Analytica Chimica Acta 649 (2009) 24–32

Contents lists available at ScienceDirect

Analytica Chimica Acta

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

Potential antioxidant compounds in *Mallotus* species fingerprints. Part I: Indication, using linear multivariate calibration techniques

C. Tistaert[a], B. Dejaegher[a], N. Nguyen Hoai[a,b], G. Chataigné[c], C. Rivière[c], V. Nguyen Thi Hong[b,c], M. Chau Van[b], J. Quetin-Leclercq[c], Y. Vander Heyden[a,*]

[a] *Department of Analytical Chemistry and Pharmaceutical Technology, Vrije Universiteit Brussel - VUB, Laarbeeklaan 103, B-1090 Brussels, Belgium*
[b] *Institute of Natural Products Chemistry, Vietnamese Academy of Science and Technology, 18 Hoang Quoc Viet Road, Hanoi, Vietnam*
[c] *Analytical Chemistry, Drug Analysis and Pharmacognosy Unit, Louvain Drug Research Institute, Université catholique de Louvain, Avenue E. Mounier 72, B-1200 Brussels, Belgium*

ARTICLE INFO

Article history:
Received 26 March 2009
Received in revised form 17 June 2009
Accepted 8 July 2009
Available online 23 July 2009

Keywords:
Fingerprints
Antioxidant activity
Multivariate calibration
Indication of peaks

ABSTRACT

Some *Mallotus* species are used in traditional medicine in Vietnam and China. Some also show interesting activities, such as antioxidant and cytotoxic ones. Combining fingerprint technology with data-handling techniques allows indicating the peaks potentially responsible for given activities. In this study it is aspired to indicate from chromatographic fingerprints the peaks potentially responsible for the antioxidant activity of several *Mallotus* species. Relevant information was extracted using linear multivariate calibration techniques, both before and after alignment of the fingerprints with correlation optimized warping (COW). From the studied techniques, Stepwise Multiple Linear Regression is least recommended as it made an inadequate variable selection. Principal Component Regression theoretically can take largely varying variables uncorrelated to the antioxidant activity into account. However, in practice in the actual case study this problem was limited. These problems in principle do not occur using Partial Least Squares (PLS) models. Of the tested PLS methods, Orthogonal Projections to Latent Structures was preferred because of its simplicity, reproducibility, reduced model complexity and improved interpretability of the regression coefficients, yielding a clearer view on the individual contribution of the compounds. Furthermore, reducing analysis times from 60 min to 35 and 22.5 min resulted in the same main compounds, indicated responsible for the antioxidant activity. Models built after alignment by COW did not result in additional information.

1. Introduction

Traditional medicines (TM), including plant-, animal- and mineral-based products, are used to cover a large part of the primary health care needs in Asia, Africa and Latin America. Also in industrialized countries, adaptations of TM, defined as complementary or alternative medicines (CAM), are gaining importance. The main problem is that the quality of the TM is not always sufficiently evaluated. To ensure the patients safety it is extremely important to use medicines which are identified and of which the quality is assessed [1–3].

Although interesting, identification of just a few compounds hardly describes the complex nature of herbal medicines and ignores synergic interactions between the compounds. Moreover, the concentrations of the herbal constituents may vary significantly depending on the harvest season, the cultivation conditions and the drying processes, making it difficult to determine and isolate compounds of interest [4,5]. Therefore, quality control of TM in general, by assaying just a few compounds, is unreliable.

Recently, the World Health Organization (WHO) has introduced and accepted chromatographic fingerprint techniques as a strategy for the assessment of herbal medicines [2,6]. A fingerprint obtained by, for instance, High-Performance Liquid Chromatography (HPLC) characterizes the composition of the herbal sample. It can be used to evaluate the authenticity and stability of herbal samples [7–12]. Nowadays, the combination of (hyphenated) chromatographic instruments and chemometrical approaches for data (pre-)treatment allows a fast investigation of the herbal samples [13–17].

Some *Mallotus* species, belonging to the family of the *Euphorbiaceae*, are used in TM in Vietnam and China. The roots, stem barks, leaves and fruits are used for the treatment of chronic hepatitis and enteritis since hundreds of years [18] and provide a broad basis for researchers looking for new pharmaceutical active compounds. Many studies are performed about the chemical components of

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

0003-2670/$ – see front matter © 2009 Elsevier B.V. All rights reserved.
doi:10.1016/j.aca.2009.07.020

given *Mallotus* species and several pharmacologically active constituents were determined [19–24].

In this study, it is aspired to indicate and identify peaks potentially responsible for the antioxidant activity of some *Mallotus* species. The antioxidant activity of the herbal extracts was determined with a 1,1-diphenyl-2-picrylhydrazyl (DPPH) radical scavenging activity test [25]. Combining this information with the chromatographic fingerprints from the corresponding herbal samples allows constructing a multivariate regression model [26]. In the actual study, several linear multivariate calibration techniques were evaluated and compared, i.e. Stepwise Multiple Linear Regression [27], Principal Component Regression [28–30], Partial Least Squares [28–30], Uninformative Variables Elimination-Partial Least Squares [31] and Orthogonal Projections to Latent Structures [32]. Since alignment is strongly recommended to correct for peak shifts caused by variations in mobile phase composition, column ageing and instrument instability, results with and without alignment were compared. Several aligning techniques are described in the literature, such as dynamic time warping [33], parametric time warping [34] and fuzzy warping [35]. In our application, correlation optimized warping [33,34,36] was applied on the analytical signals.

The emphasis of this paper lies on the interpretation of the regression coefficients related to the different models in order to indicate peaks potentially responsible for the antioxidant activity of the *Mallotus* samples. In a next paper (Part II), the results of LC-MS analyses on the herbal samples will be discussed in combination with the indicated potential antioxidant compounds.

2. Theory

2.1. Data preprocessing

Chromatographic fingerprints can be organized in an $n \times p$ data matrix **X**, where the n objects (herbal samples) constitute the rows and the p variables (measuring time points) the columns. The results of chemometric data treatment are influenced by the applied preprocessing. In this study, different methods to pretreat the data are applied and compared, i.e. column centering, normalization, standard normal variate and alignment of the data.

Useful information resides in the between-sample variation of the variables, not in their absolute levels. To remove the level differences, column centering is a generally applied preprocessing technique. By removing the column mean from each corresponding value, every centered variable has a mean of zero. Normalization of the signal is also commonly applied. Normalization removes undesired effects due to unequal amounts of injected samples. It divides each row, corresponding to a fingerprint, by its norm. Finally, standard normal variate (SNV) transformation, used to remove slope variation, was also examined. Hereby each row is corrected individually by row centering followed by row scaling, meaning that the row mean-corrected fingerprints are scaled by the standard deviation calculated for each fingerprint individually [30].

