



(12) 发明专利申请

(10) 申请公布号 CN 113122517 A

(43) 申请公布日 2021.07.16

(21) 申请号 202110316008.9

C12Q 1/6869 (2018.01)

(22) 申请日 2021.03.24

C12Q 1/6844 (2018.01)

C12R 1/19 (2006.01)

(71) 申请人 深圳清华大学研究院

地址 518000 广东省深圳市南山区科技园
高新南七道19号清华研究院

申请人 安序源生物科技(深圳)有限公司

(72) 发明人 高亚平 何筠 陈丽伊 田晖

(74) 专利代理机构 广州嘉权专利商标事务所有
限公司 44205

代理人 刘焱

(51) Int. Cl.

C12N 9/12 (2006.01)

C12N 15/54 (2006.01)

C12N 15/70 (2006.01)

C12N 1/21 (2006.01)

权利要求书1页 说明书10页

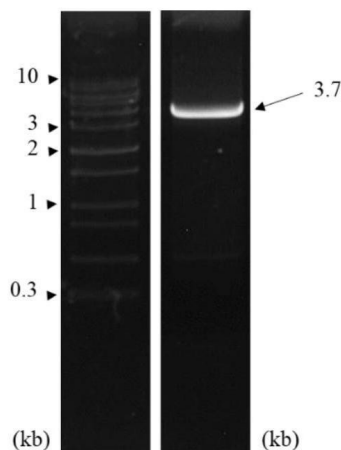
序列表16页 附图3页

(54) 发明名称

聚合酶突变体及其应用

(57) 摘要

本申请公开了聚合酶突变体及其应用。该聚合酶突变体包括如SEQ ID No.1所示的氨基酸序列经过以下突变后的氨基酸序列:第368位的苏氨酸突变为苯丙氨酸、第372位的苏氨酸突变为亮氨酸、第375位的谷氨酸突变为丝氨酸或赖氨酸、第478位的赖氨酸突变为酪氨酸、第484位的丙氨酸突变为谷氨酸、第512位的赖氨酸突变为酪氨酸。该聚合酶突变体至少具有如下有益效果:发明人意外地发现,具有如SEQ ID No.1所示氨基酸序列的Phi29聚合酶的上述位点发生突变后获得的突变体对于分子结构更复杂、分子量更大的核苷酸类似物底物具有更高的利用效率,可以有效利用核苷酸类似物进行模板DNA的复制和延伸。



1. 聚合酶突变体,其特征在于,包括如SEQ ID No.1所示的氨基酸序列经过以下突变后的氨基酸序列:第368位的苏氨酸突变为苯丙氨酸、第372位的苏氨酸突变为亮氨酸、第375位的谷氨酸突变为丝氨酸或赖氨酸、第478位的赖氨酸突变为酪氨酸、第484位的丙氨酸突变为谷氨酸、第512位的赖氨酸突变为酪氨酸;

优选的,所述聚合酶突变体的氨基酸序列如SEQ ID No.2~3中任一种所示。

2. 分离的多核苷酸,其特征在于,包括:

(a) 编码权利要求1所述的聚合酶突变体的核苷酸序列;或

(b) 与(a)所述的核苷酸序列互补的核苷酸序列;

优选的,所述多核苷酸的序列如SEQ ID No.5~6中任一种所示。

3. 重组载体,其特征在于,包括权利要求2所述的多核苷酸。

4. 宿主细胞,其特征在于,包括权利要求3所述重组载体,或包括权利要求2所述的多核苷酸。

5. 权利要求1所述的聚合酶突变体在核酸的复制、扩增或测序中的应用。

6. DNA模板的测序方法,其特征在于,包括以下步骤:

(1) 提供DNA模板、引物、核苷酸底物及权利要求1所述的聚合酶突变体;

(2) 在所述聚合酶突变体催化下,引导所述核苷酸底物掺入所述DNA模板,确定所述核苷酸底物的掺入顺序。

7. 根据权利要求6所述的测序方法,其特征在于,所述核苷酸底物为核苷酸、核苷酸类似物中的至少一种。

8. 根据权利要求7所述的测序方法,其特征在于,所述核苷酸类似物的分子量在1500~15000。

9. 组合物,其特征在于,包括权利要求1所述的聚合酶突变体。

10. 测序系统,其特征在于,所述测序系统包括权利要求1所述的聚合酶突变体。

聚合酶突变体及其应用

技术领域

[0001] 本申请涉及测序技术领域,尤其是涉及聚合酶突变体及其应用。

背景技术

[0002] 自1977年第一代测序技术的发明,测序技术已历经三代更迭发展。为适应广大的市场需求,纳米孔测序技术,凭借其小型化、自动化、低成本、高通量、高灵敏度等诸多优势,成为世界公认的第四代测序技术。与传统的测序技术不同,纳米孔测序可以在不对DNA进行生物或化学处理的前提下,直接采用物理办法读出DNA序列。其原理可以描述为:单个碱基通过纳米孔通道时,会引起通道电学参数的变化,对于A、C、T、G四种不同的碱基,其化学性质的差异所引起的电学参数的变化量也不同,对这些变化进行检测可以得到相应碱基的类型从而实现测序。

[0003] DNA链高速通过纳米孔的特性使得纳米孔测序成为可能,但其速度一般在微秒级别,过快的速度使得检测的信号质量较低甚至无法被检测到。相比较而言,DNA与DNA酶发生结合时,受酶催化反应的速度限制,其速度在毫秒级别左右,这提示可以通过DNA酶来控制DNA链在纳米孔中的通过速率,从而提高检测信号质量。因此,研发人员尝试将DNA外切酶与纳米孔复合,当DNA链靠近纳米孔时,DNA外切酶将核苷酸逐个剪下,剪下的核苷酸在电压驱动下通过纳米孔,相比于直接以DNA链通过的方式,速率放缓,对不同核苷酸的检测信号更容易区分。但利用外切酶测序过程中仍有许多需要改进的地方。因此,更多的研究针对DNA聚合酶控制DNA的通过速率。例如,有报道提出一种使用经过修饰的核苷酸类似物和纳米孔进行边合成边测序的方法,在该方法中,DNA聚合酶和纳米孔复合,当DNA模板链上具有部分杂交的引物并与纳米孔接触时,DNA聚合酶催化经过修饰的核苷酸类似物合成互补链,通过对该过程中引起的纳米孔电学参数的变化进行检测以完成测序。

[0004] Phi29 DNA聚合酶即嗜热脂肪芽孢杆菌的噬菌体Phi29中的p2蛋白。体外条件下,Phi29能够催化寡核苷酸引物起始的延伸反应,而且特有的结构域组成和折叠使其具有高保真性和卓越的链置换能力,在众多的商业化酶中,持续合成能力最高(>70kb)、延伸速率最快(50-200bp/s)。然而,相比于一般的核苷酸,上述纳米孔测序方法中所采用的核苷酸类似物的分子结构与天然核苷酸底物明显不同,野生型Phi29 DNA聚合酶对经修饰的核苷酸类似物底物的利用效率较低。

发明内容

[0005] 本申请旨在至少解决现有技术中存在的技术问题之一。为此,本申请提出一种能够对核苷酸类似物具有较高利用效率的聚合酶突变体及其应用。

[0006] 本申请的第一方面,提供聚合酶突变体,该聚合酶突变体包括如SEQ ID No.1所示的氨基酸序列经过以下突变后的氨基酸序列:第368位的苏氨酸突变为苯丙氨酸(T368F)、第372位的苏氨酸突变为亮氨酸(T372L)、第375位的谷氨酸突变为丝氨酸或赖氨酸(E375S或E375K)、第478位的赖氨酸突变为酪氨酸(K478Y)、第484位的丙氨酸突变为谷氨酸

(A484E)、第512位的赖氨酸突变为酪氨酸(K512Y)。

[0007] 其中,SEQ ID No.1所示的氨基酸序列如下:

[0008] MKHMPRKMYSCDFETTTKVEDCRVWAYGYMNIEDHSEYKIGNSLDEFMAWVLKVQADLYFHNLFKFDGA
FIINWLERNGFKWSADGLPNTYNTIISRGMQWYMIDICLGKGRKIHTVIYDSLKKLPFPVKKIAKDFKLTVLKG
DIDYHKERPVGKITYPEEYAYIKNDIQIIAEALLIQFKQGLDRMTAGSDSLKGFKDIITTKFKKVFPTLSLGLDK
EVRYAYRGGFTWLNDRFKEKEIGEGMVFVDNSLYPAQMYSRLLPYGEPVFEFGKYVWDEDYPLHIQHIRCEFELKE
GYIPTIQIKRSFYKGYNEYLKSSGGEIADLWLSNVDLELMKEHYDLYNVEYISGLKFKATTGLFKDFIDKWTYIKT
TSEGAIKQLAKLMLNSLYGKFASNPDVTGKVPYLKENGALGFRLGEEETKDPVYTPMGVFITAWARYTTITAAQAC
YDRIIYCDTDSIHLTGTEIPDVIKDIVDPKKLGYWAHESTFKRAKYLRQKTYIQDIYMKEVDGKLVEGSPDDYTDI
KFSVKCAGMTDKIKKEVTFENFKVGFSRKMKPKPVQVPGGVVLVDDTFTIKGTGSGALKTLESIVGDLEKADELKR
KYGSASAVRRLPVEELRELGFSDDEIAEIKGIPKKLREAFDLETAELYERYGSLKEIGRRLSYDDLLELGATPKA
AAEIKGPEFKFLLNIEGVGPKLAERILEAVDYDLERLASLNPEELAEKVEGLGEELAERVVYAARERVESRRKSGR
QERSEEWKEWLERKVGEGRARRLIEYFGSAGEVGKLVENAESKLLLEVPVIGDEAVARLVPGYKTLRDAGLTPAE
AERVLKRYGSVSKVQEGATPDELRELGLGDAKIARILGLRSLVNKRLDVTAYELKRRYGSVSAVRKAPVKELREL
GLSDRKIARIKIPETMLQVRGMSVEKAERLLERFDTWTKVKEAPVSELVRVPGVGLSLVKEIKAQVDPALKALLD
VKGVSPELADRLVEELGSPYRVLTAKKSDLMRVERVGPPLAERIRAAG。

[0009] 根据本申请实施例的聚合酶突变体,至少具有如下有益效果:

[0010] 发明人意外地发现,具有如SEQ ID No.1所示氨基酸序列的Phi29EL聚合酶的上述位点发生突变后获得的突变体对于分子结构更复杂、分子量更大的核苷酸类似物底物具有更高的利用效率,可以有效利用核苷酸类似物进行模板DNA的复制和延伸。

[0011] 根据本申请的一些实施方式,该聚合酶突变体的氨基酸序列如SEQ ID No.2~3中任一种所示。

[0012] 其中,SEQ ID No.2所示的氨基酸序列如下:

[0013] MKHMPRKMYSCDFETTTKVEDCRVWAYGYMNIEDHSEYKIGNSLDEFMAWVLKVQADLYFHNLFKFDGA
FIINWLERNGFKWSADGLPNTYNTIISRGMQWYMIDICLGKGRKIHTVIYDSLKKLPFPVKKIAKDFKLTVLKG
DIDYHKERPVGKITYPEEYAYIKNDIQIIAEALLIQFKQGLDRMTAGSDSLKGFKDIITTKFKKVFPTLSLGLDK
EVRYAYRGGFTWLNDRFKEKEIGEGMVFVDNSLYPAQMYSRLLPYGEPVFEFGKYVWDEDYPLHIQHIRCEFELKE
GYIPTIQIKRSFYKGYNEYLKSSGGEIADLWLSNVDLELMKEHYDLYNVEYISGLKFKATTGLFKDFIDKWFYIKL
TSSGAIKQLAKLMLNSLYGKFASNPDVTGKVPYLKENGALGFRLGEEETKDPVYTPMGVFITAWARYTTITAAQAC
YDRIIYCDTDSIHLTGTEIPDVIKDIVDPYKLGWEHESTFKRAKYLRQKTYIQDIYMKEVDGYLVEGSPDDYTDI
KFSVKCAGMTDKIKKEVTFENFKVGFSRKMKPKPVQVPGGVVLVDDTFTIKGTGSGALKTLESIVGDLEKADELKR
KYGSASAVRRLPVEELRELGFSDDEIAEIKGIPKKLREAFDLETAELYERYGSLKEIGRRLSYDDLLELGATPKA
AAEIKGPEFKFLLNIEGVGPKLAERILEAVDYDLERLASLNPEELAEKVEGLGEELAERVVYAARERVESRRKSGR
QERSEEWKEWLERKVGEGRARRLIEYFGSAGEVGKLVENAESKLLLEVPVIGDEAVARLVPGYKTLRDAGLTPAE
AERVLKRYGSVSKVQEGATPDELRELGLGDAKIARILGLRSLVNKRLDVTAYELKRRYGSVSAVRKAPVKELREL
GLSDRKIARIKIPETMLQVRGMSVEKAERLLERFDTWTKVKEAPVSELVRVPGVGLSLVKEIKAQVDPALKALLD
VKGVSPELADRLVEELGSPYRVLTAKKSDLMRVERVGPPLAERIRAAG。

[0014] SEQ ID No.3所示的氨基酸序列如下:

[0015] MKHMPRKMYSCDFETTTKVEDCRVWAYGYMNIEDHSEYKIGNSLDEFMAWVLKVQADLYFHNLFKFDGA

FIINWLERNGFKWSADGLPNTYNTIISRMGQWYIMIDICLGKGRKIHTVIYDSLKKLPFPVKKIAKDFKLTVLKG
DIDYHKERPVGKITYPEEYAYIKNDIQIIAEALLIQFKQGLDRMTAGSDSLKGFKDIITTKFKKFVPTLSLGLDK
EVRYAYRGGFTWLNDRFKEKEIGEGMVFDVNSLYPAQMYSRLPYGEPVFEKGKYVWDEDYPLHIQHIRCEFELKE
GYIPTIQIKRSFYKGYNEYLKSSSGEIALDLWLSNVDLELMKEHYDLNVEYISGLKFKATTGLFKDFIDKWFYIKL
TSKGAIKQLAKMLNSLYGKFASNPDVTGKVPYLGKENGALGFRLGEEETKDPVYTPMGVFITAWARYTTITAAQAC
YDRIIYCDTDSIHLTGTEIPDVIKDIVDPYKLGWEHESTFKRAKYLRQKTYIQDIYMKEVDGYLVEGSPDDYTDI
KFSVKCAGMTDKIKKEVTFENFKVGFSRKMKPKPVQVPGGVVLVDDTFTIKGTGSGALKTLESIVGDLEKADELKR
KYGSASAVRRLPVEELRELGFSDDEIAEIKGIPKKLREAFDLETAALYERYGSLKEIGRRLSYDDLLELGATPKA
AAEIKGPEFKFLLNIEGVGPKLAERILEAVDYDLERLASLNPEELAEKVEGLGEELAERVVYAARERVESRRKSGR
QERSEEEWKEWLERKVGEGRRRLIEYFGSAGEVGKLVENAESKLLVPGIGDEAVARLVPGYKTLRDAGLTPAE
AERVLKRYGSVSKVQEGATPDELRELGLGDAKIARILGLRSLVNKRLDVTAYELKRRYGSVSAVRKAPVKELREL
GLSDRKIARIKIPETMLQVRGMSVEKAERLLERFDTWTKVKEAPVSELVRVPGVGLSLVKEIKAQVDPAWKALLD
VKGVSPELADRLVEELGSPYRVLTAKKSDLMRVERVGPCKLAERIRAAG。

[0016] 其中,上述的聚合酶突变体中,突变前的Phi29EL聚合酶(SEQ ID No.1)中,在Phi29聚合酶的C端通过连接肽嵌合有坎德勒氏甲烷嗜热菌拓扑异构酶V的E~L的HhH基序。该拓扑异构酶V的E~L的HhH基序的氨基酸序列如SEQ ID No.4所示:

[0017] LKTLESIVGDLEKADELKRKYGSASAVRRLPVEELRELGFSDDEIAEIKGIPKKLREAFDLETAALY
ERYGSLKEIGRRLSYDDLLELGATPKAAAEIKGPEFKFLLNIEGVGPKLAERILEAVDYDLERLASLNPEELAEKV
EGLGEELAERVVYAARERVESRRKSGRQERSEEEWKEWLERKVGEGRRRLIEYFGSAGEVGKLVENAESKLLV
PGIGDEAVARLVPGYKTLRDAGLTPAEAERVLKRYGSVSKVQEGATPDELRELGLGDAKIARILGLRSLVNKRLDV
DTAYELKRRYGSVSAVRKAPVKELRELGLSDRKIARIKIPETMLQVRGMSVEKAERLLERFDTWTKVKEAPVSEL
VRVPGVGLSLVKEIKAQVDPAWKALLDVKGVSPELADRLVEELGSPYRVLTAKKSDLMRVERVGPCKLAERIRAAG。

[0018] 在利用聚合酶进行纳米孔测序时,较高的盐浓度对于减缓过孔速率有重要作用,能够有效提高检测分辨率。但是,常规的聚合酶在高盐浓度条件下活性降低,因此,通过引入坎德勒氏甲烷嗜热菌拓扑异构酶V的HhH基序提高Phi29聚合酶对高盐溶液的耐受性,保证测序质量和速度。

[0019] 本申请的第二方面,提供分离的多核苷酸,该多核苷酸包括:

[0020] (a) 编码上述的聚合酶突变体的核苷酸序列;

[0021] (b) 与(a)所述的核苷酸序列互补的核苷酸序列。

[0022] 根据本申请的一些实施方式,分离的多核苷酸的序列如SEQ ID No.5~6中任一种所示。

[0023] 本申请的第三方面,提供重组载体,该重组载体包括上述的多核苷酸。

[0024] 本申请的第四方面,提供宿主细胞,该宿主细胞包括上述的重组载体,或包括上述的多核苷酸。

[0025] 本申请的第五方面,提供上述的聚合酶突变体在核酸的复制、扩增或测序中的应用。

[0026] 本申请的第六方面,提供DNA模板的测序方法,该测序方法包括以下步骤:

[0027] (1) 提供DNA模板、引物、核苷酸底物及上述的聚合酶突变体;

[0028] (2) 在聚合酶突变体催化下,引导核苷酸底物掺入DNA模板,确定核苷酸底物的掺

入顺序。

[0029] 其中,引物是指天然或人工合成的寡核苷酸,在适当条件下(例如在缓冲液中提供核苷酸底物、聚合酶等反应原料并提供特定的温度条件)能够作为核酸合成的起始点进行延伸,从而使核苷酸底物掺入DNA模板,合成DNA模板的互补链。核苷酸底物是指由戊糖、磷酸酯及含氮杂环碱基形成的核苷酸单元或这些核苷酸单元的修饰产物。修饰产物具体是指通过对核苷酸进行修饰标记,以使不同类别的核苷酸(包括脱氧腺苷、脱氧胸苷、脱氧胞苷、脱氧鸟苷、脱氧尿苷)的大小、质量、电荷量等产生明显区别,从而提高检测过程中不同类别核苷酸之间的区分度。一方面,核苷酸类似物可以是包含多个磷酸基团的核苷酸,具体可以是二磷酸核苷酸、三磷酸核苷酸、四磷酸核苷酸、五磷酸核苷酸、六磷酸核苷酸等多磷酸核苷酸。另一方面,为了增强对核苷酸类似物的鉴别,多磷酸核苷酸上还可以进一步针对不同碱基在磷酸基团上引入不同长度/质量的标签,具体的标签可以是任选的聚合物接头如聚乙二醇或其衍生物,以及染料分子如香豆素等。

[0030] 根据本申请的一些实施方式,在掺入过程中,聚合酶突变体复合在纳米孔上或与纳米孔处于相对接近的位置。其中,纳米孔(nanopore)是指具有纳米尺度孔隙的材料,具体可以包括生物纳米孔、固态纳米孔、以及生物纳米孔和固态纳米孔的复合纳米孔。生物纳米孔具体可以例举出包括但不限于金黄色葡萄球菌 α -溶血素蛋白纳米孔(α -HL)、耻垢分枝杆菌孔道蛋白A(MspA)、噬菌体Phi29连接器(Phi29 connector)。固态纳米孔具体可以例举出包括但不限于以氮化硅、二氧化硅、氧化铝、玻璃毛细管、二硫化钼、石墨烯等材料制备得到的纳米孔。

[0031] 根据本申请的一些实施方式,确定核苷酸底物的掺入顺序的具体方法包括但不限于根据不同的核苷酸底物的化学性质的差异,分析掺入过程中体系的电学信号和/或光学信号,从而得到不同的核苷酸底物的掺入顺序。其中,电学信号的非限制性实例包括电流信号,具体的检测方式例如检测离子电流、隧道电流等。光学信号的非限制性实例包括荧光信号,具体的检测方式例如通过荧光共振能量转移检测等。

[0032] 根据本申请的一些实施方式,核苷酸底物为核苷酸、核苷酸类似物中的至少一种。

[0033] 核苷酸类似物作为底物在聚合酶催化作用下合成与DNA模板链互补的产物链时,修饰的标签通过聚合酶从核苷酸类似物中释放。释放出的标签通过跨孔电压等方式的介导,平移通过纳米孔,从而改变跨孔电流。通过检测其电流变化,可以推导出DNA链的碱基序列。

[0034] 根据本申请的一些实施方式,核苷酸类似物的分子量在1500~15000。当核苷酸类似物的分子量达到1500~15000时,标签的长度足够长,使过孔速率大大放缓,对纳米孔跨孔电流信号的检测分辨率明显提高。但同时,对于聚合酶对核苷酸类似物底物的利用效率也提出了更高的要求。另外,不同的核苷酸类似物携带有不同的标签,而不同标签的长度和结构的不同使其过孔时的阻滞电流和阻滞时间也不相同,从而能够区别不同的底物类型。优选的,核苷酸类似物的分子量为2000~15000、3000~15000、5000~15000。

[0035] 根据本申请的一些实施方式,核苷酸类似物为多磷酸核苷酸。

[0036] 根据本申请的一些实施方式,核苷酸类似物的磷酸基团在四个以上。

[0037] 根据本申请的一些实施方式,核苷酸类似物的磷酸基团在六个以上。磷酸基团越多,对于聚合酶的要求也越高,当核苷酸类似物的磷酸基团在六个以上时,本申请实施例中

所提供的聚合酶突变体仍然对其有较好的利用率。

[0038] 本申请的第七方面,提供组合物,该组合物包括上述的聚合酶突变体。该组合物具体可以是纳米孔和聚合酶突变体的组合物,纳米孔和聚合酶突变体通过共价连接或其它本领域熟知的方式结合到一起。在另外一些情况下,该组合物还可以是核苷酸底物和聚合酶突变体的组合物。在进行纳米孔测序或其它测序过程中,加入包含聚合酶突变体和核苷酸底物的混合液进行反应。

