

[19] 中华人民共和国国家知识产权局

[51] Int. Cl.

C12N 9/12 (2006.01)

C12Q 1/68 (2006.01)



[12] 发明专利说明书

专利号 ZL 01815238.4

[45] 授权公告日 2009 年 2 月 18 日

[11] 授权公告号 CN 100462433C

[22] 申请日 2001.7.9 [21] 申请号 01815238.4

审查员 曾繁辉

[30] 优先权

[74] 专利代理机构 北京市柳沈律师事务所

[32] 2000.7.7 [33] US [31] 60/216594

代理人 巫肖南 封新琴

[86] 国际申请 PCT/US2001/021811 2001.7.9

[87] 国际公布 WO2002/004680 英 2002.1.17

[85] 进入国家阶段日期 2003.3.6

[73] 专利权人 维西根生物技术公司

地址 美国得克萨斯州

[72] 发明人 苏珊·H·哈丁

詹姆斯·M·布里格斯 涂晓春

高肖联 理查德·威尔森

[56] 参考文献

WO0036151A1 2000.6.22

权利要求书 2 页 说明书 89 页 附图 14 页

WO0036152A1 2000.6.22

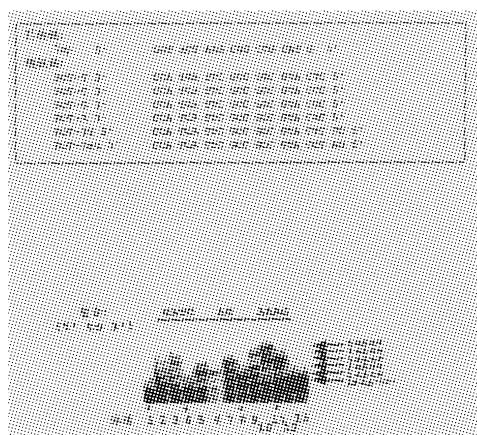
BIOCHEMISTRY 1998.1.1

[54] 发明名称

实时序列测定

[57] 摘要

本发明公开了一种测序方法，其可以对单个的 DNA 或 RNA 分子或其部分进行直接或基本上实时的测序。所述方法涉及用原子和/分子标记改造聚合酶和/或 dNTP，所述标记具有可被检测系统检测的性质。



1. 一种组合物，其包含聚合酶、引物、模板、以及对应于所述聚合酶的单体类型，所述聚合酶包括与该聚合酶上的位点共价连接的分子标记，其中至少一种单体类型包括与所述单体的 β 、 γ 或末端磷酸基共价连接的分子标记，其中至少一种标记是具有荧光特性的荧光标记，所述荧光特性在一系列单体掺入中的每一个掺入之前，之中和/或之后发生改变，其中单体的掺入顺序对应于模板中单体的相应顺序的互补形式。
2. 权利要求 1 的组合物，其中所述聚合酶是逆转录酶。
3. 权利要求 1 的组合物，其中所述聚合酶选自 Taq DNA 聚合酶 I, T7 DNA 聚合酶, 测序酶, 或大肠杆菌 DNA 聚合酶 I 的 Klenow 片段。
4. 权利要求 2 的组合物，其中所述逆转录酶是 HIV-1 逆转录酶。
5. 权利要求 1 的组合物，所述聚合酶标记包含供体荧光标记，其中每一种单体类型包括与该单体的 β 、 γ 或末端磷酸基共价连接的不同受体荧光标记，且其中所述荧光特性是在一系列单体掺入的每一个掺入期间发生在所述聚合酶标记与插入的单体上的标记之间的荧光共振能量转移(FRET)。
6. 权利要求 1 的组合物，其中所有标记都是荧光标记，并且所述荧光特性包含发射荧光的强度和/或频率。
7. 权利要求 6 的组合物，其中所述荧光特性是荧光共振能量转移(FRET)，其中所述单体标记或聚合酶标记是供体荧光标记，并且所述聚合酶标记或单体标记是受体荧光标记，且其中 FRET 在所述两种标记彼此紧密靠近时发生。
8. 一种组合物，其包含聚合酶、引物、模板、对应于该聚合酶的单体类型，所述聚合酶包括共价连接到该聚合酶上位点的供体荧光标记，所述模板含有核苷酸的未知序列，其中每个单体类型包括共价连接到该单体的 β 或 γ 磷酸基的不同受体荧光标记，其中在一系列单体掺入的每一个掺入期间，在所述聚合酶供体荧光标记和掺入单体上的荧光受体标记之间发生荧光共振能量转移(FRET)，并且其中单体的掺入顺序对应于模板中单体的相应顺序的互补形式。
9. 权利要求 8 的组合物，其中所述聚合酶是逆转录酶。
10. 权利要求 8 的组合物，其中所述聚合酶选自 Taq DNA 聚合酶 I, T7

DNA 聚合酶，测序酶，或大肠杆菌 DNA 聚合酶 I 的 Klenow 片段。

11. 权利要求 9 的组合物，其中所述逆转录酶是 HIV-1 逆转录酶。

12. 单分子测序装置，其包含：

第一小室，在该第一小室中有至少一种含有聚合酶、引物和模板的复制复合物，所述聚合酶具有共价连接到该聚合酶上位点的供体荧光标记，

第二小室，该第二小室中有对应于所述聚合酶或逆转录酶的脱氧核昔三磷酸(dNTP)类型，所述类型中的一些或全部具有与其 β 或 γ 磷酸基共价连接的受体荧光标记，

连接这些小室的通道，其中在一系列 dNTP 掺入的每一个掺入期间，在所述聚合酶标记和掺入的 dNTP 的标记之间发生荧光共振能量转移(FRET)，并单体掺入顺序对应于模板中单体的相应顺序的互补形式，和

检测器，其适于检测每种复制复合物每次发生 FRET 时产生的荧光，从而在 CCD 照相机视野中产生每种复制复合物的 FRET 事件，并将这些 FRET 事件转换成每种复制复合物掺入 dNTP 的顺序。

13. 权利要求 12 的装置，其中的第二小室包括四个小室，每一种 dNTP 一个小室，且每种 dNTP 包括不同的受体荧光团标记。

14. 在单分子水平进行核酸测序的方法，其包含：

将包含核昔酸的未知序列的核酸模板与测序组合物接触，该组合物包括聚合酶、引物、以及对应于该聚合酶的单体，所述聚合酶具有共价连接到聚合酶上位点的荧光供体，且每种单体具有与其 β 、 γ 或末端磷酸基共价连接的独特荧光受体；

用激发光源的光激发荧光供体；

检测在一系列单体掺入的每一个掺入期间，在所述聚合酶荧光供体和掺入的单体荧光受体之间发生的荧光共振能量转移所发射的荧光；和

将所述单体掺入顺序转化为模板中单体的相应顺序的读取值。

15. 权利要求 14 的方法，其中每种复制复合物的所述聚合酶、模板或引物固定在支持物上。

16. 权利要求 12 的方法，其中每种复制复合物的所述聚合酶、模板或引物固定在支持物上的区域，范围，孔，凹槽或通道中。

17. 权利要求 12 的方法，其还包括多种聚合酶、引物和模板，以形成多种复制复合物，所述聚合酶具有与该聚合酶的位点共价连接的荧光供体。

实时序列测定

发明背景

1. 发明领域

本发明涉及一种单分子测序装置和方法。

更具体地，本发明涉及一种采用标记的聚合剂和/或标记的单体的单分子测序装置和方法，其中所述标记的聚合剂和/或标记的单体在单体插入到增长的聚合链之前，之中和/或之后可检测的性质发生变化。所述装置和方法很适合于在接近实时或实时条件下对 DNA, RNA, 多肽, 糖类，或类似的生物分子的序列进行测序。

本发明还涉及一种采用标记的解聚剂 (depolymerizing agents) 和/或标记的可解聚聚合物 (depolymerizable polymer) 的单分子测序装置和方法，其中所述标记的解聚剂和/或标记的可解聚聚合物在单体脱离可解聚的聚合链之前，之中和/或之后可检测的性质发生变化。所述装置和方法很适合于对 DNA, RNA, 多肽, 糖类，或类似的生物分子序列进行测序。本发明还涉及检测证明标记的聚合剂或解聚剂和标记的或未标记的聚合亚单位如单体或单体集合间相互作用的信号，其中所述被检测的信号提供有关单体顺序的信息。在优选的实施方案中，所述方法在接近实时或实时条件下进行。

2. 相关领域

传统 DNA 测序总结

快速准确测定 DNA 片段的碱基或‘序列’顺序是关键的技术进展，其重要性十分突出。对 DNA 序列的了解使对生命的分子基础有了更广泛的了解。有关 DNA 序列的信息为科学家提供了大量生物过程的关键信息。DNA 的碱基顺序限定了 RNA 的碱基顺序，RNA 是细胞中直接编码蛋白质信息内容的分子。一般用 DNA 序列信息推断蛋白质序列信息。碱基顺序指示 DNA 结构和其功能，并提供了分子程序，它可以限定正常的发育，遗传疾病或癌症的表现。

对 DNA 序列的了解以及对这些序列进行操纵的能力加速了生物技术的发展并导致了分子技术的发展，其提供了提出并回答重要的科学问题的工具。聚合酶链式反应 (PCR) 是一种重要的生物技术，其为核酸的序列特异性检测提供了便利，其依赖于序列信息。DNA 测序方法使科学家可以确定 DNA 中是否引入了改变，并分析所述改变对生物体的生物学变化的影响，而不考虑研究的生物体类型。最终，DNA 序列信息可以提供唯一地鉴定个体的方法。

为理解所述的 DNA 测序过程，我们必须了解 DNA 一些知识。首先，DNA 分子由四种碱基构成，腺嘌呤 (A)，鸟嘌呤 (G)，胞嘧啶 (C)，胸腺嘧啶 (T)。这些碱基通过氢键相互之间以非常特异性的方式作用，例如 A 与 T 作用，G 与 C 作用。这些碱基间的特异性相互作用称为碱基对。事实上，这些碱基对(以及碱基堆积作用)稳定了双螺旋 DNA。DNA 分子的两条链呈现反平行方向，其中一条链是 5' 到 3' 方向，另一条链是 3' 到 5' 方向。术语 5' 和 3' 指 DNA 骨架的方向性，并且是描述碱基顺序的关键。按照惯例，描述 DNA 序列的碱基顺序采用 5' 到 3' 方向，由左向右书写。这样，如果知道一条 DNA 链的序列，可以推断出其互补序列。

桑格 DNA 测序法 (Sanger DNA Sequencing) (Enzymatic Synthesis)

桑格测序法是当今最为普遍的 DNA 测序法 (Sanger et al., 1977)。该方法利用了聚合酶的几个特点：其制造精确 DNA 分子拷贝的能力，其合成的方向性 (5' 到 3' 方向)，其需要从 DNA 链 ('引物') 开始合成，其需要所述引物末端的 3' OH。如果没有 3' OH，聚合酶不能延长 DNA 链。如果双脱氧核苷酸 (ddNTP；ddATP, ddTTP, ddGTP, ddCTP)，缺少 3' OH 的碱基类似物，加入到酶测序反应中，其整合到通过聚合酶增长的链中。但是，一旦 ddNTP 整合，聚合酶不能再在该链的末端添加任何其他的碱基。

重要地是，采用相同的碱基整合法则通过聚合酶整合 ddNTP 到 DNA 链，所述法则指示天然核苷酸的整合，其中 A 限定整合 T, G 限定整合 C(反之亦然)。

荧光 DNA 测序法

自动 DNA 测序仪的引入产生了测定 DNA 序列信息的巨大进步 (Smith et

al., 1986)。使用自动测序仪分离测序反应产物，测定并采集(通过计算机)反应数据，并分析碱基顺序以自动推导出 DNA 片段的碱基序列。自动测序仪检测含有荧光标记的延伸产物。采用自动测序仪得到的序列读取长度有赖于多种参数，但一般长度为 500 到 1000 个碱基(3-18 小时数据采集)。在最大能力下自动测序仪可以平行从 96 个样品采集数据。

当染料标记的终止子化学用于测定测序产品时，通过 ddNTP 结合的荧光标记的颜色确定碱基的身份。在对反应组装并加工适当数量的循环(3-12 小时)后，制备的延伸产物上样到自动测序仪的单个泳道中(除去未整合的，染料标记的 ddNTP，反应被浓缩；1-2 小时)。染料终止子化学的优势在于只有在用染料标记的 ddNTP 终止时延伸产物可见；过早终止的产物不能被检测。因此，这样的化学一般得到减少的背景噪音。

现有技术中，染料终止子化学采用 4 种能量转移荧光染料(Rosenblum et al., 1997)。这些终止子包括与 4 种不同双氯若丹明(d 罗丹明 dRhodamine)受体染料之一相连的荧光素供体染料(6-FAM)。所述与终止子相连的 d-若丹明受体染料对于 G-, A-, T- 或 C- 分别是双氯[R110]，双氯[R6G]，双氯[TAMRA]或双氯[ROX]。在自动测序仪中所述供体染料(6-FAM)有效吸收氩离子激光的能量，并将此能量传递给连接的受体染料。优化连接终止子供体和受体部分的接头的间隔以达到基本上 100% 有效的能量传递。从这些受体发射的荧光信号显现出最小重叠的光谱，采用 10 nm 有效滤光器(virtual filters)由 ABI PRISM 377 DNA 序列仪采集，对应于 G-, A-, T- 或 C- 终止子所述滤光器分别以 540, 570, 595 和 625 nm 为中心。因此，能量传递染料标记的终止子产生更强的信号并提高了光谱分辨率。这些改进产生了更为准确的 DNA 序列信息。

在自动 DNA 测序反应中所用的主要的酶是水生栖热菌(*Thermus aquaticus*)的 DNA 聚合酶 I 的基因工程化形式。优化此酶，AmpliTaq DNA 聚合酶，FS，以更为有效地将 ddNTP 整合并消除 3' 到 5' 和 5' 到 3' 核酸外切酶活性。在水生栖热菌 DNA 聚合酶中以酪氨酸代替在 667 位天然存在的苯丙氨酸使得该酶减弱了相对于 ddNTP 来说对 dNTP 的优先整合(Tabor and Richardson, 1995；

Reeve and Fuller, 1995)。因此，该酶中单个羟基基团负责区别 dNTP 和 ddNTP。消除 3' 到 5' 核酸外切酶活性，该活性使聚合酶能从新复制的 DNA

链除去错误整合的碱基(校对活性)，因为它也使聚合酶能除去整合的 ddNTP。消除 5' 到 3' 核酸外切酶活性是因为它从反应产物的 5' 端驱除碱基。由于在凝胶电泳中所述反应产物按大小分离，只有当所述反应产物具有共同的端点时才能得到可供判断的序列数据。更具体地，引物限定了延伸产物的 5' 端，整合的，颜色编码的 ddNTP 限定了所述分子的 3' 端的碱基身份。因此，传统的 DNA 测序涉及分析大量具有相同 5' 端点的 DNA 分子，但是这些分子在所述 DNA 链的 3' 端的 ddNTP 的位置不同。

基因组测序

研究人员经常需要测定 DNA 片段的序列，所述片段大于 500-1,000 个碱基平均测序读取长度。不令人惊奇的是，已经开发了完成此工作的方法。这些方法分为两类，随机的或定向的 (random or directed)，根据待测片段的大小确定采用的方法。

在随机或鸟枪 DNA 测序中，大的 DNA 片段(一般大于 20,000 碱基对)断裂为较小的片段，它们插入到克隆载体中。假定这些较小克隆内包含的信息总量等于原始 DNA 片段内包含的信息。随机挑选多个较小克隆，制备进行测序反应的 DNA 模板，与插入片段相邻的载体 DNA 序列碱基配对的引物用于启动测序反应(对于 20 kbp 的插入片段需要 2-7 天)。然后，检查每一加入的碱基(人工或通过软件启动进行) (PHRED, Ewing et al., 1998)；每一序列反应 1-10 分钟)，通过对从较小 DNA 片段得到的序列进行计算机组合重建原始 DNA 片段序列。根据时间估计，如果采用鸟枪测序法 (shotgun sequencing)，插入片段在 20 kbp 预计在 3-10 天中完成。此方法广泛用于测定有序片段的序列，所述片段代表整个人类基因组 (<http://www.nhgri.nih.gov/HGP/>)。但是，此随机方法一般不足以完成序列确定，由于计算机组合后在序列中经常留下间隙。定向的策略(如下所述)经常用于完成测序项目。

定向或引物步行(primer-walking)测序策略可用于填充大的片段测序的随机相(random phase)之后残留的间隙，以及作为对较小 DNA 片段测序的有效方法。此策略采用的 DNA 引物退火到模板的单个位点，并作为链延伸的起始位点。此方法需要一些序列信息以设计引物。得自第一反应的序

列用作设计用于下一反应的引物，重复进行这些步骤直到确定完整的序列。因此，基于引物的策略涉及从已知到未知 DNA 区的重复的测序步骤，所述方法最小化冗余，并不需要额外的克隆步骤。但是，此策略需要对每一测序循环合成新的引物。

设计和合成新的引物，以及合成它们所需的费用和时间限制了对大 DNA 片段测序时引物步行法的常规应用。研究人员已提示采用短引物库以消除常规引物合成的需要 (Studier, 1989; Siemieniak and Slightom, 1990; Kieleczawa et al., 1992; Kotler et al., 1993; Burbelo and Iadarola, 1994; Hardin et al., 1996; Raja et al., 1997; Jones and Hardin, 1998a, b; Ball et al., 1998; Mei and

Hardin, 2000; Kraltcheva and Hardin, 2001)。由于每一引物用于引发多个反应，所以引物库的可获得性最小化引物浪费，并且可以立即获得下一测序引物。

人类基因组计划的最初目的之一是到 2005 年完成整个人类基因组的序列测定 (<http://www.nhgri.nih.gov/HGP>)。但是，此计划提前了，在 2001 年 2 月出版了人体基因组的‘工作草图’。由于在几个学科的技术进展，预计在 2003 年完成基因组测序，提前两年。所有这些方面的进展涉及 DNA 操纵(特别是大的 DNA 片段的操纵和增殖)，更快和更好的 DNA 测序方法的出现 (<http://www.abrf.org>)，能操纵和分析所述数据的计算机硬件和软件的发展(生物信息学)，以及与产生和分析 DNA 序列有关的程序的自动化，这些进展是使预定时间提前的原因。

单个分子 DNA 测序

传统的 DNA 测序方法是可靠的，但是耗时，工作量大且费用高。为解决这些问题，一些研究人员研究了基于荧光的，单分子测序方法，所述方法采用酶降解，然后进行单个 dNMP 检测和鉴定。包含荧光标记的核苷酸的所述 DNA 聚合物由核酸外切酶消化，由流式细胞术检测和鉴定标记的核苷酸 (Davis et al., 1991 ; Davis et al., 1992; Goodwin et al., 1997; Keller et al., 1996; Sauer et al., 1999; Werner et al., 1999)。此方法要求合成的 DNA 链包含荧光标记的碱基。此要求限制了能测序的序列长度，增加了得到任何序列数据前进行操作的数量。一种相关的方法提

议从 DNA 的一条链上顺序分开单个(未标记的)核苷酸，按照它们初始的次序限制在固体基质上，并检测分隔的核苷酸的光谱发射以重建 DNA 序列信息 (Ulmer, 1997; Mitsis and Kwagh, 1999; Dapprich, 1999)。这是 Praelux, Inc. 开发的方法，此公司目标是开发单个分子 DNA 测序。理论上，后一方法不象前一方法那样易受到长度的限制，但是在得到任何序列信息前需要多个操作。

Li-cor, Inc. 基于 PCT 申请 WO 00/36151 提出的单分子测序法开发基于酶合成的策略。所述 Li-cor 方法涉及多个修饰的 dNTP，所述修饰通过将荧光标记连接到 γ -磷酸上，淬灭部分 (quenching moiety) 连接到 dNTP 的另一位点上，优选碱基上。加入所述淬灭部分以阻止连接到未整合 dNTP 的荧光标记发射。一旦整合，荧光标记和淬灭部分分离，标记产生发射。所述标记 (包含在所述焦磷酸上) 从聚合酶活性位点移开，但是修饰的 (淬灭的) 碱基成为 DNA 聚合物的一部分。

尽管一些单分子测序系统已经被公开，但它们中的许多预计或需要碱基修饰。参见 WO 01/16375 A2, WO 01/23610 A2, WO 01/25480, WO 00/06770, WO 99/05315, WO 00/60114, WO 00/36151, WO 00/36512, 和 WO 00/70073，在此引入作为参考。碱基修饰可能扭曲 DNA 结构 (其通常由 A-型 DNA 构成，该 A-型 DNA 最接近酶的活性位点；Li et al., 1998a)。由于 dNTP 和新合成的链中 3'-最近的约 7 个碱基与聚合酶的内部区接触 (Li et al., 1998a)，对于最大化酶和 DNA 与小沟之间的接触，A-型 DNA 是重要的。如果由于碱基修饰影响 DNA 结构，酶忠实性和/或功能可能改变。因此，本领域仍需要快速并有效的用于单个分子 DNA 序列的酶 DNA 测序系统。

发明概述

单分子测序

本发明提供了由至少一种分子或原子标记修饰的聚合剂，所述标记位于或靠近，连接于或共价连接到聚合剂上的位点，其中标记的可检测性质在单体整合之前，之中和/或之后经历变化。所述单体可以是有机的，无机的或生物有机单体如 DNA, RNA, 或混合的 DNA/RNA 序列的核苷酸，氨基酸，单糖，天然存在的核苷酸的合成类似物，天然存在的氨基酸的合成类似物，或天然存在的单糖的合成类似物，合成的有机或无机单体，或类似物。

本发明提供了解聚剂，所述试剂用至少一个分子或原子标记修饰，所述标记位于或靠近，与之相连或共价连接到解聚剂上的位点，其中所述标记在除去单体之前，之中和/或之后，其可检测性质经历变化。所述聚合物可以是 DNA, RNA, 混合 DNA/RNA 序列，其只包含天然存在的核苷酸或天然存在的核苷酸和其合成类似物的混合物，多肽序列，其只包含天然存在的氨基酸或天然氨基酸和其合成类似物的混合物，多糖或糖类序列，其只包含天然存在的单糖或天然存在的单糖与其合成类似物的混合物，或包含合成的有机或无机单体的聚合物，或类似的。

本发明还提供了能检测信号的系统，所述信号对应可检测性质，所述性质可证明在合成剂/聚合剂或解聚剂(分子)和其基体(单体或可解聚的聚合物)之间的相互作用，所述系统还能解码信号为单体顺序特异的信息或单体序列信息，优选在实时或接近实时的情况下。

单个位点标记的聚合酶

本发明提供了用至少一种分子或原子标记修饰的聚合酶，所述标记位于或靠近，与之相连，或共价连接到聚合酶的位点，其中在单体整合之前，之中和/或之后，所述标记的可检测性质经历改变。所述单体可以是 DNA, RNA 的核苷酸或混合的 DNA/RNA 单体或通过聚合酶可聚合的合成的类似物。

本发明提供了用至少一种分子或原子标记修饰的外切核酸酶，所述标记位于或靠近，与之相连，或共价连接到外切核酸酶的位点，其中在释放单体之前，之中和/或之后，所述标记的可检测性质经历改变。所述聚合物可以是 DNA, RNA 或混合的 DNA/RNA 序列，所述序列包含通过外切核酸酶可解聚的天然存在的单体或合成的类似物。

本发明提供了用至少一种分子或原子标记修饰的聚合酶，所述标记位于或靠近，与之相连，或共价连接到一位点，所述位点在单体整合之前，之中和/或之后经历构象变化，其中所述标记在聚合酶处于第一构象状态时具有第一可检测倾向，在聚合酶处于第二构象状态时具有第二可检测倾向。

本发明提供了用至少一种发色团修饰的聚合酶，所述发色团位于或靠近，与之相连，或共价连接到一位点，所述位点在单体整合之前，之中和/或之后经历构象变化，其中在聚合酶处于第一构象状态时发色团的发射光强度和/或频率具有第一值，在聚合酶处于第二构象状态时发色团的发射光

强度和/或频率具有第二值。

本发明提供了用至少一种荧光活性分子标记修饰的聚合酶，所述标记位于或靠近，与之相连，或共价连接到一位点，所述位点在单体整合之前，之中和/或之后经历构象变化，其中所述标记在聚合酶处于第一构象状态时具有第一荧光倾向，在聚合酶处于第二构象状态时具有第二荧光倾向。

本发明提供了用分子标记修饰的聚合酶，所述标记位于或靠近，与之相连，或共价连接到一位点，所述位点在单体整合之前，之中和/或之后经历构象变化，其中在聚合酶处于第一构象状态时所述标记是可充分被检测的，当在聚合酶处于第二构象状态时所述标记是基本检测不到的，或者在聚合酶处于第一构象状态时所述标记是基本检测不到的，当在聚合酶处于第二构象状态时所述标记是可充分被检测的。

本发明提供了用至少一个分子或原子标记修饰的聚合酶，所述标记位于或靠近，与之相连，或共价连接到一位点，所述位点与释放的焦磷酸基上的标记相互作用，其中所述聚合酶标记在与释放的焦磷酸基上的标记相互作用前具有第一检测倾向，当与释放的焦磷酸基上的标记相互作用时具有第二检测倾向。在优选的实施方案中，在每一焦磷酸基释放时，此检测倾向的变化循环发生。

本发明提供了用至少一个发色团修饰的聚合酶，所述发色团位于或靠近，与之相连，或共价连接到一位点，所述位点与释放的焦磷酸基上的标记相互作用，其中在发色团与释放的焦磷酸基上的标记相互作用前发色团发射的光强度和/或频率具有第一值，当与释放的焦磷酸基上的标记相互作用时具有第二值。在优选的实施方案中，每一焦磷酸基释放时此检测倾向的变化循环发生。

本发明提供了用至少一个荧光活性分子标记修饰的聚合酶，所述标记位于或靠近，与之相连，或共价连接到一位点，所述位点与释放的焦磷酸基上的标记相互作用，其中所述聚合酶标记在焦磷酸基释放前从第一状态变化，并且在所述焦磷酸基从释放位点扩散开时从第二种状态变化。在优选的实施方案中，每一焦磷酸基释放时此检测倾向的变化循环发生。

本发明提供了用分子标记修饰的聚合酶，所述标记位于或靠近，与之相连，或共价连接到一位点，所述位点与释放的焦磷酸基上的标记相互作用，其中所述聚合酶标记从在焦磷酸基释放前的可充分检测状态变化为释

放所述基团后聚合酶标记与焦磷酸基上的标记相互作用时的基本检测不到的状态，或者从焦磷酸基释放前的基本检测不到的状态变化为释放所述基团后聚合酶标记与焦磷酸基上的标记相互作用时的可充分检测状态。

多个位点标记的聚合剂或解聚剂

本发明提供了用至少一对分子和/或原子标记修饰的单体聚合试剂，所述标记位于或靠近，与之相连，或共价连接到聚合剂上的位点，其中所述标记对的至少一个标记的可检测性质在单体整合之前，之中和/或之后发生改变，或者其中所述标记对的至少一个标记的可检测性质由于内标记的相互作用在单体整合之前，之中和/或之后发生变化。

本发明提供了用至少一对分子和/或原子标记修饰的解聚剂，所述标记位于或靠近，与之相连，或共价连接到解聚剂上的位点，其中所述标记对的至少一个标记的可检测性质在单体释放之前，之中和/或之后发生改变，或者其中所述标记对的至少一个标记的可检测性质由于内标记的相互作用在单体释放之前，之中和/或之后发生变化。

本发明提供了用至少一对分子和/或原子标记修饰的单体聚合试剂，所述标记位于或靠近，与之相连，或共价连接到聚合剂上的位点，其中所述标记对的至少一个标记的可检测性质在所述聚合剂处于第一种状态时具有第一个值，且在所述聚合剂处于第二种状态时具有第二个值，其中在单体整合循环中所述聚合剂从第一种状态变为第二种状态再变回第一种状态。

本发明提供了用至少一对分子和/或原子标记修饰的单体解聚剂，所述标记位于或靠近，与之相连，或共价连接到解聚剂上的位点，其中所述标记对的至少一个标记的可检测性质在所述解聚剂处于第一种状态时具有第一个值，且在所述解聚剂处于第二种状态时具有第二个值，其中在单体释放循环中所述解聚剂从第一种状态变为第二种状态再变回第一种状态。

优选地，所述第一种状态和第二种状态是不相同的，以使检测信号发生变化。但是，未变化的结果可证实聚合介质(media)或解聚介质的其它性质。

多个位点标记的聚合酶

本发明提供了用至少一对分子标记修饰的聚合酶，所述标记位于或靠

近，与之相连，或共价连接到位点上，所述标记中的至少一个在单体整合过程中经历变化，其中聚合酶处于第一种状态时所述标记对的可检测性质具有第一个值，且在所述聚合酶处于第二种状态时可检测性质具有第二个值，其中在单体整合循环中所述聚合酶从第一种状态变为第二种状态再变回第一种状态。

本发明提供了用至少一对分子标记修饰的聚合酶，所述标记位于或靠近，与之相连，或共价连接到位点上，所述标记中的至少一个在单体整合过程中经历构象变化，其中聚合酶处于第一种构象状态时所述标记对的可检测性质具有第一个值，且在所述聚合酶处于第二种构象状态时可检测性质具有第二个值，其中在单体整合循环中所述聚合酶从第一种构象状态变为第二种状态再变回第一种构象状态。

本发明提供了用至少一对分子或原子修饰的聚合酶，其位于或靠近，与之相连，或共价连接到位点上，所述标记中的至少一个在单体整合过程中经历构象变化，其中聚合酶处于第一种构象状态或第二种状态时所述标记对相互作用形成生色团，其中在单体整合循环中所述聚合酶从第一种状态变为第二种状态再变回第一种状态。

本发明提供了用至少一对分子标记修饰的聚合酶，所述标记位于或靠近，与之相连，或共价连接到位点上，所述标记中的至少一个在单体整合过程中经历构象变化，其中在所述聚合酶处于第一种构象状态时所述标记具有第一种荧光倾向，在所述聚合酶处于第二种构象状态时所述标记具有第二种荧光倾向，其中在单体整合循环中所述聚合酶从第一种构象状态变为第二种状态再变回第一种构象状态。