Alignment or warping has an important place in the preprocessing of fingerprints. Along the time axis of chromatograms, peak shifts occur caused by variations in mobile phase composition, column ageing and instrument instability. Warping corrects for these shifts, aligning corresponding peaks. In this study, correlation optimized warping (COW), was applied [33,34,36]. COW aligns two signals by means of piecewise linear stretching and compression of the chromatogram to match it as good as possible with a target chromatogram. At the beginning of the procedure, both signals, the profile to be aligned (P) and the target profile (T) are divided into a user-specified number of equal-length sections (N). Each section of the profile P has its length stretched or shortened by shifting the position of its section end point by a limited number of data points, i.e. the user-specified slack parameter (t) [34,36]. The slack allows the section end points to shift from $-t$ to t points. For each section of P, the stretched or shortened sections are interpolated to the corresponding section of T and the correlation coefficient between both sections is computed. More detailed information can be found in ref. [36].

2.2. Exploratory analysis: Principal Component Analysis

Principal Component Analysis (PCA) [28–30] reduces the number of variables and allows visualizing information included in the $n \times p$ matrix **X**. By linear combinations of the original variables, PCA produces the so-called latent variables or principal components (PC), in such a way that they describe the largest possible variation in **X**. PCA keeps determining PCs describing the largest remaining variation in **X**, orthogonal to the earlier defined, until a maximal number of PCs equal to $n-1$ (when $n < p$). The projections of the n objects in the original data space on PC_i are the scores on PC_i. The contribution of each original variable to the score is reflected by its loading. Both the obtained scores and the loadings can be used for exploratory analysis of the original data. Score plots give information regarding the (dis)similarity of the objects, e.g. about their clustering tendency, while loading plots provide information about the contribution of the original variables.

2.3. Linear multivariate calibration techniques

In general, linear multivariate calibration techniques study the relationship between an $n \times p$ data matrix **X** and an $n \times 1$ response vector **y**. Over the years several techniques have been described [27,28,31,32], but there is no single best method to analyze all possible data.

The relationship between **X** and **y** can be described as:

$$\mathbf{y} = \mathbf{Xb} + \mathbf{f} \tag{1}$$

where **b** represents a $p \times 1$ vector of regression coefficients that express the contribution of the variables to the final model, and **f** the $n \times 1$ residual vector containing information that is not explained by the regression coefficients. In this study, the regression coefficients from several linear multivariate calibration techniques are evaluated.

2.3.1. Stepwise Multiple Linear Regression

Multiple Linear Regression (MLR) [27] produces a linear model (Eq. (2)) describing a quantitative property (dependent variable) by means of independent variables. The regression coefficients b_i are obtained using the least squares method to minimize the residuals (Eq. (3)). Each variable x_i is then multiplied by its regression coefficient b_i to obtain the predicted value for y, noted as $\hat{y}$.

$$\hat{y} = b_0 + b_1 x_1 + b_2 x_2 + \ldots + b_n x_n \tag{2}$$

$$\mathbf{b} = (\mathbf{X^T X})^{-1} \mathbf{X^T y} \tag{3}$$

A requirement to apply classical MLR is that the number of objects n (i.e. fingerprints) is larger than the number of independent variables p. When this condition is not met, a number of variables to include in the regression model should be selected. Stepwise Multiple Linear Regression includes a stepwise selection procedure to achieve this [27].

2.3.2. Principal Components Regression

Principal Components Regression (PCR) uses the latent variables created by PCA to build an MLR model. The optimal model complexity, i.e. the optimal number of PCs in the model, is based on their correlation with **y** [37] or their ability to predict **y** [38]. For

this study, the latter was chosen based on best-subset selection PCR (BSS-PCR) [39]. The PCs to be retained and the sequence in which they have to be entered are selected using Step-MLR and the regression coefficients **b** are obtained using Eq. (3).

The optimal model complexity is then determined by the leave-one-out cross-validation procedure (LOO-CV). During LOO-CV each *i*th object is left out once and for the remaining objects the model is built. The root mean squared error of cross-validation (RMSECV) (Eq. (4)) is then calculated for models with different complexities [28]:

$$\mathrm{RMSECV}(f) = \sqrt{\sum_{i=1}^{N} \frac{(\hat{y}_{\mathrm{cv},i} - y_i)^2}{N}} \quad (4)$$

where f is the model complexity, N the number of calibration samples, y_i the measured response of the ith sample, and $\hat{y}_{\mathrm{cv},i}$ the corresponding response predicted from the calibration model obtained without the ith sample. The optimal model complexity corresponds to the number of latent factors resulting in nearly the lowest RMSECV.

2.3.3. Partial Least Squares

Partial Least Squares [28–30] is another latent-variables technique to express the relation between **X** and **y**, which maximizes the covariance between **X** and **y**. The technique uses the non-linear iterative partial least squares algorithm (NIPALS) [28]. The PLS model can be presented as follows:

$$\mathbf{X} = \mathbf{TP^T} + \mathbf{E} \quad (5)$$

$$\mathbf{y} = \mathbf{TP^T b} + \mathbf{f} = \mathbf{Tq} + \mathbf{f} \quad (6)$$

$$\mathbf{b} = \mathbf{Pq} \quad (7)$$

where **T** represents the $n \times n$ score matrix for **X** and **y**, **P** the $p \times n$ loading matrix representing the regression coefficients of **X** on **T**, **E** the $n \times p$ residual matrix of **X**, **b** the $p \times 1$ vector of PLS regression coefficients, **q** the $n \times 1$ loading vector representing the regression coefficients of **y** on **T**, and **f** the $n \times 1$ residual vector of **y**. The optimal model complexity is determined by the LOO-CV procedure.

2.3.4. Uninformative variable elimination by PLS

Uninformative variable elimination by PLS (UVE-PLS) [31] aims to remove uninformative variables in multivariate data, i.e. those not containing more information than random noise. While constructing the PLS model, a matrix **R**, containing artificial noise variables is added to the original data. All experimental variables that are not more important than the added noise are consequently removed. The procedure is repeated until the number of informative variables is stabilised, resulting in the data matrix $\mathbf{X_{new}}$. A final PLS model is built using $\mathbf{X_{new}}$ and **y** (Eqs. (5–7)) and its complexity is optimized using LOO-CV. More detailed information can be found in ref. [31].

2.3.5. Orthogonal Projections to Latent Structures

Orthogonal Projections to Latent Structures (O-PLS) [32] makes use of a modified NIPALS algorithm which removes the variation in **X** that is not correlated to **y**. This is done by subtracting PLS components, orthogonal to **y**, from the original **X** data. Consequently, the original data is split into two data sets, one that contains the **y**-relevant information and another with the orthogonal data.

An O-PLS model can be written as follows:

$$\mathbf{X} = \mathbf{TP^T} + \mathbf{T_{Yosc}P_{Yosc}^T} + \mathbf{E} \quad (8)$$

$$\mathbf{y} = \mathbf{TP^T b} + \mathbf{f} = \mathbf{Tq} + \mathbf{f} \quad (9)$$

$$\mathbf{b} = \mathbf{Pq} \quad (10)$$

where **T** represents the orthonormal $n \times n$ score matrix for **X** and **y**, **P** the orthonormal $p \times n$ loading matrix representing the regression coefficients of **X** on **T**, $\mathbf{T_{Yosc}}$ the orthogonal $n \times n$ score matrix for **X** and **y**, and $\mathbf{P_{Yosc}}$ its corresponding orthogonal $p \times n$ loading matrix, **E** the $n \times p$ residual matrix of **X**, **b** the $p \times 1$ vector of regression coefficients calculated, **q** the $n \times 1$ loading vector representing the regression coefficients of **y** on **T** and **f** the $n \times 1$ residual vector of **y**.

