![SHIMADZU Excellence in Science]

Technical Report

Development of a Novel Automated Identification and Quantification System with a Database for GC-MS

Kiwao Kadokami

Abstract:

We have developed a new, fully automated identification and quantification database system (AIQS-DB) to permit quantification of large numbers of chemical substances using a gas chromatograph-mass spectrometer (GC-MS). The AIQS-DB, containing mass spectra, retention times, and calibration curves, allows identification and quantification of more than 1000 substances, without the use of chemical standards. With a projected retention time error of ±3 seconds during sample measurement, reliable identification was possible in conjunction with the mass spectrum. Reproducibility of quantitative values was less than 20%, which is comparable to conventional sample measurement using the calibration curve method. Furthermore, more than 90 percent of the registered substances were below the detection limit of 10 pg. Because the AIQS-DB permits measurement of a large number of chemical substances without using chemical standards, it can be applied to a wide range of applications.

Keywords: GC-MS, AIQS-DB, Semi-Volatile Organic Compounds, Retention Index, System Performance Evaluation Standards

1. Introduction

The number of chemical substances registered under the Chemical Abstracts Service exceeded 64 million compounds in January 2012, and more than 100,000 of these are currently available on the market. Furthermore, production of these compounds increased 8 fold during three decades from 1970 to 1998, with the quantity and variety of compound types increasing rapidly. Despite the essential role that chemicals play in modern life, they are also associated with adverse effects. There have been many reports regarding the impact of residual organic pollutants, pesticides, and illegal drugs on human health and wildlife, and there are real concerns how about the possible adverse effects of new compounds. Irrespective of the presence or absence of regulations to mitigate these adverse effects and ensure their safe and secure use, a large number of chemical compounds must be analyzed. Currently, the most commonly used instrument for the analysis of these chemicals is the gas chromatograph-mass spectrometer (GC-MS), and analysis of specifically targeted chemical substances requires the use of chemical standards. This is primarily because retention times and response values vary with the state of the GC-MS at the time of analysis. It is therefore necessary to measure standard substances prior to actual analysis of the sample to verify the relationship between the retention time, injection volume, and response, in other words, to generate a calibration curve. However if the state of the GC-MS always remains relatively constant, it becomes possible to identify and quantitate target compounds without having to analyze chemical standards for each sample measurement. We have developed a method for evaluating the GC-MS performance and for predicting the retention times of target compounds. The result is that more than 1000 types of chemical substances can now be analyzed simultaneously using a fully automated identification and quantification database system (AIQS-DB), without the use of chemical standards. In this report, we describe the structure of the AIQS-DB developed for semi-volatile chemicals, the results of our evaluation of the identification and quantification performance, as well as a method used for maintaining performance.

2. Fully Automated Identification and Quantification Database System (AIQS-DB)

Typically, when standard GC-MS analysis is carried out, a calibration curve is generated prior to actual sample analysis to determine chemical compositions. However, generating a calibration curve involves extra cost, time, and effort, not to mention the problem of substance and solvent disposal. Furthermore, the number of substances that can be measured simultaneously using the calibration curve method is usually no more than about 100 substances.

Analysis using the AIQS-DB, a 3-component database comprising retention times, mass information, and calibration curves, essentially eliminates the requirement to measure standard substances. Since the process of measuring chemical standards is replaced with use of the data that is pre-registered in the database, All substances measurable under the same GC-MS conditions as those used for initial registration in the database can theoretically be measured using the system.

Also, since conventional analysis involves the use of high-sensitivity selected ion monitoring (SIM acgusition), searching for unknown

substances is not possible. With the AIQS-DB, however, searches can be conducted in sample measurement not only for target substances, but for unknown substances as well, thanks to the recent advances in high-sensitivity rapid scanning (Total Ion Monitoring, TIM) by GC-MS in recent years.

3. Semi-Volatile Chemical Substances Registered in AIQS-DB

Table 1 shows the numbers of substances registered in the AIQS-DB for semi-volatile chemical substances. This database has been developed for the purpose of measuring pesticide residues in food and environmental pollutants, and accordingly, includes the majority of substances regulated by various countries, substances that have been detected in the environment, as well as pesticides that can be measured by GC. Additionally, internal standards comprising 8 kinds of deuterium-labeled polycyclic aromatic hydrocarbons are included in the database.

