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

Yuliana, N.D.

Chapter 7

Comprehensive extraction integrated with NMR metabolomics as a new way of bioactivity guided identification of biologically active compounds: compounds binding to adenosine A1 receptor from *Orthosiphon stamineus* Benth leaves

Nancy Dewi Yuliana[1,2], Alfi Khatib[1,3], Muhammad Jahangir[1], Young Hae Choi[1], Robert Verpoorte[1]

[1]Div. Pharmacognosy, Section Metabolomics, Institut of Biology, Leiden University, Einsteinweg 55, 2333 CC Leiden, The Netherlands.
[2]Dept. Food Science and Technology, Bogor Agricultural University, IPB Dramaga Campus, Bogor 16680, Indonesia.
[3]Center of Excellence for Food Safety Research, Faculty of Food Science and Technology, University Putra Malaysia, 43400 Serdang, Selangor Darul Ehsan, Malaysia

Abstract

A novel approach by using NMR based metabolomics coupled to multivariate data analysis was used to identify bioactivity-related metabolites in a plant extract. A comprehensive extraction method was developed by using gradients of two different solvent combinations to provide a broad-ranging NMR metabolite profile of *Orthosiphon stamineus*. Different NMR solvents were used for each combination. Partial least square (PLS) and orthogonal partial least square (OPLS) analysis were applied to study the correlation between the chemical profiles of *Orthosiphon stamineus* fractions and its adenosine A1 binding activity. The combination of *n*-hexane, acetone, and water for the extraction, and DMSO-d_6 for NMR measurements was found to give more chemical variation than the gradient of ethyl acetate/methanol (1:1) and methanol-methanol/water (1:1), with methanol-d_4 as NMR solvent. The first method thus offers better possibilities for identification of active compounds. Identification of compounds was performed by cross-checking ^{1}H NMR, J-res, COSY and HMBC NMR data from flavonoids previously reported to be active to the adenosine A1 receptor with the spectra of the active fraction obtained from comprehensive extraction. Orthogonal

partial least square (OPLS) analysis was found to be the most suitable tool to study the chemical-profile-activity correlation. The approach reported here is a rapid way for developing new lead compounds from plants for drug development since it allows rapid identification of active compounds from crude plant extracts.

Keywords: *Orthosiphon stamineus,* comprehensive extraction, partial least square analysis, orthogonal partial least square analysis, adenosine A1 receptor

In preparation to be submitted to Analytical Chemistry: Yuliana *et al*., Comprehensive extraction method integrated with NMR metabolomics - a new bioactivity screening procedure for plants: adenosine A1 receptor binding compounds in *Orthosiphon stamineus* Benth.

Introduction

The introduction of high throughput screening in the 1990's aimed to shorten the drugs discovery route. To fully use its high potential, this technology requires a large number of compounds to screen, in the order of thousands of samples per day (334). Thus, increasing the number of chemicals for initial screening is high on the agenda of pharmaceutical companies. High throughput synthesis and combinatorial chemistry were developed to address this demand. However, these technologies failed to increase new lead compounds. Sorafenib, a multikinase inhibitor indicated for advanced renal cancer, is so far the single synthetic drug derived from combinatorial chemistry approved by the Food and Drug Administration for clinical use (1).

Apparently, not providing sufficient relevant chemical diversity is one of the reasons for the failure of the aforementioned new technologies to deliver new lead compounds. On the other hand, it is well known that an enormous molecular diversity and biological functionality are two important features which distinguish plant extracts as a drug source from synthetic chemicals (345). Looking back to the past two decades, of the 1184 novel medicines developed between 1981 and 2006, 5% were NP, 23% were NP derivatives and 24% were synthetic products developed on the basis of a NP model indicating that NP is an important source of novel leads for therapeutical drugs (1).

However, a natural products-based drug discovery project also poses some challenges, mostly connected with the presence of an active compound as a complex matrix with all kind of compounds (335). An elaborative purification to identify and to isolate active compounds is thus needed. The possibility of antagonism or synergism between metabolites present in the extracts, and the fact that some common plant products have been found to be active in a number of test systems such as unsaturated fatty acid on adenosine, opioid, and GABA receptors, or tannins on enzymes based bioassays (249, 431), pose a major challenge.

Novel approaches which can address these issues are urgently needed. Such a novel approach is the use of metabolomics coupled to multivariate data analysis to correlate the chemical profiles of the plant materials with bioactivity profiles, allowing the identification of metabolites that are related to activity. Recently, some studies

introducing this method have been reported. For example by using metabolomics combined with projection-based multivariate data analysis (PCA, PLS-DA, PLS) (2-5).

Metabolomics allows a systematic study of complex mixtures such as plant extracts. In this approach, the metabolites can be linked to results from biological testing without the need of isolating the active principles. Reproducibility is the most important criteria for developing a metabolomics technology platform, and in that respect NMR is the most suited method even though its sensitivity is not as high as MS and chromatography based metabolomics platforms (1 μM – 1 mM in NMR tube) (12, 13). The simple and fast sample preparation, short measurement time, plus the possibility to elucidate structures of known or unknown compounds in a complex mixture using advanced two-dimensional (2D) NMR methods, are further advantages (13).

