LETTER

High-Throughput Active Compound Discovery using Correlations between Activity and Mass Profiles

Kyu Hwan Park, Kyo Joong Yoon, Kyung-Hoon Kwon, and Hyun Sik Kim*

Division of Mass Spectrometry Research, Korea Basic Science Institute, Ochang 363-883, Korea

Abstract: The active components in a plant extract can be represented as mass profiles. We introduce here a new, multi-compound discovery method known as Scaling of Correlations between Activity and Mass Profiles (SCAMP). In this method, a correlation coefficient is used to quantify similarities between the extract activity and mass profiles. The method was evaluated by first measuring the anti-oxidation activity of eleven fractions of an Astragali Radix extract using DPPH assays. Next, 15 T Fourier-transform ion cyclotron resonance (FT-ICR) MS was employed to generate mass profiles of the eleven fractions. A comparison of correlation coefficients indicated two compounds at *m/z* 285.076 and 286.076 that were strong antioxidants. Principal component analyses of these profiles yielded the same result. FT-ICR MS, which offers a mass resolving power of 500,000, was used to discern isotopic fine structures and indicated that the molecular formula corresponding to the peak at *m/z* 285.076 was $C_{16}H_{13}O_5$. SCAMP in combination with high-resolution MS can be applied to any type of mixture to study pharmacological activity and is a powerful tool for active compound discovery in plant extract studies.

Keywords: Active compound discovery, Activity-mass profile correlation coefficient, High resolution mass spectrometry, Mass profile, Plant extract, Astragali Radix

Introduction

Plant extracts contain numerous ingredients that may include valuable, biologically active compounds for use in pharmaceutical development. One of the main goals of plant extract studies is the discovery of novel active compounds. A plant extract can be divided into multiple fractions to reduce its complexity. Generally, a fraction showing strong activity is selected and further fractioned until only one or two compounds remain. Since several purification steps are typically required prior to compound identification, this method is time consuming and not suitable for multiplex applications.[1,2] In addition, since identification is performed at the final stage of extract processing, it is not uncommon for the isolated compound to be one that is already known. Therefore, new methods that reduce analysis times and allow high-throughput, multi-compound discovery are strongly desired.

Fourier transform ion cyclotron resonance mass spectrometry (FT-ICR MS) has a mass resolution that is high enough to overcome signal overlap in analyses of complex mixtures. It is therefore a powerful tool for plant extract analyses since it does not require costly separation steps prior to the accurate assignment of molecular formulae.[3–5] The 15 T FT-ICR MS used in the current study has a resolving power high enough to observe isotopic fine structure (IFS), which is a unique pattern of mass spectral peaks produced by molecular ions as a result of isotopic distributions among their constituent atoms, including minor isotopic atoms such as ^2H, ^{15}N, ^{17}O, ^{33}S, ^{34}S, and etc. Fast and precise molecular formula determinations can be achieved using the IFS fingerprint and ultra-high resolution mass measurements for all of the compounds in a particular mixture. The molecular weights and relative abundance of compounds in a mixture can be measured by electrospray ionization (ESI) FT-ICR MS to generate a mass profile of the mixture. Since the activity of a mixture is related to the amount of active compounds present, the mass profiles of certain fractions may correlate with their respective activity profiles. Several studies have examined the correlation between activity and composition profiles as a means of enhancing discovery efficiency.[6–9] However, these studies were limited to specific activities and certain known compounds.

Here we introduce a new method named Scaling of Correlation between Activity and Mass Profiles (SCAMP), which evaluates correlations between activity and mass profiles in order to identify multiple, pharmacologically active compounds. The SCAMP method was used to assess the anti-oxidative activity of an Astragali Radix (AR) extract made from the dried roots of *Astragalus membranaceus* Bunge. AR is a well-known, traditional Asian herbal medicine that exhibits anti-inflammatory, anti-oxidation, and anti-tumor activities.[10,11]

Materials and Methods

Extraction and fractionation

A five-year-old sample of AR was obtained from the Oriental

*Reprint requests to Dr. Hyun Sik Kim
 E-mail: fticr@kbsi.re.kr

Kyu Hwan Park, Kyo Joong Yoon, Kyung-Hoon Kwon, and Hyun Sik Kim

Medicine Center in Chungbuk Techno Park, a Korean government institute for the regulation of medicinal plants. All chemicals used in this work were of analytical grade and purchased from Sigma Chemical Co. (Yongin, South Korea) unless otherwise noted. An ethanol extract of AR was prepared by dissolving 20 g of powdered AR in 500 mL of ethanol and heating under reflux for 4 h at 40 ^{o}C and concentrating to a final volume of 20 mL. The particle size of AR powder was controlled below 20 mesh. The entire AR extract was injected onto a SunFire HPLC column (C18, 5 μm, 250 × 4.6 mm). The flow rate was 2 mL/min with a binary mobile phase composed of solvents A (100% water) and B (100% ethanol). The applied solvent gradient was 0% B for 0–10 min and 100% B for 10–110 min; the eluate was collected every 10 min, yielding eleven fractions.

