



Universiteit
Leiden

The Netherlands

Comprehensive extraction and NMR-based Metabolomics : novel approaches to natural products lead finding in drug discovery

Yuliana, N.D.

Citation

Yuliana, N. D. (2011, June 9). *Comprehensive extraction and NMR-based Metabolomics : novel approaches to natural products lead finding in drug discovery*. Retrieved from <https://hdl.handle.net/1887/17704>

Version: Not Applicable (or Unknown)
License: [Leiden University Non-exclusive license](#)
Downloaded from: <https://hdl.handle.net/1887/17704>

Note: To cite this publication please use the final published version (if applicable).

Chapter 6

Comprehensive extraction: a novel tool to provide chemical diversity for metabolomics-based lead finding studies

Nancy Dewi Yuliana^{1,2}, Alfi Khatib³, Muhammad Jahangir¹, Young Hae Choi¹, Robert Verpoorte¹

¹Div. Pharmacognosy, Section Metabolomics, Institut of Biology, Leiden University, Einsteinweg 55, 2333 CC Leiden, The Netherlands.

²Dept. Food Science and Technology, Bogor Agricultural University, IPB Dramaga Campus, Bogor 16680, Indonesia.

³Center of Excellence for Food Safety Research, University Putra Malaysia, 43400 Serdang, Selangor Darul Ehsan, Malaysia

Abstract

Metabolomics-based lead finding aims at the identification of active compounds from fractionated plant extracts. A new method, namely comprehensive extraction was developed to support this approach. In this method, a powdered plant material was packed in an empty column. A gradient of solvents with increasing polarity was then run through the column. In this method the total metabolome will be obtained divided over a series of fractions, thus avoiding the problem of many overlapping compounds in the NMR - metabolomics analysis when using a classical liquid-liquid partitioning approach. The reproducibility of the method was assessed by hierarchical cluster analysis (HCA) and partial least square-discriminant analysis (PLS-DA). The results show that the comprehensive extraction is reproducible and can be applied in NMR-based metabolomics to rapidly find compounds in plants that relate to activity.

Keywords: *Orthosiphon stamineus*, *Morus alba*, comprehensive extraction, hierarchical cluster analysis, partial least square-discriminant analysis

Introduction

Metabolomics aims at the qualitative and quantitative analysis of all metabolites in an organism under specific conditions (13, 376). It has been applied in various phytochemical studies, particularly in combination with multivariate data analysis, e.g. to identify the response of plants to stress at the level of metabolites (339), quality control of herbal medicines (406), and phylogenetic studies (410). Recently this approach was also used to correlate the metabolite profile of plant extracts with their bioactivities (3, 4). However, most studies focused on quality control rather than on identification of active principles from the extract.

The problem in finding biological active compounds in a plant extract is the complexity of the mixture, hundreds of compounds are present in very different quantities. In bioassay guided fractionation one may also lose activity in the separation process in case of synergism. As we have shown before by means of analyzing different accessions of plants which exhibited different levels of activity, it was possible to identify the active compounds by means of supervised multivariate analysis of all the NMR-data of the crude extracts. Based on this, we hypothesized that by either making different extracts from the same plant or fractionation of an extract and measuring the activities of all extracts or fractions, one may identify the compounds related to activity by means of NMR-based metabolomics in combination with multivariate analysis. Ingkaninan *et al.* described the use of centrifugal partitioning chromatography for the first step fractionation of an extract before bioassays (421). However, there is no solvent capable of extracting all metabolites from a plant, because of the differences in polarity of the metabolites. To achieve the goal of extracting all metabolites we have developed a comprehensive extraction, in which the plant material is continuously extracted with a gradient of solvents of increasing polarity. The fractions obtained will in part overlap which then offers the possibility of relating activity with the signals present in the NMR spectra. This approach avoids the need of elaborate liquid-liquid partition fractionation and chromatographic separation which are the most common methods in plant-based lead finding studies.

Clustering analysis is a mathematical method to find a natural classification of a data set so that there is more similarity (less variation) within the group but less

similarity (more variation) between the groups (422, 423). The Ward's method was chosen in this study to assess the similarity between fractions. In this method, the distance between clusters is quantified as the error sum of squares (ESS). The clustering step begins with considering each sample as a cluster. The number of clusters is then reduced by computing the ESS values of each; two clusters which have the smallest increase of ESS are then merged. This step is repeated until a cluster-profile containing all objects is obtained (423). Hierarchical cluster analysis is applied in studies aiming at classifying varieties of plants based on their metabolome, e.g. between and within tomato fruits from 94 different genotypes (388) and to generate chemical finger prints of traditional herbal medicines (424, 425).

