Journal of Chromatography B, 878 (2010) 2733–2740

Contents lists available at ScienceDirect

Journal of Chromatography B

journal homepage: www.elsevier.com/locate/chromb

Prediction and interpretation of the antioxidant capacity of green tea from dissimilar chromatographic fingerprints

M. Dumarey, I. Smets, Y. Vander Heyden*

Vrije Universiteit Brussel (VUB), Department of Analytical Chemistry and Pharmaceutical Technology, Laarbeeklaan 103, 1090 Brussels, Belgium

ARTICLE INFO

Article history:
Received 14 December 2009
Accepted 14 August 2010
Available online 21 August 2010

Keywords:
Green tea
Fingerprints
Multivariate calibration
Dissimilar chromatographic systems
Orthogonal chromatographic systems
Orthogonal Projections to Latent Structures

ABSTRACT

Previously, multivariate calibration techniques have been successfully applied to model and predict the antioxidant activity of green tea from its chromatographic fingerprint. Since the selectivity differences between dissimilar chromatographic systems have already been valuably used in several applications, in this paper it is studied whether combining the complementary information contained in two dissimilar fingerprints can improve the predictive capacity of the multivariate calibration model. The simplest way of combining the data is concatenating both fingerprints for each sample. The resulting matrix can then be subjected to Orthogonal Projections to Latent Structures (O-PLS). Unfortunately, this approach resulted in a more complex model with a prediction error of about the average of the errors obtained with the individual fingerprints. Secondly, only the peaks with high loading and low orthogonal loading from both chromatograms were included in the O-PLS model. This resulted in a reduced complexity, but not in better predictions, probably due to a lack of complementarity of the information concerning the antioxidant capacity. Finally, the concatenated fingerprints were subjected to stepwise multiple linear regression (MLR) in order to build a model based on the variables most correlated with the antioxidant capacity. The obtained prediction error was lower than those of both previous approaches, but still higher than the error of the model based on a single analysis. This is probably again caused by a lack of complementarity in the variables. Nevertheless, it was advantageous to develop fingerprints on dissimilar system, because it enables to choose the most suited chromatographic profile to build a multivariate calibration model for the considered purpose. In contrast to what was expected, the study showed that the most simple (so the worst separated) fingerprints resulted in the best predictions. On the other hand, a more complex fingerprint in which more compounds are separated is still important to improve the interpretability of the model.

© 2010 Elsevier B.V. All rights reserved.

1. Introduction

Green tea, originating from *Camillia Sinensis*, is associated with many beneficial effects, e.g. protection against cancer and cardiovascular diseases [1–4]. The health effects would be caused by the flavonoids present in the tea, for instance, epigallocatechin gallate, (−)-epigallocatechin (EGC), (−)-epicatechin gallate and epicatechin. These compounds are able to scavenge free radicals [1–4] and are thus responsible for the antioxidant capacity of green tea.

The quality control and identification of herbal products, as green tea, is often based on their fingerprints, because their composition is too complex to identify and quantify each component individually [5]. A fingerprint of a herbal extract is an experimentally obtained pattern showing the pharmacological active and/or chemically characteristic components [5] and can, amongst others, be developed by chromatographic techniques as, for instance, high-performance liquid chromatography (HPLC), or by capillary electrophoresis [5,6]. The resulting chromatographic or electrophoretic profiles are then compared with the fingerprint of a standardized extract to verify the identity and the quality of the extract [7].

An important quality criterion of green tea, its antioxidant capacity, cannot be directly derived from the fingerprints. Most commonly, spectroscopic methods, as the Trolox-equivalent antioxidant capacity (TEAC) assay [8,9], or the 2,2-Diphenyl-1-pikrylhydrazyl (DPPH) test [10], are applied to quantitatively estimate the antioxidant capacity of green tea. Earlier, Massarts group [11] managed to model the antioxidant capacity as a function of spectral information, e.g. near infrared data. van Nederkassel et al. [12] succeeded in applying two different multivaritate calibration methods, i.e. partial least squares (PLS) [13] and Uninformative Variable Elimination (UVE)-PLS [14], to predict the antioxidant capacity of green tea extracts from their chromatographic fingerprints. Later, we explored several multivariate

* Corresponding author. Tel.: +32 2 477 47 34; fax: +32 2 477 47 35.
E-mail address: yvanvdh@vub.ac.be (Y. Vander Heyden).

1570-0232/$ – see front matter © 2010 Elsevier B.V. All rights reserved.
doi:10.1016/j.jchromb.2010.08.012

calibration techniques for the same purpose [15], which resulted in a preference for the Orthogonal Projections to Latent Structures (O-PLS) technique [16]. The use of fingerprints for determining the antioxidant capacity of green tea samples has two advantages compared to the TEAC assay: first valuable information for quality control is obtained and second the compounds responsible for antioxidant capacity can be indicated.

The aim of this paper is to evaluate whether the information contained in fingerprints, measured on two dissimilar chromatographic systems, can be included in the multivariate calibration model to improve its predictive capacity. Dissimilar systems exhibit considerable differences in selectivity and may therefore provide complementary information about a sample [17,18]. Two compounds not separated, for instance, with the first system potentially may be separated with the second, and vice versa. This might result in better predictions and/or interpretability of the model.

Green tea samples were analyzed on two dissimilar reversed-phase liquid chromatography systems in order to test the above hypothesis. The resulting chromatograms were aligned per method with correlation optimized warping (COW). Then principal components analysis (PCA) was applied to detect outliers, which should be removed and eventually the remaining samples were subjected to the Kennard and Stone algorithm to select a representative calibration set of tea samples. For each calibration sample, both dissimilar (aligned) fingerprints were concatenated resulting in one matrix, which was subjected to O-PLS. In order to reduce data complexity, it was also tried to select the most important variables from the latter matrix prior to O-PLS or multiple linear regression (MLR).

