



(12) 发明专利申请

(10) 申请公布号 CN 116555216 A

(43) 申请公布日 2023. 08. 08

(21) 申请号 202210107337.7

C12N 1/21 (2006.01)

(22) 申请日 2022.01.28

C12P 19/34 (2006.01)

C12R 1/19 (2006.01)

(71) 申请人 中国科学院天津工业生物技术研究所

地址 300308 天津市滨海新区空港经济区
西七道32号

(72) 发明人 江会锋 逯晓云 李从雨 程健
卢丽娜 姜倩倩

(74) 专利代理机构 北京知元同创知识产权代理
事务所(普通合伙) 11535

专利代理师 刘元霞

(51) Int. Cl.

C12N 9/12 (2006.01)

C12N 15/54 (2006.01)

C12N 15/70 (2006.01)

权利要求书2页 说明书24页

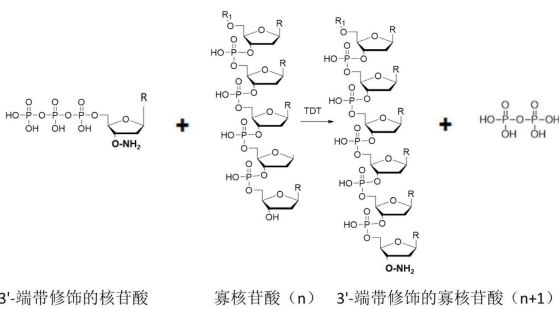
序列表9页 附图2页

(54) 发明名称

可控合成单链DNA的末端转移酶变体及应用

(57) 摘要

本发明公开了用于从头合成具有受控的序列的多核苷酸并且不依赖模板的情况下高效可控的合成核酸分子的多种末端脱氧核苷酸转移酶(TdT)及其变体,发现TdT催化结构域的一些氨基酸残基可以被特异性修饰以提高这样的修饰的TdT合成多核苷酸的能力。



1. 末端脱氧核苷酸转移酶 (TdT) 的变体, 其特征在于, 所述突变体包含选自Y178、F186、I210、I228、V302、D324、R353、D324、T330、G332、F333、R335、K337、I329、I339、G340、H341、I343、R435、G452、W453、T454、G455、S456、R457、N477、H478中至少一个位置处的残基或功能等同的残基的至少一个突变, 所示位置通过与SEQ ID NO.1比对确定。

优选地, 所述突变体的氨基酸序列包含选自Y178、F186、I210、I228、V302、T330、R335、K337、I339、G340、R435、T454、G455、S456、R457、N477、H478中至少一个位置处的残基或功能等同的残基的至少一个突变。

更优选地, 所述突变体的氨基酸序列包含选自Y178A、F186R、I210L、I228L、V302G、T330A、R335A、R335L、K337G、K337L、I339V、G340G、G340L、R435L、T454A、G455V、S456V、R457G、N477L、H478G中至少一个位置处的残基或功能等同的残基的至少一个突变。

优选地, 所述突变体的氨基酸序列包含Y178、F186、I210、I228、V302、D324、R353、D324、T330、G332、F333、R335、K337、I329、I339、G340、H341、I343、R435、G452、W453、T454、G455、S456、R457、N477、H478、N477中任意两个位点的氨基酸残基的突变。

优选地, 所述突变体的氨基酸序列包含Y178、F186、I210、I228、V302、T330、R335、K337、I339、G340、R435、T454、G455、S456、R457、N477、H478中任意两个位点的氨基酸残基的突变。

更优选地, 所述突变体的氨基酸序列包含Y178、F186、I210、I228、V302、T330、R335、K337、I339、G340、R435、T454、G455、S456、R457、N477、H478中两个位点的氨基酸残基的突变。

优选地, 所述突变体的氨基酸序列包含如下两个位点的突变: Y178/F186、I210/I228、R335/K337、I339/G340、S456/R457。

优选地, 所述突变体的氨基酸序列包含如下两个位点的突变: Y178A/F186R、I210L/I228L、R335A/K337G、R335L/K337L、I339V/G340V、S456V/R457L。

优选地, 所述变体与SEQ ID NO.1或功能等同序列具有至少80%的同一性, 更优选地与SEQ ID NO.1或功能等同序列具有至少85%、90%、95%、96%、97%、98%或99%的同一性。

2. 根据权利要求1的变体, 其特征在于, 所述变体还包含在序列330位到343位X1X2X3X4RX5GX6NX7X8X9DX10 (SEQ ID NO:1) 的至少半保守区域中的残基的至少一个突变, 其中,

X1表示选自A、I、L、V、P、Q、T、K、E的残基,

X2表示选自G、L、V、F、Y、C、H、R、D的残基,

X3表示选自A、I、L、V、P、Q、T、K、E的残基,

X4表示选自G、L、V、F、Y、C、H、R、D的残基,

X5表示选自A、I、L、V、P、Q、T、K、E的残基,

X6表示选自G、L、V、F、Y、C、H、R、D的残基,

X7表示选自A、I、L、V、P、Q、T、K、E的残基,

X8表示选自G、L、V、F、Y、C、H、R、D的残基,

X9表示选自A、I、L、V、P、Q、T、K、E的残基,

X10表示选自G、L、V、F、Y、C、H、R、D的残基。

3. 根据权利要求1或2的变体, 其特征在于, 还包含在序列449位到460位ALLX1X2X3X4X5X6QFG的至少半保守区域中的残基的至少一个突变, 其中,

X1表示选自A、I、L、V、P、Q、T、K、E的残基，

X2表示选自G、L、V、F、Y、C、H、R、D的残基，

X3表示选自A、I、L、V、P、Q、T、K、E的残基，

X4表示选自G、L、V、F、Y、C、H、R、D的残基，

X5表示选自A、I、L、V、P、Q、T、K、E的残基，

X6表示选自G、L、V、F、Y、C、H、R、D的残基。

优选地，所述变体包括在SEQ ID NO:1基础上增加或删除部分氨基酸，依旧有利用3' O-修饰的核苷酸合成核酸的功能的变体。优选，所述变体的序列为SEQ ID NO.2。

优选地，所述TdT的变体，包括不会影响蛋白活性的变化，如在上述变体或功能等同序列的N末端和/或C末端处包含标签序列，诸如MBP标签、HIS标签、GST等标签。优选，所述变体的序列为SEQ ID NO.3或SEQ ID NO.4。

优选地，TdT的变体能够掺入修饰的核苷酸，优选地，掺入修饰的3' O-核苷酸，并且更优选地掺入3' O-封闭的核苷酸。

所述变体包含选自表1中公开的取代的组合的至少一种取代的组合或功能等同残基。

4. 一种核酸分子，所述核酸分子编码如在权利要求1至3中任一项中的变体。

5. 一种表达载体，所述表达载体包含权利要求4所述的核酸分子。

6. 一种宿主细胞，所述宿主细胞包含权利要求4所述的核酸分子或权利要求5所述的表达载体。

7. 用于产生如权利要求1至3中任一项中所述变体的方法，其中根据权利要求6所述的宿主细胞在允许编码所述变体的核酸表达的培养条件下被培养，并且其中所述变体任选地被回收。

8. 如权利要求1至3中任一项中所述的变体用于在没有模板的情况下用3' O-修饰的核苷酸合成核酸分子的用途。

9. 用于在没有模板的情况下合成核酸分子的方法，所述方法包括使核酸引物与至少一种核苷酸，优选地至少一种3' O-修饰的核苷酸和如权利要求1-3中任一项中所述变体二者接触的步骤。

10. 一种用于进行核苷酸掺入反应的试剂盒、酶或组合物，其特征在于，包含如权利要求1至9中任一项中所定义的TdT的变体。

优选地，还包含一种或更多种核苷酸或核酸分子。更优选地，还包含一种或更多种3' O-修饰的核苷酸。

更优选地，还任选地至少一种DNA模板或核酸引物。

可控合成单链DNA的末端转移酶变体及应用

技术领域

[0001] 本发明涉及末端脱氧核苷酸转移酶及其应用,尤其涉及可控合成单链DNA的末端转移酶变体及应用,属于生物技术领域。

背景技术

[0002] DNA是生命信息的载体,获得DNA是研究、改造、和创造生命的首要步骤。DNA合成是合成生物学的核心使能技术之一,可对基因组DNA进行大规模合成,有助于提高我们对生命体的理解、预测和调控,并赋予我们设计和改造细胞功能甚至重构人工生命的能力。

[0003] 上世纪50年代至今,有大量科研工作者尝试通过化学和酶促的方法合成DNA。首先获得成功的是化学合成技术。经过多年优化和改进,DNA化学合成经历了从柱式合成到微芯片合成的变革发展,并得到了广泛市场化应用。但是化学方法一方面,偶联效率和副反应使寡核苷酸合成长度局限于300nt以内,难以到达kb级的基因长度。因此更长片段则需通过生物组装技术拼接寡核苷酸片段,直至获得基因、染色体或基因组长度的DNA。另一方面,反应步骤繁琐、单个循环耗时较长(6-8分钟)、化学试剂消耗较多且成本较高,大量使用有毒、易燃的有机试剂,污染较大。

[0004] 因此,生物酶合成DNA成为研究的焦点。但是常规的DNA聚合酶具有模板依赖性,无法从头合成DNA。寻找非模板依赖性的DNA聚合酶成为酶法合成的首要任务。1962年Bollum首先发现末端脱氧核苷酸转移酶(TdT),并提出TdT可用于单链寡核苷酸合成。目前,酶促寡核苷酸合成采用3位加修饰基团的核苷酸为底物,该底物与引物链的5位磷酸基团反应后,由于该修饰基团无法与下一个核苷酸5位的磷酸基团反应,阻断了核苷酸在寡核苷酸链的进一步掺入。每次酶促反应只能催化一个带修饰基团的核苷酸掺入,随后采用化学或生物方法除去修饰基团,使3'端重新恢复为OH,以便进行下一轮合成循环。但该过程野生型蛋白的催化效率低,正确率达不到,因此寻找一种具有高效的催化活性和相对稳定的酶是迫在眉睫的需求。

发明内容

[0005] 为解决上述问题,本申请的发明人通过研究用于从头合成具有受控的序列的多核苷酸并且不依赖模板的情况下高效可控的合成核酸分子的多种末端脱氧核苷酸转移酶(TdT)及其变体,发现TdT催化结构域的一些氨基酸残基可以被特异性修饰以提高TdT变体合成多核苷酸的能力。

[0006] 本发明采用以下技术方案:

[0007] 本发明提供多种末端脱氧核苷酸转移酶突变体,该突变体具有催化3'端带O-氨基修饰基团的脱氧核糖核苷酸的功能,所述突变体包含选自Y178、F186、I210、I228、V302、D324、R353、D324、T330、G332、F333、R335、K337、I329、I339、G340、H341、I343、R435、G452、W453、T454、G455、S456、R457、N477、H478中至少一个位置处的残基或功能等同的残基的至少一个突变,所示位置通过与SEQ ID NO.1比对确定。

[0008] 根据本发明,“功能等同残基”、“功能等同的残基”意指在与SEQ ID NO.1序列同源的TdT的序列中的并且具有相同功能作用的残基。功能等同残基使用序列比对,例如使用Mutalin在线比对软件(<http://multalin.toulouse.inra.fr/multalin/multalin.html>; 1988, Nucl. Acids Res., 16(22), 10881-10890)来鉴定。此外,可以通过蛋白质三维结构比对确定功能等同残基。比对后,功能等同残基处于所考虑的不同序列上的同源位置处。序列比对和功能等同残基的鉴定可以在任何TdT与其天然变体,包括种间变体之间进行。

