



## (12) 发明专利

(10) 授权公告号 CN 109100477 B

(45) 授权公告日 2022.10.21

(21) 申请号 201710588321.1

(51) Int.Cl.

(22) 申请日 2017.07.19

G01N 33/03 (2006.01)

(65) 同一申请的已公布的文献号

审查员 胡媛媛

申请公布号 CN 109100477 A

(43) 申请公布日 2018.12.28

(30) 优先权数据

15/628,043 2017.06.20 US

(73) 专利权人 香港理工大学

地址 中国香港九龙红磡

(72) 发明人 姚钟平 伍卓知 吴子浚 李素莹

(74) 专利代理机构 北京易光知识产权代理有限公司 11596

专利代理人 耿晓兵

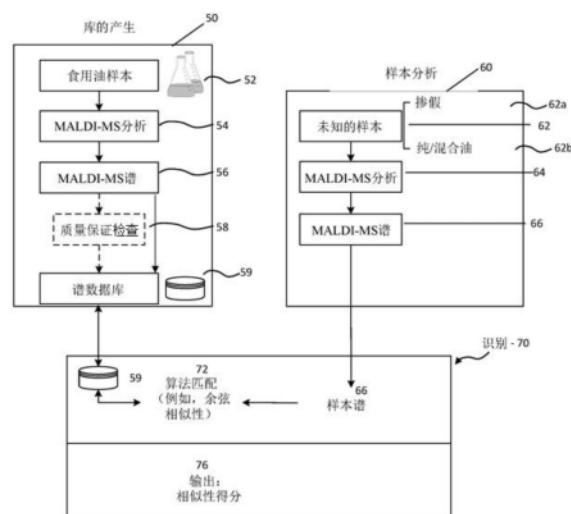
权利要求书2页 说明书11页 附图13页

## (54) 发明名称

食用油分析方法、识别系统、产生库的方法及数据载体

## (57) 摘要

本发明提供一种用于分析一个或多个食用油样本的方法和食用油样本识别系统。在一个实施例中，本发明提供：校准针对一个或多个食用油样本获得的基质辅助激光解吸/电离质谱法(MALDI-MS)数据以获得用于一个或多个食用油样本的校准的谱数据；以及将所述校准的谱数据与用于多个参考食用油样本的校准的MALDI-MS谱的库进行比较以确定一个或多个食用油样本的最可能组成。本发明还提供一种产生用于多个不同类型的多个参考食用油样本的基质辅助激光解吸/电离质谱法(MALDI-MS)数据的库以用于识别除所述参考食用油样本外的样本的方法以及一种其上具有指令的数据载体。



B

CN 109100477 B

1.一种用于分析一个或多个食用油样本的方法,所述方法包括:

校准针对一个或多个食用油样本获得的基质辅助激光解吸/电离质谱法(MALDI-MS)数据,其中通过使用从所述一个或多个食用油样本的所述基质辅助激光解吸 / 电离质谱法数据的三酰甘油(TAG)峰选择的参考峰来执行所述MALDI-MS数据的校准,以获得用于一个或多个食用油样本的经校准谱数据;以及

将所述经校准谱数据与用于多个参考食用油样本的校准的MALDI-MS谱的库进行比较,其中通过使用从所述MALDI-MS数据的TAG峰选择的参考峰来执行所述MALDI-MS谱的校准,其中所述比较确定所述一个或多个食用油样本的最可能组成,并且所述比较使用余弦相似性方法来执行。

2.根据权利要求1所述的用于分析一个或多个食用油样本的方法,其中,从所述一个或多个食用油样本的所述MALDI-MS数据的TAG峰中选择的参考峰包括908Da和882Da处的TAG峰。

3.根据权利要求2所述的用于分析一个或多个食用油样本的方法,其中余弦相似性得分是根据所述参考食用油样本存在于所述一个或多个食用油样本中的可能性,从最可能到最不可能来进行排名的;以及

基于最高排名的余弦相似性得分而确定所述一个或多个食用油样本的最可能组成。

4.根据权利要求2或3中任一权利要求所述的用于分析一个或多个食用油样本的方法,其中通过以下方式进行所述余弦相似性方法:将用于所述一个或多个食用油样本的所述校准的谱数据的一个或多个区与所述校准的MALDI-MS谱的库中所述参考食用油样本的一个或多个对应区进行比较,其中所述一个或多个区选自包括高质量、低质量和三酰甘油区的群组。

5.根据权利要求1所述的用于分析一个或多个食用油样本的方法,其中在校准后且在与用于多个参考食用油样本的所述校准的MALDI-MS谱的库进行比较之前,通过数据分箱将所述校准的谱数据量化。

6.根据权利要求5所述的用于分析一个或多个食用油样本的方法,其中通过将整个所述校准的谱数据划分为 $0.5m/z$ 的间隔,使每一箱内的所有读数的强度平均化,且将 $m/z$ 读数设定为在间隔的中间的 $m/z$ 值来执行所述数据分箱。

7.根据权利要求5或权利要求6所述的用于分析一个或多个食用油样本的方法,其中在分箱后,通过将每一箱的强度除以所有箱的最大强度或总强度、乘以适当按比例缩放的参数且舍入到最近整数,来使所述校准的谱数据正规化。

8.根据权利要求1所述的用于分析一个或多个食用油样本的方法,其中使用选自包括特性峰匹配方法、偏最小二乘判别分析(PLS-DA)或基于决策树的方法的群组的统计测试来执行所述校准的谱数据与用于多个参考食用油样本的校准的MALDI-MS谱的库之间的所述比较。

9.一种食用油样本识别系统,其包括:

库,其包括用于多个参考食用油样本的多个校准的MALDI-MS谱数据,其中所述多个参考食用油样本的库使用从所述MALDI-MS谱数据的TAG峰选择的参考峰来执行校准;

处理器,其经配置以接收至少一个或多个食用油样本的MALDI-MS数据,并且校准从所述一个或多个食用油样本获得的所述MALDI-MS数据,其中所述一个或多个食用油样本的所

述MALDI-MS数据通过使用从所述一个或多个食用油样本的所述MALDI-MS数据的TAG峰选择的参考峰来执行校准；并且

将所述一个或多个食用油样本的所述校准的MALDI-MS数据与所述库进行比较以确定所述一个或多个食用油样本的预测组成，其中所述比较使用余弦相似性方法来执行。

10. 根据权利要求9所述的食用油样本识别系统，其中所述处理器经配置以输出确定的余弦相似性得分。

11. 根据权利要求10所述的食用油样本识别系统，其中所述处理器经配置以对从所述一个或多个食用油样本获得的校准的所述MALDI-MS数据的一个或多个区以及所述库中的所述参考食用油样本的一个或多个对应区进行所述余弦相似性方法，其中所述一个或多个区选自包括高质量、低质量和TAG区的群组。

12. 根据权利要求9所述的食用油样本识别系统，其中在校准后且在与所述库的所述MALDI-MS谱数据进行比较之前，所述处理器经配置以通过数据分箱使从所述一个或多个食用油样本获得的校准的所述MALDI-MS数据量化。

13. 一种其上具有指令的数据载体，所述指令在由处理器执行时使所述处理器进行以下操作：

接收针对一个或多个食用油样本获得的MALDI-MS数据；

校准针对所述一个或多个食用油样本获得的所述MALDI-MS数据，其中通过使用从所述一个或多个食用油样本的所述MALDI-MS数据的TAG峰选择的参考峰来执行所述校准；以及

将从所述一个或多个食用油样本获得的校准的所述MALDI-MS数据与用于多个参考食用油样本的校准的MALDI-MS谱的库进行比较，其中通过使用从所述MALDI-MS数据的TAG峰选择的参考峰执行所述MALDI-MS谱的校准，以确定所述一个或多个食用油样本的最可能组成，其中所述比较使用余弦相似性方法来执行。

## 食用油分析方法、识别系统、产生库的方法及数据载体

### 技术领域

[0001] 本发明涉及用于使用光谱分析、具体来说基质辅助激光解吸/电离质谱法(MALDI-MS)识别样本中存在的食用油以及使样本数据与潜在候选者的谱库匹配的系统和方法。

### 背景技术

[0002] 消费者健康意识的提高以及对所消耗食物中存在的配料的关注度的提高均意味着食物供应链中的所有实体都需要证明所使用配料的完整性和纯度。

[0003] 食用油、特别是用于烹调的食用油的完整性和纯度已经得到显著的关注。在中国的广东地区、中国香港、中国台湾地区,已经存在许多掺假的油(“地沟油”)的情况,其中使用了再回收的油代替高质量食用油。

[0004] 确定食用油样本的真实性的一个方法是关注于在样本-标记物中检测食物残余标记物,其可包含辣椒素(辣椒的标记物)、丁香酚(调味品的标记物)、十一烷酸(受热蔬菜油的标记物)以及13-甲基十四烷酸(动物油的标记物)。然而,单独关注于这些标记物的存在或不存在并不决定食用油样本的真实性;因为无良的供应商可以简单地移除这些标记物。

[0005] 因此,使用气相色谱法-火焰电离检测器(GC-FID)的分析是验证食用油的ISO标准方法。在该方法中,首先使用氢氧化钠水解油样本,然后使用三氟化硼催化剂将其转换成脂肪酸的甲酯。随后使用GC-FID分离且检测脂肪酸甲酯。

[0006] 随后根据食用油样本的脂肪酸组成来确认食用油样本的身份。然而,这些过程是相对耗时且劳动密集的,意味着处置大量样本是不可能的。例如基于液相色谱法(LC)的方法等其它方法也具有相似的可扩展性问题。

[0007] 努力解决以上缺陷的一种方法是,已经开发使用基质辅助激光解吸/电离质谱法(MALDI-MS)的简单分析规程。

[0008] (参见例如Ng,T.T.、So,P.K.、Zheng,B.、Yao,Z.P.的“通过基质辅助激光解吸/电离质谱法对混合的食用油和地沟油的快速筛查(Rapid screening of mixed edible oils and gutter oils by matrix-assisted laser desorption/ionization mass spectrometry)”, AnalyticaChimicaActa,2015,884,70-76。

[0009] MALDI-MS广泛用于包含脂质的生物样本的分析,且具有简单的样本制备、较短的分析时间,以及高灵敏度而无需色谱分离。从MALDI-MS获得的食用油的质谱广泛公认为用于区分和验证的“指纹”。