2. Theory

2.1. Trolox-equivalent antioxidant capacity assay

The TEAC assay measures the decolorization caused by the scavenging of the blue 2,2′-azino-bis-(3-ethylbenzothiazolin-6-sulfonic acid) (ABTS) radicals by antioxidizing compounds. This decolorization is determined spectrophotometrically at 729 nm and is a measure for the antioxidant capacity. In this study, the TEAC value reflects the scavenging capacity of a 1% (m/v) green tea extract expressed as the equivalent concentration (in mM) of Trolox, a water-soluble vitamin E analogue. More details about the TEAC assay can be found in Ref. [8].

2.2. Correlation optimized warping

COW aligns two chromatograms piecewise by maximizing the correlation between them [19]. First the chromatogram is divided in a number of segments. The analyst needs to define two parameters, i.e. the number of chromatographic points in each segment (N) and the slack parameter (t). The latter restricts the number of points that the segment end point can be moved. It should be large enough to allow a flexible warping, but it should also be kept small enough to avoid the matching of non-corresponding peaks. Warping is then achieved by moving the endpoint of each segment and compressing or stretching the segment maximizing the correlation with the corresponding segment in the target chromatogram, i.e. a representative chromatogram with which all other fingerprints are aligned.

2.3. Principal components analysis

PCA is a variable reduction technique, which enables to visualize multivariate data in a low-dimensional space [20,21]. This is achieved by calculating new latent variables, called principal components (PC's), which are linear combinations of the original variables orientated in the directions of the largest remaining variation. The PC's are mutually orthogonal and maximally $(n - 1)$ PC's can be constructed, where n is the number of original variables. However, in most cases almost all variation is explained by the first few PC's.

2.4. Kennard and Stone algorithm

This algorithm can be applied to select a representative subset of samples, which are uniformly distributed in the experimental space [22,23]. Often the first arbitrarily selected object is the center point, i.e. the point closest to the mean. Then the sample located furthest from this center point is selected. An alternative is that the two objects situated furthest from each other are selected first. The third sample included in the set, is the one furthest from the first two selected, etc. This procedure can be continued until the desired number of samples is selected.

2.5. Orthogonal Projections to Latent Structures

O-PLS is a modified version of ordinary partial least squares (PLS) [16]. It removes information, not correlated with the response. This is achieved by subtracting orthogonal components from the original data. The eventual O-PLS model may be built with one PLS-factor and will thus be much simpler than the corresponding PLS model. Another advantage is that the interpretation of the contribution of the original variables to the model and to the predicted response values is simpler [15].

2.6. Stepwise multivariate linear regression

In MLR a linear model (Eq. (1)) is built with given descriptive variables using the least squares method to minimize the residuals

$$y = b_0 + b_1x_1 + b_2x_2 + \cdots + b_nx_n, \quad (1)$$

where y is the quantitative property to predict (dependent variable), x_i an independent (descriptive) variable, b_0 the intercept and b_i the regression coefficient for x_i [23]. In the stepwise approach the most important variables are selected by a stepwise selection procedure, which combines the forward selection and backward elimination approaches [21].

3. Experimental

3.1. Tea extract

The green tea samples were purchased as dried tea leaves, as gunpowder or ground in teabags from supermarkets or specialized tea shops or were received as gifts. The preparation of the tea extracts started by grinding (Jank & Kunkel Type 10A, Staufe, Germany) 0.4 g dry tea three times for 10 s followed by the sieving of the resulting powder through a 500 μm sieve (Retsch, Haan, Germany). Then, 0.1 g of the sieved tea was infused with 20 ml of initially boiling milli-Q water (Millipore, Milford, USA) for 7 min in a dark environment. The infusion was sieved through a 100 μm sieve (Retsch) and filtered through a 0.2 μm membrane filter (Pall Gelman Laboratory, Karlstein/Main, Germany). Finally, the volume of the extract was adjusted to 20.0 ml with water. The tea extract was stored in a dark recipient in the refrigerator until analyses were started (24 h maximum).

3.2. TEAC assay

A solution of ABTS radicals was prepared by storing an aqueous solution containing 7 mM ABTS (Sigma–Aldrich, Steinheim, Germany) and 2.45 mM $K_2S_2O_8$ (Merck, Darmstadt, Germany) in a dark

room at room temperature for 12 h. At the start of the TEAC analyses, the latter solution was diluted until it exhibited an absorbance between 1.5 and 1.7 at 729 nm.

An antioxidant calibration line was constructed daily, with 20, 25, 30, 35 and 40 μM aqueous Trolox standards, diluted from a 5 mM Trolox (Sigma–Aldrich) solution in ethanol (Fisher Scientific, Leicestershire, UK). 0.3 ml of the least concentrated Trolox standard was added to 1.0 ml of the diluted ABTS solution and mixed for 10 s in the cuvet with a micropipette. The absorbance (A) was measured, before and 60 s after the addition of Trolox with a UV-2101PC-spectrophotometer (Shimadzu, Tokyo, Japan) at 729 nm and the difference between both (ΔA) was calculated. This procedure was repeated three times. The other Trolox standards, with increasing concentration, were treated analogueously. The average ΔA of the four replicate measurements was plotted versus the concentration of the standard to construct the calibration line.

