

熵最小算法解析玫瑰精油 GC/MS 混合谱中的未知组分

林光川¹, 龙海荣², 张华俊^{1,3}, 冀晓雯², 卢波³, 谷筱玉²

(1. 苏州云溪分析技术有限公司, 江苏 苏州 215123;

2. 广西药用植物园, 西南濒危药材资源开发国家工程实验室, 广西 南宁 530023;

3. 广西科学院, 国家非粮生物质能源工程技术研究中心, 广西 南宁 530002)

摘要:熵最小算法是一种化学计量学算法,它根据实验数据本身就可以从混合谱中重建出纯谱。本研究采用熵最小算法对玫瑰精油 GC/MS 图谱中的 2 个共流出峰进行分析,从这些共流出峰中重建出 8 个纯谱和它们的总离子流浓度,并通过 NIST 数据库比对定性了其中 4 个组分。熵最小算法能够快速分析共流出峰中的组分,提高天然产物的分析效率,可应用于天然产物、中药、食品等体系的快速分析。

关键词:气相色谱-质谱法(GC/MS);熵最小算法;重建纯谱;定性分析

中图分类号:O657.63

文献标志码:A

文章编号:1004-2997(2015)05-0448-06

doi:10.7538/zpxb.youxian.2015.0027

Analysis of Unknown Components in Rose Essential Oil by Entropy Minimization Algorithm

LIN Guang-chuan¹, LONG Hai-rong², ZHANG Hua-jun^{1,3},

JI Xiao-wen², LU Bo³, GU Xiao-yu²

(1. Suzhou ChemoPower Technology Co., Ltd., Suzhou 215123, China;

2. National Engineering Laboratory of Southwest Endangered Medicinal Resources Development,
Guangxi Botanical Garden of Medicinal Plants, Nanning 530023, China;

3. National Engineering Research Center for Non-food Biorefinery,
Guangxi Academy of Science, Nanning 530002, China)

Abstract: Entropy minimization algorithm (EM) is a chemometrics method, which reconstructs pure spectra out from mixture spectra based on experimental data only, and no prior information needed. A rose essential oil was analyzed by GC/MS, two co-eluted mixture spectra were analyzed by EM, and unknown pure spectra were estimated and identified. The result shows that 8 pure spectra and their TIC concentration are

收稿日期:2014-11-11;修回日期:2015-02-02

基金项目:广西科学研究与技术开发计划项目(桂科合 1298014-15)资助

作者简介:林光川(1989—),女(壮族),广西南宁人,硕士研究生,从事 GC/MS 应用研究。E-mail: lin_gc@chemopower.com

通信作者:张华俊(1972—),男(汉族),江苏宜兴人,研究员,从事化学计量方法研究。E-mail: george@chemopower.com

谷筱玉(1977—),女(回族),安徽凤阳人,高级工程师,从事中药材质量标准研究。E-mail: 13597342807@163.com

网络出版时间:2015-06-02;网络出版地址: <http://www.cnki.net/kcms/detail/11.2979.TH.20150602.0851.012.html>

estimated, and 4 of these pure spectra are identified by NIST library match in these two co-eluted peaks. EM algorithm can analyze unknown components inside co-eluted peaks quickly, enhance the efficiency of nature product analyses, which is applicable in quick analyses of nature product, traditional Chinese medicine and food etc.

Key words: gas chromatography-mass spectrometry (GC/MS); entropy minimization; spectral reconstruction; qualitative analysis

气相色谱-质谱(GC/MS)联用法常用于分离和鉴定天然产物中的挥发性成分^[1],具有快速、高效、灵敏度高等优点。GC/MS 技术首先利用色谱进行组分分离,然后利用质谱检测组分得到的纯谱信息,最后将纯谱与数据库比对得到成分信息。虽然现代色谱的分离效率很高,但由于天然产物的成分复杂,仪器难以将其完全分离,导致 GC 谱图中出现多种组分的共流出峰。为了分离这些共流出峰,通常的方法是更改 GC 程序升温条件,或更换其他类型的色谱柱。但是,由于最佳分离条件难以估计,并非所有的混合峰都能通过常规的方法分开,且分离过程对仪器和人员的要求均较高。因此,在天然产物分析过程中,快速、高效、全成分的分析方法已成为研究的热点和难点^[2]。