[0009] 根据本发明,所述突变体的氨基酸序列包含选自Y178、F186、I210、I228、V302、T330、R335、K337、I339、G340、R435、T454、G455、S456、R457、N477、H478中至少一个位置处的残基或功能等同的残基的至少一个突变。

[0010] 根据本发明,所述突变体的氨基酸序列包含选自Y178A、F186R、I210L、I228L、V302G、T330A、R335A、R335L、K337G、K337L、I339V、G340G、G340L、R435L、T454A、G455V、S456V、R457G、N477L、H478G中至少一个位置处的残基或功能等同的残基的至少一个突变。

[0011] 根据本发明,所述突变体的氨基酸序列包含Y178、F186、I210、I228、V302、D324、R353、D324、T330、G332、F333、R335、K337、I329、I339、G340、H341、I343、R435、G452、W453、T454、G455、S456、R457、N477、H478中任意两个位点的氨基酸残基的突变。

[0012] 根据本发明,所述突变体的氨基酸序列包含Y178、F186、I210、I228、V302、T330、R335、K337、I339、G340、R435、T454、G455、S456、R457、N477、H478中任意两个位点的氨基酸残基的突变。

[0013] 根据本发明,所述突变体的氨基酸序列包含Y178、F186、I210、I228、V302、T330、R335、K337、I339、G340、R435、T454、G455、S456、R457、N477、H478中两个位点的氨基酸残基的突变。

[0014] 根据本发明,所述突变体的氨基酸序列包含如下两个位点的突变:Y178/F186、I210/I228、R335/K337、I339/G340、S456/R457。

[0015] 根据本发明,所述突变体的氨基酸序列包含如下两个位点的突变:Y178A/F186R、I210L/I228L、R335A/K337G、R335L/K337L、I339V/G340V、S456V/R457L。

[0016] 根据本发明,TdT的变体具有以上描述的取代或取代的组合,并且与SEQ ID NO.1或功能等同序列具有至少80%的同一性,优选地与SEQ ID NO.1或功能等同序列具有至少85%、90%、95%、96%、97%、98%或99%的同一性。

[0017] 根据本发明,如以上公开的TdT的所有变体都能够既在没有模板的情况下合成核酸片段又能够将修饰的核苷酸掺入到核酸片段中。有利地,所述变体与SEQ ID NO.1的TdT相比具有增加的将修饰的核苷酸,优选地3'-修饰的核苷酸掺入到核酸片段中的能力。

[0018] 在以上描述的一些实施方案中,变体TdT在掺入3'-修饰的核苷三磷酸方面的效率是序列SEQ ID NO.1的野生型TdT的效率的至少110%;在其他实施方案中,变体TdT在掺入3'-修饰的核苷三磷酸方面的效率是序列SEQ ID NO.1的野生型TdT的效率的至少150%;在其他实施方案中,变体TdT在掺入3'-修饰的核苷三磷酸方面的效率是序列SEQ ID NO.1的野生型TdT的效率的至少200%。在一个具体实施方案中,变体还包含在序列330位到343位X1X2X3X4RX5GX6NX7X8X9DX10的至少半保守区域中的残基的至少一个突变,其中,

[0019] X1表示选自A、I、L、V、P、Q、T、K、E的残基,

[0020] X2表示选自G、L、V、F、Y、C、H、R、D的残基,

- [0021] X3表示选自A、I、L、V、P、Q、T、K、E的残基，
- [0022] X4表示选自G、L、V、F、Y、C、H、R、D的残基，
- [0023] X5表示选自A、I、L、V、P、Q、T、K、E的残基，
- [0024] X6表示选自G、L、V、F、Y、C、H、R、D的残基，
- [0025] X7表示选自A、I、L、V、P、Q、T、K、E的残基，
- [0026] X8表示选自G、L、V、F、Y、C、H、R、D的残基，
- [0027] X9表示选自A、I、L、V、P、Q、T、K、E的残基，
- [0028] X10表示选自G、L、V、F、Y、C、H、R、D的残基。
- [0029] 在另一个具体实施方案中，还包含在序列449位到460位ALLX1X2X3X4X5X6QFG的至少半保守区域中的残基的至少一个突变，其中，
- [0030] X1表示选自A、I、L、V、P、Q、T、K、E的残基；
- [0031] X2表示选自G、L、V、F、Y、C、H、R、D的残基；
- [0032] X3表示选自A、I、L、V、P、Q、T、K、E的残基；
- [0033] X4表示选自G、L、V、F、Y、C、H、R、D的残基；
- [0034] X5表示选自A、I、L、V、P、Q、T、K、E的残基；
- [0035] X6表示选自G、L、V、F、Y、C、H、R、D的残基。
- [0036] 根据本发明，所述变体包括在SEQ ID NO.1基础上增加或删除部分氨基酸，依旧有利用3' O-修饰的核苷酸合成核酸的功能的变体。优选，所述变体的序列为SEQ ID NO.2。
- [0037] 根据本发明，所述TdT的变体，包括不会影响蛋白活性的变化，如在上述变体或功能等同序列的N末端和/或C末端处包含标签序列，诸如MBP标签、HIS标签、GST等标签。优选，所述变体的序列为SEQ ID NO.3或SEQ ID NO.4。
- [0038] 根据本发明，TdT的变体能够掺入修饰的核苷酸，优选地掺入修饰的3' O-核苷酸，并且更优选地掺入3' O-封闭的核苷酸。
- [0039] 在本发明的上下文中，表述“修饰的核苷酸”是指含有与三个磷酸基团结合的核苷（即附接到脱氧核糖或核糖糖分子上的碱基）的分子，所述核苷在其一个末端（2'、3'、5'或碱基）上具有至少一个另外的基团。所述另外的基团通过阻止任何磷酸二酯键（3' O-修饰、2'或2' O-修饰）的形成，或通过空间上阻止聚合酶附接到在其3'末端包含这样的修饰的核苷酸（5'或碱基修饰）的任何核酸片段来阻断核苷酸的进一步添加。此外，所述另外的基团有利地具有可逆性质，允许该基团通过特定的裂解反应被去除。
- [0040] 根据本发明，所述变体包含选自表1中公开的取代的组至少一种取代的组合或功能等同残基。
- [0041] 本发明的另一个目的是提供多种具有催化非天然3'-OH修饰的核苷酸的末端脱氧核苷酸转移酶变体。所述末端转移酶变体无自然解析的晶体结构，三维结构为通过Alpha fold2进行同源建模获得。
- [0042] 本发明还提供编码如上文定义的TdT的变体的核酸分子。
- [0043] 本发明还提供一种表达载体，其包含本发明所述的核酸分子。
- [0044] 本发明还提供一种宿主细胞，其特征在于，所述宿主细胞包含本发明所述的核酸分子或者所述表达载体。
- [0045] 本发明进一步提供用于产生TdT的变体的方法，其中，如上文定义的宿主细胞在允

许编码所述变体的核酸表达的培养条件下被培养,并且其中变体任选地被回收。

[0046] 本发明还涉及TdT的变体用于在没有模板的情况下用一种或更多种3'-O-修饰的核苷酸合成核酸分子的用途。

[0047] 在一些实施方案中,这样的方法包括以下步骤:(a)提供包含具有游离3'-羟基的寡核苷酸的起始片段;(b)在酶促延伸条件下,在3'-O-可逆封闭的核苷的存在下,使本发明的TdT变体与起始片段或延伸的起始片段反应。

[0048] 在一些实施方案中,这样的方法还包括步骤(c)使延伸的起始片段去封闭以形成具有游离3'-羟基的延伸的起始片段,和(d)重复步骤(b)和(c),直到合成预定序列的核酸分子。

[0049] 本发明的又一个目的是提供一种不需要模板的合成核酸分子的方法,该方法包括使核酸引物与至少一种核苷酸,优选地至少一种3'-O-修饰的核苷酸和根据本发明的TdT的变体接触的步骤。

[0050] 本发明还提供了一种用于进行核苷酸掺入反应的试剂盒、酶或核苷酸组合物,其特征在于:包含本发明的TdT的变体。

[0051] 根据本发明,所述试剂盒、酶或核苷酸组合物还包括:一种或多种核苷酸(优选地一种或多种3'-O-修饰的核苷酸),或核苷分子。

[0052] 根据本发明,所述修饰的核苷酸被荧光标记以允许其检测。

[0053] 进一步地,本发明所述试剂盒、酶或核苷酸组合物还包括:任选地至少一种DNA模板分子或核酸引物。

[0054] 优选地,所述进行核苷酸掺入反应的试剂盒、酶或核苷酸组合物,包含:1)根据本发明提供的TdT的变体;2)一种或多种核苷酸(优选地一种或多种3'-O-修饰的核苷酸);3)任选地至少一种核酸引物。

[0055] 有益的效果:

[0056] 1)本发明提供了多种末端脱氧核苷酸转移酶突变体,可以在不依赖模板的情况下高效可控的合成核酸分子,并惊讶的发现多种突变体的效率显著高于未改造前的野生型末端脱氧核苷酸转移酶。

[0057] 2)本发明提供的多种末端脱氧核苷酸转移酶突变体,可以应用于单链DNA的可控精准合成。

附图说明

[0058] 图1为TdT及变体催化反应示意图。

[0059] 图2为部分TdT变体催化3端带可逆终止修饰基团的相对活性。

[0060] 图3为应用TdT变体完成10个核苷酸合成循环的变性PAGE结果。

[0061] 发明描述

[0062] DNA聚合酶家族根据其序列同源性和晶体结构被分为7个家族。其中,PolX家族的聚合酶代表了从复制聚合酶到末端转移酶的宽范围的聚合酶。来自PolX家族的聚合酶存在于非常宽范围的真核生物体中。来自PolX家族的聚合酶参与众多种生物过程,并且特别是DNA损伤修复机制或错误纠正机制。PolX家族再分组为聚合酶 β (Pol β)、 μ (Pol μ)、 λ (Pol λ)、来自酵母的IV (Pol IV) 和末端脱氧核苷酸转移酶(TdT)。TdT天然参与DNA修复和维持机制。

特别地,即使在不存在模板链的情况下,TdT也具有保存核苷酸聚合活性的独特能力。在特定条件下并且使用天然核苷酸,在不存在任何互补链的情况下,TdT能够使DNA片段延伸几百个核苷酸。然而,野生型TdT完全不能有效掺入糖修饰的核苷酸。

[0063] 因此,本发明的目的是提供具有被靶向的突变的TdT的变体,该被靶向的突变允许TdT的变体在核苷酸片段的合成期间将修饰的核苷酸掺入到所述核酸片段中。更特别地,本发明人已鉴定了可以单独或组合地被有利地取代的特定的氨基酸残基,以提高酶合成不同长度且具有预定序列的核酸片段的能力,包括通过使用修饰的核苷酸合成不同长度且具有预定序列的核酸片段的能力。

[0064] 定义

[0065] 如其中使用的,术语“突变体”和“变体”可以互换使用,以指来源于SEQ ID NO.1并且在一个或多个(例如几个)位置处包含修饰或改变,即取代、插入和/或缺失并且既具有在没有模板的情况下的聚合酶活性又具有掺入一个或多个修饰的终止剂核苷酸的能力的多肽。变体可以通过本领域熟知的多种技术获得。特别地,用于改变编码野生型蛋白的DNA序列的技术的实例包括但不限于定点诱变、随机诱变和合成寡核苷酸构建。诱变活性包括对蛋白或者在本发明的情况下对聚合酶的序列中的一个或多个氨基酸的缺失、插入或取代。被靶向的氨基酸可以伴随或分布在聚合酶的整个序列上。例如,特定的基序或结构特征可以被靶向。

