



(12) 发明专利申请

(10) 申请公布号 CN 119317710 A

(43) 申请公布日 2025.01.14

(21) 申请号 202380039861.8

(22) 申请日 2023.04.13

(30) 优先权数据

63/331,145 2022.04.14 US

(85) PCT国际申请进入国家阶段日

2024.11.12

(86) PCT国际申请的申请数据

PCT/US2023/065719 2023.04.13

(87) PCT国际申请的公布数据

W02023/201294 EN 2023.10.19

(71) 申请人 摩登纳特斯有限公司

地址 美国

(72) 发明人 埃米·E·拉比多

玛格丽特·富兰克林

阿萨纳西奥斯·杜西斯

(74) 专利代理机构 北京集佳知识产权代理有限公司 11227

专利代理师 张福誉 韩晓帆

(51) Int.Cl.

C12N 9/12 (2006.01)

权利要求书3页 说明书52页

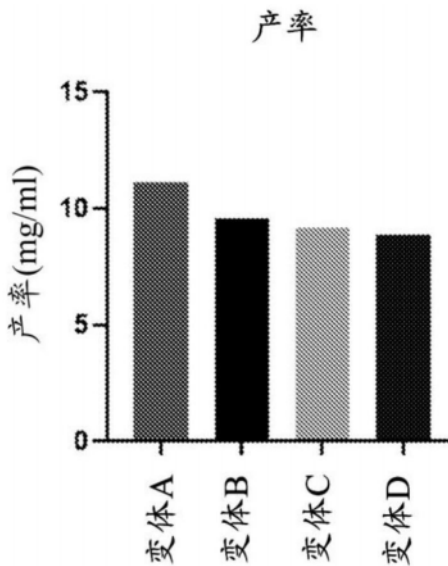
序列表(电子公布) 附图4页

(54) 发明名称

RNA聚合酶变体

(57) 摘要

RNA聚合酶变体能够实现RNA的高效转录。在一些实施方案中,所述RNA聚合酶变体能够实现具有高加帽效率和/或低水平的双链RNA污染的RNA转录。



1. 一种核糖核酸 (RNA) 聚合酶变体, 所述RNA聚合酶变体包含与SEQ ID NO:2-9中任一个的氨基酸序列具有至少90%序列同一性的氨基酸序列, 其中相对于包含SEQ ID NO:1的氨基酸序列的RNA聚合酶, 所述氨基酸序列包含位置D351处的氨基酸取代和至少两个额外的氨基酸取代。

2. 如权利要求1所述的RNA聚合酶变体, 所述RNA聚合酶变体包含与SEQ ID NO:2-9中任一个的氨基酸序列具有至少95%序列同一性的氨基酸序列。

3. 如权利要求1或2所述的RNA聚合酶变体, 所述RNA聚合酶变体包含氨基酸序列, 所述氨基酸序列包含SEQ ID NO:2-9中任一个的氨基酸序列。

4. 如权利要求1-3中任一项所述的RNA聚合酶变体, 相对于包含SEQ ID NO:1的氨基酸序列的野生型T7 RNA聚合酶, 所述RNA聚合酶变体包含至少一个、至少两个、至少三个或至少四个氨基酸取代。

5. 一种核糖核酸 (RNA) 聚合酶变体, 所述RNA聚合酶变体包含SEQ ID NO:6的氨基酸序列。

6. 一种核糖核酸 (RNA) 聚合酶变体, 所述RNA聚合酶变体包含SEQ ID NO:7的氨基酸序列。

7. 一种核糖核酸 (RNA) 聚合酶变体, 所述RNA聚合酶变体包含SEQ ID NO:8的氨基酸序列。

8. 一种核糖核酸 (RNA) 聚合酶变体, 所述RNA聚合酶变体包含SEQ ID NO:9的氨基酸序列。

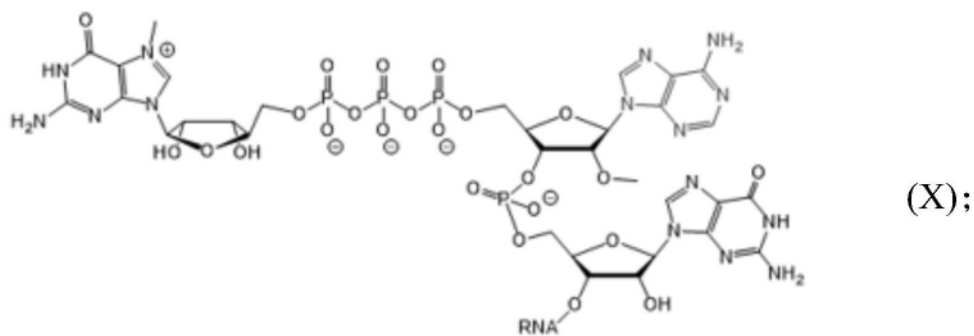
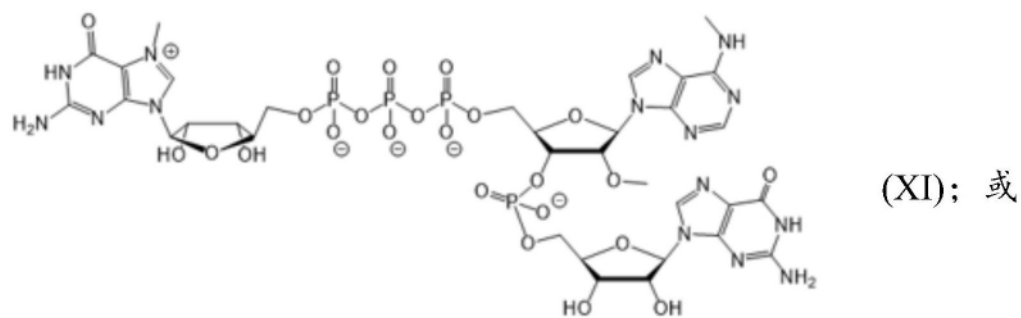
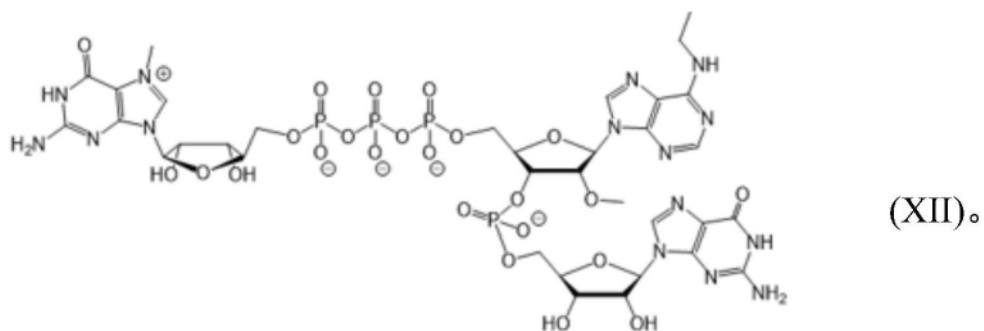
9. 一种方法, 所述方法包括: 在体外转录反应中产生信使RNA (mRNA), 所述转录反应包括DNA、核苷三磷酸、如权利要求1-8中任一项所述的RNA聚合酶变体。