To interprete high dimensional data resulting from metabolomics analysis, multivariate data analysis is commonly applied. Most metabolomics data analysis methods are based on the classification of samples into different groups (e.g. treatment, genotype), both by unsupervised, e.g. principal component analysis (PCA), hierarchical cluster analysis (HCA), or supervised data analysis methods, e.g. PLS-DA, PLS, OPLS (379). It is possible to use multivariate data analysis (MVDA) to make a regression modeling between two blocks of data, usually denoted as X and Y. In metabolomics based natural product studies, X may represent spectroscopic or chromatographic signals from different metabolites present in plant extracts which are sampled at regular time intervals and Y represents the responses, e.g. quality of product, bioactivity, yield, composition. The model then can be used to predict Y from X which is obtained from new observations. The most common MVDA method for this modeling is the partial least square (PLS) method (14). Some extended versions of PLS to improve the predictability power were developed recently, such as orthogonal signal correction (OSC), orthogonal-PLS (OPLS), and orthogonal-orthogonal-PLS (O2PLS) (393, 394). For studying plants for the presence of biologically active compounds, such an approach requires that one has either different accessions with different biological activity, or otherwise different extracts or fractions of the plant material that have different activity and different (partly overlapping) chemical profiles.

Here an integrated approach to identify active compounds from plant extracts based on NMR metabolomics and multivariate data analysis is presented. A novel comprehensive extraction method was used rather than a conventional simple extraction coupled to conventional liquid-liquid partition. The aim is to cover a wide range of metabolites with good resolution without the use of chromatographic supports. Partial least square and orthogonal- partial least square were used to find possible correlations between metabolite profiles and bioactivity. Statistical validations (CV-ANOVA, permutation test, and external validation by dividing the data into calibration group and test group) were performed to select the most reliable method. Some evaluation tools available in the corresponding multivariate data methods were used to select important signals. Based on that, compounds highly correlated with the activity were detected and identified by means of 2D NMR data (J-resolved, COSY, and HMBC). Chemical validation was also accomplished by cross-checking the presence of signals of previously isolated active pure compounds in one of the fraction which showed a good activity profile.

Materials and methods

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), [^{3}H]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 SiO_2) was bought from Fluka Analytical/Sigma Aldrich Chemie GmbH (Steinheim, Germany). All solvents and reagents were of analytical grade.

Plant material

Orthosiphon stamineus was obtained from van der Pigge Drugstore. The plant materials were identified by one of the authors (Nancy Dewi Yuliana), and a voucher specimen (ORST-Fcog-NL-230506) was deposited in the Division of Pharmacognosy, Institute of Biology, 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* 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 Waters 600E pump (Waters Chromatoghraphy Div., Milford, MA, USA). Organic solvents were sonicated and degassed, while 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 as has been described in chapter 6. For ES1, 1.5 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. For ES2, every fraction was then dried, and dissolved in DMSO-d_6, subsequently, and further subjected to ^{1}H NMR measurement and adenosine A1 bioassay test. Each extraction was performed in three replicates and each extract was tested for the bioactivity in duplo.

NMR measurement

The NMR measurement was performed according to the method described by 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. Methanol-d_4 was used as the internal lock. Each ^{1}H 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 H_2O signal with low power selective irradiation at the H_2O 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). Two dimensional J-resolved NMR spectra were acquired using 8 scans per 128 increments for F1 and 8 k for F2 using spectral widths of 5000 Hz in F2 (chemical shift axis) and 66 Hz in F1 (spin–spin coupling constant axis). A 1.5 s relaxation delay was employed, giving a total acquisition time of 56 min. Datasets were zero-filled to 512 points in F1 and both

dimensions were multiplied by sine-bell functions (SSB = 0) prior to double complex FT and were recorded on a 600 MHz Bruker DMX-600 spectrometer (Bruker). The ^{1}H-^{1}H correlation (432) and heteronuclear multiple bonds coherence (HMBC) spectra were acquired with 1.0 sec relaxation delay, 6361 Hz spectral width in both dimensions. Window function for COSY spectra was sinebell (SSB = 0). The HMBC spectra were obtained with 1.0 sec relaxation delay, 30183 Hz spectral width in F2 and 27 164 Hz in F1. The optimized coupling constants for HMBC were 18 Hz. All NMR experiments were performed at 25°C. Chemical shifts (δ) are given in ppm.

Data analysis

The ^{1}H 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 – 4.90 and δ 3.28 – 3.34 were excluded from the analysis because of the residual signal of H_2O 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 H_2O (in DMSO-d_6), respectively. Partial least square projection to latent structure (PLS) and orthogonal partial least square analysis (OPLS) were performed with the SIMCA-P software (version 12.0, Umetrics, Umeå, Sweden) with scaling based on the Pareto method.

Adenosine A1 receptor assay

The assay was performed as previously described by Chang *et al*. (290) except that the volume of the total mixture in the assay was 200 μL. Membranes were prepared from Chinese hamster ovary (CHO) cells stably expressing human adenosine receptors by a method previously described by Dalpiaz *et al.* (291). More details on the bioassay method were described in chapter 2.

Results and discussion

Comprehensive extraction

We developed a method which combines extraction and fractionation in one. We named this method comprehensive extraction. It is based on continuous extraction of plant material with a mixture of solvents of increasing polarity, and collecting the extract in fractions. In comprehensive extraction, the solvent is delivered directly into the column filled with powdered plant material mixed with kieselguhr. In ES1, we focused on relatively polar metabolites. The solvent gradient started from ethyl acetate-methanol (1:1) as nonpolar part to methanol-water (1:1) as a polar part. In ES2 a wider metabolite profile was obtained by using a solvent gradient *n*-hexane-acetone-water. Apparently ES2 gives a better chemical diversity as can be observed from the bioactivity profile which shows peaks of activity in both nonpolar (fraction 6, 7, 8) and polar (fraction 13, 14, 15) area (Fig. 1B). While in ES1 using a more polar solvent mixture, the activity is concentrated in the first 7 fractions (Fig. 1A).