Table 1 Semi-Volatile Chemical Substances Registered in AIQS-DB[Note]

Category 1 Substances	Number of Substances	Category 2 Substances	Number of Substances
CH-containing substances	194	Polycyclic aromatic	79
		PCBs	62
		Other	53
Oxygen-containing compounds	150	Phenols	50
		Other	100
Nitrogen-containing compounds	113	Aromatic amines	43
		Nitrogen compounds	42
		Other	28
Sulfur-containing compounds	12		12
Phosphorus-containing compounds	8		8
PPCPs	14		14
Pesticides	451	Insecticides	184
		Herbicides	118
		Fungicides	116
		Other	33
Total	942		
Internal Standards	8		

(Note) The GC-MS Database Software for Simultaneous Analysis (2nd Edition), from Shimadzu Corporation

4. Instrument and Measurement Conditions

The GC-MS measurement conditions used with this AIQS-DB are shown in Table 2. Measurement was conducted using the Shimadzu GCMS-QP2010, and the capillary column selected was the low polarity J&W DB-5ms, considered suitable for measurement of semi-volatile chemical substances, with a suitable length, inner-diameter, and film thickness. Taking hexane as the sample solvent, the initial column temperature was set to 40°C for the ideal

solvent effect, with the temperature rising at 8°C per minute up to 310°C. The majority of semi-volatile chemicals can be measured by increasing the temperature up to 310°C. The injector temperature was set to the most generally useful temperature of 250°C.

Table 2 GC-MS Measurement Conditions for Semi-Volatile Chemicals Using AIQS-DB

GC-MS	: Quadrupole GC-MS (e.g. Shimadzu GCMS-QP2010)
Column	: J&W DB-5ms
	(5% phenyl-95% methylsilicone) fused silica capillary column, 30 m × 0.25 mm i.d., 0.25 μm film
[GC]	
Injection volume	: 1 μL
Injection temp.	: 250°C
Column oven temp.	: 40°C (2 min) - 8°C /min - 310°C (5 min)
Control mode	: Constant velocity (40 cm/s)
Injection mode	: Splitless
Sampling time	: 1 min
Carrier gas	: He
[MS]	
Interface temp.	: 300°C
Ion source temp.	: 200°C
Ionization method	: EI
Tuning method	: Target tuning for US EPA method 625
Mass range	: m/z 45 to 600
Event time	: 0.3 sec

5. Construction of AIQS-DB and Method of Registering Substances

Here we describe the procedure for registering new substances to the AIQS-DB. After setting the GC-MS conditions as shown in Table 2, MS tuning is conducted using a method adopted by the U.S. Environmental Protection Agency as a method of analyzing semi-volatile chemicals (or EPA Method 625) in order to satisfy the fragmentation pattern of decafluorotriphenylphosphine (DFTPP). Next, after checking the retention times of n-alkanes and verifying the performance of the GC-MS following measurement of the instrument performance evaluation substance described later, a standard solution of substances to be registered is measured for generating the calibration curve. The information registered in the database consists of retention times, mass spectra, calibration curves obtained using the internal standard method, and the retention times of n-alkanes.

6. Sample Measurement and Method of Identification and Quantification

For sample measurement, the same procedure with construction of AIQS-DB described in section 5 above is conducted up to the performance check using the performance evaluation substance. Then, the internal standard mixed solution is added to the measurement sample solution at the same volume used when the substance was registered, and measurement is conducted.

For identification of a target compound, the peak corresponding to the quantification ion of that compound must be present at the expected retention time within a fixed time range, and if it does exist, the degree of similarity between the mass spectrum of that peak and each of the registered spectra is calculated. If the degree of similarity obtained is greater than a predetermined value, the target compound is judged to be present. For a substance that is

