Comprehensive extraction integrated with NMR metabolomics: a new way of bioactivity screening methods for plants, adenosine A1 receptor binding compounds in Orthosiphon stamineus Benth

Yuliana, N.D.; Khatib, A.; Verpoorte, R.; Choi, Y.H.

analytical chemistry

TECHNICAL NOTE

pubs.acs.org/ac

Comprehensive Extraction Method Integrated with NMR Metabolomics: A New Bioactivity Screening Method for Plants, Adenosine A1 Receptor Binding Compounds in *Orthosiphon stamineus* Benth

Nancy Dewi Yuliana,[†,‡] Alfi Khatib,[§] Robert Verpoorte,[†] and Young Hae Choi*[,†]

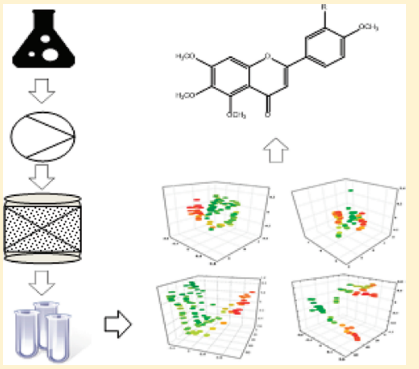
[†]Natural Products Laboratory, Institute of Biology, Leiden University, Einsteinweg 55, 2300 RA, Leiden, The Netherlands

[‡]Department of Food Science and Technology, Bogor Agricultural University, IPB Dramaga Campus, Bogor 16680, Indonesia

[§]Center of Excellence for Food Safety Research, Faculty of Food Science and Technology, University Putra Malaysia, 43400 Serdang, Selangor Darul Ehsan, Malaysia

Supporting Information

ABSTRACT: A large number of plant metabolites has provided as an incomparable chemical source for drug development. However, the wide range of the polarity of metabolites has been a big obstacle for full use of the chemical diversity. The initial step conventional extraction method by a single solvent does not make use of all the metabolites contained in plants. Also, it takes a long time to confirm the target activity of a single compound because of tedious separation steps. To solve the problem, a new extraction method coupled to NMR-based metabolomics is applied to identify bioactive natural products. A comprehensive extraction method consisting of a continuous flow of solvent mixtures through plant material was developed to provide extracts with a wider chemical variety than those yielded with a single solvent extraction. As the model experiment, ^{1}H NMR spectra of the extracts obtained from the comprehensive extraction of *Orthosiphon stamineus* were subjected to multivariate data analysis to find its adenosine A1 binding activity. On the basis of the results, two flavonoids from a large number of chemicals were clearly verified to show the adenosine A1 binding activity without any further purification steps. This method could provide a solution to the major drawbacks of natural products in drug development.



The introduction of high-throughput screening in the 1990s was designed to shorten the route to the discovery of drugs. To fully use its high potential, this technology requires a large number of compounds in order to screen thousands of samples per day.[1] Thus, the first target of pharmaceutical companies involved in the discovery of new drugs is to screen as many compounds as possible to increase the possibilities of finding one that can make it to the next stage. High-throughput synthesis and combinatorial chemistry were developed and applied to address this demand. These technologies, however, have not met the expectations built around them and there has not been a significant increase in new lead compounds delivered to the market. Sorafenib, a multikinase inhibitor indicated for advanced renal cancer, is the single synthetic drug derived from combinatorial chemistry approved by the Food and Drug Administration[2] for clinical use.[3]

Apparently, insignificant achievement in providing a chemical diversity is one of the reasons for the catastrophic failure of the aforementioned new technologies. On the other hand, it is well-known that the incomparable molecular diversity and biological functionality that characterize the metabolites present in plant extracts are the major features that distinguish them as a potential drug source as compared to combinatorial chemistry.[4] As reviewed recently,[3,5] during the past 2 decades, of the 877 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, indicating that NP is still a promising source of novel leads for therapeutic drugs. A natural products-based drug discovery project, however, also poses some challenges, mostly connected with their inclusion in a complex matrix composed of mostly uncharacterized compounds with a large dynamic range (e.g., the difference in the levels of major compounds as compared to that of minor compounds, differences in polarity, in boiling poits, or in melting points, etc.).

