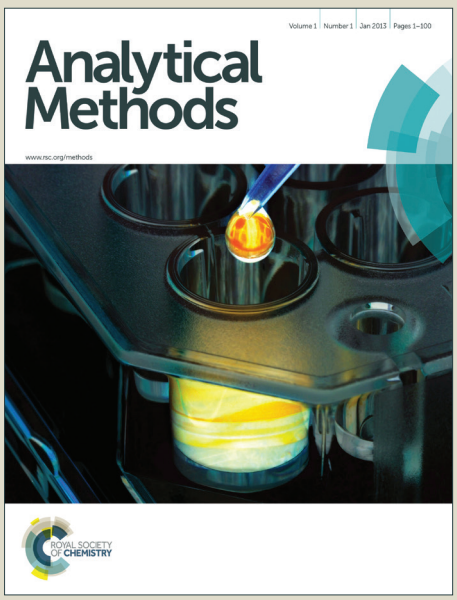
Analytical Methods

Accepted Manuscript



ROYAL SOCIETY OF CHEMISTRY

www.rsc.org/methods

Automatic integration method for single and multiple peaks in the GC and GC-MS chromatograms of characteristic oil compounds

Xinping Wang[a, b, c], Yuhui Zhao [c], Peiyan Sun [c], Min Ji [d], Mutai Bao[a,b*]

[a] *Key Laboratory of Marine Chemistry Theory and Technology, Ministry of Education, Ocean University of China, Qingdao 266100, China*

[b] *College of Chemistry and Chemical Engineering, Ocean University of China, Qingdao 266100, China*

[c] *Key Laboratory of Marine Spill Oil Identification and Damage Assessment Technology, North China Sea Environmental Monitoring Center of State Oceanic Administration, Qingdao, Shandong 266033, China*

[d] *College of Geomatics, Shandong University of Science and Technology, Qingdao 266590, China*

ABSTRACT: In oil fingerprinting studies, hundreds of compound peaks (including saturated and aromatic hydrocarbons) need to be integrated for the identification and quantification of characteristic oil compounds. The speed and quality of integration are the key factors that influence peak identification and quantification. This influence is observed not only because of the time-consuming nature of manual peak integration but also because different instrument operators may obtain different results due to different peak integration skills, especially for alkylated-PAHs, which need to be integrated into a series of peaks from one

* Corresponding author: Mutai Bao, *E-mail address*: mtbao@ouc.edu.cn (M. Bao); *Full postal address*: 238 Songling Road, Qingdao 266100, Shandong, PR China; *Tel:/Fax*: +86-532-66782509.

whole peak group. This paper describes an automatic integration method developed for characteristic oil fingerprinting compounds, including auto-recognition of single and multiple peaks, identification of bifurcated peaks, determination of the baseline and area integration for single peaks, multiple peaks and UCM (unresolved complex mixture) based on trapezoid summation theory. This method has been programmed and used in an oil data analysis and identification system in China. More importantly, this method has been applied for four years in a number of real-world oil spill case investigations and has been demonstrated to be accurate and efficient.

KEYWORDS: GC and GC-MS chromatogram; automatic integration method; single peak; multiple peaks; overlapped peaks; characteristic oil compounds

1. Introduction

With the development of the economy, marine transportation and oil exploration, the risk of offshore oil spills is increasing, and mysterious oil spills also occur more frequently. For many ocean oil spill accidents, it is difficult to find the source of the spilled oil quickly, leading to pollution disputes, plague victims, and problems for transportation and oil development companies as well as administrative departments. Therefore, identifying oil spill sources plays a vital role in catching the perpetrators and taking effective emergency response measures to mitigate the losses of an oil spill.

There are many instrumental and non-instrumental techniques that are currently used in the analysis of oil hydrocarbons, including gas chromatography (GC), gas chromatography–mass spectrometry (GC–MS), high-performance liquid chromatography (HPLC), size exclusion HPLC, infrared spectroscopy (IR), supercritical fluid chromatography

(SFC), thin layer chromatography (TLC), ultraviolet (UV) and fluorescence spectroscopy, isotope ratio mass spectrometry, and gravimetric methods.[1]

GC and GC/MS are the most widely used methods. Petroleum contains thousands of different organic compounds.[2,3] Of these, there are many characteristic compounds that can be determined by GC and GC-MS. Based on the unique distributions of hydrocarbons in different oils, GC and CG-MS can identify spilled oil and its source. In most situations, it is necessary to integrate peaks for quantitative analysis to obtain unbiased and valid results.[4] After GC and GC-MS analysis, it is important to check and revise the peak integration, as the instrument workstation software is not always perfect, which can lead to integration that is not always logical and may cause variability.[5] There are many subjective factors involved, such that different chemists can have varying integration results. Therefore, integration is very time-consuming, and an experienced chemist is needed to do this job to eliminate possible integration errors, significantly reducing work efficiency. Thus, a smart and automatic integration program is badly needed for oil spill analysis.

Much research has been conducted on peak integration, and many peak functions have been created to stimulate peak shape; the integration method and software are quite developed.[6,7] There are many integration programs, as every instrumental company, such as Agilent, ThermoFisher, Shimadzu, Waters, and Dionex, has their own workstation for integration. Additionally, independent software for peak integration has been developed, including PowerChrom and ezdata. These programs are designed for multipurpose and accurate integration, and many chromatograph and integration theories are used when performing the calculations. However, real peaks can have a variety of shapes, and if too

many factors are considered in the calculation, it could lead to a lot of uncertainty and result in unreasonable. Thus, we attempted to use a simple way to identify peaks and calculate their areas based on the operator's intuitive perception, not recondite integration theories.

Based on these considerations, an auto-integration method for GC and GC-MS chromatograms was designed and programmed. A reasonable baseline drawing method was designed that can ensure that every integrated peak has the appropriate baseline, which would be accurate and require little manual revising of peak heights and areas. Furthermore, an easy and convenient multi-peak integration method was designed to integrate the multi-peak of PAHs quickly and automatically. An unresolved complex mixture (UCM) integration method was also designed to quickly capture the area of UCMs on GC chromatograms. All of these methods were programmed and used in an oil data analysis system, which has been used for four years in our lab, thereby demonstrating its actual effectiveness. And the comparison with 2 instrument workstations (Chemstation of Agilent and Labsolution of Shimadzu) proved that this program is very accurate in integration and convenience in operation.

2. Theory and Method for Oil Chromatogram Integration

2.1. Rationale

The target of this work was to develop an automatic integration program for oil fingerprint data treatment in our oil fingerprint database system. This can be done by any instrument workstation, but too many unreasonable integration results are obtained by workstations, and much manual revising is usually required, especially when peak shapes are not satisfactory. The goals of the program to be developed were to obtain reasonable results quickly and

minimize manual operation. The integration method was created according to the intuitive judgment of the integration technician rather than chromatographic theory, thus the result approximates manual integration.

As the GC method and target compounds are relatively fixed, the mission is simple, that is, to find the target peaks according to retention times, and calculate their peak areas. There are also some special demands in oil fingerprint data integration, such as the fact that homologous series of PAHs need to be treated as a single peak, and the areas of UCMs need to be integrated. Thus, the main functions of the program are as follows:

Recognition of single peaks (determine the beginning, end, and baseline);

Recognition of multi-peaks;

Peak area calculation;

Ensure the integration method is sufficient for small and unresolved peaks.

2.2. Auto recognition of single peaks

2.2.1. Identification of the peak. Oil chromatograms were acquired using analytical instruments, and the coordinate informations of chromatograms were stored in a database as a binary data stream. The chromatogram point information of the ions in a sample was read and wrote to a double-typed 2-dimensional array.

In the array, the first dimensionality represents the retention time (min) and the second dimensionality represents the signal intensity of a chromatogram point. When the X and Y data were read, a chromatogram plot could be drawn in the chromatogram module, and a series of calculations could be performed based on the chromatogram plot.

Normally, the peak beginning and end are determined based on the slope. In this program, because all of the peak points are stored, we can easily determine the peak beginning and end by comparing the intensity between adjacent points.

A whole peak consists of a peak top, left valley and right valley. When conducting peak recognition, the peak top should be identified first. This was done as follows:

The chromatogram curve consists of many points: (x_1, y_1), (x_2, y_2),…, (x_n, y_n). If $y_n > y_{n-1}$, and $y_n > y_{n+1}$, then the point (x_n, y_n) is a peak top.

Likewise, we can identify a peak valley as follows:

If $y_n < y_{n-1}$, and $y_n < y_{n+1}$, then the point (x_n, y_n) is a peak valley.

The retention time of the target peak should be set as a parameter before recognition. From setting the retention time toward the left and right, the first two peak tops are found as described above. By comparing the 2 peak tops and selecting the nearest one to the set retention time as the peak top of the target peak, the corresponding *X* is the real retention time of the peak. From the retention time toward the left and right, the *Y* values of the points on the chromatogram curve are compared and we can find the left valley. The corresponding *X* is the starting time of this peak. In the same way, we can find the right valley and the end time.

