

[19] 中华人民共和国国家知识产权局

[51] Int. Cl.
C12Q 1/68 (2006.01)



[12] 发明专利说明书

专利号 ZL 98803287.2

[45] 授权公告日 2008 年 11 月 19 日

[11] 授权公告号 CN 100434531C

[22] 申请日 1998.1.15 [21] 申请号 98803287.2
[30] 优先权

[32] 1997.1.15 [33] GB [31] 9700746.2
[32] 1997.8.28 [33] GB [31] 9718255.4
[32] 1997.12.19 [33] GB [31] 9726953.4

[86] 国际申请 PCT/GB1998/000127 1998.1.15

[87] 国际公布 WO1998/031830 英 1998.7.23

[85] 进入国家阶段日期 1999.9.13

[73] 专利权人 X 齐里昂有限两合公司
地址 德国美因河畔法兰克福

[72] 发明人 G·施密特 A·H·汤普森
R·A·W·约翰斯通

[56] 参考文献
CN1131440A 1996.9.18
审查员 林峻凯

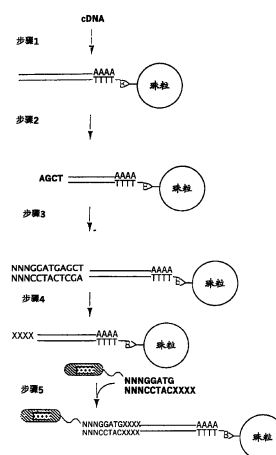
[74] 专利代理机构 上海专利商标事务所有限公司
代理人 陈文青

权利要求书 4 页 说明书 34 页 附图 30 页

[54] 发明名称
与质量标记连接的杂交探针

[57] 摘要

一种杂交探针阵列，每个探针包含与既定长度的已知碱基序列相连的质量标记，其中阵列的每个质量标记(任选的与已知碱基序列在一起)通过质谱法与该碱基序列关联。



1. 杂交探针在通过质谱法分析质量标记来测定探针杂交的方法中的应用, 该探针包含与既定长度的已知碱基序列相连的质量标记, 其特征在于, 质量标记与碱基序列可断裂性相连, 并通过碰撞诱导分解从探针上断裂下来。

2. 杂交探针阵列在通过质谱法分析质量标记来测定探针杂交的方法中的应用, 其中每个杂交探针包含与既定长度的已知碱基序列相连的质量标记, 其特征在于, 每个质量标记与其各自的已知碱基序列可断裂性相连, 其中质量标记通过碰撞诱导分解从杂交探针上断裂下来。

3. 根据权利要求 2 所述的应用, 其中每个质量标记对于阵列中其它各质量标记来说均是可唯一识别的。

4. 根据权利要求 1 或 2 所述的应用, 其中碱基序列的既定长度为 2 至 25。

5. 根据权利要求 1 或 2 所述的应用, 其中这个或各个质量标记通过可碰撞断裂的接头与已知碱基序列可断裂性相连。

6. 根据权利要求 1 或 2 所述的应用, 其中接头在质谱仪中断裂。

7. 根据权利要求 1 或 2 所述的应用, 其中接头在质谱仪的电离室内断裂。

8. 根据权利要求 1 或 2 所述的应用, 其中这个或各个质量标记通过接头与已知碱基序列可断裂性相连, 该接头对电子电离的稳定性比质量标记差。

9. 根据权利要求 1 或 2 所述的应用, 其中这个或各个质量标记在电离条件下带负电荷。

10. 根据权利要求 1 或 2 所述的应用, 其中这个或各个质量标记可在质谱仪中从已知碱基序列中分辨出来。

11. 根据权利要求 1 或 2 所述的应用, 其中这个或各个质量标记对于 50 伏的电子电离稳定。

12. 根据权利要求 1 或 2 所述的应用, 其中这个或各个已知碱基序列包含衔接寡核苷酸的粘端, 该寡核苷酸含有限制性内切核酸酶的识别位点, 该内切酶在距识别位点的既定距离处切割。

13. 根据权利要求 1 或 2 所述的应用, 其中这个或各个已知碱基序列连接了多个相同的质量标记。

14. 根据权利要求 1 或 2 所述的应用, 其中它用于在聚合酶链反应或连接酶链反应中测定探针的杂交。

15. 根据权利要求 1 或 2 所述的应用, 其中方法是专有的在线形式。

16. 一种测定探针阵列与靶核酸杂交的方法，该方法包括：

(a)在探针与靶核酸杂交的条件下使靶核酸与阵列的每个杂交探针接触，任选地除去未杂交物质，其中每个探针含有与既定长度的已知碱基序列连接的质量标记，阵列的每个质量标记与其各自的已知碱基序列可断裂性相连；和

(b)用碰撞诱导分解将质量标记从杂交的探针上断裂下来，以释放质量标记；和

(c)通过用质谱法鉴定释放的质量标记来鉴定杂交的探针。

17. 根据权利要求 16 所述的方法，其中每个质量标记对于阵列中其它各质量标记来说均是可唯一识别的。

18. 一种测定探针与靶核酸杂交的方法，该方法包括

(a)在探针与靶核酸杂交的条件下使靶核酸与杂交探针接触，该杂交探针含有与既定长度已知碱基序列相连的质量标记，任选地除去未杂交的物质，其中质量标记与其各自的已知碱基序列可断裂性相连；和

(b) 用碰撞诱导分解将质量标记从杂交的探针上断裂下来，以释放质量标记；和

(c) 通过用质谱法鉴定释放的质量标记来鉴定杂交的探针。

19. 根据权利要求 16 或 18 所述的方法，其中用电喷射电离或基质辅助激光解吸附电离来制备质谱法的这个或各个样品。

20. 根据权利要求 16 或 18 所述的方法，其中碱基序列的既定长度为 2 至 25。

21. 根据权利要求 16 或 18 所述的方法，其中这个或各个质量标记通过可碰撞断裂的接头与已知碱基序列可断裂性相连。

22. 根据权利要求 16 或 18 所述的方法，其中接头在质谱仪内断裂。

23. 根据权利要求 16 或 18 所述的方法，其中接头在质谱仪的电离室中断裂。

24. 根据权利要求 16 或 18 所述的方法，其中每个质量标记通过接头与其各自的已知碱基序列可断裂性相连，该接头对电子电离的稳定性比质量标记差。

25. 根据权利要求 16 或 18 所述的方法，其中每个质量标记在电离条件下带负电荷。

26. 根据权利要求 16 或 18 所述的方法，其中质量标记和已知碱基序列在进入质谱仪之前没有分开。

27. 根据权利要求 16 或 18 所述的方法，其中质量标记对于 50 伏的电子电离稳定。

28. 根据权利要求 16 或 18 所述的方法, 其中已知碱基序列包含衔接寡核苷酸的粘端, 该寡核苷酸含有限制性内切核酸酶的识别位点, 该内切酶在距识别位点既定距离处切割。

29. 根据权利要求 16 或 18 所述的方法, 其中已知碱基序列与多个相同的质量标记相连。

30. 根据权利要求 16 或 18 所述的方法, 其中该方法在线进行。

31. 权利要求 2 所述的杂交探针阵列在读取寡核苷酸芯片中的应用, 其中质量标记通过碰撞诱导分解从探针上断裂下来。

32. 权利要求 2 所述的杂交探针阵列在竞争性结合试验中鉴定寡核苷酸结合剂的应用, 其中质量标记通过碰撞诱导分解从探针上断裂下来。

33. 权利要求 2 所述的探针阵列在聚合酶链反应或连接酶链反应中探测既定序列的应用, 其中质量标记通过碰撞诱导分解从探针上断裂下来。

34. 一种对 cDNA 作特性分析的方法, 该方法包括:

(a)用限制性内切核酸酶切割一群含有一种或多种 cDNA 的样品, 分离出携带有 cDNA 一个末端的片段, 其限制性位点在毗邻 cDNA 末端的参照位点处;

(b)用第一取样内切核酸酶在距参照位点已知距离的第一取样位点处切割分离的片段, 产生第一和第二亚片段, 每个亚片段含有既定长度且序列未知的粘端序列, 第一亚片段有 cDNA 的该末端;

(c)将第一或第二亚片段根据其粘端序列分类成亚群, 记录下每个亚群的粘端序列作为第一粘端;

(d)用第二取样内切核酸酶在第二取样位点切割每个亚群中的亚片段以产生各亚片段的次亚片段, 所述的第二取样内切核酸酶与第一取样内切核酸酶相同或不同, 所述的第二取样位点距第一取样位点已知距离, 所述的次亚片段包含具有既定长度、序列未知的第二粘端序列; 和

(e)测定各个第二粘端序列;

其中每个亚片段的第一和第二粘端序列的总长度为 6 至 10, 参照位点和第一及第二粘端的序列和相对位置代表了该 cDNA 或各个 cDNA 的特征, 第一取样内切核酸酶与第一识别位点结合, 并在距该限制性内切核酸酶限制性位点既定距离的第一取样位点处切割, 且其中第一和/或第二识别位点提供在权利要求 12 所述杂交探针阵列的第一和/或第二衔接寡核苷酸中, 并与分离片段的限制性位点杂交;

其中第一和/或第二粘端序列的测定是, 通过碰撞诱导分解使质量标记从第一

和/或第二衔接寡核苷酸上断裂下来，用质谱法鉴定断裂下来的质量标记。

35. 一种对核酸进行测序的方法，该方法包括：

(a)获得包含核酸片段的靶核酸群，其中每个片段以唯一量存在，并且一端携带既定长度且序列未知的粘端序列，

(b)保护每个片段的另一端，和

(c)通过下列方式对每个片段测序，

(i)使片段在连接酶存在时、杂交条件下与权利要求 12 所述的杂交探针阵列接触，阵列的碱基序列的既定长度与粘端序列相同，阵列含有该既定长度所有可能的碱基序列；除去所有连接的衔接寡核苷酸，通过碰撞诱导分解释放质量标记和用质谱法鉴定释放的质量标记来记录所有连接的衔接寡核苷酸的量；

(ii)使连接的衔接寡核苷酸与测序酶接触，该测序酶与识别位点结合并切割片段显露出新的粘端序列，该粘端与前一粘端序列毗邻或重叠；和

(iii)重复步骤(i)和(ii)足够多次，通过比较各粘端序列记录的量来确定片段序列。

与质量标记连接的杂交探针

发明领域

本发明涉及杂交探针阵列，杂交探针的用途，测定该探针阵列杂交的方法以及对 cDNA 作特性分析和对核酸测序的方法。

发明背景

质谱法是测定分子质量的高度灵敏的方法，它非常灵敏，能提供详细的结构信息。本质上是使待分析的分子汽化并电离到真空中。气相离子通过电磁场加速，其质荷比通过分析电磁场中的分子行为来确定。存在各种质谱技术，它们由系统的主要目标或由它们采用的各种电离方法来确定。总之，质谱法是用来直接分析分子以便确定其质量；鉴定它们或获得结构信息。(关于质谱法的教科书参见参考文献 1)。

组合化学(关于该领域的回顾参见参考文献 2)导致了对间接分析分子更特殊的要求。现已存在各种方法，它们以组合方式用固相合成技术来生产大量相关分子。由于大多数系统在各珠粒上产生单个分子，因此它们可根据所需性能来筛选。然而，通常的情况是不能直接回收所筛选的分子，或由于其它原因很难进行直接分析，因此建议用间接标记的珠粒及其分子作为解决手段。用于“编码”(参见参考文献 3)组合文库的大多数方法好象涉及采用在某种意义下能被“测序”(参见参考文献 4)的标记，例如常用氨基酸和核酸来编码文库，因为对于短肽和寡核苷酸来说，对它们进行测序的方法是常规的，且相对快速，最近也常用质谱法来进行分析。其它有机实体(如卤化苯和仲胺)是可测序的，它们可用于这些目的(参见参考文献 5 和 6)。

另一种方法(参见参考文献 7)采用各种富含特定同位素的组合单体来产生在质谱中提供独特同位素特征(signature)的标记。该方法能产生大量标记，这些标记在质谱的限定区域内有独特方式的同位素峰。这种方法对于特异性地鉴定一种化合物(化合物的珠粒已从大的组合文库中分离出来)来说是理想的，但是要同时分辨大量分子总存在一些问题。

参考文献 15 至 17 公开了用质谱法检测不同配体结合的应用。

发明概述

本发明提供一种杂交探针的阵列，每个探针包含一个与既定长度的已知碱基序列相连的质量标记，其中阵列中每个质量标记(任选地和已知碱基序列在一起)，可通过质谱法与此碱基序列相关联。较佳的，每个杂交探针包含一个与既定长度的已知碱基序列可断裂性相连的质量标记，其中当阵列的每个质量标记从各自碱基序列中释放下来时，可通过质谱法来与该碱基序列相关联，通常是通过质荷比，该质荷比相对于阵列中所有其它质量标记来说最好是可唯一鉴定的。

本发明还提供了杂交探针的用途，该探针包含与既定长度的已知碱基序列相连的质量标记，其用法是通过质量标记以及任选的已知碱基序列的质谱来测定探针的杂交。较佳的，杂交探针包含一个与既定长度的已知碱基序列可断裂性相连的质量标记。

本发明还提供了一种测定探针与靶核酸杂交的方法，该方法包括：

- (a)在使探针与靶核酸杂交的条件下使靶核酸接触杂交探针，该探针包含一个与既定长度的已知碱基序列相连的质量标记，以及任选地除去未杂交的物质；和
- (b)用质谱法鉴定探针。

本发明还提供了一种测定探针阵列与靶核酸杂交的方法，该方法包括

- (a)在探针与靶核酸杂交的条件下使靶核酸与阵列中的每个杂交探针接触，以及任选地除去未杂交的物质，其中每个探针包含一个与既定长度的已知碱基序列相连的质量标记；和
- (b)用质谱法鉴定探针。

较佳的，这个或每个质量标记与其各自己知碱基序列可断裂性相连，每个杂交的探针断裂时释放出质量标记，用质谱法鉴定释放出的标记。

碱基序列的既定长度通常为 2 至 25。

每个质量标记可通过一个接头与已知碱基序列可断裂性相连，该接头可以是可光断裂的接头、可化学断裂的接头或可热断裂的接头。根据一个实例，接头在质谱仪内断裂，例如在质谱仪的电离室内断裂。其优点是不需要使接头在质谱仪外断裂。通过适当地选择接头，可实现质谱仪中的断裂，以使已知碱基序列与质量标记迅速分离，从而很容易鉴定质量标记。接头对于电子电离的稳定性最好比质量标记差。这就可使接头在质谱仪内断裂，而质量标记的任何部分不产生碎片。

在一个较佳的实例中，质量标记在 50 伏下(较佳的在 100 伏下)对电子电离稳定。质谱仪内发生电子电离的条件可使分子产生碎片，因此可方便地根据在特定

电压下其抵抗电子电离的能力来测定质量标记的稳定性。对电子电离的稳定性也是显示分子在质谱仪中碰撞诱导解离条件下稳定性的一个有用的指南。

较佳的，在质谱法中可从已知碱基序列中分辨出质量标记。这是有利的，因为这样就不必将每个质量标记与其各自的碱基序列分开或纯化出来。因此，在一个较佳的实例中，质量标记和已知碱基序列在进入质谱仪之前没有分开。

在另一个较佳的实例中，该方法是专门在线(on-line)的。“在线”是指该方法中没有一个阶段有离线进行的步骤。这是有好处的，因为该方法可作为连续方法而进行，且容易自动化。

在一个实例中，每个质量标记被设计成在电离条件下带负电荷。其优点是可以在安排缓冲液的条件使伴随质量标记的核酸带正电。这就使得质量标记易在质谱仪中与 DNA 分开，并且在质谱中产生较少的背景噪扰。

较佳的，已知碱基序列与多个相同质量的标记相连。采用多个相同质量的标记的优点是多个质量标记同时断裂会产生更高的信号，因为可测得的质量标记的浓度更高。

在一个实例中，已知碱基序列包含一个衔接寡核苷酸的粘端，它含有限制性内切核酸酶的识别位点，该内切酶在距识别位点既定距离处切割。

本发明提倡用质谱性能表现良好的标记在单个质谱中鉴定出相当大量的分子。“表现良好”指，通过防止多种电离状态和不采用特别不稳定的基团最大程度地减少分子在质谱中产生的峰数。有机化学中的几十年质谱法经验已经确定某些分子特征有利于这些应用，而某些特征则应予避免。

分析标记分子的质谱法：

可以用“质量”作为分子身份的指示物来标记分子，尤其是生物分子。使分子质量与其身份相关的密码很容易产生，例如，若需要鉴定的一组分子，可为待鉴定的每个不同的分子简单地选择渐增的质量。显然，许多分子单单根据其质量就可被鉴定，作标记看上去可能是多余的。但是在一些情况下，某些分子(尽管是唯一的)可能具有接近的质量并且可多次电离，这就使质谱仪中的分辨很困难，因而应采用质量标记。对于经常是同质异位但仍不同的核酸(例如序列 TATA 与 TTAA、TAAT 等不同，但在质谱仪中很难分辨)尤其希望采用质量标记。另外，可能希望待鉴定的分子可执行某种功能并且可被检测到，这就意味着不能直接检测，因此一种可被独立检测到的可去除的标记非常有用。这样就能同时分析大量非常相似的分子，以达到大规模筛选的目的。

本发明描述了用质量标记文库鉴定共价连接的核酸探针的序列。对于合格的有机化学家来说，质量标记的构建相对简单。这使得很容易生产这样一些标记，这些标记可以可控制地从其各自探针上除去，并具有好的物理性能，有助于电离进入质谱仪，有助于在大范围内、相当大量的标记中检测和分辨多个标记。

现在仅通过实施例并参照附图来进一步详细描述本发明，其中：

图 1a 和 1b 显示了本发明的带质量标记的杂交探针用于基因表达描述的方法；

图 2a 和 2b 显示了本发明的带质量标记的杂交探针用于基因表达描述的另一种方法；

图 3a 和 3b 显示了本发明的带质量标记的杂交探针用于基因表达描述的另一种方法；

图 4 显示了适用于本发明的正交飞行时间(orthogonal time of flight)质谱仪的示意图；

图 5 显示了适用于本发明的可光断裂的接头；

图 6 显示了生产用于本发明带质量标记的碱基的反应方案；

图 7 显示了适用于本发明的可片段化接头；

图 8 显示了用于本发明的质量标记的结构；

图 9 显示了用于本发明的可变基团和质量系列修饰基团；

图 10 显示了适用于本发明的溶解性和携带电荷的基团；

图 11 显示了模型化合物 AG/1/75 在负离子模式中的质谱；

图 12 显示了模型化合物 AG/1/75 在正离子模式中的质谱；

图 13 显示了模型化合物 AG/1/75 在正离子模式中的另一种质谱；

图 14 和 15 显示了各种缓冲液中的 PCR 产物在正离子和负离子模式中的质谱；

图 16 和 17 显示了带 AG/1/75 的 PCR 产物在负离子和正离子模式中的质谱；

图 18 和 19 显示了带 AG/1/75 的 PCR 产物在信号加工后的质谱；

图 20 和 21 显示了带质量标记的碱基 FT23 在负离子和正离子模式中的质谱；

图 22 和 23 显示了有寡核苷酸背景的 FT23 在负离子和正离子模式中的质谱；

图 24 显示了本发明中带质量标记的碱基 FT9 和 FT17；和

图 25 显示了本发明中带质量标记的碱基 FT18 和 FT23。

质量标记技术的应用:

主要有两种质谱电离技术是生物分析中通常采用的。它们是电喷射质谱法(ESMS)和 MALDI TOF 质谱法。ESMS 本质上是允许从液相电离成气相的技术,而 MALDI 本质上是允许从固相电离成气相的技术。许多分子生物学试验在液相中进行,或采用固相化学原理在液体基质中进行,通过该基质可以加入试剂和从固定在固相支持物上的分子中除去试剂。从某种意义上讲,这两种技术是互补的,因而能够分析固相和液相的元素。

用带质量标记的衔接分子描述基因:

参考文献 8 中描述的基因描述方法提供了这样一种方法,采集该细胞群中每个 cDNA 样品来分析细胞中的基因表达方式。根据本专利申请,提供了一种对 cDNA 进行特性分析的方法。该方法包括:

(a)用第一取样内切核酸酶在第一取样位点切割一群含一种或多种 cDNA 或其分离片段的样品,以产生各 cDNA 或其分离片段的第一和第二亚片段,每个 cDNA 或其分离片段都包含 cDNA 的一个末端(如 poly-A 尾序列),所述第一取样位点距 cDNA 末端最近的参照位点已知距离,所述第一和第二亚片段各包含既定长度但序列未知的粘端序列,第一亚片段具有 cDNA 的末端;

(b)根据粘端序列将第一或第二亚片段分类成亚群,并记录各亚群的粘端序列作为第一粘端;

(c)在每个亚群中用第二取样内切核酸酶在第二取样位点切割亚片段以产生各亚片段的次亚片段,所述的第二取样内切酶与第一取样内切核酸酶相同或不同,所述的第二取样位点相距第一取样位点已知距离,所述的次亚片段包含具有既定长度、序列未知的第二粘端序列;和

(d)测定各第二粘端的序列;

其中各亚片段的第一和第二粘端序列长度之和为 6 至 10;参照位点和第一及第二粘端的序列和相对位置代表了该 cDNA 或各个 cDNA 的特征。

用第一取样内切核酸酶切割的样品宜包含如下所得的 cDNA 的分离片段,即用限制性内切核酸酶切割一群含一种或多种 cDNA 的样品,并分离限制性位点在参照位点上的片段。

第一取样内切核酸酶宜与第一识别位点结合,并在距限制性内切核酸酶的限制性位点既定长度的第一取样位点处切割。根据本发明的这个方面,第一识别位

点提供在上述第一个带质量标记的衔接寡核苷酸中，它与分离片段的限制性位点杂交。根据本发明，各亚片段的第一和第二粘端序列的长度之和最好是 8。

在一个实例中，取样系统在群体每个 cDNA 中取出两个 4bp 的样品，并根据确定的参照点确定其序列。为了实现这一点，固定群体中的每个 cDNA，并可用限制性内切核酸酶断裂。将一个衔接子连接到所得的已知粘端上。衔接子被设计成携带了 II 型限制性内切核酸酶的结合位点。用 II 型限制性内切核酸酶在群体中每个 cDNA 的衔接子端显露出多义(ambiguous)的 4bp 粘端。用一组衔接子分子来探测那四个外露的碱基。如参考文献 8 中所讨论的，采用荧光为基的系统来探测一群 cDNA 时，一次只能加入 256 种可能的探针分子中的四种。显然，对于测定 4 碱基对序列来说，这是一种很慢的方法。采用了质量标记的衔接子，就可在一群外露的 cDNA 中同时加入所有 256 种可能的 4bp 衔接子，这大大加快了描述基因的发明进程。这是商业上可行技术所必需的。

可使这种系统相容于 ESMS。在基因描述发明中，cDNA 群体根据 II 型限制性内切核酸酶产生的外露序列而被分成 256 个亚组。这种分类产生了在 256 个孔中 cDNA 的 256 个群体。通过 II 型限制性内切核酸酶的第二次断裂，每个 cDNA 的第二个 4bp 序列可能外露，然后可通过与带质量标记的衔接子连接来确定这 4 个碱基。

以质谱法为基础的寡核苷酸芯片(chip)读数器(MALDI):

寡核苷酸阵列:

利用在平板式固相基质(如玻片)上合成的寡核苷酸阵列，可以进行各种核酸测定。该阵列通常是这样构成的，将玻片分成不同的区域，使每个区域中只有一个寡核苷酸。通过测定阵列的每个区域来自标记核酸的信号，以确定标记的核酸与阵列是否杂交。mRNA 水平的测定可用许多方式来实现。用逆转录法容易将具有 poly-A 的 mRNA 转变成 cDNA。逆转录酶 PCR(RTPCR)方法能定量地测定待测的单个 RNA，但是精度水平相对低。寡核苷酸阵列对于核酸分析来说是一种相当新的方法，它能分析突变，通过杂交来测序，并能分析 mRNA 的表达。构建该阵列的方法已经被开发出来(例如参见参考文献 9、10、11)，预计还可有其它方法。

上述类型的标记的核酸与寡核苷酸阵列的杂交通常用荧光标记来检测。寡核苷酸或 cDNA 的阵列可用经荧光标记物标记的核酸来探测。对于寡核苷酸芯片，通过含有与标记核酸杂交的寡核苷酸的阵列区域中显现荧光，就可揭示哪个寡核苷酸与标记的核酸互补。如果用质量标记来标记与寡核苷酸阵列杂交的核酸，那

么就可用 MALDI 质谱法来读取该寡核苷酸阵列。质量标记最好通过可光断裂的接头与其相应的核酸连接。这些质量标记可在它们的结构中引入激光可激发的因素,或者可在进行杂交反应后,用合适的解吸附剂(如 3-羟基甲基吡啶酸)处理寡核苷酸阵列。带质量标记的核酸一旦与芯片杂交,就可通过施加合适频率的激光使质量标记和核酸间的接头断裂。然后,用合适频率的激光扫描那些区域,就可从寡核苷酸阵列的特定区域中解离出标记。然后,根据从阵列区域中解离出的标记的质量就可以确定寡核苷酸阵列特定区域处的杂交核酸的身份。

这种方法优于荧光为基础的系统的优点单单在于可获得的标记数目。荧光为基础的技术受到光谱重叠问题的严重限制,从而限制了同时用于荧光为基础的读数器可产生的染色色泽数。用质谱法作为标记检测系统可以产生非常大量的质量标记。

