



(12) 发明专利

(10) 授权公告号 CN 101821402 B

(45) 授权公告日 2015. 03. 04

(21) 申请号 200880101253. 0

(22) 申请日 2008. 07. 31

(30) 优先权数据

60/953, 171 2007. 07. 31 US

(85) PCT国际申请进入国家阶段日

2010. 01. 29

(86) PCT国际申请的申请数据

PCT/US2008/071771 2008. 07. 31

(87) PCT国际申请的公布数据

W02009/018449 EN 2009. 02. 05

(73) 专利权人 BP 法人北美有限公司

地址 美国印第安纳州

(72) 发明人 X·谭

(74) 专利代理机构 北京纪凯知识产权代理有限公司

公司 11245

代理人 赵蓉民 陈建芳

(51) Int. Cl.

C12N 15/00(2006. 01)

(56) 对比文件

US 7202086 B2, 2007. 04. 10,

US 20020086322 A1, 2002. 07. 04,

US 6773900 B2, 2004. 08. 10,

US 20060223066 A1, 2006. 10. 05,

US 2007026438 A1, 2007. 02. 01,

John Neidhardt 等. Different Amino Acid Substitutions at the Same Position in Rhodopsin Lead to Distinct Phenotypes. 《Invest Ophthalmol Vis Sci. 》. 2006, 第 47 卷 (第 4 期),

审查员 韦炜

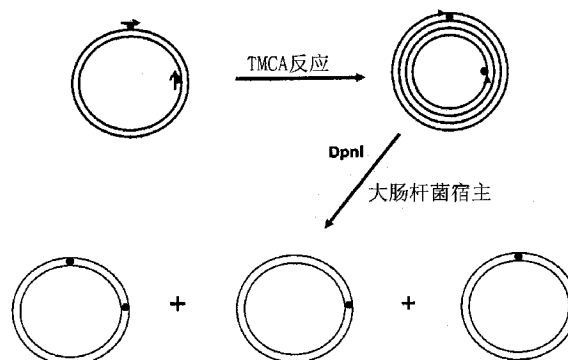
权利要求书3页 说明书19页 附图9页

(54) 发明名称

修剪式多位点组合装配

(57) 摘要

本发明提供通过修剪式多位点组合装配产生多个修饰的多核苷酸的新方法, 所述修饰的多核苷酸在多个位点具有各种突变的不同组合, 该方法包括在单一反应混合物中将至少两种或至少三种引物添加至双链模板多核苷酸, 其中所述引物不重叠, 并且其中所述引物的每一种包括至少一个不同于其他引物的突变, 其中至少一种引物是可退火至所述模板的负链的正向引物以及至少一种引物是可退火至所述模板的正链的反向引物, 并使所述反应混合物进行聚合酶延伸反应以从所述至少三种引物产生多个延伸的修饰的多核苷酸。该方法可在将延伸的修饰的多核苷酸转化至细胞之前在不使用连接步骤的情况下进行。所述多个延伸的修饰的多核苷酸可用酶处理, 以在转化至细胞之前破坏模板多核苷酸。



1. 一种产生多个突变的多核苷酸的方法,其包括:

(a) 获得模板多核苷酸的序列信息,并鉴定在所述模板多核苷酸上的三个或更多个感兴趣的突变;

(b) 在单一反应混合物中将至少三种引物添加至双链模板多核苷酸,其中所述至少三种引物不重叠,并且其中所述至少三种引物的每一种包括至少一个不同于其他引物的突变,其中至少一种引物是可与所述模板的负链退火的正向引物以及至少一种引物是可与所述模板的正链退火的反向引物,和

(c) 使所述反应混合物进行聚合酶延伸反应以从所述至少三种引物产生多个延伸的突变的多核苷酸,

其中在步骤(c)之前进一步包括:

根据它们在所述模板多核苷酸上的位置,将所述引物编成多个组,其中覆盖所述模板上的相同选择区域的引物在一个组中,

将每组中分入的引物归一化至相等的浓度,

将一组中的所述正向引物富集成正向组并将每一正向引物组之间的浓度归一化至相等,

将一组中的所述反向引物富集成反向组并将每一反向引物组之间的浓度归一化至相等,和

将相等量的所富集的正向和反向引物添加至所述反应。

2. 权利要求1的方法,其进一步包括用未经连接酶处理的所述多个延伸的产物转化细胞。

3. 权利要求2的方法,其进一步包括从所述细胞回收所述多个延伸的突变的多核苷酸。

4. 权利要求3的方法,其进一步包括分析所述多个延伸的突变的多核苷酸。

5. 权利要求4的方法,其中分析包括表达所述多个延伸的突变的多核苷酸中的至少一个并分析由其表达的多肽。

6. 权利要求5的方法,其进一步包括选择包括感兴趣的突变的所述多个延伸的突变的多核苷酸。

7. 权利要求1的方法,其进一步包括分析由所述聚合酶延伸产生的所述多个延伸的突变的多核苷酸。

8. 权利要求1的方法,其进一步包括用限制酶处理所述多个延伸的突变的多核苷酸,从而破坏所述模板多核苷酸,将经处理的延伸的突变的多核苷酸转化入原核细胞,从所述原核细胞回收所述多个延伸的突变的多核苷酸,并选择包括感兴趣的突变的所述多个延伸的突变的多核苷酸。

9. 权利要求8的方法,其中所述细胞是大肠杆菌细胞。

10. 权利要求9的方法,其中所述酶是限制酶。

11. 权利要求10的方法,其中所述限制酶是 DpnI 限制酶并且所述细胞是大肠杆菌细胞。

12. 权利要求1的方法,其中添加至少四种引物。

13. 权利要求1的方法,其中添加至少五种引物。

14. 权利要求 1 的方法,其中添加至少六种引物。
15. 权利要求 1 的方法,其中添加至少八种引物。
16. 权利要求 1 的方法,其中添加至少十二种引物。
17. 权利要求 1 的方法,其中每种引物包括单个点突变。
18. 权利要求 1 的方法,其中至少两种正向引物在所述模板多核苷酸的相同位置上包括不同的变化。
19. 权利要求 1 的方法,其中至少两种反向引物在所述模板多核苷酸的相同位置上包括不同的变化。
20. 权利要求 1 的方法,其中至少一种引物在所述模板多核苷酸的不同位置上包括至少两个变化。
21. 权利要求 1 的方法,其中至少一种引物在所述模板多核苷酸的不同位置上包括至少两个变化并且至少两种正向或两种反向引物在所述模板多核苷酸的相同位置上包括不同变化。
22. 权利要求 1 的方法,其中所述至少一个突变选自一个或多个核苷酸的变化、插入和缺失。
23. 权利要求 1 的方法,其中所述模板多核苷酸是环状双链 DNA。
24. 权利要求 1 的方法,其中至少一种引物是一组各包括简并位置的简并引物,其中所述感兴趣的突变是一系列在所述简并位置上的不同核苷酸。
25. 权利要求 1 的方法,其中至少一种引物是一组简并引物,其包括对应于所述模板多核苷酸的至少一个密码子的至少一个简并密码子以及与所述模板多核苷酸的密码子相邻的序列同源的至少一个相邻序列。
26. 权利要求 25 的方法,其中所述简并密码子是编码天然发生的氨基酸的 N,N,N,其中所述 N,N,N 表示三联体,其中“N”可以是 A、C、T 或 G。
27. 权利要求 25 的方法,其中所述简并密码子可编码小于 20 种天然发生的氨基酸。
28. 权利要求 1 的方法,其中所述正向引物被分入正向组,而所述反向引物被分入反向组,并且所述正向组中的引物和所述反向组中的引物各自独立地在相应组中被归一化为相等浓度,而不考虑在所述模板多核苷酸上的位置,并且其中在归一化后,将相同量的所述正向和反向引物添加至所述反应中。
29. 权利要求 1 的方法,其进一步包括进行两步的步骤 (b) 至 (c),并使用第一轮产生的多核苷酸作为第二轮的模板多核苷酸。
30. 产生包括感兴趣的突变的多个修饰的多核苷酸的方法,其包括:
 - (a) 在单一反应混合物中将至少两种引物添加至双链模板多核苷酸,其中所述至少两种引物不重叠,并且其中所述至少两种引物的每一种包括至少一个不同于其他引物的突变,其中至少一种引物是可与所述模板的负链退火的正向引物以及至少一种引物是可与所述模板的正链退火的反向引物,
 - (b) 使所述反应混合物进行聚合酶延伸反应以从所述至少两种引物产生多个延伸的修饰的多核苷酸,
 - (c) 用限制酶处理所述多个延伸的修饰的多核苷酸,从而破坏所述模板多核苷酸,
 - (d) 将所述处理的延伸的修饰的多核苷酸转化入原核细胞,所述多核苷酸未用连接酶

处理，

(e) 从所述原核细胞回收所述多个延伸的修饰的多核苷酸，以及

(f) 选择包括感兴趣的突变的所述多个延伸的修饰的多核苷酸；

其中在步骤 (b) 之前进一步包括：

根据它们在所述模板多核苷酸上的位置，将所述引物编成多个组，其中覆盖所述模板上的相同选择区域的引物在一个组中，

将每组中分入的引物归一化至相等的浓度，

将一组中的所述正向引物富集成正向组并将每一正向引物组之间的浓度归一化至相等，

将一组中的所述反向引物富集成反向组并将每一反向引物组之间的浓度归一化至相等，和

将相等量的所富集的正向和反向引物添加至所述反应。

修剪式多位点组合装配

[0001] 相关申请的交叉引用

[0002] 本申请要求 2007 年 7 月 31 日提交的美国临时申请序列号 60/953,171 的优先权，其内容通过引用并入本文。

技术领域

[0003] 本发明总体上涉及修剪式多位点组合装配 (tailored multi-site combinatorial assembly, “TMCA”) 的方法, 其作为产生多个子代多核苷酸和对基因进行特异变化以及在多个位点重装配突变或变化的方法。在短寡核苷酸上设计并合成突变或变化。寡核苷酸与包括野生型基因的模板 DNA 退火。DNA 聚合酶用于扩增全 DNA。从宿主回收所产生的扩增的 DNA。该方法的优势在于速度、技术简单以及控制装配的能力。

背景技术

[0004] 已公布的进行基因变化的方法采用, 例如, 易错 PCR、Invitrogen 的 Gene Tailor site-directed Mutagenesis Kit™、Stratagene 的 QuickChange Mutagenesis Kit™、重叠 PCR 和基于 PCR 的连接 / 重组。对已知方法的研究表明这些方法往往面对一个主要困难: 在单个位点 / 临近区域产生突变和 / 或修饰和 / 或这些方法在多个区域进行修饰是费力的。

[0005] 美国专利 7,202,086 (“’086 专利”) 要求保护诱变方法, 其使用至少 5 个重叠或不重叠的寡核苷酸以及 dsDNA (质粒) 以产生突变基因的文库, 其中每个突变平均存在于文库中 1/5 以下的基因中。’086 专利描述了所公开的发明不同于现有技术, 因为’086 专利要求控制突变的频率以避免在一个 DNA 分子中的“过量突变”(第 5 栏, 28-45 行)。所期望的是获得各含有一个突变的突变体。为了达到该目的, 每个突变体寡核苷酸的量与模板量之间的比例必须在 0.01-100 之间 (第 5 栏, 28-45 行)。该特征区别于现有技术, 其中同时使用几个寡核苷酸使得每条引物并入的水平大于 75% (第 5 栏, 28-45 行)。’086 专利要求控制突变的频率以避免在一个 DNA 分子中的“过量突变”并产生各含有一个突变的突变体。

[0006] 美国专利 7,132,265 (“’265 专利”) 和美国专利公开 2003/0064516 要求保护将突变引入单链 DNA (“ssDNA”) 分子的方法, 其包括退火引物、合成 DNA 链和消化 DNA 分子。TCMA 方法使用双链 DNA (“dsDNA”) 作为模板。’265 专利在使用 ssDNA 和 dsDNA 作为起始底物用于诱变方法之间进行了明确区分 (参见, 例如, 第 6 栏, 45-55 行)。

[0007] 美国已公布的申请 2003/0194807 涉及文库, 其中蛋白质的突变体包括在确定的区域的一个或多个位置上的单个预定氨基酸, 其中所述确定的区域是至少三个氨基酸。只允许在位置上的单个变化, 即, 排除具体氨基酸位置的简并变化。

[0008] 美国已公布的申请 2006/0051748、2006/0134624、2004/0248131 和 2002/0083488 以及美国专利 6,673,610、6,335,160 和 5,354,670 中的每一个都要求连接合成的 DNA 以产生具有突变的子代环状 DNA。