10. 如权利要求9所述的方法, 其中所述反应进一步包括帽类似物。

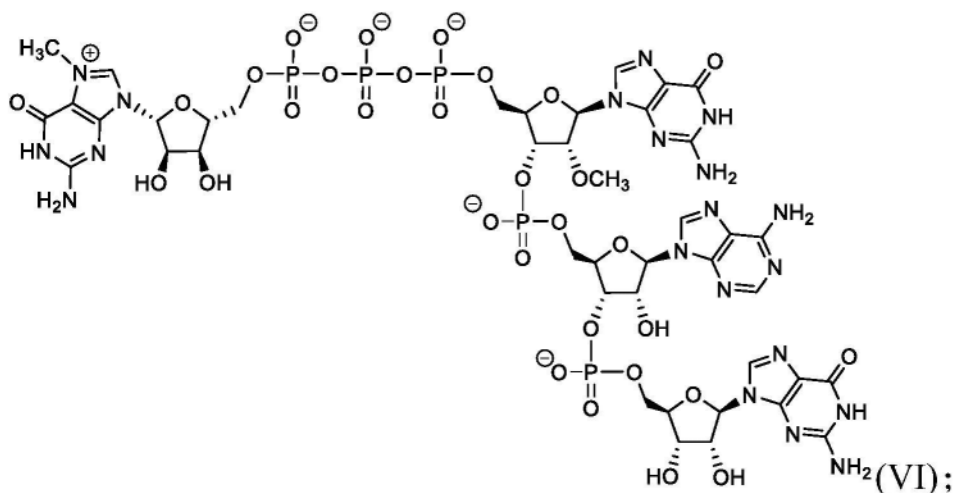
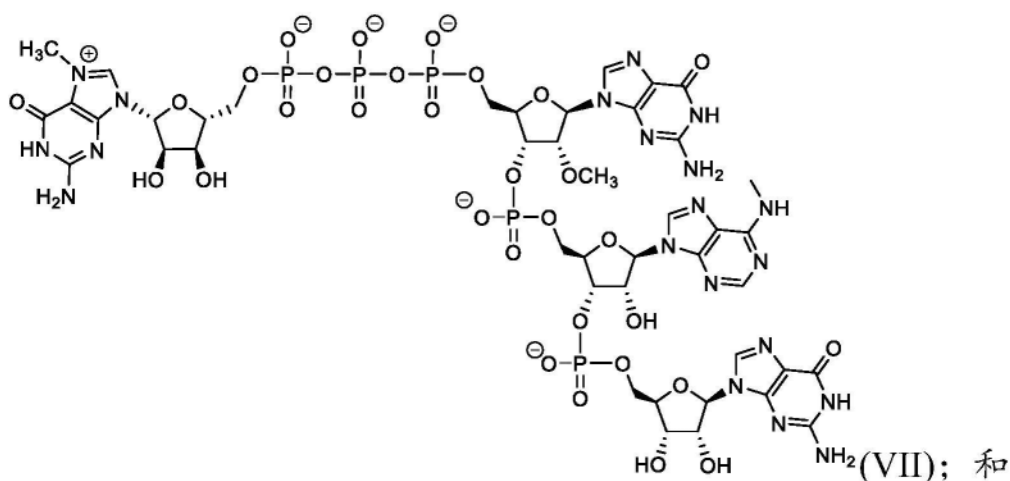
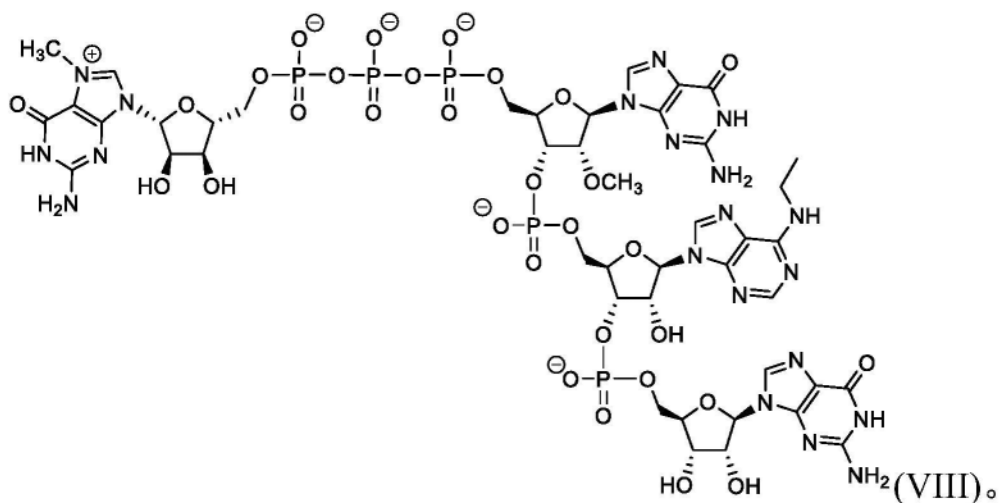
11. 如权利要求10所述的方法, 其中所述帽类似物是二核苷酸帽类似物、三核苷酸帽类似物或四核苷酸帽类似物。

12. 如权利要求11所述的方法, 其中所述帽类似物是包含GAG序列的三核苷酸帽类似物。

13. 如权利要求11所述的方法, 其中所述帽类似物包含选自以下的化合物:



14. 如权利要求11所述的方法,其中所述四核苷酸帽类似物包含GGAG序列。
15. 如权利要求11所述的方法,其中所述四核苷酸帽类似物包含选自以下的化合物:



16.如权利要求9-15中任一项所述的方法,其中所述DNA包含位置+1处的2'-脱氧胸苷残基或2'-脱氧胞苷残基。

17.一种组合物或试剂盒,所述组合物或试剂盒包括如权利要求1-9中任一项所述的RNA聚合酶变体和选自由DNA、核苷三磷酸和帽类似物组成的组的体外转录 (IVT) 试剂。

18. 一种核酸, 所述核酸编码如权利要求1-9中任一项所述的RNA聚合酶变体。

RNA聚合酶变体

[0001] 相关申请

[0002] 本申请根据35U.S.C.§119(e) 要求2022年4月14日提交的美国临时申请号63/331,145的权益,所述申请的内容以全文引用的方式并入本文中。

[0003] 电子序列表的引用

[0004] 电子序列表(M137870217W000-SEQ-HJD.xml;大小:31,370字节;以及创建日期:2023年4月12日)的内容以全文引用的方式并入本文中。

背景技术

[0005] 基于核糖核酸(RNA)的治疗剂的出现需要产生RNA并且异常活性产生的副产物很少的聚合酶。由使用噬菌体T7 RNA聚合酶进行体外转录产生的转录物表现出通常是不期望的且不可控制的免疫刺激活性。T7转录物的这种免疫刺激活性是由其从无启动子脱氧核糖核酸(DNA)端启动转录的异常活性促成的。这种活性导致与预期有义RNA产物完全互补的反义RNA的产生,并因此产生可强有力刺激非预期的免疫反应的长双链RNA(dsRNA)。此外,在帽类似物的存在下,噬菌体T7 RNA聚合酶产生具有低5'端加帽效率的T7转录物,部分原因是聚合酶对帽类似物具有低结合亲和力。

发明内容

[0006] 一些方面包括T7 RNA聚合酶变体和使用这些变体的体外转录方法,相对于对照(例如,包含SEQ ID NO:1的氨基酸序列的野生型T7 RNA聚合酶),所述变体已显示减少dsRNA污染物和/或提高共转录5'端加帽效率。

[0007] 一些方面提供了一种核糖核酸(RNA)聚合酶变体,所述RNA聚合酶变体包含与SEQ ID NO:2-9中任一个的氨基酸序列具有至少90%序列同一性的氨基酸序列,其中相对于包含SEQ ID NO:1的氨基酸序列的RNA聚合酶,所述氨基酸序列包含位置D351处的氨基酸取代和至少两个额外的氨基酸取代。

[0008] 一些方面提供了一种RNA聚合酶变体,所述RNA聚合酶变体包含氨基酸序列,相对于野生型T7 RNA聚合酶(例如,包含SEQ ID NO:1的氨基酸序列的野生型T7 RNA聚合酶),所述氨基酸序列包含至少一个、至少两个、至少三个或至少四个氨基酸取代。

[0009] 一些方面提供了一种RNA聚合酶变体,所述RNA聚合酶变体包含与SEQ ID NO:2-9中任一个的氨基酸序列具有至少90%、至少95%、至少98%或100%序列同一性的氨基酸序列。

[0010] 一些方面提供了一种RNA聚合酶变体,所述RNA聚合酶变体包含:氨基酸序列,相对于包含SEQ ID NO:1的氨基酸序列的野生型T7 RNA聚合酶,所述氨基酸序列包含(i)位置E350处的氨基酸取代,(ii)D351处的氨基酸取代,和(iii)位置K387、位置N437处或位置K387和位置N437处的氨基酸取代。