[0039] 本申请的第八方面,提供测序系统,该测序系统包括上述的聚合酶突变体。该测序系统能够以修饰后的核苷酸类似物作为底物参与待测序DNA模板的互补链的合成,有效进行延伸反应,从而完成测序。

[0040] 根据本申请的一些实施方式,该测序系统为纳米孔测序系统,该纳米孔测序系统在纳米孔上复合有上述的聚合酶突变体。

[0041] 本申请的附加方面和优点将在下面的描述中部分给出,部分将从下面的描述中变得明显,或通过本申请的实践了解到。

附图说明

[0042] 图1是本申请的一个实施例的pGH-18B的PCR产物的电泳结果,左侧为marker,右侧为pGH-18B的条带。

[0043] 图2是本申请的一个实施例的pET28a-Phi29EL、pGH-14B和pGH-18B的质粒双酶切产物的电泳结果,(A)为pET28a-Phi29EL,(B)为pGH-14B,(C)为pGH-18B;左侧为marker,右侧为对应质粒的条带。

[0044] 图3是本申请的一个实施例的表达载体pET28a-Phi29EL-14B和pET28a-Phi29EL-18B的质粒双酶切产物的电泳结果,(A)为pET28a-Phi29EL-14B,(B)为pET28a-Phi29EL-18B;左侧为marker,右侧为对应表达载体的条带。

[0045] 图4是本申请的一个实施例的聚合酶突变体的12% SDS-PAGE电泳结果,(A)为Phi29EL-14B聚合酶,(B)为Phi29EL-18B聚合酶。

[0046] 图5是本申请的一个实施例中部分核苷酸类似物的结构式示意图。

[0047] 图6是本申请的一个实施例的发卡模板示意图。

[0048] 图7是本申请的一个实施例中发卡延伸反应实验结果电泳图。

具体实施方式

[0049] 以下将结合实施例对本申请的构思及产生的技术效果进行清楚、完整地描述,以充分地理解本申请的目的、特征和效果。显然,所描述的实施例只是本申请的一部分实施例,而不是全部实施例,基于本申请的实施例,本领域的技术人员在不付出创造性劳动的前提下所获得的其他实施例,均属于本申请保护的范围。

[0050] 下面详细描述本申请的实施例,描述的实施例是示例性的,仅用于解释本申请,而不能理解为对本申请的限制。

[0051] 在本申请的描述中,若干的含义是一个以上,多个的含义是两个以上,大于、小于、超过等理解为不包括本数,以上、以下、以内等理解为包括本数。如果有描述到第一、第二只是用于区分技术特征为目的,而不能理解为指示或暗示相对重要性或者隐含指明所指示的

技术特征的数量或者隐含指明所指示的技术特征的先后关系。

[0052] 本申请的描述中,参考术语“一个实施例”、“一些实施例”、“示意性实施例”、“示例”、“具体示例”、或“一些示例”等的描述意指结合该实施例或示例描述的具体特征、结构、材料或者特点包含于本申请的至少一个实施例或示例中。在本说明书中,对上述术语的示意性表述不一定指的是相同的实施例或示例。而且,描述的具体特征、结构、材料或者特点可以在任何一个或多个实施例或示例中以合适的方式结合。

[0053] 下述实施例中使用的材料说明如下:

[0054] Phi29-EL:将野生型Phi29聚合酶引入坎德勒氏甲烷嗜热菌拓扑异构酶V的HhH基序的E~L基序得到的嵌合体聚合酶(具体参考中国专利CN110747191A)。

[0055] pET28a-Phi29EL:在pET28a质粒载体上插入Phi29-EL的基因片段得到的重组载体(具体参考中国专利CN110747191A)。

[0056] 定点突变试剂盒 **Q5[®]** Site-Directed Mutagenesis Kit:购自NEB公司,货号E0552。

[0057] 大肠杆菌表达菌株BL21:购自Transgene公司。

[0058] 普通DNA产物纯化试剂盒:购自天根公司,货号DP204-02。

[0059] KpnI限制性内切酶:购自NEB公司,货号R0142S。

[0060] BstBI限制性内切酶:购自NEB公司,货号R0519S。

[0061] 连接溶液Solution I:购自Takara公司,货号6022Q。

[0062] 质粒提取试剂盒 **Wolact[®]** Plasmid DNA Purification Kit:购自威邦生命科学(香港)有限公司,货号NP11001.50/200。

[0063] 实施例1

[0064] 本实施例提供聚合酶突变体,该聚合酶突变体的制备过程如下:

[0065] 1实验步骤

[0066] 1.1聚合酶突变体Phi29EL-14B基因片段合成

[0067] 以Phi29-EL的核苷酸序列为基础,通过全基因合成获得包含6个突变位点(T368F、T372L、E375S、K478Y、A484E、K512Y)的基因片段(记为pGH-14B,长度858bp),pGH-14B片段的5'和3'端分别包含BstBI和KpnI酶切位点,并连接到pGH载体上。

[0068] 1.2点突变生成pGH-18B基因片段

[0069] 以pGH-14B为模板,设计引物,利用定点突变试剂盒 **Q5[®]** Site-Directed Mutagenesis Kit进行S375K定点突变,得到相对于Phi29-EL包含6个突变位点(T368F、T372L、E375K、K478Y、A484E、K512Y)的基因片段(记为pGH-18B)。准备PCR反应体系,如表1所示:

[0070] 表1.点突变反应体系

	试剂	用量
	Q5 Hot Start High-Fidelity 2× Master Mix	12.5 μL
[0071]	模板 (10 ng/μL)	1 μL
	正向引物 (10 μM)	1.25 μL
	反向引物 (10 μM)	1.25 μL
	无核酸酶纯水	加至总体积 50μL

[0072] 突变正向引物的序列为:AAAGGAGCGATCAAGCAAC (SEQ ID No.7);

[0073] 突变反向引物的序列为:TGATGTCAACTTGATGTAAAAC (SEQ ID No.8)。

[0074] PCR扩增的条件为98℃变性30s,循环周期为98℃,10s,55℃,22s,72℃,2min,30个循环;72℃,2min。

[0075] 取PCR产物,进行Kinase-Ligase-DpnI (KLD) 处理去除模板、连接片段。反应体系如表2所示:

[0076] 表2.PCR产物处理体系

	试剂	用量
	PCR 产物	1 μl
[0077]	2× KLD Reaction Buffer	5 μl
	10× KLD Enzyme Mix	1 μl
	无核酸酶纯水	3 μl

[0078] 将反应体系混匀,室温孵育5min。转化大肠杆菌表达菌株BL21,将测序验证正确的载体pGH-18B,用于下一步实验。

[0079] 1.3载体酶切

[0080] 选取pET28a-Phi29EL、pGH-14B和pGH-18B的质粒进行双酶切。准备酶切反应体系,进行分步酶切。第一步酶切如下:

[0081] 表3.Kpn I酶切反应体系

	试剂	pET28a-Phi29EL 质粒	pGH-14B (或 pGH-18B) 质粒
	10×Buffer 2.1	8 μL	8 μL
[0082]	质粒	5μg	5μg
	KpnI (10U/μL)	4μL	4μL
	ddH ₂ O	加至总体积 80 μL	加至总体积 80 μL

[0083] 酶切反应条件为37℃,16h。选取酶切产物,准备第二步酶切体系,如表4:

[0084] 表4.BstBI酶切反应体系

	试剂	pET28a-Phi29EL 质粒	pGH-14B (或 pGH-18B) 质粒
[0085]	KpnI 酶切产物	80 μ L	80 μ L
	BstBI (10 U/ μ L)	4 μ L	4 μ L

[0086] 酶切反应条件为:65℃,4h。

[0087] 酶切产物用琼脂糖凝胶电泳检测,目标条带选取凝胶用普通DNA产物纯化试剂盒进行回收。

[0088] 1.4片段连接

[0089] 连接Phi29EL片段、pGH-14B片段和pGH-18B,用1.3中pGH-14B(或pGH-18B)质粒酶切回收产物片段替换野生型Phi29EL片段中对应片段,得到Phi29EL-14B(或Phi29EL-18B)。准备连接反应体系,如表5:

[0090] 表5.Phi29EL-14B(或Phi29EL-18B)连接反应体系

	试剂	用量
	Solution I	5 μ L
[0091]	Phi29EL 片段酶切回收产物	2.5 μ L
	14B (或 18B) 片段酶切回收产物	2.5 μ L

[0092] 连接溶液为Solution I。连接反应条件为:16℃,30min。

[0093] 取连接产物转化大肠杆菌Top10感受态细胞,挑取阳性克隆。选用质粒提取试剂盒Wolact[®]Plasmid DNA Purification Kit提取质粒,进行测序,获得序列正确的突变型Phi29DNA聚合酶质粒pET28a-Phi29EL-14B(或pET28a-Phi29EL-18B)。

[0094] 1.5蛋白表达纯化

[0095] 取质粒pGEX6P-1-Phi29、pET28a-Phi29EL、pET28a-Phi29EL-14B和pET28a-Phi29EL-18B,转化大肠杆菌表达菌株BL21,表达野生型Phi29 DNA聚合酶、Phi29-EL DNA聚合酶、Phi29EL-14B DNA聚合酶和Phi29EL-18B DNA聚合酶。蛋白诱导表达选择低温诱导,诱导条件为16℃,16h。

[0096] 蛋白纯化步骤包括:

[0097] ①收集菌体,超声破碎细胞。功率50~100W,超声5~15s,间隔10~30s。共持续20~40min。

[0098] ②离心去除细胞碎片,取上清。分别采用采用Glutathione Sepharose 4B和Ni Sepharose亲和层析柱纯化步骤裂解粗产物。

[0099] ③用12%SDS-PAGE检测蛋白纯度和浓度。

[0100] 2.实验结果

[0101] 2.1pGH-18B点突变结果

[0102] 图1是pGH-18B的PCR产物的电泳结果,扩增产物约为4kb。从图中可以看出,该次PCR已扩增出目的条带。

[0103] 2.2酶切结果

[0104] 图2是质粒双酶切结果。其中, (A) 为pET28a-Phi29EL, (B) 为pGH-14B, (C) 为pGH-18B; (A)、(B)、(C) 中左侧为marker, 右侧为对应质粒的两个重复的条带。结合图2, pET28a-Phi29EL、pGH-14B和pGH-18B经质粒双酶切分别回收图中所示的7000bp左右和855bp左右的片段。

[0105] 2.3pET28a-Phi29EL-14B载体酶切验证

[0106] 图3是表达载体双酶切结果。其中, (A) 为pET28a-Phi29EL-14B, (B) 为pET28a-Phi29EL-18B; (A)、(B) 中左侧为marker, 右侧为对应表达载体的条带。其中, pET28a-Phi29EL-14B用KpnI和BstBI双酶切, pET28a-Phi29EL-18B用EcoRI和NotI双酶切, 并对正确的克隆菌株测序验证。

[0107] 2.4Phi29EL-14B和Phi29EL-18B蛋白检测

[0108] 图4是聚合酶突变体的12%SDS-PAGE电泳结果, (A) 为Phi29EL-14B聚合酶的结果, (B) 为Phi29EL-18B聚合酶的结果。其中, M为蛋白分子量Marker; 泳道1和泳道2为突变型聚合酶的两个重复。结合图4, 上述聚合酶突变体的分子量均为116.6kD。

[0109] 其中, Phi29EL-14B的氨基酸序列如SEQ ID No.2所示, 核苷酸序列如SEQ ID No.5所示; Phi29EL-18B的氨基酸序列如SEQ ID No.3所示, 核苷酸序列如SEQ ID No.6所示。