[0010] 通过MALDI-MS产生的谱是由对应于像二酰基甘油(DAGs)的片段和三酰基甘油(TAGs)的钠加成物的信号来控制。也已经探究较低质量区(100-500Da)和较高质量区(1000-2000Da),但通常未获得显著信号。

[0011] 食用油的加热和混合改变食用油的组成,从而更改其MALDI-MS谱的TAG图样。油的长期加热通常造成TAG的降解以及例如二酰基甘油、单酰基甘油、游离脂肪酸、氧化TAG和TAG聚合物等化合物的形成。此改变可通过与参考谱进行比较来检测以确定正被测试的样

本的身份和/或组成。

[0012] 在处理MALDI-MS谱以确定样本的真实性/身份的一种方法中,使用主成分分析(PCA) (一种统计分析形式)来对正被分析的食用油样本中的TAG峰进行谱的比较。

[0013] 然而,手动地处理MALDI-MS样本以及使用PCA分析来识别样本或检验样本的真实性包含若干大型步骤。相关峰的提取的初始步骤需要经训练的操作者的手动选择,所述操作者已经被训练从而知道选择哪些峰。

[0014] PCA分析过程自身也可能是误差/复杂性的来源,因为对于例如葵花油和葡萄籽油等一些类型的油,其TAG组成类似于其它油品种。这意味着需要若干PCA步骤来区分油,这个过程可能是极冗长的。

[0015] 最后,PCA过程的输出是输入数据到PCA绘图上的投影,其比较样本与数据库中的数据之间的相似性。然而,无法直接显示样本的身份。根据PCA分析,经训练的操作者仍需要从结果归纳出合理的结论。

[0016] TAG图样被认为是食用油的“指纹”。然而,一些品种共享高度相似的TAG图样(例如橄榄油和鳄梨油),这无法容易地通过PCA来区分。

[0017] 使用PCA分析的先前的方法仅考虑PCA分析中的TAG相关的峰。例如氧化相关以及环肽相关的峰等其它峰也可能是有用的。然而,无法使用所建立的PCA模型直接适应那些额外的峰,从而要求PCA的重新建模来配合新信息。此过程同样需要有经验和经训练的操作者。

[0018] 因此,当使用PCA来分析MALDI-MS谱时,所述谱仍需要由经高度训练的人员处理且手动地匹配,这是昂贵、费力且耗时的,且限制了扩展MALDI-MS系统用于食用油分析的能力。

[0019] 因此,使用MALDI-MS方法对食用油的潜在分析需要达到的状态与现有的当前状态之间存在显著差距。

## 发明内容

[0020] 本发明的目的是提供一种系统和方法,其至少解决或改善上述问题中的一些或者为公众提供替代的选择。本发明的特征和优点将在以下描述中加以阐述,并且在某种程度上根据所述描述本发明的特征和优点将是明显的,或者可以通过对本文中所公开的原理的实践习得本发明的特征和优点。借助于在所附权利要求书中具体指出的装置和组合可实现且获得本发明的特征和优点。

[0021] 根据本发明的第一方面,提供一种用于分析一个或多个食用油样本的方法,所述方法包括:

[0022] 校准针对一个或多个食用油样本获得的基质辅助激光解吸/电离质谱法(MALDI-MS)数据以获得用于一个或多个食用油样本的校准的谱数据;以及

[0023] 将所述校准的谱数据与用于多个参考食用油样本的校准的MALDI-MS谱的库进行比较以确定所述一个或多个食用油样本的最可能组成。

[0024] 所述校准的谱数据与用于多个参考食用油样本的校准的MALDI-MS谱的库进行的所述比较可以使用余弦相似性测试来执行。

[0025] 余弦相似性得分可以根据所述参考食用油样本存在于所述一个或多个食用油样

本中的可能性,从最可能到最不可能来进行排名;以及基于最高排名的余弦相似性得分而确定所述一个或多个食用油样本的最可能组成。

[0026] 任选地,可以通过以下方式进行所述余弦相似性测试:将用于所述一个或多个食用油样本的所述校准的谱数据的一个或多个区与所述校准的MALDI-MS谱的库中所述参考食用油样本的一个或多个对应区进行比较,其中所述一个或多个区选自包括高质量、低质量和三酰甘油(TAG)区的群组。

[0027] 可通过使用三酰甘油(TAG)峰和二羟基苯甲酸(DHB)基质峰中的至少一个作为参考峰来执行校准。

[0028] 在校准后且在与用于多个参考食用油样本的所述校准的MALDI-MS谱的库进行比较之前,可通过数据分箱(data binning)将所述校准的谱数据量化。可通过将整个所述校准的谱数据划分为 $0.5m/z$ 的间隔,使每一箱(bin)内的所有读数的强度平均化,且将 $m/z$ 读数设定为在间隔的中间的 $m/z$ 值来执行数据分箱。

[0029] 在分箱后,可通过将每一箱的强度除以所有箱的最大强度或总强度、乘以适当按比例缩放的参数且舍入到最近整数而使所述校准的谱数据正规化。

[0030] 也可以使用选自包括特性峰匹配方法、偏最小二乘判别分析(PLS-DA)或基于决策树的方法的群组的统计测试来执行所述校准的谱数据与用于多个参考食用油样本的校准的MALDI-MS谱的库之间的所述比较。

[0031] 在本发明的另一方面中,提供一种食用油样本识别系统,其包括:

[0032] 输入系统,其经配置以接收一个或多个食用油样本的基质辅助激光解吸/电离质谱法(MALDI-MS)数据;

[0033] 库,其包括用于多个参考食用油样本的多个校准的MALDI-MS谱数据;

[0034] 处理器,其经配置以校准从所述一个或多个食用油样本获得的所述MALDI-MS数据且将其与所述库进行比较以确定所述一个或多个食用油样本的预测组成。

[0035] 所述处理器可经配置以使用余弦相似性测试将从所述一个或多个食用油样本获得的校准的MALDI-MS数据与所述库的参考食用油样本进行比较,且输出确定的余弦相似性得分。

[0036] 所述处理器可经配置以对从所述一个或多个食用油样本获得的校准的MALDI-MS数据的一个或多个区以及所述库中的参考食用油样本的一个或多个对应区进行所述余弦相似性测试,其中所述一个或多个区选自包括高质量、低质量和TAG区的群组。

[0037] 可通过使用TAG峰和DHB基质峰中的至少一个作为参考峰来执行对从所述一个或多个食用油样本获得的MALDI-MS数据和所述库的MALDI-MS谱数据的校准。

[0038] 在校准后且在与所述库的MALDI-MS谱数据进行比较之前,所述处理器可经配置以通过数据分箱使从所述一个或多个食用油样本获得的校准的MALDI-MS数据量化。

[0039] 在本发明的再一方面中,提供一种产生用于多个不同类型的多个参考食用油样本的基质辅助激光解吸/电离质谱法(MALDI-MS)数据的库以用于识别除所述参考食用油样本外的样本的方法,所述方法包括:

[0040] 提供具有已知类型的多个参考食用油样本;

[0041] 通过使用TAG峰和DHB基质峰中的至少一个作为参考峰来校准用于所述多个参考食用油样本中的每一个的所述MALDI-MS数据;

- [0042] 通过数据分箱使用于所述参考食用油样本的所述MALDI-MS数据量化；
- [0043] 通过将每一箱的强度除以最大箱的强度或所有箱的总强度以将所述参考食用油样本的所述MALDI-MS数据正规化；以及
- [0044] 使正规化的所述MALDI-MS数据与食用油类型关联。
- [0045] 本发明还可包含其上具有指令的数据载体，所述指令在由处理器执行时使所述处理器进行以下操作：
- [0046] 校准针对一个或多个食用油样本获得的基质辅助激光解吸/电离质谱法 (MALDI-MS) 数据；以及
- [0047] 将从所述一个或多个食用油样本获得的校准的所述MALDI-MS数据与用于多个参考食用油样本的校准的MALDI-MS谱的库进行比较以确定所述一个或多个食用油样本的最可能组成。

## 附图说明

[0048] 为了描述可以获得本发明的上述和其它优势和特征的方式，将参考附图中所说明的上文所简单描述的原理的特定实施例来呈现所述原理的更具体描述。在理解这些图式仅描绘本发明的示范性实施例并且因此不应认为其限制本发明的范围的情况下，将通过使用附图来以附加的特征和细节来描述并解释本文的原理。

[0049] 下文将通过实例并且参考附图来进一步详细解释本发明的优选实施例，在附图中：

- [0050] 图1a描绘用于菜籽油的示范性MALDI质谱，其示出了来自500-1000Da的全谱；
- [0051] 图1b是图1a的TAG区的放大视图；
- [0052] 图1c示出了菜籽、花生、橄榄和葵花食用油的特性峰，聚焦于这些油的TAG区；
- [0053] 图2示出了各种再回收的油的示范性MALDI-MS谱；
- [0054] 图3a到图3e描绘菜籽油和橄榄油的混合物在各种浓度下的示范性MALDI-MS谱；
- [0055] 图4a描绘不同类型的油的示范性PCA绘图，其中所述油可被划分成4个群组 (通过层次聚类而获得的相同结果)；
- [0056] 图4b描绘图4a的绘图的群组2的示范性后续PCA分析，其中进行进一步的PCA 以进一步识别不同类型的油；
- [0057] 图5a描绘本发明的系统的实施例的示范性流程图；
- [0058] 图5b描绘图5a的流程图的核心以及工作流如何进行；
- [0059] 图6a示出了库的针对已知样本存储的特性谱的示范性数据，以图形方式向用户描绘；
- [0060] 图6b示出了图6a的样本的谱区的放大部分的示范性表示；
- [0061] 图7a描绘纯油的未知样本与存储在库中的棕榈油的参考样本之间的强相关性；
- [0062] 图7b描绘掺假油的未知样本的分析的结果以及与参考样本中的花生或橄榄谱的较低相关性得分；
- [0063] 图8a是不会存储在数据库中的葵花油的不良TAG谱的示范性样本；
- [0064] 图8b是将作为参考样本存储在数据库中的葵花油的良好TAG谱的示范性样本；
- [0065] 图9a是玉米油的质谱分析的示范性样本，其中尚未恰当地进行校准；