[0011] 在一些实施方案中,变体的氨基酸序列包含位置K387处的氨基酸取代。

[0012] 在一些实施方案中,变体的氨基酸序列包含位置N437处的氨基酸取代。

- [0013] 在一些实施方案中,变体的氨基酸序列包含位置K387处和位置N437处的氨基酸取代。
- [0014] 在一些实施方案中,位置K387处的氨基酸取代是极性的中性氨基酸。
- [0015] 在一些实施方案中,极性中性氨基酸选自天冬酰胺(N)、半胱氨酸(C)、谷氨酰胺(Q)、蛋氨酸(M)、丝氨酸(S)和苏氨酸(T)。
- [0016] 在一些实施方案中,极性中性氨基酸是天冬酰胺(K387N)。
- [0017] 在一些实施方案中,极性中性氨基酸是半胱氨酸(K387C)。
- [0018] 在一些实施方案中,极性中性氨基酸是谷氨酰胺(K387Q)。
- [0019] 在一些实施方案中,极性中性氨基酸是蛋氨酸(K387M)。
- [0020] 在一些实施方案中,极性中性氨基酸是丝氨酸(K387S)。
- [0021] 在一些实施方案中,极性中性氨基酸是苏氨酸(K387T)。
- [0022] 在一些实施方案中,位置N437处的氨基酸取代是芳族氨基酸。
- [0023] 在一些实施方案中,芳族氨基酸选自色氨酸(W)、酪氨酸(Y)和苯丙氨酸(F)。
- [0024] 在一些实施方案中,芳族氨基酸是色氨酸(N437W)。
- [0025] 在一些实施方案中,芳族氨基酸是酪氨酸(N437Y)。
- [0026] 在一些实施方案中,芳族氨基酸是苯丙氨酸(N437F)。
- [0027] 其他方面提供了一种RNA聚合酶变体,所述RNA聚合酶变体包含氨基酸序列,相对于包含SEQ ID NO 1的氨基酸序列的野生型T7 RNA聚合酶,所述氨基酸序列包含(i)位置E350处的氨基酸取代,(ii)D351处的氨基酸取代,和(iii)位置D653处的氨基酸取代。
- [0028] 在一些实施方案中,位置D653处的氨基酸取代是芳族氨基酸。
- [0029] 在一些实施方案中,芳族氨基酸选自色氨酸(W)、酪氨酸(Y)和苯丙氨酸(F)。
- [0030] 在一些实施方案中,芳族氨基酸是色氨酸(D653W)。
- [0031] 在一些实施方案中,芳族氨基酸是酪氨酸(D653Y)。
- [0032] 在一些实施方案中,芳族氨基酸是苯丙氨酸(D653F)。
- [0033] 在一些实施方案中,位置E350处的氨基酸取代是芳族氨基酸。
- [0034] 在一些实施方案中,芳族氨基酸选自色氨酸(W)、酪氨酸(Y)和苯丙氨酸(F)。
- [0035] 在一些实施方案中,芳族氨基酸是色氨酸(E350W)。
- [0036] 在一些实施方案中,芳族氨基酸是酪氨酸(E350Y)。
- [0037] 在一些实施方案中,芳族氨基酸是苯丙氨酸(E350F)。
- [0038] 在一些实施方案中,位置D351处的氨基酸取代是非极性脂肪族氨基酸。
- [0039] 在一些实施方案中,非极性脂肪族氨基酸选自丙氨酸(A)、甘氨酸(G)、异亮氨酸(I)、亮氨酸(L)、脯氨酸(P)和缬氨酸(V)。
- [0040] 在一些实施方案中,非极性脂肪族氨基酸是丙氨酸(D351A)。
- [0041] 在一些实施方案中,非极性脂肪族氨基酸是甘氨酸(D351G)。
- [0042] 在一些实施方案中,非极性脂肪族氨基酸是异亮氨酸(D351I)。
- [0043] 在一些实施方案中,非极性脂肪族氨基酸是亮氨酸(D351L)。
- [0044] 在一些实施方案中,非极性脂肪族氨基酸是脯氨酸(D351P)。
- [0045] 在一些实施方案中,非极性脂肪族氨基酸是缬氨酸(D351V)。
- [0046] 其他方面提供了一种RNA聚合酶变体,所述RNA聚合酶变体包含:与SEQ ID NO:1的

氨基酸序列具有至少70%同一性的氨基酸序列,其中相对于包含SEQ ID NO:1的氨基酸序列的野生型T7 RNA聚合酶,所述变体的氨基酸序列包含(i)位置E350处的氨基酸取代,(ii)D351处的氨基酸取代,和(iii)位置K387、位置N437处或位置K387和位置N437处的氨基酸取代。

[0047] 在一些实施方案中,氨基酸序列与SEQ ID NO:1的氨基酸序列具有至少75%、至少80%、至少85%、至少95%或至少98%同一性。

[0048] 在一些实施方案中,变体的氨基酸序列包含位置K387处的氨基酸取代。

[0049] 在一些实施方案中,变体的氨基酸序列包含位置N437处的氨基酸取代。

[0050] 在一些实施方案中,变体的氨基酸序列包含位置K387处和位置N437处的氨基酸取代。

[0051] 在一些实施方案中,位置K387处的氨基酸取代是极性的中性氨基酸。

[0052] 在一些实施方案中,极性中性氨基酸选自天冬酰胺(K387N)、半胱氨酸(K387C)、谷氨酰胺(K387Q)、蛋氨酸(K387M)、丝氨酸(K387S)和苏氨酸(K387T)。

[0053] 在一些实施方案中,位置N437处的氨基酸取代是芳族氨基酸。

[0054] 在一些实施方案中,芳族氨基酸选自色氨酸(N437W)、酪氨酸(N437Y)和苯丙氨酸(N437F)。

[0055] 其他方面提供了一种RNA聚合酶变体,所述RNA聚合酶变体包含:与SEQ ID NO:1的氨基酸序列具有至少70%同一性的氨基酸序列,其中相对于包含SEQ ID NO:1的氨基酸序列的野生型T7 RNA聚合酶,所述变体的氨基酸序列包含(i)位置E350处的氨基酸取代,(ii)D351处的氨基酸取代,和(iii)位置D653处的氨基酸取代。

[0056] 在一些实施方案中,氨基酸序列与SEQ ID NO:1的氨基酸序列具有至少75%、至少80%、至少85%、至少95%或至少98%同一性。

[0057] 在一些实施方案中,位置D653处的氨基酸取代是芳族氨基酸。

[0058] 在一些实施方案中,芳族氨基酸选自色氨酸(D653W)、酪氨酸(D653Y)和苯丙氨酸(D653F)。

[0059] 在一些实施方案中,位置E350处的氨基酸取代是芳族氨基酸。

[0060] 在一些实施方案中,芳族氨基酸选自色氨酸(E350W)、酪氨酸(E350Y)和苯丙氨酸(E350F)。

[0061] 在一些实施方案中,位置D351处的氨基酸取代是非极性脂肪族氨基酸。

[0062] 在一些实施方案中,非极性脂肪族氨基酸选自丙氨酸(D351A)、甘氨酸(D351G)、异亮氨酸(D351I)、亮氨酸(D351L)、脯氨酸(D351P)和缬氨酸(D351V)。

[0063] 一些方面提供了一种RNA聚合酶变体,所述RNA聚合酶变体包含SEQ ID NO:2的氨基酸序列,其中 X^1 是芳族氨基酸,任选地选自W、Y、和F; X^2 选自非极性脂肪族氨基酸,任选地选自A、G、I、L、P和V; X^3 是极性中性氨基酸,任选地选自N、C、Q、M、S和T;并且 X^4 是芳族氨基酸,任选地选自W、Y和F。在一些实施方案中,RNA聚合酶变体包含SEQ ID NO:6的氨基酸序列。

[0064] 一些方面提供了一种RNA聚合酶变体,所述RNA聚合酶变体包含SEQ ID NO:3的氨基酸序列,其中 X^1 是芳族氨基酸,任选地选自W、Y、和F; X^2 选自非极性脂肪族氨基酸,任选地选自A、G、I、L、P和V;并且 X^4 是芳族氨基酸,任选地选自W、Y和F。在一些实施方案中,RNA聚合酶变体包含SEQ ID NO:7的氨基酸序列。

[0065] 一些方面提供了一种RNA聚合酶变体,所述RNA聚合酶变体包含SEQ ID NO:4的氨基酸序列,其中 X^1 是芳族氨基酸,任选地选自W、Y、和F; X^2 选自非极性脂肪族氨基酸,任选地选自A、G、I、L、P和V;并且 X^3 是极性中性氨基酸,任选地选自N、C、Q、M、S和T。在一些实施方案中,RNA聚合酶变体包含SEQ ID NO:8的氨基酸序列。