[0066] 如本文关于氨基酸的位置使用的术语“修饰”或“改变”意指,特定位置中的氨基酸与野生型蛋白的氨基酸相比被修饰。

[0067] “取代”意指一个氨基酸残基被另一个氨基酸残基替代。优选地,术语“取代”是指一个氨基酸残基被选自以下的另一个氨基酸残基替代:20种标准的天然存在的氨基酸残基、罕见的天然存在的氨基酸残基(例如,羟脯氨酸、羟赖氨酸、别羟基赖氨酸、6-N-甲基赖氨酸、N-乙基甘氨酸、N-甲基甘氨酸、N-乙基天冬酰胺、别-异亮氨酸、N-甲基异亮氨酸、N-甲基缬氨酸、焦谷氨酰胺、氨基丁酸、鸟氨酸、正亮氨酸、正缬氨酸)和通常经合成产生的非天然存在的氨基酸残基(例如,环己基丙氨酸)。优选地,术语“取代”是指一个氨基酸残基被选自20种标准的天然存在的氨基酸残基的另一个氨基酸残基替代。符号“+”指示取代的组合。

[0068] 在本文中,氨基酸由根据以下命名法的其单字母或三字母代码代表:A:丙氨酸(Ala);C:半胱氨酸(Cys);D:天冬氨酸(Asp);E:谷氨酸(Glu);F:苯丙氨酸(Phe);G:甘氨酸(Gly);H:组氨酸(His);I:异亮氨酸(Ile);K:赖氨酸(Lys);L:亮氨酸(Leu);M:甲硫氨酸(Met);N:天冬酰胺(Asn);P:脯氨酸(Pro);Q:谷氨酰胺(Gln);R:精氨酸(Arg);S:丝氨酸(Ser);T:苏氨酸(Thr);V:缬氨酸(Val);W:色氨酸(Trp)和Y:酪氨酸(Tyr)。

[0069] 在本文件中,使用以下术语指示取代:L238A表示亲本序列的位置238处的氨基酸残基(亮氨酸,L)被改变为丙氨酸(A)。A132V/I/M表示亲本序列的位置132处的氨基酸残基(丙氨酸,A)被以下氨基酸之一取代:缬氨酸(V)、异亮氨酸(I)或甲硫氨酸(M)。取代可以是保守取代或非保守取代。保守取代的实例在碱性氨基酸(精氨酸、赖氨酸和组氨酸)、酸性氨基酸(谷氨酸和天冬氨酸)、极性氨基酸(谷氨酰胺、天冬酰胺和苏氨酸)、疏水性氨基酸(甲硫氨酸、亮氨酸、异亮氨酸、半胱氨酸和缬氨酸)、芳香族氨基酸(苯丙氨酸、色氨酸和酪氨酸)和小氨基酸(甘氨酸、丙氨酸、和丝氨酸)的组内。

[0070] 如本文使用的,术语“序列同一性”或“同一性”是指两个多肽序列之间匹配(相同

氨基酸残基)的数目(或以百分比%表示的分数)。序列同一性通过在对齐以便使重叠和同一性最大化同时使序列空位最小化时比较序列来确定。特别地,序列同一性可以根据两个序列的长度使用多种数学全局或局部比对算法中的任何一种来确定。相似长度的序列优选地使用全局比对算法(例如,Needleman和Wunsch算法;Needleman和Wunsch,1970)进行比对,该全局比对算法在整个长度上最佳地对齐序列,而实质上不同长度的序列优选地使用局部比对算法(例如,Smith和Waterman算法(Smith和Waterman,1981)或Altschul算法(Altschul等人,1997;Altschul等人,2005年)进行比对。出于确定氨基酸序列同一性百分比的目的的比对可以以本领域技术范围内的多种方式实现,例如,使用互联网网站诸如<http://blast.ncbi.nlm.nih.gov/>或<http://www.ebi.ac.uk/Tools/emboss/>上可得的公众可得的计算机软件。本领域技术人员可以确定用于测量比对的适当参数,包括在所比较的序列的全长上实现最大比对所需的任何算法。出于本文的目的,氨基酸序列同一性%值是指使用成对序列比对程序EMBOSS Needle生成的值,该程序使用Needleman-Wunsch算法创建两个序列的最佳全局比对,其中所有搜索参数被设置为默认值,即评分矩阵=BLOSUM62,空位开放=10,空位延伸=0.5,末端空位罚分=假,末端空位开放=10,且末端空位延伸=0.5。

[0071] 本文中,术语“肽”、“多肽”、“蛋白”、“酶”是指通过肽键连接的氨基酸链,而与形成所述链的氨基酸的数目无关。

[0072] 除非另有说明,本申请中公开的位置参考SEQ ID NO.1中列出的氨基酸序列编号。

[0073] TdT的变体

[0074] 本发明提供了TdT酶的变体,其可用于在不使用模板链的情况下合成预定序列的多核苷酸,诸如DNA或RNA。本发明的TdT变体允许在酶介导的多核苷酸合成方法中使用修饰的核苷酸,并且特别是使用3' O-修饰的核苷酸。

[0075] 在本发明的上下文中,“修饰的末端脱氧核苷酸转移酶”、“修饰的TdT”、“末端脱氧核苷酸转移酶的变体”和“TdT的变体”是指与TdT的氨基酸序列共有至少25%的同一性。优选地,TdT的变体与SEQ ID NO.1共有至少40%的同一性。

[0076] 已知TdT由从N末端至C末端的不同结构域构成,分别对应于核定位结构域(NLS)、BRCT样结构域和催化结构域(C-TdT)。

[0077] 本发明的变体根据其在特定残基上的突变来描述,其位置通过与酶序列SEQ ID NO.1比对或参考酶序列SEQ ID NO.1来确定。然而,在本发明的上下文中,具有与SEQ ID NO.1功能等同序列的任何变体也是本发明的一部分。以相同方式,在功能等同残基上携带相同突变的任何变体也是本发明的一部分。

[0078] 在本发明的上下文中,“功能等同序列”是指与SEQ ID NO.1同源的TdT的序列。“功能等同残基”意指在与SEQ ID NO.1序列同源的TdT的序列中的并且具有相同功能作用的残基。功能等同残基使用序列比对,例如使用Multalin在线比对软件(<http://multalin.toulouse.inra.fr/multalin/multalin.html>;1988,Nucl.Acids Res.,16(22),10881-10890)来鉴定。比对后,功能等同残基处于所考虑的不同序列上的同源位置处。序列比对和功能等同残基的鉴定可以在任何TdT与其天然变体,包括种间变体之间进行。

[0079] TdT可以发现于许多其他生物体或微生物中。所有这些TdT都是用于进行本发明的良好候选者。特别地,改变特定TdT序列以赋予所述聚合酶增加的掺入修饰的核苷酸的能力

的修饰,可以靶向任何其他TdT序列。因此,本文参考SEQ ID NO.1,并且更特别地,参考对应于SEQ ID NO.1的氨基酸残基330至343、449至460的突变或组合,可以转座至任何其他TdT序列。

[0080] 本发明更特别地提供了如表1中列出的至少一个取代或取代的组的TdT的变体。本发明的变体至少包含左栏中列出的且称为“可变突变”的氨基酸取代或功能等同残基,以及任选地右栏中列出的且称为“任选的恒定突变”的一种或两种取代的组合或功能等同序列。

[0081] 根据本发明,TdT的变体具有以上描述的取代或取代的组合,并且与SEQ ID NO.1或功能等同序列具有至少80%的同一性,优选地与SEQ ID NO.1或功能等同序列具有至少85%、90%、95%、96%、97%、98%或99%的同一性。

[0082] 根据本发明,如以上公开的TdT的所有变体都能够既在没有模板的情况下合成核酸片段又能够将修饰的核苷酸掺入到核酸片段中。有利地,所述变体与SEQ ID NO.1的TdT相比具有增加的将修饰的核苷酸,优选地3' O-修饰的核苷酸掺入到核酸片段中的能力。

[0083] 在以上描述的一些实施方案中,变体TdT在掺入3' O-修饰的核苷三磷酸方面的效率是序列SEQ ID NO.1的野生型TdT的效率的至少110%;在其他实施方案中,变体TdT在掺入3' O-修饰的核苷三磷酸方面的效率是序列SEQ ID NO.1的野生型TdT的效率的至少150%;在其他实施方案中,变体TdT在掺入3' O-修饰的核苷三磷酸方面的效率是序列SEQ ID NO.1的野生型TdT的效率的至少200%。

[0084] 另外的修饰

[0085] 在实施方案中,TdT的变体还包括在其N末端、C末端或两个末端中的任何类型的标签肽(tagging peptide),诸如His-标签序列。所述标签肽可以用于纯化、鉴定、增加表达、分泌性(secretability)或增加催化活性。将理解,这样的不同的标签在文献中被广泛描述,并且因此本发明涵盖了技术人员已知的所有标签。

[0086] 本发明的变体还可以在蛋白的N末端区域和/或C末端区域包含一个或更多个外源或异源特征,用于例如重组聚合酶的纯化。

[0087] 本发明的变体还可以包含位置330至343、449至460之间的残基或功能等同残基(其中位置通过参考SEQ ID NO.1中列出的氨基酸序列编号)的取代。在另一种特定实施方案中,TdT的变体至少包含氨基酸序列SEQ ID NO.1或功能等同序列的取代。

[0088] 修饰的核苷酸

[0089] 根据本发明,TdT的变体能够掺入修饰的核苷酸,优选地掺入修饰的3' O-核苷酸,并且更优选地掺入3' O-封闭的核苷酸。

[0090] 在本发明的上下文中,表述“修饰的核苷酸”是指含有与三个磷酸基团结合的核苷(即附接到脱氧核糖或核糖糖分子上的碱基)的分子,所述核苷在其一个末端(2'、3'、5'或碱基)上具有至少一个另外的基团。所述另外的基团通过阻止任何磷酸二酯键(3' O-修饰、2'或2' O-修饰)的形成,或通过空间上阻止聚合酶附接到在其3'末端包含这样的修饰的核苷酸(5'或碱基修饰)的任何核酸片段来阻断核苷酸的进一步添加。此外,所述另外的基团有利地具有可逆性质,允许该基团通过特定的裂解反应被去除。

[0091] 核苷酸或核苷三磷酸包括脱氧腺苷三磷酸(dTAP)、脱氧鸟苷三磷酸(dGTP)、脱氧胞苷三磷酸(dCTP)或脱氧胸苷三磷酸(dTTP),作为含有脱氧核糖的核苷酸的实例。腺苷三

磷酸(ATP)、鸟苷三磷酸(GTP)、胞苷三磷酸(CTP)或尿苷三磷酸(UTP)是含有核糖的核苷三磷酸的另外的实例。其他类型的核苷,诸如天然存在的修饰的核苷和人工核苷可以与三个磷酸结合形成核苷三磷酸。在特定实施方案中,修饰的核苷酸是3'-O-封闭的核苷酸,其包含可逆地附接到核苷三磷酸的3'末端的基团,以阻止进一步的核苷酸添加。所述基团可以具有不同的化学性质,诸如叠氮基甲基、氨基氧基和烯丙基。

[0092] 在一些实施方案中,修饰的核苷酸包括包含嘌呤碱基或嘧啶碱基和以下核糖或脱氧核糖糖部分的修饰的核苷酸或核苷分子,所述核糖或脱氧核糖糖部分具有与其共价地附接的可去除的3'-OH封闭基团,使得3'碳原子已附接可逆结构(-O-Z)的基团,