To determine the antioxidant (AO) capacity of a green tea sample, 0.3 ml of a 200 times diluted tea extract (prepared as in Section 3.1) was added to 1.0 ml diluted ABTS solution. Then the same procedure as for the measurement of the Trolox standards was performed. When introducing the average ΔA of a sample in the equation of the calibration line, the equivalent Trolox concentration could be estimated. This concentration is equivalent to 0.3 ml of a diluted tea extract solution. In order to estimate the real TEAC value, i.e. the Trolox concentration exhibiting an antioxidant capacity equivalent to that of a 1% (m/v) green tea extract, the obtained value should be multiplied by 20/0.3 to correct for the fact that only 0.3 ml was measured and by 200 to correct for the dilution. Finally, the resulting number should be multiplied by 1 g and divided by 5 times the effectively weighted mass of ground tea to obtain the equivalent for a 1% extract.

3.3. Chromatographic analyses

As stated in the introduction the chromatographic analyses were performed on two dissimilar columns: the Chromolith Performance RP-18 (100 mm × 4.6 mm) (Merck) (method 1) and the Waters XTerra RP-18 (100 mm × 4.6 mm) (Waters, Milford, USA) (method 2). For both methods the applied high-performance liquid chromatography (HPLC) system consisted of an L-7100 pump, an L-7612 solvent degasser, an L-7250 autosampler, an L-7350 column oven, an L-7400 UV-detector and a D-7000 interface, all from Merck-Hitachi (Tokyo, Japan). This system was operated with LaChrom D-7000 HPLC Manager Software (Merck-Hitachi). The column oven temperature was 30 °C and the detection wavelength 280 nm. The injection volume and the data sampling rate were 15 μl and 200 ms, respectively. The tea samples were analyzed with gradient HPLC, where the organic phase was ACN (Fisher Scientific) with 0.05% trifluoroacetic acid (TFA, Sigma–Aldrich, St-Louis, USA) and the aqueous phase milli-Q water also with 0.05% TFA. The return to the initial conditions took 3 min and finally the column was reconditioned for 6 min.

3.3.1. Method 1

A Chromolith Guard RP-18 column (5 mm × 4.6 mm) (Merck) was placed before the Chromolith Performance to protect the analytical column against contamination. A flow rate of 2 ml/min was applied. The organic fraction increased from 2% to 26% within 10 min and then remained constant at 26% for 1 min.

3.3.2. Method 2

On the Waters XTerra column a flow rate of 1 ml/min was applied. The organic fraction increased from 2% to 26% within 10 min and then remained constant at 26% for 3 min.

3.4. Software

All data processing methods were performed with subroutines developed under Matlab 7.0.1 software (Mathworks, Natick, USA).

4. Results and discussion

The first chromatographic system applied in this study was the one developed in Ref. [12]. Since the stationary phase has a major influence on selectivity [17,24], a dissimilar system was chosen by performing the same analysis of one representative tea sample on several columns, selected based on the color maps presented in [25]. However, the zirconium based columns (Zirchrom PS and Zirchrom PBD), which are most dissimilar to the Chromolith Performance according to the color map, did not produce good separations of the tea extracts. Very few peaks were observed in the resulting chromatograms. The other selected dissimilar columns (Zorbax Extend C18, Betasil Phenyl Hexyl and Zorbax Eclipse XCB C8) produced fingerprints very similar to those obtained with Chromolith Performance. Since none of the dissimilar columns were suitable for our aim, we also tested the Waters XTerra RP 18 column. In previous impurity profiling applications [26] this column resulted in highly efficient separations dissimilar to those obtained on the Zorbax Extend C18. Since the latter column produced a similar separation as the Chromolith Performance, the Waters XTerra RP 18 might also be dissimilar to the Chromolith Performance. The fingerprint on the Waters XTerra RP18 column indeed resulted in a profile visually different from that on the Chromolith Performance. In general, when several columns result in more or less equally suitable dissimilar profiles, one may select the one resulting in a separation with the lowest correlation with the first chromatogram [7].

After the selection step, replicate fingerprints were developed on each of the two dissimilar systems for 63 green tea extracts. The fingerprints were aligned per chromatographic method [19] with COW in order to correct for peak shifts caused by instrumental variability and small changes in mobile phase composition. Profiles, visually very different from the majority of the fingerprints and/or outlying in the PCA score plots [20] (not shown), should be removed. Then the average of the two replicate chromatograms was calculated to obtain one profile per sample. A calibration set of 40 samples, which should be representative for the studied tea samples and which was used to build the multivariate calibration model with O-PLS, was selected with the Kennard & Stone algorithm [21,22].

In this study, the chromatograms obtained with the two methods were first treated individually with O-PLS as described in [15]. Then, for every sample both profiles, were concatenated in such a way that they form one row in the matrix. This is the simplest method to merge the information from both fingerprints dissimilar analyses prior to O-PLS. However, this approach is expected to increase the complexity of the data. Therefore, in a second instance, only the peaks with high loadings and low orthogonal loadings (see Section 4.6) were included in the matrix. Finally, stepwise MLR was applied on the matrix containing the concatenated chromatograms in order to build a model with a limited number of measured variables, which are most correlated with the antioxidant capacity. For all approaches based on the combined data, it was verified whether the resulting model indeed exhibited better predictive capacities than those based on the individual fingerprints. The prediction error was estimated with an external test set, i.e. the remainder of the data set without the 40 calibration samples.

4.1. TEAC results

The measured TEAC values of the 63 green tea samples ranged from 160.9 to 5395. However, it was noticed that only three samples

exhibited extreme low values. When excluding these, the lowest TEAC value is 1959, which indicates the outlying aspect of the samples with low antioxidant capacities (which further is confirmed from the fingerprint evaluation).