本发明提供了用至少一对分子标记修饰的聚合酶，所述标记位于或靠近，与之相连，或共价连接到位点上，所述标记中的至少一个在单体整合过程中经历构象变化，其中在所述聚合酶处于第一种构象状态时所述标记对是基本活化的，在所述聚合酶处于第二种构象状态时所述标记是基本不活化的，或者在所述聚合酶处于第一种构象状态时是基本不活化的，在所述聚合酶处于第二种构象状态时所述标记是基本活化的，其中在单体整合循环中所述聚合酶从第一种构象状态变为第二种状态再变回第一种构象状态。

本发明提供了用至少一对分子标记修饰的聚合酶，所述标记位于或靠

近，与之相连，或共价连接到位点上，在单体整合过程中在焦磷酸释放之中和/或之后所述标记中的至少一个经历变化，其中在所述焦磷酸释放之前所述标记处于第一种状态时所述标记对的可检测性质具有第一个值，在所述焦磷酸释放之中和/或之后所述标记处于第二种状态时具有第二个值，其中在单体整合循环中所述标记从第一种状态变为第二种状态再变回其第一种状态。

本发明提供了用至少一对分子标记修饰的聚合酶，所述标记位于或靠近，与之相连，或共价连接到位点上，由于在焦磷酸释放过程中所述聚合酶的构象变化所述标记中的至少一个经历位置变化，其中当所述标记位于其第一个位置时所述标记对的可检测性质具有第一个值，当所述标记位于其第二个位置时具有第二个值，其中在释放循环中所述标记从第一个位置变到第二个位置再变回其第一个位置。

本发明提供了用至少一对分子或原子修饰的聚合酶，其位于或靠近，与之相连，或共价连接到位点上，其中由于在焦磷酸释放过程中所述聚合酶的构象变化所述标记改变相对间隔，其中所述标记相互作用形成的生色团在所述标记的第一种间隔距离时具有第一种发射图谱，当在所述标记的第二种间隔距离时具有第二种发射图谱，在焦磷酸释放循环中所述间隔距离从其第一种状态变为其第二种状态再变回其第一种状态。

本发明提供了用至少一对分子标记修饰的聚合酶，所述标记位于或靠近，与之相连，或共价连接到位点上，其中由于在焦磷酸释放过程中所述聚合酶的构象变化所述标记改变相对间隔，其中当所述聚合酶处于第一种构象状态时所述标记具有第一种荧光倾向，当所述聚合酶处于第二种构象状态时所述标记具有第二种荧光倾向，其中在焦磷酸释放循环中所述倾向从其第一个值变为第二个值再变回其第一个值。

本发明提供了用至少一对分子标记修饰的聚合酶，所述标记位于或靠近，与之相连，或共价连接到位点上，其中由于在焦磷酸释放过程中所述聚合酶的构象变化所述标记改变相对间隔，其中当所述标记具有第一种间隔时所述标记对是充分荧光活性的，当所述标记具有第二种间隔时所述标记对基本是无荧光活性的，或者当所述标记具有第一种间隔时所述标记对基本是无荧光活性的，当所述标记具有第二种间隔时所述标记对是充分荧光活性的，其中在焦磷酸释放循环中所述荧光活性经历一个循环。

应该认识到当性质从第一种状态改变为第二种状态再变回第一种状态时，所述性质经历一个循环。优选地，第一种和第二种状态是不同的，这样所述检测信号可以发生变化。但是，没有变化的结果可以证明聚合介质或解聚介质的其它性质。

采用标记的聚合剂的方法

本发明提供了确定单体何时整合到增长的分子链中的方法，其包含监测原子或分子标记的可检测性质的步骤，其中所述标记位于或靠近，与之相连，或共价连接到聚合剂上的位点，其中所述标记的可检测性质在单体整合之前，之中和/或之后经历改变。

本发明提供了确定单体何时整合到增长的分子链的方法，其包含监测原子或分子标记的可检测性质的步骤，其中所述标记位于或靠近，与之相连，或共价连接到聚合剂上的位点，其中所述标记的可检测性质在所述聚合剂处于第一种状态时具有第一个值，在所述聚合剂处于第二种状态时具有第二个值，其中在单体整合循环中，所述试剂从第一种状态变化为第二种状态再变回第一种状态。

优选地，第一和第二种状态是不同的，这样检测信号可以发生变化。但是，没有变化的结果可以证明聚合介质的其它性质。

采用标记的聚合酶的方法

本发明提供了确定单体何时或是否整合到增长的分子链的方法，其包含监测标记的可检测性质的步骤，其中所述标记位于或靠近，与之相连，或共价连接到聚合酶上的位点，其中在单体整合过程中所述位点经历变化，且当聚合酶处于第一种状态时所述可检测性质具有第一个值，当所述聚合酶处于第二种状态时具有第二个值，其中这些值表示所述位点经历改变，并且在单体整合循环中所述聚合酶从第一种状态变化为第二种状态再变回第一种状态。

本发明提供了确定单体何时或是否整合到增长的分子链的方法，其包含监测标记的可检测性质的步骤，其中所述标记位于或靠近，与之相连，或共价连接到聚合酶上的位点，其中在单体整合过程中所述位点经历构象

变化，其中在所述聚合酶处于第一种构象状态时所述可检测性质具有第一个值，当所述聚合酶处于第二种构象状态时具有第二个值，其中所述值表示所述位点经历改变，并且在单体整合循环中所述聚合酶从第一种状态变化为第二种状态再变回第一种状态。

本发明提供了确定单体何时或是否整合到增长的分子链的方法，其包含将标记的聚合酶暴露于光，监测所述标记的聚合酶发射的荧光强度和/或频率的步骤，其中所述标记的聚合酶包含聚合酶，其包括的标记位于或靠近，与之相连，或共价连接到位点，所述位点在单体整合过程中经历构象变化，且其中当所述聚合酶处于第一种构象状态时所述标记以第一种强度和/或频率发射荧光，当所述聚合酶处于第二种构象状态时以第二种强度和/或频率发射荧光，其中所述强度和/或频率的变化表示所述位点经历变化且在单体整合循环中所述聚合酶从第一种状态变化为第二种状态再变回第一种状态。

本发明也提供了采用多个标记的聚合酶的上述方法，其可以同时并行和/或整体上并行测序。这样的平行性可用于确保置信度。这样的平行性也可用于对给定的跨种类基因 (gene across species) 快速检测 DNA 序列的同源性程度，或用于快速筛选患者 DNA 的特异性基因特征或用于快速筛选 DNA 序列多态性。

本发明提供了确定单体何时或是否整合到与聚合酶有关的增长的 DNA 链上，其中有标记位于所述聚合酶上以使当碱基整合后焦磷酸基释放时，在其扩散离开所述聚合酶之前，所述聚合酶标记与焦磷酸上的标记相互作用引起所述标记之一的可检测性质或在荧光对的情况下与两个标记都有关的可检测性质的变化。

优选地，第一和第二种状态不同，以使检测信号发生变化。但是，没有变化的结果可能证明所述聚合介质的其它特性。

采用标记的聚合剂的设备

本发明提供了单分子测序装置，其包含其上沉积了至少一种标记的聚合剂的基片 (substrat)。所述标记的聚合剂可被放置在合适的聚合介质中的基片表面或所述聚合剂可被限制在基片上的区域，范围 (area)，孔 (well)，凹槽，通道或其它类似的结构内。所述基片也能包括单体区域，

范围(area)，孔(well)，凹槽，通道，容器(reservoir)或基体上的其它类似的结构，其通过能支持单体分子运输到聚合剂的至少一个连接结构如通道，凹槽，或类似的连接到聚合剂限制结构(confinement structure)。备选地，所述基体能包括包含每一单体的结构，其中每一结构通过能支持单体分子运输到聚合剂的连接结构连接到聚合剂限制结构。所述基体也能细分为多种聚合剂限制结构，其中每一结构连接到单体容器。备选地，每一聚合剂限制结构能具有其自己的单体容器或足够的单体容器以使每一容器包含特异性单体。

本发明也提供了单分子测序装置，其包含基体，所述基体具有通过分子系链(molecular tether)或连接基团连接到所述基体表面的至少一个标记的聚合剂，其中所述系链或连接基团的一端结合到(bond to)基片表面上的位点上，另一端结合到聚合剂的位点上或与所述聚合剂坚固连接的分子上的位点上。在此处，术语'结合到'意指化学和/或物理相互作用，其在通常的聚合条件下足以保持聚合剂在给定的基体区域内。所述的化学和/或物理相互作用包括，但不限于，共价结合(covalent bonding)，离子结合，氢键结合，非极性结合，静电吸引，偶极间相互作用，或其它任何足以完全保持聚合剂在所需的基体区域内电子或量子力学的相互作用。其上连接有被束缚的标记的聚合剂的所述基体可以放置在容器中，所述容器包括合适的聚合介质。备选地，所述标记的聚合剂可被束缚或锚定在基体上的区域，范围，孔，凹槽，通道或其它类似的结构之上或之内，它们可以被合适的聚合介质填充。所述基体也能包括基体上的单体区域，范围，孔，凹槽，通道或其它类似的结构，它们可与聚合剂结构通过至少一种能支持单体分子运输到所述聚合剂的连接结构连接。备选地，所述基体包括包含每一单体的结构，其中每一结构通过至少一种能支持单体分子运输到所述聚合剂的连接结构连接到聚合剂结构。所述基体还可被细分为多种聚合剂结构，每一结构具有至少一种被束缚的聚合剂，其中每一结构连接到单体贮存器(monomer reservoir)。备选地，每一聚合剂结构可具有其自己的单体贮存器或足够多的单体贮存器，每一特异性单体都有一个储存器。

这些装置所用的单体包括但不限于，dNTP，标记的dNTP，ddNTP，标记的ddNTP，氨基酸，标记的氨基酸，单糖，标记的单糖或根据待测序聚合物的类型其合适的混合物或组合。

采用标记的聚合酶的装置

本发明提供了单分子测序装置，其包含基体，所述基体具有至少一种标记的聚合酶沉积其上。所述标记的聚合酶能被放置在合适的聚合介质中，基体的表面，或者所述聚合酶可被限制在基体上的区域，范围，孔，凹槽，通道或其它类似的结构内，它们可以被合适的聚合介质填充。所述基体也能包括基体上的单体区域，范围，孔，凹槽，通道或其它类似的结构，它们可与聚合剂限制结构通过至少一种能支持单体分子运输到所述聚合酶的连接结构连接。备选地，所述基体包括包含每一单体的结构，其中每一结构通过至少一种能支持单体分子运输到所述聚合酶的限制结构中的聚合酶的连接结构连接到聚合酶限制结构。所述基体还可被细分为多种聚合酶限制结构，其中每一结构都与单体贮存器连接。备选地，每一聚合酶限制结构可具有其自己的单体贮存器或4个单体贮存器，每一个储存器包括特异性单体。

本发明也提供了单分子测序装置，其包含基体，所述基体具有通过分子系链或连接基团连接到所述基体表面的至少一种标记的聚合酶，其中所述系链或连接基团的一端结合到基体的表面，另一端结合到(直接地或间接地)所述聚合酶的位点上，或与所述聚合酶坚固连接的分子上的位点上。在此处，术语'结合到'意指化学和/或物理相互作用，其在通常的聚合条件下足以保持聚合酶在给定的基体区域内。所述的化学和/或物理相互作用包括，但不限于，共价结合，离子结合，氢键结合，非极性结合，静电吸引，偶极间相互作用，或其它任何足以完全保持聚合酶在所需的基体区域内的电子或量子力学相互作用。具有被束缚的标记的聚合剂连接其上的所述基体可以放置在容器中，所述容器包括合适的聚合介质。备选地，它们的聚合剂可被束缚或锚定在基体上的区域，范围，孔，凹槽，通道或其它类似的结构内，它们可以被合适的聚合介质填充。所述基体也能包括基体上的单体区域，范围，孔，凹槽，通道或其它类似的结构，它们可通过至少一通道连接到聚合酶结构。备选地，所述基体包括包含每一单体的结构，其中每一结构通过至少一种能支持单体分子运输到在聚合酶限制结构中的聚合酶的连接结构连接到聚合酶结构。所述基体可被细分为多种聚合酶结构，每一结构具有至少一种束缚的聚合酶，其中每一结构连接到单体贮存器。

备选地，每一聚合酶结构能具有其自己的单体贮存器或 4 个贮存器，每一储存器包含特异性单体。

这些装置所用单体包括但不限于，dNTP，标记的 dNTP，ddNTP，标记的 ddNTP，或其混合物或组合。

采用单分子测序装置的方法

本发明提供了用于单分子测序的方法，其包含步骤为：提供多种单体给限制或束缚于基体的标记的聚合剂，监测所述标记可检测性质随时间的变化。该方法也包括将可检测性质与发生单体添加(计时)和/或每一整合的单体的身份和/或对整合单体的序列的近乎同时的确定相联系。

本发明提供了用于单分子测序的方法，其包含步骤为：提供多种单体给限制或束缚于基体的标记的聚合剂，将所述标记的聚合剂持续或间歇暴露于光，测定所述标记发射的荧光强度和/或频率随时间的变化。该方法可进一步包含将测定的所述标记发射的荧光强度和/或频率随时间的变化与发生单体添加(计时)和/或每一整合的单体的身份和/或对整合单体的序列的近乎同时的确定相联系。

本发明提供了用于单分子测序的方法，其包含步骤为：提供多种单体给限制或束缚于基体的标记的聚合酶，监测所述标记可检测性质随时间的变化。该方法也包括将可检测性质与发生单体添加(计时)和/或每一整合的单体的身份和/或对整合单体的序列的近乎同时的确定相联系。

本发明提供了用于单分子测序的方法，其包含步骤为：提供多种单体给限制或束缚于基体的标记的聚合酶，将所述标记的聚合酶持续或间歇暴露于光，测定所述标记发射的荧光强度和/或频率随时间的变化。该方法可进一步包含将测定的所述标记发射的荧光强度和/或频率随时间的变化与发生单体添加(计时)和/或每一整合的单体的身份和/或对整合单体的序列的近乎同时的确定相联系。

协同标记系统

本发明提供了协同标记的聚合剂和标记的单体，其中所述至少一个标记的可检测性质在所述标记在单体插入之前，之中和/或之后相互作用时发生改变。在一个优选的实施方案中，定位聚合酶上的所述标记，以使所述

标记在单体插入之前，之中和/或之后相互作用。在单体插入之后从所述单体释放的标记如， β 和/或 γ 磷酸标记的 dNTP，即位于 β 和/或 γ 磷酸基上的标记的情况下，所述聚合剂上的标记可被设计为：只在所述标记从所述聚合剂释放后与所述单体上的标记相互作用，其中所述释放在单体插入后。可以通过将所述聚合酶标记连接到聚合酶上的位点，使所述位点在整合发生过程中发生移动，从而改变所述两标记间的相对间隔，优化聚合剂内的标记的放置来增强所述聚合酶标记和 dNTP 标记之间的相互作用，或在碱基整合过程中从所述聚合剂释放时和扩散离开所述聚合剂之前，优化聚合剂内的标记的放置，来增强所述聚合酶标记和所述焦磷酸上的标记的相互作用。

本发明提供了协同标记的聚合剂和标记的单体，其中当所述标记在足以引起可检测性质的可测量改变的距离内时，所述至少一个标记的可检测性质发生改变。如果可检测性质是通过能量传递到其它标记或由于一个标记淬灭了另一标记的荧光或引起了荧光强度和/或频率的可测量的变化，从而在一种标记中诱发了荧光，那么通过使所述标记相互接近，即缩短各标记的间隔距离可引起可测量的改变。通常，引起所述可检测性质的可测量的改变所需的距离在约 100\AA 之内(小于或等于)，优选在 50\AA 之内，具体为 25\AA 之内，具体为 15\AA 之内，最优选在 10\AA 之内。当然，本领域技术人员可以认识到足以引起标记的可检测性质的可测量改变的距离有赖于许多参数，包括标记的位置，标记的性质，溶剂系统，外部环境，激发源强度和频率波段的宽度，温度，压力等。

本发明提供了标记的聚合剂和一个或多个标记的单体前体，其中在单体插入之前，之中和/或之后所述标记相互作用时，至少一个标记发射的荧光强度和/或频率改变。

本发明提供了协同标记的解聚剂和标记的可解聚聚合物，其中在单体释放之前，之中和/或之后所述标记相互作用时，至少一个标记的可检测性质发生改变。解聚剂上的标记可被设计为在每一单体释放之前，之中和/或之后相互作用。

本发明提供了协同标记的解聚剂和标记的聚合物，其中当所述标记在足以引起所述可检测性质的可测量的改变的距离内时，所述至少一个标记的可检测性质发生改变。如果所述的可检测性质是通过能量传递到其它标

记或由于一个标记淬灭了另一标记的荧光或引起了荧光强度和/或频率的可测量的变化，从而在一个标记中诱发了荧光，那么通过使所述标记相互接近，即缩短各标记的间隔距离可引起可测量的改变。通常，引起所述可检测性质的可测量的改变所需的距离在约 100\AA 之内(小于或等于)，优选在 50\AA 之内，具体为 25\AA 之内，具体为 15\AA 之内，最优选在 10\AA 之内。当然，本领域技术人员可以认识到足以引起标记的可检测性质的可测量改变的距离有赖于许多参数，包括标记的位置，标记的性质，溶剂系统，外部环境，激发源强度和频率波段的宽度，温度，压力等。

本发明提供了标记的解聚剂和标记的聚合物，其中在单体释放之前，之中和/或之后所述标记相互作用时，至少一个标记发射的荧光强度和/或频率改变。

采用聚合酶的协同标记系统

本发明提供了协同标记的聚合酶和标记的单体，其中所述至少一个标记的可检测性质在所述标记在单体插入之前，之中和/或之后相互作用时发生改变。聚合酶上的所述标记可被设计为所述标记在单体插入之前，之中和/或之后相互作用。在单体插入之后从所述单体释放的标记如， β 和/或 γ 磷酸标记的 dNTP，即位于 β 和/或 γ 磷酸基上的标记的情况下，所述聚合剂上的标记可被设计为：只在所述标记从所述聚合剂释放后与所述单体上的标记相互作用，其中所述释放在单体插入后。在第一种情况下，所述聚合酶的标记必须位于聚合酶的位点上，该位点在单体插入过程中一开始结合和键合到增长的聚合物中时使所述聚合酶标记与单体标记相互作用。而在第二种情况下，所述聚合酶标记必须位于聚合酶位点上，在焦磷酸扩散离开聚合酶进入聚合介质之前，该位点使聚合酶标记与现位于焦磷酸上的单体标记相互作用。

本发明提供了协同标记的聚合酶和标记的单体，其中当所述标记在足以引起所述可检测性质的可测量改变的距离内时，所述至少一个标记的可检测性质发生改变。如果所述的可检测性质是通过能量传递到其它标记或由于一个标记淬灭了另一标记的荧光或引起了荧光强度和/或频率的可测量的变化，从而在一个标记中诱发了荧光，那么通过使所述标记相互接近，即缩短各标记的间隔距离可引起可测量的改变。通常，所述的距离或极接

近的距离在约 100Å 到约 10Å 之间。备选地，所述距离为小于或等于约 100Å，优选小于或等于约 50Å，具体为小于或等于约 25Å，具体为小于或等于约 15Å，最优先小于或等于约 10Å。当然，本领域技术人员可以认识到足以引起标记的可检测性质的可测量改变的距离有赖于许多参数，包括标记的位置，标记的性质，溶剂系统(聚合介质)，外部环境，激发源强度和频率波段的宽度，温度，压力等。

本发明提供了标记的聚合酶和标记的单体前体，其中所述标记形成荧光活性对如供体-受体对，并且当标记相互作用时，至少一个标记(通常为供体-受体对中的受体标记)发射的荧光的强度和/或频率发生改变。

本发明提供了标记的聚合酶和标记的单体前体，其中所述标记形成荧光活性对如供体-受体对，并且当标记间的距离足以或接近改变所述荧光的或者是强度和/或频率时，至少一个标记(通常为供体-受体对中的受体标记)发射的荧光的强度和/或频率发生改变。通常，所述的距离或极接近的距离在约 100Å 到约 10Å 之间。备选地，所述距离为小于或等于约 100Å，优选小于或等于约 50Å，具体为小于或等于约 25Å，具体为小于或等于约 15Å，最优先小于或等于约 10Å。当然，本领域技术人员可以认识到足以引起标记的可检测性质的可测量改变的距离有赖于许多参数，包括标记的位置，标记的性质，溶剂系统，外部环境，激发源强度和频率波段的宽度，温度，压力等。

本发明提供了单分子测序系统，其包含一容器，其中具有限制或束缚在其内表面上的至少一种标记的聚合酶，并具有包含多种与内表面接触的标记的单体的溶液。

分子数据流读取方法和装置

本发明提供了单分子测序方法，包括如下步骤：将多种标记的单体提供给限制在容器内表面上的标记的聚合酶，暴露所述标记的聚合酶于光，测定在每一后继的单体添加或插入到增长的聚合物链过程中所述标记的聚合酶发射的荧光的强度和/或频率。该方法还进一步包含将测定的发射的荧光强度和/或频率与整合的发生和/或每一插入或添加的单体的鉴定相联系，这得到对增长的核酸序列-DNA 序列，RNA 序列或混合的 DNA/RNA 序列的接近实时或实时的序列读出。

本发明提供了重现储存的信息的系统，其包含具有代表数据流的已知元件(element)的序列的分子，单分子测序仪，其包含聚合酶，该酶具有至少一个与之连接的标记，适于激发聚合酶上至少一个标记的激发源，适于检测聚合酶上被激发的标记的响应的检测器，其中所述来自至少一个标记的所述响应在元件的互补序列聚合过程中变化，且所述变化代表数据流的内容。

本发明提供了从单个分子测定序列信息的系统，包含具有已知元件的序列的分子，单分子测序仪，其包含聚合酶，该酶有至少一个与之连接的标记，适于激发聚合酶上至少一个标记的激发源，适于检测聚合酶上被激发的标记的响应的检测器，其中所述来自至少一个标记的所述响应在元件的互补序列聚合过程中变化，所述元件的互补序列代表分子的元件序列。

本发明提供了从单个分子测定序列信息的系统，包含具有已知元件的序列的分子，单分子测序仪，其含有连接了至少一个荧光标记的聚合酶，适于激发聚合酶和/或单体上的至少一个荧光标记的激发光源，适于检测所述聚合酶和/或单体上至少一个荧光标记发射的荧光的至少一种荧光强度检测器，其中每次新的核苷酸或核苷酸类似物聚合到互补序列时所述信号强度发生变化，发射的持续时间或缺少发射或发射光的波长范围证明聚合到序列中的具体的核苷酸或核苷酸类似物，以使测序的数据流完整重现。

本发明提供了储存和重现数据的系统，包含代表给定的数据流的核苷酸或核苷酸类似物的序列；单分子测序仪，其包含共价连接了至少一个荧光标记的聚合酶；适于激发聚合酶和/或单体上至少一个荧光标记的激发光源；适于检测所述聚合酶和/或单体上的至少一个荧光标记发射的荧光的荧光检测器，其中在每次新的核苷酸或核苷酸类似物聚合到互补序列时至少一个荧光标记发射或不发射荧光，并且发射的持续时间或缺少发射或发射光的波长范围证明聚合到序列的具体的核苷酸或核苷酸类似物，以使在完整测序时重现数据流。

此处所用术语单体意指通过给定的聚合酶能整合到增长的分子链的任何化合物。这样的单体包括但不限于，天然存在的核苷酸(例如，ATP, GTP, TTP, UTP, CTP, dATP, dGTP, dTTP, dUTP, dCTP, 合成的类似物)，每种核苷酸的前体，非天然存在的核苷酸和它们的前体或任何可借助给定的聚合酶整合到增长的聚合链中的其它分子。此外，用于蛋白或蛋白类似物合

成的氨基酸(天然的或合成的), 用于糖类合成的单糖或其它单体的综合体(syntheses)。

本文所用术语聚合酶意指任何分子或分子组合, 其能聚合一组单体到具有预先确定的单体序列的聚合物中, 包括但不限于天然存在的聚合酶或逆转录酶, 突变的天然聚合酶或逆转录酶, 其中所述突变包含用其它氨基酸取代一个或多个或许多氨基酸, 所述聚合酶或逆转录酶的一个或多个或许多氨基酸的插入或缺失, 或一个或多个聚合酶或逆转录酶, 非天然存在的聚合酶或逆转录酶的部分偶联。所述术语聚合酶也包含合成的分子或分子组合, 其能聚合聚合物, 所述聚合物具有预先确定的单体序列, 或任何其它分子或分子组合, 其具有额外的序列以利于标记的纯化和/或固定和/或分子相互作用, 所述分子或分子组合能聚合聚合物, 所述聚合物具有预先确定的或限定的或模板的单体序列。

单个位点标记的聚合剂或解聚剂

本发明提供了包含聚合剂的组合物, 所述聚合剂包括至少一种分子和/或原子标记, 所述标记位于或靠近, 与之相连或共价连接到所述试剂上的位点, 其中所述标记的可检测性质在单体整合之前, 之中和/或之后经历改变。

本发明提供了包含聚合剂的组合物, 所述聚合剂包括至少一种分子和/或原子标记, 所述标记位于或靠近, 与之相连或共价连接到所述试剂上的位点, 其中所述标记的可检测性质在单体聚合过程中当所述聚合酶处于第一种状态时具有第一个值, 当所述聚合酶处于第二种状态时具有第二个值。

本发明提供了包含解聚剂的组合物, 所述解聚剂包括至少一种分子和/或原子标记, 所述标记位于或靠近, 与之相连或共价连接到所述试剂上的位点, 其中所述标记的可检测性质在除去单体之前, 之中和/或之后经历改变。

本发明提供了包含聚合剂的组合物, 所述聚合剂包括至少一种分子和/或原子标记, 所述标记位于或靠近, 与之相连或共价连接到所述试剂上的位点, 其中所述标记的可检测性质在除去单体过程中当所述聚合酶处于第一种状态时具有第一个值, 当所述聚合酶处于第二种状态时具有第二个值。

单个位点标记的聚合酶

本发明提供了包含聚合酶的组合物，所述聚合酶包括至少一种分子和/或原子标记，所述标记位于或靠近，与之相连或共价连接到所述聚合酶上的位点，其中所述标记的可检测性质在单体整合之前，之中和/或之后经历改变。

本发明提供了包含聚合酶的组合物，所述聚合酶包括至少一种分子和/或原子标记，所述标记位于或靠近，与之相连或共价连接到所述聚合酶上的位点，其中所述标记的可检测性质在单体整合过程中当所述聚合酶处于第一种状态时具有第一个值，当所述聚合酶处于第二种状态时具有第二个值。

本发明提供了包含外切核酸酶的组合物，所述外切核酸酶包括至少一种分子和/或原子标记，所述标记位于或靠近，与之相连或共价连接到所述试剂上的位点，其中所述标记的可检测性质在除去单体之前，之中和/或之后经历改变。

本发明提供了包含外切核酸酶的组合物，所述外切核酸酶包括至少一种分子和/或原子标记，所述标记位于或靠近，与之相连或共价连接到所述试剂上的位点，其中所述标记的可检测性质在除去单体过程中当所述聚合酶处于第一种状态时具有第一个值，当所述聚合酶处于第二种状态时具有第二个值。

本发明提供了包含酶的组合物，所述酶被修饰以在与合适的修饰的单体作用之前，之中，和/或之后产生可检测的响应，其中所述单体是核苷酸，核苷酸类似物，氨基酸，氨基酸类似物，单糖，单糖类似物或其混合物或结合。

本发明提供了包含聚合酶的组合物，所述聚合酶包括至少一个分子标记，所述标记位于或靠近，与之相连或共价连接到位点，在单体整合过程中所述位点经历构象改变，其中当所述聚合酶处于第一种构象状态时所述标记具有第一种检测倾向，当所述聚合酶处于第二种构象状态时具有第二种检测倾向。

本发明提供了包含聚合酶的组合物，所述聚合酶包括至少一个生色团，所述生色团位于或靠近，与之相连或共价连接到位点，在单体整合过程中所述位点经历改变，其中当所述聚合酶处于第一种构象状态时所述标记发

射光的强度和/或频率具有第一个值，当所述聚合酶处于第二种构象状态时具有第二个值。

本发明提供了包含聚合酶的组合物，所述聚合酶包括至少一个分子标记，所述标记位于或靠近，与之相连或共价连接到位点，在单体整合过程中所述位点经历构象改变，其中当所述聚合酶处于第一种构象状态时所述标记具有第一种荧光倾向，当所述聚合酶处于第二种构象状态时具有第二种荧光倾向。

本发明提供了包含聚合酶的组合物，所述聚合酶包括分子标记，所述标记位于或靠近，与之相连或共价连接到位点，在单体整合过程中所述位点经历构象改变，其中当所述聚合酶处于第一种构象状态时所述标记是基本活化的，当所述聚合酶处于第二种构象状态时所述标记是基本失活的，或当所述聚合酶处于第一种构象状态时所述标记是充分失活的，当所述聚合酶处于第二种构象状态时所述标记是基本活化的。

多个位点标记的聚合和解聚剂

本发明提供了包含聚合剂的组合物，所述聚合剂包括至少一对分子标记，所述标记位于或靠近，与之相连或共价连接到所述试剂上的位点，其中所述标记中至少一个的可检测性质在单体整合之前，之中和/或之后经历改变。

本发明提供了包含聚合剂的组合物，所述聚合剂包括至少一对分子标记，所述标记位于或靠近，与之相连或共价连接到所述试剂上的位点，其中所述标记中至少一个的可检测性质在单体聚合过程中当所述聚合酶处于第一种状态时具有第一个值，当所述聚合酶处于第二种状态时具有第二个值。

本发明提供了包含解聚剂的组合物，所述解聚剂包括至少一对分子标记，所述标记位于或靠近，与之相连或共价连接到所述试剂上的位点，其中所述标记中至少一个的可检测性质在除去单体之前，之中和/或之后经历改变。

本发明提供了包含解聚剂的组合物，所述解聚剂包括至少一对分子标记，所述标记位于或靠近，与之相连或共价连接到所述试剂上的位点，其中所述可检测性质在除去单体过程中当所述聚合酶处于第一种状态时具有