- [0066] 图9b示出了在已恰当地进行校准之后的玉米油的同一样本；
- [0067] 图9c示出了用于参考的玉米油的TAG谱的参考样本；
- [0068] 图9d示出了另一样本，其中山茶油的TAG谱被较差地校准且具有较差的分辨率；
- [0069] 图9e示出了图9d的样本，其中已矫正分辨率和校准；
- [0070] 图10a示出了误标记为亚麻籽油的示范性样本；
- [0071] 图10b示出了样本的GC-FID分析，其支持此样本已被误标记的结论。

## 具体实施方式

[0072] 下文详细论述本发明的各种实施例。虽然讨论了特定实施方案，但是应了解，这样做是为了说明性目的。相关领域的普通技术人员将认识到，可以在不脱离本发明的范围的情况下使用其它组件和配置。

[0073] 所公开的技术解决了此项技术中对于用于食用油分析的可扩展、可靠的技术的需要。

- [0074] 用于样本(库和未知样本)的分析的典型MALDI规程

[0075] 典型MALDI-MS规程涉及基质溶液与样本溶液的混合，其随后被允许干燥到MALDI板上。在样本和基质晶体的形成之后，将MALDI板插入到MALDI-MS仪器中用于分析。

- [0076] MALDI-MS分析中不涉及色谱分离，因此允许食用油样本的快速分析。

[0077] 在示范性MALDI-MS规程中，可将食用油样本直接加载到预沉积有基质层的MALDI板上以用于MALDI-MS分析。用于一个样本的样本加载可在几秒内完成，且大约三百个样本可加载到同一MALDI板上。

[0078] 在根据示范性方法的样本加载的特定步骤中，将丙酮中100mg mL<sup>-1</sup> DHB的1μL的使用的等分部分加载到MADLI板的点上且风干以形成基质层。随后通过吸液管尖端或棉签尖端转移每一油样本的大约0.2μL以在基质层上形成薄油层。

- [0079] 随后将该板引入到质谱仪中用于MALDI-MS分析。

[0080] 使用UltraflexXtreme MALDI-TOF/TOF质谱仪(德国布鲁克道尔顿公司)用于分析。MALDI源的激光是在2000Hz的频率下操作的智能束-II(Nd:YAG, 355nm)脉冲激光。质谱仪在正反射模式中操作。用于离子源1、源2、透镜、反射器1、反射器2和脉冲离子提取的正反射模式的设定分别是20.00kV、17.75kV、7.00kV、21.10kV、10.85 kV和140ns。取样率和数字化器设定成5.00GS/s。在必要时采用扩展的质量范围。

- [0081] 以PEG溶液混合物(PEG600/PEG1000/PEG2000/NaI=1/2/2/5(v/v))校准质谱仪。

[0082] 使用flexControl 3.4(德国布鲁克道尔顿公司)程序执行谱采集。使用flexAnalysis 3.4(德国布鲁克道尔顿公司)程序分析质谱。使用质心算法用于峰检测。

- [0083] 参考图1a到图1c，示出了产生于选定食用油样本的MALDI-MS分析的典型谱。

[0084] 现参看图1a，示出了产生于来自500-1200m/z的菜籽油样本的MALDI-MS谱的全谱的结果。可以看到在907.8Da处的主要三酰甘油峰连同在881.8Da处的较小峰。还存在片段区中群集于603.6Da周围的若干峰。在920Da后方的峰是典型的氧化/热产物。

[0085] 群集于935.8Da周围的峰可为氧化的TAG，且在1059.8Da处的峰似乎与TAG 片段群集离子相关。

- [0086] 在作为来自850-920m/z的图1a的区的放大部分的图1b中，可看见相同的两个 TAG

峰,其中字母表示各种脂肪酸,包含P(棕榈酸)、O(油酸)、L(亚麻油酸)和 Ln(次亚麻油酸)。

[0087] 最后,图1c描绘不同类型的若干其它食用油样本的MALDI MS谱的峰,包含:(i) 菜籽,(ii) 花生,(iii) 橄榄,(iv) 葵花,具有其特征TAG峰(且按较轻微程度来说,其像DAG的峰)。

[0088] 图2a-图2f描绘从地处广州的国家分析中心获得的且TAG图样似乎相当不同于数据库的纯食用油的再回收油的各种样本的TAG峰。

[0089] 图3a到图3d描绘通过橄榄油和菜籽油的各种混合物的分析已获得的典型 MALDI-MS谱。特别强调的是随着混合物的组分的相对比例改变而在特性峰中出现改变。可以看到在908Da处的特性TAG峰和在882Da处的次级峰是图3a中的菜籽油的存在的独特指示符。如在图3a到图3e处所示,当橄榄油的组成增加时,在882Da 处的峰的强度增加。谱的改变允许对橄榄油和菜籽油的各种混合物进行区分。

[0090] 现参看图4a,示出了在已针对样本获得例如图1、图2和图3的那些峰之后,来自现有技术方法的输出,其中使用PCA进行样本识别。

[0091] PCA使用正交线性变换将可能相关的变量的观测值转换为线性不相关变量的一组值。此技术允许高维度数据集的视觉化和处理,但同时保持数据集的尽可能多的变化。

[0092] 在进行样本的主成分分析时,从样本的第一和第二主要TAG组分产生得分绘图。

[0093] 对TAG组分使用PCA的结果示出了来自同一品种的样本单独地群集,且不同的植物油品种可以彼此清楚地区分。(参见Ng, T.T.、So, P.K.、Zheng, B.、Yao, Z.P.的“通过基质辅助激光解吸/电离质谱法对混合的食用油和地沟油的快速筛查(Rapid screening of mixed edible oils and gutter oils by matrix-assisted laser desorption/ionization mass spectrometry)”,*AnalyticaChimicaActa*,2015,884,70-76。)

[0094] 在图4a中,对样本的MALDI-MS的结果可划分成基本上四个群组,如由轮廓线指示。

[0095] 群组1(10)可能是花生油,群组2(12)可能是亚麻籽,群组3(14)可能是具有类似于橄榄油的TAG图样的蔬菜油,且群组4(16)可能是其它蔬菜油。

[0096] 参考图4b,可以看到进一步的PCA分析的结果,其中进一步分析子群组2。菜籽/ 油菜籽油、芝麻油、米糠油和棉籽油的样本(位于下半部的点)可能与子群组2中的其它油(位于上半部的点)区分。子群组2中的其它油将需要进一步PCA分析以完全区分。

[0097] 然而,如本发明的背景中所提到,PCA分析存在若干缺陷,这意味着其是不可扩展的且必须由有技巧的操作者执行。

[0098] 因此,本发明提供一种方法和系统,其解决了这些缺陷,且实现用于食用油的分析、检验和识别的稳健、可扩展的技术。

[0099] 参考图5a,可以看到本发明的方法和系统中的步骤的三个主要群组。

[0100] 参照组成部分50,公开了与产生MALDI-MS谱数据的库相关联的一系列步骤。在步骤52处,选择具有已知类型和发源的多个食用油样本用于评估。在产生的库中,我们已选择来自各种供应商的多达六百个样本,包含来自中国大陆、中国台湾、中国香港以及美国等国家和地区 的西格玛-奥德里奇公司。在许多情况下已选择来自不同来源的一个类型的油的多个实例以提供完整数据库。

[0101] 如在步骤54中在本发明的规程下所公开的,这些样本经受MALDI-MS分析。

[0102] 在步骤56处从分析获得的MALDI-MS谱通过在步骤58处的任选的质量保证检查,其

涉及由经训练的操作者来检查。在此检查期间,操作者检查谱的校准和分辨率以便确保最佳参考谱包含在库中。高水平的氧化产物和/或较差的校准将造成数据被检查者拒绝。将了解,虽然此步骤是任选的,特别是在谱数据的库的产生中是任选的,但对谱数据进行检查以确保拒绝显而易见的误差或被污染的样本,这又增加了所获得谱数据的库的完整性。

[0103] 在质量保证检查后,在步骤59处将MALDI-MS谱存储在谱数据库中,从而形成食用油样本的谱数据的库。任选地,为了增加使用库中的样本进行假定识别的准确性,可从若干不同制造商获得已知油类型的多个不同样本,预期不管来自若干不同制造商的油的发源如何,都将观测到相当相似的谱。

[0104] 本发明的系统的另一部分包含图5a的框60处涉及的样本分析系统。将了解,样本分析允许检测掺假的油(62a)(与较贵的油混合的较便宜的油,且被误标记为纯的、常常更贵的油)和已经用于烹调且已经再回收的“地沟油”,以及来自未知样本的纯食用油或来自未知样本的食用油混合物的识别(62b)。

[0105] 根据本发明的较早描述的规程在步骤64未知样本经受MALDI-MS分析以在步骤66产生样本MALDI-MS谱。

[0106] 现参看本发明的识别组成部分70,概述地描述示范性识别系统中的一系列步骤。

[0107] 在算法匹配步骤72通过计算机的处理器匹配样本谱66。算法匹配可为一种比较的形式,例如余弦相似性测试,或类似方法,通过该类似方法将样本数据66与在库的产生组成部分50中针对多个食用油样本已获得的谱数据库59进行比较。参考图5b更详细地公开了该过程。

[0108] 样本谱66与谱数据库59之间的算法匹配的结果是根据库中存在的食用油样本的一个或多个相似性得分。

[0109] 如所属领域的技术人员所了解,无论通过哪种算法用于使样本的谱与食用油库中的对应谱匹配,产生的相似性得分越高,未知食用油样本的假定识别的机会越高。

[0110] 将了解,不管未知食用油样本是否为掺假的食用油样本、纯食用油样本或食用油的混合物,此过程都是相似的,即,通过算法进行与库的匹配并得到假定识别将是大体上相同的。

[0111] 现参看图5b,更详细地公开用于未知样本的谱的处理的过程。在66处,用户提供原始MALDI-MS谱用于分析。将了解,这可作为数据文件提供到在地理上远离执行 MALDI-MS谱的地点的中央处理设施。替代地,所述谱和处理可在同一设施处进行,而不偏离本发明。

[0112] 一旦已经从用户输入获得原始MALDI-MS谱,就在步骤72a处校准该谱。通常使用突出的TAG峰和DHB基质峰作为参考来进行校准。此校准过程确保样本被标准化,以便于对照标准化食用油谱峰的库的分析和分析的再现性。

[0113] 也可使用替代的基质,例如CHCA( $\alpha$ -氰基-4-羟基肉桂酸)和SA(芥子酸)。基于无论使用何种基质,也可通过特性峰执行适当校准。然而,应注意,如果使用CHCA则背景噪声较高,且如果芥子酸用作基质则信号强度非常差。