To determine the reproducibility of the comprehensive extraction method we developed, we applied HCA with Ward's method. Similarity between replicate fractions obtained from comprehensive extraction should result in a clustering of these fractions in the same group.

Materials and method

Chemicals and reagents

Methanol, *n*-hexane, acetone, ethyl acetate, HCl, NaOH, and DMSO were purchased from Biosolve BV (Valkenswaard, The Netherlands). Tris buffer was purchased from Gibco BRL (New York, NY, USA), [³H]DPCPX (8-cyclopentyl-1,3-dipropylxanthine) was from DuPont NEN, and CPA (N⁶ cyclopentyladenosine) was from RBI Inc. (Zwijndrecht, The Netherlands). Kieselguhr (calcined and purified SiO₂) was bought from Fluka Analytical/Sigma Aldrich Chemie GmbH (Steinheim, Germany). All solvents and reagents were of analytical grade.

Plant material

Orthosiphon stamineus dried leaves were obtained from van der Pijpe Drugstore. The plant materials were identified by Nancy Dewi Yuliana (Leiden University), *Morus alba* stem bark was obtained from Korean Export and Import Federation of Drugs, Seoul, South Korea and identified by Dr. Young Hae Choi (Leiden University).

Comprehensive extraction

Extraction was performed in a metal column (4.00 cm length, 1.80 cm diameter) where the mixture of 0.70 g dried powdered *O. stamineus* or *M. alba* bark and 0.10 g Kieselguhr were placed. The column was closed on the top and the bottom with fat-free cotton, and then connected to the pump (Waters 600E, Waters Chromatography Div., Milford, MA, USA). Organic solvents were sonicated and degassed, while milipore water was filtered before use. The fractions were collected in 10 mL tubes every 2 minutes by using an automatic fraction collector. Two extraction schemes were developed.

Scheme 1 (ES1): First very non-polar lipids were removed from the powdered plant material before the comprehensive extraction starts by adding: 3 mL of *n*-hexane to powdered plant material, followed by sonication for 15 minutes, and then the solvent was removed. This was repeated 2 times with 2 mL of *n*-hexane. The plant material was dried under N₂, mixed with Kieselguhr, and loaded into the extraction column. The combination of solvents used for the comprehensive extraction was ethyl acetate-methanol (1:1) as solvent A, and methanol-water (1:1) as solvent B. The flow rate was 4 mL/min and fractions were collected every two minutes. Thirty fractions of 8 mL each were obtained at the end of the extraction. One and half mL of the extract was sampled from each fraction to be used in the bioassay. The remaining volume (6.5 mL) was used for NMR measurement where methanol-*d*₄ was used as a solvent. The extraction was performed in three replicates.

Scheme 2 (ES2): No defatting was applied in this method and the plant material mixed with Kieselguhr was directly loaded on the column. The combination of solvents used for the comprehensive extraction was *n*-hexane (A), acetone (B), and water (C). The flow rate was 3 mL/min and fractions were collected every two minutes. Twenty fractions were obtained at the end of the extraction. Every fraction was dried, and dissolved in DMSO-*d*₆, and subsequently used for ¹H NMR measurement and the adenosine A1 bioassay. The extraction was performed in three replicates.

Table 1. Extraction scheme 1 (ES1): T= time (min), F = flow rate (mL), A = ethyl acetate-methanol (1:1), B = methanol-water (1:1), Extraction scheme 2 (ES2): T= time (min), F = flow rate (mL), A= *n*-hexane, B = acetone, C = water.

ES1			ES2		
T (min)	F (mL)	Solvent gradient	T (min)	F (mL)	Solvent gradient
0-10	4	100% A	0-12	3	100% A
10-50	4	100% A to 100% B	12-32	3	100% A to 100% B
50-60	4	100% B	32-44	3	100% B
-	-	-	44-64	3	100% B to 100% C
-	-	-	64-80	3	100% C