寡核苷酸阵列可直接用于参考文献 12 公开的基因描述技术。在表面上确定位点处携带了所有 256 种可能的 4 碱基寡核苷酸的阵列,可用于实现上述发明所要求的分类步骤。为了使该芯片为基础的描述系统与质谱分析相容,衔接子上用于测定第二个 4 碱基样品序列的标记应与 MALDI 相容,这样就可用 MALDI 质谱仪中的紫外激光扫描寡核苷酸芯片。这就能用取自单一固定来源的单份 DNA 样品测定群体中每个 cDNA 的 8 碱基签名(signature),并对其在一系列激光扫描中进行分析。一组标记从某芯片区解离出来,此芯片区就确定了签名的头 4 个碱基,而标记的组成确定了签名的第二组 4 碱基和各 cDNA 的相对量。

用液相层析质谱法描述基因:

基因描述方法以两步过程来操作:分子签名分类,然后分析与分类签名连接的探针分子。MALDI 方法通过寡核苷酸阵列来实现签名的分类。阵列的另一种用途是亲和层析。亲和柱根据既定长度的多义性粘端对签名分类。为了用 4bp 的多义性粘端对签名分类,可用位于粘端的 256 种可能的 4 聚物对适用于 HPLC 形式的珠粒进行衍生处理。该柱中可加入签名,该签名溶解在缓冲液中,这有利于 4 聚物杂交到经衍生的珠粒上。这将推动杂交反应的平衡向有利于杂交的方向移动。然后用浓度逐渐增加的抑制杂交的缓冲液洗柱。以 AAAA 或 TTTT 粘端结尾的签名会首先释放,而 GGGG 和 CCCC 签名将最后释放。为了确保彼此互补签名的分离,可用碱基类似物来衍生处理珠粒,使签名中的鸟嘌呤对珠粒上胞嘧啶的杂交亲和力与签名粘端中的胞嘧啶对珠粒上鸟嘌呤的杂交亲和力不同。另外,可保证珠粒上存在的各种 4 聚体的浓度相互不同。

该亲和柱能根据其多义性粘端序列将一群签名分类成 256 种组分。然后可将

这些组分直接加入电喷射质谱仪中进行分析。

用带质量标记的衔接分子进行 DNA 测序:

参考文献 13 中描述了一种测序技术, 其中提供了一种核酸测序方法, 该方法包括:

(a) 获得一群包含核酸片段的靶核酸, 其中每个片段以唯一量存在, 一端携带既定长度但序列未知的粘端效率,

(b) 保护每个片段的另一端, 和

(c) 用以下方法对每个片段测序,

(i) 在一轮循环中使片段与衔接寡核苷酸阵列接触, 每个衔接寡核苷酸具有一个标记、测序酶识别位点以及既定长度与粘端序列相同的已知的独特碱基序列, 该阵列含有既定长度的所有可能的碱基序列; 其中循环步骤包括在杂交条件和连接酶存在下, 使阵列的每个衔接寡核苷酸依次接触核酸片段, 取出所有连接的衔接寡核苷酸, 通过检测标记记录已连接的衔接寡核苷酸的量, 然后重复该循环直至阵列中所有衔接子被测定;

(ii) 使连接的衔接寡核苷酸与测序酶接触, 该测序酶与识别位点结合并切割片段显露出新的粘端序列, 该粘端与前一粘端序列毗邻或重叠;

(iii) 重复步骤(i)和(ii)足够多次, 比较各粘端序列记录的量来确定片段序列。较佳的, 粘端碱基序列的既定长度为 3 至 5。如上所述, 根据本发明, 每个衔接寡核苷酸携带了一个质量标记。这在原理上与参考文献 8 中描述的基因描述系统相似, 因为在一个重复的循环中, DNA 分子是固定的, 在其末端具有 II 型限制性内切核酸酶所显露的 4 碱基序列。它们也用衔接分子来探测, 因此由于与基因描述相同的理由, 采用带质量标记的衔接子是有利的, 但是与液相系统相容的标记将更为适合, 例如和电喷射质谱系统一起使用, 因为本测序发明是一种反复的过程, 序列样品是连续分析的, 而不是如基因描述系统中那样只分析一次。

杂交实验:

参考文献 14 公开了一种鉴定 RNA 三级结构中寡核苷酸可及(accessible)位点的方法, 该方法不需要扩增寡核苷酸或进行任何形式的电泳。检测短的寡核苷酸探针(较佳的是 4 聚体)与 mRNA 的结合, 结合方式与 mRNA 的一级结构相关联。可及区域会有许多探针与之高亲和力结合, 那些探针的序列应与可及区域的一级序列互补, 探针的序列也应重叠。在上述专利申请中, mRNA 或探针被固定在固

相基质上, 标记的探针或 mRNA 分别与捕获的核酸杂交。参考文献 14 中公开的较佳的标记方法是荧光标记, 但是显然也可使用经质量标记的核酸。

许多杂交为基础的方法是本领域中所知的, 但是特别重要的是 Southern 印迹法和检测样品中特异性序列存在的其它方法。对于本领域技术人员来说, 带质量标记的杂交探针显然可用于这些目的。另外很显然的是, 采用质量标记的杂交探针的优点是能用多个独特的质量标记的核酸杂交探针同时探测多个序列。

参考文献 15 讨论了与质量标记的核酸探针相容的各种杂交试验。

用质谱法分析质量标记的核酸:

质谱法的基本特征如下:

输入系统→离子来源→质量分析仪→离子检测器→数据捕获系统。为了分析生物分子(在本申请中是带质量标记的核酸探针), 输入系统和离子来源是关键所在。为进行生物学分析的其它重点是质量分析仪/检测器排列的灵敏度及其定量测定分析物分子的能力。

电离技术:

许多生物学质谱应用采用了所谓的“软”电离技术。这就使得大分子(如蛋白和核酸)能电离而基本上不会片段化。液相技术能使大分子在温和的 pH 溶液中以低浓度进入质谱仪。许多技术适用于本发明, 包括(但不局限于)电喷射电离、快速原子轰击和基质辅助的激光解吸附电离(MALDI)。

电喷射电离:

电喷射电离需要将生物分子的稀释液喷雾到质谱仪中, 即以细喷雾形式注入。例如, 溶液可在静电场影响下靠一股干燥的氮气从毛细管尖头喷出。对于电离机理并未完全了解, 但是认为其大致如下。溶剂在氮气气流下蒸发。随着液滴变小, 生物分子的浓度增加。在喷射条件下, 大多数生物分子携带了正或负的净电荷, 这就提高了溶解的生物分子之间的静电斥力。随着蒸发的继续, 该斥力最终变得大于液滴的表面张力, 使得液滴爆炸成更小的液滴。静电场有助于进一步克服液滴的表面张力并促进喷射过程。更小的液滴继续蒸发, 从而反复爆炸直至生物分子基本上处于全部为溶剂的气相中。该技术对于使用质量标记来说特别重要, 因为它赋予了离子非常少的额外内能, 以致群体中的内能分布趋向于在较窄的范围内。在施加的电场梯度的影响下, 离子被加速冲出电离室外。该电场梯度的方向决定了是正离子还是负离子进入质量分析仪。电场强度为它们提供了动能。这继而导致在离子和中性分子碰撞时有或多或少的能量转移, 从而可能引起

片段化。在考虑离子在质谱仪内片段化时，这是很重要的。赋予一群离子的能量越多，则分析物分子通过与离子源中存在的槽气(bath gas)或溶剂蒸汽碰撞产生片段的可能就越大。通过调节电离室中用于加速离子的电压，就可控制离子的片段化。当离子片段化被用作从质量标记核酸中切断标记物的手段时，这一现象很有利。

基质辅助的激光解吸附电离(MALDI):

MALDI 需要将生物分子包埋在大的摩尔过量的光活性“基质”中。施加合适频率的激光(对于烟酸为 266 纳米)导致基质激发，从而使包埋的生物分子激发并电离。该方法为离子提供了大量的平动能，但是不会诱导过度的片段化。该方法也可用电场来控制片段化。MALDI 方法可以两种方式使用。可将带质量标记的 DNA 包埋在基质中，这样标记本身不会被激光特异性激发，或者可将标记构建成含有允许激光激发的所需基团。后一方法意味着在进行质谱法之前不必将标记包埋在基质内。这些基团包括烟酸、芥子酸或肉桂酸部分。MALDI 为基础的标记断裂对于可光断裂的接头来说可能是最有效的，因为这将避免在进行 MALDI 质谱法之前的断裂步骤。不同的可激发电离剂有不同的激发频率，因此可以选择不同的频率来引发用来断裂可光解接头从那些物质上电离下来。这些可激发部分可用有机化学的标准合成方法来衍生获得，以提供具有一定质量范围的各种标记。该范围可以组合方式构建。

快速原子轰击:

快速原子轰击描述了许多蒸发和电离相对无挥发性的分子的技术。这些技术的基本原理是使样品与加速的原子或离子(通常是氩原子或铯离子)碰撞，从表面上解吸附下样品。和 MALDI 一样，可将样品涂布在固体表面上，但是不需要复杂的基质。这些技术也适用于液相输入系统—使毛细管电泳入口或高压液相层析的洗脱液经过釉料(frit)，基本上是将分析物溶液涂布在釉料表面，使分析物受原子轰击从釉料表面电离下来。

定量测定和质谱法:

通常，许多生物化学和分子生物学试验是定量的。质谱仪并不是一种简单的定量测定装置，但是采用合适的仪器可获得很高的灵敏度。达到质谱仪检测器的离子数目并不直接表示实际在离子源中的分子数目。离子数目和生物分子的最初浓度间的关系是电离行为的复杂函数。定量可通过扫描质谱和计数每一扫描的质荷比处的离子数来实现。对计数求和，给出在给定时间内质谱中每个点的总数。这些计数可反过来与样品中的来源分子的原始量相关联。关联离子计数或电流和

源分子的量的方法可以不同。外部标准法(external standard)是一种方法,其中在测定未知样品前先确定样品分子的行为。通过测定系列稀释的样品分子在加入所用仪器结构中时的离子流,就可测得每个样品分子的标定曲线。

内部标准法可能是一种比外部标准法更佳的方法,因为内部标准品所受的试验条件与样品相同,因此任何试验变化对内部对照和样品分子均有影响。为了确定样品中的基质量,在样品中加入已知量的内部标准品。选择内部标准品,使其电离行为与待测基质的相似。可用样品离子计数与内部标准品离子计数之比来确定样品量。该方法主要的困难是选择合适的标准品。内部标准品应与基质相似,但是质量不同。最佳的方法是采用同位素标记的内部标准品。如果需要大量质量标记,则该方法不如采用外部标准品理想,因为合成合适的内部标准品的成本很高。然而,这些标记的定量测定结果比外部标准品更佳。同位素标记的替换方案是找出一种内部标准品,该标准品的化学行为与样品在质谱仪中的相似但不相同。寻找这种类似物很困难,它可能是质量标记大家族的重要任务。

一种折中的方法可能是合适的,因为通过组合方式合成的质量标记大家族在化学上是相关的。可采用少量内部对照品,其中每个独立的对照品确定了许多质量标记的量。内部标准品与每个质量标记间的精确关系可在外部标定试验中确定,以补偿它们之间电离特征的差别。

质谱仪的构造对于确定实际离子计数来说很关键。在这方面,电离和质量分离方法是特别灵敏的。某些质量分析方法作为“质量过滤器”起作用。例如,四极矩质谱仪每一次只允许具有特定质荷比的离子通过。这意味着相当大比例的离子决不会到达检测器。大多数质谱仪一次只检测一部分质谱。假定大部分质谱是空的或不相关的,但通常仍被扫描,这就意味着更大部分的样品被浪费。这些因素可能是检测丰度非常低的离子存在的问题,但是这些大部分可通过纠正仪器的构造而克服。

为了确保更佳的定量测定结果,应试图确保所有离子受到检测。Mattauch-Herzog 几何扇区仪(geometry sector instrument)可实现这一点,但是有许多限制。扇区仪由不同区域(扇区)组成执行某些功能。通常,离子在离子源中产生,来自发散束,该光束通过可调节的狭缝时变窄。然后,这种明确的光束通过无场区域进入电扇区中被聚焦。通过狭缝时一些离子损失,因而对样品来说其灵敏度降低。聚焦的离子束经过第二个无场区进入磁扇区。最后的这个扇区根据离子的质荷比来对离子束聚焦。在质量分离束裂缝间可放置照相板,它可用来测定离子丰度及其质荷比。不幸的是,照相板在饱和前只有较小的动态灵敏度范围,并且很

笨重。采用电子倍增器阵列可实现更佳的动力范围，但代价是分辨力会有所损失。利用该阵列，就可监测特性经充分鉴定的质量标记家族。通常，阵列检测器可同时连续监测质谱的多个区域。此阵列限于分辨质谱中间隔较近的区域可能限制了可采用标记的数目。对于“选择离子监测”(SIM)，四极矩装置比许多其它构造更有利，因为能非常迅速地改变分离不同质荷比的离子的电场，从而可对较少的感兴趣的峰作高速取样。

质量分析仪几何结构:

质谱法是一门高度多样化的学科，存在许多质量分析仪构造，它们通常能与组合在各种几何结构中，从而能分析复杂的有机分子。典型的单阶段质量分析仪是四极矩或飞行时间(time-of-flight)仪器，它们均与本发明相容。扇区仪也是适用的。

正交 TOF 质谱法:

对于生物学应用来说，测定的灵敏度和样品定量非常重要。灵敏度与阵列几何结构相当的一种方法是正交飞行时间质谱仪。该几何结构能非常迅速地对离子束取样，然后几乎瞬时检测所有的离子种类。离子流离开离子源(对于许多生物应用来说可能是电喷射来源)，经过一与离子束垂直放置的平板电极。该电极本质上是电门。脉冲电势使部分离子束呈“正交”偏移到飞行时间质量分析仪中。当电门“闭合”使离子偏移到 TOF 分析仪中时，触发一计时器。记录下偏移离子的飞行时间，这样就足以确定其质荷比。该门每次通常只将离子短脉冲传送到 TOF 分析仪中。由于记录了所有离子的到达且 TOF 分离非常迅速，因此可以同时有效地测定全部质谱。另外，门电极可以极高的频率对离子束取样，因此可在非常短的时间间隔内累积多个质谱。当离子源中的样品浓度低或只能维持很短时间时，这非常重要。正交 TOF 几何结构非常灵敏。

用串联质谱法分析带质量标记的核酸:

串联质谱法描述了许多技术，其中用第一个质量分析仪根据离子的质荷比从样品中选出离子，通过诱导所选的那些离子片段化，作进一步分析。用第二个质量分析仪分析片段化产品。串联仪器中的第一个质量分析仪作为过滤器选择待研究的离子。所选离子在离开第一个质量分析仪时通过含有中性气体的碰撞室，致使其中一些离子片段化。

离子源→MS1→碰撞单元(cell)→MS2→离子检测器

质量标记的诱导断裂:

这些年来已经开发了各种分析方法来促进离子的片段化,用于结构研究和根据片段化“指纹(fingerprints)”来明确地鉴定分子。大多数电离方法产生了一些片段,但是软电离方法几乎不产生片段离子。然而,例如化学电离方法的变化方案可用来帮助片段化。同样,可对电喷射电离稍加改进来促进片段化,加入一个电晕放电电极来电离更多的样品分子或增加分子性离子的片段化。该方法已被称为大气压化学电离(APCI)。

一种更活泼的片段化方法需要诱导分子性离子分解,例如通过碰撞诱导分解(CID)。CID利用质谱仪构造来分离出一套选定的离子,然后通过和中性气体碰撞诱导离子片段化;所得片段离子用第二质谱仪分析。

其它诱导断裂的方法也适用于质量标记方法学。如较早讨论的一种较佳的方法是光子诱导的分解,它涉及采用可光断裂的质量标记。典型的结构采用了与CID中所用类似的串联质量分析仪构型,但是碰撞单元以光激发室代替,在光激发室中用激光来激发离开第一质量分析仪的离子流。需要高强度激光来确保大部分快速移动的离子流与光子反应适当地诱导断裂。激光安置的位置对保证离子流暴露有效的时间极其重要。将激光调频至特定频率就可精确地控制诱导断裂的键。因此,通过合适的可光断裂接头与其探针连接的质量标记可在质谱仪内断裂。光断裂阶段不需要串联结构,光断裂室可在离子源中或紧随其后。

分子性离子片段化的另一种可用的技术是表面诱导分解。表面诱导分解是一种串联分析仪技术,它涉及产生一股离子束,该离子束在第一分析仪内分离成选定的 m/z 比。所有选定的离子以掠射角与固体表面碰撞。然后用第二质谱仪分析所得碰撞片段。

一类串联质谱仪采用三重四极矩集合,它包含三个四极矩质量分析仪,其中一个作为碰撞室。碰撞室四极矩不仅作为碰撞室,还是其它两个四极矩质量分析仪间的离子导引器。在中间的四极矩中可引入气体,以使气体分子与从第一个质量分析仪进入的离子碰撞。片段离子在第三个四极矩中分离。诱导断裂可用除采用串联扇区或四极矩分析仪外的结构来进行。可采用离子捕获质谱仪,它可通过在捕获器中引入缓冲液或“槽气”来促进片段化。所有被捕获的离子与缓冲液气体分子碰撞,产生的能量转移可能会导致碰撞。通过加速捕获离子可增加碰撞能。离子捕获器中可用氦或氖作为槽气。同样,光子诱导的片段化可适用于捕获的离子。另一种有利的结构是四极矩/正交飞行时间仪器,其中四极矩的高速扫描与TOF质量分析仪的更高灵敏度相结合以鉴定片段化的产品。

常规的“扇区”仪器是用于串联质谱法的另一种常用结构。一个扇区质量分析仪包含两个分开的“扇区”，电扇区利用电场将离开离子源的离子束会聚成有相同动能的离子流。磁扇区根据离子质量来分离离子，在检测器中产生质谱。对于串联质谱法，可采用两个此类扇区质量分析仪，其中电扇区提供第一质量分析仪，磁扇区提供第二质量分析仪，两个扇区间放置一个碰撞单元。该结构对于断裂带质量标记的核酸中的标记可能是非常有效的。由碰撞单元分开的两个完整的扇区质量分析仪也可用来分析带质量标记的核酸。

离子捕获:

离子捕获质谱仪是与四极矩质谱仪相关的仪器。离子捕获器通常有3电极结构——一个“圆柱形(torroidal)”电极和每个末端形成空腔的“帽”电极(离子捕获器)。在圆柱形电极上施加正弦射频电势，而帽电极用DC或AC电势加偏压。喷射到空腔内的离子由于圆柱形电极的振荡电场而被限制在稳定的圆形轨道内。然而，对于给定振幅值的振荡电势，某些离子会有不稳定的轨道，并会从捕获器中射出。通过改变振荡射频电势，就可根据其质荷比将射入捕获器中的离子样品依次射出捕获器。然后可检测射出的离子，产生质谱。

离子捕获器通常通过离子捕获器空腔内的少量“槽气”(如氦)来操作。这提高了该装置的分辨力和灵敏度，因为进入捕获器的离子通过与槽气分子碰撞而基本上被冷却至槽气的环境温度。碰撞减弱了离子轨道的幅值和速度，使得它们更靠近捕获器的中央。这就意味着，当改变振荡电势时，轨道变得不稳定的离子比减幅循环离子更迅速地获得能量，并以更紧密的集束离开捕获器，从而提供更高的分辨力。

离子捕获器能模仿串联扇区质谱仪的结构。实际上，它们能模仿多质谱仪的结构，从而能对被捕获离子进行复杂的分析。捕获器内可保留单一质量的样品，即，可将所有其它物质射出，然后，在第一个振荡频率上叠加第二个振荡频率，小心地激发保留下来的物质。动力学上激发的离子与槽气分子碰撞，如果离子充分激发则会产生片段。然后可进一步分析片段。这是MS/MS或MS²。通过射出所有其它离子然后动力学激发片段，使片段与槽气分子碰撞后，就可对片段离子作进一步分析(MS/MS/MS或MS³)。该过程可重复进行，只要有充足的样品存在而进行进一步分析(MSⁿ)。应当注意，在诱导片段化后，离子捕获器通常保留了大部分片段离子。这些仪器和FTIGR质谱仪(见下述)代表了时间分辨串联质谱法，而不是在线性质谱仪中所见的空间分辨串联质谱法。

傅里叶变换离子回旋共振质谱法(FTICR MS):

FTICR 质谱法与离子捕获器的相似特征在于将离子样品保留在空腔内,但是在 FTICR MS 中通过正交电磁场离子被捕获在高真空容器(ICR 单元)内。电场由一对平板式电极产生,这对电极形成了一个盒子的两侧面。该盒子被包含在一磁场中,该磁铁与两块平板(捕获板)连接,将射入的离子强制在圆滚形轨道内。通过在两块“传递板”上施加射频脉冲,离子可在动力学上被激发进入更大的圆滚形轨道。离子的圆滚形运动在盒子其余两个相对侧面(板,它们是“接收板”)上产生了相应的电场。激发脉冲从动力学上激发离子进入更大的轨道,该轨道随着离子相干运动(coherent motion)的丧失(由于离子与中性气体分子的碰撞)而衰减。通过傅里叶变换分析将接收板检测到的相应信号转变成质谱。

对于诱导片段化的实验,这些仪器可以类似于离子捕获器的方式进行工作——将除感兴趣的单一物质外的所有离子射出 ICR 单元。可将碰撞气体引入捕获器中并诱导片段化。随后可分析片段离子。如果用 FT 分析“接受板”检测到的信号,通常由于片段化产物与槽气混合导致较差的分辨力。然而,可将片段离子射出单元,然后用(例如)四极矩串联结构进行分析。

带质量标记的杂交探针

为了从质量标记获得所需的行为,希望这些标记有某些化学性质,它们可用以多种方式掺入质量标记的特殊的分子基团或部分为代表。

带质量标记的杂交探针的结构

带质量标记的杂交探针可以有如下列基本结构。

Nu-M

Nu-L-M

其中 Nu 是核酸探针, L 是连接核酸探针和质量标记 M 的接头基团,接头基团(L)是任选的,质量标记中可加入所需特征的接头。当需要不可断裂的带质量标记的杂交探针时不需要接头基团。核酸是核苷酸的线性聚合物(核苷酸的天然存在的种类数量相当少,但是化学合成的类似物的数量不断增加),其许多位置可与接头基团连接。下面将讨论这些可能性。

接头:

接头基团可有下列结构特征:

手柄(handle)1-[可断裂基团]-手柄 2

手柄 1、2 是化学基团, 它能使接头一端与核酸探针相连, 另一端与质量标记相连。至少需要有一种可断裂基团在手柄间或作为手柄的一部分, 从而使质量标记能从其相联的核酸探针上可控制地除去。

质量标记:

质量标记可有下列结构:

手柄-质量标记

其中手柄是一种基团, 它能使质量标记连接其相应的核酸探针、或与质量标记与核酸探针之间的接头连接。

质量标记的性质:

为了获得本发明质谱法的最优性能, 对于这些质量标记来说, 高达 2000 至 3000 单位的质荷比是合适的范围, 因为该范围对应于能可靠地、最灵敏地被检测到的单电荷离子的范围。然而, 质量小于 200 至 300 道尔顿的质量标记并不理想, 因为所有质谱的低质量端有溶剂分子、小分子杂质、多个电离峰和片段峰聚集的倾向。而且, 每个标记应与其邻居分开至少约 4 道尔顿, 以避免碳、氮和氧同位素峰产生重叠。

质量标记应能电离和分离形成主要一种物质(没有片段化)。

质量标记应容易电离, 以确保能检测到尽可能多的断裂的质量标记。

为了允许检测, 标记需要有净电荷, 但是最好不是多次电离, 即它们应有单电荷。而且, 标记应能抵抗片段化作用, 从而使质谱扫描中的每个峰仅仅或唯一对应于一个单一标记; 这就简化了数据分析, 减少了测定标记量时的模糊性, 这是本发明开发的一些应用中非常重要的标准。

存在各种化学官能团, 它们携带或可能携带正离子质谱法所要求的正电荷。它们包括, 但不局限于, 胺(特别是叔胺和季铵)、磷和硫醚。季铵基团携带了单个正电荷, 它不需要进一步电离。对于正离子质谱法来说, 这些预先电离的物质具有较高的灵敏度。因此, 较佳的正离子质量标记应携带至少一个这种基团。冠醚是另一类可用来携带正电荷的化合物。

可采用各种化学官能团来携带进行负离子质谱法所需的负电荷, 它们包括,

但不局限于，羧酸、磷酸、磷酸和磺酸、酚羟基类、磺胺类、磺酰脲类、四唑和全氟醇类。

质量标记从核酸探针中电离和分离出来:

DNA 和其它核酸在质谱仪中有广泛片段化的趋势。希望确保所得质谱中的DNA 片段峰没有使质量标记引起的峰变模糊。最好保证核酸探针片段在断裂后与质量标记分离。为了实现这一点，可采用在电离时形成负离子的质量标记，它可通过负离子质谱来分离。尽管核酸具有带负电荷的骨架，但是它在电离(特别是电喷射和相关的液相到气相的电离技术)时有质子化趋势。这意味着，如果质谱仪的结构用于进行负离子质谱法，则质谱中仅出现带负电荷的质量标记。大多数核酸片段不会到达检测器。

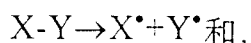
如果采用这种方法，可用合适的缓冲液促进核酸探针质子化，从而确保核酸以预先存在的正电荷形式广泛存在。

质谱仪内的片段化:

片段化是质谱仪的一个非常显著的特征。根据本发明，考虑怎样来鉴定质量标记很重要。在一种极端情况下，质量标记可被设计成它们非常不易片段化，标记通过质谱中标记的分子性离子的出现来确定。在这种情况下，需要设计出具有唯一分子性离子的标记家族。在另一种极端情况下，可以设计具有高度特征性片段化方式的质量标记，从而可通过该方式来鉴定它。在这种情况下，必须设计出有非重叠方式的标记家庭，或每个标记有至少一个独特的片段化物质。片段化是最初分子以及用来从分子产生离子的电离方法的性质。不同的方法赋予了最初形成的离子不同的能量，且离子的化学环境变化很大。因此，适用于一种质谱法的标记可能不适用于另一种质谱法。较佳的方法是设计有抗片段化能力的分子，尽管有一些片段化作用是不可避免的。这意味着一个目的是用单种主要物质(可以是分子性离子或单个易产生的片段离子)来鉴定分子。

在质谱仪中测定键的稳定性

在中性分子中，通过考虑键强度就可很简单地确定分子是否有抗片段化作用。然而，当分子被电离时，键强度可能增大或减小，很难预测那个优先。例如，对于给定的键，X-Y 在其非电离的形式：



所以 $D(X-Y) = \Delta H(X^\circ) + \Delta H(Y^\circ) - \Delta H(X-Y)$

其中 D 表示合适单位的键解离能。

但是, 对于电离的物质(在本例中为正电荷),

$D(X-Y)^+ = \Delta H(X^+) + \Delta H(Y^\circ) - \Delta H(X-Y^+)$

所以 $D(X-Y) - D(X-Y)^+ = \Delta H(X^\circ) - \Delta H(X^+) - \Delta H(X-Y) + \Delta H(X-Y^+)$

因为

$I(X^\circ) = \Delta H(X^+) - \Delta H(X^\circ)$, 其中 I 是电离能,

$I(X-Y) = \Delta H(X-Y^+) - \Delta H(X-Y)$

所以, $D(X-Y) - D(X-Y)^+ = I(X-Y) - I(X^\circ)$

这意味着如果 $I(X-Y) > I(X^\circ)$, 则 $D(X-Y) - D(X-Y)^+ > 0$, 但是,

同样, 如果 $I(X-Y) < I(X^\circ)$, 则 $D(X-Y) - D(X-Y)^+ < 0$ 。

因为 $I(X-Y)$ 和 $I(X^\circ)$ 都是正的, 所以若 $I(X-Y) < I(X^\circ)$, 形成的键更强

当 $I(X-Y) > I(X^\circ)$, 形成的键较弱。

在上述方程式中, $D(A-B)$ 指括号内物质的键解离能, $I(N)$ 指括号内物质的电离能, ΔH 是括号内物质的形成热函。为本发明的目的 $\Delta S = 0$, 因此 $\Delta G = \Delta H$ 。上述方程式的结论是, 为了预测键是否能在给定的一组电离条件下稳定, 需要知道分子的电离能和讨论的键片段化所产生的中性片段的电离能。

例如, 考虑苯胺中的 C-N 键:

$I(NH_2^\bullet) = 11.14$ 电子伏特(eV), $I(C_6H_5NH_2) = 7.7$ eV

所以 $I(C_6H_5NH_2)$ 比 $I(NH_2^\bullet)$ 小 3.44 eV

该键中的另一个断裂是:

$I(C_6H_5^\bullet) = 9.35$ eV, $I(C_6H_5NH_2) = 7.7$ eV

所以 $I(C_6H_5NH_2)$ 比 $I(C_6H_5^\bullet)$ 小 1.65 eV

因此, 该键在离子中不容易断裂。

如果苯胺有足够的起始能量来片段化, 则通常发现它会断裂释放出 HCN, 而不是 C-N 键断裂。类似的考虑适用于苯酚:

$I(OH^\bullet) = 13$ eV, $I(C_6H_5OH) = 8.47$ eV

所以 $I(C_6H_5OH)$ 比 $I(OH^\bullet)$ 小 4.53 eV

该键中的另一个断裂是

$I(C_6H_5^\bullet) = 9.35$ eV, $I(C_6H_5OH) = 8.47$ eV

所以 $I(C_6H_5OH)$ 比 $I(C_6H_5^\bullet)$ 小 0.88 eV

在苯酚的分子性正离子中没有观察到 C-O 键断裂。

测定分子和中性片段的电离能之差是一种常用的工作原理，它可用来预测可能的离子键强度。如果电离时加入的能量低于离子键强度，则不会见到片段化。具有良好强度的典型的离子键包括芳基-氧、芳基-氮、芳基-硫键，它们由于电子的离域作用而稳定。通常，脂族型键在离子形式时较不稳定。因此，碳碳单键在离子中较弱，但碳碳双键仍相当强。由于和芳基-氧等相同的原因，芳基-C=C 键强度过大。芳基或芳基-F 键在离子中也强，它很适合质量标记，因为氟碳化合物的生产成本很低，是化学惰性的，相对于碳氢分子它有可检测的质量亏损(defect)，且氟只有一个天然存在的同位素 ^{19}F 。

应用于负离子可作类似的考虑，只是在上述方程式中需要用电子亲和力。

接头的性质：

质量标记从其相联的核酸探针上可控制地释放下来可通过以下各种方式来实现：

- 光断裂
- 化学断裂
- 热断裂
- 质谱仪内诱导片段化

很容易设计出可光断裂和可化学断裂的接头用于所述应用。图 5 显示了一系列典型的可光断裂接头。

邻硝基苄基是本领域中众所周知的作为可光断裂接头的基团，它在苄基胺键处断裂。关于可断裂接头的回顾参见参考文献 18，该文讨论可各种可光断裂和可化学断裂的接头。

热断裂通过热诱导重排进行。图 6 显示了一个质量标记例子的合成，该标记通过可热断裂接头与胸苷残基的 3'-OH 位相连。图 6 还显示了将标记从其相联的核苷酸上断裂下来的热诱导重排。显然，本例中的基团 X 可以是下文讨论的芳基醚聚合物。较佳的，这个可热断裂基团也产生了大量适用于负离子质谱法的负离子。该分子的热分解需要接头中有 S=O 基团。这里，S 可用 N 或 C 代替，O 可用 S 代替。关于其它例子参见参考文献 28。

质量标记在质谱仪中的断裂：

一种较佳的断裂方法是用电离过程诱导标记片段化。接头可设计成在电离过程中高度稳定，这样当与其相连的分子在质谱仪中电离时会断裂。在采用该方法

控制断裂时要考虑两种因素：(1)在电离过程中有多少过多的能量存放(deposit)在离子中，和(2)该过量是否足以克服离子中的任何一个键能。存放的过量能量在很大程度上取决于采用的电离技术。为了使存放能量能有效断裂一个键，该能量必须是振荡/旋转模式，并且必须足以克服键的解离能。显然，键能由待分析分子的化学结构确定。键能将在以后讨论。一般来说，电离过程中能量以电子、振荡、旋转和平动能方式赋予。在很短的电离时间内，这一过量的内能的大多数已通过系统间和状态间的交叉转化成振动能和旋转能。过量的内部再振动能可能会或不会导致键断裂。为了赋予移动离子更多的内部振动能，可使它们与槽气碰撞，以使离子片段化。在电喷射源中有槽气和挥发性溶剂。离子可通过电场加速以增加与槽气碰撞的能量。加速给予离子动能。如果赋予离子足够的动能，则离子与槽气碰撞后会片段化。所需的动能量取决于离子中键的强度，但是通过调节加速电势就可控制赋予的能量。

为了产生电离时在预定键断裂的用于质量标记的接头，接头中需要有一个弱键，而其余为强键。某些基团有特别的抗片段化能力，而其它基团(如脂族类键)却相当容易断裂。为了设计出在指定位置断裂的接头，可以这样来设计分子：它总体上能抗片段化，但是含有一个“弱键”。当断裂发生在离子的某些键上时，发现某些结构特征可使片段离子稳定。直链烷的片段化相对随机，但是具有仲烷基和叔烷基的分子因为仲碳和叔碳增加稳定化作用而最常在分子的支链点断裂。同样，双键通过共振或离域效应使毗邻的正电荷或负电荷稳定。发现毗邻芳基的键有类似效应。图7中显示了可通过碰撞或其它作用诱导片段化的一些可断裂接头。它们以其不稳定性的增加按次序编号。可断裂键左侧基团是熟知的良好的离去基团，用来保护分子中的反应位置。因此，它们易在某些条件下受化学断裂。可选用的精确结构取决于探针的用途和化学环境。图7中的接头(4)非常易受质子化学攻击，因此它只能用作可片段化接头(如果探测反应不是酸性)。接头(1)相当不易光解断裂。显然，可有意选择这些基团按需进行化学断裂。从图7可以看出，这些接头也可作为离域的芳基-醚聚合物系统的一部分。可断裂键右侧的基团基本使负电荷稳定，这是有利的，因为它促进了该位点的键断裂，并可提供可检测的负离子。该位置还可使用其它电荷稳定基团。此图和其它图中的“手柄”通常代表用于质量标记的碱基序列合成中的反应基团，它们也可以不存在于合成的带质量标记的分子中。

核酸探针:

将基团与核酸连接:

质量标记及其接头可与核酸的多个位置连接。对于常规的固相合成,核糖的5'羟基最容易衍生。其它有利的修饰位置是嘧啶中的5'位和嘌呤中的7'和8'位。这些是连接可断裂质量标记和不可断裂的质量标记的较佳的位置。

糖的2'位可进行质量修饰,但更适合不打算除去的小质量修饰。

天然核酸中的磷酸键也可作很大程度的修饰,包括用质量标记衍生。

杂交探针:

根据用途,可能要采用修饰过的核酸,它含有许多不同的类似物,被修饰成有杂交行为。当同时采用杂交探针的基团时,这是特别重要的。修饰探针基团的杂交行为是需要的,因此正确杂交的探针的熔点非常接近或至少高于某临界值。不正确杂交探针的熔点宜低于该临界值。这样就能同时使用探针的各基团,并同时确保杂交反应的严谨性。

含有所有 Watson-Crick 碱基对的短寡核苷酸双链体的稳定性间有主要的差别。例如,与仅含鸟嘌呤和胞嘧啶的双链体相比,仅含腺嘌呤和胸腺嘧啶的双链体不稳定。当试图使短寡核苷酸混合物与靶 RNA 杂交时,稳定性的这些差别可能会产生问题。富含 A-T 序列的杂交需要低温,但是在这些温度下,富含 G-C 的序列与不完全互补的序列杂交。这就意味着,可能会发生一些错配,富含 G-C 的序列会丧失特异性。在更高的温度下,富含 G-C 的序列会特异性地杂交,但是富含 A-T 的序列则不杂交。

为了使这些效果标准化,可对核酸作修饰。这些修饰分为三大类:碱基修饰、骨架修饰和糖修饰。

碱基修饰

对标准的 Watson-Crick 碱基可作多种修饰。下面是标准化碱基配对能量到某种程度的修饰例子,但是它们没有限制性:

- ° 腺嘌呤类似物 2,6-二氨基嘌呤与胸腺嘧啶形成三个氢键(而不是两个),因此形成了更稳定的碱基对。

- ° 胸腺嘧啶类似物 5-丙炔基脱氧尿苷与腺嘌呤形成更稳定的碱基对。

- ° 鸟嘌呤类似物次黄嘌呤与胞嘧啶形成两个氢键(而不是三个),因此形成了较不稳定的碱基对。

这些和其它可能的修饰使压缩短寡核苷酸与其互补序列特异性杂交的温度

范围成为可能。

骨架修饰:

核苷酸的磷酸部分很容易修饰。在某些条件(如低盐浓度)下,类似物如甲基磷酸、三酯和氨基磷酸酯已经显示可增加双链体稳定性。这些修饰可能也具有增加的耐受核酸酶能力。其它磷酸修饰包括磷酸二醚合物(phosphodithirite)和甲硼磷酸盐,它们均能增加寡核苷酸抗外切核酸酶的稳定性。

用硫代替磷的电子等排替换提供了抗核酸酶的寡核苷酸(参见参考文献 19)。也可用碳替代磷或连接的氧。

糖修饰:

在糖部分的 2'位可作各种修饰(参见参考文献 20 和 21)。糖可被不同的糖(如己糖)替换,或者整个糖磷酸骨架可被新的骨架(如肽核酸(PNA)中的骨架)完全替换。讨论参见参考文献 22。PNA 形成了至今发现的所有类似物的具有最高热稳定性的双链体。

疏水修饰:

在寡核苷酸 3'和 5'端加入疏水性基团,通过隔离水和碱基来提高双链体的稳定性,从而减少复合物的“磨损”,即,疏水性基团减少了末端碱基的溶剂化。

人为错配:

杂交反应中的错误的一个主要来源是引物与靶序列及其它未知碱基的杂交严谨性。如果作用于靶的引物携带单个人为导入的错配,则系统的辨别力非常高(参见参考文献 23)。当采用完全互补的引物时,其它错配忍受的程度与单个错配的忍受程度不相同。通常发现,有一个错配的双链体与有两个错配的双链体间的解链温度差别大于正确杂交的双链体与有一个错配的双链体间的解链温度差。预计这是本申请公开的杂交探针的一个重要特征。如果核酸探针有一个关键性碱基,即,为了检测单核苷酸多态现象(polymorphism),在远离该关键性碱基的 1 处螺旋转角处引入 1 个人为的错配,使双链螺旋在很大程度上变得不稳定(如果探针位点有第二个错配的话)。

杂交程序:

关于核酸探针的杂交条件(尤其是缓冲液和温度)的影响的详细情况参见参考文献 24 至 26。

寡核苷酸合成:

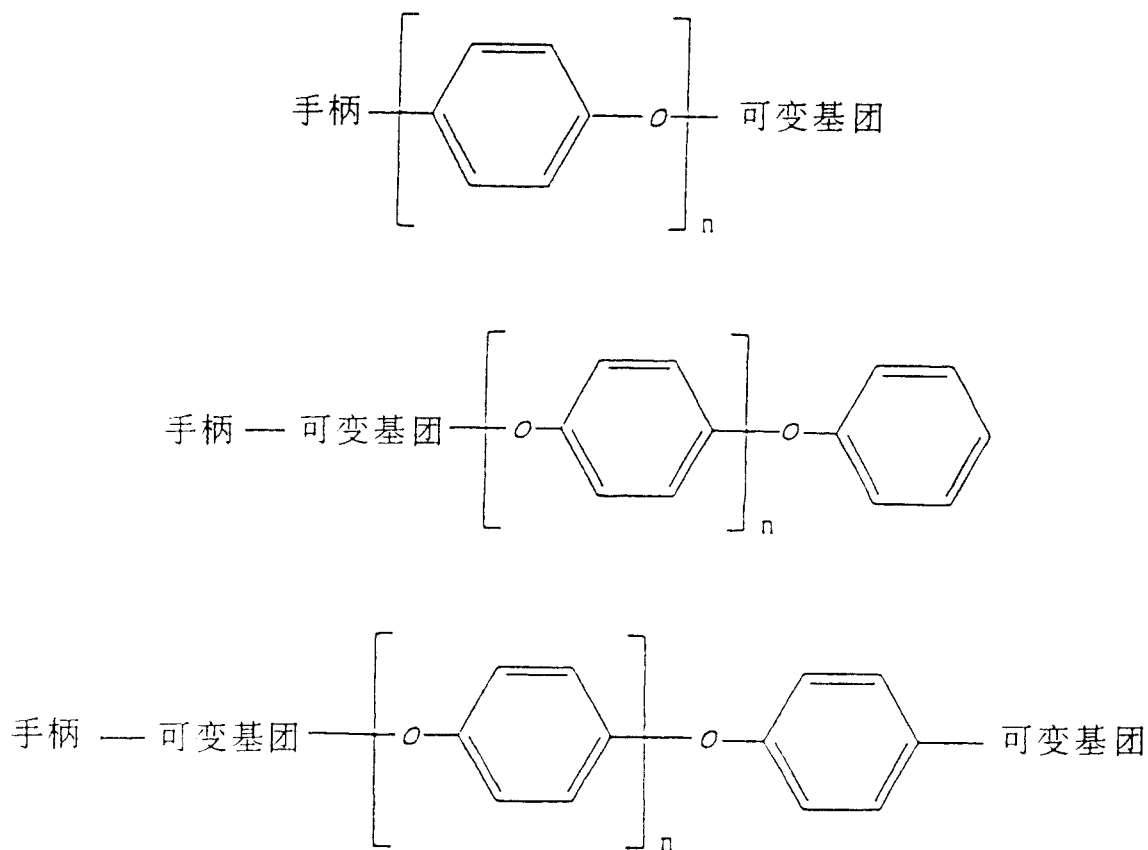
寡核苷酸的合成方法是本领域中熟知的(参见参考文献 27 和 28)。

质量标记的合成:

对于所有在实践或工业上有用的系统,重要的是,标记的构建应尽可能的简单,采用试剂和加工步骤应尽可能的少。组合方法是理想的,其中可将一系列单体分子单元用于彼此之间的多个组合。

可用有机化学技术合成质量标记。该标记可携带具有单个电荷的基团,并可在所用的质谱法中抵抗片段化作用。如果采用正离子质谱法,则胺衍生物、季铵离子或带正电荷的硫中心是良好的电荷载体。它们有极好的检测性质,产生了清晰的明显的信号。同样,也可采用负电荷离子,这样,有羧酸、磺酸和其它部分的分子适用于负离子质谱。用于 MALDI 质谱法的标记可通过衍生处理可被紫外可见激光激发的已知分子(例如芥子酸或肉桂酸,它们已有许多衍生物可从商业上购得)来产生。上面讨论了抗片段化的基团。关于有机化学的课本参见参考文献 29 或 30。

这些标记的组合合成可以相当简单的方式来实现。下面显示了较佳的质量标记结构。



这些聚芳基醚结构非常耐片段化,并产生了良好的负离子,因为电子的离域作用超过分子可有效地稳定负电荷。这些分子也是热稳定的,因此特别适用于热断裂的接头和在质谱仪中通过碰撞过程断裂的接头。聚芳基醚任一端的“可变基

团”宜为修饰质量标记性能的取代的芳基醚(图 9)。这些修饰基团包括“质量系列修饰(mass series modifying)”基团(见图 9)、稳定化基团、带电荷基团(见图 10)和质量亏损基团(见图 8)。线性聚芳基醚的聚合物分子中每增加一个“苯氧基”残基,就增加 92 个质量单位。为了充分利用质谱,质量标记只需相差大约 4 道尔顿。为了产生相差 4 道尔顿的质量标记,每个质量标记宜含有一个调整(shift)每一系列芳基醚质量的基团。该质量系列修饰基团(MSM)(见图 9)的作用是使每一系列芳基醚聚合物的质量偏移到其它聚合物。对于芳基醚的线性聚合物,每个单体增加 92 道尔顿,如果每一系列质量标记相差 4 个质量单位,那么质量不符合的将最多有 23 个系列。为了产生 256 个质量标记,例如,需要产生 23 个 MSM 基团,以连接芳基醚聚合物和最多 12 个连续的苯氧基重复单元。这将提供总共 276 个质量标记。

显然,可以产生包含许多不同亚单元的聚合物,那些亚单元以不同的顺序出现。另外,也可以是支链结构,但是为了便于描述,只显示了线性聚合物。为了便于合成,选择所示的较佳的结构。相同亚单元的不同顺序的生产并不是更困难,但是最好是在尽可能少的合成步骤内产生尽可能多的标记。一种较佳的合成策略是产生高达 12 个重复单元的聚芳基醚,然后用许多不同的 MSM 基团对其作衍生处理,使其质量理想地相差 4 道尔顿,以避免同位素峰的重叠。可用同位素取代细微调节 MSM 基团中的变化;例如,用 4 个氘原子代替分子中的 4 个氢将产生 4 道尔顿的质量差别。

本发明质量标记的其它例子包括单体、寡聚或多聚形式的芳族、苯酚类、苯胺类及其异种类似物,含有 $C=C$ 或 $C\equiv C$ 或其异种类似物的其它部分,以及它们的寡聚或多聚对应物。必须避免含有 $C-H$ 或 $C-$ 卤素(非氟)的分子或其部分,除了以上讨论的聚酯,还可用硫醚、胺、磷酸酯、膦酸酯、硫代磷酸酯、硅烷、硅氧烷、磺酸酯、氨磺酰类和那些具有 $C=C$ 、 $C\equiv C$ 和 $C=N$ 的质量标记类似物。

当采用芳族或异种芳族时,它们可被取代或未取代。如果取代,则取代基必须也抗片段化,它可选自上述任一种类。

如上文讨论的,质量标记最好能抗片段化,并且最好应在 50 伏的电子电离条件下具有稳定性。

当标记从其核酸上断裂下来后用高分辨质量分析标记时,该技术的一个有利的实例是采用氟化的质量标记。整体质量为 100 碳氢化合物分子将有一个在级分上精度较高的质量。相反,整体质量为 100 的氟化分子有一个在级分上精度较低的质量。在高分辨质量分析中,质量上的这种差别是可以区分的,如果整体质量

相同但组成不同的两个分子有不同程度的氢-碳和氟-碳，则它们会在质谱中产生不同的峰。氟化分子被称为具有“质量亏损”。由于氟化分子在活系统中并不常见，因此这就意味着，即使有污染性峰(由于核酸片段化引起或来自缓冲液)存在，在质谱中仍可辨别氟化的质量标记，只要所用的核酸和试剂本身没有氟化。将质量亏损引入质量标记中的一种简单方法是掺入许多单位的氟化芳基醚(见图 8)。用一系列分开的质量亏损基团的另一种方案是用其氟化类似物替代正常芳基醚的聚合物。

氨基酸:

对于少量氨基酸(如甘氨酸、丙氨酸和亮氨酸)，可用本领域熟知的标准肽合成技术产生大量质量不同的短肽。氨基酸越多，可合成的标记就越多得多。这不必局限于天然的氨基酸。任两种手性形式以及不同的非天然侧链均可接受。(参见参考文献 31)。

实施例 1

负离子形成物质的合成

材料:

BSA(2-磺基苯甲酸环酐)-100 毫克, 0.54 毫摩尔

苄醇-2 毫升

碳酸钠-1.1 当量, 63 毫克

方法:

溶解碳酸盐和 BSA，加入苄醇。温热以开始反应(放出 CO₂)。搅拌直至泡腾停止。加入乙醚，过滤并沉淀产物。搅拌 10 分钟，过滤分离产物。产物是白色固体。该分子称为 AG/1/75(见图 11)。

质谱法: 负离子模式

图 11 中显示了前面合成的分子 AG/1/75 的负离子质谱。该质谱由 10ng/μl 的分子产生。溶剂是 1: 1 的甲醇和水。质谱由与电喷射输入系统连接的扫描四极矩质谱仪产生。插图显示了对应于 AG/1/75 分子阴离子(一种在 m/z 291 道尔顿 [M-Na]⁻带单个电荷的负离子)的质量峰。注意到在离准分子离子峰约三道尔顿内，同位素峰是显著的。

图 12 显示了 AG/1/75 的正离子谱。在该谱中没有可检测的分子性离子(molecular ion)，因此该分子最好用作负离子模式标记。上述两谱均用电喷射源中 45 伏的锥形电压产生。

图 13 显示了前述质谱同一溶液中的 AG/1/75 的负离子谱, 只是采用 75 伏的锥形电压。该电压足以在分子中产生明显的片段化, 在 m/z 156 道尔顿处产生一个主要片段负离子峰, 其对应于图 13 插入结构中所示位置处的断裂。

图 14 和 15 显示了正模式和负模式中不同溶液中的“非条件化(unconditioned)” PCR 产物的质谱。PCR 产物是“非条件化的”, 因为并没有进行努力将 DNA 从缓冲液和反应物中分离出来(这超出了凝胶电泳通常的做法)。没有打算将铵离子换成金属离子加成物或产生纯的 DNA(这些也是质谱法常用的实践)。图 16 和 17 显示与 AG/1/75 相同的 PCR 产物, 它在负离子模式中可清楚地检测到, 而在正离子模式中却检测不到。图 18 和 19 显示了信号加工后底物背景噪声的同一质谱, 显然在负离子模式中易检测到 AG/1/75。

实施例 2

带芳基醚质量标记的碱基的合成

下面是合成一系列胸腺嘧啶核苷酸的芳基醚的步骤。这些化合物的结构显示在图 24 和 25 中。

FT 9(见图 24)

用 N-甲基吗啉(27 μ L, 0.25 毫摩尔)和 2-氯-4,6-二甲氧基三嗪(44 毫克, 0.25 毫摩尔)处理以二氯甲烷(4 毫升)配的 5'-O-(4,4'-二甲氧基三苯甲游基-3'-琥珀酰基胸苷)(161 毫克, 0.25 毫摩尔), 全部试剂在室温下搅拌 1 小时。然后, 加入苯氧基苯酚(51 毫克, 0.27 毫摩尔), 连续搅拌 5 天。用二氯甲烷稀释反应混合物, 用柠檬酸水溶液(10% w/v)洗涤, 再用水洗两次。用硫酸钠干燥有机相, 然后减压除溶剂。用快速色谱(以含 1% 三乙胺的乙酸乙酯/正己烷(2:1)作为洗脱液)纯化残余物, 获得 86 毫克(得率为 42%)无色泡沫状 FT 9。 $^1\text{H NMR}(\text{CDCl}_3)$: δ 1.39(3H,m); 2.46(2H,m); 2.75(2H,m); 2.86(2H, m) 3.48(2H, m); 3.78(6H,s); 4.14(1H,m); 5.52(1H, m); 6.44(1H, m); 6.75-7.45(22H, m); 7.60(1H, d)。MS(FAB), m/z 812(M^+)。 $\text{C}_{47}\text{H}_{44}\text{N}_2\text{O}_{11}$ 计算值: C 69.44; H 5.46; N 3.46%, 测定值: C 69.66; H 5.53; N 3.24%。

