

НАДЕЖНОЕ СОПОСТАВЛЕНИЕ МАСС-СПЕКТРОВ, ЗАРЕГИСТРИРОВАННЫХ В ИДЕНТИЧНЫХ УСЛОВИЯХ, С ПОМОЩЬЮ МЕТОДА ГЛАВНЫХ КОМПОНЕНТ

А.С. Самохин, И.А. Ревельский

Московский государственный университет имени М.В. Ломоносова, химический факультет, 119991 Москва, Ленинские горы, д. 1, стр. 3
andrey.s.samokhin@gmail.com

Поступила в редакцию 8 июня 2012 г

Рассмотрена возможность использования метода главных компонент для проведения надежного сопоставления масс-спектров, зарегистрированных в идентичных условиях и содержащих один и тот же набор значений m/z . Описанный в работе подход к сопоставлению масс-спектров основан на совместном использовании метода главных компонент и t -критерия Стьюдента. В качестве модельных соединений были рассмотрены алифатические спирты нормального строения (характеризующиеся весьма близкими масс-спектрами ионизации электронами).

Ключевые слова: сопоставление масс-спектров, базы данных масс-спектров, достоверная идентификация, метод главных компонент.

Самохин Андрей Сергеевич – аспирант Химического факультета Московского государственного университета им. М.В. Ломоносова.

Область научных интересов – хромато-масс-спектрометрия, хемометрика.

Опубликовано 10 печатных работ.

Ревельский Игорь Александрович – д.х.н., профессор Химического факультета Московского государственного университета им. М.В. Ломоносова.

Область научных интересов – хромато-масс-спектрометрия, анализ сложных многокомпонентных смесей, определение следовых количеств веществ.

Опубликовано более 400 печатных работ, в том числе 1 монография и 35 авторских свидетельств и патентов.

Введение

Метод газовой хроматографии/масс-спектрометрии (ГХ/МС) с ионизацией электронами (ИЭ) применяется для идентификации компонентов смесей органических соединений. Широкое использование метода обусловлено высокой чувствительностью и селективностью, а также наличием баз данных индексов удерживания и масс-спектров ИЭ (использование которых значительно упрощает идентификацию неизвестных органических соединений). В ряде случаев (особенно при наличии дополнительной информации об исследуемом образце) совместное использование баз данных масс-спектров ИЭ и индексов удерживания позволяет провести однозначную идентификацию неизвестного соединения. Однако в общем случае идентификация, основанная на использовании библиотечных (или расчетных в случае индексов удерживания) данных, носит предположительный характер, что объясняется рядом причин.

В современных базах данных масс-спектров ИЭ представлены спектры 200–600 тысяч органических соединений [1, 2] (число соединений, представленных в базах данных индексов удерживания,

– в несколько раз меньше). В свою очередь, число низкомолекулярных ($MM < 600$ Да) соединений, зарегистрированных в базе данных CAS (по данным на конец 2009 года), составляет около 35 млн [3].

Масс-спектры ИЭ зависят от условий их регистрации (тип и модель масс-спектрометра, метод калибровки прибора, температура источника ионов и др.).

Результаты поиска по базам данных масс-спектров ИЭ следует рассматривать как ориентировочные, поскольку не существует однозначной корреляции между величиной степени совпадения (или позицией в списке возможных кандидатов) и надежностью идентификации [4, 5].

Индексы удерживания, приведенные в базах данных для одного и того же вещества, полученные на одной и той же стационарной фазе, могут в значительной степени отличаться друг от друга [6]. Например, в базе данных NIST'11 [1] как для 2,4-диметилфенола, так и для 2,5-диметилфенола представлено около 10 экспериментальных значений индексов удерживания (полученных при использовании капиллярных колонок, покрытых фазой 5 % – фенил и 95 % – метилполисилоксан). Средние значения индексов удерживания 2,4- и

2,5-диметилфенола составляют 1152 и 1154, а стандартные отклонения – 5 и 7 соответственно. Таким образом, в данном случае однозначная идентификация, основанная на использовании только библиотечных данных, невозможна, если в исследуемом образце присутствует только один из указанных изомеров.

При идентификации неизвестных органических соединений присутствие пика молекулярного иона позволяет определить молекулярную массу (а в ряде случаев и брутто формулу), что значительно увеличивает надежность идентификации. Однако пик молекулярного иона отсутствует в масс-спектрах ИЭ около четверти органических соединений [7].

