

The Definition of the reliability of identification of complex organic compounds using HPLC and base chromatographic and spectral data

Yu. A. Kudiyarov, DCs, E.V. Kulyabina, PhD
Russian Research Institute for Metrological Service (VNIIMS)

Abstract

In this work a method for determining the reliability of the identification of complex organic compounds using HPLC is proposed. This method is based on the use of Student's statistical criterion. The advantage of this method is reliable identification with values of retention parameter and measuring information about the spectrum of substance without the use of traditional calibration.

Key words: HPLC, identification, reliability, Student criterion

Introduction

This article describes a method for identification and determination of complex organic compounds. More precisely, this particular method uses high-performance liquid chromatography with UV detection without the use of traditional calibration with standard samples of analyses [1]. It is hard to reliably identify complex organic compounds due to absence of standard samples. A method of analysis [2] and multi-component test mixture [1] was developed to solve these tasks and is summarized in this article.

Materials and Methods

To carry out this research, we used high performance liquid chromatograph "MiliChrome A-02" (manufactured by Institute of Chromatography "EcoNova", Novosibirsk), UV spectrophotometric detector which can record spectrum and detect substance at 8 wavelengths. The chromatograph has a chromatographic column with volume variability no more than ± 2 % from instrument to instrument and the column efficiency variability no more than ± 5 %.

The test mixture "DB-2012" was used to check the stability parameters of the chromatograph. The database of spectral and chromatographic data of 500 compounds (DB-2012) was used for the purpose of identification.

The method of identification and quantification of complex organic compounds is shown below.

1. Pass the multi-component mixture through the chromatograph (each component of the mixture is designed to test a particular parameter (s) of the chromatograph).

2. Check if the resulting chromatogram of the test mixture agrees with the certified values within the certified tolerances.

3. Pass the sample to be analyzed (analyte) through chromatograph, keeping all settings the same as in the previous step.
4. Initial identification of the analyte can be carried out in terms of retention parameter. Final identification of the analyte can be carried out by analyzing spectral ratios.
5. If a unique substance is identified, amount of substance in the sample can be determined.

To test of the statistical hypothesis, we used Student's criterion.

Results and Discussion

In this work we have adopted the following definition of identification: the substance is considered to be identified if there is a one-to-one match characteristics of the analyte with the characteristics of the substances from database.

The database itself is derived from chromatograms of various substances. In particular, from chromatograms we obtained chromatographic characteristics of substances - a retention time and volume, peak height and area, and the ultra-violet spectrum (and the spectral ratios) were recorded in the range from 190 nm to 400 nm.

For identification there we consider two types of data [3].

The first type of data are data from DB-2012, which are chromatographic characteristics and spectra of standard samples of pure substances. DB-2012 shows the average values of these characteristics $a_R(i, j)$, ($i = 1, 2, \dots, m$, $j = 1, 2, \dots, k$), the index i means index of the substance in the database, j - the index of characteristics. The characteristics of substances were retention volume and spectral ratios determined by the experimentally measured ultraviolet spectra. Each characteristics - volume retention and spectral relationships have values $\sigma_R(i, j)$ and corresponding tolerances $d(i, j)$ to their values.

Recommended range of tolerances [4], is given by

$$2\sigma_R(i, j) < d(i, j) < 3\sigma_R(i, j). \quad (1)$$

The second type of data is the experimentally obtained characteristics of substances $x(i)$, such as average values of retention volumes, spectral ratio $b_x(i, j)$ and the corresponding values of the mean square deviation $\sigma_x(i, j)$.

To use the statistical methods to identify substances, is necessary to determine the limits of applicability of the theory.

We assume that the measurements of substance parameters behave as random variables and their distribution is normal. However, characteristics of the analyte cannot be a random quantity, but the difference between the retention

times or the corresponding spectral ratios is a random variable and it can be applied to adopt methods of mathematical statistics.

The distance between measurement and reference in this case is the difference $\Delta_x(i, j)$ between the measured average value characteristics and the value of this feature in the database:

$$\Delta_x(i, j) = |a_R(i, j) - b_x(i, j)|. \quad (2)$$

Identification is considered effective if the difference between the measured value and characteristics listed in the database will be less or equal than some criterion (in our case - threshold for this difference), that is if the following condition is true

$$\Delta_x(i, j) \leq d(i, j). \quad (3)$$

Condition (3) is necessary but not sufficient for unambiguous identification. For implementation of identification, it is need to formulate a statistical hypothesis H_0 (null hypothesis) and H_1 (alternative hypothesis).

The null hypothesis H_0 is the statement that there is no substance in the sample.

The error of the 1st kind is made with a false positive result, then the hypothesis H_0 is rejected and the hypothesis H_1 accepted.