FT 17(见图 24)

用三滴吡啶处理二氯甲烷(3 毫升)配的 5'-O-(叔-丁基二甲基甲硅烷基)-3'-琥珀酰基胸苷(288 毫克, 0.5 毫摩尔)液, 然后滴入二氯甲烷配的草酰氯(2M; 0.3mL, 0.6 毫摩尔)。在室温下搅拌反应混合物 90 分钟。将形成的酸性氯化物溶液滴加入

冰冷的二氯甲烷(3 毫升)配的 4-苯氧基苯酚(110 毫克, 0.59 毫摩尔)和吡啶(0.3 毫升)中。30 分钟后, 再加入一部分二氯甲烷(0.7 毫升)配的 4-苯氧基苯酚(135 毫克, 0.19 毫摩尔), 继续搅拌 4 小时。用二氯甲烷稀释反应混合物, 用碳酸氢钠水溶液(5% w/v)洗涤, 再用水洗两次。用硫酸钠干燥有机相, 然后减压除溶剂。用快速色谱(以乙酸乙酯/正己烷(1:1)作为洗脱液)纯化残余物, 获得 145 毫克(得率为 47%)无色泡沫状 FT 17。 ^1H NMR(CDCl_3): δ 0.12(6H); 0.92(9H); 1.92(3H,s); 2.12(1H,m); 2.40(1H,m); 2.77(2H,m); 2.89(2H); 3.90(2H,d); 4.11(1H,d); 5.30(1H,d); 6.36(1H,dd); 7.00-7.27(9H,m); 7.54(1H,d); 8.28(1H,br s)。MS (FAB) m/z 625 $[\text{M}+\text{H}]^+$ 。 $\text{C}_{32}\text{H}_{40}\text{N}_2\text{O}_9\text{Si}$ 的计算值: C 61.52; H 6.45; N 4.48%。测定值: C 61.60; H 6.45; N 4.45%。

FT 18/1(见图 25)

用三滴吡啶处理二氯甲烷(3 毫升)配的戊二酸 4-苯氧基苯酯(180 毫克, 0.6 毫摩尔), 然后滴加入二氯甲烷配的草酰氯溶液(2M; 0.35 毫升, 0.7 毫摩尔)。在室温下搅拌反应混合物 90 分钟。将形成的酸性氯化物溶液滴加入冰冷的二氯甲烷(3 毫升)配的 5'-O-(叔丁基二甲基甲硅烷基)胸苷(228 毫克, 0.5 毫摩尔)和吡啶(0.3 毫升)中。在室温下继续搅拌 5 小时。用二氯甲烷稀释反应混合物, 用碳酸氢钠水溶液(5% w/v)洗涤, 再用水洗两次。用硫酸钠干燥有机相, 然后减压除溶剂。用快速色谱(以乙酸乙酯/正己烷(1:1)作为洗脱液)纯化残余物, 获得 111 毫克(得率为 35%)无色油状 FT 18/1。 ^1H NMR(CDCl_3): δ 0.12(6H); 0.92(9H,s); 1.92(3H,s); 2.02-2.30(3H,m); 2.35-2.75(5H,m); 3.92(2H,d); 4.10(1H,d); 5.29(1H,d); 6.36(1H,dd); 6.97-7.37(9H,m); 7.54(1H,d); 8.65(1H,br s)。MS (FAB), m/z 639 $[\text{M}+\text{H}]^+$ 。 $\text{C}_{33}\text{H}_{42}\text{N}_2\text{O}_9\text{Si}$ 的计算值: C 60.35; H 6.75; N 4.26%。测定值: C 60.57; H 6.60; N 4.18%。

F23 (见图 25)

用三滴吡啶处理二氯甲烷(3 毫升)配的 5'-O-(叔丁基二甲基甲硅烷基)-3'-琥珀酰基胸苷(288 毫克, 0.5 毫摩尔), 然后滴入二氯甲烷配的草酰氯(2M; 0.3mL, 0.6 毫摩尔)。在室温下搅拌反应混合物 90 分钟。将形成的酸性氯化物溶液滴加入冰冷的二氯甲烷(3 毫升)配的(4'-苯氧基)-4-苯氧基苄醇(146 毫克, 0.5 毫摩尔)和吡啶(0.3 毫升)中。在室温下继续搅拌 4 小时。用乙酸乙酯稀释反应混合物, 用碳酸氢钠水溶液(5% w/v)洗涤, 再用水洗两次。用硫酸钠干燥有机相, 然后减压除溶

剂。用快速色谱(以乙酸乙酯/正己烷(1:1)作为洗脱液)纯化残余物, 获得 73 毫克(得率为 20%)FT 23。¹H NMR(CDCl₃): δ 0.13(6H,s); 0.92(9H,s); 1.92(3H,s); 2.11(1H,m); 2.39(1H,m); 2.68(4H,s); 3.90(2H,d); 4.06 (1H,d); 5.11(2H,s); 5.27(1H,d); 6.34(1H,m); 6.95-7.37(13H,m); 7.35(1H,d); 8.27(1H,br s)。MS (FAB) m/z 731 [M+H]⁺。C₃₉H₄₆N₂O₁₀Si 的计算值: C 64.08; H 6.34; N 3.85%。测定值: C 64.32; H 6.38; N 3.79%。

质量标记的碱基 FT23 的质谱法

以 FT23 作为质量标记碱基的行为模型, 在有和没有寡核苷酸背景下进行质谱法研究。这些研究的结果显示在图 20-23 中。每个图中显示了用 45 伏锥形电压的电喷射离子源在 Platform-LC 四极矩扫描质谱仪(Micromass UK)中产生的质谱。在每个例子中, FT23 以 4pmol/ μ l 形式存在。图 20 显示了负离子模式中的质谱, 729.3 处的突出的峰对应于[M-H]⁻离子。图 21 显示在正离子模式中相应质谱有多个突出的峰。

图 22 和 23 分别显示了在图 20 所示相同条件下产生的负离子和正离子模式质谱, 只是每种情况中还存在 4pmol/ μ l 的分子量约为 3000 的寡核苷酸样品。在负离子模式(图 22)中, 在 729.3 处可分辨出明显的峰。在正离子模式(图 23)中, 再次检测到有许多峰。

这些结果表明, 即使在等摩尔数(污染性)寡核苷酸存在下, 在负离子模式质谱法中很容易能检测到质量标记的 FT23。

参考文献

1. R.A.W. Johnstone 和 M.E. Rose, "化学家和生物化学家用质谱法", 第2版, Cambridge University Press, 1996
2. G. Jung 和 A. G. Beck-Sickinger, *Angew. Chem. Int. Edit.*, Engl. 31, 367-383
3. S. Brenner 和 R.A. Lerner, "编码的组合化学", *Proc. Natl. Acad. Sci. USA* 89, 5381-5383
4. M.J. Bishop 和 C.J. Rawlings, 编者, "核酸和蛋白质序列分析: 一种实用方法", IRL Press, Oxford, 1991
5. P.H. Nestler, P.A. Bartlett 和 W.C. Still, "组合化学编码文库分子标记的一种通用方法", *J. Org. Chem.* 59, 4723-4724, 1994
6. Z.-J. Ni 等, "用化学上坚固的次级胺标记编码组合性有机合成的多样化方法", *J. Med. Chem.* 39, 1601-1608, 1996
7. H.M. Geysen 等, "同位素或质量编码组合文库", *Chemistry and Biology* 3, 679-688, 1996 年 8 月
8. 英国专利申请 No. 9618544.2
9. A.C. Pease 等, *Proc. Natl. Acad. Sci. USA* 91, 5022-5026, 1994
10. U. Maskos 和 E.M. Southern, *Nucleic Acids Research* 21, 2269-2270, 1993
11. E.M. Southern 等, *Nucleic Acids Research* 22, 1368-1373, 1994
12. PCT/GB97/02403
13. 英国专利申请 No. 9620769.1
14. PCT/GB97/02722
15. WO97/27325
16. WO97/27327
17. WO97/27331
18. Lloyd-Williams 等, *Tetrahedron* 49:11065-11333, 1993
19. J.F. Milligan, M.D. Matteucci, J.C. Martin, *J. Med. Chem.* 36(14), 1923-1937, 1993
20. C.J. Guinosso, G.D. Hoke, S.M. Freier, J.F. Martin, D.J. Ecker, C.K. Mirabelle, S.T. Crooke, P.D. Cook, *Nucleosides Nucleotides* 10, 259-262, 1991
21. M. Carmo-Fonseca, R. Pepperkok, B.S. Sproat, W. Ansorge, M.S. Swanson, A.I. Lamond, *EMBO J.* 7, 1863-1873, 1991

22. (8)P.E. Nielsen, Annu. Rev. Biophys. Biomol. Struct. 24, 167-183, 1995
23. Zhen Guo 等, Nature Biotechnology 15, 331-335, 1997 年 4 月
24. Wetmur, Critical Reviews in Biochemistry and Molecular Biology, 26, 227-259, 1991
25. Sambrook 等, "分子克隆: 实验室手册, 第 2 版", Cold Spring Harbour Laboratory, New York, 1989
26. Hames, B.D., Higgins, S.J., "核酸杂交: 一种实用方法", IRL Press, Oxford, 1988
27. Gait, M.J.编者, "寡核苷酸合成: 一种实用方法", IRL Press, Oxford, 1990
28. Eckstein, 编者, "寡核苷酸和类似物: 一种实用方法", IRL Press, Oxford, 1991
29. Vogel's "有机化学课本", 第 4 版, B. S. Furniss, A.J. Hannaford, V. Rogers, P.W.G. Smith & A. R. Tatchell, Longman, 1978
30. 高级有机化学, J. March
31. E. Atherton 和 R.C. Sheppard, 编者, "固相肽合成: 一种实用方法", IRL Press, Oxford

附图图解

图 1

- 步骤 1: 产生被捕获在固相载体上的 cDNA, 例如用生物素化的 poly-T 引物
- 步骤 2: 用“参照酶”处理被保留的携带 poly-A 的 cDNA, 洗去松散的片段
- 步骤 3: 加入衔接子, 该衔接子的粘端与“参照酶”粘端互补并携带了“取样酶”的结合位点。衔接子还可携带使模板线性扩增的引物序列
- 步骤 4: 用“取样内切核酸酶”处理衔接的 cDNA, 洗去松散的片段
- 步骤 5: 加入衔接子, 该衔接子的粘端与“参照酶”的粘端互补并携带了“取样酶”的结合系列。该衔接子还应携带有可光断裂接头的质量标记
- 步骤 6: 加入“取样酶”
- 步骤 7: 取出签名片段已被释放到其中的液相, 连接到玻璃芯片分开的位置携带所有 256 种可能的 4 聚体的寡核苷酸阵列上
- 步骤 8: 将连接的签名包埋到 MALDI MATRIX 中。将带有连接签名的芯片转移到 MALDI 质谱仪中

步骤 9：用激光仪扫描芯片，使芯片上一个区域中的签名的质量标记断裂。用 UV 激光以第二频率扫描同一区域，使已经断裂的质量标记电离，以便进行质谱分析

图 2a

步骤 1：通过带有生物素标记的 poly-T 的基质，基质与亲和素包被的珠粒结合

步骤 2：用“参照内切核酸酶”处理被截留的携带 poly-A 的 cDNA，洗去松散的片段

步骤 3：加入衔接子，该衔接子的粘端与“参照酶”的粘端互补并携带“取样内切核酸酶”的结合位点

步骤 4：加入“取样酶”

步骤 5：加入衔接子，该衔接子的粘端与所有 4 种可能的碱基粘端互补并携带“取样内切核酸酶”的结合位点。这些衔接子还携带“质量标记”以确定它们所鉴定的多义性粘端的序列

图 2b

步骤 6：加入“取样酶”

步骤 7：取出签名片段已被释放入其中的液相，并分置入 256 个孔

步骤 8：使签名与孔中的珠粒连接。每个孔中含有的珠粒对应于一种可能的粘端。洗去每个孔中未连接的签名。

步骤 9：断裂固定化签名片段的质量标记，从而将其释放到液相中，用电喷射质谱法分析

图 3a

步骤 1：通过带有生物素标记的 poly-T 的基质，该 poly-T 与亲和素包被的珠粒结合

步骤 2：用“参照内切核酸酶”处理截留的携带 poly-A 的 cDNA，洗去松散的片段

步骤 3：加入衔接子，该衔接子的粘端与“参照酶”粘端互补并携带了“取样内切核酸酶”的结合位点

步骤 4：加入“取样酶”

步骤 5：加入衔接子，该衔接子的粘端与所有可能的 4 碱基粘端互补并携带了“取样内切核酸酶”的结合位点。这些衔接子还携带了“质量标记”来鉴定它们所鉴定的多义性粘端序列

图 3b

步骤 6: 加入“取样酶”

步骤 7: 取出签名片段已经释放入其中的液相, 并上样于 HPLC 亲和柱中, 根据粘端将片段分类成 256 个亚组

步骤 8: 柱应当将签名分类成携带相同粘端的组分。然后必须使这些组分暴露于激光来断裂质量标记

步骤 9: 然后可将断裂的质量标记和签名片段直接注入电喷射质谱仪内进行分析。标记的电荷可设计成与寡核苷酸签名相反。因此, 如果标记为负电荷, 则可用负离子质谱法分析标记。

图 4

A 离子源

B 离子流

C 电门

D 反射器(reflectron)

E 检测器

(1)、(2)是较佳的可光断裂接头

图 8

(1)-(3)是较佳的质量标记结构物, 其中 n 大于等于 0

(4) 含质量标记的质量亏损, 其中 n 大于等于 0, m 大于等于 0, X 宜为 F 或 H

图 9

(1) 较佳的终端变量或质量系列修饰基团

(2) 较佳的内部变量或质量系列修饰基团, 其中 n 大于等于 0, R 可以是任意基团。对于质量系列修饰基团, R 基团最好应不电离或片段化。电离基团显示在另一图中。

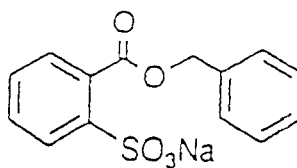
图 10

(1) 负离子模式基团

(2) 正离子模式基团

图 11

图例: 样品 AG/1/75, 10ng/ μ l, 1: 1 MeOH:水, CV=45V LIVER01 1 (0.997) Sm (SG, 2x0.60), Scan ES-1.79e8, 其中 AG/1/75 是



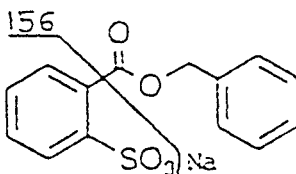
(M-314)

图 12

图例: AG/1/75 5×10^{-7} M 20 μ l/分 在 MeOH/H₂O 1:1 LPOOL3 13(0.496) Cm(9:13)中输注, Scan ES+1.89e6

图 13

图例: 样品 AG/1/75, 10ng/ μ L, 1:1 MeOH:水, CV=75V LIVER02 1(0.998) Sm(SS, 2x0.60), Scan ES-4.37e7, 其中 AG/1/75 是



(M-314)

图 14

图例: DNA 1:5D 在 MeOH:H₂O+0.2%甲酸中 45 伏+/-开关
(1): LPOOL5 9(0.628) Cm(2:13), 1:Scan ES-4.56e3
(2): LPOOL5 3(0.243) Cm(3:10), 2:Scan ES+1.13e5

图 15

图例: DNA 1:5D 在 MeOH:H₂O+0.2%氨中 45 伏+/-开关
(1): LPOOL6 11(0.761) Cm(4:12), 1:Scan ES-1.37e4
(2): LPOOL6 10(0.726) Cm(2:11), 2:Scan ES+8.13e4

图 16

图例: DNA+AG/1/75+0.2%甲酸 LOOP INJ+/-ES
(1): LPOOL9 14(0.800) Cm(11:18), 2:Scan ES+1.04e6
(2): LPOOL9 14(0.771) Cm(12:18), 1:Scan ES-4.20e3

图 17

图例: DNA+AG/1/75+0.2%甲酸 LOOP INJ+/-ES
(1): LPOOL10 13(0.747) Cm(11:17), 2:Scan ES+1.86e6

(2): LPOOL10 11(0.608) Cm(11:17), 1:Scan ES-3.23e3

图 18

图例: DNA+AG/1/75+0.2%甲酸 LOOP INJ+/-ES

(1): LPOOL9 14 (0.800) Cm(13:15-(22:29+4:7)), 2:Scan ES+1.02e6(减去背景)

(2): LPOOL9 16 (0.881) Cm(16:19-(23:29+9:13)), 1:Scan ES-2.70e3(减去背景)

图 19

图例: DNA+AG/1/75+0.2%氨 LOOP INJ+/-ES

(1): LPOOL10 13 (0.747) Cm(13:14-(6:8+22:25)), 2:Scan ES+2.93e6(减去背景)

(2): LPOOL10 11 (0.608) Cm(11:16-(8+26)), 1:Scan ES-1.03e3(减去背景)

图 20

图例: FT23(仅仅)(-ve 离子)4pmol/ μ l, LPOOL2 3(0.266) Cm(2:24), 2:Scan ES-7.35e5

图 21

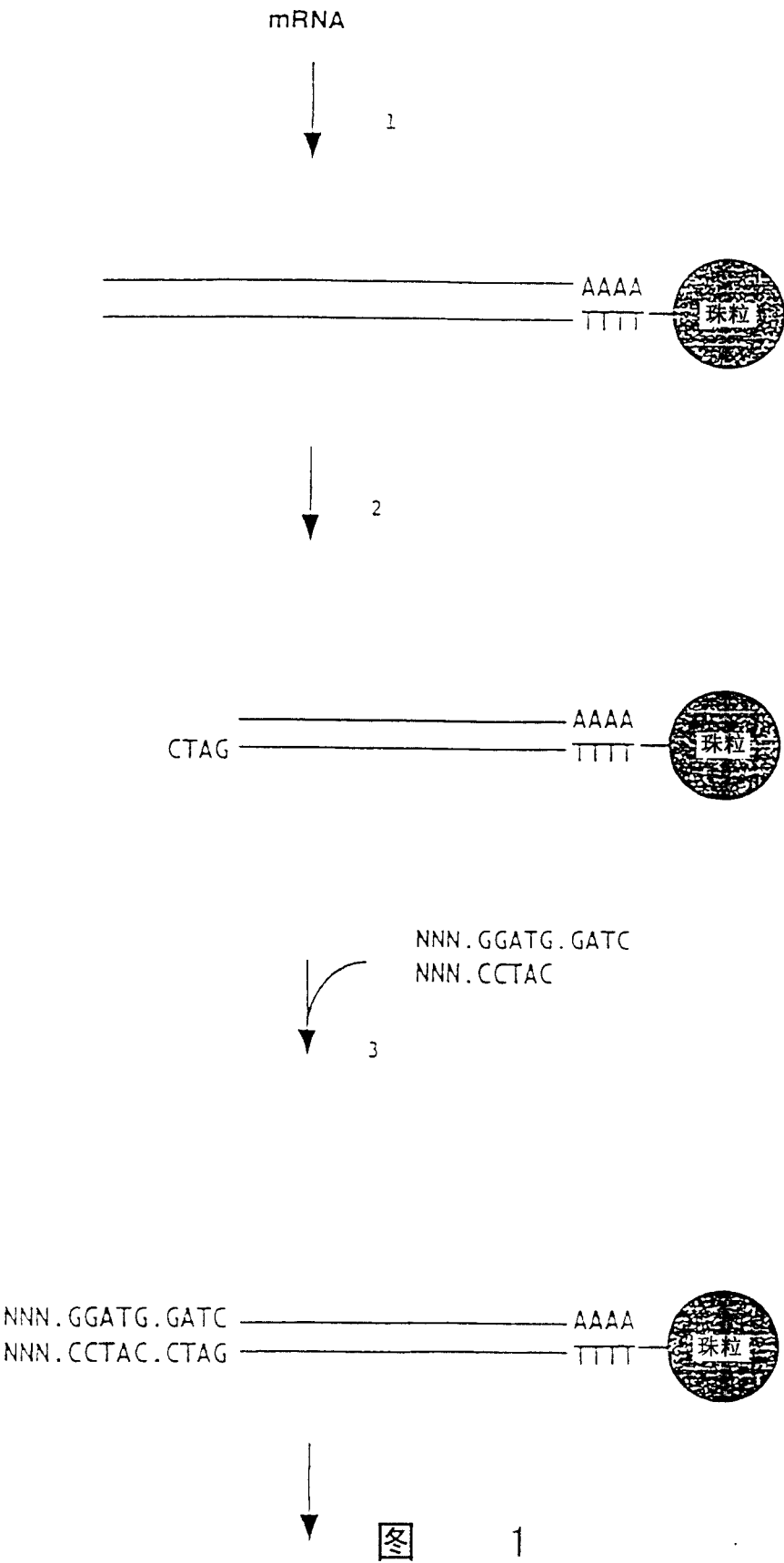
图例: FT23(仅仅)(+ve 离子)4pmol/ μ l, LPOOL2 5 (0.381)Cm(2:24), 1:Scan ES+3.68e6

图 22

图例: F23/OLIGO(-ve 离子)4pmol/ μ l, LPOOL1 18(1.405) Cm(3:25), (寡核苷酸分子量为 3000), 2:Scan ES-3.21e5

图 23

图例: F23/OLIGO(+ve 离子)4pmol/ μ l, LPOOL1 11(0.830) Cm(4:26), (寡核苷酸分子量为 3000), 1:Scan ES+2.03e6



