

[19] 中华人民共和国国家知识产权局

[51] Int. Cl⁷

C12N 15/52

C12N 15/54 C12N 9/00

C12N 9/12



[12] 发明专利申请公开说明书

[21] 申请号 200310114395. X

[43] 公开日 2004 年 12 月 8 日

[11] 公开号 CN 1552869A

[22] 申请日 1996. 12. 26

[21] 申请号 200310114395. X

分案原申请号 96180125. 5

[30] 优先权

[32] 1995. 12. 27 [33] JP [31] 353778/1995

[71] 申请人 宝酒造株式会社

地址 日本京都府京都市

[72] 发明人 上森隆司 石野良纯 加藤郁之进

[74] 专利代理机构 中国专利代理(香港)有限公司

代理人 张广育

权利要求书 1 页 说明书 49 页 附图 6 页

[54] 发明名称 新型 DNA 聚合酶

[57] 摘要

本发明涉及具有如下性质的 DNA 聚合酶：(1) 在以单链模板与引物退火形成的复合物为底物时测定的聚合酶活性，比以活化的 DNA 作模板时显示的聚合酶活性高。(2) 具有 3' → 5' 外切核酸酶活性。(3) 在使用 λ DNA 作为模板进行 PCR 反应时，能够扩增约 20 千碱基对的 DNA 片段而不需要加入任何其他酶；构成该 DNA 聚合酶的蛋白质；含有编码该酶的碱基序列的 DNA；以及制备该 DNA 聚合酶的方法。本发明所提供的是一种同时含有高引物延伸活性和 3' → 5' 外切核酸酶活性的新型 DNA 聚合酶。

I S S N 1 0 0 8 - 4 2 7 4

1. 一种 DNA，它在如下条件下与含有编码 SEQ ID NO: 1 所示的氨基酸序列或在 SEQ ID NO: 1 中有一个或几个氨基酸缺失、插入、
5 增加或置换的氨基酸序列的碱基序列的 DNA 杂交：6×SSC 溶液中含有 0.5% SDS、0.1% 牛血清白蛋白、0.1% 聚乙烯吡咯烷酮、0.1% Ficoll 400 和 0.01% 变性鲑精 DNA，50°C，12 小时，所述 DNA 含有编码一种 DNA 聚合酶组成蛋白的碱基序列，所述蛋白在含有 SEQ ID NO: 3 的氨基酸序列的 DNA 聚合酶组成蛋白共存时显示 DNA 聚合酶活性。

10

2. 一种 DNA，它在如下条件下与含有编码 SEQ ID NO: 3 所示的氨基酸序列或在 SEQ ID NO: 3 中有一个或几个氨基酸缺失、插入、
增加或置换的氨基酸序列的碱基序列的 DNA 杂交：6×SSC 溶液中含有 0.5% SDS、0.1% 牛血清白蛋白、0.1% 聚乙烯吡咯烷酮、0.1% Ficoll
15 400 和 0.01% 变性鲑精 DNA，50°C，12 小时，所述 DNA 含有编码一种 DNA 聚合酶组成蛋白的碱基序列，所述蛋白在含有 SEQ ID NO: 1 的氨基酸序列的 DNA 聚合酶组成蛋白共存时显示 DNA 聚合酶活性。

新型 DNA 聚合酶

技术领域

5 本发明专利申请是申请日为 1996 年 12 月 26 日的中国专利申请 96180125.5 的分案申请。

本发明涉及作为遗传工程用试剂有用的 DNA 聚合酶及其制造方法，以及编码该酶的基因。

10 背景技术

DNA 聚合酶是作为遗传工程用试剂有用的酶，并被广泛用于 DNA 碱基序列测定、标记、位点特异性诱变等方法。而且，由于近来聚合酶链反应（PCR）方法的发展，耐热 DNA 聚合酶引起了人们的注意。已经有多种适用于 PCR 方法的 DNA 聚合酶得到了发展并获得了商业
15 化。

现在已知的 DNA 聚合酶根据核苷酸序列的同源性可大体上分为 4 个家族，其中家族 A（pol I 型酶）和家族 B（ α 型酶）占大多数。虽然属于各家族的 DNA 聚合酶一般具有相似的生物化学特征，但详细的比较显示，每种酶在底物特异性、底物类似物的整合、引物延伸性的
20 强度和速度、DNA 合成的模式、外切核酸酶活性的伴随、温度、pH 等最适反应条件、以及对抑制剂的敏感性等方面都有独特的性质。因此，在相应的实验操作中应在各种现有的 DNA 聚合酶中选用具有最适性质的 DNA 聚合酶。

25 发明内容

本发明的目的是提供不属于上述家族的、且具有现有 DNA 聚合酶没有的生物化学特性的新型 DNA 聚合酶。例如，引物延伸活性与 3'→5' 外切核酸酶活性被认为是相反的特性，现存的 DNA 聚合酶中具备强的引物延伸活性的酶不具有对合成忠实性起重要校正功能作用的 3'→5' 外切核酸酶活性。同时校正功能优秀的酶的引物延伸活性弱。因此，
30 开发具有强的 3'→5' 外切核酸酶活性并且引物延伸活性也很强的 DNA 聚合酶，对试管内 DNA 合成是很有意义的。

本发明的另一目的是提供上述新型 DNA 聚合酶的制造方法。

本发明的另一目的是提供编码本发明的 DNA 聚合酶的基因。

发明人锐意研究的结果是，在极端嗜热性古细菌激烈热球菌 (*Pyrococcus furiosus*) 中发现了新型 DNA 聚合酶的基因，在克隆之后，弄清楚了只有两种新型蛋白质同时存在时，其新型 DNA 聚合酶活性才能显示。而且，本发明还制备了导入了上述基因的转化子，并成功地大量生产了这种复合型 DNA 聚合酶，至此本发明完成。

即，本发明的要点涉及：

(1) 具有下列特性的 DNA 聚合酶：

①在以单链模板 DNA 与引物退火形成的复合物为底物时，测定的聚合酶活性比以活化的 DNA 为底物时显示的活性高。

②具有 3'→5'外切核酸酶活性。

③在按下述条件以 λ -DNA 为模板进行聚合酶链反应 (PCR) 时，能够扩增约 20 千碱基对的 DNA 片段。

PCR 的条件：

(a) 反应液组成：含 10 mM Tris-盐酸 (pH9.2)、3.5 mM MgCl₂、75 mM KCl、400 μ M dATP、dCTP、dGTP、dTTP、0.01%小牛血清白蛋白、0.1% Triton X-100、5.0 ng/50 μ l λ -DNA、10 pmol/50 μ l 引物 λ 1 (序列表的序列号 8)、以及引物 λ 11 (序列表的序列号 9)、3.7 单位/50 μ l DNA 聚合酶。

(b) 反应条件：以每循环 98°C、10 秒~68°C、10 分钟进行 30 个循环的 PCR 反应。

(2) 前面 (1) 记载的 DNA 聚合酶，其特征为与 Taq DNA 聚合酶相比，在 DNA 合成时的错误频率低；

(3) 前面 (1) 或 (2) 记载的 DNA 聚合酶，用凝胶过滤法测得的分子量为约 220 千道尔顿或约 385 千道尔顿；

(4) 前面 (1) ~ (3) 记载的任何 DNA 聚合酶，其特征为在两种 DNA 聚合酶组成蛋白 (第 1 种 DNA 聚合酶组成蛋白和第 2 种 DNA 聚合酶组成蛋白) 共存时能显示活性；

(5) 前面 (4) 记载的 DNA 聚合酶，其特征是，第 1 种 DNA 聚合酶组成蛋白和第 2 种 DNA 聚合酶组成蛋白用 SDS-PAGE 测得的分子量分别为约 90,000 道尔顿和约 140,000 道尔顿；

(6) 前面(4)或(5)中记载的DNA聚合酶,其特征为,构成前面(4)或(5)中记载的DNA聚合酶的第1种DNA聚合酶组成蛋白包括序列表的序列号1中所示的氨基酸序列或是在该氨基酸序列中有一个以上的氨基酸缺失、插入、附加或置换并具有基本相同的活性的功能等价物;

(7) 前面(4)或(5)中记载的DNA聚合酶,其特征为,构成前面(4)或(5)中记载的DNA聚合酶的第2种DNA聚合酶组成蛋白包括序列表的序列号3中所示的氨基酸序列或是在该氨基酸序列中有一个以上的氨基酸缺失、插入、附加或置换并具有基本相同的活性的功能等价物;

(8) 前面(4)或(5)中记载的DNA聚合酶,其特征为,构成前面(4)或(5)中记载的DNA聚合酶的第1种DNA聚合酶组成蛋白包括序列表的序列号1中所示的氨基酸序列或是在该氨基酸序列中有一个以上的氨基酸缺失、插入、附加或置换并具有基本相同的活性的功能等价物;构成前面(4)或(5)中记载的DNA聚合酶的第2种DNA聚合酶组成蛋白包括序列表的序列号3中所示的氨基酸序列或是在该氨基酸序列中有一个以上的氨基酸缺失、插入、附加或置换并具有基本相同的活性的功能等价物;

(9) 构成前面(4)或(5)中记载的DNA聚合酶的第1种DNA聚合酶组成蛋白,包括序列表的序列号1中所示的氨基酸序列,或包括在该氨基酸序列中有一个以上的氨基酸缺失、插入、附加或置换且具有基本相同的活性的功能等价物的氨基酸序列;

(10) 构成前面(4)或(5)中记载的DNA聚合酶的第2种DNA聚合酶组成蛋白,包括序列表的序列号3中所示的氨基酸序列,或包括在该氨基酸序列中有一个以上的氨基酸缺失、插入、附加或置换且具有基本相同的活性的功能等价物的氨基酸序列;

(11) 含有编码前面(9)中记载的第1种DNA聚合酶组成蛋白的碱基序列的DNA,其特征为含有编码序列表的序列号1中所示的氨基酸序列的碱基序列的全部或一部分,或者编码含有在序列号1的氨基酸序列中有一个以上的氨基酸缺失、插入、附加或置换的氨基酸序列且具有第1种DNA聚合酶组成蛋白的功能的蛋白质;

(12) 含有编码前面(9)中记载的第1种DNA聚合酶组成蛋白

的碱基序列的 DNA, 其特征为含有序列表的序列号 2 中所示的碱基序列的全部或一部分, 或者含有在严格条件下能与该序列杂交的碱基序列;

5 (13) 含有编码前面 (10) 中记载的第 2 种 DNA 聚合酶组成蛋白的碱基序列的 DNA, 其特征为含有编码序列的序列号 3 中所示的氨基酸序列的碱基序列的全部或一部分, 或者编码含有在序列号 3 的氨基酸序列中有一个以上的氨基酸缺失、插入、附加或置换的氨基酸序列且具有第 2 种 DNA 聚合酶组成蛋白的功能的蛋白质;

10 (14) 含有编码前面 (10) 中记载的第 2 种 DNA 聚合酶组成蛋白的碱基序列的 DNA, 其特征为含有序列表的序列号 4 中所示的碱基序列的全部或一部分, 或者含有在严格条件下能与该序列杂交的碱基序列;

(15) DNA 聚合酶的制造方法, 其特征为, 培养含有编码前面 (11) 或 (12) 中记载的第 1 种 DNA 聚合酶组成蛋白的基因和编码前面 (13) 或 (14) 中记载的第 2 种 DNA 聚合酶组成蛋白的基因两者的转化子, 15 从该培养物中收集 DNA 聚合酶; 和

(16) DNA 聚合酶的制造方法, 其特征为分别培养含有编码前面 (11) 或 (12) 中记载的第 1 种 DNA 聚合酶组成蛋白的基因的转化子和含有编码前面 (13) 或 (14) 中记载的第 2 种 DNA 聚合酶组成蛋白 20 的转化子, 混合培养物中所含的 DNA 组成蛋白, 收集 DNA 聚合酶。

附图说明

图 1 是实施例 1 中获得的粘粒克隆 No. 264 和粘粒克隆 No. 491 中插入的 DNA 片段的限制酶切图谱。

25 图 2 是质粒 pFU1001 中插入的 XbaI-XbaI DNA 片段的限制酶切图谱。

图 3 是本发明的 DNA 聚合酶的最适 pH 的示意图。

图 4 是本发明的 DNA 聚合酶的热稳定性示意图。

图 5 是本发明的 DNA 聚合酶的 3'→5'外切核酸酶活性的示意图。

30 图 6 是本发明的 DNA 聚合酶的引物延伸活性的放射性自显影图。

具体实施方案

(1) 本发明的 DNA 聚合酶及其组成蛋白

本发明的 DNA 聚合酶的一个例子具有以下性质:

①在以单链模板与引物退火形成的复合物为底物时测定的聚合酶活性, 比以活化的 DNA (DNase I 处理的小牛胸腺 DNA) 作模板时显示 5 的聚合酶活性高。

②具有 3'→5'外切核酸酶活性。

③最适 pH: 6.5-7.0(磷酸钾缓冲液中、75°C)。

④80°C、30 分钟热处理后能显示 80%的残存活性。

⑤并且, 在以 λ -DNA 为模板进行聚合酶链反应 (PCR) 时, 能够扩 10 增约 20 千碱基对的 DNA 片段而不需要添加其他的酶。PCR 的条件如下:

(a) 反应液组成: 含 10 mM Tris-盐酸 (pH9.2)、3.5 mM MgCl₂、75 mM KCl、400 μ M dATP、dCTP、dGTP、dTTP、0.01%小牛血清白蛋白 (BSA)、0.1% Triton X-100、5.0 ng/50 μ l λ -DNA、10 pmol/50 15 μ l 引物 λ 1(序列表的序列号 8)、以及引物 λ 11(序列表的序列号 9)、3.7 单位/50 μ l DNA 聚合酶。在这里, DNA 聚合酶的 1 个单位按如下方法确定。制备 50 μ l 含有待测定活性样品的反应液[20 mM Tris-盐酸 (pH7.7)、15 mM MgCl₂、2 mM 2-巯基乙醇、0.2 mg/ml 活化 DNA、40 μ M dATP、dCTP、dGTP、dTTP、60 nM [³H]dTTP (Amersham 公司 20 生产)], 在 75°C 反应 15 分钟。将此反应液 40 μ l 点于 DE 滤纸(Whatman 公司生产), 用 5% Na₂HPO₄ 洗涤 5 次, 然后用液闪计数器测定 DE 滤纸上残留的放射活性, 在 30 分钟内将 10 nmol 的 [³H]dTTP 掺入底物 DNA 的酶量定义为 1 个单位。

(b) 反应条件: 以每循环 98°C、10 秒~68°C、10 分钟进行 30 25 个循环的 PCR 反应。

⑥本发明的 DNA 聚合酶在引物延伸活性、DNA 合成的正确性两方面与 Taq DNA 聚合酶比较都更优秀。即, 本发明的 DNA 聚合酶的 DNA 合成反应的延伸活性, 例如在 PCR 反应中被扩增 DNA 的链长, 和反应的正确性 (DNA 合成时错误频率低) 与作为耐热性 DNA 聚合酶代表的 30 Taq DNA 聚合酶 (例如宝酒造社生产的 TaKaRa Taq) 相比都更为优秀。

本发明的 DNA 聚合酶的分子量用凝胶过滤法测定约为 220 千道尔顿或约 385 千道尔顿, 在 SDS-PAGE 上显示的是与约 90,000 道尔顿

和约 140,000 道尔顿相当的两条带,它是由两种蛋白质组成的酶。在本发明中,约 90,000 道尔顿的蛋白质(相当于下文的 ORF3)被称为第 1 种 DNA 聚合酶组成蛋白,约 140,000 道尔顿的蛋白质(相当于下文的 ORF4)被称为第 2 种 DNA 聚合酶组成蛋白。经过推定,本发明的 DNA 聚合酶是由第 1 种 DNA 聚合酶组成蛋白与第 2 种 DNA 聚合酶组成蛋白以 1:1 或 1:2 的摩尔比通过非共价结合形成的复合体。

构成本发明的 DNA 聚合酶的第 1 种 DNA 聚合酶组成蛋白包括序列表的序列号 1 中所示的氨基酸序列,或也可以是具有与其基本相同活性的功能等价物,而第 2 种 DNA 聚合酶组成蛋白含有序列表的序列号 3 中所示的氨基酸序列,或也可以是具有与其基本相同活性的功能等价物。

在本说明书中记载的“功能等价物”的含义如下。除了天然存在的蛋白质的编码基因的多形性及变异之外,生成的蛋白质因在体内或纯化中的修饰反应等可在氨基酸序列中引起氨基酸的缺失、插入、附加、置换等变异,但已知某些蛋白质能显示与没有变异的蛋白质基本上相同的生理活性和生物学活性。这种在构造上有差异但在功能和活性上没有发现大的差异的蛋白质就被称为功能等价物。在这里,只要基本上能显示同等的生理学和生物学活性,变异氨基酸的数目没有特别的限制。具体的例子为 1 个以上,如 1 个或几个,更具体为 1 个至约 10 个变异(缺失、插入、附加、置换等)。

人为地在蛋白质的氨基酸序列中引入上述变异时也一样,在这种情况下,可以制备各种各样的变异体。例如,虽然在大肠杆菌中表达的蛋白质的 N 末端存在的甲硫氨酸残基在多数情况下用甲硫氨酸氨肽酶来去除。但这种去除是否完全取决于蛋白质的种类,带有或不带有甲硫氨酸的蛋白质两者都可能生成。尽管如此,在大多数情况下,甲硫氨酸残基的有无并不影响蛋白的活性。而且,已经知道白细胞介素 2 (IL-2) 的氨基酸序列中的特定半胱氨酸残基被丝氨酸置换时,仍能保持白细胞介素 2 的活性[《科学》(Science),第 224 卷,1431 页(1984)]。