We selected three samples with a low (1959), intermediate (3279) and high TEAC value (4240) respectively in order to calculate, as in Ref. [12], the pooled standard deviation of the TEAC analysis, i.e. an estimate of the experimental variability. For each sample four replicate measurements were performed resulting in a pooled standard deviation of 26.5. A better precision than in Ref. [12] (pooled standard deviation = 143) was thus achieved.

4.2. Warping the chromatograms

Peak shifts occurred for both analyses and therefore COW was applied to align the individual chromatograms with a target chromatogram. The latter was the fingerprint exhibiting the highest correlation with all other on a given column. N, expressed as the length of the segment in this study, was set to 150 data points and t was varied between 2 and 60 data points. For some fingerprints too low t values did not allow obtaining a high correlation with the target chromatogram and some peaks were not warped, while for others too high values of t caused the matching of non-corresponding peaks. The suitability of a given t was evaluated visually for each chromatogram individually. The COW procedure resulted in well-aligned chromatograms (Figs. 1 and 2), which exhibited high correlation with the target chromatogram (mostly above 0.90).

4.3. Comparison of dissimilar fingerprints

The chromatographic profiles of the green tea samples on the two dissimilar systems (Figs. 1b and 2b) showed some selectivity differences. The first obvious difference is that caffeine was partially co-eluting with another compound applying method 1, while this was not observed with method 2. A second noticeable difference is that more peaks appeared between the caffeine and EGC peaks in the second fingerprint. Also in general, more peaks were resolved with method 2.

4.4. Outlier detection

When studying the individual fingerprints, it was observed that the samples with extreme low TEAC values also exhibited a very different chromatographic profile compared to the other tea samples on both systems (not shown). Probably, these three samples were another type of tea and therefore they were definitively excluded from the data sets used for modeling and prediction.

The scores of the 60 remaining fingerprints (two replicates) were distributed quite homogenously, both in the PC1–PC2 and in the PC1–PC3 score plots. No further outlying objects were observed for the two dissimilar analyses. It can thus be concluded that all remaining fingerprints may be subjected to the multivariate calibration.

4.5. Multivariate calibration

Prior to the modeling, the average of the two replicate chromatograms for each tea extract was calculated resulting in one fingerprint per sample and per analysis method. Then 40 green tea samples were selected with the Kennard and Stone algorithm, based on their fingerprints (obtained with method 1), in order to obtain a representative calibration set. The 20 remaining samples were used as external test set. The last pretreatment step consisted of removing uninformative baseline points from the fingerprints. For method 1 the chromatographic profile between 1.00 and 10.00 min was retained, while for method 2 the data from 1.66 min till the end (13 min) were used.

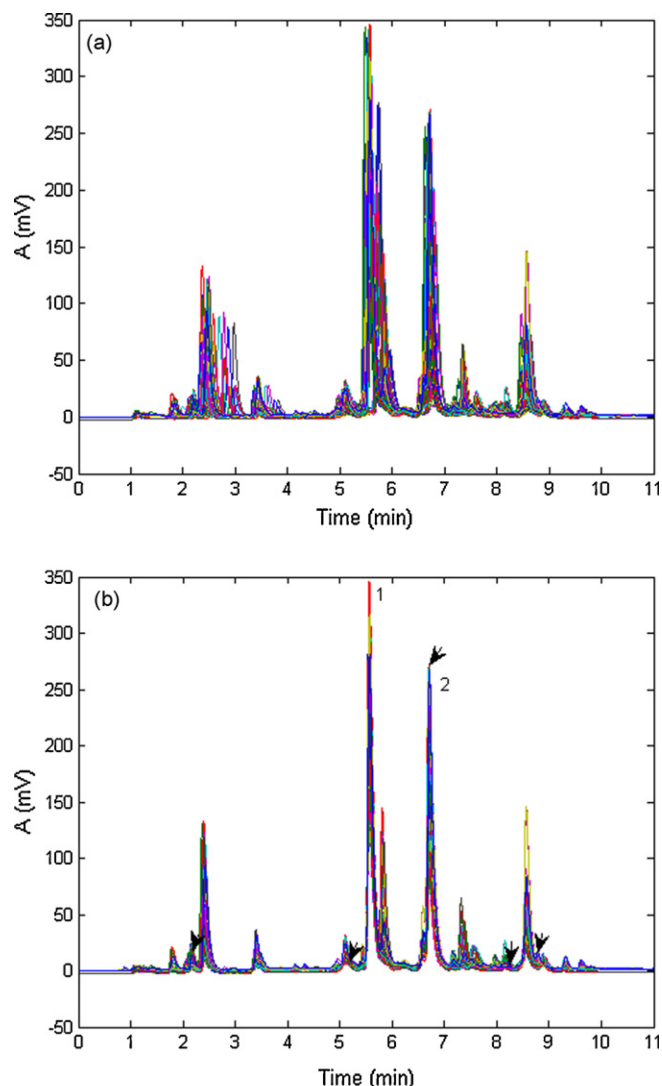

Fig. 1. The (a) raw and (b) warped fingerprints of 60 green tea extracts measured on the Chromolith Performance column (method 1); 1 caffeine and 2 EGC. The arrows indicate the variables selected with stepwise MLR.

When applying O-PLS on the pretreated data of method 1, three orthogonal components were removed based on leave-one-out cross validation (LOO-CV) [13]. This resulted in worse predictions (Table 1) than in Ref. [12], what can be explained by the fact that the tea samples included in this study were more diverse. For instance, only gunpowder tea was present in the dataset of Ref. [12], while in this study also ground tea in bags was considered. Moreover,

Table 1
The complexity, the root mean squared error of cross validation (RMSCV) obtained with LOO-CV, the root mean squared error for the calibration set (RMS) and the root mean squared error of prediction for the external validation set (RMSEP) for the models obtained with the individual fingerprints, the combined fingerprints and the fingerprints from Ref. [12].