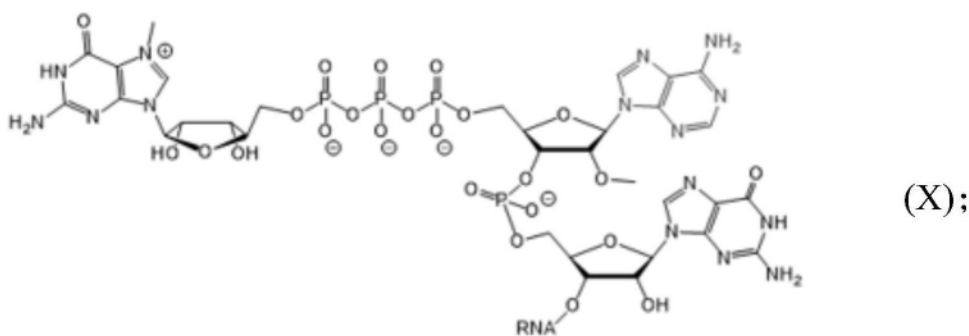
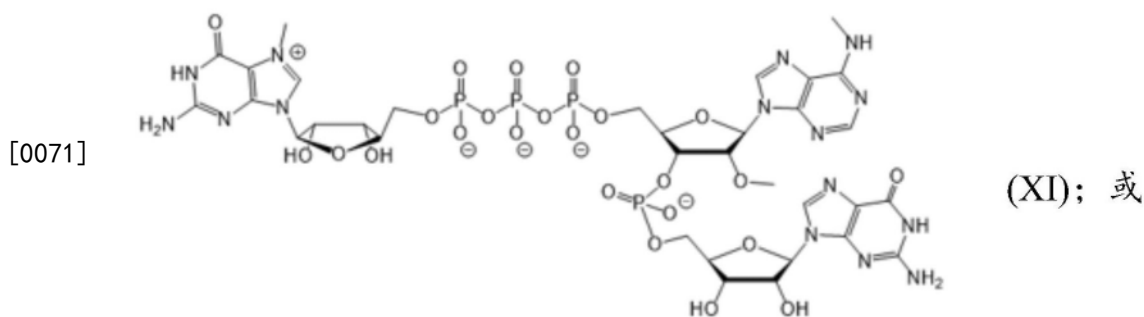
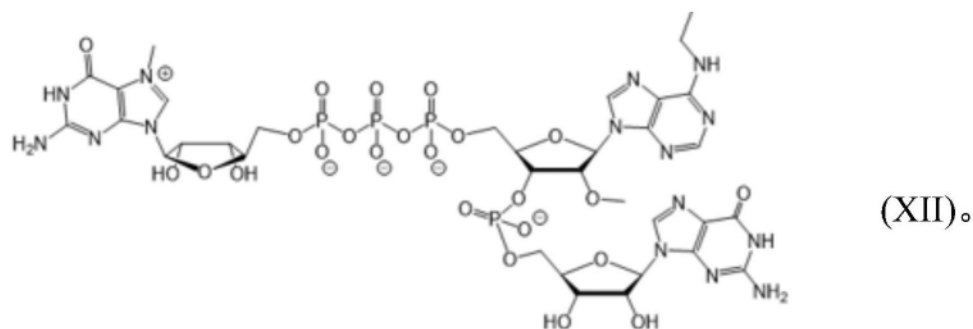
[0066] 一些方面提供了一种RNA聚合酶变体,所述RNA聚合酶变体包含SEQ ID NO:5的氨基酸序列,其中 X^1 是芳族氨基酸,任选地选自W、Y、和F; X^2 选自非极性脂肪族氨基酸,任选地选自A、G、I、L、P和V;并且 X^5 是芳族氨基酸,任选地选自W、Y和F。在一些实施方案中,RNA聚合酶变体包含SEQ ID NO:9的氨基酸序列。

[0067] 一些方面提供了一种方法,所述方法包括:在体外转录反应中产生信使RNA(mRNA),所述转录反应包括DNA、核苷三磷酸、前述段落中任一项的RNA聚合酶变体和任选的帽类似物。

[0068] 在一些实施方案中,反应包括帽类似物。

[0069] 在一些实施方案中,帽类似物是二核苷酸帽类似物、三核苷酸帽类似物或四核苷酸帽类似物。在一些实施方案中,帽类似物是四核苷酸帽类似物。

[0070] 在一些实施方案中,帽类似物是包含GAG序列的三核苷酸帽类似物。在一些实施方案中,GAG帽类似物包含选自以下的化合物:



[0072] 在一些实施方案中,四核苷酸帽类似物包含GGAG序列。在一些实施方案中,四核苷

附图说明

[0078] 图1A-1D示出了描绘由涉及示例性RNA聚合酶变体的体外转录 (IVT) 反应产生的转录RNA产物的功能特性的图表。寡dT纯化后,分析转录RNA产物的产率(图1A)、加帽RNA百分比(图1B)、根据Tris RP (反相) 方法的加尾百分比(即,包含多聚A尾的RNA的百分比)(图1C)和dsRNA的量(图1D)。

[0079] 图2A-2C示出了描绘在不同水平的GGAG帽类似物的存在下由涉及示例性RNA聚合酶变体的体外转录 (IVT) 反应产生的转录RNA产物的功能特性的图表。寡dT纯化后,分析转录RNA产物的加帽RNA百分比(图2A)、产率(图2B)和根据Tris RP (反相) 方法的加尾百分比(即,包含多聚A尾的RNA的百分比)(图2C)。

具体实施方式

[0080] RNA聚合酶(例如,DNA依赖性RNA聚合酶)是催化核糖核苷酸依序添加至生长RNA链的3'端(5'→3'方向上的RNA转录)的酶,其中核苷三磷酸(NTP)充当酶的底物并且核苷酸的序列由DNA模板指定。转录依赖于碱基的互补配对。双螺旋的两条链局部分离,并且分离的链之一充当模板(DNA模板)。然后RNA聚合酶通过模板中的互补碱基催化DNA模板上游离核苷酸的对齐。因此,如果RNA聚合酶催化核糖核苷酸依序添加至生长RNA链的3'端,则认为RNA聚合酶具有RNA聚合酶活性。

[0081] DNA指导的RNA聚合酶能够在没有引物的情况下启动RNA的合成;启动的第一个催化阶段被称为从头RNA合成。从头合成是转录周期中的独特阶段,其中RNA聚合酶结合两个核苷酸,而不是新生RNA聚合物和单个核苷酸。对于噬菌体T7 RNA聚合酶,转录明显优选从+1和+2位置处的GTP开始。起始核苷酸与RNA聚合酶结合的位置与针对延伸复合物所描述的位置不同(Kennedy WP等人J Mol Biol.2007;370(2):256-68)。有利于GTP作为起始核苷酸的选择偏差是通过形状互补性、广泛的蛋白质侧链和酶活性位点中的鸟嘌呤部分的强碱基堆积相互作用实现的。因此,起始GTP为开放启动子构象提供了最大的稳定力(Kennedy等人2007)。在一些实施方案中,RNA聚合酶变体包含用于从头RNA合成的一个或多个结合位点残基处的一个或多个氨基酸取代,不受理论的束缚,其改变了对体外转录反应的帽类似物的RNA聚合酶亲和力,例如使得在低帽类似物浓度下的加帽效率有所提高。

[0082] 因此,在一些方面,RNA聚合酶变体包括RNA聚合酶,所述RNA聚合酶包含用于从头RNA合成的结合位点残基处的两个或更多个氨基酸取代。RNA聚合酶变体是具有RNA聚合酶活性和相对于对应的野生型RNA聚合酶的至少一个取代和/或修饰的酶。在一些实施方案中,相对于野生型RNA聚合酶,氨基酸取代在选自位置350、351、387、437和653的位置处,其中野生型RNA聚合酶包含SEQ ID NO:1的氨基酸序列。

[0083] T7 RNA聚合酶的结构研究显示,N末端结构域的构象在转录的起始阶段与延伸阶段之间发生了实质变化。N末端结构域包含C-螺旋子结构域和启动子结合结构域,所述启动子结合结构域包括由子结构域H分隔的两个区段。启动子结合结构域和结合的启动子在8-nt RNA转录物的合成时旋转大约45度,从而允许在活性位点扩展以适应生长的异源双链体时维持启动子接触。C-螺旋子结构域向其延伸构象适度移动,而子结构域H保持在其起始阶段位置-而不是其距离超过70埃的延伸阶段位置。T7 RNA聚合酶起始和延伸复合物的结构比较揭示了N末端267个残基(N末端结构域)内的广泛构象变化和RNA聚合酶的其余部分的

极少变化。启动子结合结构域的刚体旋转以及N末端C-螺旋(残基28-71)和H(残基151-190)子结构域的重折叠负责消除启动子结合位点,扩大活性位点并为RNA转录物产生输出通道(exit tunnel)。特别地,T7 RNA聚合酶的残基E42-G47(在起始复合物中以 β -环结构存在)在延伸复合物中采用 α -螺旋结构。N末端结构域内的结构变化解释了延伸复合物的稳定性和持续合成能力的增加(参见,例如,Durniak,K.J.等人,Science 322(5901):553-557,2008,所述文献以引用的方式并入本文中)。T7 RNA聚合酶还包含‘N螺旋’(残基374-409),其功能是在RNA转录物与模板分离并影响延伸复合物的稳定性和持续合成能力时偏转RNA转录物的5'端的方向(例如,通过残基385-395与核糖主链之间的相互作用)。RNA聚合酶的‘O螺旋’(残基627-640)功能是在插入期间稳定传入的NTP并在RNA转录物的合成期间防止回溯。最后,‘Y螺旋’(残基644-661)功能是将模板碱基稳定在生长RNA转录物的n+1位置处。

[0084] 在一些方面,是促进从RNA聚合酶起始复合物至RNA聚合酶延伸复合物的构象变化的RNA聚合酶变体(例如,T7 RNA聚合酶变体)。在一些实施方案中,相对于野生型RNA聚合酶,RNA聚合酶变体包含至少一个、至少两个、至少三个或至少四个氨基酸修饰,这在RNA聚合酶变体从起始复合物转变为延伸复合物时引起RNA聚合酶变体的至少一个三维环结构经历构象变化为螺旋结构。因此,在一些实施方案中,相对于野生型氨基酸,至少一个氨基酸修饰具有高螺旋倾向。

