



(12)发明专利

(10)授权公告号 CN 105283558 B

(45)授权公告日 2019.05.10

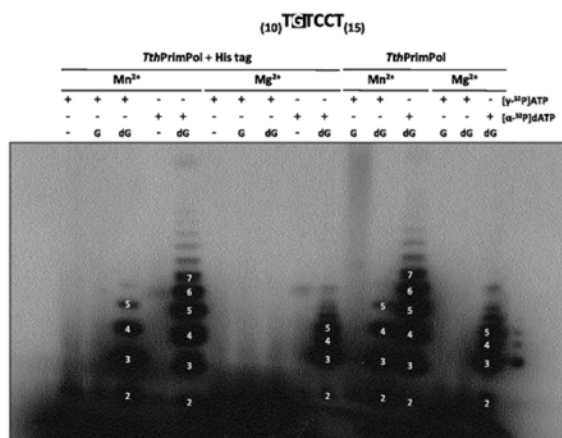
(21)申请号 201480026962.2  
 (22)申请日 2014.03.14  
 (65)同一申请的已公布的文献号  
 申请公布号 CN 105283558 A  
 (43)申请公布日 2016.01.27  
 (30)优先权数据  
 13159629.8 2013.03.15 EP  
 (85)PCT国际申请进入国家阶段日  
 2015.11.11  
 (86)PCT国际申请的申请数据  
 PCT/EP2014/055158 2014.03.14  
 (87)PCT国际申请的公布数据  
 W02014/140309 EN 2014.09.18  
 (73)专利权人 西格尼斯生物技术有限责任公司  
 地址 西班牙马德里卡斯蒂利亚大道123号  
 3°B  
 (72)发明人 安赫尔·J·切尔  
 路易斯·布兰科  
 (74)专利代理机构 北京布瑞知识产权代理有限公司 11505  
 代理人 张丹

(51)Int.Cl.  
 C12Q 1/6869(2018.01)  
 C12N 15/10(2006.01)  
 C12N 9/12(2006.01)  
 (56)对比文件  
 CN 87105787 A,1988.05.11,  
 Uniprot.ACCESSION NO:Q72JW2 entry  
 version 34.《DATABASE UNIPROT》.2012,  
 Uniprot.ACCESSION NO:Q72JW2 entry  
 version 34.《DATABASE UNIPROT》.2012,  
 Lipps et al.A novel type of  
 replicative enzyme harbouring ATPase,  
 primase and DNA polymerase activity.《The  
 EMBO Journal》.2003,第22卷(第10期),2516-  
 2525.  
 Picher et al.TruePrime is a novel  
 method for whole-genome amplification  
 from single cells based on TthPrimPol.  
 《NATURE COMMUNICATIONS》.2016,第7卷1-16.  
 审查员 张丽颖

权利要求书2页 说明书27页  
 序列表8页 附图20页

(54)发明名称  
 使用热稳定的TthPrimPol扩增和测序的方法

(57)摘要  
 本发明公开了一种使用从极端嗜热菌HB27中提取的热稳定的、双功能复制酶“TthPrimPol”进行核酸复制、扩增和测序的方法。该TthPrimPol酶对于核苷酸改变的模板核酸具有极强的耐受性。因此,本发明还一方面公开了一种用于受损的聚核苷酸模板的复制、扩增和测序方法。



CN 105283558 B

1. 一种用于在无引物条件下DNA的复制、扩增或测序的方法,包括以下步骤:
  - a) 提供
    - i. 聚合酶,或
    - ii. 聚合酶缀合物,其中,聚合酶或聚合酶缀合物的聚合酶部分的序列如序列SEQ ID NO:1所示,并且进一步包括聚合酶和引发酶活性,和
  - b) 提供模板核酸,和
  - c) 提供用于结合到核酸互补链的核苷酸和/或核苷酸衍生物,和
  - d) 提供合适的缓冲液,和
  - e) 提供第二聚合酶,和
  - f) 使步骤a至e中的材料接触适当的时间。
2. 如权利要求1所述的方法,其中所述步骤e)中提供的第二聚合酶选自以下组:原核A家族聚合酶,原核B家族聚合酶。
3. 如权利要求2所述的方法,其中所述第二聚合酶是热稳定性聚合酶。
4. 如权利要求3所述的方法,所述热稳定性聚合酶选自以下组:Taq聚合酶,Tth聚合酶,Pfu聚合酶,Vent聚合酶。
5. 如权利要求1所述的方法,其中所述步骤e)中提供的第二聚合酶为Phi29型DNA聚合酶或其缀合物。
6. 如权利要求1-5任一项所述的方法,其中步骤b)中核酸为受损的核酸。
7. 如权利要求6所述的方法,其中受损的核酸如下定义:
  - a) 包括单链或双链断裂,和/或
  - b) 包括化学改性核苷酸和/或
  - c) 来源于甲醛或para-聚甲醛嵌入的组织。
8. 如权利要求7所述的方法,所述化学改性核苷酸选自氧化的核苷酸和/或脱氨基的核苷酸。
9. 如权利要求7或8所述的方法,所述化学改性核苷酸选自以下组:7,8-二氢-8-氧桥腺嘌呤(7,8-dihydro-8-oxoadenine),7,8-二氢-8-氧桥鸟嘌呤(7,8-dihydro-8-oxoguanine),胸腺嘧啶乙二醇(thymine glycol),5-羟基胞嘧啶(5-hydroxy cytosine)或5-羟基尿嘧啶(5-hydroxy uracil)。
10. 如权利要求1所述的方法,其中,步骤c)中至少一种核苷酸衍生物为化学改性的核苷酸。
11. 如权利要求10所述的方法,其中,步骤c)中的核苷酸衍生物包括至少一种选自以下组的核苷酸衍生物:
  - 氧化的核苷酸衍生物,和/或
  - 脱氨基的核苷酸,
  - 包括至少一种空间位阻侧基的衍生物。
12. 如权利要求1所述的方法,其中,步骤b)中模板核酸来自少于100个细胞。
13. 如权利要求12所述的方法,其中,步骤b)中模板核酸来自单细胞。
14. 如权利要求1所述的方法,其中污染的DNA的扩增被抑制。

15. 序列SEQ ID NO:1所示的聚合酶或聚合酶缀合物在权利要求1-14任一项所述方法中的应用。

16. 一种用于检测改性的方法,尤其是用于检测氧化过程中的产物,在DNA模板中,包括以下步骤a)-f)所示的DNA的复制:

a) 提供

i. 聚合酶,或

ii. 聚合酶缀合物,

其中,聚合酶或聚合酶缀合物的聚合酶部分的序列如SEQ ID NO:1所示,并且进一步包括聚合酶和引发酶活性,和

b) 提供模板核酸,和

c) 提供用于结合到核酸互补链的核苷酸和/或核苷酸衍生物,和

d) 提供合适的缓冲液,和

e) 任选地提供一种或多种引物,和

f) 使步骤a至e中的材料接触适当的时间,任选地,在高温下,和

g) 根据多条序列读数计算结合至每个核苷酸位点的核苷酸在步骤a) 整个序列读数中的相对比例,和

h) 将结合的核苷酸的总比例与模板中DNA改性的原始种类相关联。

17. 如权利要求16所述的方法,其中所述高温为温度超过40°C。

18. 如权利要求17所述的方法,其中所述高温为温度超过50°C。

19. 如权利要求16所示的方法,进一步包括如权利要求1-5任一项所定义的方法中的a)到f)步骤。

20. 一种用于随机突变的方法,包括以下步骤:

a) 将核酸与氧化化合物相接触,其中,所述接触导致核酸碱基的氧化,和

b) 通过如序列SEQ ID NO:1所示的聚合酶或聚合酶缀合物复制被氧化的核酸。

## 使用热稳定的TthPrimPol扩增和测序的方法

### 技术领域

[0001] 广义上说,本发明涉及一种极端嗜热菌(*Thermus thermophilus*) HB27的热稳定引发酶/聚合酶蛋白TthPrimPol,及其在核酸合成、扩增和测序方法中的应用。此外,本发明公开了无引物方法,包括TthPrimPol,用于受损的DNA或RNA的扩增,还公开了包含该蛋白的器械和试剂盒。

### 背景技术

[0002] 在迄今已知的几乎所有的DNA复制系统中,复制的起始需要一个由DNA聚合酶提供的游离羟基作为核苷酸的受体。为实现该目的,几种不同的解决方案已经描述了:在所有细胞生命形式以及很多病毒和质粒中,一种专门的RNA聚合酶称为“引发酶”聚合核糖核苷酸,生成一条短链RNA引物随后通过DNA聚合酶延伸(Lodish,Berk等,1999)。在逆转录因子和逆转录病毒中,逆转录酶使用tRNA来启动复制(Lodish,Berk等,1999)。在病毒如噬菌体phi29的蛋白引发(protein-primed)系统中,特定丝氨酸、苏氨酸或酪氨酸侧链的OH基团,作为DNA复制的起始位点(Salas,1991)。最后,在遵循滚环复制模型的病毒中,核酸内切酶在一条DNA链中制造一个缺口,产生游离3OH基团,通过DNA聚合酶延伸(Ilyina和Koonin,1992; Noiro-Gros和Ehrlich,1996)。

[0003] 引发酶可被分为两种进化上不相关的家族:DnaG类引发酶(细菌)和AEP类引发酶(古细菌和真核生物)(Aravind,Leipe等,1998;Iyer,Koonin等,2005)。最近,一个新的AEP的亚种称为“PrimPol”(Lipps 2004;Lipps,Weinzierl等,2004)已经被公开,该亚种零星的存在于革兰氏阳性细菌质粒中。PrimPol同时具有DNA聚合酶和RNA引物酶活性,并且通常与解旋酶相关,以形成一个复制起始复合物。这些特性能使相同的酶在一个系统中同时作用于启动和延伸阶段。就在最近,一种同样包括引发酶和聚合酶的解旋酶的活性已被描述(Sanchez-Berrondo,Mesa等,2011)。不像真核引发酶,古菌质(archaea)引发酶能够分别实现引发使RNA和DNA链长从1KB扩增至7KB(Chemnitz Galal,Pan等,2012;Lao-Sirieix和Bell 2004;Lao-Sirieix,Pellegrini等,2005)。

[0004] 最近,PrimPol具体族成员的酶学性质已被进一步记载,例如García Gómez指出人类PrimPol(HsPrimPol),由人类PRIMPOL基因(也称为CCDC111)编码,能够绕过如氧化核苷酸修饰的DNA模板中脱碱基位点,作用于DNA的维持和被遏止复制叉的再引发(García Gómez,Martínez Jiménez等,2012),而嗜热鸚鵡螺30/1(*Thermococcus nautilus* 30/1)质粒PTN2的耐高温primpol包括末端转移酶活性(WO2011/098588)。

[0005] 来自极端嗜热菌HB27的假设保守蛋白AAS81004.1,预计将包含一个引发-聚合酶(PrimPol),该酶类型为从古菌质粒得到的双功能型复制酶的一种,包括从硫化苔(*Sulfolobus islandicus*)得到的crenarchaeal质粒pRN1(pRN1 primpol)的ORF904蛋白,被克隆。表达于可溶和活性形式的大肠杆菌(*E.coli*)中,并纯化至均一。当深度表征分离的蛋白质(以下称为TthPrimPol)的酶学性质时,本发明人惊奇地发现,除了所预测的引发酶的活性,该蛋白质揭示对于损坏或污染的模板意想不到的高耐受性,有助于发展新的扩增

和测序方法。

[0006] 核酸的扩增,例如通过由Mullis (US5656493) 介绍地聚合酶链式反应(PCR)进行,这是在一项医学和生物学研究中所不可或缺的技术。它已成功地应用于如核酸的克隆、操纵或测序,基于DNA的功能化和基因系统化发育分析,疾病的检测和诊断以及法医和亲子鉴定测序等多个方面。

[0007] 至今为止,存在许多不同的测序技术,通常归类为第一代测序,第二代测序或通常被称为下一代测序(NGS),以及第三代或单分子测序(SMS)。第一代测序主要是指Maxam和Gilbert (Maxam和Gilbert,1977)或Sanger的方法(Sanger,Nicklen等,1977;Sanger和Coulson,1978),只有后者被沿用至今。

[0008] 第二或下一代测序是指在产生许多序列的同时使用先进技术(光学)检测碱基位置方法的技术。现有的方法在一个综述中给出(Metzker,2009)。

[0009] 第三代或单分子测序(SMS)的技术不需要事先扩增,序列无需克隆,也不需要DNA基因组数据库,仅仅是单分子,通常是“实时”检测,在线记录聚合酶的活性(Sam,Lipson等,2011;Thompson和Milos,2011)。

[0010] 尽管如此,仍需要一个新的热稳定的、高过程性的属于primpol DNA/RNA聚合酶家族的聚合酶,其可在高温下工作,能够生产出长延伸产物,优选多于5kb的长度,无需额外的蛋白质的存在。

[0011] 一种易于生产、具有DNA引发酶活性,能够合成或扩增未知的DNA模板,优选在无引物的情况下能够产生具有保真度长延伸产物的热稳定的DNA聚合酶是尤其令人期待的。

[0012] 特别是在法医领域和临床扩增应用,以及第二和第三代测序中,对各种核苷酸底物具有高度耐受的,能够操纵质量往往较差的经修饰的模板的聚合酶是尤其令人感兴趣的。

[0013] 因此,本发明的目的之一是提供一种扩增方法,特别是用于DNA或RNA模板的无引物扩增,即使该模板质量较差。

[0014] 本发明的另一个目的是,提供一种具有保真度的RNA的逆转录方法,即使在高温下。

[0015] 本发明权利要求书及下述公开的内容的主题致力于解决本发明的技术问题。

## 发明内容

[0016] 本发明涉及如本发明和权利要求中所定义的主题。所述主题包括一种用于核苷酸的复制,扩增或测序的方法,包括如下步骤:提供TthPrimPol酶;提供模板核酸;提供用于结合至核酸互补链的核苷酸和/或核苷酸类似物;并提供合适的缓冲液;任选地,提供一种或多种引物;使上述材料接触一段适当的时间;任选地,在高温下。

[0017] 在另一方面,本发明涉及一种用于DNA的无引物扩增的方法-任选地,包括通过用于扩增反应的两种聚合酶协作合成DNA-包括如下步骤:提供TthPrimPol酶;并提供第二聚合酶,特别是Phi29型聚合酶;提供模板核酸;提供用于结合到核酸互补链的核苷酸和/或核苷酸类似物;并提供合适的缓冲液;使上述材料接触一段适当的时间;任选地,在高温下。

[0018] 在其它方面,本发明涉及一种用于cDNA的合成或RT PCR的方法,包括如下步骤:提供TthPrimPol酶;提供模板RNA;提供dNTPs;并提供合适的缓冲液;任选地,提供一种或多种

引物;使上述材料接触一段适当的时间;任选地,在高温下。

### 附图说明

[0019] 图1显示了在单链寡核苷酸上TthPrimPol的DNA引发酶的活性,其中,潜在引发酶识别序列(GTCC)侧接在胸腺嘧啶残基上(Cavanaugh和Kuchta,2009),该嘧啶序列为用于通过几种病毒、原核和真核RNA引发酶启动引发反应的优选的模板范围(Holmes, Cheriathundam等,1985;Parker和Cheng 1987;Frick和Richardson 2001)。引发仅发生在“TC”序列之前,而在聚dT示踪的对面不发生引发。在锰存在下,作为引物(5'端)的核苷酸可以是核糖核苷酸(ATP)或脱氧核苷酸(dATP),但当镁作为金属辅因子时,仅可为脱氧核苷酸(dATP)。不考虑金属辅因子,添加的核苷酸(3'端)必须为脱氧核苷酸(dGTP)。条件:1 $\mu$ M DNA (SEQ ID NO:3),400nM TthPrimPol,5mM MgCl<sub>2</sub>,1mM MnCl<sub>2</sub>,16nM[ $\gamma$ -<sup>32</sup>P]ATP,16nM[ $\alpha$ -<sup>32</sup>P]dATP,10 $\mu$ M GTP/dGTP,60',55 $^{\circ}$ C。

[0020] 图2显示了位于富含胸苷的ssDNA上的CTC是TthPrimPol优选的模板启动位点,其中ssDNA的序列参见序列表中SEQ ID No:4。模板TC起始位点之前的核苷酸具有很强的影响,当C是起始位点之前的核苷酸时,具有最高的引发酶活性。条件:1 $\mu$ M DNA (SEQ ID NO:4),400nM TthPrimPol+His tag,5mM MgCl<sub>2</sub>,1mM MnCl<sub>2</sub>,16nM[ $\gamma$ -<sup>32</sup>P]ATP,1 $\mu$ M dGTP,在55 $^{\circ}$ C温度下反应60min。

[0021] 图3显示了单链环状DNA模板(M13mpl8ssDNA)上的TthPrimPol的DNA引发酶的活性。TthPrimPol用于合成时严格依赖于脱氧核苷酸,而在核糖核苷酸的存在下几乎没有产品被合成。条件:400ng M13mpl8ssDNA,400nM TthPrimPol+His tag,1mM MnCl<sub>2</sub>,16nM[ $\gamma$ -<sup>32</sup>P]ATP,100nM ATP/dATP,1 $\mu$ M CTP/dCTP,1 $\mu$ M GTP/dGTP,1 $\mu$ M UTP/dTTP,在55 $^{\circ}$ C温度下反应60min。

[0022] 图4显示了TthPrimPol的DNA聚合活性。通过由聚合脱氧核苷酸按照所提供的模板(SEQ ID NO:5)该酶能够完全延伸引物分子(SEQ ID NO:6)。镁和锰是进行DNA聚合有效的金属辅因子,虽然锰似乎是更有效的。条件:5nM DNA,1mM MnCl<sub>2</sub>,5mM MgCl<sub>2</sub>,100 $\mu$ M dNTPs,在40 $^{\circ}$ C温度下反应60min。

[0023] 图5显示了在有缺口的DNA双链(序列为SEQ ID NOs 7-9)上TthPrimPol的DNA/RNA聚合活性。在RNA聚合的情况下显示出低效率,并似乎严格依赖于锰。此外,对于有缺口的DNA(gapped DNA),锰极强地促进链置换合成,允许使用整个模板序列。条件:5nM DNA,1mM MnCl<sub>2</sub>,5mM MgCl<sub>2</sub>,100 $\mu$ M NTPs/dNTPs,在40 $^{\circ}$ C温度下反应30min。

[0024] 图6描述了TthPrimPol也是一种RNA指导的聚合酶。图中显示了从DNA引物(SEQ ID NO:8)为起始的序列为SEQ ID NO:12的RNA模板的逆转录。反应的效率低于使用DNA模板所达到的效果,RNA指导的DNA聚合酶活性似乎严格依赖于锰。使用相同的引物和DNA模板(SEQ ID NO:10)进行对比,通过对比结果得到DNA指导聚合酶活性。条件:5nM DNA/RNA,1mM MnCl<sub>2</sub>,5mM MgCl<sub>2</sub>,1 $\mu$ M dNTPs,在40 $^{\circ}$ C温度下反应60min。

[0025] 图7描述了采用TthPrimPol的高保真度DNA合成。无论是在镁或锰离子存在下,4种dNTPs被分别单独地测定作为一种底物与相对的四种可能的模板碱基相结合(模板序列参见SEQ ID NO:13;引物序列参见SEQ ID NO:14)。在所有的情形中,优先地,由第一个可用的模板碱基指导TthPrimPol高保真地插入互补核苷酸。即使含有锰,当延伸至正确配对的引

物对末端时,在由TthPrimPol提供的高保真度DNA聚合期间进行碱基辨别从而形成正确的Watson和Crick碱基对。条件:10nM DNA,在40℃温度下反应20min,5mM MgCl<sub>2</sub>,200nM TthPrimPol,1μM dNTPs(图中较高的面);1mM MgCl<sub>2</sub>,20nM TthPrimPol,100nM dNTPs(图中较低的面)。

[0026] 图8显示了相对于它的底物脱氧核糖核苷酸dTTP,dGTP,dCTP和dATP,TthPrimPol亲和性相类似。在单循环条件下,酶浓度高于DNA浓度,dNTPs的结合清楚地表明TthPrimPol对于dNTPs具有的高亲和性,且对于任意一种dNTP并无特定的倾向性。模板序列参见SEQ ID NO:13;引物序列参见SEQ ID NO:14。条件:10nM DNA,1mM MnCl<sub>2</sub>,50nM TthPrimPol,在40℃温度下反应20min。

[0027] 图9描述了TthPrimPol对于受损的DNA的高耐受性。蛋白由列表SEQ ID:13所示的模板提供,包括碱基位点(AP),7,8-二氢-8-氧桥腺嘌呤(8oxoA),7,8-二氢-8-氧桥鸟嘌呤(8oxoG),胸腺嘧啶乙二醇(Tg),5-羟基胞嘧啶(5OHC)或5-羟基尿嘧啶(5OHU)。如图所示,TthPrimPol能够通过损伤点的对面引物(SEQ ID NO:14)的模板依赖性延伸或通过跳过损伤点并复制下一个模板碱基(dG的插入)从而进行高效的病变旁路。条件:10nM DNA,1mM MnCl<sub>2</sub>,50nM TthPrimPol,100μM dNTPs,在40℃温度下反应20min。

[0028] 图10描述了8oxoG(7,8-二氢-8-氧桥鸟嘌呤)的高效旁路,这是在DNA氧化应激的后果中最常见损害形式之一(Berquist和Wilson,2012),主要为引入的dATP的顺式异构体。dATP插入损伤点的对面与dATP插入未受损的T的动力学特征是一致的。模板参见序列表中SEQ ID NO:13,引物参见序列表中SEQ ID NO:14。条件:10nM DNA,1mM MnCl<sub>2</sub>,50nM TthPrimPol,在40℃温度下反应20min。

[0029] 图11显示了TthPrimPol缺乏末端转移酶活性以及外切酶活性。TthPrimPol不能将任何核苷酸添加至同聚的ssDNA分子(SEQ ID NOs 15和16)的3'末端,这意味着TthPrimPol完全是一个模板依赖性酶。控制反应条件不包括dNTPs,也证实了TthPrimPol缺乏3'-5'核酸外切酶活性,由于DNA分子未退化,与之相同的,也缺乏ExoI,ExoII及ExoIII共有基序,能够在几种DNA聚合酶成员上形成一个进化上保守的3'-5'外切酶的活性位点(Bernad,Blanco等,1989)。条件:5nM DNA,400nM TthPrimPol,1mM MnCl<sub>2</sub>,5mM MnCl<sub>2</sub>,100μM dNTPs,在58℃温度下反应60min。

[0030] 图12描述了通过TthPrimPol的ssDNA高效无引物扩增。在dNTP底物的存在下,TthPrimPol(+/-His tag)的全部情形都能够大量扩增单链M13mp18模板DNA,但是如果不额外添加特定或随机引物,也清楚地表明TthPrimPol具有同样的DNA引发酶和DNA聚合酶的作用。另一方面,核糖核苷酸(NTPs)似乎是通过TthPrimPol的DNA扩增的不良底物。条件:50mM Tris-HCl pH 7.5,1mM DTT,0.1μg/μl BSA,50mM NaCl,5%甘油,1mM MnCl<sub>2</sub>,200ng M13mp18ssDNA,在65℃温度下反应6h。

[0031] 图13显示了TthPrimPol依赖于金属辅因子。在1mM Mn a)或5mM Mg b)的存在下,扩增效率最高,而在任意测试浓度下,Co和Ca都不适合作为辅因子c)和d)。条件:50mM Tris-HCl pH 7.5,1mM DTT,0.1μg/μl BSA,32mM NaCl,5%甘油,200ng M13mp18ssDNA,400nM TthPrimPol,在65℃下反应3h。