[0114] 为了帮助获得再现的结果,按照最佳实验室实践,还应当在进行分析之前以PEG溶液混合物(PEG600/PEG1000/PEG2000/NaI=1/2/2/5(v/v))来校准质谱仪。

[0115] 在校准后,在步骤72b处对数据进行分箱过程。如所属领域的技术人员所知,数据分箱或分桶(data binning or bucketing)是使数据量化的数据预处理技术。基本上,在数

据分箱中,落入给定小窗口或箱(bin)中的原始数据系列值由代表该间隔的值(常常是中心值)代替。

[0116] 如此项技术中已知,数据分箱减少了数据量(必定丢失信息)但促进了分析。在 MALDI-MS 谱分析中,箱的典型大小是 0.5m/z,但也可使用其它大小。所属领域的技术人员将了解,增加的箱大小将减小所获得数据的分辨率。箱的大小影响匹配的准确性和质量,且箱的最佳大小表示太多细节与太低分辨率之间的平衡。

[0117] 在本发明的示范性实施例中,使用的数据分箱过程是将整个谱划分为 0.5m/z 的间隔,使每一箱内的所有读数的强度平均化,且将 m/z 读数设定为在间隔的中间的 m/z 值。

[0118] 在步骤 72b 处的数据分箱过程之后,通过将每一箱的强度除以所有箱的最大强度、接着乘以 10000 且舍入到最近整数来将数据正规化。

[0119] 替代地,在正规化的替代过程中,可通过将每一箱的强度除以所有箱的总强度,并且然后将结果乘以 10000 且进行舍入来将样本数据正规化。

[0120] 在正规化过程之后,在步骤 72d 中,将已经校准、分箱且正规化的未知样本数据与食用油的参考谱的数据库(该数据库中的信息是用于已经被类似地校准的、分箱的且正规化的食用油)进行比较。该比较可使用余弦相似性方法来进行。

[0121] 如此项技术中已知,余弦相似性匹配是相似性量度,其中未知样本谱的样本表示为向量;且获得所述样本与表示库中的样本数据的多个向量的点积。

[0122] 余弦相似性计算两个向量之间的角度的余弦。

[0123] 在向量处于大体上同一方向上的情况下,两个向量之间的角度的余弦近似为 100% 或 1。然而,在向量(谱数据的抽象化)定向于不同方向上(近似为正交)的情况下,通过点积求解获得的两个向量之间的角度的余弦近似为零,即 0%。此外,在向量处于完全不同(相反)方向上的情况下,获得的余弦相似性是 -1。

[0124] 因此,余弦相似性是获得表示两个谱之间的相似度的数值得分的有用算法。

[0125] 任选地,也可使用其它数据处理技术,例如特性峰匹配方法(主要用于检测氧化产物或环肽的存在或不存在)、偏最小二乘判别分析(PLS-DA)或基于决策树的技术(例如,随机森林),以将经校准的、分箱的且正规化的 MALDI-MS 样本数据与参考谱的类似地经校准的、分箱的且正规化的数据进行比较。

[0126] 在步骤 76 处的过程的结果是油的识别。

[0127] 有利的是,如连续图中描绘的,其可以库的多个样本的余弦相似性的得分的排名系列的形式来提供,或替代地可简单地被选择为最高得分识别。

[0128] 现参看图 6a,可以看到食用油样本 S88(菜籽)的示范性参考谱。突出的 TAG 峰区 80 在图 6a 中被圈出以便于参考,并且在图 6b 中以放大视图进一步提取。

[0129] 在样本的左侧,可以看到存在菜籽油 82 的多个样本,其中已经在 84 处记录制造商和收集源的细节。

[0130] 还任选地存在通过链接 86 的选择而以文本格式查看原始数字数据的能力。

[0131] 大多数食用油具有在 870-885Da 周围的一个群组的 TAG 峰,以及在 900-910Da 处的另一群组的 TAG 峰。例如椰子油等一些食用油将具有其在不同区的 TAG 峰。对于每种类型的油,每一 TAG 的比率大部分由母代品种的酶类决定。因此,TAG 区内的峰的相对强度对于每一类型的油是特定的,且峰的强度对于每一类型的油形成独特形状。因此,TAG 区的位置和形

状可用作指纹以识别未知的油类型。

[0132] 因此,本发明提供了用户例如通过浏览多个不同类型食用油的参考谱且在许多情况下浏览特定食用油的多个样本的参考谱来进行查看的能力。

[0133] 此样本参考数据可被存储为一系列原始数字数据,但为了便于人为解释而以描绘的图形格式表示。

[0134] 现参看图7a,可以看到来自未知油样本的验证过程的输出。有利的是,文件可包含在由88指示的屏幕的部分以及输出由在90处下方的区中所示的相关性得分指示。在此实例中,呈文本格式的未知的油谱已被假定地识别为棕榈油,其具有0.9959的相对较强相关性得分。

[0135] 现参看图7b,可以看到存在两个样本,其已被分别提供在92和94处呈现的输出处。

[0136] 对于在92处被识别为黄油的样本,相关性得分相对较低(0.9410),意味着对该假定的识别没有太多的信心。

[0137] 类似地在94处,经识别的食用油是南瓜籽油,然而相关性得分也相对较低。

[0138] (注意:‘低得分’可取决于用法。如果用户仅想要识别未知物而无其它信息,那么用户可满足于0.97的得分,因为其将为最接近匹配。)

[0139] 然而,如果用户想要知道已知类型的油是否已经掺假或受热一段时间,那么用于正确匹配的阈值需要更高,因为较低得分将意味着油中的TAG已在某种程度上改变。

[0140] 一般来说,基于到目前为止获得的实验数据,0.97的阈值得分似乎为用于大多数目的的合理水平。

[0141] 这些结果可与图7a中针对棕榈油所示的高相关性得分进行比较。

[0142] 因此,考虑到弱相关性得分,查阅这些结果的用户将较不确信后面的样本的假定识别。

[0143] 现参看图8a,可以看到较差质量的示例性谱。如在917.7和933.7处的峰(分别为100和102)表示,样本中存在较高水平的氧化产物。因此,通过图5a的人为质量保证步骤(58),在建立谱数据库时将不包含此样本。

[0144] 替代地,如果这些样本是从未知的样本获得,那么系统将对照数据库中的油谱而产生较差的相关性得分,且用户将知道油已在某种程度上被改变(掺假、受热、储存太久等)。

[0145] 现参看图8b,可以看到正常葵花油样本的好得多的谱,其中氧化产物100、102明显不存在。

[0146] 类似地,现参看图9a,描绘玉米油的较差校准的谱,其中与玉米油相关联的典型TAG峰已经向右移位0.3Da。

[0147] 当将图9a的谱或较差校准的谱与图9b的参考谱进行比较时,TAG峰110和112显然移位,这是明显的,且因此在校准之后图9a的谱被调整0.3Da以提供恰当校准的谱,如图9c中所描绘。

[0148] 图9d中可以看到较差校准和分辨率的另一实例,其中描绘山茶油的TAG图样。查看此山茶油样本的特性峰114和116可以看到该样本已经再次被较差地校准,且在900与910之间的区118中经受较差的分辨率。所获得的峰应当为尖锐的且基线应当为较低的。图9d中的特性峰看起来太宽且基线被提高。

[0149] 因此,在操作者执行的检查和质量保证之后,如图9e中所描绘,TAG山茶油样本的重新分析给出好得多的更稳健的谱,其中校准和分辨率问题已经解决。

[0150] 参考图10a,存在对形成参考库的样本进行质量保证的必要性的又一实例。

[0151] 可以看到10a(i)中的谱的有问题的样本具有与10a(ii)和10a(iii)中描绘的亚麻籽油样本完全不同的峰图样。在对亚麻籽油的所声称的参考样本与参考库中的其它参考样本之间的误匹配进行识别之后,有问题的样本可经受GC-FID分析。如图10b中所描绘,此分析的结果揭露了该样本实际上被误标记,且该样本将不再作为亚麻籽油包含在参考数据库中。

[0152] 因此,可增加用于食用油样本的识别的参考库的完整性。

[0153] 本发明提供识别食用油的有利、潜在地可扩展的方法。这实现误标记的食用油的快速检测、掺假油和地沟油的识别,以及验证经标记的油谱的能力。

[0154] 也可能通过与参考库中的参考样本(例如图3a到3e中描绘的那些)的比较而识别混合油的主要元素和少量元素。

[0155] 在该使用中,用户可以实际所检测的组成来检查食用油的标签上所声称的相对比例。

[0156] 本发明的另一优点是可在分析实验室进行MALDI-MS分析,且参考库可以必要地位于在地理上远离分析实验室的位置。假定的识别可随后在因特网上进行,其中数据简单地上载到适当网站上。

[0157] 食用油的分析和假定识别可通过例行实验室程序而实行,而不需要不间断的训练或穷尽性统计分析。

[0158] 用于匹配谱的算法提供了对照参考库假设地识别广泛多种食用油的可靠、可扩展且高效的方式。

[0159] 包含自动数据匹配过程消除了对参考谱的人工匹配的需要。

[0160] 不同于先前的PCA方法,若新类型的油添加到数据库,所述算法不需要修改。

[0161] 结果也可自动显示给用户,这对于先前的方法是不可能的。

[0162] 上文实施例仅作为实例描述。在不脱离本发明的如所附权利要求书中所限定的范围的情况下,许多变化是可能的。

[0163] 为了清楚解释,在一些情况下本发明技术可呈现为包含单独的功能块,其包括装置、装置组件、体现在软件或硬件与软件的组合的方法中的步骤或例程。

[0164] 根据上述实例的方法可使用经存储的或另外从计算机可读介质可用的计算机可执行过程来实施。这些过程可包括例如致使或另外配置通用计算机、专用计算机或专用处理装置执行某一功能或功能群组的指令和数据。所使用计算机资源的部分可通过网络来接入。计算机可执行指令可以例如是二进制、例如汇编语言的中间格式指令、固件,或源代码。可用以存储指令、所使用信息和/或在根据所描述实例的方法期间产生的信息的计算机可读介质的实例包含磁盘或光盘、快闪存储器、具备非易失性存储器的通用串行总线(USB)装置、联网存储装置等。