近年来,数学算法在化学分析领域中扮演着重要的角色^[3],它能解决许多实验不能解决的问题。熵最小算法^[4](entropy minimization, EM)是一种化学计量学算法,主要包括锁峰熵最小方法(BTEM)^[4]、加权多重锁峰熵最小方法(*t*BTEM)^[5]和多重重建熵最小方法(MREM)^[6]等。使用熵最小算法解析 GC/MS 共流出谱图,可以在没有先验信息的情况下,根据混合谱数据本身,从混合谱中重建出每个纯组分的谱图,并通过峰面积归一化法进行半定量分析。该算法已应用于质谱^[7]、红外^[8]、紫外^[9]和拉曼^[10]等谱图的解析。

目前,应用于共流出谱图解析的化学计量学方法已有报道^[11-13],这些方法一般需要一些先验信息协助解析共流出谱图,而且对共流出的各个组分的流出峰型有一定的要求,所以对完全未知体系(黑色体系)的应用有一定的难度,对实验设定的要求也较高。现阶段,用于共流出谱图解析的方法有:直观推导式演进特征投影法(HELP),该方法利用各种算法和多步骤,将黑色体系转化成灰色体系,直至白色体

系;熵最小算法,利用统计学的熵和优化方法搜索代表化合物信息向量的最优值,直接将黑色体系变成白色体系,简化了解析共流出谱图的过程。实际使用中,熵最小算法直接利用实验数组中的离子碎片峰信息进行定性分析,再利用重建出的峰面积进行定量分析。

通过仪器分析和数学算法相结合的熵最小算法,不仅可用于天然产物的分析,还可用于中药、食品以及其他复杂混合组分的快速分离、分析和鉴定过程。该方法不仅在一定程度上简化了样品的前处理过程,降低了分析成本,还具有通用性强的优点。

本研究将基于熵最小算法的 Matlab 程序,从玫瑰精油 GC/MS 的原始谱图中选取 2 个混合峰区间,从中重建出纯谱,再通过比对 NIST 数据库,得到部分组分的定性信息。

1 实验部分

1.1 主要仪器及材料

7820A-5977E 气相色谱-质谱联用仪;美国 Agilent 公司产品,配有电子电离源(EI)。

玫瑰精油:甘肃皓思玫瑰制品有限公司产品,产地为甘肃兰州市永登县苦水镇,样品不经稀释,直接进样分析。

1.2 实验条件

1.2.1 色谱条件 HP-5MS 石英毛细管柱(30 m×0.25 mm×0.25 μm);升温程序:柱温 40 °C 保持 1 min,以 5 °C/min 升至 280 °C,保持 10 min;载气(氮气)流速:1 mL/min;进样口温度 260 °C;分流比为 100:1;进样量 0.2 μL。

1.2.2 质谱条件 电子轰击(EI)离子源,电子能量 70 eV,采集模式为全扫描,扫描区间为 50~250 u,离子源温度 230 °C,四极杆温度 150 °C。

1.3 数据处理

样品经 GC/MS 分析后,用质谱工作站软件将数据导出成 CDF 格式,然后导入 Matlab

程序,截取其中色谱保留时间1 687~1 697 s (A 区间)以及 1 698~1 712 s (B 区间)的共流出峰,进行熵最小计算,得到各个重建纯谱。将重建纯谱数据导出成 msp 格式,再导入 NIST 数据库进行比对,得到相应的定性结果。

2 结果与讨论

2.1 熵最小算法简介

当某个样品经 GC/MS 分析后,得到对应的实验数据。对于数据中某一混合峰区间,其数字信息经导出后得到一套混合谱数据 $A_{k \times v}$, 其中 k 为仪器的采样次数, v 为数据通道。对 GC/MS 数据而言, k 为不同保留时间的采样点数目, v 为不同质荷比(m/z)的采样点数目。在这样的共流出谱图数据上,熵最小算法的计算步骤如下:

1) 首先对 $A_{k \times v}$ 进行奇异值分解(SVD),将奇异值矩阵 $V_{v \times v}^T$ 中无物理意义的部分、矩阵 U 和对角矩阵 $S_{k \times v}$ 全零的部分去除,得到修正后的矩阵 $V_{k \times v}^T$ 和 $S_{k \times k}$ 。

$$A_{k \times v} = U_{k \times k} \times S_{k \times v} \times V_{v \times v}^T \quad (1)$$

2) 观察矩阵 $V_{k \times v}^T$ 中的每个行向量,如果认为第 j 行以后的向量中数据为噪声,则将 $(j+1)$ 到 k 行的向量全部去掉,得到矩阵 $V_{j \times v}^T$ 和 $S_{j \times j}$ 。

