pH between samples can also lead to displacements of chemical shifts for molecules with ionizable groups.

Figure 2 depicts a set of 60 tea spectra in the low field region before and after peak alignment. The improvement in data registration after peak alignment is obvious. Note that the local alignment is a necessary operation: this experimental problem seriously affected the covariance matrix PCA results as shown in **Figure 3**. The pattern of the covariance matrix PCA results before alignment almost described an "arabesque" style figure in both PC1/PC2 and PC2/PC3 scores plots (**Figure 3A**,**B**). The cause of this behavior is related to the way some individual peaks (especially those of caffeine) are shifted from sample to sample (*36*). This example illustrates the importance of peak alignment before chemometric analysis in order to obtain sound results.

Instrumental Repeatability. NMR spectra of all the green tea extracts were run twice during one acquisition session. **Figure 4** displays the scores of 140 samples along with 14 instrumental replicates. The overall spread of the scores of the large number of samples can be compared visually to the closeness of the scores of the 14 original spectra selected for this display to their replicates. This PCA plot shows that the instrumental repeatability is satisfactory.

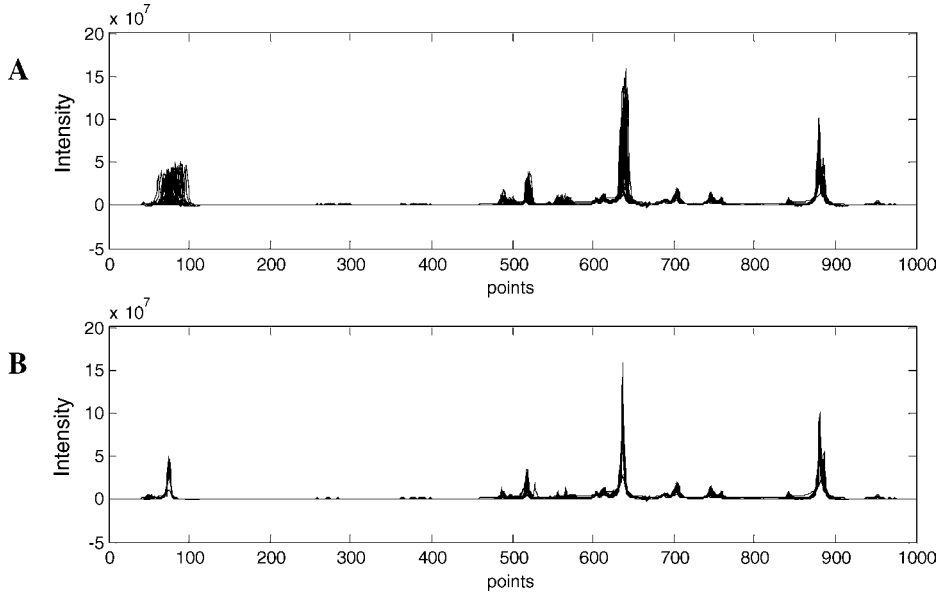


Figure 2. Example of 60 green tea spectra before (**A**) and after (**B**) local peak alignment.

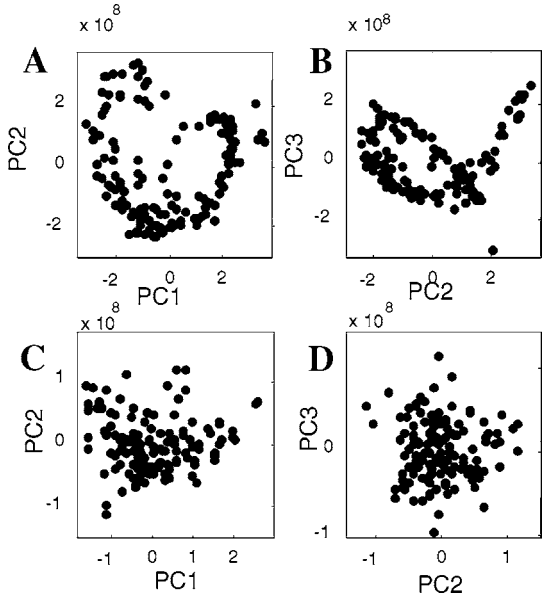


Figure 3. Demonstration on the importance of alignment of shifting peaks. PCA on 140 green tea ^{1}H NMR spectra with nonaligned peaks (**A** and **B**) and PCA on the same spectra after peak alignment (**C** and **D**). Covariance matrixes were used.