[0165] 实施根据这些公开内容的方法的装置可包括硬件、固件和/或软件,且可采取任何多种形状因数。这些形状因数的典型实例包含便携式计算机、智能电话、小形状因数个人计算机等。本文所描述的功能性也可体现在外围设备或附加卡中。借助于进一步的实例,这些

功能性也可实施于电路板上的不同芯片中或在单个装置中执行的不同过程中。

[0166] 指令、用于递送这些指令的介质、用于执行指令的计算资源以及用于支持这些计算资源的其它结构是用于提供这些公开内容中描述的功能的方式。

[0167] 虽然使用多种实例和其它信息来阐释所附权利要求书的范围内的各个方面,但不应当基于这些实例中的特定特征或布置而暗示对权利要求书的限制,因为所属领域的技术人员将能够使用这些实例以推导出广泛的多种实施方案。

[0168] 此外虽然可能已经用结构特征和/或方法步骤的实例所特定的语言描述了一些主题,但应了解,所附权利要求书中限定的主题不一定限于这些描述的特征或动作。举例来说,这些功能性可在除本文识别的那些组件外的组件中不同地分布或执行。实际上,所描述的特征和步骤被公开作为所附权利要求书的范围内的系统和方法的组成部分的实例。

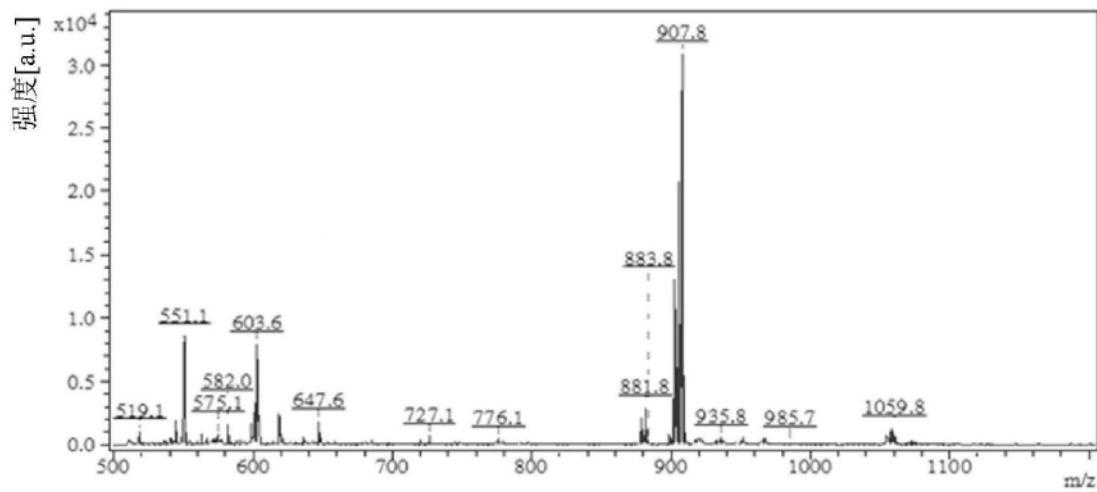


图1a

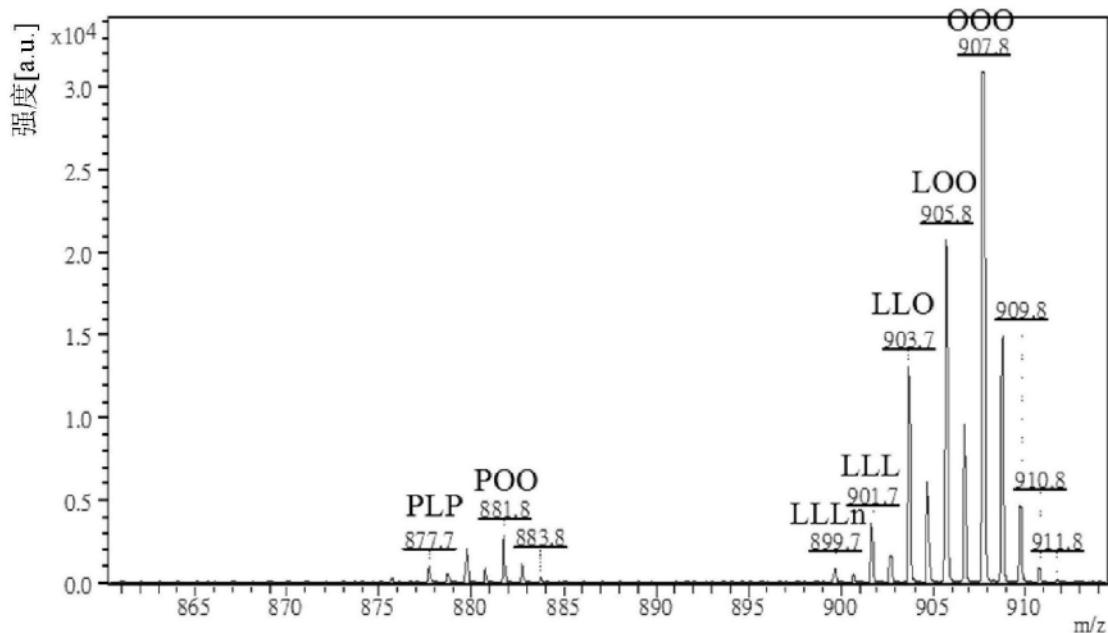


图1b

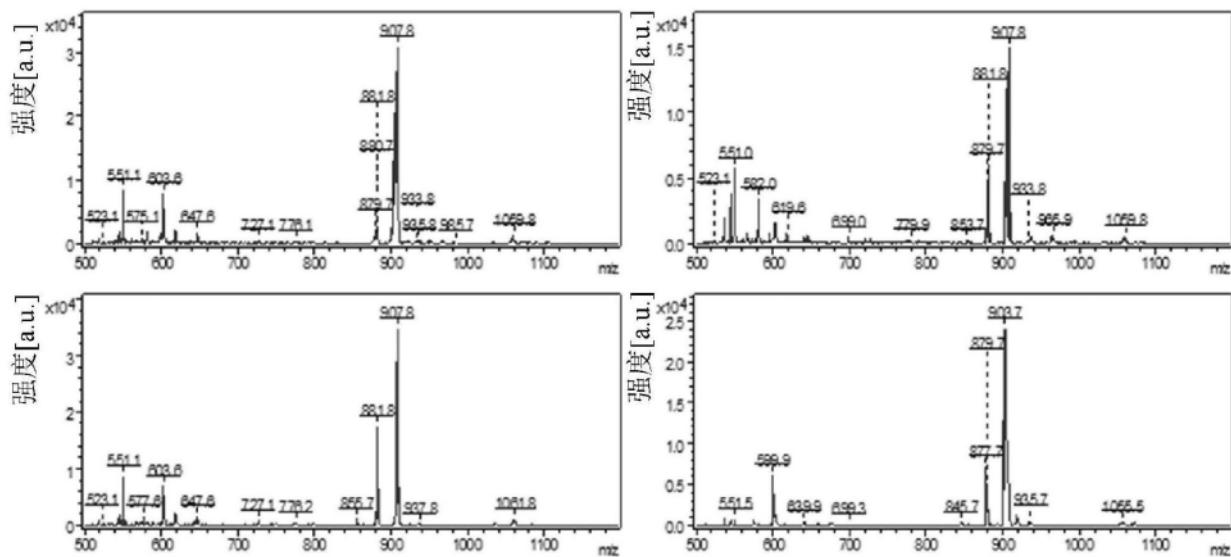
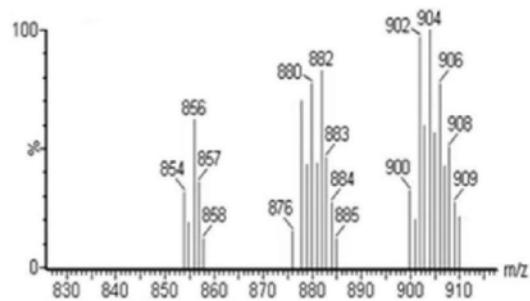
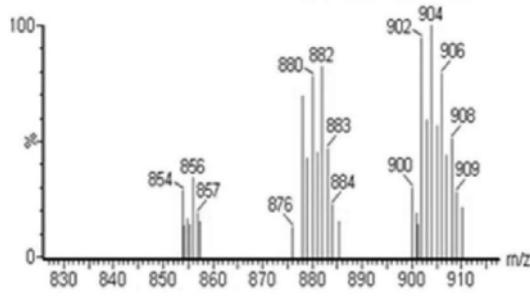