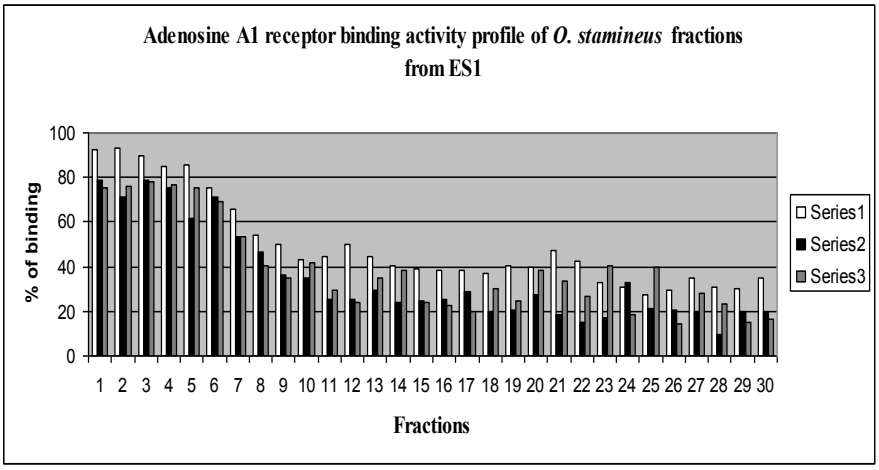
Multivariate data analysis

While PCA is an unsupervised data analysis method to get a general overview about systematic variation in data table X, PLS is an extended version of PCA. It tries to fit two PCA-like models; X (factors or predictors) and Y (response) at the same time and simultaneously tries to establish linear or polynomial relationship between these two blocks of matrixes. The aim is to model X and Y, and to predict Y from X (15). Score plots of the PLS first 3 components for ES1 fractions shows a group of active fractions separated from the non-active ones in PC1 (Fig. 2A), while for ES2 fractions there are 2 groups of active fractions; the non-polar ones separated in PC1 while the polar ones separated in PC2 (Fig. 2B).

Orthogonal-PLS (OPLS) is an extension of PLS which divides systematic variation in X-block into two; one predictive component which models the correlation between X and Y, and orthogonal component(s) which expresses X-variation unrelated to Y (15, 433). As expected, active fractions from both extraction schemes are clearly separated from non-active ones in the predictive component (Fig. 2C). Particularly, the

two active fractions from ES2 are now clustered together at the same score of a predictive component, separated from the non-active ones (Fig. 2D).

A.



B.

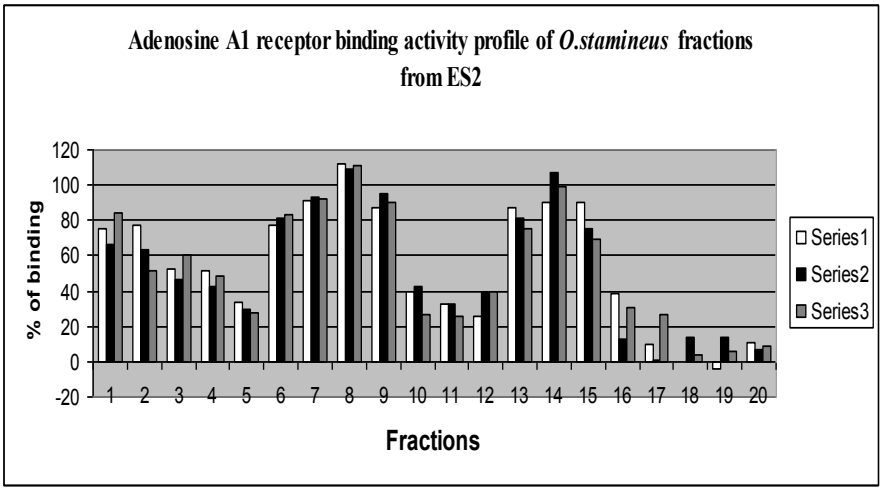


Figure 1. A. Adenosine A1 binding activity profile of fractions obtained from extraction scheme 1 (ES1): Different bar colours represent replications. **B.** Adenosine A1 binding activity profile of fractions obtained from scheme 2 (ES2).

Statistical validation

The most common ways to interprete the multivariate model are by means of R^2Y and cumulative Q^2Y after cross validation (15). Cross validation testing evaluates the predictive power of the model. The R^2Y shows the amount of Y variables explained by the model after cross validation and gives an overview about the fitting of the model, while cumulative Q^2Y gives information about the predictive quality of the model. All values close to 1 resembles a good model (15). Overall, as presented in Table 1, all models show good performance with R^2Y and Q^2Y values above 0.80.

The PLS models were further validated using response of permutation test through 100 permutations. The permutation test assesses the statistical significance of the estimated predictive power previously calculated by cross validation test (15). The test compares the distribution of R^2Y and Q^2Y when the Y-data is randomly permuted (while X-data are left intact), with R^2Y and Q^2Y value of the non-permuted Y model (434). The validation plot is constructed with the Y-axis representing R^2Y and Q^2Y of the original and permutated models, while the X-axis represents correlation coefficients between permutated and original models. Regression lines were then fitted among R^2Y and Q^2Y points. To be a valid model, the original models should have higher R^2Y and Q^2Y value than the permutated models. To achieve this, Errikson *et al*. mentioned that the intercepts of R^2Y should be less than 0.3-0.4, while Q^2Y should not exceed 0.05 (15). Both PLS models from ES1 and ES2 are shown as good models as the intercept of R^2Y are 0.06 and 0.18 respectively, intercept of Q^2Y are -0.20 and -0.37, respectively (Fig. 3A – 3B).

