Journal of Chromatography B, 910 (2012) 114–121

Contents lists available at SciVerse ScienceDirect

Journal of Chromatography B

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

Potentially antioxidant compounds indicated from *Mallotus* and *Phyllanthus* species fingerprints☆

S. Thiangthum[a,b], B. Dejaegher[a], M. Goodarzi[a], C. Tistaert[a], A.Y. Gordien[c], N. Nguyen Hoai[d], M. Chau Van[d], J. Quetin-Leclercq[c], L. Suntornsuk[b], Y. Vander Heyden[a,*]

[a] *Department of Analytical Chemistry and Pharmaceutical Technology, Center for Pharmaceutical Research (CePhaR), Vrije Universiteit Brussel (VUB), Laarbeeklaan 103, B-1090 Brussels, Belgium*
[b] *Department of Pharmaceutical Chemistry, Faculty of Pharmacy, Mahidol University, 447 Sri-Ayudhaya Road, Rajathevee, Bangkok 10400, Thailand*
[c] *Analytical Chemistry, Drug Analysis and Pharmacognosy Unit, Louvain Drug Research Institute, Université Catholique de Louvain (UCL), Avenue E. Mounier 72, B-1200 Brussels, Belgium*
[d] *Institute of Marine Biochemistry, Vietnam Academy of Science and Technology (VAST), 18 Hoang Quoc Viet, Cau Giay, Hanoi, Vietnam*

ARTICLE INFO

Article history:
Received 18 February 2012
Accepted 21 June 2012
Available online 28 June 2012

Keywords:
Chromatographic fingerprints
Antioxidant activity
Multivariate calibration
Mallotus species
Phyllanthus species

ABSTRACT

The genera of *Mallotus* and *Phyllanthus* contain several species that are commonly used as traditional medicines in oriental countries. Some species show interesting pharmaceutical activities, such as an antioxidant activity. To produce clinically useful medicines or food supplements (nutraceuticals) from these herbs, the species should be identified and a thorough quality control should be implemented. Nowadays, the integration of chromatographic and chemometric approaches allows a high-throughput identification and activity prediction of medicinal plants. In this study, Principal Component Analysis (PCA) and Hierarchical Cluster Analysis (HCA) were applied and compared to distinguish *Mallotus* and *Phyllanthus* species. Moreover, peaks from their chromatographic fingerprints, which were responsible for their antioxidant activity were assigned. For the latter purpose, the relevant information was extracted from the chromatographic fingerprints using linear multivariate calibration techniques, i.e., Partial Least Squares (PLS) and Orthogonal Projections to Latent Structures (O-PLS). Results reveal that exploratory analysis using PCA shows somewhat diverging clustering tendencies between *Mallotus* and *Phyllanthus* samples than HCA. However, both approaches mainly confirm each other. Concerning the multivariate calibration techniques, both PLS and O-PLS models demonstrate good predictive abilities. By comparing the regression coefficients of the models with the chromatographic fingerprints, the peaks that are potentially responsible for the antioxidant activity of the extracts could be confirmed.

1. Introduction

Herbal medicines have been used by many people around the world for thousands of years. Unfortunately, their quality, safety, and efficacy are not always sufficiently evaluated [1,2]. Because of their complex composition, the development of a suitable analytical procedure to separate all or as many compounds as possible from such herbal sample is a challenging task. Furthermore, the concentrations of the herbal components can vary significantly depending on the cultivation conditions of the plant, the drying process, and the harvest season [2]. Some researchers only use information from one or a limited number of compounds, the so-called markers, to evaluate the quality of herbal medicines. However, it seems doubtful that only focusing on some compounds will describe and evaluate the complexity of the herbal sample properly.

The World Health Organization (WHO) has accepted chromatographic fingerprint analysis as a strategy for the assessment of herbal medicines [3]. A chromatographic fingerprint can be obtained by, for example, reversed-phase high-performance liquid chromatography (RPLC), and typifies the complete composition of a herbal medicine. A fingerprint represents a chromatographic profile in which the detectable chemical constituents are separated as much as possible. The obtained fingerprints can be used as a unique identification tool to evaluate the authenticity of a herbal sample, the quality and assurance of the consistency, and the stability of a herbal medicine. Nowadays, the combination of (hyphenated) chromatographic instruments and chemometrical approaches for data (pre-) treatment allows a fast investigation of herbal samples [4–11]. Moreover, chemometric treatment of

☆ This paper belongs to the Special Issue Chemometrics in Chromatography, edited by Pedro Araujo and Bjørn Grung.

* 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 © 2012 Elsevier B.V. All rights reserved.
http://dx.doi.org/10.1016/j.jchromb.2012.06.025

the chromatographic fingerprints also allows modeling and predicting pharmacological activities (e.g. antioxidant and cytotoxic activities) and/or indicating peaks potentially responsible for the modeled activities [2,4,12–16].

The *Mallotus* and *Phyllanthus* genera, belonging to the family of the *Euphorbiaceae*, are widely distributed in most tropical and sub-tropical countries. They have been extensively used in folk medicine in India, China, Vietnam and other countries for thousands of years for the treatment of a broad spectrum of diseases. For instance, the genus *Mallotus* has been used for the treatment of chronic hepatitis and enteritis [17,18], while the genus *Phyllanthus* has been used for kidney, urinary bladder, and intestinal infections, and for diabetes [19,20]. Recently, many studies were performed concerning the chemical components of *Mallotus* and *Phyllanthus* species and several pharmacologically active constituents were determined [12,21–25]. Cytotoxic (*Mallotus apelta* [22]), antimicrobial (*Mallotus peltalus* [22], *Phyllanthus emblica* [24]), anti-inflammatory (*Mallotus peltalus* [22], *Mallotus spodocarpys* [22]), and antioxidant activities (*Mallotus metcalfianus* [23], *Phyllanthus emblica* [21], *Phyllanthus niruri* [25]) have been reported for both genera.