Relationship between Country of Origin and ^{1}H NMR Chemical Profile of Green Tea Extracts. All of the 191 green teas were used in a first approach for data exploration (see Materials) carried out using PCA. In the PC1/PC2 plot (**Figure 5**), a clustering of the scores of non-Chinese teas occurs along the PC1 axis while the Chinese tea scores are scattered evenly within the plot. There is no further obvious clustering according to the country of origin. Thus, although the distribution of Chinese and non-Chinese samples in the scores plot clearly suggests a country of origin effect, it is not possible to achieve a more specific allocation to individual countries with this sample selection. This is partly due to the small number of samples from each country apart from China.

Cluster analysis was then applied to the ^{1}H NMR spectra of 10 Chinese teas (covering the range of sample types but otherwise randomly selected) and the 23 other non-Chinese teas. Details of this selection are given in **Table 2**. The dendrograms

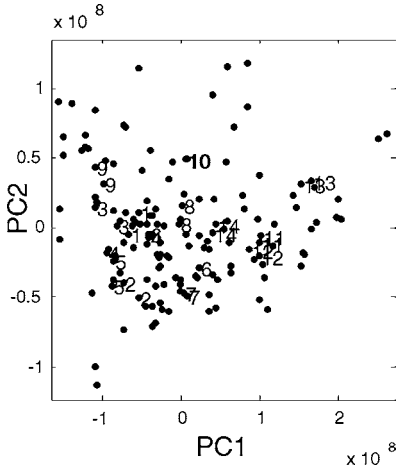


Figure 4. Instrumental repeatability evaluation. PCA applied on a total of 154 ^{1}H NMR spectra of green tea extracts (covariance matrix). Fourteen original spectra are paired (numbers) to their instrumental replicates.

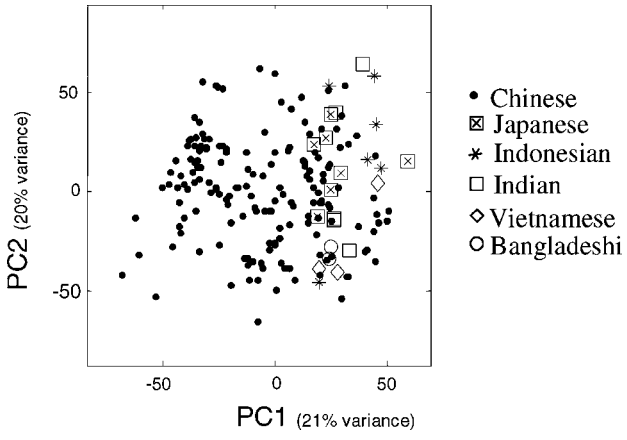


Figure 5. PCA on the ^{1}H NMR spectra of 191 green teas originating from different countries (correlation matrix).

obtained here were used as another means of representing the multivariate data, this time on a geographically balanced set of teas. The input for the cluster analysis consisted of the scores from a PCA applied to the selected data set. This method provided a simple representation of the data that included

Table 2. List of Teas Used for Cluster Analysis

database	country of origin	type of green tea	name	grade[a]
1	China	green	Xihu Longjing	SP
2	China	green	Zhejiang Longjing	sec
3	China	green	Xueshuiyunlu	no class
4	China	green	Wuyi	first
5	China	green	Putuo Buddha Tea	first
6	China	green	Wansheng	SP
7	China	green	Gunpowder	leaf
8	China	green	Hubei Tengda	first
9	China	jasmine	Chonqing	SP
10	China	jasmine	Lion (special)	
11	Japan	green	Gyokuro	SP
12	Japan	green	Bancha	(coarse leaf)
13	Japan	green	Fukaamushi Sencha	(deep steamed)
14	Japan	green	Sencha	regular
15	Japan	green	Sencha	premium
16	Japan	green	Sencha	sec
17	Japan	green	Sencha	first
18	Indonesia	green	Chunmee	
19	Indonesia	green	Chunmee	
20	Indonesia	green	Chunmee	
21	Indonesia	green	Gunpowder	
22	Indonesia	green	Gunpowder	
23	Indonesia	green	Gunpowder	
24	South India	green		
25	India	green	Chunmee	
26	India	green	Chunmee	
27	India	green	Chunmee	
28	India	green	Gunpowder	
29	Vietnam	green		
30	Vietnam	jasmine		
31	Vietnam	jasmine		
32	Bangladesh	green		
33	Bangladesh	green		

[a] Abbreviations: SP, superfine; sec, secondary grade.