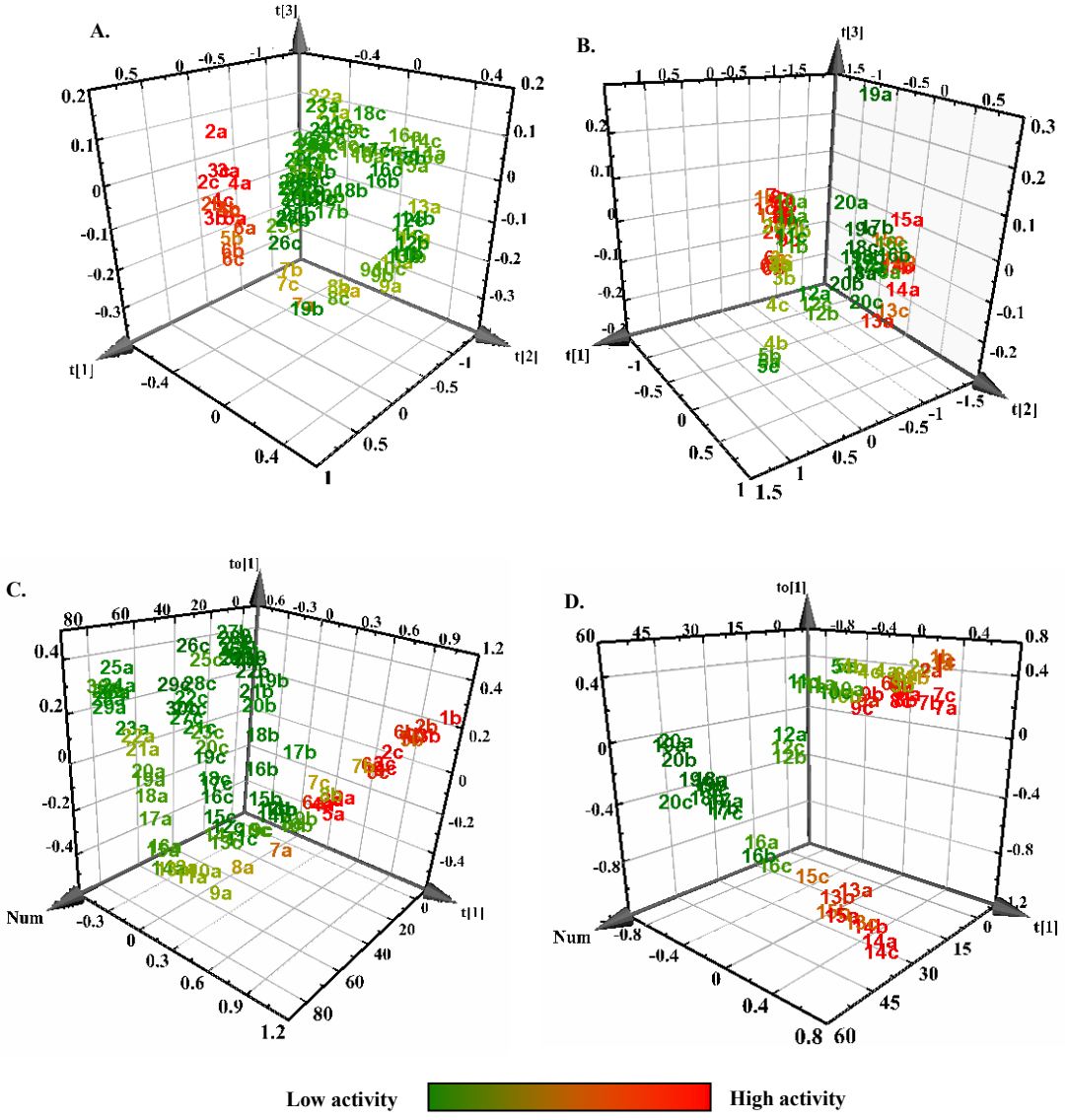


Figure 2. Scores scatter plot of PLS and OPLS of fractions obtained from ES1 and ES2. Fractions are labeled by number with a-c representing replications and coloured in gradient from green to red representing the low to the high activity. Active fractions are well separated from the non-active ones: **A.** The first 3 components of PLS of ES1 fractions (R^2X_1=56.1%, R^2X_2=23.9%, R^2X_3=2.9%) **B.** The first 3 components of PLS of ES2 fractions (R^2X_1=58.7%, R^2X_2=21.4%, R^2X_3=2.3%) **C.** The predictive and first orthogonal component of

OPLS of ES1 fractions (R^2X=55.3%, $R^2XOrthogonal_1$=24.8%) **D.** The predictive and first orthogonal component of OPLS of ES2 fractions (R^2X=19.2%, $R^2XOrthogonal_1$=51.4%).

External validation was further conducted to assess the performance of the models. Each group of the ES1 and ES2 data was divided into two: calibration group and test group. The accuracy of calibration and test groups is assessed by the value of root-mean-square errors of estimation (RMSEE) and prediction (RMSEP), respectively, which are expressed in units of the original measurement (%). The ES1 PLS and OPLS are shown as good models with RMSEP and RMSEE not so much different from each other (Table 1). The same applies for ES2 models although PLS shows a poorer model as compared to OPLS (Table 1). Overall, ES1 model gives better external validation results as can be seen from lower values of RMSEP and RMSEE.

The CV-ANOVA has been used to test the significance of PLS and OPLS models and the results were found to be consistent with other validation methods such as response permutation test and external validation (434). Both PLS and OPLS models from ES1 (Table 1A) and ES2 (Table 1B) are shown as significant models according to CV-ANOVA validation with p value less than 0.05.

Table 1. Statistical validations (CV-ANOVA and external validation): A. ES1 B. ES2

A.

MVDA method	ES1				
	R^2Y	Q^2Y	p CV-ANOVA	RMSEP	RMSEE
PLS	0.90	0.87	1.04×10^{-31}	8.40	8.44
OPLS	0.85	0.84	2.09×10^{-33}	8.52	8.74

B.