In a parallel study [26], RPLC fingerprints of 36 samples, i.e., 10 *Mallotus* and 26 *Phyllanthus* samples, were developed. Then, unsupervised and supervised classification methods were used to classify the *Mallotus* and *Phyllanthus* samples according to genera (*Mallotus* and *Phyllanthus*) and species (*Mallotus apelta*, *Mallotus paniculatus, Phyllanthus emblica*, *Phyllanthus reticulatus*, *Phyllanthus urinaria* L., *Phyllanthus amarus*). As unsupervised techniques, Principal Component Analysis (PCA) and Hierarchical Cluster Analysis (HCA) were evaluated, while as supervised classification techniques, Linear Discriminant Analysis, Quadratic Discriminant Analysis, and Classification and Regression Trees, were compared. The methods were applied to classify the samples in two classes (i.e. the two genera) or in 6 classes (i.e. the 6 species). Results showed the applicability of both unsupervised and supervised methods to discriminate between the samples.

The goal of this study was to model the antioxidant activity of the 36 *Mallotus* and *Phyllantus* samples, originating from different genera, species, origins and/or collection times, as a function of their chromatographic fingerprints. The goal of this modeling is not to use the model to predict the activity of future samples but to indicate peaks potentially responsible for the antioxidant activity [16]. The antioxidant activity of the herbal extracts was determined with a 1,1-diphenyl-2-picrylhydrazyl (DPPH) radical scavenging activity test and expressed in term of the median inhibition concentration (IC_{50}). First, an unsupervised data analysis using PCA and HCA was performed to verify whether the antioxidant samples could be distinguished from the less or non-active samples. Then, the antioxidant activity was modeled as a function of the earlier developed HPLC fingerprints [26] using the multivariate calibration techniques Partial Least Squares (PLS) [27] and Orthogonal Projections to Latent Structures (O-PLS) [28]. The regression coefficients of the resulting models were evaluated to indicate the peaks possibly responsible for the antioxidant activity.

2. Theory

2.1. Data preprocessing

Prior to data analysis, the chromatographic fingerprints are organized in an $n \times p$ data matrix **X**, where the n objects (herbal samples) constitute the rows and the p variables (time points) the columns. In each cell of the matrix, the detector signal intensity at a given time point is presented.

The results of the data analysis are influenced by the applied data preprocessing method. In this study, different methods to pretreat the data, i.e. column centering, normalization followed by column centering, and standard normal variate (SNV) followed by column centering, were applied and compared [4,27,29,30].

Column centering removes the column mean from each corresponding column element. Normalization of the chromatograms scales the rows to a constant total by dividing each row by its corresponding norm. SNV corresponds to row centering, followed by row scaling, where row centering removes the row mean from each corresponding row element and row scaling divides each row element by its corresponding row standard deviation.

2.2. Unsupervised exploratory data analysis

Unsupervised data analysis only uses information contained in the data matrix **X**, i.e. the fingerprints, and does not use the information contained in the response vector **y**.

2.2.1. Principal Component Analysis

Principal Component Analysis (PCA) reduces the number of variables and visualizes the information included in the $n \times p$ data matrix **X** [27,30]. PCA makes linear combinations of the original variables, thus creating the so-called latent variables or principal components (PCs), in such a way that the latter describe the largest possible remaining variation in **X** and are orthogonal. The projection of an object on a PC is called a score on this PC, while the projection of each original variable to the PC is called a loading. A score plot represents the scores on two PC's and shows information regarding the (dis)similarity of the objects, while a loading plot provides information on the contribution of the original variables to the considered PC's.

2.2.2. Hierarchical Cluster Analysis

Hierarchical Cluster Analysis (HCA) is a clustering method applied to reveal the underlying structure of objects through an iterative process that associates (agglomerative method) or dissociates (divisive method) the data set object by object, and that is stopped when all objects have been processed [30,31].

A divisive method starts with all objects in one cluster and divides them into subsets, continuously making smaller clusters until all objects are individually in a cluster [30,32]. An agglomerative procedure, on the other hand, starts with each object in a separate cluster and combines the clusters sequentially, reducing the number of clusters at each step until all objects belong to only one cluster [30]. The hierarchical clustering process can be represented as a tree or dendrogram, where each step in the clustering process is illustrated by a joint of clusters.

In this study, an agglomerative HCA was selected to visualize the data contained in the fingerprint matrix **X** and give insight in the clustering tendency of the data. Several (dis)similarity measures to cluster the objects can be used, e.g. Euclidean distance, Mahalanobis distance, Pearson correlation distance, and Spearman's rank correlation coefficient [30].

In this work, the Euclidean distance and the Pearson correlation distance, which is calculated as "1 − Pearson correlation coefficient r", were evaluated as distance measures [30]. Consider $\mathbf{x}$ ($x_1, x_2, \ldots, x_n$) and $\mathbf{z}$ ($z_1, z_2, \ldots, z_n$) as two sets of n measurements/variables (fingerprints) with means (averages) $\bar{x}$ and $\bar{z}$. The Euclidean distance (ED) between $\mathbf{x}$ and $\mathbf{z}$ is calculated as follows

$$\mathrm{ED} = \sqrt{\sum_{i=1}^{n} (x_i - z_i)^2} \quad (1)$$

where x_i and z_i are the ith elements of $\mathbf{x}$ and $\mathbf{z}$.