Metabolomics, which is defined as the comprehensive qualitative and quantitative analysis of all metabolites (the metabolome) present in an organism,[6,7] could potentially be instrumental in the approach needed to study the complex mixture of plant extracts in this context. To facilitate this, a specific extraction method that is able to extract the widest possible range of metabolites is needed. The next requirement is the selection of

Received: June 9, 2011
Accepted: July 29, 2011
Published: July 29, 2011

a suitable tool to detect the broad range of these extracted metabolites. Reproducibility is the most important criteria for developing a metabolomics technology platform, and in that context, NMR is one of the most suited methods even though its sensitivity is not as high as MS and chromatography based metabolomics platforms (1 μM to 1 mM in an NMR tube).[7,8] Thus, the choice of the extraction method is an essential step to obtain the real metabolic profile. The simple and fast sample preparation, short measurement time, plus the possibility of elucidating structures of known or unknown compounds in a complex mixture using advanced two-dimensional (2D) NMR methods are further advantages.[7] Finally, we need an advanced data analysis tool to extract information from the high dimensional data set resulting from these analyses. Such a method should additionally be able to link the chemical profile of the plant extract to its bioactivity data. Recently, some studies introducing this concept have been reported using, for example, metabolomics combined with projection-based multivariate data analysis (PLS-DA, PLS).[9−12]

In the search for biologically active compounds in plants, it is necessary to count on different accessions of plants with different biological activities or extracts or fractions of the plant material that have different activities and different (partly overlapping) chemical profiles.

In this paper an integrated approach is described. A novel comprehensive extraction method was used rather than a conventional single extraction coupled to a conventional liquid−liquid partition. The aim was to cover a wide range of metabolites with good resolution. In this method, a solvent mixture with a gradient of polarity was continuously delivered into a column packed with the powdered plant material. No other substance was added to the column except a small quantity of Kieselguhr that was mixed in with the powdered plant material to avoid clogging of the column. The fractions were collected according to specified time intervals and subjected to NMR measurement and bioactivity testing. Multivariate data analysis tools such as PLS and OPLS were used to find possible correlations between the metabolite profile and bioactivity. Statistical validations (CV-ANOVA, permutation test, and external validation) were performed to select the most reliable method. Some evaluation tools available in the corresponding multivariate data methods were used to select important signals. On the basis of that, compounds highly correlated with the activity were detected and identified by means of 2D NMR data. Chemical validation was also accomplished by the detection of signals corresponding to previously isolated active compounds in those fractions which showed a good activity profile. As a model activity test, the adenosine A1 binding activity was selected since the obtained data (% inhibition) can be easily integrated with NMR data, and the plants (*Orthosiphon stamineus*) used in this study already showed significant activity in this model system.[13]

EXPERIMENTAL SECTION

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), [^{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.

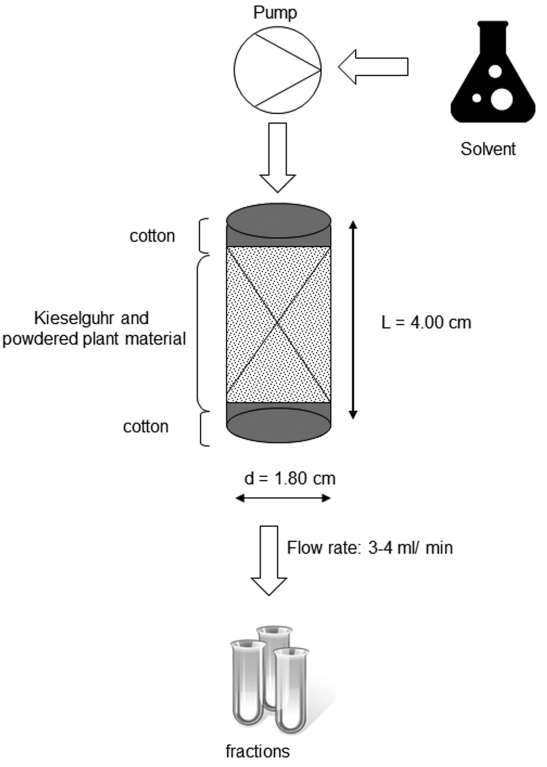


Figure 1. Schematic diagram of a comprehensive extraction kit employed in this study.