[0085] 此外,在一些方面,是相对于野生型T7 RNA聚合酶,增加延伸复合物的稳定性和持续合成能力,防止回溯并稳定传入的NTP和模板的RNA聚合酶变体(例如,T7 RNA聚合酶变体)。在一些实施方案中,相对于野生型RNA聚合酶,RNA聚合酶变体包含至少一个、至少两个、至少三个或至少四个氨基酸修饰,其增加延伸复合物的稳定性和持续合成能力,防止回溯并稳定传入的NTP和模板。在一些实施方案中,相对于野生型RNA聚合酶,RNA聚合酶变体在‘N螺旋’(残基374-409)中包含至少一个、至少两个、至少三个或至少四个氨基酸修饰(例如,以增加延伸复合物的稳定性和持续合成能力)。在一些实施方案中,相对于野生型RNA聚合酶,RNA聚合酶变体在‘O螺旋’(残基627-640)中包含至少一个、至少两个、至少三个或至少四个氨基酸修饰(例如,以在插入期间稳定传入的NTP并防止回溯)。在一些实施方案中,相对于野生型RNA聚合酶,RNA聚合酶变体在‘Y螺旋’(残基644-661)中包含至少一个、至少两个、至少三个或至少四个氨基酸修饰(例如,以稳定生长RNA转录物)。

[0086] 因此,一些方面提供了相对于野生型RNA聚合酶,包含多个氨基酸取代和/或修饰的RNA聚合酶变体。在一些实施方案中,RNA聚合酶变体包含氨基酸序列,所述氨基酸序列包含(a)用于从头RNA合成的结合位点残基处的氨基酸取代,和(b)促进从RNA聚合酶起始复合物至RNA聚合酶延伸复合物的构象变化的氨基酸取代。

[0087] 在一些实施方案中,相对于对照RNA聚合酶,在体外转录反应中RNA聚合酶变体的使用增加了转录效率。举例来说,RNA聚合酶变体的使用可以使转录效率(例如,RNA产率和/或转录速率)增加至少20%。在一些实施方案中,RNA聚合酶变体的使用使转录效率(例如,RNA产率和/或转录速率)增加至少30%、至少40%、至少50%、至少60%、至少70%、至少80%、至少90%或至少10%。在一些实施方案中,RNA聚合酶变体的使用使转录效率增加20-100%、20-90%、20-80%、20-70%、20-60%、20-50%、30-100%、30-90%、30-80%、30-70%、30-60%、30-50%、40-100%、40-90%、40-80%、40-70%、40-60%、40-50%、50-100%、50-90%、50-80%、50-70%或50-60%。在一些实施方案中,RNA聚合酶变体的使用使

总RNA产率增加至少30%、至少40%、至少50%、至少60%、至少70%、至少80%、至少90%或至少10%。在一些实施方案中,RNA聚合酶变体的使用使总RNA产率增加20-100%、20-90%、20-80%、20-70%、20-60%、20-50%、30-100%、30-90%、30-80%、30-70%、30-60%、30-50%、40-100%、40-90%、40-80%、40-70%、40-60%、40-50%、50-100%、50-90%、50-80%、50-70%或50-60%。在一些实施方案中,RNA聚合酶变体的使用使转录速率增加至少30%、至少40%、至少50%、至少60%、至少70%、至少80%、至少90%或至少10%。在一些实施方案中,RNA聚合酶变体的使用使转录速率增加20-100%、20-90%、20-80%、20-70%、20-60%、20-50%、30-100%、30-90%、30-80%、30-70%、30-60%、30-50%、40-100%、40-90%、40-80%、40-70%、40-60%、40-50%、50-100%、50-90%、50-80%、50-70%或50-60%。在一些实施方案中,对照RNA聚合酶是包含SEQ ID NO:1的氨基酸序列的野生型RNA聚合酶(“野生型T7 RNA聚合酶”)。

[0088] 令人惊讶的是,RNA聚合酶变体使得能够在体外转录反应中使用低得多的浓度(量)的帽类似物,产生与使用野生型T7 RNA聚合酶产生的加帽RNA等量的加帽RNA。参见例如如图1A-2C和实施例1-2。在一些实施方案中,当在体外转录反应中使用一半浓度的帽类似物时,体外转录反应中RNA聚合酶变体的使用增加了加帽RNA的产率。在一些实施方案中,当在体外转录反应中使用仅25%、50%或75%浓度的帽类似物时,体外转录反应中RNA聚合酶变体的使用增加了加帽RNA的产率。举例来说,当在体外转录反应中使用仅25%、50%或75%浓度的帽类似物时,RNA聚合酶变体的使用可以使加帽RNA的产率增加至少20%。在一些实施方案中,当在体外转录反应中使用仅25%、50%或75%浓度的帽类似物时,RNA聚合酶变体的使用使加帽RNA的产率增加至少30%、至少40%、至少50%、至少60%、至少70%、至少80%、至少90%或至少100%。在一些实施方案中,当在体外转录反应中使用仅25%、50%或75%浓度的帽类似物时,RNA聚合酶变体的使用使加帽RNA的产率增加20-100%、20-90%、20-80%、20-70%、20-60%、20-50%、30-100%、30-90%、30-80%、30-70%、30-60%、30-50%、40-100%、40-90%、40-80%、40-70%、40-60%、40-50%、50-100%、50-90%、50-80%、50-70%或50-60%。在一些实施方案中,对照RNA聚合酶是野生型T7 RNA聚合酶。

[0089] 在一些实施方案中,RNA聚合酶变体的使用使加帽RNA的总产率增加至少30%、至少40%、至少50%、至少60%、至少70%、至少80%、至少90%或至少10%。在一些实施方案中,RNA聚合酶变体的使用使加帽RNA的总产率增加20-100%、20-90%、20-80%、20-70%、20-60%、20-50%、30-100%、30-90%、30-80%、30-70%、30-60%、30-50%、40-100%、40-90%、40-80%、40-70%、40-60%、40-50%、50-100%、50-90%、50-80%、50-70%或50-60%。

[0090] 在一些实施方案中,体外转录反应中RNA聚合酶变体的使用增加了共转录加帽效率。举例来说,RNA聚合酶变体的使用可以使共转录加帽效率(例如,包含帽类似物的转录物的百分比)增加至少20%。在一些实施方案中,RNA聚合酶变体的使用使共转录加帽效率(例如,包含帽类似物的转录物的百分比)增加至少30%、至少40%、至少50%、至少60%、至少70%、至少80%、至少90%或至少100%。在一些实施方案中,RNA聚合酶变体的使用使共转录加帽效率增加20-100%、20-90%、20-80%、20-70%、20-60%、20-50%、30-100%、30-90%、30-80%、30-70%、30-60%、30-50%、40-100%、40-90%、40-80%、40-70%、40-

60%、40-50%、50-100%、50-90%、50-80%、50-70%或50-60%。在一些实施方案中,对照RNA聚合酶是野生型T7 RNA聚合酶。

[0091] 在一些实施方案中,至少50%的mRNA包含功能性帽类似物。举例来说,至少50%、至少55%、至少60%、至少65%、至少70%、至少75%、至少80%、至少85%、至少95%或100%的mRNA可包含帽类似物。在一些实施方案中,50%-100%、50%-90%、50%-80%或50%-70%的mRNA包含帽类似物。

[0092] 在一些实施方案中,在体外转录反应中使用的一半浓度的帽类似物下,体外转录反应中RNA聚合酶变体的使用提高了RNA的3'同质性。举例来说,当在体外转录反应中使用仅25%、50%或75%浓度的帽类似物时,RNA聚合酶变体的使用可以使RNA的3'同质性提高至少20%。在一些实施方案中,当在体外转录反应中使用仅25%、50%或75%浓度的帽类似物时,RNA聚合酶变体的使用使3'同质性提高至少30%、至少40%、至少50%、至少60%、至少70%、至少80%、至少90%或至少100%。在一些实施方案中,当在体外转录反应中使用仅25%、50%或75%浓度的帽类似物时,RNA聚合酶变体的使用使3'同质性提高20-100%、20-90%、20-80%、20-70%、20-60%、20-50%、30-100%、30-90%、30-80%、30-70%、30-60%、30-50%、40-100%、40-90%、40-80%、40-70%、40-60%、40-50%、50-100%、50-90%、50-80%、50-70%或50-60%。在一些实施方案中,对照RNA聚合酶是野生型T7 RNA聚合酶。

[0093] 在一些实施方案中,体外转录反应中产生的包含RNA聚合酶变体的至少50%的mRNA表现出3'同质性。举例来说,至少55%、至少60%、至少65%、至少70%、至少75%、至少80%、至少85%、至少95%或100%的mRNA表现出3'同质性。在一些实施方案中,50%-100%、50%-90%、50%-80%或50%-70%的mRNA表现出3'同质性。