	PLS factors	Orthogonal factors	RMSCV	RMS	RMSEP
Ref. [12]	8	/	159	81	174
Method 1	1	3	370	285	262
Method 2	1	3	499	386	332
Merged data	1	6	333	189	300
Selected data	1	4	413	294	312
Stepwise MLR	13 variables		129	83	286

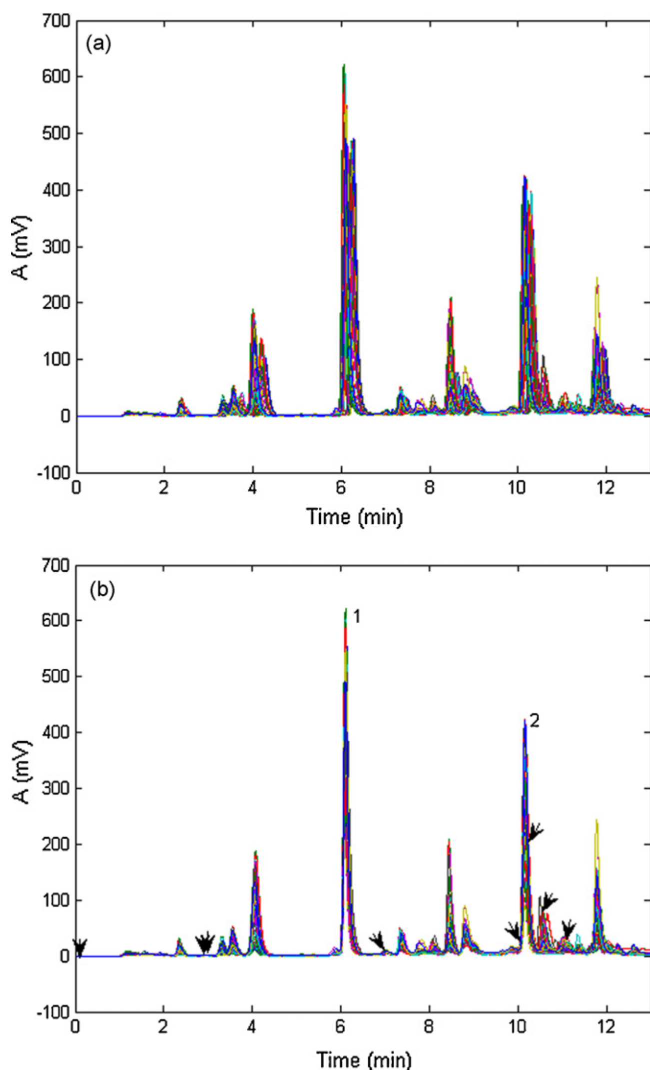

Fig. 2. The (a) raw and (b) warped fingerprints of 60 green tea extracts measured on (a) on the XTerra column (method 2); 1 caffeine and 2 EGC. The arrows indicate the variables selected with stepwise MLR.

we also included green tea with extra aromas as, for instance, earl grey. This might cause the appearance of additional peaks in the chromatograms. Another observation possibly causing worse predictions was that the peaks are broader and as a consequence less high than in Ref. [12]. Column ageing was probably responsible for this observation.

When a one-factor O-PLS model was built based on the pretreated data from method 2, also three orthogonal components were removed. Although the tea extracts seemed better separated with method 2, the prediction errors (Table 1) were higher than with method 1. This can be explained by the fact that it is more difficult to model the larger variation in the fingerprints resulting from method 2. Another possible explanation is that co-eluting compounds in method 1 probably have both similar structures and similar antioxidant capacities. Therefore their separation was less useful for modeling and only complicated the modeling. These results also confirm the findings in Ref. [12], in which longer chromatograms with better separated peaks resulted in worse predictions than the five times shorter ones obtained on the same column.

When concatenating the pretreated data of both analyses in one matrix and applying O-PLS, six orthogonal components were subtracted. This resulted in a higher root mean squared error of

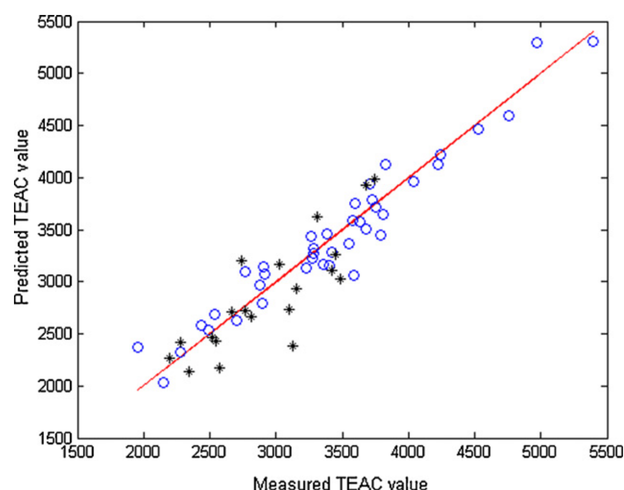

Fig. 3. The measured TEAC values of 60 green tea extracts versus their values predicted with an O-PLS model based on the combined fingerprints; (○) calibration samples and (*) test set samples.

prediction for the external test set (RMSEP) than for method 1, but lower than for method 2. It can be noticed that in fact the RMSEP obtained with the combined data was approaching the mean RMSEP of both individual analyses. The combination of the complementary information did thus not result in better predictions due to an increased complexity of the data, which is more difficult to model. This increased complexity was also observed from the fact that more orthogonal components needed to be removed. In Table 1 it also can be noticed that the RMSEP was considerably higher than the root mean squared error of prediction of the calibration set (RMS). This might indicate over fitting of the model. However, in models where a lower number of orthogonal components was subtracted, no reduced RMSEP was obtained. In the plot which shows the predicted (based on the model from the combined fingerprints) versus the measured TEAC values (Fig. 3), also no indication of over fitting was observed.