此外,在用遗传工程方法进行蛋白质的生产时,常以融合蛋白的形式来进行表达。例如,为了增加目的蛋白的表达量,在目的蛋白的 N 末端附加来自其他蛋白质 N 末端的肽链,或在 N 末端或 C 末端附加

适当的肽链，表达该蛋白，通过使用与此类肽链有亲和性的载体来使目的蛋白的纯化更加容易进行。因此，与本发明的 DNA 聚合酶有部分氨基酸的差别的 DNA 聚合酶，只要能显示与本发明的 DNA 聚合酶基本上相同的活性，该酶作为“功能等价物”也属于本发明的范围内。

5

(2) 本发明的 DNA 聚合酶的基因

编码构成本发明的 DNA 聚合酶的第 1 种 DNA 聚合酶组成蛋白的 DNA，可举出含有编码序列的序列号 1 中所示氨基酸序列的碱基序列的全部或一部分的 DNA，例如，含有序列的序列号 2 中所示的碱基序列的全部或一部分的 DNA。即，含有编码序列号 1 中所示氨基酸序列的碱基序列的一部分的 DNA，例如，含有序列的序列号 2 中所示碱基序列的一部分的 DNA，编码具有第 1 种 DNA 聚合酶组成蛋白的功能的蛋白质的碱基序列，也在本发明的范围内。并且，还可举出编码包括在序列号 1 的氨基酸序列中有一个以上如 1 个或几个氨基酸缺失、插入、附加或置换的氨基酸序列、并具有第 1 种 DNA 聚合酶组成蛋白功能的蛋白质的 DNA。并且，能与上述序列在严格条件下杂交、并编码具有第 1 种 DNA 聚合酶组成蛋白功能的蛋白质的碱基序列，也在本发明的范围内。另外，编码本发明的 DNA 聚合酶的第 2 种 DNA 聚合酶组成蛋白的 DNA，可举出含有编码序列的序列号 3 中所示氨基酸序列的碱基序列的全部或一部分的 DNA，例如，含有序列的序列号 4 中所示的碱基序列的全部或一部分的 DNA。即，含有编码序列号 3 中所示氨基酸序列的碱基序列的一部分的 DNA，例如，含有序列的序列号 4 中所示的碱基序列的一部分的 DNA，编码具有第 2 种 DNA 聚合酶组成蛋白的功能的蛋白质的碱基序列，也在本发明的范围内。并且，还可举出编码包括在序列号 3 的氨基酸序列中有一个以上如 1 个或几个氨基酸缺失、插入、附加或置换的氨基酸序列、并具有第 2 种 DNA 聚合酶组成蛋白的功能的蛋白质的 DNA。并且，能与上述序列在严格条件下杂交、并编码具有第 2 种 DNA 聚合酶组成蛋白功能的蛋白质的碱基序列，也在本发明的范围内。

在这里，“具有第 1 种 DNA 聚合酶组成蛋白的功能的蛋白质”或“具有第 2 种 DNA 聚合酶组成蛋白的功能的蛋白质”是指这两种蛋白质共存时，能够提供具有前面①~⑥中所示各种理化性质的 DNA 聚合

酶活性。

在这里，“能在严格条件下杂交”是指在含有 0.5%SDS、0.1%小牛血清白蛋白 (BSA)、0.1%聚乙烯吡咯烷酮、0.1% Ficol 400、0.01% 变性鲑精 DNA 的 6×SSC (1×SSC 表示 0.15 M NaCl、0.0015 M 柠檬酸钠、pH7.0) 中与探针在 50°C 保温 12-20 小时后，与探针杂交。

在此对本说明书中记载的“含有编码氨基酸序列的碱基序列的 DNA”这一用语进行如下说明：已知在基因上指导氨基酸的密码子 (3 碱基组合) 依氨基酸的种类存在 1-6 种。因此编码氨基酸的 DNA 根据氨基酸序列可有多种存在方式。在自然界中，基因并不一定稳定存在，基因的碱基序列上发生变异并不少见，有时基因上碱基序列的变异并不引起所编码氨基酸序列的变异 (称为无声突变)。在这种情况下，产生了编码相同氨基酸序列的不同基因。因此，即使在已经分离了的编码特定氨基酸序列的基因上，在经过含有此基因的生物传代后，也不能否定编码相同氨基酸的多种基因的可能性。

而且，采用各种遗传工程的方法人工制备编码相同氨基酸的多种基因也不困难。例如，用遗传工程方法生产蛋白质时，编码目的蛋白质的原始基因上使用的密码子在宿主中的使用频率很低，此时蛋白质的表达量也低。在这种情况下，在不引起所编码的氨基酸序列变化的前提下，人工用宿主高频率使用的密码子进行变换，能使目的蛋白高表达 (例如，特公平 7-102146 号公报)。因此人工制备编码特定氨基酸序列的多种基因是可能的，在自然界中也能生成。因此含有与本发明公开的碱基序列相同的碱基序列的基因，以及编码本发明公开的氨基酸序列的基因都包括在本发明中。

(3) 本发明的 DNA 聚合酶的制造方法

发明人从极端嗜热性古细菌激烈热球菌中发现了新型 DNA 聚合酶的基因，通过克隆，清楚了该基因编码新型 DNA 聚合酶，该基因上的两种蛋白质在共存时能显示活性。可以通过制备导入了这种基因的转化子来大量生产本发明的 DNA 聚合酶。在这种情况下，，可以通过培养含有前面的编码第 1 种 DNA 聚合酶组成蛋白的基因和编码第 2 种 DNA 聚合酶组成蛋白的基因两者的转化子，从培养物中收集 DNA 聚合酶的方法来进行。也可以通过分别培养含有编码第 1 种 DNA 聚合酶组

成蛋白的基因的转化子和含有编码第 2 种 DNA 聚合酶组成蛋白的基因的转化子，然后将培养物中所含的 DNA 聚合酶组成蛋白混合，然后收集 DNA 聚合酶的方法来进行。

- 5 在这里，“含有编码第 1 种 DNA 聚合酶组成蛋白的基因和第 2 种 DNA 聚合酶组成蛋白的基因两者的转化子”，可以是分别含有相应基因的两种表达载体同时转化而来的转化子，或者是通过将两种基因重组到能分别表达各自蛋白的一个表达载体上而制备的转化子。

- 10 (4) 下面详细说明本发明的 DNA 聚合酶基因的克隆、所获克隆的分析、表达产物 DNA 聚合酶的理化性质、活性、对 PCR 方法的适用性。

本发明使用的菌株没有特别的限制，例如属于热球菌属的激烈热球菌 (*pyrococcus furiosus*) DSM3628，该菌株可由 Deutsche Sammlung von Mikroorganismen und Zellkulturen GmbH 获得。发明人在适当的增殖培养基上培养该菌株，制备培养物的粗抽提物，进行聚丙烯酰胺凝胶电泳时，发现凝胶内存在多条能显示 DNA 聚合酶活性的蛋白质带，推测也存在各自对应的基因。即，新型 DNA 聚合酶基因及其产物可按如下举例程序进行克隆。

- ①从激烈热球菌中抽提 DNA。
- ②用适当的限制酶消化①的 DNA，用质粒、粘粒等作载体制备 DNA 文库。
- 20 ③将②中制备的文库导入大肠杆菌，外源基因表达后，收集各克隆的粗抽提物制备蛋白文库。
- ④使用③中制备的蛋白文库测定 DNA 聚合酶活性，从粗抽提物中有活性的大肠杆菌克隆中提取外源 DNA。
- 25 ⑤分析提取的质粒或粘粒中所含的激烈热球菌 DNA 片段，缩小能编码显示 DNA 聚合酶活性的蛋白质的区域。
- ⑥确定预计编码能显示 DNA 聚合酶活性的蛋白质的区域，推导蛋白质的一级结构。
- ⑦构建便于在大肠杆菌中表达⑥中推导的蛋白质的表达载体，纯化所产生的蛋白质，并分析其特性。
- 30

上述 DNA 供体激烈热球菌 DSM3638 是一种极端嗜热性古细菌，将此细菌在 95°C 厌氧培养。破碎增殖的菌体后抽提、纯化 DNA 的方法

和用限制酶将所获得的 DNA 切断的方法等都可采用众所周知的方法。这些方法详见 1982 年冷泉港实验室出版、T.Maniatis 等著《分子克隆实验指南》(Molecular Cloning, A Laboratory Manual) 第 75-178 页中的记载。

- 5 制备 DNA 文库时,可以使用例如三股螺旋粘粒载体(Stratagene 公司生产)。激烈热球菌的 DNA 用 Sau 3AI (宝酒造社生产) 部分消化,密度梯度离心所获得的长链 DNA 片段,与上述载体的 BamHI 位点连接,然后体外包装。将用此方法制备的 DNA 文库所获得的转化子分别培养,收集菌体,然后超声波处理以破碎菌体,进行热处理使大肠杆菌含有的 DNA 聚合酶失活。用离心法可以获得含有耐热性 DNA 聚合酶的上清。该上清即命名为粘粒蛋白文库。用部分上清测定 DNA 聚合酶活性,可以获得能表达来自激烈热球菌的 DNA 聚合酶的克隆。DNA 聚合酶活性的测定方法可以采用众所周知的方法,该方法在 Harpar and Row 公司出版、D.R.Davis 编辑的《大肠杆菌的 DNA 聚合酶》(DNA polymerase from *Escherichia coli*) 第 263-276 页(C.C.Richardson 著)中记载。
- 10
- 15

- 发明人已经克隆了一种激烈热球菌的 DNA 聚合酶基因,并清楚了其结构,见《核酸研究》(Nucleic Acids Research) 第 21 卷 259-265 页(1993)中记载。该基因的翻译产物由 775 个氨基酸组成,是分子量为 90,000 道尔顿的多肽,该氨基酸序列明显含有 α -型 DNA 聚合酶的保守序列。实际上,该基因产物所显示的 DNA 聚合酶活性可被 α -型 DNA 聚合酶的特异性抑制剂蚜栖菌素(aphidicolin)抑制,与本发明的 DNA 聚合酶不一样。因此,需从所获得的显示耐热性 DNA 聚合酶活性的克隆中将上述已经存在的基因除去,这可通过消化含有各克隆的粘粒,以上述基因作探针进行杂交,选择不能杂交的克隆来进行。对含有所获得的新型 DNA 聚合酶基因克隆的消化粘粒中的 DNA 插入片段作限制酶切图谱。下一步,将插入 DNA 片段按限制酶切图谱切成各种区域,分别在质粒中亚克隆,导入大肠杆菌后测定所表达的耐热性 DNA 聚合酶的活性,从而确定该 DNA 片段聚合酶基因的位置。通过这种方法能够获得含有 DNA 聚合酶基因的约 10 千碱基对的 XbaI-XbaI DNA 片段。
- 20
- 25
- 30

含有整合了该 DNA 片段的质粒的重组大肠杆菌,其抽提物经

90°C、20 分钟热处理后仍有足够的 DNA 合成活性，而没有整合该 DNA 片段的质粒则没有此活性。因此结论是，在该 DNA 片段上存在产生耐热性 DNA 聚合酶的信息，且在大肠杆菌中带有该信息的基因能够表达。该 DNA 片段整合进载体 pTV118N（宝酒造社生产）所得到的质粒
5 被命名为 pFU1001，用该质粒转化的大肠杆菌 JM109 被命名为 *Escherichia coli* JM109/pFU1001。该菌已在 1995 年 8 月 11 日（原始保藏日）按照布达佩斯条约，在日本茨城县つくば市东一丁目 1 番 3 号（邮编 305）的通产省工业技术院生命工学工业技术研究所，以保藏号 FERM BP-5579 进行了保藏。

10 质粒 pFU1001 中所插入的 DNA 片段的碱基序列，可以用常规方法，例如双脱氧法来测定。并且，通过分析所获得的碱基序列，就可以推导出该碱基序列中编码蛋白质的区域-开放阅读框（ORF）。

质粒 pFU1001 中所插入的约 10 千碱基对的 XbaI-XbaI DNA 片段中 8450 碱基对的序列在序列表的序列号 5 中表示。在该碱基序列中
15 存在 6 个连续的 ORF，自 5' 端依次被命名为 ORF1、ORF2、ORF3、ORF4、ORF5、ORF6。图 2 显示了上述 XbaI-XbaI 片段的限制酶切图谱，以及在该片段上存在的 ORF 的位置（自图中左侧，ORF1-ORF6）。

上述 6 个 ORF 中没有与已知 DNA 聚合酶显示同源的序列。但是在 ORF1 和 ORF2 中存在与啤酒酵母（*Saccharomyces cerevisiae*）中发
20 现的 CDC6 蛋白或粟酒裂殖酵母（*Schizosaccharomyces pombe*）中发现的 CDC18 蛋白质同源的序列。CDC6、CDC18 据推测是酵母菌在向细胞周期的 DNA 合成期（S 期）转变过程中必要的蛋白质，这两种蛋白质调节 DNA 复制的起始。而 ORF6 含有与已知在酵母 DNA 损伤修复及体细胞分裂期和减数分裂期的重组过程中起作用的 RAD51 蛋白质及
25 其在减数分裂期中的特异性同源物 Dmc1 蛋白质同源的序列。已知编码 RAD51 蛋白质的基因在细胞周期由 G1 向 S 期转变时也被诱导表达。对于其他 ORF，即 ORF3、ORF4、ORF5，没有发现与已知蛋白质有同源性的序列。

可以通过将缺失各区域的 DNA 片段插入来制备重组质粒，测定用
30 这种质粒转化获得的转化子的耐热聚合酶活性，来调查耐热性 DNA 聚合酶的活性来自上述哪个 ORF。用插入了由上述约 10 千碱基对的 XbaI-XbaI DNA 片段中缺失了 ORF1、ORF2 或 ORF5、ORF6 的重组质粒

转化的转化子,保持了耐热性 DNA 聚合酶活性,而缺失了 ORF3 或 ORF4 则活性丧失。由此可以预测 ORF3 或 ORF4 编码 DNA 聚合酶活性。

ORF3、ORF4 究竟哪一个编码 DNA 聚合酶活性,可以通过将各 ORF 分别插入来制备重组质粒。然后调查所获得的转化子中的耐热性 DNA 聚合酶活性的表达来进行调查。令人意外的是,单独含有 ORF3 或 ORF4 的转化子获得的粗抽提物中仅检测到极弱的 DNA 聚合酶活性。而在将两种抽提物混合时,可以得到与含有 ORF3、ORF4 两者的转化子同样的耐热性 DNA 聚合酶活性,因此这显示,本发明的新型 DNA 聚合酶必须有这两种 ORF 的翻译产物起作用。是这两种蛋白质形成复合体后显示 DNA 聚合酶活性,还是其中一方修饰另一方转变成有活性的酶,可以通过测定 DNA 聚合酶的分子量来了解。用凝胶过滤法测定该 DNA 聚合酶分子量的结果表明,上述两种蛋白质形成复合体。

序列表的序列编号 2 中显示 ORF3 的碱基序列,序列编号 1 中显示由该碱基序列推导的 ORF3 的翻译产物,即第 1 种 DNA 聚合酶组成蛋白的氨基酸序列。而序列表的序列编号 4 中显示 ORF4 的碱基序列,序列编号 3 中显示由该碱基序列推导的 ORF4 的翻译产物,即第 2 种 DNA 聚合酶组成蛋白的氨基酸序列。

本发明的 DNA 聚合酶可以在常规的培养条件下,例如含有 100 $\mu\text{g/ml}$ 氨苄青霉素的 LB 培养基(蛋白胨 10 g/升、酵母抽提物 5g/升、NaCl 5g/升、pH7.2)中,培养用整合了 ORF3 及 ORF4 的重组质粒转化的转化子,如大肠杆菌 JM109/pFU1001,来在菌体进行表达。该聚合酶可以从上述培养的菌体中,采用诸如超声波处理、热处理、以及用阴离子交换柱(RESOURCE Q 柱,Pharmacia)、Heparin Sepharose 柱(HiTrap Heparin, Pharmacia 公司生产)、凝胶过滤柱(Superose 6HR, Pharmacia 公司生产)等进行色谱,纯化至在 SDS-聚丙烯酰胺凝胶电泳(SDS-PAGE)上基本上显示两种 DNA 聚合酶组成蛋白的两条带。而且,也可将单独含有 ORF3、ORF4 的转化子分别培养后,获得培养菌体,混合其抽提物、或纯化的 DNA 聚合酶组成蛋白来获得目的 DNA 聚合酶。在混合两种 DNA 聚合酶组成蛋白时,不需要特别的操作。可以将相应的两种转化子的抽提物或由其纯化而来的两种蛋白质简单地以适当量混合来获得有活性的 DNA 聚合酶。

由此获得的本发明的 DNA 聚合酶在 SDS-PAGE 上显示与约 90,000

道尔顿、约 140,000 道尔顿相当的两条带，分别与第 1 及第 2 种 DNA 聚合酶组成蛋白相对应。

如图 3 所示，本发明的 DNA 聚合酶在磷酸钾缓冲液中、75°C 时，显示最适 pH 在 pH6.5-7.0 附近。而且在不同的温度测定该酶的酶活
5 时，该酶在 75-85°C 显示最高活性，但由于在高温时，用作活性测定底物的活化 DNA 的双链结构会被破坏，该酶活性的准确最适温度没有测定。该 DNA 聚合酶有高热稳定性，图 4 显示在经过 80°C、30 分钟的热处理后，仍保持 80% 以上的残存活性。这种热稳定性使该酶能用于 PCR 方法。而且，在调查已知的 α 型聚合酶的特异性抑制剂蚜栖菌素 (aphidicolin) 的影响时，该 DNA 聚合酶的活性在 2 mM 蚜栖菌素存在下仍不被抑制。
10