NMR measurement

The NMR measurement and data analysis were performed according to Kim *et al.* (426). NMR spectra were recorded on a 500-MHz Bruker DMX 500 Spectrometer (Bruker, Karlsruhe, Germany) operating at a proton NMR frequency of 500.13 MHz. MeOH-*d*₄ was used as the internal lock. Each ¹H NMR spectrum consisted of 128 scans requiring 10 min and 26 s acquisition time with the following parameters: 0.16 Hz/point, pulse width (PW) = 30° (11.3 μs), and relaxation delay (RD) = 1.5 s. A pre-saturation sequence was used to suppress the residual H₂O signal with low power selective irradiation at the H₂O frequency during the recycle delay. The FIDs were Fourier transformed with LB = 0.3 Hz. The resulting spectra were manually phased and baseline corrected, and calibrated to DMSO-*d*₆ at 2.49 ppm or methanol-*d*₄ at 3.33 ppm, using XWIN NMR (version 3.5, Bruker).

Data analysis

The ¹H NMR spectra were automatically reduced to ASCII files. Bucketing was performed by AMIX software (Bruker). Spectral intensities were scaled to total intensity and reduced to integrated regions of equal width (0.04) corresponding to the region of δ 0.3–10.0. For ES1 samples, the regions of δ 4.75–4.9 and δ 3.28–3.40 were excluded from the analysis because of the residual signal of water and methanol-*d*₄,

respectively. For ES2 samples, the region of δ 2.44 – 2.56 and δ 3.28–3.36 was excluded from the analysis because of the residual signal of DMSO- d_6 and water (in DMSO- d_6), respectively. Hierarchical clustering analysis (HCA) based on Ward's method and PLS-DA were performed with the SIMCA-P software (version 12.0, Umetrics, Umeå, Sweden) with scaling based on the Pareto method. Identification of primary metabolites in all fractions and methoxy flavonoids from *O. stamineus* were performed by comparing NMR data previously reported by our group (339, 420). Identification of diterpenoids from *O. stamineus* was performed by comparing with NMR data from a previous report (427), similarly alkaloids and prenylflavonoids from *M. alba* were identified (303, 428, 429).

Results and discussion

The most popular method in plant based drug discovery is bioassay guided fractionation of a plant extract in which single solvent extraction followed by solid supported chromatography is widely used. The problem with this approach is that one needs to make more than one extract of a plant to cover most of the metabolites. Moreover, metabolites may interact with the solid phase and are not eluted. It is also a time consuming and costly approach. To identify active metabolites from spectra or chromatograms of crude extracts obtained from classical extraction methods is not easy since a huge number of overlapping peaks is obtained. Schripsema *et al.* (430) mentioned that to reduce the risk of missing active compounds for bioactivity tests, one needs to develop a more elaborate extraction scheme yielding several extracts from one plant which each extract contains a limited number of compounds. Therefore we developed a method which combines extraction and fractionation in one. The aim is to cover a wide range of metabolites with good resolution without the use of chromatographic supports. We named this method comprehensive extraction. It is based on continuous extraction of plant material with a mixture of miscible solvents of increasing polarity, and collecting the extract in fractions. No adsorbents material was filled into the column except a small quantity of kieselguhr which is mixed with the powdered plant material to avoid clogging of the column. The fractions were collected according to a specific time interval and further subjected to NMR.

Although nonpolar solvents are preferable to extract metabolites as previously mentioned, it raises the solubility problem in various bioassay systems for which aqueous solutions are needed (430). Moreover, some nonpolar compounds give unselective responses in certain receptor or enzymatic based bioassays, e.g. unsaturated fatty acids for adenosine and opioid receptors (249). To overcome these problems, two extraction schemes were developed (ES1 and ES2). In ES1, the first step was a defatting process by hexane extraction. Comprehensive extraction by using a relatively polar solvent system was then applied. The solvent gradient started from ethyl acetate-methanol (1:1) as nonpolar and methanol-water (1:1) as the polar solvent. Since the fractions obtained consist of relatively polar metabolites, methanol- d_4 was used for NMR measurement. In ES2, no defatting step was applied but a solvent combination with a wider range of polarity (*n*-hexane-acetone-water) was used and DMSO- d_6 was used for NMR measurement. The described gradients, extraction times, and flow rate came out as optimal after testing different schemes. One clear limitation we found is for the flow rate as clogging of the column occurs at higher pressure needed for higher flow rates.