In vitro activity assays

Anti-oxidation activity profiles were generated by measuring the radical scavenging activity of each fraction with 1,1-diphenyl-2-picrylhydrazyl (DPPH).[12] Each fraction was lyophilized and reconstituted in 50% ethanol to make a 1 mL 0.1 mg/mL stock solution. The eleven stock solutions were diluted twice with ethanol and mixed vigorously with 0.1 mL of 60 μM aqueous DPPH solution. After 30 min, the absorbance of the remaining DPPH was measured at 517 nm. Chlorogenic acid was used as a standard agent.

Mass profiling by FT-ICR MS

Mass profiling was performed on a 15 T FT-ICR mass spectrometer (Apex-Qe; Bruker Daltonics, Bremen, Germany) in positive ESI mode. Stock solutions of the eleven fractions were diluted 100 fold with 50% methanol and introduced into the FT-ICR mass spectrometer using direct infusion with a flow of 2 mL/min by a syringe pump (Cole Palmer, Chicago, IL, USA). The MS parameters were: capillary voltage 4800 V, mass range m/z 185–2000, accumulation time 0.7 s, transient length 1.68 s, scan number 200, and a sine bell apodization window function. External calibration was performed with a 10 μg/mL arginine solution. All data were processed using DataAnalysis (ver. 3.4), a FT-ICR MS data processing program (Bruker Daltonics). The masses and abundances of the isotopes used for theoretical mass calculations were taken from the National Institute of Standards and Technology.[13] Peaks with the Signal-to-Noise ratio (S/N) >10 on the MS spectra were selected for mass profiling.

Calculation of correlation coefficients

A mass profile can be divided into multiple m/z bins in accordance with the resolving power of the equipment. For example, the m/z range from 200 to 2000 can be divided into 1801 bins with unit resolution. A mass profile of the jth m/z bin can be represented by the vector V, $V_j = \{m_{1j}, m_{2j}, m_{3j}, ..., m_{Nj}\}$, where m_{kj} = the MS intensity of the jth m/z bin of fraction k and N is the total number of fractions. If there is more than one peak within an m/z bin, their intensities are summed to

assign the value of m_{kj}. In the same way, the activity profile vector is defined as A = $\{a_1, a_2, a_3, ..., a_N\}$, where a_k is the activity of fraction k. The activity and mass profile vectors were normalized such that $|A^o| = 1$ and $|V^o_j| = 1$ prior to the calculation of correlation coefficients. The normalized vector components were labeled as m^o and a^o, respectively. The correlation coefficient, C_j, is the sum of values obtained as the vector product between the normalized activity and the normalized MS intensity. Thus, we can write, as shown in eq 1,

$$C_j = \sum_{k=1}^{N} (a_k^o \times m_{kj}^o) \tag{1}$$

where j is the jth m/z bin, k is the fraction number, a^o_k is the normalized activity of fraction k, and m^o_{kj} is the normalized MS intensity of the jth m/z bin in fraction k. C_j is greater if the pattern of the jth m/z profile is similar to the activity profile. Using this set of normalized activities and MS intensities, a principal component analysis (PCA) (R package)[14] was performed.