[0009] 另外, 美国已公布的申请 2006/0051748 要求使用瓣状核酸内切酶并将所有引物

与同一 DNA 链退火。美国已公布的申请 2006/0134624 要求依次（即，不在一个反应中）使用两种引物。在美国已公布的申请 2004/0248131 中，引物与两条链退火，其中所述引物必须包括 2-4 个互补碱基对。美国专利 6,673,610 和美国已公布的申请 2002/0083488 要求使用通过亲代 DNA 链消化产生的片段作为大引物（megaprimer）以获得用于转化的环状 DNA。美国专利 6,335,160 涉及来自重叠片段和产生重组文库的基因装配。最后，美国专利 5,354,670 要求两个转化步骤和采用限制性（restriction）进行的中间处理。

[0010] 美国专利 7,176,004、6,713,285、6,391,548 和 5,932,419 以及美国已公布的申请 20040253729 和 20030032037 中的每一个都要求两种引物以与两条不同的链退火，用于在相反的方向起始扩增（即，正向和反向引物）并要求具有互补区。美国专利 7,078,389 和 5,935,830 要求引物以包括诱变剂（例如，补骨脂素），其与模板相互作用以便形成三链分子。

[0011] 美国已公布的申请 2006/0228786 要求在两个不同的反应中使用两个不同的引物进行两条链的聚合，然后将合成的 ssDNA 分子退火。美国已公布的申请 2003/0077613 涉及基因装配和产生文库的方法，其中装配的基因（ssDNA）与支架 DNA 退火以填充间隙并产生 dsDNA，其被亚克隆至载体。

[0012] 美国已公布的申请 2004/0002057 描述了在样品中检测不包括诱变的配体的方法。美国已公布的申请 2004/0002057 描述了通过使用诱变剂在培养的细胞中建立突变体大肠杆菌（*E. coli*）菌株的方法。美国已公布的申请 2006/0199222 描述了定向进化的一般方法，其中将突变的 DNA 转化至具体的杆菌（*Bacillus*）菌株。

[0013] 仍然存在对更好的且更有效的方法的需要，所述方法有效且快速地产生特异的基因变体和组合基因文库。

发明内容

[0014] 除非特别定义，所有用于本文的技术和科学术语具有本领域普通技术人员例如化学、生物化学、细胞生物学、分子生物学或医学领域结合上下文来看所通常理解的含义。

[0015] 因此，本发明的一个目的是提供通过修剪式多位点组合装配产生多个修饰的多核苷酸的方法，所述修饰的多核苷酸在多个位点具有各种突变的不同组合。本发明允许对基因进行特异变化并在所述基因的多个位点重装配突变或变化。本发明的这些和其他目的——其与下列优选实施方式的详细描述（单独或其组合）结合而变得更加明显——通过发现如下方法而得以满足，所述方法包括：

[0016] (a) 在单一反应混合物中将至少三种引物添加至双链模板多核苷酸，其中所述至少三种引物不重叠，并且其中所述至少三种引物的每一种包括至少一个不同于其他引物的突变，其中至少一种引物是可与所述模板的负链退火的正向引物以及至少一种引物是可与所述模板的正链退火的反向引物，和

[0017] (b) 使所述反应混合物进行聚合酶延伸反应以从所述至少三种引物产生多个延伸的修饰的多核苷酸。

[0018] 在另一个实施方式中，产生多个包括感兴趣突变的修饰的多核苷酸的方法包括：

[0019] (a) 在单一反应混合物中将至少两种引物添加至双链模板多核苷酸，其中所述至少两种引物不重叠，并且其中所述至少两种引物的每一种包括至少一个不同于其他引物的

突变,其中至少一种引物是可与所述模板的负链退火的正向引物以及至少一种引物是可与所述模板的正链退火的反向引物,

[0020] (b) 使所述反应混合物进行聚合酶延伸反应以从所述至少两种引物产生多个延伸的修饰的多核苷酸,

[0021] (c) 用酶处理所述多个延伸的修饰的多核苷酸,从而破坏所述模板多核苷酸,

[0022] (d) 将所述处理的延伸的修饰的多核苷酸转化入细胞,所述多核苷酸未用连接酶处理,

[0023] (e) 从所述细胞回收所述多个延伸的修饰的多核苷酸,和

[0024] (f) 选择包括感兴趣的突变的所述多个延伸的修饰的多核苷酸。

[0025] 本发明的一个或多个实施方式的细节在如下所附描述中阐述。虽然类似或等同于本文所述的任何方法和材料可用于本发明的实践或测试,但是现描述优选的方法和材料。本发明的其他特征、目的和优势通过说明书和权利要求将变得明显。在说明书和所附权利要求中,单数形式包括复数指代,除非上下文明确另外指出。除非上下文明确地另外说明,本文使用或考虑的技术是本领域普通技术人员熟知的标准方法。实施方式的实例仅用于说明目的。

附图说明

[0026] 本发明的更完整的理解和其许多附随优势将易于获得,因为当与附图结合考虑时,参考下列详述,它们得到更好的理解,其中:

[0027] 图 1. GSSMSM 的图示。

[0028] 图 2. 进化 -GSSMSM 方法流程的图示。

[0029] 图 3. 修剪式多位点组合装配的图示。

[0030] 图 4A-D. TMCA 反应中的引物组合。

[0031] 图 5. 六个突变装配中引物退火的图谱。

[0032] 图 6. 具有六个突变位点的可能组合的分布。标准:在理想状态下计算的变体分布;1A:采用大肠杆菌菌株 XL1-Blue 的反应条件 1;7A:采用大肠杆菌菌株 XL1-Blue 的反应条件 2;13A:采用大肠杆菌菌株 XL1-Blue 的反应条件 3;总计:来自 1A、7A 和 13A 组合的数据;0x:无突变;1x:单个突变;2x:两个突变;3x:三个突变;4x:四个突变;5x:五个突变;6x:六个突变。

[0033] 图 7. 六突变装配中统计计算与实验数据。

[0034] 图 8. 四突变装配中引物退火的图谱。

[0035] 图 9. 具有四个突变位点的可能组合的分布。标准:在理想状态下计算的变体分布;2A:采用大肠杆菌菌株 XL1-Blue 的反应条件 1;8A:采用大肠杆菌菌株 XL1-Blue 的反应条件 2;14A:采用大肠杆菌菌株 XL1-Blue 的反应条件 3;总计:来自 2A、8A 和 14A 组合的数据;0x:无突变;1x:单个突变;2x:两个突变;3x:三个突变;4x:四个突变。

[0036] 图 10. 具有四个位点的可能组合的分布。标准:在理想状态下计算的变体分布;2B:采用大肠杆菌菌株 Stb12 的反应条件 1;8B:采用大肠杆菌菌株 Stb12 的反应条件 2;14B:采用大肠杆菌菌株 Stb12 的反应条件 3;总计:来自 2B、8B 和 14B 组合的数据;0x:无突变;1x:单个突变;2x:两个突变;3x:三个突变;4x:四个突变。

[0037] 图 11. 三突变装配中引物退火的图谱。

[0038] 图 12. 具有三个突变位点的可能组合的分布。标准：在理想状态下计算的变体分布；15A：采用大肠杆菌菌株 XL1-Blue 的反应条件 3；9B：采用大肠杆菌菌株 Stb12 的反应条件 2；15B：采用大肠杆菌菌株 Stb12 的反应条件 3；总计：来自 15A、9B 和 15B 组合的数据；0x：无突变；1x：单个突变；2x：两个突变；3x：三个突变。

[0039] 图 13. 具有 5 个突变位点和 13 个突变体的引物退火图谱。

[0040] 图 14A、B、C. 五个突变位点和 13 个引物装配的独特变体组合。在 TMCA 第 I 轮中的变体破坏 (A)、第 II 轮后的变体破坏 (B)，引物退火图谱 (C)。

[0041] 优选实施方式详述

[0042] TMCA 方法可高效快速地产生包括多个变化的特异基因变体或产生组合基因文库；要求最低的财力和人力；并可以被修改来根据“需要”产生偏好的组合文库。TMCA 方法可不使用连接步骤而进行，因此简化产生多个突变的方法。对具体文库的“需要”随实验而变化。潜在突变位点——“需要”——例如，可以是 1) 合理设计的氨基酸变化或 2) 经验性确定来产生所期望的对酶的作用的单独氨基酸改变（由 GSSMSM 和筛选确定）。产生的每个文库具有具体数目的潜在突变位点。可优选产生文库，其偏好于在潜在突变位点具有更多或更少突变的子代。同样，可优选产生文库，其中更倾向于或更不倾向于具体突变或突变位点。

[0043] 所有类似或等同于本文所述的方法和材料可用于本发明的实践或测试，其中合适的方法和材料在本文进行了描述。本文提及的所有出版物、专利申请、专利和其他参考文献通过引用全部并入本文。另外，材料、方法和实例仅是阐述性的并且不意图是限制性的，除非另外说明。

[0044] 在该申请中，本发明人设计了在图 3 中一般性示出的修剪式多位点组合装配方法。作为比较，图 1 和图 2 图解了进化-基因位点饱和诱变 (Gene Site Saturated Mutagenesis, “GSSM”), 其中对于不同氨基酸，每个突变位置可含有两个或更多个突变。进化-GSSMSM 可用于将核苷酸变化引入具体基因并每次对于一个残基或更多个残基将开放阅读框的每个密码子突变为所有其他氨基酸。因此，产生 GSSMSM 文库，其中单个克隆包括具有一个变化的 DNA，而由 TMCA 方法产生的文库中的子代多肽可包括多个突变，优选两个或更多个、更优选三个或更多个、更优选四个或更多个、五个或更多个、六个或更多个、八个或更多个、十个或更多个以及更优选十二个或更多个突变。

[0045] 使用 GSSMSM 技术，每次改变一个残基以覆盖所有 20 种氨基酸。筛选文库并鉴定高表达突变型。设计 TMCA 反应，用于在一个分子的多个位点产生突变。TMCA 反应可用于组合从 GSSMSM 文库鉴定的高表达突变型。在 TMCA 反应条件下，预期形成多种 PCR 产物。该 PCR 产物不被预期转化入细胞并得到扩增。

[0046] 在本发明的上下文中，如本文所用，术语“氨基酸”是指含有氨基 (-NH₂) 和羧基 (-COOH) 的任何有机化合物；优选地作为自由基或可选地在缩合后作为肽键的一部分。“二十种天然发生的氨基酸”是本领域所理解的，其是指：丙氨酸 (ala 或 A)、精氨酸 (arg 或 R)、天冬酰胺 (asn 或 N)、天冬氨酸 (asp 或 D)、半胱氨酸 (cys 或 C)、谷氨酸 (glu 或 E)、谷氨酰胺 (gln 或 Q)、甘氨酸 (gly 或 G)、组氨酸 (his 或 H)、异亮氨酸 (ile 或 I)、亮氨酸 (leu 或 L)、赖氨酸 (lys 或 K)、甲硫氨酸 (met 或 M)、苯丙氨酸 (phe 或 F)、脯氨酸 (pro 或 P)、丝

氨酸 (ser 或 S)、苏氨酸 (thr 或 T)、色氨酸 (trp 或 W)、酪氨酸 (tyr 或 Y) 和缬氨酸 (val 或 V)。

[0047] 术语“扩增”(“聚合酶延伸反应”)意思是多核苷酸拷贝数增加。

[0048] 术语“对应于”用于本文意思是多核苷酸序列与参考多核苷酸序列的全部或部分同源(即,相同,但不是严格进化相关),或多肽序列与参考多肽序列相同。作为对比,术语“互补于”用于本文意思是互补序列与参考多核苷酸序列的全部或部分同源。作为说明,核苷酸序列“TATAC”对应于参考“TATAC”并互补于参考序列“GTATA”。

[0049] “引物”在本文定义为可与模板核酸退火并作为 DNA 扩增起点的核酸链。引物可完全或部分互补于模板多核苷酸的特定区域。非互补核苷酸在本文定义为错配。错配可位于引物内或在引物的一端。优选地,单个核苷酸错配;更优选两个,以及更优选地,三个或更多连续或不连续核苷酸错配位于引物内。引物具有 6 至 200 个核苷酸,优选 20 至 80 个核苷酸,以及更优选 43 至 65 个核苷酸。更优选地,引物具有 10、15、20、25、30、35、40、45、50、55、60、65、70、75、80、85、90、95、100、105、110、115、120、125、130、135、140、145、150、155、160、165、170、175、180、185 或 190 个核苷酸。如本文所定义,“正向引物”是互补于模板多核苷酸负链的引物。如本文所定义,“反向引物”是互补于模板多核苷酸正链的引物。优选地,正向和反向引物不包括重叠核苷酸序列。如本文所定义,“不包括重叠核苷酸序列”意思是正向和反向引物分别不能与模板多核苷酸负链和正链区域退火,在所述区域中正链和负链互相互补。关于与模板多核苷酸的相同链退火的引物,“不包括重叠核苷酸序列”意思是引物不包括互补于模板多核苷酸相同链相同区域的序列。