information from more PCs than can be displayed on a single scores plot. Reduction in the number of Chinese samples also greatly simplified the appearance of the plots. The outputs of several models based on 3, 5, 8, and 10 PCs (presented in **Figure 6A−D**) are compared. The figures for cumulative variance explained by each model (**Figure 6**) suggest that it may be beneficial to look beyond the model for three PCs. The dendrograms differ from each other, but there are a number of samples that remain clustered regardless of the number of PCs used, four Japanese teas labeled 13 to 16, two pairs of Indian teas (25, 26 and 27, 28), two Indonesian teas (22 and 23), two Vietnamese (29 and 31), and two Chinese teas (1 and 2) (**Figure 6A−D**). The four Japanese teas are all Sencha (see **Table 2** for the description of the different types of green tea), and the fifth Sencha sample (labeled 17) falls close to their cluster while the two Chinese teas belong to the same class (Longjing). These results highlight the influence of the class of the green teas along with the influence of their geographical origin. The latter is illustrated by the case of the Indian teas where four samples are grouped in two pairs in the first two models (**Figure 6A**,**B**) and then later clustered together in the two other models (**Figure 6C**,**D**). The fifth Indian tea remains on its own, and it is the only one grown in South India (**Table 2**). Note that the large majority of Indian teas are grown in the North (Darjeeling and Assam). As for the Vietnamese and Bangladeshi teas, they are clustered according to their respective country of origin in most models, and Indonesian teas show a tendency to cluster in the models using three and five PCs.

It therefore appears that the geographical origin may be indicated by the chemical composition of green teas (however, as the number of non-Chinese teas is limited, it is difficult to give a final conclusion). This provisional conclusion is in

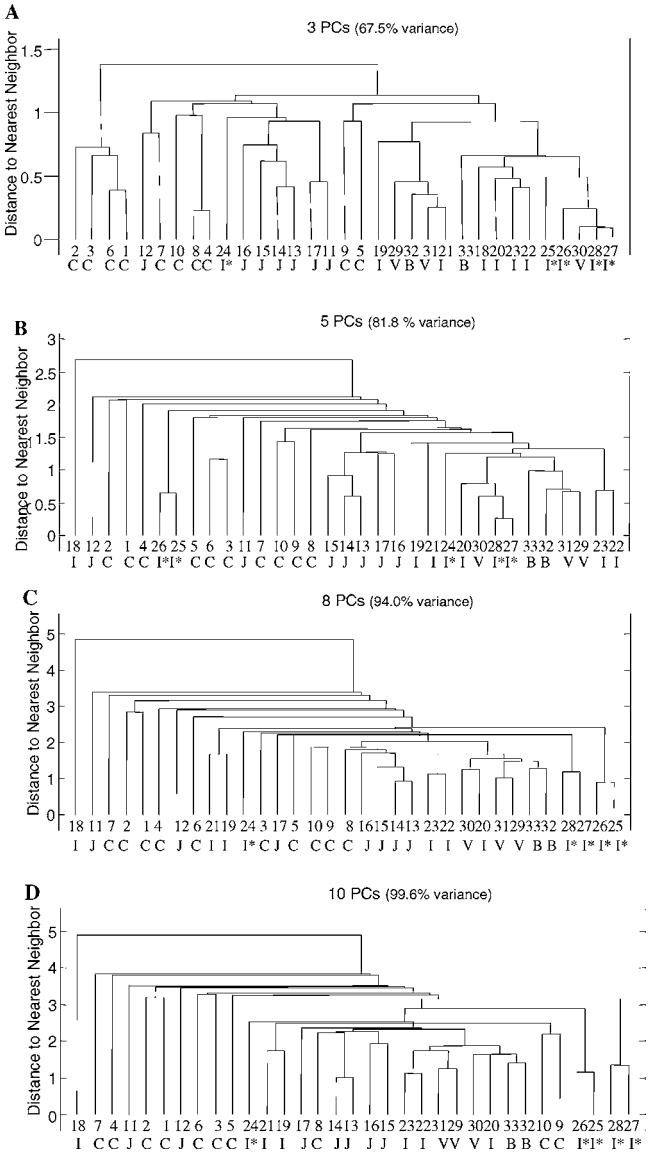


Figure 6. Dendrograms obtained from cluster analysis applied on 33 green tea ^{1}H NMR spectra using three (**A**), five (**B**), eight (**C**), and 10 (**D**) principal components. Key: B, Bangladeshi; C, Chinese; J, Japanese; I, Indonesian; I*, Indian; and V, Vietnamese.

agreement with the work described by Kim et al. (*37*), who, on the basis of NIR spectra, established that it was possible (i) to discriminate Japanese from Korean green teas and (ii) to discriminate teas from different processing methods, although Korean steamed teas were closer to their roasted counterparts than to the Japanese steamed ones. In fact, several articles refer to the possibility of classifying teas (green and black sometimes associated into one data set) on the basis of their country of origin (*38−40*), but all have in common few samples to represent the different populations of teas.