MVDA method	ES2				
	R^2Y	Q^2Y	p CV-ANOVA	RMSEP	RMSEE
PLS	0.84	0.75	1.63×10^{-9}	23.12	21.45
OPLS	0.85	0.81	4.67×10^{-13}	14.79	12.40

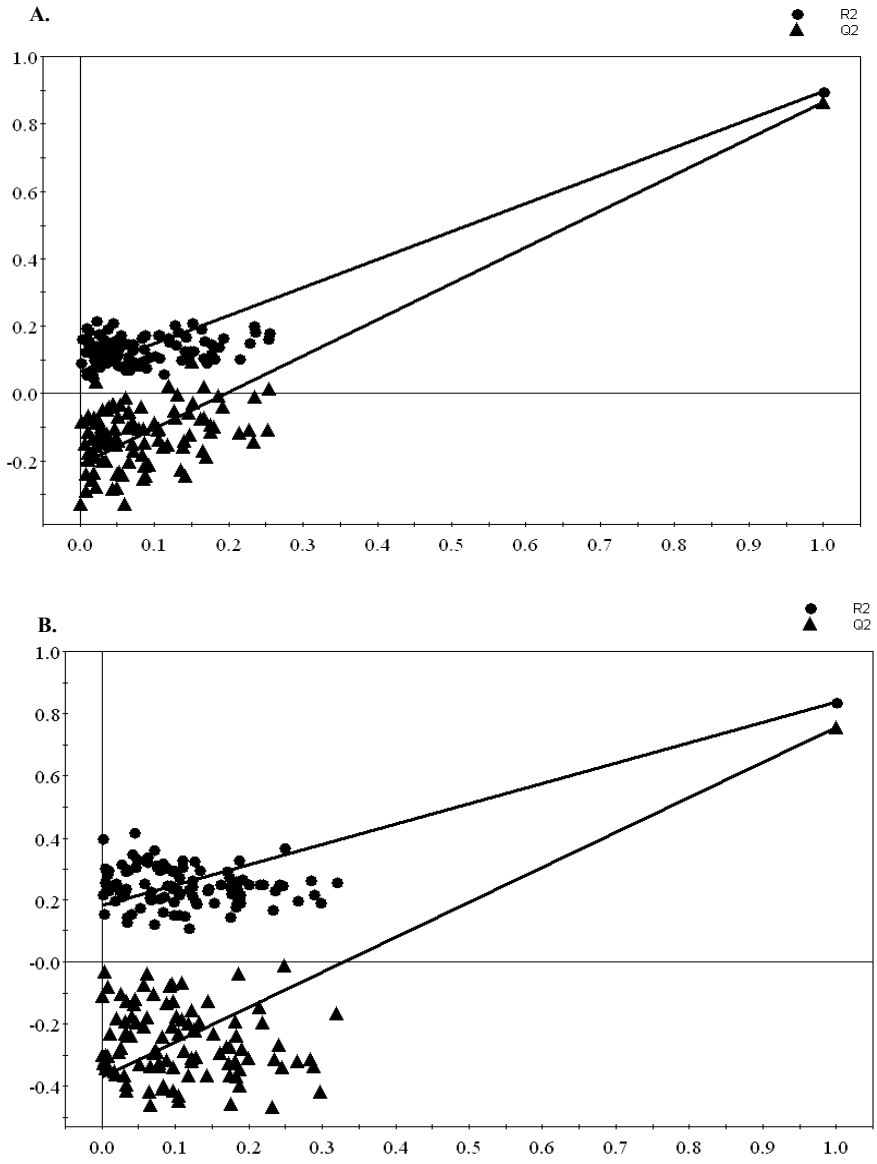


Figure 3. Response of permutation test **A**. PLS of ES1 fractions: R^2Y (0.0, 0.06), Q^2Y (0.0, -0.20), 100 permutations **B**. PLS of ES2 fractions from extraction: R^2Y (0.0, 0.18), Q^2Y (0.0, -0.37), 100 permutations.

Model interpretation

We have chosen the OPLS model to select which signals are important for the activity. By partitioning the uncorrelated variations orthogonally from the predictive ones, OPLS is easier to interprete. Based on the loading bi-plot profile, ES1 is strongly affected by fatty acids signals (Fig. 4A), while ES2 seems more interesting since more methoxy and aromatic signals are close to the activity (Fig. 4B). Interesting signals in ES2 are laid between δ 1 – 2 (**1**), δ 3.70 – 4.00 (**2**), δ 4.5 – 5 (**1**) and δ 6.80 – 8.00 (**3**), which are close to the activity (upper area of the plot). Area **1** represents typical signals for unsaturated fatty acids and diterpenoids, area **2** is methoxy signals of flavonoids, and area **3** for other flavonoids signals in the aromatic region.

It has been reported that unsaturated fatty acids, which are ubiquitously present in plants extract, bind unspecifically to adenosine A1 receptors (249). These compounds may mask the real active principles and mislead the identification step. In the method proposed in this study, it can be clearly observed if the activity comes from false positive compounds or from novel actives. However, the choice of method applied in defatting step should be taken into consideration. To include a defatting step into the extraction process by using hexane (ES2) was shown to be the better choice than using a separate defatting step (ES1). By comparing the loading bi-plots of ES1 and ES2, it can be seen that by observing ES1 loading bi-plot may lead to the wrong conclusion that adenosine A1 binding activity of the *O. stamineus* extract was only due to the presence of fatty acids. Strong signals of fatty acids suppressed signals of other compounds which are actually close to the activity, such as methoxy signals and other flavonoids signals in the aromatic region. While in ES2 loading bi-plot, beside fatty acids, these flavonoids signals, particularly methoxy signals can be identified as important for the activity. Apparently separate defatting does not extract all the unsaturated fatty acids thus in ES1 eluted together with flavonoids.