Plant Material. *Orthosiphon stamineus* was obtained from van der Pigge Drugstore. The plant materials were identified by one of the authors (N. D. Yuliana), and a voucher specimen (ORST-Fcog-NL-230506) was deposited in the Division of Pharmacognosy, Institute of Biology, Leiden University. The plant materials were ground into a powder in a blender. The particle size was under 0.71 mm.

Comprehensive Extraction. Extraction was performed in a stainless steel column (4.00 cm length, 1.80 cm diameter) packed with the mixture of 0.70 g of dried powdered *O. stamineus* and 0.10 g of Kieselguhr (Figure 1). The column was closed at both ends with fat free cotton and connected to a Waters 600E pump (Waters, Milford, MA). Organic solvents and filtered Millipore water (500 mL each) were ultrasonicated and degassed before use. The fractions were collected in 10 mL tubes every 2 min with an automatic fraction collector. Two extraction schemes were developed.

Scheme 1 (ES1). Lipids were removed from the powdered plant material as follows: 3 mL of *n*-hexane were mixed with the powdered plant material, ultrasonicated for 15 min, and filtered. This was repeated twice with 2 mL of *n*-hexane. The plant material was then dried under N_2, mixed with Kieselguhr, and loaded into the extraction column. The combination of solvents used for extraction scheme 1 was ethyl acetate−methanol (1:1) as solvent A and methanol−water (1:1) as solvent B. A total of 30 fractions were obtained at the end of the extraction. An aliquot of 1.5 mL of the extract was kept from each fraction to be used in the bioassay. The remaining volume (6.5 mL) was used for NMR measurement with methanol-d_4 as a solvent. Fractions were taken to dryness under nitrogen flow. Three replicates were done of each extraction and bioactivity of each extract was tested in duplicate.

Scheme 2 (ES2). The combination of solvents used for the extraction scheme 2 was *n*-hexane (A), acetone (B), and water (C).

