

[12] 发明专利申请公开说明书

[21]申请号 95116257.8

[51]Int.Cl⁶

G01N 33 / 50

[43]公开日 1996 年 7 月 31 日

[22]申请日 95.9.6

[30]优先权

[32]94.9.7 [33]JP[31]214055 / 94

[71]申请人 株式会社日立制作所

地址 日本东京都

[72]发明人 神原秀记 冈野和宜

[74]专利代理机构 中国国际贸易促进委员会专利商
标事务所

代理人 李 瑛

C12Q 1 / 68

权利要求书 6 页 说明书 22 页 附图页数 7 页

[54]发明名称 DNA分离、分级分离和分析方法及其系统

[57]摘要

将长 DNA 消化成小的片段，在没有亚克隆情况下使用 16 种引物的小文库进行测序。可联合使用凝胶电泳分离和与 DNA 聚合酶反应结合在一起的杂交技术分离各片段，其并没有使用亚克隆方法，因此对方法的全自动化是有宜的。本发明公开的 DNA 分离、分级分离和分析方法使用包括固定所说 DNA 探针之固相的容器的 DNA 分离和分级分离系统，以及包括为所说的容器内装添所需物质的工具、温度控制工具、样品转移工具及分级分离已分离之靶 DNA 片段的工具的机器人系统。

权 利 要 求 书

1. *DNA* 分离和分级分离方法,其包括:

i) 制备检测板的操作步骤,所说的检测板中 *DNA* 探针被固定在用于每种选择性序列的固相表面上——所说的 *DNA* 探针包括有寡聚体,该寡聚体含有所有探针共有的序列和 3' 末端上用于选择片段的 1 至 3 个碱基的选择性序列,或者将探针滞留在分有间隔的容器中的操作步骤;

ii) 使多个 *DNA* 片段与所说的 *DNA* 探针杂交的操作步骤;

iii) 通过 *DNA* 聚合酶反应只延伸 *DNA* 探针来稳定 *DNA* 探针与所说的 *DNA* 片段间之杂交体的操作步骤,所说的 *DNA* 探针中在已知序列后的 *DNA* 探针 3' 末端侧的选择性序列互补于所说的 *DNA* 片段相对应的序列部分;以及

iv) 通过在操作步骤 iii) 后的加热处理,根据杂交体稳定性的差异来分离和分级分离特异性 *DNA* 片段的操作步骤。

2. *DNA* 分离和分级分离方法,其包括:

i) 使具有已知序列的寡聚体与多个 *DNA* 片段在 3' 末端上结合的操作步骤;

ii) 制备固相如检测板或小珠的操作步骤,其中将包含寡聚体的 *DNA* 探针固定在分别用于每种选择序列的固相表面上——所说的寡聚体含有与已知序列互补的序列和包括 3' 末端侧之 1 至 3 个碱基的碱基选择性序列,或者将其保持在带有分隔的容器内的

操作步骤；

iii)使用固定步骤 ii)中得到之 DNA 探针的固相,使步骤 i)中得到的 DNA 片段与所说固相上寡聚体杂交的操作步骤；

iv)通过只延伸 DNA 探针来稳定 DNA 探针与步骤 i)中得到的 DNA 片段间之杂交体的操作步骤,其中所说的 DNA 探针中在互补于已知序列的序列之后是 DNA 探针 3' 末端侧上的选择性序列；以及

v)根据在步骤 iv)后的加热处理期间杂交体稳定性的差异来分离和分级分离特异性 DNA 片段的操作步骤。

3. DNA 分析方法,其包括在完成权利要求 2 的所有操作步骤后,经扩增,或者没有扩增已分离和分级分离的特异性 DNA 片段的拷贝数目,分析已经分离和分级分离之 DNA 片段的 DNA 序列的操作步骤。

4. 根据权利要求 1 的 DNA 分析方法,其包括在步骤 ii)之前用限制性酶消化靶 DNA 序列,以制备所说的多个 DNA 片段的步骤 i')。

5. DNA 分析方法,其包括在完成权利要求 4 的所有步骤之后,经扩增,或没有扩增已分离和分级分离之特异性 DNA 片段的拷贝数目,分析已分离和分级分离之 DNA 片段的 DNA 序列的操作步骤。

6. 根据权利要求 2 的 DNA 分离和分级分离方法,其进一步包括:

v')经加热步骤 1)中制得的 DNA 片段和固相层上所说 DNA 探针间的杂交体,解离没有经受互补链延伸的杂交体的操作步骤,

在完成步骤 iv) 后步骤 v) 时发生所说步骤;

iv) 降低杂交体和步骤 v) 中得到的 DNA 片段的温度, 以及固相层上没有杂交的 DNA 探针的混合物的温度, 重复步骤 iii)、iv) 和 v') 的步骤; 以及

vii) 在完成步骤 iv) 后用 PCR 方法扩增特定 DNA 片段, 以及将处理中使用的具有与已知序列互补之序列的寡聚体加到混合物中以分离和分级分离它们的操作步骤。

7. 根据权利要求 1 的 DNA 分离和分级分离方法, 其包括在步骤 ii) 之前或步骤 iv) 之后, 经凝胶电泳分离和分级分离 DNA 片段的操作步骤。

8. 根据权利要求 2 的 DNA 分离和分级分离方法, 其包括在操作步骤 iii) 之前或步骤 v) 之后经凝胶电泳分离和分级分离 DNA 片段的操作步骤。

9. 根据权利要求 2 的 DNA 分离和分级分离方法, 其中步骤 i) 中所说的 DNA 探针是荧光团标记的。

10. 根据权利要求 3 的 DNA 分析方法, 其中分析所说的 DNA 序列的步骤是一种凝胶电泳方法。

11. 根据权利要求 5 的 DNA 分离方法, 其中分析所说 DNA 序列的步骤是一种凝胶电泳方法。

12. DNA 分离和分级分离系统, 该系统包括

i) DNA 片段分级分离亚系统, 其包括含有检测板的容器, 检测板中包含已知序列和其中 3' 末端侧之 1 至 3 个碱基是选择性序列的寡聚体被分成间隔并根据每种选择性序列固定在固相表面上, 或者包括在这样一种放在分隔小室内之小珠的固相的表面上

保留有寡聚体的容器,以及

ii) 装配有将含有靶 DNA 片段的溶液倒入亚系统 i) 中并进一步将 DNA 聚合酶放入其中的装置、根据特定温度曲线发出控制亚系统 i) 中温度之指令的装置,以及只分级分离结合到所说固相上的余留 DNA 片段的装置和样品转移装置的机器人系统。

13. 根据权利要求 12 的 DNA 分离和分级分离系统,其进一步包括:

iii) 分离和分级分离 DNA 片段之长度的凝胶电泳仪。

14. DNA 分析系统,其包括权利要求 12 的 DNA 分离和分级分离系统,并进一步包括:

iv) DNA 测序仪。

15. DNA 分析系统,其包括权利要求 13 的 DNA 分离和分级分离系统并进一步包括:

iv) DNA 测序仪。

16. 根据权利要求 14 的 DNA 分析系统,其中所说的 DNA 测序仪是一种凝胶电泳仪。

17. 根据权利要求 15 的 DNA 分析系统,其中所说的 DNA 测序仪是一种凝胶电泳仪。

18. DNA 分析方法,其包括:

i) 将已知序列在其 3' 末端引入靶 DNA 片段的操作步骤;

ii) 制备引物库的操作步骤,所说的引物库包含针对每种限制性酶的至少 16 种引物且少于 64 种引物,所说的引物包括与在靶 DNA 片段上引入的已知序列互补的序列和限制性酶之识别序列的一部分,以及 3' 末端的 1 至 3 个碱基的选择性序列;

iii)使探针与片段杂交的完成 DNA 测序的操作步骤。

19. 测序引物库,其至少包含 4 种具有序列 $5' \cdots X_1 \cdots X_n Y_1 3'$ 的引物,其中 $X_1 \sim X_n (n=2 \sim 6)$ 代表限制性酶之识别序列的一部分,且 Y_1 是 A、C、G 或 T 的选择性碱基。

20. 测序引物库,其至少包含 16 种具有序列 $5' \cdots X_1 \cdots X_n Y_1 Y_2 3'$ 的引物,其中 $X_1 \sim X_n (n=2 \sim 6)$ 代表限制性酶之识别序列的一部分,且 $Y_1 Y_2 (Y_1, Y_2 = A, C, G \text{ 或 } T)$ 的选择性序列。

21. 测序引物库,其至少包含 64 种具有序列 $5' \cdots X_1 \cdots X_n Y_1 Y_2 Y_3 3'$ 的引物,其中 $X_1 \sim X_n (n=2 \sim 6)$ 代表限制性酶之识别序列的一部分,且 $Y_1 Y_2 Y_3 (Y_1 Y_2 Y_3 = C, C, G \text{ 或 } T)$ 是选择性序列。

22. 测序引物库,其至少包含 16 种具有序列 $5' \cdots X_1 \cdots X_n Z_1 \cdots X_m Y_1 Y_2 3'$ 的引物,其中 $X_1 \sim X_n (n=2 \sim 6)$ 代表限制性酶之识别序列的一部分, $Z_1 \sim Z_m (m=1 \sim 4)$ 代表核苷酸类似物,且 $Y_1 Y_2 (Y_1, Y_2 = A, C, G \text{ 或 } T)$ 的选择性序列。