[0094] 在一些实施方案中,体外转录反应中产生的包含RNA聚合酶变体的mRNA具有大于阈值的3'同质性。在一些实施方案中,阈值为50%或至少50%。举例来说,阈值可以是55%、60%、65%、70%、75%、80%、85%或90%。

[0095] 在一些实施方案中,体外转录反应中RNA聚合酶变体的使用提高了转录保真度(例如,突变率)。举例来说,RNA聚合酶变体的使用可以使转录保真度提高至少20%。在一些实施方案中,RNA聚合酶变体的使用使转录保真度提高至少30%、至少40%、至少50%、至少60%、至少70%、至少80%、至少90%或至少100%。在一些实施方案中,RNA聚合酶变体的使用使转录保真度提高20-100%、20-90%、20-80%、20-70%、20-60%、20-50%、30-100%、30-90%、30-80%、30-70%、30-60%、30-50%、40-100%、40-90%、40-80%、40-70%、40-60%、40-50%、50-100%、50-90%、50-80%、50-70%或50-60%。提高转录保真度的RNA聚合酶变体将产生比对照RNA聚合酶具有更低的突变率或突变总数的RNA转录物(例如,mRNA转录物)。在一些实施方案中,对照RNA聚合酶是野生型T7 RNA聚合酶。

[0096] 在一些实施方案中,相对于DNA模板,使用RNA聚合酶变体产生的mRNA每100个核苷酸具有少于1个突变。举例来说,相对于DNA模板,产生的mRNA每200、300、400、500、600、700、800、900或1000个核苷酸可以具有少于1个突变。

[0097] 在一些实施方案中,体外转录反应中RNA聚合酶变体的使用降低了体外转录反应中双链RNA(dsRNA)污染的量。举例来说,RNA聚合酶变体的使用可以使体外转录反应中dsRNA污染的量降低至少20%。在一些实施方案中,RNA聚合酶变体的使用使体外转录反应

中dsRNA污染的量降低至少30%、至少40%、至少50%、至少60%、至少70%、至少80%、至少90%或至少100%。在一些实施方案中, RNA聚合酶变体的使用使体外转录反应中dsRNA污染的量降低20-100%、20-90%、20-80%、20-70%、20-60%、20-50%、30-100%、30-90%、30-80%、30-70%、30-60%、30-50%、40-100%、40-90%、40-80%、40-70%、40-60%、40-50%、50-100%、50-90%、50-80%、50-70%或50-60%。在一些实施方案中, 对照RNA聚合酶是野生型T7 RNA聚合酶。

[0098] 在一些实施方案中, dsRNA污染的浓度小于10ng/25μg的mRNA产物。在一些实施方案中, dsRNA污染的浓度小于5ng/25μg的mRNA产物。举例来说, dsRNA污染的浓度可以小于4ng/25μg的mRNA产物、小于3ng/25μg的mRNA产物、小于2ng/25μg的mRNA产物或小于1ng/25μg的mRNA产物。在一些实施方案中, dsRNA污染的浓度为0.5-1、0.5-2、0.5-3、0.4或0.5-5ng/25μg的mRNA产物。

[0099] 在一些实施方案中, 体外转录反应中产生的包含RNA聚合酶变体的mRNA具有低于阈值数量的dsRNA。在一些实施方案中, 阈值为10ng。在一些实施方案中, 阈值为5ng。在一些实施方案中, 阈值为4ng、3ng、2ng或1ng。

[0100] 氨基酸取代和修饰

[0101] 相对于野生型(WT) RNA聚合酶, RNA聚合酶变体包含至少一个氨基酸取代, 优选至少两个氨基酸取代。举例来说, 关于具有SEQ ID NO:1的氨基酸序列的WT T7 RNA聚合酶, 位置350处的谷氨酸(E)被认为是“野生型氨基酸”, 而位置350处的谷氨酸取代色氨酸被认为是“氨基酸取代”。在一些实施方案中, 相对于WT RNA聚合酶(例如, 具有SEQ ID NO:1的氨基酸序列的WT T7 RNA聚合酶), RNA聚合酶变体是包含至少一个(一个或多个)氨基酸取代的T7 RNA聚合酶变体。

[0102] 在一些实施方案中, RNA T7聚合酶变体包含至少两个氨基酸取代。在一些实施方案中, RNA T7聚合酶变体包含至少三个氨基酸取代。在一些实施方案中, RNA T7聚合酶变体包含至少四个氨基酸取代。在一些实施方案中, RNA T7聚合酶变体包含至少五个氨基酸取代。

[0103] 在一些实施方案中, RNA聚合酶变体包含氨基酸序列, 所述氨基酸序列包含(至少一个)氨基酸修饰, 这在RNA聚合酶变体从起始复合物转变为延伸复合物时引起RNA聚合酶变体的环结构经历构象变化为螺旋结构。在一些实施方案中, 氨基酸取代是高倾向性(high propensity)氨基酸取代。高螺旋倾向性氨基酸的实例包括丙氨酸、异亮氨酸、亮氨酸、精氨酸、蛋氨酸、赖氨酸、谷氨酰胺和/或谷氨酸。

[0104] 在一些实施方案中, RNA聚合酶变体包含氨基酸序列, 所述氨基酸序列包含(至少一个)氨基酸取代以引入极性中性氨基酸。在一些实施方案中, 极性中性氨基酸选自天冬酰胺(N)、半胱氨酸(C)、谷氨酰胺(Q)、蛋氨酸(M)、丝氨酸(S)和苏氨酸(T)。在一些实施方案中, RNA聚合酶变体包含氨基酸序列, 所述氨基酸序列包含(至少一个)氨基酸取代以引入芳香族氨基酸。在一些实施方案中, 芳香族氨基酸选自色氨酸(W)、酪氨酸(Y)和苯丙氨酸(F)。在一些实施方案中, RNA聚合酶变体包含氨基酸序列, 所述氨基酸序列包含(至少一个)氨基酸取代以引入非极性脂肪族氨基酸。在一些实施方案中, 非极性脂肪族氨基酸选自丙氨酸(A)、甘氨酸(G)、异亮氨酸(I)、亮氨酸(L)、脯氨酸(P)和缬氨酸(V)。在一些实施方案中, RNA聚合酶变体包含氨基酸序列, 所述氨基酸序列包含(至少一个)氨基酸取代以引入带正电荷的氨

氨酸。在一些实施方案中,带正电荷的氨基酸选自赖氨酸(K)、精氨酸(R)和组氨酸(H)。在一些实施方案中,RNA聚合酶变体包含氨基酸序列,所述氨基酸序列包含(至少一个)氨基酸取代以引入带负电荷的氨基酸。在一些实施方案中,带负电荷的氨基酸选自天冬氨酸(D)和谷氨酸(E)。

[0105] 在一些实施方案中,RNA聚合酶变体包含氨基酸序列,所述氨基酸序列包含在非保守氨基酸残基的位置处的(至少一个)氨基酸修饰。保守氨基酸残基是通常在相同蛋白质的多个同源序列之间共享的氨基酸或氨基酸类型(例如,单个氨基酸诸如Gly或Ser,或共享类似性质诸如具有酸性官能团的氨基酸的氨基酸组)。可以使用同源氨基酸序列的序列比对来鉴定保守氨基酸残基。使用基本局部比对(Basic Local Alignment)搜索获得的大约1000个RNA聚合酶序列的序列比对允许确定最有可能在RNA聚合酶序列中保守的SEQ ID NO:1的240个位置。最有可能在RNA聚合酶序列中保守的SEQ ID NO:1的这240个位置在位置5-6、39、269-277、279、281-282、323-333、411-448、454-470、472-474、497-516、532-560、562-573、626-646、691、693-702、724-738、775-794、805-820、828-833、865-867和877-879处。因此,在一些实施方案中,RNA聚合酶变体包括RNA聚合酶,所述RNA聚合酶包含在非SEQ ID NO:1的位置5-6、39、269-277、279、281-282、323-333、411-448、454-470、472-474、497-516、532-560、562-573、626-646、691、693-702、724-738、775-794、805-820、828-833、865-867和877-879中的一个的位置处的(至少一个)氨基酸修饰。在一些实施方案中,RNA聚合酶变体可进一步包含在非SEQ ID NO:1的位置5-6、39、269-277、279、281-282、323-333、411-448、454-470、472-474、497-516、532-560、562-573、626-646、691、693-702、724-738、775-794、805-820、828-833、865-867和877-879中的一个的任何数量的位置处的任何数量的氨基酸修饰。在一些实施方案中,包含SEQ ID NO:2-9中任一个的氨基酸序列的RNA聚合酶变体可进一步包含在非位置5-6、39、269-277、279、281-282、323-333、411-448、454-470、472-474、497-516、532-560、562-573、626-646、691、693-702、724-738、775-794、805-820、828-833、865-867和877-879中的一个的位置处的(至少一个)额外的氨基酸修饰。相反地,非保守的氨基酸位置最有可能发生修饰或突变。因此,在一些实施方案中,RNA聚合酶变体包括RNA聚合酶,所述RNA聚合酶包含在位置1-4、7-38、40-268、278、280、283-322、334-410、449-453、471、475-496、517-531、561、574-625、647-690、692、703-723、739-774、795-804、821-827、834-864、868-876和880-883处的(至少一个)氨基酸修饰。在一些实施方案中,包含SEQ ID NO:2-9中任一个的氨基酸序列的RNA聚合酶变体可进一步包含在位置1-4、7-38、40-268、278、280、283-322、334-410、449-453、471、475-496、517-531、561、574-625、647-690、692、703-723、739-774、795-804、821-827、834-864、868-876和880-883处的(至少一个)额外的氨基酸修饰。