The origin of the clustering of the samples shown in this study may be related to several factors: the genetic strain, the climate, the altitude of growth, the type of soil, the use of fertilizer, or even the processing of the tea leaves. All have an influence on the overall chemical content of tea. Both PCA and cluster analysis seem to indicate that the green teas can partially be classified according to their country of origin but pinning down the underlying factor(s) responsible for such a classification is difficult. The difficulty of such a study (apart from having a representative number of samples for each population) is that

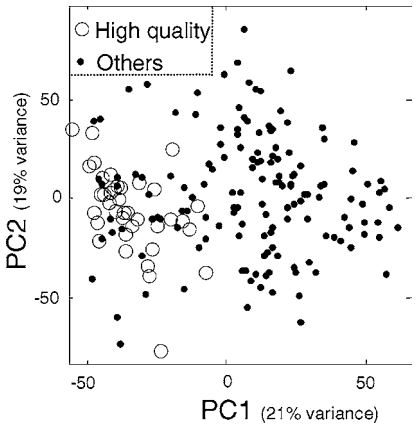


Figure 7. PCA on the ^{1}H NMR spectra of 184 green teas labeled as two groups: high quality teas (38 Longjing) and other teas (146) (correlation matrix).

there are many factors to take into account to attempt a classification not just the country of origin. For example, it could certainly be argued that teas from a country as large as China with its diversity of climates and soil types do not form a single homogeneous class. Therefore, although some degree of classification is observed for this set of data, the interpretation of the results remains open to debate.

Relationship between Quality and ^{1}H NMR Chemical Profile of the Chinese Green Tea Extracts. Some factors that are not environment or process related affect the chemical composition of green tea. The size (age) of the leaves is actually the main criterion for grading green tea, and unlike black tea, green tea quality is directly related to the grading. Teas made of the youngest leaves and the bud are the most expensive. In the present data set, 38 high quality teas (Longjing or Dragon well), graded from superfine to fourth grade, were priced from £12−15 per 50 g (superfine) down to £4−6 per 50 g (fourth grade). In comparison, the 77 other Chinese teas purchased from the Tea Research Institute, Hangzhou, cost £2.80 per 50 g (superfine) to £0.80 per 50 g (fourth grade). The cost of the teas supplied by the industrial company is not known, but none of the teas are Longjing and the majority fall under styles such as "roasted", "Maofeng", "imperial", or "gunpowder", which are known to be reasonably cheap (i.e., under £4 per 50 g). With such a data set, it was therefore interesting to compare the two groups of teas represented by (i) the Longjing and (ii) all the others in order to see if they could be discriminated and, if so, to determine which metabolites could be involved in the group separation.

PCA was applied to a matrix of 184 ^{1}H NMR spectra (the seven Japanese teas, which were steamed, were removed). A clear division was obtained on the first PC with the Longjing teas having negative scores and most of the others positive scores (**Figure 7**). The corresponding loading (**Figure 8**) is negative at the positions of theanine, theogallin, the minor sugars between 5.36 and 5.79 ppm, epicatechin gallate, gallic acid, caffeine, and theobromine and positive at the levels of fatty acids, quinic acid, sucrose, and epigallocatechin. All of the compounds with negative signals are more prominent in teas with negative scores (that is the Longjing teas), and all compounds with a positive loading value are present at higher levels in the teas with positive scores (other teas). These results are in agreement with the previous work. Horie and Kohata (*5*), who analyzed green tea leaves according to their maturity, suggested that high levels of epigallocatechin gallate, epicatechin gallate, theanine, and caffeine and low levels of epigallocatechin and epicatechin are a good guide to quality. Lin et al. (*41*) also found less caffeine and more catechins in older leaves. The results by Goto et al. (*42*) and Saijo and Takeda (*32*) confirmed that higher levels of epigallocatechin are present in lower quality teas, and the former authors (*42*) indicate that high quality teas contain more caffeine and less catechins. Sakata et al. (*20*) stated that 2-*O*-(*β*-L-arabinopyranosyl)-myo-inositol is present in a higher amount in high quality teas, but it is reported as essentially tasteless. Finally, Aucamp et al. (*43*) indicate that gallic acid contributes to quality along with theanine and caffeine

ANOVA was performed on the selected peak heights of 18 compounds of interest. For this, the green teas were divided into high quality (38 Longjing teas) and other (146) groups. The results are displayed in **Table 3**. The degrees of freedom to consider for the calculation are 1 and 182. The *F* critical values for those degrees of freedom are $F_{0.05} = 3.8$, $F_{0.01} = 6.6$, and $F_{0.001} = 10.8$. It was decided not to consider significant *F* values below 4 (only quinic acid in **Table 3**).