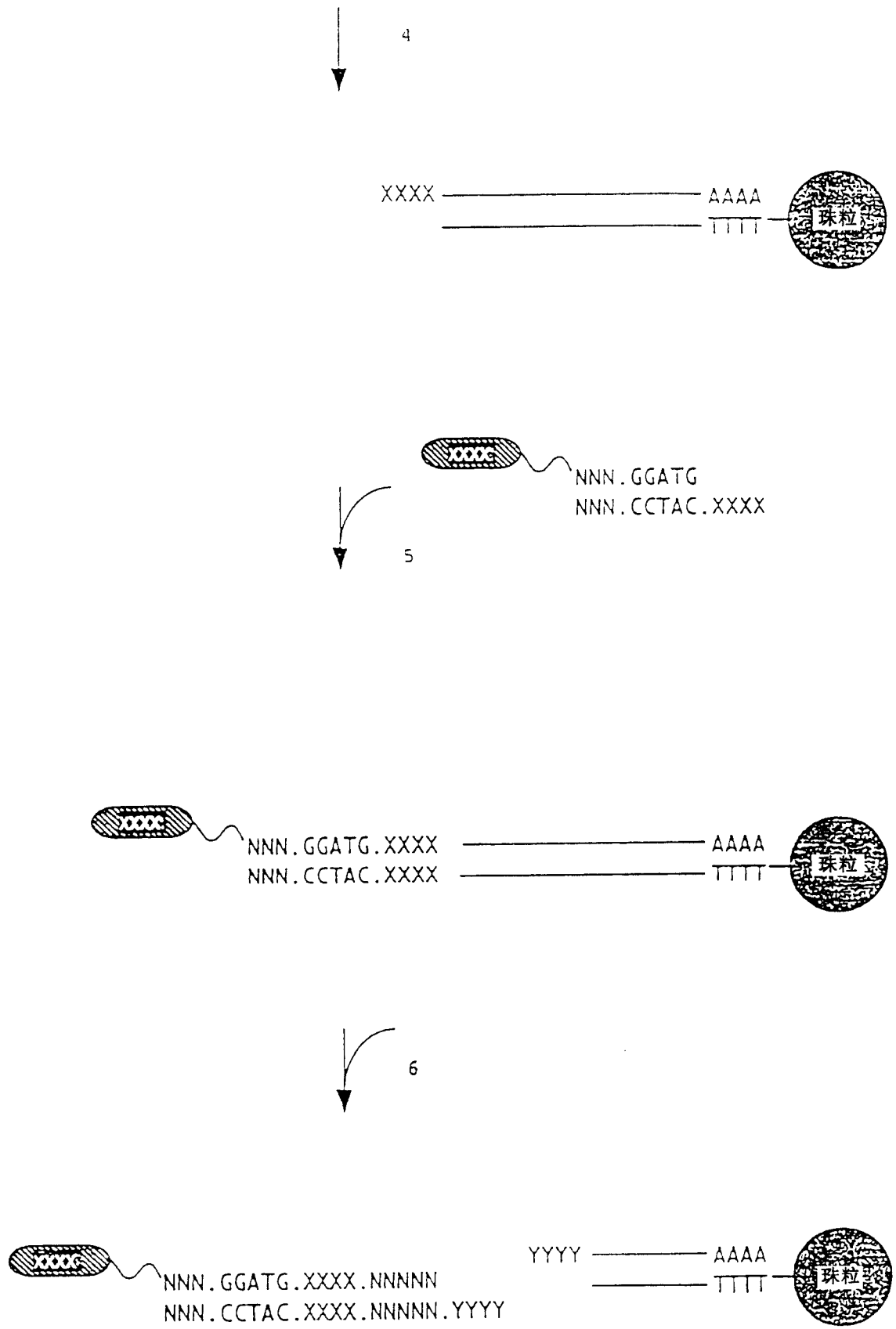


图 1续

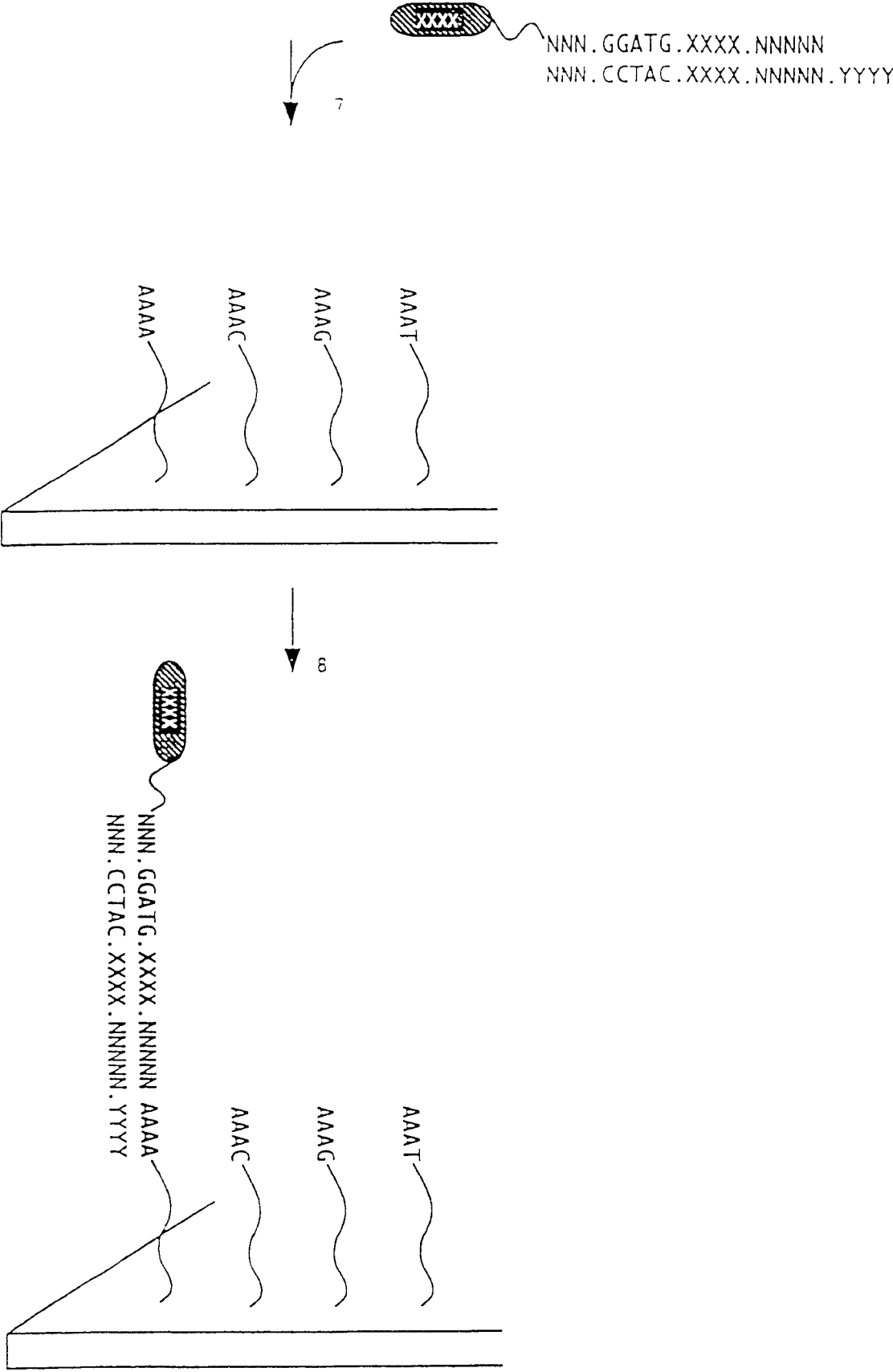


图 1 续

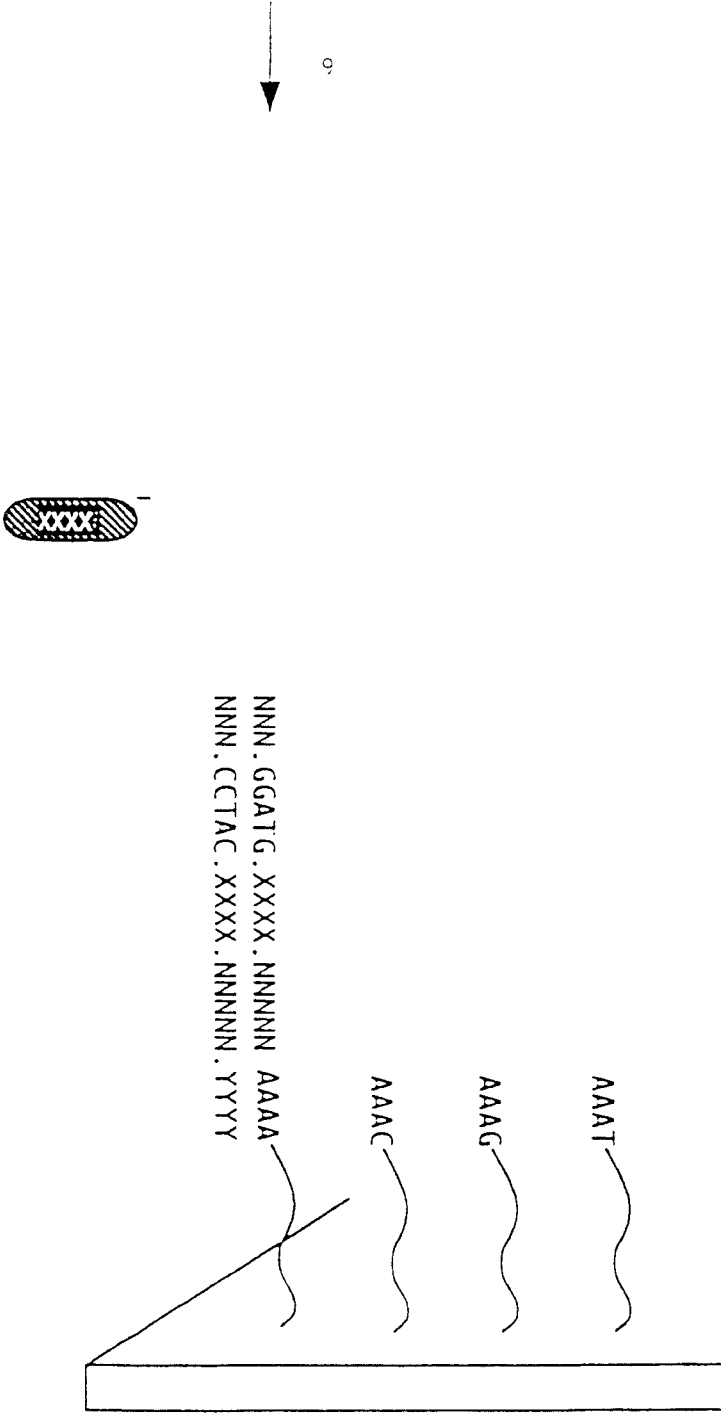


图 1续

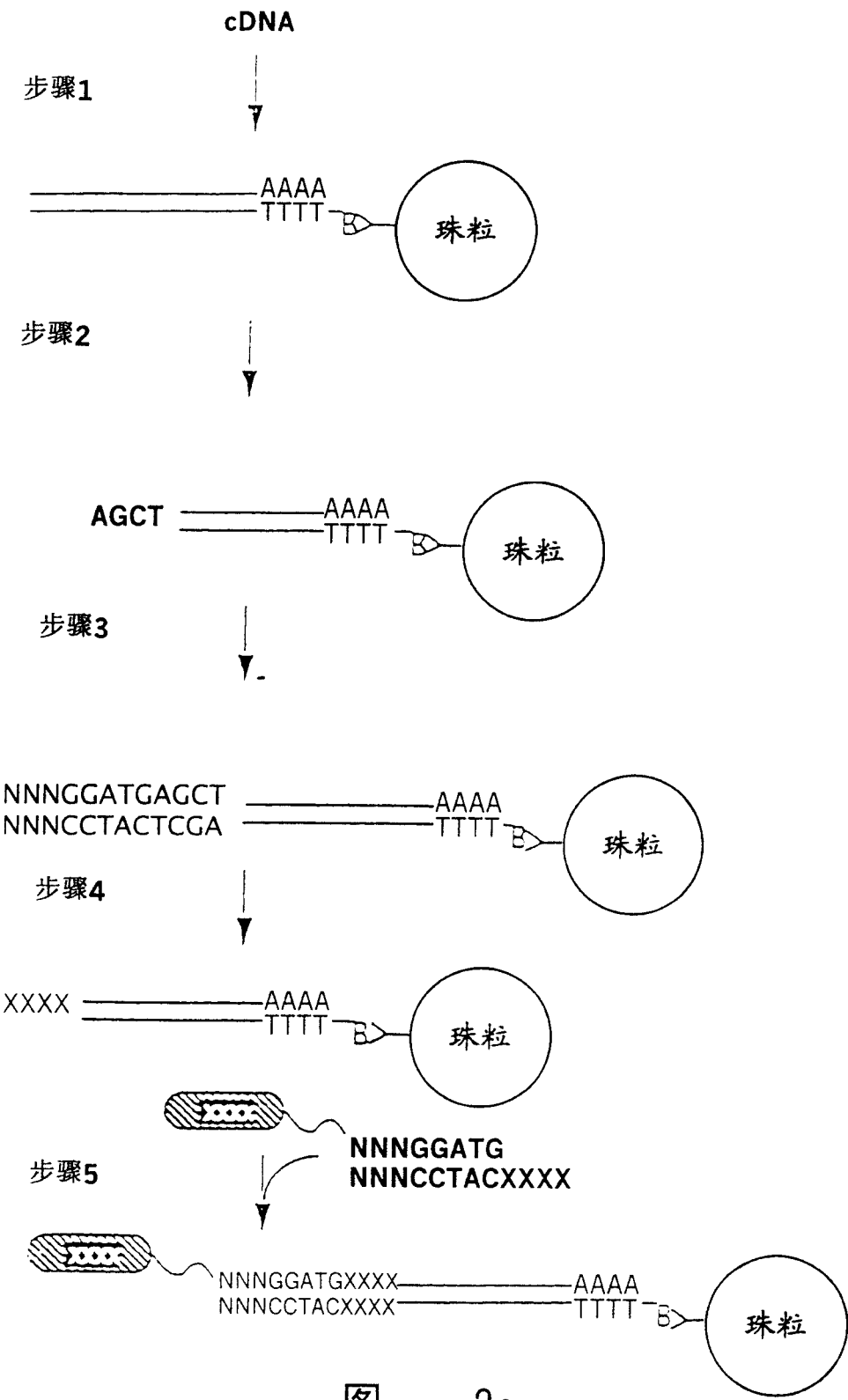


图 2a

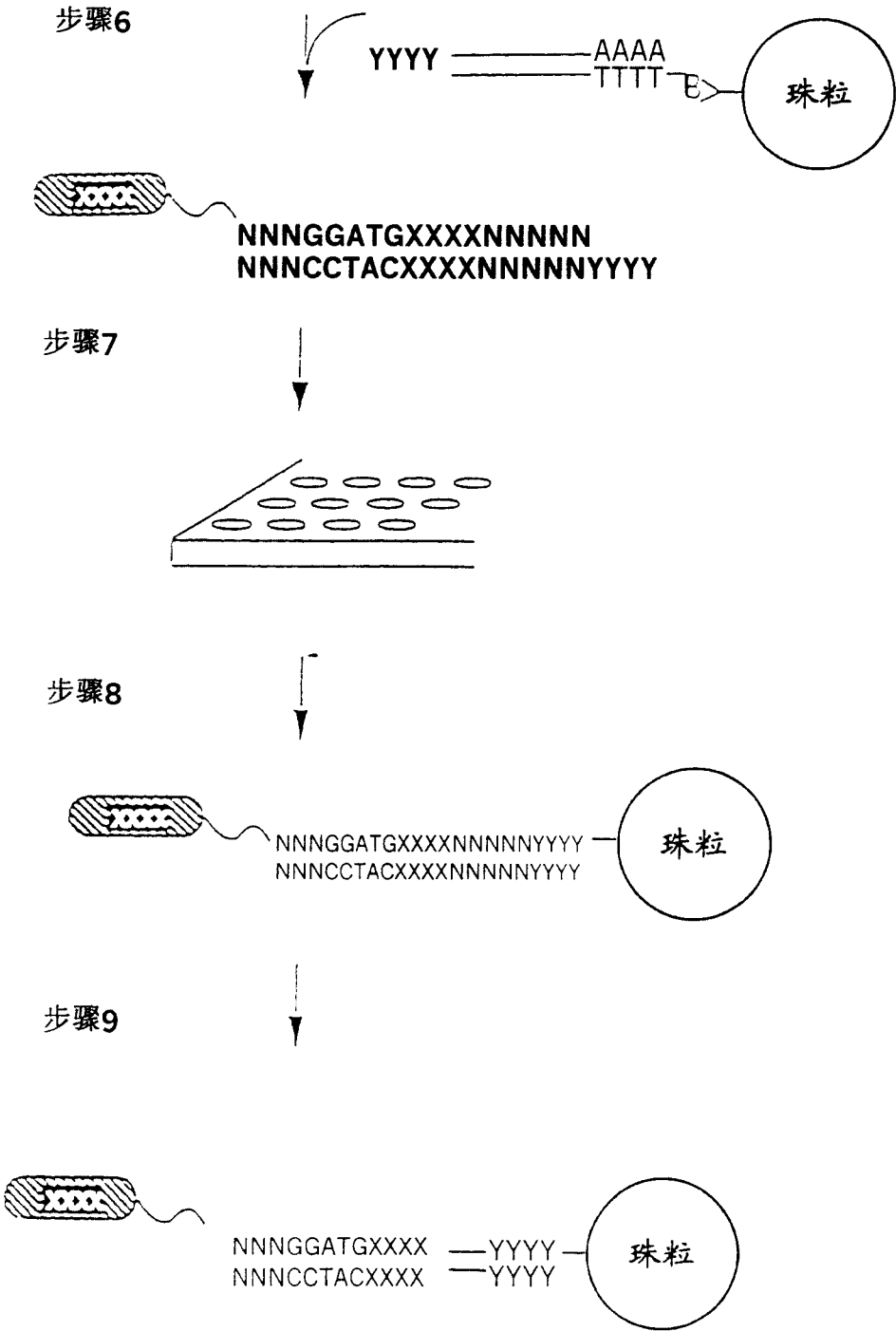


图 2b

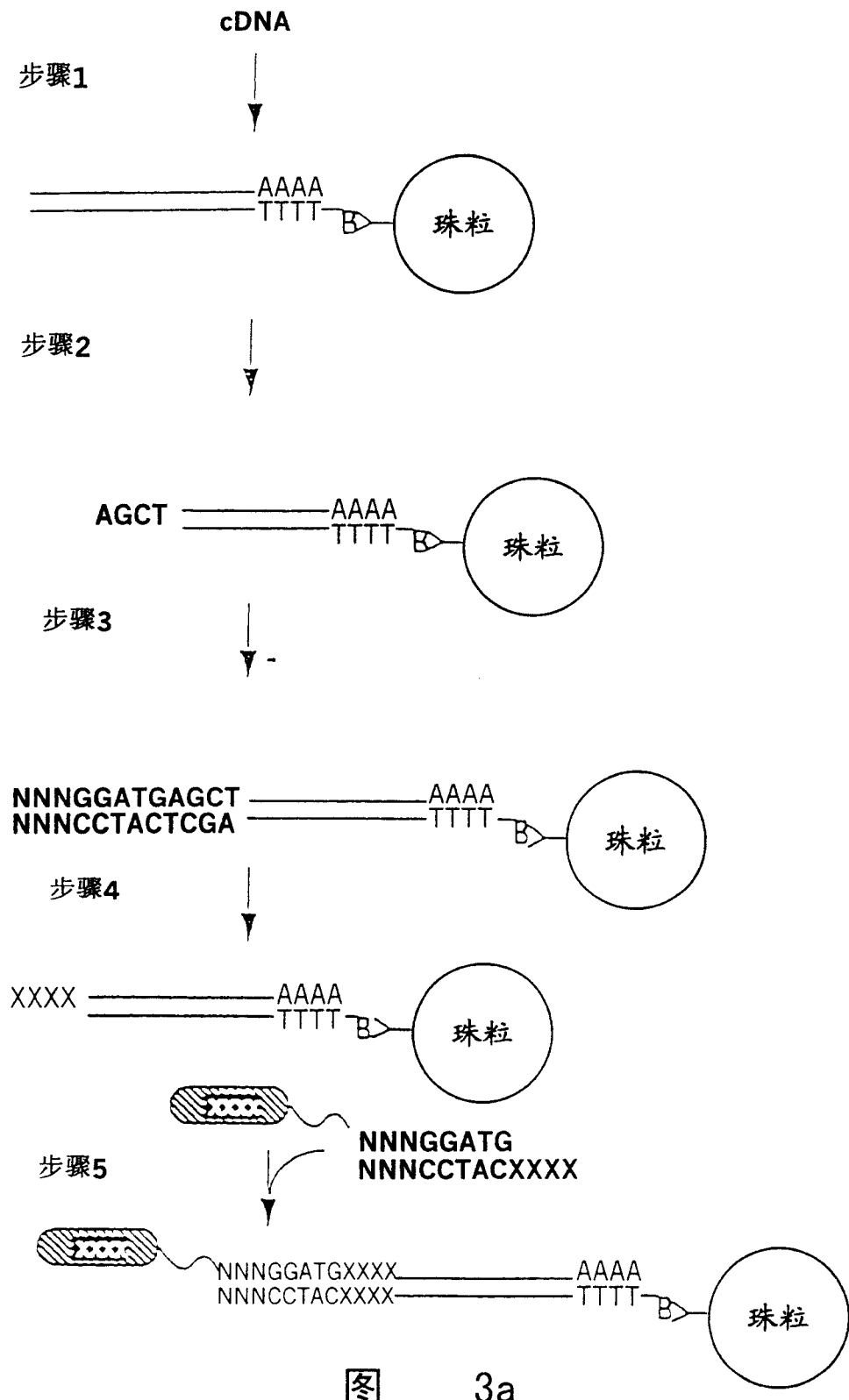
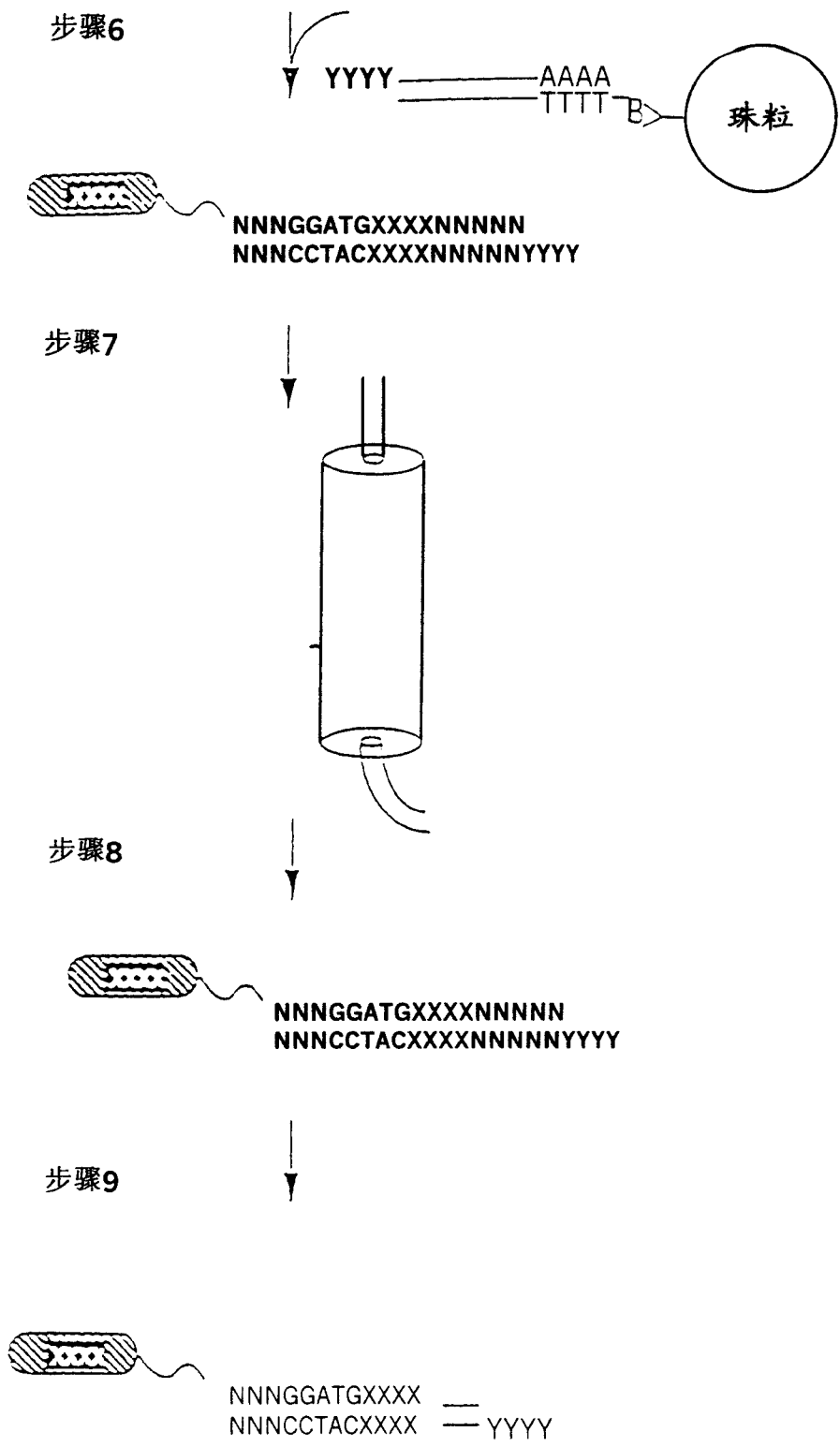


图 3a



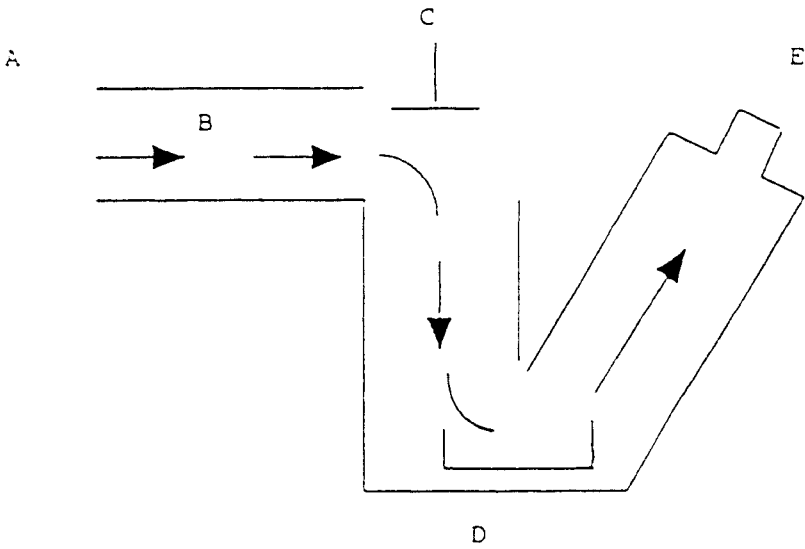


图 4

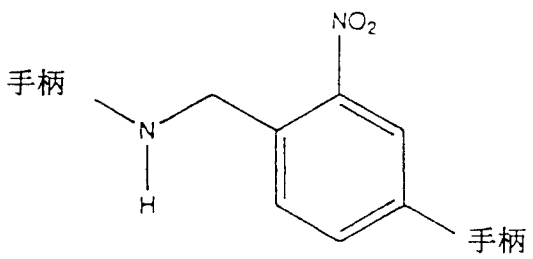
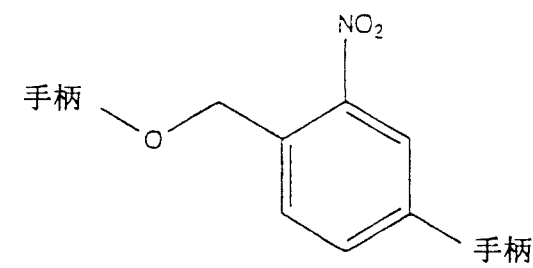