We also tried to reduce the complexity of the merged data in order to decrease the number of orthogonal components of the calibration model, to improve the predictive capacity and to evaluate the complementary information from both fingerprints. Therefore the chromatographic regions with high loadings and low orthogonal loadings from both dissimilar chromatograms were selected to build the O-PLS model. As described in Ref. [15], these chromatographic regions are highly correlated with the antioxidant capacity and might as a consequence represent anti-oxidizing compounds. For method 1, these were the regions around the peaks at 3.5 and 8.6 min and around the EGC peak. For method 2 four regions were selected: the region between 3.5 and 4.3 min containing two peaks, the region between 6.8 and 8.3 min containing a number of smaller peaks and the regions around the EGC peak and the peak at about 11.9 min. Although the number of orthogonal components could be reduced to four, the model based on these selected regions resulted even in a slightly higher prediction error (RMSEP = 312) (Table 1). A possible explanation for this observation is that probably only 2 compounds, i.e. EGC and the one represented by the last, big peak, determine most of the antioxidant capacity, while the lower concentrated antioxidizing components only have a minor influence. Since the main difference between both dissimilar chromatograms lays in the appearance of small peaks, their information concerning antioxidant capacity was not complementary and the merging of the data thus of limited use. However, in applications where the property of interest is determined by many, rather equally important compounds, not separable in a single chromatographic analysis, the suggested approach might be beneficial.

From the latter observations it can also be deduced that the worse predictive capacity of the model based on method 2 is probably caused by a higher noise level or by the co-elution of either EGC or the other main antioxidizing compound, with compounds with no or little antioxidant capacity.

Another way to reduce data complexity is selecting stepwise the variables most correlated with the antioxidant capacity and therefore stepwise MLR was applied on the merged data. This resulted in an RMSEP value of 286, which was lower than the prediction error of the O-PLS model based on the merged data, but higher than that of the model based on method 1. Moreover, the variable selection (indicated with arrows in Figs. 1b and 2b) seemed inadequate, because also a few baseline points and points of low signal intensity are included instead of the more expected peak maxima. This implied that also noise is modeled, which decreases the reliability of the model and explains probably the large difference between RMS and RMSEP. For stepwise MLR, in contrast to O-PLS, it was advantageous to include data from both dissimilar analyses. The obtained RMSEP value was considerably lower than those of the models based on the individual analyses (357 and 628). The stepwise MLR procedure also confirmed the importance of EGC, since it was the only peak maximum included in the model.

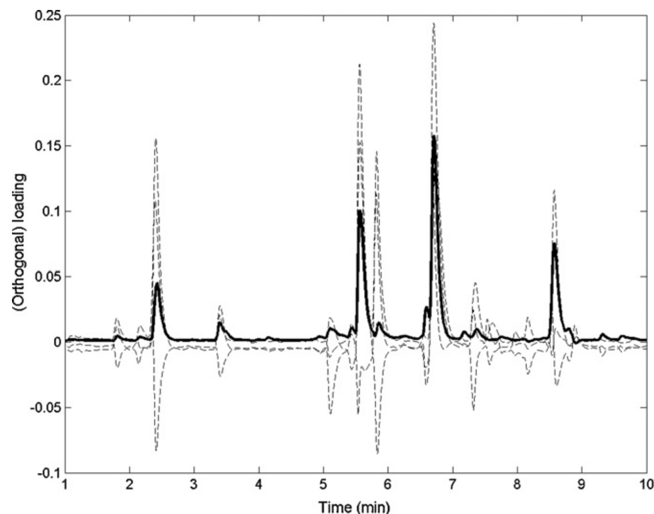

Fig. 4. The loadings (—) and orthogonal loadings (- - -) of the O-PLS model based on the 60 fingerprints obtained with method 1.

4.6. Interpretation of the models

When studying the loadings of the O-PLS model based on the method 1 fingerprints (Fig. 4), the EGC peak, an anti-oxidizing compound, is seen to have the largest loading. The peak eluting at about 2.4 min had a loading proportional to its peak height, while the peak at about 8.6 min had a relatively higher loading. The caffeine peak, with a higher peak in the chromatogram, exhibited a relatively lower loading. This is logic, since caffeine does not possess antioxidant capacity [12]. It could also be observed that the peak partially co-eluting with caffeine had very low loadings.

In order to interpret the model further also the orthogonal loadings, i.e. the loadings of the variables on the removed, orthogonal components, were studied (Fig. 3). The first orthogonal loading, confirmed the earlier findings: the caffeine and the partially co-eluting compound showed high variation not correlated with the antioxidant capacity, while the opposite was true for the EGC and the compound eluting around 8.6 min. The peak eluting at about 2.4 min had also a relatively high orthogonal loading. When studying the loading plots in more detail, it was seen that the loadings of the two small peaks eluting about 1.8 and 2.2 min were rather low, while their orthogonal loadings were relatively higher. The same was observed for the two peaks eluting just after EGC. These orthogonal loadings indicate the presence of compounds with limited or no antioxidant activity. The fact that the loadings were not zero for these compounds might be caused by co-elution of a compound with anti-oxidizing compounds. Another, more probable explanation, is that the concentration of the compounds were somewhat correlated with the concentration of some anti-oxidizing compounds present in the tea. As a consequence the concentration of the compound was also somewhat correlated with the antioxidant capacity and thus this information is partly included in the model.