Results and Discussion

MS and activity profiles

The ESI FT-ICR MS spectra of the eleven fractions showed, at worst, a 0.5 ppm mass accuracy in the range of m/z 200–1700 and a mass resolution of 500,000 at m/z 400. As shown in Figure 1, more than 5,000 peaks were detected in each spectrum without considerable signal overlap; these peaks were used to generate a compositional profile of each fraction. The DPPH radical scavenging activity was highest in fraction 6, followed by fraction 7 (Table 1). The anti-oxidation activity of fraction 6 was higher than that of chlorogenic acid, suggesting that fraction
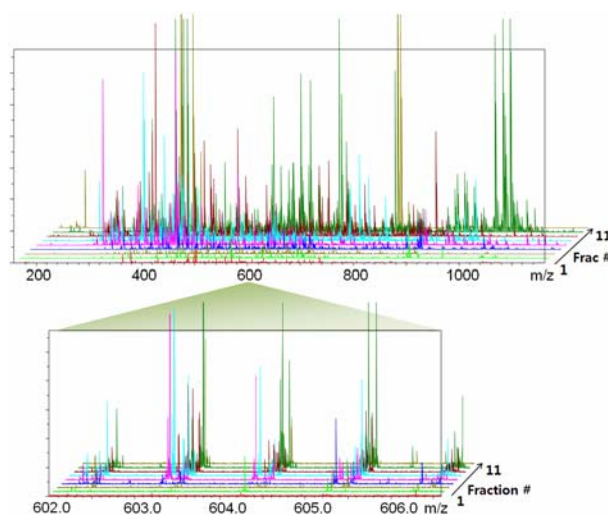

Figure 1. Stacked HR mass profiles of the eleven fractions obtained by 15 T FT-ICR MS. The mass resolution was 500,000 at m/z 400. The magnified spectra at the bottom show that the resolution was high enough to resolve all peaks without overlapping.

Table 1. Anti-oxidation activities of the fractions as measured by DPPH assay. The removal of radicals was monitored by measuring the UV absorbance at 517 nm. Chlorogenic acid was used as a standard sample. DPPH (−) and (+) refer to positive and negative controls of DPPH, respectively

Sample	Activity	Sample	Activity
Chlorogenic acid	122.5	Fraction #7	8.1
Fraction #1	5.1	Fraction #8	2.3
Fraction #2	2.0	Fraction #9	3.3
Fraction #3	4.3	Fraction #10	1.0
Fraction #4	3.1	Fraction #11	0.0
Fraction #5	5.1	DPPH (−)	131.0
Fraction #6	100.0	DPPH (+)	0.0

6 contained active antioxidants. Chlorogenic acid was used as a reference agent at 25 μg/mL in aqueous solution.[15,16]

Correlation coefficients

ESI MS spectra require strict alignment to adjust for subtle m/z shifts that may occur during the experiments. This is especially important for high-resolution spectra in which the large number of peaks and tiny m/z shifts can make identical peaks appear very different. In this work, the peak picking threshold was S/N>10 with a 1 mDa mass tolerance applied to the averaged m/z value of the eleven ESI MS spectra. If there were any active compounds with strong activity but the amount of the active compounds were too small to detect by MS, SCAMP may not be able to get proper profiles and the result may be deviated.

Correlation coefficients for each m/z were calculated according to Eq. (1) following vector normalization of the activities and alignment of the MS spectra. The relative activities of compounds in a plant mixture can be evaluated and sorted by C_j. The coefficient at m/z 286.076 was the greatest (76.47), indicating that a compound with a molecular weight of 286.076 Da was the most potent antioxidant (Table 2). It is reasonable that the coefficient at m/z 285.076 was also high since m/z 286.076 is an isotopic peak of m/z 285.076. The similarity in C_j values indicates that the compound with a molecular weight of 285 Da was also a potent antioxidant.

Principal component analysis

The activity profile vectors were included in the data set of MS intensity profile vectors. PCA of the merged data set showed that the mass profile of m/z 285.076 was very close to, and exhibited a pattern similar to, the activity profile vector indicated by a red dot in Figure 2. This suggests that the 285.076 Da compound is a strong antioxidant. m/z 286.076 is also located near the point of the activity profile vector. This result agrees with those of our correlation coefficient comparison above. In Figure 3a, the mass profile of m/z 285.076, indicated by a shaded square in the stacked raw MS spectra, shows a pattern nearly identical to that of the activity profile shown to the left of the mass profiles.
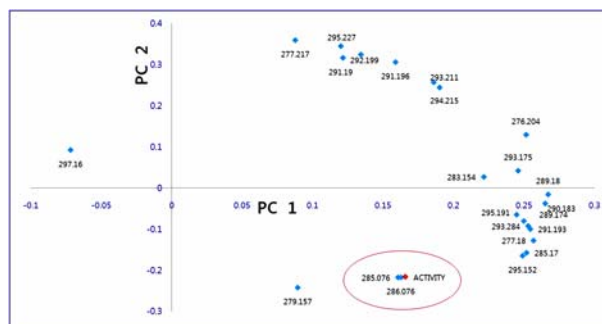

Figure 2. PCA plot of the data set composed of vector normalized activities and MS intensities. In the red circle are the expected active compounds whose m/z points were located near the activity point (red dot).