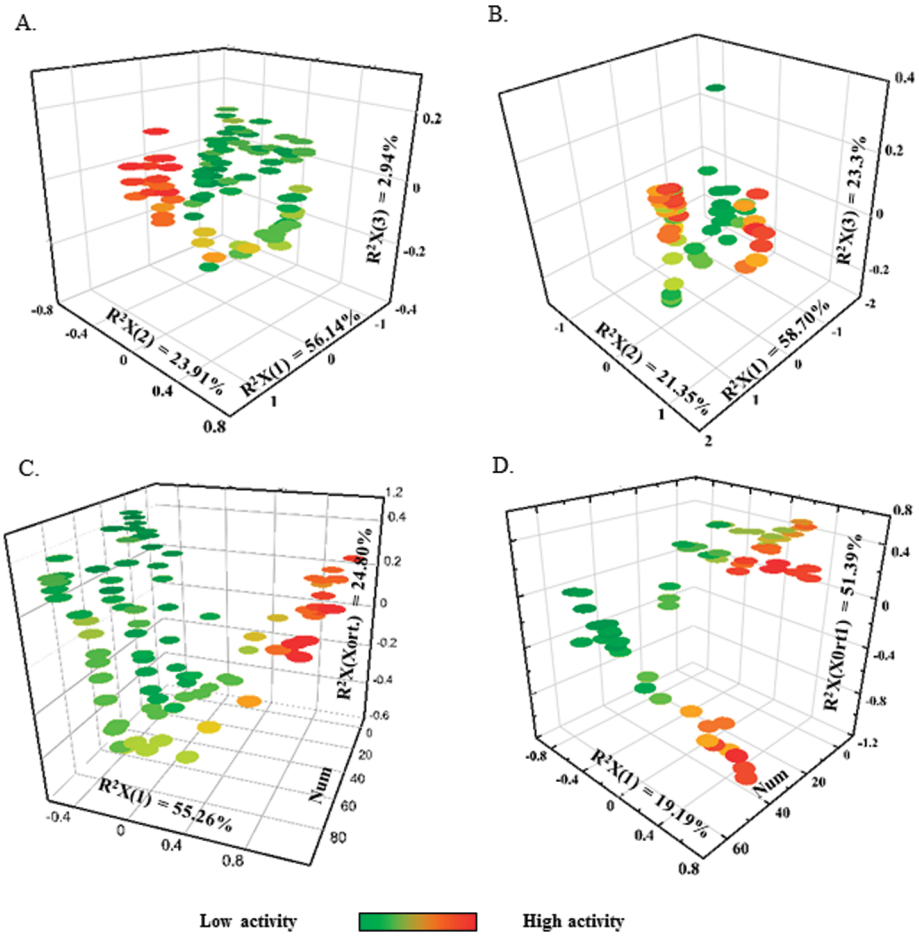


Figure 2. Score plot of PLS and OPLS of fractions obtained from ES1 and ES2. A color gradient from green to red represents an increase of adenosine A1 receptor binding activity of each fraction: (A) first 3 PLS components of ES1 fractions; (B) first 3 PLS components of ES2 fractions; (C) predictive and first orthogonal component of OPLS of ES1 fractions; and (D) predictive and first orthogonal component of OPLS of ES2 fractions.

A mixture of powdered plant material and Kieselguhr was directly loaded into the column without the defatting step. A total of 20 fractions were obtained at the end of the extraction. Every fraction was then dried under nitrogen flow and dissolved in DMSO-d_6 and subsequently subjected to ^{1}H NMR measurement and the adenosine A1 bioassay test. Three replicates were made with each extraction and each bioactivity test was done in duplicate.

NMR Measurement. NMR measurements were performed according to Kim et al.[14] requiring minor modification as can be seen in the Supporting Information.

Data Analysis. The ^{1}H NMR spectra were automatically reduced to ASCII files. Bucketing was performed by AMIX software (Bruker, Karlsruhe, Germany). 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 HDO 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 residual HDO, respectively. Partial least-squares projection to latent structure (PLS) and orthogonal-PLS (OPLS) were performed with the SIMCA-P software (v. 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.[15] 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.[16]

RESULTS AND DISCUSSION

It is quite a common practice in current natural products research, and particularly more so in the case of drug discovery research, to use a single solvent to obtain the initial plant extract. The drawback to this relatively simple procedure is that it is not possible to know if most of the important and/or relevant metabolites have been extracted, since given the broad spectrum of the polarity of compounds, many of them could not necessarily have affinity with the selected solvent. We therefore developed a new extraction method, which we called comprehensive extraction. It is based on the continuous extraction of plant material with a mixture of solvents of increasing polarity and the collection of fractions at predetermined intervals. In comprehensive extraction the solvent is delivered directly into the column filled with powdered plant material mixed with Kieselguhr to remove void volume. In ES1, we focused on relatively polar metabolites while in ES2 a wider metabolite profile was targeted. In any case of the

comprehensive extraction scheme, diverse compounds with a large range of polarity, e.g., sugars and lipids in a single run were extracted, which cannot be expected from a conventional single step extraction. The chemical diversity obtained from the comprehensive extraction made it possible to test a broad range of compounds for the bioactivity test without a further additional extraction step.

For the next step, the chemical profiles of ^{1}H NMR spectra were correlated with an adenosine A1 receptor binding test by multivariate data analysis. The % inhibition of adenosine A1 receptor of the fractions obtained ES1 and EX2 are listed in Supporting Information Table 1. For the correlation to find a bioactive compound in the mixture, all the fractions should be dissolved in common deuterated solvent to provide the same variables for multivariate data analysis, chemical shift in the ^{1}H NMR case. Diverse solvents such as D_2O, methanol-d_4, DMSO-d_6, and their mixtures were tested. Of the solvents employed in this study, methanol-d_4 or DMSO-d_6 showed promising results in terms of resolution and variety of signals. In Figure 2A, score plots of the PLS first three components for ES1 fractions shows the separation between active and nonactive fractions in PC1. For ES2 fractions, the nonpolar active fractions are separated from others in PC1, while the polar ones are separated in PC2 (Figure 2B). Apparently ES2 provides an ampler chemical diversity as can be observed from the bioactivity profile which shows two peaks of activity: one in the nonpolar area and another in the polar area. Meanwhile in ES1, with the use of a more polar solvent mixture, the activity was concentrated in the first seven fractions.

In OPLS, systematic variations in the *X*-block were divided into two; one predictive component which models the correlation between *X* and *Y* and an orthogonal component(s) which expresses *X*-variation unrelated to *Y*.[17,18] As expected, active fractions from both extraction scheme were clearly separated from nonactive ones in the predictive component (Figure 2C,D). Particularly, the two active fractions from ES2 were now clustered together in a predictive component, separated from nonactive ones.