Removing the orthogonal information of the original data set reduces the number of PLS components in the O-PLS model to a single component, allowing an improved interpretability of the regression coefficients.

3. Experimental

3.1. Herbs and preparation of the herbal extracts

39 *Mallotus* samples, from at least 17 different species, were collected in different Vietnamese regions (Table 1). Six samples were unidentified. Depending on the species and the applicable nature conservation laws, the leaves, roots and/or bark were used. All samples were authenticated by Professor Nguyen Nghia Thin (Hanoi National University, Vietnam) and deposited at the Institute of Natural Products Chemistry, Hanoi, Vietnam.

Extracts were prepared by weighing 2.5 g plant sample and extracting three times with 25 mL methanol in an ultrasonic bath (Branson Ultrasonic Corporation, Connecticut, USA) at a temperature between 40 and 50 °C during 1 h. The extract was filtered through a 240 nm pore size filter paper (Whatman, Hanoi, Vietnam) and evaporated at decreased pressure (60 Pa) and elevated temperature (50 °C). The obtained crude extract was divided over three sample tubes, i.e. one for the DPPH radical scavenging assay, one for HPLC analysis, and one as a library sample.

3.2. HPLC

3.2.1. Equipment, chemicals and reagents

An Agilent 1050 HPLC system (Waldbronn, Germany), consisting of a vacuum degasser, quaternary pump, autosampler and a variable wavelength UV detector, was used. Two coupled Chromolith™ Performance RP-18e columns (100 mm × 4.6 mm I.D.) with a Chromolith™ RP-18e guard column (5 mm × 4.6 mm I.D.) contained the stationary phase. Data acquisition and processing was done with Chemstation for LC (Agilent). HPLC grade methanol, acetonitrile (both Fisher Scientific, Leicestershire, UK), trifluoroacetic acid (TFA) (Sigma–Aldrich, Steinheim, Germany) and MilliQ water, obtained from a MilliQ purification system (Millipore, Bedford, MA), were used to prepare the mobile phases. All solvents were degassed during 15 min on an ultrasonic bath (Branson Ultrasonic Corporation, Danbury, CT) prior to HPLC analyses.

3.2.2. Sample preparation

Samples for HPLC analysis were prepared diluting 50.0 mg crude extract in 2.0 mL methanol. The solution was mixed during 15 min at 400 rpm on a shaking bath (Edmund Bühler, Hechingen, Germany) and afterwards filtered through a 2 μm pore size filter (Schleicher & Schuell, Dassel, Germany) followed by filtration through a 25 mm syringe polypropylene membrane with 0.2 μm pore size (VWR International, Leuven, Belgium).

For the standard compounds mallonanoside A and B [40], quercetin and myricetin, 1.0 mg standard was weighed and dissolved in 10.0 mL methanol. Then, the same procedure was followed as for the crude extracts.

3.2.3. Chromatographic conditions

The conditions developed in [26] were chosen. The mobile phase consisted of (A) 0.05% TFA in ACN, and (B) 0.05% TFA in MilliQ

Table 1
The *Mallotus* samples with their voucher number, species, origin, collection time, used part of the plant and the DPPH scavenging activity results indicated.

Sample	Voucher number	Species	Origin	Collection time	Part of plant	%DPPH$_{rem}$ (n = 3)	s
1	01	*Mallotus luchenensis*	Son La	July 2006	Leaves	82.0	12.1
2	02	*Mallotus microcarpus*	Son La	July 2006	Leaves	63.6	13.0
3	03	*Mallotus barbatus*	Son La	July 2006	Leaves	79.4	9.7
4	MA07	*Mallotus sp1*	Van Hoa	April 2006	Leaves	113.0	20.5
5	NT01	*Mallotus barbatus*	Hagiang	November 2006	Leaves	77.2	10.4
6	NT02	*Mallotus paniculatus*	Hagiang	November 2006	Leaves	82.2	5.5
7	NT03	*Mallotus metcalfianus*	Hagiang	November 2006	Leaves	51.1	14.6
8	MA01	*Mallotus apelta (Ma1)*	Tam Dao	July 2006	Leaves	94.5	0.4
9	MA02	*Mallotus apelta (Ma2)*	Tam Dao	December 2006	Leaves	92.5	3.3
10	MA03	*Mallotus paniculatus*	Tam Dao	April 2006	Leaves	58.4	5.4
11	SP4	*Mallotus sp2*	Langson	March 2006	Leaves	56.8	3.9
12	SP5	*Mallotus philippinensis*	Langson	March 2006	Leaves	98.9	12.7
13	MA11	*Mallotus macrostachyus*	Langson	March 2006	Leaves	75.7	2.2
14	MA12	*Mallotus microcarpus*	Quangbinh	March 2006	Leaves	83.1	2.0
15	MA13	*Mallotus pallidus*	Quangbinh	March 2006	Leaves	65.3	1.9
16	**MA14**	***Mallotus oblongifolius***	**Quangtri**	**March 2006**	**Leaves**	**6.7**	**0.3**
17	**MA15**	***Mallotus floribundus***	**Langson**	**November 2006**	**Leaves**	**6.4**	**0.2**
18	MA16	*Mallotus cuneatus*	Langson	November 2006	Leaves	86.9	3.2
19	**MA17**	***Mallotus cuneatus***	**Quangbinh**	**December 2006**	**Leaves**	**10.3**	**4.1**
20	MA18	*Mallotus sp3*	Quang tri	December 2006	Leaves	91.6	4.1
21	MA19	*Mallotus yunnanensis*	Lang Son	November 2006	Leaves	91.6	6.6
22	MA20	*Mallotus poilanei*	Ke Bang	March 2006	Leaves	90.5	7.0
23	MA22	*Mallotus hookerianus*	Dakrong	March 2006	Leaves	50.0	4.6
24	MA23	*Mallotus nanus*	Daclak	March 2006	Leaves	78.4	9.5
25	MA24	*Mallotus sp4*	Daclak	March 2006	Leaves	56.9	11.7
26	M25	*Mallotus oreophilus*	LaoCai	June 2006	Leaves	88.8	10.5
27	**MA28**	***Mallotus philippinensis***	**Cucphuong**	**December 2006**	**Leaves**	**22.3**	**10.0**
28	**MA29**	***Mallotus barbatus***	**Cucphuong**	**December 2006**	**Leaves**	**11.3**	**4.8**
29	MP31L	*Mallotus paniculatus*	VQG Pumat	September 2006	Leaves	73.5	8.5
30	MP3R	*Mallotus paniculatus*	VQG Pumat	September 2006	Roots	91.5	5.7
31	MP33L	*Mallotus paniculatus*	Bach Ma-TTH	October 2006	Leaves	81.5	3.8
32	MP34R	*Mallotus paniculatus*	Bach Ma-TTH	October 2006	Roots	83.5	6.6
33	**MP35R**	***Mallotus paniculatus***	**Cucphuong**	**December 2006**	**Roots**	**27.9**	**11.3**
34	MP36L	*Mallotus paniculatus*	Cucphuong	December 2006	Leaves	75.3	8.8
35	**MN37R**	***Mallotus nanus***	**VQG-Bachma**	**May 2006**	**Roots**	**12.2**	**1.7**
36	**MN37L**	***Mallotus nanus***	**VQG-Bachma**	**May 2006**	**Leaves**	**4.5**	**1.0**
37	**MN39C**	***Mallotus nanus***	**VQG-Bachma**	**May 2006**	**Bark**	**27.1**	**4.7**
38	M40L	*Mallotus sp5*	VQG Bavi	August 2006	Leaves	73.7	8.7
39	M41C	*Mallotus sp6*	VQG Bavi	August 2006	Bark	65.6	8.0