[0032] 图14描述了在无引物分析中M13mp18ssDNA的扩增效率与所使用的TthPrimPol的量成比例,无论是否有Mg<sup>2+</sup>a)或Mn<sup>2+</sup>b)的存在。条件:50mM Tris-HCl pH 7.5,1mM DTT,0.1μ

g/ $\mu$ l BSA, 50mM NaCl, 5% 甘油, 100mM dNTPs, 200ng M13mpl8ssDNA, 400nM TthPrimPol, 在65 $^{\circ}$ C下反应3h。

[0033] 图15描述了单链模板DNA浓度在DNA扩增效率中的作用。M13mpl8ssDNA通过TthPrimPol的无引物扩增。产物的量与最初模板DNA的量成比例, 无论是否有Mg<sup>2+</sup> (左侧) 或Mn<sup>2+</sup> (右侧) 的存在。条件: 50mM Tris-HCl pH 7.5, 1mM DTT, 0.1 $\mu$ g/ $\mu$ l BSA, 32mM NaCl, 5% 甘油, 100 $\mu$ M dNTPs, 400nM TthPrimPol, 5mM MgCl<sub>2</sub>, 1mM MnCl<sub>2</sub>, 在65 $^{\circ}$ C下反应3h。

[0034] 图16显示了在通过TthPrimPol的M13mpl8ssDNA无引物扩增中, Mg<sup>2+</sup>a) 或Mn<sup>2+</sup>b) 的存在下, 反应时间依赖于扩增产物的产率。在其他情况下, 扩增在4h后达到饱和, 或许是由于dNTP匮乏。条件: 50mM Tris-HCl pH7.5, 1mM DTT, 0.1 $\mu$ g/ $\mu$ l BSA, 50mM NaCl (当MgCl<sub>2</sub>作为辅因子时, 为32mM), 5% 甘油, 100 $\mu$ M dNTPs, 400nM TthPrimPol, 200ng M13mpl8ssDNA, 5mM MgCl<sub>2</sub>, 1mM MnCl<sub>2</sub>, 65 $^{\circ}$ C。

[0035] 图17显示了扩增产物的数量变化和迁移率变化作为dNTP浓度的函数。对于每种金属辅因子, 最佳的dNTP浓度范围为100 $\mu$ M (使用锰时) 至500 $\mu$ M (使用镁时)。过高的浓度会抑制反应。在Mg<sup>2+</sup> (左侧) 和Mn<sup>2+</sup> (右侧) 分别存在下, 对于各金属来说, 在有效区间的dNTP中, 扩增产物具有相似的迁移率及产率。条件: 50mM Tris-HCl pH 7.5, 1mM DTT, 0.1 $\mu$ g/ $\mu$ l BSA, 32mM NaCl, 5% 甘油, 400nM TthPrimPol, 200ng M13mpl 8ssDNA, 5mM MgCl<sub>2</sub>, 1mM MnCl<sub>2</sub>, 在65 $^{\circ}$ C下反应3h。

[0036] 图18描述了孵育温度对ssDNA扩增中使用的金属辅因子的影响。最佳温度范围, 相对于镁(55-60 $^{\circ}$ C)来说, 锰的范围更大(52-62 $^{\circ}$ C)。条件: 50mM Tris-HCl pH 7.5, 1mM DTT, 0.1 $\mu$ g/ $\mu$ l BSA, 32mM NaCl, 5% 甘油, 400nM TthPrimPol, 200ng M13mpl 8ssDNA, 100 $\mu$ M dNTPs, 5mM MgCl<sub>2</sub>, 1mM MnCl<sub>2</sub>, 3h。

[0037] 图19显示了随机引物(random primers, RP) 被TthPrimPol所取代, 随后通过Phi29 DNA聚合酶进行多重置换扩增。这两种酶的组合能够熟练扩增双链质粒DNA。在200nM TthPrimPol的存在下, 扩增产物量的增加依赖于TthPrimPol剂量, 扩增的DNA的产率和迁移率与所使用的随机寡核苷酸引物(RP) 相当。条件: 1ng pRSET DNA (3kb), 50 $\mu$ M 随机引物(RP), 40ng Phi29 DNA聚合酶, 400nM TthPrimPol, 10mM MgCl<sub>2</sub>, 500 $\mu$ M dNTPs, 在30 $^{\circ}$ C下反应5h。

[0038] 图20描述了在Phi29全基因组扩增(WGA) 的实验系(N=3) 中, 扩增产物的TthPrimPol剂量-依赖性上升。TthPrimPol和Phi29 DNA聚合酶(DNApol) 的组合, 能够熟练地扩增基因组DNA, 其产率与公认的使用随机引物(RP) 的WGA方法相当。条件: 650ng Phi29 DNApol (QualiPhi), 400nM TthPrimPol, 1ng基因组DNA (genomic DNA), 50 $\mu$ M 随机引物(random primers), 10mM MgCl<sub>2</sub>, 500 $\mu$ M dNTPs, 16h 30 $^{\circ}$ C。

[0039] 图21显示了TthPrimPol具有长时间的稳定性和活性。扩增的核酸的量与2011年7月7日的采用相同批次TthPrimPol酶所扩增的核酸的量相当, 超过19个月(2013年3月6日)。图中所示为单链M13mpl8ssDNA的无引物扩增a); 存储TthPrimPol超过16个月不改变其模板依赖性的延伸引物的能力b)。条件: a) 50mM Tris-HCl pH7.5, 1mM DTT, 0.1 $\mu$ g/ $\mu$ l BSA, 32mM NaCl, 5% 甘油, 400nM TthPrimPol, 200ng M13mpl8ssDNA, 100 $\mu$ M dNTPs, 3h 65 $^{\circ}$ C; b) 5nM DNA, 5mM MgCl<sub>2</sub> 100 $\mu$ M dNTPs, 60' 40 $^{\circ}$ C, 模板和引物序列参见SEQ ID NOS: 5和6。

[0040] 图22描述了TthPrimPol是比HsPrimPol效率更高的DNA聚合酶。如图所示是在NTPs



和dNTPs的存在下M13mpl8ssDNA的扩增。甚至较小量(200nM)的TthPrimPol产生比Hs方案更多扩增的核酸。另一方面,这两种酶利用NTPs的效率比dNTPs低,但在类似的范围内。

[0041] 图23显示了TthPrimPol能够结合荧光标记的sNTPs。如图所示为Cy5标记的dCTP。条件:10nM DNA (SEQ ID NOS:10和8),100nM TthPrimPol,1mM MnCl<sub>2</sub>,5mM MgCl<sub>2</sub>,在40℃下反应20min。

[0042] 图24描述了本发明所用的氨基酸和核酸。

[0043] 图25描述了通过TthPrimPol和Phi29 DNA聚合酶的组合进行有效扩增时模板DNA所需的最小量。

[0044] TthPrimPol和野生型Phi29 DNA聚合酶的组合能够从最少10fg的模板DNA中熟练地扩增人基因组DNA。如图所示为两个独立(图25a和图25b)的实验,每个反应条件重复三次。孵育混合物包括:50μL,50mM Tris-HCl pH 7.5,50mM KCl,1mM DTT,10mM MgCl<sub>2</sub>,500μM dNTPs,700ng野生型Phi29 DNA pol,400nM TthPrimPol以及指定量的人基因组DNA。NTC(无模板对照)表示无模板DNA参加反应。反应混合物在30℃下孵育6小时。扩增产物使用PicoGreen试剂(Quant-iT™PicoGreen dsDNA reagent,Invitrogen)定量。星号表示在三个反应中有两个未能进行DNA扩增。X轴:人基因组模板DNA的量;Y轴:DNA产率,μg。

[0045] 图26描述了TthPrimPol和野生型Phi29 DNA聚合酶的组合能够抑制非特异性的背景DNA扩增。

[0046] TthPrimPol和野生型Phi29 DNA聚合酶的组合,在没有外部加入模板DNA的情况下,任何孵育时间下,都不能进行DNA扩增(图26a)。另一方面,Phi29 DNA与随机引物的结合引起显著的DNA产率增加,即使反应仅孵育3小时。因此,使用TthPrimPol能够防止在没有外部添加模板DNA时非特异性背景DNA扩增。相比之下,TthPrimPol和Phi29 DNA聚合酶的组合,在外部添加模板DNA分子(图26b)的情况下,能够产生可比量的扩增的DNA。

[0047] 每个反应条件重复三次。孵育混合物包括:50μL,50mM Tris-HCl pH 7.5,50mM KCl,1mM DTT,10mM MgCl<sub>2</sub>,500μM dNTPs,700ng QualiPhi Phi29 DNA pol(改良版酶),400nM TthPrimPol或50μM随机引物。NTC(无模板对照)表示无模板DNA从外部加入反应。DNA表示1g人基因组DNA作为模板加入反应。反应混合物在30℃下孵育几小时。扩增产物使用PicoGreen试剂(Quant-iT™PicoGreen dsDNA reagent,Invitrogen)定量。

[0048] 图27描述了TthPrimPol和野生型Phi29 DNA聚合酶的组合对酵母基因组DNA的扩增。TthPrimPol和野生型Phi29 DNA聚合酶的组合能够熟练地扩增酵母基因组DNA,产率与成熟的使用随机引物(RP)的WGA方法相当。

[0049] 孵育混合物包括:体积50μL,50mM Tris-HCl pH 7.5,50mM KCl,1mM DTT,10mM MgCl<sub>2</sub>,500μM dNTPs,1ng酵母菌属酵母基因组DNA(Saccharomyces cerevisiae genomic DNA),750ng野生型Phi29 DNA pol(Repli-G single cell version,Qiagen),50μM随机引物或指定量的TthPrimPol。反应混合物在30℃下孵育3小时。扩增产物使用PicoGreen试剂(Quant-iT™PicoGreen dsDNA reagent,Invitrogen)定量。

[0050] 图28显示了延伸获得的MiSeq序列读取内容映射到参考基因组。这详细说明了每个样本读取的总量、映射到基因组的读取的量以及没有被映射到参考基因组的读取的量。它进一步描述了映射和未映射读取的每种样品碎片,质量剪切后的平均读取长度,以及映射或没有被映射到基因组的碱基的碎片的量。

[0051] 最显著的是,非扩增样本及Tth/Phi扩增样本映射到相同的延伸度(>90%),而仅83.92%的随机引物样本(RP)映射到参考物。

[0052] 图29显示了在个人染色体基础上特征的过量、和代表性不足性分析。比较随机引物(RP)和Tth/Phi扩增的样品与非扩增的样本(NA)揭示了,从NA样本到Tth/Phi样本,相对差下降了17%。

### 具体实施方式

[0053] 本发明从以下几方面进行描述,并在权利要求中对其进行限定。

[0054] 从极端嗜热菌HB27中的获得的假设保守的蛋白AAS81004.1,预计包含一个发现于来自古细菌质粒中的双功能复制酶的引发酶-聚合酶(PrimPol)型区域,用于克隆,以可溶和活性形式高度表达于大肠杆菌中,并能高度纯化,古细菌质粒包括来自硫化叶菌(*Sulfolobus islandicus*) (pR 1pripol)的crenarchaeal质粒pR 1的ORF904蛋白。

[0055] 发明人发现纯化后的TthPrimPol在胸腺嘧啶残基两侧的具有潜在引发酶识别序列(GTCC)的单链寡核苷酸中显示出很强的DNA引发酶活性(Cavanaugh和Kuchta,2009)。这样的嘧啶序列已被证明是通过几种病毒、原核和真核RNA引发酶启动引发反应的优选的模板(Holmes,Cheriathundam等,1985;Parker和Cheng,1987;Frick和Richardson,2001)。出乎意料的是,发明人发现引发仅发生在“TC”序列之前,而在聚dT示踪的对面不发生引发。进一步的模板序列需求分析揭示了在模板起始位点之前的核苷酸的TthPrimPol引发酶活性效果-C优于A,G或T。虽然TthPrimPol优选CTC作为模板启示位点,但它通常能够作为同类形式XTC的任意序列的引发酶,X选自A,C,G或T。合适的序列需求是几乎所有天然模板的随机引发的良好基础。

[0056] 发明人进一步发现TthPrimPol表现出DNA-依赖性DNA聚合酶活性,这也被其它AEP引发酶所证实,能够实现RNA和DNA链的引发和延伸,将其分别延伸至1kb或7kb(Chemnitz Galal,Pan等,2012;Lao-Sirieix和Bell,2004;Lao-Sirieix,Pellegrini等,2005)。此外,发明人发现TthPrimPol能够将RNA逆转录为DNA,因此能够作为逆转录酶,当作用在有缺口的模板上时,具有链置换活性。然而,在RNA聚合酶的情况下,TthPrimPol严格依赖于锰作为金属辅因子,且效率较低。

[0057] 这些研究结果给出TthPrimPol利用多种模板和底物的能力的提示。因此,发明人测试了TthPrimPol对于受损的DNA的耐受性。蛋白的模板包括碱基位点(AP),7,8-二氢-8-氧桥腺嘌呤(8oxoA),7,8-二氢-8-氧桥鸟嘌呤(8oxoG),胸腺嘧啶乙二醇(Tg),5-羟基胞嘧啶(5OHC)或5-羟基尿嘧啶(5OHU)。发明人惊奇的发现TthPrimPol能够通过将核苷酸插入损伤点的对面或通过跳过损伤点并复制下一个模板基从而进行高效的病变旁路。8oxoG,这是在DNA氧化应激的后果中最常见损害形式之一(Berquist和Wilson,2012),是引入的dATP的顺式异构体的主要和有效的旁路。

[0058] 进一步的,发明人证实了TthPrimPol具体强的链置换活性使之成为一个对于全基因扩增(WGA)来说有吸引力的聚合酶,即使在无引物环境中或用于共价闭合的DNA滚环扩增(RCA)中也一样。

[0059] 虽然上述研究结果显示了TthPrimPol的普遍酶性质,但酶在实验室日常应用中也依赖于碱基蛋白的性质如溶解度和稳定性。不像HsPrimPol,一旦被隔离立刻会降低引发酶

活性,这可能是由于更复杂的结构,在靠近C-端存在锌指结构域(García Gómez,Martínez Jiménez等,2012),发明人发现无锌指结构的TthPrimPol具有高度稳定性,可在很长时间内保持引发酶和聚合酶活性。令人惊奇的是,发明人发现当在-80℃下存储超过20个月(图21),无论是TthPrimPol的引发酶还是聚合酶活性都不会显著降低,这使其成为一个有吸引力的蛋白,能够用于开发出一个所有日常实验均可用的可信赖的方法。

[0060] 本文所述的研究结果是为了开发出一种新的,尤其是一种无引物的,用于核酸扩增、复制和测序的可信赖的方法。进一步的,这些研究成果也为了开发一种新的随机突变的方法。

[0061] 因此,在第一方面,本发明提供一种用于核酸复制、扩增或测序的方法,包括以下步骤:

[0062] a. 提供

[0063] i. 聚合酶,或

[0064] ii. 聚合酶缀合物,

[0065] 其中,聚合酶或聚合酶缀合物的聚合酶部分包含与序列SEQ ID NO:1至少70%同源的序列并且进一步包括聚合酶和引发酶活性,和

[0066] b. 提供模板核酸,和

[0067] c. 提供用于结合到核酸互补链的核苷酸和/或核苷酸类似物,和

[0068] d. 提供合适的缓冲液,和

[0069] e. 任选地,提供一种或多种引物,和

[0070] f. 使步骤a至e中的材料接触适当的时间,任选地在高温下。

[0071] 在一个优选的实施方式中,本发明提供一种无引物情况下,用于DNA复制、扩增或测序的方法,包括以下步骤:

[0072] a. 提供

[0073] i. 聚合酶,或

[0074] ii. 聚合酶缀合物,

[0075] 其中,聚合酶或聚合酶缀合物的聚合酶部分包含与序列SEQ ID NO:1至少70%同源的序列并且进一步包括聚合酶和引发酶活性,和

[0076] b. 提供模板核酸,和

[0077] c. 提供用于结合到核酸互补链的核苷酸和/或核苷酸衍生物,和

[0078] d. 提供合适的缓冲液,和

[0079] e. 提供第二种聚合酶,和

[0080] f. 使步骤a至e中的材料接触适当的时间。

[0081] 术语复制,扩增,测序,聚合酶或“聚合酶活性”,以及引物或核酸均为现有技术所公知。

[0082] 术语核酸例如是DNA或RNA的总称,是多核苷酸的同义词。核酸被发现存在于所有的生命中,用于遗传信息编码、传递和表达,是遗传的基础。

[0083] 例如,术语DNA聚合酶活性代表酶催化脱氧核糖核苷酸聚合形成DNA链的活性。DNA聚合酶以其在DNA复制中的作用而出名,DNA的复制是复制一条DNA链的过程,聚合酶“读取”完整的DNA链作为模板并使用它来合成新的链。聚合酶活性的测定可采用例如根据实施例

1) 的方法或任何本领域技术人员所熟知DNA聚合酶活性测定方法。

[0084] 一般而言,术语扩增是指基因能够过表达的多种方法之一。此处所使用的术语扩增主要是指DNA片段或整个DNA分子在体外能够复制或成倍增加的多种方法之一。体外扩增方法的非限制例子包括PCR (Polymerase Chain Reaction, 聚合酶链式反应), LAMP (Loop mediated Isothermal Amplification, 环介导等温扩增), RDC (Reaction displacement chimeric, 反应迁移嵌合)、NASBA或包含Phi29聚合酶的等温扩增。(US5001050; WO/2011/000997) 主要表现为如在共价闭合DNA情况下的f.e.滚环扩增(RCA)或线性基因组DNA的全基因组扩增(WGA)的链置换扩增。

[0085] 由于其热稳定性和链置换能力, TthPrimPol有利于在需要强链置换能力的恒温条件下以及PCR方法中使用。在两种情况下, 由于TthPrimPol允许复制叉的重启动, 特别有益于受损DNA的扩增, 本发明的方法相较于传统的扩增方法更有利。

[0086] 术语DNA测序是指测定DNA分子中核苷酸确切顺序的程序, 包括任意用于测定DNA链中四种碱基-腺嘌呤、鸟嘌呤、胞嘧啶和尿嘧啶顺序的方法和技术。迄今为止, 存在许多不同的测序技术, 通常归类为第一代测序、第二代测序、第三代测序或单分子测序(SMS)。第一代测序主要是指Maxam和Gilbert (Maxam和Gilbert, 1977) 或Sanger (Sanger, Nicklen等, 1977; Sanger和Coulson, 1978) 的方法, 其中只有后者被沿用至今。

[0087] 第二或下一代测序是指在产生许多序列的同时使用先进技术(光学)检测碱基位置方法的技术。现有的方法在综述中给出 (Metzker, 2009)。

[0088] NGS技术是循环测序技术, 也称为通过合成测序技术: Illumina平台 (Bentley, Balasubramanian等, 2008), 生命技术离子激流平台 (Life technologies Ion Torrent) ([www.iontorrent.com/technology-scalability-simplicity-speed](http://www.iontorrent.com/technology-scalability-simplicity-speed))。焦磷酸测序 (Ronaghi, Karamohamed等, 1996; Ronaghi, Uhlen等, 1998) 是最早的下一代测序技术之一。这项技术基于链延伸过程中释放的焦磷酸 (PPi) 的检测。可见光与结合的核苷酸数目成正比。释放的PPi通过ATP硫酸化酶转换为ATP。然后, ATP提供能量使荧光素酶氧化荧光素从而产生闪光。由于核苷酸按预先确定的顺序加入, DNA序列能够被检测。

[0089] 测序程序中使用的荧光修饰的核苷酸经常会给聚合酶的利用带来难度, 因为它们表现得不像用于酶的天然底物。这些核苷酸必须被插入到具有高度特异性的生长链中, 荧光基团被切割掉之后接着成像, 并在下一循环中有效地延伸。在循环测序技术, 修饰的核苷酸的加入导致反应的可逆终止, 因此归类为“可逆终止子”。存在两组可逆终止子: 3'-封闭终止子, 通过可裂解基团键合到2'-脱氧核糖糖3'-氧端以及3'-未封闭终止子。可能的封闭基团包括3'-O-烯丙基 (例如f 3'-O-allyl-dCTP-allyl-bodipy-FL-510 [ $\lambda_{\text{abs}}$  (最大) = 502nm;  $\lambda_{\text{em}}$  (最大) = 510nm], 3'-O-allyl-dUTP-allyl-R6G [ $\lambda_{\text{abs}}$  (最大) = 525nm;  $\lambda_{\text{em}}$  (最大) = 550nm], 3'-O-allyl-dATP-allyl-ROX [ $\lambda_{\text{abs}}$  (最大) = 585nm;  $\lambda_{\text{em}}$  (最大) = 602nm] 3'-O-allyl-dGTP-allyl-bodipy-650 [ $\lambda_{\text{abs}}$  (最大) = 630nm;  $\lambda_{\text{em}}$  (最大) = 650nm] (Ju, Kim等, 2006), 目前被应用于智能生物系统 Intelligent Biosystems / Qiagen ([www.intelligentbiosystems.com](http://www.intelligentbiosystems.com))。使用的另一个封闭基团是3'-O-azidomethyl (Bentley, Balasubramanian等, 2008; Guo, Xu等, 2008), 用于Illumina。这些核苷酸的例子包括: ddCTP-N3-Bodipy-FL-510 ( $\lambda_{\text{abs}}$  (最大) = 502nm;  $\lambda_{\text{em}}$  (最大) = 510nm), ddUTP-N3-R6G ( $\lambda_{\text{abs}}$  (最大) = 525nm;  $\lambda_{\text{em}}$  (最大) = 550nm), ddATP-N3-ROX ( $\lambda_{\text{abs}}$  (最大) = 585nm;  $\lambda_{\text{em}}$  (最大) =

602nm), 和ddGTP-N3-Cy5 ( $\lambda_{\text{abs}}$  (最大) = 649nm;  $\lambda_{\text{em}}$  (最大) = 670nm)。与未修饰的底物相比, 将较大的基团附加到的核苷酸的3'端对于这些核苷酸的加入是不利的。这一问题的通用解决方案是突变DNA聚合酶, 使其更好的耐受3'-终止子核苷酸。

[0090] 3'-未封闭可逆终止子给聚合酶带来更少的问题, 往往与天然底物具有相似的结合特性(Wu, Stupi等, 2007)。这些核苷酸是N6-烷基化及光可裂解的。

[0091] 最近, 一种可能提高准确度和读取长度的更有利的化学作用已经被介绍(Litosh, Wu等, 2011)。这种终止子修饰是基于5-羟甲基-2'-脱氧尿苷三磷酸(HOMedUTP, 5-hydroxymethyl-2'-deoxyuridine triphosphate)。

[0092] 其他的3'-未封闭终止子使用大型染料基团的空间位阻来抑制在第一个改性的核苷酸已结合后再去结合额外核苷酸(Turcatti, Romieu等, 2008)。

[0093] 在本发明第一方面的一个具体的实施方式中, TthPrimPol将很好的应用于下一代测序, 由于TthPrimPol的高持续合成能力、对模板改性DNA的不敏感性及对所使用的核苷酸的高耐受性, TthPrimPol因而是非常合适的。

[0094] 单分子测序技术依赖于核苷酸, 其中荧光基团与末端的磷酸盐相连, 而非碱基(Life Technologies/VisiGen和Pacific Biosciences (Eid, Fehr等, 2009)), 通过聚合酶结合并不是一个大问题, 天然碱基被插入生长的DNA链中。

[0095] 第三代测序或单分子测序(SMS)技术不需要预先扩增, 排序无需克隆或全DNA, 仅需要单分子即可, 通常是“即时”, 这意味着可在线记录聚合酶的活性(Sam, Lipson等, 2011; Thompson和Milos, 2011)。

[0096] 重要的SMS测序平台包括基于FRET的生命技术(Life Technologies FRET-based), 采用量子点与聚合酶连接; Pacific Biosciences, 将Phi29 DNAPol固定在零模(Zero-mode)波导器件的中心, 测量荧光耦合核苷酸的近似值; 或Helicos Biosciences, 引物固定在玻璃表面。