[0050] 正链与有义链相同并也可称为编码链或非模板链。这是与 mRNA 具有相同序列的链(除了它具有 T 而不是 U)。另一条链称为模板链、负链或反义链,互补于 mRNA。

[0051] “覆盖模板多核苷酸同一选择区域的引物”在本文定义为一组简并引物,各引物包括至少一个简并位置,其中感兴趣的突变是一系列在简并位置的不同核苷酸;或定义为一组简并引物,其包括至少一个简并密码子,该简并密码子对应于模板多核苷酸的至少一个密码子;或者它们的组合。例如,用于三密码子突变 Y276F/S282L、H、P、R 或 C/L284F(参见例如,图 4、15 或 16)的所有可能组合的一组引物是覆盖模板的同一区域的引物。“覆盖模板多核苷酸同一选择区域的引物”也可以是,例如,具体引物的组合。

[0052] DNA “消化”是指用仅在 DNA 的某些序列上起作用的限制酶催化切割 DNA。本文所用的各种限制酶是可购得的,并且它们的反应条件、辅因子和其他要求得以采用,如本领域普通技术人员所知的。用于分析目的,通常地 1 μ g 质粒或 DNA 片段在约 20 μ l 缓冲溶液中与约 2 个单位的酶一起使用。用于分离用于质粒构建的 DNA 片段的目的,通常在更大体积中用 20 至 250 个单位的酶消化 5 至 50 μ g 的 DNA。对具体限制酶的适当缓冲液和底物量由制造厂商规定。通常在 37°C 使用约 1 小时的温育时间,但可根据供应商说明而变化。消化后,可以将反应物在凝胶上进行电泳。

[0053] “重组”酶是指重组 DNA 技术产生的酶,即,由编码所期望酶的外源 DNA 构建体转化的细胞产生的酶。“合成”酶是由化学合成制备的酶。

[0054] 术语“限制位点”是指识别序列,其对限制酶作用的发挥是必需的,并包括催化切割位点。可理解的是,切割位点可以包括在或不包括在限制位点的一部分中,该限制位点包括低两可性序列(a lowambiguity sequence)(即含有限制位点存在频率的主要决定

子的序列)。因此,在许多情况下,相关限制位点仅含有低两可性序列,其具有内部切割位点(例如EcoRI 位点中的G/AATTC)或直接相邻切割位点(例如EcoRII 位点中的/CCWGG)。在其他情况下,相关限制酶(例如Eco57I 位点或CTGAAG(16/14))含有低两可性序列(例如Eco57I 位点中的CTGAAG 序列),其具有外部切割位点(例如Eco57I 位点中的N₁₆ 部分)。当酶(例如限制酶)被称为“切割”多核苷酸时,应理解的意思是该限制酶催化或促进多核苷酸的切割。

[0055] 限制位点中的“多义碱基要求(ambiguity base requirement)”是指不是最完全程度上明确的核苷酸碱基要求,即不是具体碱基(例如,在非限制示例中,具体碱基选自A、C、G 和T),而可以是至少两个或更多个碱基中的任何一个。本领域以及本文用来表示碱基多义性的通常可接受的缩写包括下列:R = G 或 A;Y = C 或 T;M = A 或 C;K = G 或 T;S = G 或 C;W = A 或 T;H = A 或 C 或 T;B = G 或 T 或 C;V = G 或 C 或 A;D = G 或 A 或 T;N = A 或 C 或 G 或 T。

[0056] “参考序列”是用作序列比较基础的限定序列;参考序列可以是较大序列的亚群,例如,序列表中给定的全长cDNA 或基因序列的片段,或可包括完整的cDNA 或基因序列。通常,参考序列的长度为至少20 个核苷酸,通常地长度为至少25 个核苷酸,以及通常地长度为至少50 个核苷酸。由于两个多核苷酸可各自(1) 包括两个多核苷酸之间相似的序列(即,完整多核苷酸序列的一部分),以及(2) 可进一步包括两个多核苷酸之间不同的序列,两个(或更多个)多核苷酸之间的序列比较通常通过在“比较窗”的范围内比较两个多核苷酸的序列来进行以鉴定并比较具有序列相似性的局部区域。

[0057] “比较窗”如本文所用是指至少20 个连续核苷酸位置的概念上的片段,其中多核苷酸序列可以与具有至少20 个连续核苷酸的参考序列比较,并且其中对于两个序列的最适比对,所述多核苷酸序列在比较窗中的部分相比于参考序列(其不包括插入或缺失)可包括百分之20 以下的插入或缺失(即,缺口)。用于比对比较窗的序列最适比对可以通过Smith 的局部同源算法(Smith 和Waterman, Adv Appl Math,1981;Smith 和Waterman, J Teor Biol,1981;Smith 和Waterman, J Mol Biol,1981;Smith 等, J Mol Evol,1981)、Needleman 的同源比对算法(Needleman 和Wunsch,1970)、Pearson 的相似性检索方法(Pearson 和Lipman,1988)、这些算法的计算机执行(Wisconsin Genetics SoftwarePackage Release 7.0 中的GAP、BESTFIT、FASTA 和TFASTA, GeneticsComputer Group,575 Science Dr., Madison, Wis.) 或通过目测来进行,并选择由各种方法产生的最佳比对(即,在比较窗范围内得到最高同源性百分比)。

[0058] “保守氨基酸置换”是指具有类似侧链的残基的可交换性。例如,具有脂肪族侧链的一组氨基酸是甘氨酸、丙氨酸、缬氨酸、亮氨酸和异亮氨酸;具有脂肪族-羟基侧链的一组氨基酸是丝氨酸和苏氨酸;具有含氨基侧链的一组氨基酸是天冬酰胺和谷氨酰胺;具有芳香族侧链的一组氨基酸是苯丙氨酸、酪氨酸和色氨酸;具有碱性侧链的一组氨基酸是赖氨酸、精氨酸和组氨酸;以及具有含硫侧链的一组氨基酸是半胱氨酸和甲硫氨酸。优选的保守氨基酸置换组是:缬氨酸-亮氨酸-异亮氨酸、苯丙氨酸-酪氨酸、赖氨酸-精氨酸、丙氨酸-缬氨酸以及天冬酰胺-谷氨酰胺。

[0059] 当提及参考多肽时,术语“片段”、“衍生物”和“类似物”包括多肽,其保持与参考多肽至少基本相同的至少一种生物功能或活性。另外,术语“片段”、“衍生物”或“类似物”

由“前原-形式”分子示例,例如低活性前蛋白,其通过可切割修饰以产生具有明显更高活性的成熟酶。

[0060] 术语“基因”意思是参与产生多肽链的 DNA 片段;它包括编码区之前和之后的区域(前导区和非转录尾区)以及各编码片段(外显子)之间的插入序列(内含子)。

[0061] 术语“异源”意思是一个单链核酸序列不能与另一个单链核酸序列或它的互补链杂交。因此异源区意思是多核苷酸区或多核苷酸在它们的序列中具有不能与另一个核酸或多核苷酸杂交的区域或区。这些区域或区例如是突变区。

[0062] 术语“同源的”或“部分同源的”意思是一个单链核酸序列可与互补单链核酸序列杂交。杂交的程度可取决于许多因素,包括序列之间同一性的量以及杂交条件,例如如下所讨论的温度和盐浓度。优选地,同一性区域大于约 5bp,更优选地,同一性区域大于 10bp。

[0063] 术语“相同”或“同一性”意思是两个核酸序列具有相同序列或互补序列。因此,“同一性区”意思是多核苷酸的区域或区或全部多核苷酸与另一个多核苷酸的区或该多核苷酸相同或互补。

[0064] 术语“分离的”意思是材料从其原始环境(例如,自然环境,如果它是天然发生的)中去除。例如,活体动物中存在的天然发生的多核苷酸或酶不是分离的,但是从自然系统中一些或所有共存材料中分离的同一多核苷酸或酶则是分离的。这些多核苷酸可以是载体的一部分和/或这些多核苷酸或酶可以是组合物的一部分,以及仍可以是分离的,因为这些载体或组合物不是其自然环境的一部分。

[0065] “分离的核酸”意思是核酸,例如 DNA 或 RNA 分子,其不与 5' 和 3' 侧翼序列直接相邻,当在所述核酸来源的生物体的天然发生的基因组中存在时,所述核酸通常与所述 5' 和 3' 侧翼序列直接相邻。因此该术语描述,例如,并入载体如质粒或病毒载体的核酸;并入异源细胞基因组(或同源细胞基因组,但在不同于其天然发生的位点)的核酸;以及作为分离的分子存在的核酸,例如 PCR 扩增或限制酶消化产生的 DNA 片段或体外转录产生的 RNA 分子。该术语还描述了重组核酸,其形成编码另外的多肽序列的杂合基因的一部分,所述多肽序列可用于例如融合蛋白的产生。

[0066] “连接”是指在核酸链之间形成磷酸二酯键的过程(Sambrook 等,1982, p. 146; Sambrook,1989)。DNA 连接酶可将两个具有单链断裂(在 DNA 的两个互补链中的断裂)的 DNA 链连接在一起。可选的、双-链断裂使用互补链作为模板通过不同类型的 DNA 连接酶进行修复,但还需要 DNA 连接酶来产生最终的磷酸二酯键以完全修复 DNA。除非另外提供,可使用已知缓冲液和条件,每 0.5 μ g 大概等摩尔量的待连接 DNA 片段用 10 个单位的 T4DNA 连接酶(“连接酶”)完成连接。“产物不被连接”是指在通过使用引物来扩增完整的环状双链模板多核苷酸而获得的核酸末端之间不形成磷酸二酯键。

[0067] 术语“突变”被定义为野生型序列或亲代核酸序列的变化或肽序列的变化。这些突变可以是点突变例如转换或颠换。突变可以是一个或多个核苷酸或编码的氨基酸序列的变化。突变可以是缺失、插入或复制。

[0068] 如本文所用,简并“N, N, N”核苷酸序列表示三联体,其中“N”可以是 A、C、G 或 T。

[0069] 如本文所用,术语“天然发生”当应用于对象时是指对象可在自然界中发现这一事实。例如,存在于生物体(包括病毒)中,可从自然界来源中分离的并没有经过实验室人工刻意修饰的多肽或多核苷酸序列是天然发生的。通常,术语天然发生的是指存在于非病原

(无疾病的)个体的对象,例如对该物种将具有典型性。

[0070] 如本文所用,“核酸分子”包括至少一个碱基或一个碱基对,其分别取决于它是单链或双链。另外,核酸分子可以专有地或嵌合地属于含核苷酸的分子的任何组,例如但不限于下列核酸分子组:RNA、DNA、基因组核酸、非基因组核酸、天然发生的和非天然发生的核酸以及合成核酸。这包括——作为非限制实例——与任何细胞器例如线粒体相关的核酸、核糖体 RNA 和嵌合地包括与天然发生的成分非天然一起发生的一种或多种成分的核酸分子。另外,“核酸分子”可部分含有一个或多个非基于核苷酸-的组分,例如但不限于氨基酸和糖类。因此,作为实例而非限制,部分基于核苷酸-部分基于蛋白质的核酶被认为是“核酸分子”。另外,作为实例而非限制,用可检测的部分标记(例如放射性标记或可选地非放射性标记的核酸分子)同样被认为是“核酸分子”。

[0071] 术语“编码特定蛋白质或多肽的核酸序列”或“编码特定蛋白质或多肽的序列的 DNA”或“编码特定蛋白质或多肽的核苷酸序列”是指 DNA 序列,当被置于适当的调节序列的调控下时,其被转录并翻译成蛋白质或多肽。“启动子序列”是能够在细胞中结合 RNA 聚合酶并启动下游(3' 方向)编码序列转录的 DNA 调节区域。启动子是 DNA 序列的一部分。该序列区域在它的 3' 末端具有起始密码子。启动子序列包括最小数目的碱基,其中元件是以高于背景的可检测水平启动转录必需的。但是,在 RNA 聚合酶结合序列并在起始密码子(启动子的 3' 末端)启动转录后,转录以 3' 方向向下游进行。在启动子序列中,将发现转录起始位点(通过用核酸酶 S1 作图进行方便地确定)以及负责 RNA 聚合酶结合的蛋白结合结构域(共有序列)。

[0072] 术语“编码蛋白质或肽的核酸”或“编码蛋白质或肽的 DNA”或“编码蛋白质或肽的多核苷酸”和其他同义术语包括仅包含蛋白质或肽的编码序列的多核苷酸,以及包括包含另外的编码和/或非编码序列的多核苷酸。