judged to be present, the ratio between the peak area of that substance and that of the internal standard is obtained, and the amount detected is calculated from the calibration curve recorded in the AIQS-DB.
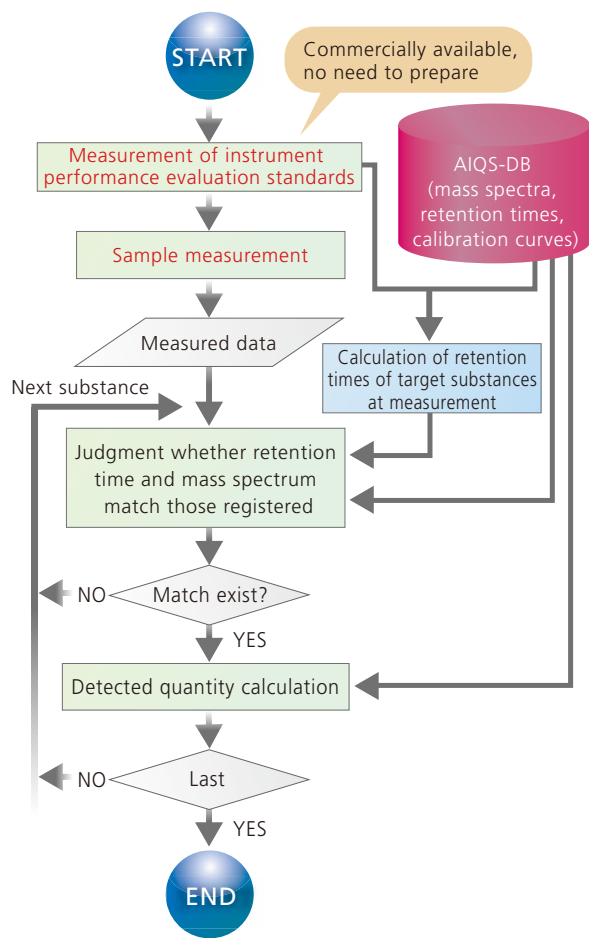

Fig. 1 Chemical Substance Measurement Flow Using AIQS-DB

7. Retention Time Prediction

The most important piece of information for target substance identification using GC is the retention time. When using the AIQS-DB, the predicted retention time of a target substance at sample measurement is calculated using the following equation, with the retention times of n-alkanes at sample measurement and the retention times of the target substance and n-alkanes recorded in the AIQS-DB.

$RT_{T'} = RT_{C'n} + (RT_T - RT_{Cn}) \times (RT_{C'n+1} - RT_{C'n}) / (RT_{Cn+1} - RT_{Cn})$ …
Equation 1

Here, RT_T, RT_{Cn}, and RT_{Cn+1} are the AIQS-DB-registered retention times of target substance T and those of n-alkanes C_n and C_{n+1} that appear before and after T. In addition, $RT_{T'}$, $RT_{C'n}$, and $RT_{C'n+1}$ are the retention times of target substance T and n-alkanes C_n and C_{n+1} at the time of sample measurement.

8. AIQS-DB Identification and Quantification Performance

To reliably conduct identification and quantification using the AIQS-DB, two conditions must be satisfied. The first is that database-recorded substances in the measurement sample are detected with certainty, while database-recorded substances that are not in the sample are not mistakenly detected. In other words, there must be few false detections, and absolutely no non-detections.

In order to satisfy these contradictory conditions, the peak detection retention time must be accurately predicted, and the mass spectrum of the target substance must be reliably determined to exist at that retention time.

The second condition is a high reliability of the quantitative value. If only the presence or absence of a substance needs to be confirmed, even existing mass spectral databases can be used to some extent. If these two conditions are satisfied, even without measuring standard substances at the time of sample measurement, identification and quantification of substances registered in the AIQS-DB can be accomplished, thereby permitting reduced usage of hazardous substances, less time required for analysis, and an overall reduction in the cost of analysis.

8-1. Instrument Performance Evaluation and Maintenance Using Instrument Performance Evaluation Standard Substances

In order to obtain reliable measurement values using the AIQS-DB, the instrument performance should be as close as possible to that at the time the AIQS-DB was constructed. Factors that affect the identification and quantification include MS tuning, capillary column performance (dirt, degradation, etc.), degree of inertness and contamination of the injection unit, state of the ion source (fouling), and the impact of contaminants in the measured sample. Therefore, we investigated performance probe compounds that could be used to evaluate changes in the instrument state that occur due to repeated sample measurement.

After conducting 160 injections of a food sample (typically containing many contaminants) into the GC-MS, comprehensive maintenance was conducted, including replacement of the injection unit insert, column removal, and ion source cleaning and replacement. In addition, during these procedures, 220 different performance evaluation substance candidates were measured a total of 29 times each to evaluate changes in sensitivity and retention time.