[0106] 在一些实施方案中,包含SEQ ID NO:2-9中任一个的氨基酸序列的RNA聚合酶变体可进一步包含在不破坏RNA聚合酶蛋白的二级或三级结构的任何氨基酸位置处的(至少一个)氨基酸修饰。在一些实施方案中,包含SEQ ID NO:2-9中任一个的氨基酸序列的RNA聚合酶变体可进一步包含在不破坏RNA聚合酶蛋白折叠的能力的任何氨基酸位置处的(至少一个)氨基酸修饰。在一些实施方案中,包含SEQ ID NO:2-9中任一个的氨基酸序列的RNA聚合酶变体可进一步包含在不破坏RNA聚合酶蛋白结合核酸(例如,DNA)的能力的任何氨基酸位置处的(至少一个)氨基酸修饰。

[0107] 在一些实施方案中,相对于野生型RNA聚合酶,RNA聚合酶变体包含氨基酸序列,所述氨基酸序列包含位置437处的氨基酸取代(例如,N437Y)、位置387处的氨基酸取代(例如,K387S)、位置350处的氨基酸取代(例如,E350W)和位置351处的氨基酸取代(例如,D351V),其中野生型RNA聚合酶包含SEQ ID NO:1的氨基酸序列。在一些实施方案中,相对于野生型RNA聚合酶,RNA聚合酶变体包含N437Y、K387S、E350W和D351V取代,其中野生型RNA聚合酶包含SEQ ID NO:1的氨基酸序列。

[0108] 在一些实施方案中,相对于野生型RNA聚合酶,RNA聚合酶变体包含氨基酸序列,所述氨基酸序列包含位置437处的氨基酸取代(例如,N437Y)、位置350处的氨基酸取代(例如,E350W)和位置351处的氨基酸取代(例如,D351V),其中野生型RNA聚合酶包含SEQ ID NO:1的氨基酸序列。在一些实施方案中,相对于野生型RNA聚合酶,RNA聚合酶变体包含N437Y、E350W和D351V取代,其中野生型RNA聚合酶包含SEQ ID NO:1的氨基酸序列。

[0109] 在一些实施方案中,相对于野生型RNA聚合酶,RNA聚合酶变体包含氨基酸序列,所述氨基酸序列包含位置387处的氨基酸取代(例如,K387S)、位置350处的氨基酸取代(例如,E350W)和位置351处的氨基酸取代(例如,D351V),其中野生型RNA聚合酶包含SEQ ID NO:1的氨基酸序列。在一些实施方案中,相对于野生型RNA聚合酶,RNA聚合酶变体包含K387S、E350W和D351V取代,其中野生型RNA聚合酶包含SEQ ID NO:1的氨基酸序列。

[0110] 在一些实施方案中,相对于野生型RNA聚合酶,RNA聚合酶变体包含氨基酸序列,所述氨基酸序列包含位置653处的氨基酸取代(例如,D653W)、位置350处的氨基酸取代(例如,E350W)和位置351处的氨基酸取代(例如,D351V),其中野生型RNA聚合酶包含SEQ ID NO:1的氨基酸序列。在一些实施方案中,相对于野生型RNA聚合酶,RNA聚合酶变体包含D653W、E350W和D351V取代,其中野生型RNA聚合酶包含SEQ ID NO:1的氨基酸序列。

[0111] 在一些实施方案中,位置K387处的氨基酸取代是极性中性氨基酸。在一些实施方案中,极性中性氨基酸选自天冬酰胺(N)、半胱氨酸(C)、谷氨酰胺(Q)、蛋氨酸(M)、丝氨酸(S)和苏氨酸(T)。因此,在一些实施方案中,位置K387处的氨基酸取代是K387N、K387C、K387Q、K387M、K387S或K387T。

[0112] 在一些实施方案中,位置N437处的氨基酸取代是芳族氨基酸。在一些实施方案中,芳族氨基酸选自色氨酸(W)、酪氨酸(Y)和苯丙氨酸(F)。因此,在一些实施方案中,位置N437处的氨基酸取代是N437W、N437Y或N437F。

[0113] 在一些实施方案中,位置D653处的氨基酸取代是芳族氨基酸。在一些实施方案中,芳族氨基酸选自色氨酸(W)、酪氨酸(Y)和苯丙氨酸(F)。因此,在一些实施方案中,位置D653处的氨基酸取代是D653W、D653Y或D653F。

[0114] 在一些实施方案中,位置E350处的氨基酸取代是芳族氨基酸。在一些实施方案中,芳族氨基酸选自色氨酸(W)、酪氨酸(Y)和苯丙氨酸(F)。因此,在一些实施方案中,位置E350处的氨基酸取代是E350W、E350Y或E350F。

[0115] 在一些实施方案中,位置D351处的氨基酸取代是非极性脂肪族氨基酸。在一些实施方案中,非极性脂肪族氨基酸选自丙氨酸(A)、甘氨酸(G)、异亮氨酸(I)、亮氨酸(L)、脯氨酸(P)和缬氨酸(V)。因此,在一些实施方案中,位置D351处的氨基酸取代是D351A、D351G、D351I、D351L、D351P或D351V。

[0116] 在一些实施方案中,相对于野生型RNA聚合酶,RNA聚合酶变体包含氨基酸序列,所

述氨基酸序列包含位置R379处的氨基酸取代(例如,R379A),其中野生型RNA聚合酶包含SEQ ID NO:1的氨基酸序列。在一些实施方案中,位置R379处的氨基酸取代是非极性脂肪族氨基酸;极性中性氨基酸;芳族氨基酸或带电荷的氨基酸。在一些实施方案中,非极性脂肪族氨基酸选自丙氨酸(A)、甘氨酸(G)、异亮氨酸(I)、亮氨酸(L)、脯氨酸(P)和缬氨酸(V)。在一些实施方案中,芳族氨基酸选自色氨酸(W)、酪氨酸(Y)和苯丙氨酸(F)。在一些实施方案中,带电荷的氨基酸是带正电荷的氨基酸(例如,赖氨酸(K)或组氨酸(H))或带负电荷的氨基酸(例如,谷氨酸(E)或天冬氨酸(D))。在一些实施方案中,位置R379处的氨基酸取代是R379A、R379K、R379E或R379W。

[0117] 在一些实施方案中,相对于野生型RNA聚合酶,RNA聚合酶变体包含氨基酸序列,所述氨基酸序列包含位置Y385处的氨基酸取代(例如,Y385A),其中野生型RNA聚合酶包含SEQ ID NO:1的氨基酸序列。在一些实施方案中,位置Y385处的氨基酸取代是非极性脂肪族氨基酸;极性中性氨基酸;芳族氨基酸或带电荷的氨基酸。在一些实施方案中,非极性脂肪族氨基酸选自丙氨酸(A)、甘氨酸(G)、异亮氨酸(I)、亮氨酸(L)、脯氨酸(P)和缬氨酸(V)。在一些实施方案中,芳族氨基酸选自色氨酸(W)和苯丙氨酸(F)。在一些实施方案中,带电荷的氨基酸是带正电荷的氨基酸(例如,赖氨酸(K)、组氨酸(H)或精氨酸(R))或带负电荷的氨基酸(例如,谷氨酸(E)或天冬氨酸(D))。在一些实施方案中,位置Y385处的氨基酸取代是Y385A、Y385K、Y385W或Y385V。

[0118] 在一些实施方案中,相对于野生型RNA聚合酶,RNA聚合酶变体包含氨基酸序列,所述氨基酸序列包含位置R386处的氨基酸取代(例如,R386A),其中野生型RNA聚合酶包含SEQ ID NO:1的氨基酸序列。在一些实施方案中,位置R386处的氨基酸取代是非极性脂肪族氨基酸;极性中性氨基酸;芳族氨基酸或带电荷的氨基酸。在一些实施方案中,非极性脂肪族氨基酸选自丙氨酸(A)、甘氨酸(G)、异亮氨酸(I)、亮氨酸(L)、脯氨酸(P)和缬氨酸(V)。在一些实施方案中,带电荷的氨基酸是带正电荷的氨基酸(例如,赖氨酸(K)或组氨酸(H))或带负电荷的氨基酸(例如,谷氨酸(E)或天冬氨酸(D))。在一些实施方案中,芳族氨基酸选自色氨酸(W)、酪氨酸(Y)和苯丙氨酸(F)。在一些实施方案中,位置R386处的氨基酸取代是R386A、R386K或R386Y。