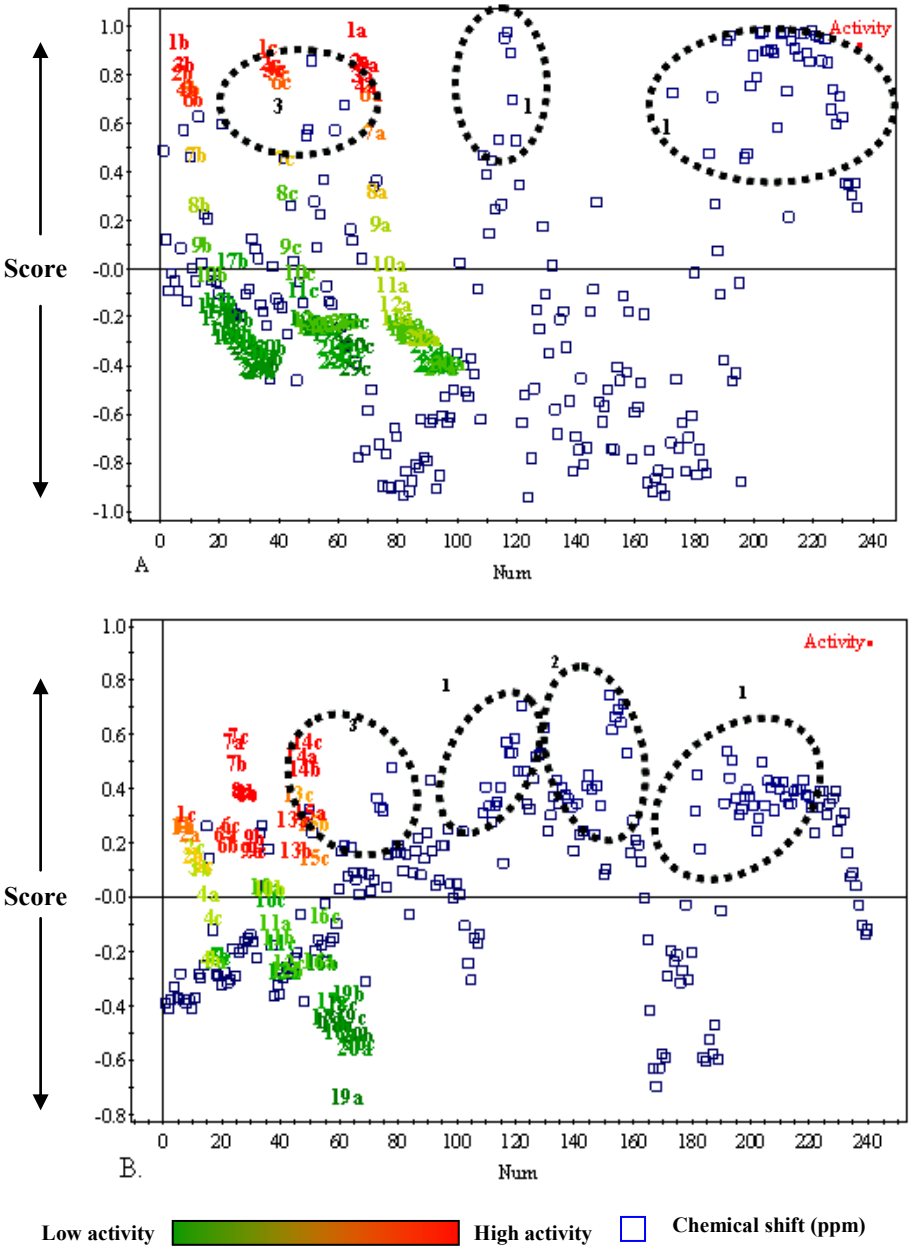


Figure 4. A. ES1 OPLS Loading bi-plot **B.** ES2 OPLS Loading bi-plot (1 = fatty acids and diterpenoids signals 2 = methoxy signals of flavonoids 3 = other flavonoids signals in aromatic area). Fractions are labeled by number with a-c representing replications and colour gradient from green to red representing low to the high activity. Chemical shift with a score value similar to certain fractions are abundant in the respective fraction(s).

Chemical validation

Previously our group has isolated 7 methoxy flavonoids which have binding activity to the adenosine A1 receptor with a pKi value in the μM (420).We checked the presence of signals belonging to these compounds in the loading plot and how they correlate to activity. For the last purpose we used two tools available in OPLS; Y-related coefficient (Ycoeff.) and variable of importance (VIP). A VIP value higher than 0.7-0.8 is considered to have a strong influence to the model but the limitation is that it gives positive values to all kind of correlations while in Y-related coefficient either positive or negative correlation is shown (15). The summary is given in table 2.

As can be seen from table 2, all flavonoids signals are found in the studied fractions and are positively correlated with the activity (positive Y-related coefficient). The methoxy signals have a higher VIP value than signals in the aromatic area. The last are less than 0.7.

We have chosen fraction 6c (82.83% binding activity) for 2D NMR measurements to elucidate the possible compounds responsible for the activity, although it is not the most active fraction but it has interesting signals other than methoxy flavonoids (δ 0.50 – 2.00, δ 4.20 – 5.80 ppm) which are also close to the activity (Fig. 5). Fraction 8c and 14c are more active than 6c, but in 8c (as well as 8a and b, spectra not shown) methoxy flavonoids signals are predominant while in 14c (as well as 14a and b, spectra not shown) there is a broadening water signal which hampers interpretation of the spectra.

The data from 2D NMR (J-res, COSY, and HMBC) confirmed the presence of these methoxy flavonoids in fraction 6c. Doublet of doublets signals at H-6' (δ 7.5 – 8.0) due to the proton-proton correlations between H-6' which is in *ortho*-position to H-5' and of *meta*-position to H-2', are observed in the COSY spectra of fraction 6c. Singlets at δ 3.7 – 4.0 assigned to methoxy groups were also confirmed by HMBC spectra which showed the coupling of the protons of these methoxy groups to the aromatic carbons (δ 140.0 – 162.0).