2.2.2. Setting the baseline. Generally, there are 2 types of baselines in chromatogram integration: a horizontal baseline and a sloped baseline. To simplify the process, only horizontal baselines were allowed to be auto-integrated in this program. In this integration method, the baseline was set according to the minimum height around the target peak. In the program, a time band should be set for baseline drawing, and the band was set as 0.2 min by default and can be modified if necessary. The program will search for the minimum *Y* in the

time range from (retention time - band) to (retention time + band), and the *Y* is the baseline height.

2.2.3. Treatment of bifurcate peaks. Bifurcate peaks (Fig. 1) are often observed in chromatograms, some are due to partly overlapped enantiomeric fractions,[8] and some are due to low concentration single peaks. Many models and programs have been used for treating overlapping peaks.[9-12] However, in this program, it is not important to deconvolute the overlapped peaks, but rather to compare between samples. In this program, the main question was to determine whether there is one single peak or there are 2 (or more) separated peaks. To resolve this problem, a principle was set according to the resolving degree, i.e., resolving degree of peak height smaller than 1/3 will be treated as one single peak, otherwise it will be treated as separated peaks. If the bifurcate peak was identified as 2 separated peaks, the common valley drop method (VDM) was applied to draw the peak boundaries.

Once the target peak has been identified, the program will find its adjacent peaks and determine if they are the bifurcate peaks. From the left (right) valley toward the left (right), it will find the left (right) adjacent peak of the target peak. The target peak and its left and right adjacent peaks are obtained. The height between the top of the target peak and the left valley (H_0) are calculated. The height of the target peak is represented by H. If $H_0/H<1/3$, the 2 peaks are considered to be one bifurcate peak, and the 2 peaks are combined, and the summation of the areas of the 2 peaks will be the area of the bifurcate peak. Then, this bifurcate peak is set as the target peak, and its adjacent peaks are found and they are determined to be two separate or one bifurcate peak. This process is repeated until no more bifurcate peaks are found.

2.2.4. Manual revision of the auto-recognition results. No program can replace a human entirely, and this program allows users to check and revise the auto-recognition results. The operation of manual revising is very simple; users just need to click the mouse on the starting point and drag to the ending point, and then release the mouse. Both horizontal and sloped baselines are allowed in manual revising. If we choose a sloped baseline, the program will use a straight line from the starting point to the ending point as the baseline of the peak.

2.3. Auto-recognition of multi-peaks

In PAH integration for oil fingerprint analysis, a series of homologous peaks are always treated as a single peak. Auto-recognition of multi-peaks is relatively simpler than that of single peaks. The difference between multi-peak recognition and single peak recognition is the time setting. As there are many peaks in the time range of a multi-peak, and the peak amount and distribution patterns are not fixed for all samples, it is not possible or necessary to identify every single peak of the multi-peak. We only need to define the starting and the ending time of the multi-peak. When chromatogram data are obtained, we define the points nearest to the set starting and ending times as the starting and ending points of the target multi-peak. Then, the program searches for the highest point, which is the top of the multi-peak, and searches for the lowest point. The horizontal line passing through the lowest point is the baseline (Fig. 2). Integration checking and manual revision are also allowed for multi-peak integration.

The auto-integration results for multi-peaks may appear to be inaccurate, as the peak beginning and end may not be right in the valley. However, because the area of the

multi-peak is often very large, the slight deviation of the peak beginning and end does not lead to significant error, so manual revision might not be necessary.

3. Peak Area Calculation

3.1. Auto-integration for single and multi-peaks (horizontal baseline)

In manual peak area calculations, peak areas can be estimated by peak height and peak widths based on peak models.[13] These methods can be conveniently used in manual calculation and are quite accurate for the Gaussian or exponentially modified Gaussian peaks. In digital peak integration, because the curve actually consists of disscatered points and the computer can perform calculations quickly, it is easy to obtain the actual area covered by a peak curve.

Peak areas are determined according to the starting time, ending time, and baseline. As Fig. 3a shows, the curve between *PL* and *PR* is a whole peak. The horizontal line between *PL* and *PR* is the baseline, *h* is the height, and *pl* and *pr* represent 2 adjacent data points on the peak. The line from *pl* to *pr*, the vertical lines passing through *pl* and *pr*, and the baseline form a trapezoid. A peak consists of several trapezoids (mostly and one or two triangles). By calculating the summation of the areas of all trapezoids (and triangles), we obtain the peak area. The calculation process is as follows:

The *x*-axis points (retention time) of a peak are t_0, t_1, … t_n, and the heights (*y*-axis minus the baseline) are I_0, I_1, …… I_n, (t_0, I_0) and (t_n, I_n) are the starting and ending points, respectively. The area is calculated using formula (1).

$$A = (I_1+I_0)\times(t_1-t_0)/2 + (I_2+I_1)\times(t_2-t_1)/2 +\ldots\ldots +(I_{n-1}+I_n)\times(t_{n-1}-t_n)/2 \qquad (1)$$

As the time interval between 2 points is approximately equal, the area was calculated using formula (2).

$$A = (t_n - t_0) \times (I_0/2 + I_n/2 + I_1 + I_2 + \ldots\ldots + I_{n-1})/n \quad (2)$$

The area obtained using this formula is an approximate result because the real peak does not consist of only trapezoids under ideal conditions, the curve is shaped by some arcs, not only straight lines. Thus, the error depends on the sampling frequency of the chromatogram points; the more points the peak consists of, the smaller error is.

The method of multi-peak integration is the same as fir single peak, although the calculated time range is much larger, and many more peak tops and valleys are involved.

3.2. Integration for sloped baselines

The integration of sloped baselines is similar to that of horizontal baselines, with only slightly more work needed. First, the peak is integrated for a horizontal baseline. Second, the area of the triangle formed between the horizontal baseline and sloped baseline (Fig. 3b) is calculated. Then, the area of the triangle is subtracted from the area of the peak integrated with a horizontal baseline.

3.3. Integration for UCMs

After setting the starting and ending times of a UCM, the program will look for all of the valleys on the chromatogram curve during the set time range, connect all the valleys, and produce a new curve. Then, all the valleys on the new curve are connected to produce another new curve.. This process is repeated several times until an appropriate UCM curve is obtained. The number of repeat times must be set before the calculation. Generally, 3 or 4 repeats is satisfactory. The UCM area is obtained by setting the baseline as high as the lowest

point on the UCM curve, and integrating the area between the UCM curve and the baseline (Fig. 4).

4. Application of the Methods

These methods were programmed and used in our oil data analysis and identification system. This system has been used in our lab for oil data analysis for 4 years and over 3000 samples were treated by this program successfully.

Before using this system for auto-integration, we must first create a method table. The integration method table includes 12 columns: method name, ion, starting time, retention time, ending time, integration type, compound type, band, sequence number, editing user, and editing time. The first 10 columns must be edited by the user and the editing user and editing time are generated automatically by the system. There can be many rows in this table, with one for each compound. When we use this system for auto-integration, we select a sample and display its chromatogram, then select a previously created method, and press the integration button. Next, the program will conduct auto-integration for all the compounds specified in the method table, and generate a result table, which includes the same rows as method table, one for each compound, with compound name, height and area included in the result table.

4.1. Application in real spill case

We have used this system to conduct auto-integration for an oil spill case. 12 samples were analyzed. The weathering extents of all of the samples were very similar, and all of the samples were identical. We used the system to conduct auto-integration for terpanes, steranes, and PAHs, and the results were not manually modified. The area of each compound was

normalized to a stable and large peak (hopane for terpanes and steranes, and C4-phenanthrenes for PAHs).

The normalized peak areas of terpanes and steranes are displayed in Table 1.

As can be observed in Table 1, the relative abundances of most compounds were very similar for all 12 samples, and the RSDs were quite low. However, some of the RSDs were relatively higher, suggesting that the integrations should be checked.

Among all of the RSDs, the RSD for 5α,14β,17β,20*S*-cholestane was the largest (59%), which is obviously unacceptable. The integration of the largest peak (S4) and smallest peak (S10) were checked, and the reason why the difference is so large may be due to small peaks that can easily be incorrectly identified, which should be treated as "not detected" (Fig. 5a).

The RSD of 13β, 17α, 20*R*-cholestane was the second largest (11.1%). The integration (Fig. 5b) were checked and found to be reasonable, even though the peaks were very small and their shapes were irregular. Thus, the variability between samples is unavoidable. From Fig. 5b, it is clear that the program made a very reasonable identification of bifurcate peaks.

The RSD of 17α, 21β-25-norhopanehopane was the third largest (10.6%), and the integration was checked (Fig. 5c). The baselines are not consistent between the 2 samples, and the peaks are also very small, which led to the observed differences.

According to the analysis integration results, we conclude that the auto-integration program is satisfactory for terpanes and steranes.

The normalized peak areas of PAHs are displayed in Table 2.

The relative abundances of most compounds were very similar for all 12 samples (Table 2). The RSDs were quite low, with a few relatively high RSDs (naphthalene, C1-naphthalenes,

and benzo(a)fluorine). The relatively high RDSs were observed because these compounds are notably sensitive to evaporation and degradation.