***Morus alba* stem bark ES1 extraction**

The result of the HCA is shown in Figure 1. It was found that all fractions can be divided into 4 main groups. Good reproducibility is achieved when all fractions are in the same group with their replicates. In general, most of the *M. alba* fractions fulfilled this criteria (Table 2), although a few outliers are also found, such as fraction 11a located in group 2 while their replicates are in group 3; fraction 21c and 22a located in group 4 while their respective replicates are in group 3. With the SIMCA software, it is possible to perform PLS-DA analysis based on the grouping found in the HCA extraction scheme (Fig. 2). This allows to analyze the reason for the grouping, and can also be used to validate the grouping. Validation by response permutation showed that these 4 groups are significant since all have intercepts of R^2Y less than 0.3-0.4 and intercepts of Q^2Y not exceeding 0.05.

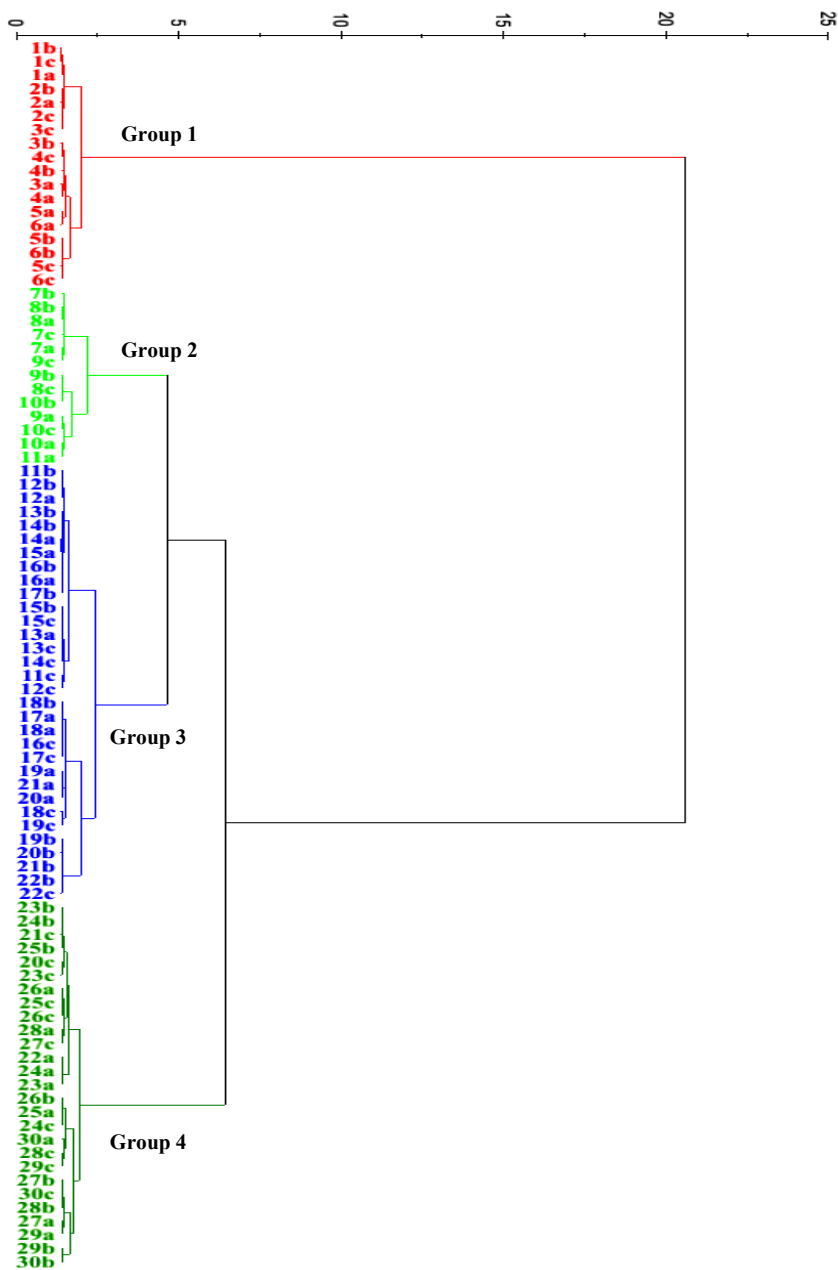
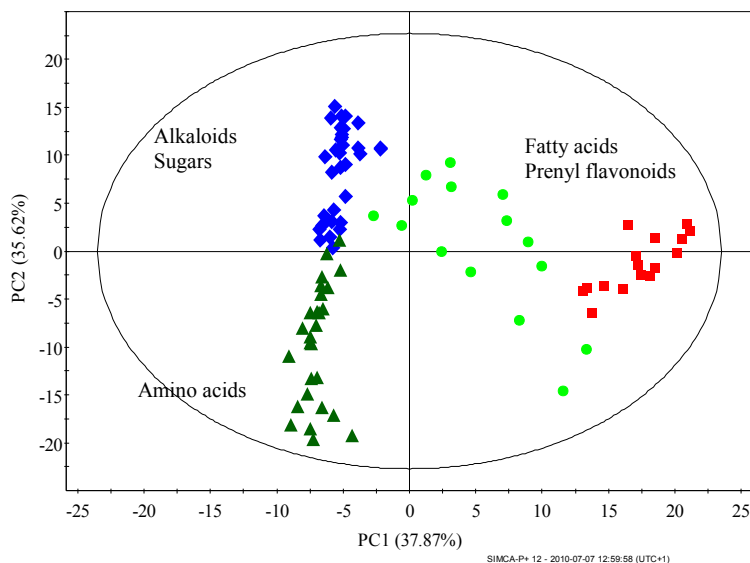