To give a clearer idea of the distribution of the compound levels in the two groups, box plots and histograms of peak intensities are displayed in **Figure 9** for two examples. The three horizontal lines of each box are at (from bottom to top) the 25th percentile, median, and 75th percentile values. The notches represent a robust estimate of uncertainty about the median (graphical equivalence of a *t*-test). The length of the whiskers indicates the range of the data, except that any values at a distance of more than 1.5 times the interquartile range from the top or bottom of the box are considered outliers and marked as crosses. Any imbalance in the length of the boxes or the notches is an indication of skewness. The histograms give

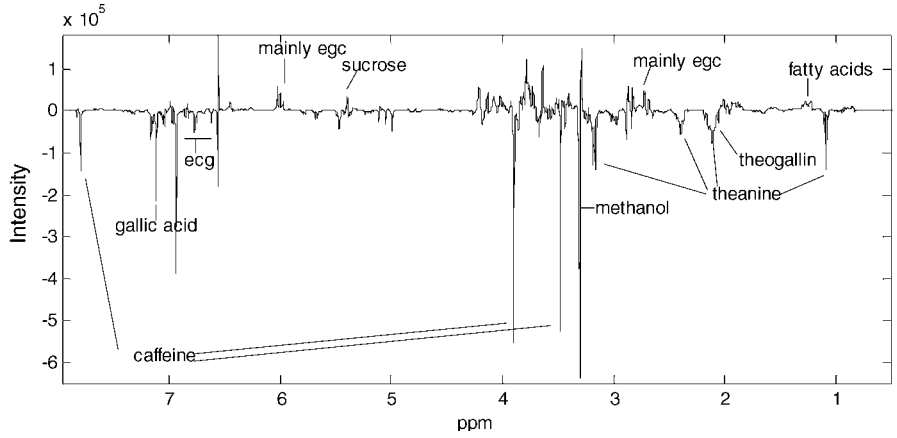


Figure 8. Loading corresponding to PC1 of **Figure 7**. See **Figure 1** caption for key to assignment.

Table 3. ANOVA Results for Selected Signals Comparing Longjing (L) and Other Teas (O)

compound[a]	*F* value	level
theanine	36.9	L > O
fatty acids[b]	41.1	O > L
quinic acid	3.71	O = L
theogallin	84.0	L > O
caffeine	32.2	L > O
EGCG	14.9	L > O
ECG	20.4	L > O
ara	9.6	L > O
EGC	63.9	O > L
sucrose	15.7	O > L
sugar at 5.36 ppm	56.7	L > O
sugar at 5.38 ppm	56.6	L > O
sugar at 5.42 ppm	41.1	L > O
signal at 5.64 ppm	55.6	L > O
sugar at 5.67 ppm	38.9	L > O
sugar at 5.79 ppm	97.5	L > O
theobromine	10.0	L > O
gallic acid	68.3	L > O

[a] Key: ara, 2-*O*-(β-L-arabinopyranosyl)-myo-inositol. [b] Includes other fatty acids than linolenic acid.

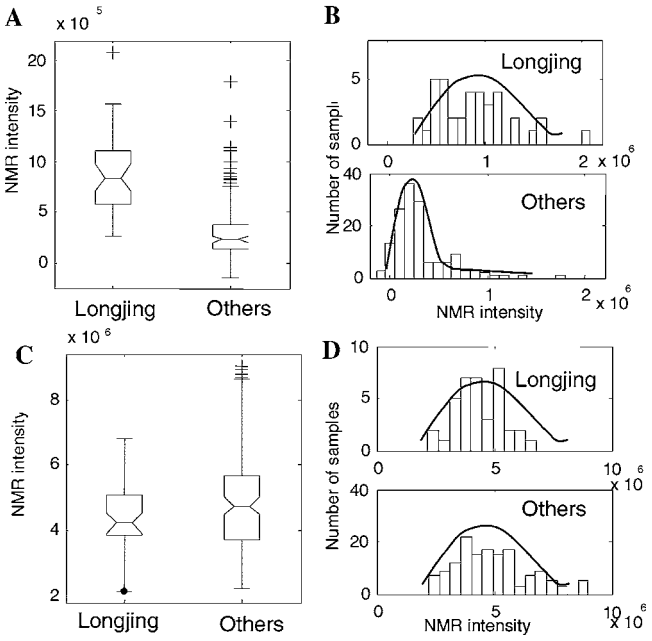


Figure 9. ANOVA box plots (**A** and **C**) and histograms of signal intensity (**B** and **D**) for the minor sugar at 5.79 ppm (**A** and **B**) and quinic acid (**C** and **D**).

another insight into the distribution of peak intensities. **Figure 9** shows the two extremes observed for the 18 compounds involved in the ANOVA study. There is only one selected compound that did not show a significant difference in level (quinic acid) despite the indication given by the loading. This emphasizes the caution needed when interpreting the loadings and the advisability of using additional means of analyzing the data (i.e., ANOVA). The levels of the rest of the compounds show significant differences between Longjing and other teas as shown in **Table 3**.