[0073] 因此,在非限制实施方式中,“核酸文库”包括一个或多个核酸分子的基于载体的集合。在另一个优选的实施方式中,“核酸文库”包括核酸分子的非基于载体的集合。还在另一个优选的实施方式中,“核酸文库”包括部分基于载体和部分非基于载体的核酸分子的组合集合。优选地,组成文库的分子集合是根据各核酸分子种类可检索并可分离的。

[0074] “寡核苷酸”(或同义为“寡聚体”)是指单链多脱氧核苷酸或两个互补多脱氧核苷酸链,其可以是化学合成的。这些合成寡核苷酸可以具有或不具有 5' 磷酸。不具有 5' 磷酸的合成寡核苷酸在激酶存在的情况下,不添加磷酸与 ATP,不能与另一个寡核苷酸连接。合成的寡核苷酸将连接至未去磷酸化的片段。

[0075] 如本文所用,术语“亲代多核苷酸组”是包括一个或多个不同的多核苷酸种类的组。通常该术语用于指代子代多核苷酸组,其优选地通过亲代组的诱变处理而获得,其中术语“亲代”、“起始”和“模板”可互换使用。

[0076] 如本文所用,术语“生理条件”是指温度、pH、离子强度、粘度和与活生物体相容的和/或通常在活的培养的酵母细胞或哺乳动物细胞中胞内存在的类似生物化学参数。例如,在通常的实验室培养条件下生长的酵母细胞的胞内条件是生理条件。体外转录混合物的合适体外反应条件通常是生理条件。一般而言,体外生理条件包括 50-200mM NaCl 或 KCl, pH 6.5-8.5, 20-45°C 和 0.001-10mM 二价阳离子(例如, Mg^{++} 、 Ca^{++}); 优选约 150mM NaCl 或 KCl, pH 7.2-7.6, 5mM 二价阳离子并通常包括百分之 0.01-1.0% 非特异性蛋白质(例

如, BSA)。非离子去污剂 (Tween, NP-40, Triton X-100) 可一般以约 0.001 至 2%, 通常以 0.05-0.2% (v/v) 存在。具体水性条件可以由实践者根据常规方法选择。作为一般指导: 下列缓冲水性条件是可应用的: 10-250mM NaCl, 5-50mM Tris-HCl, pH 5-8, 任选添加二价阳离子 (一种或多种) 和 / 或金属螯合剂和 / 或非离子去污剂和 / 或膜组分和 / 或防泡剂和 / 或闪烁体。

[0077] 标准规则 (5' 至 3') 在本文用来描述双链多核苷酸的序列。

[0078] 术语“相关多核苷酸”意思是多核苷酸的区域或区相同以及多核苷酸的区域或区异源。

[0079] “特异杂交”在本文定义为在第一多核苷酸和第二多核苷酸 (例如, 与第一多核苷酸不同但基本相同的多核苷酸) 之间形成杂交体, 其中基本上不相关的多核苷酸序列在混合物中不形成杂交体。

[0080] “严格杂交条件”意思是只有序列之间存在至少 90% 同一性, 优选至少 95% 同一性以及最优选至少 97% 同一性时, 才发生杂交。参见 Sambrook 等, 1989, 其通过引用于此全部并入。

[0081] 术语“野生型”意思是多核苷酸不包括任何突变。“野生型”蛋白意思是蛋白质以自然界中发现的活性水平具有活性, 并包括自然界中发现的氨基酸序列。

[0082] 原始多核苷酸的来源可从单独的生物体 (“隔离群”)、已生长在确定成分培养基 (“富集培养物”) 中的生物体集合或最优选地未培养的生物体 (“环境样品”) 中分离。使用不依赖培养的方法以从环境样品中衍生出编码新生物活性的多核苷酸是最优选的, 原因在于它可获得未被利用的生物多样性资源。

[0083] 可从中制备多核苷酸的微生物包括原核微生物, 例如真细菌和古细菌, 以及低等真核微生物例如真菌、一些藻类和原生动物。多核苷酸可以从环境样品中分离, 在该情况下核酸可以在不培养生物体的情况下加以回收或从一种或多种培养的生物体加以回收。一方面, 这些微生物可以是嗜极端生物, 例如超嗜热菌、嗜冷菌 (psychrophiles)、嗜冷生物 (psychrotrophs)、嗜盐菌、嗜压微生物和嗜酸微生物。

[0084] 将如上文所述选择和分离的多核苷酸引入合适的宿主细胞。选择的多核苷酸优选已经存在于包括适当的控制序列的载体中。宿主细胞可以是高等真核细胞例如哺乳动物细胞, 或低等真核细胞例如酵母细胞, 或者优选地, 宿主细胞可以是原核细胞, 例如细菌细胞。将构建体引入宿主细胞可通过磷酸钙转染、DEAE-Dextran 介导的转染或电穿孔来实现 (Davis 等, 1986)。

[0085] 作为适当宿主的典型实例, 可提及的是: 细菌细胞, 例如大肠杆菌和荧光假单胞菌 (*Pseudomonas fluorescens*); 噬菌体; 真菌细胞, 例如酵母、毕赤酵母 (*Pichia pastoris*) 和黑曲霉 (*Aspergillus niger*); 昆虫细胞, 例如果蝇 (*Drosophila*) S2 和夜蛾 (*Spodoptera*) Sf9; 动物细胞, 例如 CHO、COS 或 Bowes 黑色素瘤; 腺病毒; 和植物细胞。TMCA 文库可例如以质粒形式在大肠杆菌细胞中产生, 然后可进一步将文库引入其他宿主。通过本文的教导, 适当宿主的选择被认为在本领域普通技术人员的范围内。

[0086] 当具体提及各种可应用以表达重组蛋白的哺乳动物细胞培养系统时, 哺乳动物表达系统的实例包括猴肾成纤维细胞的 COS-7 系, 描述为“SV40- 转化的猿猴细胞支持早期 SV40 突变体的复制” (Gluzman, 1981), 以及其他能够表达相容性载体的细胞系, 例如 C127、

3T3、CHO、HeLa 和 BHK 细胞系。哺乳动物表达载体将包括复制起点、合适的启动子和增强子，且还包括任何必需的核糖体结合位点、多聚腺苷化位点、剪接供体和受体位点、转录终止序列以及 5' 侧翼非转录序列。源自 SV40 剪接的 DNA 序列和多聚腺苷化位点可用于提供所需的非转录基因元件。

[0087] 含有感兴趣的多核苷酸的宿主细胞可在常规营养培养基中培养，该常规营养培养基被修饰为适合于激活启动子、选择转化子或扩增基因。培养条件，例如温度、pH 等等是以往用于筛选表达的宿主细胞的那些条件，并且对本领域普通技术人员是明显的。

[0088] 作为可使用的表达载体的代表性实例，可以提到病毒颗粒、杆状病毒、噬菌体、质粒、噬菌粒、黏粒、F 黏粒、细菌人工染色体、病毒 DNA（例如牛痘病毒、腺病毒、口蹄疫病毒、狂犬病毒和 SV40 衍生物）、基于 P1 的人工染色体、酵母质粒、酵母人工染色体和对感兴趣的具体宿主（例如杆菌、曲霉菌和酵母）特异的任何其他载体。因此，例如，DNA 可以包括在用于表达多肽的多个表达载体中的任何一个。这些载体包括染色体 DNA 序列、非染色体 DNA 序列和合成 DNA 序列。大量合适的载体是本领域普通技术人员所知的并可购得。下列载体作为实例提供；细菌：pQE 载体（Qiagen）、pBluescript 质粒、pNH 载体、（lambda-ZAP 载体（Stratagene）；ptrc99a、pKK223-3、pDR540、pRIT2T（Pharmacia）；真核：pXT1、pSG5（Stratagene）、pSVK3、pBPV、pMSG、pSVLSV40（Pharmacia）。但是，可使用任何其他质粒或其他载体，只要它们可在宿主中复制并存活。低拷贝数或高拷贝数载体可用于本发明。

[0089] 用于本发明的优选类型的载体含有 f⁻ 因子起点复制。大肠杆菌中的 f⁻ 因子（或致育因子）是在接合过程中实现高频转移并使其自身的细菌染色体低频转移的质粒。特别优选的实施方式是使用克隆载体，其被称为“F 黏粒”或细菌人工染色体（BAC）载体。这些源自大肠杆菌 f⁻ 因子，其能够稳定整合基因组 DNA 的大片段。当与来自混合的未培养的环境样品的 DNA 整合时，这使得有可能获得稳定“环境 DNA 文库”形式的大基因组片段。

[0090] 用于本发明的另一个优选类型的载体是黏粒载体。黏粒载体最初设计来克隆并增殖基因组 DNA 的大片段。克隆至黏粒载体在“MolecularCloning:A laboratory Manual”（Sambrook 等，1989）中详述。

[0091] 表达载体中的 DNA 序列被可操纵地连接至适当的表达控制序列（启动子）以引导 RNA 合成。具体命名的细菌启动子包括 lacI、lacZ、T3、T7、gpt、 λ P_R、P_L 和 trp。真核启动子包括 CMV 即时早期、HSV 胸苷激酶、早期和晚期 SV40、来自逆转录病毒的 LTR 和小鼠金属硫蛋白-I。适当的载体和启动子的选择完全在本领域普通技术人员水平之内。表达载体还含有用于翻译起始的核糖体结合位点以及转录终止子。载体还可包括用于扩增表达的适当序列。启动子区域可选自任何期望的基因，其使用 CAT（氯霉素转移酶）载体或具有可选择标记的其他载体。

[0092] 另外，表达载体优选地含有一个或多个可选择的标记基因以提供表型性状用于筛选转化的宿主细胞，例如用于真核细胞培养的二氢叶酸还原酶或新霉素抗性或例如在大肠杆菌中的四环素或氨苄青霉素抗性。

[0093] 通常，重组表达载体将包括复制起点和允许宿主细胞转化的可选择标记，例如大肠杆菌的氨苄青霉素抗性基因和酿酒酵母（*S. cerevisiae*）TRP1 基因，以及源自高表达基因的启动子以引导下游结构序列转录。这些启动子可源自操纵子，其编码糖分解酶例如

3-磷酸甘油酸酯激酶 (PGK)、 α -因子、酸性磷酸酶或热休克蛋白等等。异源结构序列在适当的阶段与翻译起始和终止序列装配,以及优选地与能够引导翻译的蛋白质分泌至外周胞质空间或细胞外培养基的前导序列装配。

[0094] 该克隆策略允许通过载体驱动启动子和内源启动子进行表达;载体启动可能对基因表达是重要的,该基因的内源启动子在大肠杆菌中不发挥作用。

[0095] 在探测所选择的 DNA 之前,分离自或源自微生物的 DNA 可优选地插入载体或质粒。这些载体或质粒优选地是那些含有表达调节序列——包括启动子、增强子等——的载体或质粒。这些多核苷酸可以是载体和 / 或组合物的一部分并且仍可以是分离的,因为这些载体或组合物不是其天然环境的一部分。特别优选的噬菌体或质粒和用于引入并包装至它们中的方法在本文描述的步骤中详述。

[0096] 纯化形式的任何来源的核酸可用作起始核酸(也定义为“模板多核苷酸”)。因此,该方法可利用 DNA 或 RNA,包括信使 RNA,该 DNA 或 RNA 可以是单链并优选是双链。另外,含有各自一条链的 DNA-RNA 杂合体可以被利用。核酸序列可以具有各种长度,其取决于待突变的核酸序列的大小。优选地,具体核酸序列为 50 至 50000 个碱基对,以及更优选 50-11000 个碱基对。

[0097] 核酸可以从任何来源获得,例如,从质粒获得,例如 pBR322,从克隆的 DNA 或 RNA 或从来自任何来源的天然 DNA 或 RNA 获得,该任何来源包括细菌、酵母、病毒或高等生物例如植物或动物。DNA 或 RNA 可以从血液或组织材料中提取。模板多核苷酸可以使用多核苷酸链式反应 (PCR, 参见美国专利号 4,683,202 和美国专利号 4,683,195) 通过扩增获得。可选地,多核苷酸可存在于在细胞中存在的载体中,以及通过本领域已知方法培养细胞并从细胞提取核酸,可以获得足够的核酸。

[0098] 具有突变的具体核酸序列的起始较小群体可用大量不同方法产生。突变可由易错 PCR 产生。易错 PCR 使用低保真度聚合条件以在较长序列上随机引入较低水平的点突变。可选地,突变可通过寡核苷酸定向诱变被引入模板多核苷酸。在寡核苷酸定向诱变中,使用限制酶消化将多核苷酸的较短序列从多核苷酸中去除并用合成的多核苷酸替代,在该合成的多核苷酸中各种碱基已从原始序列改变。多核苷酸序列也可通过化学诱变改变。化学诱变剂包括例如亚硫酸氢钠、亚硝酸、羟胺、肼或甲酸。其他试剂——其是核苷酸前体的类似物——包括亚硝基胍、5-溴尿嘧啶、2-氨基嘌呤或吡啶。通常,将这些试剂添加至 PCR 反应以替代核苷酸前体,从而突变序列。嵌入剂例如原黄素、吖啶黄、奎纳克林等等也可使用。多核苷酸序列的随机诱变可通过 X-射线或紫外光照射而实现。通常,如此诱变的质粒多核苷酸被引入大肠杆菌并增殖作为杂合质粒库或文库。