The multivariate model is most commonly interpreted in two ways, i.e., by means of R^2 and Q^2 cumulative after cross validation.[17] The R^2 informs the amount of *Y* variables explained by the model after cross validation and gives an overview on the fitness of the model, while Q^2 gives information about the prediction quality of the model. All values close to 1 resemble a good model.[17] In this study, all models show good performance with an R^2 and Q^2 value between 0.75 and 0.92.

The PLS models were also validated using the response of the permutation test through 100 permutations. External validation was further conducted to assess the performance of the models. Each group of ES1 and ES2 data were separated into a calibration group and a test group. For ES1, PLS and OPLS were shown as good models with the root-mean-square error of prediction (RMSEP) and root-mean-square error of estimation (RMSEE) that do not differ much. The same applies for ES2 models although PLS showed a poorer model as compared to OPLS. Overall, ES1 models gave better external validation results as can be seen from the lower value of RMSEP and RMSEE. In addition, cross validation (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.[19] Both PLS and OPLS models from ES1 and ES2 were shown as significant models according to CV-ANOVA validation with $p < 0.05$.

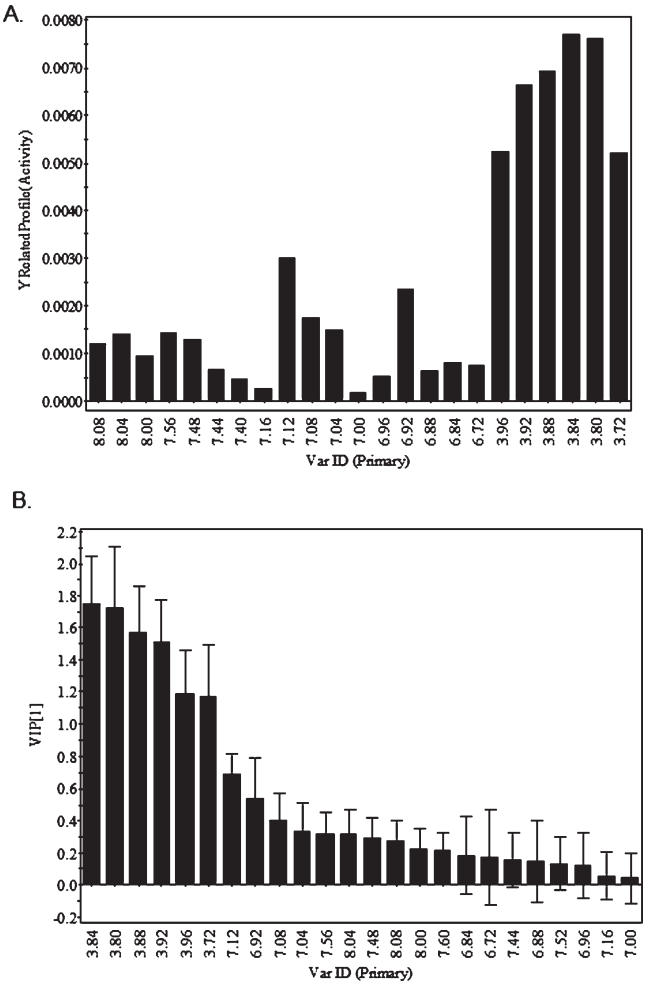


Figure 3. (A) Plot of OPLS *Y*-related coefficient values of important signals in ES2; (B) plot of OPLS VIP value of important signals in ES2.

We chose the OPLS model to select which signals were important for the activity. It was easier to interpret OPLS by partitioning the uncorrelated variations orthogonally from the predictive ones, based on the loading S-plot profile (Supporting Information Figures 1 and 2), ES2 seemed more interesting since more methoxy and aromatic signals were close to the activity while loading S-plot of ES1 was strongly affected by fatty acids signals. It has been reported that unsaturated fatty acids bind unspecifically to adenosine A1 receptors.[20] In the loading S-plot of ES2, there were interesting signals between δ 3.70 and 4.00 and δ 6.80 and 8.00 related to the activity. on the basis of the OPLS Xvar plot (not shown), these signals were found to be abundant in nonpolar active fractions from ES2. We chose fraction 6c (82.8% binding activity) for 2D NMR measurement to elucidate the possible compounds responsible for the activity. Using J-resolved spectra, several singlets between δ 3.70 and 4.00, which are typical signals of a methoxy substituent, were detected. Some singlets between δ 6.70 and 7.00, double doublets at δ 7.57 and δ 7.65, and several doublets between δ 7.07 and δ 8.10 were also observed, indicating the presence of methoxy-flavonoids.