23. DNA 分析方法,其包括:

i)使用含有多个 DNA 片段的 DNA 样品,借助连接或末端转移酶加入碱基,将具有已知序列的寡聚体引入这些 DNA 片段的末端,以建立用于 DNA 探针杂交的区域的操作步骤;

ii)使多个 DNA 片段与 DNA 探针杂交的步骤,其中将对 DNA 片段进引选择的包括与所说已知序列互补序列及 3' 末端侧的 1 至 3 个碱基是给定序列的 DNA 探针分成间隔并固定在检测板上,或者其中将探针保留在分隔的容器中的操作步骤;

iii)通过 DNA 聚合酶经互补链延伸,只延伸其中在已知序列

后的 3' 末端侧上用于选择的序列与所说 DNA 片段所对应的序列准确互补的 DNA 探针, 促进 DNA 探针和所说 DNA 序列间杂交的操作步骤, 其中 DNA 探针以外部分在步骤 ii) 发生了杂交; 以及

iv) 鉴定在步骤 iii) 中经受互补链延伸的探针和使用基本上具有相同序列的引物基于 Sanger 氏方法进行 DNA 序列分析, 以及分级分离和分析与经受互补链延伸的 DNA 探针杂交的 DNA 片段的操作步骤。

说 明 书

DNA 分离、分级分离和分析方法及其系统

本发明涉及样品制备方法, *DNA* 分析方法及其用于 *DNA* 序列测定或 *DNA* 分析的分离、分级分离和分析方法的系统。

为了按照常规方法测定长的 *DNA* 序列, 将靶 *DNA* 克隆在载体如质粒中, 从而经培养而增加有广泛用途的 *DNA* 拷贝数目。为了测定所说 *DNA* (在本文中为 2—6 升碱基), (1) 用限制性酶将 *DNA* 消化成较小的片段, (2) 再克隆(亚克隆)后将其感染到大肠杆菌中, 以及(3)将所说的大肠杆菌培养成菌落。然后经菌落筛选来分级分离 *DNA* 片段。培养增殖所选择的菌落中的大肠杆菌并提取 *DNA* 以得到包埋于质粒中之靶 *DNA* 的许多拷贝。也就是说, 用限制性酶消化 *DNA* 以通过亚克隆和继后的菌落培养来纯化、分级分离和扩增 *DNA*。如上所述, 在琼脂上培养大肠杆菌以形成菌落。各菌落包含有携带质粒的大肠杆菌, 质粒中则含有 *DNA* 片段的单一拷贝。从各菌落中挑取大肠杆菌并培养在试管中以增加拷贝数目。然后提取 *DNA* 用作进行序列测定的样品。为了分析各种不同的 *DNA* 片段, 分析得自各不同种之菌落的 *DNA* 片段。因为有反复分析含同一 *DNA* 片段之菌落的情况, 所以分析过程冗长而且常常使菌落的数目平均为不同种 *DNA* 片段数目的 4 倍左右。

使用这样的亚克隆方法进行 *DNA* 制备可使 *DNA* 片段在一个过程中得以分离和扩增。尽管有这一优点,但该方法也有花费过多时间和劳动,并且不适于自动化操作的缺点。此外,其另一个缺点是必须重复分析同一 *DNA* 片段,因为预先不可能知道在各菌落的大肠杆菌中含有何种 *DNA* 片段。

本发明的目的是解决现有技术所说的的问题,并提供没有使用克隆和培养的 *DNA* 分级分离、增殖和分析方法,从而取代耗费时间和不适于自动化的亚克隆及培养方法。

为了实现所说的目的,本发明使 *DNA* 片段与配有 3' 末端侧有选择 *DNA* 片段之序列的 1 至 3 个碱基的 *DNA* 探针间发生杂交;从而,可根据是否发生探针 *DNA* 链延伸反应来确定可用作测定各 *DNA* 序列之引物的 *DNA* 探针。本发明包括使用所说的方法对 *DNA* 片段进行分级分离的方法。本文所用的术语 *DNA* 与常规定义几乎没有区别。两个或多个 *DNA* 探针的大部分序列都是一样的,只是 3' 末端侧的 1 至 3 个碱基不同。目的是观察是否存在互补链延伸,是否有杂交。必要时用 *PCR* 方法(聚合酶链反应)扩增 *DNA* 片段,并与所说的新建立的方法平行进行,经凝胶电泳按照 *DNA* 片段的长度进行分离和分级分离,从而完成 *DNA* 分析。

如上文所讨论的,在克隆和亚克隆方法中,分级分离各个不同种的 *DNA* 片段以增加拷贝的数目。如果可用适合于自动化的便利方法代替分级分离和扩增两种方法,则对于大批量 *DNA* 序列测定将是很有帮助的。已经广泛使用 *PCR* 方法作为增加 *DNA* 片段拷贝数的手段。所以,如果在混合的条件下对各种不同的 *DNA* 片段进行分级分离或 *DNA* 序列测定,则将会解决这一问题。已经公开

了一些分级分离 *DNA* 片段的手段,包括凝胶电泳分离法或在固相表面上固定寡聚体以利用有无杂交的方法。然而,凝胶电泳法不能分离有差不多相同长度的 *DNA* 片段或互补链。再者,当使用杂交法时,要根据某些相关 *DNA* 片段组是否具有特异性序列来进行鉴定;因此,并不是所有序列都能分级分离未知 *DNA* 组中的各个片段。为了解决这一问题,提出了一个其中使用在各片段末端上的几个碱基作为标志来分级分离和分析各片段的系统。也就是说,当用限制性酶消化各片段的末端时,在所有片段的切割位点处都有酶的识别序列,但是预期根据各片段不同与之相邻的 1 至 3 个碱基的序列有所不同;这一点即在所提出的方法中用来进行选择。或者当用超声波等在随机位置上消化该 *DNA* 片段时,注意到在任意选择的末端上有 3 至 4 个碱基之特异序列片段,以代替限制性酶识别序列(当选择 *GACT* 作为 4 个碱基时,其将以 256 个碱基之一的比例出现);当利用与之相邻的 1 至 3 个碱基序列的差异鉴定和分级分离具有以这一恒定频率出现之末端序列的片段时,则按照另一种可替代的方法进行。为了认识这一点,本发明使用凝胶电泳法对片段进行粗略分离和分级分离。为了鉴定各部分中的 *DNA* 片段,利用了末端几个碱基的差异,在许多情况下是利用 4 至 7 个碱基的差异。在这 4 至 7 个碱基中,2 至 4 个是限制性酶的识别序列或出现在规定间隔处的序列,而 3' 末端其余的 1 至 3 个碱基则用于选择。试图制备具有与该 4 至 7 个碱基部分互补之序列的 *DNA* 探针,并根据是否精确地发生杂交对其进行分级分离,没有获得成功的结果。这是因为下述原因:首先,4 至 7 个碱基杂交不能稳定地形成杂交体;其次,杂交碱基不能达到不包括一个碱基错配的准确分级分

离。因此,使用迄今所用的简单 *DNA* 杂交方法不能鉴定或分离未知的 *DNA* 片段。根据这一提议,为了确保稳定的杂交,在 *DNA* 片段的 3' 末端加上已知序列的 *DNA* 寡聚体以延长杂交区域。另外,使其 3' 末端有选择之序列的用于选择的探针与片段的 3' 末端杂交,并且不仅根据是否发生杂交,而且还根据杂交的 *DNA* 探针是否通过 *DNA* 互补链延长反应引起 *DNA* 链的延伸来分级分离 *DNA* 片段。对此考虑到下述事实:是否发生互补链延伸取决于 3' 末端的 1 至 3 个碱基是否实现了完全匹配。使用上文讨论的 *DNA* 探针的 *DNA* 分析方法包括:1)消化靶 *DNA*,2)凝胶电泳后分收分离经消化的 *DNA* 片段,3)在 3' 末端加上寡聚体以修饰 *DNA* 片段,4)使 *DNA* 探针与该片段杂交,5)通过当探针的 3' 末端与片段完全匹配时发生的聚合酶反应来延伸 *DNA* 探针,6)经检测延伸的引物,确定测序引物,和/或用经过延伸的 *DNA* 探针分级分离 *DNA* 片段,以及 7)使用已确定的引物测定 *DNA* 片段混合物中之 *DNA* 片段的序列,或者用经过延伸的 *DNA* 探针测定分级分离之片段的 *DNA* 序列。方法的关键是步骤(5)和(6)。例如,当选择的序列是 2 个碱基单体时,有 16 种 *DNA* 探针。将 *DNA* 样品分成 16 个相等的部分,并将彼此不同的 *DNA* 探针插入各部分以进行互补链延伸。除了 3' 末端的两个碱基外所有序列都是共有的;可与任何片段发生杂交。但只有在末端的两个碱基准确杂交时才发生有效的互补链延伸。用荧光团标记的 *DNA* 探针研究 *DNA* 探针引导延伸的能力。经凝胶电泳分析反应产生之产物的长度。因此,根据迁移速度检查长度可鉴定有 *DNA* 片段测序引物功能的 *DNA* 探针,以及 *DNA* 片段的近似长度。当用生物素标记 *DNA* 探针时,可用固定有抗生蛋