[0110] 实施例2

[0111] 聚合酶突变体底物结合能力检测

[0112] 天然核苷酸底物由碱基、核糖和三个磷酸组成, 每个核苷酸的整合, 伴随着DNA链的延伸和焦磷酸的释放。在DNA聚合酶辅助的纳米孔测序技术中, 单个DNA聚合酶分子共价结合到 α -溶血素七聚物纳米孔上, DNA聚合酶结合模板引物复合物后, 以携带标记的核苷酸为底物, 进行边合成边测序。纳米孔依次捕获和检测携带标记的核苷酸, 从而产生核苷酸特异的电流信号。根据这一原理, 本实施例中设计合成了包含六个磷酸基团、聚乙二醇(PEG)侧链和香豆素(Coumarin)的核苷酸类似物, 记为Coumarin-PEG_m-dN6P (CP_m-dN6P), 其中, m表示核苷酸类似物中聚乙二醇单元的重复数量, N表示A、T、C、G中的任一种。

[0113] 具体如下: Coumarin-PEG12-dA6P (CP12-dA6P)、Coumarin-PEG24-dC6P (CP24-dC6P)、Coumarin-PEG24-dA6P (CP24-dA6P)、Coumarin-PEG24-dT6P (CP24-dT6P)、Coumarin-PEG24-dG6P (CP24-dG6P)、Coumarin-PEG36-dT6P (CP36-dT6P) 和Coumarin-PEG44-dG6P (CP44-dG6P), 部分核苷酸类似物的结构式如图5中的(A)~(D)所示。

[0114] 取等量的Phi29EL DNA聚合酶、Phi29EL-14B DNA聚合酶和Phi29EL-18B DNA聚合酶, 以核苷酸类似物为底物, 进行发夹延伸实验, 检测突变体底物结合能力。发夹模板如图6所示, 长度为53nt, 延伸产物长度为74nt。延伸反应结束后, 模板链含量减少, 生成产物链。根据有无产物链以及产物链的含量, 即可判断DNA聚合酶对特异性底物的结合能力。

[0115] 将200nM实施例1中制得的DNA聚合酶、500nM的发夹模板及反应缓冲液混合, 得到反应体系, 另外以Phi29EL DNA聚合酶作为对照。反应体系包括: dNTP (-KC1) 组、dNTP (+KC1) 组、CP-dN6P (-KC1) 组、CP-dN6P (+KC1) 组, 具体组成如表6所示:

[0116] 其中, 反应缓冲液由50mM的Tris-HCl、10mM的MgCl₂、10mM的(NH₄)₂SO₄、4mM的DTT及200 μ M dNTPs组成, 反应缓冲液在25℃条件下的pH为7.5。将反应体系在温度为30℃的条件下反应10min, 得到反应产物。

[0117] 表6发夹延伸反应体系

试剂	dNTP	dNTP	CP-dN6P	CP-dN6P
	(-KCl)	(+KCl)	(-KCl)	(+KCl)
10×Buffer	2 μ L	2 μ L	2 μ L	2 μ L
牛血清白蛋白 (20 mg/ml)	0.1 μ L	0.1 μ L	0.1 μ L	0.1 μ L
[0118] dNTP (1 mM)	0.4 μ L	0.4 μ L	-	-
dNTP 类似物 (1mM)	-	-	0.8 μ L	0.8 μ L
酶 (2 μ M)	1 μ L	1 μ L	1 μ L	1 μ L
KCl (2 M)	-	3 μ L	-	3 μ L
H ₂ O	到 20 μ L	到 20 μ L	到 20 μ L	到 20 μ L

[0119] 将反应产物加入6×DNA loading后,上样到20%TBE-Urea聚丙烯酰胺凝胶中;电泳条件为180V,3h。电泳结束后,将PAGE胶浸泡在SYBR Gold染液中,缓慢水平震荡约20min。然后利用Azure Biosystems成像检测模板链和产物链的亮度。通过模板链和产物链的亮度确定实施例1制备的DNA聚合酶的活性。

[0120] 实验结果如图7所示,从图中可以看出,实施例1制备的DNA聚合酶的具有DNA酶的活性,且Phi29EL-14B和Phi29EL-18B分别能够以CP24-dC6P、CP24-dA6P、CP24-dT6P和CP24-dG6P的混合物为底物,还能够以CP12-dA6P、CP24-dC6P、CP36-dT6P和CP44-dG6P混合物为底物,合成全长的延伸产物。而作为对照的Phi29EL DNA聚合酶以两类核苷酸类似物为底物时,全部或者大部分产物为不完全延伸产物。

[0121] 比较不同KCl浓度对反应的影响,从图中可以看出,添加0.3M KCl时,三种DNA聚合酶的活性均有一定程度的下降,但仍然能够利用核苷酸类似物合成产物链。其中,Phi29EL-14B和Phi29EL-18B这两种DNA聚合酶能够利用天然底物和修饰底物,合成少量全长产物;而Phi29EL聚合酶仅能生成少量的不完全延伸产物。证实Phi29EL-14B和Phi29EL-18B这两种DNA聚合酶在低盐和高盐浓度下,均能有效利用包含PEG侧链的核苷酸类似物为底物,进行DNA复制和延伸。

[0122] 上面结合实施例对本申请作了详细说明,但是本申请不限于上述实施例,在所属技术领域普通技术人员所具备的知识范围内,还可以在不脱离本申请宗旨的前提下作出各种变化。此外,在不冲突的情况下,本申请的实施例及实施例中的特征可以相互组合。

SEQUENCE LISTING

<110> 深圳清华大学研究院

安序源生物科技(深圳)有限公司

<120> 聚合酶突变体及其应用

<130> 1

<160> 8

<170> PatentIn version 3.5

<210> 1

<211> 1028

<212> PRT

<213> 人工序列

<400> 1

```

Met Lys His Met Pro Arg Lys Met Tyr Ser Cys Asp Phe Glu Thr Thr
1           5           10           15
Thr Lys Val Glu Asp Cys Arg Val Trp Ala Tyr Gly Tyr Met Asn Ile
          20           25           30
Glu Asp His Ser Glu Tyr Lys Ile Gly Asn Ser Leu Asp Glu Phe Met
          35           40           45
Ala Trp Val Leu Lys Val Gln Ala Asp Leu Tyr Phe His Asn Leu Lys
          50           55           60
Phe Asp Gly Ala Phe Ile Ile Asn Trp Leu Glu Arg Asn Gly Phe Lys
65           70           75           80
Trp Ser Ala Asp Gly Leu Pro Asn Thr Tyr Asn Thr Ile Ile Ser Arg
          85           90           95
Met Gly Gln Trp Tyr Met Ile Asp Ile Cys Leu Gly Tyr Lys Gly Lys
          100          105          110
Arg Lys Ile His Thr Val Ile Tyr Asp Ser Leu Lys Lys Leu Pro Phe
          115          120          125
Pro Val Lys Lys Ile Ala Lys Asp Phe Lys Leu Thr Val Leu Lys Gly
          130          135          140
Asp Ile Asp Tyr His Lys Glu Arg Pro Val Gly Tyr Lys Ile Thr Pro
145          150          155          160
Glu Glu Tyr Ala Tyr Ile Lys Asn Asp Ile Gln Ile Ile Ala Glu Ala
          165          170          175
Leu Leu Ile Gln Phe Lys Gln Gly Leu Asp Arg Met Thr Ala Gly Ser
          180          185          190
Asp Ser Leu Lys Gly Phe Lys Asp Ile Ile Thr Thr Lys Lys Phe Lys
          195          200          205

```

Lys Val Phe Pro Thr Leu Ser Leu Gly Leu Asp Lys Glu Val Arg Tyr		
210	215	220
Ala Tyr Arg Gly Gly Phe Thr Trp Leu Asn Asp Arg Phe Lys Glu Lys		
225	230	235 240
Glu Ile Gly Glu Gly Met Val Phe Asp Val Asn Ser Leu Tyr Pro Ala		
	245	250 255
Gln Met Tyr Ser Arg Leu Leu Pro Tyr Gly Glu Pro Ile Val Phe Glu		
	260	265 270
Gly Lys Tyr Val Trp Asp Glu Asp Tyr Pro Leu His Ile Gln His Ile		
	275	280 285
Arg Cys Glu Phe Glu Leu Lys Glu Gly Tyr Ile Pro Thr Ile Gln Ile		
	290	295 300
Lys Arg Ser Arg Phe Tyr Lys Gly Asn Glu Tyr Leu Lys Ser Ser Gly		
305	310	315 320
Gly Glu Ile Ala Asp Leu Trp Leu Ser Asn Val Asp Leu Glu Leu Met		
	325	330 335
Lys Glu His Tyr Asp Leu Tyr Asn Val Glu Tyr Ile Ser Gly Leu Lys		
	340	345 350
Phe Lys Ala Thr Thr Gly Leu Phe Lys Asp Phe Ile Asp Lys Trp Thr		
	355	360 365
Tyr Ile Lys Thr Thr Ser Glu Gly Ala Ile Lys Gln Leu Ala Lys Leu		
	370	375 380
Met Leu Asn Ser Leu Tyr Gly Lys Phe Ala Ser Asn Pro Asp Val Thr		
385	390	395 400
Gly Lys Val Pro Tyr Leu Lys Glu Asn Gly Ala Leu Gly Phe Arg Leu		
	405	410 415
Gly Glu Glu Glu Thr Lys Asp Pro Val Tyr Thr Pro Met Gly Val Phe		
	420	425 430
Ile Thr Ala Trp Ala Arg Tyr Thr Thr Ile Thr Ala Ala Gln Ala Cys		
	435	440 445
Tyr Asp Arg Ile Ile Tyr Cys Asp Thr Asp Ser Ile His Leu Thr Gly		
	450	455 460
Thr Glu Ile Pro Asp Val Ile Lys Asp Ile Val Asp Pro Lys Lys Leu		
465	470	475 480
Gly Tyr Trp Ala His Glu Ser Thr Phe Lys Arg Ala Lys Tyr Leu Arg		
	485	490 495
Gln Lys Thr Tyr Ile Gln Asp Ile Tyr Met Lys Glu Val Asp Gly Lys		
	500	505 510
Leu Val Glu Gly Ser Pro Asp Asp Tyr Thr Asp Ile Lys Phe Ser Val		

515	520	525
Lys Cys Ala Gly Met Thr Asp	Lys Ile Lys Lys Glu Val Thr Phe Glu	
530	535	540
Asn Phe Lys Val Gly Phe Ser Arg	Lys Met Lys Pro Lys Pro Val Gln	
545	550	555
Val Pro Gly Gly Val Val Leu Val	Asp Asp Thr Phe Thr Ile Lys Gly	
565	570	575
Thr Gly Ser Gly Ala Leu Lys Thr	Leu Glu Ser Ile Val Gly Asp Leu	
580	585	590
Glu Lys Ala Asp Glu Leu Lys Arg	Lys Tyr Gly Ser Ala Ser Ala Val	
595	600	605
Arg Arg Leu Pro Val Glu Glu Leu	Arg Glu Leu Gly Phe Ser Asp Asp	
610	615	620
Glu Ile Ala Glu Ile Lys Gly Ile	Pro Lys Lys Leu Arg Glu Ala Phe	
625	630	635
Asp Leu Glu Thr Ala Ala Glu Leu	Tyr Glu Arg Tyr Gly Ser Leu Lys	
645	650	655
Glu Ile Gly Arg Arg Leu Ser Tyr	Asp Asp Leu Leu Glu Leu Gly Ala	
660	665	670
Thr Pro Lys Ala Ala Ala Glu Ile	Lys Gly Pro Glu Phe Lys Phe Leu	
675	680	685
Leu Asn Ile Glu Gly Val Gly Pro	Lys Leu Ala Glu Arg Ile Leu Glu	
690	695	700
Ala Val Asp Tyr Asp Leu Glu Arg	Leu Ala Ser Leu Asn Pro Glu Glu	
705	710	715
Leu Ala Glu Lys Val Glu Gly Leu	Gly Glu Glu Leu Ala Glu Arg Val	
725	730	735
Val Tyr Ala Ala Arg Glu Arg Val	Glu Ser Arg Arg Lys Ser Gly Arg	
740	745	750
Gln Glu Arg Ser Glu Glu Glu Trp	Lys Glu Trp Leu Glu Arg Lys Val	
755	760	765
Gly Glu Gly Arg Ala Arg Arg Leu	Ile Glu Tyr Phe Gly Ser Ala Gly	
770	775	780
Glu Val Gly Lys Leu Val Glu Asn	Ala Glu Val Ser Lys Leu Leu Glu	
785	790	795
Val Pro Gly Ile Gly Asp Glu Ala	Val Ala Arg Leu Val Pro Gly Tyr	
805	810	815
Lys Thr Leu Arg Asp Ala Gly Leu	Thr Pro Ala Glu Ala Glu Arg Val	
820	825	830