Для увеличения надежности идентификации неизвестное и предполагаемое соединения должны быть проанализированы в одних и тех же условиях, а полученные экспериментальные индексы (или времена) удерживания и масс-спектры ИЭ сопоставлены. Индексы (или времена) удерживания являются одномерными аналитическими сигналами, поэтому их сравнение не вызывает трудностей. В свою очередь, масс-спектры ИЭ – многомерны, и их сравнение является более сложной задачей. Из литературы известно несколько способов сравнения масс-спектров (зарегистрированных при одних и тех же условиях), однако все они имеют весьма серьезные ограничения:

- визуальное сравнение масс-спектров необъективно и не позволяет оценить «похожесть» масс-спектров количественно;
- сравнение масс-спектров, основанное на расчете величины степени совпадения между спектрами (с использованием алгоритмов, применяемых в программах библиотечного поиска), позволяет найти различия между визуально одинаковыми масс-спектрами [8]. Однако для установления критического (обеспечивающего надежную идентификацию) значения величины степени совпадения необходимо проведение большого числа дополнительных экспериментов. Кроме того, критическое значение величины степени совпадения может отличаться для разных классов органических соединений и изменяться во времени (например, в результате загрязнения источника ионов);
- сравнение интенсивностей отдельных масс-спектральных пиков используется при проведении идентификации предполагаемых соединений [9–11]. Однако только несколько значений m/z (обычно три) принимаются во внимание. Следует также отметить, что сопоставление интенсивностей только небольшого числа масс-спектральных пиков увеличивает вероятность ложноположительной идентификации [12] по сравнению с рассмотрением полных спектров.

В ряде работ была показана возможность использования метода главных компонент для дифференциации органических соединений,

обладающих практически одинаковыми масс-спектрами (содержащими один и тот же набор фрагментных ионов). Авторами были рассмотрены как масс-спектры ИЭ [13–16], так и другие виды масс-спектров [17–20]. В принципе методы классификации с обучением (такие, например, как линейный дискриминантный анализ или метод SIMCA) могут быть использованы, чтобы сделать обоснованный вывод, является ли неизвестное соединение предполагаемым кандидатом или нет [13, 17, 19]. Однако для построения и проверки классификационной модели необходимо предварительно зарегистрировать десятки масс-спектров. По этой причине методы классификации с обучением не подходят для проведения быстрого сопоставления масс-спектров неизвестного соединения и предполагаемого кандидата. Ранее нами был предложен подход к сравнению масс-спектров, зарегистрированных при одних и тех же условиях, основанный на совместном использовании метода главных компонент и t -критерия Стьюдента [21]. Целью настоящей работы являлось изучение возможности использования данного подхода для дифференциации алифатических спиртов нормального строения, основанной на попарном сравнении соответствующих масс-спектров ИЭ (зарегистрированных при одних и тех же условиях).

Экспериментальная часть

В работе использовали раствор алифатических спиртов нормального строения (1-октанол, 1-нонанол, 1-деканол, 1-ундеканол, 1-додеканол, 1-тетрадеканол (PolyScience Corporation)) в *n*-гексане. Концентрация аналитов составляла 3 нг•мкл⁻¹. Масс-спектры ИЭ были зарегистрированы при использовании времяпролетного масс-спектрометра Pegasus HT (Leco), соединенного с газовым хроматографом Agilent 7890A (Agilent Technologies). Ввод пробы объемом 1 мкл осуществляли в режиме с делением потока 1:10. Температура инжектора хроматографа составляла 250 °С. Разделение модельных соединений осуществляли на капиллярной колонке Restek RTX-5 (10 м × 0.1 мм × 0.1 мкм). Температуру термостата колонок изменяли по следующей программе: изотерма 50 °С в течение 3 мин; нагрев со скоростью 50 °С•мин⁻¹ до 250 °С; изотерма 250 °С в течение 3 мин. В качестве газа носителя использовали гелий марки 60 (0.5 мл•мин⁻¹). Температура источника ионов составляла 250 °С. Энергия ионизирующих электронов 70 эВ. Диапазон сканируемых значений m/z составлял 50–250 при скорости сканирования 20 скан•с⁻¹. Выделение «чистых» масс-спектров проводили с помощью программного обеспечения ChromaTOF 4.32. Обработку полученных данных проводили с помощью программного обеспечения Microsoft Excel 2007 и надстройки Chemometrics Add-In [22].