Next we tried to identify the availability of diterpenoids signals in fraction 6c and check their importance for the activity. The presence of this class of compounds in

O. stamineus was previously reported by several authors (315, 427). Typical signals of these compounds are present in area **1** of the ES2 loading bi-plot and shown to be positively correlated with the activity; those are δ 0.7 – 1.6 for tertiary methyls, δ 1.90 – 3.1 for aliphatic methylenes, δ 4.00 – 5.00 for oxygenated aliphatic methines and methylenes, δ 5.5 – 6.5 for vinyls, and also signals in aromatic area since these diterpenoids may contain olefinic protons and benzoyl groups (Fig. 4B). From the proton-proton and proton-carbon correlations deduced from COSY and HMBC spectra, the presence of a diterpenoid skeleton can be inferred. Tertiary methyls and aliphatic methylenes have VIP value between 0.6 – 1.5, while vinyls, oxygenated aliphatic methines and methylenes have less VIP value which are vary between 0.2 – 0.7. However, chemical validation by testing the reference compounds is required to confirm this prediction. There are some signals which are negatively correlated with the activity, such as δ 2.32 – 2.64, δ 2.80 – 3.48, and δ 5.72 – 5.92. From ES2 OPLS Xvar plot, these signals are abundant in polar fractions, thus can be attributed to polar metabolites such as sugars and amino acids.

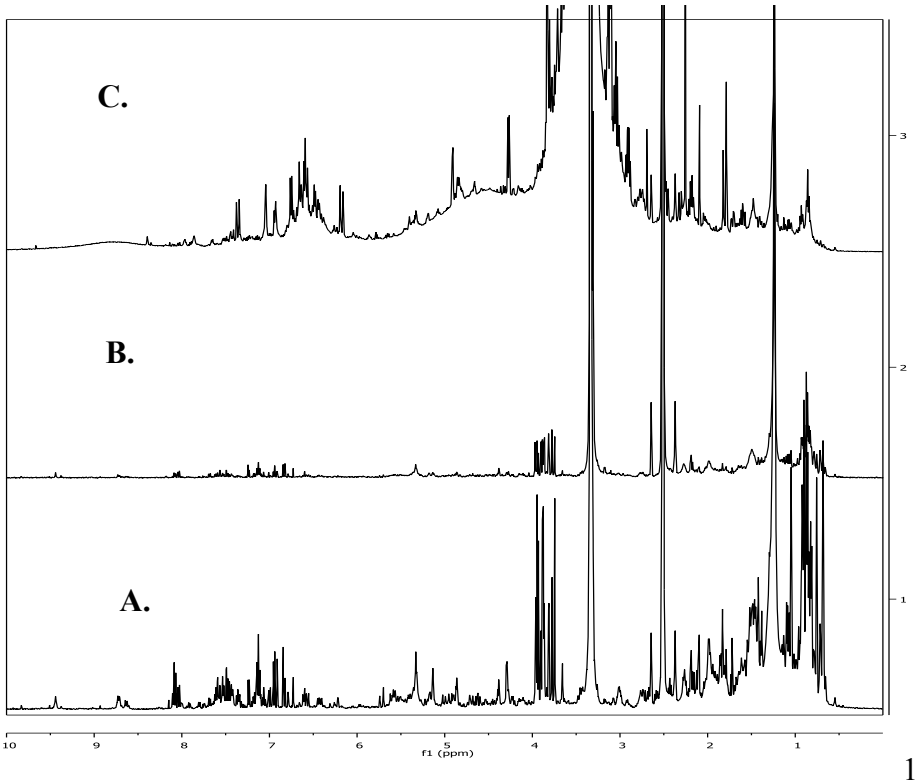


Figure 5. The ^{1}H NMR (DMSO-d_6) spectra of ES2 fraction 6c (A), 8c (B), and 14c (C)

Table 2. The adenosine A1 binding activity (*p*Ki), Y-related coefficient and VIP values of methoxy flavonoids signals isolated from *O. stamineus*.