第一个值，当所述聚合酶处于第二种状态时具有第二个值。

多个位点标记的聚合酶

本发明提供了包含聚合酶的组合物，所述聚合酶包括至少一对分子标记，所述标记位于或靠近，与之相连或共价连接到所述聚合酶上的位点，其中所述标记中至少一个的可检测性质在单体整合之前，之中和/或之后经历改变。

本发明提供了包含聚合酶的组合物，所述聚合酶包括至少一对分子标记，所述标记位于或靠近，与之相连或共价连接到所述聚合酶上的位点，其中所述标记中至少一个的可检测性质在单体聚合过程中当所述聚合酶处于第一种状态时具有第一个值，当所述聚合酶处于第二种状态时具有第二个值。

本发明提供了包含外切核酸酶的组合物，所述外切核酸酶包括至少一对分子标记，所述标记位于或靠近，与之相连或共价连接到所述聚合酶上的位点，其中所述标记中至少一个的可检测性质在除去单体之前，之中和/或之后经历改变。

本发明提供了包含外切核酸酶的组合物，所述外切核酸酶包括至少一对分子标记，所述标记位于或靠近，与之相连或共价连接到所述聚合酶的位点，其中所述可检测性质在除去单体过程中当所述聚合酶处于第一种状态时具有第一个值，当所述聚合酶处于第二种状态时具有第二个值。

本发明提供了包含聚合酶的组合物，所述聚合酶包括至少一对分子标记，所述标记位于或靠近，与之相连或共价连接到位点，在单体整合过程中所述位点经历构象改变，其中当所述聚合酶处于第一种构象状态时所述标记对的可检测性质具有第一个值，当所述聚合酶处于第二种构象状态时具有第二个值。

本发明提供了包含聚合酶的组合物，所述聚合酶包括至少一对分子标记，所述标记位于或靠近，与之相连或共价连接到位点，在单体整合过程中所述位点经历构象改变，其中在所述聚合酶处于第一种构象状态或第二种构象状态时所述标记对相互作用形成生色团。

本发明提供了包含聚合酶的组合物，所述聚合酶包括至少一对分子标记，所述标记位于或靠近，与之相连或共价连接到位点，在单体整合过程

中所述位点经历构象改变，其中当所述聚合酶处于第一种构象状态时所述标记具有第一种荧光倾向，当所述聚合酶处于第二种构象状态时具有第二种荧光倾向。

本发明提供了包含聚合酶的组合物，所述聚合酶包括至少一对分子标记，所述标记位于或靠近，与之相连或共价连接到位点，在单体整合过程中所述位点经历构象改变，其中当所述聚合酶处于第一种构象状态时所述标记对是基本活化的，当所述聚合酶处于第二种构象状态时所述标记对是基本失活的，或当所述聚合酶处于第一种构象状态时所述标记对是基本失活的，当所述聚合酶处于第二种构象状态时所述标记对是充分活化的。

采用标记的聚合酶的方法

本发明提供了确定单体何时整合到增长的分子链的方法，其包含监测标记的可检测性质的步骤，其中所述标记位于或靠近，与之相连，或共价连接到聚合酶上的位点或与之相连，或共价连接到单体上的位点，其中在单体整合过程中所述位点经历变化，且当聚合酶处于第一种状态时所述可检测性质具有第一个值，当所述聚合酶处于第二种状态时具有第二个值，并且在每一单体添加过程中从第一个值循环变化为第二个值。

本发明提供了确定单体何时整合到增长的分子链的方法，其包含监测标记的可检测性质的步骤，其中所述标记位于或靠近，与之相连，或共价连接到聚合酶上的位点或与之相连，或共价连接到单体上的位点，其中在单体整合过程中所述位点经历构象变化，其中在所述聚合酶处于第一种构象状态时所述可检测性质具有第一个值，当所述聚合酶处于第二种构象状态时具有第二个值，并且在添加每一单体过程中从第一个值循环变化为第二个值。

本发明提供了确定单体何时整合到增长的分子链的方法，其包含暴露标记的聚合酶于光，监测所述标记的聚合酶和/或单体发射的荧光强度和/或频率的步骤，其中所述标记的聚合酶包含聚合酶，其包括的标记位于或靠近，与之相连，或共价连接到位点，所述位点在单体整合过程中经历构象变化，或所述标记或与之相连，或共价连接到单体上的位点，且其中当所述聚合酶处于第一种构象状态时所述标记以第一种强度和/或频率发射荧光，当所述聚合酶处于第二种构象状态时以第二种强度和/或频率发射荧

光，并且在添加每一单体的过程中，从第一个值循环变化为第二个值。

采用标记的聚合酶的单分子测序装置

本发明提供了包含单分子测序装置的组合物质，所述装置包含具有小室或片状表面的基体，其中至少一种标记的聚合酶被限制于此，还有多种小室(每一个小室都包括一种特异性单体)和多种连接这些小室的通道，其中每一复制复合物相距足够远以实现单独地从每一复合物收集数据。

本发明提供了单分子测序方法，包含如下步骤：提供多种单体给限制在基体上的标记的聚合酶，将标记的聚合酶暴露于光，测定标记的聚合酶发生的荧光的强度和/或频率。所述方法还可包括将测定的发射的荧光强度和/或频率与特定引物向增长的DNA链中的整合相联系的步骤。

协同标记的单体和标记的聚合剂

本发明提供了包含协同标记的聚合剂和标记的单体的组合物，其中当所述标记相互作用时，所述标记中至少一个的可检测性质发生变化。

本发明提供了包含协同标记的解聚剂和标记的可解聚单体的组合物，其中当所述标记相互作用时，所述标记中至少一个的可检测性质发生变化。

协同标记的单体和标记的聚合酶

本发明提供了包含协同标记的聚合酶和标记的单体的组合物，其中当所述标记相互作用时，所述标记中至少一个的可检测性质发生变化。

本发明提供了包含协同标记的聚合酶和标记的单体的组合物，其中当所述标记位于足以引起发射的荧光强度和/或频率的变化的距离内时，所述标记中至少一个的可检测性质发生变化。

本发明提供了包含标记的聚合酶和标记的单体前体的组合物，其中当所述标记相互作用时，所述标记中至少一个发射的荧光强度和/或频率发生变化。

本发明提供了包含标记的聚合酶和标记的单体前体的组合物，其中当所述标记位于足以引起发射的荧光强度和/或频率的变化的距离内时，所述标记中至少一个发射的荧光强度和/或频率发生变化。

本发明提供了单分子测序装置，其包含容器，容器内表面上限制有至

少一种标记的聚合酶，且有包含大量与内表面接触的标记的单体的溶液，或一亚组标记的单体和一亚组未标记的单体，它们一起提供了用于聚合的所有单体前体。

本发明提供了用于单分子测序的方法，包括如下步骤：提供大量标记的单体给限制在容器内表面的标记的聚合酶，将标记的聚合酶暴露于光，测定所述标记的聚合酶发射的荧光的强度和/或频率。所述方法还包括将测定的发射的荧光的强度和/或频率与特定的单体向增长的DNA链中的整合相关联。

本发明提供了再现储存的信息的系统，包括(a)包含具有代表数据流的元件(element)的序列的分子；(b)单分子测序仪，其包含聚合酶，该酶具有至少一个与之连接的标记；(c)适于激发聚合酶上至少一个标记的激发源；(d)适于检测聚合酶或单体上标记的响应的检测器，其中在元件的互补序列聚合过程中来自至少一个标记的所述响应变化，且所述变化代表数据流的内容。

本发明提供了从单个分子测定序列信息的系统，包含(a)具有元件的序列的分子；(b)单分子测序仪，其包含聚合酶，该酶有至少一个与之连接的标记；(c)适于激发聚合酶或单体上至少一个标记的激发源；(d)适于检测聚合酶上所述标记的响应的检测器，其中在元件的互补序列聚合过程中来自至少一个标记的所述响应变化，且代表分子的元件的序列。

本发明提供了从单个分子测定序列信息的系统，包含(a)具有元件的序列的分子；(b)单分子测序仪，其含有连接了至少一个荧光标记的聚合酶；(c)适于激发聚合酶和/或单体上的至少一个荧光标记的激发光源；(d)适于检测所述聚合酶上至少一个荧光标记发射的荧光的至少一种强度的检测器，其中每次新的核苷酸或核苷酸类似物聚合到互补序列时所述荧光标记的发射强度发生变化或不发射荧光，发射的持续时间或缺少发射或发射光的波长范围证明聚合到序列中的具体的核苷酸或核苷酸类似物，以使测序完成时数据流重现。

本发明提供了储存和重现数据的系统，包含(a)代表给定数据流的核苷酸或核苷酸类似物的序列；(b)单分子测序仪，其包含共价连接了至少一个荧光标记的聚合酶；(c)适于激发聚合酶上至少一个荧光标记的激发光源；(d)适于检测所述聚合酶的至少一个荧光标记发射的荧光的荧光检测器，其

中在每次新的核苷酸或核苷酸类似物聚合到互补序列时至少一个荧光标记发射或不发射荧光，发射的持续时间或缺少发射或发射光的波长范围证明聚合到序列的具体的核苷酸或核苷酸类似物，以使在完成测序时重现数据流。

本发明提供了储存和重现数据的系统，包含(a)代表给定数据流的核苷酸或核苷酸类似物的序列；(b)单分子测序仪，其包含共价连接了至少一个荧光标记的聚合酶；(c)适于激发聚合酶或单体上至少一个荧光标记的激发光源；(d)适于检测所述聚合酶或单体上至少一个荧光标记发射的荧光的荧光检测器；其中在每次新的核苷酸或核苷酸类似物聚合到互补序列时至少一个荧光标记发射或不发射荧光，发射的持续时间或缺少发射或发射光的波长范围证明聚合到序列的具体的核苷酸或核苷酸类似物，以使在完成测序时重现数据流。

本发明提供了对分子序列测序的方法，包含(a)代表给定数据流的核苷酸或核苷酸类似物的序列；(b)单分子测序仪，其包含共价连接了至少一个荧光标记的聚合酶；(c)适于激发聚合酶或单体上至少一个荧光标记的激发光源；(d)适于检测所述聚合酶上的至少一个荧光标记发射的荧光的荧光检测器，其中在每次新的核苷酸或核苷酸类似物聚合到互补序列时至少一个荧光标记发射或不发射荧光，发射的持续时间或缺少发射或发射光的波长范围证明聚合到序列的具体的核苷酸或核苷酸类似物，以使在完成测序时重现数据流。

本发明提供了合成 γ 磷酸修饰的核苷酸的方法，包含将分子标记连接于焦磷酸基，将修饰的焦磷酸与dNMP相接触以产生 γ 磷酸标记的dNTP。

本发明提供了用于5'末端标记生物分子的方法，包括使生物分子与能将 γ 磷酸标记的ATP的 γ 磷酸基传递到生物分子的5'末端的激酶接触，产生共价修饰的生物分子。

本发明提供了末端标记多肽或糖类的方法，包括如下步骤：将多肽或糖类与能传递原子或分子标记到蛋白或多肽的羧基或氨基末端，或 γ 磷酸标记的ATP的 γ 磷酸基到生物分子的5'末端的试剂接触，产生共价修饰的生物分子。

附图简述

参照下述说明与附图可以更好的理解本发明，附图中相似的元件已作了相同的编号。

图 1 表示 F R E T 活性为分隔荧光供体和受体的距离的函数；

图 2 表示 *Taq* DNA pol I (Klentaq 1) 的大片段的开放和闭合的三元复合体形式；

图 3 表示大片段 *Taq* DNA 聚合酶 I，在 3ktq (闭合的‘黑’) 和 1 tau (开放的‘浅蓝’) 之间的重叠；

图 4 表示 20% 变性聚丙烯酰胺图像，其含有来自 DNA 延伸实验的按大小分离的放射标记产物，所述实验 γ -ANS-磷酸-dATP；

图 5 表示 (A) 实际涉及的图像；(B) 减弱的磷光图像；(C) 采用 γ -ANS-磷酸-dNTP 的 DNA 延伸反应中产生的产物的增强的磷光图像；

图 6 表示 (A) 6% 变性的聚丙烯酰胺凝胶图像；(B) 实际胶的减弱的磷光图像；(C) 实际胶的增强的磷光图像，其包含采用 γ -ANS-磷酸-dNTP 的 DNA 延伸反应中产生的产物。

发明详述

本发明设计了一种方法，该方法采用标记的单体如 dNTP 和/或标记的聚合剂如聚合酶和/或与聚合剂相连接的标记试剂如聚合酶相连蛋白 (polymerase associated proteins) 或探针以在所述聚合酶具备活性的过程中直接读取准确的单体序列如 RNA 或 DNA 序列的碱基序列。本发明的方法适用于蛋白质合成或糖类合成或任何分子序列的合成，其中所述单体序列提供有用的信息如 RNA 或 DNA 分子，蛋白，糖类，混合的生物分子的序列，或单体的无机或有机序列，其储存了数据流。设计采用这些方法的方法和装置以创造出新的方法解决基本的研究问题，如监测复制过程中构象的变化，在多种序列中分析聚合酶整合的忠实度。设计本发明的单个分子检测系统以改进荧光分子化学，计算机建模，选择碱基 (base-calling) 的算法，生物分子的基因工程化，特别是实时或接近实时测序。本发明人也发现了所述方法可适用于解聚剂如外切核酸酶，其中所述聚合物序列通过解聚而非聚合确定。而且，本发明的单个分子系统可修正用于平行和/或大量平行检测，其中所述标记的聚合酶在基体上排成列。采集自这些列的数据可用于改进序列置信度和/或同时测序许多不同来源的 DNA 区以鉴定

类似性或不同。

本发明的单分子测序系统可能用于替代当今的 DNA 测序技术，因为这种方法可以缩短时间，降低与测序过程有关的劳动，和费用，并且可以产生可高度放大的测序系统，其以每个反应至少 1 到 2 个数量级改进了所述 DNA 序列的发现方法。

收集发射信号的模式，可以是直接地，如通过 Intensified Charge Coupled Devise (ICCD) 或在电子检测前通过一个中间体或一系列中间体扩增信号，其中所述信号被解码，为每一碱基指定了置信度值以揭示互补于模板的序列的序列。因此，本发明也提供了扩增荧光标记发射的荧光的技术，其采用物理光扩增技术或分子级联试剂 (molecular cascading agent) 扩增单个分子荧光反应产生的光。

本发明的单分子测序技术能：(1)使其易于通过简单的测序其基因组或其部分对生物体分类或鉴定生物体内的变异；(2)更易于快速鉴定病原体或基因改变的病原体，特别是在极端环境下如战争所用的病原体；和(3)为法律强制或军事应用快速鉴定人。

本发明的单分子测序技术的一个实施方案涉及策略性定位于 DNA 聚合酶上的一对标记以使在聚合反应过程中当 dNTP 整合时，所述标记改变相对间隔。此相对改变引起可检测性质的改变，如一个或两个标记的荧光强度和 / 或频率的改变。所述可检测性质的这些变化的时间图谱证明每一单体整合的发生，并提供有关每一整合发生时哪一种特定的 dNTP 整合的证明。所述引物对不是必须共价连接到聚合酶，而是可连接到与所述聚合酶连接的分子，以这样的方法连接：在碱基整合过程中所述碱基的相对间隔发生改变。

本发明的单分子测序技术的另一个实施方案涉及策略性定位于 DNA 聚合酶上的单个标记，其与 dNTP 上的标记或每一 dNTP 上的分隔的标记相互作用。对于每一 dNTP 所述标记可以是不同的，如色彩编码的标记，其发射不同色彩的荧光。在聚合过程中当下一 dNTP 整合时，通过特征荧光信号 (颜色) 或荧光信号的强度和 / 或频率的改变指示碱基的身份。聚合酶的整合速度可以是不同的和 / 或可控的以产生基本“实时”或接近“实时”或实时读取聚合酶活性和碱基序列。可以每小时 >100,000 碱基的速度从每一聚合酶收集序列数据。

本发明的单分子测序技术的另一个实施方案中，每一标记的聚合酶包括位于所述聚合酶上或之内的供体标记和受体标记，其中当 dNTP 结合，dNTP 整合和 / 或链延伸过程中所述标记间距离改变。此内标 (inter-tag) 距离的改变导致荧光标记发射的荧光的强度和 / 或波长的变化。监测发射光的强度和 / 或波长的变化可以提供有关聚合发生以及整合碱基的身份的信息或数据。

在另一实施方案中，设计聚合酶上的标记与 dNTP 上的标记相互反应，其中所述相互反应改变了所述标记之一个或两个的可检测性质。采用标记的 dNTP 监测每一荧光标记的聚合酶的聚合反应以测定从其得到的碱基整合数据的效力。已开发了特定的检测和方法以及特定的分析设备以测定和定量荧光数据，使每一整合的 dNTP 得到检测并证实。同时，本发明人证实了标记的 dNTP，其由合适的聚合酶聚合，并开发了分析所述反应发射的荧光并解译碱基身份的软件。本领域技术人员将了解本领域已知的合适的荧光活性对，它们从商家可以获得，如位于 Oregon 的 Molecular Probes 或 Biosearch Technologies, Inc. 位于 Novato, CA。

本发明所用标记的 DNA 聚合酶通过基因工程化提供一个或多个标记的结合位点，所述位点可以允许本发明实施不同的实施方案。一旦找到合适候选聚合酶，突变和 / 或修饰所述聚合酶内特定的氨基酸，该反应为本领域所知；然而前提是，所述突变和 / 或修饰不对聚合效率产生显著不利影响。使所述突变的和 / 或修饰的氨基酸适于促进标记的结合如染料或荧光供体或受体分子，如果是在光活化的标记情况下。工程化聚合酶一旦形成，可以与一个或多个合适的标记接触，并用于本发明的装置和方法中。

工程化聚合酶使其起 DNA 碱基身份的直接分子感受器的作用，这为快速且可能实时的酶 DNA 测序系统提供了一条途径。本发明所述单分子测序系统可以显著缩短时间，降低与测序过程有关的劳动强度和费用，并可高度放大。本发明所述单分子测序系统：(1) 可以每个反应至少两个数量级地改进序列发现方法；(2) 不受长度的限制，所述限制与降解碱基的单个分子方法有关；(3) 可以对所需 (靶) DNA 序列直接测序，特别是基因组，而无需克隆或 PCR 扩增，此两种方法都引入错误到序列中。本发明系统可以简单的通过对目的基因组或基因组任何所需部分测序容易地进行生物体分类或鉴定生物体内的变异。本发明系统适于快速鉴定病原体或工程化病原体，其

对于评价与健康有关的影响，普通 DNA 诊断，包括癌症检测和 / 或定性，基因组分析，或更广泛形式的基因变异检测具有重要意义。本发明所述单个分子 DNA 测序系统可以成为单个分子基因分析的启动平台 (enabling platform) 技术。

本发明的单分子测序系统具有如下优点：(1) 所述系统取消了测序反应处理，凝胶或毛细管加载，电泳，数据汇编；(2) 所述系统显著节约了劳力，时间和成本；(3) 所述系统可以接近实时或实时获得数据，处理数据并确定整合的发生 (计时，持续时间，等)，碱基序列等；(4) 所述系统可以微阵列的形式对平行或大量平行的样品进行处理；(5) 所述系统可以在一天或更少的时间内进行快速基因组测序；(6) 所述系统需要很少量分析材料；(7) 所述系统可以进行动物的快速基因鉴定，筛选和表征，包括人或病原体；(8) 所述系统可以大幅提高序列的容许能力；(9) 所述系统可以避免 PCR, RT-PCR 和转录过程中引入的错误；(10) 所述系统可以准确地获得序列信息用于检测等位基因特异的突变 (allele-specific mutation)；(11) 所述系统可以快速进行医疗诊断，例如单核苷酸多态性 (SNP) 检测；(12) 所述系统可以改进碱基检索，例如在大量不同序列中测定聚合酶的整合速率；分析不同的前后序列中的错误；后生型分析；蛋白糖基化分析；蛋白质鉴定；(13) 所述系统可以创造新的加强的 (robust) (坚固的 (rugged)) 单个分子检测装置；(14) 所述系统可以开发与生物分子相适合的系统和程序；(15) 所述系统可以开发基因的毫微级设备 (nanomachines) 或毫微米技术；(16) 所述系统可以构建大的基因数据库，且(17) 所述系统对于低的突变发生检测具有高灵敏度。

单个分子 DNA 测序的简述

在本发明的单个分子 DNA 测序系统的一个实施方案中，单个标记连接到聚合酶上的适当位点，一个独特的标记连接到四种核苷酸的每一个上：dATP, dTTP, dCTP, dGTP。每一 dNTP 上的标记被设计为具有独特的发射特征 (signature) (即，不同的发射频率光谱或颜色)，整合时可以直接检测到。当标记的 dNTP 整合到增长的 DNA 聚合体上时，由于聚合酶标记和 dNTP 标记的相互作用，特征性荧光信号或碱基发射的特征波行被发射。所述荧光信号，即发射强度和 / 或频率，被检测和分析以确定 DNA 碱基序列。

挑选用于本发明的标记的聚合酶和 / 或 dNTP 的标准是聚合酶和 / 或 dNTP 上的标记不干扰 Watson-Crick 的碱基配对或对聚合酶活性无显著不良影响。本发明者发现包含连接到末端 (γ) 磷酸的标记的 dNTP 通过天然 Taq 聚合酶与未标记的 dNTP 一起或只采用标记的 dNTP 整合。因为所得 DNA 链在它们的分子组成上不包括任一 dNTP 标记，所以优选在 β 和 / 或 γ 磷酸基上标记 dNTP，这样可最小化酶变形 (enzyme distortion) 和背景荧光。

本发明的测序系统的一个实施方案涉及将荧光供体如荧光素或荧光素型分子放置于聚合酶上，独特的荧光受体如 d-若丹明或类似的分子放置于每一 dNTP 上，其中每一独特的受体，当与聚合酶上的供体相互作用时，产生的荧光光谱包括至少一种可进行区别的频率或光谱特征。当输入的，标记的 dNTP 与聚合酶结合用于 DNA 延长时，分析检测到的荧光信号或光谱，确定整合的碱基的身份。

本发明的测序系统的另一个实施方案涉及聚合酶上的荧光标记和 dNTP 上的独特的淬灭剂，其中所述淬灭剂优选对聚合酶标记具有可进行区别的淬灭效果。结果，每一输入的淬灭剂标记的 dNTP 的身份根据其对聚合酶荧光标记的发射的独特淬灭效果而确定。再一次，检测整合过程中产生的信号并分析确定所整合的每一碱基，其序列产生 DNA 碱基序列。

试剂

本发明所用合适的聚合剂包括但不限于聚合相对于特定模板的单体的任何聚合剂如 DNA 或 RNA 聚合酶，逆转录酶，或类似的或以逐步的方式聚合单体的聚合剂。

本发明所用合适的聚合酶包括但不限于可从其宿主分离的任何聚合酶，所述分离的量足以用于纯化和使用和 / 或以足够本发明使用的量基因工程化到其它生物用于表达，分离和纯化，如 DNA 或 RNA 聚合酶，所述聚合酶聚合 DNA，RNA 或混合序列到延伸的核酸聚合体中。本发明所用的优选的聚合酶包括天然聚合酶的突变体或突变的变体，其中所述突变体具有一个或多个氨基酸取代，所述取代可用于连接原子或分子标记，所述标记具有可检测性质。DNA 聚合酶的实例包括但不限于，采用 RNA 或 DNA 模板的 HIV-1 逆转录酶，来源于水生栖热菌或大肠杆菌的 DNA 聚合酶 I，噬菌体 T4 DNA 聚合酶，T7DNA 聚合酶或类似的。RNA 聚合酶例如但不限于 T7 RNA 聚合

酶或类似的。

本发明所用的合适的解聚剂包括但不限于任何以逐步方式解聚单体的解聚剂，如 DNA, RNA 或混合的 DNA / RNA 聚合体情况下的核酸外切酶，多肽和酶情况下的蛋白酶，或可以顺序解聚多糖的酶系统。

本发明所用的合适的单体包括但不限于任何采用聚合剂可以逐步聚合到聚合体的单体。本发明所用的合适的核苷酸包括但不限于天然存在的核苷酸，合成的其类似物，连接了原子和 / 或分子标记的类似物，或其混合物或组合。

本发明所用合适的原子标记包括但不限于任何可与聚合剂或 dNTP 特定位点相连的原子成分，特别是铕位移试剂 (Europium shift agent)，nmr 活性原子或类似的。

本发明所用合适的原子标记包括但不限于任何可与聚合剂或 dNTP 上特定位点连接的原子成分，特别是荧光染料如 d-若丹明受体染料包括二氯 [R110]，二氯 [R6G]，二氯 [TAMRA]，二氯 [ROX] 或类似的，荧光素供体染料包括荧光素，6-FAM，或类似的；吖啶包括吖啶橙，吖啶黄，二氨基吖啶，pH7，或类似的；芳香烃包括 2-甲基苯并 哒，乙基 p-二甲基氨基苯甲酸盐，苯酚，吡咯，苯，甲苯，或类似的；芳基次甲基染料包括金胺 0，结晶紫，H₂O，结晶紫，甘油，孔雀绿或类似的；香豆素染料包括 7-甲氧基香豆素-4-乙酸，香豆素 1，香豆素 30，香豆素 314，香豆素 343，香豆素 6 或类似的；青色素 (cyanine) 染料包括 1, 1' -二乙基-2, 2' 碘化青色素，隐青色素 (cryptocyanine)，靛炭青色素 (Indocarbocyanine) (C3) 染料，靛二碳青色素 (Indodicarbocyanine) (C5) 染料，靛三碳青色素 (Indotricarbocyanine) (C7) 染料，氧碳青色素 (Oxacarbocyanine) (C3) 染料，氧二碳青色素 (Oxadicarbocyanine) (C5) 染料，氧三碳青色素 (Oxatricarbocyanine) (C7) 染料，碘化松柏氰醇 (Pinacyanol iodide), Stains all, 硫碳青色素 (Thiacarbocyanine) (C3) 染料，乙醇，硫碳青色素 (Thiacarbocyanine) (C3) 染料，n-丙醇，硫二碳青色素 (Thiadicarbocyanine) (C5) 染料，硫三碳青色素 (Thiatricarbocyanine) (C7) 染料，或类似的；Dipyrrin 染料包括 N, N'-Di fluoroboryl-1, 9-二甲基-5- (4-碘代苯基)-dipyrrin, N, N'-Difluoroboryl-1, 9-二甲基-5-[(4-(2-三甲基甲硅烷乙炔基 (trimethylsilylethyanyl)), N,

N'-Difluoroboryl-1, 9-二甲基-5phenyldipyrrin, 或类似的； 部花青 (Merocyanine), 包括 4- (二氟亚甲基)-2-甲基-6-(p-二甲基氨基苯乙烯基)-4H-吡喃 (DCM), 乙腈, 4- (二氟亚甲基)-2-甲基-6-(p-二甲基氨基苯乙烯基)-4H-吡喃 (DCM), 甲醇, 4-二甲胺-4'-硝基 1, 2-二苯乙烯, 部花青 540, 或类似的； 混杂 (Miscellaneous) 染料, 包括 4', 6-二脒基 (Diamidino)-2-苯基吲哚 (DAPI), 4', 6-二脒基-2-苯基吲哚 (DAPI), 二甲亚砜, 7-苯甲胺-4-硝基苯-2-氧-1, 3-二唑, 丹酰甘氨酸 (Dansyl glycine), H₂O, 丹酰甘氨酸, 二 烷 (dioxane), Hoechst 33258, DMF, Hoechst 33258, H₂O, 荧光黄 CH, 吡氧噻嗪 (Piroxicam), 硫酸奎宁 (Quinine sulfate), 0.05 M H₂SO₄, 硫酸奎宁, 0.5 M H₂SO₄, Squarylium 染料 III, 或类似的； 寡亚苯 (Oligophenylenes) 包括 2, 5-二苯基𫫇唑 (PP0), 联苯, POPOP, p- 四联苯, p-三联苯, 或类似的； 呕嗪, 包括紫色高氯酸羟甲苯酯 (Cresyl violet perchlorate), 尼罗蓝 (Nile Blue), 甲醇, 尼罗红 (Nile Red), 尼罗蓝, 乙醇, 呕嗪 1, 呕嗪 170, 或类似的； 多环芳香烃, 包括 9, 10-双(苯乙炔基)蒽, 9, 10-二苯蒽, 蒽, 萘, 菲, 芘, 或类似的； 多烯/聚炔烃, 包括 1, 2-二苯乙炔, 1, 4-二苯丁二烯, 1, 6-二苯己三烯, β-胡萝卜素, 茜 (Stilbene), 或类似的； 氧化还原活性的生色团, 包括蒽醌, 偶氮苯, 苯醌, 二茂铁, 核黄素, Tris (2, 2'-二吡啶基) 钉 (II), 四吡咯 (Tetrapyrrole), 胆红素, 叶绿素 a, 二乙醚, 叶绿素 a, 甲醇, 叶绿素 b, 二质子化的四苯卟啉 (Diprotonated tetraphenylporphyrin), 氧化苏木精, Magnesium octaethylporphyrin, Magnesium octaethylporphyrin (MgOEP), 酞氯镁 (Magnesium phthalocyanine) (MgPc), PrOH, 酞氯镁 (MgPc), 噻啶, Magnesium tetramesitylporphyrin (MgTMP), 四苯基卟啉镁 (Magnesium tetraphenylporphyrin) (MgTPP), 八乙基卟啉, 酞氯 (Pc), 吲哚, 四-t-丁基氮杂卟啉 (Tetra-t-butylazaporphine), 四-t-丁基萘青色素 (Tetra-t-butylnaphthalocyanine), 四 (2, 6-二氯苯基) 吲啉 (Tetrakis (2, 6-dichlorophenyl) porphyrin), 四 (o-胺苯基) 吲啉 (Tetrakis (o-aminophenyl) porphyrin), Tetramesitylporphyrin (TMP), 四苯卟啉 (Tetraphenylporphyrin) (TPP), 维生素 B12, 八乙基卟啉锌 (Zinc octaethylporphyrin) (ZnOEP), 酞氯锌 (Zinc phthalocyanine) (ZnPc), 噻啶, Zinc tetramesitylporphyrin (ZnTMP), Zinc tetramesitylporphyrin

radical cation, 四苯基卟啉锌(ZnTPP), 或类似的; 氧杂蒽(Xanthenes), 包括曙红 Y, 荧光素, 碱性乙醇(basic ethanol), 荧光素, 乙醇, 若丹明 123, 若丹明 6G, 若丹明 B, 玫瑰红(Rose Bengal), Sulforhodamine 101, 或类似的; 或其混合物或组合或其合成的衍生物或 FRET 荧光团-淬灭剂对, 包括 DL0-FB1 (5'-FAM/3'-BHQ-1) DL0-TEB1 (5'-TET/3'-BHQ-1), DL0-JB1 (5'-JOE/3'-BHQ-1), DL0-HB1 (5'-HEX/3'-BHQ-1), DL0-C3B2 (5'-Cy3/3'-BHQ-2), DL0-TAB2 (5'-TAMRA/3'-BHQ-2), DL0-RB2 (5'-ROX/3'-BHQ-2), DL0-C5B3 (5'-Cy5/3'-BHQ-3), DL0-C55B3 (5'-Cy5.5/3'-BHQ-3), MB0-FB1 (5'-FAM/3'-BHQ-1), MB0-TEB1 (5'-TET/3'-BHQ-1), MB0-JB1 (5'-JOE/3'-BHQ-1), MB0-HB1 (5'-HEX/3'-BHQ-1), MB0-C3B2 (5'-Cy3/3'-BHQ-2), MB0-TAB2 (5'-TAMRA/3'-BHQ-2), MB0-RB2 (5'-ROX/3'-BHQ-2); MB0-C5B3 (5'-Cy5/3'-BHQ-3), MB0-C55B3 (5'-Cy5.5/3'-BHQ-3) 或类似得自 Biosearch Technologies Inc. of Novato, CA, 的 FRET 对, 具有 nmr 活性集团的标记, 具有光谱特征的标记, 所述特征可以被容易地鉴定, 如 IR, 远 IR, 可见 UV, 远 UV 或类似的。