Leu Lys Arg Tyr Gly Ser Val Ser Lys Val Gln Glu Gly Ala Thr Pro																			
835																			
Asp Glu Leu Arg Glu Leu Gly Leu Gly Asp Ala Lys Ile Ala Arg Ile																			
850																			
Leu Gly Leu Arg Ser Leu Val Asn Lys Arg Leu Asp Val Asp Thr Ala																			
865																			
Tyr Glu Leu Lys Arg Arg Tyr Gly Ser Val Ser Ala Val Arg Lys Ala																			
Pro Val Lys Glu Leu Arg Glu Leu Gly Leu Ser Asp Arg Lys Ile Ala																			
900																			
Arg Ile Lys Gly Ile Pro Glu Thr Met Leu Gln Val Arg Gly Met Ser																			
915																			
Val Glu Lys Ala Glu Arg Leu Leu Glu Arg Phe Asp Thr Trp Thr Lys																			
930																			
Val Lys Glu Ala Pro Val Ser Glu Leu Val Arg Val Pro Gly Val Gly																			
945																			
Leu Ser Leu Val Lys Glu Ile Lys Ala Gln Val Asp Pro Ala Trp Lys																			
Ala Leu Leu Asp Val Lys Gly Val Ser Pro Glu Leu Ala Asp Arg Leu																			
980																			
Val Glu Glu Leu Gly Ser Pro Tyr Arg Val Leu Thr Ala Lys Lys Ser																			
995																			
Asp Leu Met Arg Val Glu Arg Val Gly Pro Lys Leu Ala Glu Arg																			
1010																			
Ile Arg Ala Ala Gly																			
1025																			
<210> 2																			
<211> 1028																			
<212> PRT																			
<213> 人工序列																			
<400> 2																			
Met Lys His Met Pro Arg Lys Met Tyr Ser Cys Asp Phe Glu Thr Thr																			
1																			
Thr Lys Val Glu Asp Cys Arg Val Trp Ala Tyr Gly Tyr Met Asn Ile																			
20																			
Glu Asp His Ser Glu Tyr Lys Ile Gly Asn Ser Leu Asp Glu Phe Met																			
35																			
Ala Trp Val Leu Lys Val Gln Ala Asp Leu Tyr Phe His Asn Leu Lys																			
50																			

Phe	Asp	Gly	Ala	Phe	Ile	Ile	Asn	Trp	Leu	Glu	Arg	Asn	Gly	Phe	Lys
65					70					75					80
Trp	Ser	Ala	Asp	Gly	Leu	Pro	Asn	Thr	Tyr	Asn	Thr	Ile	Ile	Ser	Arg
				85						90					95
Met	Gly	Gln	Trp	Tyr	Met	Ile	Asp	Ile	Cys	Leu	Gly	Tyr	Lys	Gly	Lys
				100						105					110
Arg	Lys	Ile	His	Thr	Val	Ile	Tyr	Asp	Ser	Leu	Lys	Lys	Leu	Pro	Phe
				115						120					125
Pro	Val	Lys	Lys	Ile	Ala	Lys	Asp	Phe	Lys	Leu	Thr	Val	Leu	Lys	Gly
				130						135					140
Asp	Ile	Asp	Tyr	His	Lys	Glu	Arg	Pro	Val	Gly	Tyr	Lys	Ile	Thr	Pro
145					150					155					160
Glu	Glu	Tyr	Ala	Tyr	Ile	Lys	Asn	Asp	Ile	Gln	Ile	Ile	Ala	Glu	Ala
					165					170					175
Leu	Leu	Ile	Gln	Phe	Lys	Gln	Gly	Leu	Asp	Arg	Met	Thr	Ala	Gly	Ser
					180					185					190
Asp	Ser	Leu	Lys	Gly	Phe	Lys	Asp	Ile	Ile	Thr	Thr	Lys	Lys	Phe	Lys
					195					200					205
Lys	Val	Phe	Pro	Thr	Leu	Ser	Leu	Gly	Leu	Asp	Lys	Glu	Val	Arg	Tyr
					210					215					220
Ala	Tyr	Arg	Gly	Gly	Phe	Thr	Trp	Leu	Asn	Asp	Arg	Phe	Lys	Glu	Lys
225					230					235					240
Glu	Ile	Gly	Glu	Gly	Met	Val	Phe	Asp	Val	Asn	Ser	Leu	Tyr	Pro	Ala
					245					250					255
Gln	Met	Tyr	Ser	Arg	Leu	Leu	Pro	Tyr	Gly	Glu	Pro	Ile	Val	Phe	Glu
					260					265					270
Gly	Lys	Tyr	Val	Trp	Asp	Glu	Asp	Tyr	Pro	Leu	His	Ile	Gln	His	Ile
					275					280					285
Arg	Cys	Glu	Phe	Glu	Leu	Lys	Glu	Gly	Tyr	Ile	Pro	Thr	Ile	Gln	Ile
					290					295					300
Lys	Arg	Ser	Arg	Phe	Tyr	Lys	Gly	Asn	Glu	Tyr	Leu	Lys	Ser	Ser	Gly
305					310					315					320
Gly	Glu	Ile	Ala	Asp	Leu	Trp	Leu	Ser	Asn	Val	Asp	Leu	Glu	Leu	Met
					325					330					335
Lys	Glu	His	Tyr	Asp	Leu	Tyr	Asn	Val	Glu	Tyr	Ile	Ser	Gly	Leu	Lys
					340					345					350
Phe	Lys	Ala	Thr	Thr	Gly	Leu	Phe	Lys	Asp	Phe	Ile	Asp	Lys	Trp	Phe
					355					360					365
Tyr	Ile	Lys	Leu	Thr	Ser	Ser	Gly	Ala	Ile	Lys	Gln	Leu	Ala	Lys	Leu

370	375	380
Met Leu Asn Ser Leu Tyr Gly Lys Phe Ala Ser Asn Pro Asp Val Thr		
385	390	395
Gly Lys Val Pro Tyr Leu Lys Glu Asn Gly Ala Leu Gly Phe Arg Leu		400
	405	410
Gly Glu Glu Glu Thr Lys Asp Pro Val Tyr Thr Pro Met Gly Val Phe		415
	420	425
Ile Thr Ala Trp Ala Arg Tyr Thr Thr Ile Thr Ala Ala Gln Ala Cys		430
	435	440
Tyr Asp Arg Ile Ile Tyr Cys Asp Thr Asp Ser Ile His Leu Thr Gly		445
	450	455
Thr Glu Ile Pro Asp Val Ile Lys Asp Ile Val Asp Pro Tyr Lys Leu		460
	465	470
Gly Tyr Trp Glu His Glu Ser Thr Phe Lys Arg Ala Lys Tyr Leu Arg		475
	485	490
Gln Lys Thr Tyr Ile Gln Asp Ile Tyr Met Lys Glu Val Asp Gly Tyr		495
	500	505
Leu Val Glu Gly Ser Pro Asp Asp Tyr Thr Asp Ile Lys Phe Ser Val		510
	515	520
Lys Cys Ala Gly Met Thr Asp Lys Ile Lys Lys Glu Val Thr Phe Glu		525
	530	535
Asn Phe Lys Val Gly Phe Ser Arg Lys Met Lys Pro Lys Pro Val Gln		540
	545	550
Val Pro Gly Gly Val Val Leu Val Asp Asp Thr Phe Thr Ile Lys Gly		555
	565	570
Thr Gly Ser Gly Ala Leu Lys Thr Leu Glu Ser Ile Val Gly Asp Leu		575
	580	585
Glu Lys Ala Asp Glu Leu Lys Arg Lys Tyr Gly Ser Ala Ser Ala Val		590
	595	600
Arg Arg Leu Pro Val Glu Glu Leu Arg Glu Leu Gly Phe Ser Asp Asp		605
	610	615
Glu Ile Ala Glu Ile Lys Gly Ile Pro Lys Lys Leu Arg Glu Ala Phe		620
	625	630
Asp Leu Glu Thr Ala Ala Glu Leu Tyr Glu Arg Tyr Gly Ser Leu Lys		635
	645	650
Glu Ile Gly Arg Arg Leu Ser Tyr Asp Asp Leu Leu Glu Leu Gly Ala		655
	660	665
Thr Pro Lys Ala Ala Ala Glu Ile Lys Gly Pro Glu Phe Lys Phe Leu		670
	675	680
		685

Leu Asn Ile Glu Gly Val Gly Pro Lys Leu Ala Glu Arg Ile Leu Glu			
690	695	700	
Ala Val Asp Tyr Asp Leu Glu Arg Leu Ala Ser Leu Asn Pro Glu Glu			
705	710	715	720
Leu Ala Glu Lys Val Glu Gly Leu Gly Glu Glu Leu Ala Glu Arg Val			
	725	730	735
Val Tyr Ala Ala Arg Glu Arg Val Glu Ser Arg Arg Lys Ser Gly Arg			
	740	745	750
Gln Glu Arg Ser Glu Glu Glu Trp Lys Glu Trp Leu Glu Arg Lys Val			
	755	760	765
Gly Glu Gly Arg Ala Arg Arg Leu Ile Glu Tyr Phe Gly Ser Ala Gly			
	770	775	780
Glu Val Gly Lys Leu Val Glu Asn Ala Glu Val Ser Lys Leu Leu Glu			
785	790	795	800
Val Pro Gly Ile Gly Asp Glu Ala Val Ala Arg Leu Val Pro Gly Tyr			
	805	810	815
Lys Thr Leu Arg Asp Ala Gly Leu Thr Pro Ala Glu Ala Glu Arg Val			
	820	825	830
Leu Lys Arg Tyr Gly Ser Val Ser Lys Val Gln Glu Gly Ala Thr Pro			
	835	840	845
Asp Glu Leu Arg Glu Leu Gly Leu Gly Asp Ala Lys Ile Ala Arg Ile			
850	855	860	
Leu Gly Leu Arg Ser Leu Val Asn Lys Arg Leu Asp Val Asp Thr Ala			
865	870	875	880
Tyr Glu Leu Lys Arg Arg Tyr Gly Ser Val Ser Ala Val Arg Lys Ala			
	885	890	895
Pro Val Lys Glu Leu Arg Glu Leu Gly Leu Ser Asp Arg Lys Ile Ala			
	900	905	910
Arg Ile Lys Gly Ile Pro Glu Thr Met Leu Gln Val Arg Gly Met Ser			
	915	920	925
Val Glu Lys Ala Glu Arg Leu Leu Glu Arg Phe Asp Thr Trp Thr Lys			
	930	935	940
Val Lys Glu Ala Pro Val Ser Glu Leu Val Arg Val Pro Gly Val Gly			
945	950	955	960
Leu Ser Leu Val Lys Glu Ile Lys Ala Gln Val Asp Pro Ala Trp Lys			
	965	970	975
Ala Leu Leu Asp Val Lys Gly Val Ser Pro Glu Leu Ala Asp Arg Leu			
	980	985	990
Val Glu Glu Leu Gly Ser Pro Tyr Arg Val Leu Thr Ala Lys Lys Ser			