We conclude that the auto-integration program is suitable for oil data treatment; the integration is highly accurate, and notably little manual revising is needed. In this program, the recondite integration theories was abandoned and the integration method was just designed based on the peak shape and experts' intuitive perception, the goals of this program are accuracy of integration and convenience of operation, and the integration application on thousands of samples showed that the goals were successfully achieved.

4.2. Comparison with 2 instrument workstations

To prove the practicability of the program, a sample was treated by this program and 2 instrument workstations (Chemstation of Agilent and Labsolution of Shimadzu) and the integration results were compared, the retention times for each compounds in this program and the compared instrument workstation are just the same.

The comparison result with Chemstation was shown in Table 3. The automatic integration result by this program is far more close to the revised result than the automatic Chemstation integration. The automatic Chemstation integration result needs much manual revising because the workstation always gave unreasonable integration result, see Figure 6.

The comparison result with Labsolution was in Table 4. The automatic integration result by this program and Labsolution are very similar, however, the integration result by this program is a little bit more close to the revised result. As an example, the automatic Labsolution integration (Figure 7a) for compound "18α-22,29,30- trisnorhopane" is not very correct, in my opinion, the correct integration should be as Figure 7b. The automatic

integration by this program is more correct than Labsolution.

From the comparison, it is cleat that the automatic integration result by this program is very reasonable and correct. In Chemstation and Labsolution, the automatic integration can be adjusted through revising the parameter such as “drift” and “T.DBL”, but the pamameter adjusting could affect all the compounds, it’s not easy to set a set of parameters suit for all the compounds. In this program, the parameter “drift” and “T.DBL” were abandoned, but the integration result is very reasonable. And there are some other advantages compared to Chemstation and Labsolution, such as, the checking and revising on integration result is more convenient than Chemstation, manual baseline setting is more convenient than Labsolution, and the integration of UCMs and multi-peaks is much more easy in this program.

5. Conclusions

An auto-integration method for oil chromatogram data was designed and programmed. This method demonstrated certain advantages: (1) appropriate baselines can be drawn for a peak; (2) multi-peaks can be conveniently integrated; (3) bifurcate peaks can be intelligently identified; (4) UCMs can be identified easily; (5) manual revision of peak integration (except for UCMs) is convenient, though little manual revision is necessary; (6) integration parameters and results are listed in 2 tables, and it is very convenient to set parameters and for data analysis.

Unlike popular integration software, this program is a specified method for oil fingerprint databases. It was developed with the goal of obtaining visually reasonable results using simple and direct principles and methods. The results demonstrated that it achieved its goal.

The auto-integration program has been used in practical oil spill response for 4 years in our lab. Thousands of real samples was tested the function of the program, and it proved to be effective.

Acknowledgments

This study was supported by Public Science and Technology Research Funds Projects of Ocean (201205012), the National Natural Science Foundation of China (41376084, 21077028), and the Open Foundation of Key Laboratory of Marine Spill Oil Identification and Damage Assessment Technology of SOA" (201402).

References

1 Z. D. Wang, M. Fingas, D.S. Page, *J. Chromatogr. A*, 1999, **843**, 369–411.

2 P. Y. Sun, M. T. Bao, G. M. Li, X. P. Wang, Y.H. Zhao, Q. Zhou, L.X. Cao, *J. Chromatogr. A*, 2009, **1216**, 830–836.

3 Z. D. Wang, M. Fingas, *Mar. Pollut. Bull.*, 2003, **47**, 423–452.

4 S. A. Stout, A. D. Uhler, K.J. McCarthy, *Environ. Forensics*, 2001, **2**, 87–98.

5 CEN, CEN/TR 15522-2 version 2, 2011.

6 J. P. Foley, J. G. Dorsey, *Anal. Chem.*, 1983, **55**, 730–737.

7 J. W. Li, *Anal. Chem.*, 1997, **69**, 4452–4462.

8 B. J. Asher, L. A. D'Agostino, J. D. Way, C. S. Wong, J. J. Harynuk, *Chemosphere*, 2009, **75**, 1042–1048.

9 A. W. Westerberg, *Anal. Chem.*, 1969, **40**, 1595.

10 J. W. Li, *J. Chromatogr. A*, 2002, **952**, 63–70.

11 H. W. Kong, F. Ye, X. Lu, L. Guo, J. Tian, G. W. Xu, *J. Chromatogr. A*, 2005, **1086**,

160–164.

12 R. Koradi, M. Billeter, M. Engeli, P. Gu¨ntert, K. Wüthrich, *J. Magn. Reson.*, 1998, **135**, 288–297.

13 E. Voigtman, *Appl. Spectrosc.*, 1991, **45**, 237–241.

Figure captions

Figure 1 Recognition of bifurcate peak

Figure 2 Area integration for multi-peak

Figure 3a Area integration for single peak (horizontal baseline)

Figure 3b Area integration for single peak (sloped baseline)

Figure 4 Drawing of UCM curve

Figure 5a Integration of 17α,21β- 25-norhopanehopane for S4 (left) and S10 (right)

Figure 5b Integration of 13β , 17α, 20R - cholestane (diasterane) for S9 (left) and S12 (right)

Figure 5c Integration of 5α, 14β, 17β, 20S-cholestane for S4 (left) and S11 (right)

Figure 6 A sample of unreasonable peak integration given by Agilent Chemstation

Figure 7 Integration result of automatic labsolution integration (a), manual integration (b), and automatic program integration (c)