3) 观察矩阵 $V_{j \times v}^T$ 中的每个行向量,找出感兴趣的碎片信息峰或者区间进行运算。

4) 优化方法将首先生成一个随机向量 $T_{1 \times j}$ 。通过方程(2),计算得到一个与纯谱有关的向量 a^{est} 。在优化过程中,将不断地产生新的 $T_{1 \times j}$,直到优化结束,最终的 a^{est} 即为目标纯谱。

$$a_{1 \times v}^{\text{est}} = T_{1 \times j} \times (S_{j \times j} \times V_{j \times v}^T) \quad (2)$$

5) 用锁定的峰或区间中的最高峰按方程(3)将 a^{est} 单位化,式中 a' 和 a'' 是锁定的峰或区间,这一步叫做锁峰。BTEM 方法锁住 1 个峰或区间,而 t BTEM 方法锁住 2 个峰。

$$\hat{a}_{v \times 1} = \frac{a_{1 \times v}^{\text{est}}}{\max(a') + \max(a'')} \quad (3)$$

6) 接着,按方程(4)得到每个重建纯谱的伪重建浓度信息,然后按照方程(5)构建优化方法的目标函数和罚函数。

$$C_{k \times 1} = A_{k \times v} \times \hat{a}_{v \times 1}^T \times (\hat{a}_{v \times 1} \times \hat{a}_{v \times 1}^T)^{-1} \quad (4)$$

$$\min(G) = f(\hat{a}_{v \times 1}) + P(\hat{a}_{v \times 1}, C_{v \times 1}) \quad (5)$$

7) 检测目标函数的值,如果达到了优化终点,则输出 a^{est} 和 $C_{k \times 1}$, 其中, a^{est} 为重建纯谱, $C_{k \times 1}$ 为重建伪浓度;如果没有达到优化终点,将由优化方法重新生成一个 $T_{1 \times j}$, 返回到步骤 4)~7)重新计算,直至优化结束。

8) BTEM 和 t BTEM 每次计算得到一个重建纯谱,对不同的峰或区间进行锁峰计算,有时会得到相同的重建纯谱。当共流出峰里的主要成分都重建后,通过对这些重建纯谱和原始数据进行计算,得到各重建纯谱相对应的重建浓度,示于方程(6),其中 n 是共流出中的主要成分数。

$$C_{k \times n} = A_{k \times v} \times \hat{a}_{v \times n}^T \times (\hat{a}_{v \times n} \times \hat{a}_{v \times n}^T)^{-1} \quad (6)$$

将重建纯谱的质谱数据与数据库(例如 NIST)进行比对,可得到该重建纯谱所代表的物质结构信息。

2.2 利用熵最小算法对玫瑰精油 GC/MS 谱图进行分析

玫瑰精油样品的 GC/MS 总离子流图示于图 1, 其中, 图 1 中的小图为保留时间在 1 624~

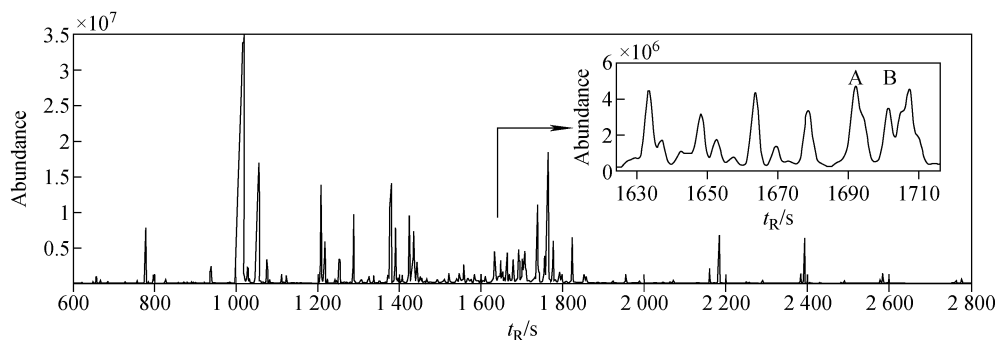


图 1 玫瑰精油样品的总离子流图

Fig. 1 TIC of rose essential oil

1 716 s 的放大图。

从图 1 中可以看到很多次要成分的小峰,其中,保留时间在 1 624~1 716 s 区间的共流出谱峰为由两种或两种以上的组分混合而成,这些峰的峰形不对称、相互交错重叠,且比体系中其他主要组分的丰度低。对该区间的各峰使用 NIST 数据库检索时,由于这些峰都是共流出峰,检索结果的可信度很低,故无法获得此区间内的组分信息。