Compounds	*p*Ki (μM)	^{1}H NMR signals (ppm, DMSO-d_6)	Ycoeff.	VIP
3'-hydroxy-4',5,6,7 tetramethoxyflavone (**1**)	5.4 ± 0.2	6.57 (H-3, *s*)	1.5 x 10 $^{-3}$	0.3
		3.99 (5-OMe, *s*)	5.0 x 10 $^{-3}$	1.2
		3.74 (6-OMe, *s*)	7.0 x 10 $^{-3}$	1.7
		3.93 (7-OMe, *s*)	6.0 x 10 $^{-3}$	1.5
		7.15 (H-8, *s*)	3.0 x 10 $^{-3}$	0.7
		7.41 (H-2', *d, J*= 2.0 Hz)	4.0 x 10 $^{-4}$	0.2
		3.84 (4'-OMe, *s*)	7.0 x 10 $^{-3}$	1.7
		7.06 (H-5', *d, J*=8.5 Hz)	1.7 x 10 $^{-3}$	0.4
		7.50 (H-6', *dd, J*=8.5 Hz, *J*=2.0 Hz)	1.3 x 10 $^{-4}$	0.1
3',5-dihydroxy-4',6,7-trimethoxyflavone (**2**)	5.5 ± 0.1	6.91 (H-3, *s*)	2.0 x 10 $^{-3}$	0.1
		3.71 (6-OMe, *s*)	5.0 x 10 $^{-3}$	1.2
		3.85 (7-OMe, *s*)	7.0 x 10 $^{-3}$	1.5
		6.81 (H-8, *s*)	7.0 x 10 $^{-4}$	0.2
		7.46 (H-2', *d, J*= 2.0 Hz)	1.2 x 10 $^{-3}$	0.3
		3.91 (4'-OMe, *s*)	6.0 x 10 $^{-3}$	1.5
		7.06 (H-5', *d, J*=8.5 Hz)	1.7 x 10 $^{-3}$	0.4
		7.56 (H-6', *dd, J*=8.5 Hz, *J*=2.0 Hz)	1.4 x 10 $^{-3}$	0.3
4',5,6,7-tetramethoxyflavone (**3**)	5.4 ± 0.1	6.70 (H-3, *s*)	7.0 x 10 $^{-4}$	0.2
		3.74 (5-OMe, *s*)	7.0 x 10 $^{-3}$	1.7
		3.78 (6-OMe, *s*)	7.0 x 10 $^{-3}$	1.7
		3.83 (7-OMe, *s*)	7.0 x 10 $^{-3}$	1.7
		7.20 (H-8, *s*)	2.4 x 10 $^{-5}$	0.01
		8.10 (H-2', *d, J*=9.0 Hz)	1.2 x 10 $^{-3}$	0.3
		7.08 (H-3', H5', *d, J*=9.0 Hz)	1.7 x 10 $^{-3}$	0.4
		3.93 (4'-OMe, *s*)	6.0 x 10 $^{-3}$	1.5
		7.01 (H-6', *d, J*=9.0 Hz)	1.4 x 10 $^{-3}$	0.3
3',5-dihydroxy- 4', 7-dimethoxyflavone (**4**)	4.3 ± 0.1	6.70 (H-3, *s*)	7.0 x 10 $^{-4}$	0.1
		6.50 (H-6, *d, J*=2.0 Hz)	7.0 x 10 $^{-4}$	0.1
		3.93 (7-OMe, *s*)	6.0 x 10 $^{-3}$	1.5
		6.88 (H-8, *d*, *J*=2.0 Hz)	6.0 x 10 $^{-4}$	0.1
		8.04 (H-2', *d, J*=2.0 Hz)	1.4 x 10 $^{-3}$	0.3
		3.83 (4'-OMe, *s*)	7.0 x 10 $^{-3}$	1.7
		7.10 (H-5', *d, J*=8.5 Hz)	3.0 x 10 $^{-3}$	0.7
5,6-dihydroxy- 4', 7-dimethoxyflavone (**5**)	5.1 ± 0.1	6.70 (H-3, *s*)	7.0 x 10 $^{-4}$	0.1
		6.52 (H-6, *d*, *J*=2 Hz)	7.0 x 10 $^{-4}$	0.1
		3.90 (7-OMe, *s*)	7.0 x 10 $^{-3}$	1.7

		7.94 (H-8, *d*, *J*=2 Hz)	6.0 x 10^{-4}	0.1
		8.04 (H-2', *d, J*=8.5 Hz)	1.4 x 10^{-3}	0.3
		7.10 (H-3', H-5', *d, J*=8.5 Hz)	3.0 x 10^{-3}	0.7
		3.84 (4'-OMe, *s*)	7.0 x 10^{-3}	1.7
		6.90 (H-5', *d, J*=8.5 Hz)	2.3 x 10^{-3}	0.5
3',3-dihydroxy- 4', 5, 6, 7-tetramethoxyflavone (**6**)	5.4 ± 0.1	3.74 (6-OMe, *s*)	7.0 x 10^{-3}	1.7
		3.93 (7-OMe, *s*)	6.0 x 10^{-3}	1.5
		6.57 (H-8, *s*)	1.5 x 10^{-3}	0.3
		7.59 (H-2',*d*, *J*=2.0Hz)	9.3 x 10^{-4}	0.2
		3.77 (3'-OMe, *s*)	7.0 x 10^{-3}	1.7
		6.90 (H-5', *d*, *J*=8.5)	2.3 x 10^{-3}	0.5
		7.50 (H-6', *dd, J*=2.0 Hz, *J*=8.5 Hz)	1.3 x 10^{-3}	0.3
		7.21 (H-3, *s*)	2.3 x 10^{-5}	0.01
3',4',5,6,7-pentamethoxyflavone (**7**)	5.5 ± 0.1	3.94 (5-OMe, *s*)	6.0 x 10^{-3}	1.5
		3.75 (6-OMe, *s*)	7.0 x 10^{-3}	1.7
		3.87 (7-OMe, *s*)	7.0 x 10^{-3}	1.7
		6.79 (H-8, *s*)	7.0 x 10^{-4}	0.1
		7.14 (H-2', *d, J*=2.0 Hz)	3.0 x 10^{-3}	0.4
		3.78 (3'-OMe, *s*)	7.0 x 10^{-3}	1.7
		3.83 (4'-OMe, *s*)	7.0 x 10^{-3}	1.7
		7.55 (H-5', *d*, *J*=8.5 Hz)	1.4 x 10^{-3}	0.3
		7.62 (H-6', *dd, J*=2.0 Hz, *J*=8.5 Hz)	1.3 x 10^{-4}	0.1

Conclusion

In the present report, applying comprehensive extraction coupled to NMR metabolomics and multivariate data analysis, it is shown that it is possible to identify active compounds from plant extracts without any chromatographic steps. The most chemical diversity was obtained when comprehensive extraction with a broad polarity was applied (ES2). The choice of NMR solvent must be taken into consideration since it affects the solubility of important compounds to be detected. The same solvent as the one which is used in the bioassay is recommended. Dimethyl sulfoxide (DMSO) is the most common solvent for various bioassay systems. Although in this report data from ES2 is easier to interprete, ES1 showed better results in the statistical validation. The reason could be because with DMSO-d_6 there is more noise in the baseline as compared to methanol-d_4. Overall, based on statistical and chemical validations, this method seems promising as alternative to conventional bioassay guided fractionation.