Table captions

Table 1 Normalized Peak Areas of Terpanes and Steranes

Table 2 Normalized Peak Areas of PAHs

Table 3 Comparison result with Agilent Chemstation

Table 4 Comparison result with Shimadzu Labsolution

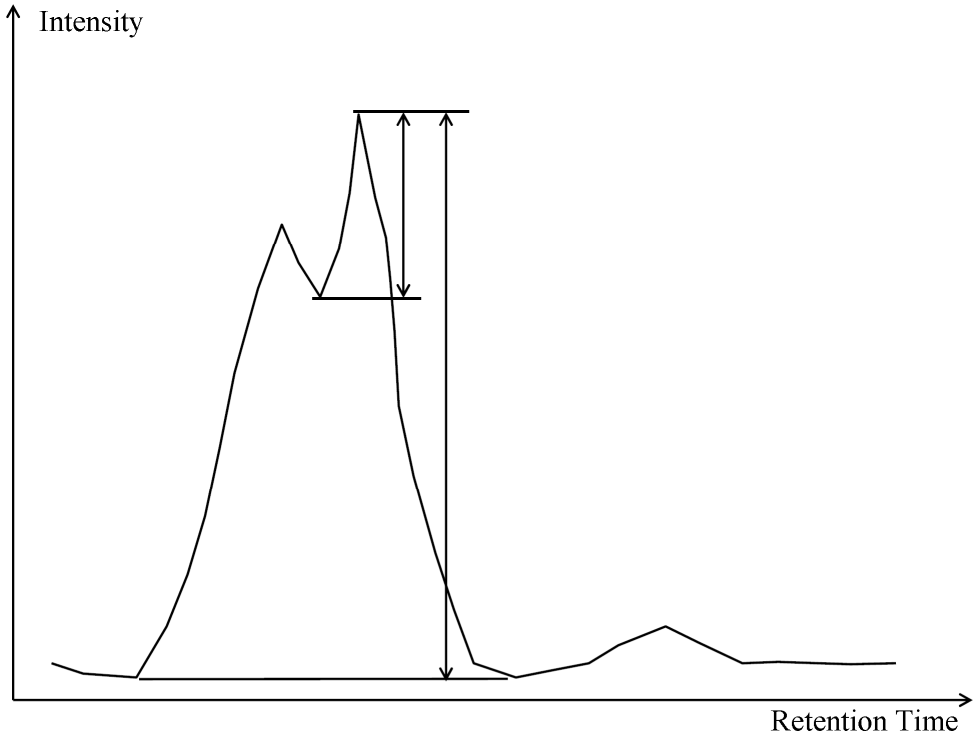


Figure 1 Recognition of bifurcate peak

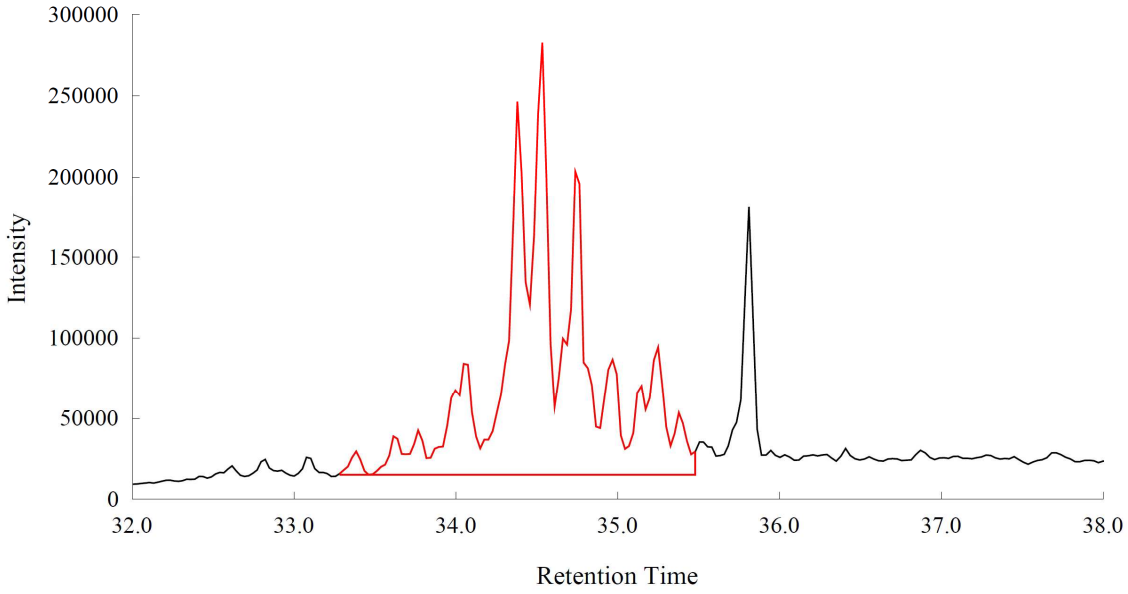


Figure 2 Area integration for multi-peak

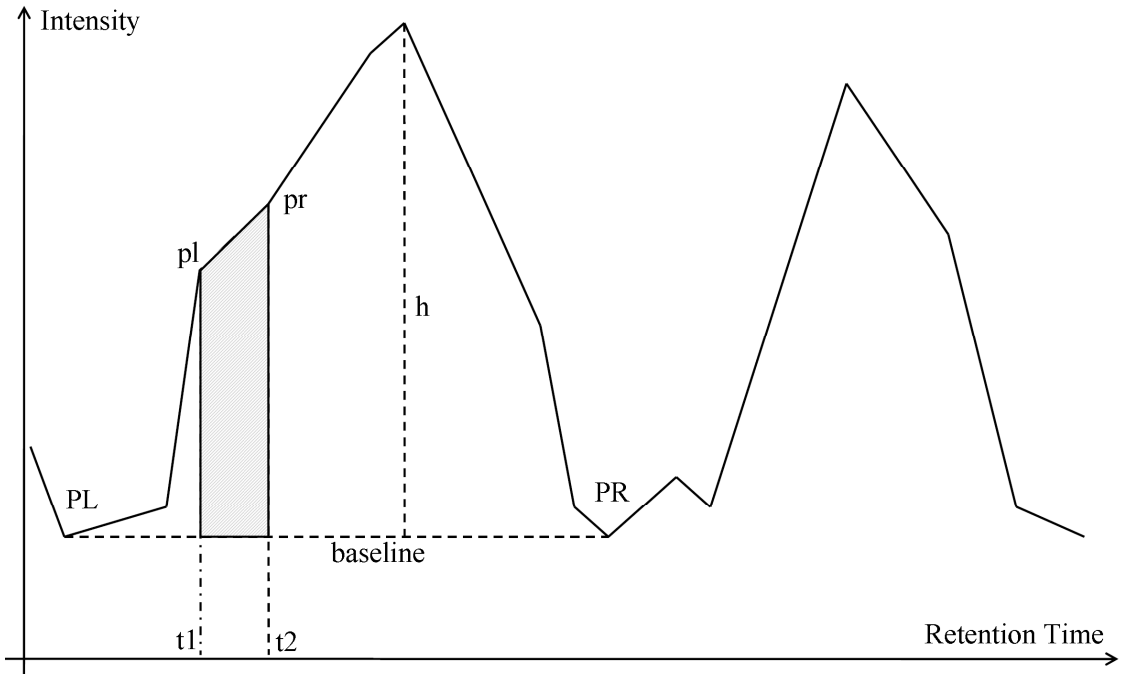


Figure 3a Area integration for single peak (horizontal baseline)

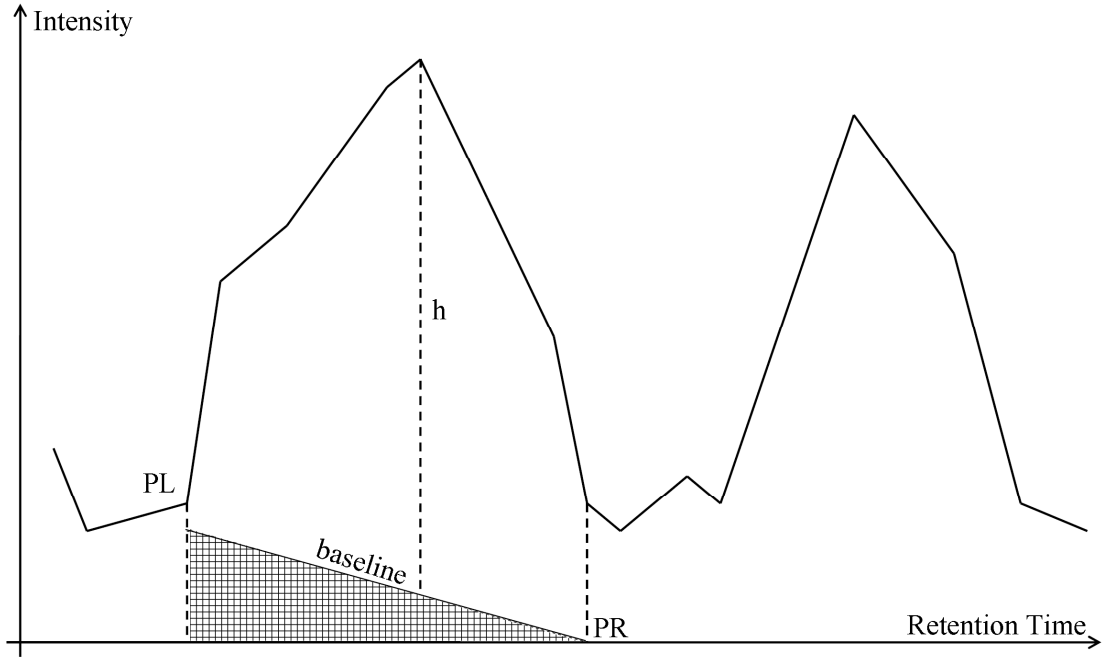


Figure 3b Area integration for single peak (sloped baseline)