白链菌素的磁珠挑出 DNA 片段。如果在这种情况下将温度定在摄氏 80—85 度，与没有互补链延伸的 DNA 探针杂交的 DNA 片段便不稳定，而离开 DNA 探针并被除掉。与经过延伸的 DNA 探针杂交的 DNA 片段并不从经延伸的 DNA 探针上离去并因而被分级分离。为了使这一方法更为有效，在固相表面上提供 16 个凹槽并使 16 种 DNA 探针分别固定到这些凹槽内，以形成用于本目的的检测板。当然，也可以将 DNA 探针固定到小珠上以代替检测板，并根据所固定的 DNA 的种的不同将它们包封在不同的凹槽内。也可使用不同的容器。将 DNA 片段混合物放在包含检测板的容器内，并使固定的寡聚体间发生杂交。这一过程后使用 DNA 聚合酶进行 DNA 探针的互补链延伸。当含有 DNA 探针 3' 末端之 2 个碱基的选择性序列互补于已杂交之 DNA 的序列时，即发生互补链延伸以确保杂交有更大的稳定性。当使温度升至 80 至 85℃ 时，连接到没有经受互补链延伸之 DNA 探针上的 DNA 片段即从 DNA 探针上离开并被溶剂洗掉，然后除去之。如此导致只有用已延伸的 DNA 探针固定 DNA 片段。经鉴定这些 DNA 探针的不同种，即有可能选择出用来从 DNA 片段混合物中确定特异性片段之序列的引物。也有可能分离分离根据每个种陷入凹槽中的 DNA 片段，以将其用作序列测定模板。以上主要讨论了 DNA 分析。当然，只分级分离与由于 DNA 链延伸而完全匹配的 DNA 探针杂交的 DNA，也是一种有效的方法。

换句话说，通过使用 DNA 探针杂交和其确保的互补链延伸，本发明有可能根据末端序列获得经鉴定 DNA 片段进行分析的信息，以及分级分离 DNA 片段，从而提供对具有未知序列的 DNA 进

行序列测定的方法。也有可能按照需要在扩增了它们的拷贝后，测定按所说的步骤分离和分级分离之 *DNA* 片段的序列。可以用任何处理方法制备作为按上述步骤进行分离和分级分离的起动物质的 *DNA* 片段；例如，用超声波消化或用限制性酶进行酶促 *DNA* 消化。用于这一目的的限制性酶包括 *Hha* I、*Hind* III、*Sau*3A1、*Not*I 和 *Eco*RI。从中适当地选择一种或几种符合相关 *DNA* 序列的酶。另外，可使用核酸外切酶或核酸内切酶消化、用超声波消化或经化学反应消化等方法产生 *DNA* 片段。

再者，该 *DNA* 分离和分级分离方法可以在使用所说的步骤 1) 至 7) 的方法上再加入下述扩增 *DNA* 片段之拷贝数目的方法。也就是说，该 *DNA* 分离和分级分离法至少有下列操作步骤：I) 使具有已知序列的寡聚体与多个 *DNA* 片段结合的操作步骤；II) 制备检测板的操作步骤，其中将含有寡聚体的 *DNA* 探针固定在固体表面上制得每种序列的选择性 *DNA* 探针(其中所说的寡聚体包括与已知序列部分互补的序列，并且其中 3' 末端侧的 1 至 3 个碱基是序列选择性的)；或者将探针装在分成间隔的容器中的操作步骤；以及 III) 使用操作步骤 II) 中得到的分成间隔的并固定的 *DNA* 探针，和具有与所说的已知序列之共有序列互补两序列的，并与所有片段的末端杂交的引物，经用酶进行互补链延伸以扩增所说 *DNA* 片段之拷贝数目，以及分离各个种的 *DNA* 片段以在固体表面上进行分离和分级分离的操作步骤。

具有与所说的已知序列互补之序列的 *DNA* 探针已知序列部分的碱基长度应符合通过杂交形成双链的要求。其较好为 10 至 30mer，更好应为 16 至 20mer。应注意的是，当借助末端转移酶加上

poly A(多聚腺苷酸)时,附加的链长度可达到约100mer。在这种情况下,选择DNA片段之DNA探针的长度应为10至30mer。如果长度小于10mer则可导致杂交不稳定。如果大于30mer,则倾向于与没有精确互补的部分杂交,这是不可取的。对于引物,使用约18mer(正常使用的大小)是适宜的。注意到标记加到DNA片段上的寡聚体对于继后的分离更为便利。使用荧光团、化学发光团诱导物质或放射性物质作为标记物。

可结合使用依据以凝胶电泳法测得的DNA片段的长度进行分离和分级分离的方法,使根据本发明的所说的DNA分析、分离和分级分离方法更为有效。

可使用在加有6%浓度聚丙烯酰胺(在本说明书中,凝胶浓度以总单体浓度的重量/体积(g/ml)百分比表示)的凝胶板上分离DNA带的方法,经凝胶电泳进行分离和分级分离,或者,使用将DNA从管凝胶末端洗脱到溶液中进行分级分离的方法(例如美国专利5,277,780中所公开的方法)。

下列文献中描述了制备固定根据本发明之寡聚体的固相的方法:为将寡聚体固定到固体表面上,可以用K. Beattie et al., *Clinical Chem.*, vol. 39(1993), pp. 719—722中公开的方法;或者按照M. Mistuhashi et al., *Nature*, vol. 357(1992), pp. 519—520中所述的方法完成固定,在使用装有*poly T*的塑料表面使具有与*poly A*和靶寡聚体序列互补之序列的寡聚体与*poly T*(多聚胸苷酸)杂交后,经互补链延伸在所说*poly T*的3'末端侧形成靶DNA探针。使用塑料、石英板、硅片等作为固定该DNA探针的固相基体。

如上文所讨论的, *DNA* 探针中用于选择 *DNA* 片段的选择序列应为 1 至 3 个碱基长。通过将此处理步骤重复两次或多次, 选择序列可以为 4 至 6 个碱基长, 但每次操作必须是 1 至 3 个碱基。本发明的基本原则之一是, 游离的或固定到固体上的 *DNA* 探针完成互补链延伸(所说的操作步骤 5 或 III) 以与靶 *DNA* 形成稳定的杂交体, 而只有当用于 *DNA* 选择的 *DNA* 探针之 3' 末端侧上的选择序列与 *DNA* 片段准确杂交时, 才能通过杂交体的形成从其他 *DNA* 片段中鉴定出所说的靶 *DNA*。在这种情况下, 互补链延伸能力在很大程度上取决于 3' 末端上 1 至 3 个碱基的序列是否与 *DNA* 片段序列精确互补。本发明的基本要求是, 在选择性 *DNA* 探针的 3' 末端侧上提供至可 1 个用于选择的随机碱基。另外, 当用于选择之 *DNA* 探针的 3' 末端侧上的选择序列是 4 个或更多个碱基长时, 尽管在该部分上有 1 个碱基错配, 互补链延伸还会进行, 从而导致不能从错配的 *DNA* 探针中鉴定出准确匹配的 *DNA* 探针, 这当然是不可取的。

在 70 至 75°C 的反应温度下借助 *DNA* 聚合酶完成互补链延伸, 其中将 2 单位作为 *DNA* 聚合酶的 *Tag* 和 *dNTP* (三磷酸脱氧核苷酸) 的混合物(作为 *DNA* 聚合酶反应混合物) 加到装有与靶 *DNA* 片段杂交之检测板的容器中。如果反应温度低于 70°C, 反应效率将会低下, 这是不可取的。象 *Tag* 这样的酶的活性是很重要的, 并且其量(例如根据在规定的时间内分解或合成特定量的 *DNA* 所需的酶量来限定) 由单位数表示。

一般是使用 0.1 至 1 *pmol* (10^{11} 至 10^{12} 个分子) 的靶 *DNA* 片段进行 *DNA* 测序。因此, 如果靶 *DNA* 片段的量不足够供测序使

用,则按照所说的操作步骤Ⅲ中所述用酶扩增 DNA 片段拷贝的数目。这一过程可在正常 PCR(聚合酶链反应)条件下完成。

使用荧光检测型自动 DNA 序列仪,即荧光检测型电泳装置分析 DNA 序列,并按照其标准程序确定碱基序列。

PCR 在增加 DNA 片段拷贝数目中是有效的,而凝胶电泳则在按照长度分离 DNA 片段中和对它们进行分类中是有效的。借助连接或在末端转移酶的作用下加入 dATP 将已知序列的寡聚体连接到经凝胶电泳分级分离的 DNA 片段的两个末端上。这样做是为了在片段上加入一个区域,以确保引物的稳定杂交。将含有该序列和与已知序列(如用限制性酶消化的部分)互补之序列,并且在 3' 末端侧上具有作为选择序列之两个序列选择碱基的 DNA 探针就每个末端序列的不同分成间隔,探针的 5' 末端侧固定到准备好的容器中。将含有 DNA 片段的溶液倒入容器中并使之与固相化的探针杂交。用 DNA 聚合酶延伸 DNA 探针。只有当探针的 3' 末端序列与 DNA 片段互补时才发生链延伸,而且可提高杂交体的稳定性。