As previously mentioned, it has been reported that theanine, catechins, caffeine, and gallic acid play a role in green tea quality; however, fatty acids, sucrose, theogallin, 2-*O*-(β-L-arabinopyranosyl)-myo-inositol, and the minor signals have never previously been described as marker compounds for quality. Those compounds may not have a direct influence on organoleptic properties, and their relation to quality could be related to the age of the leaves. Indeed, Maeda Yamamoto et al. (*44*) reported that the content of sucrose increases while the content of 2-*O*-(β-L-arabinopyranosyl)-myo-inositol decreases during development of tea shoots. Those variations were accompanied by a decrease of epigallocatechin gallate, epicatechin gallate, free amino acid, and caffeine contents and an increase in the contents of epigallocatechin and epicatechin. The authors suggest that the level of 2-*O*-(β-L-arabinopyranosyl)-myo-inositol in both black and green tea might be a useful indicator that distinguishes the degree of tea shoot development.

In conclusion, after an extensive assignment of spectra, NMR spectroscopy has been shown to provide a wealth of information about the main metabolites of the teas studied. PCA analysis grouped Indonesian, Indian, Vietnamese, Bangladeshi, and Japanese green teas together on the margin of the Chinese teas. An additional study (cluster analysis) also suggested that teas were clustering according to geographical origin. However, the populations of non-Chinese teas were small (possibly not representative) and the interpretation of those results remains open to debate. The database allowed a direct comparison of two groups of teas: the expensive Longjing teas vs the rest of the teas (that do not include other high quality teas). PCA showed a segregation of the Longjing teas against the rest, and a series of ANOVAs performed on individual NMR signals confirmed this division. Longjing teas showed higher levels of theanine, theogallin, gallic acid, caffeine, epigallocatechin gallate, epicatechin gallate, 2-*O*-(β-L-arabinopyranosyl)-myo-inositol, and six other minor sugar compounds, and lower levels of fatty acids, epigallocatechin, and sucrose as compared to the other teas. It was previously once mentioned that high levels of 2-*O*-(β-L-arabinopyranosyl)-myo-inositol were observed in high quality teas, but there were no data to confirm it. The quality of tea is highly related with the ages of the tea leaves used; the younger the leaves are, the better the quality is. Although some of the compounds listed above may not directly contribute to the taste, they remain good indicators of the maturity of the tea shoots used to make a given bulk lot of tea. They could prove to be useful in authenticating bulk lots of tea made from broken leaves. This study shows that NMR combined with chemometrics and univariate statistics has a useful place in metabolic profiling research and that green tea can be added to the list of products for which ^{1}H NMR constitutes a consistent, quick, and informative screening technique.

ACKNOWLEDGMENT

We thank Adrienne Davis for sharing information on theogallin assignment, Unilever (Bedford, U.K.), and the Tea Research Institute (Hangzhou, China) for the tea samples.

LITERATURE CITED

(1) McDowell, I.; Owuor, P. O. The taste of tea. *New Sci.* **1992**, *133*, 30−33.

(2) Gulati, A.; Ravindranath, S. D. Seasonal variations in quality of Kangra tea (Camellia sinensis (L) O Kuntze) in Himachal Pradesh. *J. Sci. Food Agric.* **1996**, *71*, 231−236.

(3) Liang, Y. R.; Liu, Z. S.; Xu, Y. R.; Hu, Y. L. A study on chemical composition of 2 special green teas (Camellia-Sinensis). *J. Sci. Food Agric.* **1990**, *53*, 541−548.

(4) Chu, D. C. Green tea- its cultivation, processing of the leaves for drinking materials and kinds of green tea. In *Chemistry and Applications of Green Tea*; Yamamoto, T., Juneja, R. L., Chu, D. C., Kim, M., Eds.; CRC Press: Baton Rouge, New York, 1997; pp 1−11.