图 5

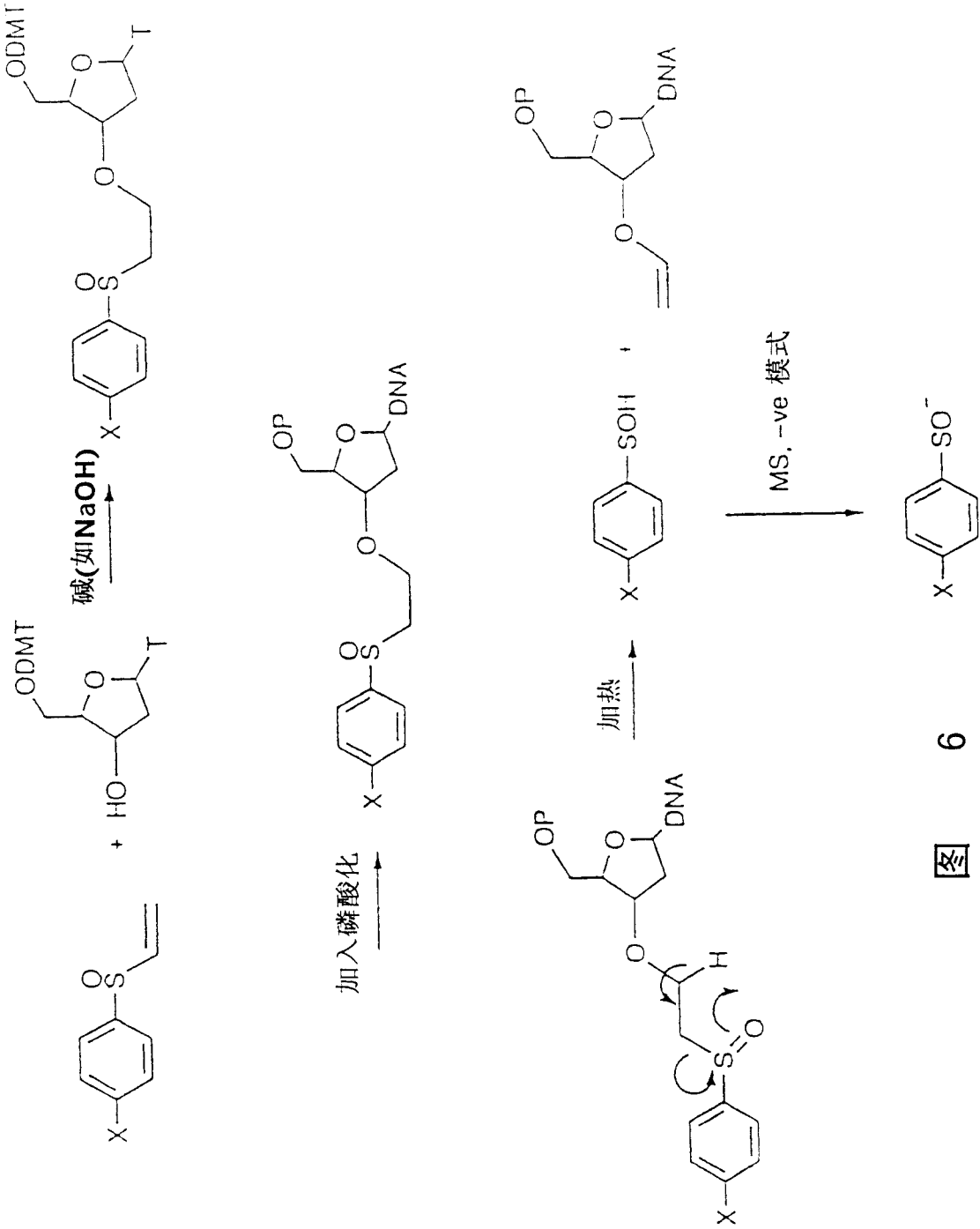


图 6

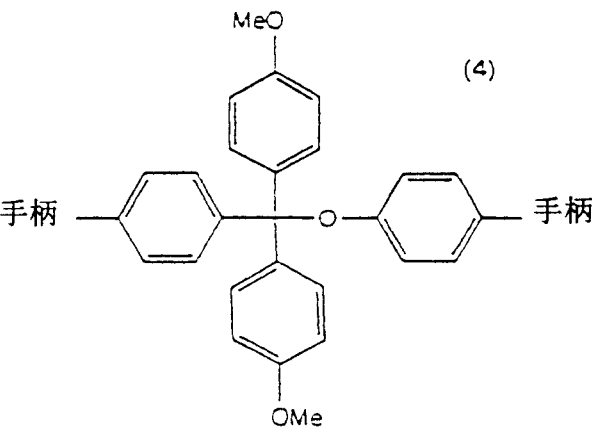
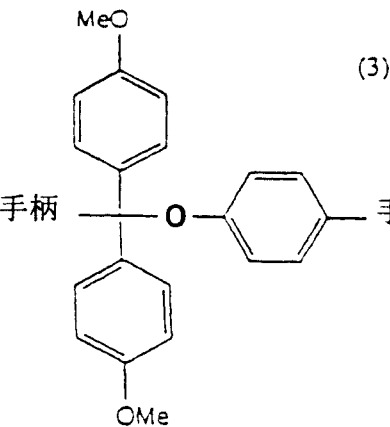
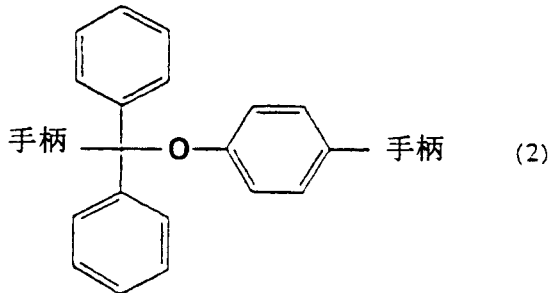
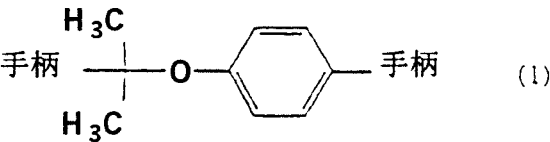
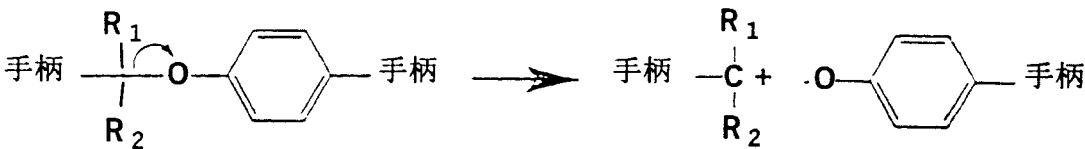


图 7

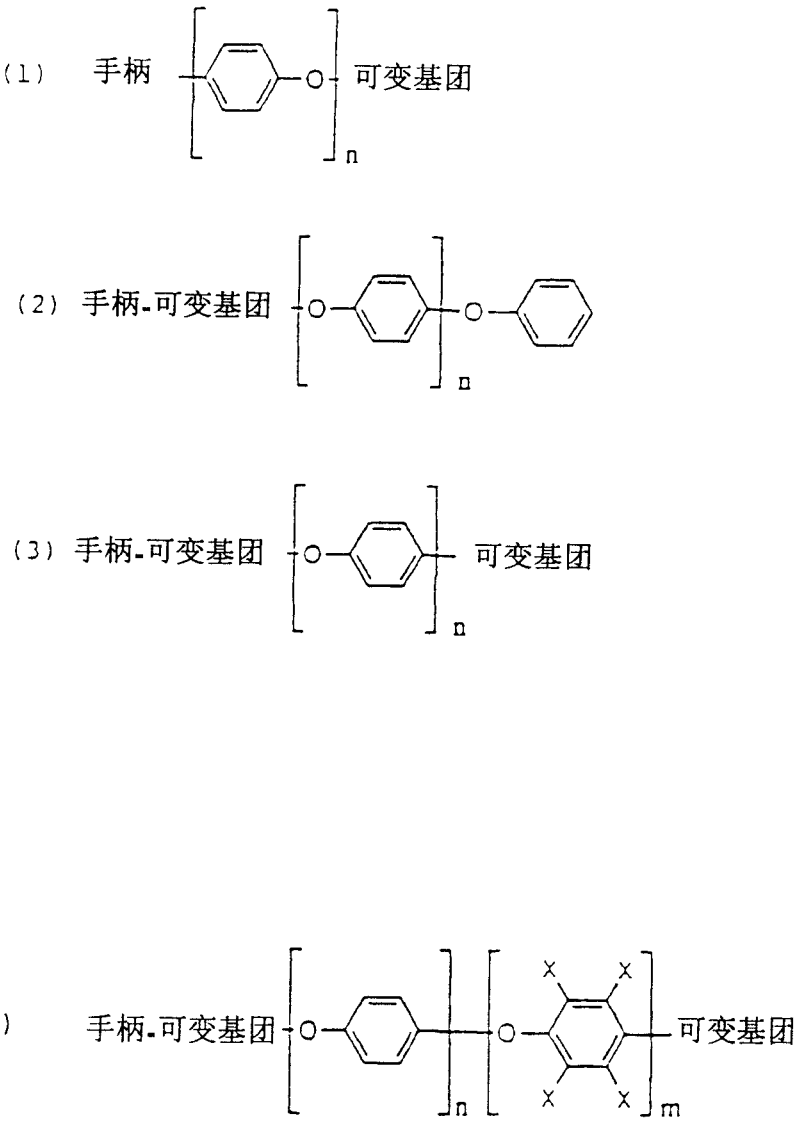


图 8

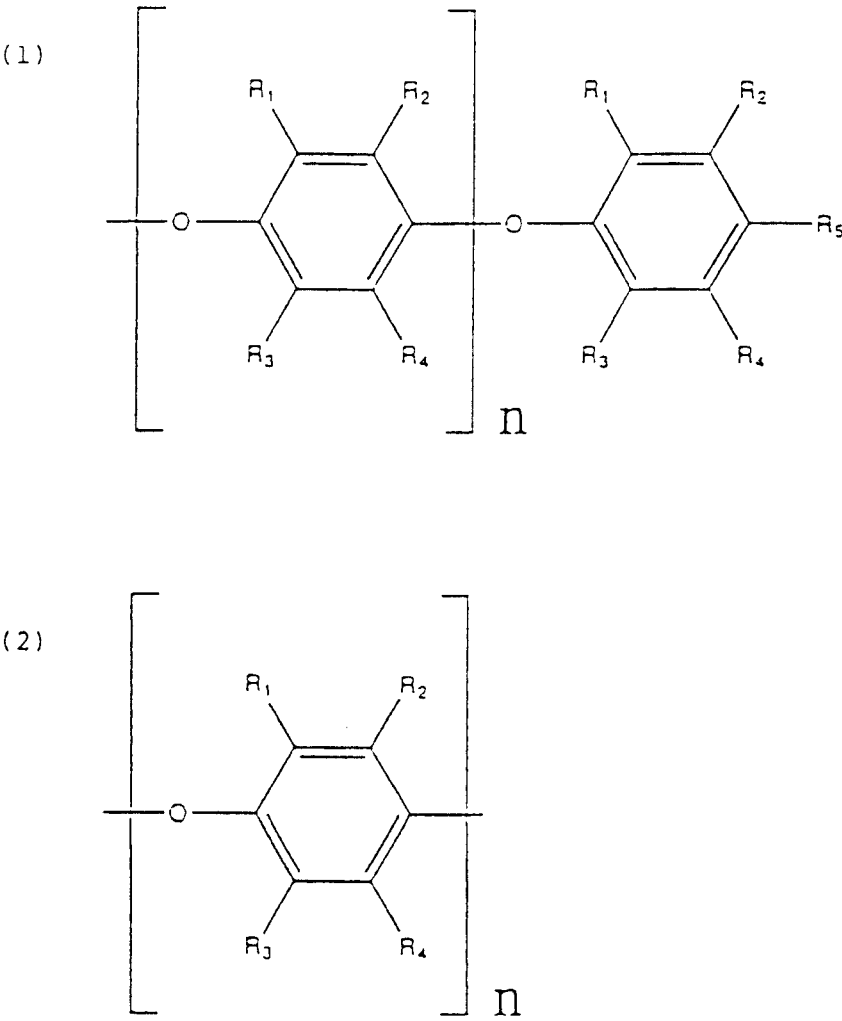


图 9

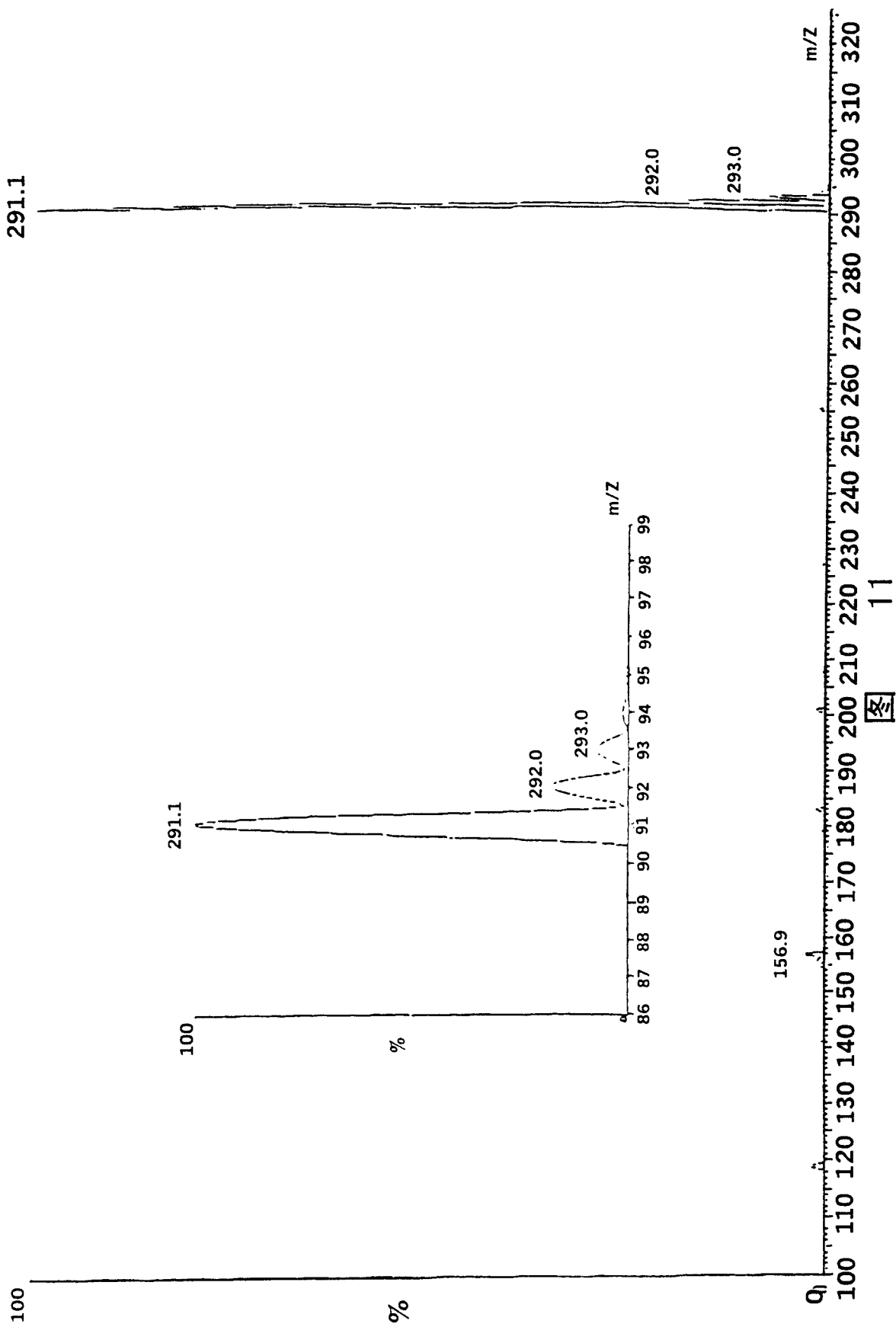


图 11

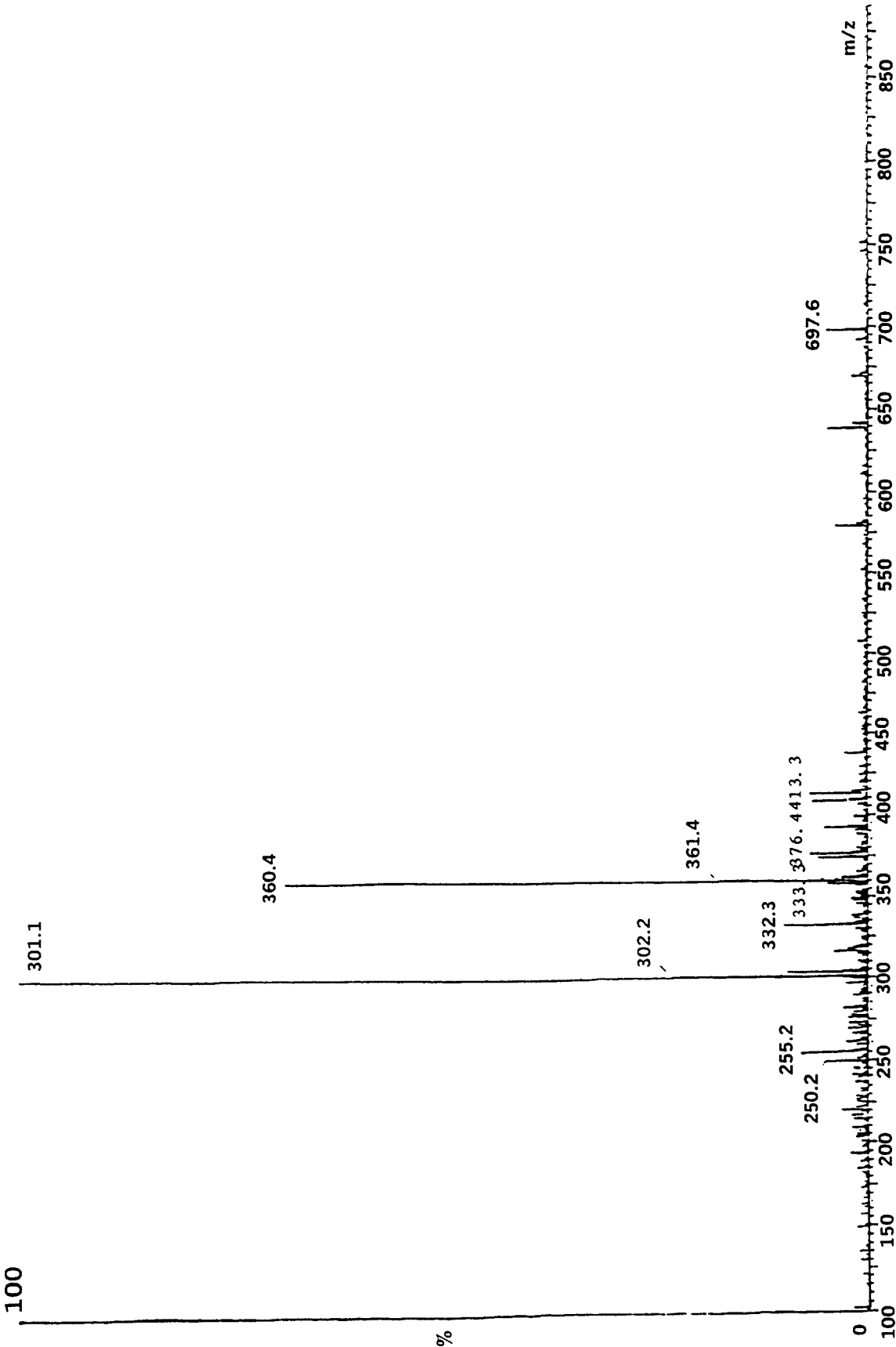


图 12

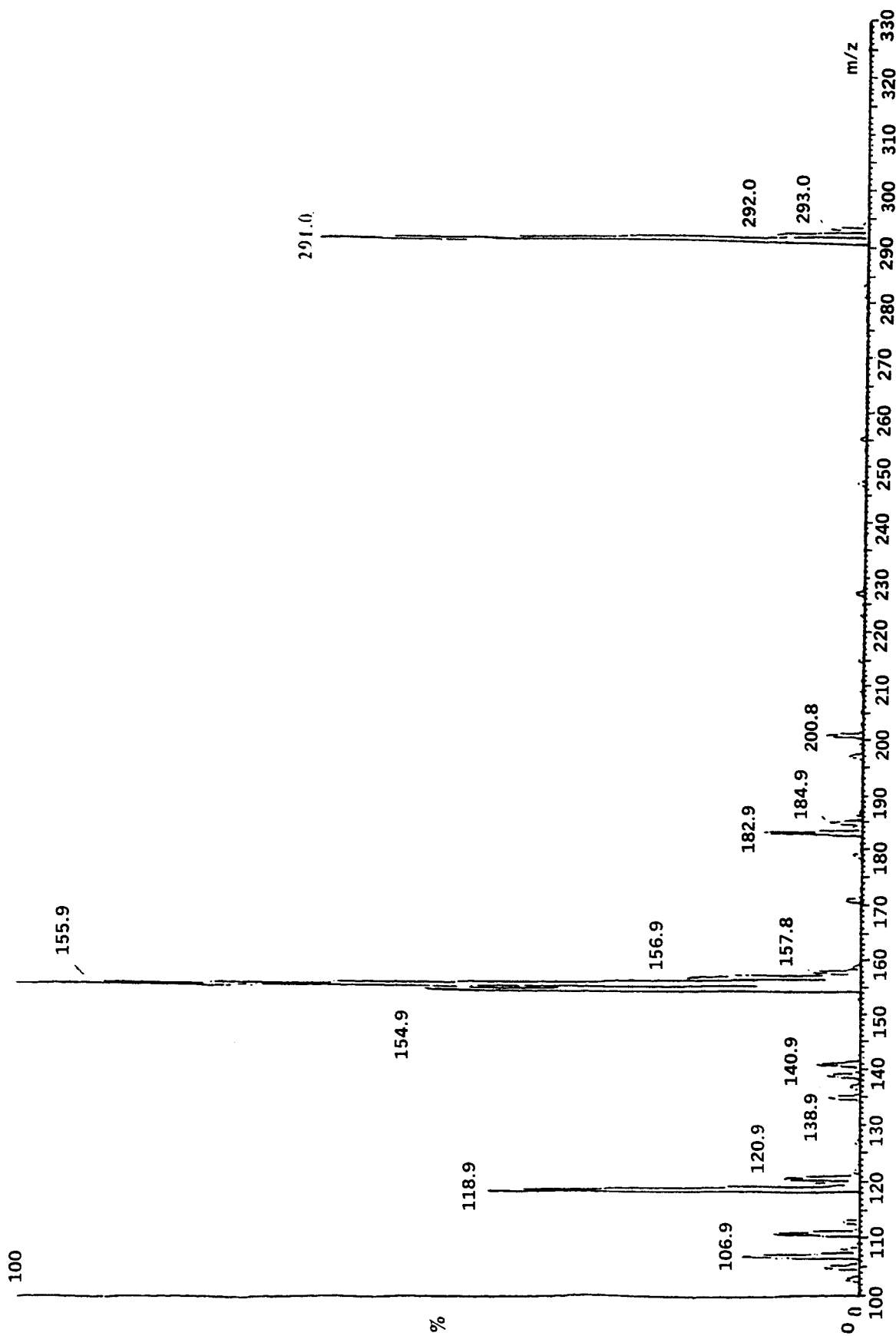


图 13

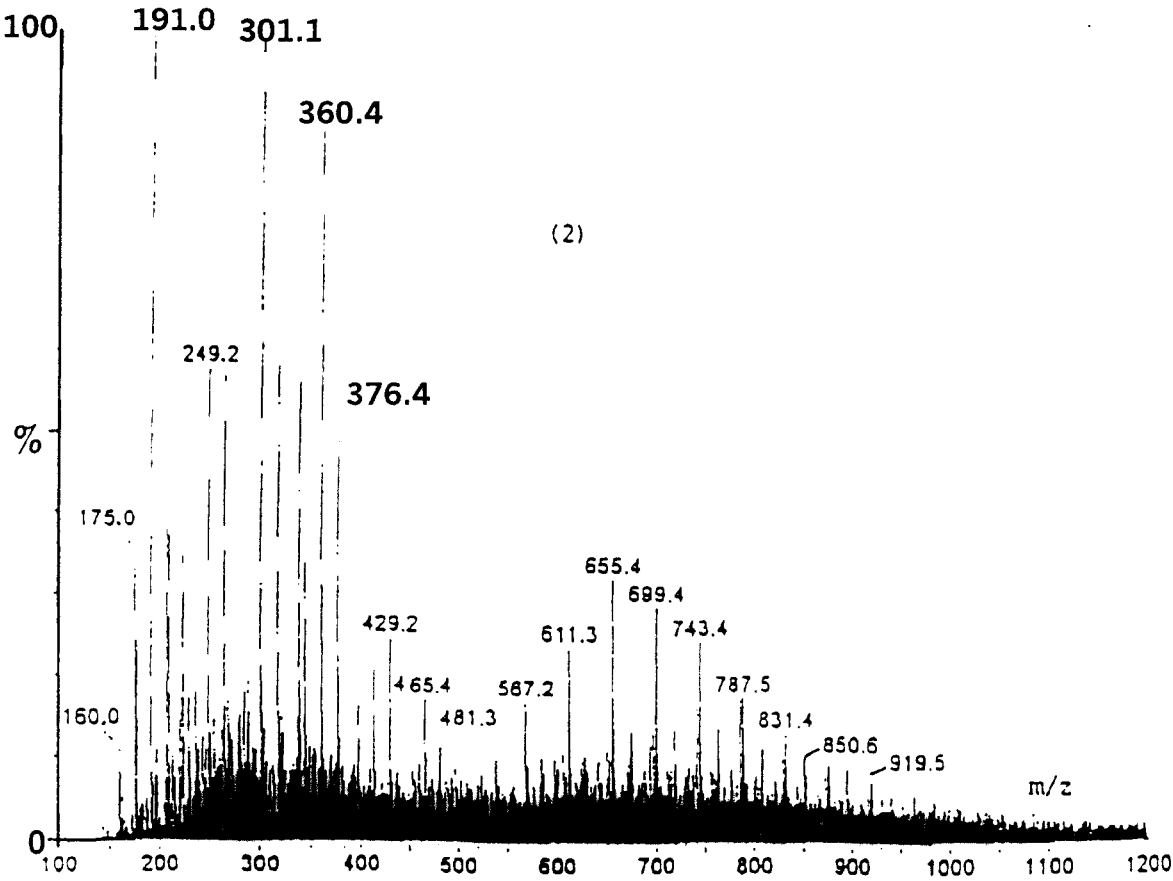
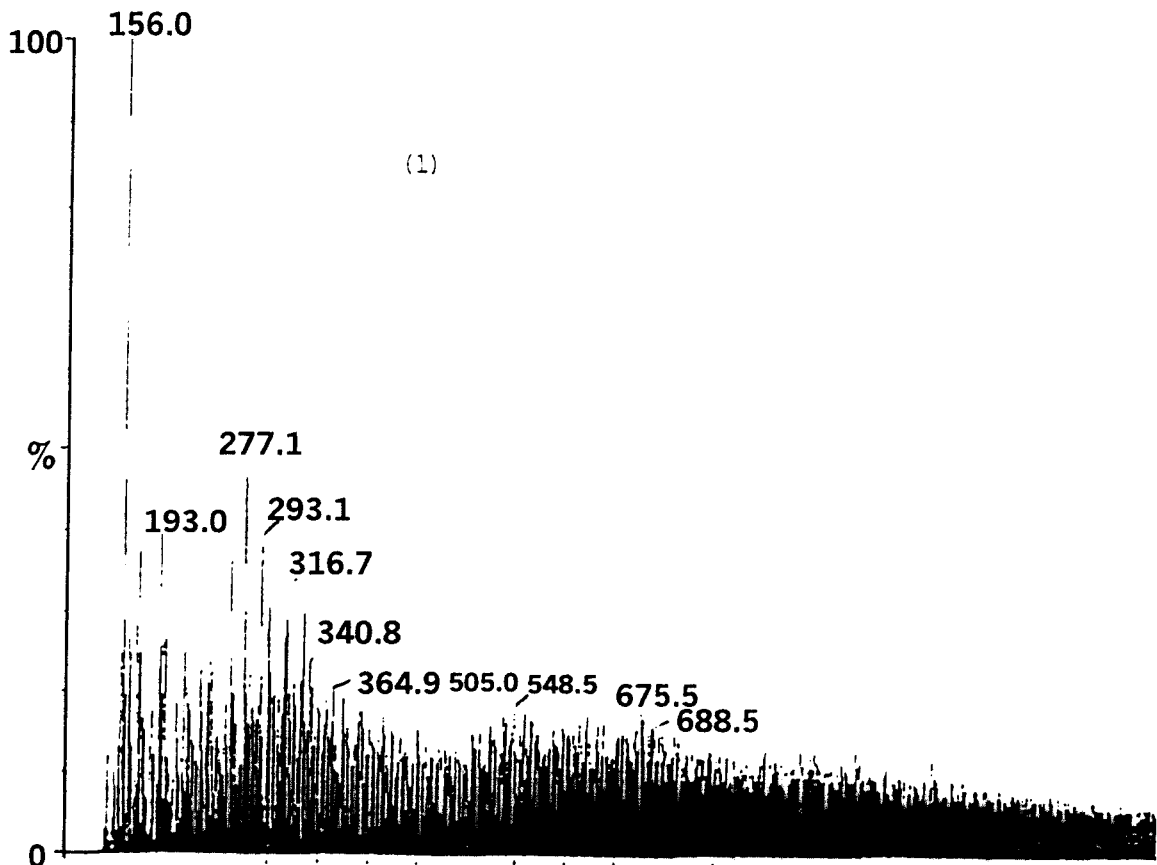