When looking at the equivalent plots for the O-PLS model based on the “method 2” fingerprints (Fig. 5), similar observations as above are made for caffeine and EGC. The peak at 12 min had a high loading and a relatively small orthogonal loading, which may indicate an anti-oxidizing compound. The highest peak eluting between caffeine and EGC, on the contrary, had a very low loading and a high orthogonal loading and therefore originates from a compound without antioxidant properties. The three small peaks eluting just before the highest peak between caffeine and EGC and

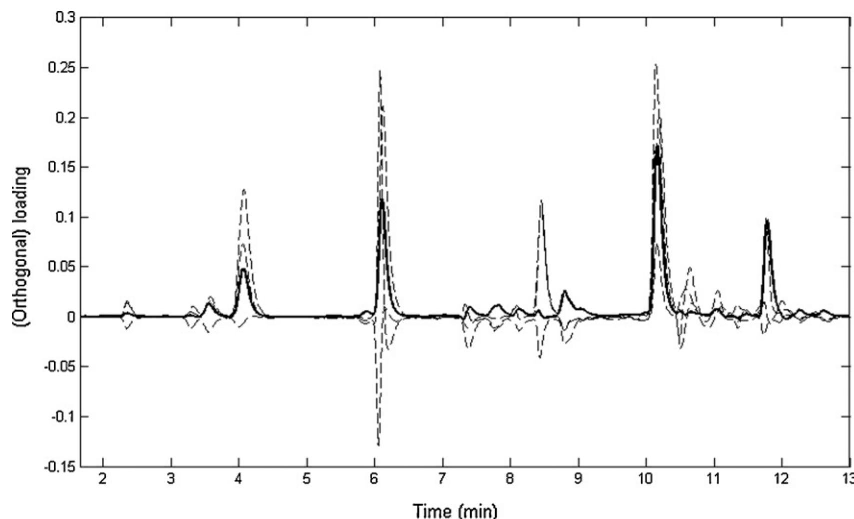

Fig. 5. The loadings (—) and orthogonal loadings (- - -) of the O-PLS model based on the 60 fingerprints obtained with method 2.

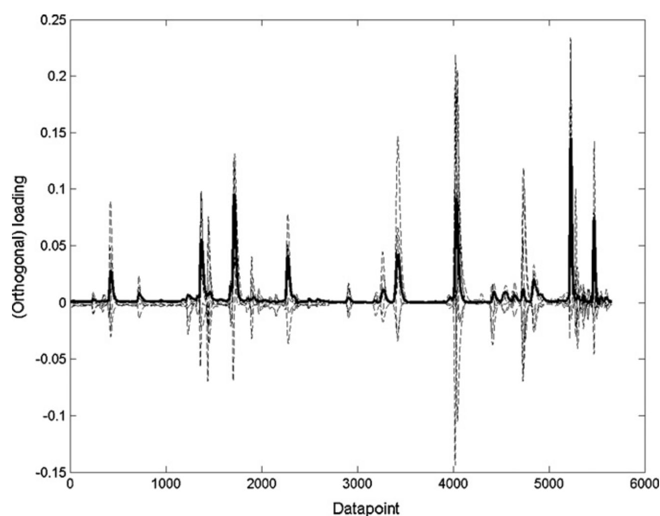

Fig. 6. The loadings (—) and orthogonal loadings (- - -) of the O-PLS model based on the combined fingerprints.

the peak eluting just behind exhibited loadings proportional to their peak height and low orthogonal loadings, which indicates the presence of at least four potentially anti-oxidizing compounds. The small peaks at about 2.4 and 3.3 min showed small loadings and relatively larger orthogonal loadings, which implied that the compounds had no antioxidant capacity. The opposite was true for the peaks eluting at about 3.6 and 4.1 min. The small peaks, appearing between EGC and the peak at 12.0 min, exhibited small loadings and relatively higher orthogonal loadings. The opposite situation is observed for the two little peaks eluting after 12 min. As seen from above, the more peaks are present in the fingerprint the more peaks with or without probable antioxidant capacity can be indicated.

When comparing the (orthogonal) loadings for methods 1 and 2, caffeine, EGC and the last eluting peak (at 8.6 min for method 1 and at 12.0 min for method 2) were behaving similarly in both fingerprints. It was more difficult to compare the loadings of other peaks, because it is impossible to match the peaks without identifying them. An obvious difference between the (orthogonal) loading plots of both analyses was that for method 2 more (orthogonal) loadings, which can easily be linked with a peak of the fingerprint, are seen. The fact that more peaks are present in the fingerprint and the corresponding (orthogonal) loading plot results thus in an improved (or more detailed) interpretability of the model.

The loadings and orthogonal loadings of the O-PLS model based on the merged fingerprints are shown in Fig. 6. The loadings based on the second fingerprint are higher, which is probably caused by the fact that higher peaks are obtained with method 2 due to the higher efficiency of the XTerra RP 18 column. Also the orthogonal loadings are relatively much higher, which confirms the noisy profile of method 2. Another observation is that the loading plot is in fact the merged loading plots of both individual analyses. It can thus be concluded that the merging of two dissimilar fingerprints does not result in a better interpretation of the model.

The variable selection in stepwise MLR also could not improve interpretability, because, as mentioned in the previous section, the selected variables have a low signal and in some cases even are baseline points.