当温度升至 80 至 85℃ 时,未延伸之引物和 DNA 片段的杂交体解离。另一方面,延伸之引物和 DNA 片段的杂交体则不解离,而且 DNA 片段保留在固相表面上。洗去与 DNA 探针分离的 DNA 片段。然后使温度增加到 95 至 98℃ 以从探针上进而从固相表面上除去保留的 DNA 片段。分级分离 DNA 片段。

如上文提到的,本方法的关键点是 DNA 聚合酶反应发生在 DNA 探针上。聚合酶反应对 DNA 探针之 3' 末端上的碱基序列是很敏感的。当 DNA 探针的末端两碱基序列与 DNA 片段的相应序列互补(末端两个碱基与 DNA 序列匹配)时,聚合酶反应即发生在

探针上。反之,则不能发生或很难发生,导致杂交体在 95 至 98℃ 的高温下解离。因而依据 3' 末端序列分离 DNA 片段并进而分级分离之。两碱基序列有 16 种组合。许多 DNA 片段不能一次用这种杂交和探针延伸方法分离和分级分离。但将凝胶电泳分离与该方法联合起来,则可分离和分级分离许多 DNA 片段。用 PCR 方法扩增以这种方法分级分离的 DNA 片段并用于 DNA 测序反应。

用凝胶电泳法正常完成 DNA 序列测定。日本专利申请 No. 4-180095(申请日 1992 年 7 月 7 日)的说明书中公开了使用 DNA 探针测定长 DNA 序列的方法(该日本专利相当于申请日为 1993 年 7 月 7 日的美国专利申请序号 08/086,892)。这种方法只根据 DNA 探针和 DNA 片段间是否存在杂交来选择引物,并且不能选择用于以足够的准确性进行 DNA 互补延伸的引物;有些情况是 3' 末端序列错配并且探针不能被使用。在某些场合中用已报导的方法,已知序列部分上的碱基长度不足以产生稳定的杂交体。再者,本方法并不涉及 DNA 的分离和分级分离,并且基本上不同于所说的现有技术的方法。

所说的根据本发明的 DNA 分离和分级分离具有在初始操作过程中使含有已知序列的寡聚体与多个 DNA 片段结合的步骤。该步骤可以因使用检测板的双股寡聚体而删去。

用于根据本发明的 DNA 分离和分级分离的系统有:i)DNA 片段分级分离亚系统,其包括含有检测板的容器,其中将包括已知序列并且 3' 末端侧的 1 至 3 个碱基为序列选择性的寡聚体分成间隔,并就各个序列选择种的不同固定在固相表面上,或者包括将寡聚体装在由间壁分隔的凹槽中的容器,其中寡聚体固定在这样一种

作为小珠的固相的表面上；以及 *ii*) 配备有将含有靶 *DNA* 片段的溶液倒入亚系统 *i*) 并进一步将 *DNA* 聚合酶倒入其中的装置、根据特定的温度曲线发出控制亚系统 *i*) 中之温度的指令的装置，只分级分离结合到所说固相上之余留 *DNA* 片段的装置，以及样品转移装置的机器人系统。可将机器人系统上配备的温度控制指令装置安装在亚系统 *i*) 上。此外，需要时，可在所说的亚系统 *i*) 和机器人 *ii*) 上增加一个用于根据本发明进行 *DNA* 分离和分级分离的凝胶电泳仪 *iii*)。

根据本发明的 *DNA* 分析系统包括加到所说的亚系统 *i*) 和机器人 *ii*) 上的 *DNA* 测序仪 *iv*)。必要时，可以加入另一个用于 *DNA* 片段分离和分级分离的凝胶电泳仪 *iii*)。如果可以达到目的，则该凝胶电泳仪可以是任何类型的。例如，常规可使用荧光检测型电泳仪。图 8 举例说明了根据本发明的 *DNA* 分析系统。从图 8 中去掉 *DNA* 序列分析仪即给出根据本发明的 *DNA* 分离和分级分离系统。

图 1 是代表根据本发明之 *DNA* 分离、分级分离和分析方法的流程图；

图 2 是图解说明用限制性酶消化样品 *DNA* 并将具有已知序列的寡聚体导入所得 *DNA* 片段之末端的方法；

图 3 是图解说明用另一种限制性酶消化样品 *DNA* 并将具有已知序列的寡聚体导入所得 *DNA* 片段之末端的方法；

图 4 是图解显示固定到固相表面上之探针的结构；

图 5 是图解显示探针检测板的结构；图 6 是图解说明随迁移时间改变的荧光强度。

图 7 是图解显示用于 DNA 片段测序之引物(或探针)的结构;

图 8 是图解显示用于根据本发明的 DNA 分离、分级分离和分析方法中之系统的配置;

图 9 是显示实例 4 之分析方法的流程图;

图 10 是对用 *HhaI* 酶消化所得到的 DNA 片段进行琼脂糖凝胶电泳后获得的低分离能力电泳图。

下面参照实例对本发明进行具体描述。

实施例 1

图 1 是代表 DNA 分析的流程图。已选择了入噬菌体 DNA 作为样品 11。使用 6 碱基识别酶,例如 *Hind* III 消化之。可使用任何类型的限制性酶,其包括 *Aat* I、*Acc* I、*Acy* III、*Alu* I、*Apa* I、*Asp* I、*Ban* I、*Bra* I、*Dra* III、*Hae* III、*Ecor* I、*Minf* I、*Mae* I、*Not* I、*Pst* I、*Hha* I、*Sau3A* I、*Sca* I、*Taq* I 等。这些酶都列在生物化学公司的生物化学品目录中,例如参见 *Boehringer Mannheim* 1990 的“*Biochemicals for Molecular Biology*”,或 *Sambrook, Fristch* 和 *Maniatis* 的“*Molecular Cloning* (pp. 5.3—5.9)”。该酶识别图 2 中所示的序列 10 并在图示的位置上消化 DNA。通过连接使经过消化的 DNA 片段 12 结合到具有已知序列的 DNA 寡聚体 13(图 2 中所示的)上。对于下述分级分离操作,使用标记了荧光团 14 的寡聚体是有效的。本发明人使用了用硫氰胺 101(*Sulforhodamine 101*, 为有 620nm 发射波长的德克萨斯红)标记的寡聚体。通过丙烯酰胺凝胶电泳分离和分级分离连接产物。各部分(管)中 DNA 片段的数目取决于凝胶电泳的分离能力。当欲分离有 2 至 6 千碱基长度的 DNA 片段时,以约占长度之 1% 的准确性分级分离是可能的。如图

1 中 16 所示,可在各部分中包括 1 至 6 种或较少种的 *DNA* 片段。如图 1 中 2 所指示的,对每个部分再次用另一种限制性酶消化这些 *DNA* 片段。本发明人使用了 4 碱基识别酶如 *Sau* 3*AI* 作为限制性酶。通过连接反应使经过消化的部分与具有已知序列的寡聚体结合。可用末端转移酶将 *poly A* 尾部引入 3' 末端,但是如果用 *Sau* 3*AI* 消化,在被消化产物的 3' 末端便不保留有识别序列(在 5' 末端上提供识别序列)。因为在 3' 末端没有已知序列,所以经互补链延伸对被消化部分的 3' 末端进行一次延伸,以在被消化部分的 3' 末端上产生序列互补链;然后连接一个 *polyA* 尾部。图 3 图解显示了被消化的部分和已知的寡聚体序列。“122”代表经消化的 *DNA* 序列片段。可根据该已知寡聚体 113 选择不同于所说的已知序列 13 的序列。将荧光团标记 14 连接到寡聚体上以利于检测。“15”和“115”代表其中已引入了用荧光团标记之已知序列 *DNA* 寡聚体的 *DNA* 序列。图 2 中的“N”是指腺嘌呤(A)、鸟嘌呤(G)、胞嘧啶(C)或胸腺嘌呤(T)之一。

为了分级分离所得到的 *DNA* 片段,制备如图 1 中所示的装有寡聚体的检测板 30。如图 4 所示,用检测板上的寡聚体通过接头 181 将其 5' 末端侧固定到固相表面 180 上,并包含通过连接反应连接到 *DNA* 片段样品上的寡聚体序列 32,以及其 3' 末端侧上限制性酶识别序列 33 的一部分。该寡聚体与所有 *DNA* 片段形成杂交体。为了鉴定不同种的 *DNA* 片段,使用在 3' 末端侧上例如有另外两个用于选择序列 34 的碱基(也可以是 1 至 3 个碱基)的寡聚体。可以从 16 种组合中选择两个序列选择碱基。如图 5 中所示,借助分离间壁 202 将检测板表面分成 16 个部分,并将单一种的寡聚体固定