Figure 1. Dendrogram of clustering of *M. alba* stem bark fractions obtained from comprehensive extraction ESI using Ward's method. Fractions are labeled by numbers with a, b, and c represent the replications.

Table 2. Grouping of the *M. alba* fractions obtained from comprehensive extraction ES1.

Group	Members
1	1a-c, 2a-c, 3a-c, 4a-c, 5a-c, 6a-c,
2	7a-c, 8a-c, 9 a-c, 10a – c, 11a
3	11b, 11c, 12a-c, 13a-c, 14a-c, 15a-c, 16a-c, 17a-c, 18 a-c, 19a-c, 20a-c, 21a-b, 22b-c
4	21c, 22a, 23a-c, 24a-c, 25a-c, 26a-c, 27a-c, 28a-c, 29a-c, 30a-c

**Figure 2.** The score plot of PLS-DA. ■ Group 1 ● Group 2 ◆ Group 3 ▲ Group 4

From the PLS-DA loading plot, signals of prenylflavonoids and fatty acids are found to be mainly in group 1 and to lesser extent in some of fractions in group 2. Alkaloid and sugar signals are higher in group 3. Fraction 11a contains higher amount of fatty acids than its replicates, therefore it is clustered in group 2. Fraction 21c and 22a are clustered in group 4 because alkaloid signals are found to be less in these two fractions than their respective replicates. However, these outliers are actually grouped not so far away from their replicates. Defatting step which was performed before the

extraction seems to be inefficient since the effect of fatty acids signals in the PLS-DA loading plot is very dominant.

***Orthosiphon stamineus* ES1 extraction**

The dendrogram of HCA of *O. stamineus* is presented in Figure 3. At the same distance (4.45), three groups were obtained. The PLS-DA analysis was performed based on this grouping and from the score plot all fractions are clearly clustered in 3 groups (Fig. 4). From the response permutation test, these 3 groups were found to be significant.

Compounds responsible for the separation of these 3 groups were identified based on the loading plot profiles. Fatty acid and diterpenoid signals are found to be higher in group 1, while methoxy flavonoids are higher in group 2. In group 3 polar metabolites such as glycoside of flavonoids, amino acids and sugars are predominant. Similar to *M. alba*, fatty acids signals were found to strongly affect the loading plot, indicating that the separate defatting step is not so efficient.

***Orthosiphon stamineus* ES2 extraction**

The chemical profile of *O. stamineus* fractions obtained from ES2 showed better reproducibility as compared to ES1 fractions (Fig. 5). Fractions are clustered in the same group with their replicates (Table 4). Three significant groups validated by response permutation were obtained (Fig. 6). The loading plot of PLS-DA shows that fractions in group 1 contain most non-polar metabolites such as fatty acids, diterpenoids, and methoxy flavonoids. Signals of phenolics and sugars are found together in group 2, indicating the presence of glycoside flavonoids. Group 3 contains mainly polar metabolites such as sugars and amino acids. As compared to ES1, fatty acids signals are not so dominant in the PLS-DA loading plot. These metabolites were found to be higher in the few first fractions and then gradually reduced in the next ones, indicating that it is more efficient to include defatting step in the extraction scheme.