The results indicated that the quantitative values of 115 substances changed by more than ±20% from their respective initial values with sample injection, the retention times of 5 substances showed discrepancies greater than 6 seconds with respect to the expected retention times, and as a result of ion source maintenance, the quantitative values of 77 substances decreased to less than 70% of their initial values. The performance evaluation substances shown in Table 3 were designed based on the above-mentioned

results and the materials that are adopted by the EPA.

The actual evaluation procedure starts with generation of a calibration curve (for example 0.2 ppm – 2.0 ppm) using the GC-MS when its state of performance is good. Then, the instrument performance is evaluated by measuring a 1.0 ppm standard solution at the time of sample measurement and then checking the measurement values, degree of tailing, mass patterns, and some other parameters.

Table 3 Instrument Performance Evaluation Standard

Substance	Evaluation Item
2,3,5,6-Tetrachlorophenol	Injection unit insert, column, ion source
Pentachlorophenol	Injection unit insert, column, ion source
Isofenphos oxon	Injection unit insert, column, ion source
Benzidine	Injection unit insert, column
Thenylchlor	Injection unit insert, column
4,4'-DDT	Injection unit insert
Butylbenzyl phthalate	Injection unit insert
Chlorothalonil (TPN)	Injection unit insert
1,4-Benzenediol	Column, ion source, retention time
4-Nitrophenol	Column, ion source
Captafol	Column, ion source
Dibutylamine	Column
2,6-Dimethylaniline	Column
2,4-Dichloroaniline	Column
2,6-Dimethylphenol	Column
2,6-Dichlorophenol	Column
Octanol	Column
Nitrobenzene	Ion source
2,4-Dinitrotoluene	Ion source
Tributyl phosphate	Ion source
2,4,6-Trinitrotoluene	Ion source
2,4-Dinitroaniline	Retention time
Tris (2-chloroethyl) phosphate	Retention time
1,2,4,5-Tetrabromobenzene	Mass pattern
DFTPP	Mass pattern
trans-Nonachlor	Mass pattern
Triphenylmethane	Quantitative value stability
Dimethyl phthalate	Quantitative value stability
Atrazine	Quantitative value stability
n-Alkanes (C9 - C33)	Quantitative value stability

8-2. Retention Time Prediction Accuracy

In temperature-programmed GC, a method based on the programmed-temperature retention index using n-alkanes has been developed, and some mass spectrum databases use this technique for identification. Also, the method of fixing retention times based on the relationship between the injection unit pressure and the retention times of specific compounds is implemented. A retention time prediction method developed for the AIQS-DB is similar to that of the programmed-temperature retention index. In other words, a series of n-alkanes is measured together with the target substance under a given set of GC conditions, and the retention time of the target substance as well as the retention times of the two n-alkanes that sandwich the target are registered together in the database. When conducting actual sample analysis, only the series of n-alkanes is measured, and the retention time of the target substance is predicted using the above-mentioned Equation 1 based on those retention times.

Factors that determine the retention time in capillary GC analysis, with respect to the column, include column length, internal diameter, liquid phase, and film thickness, and, with respect to the carrier gas, include heating conditions as well as the gas type and linear velocity. Regarding the sample injection method selected for use with the AIQS-DB, we adopted the splitless injection method because it is most widely used for environmental and food analyses, although retention time may be affected depending on the solvent injected. Among these, we examined the factors that could be changed at the time of GC-MS analysis to determine the effect on retention time prediction accuracy.

First, we examined the accuracy of the predicted retention time value by changing only the column. A comparison of the predicted retention time with the retention time of the performance evaluation substance using 3 different columns indicated that the prediction error was within ±3 seconds in all cases, thus confirming that if the GC conditions remain the same, except for the column, as long as the quality of the mass spectra obtained from the sample remains constant, the first type of error (false detection) can certainly be prevented.

Next, we examined the effects of changing the programmed-temperature conditions, column length, film thickness, carrier gas linear velocity, and injection solvent. As a result, it was confirmed that major changes to those parameters cause a large discrepancy between the predicted retention time and actual retention time. In particular, internal standard substances associated with a large discrepancy were the deuterated polycyclic aromatic hydrocarbons.

Therefore, after exchanging or cutting the column, there was a difference in the behavior of the target substance and n-alkanes used for predicting retention time, such that there was a discrepancy between the predicted and actual retention times of some of the substances. This increases the possibility that the second type of error (non-detection error) will occur. The magnitude of such an error in the predicted value can be determined by measuring the performance evaluation substance. In other words, the search range for identification can be confirmed by determining the greatest error in the predicted retention time associated with the internal standard (typically perylene-d12). Furthermore, if necessary, the GC conditions for correctly predicting the retention time, for example, the carrier gas linear velocity suitable for the column length, can be obtained.