(5) Horie, H.; Kohata, K. Application of capillary electrophoresis to tea quality estimation. *J. Chromatogr. A* **1998**, *802*, 219−223.
(6) Willson, K. C. Coffee, cocoa and tea. In *Crop Production Science in Horticulture Series*, 8; Atherton, J., Rees, A., Eds.; Cabi International: Oxon, 1999; pp 248−249.
(7) Ekborg-Ott, K. H.; Taylor, A.; Armstrong, D. W. Varietal differences in the total and enantiomeric composition of theanine in tea. *J. Agric. Food Chem.* **1997**, *45*, 353−363.
(8) Higdon, J. V.; Frei, B. Tea catechins and polyphenols: Health effects, metabolism, and antioxidant functions. *Crit. Rev. Food Sci. Nutr.* **2003**, *43*, 89−143.
(9) Ding, M.; Yang, H.; Xiao, S. Rapid, direct determination of polyphenols in tea by reversed-phase column liquid chromatography. *J. Chromatogr. A* **1999**, *849*, 637−640.
(10) Dalluge, J. J.; Nelson, B. C. Determination of tea catechins. *J. Chromatogr. A* **2000**, *881*, 637−640.
(11) Wang, H. F.; Helliwell, K.; You, X. Q. Isocratic elution system for the determination of catechins, caffeine and gallic acid in green tea using HPLC. *Food Chem.* **2000**, *68*, 115−121.
(12) Lin, J. K.; Lin, C. L.; Liang, Y. C.; Lin-Shiau, S. Y.; Juan, I. M. Survey of catechins, gallic acid, and methylxanthines in green, oolong, Pu-erh, and black teas. *J. Agric. Food Chem.* **1998**, *46*, 3635−3642.
(13) Horie, H.; Fukatsu, S.; Mukait, T.; Goto, T.; Kawanaka, M.; Shimohara, T. Quality evaluation on green tea. *Sens. Actuators, B* **1993**, *13*, 451−454.
(14) Togari, N.; Kobayashi, A.; Aishima, T. Pattern recognition applied to gas chromatographic profiles of volatile components in three tea categories. *Food Res. Int.* **1995**, *28*, 495−502.
(15) Shimoda, M.; Shigematsu, H.; Shiratsuchi, H.; Osajima, Y. Comparison of volatiles compounds among different grades of green tea and their relations to odor attributes. *J. Agric. Food Chem.* **1995**, *43*, 1621−1625.
(16) Liu, X. D.; Vanespen, P.; Adams, F.; Yan, S. H.; Vanbelle, M. Classification of Chinese tea samples according to origin and quality by Principal Component techniques. *Anal. Chim. Acta* **1987**, *200*, 421−430.
(17) Ikegaya, K. Determination of chemical constituents in processed green tea by near-infrared. *Jpn. Agric. Res. Q.* **1990**, *24*, 49−53.
(18) Schulz, H.; Engelhardt, U. H.; Wegent, A.; Drews, H. H.; Lapczynski, S. Application of near-infrared reflectance spectroscopy to the simultaneous prediction of alkaloids and phenolic substances in green tea leaves. *J. Agric. Food Chem.* **1999**, *47*, 5064−5067.
(19) Luypaert, J.; Zhang, M. H.; Massart, D. L. Feasibility study for the use of near-infrared spectroscopy in the qualitative and quantitative analysis of green tea, Camellia sinensis (L.). *Anal. Chim. Acta* **2003**, *478*, 303−312.
(20) Sakata, K.; Yamauchi, H.; Yagi, A.; Ina, K.; Parkanyi, L.; Clardy, J. Chemical studies on tea by ^{13}C NMR spectroscopy. 2-*O*-(β-L-Arabinopyranosyl)-myo-inositol as a main constituent of tea (Camellia-Sinensis). *Agric. Biol. Chem. Tokyo* **1989**, *53*, 2975−2979.
(21) Kuhr, S.; Engelhardt, U. H. Determination of flavonols, theogallin, gallic acid and caffeine in tea using HPLC. *Z. Lebensm. Unters. Forsch.* **1991**, *192*, 526−529.
(22) Bailey, R. G.; McDowell, I.; Nursten, H. E. Use of an HPLC photodiode-array detector in a study of the nature of a black tea liquor. *J. Sci. Food Agric.* **1990**, *52*, 509−525.
(23) Shao, W. F.; Powell, C.; Clifford, M. N. The analysis by HPLC of green, black and Puer teas produced in Yunnan. *J. Sci. Food Agric.* **1995**, *69*, 535−540.
(24) Le Gall, G.; Puaud, M.; Colquhoun, I. J. Discrimination between orange juice and pulp wash by H-1 nuclear magnetic resonance spectroscopy: Identification of marker compounds. *J. Agric. Food Chem.* **2001**, *49*, 580−588.
