[19] 中华人民共和国国家知识产权局

[51] Int. Cl[7]
C12Q 1/68

[12] 发明专利说明书

[21] ZL 专利号 96121784.7

[43] 授权公告日 2003 年 4 月 16 日

[11] 授权公告号 CN 1105782C

[22] 申请日 1996.11.29 [21] 申请号 96121784.7

[30] 优先权

[32] 1995.11.30 [33] JP [31] 311949/1995

[71] 专利权人 株式会社日立制作所

地址 日本东京

[72] 发明人 神原秀记 冈野和宣 植松千宗

[56] 参考文献

CN1038309A 1989.12.27 C12Q1/68

EP0630972A2 1994.12.28 C12Q1/68

WO9317126 1993.02.02 C12Q1/68

分子克隆实验指南第二版 1992－12－01 J,萨姆布鲁克等科学出版社

审查员 孟海燕

[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

权利要求书 1 页 说明书 19 页 附图 4 页

[54] 发明名称 分析或检测多核苷酸的方法及其分析仪或装置

[57] 摘要

一种分析或测定核苷酸的方法，该方法包括：(1)用限制性酶消化 DNA 的步骤；(2)用 DNA 探针辨别得自以上步骤(1)的 DNA 片段在其 3′末端附近的序列的差异并通过互补链合成延伸 DNA 探针而将 DNA 片段分级分离成几组的步骤；和(3)测量属于所述组的 DNA 片段长度或通过所述互补链延伸反应延伸的 DNA 探针长度的步骤；其中对于该 DNA 片段 3′末端附近的每个碱基序列所测得的长度被用作指纹。

ISSN1008-4274

1．一种分析或测定核苷酸的方法，该方法包括：

(1) 用限制性酶消化DNA以得到DNA片段；

(2) 用标记的DNA探针辨别DNA片段在其限制酶识别位点附近1—3个碱基的序列的差异，其中所述探针在其3′末端带有由选自A、C、G和T之中1—3个碱基的所有组合；

(3) 用互补链合成将标记的DNA探针延伸，以将DNA片段分级分离成几个组；和

(4) 测量属于所述组的DNA片段长度，或测量通过所述互补链合成延伸的标记的DNA探针长度；其中分级分离入各个组的DNA片段所测得的长度被用作指纹，所述分级分离是通过DNA片段邻近限制酶识别位点的1—3个碱基的序列差异进行的。

分析或检测多核苷酸的方法及其分析仪或装置

本发明涉及用 DNA 进行诊断，检测 DNA 的性质以及分析或检测 DNA 的方法。

用 DNA 诊断疾病已日益普遍。在这种疾病诊断方法中，（1）制备具有与靶 DNA 互补序列的 DNA 探针并进行探针检测，看这种 DNA 探针是否能与靶 DNA 杂交；（2）选出编码靶 DNA 的某一序列区并用两个 DNA 探针（引物）对其进行聚合酶链式反应扩增，阅读所得的 DNA 片段或检验其长度，并根据获得的 DNA 片段信息进行检测；等等。将这样得到的结果用于诊断等方面。这些方法适用于分析或检测单一类型的 DNA、至多是几种类型的 DNA，但不适合于大量 DNA 片段的 DNA 检测或卡 DNA 的全面评估。

不过，DNA 或基因组通过在体内相互作用而发挥其功能。所以非常需要从总体上鉴定染色体或其中所含的全部 DNA。例如，在基因组计划中引人关注的 cDNA 中，已经尝试对互补于 mRNA 的 cDNA 的类型和数量进行检测从而从整体上了解一种生物体，注意到 DNA 在生物体中发挥作用的机理，当 DNA 在生物体中发挥作用时，DNA 信息首先被转录到 mRNA 上，基于这一信息合成蛋白质，以对生物体发挥作用。在这种尝试中，从活样品中取出 cDNA，对各种 cDNA 进行测序以分析每种 cDNA 在一组织中出现的频率（机体作图）。

机体作图包括下列步骤。首先，从 mRNA 制备 cDNA（各种 cDNA 的混合物），再进行克隆。把含有各种 cDNA 的大肠埃希氏杆菌（E.coli）在琼脂糖平板上涂布并培养以获取菌落，每一菌落含有所需 cDNA 中的一种。挑出所需 cDNA 并测序以鉴定 cDNA 的类型。按相似方式从每一菌落挑出所需 cDNA 并测序，这样就经常出现相同的 cDNA。当关注一种组织中存在的某一特定 cDNA 时，这一特定 cDNA 的数量越大，特定

cDNA 与在组织中强烈表达的基因的相关性就越高，所以该基因在菌落中出现的频率也就越高。因而提出了一种测定 cDNA 出现频率的方法，该方法包括：在许多菌落中进行 cDNA 测序，观察某一特定 cDNA 在各菌落中出现多少次（ Katsuji, Murakawa 等， Genomics, 23, 379-389（1994））。

另一方面，还提出了用于 DNA 诊断的另一种尝试，它关注的是一个基因组（所有染色体中的 DNA）或一条特定染色体的整体分布。一种称之为基因扫描、也称为 Landmark 基因组扫描（LGS 方法）的指纹技术涉及下列步骤：用象 Not I 等等这样的 8 碱基切割限制性酶（每 48 ~ 64kbs 消化一次）选择性消化 DNA；把消化过的片段结合到放射性同位素标记或用荧光团标记的核苷酸上；用电泳法分离这些片段；然后，用 4 碱基切割限制性酶（每 44 ~ 256 碱基消化一次）；对聚丙烯酰胺板凝胶的上端的 DNA 片段进行双向电泳。把这样获得的图样用作指纹，借此了解 DNA 的整体分布。一种尝试包括图样的诊断用途，已知正常细胞中的 DNA 图样与患有癌症等的异常细胞的 DNA 图样不同。

不过，迄今还没有发现一种优良技术，因为在前述方法中应该检验长 DNA 的哪一特定位点存在异常。

对疾病的早期检测或了解细胞中 DNA 的功能而言，检验长、大 DNA 或弄清含不同 cDNA 的样品的整体分布是很重要的。然而，如前所述，尚未开发出足以达到这一目的优良技术。根据上文中阐述的技术，必须测定众多克隆的碱基序列。所要求的人力和时间使这些技术不可能实际用于各种样品。常规的 DNA 探查只足以在一个操作周期中检验至多几种到几十种 DNA，但不适合于在一个操作周期中检测几百到上千种 cDNA 或 DNA 片段。此外，上述 cDNA 分析法不适合于存在异常的长 DNA 的检测。

另一方面，基因扫描技术可以满足前述要求，但也遇到一些问题，即在用限制性酶第二次消化过程中和双向电泳中要消耗大量酶、作为第一次消化中出现的 DNA 片段长度的尺度的横坐标和作为最终片段的长度的某种尺度的纵坐标并不总是定量，所以难以用这些数据建立数据库。