[0097] 一项完全不同的基于“纳米孔”(nanopores)的技术允许DNA或核苷酸的碱基转换通过隔膜进行(Branton, Deamer等, 2008)。这项技术主要市场为Oxford Nanopore Technologies ([www.nanoporetech.com](http://www.nanoporetech.com))。最初的想法是在合成时通过孔穿成一条DNA链, 测量穿过孔的离子电流的变化, (Maglia, Restrepo等, 2008), 另一种可能涉及的方式是使用核酸外切酶, 并检测穿过孔的释放的核苷酸转变(Clarke, Wu等, 2009)。这种技术可以彻底改变测序的价位, 因为不再需要昂贵的荧光标记的核苷酸。

[0098] 这些纳米孔使用的蛋白质包括 $\alpha$ -溶血素(alpha-hemolysin) (Maglia, Restrepo等, 2008; Jetha, Wiggin等, 2009) 或MspA (Derrington, Butler等, 2010; Manrao, Derrington等, 2011; Manrao, Derrington等, 2012; Pavlenok, Derrington等, 2012; Butler, Pavlenok等, 2008)。

[0099] 迄今为止, Liebermann, Manrao或Wendell已描述了Phi29 DNAPol的主要应用(Lieberman, Cherf等, 2010; Manrao, Derrington等, 2012; Wendell, Jing等, 2009)。在本发明第一方面的一个具体实施方式中, TthPrimPol用于联结或共价连接至 $\alpha$ -溶血素(alpha-hemolysin) 或MspA或类似蛋白用于改善单分子纳米孔测序。在该实施方式中使用TthPrimPol的一个特别的优势是对于反应中不同的核苷酸具有高耐受性(例如dNTPs, NTPs, oxoNTPs)。由于大多数应用使用不同改性的核苷酸(如荧光标记), TthPrimPol对改性

核苷酸可能提供更强的持续合成能力和更高的品质。

[0100] 酶是具有特定的生物学功能的复合蛋白生物分子。无损的蛋白质为一级、二级、三级甚至有时为四级结构。

[0101] 本领域技术人员公知,蛋白的生物学功能甚至分子稳定性依赖于其结构的保存,受到许多因素例如温度、冻结或解冻周期、pH、蛋白浓度、盐条件、溶剂或类似因素的影响。此外,蛋白的结构和稳定性甚至受到个体氨基酸的氧化的影响,主要包括蛋氨酸、半胱氨酸、色氨酸、酪氨酸和组氨酸的氧化(Kim,Berry等,2001)。所使用的技术中,蛋白质稳定性通过标准吉布斯自由能变化、 $\Delta G^0$ 定量表征,涉及展开的唯一结构、三维结构或随机盘绕多肽链(Hinz,Steif等,1993)。

[0102] 相比之下,本文使用的术语蛋白或酶的稳定性主要是指蛋白的酶活性在长时间内的保存。

[0103] 在本文中,酶活性的保存优选地倾向于引发酶和聚合酶活性被保存至少为刚离体TthPrimPol活性的30%,40%,50%,优选地,至少为60%,70%,80%,更优选地,至少85%,90%,95%例如近乎100%。

[0104] 用于储存和进行酶促反应的缓冲液是影响酶稳定性和活性的一个因素。在本发明方法中对使用的缓冲液不进行特别的限定,本领域技术人员可根据需要优化本发明方法中用于储存和反应的缓冲液的条件。

[0105] 本发明任一方法中使用的蛋白因此被表征为蛋白保持引发酶和聚合酶活性,如本文中定义,当储存在合适的条件下时,可保持很长一段时间。本领域技术人员公知,蛋白的稳定性和保存期限同时依赖于蛋白的内在属性和存储条件,现有技术中多种方法和技术已经用于优化储存条件。全面的概述,可参考Simpson(Simpson,2005)。

[0106] 在一个具体的实施方式中,本发明第一方面所述的方法因而包括TthPrimPol酶优选地维持其活性,在合适的储存条件下,至少超过4,5,6个月,优选为至少7,8,9个月,优选为,至少10,11,12个月,优选为至少13,14,15个月,优选为至少16,17,18个月,优选为至少19,20,21个月,优选为至少22,23,24个月。

[0107] 在一个具体的实施方式中,本发明方法包括TthPrimPol酶在合适的储存条件维持其活性超过2年。

[0108] 在本文中,本发明任一方法中所使用的蛋白,其序列与SEQ ID NO:1序列至少70%同源,例如与SEQ ID NO:1序列71%,72%,73%,74%,75%,76%,78%,79%,80%,81%,82%,83%,84%,85%,86%,87%,88%,89%,90%,91%,92%,93%,94%,95%,96%,97%,98%,99%同源,任选的,包括聚合酶和/或引发酶活性。

[0109] 在一个最优选的实施方式中,本发明任一方法中所使用的蛋白具有SEQ ID NO:1所示的序列。

[0110] 本领域技术人员很容易确定其比例。作为一个非限定例子,对于包含20个氨基酸的肽,根据第一方面,70%同源于该序列表示14-20个氨基酸是相同的。此外,最主要的生物信息网站,例如NCBI或EBI,提供氨基酸序列相同和/或相似计算以及核苷酸序列相同和/或相似计算的服务。本领域技术人员将知晓通用的BLAST(Basic Local Alignment Search Tool)软件适用于这样的计算(Altschul,Gish等,1990)。

[0111] 本领域技术人员能够知晓本发明任一方法中使用的蛋白,如本文中定义,可以是

含有SEQ ID NO:1蛋白的衍生物,可改变一个或多个氨基酸,和/或插入一个或多个连续的氨基酸,和/或插入连续的氨基酸,和/或去除氨基酸。在本发明任一方法中使用的蛋白衍生物可进一步包括一个或多个共价修饰,如本领域技术人员知晓,本文并不对其进行特别限定。优选地,所述一个或多个共价修饰选自以下组:乙酰化、酰胺化、二硫键形成、甲酰化、糖基化、甲基化、磷酸化、硫酸化。

[0112] 通常,本发明任一方法所使用的蛋白或蛋白衍生物可为嵌合蛋白,嵌合蛋白衍生物或嵌合蛋白缀合物。嵌合肽等并不做特别限定,为本领域技术人员所熟知。

[0113] 优选地,嵌合蛋白或嵌合蛋白衍生物是一种蛋白缀合物,包括本发明任一方法所使用的蛋白或蛋白衍生物,如上文所述。在本文中,“蛋白缀合物”优选指的是一种缀合物包括如上所述的蛋白或蛋白衍生物及其它基团。所述其它基团,在本发明方法中使用的,也可指多于一个上述基团。因此,蛋白缀合物也可包括多于一个上述基团,例如指定型基团两个或更多个基团和/或不同类型基团的两个或更多基团。所述基团优选为能够与核苷酸相互作用的已知基团例如螺旋-发卡-螺旋(Helix-Hairpin-Helix)基团,DNA解旋(DNA unwinding)基团例如促旋酶(gyrase)基团,或单链键合基团例如E.coli's PriB蛋白基团或单纯疱疹(Herpes simplex)的icp8基团。

[0114] 通常,本领域技术人员将知晓上述衍生物或其缀合物相比于具有SEQ ID NO:1的序列的蛋白,可具有增强的引发酶和/或聚合酶活性,例如增强至少10%,至少20%,至少30%,至少40%,至少50%,至少60%,至少70%,至少80%,至少90%,至少100%,至少150%,至少200%。

[0115] 与本发明一致,本发明任一方法中使用的蛋白,蛋白衍生物或缀合物,可因而包括可探测标记或标签使蛋白、蛋白衍生物或及缀合物易于分离,例如存在于具有SEQ ID NO:2所示序列的分子中的多聚组氨酸标签(His-tag)。

[0116] 本领域技术人员易于得到上述蛋白衍生物或蛋白缀合物-例如其方法包括化学合成、寄主细胞产生或两者的结合。

[0117] 一个分子生物学中长期存在的问题是RNA转录成DNA(逆转录)。这是cDNA图书馆的构建、基于PCR的分析等所必须的。到目前为止,逆转录通过许多合适的酶来实现,例如从逆转录病毒中原始分离出的MMLV RT,AMV RT,HIV-1RT。由于复杂的二级结构导致在这些酶在正常反应温度下(37或42℃)不能被融解,迄今为止RTs的一个难题是全细胞mRNA的非均匀转录。当试图逆转录包含发卡结构例如miRNA的mRNA时,尤其困难。

[0118] 一些RT酶的更耐热型已经被介绍:MMLV-RT突变异种(Gerard,Fox等,1997)(Maxima Reverse Transcriptase,突变异种MMLV RT([www.thermoscientificbio.com/reverse-transcription-rtqpcr-rtqpcr/maxima-reverse-transcriptase/by](http://www.thermoscientificbio.com/reverse-transcription-rtqpcr-rtqpcr/maxima-reverse-transcriptase/by) Thermo Scientific),RocketScript™RT,突变异种MMLV RT在70℃下工作([us.bioneer.com/products/accupower/accupower-rocketscript-overview.aspx](http://us.bioneer.com/products/accupower/accupower-rocketscript-overview.aspx) by Bioneer),PyroPhage RT([www.lucigen.com](http://www.lucigen.com)) (Schoenfeld,Patterson等,2008;Schoenfeld,Liles等,2009; Moser,DiFrancesco等,2012),Tth逆转录酶/DNA聚合酶(Thermus thermophilus(Tth) Pol I在锰离子的存在下也能转录RNA,但是是极不准确的)(Myers和Gelfand,1991)。

[0119] TthPrimPol具有高热稳定性和高度持续合成能力,对逆转录复杂RNA的这个难题提供了出色的解决方案。

- [0120] 因此,在本发明一个优选地实施方式中,TthPrimPol将在合适的缓冲液中用作RT。
- [0121] 如上所述,TthPrimPol将在PCR反应中同时作为RT和DNA聚合酶。其优势是RNA能够直接作为模板,避免两个独立反应带来的额外劳动和花费,也能避免两个独立反应可能带来的偏差和污染。为专家所知的是,典型温度曲线可用于上述组合反应(例如,94℃初始变性,70℃下进行RT反应30min,用于PCR反应的在94,50和70℃之间的额外循环)。
- [0122] 在本文中,术语核苷酸或核苷酸类似物不限于特定形式,这意味着包括脱氧核糖三磷酸盐(deoxynucleoside triphosphates,dNTPs)例如,包括但不限于dATP,dCTP,dGTP,dTTP,dITP,dUTP或其衍生物。作为非限定性的例子,所述衍生物可参考双脱氧核苷酸例如ddATP,ddCTP,ddGTP,ddTTP,ddITP,ddUTP,或被氧化的衍生物例如8oxoA,8oxoG,50HC,50HU,或标记的衍生物例如荧光标记氧生物或更复杂的标记例如用于第三代测序的标记。
- [0123] 在本文中,术语困难模板、受损DNA、受损核酸或污染的DNA可相互互换,指代改变了核苷酸碱基的核酸,以至于难于通过常规扩增或测序方法来处理。作为非限制性的例子,可参考生理学上发生的改变例如由氧化应力引起的改变例如8oxoA和8oxoG,或例如甲醛等交联剂引起的改变。
- [0124] 由于TthPrimPol能够高度接受不同的改性DNA,本发明第一方面进一步的实施方式包括TthPrimPol用于例如受损DNA测序或扩增。
- [0125] 测序用FFPE样品是本领域一大需求,如许多临床样品以这样的形式储存。交联剂如甲醛或多聚甲醛能够引起DNA变性,难以采用常规手段处理。因此,在一个具体的实施方式中,核酸测序或扩增的方法包括来自FFPE样品的DNA或RNA模板。
- [0126] 在一个更具体的实施方式中,核酸测序或扩增的方法包括氧化的核苷酸碱基例如8oxoA,8oxoG,50HC,50HU。
- [0127] 同样的,由于其高耐受性,TthPrimPol在模板组合物中能够被用于法医DNA物质或考古样品中DNA的扩增和测序。
- [0128] 在其它实施方式中,TthPrimPol能够用于翻译生理中出现的改性DNA(DNA-modification)。改性修饰物的例子包括8oxoA,8oxoG,50HC,50HU。以8oxoG为例进行表征,TthPrimPol能够通过插入核苷酸“读取”修饰物,该核苷酸一定概率能够结合至合成链中。在8oxoG条件下,在受损位点的对面,dA的插入频率将比dC至少多五倍。根据改性修饰物的不同选取特定分布特征或插入偏好。当进行深度测序时,由于相同序列的多种不同的读数可以获知(例如30-50个读数),在位置X上的原始DNA改性修饰物能够从该位置上读取的核苷酸的分数的分布推断出来。同样的,对于Thymidin二聚物,DNA改性修饰物能够根据原始序列中删除的碱基来确定。
- [0129] 在老龄化领域,这项应用是具备优势的,通过确定DNA改性修饰物的性质和区域监控药效。同样的,上述应用在肿瘤学领域中也是有益的,用于检测特定疗法的效果和癌症细胞分级。
- [0130] 基础基因组研究和临床诊断应用均转向越来越少的样品分析再到单细胞的分析。同样,法医分析依赖的遗传物质通常也非常少。在所有这些情况下,从样品开始处理到用于下游目的(例如测序或DNA阵列杂交)的遗传信息扩增的任意过程中,均无核酸产生是非常重要的。
- [0131] 一个难以避免的污染源往往是核酸物质自带的酶制剂或试剂(例如缓冲液,核苷



酸等)。

[0132] 由于Phi29聚合酶的高保真性和高持续合成能力,使用phi29聚合酶和随机六聚物引物的多重置换扩增(MDA)已成为单细胞WGA的首选方法(Macaulay IC, Voet T, 2014)。然而,其它利用随机引物的较不受欢迎的方法也是存在的,例如以PEP-PCR或DA-PCR为基础的应用,如MALBAC(Zhang L等,1992)。

[0133] 例如,Woybe等发现细菌污染起源于常规的MDA试剂。Blainey和colleagues(2011)采用新的灵敏的数字MDA(dMDA)检测方法也发现DNA在三种商业可用的Phi29聚合酶制剂中存在。Yokouchi及同事也发现Phi29聚合酶被E.coli DNA污染(Yokouchi H等,2006)。

[0134] 在使用Taq聚合酶及其变体的PCR型反应中存在表面DNA污染已被文献广泛报道,例如Hein等(2007);Corless等(2000)和Champlot S等(2010)。

[0135] 本发明实施例15描述了采用PrimPol代替随机引物从而抑制污染物扩增。该实施例描述了含Phi9 DNA聚合酶、其它聚合酶(例如Taq聚合酶)的结合物也能够在此应用。

[0136] 扩增极少量DNA用于下游分析的需求不断增长。这适用于法医或病理学样本,同样适用于单细胞或少数几个细胞。一项非常感兴趣的应用是,对上述已扩增物质进行测序,从而获取健康组织或疾病组织细胞之间的遗传差异。药学上,这在肿瘤学领域是极为有价值的,药物的选择(例如,激酶抑制剂)是基于所确定的突变进行。因此,扩增物质反映出与样品的原始序列特征具有最大保真度是极为重要的。

[0137] 然而,目前用于从少量样品中扩增遗传物质的方法是不完善的。代表性的原始核酸物质是存在偏差的,这已被大量科技出版物所系统评估过。

[0138] 例如,Pinard及其同事(Pinard等,2006)研究了包括多重置换扩增(MDA)在内的不同的扩增方法。基因组的覆盖范围与未扩增的样品之间,存在显著偏差。GC含量也与未扩增样品之间存在偏差。

[0139] Paez等(2004)也报道了局部代表性不足及扩增核酸物质时的丢失问题。Raghunathan等(2005)发现从单E.coli细胞中通过MDA进行qPCR DNA扩增的方法存在偏差。此外,人及酵母菌染色体邻近末端区域的序列丢失也已被报道(Lage等(2003))。Hosono等(2003)描述了通过qPCR判断MDA型扩增导致基因组位点的过度代表或代表性不足。

[0140] Han等(2012)发现通过MDA扩增的人DNA中存在表观拷贝数变异(CNVs)的变化。他们发现那些持续进行被控制的不完全扩增的区域也有较高的GC含量。Pugh等(2008)也证实显著表观改变的CNVs与邻近染色体末端的GC含量有关。

[0141] 这些由MDA过程引入的人工制品带来了一个显著的问题,特别是用于诊断和治疗决策的临床用途时。很明显,由于所使用随机引物类型的改变会修正被引入的偏差类型,这种偏差至少一部分是由包含Phi29 DNA聚合酶的MDA过程所必需的随机引物导致(Alsmadi等,2009年)。因此,一种替代的引入较少或至少不同偏差型的扩增工艺是非常必要的。

[0142] 实施例16提供了这样一种替代方案,在MDA型扩增中TthPrimPol替代了随机引物。

[0143] 因此,在一个优选实施方案中,本发明涉及一种根据本发明的方法,其中受污染DNA的扩增得到抑制。

[0144] 实施例

[0145] 下面,本发明将通过以下内容作为参考进一步描述本发明,但并不限于实施例。

[0146] 实施例1:特定材料和方法

[0147] TthPrimPol的克隆:极端嗜热菌HB27基因组 (DDBJ/EMBL/GeneBank AE017221.1; GI:46197919)的序列分析显示出,ORF TTC0656,编码了属于古细菌/eukarya1引发酶(AEP)超家族的蛋白。使用此序列信息,我们合成了两种引物SEQ ID 11和17)用于TthPrimPol扩增,其中,TthPrimPol通过PCR从极端嗜热菌基因组DNA中提取。该通过PCR扩增的基因片段,使用延伸的高保真聚合酶(Roche)通过TA克隆连接到pGEM T-easy载体(Promega)上,经测序证实。使用NdeI和EcoRI位点,该带有靶基因的片段连接到pET21和pET28载体(Novagen)中。pET28载体允许作为融合物的重组蛋白的表达,融合物包括含有六组氨酸(hexahistidyl)序列的多官能前导肽用于在Ni<sup>2+</sup>亲和树脂柱中纯化。

[0148] TthPrimPol的过度生产和纯化:TthPrimPol的表达在大肠杆菌菌株BL21-CodonPlus (DE3)-RIL (Stratagene)中进行,伴随额外的argU,ileY和leuW tRNA基因的复制。TthPrimPol的诱导表达通过加入1mM IPTG至30℃下1.5升数期大肠杆菌细胞在LB中生长至Abs600nm为0.5。诱导后,细胞在30℃下孵育5小时。随后,收集培养好的细胞,将沉淀的细胞称重,并冷冻(-20℃)。

[0149] TthPrimPol-His标签纯化:刚纯化前,在4℃下进行,冷冻细胞(3,7gr)解冻并重悬于20ml补充有1M NaCl的缓冲液A(50mM Tris-HCl,pH 7.5,5%甘油,0.5mM EDTA,1mM DTT),然后在冰浴中通过超声破碎。细胞碎片以3000rpm离心5min后丢弃。不溶性物质的混合物经11000rpm离心20min成粒状。DNA在0.4%聚乙烯亚胺(10%在水中的储液,pH 7.5)中沉淀,经11000rpm离心20min得到沉淀物。用缓冲液A将上清液稀释至氯化钠的终浓度为0.25M,沉淀硫酸铵至30%饱和度,得到无聚乙烯亚胺的蛋白沉淀。该沉淀重悬于补充有0.25%吐温-20的缓冲液A中,并装入HiTrap Heparin HP柱(5ml,GE Healthcare)中,该HP柱已通过补充有0.2M NaCl和0.25%吐温-20的缓冲液A预先平衡。用补充有0.2M NaCl和0.25%吐温-20的缓冲液A彻底洗脱后,蛋白质用0.2-0.8M NaCl线性梯度洗脱。含TthPrimPol的洗脱液用缓冲液A稀释至NaCl终浓度为0.2M,并装入MONOS4.6/100PE柱(1.7ml,GE Healthcare)中,该PE柱已通过补充有0.2M NaCl和0.25%吐温-20的缓冲液A预先平衡。洗涤该柱,蛋白以0.1-1M NaCl线性梯度洗脱。含TthPrimPol馏分被收集,稀释至0.2M NaCl,装入HiTrap Heparin HP柱(5ml,GE Healthcare)中,以相同的缓冲液预先平衡。洗涤该柱,并用补充有1M NaCl和0.25%吐温-20的缓冲液洗脱蛋白。该馏分含有高纯度(>99%)TthPrimPol。蛋白质浓度通过考马斯蓝染色的10%SDS-聚丙烯酰胺凝胶光密度估算,使用已知浓度的标准。终馏分,调整到50%(v/v)甘油,保存在-80℃下。

[0150] TthPrimPol+His标签纯化:刚纯化前,在4℃下进行,冷冻细胞(3.5g)解冻并重悬于20ml补充有1M NaCl的缓冲液A(50mM Tris-HCl,pH 7.5,5%甘油,0.5mM EDTA,1mM DTT)中,然后通过冰上超声破碎。细胞碎片以40000g离心50min后丢弃。上清液置于HisTrap粗FF柱中(5ml,GE Healthcare),该FF柱已通过补充有1M NaCl和0.25%吐温-20及30mM咪唑的缓冲液A预先平衡。用补充有1M NaCl和0.25%吐温-20及30mM咪唑的缓冲液A彻底洗脱后,蛋白质用30-250mM咪唑线性梯度洗脱。含TthPrimPol的洗脱液用补充有0.25%吐温-20的缓冲液A稀释至NaCl终浓度为0.1M,并装入HiTrap Heparin HP柱(5ml,GE Healthcare)中,该PE柱已通过补充有0,1M NaCl和0.25%吐温-20的缓冲液A预先平衡。洗涤该柱,蛋白以补充有1M NaCl和0.25%吐温-20的缓冲液A洗脱。该馏分含有高纯度(>99%)TthPrimPol。蛋白质浓度通过考马斯蓝染色的10%SDS-聚丙烯酰胺凝胶光密度估算,使用已知浓度的标

准。终馏分,调整到50% (v/v) 甘油,保存在-80℃下。

[0151] DNA底物:通过PAGE (获自Sigma) 获得纯化的合成寡核苷酸。为评估TthPrimPol的聚合酶活性,模板/引物分子通过PI引物(5'CTGCAGCTGATGCGCC;SEQ ID NO:6)退火至T1模板(5'GTACCCGGGATCCGTACGGCGCATCAGCTGCAG;SEQ ID NO:5)而生成,或P2引物(5'GTACCCGGGATCCGTAC;SEQ ID NO:14)至T2模板(5'CTGCAGCTGATGCGCXGTACGGATCCCCGGGTAC;SEQ ID NO:13),其中X为A,C,G,T,脱碱基位点(AP),7,8-二氢-8-氧桥腺嘌呤(8oxoA),7,8-二氢-8-氧桥鸟嘌呤(8oxoG),胸腺嘧啶乙二醇(Tg),5-羟基胞嘧啶(5OH)或5-羟基尿嘧啶(5OHU),或P3引物(5'GATC AC AGTGAGT AC;SEQ ID NO:8)至T3模板(T3DNA 5'AGAAGTGTATCTTGTACTCACTGTGATC;SEQ ID NO:10 or T3RNA 5'AGAAGUGUAUCUUGUACUCACUGUGAUC;SEQ ID NO:12)。通过退火P3引物(SEQ ID NO:8)至T4模板(5'ACTGGCCGTCGTTCTATTGTACTCACTGTGATC;SEQ ID NO:7)或至下游寡核苷酸DG5P(5'-磷酸盐基团的5'AACGACGGCCAGT;SEQ ID NO:9)产生带缺口的5'-核苷酸分子。荧光基团(Cy5)标记在引物的5'-端。在50mM Tris-HCl,pH 7.5和0.3M NaCl的存在下,在80℃下加热10min,缓慢冷却至室温过夜,各引物杂交到模板或模板和下游寡核苷酸上以产生不同的DNA分子。为测定引发酶的活性,我们使用XTCC寡核苷酸(5'T15CCTXT10,其中X为A,C,G或T;SEQ ID NO:4),包含一个推定的疱疹病毒引发起始位点(Cavanaugh和Kuchta,2009)。