995	1000	1005
Asp Leu Met Arg Val Glu Arg Val Gly Pro Lys Leu Ala Glu Arg		
1010	1015	1020
Ile Arg Ala Ala Gly		
1025		
<210> 3		
<211> 1028		
<212> PRT		
<213> 人工序列		
<400> 3		
Met Lys His Met Pro Arg Lys Met Tyr Ser Cys Asp Phe Glu Thr Thr		
1	5	10
Thr Lys Val Glu Asp Cys Arg Val Trp Ala Tyr Gly Tyr Met Asn Ile		
20	25	30
Glu Asp His Ser Glu Tyr Lys Ile Gly Asn Ser Leu Asp Glu Phe Met		
35	40	45
Ala Trp Val Leu Lys Val Gln Ala Asp Leu Tyr Phe His Asn Leu Lys		
50	55	60
Phe Asp Gly Ala Phe Ile Ile Asn Trp Leu Glu Arg Asn Gly Phe Lys		
65	70	75
Trp Ser Ala Asp Gly Leu Pro Asn Thr Tyr Asn Thr Ile Ile Ser Arg		
85	90	95
Met Gly Gln Trp Tyr Met Ile Asp Ile Cys Leu Gly Tyr Lys Gly Lys		
100	105	110
Arg Lys Ile His Thr Val Ile Tyr Asp Ser Leu Lys Lys Leu Pro Phe		
115	120	125
Pro Val Lys Lys Ile Ala Lys Asp Phe Lys Leu Thr Val Leu Lys Gly		
130	135	140
Asp Ile Asp Tyr His Lys Glu Arg Pro Val Gly Tyr Lys Ile Thr Pro		
145	150	155
Glu Glu Tyr Ala Tyr Ile Lys Asn Asp Ile Gln Ile Ile Ala Glu Ala		
165	170	175
Leu Leu Ile Gln Phe Lys Gln Gly Leu Asp Arg Met Thr Ala Gly Ser		
180	185	190
Asp Ser Leu Lys Gly Phe Lys Asp Ile Ile Thr Thr Lys Lys Phe Lys		
195	200	205
Lys Val Phe Pro Thr Leu Ser Leu Gly Leu Asp Lys Glu Val Arg Tyr		
210	215	220
Ala Tyr Arg Gly Gly Phe Thr Trp Leu Asn Asp Arg Phe Lys Glu Lys		

225	230	235	240
Glu Ile Gly Glu Gly Met Val Phe Asp Val Asn Ser Leu Tyr Pro Ala			
	245	250	255
Gln Met Tyr Ser Arg Leu Leu Pro Tyr Gly Glu Pro Ile Val Phe Glu			
	260	265	270
Gly Lys Tyr Val Trp Asp Glu Asp Tyr Pro Leu His Ile Gln His Ile			
	275	280	285
Arg Cys Glu Phe Glu Leu Lys Glu Gly Tyr Ile Pro Thr Ile Gln Ile			
	290	295	300
Lys Arg Ser Arg Phe Tyr Lys Gly Asn Glu Tyr Leu Lys Ser Ser Gly			
305	310	315	320
Gly Glu Ile Ala Asp Leu Trp Leu Ser Asn Val Asp Leu Glu Leu Met			
	325	330	335
Lys Glu His Tyr Asp Leu Tyr Asn Val Glu Tyr Ile Ser Gly Leu Lys			
	340	345	350
Phe Lys Ala Thr Thr Gly Leu Phe Lys Asp Phe Ile Asp Lys Trp Phe			
	355	360	365
Tyr Ile Lys Leu Thr Ser Lys Gly Ala Ile Lys Gln Leu Ala Lys Leu			
	370	375	380
Met Leu Asn Ser Leu Tyr Gly Lys Phe Ala Ser Asn Pro Asp Val Thr			
385	390	395	400
Gly Lys Val Pro Tyr Leu Lys Glu Asn Gly Ala Leu Gly Phe Arg Leu			
	405	410	415
Gly Glu Glu Glu Thr Lys Asp Pro Val Tyr Thr Pro Met Gly Val Phe			
	420	425	430
Ile Thr Ala Trp Ala Arg Tyr Thr Thr Ile Thr Ala Ala Gln Ala Cys			
	435	440	445
Tyr Asp Arg Ile Ile Tyr Cys Asp Thr Asp Ser Ile His Leu Thr Gly			
	450	455	460
Thr Glu Ile Pro Asp Val Ile Lys Asp Ile Val Asp Pro Tyr Lys Leu			
465	470	475	480
Gly Tyr Trp Glu His Glu Ser Thr Phe Lys Arg Ala Lys Tyr Leu Arg			
	485	490	495
Gln Lys Thr Tyr Ile Gln Asp Ile Tyr Met Lys Glu Val Asp Gly Tyr			
	500	505	510
Leu Val Glu Gly Ser Pro Asp Asp Tyr Thr Asp Ile Lys Phe Ser Val			
	515	520	525
Lys Cys Ala Gly Met Thr Asp Lys Ile Lys Lys Glu Val Thr Phe Glu			
	530	535	540

Asn Phe Lys Val Gly Phe Ser Arg Lys Met Lys Pro Lys Pro Val Gln			
545	550	555	560
Val Pro Gly Gly Val Val Leu Val Asp Asp Thr Phe Thr Ile Lys Gly			
	565	570	575
Thr Gly Ser Gly Ala Leu Lys Thr Leu Glu Ser Ile Val Gly Asp Leu			
	580	585	590
Glu Lys Ala Asp Glu Leu Lys Arg Lys Tyr Gly Ser Ala Ser Ala Val			
	595	600	605
Arg Arg Leu Pro Val Glu Glu Leu Arg Glu Leu Gly Phe Ser Asp Asp			
	610	615	620
Glu Ile Ala Glu Ile Lys Gly Ile Pro Lys Lys Leu Arg Glu Ala Phe			
625	630	635	640
Asp Leu Glu Thr Ala Ala Glu Leu Tyr Glu Arg Tyr Gly Ser Leu Lys			
	645	650	655
Glu Ile Gly Arg Arg Leu Ser Tyr Asp Asp Leu Leu Glu Leu Gly Ala			
	660	665	670
Thr Pro Lys Ala Ala Ala Glu Ile Lys Gly Pro Glu Phe Lys Phe Leu			
	675	680	685
Leu Asn Ile Glu Gly Val Gly Pro Lys Leu Ala Glu Arg Ile Leu Glu			
	690	695	700
Ala Val Asp Tyr Asp Leu Glu Arg Leu Ala Ser Leu Asn Pro Glu Glu			
705	710	715	720
Leu Ala Glu Lys Val Glu Gly Leu Gly Glu Glu Leu Ala Glu Arg Val			
	725	730	735
Val Tyr Ala Ala Arg Glu Arg Val Glu Ser Arg Arg Lys Ser Gly Arg			
	740	745	750
Gln Glu Arg Ser Glu Glu Glu Trp Lys Glu Trp Leu Glu Arg Lys Val			
	755	760	765
Gly Glu Gly Arg Ala Arg Arg Leu Ile Glu Tyr Phe Gly Ser Ala Gly			
	770	775	780
Glu Val Gly Lys Leu Val Glu Asn Ala Glu Val Ser Lys Leu Leu Glu			
785	790	795	800
Val Pro Gly Ile Gly Asp Glu Ala Val Ala Arg Leu Val Pro Gly Tyr			
	805	810	815
Lys Thr Leu Arg Asp Ala Gly Leu Thr Pro Ala Glu Ala Glu Arg Val			
	820	825	830
Leu Lys Arg Tyr Gly Ser Val Ser Lys Val Gln Glu Gly Ala Thr Pro			
	835	840	845
Asp Glu Leu Arg Glu Leu Gly Leu Gly Asp Ala Lys Ile Ala Arg Ile			

850	855	860
Leu Gly Leu Arg Ser	Leu Val Asn Lys Arg	Leu Asp Val Asp Thr Ala
865	870	875
Tyr Glu Leu Lys Arg	Arg Tyr Gly Ser Val	Ser Ala Val Arg Lys Ala
885	890	895
Pro Val Lys Glu Leu Arg	Glu Leu Gly Leu Ser	Asp Arg Lys Ile Ala
900	905	910
Arg Ile Lys Gly Ile Pro	Glu Thr Met Leu Gln	Val Arg Gly Met Ser
915	920	925
Val Glu Lys Ala Glu Arg	Leu Leu Glu Arg Phe	Asp Thr Trp Thr Lys
930	935	940
Val Lys Glu Ala Pro Val	Ser Glu Leu Val Arg	Val Pro Gly Val Gly
945	950	955
Leu Ser Leu Val Lys Glu	Ile Lys Ala Gln Val	Asp Pro Ala Trp Lys
965	970	975
Ala Leu Leu Asp Val Lys	Gly Val Ser Pro Glu	Leu Ala Asp Arg Leu
980	985	990
Val Glu Glu Leu Gly Ser	Pro Tyr Arg Val Leu	Thr Ala Lys Lys Ser
995	1000	1005
Asp Leu Met Arg Val Glu	Arg Val Gly Pro Lys	Leu Ala Glu Arg
1010	1015	1020
Ile Arg Ala Ala Gly		
1025		
<210> 4		
<211> 447		
<212> PRT		
<213> 人工序列		
<400> 4		
Leu Lys Thr Leu Glu Ser	Ile Val Gly Asp Leu	Glu Lys Ala Asp Glu
1	5	10
Leu Lys Arg Lys Tyr Gly	Ser Ala Ser Ala Val	Arg Arg Leu Pro Val
20	25	30
Glu Glu Leu Arg Glu Leu	Gly Phe Ser Asp Asp	Glu Ile Ala Glu Ile
35	40	45
Lys Gly Ile Pro Lys Lys	Leu Arg Glu Ala Phe	Asp Leu Glu Thr Ala
50	55	60
Ala Glu Leu Tyr Glu Arg	Tyr Gly Ser Leu Lys	Glu Ile Gly Arg Arg
65	70	75
Leu Ser Tyr Asp Asp Leu	Leu Glu Leu Gly Ala	Thr Pro Lys Ala Ala