在各个寡聚体固定凹槽上。用此检测板分离和分级分离 DNA 片段。杂交的稳定性主要取决于是否因 DNA 片段之限制性酶识别序列后的两碱基序列差异而发生互补链延伸；这一发现被用于分离和分级分离。在检测板上的寡聚体 3' 末端，除了两个碱基以外，DNA 片段完全互补。根据杂交的程度，很难确定末端的两个碱基是否互补。然而，使用 DNA 聚合酶进行的互补链延伸反应却在很大程度上取决于用作引物的寡聚体之 3' 末端上的两个碱基。经受互补链延伸的寡聚体与 DNA 片段间的结合力远大于未经延伸的寡聚体与 DNA 片段间的结合力。所以在寡聚体 31 固定在图 1 所示寡聚体固相化检测板 30 的表面上，并且有已知序列的 DNA 片段 37 和 38 发生了杂交之后，即完成 DNA 互补链延伸反应。只得到连接到链延伸之寡聚体 39 上的 DNA 片段即完成分离和分级分离。将 DNA 片段放在其底部有检测板的容器中（检测板可以放在容器的底上），从而形成与寡聚体的杂合体。图 1 中的参考编号 50、51 和 52 代表固定有不同寡聚体的检测板上寡聚体固相化区域。向其中加入 *Tag* DNA 聚合酶及互补链延伸测序反应混合物 *dNTP*（三磷酸脱氧核苷酸）以促进互补链延伸。用其中寡聚体 3' 末端侧上的两个碱基与 DNA 样品之序列精确互补的杂合体 37，互补链延伸继续进行，并且固定到表面上的寡聚体链延伸。而用其中寡聚体 3' 末端侧上的两个碱基没有精确互补的杂合体 38，则互补链延伸不能进行下去，或者即使互补链延伸发生也很慢。即使末端的两个碱基不互补，也可以发生互补链延伸。特别是当这两个碱基含有 G 和 C 的组合，并且 1 个碱基被另一个碱基取代时更是这样。在这种情况下，与寡聚体的两个末端碱基相邻的第三个碱基被次黄苷取

代或错配的碱基,以减弱末端的杂交力,其中在这两个碱基处的错配对于互补链延伸更为严重。结果,实际上互补链延伸只有在末端的两个碱基互补于 DNA 片段(匹配)时才发生。在对于 *Taq* 聚合酶反应最适的 75℃ 下进行反应。反应后,将溶液温度升高到 80—85℃ 以解离没有经受互补链延伸的杂交体。如就 DNA 片段 40 来说,那些作为选择序列的末端上的两个碱基没有精确互补的片段被除去。另一方面,经受互补链延伸的杂交体,如具有被经受链延伸的寡聚体保留之 DNA 序列的检测板 41,随着被延伸而固定到固相表面上。DNA 片段的分级分离效率经重复进行该反应循环而增加。例如,具有 16 个种的 DNA 寡聚体与特异性 DNA 片段间在一次反应型全杂交的可能性是 1/16。这就意味着只有 1/16 的 DNA 片段可被适当的 DNA 寡聚体捕获以进行链延伸。在第二次反应循环试探后,1/16 残余的未被捕获的 DNA 片段(量为总 DNA 片段的 15/16)与寡聚体适当杂交并在互补链延伸后补捕获。重复十次后,这些片段中有一半被经过延伸的寡聚体捕获在特定区域。为了只分级分离以这种方式被分级分离的 DNA,使溶液温度升高至大约 85℃ 以解离没有经受互补链延伸的杂交体,并用缓冲液洗掉。于 95℃ 洗脱滞留在各凹槽内的 DNA 片段并使用毛细管等分级分离各凹槽内的样品。在某些情况下,可使用无填凝胶的毛细管进行分离,从而保证更清晰地分离和分级分离。这里使用凝胶电泳法,这是因为例如就回收的 DNA 片段 42,43 和 44 来说,可能存在 3' 末端的两个碱基是相同的不同 DNA 片段。再者,得到关于 DNA 片段长度的信息,便有可能确定是否可以一次进行 DNA 序列测定,这将为序列测定提供有用的信息。当 DNA 片段的长度是 1kb(1000 碱

基长)或较少时,使用长凝胶(50至60cm)电泳(可分离近1000个碱基),或者将序列读至500—600个碱基长,并使用新的引物确定下面的序列;这种方法(普通熟知的引走法)是适宜的。另一方面,当长度超过1kb时,用另一种限制性酶消化该片段以缩短长度,或者在克隆到质粒等载体上之后从末端删除(普通熟知的缺失法)。使用这样的手段,产生并分析有不同长度的DNA;该方法是值得推荐的。在DNA测序中,依据相关DNA片段的长度不同所用分析法也不同,所以确定用于凝胶电泳的DNA片段的长度是很重要的。为了用凝胶电泳法进行分离和分级分离,本发明人使用在内径为0.3mm且长度为20cm的玻璃毛细管内产生的柱凝胶(毛细管凝胶:4%)分级分离了DNA,同时借助荧光检测型电泳仪监测从毛细管底部熔融的DNA。包含在各部分中的加至DNA片段3'末端侧上之限制酶识别序列上的两个碱基是已经知道的;因此,使用除了有被连接之寡聚体的或poly A尾部的共同部分外在3'末端还具有与这两个碱基互补之序列的引物,以及只有与已知的共同部分序列(经连接而加入的寡聚体的一部分)互补之序列的引物进行PCR扩增。增加双股DNA片段拷贝的数目以用作测序模板,并进行分析。

图1中参考编号1是指片段分割过程和用第一种限制性酶粗略分离样品DNA的方法,2是显示用第二种限制性酶再次进行片段分割的过程;而3则是指用寡聚体固相化检测板分离DNA片段的方法。

在本实施例中,在使用保留寡聚体的检测板就各自末端序列分级分离各种不同的DNA片段后进行电泳。这一次序也可以倒过来。

也就是说，用凝胶电泳法根据长度分离不同类的 DNA 片段。当 DNA 片段 1kb 长或较短时，根据长度分离的准确性为 1 至 2 个碱基。就 DNA 片段混合物来说，对每个部分使用 2 至 6 种是容易的（DNA 片段是作为两条链来处理，所以进行两种 DNA 片段切割是最少的）。使用 16 种寡聚体对每个部分中 DNA 片段的末端两个碱基（限制性酶消化序列之后的两个碱基）进行 DNA 互补链延伸。根据是否有互补链延伸来鉴定并分级分离末端碱基，并进行 PCR 扩增以提供 DNA 测序所需的样品。

图 8 显示用于本实例之 DNA 分析系统配置的一个例子。

实施例 2

图 6 是代表碱基序列测定之一例子的 DNA 片段图谱。该图显示对经用 *Hind* III 和 *Sau* 3AI 消化 M13 噬菌体 DNA 所得到的片段测序的结果。如图下所示，测序引物具有与共同序列 300 的部分互补的序列和在末端用于选择 DNA 片段 301 的两个碱基。如果用在末端有错配的两碱基的引物，互补链延伸就完全不能进行，所以也就不能获得序列信息（图 6—(5)）。如果引物与只在该部分中的 DNA 片段的一种匹配，则使用有两碱基选择序列的匹配的引物经碱基测序反应得到图 6—(1) 中所示的萤光谱。这表明可根据 3' 末端序列差异成功地使用引物从混合物 310 中选择片段。使用基于凝胶电泳的 DNA 序列仪保证测序的完成。另一方面，当不进行 DNA 片段的电泳分离，或因此而在此部分（管）中有许多片段时，则很可能在一个部分（管）中包括两种或多种具有相同末端 2 碱基序列的 DNA 片段。例如，当含有与一个引物匹配的两 DNA 片段时，如图 6—(2) 中所示的用一个引物获得的测序谱中可归属于两

种 DNA 片段的峰即彼此交迭,这样就使情况复杂化。使用 3' 末端上具有用于从混合物中选择片段之三碱基序列的引物 311 确定序列,并完成 DNA 测序反应;然后即可因此而获得关于个别碱基序列的信息(图 6—(3)和(4))。如果将具有几万个碱基或更长的长 DNA 消化成例如有几百个碱基的短片段。将形成许多种片段,这将花费大量时间。这样使用制得平均 4 千碱基片段的 6 碱基识别酶消化 DNA,即没有在一次产生许多种片段,并可用电泳法分级分离,然后用 4 碱基识别酶消化各片段,以制得有允许分析之大小的片段。然后用凝胶和 *Oligo*—检测板分级分离;这种方法是可取的。

实施例 3

本实例代表原始 DNA 拷贝的数目小的情况。用酶或其他方法将 DNA 切成片段,经连接反应结合寡聚体。按照其长度用凝胶电泳法分级分离各片段。在一个部分(管)中包含几种 DNA 片段。将它们倒入其底部固定有用于分级分离 DNA 之寡聚体的容器中。与 DNA 片段杂交后使之发生互补链延伸。然后将温度升至 80℃ 以解离没有经历互补链延伸的杂交体。降低温度后,重复进行互补链延伸。将所说的过程重复几次后,加入与用于连接的寡聚体有同样序列的引物,并发生 PCR 扩增。扩增发生在溶液中和固相表面上,但稍稍降低所加引物的量将增加固相表面上互补链延伸的百分比例。此外,如果将不同序列的寡聚体结合到 DNA 片段的两末端,则 PCR 扩增只发生在固相表面上。即是说,将一侧有寡聚体序列的引物固定到固相表面,而将与结合到另一个 DNA 末端上的寡聚体有相同序列的寡聚体放入溶液中。只有在两引物制成后才发生 PCR 扩增;因此, DNA 片段扩增只发生在固相表面上。

已在本实例中使用了平板形成的固相表面,有可能使用小珠、毛细管或多孔固相以增加表面积,从而增被俘获的寡聚体的数目。

如上文讨论的,当 *DNA* 片段的扩增只发生在固相表面上时,用固定到固相表面上的寡聚体进行 *PCR*,以扩增所需的 *DNA* 并挑选之。这样便有可能在分立的状态下同时增加各个部分的不同的 *DNA*。这样不仅便于分级分离,而且可以节省试剂。