[0152] 引发酶测定:M13mpl8的ssDNA(20ng/μl)或XTCC寡核苷酸(1μM)用作模板,以测定引发酶的活性。反应混合物(20μl)包括:50m Tris-HCl pH 7.5,75mM NaCl,5mM的MgCl<sub>2</sub>或1mM MnCl<sub>2</sub>,1mM DTT,2.5%甘油,0.1mg/ml BSA,[ $\alpha$ -<sup>32</sup>P]dATP(16nM;3000Ci/mmol)或[ $\gamma$ -<sup>32</sup>P]ATP(16nM;3000Ci/mmol),每种指定量的dNTP或NTP,在TthPrimPol(400nm)存在下。在55℃下反应60min后,反应通过加入装载甲酰胺的缓冲液(10mM EDTA,95%v/v甲酰胺,0.3%w/v二甲苯-苯胺)停止。向反应中加入8M尿素含量20%聚丙烯酰胺测序凝胶。电泳后,从头合成的多核苷酸(引物)通过放射自显影法检测。

[0153] DNA和RNA聚合酶测定:孵育混合物包括:20μl,50mM Tris-HCl pH 7.5,5mM MgCl<sub>2</sub>or 1mM MnCl<sub>2</sub>,1mM DTT,5%甘油,0.1mg/ml BSA,每种条件下显示出5nM杂交DNA,dNTP或NTP为指定浓度,及TthPrimPol为指定量。反应混合物在40℃下孵育指定时间,通过添加10μl终止液(10mM EDTA和97.5%的去离子甲酰胺)终止反应。标记引物链的延伸通过8M尿素和20%PAGE进行分析,使用Typhoon 9410扫描器(GE Healthcare)进行可视化检测。

[0154] 环状DNA扩增反应:M13mpl8ssDNA和pET28dsDNA作为模板以测定TthPrimPol的扩增能力。孵育混合物包括,20μl,50mM Tris-HCl pH 7.5,1mM DTT,5%甘油,0.1mg/ml BSA,指定量的模板DNA,指定浓度的MgCl<sub>2</sub>或MnCl<sub>2</sub>,指定浓度的dNTP或NTPs和指定量的TthPrimPol。反应混合物在不同温度(35~85℃)下孵育指定时间,通过冰冻降温停止反应。扩增产物通过非变性凝胶电泳分析。

[0155] 随机引物合成用于Phi29 DNApol为基础的滚环扩增(RCA):孵育混合物包括:12,5μl,40mM Tris-HCl pH7.5,50mM KCl,45mM(NH<sub>4</sub>)<sub>2</sub>SO<sub>4</sub>,10mM MgCl<sub>2</sub>,0.025%Tween-20,500μM dNTPs,50μM随机六聚体(random hexamers),1ng pRSET,40ng Phi29 DNApol和指定量的TthPrimPol。加入酶之前,反应混合物煮沸(3min,95℃)以变性模板DNA。然后,反应混合物在30℃下孵育5h。扩增产物用HindIII消解,通过非变性凝胶电泳进行分析。

[0156] 随机引物合成用于Phi29 DNApol为基础的全基因组扩增(WGA):孵育混合物包括,

50 $\mu$ l, 40mM Tris-HCl pH 7.5, 50mM KCl, 45mM (NH<sub>4</sub>)<sub>2</sub>SO<sub>4</sub>, 10mM MgCl<sub>2</sub>, 0.025%吐温-20, 500 $\mu$ M dNTPs, 50 $\mu$ M随机六聚体, 1ng人基因组DNA, 650ng Phi29 DNApol, 以及指定量的 TthPrimPol。反应混合物在30 $^{\circ}$ C下孵育16h。扩增产物使用PicoGreen试剂(Quant-iT<sup>TM</sup>PicoGreen dsDNA试剂, Invitrogen)定量。

[0157] 酵母增殖及基因组DNA的分离: 酵母菌株BY4741取自Euroscarf (Institute of Molecular Biosciences, J.-W.Goethe-University Frankfurt), 30 $^{\circ}$ C下在YPD(10g酵母提取物, 20g蛋白胨, 20g葡萄糖每升) 介质中增殖。根据Qiagen公司Tips 100/G (QIAGEN Genomic DNA Handbook 04/2012) 所示的方法, 2份独立过夜培养物(OD 600=1;  $\sim$ 1.5 $\times$ 10<sup>9</sup>细胞) 各50毫升用于基因组DNA的分离。

[0158] 分离的DNA在1%琼脂糖凝胶上分析并测量OD值: OD<sub>260/280</sub>和OD<sub>260/230</sub>。最后测试DNA消化率: 100ng的ygDNA使用1U EcoRI在37 $^{\circ}$ C下消解1h。

[0159] 分离浓度为200ng/ $\mu$ l的两份酵母DNA样品。#1样品用于下游实验。

[0160] 酵母DNA的MiSeq测序: 起始于2 $\mu$ g任意非扩增的酵母DNA(样品NA), 常规扩增的酵母DNA(样品RP), 及通过增加量TthPrimPol(分别为100, 200和400ng) 扩增的酵母DNA, 与Phi29聚合酶(sample names Tth-100/Phi, Tth-200/Phi和Tth-400/Phi) 结合, 贮库制剂, 五组样品的多路复用和测序由已授权的服务提供商GATC(GATC Biotech AG, Germany) 采用IsSeq测序仪(Illumina, USA) 完成Illumina测序。测序参数如下: 配对末端读取, 读取长300bp。GATC返回10份FASTQ文件, 每个样品和阅读方向各一份。

[0161] 实施例2:

[0162] TthPrimPol是一种活性引发酶能够与dNTPs启动核酸合成

[0163] 经提纯的TthPrimPol显示出在单链寡核苷酸上的DNA引发酶活性, 其中, 具有潜在引发酶识别序列(GTCC) 侧接在胸腺嘧啶上(Cavanaugh和Kuchta 2009)。该嘧啶序列显示其为用于启动通过几种病毒、原核和真核RNA引发酶的引发反应的优选的模板范围(Holmes, Cheriathundam等, 1985; Parker和Cheng, 1987; Frick和Richardson, 2001)。如图1所示, 引发仅发生在“TC”序列之前, 而在聚dT示踪的对面不发生引发。在锰存在下, 作为引物(5'端)的核苷酸可以是核糖核苷酸(ATP) 或脱氧核苷酸(dATP), 但当镁作为金属辅因子时, 仅可为脱氧核苷酸(dATP)。无论是否金属辅因子, 添加的核苷酸(3'端) 必须为脱氧核苷酸(dGTP)。对核苷酸序列需求分析的结果表明起始位点之前的核苷酸显示出强烈影响。如图2所示, 当C是起始位点之前的核苷酸时, 具有最高的引发酶活性。因此, TthPrimPol优选CTC作为起始位点。此外, 如图3所示, 在单链环状DNA模板(M13mpl8ssDNA) 上TthPrimPol显示出DNA引发酶的活性。有趣的是, 蛋白的合成严格依赖于脱氧核苷酸, 而在核糖核苷酸的存在下几乎没有产品被合成。

[0164] 通常, 引发酶作为RNA引物, 但古细菌和一些细菌中AEP相关的引物除外(Sanchez-Berrondo, Mesa等, 2011; Lao-Sirieix, Pellegrini等, 2005), 使用dNTPs作为引发反应有效的底物。这种例外的RNA/DNA引发酶活性被证实为TthPrimPol所固有。

[0165] 实施例3:

[0166] TthPrimPol同样是一种有效的聚合酶

[0167] 不像真核引发酶, 古菌质(archaea) 引发酶能够分别实现引发使RNA和DNA链长从1KB至7KB的扩增(Chemnitz Galal, Pan等, 2012; Lao-Sirieix和Bell 2004; Lao-Sirieix,

Pellegrini等,2005),这些酶同时为引发酶和聚合酶或“PrimPols”(Lipps,Rother等,2003)。因此,我们首先测试,除了它的DNA/RNA的引发酶的活性,TthPrimPol是否具有DNA依赖性的DNA聚合酶活性。如图4所示,TthPrimPol能够通过聚合作用彻底延伸一个模板/引物分子至18个脱氧核苷酸。镁和锰是进行DNA聚合有效的金属辅因子,虽然锰似乎是更有效的。如图5所示,TthPrimPol也能够有缺口的dNA分子上进行DNA或RNA合成,而在核糖核酸聚合的情况下具有更低的效率。

[0168] 然而,RNA聚合酶活性似乎严格依赖于锰,对于有缺口的DNA,锰极强地促进链置换合成,允许使用可用的整个模板序列(+18)。

[0169] 实施例3:

[0170] TthPrimPol具有逆转录酶活性

[0171] 如图6所示,TthPrimPol能够使用RNA模板合成互补的DNA链,尽管反应的效率地犹使用DNA模板所达到的效果。然而你,RNA指导的DNA聚合酶活性似乎严格依赖于锰。

[0172] 实施例4:

[0173] TthPrimPol与dNTPs结合的保真性

[0174] 为分析TthPrimPol保真地催化DNA合成的能力,无论是在镁或锰离子存在下,4种dNTP被分别单独地测定作为一种底物与相对的四种可能的模板碱基的相结合。图7显示出在所有的情形中,优先地,由第一个可用的模板碱基指导TthPrimPol高保真的插入互补核苷酸。即使含有锰,当延伸至正确配对的引物对末端时,在由TthPrimPol提供强偏移DNA聚合期间进行碱基辨别从而形成正确的Watson和Crick碱基对。此外,测定了TthPrimPol对于dNTPs的亲合力。在单循环条件下,酶浓度高于DNA的浓度(图8),由于在低浓度下酶能够有效地插入互补的核苷酸链,dNTP的结合清楚的揭示了TthPrimPol对于dNTPs具有极高的亲合力。

[0175] 实施例5

[0176] TthPrimPol对于受损的DNA具有高耐受性

[0177] 为测定TthPrimPol对于受损DNA的耐受性,提供的模板蛋白包括:碱基位点(AP),7,8-二氢-8-氧桥腺嘌呤(8oxoA),7,8-二氢-8-氧桥鸟嘌呤(8oxoG),胸腺嘧啶乙二醇(Tg),5-羟基胞嘧啶(5OHc)或5-羟基尿嘧啶(5OHU)。如图9所示,TthPrimPol能够通过将核苷酸插入损伤点的对面或通过跳过损伤点并复制下一个模板基(dG的插入)从而进行高效的病变旁路。氧化态的鸟嘌呤(8oxoG),是在DNA氧化应激的后果中最常见形式的损害之一(Berquist和Wilson,2012),主要和高效地绕过引入的dATP的顺式异构体。如图10所示,dATP插入损伤点的对面与dATP插入未受损的T的动力学特征是一致的。

[0178] 由于TthPrimPol能够在模板-引物和带缺口的DNA分子中合成,它不能添加任何的核苷酸至同聚的ssDNA的3'末端(图11),这意味着TthPrimPol完全是一个模板依赖性酶。控制反应条件不包括dNTPs,也证实了TthPrimPol缺乏3-5核酸外切酶活性,由于DNA分子未退化,与之相同的,也缺乏ExoI,ExoII及ExoIII共有基序,能够在几种DNA聚合酶成员上形成一个进化上保守的3-5外切酶的活性位点(Bernad,Blanco等,1989)。

[0179] 实施例6:

[0180] TthPrimPol具有高效的环DNA扩增

[0181] 由于TthPrimPol结合了高效的DNA引发酶和DNA聚合酶活性,我们假设这种酶能够

熟练的扩增DNA。为测定TthPrimPol生物扩增能力,在TthPrimPol和dNTP或NTPs的存在下,未添加任何特定或随机引物,采用M13mp18ssDNA作为模板DNA进行检测。如图12所示,在dNTP底物的存在下,TthPrimPol (+/-His tag)的全部情形都能够大量扩增供给的DNA。另一方面,核糖核苷酸(NTPs)似乎是通过TthPrimPol的DNA扩增的不良底物。

[0182] 实施例7:

[0183] 辅因子

[0184] 如图13a-d所示,在扩增反应中,只有锰和镁能够作为金属付银子,而Co和Ca则不合作为辅因子,当锰的浓度为0.5-2.5nM及镁浓度为2.5-10nM时,扩增产量最大。

[0185] 接下来我们分析了为得到扩增产物所需的蛋白量。如图14a-b所示,采用镁或锰作为金属辅因子时,扩增反应的产率与反应中使用的TthPrimPol的量成比例。同样的,产物的量依赖于初始DNA模板的量(参见图15)。而反应时间,参见图16a-b,分别使用金属引物( $Mn^{2+}$ 或 $Mg^{2+}$ ),也许是由于缺乏dNTP,反应4h后达到饱和。图17显示了扩增产物的数量变化和迁移率变化作为dNTP浓度的函数。对于每种金属辅因子,最佳的dNTP浓度范围为 $100\mu M$ (使用锰时)至 $500\mu M$ (使用镁时)。过高的浓度会抑制反应。对于各金属来说,在有效区间的dNTP中,扩增产物( $Mg^{2+}$ 和 $Mn^{2+}$ )具有相似的迁移率及产率。如图18所示,对于每种金属辅因子,当使用镁时最佳温度区间为 $55-60^{\circ}C$ ,当使用锰时为 $52-62^{\circ}C$ 。

[0186] 实施例8:

[0187] TthPrimPol能够引发DNA用于随后的采用Phi29 DNA聚合酶的扩增反应

[0188] 作为反应起始的基础,TthPrimPol是一种有效和通用的引发酶,我们分析了TthPrimPol用于随机引物环dsDNA的能力,该引发的DNA能够有效用于采用Phi29 DNApol的多重置换扩增。之后,通过Phi29 DNApol生成的置换链将构成新的TthPrimPol的底物,其中酶可以随机合成新的引物,从而为Phi29 DNApol产生新的起始位点,实现呈指数的DNA扩增。如图19所示,这两种酶的组合能够熟练扩增质粒DNA。扩增产物量的增加依赖于TthPrimPol剂量,使得无需加入随机合成引物。为分析TthPrimPol有效引发基因组DNA的能力,发明人进行了一系列全基因扩增试验。发明人再次得到依赖于TthPrimPol剂量的扩增的DNA,当TthPrimPol浓度为400nM时,达到最大值。更高的量未进行测试。这些实施例清楚地表明无论是环状DNA的扩增还是线性DNA的扩增,TthPrimPol可代替合成六聚核苷酸(图20)。

[0189] 实施例9:

[0190] TthPrimPol是比HsPrimPol更有效的聚合酶

[0191] 尽管TthPrimPol具有热稳定型,该酶与已被很好地表征的(参见Garcia Gomez, 2012)HsPrimPol相比,仍具有其绝大多数的基本功能。我们因此希望知道在相同条件下,未额外添加随机引物或特定引物,这两种酶是否都能用于扩增DNA。因此,在无引物的情况下,在各自最优的反应温度下,相当量的TthPrimPol和HsPrimPol用于扩增M13mp18ssDNA。在NTPs的存在下,TthPrimPol和HsPrimPol产生相当量的产物。但是相比之下,在dNTPs的存在下,TthPrimPol相比于HsPrimPol酶具有更高的效率,导致一个明显更多量的扩增的DNA(图22)。

[0192] 实施例10:

[0193] TthPrimPol用作试剂盒化合物

[0194] 尤其对于常常被存储在实验室中较长时间的试剂盒的发展,化合物的稳定性始终是一个重要因素。为了解决这个问题,分离的TthPrimPol在-20°C下可以在存储缓冲液中保存数月。经常反复实验,如本文所述(参见例如图21),测试TthPrimPol的引发酶和聚合酶活性,测量的活性并无显著的降低。

[0195] 实施例11:

[0196] TthPrimPol作为引发酶与不同的聚合酶联合使用

[0197] 本发明的一种具体实施方式中,使用TthPrimPol作为引发酶与其他聚合酶一起用于DNA扩增。目前,全基因组扩增(WGA)优选使用Phi29 DNApol和随机寡核苷酸(MDA,多重置换扩增)(Spits,Le Caignec等,2006;Spits,Le Caignec等,2006;Silander和Saarela,2008;Alsmadi,Alkayal等,2009)。WGA是一种链置换扩增,优选用于线性基因组DNA。另一种链置换扩增是滚换扩增(RCA),用于共价封闭的环状DNA的扩增。Phi29 DNApol是用于RCA方法优选的酶。TthPrimPol能够同时代替寡核苷酸、WGA和RCA,扩增方法如图19至20所示。由于其遏制复制叉的重启动的负作用,TthPrimPol将有利于与常规PCR方法配合使用。

[0198] 实施例12:

[0199] DNA标签

[0200] 在一个具体实施方式中,TthPrimPol用于一种复制和扩增核酸的方法,包括:带标记的核苷酸结合至互补的DNA链。例如,荧光基团标记的核苷酸(图23)或放射性核苷酸可标记DNA分子用于在分析中测定。由于其对于核苷酸的高持续合成能力和高耐受性,TthPrimPol同样适用于结合。上述标签能够用于DNA分析的样品制备,例如Affymetrix分析。

[0201] 实施例13:

[0202] DNA突变

[0203] TthPrimPol能够用于突变DNA,例如通过缓慢氧化模板链,使TthPrimPol扩增亲代DNA。通过与不同核苷酸的结合(图7和8),可实现随机突变。

[0204] 实施例14:

[0205] DNA和RNA的扩增

[0206] 由于其高持续合成能力,在有/无引物情况下,TthPrimPol能够用于扩增DNA,相似的应用中,Phi29 DNA聚合酶被使用。然而,由于其作为一种RNA-依赖性的RNA聚合酶,TthPrimPol能够用于扩增RNA。当试图获得单个细胞或少数细胞的表达图谱时,这是尤为必要的,例如,取自CNS中的神经元或肿瘤研究中的细胞时。至今为止,RNA聚合方案依赖于许多步骤,例如(Rossner,Hirrlinger等,2006;Newrzella,Pahlavan等,2007)。TthPrimPol基于RNA的扩增将产生均匀的,无偏差的扩增产物。另一个好处是高温下扩增反应仍可以进行,能够抑制逆向RNases,分解二级结构。

[0207] 实施例15:

[0208] 非特异性背景DNA的抑制

[0209] 令人惊奇的是,发明人发现TthPrimPol和Phi29 DNA聚合酶,在任意孵育时间下进行测试(直至16小时),两种酶的结合在无额外添加模板DNA分子的情况下(图26a)不会产生DNA的扩增。在上述条件下,TthPrimPol引发酶活性似乎依赖于变性DNA的存在(单链DNA)。

[0210] 因此,在进一步的实施方式中,TthPrimPol和Phi29 DNA聚合酶的结合用于避免在

等温MDA扩增中受污染的DNA的扩增。由于酶制剂或试剂中带核酸物质,上述污染的DNA往往很难避免。

[0211] 实施例16

[0212] 通过TthPrimPol和Phi29polymerase的联合使用扩增DNA探针的全基因组序列

[0213] 使用TthPrimPol和Phi29-pol (Tth/Phi) 协同扩增允许极小量的DNA的扩增(实施例14),使其成为能通过有限的模板DNA生成全用途DNA探针的一种有效方法。由于常规MDA会引入一些偏差,主要是拷贝数偏差,依赖于靶DNA和随机引物的组成成分(Alsmadi 2009),我们测试了DNA的Tth/Phi扩增是否能够允许更少的扩增偏差。

[0214] 因此,为确定Tth/Phi扩增的DNA的性能,我们在Miseq(Illumina,USA) 测序仪上,对非扩增酵母(strain BY7471) DNA (NA)、扩增(随机引物) 酵母DNA (RP) 和通过Tth/Phi扩增的DNA(分别使用不同的TthPrimPol量:100ng,200ng和400ng) 进行测序。

[0215] 基本性能的评估采用FastX-toolkit(hannonlab.cshl.edu/fastx\_toolkit) 完成,但是在所有的测试样本中,并未显示出显著的不同,这推断出Tth/Phi扩增方法也适用于下一代测序(NGS) 探针的制备。

[0216] 为了识别源于序列库制剂的变化,相同量(140万) 随机选择,高品质(Phred values>20;Ewing B.et a.1998) 配对末端序列读数(280万单端读数) 被用于进一步分析。

[0217] 软件'CLC Genomic Workbench' Version 7.0 (CLC Bio;Denmark) 被用于进一步分析NGS数据。该软件时用于NGS测序分析的特征丰富的完整的软件解决方案,不仅能够映射所有读取内容到参考基因组,而且还(基于所述映射) 来确定过代表和代表性不足的基因组区域。采用上述映射分析所有样品,本发明人意外地发现,全部NA的90%以上(90.48%) 和Tth/Phi(91.92%和91.90%) 9读取内容映射到参考基因组,而只有83.92%的RP映射到参考基因组(图28)。Tth/Phi和RP之间的这种差别也能通过在染色体水平中过代表和代表性不足的区域显著差异( $p=0.00029$ ) 所证实。

[0218] 在扩增的DNA的所有情形下,过代表和代表性不足区域的总数是由测序和扩增方法引入的总偏差的量度。与此相反,对于NA样品,过代表和代表性不足区域的总数仅取决于测序方法。

[0219] 如果对比RP到NA和Tth/Phi到NA的相对差异,可以仅测量完全由扩增方法产生的偏差。Tth/Phi扩增引物的偏差比RP扩增多17%,Tth/Phi扩增相比RP扩增具有明显的优势(图29)。

[0220] 参考文献

[0221] Alsmadi,O.,F.Alkayal等(2009). "Specific and complete human genome amplification with improved yield achieved by phi29 DNA polymerase and a novel primer at elevated temperature." BMC Res Notes 2:48.

[0222] Altschul,S.F.,W.Gish等(1990). "Basic local alignment search tool." J Mol Biol 215(3):403-10.

[0223] Aravind,L.,D.D.Leipe等(1998). "Toprim~a conserved catalytic domain in type IA and II topoisomerases,DnaG-type primases,OLD family nucleases and RecR proteins." Nucleic Acids Res 26(18):4205-13.

[0224] Bentley,D.R.,S.Balasubramanian等(2008). "Accurate whole human genome



sequencing using reversible terminator chemistry."Nature 456(7218):53-9.

[0225] Bernad,A.,L.Blanco等(1989). "A conserved 3'—5' exonuclease active site in prokaryotic and eukaryotic DNA polymerases."Cell 59(1):219-28.

[0226] Berquist,B.R.and D.M.Wilson,3rd(2012) "Pathways for repairing and tolerating the spectrum of oxidative DNA lesions."Cancer Lett 327(1-2):61-72.

[0227] Blainey et.al(2011)Nucleic Acids Research Vol.39,No.4.

[0228] Branton,D.,D.W.Deamer等(2008). "The potential and challenges of nanopore sequencing."Nat Biotechnol 26(10):1146-53.

[0229] Butler,T.Z.,M.Pavlenok等(2008). "Single-molecule DNA detection with an engineered MspA protein nanopore."Proc Natl Acad Sci U S A 105(52):20647-52.

Cavanaugh,N.A.and R.D.Kuchta(2009). "Initiation of new DNA strands by the herpes simplex virus-1primase-helicase complex and either herpes DNA polymerase or human DNA polymerase alpha."J Biol Chem 284(3):1523-32.

[0230] Champ lot S,Berthelot C,Pruvost M,Bennett EA,Grange T等(2010) "An Efficient Multistrategy DNA Decontamination Procedure of PCR Reagents for Hypersensitive PCR Applications".PLoS ONE 5(9)

[0231] Chemnitz Galal,W.,M.Pan等(2012) "Characterization of DNA primase complex isolated from the archaeon,Thermococcus kodakaraensis."J Biol Chem 287(20):16209-19.

[0232] Clarke,J.,H.C.Wu等(2009). "Continuous base identification for single-molecule nanopore DNA sequencing."Nat Nanotechnol 4(4):265-70.