[0093] 其中-Z是-C(R')₂-O-R''、-C(R')₂-N(R'')₂、-C(R')₂-N(H)R''、-C(R')₂-S-R''和-C(R')₂-F的任何一种,其中每个R''是可去除的保护基团或是可去除的保护基团的一部分;每个R'独立地是氢原子、烷基、取代的烷基、芳基烷基、烯基、炔基、芳基、杂芳基、杂环、酰基、氰基、烷氧基、芳基氧基、杂芳基氧基或酰胺基团,或通过连接基团附接的可检测标记物;条件是在一些实施方案中,这样的取代基具有至多10个碳原子和/或至多5个氧或氮杂原子;或(R')₂代表式=C(R'')₂的次烷基基团(alkylidene),其中每个R''可以是相同的或不同的,并且选自包含氢和卤素原子以及烷基基团的组,条件是在一些实施方案中,每个R''的烷基具有1至3个碳原子;并且其中分子可以反应以产生中间体,所述中间体中每个R''被交换为H,或其中Z是-(R')₂-F,F被交换为OH、SH或NH₂,优选地交换为OH,所述中间体在水性条件下解离以提供具有游离3'-OH的分子;条件是当Z是-C(R')₂-S-R''时,两个R'基团均不是H。在某些实施方案中,修饰的核苷酸或核苷的R'是烷基或取代的烷基,条件是这样的烷基或取代的烷基具有1至10个碳原子和0至4个氧或氮杂原子。在某些实施方案中,修饰的核苷酸或核苷的-Z是式-C(R')₂-N₃。在某些实施方案中,Z是叠氮基甲基。

[0094] 在一些实施方案中,Z是具有或不具有杂原子的分子量为200或更小的可裂解有机部分。在其他实施方案中,Z是具有或不具有杂原子的分子量为100或更小的可裂解有机部分。在其他实施方案中,Z是具有或不具有杂原子的分子量为50或更小的可裂解有机部分。

[0095] 在另外的特定实施方案中,“3'-O修饰的核苷酸”是指在3'末端处携带3'-O-甲基、3'-叠氮基、3'-O-叠氮基甲基、3'-O-氨基、3'-氨基氧基或3'-O-烯丙基基团的核苷三磷酸。在另外的实施方案中,3'-封闭的核苷三磷酸被3'-O-叠氮基甲基、3'-氨基氧基或3'-O-烯丙基基团封闭。在其他实施方案中,“3'-O修饰的核苷酸”是指在3'末端处携带酯、醚、腈、磷酸酯(phosphates)、碳酸酯(carbonate)、氨基甲酸酯、羟胺、硼酸酯(borate)、硝酸酯(nitrate)、糖、磷酰胺、磷酰胺酸酯、苯磺酸酯、硫酸酯、砒类或氨基酸的核苷三磷酸。在一些实施方案中,前述3'-O-封端基团具有100或更小的分子量。

[0096] 在其他实施方案中,本发明的3'-O-封端基团包括甲基、3'-O-(2-硝基苄基)、烯丙基、胺、叠氮基甲基、叔丁氧基乙氧基或炔丙基。

[0097] 在另外的特定实施方案中,“3'-O修饰的核苷酸”是指具有终止剂效应物修饰基团的核苷三磷酸,诸如在W02016034807中描述的那些。

[0098] 有趣的是,本发明的变体与野生型TdT相比表现出对修饰的核苷酸的增加的亲和力,并且从而表现出增加的在核酸合成期间将这样的修饰的核苷酸掺入核酸序列的能力。更特别地,本发明的变体能够使用修饰的3'-O-核苷酸(并且更特别地,3'-O-封闭核苷酸)并高效地将其掺入到核酸序列中,这在野生型小鼠TdT的情况中是不可能的(参见Knapp等人

Chem. Eur. J., 2011, 17:2903)。

[0099] 根据特定方面,本发明涉及能够在核酸酶促合成方法中与修饰的核苷酸,特别地与3' O-修饰的核苷酸(例如,3' O-封闭的核苷酸)一起工作,并且具有产生长长度核酸分子或核酸分子衍生物的能力的TdT的变体。

[0100] 核酸的酶促合成

[0101] 本发明的目的是提供可以用于核酸合成的TdT的变体,诸如在Ybert等人W02015/159023;Jensen等人,Biochemistry, 57:1821-1832 (2018);Hiatt等人,美国专利第5808045号中描述的。更特别地,本发明的目的是提供适于向起始核酸链添加修饰的核苷酸的TdT的变体。然后可以除去封闭基团,以允许添加新的修饰的核苷酸。

[0102] 根据本发明,通过使用本发明的变体,实施包括添加和去保护的连续循环是可能的。因此,该方法将允许通过添加可逆修饰的核苷酸和进一步去除封闭基团的多个循环,以允许起始核酸链受控制地延伸为限定的序列。

[0103] 本发明预期了在任何酶促核酸合成方法中使用根据本发明的修饰的TdT。

[0104] 本发明的又一个目的是提供一种用于在没有模板的情况下合成核酸分子的方法,该方法包括使核酸引物与至少一种核苷酸,优选地至少一种3' O-修饰的核苷酸和本发明的变体二者接触的步骤。

[0105] 本发明预期了酶促核酸合成方法的概念。在这样的方法中,核酸分子在不存在任何模板链的情况下从头合成。因此,核苷酸的有序序列与起始核酸片段借助于本发明的变体偶联。将理解,每个核苷酸与生长的核酸链的定量偶联和更普遍的高偶联效率是非常重要的。还将理解,非终止剂核苷酸,诸如天然核苷酸或永久标记的核苷酸将不允许对合成的序列进行任何控制,并且将导致例如不受控制的且不期望的多聚添加(poly-addition)。

[0106] 根据特定实施方案,酶促核酸合成方法包括:

[0107] a. 提供与固体支持物连接的核酸分子;

[0108] b. 使先前的核酸分子与可逆终止剂修饰的核苷酸和根据本发明的TdT的变体反应;

[0109] 根据另一个特定实施方案,酶促核酸方法包括:

[0110] a. 提供与固体支持物连接的核酸分子;

[0111] b. 添加可逆修饰的核苷酸和根据本发明的TdT的变体;

[0112] c. 首先从固体支持物中去除一种或几种试剂;

[0113] d. 使可逆修饰的核苷酸的可逆部分反应,以使可逆修饰的核苷酸去保护,用于进一步的后续延伸;

[0114] e. 其次从固体支持物中去除一种或几种试剂;

[0115] f. 任选地并最终从固体支持物上裂解核酸分子。

[0116] 根据另一个特定实施方案,酶促核酸方法包括以以下方式细分的循环:

[0117] a. 称为第一阶段的使 X_i 个核苷酸延伸到所述片段一个末端的阶段, X 在1与5之间,优选地在1与3之间是可能的, i 是循环数,使得获得包含 $n+X_i$ 个核苷酸的片段成为可能,并且包括以下阶段:

[0118] - 将包括 n 个核苷酸的初始核酸片段或延伸过程中的核酸片段的第一末端附接到第一支持物上的第一附接阶段,

- [0119] -添加TdT的变体所需的试剂的阶段,
- [0120] -TdT的变体向所述核酸片段的第二末端添加 X_i 个核苷酸的阶段, X 在1与5之间,优选地1与3之间, i 是循环数,
- [0121] -从反应介质中去除不期望的试剂的任意的阶段,
- [0122] -使包含 $n+X_i$ 个核苷酸的所述片段从所述第一支持物上脱离的阶段,
- [0123] -转移包含 $n+X_i$ 个核苷酸的所述片段的第一转移阶段,
- [0124] b. 称为第二阶段的纯化具有包含 $n+X_i$ 个核苷酸的正确序列的片段的阶段,包括以下连续阶段:
- [0125] -将包含 $n+X_i$ 个核苷酸的所述片段附接到第二支持物上的第二附接阶段,所述附接通过携带在第一阶段期间添加的 X_i 个核苷酸的末端进行,
- [0126] -去除没有被附接到第二支持物上的片段的阶段,
- [0127] -使包含 $n+X_i$ 个核苷酸的所述片段从所述第二支持物上分离的阶段,
- [0128] -从反应介质中去除不期望的残留试剂的任意的阶段;
- [0129] c. 称为第三阶段的任意的对具有包含 $n+X_i$ 个核苷酸的正确序列的片段进行扩增,优选地酶促扩增,诸如通过PCR扩增的阶段,包括以下连续阶段:
- [0130] -添加扩增所需的试剂的阶段,
- [0131] -包含 $n+X_i$ 个核苷酸的片段以倍增因子 Y_i 倍增的阶段(任选地由使该过程成为可能的子阶段组成), i 是循环数, Y 在1与 4×10^{10} 之间,优选地在1与 1×10^9 之间是可能的,
- [0132] -转移包含 $n+X_i$ 个核苷酸的片段的阶段,
- [0133] 每个循环都在与酶促添加和酶促扩增相容的反应介质,诸如水性介质中进行,合成方法还包括在所有 i 个延伸循环结束时以倍增因子 Y_f 进行最终扩增的阶段。
- [0134] 在一些实施方案中,合成多核苷酸的方法包括以下步骤:(a) 提供具有游离3'-羟基的起始片段;(b) 在延伸条件下,在3'-O-封闭的核苷三磷酸存在下,使起始片段或具有游离3'-羟基的延伸中间体与本发明的变体TdT反应,以产生3'-O-封闭的延伸中间体;(c) 使延伸中间体去封闭以产生具有游离3'-羟基的延伸中间体;和(d) 重复步骤(b) 和(c),直到合成多核苷酸。
- [0135] 在一些实施方案中,合成多核苷酸的方法包括以下步骤:(a) 提供附接到固体支持物的起始片段,引发剂是具有游离3'-羟基的寡核苷酸;(b) 在延伸条件下,在3'-O-封闭的核苷三磷酸存在下,使起始片段或具有游离3'-羟基的延伸中间体与本发明的变体TdT反应,以产生3'-O-封闭的延伸中间体;(c) 洗涤固体支持物以去除未掺入的3'-O-封闭的核苷三磷酸;(d) 通过将固体支持物暴露于去封闭剂使延伸中间体去封闭,以产生具有游离3'-羟基的延伸中间体;和(e) 重复步骤(b) 和(d),直到合成多核苷酸。该方法可以包括从固体支持物上裂解完成的多核苷酸的另外的步骤。
- [0136] 在一些实施方案中,对于步骤(b),TdT催化的添加反应,酶促条件可以含有约0.20 μM 至约200 μM 的具有保护3'-羟基的可去除的封闭部分的核苷酸和约0.20 μM 至200 μM 的来源于起始底物的游离和未修饰的3'-羟基。在一些实施方案中,反应缓冲液含有约10mM至约500mM的二甲胍酸钾(potassium cacodylate)缓冲液(pH在6.5与7.5之间)和约0.01mM至约10mM的二价阳离子(例如, CoCl_2 或 MnCl_2)。其他缓冲液组合物和组分可以适于本发明的特定期望的实施方案。

[0137] 在本发明的上下文中,表述“裂解反应”是指能够裂解先前描述的可逆修饰的核苷酸上的附加基团的物质或物理条件的任何作用。本领域技术人员能够确定任何先前列出的基团的裂解反应。