酶的选择

本发明人发现水生栖热菌(*Thermus aquaticus*)的 DNA 聚合酶-*Taq* DNA 聚合酶 I-是很适于本发明的单个分子装置, 系统和方法。*Taq* DNA 聚合酶有时在本文中简写为 *Taq*, 其有许多特点, 发明人可以利用来构建标记的聚合酶, 用于本申请公开的本发明。当然, 本领域普通技术人员将认识到其它聚合酶也能适用于本发明的单分子测序系统。

由于 *Taq* DNA 聚合酶 I 耐受在其活性位点内或附近有许多突变(如 Patel et al, J. Mol Biol., volume 308, pages 823-837, 引入作为参考), 所以所述酶更能耐受酶标记修饰并也能整合较大范围的修饰的核苷酸基体。

可获得 *Taq* DNA 聚合酶的晶体结构

对 *Taq* DNA 聚合酶有 13 种结构解释, 有或没有 DNA 模板/引物, dNTP, 或 ddNTP, 其允许足够的信息以在聚合酶内部挑选氨基酸位点, 使该位点上

连接原子和/或分子标记，如荧光标记，而不对聚合酶活性产生不良影响。参见，例如，Eom 等.，1996，Li 等，1998a，Li 等，1998b。此外，本发明人具有一书写的程序以帮助鉴定优化的标记添加位点。所述程序对比与 *Taq* 聚合酶的开放和闭合形式有关的结构数据以鉴定出聚合酶结构中的某些区域，这些位置可用来优化所述聚合酶上标记和 dNTP 间的极端构象 (conformation extremes) 的不同或优化所述聚合酶上两个标记之间间距的变化，从而提高或最大化所述标记或标记对的可检测性质的变化。

Taq DNA 聚合酶在大肠杆菌中有效表达

所述 *Taq* DNA 聚合酶在大肠杆菌中有效表达，可以有效生产和纯化新生的聚合酶和其变体用于快速鉴定，表征和优化，用于本发明的单个分子 DNA 测序系统的工程化 *Taq* DNA 聚合酶。

蛋白序列中没有半胱氨酸

所述 *Taq* DNA 聚合酶不含有半胱氨酸，这可以容易的产生含半胱氨酸的突变体，其中单个半胱氨酸在策略位点被放置或取代现存的氨基酸，其中所述插入的半胱氨酸作为标记结合位点。

酶的处理能力可以被改变

尽管天然 *Taq* DNA 聚合酶可以不代表本发明的测序系统的优化的聚合酶，因为它不是处理能力很强的聚合酶(在解离前整合 50-80 个核苷酸)，通过适当修饰的碱基选择软件 (base calling software)，可以补偿低处理能力。备选地，所述 *Taq* DNA 聚合酶的处理能力能通过将处理能力增强序列插入到聚合酶基因的基因工程化而得到增强。预计高处理能力聚合酶可以最小化可能由模板解离效应引起的并发症 (complication)，其能改变聚合速率。通过引入 T7 DNA 聚合酶的 76 氨基酸 ‘处理能力结构域 (processivity domain)’ 到 *Taq* 的 H 和 H₁ 螺旋 (位于聚合酶内 ‘拇指 (thumb)’ 区的顶端) 之间，可以在基因水平上改变 *Taq* 的处理能力。所述处理能力结构域也包括 T7 DNA 聚合酶的硫氧化蛋白结合域 (TBD)，导致 *Taq* 聚合酶依赖硫氧化蛋白提高 *Taq* 聚合酶的处理能力和比活性。参见 Bedford et al., 1997; Bedford et al., 1999。

Taq DNA 聚合酶具有 5' 到 3' 外切核酸酶活性且是热稳定的单链 M13 DNA 和合成的寡核苷酸用于开始的研究。优化聚合酶活性之后，所述测序系统用于从分离的染色体-双链 DNA 分子直接测定序列信息。通常，加热双链 DNA 分子的样品足以产生或保持双链 DNA 分子的链式 DNA 形式用于测序。

为利于单链状态，保持用于单分子 DNA 测序的工程化酶中的天然 *Taq* DNA 聚合酶的 5' 到 3' 外切核酸酶活性。*'TaqMan'* 分析利用聚合酶的这种活性。这种外切核酸酶活性采用切口转平移反应机制 (nick-translation reaction mechanics) 除去可能从复制起点的下游发生变性的一条双螺旋的链 (duplex strand)。通过合成的寡核苷酸引物 (如果需要特异性反应启动) 或 DNA 分子上的切口 (如果处理多个反应) 可以启动从工程化聚合酶的合成以确定整个 DNA 分子的序列。

聚合酶没有 3' 到 5' 外切核酸酶活性

所述 *Taq* DNA 聚合酶不包含 3' 到 5' 外切核酸酶活性，这意味着所述聚合酶不能使检测到荧光信号的碱基被产生另一特征荧光信号的碱基所取代。

所有聚合酶都会产生复制错误。所述 3' 到 5' 外切核酸酶活性用于校对新复制的 DNA 链。由于 *Taq* DNA 聚合酶缺少此校对功能，碱基整合中的错误会成为 DNA 复制中的错误。*Taq* DNA 聚合酶的错误率是每合成 ~100,000 个碱基有 1 个错误，这样低的错误率是以确保相对高的忠实度。例如参见 Eckert and Kunkel, 1990 ; Cline et al., 1996。已提示和证明对于聚合酶，此外切核酸酶活性的消除揭示了整合中降低的忠实度。因此，*Taq* 聚合酶必须-必要地-在初始核苷酸的选择和 / 或整合过程中更准确，因此是本发明所用的极好选择。

通过确定它们合成已知序列的错误率分析本发明的工程化聚合酶的错误率。所述错误率确定平行进行的最佳反应数量，以赋予测序信息的连贯性置信值。优选的数目是 1 或 10 或更多。例如，本发明人已经发现碱基 (base context) 影响聚合酶的准确度和反应动力学，此信息用于赋予置信值给单个配对碱基 (base calls)。但是，根据特定测序项目的目的，尽快产生基

因组序列可能更为重要。例如为了初始的鉴定目的或对潜在病原体的快速筛选，优选以降低的准确度产生，或绘制，病原体的基因组序列。

Taq DNA 聚合酶是挑选的用于单个分子 DNA 测序的酶

工程化聚合酶使其起 DNA 碱基身份的直接分子检测器的作用使最快速酶性 DNA 测序系统成为可能。对于如上所述的原因，*Taq* DNA 聚合酶是在基因水平进行修饰和改造，以适于单个分子 DNA 测序的最佳酶。此外，复制中有关 DNA 聚合酶结构和功能的基本研究问题可采用此技术领先的单个分子检测系统和在其它学科的分子模型解决。本发明人发现天然 *Taq* DNA 聚合酶整合 γ -标记的 dNTP，产生延伸的 DNA 聚合体。重要的是，修饰的核苷酸的整合不损及聚合酶的活性，通过整合 γ -标记的核苷酸进行的引物链延伸遵守 Watson-Crick 碱基配对法则。

检测标记的聚合酶-核苷酸之间的相互作用

检测聚合酶-核苷酸相互作用的优选方法涉及基于荧光共振的能量传递 (fluorescence resonance energy transfer-based) (FRET-based) 的方法，最大化信号和最小化噪音。当从受体的发射比从供体的发射更强时，即，在激发频率时受体的荧光量子产出高于供体时，存在基于 FRET 的方法。可从计算模型估计 FRET 方法的效率。例如参见 Furey et al., 1998; Clegg et al., 1993; Mathies et al., 1990。能量传递的效率 (E) 可从公式 (1) 计算：

$$E = 1 / (1 + [R / R_0]^6) \quad (1)$$

其中在 $E=0.5$ 时， R_0 是 Forester 必需距离 (Forester critical distance)。 R_0 可从公式 (1) 计算：

$$R_0 = (9.79 \times 10^3) (\kappa^2 n^{-4} Q_D J_{DA})^{1/6} \quad (2)$$

其中 n 是介质的折光指数 (对于水溶液 $n=1.4$)， κ^2 是涉及两个转变偶极 (transition dipoles) 的相对角度的几何方向因子 (geometric orientation factor) (κ^2 通常为 $2/3$)， J_{DA} [$M^{-1} \text{cm}^3$] 重叠积分代表供体发射和受体吸收的标准化的光谱重叠， Q_D 是量子产出 (quantum yield)。所述重叠积分由公式 (3) 得到：

$$J_{DA} = [\int F_D(\lambda) \epsilon_A(\lambda) \lambda^4 d\lambda] / [\int F_D(\lambda) d\lambda] \quad (3)$$

其中 F_D 是供体发射， ϵ_A 是受体吸收。 Q_D 从公式 (4) 计算：

$$Q_D = Q_{RF} (I_D / I_{RF}) (A_{RF} / A_D) \quad (4)$$

其中 I_D 和 I_{RF} 是供体和参照化合物 (0.1N NaOH 中荧光素) 的荧光强度， A_{RF} 和 A_D 是参照化合物和供体的吸收。 Q_{RF} 是 0.1N NaOH 中荧光素的量子产出，记为 0.90。

供体和受体之间的距离 R 视聚合酶的不同构型 (例如构象) 来测定，其目的是为了得到构象平均值。如果两个标记都在聚合酶上，则 R 是处于开放和闭合构象的供体和受体之间的距离，而如果供体在聚合酶上，受体在 dNTP 上，当 dNTP 与聚合酶结合且所述聚合酶处于闭合形式时， R 是供体和受体之间的距离。

标记的 γ -磷酸和所选用于标记的开放和闭合的聚合酶构象中的氨基酸位点之间的距离决定了最佳的染料组合。如果所述在供体和受体之间的距离 (R) 与 R_0 相同 (R_0 是 Forster 必需距离)，FRET 效率 (E) 是 50%。如果 R 大于 $1.5 R_0$ ，所述能量传递效率可忽略不计 ($E < 0.02$)。鉴定酶内的位点，在此位点开放形式的 R/R_0 和闭合形式的 R/R_0 之间的差异超过 1.6，如果需要，这些距离和 / 或距离差别可以通过基因工程提高。图 1 是 FRET 效率相对于距离的图示。

荧光染料选择方法

当聚合酶处于其闭合构型时，挑选染料组以最大化标记的 dNTP 和聚合酶上标记之间能量的传递效率，当聚合酶处于开放构型时，最小化 dNTP 上标记 (非有效结合或在溶液中) 和聚合酶上标记之间的能量传递。如果反应介质中每一核苷酸的摩尔浓度不超过约 $1 \mu M$ ，计算所得标记的核苷酸间平均距离大于或等于约 250 \AA 。因为此距离几倍大于聚合酶在其开放到闭合构象中分离的位点间的距离，可观察到聚合酶和游离 dNTP 间的最小化的 FRET 背景。优选地，核苷酸浓度降低到 $1 \mu M$ 以下。降低 dNTP 浓度到至少 K_m 的 10% 以下可进一步最小化背景荧光，并提供方便的控制聚合酶反应速率的方法用于实时监控。在这样的条件下，聚合反应速度与 dNTP 浓度成线性比例，因此，对调节高度敏感。此外，使用单个激发波长可以改进对每一 dNTP 上独特的标记的鉴定。较低波长激发激光用于达到高选择性。

在一优选的实施方案中，荧光供体与聚合酶上位点连接，所述聚合酶包含更易于供体连接的取代的氨基酸，如半胱氨酸，四个独特的荧光受体连接到每一 dNTP 上。例如将荧光素连接到聚合酶上的位点，若丹明，若丹明衍生物和/或荧光素衍生物连接到每一 dNTP。设计每一供体-受体荧光团对的吸收光谱使其足以与其它荧光团对区分，以使激发后进行分离鉴定。优选的，选择所述供体以使激发光活化供体，其然后有效地传递激发能给受体之一。能量传递后，受体发射独特特征的荧光。荧光供体的发射光谱必须与荧光受体的吸收光谱显著重叠才能有效地进行能量传递。但是本发明的方法也能采用二，三或四个独特的荧光供体-受体对通过并行反应进行。

对荧光团的选择不仅有赖于其酶的适应性 (compatibility)，还有其光谱和光化学性质。例如，受体荧光团在供体荧光团的激发波长不具有任何显著的吸收是关键的，稍不重要的(也是需要的)是，供体荧光团在受体荧光团的检测波长不具有发射。这些光谱特征可通过对荧光团的环系统进行的化学修饰削弱。

尽管可修改 dNTP 以在几个位点进行标记，包括碱基，糖和磷酸基，但优选对 dNTP 标记于 β 和/或 γ 磷酸。标记 dNTP 的末端磷酸具有独特的优点。当输入的，标记的 dNTP 结合到聚合酶的活性位点时，从聚合酶上的供体到 dNTP 上的受体出现显著的 FRET。受体独特的荧光鉴定整合了哪一种 dNTP。一旦标记的 dNTP 被整合到增长的 DNA 链，所述荧光受体，其目前是结合到焦磷酸基上，随着焦磷酸基的裂解被释放到介质中。实际上，增长的 DNA 链根本不包括荧光受体分子。基本上，FRET 只发生在聚合酶上的供体和输入的受体标记的 dNTP 之间，一次一个。此方法好于受体备选连接到 dNTP 的 dNMP 部分内的位点，或者使用具有多个修饰的 dNTP。如果受体连接的位点不是 β 或 γ 磷酸基，其成为增长的 DNA 链的一部分，所述 DNA 链将包含多个荧光受体。可能发生对聚合反应和 FRET 检测的干扰。

如果聚合介质(背景)中标记的 dNTP 的荧光有问题，可将碰撞淬灭剂加入到聚合介质中，所述淬灭剂不与 dNTP 上受体共价作用但淬灭介质中标记的 dNTP 的荧光。当然，也可改造所述淬灭剂，使其与聚合酶上的供体没有明显接触。为使碰撞淬灭剂与聚合酶上的供体之间的相互作用最小化，可优选使聚合酶的标记位于内部且避开碰撞淬灭剂，或可将所述碰撞淬灭剂

制备成大的空间体积或与空间体积大的基团相连以降低所述淬灭剂与聚合酶之间的相互作用。检测聚合酶-核昔酸之间相互作用的另一优选方法涉及应用核昔酸特异的淬灭剂，以淬灭聚合酶上荧光标记的发射。这样，用荧光团标记聚合酶，而每一 dNTP 用荧光团的淬灭剂标记。一般，DABCYL (4-(4'-二甲基氨基苯基偶氮(dimethylaminophenylazo))) 安息香酸是通用的淬灭剂，其吸收荧光团的能量，所述荧光团如 5-(2'-氨基)氨基萘-1-磺酸(AEANS)，并消散热量。优选的，为每一 dNTP 选择淬灭剂，以使当每一淬灭剂靠近荧光团时，得到不同的淬灭效果。因此，当其被整合到增长的 DNA 链中时，淬灭的程度用于鉴定每一 dNTP。此优选的检测方法的一个优点是来自单个源的荧光发射使背景噪音可忽略不计。尽管较不优选，如果只有二或三种合适的淬灭剂被鉴定，然后 4 个 dNTP 中有 2 或 3 个被标记，每次采用标记的 dNTP 的不同的对进行一系列的聚合反应。组合这些反应结果，得到所述 DNA 分子的完整序列。

标记 *Taq* 聚合酶和 dNTP 的位点选择

尽管本发明可以结合具有可检测性质的任何类型的原子和 / 或分子标记，但仍然描述了采用优选类型的标记，即荧光标记进行位点选择和标记结合的方法。

聚合酶和 / 或 dNTP 的荧光标记

设计连接到聚合酶或 dNTP 的荧光探针或淬灭剂，以最小化对 DNA 聚合反应的负面影响。本发明人已经开发了合成方法用于采用荧光探针或淬灭剂化学标记聚合酶和 dNTP。

通常，通过用编码更易与分子标记反应的氨基酸，如半胱氨酸(通过诱变)，的密码子来取代编码聚合酶的 DNA 序列中所选的氨基酸的密码子来标记聚合酶。一旦制备了突变的 DNA 序列，所述突变体插入到大肠杆菌中用于表达。表达后，分离并纯化突变体聚合酶。然后检测纯化的突变体聚合酶的聚合酶活性。活性确认后，突变体聚合酶与摩尔稍过量的所需标记反应以得到靠近化学计量的标记。备选地，所述聚合酶能用过量的标记处理并根据时间函数进行标记。当得到接近化学计量的标记时，标记反应停止。

如果突变体聚合酶包括几个位点，所述位点包括能用所需分子标记进

行标记的靶残基，那么标记反应也能在特定的反应条件下进行，如采用保护基团或竞争性抑制剂和可逆的阻断基因，其随后被除去。如果突变体聚合酶中靶氨基酸残基靠近活性 dNTP 结合位点，首先加入饱和水平的保护基团或竞争性抑制剂以保护靶残基，随后加入可逆阻断基团以使非靶残基失活。然后从靶残基除去保护基团或竞争性抑制剂，用所需标记处理突变体聚合酶以标记靶残基。最后，从突变体聚合酶中的非靶残基化学除去阻断基团，由于靶残基上标记基本上完全分离得到标记的突变体聚合酶。

备选地，如果靶残基不靠近活性位点，所述聚合酶可用阻断基团处理以使非靶残基失活。除去未反应的阻断基团后，用标记靶残基的所需标记处理突变体聚合酶。最后，从突变体聚合酶中非靶残基化学除去阻断基团，以得到标记的突变体聚合酶。

选择 *Taq* 聚合酶的氨基酸位点

发明人已鉴定了 *Taq* 聚合酶中的某些氨基酸，所述氨基酸有可能经受得住突变和随后的标记结合如荧光标记的连接。而许多位点能进行半胱氨酸取代和标记结合，采用下列标准鉴定聚合酶中优选的位点：(1)它们不与其它蛋白接触；(2)它们不改变聚合酶的构象或折叠；和(3)它们不涉及蛋白功能。采用突变研究的组合完成挑选，包括序列分析数据，计算机研究包括分子停靠(molecular docking)数据和分析聚合酶活性和忠实度。位点突变后，采用计算机研究来精制分子模型并帮助鉴定其它潜在的突变位点。

间接地，通过采用进化追踪方法(evolutionary trace method)根据进化时间和蛋白功能之间的关系研究序列中的改变，鉴定出对功能并不重要的蛋白表面的区，参见，例如 Lichtarge et al., 1996。在此方法中，通过将进化突变和结构同源物对比，发现对结构或功能重要的氨基酸残基。所述聚合酶是进行此类研究的理想系统，因为有许多晶体和共晶体结构和许多可得的序列。本发明人突变/标记位点的选择排除了结构/功能重要的区域。此外，对不同构象状态的可得的聚合酶晶体结构的视觉观察和覆盖，对鉴定靠近 dNTP 结合位点的氨基酸位点有进一步的帮助。所选氨基酸位点中的一些位点定位于稍微偏内部，优选环绕聚合酶中的活性区，其在碱基整合中经历改变，如 dNTP 结合区，碱基整合区，焦磷酸释放区，等。优选这些内部位点，是因为这些位点上的标记在检测过程中显示降低的背景信

号,即当使用荧光标记的 dNTP 时,减少了聚合酶和非特异性连接标记的 dNTP 之间的相互作用。

一旦制备了标记的突变体聚合酶,且最小化了全溶剂环境中的能量,由于突变和/或标记对聚合酶结构的影响的估算可提供有关相对的标记定位和分隔的信息。然后此数据用于检测前估计 FRET 效率。当然,如果 dNTP 用猝灭剂标记,这些考虑是不重要的。

本发明的另一方面涉及构建对于用于标记 dNTP 和聚合酶的原子和/分子标记如荧光标记的分子力学力场参数(molecular mechanics force field parameters),和对于聚合酶上荧光标记的氨基酸和/或 dNTP 的参数。力场参数采用量子力学研究以获得相关的分子内构象的局部电荷分布和能量(即对于二面角限定),所述构象得自己知的聚合酶晶体结构。

采用静电模型评估每一可离子化残基的离子化状态,其中采用 UHBD 程序处理蛋白为低介电区,溶剂为高电介质。参见例如, Antosiewicz et al., 1994; Briggs and Antosiewicz, 1999; Madura et al., 1995。通过对每一残基解 Poisson-Boltzmann 方程计算每一可离子化残基的离子化的静电自由能。调整这些离子化的自由能以将偶合的滴定方式(titration behavior)考虑在内,所述方式导致一组前后一致的(self-consistent)预计的离子化状态。然后重新计算这些预计的离子化自由能。重新计算这些预计的离子化自由能以考虑到由于 dNTP 结合引起的离子化的改变。对未预期的离子化状态进行进一步的计算和实验研究,产生蛋白中每一残基的一组局部电荷,即,根据结合标记或氨基酸取代的类型,蛋白中的每一可离子化的残基都能具有不同的带电状态。

为进一步有助于氨基酸位点选择,主要采用 UHBD 程序通过溶剂,和任选通过溶解的离子(即,离子强度)筛选,从 *Taq* 聚合酶/DNA 复合物的分子表面性质得到静电势图。所述图提供了有关 dNTP 结合位置和在建议的突变/标记位点处的静电环境的指导。

设计产生的分子模型,考虑到新的实验数据进行持续的改进,从而改进分子模型,改进分子动力学计算和改进力场参数,以使所述模型能较好地预测系统行为以改进标记化学和/或标记定位,预计新的聚合酶突变体,碱基整合速率和聚合酶忠实度。

分子停靠模拟用于预计在聚合酶结合袋(binding pocket)内,天然和

荧光标记的 dNTP 的停靠方向。最佳停靠构型是在明确的溶剂环境存在下能量最小化的。与所选进行标记的聚合酶中氨基酸位点相结合，停靠研究用于分析标记之间如何相互作用，以及预计每一选择的氨基酸位点的 FRET 效率。

除静电计算外，所有的停靠，量子力学，分子力学，和分子动力学计算都是和将是采用 HyperChem(v6.0) 计算机程序计算。HyperChem 软件在 PC 中运行于 Windows 操作系统下。多个用于数据分析或用于 FRET 预测(如下所述)的计算程序是和将是在采用 Linux 操作系统和运行于 Linux 下的 UHBD 程序在 PC 上完成的。

聚合酶结构分析

对于 DNA 聚合酶 I(DNA pol I)(来源于大肠杆菌，水生栖热菌(*T. aquaticus*)，嗜热脂肪芽孢杆菌(*B. stearothermophilus*)，T7 噬菌体)，和人 pol β 所得到的共结晶结构表明(复制型)聚合酶具有共同的力学和结构特点。聚合酶保持‘开放式’(非有效的)构象和‘闭合’(有效的)构象的使 *Taq* DNA 所述这些结构对于鉴定在碱基整合中发生变化的聚合酶的区域有重要意义。添加核苷酸到聚合酶/引物/模板复合物中是从其从开放构象转变为闭合构象的原因。这些结构的对比提供了有关在核苷酸整合中聚合酶内部发生的构象变化的信息。具体地，在闭合构象中，手指状结构域(finger domain)的顶端向内旋转 46°，从而定位 dNTP 于聚合酶活性位点中的引物链的 3' 末端。此末端碱基对的几何形状精确相符合于其结合袋的形状。正确，互补的碱基的结合有利于闭合构象的形成，而错误的 dNTP 结合不诱发此构象变化。当所述酶处于闭合构象时，发生化学反应。

参考图 2，*Taq* DNA pol I 的大片段(KLenTaq 1)开放和闭合的三元复合体(ternary complex)形式以它们的 C α 绘图的叠印形式显示。所述三元复合体包括酶，ddCTP 和引物/模板双螺旋 DNA。开放式结构显示为洋红色，闭合结构显示为黄色。蛋白左上方部分外观杂乱显示‘手指’区处于开放和闭合构象的运动中。

从包含引物和结合的 ddGTP 的 *Taq* 聚合酶的两种不同晶体结构，采用程序确定氨基酸位置在聚合酶的开放和闭合构象中相对于结合的 ddGTP 的 γ 磷酸的变化，鉴定突变和标记位置发生最大变化的 20 个氨基酸位点。对

于每一氨基酸计算它们的 α 和 β 碳原子与结合的 ddGTP 的 γ 磷酸基之间的距离。表 I, II, III 和 IV 给出了对于 *Taq* 聚合酶得自两个不同组的结晶学数据的列表。

表 I: 相对于残基的 α 碳在聚合酶的开放形式到闭合形式之间的 $2ktq$
数据经历最大位置变化的 20 个氨基酸位点

| 残基位置 | 残基身份 | 距离的变化 (Å) | 残基位置 | 残基身份 | 距离的变化 (Å) |
|------|------|--------------|------|------|--------------|
| 517 | 丙氨酸 | 9.10 | 491 | 谷氨酸 | 2.90 |
| 516 | 丙氨酸 | 6.86 | 486 | 丝氨酸 | 2.78 |
| 515 | 丝氨酸 | 6.53 | 490 | 亮氨酸 | 2.62 |
| 513 | 丝氨酸 | 6.40 | 586 | 缬氨酸 | 2.61 |
| 518 | 缬氨酸 | 5.12 | 492 | 精氨酸 | 2.60 |
| 514 | 苏氨酸 | 3.94 | 462 | 谷氨酸 | 2.59 |
| 488 | 天冬酰胺 | 3.73 | 483 | 天冬酰胺 | 2.47 |
| 487 | 精氨酸 | 3.50 | 685 | 脯氨酸 | 2.46 |
| 489 | 谷氨酰胺 | 3.13 | 587 | 精氨酸 | 2.44 |
| 495 | 苯丙氨酸 | 3.05 | 521 | 丙氨酸 | 2.38 |

表 II: 相对于残基的 β 碳在聚合酶的开放形式到闭合形式之间的 $2ktq$
数据经历最大位置变化的 20 个氨基酸位点

| 残基位置 | 残基身份 | 距离的变化 (Å) | 残基位置 | 残基身份 | 距离的变化 (Å) |
|------|------|--------------|------|------|--------------|
| 517 | 丙氨酸 | 10.98 | 491 | 谷氨酸 | 3.41 |
| 516 | 丙氨酸 | 9.05 | 587 | 精氨酸 | 3.39 |
| 515 | 丝氨酸 | 8.02 | 521 | 丙氨酸 | 3.33 |
| 513 | 丝氨酸 | 7.46 | 498 | 亮氨酸 | 3.21 |
| 518 | 缬氨酸 | 5.47 | 489 | 谷氨酰胺 | 3.08 |
| 685 | 脯氨酸 | 5.16 | 514 | 苏氨酸 | 2.97 |
| 487 | 精氨酸 | 4.24 | 581 | 亮氨酸 | 2.93 |
| 495 | 苯丙氨酸 | 3.94 | 483 | 天冬酰胺 | 2.92 |
| 488 | 天冬氨酸 | 3.88 | 497 | 谷氨酸 | 2.91 |
| 520 | 谷氨酸 | 3.66 | 462 | 谷氨酸 | 2.83 |

表 III: 相对于残基的 α 碳在聚合酶的开放形式到闭合形式之间的 3ktq
数据经历最大位置变化的 20 个氨基酸位点

| 残基位置 | 残基身份 | 距离的变化 (Å) | 残基位置 | 残基身份 | 距离的变化 (Å) |
|------|------|--------------|------|------|--------------|
| 517 | 丙氨酸 | 8.95 | 515 | 丝氨酸 | 6.36 |
| 656 | 脯氨酸 | 8.75 | 653 | 丙氨酸 | 6.16 |
| 657 | 亮氨酸 | 8.59 | 661 | 丙氨酸 | 5.94 |
| 655 | 天冬氨酸 | 8.05 | 652 | 谷氨酸 | 5.44 |
| 660 | 精氨酸 | 7.35 | 647 | 苯丙氨酸 | 5.25 |
| 658 | 甲硫氨酸 | 7.06 | 649 | 缬氨酸 | 5.22 |
| 659 | 精氨酸 | 6.69 | 518 | 缬氨酸 | 5.15 |
| 654 | 缬氨酸 | 6.60 | 644 | 丝氨酸 | 5.08 |
| 513 | 丝氨酸 | 6.59 | 643 | 丙氨酸 | 5.01 |
| 516 | 丙氨酸 | 6.57 | 650 | 脯氨酸 | 4.72 |