[0233] Corless等,JOURNAL OF CLINICAL MICROBIOLOGY,May 2000,p.1747-1752Derrington,I.M.,T.Z.Butler,et al.(2010) "Nanopore DNA sequencing with MspA."Proc Natl Acad Sci U S A 107(37):16060-5.

[0234] Eid,J.,A.Fehr等(2009). "Real-time DNA sequencing from single polymerase molecules."Science 323(5910):133-8.

[0235] Ewing B,Hillier L,Wendl MC,Green P(1998). "Base-calling of automated sequencer traces using phred.I.Accuracy assessment".Genome Res.8(3):175-185

[0236] Ewing B,Hillier L,Wendl MC,Green P(1998). "Base-calling of automated sequencer traces using phred.II.Accuracy assessment".Genome Res.8(3):186-194

[0237] Frick,D.N.和C.C.Richardson(2001). "DNA primases."Annu Rev Biochem 70:39-80.

[0238] García Gómez,S.I.,M.I.Martínez Jiménez等(2012).PrimPol a two in one enzyme:primase and bypass polymerase.From Single Molecules to System Biology.

[0239] Sevilla,IUBMB FEBS.

[0240] Gerard,G.F.,D.K.Fox等(1997). "Reverse transcriptase.The use of cloned Moloney murine leukemia virus reverse transcriptase to synthesize DNA from RNA."Mol Biotechnol 8(1):61-77.

[0241] Guo,J.,N.Xu等(2008). "Four-color DNA sequencing with 3a€2-0-modified

nucleotide reversible terminators and chemically cleavable fluorescent dideoxynucleotides."Proceedings of the National Academy of Sciences 105(27): 9145-9150.

[0242] Han等(2012)BMC Genomics,13:217)

[0243] Hein等,Journal of Microbiological Methods 71(2007)336-339

[0244] Hinz,H.J.,C.Steif等(1993)."Fundamentals of protein stability."Pure& Appl.Chem.65(5):5.

[0245] Holmes,A.M.,E.Cheriathundam等(1985)."Initiation of DNA synthesis by the calf thymus DNA polymerase-primase complex."J Biol Chem 260(19):10840-6.

[0246] Hosono et al.(2003)Genome Res 13:954-964)

[0247] Ilyina,T.V.and E.V.Koonin(1992)."Conserved sequence motifs in the initiator proteins for rolling circle DNA replication encoded by diverse replicons from eubacteria,eucaryotes and archaeobacteria."Nucleic Acids Res 20(13):3279-85.

[0248] Iyer,L.M.,E.V.Koonin等(2005)."Origin and evolution of the archaeo-eukaryotic primase superfamily and related palm-domain proteins:structural insights and new members."Nucleic Acids Res 33(12):3875-96.

[0249] Jetha,N.N.,M.Wiggin等(2009)."Forming an alpha-hemolysin nanopore for single-molecule analysis."Methods Mol Biol 544:113-27.

[0250] Ju,J.,D.H.Kim等(2006)."Four-color DNA sequencing by synthesis using cleavable fluorescent nucleotide reversible terminators."Proceedings of the National Academy of Sciences 103(52):19635-19640.

[0251] Kim,Y.H.,A.H.Berry等(2001)."Comparing the effect on protein stability of methionine oxidation versus mutagenesis:steps toward engineering oxidative resistance in proteins."Protein Eng 14(5):343-7.

[0252] Kornberg,A.and T.Baker(1991).DNA Replication,2nd edn,W.H.Freeman& Company,NY.

[0253] Lage et al,(2003)Genome Res13:294-307.

[0254] Lao-Sirieix,S.H.和S.D.Bell(2004)."The heterodimeric primase of the hyperthermophilic archaeon Sulfolobus solfataricus possesses DNA and RNA primase,polymerase and 3'-terminal nucleotidyl transferase activities."J Mol Biol344(5):1251-63.

[0255] Lao-Sirieix,S.H.,L.Pellegrini等(2005)."The promiscuous primase."Trends Genet 21(10):568-72.

[0256] Lieberman,K.R.,G.M.Cherf等(2010)"Processive replication of single DNA molecules in a nanopore catalyzed by phi29 DNA polymerase."J Am Chem Soc 132(50):17961-72.

[0257] Lipps,G.(2004)."The replication protein of the Sulfolobus islandicus plasmid pRN1."Biochem Soc Trans 32(Pt2):240-4.

- [0258] Lipps,G.,S.Rother等(2003).“A novel type of replicative enzyme harbouring ATPase,primase and DNA polymerase activity.”EMBO J 22(10):2516-25.
- [0259] Lipps,G.,A.O.Weinzierl等(2004).“Structure of a bifunctional DNA primase-polymerase.”Nat Struct Mol Biol11(2):157-62.
- [0260] Litosh,V.A.,W.Wu等(2011).“Improved nucleotide selectivity and termination of 3’-OH unblocked reversible terminators by molecular tuning of 2-nitrobenzyl alkylated HOMedU triphosphates.”Nucleic Acids Res 39(6):e39.
- [0261] Lodish,H.B.,A.;Zipurski,S.L.,Matsudaira,P.;Baltimore,D.;Darnell,E.J.(1999).Molecular Cell Biology,W.H.Freemann&Co.,NY.
- [0262] Macaulay IC,Voet T(2014)“Single Cell Genomics:Advances and Future Perspectives”.PLoS Genet 10(1).
- [0263] Maglia,G.,M.R.Restrepo等(2008).“Enhanced translocation of single DNA molecules through alpha-hemolysin nanopores by manipulation of internal charge.”Proc Natl Acad Sci U S A 105(50):19720-5.
- [0264] Manrao,E.A.,I.M.Derrington等(2012)“Reading DNA at single-nucleotide resolution with a mutant MspA nanopore and phi29 DNA polymerase.”Nat Biotechnol 30(4):349-53.
- [0265] Manrao,E.A.,I.M.Derrington等(2011)“Nucleotide discrimination with DNA 1 mMobilized in the MspA nanopore.”PLoS One 6(10):e25723.
- [0266] Maxam,A.M.and W.Gilbert(1977).“A new method for sequencing DNA.”Proc Natl Acad Sci U S A 74(2):560-4.
- [0267] Metzker,M.L.(2009).“Sequencing technologies-the next generation.”Nat Rev Genet 11(1):31-46.
- [0268] Moser,M.J.,R.A.DiFrancesco等(2012).“Thermostable DNA polymerase from a viral metagenome is a potent RT-PCR enzyme.”PLoS One 7(6):e38371.
- [0269] Myers,T.W.和D.H.Gelfand(1991).“Reverse transcription and DNA amplification by a Thermus thermophilus DNA polymerase.”Biochemistry 30(31):7661-6.
- [0270] Newrzella,D.,P.S.Pahlavan等(2007).“The functional genome of CA1 and CA3 neurons under native conditions and in response to ischemia.”BMC Genomics 8:370.
- [0271] Noirot-Gros,M.F.和S.D.Ehrlich(1996).“Change of a catalytic reaction carried out by a DNA replication protein.”Science 274(5288):777-80.
- [0272] Paez等(2004)Nucleic Acids Res 32:e71.
- [0273] Parker,W.B.和Y.C.Cheng(1987).“Inhibition of DNA primase by nucleoside triphosphates and their arabinofuranosyl analogs.”Mol Pharmacol 31(2):146-51.
- [0274] Pavlenok,M.,I.M.Derrington等(2012)“MspA nanopores from subunit dimers.”PLoS One 7(6):e38726.
- [0275] Pinard等(2006)BMC Genomics 7:216.

- [0276] Pugh等(2008)Nucleic Acids Research Vol.36,No.13.
- [0277] Raghunathan et al.(2005)Appl Environ Microbiol 71:3342-3347.
- [0278] Ronaghi,M.,S.Karamohamed等(1996).“Real-time DNA sequencing using detection of pyrophosphate release.”Anal Biochem 242(1):84-9.
- [0279] Ronaghi,M.,M.Uhlen等(1998).“A sequencing method based on real-time pyrophosphate.”Science 281(5375):363,365.
- [0280] Rossner,M.J.,J.Hirrlinger等(2006).“Global transcriptome analysis of genetically identified neurons in the adult cortex.”J Neurosci 26(39):9956-66.
- [0281] Salas,M.(1991).“Protein-priming of DNA replication.”Annu Rev Biochem 60:39-71.Sam,L.T.,D.Lipson等(2011)“A comparison of single molecule and amplification based sequencing of cancer transcriptomes.”PLoS One 6(3):e17305.
- [0282] Sanchez-Berrondo,J.,P.Mesa等(2011)“Molecular architecture of a multifunctional MCM complex.”Nucleic Acids Res 40(3):1366-80.
- [0283] Sanger,F.和A.R.Coulson(1978).“The use of thin acrylamide gels for DNA sequencing.”FEBS Lett 87(1):107-10.
- [0284] Sanger,F.,S.Nicklen等(1977).“DNA sequencing with chain-terminating inhibitors.”Proc Natl Acad Sci U S A 74(12):5463-7.
- [0285] Schoenfeld,T.,M.Liles等(2009).“Functional viral metagenomics and the next generation of molecular tools.”Trends Microbiol 18(1):20-9.
- [0286] Schoenfeld,T.,M.Patterson等(2008).“Assembly of viral metagenomes from yellowstone hot springs.”Appl Environ Microbiol 74(13):4164-74.
- [0287] Silander,K.and J.Saarela(2008).“Whole genome amplification with Phi29 DNA polymerase to enable genetic or genomic analysis of samples of low DNA yield.”Methods Mol Biol 439:1-18.
- [0288] Simpson,R.J.(2005).“Stabilization of proteins for storage.”Cold Spring Harb Protoc 2010(5):pdb top79.
- [0289] Spits,C,C.Le Caignec等(2006).“Optimization and evaluation of single-cell whole-genome multiple displacement amplification.”Hum Mutat 27(5):496-503.
- [0290] Spits,C,C.Le Caignec等(2006).“Whole-genome multiple displacement amplification from single cells.”Nat Protoc 1(4):1965-70.
- [0291] Thompson,J.F.and P.M.Milos(2011)“The properties and applications of single-molecule DNA sequencing.”Genome Biol 12(2):217.
- [0292] Turcatti,G.,A.Romieu等(2008).“A new class of cleavable fluorescent nucleotides:synthesis and optimization as reversible terminators for DNA sequencing by synthesis.”Nucleic Acids Research 36(4):e25.
- [0293] Wendell,D.,P.Jing等(2009).“Translocation of double-stranded DNA

through membrane-adapted phi29 motor protein nanopores." Nat Nanotechnol 4(11): 765-72.

[0294] Woyke T, Sczyrba A, Lee J, Rinke C, Tighe D 等 (2011) "Decontamination of MDA Reagents for Single Cell Whole Genome Amplification". PLoS ONE 6(10)

[0295] Wu, W., B.P. Stupi 等 (2007). "Termination of DNA synthesis by N6-alkylated, not 3'-O-alkylated, photocleavable 2'-deoxyadenosine triphosphates." Nucleic Acids Res 35(19): 6339-49.

[0296] Zhang L, Cui X, Schmitt K, Hubert R, Navidi W 等 (1992), "Whole genome amplification from a single cell: implications for genetic analysis". Proc Natl Acad Sci U S A 89: 5847-5851

[0297] 本发明进一步包括如下1-23所述的方法:

[0298] 1. 一种用于核酸复制、扩增或测序的方法, 包括以下步骤:

[0299] a) 提供

[0300] i. 聚合酶, 或

[0301] ii. 聚合酶缀合物,

[0302] 其中, 聚合酶或聚合酶缀合物的聚合酶部分包含与序列SEQ ID NO:1至少70%同源的序列并进一步包括聚合酶和引发酶活性, 和

[0303] b) 提供模板核酸, 和

[0304] c) 提供用于结合到核酸互补链的核苷酸和/或核苷酸衍生物, 和

[0305] d) 提供合适的缓冲液, 和

[0306] e) 任选地, 提供一种或多种引物, 和

[0307] f) 使步骤a-e中的材料接触适当的时间, 任选地在高温下, 优选地, 温度超过40°C, 更优选地, 温度超过50°C。

[0308] 2. 如条目1所述的方法, 其中步骤a) i) 中聚合酶或步骤a) ii) 中聚合酶缀合物中的聚合酶部分包含与序列SEQ ID NO:1至少80%同源的序列。

[0309] 3. 如条目1所述的方法, 其中步骤a) i) 中聚合酶或步骤a) ii) 中聚合酶缀合物中的聚合酶部分包含与序列SEQ ID NO:1至少90%同源的序列。

[0310] 4. 如条目1所述的方法, 其中步骤a) i) 中聚合酶或步骤a) ii) 中聚合酶缀合物中的聚合酶部分具有如SEQ ID NO:1所示的序列。

[0311] 5. 如条目1-4任一项所述的方法, 其中步骤b) 中核酸为受损的核酸。

[0312] 6. 如条目5所述的方法, 其中受损的核酸如下定义:

[0313] a) 包括单链或双链断裂, 和/或

[0314] b) 包括化学改性核苷酸, 优选自氧化的核苷酸和/或脱氨基的核苷酸, 最优选的, 选自以下组: 7,8-二氢-8-氧桥腺嘌呤 (7,8-dihydro-8-oxoadenine), 7,8-二氢-8-氧桥鸟嘌呤 (7,8-dihydro-8-oxoguanine), 胸腺嘧啶乙二醇 (thymine glycol), 5-羟基胞嘧啶 (5-hydroxy cytosine) 或 5-羟基尿嘧啶 (5-hydroxy cytosine), 和/或

[0315] c) 来源于甲醛或para-聚甲醛嵌入的组织。

[0316] 7. 如条目1-6任一项所述的方法, 其中, 步骤c) 中至少一种核苷酸衍生物为化学改性的核苷酸。

- [0317] 8. 如条目7所述的方法,其中,步骤c)中的核苷酸衍生物包括至少一种选自以下组的核苷酸衍生物:
- [0318] 氧化的核苷酸衍生物,和/或
- [0319] 脱氨基的核苷酸,
- [0320] 包括至少一种空间位阻侧基的衍生物。
- [0321] 9. 如条目1-8任一项所述的方法,其中,步骤b)中模板核酸来自少于100个细胞,优选来自单细胞。
- [0322] 10. 如条目1-9任一项所述的方法,其中步骤e)中无引物。
- [0323] 11. 如条目1-10任一项所述的方法,其中步骤b)中核酸为脱氧核糖核酸(DNA)。
- [0324] 12. 如条目1-10任一项所述的方法,其中步骤b)中核酸为核糖核酸(RNA)。
- [0325] 13. 如条目12所述的方法,包括RNA的逆转录。
- [0326] 14. 如条目1-12任一项所述的方法,包括第二聚合酶。
- [0327] 15. 如条目14所述的方法,其中所述第二聚合酶选自以下组:原核A家族聚合酶,原核B家族聚合酶。
- [0328] 16. 如条目14所述的方法,其中所述第二聚合酶选自以下组:Taq聚合酶,Tth聚合酶,Pfu聚合酶,Vent聚合酶。
- [0329] 17. 如条目14所述的方法,其中所述第二聚合酶为Phi29型DNA聚合酶或其缀合物。
- [0330] 18. 如条目1-4任一项所定义的一种聚合酶或聚合酶缀合物在条目1-17任一项所述方法中的应用。
- [0331] 19. 一种用于检测改性的方法,尤其是用于检测氧化过程中的产物,在DNA模板中,包括以下步骤:
- [0332] a) 根据条目1-4任一项得到多条序列读数,和
- [0333] b) 计算结合至每个核苷酸位点的核苷酸在步骤a)整个序列读数中的相对比例,和
- [0334] c) 将结合的核苷酸的总比例与模板中DNA改性的原始种类相关联。
- [0335] 20. 一种用于随机突变的方法,包括以下步骤:
- [0336] a) 将核酸与氧化化合物相接触,其中,所述接触导致核酸碱基的氧化,和
- [0337] b) 通过如条目1-4任一项所定义的聚合酶或聚合酶缀合物复制被氧化的核酸。
- [0338] 21. 一种用于扩增RNA分子的试剂盒,在合适的条件下耐储性超过20个月,包括:
- [0339] b) 第一容器包括如条目1-4任一项所定义的聚合酶;
- [0340] b) 第二容器包括一种或多种脱氧核苷三磷酸。
- [0341] 22. 一种用于扩增DNA分子的试剂盒,在合适的条件下耐储性超过20个月,包括:
- [0342] a) 第一容器包括如条目1-4任一项所定义的聚合酶;
- [0343] b) 第二容器包括一种或多种核糖核苷三磷酸。
- [0344] 23. 一种用于扩增DNA分子的试剂盒,在合适的条件下耐储性超过20个月,包括:
- [0345] a) 第一容器包括如条目1-4任一项所定义的聚合酶;
- [0346] b) 第二容器包括一种或多种脱氧核糖核苷三磷酸;
- [0347] c) 第三容器包括一种或多种脱氧核苷三磷酸。

## SEQUENCE LISTING

<110> Sygnis Biotech S.L.U.

<120> 使用热稳定的 TthPrimPol 扩增和测序的方法

<130> BL68504PC

<160> 17

<170> BiSSAP 1.2

<210> 1

<211> 293

<212> PRT

<213> *Thermus thermophilus*

<400> 1

Met Arg Pro Ile Glu His Ala Leu Ser Tyr Ala Ala Gln Gly Tyr Gly

1                    5                    10                    15

Val Leu Pro Leu Arg Pro Gly Gly Lys Glu Pro Leu Gly Lys Leu Val

                  20                    25                    30

Pro His Gly Leu Lys Asn Ala Ser Arg Asp Pro Ala Thr Leu Glu Ala

                  35                    40                    45

Trp Trp Arg Ser Cys Pro Arg Cys Gly Val Gly Ile Leu Pro Gly Pro

                  50                    55                    60

[0001]

Glu Val Leu Val Leu Asp Phe Asp Asp Pro Glu Ala Trp Glu Gly Leu

65                    70                    75                    80

Arg Gln Glu His Pro Ala Leu Glu Ala Ala Pro Arg Gln Arg Thr Pro

                  85                    90                    95

Lys Gly Gly Arg His Val Phe Leu Arg Leu Pro Glu Gly Val Arg Leu

                  100                    105                    110

Ser Ala Ser Val Arg Ala Ile Pro Gly Val Asp Leu Arg Gly Met Gly

                  115                    120                    125

Arg Ala Tyr Val Val Ala Ala Pro Thr Arg Leu Lys Asp Gly Arg Thr

                  130                    135                    140

Tyr Thr Trp Glu Ala Pro Leu Thr Pro Pro Glu Glu Leu Pro Pro Val

145                    150                    155                    160

Pro Gln Ala Leu Leu Leu Lys Leu Leu Pro Pro Pro Pro Pro Arg

                  165                    170                    175

Pro Ser Trp Gly Ala Val Gly Thr Ala Ser Pro Lys Arg Leu Gln Ala

                  180                    185                    190

Leu Leu Gln Ala Tyr Ala Ala Gln Val Ala Arg Thr Pro Glu Gly Gln

                  195                    200                    205

Arg His Leu Thr Leu Ile Arg Tyr Ala Val Ala Ala Gly Gly Leu Ile

                  210                    215                    220

Pro His Gly Leu Asp Pro Arg Glu Ala Glu Glu Val Leu Val Ala Ala

225                    230                    235                    240  
 Ala Met Ser Ala Gly Leu Pro Glu Trp Glu Ala Arg Asp Ala Val Arg  
                           245                    250                    255  
 Trp Gly Leu Gly Val Gly Ala Ser Arg Pro Leu Val Leu Glu Ser Ser  
                           260                    265                    270  
 Ser Lys Pro Pro Glu Pro Arg Thr Tyr Arg Ala Arg Val Tyr Ala Arg  
                           275                    280                    285  
 Met Arg Arg Trp Val  
                           290  
  
 <210> 2  
 <211> 313  
 <212> PRT  
 <213> Thermus thermophilus  
  
 <220>  
 <223> PrimPol + His Tag  
  
 <400> 2  
 Met Gly Ser Ser His His His His His Ser Ser Gly Leu Val Pro  
 1                    5                    10                    15  
 [0002] Arg Gly Ser His Met Arg Pro Ile Glu His Ala Leu Ser Tyr Ala Ala  
                           20                    25                    30  
 Gln Gly Tyr Gly Val Leu Pro Leu Arg Pro Gly Gly Lys Glu Pro Leu  
                           35                    40                    45  
 Gly Lys Leu Val Pro His Gly Leu Lys Asn Ala Ser Arg Asp Pro Ala  
                           50                    55                    60  
 Thr Leu Glu Ala Trp Trp Arg Ser Cys Pro Arg Cys Gly Val Gly Ile  
 65                    70                    75                    80  
 Leu Pro Gly Pro Glu Val Leu Val Leu Asp Phe Asp Asp Pro Glu Ala  
                           85                    90                    95  
 Trp Glu Gly Leu Arg Gln Glu His Pro Ala Leu Glu Ala Ala Pro Arg  
                           100                    105                    110  
 Gln Arg Thr Pro Lys Gly Gly Arg His Val Phe Leu Arg Leu Pro Glu  
                           115                    120                    125  
 Gly Val Arg Leu Ser Ala Ser Val Arg Ala Ile Pro Gly Val Asp Leu  
                           130                    135                    140  
 Arg Gly Met Gly Arg Ala Tyr Val Val Ala Ala Pro Thr Arg Leu Lys  
 145                    150                    155                    160  
 Asp Gly Arg Thr Tyr Thr Trp Glu Ala Pro Leu Thr Pro Pro Glu Glu  
                           165                    170                    175  
 Leu Pro Pro Val Pro Gln Ala Leu Leu Leu Lys Leu Leu Pro Pro Pro  
                           180                    185                    190



Pro Pro Pro Arg Pro Ser Trp Gly Ala Val Gly Thr Ala Ser Pro Lys  
 195 200 205  
 Arg Leu Gln Ala Leu Leu Gln Ala Tyr Ala Ala Gln Val Ala Arg Thr  
 210 215 220  
 Pro Glu Gly Gln Arg His Leu Thr Leu Ile Arg Tyr Ala Val Ala Ala  
 225 230 235 240  
 Gly Gly Leu Ile Pro His Gly Leu Asp Pro Arg Glu Ala Glu Glu Val  
 245 250 255  
 Leu Val Ala Ala Ala Met Ser Ala Gly Leu Pro Glu Trp Glu Ala Arg  
 260 265 270  
 Asp Ala Val Arg Trp Gly Leu Gly Val Gly Ala Ser Arg Pro Leu Val  
 275 280 285  
 Leu Glu Ser Ser Ser Lys Pro Pro Glu Pro Arg Thr Tyr Arg Ala Arg  
 290 295 300  
 Val Tyr Ala Arg Met Arg Arg Trp Val  
 305 310

<210> 3

<211> 29

<212> DNA

<213> Artificial Sequence

[0003]

<220>

<221> source

<222> 1..29

<223> /mol\_type="unassigned DNA"

/note="template DNA"

/organism="Artificial Sequence"

<400> 3

tttttttt ttttctgt tttttttt

29

<210> 4

<211> 29

<212> DNA

<213> Artificial Sequence

<220>

<221> source

<222> 1..29

<223> /mol\_type="unassigned DNA"

/note="template DNA"

/organism="Artificial Sequence"

<400> 4

tttttttt ttttcctnt tttttttt

29

<210> 5

<211> 34

<212> DNA

<213> Artificial Sequence

<220>

<221> source

<222> 1..34

<223> /mol\_type="unassigned DNA"

/note="template DNA"

/organism="Artificial Sequence"

<400> 5

gtaccggggg atccgtacgg cgcatcagct gcag

34

[0004]

<210> 6

<211> 16

<212> DNA

<213> Artificial Sequence

<220>

<221> source

<222> 1..16

<223> /mol\_type="unassigned DNA"