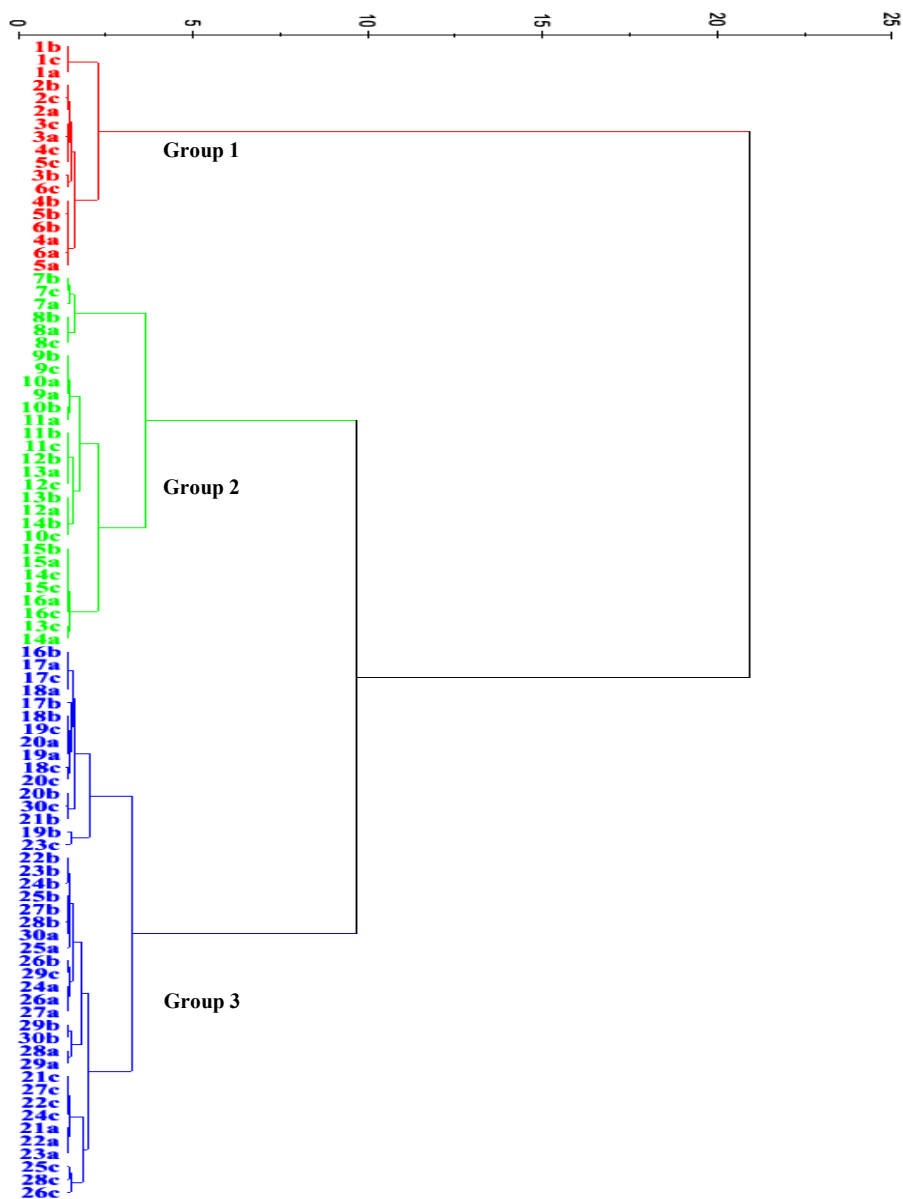


Figure 3. Dendrogram of clustering of *O. stamineus* fractions obtained from comprehensive extraction ES1 using Ward's method. Fractions are labeled by numbers with a, b, and c represent the replications.

Table 3. Grouping of *O. stamineus* fractions obtained from comprehensive extraction ES1.

Group	Members
1	1a-c, 2a-c, 3a-c, 4a-c, 5a-c, 6a-c
2	7 a-c, 8 a-c, 9 a-c, 10a – c, 11a-c, 12a-c, 13a-c, 14a-c, 15a-c, 16a, 16c
3	16b, 17a-c, 18a-c, 19a-c, 20a-c, 21a-c, 22a-c, 23a-c, 24a-c, 25a-c, 26a-c, 27a-c, 28a-c, 29a-c, 30a-c