Table 2. A partial list of correlation coefficients produced using eq. (1) after vector normalization of the activities and MS intensities

m^o_{kj}	276.204	277.180	277.217	279.157	279.160	283.154	285.076	285.170	286.076	289.175	289.180	290.183	291.190	291.193	291.196	291.199	292.175	293.211	293.284	294.215	295.152	295.191	295.227	A^o_{kj}
Frc#1	0.00	0.01	0.02	0.00	0.82	0.01	0.01	0.00	0.02	0.01	0.02	0.02	0.00	0.01	0.02	0.00	0.00	0.00	0.01	0.01	0.02	0.02	0.00	0.04
Frc#2	0.01	0.02	0.00	0.00	0.00	0.01	0.01	0.01	0.00	0.02	0.01	0.02	0.02	0.00	0.01	0.02	0.01	0.02	0.02	0.00	0.01	0.02	0.00	0.02
Frc#3	0.01	0.02	0.02	0.00	0.01	0.01	0.02	0.02	0.00	0.01	0.02	0.01	0.00	0.01	0.02	0.02	0.00	0.01	0.02	0.00	0.00	0.02	0.01	0.03
Frc#4	0.00	0.00	0.00	0.01	0.01	0.01	0.00	0.02	0.01	0.02	0.02	0.00	0.00	0.00	0.02	0.00	0.01	0.01	0.02	0.02	0.00	0.01	0.02	0.02
Frc#5	0.01	0.02	0.02	0.02	0.01	0.00	0.00	0.00	0.01	0.00	0.00	0.02	0.02	0.01	0.01	0.01	0.00	0.00	0.01	0.01	0.00	0.01	0.00	0.04
Frc#6	0.32	0.66	0.03	0.81	0.00	0.61	1.00	0.78	1.00	0.48	0.51	0.53	0.17	0.55	0.23	0.18	0.56	0.15	0.53	0.18	0.80	0.42	0.03	0.76
Frc#7	0.82	0.75	0.23	0.01	0.00	0.52	0.02	0.63	0.03	0.88	0.83	0.82	0.29	0.83	0.38	0.30	0.59	0.59	0.85	0.58	0.60	0.91	0.40	0.06
Frc#8	0.41	0.00	0.64	0.00	0.02	0.00	0.01	0.01	0.01	0.00	0.24	0.23	0.50	0.00	0.63	0.67	0.59	0.75	0.00	0.76	0.00	0.00	0.53	0.02
Frc#9	0.25	0.00	0.73	0.00	0.50	0.60	0.00	0.00	0.00	0.00	0.06	0.00	0.80	0.00	0.64	0.66	0.00	0.26	0.00	0.24	0.00	0.00	0.75	0.02
Frc#10	0.00	0.02	0.00	0.39	0.17	0.00	0.00	0.00	0.00	0.01	0.01	0.00	0.00	0.00	0.00	0.00	0.00	0.01	0.01	0.01	0.00	0.02	0.01	0.01
Frc#11	0.00	0.01	0.01	0.43	0.20	0.00	0.00	0.82	0.01	0.01	0.00	0.02	0.01	0.02	0.02	0.00	0.01	0.02	0.00	0.00	0.00	0.01	0.01	0.00
C_j	30.78	55.19	6.36	62.18	4.60	51.39	76.43	63.08	76.47	41.85	44.42	45.80	17.24	47.44	22.43	18.52	47.30	16.71	45.33	18.97	64.98	37.33	7.37	

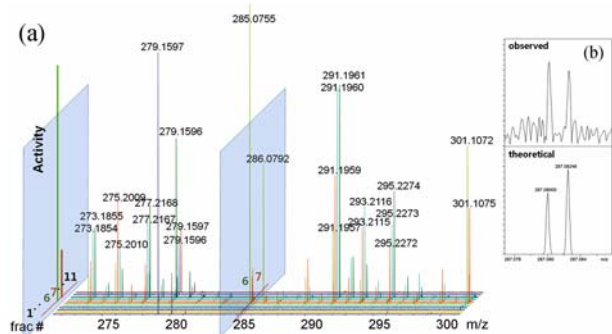Kyu Hwan Park, Kyo Joong Yoon, Kyung-Hoon Kwon, and Hyun Sik Kim

Figure 3. (a) The profile of the *m/z* 285 bin is similar to the activity profile, indicating that the component with *m/z* 285 was highly active. (b) The [M + 2] IFS of *m/z* 285 (above) is similar to the theoretical IFS of $C_{16}H_{13}O_5$ (below). The comparison of experimental and theoretical IFS is critical to determine the molecular formula.