/note="primer"

/organism="Artificial Sequence"

<400> 6

ctgcagctga tgcgcc

16

<210> 7

<211> 33

<212> DNA

<213> Artificial Sequence

	<220>	
	<221> source	
	<222> 1..33	
	<223> /mol_type="unassigned DNA"	
	/note="template DNA"	
	/organism="Artificial Sequence"	
	<400> 7	
	actggccgtc gttctattgt actcactgtg atc	33
	<210> 8	
	<211> 15	
	<212> DNA	
	<213> Artificial Sequence	
	<220>	
	<221> source	
	<222> 1..15	
[0005]	<223> /mol_type="unassigned DNA"	
	/note="primer"	
	/organism="Artificial Sequence"	
	<400> 8	
	gatcacagtg agtac	15
	<210> 9	
	<211> 13	
	<212> DNA	
	<213> Artificial Sequence	
	<220>	
	<221> source	
	<222> 1..13	
	<223> /mol_type="unassigned DNA"	
	/note="primer"	
	/organism="Artificial Sequence"	
	<400> 9	
	aacgacggcc agt	13

<210> 10  
 <211> 28  
 <212> DNA  
 <213> Artificial Sequence  
  
 <220>  
 <221> source  
 <222> 1..28  
 <223> /mol\_type="unassigned DNA"  
       /note="template DNA"  
       /organism="Artificial Sequence"  
  
 <400> 10  
 agaagtgtat cttgtactca ctgtgatc 28  
  
 <210> 11  
 <211> 31  
 <212> DNA  
 <213> Artificial Sequence  
  
 [0006] <220>  
       <221> source  
       <222> 1..31  
       <223> /mol\_type="unassigned DNA"  
           /note="primer"  
           /organism="Artificial Sequence"  
  
 <400> 11  
 ccggcccata tgaggccgat tgagcacgcc c 31  
  
 <210> 12  
 <211> 28  
 <212> RNA  
 <213> Artificial Sequence  
  
 <220>  
 <221> source  
 <222> 1..28  
 <223> /mol\_type="unassigned RNA"  
       /note="template RNA"  
       /organism="Artificial Sequence"

	<400> 12	
	agaaguguau cuuguacuca cugugauc	28
	<210> 13	
	<211> 34	
	<212> DNA	
	<213> Artificial Sequence	
	<220>	
	<221> source	
	<222> 1..34	
	<223> /mol_type="unassigned DNA"	
	/note="template DNA"	
	/organism="Artificial Sequence"	
	<400> 13	
	ctgcagctga tgcgcngtac ggatccccgg gtac	34
[0007]	<210> 14	
	<211> 18	
	<212> DNA	
	<213> Artificial Sequence	
	<220>	
	<221> source	
	<222> 1..18	
	<223> /mol_type="unassigned DNA"	
	/note="primer"	
	/organism="Artificial Sequence"	
	<400> 14	
	gtaccggggg atccgtac	18
	<210> 15	
	<211> 14	
	<212> DNA	
	<213> Artificial Sequence	
	<220>	

	<221> source	
	<222> 1..14	
	<223> /mol_type="unassigned DNA"	
	/note="poly-t"	
	/organism="Artificial Sequence"	
	<400> 15	
	ttttttttt tttt	14
	<210> 16	
	<211> 14	
	<212> DNA	
	<213> Artificial Sequence	
	<220>	
	<221> source	
	<222> 1..14	
	<223> /mol_type="unassigned DNA"	
	/note="poly-a"	
[0008]	/organism="Artificial Sequence"	
	<400> 16	
	aaaaaaaaaa aaaa	14
	<210> 17	
	<211> 35	
	<212> DNA	
	<213> Artificial Sequence	
	<220>	
	<221> source	
	<222> 1..35	
	<223> /mol_type="unassigned DNA"	
	/note="primer"	
	/organism="Artificial Sequence"	
	<400> 17	
	gcgcgcgaat tctcatacc acctctcat ccggg	35