在本实施例中,用限制性酶消化 *DNA*,并通过连接将已知序列的寡聚体结合到经消化的 *DNA* 上,或者借助末端转移酶在 3' 末端上形成 *poly A* 尾部并使用之。应注意的是,可以使用末端转移酶将碱基一个接一个地加到单链 *DNA* 链的 3' 末端上。

作为本方法的一种可代替方法,可以用超声波等随机消化 *DNA*,并借助末端转移酶在 3' 末端形成 *poly A* 尾部。使用与该 *poly A* 尾部相邻的几个碱基的序列作为选择序列,从而分级分离特异性 *DNA*。在这种情况下,在第一次选择中很难得到单一种 *DNA* 片段,所以应重复选择几次。也就是,如果第一次是借助末端的两个碱基(如 *GG*)选择,然后通过包括这两个碱基在内的末端 4 个碱基(如 *GAGG*)进行选择。对于第三次,则可通过包括此 4 个碱基在内的末端 6 个碱基选择。以这种方式重复选择。通过互补链延伸进行特异链选择的选择性主要取决于末端 1 和 2 个碱基杂交的匹配程度。重复进行末端两碱基的选择对于改善分级分离精确性是有效的。

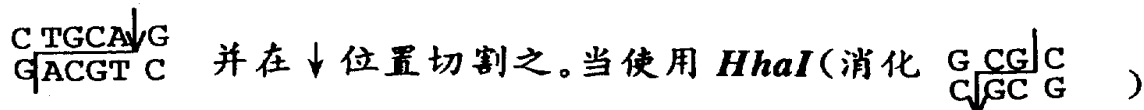
再者,将生物素标记物放在寡聚体中并使之与 *DNA* 杂交。完成互补链延伸后,与固定在固相层上的抗生物素蛋白形成生物素-抗生物素蛋白键合,从而使之被固相层捕获。

实施例 4

选择 *PUC DNA* 作为 *DNA* 测序的样品 *DNA*。 *PUC* 的 *DNA* 大小约为 2.7 千碱基,这一长度对于一次序列测定显得太长。根据本发明,用酶将 *DNA* 消化成可经一次操作进行序列测定的小片段。使用 16 个引物的文库(两碱基选择性引物)平等地测定所产生的许多小片段的序列。图 9 显示该实例的流程,用限制性酶 *HhaI* 消化 *DNA*(如消化成 $5' \cdots \text{GCGC} \cdots 3'$ $3' \cdots \text{CGCG} \cdots 5'$),并以凝胶电泳法分级分离所得片段。当然,可使用任何类型的限制性酶。图 10 显示电泳图谱。进行分级分离将其大致分成三类(级分 1 至级分 3)。然后将寡聚体导入末端。使用作为 *DNA* 修饰酶的连接酶将寡聚体与 *DNA* 连接。或者使用末端转移酶将 *poly A* 尾部接到 *DNA* 片段的 3' 末端上完成导入。按照连接酶方法,被结合之寡聚体的长度是恒定的,并且产生与原始长度一致的恒定长度的片段。所以该操作可在电泳之前完成。另一方面,在 *poly A* 尾部方法中,由于末端转移酶活性在每次使用中的 *poly A* 的长度有所不同。在以电泳法分离时,带将延伸并且能够得到准确的分离长度。所以用凝胶电泳法分离 *DNA* 片段,然后接上 *poly A* 尾部。级分 1 至 3 中包含多种 *DNA* 片段。用选择性 *DNA* 探针分析它们或对每种 *DNA* 片段进行分级分离。

选择性 *DNA* 探针当然被用作聚合物反应的引物,因此也用作 *DNA* 测序引物。引物库由其中具有与被导入之序列互补的序列、限制性酶的识别序列的一部分($X_1 - X_2 \cdots X_n$; $n = 2 - 6$)及选择性序列的引物所组成。引物可被表示为 $5' \cdots X_1 \cdots X_n \cdot Y_1 \cdots Y_m 3'$ 。文库中引物的特征是每个序列 3' 末端都有选择 *DNA* 片段的 1 至 3 个碱基,及与被导入之序列相邻的一部分识别序列。当引物与片段杂交时,

序列 $X_1 \cdots X_n$ 识别片段末端。至少在聚合酶反应期间被导入的序列使杂合体稳定。 $Y_1 \cdots Y_m$ 的选择性序列根据是否与片段序列匹配,通过促进或抑制聚合酶反应来区分各不同片段。酶 *Pst I* 识别序列



时,序列则变成 *CGC*。在选择性序列中, $Y_1 \cdots Y_m$ 可以取 *A, C, G* 或 *T* 中的任何一种碱基。当 $m=1$ 时, $Y_1 = A, C, G$ 或 *T* 且文库中有4个引物。当 $m=2$ 时, $y_1 \cdots y_2 = AA, AC, AG, AT, CA, CG, \cdots TT$ 且文库中有16个引物。同样,当 $m=3$ 时,文库中有64个引物。

可以制备每种限制性酶的文库。实践中,有3至4种酶的引物文库即足够用于各种 *DNA* 片段的序列选择。有两碱基选择序列的16种引物的引物库是最有效的,因为引物的数目并不太大,而且可有效地进行片段选择。

在第一个例子中,用末端2碱基选择引物确定工作引物。其显示两种情况,即直接从混合物中测定序列以及在分级分离后测定序列。从每管中取一部分并分成16等份。向各等份中加入选择性 *DNA* 探针 *Tag* 聚合酶和三磷酸脱氧核苷酸(*dNTP*),从而发生 *DNA* 聚合酶链反应。选择性 *DNA* 探针是荧光团标记的引物,该引物在靠近5'末端接有生物素和硫氰胺101,并且其序列是5'*TT*...*TCGCXY* ($X, Y: A, C, G$ 或 *T*) 中的任何一种。只有含匹配之引物的部分才产生延伸的引物(作为反应产物)。经用凝胶电泳法分析该产物的长度后,弄清与 *DNA* 片段准确杂交(匹配)的引物及片段的长度。然后,借助表面上结合有抗生物素蛋白的小珠捕获这些 *DNA* 探针与 *DNA* 片段间形成的杂交体,并分级分离之。当然,就测序来说可将此捕获步骤删除,因为有可能直接从片段混合物中测知 *DNA* 序

列。

使用与所用的 *DNA* 探针有相同序列的引物，按照常规方法制备 *DNA* 样品，以进行 *DNA* 序列测定。用作测序模板的 *DNA* 片段可以是混合状态的，但在借助小珠将其捕获后再使用则可获得更好的结果。用上述16种引物的小文库确定各片段的序列。将各序列与片段之切割位点的相邻序列的信息结合起来以构建完整序列。在这种情况下，测定用其他酶消化之片段的序列并且于连接。

在任何情况下，该方法都不需要克隆步骤，也不需要使用象用于以引物文库完成引物行走那样的多至几百或几千个引物的文库。

如上文所讨论的，根据本发明，将根据 *DNA* 片段的长度进行分离和根据 *DNA* 末端序列进行分离结合起来，可以很容易地用分立的引物和未经克隆过程即可得到的单个 *DNA* 片段完成 *DNA* 序列测定。结合使用 *PCR* 方法得以在没有经过克隆的情况下分割长 *DNA*，并得以分离这些切割得到的片段，从而增加进行碱基测序分析所需的拷贝数目。