[0099] 可选地,具体核酸的较小混合群体可在自然界中发现,原因在于它们可由相同基因的不同等位基因或来自不同相关物种的相同基因(即,同源基因)组成。可选地,它们可以是在一个物种中发现的相关 DNA 序列,例如,免疫球蛋白基因。

[0100] 一旦产生具体核酸序列的混合群体,可直接使用多核苷酸或使用本领域熟知技术将多核苷酸插入适当的克隆载体。

[0101] 载体的选择取决于多核苷酸序列的大小和在本发明方法中使用的宿主细胞。本发明模板可以是质粒、噬菌体、黏粒、噬菌粒、病毒(例如逆转录病毒、副流感病毒、疱疹病毒、呼肠病毒、副黏病毒等等)或其选择的部分(例如,包被蛋白、突起糖蛋白、衣壳蛋白)。例

如,当待突变的具体核酸序列较大时,黏粒和噬菌粒是优选的,原因在于这些载体能够稳定地扩增大的多核苷酸。

[0102] 为简便起见,本发明的 TMCA 方法将被解释为意图在六个不同位点装配六个点突变。

[0103] 首先,设计六种引物并合成。与野生型序列相比,每种引物含有一个点突变。三个寡聚体被设计为正向引物,而三个寡聚体被设计为反向引物以与基因退火(图 5)。将所有六个寡聚体混合在一起并用于在实施例中详述的条件下启动 TMCA 反应。然后,通过琼脂糖凝胶确认完成的 TMCA 反应以确定反应是否成功。将 DpnI 限制酶添加至 TMCA 反应以破坏模板环状 DNA。为使 DpnI 发挥作用,模板 DNA 必须来自可甲基化 DNA 的大肠杆菌宿主。将 DpnI- 处理的反应物转化入大肠杆菌细胞以回收具有期望突变的 DNA。通过测序或期望的分析来筛选转化体。

[0104] 本发明的方法不限于六个位点。可通过本方法装配更多或更少数量的位置。还不限于在一个位置的单个变化。可设计多个引物以在相同位置覆盖不同变化,其中在每个引物上具有单个变化。大肠杆菌已用于实证;但是,其他细菌宿主可用于本方法。本发明的方法不但可以引入点突变,而且可以用简并引物产生缺失或插入或多个突变。

[0105] TMCA 反应可随引物浓度、引物 T_m (与模板的退火温度)、DNA 聚合酶、模板浓度、引物组合以及控制如何装配在不同位点的变化的不同宿主而变化。

[0106] 装配可在体外或体内或二者的组合发生。图 5 图解了引物退火至基因的图谱。本发明方法的主要用途是 GSSMSM 高表达突变型的组合重装配。但是,本发明的方法也可用于例如以下所列的任何其他应用。

[0107] 1. TMCA 可用于对基因进行特异变化,包括突变、缺失和插入。

[0108] 2. TMCA 可用于产生基于野生型基因的特异基因变体。

[0109] 3. TMCA 可用于组合突变、缺失或插入。

[0110] 4. TMCA 可用于以可控制的方式产生突变、缺失或插入的组合文库。

[0111] 5. TMCA 可用于产生组合多位点 GSSMSM 文库。

[0112] 一般而言,本发明提供产生多个子代多核苷酸的方法,所述子代多核苷酸在多个位点具有各种突变的不同组合。该方法可部分地通过下列步骤的至少一个或多个的组合来进行:

[0113] 获得多核苷酸(“第一”或“模板”)的序列信息。例如,序列可以是野生型、突变的野生型或非天然发生的序列。序列信息可以是完整的多核苷酸或感兴趣的部分区域,例如编码结合位点、结合特异性位点、催化位点或底物特异性位点的序列。多核苷酸可包括序列,例如开放阅读框、基因、多肽编码序列或酶编码序列,其具有或不具有信号或分泌序列。

[0114] 沿着多核苷酸序列,鉴定三个或更多个感兴趣的突变,例如在 3、4、5、6、8、10、12、20 或更多个位置的突变。突变可以在多核苷酸序列水平或对由多核苷酸序列例如密码子编码的氨基酸序列的突变。位置可由绝对位置或周围残基或同源性情况加以预先确定。突变位置任何一侧的序列优选是已知的。例如对于不同氨基酸,每个突变位置可含有两个或更多个突变。这些突变可以使用基因位点饱和诱变(GSSM)鉴定,如上所述以及美国专利号 6,171,820、6,562,594 或 6,764,835 中所述。

[0115] 相对于模板序列,提供包括感兴趣突变的引物。引物可以是合成的寡核苷酸。优

选地,对每个感兴趣的突变提供引物。突变可以是一个或多个核苷酸或所编码的氨基酸序列的变化、插入或缺失。因此,具有 3 个感兴趣的突变的位置可在该位置使用 3 种引物。也可以含有简并位置的引物库提供引物,以使感兴趣的突变处于任何核苷酸或天然发生的氨基酸的范围内或该范围的亚群中。例如,可提供引物库,其偏好用于脂肪族氨基酸残基的突变。

[0116] 引物可被制备为正向或反向引物,优选地至少一个正向引物和至少一个反向引物,而更优选每种引物具有相对平衡的数量(例如 3 种正向引物和 4 种反向引物)。3 种正向引物可以选择相对邻近的,与类似邻近的反向引物一致,例如 1F、2F、3F、4R、5R、6R、7R。当突变紧密地布置在一起时,可以方便地使用这样地引物,其含有多于一个位置的突变或在多个位置的突变的不同组合。

[0117] 提供含有模板多核苷酸的多核苷酸。多核苷酸优选地是环状的,更优选地是超螺旋,例如用于克隆、测序或表达的质粒或载体。多核苷酸可以是单链(“ssDNA”)以及优选地双链(“dsDNA”)。例如,TCMA 方法将超螺旋(“sc”)dsDNA 模板进行 95°C 1min 的加热步骤,模板没有变成 ssDNA(参见 Levy, NAR, 28(12):e57(i-vii)(2000),其表明将 sc dsDNA 加热至 95°C 5min 不能产生 ssDNA 分子并且如果在加热后冷却该分子则是可逆的(第 ii-iii 页,图 2))。

[0118] 在使得引物与多核苷酸退火的条件下,在反应混合物中将引物添加至模板多核苷酸。优选地,在单一反应混合物中将引物添加至多核苷酸,但可根据实验设计在多个反应中添加。

[0119] 进行引物的聚合酶延伸,优选地使得延伸沿环状模板分子进行完全。延伸产物(如本文所定义,“子代”或“修饰的延伸的多核苷酸”)可通过常规方法扩增。

[0120] 可对产物进行长度、序列、所期望的核酸性质或多核苷酸和/或多肽表达分析。其他分析方法包括原位杂交、序列筛选或表达筛选。分析可包括一轮或多轮期望性质的筛选和选择。

[0121] 产物也可转化入细胞或其他表达系统,例如无细胞系统。无细胞系统可含有与 DNA 复制、修复、重组、转录或翻译相关的酶。示例性宿主包括细菌、酵母、植物和动物细胞以及细胞系,并包括大肠杆菌、荧光假单胞菌、毕赤酵母和黑曲霉。例如,大肠杆菌的 XL1-Blue 或 Stb12 菌株可用作宿主。当使用大肠杆菌与 Dpn1(其可用于在反应后去除不期望的模板)时,模板 DNA 可来自可甲基化 DNA 的大肠杆菌宿主。细胞可用于子代多核苷酸的表达。

[0122] 多核苷酸或多肽表达产物可以从细胞中回收并分析长度、序列、所期望的核酸性质或表达多肽。分析可包括一轮或多轮期望性质的筛选和选择。

[0123] 本发明的方法可在不同反应条件下与相同或不同引物一起使用,以促进具有不同组合或突变数量的产物,例如在实施例所述的条件 1A、7A 和 13A 下。

[0124] 通过进行上述方法,本发明还提供由该方法产生的一种或多种多核苷酸,其可被筛选或选择所期望的性质。一种或多种子代多核苷酸可以表达为多肽,以及任选地被筛选或选择期望的性质。因此,本发明提供通过本发明的方法产生的多核苷酸和多肽,以及这些多核苷酸和多肽的文库。本发明进一步提供通过筛选或选择文库来筛选文库以获得一种或多种多核苷酸或多肽。

[0125] 在本发明的一个方面,产生多个修饰的多核苷酸的优选方法包括:

[0126] (a) 在单一反应混合物中将至少三种引物添加至双链模板多核苷酸,其中所述至少三种引物不重叠,并且其中所述至少三种引物的每一种包括至少一个不同于其他引物的突变,其中至少一种引物是可与所属模板的负链退火的正向引物以及至少一种引物是可与所述模板的正链退火的反向引物,和

[0127] (b) 使所述反应混合物进行聚合酶延伸反应以从所述至少三种引物产生多个延伸的修饰的多核苷酸。

[0128] 在本发明的另一方面,用未经连接酶处理的所述多个延伸的产物转化细胞。在本发明的另一方面,从所述细胞回收所述多个延伸的修饰的多核苷酸。在另一个实施方式中,分析所述回收的多个延伸的修饰的多核苷酸,例如,通过表达所述多个延伸的修饰的多核苷酸中的至少一个并分析其表达的多肽。在另一个实施方式中,选择包括感兴趣的突变的所述多个延伸的修饰的多核苷酸。

[0129] 在一个实施方式中,模板多核苷酸是环状 DNA,例如质粒或载体 DNA。在另一个实施方式中,环状 DNA 是超螺旋 DNA。

[0130] 另一方面,可获得关于模板多核苷酸的序列信息,以及沿着模板多核苷酸上可鉴定三个或更多个感兴趣的突变。在另一个实施方式中,在将多个延伸的修饰的产物转化入细胞之前,可分析聚合酶延伸获得的产物。

[0131] 在本发明的另一方面,用酶,优选地用限制酶以及更优选地用 DpnI 限制酶处理通过聚合酶延伸获得的产物,从而破坏模板多核苷酸序列。将处理的产物转化入细胞,优选大肠杆菌细胞。

[0132] 在另一个实施方式中,可使用至少两种,优选地至少三种,更优选地,至少四种、至少五种、至少六种、至少七种、至少八种、至少九种、至少十种、至少十一种、至少十二种或更多种引物。在一个实施方式中,每种引物包括单点突变(图 4A)。在另一个实施方式中,两种正向引物或两种反向引物在模板多核苷酸的相同位置上包括不同变化(图 4B)。在本发明的另一方面,至少一种引物在模板多核苷酸的不同位置上包括至少两个变化(图 4C)。还在另一个实施方式中,至少一种引物在模板多核苷酸的不同位置上包括至少两个变化并且至少两种正向或两种反向引物在模板多核苷酸的相同位置上包括不同变化(图 4D)。

[0133] 在一个实施方式中,正向引物被分入正向组,而反向引物被分入反向组,并且正向组中的引物和反向组中的引物各自独立地在相应组中被归一化为相等浓度而不考虑模板多核苷酸的位置,并且其中在归一化后,将相同量的正向和反向引物添加至反应中。在该归一化方法中,可以偏好一些位置的组合。偏好可以归因于例如,与含有多个引物的位置相比,含有单个引物的一个位置具有相对较低的引物浓度。“位置偏好”是指产生的多核苷酸显示在它的正向或反向引物组中对于引物在单个位置的并入相对于其他位置具有强的偏好性。这产生修饰的多核苷酸组合,其可在单个引物位置中具有高的突变百分比,但是在正向或反向引物组中另一个位置具有较低的突变百分比。当 TMCA 的目标是产生包括模板变化的所有可能组合的子代多核苷酸时,该偏好是不利的。可纠正该偏好,例如通过将引物归一化成在每个位置都相同的库。进行两轮 TMCA 方法可提高所期望的子代多核苷酸产量,所述子代多核苷酸包括对模板的多个变化,其中第 II 轮使用第 I 轮得到的一些变体。

[0134] 在另一个实施方式中,通过将所述引物编成多个组进行引物归一化,这取决于它们在模板多核苷酸上的位置,其中覆盖模板上的相同选择区域的引物在一个组中;将每组