分析纯化的 DNA 聚合酶的生物化学活性特性的结果为，本发明的 DNA 聚合酶在试管内有非常优秀的引物延伸活性，如表 1 所示，在用引物与单链 DNA 退火形成的底物 (M13-HT 引物) 测定 DNA 聚合酶的
15 活性时，能获得比通常活性测定所用的活化 DNA (DNAaseI 处理的小牛胸腺 DNA) 高的核苷酸掺入活性。与使用上述 M13-HT 引物底物的其他 DNA 聚合酶的引物延伸能力进行比较，本发明的 DNA 聚合酶能显示与已知的来自激烈热球菌的 DNA 聚合酶 (Pfu DNA 聚合酶, Stratagene 生产) 和来自水生栖热菌 (*Thermus aquaticus*) 的 Taq DNA 聚合酶
20 (TaKaRa Taq, 宝酒造社生产) 相比更优的延伸活性。更进一步，在反应体系中加入活化 DNA 作为竞争底物时，上述 2 种 DNA 聚合酶的引物延伸活性被强烈抑制，而本发明的 DNA 聚合酶只被轻微抑制，这表明该酶对引物延伸型底物具有高亲和性 (图 6)。

表 1

底物	相对活性		
	本发明的 DNA 聚合酶	P f u D N A 聚合酶	T a q D N A 聚合酶
活化 DNA	1 0 0	1 0 0	1 0 0
热变性 DNA	3 4 0	8 7	1 3 0
M13-HT 引物	1 7 0	2 3	9 0
M13-RNA 引物	5 2	0.4 9	3 8
poly dA-Oligo dT	9 4	3 9 0	2 9 0

poly A-Oligo dT	0.085		0.063
-----------------	-------	--	-------

而且,本发明的DNA聚合酶在用于PCR方法时,能显示极优的性能。PCR方法中频繁使用的、来自水生栖热菌的DNA聚合酶在单独使用时,扩增10千碱基对以上的DNA片段非常困难,在与其他DNA聚合酶组合使用时,也难以扩增20千碱基对以上的DNA片段《美国国家科学院院报》(Proceeding of the National Academy of Sciences of the USA,第91卷,第2216-2220页(1994))。而且,据报道,使用Pfu DNA聚合酶所能扩增的DNA链长最大为3千碱基对左右。在使用本发明的DNA聚合酶时,在不添加其他酶的情况下,能够单独扩增20千碱基对的DNA片段。

而且,本发明的DNA聚合酶还伴随有3'→5'外切核酸酶活性,相对于DNA聚合酶活性,外切核酸酶活性很强,已知这种活性强因而DNA合成的准确性非常高,在这点上,本发明的DNA聚合酶能与Pfu DNA聚合酶的活性相媲美(图5)。而且,经测定本发明的DNA聚合酶在DNA合成反应中引起的误差频率比Taq DNA聚合酶低。上述各种性质显示,本发明的DNA聚合酶作为PCR方法等遗传工程用试剂非常优秀。

而且,本发明的新型DNA聚合酶基因的发现还提供了更有趣的知识。发明人为了了解含有编码新型DNA聚合酶的ORF3、4的基因的区域在细胞内是怎样转录的,用Northern印迹法、RT PCR法以及引物延伸法对由激烈热球菌中制备的RNA部份进行了分析,结果证明,ORF1-ORF6以一条信使RNA(mRNA)从紧接ORF1的上游开始转录。由此可以设想ORF1和ORF2在细胞内的产生与ORF3、4受到同样的控制,同时考虑到ORF1、2、5、6与同酵母的DNA复制起始调节有关的CDC6、CDC18有序列上的同源性,本发明的新型DNA聚合酶很可能是DNA复制过程中重要的DNA聚合酶。而且激烈热球菌属于古细菌,据推测古细菌的DNA复制系统与真核生物非常接近,由此可以预测在真核生物中存在还没有被发现的、与本发明的DNA聚合酶类似的重要的DNA聚合酶。

而且,还可以预期与本发明的DNA聚合酶类似的耐热性DNA聚合酶,可在与激烈热球菌同属于极端嗜热性古细菌的细菌中,例如除激烈热球菌以外属于热球菌属的细菌、热网菌属(*Pyrodictium*)、热

球菌属 (*Thermococcus*)、葡萄嗜热菌属 (*staphylothermus*) 等属的细菌内产生。当这些酶与本发明的 DNA 聚合酶一样由 2 种 DNA 聚合酶组成蛋白组成时, 可以期待这些酶的一个 DNA 聚合酶组成蛋白与相当于另一个 DNA 聚合酶组成蛋白的本发明的 DNA 聚合酶组成蛋白组合, 能够表现同样的 DNA 聚合酶活性。

可以期待上述极端嗜热性古细菌生产的、与本发明的 DNA 聚合酶相似的耐热性 DNA 聚合酶在有关氨基酸序列、及编码其的基因的碱基序列上, 与本发明的 DNA 聚合酶的氨基酸序列、及编码其的基因的碱基序列之间存在同源性。因此, 以本发明中分离的基因或其碱基序列的一部分做探针, 用杂交的方法从上述嗜热性古细菌中获得 DNA 片段, 将其导入适当的微生物, 用前面记载的粘粒文库同样的方法制备热处理裂解液, 用适当的方式检测其 DNA 聚合酶活性, 就能够获得与本发明的 DNA 聚合酶的碱基序列不同, 但具有同样酶活性, 且与本发明的酶类似的耐热性 DNA 聚合酶的基因。

上述杂交可以按下列条件进行。即, 固定了 DNA 的膜与探针在含有 0.5% SDS、0.1% 小牛血清白蛋白 (BSA)、0.01% 聚乙烯吡咯烷酮、0.1% Ficol 400、0.01% 变性鲑精 DNA 的 6×SSC (1×SSC 表示 0.15 M NaCl、0.015 M 柠檬酸钠、pH7.0) 中, 50°C 温育 12-20 小时。温育终止后, 于含有 0.05% SDS 的 2×SSC 中、在 37°C 开始洗涤, 将 SSC 的浓度变化至 0.1×、温度变化至 50°C, 洗净至来源于固定 DNA 的信号与背景可以区别。

这样, 以本发明中分离的基因或其碱基序列的一部分作引物, 由上述嗜热性古细菌中获得的 DNA 为模板进行基因扩增, 将得到的片段或以该片段为探针进行杂交而从上述嗜热性古细菌中获得的片段导入适当的微生物, 按与前面一样的方法检测 DNA 聚合酶活性, 就能够获得与本发明的 DNA 聚合酶的活性不一样、但具有同一水平的、与本发明的酶类似的耐热性 DNA 聚合酶基因。

下面说明本发明的实施例, 但本发明并不限于下面的实施例。而且, 实施例中的 % 的意思是重量 %。

实施例 1

(1) 激烈热球菌基因组 DNA 的制备

激烈热球菌 DSM3638 的培养按下面的方法进行:

将组成为蛋白胨 1%、酵母抽提物 0.5%、可溶性淀粉 1%、Jarmarin S Solid (Jamarin 实验室生产) 3.5%、Jarmarin S Liquid (Jamarin 实验室生产) 0.5%、 MgSO_4 0.003%、 NaCl 0.001%、 $\text{FeSO}_4 \cdot 7\text{H}_2\text{O}$ 0.0001%、
5 CoSO_4 0.0001%、 $\text{CaCl}_2 \cdot 7\text{H}_2\text{O}$ 0.0001%、 ZnSO_4 0.0001%、 $\text{CuSO}_4 \cdot 5\text{H}_2\text{O}$ 0.1ppm、 $\text{KAl}(\text{SO}_4)_2$ 0.1ppm、 H_3BO_3 0.1ppm、 $\text{NaMoO}_4 \cdot 2\text{H}_2\text{O}$ 0.1ppm、 $\text{NiCl}_2 \cdot 6\text{H}_2\text{O}$ 0.25ppm 的培养基装入 2 升的培养瓶中, 120°C、20 分钟灭菌, 充入氮气以去除溶解氧后, 接种上述细菌, 在 95°C 静置培养 16 小时, 培养后, 离心收集菌体。

10 然后, 在 4 ml 含 25%蔗糖的 0.05 M Tris-HCl (pH8.0) 中悬浮收集的菌体, 在悬浮液中加入 0.8 ml 的溶菌酶[5 mg/ml、0.25 M Tris-HCl(pH8.0)]和 2 ml 的 0.2 M EDTA, 在 20°C 保温 1 小时。然后加入 24 ml 的 SET 溶液[150 mM NaCl、1 mM EDTA、20 mM Tris-HCl (pH8.0)], 再加入 5% 的 SDS 4 ml 和蛋白酶 K (10 mg/ml) 400
15 μl , 37°C 反应 1 小时。反应终止后, 用酚氯仿抽提, 接着用乙醇沉淀, 制备了约 3.2 mg 的基因组 DNA。

(2) 粘粒蛋白文库的制备

激烈热球菌 DSM3638 的基因组 DNA 400 μg 用 Sau 3A1 部分消化,
20 用密度梯度离心分离成 35-50 kb 大小的片段。接着, 1 μg 的三股螺旋粘粒载体 (Stratagene 公司生产) 用 XbaI 消化, 然后用碱性磷酸酶 (宝酒造社生产) 去磷酸化, 并用 BamHI 消化, 与上面分级的 30-50 kb 的 DNA 140 μg 混合, 进行连接, 使用 GIGAPACK GOLD (Stratagene 公司生产) 按体外包装法将激烈热球菌的基因组 DNA 片段包装入 λ 噬
25 菌体颗粒中, 制备文库。接着, 将所得到的文库的一部分转入大肠杆菌 DH5 α MCR 中。选择几个获得的转化子制备粘粒 DNA, 在证明有适当大小的插入片段后, 另从上述文库中选择 500 个转化子, 分别接种在 150 ml 含有 100 $\mu\text{g}/\text{ml}$ 氨苄青霉素的 LB 培养基 (蛋白胨 10 g/升、酵母抽提物 5g/升、NaCl 5g/升、pH7.2) 中培养。将该培养物离心,
30 收集的菌体悬浮于 1 ml 的 20 mM Tris-HCl pH8.0 中, 在 100°C 热处理 10 分钟。接着进行超声波处理, 然后再次在 100°C 热处理 10 分钟。离心后的上清即获得的裂解物作为粘粒的蛋白文库。

(3) DNA 聚合酶活性的测定

DNA 聚合酶活性测定用 DNAaseI 处理而活化的小牛胸腺 DNA (Worthington 生产) (活化 DNA) 作底物。DNA 的活化和 DNA 聚合酶活性的测定按 Harpar and Row 公司出版、D.R.Davis 编辑的《大肠杆菌的 DNA 聚合酶》(DNA polymerase from *Escherichia coli*) 第 263-276 页 (C.C.Richardson 著) 中记载的方法进行。

酶活性的测定按下面的方法进行。即, 制备 50 μ l 含有待测定活性样品的反应液 [20 mM Tris-盐酸 (pH7.7)、15 mM MgCl₂、2 mM 2-巯基乙醇、0.2 mg/ml 活化 DNA、40 μ M dATP、dCTP、dGTP、dTTP、60 nM [³H]dTTP (Amersham 公司生产)], 在 75°C 反应 15 分钟。将此反应液 40 μ l 点于 DE 滤纸 (Whatman 公司生产), 用 5% Na₂HPO₄ 洗涤 5 次, 然后用液闪计数器测定 DE 滤纸上残留的放射活性。按上述酶活性测定方法, 在 30 分钟内将 10 nmol 的 [³H]dTTP 掺入底物 DNA 的酶量定义为 1 个单位。

(4) 含有 DNA 聚合酶基因的粘粒文库的选择

制备含有 20 mM Tris-盐酸 (pH7.7)、2 mM MgCl₂、2 mM 2-巯基乙醇、0.2 mg/ml 活化 DNA、40 μ M dATP、dCTP、dGTP、dTTP、60 nM [³H]dTTP (Amersham 公司生产) 的反应液, 在 45 μ l 此反应液中加入分别来自粘粒文库的 5 个克隆的抽提物各 1 μ l, 即每个反应加入 5 μ l, 在 75°C 反应 15 分钟后。分别在 DE 滤纸 (Whatman 公司生产) 点样 40 μ l, 用 5% Na₂HPO₄ 洗涤 5 次, 然后用液闪计数器测定 DE 滤纸上残留的放射活性。在第一次检测中被确认含有活性的含有 5 个克隆的一组, 接着将 5 个克隆分成单个克隆进行第二次检测。粘粒文库中含有已知的 DNA 聚合酶的克隆, 这是以该基因为探针进行杂交而推测的, 它们分别为克隆 No. 57、154、162、363。除此之外, 还发现 5 个克隆 No. 41、153、264、462、491 含有 DNA 合成活性。

(5) 限制酶切图谱的制备

从上述 5 个克隆中分离粘粒, 分别用 BamHI 消化, 检测电泳图谱, 发现几条相同的带。预计 5 个克隆整合有重叠和轻微错动的区域。因

而制作克隆 No. 264 和 491 的插入 DNA 片段的限制酶切图谱。从两个克隆制备的粘粒用各种限制酶消化的结果是, KpnI、NotI、PstI、SmaI、XbaI、XhoI (均为宝酒造社生产) 都能切出适当大小的片段, 其相应的酶切位点确定的结果就获得了图 1 所示的图谱。

5

(6) DNA 聚合酶基因的亚克隆

在图 1 所示的限制酶切图谱的基础上, 由来自克隆 264 或 491 的粘粒切下 10 千碱基对左右的各种 DNA 片段, 亚克隆到 pTV118N 或 pTV119N 载体 (宝酒造社生产)。对来自所获重组质粒的转化子检测耐热性 DNA 聚合酶活性的结果, 证明约 10 千碱基对的 XbaI-XbaI 片段上有能产生强耐热性 DNA 聚合酶的基因。重组了该 XbaI-XbaI 片段的 pTV118N 载体被命名为 pFU1001, 用该质粒转化的大肠杆菌 JM109 被命名为 *Escherichia coli* JM109/pFU1001。

15 实施例 2

含有新型 DNA 聚合酶基因的 DNA 片段的碱基序列的测定

从实施例 1 中获得的质粒 pFU1001 中再次用 XbaI 将含有 DNA 聚合酶基因的上述 XbaI-XbaI 片段切出, 用 DNA 平末端化试剂盒 (宝酒造社生产) 平末端化, 然后与用 SmaI 线性化的新载体 pTV118N 连接, 来制备用于制备缺失突变体的质粒, 依插入片段方向的不同, 质粒被分别命名为 pFU1002 和 pFU1003。使用这两种质粒, 由插入 DNA 片段的两端开始, 顺次制备缺失突变体。制备应用 Henikoff 的方法 (《基因》(Gene) 第 28 卷、第 351-359 页), 并使用 kilo-sequence 用缺失试剂盒 (宝酒造社生产)。3'突出型和 5'突出型限制酶分别采用 PstI 和 XbaI。以得到的各种缺失突变体为模板, 用 BcaBE-ST 双脱氧测序试剂盒 (宝酒造社生产) 按双脱氧法确定插入 DNA 片段的碱基序列。

序列表的序列号 5 显示了所获得的碱基序列的 8450 碱基对的序列。对该碱基序列进行分析的结果发现了 6 个编码蛋白质的开放阅读框 (ORF), 分别存在于序列表的序列号 5 中所示的碱基序列的碱基号 123-614、(ORF1)、611-1381 (ORF2)、1384-3222 (ORF3)、3225-7013 (ORF4)、7068-7697 (ORF5)、7711-8385 (ORF6) 的位

置。图 2 中显示了重组在质粒 pFU1001 中的约 10 千碱基对的 XbaI-XbaI DNA 片段的限制酶切图谱和该片段上上述 ORF 的位置。

此外，用上述各种缺失突变体检测了 DNA 聚合酶活性，结果显示不管是从上游缺失还是从下游缺失，当 ORF3 和 ORF4 区域缺失时，DNA 聚合酶的活性都丧失。这显示 ORF3 和 ORF4 的翻译产物对 DNA 聚合酶活性的表达都非常重要。ORF3 的碱基序列在序列表的序列号 2 中显示，由该碱基序列推导的 ORF3 的翻译产物的氨基酸序列在序列号 1 中显示。ORF4 的碱基序列在序列表的序列号 4 中显示，由该碱基序列推导的 ORF4 的翻译产物的氨基酸序列在序列号 3 中显示。