说明书附图

图 1

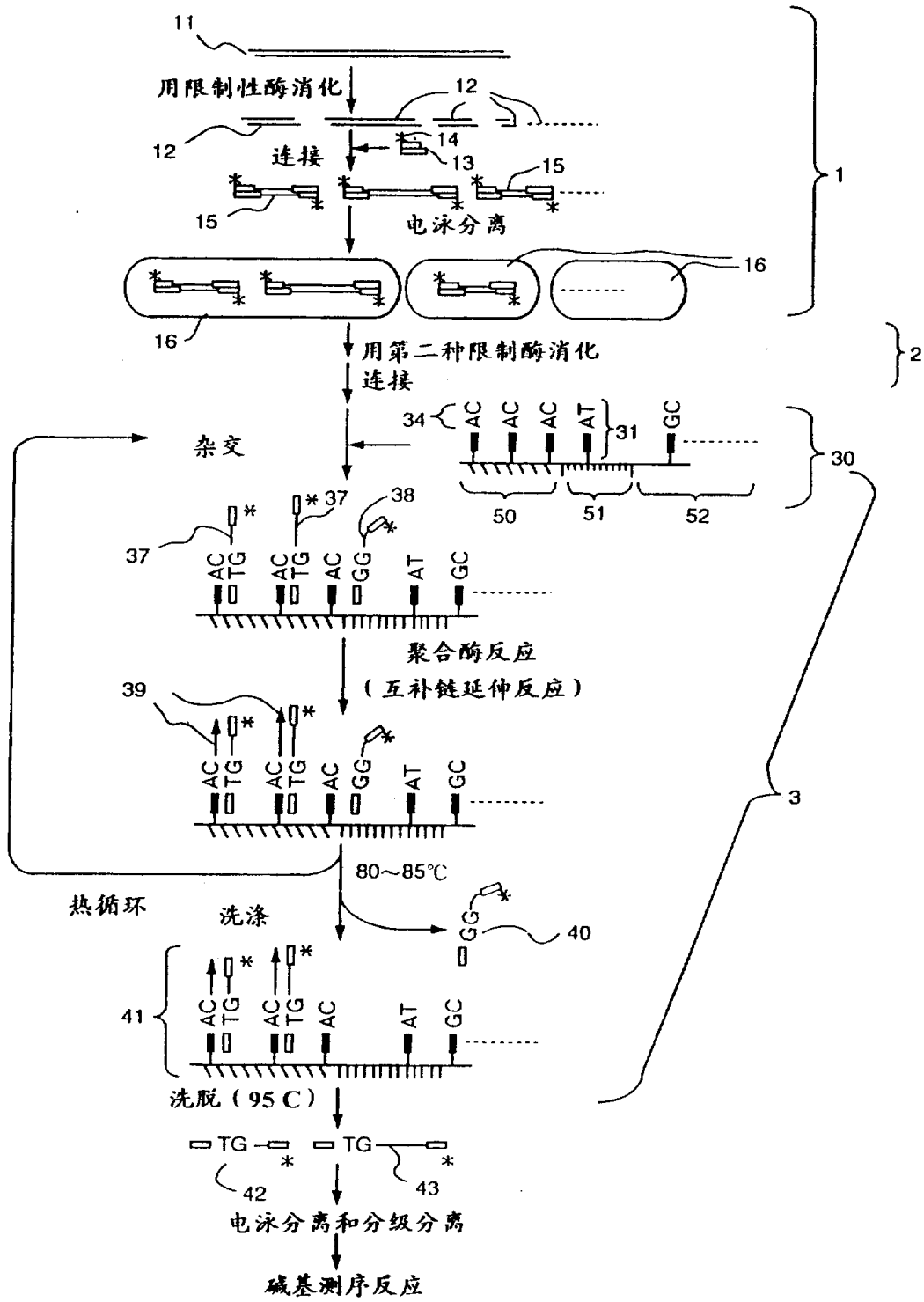


图 2

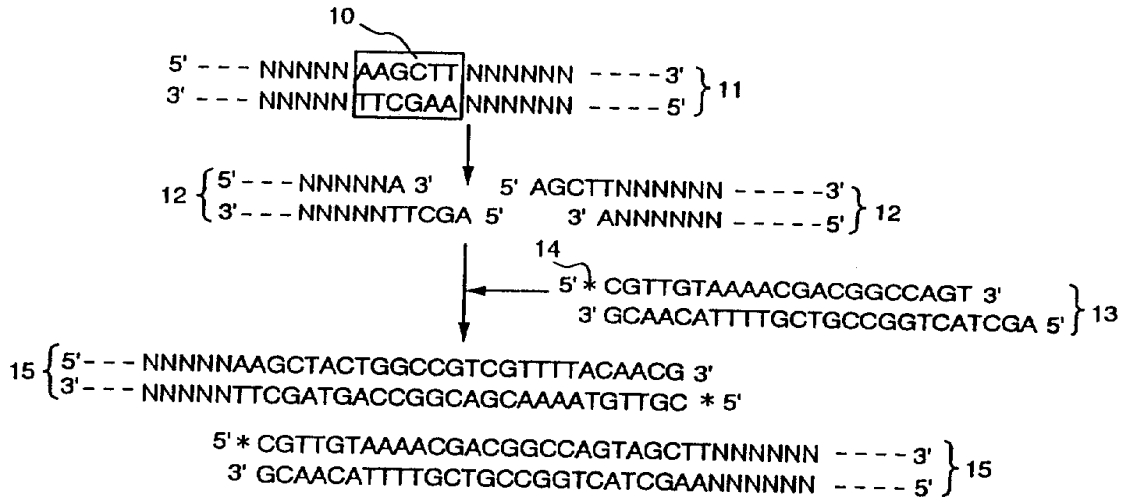


图 3

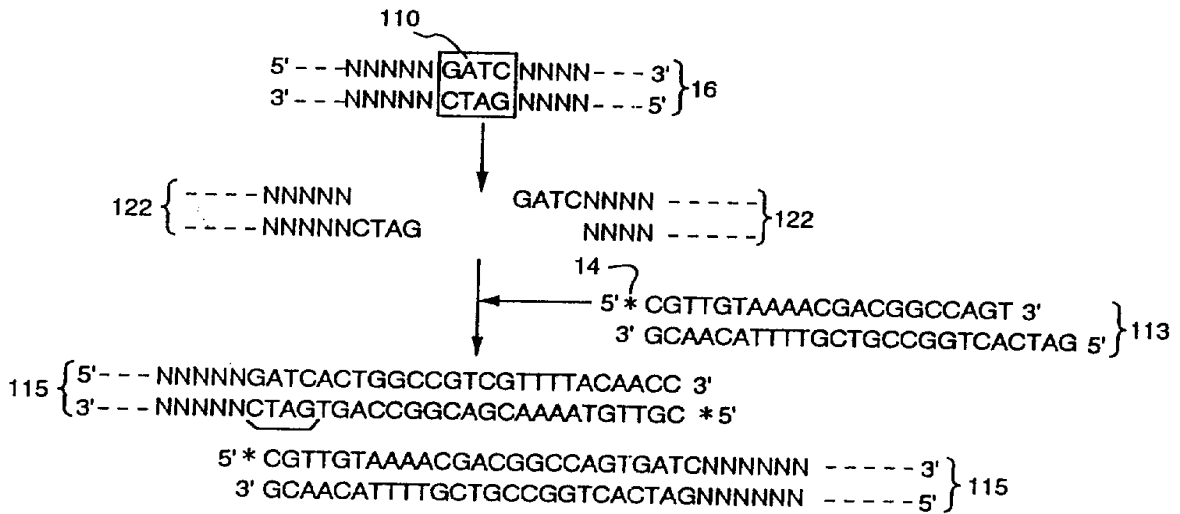


图 4

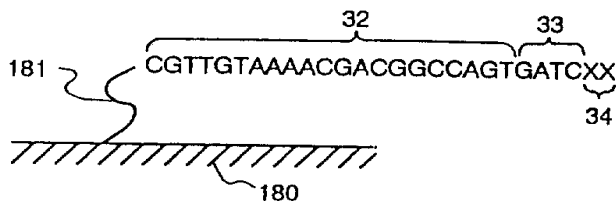


图 5

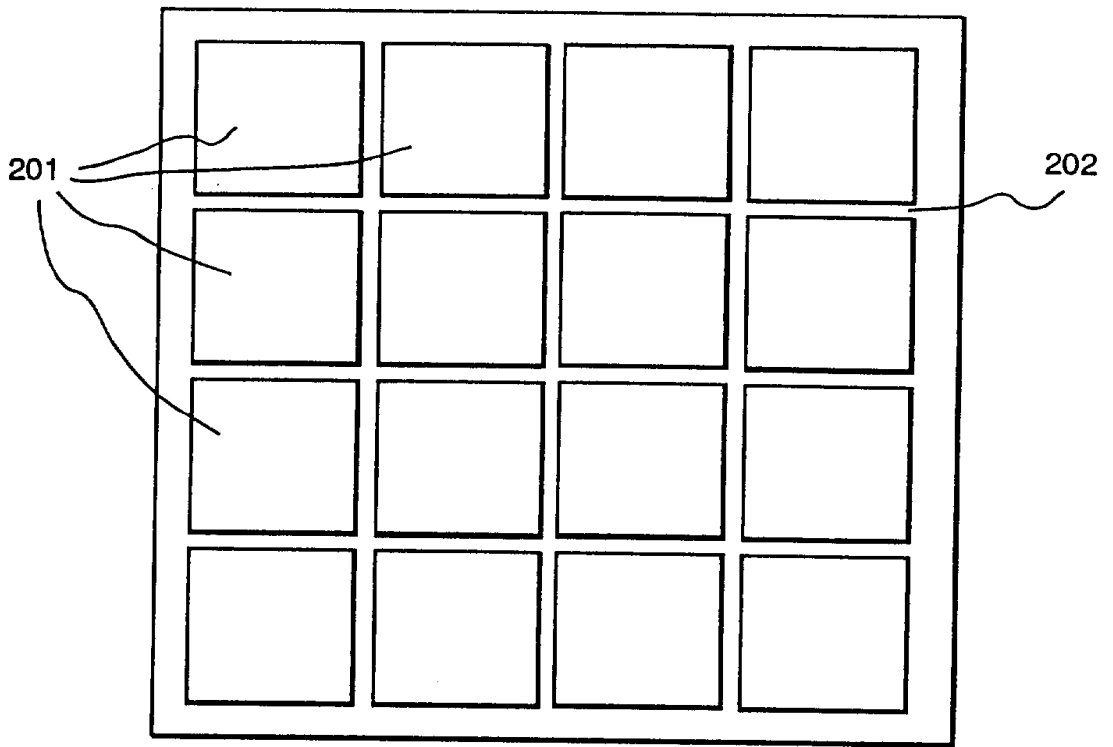