[0138] 在一种实施方案中,裂解剂是化学裂解剂。在替代实施方案中,裂解剂是酶促裂解剂。

[0139] 本领域技术人员将理解,裂解剂的选择取决于所使用的3'-核苷酸封闭基团的类型。例如,三(2-羧乙基)膦(TCEP)可以用于裂解3'-O-叠氮基甲基基团,钡络合物可以用于裂解3'-O-烯丙基基团,或者亚硝酸钠可以用于裂解3'-O-氨基基团。在特定实施方案中,裂解反应包括:TCEP、钡络合物或亚硝酸钠。

[0140] 在特定实施方案中,裂解反应在附加组分诸如变性剂(例如尿素、盐酸胍、甲酰胺或甜菜碱)的存在下进行。在另外的实施方案中,裂解反应用一种或更多种缓冲液进行。本领域技术人员将理解,缓冲液的选择取决于确切的反应机理。

[0141] 本发明涉及具有以定量方式掺入修饰的核苷酸的能力的TdT的变体。“定量方式”或“定量反应”意指一种进行完全的反应,即其中反应物完全转化为产物的反应。以定量方式掺入可逆修饰的核苷酸的聚合酶是能够用所有可用的核苷酸延伸每个核酸片段,导致长度为n的所有起始片段转化为长度为n+1的片段的聚合酶。

[0142] 如本文使用的,“起始片段”是指具有游离3'末端的可以被进一步延伸的短寡核苷酸序列。在一种实施方案中,起始片段是DNA起始片段。在替代实施方案中,起始片段是RNA起始片段。

[0143] 在一种实施方案中,起始片段具有3与100个之间的核苷酸,特别是3与20个之间的核苷酸。

[0144] 在一种实施方案中,起始片段是单链的。在替代实施方案中,起始片段是双链的。

[0145] 在一种实施方案中,起始片段被固定在固体支持物上。起始片段可以通过多种方法被附接到固体支持物上,产生在片段将经历的多种酶促或合成反应条件下稳定的。

[0146] 在一种实施方案中,起始片段通过可逆相互作用部分被固定在固体支持物上,所述可逆相互作用部分诸如化学可裂解接头、抗体/免疫原性表位、生物素/生物素结合蛋白或谷胱甘肽-GST标签。在另外的实施方案中,起始片段通过化学可裂解接头被固定在固体支持物上,所述化学可裂解接头诸如二硫化物、烯丙基或叠氮化物掩蔽的半胺缩醛醚(hemiaminal ether)接头。

[0147] 在起始片段中,被固定的部分含有至少一个限制性位点。文献中描述了使用限制性酶和限制性位点在特定位点处选择性水解核酸链。任何技术人员都能够选择与起始片段裂解位点序列匹配的适当的限制性酶。

[0148] 在替代实施方案中,起始片段含有至少一个尿苷。用尿嘧啶-DNA糖基化酶(UDG)处理产生无碱基位点。用无嘌呤/无嘧啶(AP)位点内切核酸酶对适当的底物的处理将提取核酸链。

[0149] 核酸分子

[0150] 本发明的又一个目的是提供编码本发明的变体的核酸分子。如本文使用的,“核酸分子”是指核苷的聚合物。在一种实施方案中,核酸是DNA。在替代实施方案中,核酸是RNA。在替代实施方案中,核酸是XNA。

[0151] 本领域技术人员将理解,先前列出的每个核酸分子可以在构成聚合分子的核苷酸的碱基上携带修饰。这样的修饰可以是天然修饰,诸如表观遗传修饰,或非天然修饰,诸如标记物。

[0152] 在一种实施方案中,核酸分子是携带天然存在的表观遗传修饰诸如甲基化、羟甲基化、甲酰化或5-羧基化的DNA、RNA或XNA。

[0153] 在一种实施方案中,核酸分子是携带非天然存在的修饰诸如荧光标签、荧光标记物、相互作用基团的DNA、RNA或XNA。

[0154] 在一种实施方案中,核酸分子是具有多于50个、100个、200个、300个、400个、500个、600个、700个、800个、900个、1 000个、2 000个、3 000个、4 000个、5 000个、6 000个、7000个、8 000个、9 000个、10 000个、15 000个、20 000个、30 000个、40 000个、50 000个或个100 000个核苷酸的长度的聚合物分子。

[0155] 应用

[0156] 本文描述了TdT的变体用于核酸合成、寡核苷酸合成、探针合成、加标签、核酸扩增、适配体、治疗性核酸分子、药物靶发现和验证、疾病诊断、代谢工程、数据存储、作物改良、文库设计、测序池、核酸标记或附接或涉及核酸分子的任何其他应用的用途。

[0157] 变体TdT的产生

[0158] 本发明的变体可以通过对已知的参考或野生型TdT的编码多核苷酸进行突变,然后使用常规分子生物学技术使其表达来产生。例如,小鼠TdT基因(SEQ ID NO.1)可以使用常规分子生物学技术,例如使用由Stemmer等人,Gene,164:49-53(1995);Kodumal等人,Proc.Natl.Acad.Sci.,101:15573-15578(2004)等描述的方案由合成片段组装;或者其可以使用Boule等人,Biotech NOlogy,10:199-208(1998),或Bentolila等人,EMBO J.,14:4221-4229(1995)等描述的方案从小鼠细胞直接克隆。

[0159] 例如,可以将分离的TdT基因插入表达载体,诸如pET32(NO.vagen)中,得到载体pCTdT,然后该载体pCTdT可以被用于使用常规方案制备和表达变体TdT蛋白。具有正确序列的载体可以在大肠杆菌(E.coli)生产菌株中转化。

[0160] 使用常规技术将转化的菌株培养成团块(pellets),从其中提取TdT蛋白。例如,将先前制备的团块在30℃至37℃水浴中解冻。完全解冻后,将团块重悬于裂解缓冲液中,该缓冲液由50mM Tris-HCl(Sigma)pH 7.5、150mM NaCl(Sigma)、0.5mM巯基乙醇(Sigma)、5%甘油(Sigma)、20mM咪唑(Sigma)和1片100mL蛋白酶混合物抑制剂(ThermoFisher)组成。仔细进行重悬,以避免聚集体过早溶解和残留。通过几个循环的弗氏压碎器(French press)裂解重悬的细胞,直到获得完全的颜色均匀度。通常使用的压力是14,000psi。然后将裂解物以10,000rpm离心1h至1h 30min。使离心液(centrifugate)通过0.2μm过滤器,以去除任何碎片,然后柱纯化。

[0161] 可以以一步亲和程序从离心液中纯化TdT蛋白。例如,使用Ni-NTA亲和柱(GEHealthcare)结合聚合酶。最初,柱已被洗涤,并用15个柱体积的50mM Tris-HCl(Sigma)pH 7.5、150mM NaCl(Sigma)和20mM咪唑(Sigma)平衡。平衡后,使聚合酶与柱结合。然后,将由50mM Tris-HCl(Sigma)pH 7.5、500mM NaCl(Sigma)和20mM咪唑(Sigma)组成的洗涤缓冲液施加到柱上,持续15个柱体积。洗涤后,将聚合酶用50mM Tris-HCl(Sigma)pH 7.5、500mM NaCl(Sigma)和0.5M咪唑(Sigma)洗脱。对应于最高浓度的感兴趣的聚合酶的级分被收集并

合并在单一样品中。将合并的级分相对于透析缓冲液(20mM Tris-HCl pH 6.8、200mM NaCl、50mM MgOAc、100mM $[\text{NH}_4]_2\text{SO}_4$)透析。随后借助浓缩过滤器(Amicon Ultra-30, MerkMillipore)浓缩透析液。将浓缩的酶分配在小等分试样中,最终添加50%的甘油,并且然后将这些等分试样冷冻在-20℃并长期储存。在SDS-PAGE凝胶中分析5μL的纯化的酶的不同级分。

[0162] 试剂盒、酶和核苷酸组合物

[0163] 本发明的特定方面涉及包含根据本发明或本发明的任何特定方面的TdT的变体以及任选地具有选自以下的一种或更多种组分的任何组合的组合物和试剂盒的用途:起始片段、一种或更多种可逆终止剂核苷酸、另外的酶和裂解反应中使用的试剂。所述试剂盒可以用于在酶促核酸合成的方法中使用。

[0164] 本发明涵盖包含根据本发明或本发明任何特定方面的TdT变体的物质组合物,其中可逆修饰的核苷酸与适当的缓冲液和比例浓度混合。

具体实施方式

[0165] 下述实施例中所使用的实验方法如无特殊说明,均为常规方法。

[0166] 下述实施例中所用的材料、试剂等,如无特殊说明,均可从商业途径得到。

[0167] 在下面的实施例中进一步说明本发明,但并不限制本发明的范围。部分分子克隆方法细节依据试剂、酶或试剂盒提供商不同而有所差别,应当按照产品说明进行操作,在实施例中不再详细描述。实施例中所述各种详细的表达条件、培养条件、反应条件、检测条件仅为示例,本发明不仅限于此,所有可以实现相应效果的条件均包括。

[0168] 实施例1、TdT蛋白的物种来源

[0169] 本发明采用的物种来源及氨基酸序列:SEQ ID NO.1(白喉雀,ZaTdT)。

[0170] 本实施例采用的酶,基因均由金唯智合成,均构建至pET-28a载体,位于酶切位点NdeI和XhoI之间,并在大肠杆菌中表达纯化蛋白。ZaTdT蛋白的构建及表达纯化载体以PET-28a载体为例,宿主菌以大肠杆菌,本发明包含但不仅限于此,所有可以成功表达纯化ZaTdT蛋白的载体及宿主菌均应包含。

[0171] 本发明还采用将SEQ ID NO.1做N端截短部分氨基酸,序列如SEQ ID NO:2(ZaTdT-JD)所示;本发明还采用将SEQ ID NO.1或SEQ ID NO:2的N端添加部分氨基酸,序列如SEQ ID NO:3(MBP-ZaTdT)或SEQ ID NO:4(MBP-ZaTdT-JD)所示。

[0172] 实施例2、高活性变体的获得

[0173] 本发明对来自白喉雀的ZaTdT进行同源建模获得其三维蛋白结构,基于发明人已经验,通过计算机模拟理性设计、单点或多点饱和突变,结合高通量筛选等手段,获得具有高催化效率和底物特异性的变体,本发明所包含的变体如下表所示。

[0174] 表1本发明所包含的突变体

[0175]

编号	突变的组合
S1	Y178A
S2	Y178I
S3	Y178L
S4	Y178V
S5	Y178P
S6	Y178Q
S7	Y178T
S8	Y178K
S9	Y178E
S10	F186G
S11	F186L
S12	F186V
S13	F186Y
S14	F186C
S15	F186H
S16	F186R
S17	F186D
S18	Y178A-F186G
S19	Y178A-F186R
S20	Y178L-F186G
S21	Y178L-F186R
S22	I210A
S23	I210L
S24	I210V
S25	I210G
S26	I210H
S27	I210K
S28	I210P
S29	I210D

[0176]

S30	I210Y
S31	I228A
S32	I228L
S33	I228G
S34	I228H
S35	I228K
S36	I228P
S37	I228D
S38	I228Y
S39	I210A-I228L
S40	I210L-I228L
S41	V302A
S42	V302L
S43	V302G
S44	D324G
S45	D324L
S46	D324V
S47	D324F
S48	D324Y
S49	D324C
S50	D324H
S51	D324R
S52	R353G
S53	R353L
S54	R353V
S55	R353F
S56	R353Y
S57	R353C
S58	R353H
S59	R353D
S60	D324G-R353G
S61	D324G-R353L
S62	D324F-R353G

[0177]