The probability of error of the 1st kind denoted α - probability of acceptance of alternative hypothesis or significance level. Alternative situation is adoption of hypothesis H_1 .

The error of the 2nd kind is made with a false negative results, then the hypothesis H_1 is rejected and the hypothesis H_0 accepted. The probability of error of the 2nd kind is indicated by β .

The accuracy of the identification, the function P associated with α and β [4] by ratio

$$P = 1 - \alpha - \beta. \quad (4)$$

Example

For example, suppose after performing an experiment, we obtained the following retention volumes

$$V(1,1) = 3300 \text{ mcl}, \quad V(2,1) = 3572 \text{ mcl}, \quad V(3,1) = 3306 \text{ mcl} \quad (5)$$

By comparing these values with the corresponding values of retention volumes in the database, we see that the first material can be identified as pyrene ($V_p(1,1) = 3301$), the second - as ionol ($V_R(2,1) = 3569$), the third - as isoamyl benzoate ($V_R(3,1) = 3304$).

The identification of the second substance is not complicated, but the identification of the first and third materials have problems as its can be identi-

fied as pyrene, and as isoamyl benzoate, wherein the error of the second kind is low. Therefore, it is not possible to uniquely identify these two substances from only the retention volumes.

The above results indicate that for the unambiguous identification we need to consider other properties of substances, for example, spectral ratios, in addition to the values of retention volumes.

To identify we use an average value of the spectral relations \bar{s} of a particular substance. Average value of spectral ratios \bar{s} substance is simply the arithmetic average of a set of values of the spectral relations.

We introduce the difference between the base and the measured spectral ratios [3]

$$\Delta(0, i, j) = s(0, i) - s(i, j) . \quad (6)$$

Substance is considered identified if the following condition is present

$$\Delta_{cp}(0, j) < d \quad \text{or} \quad \bar{s}(0) - \bar{s}(j) < d$$

where d - the tolerance on the average base spectral ratios:

$$\bar{s}(0) - d < \bar{s}(0) < \bar{s}(0) + d .$$

Hypothesis H_0 (null hypothesis) is the assertion that the considered substances in the sample is missing.

This means that in this case, the inequality

$$\Delta_{cp}(0, j) - d > 0 . \quad (7)$$

Hypothesis H_1 (alternative hypothesis) implies substance is present in the sample.

Condition for presence of substances is expressed by the inequality

$$\Delta_{cp}(0, j) - d < 0 . \quad (8)$$

We determine which hypothesis is realized using Student's criterion. Quantile of the Student distribution $t_{N-1, P}$ corresponding to probability $P = 0.95$ for $N - 1 = 6$ ($N = 7$) of degrees of freedom is equal to 2.4470. This means that it is the value of a variable $\Delta_{cp}(0, j) - d$ is implemented with probability $P = 0.95$.

If the experimental value of the quantile $t(0, j)$ is less than the value of the table, the condition

$$t(0, j) < t_{N-1, P} , \quad (9)$$

means that the corresponding hypothesis is rejected, the alternative - was adopted.

Taking into account the spectral ratios let us calculate the Student's coefficient for the example above.

For substances 2 and 3 the hypothesis is true (these substances in the sample are not present), and for the substance 1 this hypothesis should be rejected, that is it may be identified as pyrene. The level of significance (the 1st kind error or the probability of false identification) is to be equal to 0.05 and the corresponding probability of identification $P = 0.95$.

A similar result was obtained in the analysis of samples within the hypotheses .

Conclusions

Thus, it should be noted that the above results suggest that the use of the spectra of compounds and spectral ratios allow quantifying the reliability of identification. It is also possible to extend this method to determine the reliability of the identification of a set of characteristics jointly, as opposed to looking at only one characteristic.

Literature

1. E.V. Kulyabina. Development and study of methods of identification and quantification of complex organic compounds through the use of complex substances with standardized chromatographic and spectral parameters, Dissertation // FSUE "VNIIMS", Moscow, 2013. – P.140.
2. M.A. Grachev, G.I. Baram, I.N. Azarova. MVI "Mass concentration of UV-absorbing substances. Methods of measurement by HPLC." Limnological Institute of SO RAS. Irkutsk, 2003. Attestation Certificate № 37-03 dated 10.12.2003 number FR.1.31.2003.00950.
3. Y.A. Kudeyarov, E.V. Kulyabina, O.L. Rutenberg, Application of Student's criterion to determine the reliability of the identification of substances in chromatographic analysis, "Legislative and applied metrology", № 3, 2013 pp. 44-48.
4. V.I. Vershinin, B.G. Derendyaev, K.S. Lebedev. Computer identification of organic compounds. M: "Science". - 2002. - P.180.