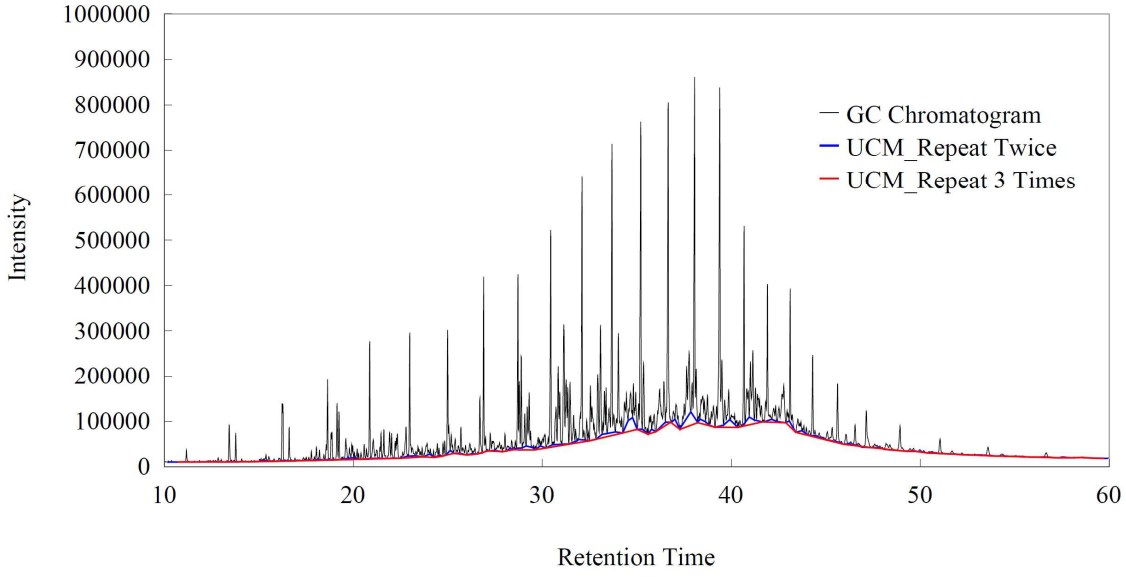
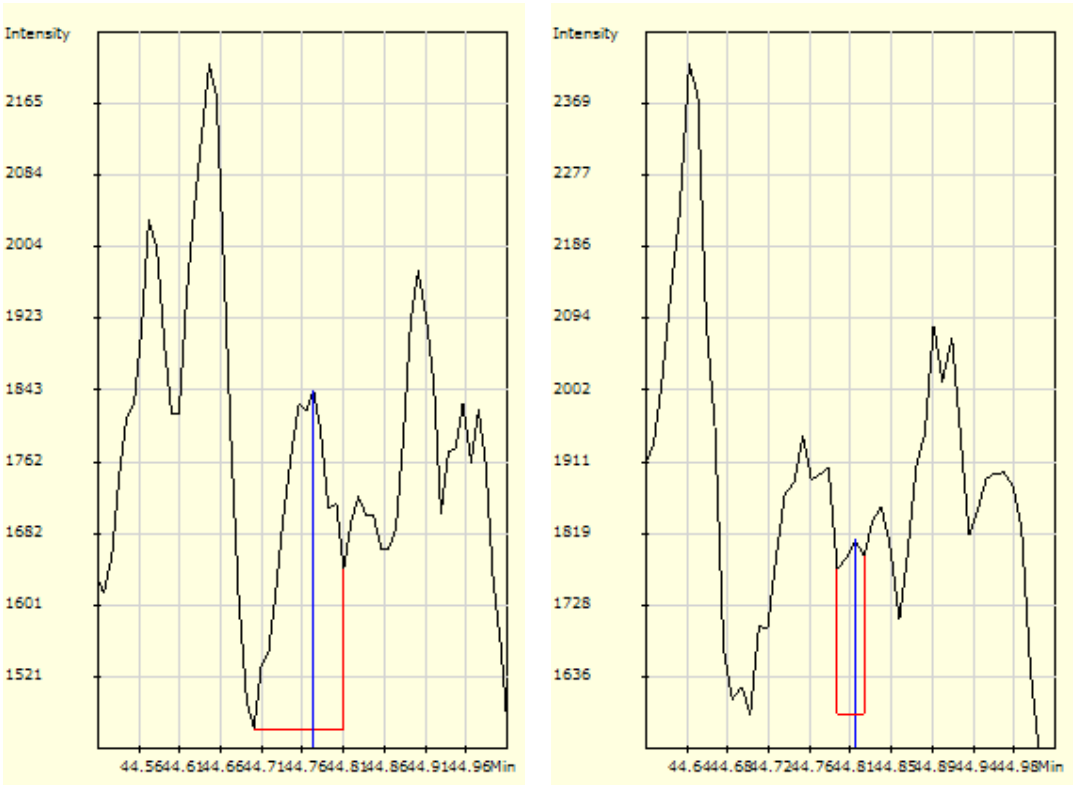
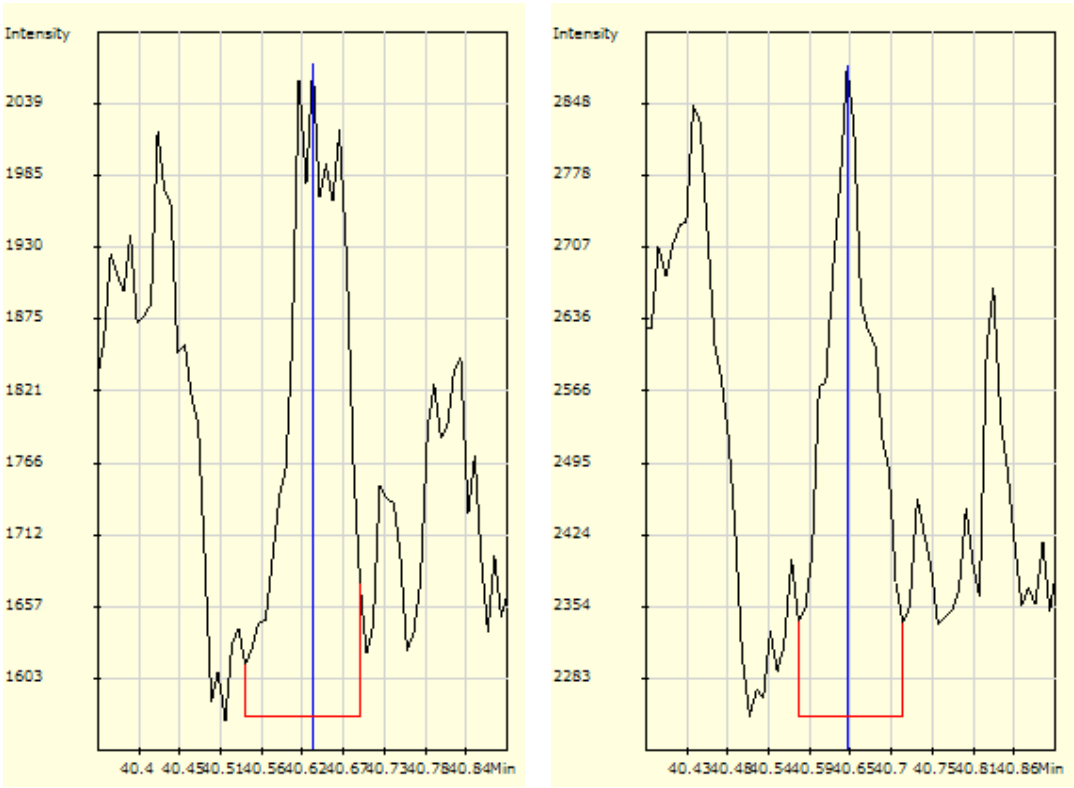


Figure 4 Drawing of UCM curve



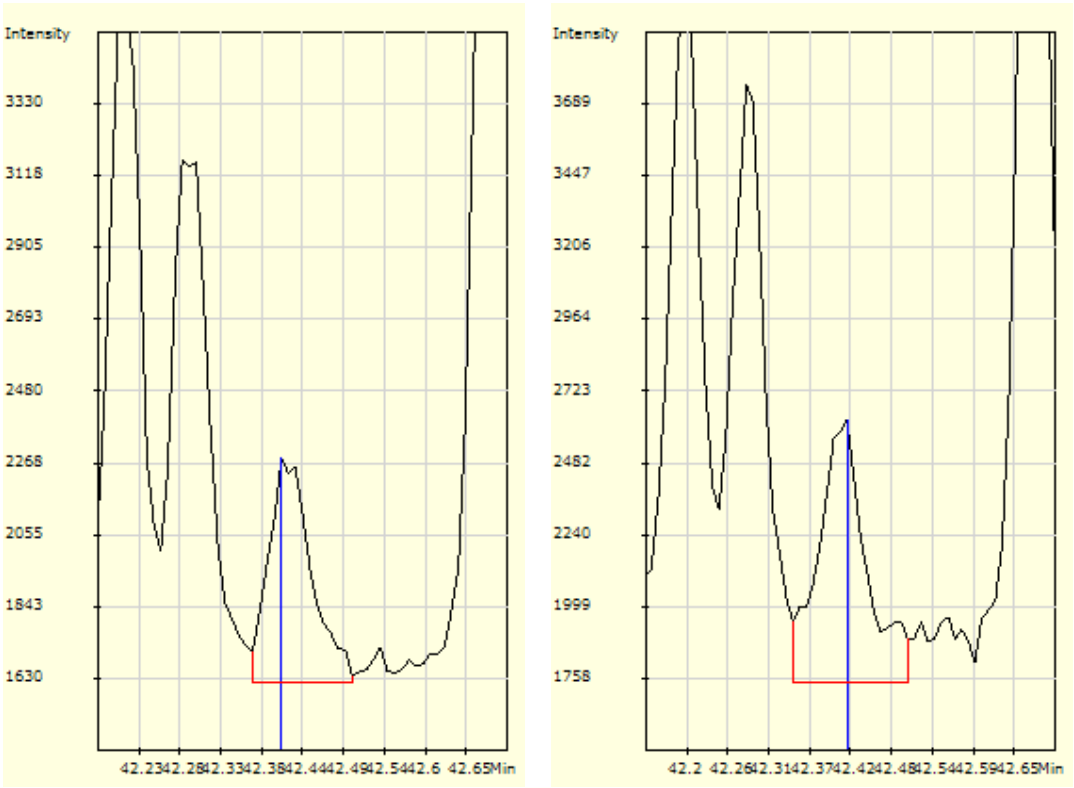
Integration for S4 Integration for S10

Figure 5a Integration of 17α, 21β-25-norhopanehopane for S4 (left) and S10 (right)



Integration for S9 Integration for S12

Figure 5b Integration of 13β, 17α, 20R-cholestane (diasterane) for S9 (left) and S12 (right)



Integration for S4 Integration of for S11

Figure 5c Integration of 5α, 14β, 17β, 20S-cholestane for S4 (left) and S11 (right)