本发明的一个目的是克服上述问题、提供一种新的指纹技术即DNA分析法，该方法适用于各种样品并且适合于测定大量cDNA或DNA片段和测定长DNA。

本发明的DNA分析或测定方法具有下列特征。首先，用限制性酶切割样品中所含的DNA而产生具有样品固有长度的DNA片段。将带有已知序列的寡聚体连接到产生的DNA片段的3′末端上以形成一个引导区。制备一种能与这个引导区、识别序列和与识别序列相邻的几个碱基杂交的DNA探针。这一DNA探针由一套16个DNA探针组成，例如，其中3′末端的两个碱基是基本上所有碱基的组合。使这16个DNA探针独立地与样品DNA杂交，进行互补链合成。所合成的链用荧光团等标记。引发互补链合成时，对互补链合成产物针对所用每一种DNA探针进行凝胶电泳，并测定每个片段的长度以获得指纹。下文简要描述本发明的组成。

A.根据本发明的分析或测定核苷酸的方法的第一方面包括：

（1）用限制性酶消化DNA的步骤；

（2）用标记的DNA探针（因互补链合成而掺入的标记的核苷酸也可用作标记的DNA链）的3′末端序列辨别得自以上步骤（1）的DNA片段在其3′末端附近的序列的差异并只有在末端序列配对时才通过互补链合成延伸标记的DNA探针而将DNA片段分级分离成几组的步骤；和

（3）测量属于所述组的DNA片段长度或通过所述互补链延伸反应延伸的DNA探针长度的步骤；

其中对于该DNA片段3′末端附近的每个碱基序列所测得的长度被用作指纹。

B.根据本发明的分析或测定核苷酸的方法的第二方面至少包括：

（1）用限制性酶消化DNA的步骤；

（2）将脱氧核苷酸或含有脱氧核苷酸，类似物的寡聚体与得自以上步骤（1）的DNA片段在其3′末端连接的步骤；

（3）用一套4个或16个标记的DNA探针进行互补链合成使所述标记的DNA探针延伸的步骤，标记的DNA探针具有互补于前述连接的

寡核苷酸或互补于其3′末端的1个或2个碱基的一部分识别序列的任意序列并具有不同的末端序列；和

（4）对通过互补链合成而延伸的链进行电泳以检测DNA片段的步骤。

在上述方法A和B中，本发明的特征至少在于：标记物是生物素、化学发光试剂或荧光团；通过检测荧光团来检测DNA探针或DNA片段；以及标记的DNA探针含有一套16个标记的DNA探针，其中至少两个末端碱基由基本上所有碱基种类或其类似物构成。

C.根据本发明的分析或测定核苷酸的方法的第三个方面至少包括：

（1）用限制性酶消化DNA的步骤；

（2）将脱氧核苷酸或含有脱氧核苷酸类似物的寡聚体与得自以上步骤（1）的DNA片段在其3′末端连接的步骤；

（3）使用在其3′末端具有由任意碱基种类组成的1～3个碱基的所有组合的每个标记的DNA探针进行互补链合成而使每个所述标记的DNA探针互补链延伸的步骤，标记的DNA探针是与以上步骤（2）中连接的寡核苷酸或与所述限制性酶所识别的一部分识别序列互补的；

（4）将通过每个标记的DNA探针的互补链合成延伸的DNA链按照末端序列进行分类和分离的步骤；

（5）将第二个寡聚体引入通过互补链合成在其3′末端延伸的DNA链中的步骤；

（6）用引入了第二个寡聚体的DNA链作为模板进行步骤（3）和（4）的步骤；和

（7）分级分离并根据长度，检测通过互补链合成延伸的DNA链的步骤。

在上述方法C中，本发明的特征在于：通过重复上述步骤（3）和（4）的，DNA片段或DNA片段的互补链相互间可以分级分离从而检测DNA，因为标记的第二个DNA探针的每个末端碱基序列具有第二个寡聚体的互补链。通过重复前述步骤（3）和（4），可以增加级分的数目，也就可以进行更为详细的分析。而且，方法C适用于含有更多片段的样品。

D.根据本发明的分析或测定核苷酸的方法的第四个方面至少包括：

（1）用限制性酶消化DNA的步骤；

（2）将脱氧核苷酸或含有脱氧核苷酸类似物的寡聚体与得自以上步骤（1）的DNA片段在其3′末端连接的步骤；

（3）使用在其3′末端具有由任意碱基种类组成的1～3个碱基的所有组合的第一个标记的DNA探针进行互补链合成而使第一个标记的DNA探针的互补链延伸的步骤。标记的DNA探针是与以上步骤（2）中连接的寡核苷酸（或寡聚体）或与所述限制性酶所识别的一部分识别序列互补的；

（4）将第一个标记的DNA探针的每个末端碱基通过互补链合成而延伸的DNA链之互补DNA链分类和分离的步骤；

（5）用不同于以上步骤（3）中所用的第一个标记的DNA探针的第二个标记的DNA探针再次进行上述步骤（3）的步骤；和

（6）分级分离并根据长度检测经互补链合成延伸的DNA链的步骤，该DNA是由以上步骤（5）中得到的第二个标记的DNA探针生成的。

在上述方法A－D中，本发明的特征在于：所述限制性酶是用于形成3′－突出端或平端型片段的限制性酶。方法A－D中，本发明的特征还在于：其中所用限制性酶是用于产生5′－突出端的限制性酶，而且这些方法还包括，在以上步骤（1）之后，用DNA聚合酶填补用限制性酶消化的部分以形成双链的步骤或者通过连接引入含有被消化部分的序列的寡聚体的步骤。

在方法A中，本发明的特征在于：基于两个碱基的差别分级分离的DNA片段的互补链至少用荧光团或化学发光试剂进行标记并用具有在其3′末端的任意两个碱基所有可能组合的标记的DNA探针通过互补链合成而产生。

E.在方法A－D中，使用的标记的DNA探针是一套至少16个DNA探针，它们是N1，…Nn的所有组合形式（其中5≤n≤2（7），X1，…，Xm（1≤m≤6）；Y_1Y_2是A，C，G和T任何一种；所述标记的DNA探针是用于互补链合成的引物并用下式表示：

$5' - N_1 \cdots Nn, X_1 \cdots XmY_1Y_2 - 3'$

其中 $N_1 \cdots Nn$ 具有基本上互补于与 DNA 片段连接的寡聚体的序列；$X_1 \cdots Xm$ 基本上互补于用限制性酶消化的部分的一部分序列；Y_1Y_2 由 A，C，G 和 T 中任意两个的组合构成。

F.在方法 A－D 中，使用的标记的 DNA 探针是一套至少 16 个 DNA 探针，它们是 N1，…，Nn 的所有组合形式（其中 5≤n≤2（7），X1，Xm（1≤m≤6）；Y_1Y_2 是 A，C，G 和 T 任何一种；Zi（1≤i≤3）是可与多种核苷酸杂交的核苷酸似物；所述标记的 DNA 探针是用于互补链合成的引物并用下式表示：

$5' - N_1 \cdots Nn, X_1 \cdots Xm\ Zi \cdots Y_1Y_2 - 3'$

其中 $N_1 \cdots Nn$ 具有基本上互补于与 DNA 片段连接的寡聚体的序列；$X_1 \cdots Xm$ 基本上互补于用限制性酶消化的部分的一部分序列；Y_1Y_2 由 A，C，G 和 T 中任意两个的组合构成。

F 中的 DNA 探针中，本发明的特征在于：核苷酸 N_1，…，Nn 任一个用生物素或荧光团进行标记。

G.在方法 A－D 中，用限制性酶消化的 DNA 是用下列步骤获得的 DNA：（1）从组织样品收集 mRNA 以制备 mRNA－cDNA 互补对的步骤；和（2）从 mRNA－cDNA 互补对制备双链 DNA 以给出有关 mRNA 的种类和量的信息的步骤。