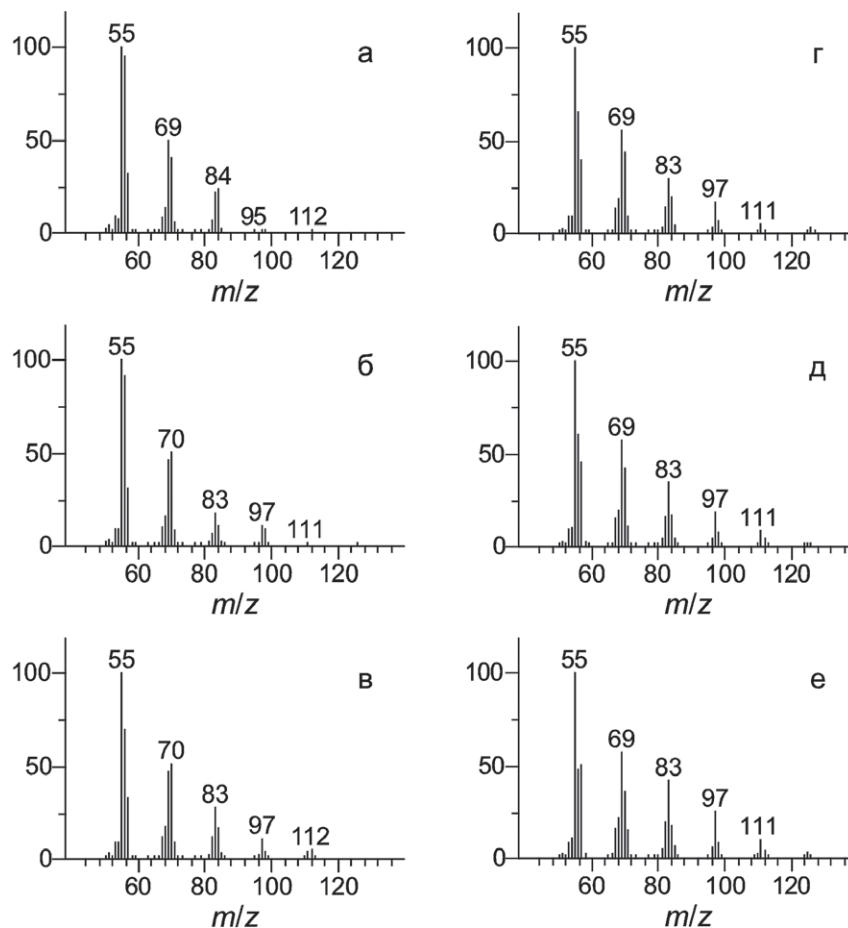


Рис. 1. Усредненные ($n = 5$) масс-спектры ИЭ 1-октанола (а), 1-нонанола (б), 1-деканола (в), 1-ундеканола (г), 1-додеканола (д) и 1-тетрадеканола (е). В хроматографическую колонку было введено около 0.3 нг аналитов

Обсуждение результатов

Гомологи алифатических спиртов нормального строения были выбраны в качестве модельных соединений, поскольку их масс-спектры ИЭ весьма похожи (рис. 1) и содержат практически идентичный набор масс-спектральных пиков. Разделение хроматографических пиков аналитов до базовой линии позволило зарегистрировать масс-спектры ИЭ всех исследуемых соединений за один анализ. Модельная смесь была проанализирована 5 раз при одних и тех же экспериментальных условиях. С помощью программного обеспечения ChromaTOF для каждого компонента были выделены 5 масс-спектров ИЭ низкого разрешения (каждому хроматографическому пику соответствовал только один «очищенный» масс-спектр). Из масс-спектров вручную были исключены малоинтенсивные значения m/z , для которых соотношение S/N для хроматографического пика на соответствующей масс-хроматограмме было меньше 5. В результате при сопоставлении масс-спектров нами были рассмотрены только 13 масс-спектральных пиков ($m/z = 53-57, 67-71$ и $82-84$), присутствующих в масс-спектрах ИЭ каждого из аналитов. Масс-спектры ИЭ алифатических спиртов были сопоставлены

попарно. Общее число рассмотренных пар соединений составило $\frac{6!}{2! \cdot (6-2)!} = 15$.