S63	D324Y-R353D
S64	T330A
S65	T330L
S66	T330A-G331V
S67	T330L-G331V
S68	T330K-G331V
S69	T330R-G331H
S70	G332A-F333G
S71	G332L-F333G
S72	G332A-F333V
S73	G332L-F333V
S74	F333G
S75	F333L
S76	R335A
S77	R335A-K337G
S78	R335A-K337A
S79	R335A-K337L
S80	R335A-K337V
S81	R335A-K337F
S82	R335A-K337Y
S83	R335A-K337C
S84	R335A-K337C
S85	R335A-K337H
S86	R335A-K337R
S87	R335A-K337D
S88	R335I
S89	R335I-K337G
S90	R335I-K337L
S91	R335I-K337V
S92	R335I-K337F
S93	R335I-K337Y
S94	R335I-K337C
S95	R335I-K337H

[0178]

S96	R335I-K337R
S97	R335I-K337D
S98	R335L
S99	R335L-K337A
S100	R335L-K337G
S101	R335L-K337L
S102	R335L-K337V
S103	R335L-K337F
S104	R335L-K337Y
S105	R335L-K337C
S106	R335L-K337H
S107	R335L-K337R
S108	R335L-K337D
S109	R335V
S110	R335V-K337G
S111	R335V-K337L
S112	R335V-K337V
S113	R335V-K337F
S114	R335V-K337Y
S115	R335V-K337C
S116	R335V-K337H
S117	R335V-K337R
S118	R335V-K337D
S119	R335P
S120	R335P-K337G
S121	R335P-K337L
S122	R335P-K337V
S123	R335P-K337F
S124	R335P-K337Y
S125	R335P-K337C
S126	R335P-K337H
S127	R335P-K337R
S128	R335P-K337D

[0179]

S129	R335Q
S130	R335Q-K337G
S131	R335Q-K337L
S132	R335Q-K337V
S133	R335Q-K337F
S134	R335Q-K337Y
S135	R335Q-K337C
S136	R335Q-K337H
S137	R335Q-K337R
S138	R335Q-K337D
S139	R335T
S140	R335T-K337G
S141	R335T-K337L
S142	R335T-K337V
S143	R335T-K337F
S144	R335T-K337Y
S145	R335T-K337C
S146	R335T-K337H
S147	R335T-K337R
S148	R335T-K337D
S149	R335K
S150	R335K-K337G
S151	R335K-K337L
S152	R335K-K337V
S153	R335K-K337F
S154	R335K-K337Y
S155	R335K-K337C
S156	R335K-K337H
S157	R335K-K337R
S158	R335K-K337D
S159	R335E
S160	R335E-K337G
S161	R335E-K337L

[0180]

S162	R335E-K337V
S163	R335E-K337F
S164	R335E-K337Y
S165	R335E-K337C
S166	R335E-K337H
S167	R335E-K337R
S168	R335E-K337D
S169	K337G
S170	K337L
S171	K337V
S172	K337F
S173	K337Y
S174	K337C
S175	K337H
S176	K337R
S177	K337E
S178	K337C
S179	K337D
S180	I329G
S181	I329L
S182	I329V
S183	I329F
S184	I329Y
S185	I329C
S186	I329H
S187	I329R
S188	I329D
S189	I339A
S190	I339L
S191	I339A-G340L
S192	I339L-G340V
S193	I339V-G340L
S194	I339V-G340V

[0181]

S195	I339P-G340L
S196	G340L
S197	H341A
S198	H341L
S199	H341A-I343G
S200	H341A-I343L
S201	H341L-I343G
S202	H341L-I343L
S203	H341L-I343V
S204	H341L-I343R
S205	H341V-I343G
S206	H341V-I343V
S207	H341V-I343L
S208	I343G
S209	R435L
S210	R435G
S211	R435A
S212	G452A-W453G
S213	G452A-W453L
S214	G452A-W453V
S215	G452A-W453K
S216	G452L-W453G
S217	G452L-W453L
S218	G452L-W453V
S219	G452L-W453P
S220	G452V-W453G
S221	G452V-W453L
S222	G452V-W453V
S223	G452V-W453K
S224	W453G
S225	W453L
S226	W453V
S227	T454A

[0182]

S228	T454A-G455L
S229	T454A-G455V
S230	T454A-G455F
S231	T454L-G455L
S232	T454L-G455V
S233	T454L-G455K
S234	T454V-G455L
S235	T454V-G455V
S236	T454V-G455F
S237	G455V
S238	G455L
S239	S456A
S240	S456L
S241	S456V
S242	S456A-R457G
S243	S456A-R457L
S244	S456A-R457V
S245	S456A-R457P
S246	S456L-R457G
S247	S456L-R457L
S248	S456L-R457V
S249	S456L-R457K
S250	S456V-R457G
S251	S456V-R457L
S252	S456V-R457V
S253	S456V-R457P
S254	R457G
S255	R457L
S256	R457V
S257	N477A
S258	N477I
S259	N477L
S260	N477G

[0183]	S261	N477K
	S262	N477R
	S263	N477D
	S264	N477Y
	S265	H478G
	S266	H478L
	S267	H478V
	S268	H478F
	S269	H478Y
	S270	H478C
	S271	H478D
	S272	H478R
	S273	N477L-H478G
	S274	N477R-H478G
	S275	N477K-H478G
	S276	N477L-H478F
	S277	N477L-H478C

[0184] 实施例3、ZaTdT变体蛋白的获得

[0185] 为了体外检测ZaTdT变体酶活性,在大肠杆菌中对该酶进行外源表达及纯化。实施例中所述的宿主菌为E.coli BL21 (DE3),但是宿主菌并不仅限于此,所有可以用于表达蛋白的宿主均包括。蛋白表达、纯化的方法参照专利:201911034444.6。

[0186] 实施例4、功能验证

[0187] 本发明实施例所述采用底物均以脱氧核糖核苷酸(dNTPs)为基础,3'-OH端变更为可逆终止基团,包括腺嘌呤脱氧核糖核苷酸、鸟嘌呤脱氧核糖核苷酸、胞嘧啶脱氧核糖核苷酸、胸腺嘧啶脱氧核糖核苷酸。优选地,3'端的修饰基团为氧氨基。

[0188] 1、示意图

[0189] 在起始序列存在的情况下,ZaTdT催化3'端带O-氨基修饰基团的脱氧核糖核苷酸加入起始序列3端的催化反应,如图1所示。

[0190] 2、体外纯酶反应

[0191] 反应体系:100mM NaCl,0.25mM CoCl₂,50mM KAc,10mM Mg (Ac)₂,pH 6.8。底物:1μM 起始链(16nt:ACTAGGACGACTCGAATT),100μM 3'端带O-氨基修饰基团的脱氧核糖核苷酸。酶:50μM TdT蛋白及突变体。

[0192] 反应条件:加入酶后开始反应,30℃反应5不同时间,随后95℃灭活蛋白,离心去沉淀。离心后上清液样品采用20%尿素-聚丙烯酰胺凝胶(变性PAGE)检测分析,300V电压跑3h。凝胶经1*SYBR Gold染色20min,进一步用Tanon的荧光凝胶成像仪分析图片各个DNA条带的净光密度。用掺入核苷酸之后的DNA条带的净光密度(17nt对应的净光密度)除以总体DNA条带的净光密度(16nt和17nt对应的净光密度),获得核苷酸掺入效率。掺入效率乘以底

物浓度除以反应时间获得变体酶反应的催化效率。

[0193] ZaTdT的N端氨基酸缺失或添加的变性PAGE分析结果如表2所示,结果显示,ZaTdT、ZaTdT的N端氨基酸缺失或添加的变体均能够和底物结合,并且经计算,N端缺失或添加部分氨基酸不影响ZaTdT的催化效率。

[0194] ZaTdT及部分单点或多点变体的催化效率如表3所示。ZaTdT及部分单点或多点变体的催化效率除以野生型TdT的催化效率获得TdT突变体的相对活性,部分变体的相对活性如图2所示。

[0195] 结果显示,与野生型相比,多种变体的催化效率得到了显著的提升,最高提高倍数约60倍。

[0196] 表2 ZaTdT N端缺失或添加氨基酸后的催化效率

[0197]

酶	催化效率nM/min
ZaTdT	4.40 ± 0.23
ZaTdT-JD	4.35 ± 0.18
MBP-ZaTdT	4.65 ± 0.31
MBP-ZaTdT-JD	4.52 ± 0.27

[0198] 表3 ZaTdT及部分突变体的催化效率

[0199]

酶	催化效率nM/min
WT	4.40
S1	24.82
S16	20.68
S19	48.88
S23	31.68
S32	20.24
S40	42.37
S43	6.78
S64	15.52
S76	88.22
S78	124.65
S98	189.95
S100	260.61
S101	175.47
S102	167.95
S169	188.85
S170	164.03
S194	26.93
S209	39.42
S227	23.32
S237	28.78
S251	41.54

S254	21.65
S259	88.66
S265	156.68

[0200] 实施例5、TdT变体在单链核苷酸合成中的应用

[0201] (1) 固定, 带有生物素标记的起始链与链霉亲和素磁珠结合, 30℃ 20min;

[0202] (2) 反应, 将结合有起始链的磁珠置于如下反应体系: 50mM PBS, 0.25mM CoCl_2 , 0.25mM 3'-OH被O-氨基修饰的核苷酸, 1.5mg/mL TdT变体(S100)。30℃反应10min。

[0203] (3) 脱保护, 用0.7M亚硝酸溶液脱保护; 再用磷酸盐Buffer洗两遍。

[0204] (4) 重复(2)(3)步骤, 循环9次。

[0205] (5) 每个循环样品进行20%尿素-聚丙烯酰胺凝胶(变性PAGE)检测, 经1*SYBR Gold染色后结果如图3所示。

[0206] 结果显示, TdT变体S100可以完成10个核苷酸的合成, 可以应用到单链核苷酸合成中。由此可知, 表1或表3中包含的与S100变体相同的具有在单链寡核苷酸末端掺入修饰核苷酸功能的变体, 均可以应用到单链DNA合成中。

[0207] 以上, 对本发明的实施方式进行了说明。但是, 本发明不限于上述实施方式。凡在本发明的精神和原则之内, 所做的任何修改、等同替换、改进等, 均应包含在本发明的保护范围之内。