The highly antioxidant samples are marked in bold.

water. Gradient elution was applied. For 60 min fingerprints, the gradient was 5–20% A in 0–25 min, 20–95% A in 25–50 min, and 95% A during 50–60 min. For 35 min fingerprints, it was 5–25% A in 0–12.5 min, 25–95% A in 12.5–25 min, and 95% A during 25–35 min, while for 22.5 min fingerprints 10–30% A in 0–6.25 min, 30–95% A in 6.25–12.5 min, and 95% A during 12.5–22.5 min was used. Column temperature was 25 °C, flow rate 1.0 mL min^{-1}, injection volume 10 μL, and detection wavelength 254 nm.

3.3. DPPH radical scavenging test

The DPPH antioxidant activity scavenging test [25] measures the capacity to scavenge the stable 1,1-diphenyl-2-picrylhydrazyl radical (DPPH°). In its radical form, DPPH has an absorption band at 515 nm, which disappears upon reduction by an antiradical compound. The remaining DPPH° concentration in the reaction medium is then estimated from a calibration curve. The percentage of remaining DPPH° (%[DPPH°$_{rem}$]) is expressed as follows:

$$\%[\mathrm{DPPH}^\circ_{rem}] = \frac{[\mathrm{DPPH}^\circ_{20\,min}]}{[\mathrm{DPPH}^\circ_{0\,min}]} \times 100 \quad (11)$$

where [DPPH°$_{0\,min}$] is the starting concentration of DPPH radicals, and [DPPH°$_{20\,min}$] the remaining concentration after 20 min of incubation with the sample.

An aliquot (50 μL at a concentration of 20 μg mL^{-1}) of a MeOH solution containing sample or a positive control (tocopherol) was added to 2.5 mL of daily prepared DPPH° solution (25 μg mL^{-1} in methanol). An equal volume (50 μL) of the solvent (methanol) was added to control tubes ([DPPH°$_{0\,min}$]). The DPPH° concentration in the reaction medium was calculated from a linear calibration curve at concentrations ranging from 1 to 50 μg mL^{-1}. Absorbance at 515 nm was measured on a Uvikon 933 spectrophotometer 20 min after starting the reaction. All experiments were performed in triplicate. The reported results (Table 1) are the averages and standard deviations of three independent measurements. The average standard deviation for the 39 *Mallotus* samples was found to be 6.7.

3.4. Data analysis

Computations were performed on a PC with an Intel Core 2 Duo E6750 processor containing 2 GB RAM and running Microsoft Windows XP and Matlab™ 7.1 (The Mathworks, Natick, MA). All data (pre)processing is performed using m-files written for Matlab™ 7.1.

4. Results and discussion

4.1. DPPH radical scavenging test

The results (Table 1) show that only nine samples are considered to have a high antioxidant activity (%DPPH$_{rem}$ < 30), i.e. *Mallotus oblongifolius* (sample 16), *Mallotus floribundus* (17), *Mallotus cuneatus* (19), *Mallotus philippinensis* (27), *Mallotus barbatus* (28), *Mallotus paniculatus* (33) and three *Mallotus nanus* samples (35–37). Two are considered to have an intermediate antioxidant

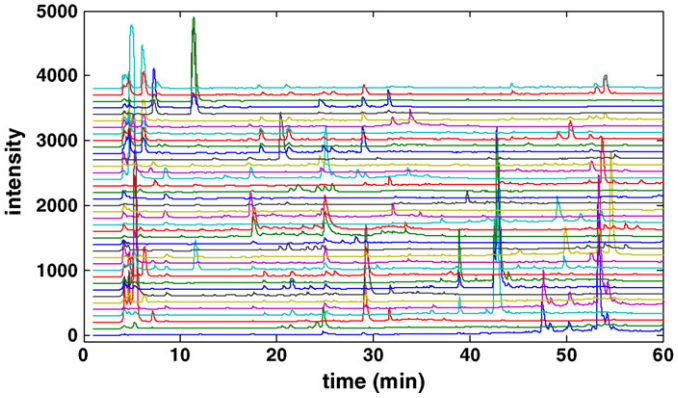

Fig. 1. 60 min fingerprints of the *Mallotus* extracts.

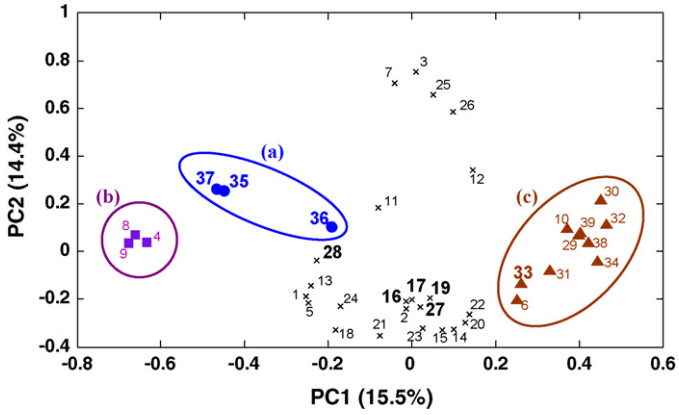

Fig. 2. PC1–PC2 score plot for 60 min fingerprints of the 39 *Mallotus* samples, normalized and column centered. Three groups are distinguished, i.e. containing (a) samples 35-36-37 (●), (b) samples 4-8-9 (■), and (c) samples 6-10-29-30-31-32-33-34-38-39 (▲). The nine highly active antioxidant samples are marked in bold. The remaining samples are indicated with **x**.

activity ($30 < \%DPPH_{rem} < 50$), i.e. *Mallotus metcalfianus* (7) and *Mallotus hookerianus* (23).

M. philippinensis and *M. barbatus* both have antioxidant and non-antioxidant samples present, with the antioxidant samples having their origin in Cucphuong. From eight *M. paniculatus* samples, only the roots from the species with origin in Cucphuong present high antioxidant activity. For *M. nanus*, only the three samples (roots, leaves and bark) with origin in Bachma possess a high antioxidant activity, while those from Daclak did not. Further results and discussion about the variability within the species and possible influences of origin and collection of the samples will be reported in a next paper.

4.2. HPLC fingerprints

Fingerprints with different length (60, 35 and 22.5 min) have been developed. Because there are at least 17 different *Mallotus* species present (Table 1), the fingerprints are very different (Fig. 1). Therefore, it is most likely that different pharmaceutical activities can be attributed to different species.

As the shorter fingerprints show more co-eluting or overlapping peaks, it will be evaluated whether this will lead to a significant loss of information during the data analysis.