Масс-спектры каждой пары соединений были нормированы на полный ионный ток и представлены в матричном виде (строки соответствовали масс-спектрам, а столбцы – значениям m/z). Каждая из 15 матриц была центрирована и проанализирована с помощью метода главных компонент [23–25]. В качестве примера на рис. 2 представлен график счетов, полученный при сопоставлении масс-спектров ИЭ 1-ундеканола и 1-додеканола. На

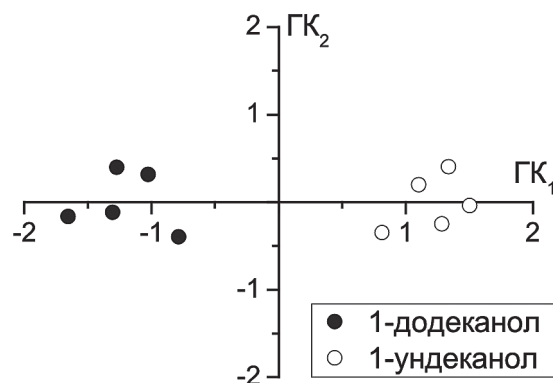


Рис. 2. График счетов. Каждая точка представляет собой масс-спектр. ГК₁ и ГК₂ – оси первой и второй главных компонент

плоскости главных компонент 1 и 2 можно выделить две группы точек, которые отвечают масс-спектрам 1-ундеканолола и 1-додеканолола соответственно. Первая главная компонента разделяет объекты, относящиеся к различным группам, и объясняет 88 % исходной вариации. Вторая и более высокие компоненты содержат только шум и могут не рассматриваться в дальнейшем (вторая главная компонента объясняет менее 5 % исходной вариации, третья – менее 3 %).

Краткое описание предложенного нами ранее подхода [21] к сопоставлению масс-спектров органических соединений (зарегистрированных при одних и тех же условиях) представлено ниже. Первая главная компонента представляет собой линейную комбинацию исходных переменных (а именно значений интенсивностей отдельных масс-спектральных пиков). Обычно полагают, что интенсивность масс-спектральных пиков имеет нормальное распределение. В этом случае первая главная компонента также имеет нормальное распределение (поскольку линейная комбинация независимых случайных величин распределена нормально). В результате для сравнения масс-спектров (отвечающих, например, предполагаемому кандидату и неизвестному соединению) может быть использован *t*-критерий Стьюдента.

Средние значения и стандартные отклонения первой главной компоненты были рассчитаны отдельно для масс-спектров ИЭ, относящихся к разным соединениям. Для сравнения средних значений использовали *t*-критерий Стьюдента. Рассчитанные значения ξ представлены в табл. 1. Пара соединений 1-ундеканол и 1-додеканол характеризовалась наименьшим значением $\xi = 13$, однако даже в этом случае оно превышало табличное значение $t_{крит.} (P = 0.999; f = 8) = 5$.

Таблица 1

Значения $\xi (n_1 = n_2 = 5)$, рассчитанные для каждой пары соединений

	1-октанол	1-нонанол	1-деканол	1-ундеканол	1-додеканол	1-тетрадеканол
1-октанол	–	72	45	64	71	124
1-нонанол	72	–	30	44	50	78
1-деканол	45	30	–	26	28	68
1-ундеканол	64	44	26	–	13	40
1-додеканол	71	50	28	13	–	23
1-тетрадеканол	124	78	68	40	23	–

Набор наиболее информативных значений *m/z* (вносящих наибольший вклад в дифференциацию спиртов) отличался в зависимости от рассматриваемой пары соединений. Например, для пары соединений 1-октанол/1-нонанол различие между масс-спектрами в основном было обусловлено значениями *m/z* 70 и 84; для пары соединений 1-ундеканол/1-додеканол – значениями *m/z* 56, 57 и 83.

Описанный в работе подход к сопоставлению масс-спектров ИЭ может быть использован для увеличения надежности идентификации неизвестных соединений, когда заключающим этапом идентификации является анализ неизвестного и предполагаемого соединений методом ГХ/МС в идентичных экспериментальных условиях [21]. Одним из основных вопросов, возникающих при использовании данного подхода, является выбор оптимального числа параллельных экспериментов. С точки зрения цены анализа предпочтительным является сокращение числа экспериментов, однако это неизбежно приведет к увеличению коэффициента Стьюдента и, в общем случае, к увеличению числа ложноположительных ответов.