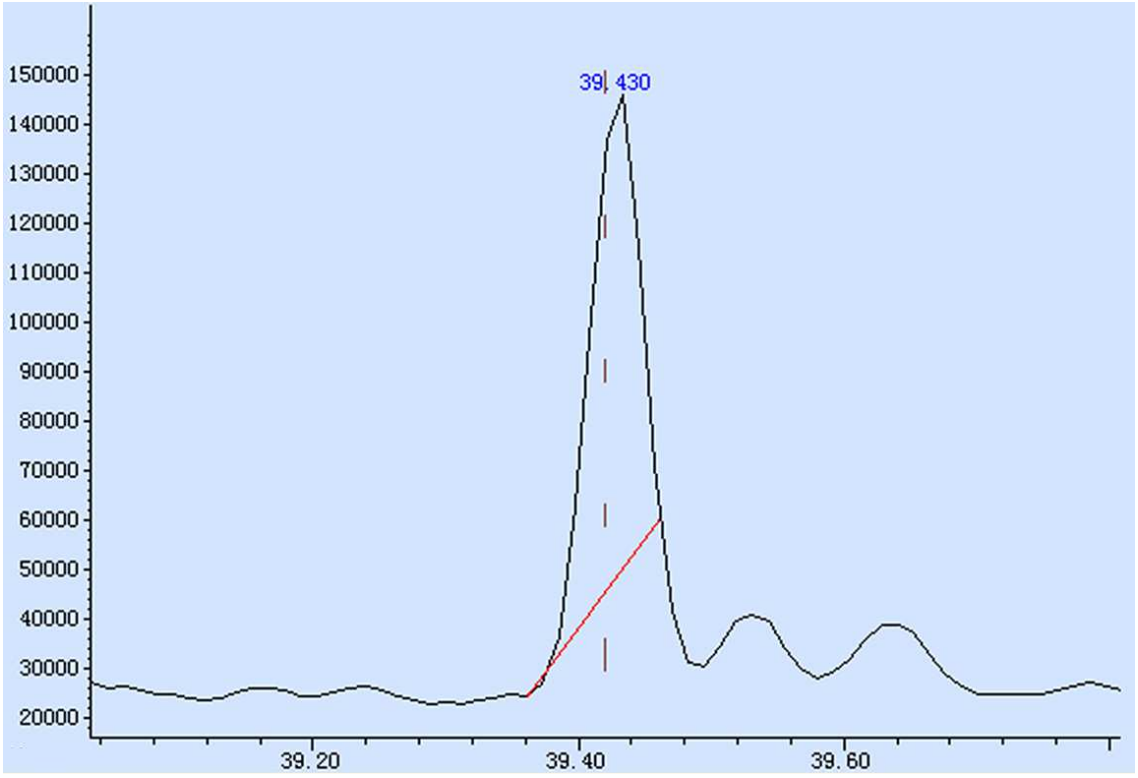


Figure 6 An example of unreasonable peak integration given by Agilent Chemstation

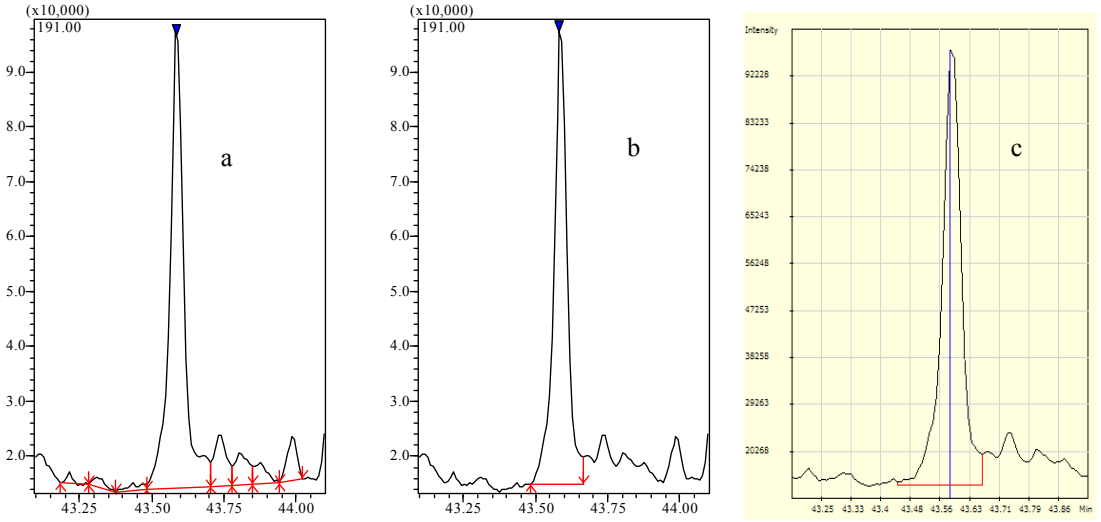
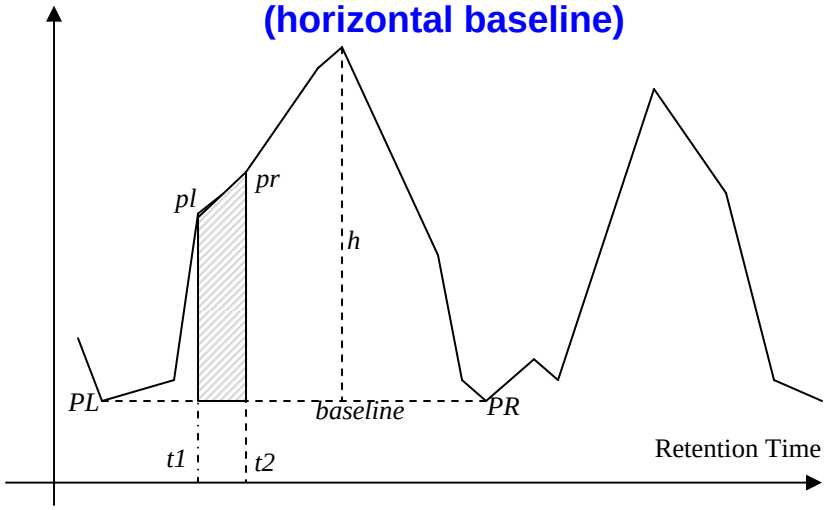
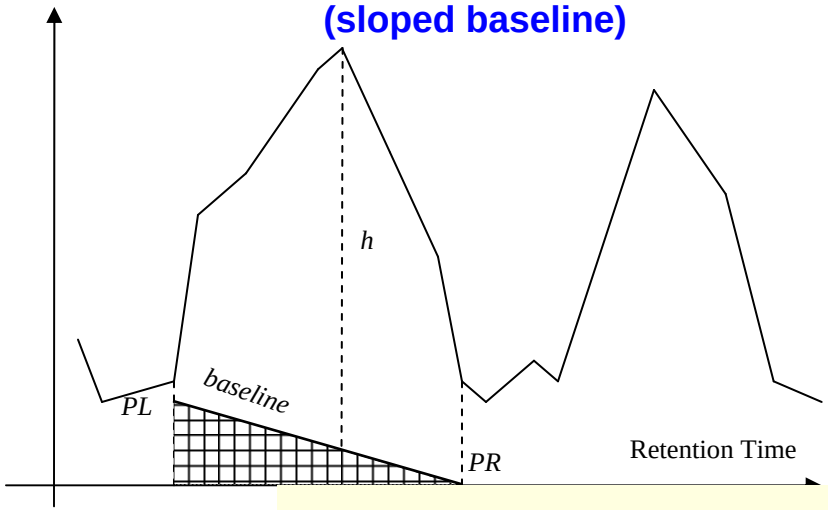


Figure 7 Integration result of automatic Shimadzu Labsolution integration (a), manual integration (b), and automatic integration (c) of this program

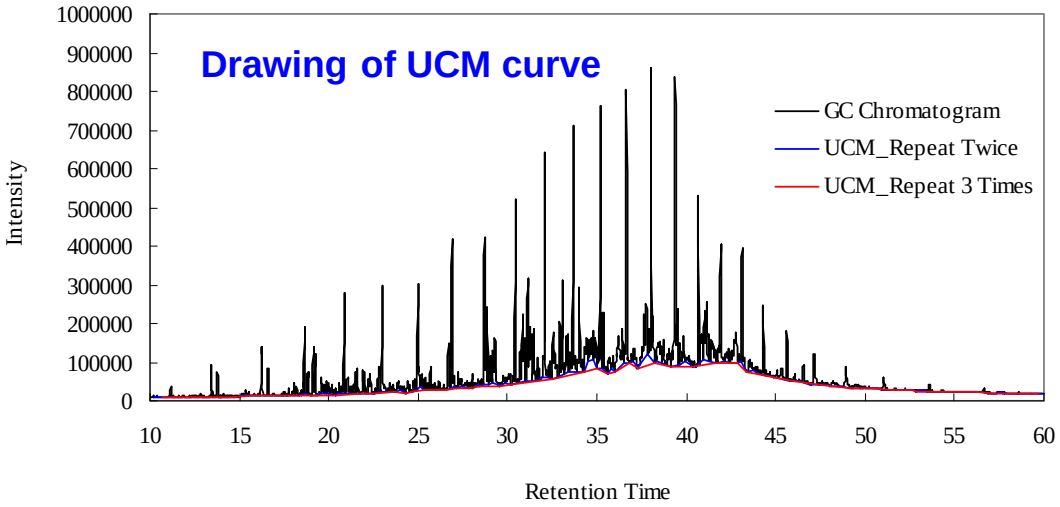
Area integration for single peak (horizontal baseline)



Area integration for single peak (sloped baseline)