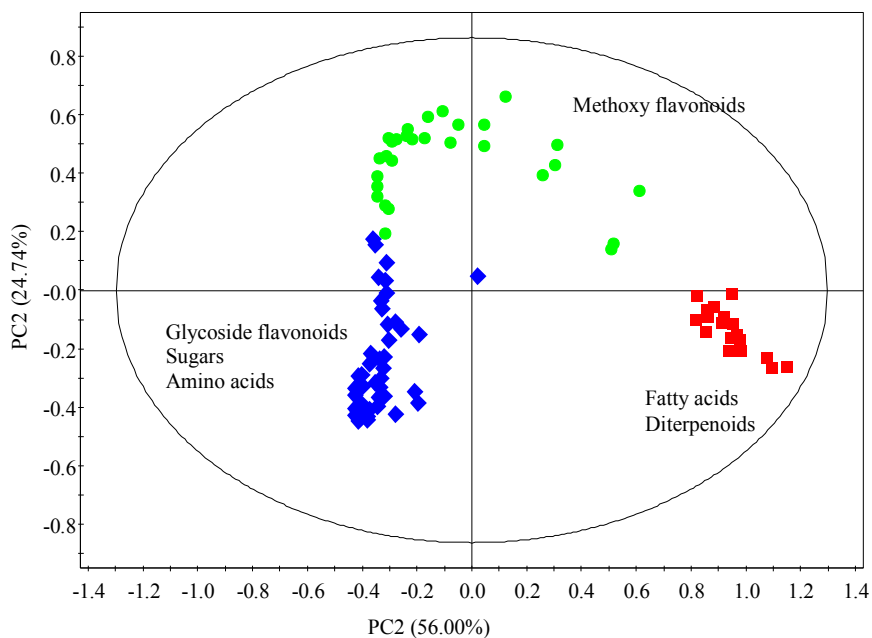


Figure 4. The score plot of PLS-DA. ■ Group 1 ● Group 2 ◆ Group 3

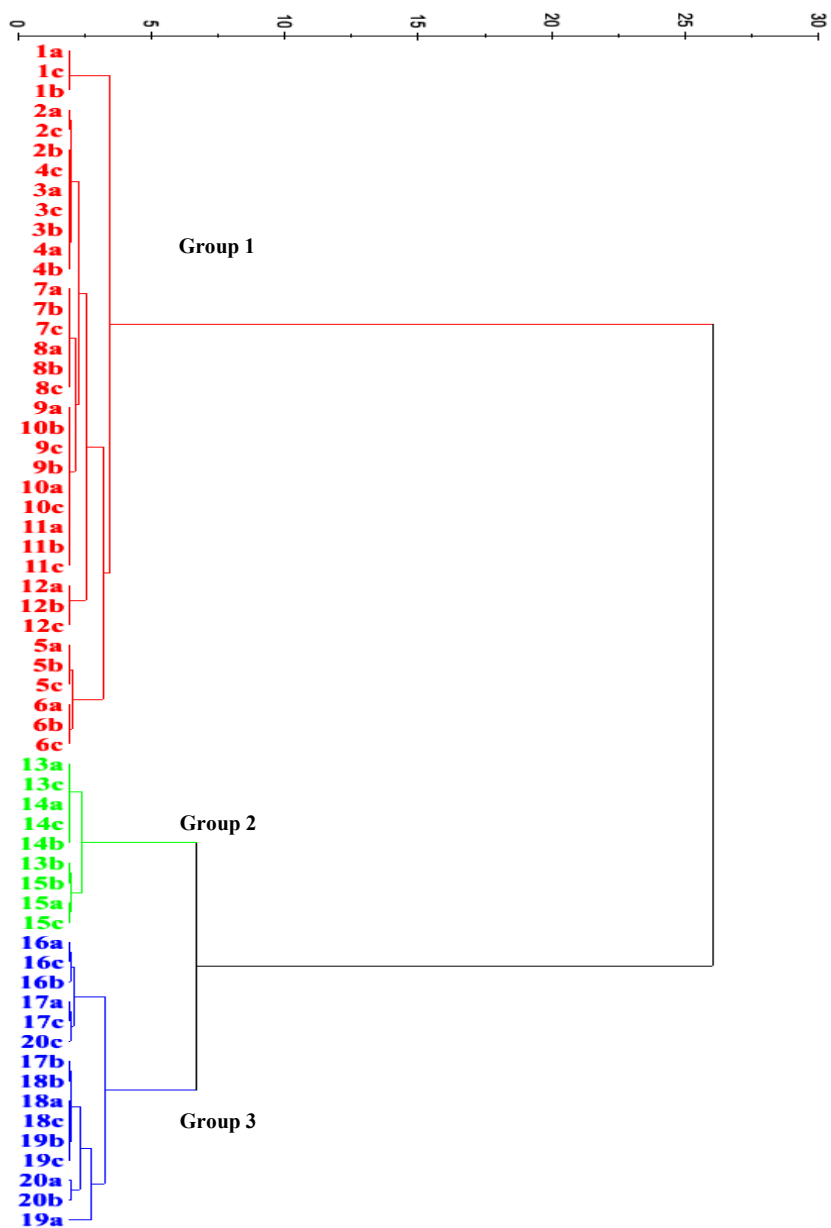
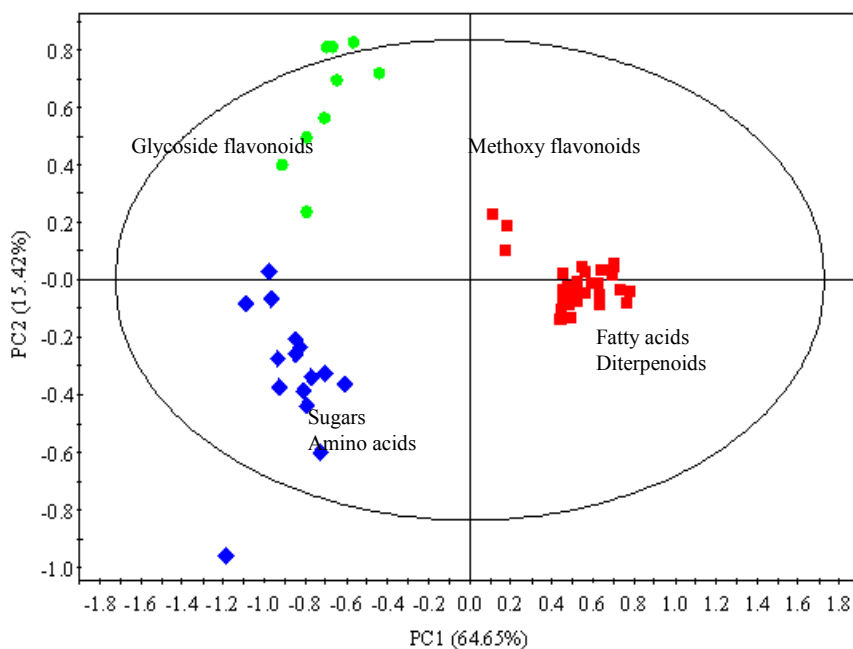