图 14

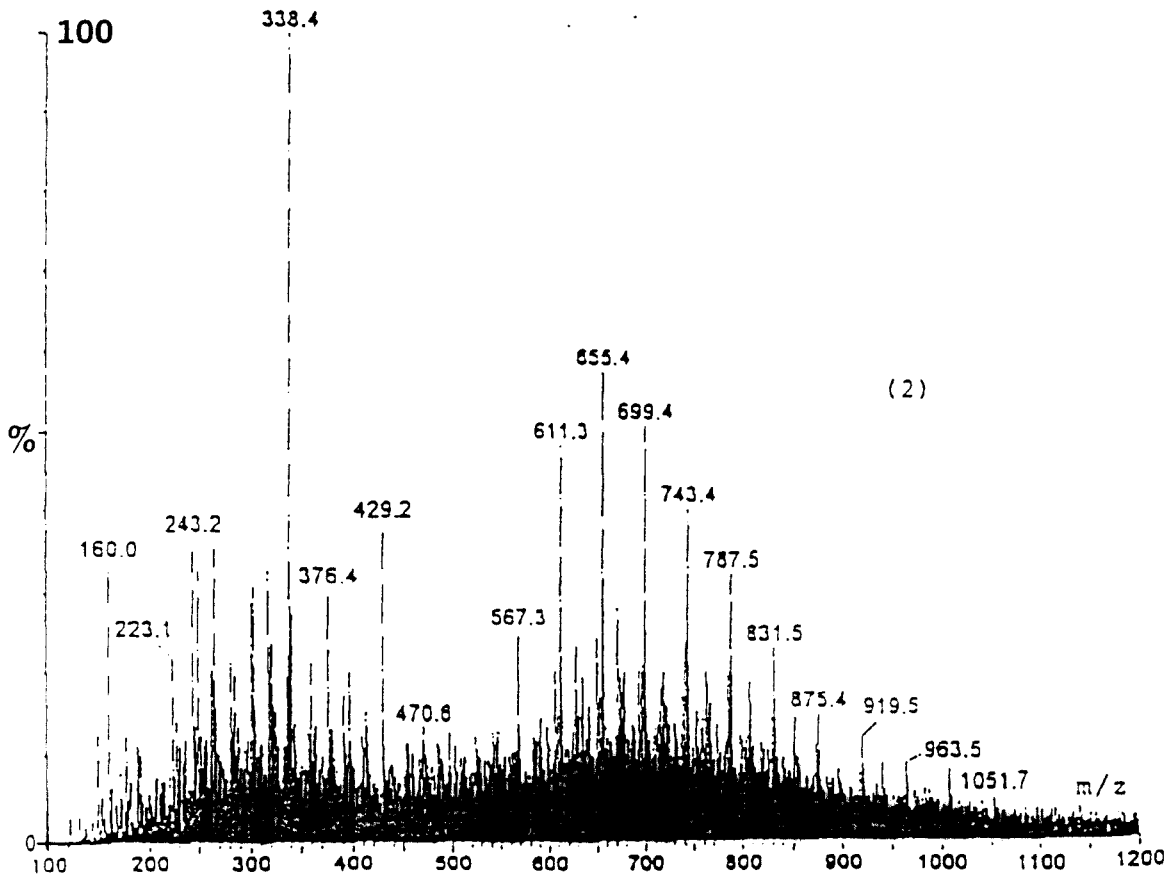
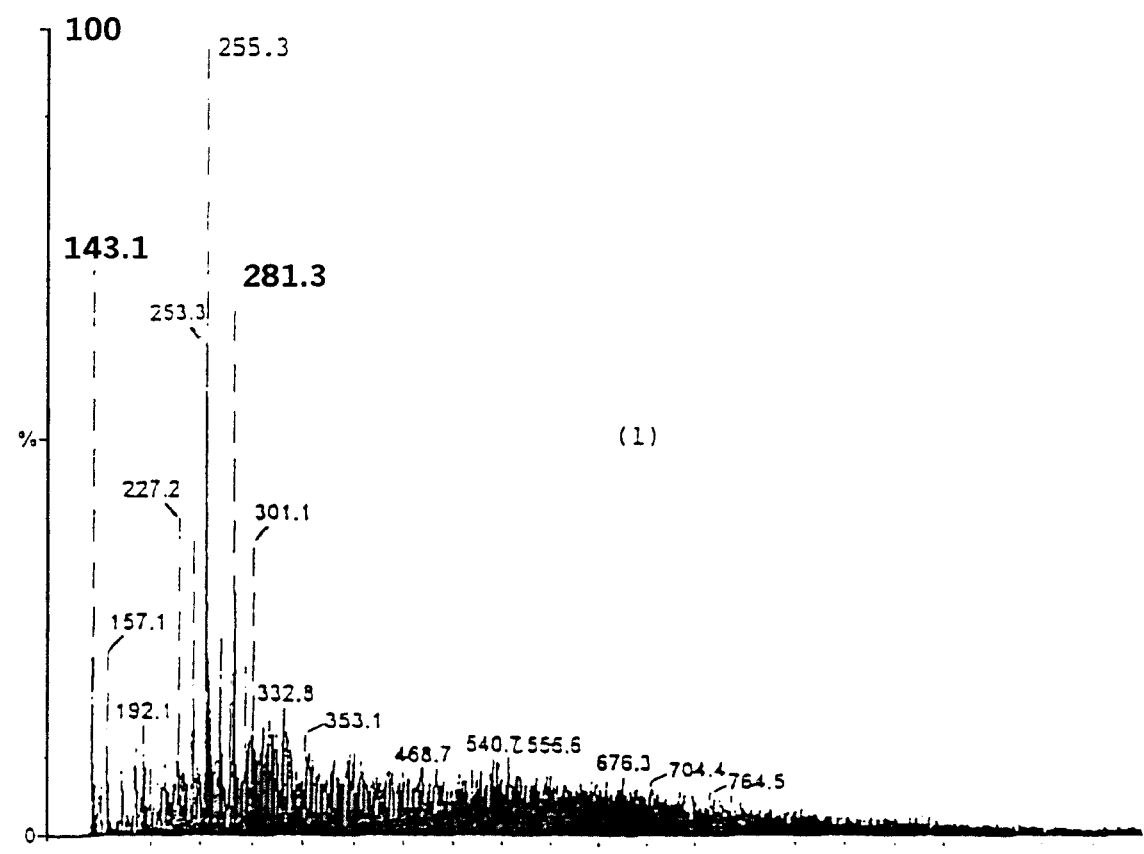


图 15

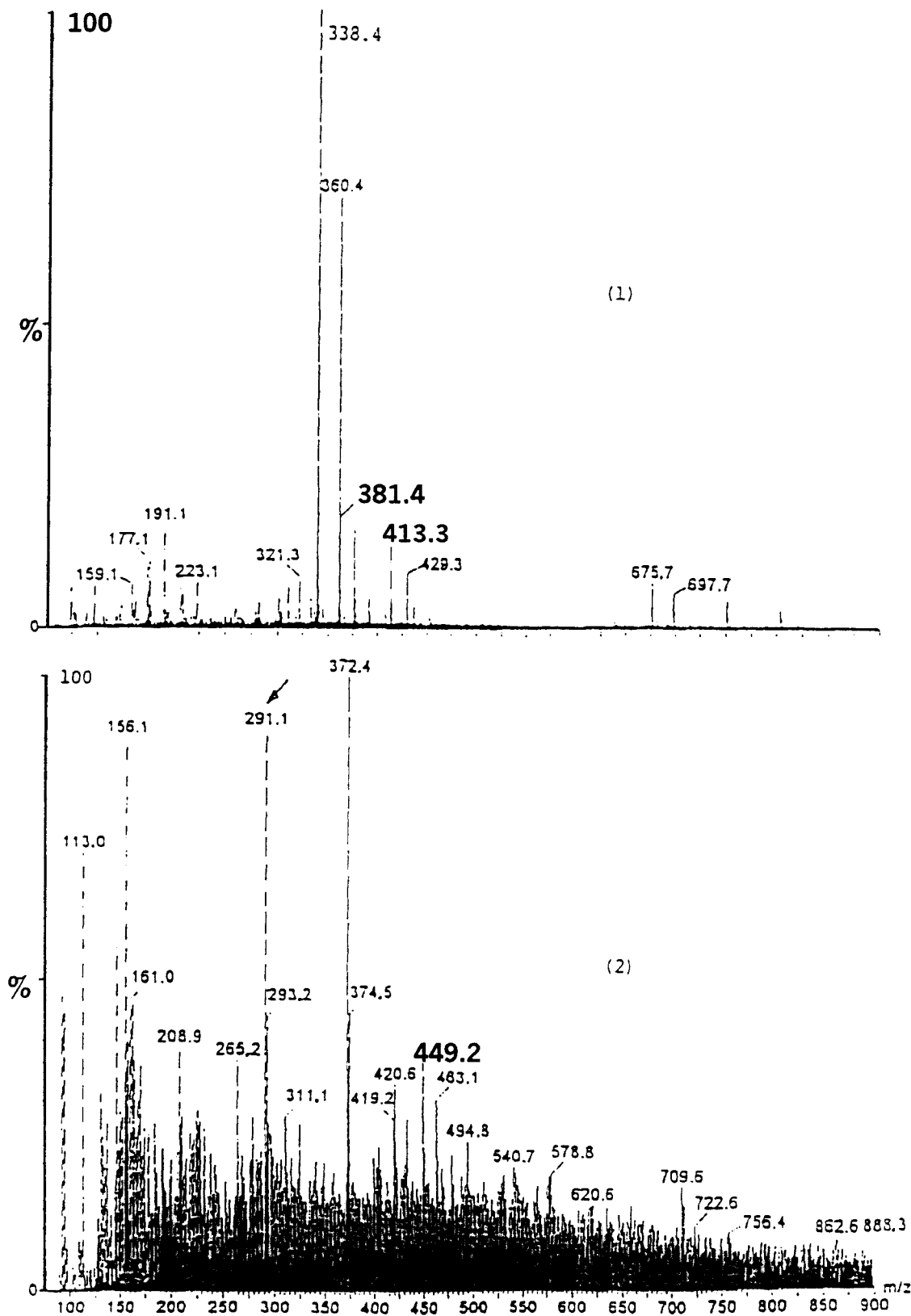


图 16

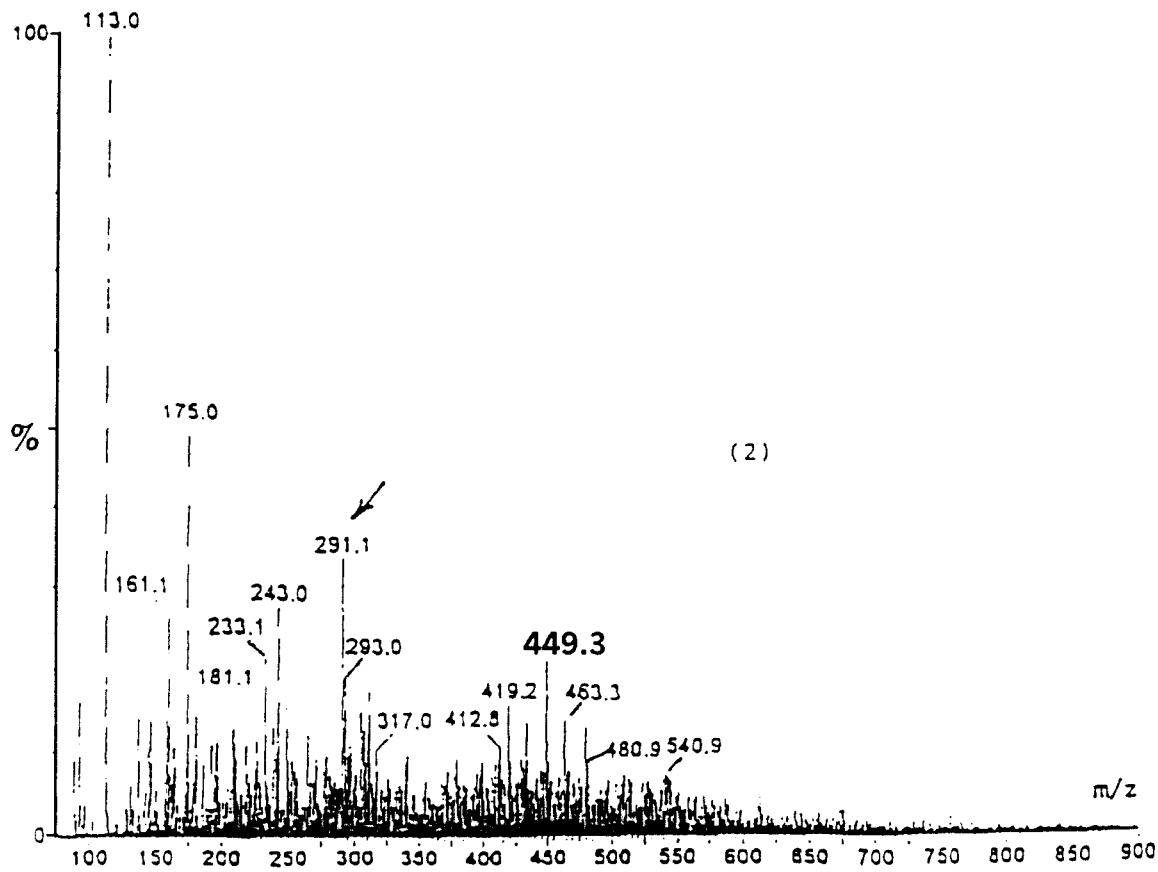
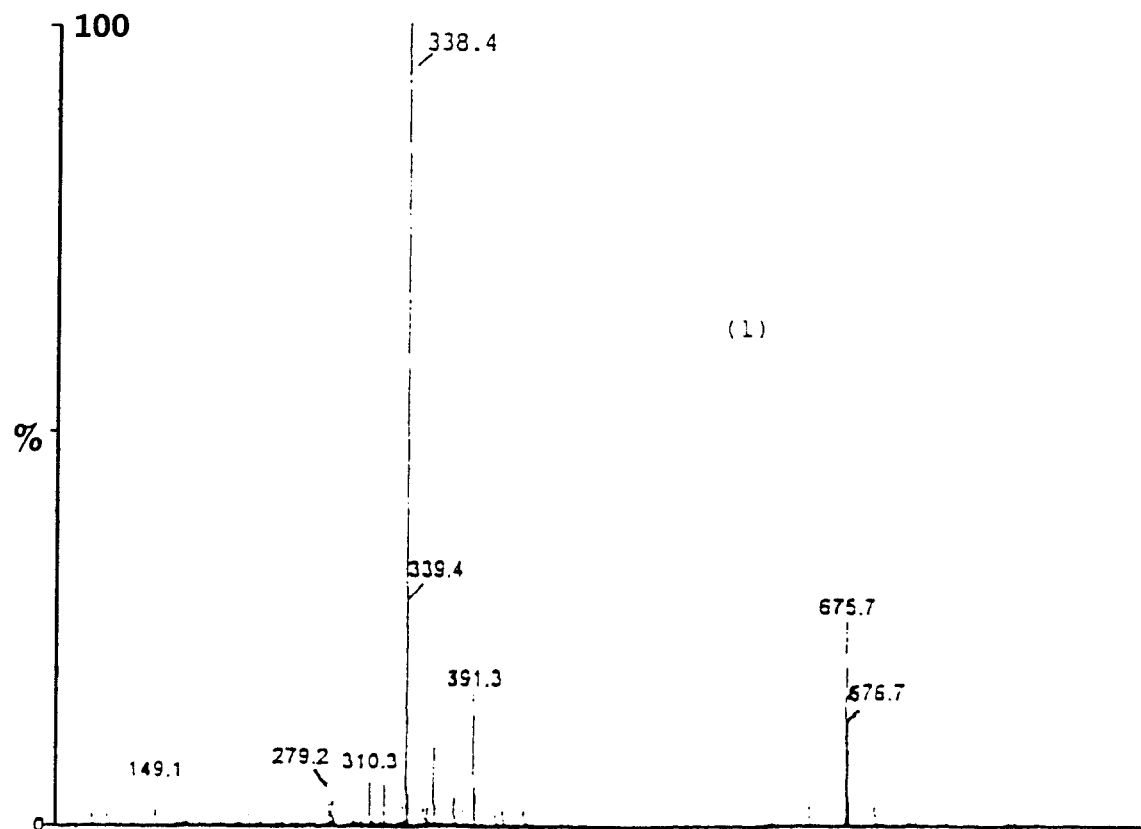


图 17

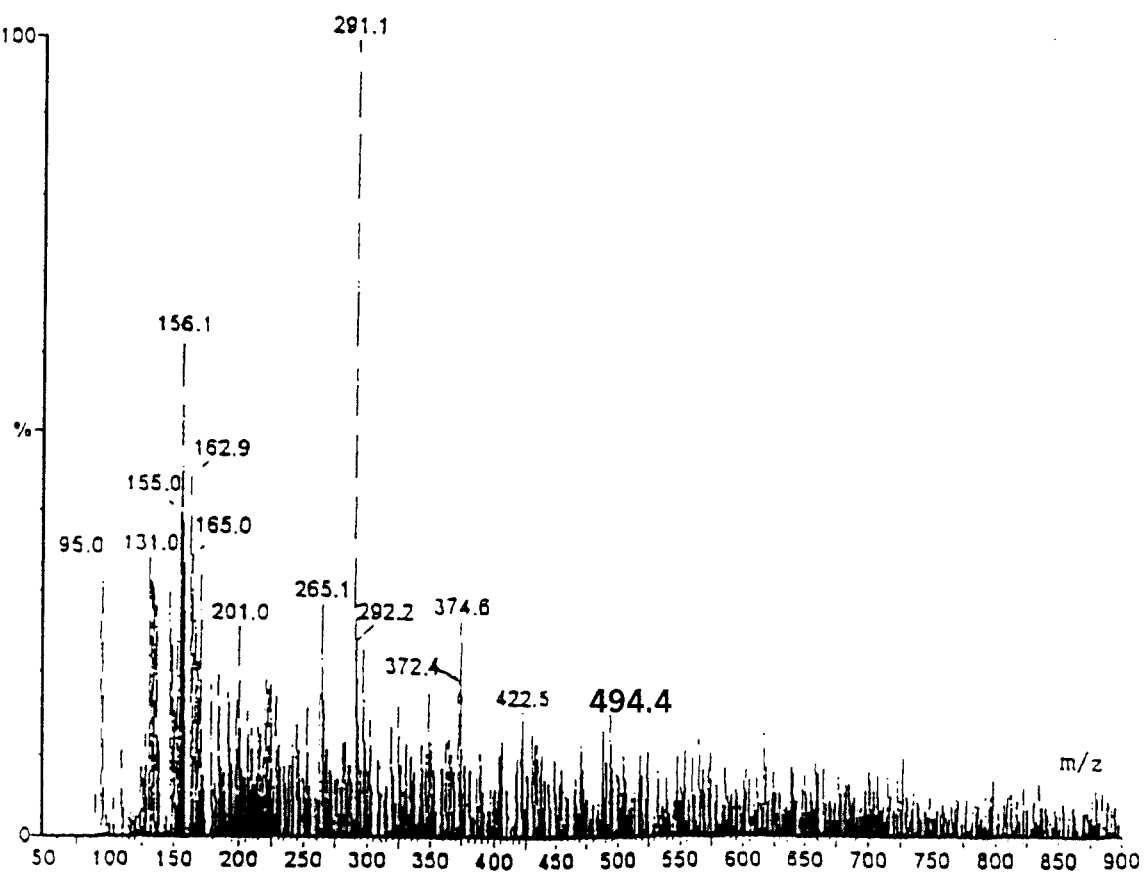
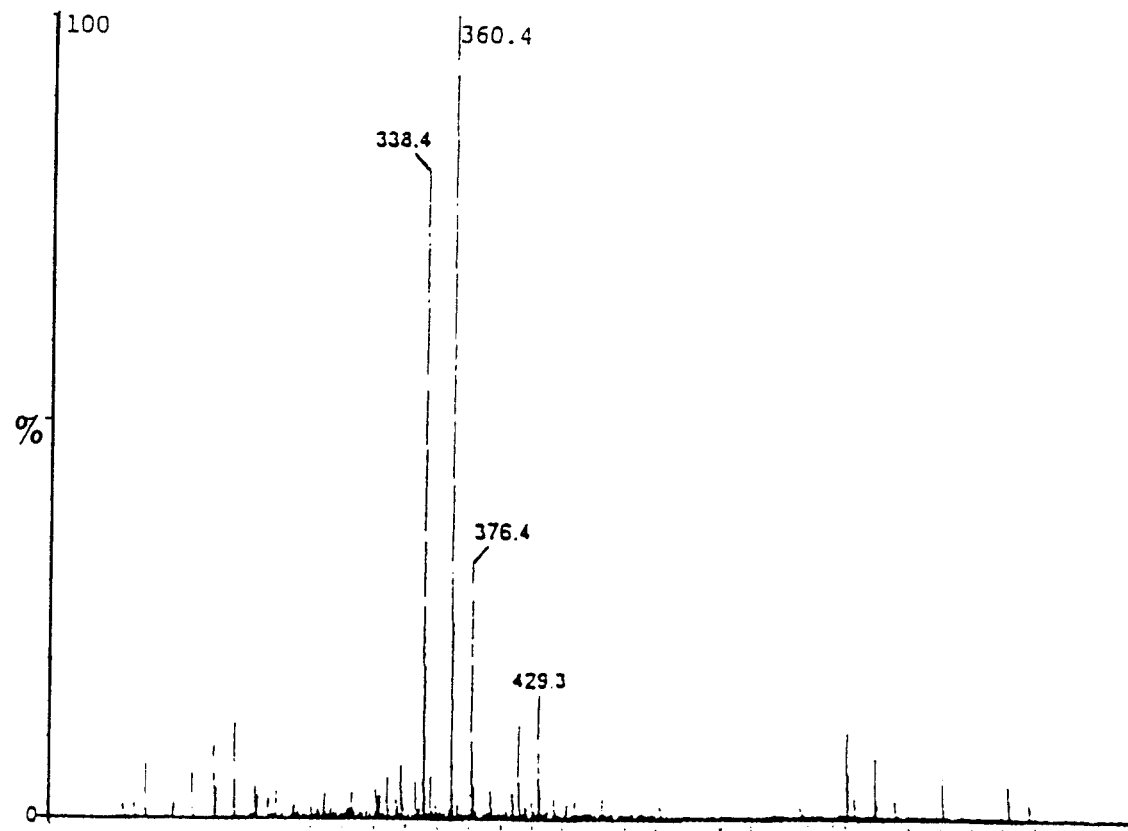