Drawing of UCM curve



Integration of 13 β , 17 α , 20R-cholestane (diasterane) for S9

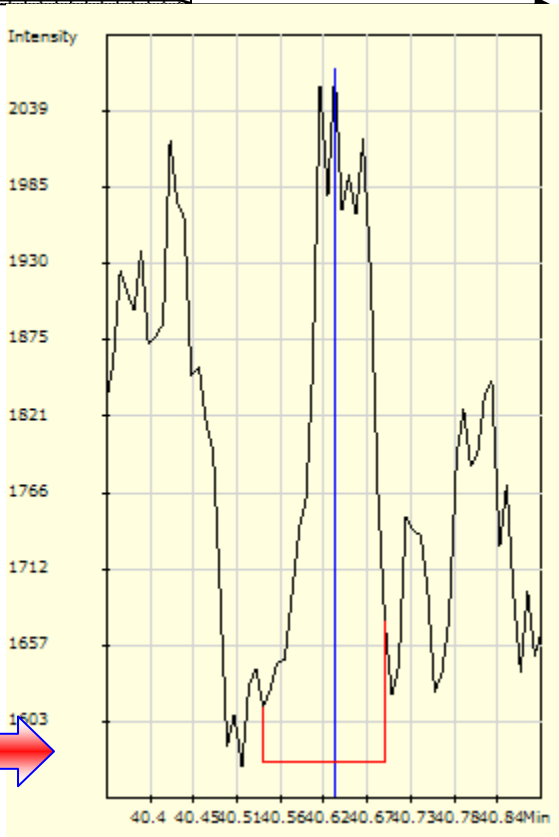


Table 1 Normalized Peak Areas of Terpanes and Steranes

Compound name	S1	S2	S3	S4	S5	S6	S7	S8	S9	S10	S11	S12	RSD %
C21 Tricyclic diterpane	20.5	19.6	20.0	19.8	19.7	20.4	21.0	19.8	19.3	20.4	20.1	19.7	2.3
C22 Tricyclic diterpane	19.5	20.0	19.9	20.3	20.7	20.8	20.1	18.5	19.5	20.7	19.9	19.2	3.5
C23 Tricyclic diterpane	29.2	29.6	29.8	29.0	29.2	28.7	29.7	28.2	29.5	30.0	29.1	28.2	2.0
C24 Tricyclic diterpane	10.4	11.3	10.7	11.3	11.3	11.1	11.5	10.6	9.2	10.5	10.6	10.0	6.0
C25 Tricyclic diterpane	6.4	6.2	6.5	6.1	6.4	6.3	6.4	6.3	5.8	6.6	6.2	6.5	3.7
C26 Tricyclic diterpane	2.3	2.6	2.2	2.3	2.5	2.5	2.6	**2.8**	2.4	**2.1**	2.4	2.2	**7.9**
C26 Tricyclic diterpane	13.1	12.9	12.9	12.5	12.7	13.1	12.8	12.4	13.2	12.9	13.0	13.4	2.1
18α-22,29,30-trisnorhopane	14.0	13.7	14.8	12.5	14.5	13.7	13.9	13.9	12.7	11.9	14.0	13.1	6.3
17α-22,29,30-trisnorhopane	33.1	33.1	34.4	33.3	32.5	33.1	33.6	31.9	32.0	34.0	32.7	33.5	2.3
17α,21β- 25-norhopanehopane	1.9	1.3	1.8	2.0	0.3	0.3	1.3	0.5	0.6	0.4	1.0	1.8	**59.1**
17α,21β-30-norhopane + 18α-30-norneohopane	82.4	82.4	84.5	80.9	80.2	82.1	81.2	78.7	79.1	83.6	81.3	81.8	2.1
15α-methyl-17α-27-norhopane (diahopane)	5.6	6.0	5.7	5.6	5.8	5.9	5.6	5.6	5.5	5.9	5.1	5.3	4.4
17β,21α-30-norhopane (normoretane)	31.5	31.4	33.1	31.5	30.9	31.8	31.4	30.4	30.3	32.7	31.3	31.4	2.6
18α-oleanane	4.1	4.0	4.6	4.1	3.9	4.4	4.1	4.2	3.9	4.4	4.1	4.1	5.2
17α,21β- hopane	100	100	100	100	100	100	100	100	100	100	100	100	0.0
17β,21α--hopane (moretane)	44.9	45.7	47.8	44.4	43.5	45.1	43.6	42.8	42.7	46.5	45.6	45.2	3.4
17α,21β, 22S-homohopane	26.5	25.5	26.2	25.6	24.9	25.2	25.3	25.1	24.8	26.0	26.0	25.5	2.1
17α,21β, 22R-homohopane	24.8	24.8	25.2	25.3	24.0	24.5	24.4	23.9	23.7	24.6	25.5	24.5	2.3
Gammacerane	18.6	18.0	17.2	**18.7**	15.8	17.6	16.1	15.6	**15.1**	17.6	18.3	16.5	**7.2**
17α,21β, 22S- bishomohopane	17.1	17.3	17.9	16.8	17.0	17.3	17.4	16.7	16.9	17.6	17.2	17.1	2.0
17α,21β, 22R-bishomohopane	17.7	17.8	18.2	17.7	17.2	16.9	17.8	16.9	17.3	18.4	17.9	17.6	2.7
17α,21β, 22S- trishomohopane	12.9	12.8	13.9	13.3	13.2	13.6	13.1	12.8	12.0	13.0	13.8	13.9	4.2
17α,21β, 22R-trishomohopane	10.8	10.5	10.8	10.6	10.4	10.7	10.7	9.8	10.2	11.1	11.0	10.9	3.4
17α,21β, 22S- tetrakishomohopane	10.6	10.3	11.0	12.1	10.1	10.7	10.8	9.4	10.9	11.3	10.9	10.8	6.3
17α,21β, 22R-tetrakishomohopane	6.4	7.1	6.7	7.0	6.3	6.8	6.1	6.5	7.5	6.8	6.4	6.5	5.8
13β , 17α, 20S - cholestane (diasterane)	4.7	4.6	4.6	4.5	4.6	4.9	4.7	**3.9**	**5.3**	4.8	4.1	4.5	**7.8**
13β , 17α, 20R - cholestane (diasterane)	**2.8**	3.3	3.9	3.3	**3.8**	3.6	3.1	3.3	**3.8**	3.2	3.2	**2.8**	**11.1**
5α, 14α, 17α, 20S- cholestane	7.9	8.2	8.2	8.3	8.5	8.7	8.1	8.6	8.4	8.3	9.2	8.3	4.1
5α , 14β, 17β, 20R-cholestane	6.7	7.0	7.1	7.1	6.9	7.2	7.0	7.0	6.4	6.9	7.1	7.0	3.1
5α , 14β, 17β, 20S-cholestane	3.4	3.7	3.6	**3.1**	**3.1**	3.7	3.2	4.1	3.3	3.4	**4.3**	3.3	**10.6**
5α, 14α, 17α, 20R- cholestane	15.0	15.3	16.0	15.1	15.7	15.2	14.2	15.0	14.1	14.6	15.2	14.3	3.9
24-methyl-5α, 14α, 17α, 20S- cholestane	5.6	5.8	6.0	5.6	5.3	6.5	6.0	5.7	5.7	6.5	6.0	5.8	5.8
24-methyl-5α, 14β , 17β, 20R- cholestane	8.9	8.9	8.4	8.0	8.4	7.9	8.4	8.2	8.2	8.1	7.9	8.0	4.1
24-methyl-5α, 14β , 17β, 20S- cholestane	5.4	5.2	**5.8**	5.2	5.0	5.1	5.5	5.6	4.9	**4.1**	4.2	4.9	**10.3**
24-methyl-5α, 14α, 17α, 20R- cholestane	15.3	15.2	15.9	15.0	14.7	14.6	15.7	15.4	14.9	14.8	15.0	15.3	2.6
24-ethyl-5α, 14α, 17α, 20S- cholestane	12.9	13.4	14.0	13.1	13.5	13.1	14.3	13.2	13.6	13.1	13.7	12.9	3.3
24-ethyl-5α, 14β , 17β, 20R- cholestane	10.7	10.6	11.1	10.6	10.6	10.8	10.8	10.1	10.2	11.3	10.6	10.7	3.0
24-ethyl-5α, 14β , 17β, 20S- cholestane	16.2	15.5	16.2	16.7	16.0	15.7	15.4	15.9	15.0	16.6	15.6	16.2	3.2
24-ethyl-5α, 14α, 17α, 20R- cholestane	33.5	34.4	34.0	34.5	32.5	35.0	32.6	33.4	31.6	35.3	34.3	35.0	3.4