4.3. Evaluation of antioxidant activity

4.3.1. Data preprocessing

Different preprocessing methods, i.e. column centering, normalization followed by column centering and standard normal variate followed by column centering, were evaluated. It was found that for all applied linear multivariate calibration techniques, normalization followed by column centering gave the best results for this specific data set (see further). All further discussed results are acquired by data preprocessed in this way.

When shifts in retention times are observed between chromatograms, alignment of the corresponding peaks is needed. However, because of the great divergence in the HPLC fingerprints of the different *Mallotus* species, peak alignment turned out to be far from evident. It is in fact difficult to know which peaks correspond and should be aligned, since no diode array detector (DAD) or mass spectrometry (MS) data were available for the entire data set. Nonetheless, an attempt to align the long fingerprints (60 min) applying correlation optimized warping was performed.

As reference chromatogram, the mean of all 39 fingerprints was calculated. Correlation to the mean chromatogram ranged between 0.53 and 0.83 prior to alignment, and between 0.56 and 0.88 after alignment. On average, correlation to the mean increased from 0.69 to 0.74.

The warped fingerprints (not shown) did not result in improved models or regression coefficient plots. Moreover, one may wonder about the validity of the alignment because of the high complexity of, and the differences between, the *Mallotus* samples. The use of a diode array detector or mass spectrometer may provide information to obtain a more correct alignment of the chromatograms. However, given the above results, in the rest of the paper we worked on the unaligned data set.

4.3.2. Exploratory analysis: Principal Component Analysis

To verify whether groups of samples, occasionally with similar antioxidant activity, could be distinguished, Principal Components Analysis was applied.

When examining the PC1–PC2 score plot (Fig. 2) of the 60 min fingerprints, the samples with antioxidant activity (marked in bold) are not densely clustered. Nevertheless, they are all situated more or less centrally in the plot. Combining the proximity of the samples on the score plot and the a priori knowledge of their recorded fingerprint profiles and species results in the distinction of three groups, i.e. containing (a) samples 35-36-37 (●), (b) samples 4-8-9 (■), and (c) samples 6-10-29-30-31-32-33-34-38-39 (▲). The remaining samples were labelled with **x** and according to the above used criteria cannot be further split in subgroups.

These groups allowed extracting additional information about samples of unknown species. From the fingerprints of group (b) it can be assumed that *Mallotus sp1* (sample 4) belongs to the species *Mallotus apelta* since similar fingerprints are obtained. Similarly, group (c) consists of the fingerprints of eight *M. paniculatus* samples (6-10-29-30-31-32-33-34) and two unidentified samples (38 and 39). As these two latter have similar profiles, they probably belong to the *M. paniculatus* species.

4.3.3. Linear multivariate calibration techniques

Models were built, using several linear multivariate calibration techniques, with data matrix **X** consisting of the 39 fingerprints, and response vector **y**, representing the DPPH radical scavenging test results. No division of the data into a calibration set and a test set was made since the data set is not large enough and prediction of the antioxidant activity of new samples is not the primary concern. The data set also contains too few samples with antioxidant activity versus too many samples without to be very suitable as calibration set to build predictive models.

All models were built for the three fingerprints lengths. The optimal model complexity was determined from the RMSECV obtained from LOO-CV. For all applied techniques, the simplest model with

Table 2
Number of components and RMSECV for the calibration models build based on the 22.5, 35 and 60 min fingerprints using normalization and column centering as preprocessing.

Calibration technique	22.5 min fingerprints		35 min fingerprints		60 min fingerprints	
	# of components	RMSECV	# of components	RMSECV	# of components	RMSECV
Step-MLR	10	15.3	14	16.3	9	16.2
PCR	3	21.8	7	15.1	6	15.1
PLS	6	11.4	7	13.7	5	12.7
UVE-PLS	4	7.5	4	9.0	3	11.6
O-PLS	1 (1*)	13.4	1 (1*)	14.2	1 (1*)	13.8

For O-PLS, the number of removed orthogonal components is given between brackets.

Table 3
Results from the DPPH radical scavenging assay and predictions from the models built with the 60 min fingerprints. Preprocessing: normalization and column centering.

Sample no.	DPPH	Step-MLR	PCR	PLS	UVE-PLS	O-PLS	O-PLS (aligned)
16	6.7	−35.6	8.2	5.4	6.1	0.5	−1.1
17	6.4	20.9	19.3	27.6	19.6	15.5	−13.3
19	10.3	38.8	33.5	21.9	25.9	28.7	31.9
27	22.3	12.3	48.5	41.3	30.6	44.8	50.6
28	11.3	54.5	56.7	46.6	58.8	47.3	63.8
33	27.9	64.7	53.5	49.5	38.8	48.8	56.6
35	12.2	27.8	12.9	24.5	16.4	9.1	10.1
36	4.5	−16.6	29.9	19.1	17.2	24.5	23.0
37	27.1	18.0	14.4	18.8	23.0	38	4.2
Mean bias	/	27.7	23.4	18.6	18.3	21.0	26.2

nearly the lowest RMSECV was chosen. The differences in prediction error and model complexity were evaluated for models selected for the different techniques. However, the main focus of this study was on the interpretability of the regression coefficients given by the different multivariate calibration models as indication for peaks with potential antioxidant activity.

Table 2 shows the calibration results for all fingerprint lengths after application of normalization and column centering as preprocessing approach.

With the exception of the PCR model, those with the shortest fingerprints have the smallest RMSECV, which is most pronounced for the UVE-PLS technique, 7.5 (against the average standard deviation of 6.7 for the DPPH reference test). However, to indicate interesting peaks it is better to evaluate the longer fingerprints where peaks are not or less overlapping.

For the nine highly antioxidant samples (16, 17, 19, 27, 28, 33, 35, 36 and 37) none of the techniques predicted all nine as being highly active ($\%DPPH_{rem} < 30$), with a range between five and seven out of nine (Table 3). For instance, PLS and O-PLS predicted the same three samples as intermediate active ($30 < \%DPPH_{rem} < 50$), Step-MLR predicted one intermediate active and two non-active ($\%DPPH_{rem} > 50$) samples, PCR predicted two intermediate active and two non-active samples and UVE-PLS predicted one intermediate and one non-active sample. None of the models predicted antioxidative activity for non-active samples. Furthermore, prediction of the antioxidant samples by Step-MLR shows a large bias compared to the other models (Table 3). This might have implications on the reliability of the regression coefficients.

Concerning model complexity, O-PLS resulted in the lowest complexity as it removes the orthogonal information to obtain a single-component PLS model. By doing so, only the information of data matrix **X** (fingerprints) correlated to the response vector **y** (antioxidant activity) is kept resulting in a significant amelioration of the RMSECV compared to the matching one-component PLS model (RMSECV = 13.8 versus 17.1 for one orthogonal projection removed).

4.3.4. Regression coefficients

The main focus of this study is to indicate those peaks in the fingerprints potentially responsible for the antioxidant activity of the measured samples. For this purpose, the regression coefficients of the models are examined. On the regression plots (Fig. 3), the coefficients indicating peaks corresponding to potential antioxidant compounds or to those representing a similar behaviour, i.e. that are present at high concentration when the antioxidant activity is high, are negative as the DPPH radical scavenging test results decrease with increasing activity. Positive regression coefficients represent compounds that show an opposite behaviour to the antioxidant activity. To evaluate which modelling technique gives the best interpretable regression coefficients, they are compared for the 60 min fingerprints of the antioxidant samples (Fig. 3).