Figure 5. Dendrogram of clustering of *O. stamineus* fractions obtained from comprehensive extraction ES2 using Ward's method. Fractions are labeled by numbers with a, b, and c represent the replications.

Table 4. Grouping of *O. stamineus* fractions obtained from comprehensive extraction ES2.

Group	Members
1	1a-c, 2a-c, 3a-c, 4a-c, 5a-c, 6a-c, 7 a-c, 8 a-c, 9 a-c, 10a – c, 11a-c, 12a-c,
2	13a-c, 14a-c, 15a-c
3	16 a-c, 17a-c, 18a-c, 19a-c, 20a-c

**Figure 6.** The score plot of PLS-DA. ■ Group 1 ● Group 2 ◆ Group 3

Conclusion

As compared to liquid-liquid partition or simple extraction coupled to chromatographic separation, comprehensive extraction is faster and more efficient in

identifying a wide range of metabolites. Non-polar to polar compounds are eluted subsequently according to solvent gradient, resulting in more clear spectra of chemical profile of the plant material which are easier to interpret only by ^1H NMR spectra. The defatting step which was included in ES2 by using *n*-hexane as the first solvent in comprehensive extraction is more efficient than performing a separate *n*-hexane extraction before the comprehensive extraction (ES1). The use of wider range of solvent polarity (ES2) gave better reproducibility than the narrower range (ES1). Based on the result of the hierarchical cluster analysis coupled to PLS-DA, the reproducibility and the ability to produce chemical variation with this new extraction method seems a promising approach to relate compounds to activity.

As shown in chapter 5, activity can be related to signals of metabolomics analysis of different fractions. For that approach some variability in the fractionation is not a problem, and may even be an advantage as correlations are based on matching activity and concentration of active compound(s). The comprehensive extraction is thus an efficient extraction and fractionation method for identification of biological active compounds in crude extracts.