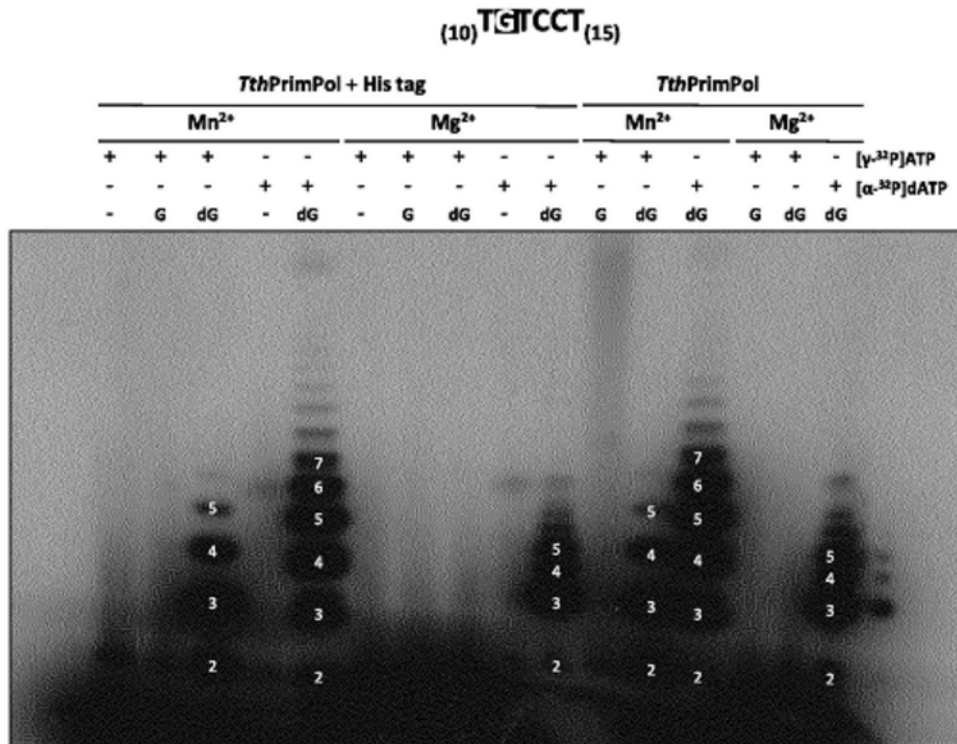


图1

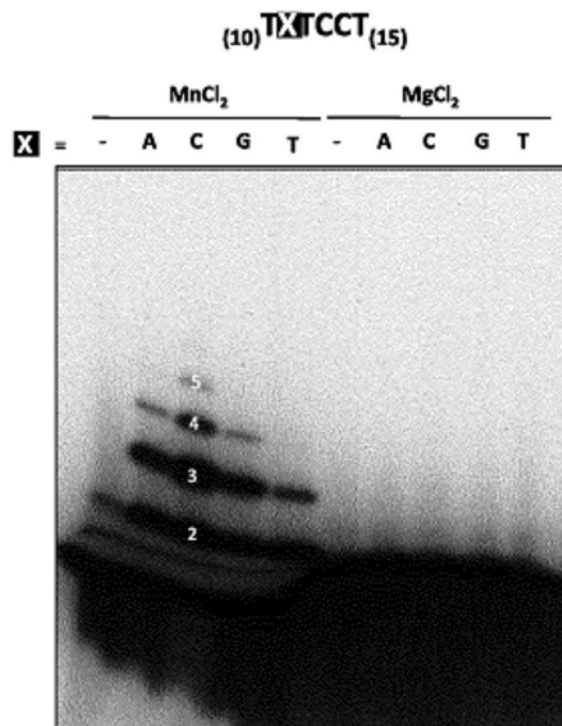


图2

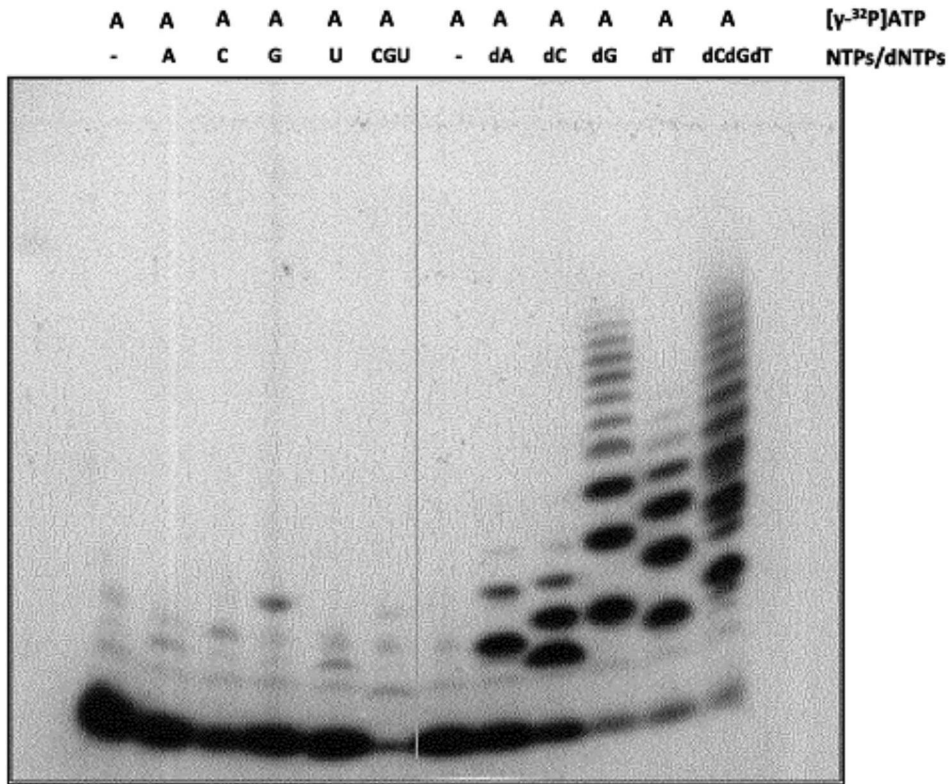


图3

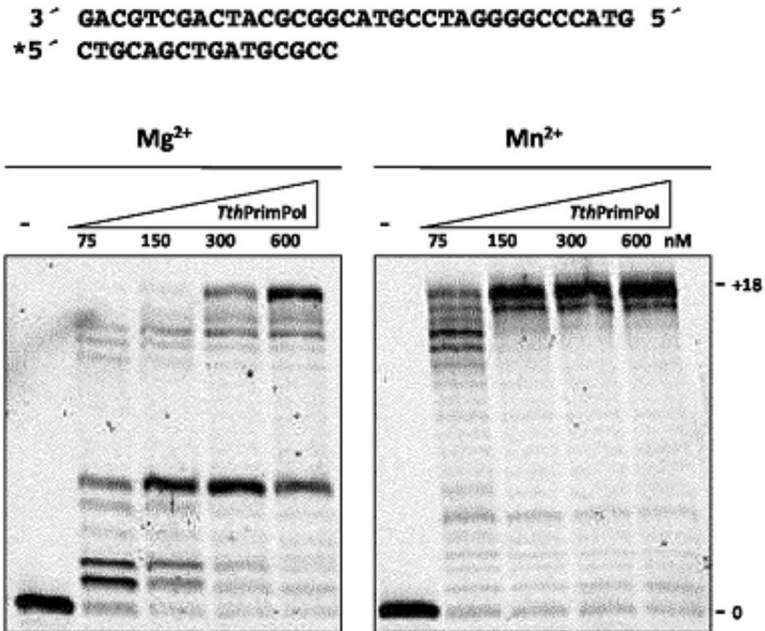


图4



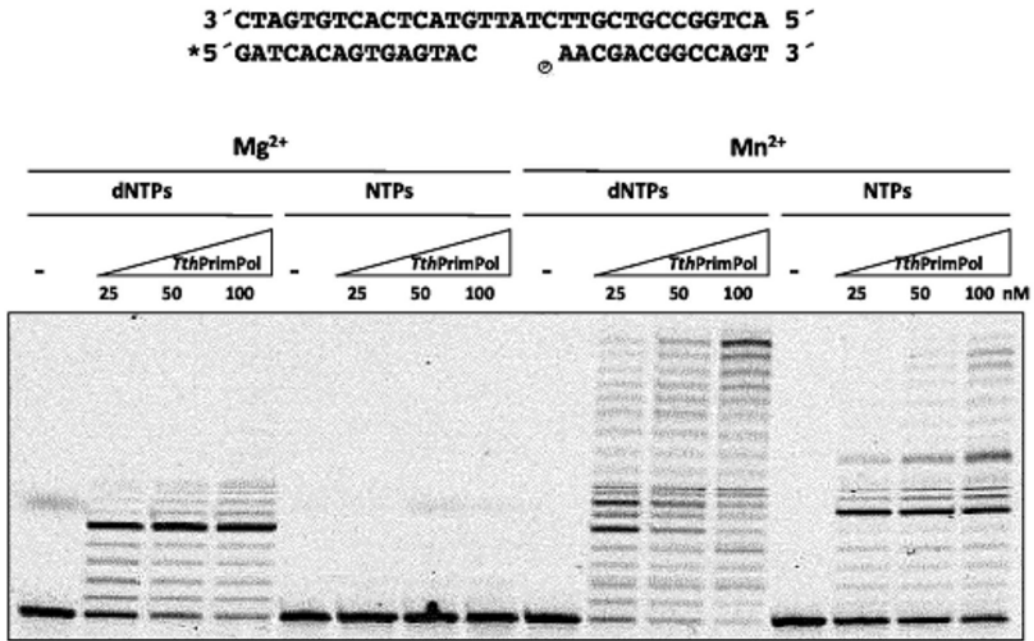


图5

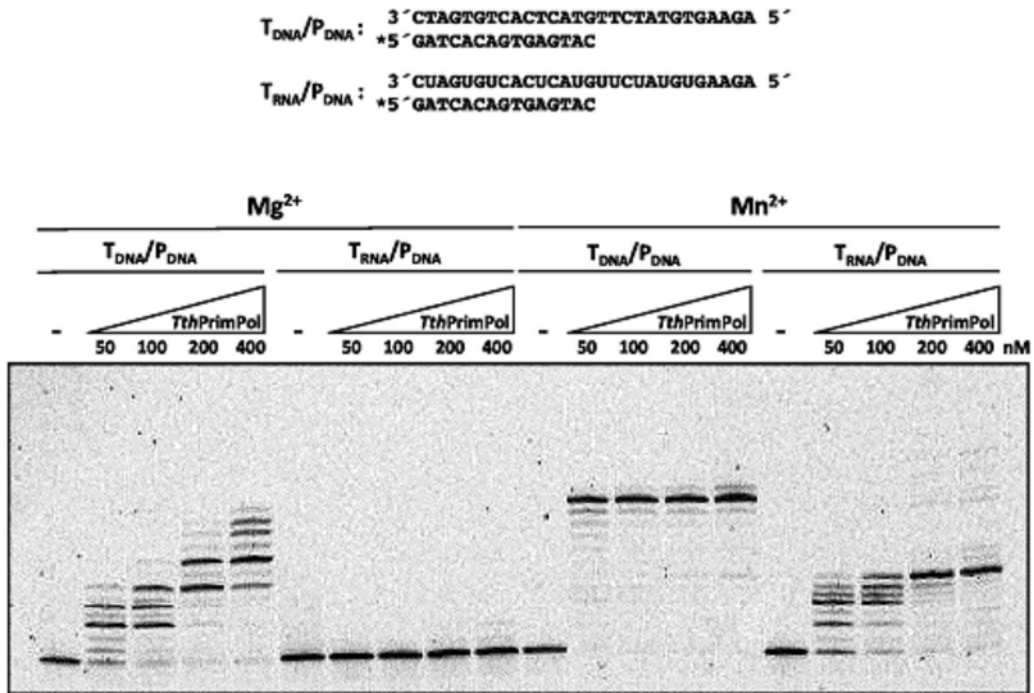


图6

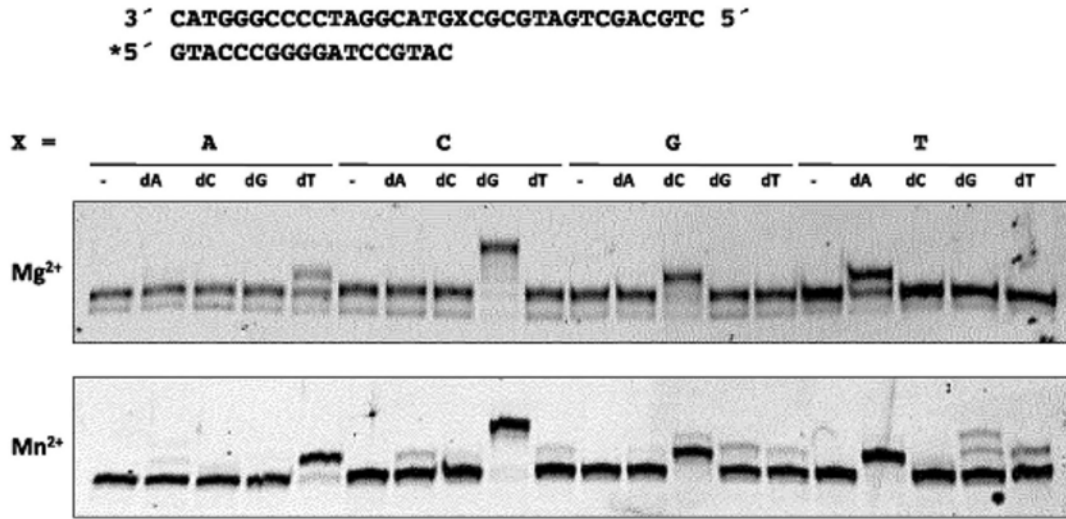


图7

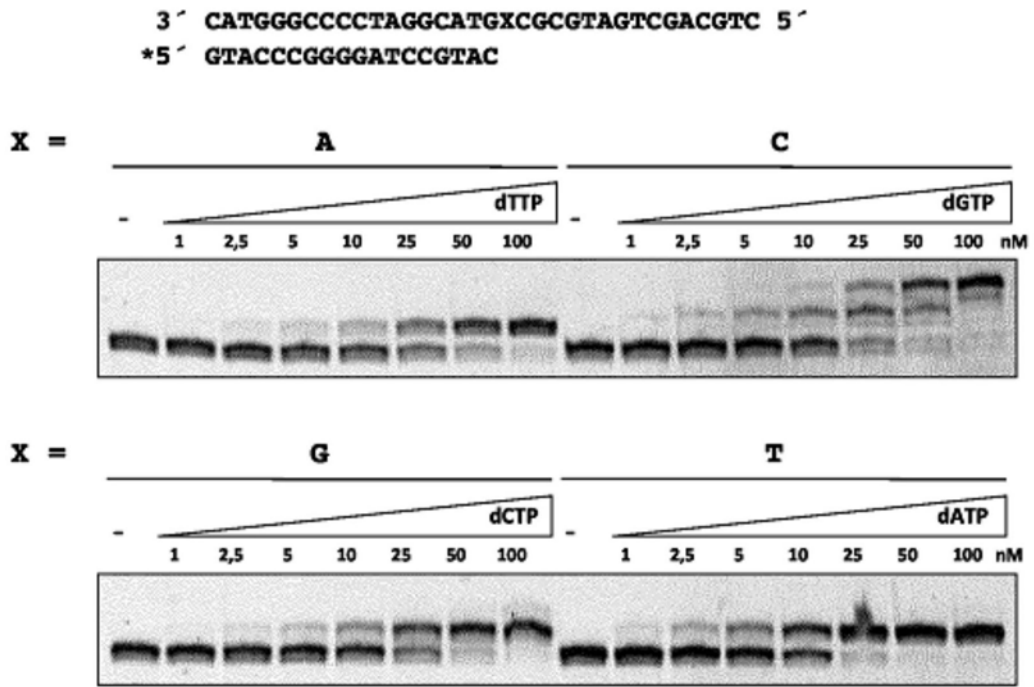


图8

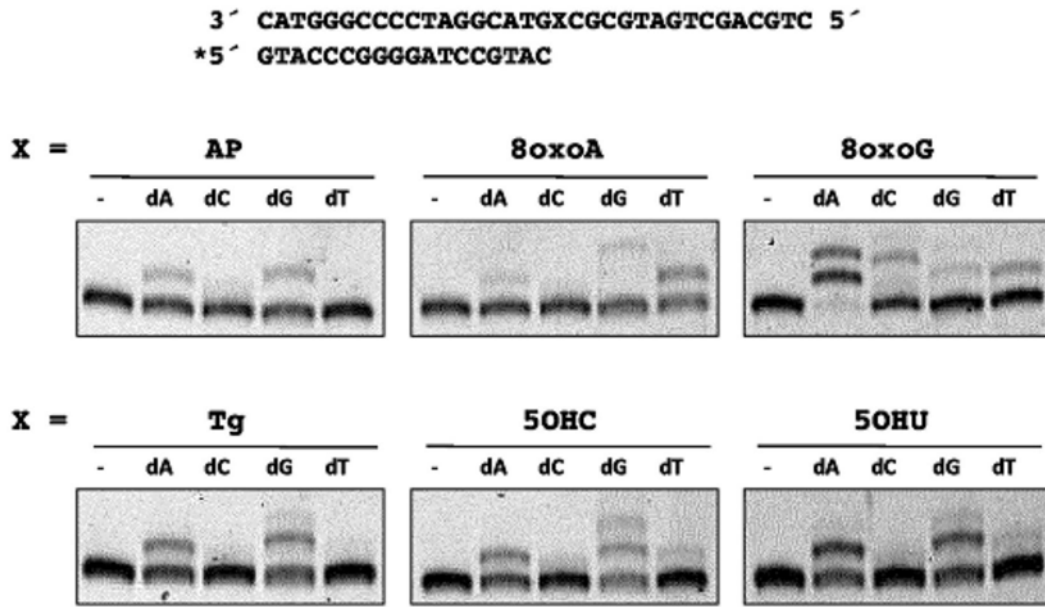


图9

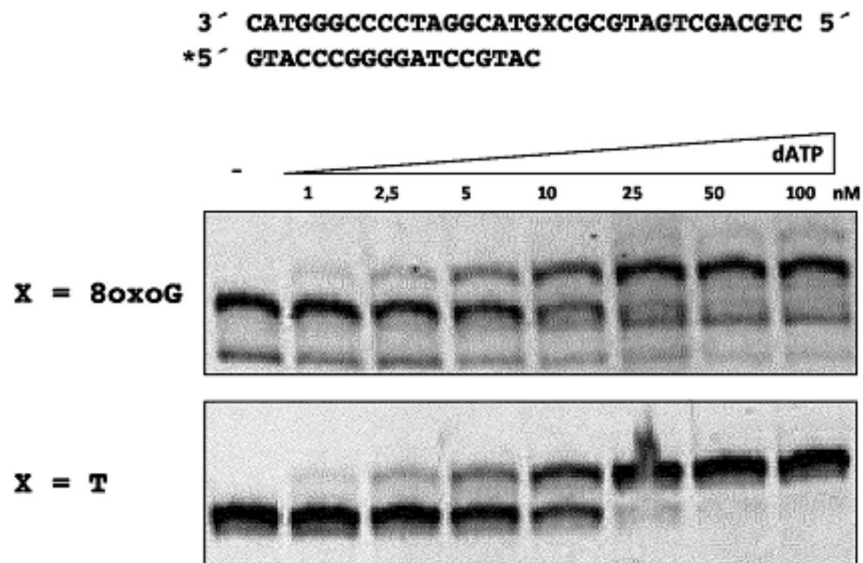


图10

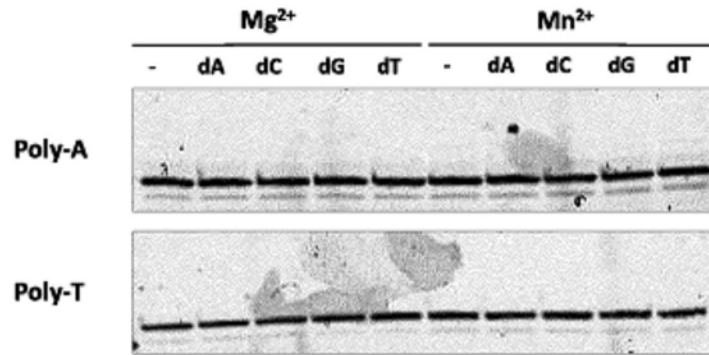


图11

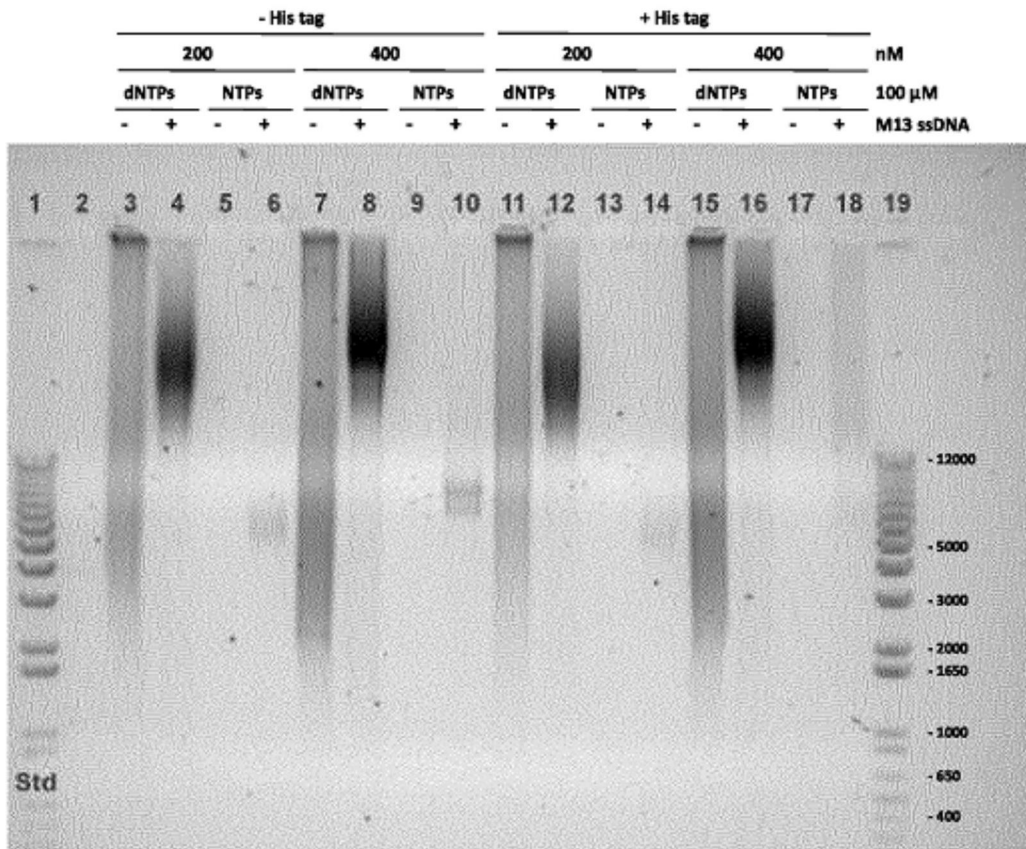


图12

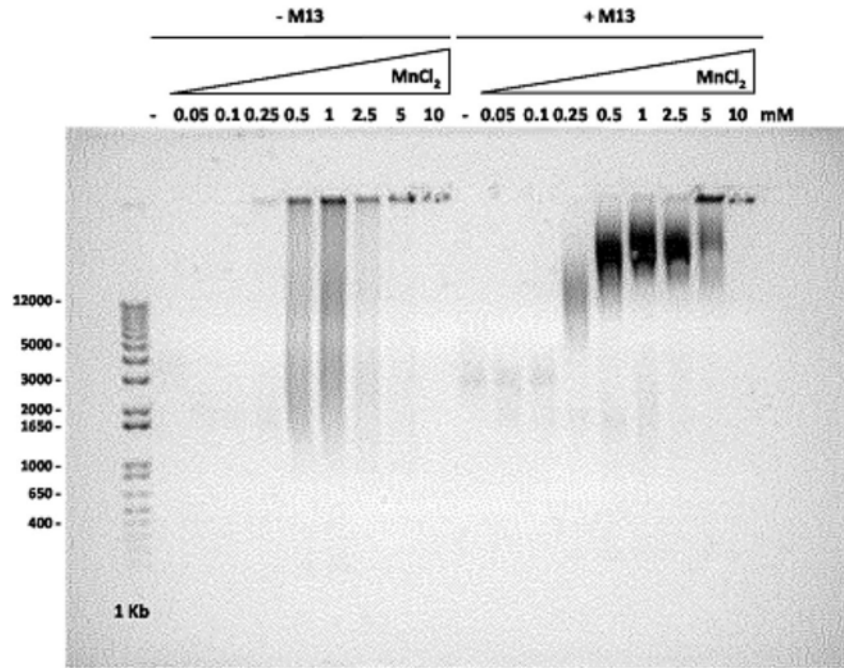


图13a

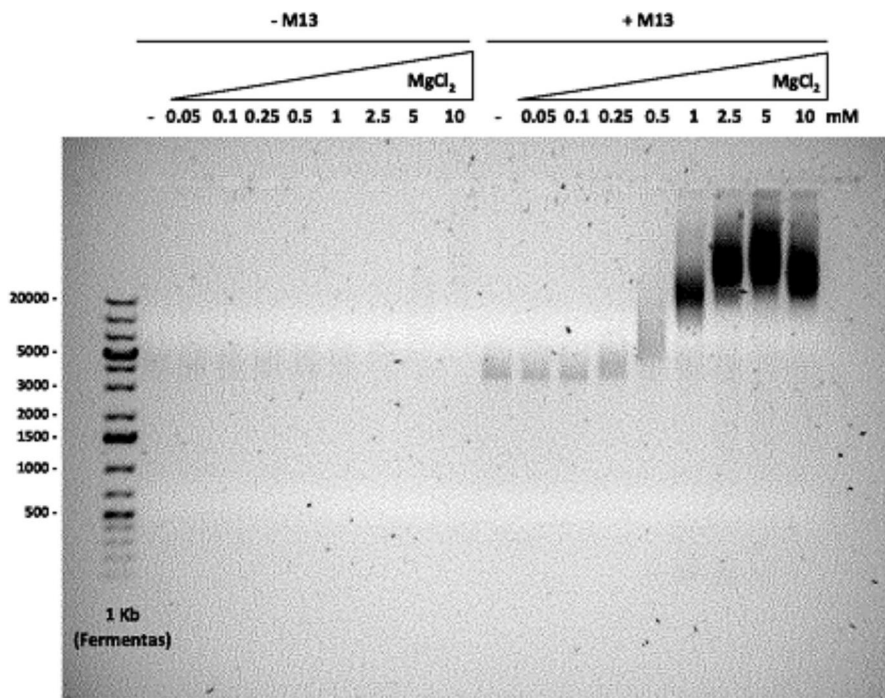


图13b

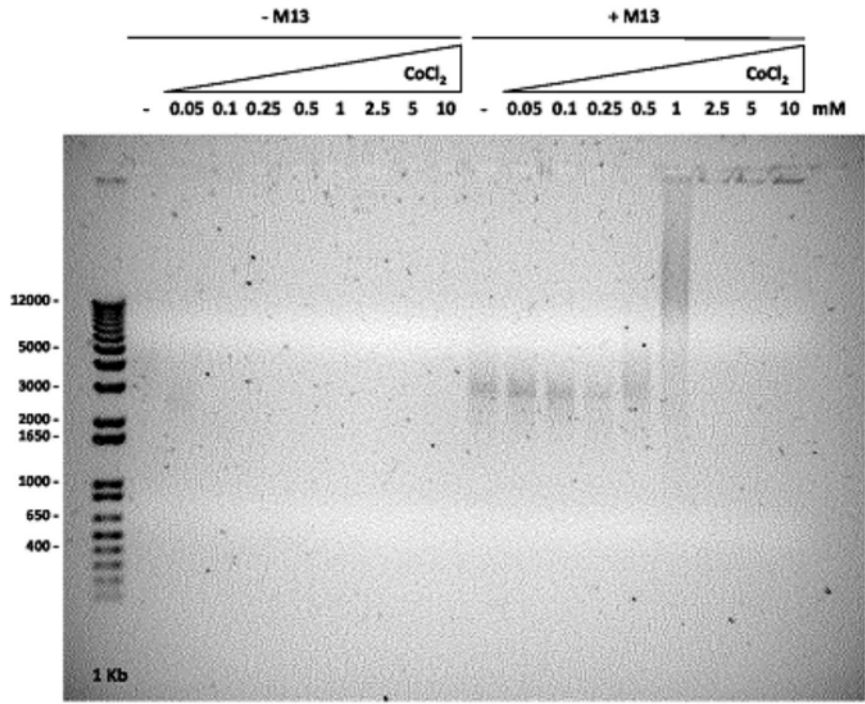


图13c

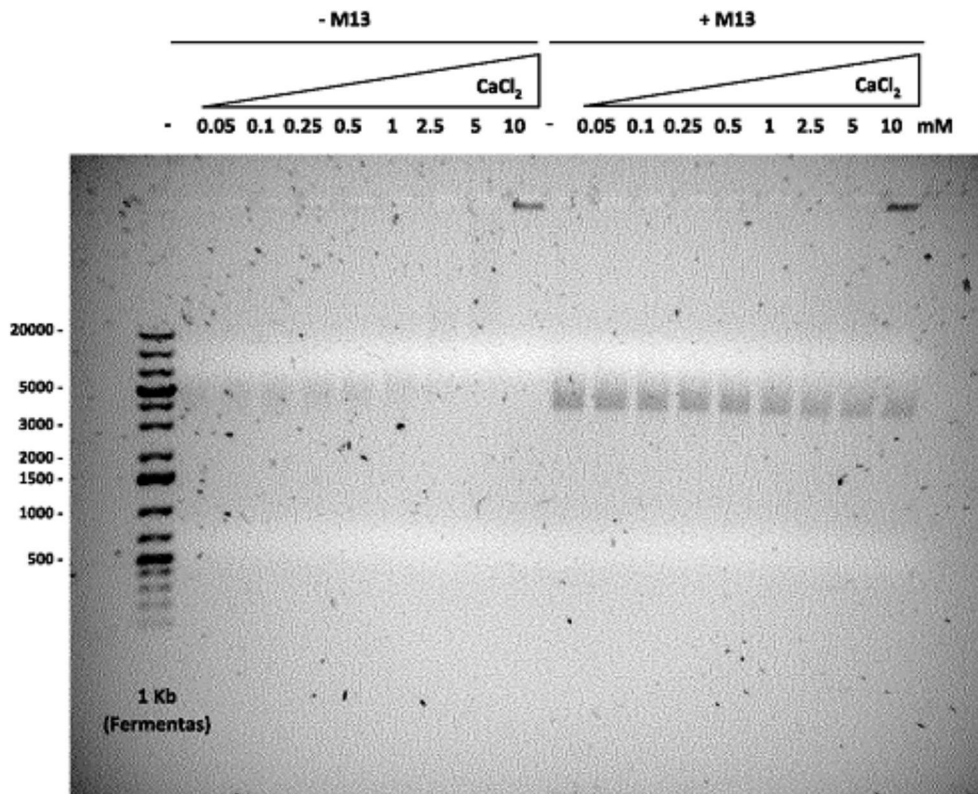


图13d

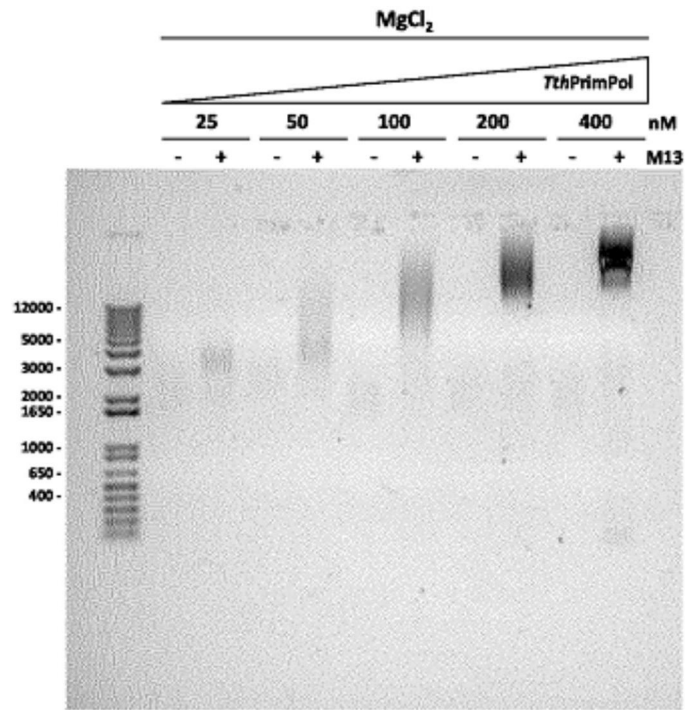


图14a

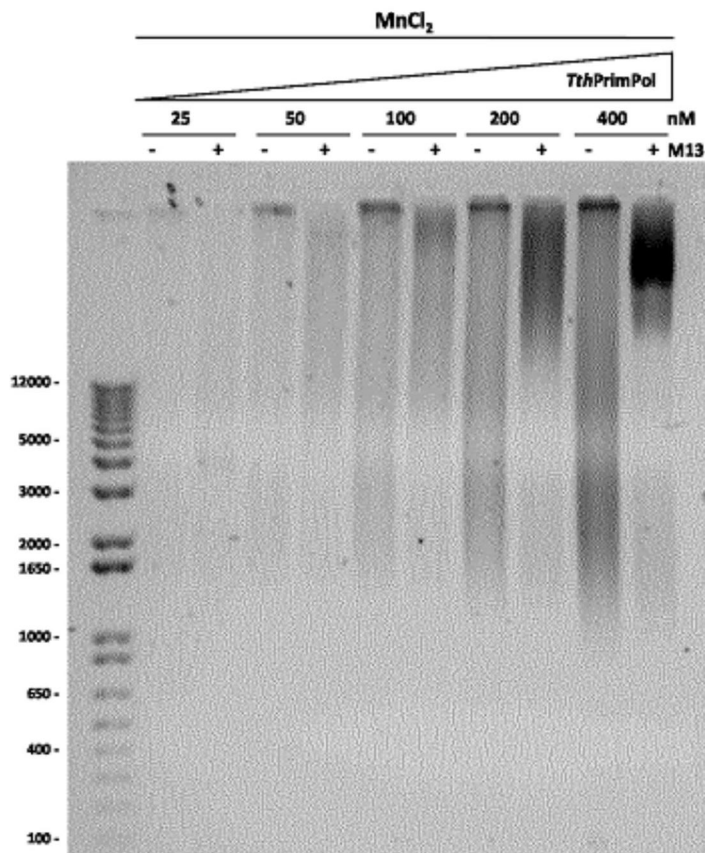


图14b

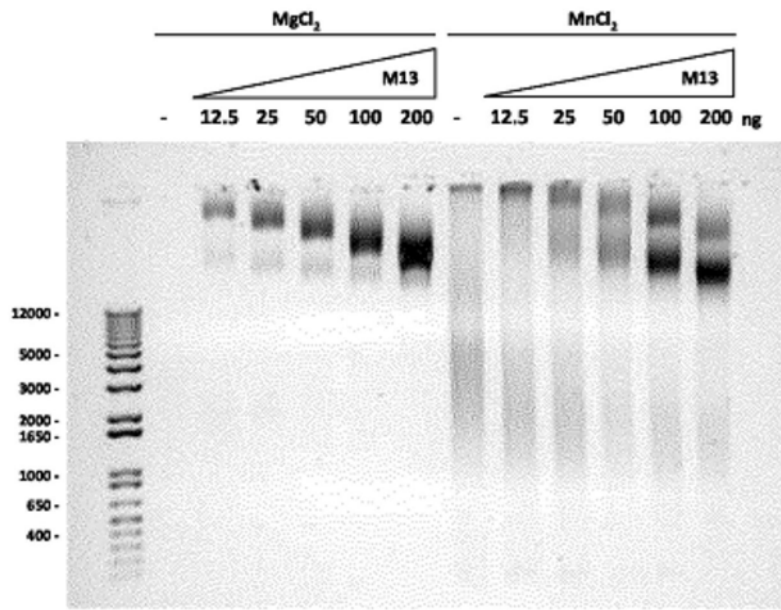


图15

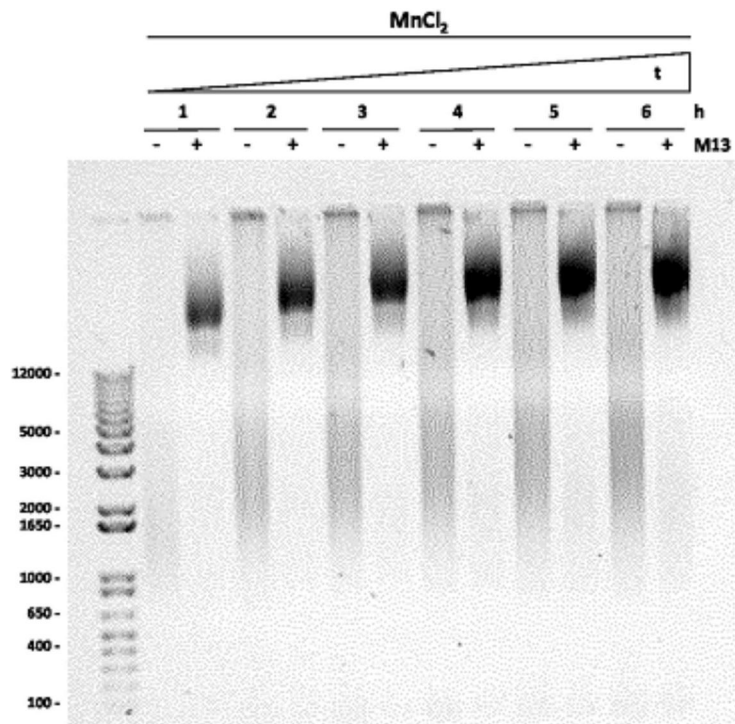


图16a



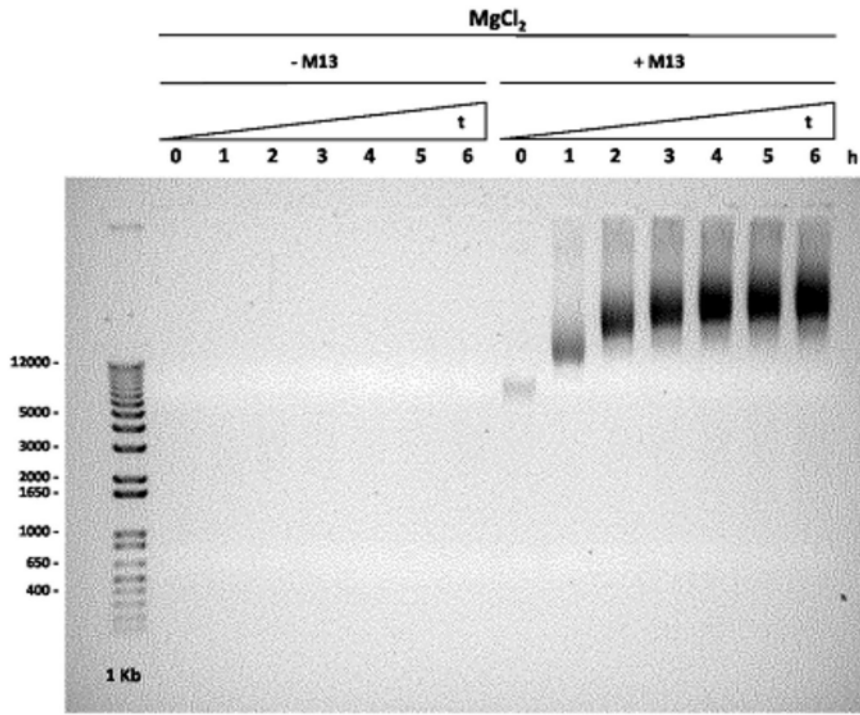


图16b

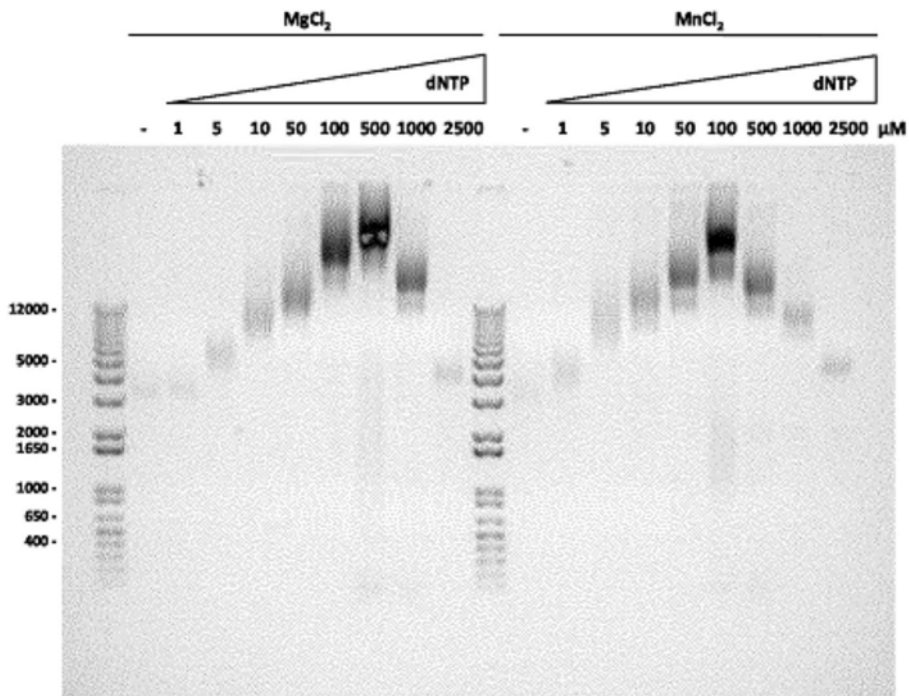


图17

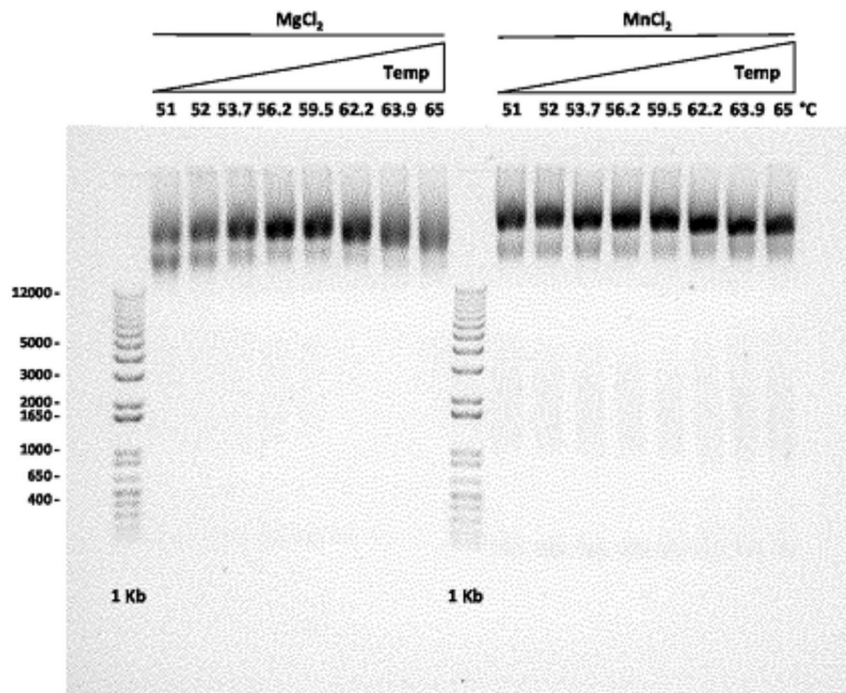


图18

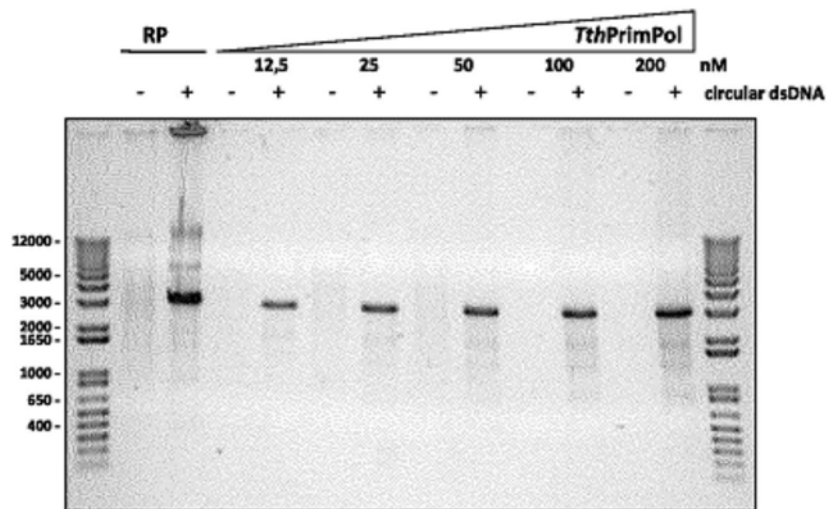


图19

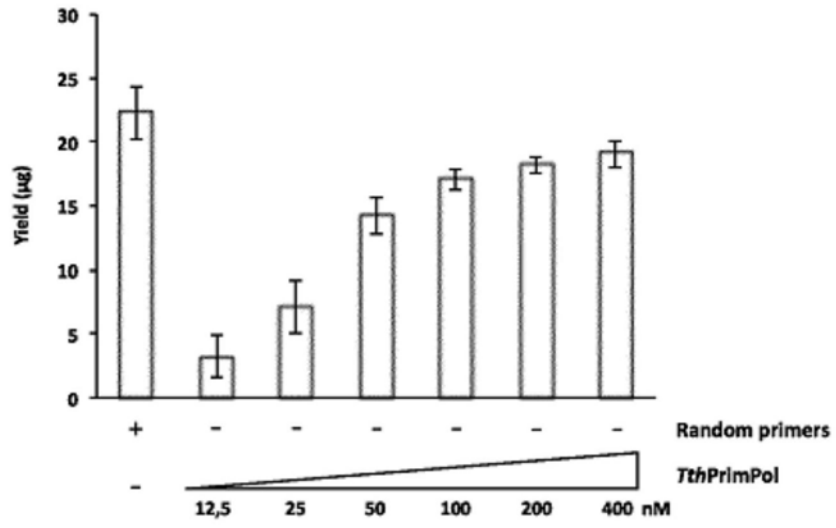


图20

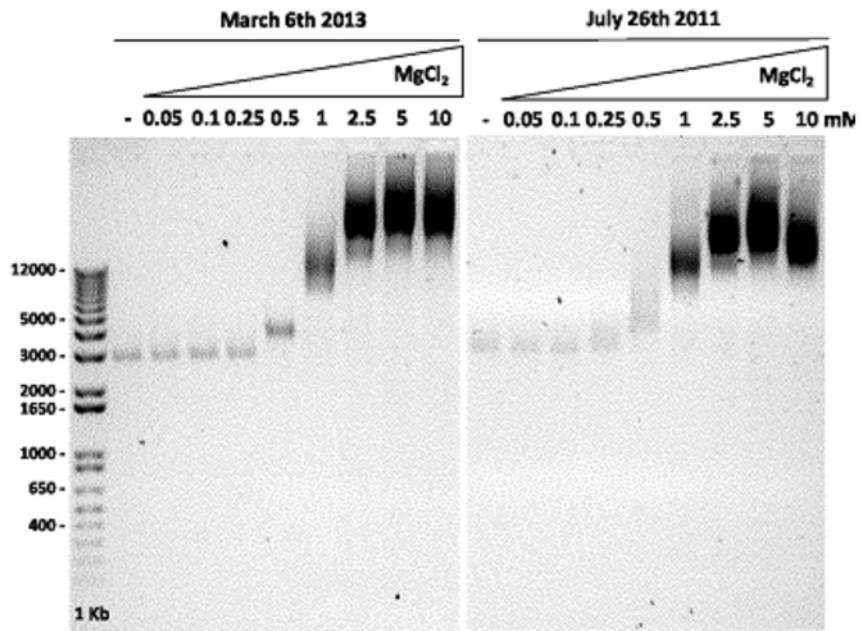


图21a

3' GACGTCGACTACGCGGCATGCCTAGGGGCCCATG 5'  
 \*5' CTGCAGCTGATGCGCC

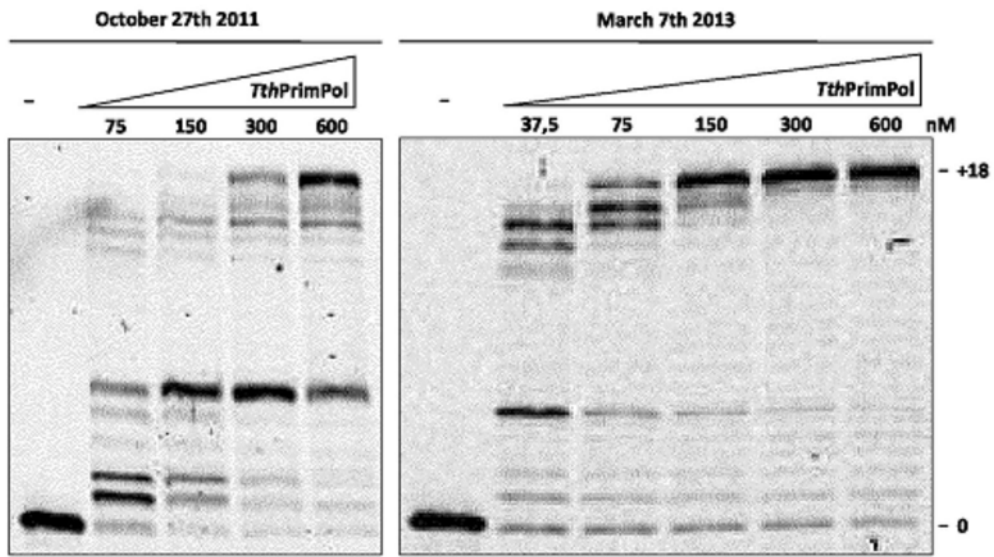


图21b

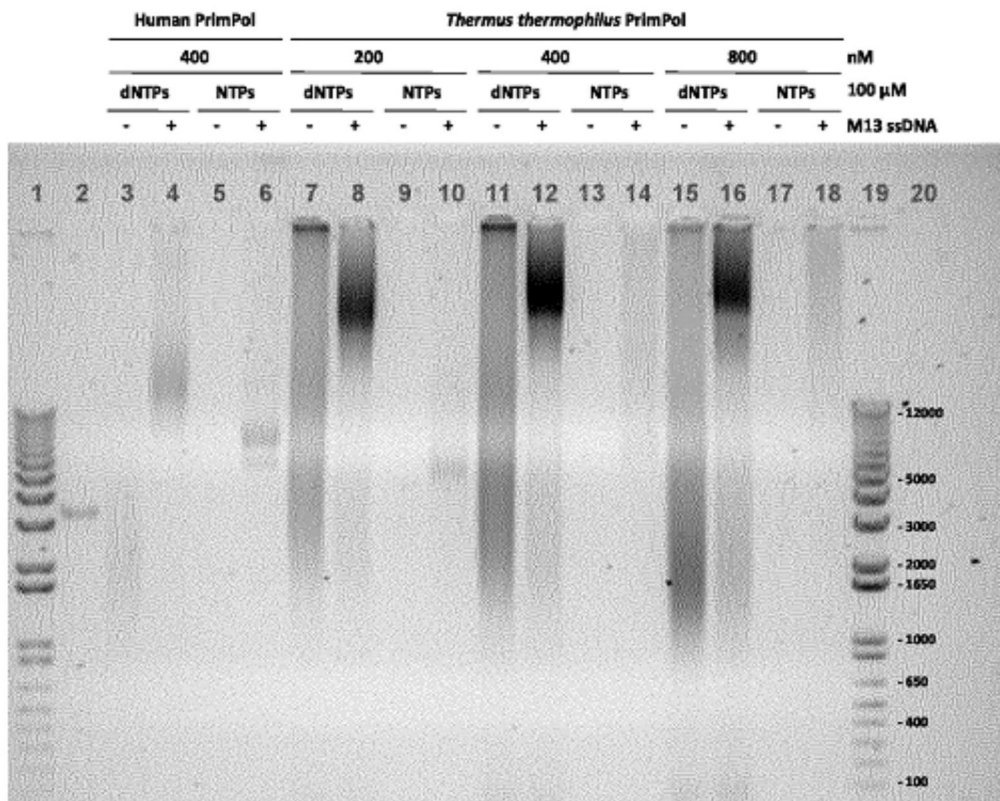


图22

3' CTAGTGTCACCTCATGGTCTATGTGAAGA 5'  
\*5' GATCACAGTGAGTAC AGATACTTCT 3'

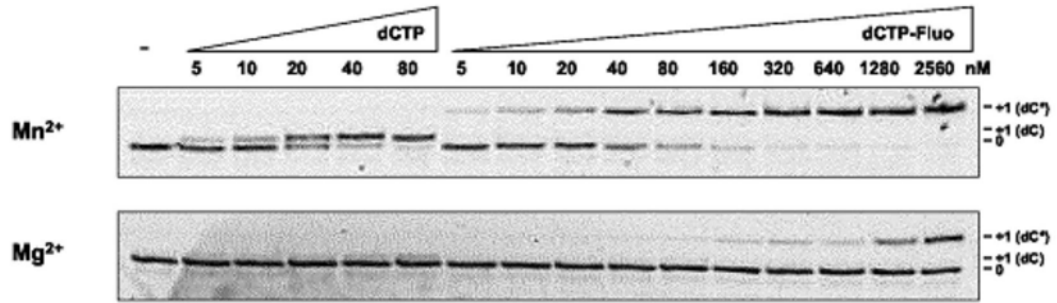


图23

```
>SEQ ID NO: 1
MRPIEHALSYAAQGYGVLPRLPGGKEPLGKLVPHGLKNASRDPATLEAWWRSCPRCGVILP
GPEVLVLDLDFDDPEAWGLRQEHFALEAAAPRQRTPKGGRHVFLRLPEGVRLSASVRAIPGVDL
RGMGRAYVVAAPTRLKDGRTYTWEAPLTPPEELPPVPQALLLKLPPPPPPRPSWGA VGTAS
PKRLQALLQAYAAQVARTPEGQRHLTLIRYAVAAGGLIPHGLDPREAEV LVAAM SAGLPE
WEARDAVRWGLGVGASRPLVLESSSKPPEPRTYRARVYARMRRWV

>SEQ ID NO: 2
MGSSHHHHSSSGLVPRGSHMRPIEHALSYAAQGYGVLPRLPGGKEPLGKLVPHGLKNASRDP
ATLEAWWRSCPRCGVILPGPEVLVLDLDFDDPEAWGLRQEHFALEAAAPRQRTPKGGRHVFLRL
PEGVRLSASVRAIPGVDLRGMGRAYVVAAPTRLKDGRTYTWEAPLTPPEELPPVPQALLLKL
PPPPPPRPSWGA VGTASPKRLQALLQAYAAQVARTPEGQRHLTLIRYAVAAGGLIPHGLDPRE
AEV LVAAM SAGLPEWEARDAVRWGLGVGASRPLVLESSSKPPEPRTYRARVYARMRRWV

>SEQ ID NO: 3
TTTTTTTTTTTTTTCCTGTTTTTTTTTT

>SEQ ID NO: 4
TTTTTTTTTTTTTTCCTXTTTTTTTTTTT

>SEQ ID NO: 5
GTACCCGGGGATCCGTACGGCGCATCAGCTGCAG

>SEQ ID NO: 6
CTGCAGCTGATGCGCC

>SEQ ID NO: 7
ACTGGCCGTCGTTCTATTGTACTCACTGTGATC

>SEQ ID NO: 8
GATCACAGTGAGTAC

>SEQ ID NO: 9
AACGACGGCCAGT

>SEQ ID NO: 10
AGAAGTGTATCTTGTACTCACTGTGATC

>SEQ ID NO: 11
CCGGCCCATATGAGGCCGATTGAGCACGCC

>SEQ ID NO: 12
AGAAGUGUAUCUUGUACUCACUGUGAUC

>SEQ ID NO: 13
CTGCAGCTGATGCGCXGTACGGATCCCCGGGTAC