10

实施例 3

纯化 DNA 聚合酶标准品的制备

将实施例 1 中获得的 *Escherichia coli* JM109/pFU1001 在含有浓度为 100 $\mu\text{g/ml}$ 的氯苄青霉素的 500 ml LB 培养基（蛋白胨 10 g/升、酵母抽提物 5g/升、NaCl 5g/升、pH7.2）中培养。当培养物的浊度达到 0.6 A_{600} 时，加入诱导物异丙基- β -D-硫代半乳糖苷 (IPTG)，培养 16 小时，收集菌体，然后用 37 ml 的超声缓冲液（50 mM Tris-HCl、pH8.0、2 mM 2-巯基乙醇、10%甘油、2.4 mM PMSF（苯甲基磺酰氟））悬浮菌体，用于超声波破碎仪。12000 rpm、10 分钟离心，回收 42 ml 作为粗抽提物的上清，然后在 80°C 热处理 15 分钟。然后再以 12000 rpm、10 分钟离心，获得 33 ml 的热处理酶液。接着将此溶液以 800 ml 缓冲液 A（50 mM 磷酸钾、pH6.5、2 mM 2-巯基乙醇、10%甘油）作外液透析 2 小时 $\times$ 4。透析后的酶液 32 ml 上用缓冲液 A 平衡的 RESOURCE Q 柱（Pharmacia 公司生产），使用 FPLC 系统（Pharmacia 公司生产）进行色谱。展开用 0 $\rightarrow$ 500 mM 线性浓度梯度的 NaCl 进行。DNA 聚合酶活性在 340 mM NaCl 时洗脱。

收集含活性的级分所得到的 10 ml 酶液用超滤进行脱盐、浓缩，溶于缓冲液 A+150 mM NaCl，得到 3.5 ml 酶液，然后将其上用同样的缓冲液平衡的 HiTrap Heparin 柱（Pharmacia 公司生产），用 FPLC 系统以 150 $\rightarrow$ 650 mM 线性浓度梯度的 NaCl 进行展开。在 400 mM NaCl 的洗脱级分中获得活性成分。将此级分 5 ml 用超滤反复脱盐、浓缩，使之最后浓缩为 120 μl 含 50 mM 磷酸钾、pH6.5、2 mM 2-巯基乙醇、

75 mM NaCl 的溶液。将此溶液上用同一缓冲液平衡的 Superose 6 凝胶过滤柱 (Pharmacia 公司生产), 并用同一缓冲液进行洗脱, 结果, DNA 聚合酶活性在保留时间为 34.7 分钟和 38.3 分钟的位置洗脱出来。与同一条件下分子量标准的洗脱位置进行比较的结果, 预计此活性具有的分子量分别为约 385 千道尔顿和约 220 千道尔顿, 这些分子量相当于 ORF3 的翻译产物与 ORF4 的翻译产物以 1: 2 的摩尔比形成的复合体以及以 1: 1 的摩尔比形成的复合体。由于在分子量时, 分子量测定的误差较大, 不能否定前者是 2 种翻译产物以 2: 2 的摩尔比形成复合体的可能性。

10

实施例 4

(1) DNA 聚合酶的生物化学特性

对于实施例 3 中获得的 ORF3 和 ORF4 的翻译产物, 即第 1 种 DNA 聚合酶组成蛋白和第 2 种 DNA 聚合酶组成蛋白以 1: 1 形成的复合体的 DNA 聚合酶标准品, 首先寻找其最适 $MgCl_2$ 和 KCl 浓度。在反应体系为 20 mM Tris-盐酸、pH7.7、2 mM 2-巯基乙醇、0.2 mg/ml 活化 DNA、40 μM dATP、dCTP、dGTP、dTTP, 并存在 2 mM $MgCl_2$ 时, KCl 的浓度在 0-200 mM 的范围内以 20 mM 进行增加, 测定 DNA 聚合酶活性。结果为在 KCl 为 60 mM 时显示最大活性。然后用上述反应体系, 并在 60 mM KCl 的条件下, 这次 $MgCl_2$ 在 0.5-25 mM 范围内每次增加 2.5 mM 进行比较, 结果在 10 mM 时显示最大活性, 在不含 KCl 的情况下进行比较时, 在 17.5 mM 时显示最大活性。

然后调查最适 pH。使用各种 pH 的磷酸钾缓冲液, 制备含有 20 mM 磷酸钾、15 mM $MgCl_2$ 、2 mM 2-巯基乙醇、0.2 mg/ml 活化 DNA、40 μM dATP、dCTP、dGTP、dTTP、60 nM [3H]dTTP 的反应液, 用来在 75°C 时测定 DNA 聚合酶活性。其结果在图 3 中显示。图中横轴表示 pH, 纵轴表示掺入高分子 DNA 的放射活性。图中显示本发明的 DNA 聚合酶在 pH6.5-7.0 时显示最大活性, 在用 Tris-盐酸取代磷酸钾的情况下, 在碱性一侧活性较高, 在测定所用最高的 pH8.02 时, 显示最大活性。

本发明的 DNA 聚合酶的热稳定性按以下的方法进行调查, 制备含有纯化 DNA 聚合酶的 20 mM Tris-盐酸 (pH7.7)、2 mM 2-巯基乙醇、10%甘油、0.1% 小牛血清白蛋白的溶液, 在各种温度下保温 30 分钟,

检测残存的 DNA 聚合酶活性。结果示于图 4。图中横轴表示保温温度，纵轴表示残存的活性。图中显示，本发明的酶在 80°C、30 分钟的热处理后，仍能保持 80%以上的残存活性。

为比较抑制剂的抑制模式，使用 α 型 DNA 聚合酶的特异性抑制剂
5 蛭蛭菌素，将本发明的 DNA 聚合酶与已知的来自激烈热球菌的 α 型 DNA 聚合酶（Pfu DNA 聚合酶，Stratagene 公司生产）的抑制方式进行比较。在 20 mM Tris-盐酸（pH7.7）、15 mM MgCl₂、2 mM 2-巯基乙醇、0.2 mg/ml 活化 DNA、40 μ M dATP、dCTP、dGTP、dTTP 存在下，蛭蛭菌素从 0 到 2.0 mM 增加，检测活性的变化，Pfu DNA 聚合酶在
10 1.0 mM 时活性减少至 20%，而与之相对，本发明的新型 DNA 聚合酶在 2.0 mM 时，也完全未被抑制。

（2）引物延伸反应

接着，为比较本发明的 DNA 聚合酶对底物 DNA 形式的选择性，检
15 测了下面的模板-引物。除常规的活性检测中使用的活化 DNA 外，制备的底物还包括：85°C 热处理 5 分钟的活化 DNA（热变性 DNA）、M13 噬菌体单链 DNA（M13mp18ss DNA，宝酒造社生产）以序列表的序列号 6 所示序列的 45 个碱基的合成脱氧核糖寡核苷酸为引物退火的产物（M13-HT 引物）、相同的 M13 噬菌体单链 DNA 以序列表的序列号 7
20 中所示序列的 17 个碱基的合成脱氧核糖寡核苷酸为引物退火的产物（M13-RNA 引物）、聚脱氧腺苷酸（poly dA，Pharmacia 公司生产）与寡聚胸苷酸（Oligo dT，Pharmacia 公司生产）以 20: 1 的摩尔比混合（polydA-Oligo dT）的产物、聚腺苷酸（poly A，Pharmacia 公司生产）与寡聚脱氧胸苷酸以 20: 1 的摩尔比混合的产物
25 （polyA-Oligo dT）。

用这些底物代替活化 DNA 测定 DNA 聚合酶活性，以底物用活化 DNA 时所获得的活性为 100，表 1 中显示用各底物时的相对活性。为了比较，对 Pfu DNA 聚合酶、来自水生栖热菌的 Taq DNA 聚合酶（TaKaRa Taq，宝酒造社生产）也进行了同样的检测。如表 1 所示，本发明的
30 新型 DNA 聚合酶与其他 DNA 聚合酶相比，在用 M13-HT 引物而不是以活化 DNA 为底物时，所得到的活性高，因而对引物延伸反应有特别高的适用性。

接着对引物延伸活性进行更详细的检测。使用 M13-HT 引物作底物，引物的 5'端用 $[\gamma\text{-}^{32}\text{P}]\text{ATP}$ (Amersham 公司生产) 和 T4 多聚核苷酸激酶 (宝酒造社生产) 标记。制备 10 μl 含终浓度为 0.05 $\mu\text{g}/\mu\text{l}$ 的上述底物和在以活化 DNA 为底物的活性测定中的酶量为 0.05 单位
5 的各种聚合酶的反应液 (20 mM Tris-盐酸 (pH7.7)、15 mM MgCl_2 、2 mM 2-巯基乙醇、270 μM dATP、d CTP、dGTP、dTTP)，在 75°C 反应 1、2、3、4 分钟，反应结束后，加入 2 μl 的反应终止液 (95% 甲醛、20 mM EDTA、0.05% 溴酚兰、0.05% 二甲苯青)，在 95°C 热变性处理 3 分钟，然后取 2 μl 反应液用含 8M 尿素的聚丙烯酰胺进行凝
10 胶电泳，然后制备放射性自显影。并且，为调查在作为竞争底物的活化 DNA 存在下的引物延伸活性，在上述反应液中加入活化 DNA 至终浓度为 0.4 $\mu\text{g}/\mu\text{l}$ ，按上面同样的方法制备放射性自显影，所得到的放射性自显影在图 6 中显示。

在图中，Pol、Pfu、Taq 分别表示本发明的 DNA 聚合酶、Pfu DNA
15 聚合酶、Taq DNA 聚合酶所获得的结果，1、2、3、4 分别表示反应时间 (分钟)。而且，图中的-、+分别代表在活化 DNA 不存在和存在下所获得的结果。而图左端用 G、A、T、C 表示的条带是用与上面相同的底物用双脱氧法终止反应所得到的反应物的电泳结果，用来推定延伸反应的长度。图中显示本发明的 DNA 聚合酶与 Pfu DNA 聚合酶、Taq
20 DNA 聚合酶相比显示更优的引物延伸活性。而且还显示，在活化 DNA 不存在时显示比较高的引物延伸活性的 Taq DNA 聚合酶在加入大量过剩的活化 DNA 时明显被抑制，而与之相对，本发明的 DNA 聚合酶不被活化 DNA 抑制。这证明，本发明的 DNA 聚合酶特别对引物与单链模板退火形成的引物延伸型底物具有高亲和性。

25

(3) 伴随外切核酸酶活性的有无

本发明的 DNA 聚合酶伴随的外切核酸酶活性用下述方法进行调
查。5'→3'外切核酸酶活性检测所用的底物是：pUC119 载体 (宝酒造社生产) 用 SspI (宝酒造社生产) 消化，用琼脂糖凝胶电泳分离所获
30 得的 386 碱基对的 DNA 片段，纯化，用 $[\gamma\text{-}^{32}\text{P}]\text{ATP}$ 和多聚核苷酸激酶制备 5'末端 ^{32}P 标记的 DNA 片段。3'→5'外切核酸酶活性检测所用的底物是：pUC119 载体 (宝酒造社生产) 用 Sau3AI 消化，所获得的 341

碱基对的 DNA 片段用琼脂糖凝胶电泳分离、纯化，用 $[\gamma\text{-}^{32}\text{P}]\text{CTP}$ (Amersham 公司生产) 和 Klenow 片段 (宝酒造社生产) 进行补平反应制备 3'末端 ^{32}P 标记的 DNA 片段。标记 DNA 用 NICK 柱 (Pharmacia 公司生产) 凝胶过滤进行纯化，并用于以下反应。在含有这种标记 DNA 1 ng 的反应液 (20 mM Tris-盐酸 (pH7.7)、15 mM MgCl_2 、2 mM 2-巯基乙醇) 中，加入 0.015 单位的 DNA 聚合酶，在 75°C 反应 2.5、5、7.5 分钟后，加入乙醇沉淀 DNA。上清中存在的放射性活性用液体闪烁计数器进行测定，求出外切核酸酶活性的分解量。在本发明的 DNA 聚合酶中没有发现 5'→3'外切核酸酶活性，而显示带有很强的 3'→5'外切核酸酶活性。已知 3'→5'外切核酸酶活性很强的 Pfu DNA 聚合酶也用同样的方法测定 3'→5'外切核酸酶活性，其结果一起在图 5 中显示。

图中横轴表示反应时间，纵轴表示上清中游离的放射性活性在全部反应液的放射性活性中的比例。而图中空心圈表示本发明的 DNA 聚合酶所测得的结果，实心圈表示 Pfu DNA 聚合酶所测得的结果。图中显示本发明的 DNA 聚合酶具有与 Pfu DNA 聚合酶相同程度的 3'→5'外切核酸酶活性，已知由于具有很强 3'→5'外切核酸酶活性的 Pfu DNA 聚合酶的 DNA 合成的忠实性很高。

20 (4) DNA 合成反应的正确性的比较

DNA 聚合酶在 DNA 合成反应中的正确性，以部分单链化的 pUC118 载体 (宝酒造社生产) (开环型双螺旋质粒) 为模板来检测。单链 pUC118 以大肠杆菌 MV1184 (宝酒造社生产) 为宿主，使用辅助噬菌体 M13K07 (宝酒造社生产)，按 1989 年冷泉港实验室出版、T. Maniatis 等编辑、《分子克隆：实验室指南第 2 版》4.44-4.48 中记载的方法，进行制备。而双链 DNA 用 PvuII (宝酒造社生产) 消化 pUC118 载体后，进行琼脂糖凝胶电泳，回收约 2.8 千碱基对的 DNA 片段，进行制备。

将上述单链 DNA 1 μg 和双链 DNA 2 μg 混合在 180 μl 的灭菌蒸馏水中，在 70°C 保温 10 分钟，然后加入 20 μl 的 20×SSC，在 60°C 静置 10 分钟。进行乙醇沉淀以回收 DNA，取一部分进行琼脂糖凝胶电泳，证明获得了开环型双链质粒。制备含 1/10 量的所获开环型双链 DNA 质粒的反应液 (10 mM Tris-盐酸 pH8.5、50 mM KCl、10 mM MgCl_2 、

1 mM dATP、d CTP、dGTP、dTTP) 30 μ l, 在 70°C 保温 3 分钟后, 加入 0.5 单位的 DNA 聚合酶, 在 70°C 进行 10 分钟的 DNA 合成反应。反应结束后, 用 10 μ l 的反应液转化大肠杆菌 DH5 α (BRL 公司生产)。在含有 100 μ g/ml 氨苄青霉素、0.1 mM IPTG、40 μ g/ml 的 5-溴-4-
5 氯-3-吲哚- β -D 半乳糖苷的 LB 平板上, 于 37°C 培养 18 小时。记录平板上出现的白色及青色克隆数。计算 DNA 合成时由于错误而产生的白色克隆的出现率。结果为, 用 Taq DNA 聚合酶作为 DNA 聚合酶时的出现率为 3.18%, 而相应地, 用本发明的 DNA 聚合酶的出现率为较低的数值 1.61%。

10

(5) PCR 中的应用

在本发明的 DNA 聚合酶在 PCR 反应中的性能与 Taq DNA 聚合酶进行比较时, 以 λ -DNA 为模板进行 PCR 反应。本发明的 DNA 聚合酶所用的反应液的组成包括 10 mM Tris-盐酸 (pH9.2)、3.5 mM MgCl₂、
15 75 mM KCl、400 μ M dATP、d CTP、dGTP、dTTP、0.01%牛血清白蛋白、0.1%Triton X-100, Taq DNA 聚合酶所用的反应液的组成包括 10 mM Tris-盐酸 (pH8.3)、1.5 mM MgCl₂、50 mM KCl、400 μ M dATP、d CTP、dGTP、dTTP。制备 50 μ l 含有 5.0 ng/50 μ l λ -DNA (宝酒造社生产)、10 pmol/50 μ l 引物 λ 1 和引物 λ 11、3.7 单位/50 μ l DNA
20 聚合酶的反应液。引物 λ 1 和引物 λ 11 的碱基序列分别表示在序列表的序列号 8 和序列号 9 中。上述反应液按每循环 98°C、10 秒~68°C、10 分钟进行 30 个循环的 PCR 反应, 然后取 5 μ l 反应液进行琼脂糖凝胶电泳, 用溴化乙啶染色来确认扩增的 DNA 片段。结果证明, 在使用 Taq DNA 聚合酶时没有发现 DNA 片段的扩增, 而用本发明的 DNA
25 聚合酶时, 有约 20 千碱基对的 DNA 片段的扩增。

接着, 将引物变为引物 λ 1 和引物 λ 10 进行实验。引物 λ 10 的碱基序列在序列表的序列号 10 中表示。按上面同样的反应液组成, 制备 25 μ l 含有 2.5 ng 的 λ -DNA、10 pmol 的引物 λ 1 和引物 λ 10、3.7 单位的 DNA 聚合酶的反应液。按上面同样的反应条件进行 5 个循环
30 的反应, 取 5 μ l 反应液进行琼脂糖凝胶电泳, 用溴化乙啶进行染色。结果证明, 在使用 Taq DNA 聚合酶时没有发现特异性扩增, 而用发明的 DNA 聚合酶扩增了约 15 千碱基对的 DNA 片段。

实施例 5

(1) 用于单独 ORF3 翻译产物表达的质粒的构建

以通过将实施例 2 中记载的质粒 pFU1002 (在该载体上 ORF1-6 位于载体上的 lac 启动子的下游) 中插入 DNA 片段中 ORF3 的紧邻下游部分缺失而制备的突变体质粒 6-82 为模板, 用引物 M4 (宝酒造社生产) 及碱基序列示于序列表的序列号 11 的引物 N03 进行 PCR 反应。并在 PCR 反应中使用 DNA 聚合酶的合成反应正确性较高的 Pfu DNA 聚合酶 (Stratagene 公司生产)。制备 100 μ l 含有 1 ng 模板 DNA、各 25 pmol 的两种引物和 2.5 单位的 Pfu DNA 聚合酶的 PCR 反应液 [20 mM Tris-盐酸、pH8.2、10 mM KCl、20 mM MgCl₂、6 mM (NH₄)₂SO₄、0.2 mM dATP、dCTP、dGTP、dTTP、1% Triton X-100、0.01% 小牛血清白蛋白], 以每循环 94°C、0.5 分钟~55°C、0.5 分钟~72°C、2 分钟进行 25 个循环的 PCR 反应。所扩增的约 2 千碱基对的 DNA 片段用 NcoI、SphI (均为宝酒造社生产) 消化, 然后整合到 pTV118N 载体 (宝酒造社生产) 的 NcoI-SphI 位点间, 制备质粒 pFU-ORF3。该质粒的插入 DNA 片段中含有可翻译的单独的 ORF3。