The type of solvent to be injected also affects retention time. Hexane solution is measured when the AIQS-DB is created. Therefore, if the solvent used has a boiling point higher than that of hexane, the retention time value for a substance with a short retention time will be larger than the predicted value. In such a case, the retention time is predicted by measuring n-alkanes prepared with the same solvent as that in the sample.

Changes in retention time due to a similar mechanism were also seen in samples that contained a great many contaminants, for example food samples. In this case, the search range must be widened because the retention time for a given substance will be longer than expected.

8-3. Reliability of Identification

Identification by GC-MS is conducted using the mass and retention time values. The retention time prediction accuracy, as described above, is comparable to that obtained using the conventional standard substance measurement method. As for mass-based identification, AIQS-DB utilizes scan measurement (total ion monitoring, TIM) to allow the obtaining of mass spectra, which provides better identification reliability than that obtained using the conventional method that relies on SIM.

Improving the accuracy of identification can be achieved by obtaining better quality mass spectra and more precise retention times. In other words, a target peak of extremely small intensity is affected by the background, and if overlapped by a large peak, identification accuracy will be adversely affected by its spectrum. Also, if the MS tuning conditions are very different from those used when database measurement was conducted, the mass spectra themselves will be different.

The following conclusions were drawn as a result of examining these situations. When MS tuning conditions for sample measurement were significantly different from those used for database construction, the measured mass spectra differed from those recorded in the database, sometimes causing non-detection due to extremely low similarity indices. Therefore, MS tuning conditions for sample measurement must be the same as those used when the AIQS-DB was constructed. Also, background subtraction was effective in minimizing the effect of background or other substance interference, allowing good spectra to be obtained. However, a high similarity index could not be obtained in cases where a peak with extremely large intensity overlapped the target peak. In such cases, the presence or absence of the target substance must be determined visually. Upon comparison of the identification accuracy using both forward and reverse searching, higher accuracy of identification was obtained with the reverse search in cases where the quality of the mass spectrum was poor due to small peak intensity or interfering peaks.

8-4. Reliability of Quantitative Values

An important objective of the AIQS-DB is that once calibration curves are registered to the database, accurate quantitative values should be obtainable anytime and anywhere. Therefore, using 114 substances having a wide range of physiochemical properties, including those that are difficult to measure with a low polarity column, we examined the variation in measurement values and its causes by measuring repeatability, intermediate precision, and reproducibility with 4 GC-MS instruments at 4 institutions. Here, the

reproducibility results were separated into 3 categories, and are shown in Table 4.

Table 4 Reproducibility Test Results

Reproducibility	RSD, % (Number of Substances)
Up to 10%	41% (47)
10 − 20%	37% (42)
Greater than 20%	22% (25)

(I) If the GC-MS is maintained at a level that provides acceptable performance, among phthalates, organochlorine pesticides, heterocyclic aromatics, polycyclic aromatic hydrocarbons, benzenes, aromatic amines, phenols, and nitrobenzenes, many of the substances having one highly polar functional group (hydroxyl group, amino group, nitro group) show measurement accuracy less than RSD 10%, and are measured with accuracy comparable to that by the internal standard method. (II) The following substances show a range of RSD of 10 − 20%: Polycyclic aromatic hydrocarbons with a high boiling point; substances with a quantification ion m/z value of 250 or more containing large amounts of chlorine or bromine; phosphate esters; and substances with relatively high water solubility (log $P < 3$) containing a high-polarity functional group such as a hydroxyl group, amino group, or nitro group. Although quantification was inferior to that provided by the conventional method, it is feasible depending on the purpose. (III) With the current GC-MS performance and GC conditions for the AIQS-DB for semi-volatile chemical substances, in particular with the low polarity column, the range of RSD is 20 − 70% with substances having at least 2 high-polarity functional groups such as m-aminophenol, 2,6-diaminotoluene, 4-methyl-3-nitrophenol, and bentazone, and some alcohols and phosphate esters such as 2-ethyl-1-hexanol and tris (2-ethylhexyl) phosphate. However, although quantification with good accuracy is difficult, the method is acceptable for use in screening.