图 18

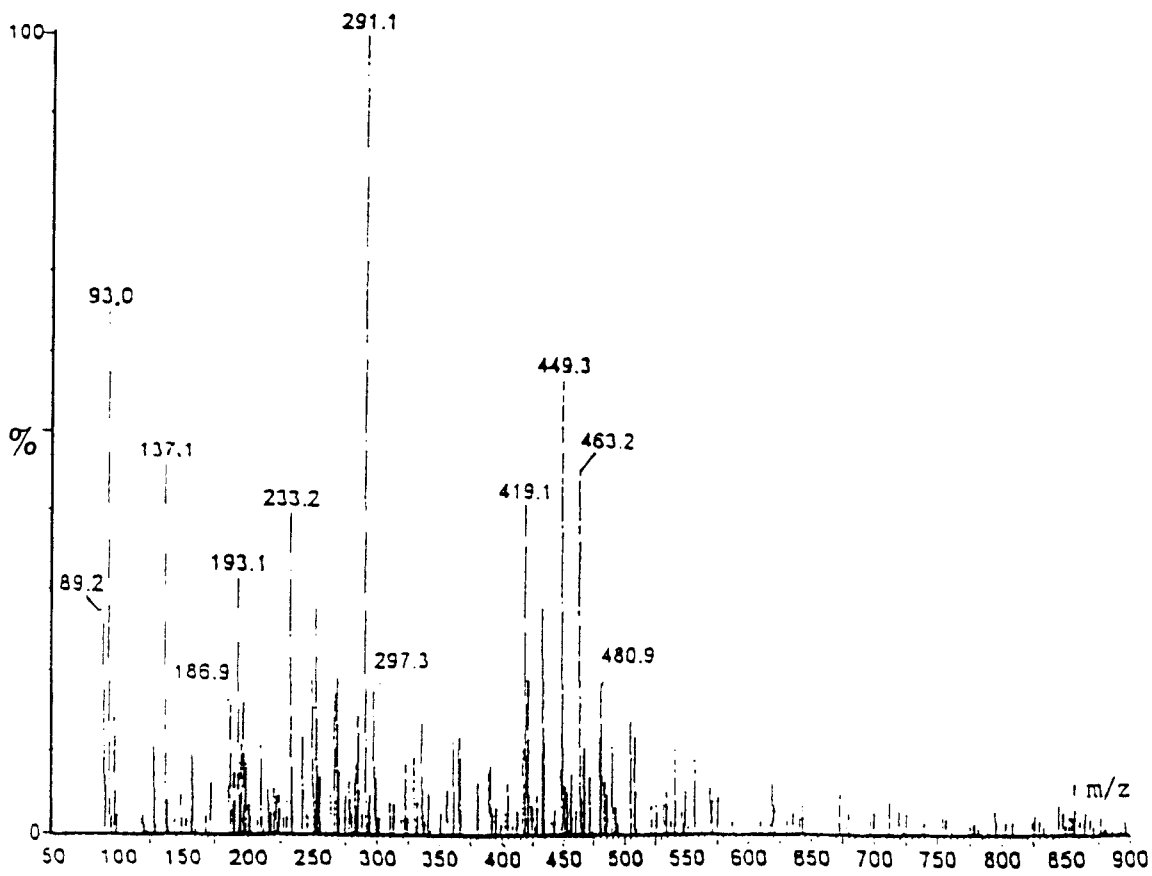
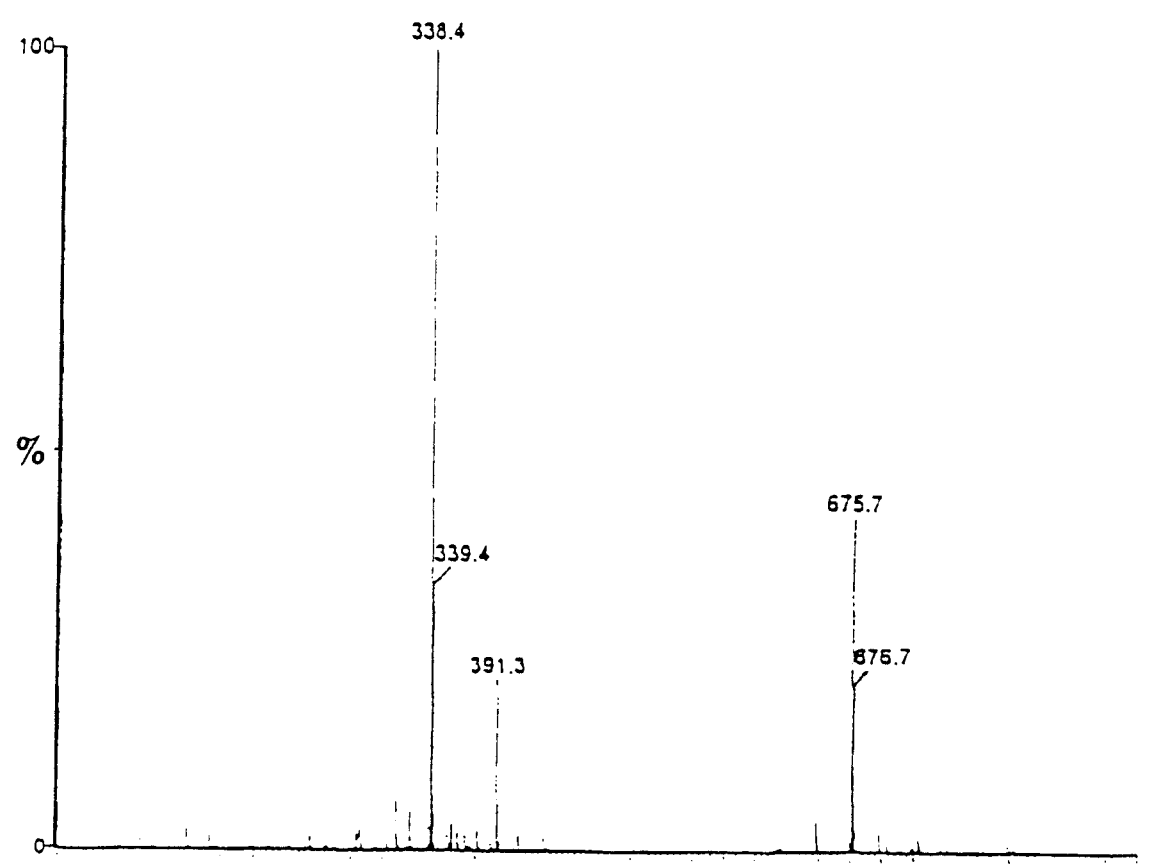


图 19

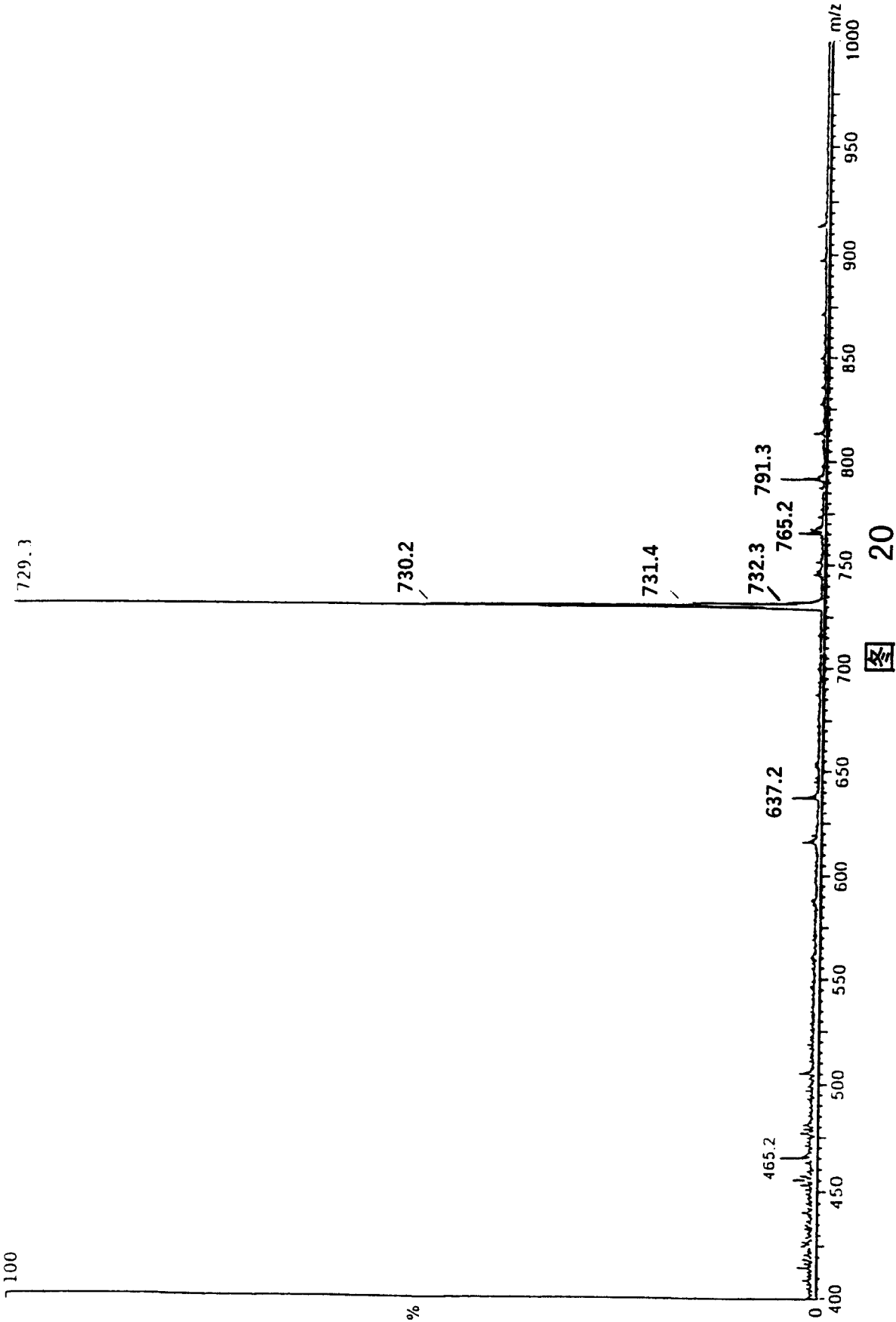


图 20

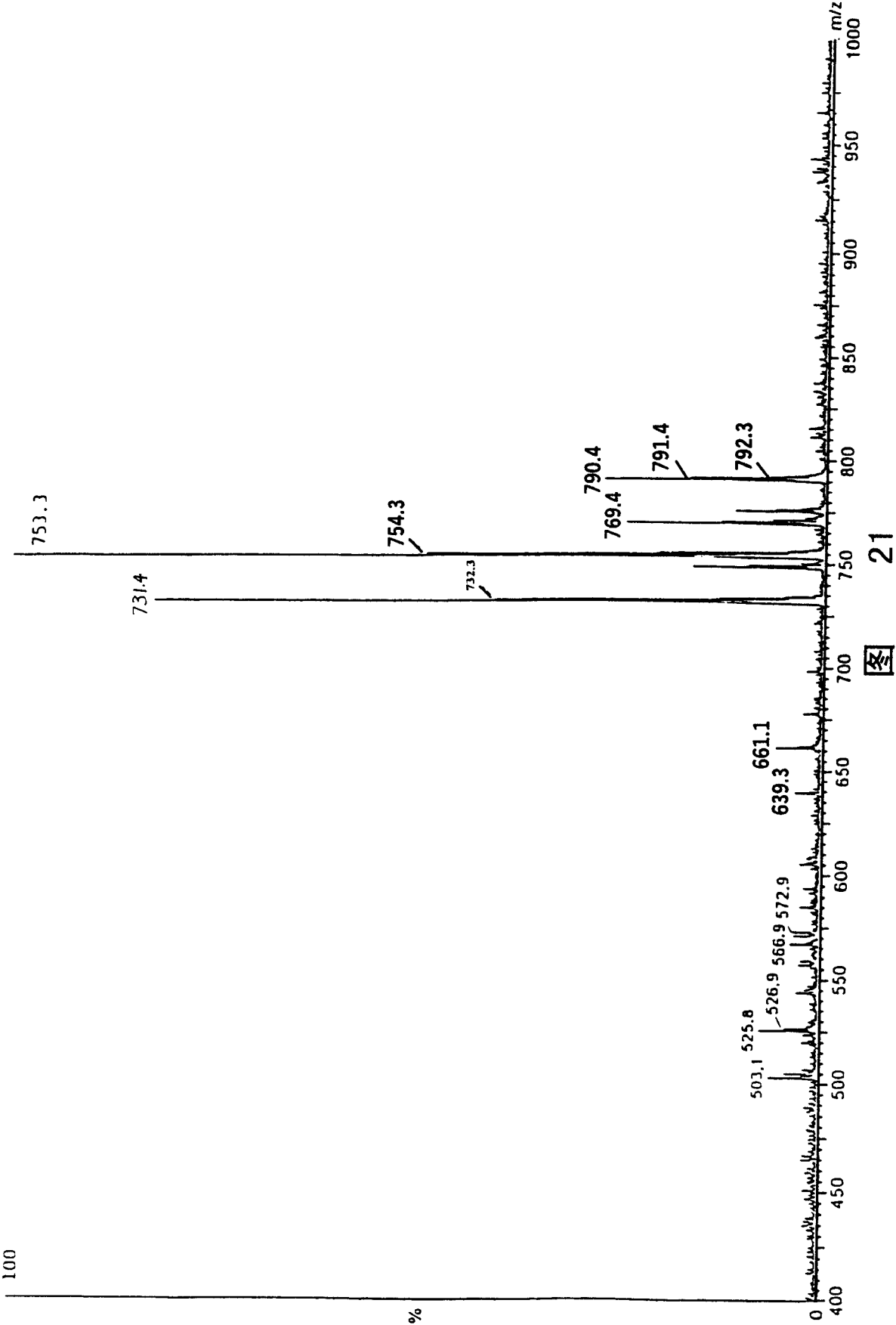
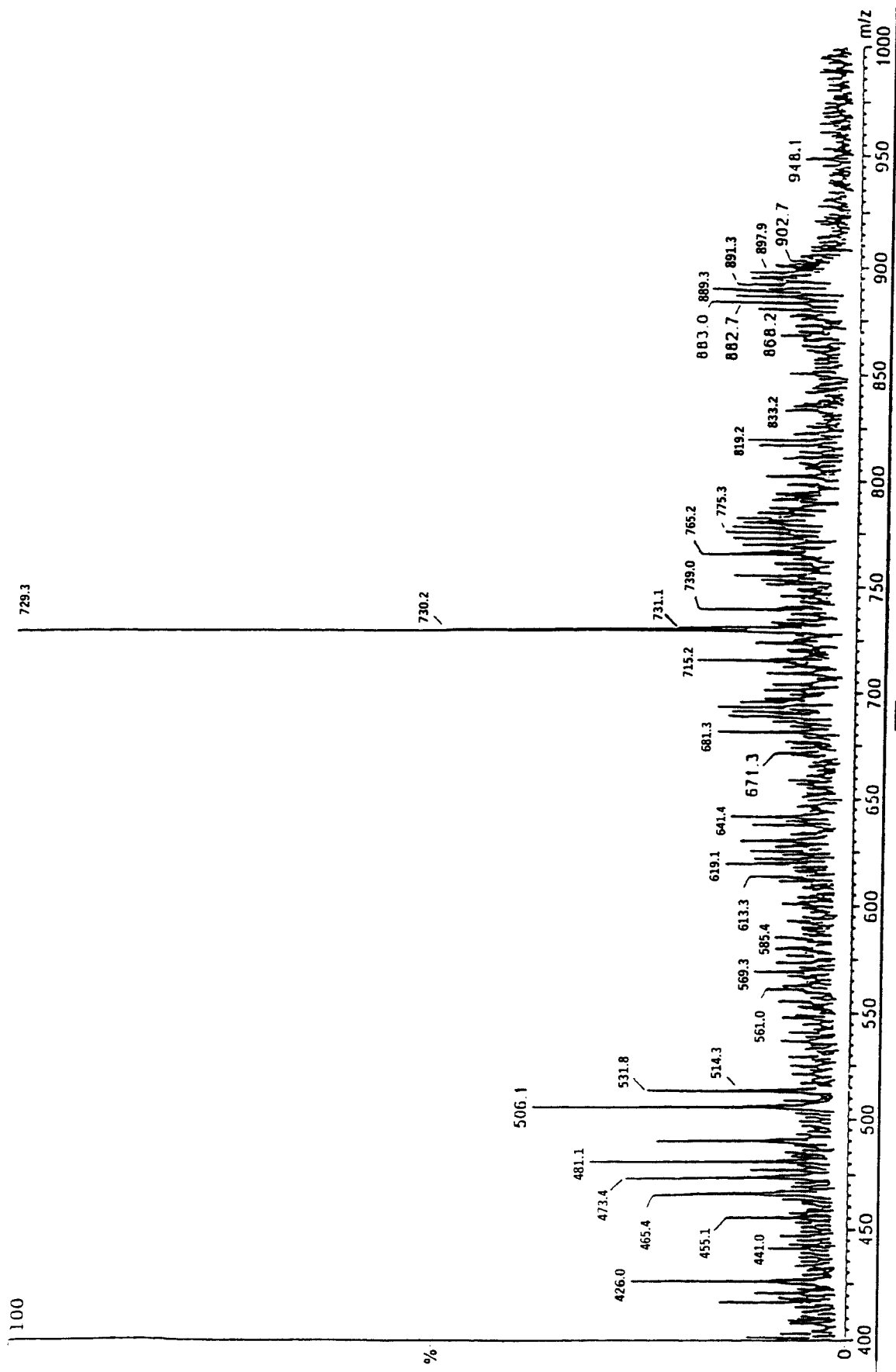


图 21



22

22

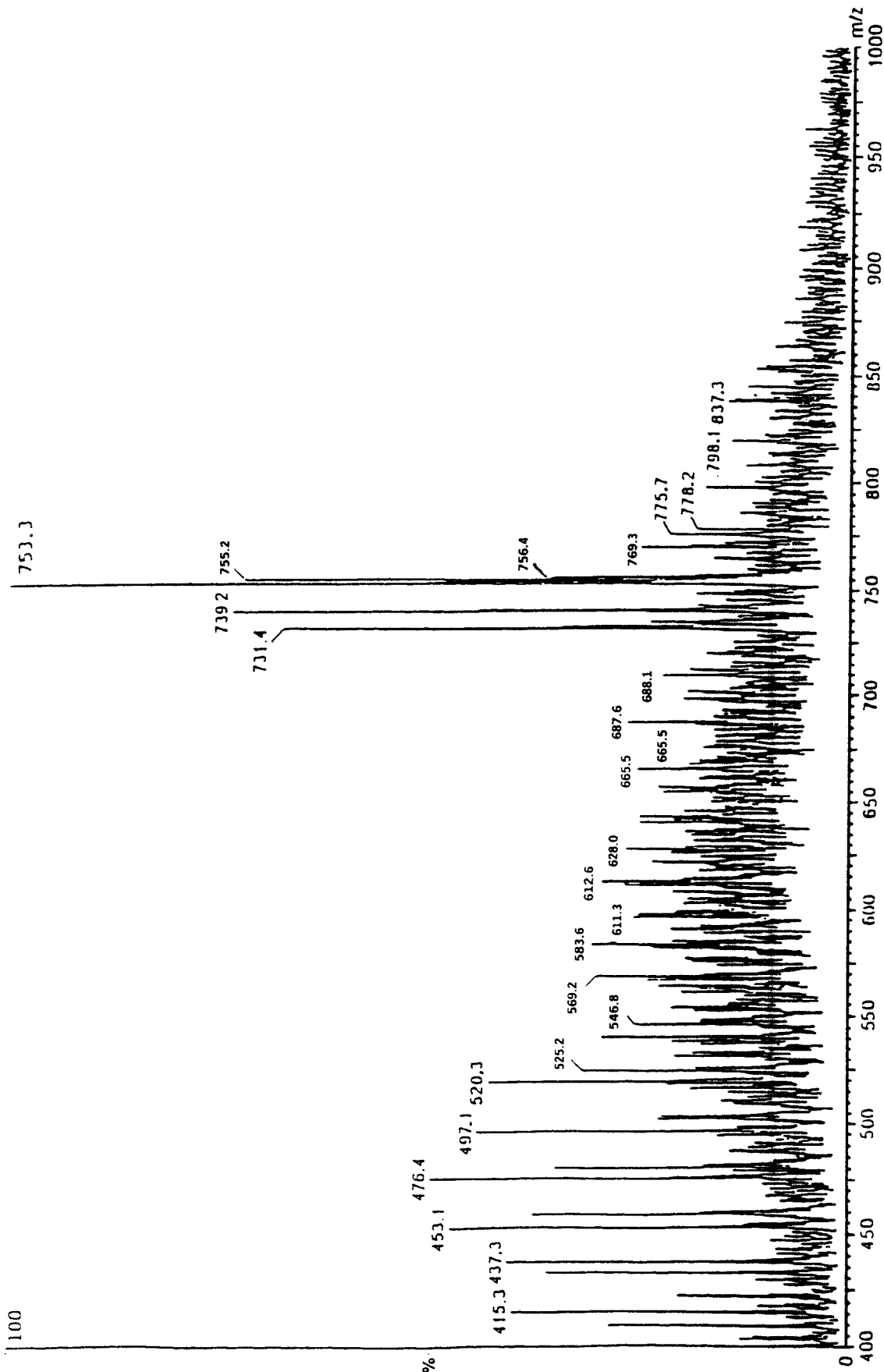
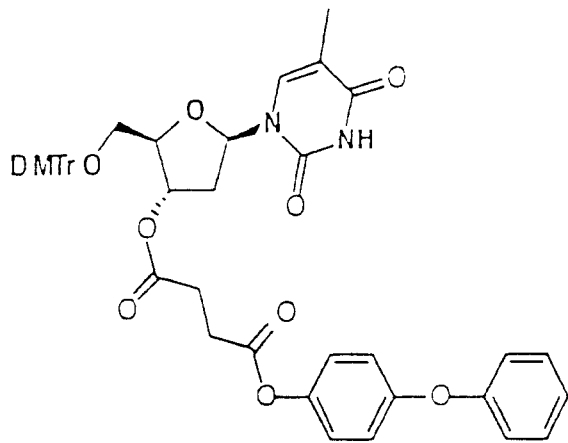


图 23

FT9



FT17

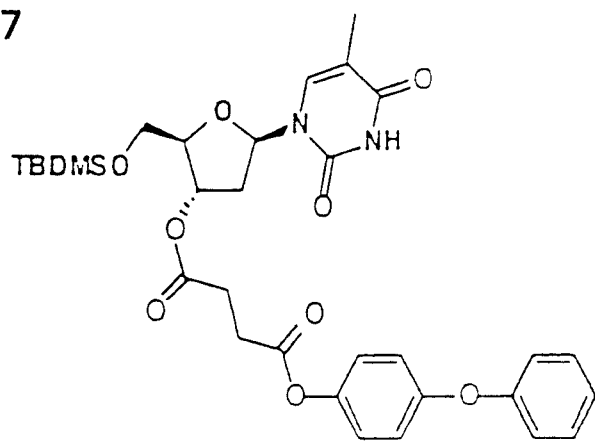
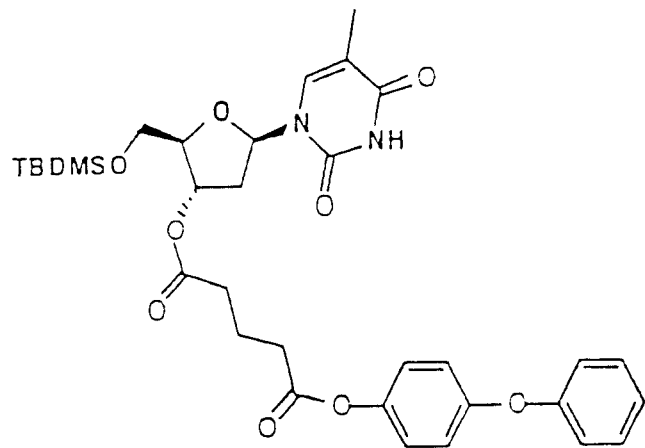


图 24

FT 18/1



FT23

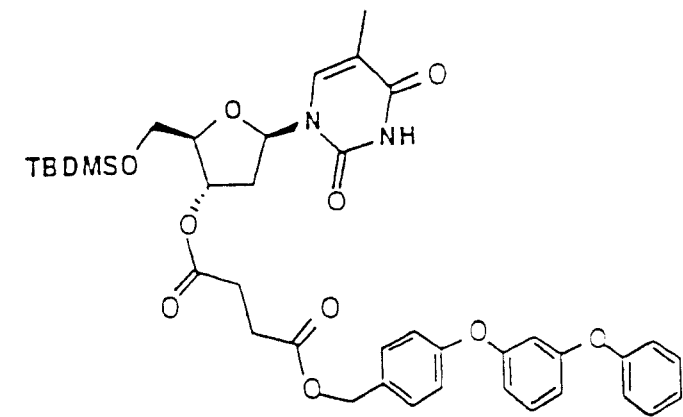


图 25