(2) 用于单独 ORF4 翻译产物表达的质粒的构建

以通过将上述质粒 pFU1002 中的插入 DNA 片段中 ORF4 中央部分的下游部分缺失而制备的突变体质粒 6-2 为模板, 用引物 M4 及碱基序列示于序列表的序列号 12 的引物 N04 进行 PCR 反应。除模板 DNA 变为质粒 6-2、引物 N03 变为 N04 外, 按上述实施例 5 (1) 相同的条件进行反应。所扩增的 DNA 片段用 NcoI、NheI (宝酒造社生产) 消化, 获得约 1.6 千碱基对的 DNA 片段, 将此片段与由上述质粒 pFU1002 中分离的约 3.3 千碱基对的 NheI-SalI 片段 (含有 ORF4 的后半部分) 一起重组到 pET15b 载体 (Novagen 公司生产) 的 NcoI-XhoI 位点间, 制备质粒 pFU-ORF4。该质粒的插入 DNA 片段中含有可翻译的单独的 ORF4。

(3) 由 ORF3、ORF4 的翻译产物重新组成 DNA 聚合酶

分别培养用上述质粒 pFU-ORF3 转化大肠杆菌 JM109 所得到的

Escherichia coli JM109/pFU-ORF3 及用上述质粒 pFU-ORF4 转化大肠杆菌 HMS174 所得到的 Escherichia coli HMS174/pFU-ORF4, 纯化菌体中表达的两种 ORF 的翻译产物。转化子的培养及粗抽提物的制备按实施例 3 中记载的方法进行。而两种翻译产物的纯化用 RESOURCE
5 Q、Hitrap Heparin、Superose 6 等柱进行, 并用 SDS-PAGE 上翻译产物的表现进行确定。已经确认, 虽然这样纯化的两种 ORF 翻译产物单独都不显示 DNA 聚合酶活性, 但将 2 者混合时, 能表现出耐热性 DNA 聚合酶活性。

工业上的利用可能性

10 本发明提供具有高的引物延伸活性和 3'→5'外切核酸酶活性的新型 DNA 聚合酶。该酶适用于 PCR 方法, 在遗传工程研究中非常有用。而且该酶可通过遗传工程的方法用编码本发明的 DNA 聚合酶的基因进行生产。

序列表

序列号: 1

序列长度: 613

序列类型: 氨基酸

5 链型: 单链

拓扑构型: 直链型

序列种类: 多肽

序列:

	Met Asp Glu Phe Val Lys Ser Leu Leu Lys Ala Asn Tyr Leu Ile
10	5 10 15
	Thr Pro Ser Ala Tyr Tyr Leu Leu Arg Glu Tyr Tyr Glu Lys Gly
	20 25 30
	Glu Phe Ser Ile Val Glu Leu Val Lys Phe Ala Arg Ser Arg Glu
15	35 40 45
	Ser Tyr Ile Ile Thr Asp Ala Leu Ala Thr Glu Phe Leu Lys Val
	50 55 60
	Lys Gly Leu Glu Pro Ile Leu Pro Val Glu Thr Lys Gly Gly Phe
20	65 70 75
	Val Ser Thr Gly Glu Ser Gln Lys Glu Gln Ser Tyr Glu Glu Ser
	80 85 90
	Phe Gly Thr Lys Glu Glu Ile Ser Gln Glu Ile Lys Glu Gly Glu
	95 100 105
25	Ser Phe Ile Ser Thr Gly Ser Glu Pro Leu Glu Glu Glu Leu Asn
	110 115 120
	Ser Ile Gly Ile Glu Glu Ile Gly Ala Asn Glu Glu Leu Val Ser
	125 130 135
30	

Asn Gly Asn Asp Asn Gly Gly Glu Ala Ile Val Phe Asp Lys Tyr		
140	145	150
Gly Tyr Pro Met Val Tyr Ala Pro Glu Glu Ile Glu Val Glu Glu		
155	160	165
Lys Glu Tyr Ser Lys Tyr Glu Asp Leu Thr Ile Pro Met Asn Pro		
170	175	180
Asp Phe Asn Tyr Val Glu Ile Lys Glu Asp Tyr Asp Val Val Phe		
185	190	195
Asp Val Arg Asn Val Lys Leu Lys Pro Pro Lys Val Lys Asn Gly		
200	205	210
Asn Gly Lys Glu Gly Glu Ile Ile Val Glu Ala Tyr Ala Ser Leu		
215	220	225
Phe Arg Ser Arg Leu Lys Lys Leu Arg Lys Ile Leu Arg Glu Asn		
230	235	240
Pro Glu Leu Asp Asn Val Val Asp Ile Gly Lys Leu Lys Tyr Val		
245	250	255
Lys Glu Asp Glu Thr Val Thr Ile Ile Gly Leu Val Asn Ser Lys		
260	265	270
Arg Glu Val Asn Lys Gly Leu Ile Phe Glu Ile Glu Asp Leu Thr		
275	280	285
Gly Lys Val Lys Val Phe Leu Pro Lys Asp Ser Glu Asp Tyr Arg		
290	295	300
Glu Ala Phe Lys Val Leu Pro Asp Ala Val Val Ala Phe Lys Gly		
305	310	315
Val Tyr Ser Lys Arg Gly Ile Leu Tyr Ala Asn Lys Phe Tyr Leu		
320	325	330

Pro Asp Val Pro Leu Tyr Arg Arg Gln Lys Pro Pro Leu Glu Glu			
	335	340	345
Lys Val Tyr Ala Ile Leu Ile Ser Asp Ile His Val Gly Ser Lys			
	350	355	360
Glu Phe Cys Glu Asn Ala Phe Ile Lys Phe Leu Glu Trp Leu Asn			
	365	370	375
Gly Asn Val Glu Thr Lys Glu Glu Glu Glu Ile Val Ser Arg Val			
	380	385	390
Lys Tyr Leu Ile Ile Ala Gly Asp Val Val Asp Gly Val Gly Val			
	395	400	405
Tyr Pro Gly Gln Tyr Ala Asp Leu Thr Ile Pro Asp Ile Phe Asp			
	410	415	420
Gln Tyr Glu Ala Leu Ala Asn Leu Leu Ser His Val Pro Lys His			
	425	430	435
Ile Thr Met Phe Ile Ala Pro Gly Asn His Asp Ala Ala Arg Gln			
	440	445	450
Ala Ile Pro Gln Pro Glu Phe Tyr Lys Glu Tyr Ala Lys Pro Ile			
	455	460	465
Tyr Lys Leu Lys Asn Ala Val Ile Ile Ser Asn Pro Ala Val Ile			
	470	475	480
Arg Leu His Gly Arg Asp Phe Leu Ile Ala His Gly Arg Gly Ile			
	485	490	495
Glu Asp Val Val Gly Ser Val Pro Gly Leu Thr His His Lys Pro			
	500	505	510
Gly Leu Pro Met Val Glu Leu Leu Lys Met Arg His Val Ala Pro			
	515	520	525

Met Phe Gly Gly Lys Val Pro Ile Ala Pro Asp Pro Glu Asp Leu
530 535 540
Leu Val Ile Glu Glu Val Pro Asp Val Val His Met Gly His Val
5 545 550 555
His Val Tyr Asp Ala Val Val Tyr Arg Gly Val Gln Leu Val Asn
560 565 570
Ser Ala Thr Trp Gln Ala Gln Thr Glu Phe Gln Lys Met Val Asn
10 575 580 585
Ile Val Pro Thr Pro Ala Lys Val Pro Val Val Asp Ile Asp Thr
590 595 600
Ala Lys Val Val Lys Val Leu Asp Phe Ser Gly Trp Cys
15 605 610

序列号: 2

序列长度: 1839

序列类型: 核酸

链型: 双链

20 拓扑构型: 直链型

序列种类: 基因组 DNA

序列:

ATGGATGAAT TTGTAAAATC ACTTCTAAAA GCTAACTATC TAATAACTCC CTCTGCCTAC 60
25 TATCTCTTGA GAGAATACTA TGAAAAAGGT GAATTCTCAA TTGTGGAGCT GGTAAAATTT 120
GCAAGATCAA GAGAGAGCTA CATAATTACT GATGCTTTAG CAACAGAATT CCTTAAAGTT 180
AAAGGCCTTG AACCAATTCT TCCAGTGGAA ACAAAGGGGG GTTTTGTTC CACTGGAGAG 240
TCCCAAAAAG AGCAGTCTTA TGAAGAGTCT TTTGGGACTA AAGAAGAAAT TTCCCAGGAG 300
30 ATTAAAGAAG GAGAGAGTTT TATTTCCACT GGAAGTGAAC CACTTGAAGA GGAGCTCAAT 360

```

AGCATTGGAA TTGAGGAAAT TGGGGCAAAT GAAGAGTTAG TTTCTAATGG AAATGACAAT 420
GGTGGAGAGG CAATTGTCTT TGACAAATAT GGCTATCCAA TGGTATATGC TCCAGAAGAA 480
ATAGAGGTTG AGGAGAAGGA GTACTCGAAG TATGAAGATC TGACAATACC CATGAACCCC 540
GACTTCAATT ATGTGGAAAT AAAGGAAGAT TATGATGTTG TCTTCGATGT TAGGAATGTA 600
AAGCTGAAGC CTCCTAAGGT AAAGAACGGT AATGGGAAGG AAGGTGAAAT AATTGTTGAA 660
GCTTATGCTT CTCTCTTCAG GAGTAGGTTG AAGAAGTTAA GGAAAATACT AAGGGAAAAT 720
CCTGAATTGG ACAATGTTGT TGATATTGGG AAGCTGAAGT ATGTGAAGGA AGATGAAACC 780
GTGACAATAA TAGGGCTTGT CAATTCCAAG AGGGAAGTGA ATAAAGGATT GATATTTGAA 840
ATAGAAGATC TCACAGGAAA GGTAAAGTT TTCTTGCCGA AAGATTCGGA AGATTATAGG 900
GAGGCATTTA AGGTTCTTCC AGATGCCGTC GTCGCTTTTA AGGGGGTGTA TTCAAAGAGG 960
GGAATTTTGT ACGCCAACAA GTTTTACCTT CCAGACGTTT CCCTCTATAG GAGACAAAAG 1020
CCTCCACTGG AAGAGAAAGT TTATGCTATT CTCATAAGTG ATATACACGT CGGAAGTAAA 1080
GAGTTCTGCG AAAATGCCTT CATAAAGTTC TTAGAGTGGC TCAATGGAAA CGTTGAAACT 1140
AAGGAAGAGG AAGAAATCGT GAGTAGGGTT AAGTATCTAA TCATTGCAGG AGATGTTGTT 1200
GATGGTGTG GCGTTTATCC GGGCCAGTAT GCCGACTTGA CGATTCCAGA TATATTGAC 1260
CAGTATGAGG CCCTCGCAA CTTTCTCTCT CACGTTCCTA AGCACATAAC AATGTTTATT 1320
GCCCCAGGAA ACCACGATGC TGCTAGGCAA GCTATTCCCC AACCAGAATT CTACAAAGAG 1380
TATGCAAAAC CTATATACAA GCTCAAGAAC GCCGTGATAA TAAGCAATCC TGCTGTAATA 1440
AGACTACATG GTAGGGACTT TCTGATAGCT CATGGTAGGG GGATAGAGGA TGTCGTTGGA 1500
AGTGTTCTTG GGTGACCCA TCACAAGCCC GGCCTCCCAA TGGTTGAACT ATTGAAGATG 1560
AGGCATGTAG CTCCAATGTT TGGAGGAAAG GTTCCAATAG CTCCTGATCC AGAAGATTTG 1620
CTTGTTATAG AAGAAGTTCC TGATGTAGTT CACATGGGTC ACGTTCACGT TTACGATGCG 1680
GTAGTTTATA GGGGAGTTCA GCTGGTTAAC TCCGCCACCT GGCAGGCTCA GACCGAGTTC 1740
CAGAAGATGG TGAACATAGT TCCAACGCCT GCAAAGGTTT CCGTTGTTGA TATTGATACT 1800
GCAAAAGTTG TCAAGGTTTT GGACTTTAGT GGGTGGTGC 1839

```

序列号: 3

序列长度: 1263

序列类型: 氨基酸

链型: 单链

5 拓扑构型: 直链型

序列种类: 多肽

序列:

	Met Glu Leu Pro Lys Glu Ile Glu Glu Tyr Phe Glu Met Leu Gln	
10	5 10 15	
	Arg Glu Ile Asp Lys Ala Tyr Glu Ile Ala Lys Lys Ala Arg Ser	
	20 25 30	
	Gln Gly Lys Asp Pro Ser Thr Asp Val Glu Ile Pro Gln Ala Thr	
15	35 40 45	
	Asp Met Ala Gly Arg Val Glu Ser Leu Val Gly Pro Pro Gly Val	
	50 55 60	
	Ala Gln Arg Ile Arg Glu Leu Leu Lys Glu Tyr Asp Lys Glu Ile	
	65 70 75	
20	Val Ala Leu Lys Ile Val Asp Glu Ile Ile Glu Gly Lys Phe Gly	
	80 85 90	
	Asp Phe Gly Ser Lys Glu Lys Tyr Ala Glu Gln Ala Val Arg Thr	
	95 100 105	
25	Ala Leu Ala Ile Leu Thr Glu Gly Ile Val Ser Ala Pro Leu Glu	
	110 115 120	
	Gly Ile Ala Asp Val Lys Ile Lys Arg Asn Thr Trp Ala Asp Asn	
	125 130 135	
30	Ser Glu Tyr Leu Ala Leu Tyr Tyr Ala Gly Pro Ile Arg Ser Ser	

140	145	150
Gly Gly Thr Ala Gln Ala Leu Ser Val Leu Val Gly Asp Tyr Val		
155	160	165
Arg Arg Lys Leu Gly Leu Asp Arg Phe Lys Pro Ser Gly Lys His		
170	175	180
Ile Glu Arg Met Val Glu Glu Val Asp Leu Tyr His Arg Ala Val		
185	190	195
Ser Arg Leu Gln Tyr His Pro Ser Pro Asp Glu Val Arg Leu Ala		
200	205	210
Met Arg Asn Ile Pro Ile Glu Ile Thr Gly Glu Ala Thr Asp Asp		
215	220	225
Val Glu Val Ser His Arg Asp Val Glu Gly Val Glu Thr Asn Gln		
230	235	240
Leu Arg Gly Gly Ala Ile Leu Val Leu Ala Glu Gly Val Leu Gln		
245	250	255
Lys Ala Lys Lys Leu Val Lys Tyr Ile Asp Lys Met Gly Ile Asp		
260	265	270
Gly Trp Glu Trp Leu Lys Glu Phe Val Glu Ala Lys Glu Lys Gly		
275	280	285
Glu Glu Ile Glu Glu Ser Glu Ser Lys Ala Glu Glu Ser Lys Val		
290	295	300
Glu Thr Arg Val Glu Val Glu Lys Gly Phe Tyr Tyr Lys Leu Tyr		
305	310	315
Glu Lys Phe Arg Ala Glu Ile Ala Pro Ser Glu Lys Tyr Ala Lys		
320	325	330
Glu Ile Ile Gly Gly Arg Pro Leu Phe Ala Gly Pro Ser Glu Asn		

335	340	345
Gly Gly Phe Arg Leu Arg Tyr Gly Arg Ser Arg Val Ser Gly Phe		
350	355	360
Ala Thr Trp Ser Ile Asn Pro Ala Thr Met Val Leu Val Asp Glu		
365	370	375
Phe Leu Ala Ile Gly Thr Gln Met Lys Thr Glu Arg Pro Gly Lys		
380	385	390
Gly Ala Val Val Thr Pro Ala Thr Thr Ala Glu Gly Pro Ile Val		
395	400	405
Lys Leu Lys Asp Gly Ser Val Val Arg Val Asp Asp Tyr Asn Leu		
410	415	420
Ala Leu Lys Ile Arg Asp Glu Val Glu Glu Ile Leu Tyr Leu Gly		
425	430	435
Asp Ala Ile Ile Ala Phe Gly Asp Phe Val Glu Asn Asn Gln Thr		
440	445	450
Leu Leu Pro Ala Asn Tyr Val Glu Glu Trp Trp Ile Gln Glu Phe		
455	460	465
Val Lys Ala Val Asn Glu Ala Tyr Glu Val Glu Leu Arg Pro Phe		
470	475	480
Glu Glu Asn Pro Arg Glu Ser Val Glu Glu Ala Ala Glu Tyr Leu		
485	490	495
Glu Val Asp Pro Glu Phe Leu Ala Lys Met Leu Tyr Asp Pro Leu		
500	505	510
Arg Val Lys Pro Pro Val Glu Leu Ala Ile His Phe Ser Glu Ile		
515	520	525
Leu Glu Ile Pro Leu His Pro Tyr Tyr Thr Leu Tyr Trp Asn Thr		