图1c



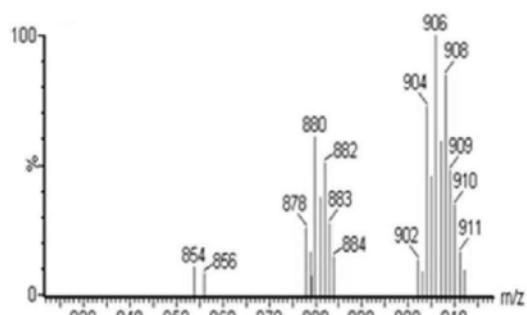
再回收的油样本6

图2a



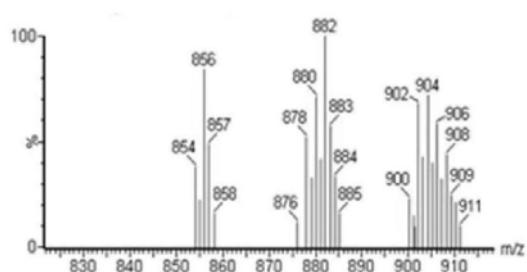
再回收的油样本9

图2b



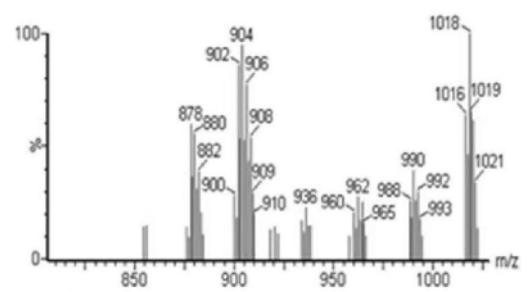
再回收的油样本15

图2c



再回收的油样本18

图2d



再回收的油样本19

图2e

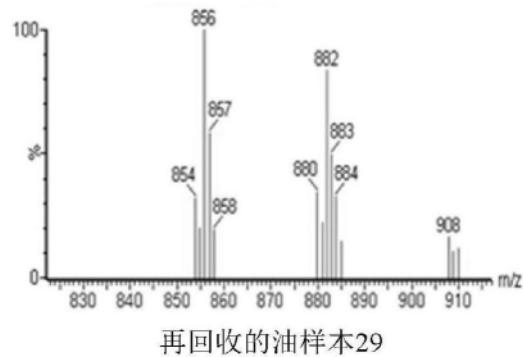


图2f

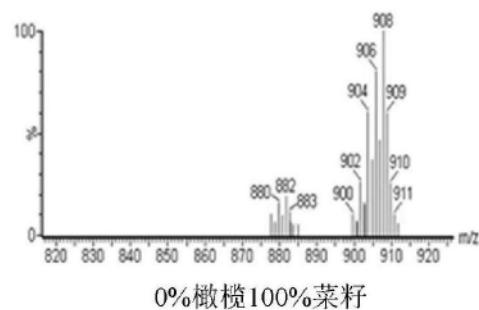


图3a

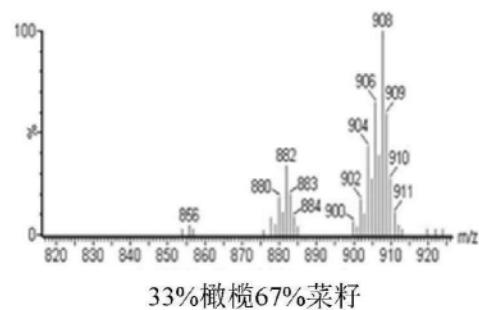


图3b

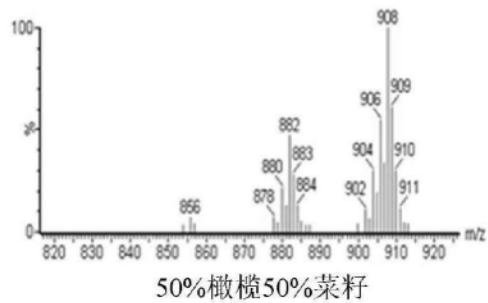


图3c

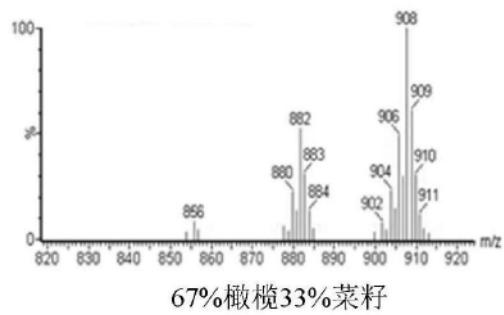


图3d

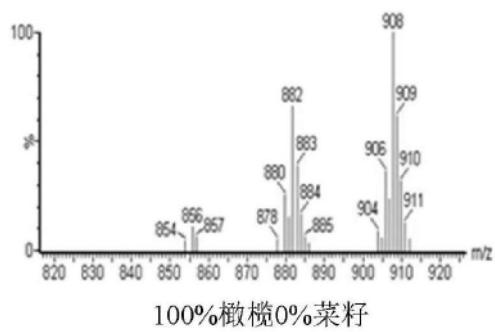


图3e

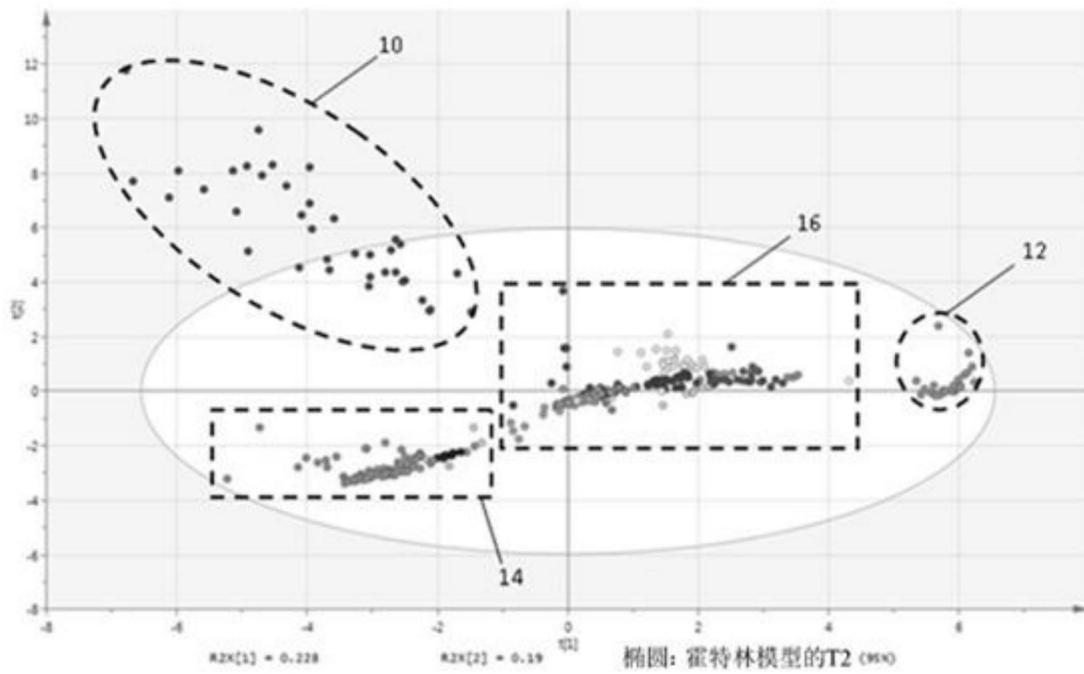


图4a (现有技术)

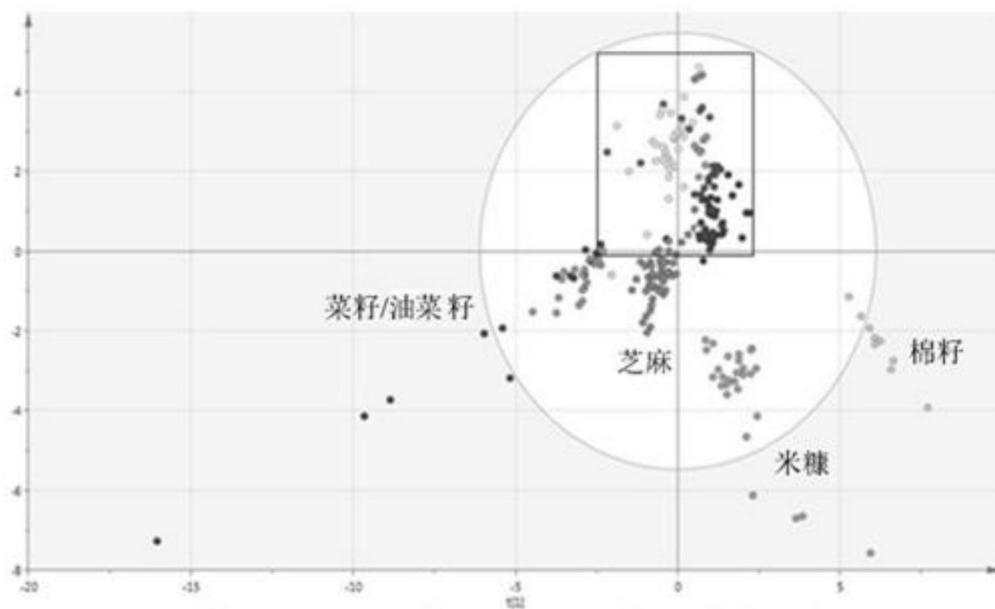
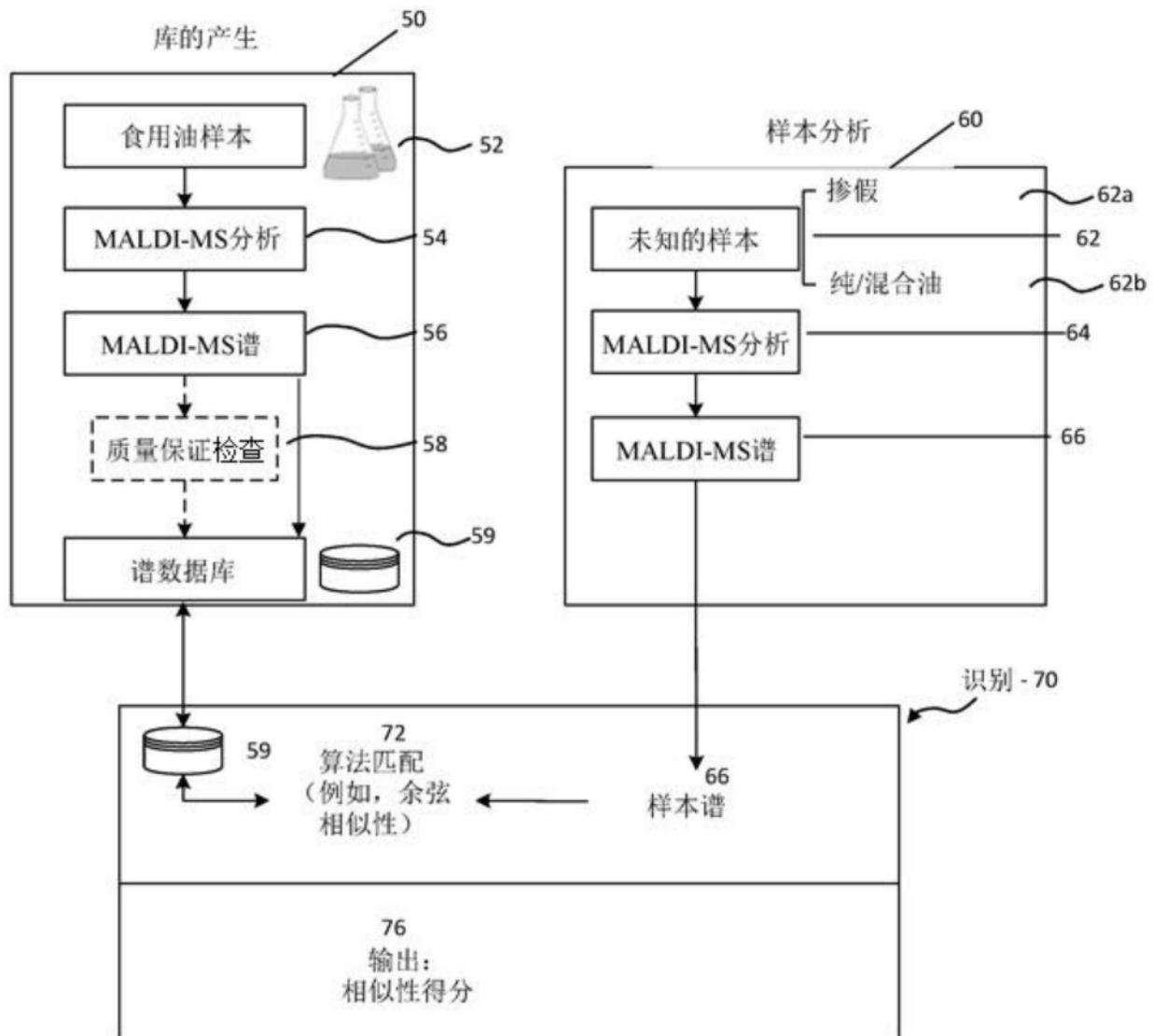