There are measures that can be taken for substances with poor reproducibility. In the case of low-polarity substances such as n-paraffin and halogenated substances belonging to group (II), these can be added as performance evaluation substances to help easily discover the nature of the problem, thereby permitting implementation of measures that should improve the reproducibility. On the other hand, for substances belonging to group (III), the problem extends further than the AIQS-DB method, requiring construction of a database using different GC conditions to accurately measure these substances.

The official method specified by Japan's Ministry of the Environment for measurement of toxic substances such as dioxins using GC-MS requires that a calibration curve be generated at the time of sample measurement, and that the variation of obtained values be within 20% of the expected value.

The results of this study indicate that this criterion is satisfied for the majority of substances, and that by properly managing the instrument, accurate measurement values can be obtained even without measuring standard solutions.

8-5. Detection Limit

The limit of detection of the instrument is inferior to that of SIM since the AIQS-DB system employs scan measurement (total ion monitoring, TIM). However, the S/N = 3 for approximately 90% of the substances was equivalent to less than 10 pg. In other words, if 1 L of aqueous sample were concentrated to 1 mL, the detection limit would be less than 10 ppt. This value, with the exception of the ultra-trace analysis that is required for some chemical substances, such as dioxins, provides a sufficiently low detection limit.

9. Summary

We have developed an AIQS-DB system for semi-volatile chemicals, and evaluated its identification and quantification performance. In the study, retention time was predicted accurately using the programmed-temperature retention index, and reliable identification was possible based on the predicted retention time in conjunction with the mass spectrum. Also, fairly accurate quantification of non-polar to medium-polarity substances was achieved with a low polarity column by maintaining instrument performance at a consistent level using performance evaluation substances. Currently, the AIQS-DB for semi-volatile chemical substances can be used for measurement of about 1000 substances, but by building similar databases with increased numbers of registered substances as well as support for other col-

umns, measurement of more generic as well as comprehensive chemical substances will become possible.

The advantage of the AIQS-DB is that it can be used for identification and quantification of more than 1000 substances, without using chemical standards. Thus, low-cost, labor-saving, low-resource, waste-reducing, and quick measurements can be conducted. In environmental analysis, the overall picture of the environment can be grasped as never before possible. In other words, the system helps identify the source of chemical substances at the site of examination, for example, the ratios of substances associated with agriculture, typical living activities, and industry, as well as determine business categories related to those substances. Also, it can effectively assist in ascertaining the cause of environmental contamination accidents, as well as confirming safety in emergencies. In analysis of foods and drinks, it is common knowledge that the higher the number of substances to be measured, the more effective the safety verification. Since chemical standards are not required, this system can be applied to analysis of difficult-to-obtain illicit drugs and explosives, as well as the analysis of metabolites, suggesting its usefulness in the field of medicine. With advancements in technology, progress is being made toward higher sensitivity, higher selectivity, and better stability of analytical instruments. And, with the features presented here, conducting analysis using the AIQS-DB system may become commonplace in the future.

References

1. *Development of a novel GC/MS database for simultaneous determination of hazardous chemicals,* Kiwao Kadokami, Kyoko Tanada, Katsuyuki Taneda, Katsuhiro Nakagawa, *BUNSEKI KAGAKU,* 53, pp 581-588, 2004.

2. *Novel gas chromatography-mass spectrometry database for automatic identification and quantification of micropollutants,* Kiwao Kadokami, Kyoko Tanada, Katsuyuki Taneda, Katsuhiro Nakagawa, *J. Chromatogr A,* 1089, pp 219-226, 2005.

3. *Reproducibility of measurement results by automated identification and quantification system with database for GC-MS,* Terumi Miyazaki, Kiwao Kadokami, Yuichi Sonoda, Daisuke Jinya, Takashi Yamagami, Kenichi Toubou, Hiroaki Ogawa, *BUNSEKI KAGAKU,* 60, pp 543-556, 2011.

4. *System performance check standards for gas chromatography-quadrupole mass spectrometry in simultaneous analysis of semi-volatile compounds*, Daisuke Jinya, Tomomi Iwamura, Kiwao Kadokami, Tetsuya Kusuda, Haruhiko Miyagawa, Katsuhiro Nakagawa, Tomoaki Kondo, *BUNSEKI KAGAKU,* 60, pp 965-975, 2011.

SHIMADZU

Shimadzu Corporation

www.shimadzu.com/an/