H.本发明的另一方面涉及 DNA 分析仪或 DNA 测定义，它包括：

用限制性酶消化 DNA 的第一反应容器；

进行第二个反应的多个第二反应容器，第二个反应包括：

在多个第二反应容器的每一个中混合在其 3′末端附近具有不同碱基序列的多个标记的 DNA 探针的任一个，其中将消化产物分级分离并装入第二反应容器中，

用标记的 DNA 探针辨别通过以上消化作用得到的 DNA 片段在其 3′末端附近的碱基序列的差异；

只有在末端序列配对时才通过互补链合成延伸标记的 DNA 探针，将 DNA 片段分级分离成几组；和

用于测量属于各组的 DNA 片段长度的电泳装置或，用经互补链合

成延伸的标记的DNA探针作为样品并测量该样品的长度;

由此得到所测量的DNA片段在其3′末端附近的每个碱基序列样品的长度作为指纹。

1.本发明的再一方面涉及DNA分析仪或DNA测定仪，它包括:

用限制性酶消化DNA并将脱氧核苷酸或含有脱氧核苷酸类似物的寡聚体与以上消化中得到的DNA片段在其3′末端连接的第一反应容器;

进行第二个反应的多个第二反应容器，第二个反应包括:

在多个第二反应容器的每一个中混合一套4或16个标记的DNA探针的任一个，其中将消化产物分级分离并装入第二反应容器中。标记的DNA探针具有互补于以上连接的寡核苷酸（或寡聚体）或互补于在其3′末端的1个或2个碱基的识别序列一部分的任意序列; 和

用于用所延伸的DNA链作为样品测量经互补链合成延伸的DNA链长度的电泳装置;

从而得到每个标记的DNA探针样品的长度作为指纹。

在DNA分析仪或测定仪H和I中，本发明的特征在于: 长度标记物和样品一起在所述电泳装置的凝胶电泳泳道中进行电泳从而对经电泳分离的样品片段进行分级分离和辨别，并用荧光团来标记样品和长度标记物，从而根据荧光团标记物的荧光波长加以辨别。

J.本发明的又一方面涉及DNA分析仪或核苷酸测定仪，其中，提供了多个反应容器和在控制温度的条件下进行化学反应的一个化学反应容器，并将至少一套16个标记的DNA探针分别装入多个反应容器中，并使分别装入多个反应容器中的每个DNA片段与标记的DNA探针同时反应。

尽管样品中存在着许多待分析的DNA片段，但在大多数情况中，每个DNA片段具有的序列不同于邻近于识别序列的序列。因此，根据邻近于识别序列的序列的差异将DNA片段初步分级分离成几组。重复进行这种分组操作，直到每组中DNA片段的数量降低到不会对用凝胶电泳分级分离DNA片段造成不便的水平时为止。最后，借助每个DNA片段的长度通过凝胶电泳检测每组中的DNA片段。

用具有两个末端碱基所有组合的一套16个DNA探针通过互补链合成对样品中用限制性酶消化DNA而得到的DNA片段进行分级分离和分类。当两个末端碱基完全与靶DNA杂交时进行互补链合成，否则不能进行。通过互补链合成而延伸的DNA链其长度基本上与原DNA链（用限剂性酶消化而得到的DNA片段）相同。当这样设计时，即当标记物被接纳到由互补链合成所延伸的DNA链中而使标记的DNA链能与由互补链合成而延伸的其它DNA链明显区分开时，对末端的两个碱基进行辨别以确定原DNA链的长度。

在单一分级分离不足以辨别在样品中用限制性酶消化DNA而得到的DNA片段的情况中，把生物素标记到DNA探针上并仅仅探查出延伸的链。再一次把分级分离用于探查出的DNA片段（由互补链合成而延伸的DNA链）。在这种情形中，所用的DNA探针是连接到互补链3′末端（相应于原DNA的5′末端）上的一种DNA探针，所以可以进行不同于第一个操作的分类。

也就是说，根据本发明，有关用限制性酶消化样品中DNA而得到的许多DNA片段的长度的信息并不用作指纹。而是借助末端碱基序列对用限制性酶消化而得到的DNA片段进行分级分离并分成几组。对每组中存在的DNA片段，用凝胶电泳测定片段长度并将片段长度用作指纹。所以本发明的优点在于：片段长度不会相互重叠但能达到高分辨力。

本发明适合于多种样品而无需消耗大量人力和时间，并可以提供适于测定许多DNA或DNA片段或者分析或测定长DNA的指纹图谱法，以及其中所用到的分析仪或装置。

根据本发明，不再需要克隆和培养步骤，即这些步骤在现有技术中为基因表达分析而进行的mRNA显性是必需的而且所需时间长，这样就可以在更短的时间内进行显性，所需时间少于现有技术的十分之一。此外，所获得的数据可以换算成数字数据，数字数据是计算中优选的数据库。

下面将参照图1描述本发明。用限制性酶HhaI消化样品1，并把聚腺苷酸加到每个DNA片段的3′末端碱基序列4，制备一套16个具有RXY结构（其中X和Y代表A，C，G或T）的荧光团标记的DNA

探针 6。把含有由所有片段组成的样品 DNA 的溶液分成 16 个级分 15，把不同的 DNA 探针 6 加到各级分中以进行互补链合成。DNA 探针与 DNA 片段杂交，但如 11 所示只有在 3′末端完全杂交的 DNA 探针（*R-AA）才通过互补链合成而得以延伸。用荧光凝胶电泳装置鉴别出具有延伸的 DNA 序列相同长度即延伸的互补链长度的互补链 13。在凝胶电泳图谱中，片段长度不会相互重叠，但能获得长度的高分辨力。所以，本方法和分析仪或装置适合于多种样品。

图 1 是解释本发明实施例 1 中操作程序的图示。

图 2A 是由混合的 DNA 片段测定的电泳图谱，混合的 DNA 片段是通过把本发明实施例中的操作程序用于样品－DNA 而获得的。

图 2B 是由按照末端碱基序列分类的 DNA 片段测定的电泳图谱，分类的 DNA 片段是用一套 16 个 DNA 探针通过把本发明实施例 1 中的操作程序用于样品－DNA 而获得的。

图 3A 表示 PUC19 的电泳图谱，它是本发明实施例 2 得到的。

图 3B 表示 PUC118 的电泳图谱，它是本发明实施例 2 得到的。

图 4 是解释本发明实施例 2 中的操作程序的图示。

本发明适合于象 DNA 或 RNA 这样的多核苷酸样品。通过指纹分析或采用指纹法的测定技术进行多核苷酸样品的分析或测定。本发明还涉及在 DNA 分析或 DNA 测定中应用的 DNA 分析仪或 DNA 测定仪。