Determination of the molecular formulae

The mass resolution afforded by FT-ICR MS was sufficient to reveal the IFS of small molecules such as phytochemicals, lipids, and metabolites. Using the IFS fingerprint, molecular formulae can be determined conclusively without any other information. The molecular formula corresponding to the peak at *m/z* 285.076 was $[C_{16}H_{12}O_5 + H]^+$ (Figure 3b). There are two candidate structures within the 5-ppm mass error of *m/z* 285.076: $C_{16}H_{13}O_5$ (285.0758 Da) and $C_{17}H_9N_4O$ (285.0771 Da). As shown in Figure 3b, the observed M+2 IFS of *m/z* 285.076 is the same as the theoretical M+2 IFS of $C_{16}H_{13}O_5$. In contrast, the M+2 IFS of $C_{17}H_9N_4O$ has a multiplet structure (data not shown). In addition, there are several phytochemicals with the molecular formula $C_{16}H_{13}O_5$, including wogonin, brazilein, and biochanin A.[17] Proper identification of the $C_{16}H_{13}O_5$ molecule is possible with further analysis using MS/MS fragmentation and nuclear magnetic resonance experiments.

Conclusions

SCAMP coupled with high-resolution MS was successfully applied to eleven fractions of an AR extract. Two compounds at *m/z* 285.076 and 286.076 exhibited strong antioxidative behavior, even though one was an isotopic peak of the other. This result was confirmed by PCA. The molecular formula of *m/z* 285.076 was $C_{16}H_{13}O_5$, as determined by close inspection of the IFS fingerprint. Mass profiling using 15 T FT-ICR MS yielded highly refined mass profiles, revealed hidden compounds, and allowed determination of the molecular formulae for almost all compounds in the extract.

In principle, the SCAMP method described herein can be used to analyze correlations between various combinations of mass and activity profiles and can evaluate mass correlations for multiple activity profiles simultaneously. This method reduces the necessity for costly and time-consuming separation and concentration procedures compared to conventional methods.

References

1. Harvey, A. *Drug Discovery Today* **2000**, 5, 294.
2. Strege, M. A. *J. Chromatogr. B Biomed. Sci. Appl.* **1999**, 725, 67.
3. Brown, S. C.; Kruppa, G.; Dasseux, J. L. *Mass Spectrom. Rev.* **2005**, 24, 223.
4. Miura, D.; Tsuji, Y.; Takahashi, K.; Wariishi, H.; Saito, K. *Anal. Chem.* **2010**, 82, 5887.
5. Cooper, H. J.; Marshall, A. G. J. *Agric. Food Chem.* **2001**, 49, 5710.
6. Senorans, F. J.; Ibanez, E.; Cavero, S.; Tabera, J.; Reglero, G. J. *Chromatogr. A* **2000**, 870, 491.
7. Wang, Y.; Jin, Y.; Zhou, C.; Qu, H.; Cheng, Y. *Med. Biol. Eng. Comput.* **2008**, 46, 605.
8. Aggio, R.; Ruggiero, K.; Villas-Boas, S. G. *Bioinfomatics* **2010**, in press.
9. Anjum, T.; Bajwa, R. *Nat. Prod. Res.* **2010**, 24, 1783.
10. Toda, S.; Shirataki, Y. J. *Ethnopharmacol.* **1999**, 68, 331.
11. Xuejiang, W.; Ichikawa, H.; Konishi, T. *Biol. Pharm. Bull.* **2001**, 24, 558.
12. Kitts, D. D.; Wijewickreme, A. N.; Hu, C. *Mol. Cell. Biochem.* **2000**, 203, 1.
13. http: //www.nist.gov/physlab/data/comp.cfm.
14. The R foundation for Statistical Computing (http://www.r-project.org/)
15. Kono, Y.; Shibata, H.; Kodama, Y.; Sawa, Y. *Biochem. J.* **1995**, 312, 947
16. Ozgen, M.; Reese, R. N.; Tulio Jr., A. Z.; Scheerens, J. C.; Miller, A. R. J. *Agric. Food Chem.* **2006**, 54, 1151.
17. Huang, W. H.; Lee, A. R.; Yang, Biosci. C. H. *Biotechnol. Biochem.* **2006**, 70, 2371.