The stepwise selection procedure included only 9 variables of the 18,000 available data points. Only three variables are selected at the maximum of a peak, while the others are located at the beginning or tail of a peak, as well as at very minor peaks or at the baseline. One would expect the variables to be selected at the maximum of a peak, since higher concentration correlates to higher contribution to the antioxidant activity. Furthermore, the regression coefficients did not correspond to any peak of the highly active samples 16, 17, 19 and 27. One may question the reliability of these variables and their contribution to the antioxidant activity.

For PCR, the selection procedure included six PCs (3, 4, 5, 6, 8 and 12). Several major negative (and positive) peaks can be noticed in the regression plot. However, the true correlation of the coefficients to the antioxidant activity may a priori be questioned. The selected PCs are created based on the largest remaining variance within the data matrix **X**. Only afterwards, the selection procedure includes the PCs based on their ability to predict **y**. Anyway, the PCR regression coefficients correlate rather well to the O-PLS coefficients (see further) that are obtained taking the **y**-information into account and which were further considered as best to interpret and to indicate relevant peaks.

In the regression plot of the PLS model, large and small coefficients, both negative and positive, can be noticed. The large coefficients correspond to compounds potentially important for the modelled activity, while the small are caused by the orthogonal variation present in the data matrix **X** [32]. Therefore, they are considered unimportant, but the presence of these small coefficients renders interpretation of the regression plot harder in regards of identifying the potentially interesting peaks. However, theoretically this regression plot gives a better representation of the true contributions of components to the antioxidant activity compared to

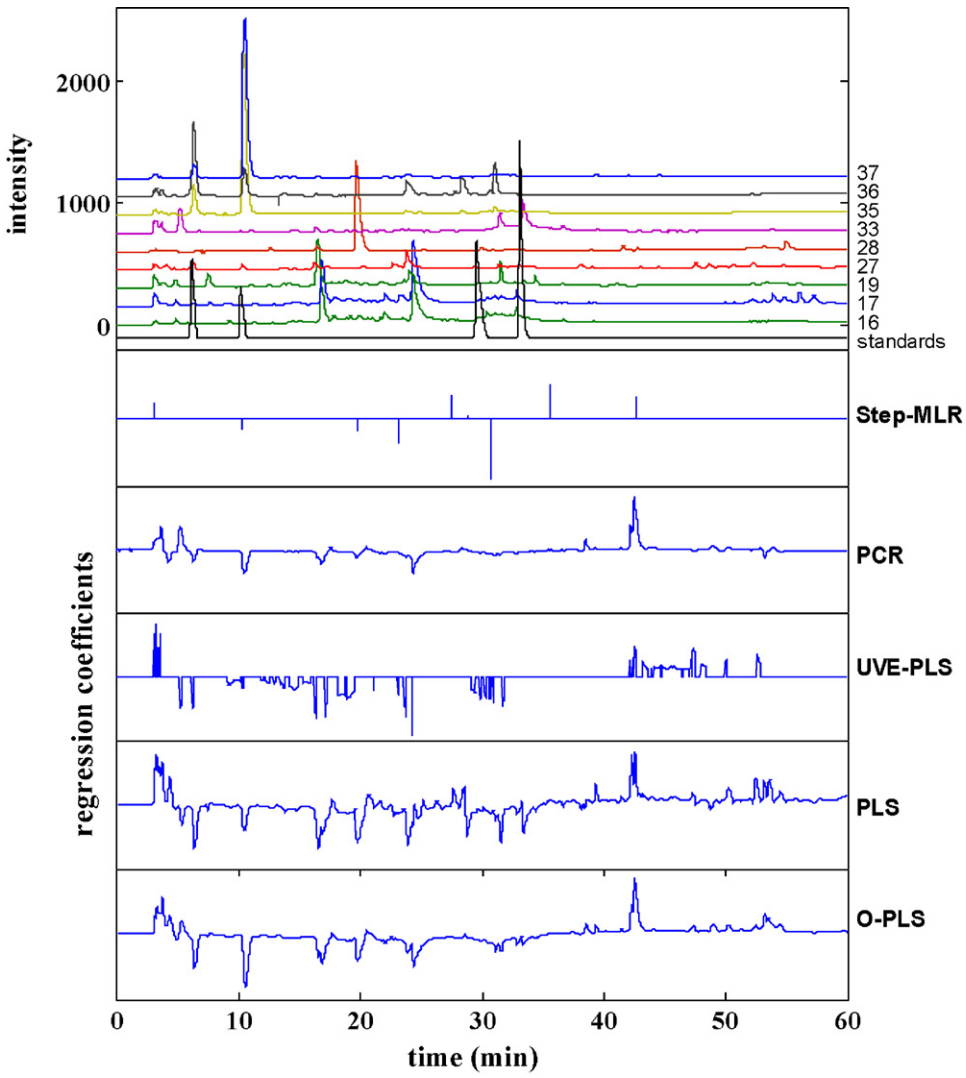

Fig. 3. Plot of the 60 min fingerprints of the antioxidant samples with compounds mallonanoside B (first peak of standards), mallonanoside A (second peak), myricetin (third peak) and quercetin (fourth peak). The bottom plots show the regression coefficients from Step-MLR, PCR, PLS, UVE-PLS and O-PLS, preprocessed with normalization and column centering.

PCR as the correlation between the chromatographic data (**X**) and the antioxidant activity (**y**) is also taken into account. For example, at about 55 min, the PLS model present some regression peaks, while PCR does not. However, overall for this case study the PCR coefficients seem to agree rather well with the PLS ones.

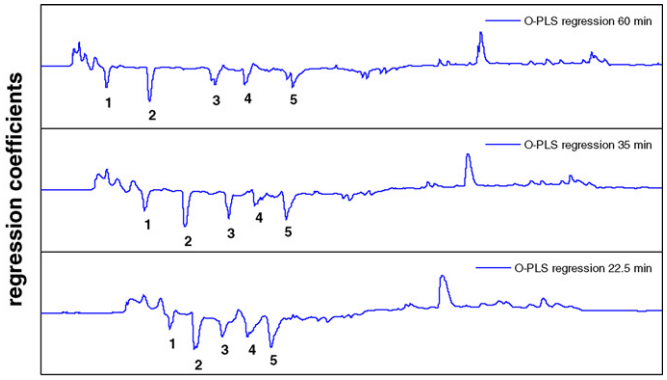

Fig. 4. The O-PLS regression plots for the 60, 35 and 22.5 min fingerprints.

When correlating the regression plot of UVE-PLS to the fingerprints, not all regression coefficient peaks correspond to substance peaks, but occasionally also to baseline, e.g. some negative regression peaks between 29.8 and 31.1 min (Fig. 3). Furthermore, several minor peaks seen in the fingerprints were considered as uninformative by the stability criterion of the technique and thus eliminated from the final data set. One may question the reliability of the stability criterion in our context as uninformative data points might be included, while elimination of real peaks may result in a loss of information. Decreasing the considered cut-off value in UVE-PLS to select more variables did not result in inclusion of the eliminated peaks, while even more baseline variables were selected.