			85					90					95				
Ala	Glu	Ile	Lys	Gly	Pro	Glu	Phe	Lys	Phe	Leu	Leu	Asn	Ile	Glu	Gly		
			100					105					110				
Val	Gly	Pro	Lys	Leu	Ala	Glu	Arg	Ile	Leu	Glu	Ala	Val	Asp	Tyr	Asp		
			115					120					125				
Leu	Glu	Arg	Leu	Ala	Ser	Leu	Asn	Pro	Glu	Glu	Leu	Ala	Glu	Lys	Val		
			130					135					140				
Glu	Gly	Leu	Gly	Glu	Glu	Leu	Ala	Glu	Arg	Val	Val	Tyr	Ala	Ala	Arg		
145			150					155					160				
Glu	Arg	Val	Glu	Ser	Arg	Arg	Lys	Ser	Gly	Arg	Gln	Glu	Arg	Ser	Glu		
			165					170					175				
Glu	Glu	Trp	Lys	Glu	Trp	Leu	Glu	Arg	Lys	Val	Gly	Glu	Gly	Arg	Ala		
			180					185					190				
Arg	Arg	Leu	Ile	Glu	Tyr	Phe	Gly	Ser	Ala	Gly	Glu	Val	Gly	Lys	Leu		
195			200					205									
Val	Glu	Asn	Ala	Glu	Val	Ser	Lys	Leu	Leu	Glu	Val	Pro	Gly	Ile	Gly		
210			215					220									
Asp	Glu	Ala	Val	Ala	Arg	Leu	Val	Pro	Gly	Tyr	Lys	Thr	Leu	Arg	Asp		
225			230					235					240				
Ala	Gly	Leu	Thr	Pro	Ala	Glu	Ala	Glu	Arg	Val	Leu	Lys	Arg	Tyr	Gly		
			245					250					255				
Ser	Val	Ser	Lys	Val	Gln	Glu	Gly	Ala	Thr	Pro	Asp	Glu	Leu	Arg	Glu		
			260					265					270				
Leu	Gly	Leu	Gly	Asp	Ala	Lys	Ile	Ala	Arg	Ile	Leu	Gly	Leu	Arg	Ser		
			275					280					285				
Leu	Val	Asn	Lys	Arg	Leu	Asp	Val	Asp	Thr	Ala	Tyr	Glu	Leu	Lys	Arg		
290			295					300									
Arg	Tyr	Gly	Ser	Val	Ser	Ala	Val	Arg	Lys	Ala	Pro	Val	Lys	Glu	Leu		
305			310					315					320				
Arg	Glu	Leu	Gly	Leu	Ser	Asp	Arg	Lys	Ile	Ala	Arg	Ile	Lys	Gly	Ile		
			325					330					335				
Pro	Glu	Thr	Met	Leu	Gln	Val	Arg	Gly	Met	Ser	Val	Glu	Lys	Ala	Glu		
			340					345					350				
Arg	Leu	Leu	Glu	Arg	Phe	Asp	Thr	Trp	Thr	Lys	Val	Lys	Glu	Ala	Pro		
355			360					365									
Val	Ser	Glu	Leu	Val	Arg	Val	Pro	Gly	Val	Gly	Leu	Ser	Leu	Val	Lys		
370			375					380									
Glu	Ile	Lys	Ala	Gln	Val	Asp	Pro	Ala	Trp	Lys	Ala	Leu	Leu	Asp	Val		
385			390					395					400				

Lys	Gly	Val	Ser	Pro	Glu	Leu	Ala	Asp	Arg	Leu	Val	Glu	Glu	Leu	Gly		
					405						410						415
Ser	Pro	Tyr	Arg	Val	Leu	Thr	Ala	Lys	Lys	Ser	Asp	Leu	Met	Arg	Val		
					420						425						430
Glu	Arg	Val	Gly	Pro	Lys	Leu	Ala	Glu	Arg	Ile	Arg	Ala	Ala	Gly			
					435						440						445

<210> 5

<211> 3093

<212> DNA

<213> 人工序列

<400> 5

```

atgccgagaa agatgtatag ttgtgacttt gagacaacta ctaaagtgga agactgtagg 60
gtatgggcgt atggttatat gaatatagaa gatcacagtg agtacaaaat aggtaatagc 120
ctggatgagt ttatggcgtg ggtgttgaag gtacaagctg atctatattt ccataacctc 180
aaatttgacg gagcttttat cattaactgg ttggaacgta atggttttta gtggtcggct 240
gacggattgc caaacacata taatacgatc atatctcgca tgggacaatg gtacatgatt 300
gatatatgtt taggctacaa agggaaacgt aagatacata cagtgatata tgacagctta 360
aagaaactac cgtttcctgt taagaagata gctaaagact ttaaactaac tgttcttaaa 420
ggtgatattg attaccacaa agaaagacca gtcggctata agataacacc cgaagaatac 480
gcctatatta aaaacgatat tcagattatt gcggaagctc tgttaattca gtttaagcaa 540
ggttttagacc ggatgacagc aggcagtgac agtctaaaag gtttcaagga tattataacc 600
actaagaaat tcaaaaaggt gtttcctaca ttgagtcttg gactcgataa ggaagtgaga 660
tacgcctata gaggtggttt tacatggtta aatgataggt tcaaagaaaa agaaatcgga 720
gaaggcatgg tcttcgatgt taatagtcta tatcctgcac agatgtatag ccgtctcctt 780
ccatatggtg aacctatagt attcgagggt aaatacgttt gggacgaaga ttaccacta 840
cacatacagc atatcagatg tgagttcgaa ttgaaagagg gctatatacc cactatacag 900
ataaaaagaa gtaggtttta taaaggtaat gactaccta aaagtagcgg cggggagata 960
gccgacctct ggttgtcaaa ttagaccta gaattaatga aagaacacta cgatttatat 1020
aacgttgaat atatcagcgg cttaaaattt aaagcaacta caggtttggt taaagatttt 1080
atagataaat ggttttacat caagttgaca tcaagcggag cgatcaagca actagcaaaa 1140
ctgatgttaa acagtctata cggtaaattc gctagtaacc ctgatgttac agggaaagtc 1200
ccttatttta aagagaatgg ggcgctaggt ttcagacttg gagaagagga aacaaaagac 1260
cctgtttata cacctatggg cgttttcatc actgcatggg ctagatacac gacaattaca 1320
gcggcacagg cttgttatga tcggataata tactgtgata ctgacagcat acatttaacg 1380
ggtacagaga tacctgatgt aataaaaagat atagttgacc cttacaaatt gggatactgg 1440
gaacatgaaa gtacattcaa aagagctaaa tatctgagac agaagaccta tatacaagac 1500
atctatatga aagaagtaga tggttattta gtagaagga gtccagatga ttacactgat 1560
ataaaattta gtgttaaatg tgcgggaatg actgacaaga ttaagaaaga ggttacgttt 1620
gagaatttca aagtcggatt cagtcggaaa atgaagccta agcctgtgca agtgccgggc 1680

```

ggggtggttc tggttgatga cacattcaca atcaaaggta cgggctctgg cgccctgaag 1740
 accctggaga gcatagtagg ggatctggag aaggccgacg agctgaagcg gaagtacgga 1800
 tccgcgtccg cggttcgacg tctgcccgtg gaggagctac gcgaactcgg gttctccgac 1860
 gatgagatcg ccgagatcaa ggggatacct aagaagctcc gggaggcctt cgaccttgag 1920
 accgccgcgg aactctacga gcggtacggg tcgctgaaag agatcggtcg ccgactctct 1980
 tacgacgatc tactcgagct cgggtcgact ccgaaggccg cggccgagat caaggggccc 2040
 gagttcaagt tcctcctgaa catcgaaggg gtcggaccga aactcgctga gcggatactc 2100
 gaggccgtgg attatgacct cgagcgactg gcttccctga atcccaggga acttgcgagg 2160
 aaggtggaag gactgggcga agagctcgcg gagcgcgtcg tgtacgctgc tagggagcgc 2220
 gtagaaaagtc gcaggaaagtc cggccgccag gagcggtcgg aggaagaatg gaaggagtgg 2280
 ctcgagcgta aggtcggcga ggggagggct cgccggttga ttgagtattt cggctccgcg 2340
 ggtgaagtag gaaaagctggg cgagaacgcc gaggtgtcga agctactgga ggtcccgggt 2400
 ataggcgacg aggccgtcgc taggtctgta ccgggctaca agaccctacg agacgccggt 2460
 ctcacgccgg ccgaagcgga gcgcgtgctg aaacggtagc gctcggtctc caaagtgcag 2520
 gaaggagcca ctccggacga gttacgcgag ctcggcctcg gcgacgcaa gatcgcgagg 2580
 atcctgggcc tgcgcagcct ggtgaacaag aggctggacg tggacaccgc gtacgagctc 2640
 aagcgtagat acggttccgt ctccgccgtc cggaaggccc cggtgaaaga actgcgcgag 2700
 ctcggcctct ccgatcgga gatcgcacgt atcaaggga tcccggagac gatgcttcag 2760
 gtccgagggg tgagcgtgga gaaagcggag cggctgctgg agcgtttcga tacctggacc 2820
 aaggtgaagg aagctcccgt ctcgagctg gtgagagtcc cgggtgtcgg attgagtttg 2880
 gtgaaggaga tcaaggctca ggtggatccg gcctggaagg cacttctgga tgtcaaaggg 2940
 gtcagtccgg agctggccga ccggctcgtc gaggagctcg gcagcccgta tcgggtgctg 3000
 acggccaaga aatccgacct gatgagagtc gagagagtc gaccgaagct cgccgagcga 3060
 atccgggccc cgggctaagc ggccgcatcg tga 3093

<210> 6

<211> 3093

<212> DNA

<213> 人工序列

<400> 6

atgccgagaa agatgtatag ttgtgacttt gagacaacta ctaaagtgga agactgtagg 60
 gtatgggctg atggttatat gaatatagaa gatcacagt agtacaaaat aggtaatagc 120
 ctggatgagt ttatggcgtg ggtgttgaag gtacaagctg atctatattt ccataacctc 180
 aaatttgacg gagcttttat cattaactgg ttggaacgta atggttttaa gtggtcggct 240
 gacggattgc caaacacata taatacgatc atatctcgca tgggacaatg gtacatgatt 300
 gatatatgtt taggctacaa agggaaacgt aagatacata cagtgatata tgacagctta 360
 aagaaactac cgtttcctgt taagaagata gctaaagact ttaaactaac tgtttctaaa 420
 ggtgatattg attaccacaa agaaagacca gtcggctata agataacacc cgaagaatac 480
 gcctatatata aaaacgatat tcagattatt gcggaagctc tgttaattca gtttaagcaa 540
 ggttttagacc ggatgacagc aggcagtgac agtctaaaag gtttcaagga tattataacc 600