SEQUENCE LISTING

<110> 中国科学院天津工业生物技术研究所

<120> 可控合成单链DNA的末端转移酶变体及应用

<130> CPCN21411298

<160> 4

<170> PatentIn version 3.5

<210> 1

<211> 513

<212> PRT

<213> 人工合成

<400> 1

```

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

```

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

<210> 2

<211> 382

<212> PRT

<213> 人工合成

<400> 2

```

Leu Pro Leu Asn Met Pro Ala Leu Glu Met Pro Ala Phe Ile Ala Thr
1           5           10           15
Lys Val Ser Gln Tyr Ser Cys Gln Arg Lys Thr Thr Leu Asn Asn Tyr
          20           25           30
Asn Lys Lys Phe Thr Asp Ala Phe Glu Val Met Ala Glu Asn Tyr Glu
          35           40           45
Phe Lys Glu Asn Glu Ile Phe Cys Leu Glu Phe Leu Arg Ala Ala Ser
          50           55           60
Leu Leu Lys Ser Leu Pro Phe Ser Val Thr Arg Met Lys Asp Ile Gln
65           70           75           80
Gly Leu Pro Cys Val Gly Asp Gln Val Arg Asp Ile Ile Glu Glu Ile
          85           90           95
Ile Glu Glu Gly Glu Ser Ser Arg Val Asn Glu Val Leu Asn Asp Glu
          100          105          110
Arg Tyr Lys Ala Phe Lys Gln Phe Thr Ser Val Phe Gly Val Gly Val
          115          120          125
Lys Thr Ser Glu Lys Trp Tyr Arg Met Gly Leu Arg Thr Val Glu Glu
          130          135          140
Val Lys Ala Asp Lys Thr Leu Lys Leu Ser Lys Met Gln Lys Ala Gly
145          150          155          160
Leu Leu Tyr Tyr Glu Asp Leu Val Ser Cys Val Ser Lys Ala Glu Ala
          165          170          175
Asp Ala Val Ser Leu Ile Val Lys Asn Thr Val Cys Thr Phe Leu Pro
          180          185          190
Asp Ala Leu Val Thr Ile Thr Gly Gly Phe Arg Arg Gly Lys Asn Ile
          195          200          205
Gly His Asp Ile Asp Phe Leu Ile Thr Asn Pro Gly Pro Arg Glu Asp
          210          215          220
Asp Glu Leu Leu His Lys Val Ile Asp Leu Trp Lys Lys Gln Gly Leu
225          230          235          240
Leu Leu Tyr Cys Asp Ile Ile Glu Ser Thr Phe Val Lys Glu Gln Leu
          245          250          255
Pro Ser Arg Lys Val Asp Ala Met Asp His Phe Gln Lys Cys Phe Ala
          260          265          270
Ile Leu Lys Leu Tyr Gln Pro Arg Val Asp Asn Ser Thr Cys Asn Thr

```

275	280	285
Ser Glu Gln Leu Glu Met Ala Glu Val Lys Asp Trp Lys Ala Ile Arg		
290	295	300
Val Asp Leu Val Ile Thr Pro Phe Glu Gln Tyr Pro Tyr Ala Leu Leu		
305	310	315
Gly Trp Thr Gly Ser Arg Gln Phe Gly Arg Asp Leu Arg Arg Tyr Ala		
325	330	335
Ala His Glu Arg Lys Met Ile Leu Asp Asn His Gly Leu Tyr Asp Arg		
340	345	350
Arg Lys Arg Ile Phe Leu Lys Ala Gly Ser Glu Glu Glu Ile Phe Ala		
355	360	365
His Leu Gly Leu Asp Tyr Val Glu Pro Trp Glu Arg Asn Ala		
370	375	380
<210> 3		
<211> 896		
<212> PRT		
<213> 人工合成		
<400> 3		
Met Lys Ile Glu Glu Gly Lys Leu Val Ile Trp Ile Asn Gly Asp Lys		
1	5	10
Gly Tyr Asn Gly Leu Ala Glu Val Gly Lys Lys Phe Glu Lys Asp Thr		
20	25	30
Gly Ile Lys Val Thr Val Glu His Pro Asp Lys Leu Glu Glu Lys Phe		
35	40	45
Pro Gln Val Ala Ala Thr Gly Asp Gly Pro Asp Ile Ile Phe Trp Ala		
50	55	60
His Asp Arg Phe Gly Gly Tyr Ala Gln Ser Gly Leu Leu Ala Glu Ile		
65	70	75
Thr Pro Asp Lys Ala Phe Gln Asp Lys Leu Tyr Pro Phe Thr Trp Asp		
85	90	95
Ala Val Arg Tyr Asn Gly Lys Leu Ile Ala Tyr Pro Ile Ala Val Glu		
100	105	110
Ala Leu Ser Leu Ile Tyr Asn Lys Asp Leu Leu Pro Asn Pro Pro Lys		
115	120	125
Thr Trp Glu Glu Ile Pro Ala Leu Asp Lys Glu Leu Lys Ala Lys Gly		
130	135	140
Lys Ser Ala Leu Met Phe Asn Leu Gln Glu Pro Tyr Phe Thr Trp Pro		
145	150	155
Leu Ile Ala Ala Asp Gly Gly Tyr Ala Phe Lys Tyr Glu Asn Gly Lys		

	165		170		175
Tyr Asp Ile Lys Asp Val Gly Val Asp Asn Ala Gly Ala Lys Ala Gly					
	180		185		190
Leu Thr Phe Leu Val Asp Leu Ile Lys Asn Lys His Met Asn Ala Asp					
	195		200		205
Thr Asp Tyr Ser Ile Ala Glu Ala Ala Phe Asn Lys Gly Glu Thr Ala					
	210		215		220
Met Thr Ile Asn Gly Pro Trp Ala Trp Ser Asn Ile Asp Thr Ser Lys					
225		230		235	240
Val Asn Tyr Gly Val Thr Val Leu Pro Thr Phe Lys Gly Gln Pro Ser					
	245		250		255
Lys Pro Phe Val Gly Val Leu Ser Ala Gly Ile Asn Ala Ala Ser Pro					
	260		265		270
Asn Lys Glu Leu Ala Lys Glu Phe Leu Glu Asn Tyr Leu Leu Thr Asp					
	275		280		285
Glu Gly Leu Glu Ala Val Asn Lys Asp Lys Pro Leu Gly Ala Val Ala					
	290		295		300
Leu Lys Ser Tyr Glu Glu Glu Leu Val Lys Asp Pro Arg Ile Ala Ala					
305		310		315	320
Thr Met Glu Asn Ala Gln Lys Gly Glu Ile Met Pro Asn Ile Pro Gln					
	325		330		335
Met Ser Ala Phe Trp Tyr Ala Val Arg Thr Ala Val Ile Asn Ala Ala					
	340		345		350
Ser Gly Arg Gln Thr Val Asp Glu Ala Leu Lys Asp Ala Gln Thr Asn					
	355		360		365
Ser Ser Ser Asn Asn Asn Asn Asn Asn Asn Asn Asn Asn Leu Gly Met					
	370		375		380
Asp Arg Phe Lys Ala Pro Ala Val Ile Ser Gln Arg Lys Arg Gln Lys					
385		390		395	400
Gly Leu His Ser Pro Lys Leu Ser Cys Ser Tyr Glu Ile Lys Phe Ser					
	405		410		415
Asn Phe Val Ile Phe Ile Met Gln Arg Lys Met Gly Leu Thr Arg Arg					
	420		425		430
Met Phe Leu Met Glu Leu Gly Arg Arg Lys Gly Phe Arg Val Glu Ser					
	435		440		445
Glu Leu Ser Asp Ser Val Thr His Ile Val Ala Glu Asn Asn Ser Tyr					
	450		455		460
Leu Glu Val Leu Asp Trp Leu Lys Gly Gln Ala Val Gly Asp Ser Ser					
465		470		475	480

Arg Phe Glu Leu Leu Asp Ile Ser Trp Phe Thr Ala Cys Met Glu Ala		
485	490	495
Gly Arg Pro Val Asp Ser Glu Val Lys Tyr Arg Leu Met Glu Gln Ser		
500	505	510
Gln Ser Leu Pro Leu Asn Met Pro Ala Leu Glu Met Pro Ala Phe Ile		
515	520	525
Ala Thr Lys Val Ser Gln Tyr Ser Cys Gln Arg Lys Thr Thr Leu Asn		
530	535	540
Asn Tyr Asn Lys Lys Phe Thr Asp Ala Phe Glu Val Met Ala Glu Asn		
545	550	555
Tyr Glu Phe Lys Glu Asn Glu Ile Phe Cys Leu Glu Phe Leu Arg Ala		
565	570	575
Ala Ser Leu Leu Lys Ser Leu Pro Phe Ser Val Thr Arg Met Lys Asp		
580	585	590
Ile Gln Gly Leu Pro Cys Val Gly Asp Gln Val Arg Asp Ile Ile Glu		
595	600	605
Glu Ile Ile Glu Glu Gly Glu Ser Ser Arg Val Asn Glu Val Leu Asn		
610	615	620
Asp Glu Arg Tyr Lys Ala Phe Lys Gln Phe Thr Ser Val Phe Gly Val		
625	630	635
Gly Val Lys Thr Ser Glu Lys Trp Tyr Arg Met Gly Leu Arg Thr Val		
645	650	655
Glu Glu Val Lys Ala Asp Lys Thr Leu Lys Leu Ser Lys Met Gln Lys		
660	665	670
Ala Gly Leu Leu Tyr Tyr Glu Asp Leu Val Ser Cys Val Ser Lys Ala		
675	680	685
Glu Ala Asp Ala Val Ser Leu Ile Val Lys Asn Thr Val Cys Thr Phe		
690	695	700
Leu Pro Asp Ala Leu Val Thr Ile Thr Gly Gly Phe Arg Arg Gly Lys		
705	710	715
Asn Ile Gly His Asp Ile Asp Phe Leu Ile Thr Asn Pro Gly Pro Arg		
725	730	735
Glu Asp Asp Glu Leu Leu His Lys Val Ile Asp Leu Trp Lys Lys Gln		
740	745	750
Gly Leu Leu Leu Tyr Cys Asp Ile Ile Glu Ser Thr Phe Val Lys Glu		
755	760	765
Gln Leu Pro Ser Arg Lys Val Asp Ala Met Asp His Phe Gln Lys Cys		
770	775	780
Phe Ala Ile Leu Lys Leu Tyr Gln Pro Arg Val Asp Asn Ser Thr Cys		

785	790	795	800
Asn Thr Ser Glu Gln Leu Glu Met Ala Glu Val Lys Asp Trp Lys Ala			
	805	810	815
Ile Arg Val Asp Leu Val Ile Thr Pro Phe Glu Gln Tyr Pro Tyr Ala			
	820	825	830
Leu Leu Gly Trp Thr Gly Ser Arg Gln Phe Gly Arg Asp Leu Arg Arg			
	835	840	845
Tyr Ala Ala His Glu Arg Lys Met Ile Leu Asp Asn His Gly Leu Tyr			
	850	855	860
Asp Arg Arg Lys Arg Ile Phe Leu Lys Ala Gly Ser Glu Glu Glu Ile			
865	870	875	880
Phe Ala His Leu Gly Leu Asp Tyr Val Glu Pro Trp Glu Arg Asn Ala			
	885	890	895
<210> 4			
<211> 765			
<212> PRT			
<213> 人工合成			
<400> 4			
Met Lys Ile Glu Glu Gly Lys Leu Val Ile Trp Ile Asn Gly Asp Lys			
1	5	10	15
Gly Tyr Asn Gly Leu Ala Glu Val Gly Lys Lys Phe Glu Lys Asp Thr			
	20	25	30
Gly Ile Lys Val Thr Val Glu His Pro Asp Lys Leu Glu Glu Lys Phe			
	35	40	45
Pro Gln Val Ala Ala Thr Gly Asp Gly Pro Asp Ile Ile Phe Trp Ala			
	50	55	60
His Asp Arg Phe Gly Gly Tyr Ala Gln Ser Gly Leu Leu Ala Glu Ile			
65	70	75	80
Thr Pro Asp Lys Ala Phe Gln Asp Lys Leu Tyr Pro Phe Thr Trp Asp			
	85	90	95
Ala Val Arg Tyr Asn Gly Lys Leu Ile Ala Tyr Pro Ile Ala Val Glu			
	100	105	110
Ala Leu Ser Leu Ile Tyr Asn Lys Asp Leu Leu Pro Asn Pro Pro Lys			
	115	120	125
Thr Trp Glu Glu Ile Pro Ala Leu Asp Lys Glu Leu Lys Ala Lys Gly			
	130	135	140
Lys Ser Ala Leu Met Phe Asn Leu Gln Glu Pro Tyr Phe Thr Trp Pro			
145	150	155	160
Leu Ile Ala Ala Asp Gly Gly Tyr Ala Phe Lys Tyr Glu Asn Gly Lys			