Also, we did not experience UVE-PLS as a user-friendly technique. First one has to define the number of added noise variables and the cut-off level to consider. As it concerns an iterative method, calculations should be repeated till the stability criterion is met and the number of variables is constant. Moreover, small differences could be distinguished in the model regarding the selected variables, the RMSECV, the model complexity and the regression coefficients upon changing the size of **R** or repeating the calculations.

The regression plot obtained from the one-component O-PLS model after removal of one orthogonal projection, turns out much smoother than the others (with the exception of PCR). Due to the removal of the orthogonal information of the original data, small contributions, rendering the interpretation of the regression coefficients hard, are avoided. The (negative) regression coefficient peaks clearly correspond to substance peaks in the fingerprints. Removal of more than one orthogonal projection did not result in changed regression.

The fingerprints of the highly active antioxidant samples and the O-PLS regression coefficients are compared (Fig. 3). The peaks potentially responsible for the antioxidant activity can clearly be distinguished in the fingerprints of the active samples. Two of the standard compounds, mallonanoside A and B, align with peaks identified as possibly antioxidative. These two compounds are identified in *M. nanus* species [40], which are also the species involved in the fingerprints shown (samples 35–37). Further investigation is needed to identify the compounds as indeed being mallonanoside A and B and to evaluate their real antioxidant activity, as well as identification and evaluation of the underlying unknown compounds of the indicated peaks (see Part II).

The antioxidant samples have no major peaks present at retention times corresponding to the positive coefficient peaks in the regression plot. Some of these coefficients match compounds present in the fingerprints of *M. apelta* (samples 4, 8 and 9). It is known that *M. apelta* possesses cytotoxic activity [41–43]. Further, none of the negative regression peaks correspond to a substance peak in the *M. apelta* samples, which is in accordance with the results of the DPPH test.

Calculations for O-PLS were also made for the 35 and 22.5 min fingerprints. Both analysis and calculation times are reduced considerably. On the regression plots (Fig. 4) the major negative coefficients corresponding to the potentially antioxidant compounds are found, also on the shorter fingerprints. For this case study, the shorter fingerprints still seem to allow indicating the peaks of interest. However, the coefficient peaks 3 and 5 are splitted for the 60 min fingerprints. Further analysis (e.g. by LC-MS) should be performed to verify whether these split peaks correspond to different compounds in the fingerprint or are caused by experimental shifts in retention time. The results will be discussed in Part II.

5. Conclusions

Several linear multivariate calibration techniques were applied to fingerprints of *Mallotus* extracts to indicate the peaks responsible for the antioxidant activity. This was done by examining the regression coefficients of the different calibration models after using several preprocessing methods and alignment by Correlation Optimized Warping. Normalization followed by column centering resulted in better models for this particular data set, while alignment turned out to be far from evident due to the high complexity of and the large differences between the *Mallotus* samples.

Large differences in the interpretability of the coefficients could be noticed. From the examined techniques, Stepwise Multiple Linear Regression is least recommended as prediction of the highly active samples showed a large bias and the variable selection was of few use for our purpose. Principal Component Regression potentially can take large variations uncorrelated to the antioxidant activity into account. However, in our actual case study PCR did not result in such bad results. These problems theoretically do not occur using Partial Least Squares models. Ordinary PLS gave rise to difficulties interpreting the regression coefficients because of the presence of small positive and negative contributions, possibly caused by orthogonal information of the original data set. Uninformative Variables Elimination-PLS reduced the number of variables taken into account, but one could question the reliability of the selection criterion in our study. Moreover, UVE-PLS leads to different results when repeating the procedure and is not very user-friendly.

In this study, Orthogonal Projection to Latent Structures was found to be the better performing technique to indicate the potential antioxidant active compounds in the *Mallotus* extracts due to its simplicity and repeatability, and to remove the orthogonal information in the original data set. O-PLS resulted in a decreased model complexity contributing to an improved interpretability of the regression coefficients. No changes in the regression were observed upon removing more then one orthogonal projection.

The retention time of two peaks indicated as potentially interesting align to the standard compounds mallonanoside A and B. Further investigation using LC-MS is needed to identify them as well as other unknown but interesting peaks. In Part II, the results of LC-MS analyses will be discussed.

Acknowledgement

Bieke Dejaegher is a post-doctoral fellow of the Fund for Scientific Research (FWO), Flanders, Belgium.

References

[1] Traditional Medicine Strategy 2002–2005, World Health Organization, Geneva, 2005, accessible at http://www.who.int/medicines/publications/traditionalpolicy/en/ (accessed on June 10, 2008).
[2] Guidelines for the Assessment of Herbal Medicine, World Health Organization, Munich, Germany, 1991, accessible at http://whqlibdoc.who.int/HQ/1991/WHO_TRM_91.4.pdf (accessed on January 10, 2009).
[3] Quality control methods for medicinal plant materials, World Health Organization, Geneva, 1998, accessible at http://whqlibdoc.who.int/publications/1998/9241545100.pdf (accessed on March 5, 2009).
[4] Q. Lu, S.C. Chow, S.K. Tse, Assessing the consistency of traditional Chinese medicine with multiple correlative active components, Journal of Biopharmaceutical Statistics 17 (2007) 791–808.
[5] V. Bankova, Chemical diversity of propolis and the problem of standardization, Journal of Ethnopharmacology 100 (2005) 114–117.
[6] WHO, General Guidelines for Methodologies on Research and Evaluation of Traditional Medicines, Geneva, Switzerland, 2000, accessible at http://www.who.int/medicinedocs/en/d/Jwhozip42e/#Jwhozip42e (accessed on December 8, 2008).
[7] G. Alaerts, N. Matthijs, J. Smeyers-Verbeke, Y. Vander-Heyden, Chromatographic fingerprint development for herbal extracts. A screening and optimization methodology on monolithic columns, Journal of Chromatography A 1172 (2007) 1–8.
[8] M. Obradovič, S. Strgulc Krajšek, M. Dermastia, S. Kreft, A new method for the authentication of plant samples by analyzing fingerprint chromatograms, Phytochemical Analysis 18 (2007) 123–132.
[9] F. Zeng, X.-M. Wang, M. Yang, Z.-H. Lu, D.-A. Guo, Fingerprint analysis of different *Panax* herbal species by HPLC-UV method, Journal of Chinese Pharmaceutical Sciences 16 (2007) 272–276.
[10] X.-Q. Liu, L.-L. Du, W.-W. Li, H.-F. Guan, F. Liu, Simultaneous qualitative and quantitative analysis of commercial *Bistorta* Rhizome and its differentiation from closely related herbs using TLC and HPLC-DAD fingerprinting, Chemical and Pharmaceutical Bulletin 56 (2008) 75–78.
[11] C.-F. Qiao, Q.-B. Han, J.-Z. Song, S.-F. Mo, L.-D. Kong, H.-F. Kung, H.-X. Xu, Chemical fingerprint and quantitative analysis of Fructus Psoraleae by high-performance liquid chromatography, Journal of Separation Science 30 (2007) 813–818.
[12] Y. Xie, Z.-H. Jiang, H. Zhou, X. Cai, Y.-F. Wong, Z.-Q. Liu, Z.-X. Bian, H.-X. Xub, L. Liu, Combinative method using HPLC quantitative and qualitative analyses for quality consistency assessment of a herbal medicinal preparation, Journal of Pharmaceutical and Biomedical Analysis 43 (2007) 204–212.
[13] A.M. van Nederkassel, M. Daszykowski, D.L. Massart, Y. Vander Heyden, Prediction of total green tea antioxidant capacity from chromatograms by multivariate modelling, Journal of Chromatography A 1096 (2005) 177–186.
[14] M. Dumarey, A.M. van Nederkassel, E. Deconinck, Y. Vander Heyden, Exploration of linear multivariate calibration techniques to predict the total antioxidant capacity of green tea from chromatographic fingerprints, Journal of Chromatography A 1192 (2008) 81–88.
[15] A.M. van Nederkassel, C.J. Xu, P. Lancelin, M. Sarraf, D.A. MacKenzie, N.J. Walton, F. Bensaid, M. Lees, G.J. Martin, J.R. Desmurs, D.L. Massart, J. Smeyers-Verbeke, Y. Vander Heyden, Chemometric treatment of vanillin fin-