利用熵最小算法对选择的 2 个共流出峰区间 1 687~1 697 s(区间 A)和 1 698~1 712 s(区间 B)进行分析,然后利用 Matlab 程序导出这 2 个区间的数据。在 A 区间,仪器在不同保留时间上共采样 33 次,采样范围为 m/z 50~250,采样间隔为 m/z 1,共有 201 个数据通道,得到了混合谱图数据 $A_{33 \times 201}$;类似的,在 B 区间,导出得到的混合谱数据为 $B_{49 \times 201}$ 。通过观察 V^T ,在 A 和 B 区间的运算中使用了 5 个 V^T 向量(步骤 2)。

通过观察 V^T 中的峰,锁住不同的质谱碎片峰(步骤 3),并进行多次运算得到多个重建纯谱,将其中相同的重建纯谱去除后,

在每个区间都发现了 4 个纯组分,按步骤 8 计算它们所对应的重建浓度。区间 A 和 B 的实验 TIC 浓度和对应的重建 TIC 浓度示于图 2。

从图 2 可以看出,使用熵最小算法对共流出混合谱图进行解析后,从实验得到的共流出峰中,A 和 B 区间都重建出 4 个纯谱以及相应的 4 个重建浓度。将各个重建浓度加和,得到的总重建浓度与实验得到的 TIC 浓度峰面积比分别为 95.8%(区间 A)和 99.4%(区间 B),且峰型一致。图 2 中每个重建浓度的峰宽都在 5 s 左右,这与同一实验谱图中丰度相近的纯峰的峰宽一致。由于总重建 TIC 浓度与实验 TIC 浓度非常一致,利用其中的每个重建纯谱的 TIC 面积对共流出峰中的组分进行定量分析就成为可能。因此,可利用现行的定量方法^[14](标准物浓度标准曲线法),按照计算出的重建浓度的面积,利用标准品对其中的各个重建纯组分进行定量分析。

将重建出的 8 个纯谱数据与 NIST 数据库比对,其中 4 个纯谱的比对结果较好,分别为重建纯谱 A1、A3 和 B2、B3,其结果列于表 1。

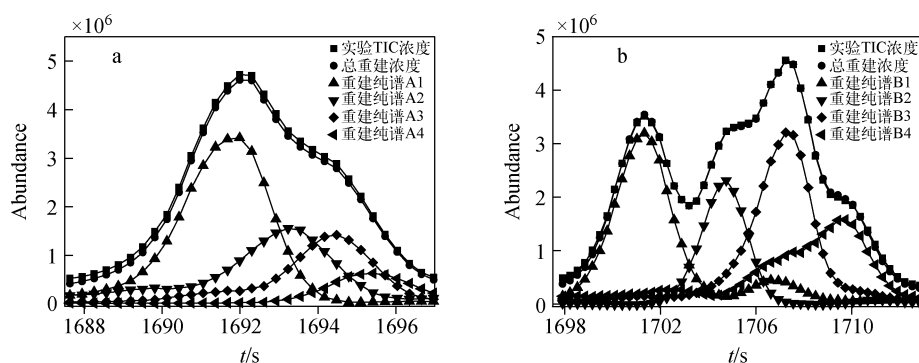


图 2 区间 A(a)和区间 B(b)的重建 TIC 浓度与实验 TIC 浓度对比

Fig. 2 Comparison of estimated TIC and experimental TIC with concentration of A (a) and B (b) area

表 1 区间 A 和 B 的重建纯谱和 NIST 比对结果

Table 1 NIST match results of estimated spectra from A and B area

重建纯谱序号	保留时间/s	化合物名称	保留指数/i. u.	分子式	NIST(R. Match)
A1	1692.0	α -Bisabolol	1665(估值 1621)	$C_{15}H_{26}O$	916
A3	1694.5	(Z,E)-Farnesol	估值 1661	$C_{15}H_{28}O$	833
B2	1704.8	2-Pentadecanone	1682	$C_{15}H_{30}O$	827
B3	1707.2	trans-Farnesol	1690	$C_{15}H_{26}O$	827