530	535	540
Val Asn Pro Lys Asp Val Glu Arg Leu Trp Gly Val Leu Lys Asp		
545	550	555
Lys Ala Thr Ile Glu Trp Gly Thr Phe Arg Gly Ile Lys Phe Ala		
560	565	570
Lys Lys Ile Glu Ile Ser Leu Asp Asp Leu Gly Ser Leu Lys Arg		
575	580	585
Thr Leu Glu Leu Leu Gly Leu Pro His Thr Val Arg Glu Gly Ile		
590	595	600
Val Val Val Asp Tyr Pro Trp Ser Ala Ala Leu Leu Thr Pro Leu		
605	610	615
Gly Asn Leu Glu Trp Glu Phe Lys Ala Lys Pro Phe Tyr Thr Val		
620	625	630
Ile Asp Ile Ile Asn Glu Asn Asn Gln Ile Lys Leu Arg Asp Arg		
635	640	645
Gly Ile Ser Trp Ile Gly Ala Arg Met Gly Arg Pro Glu Lys Ala		
650	655	660
Lys Glu Arg Lys Met Lys Pro Pro Val Gln Val Leu Phe Pro Ile		
665	670	675
Gly Leu Ala Gly Gly Ser Ser Arg Asp Ile Lys Lys Ala Ala Glu		
680	685	690
Glu Gly Lys Ile Ala Glu Val Glu Ile Ala Phe Phe Lys Cys Pro		
695	700	705
Lys Cys Gly His Val Gly Pro Glu Thr Leu Cys Pro Glu Cys Gly		
710	715	720
Ile Arg Lys Glu Leu Ile Trp Thr Cys Pro Lys Cys Gly Ala Glu		

725	730	735
Tyr Thr Asn Ser Gln Ala Glu Gly Tyr Ser Tyr Ser Cys Pro Lys		
740	745	750
Cys Asn Val Lys Leu Lys Pro Phe Thr Lys Arg Lys Ile Lys Pro		
755	760	765
Ser Glu Leu Leu Asn Arg Ala Met Glu Asn Val Lys Val Tyr Gly		
770	775	780
Val Asp Lys Leu Lys Gly Val Met Gly Met Thr Ser Gly Trp Lys		
785	790	795
Ile Ala Glu Pro Leu Glu Lys Gly Leu Leu Arg Ala Lys Asn Glu		
800	805	810
Val Tyr Val Phe Lys Asp Gly Thr Ile Arg Phe Asp Ala Thr Asp		
815	820	825
Ala Pro Ile Thr His Phe Arg Pro Arg Glu Ile Gly Val Ser Val		
830	835	840
Glu Lys Leu Arg Glu Leu Gly Tyr Thr His Asp Phe Glu Gly Lys		
845	850	855
Pro Leu Val Ser Glu Asp Gln Ile Val Glu Leu Lys Pro Gln Asp		
860	865	870
Val Ile Leu Ser Lys Glu Ala Gly Lys Tyr Leu Leu Arg Val Ala		
875	880	885
Arg Phe Val Asp Asp Leu Leu Glu Lys Phe Tyr Gly Leu Pro Arg		
890	895	900
Phe Tyr Asn Ala Glu Lys Met Glu Asp Leu Ile Gly His Leu Val		
905	910	915
Ile Gly Leu Ala Pro His Thr Ser Ala Gly Ile Val Gly Arg Ile		

920	925	930
Ile Gly Phe Val Asp Ala Leu Val Gly Tyr Ala His Pro Tyr Phe		
935	940	945
His Ala Ala Lys Arg Arg Asn Cys Asp Gly Asp Glu Asp Ser Val		
950	955	960
Met Leu Leu Leu Asp Ala Leu Leu Asn Phe Ser Arg Tyr Tyr Leu		
965	970	975
Pro Glu Lys Arg Gly Gly Lys Met Asp Ala Pro Leu Val Ile Thr		
980	985	990
Thr Arg Leu Asp Pro Arg Glu Val Asp Ser Glu Val His Asn Met		
995	1000	1005
Asp Val Val Arg Tyr Tyr Pro Leu Glu Phe Tyr Glu Ala Thr Tyr		
1010	1015	1020
Glu Leu Lys Ser Pro Lys Glu Leu Val Arg Val Ile Glu Gly Val		
1025	1030	1035
Glu Asp Arg Leu Gly Lys Pro Glu Met Tyr Tyr Gly Ile Lys Phe		
1040	1045	1050
Thr His Asp Thr Asp Asp Ile Ala Leu Gly Pro Lys Met Ser Leu		
1055	1060	1065
Tyr Lys Gln Leu Gly Asp Met Glu Glu Lys Val Lys Arg Gln Leu		
1070	1075	1080
Thr Leu Ala Glu Arg Ile Arg Ala Val Asp Gln His Tyr Val Ala		
1085	1090	1095
Glu Thr Ile Leu Asn Ser His Leu Ile Pro Asp Leu Arg Gly Asn		
1100	1105	1110
Leu Arg Ser Phe Thr Arg Gln Glu Phe Arg Cys Val Lys Cys Asn		

	1115	1120	1125
	Thr Lys Tyr Arg Arg Pro Pro Leu Asp Gly Lys Cys Pro Val Cys		
	1130	1135	1140
5	Gly Gly Lys Ile Val Leu Thr Val Ser Lys Gly Ala Ile Glu Lys		
	1145	1150	1155
	Tyr Leu Gly Thr Ala Lys Met Leu Val Ala Asn Tyr Asn Val Lys		
	1160	1165	1170
10	Pro Tyr Thr Arg Gln Arg Ile Cys Leu Thr Glu Lys Asp Ile Asp		
	1175	1180	1185
	Ser Leu Phe Glu Tyr Leu Phe Pro Glu Ala Gln Leu Thr Leu Ile		
	1190	1195	1200
15	Val Asp Pro Asn Asp Ile Cys Met Lys Met Ile Lys Glu Arg Thr		
	1205	1210	1215
	Gly Glu Thr Val Gln Gly Gly Leu Leu Glu Asn Phe Asn Ser Ser		
	1220	1225	1230
	Gly Asn Asn Gly Lys Lys Ile Glu Lys Lys Glu Lys Lys Ala Lys		
20	1235	1240	1245
	Glu Lys Pro Lys Lys Lys Lys Val Ile Ser Leu Asp Asp Phe Phe		
	1250	1255	1260

Ser Lys Arg

25 序列号: 4
 序列长度: 3789
 序列类型: 核酸
 链型: 双链
 拓扑构型: 直链型
 30 序列种类: 基因组 DNA
 序列:

ATGGAGCTTC CAAAGGAAAT TGAGGAGTAT TTTGAGATGC TTCAAAGGGA AATTGACAAA 60
 GCTTACGAGA TTGCTAAGAA GGCTAGGAGT CAGGGTAAAG ACCCCTCAAC CGATGTTGAG 120
 5 ATTCCCAGG CTACAGACAT GGCTGGAAGA GTTGAGAGCT TAGTTGGCCC TCCCGGAGTT 180
 GCTCAGAGAA TTAGGGAGCT TTTAAAAGAG TATGATAAGG AAATTGTTGC TTTAAAGATA 240
 GTTGATGAGA TAATTGAGGG CAAATTTGGT GATTTTGGAA GTAAAGAGAA GTACGCTGAA 300
 CAGGCTGTAA GGACAGCCTT GGCAATATTA ACTGAGGGTA TTGTTTCTGC TCCACTTGAG 360
 GGTATAGCTG ATGTTAAAAT CAAGCGAAAC ACCTGGGCTG ATAACCTCTGA ATACCTCGCC 420
 10 CTTTACTATG CTGGGCCAAT TAGGAGTTCT GGTGGAAGT CTCAAGCTCT CAGTGTACTT 480
 GTTGGTGATT ACGTTAGGCG AAAGCTTGGC CTTGATAGGT TTAAGCCAAG TGGGAAGCAT 540
 ATAGAGAGAA TGGTTGAGGA AGTTGACCTC TATCATAGAG CTGTTTCAAG GCTTCAATAT 600
 CATCCCTCAC CTGATGAAGT GAGATTAGCA ATGAGGAATA TTCCCATAGA AATCACTGGT 660
 15 GAAGCCACTG ACGATGTGGA GGTTTCCCAT AGAGATGTAG AGGGAGTTGA GACAAATCAG 720
 CTGAGAGGAG GAGCGATCCT AGTTTTGGCG GAGGGTGTTT TCCAGAAGGC TAAAAAGCTC 780
 GTGAAATACA TTGACAAGAT GGGGATTGAT GGATGGGAGT GGCTTAAAGA GTTTGTAGAG 840
 GCTAAAGAAA AAGGTGAAGA AATCGAAGAG AGTGAAAGTA AAGCCGAGGA GTCAAAAGTT 900
 20 GAAACAAGGG TGGAGGTAGA GAAGGGATTC TACTACAAGC TCTATGAGAA ATTTAGGCCT 960
 GAGATTGCCC CAAGCGAAAA GTATGCAAAG GAAATAATTG GTGGGAGGCC GTTATTCGCT 1020
 GGACCCTCGG AAAATGGGGG ATTTAGGCTT AGATATGGTA GAAGTAGGGT GAGTGGATTT 1080
 GCAACATGGA GCATAAATCC AGCAACAATG GTTTTGGTTG ACGAGTTCTT GGCCATTGGA 1140
 ACTCAAATGA AAACCGAGAG GCCTGGGAAA GGTGCAGTAG TGA CTCCAGC AACAAACGCT 1200
 25 GAAGGGCCGA TTGTTAAGCT AAAGGATGGG AGTGTGTGTA GGGTTGATGA TTACAACTTG 1260
 GCCCTCAAAA TAAGGGATGA AGTCGAAGAG ATACTTTATT TGGGAGATGC AATCATAGCC 1320
 TTTGGAGACT TTGTGGAGAA CAATCAAACCT CTCCTTCCTG CAACTATGT AGAGGAGTGG 1380

TGGATCCAAG AGTTCGTAAA GGCCGTTAAT GAGGCATATG AAGTTGAGCT TAGACCCTTT 1440
GAGGAAAATC CCAGGGAGAG CGTTGAGGAA GCAGCAGAGT ACCTTGAAGT TGACCCAGAA 1500
TTCTTGGCTA AGATGCTTTA CGATCCTCTA AGGGTTAAGC CTCCCGTGGA GCTAGCCATA 1560
CACTTCTCGG AAATCCTGGA AATTCTCTC CACCCATACT ACACCCTTTA TTGGAATACT 1620
GTAAATCCTA AAGATGTTGA AAGACTTTGG GGAGTATTAA AAGACAAGGC CACCATAGAA 1680
TGGGGCACTT TCAGAGGTAT AAAGTTTGCA AAGAAAATTG AAATTAGCCT GGACGACCTG 1740
GGAAGTCTTA AGAGAACCCT AGAGCTCCTG GGACTTCCTC ATACGGTAAG AGAAGGGATT 1800
GTAGTGGTTG ATTATCCGTG GAGTGCAGCT CTTCTCACTC CATTGGGCAA TCTTGAATGG 1860
GAGTTTAAGG CCAAGCCCTT CTACACTGTA ATAGACATCA TTAACGAGAA CAATCAGATA 1920
AAGCTCAGGG ACAGGGGAAT AAGCTGGATA GGGGCAAGAA TGGGAAGGCC AGAGAAGGCA 1980
AAAGAAAGAA AAATGAAGCC ACCTGTTCAA GTCCTCTTCC CAATTGGCTT GCCAGGGGGT 2040
TCTAGCAGAG ATATAAGAA GGCTGCTGAA GAGGGAAAAA TAGCTGAAGT TGAGATTGCT 2100
TTCTTCAAGT GTCCGAAGTG TGGCCATGTA GGGCCTGAAA CTCTCTGTCC CGAGTGTGGG 2160
ATTAGGAAAG AGTTGATATG GACATGTCCC AAGTGTGGGG CTGAATACAC CAATTCCCAG 2220
GCTGAGGGGT ACTCGTATTC ATGTCCAAAG TGCAATGTGA AGCTAAAGCC ATTCACAAAG 2280
AGGAAGATAA AGCCCTCAGA GCTCTTAAAC AGGGCCATGG AAAACGTGAA GGTTCATGGA 2340
GTTGACAAGC TTAAGGGCGT AATGGGAATG ACTTCTGGCT GGAAGATTGC AGAGCCGCTG 2400
GAGAAAGGTC TTTTGAGAGC AAAAAATGAA GTTTACGTCT TTAAGGATGG AACCATAAGA 2460
TTTGATGCCA CAGATGCTCC AATAACTCAC TTTAGGCCTA GGGAGATAGG AGTTTCAGTG 2520
GAAAAGCTGA GAGAGCTTGG CTACACCCAT GACTTCGAAG GGAAACCTCT GGTGAGTGAA 2580
GACCAGATAG TTGAGCTTAA GCCCAAGAT GTAATCCTCT CAAAGGAGGC TGGCAAGTAC 2640
CTCTTAAGAG TGGCCAGGTT TGTTGATGAT CTTCTTGAGA AGTTCTACGG ACTTCCCAGG 2700
TTCTACAACG CCGAAAAAT GGAGGATTTA ATTGGTCACC TAGTGATAGG ATTGGCCCTT 2760
CACACTTCAG CCGGAATCGT GGGGAGGATA ATAGGCTTTG TAGATGCTCT GGTGAGTAC 2820
GCTCACCCTT ACTTCCATGC GGCCAAGAGA AGGAACTGTG ATGGAGATGA GGATAGTGTA 2880
ATGCTACTCC TTGATGCCCT ATTGAACTTC TCCAGATACT ACCTCCCCGA AAAAAGAGGA 2940

5 GGAAAAATGG ACGCTCCTCT TGTCATAACC ACGAGGCTTG ATCCAAGAGA GGTGGACAGT 3000
GAAGTGCACA ACATGGATGT CGTTAGATAC TATCCATTAG AGTTCTATGA AGCAACTTAC 3060
GAGCTTAAAT CACCAAAGGA ACTTGTGAGA GTTATAGAGG GAGTTGAAGA TAGATTAGGA 3120
AAGCCTGAAA TGTATTACGG AATAAAGTTC ACCCAGGATA CCGACGACAT AGCTCTAGGA 3180
CCAAAGATGA GCCTCTACAA GCAGTTGGGA GATATGGAGG AGAAAGTGAA GAGGCAATTG 3240
ACATTGGCAG AGAGAATTAG AGCTGTGGAT CAACACTATG TTGCTGAAAC AATCCTCAAC 3300
10 TCCCCTTAA TTCCCGACTT GAGGGGTAAC CTAAGGAGCT TTACTAGACA AGAATTTTCGC 3360
TGTGTGAAGT GTAACACAAA GTACAGAAGG CCGCCCTTGG ATGGAAAATG CCCAGTCTGT 3420
GGAGGAAAGA TAGTGCTGAC AGTTAGCAAA GGAGCCATTG AAAAGTACTT GGGGACTGCC 3480
AAGATGCTCG TAGCTAACTA CAACGTAAAG CCATATACAA GGCAGAGAAT ATGCTTGACC 3540
15 GAGAAGGATA TTGATTCACT CTTTGAGTAC TTATTCCCAG AAGCCCAGTT AACGCTCATT 3600
GTAGATCCAA ACGACATCTG TATGAAAATG ATCAAGGAAA GAACGGGGGA AACAGTTCAA 3660
GGAGGCCTGC TTGAGAACTT TAATTCCTCT GGAAATAATG GGAAGAAAAT AGAGAAGAAG 3720
GAGAAAAAGG CAAAGGAAAA GCCTAAAAAG AAGAAAGTTA TAAGCTTGGA CGACTTCTTC 3780
20 TCCAAACGC 3789

序列号: 5

序列长度: 8450

序列类型: 核酸

25 链型: 双链

拓扑构型: 直链型

序列种类: 基因组 DNA

序列:

30 CATAACTAAA TTATTACATT TAGTTATATG GATGGGGGAA AAATTAACAA CATGTGTTAT 60
GTTTCCTCTG GAAAATTGAT CTATAATAAT CTAGGAGCAC AATTTCCAAT GGAGGGTCAT 120
CAATGAACGA AGGTGAACAT CAAATAAAGC TTGACGAGCT ATTCGAAAAG TTGCTCCGAG 180