表 IV: 相对于残基的 β 碳在聚合酶的开放形式到闭合形式之间的 3ktq
数据经历最大位置变化的 20 个氨基酸位点

| 残基位置 | 残基身份 | 距离的变化 (Å) | 残基位置 | 残基身份 | 距离的变 化(Å) |
|------|------|--------------|------|------|--------------|
| 517 | 丙氨酸 | 10.85 | 654 | 缬氨酸 | 6.25 |
| 656 | 脯氨酸 | 9.05 | 653 | 丙氨酸 | 6.14 |
| 657 | 亮氨酸 | 8.75 | 661 | 丙氨酸 | 6.04 |
| 516 | 丙氨酸 | 8.68 | 643 | 丙氨酸 | 5.74 |
| 655 | 天冬氨酸 | 8.24 | 649 | 缬氨酸 | 5.55 |
| 515 | 丝氨酸 | 7.92 | 647 | 苯丙氨酸 | 5.45 |
| 660 | 精氨酸 | 7.89 | 518 | 缬氨酸 | 5.42 |
| 513 | 丝氨酸 | 7.60 | 652 | 谷氨酸 | 5.13 |
| 659 | 精氨酸 | 6.98 | 644 | 丝氨酸 | 4.89 |
| 658 | 甲硫氨酸 | 6.77 | 487 | 精氨酸 | 4.77 |

上面所列氨基酸代表优选的氨基酸位点用于半胱氨酸取代和随后的标记连接，因为这些位点代表 *Taq* 聚合酶中的位点，在碱基整合中经历显著的变化。

为进一步改进氨基酸位点的选择，对处于开放和闭合构象端 (extreme) 的聚合酶，观察那些鉴定的氨基酸位点，以使最后挑选的氨基酸位点，当被修饰以携带荧光标记采用 FRET 方法进行分析时，最大化信号和最小化背景噪音。预计不会显著影响蛋白二级结构或活性的氨基酸改变组成了在 *Taq* 聚合酶中的一组优选的氨基酸位点以进行诱变和荧光修饰，这样，所述标记可以被遮蔽，从而免与游离的 dNTP 发生相互作用。下列 3 组说明本发明所用方法以从以上所列氨基酸列表中估选氨基酸位点，所述列表中的氨基酸在聚合酶从开放形式转化为闭合形式时在相对于结合的 ddGTP 的位置经历最大改变。

参照图 3A-C，显示了 3ktq (闭合的‘黑’) 和 1tau (开放的‘浅蓝’) 之间的重叠，*TAQ* DNA 聚合酶 I 的大片段。参看图 3A，来自 3ktq 的结合的 DNA 显示为红色而与 3ktq 结合的 ddCTP 是绿色。当聚合酶从开放 (1tau) 变为闭合 (3ktq) 时，可发现 3 个残基移动最大，即 Asp665, Pro656, 和 Leu657。基于进一步的结构分析，Pro656 显示具有加帽 (capping) O-螺旋的作用。Leu657 的侧链非常接近闭合 (3ktq) 形式的蛋白的另一部分。添加较大的侧链/标记被认为减小了聚合酶获得完全闭合的，活性构象的能力。相反地，Asp665 在聚合酶闭合和开放构象中是完全溶剂暴露的 (solvent exposed)。参看图 3B，显示了 *Taq* 聚合酶的 3ktq (闭合) 和 1tau (开放) 构象的重叠部分的活性位点的近视图。开放和闭合构象之间的大的位移是明显的。参看图 3C，3ktq (缺少 DNA 和 ddCTP) 的分子表面图像的近视图。所述分子表面两个区域被染色，对于 Asp655 为蓝色，对于 Leu657 为绿色。在此图像中，相信 Leu657 极接近所述蛋白的另一部分，因为在拇指结构域中，分子表面的绿色部分被‘连接’到手指区的部分。此图显示了聚合酶的此区域，观察手掌区，右边的手指区，左边的拇指区。

聚合酶变体的诱变和测序

得到编码 *TAQ* DNA 聚合酶的基因并在大肠杆菌菌株 DH1 的 pTTQ18 中表达。参见，例如，Engelke 等，1990。本发明人已经证实用于诱变的备选的

氨基酸包括表 I-IV 中所列氨基酸，其改进的列表或混合物或组合。本发明人采用本领域已知的标准分子学方法分别在每一靶氨基酸位点引入半胱氨酸密码子。参见 Sambrook et al., 1989 和 Allen et al., 1998。从表达突变体聚合酶的分离的克隆纯化 DNA，采用染料-终止子荧光化学测序，在 ABI PRISM377 自动测序仪上检测，采用 GeneCodes, Inc. 的 Sequencher™ 进行分析。

酶变体的表达和纯化

本发明人已经证明 *Taq* 酶能整合 γ 标记的 dNTP 以合成延伸的 DNA 序列。下一步涉及构建突变体，所述突变体能携带设计与 dNTP 上标记相互反应的标记，和优化聚合酶用于单分子测序。采用如上所述和实验部分的标准定点诱变构建突变体。然后将所述构建体插入到大肠杆菌，并表达。在充足的大肠杆菌生长以进行随后的聚合酶分离和纯化后，得到突变体 *Taq* 聚合酶。

尽管通过计算机控制基于反馈的非发酵基体的供应来培养大肠杆菌到光密度超过 100，但所得的 3kg 大肠杆菌细胞团在聚合酶优化过程中是过量的。当然，制备优化的聚合酶构建体时，需要使用如此大的产量。在产生优化的聚合酶过程中，从培养在 10L 供氧充分的采用得自 Amgen 的富集培养基的分批培养液中的大肠杆菌细胞团块得到突变体。为了快速筛选聚合酶突变体，在 2L 带挡板的摇瓶中培养大肠杆菌来制备突变体。用 6L 的制备式离心机收获细胞团，通过 French 压碎器裂解，离心去除细胞碎片。为减少大肠杆菌核酸序列的干扰，优选也除去其它核酸。采用核酸酶(随后热变性核酸酶)或，优选采用如 Murphy et al., Nature Biotechnology 17, 822, 1999，所述的基于压缩试剂的核酸沉淀改进方法。

因为 *Taq* 聚合酶的热稳定性明显大于一般的大肠杆菌蛋白，从污染 *Taq* 聚合酶的蛋白中纯化 *Taq* 聚合酶或其变体可通过粗聚合酶在 75°C 下 60 分钟的热处理来完成，这降低了大肠杆菌蛋白污染约 100 倍。此大肠杆菌蛋白污染的降低以及高的初始表达水平，在方便的起始步骤产生几乎纯的 *Taq* 聚合酶或其变体；当然突变体聚合酶仍保持天然聚合酶的热稳定性。

对于常规测序和 PCR 筛选，通常需要进一步的有限纯化。单步离子交换，一般在 pH8.0 的 Q Sepharose 上，通常足以得到对这些实验足够纯的

产品。优选的，也进行第二步纯化步骤，以确保污染不会模糊随后的实验结果。第二步纯化涉及 SDS-PAGE 和 CD-监测的解链实验。

dNTP 位点选择以接受荧光标记

采用 AutoDock 计算机程序进行分子停靠模拟以预测天然和荧光标记的 dNTP 的停靠方向 (Morris et al., 1998 ; Soares et al., 1999)。在停靠模拟中利用 AutoDock 计算机程序中采用的有效 Lamarckian Genetic 算法可允许存在构象的适应性。也进行一亚组蛋白侧链的移动以在其停靠时容纳 dNTP。那么，最好的停靠构型是在溶剂环境下能量最小化的。得到的实验数据可鉴定聚合酶活性中心的氨基酸，这些氨基酸与催化有关，并与模板/引物 DNA 链或要整合的 dNTP 接触。使用计算机辅助化学模型如停靠研究鉴定和支持 dNTP 的位点，所述位点可被标记并预计在特异性位点携带特异性标记的 dNTP 的 FRET 效率。

通常，通过将 dNTP 和所需标记反应或通过将前体如焦磷酸基或碱基与所需标记反应，然后完成 dNTP 的合成来标记 dNTP。

用于 DNA 聚合酶反应的核苷酸的化学修饰

本发明人已经开发了合成方法用于修饰荧光团和荧光能量传递化合物使之对于不同的信号检测具有不同的光学性质，对于核苷酸/核苷合成子，在碱基，糖或磷酸骨架位置的加入修饰，并产生在核碱基 (nucleobases)，糖或磷酸骨架上包含取代的四种脱氧核苷酸三磷酸 (dNTP) 的互补组。

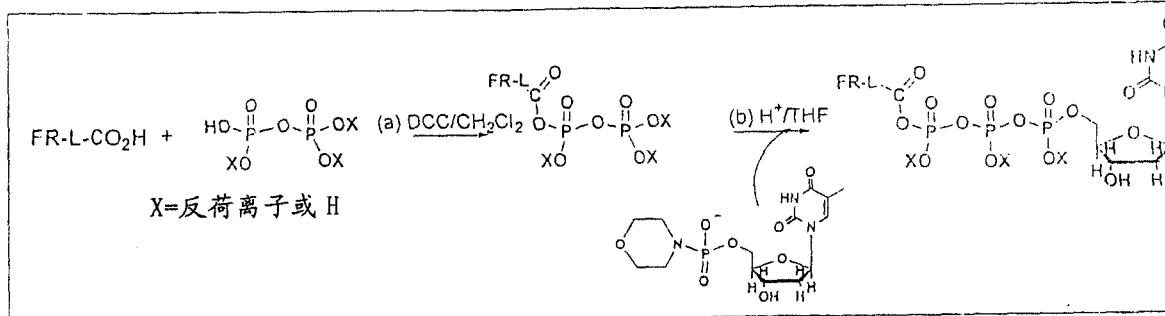
γ -磷酸修饰的 dNTP 的合成

本发明人已经发现天然 *Taq* 聚合酶能够聚合磷酸修饰的 dNTP 或 ddNTP。因为复制的 DNA 不包含非天然碱基，优选在 β 和/或 γ 磷酸基标记 dNTP 或 ddNTP，这样聚合酶活性不受显著不良影响，可产生长 DNA 链。本发明人合成了 γ -ANS-磷酸 dNTP，其中所述 ANS 通过磷酰胺键连接到磷酸。尽管这些标记的 dNTP 容易通过天然 *Taq* 聚合酶和 HIV 逆转录酶整合，但 ANS 只是多种可以通过 β 和/或 γ 磷酸基连接的标记中的一种。

本发明采用标记的 dNTP 或 ddNTP 与聚合酶联合进行信号检测。通过共价键或亲合连接在磷酸位置 (α , β 和/或 γ) 和/或其它核苷酸位置修饰 dNTP。

设计标记使记其在下一单体加入序列前从碱基除去。除去标记的一个方法是将标记放置于 γ 和/或 β 磷酸上。当焦磷酸从增长的 DNA 序列解离时，所述标记被除去。另一方法是通过可分裂键将标记结合到单体上的一个位置。在整合之后和在下一单体整合之前除去所述标记，采用光，聚合介质中的化学键裂解试剂，和/或热分裂可分裂键。

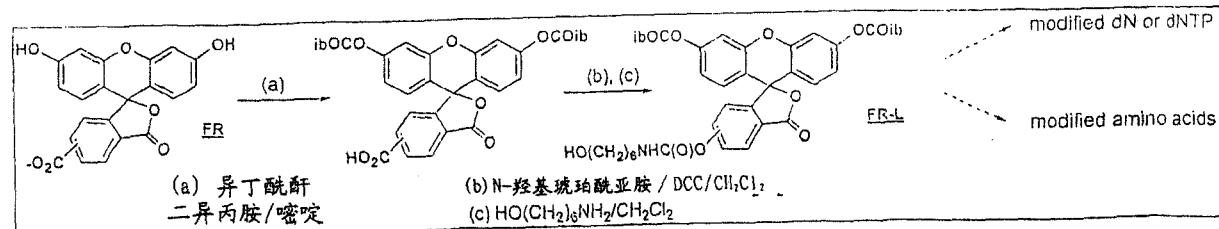
下面给出一个对于合成其它 γ -标记的 dNTP 的通用合成路线：



其中 FR 是荧光标记，L 是连接基团，X 根据反应介质的 pH 值是 H 或反荷离子 (counterion)，Z 是能与焦磷酸的羟基反应的集团，Z' 是与 dNMP 反应后的基团。优选地，Z 是 Cl, Br, I, OH, SH, NH₂, NHR, CO₂H, CO₂R, SiOH, SiOR, GeOH, GeOR, 或类似的反应功能团或离去基团，其中 R 是烷基，芳基，芳烷基，烷芳基，其卤化类似物或杂原子类似物，Z' 是 O, NH, NR, CO₂, SiO, GeO, 其中 R 是烷基，芳基，芳烷基，烷芳基，其卤化类似物或杂原子类似物。

所述合成涉及在 DCC 和二氯甲烷中将 Z 为末端的荧光标记，FR-L-Z 与焦磷酸基，P₂O₆X₃H 反应，以产生荧光标记的焦磷酸。在制备荧光标记的焦磷酸后，其与吗啉为末端的 dNMP 在酸性 THF 中反应以产生在其 γ -磷酸上具有荧光标记的 dNTP。因为最终的反应带有荧光标记，并要大于起始物质，所以可直接进行从未修饰的起始物质和标记的焦磷酸的分离。

下面图示 FR-L 基的普遍合成方法：



首先在二异丙胺存在下荧光素与嘧啶中的异丁酰酐 (isobutyryl

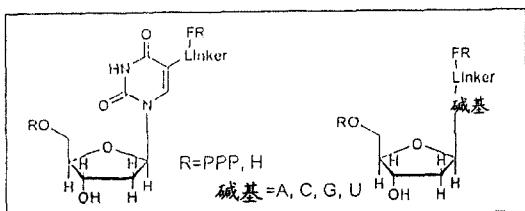
anhydride) 反应以产生具有双环羟基(已保护用于随后的接头连接)的荧光素。所述羟基保护的荧光素然后与 DDC 和二氯甲烷中的 N-羟基琥珀酰亚胺反应，随后加入 1-羟基-6-氨基己烷以产生羟基为末端的 FR-L 基。此基团然后与焦磷酸反应以在它们的 γ -磷酸基标记 dNTP 或标记氨基酸。参见，例如，Ward 等，1987；Engelhardt 等，1993；Little 等，2002；Hobbs，1991。

通过对每一 dNTP 使用不同的荧光标记，可设计标记使每一标记发射不同的发射光谱。可通过产生具有非重叠发射频率 (non-overlapping emission frequencies)-多色 (multicolor) 的标记来区分发射光谱，或者每一标记可具有非重叠发射光谱特点，如独特的发射波段，独特的吸收波段和/或独特的强度特征。对每一 dNTP 采用不同的标记的系统改进了与这样碱基的算法有关的置信度值。

上述荧光素的合成方案也适于其它染料，如四氯荧光素 (JOE)，或 N, N, N', N' - 四甲基-6-羧基若丹明 (TAMRA)。一般地，在碱性水溶液和碳化二亚胺，如 DEC 中进行 γ -磷酸标记反应。其它荧光团分子和 dNTP 可被类似地修饰。

碱基标记的 dNTP 的合成

尽管优选在 β 和/或 γ 磷酸标记 dNTP，但也可在碱基和/或糖部分标记 dNTP 同时保持它们的聚合酶反应活性。优先挑选用于修饰的位点不干扰 Watson-Crick 碱基配对。下面表示用于碱基修饰的普遍方案。



采用荧光标记酶进行聚合酶活性分析

聚合酶发展过程中监测聚合酶变体的活性。在备选氨基酸突变为半胱氨酸和荧光标记半胱氨酸之后检测聚合酶活性。用于监测天然 Taq 聚合酶整合荧光标记的 dNTP 的能力的检测也可用于筛选聚合酶变体。由于 Taq 聚

合酶变体改变了氨基酸序列，所述检测提供了突变体特征数据，如热稳定性，忠实度，聚合速率，相对于天然碱基，经修饰的碱基的亲和性。

在与用于分析通过天然 *Taq* 聚合酶将荧光标记的 dNTP 整合到 DNA 聚合物的类似的条件下进行突变体 *Taq* 聚合酶活性检测。为测定突变体 *Taq* 聚合酶活性，在聚合酶反应缓冲液中与 5' -³²P 末端标记的引物/单链模板双螺旋和适当标记的 dNTP 一起孵育纯化的突变体 *Taq* 聚合酶。在 ABI 377 DNA 测序仪（用于荧光标记的碱基），Fuji BAS1000 磷光成像系统 (phosphorimaging system)，或其它合适或类似的检测器或检测系统上，通过检测与延伸的引物有关的荧光相对量监测聚合酶的整合荧光标记的 dNTP 的能力。此检测用于确认突变体聚合酶整合了标记的 dNTP 并确认碱基整合中已获得了荧光的特征。这些检测使用末端标记引物 (end-labeled primer)，荧光标记的 dNTP 和无荧光标记的合适的碱基。然后所述产物进行按大小分离并用于延伸分析。按照需要，所述反应在固定温度反应条件或热循环下进行。

引物延伸分析

采用与通过荧光标记的 DNA 聚合酶检测单个碱基整合的类似方法检测 *TAQ* DNA 聚合酶整合 γ -磷酸 dNTP 变体的能力。参见例如 Furey 等，1998。这些实验表明不能推理带有荧光标记的聚合酶降低了聚合活性。本发明人已经表明天然 *Taq* 聚合酶整合 γ 标记的 dNTP，单独地或共同地产生长 DNA 链。

为测定聚合酶活性，在得自 Promega Corporation of Madison, Wisconsin 的聚合酶反应缓冲液中与 5' -³²P 或荧光末端标记的引物 (TOP) / 单链模板 (BOT- 'X') 双螺旋和表 V 所示合适的 dNTP 一起孵育聚合酶。按照需要，反应在固定温度反应条件或热循环下进行。所述反应产物进行按大小分离并采用磷光成像或荧光检测系统进行定量。通过与其天然对应物 (counterpart) 对比确定每一标记的 dNTP 整合的相对效率。

表 V

引物链：

顶部 5' GGT ACT AAG CCG CCG CAT G 3'

模板链：

BOT-T 3' CCA TGA TTC GCC GGC GTA CTC 5'
 BOT-C 3' CCA TGA TTC GCC GGC GTA CCC 5'
 BOT-G 3' CCA TGA TTC GCC GGC GTA CGC 5'
 BOT-A 3' CCA TGA TTC GCC GGC GTA CAC 5'
 BOT-3T 3' CCA TGA TTC GCC GGC GTA CTT TC 5'
 BOT-Sau 3' CCA TGA TTC GCC GGC GTA CCT AG 5'

表 V 中，‘TOP’ 表示检测 DNA 双螺旋的引物链。模板链的变体由‘BOT’代表。DNA 模板的相关特点在连字号后给出。例如，BOT-T, BOT-C, BOT-G, BOT-A 用于分别对 dA, dG, dC, 和 dT 的多个核苷酸或核苷酸变体监测聚合酶整合效率和忠实度。

在彻底纯化标记的 dNTP 之前，采用‘BOT-Sau’ 模板进行初步分析以确保聚合酶不被与标记的 dNTP 一起共纯化的化学物质抑制。设计‘BOT-Sau’ 模板以监测标记的 dATP 之前天然 dGTP 的整合(即，聚合酶活性的阳性对照)。对有希望的标记的核苷酸进行更进一步的纯化。类似地，进行实验以确定标记的 dNTP 整合后聚合酶是否继续延伸，实验采用相同的末端标记的‘TOP’ 引物，合适的‘BOT’ 引物，荧光标记的 dNTP 和标记的核苷酸的适当 3’ 碱基，单独或一起进行。接着按大小分离产物并分析确定相对的延伸效率。

分析 γ -磷酸标记的核苷酸整合的忠实度

TAQ DNA 聚合酶缺少 3’ 到 5’ 外切核酸酶活性(校对活性)。如果在单个分子 DNA 测序中所用聚合酶具有 3’ 到 5’ 外切核酸酶活性，所述聚合酶将能添加另一碱基以取代由于校对活性被除去的碱基。此新添加的碱基能产生特征荧光信号证实模板中加入的碱基的整合，产生错误鉴定的 DNA 序列，这样会使本发明的单分子测序系统出现问题。

如果修饰的 dNTP 的整合错误率超过 100 中约 1 个错误的阈值水平，测序反应优选并列进行，采用产生序列信息所需的优选数目，所述信息对每一由错误率确定的碱基配对具有高忠实度。较大的错误率需要更多的平行试验，而较小的错误率需要较少的平行试验。实际上，如果错误率足够低，

通常小于 1000 个整合碱基中 1 个错误, 优选 5000 中 1 个错误, 特别是 10000 中 1 个错误, 那么不需要平行试验。插入或缺失潜在的是更严重的错误类型, 要保证每个样品最少有 3 个重复。如果进行 2 个反应, 不能确定哪一个是正确的。因此, 此系统产生高质量数据需要 3 个反应。

BOT 变体模板用于表征由表 V 所示由工程化聚合酶整合每一 γ 标记的 dNTP 的正确度。寡核苷酸用作 DNA 模板, 每一个只在整合的第一碱基不同。采用这些模板的试验用于测定每一碱基的相对整合效率, 以及聚合酶区分标记的 dNTP 的能力。开始, 采用聚合酶变体的实验采用相对简单的序列, 单链 DNA 模板进行。从 University of Houston 的 Dr. Hardin's laboratory 获得大量具有序列特征的模板, 包括超过 300 个纯化模板。例如一系列包含长度可变的 polyA 或 polyT 序列。如果需要, 构建额外的限定的序列模板, 利于进行选择碱基的算法。

相对荧光强度分析

采用 SPEX 212 仪器或类似的仪器通过测量溶液荧光, 得到聚合酶对标记的 dNTP 作用的直接测定。此仪器成功地用于检测来自 ANS 标记的 γ -磷酸 dNTP 的荧光信号, 所述 γ -磷酸 dNTP 被毫微摩尔浓度的 *Taq* 聚合酶整合。所述 SPEX 212 仪器包括 450 瓦氙弧光源, 双重发射并双重激发的单色仪 (dual emission and dual excitation monochromators), 冷却的 (cooled) PMT (最近升级为同时 T-模式各向异性数据收集 (anisotropy data collection)), Hi-Tech 终止流动附件 (stopped-flow accessory)。此仪器能检测在标记的焦磷酸从 ANS 标记的 γ -磷酸 dNTP 释放时荧光强度的增长和/或吸收光谱的变化, 如对由 *Taq* 和 RNA 聚合酶和蛇毒磷酸二酯酶释放的 ANS 焦磷酸所证实的。

通过在聚合酶 (对照: 没有酶) 和 DNA 引物/模板 [poly(dA).poly(dt)] (对照: 没有引物/模板 DNA) 存在下, 合适的缓冲液 (即得自 Promega 公司的缓冲液) 中孵育 γ 磷酸标记的 dATP 或 TTP (对照: 未修饰的 dATP 和 TTP), 已经或正在进行实验。当所述聚合酶整合标记的 dNTP 时, 检测荧光强度和/或频率, 吸收和/或发射光谱, 以及 DNA 聚合体浓度的变化。实验试管与对照试管相对比时, 作为实验时间和/或温度的函数的测量值的变化可以明确确定聚合酶是否整合 γ 磷酸标记的 dNTP。可基于这些测量值的变化对每

一标记的 dNTP 进行激发和荧光发射的优化。

单个分子检测系统的开发

优选采用显微镜方法进行单个分子荧光的检测。在本发明中可以使用共焦扫描显微镜(Confocal-scanning microscopy)，但优选非扫描方法。由于聚合酶活性可用于检测荧光信号的显微镜包括任何类型的显微镜，优选浸油(oil-immersion)型显微镜。显微镜的优选环境为振动和温度的变化受到控制的环境，并配有高灵敏度数字相机。虽然很多相机可用于记录荧光信号，优选 CCD 增强型，如 Princeton Instruments 的 iPentaMax。

检测方法涉及以足以诱发标记荧光的波长照射样品，优选以内反射形式(internal-reflection format)。如果荧光标记是供体-受体对，那么激发频率必须足以激发供体。尽管可使用任何类型的光源，优选的光源是激光。将同一样品在多个荧光发射波长下快速连续或同时成像经常是有利的。对于同时多色成像，优选的图像分离器(image splitter)可以使 CCD 同时捕获所有的颜色图像。备选地，可使用多个相机，每一个透过不同波长特异性的光学滤镜观察样品。

标记检测实际上，当然，有赖于许多变量，还包括所用特异性标记，电，荧光，化学，物理，电化学，质量同位素(mass isotope)，或其它性质。利用研究级 Nikon Diaphot TMD 倒置外荧光显微镜(inverted epifluorescence microscope)得到单个分子荧光成像，用激光照射和更灵敏的相机升级。而且，单个分子技术是已充分发展并商业化的技术。参见例如，Peck et al., 1989; Ambrose et al., 1994; Goodwin et al., 1997; Brouwer et al., 1999; Castro and Williams, 1997; Davis et al., 1991; Davis et al., 1992; Goodwin et al., 1997; Keller et al., 1996; Michaelis et al., 2000; Orrit and Bernard, 1990; Orrit et al., 1994; Sauer et al., 1999; Unger et al., 1999; Zhuang et al., 2000。

外荧光显微镜可采用 488nm 氩离子激光改型为损耗波(evanescent-wave)激发。本发明人曾在核酸杂交研究的检测中使用此光照几何学(illumination geometry)。现有设置已用 12-bit 512 x 512 pixel Princeton Instruments I-PentaMAX generation IV 增强型 CCD 相机替换现有的 CCD 相机进行升级，其已成功用于大量类似的单个分子应用中。此

相机在所用染料发射波长的整个范围内达到量子效率超过 45%，且远远超过此范围。它们现存显微镜的垂直调正倾向于使振动问题最小化，且所述仪器现安装在抗震桌上。

优选的高灵敏度成像系统基于 Olympus IX70-S8F 倒置外荧光显微镜。此系统整合了低背景组件并以大于每秒 80 幅的速率捕捉单个分子荧光图像，其量子效率在荧光活性标记的发射波长范围的 60%-70% 之间。

在多个单个分子荧光成像中，优选最小化数据采集通道内多个荧光发射器的出现，如 CCD 或其它数字成像系统视野的单个像素或像素箱 (pixel-bin)。有限数量的数据采集通道如像素在任何给定的数字成像装置中可以获得。随机间隔，密集定位的荧光发射器通常产生增强的像素或像素箱的折射，它们被多重占据 (multiply-occupied)，在数据分析中是成问题的。当视野中发射器的密度提高时，成问题的数据通道也增加。而通过降低视野中发射器的密度可以降低视野内不同数据采集区的多重占据，此发射器密度的降低可以提高数据采集通道或像素的折射，使根本看不到发射器，以此导致无效的数据采集。

提高和/或最大化数据采集效率的优选方法包括控制发射器 (标记的聚合酶分子) 间隔。以几种方法达到此间隔。首先，固定所述聚合酶在基体上以使只有单个聚合酶位于每一数据采集通道内或成像系统视野内的像素区。通过锚定捕获试剂或连接基团化学连接到所述基体完成固定。通过挑选天然的大捕获试剂，通过将它们偶联或将它们结合到增强它们空间体积的或静电排斥体积的分子，或通过在选择的最大化聚合分子间聚集 (例如，低离子强度以最大化静电排斥) 的排斥条件下固定，可以间隔捕获或连接试剂到有用的距离。

备选地，所述聚合酶可以与有关蛋白相联，所述蛋白提高了聚合酶的空间体积或聚合系统的静电排斥体积以使每一聚合分子间的聚集不能接近比成像系统数据通道分辨尺寸更大的距离。

采用单个分子检测系统的聚合酶活性检测

这些检测基本如本文所述对聚合酶活性的检测进行。如上所述，用于筛选目的的检测聚合酶活性间的主要差异涉及固定聚合体的某部分，如聚合酶，靶 DNA 或引物连接蛋白到固体支持物上，以能观察到单个的复制事

件。可以有多种固定方法可供选择，包括但不限于在纳米管 (nanotube)，或其它类似的纳米结构之内或之上和/或在多孔基质之内或之上，使一种分子聚合体共价和/或非共价连接到表面如有机表面，无机表面。可以设计这些固定化技术以提供特定区域用于所述可检测性质的检测，如荧光，NMR，或类似的，其中所述间隔足以减少或最小化具有多个发射器的数据采集通道。这样，用于单分子测序的优选数据采集方法是确保在成像装置的视野内所述荧光标记的聚合酶被间隔开，以使每一数据采集通道只能观察到单个聚合酶的活性。

分析单分子测序系统的荧光信号

检测器产生的原始数据代表 1 到 4 个依赖于时间的荧光数据流，其包含波长和强度：一个数据流代表被监测的每一荧光标记的碱基。采用 PHRED 计算机程序计算碱基身份和可靠性。如果需要，本发明人可以书写计算机程序以翻译具有偏颇和重叠数据的数据流。在这样的情况下进行多个实验以根据与聚集片段的延伸序列有关的可靠性指数和困难性的变化给每一碱基身份指定置信界限。可靠性指数代表与理想值相比在观察到的荧光波长和强度之间的适合度。所述信号分析的结果是具有相关确定性概率 (associated probabilities of certainty) 的线性 DNA 序列。此外，在需要时，所述数据储存于数据库中用于动态查询进行鉴定和比较。6-10, 11-16, 21-30 个碱基的检索条件 (序列) 可与参考序列进行对比以快速鉴定完全匹配的序列或具有共同的使用者确定的同一性水平的序列。进行多重试验，以使可按照与聚集片段的延伸序列有关的可靠性指数和困难性的变化给每一碱基身份指定置信界限。可靠性指数代表与理想值相比在观察到的荧光波长和强度之间的适合度。所述信号分析的结果是具有相关确定性概率的线性 DNA 序列。

信息学：单个分子检测系统荧光信号的分析

为了确定模板或靶分子的整个序列，采集数据，使来自部分信息的数据汇集成从多个聚合酶分子得到的序列的信息。用多个单个分子得到的结果联系在一起的重要驱动力量是经过不确定的时间段从单个分子得到数据的可能性。在典型的光漂白 (photobleaching) 效率为 $2 * 10^{-5}$ 时，预计在永

久光漂白前，典型的染料分子经历 50,000 个激发/发射循环。采集自给定分子的数据可能由于系统间过渡为无光学活性(在目的时间范围时)的三重体(triplet)状态而被中断。因此即使采取了对光漂白的预防，得自任何给定分子的数据对于充分长度的模板序列必然是片段性的数据，为了得到靶 DNA 分子的全部序列需对这些亚序列共同处理。