Further confirmation using correlation spectroscopy (COSY) and heteronuclear multiple-bond correlation spectroscopy (HMBC) was performed. Many correlations were observed between phenolic protons with chemical shifts higher than δ 7.00.

Also, the long-range coupling between some protons at δ 3.70−4.00 and carbons at δ 130.0−δ 160.0 were observed in HMBC spectra and were confirmed as the correlations between methoxy protons and the phenolic carbons attached by the corresponding methoxy groups.[13]

The *Y*-related coefficient and VIP value of all the aforementioned signals were evaluated. A VIP value higher than 0.7−0.8 is considered to have a strong influence on the model but the limitation is that it gives positive values for all kinds of correlations, while in *Y*-related coefficients either a positive or negative correlation are shown.[17] As shown in Figure 3A, all signals exhibited a positive *Y*-related coefficient. The VIP values of methoxy signals were significantly higher than others (Figure 3B).

For chemical validation, we chose the two methoxy flavonoids, 4′,5,6,7-tetramethoxyflavone (tetramethylscutellarein) and 3′,4′,5,6,7-pentamethoxyflavone (sinensetin) that have been reported to have an adenosine A1 receptor binding activity with pKi values of 5.4 ± 0.1 and 5.5 ± 0.1, respectively.[13] The availability of signals present in these two compounds in fraction 6c was confirmed by comparison with the ^{1}H, ^{13}C, and 2D NMR data (in DMSO-d_6) in our in-house metabolomics library and the literature.[21]

For 4′,5,6,7-tetramethoxyflavone (tetramethylscutellarein), three methoxy signals of the compound (δ 3.76, δ 3.79, and δ 3.94) were found in fraction 6c. In the HMBC of fraction 6c, the proton of the methoxy group at position 5 (5-methoxy at δ 3.94, s) correlated with its direct carbon (C-7, δ 157.4). The *Y*-related coefficients and VIP value for this proton are 0.007 and 1.55, respectively. The proton of 7-methoxy (δ 3.79, s) correlated to C-5 (δ 153) with *Y*-related coefficients and a VIP value of 0.008 and 1.70, respectively. The proton of 6-methoxy (δ 3.76) correlated to C-6 (δ 139.8) with *Y*-related coefficients and VIP values of 0.005 and 1.20, respectively. Several protons with a meta position to 4′-methoxy are also found in fraction 6c. In the COSY spectra, proton at position 5′ (5′-H, δ 7.11, d) correlated to 6′-H (δ 8.03, d), similarly 2′-H (δ 8.03, d) correlated to 3′-H (δ 7.11, d). These two signals have positive *Y*-related coefficients, 0.004 and 0.003, respectively, and a VIP value of 1.10 and 0.99, respectively.

For 3′,4′,5,6,7-pentamethoxyflavone (sinensetin), one of the methoxy signals of this compound (6-methoxy) is present in fraction 6c (δ 3.92, s). In HMBC, this proton correlated to C-7 (δ 157.4). The *Y*-related coefficient and VIP value are 0.007 and 1.55, respectively. Similarly, the correlation between protons with a meta position to 3′-methoxy and 4′-methoxy is found in the COSY spectra, those are 5′-H (δ 7.11, d), which is correlated to 6′-H (δ 7.72, dd). The *Y*-related coefficient and VIP value are 0.004 and 0.40, respectively.

CONCLUSIONS

The application of this comprehensive extraction method coupled to NMR metabolomics and multivariate data analysis reveals the possibility of achieving the identification of active compounds from plant extracts in a single run extraction. To obtain the best chemical variation, the most effective choice was a combination of solvents that covered a broad range of polarities (ES2). Overall, on the basis of statistical and chemical validations, this method seems promising as a new paradigm for natural product-based drugs discovery. It solves the common problems related to their matrix complexity. It is quite easy to pick out compounds that are important for the tested bioactivity to be further identified while common compounds that may cause false positives can be detected and thus discarded at a very initial stage. It is also able to provide a rich chemical variety that is a prerequisite for a successful drug discovery program.