>SEQ ID NO: 14
GTACCCGGGGATCCGTAC

>SEQ ID NO: 15
TTTTTTTTTTTTT

>SEQ ID NO: 16
AAAAAAAAAAAAA

>SEQ ID NO: 17
GCGCGGAATTCTCATACCCACCTCCTCATCCGGG
```

图24

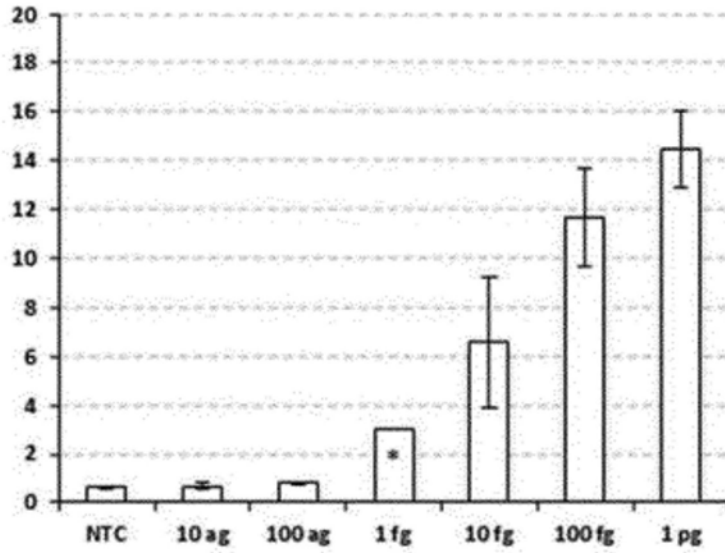


图25a

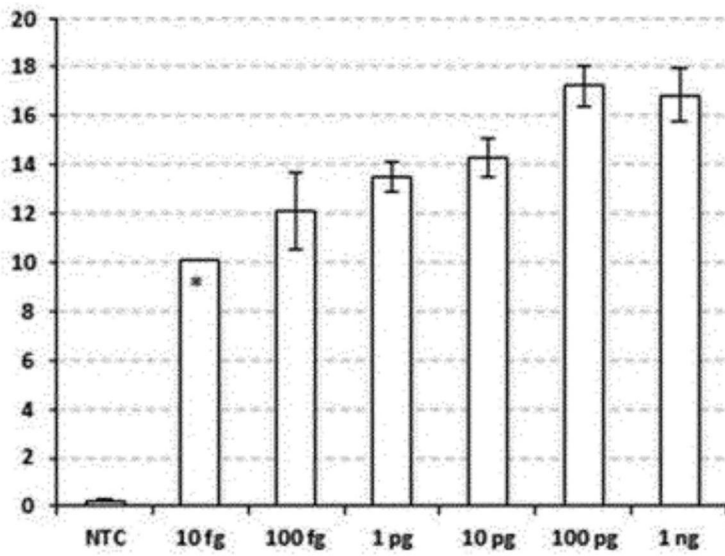


图25b

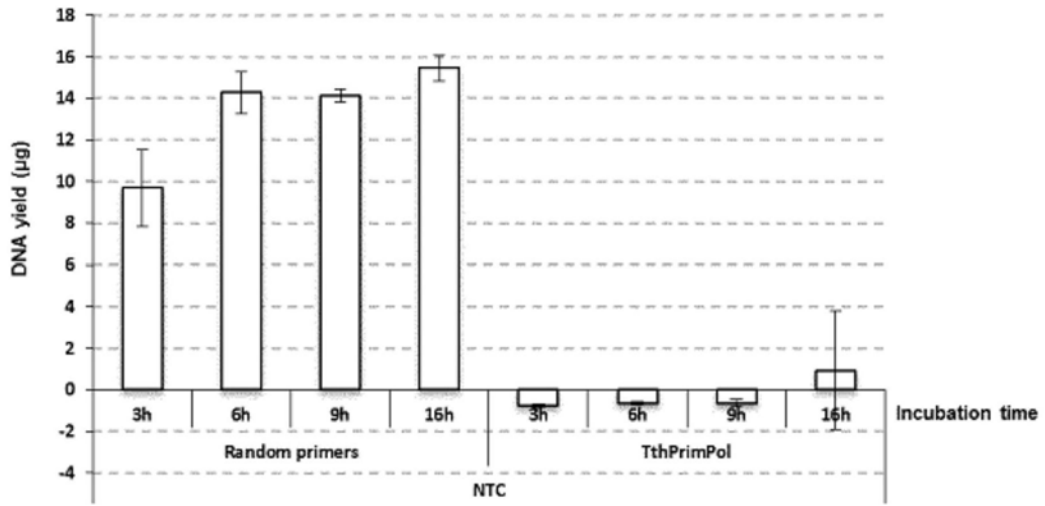


图26a

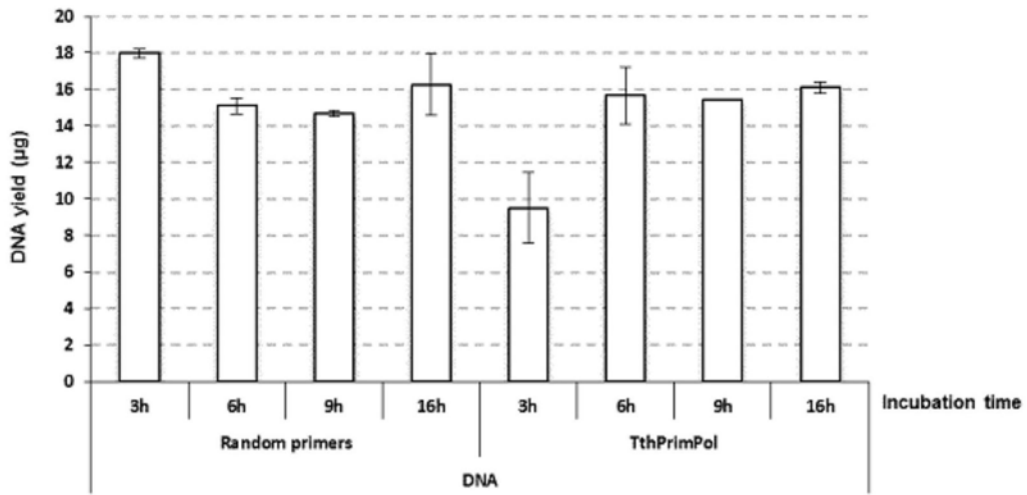


图26b



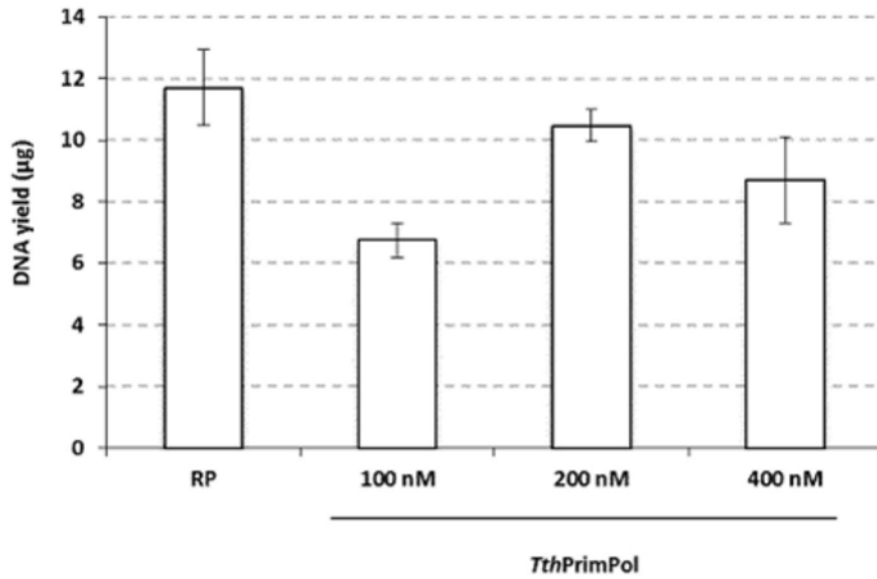


图27

	Sample	Count	Percentage of reads	Average length	Number of bases	Percentage of bases
Mapped reads	Tth-100/Phi	2.573.771	91,92%	253,21	651.711.875	93,41%
Not mapped reads		226.201	8,08%	203,14	45.949.636	6,59%
Total reads		2.799.972	100,00%	249,17	697.661.511	100,00%
Mapped reads	Tth-200/Phi	2.533.407	90,48%	254,79	645.478.540	91,25%
Not mapped reads		266.565	9,52%	232,16	61.885.052	8,75%
Total reads		2.799.972	100,00%	252,63	707.363.592	100,00%
Mapped reads	RP	2.349.664	83,92%	254,87	598.866.336	89,94%
Not mapped reads		450.258	16,08%	148,84	67.017.422	10,06%
Total reads		2.799.922	100,00%	237,82	665.883.758	100,00%
Mapped reads	NA	2.570.299	91,80%	260,15	668.666.985	95,36%
Not mapped reads		229.643	8,20%	141,75	32.551.077	4,64%
Total reads		2.799.942	100,00%	250,44	701.218.062	100,00%

图28

Reference	Tth-100/Phi			RP			NA		
	Reads	Length	Signatures	Reads	Length	Signatures	Reads	Length	Signatures
chr01_BY4741	49.200	230.221	141	43.615	230.221	173	49.902	230.221	96
chr02_BY4741	163.037	813.172	421	154.626	813.172	536	154.393	813.172	199
chr03_BY4741	60.754	316.620	225	64.010	316.620	242	64.803	316.620	142
chr04_BY4741	303.201	1.531.933	839	288.631	1.531.933	1.061	277.857	1.531.933	393
chr05_BY4741	125.522	576.874	341	115.451	576.874	441	117.228	576.874	195
chr06_BY4741	62.177	270.161	186	60.104	270.161	190	57.101	270.161	92
chr07_BY4741	221.523	1.090.941	521	202.750	1.090.941	772	202.398	1.090.941	293
chr08_BY4741	121.616	562.641	347	108.334	562.641	351	115.570	562.641	156
chr09_BY4741	99.845	439.890	257	87.148	439.890	274	90.709	439.890	139
chr10_BY4741	152.952	745.751	420	136.183	745.751	523	146.790	745.751	219
chr11_BY4741	144.347	666.813	386	129.798	666.813	482	128.956	666.813	189
chr12_BY4741	284.192	1.078.180	884	248.328	1.078.180	959	437.999	1.078.180	787
chr13_BY4741	197.248	924.428	487	173.647	924.428	556	176.070	924.428	265
chr14_BY4741	176.831	784.331	450	152.806	784.331	522	152.875	784.331	198
chr15_BY4741	218.242	1.091.288	595	199.820	1.091.288	668	204.714	1.091.288	273
chr16_BY4741	192.097	948.066	499	182.585	948.066	695	178.277	948.066	243
chr17_BY4741	987	85.779	21	1.828	85.779	24	14.657	85.779	80
Total	2.573.771	12.157.089	7.020	2.349.664	12.157.089	8.469	2.570.299	12.157.089	3.959

图29