Для двух спиртов (1-ундеканолола и 1-додеканолола), характеризующихся наиболее похожими масс-спектрами, нами была рассмотрена возможность получения ложноположительных ответов в случае сокращения числа параллельных экспериментов. Из исходного набора данных (содержащего по пять масс-спектров каждого из двух соединений) нами были сформированы всевозможные поднаборы, содержащие по два или по три параллельных масс-спектра ИЭ каждого соединения (общее число таких комбинаций как для двух, так и для трех параллельных спектров составило $\left(\frac{5!}{2! \cdot (5-2)!}\right)^2 = \left(\frac{5!}{3! \cdot (5-3)!}\right)^2 = 100$). Как и ранее, масс-спектры были нормированы, представлены в матричном виде и проанализированы с помощью метода главных компонент. Средние значения первой главной компоненты (отвечающие различным соединениям) были сопоставлены с помощью *t*-критерия Стьюдента. Полученные ре-

Таблица 2

Процент ложноположительных ответов* в зависимости от значения доверительной вероятности (*P*) и числа параллельных опытов (*n*)

<i>P</i>	Процент ложноположительных ответов*, %	
	<i>n</i> = 2	<i>n</i> = 3
0.95	1	0
0.99	56	0
0.999	98	18

Примечание: * – процент случаев, для которых статистически значимое различие между масс-спектрами ИЭ 1-ундеканолола и 1-додеканолола не было найдено (т.е. $\xi \leq t_{крит.} (P, f = 2n - 2)$).

зультаты обобщены в табл. 2. Как видно из табл. 2, при проведении трех параллельных опытов для доверительной вероятности $P = 0.99$ ни в одном из 100 случаев не было получено ни одного ложноположительного ответа (то есть статистически значимое различие между масс-спектрами 1-ундеканола и 1-додеканола было найдено во всех случаях). Однако для $n = 2$ и $P = 0.99$ число ложноположительных ответов превышало 50 %. В связи с этим при проведении идентификации целесообразно регистрировать не менее трех параллельных масс-спектров ИЭ неизвестного соединения и предполагаемого кандидата. Доверительная вероятность должна выбираться исследователем, исходя из того, что ее величина определяет критическое значение коэффициента Стьюдента ($t_{\text{крит.}}$) и напрямую влияет на число ошибок первого и второго рода.

Заключение

Рассмотрена возможность совместного использования метода главных компонент и t -критерия Стьюдента для проведения надежного сопоставления масс-спектров, зарегистрированных в идентичных условиях. Описанный в работе подход может быть использован для проведения сопоставления масс-спектров неизвестного соединения и возможного кандидата (зарегистрированных в идентичных условиях) и тем самым для увеличения надежности идентификации.

Авторы выражают благодарность компании Leco за предоставление времяпрелетного масс-спектрометра Pegasus HT (Leco) для проведения исследований.

ЛИТЕРАТУРА

1. NIST/EPA/NIH Mass Spectral Library, National Institute of Standards and Technology, 2011.
2. Wiley Registry of Mass Spectral Data, 2009, 9th electronic ed.
3. Little J.L., Cleven C.D., Brown S.D. Identification of "known unknowns" utilizing accurate mass data and chemical abstracts service databases // J. Am. Soc. Mass Spectrom. 2011. V. 22, № 2. P. 348–359.
4. Stein S.E.; Scott D.R. Optimization and testing of mass spectral library search algorithms for compound identification // J. Am. Soc. Mass Spectrom. 1994. V. 5, № 9. P. 859–866.
5. Возможность достоверной идентификации неизвестных соединений при использовании программы MS Search и коммерческой базы данных масс-спектров электронной ионизации / А.С. Самохин и [др.] // Масс-спектрометрия. 2011. Т. 8, № 1. С. 65–67.
6. Зенкевич И.Г., Уколова Е.С. Об основных факторах, влияющих на межлабораторную воспроизводимость газохроматографических индек-