(25) Brereton, R. G. In *Chemometrics. Applications of Mathematics and Statistics to Laboratory Systems*; Ellis Horwood: Chichester, U.K., 1990; pp 244−262.
(26) Krzanowski, W. J. In *Principles of Multivariate Analysis: A User's Perspective*; Clarendon Press: Oxford, 1988; pp 89−102.
(27) Wonnacott, T. H. W.; Wonnacott, R. J. W. In *Introductory Statistics*; Wiley: Chichester, U.K., 1990.
(28) Le Gall, G.; Colquhoun, I. J.; Davis, A. L.; Collins, G. J.; Verhoeyen, M. E. Metabolite profiling of tomato using 1H NMR spectroscopy as a tool to detect potential unintended effects following a genetic modification. *J. Agric. Food Chem.* **2003**, *51*, 2447−2456.
(29) Fan, T. W.-M. Metabolite profiling by one- and two-dimensional NMR analysis of complex mixtures. *Prog. Nucl. Magn. Reson. Spectrosc.* **1996**, *28*, 161−219.
(30) Harbowy, M. E.; Balentine, D. A. Tea chemistry. *Crit. Rev. Plant Sci.* **1997**, *16*, 415−480.
(31) Davis, A. L.; Cai, Y.; Davies, A. P.; Lewis, J. R. H-1 and C-13 NMR assignments of some green tea polyphenols. *Magn. Reson. Chem.* **1996**, *34*, 887−890.
(32) Saijo, R.; Takeda, Y. HPLC analysis of catechins in various kinds of green teas produced in Japan and abroad. *J. Jpn. Soc. Food Sci. Technol.-Nippon Shokuhin Kagaku Kogaku Kaishi* **1999**, *46*, 138−147.
(33) Wang, L. F.; Kim, D. M.; Lee, C. Y. Effects of heat processing and storage on flavanols and sensory qualities of green tea beverage. *J. Agric. Food Chem.* **2000**, *48*, 4227−4232.
(34) Wang, H. F.; Helliwell, K. Epimerisation of catechins in green tea infusions. *Food Chem.* **2000**, *70*, 337−344.
(35) Vogels, J. T. W. E.; Tas, A. C.; Venekamp, J.; VanderGreef, J. Partial linear fit: A new NMR spectroscopy preprocessing tool for pattern recognition applications. *J. Chemom.* **1996**, *10*, 425−438.
(36) Defernez, M.; Colquhoun, I. J. Factors affecting the robustness of metabolite fingerprinting using H-1 NMR spectra. *Phytochemistry* **2003**, *62*, 1009−1017.
(37) Kim, Y. S.; Scotter, C. N. G.; Voyiagis, M.; Hall, M. Potential of NIR spectroscopy for discriminating geographical origin of green tea from Korea and Japan. *Foods Biotechnol.* **1997**, *6*, 74−78.
(38) Marcos, A.; Fisher, A.; Rea, G.; Hill, S. J. Preliminary study using trace element concentrations and a chemometrics approach to determine the geographical origin of tea. *J. Anal. At. Spectrom.* **1998**, *13*, 521−525.
(39) Fernandez-Caceres, P. L.; Martin, M. J.; Pablos, F.; Gonzalez, A. G. Differentiation of tea (Camellia sinensis) varieties and their geographical origin according to their metal content. *J. Agric. Food Chem.* **2001**, *49*, 4775−4779.
(40) Fernandez, P. L.; Pablos, F.; Martin, M. J.; Gonzalez, A. G. Study of catechin and xanthine tea profiles as geographical tracers. *J. Agric. Food Chem.* **2002**, *50*, 1833−1839.
(41) Lin, Y. S.; Tsai, Y. J.; Tsay, J. S.; Lin, J. K. Factors affecting the levels of tea polyphenols and caffeine in tea leaves. *J. Agric. Food Chem.* **2003**, *51*, 1864−1873.
(42) Goto, T.; Yoshida, Y.; Kiso, M.; Nagashima, H. Simultaneous analysis of individual catechins and caffeine in green tea. *J. Chromatogr. A* **1996**, *749*, 295−299.
(43) Aucamp, J. P.; Hara, Y.; Apostolides, Z. Simultaneous analysis of tea catechins, caffeine, gallic acid, theanine and ascorbic acid by micellar electrokinetic capillary chromatography. *J. Chromatogr. A* **2000**, *876*, 235−242.
(44) MaedaYamamoto, M.; Tahara, N.; Yamaguchi, Y.; Tsuji, K. Changes in dietary fiber contents in tea leaves. *J. Jpn. Soc. Food Sci.* **1996**, *43*, 1309−1313.

Received for review July 24, 2003. Revised manuscript received December 9, 2003. Accepted December 14, 2003. We thank the U.K. DEFRA Link program on Agro-Food Quality for funding (Project AFQ119).

JF034828R