CTAGGAAGAT ATTCAAAAAC AAAGATGTCC TTAGGCATAG CTATACTCCC AAGGATCTAC 240
CTCACAGACA TGAGCAAATA GAAACTCTCG CCCAAATTTT AGTACCAGTT CTCAGAGGAG 300
AAACTCCATC AAACATATTC GTTTATGGGA AGACTGGAAC TGGAAAGACT GTAAGTGTAA 360
AATTTGTAACT TGAAGAGCTG AAAAGAATAT CTGAAAAATA CAACATTCCA GTTGATGTGA 420
TCTACATTAA TTGTGAGATT GTCGATACTC ACTATAGAGT TCTTGCTAAC ATAGTTAACT 480
ACTTCAAAGA TGAGACTGGG ATTGAAGTTC CAATGGTAGG TTGGCCTACC GATGAAGTTT 540
ACGCAAAGCT TAAGCAGGTT ATAGATATGA AGGAGAGGTT TGTGATAATT GTGTTGGATG 600
AAATTGACAA GTGGTAAAGA AGAGTGGTGA TGAGGTTCTC TATTCATTAA CAAGAATAAA 660
TACTGAACTT AAAAGGGCTA AAGTGAGTGT AATTGGTATA TCAAACGACC TTAAATTTAA 720
AGAGTATCTA GATCCAAGAG TTCTCTCAAG TTTGAGTGAG GAAGAGGTGG TATTCCCACC 780
CTATGATGCA AATCAGCTTA GGGATATACT GACCCAAAGA GCTGAAGAGG CCTTTTATCC 840
TGGGGTTTTTA GACGAAGGTG TGATTCCCCT CTGTGCAGCA TTAGCTGCTA GAGAGCATGG 900
AGATGCAAGA AAGGCACTTG ACCTTCTAAG AGTTGCAGGG GAAATAGCGG AAAGAGAAGG 960
GGCAAGTAAA GTAAGTAAA AGCATGTTTG GAAAGCCCAG GAAAAGATTG AACAGGACAT 1020
GATGGAGGAG GTAATAAAAA CTCTACCCCT TCAGTCAAAA GTTCTCCTCT ATGCCATAGT 1080
TCTTTTGGAC GAAAACGGCG ATTTACCAGC AAATACTGGG GATGTTTACG CTGTTTATAG 1140
GGAATTGTGC GAGTACATTG ACTTGGAACC TCTCACCCTA AGAAGGATAA GTGATCTAAT 1200
TAATGAGCTT GACATGCTTG GAATAATAAA TGCAAAAGTT GTTAGTAAGG GGAGATATGG 1260
GAGGACAAAG GAAATAAGGC TTAACGTTAC CTCATATAAG ATAAGAAATG TGCTGAGATA 1320
TGATTACTCT ATTCAGCCCC TCCTCACAAT TTCCCTTAAG AGTGAGCAGA GGAGGTTGAT 1380
CTAATGGATG AATTTGTAAA ATCACTTCTA AAAGCTAACT ATCTAATAAC TCCCTCTGCC 1440
TACTATCTCT TGAGAGAATA CTATGAAAAA GGTGAATTCT CAATTGTGGA GCTGGTAAAA 1500
TTTGCAAGAT CAAGAGAGAG CTACATAATT ACTGATGCTT TAGCAACAGA ATTCCTTAAA 1560
GTAAAGGCC TTGAACCAAT TCTTCCAGTG GAAACAAAGG GGGGTTTTGT TTCCACTGGA 1620
GAGTCCCAA AAGAGCAGTC TTATGAAGAG TCTTTTGGGA CTAAAGAAGA AATTTCCAG 1680
GAGATTAAAG AAGGAGAGAG TTTTATTTCC ACTGGAAGTG AACCCTTGA AGAGGAGCTC 1740

AATAGCATTG GAATTGAGGA AATTGGGGCA AATGAAGAGT TAGTTTCTAA TGGAAATGAC 1800
AATGGTGGAG AGGCAATTGT CTTTGACAAA TATGGCTATC CAATGGTATA TGCTCCAGAA 1860
GAAATAGAGG TTGAGGAGAA GGAGTACTCG AAGTATGAAG ATCTGACAAT ACCCATGAAC 1920
CCCGACTTCA ATTATGTGGA AATAAAGGAA GATTATGATG TTGTCTTCGA TGTTAGGAAT 1980
GTAAAGCTGA AGCCTCCTAA GGTAAAGAAC GGTAATGGGA AGGAAGGTGA AATAATTGTT 2040
GAAGCTTATG CTTCTCTCTT CAGGAGTAGG TTGAAGAACT TAAGGAAAAT ACTAAGGGAA 2100
AATCCTGAAT TGGACAATGT TGTTGATATT GGAAGCTGA AGTATGTGAA GGAAGATGAA 2160
ACCGTGACAA TAATAGGGCT TGTCAATTCC AAGAGGGAAG TGAATAAAGG ATTGATATTT 2220
GAAATAGAAG ATCTCACAGG AAAGGTAAA GTTTTCTTGC CGAAAGATTC GGAAGATTAT 2280
AGGGAGGCAT TTAAGGTTCT TCCAGATGCC GTCGTCGCTT TTAAGGGGGT GTATTCAAAG 2340
AGGGGAATTT TGTACGCCAA CAAGTTTTAC CTTCCAGACG TTCCCCTCTA TAGGAGACAA 2400
AAGCCTCCAC TGAAGAGAA AGTTTATGCT ATTCTCATAA GTGATATACA CGTCGGAAGT 2460
AAAGAGTTCT GCGAAAATGC CTTCATAAAG TTCTTAGAGT GGCTCAATGG AAACGTTGAA 2520
ACTAAGGAAG AGGAAGAAAT CGTGAGTAGG GTTAAGTATC TAATCATTGC AGGAGATGTT 2580
GTTGATGGTG TTGGCGTTTA TCCGGGCCAG TATGCCGACT TGACGATTCC AGATATATTC 2640
GACCAGTATG AGGCCCTCGC AAACCTTCTC TCTCACGTTT CTAAGCACAT AACAATGTTT 2700
ATTGCCCCAG GAAACCACGA TGCTGCTAGG CAAGCTATTC CCCAACCAGA ATTCTACAAA 2760
GAGTATGCAA AACCTATATA CAAGCTCAAG AACGCCGTA TAATAAGCAA TCCTGCTGTA 2820
ATAAGACTAC ATGGTAGGGA CTTTCTGATA GCTCATGGTA GGGGGATAGA GGATGTCGTT 2880
GGAAGTGTTT CTGGGTTGAC CCATCACAAG CCCGGCCTCC CAATGGTTGA ACTATTGAAG 2940
ATGAGGCATG TAGCTCCAAT GTTTGGAGGA AAGGTTCCAA TAGCTCCTGA TCCAGAAGAT 3000
TTGCTTGTTA TAGAAGAAGT TCCTGATGTA GTTCACATGG GTCACGTTCA CGTTTACGAT 3060
GCGGTAGTTT ATAGGGGAGT TCAGCTGGTT AACTCCGCCA CCTGGCAGGC TCAGACCGAG 3120
TTCCAGAAGA TGGTGAACAT AGTTCCAACG CCTGCAAAGG TTCCCGTTGT TGATATTGAT 3180
ACTGCAAAG TTGTCAAGGT TTTGGACTTT AGTGGGTGGT GCTGATGGAG CTTCCAAAGG 3240
AAATTGAGGA GTATTTTGAG ATGCTTCAA GGGAAATTGA CAAAGCTTAC GAGATTGCTA 3300

AGAAGGCTAG GAGTCAGGGT AAAGACCCCT CAACCGATGT TGAGATTCCC CAGGCTACAG 3360
ACATGGCTGG AAGAGTTGAG AGCTTAGTTG GCCCTCCCGG AGTTGCTCAG AGAATTAGGG 3420
AGCTTTTAAA AGAGTATGAT AAGGAAATTG TTGCTTTAAA GATAGTTGAT GAGATAATTG 3480
AGGGCAAATT TGGTGATTTT GGAAGTAAAG AGAAGTACGC TGAACAGGCT GTAAGGACAG 3540
CCTTGGCAAT ATTAAGTGA GGTATTGTTT CTGCTCCACT TGAGGGTATA GCTGATGTTA 3600
AAATCAAGCG AAACACCTGG GCTGATAACT CTGAATACCT CGCCCTTTAC TATGCTGGGC 3660
CAATTAGGAG TTCTGGTGGA ACTGCTCAAG CTCTCAGTGT ACTTGTGGT GATTACGTTA 3720
GGCGAAAGCT TGGCCTTGAT AGGTTTAAGC CAAGTGGGAA GCATATAGAG AGAATGGTTG 3780
AGGAAGTTGA CCTCTATCAT AGAGCTGTTT CAAGGCTTCA ATATCATCCC TCACCTGATG 3840
AAGTGAGATT AGCAATGAGG AATATTCCCA TAGAAATCAC TGGTGAAGCC ACTGACGATG 3900
TGGAGGTTTC CCATAGAGAT GTAGAGGGAG TTGAGACAAA TCAGCTGAGA GGAGGAGCGA 3960
TCCTAGTTTT GGCGGAGGGT GTTCTCCAGA AGGCTAAAAA GCTCGTGAAA TACATTGACA 4020
AGATGGGGAT TGATGGATGG GAGTGGCTTA AAGAGTTTGT AGAGGCTAAA GAAAAAGGTG 4080
AAGAAATCGA AGAGAGTGAA AGTAAAGCCG AGGAGTCAAA AGTTGAAACA AGGGTGGAGG 4140
TAGAGAAGGG ATTCTACTAC AAGCTCTATG AGAAATTTAG GGCTGAGATT GCCCCAAGCG 4200
AAAAGTATGC AAAGGAAATA ATTGGTGGGA GGCCGTTATT CGCTGGACCC TCGGAAAATG 4260
GGGGATTTAG GCTTAGATAT GGTAGAAGTA GGGTGAGTGG ATTTGCAACA TGGAGCATAA 4320
ATCCAGCAAC AATGGTTTTG GTTGACGAGT TCTTGGCCAT TGGAAC TCA ATGAAAACCG 4380
AGAGGCCTGG GAAAGGTGCA GTAGTACTC CAGCAACAAC CGCTGAAGGG CCGATTGTTA 4440
AGCTAAAGGA TGGGAGTGTT GTTAGGGTTG ATGATTACAA CTTGGCCCTC AAAATAAGGG 4500
ATGAAGTCGA AGAGATACTT TATTTGGGAG ATGCAATCAT AGCCTTTGGA GACTTTGTGG 4560
AGAACAATCA AACTCTCCTT CCTGCAAACT ATGTAGAGGA GTGGTGGATC CAAGAGTTCTG 4620
TAAAGGCCGT TAATGAGGCA TATGAAGTTG AGCTTAGACC CTTTGAGGAA AATCCCAGGG 4680
AGAGCGTTGA GGAAGCAGCA GAGTACCTTG AAGTTGACCC AGAATTCTTG GCTAAGATGC 4740
TTTACGATCC TCTAAGGGTT AAGCCTCCCG TGGAGCTAGC CATACACTTC TCGGAAATCC 4800
TGGAAATTCC TCTCCACCCA TACTACACCC TTTATTGGAA TACTGTAAAT CCTAAAGATG 4860

TTGAAAGACT TTGGGGAGTA TTAAAAGACA AGGCCACCAT AGAATGGGGC ACTTTCAGAG 4920
GTATAAAGTT TGCAAAGAAA ATTGAAATTA GCCTGGACGA CCTGGGAAGT CTTAAGAGAA 4980
CCCTAGAGCT CCTGGGACTT CCTCATACGG TAAGAGAAGG GATTGTAGTG GTTGATTATC 5040
CGTGGAGTGC AGCTCTTCTC ACTCCATTGG GCAATCTTGA ATGGGAGTTT AAGGCCAAGC 5100
CCTTCTACAC TGTAATAGAC ATCATTAACG AGAACAATCA GATAAAGCTC AGGGACAGGG 5160
GAATAAGCTG GATAGGGGCA AGAATGGGAA GGCCAGAGAA GGCAAAAGAA AGAAAAATGA 5220
AGCCACCTGT TCAAGTCCTC TTCCAATTG GCTTGGCAGG GGGTTCTAGC AGAGATATAA 5280
AGAAGGCTGC TGAAGAGGGA AAAATAGCTG AAGTTGAGAT TGCTTTCTTC AAGTGTCGA 5340
AGTGTGGCCA TGTAGGGCCT GAAACTCTCT GTCCCGAGTG TGGGATTAGG AAAGAGTTGA 5400
TATGGACATG TCCAAGTGT GGGGCTGAAT ACACCAATTC CCAGGCTGAG GGGTACTCGT 5460
ATTCATGTCC AAAGTGCAAT GTGAAGCTAA AGCCATTAC AAAGAGGAAG ATAAAGCCCT 5520
CAGAGCTCTT AAACAGGGCC ATGGAAAACG TGAAGGTTTA TGGAGTTGAC AAGCTTAAGG 5580
GCGTAATGGG AATGACTTCT GGCTGGAAGA TTGCAGAGCC GCTGGAGAAA GGTCTTTTGA 5640
GAGCAAAAAA TGAAGTTTAC GTCTTTAAGG ATGGAACCAT AAGATTTGAT GCCACAGATG 5700
CTCCAATAAC TCACTTTAGG CCTAGGGAGA TAGGAGTTTC AGTGGAAAAG CTGAGAGAGC 5760
TTGGCTACAC CCATGACTTC GAAGGGAAAC CTCTGGTGAG TGAAGACCAG ATAGTTGAGC 5820
TTAAGCCCCA AGATGTAATC CTCTCAAAGG AGGCTGGCAA GTACCTCTTA AGAGTGGCCA 5880
GGTTTGTGTA TGATCTTCTT GAGAAGTTCT ACGGACTTCC CAGGTTCTAC AACGCCGAAA 5940
AAATGGAGGA TTTAATTGGT CACCTAGTGA TAGGATTGGC CCCTCACACT TCAGCCGGAA 6000
TCGTGGGGAG GATAATAGGC TTTGTAGATG CTCTGGTTGG CTACGCTCAC CCCTACTTCC 6060
ATGCGGCCAA GAGAAGGAAC TGTGATGGAG ATGAGGATAG TGTAATGCTA CTCCTTGATG 6120
CCCTATTGAA CTTCTCCAGA TACTACCTCC CCGAAAAAAG AGGAGGAAAA ATGGACGCTC 6180
CTCTTGTCAT AACCACGAGG CTTGATCCAA GAGAGGTGGA CAGTGAAGTG CACAACATGG 6240
ATGTCGTTAG ATACTATCCA TTAGAGTTCT ATGAAGCAAC TTACGAGCTT AAATCACCAA 6300
AGGAACTTGT GAGAGTTATA GAGGGAGTTG AAGATAGATT AGGAAAGCCT GAAATGTATT 6360
ACGGAATAAA GTTCACCCAC GATACCGACG ACATAGCTCT AGGACCAAAG ATGAGCCTCT 6420

ACAAGCAGTT GGGAGATATG GAGGAGAAAG TGAAGAGGCA ATTGACATTG GCAGAGAGAA 6480
TTAGAGCTGT GGATCAACAC TATGTTGCTG AAACAATCCT CAACTCCCAC TTAATTCCCG 6540
ACTTGAGGGG TAACCTAAGG AGCTTTACTA GACAAGAATT TCGCTGTGTG AAGTGTAACA 6600
CAAAGTACAG AAGGCCGCCC TTGGATGGAA AATGCCCAGT CTGTGGAGGA AAGATAGTGC 6660
TGACAGTTAG CAAAGGAGCC ATTGAAAAGT ACTTGGGGAC TGCCAAGATG CTCGTAGCTA 6720
ACTACAACGT AAAGCCATAT ACAAGGCAGA GAATATGCTT GACGGAGAAG GATATTGATT 6780
CACTCTTTGA GTACTTATTC CCAGAAGCCC AGTTAACGCT CATTGTAGAT CCAAACGACA 6840
TCTGTATGAA AATGATCAAG GAAAGAACGG GGGAAACAGT TCAAGGAGGC CTGCTTGAGA 6900
ACTTTAATTC CTCTGGAAAT AATGGGAAGA AAATAGAGAA GAAGGAGAAA AAGGCAAAGG 6960
AAAAGCCTAA AAAGAAGAAA GTTATAAGCT TGGACGACTT CTTCTCCAAA CGCTGACCAC 7020
AACTTTTAAG TTCTTTCTTG AGAATAAATT CCCAGGTGGC TTAGAGAATG AAGATTGTGT 7080
GGTGTGGTCA TGCCTGCTTC TTGGTGGAGG ATAGGGGGAC TAAGATACTA ATCGATCCAT 7140
ACCCAGACGT TGATGAAGAC AGAATAGGCA AGGTCGATTA CATTCTAGTT ACCCACCAGC 7200
ACATGGATCA CTACGGTAAG ACCCCACTAA TAGCAAAGCT CAGTGATGCC GAGGTTATAG 7260
GGCCGAAAAC AGTTTATCTC ATGGCAATAA GTGATGGGCT AACAAAGGTC AGAGAGATAG 7320
AGGTGGGACA GGAAATCGAG CTGGGAGATA TTAGGGTTAG GGCATTTTTT ACAGAGCATC 7380
CAACAAGCCA GTATCCCCTG GGATATCTAA TTGAAGGAAG CAAAAGAGTG GCTCACTTGG 7440
GAGATACATA CTACAGTCCA GCTTTTACAG AGTTGAGGGG AAAGGTTGAT GTTCTTTTGG 7500
TTCCAATAGG TGGGAAGTCC ACCGCTAGTG TAAGGGAGGC TGCGGATATA GTGGAGATGA 7560
TAAGGCCCCAG GATAGCAGTT CCAATGCACT ATGGAACGTA CAGCGAGGCC GATCCTGAAG 7620
AGTTCAAGAA GGAGCTCCAA AAAAGGCCCA TATGGGTTTT AGTAAAGGAT CTTAAGCCCT 7680
ATGAGGGTTT TGAAATCTGA AGGTGTTTCA ATGCTAAATA CTGAGCTCTT AACCCTGGA 7740
GTCAAGGGGT TAGATGAGCT TTTAGGTGGT GGAGTTGCTA AGGGAGTAAT ACTCCAAGTT 7800
TACGGGCCAT TTGCCACCGG GAAGACAAC TTTGCAATGC AGGTTGGATT ATTGAATGAG 7860
GGAAAAGTGG CTTATGTTGA TACTGAGGGG GGATTCTCCC CCGAAAGGTT AGCTCAAATG 7920
GCAGAATCAA GGAACCTTGA TGTGGAGAAA GCACTTGAAA AGTTCGTGAT ATTCGAACCT 7980