此外，在某些例子中，进行与参考对照的反应，类似于微阵列分析是有用的。参考序列和试验样品之间的信号的对比用于鉴定各序列或序列组合物的不同和近似性。这样的反应用于 DNA 聚合物的快速筛选以测定各聚合物之间的同源程度，以确定 DNA 聚合物的多态性，或用于鉴定病原体。

实施例

Taq 聚合酶的克隆和诱变

克隆

从美国典型培养物保藏中心(ATCC; Manassas, VA)获得具有编码 DNA 聚合酶 I (*Taq pol I*) 的水生栖热菌全长序列的嗜菌体 λ 宿主菌株 Charon 35。从感染的大肠杆菌宿主的裂解物采用下面的 DNA 寡核苷酸引物直接扩增 *Taq pol I*:

Taq pol I 正向:

5'-gc gaattc atgaggggga tgctgccct ctttgagccc-3'

Taq pol I 反向

5'-gc gaattc accctcccttgg cggagcgc cagtcctccc-3'

每一合成的 DNA 寡核苷酸的划线部分代表紧接 *Taq pol I* 基因前后的工程化 EcoRI 限制性位点。采用如上所述的反向引物以及下述的正向引物的 PCR 扩增产生具有基因 N-末端缺失的额外的构建体:

Taq pol I -A293-trunk

5'-aatccatgggccctggaggaggc cccctggcccccg-3'

下划线部分对应具有编码丙氨酸第一密码子的工程化 NcoI 限制性位点(ATG 起始代表核糖体结合位点后的表达载体)。理想地，*Taq pol I* 基因的全长和截短的构建体连接到单个的 EcoRI 位点(全长)和 NcoI/EcoRI 消化的 pRSET-b 表达载体。大肠杆菌菌株 JM109 用作工程化载体的所有体内操纵的宿主。

诱变

一旦产生了合适的构建体，在天然 Taq 聚合酶的优选的氨基酸位点包括 513-518, 643, 647, 649 和 653-661 引入个别的半胱氨酸突变。下列氨基酸残基对应氨基酸 643 到 661 之间的氨基酸，其中 xxx 代表在天然聚合酶中插入的氨基酸残基。

643-Ala xxx xxx xxx Phe xxx Val xxx xxx Glu Ala Val Asp Pro Leu
Met Arg Arg Ala-661

重叠引物用于通过基于 PCR 的诱变(采用 Pfu DNA 聚合酶)引入点突变到天然基因。

互补正向和反向引物的每一个包含编码所需突变的氨基酸残基的密码子。采用这些引物的 PCR 产生有裂口的(knicked)，非甲基化的双链质粒，其包含所需突变。为除去模板 DNA，用 DpnI 限制酶处理整个 PCR 产物(在序列 GATC 中的甲基化鸟苷处进行酶切)。模板质粒消化后，转化突变的质粒并体内连接。

下列合成的 DNA 寡核苷酸引物用于如下所述的诱变，其中小写字母表示在指示的位点已被修饰产生所需半胱氨酸取代。然后通过自动测序筛选突变体。

丙氨酸 643 到半胱氨酸的取代

Taq Pol I⁻Ala643Cys-fwd

5'-C CAC ACG GAG ACC tgC AGC TGG ATG TTC GCC G-3'

Taq Pol I Ala643Cys-rev

5'-C GCC GAA CAT CCA CGA Gca GGT CTC CGT GTG G-3'

苯丙氨酸 647 到半胱氨酸的取代

Taq Pol I-Phe647Cys-fwd

5'-CC GCC AGC TGG ATG TgC GGC GTC CCC CGG GAG GCC-3'

Taq Pol I-Phe647Cys-rev

5'-GGC CTC CCG GGG GAC GCC GcA CAT CCA CGT GGC GG-3'

缬氨酸 649 到半胱氨酸的取代

Taq Pol I-Va1649Cys-fwd

5'-GCC AGC TGG ATG TTC GGC tgC CCC CGG GAG GCC GTG G-3'

Taq Pol I_Va1649Cys_rev

5'-C CAC GGC CTC CCG GGG Gca GCC GAA CAT CCA GCT GGC-3'

谷氨酸 652 到半胱氨酸的取代

Taq Pol I_Glu652Cys_fwd

5'-GGC GTC CCC CGG tgc GCC GTG GAC CCC CTG ATG CGC-3'

Taq Pol I_Glu652Cys_rev

5'-GCG CAT CAG GGG GTC CAC GGC gca CCG GGG GAC GCC-3'

丙氨酸 653 到半胱氨酸的取代

Taq Pol I_AlA653Cys_fwd

5'-GGC GTC CCC CGG GAG tgC GTG GAC CCC CTG ATG CGC-3'

Taq Pol I_AlA653Cys_rev

5'-GCG CAT CAG GGG GTC CAC Gca CTC CCG GGG GAC GCC-3'

缬氨酸 654 到半胱氨酸的取代

Taq Pol I_Va1654Cys_fwd

5'-GTC CCC CGG GAG GCC tgt GAC CCC CTG ATG CGC-3'

Taq Pol I_Va1654Cys_rev

5'-GCG CAT CAG GGG GTC aca GGC CTC CCG GGG GAC-3'

天冬氨酸 655 到半胱氨酸的取代

Taq Pol I_D655C_fwd

5'-CCC CGG GAG GCC GTG tgC CCC CTG ATG CGC CGG-3'

Taq Pol I_D655C_rev

5'-CCG GCG CAT CAG GGG Gca CAC GGC CTC CCG GGG-3'

脯氨酸 656 到半胱氨酸的取代

Taq Pol I_Pro656Cys_fwd

5'-CGG GAG GCC GTG GAC tgC CTG ATG CGC CGG GCG-3'

Taq Pol I_Pro656Cys_rev

5'-CGC CCG GCG CAT CAG Gca GTC CAC GGC CTC CCG-3'

亮氨酸 657 到半胱氨酸的取代

Taq Pol I_Leu657Cys_fwd

5'-GCC GTG GAC CCC tgc ATG CGC CGG GCG GCC-3'

Taq Pol I_Leu657Cys_rev

5' -GGC CGC CCG GCG CAT gca CGG GTC CAC GGC-3'

甲硫氨酸 658 到半胱氨酸的取代

Taq Pol I-Met658Cys-fwd

5' -GCC GTG GAC CCC CTG tgt CGC CGG GCG GCC-3'

Taq Pol I-Met658Cys-rev

5' -GGC CGC CCG GCG aca CAG GGG GTC CAC GGC-3'

精氨酸 659 到半胱氨酸的取代

Taq Pol I-Arg659Cys-fwd

5' -GCC GTG GAC CCC CTG ATG tGC CGG GCG GCC AAG ACC-3'

Taq Pol I-Arg659Cys-rev

5' -GGT CTT GGC CGC CCG GCa CAT CAG GGG GTC CAC GGC-3'

精氨酸 660 到半胱氨酸的取代

Taq Pol I-Arg660Cys-fwd

5' -GAC CCC CTG ATG CGC tGc GCG GCC AAG ACC ATC-3'

Taq Pol I-Arg660Cys-rev

5' -GAT GGT CTT GGC CGC gCa GCG CAT CAG GGG GTC-3'

丙氨酸 661 到半胱氨酸的取代

Taq Pol I-Ala661Cys-fwd

5' -CCC CTG ATG CGC CGG tgc GCC AAG ACC ATC AAC-3'

Taq Pol I-Ala661Cys-rev

5' -GTT GAT GGT CTT GGC gca CCG GCG CAT CAG GGG-3'

所得突变体 Taq 聚合酶然后与所需的原子或分子标记反应以通过半胱氨酸的 SH-基标记突变体结构中的半胱氨酸，并筛选天然和 / 或标记的 dNTP 整合和整合效率。碱基整合过程中聚合酶突变体也进行荧光活性筛选。因此，本发明也涉及 Taq 聚合酶突变体，其具有添加到一个或多个位点的半胱氨酸残基，所述位点选自 513-518, 643, 647, 649 和 653-661。半胱氨酸取代和采用修饰的 dNTP 证实聚合酶活性后，所述 Taq 聚合酶突变体通过插入的半胱氨酸残基的 SH-基与标记反应。

合成修饰的 dNTP

(γ -AmNS) dATP 的合成

合成核昔酸类似物，其包含连接到 γ -磷酸键的荧光团 1-氨基萘-5-磺酸 (1-aminonaphalene-5-sulfonate) (J Biol. Chem. 254, 12069-12073, 1979)。稍微改变 Yarbrough 及合作者对 (γ -AmNS) ATP 所述方法，合成 dATP 类似物- (γ -AmNS) dATP。

此实施例说明了 γ ANS 标记的 dATP 的制备，图示于图 4。

1-氨基萘-5-磺酸 (447mg, 2mmol, 40eq., 来自 Lancaster) 加入到 10ml H₂O 中，用 1 N NaOH 调 pH 为 5.8。用注射滤器除去不溶物质，得到此 pH 值的基本饱和的溶液 (~ 0.18 到 0.2M)。4 mL 12.5mM 5' 三磷酸-2'-脱氧腺苷二钠盐 (0.05 mmol, 1 eq.) 和 2mL 1M 1- (3-二甲基氨基丙基)-3-乙基-碳二亚胺盐酸盐 (DEC, 2 mmol, 40eq., 来自 Lancaster) 22°C 加入到反应容器中。加入 10ml 1-氨基萘-5-磺酸溶液引发反应，持续 2.5h。间歇加入 0.1N HCl 使 pH 保持在 5.65-5.75。2.5h 后，稀释反应液到 50ml，调整到 0.05M 三乙基铵二碳酸氢盐缓冲液中 (TEAB, pH 7.5)。所述反应产物在 50ml DEAE-SEPHADEX 离子交换柱 (A-25-120) 于低温下进行层析，采用 ~ pH 7.5 1.0 M TEAB 水溶液 (100 mL), 1.0M 碳酸氢钠水溶液 (100 mL)，和 ~ pH 7.5, 0.05 M TEAB 水溶液 (100 mL) 平衡。采用 ~ pH 7.5 TEAB 水溶液 0.05 到 0.9 M 线性梯度洗脱柱。采集约 10ml 级分。合适的稀释后得到所述级分的吸光度和荧光图 (UV 366nm)。所述荧光级分在未反应的 dATP 的峰后在 ~ 0.7M 缓冲液洗脱，显示出亮蓝色荧光。汇集含产物级分，冷冻干燥并与 H₂O/乙醇 (70/30) 共蒸发两次。残余物溶于水并冷冻干燥。³¹P NMR (¹H 去耦合 (decoupled)), 600 MHz, D₂O, Me₃PO₄ 外参比, 293 K, pH 6.1) δ (ppm) -12.60, -14.10, -25.79。参照化合物 dATP 给出下列共振峰：³¹P NMR (dATP, Na⁺) 在 D₂O 293 K, δ (ppm) -11.53 (γ), -13.92 (α), -24.93 (β)。

采用二极管阵列 UV 检测 HPLC，可以容易的通过在 366nmANS 基的不同吸光度鉴定含有所要产物的级分。此外，³¹P NMR 光谱记录了水溶液中 γ -磷酸标记的 dATP 和常规的 dATP。对于每一化合物，观察到三个特征共振，确认在 γ -标记的 dATP 中的三磷酸部分。¹H-NMR, HPLC 和 UV 光谱的结合分析提供正确化合物形成的支持信息。

相同的合成方法用于制备 γ -ANS 磷酸修饰的 dGTP, dTTP, dCTP。

通过 Taq 聚合酶整合 γ -磷酸标记的 dNTP

下述实施例表明了商业可得的 Taq 聚合酶有效整合 ANS- γ -磷酸 dNTP，如下合成并定性。

在第一个实例中，描述了 ANS- γ -磷酸 dATP 的整合以从引物模板产生延伸的 DNA 产物。在延伸缓冲液中进行所述反应，采用 20% 变性聚丙烯酰胺凝胶按大小分离所得的放射标记产物。采用磷光成像系统采集数据。参考图 5，泳道 1 在延伸缓冲液中包含 5' 放射标记的‘TOP’探针。泳道 2 包含 Taq DNA 聚合酶，50 μ M dGTP 与 DNA 双链温育(放射标记的 TOP，采用过量的‘BOT-Sau’)。泳道 3 包含 Taq DNA 聚合酶 50 μ M dATP 与 DNA 双链温育(放射标记的 TOP，采用过量的‘BOT-Sau’)。泳道 4 包含 Taq DNA 聚合酶，50 μ M ANS- γ -dATP 与 DNA 双链温育(放射标记的 TOP，采用过量的‘BOT-Sau’)。泳道 5 包含 Taq DNA 聚合酶，50 μ M dGTP 与 DNA 双链温育(放射标记的 TOP，采用过量的‘BOT-T’)。泳道 6 包含泳道 5 的溢出(spill-over)。泳道 7 包含 Taq DNA 聚合酶，50 μ M dATP 与 DNA 双链温育(放射标记的 TOP，采用过量的‘BOT-T’)。泳道 8 包含 Taq DNA 聚合酶，50 μ M ANS- γ -dATP 与 DNA 双链温育(放射标记的 TOP，采用过量的‘BOT-T’)。泳道 9 包含 Taq DNA 聚合酶，50 μ M dGTP 与 DNA 双链温育(放射标记的 TOP，采用过量的‘BOT-3T’)。泳道 10 包含 Taq DNA 聚合酶，50 μ M dATP 与 DNA 双链温育(放射标记的 TOP，采用过量的‘BOT-3T’)。泳道 11 包含 Taq DNA 聚合酶，ANS- γ -dATP 与 DNA 双链温育(放射标记的 TOP，采用过量的‘BOT-3T’)。泳道 12 在延伸缓冲液中包含 5' 放射标记的‘TOP’探针。泳道 13 在延伸缓冲液中包含 5' 放射标记的“TOP”探针和 Taq DNA 聚合酶。寡核苷酸序列表示于表 V 中。

泳道 1 和泳道 4 的定量对比表明当 ANS- γ -dATP 包含在反应中时检测到很少的非特异性，单个碱基延伸，但是第一个整合的碱基应该是 dGTP(其不添加到反应中)。通道 1 和 8 的定量分析表明当反应中包含 ANS- γ -dATP 时约 71% 的 TOP 引物通过模板指导的单个碱基延伸，第一个整合的碱基应该是 dATP。因此，Taq DNA 聚合酶整合 γ 标记的核苷酸。与聚合酶整合 γ 标记的核苷酸能力同等重要的是其在整合修饰的 dATP 之后延伸 DNA 引物的能力。泳道 1 和 11 的对比表明在整合 γ 标记的核苷酸后 DNA 链延伸。因此整合修饰的核苷酸对聚合酶活性无害。注意，通过整合 ANS- γ 核苷酸延伸引物链也是根据 Watson-Crick 碱基配对法则。实际上，通过添加此标记到

γ -磷酸，核苷酸整合的忠实度提高了至少 15 倍。

下一个实例说明了采用所有 4 种 ANS 标记的 γ -磷酸 dNTP 合成延伸的 DNA 聚合物。这些反应产生的产物在 20% 变性聚丙烯酰胺凝胶分离，干燥胶并暴露于 Fuji BAS1000 显影板 (imaging plate) 过夜显影。参考图 6，(A) 为实际胶的图像，(B) 为减弱的磷光图像和(C) 增强的磷光图像。对 A, B 和 C 的泳道描述如下：泳道 1 是包含纯化的 10-碱基的引物的对照，通过 $\alpha^{32}\text{P}$ dCTP 的模板介导的添加延伸到 11 和 12 碱基。泳道 2 包含同样的引物，其与双链质粒 DNA 在 96°C 下温育 3 分钟 (以变性模板)，将反应降为 37°C (以退火引物-模板)，加入 Taq DNA 聚合酶和所有四种天然 dNTP (每种 100 μM)，反应液在 37°C 下温育 60 分钟。泳道 3 包括同样标记的引物，其与双链质粒 DNA 在 96°C 温育 3 分钟，向反应液中加入 DNA 聚合酶和所有四种 γ -修饰的 dNTP (每种 100 μM)，并在 37°C 下 60 分钟温育该反应液。泳道 4 包括对照，纯化的 10-碱基引物，其通过 $\alpha^{32}\text{P}$ -dCTP 的添加延伸到 11 和 12 碱基，与泳道 5-8 的反应并列循环。泳道 5 包括同样的 ^{32}P -标记的引物，其与双链质粒 DNA 在 96°C 下温育 3 分钟，将反应降为 37°C 10 分钟，此期间加入 Taq DNA 聚合酶和所有四种天然 dNTP (每种 100 μM)。反应液进行 96°C 10 秒，37°C 1 分钟，70°C 5 分钟的 25 个循环。泳道 6 包括同样的 ^{32}P -标记的引物，其与双链质粒 DNA 在 96°C 下温育 3 分钟，将反应降为 37°C 10 分钟，此期间加入 Taq DNA 聚合酶和所有四种 γ 修饰的 dNTP (每种 100 μM)。反应液进行 96°C 10 秒，37°C 1 分钟，70°C 5 分钟的 25 个循环。泳道 7 包含未纯化的，10-碱基 ^{32}P -标记的引物，其与双链质粒 DNA 在 96°C 下温育 3 分钟，将反应降为 37°C 10 分钟，此期间加入 Taq DNA 聚合酶和所有四种天然 dNTP (每种 100 μM)。反应液进行 96°C 10 秒，37°C 1 分钟，70°C 5 分钟的 25 个循环。泳道 8 包含未纯化的，10-碱基 ^{32}P -标记的引物，其与双链质粒 DNA 在 96°C 下温育 3 分钟，将反应降为 37°C 10 分钟，此期间加入 Taq DNA 聚合酶和所有四种 γ 修饰的 dNTP (每种 100 μM)。反应液进行 96°C 10 秒，37°C 1 分钟，70°C 5 分钟的 25 个循环。这些涉及标记的 dNTP 的反应的证据是与涉及天然核苷酸的反应相比焦磷酸裂解 (pyrophosphorolysis) 的大幅降低。

下述实例说明了采用所有四种 ANS 标记的 γ -磷酸 dNTP 合成长 DNA 聚合物。每一引物延伸反应液分为两个部分，一部分经 20% 变性凝胶电泳 (如上所述)，而另一部分经 6% 变性凝胶电泳以更好的估计产物长度。干燥凝

胶并显影(过夜)于 Fuji BAS1000 显影板。参照图 7, (A) 为实际凝胶, (B) 为实际胶的减弱的磷光图像, (C) 为实际胶的增强的磷光图像。对泳道 A, B 和 C 的描述如下: 泳道 1 包括 123 标记物, 其大小标准指示于每一排左侧。泳道 2 包括对照, 纯化的 10-碱基引物, 其通过 α -³²P-dCTP 的模板介导的添加延伸到 11 和 12 碱基。泳道 3 包括同样的 ³²P-标记的引物, 其与双链质粒 DNA 在 96°C 下温育 3 分钟(以变性模板), 将反应降为 37°C(以退火引物-模板), 加入 Taq DNA 聚合酶和所有四种天然 dNTP(每种 100 μ M), 反应液在 37 °C 温育 60 分钟。泳道 4 包括同样的 ³²P-标记的引物, 其与双链质粒 DNA 在 96°C 下温育 3 分钟, 将反应降为 37°C 10 分钟, 此期间加入 Taq DNA 聚合酶和所有四种 γ 修饰的 dNTP(每种 100 μ M), 反应液在 37 °C 温育 60 分钟。泳道 5 包括对照, 纯化的 10-碱基引物, 其通过 α -³²P-dCTP 的模板介导的添加延伸到 11 和 12 碱基, 其与泳道 5-8 的反应并列循环。泳道 6 包括同样的 ³²P-标记的引物, 其与双链质粒 DNA 在 96°C 下温育 3 分钟, 将反应降为 37 °C 10 分钟, 此期间加入 Taq DNA 聚合酶和所有四种天然 dNTP(每种 100 μ M)。反应液进行 96°C 10 秒, 37°C 1 分钟, 70°C 5 分钟的 25 个循环。泳道 7 包括同样的 ³²P-标记的引物, 其与双链质粒 DNA 在 96°C 下温育 3 分钟, 将反应降为 37°C 10 分钟, 此期间加入 Taq DNA 聚合酶和所有四种 γ -标记的 dNTP(每种 100 μ M)。反应液进行 96°C 10 秒, 37°C 1 分钟, 70°C 5 分钟的 25 个循环。泳道 8 包含未纯化的, 10-碱基, ³²P-标记的引物, 其与双链质粒 DNA 在 96°C 下温育 3 分钟, 将反应降为 37°C 10 分钟, 此期间加入 Taq DNA 聚合酶和所有四种天然 dNTP(每种 100 μ M)。反应液进行 96°C 10 秒, 37°C 1 分钟, 70°C 5 分钟的 25 个循环。泳道 9 包含未纯化的, 10-碱基, ³²P-标记的引物, 其与双链质粒 DNA 在 96°C 下温育 3 分钟, 将反应降为 37°C 10 分钟, 此期间加入 Taq DNA 聚合酶和所有四种 γ -标记的 dNTP(每种 100 μ M)。反应液进行 96°C 10 秒, 37°C 1 分钟, 70°C 5 分钟的 25 个循环。

此反应中对天然和 γ 修饰的 dNTP 的大部分延伸产物是几百个碱基长, 这些产物中相当大的百分比由于太大不能进入凝胶。因此, 表明采用 γ -磷酸标记的 dNTP 通过 Taq 聚合酶产生了长的 DNA 聚合物, 其是无标记的或天然的 DNA 聚合物链。

不同的聚合酶与 γ -修饰的核苷酸的反应不同

在制造商建议的反应缓冲液中 37°C 温育指明的酶 (Taq DNA 聚合酶, 测序酶, HIV-1 逆转录酶, T7 DNA 聚合酶, Klenow 片段, *Pfu* DNA 聚合酶), 50 μm 指明的核苷酸 30-60 分钟, 经 20% 变性凝胶分离分析反应产物。

Taq DNA 聚合酶有效地使用了 γ 修饰的核苷酸以合成延伸的 DNA 聚合物, 准确性提高, 如图 4-6 所示。

大肠杆菌 DNA 聚合酶 I 的 Klenow 片段有效地使用了 γ 修饰的核苷酸, 但并未显示出如图 8 所示的采用其它酶观察到的忠实度的极大改进 (extreme fidelity improvement)。

Pfu DNA 聚合酶没有有效使用 γ 修饰的核苷酸, 因此不是如图 9 所示单分子测序系统的优选酶。

HIV-1 逆转录酶有效使用 γ 标记的核苷酸, 如图 10 所示显著改进了忠实度。

难于检测天然 T7 DNA 聚合酶产生的反应产物中的聚合活性 (由于存在核酸外切酶活性)。但是, 其基因修饰的衍生物, 测序酶表明相对于未修饰的核苷酸, γ 修饰的核苷酸被有效整合, 改进了整合忠实度。天然 T7 DNA 聚合酶和测序酶的试验结果示于图 11。

因此, 对于 Taq 聚合酶或 HIV 1 逆转录酶, 由于采用了本发明的 γ 修饰的 dNTP 改进了忠实度, 可以进行单个分子 DNA 测序。但是, 不是所有的聚合酶同等地利用了本发明 γ 修饰的 dNTP, 具体的说, Klenow, 测序酶, HIV 1 逆转录酶, Taq 聚合酶整合了本发明的修饰的核苷酸, 而 *Pfu* DNA 聚合酶没有表现出整合了本发明的修饰的核苷酸。

改进的 PCR-产生 DNA 长序列

核酸合成的忠实度是采用 PCR 完成长目标分子扩增的限制因素。在引物延伸产物的合成中核苷酸的错误整合限制了可以有效扩增的目标的长度。在 Huang et al., 1992, Nucl. Acids Res. 20: 4567-4573 (引入作为参考) 中记述了 3' -末端碱基与模板错配对引物延伸的影响。错误整合核苷酸的存在可能产生成熟前中断链合成, 这样降低了未来扩增循环的模板链数量, 从而降低了长目标扩增的效率。低水平的核苷酸错误整合可能对长于 10 kb 的序列是至关重要的。图 4 的数据显示对于天然 Taq 聚合酶采用 γ 标记的 dNTP 的 DNA 合成的忠实度得到改进, 使较长 DNA 延伸成为可能, 而

不需要添加具有 3' -到 5' 外切核酸酶，或“校对”活性的聚合酶，而这些是如 Cheng et al., U. S. Pat. Nos. 5,512,462 (引入作为参考) 中长距离 PCR 方法所需的。因此本发明提供了改进的 PCR 系统，用于得到提高了延伸长度的 PCR 扩增的 DNA 产物，其包含在 PCR 反应条件下将天然 Taq 聚合酶与本发明 γ 标记的 dNTP 相接触。由于使用了本发明的 γ 修饰的 dNTP 改进了碱基整合正确性，可以得到长度延伸的 PCR 产物。

信号强度和反应动力学提供有关碱基身份的信息

在延伸的 DNA 链中每一核苷酸的信号强度用于检测，确认或支持碱基身份的数据。参照图 12，实线条对应于合成反应中包含 4 种天然核苷酸 (dATP, dCTP, dGTP, dTTP) 时产生的反应产物。虚线对应于当在反应中包含本发明的，碱基修饰的核苷酸时产生的反应产物。清楚显示，序列的连贯性和碱基修饰影响反应产物的强度和/动力学，这些鉴定方式被整合到本发明的选择碱基的软件中以在每一测序位点提供碱基身份的高置信度值。

所有引用和所列文献用于参考。虽然本发明已作出了全面和完整的描述，但应当理解在本发明权利要求范围内，本发明也可不如具体所述地实施。尽管参照其优选实施方案公开了本发明，根据所读的内容本领域技术人员可以作出适当的调整和改变，这些都包括在本发明如上所述和权利要求的范围和精神中。

序列表

(1) 信息总括:

(iii) 序列号: 23

(2) SEQ ID NO: 1 的信息:

(i) 序列特征:

- (A) 长度: 碱基
- (B) 类型: 核酸
- (C) 链型: 单链
- (D) 拓扑构型: 线性

(ii) 分子类型: 合成 DNA

(ix) 特征:

- (A) 名称/关键词: Taq Pol I For
- (B) 位置:
- (D) 其他信息: 用于扩增全长 TAQ POL I 编码序列

(xi) 序列描述: SEQ ID NO: 1:

Taq Pol I 正向

5'-gc gaattc atgaggggga tgctgccccct ctttgagccc-3'

(2) SEQ ID NO: 2 的信息:

(i) 序列特征:

- (A) 长度: 碱基
- (B) 类型: 核酸
- (C) 链型: 单链
- (D) 拓扑构型: 线性

(ii) 分子类型: 合成 DNA

(ix) 特征:

- (A) 名称/关键词: Taq Pol I rev
- (B) 位置:
- (D) 其他信息: 用于扩增全长 Taq Pol I 编码序列

(xi) 序列描述: SEQ ID NO: 2:

Taq Pol I 反向

5'-gc gaattc accatccattgg cggagcgcc cagtcctccc-3'

(2) SEQ ID NO: 3 的信息:

(i) 序列特征:

- (A) 长度: 碱基
- (B) 类型: 核酸
- (C) 链型: 单链

(D) 拓扑构型: 线性

(ii) 分子类型: 合成 DNA

(ix) 特征:

(A) 名称/关键词: Taq Pol I_A293_trunk

(B) 位置:

(D) 其他信息: 用于扩增截短的 Taq DNA Pol I 编码序列

(xi) 序列描述: SEQ ID NO: 3:

Taq Pol I_A293_trunk

5'-aatccatgggcctggaggaggc cccctggccccccgc-3'

(2) SEQ ID NO: 4 的信息:

(i) 序列特征:

(A) 长度: 碱基

(B) 类型: 核酸

(C) 链型: 单链

(D) 拓扑构型: 线性

(ii) 分子类型: 合成 DNA

(ix) 特征:

(A) 名称/关键词: Taq Pol I_AlA643Cys_fwd

(B) 位置:

(D) 其他信息: 用于引导 cys 突变到 Taq DNA Pol I 编码序列

(xi) 序列描述: SEQ ID NO: 4:

半胱氨酸取代 643 位的丙氨酸

Taq Pol I_AlA643cys_正向

5'-C CAC ACG GAG ACC tgc AGC TGG ATG TTC GGC G-3'

(2) SEQ ID NO: 5 的信息:

(i) 序列特征:

(A) 长度: 碱基

(B) 类型: 核酸

(C) 链型: 单链

(D) 拓扑构型: 线性

(ii) 分子类型: 合成 DNA

(ix) 特征:

(A) 名称/关键词: Taq Pol I_AlA643Cys_rev

(B) 位置:

(D) 其他信息: 用于引导 cys 突变到 Taq DNA Pol I 编码序列

(xi) 序列描述: SEQ ID NO: 5:

Taq Pol I_Alα643Cys_反向

5'-C GCC GAA CAT CCA CGA Gca GGT CTC CGT GTG G-3'

(2) SEQ ID NO: 6 的信息:

(i) 序列特征:

- (A) 长度: 碱基
- (B) 类型: 核酸
- (C) 链型: 单链
- (D) 拓扑构型: 线性

(ii) 分子类型: 合成 DNA

(ix) 特征:

- (A) 名称/关键词: Taq Pol I_Phe647Cys_fwd
- (B) 位置:
- (D) 其他信息: 用于引导 cys 突变到 Taq DNA pol I 编码序列

(xi) 序列描述: SEQ ID NO: 6:

647 位的苯丙氨酸取代半胱氨酸

Taq Pol I_Phe647Cys_正向

5'-CC GCC AGC TGG ATG TgC GGC GTC CCC CGG GAG GCC-3'

(2) SEQ ID NO: 7 的信息:

(i) 序列特征:

- (A) 长度: 1119 碱基
- (B) 类型: 核酸
- (C) 链型: 单链
- (D) 拓扑构型: 线性

(ii) 分子类型: 合成 DNA

(ix) 特征:

- (A) 名称/关键词: Taq Pol I_Phe647Cys_rev
- (B) 位置:
- (D) 其他信息: 用于引导 cys 突变到 Taq DNA pol I 编码序列

(xi) 序列描述: SEQ ID NO: 7:

647 位的苯丙氨酸取代半胱氨酸

Taq Pol I_Phe643Cys_反向

5'-GGC CTC CCG GGG GAC GCC GcA CAT CCA CGT GGC GG-3'

(2) SEQ ID NO: 8 的信息:

(i) 序列特征:

- (A) 长度: 碱基
- (B) 类型: 核酸
- (C) 链型: 单链
- (D) 拓扑构型: 线性
- (ii) 分子类型: 合成 DNA
- (ix) 特征:
 - (A) 名称/关键词: Taq Pol I_Val649Cys_fwd
 - (B) 位置:
 - (D) 其他信息: 用于引导 cys 突变到 Taq DNA pol I 编码序列
- (xi) 序列描述: SEQ ID NO: 8:

649 位的缬 氨酸取代半胱氨酸

Taq Pol I_Val649Cys_正向

5'-GCC AGC TGG ATG TTC GGC tgc CCC CGG GAG GCC GTG G-3'

(2) SEQ ID NO: 9 的信息:

- (i) 序列特征:
 - (A) 长度: 碱基
 - (B) 类型: 核酸
 - (C) 链型: 单链
 - (D) 拓扑构型: 线性
- (ii) 分子类型: 合成 DNA
- (ix) 特征:
 - (A) 名称/关键词: Taq Pol I_Val649Cys_rev
 - (B) 位置:
 - (D) 其他信息: 用于引导 cys 突变到 Taq DNA pol I 编码序列
- (xi) 序列描述: SEQ ID NO: 9:

Taq Pol I_Val649Cys_反向

5'-C CAC GGC CTC CCG GGG Gca GCC GAA CAT CCA GCT GGC-3'

(2) SEQ ID NO: 10 的信息:

- (i) 序列特征:
 - (A) 长度: 碱基
 - (B) 类型: 核酸
 - (C) 链型: 单链
 - (D) 拓扑构型: 线性
- (ii) 分子类型: 合成 DNA
- (ix) 特征:
 - (A) 名称/关键词: Taq Pol I_Glu652Cys_fwd

(B) 位置:

(D) 其他信息: 用于引导 cys 突变到 Taq DNA pol I 编码序列

(xi) 序列描述: SEQ ID NO: 10:

半胱氨酸取代 652 位的谷氨酸

Taq Pol I_Glu652Cys_正向

5'-GGC GTC CCC CGG tgc GCC GTG GAC CCC CTG ATG CGC-3'

(2) SEQ ID NO: 11 的信息:

(i) 序列特征:

- (A) 长度: 碱基
- (B) 类型: 核酸
- (C) 链型: 单链
- (D) 拓扑构型: 线性

(ii) 分子类型: 合成 DNA

(ix) 特征:

- (A) 名称/关键词: Taq Pol I_Glu652Cys_rev
- (B) 位置:

(D) 其他信息: 用于引导 cys 突变到 Taq DNA pol I 编码序列

(xi) 序列描述: SEQ ID NO: 11:

Taq Pol I_Glu652Cys_反向

5'-GCG CAT CAG GGG GTC CAC GGC gca CCG GGG GAC GCC-3'

(2) SEQ ID NO: 12 的信息:

(i) 序列特征:

- (A) 长度: 碱基
- (B) 类型: 核酸
- (C) 链型: 单链
- (D) 拓扑构型: 线性

(ii) 分子类型: 合成 DNA

(ix) 特征:

- (A) 名称/关键词: Taq Pol I_Alpha653Cys_fwd
- (B) 位置:

(D) 其他信息: 用于引导 cys 突变到 Taq DNA pol I 编码序列

(xi) 序列描述: SEQ ID NO: 12:

半胱氨酸取代 653 位的丙氨酸

Taq Pol I_Alpha653Cys_正向

5'-GGC GTC CCC CGG GAG tgt GTG GAC CCC CTG ATG CGC-3'

(2) SEQ ID NO: 13 的信息:

(i) 序列特征:

- (A) 长度: 碱基
- (B) 类型: 核酸
- (C) 链型: 单链
- (D) 拓扑构型: 线性

(ii) 分子类型: 合成 DNA

(ix) 特征:

- (A) 名称/关键词: Taq Pol I_Alα653Cys_rev
- (B) 位置:
- (D) 其他信息: 用于引导 cys 突变到 Taq DNA pol I 编码序列

(xi) 序列描述: SEQ ID NO: 13:

Taq Pol I_Alα653Cys_反向

5'-GCG CAT CAG GGG GTC CAC Gca CTC CCG GGG GAC GCC-3'

(2) SEQ ID NO: 14 的信息:

(i) 序列特征:

- (A) 长度: 碱基
- (B) 类型: 核酸
- (C) 链型: 单链
- (D) 拓扑构型: 线性

(ii) 分子类型: 合成 DNA

(ix) 特征:

- (A) 名称/关键词: Taq Pol I_Val654Cys_fwd
- (B) 位置:
- (D) 其他信息: 用于引导 cys 突变到 Taq DNA pol I 编码序列

(xi) 序列描述: SEQ ID NO: 14:

半胱氨酸取代 654 位的缬氨酸

Taq Pol I_Val654Cys_正向

5'-GTC CCC CGG GAG GCC tgt GAC CCC CTG ATG CGC-3'

(2) SEQ ID NO: 15 的信息:

(i) 序列特征:

- (A) 长度: 碱基
- (B) 类型: 核酸
- (C) 链型: 单链
- (D) 拓扑构型: 线性

(ii) 分子类型: 合成 DNA

(ix) 特征:

(A) 名称/关键词: Taq Pol I_Val654Cys_rev

(B) 位置:

(D) 其他信息: 用于引导 cys 突变到 Taq DNA pol I 编码序列

(xi) 序列描述: SEQ ID NO: 15:

Taq Pol I_Val654Cys_反向

5'-GCG CAT CAG GGG GTC aca GGC CTC CCG GGG GAC-3'

(2) SEQ ID NO: 16 的信息:

(i) 序列特征:

(A) 长度: 碱基

(B) 类型: 核酸

(C) 链型: 单链

(D) 拓扑构型: 线性

(ii) 分子类型: 合成 DNA

(ix) 特征:

(A) 名称/关键词: Taq Pol I_D655C_fwd

(B) 位置:

(D) 其他信息: 用于引导 cys 突变到 Taq DNA pol I 编码序列

(xi) 序列描述: SEQ ID NO: 16:

半胱氨酸取代 655 位的天冬氨酸

Taq Pol I_D655C_正向

5'-CCC CGG GAG GCC GTG tgc CCC CTG ATG CGC CGG-3'

(2) SEQ ID NO: 17 的信息:

(i) 序列特征:

(A) 长度: 碱基

(B) 类型: 核酸

(C) 链型: 单链

(D) 拓扑构型: 线性

(ii) 分子类型: 合成 DNA

(ix) 特征:

(A) 名称/关键词: Taq Pol I_D655C_rev

(B) 位置:

(D) 其他信息: 用于引导 cys 突变到 Taq DNA pol I 编码序列

(xi) 序列描述: SEQ ID NO: 17:

Taq Pol I_D655C_反向

5'-CCG GCG CAT CAG GGG Gca CAC GGC CTC CCG GGG-3'

(2) SEQ ID NO: 18 的信息:

(i) 序列特征:

- (A) 长度: 碱基
- (B) 类型: 核酸
- (C) 链型: 单链
- (D) 拓扑构型: 线性

(ii) 分子类型: 合成 DNA

(ix) 特征:

- (A) 名称/关键词: Taq Pol I_Pro656Cys_fwd
- (B) 位置:
- (D) 其他信息: 用于引导 cys 突变到 Taq DNA pol I 编码序列

(xi) 序列描述: SEQ ID NO: 18:

半胱氨酸取代 656 位的脯氨酸

Taq Pol I_Pro656Cys_正向

5'-CGG GAG GCC GTG GAC tgc CTG ATG CGC CGG GCG-3'

(2) SEQ ID NO: 19 的信息:

(i) 序列特征:

- (A) 长度: 碱基
- (B) 类型: 核酸
- (C) 链型: 单链
- (D) 拓扑构型: 线性

(ii) 分子类型: 合成 DNA

(ix) 特征:

- (A) 名称/关键词: Taq Pol I_Pro656Cys_rev
- (B) 位置:
- (D) 其他信息: 用于引导 cys 突变到 Taq DNA pol I 编码序列

(xi) 序列描述: SEQ ID NO: 19:

Taq Pol I_Pro656Cys_反向

5'-CGC CCG GCG CAT CAG Gca GTC CAC GGC CTC CCG-3'

(2) SEQ ID NO: 20 的信息:

(i) 序列特征:

- (A) 长度: 碱基
- (B) 类型: 核酸

- (C) 链型: 单链
 (D) 拓扑构型: 线性
 (ii) 分子类型: 合成 DNA
 (ix) 特征:
 (A) 名称/关键词: Taq Pol I_Leu657Cys_fwd
 (B) 位置:
 (D) 其他信息: 用于引导 cys 突变到 Taq DNA pol I 编码序列
 (xi) 序列描述: SEQ ID NO: 20:

半胱氨酸取代 657 位的亮氨酸

Taq Pol I_Leu657Cys_正向

5'-GCC GTG GAC CCC tgc ATG CGC CGG GCG GCC-3'

(2) SEQ ID NO: 21 的信息:

- (i) 序列特征:
 (A) 长度: 碱基
 (B) 类型: 核酸
 (C) 链型: 单链
 (D) 拓扑构型: 线性
 (ii) 分子类型: 合成 DNA
 (ix) 特征:
 (A) 名称/关键词: Taq Pol I_Leu657Cys_rev
 (B) 位置:
 (D) 其他信息: 用于引导 cys 突变到 Taq DNA pol I 编码序列
 (xi) 序列描述: SEQ ID NO: 21:

Taq Pol I_Leu657Cys_反向

5'-GGC CGC CCG GCG CAT gca GGG GTC CAC GGC-3'

(2) SEQ ID NO: 22 的信息:

- (i) 序列特征:
 (A) 长度: 碱基
 (B) 类型: 核酸
 (C) 链型: 单链
 (D) 拓扑构型: 线性
 (ii) 分子类型: 合成 DNA
 (ix) 特征:
 (A) 名称/关键词: Taq Pol I_Met658Cys_fwd
 (B) 位置:
 (D) 其他信息: 用于引导 cys 突变到 Taq DNA pol I 编码序列

(xi) 序列描述: SEQ ID NO: 22:

半胱氨酸取代 658 位的甲硫氨酸

Taq Pol I_Met658Cys_正向

5'-GCC GTG GAC CCC CTG tgt CGC CGG GCG GCC-3'

(2) SEQ ID NO: 23 的信息:

(i) 序列特征:

- (A) 长度: 碱基
- (B) 类型: 核酸
- (C) 链型: 单链
- (D) 拓扑构型: 线性

(ii) 分子类型: 合成 DNA

(ix) 特征:

- (A) 名称/关键词: Taq Pol I_Met658Cys_rev
- (B) 位置:
- (D) 其他信息: 用于引导 cys 突变到 Taq DNA pol I 编码序列

(xi) 序列描述: SEQ ID NO: 23:

Taq Pol I_Met658Cys_反向

5'-GGC CGC CCG GCG aca CAG GGG GTC CAC GGC-3'

(2) SEQ ID NO: 24 的信息:

(i) 序列特征:

- (A) 长度: 碱基
- (B) 类型: 核酸
- (C) 链型: 单链
- (D) 拓扑构型: 线性

(ii) 分子类型: 合成 DNA

(ix) 特征:

- (A) 名称/关键词: Taq Pol I_Arg659Cys_fwd
- (B) 位置:
- (D) 其他信息: 用于引导 cys 突变到 Taq DNA pol I 编码序列

(xi) 序列描述: SEQ ID NO: 24:

半胱氨酸取代 659 位的精氨酸

Taq Pol I_Arg659Cys_正向

5'-GCC GTG GAC CCC CTG ATG tGC CGG GCG GCC AAG ACC-3'

(2) SEQ ID NO: 25 的信息:

(i) 序列特征:

- (A) 长度: 碱基
- (B) 类型: 核酸
- (C) 链型: 单链
- (D) 拓扑构型: 线性

(ii) 分子类型: 合成 DNA

(ix) 特征:

- (A) 名称/关键词: Taq Pol I_Arg659Cys_rev
 - (B) 位置:
 - (D) 其他信息: 用于引导 cys 突变到 Taq DNA pol I 编码序列
- (xi) 序列描述: SEQ ID NO: 25:

Taq Pol I_Arg659Cys_反向

5' -GGT CTT GGC CGC CCG GCa CAT CAG GGG GTC CAC GGC-3'

(2) SEQ ID NO: 26 的信息:

(i) 序列特征:

- (A) 长度: 碱基
- (B) 类型: 核酸
- (C) 链型: 单链
- (D) 拓扑构型: 线性

(ii) 分子类型: 合成 DNA

(ix) 特征:

- (A) 名称/关键词: Taq Pol I_Arg660Cys_fwd
 - (B) 位置:
 - (D) 其他信息: 用于引导 cys 突变到 Taq DNA pol I 编码序列
- (xi) 序列描述: SEQ ID NO: 26:

半胱氨酸取代 660 位的精氨酸

Taq Pol I_Arg660Cys_正向

5' -GAC CCC CTG ATG CGC tGc GCG GCC AAG ACC ATC-3'

(2) SEQ ID NO: 27 的信息:

(i) 序列特征:

- (A) 长度: 碱基
- (B) 类型: 核酸
- (C) 链型: 单链
- (D) 拓扑构型: 线性

(ii) 分子类型: 合成 DNA

(ix) 特征:

(A) 名称/关键词: Taq Pol I_Arg660Cys_rev

(B) 位置:

(D) 其他信息: 用于引导 cys 突变到 Taq DNA pol I 编码序列

(xi) 序列描述: SEQ ID NO: 27:

Taq Pol I_Arg660Cys_反向

5'-GAT GGT CTT GGC CGC gCa GCG CAT CAG GGG GTC-3'

(2) SEQ ID NO: 28 的信息:

(i) 序列特征:

(A) 长度: 碱基

(B) 类型: 核酸

(C) 链型: 单链

(D) 拓扑构型: 线性

(ii) 分子类型: 合成 DNA

(ix) 特征:

(A) 名称/关键词: Taq Pol I_Arg661Cys_fwd

(B) 位置:

(D) 其他信息: 用于引导 cys 突变到 Taq DNA pol I 编码序列

(xi) 序列描述: SEQ ID NO: 28:

半胱氨酸取代 661 位的丙氨酸

Taq Pol I_Arg660Cys_正向

5'-CCC CTG ATG CGC CGG tgc GCC AAG ACC ATC AAC-3'

(2) SEQ ID NO: 29 的信息:

(i) 序列特征:

(A) 长度: 碱基

(B) 类型: 核酸

(C) 链型: 单链

(D) 拓扑构型: 线性

(ii) 分子类型: 合成 DNA

(ix) 特征:

(A) 名称/关键词: Taq Pol I_Al661Cys_rev

(B) 位置:

(D) 其他信息: 用于引导 cys 突变到 Taq DNA pol I 编码序列

(xi) 序列描述: SEQ ID NO: 29:

Taq Pol I_Arg661Cys_反向

5'-GTT GAT GGT CTT GGC gca CCG GCG CAT CAG GGG-3'

(2) SEQ ID NO: 30 的信息:

(i) 序列特征:

- (A) 长度:
- (B) 类型: 核酸
- (C) 链型: 单链
- (D) 拓扑构型: 线性

(ii) 分子类型: Tag DNA 聚合酶多肽区

(ix) 特征:

- (A) 名称/关键词:
- (B) 位置:
- (D) 其他信息: 包含用于标记连接的突变 Taq DNA 聚合酶的多肽区

(xi) 序列描述: SEQ ID NO: 30:

```
Cys Ser Trp Met Phe Gly Val Pro Arg Glu Ala Val Asp Pro Leu
Met
643 644 645 646 647 648 649 650 651 652 653 654 655 656 657
658
Arg Arg Ala
659 660 661
```

(2) SEQ ID NO: 31 的信息:

(i) 序列特征:

- (A) 长度:
- (B) 类型: 核酸
- (C) 链型: 单链
- (D) 拓扑构型: 线性

(ii) 分子类型: Tag DNA 聚合酶多肽区

(ix) 特征:

- (A) 名称/关键词:
- (B) 位置:
- (D) 其他信息: 包含用于 Tag 附着的突变的 Tag DNA 聚合酶多肽区

(xi) 序列描述: SEQ ID NO: 31:

```
Ala Ser Trp Met Cys Gly Val Pro Arg Glu Ala Val Asp Pro Leu
Met
643 644 645 646 647 648 649 650 651 652 653 654 655 656 657
658
Arg Arg Ala
659 660 661
```

(2) SEQ ID NO: 32 的信息:

(i) 序列特征:

- (A) 长度:
- (B) 类型: 氨基酸
- (C) 链型: 单链

(D) 拓扑构型:

(ii) 分子类型: Tag DNA 聚合酶多肽区

(ix) 特征:

(A) 名称/关键词:

(B) 位置:

(D) 其他信息: 包含用于 Tag 附着的突变的 Tag DNA 聚合酶多肽区

(xi) 序列描述: SEQ ID NO: 32:

```

Ala Ser Trp Met Phe Gly Cys Pro Arg Glu Ala Val Asp Pro Leu
Met
643 644 645 646 647 648 649 650 651 652 653 654 655 656 657
658
Arg Arg Ala
659 660 661

```

(2) SEQ ID NO: 33 的信息:

(i) 序列特征:

(A) 长度:

(B) 类型: 氨基酸

(C) 链型: 单链

(D) 拓扑构型:

(ii) 分子类型: Tag DNA 聚合酶多肽区

(iii) 特征:

(A) 名称/关键词:

(B) 位置:

(D) 其他信息: 包含用于 Tag 附着的突变的 Tag DNA 聚合酶多肽区

(ix) 序列描述: SEQ ID NO: 33:

```

Ala Ser Trp Met Phe Val Pro Arg Cys Ala Val Asp Pro Leu
Met
643 644 645 646 647 648 649 650 651 652 653 654 655 656 657
658
Arg Arg Ala
659 660 661

```

(2) SEQ ID NO: 34 的信息:

(i) 序列特征:

(A) 长度:

(B) 类型: 氨基酸

(C) 链型: 单链

(D) 拓扑构型:

(ii) 分子类型: Tag DNA 聚合酶多肽区

(iii) 特征:

(A) 名称/关键词:

(B) 位置:

(D) 其他信息: 包含用于 Tag 附着的突变的 Tag DNA 聚合酶多肽区

(ix) 序列描述: SEQ ID NO: 34:

Ala Ser Trp Met Phe Gly Val Pro Arg Glu Cys Val Asp Pro Leu
 Met
 643 644 645 646 647 648 649 650 651 652 653 654 655 656 657
 658
 Arg Arg Ala
 659 660 661

(2) SEQ ID NO: 35 的信息:

(i) 序列特征:

- (A) 长度:
- (B) 类型: 氨基酸
- (C) 链型: 单链
- (D) 拓扑构型:

(ii) 分子类型: Tag DNA 聚合酶多肽区

(iii) 特征:

- (A) 名称/关键词:
- (B) 位置:
- (D) 其他信息: 包含用于 Tag 附着的突变的 Tag DNA 聚合酶多肽区

(ix) 序列描述: SEQ ID NO: 35:

Ala Ser Trp Met Phe Gly Val Pro Arg Glu Ala Cys Asp Pro Leu
 Met
 643 644 645 646 647 648 649 650 651 652 653 654 655 656 657
 658
 Arg Arg Ala
 659 660 661

(2) SEQ ID NO: 36 的信息:

(i) 序列特征:

- (A) 长度:
- (B) 类型: 氨基酸
- (C) 链型: 单链
- (D) 拓扑构型:

(ii) 分子类型: Tag DNA 聚合酶多肽区

(iii) 特征:

- (A) 名称/关键词:
- (B) 位置:
- (D) 其他信息: 包含用于 Tag 附着的突变的 Tag DNA 聚合酶多肽区

(ix) 序列描述: SEQ ID NO: 36:

Ala Ser Trp Met Phe Gly Val Pro Arg Glu Ala Val Cys Pro Leu
 Met
 643 644 645 646 647 648 649 650 651 652 653 654 655 656 657
 658
 Arg Arg Ala
 659 660 661

(2) SEQ ID NO: 37 的信息:

(i) 序列特征:

- (A) 长度:
- (B) 类型: 氨基酸
- (C) 链型: 单链
- (D) 拓扑构型:

(ii) 分子类型: Tag DNA 聚合酶多肽区

(iii) 特征:

- (A) 名称/关键词:
- (B) 位置:
- (D) 其他信息: 包含用于 Tag 附着的突变的 Tag DNA 聚合酶多肽区

(ix) 序列描述: SEQ ID NO: 37:

```

Ala Ser Trp Met Phe Gly Val Pro Arg Glu Ala Val Asp Cys Leu
Met
643 644 645 646 647 648 649 650 651 652 653 654 655 656 657
658
Arg Arg Ala
659 660 661

```

(2) SEQ ID NO: 38 的信息:

(i) 序列特征:

- (A) 长度:
- (B) 类型: 氨基酸
- (C) 链型: 单链
- (D) 拓扑构型:

(ii) 分子类型: Tag DNA 聚合酶多肽区

(iii) 特征:

- (A) 名称/关键词:
- (B) 位置:
- (D) 其他信息: 包含用于 Tag 附着的突变的 Tag DNA 聚合酶多肽区

(ix) 序列描述: SEQ ID NO: 38:

```

Ala Ser Trp Met Phe Gly Val Pro Arg Glu Ala Val Asp Pro Cys
Met
643 644 645 646 647 648 649 650 651 652 653 654 655 656 657
658
Arg Arg Ala
659 660 661

```

(2) SEQ ID NO: 39 的信息:

(i) 序列特征:

- (A) 长度:
- (B) 类型: 氨基酸
- (C) 链型: 单链
- (D) 拓扑构型:

(ii) 分子类型: Tag DNA 聚合酶多肽区

(iii) 特征:

(A) 名称/关键词:

(B) 位置:

(D) 其他信息: 包含用于 Tag 附着的突变的 Tag DNA 聚合酶多肽区

(ix) 序列描述: SEQ ID NO: 39:

```

Ala Ser Trp Met Phe Gly Val Pro Arg Glu Ala Val Asp Pro Leu
Cys
643 644 645 646 647 648 649 650 651 652 653 654 655 656 657
658
Arg Arg Ala
659 660 661

```

(2) SEQ ID NO: 40 的信息:

(i) 序列特征:

(A) 长度:

(B) 类型: 氨基酸

(C) 链型: 单链

(D) 拓扑构型:

(ii) 分子类型: Tag DNA 聚合酶多肽区

(iii) 特征:

(A) 名称/关键词:

(B) 位置:

(D) 其他信息: 包含用于 Tag 附着的突变的 Tag DNA 聚合酶多肽区

(ix) 序列描述: SEQ ID NO: 40:

```

Ala Ser Trp Met Phe Gly Val Pro Arg Glu Ala Val Asp Pro Leu
Met
643 644 645 646 647 648 649 650 651 652 653 654 655 656 657
658
Cys Arg Ala
659 660 661

```

(2) SEQ ID NO: 41 的信息:

(i) 序列特征:

(A) 长度:

(B) 类型: 氨基酸

(C) 链型: 单链

(D) 拓扑构型:

(ii) 分子类型: Tag DNA 聚合酶多肽区

(iii) 特征:

(A) 名称/关键词:

(B) 位置:

(D) 其他信息: 包含用于 Tag 附着的突变的 Tag DNA 聚合酶多肽区

(ix) 序列描述: SEQ ID NO: 41:

Ala Ser Trp Met Phe Gly Val Pro Arg Glu Ala Val Asp Pro Leu
 Met
 643 644 645 646 647 648 649 650 651 652 653 654 655 656 657
 658

Arg Cys Ala
 659 660 661

(2) SEQ ID NO: 42 的信息:

(i) 序列特征:

- (A) 长度:
- (B) 类型: 氨基酸
- (C) 链型: 单链
- (D) 拓扑构型:

(ii) 分子类型: Tag DNA 聚合酶多肽区

(iii) 特征:

- (A) 名称/关键词:
- (B) 位置:
- (D) 其他信息: 包含用于 Tag 附着的突变的 Tag DNA 聚合酶多肽区

(ix) 序列描述: SEQ ID NO: 42:

Ala Ser Trp Met Phe Gly Val Pro Arg Glu Ala Val Asp Pro Leu
 Met
 643 644 645 646 647 648 649 650 651 652 653 654 655 656 657
 658
 Arg Arg Cys
 659 660 661

(2) SEQ ID NO: 43 的信息:

(i) 序列特征:

- (A) 长度:
- (B) 类型: 氨基酸
- (C) 链型: 单链
- (D) 拓扑构型:

(ii) 分子类型: Tag DNA 聚合酶多肽区

(iii) 特征:

- (A) 名称/关键词:
- (B) 位置:
- (D) 其他信息: 包含用于 Tag 附着的突变的 Tag DNA 聚合酶多肽区

(ix) 序列描述: SEQ ID NO: 43:

Cys Thr Ser Ala Ala Val
 513 514 515 516 517 518

(2) SEQ ID NO: 44 的信息:

(i) 序列特征:

- (A) 长度:
- (B) 类型: 氨基酸

(C) 链型: 单链

(D) 拓扑构型:

(ii) 分子类型: Tag DNA 聚合酶多肽区

(iii) 特征:

(A) 名称/关键词:

(B) 位置:

(D) 其他信息: 包含用于 Tag 附着的突变的 Tag DNA 聚合酶多肽区

(ix) 序列描述: SEQ ID NO: 44:

Ser Cys Ser Ala Ala Val
513 514 515 516 517 518

(2) SEQ ID NO: 45 的信息:

(i) 序列特征:

(A) 长度:

(B) 类型: 氨基酸

(C) 链型: 单链

(D) 拓扑构型:

(ii) 分子类型: Tag DNA 聚合酶多肽区

(iii) 特征:

(A) 名称/关键词:

(B) 位置:

(D) 其他信息: 包含用于 Tag 附着的突变的 Tag DNA 聚合酶多肽区

(ix) 序列描述: SEQ ID NO: 45:

Ser Thr Cys Ala Ala Val
513 514 515 516 517 518

(2) SEQ ID NO: 46 的信息:

(i) 序列特征:

(A) 长度:

(B) 类型: 氨基酸

(C) 链型: 单链

(D) 拓扑构型:

(ii) 分子类型: Tag DNA 聚合酶多肽区

(iii) 特征:

(A) 名称/关键词:

(B) 位置:

(D) 其他信息: 包含用于 Tag 附着的突变的 Tag DNA 聚合酶多肽区

(ix) 序列描述: SEQ ID NO: 46:

Ser Thr Ser Cys Ala Val
513 514 515 516 517 518

(2) SEQ ID NO: 47 的信息:

(i) 序列特征:

(A) 长度:

(B) 类型: 氨基酸

(C) 链型: 单链

(D) 拓扑构型:

(ii) 分子类型: Tag DNA 聚合酶多肽区

(iii) 特征:

(A) 名称/关键词:

(B) 位置:

(D) 其他信息: 包含用于 Tag 附着的突变的 Tag DNA 聚合酶多肽区

(ix) 序列描述: SEQ ID NO: 47:

| | | | | | |
|-----|-----|-----|-----|-----|-----|
| Ser | Thr | Ser | Ala | Cys | Val |
| 513 | 514 | 515 | 516 | 517 | 518 |

(2) SEQ ID NO: 48 的信息:

(i) 序列特征:

(A) 长度:

(B) 类型: 氨基酸

(C) 链型: 单链

(D) 拓扑构型:

(ii) 分子类型: Tag DNA 聚合酶多肽区

(iii) 特征:

(A) 名称/关键词:

(B) 位置:

(D) 其他信息: 包含用于 Tag 附着的突变的 Tag DNA 聚合酶多肽区

(ix) 序列描述: SEQ ID NO: 48:

| | | | | | |
|-----|-----|-----|-----|-----|-----|
| Ser | Thr | Ser | Ala | Ala | Cys |
| 513 | 514 | 515 | 516 | 517 | 518 |

$$\Xi = R^{-6}/(R^{-6} + R_o^{-6})$$

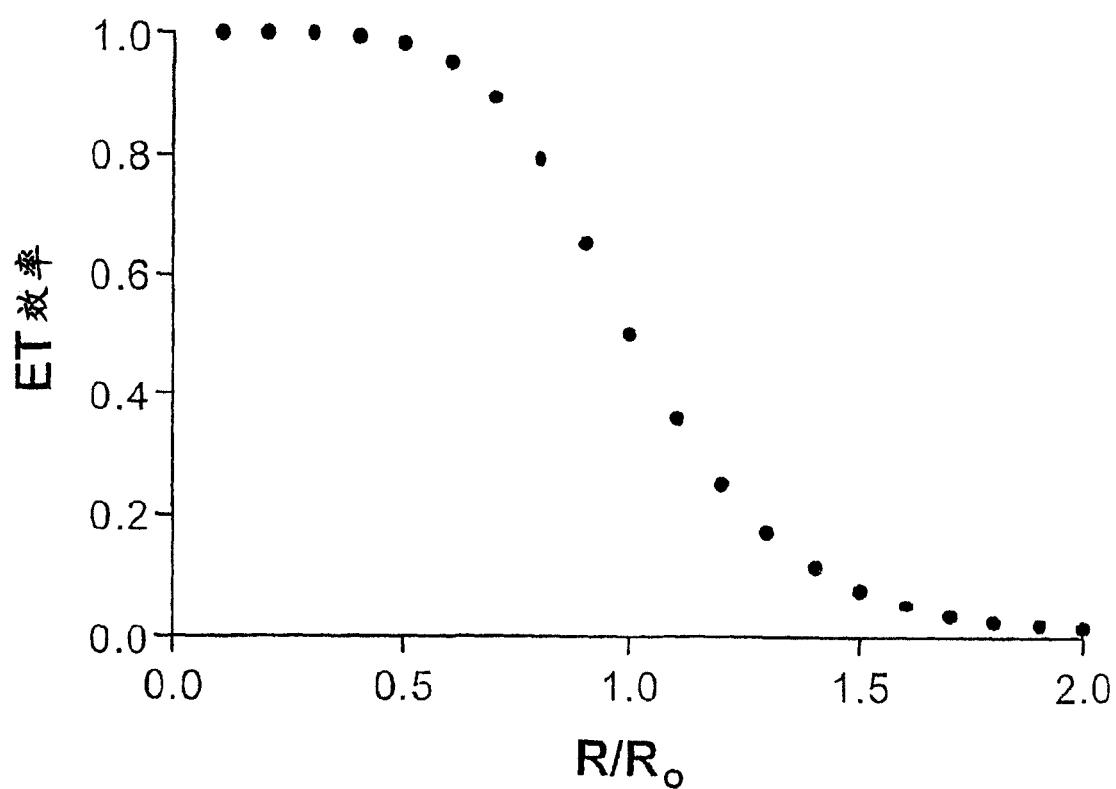


图 1



图 2



图 3A

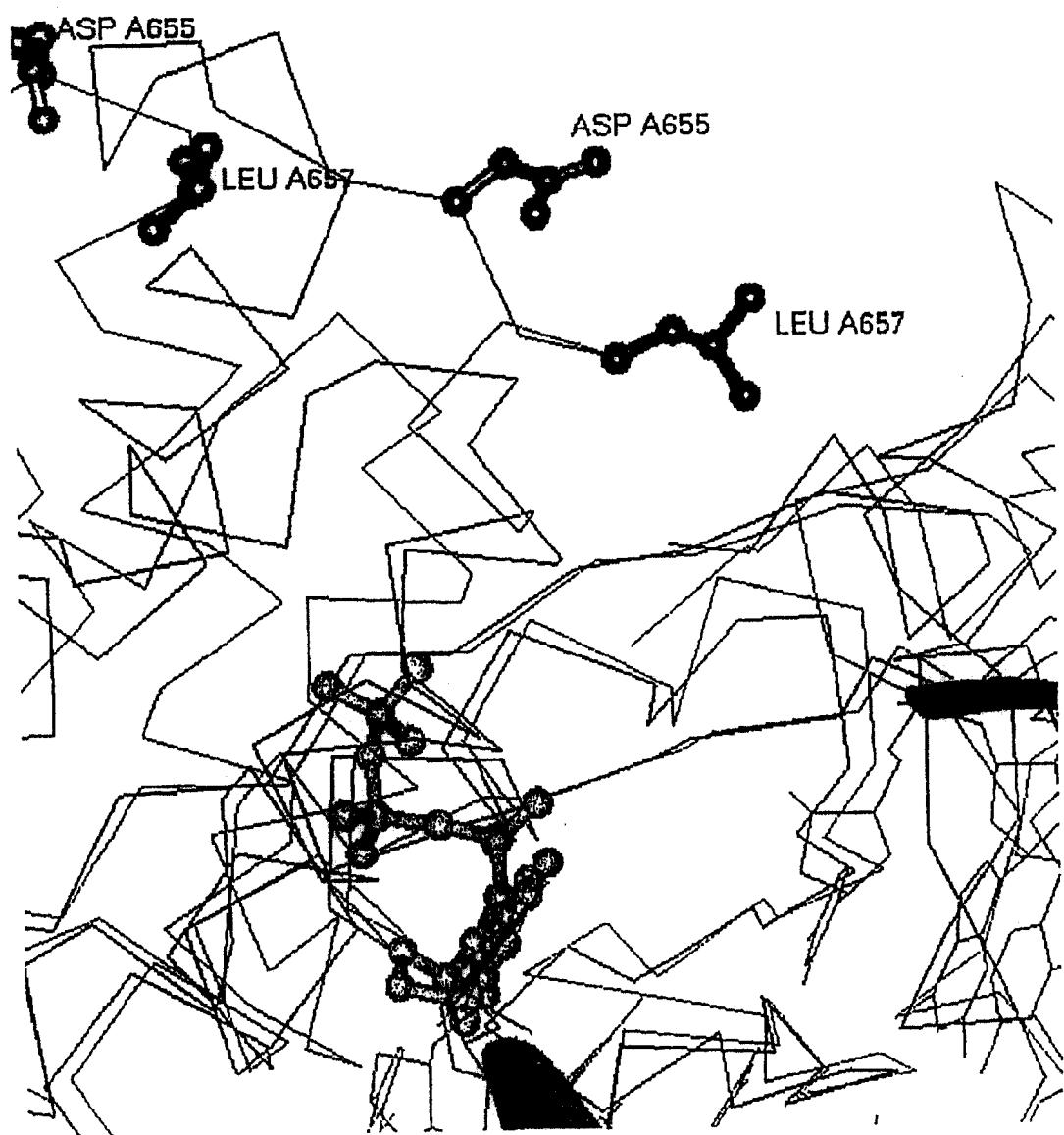


图 3B



图 3C

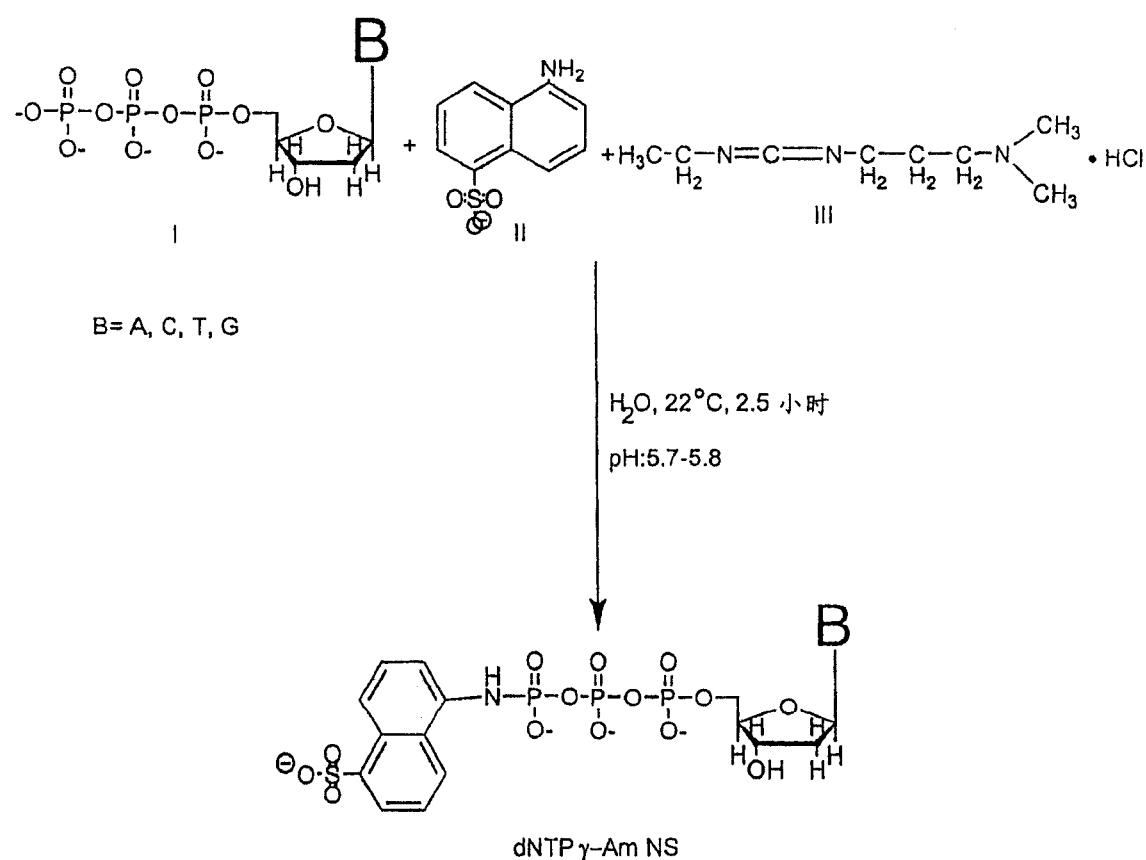


图 4

引物链:

Top 5' GGT ACT AAG CGG CCG CAT G 3'

模板链:

BOT-T 3' CCA TGA TTC GCC GGC GTA CTC 5'

BOT-C 3' CCA TGA TTC GCC GGC GTA CCC 5'

BOT-G 3' CCA TGA TTC GCC GGC GTA CGC 5'

BOT-A 3' CCA TGA TTC GCC GGC GTA CAC 5'

BOT-3T 3' CCA TGA TTC GCC GGC GTA CTT TC 5'

BOT-Sau 3' CCA TGA TTC GCC GGC GTA CCT AG 5'

整合:
(5' to 3')GATC AG AAAG

图 5

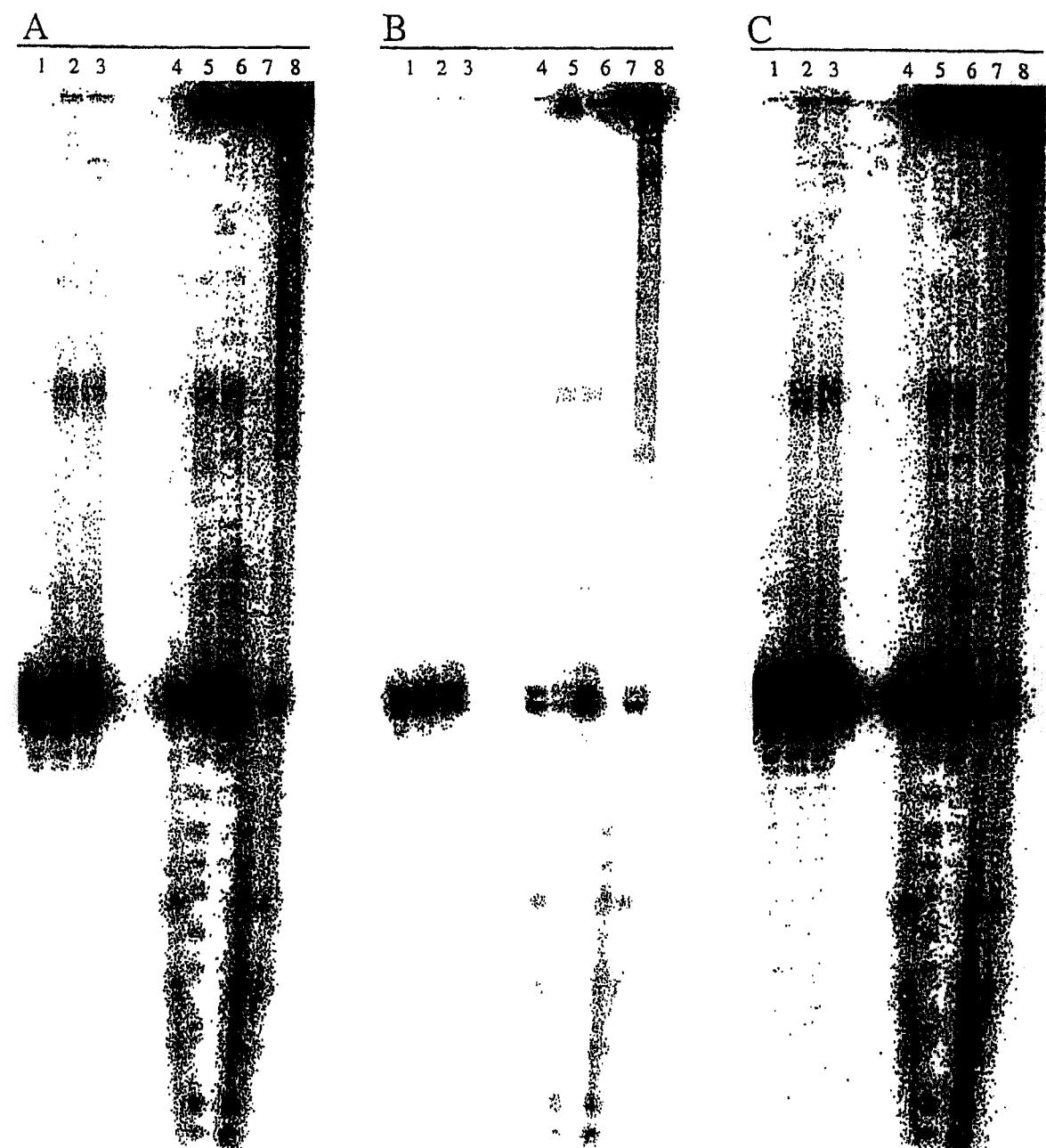


图 6

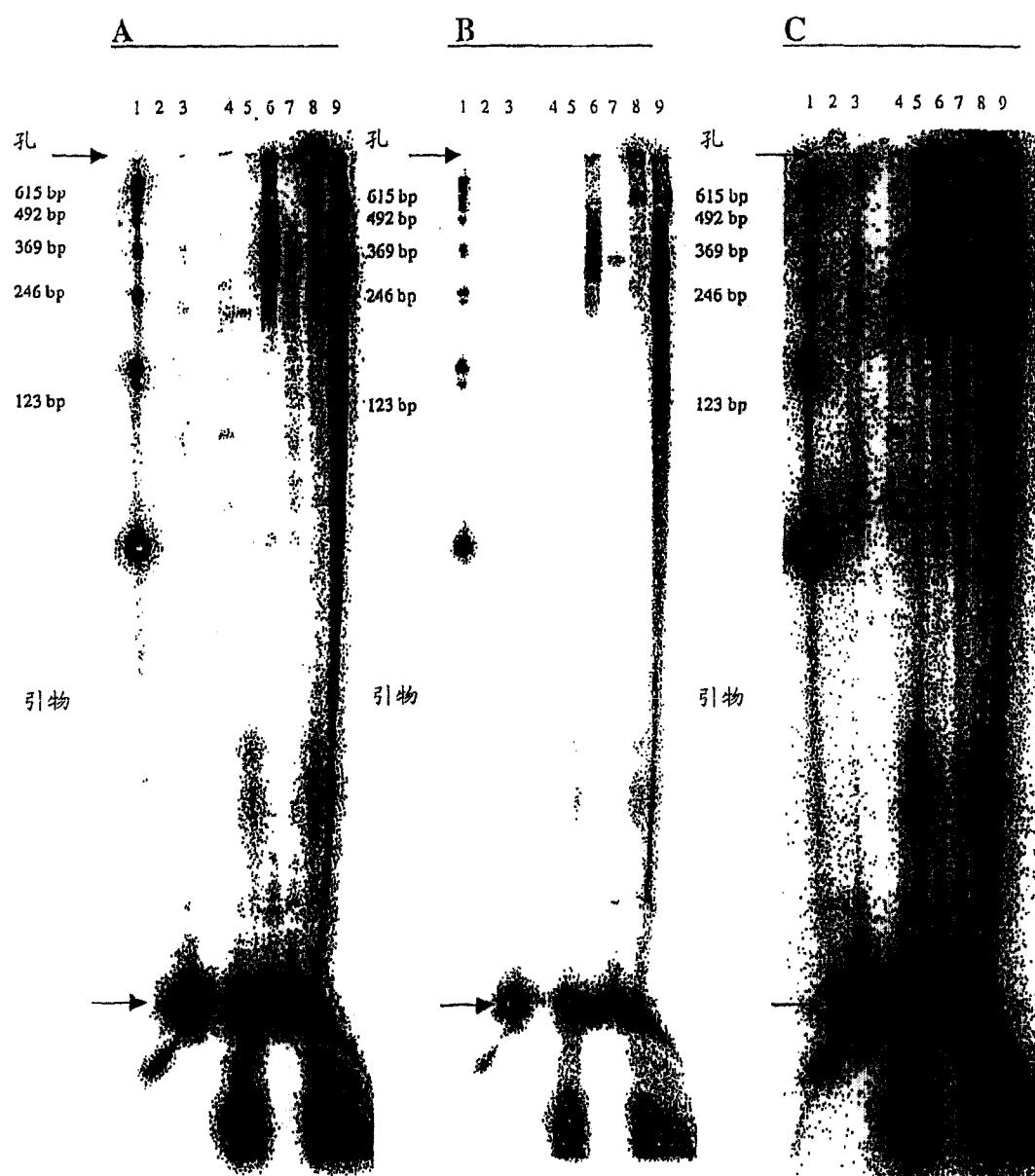


图 7

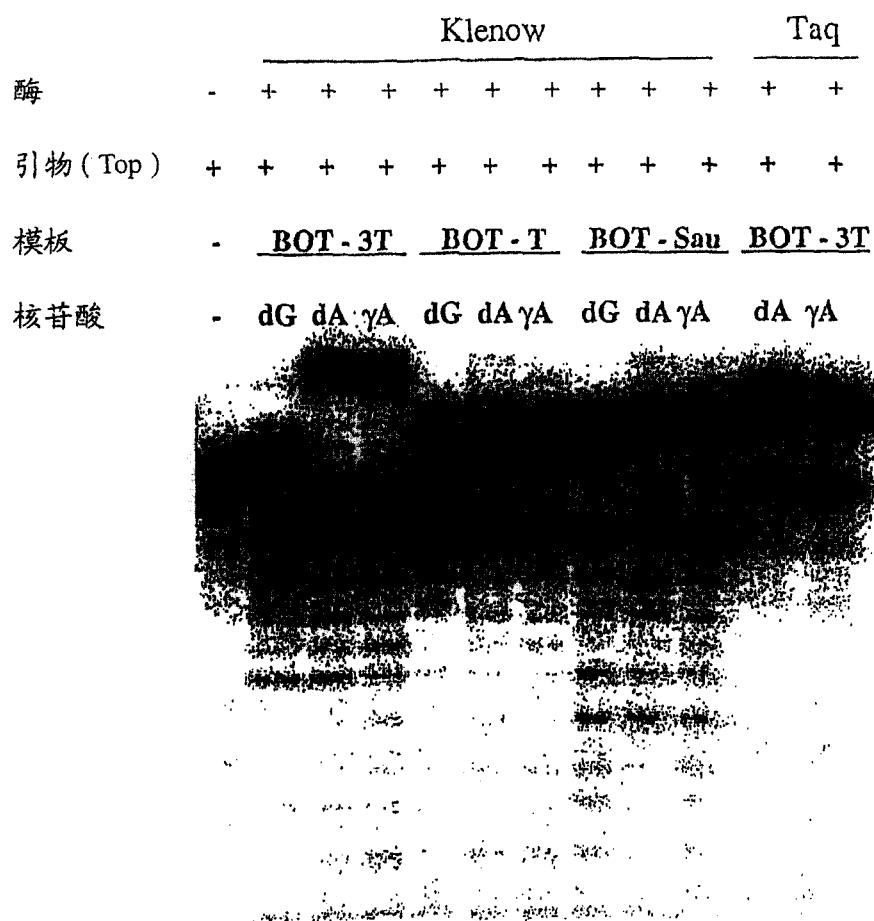


图 8

Pfu 引物延伸检测

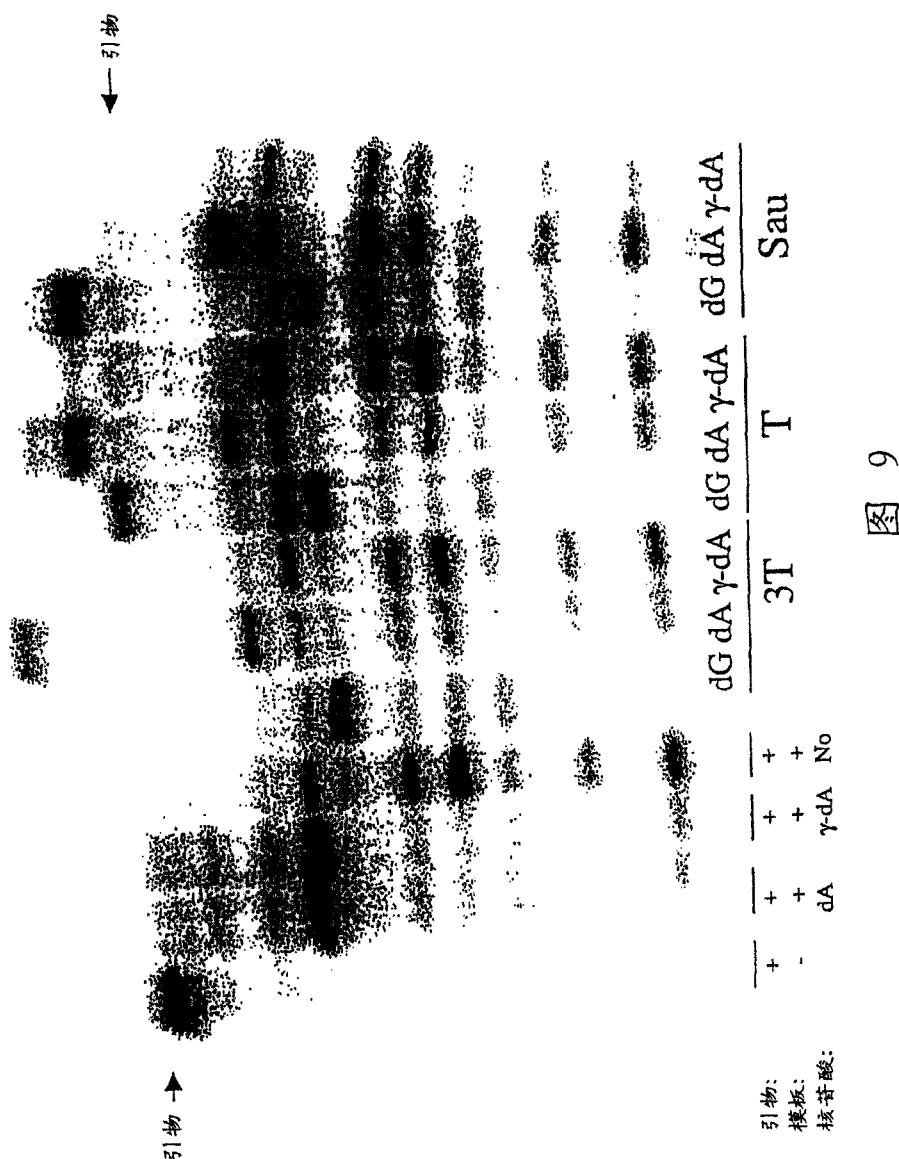


图 9

• 引物链:

Top 5' GGT ACT AAG CGG CCG CAT G 3'

• 模板链:

3T 3' CCA TGA TTC GCC GGC GTA CTT TC 5'
Sau 3' CCA TGA TTC GCC GGC GTA CCT AG 5'

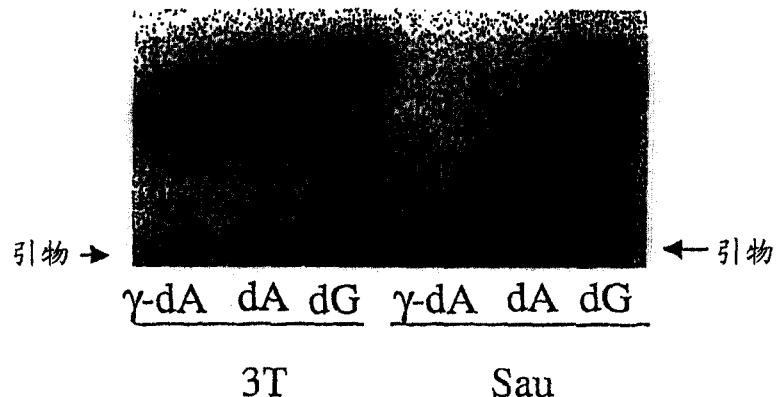


图 10

引物链:

| | | | | | | | | | |
|-----|----|-----|-----|-----|-----|-----|-----|---|----|
| Top | 5' | GGT | ACT | AAG | CGG | CCG | CAT | G | 3' |
|-----|----|-----|-----|-----|-----|-----|-----|---|----|

模板链:

| | | | | | | | | | | |
|---------|----|-----|-----|-----|-----|-----|-----|-----|----|----|
| BOT-3T | 3' | CCA | TGA | TTC | GCC | GGC | GTA | CTT | TC | 5' |
| BOT-Sau | 3' | CCA | TGA | TTC | GCC | GGC | GTA | CCT | AG | 5' |

| 酶: | 无 | T7 | T7 | Seq | Seq | T7 | | | 测序酶 | | | Taq | | | | | | |
|------|---|----|-------------|-----|-------------|--------|----|---------|-------------|----|-----|---------------|----|---------------|----|-------------|----|-------------|
| 引物: | + | + | + | + | + | + | + | + | + | + | + | + | + | | | | | |
| 模板: | - | + | - | + | | BOT-3T | | Sau | BOT-3T | | Sau | BOT-3T | | | | | | |
| 核苷酸: | - | dA | γ dA | dA | γ dA | dG | dA | (spill) | γ dA | dG | dA | γ dAdG | dA | γ dAdG | dA | γ dA | dA | γ dA |

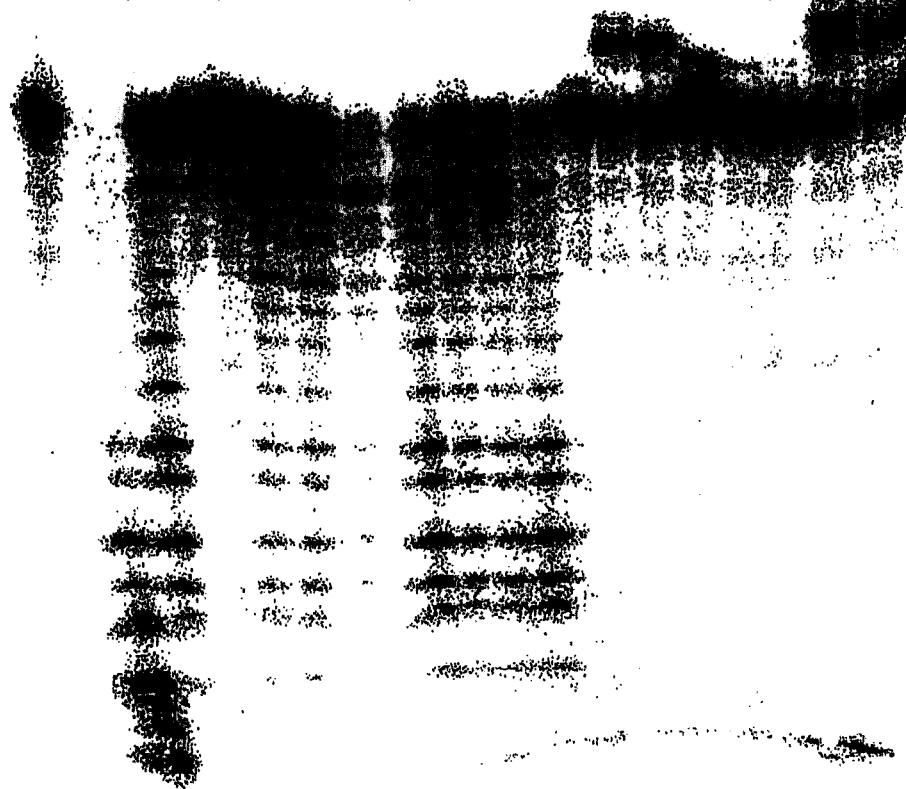
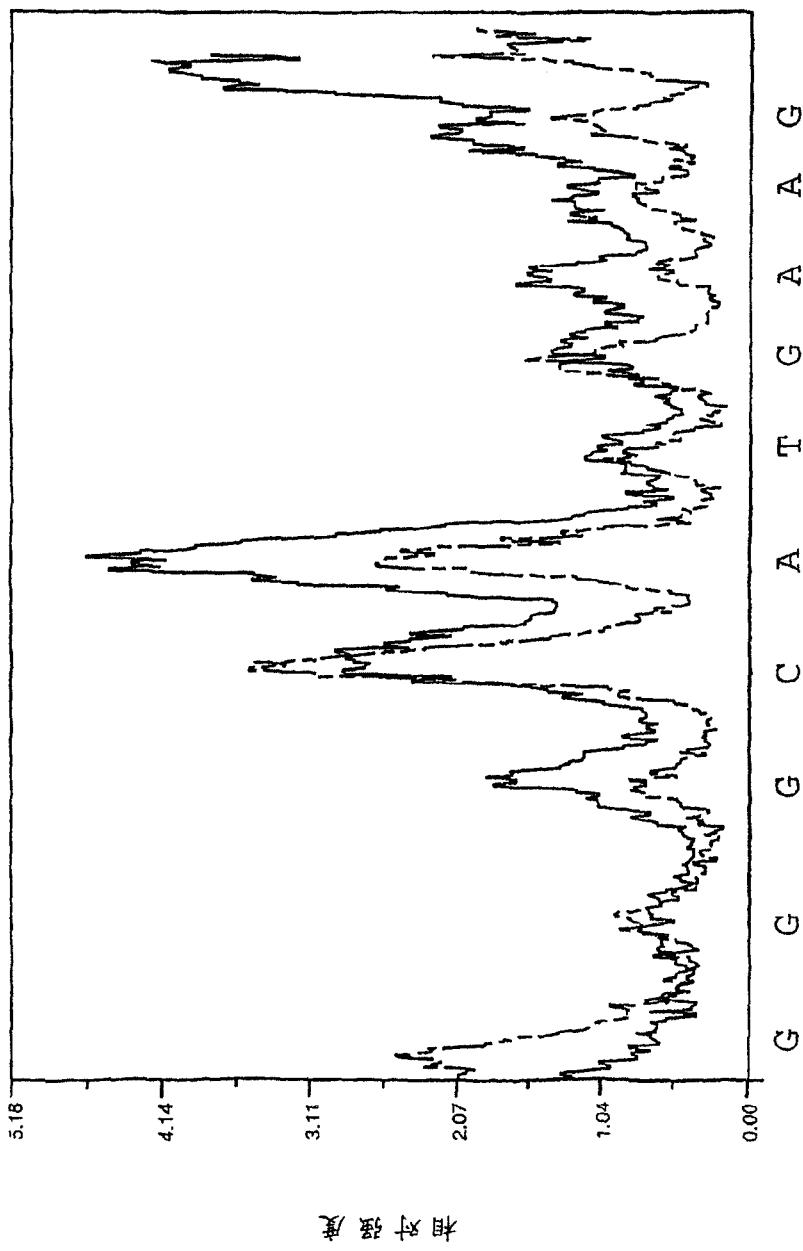


图 11



信号强度和反应动力学提供有关碱基身份的信息。延伸DNA链中的每一核苷酸的信号强度用于测定，确认或支持碱基身份的数据。其中绿实线表示当合成反应中包含四种天然核苷酸(dATP, dCTP, dGTP, dTTP)时产生的反应产物。红虚线表示当反应中包含特有的，碱基修饰的核苷酸时所产生的反应产物。如明确显示的，序列的连贯性和碱基修饰影响反应产物强度和/或反应动力学，这些鉴定特征整合到特有的选择碱基的软件中以在每一测序位置上获得有关碱基身份的高置信值。