
DATA HANDLING OF CHROMATOGRAPHIC HERBAL FINGERPRINTS

Bieke Dejaegher and Yvan Vander Heyden

Vrije Universiteit Brussel (VUB), Pharmaceutical Institute, Department Analytical Chemistry and Pharmaceutical Technology (FABI), Laarbeeklaan 103, B-1090 Brussels, Belgium,
Bieke.Dejaegher@vub.ac.be, yvanvdh@vub.ac.be

The use of herbal medicines is gaining importance, even in Western medicines. In any herbal medicine (herbs and extracts), however, there are many unknown components, which are often only present in low amounts. Identifying only one or some compounds hardly describes the complex extracts, and will thus not be reliable enough for their quality control. A possible solution is the use of 'fingerprint technology', which can, for example, be performed by means of High-Performance Liquid Chromatography (HPLC). This technology has been approved by the World Health Organization, the Chinese State Food and Drug Administration, the USA Food and Drug Administration, and the European Agency for the Evaluation of Medicinal Products. The idea is to develop a complete chromatographic pattern of the herbal extract, i.e. a chromatographic fingerprint, in which as many compounds as possible are separated.

When using the fingerprint technology, two main steps are considered, i.e. the fingerprint development, and the extraction of information from the fingerprint or the data handling. Fingerprint development can be further divided into a sample preparation step and an actual fingerprint development step (i.e. determining the optimal analytical conditions). Considering the data handling, (1) data pretreatment, (2) identification and quality control, (3) exploratory analysis, (4) pattern recognition, (5) curve resolution, and/or (6) multivariate calibration techniques can be used, depending on the goal of the study.

Often applied *preprocessing* techniques for fingerprints are column centering, normalization, and peak-aligning approaches, such as correlation optimized warping (COW). The latter approaches will align the corresponding peaks in the fingerprints, in order to correct for retention time shifts. When the goal is *identification or quality control* of herbal extracts, fingerprints of samples are compared to those of reference standards, and comparison between fingerprints then is often obtained by calculating the correlation coefficient or the so-called similarity score between both. *Exploratory analysis* of the fingerprints, e.g. using principal component analysis (PCA) or cluster analysis, will give the analyst a better insight in the data in order to see whether there is a given structure or clustering tendency in the data set, for example according to a given activity, species, origin, harvest time, etc. Another interesting application is the *classification* of herbal samples, e.g. according to a given activity, species, origin, harvest time, etc. based on their fingerprints. For this goal, techniques, such as linear and quadratic discriminant analysis (LDA and QDA), classification and regression trees (CART), and partial least squares discriminant analysis (PLS-DA), can be applied. When fingerprints are measured using HPLC – Diode Array Detection (DAD) and given compounds are of interest, *curve resolution* methods can be applied to resolve overlapping peaks in the fingerprints. Finally, *multivariate calibration* approaches, such as principal components (PCR) regression and partial least squares (PLS) regression, can be applied on fingerprint data in order to model and predict a property of interest of the samples, for example, the antioxidant, antimicrobial or cytotoxic activities. When the goal is prediction of the activity for new samples, the predictive properties of the model are essential. On the other hand, when looking for potential interesting medicinal components, the predictive properties of the model are less crucial, but then peaks, contributing to the observed activity, should be localized by evaluating the regression coefficients of the multivariate models. In a next step, the substances from these relevant peaks are then identified using HPLC – Mass Spectrometry (MS). Furthermore, they should then be isolated and further examined.