中分入的引物归一化至相等的浓度;将一组中的正向引物富集成正向组并将所述每一正向引物组之间的浓度归一化至相等;将一组中的反向引物富集成反向组并将所述每一反向引物组之间的浓度归一化至相等;以及将相等量的所富集的正向和反向引物添加至所述反应。对于位置组合,没有观察到偏好。

[0135] 在一个实施方式中,提供一组各包括简并位置的简并引物,其中感兴趣的突变是一系列在简并位置上的不同核苷酸。在另一个实施方式中,提供一组简并引物,其包括对应于所述模板多核苷酸的至少一个密码子的至少一个简并密码子以及与与所述模板多核苷酸序列的密码子相邻的序列同源的至少一个相邻序列。在另一个实施方式中,简并密码子是 N,N,N 并编码任何 20 种天然发生的氨基酸。在另一个实施方式中,简并密码子编码小于 20 种天然发生的氨基酸。

[0136] 在不同实施方式中,产生包括感兴趣突变的多个修饰的多核苷酸的优选方法包括:

[0137] (a) 在单一反应混合物中将至少两种引物添加至双链模板多核苷酸,其中所述至少两种引物不重叠,并且其中所述至少两种引物的每一种包括至少一个不同于其他引物的突变,其中至少一种引物是可与所述模板的负链退火的正向引物以及至少一种引物是可与所述模板的正链退火的反向引物,

[0138] (b) 使所述反应混合物进行聚合酶延伸反应以从所述至少两种引物产生多个延伸的修饰的多核苷酸,

[0139] (c) 用酶处理所述多个延伸的修饰的多核苷酸,从而破坏模板多核苷酸,

[0140] (d) 将所述处理的延伸的修饰的多核苷酸转化入细胞,所述多核苷酸未用连接酶处理,

[0141] (e) 从所述细胞回收所述多个延伸的修饰的多核苷酸,以及

[0142] (f) 选择包括感兴趣的突变的所述多个延伸的修饰的多核苷酸。

[0143] 下列实施例表明,在模板或基因的多个位点上的单个突变可在简单的单一反应混合物中成功地组合,这基于已知的产生突变的方法是预料不到的。来自实验的所有可能组合的分布近似地反映统计计算的分布模式。反应可以被修改以根据需要产生偏好的组合。

[0144] 在 GSSMSM 技术下, TMCA 技术不使用互补引物,该互补引物与模板多核苷酸正链和负链退火。根据含有对于该 TMCA 发明所描述的引物(在单个热循环反应中正向和反向组)的热循环延伸的合理期望将是由正向和反向引物的各个引物对确定的扩增线性多核苷酸的专门集合(exclusive collection)。TMCA 条件设置几乎与标准 PCR 条件相同。预期在 TMCA 反应中产生多种 PCR 产物,其中每种产物比起在单个反应中使用地一组引物包括的整套突变更少的突变,例如,当使用 6 种引物并且每种引物包括 1 个突变时,小于 6 个突变。另外,不期望 PCR 产物被转化入细胞并得以扩增。令人惊讶地, TMCA 方法可产生在一个分子中包括多个变化的特异基因变体,并可不使用连接步骤而进行,因此简化了产生多个突变的过程。

实施例

[0145] 实施例 1. TMCA 方法的示例性步骤显示如下:

[0146] TMCA 反应

[0147]	↓		
[0148]	<u>Dpn1 处理</u>		
[0149]	↓		
[0150]	<u>转化入宿主</u>		
[0151]	↓		
[0152]	<u>筛选</u>		
[0153]	1. 进行 TMCA 反应		
[0154]	条件 1		
[0155]	Pfu 10× 缓冲液	2.5 μ l	
[0156]	DMSO	2.5 μ l	
[0157]	dNTP (10mM)	0.5 μ l	
[0158]	模板 DNA (25ng/ μ l)	1 μ l	
[0159]	Pfu Turbo	0.5 μ l	
[0160]	水	14 μ l	
[0161]	正向引物 (5 μ M)	2 μ l	
[0162]	反向引物 (5 μ M)	2 μ l	
[0163]	总计	25 μ l	
[0164]	条件 2		
[0165]	PfxAccu 缓冲液	5 μ l	
[0166]	模板 DNA (25ng/ μ l)	1 μ l	
[0167]	Pfx Accuprime	0.4 μ l	
[0168]	水	37.6 μ l	
[0169]	正向引物 (5 μ M)	3 μ l	
[0170]	反向引物 (5 μ M)	3 μ l	
[0171]	总计	50 μ l	
[0172]	条件 3		
[0173]	PfxAccu 缓冲液	2.5 μ l	
[0174]	模板 DNA (25ng/ μ l)	1 μ l	
[0175]	Pfx Accuprime	0.2 μ l	
[0176]	水	17.3 μ l	
[0177]	正向引物 (5 μ M)	2 μ l	
[0178]	反向引物 (5 μ M)	2 μ l	
[0179]	总计	25 μ l	
[0180]	<u>循环</u>		
[0181]		<u>Robocycler</u>	<u>Perkin-Elmer</u>
[0182]	起始变性	95°C 1min	95°C ;3min
[0183]	变性	95°C 45sec	95°C ;45sec
[0184]	退火	50°C 1min	50°C ;45sec 20 个循环
[0185]	延伸	68°C 2min/kb	68°C ;2min/kb

[0186] Polish(修补) 68℃ 5min 68℃ 5min

[0187] 4℃ ;无限期 4℃ ;无限期

[0188] 2. 在琼脂糖凝胶上跑 50 的 TMCA 反应物,以测定反应是否成功。

[0189] 3. 将 10 的 DpnI 限制酶稀释在 3 μ l 水和 1 μ l 缓冲液 4(New England Biolab) 中。将 5 μ l 稀释的酶添加至每个 TMCA 反应中。在 37 度温育 4-8 小时。

[0190] 4. 通过标准转化方法将 DpnI- 处理的反应物转化入大肠杆菌细胞。

[0191] 5. 通过测序或期望的分析筛选产生的克隆。

[0192] 实施例 2.

[0193] 在第一个实验中,选择基因上的六个位点以组合(图 5)。对于三个位点,使用正向引物建立反应,而对于其余三个位点,使用三个反向引物建立反应。通过测序鉴定来自所述反应的变体。存在六十四个不同的可能组合。在条件 1 下,更多具有较低突变位点数量的变体是优选的(图 6)。在条件 2 和 3 下,更多具有较高突变位点数量的变体是优选的(图 6)。来自组合数据(总计)的所有可能组合的分布与来自统计计算的分布模式类似(图 6)。当筛选零至六百个克隆时,图 7 中的一条曲线显示变体的期望覆盖度(%),而另一条曲线显示完全覆盖度的可能性。期望覆盖度(%)曲线上的圆形(即,78%、95%、99%和 100%)显示当筛选 96、192、288 或 384 个克隆时的期望覆盖度。期望覆盖度(%)曲线下部的方形(即,70%、91%、95%和 98%)显示来自实验数据的实际覆盖度。数据显示期望覆盖度与实际覆盖度之间的近乎完美的匹配。

[0194] 表 1. 六突变装配

[0195]

组合	筛选	单一克隆	单一克隆
1A	1.5×	69% (44/64)	69% (44/64)
1A	3×	80% (51/64)	80% (51/64)
7A	1.5×	63% (40/64)	75% (42/64)
13A	1.5×	61% (39/64)	67% (43/64)
13A	3×	76% (49/64)	80% (51/64)
1A+7A	3×	91% (58/64)	91% (58/64)
1A+7A	4.5×	94% (60/64)	94% (60/64)
1A+13A	3×	91% (58/64)	91% (58/64)
1A+13A	4.5×	95% (61/64)	95% (61/64)
1A+13A	6×	98% (63/64)	98% (63/64)

[0196] 实施例 3.

[0197] 在第二个实验中,选择基因上的四个位点以组合(图8)。对于两个位点,使用正向引物建立反应,而对于其余两个位点,使用两条反向引物建立反应。通过测序鉴定来自反应的变体。存在十六种不同的可能组合。与第一个实验类似,条件1产生更多具有较低突变位点数量的变体,而条件2和3产生更多具有较高突变位点数量的变体(图9和10)。来自组合数据(总计)的所有可能组合的分布与统计计算的分布模式类似(图9和10)。

[0198] 表2. 四突变装配

[0199]

组合	筛选	单一克隆	单一克隆
2A	2×	75% (12/16)	81% (13/16)
8A	2×	62% (10/16)	87% (14/16)
14A	2×	68% (11/16)	75% (12/16)
2A+8A	4×	93% (15/16)	100% (16/16)
2A+14A	4×	93% (15/16)	93% (15/16)
2B	2×	75% (12/16)	75% (12/16)
8B	2×	50% (8/16)	68% (11/16)
14B	2×	75% (12/16)	87% (14/16)
2B+8B	4×	81% (13/16)	81% (13/16)
2B+14B	4×	87% (14/16)	87% (14/16)

[0200] 实施例4.

[0201] 在第三个实验中,选择基因上的三个位点以组合(图11)。对于两个位点,使用正向引物建立反应,而对于第三个位点,使用反向引物建立反应。通过测序鉴定来自反应的变体。在这种情况下,存在八种不同的可能组合。在9B条件下,通过测序24个克隆,回收所有8种变体。参见图12。

[0202] 实施例5.

[0203] 选择13个GSSMSM高表达突变型(5个位点)以增加热稳定性和提高脂酶的比活性(图13)。将三个位点(N168S、N171E和M176W)分在一组并被包含在一个引物中。文库大小是 $6 \times 6 \times 2 \times 2 \times 2 = 288$ 。按下列方法进行反应:将正向引物分入正向组并将反向引物分入反向组,并且正向组中的引物和反向组中的引物各自独立地在相应组中被归一化为相等浓度,而不考虑模板多核苷酸的位置,并且其中在归一化后,将相同量的正向和反向引物添加至反应中。位置1和2的组合是偏好的(图13和14A、B、C)。与位置1和3的组合或位置2和3的组合相比,对于位置1和2的组合获得较低百分比的可能独特变体(uniquevariant)。已进行两轮TMCA反应。在第II轮中,使用第I轮得到的一些变体。在

测序 720 个克隆后（文库的 $2.5\times$ 覆盖度），获得第 I 轮中的 288 个独特变体中的 46%。 $1\times$ 覆盖度测序意思是测序的变体（子代）数量等于可能的独特变体的数量，因此， $2.5\times$ 覆盖度表明测序的变体（子代）数量等于独特变体可能数量（288）的 2.5 倍。已进行两轮 TMCA 反应。在第 II 轮中，使用第 I 轮得到的一些变体作为模板多核苷酸。将用于第 II 轮中每个 TMCA 反应的引物修剪，以获得在第 I 轮未实现的变体。在第 II 轮后，获得 288 个独特变体的 95.5%。筛选后从该文库获得十个高表达突变型（图 14A、B、C）。

[0204] 实施例显示 TMCA 方法可高效产生组合文库。突变体位置限定了一些限制，但可设计可选方案来克服这些限制。新的改进的修饰明显降低偏好。可进行多轮 TMCA 反应，以克服某些偏好。TMCA 方法显示出对多种酶和载体系统有效。载体大小限制可以长至 11kb。

[0205] 显然，根据上述教导，本发明的众多修改和变化是可能的。因此，应理解的是，在所附权利要求的范围内，本发明可以不同于本文具体描述的方式进行实践。

进化-GSSM™ (1)

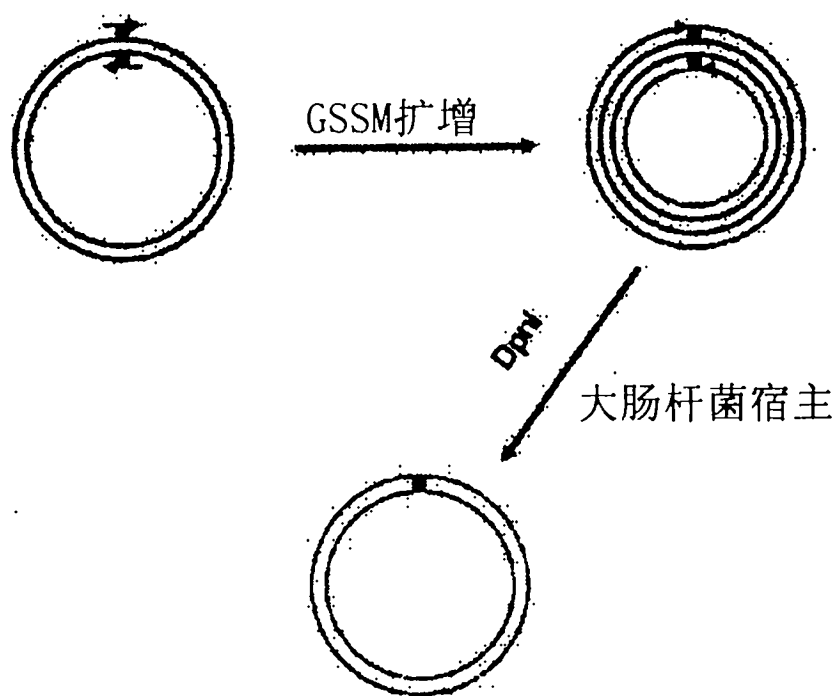


图 1

进化-GSSM™ (2)