下面将参照附图详细描述本发明的实施方案。

实施例 1

该实施例主要用于解释本发明的基本原理。对于 DNA 诊断，通常使用的技术是用限制酶消化 DNA，通过凝胶电泳分级分离所产生的 DNA 片段并根据所获得的电泳图谱进行诊断。但是，随着由限制酶产生的 DNA 片段的种类增加，这些 DNA 片段不易被彼此区分，造成诊断困难。这些片段的特征在于其碱基长度和序列。虽然每一片段都有其固有序列，但均由数个碱基组成的各个末端序列通常也各不相同。例如，4－碱基切割限制酶的消化位点通常出现在平均每隔 256bp 处。因此，当用 4－碱基切割限制酶消化一约 50kb 的双链 DNA 时，产生约 200 个双链 DNA 片段（约 400 个单链 DNA 片段）。首先用与识别序列相邻的两碱基 3′

－末端序列分级分离这些DNA片段，两碱基序列的组合数变为4×4＝16种组合。属于每一组的单链DNA片段的平均数约为25个片段。具有从数个碱基到1kb的各种长度的约25个DNA易于分级分离。当完整的DNA很长或当所要处理的样品含有许多DNA片段时，利用较长的碱基序列对DNA片段进行分组。例如，当利用末端四碱基序列来分组时，DNA片段被分级分离为256组。当按长度对每组的DNA片段进行分级分离时，具有从几个碱基到1000个碱基在内的至少100种DNA片段可被区分，因此总共20,000个以上的DNA片段可被分级分离。由此获得的数据可用作指纹图谱。下面参照图1详细解释具体的步骤。

使用λDNA作为样品1，并使用HhaI作为限制酶、HhaI是一种消化双链DNA的限制酶，它将双链DNA：

```
5' - N ... NGCGCN ... N - 3'
3' - N ... NCGCGN ... N - 5'
```

消化成：

```
5' - N ... NGCG        CN ... N - 3'
3' - N ... NC        GCGN ... N - 5'
```

其中N代表A、T、G和C中任何一种。每一DNA片段3的3′末端碱基序列4为GCG。在图1中，与每一DNA片段3的3′－末端碱基序列4相邻的两个碱基10通常用NN表示，其中N为A、C、G和T中的任何一种。

通过连接将一寡聚物加到该DNA片段的3′末端。或者，利用末端脱氧核苷酸转移酶将poly A（它可为A、C、G或T）加到该DNA片段的3′末端。当加入polyA时，3′末端应为3′－突出末端或平末端。参照图1解释了加入poly A5的实施方案。当将poly A5加到每一DNA片段的3′末端时，3′末端的序列由下式表示：

GCGAAAA ... A - 3'

当加入 polyA 时识别序列 GCG 的存在是鉴别末端位点的关键。这是因为 polyA 的长度不能被控制。因此，制备了一套具有 SEQ No.1 所示结构的 16 种 DNA 探针 6：

5' - TTTTTTTTTTTTTTTTTTTCGCXY - 3'

其中 x 和 y 均代表 A、C、G 或 T（16 种 DNA 探针是所有可能的 x 和 y 组合）。在图 1 中，这些 DNA 探针由 RXY 表示，其中 x 和 y 均代表 A、C、G 或 T，且 R 代表 TTTTTTTTTTTTTTTTTTTCGC。符号*代表荧光团标记。当 CGC 序列缺失时，XY 部分就不能发挥其区分 DNA 片段中的 TG、TC、TT、TA 等的作用（因为 XY 是与 poly T 相连的）。

DNA 探针的 CGC 5′端的序列与通过用限制酶 HhaI 消化所获得的所有 DNA 片段互补，并可与所有这些片段杂交。下面解释其中 XY（即与 3′末端碱基序列 4 相邻的两个碱基 10）为 AA 的一个实施方案。在这一例子中，XY 部分仅与具有互补链序列 TT 的片段杂交，形成双链 11，当 XY 部分 10 不具有与该片段互补的序列时，则 XY 部分不与任何片段杂交，如 12 所示。并不能仅仅通过杂交的稳定性来确定 XY 部分是否完全杂交，这是因为几乎所有该 DNA 探针的位点均可与任何片段杂交，而仅仅依靠末端两个碱基序列上的差异并不足以用于鉴定。

但是，通过利用 DNA 聚合酶进行互补链的合成，就可以鉴别出末端两碱基是否完全杂交。将一含有样品 DNA 的溶液分成 16 等分（15），该样品 DNA 由通过用限制酶 HhaI 消化所获得的所有 DNA 片段组成。将不同的每种 DNA 探针 6 加到每一等分中以进行互补链的合成。DNA 探针与 DNA 片段杂交，但是只有 3′末端完全杂交的 DNA 探针（*R-AA）（如 11 所示）未通过互补链的延伸而被延长，从而获得与 DNA 片段具有相同长度的互补链 13，当 DNA 探针被荧光团标记时，利用荧光凝胶电泳装置即可鉴别出被延长的互补链的长度。

图 2A 和 2B 显示了利用－DNA 所获得的部分结果。当试图同时测定通过用限制酶 HhaI 消化所获得的所有 DNA 片段的长度时，许多峰相互重叠（如图 2A 的电泳图谱 21 所示），使得其鉴别不可能。但是，如图 2B 所示，在利用一套 16 种 DNA 探针 23（在图 2B 中仅显示了该 DNA 探针*RXY 的 XY 部分）根据末端碱基序列分组的 DNA 片段的电泳图谱 22 中，对应于各个片段的峰彼此分开。从而可测定出分组的 DNA 片段中每一 DNA 片段的长度。根据所要分析的完整 DNA 的种类，图 2B 所示的每一 DNA 片段的电泳图谱都是固定的。如果完整 DNA 的结构不同，将出现另一种电泳图谱。因此，可利用这些电泳图谱作为指纹图谱，后者反过来可用于 DNA 的诊断等。在图 2A 和 2B 中，横座标表示碱基长度（24）。

实施例 2

图 3A 和 3B 显示了应用本发明方法鉴定不同 DNA 所获得的结果。操作步骤与实施例 1 类似。即使用的样品是基碱基序列具有某些差别的两种 PUC，即 PUC19 和 PUC118。PUC19 和 PUC118 是分别由 2686 个和 3162 个碱基组成的 DNA，其中约有 2630 个碱基构成完全相同的共同序列。用限制酶 HhaI 消化层，通过连接反应将具有已知序列的寡聚物加到消化片段的末端。随后按与实施例 1 类似的方式，利用一套 16 种引物（DNA 探针）按消化片段的末端碱基序列对 DNA 片段进行分组，并获得每一组的电泳图谱。

图 3A 中所示的电泳图谱 31 得自 PUC19。图 3B 中所示的电泳图谱 32 得自 PUC118。图 3A 和 3B 中的横座标代表碱基长度 24。作为在 PUC19 的电泳图谱中缺失而仅在 PUC118 的电泳图谱中检测到 DNA 片段检测到的是约 75 个碱基的片段（CT、GT 引物）和约 450 个碱基的片段（CC、GT 引物），作为仅在 PUC19 中测得的片段则是约 130 个碱基的片段（CT、GT 引物）。

如上所述，通过应用本发明的方法可以简单方式将彼此稍有差异的各种 DNA 区分开。当然，本发明也可应用于基因组 DNA。当用限制酶 HhaI 消化 DNA 时，平均每隔 250 个碱基发现一个消化位点，因为 HhaI 是一种 4－碱基切割限制酶。在基因组的情况下，每 1M 碱基平均产生