gerprint chromatograms: effect of different signal alignments on principal component analysis plots, Journal of Chromatography A 1120 (2006) 291–298.
[16] P. Chen, M. Ozcan, J. Harnly, Chromatographic fingerprint analysis for evaluation of *Ginkgo biloba* products, Analytical and Bioanalytical Chemistry 389 (2007) 251–261.
[17] P.K. Soares, I.S. Scarminio, Multivariate chromatographic fingerprint preparation and authentication of plant material from the genus *Bauhinia*, Phytochemical Analysis 19 (2008) 78–85.
[18] D.T. Loi (Ed.), Glossary of Vietnamese Medicinal Plants, Science & Technics Publication, Hanoi, Vietnam, 2001.
[19] J.-F. Xu, Z.-M. Feng, J. Liu, P.C. Zhang, New hepatoprotective coumarinolignoids from *Mallotus apelta*, Chemistry & Biodiversity 5 (2008) 591–597.
[20] V.S. Rana, M.S.M. Rawat, G. Pant, A. Nagatsu, Chemical constituents and antioxidant activity of *Mallotus roxburghianus* leaves, Chemistry & Biodiversity 2 (2005) 792–798.
[21] T.-Y. An, L.-H. Hu, X.-F. Cheng, Z.-L. Chen, Benzopyran derivates from *Mallotus apelta*, Phytochemistry 57 (2001) 273–278.
[22] B. Supudompol, K. Likhitwitayawuid, P.J. Houghton, Phloroglucinol derivatives from *Mallotus pallidus*, Phytochemistry 65 (2004) 2589–2594.
[23] D. Chattopadhyay, G. Arunachalam, A.B. Mandal, T.K. Sur, S.C. Mandal, S.K. Bhattacharya, Antimicrobial and anti-inflammatory activity of folklore: *Mallotus peltatus* leaf extract, Journal of Ethnopharmacology 82 (2002) 229–237.
[24] C. Rivière, V. Nguyen Thi Hong, L. Pieters, B. Dejaegher, Y. Vander Heyden, M. Chau Van, J. Quetin-Leclercq, Polyphenols isolated from antiradical extracts of *Mallotus metcalfianus*, Phytochemistry 70 (2009) 91–99.
[25] R. Aquino, S. Morelli, M.R. Lauro, S. Abdo, A. Saija, A. Tomaino, Phenolic constituents and antioxidant activity of an extract of *Anthurium versicolor* leaves, Journal of Natural Products 64 (2001) 1019–1023.
[26] N. Nguyen Hoai, B. Dejaegher, C. Tistaert, V. Nguyen Thi Hong, C. Rivière, C. van Minh, J. Quetin-Leclercq, Y. Vander Heyden, Development of HPLC fingerprints for Mallotus species extracts and evaluation of the peaks responsible for their antioxidant activity, Journal of Pharmaceutical and Biomedical Analysis, doi:10.1016/j.jpba.2009.06.016, in press.
[27] 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.
[28] 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.
[29] H. Martens, T. Næs, Multivariate Calibration, Wiley, Chichester, 1989.
[30] H. Martens, M. Martens, Multivariate Analysis of Quality, An Introduction, Wiley, Chichester, 2001.
[31] V. Centner, D.L. Massart, O.E. de Noord, S. de Jong, B.G.M. Vandeginste, C. Sterna, Elimination of uninformative variables for multivariate calibration, Analytical Chemistry 68 (1996) 3851–3858.
[32] J. Trygg, S. Wold, Orthogonal projections to latent structures (O-PLS), Journal of Chemometrics 16 (2002) 119–128.
[33] G. Tomasi, F. van den Berg, C. Andersson, Correlation optimized warping and dynamic time warping as preprocessing methods for chromatographic data, Journal of Chemometrics 18 (2004) 231–241.
[34] A.M. van Nederkassel, M. Daszykowski, P.H.C. Eilers, Y. Vander Heyden, A comparison of three algorithms for chromatograms alignment, Journal of Chromatography A 1118 (2006) 199–210.
[35] B. Walczak, W. Wu, Fuzzy warping of chromatograms, Chemometrics and Intelligent Laboratory Systems 77 (2005) 173–180.
[36] N.-P.V. Nielsen, J.M. Carstensen, J. Smedsgaard, Aligning of single and multiple wavelength chromatographic profiles for chemometric data analysis using correlation optimised warping, Journal of Chromatography A 805 (1998) 17–35.
[37] J. Sun, A correlation principal component regression analysis of NIR data, Journal of Chemometrics 9 (1995) 21–29.
[38] J.M. Sutter, J.H. Kalivas, P.M. Lang, Which principal components to utilize for principal component regression, Journal of Chemometrics 6 (1992) 217–225.
[39] J. Verdú-Andrés, D.L. Massart, Comparison of prediction- and correlation-based methods to select the best subset of principal components for principal component regression and detect outlying objects, Applied Spectroscopy 52 (1998) 1425–1434.
[40] M. Chau Van, D. Nguyen Hai, K. Phan Van, T. Nguyen Phuong, C. Nguyen Xuan, Y. Vander Heyden, J. Quetin-Leclercq, Two new *C*-glucosyl benzoic acids and flavonoids from *Mallotus nanus*, in preparation.
[41] K. Phan Van, M. Chau Van, H. Hoang Thanh, N. Nguyen Hoai, L. Jung Joon, K. Young Ho, Pentacyclic triterpenoids from *Mallotus apelta*, Archives of Pharmacal Research 27 (2004) 1109–1113.
[42] K. Phan Van, D. Nguyen Hai, B. Ha Viet, H. Hoang Thanh, M. Chau Van, H. Le Mai, L. Jung Joon, K. Young Ho, New cytotoxic benzopyrans from the leaves of *Mallotus apelta*, Archives of Pharmacal Research 28 (2005) 1131–1134.
[43] L. Chinh Van, M. Chau Van, L. Jung-Joon, J. Sang-Hun, Exploration of essential structure of MalloapeltaB for the inhibitory activity against TNF induced NF-κB activation, Archives of Pharmacal Research 29 (2006) 840–844.