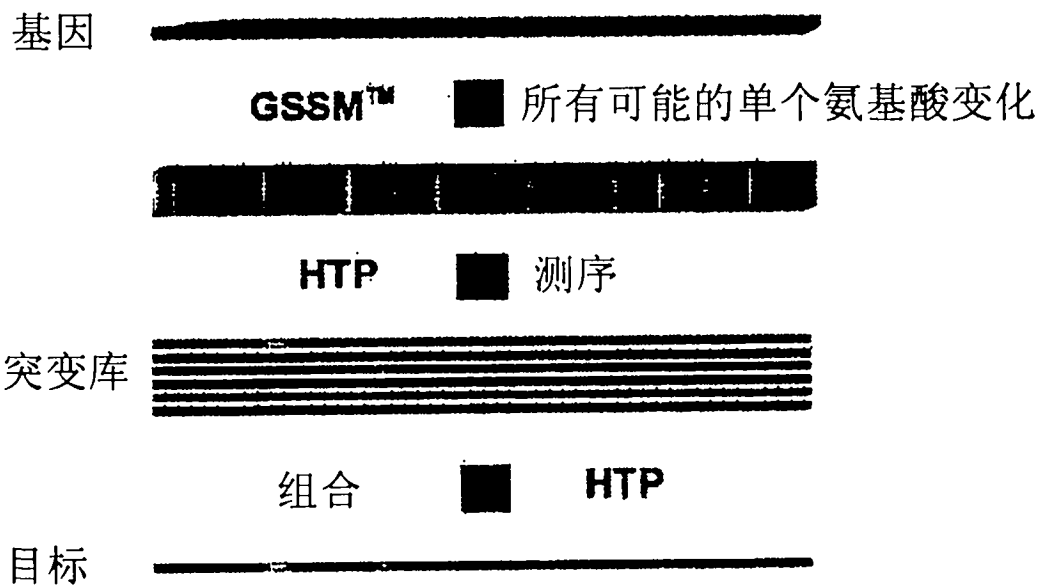


图 2

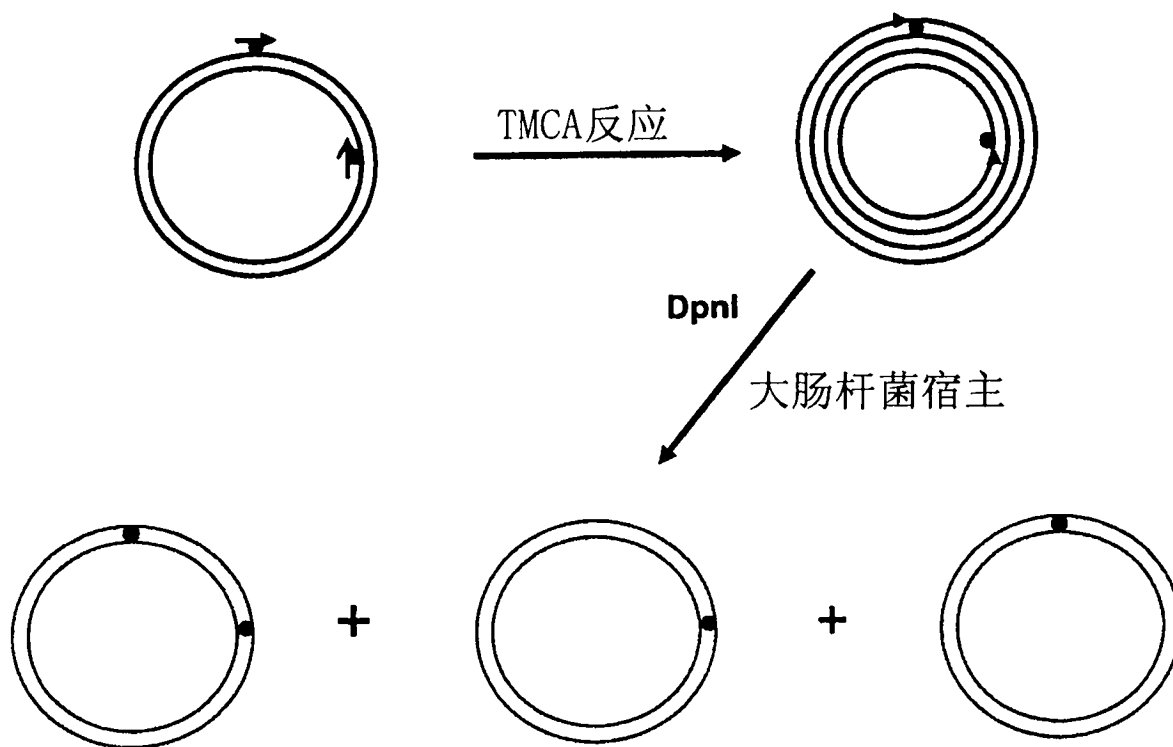


图 3

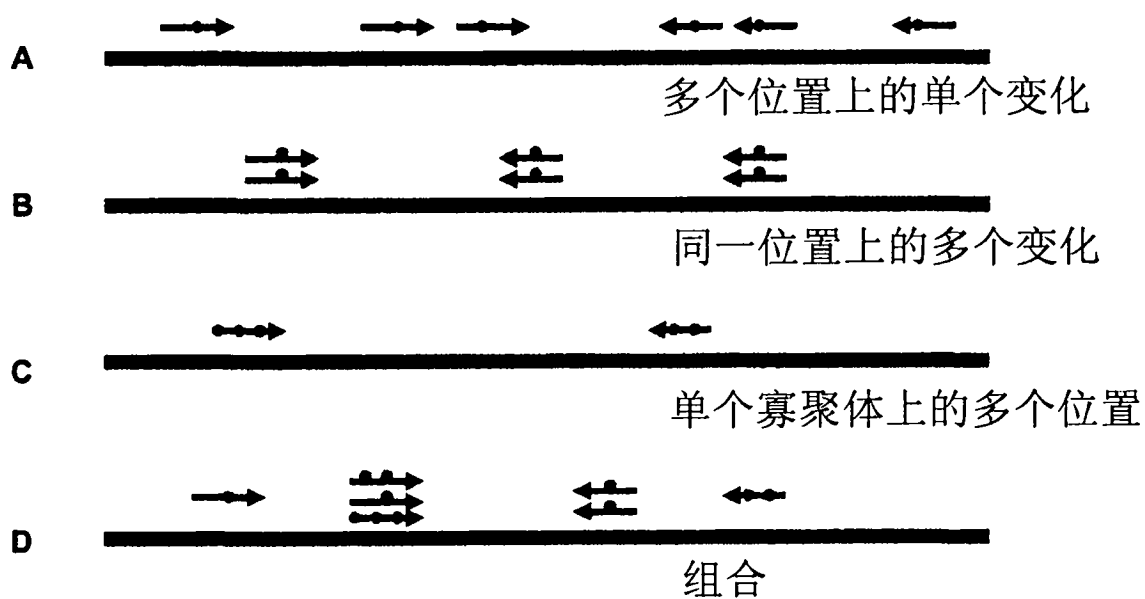


图 4

六突变装配



图 5

六突变装配

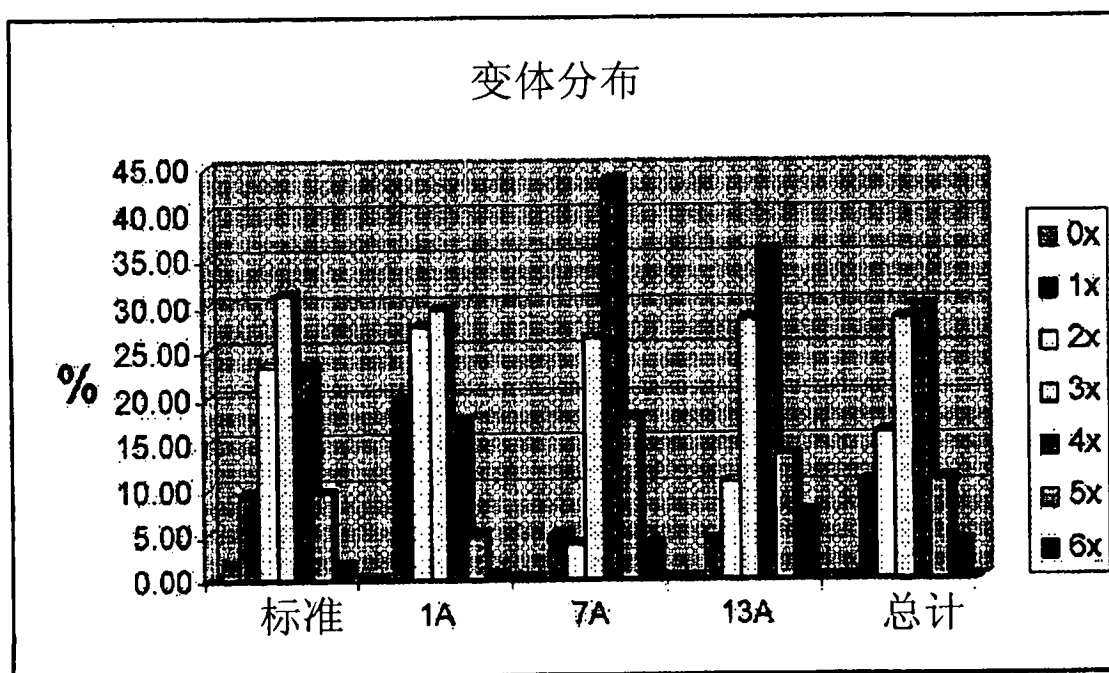


图 6

六突变装配

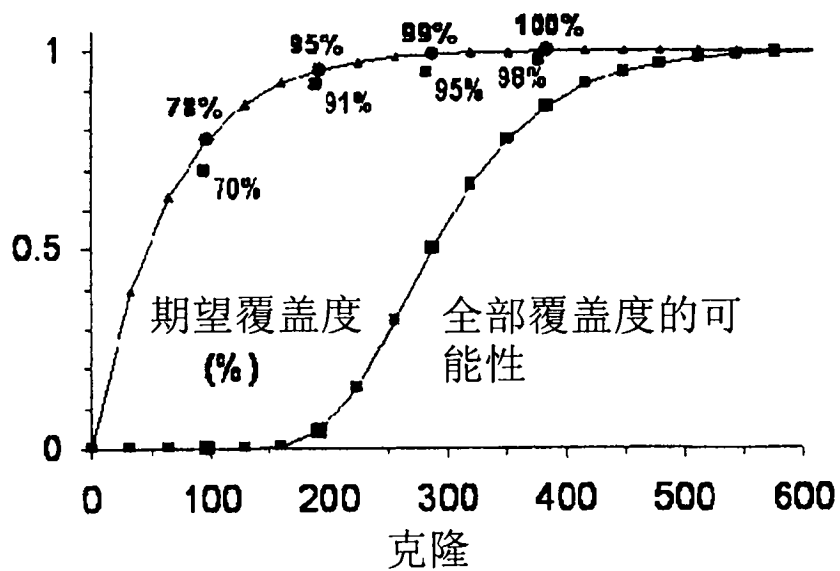


图 7

四突变装配



图 8

四突变装配

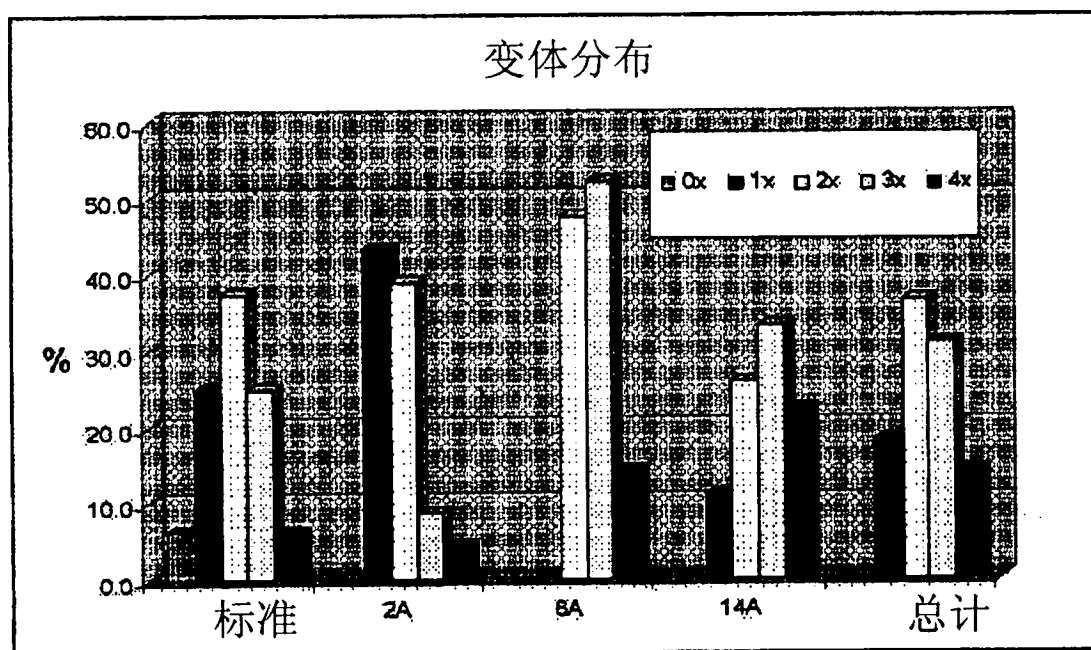


图 9