5. Conclusions

The combination of fingerprints, obtained on two dissimilar chromatographic systems, to build multivariate calibration models did neither result in better predictions of the antioxidant capacity of green tea extracts nor in a better interpretability of the model. This might on the one hand be caused by the increased complexity of the data, when the two chromatographic profiles are merged. As a consequence more orthogonal components are removed in the O-PLS model (higher complexity) and averaged prediction errors (relative to the individual models) are obtained. On the other hand, it is also possible that no or only little useful complementary information, concerning antioxidant capacity, was present in both dissimilar fingerprints. This assumption was confirmed by the fact that only few antioxidizing compounds are present in high concentration in green tea and they will make the largest contribution to the model. The lower concentrated compounds are then less important for modeling and moreover a higher number of contributing peaks increases the data complexity resulting in worse models.

However, in applications where many active compounds, occasionally not separated from inactive compounds in a single chromatographic analysis, are equally contributing to the property of interest, the merging of dissimilar profiles might be advantageous. The suggested approach remains thus an interesting research topic, which is to be explored further.

For all applications, measuring fingerprints on dissimilar systems is beneficial to select the most appropriate system for further multivariate calibration. In this paper and in Ref. [12], it was observed that the best predictions are achieved with the least complex fingerprints, i.e. the chromatogram with the lowest number of peaks and thus most co-elution. On the contrary, when the goal of the analyst is to indicate peaks (compounds) with antioxidant properties, the more complex fingerprint is preferred.

Acknowledgements

Melanie Dumarey acknowledges the Institute for the Promotion of Innovation through Science and Technology in Flanders (IWT-Vlaanderen) for the funding of her PhD project.

References

[1] C.J. Dufresne, E.R. Farnworth, J. Nutr. Biochem. 12 (2001) 404.
[2] R.L. Pastore, P. Fratellone, Explore Diet Nutr. 2 (2006) 531.
[3] H. Wang, G.J. Provan, K. Helliwel, Trends Food Sci. Technol. 11 (2000) 152.
[4] N. Khan, H. Mukhtar, Life Sci. 81 (2007) 519.
[5] Y.Z. Liang, P. Xie, K. Chan, J. Chromatogr. B 812 (2004) 53.
[6] M. Ganzera, Electrophoresis 29 (2008) 3489.
[7] Y. Vander Heyden, LC–GC Europe 21 (2008) 438.
[8] R. Re, N. Pellegrini, A. Proteggente, A. Pannala, M. Yang, C. Rice-Evans, Free Radic. Biol. Med. 26 (1999) 1231.
[9] N.J. Miller, C.A. Rice-Evans, Redox Report 2 (1996) 171.
[10] I. Koleva, T. van Beek, J. Linssen, A. de Groot, L. Evstatieva, Phytochem. Anal. 13 (2002) 8.
[11] M.H. Zhang, J. Luypaert, J.A. Fernandez Pierna, Q.S. Xu, D.L. Massart, Talanta 62 (2004) 25.
[12] A.M. van Nederkassel, M. Daszykowski, D.L. Massart, Y. Vander Heyden, J. Chromatogr. A 1096 (2005) 177.
[13] B.G.M. Vandeginste, D.L. Massart, L.M.C. Buydens, S. De Jong, P.J. Lewi, J. Smeyers-Verbeke, Handbook of Chemometrics and Qualimetrics: Part B, Elsevier, Amsterdam, 1998, p. 238.
[14] V. Centner, D.L. Massart, O.E. de Noord, S. de Jong, B.G.M. Vandeginste, C. Sterna, Anal. Chem. 68 (1996) 3851.
[15] M. Dumarey, A.M. van Nederkassel, E. Deconinck, Y. Vander Heyden, J. Chromatogr. A 1192 (2008) 81.
[16] J. Trygg, S. Wold, J. Chemometr. 16 (2002) 119.
[17] E. Van Gyseghem, S. Van Hemelryck, M. Daszykowski, F. Questier, D.L. Massart, Y. Vander Heyden, J. Chromatogr. A 988 (2003) 77.
[18] J. Pellett, P. Lukulay, Y. Mao, W. Bowen, R. Reed, M.R. Ma, C. Munger, J.W. Dolan, L. Wrisley, K. Medwid, N.P. Toltl, C.C. Chan, M. Skibic, K. Biswas, K.A. Wells, L.R. Snyder, J. Chromatogr. A 1101 (2006) 122.
[19] N.P. Vest Nielsen, J.M. Carstensen, J. Smedsgaard, J. Chromatogr. A 805 (1998) 17.
[20] D.L. Massart, Y. Vander Heyden, LC–GC Europe 17 (2004) 586.
[21] D.L. Massart, B.G.M. Vandeginste, L.M.C. Buydens, S. De Jong, P.J. Lewi, J. Smeyers-Verbeke, Handbook of Chemometrics and Qualimetrics: Part A, Elsevier, Amsterdam, 1997, pp. 263, 519.
[22] R.W. Kennard, L.A. Stone, Technometrics 11 (1969) 137.

[23] M. Daszykowski, B. Walczak, D.L. Massart, Anal. Chim. Acta 468 (2002) 91.
[24] M. Dumarey, Y. Vander Heyden, in: S. Ahuja, M.I. Jimidar (Eds.), Capillary Electrophoresis Methods for Pharmaceutical Analysis, Separation Science and Technology, vol. 9, Elsevier, New York, 2008, p. 425.
[25] E. van Gyseghem, B. Dejaegher, R. Put, P. Forlay-Frick, A. Elkihel, M. Daszykowski, K. Héberger, D.L. Massart, Y. Vander Heyden, J. Pharm. Biomed. Anal. 41 (2006) 141.
[26] E. Van Gyseghem, M. Jimidar, R. Sneyers, D. Redlich, E. Verhoeven, D.L. Massart, Y. Vander Heyden, J. Chromatogr. A 1074 (2005) 117.