ASSOCIATED CONTENT

Supporting Information. Additional information as noted in text. This material is available free of charge via the Internet at http://pubs.acs.org.

AUTHOR INFORMATION

Corresponding Author
*E-mail: y.choi@chem.leidenuniv.nl.

ACKNOWLEDGMENT

We thank Dr. E. G. Wilson for comments and advice on the work. This work was supported by a grant of the Ministry of Land, Transport, and Maritime Affairs, Korea (Grant PM56091).

REFERENCES

(1) Mishra, K. P.; Ganju, L.; Sairam, M.; Banerjee, P. K.; Sawhney, R. C. *Biomed. Pharmacother.* **2008**, *62*, 94–98.
(2) Khan, A.; Safdar, M.; Ali Khan, M. M.; Khattak, K. N.; Anderson, R. A. *Diabetes Care* **2003**, *26*, 3215–3218.
(3) Newman, D. J.; Cragg, G. M. *J. Nat. Prod.* **2007**, *70*, 461–477.
(4) Nisbet, L. J.; Moore, M. *Curr. Opin. Biotechnol.* **1997**, *8*, 708–712.
(5) Newman, D. J.; Cragg, G. M.; Snader, K. M. *J. Nat. Prod.* **2003**, *66*, 1022–1037.
(6) Hall, R. D. *New Phytol.* **2006**, *169*, 453–468.
(7) Verpoorte, R.; Choi, Y.; Kim, H. *Phytochem. Rev.* **2007**, *6*, 3–14.
(8) Colquhoun, I. J. *J. Pestic. Sci.* **2007**, *32*, 200–212.
(9) Bailey, N. J. C.; Sampson, J.; Hylands, P. J.; Nicholson, J. K.; Holmes, E. *Planta Med.* **2002**, *68*, 734–738.
(10) Bailey, N. J. C.; Wang, Y.; Sampson, J.; Davis, W.; Whitcombe, I.; Hylands, P. J.; Croft, S. L.; Holmes, E. *J. Pharm. Biomed. Anal.* **2004**, *35*, 117–126.
(11) Roos, G.; Röseler, C.; Büter, K. B.; Simmen, U. *Planta Med.* **2004**, *70*, 771–777.
(12) Cardoso-Taketa, A. T.; Pereda-Miranda, R.; Choi, Y. H.; Verpoorte, R.; Villarreal, M. L. *Planta Med.* **2008**, *74*, 1295–1301.
(13) Yuliana, N.; Khatib, A.; Link-Struensee, A.; Ijzerman, A.; Rungkat-Zakaria, F.; Choi, Y.; Verpoorte, R. *Planta Med.* **2009**, *75*, 132–136.
(14) Kim, H.; Choi, Y.; Verpoorte, R. *Nat. Protoc.* **2010**, *5*, 536–549.
(15) Chang, L.; Brussee, J.; Ijzerman, A. *Chem. Biodiversity* **2004**, *1*, 1591–1626.
(16) Dalpiaz, A.; Townsend-Nicholson, A.; Beukers, M. W.; Schofield, P. R.; Ijzerman, A. P. *Biochem. Pharmacol.* **1998**, *56*, 1437–1445.
(17) Eriksson, L.; Johansson, E.; Kettaneh-Wold, N.; Wold, S. *Multi and Megavariate Data Analysis*; Umetrics AB: Umeå, Sweden, 2006.
(18) Trygg, J.; Wold, S. *J. Chemom.* **2002**, *16*, 119–128.
(19) Eriksson, L.; Trygg, J.; Wold, S. *J. Chemom.* **2008**, *22*, 594–600.
(20) Ingkaninan, K.; von Frijtag Drabbe Kunzel, J. K.; Ijzerman, A. P.; Verpoorte, R. *J. Nat. Prod.* **1999**, *62*, 912–914.
(21) Iinuma, M.; Matsuura, S.; Kusuda, K. *Chem. Pharm. Bull.* **1980**, *28*, 708–716.