区间 A 和 B 中的其他 4 个组分(A2, A4 和 B1, B4)与 NIST 数据库中纯谱进行峰对峰比对(peak-to-peak),仍无法将这 4 个纯组分定性,所以将它们归为已知的未知物。从图 2 中 B 区间重建浓度可以看出,组分 B1 的分离度较好,与其他组分重合度低,直接比照 NIST 数据库即可得到定性结果。但使用质谱工作站软件在 B1 和 NIST 数据库比较,并没有找到匹配度较好的结果。这是因为使用常规方法对天然产物进行分析时,体系中存在大量的未知组分,基本上只能对丰度较高的组分进行定性,而类似于图 1 中的低丰度混合峰中的组分则难以定性。因此,这些低丰度组分很可能没有收录在 NIST 数据库中,也就难以得到定性结果。虽然这 4 个未知物未能定性,但是将它们的信息导入 NIST 数据库,可以方便以后的分析研究。值得一提的是,比对 NIST 数据库,并不仅仅以 NIST 数据库给出的匹配度作为唯一标准,还需要用人工观察碎片信息峰的匹配度来定性重建纯谱。

利用质谱工作站软件在 A 区间峰形最高点(1 692 s)直接比对,和与 NIST 数据库进行比对的定性结果均为 α -Bisabolol,但匹配度仅为 27.9%,该结果与重建纯谱 A1 的比对结果一致。这是因为在 A 区间中,该组分的浓度相对较高,受其他组分浓度的干扰较小,信噪比较高,可以利用峰形最高点处的质谱比对 NIST 数据库;但是,由于其他组分质谱碎片的影响,导致匹配度较低。因为无法确认共流出峰的组成,常规的分析方法则无法确认 α -Bisabolol 这样的组分。但是,经过熵最小算法的重建,在 A 区间中,由于去掉了其他组分的干扰,重建纯谱 A1 的比对结果达到了 83.3%,大大高于解谱前的比对结果(27.9%);通过算法分析,可以确认 α -Bisabolol 在共流出峰中的分布,进而确认该成分的存在;进一步利用该算法,能够确认原来无法确认的共流出峰中的其他组分。

在 B 区间中,B1 和 B4 组分重建浓度所对应的峰型为不对称峰型,且 B1 的重建浓度为肩峰,与理论情况不符。通过观察 B1 和 B4 的质谱峰,发现它们的质谱碎片峰 m/z 的位置大多是重叠的,也就是说,B1 和 B4 组分的分子组成中相同的基团很多,甚至可能为同分异构

体或同系物。对于峰位置高度重叠的体系,很难应用化学计量法进行解析。熵最小算法则能够处理高度重叠的体系^[5,9],但是对于同分异构体或者同系物体系,由于重叠的峰较多,应用该算法也有一定的难度,所以重建浓度的峰型与实际有一定的差距。

利用保留指数可以验证定性结果是否正确。理论上,各组分保留指数的大小应与出峰顺序一致。由于 NIST 数据库只给出了 A3 的保留指数估值,将 A1 和 A3 的保留指数估值做比较,结果与出峰顺序一致;对于组分 A1、B2 和 B3,其出峰顺序与实际保留指数的大小相符合;A1 和 B3 互为同分异构体,其出峰顺序与它们的保留指数一致。

常规的仪器分析方法在分析和鉴定天然产物未知成分时,其分析时间长、方法复杂、成本高。利用熵最小算法则只需要对天然产物进行简单的 GC 分离,而不需要反复调整 GC 程序升温条件,简化了样品的前处理过程,能够得到混合物中各组分重建纯谱信息,节省了人工和时间成本。将熵最小算法与分析仪器相结合,不仅可用于玫瑰精油等天然产物的分析,还可广泛用于中药、食品以及其他多种混合复杂组分的快速分离、分析和鉴定过程,具有通用性强的优点,同时也为下一步的工作提供了快速、简便的分离方法。将熵最小算法应用于中药成分分析过程中,可以简化中药成分的分离鉴定过程,避免了因样品中有效组分的变质、失效而造成的实验结果不准确,可用于中药中有效、活性成分的快速分离和提纯;同样,该算法也有助于中药成分中有害组分的快速分析鉴定,有助于解决中药分析中有效或有害组分不明确的问题,这为中药的标准化提供了一种有效的方法。综上,采用熵最小算法对 GC/MS 混合谱图进行纯谱重建和分离,通过与数据库比对得到成分信息,是对天然产物、食品、中药成分分析的一种快速、高效、简便的方法。