ATGGATTTAA ACGAGCAAAG ACAGGTAATT GCGAGGTTGA AAAATATCGT GAATGAAAAG 8040
TTTTCTTTAG TTGTGGTCGA CTCCTTTACG GCCCATTATA GAGCGGAGGG GAGTAGAGAG 8100
5 TATGGAGAAC TTTCCAAGCA ACTCCAAGTT CTTCAGTGGA TTGCCAGAAG AAAAAACGTT 8160
GCCGTTATAG TTGTCAATCA AGTTTATTAC GATTCAAACCT CAGGAATTCT TAAACCAATA 8220
GCTGAGCACA CCCTGGGGTA CAAAACAAAG GACATCCTCC GCTTTGAAAG GCTTAGGGTT 8280
GGAGTGAGAA TTGCAGTTCT GGAAAGGCAT AGGTTTAGGC CAGAGGGTGG GATGGTATAC 8340
10 TTCAAATAA CAGATAAAGG ATTGGAGGAT GTAAAAACG AAGATTAGAG CCTGTCGTAG 8400
ACCTCCTGGG CAATCCTCAG CGTTGCCTTA TAGAGCTTCT CACTAATAAT 8450

序列号: 6

序列长度: 45

15 序列类型: 核酸

链型: 单链

拓扑构型: 直链型

序列种类: 其他核酸 (合成 DNA)

序列:

20 CCGGAACCGC CTCCTCAGA GCGGCCACCC TCAGAACCGC CACCC 45

序列号: 7

序列长度: 17

序列类型: 核酸

25 链型: 单链

拓扑构型: 直链型

序列种类: 其他核酸 (合成 RNA)

序列:

30 GUUUUCCCAG UCACGAC 17

序列号: 8

序列长度: 23

	序列类型: 核酸	
	链型: 单链	
	拓扑构型: 直链型	
	序列种类: 其他核酸 (合成 DNA)	
5	序列:	
	GATGAGTTCG TGTCCGTACA ACT	23
	序列号: 9	
	序列长度: 22	
10	序列类型: 核酸	
	链型: 单链	
	拓扑构型: 直链型	
	序列种类: 其他核酸 (合成 DNA)	
	序列:	
15	ACAAAGCCAG CCGGAATATC TG	22
	序列号: 10	
	序列长度: 22	
	序列类型: 核酸	
20	链型: 单链	
	拓扑构型: 直链型	
	序列种类: 其他核酸 (合成 DNA)	
	序列:	
25	TACAATACGA TGCCCCGTTA AG	22
	序列号: 11	
	序列长度: 32	
	序列类型: 核酸	
	链型: 单链	
30	拓扑构型: 直链型	
	序列种类: 其他核酸 (合成 DNA)	
	序列:	

CAGAGGAGGT TGATCCCATG GATGAATTTG TA

32

序列号: 12

序列长度: 32

5 序列类型: 核酸

链型: 单链

拓扑构型: 直链型

序列种类: 其他核酸 (合成 DNA)

序列:

10 TTTAGTGGGT GGTGCCCATG GAGCTTCCAA AG

32

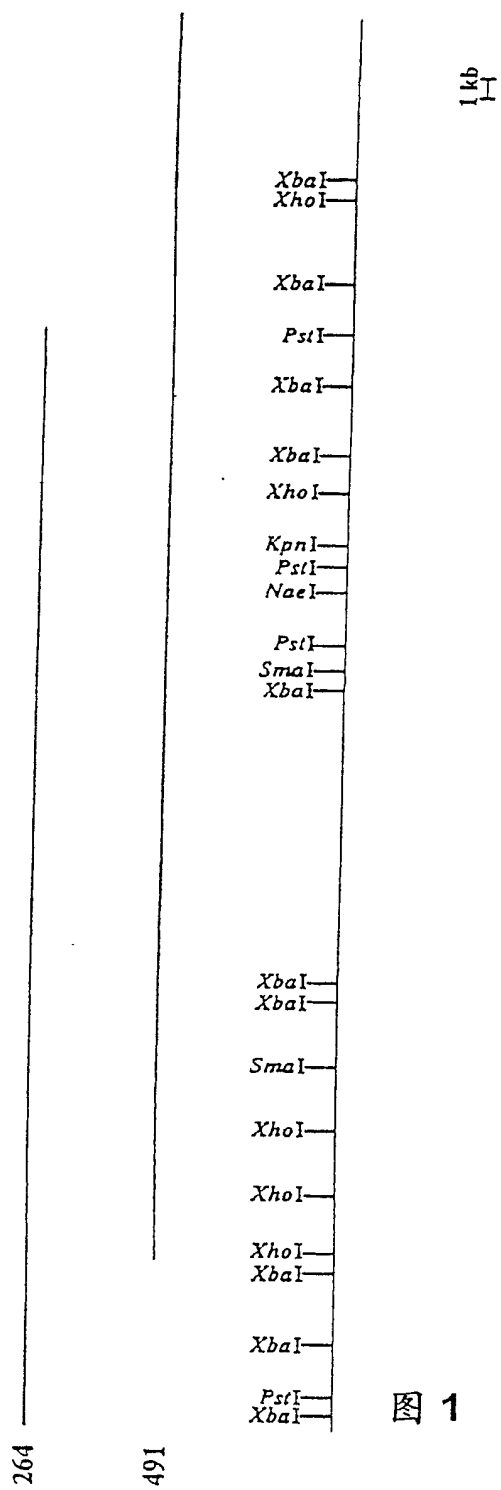
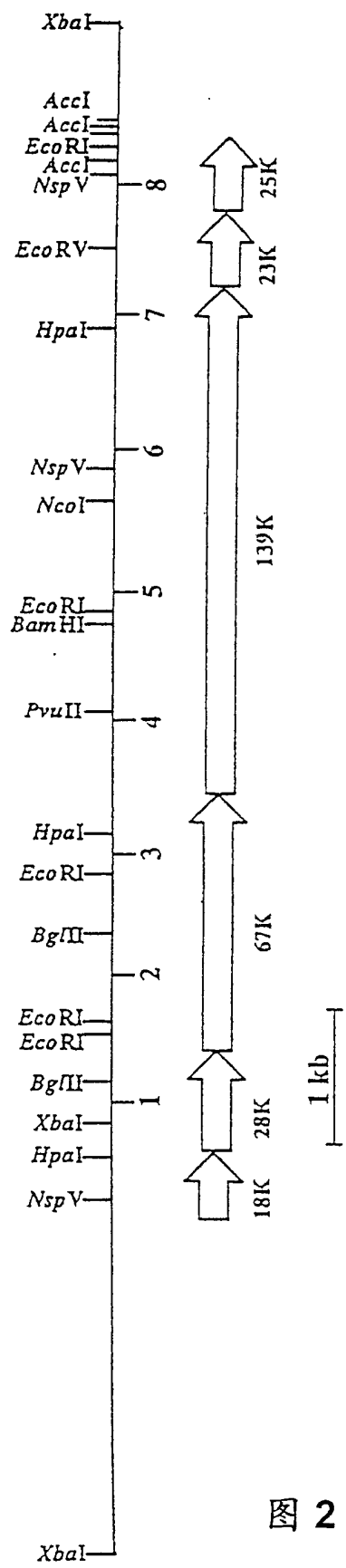


图 1



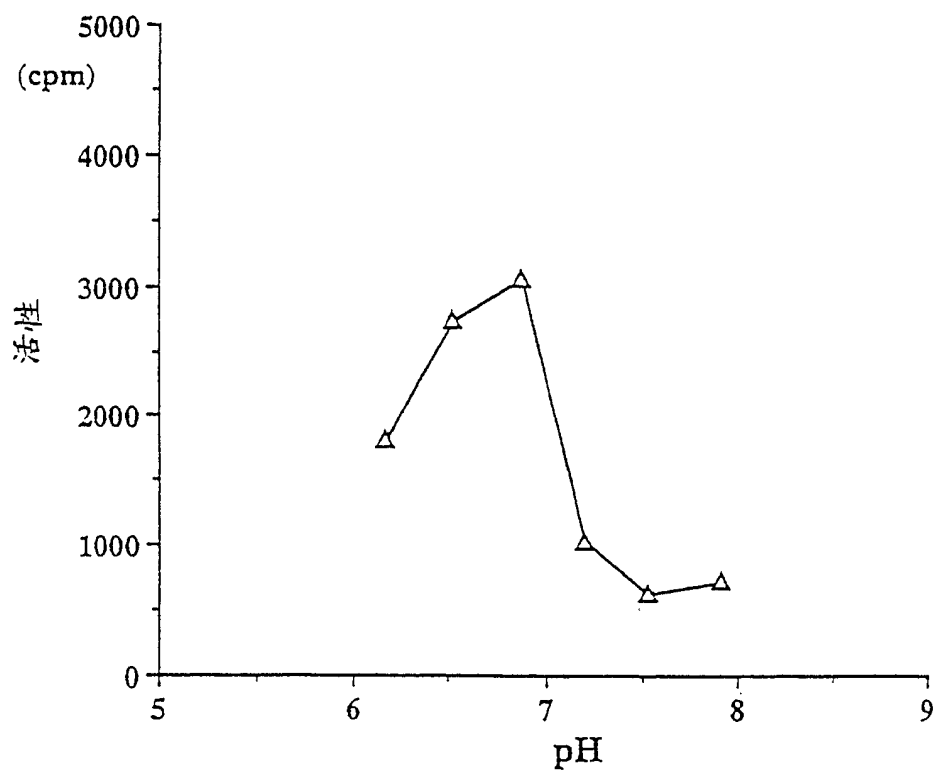


图 3

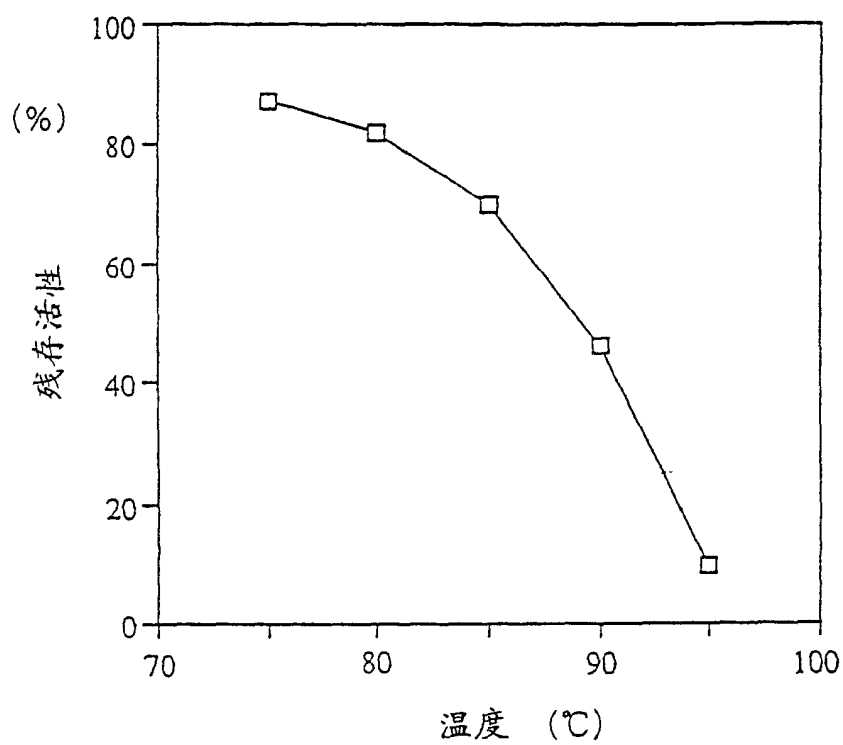


图 4

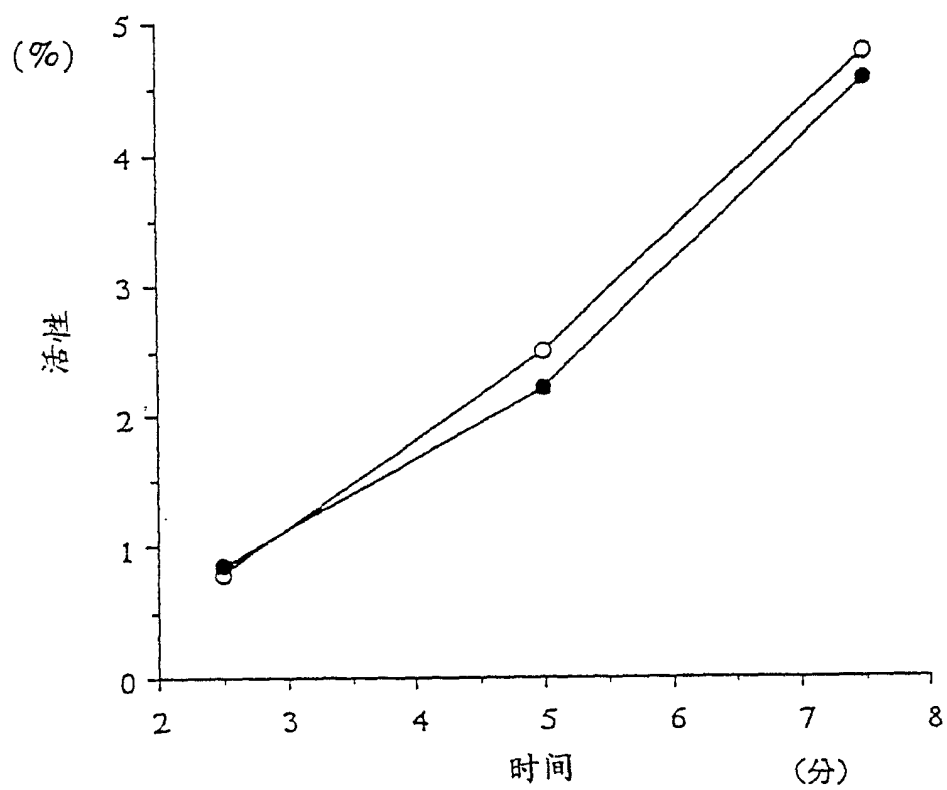


图 5

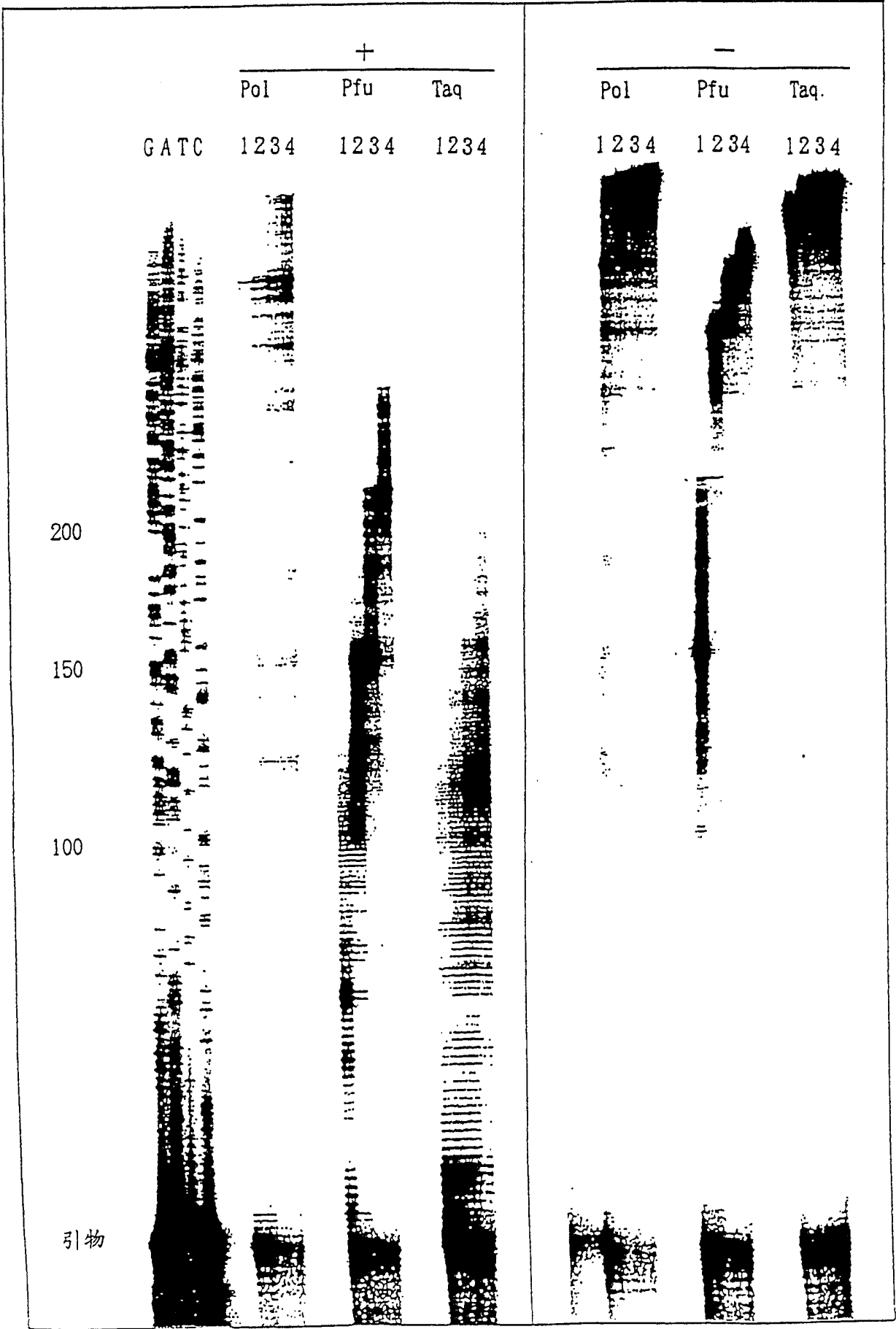


图 6