约100个双链DNA片段。当利用8－碱基切割限制酶如NotI等检测约100M碱基的一基因组时，产生几乎同样数量级的DNA片段，因此该方法同样适用于基因组。

通过凝胶电泳极难对如此大量的DNA片段（按单链计算为2000－3000个片段）进行分级分离，并发现其电泳图谱上的细微差别。因此，按照被限制酶消化的每一DNA片段的3′－末端序列将DNA片段分为16组，从而使第一组中平均含有约150个DNA片段。由于这一数目小于实施例1中所示的数目，从而可按类似于实施例1的方式分析DNA片段。

3′－末端的两碱基是否与DNA探针的靶DNA（DNA片段）完全互补对互补链的合成影响甚大。利用这一现象将所形成的DNA片段分为16组。DNA片段以下式表示：

$$5' - C'_1 \ldots C'_i X_n \ldots X_1 C_i \ldots C_1 - 3'$$

其中Xn…X_1为每一片段本身的序列；Ci…C_1和C′$_1$…C′i代表识别序列或其一部分，且为已知。18聚寡聚物A′$_1$…A′$_{18}$，A_{18}…A_1分别在该DNA片段的5′和3′末端与其连接。结果每一DN片段均具有了如下结构：

$$5' - A'_1 \ldots A'_{18} C'_1 \ldots C'_i X_n \ldots X_1 C_i \ldots C_1 A_{18} \ldots A_1 - 3'$$

A_{18}…A_1和A_1′…A_{18}′为具有已知序列的寡聚物。在每一DNA片段中，$C_iC_1A_{18}$…A_1的序列已知且与该已知序列相邻的3′末端Xn…X_1中的几个碱基被用于DNA片段的分组。

制备与该DNA片段杂交的DNA探针。该DNA探针是与DNA片段的A_1…$A_{18}C_1$…Ci互补且与特定序列x_1x_2完全杂交的序列。由于x_1x_2存在16种可能的组合，因此共制备16种探针。DNA探针以下式表示：

$$5' - A'_1 \ldots A'_{18} C'_1 \ldots C'_i Y_1 Y_2 - 3'$$

可在 C_i' 和 Y_1 之间插入一 DNA 类似物如肌苷或用其取代 C_i'，并也可使用由此产生的探针。用生物素（B）标记该 DNA 探针；根据需要，可对该标记的探针进行设计，以便通过其上固定有链霉抗生物素蛋白的微量滴定板或通过磁性珠等对其进行分离。在上述解释中，$C_1'\cdots,C_i$；C_1'，$\cdots,C_i'$；$x_i,\cdots,x_n$；Y_1, Y_2；$A_1,\cdots,A_{18}$；A_1'，$\cdots,A_{18}'$ 代表 A、C、G 和 T 中任何一种；且序列 $C_1\cdots C_i$ 和序列 $C_1'\cdots C_i'$；序列 $A_1\cdots A_{18}$ 和序列 $A_1'\cdots A_{18}'$；序列 x_1x_2 和 Y_1Y_2 形成互补链（如后面图 4 所示）。

下面参照图 4 进一步详细描述本发明。DNA 样品 38（具有已知序列的 18 聚寡聚物：$A'_1\cdots A'_{18}$ 和 $A_{18}\cdots A_1$ 分别连到每一 DNA 片段的 5′ 和 3′ 末端；它是通过用 HindIII 消化 DNA 而获得的，并将由此获得的 DNA 片段分组）被分成 16 份。将每一份单独加到容器中。向每一份中加入一生物素化的 DNA 探针 39（其中 B 代表生物素）：

$$5' - B - A'_1\ldots A'_{18}C'_1\ldots C'_iY_1Y_2 - 3'$$

进一步向其中加入 DNA 聚合酶和用于链合成的底物（dNTP；脱氧核苷酸三磷酸），完成互补链延伸反应。

可作为 DNA 聚合酶使用的有热稳定性 Taq 或热稳定性测序酶。在热循环条件下进行反应，合成与 DNA 探针末端序列完全互补的 DNA 片段互补链。互补链合成后，加入其表面具有链霉抗生物素蛋白（Av）的磁性珠，以将生物素标记的合成 DNA 链 40 捕获至该珠上。也可用其上固定有链霉抗生物素蛋白的塑料珠或滤膜代替磁性珠。除去所捕获的除 DNA 链外的未反应物质。如上所述，对于具有下列特定末端序列的每一 DNA 片段均制备了 16 个组（43）：

$$X_2X_1C_i\ldots C_1A_{18}\ldots A_1$$

在本实例中，是利用每一 DNA 片段的互补链进行分组。通过升高含有

DNA 片段的溶液的温度，杂交的完整 DNA 片段被释放并被除去。如图 4，回收电段的互补链（DNA 链）46 组在 5′末端具有共同的探针序列（在互补链合成后 DNA 探针序列位于合成链的 5′末端），并根据与探针序列相邻的两碱基的不同被分成 16 组。每一组平均含有 150 个 DNA 链。

根据 3′末端序列的差异进一步对每组 DNA 片段进行鉴定和分类（3′末端序列对应于在互补链合成中作为模板的完整 DNA 片段的 5′末端序列）。捕获在珠上的 DNA 链 46 含有通过连接反应而连接上的寡聚物序列 $A_1 \cdots A_{18}$，DNA 探针可与其杂交。一组 16 个荧光团标记的探针 44 具有与前面所述相同的序列：

$$5' - * A'_1 \ldots A'_{18} C'_1 \ldots C'_i Y_3 Y_4 - 3'$$

其中*为荧光团标记，Y_3 和 Y_4 代表 A、C、G 和 T 中的任何一种。将捕获在珠上的 DNA 片段分成 16 组。

如上述实施方案所述，向每一组中加入一 DNA 探针以完成互补链的合成（延伸反应）。共在 16 × 16 = 256 组中进行反应，并测定延伸产物 45 的电泳图谱。可改变每一探针的所述荧光团标记，用具有不同颜色的荧光团标记的探针可同时反应，并可用多色检测型 DNA 分析仪检测延伸产物 45。当使用具有四种不同颜色的荧光团时，对每一 DNA 组足以完成四次反应，即总共为 4 × 16 = 64 次。另外，64 条泳道对于电泳是足够的。在这一实施方案中，与两末端相连的寡聚物被看作是相同的，但也可在两末端使用不同的限制酶，从而可在两末端连上不同的序列。

通过对其基因组织有不同的大肠杆菌 JM109 和 C600 的电泳图谱进行比较，发现它们绝大部分相同，仅部分存在差别，就象图 3A 和 3B 所示的 pUC19 和 pUC118 的例子一样。当未分组进行电泳时则不能发现该差异。

如上所述，当一大 DNA 被限制酶消化后，根据两末端附近的序列对所得的 DNA 片段进行分组并进行电泳以获得长度分离图形，由此可以获得每一基因组所固有的电泳图谱。该图形能有效地鉴定基因组之间的

差异。尤其是，根据末端碱基种类所进行的分组也可有效地用于构建数据库。在仅由二维电泳构成的图形数据中，通过类似信息提供的 DNA长度是不确定的，因而难以构建数据库。但是，通过将按照末端碱基种类进行的分组与数字信息结合，即使通过类似信息进行的长度分离不是很精确，也可精确地区分 DNA 。