图4b (现有技术)



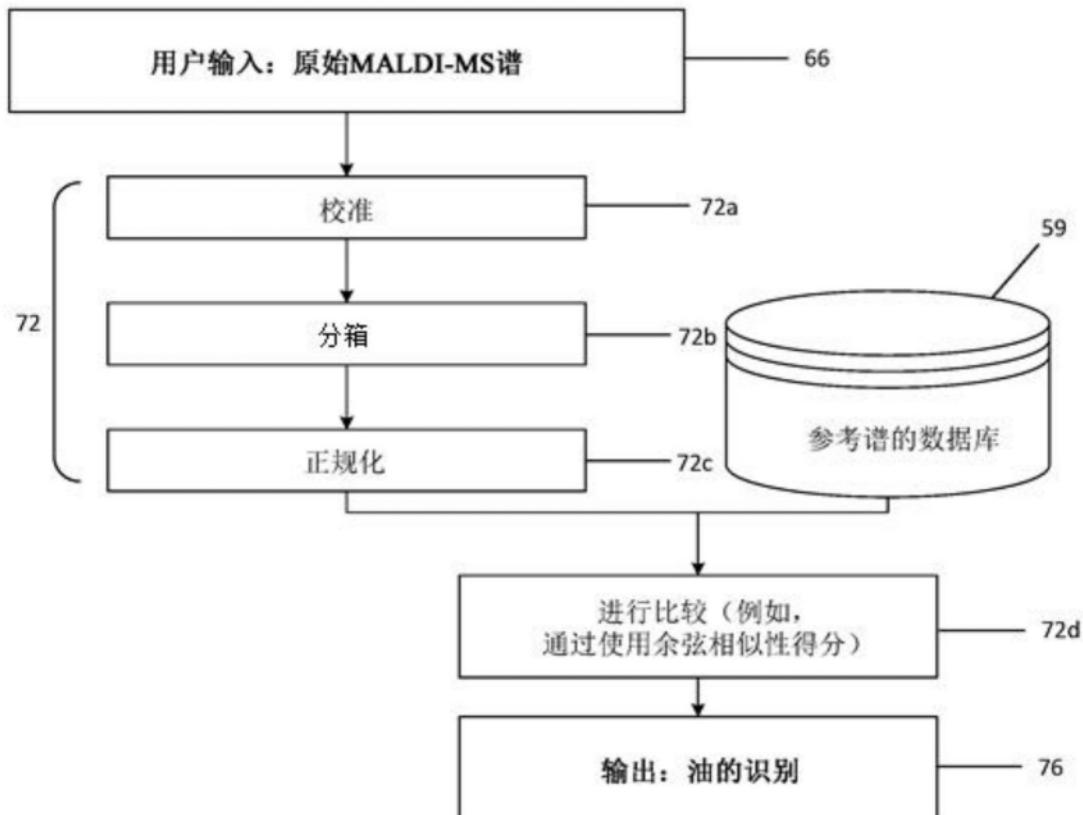


图5b

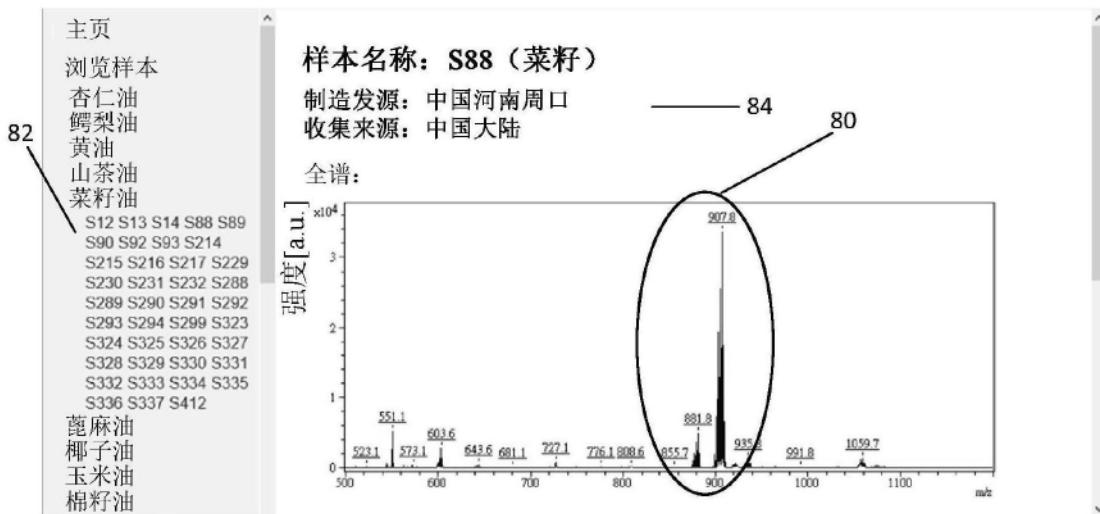


图6a

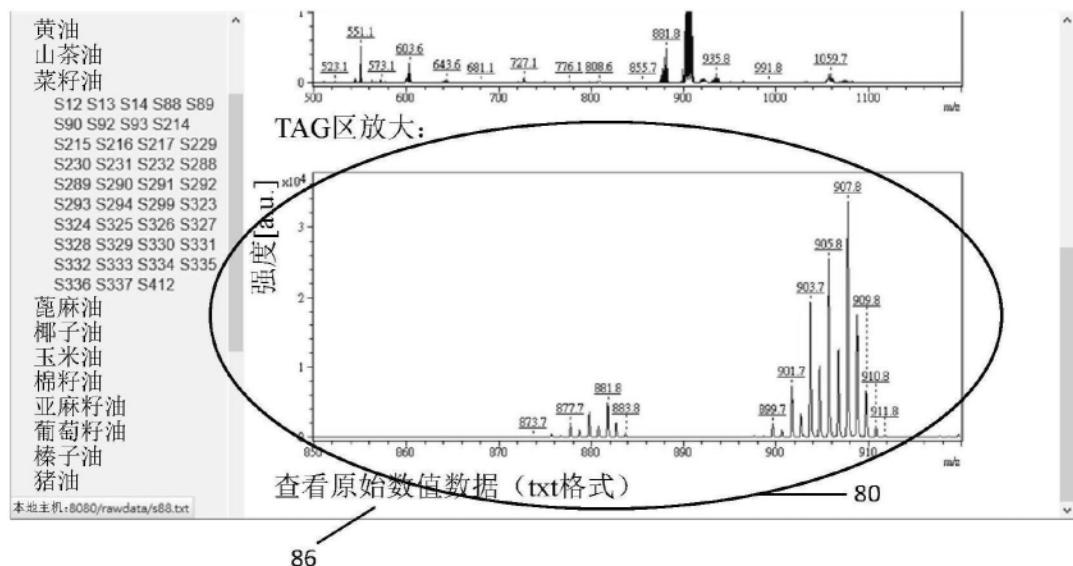


图6b



图7a

主页  
油验证  
浏览油数据库  
杏仁油  
鳄梨油  
黄油  
山茶油  
菜籽油  
蓖麻油  
椰子油  
玉米油  
棉籽油  
鱼油  
亚麻籽油  
葡萄籽油  
HAT-GWO油  
榛子油  
高油酸

油验证

请上载原始MALDI-MS数据文件:

将文件拖动到此处  
或  
点击打开文件浏览器

20170411 s470 3\_0\_G12\_1.txt  
状态: 上载完成 匹配=黄油; 相关性得分 = 0.9410 ————— 92

20170411 s486 1\_0\_H6\_1.txt  
状态: 上载完成 匹配=南瓜籽; 相关性得分 = 0.9590 ————— 94

图7b

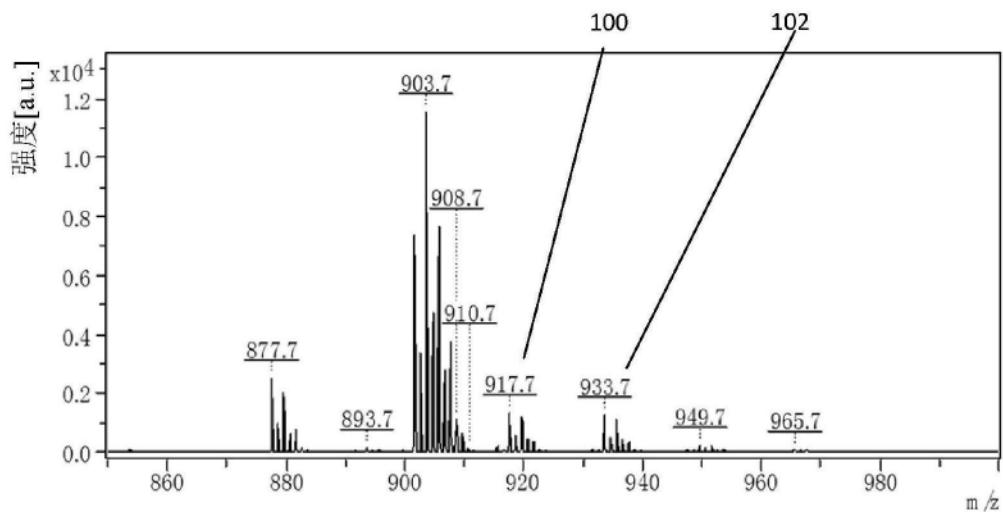


图8a

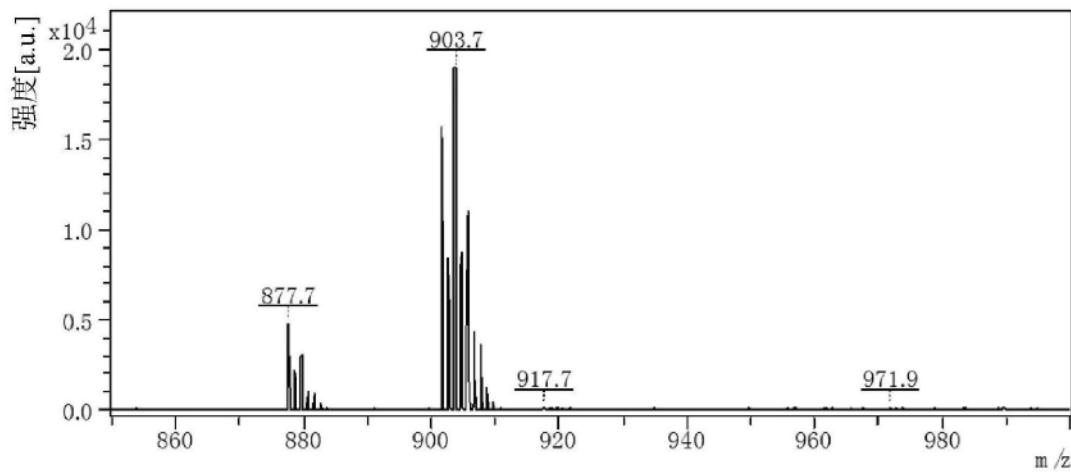


图8b

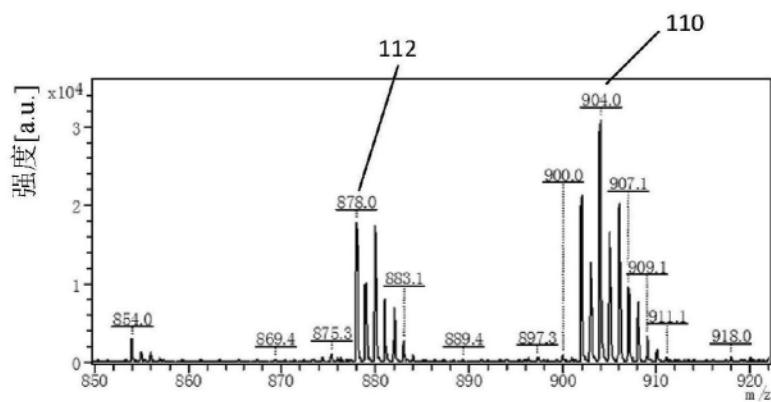


图9a

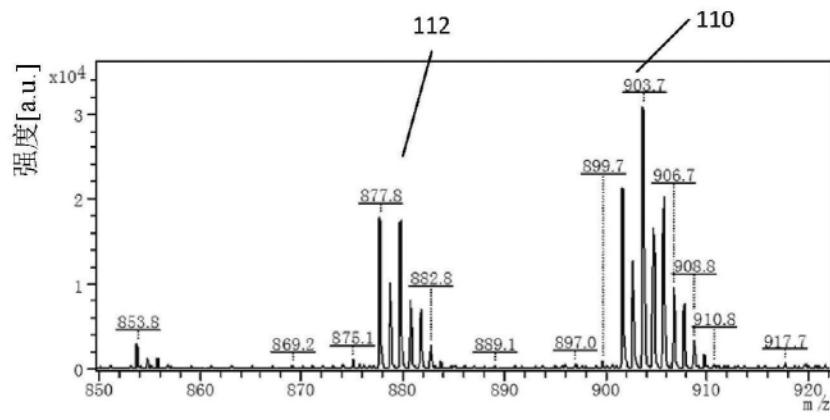


图9b

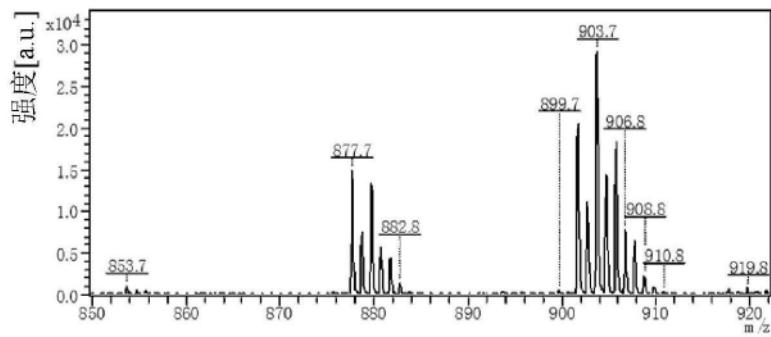


图9c

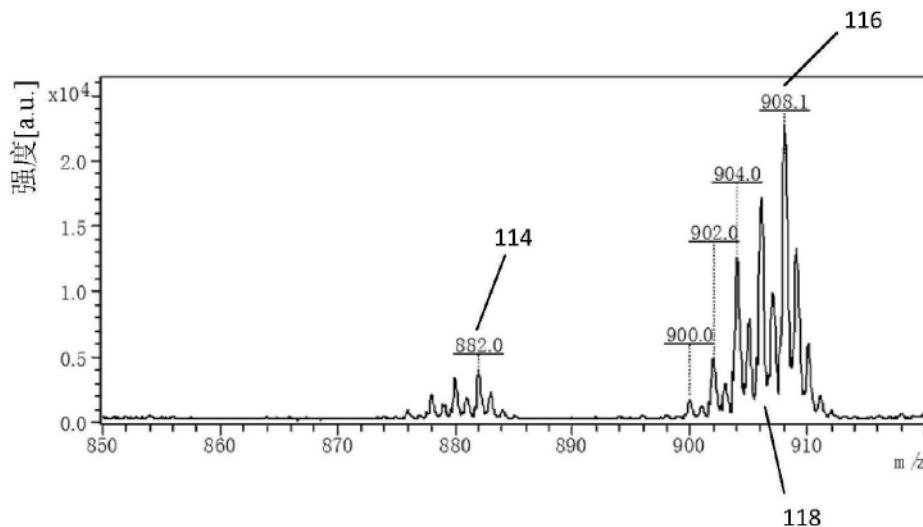


图9d

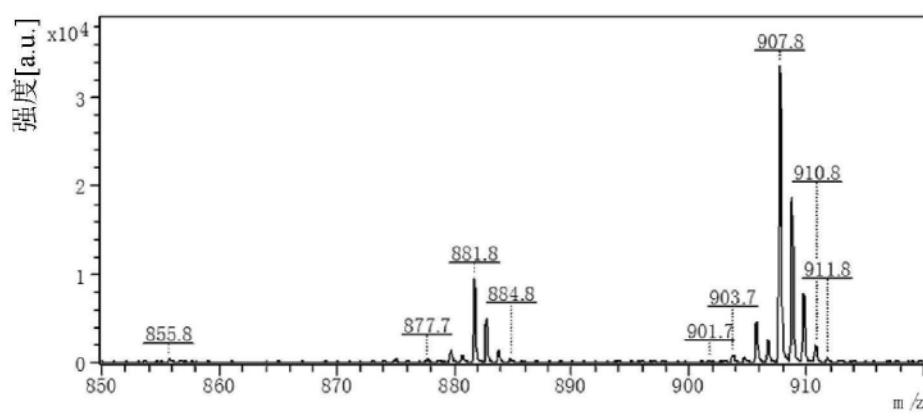
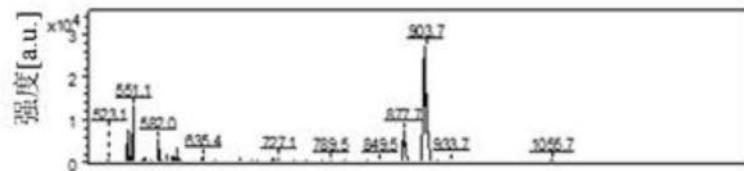
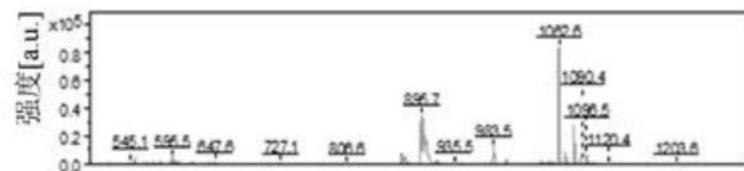


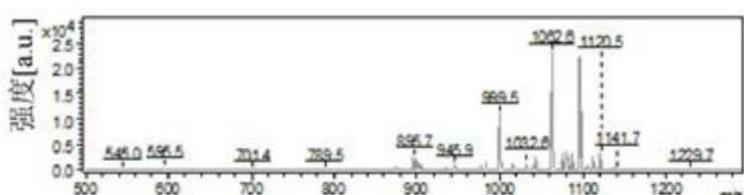
图9e



(i) 有问题的样本 (声称为亚麻籽油)



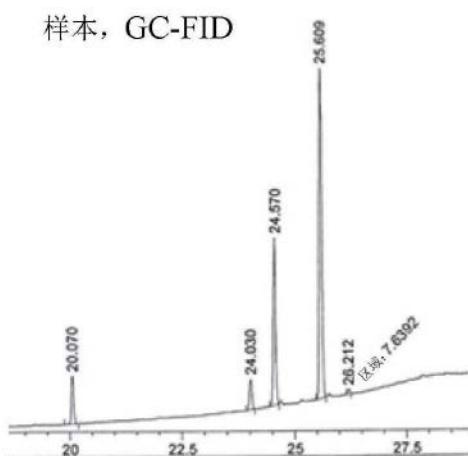
(ii) 典型的亚麻籽油1



(iii) 典型的亚麻籽油2

图10a

样本, GC-FID



脂肪酸	%组成	
	样本	亚麻籽*
C16:0	7.7	4.0-6.0
C18:0	5.1	2.0-3.0
C18:1	27.7	10.0-22.0
C18:2	58.6	12.0-18.0
C18:3	N.D.	56.0-71.0

\*[https://en.wikipedia.org/wiki/Linseed\\_oil#cite\\_note-DGF-16](https://en.wikipedia.org/wiki/Linseed_oil#cite_note-DGF-16)

图10b