actaagaaat tcaaaaaggt gtttcctaca ttgagtcttg gactcgataa ggaagtgaga 660
 tacgcctata gaggtggttt tacatggtta aatgataggt tcaaagaaaa agaaatcgga 720
 gaaggcatgg tcttcgatgt taatagtcta tatcctgcac agatgtatag ccgtctcctt 780
 ccatatgggtg aacctatagt attcgagggt aaatacgttt gggacgaaga ttaccacta 840
 cacatacagc atatcagatg tgagttcgaa ttgaaagagg gctatatacc cactatacag 900
 ataaaaagaa gtaggtttta taaaggtaat gagtacctaa aaagtagcgg cggggagata 960
 gccgacctct ggttgtcaaa ttagacctga gaattaatga aagaacacta cgatttatat 1020
 aacgttgaat atatcagcgg cttaaaattt aaagcaacta caggtttggt taaagatttt 1080
 atagataaat ggttttacat caagttgaca tcaaaaggag cgatcaagca actagcaaaa 1140
 ctgatgttaa acagtctata cggtaaattc gctagtaacc ctgatgttac agggaaagtc 1200
 ccttatttta aagagaatgg ggcgctaggt ttcagacttg gagaagagga aacaaaagac 1260
 cctgtttata cacctatggg cgttttcatc actgcatggg ctagatacac gacaattaca 1320
 gcggcacagg ctgtttatga tcggataata tactgtgata ctgacagcat acatttaacg 1380
 ggtacagaga tacctgatgt aataaaagat atagttgacc cttacaaatt gggatactgg 1440
 gaacatgaaa gtacattcaa aagagctaaa tatctgagac agaagaccta tatacaagac 1500
 atctatatga aagaagtaga tggttattta gtagaagga gtccagatga ttacactgat 1560
 ataaaattta gtgttaaatg tgcgggaatg actgacaaga ttaagaaaga ggttacgttt 1620
 gagaatttca aagtcggatt cagtcgaaa atgaagccta agcctgtgca agtgccgggc 1680
 ggggtggttc tggttgatga cacattcaca atcaaagga ccggctcttg cgccctgaag 1740
 accctggaga gcatagtagg ggatctggag aaggccgacg agctgaagcg gaagtacgga 1800
 tccgcgtccg cggttcgacg tctgcccgtg gaggagctac gcgaactcgg gttctccgac 1860
 gatgagatcg ccgagatcaa ggggatacct aagaagctcc gggaggcctt cgaccttgag 1920
 accgccgagg aactctacga gcggtacggt tcgctgaaag agatcggtcg ccgactctct 1980
 tacgacgac tactcgagct cgggtcgact ccgaaggccg cggccgagat caaggggccg 2040
 gagttcaagt tcctctgaa catcgaagggt gtcggaccga aactcgctga gcgatactc 2100
 gaggccgtgg attatgacct cgagcgactg gcttcctga atcccaggga acttgcgag 2160
 aagggtggaag gactgggcca agagctcgcg gagcgcgtcg tgtacgctgc tagggagcgc 2220
 gtagaaaagtc gcaggaaagtc cggccgccag gagcggctcg aggaagaatg gaaggagtgg 2280
 ctcgagcgta aggtcggcga ggggagggtc cgccggttga ttgagtattt cggctccgcg 2340
 ggtgaagtag gaaagctggt cgagaacgcc gaggtgtcga agctactgga ggtcccggt 2400
 ataggcgacg aggccgtcgc taggctcgta ccgggctaca agaccctacg agacgccggt 2460
 ctcacgccgg ccgaagcgga gcgctgtctg aaacggtagc gctcggtctc caaagtgcag 2520
 gaaggagcca ctccggacga gttacgcgag ctcgccctcg gcgacgcaa gatcgcgagg 2580
 atcctgggcc tgcgcagcct ggtgaacaag aggttgagc tggacaccgc gtacgagctc 2640
 aagcgtagat acggttccgt ctccgccgtc cggaaggccc cggtgaaaga actgcgcgag 2700
 ctcggcctct ccgatcgga gatcgcacgt atcaaggcca tcccggagac gatgcttcag 2760
 gtccgaggga tgagcgtgga gaaagcgag cggctgctgg agcgtttcga tacctggacc 2820
 aagggtgaagg aagctcccgt ctcggagctg gtgagagtc cgggtgtcgg attgagtttg 2880
 gtgaaggaga tcaaggctca ggtggatccg gcctggaagg cacttctgga tgtcaaagg 2940

gtcagtccgg agctggccga ccggctcgtc gaggagctcg gcagcccgta tcgggtgctg 3000
acggccaaga aatccgacct gatgagagtc gagagagtcg gaccgaagct cgccgagcga 3060
atccggggccg cgggctaagc ggccgcatcg tga 3093

<210> 7

<211> 19

<212> DNA

<213> 人工序列

<400> 7

aaaggagcga tcaagcaac 19

<210> 8

<211> 22

<212> DNA

<213> 人工序列

<400> 8

tgatgtcaac ttgatgtaaa ac 22

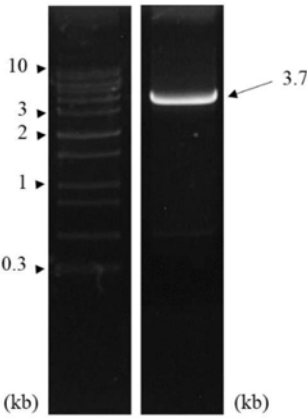


图1

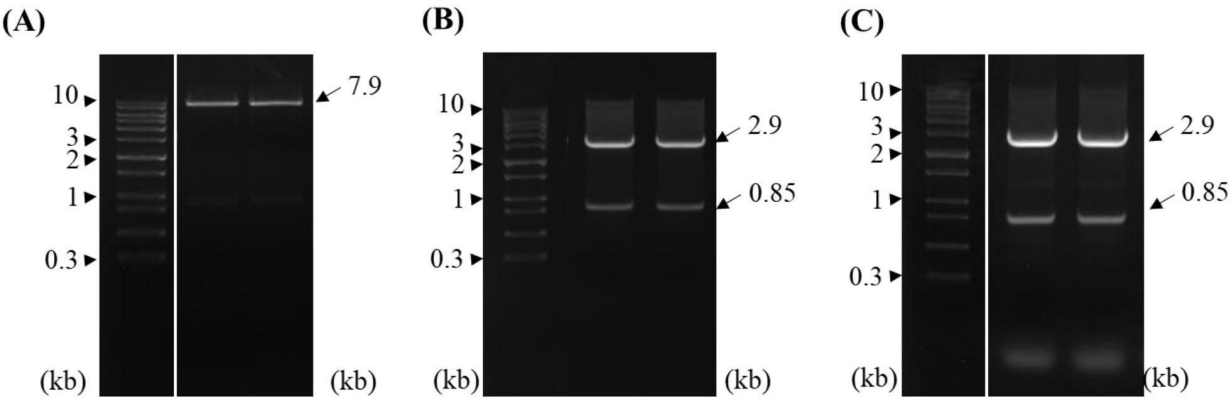


图2

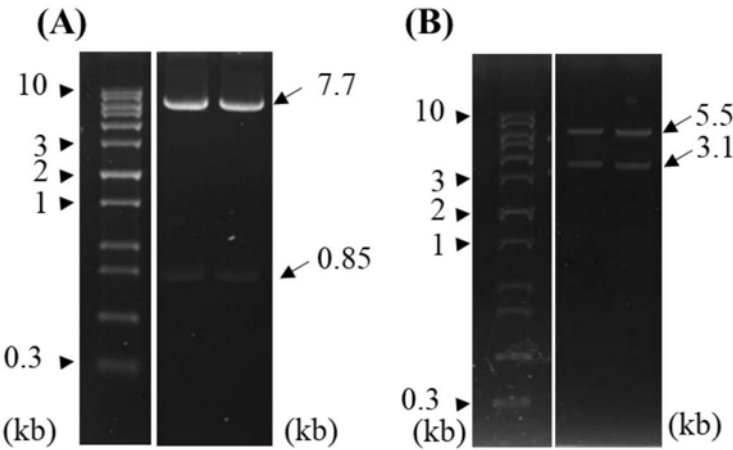


图3

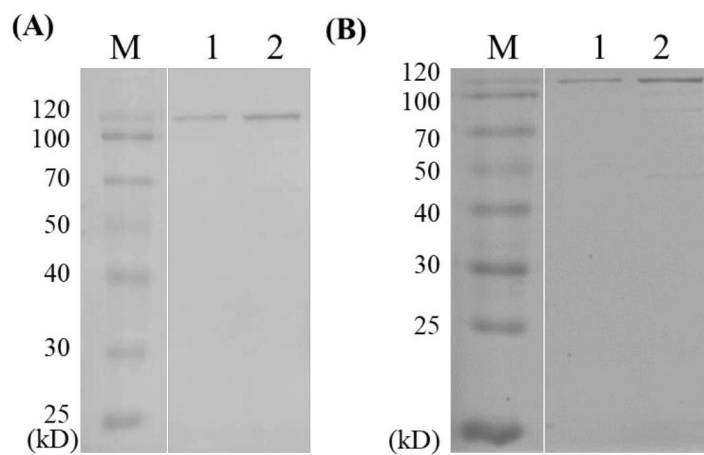
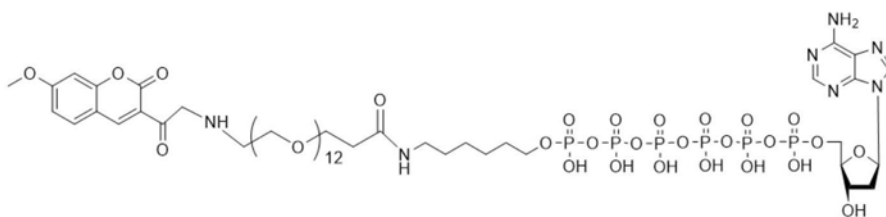
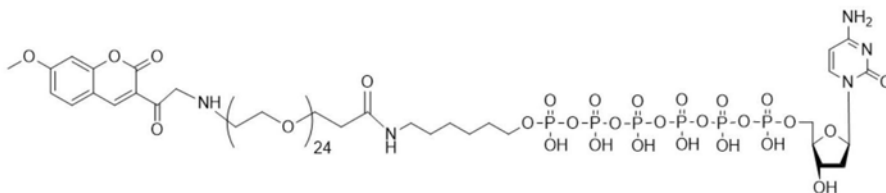


图4

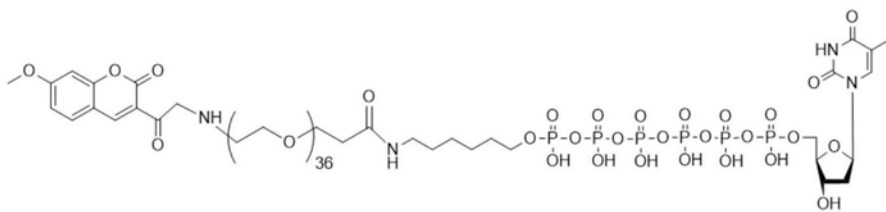
(A) CP12-dA6P: Coumarin-PEG12-dA6P



(B) CP24-dC6P: Coumarin-PEG24-dC6P



(C) CP36-dT6P: Coumarin-PEG36-dT6P



(D) CP44-dG6P: Coumarin-PEG44-dG6P

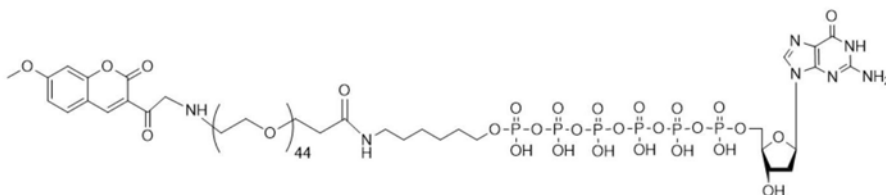


图5



图6

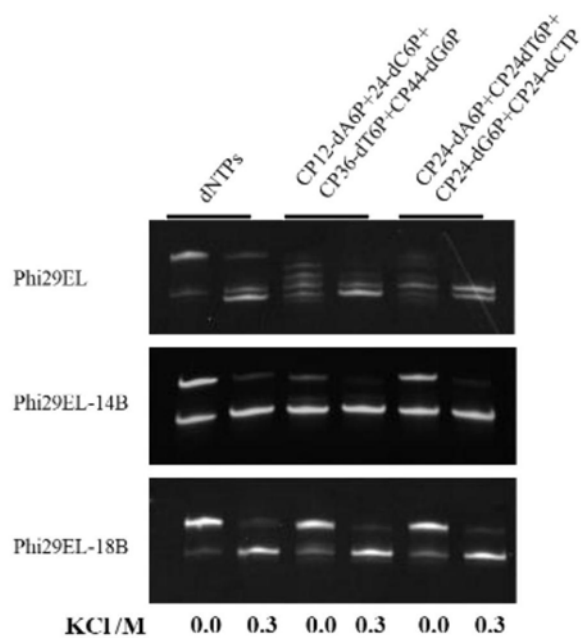


图7