在上面的解释中，通过升高含有 DNA 片段的溶液的温度可以分离和除去杂交的完整 DNA 片段。也可以在分级分离后再使用与通过互补链合成延长的 DNA 探针杂交的完整 DNA 片段。在这种情况下，首先是将溶液加热至 80 － 85 ℃，以释放并除去与未通过互补链合成延长的探针杂交的 DNA 片段，然后将溶液加热至 95 － 100 ℃，以释放和回收与通过互补链合成延长的探针杂交的 DNA 片段。

实施例 3

本实施例的目的在于获得与显示基因表达分析所需 mRNA 有关的信息。据说人类的基因数为 50,000 － 100,000 种。因此，相应地存在 50,000 － 100,000 种 mRNA（信使 RNA ）。在这些 mRNA 中，推测约有 3,000 种 mRNA 是正常工作的。因此，了解这些 mRNA 如何在各种组织中发挥作用是十分重要的。为了测定 mRNA 的功能，现有技术包括收集 mRNA，克隆 mRNA，分析每一克隆中 mRNA 的序列并检测显示 mRNA 的频率。但是，该方法需要大量的劳动。因此有必要开发用于这一分析的简便操作方法。与 mRNA 互补的 cDNA 的分析已取得进展，由此也可能制备出能够与每一 cDNA 特异性杂交的 DNA 探针并利用该 DNA 探针进行检测。但是，实际上可应用的探针仅限于约数十种到 100 种探针。当要检测不能被探针检测的 cDNA 时，利用该有限的探针则不实用。因此有必要开发一套可检测任何 cDNA 的系统。鉴于这种情况，在本实施方案中是利用本发明来显示 mRNA 。

从组织中制备出 mRNA 文库（ cDNA 文库）（ J.DNA Sequencing and Mapping,2,137-144 （ 1991 ））。按照 Nature, 357,519-520 （ 1992 ）所述技术，利用生物素化的 polyT 制备 cDNA ，并制备 mRNA － cDNA 杂交体和双链 DNA 。用限制酶 Sau 3AI 消化双链 DNA ，并将具有已知序列的寡聚物连接到消化位点，也可使用 HhaI （ GCG ↓c ）或 NlaIII

(CATG↓)作为限制酶。

通过这些步骤制成了在其5′末端带有生物素化的polyT序列且在其3′末端带有识别序列和已知寡聚物的DNA片段。向该DNA片段中加入其上固定有链霉抗生物素蛋白的磁性珠，形成生物素链霉抗生物素蛋白复合物。该复合物捕获上述DNA片段及其互补链。将其上固定了DNA片段的珠粒加到容器中，并向其中加入缓冲溶液。加热所产生的混合物以释放互补DNA片段，并使用该互补DNA片段作为样品。随后按实施2方法处理该样品，使DNA片段根据与其polyA相邻的3′末端序列进行分级分离。在本例中，与实施例2所不同的是共有12种与polyT相邻的选择序列。分级分离生物素化的DNA探针后，利用16种荧光团标记的探针合成互补链。通过凝胶电泳测定所产生的DNA的长度，获得指纹图谱。

实施例4

为了实施实施例1－3所述方法，所使用的DNA分析仪包括用于16种或256种DNA探针同时进行反应以获得互补链延伸产物的反应器，一个用于将每一探针自动加到每一反应器的自动移液管，以及一个用于电泳的装置。优选在65℃以上的温度进行互补链延伸反应，因为无论该反应进行与否它都明显受末端碱基序列的配对或误配的影响。通过将DNA探针末端碱基的每一碱基同时进行延长产物的电泳，或在通过对DNA探针的末端碱基的每一碱基同时进行长度(对照)标记物的电泳，对互补链延伸反应的产物的相对长度进行比较也是必要的。为此，尤其优选这样一种测定片段长度的装置，它将对照标记物与互补链延伸反应的每种产物一起电泳，并且利用了通过电泳分级分离的互补链延伸反应产物片段与标记物的相对位置关系。在这种情况下，通过鉴别DNA片段的长度易于完成操作，其中利用的是具有不同长度的多个标记物，根据其排列方式即都确定DNA片段属于哪一组。如果最高达1kb的DNA片段被每隔5个碱基上出现的标记物分级分离成200组，则该DNA片段甚至将分级分离成256×200－50,000组。这一组数足以用于mRNA的诊断等或用于指纹图谱分析。

在10－500碱基的长度范围内电泳中的测量精确度是很高的。当对

每2个碱基进行DNA片段分级分离时所获得的分级分离数据将更详细。在该测定中，需要用其发射波长与标记靶DNA片段的荧光团的发射波长完全不同的荧光团来标记该标记物，以便该标记物与靶DNA片段易于鉴别，荧光团在较长波长方向上具有一平的发射区。因此，重要的是通过为标记靶DNA片段的标记物选择合适种类的荧光团使得标记物不干扰所需的DNA片段测量并在较长的波长区产生荧光信号。根据序列长度的不同DNA片段可具有不同的结构，而且即使它们具有相同的碱基长度，它们在凝胶电泳中的流动性也可能不同。但是，即使在这种情况下，只要知道DNA片段在哪一级分出现，就可以利用标记物作为检测标准对DNA片段进行分级分离、也就是说，即使DNA片段具有相同长度，它们也会由于序列上的差异而以不同的级分被分级分离，但分级分离的良好再现性消除了分析上的任何困难。