Table 2 Normalized Peak Areas of PAHs

Compound name	S1	S2	S3	S4	S5	S6	S7	S8	S9	S10	S11	S12	RSD
Naphthalene	14.5	11.5	10.6	10.6	13.4	13.3	11.8	13.1	16.9	11.0	8.9	10.7	**17.8**
C1-naphthalenes	25.0	23.2	22.0	22.1	25.1	24.8	23.4	24.6	28.4	21.7	20.1	21.4	**9.5**
C2-naphthalenes	36.2	36.2	35.1	35.8	37.0	36.7	35.9	36.9	39.0	34.5	33.4	33.9	4.2
C3-naphthalenes	35.3	35.8	35.1	35.8	35.7	35.5	35.3	35.6	36.3	34.5	34.0	34.2	2.0
C4-naphthalenes	24.9	25.2	24.8	25.3	24.8	25.0	24.8	24.8	25.2	24.7	24.4	24.4	1.2
Phenanthrene	13.1	13.1	13.0	13.2	13.0	13.0	13.0	13.1	13.4	12.7	12.7	12.8	1.6
C1-phenanthrenes	51.8	51.7	51.2	51.6	50.3	51.7	51.7	50.2	51.1	50.6	50.6	50.6	1.2
C2-phenanthrenes	105.4	105.9	105.2	106.7	104.6	106.0	105.5	103.7	104.6	104.7	104.5	103.9	0.8
C3-phenanthrenes	129.9	129.4	128.9	130.0	129.1	130.0	129.4	128.5	128.4	129.6	129.4	128.8	0.4
C4-phenanthrenes	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	0.0
Dibenzothiophene	1.4	1.4	1.4	1.4	1.4	1.4	1.4	1.4	1.5	1.4	1.3	1.4	3.2
C1-dibenzothiophenes	7.5	7.4	7.3	7.5	7.3	7.5	7.5	7.3	7.6	7.3	7.2	7.3	1.6
C2-dibenzothiophenes	14.8	14.8	14.6	14.9	14.5	14.9	14.7	14.6	14.7	14.5	14.5	14.6	1.0
C3-dibenzothiophenes	22.4	22.2	22.2	22.4	22.5	22.5	21.9	22.7	22.3	22.2	22.3	22.1	0.9
Fluorene	3.7	3.7	3.5	3.6	3.7	3.7	3.6	3.7	3.9	3.4	3.4	3.5	3.8
C1-fluorenes	6.7	6.7	6.5	6.7	6.6	6.7	6.7	6.7	6.7	6.5	6.5	6.5	1.4
C2-fluorenes	11.0	11.2	10.9	11.2	10.7	11.1	11.0	10.7	10.9	10.9	10.9	10.9	1.4
C3-fluorenes	16.8	16.8	16.5	16.7	16.0	16.5	16.4	16.0	16.2	16.5	16.4	16.6	1.6
Chrysene	20.6	20.8	20.8	21.5	21.1	20.9	21.3	21.7	21.4	21.0	20.4	20.7	1.8
C1-chrysenes	86.2	87.9	87.5	88.6	89.2	88.7	88.2	90.9	90.6	86.4	85.8	85.2	2.1
C2-chrysenes	126.2	127.2	126.6	126.5	123.1	125.9	125.0	128.6	129.4	125.7	125.6	125.6	1.3
C3-chrysenes	98.7	97.1	97.9	97.0	94.3	99.3	96.0	98.7	98.7	93.4	96.0	94.3	2.1
Retene	13.6	13.6	13.5	13.5	13.7	13.7	13.6	13.6	13.7	13.7	13.6	13.5	0.5
benzo(a)fluorene	8.7	6.5	8.7	8.0	7.8	8.8	6.5	6.5	7.7	8.7	6.3	8.6	**13.3**
benzo(b+c)fluorene	4.1	4.2	4.1	4.1	4.1	4.0	4.1	4.0	4.0	4.0	4.0	3.9	1.8
2-methylpyrene	10.5	10.5	10.6	10.6	10.5	10.6	10.5	10.5	10.6	10.5	10.4	10.5	0.6
4-methylpyrene	10.9	10.9	10.8	10.9	10.9	11.0	10.9	10.9	10.9	10.7	10.7	10.7	0.9
1-methylpyrene	9.4	9.4	9.5	9.6	9.4	9.4	9.4	9.5	9.5	9.4	9.3	9.3	0.9
3-methyl phenanthrene	11.3	11.3	11.3	11.4	11.1	11.4	11.4	11.0	11.3	11.1	11.1	11.2	1.3
2-methyl phenanthrene	15.8	15.6	15.5	15.6	15.2	15.6	15.6	15.3	15.5	15.3	15.2	15.3	1.3
9/4-methyl	7.3	7.0	6.9	7.1	7.0	7.2	7.3	7.2	7.1	7.3	7.4	7.2	2.0
1-methyl phenanthrene	10.5	10.8	10.7	10.6	10.3	10.4	10.6	10.3	10.6	10.0	9.9	10.2	2.8
Methyl anthracene	5.9	6.0	5.8	5.8	5.7	5.9	5.9	5.7	5.8	5.8	5.8	5.6	2.0
4-methyl	2.1	2.1	2.1	2.1	2.0	2.1	2.1	2.1	2.1	2.1	2.0	2.1	1.5
2/3-methyl	2.7	2.7	2.6	2.7	2.6	2.7	2.6	2.6	2.6	2.6	2.6	2.6	1.5
1-methyl	1.0	0.9	0.9	0.9	0.9	1.0	0.9	1.0	0.9	0.9	0.8	1.0	5.8

Table 3 Comparison result with Agilent Chemstation

Compound name	Revised result based on Chemstation integration	Automatic Chemstation integration	Automatic integration by this program
18α-22,29,30-trisnorhopane	3679325	2595055	3985770
17α-22,29,30-trisnorhopane	2607096	2607096	2717100
17α,21β- hopane	27011720	27011720	27163250
17β,21α--hopane (moretane)	3319860	1016963	3397770
17α,21β, 22S-homohopane	8325821	3862791	8486450
17α,21β, 22R-homohopane	5519259	3163158	6669310
Gammacerane	4886579	1080413	4971760
17α,21β, 22S- bishomohopane	5494939	2309351	5569430
17α,21β, 22R-bishomohopane	4068985	2129766	4282170
17α,21β, 22S- trishomohopane	3734048	1164607	4008650
17α,21β, 22R-trishomohopane	2565843	966654	2835030
17α,21β, 22S- tetrakishomohopane	2364007	783614	2425430
17α,21β, 22R-tetrakishomohopane	1286933	389034	1429390
17α,21β, 22S-pentakishomohopane	1026517	298973	995950
17α,21β, 22R-pentakishomohopane	974761	167777	1019390
13β , 17α, 20S - cholestane (diasterane)	875154	729405	1010680
13β , 17α, 20R - cholestane (diasterane)	554617	244924	641920
5α, 14α, 17α, 20S- cholestane	1746686	1242645	1920740
5α , 14β, 17β, 20R-cholestane	1985664	333672	1912010
5α , 14β, 17β, 20S-cholestane	1439824	651749	1487250
5α, 14α, 17α, 20R- cholestane	1933364	1933364	1910700
24-methyl-5α, 14α, 17α, 20S- cholestane	1150217	1150217	1103540
24-methyl-5α, 14β , 17β, 20R- cholestane	1523910	1523910	1473020
24-methyl-5α, 14β , 17β, 20S- cholestane	1028367	1028367	1228520
24-methyl-5α, 14α, 17α, 20R- cholestane	1279681	1279681	1312430
24-ethyl-5α, 14α, 17α, 20S- cholestane	2805054	2805054	2883870
24-ethyl-5α, 14β , 17β, 20R- cholestane	1890045	1890045	1786450
24-ethyl-5α, 14β , 17β, 20S- cholestane	1506705	1773141	1221040
24-ethyl-5α, 14α, 17α, 20R- cholestane	2086082	2086082	2100170
Correlation coefficient with the revised result		**0.965278**	**0.998894**

Table 4 Comparison result with Shimadzu Labsolution

Compound name	Revised result based on Labsolution integration	Automatic Labsolution integration	Automatic integration by this program
18α-22,29,30-trisnorhopane	279773	303835	295857
17α-22,29,30-trisnorhopane	157830	168863	161057
17α,21β- hopane	1668819	1668819	1689690
17β,21α--hopane (moretane)	202810	202810	199404
17α,21β, 22S-homohopane	489772	492863	489233
17α,21β, 22R-homohopane	344790	391679	343986
Gammacerane	310971	282747	260726
17α,21β, 22S- bishomohopane	342561	342561	345251
17α,21β, 22R-bishomohopane	257429	268547	257762
17α,21β, 22S- trishomohopane	228652	228652	227508
17α,21β, 22R-trishomohopane	164773	164773	164602
17α,21β, 22S- tetrakishomohopane	144789	144789	147635
17α,21β, 22R-tetrakishomohopane	88195	88195	89300
17α,21β, 22S-pentakishomohopane	62909	62909	63264
17α,21β, 22R-pentakishomohopane	65046	53270	50516
13β , 17α, 20S - cholestane (diasterane)	79853	91181	83018
13β , 17α, 20R - cholestane (diasterane)	60463	60463	61903
5α, 14α, 17α, 20S- cholestane	119459	119459	114688
5α , 14β, 17β, 20R-cholestane	147850	147850	143927
5α , 14β, 17β, 20S-cholestane	110905	110905	109473
5α, 14α, 17α, 20R- cholestane	132957	145797	136652
24-methyl-5α, 14α, 17α, 20S- cholestane	81026	81026	72783
24-methyl-5α, 14β , 17β, 20R- cholestane	107669	107669	102237
24-methyl-5α, 14β , 17β, 20S- cholestane	91086	91086	96224
24-methyl-5α, 14α, 17α, 20R- cholestane	100518	100518	100429
24-ethyl-5α, 14α, 17α, 20S- cholestane	168245	168245	167137
24-ethyl-5α, 14β , 17β, 20R- cholestane	135452	135452	119182
24-ethyl-5α, 14β , 17β, 20S- cholestane	104786	103078	85758
24-ethyl-5α, 14α, 17α, 20R- cholestane	137873	137873	139255
Correlation coefficient with the revised result		**0.999185**	**0.999273**