3 结论

本研究采用熵最小算法对玫瑰精油 GC/MS 混合谱图进行分析,在不对称的低丰度共流出峰形中发现 8 个未知物,并得到它们的纯谱数据,通过检索 NIST 数据库,分别对其中 4

个组分进行定性,发现了4个未知物。虽然该算法不能定性全部的未知物组分,但是本结果对于化合物分离和鉴定仍具有一定的意义,它不仅确定了共流出的组分数,还确定了未知组分的TIC时间,这为进一步的分离分析提供了可靠的数据。

参考文献:

- [1] 黄朝情,郭宝林,黄文华,等. 北京妙峰山玫瑰精油化学成分的GC-MS分析[J]. 北京农学院学报,2011,26(1):46-50.
- HUANG Chaoqing, GUO Baolin, HUANG Wenhua, et al. Analysis of chemical component of essential oil from roseleaf cultivated in Beijing Miaofengmountain by GC-MS[J]. Journal of Beijing University of Agriculture, 2011, 26(1): 46-50(in Chinese).
- [2] ULUSOY S, BOSGELMEZ-TINAZ G, SECILMIS-CANBAY H. Tocopherol, carotene, phenolic contents and antibacterial properties of rose essential oil, hydrosol and absolute[J]. Current Microbiology, 2009, 59(5): 554-558.
- [3] 梁逸曾,许青松. 复杂体系仪器分析[M]. 北京: 化学工业出版社,2012:153-155.
- [4] CHEW W, WIDJAJA E, GARLAND M. Band-target entropy minimization (BTEM): An advanced method for recovering unknown pure component spectraapplication to the FTIR spectra of unstable organometallic mixtures[J]. Organometallics, 2002, 21(9): 1 982-1 990.
- [5] ZHANG H J, GARLAND M, ZENG Y Z, et al. Weighted two-band target entropy minimization for the reconstruction of pure component mass spectra: Simulation studies and the application to real systems[J]. Journal of the American Society for Mass Spectrometry, 2003, 14(11): 1 295-1 305.
- [6] ZHANG H J, CHEW W, GARLAND M. The multi-reconstruction entropy minimization method: Unsupervised spectral reconstruction of pure components from mixture spectra, without the use of a Priori information[J]. Applied Spectroscopy, 2007, 61(12): 1 366-1 372.
- [7] MEIJIA J, MESTER Z, DULIVO A. Mass spectrometric separation and quantitation of overlapping isotopologues. $H_2O/HOD/D_2O$ and $H_2Se/HDS_2/D_2Se$ mixtures[J]. Journal of the American Society for Mass Spectrometry, 2006, 17(7): 1 028-1 036.
- [8] CHEN L, GARLAND M. Use of Entropy minimization for the preconditioning of large spectroscopic data arrays: Application to in-Situ FTIR studies from the unmodified homogeneous rhodium catalyzed hydroformylationreaction[J]. Applied Spectroscopy, 2002, 56(11): 1 422-1 428.
- [9] GAO F, ZHANG H J, GUO L F, et al. Application of the BTEM family of algorithms to reconstruct individual UV-Vis spectra from multi-component mixtures[J]. Chemometrics and Intelligent Laboratory Systems, 2009, 95(1): 94-100.
- [10] WIDJAJA E, SEAH R K H. Application of Raman microscopy and band-target entropy minimization to identify minor components in model pharmaceutical tablets[J]. Journal of Pharmaceutical and Biomedical Analysis, 2008, 46(2): 274-281.
- [11] MALINOWSKI E R, HOWERY D G. Factor analysis in chemistry[M]. New York: Wiley-Interscience,1980: 24-94.
- [12] SHAO X G, YU Z L, SUN L, et al. Immune algorithms in analytical chemistry[J]. Analytical Chemistry, 2003, 22(2): 59-69.
- [13] MALINOWSKI E R. Window factor analysis: Theoretical derivation and application to flow injection analysis data[J]. Journal of Chemometrics, 1992, 6(1): 29-40.
- [14] 廖燕芝,黄辉,杨代明,等. GC-MS测定食品中丙烯酰胺含量的定量方法比较[J]. 食品与机械, 2013,29(5):91-94.
- LIAO Yanzhi, HUANG Hui, YANG Daiming, et al. Comparison of quantitative methods for determination of acrylamide in foods by GC-MS[J]. Food & Machinery, 2013, 29(5): 91-94(in Chinese).