四突变装配

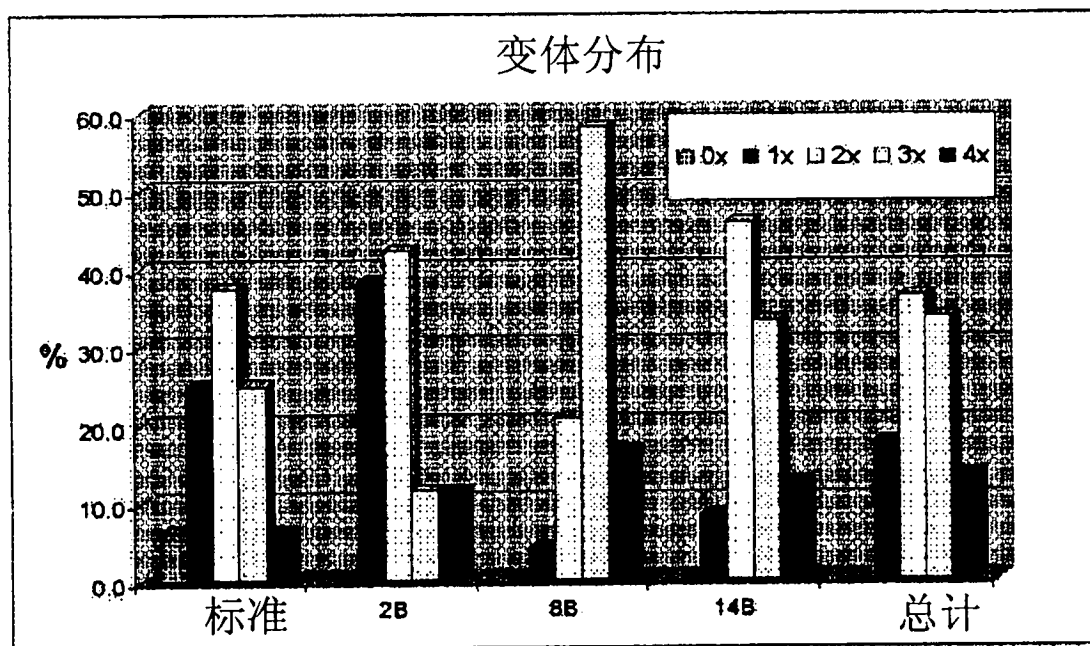


图 10

三突变装配



图 11

三突变装配

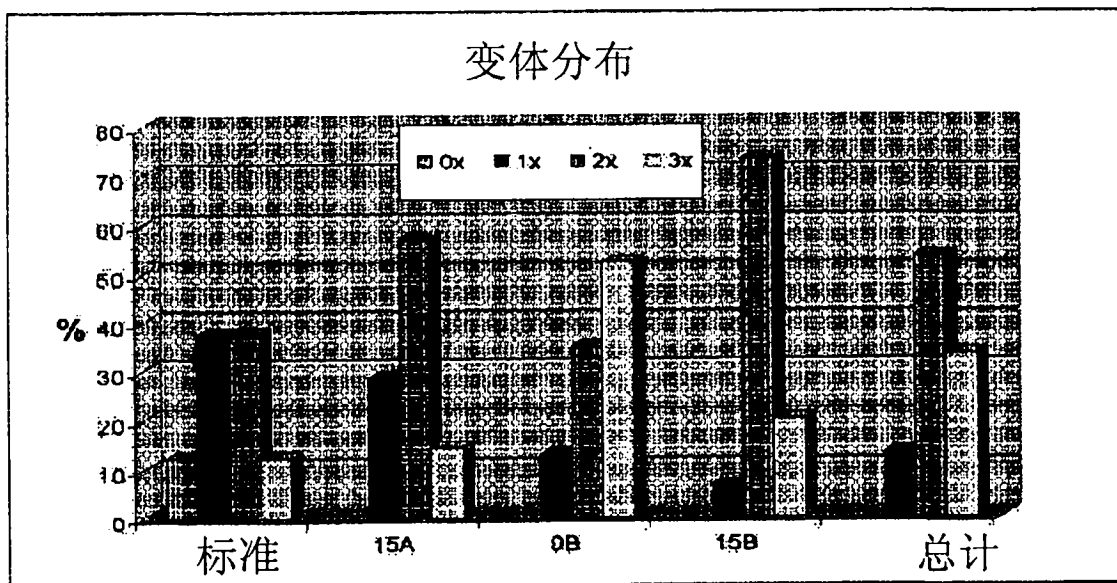


图 12

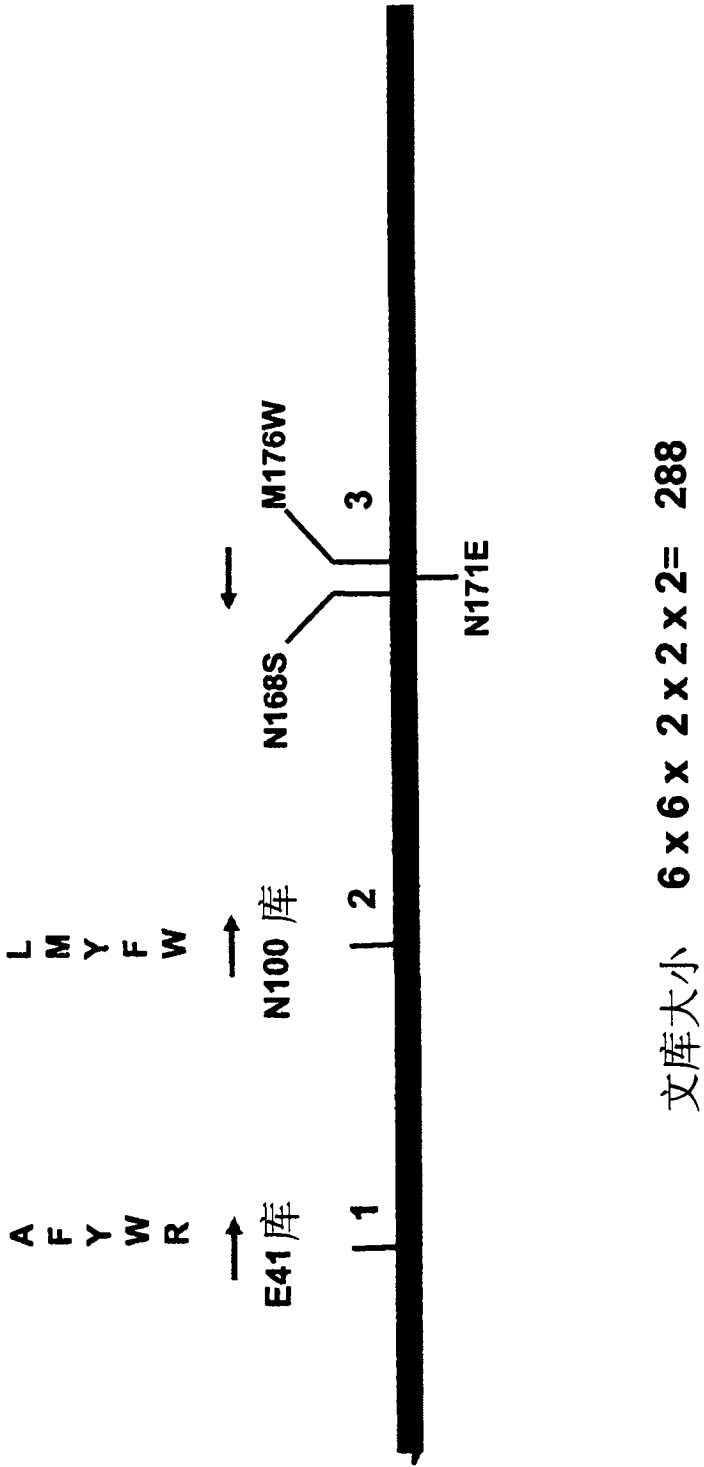


图 13

变体的破坏 - TMCA第I轮

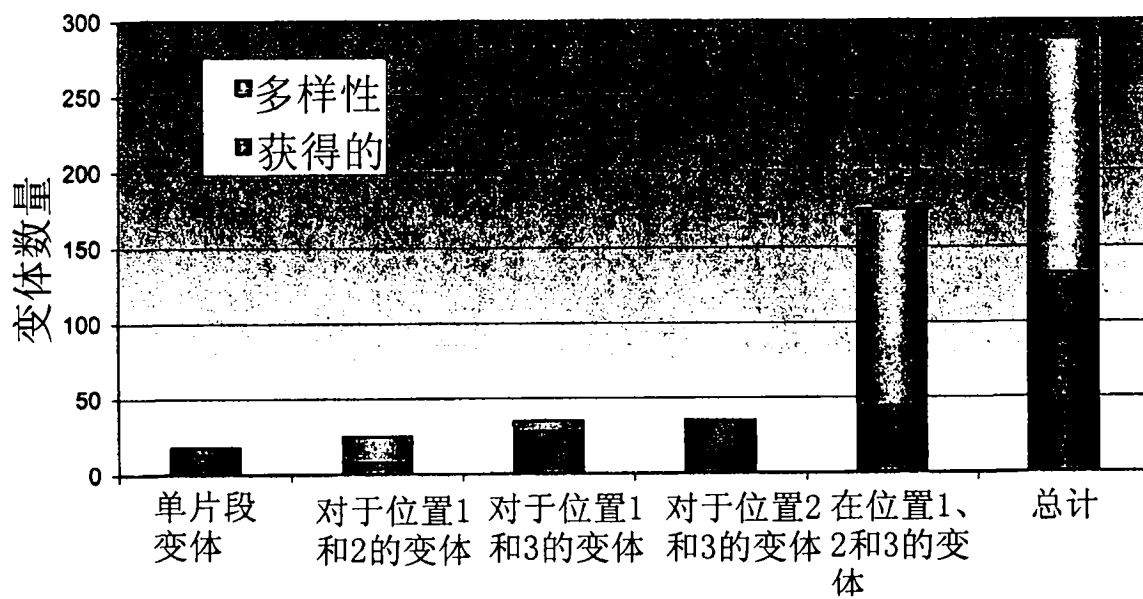


图 14A

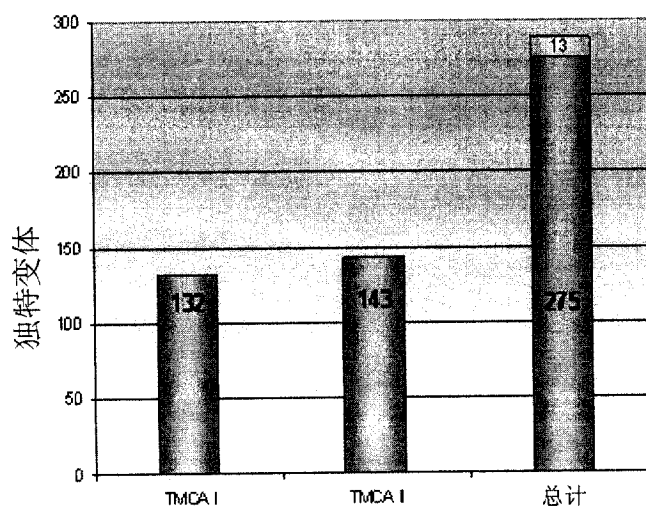


图 14B

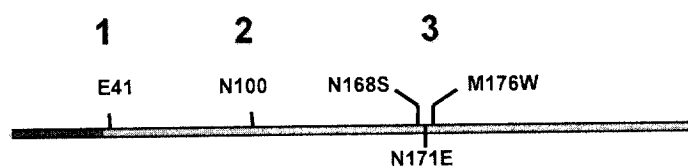


图 14C