- сов удерживания // Аналитика и контроль. 2010. Т. 14, № 4. С. 243–250.
7. Самохин А.С., Ревельский И.А. Интенсивность пика молекулярного иона в масс-спектрах электронной ионизации // Масс-спектрометрия. 2012. Т. 9, № 1. С. 58–60.
8. Самохин А.С., Ревельский И.А., Ревельский А.И. Увеличение достоверности идентификации неизвестных соединений при использовании баз данных масс-спектров электронной ионизации // Тезисы докладов Четвертой Всероссийской конференции-школы «Фундаментальные вопросы масс-спектрометрии и ее аналитические применения». Звенигород, 2010. С. 58–59.
9. Commission Decision 2002/657/EC of 12 August 2002 // Off. J. Eur. Commun. 2002. V. L221. P. 8–36.
10. Guidance for industry: mass spectrometry for confirmation of the identity of animal drug residues: final guidance. FDA, Washington, 2003.
11. WADA Technical Document – TD2010IDCR, 2011.
12. Stein S.E., Heller D.N. On the risk of false positive identification using multiple ion monitoring in qualitative mass spectrometry: large-scale inter-comparisons with a comprehensive mass spectral library // J. Am. Soc. Mass Spectrom. 2006. V. 17, № 6. P. 823–835.
13. Leth T. Chemometric analysis of mass spectra of cis and trans fatty acid picolinyl esters // Z. Lebensm.-Unters. Forsch. 1997. V. 205, № 2. P. 111–115.
14. Mjos S.A., Pettersen J. Determination of trans double bonds in polyunsaturated fatty acid methyl esters from their electron impact mass spectra // Eur. J. Lipid Sci. Technol. 2003. V. 105, № 3–4. P. 156–164.
15. Discrimination among geometrical isomers of α -linolenic acid methyl ester using low energy electron ionization mass spectrometry / L. Hejazi et [al.] // J. Am. Soc. Mass Spectrom. 2009. V. 20, № 7. P. 1272–1280.
16. Samokhin A.S., Revelsky I.A. Distinguishing by principal component analysis *o*-xylene, *m*-xylene, *p*-xylene and ethylbenzene using electron ionization mass spectrometry // Eur. J. Mass Spectrom. 2011. V. 17, № 5. P. 477–480.
17. Differentiation of betamethasone and dexamethasone using liquid chromatography/positive electrospray tandem mass spectrometry and multivariate statistical analysis / J.P. Antignac et [al.] // J. Mass Spectrom. 2002. V. 37, № 1. P. 69–75.
18. High energy collision-induced dissociation (CID) product ion spectra of isomeric polyhydroxy sugars / J.H. Scrivens et [al.] // Int. J. Mass Spectrom. 2003. V. 230, № 2-3. P. 201–208.
19. Distinguishing monosaccharide stereo- and structural isomers with TOF-SIMS and multivariate statistical analysis / E.S.F. Berman et [al.] // Anal. Chem. 2006. V. 78, № 18. P. 6497–6503.

20. Vaezian B., Anderton C.R., Kraft M.L. Discriminating and imaging different phosphatidylcholine species within phase-separated model membranes by principal component analysis of TOF-secondary ion mass spectrometry images // *Anal. Chem.* 2010. V. 82, № 24. P. 10006–10014.
21. Samokhin A.S., Revelsky I.A. Reliable comparison of electron ionization mass spectra recorded under identical experimental conditions // *Spectroscopy Letters* (submitted for publication).
22. [Электронный ресурс]: <http://rcs.chph.ras.ru/Tutorials/projection.htm> (дата обращения: 30.05.2012).
23. Jolliffe I.T. *Principal component analysis*. Second edition. New York: SpringerVerlag, 2002. 502 p.
24. Jackson J.E.. *A user's guide to principal component*. New York: John Willey, 1991. 567 p.
25. Эсбенсен К. Анализ многомерных данных: пер. с англ. С.В. Кучерявского; под ред. О.Е. Родионова. Черноголовка: ИПХФ РАН, 2005. 160 с.

RELIABLE COMPARISON OF MASS SPECTRA RECORDED UNDER IDENTICAL CONDITIONS USING PRINCIPAL COMPONENT ANALYSIS

A.S. Samokhin, I.A. Revelsky

*Lomonosov Moscow State University, Chemistry Department
Leninskie Gory 1-3, 119991 Moscow, Russia
andrey.s.samokhin@gmail.com*

In this work, we considered the possibility of using principal component analysis for the reliable comparison of mass spectra recorded under identical conditions and containing the same set of m/z values. The described approach for mass spectra comparison was based on simultaneous use of principal component analysis and Student's t -test. Normal aliphatic alcohols (having very similar electron ionization mass spectra) were used as model compounds.

Key words: mass spectra comparison, mass spectral data bases, reliable identification, principal component analysis.