序列表

SEQ ID No:1

长度： 24 个碱基对

类型： 核酸

链型： 单链

拓朴结构： 线性

分子类型： 合成 DNA， 荧光团标记

序列描述：

X， Y： 任意选自 A、 C、 G 或 T。

图.1

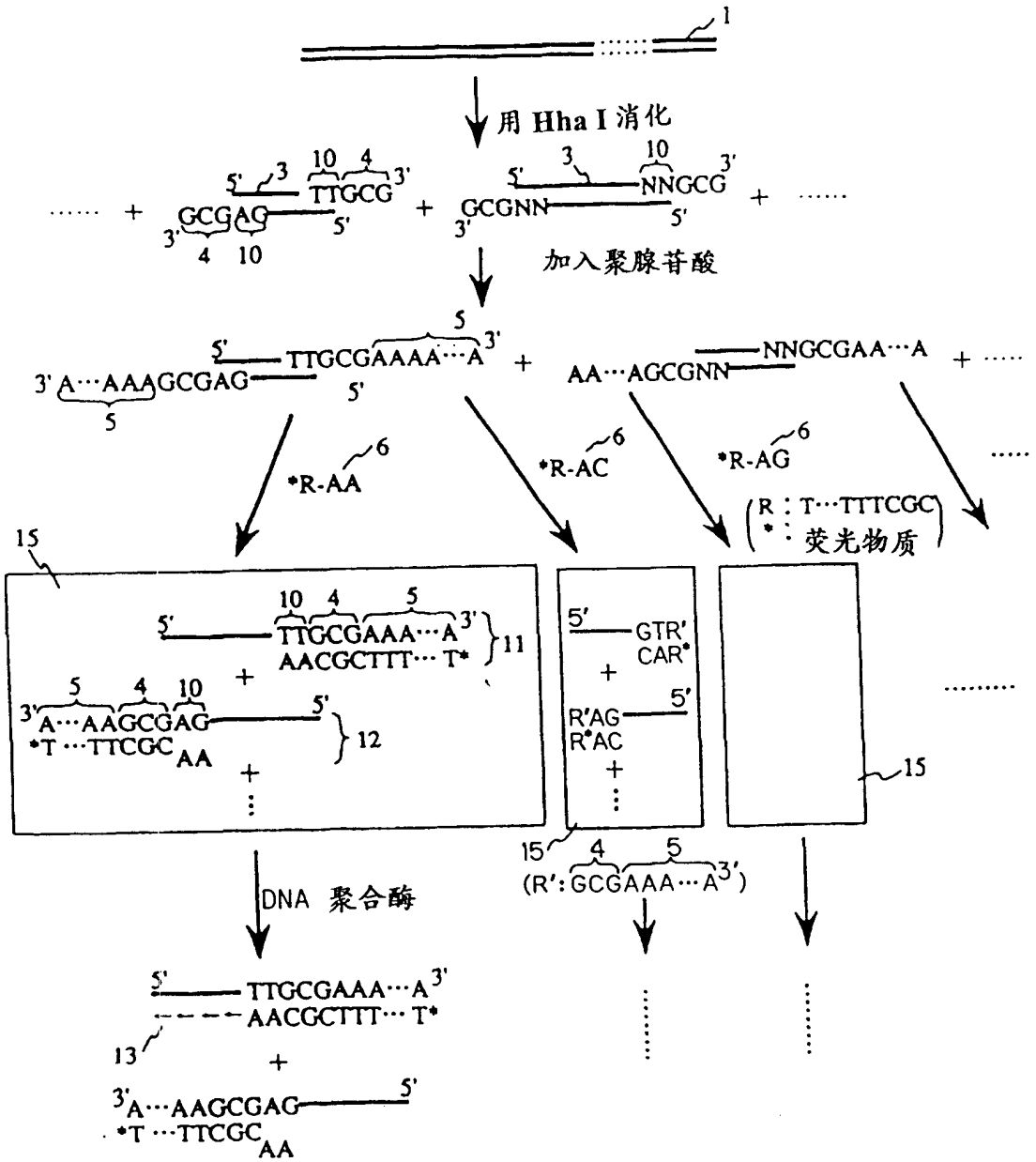


图.2A

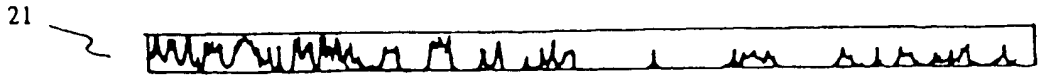

21

图.2B

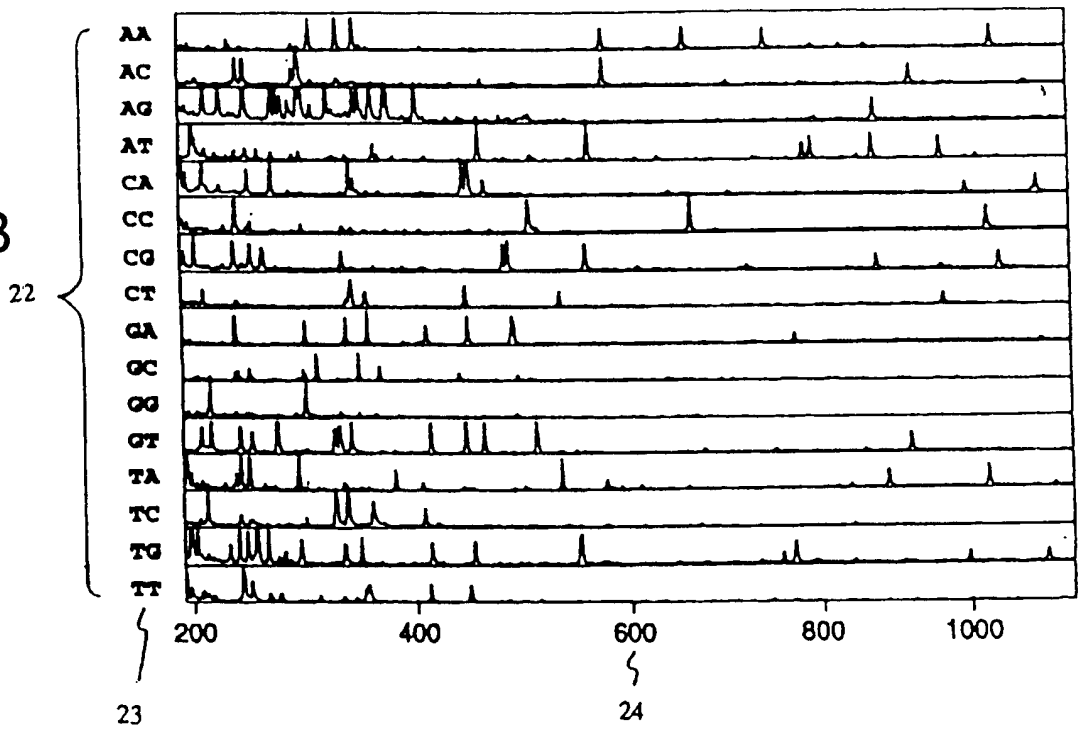

AA
AC
AG
AT
CA
CC
CG
CT
GA
GC
GG
GT
TA
TC
TG
TT
22
200
400
600
800
1000
23
24

图.3A

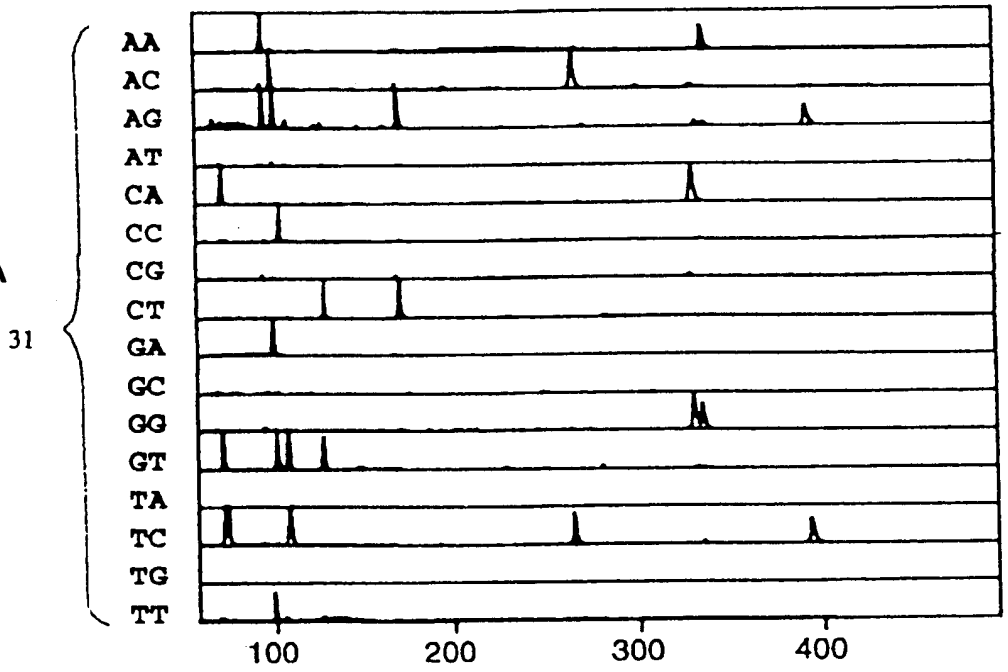

AA
AC
AG
AT
CA
CC
CG
CT
GA
GC
GG
GT
TA
TC
TG
TT
31
100
200
300
400

图.3B

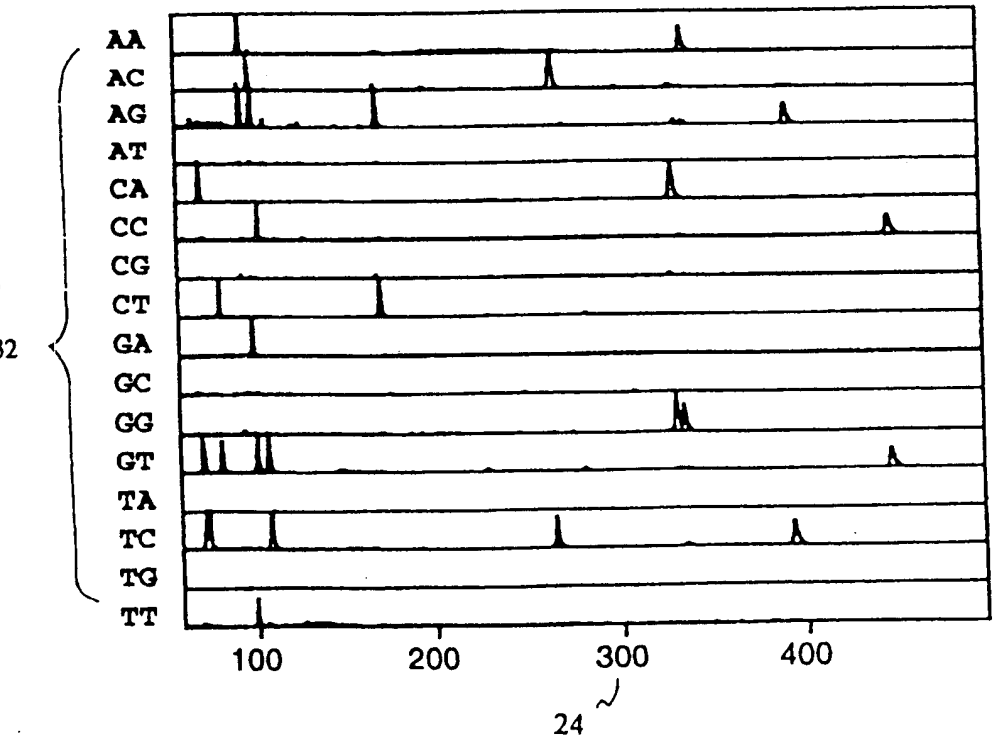
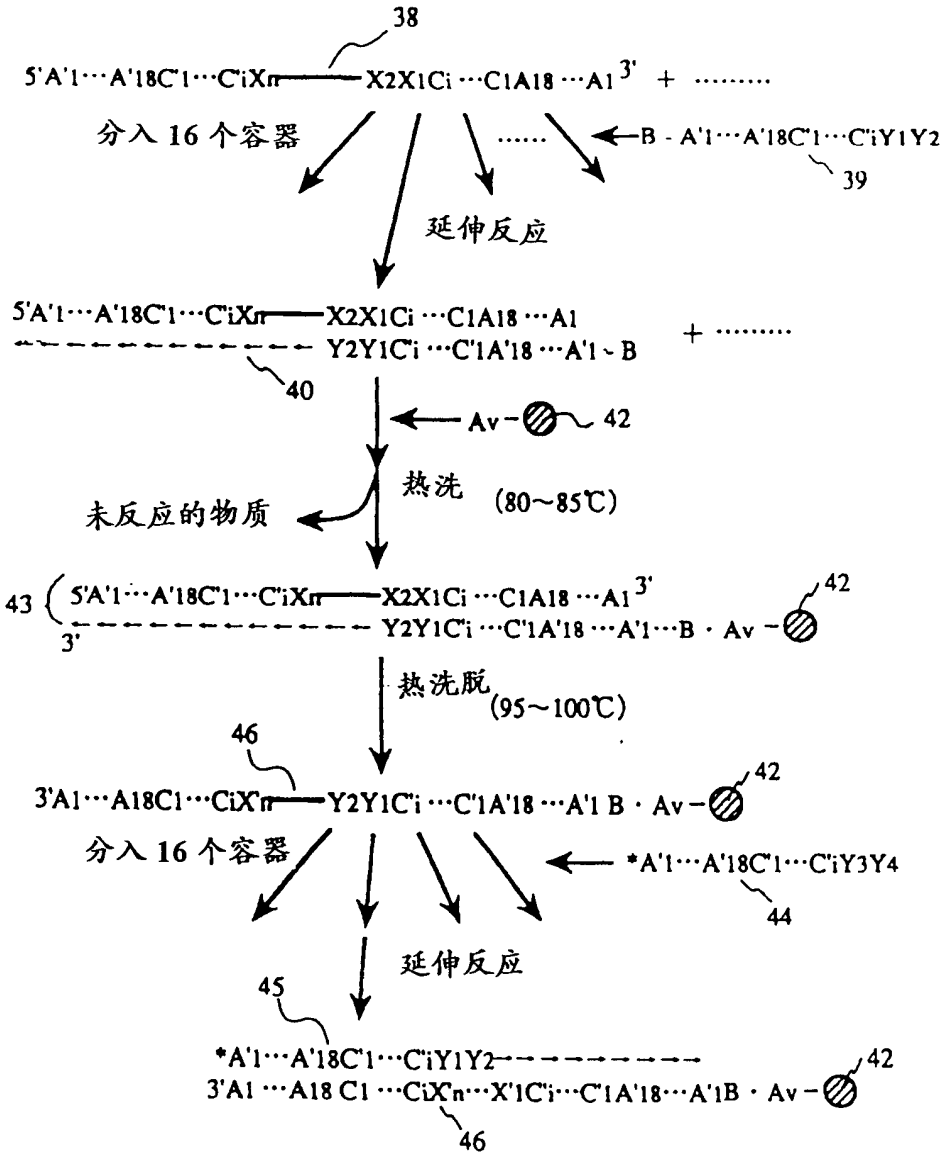

AA
AC
AG
AT
CA
CC
CG
CT
GA
GC
GG
GT
TA
TC
TG
TT
32
100
200
300
400
24

图.4


38
5'A'1…A'18C'1…C'iXn——X2X1Ci…C1A18…A1 3' + ……
分入 16 个容器
……
B - A'1…A'18C'1…C'iY1Y2
39
延伸反应
5'A'1…A'18C'1…C'iXn——X2X1Ci…C1A18…A1
Y2Y1C'i…C'1A'18…A'1 - B
+ ……
40
Av - 42
热洗 (80~85℃)
未反应的物质
43
5'A'1…A'18C'1…C'iXn——X2X1Ci…C1A18…A1 3'
3'
Y2Y1C'i…C'1A'18…A'1…B · Av -
42
热洗脱 (95~100℃)
46
3'A1…A18C1…CiX'n——Y2Y1C'i…C'1A'18…A'1 B · Av -
42
分入 16 个容器
*A'1…A'18C'1…C'iY3Y4
44
延伸反应
45
*A'1…A'18C'1…C'iY1Y2
3'A1…A18C1…CiX'n…X'1C'i…C'1A'18…A'1B · Av -
42
46