	165		170		175
Tyr Asp Ile Lys Asp Val Gly Val Asp Asn Ala Gly Ala Lys Ala Gly					
	180		185		190
Leu Thr Phe Leu Val Asp Leu Ile Lys Asn Lys His Met Asn Ala Asp					
	195		200		205
Thr Asp Tyr Ser Ile Ala Glu Ala Ala Phe Asn Lys Gly Glu Thr Ala					
	210		215		220
Met Thr Ile Asn Gly Pro Trp Ala Trp Ser Asn Ile Asp Thr Ser Lys					
225		230		235	240
Val Asn Tyr Gly Val Thr Val Leu Pro Thr Phe Lys Gly Gln Pro Ser					
	245		250		255
Lys Pro Phe Val Gly Val Leu Ser Ala Gly Ile Asn Ala Ala Ser Pro					
	260		265		270
Asn Lys Glu Leu Ala Lys Glu Phe Leu Glu Asn Tyr Leu Leu Thr Asp					
	275		280		285
Glu Gly Leu Glu Ala Val Asn Lys Asp Lys Pro Leu Gly Ala Val Ala					
	290		295		300
Leu Lys Ser Tyr Glu Glu Glu Leu Val Lys Asp Pro Arg Ile Ala Ala					
305		310		315	320
Thr Met Glu Asn Ala Gln Lys Gly Glu Ile Met Pro Asn Ile Pro Gln					
	325		330		335
Met Ser Ala Phe Trp Tyr Ala Val Arg Thr Ala Val Ile Asn Ala Ala					
	340		345		350
Ser Gly Arg Gln Thr Val Asp Glu Ala Leu Lys Asp Ala Gln Thr Asn					
	355		360		365
Ser Ser Ser Asn Asn Asn Asn Asn Asn Asn Asn Asn Asn Leu Gly Leu					
	370		375		380
Pro Leu Asn Met Pro Ala Leu Glu Met Pro Ala Phe Ile Ala Thr Lys					
385		390		395	400
Val Ser Gln Tyr Ser Cys Gln Arg Lys Thr Thr Leu Asn Asn Tyr Asn					
	405		410		415
Lys Lys Phe Thr Asp Ala Phe Glu Val Met Ala Glu Asn Tyr Glu Phe					
	420		425		430
Lys Glu Asn Glu Ile Phe Cys Leu Glu Phe Leu Arg Ala Ala Ser Leu					
	435		440		445
Leu Lys Ser Leu Pro Phe Ser Val Thr Arg Met Lys Asp Ile Gln Gly					
	450		455		460
Leu Pro Cys Val Gly Asp Gln Val Arg Asp Ile Ile Glu Glu Ile Ile					
465		470		475	480

Glu	Glu	Gly	Glu	Ser	Ser	Arg	Val	Asn	Glu	Val	Leu	Asn	Asp	Glu	Arg	485	490	495
Tyr	Lys	Ala	Phe	Lys	Gln	Phe	Thr	Ser	Val	Phe	Gly	Val	Gly	Val	Lys	500	505	510
Thr	Ser	Glu	Lys	Trp	Tyr	Arg	Met	Gly	Leu	Arg	Thr	Val	Glu	Glu	Val	515	520	525
Lys	Ala	Asp	Lys	Thr	Leu	Lys	Leu	Ser	Lys	Met	Gln	Lys	Ala	Gly	Leu	530	535	540
Leu	Tyr	Tyr	Glu	Asp	Leu	Val	Ser	Cys	Val	Ser	Lys	Ala	Glu	Ala	Asp	545	550	555
Ala	Val	Ser	Leu	Ile	Val	Lys	Asn	Thr	Val	Cys	Thr	Phe	Leu	Pro	Asp	565	570	575
Ala	Leu	Val	Thr	Ile	Thr	Gly	Gly	Phe	Arg	Arg	Gly	Lys	Asn	Ile	Gly	580	585	590
His	Asp	Ile	Asp	Phe	Leu	Ile	Thr	Asn	Pro	Gly	Pro	Arg	Glu	Asp	Asp	595	600	605
Glu	Leu	Leu	His	Lys	Val	Ile	Asp	Leu	Trp	Lys	Lys	Gln	Gly	Leu	Leu	610	615	620
Leu	Tyr	Cys	Asp	Ile	Ile	Glu	Ser	Thr	Phe	Val	Lys	Glu	Gln	Leu	Pro	625	630	635
Ser	Arg	Lys	Val	Asp	Ala	Met	Asp	His	Phe	Gln	Lys	Cys	Phe	Ala	Ile	645	650	655
Leu	Lys	Leu	Tyr	Gln	Pro	Arg	Val	Asp	Asn	Ser	Thr	Cys	Asn	Thr	Ser	660	665	670
Glu	Gln	Leu	Glu	Met	Ala	Glu	Val	Lys	Asp	Trp	Lys	Ala	Ile	Arg	Val	675	680	685
Asp	Leu	Val	Ile	Thr	Pro	Phe	Glu	Gln	Tyr	Pro	Tyr	Ala	Leu	Leu	Gly	690	695	700
Trp	Thr	Gly	Ser	Arg	Gln	Phe	Gly	Arg	Asp	Leu	Arg	Arg	Tyr	Ala	Ala	705	710	715
His	Glu	Arg	Lys	Met	Ile	Leu	Asp	Asn	His	Gly	Leu	Tyr	Asp	Arg	Arg	725	730	735
Lys	Arg	Ile	Phe	Leu	Lys	Ala	Gly	Ser	Glu	Glu	Glu	Ile	Phe	Ala	His	740	745	750
Leu	Gly	Leu	Asp	Tyr	Val	Glu	Pro	Trp	Glu	Arg	Asn	Ala				755	760	765

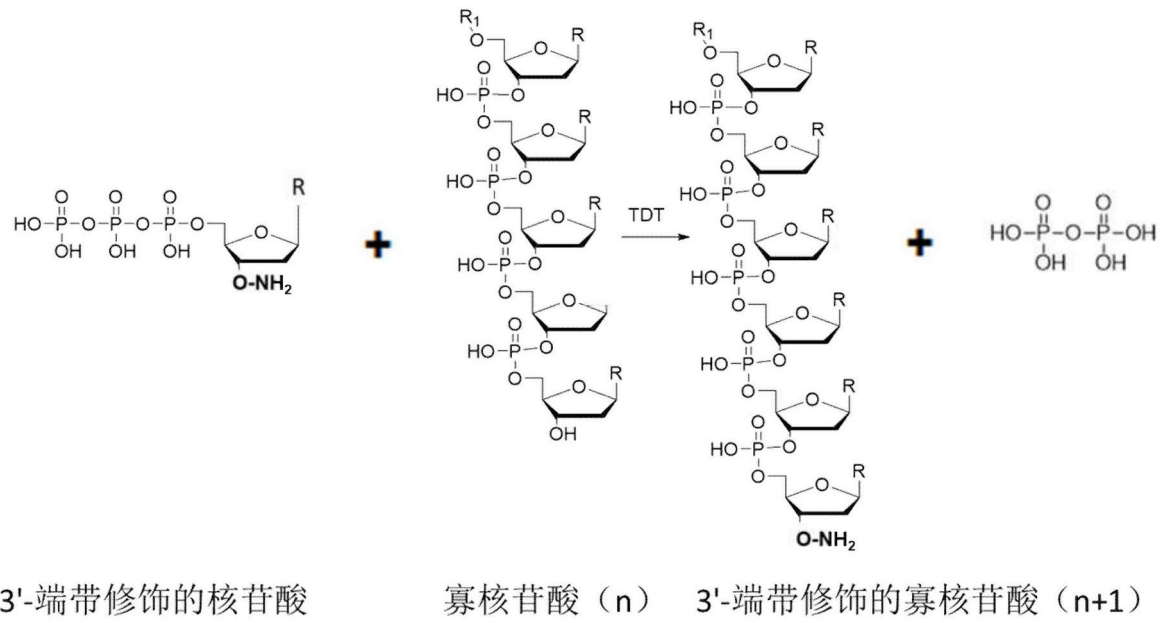


图1

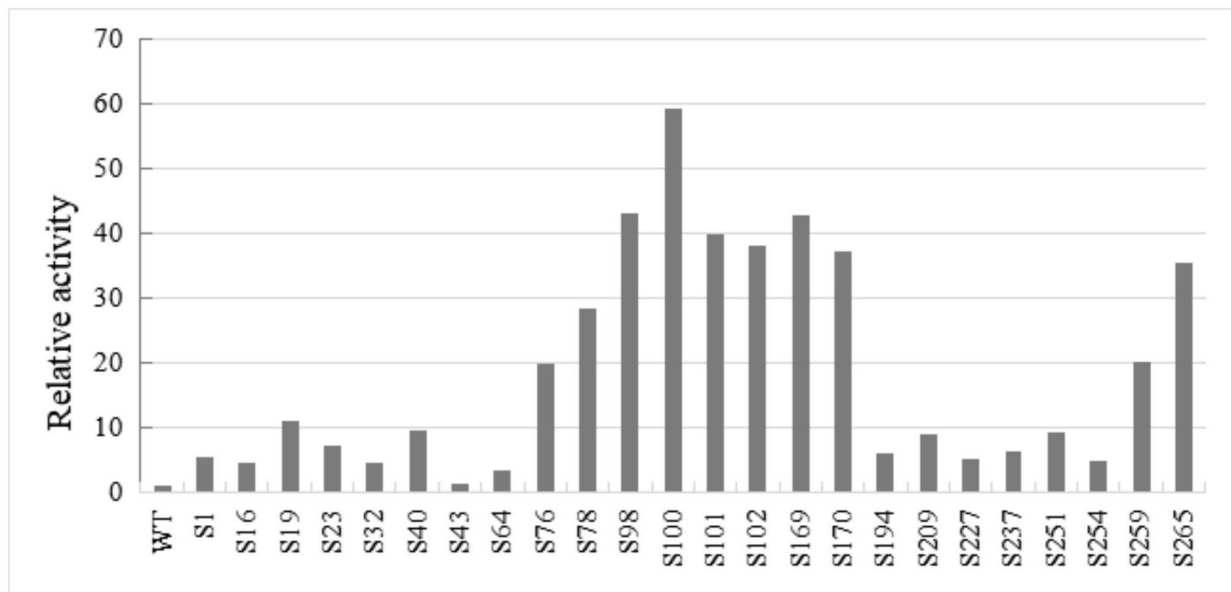


图2

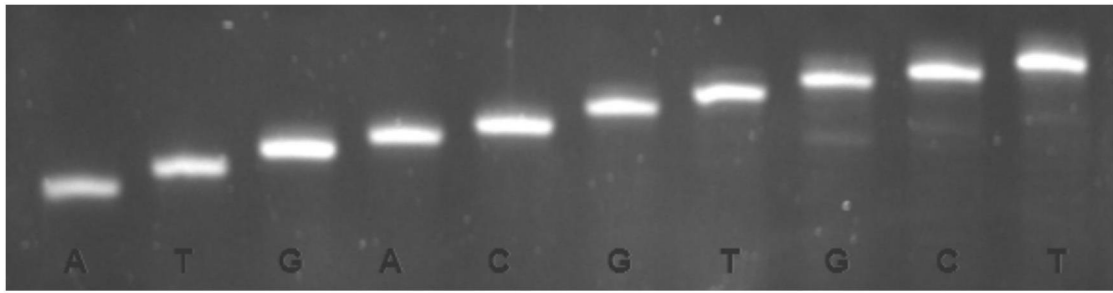


图3