图 6

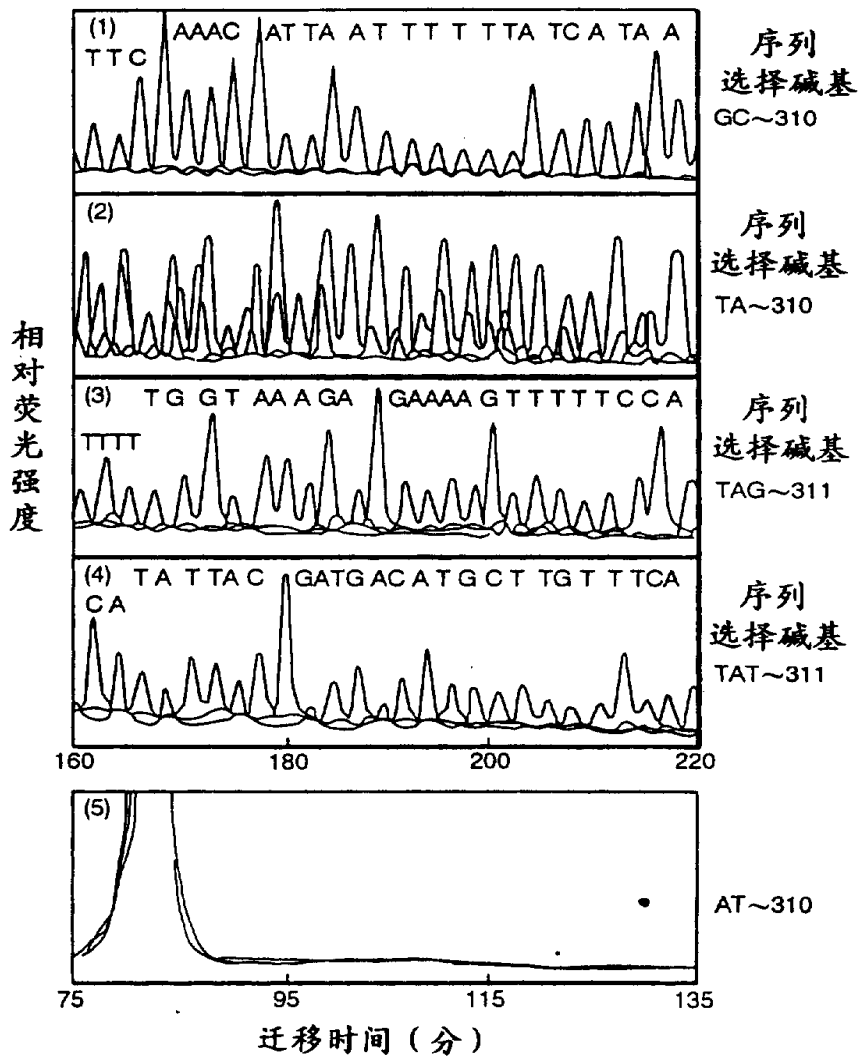


图 7

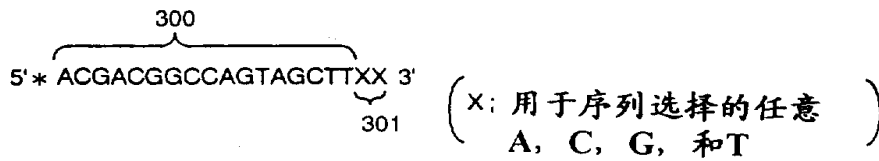


图 8

DNA片段分级
分离亚系统

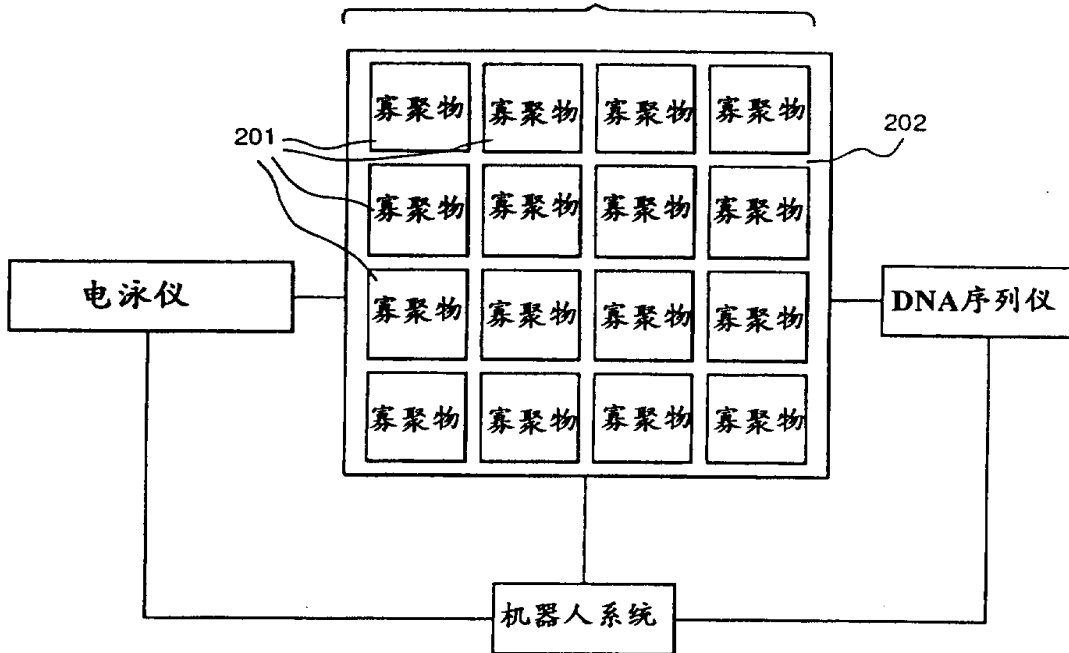


图 9

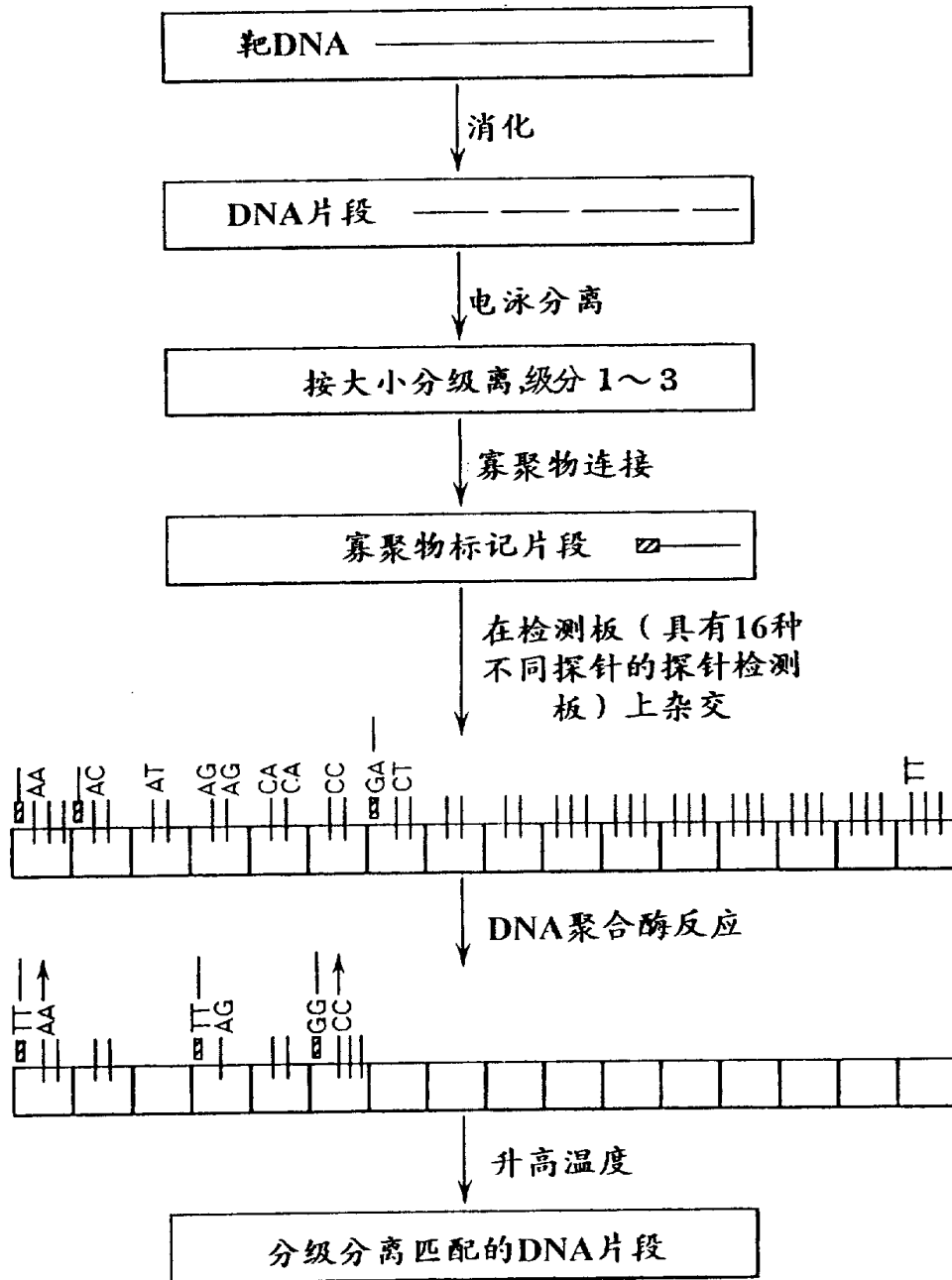


图 10