[0119] 在一些实施方案中,相对于野生型RNA聚合酶,RNA聚合酶变体包含氨基酸序列,所述氨基酸序列包含位置D388处的氨基酸取代(例如,D388A),其中野生型RNA聚合酶包含SEQ ID NO:1的氨基酸序列。在一些实施方案中,位置D388处的氨基酸取代是非极性脂肪族氨基酸;极性中性氨基酸;芳族氨基酸或带电荷的氨基酸。在一些实施方案中,非极性脂肪族氨基酸选自丙氨酸(A)、甘氨酸(G)、异亮氨酸(I)、亮氨酸(L)、脯氨酸(P)和缬氨酸(V)。在一些实施方案中,极性中性氨基酸选自天冬酰胺(N)、半胱氨酸(C)、谷氨酰胺(Q)、蛋氨酸(M)、丝氨酸(S)和苏氨酸(T)。在一些实施方案中,芳族氨基酸选自色氨酸(W)、酪氨酸(Y)和苯丙氨酸(F)。在一些实施方案中,位置D388处的氨基酸取代是D388A、D388N或D388Y。

[0120] 在一些实施方案中,相对于野生型RNA聚合酶,RNA聚合酶变体包含氨基酸序列,所述氨基酸序列包含位置K389处的氨基酸取代(例如,K389A),其中野生型RNA聚合酶包含SEQ ID NO:1的氨基酸序列。在一些实施方案中,位置K389处的氨基酸取代是非极性脂肪族氨基酸;极性中性氨基酸;芳族氨基酸或带电荷的氨基酸。在一些实施方案中,非极性脂肪族氨基酸选自丙氨酸(A)、甘氨酸(G)、异亮氨酸(I)、亮氨酸(L)、脯氨酸(P)和缬氨酸(V)。在一些

实施方案中,极性中性氨基酸选自天冬酰胺(N)、半胱氨酸(C)、谷氨酰胺(Q)、蛋氨酸(M)、丝氨酸(S)和苏氨酸(T)。在一些实施方案中,芳族氨基酸选自色氨酸(W)、酪氨酸(Y)和苯丙氨酸(F)。在一些实施方案中,位置K389处的氨基酸取代是K389A、K389S或K389Y。

[0121] 在一些实施方案中,相对于野生型RNA聚合酶,RNA聚合酶变体包含氨基酸序列,所述氨基酸序列包含位置R391处的氨基酸取代(例如,R391A),其中野生型RNA聚合酶包含SEQ ID NO:1的氨基酸序列。在一些实施方案中,位置R391处的氨基酸取代是非极性脂肪族氨基酸;极性中性氨基酸;芳族氨基酸或带电荷的氨基酸。在一些实施方案中,非极性脂肪族氨基酸选自丙氨酸(A)、甘氨酸(G)、异亮氨酸(I)、亮氨酸(L)、脯氨酸(P)和缬氨酸(V)。在一些实施方案中,位置R391处的氨基酸取代是R391A。

[0122] 在一些实施方案中,相对于野生型RNA聚合酶,RNA聚合酶变体包含氨基酸序列,所述氨基酸序列包含位置R394处的氨基酸取代(例如,R394A),其中野生型RNA聚合酶包含SEQ ID NO:1的氨基酸序列。在一些实施方案中,位置R394处的氨基酸取代是非极性脂肪族氨基酸;极性中性氨基酸;芳族氨基酸或带电荷的氨基酸。在一些实施方案中,非极性脂肪族氨基酸选自丙氨酸(A)、甘氨酸(G)、异亮氨酸(I)、亮氨酸(L)、脯氨酸(P)和缬氨酸(V)。在一些实施方案中,极性中性氨基酸选自天冬酰胺(N)、半胱氨酸(C)、谷氨酰胺(Q)、蛋氨酸(M)、丝氨酸(S)和苏氨酸(T)。在一些实施方案中,芳族氨基酸选自色氨酸(W)、酪氨酸(Y)和苯丙氨酸(F)。在一些实施方案中,位置R394处的氨基酸取代是R394A、R394Q或R394Y。

[0123] 在一些实施方案中,相对于野生型RNA聚合酶,RNA聚合酶变体包含氨基酸序列,所述氨基酸序列包含位置R395处的氨基酸取代(例如,R395A),其中野生型RNA聚合酶包含SEQ ID NO:1的氨基酸序列。在一些实施方案中,位置R395处的氨基酸取代是非极性脂肪族氨基酸;极性中性氨基酸;芳族氨基酸或带电荷的氨基酸。在一些实施方案中,非极性脂肪族氨基酸选自丙氨酸(A)、甘氨酸(G)、异亮氨酸(I)、亮氨酸(L)、脯氨酸(P)和缬氨酸(V)。在一些实施方案中,位置R395处的氨基酸取代是R395A。

[0124] 在一些实施方案中,相对于野生型RNA聚合酶,RNA聚合酶变体包含氨基酸序列,所述氨基酸序列包含位置D471处的氨基酸取代(例如,D471A),其中野生型RNA聚合酶包含SEQ ID NO:1的氨基酸序列。在一些实施方案中,位置D471处的氨基酸取代是非极性脂肪族氨基酸;极性中性氨基酸;芳族氨基酸或带电荷的氨基酸。在一些实施方案中,非极性脂肪族氨基酸选自丙氨酸(A)、甘氨酸(G)、异亮氨酸(I)、亮氨酸(L)、脯氨酸(P)和缬氨酸(V)。在一些实施方案中,带电荷的氨基酸是带正电荷的氨基酸(例如,赖氨酸(K)、组氨酸(H)或精氨酸(R))或带负电荷的氨基酸(例如,谷氨酸)。在一些实施方案中,芳族氨基酸选自色氨酸(W)、酪氨酸(Y)和苯丙氨酸(F)。在一些实施方案中,位置D471处的氨基酸取代是D471A、D471E或D471Y。

[0125] 在一些实施方案中,相对于野生型RNA聚合酶,RNA聚合酶变体包含氨基酸序列,所述氨基酸序列包含位置R627处的氨基酸取代(例如,R627A),其中野生型RNA聚合酶包含SEQ ID NO:1的氨基酸序列。在一些实施方案中,位置R627处的氨基酸取代是非极性脂肪族氨基酸;极性中性氨基酸;芳族氨基酸或带电荷的氨基酸。在一些实施方案中,非极性脂肪族氨基酸选自丙氨酸(A)、甘氨酸(G)、异亮氨酸(I)、亮氨酸(L)、脯氨酸(P)和缬氨酸(V)。在一些实施方案中,位置R627处的氨基酸取代是R627A。

[0126] 在一些实施方案中,相对于野生型RNA聚合酶,RNA聚合酶变体包含氨基酸序列,所

述氨基酸序列包含位置R631处的氨基酸取代(例如,R631A),其中野生型RNA聚合酶包含SEQ ID NO:1的氨基酸序列。在一些实施方案中,位置R631处的氨基酸取代是非极性脂肪族氨基酸;极性中性氨基酸;芳族氨基酸或带电荷的氨基酸。在一些实施方案中,非极性脂肪族氨基酸选自丙氨酸(A)、甘氨酸(G)、异亮氨酸(I)、亮氨酸(L)、脯氨酸(P)和缬氨酸(V)。在一些实施方案中,位置R631处的氨基酸取代是R631A。

[0127] 在一些实施方案中,相对于野生型RNA聚合酶,RNA聚合酶变体包含氨基酸序列,所述氨基酸序列包含位置R632处的氨基酸取代(例如,R632A),其中野生型RNA聚合酶包含SEQ ID NO:1的氨基酸序列。在一些实施方案中,位置R632处的氨基酸取代是非极性脂肪族氨基酸;极性中性氨基酸;芳族氨基酸或带电荷的氨基酸。在一些实施方案中,非极性脂肪族氨基酸选自丙氨酸(A)、甘氨酸(G)、异亮氨酸(I)、亮氨酸(L)、脯氨酸(P)和缬氨酸(V)。在一些实施方案中,芳族氨基酸选自色氨酸(W)、酪氨酸(Y)和苯丙氨酸(F)。在一些实施方案中,极性中性氨基酸选自天冬酰胺(N)、半胱氨酸(C)、谷氨酰胺(Q)、蛋氨酸(M)、丝氨酸(S)和苏氨酸(T)。在一些实施方案中,位置R632处的氨基酸取代是R632D、R632A、R632Q或R632Y。

[0128] 在一些实施方案中,相对于野生型RNA聚合酶,RNA聚合酶变体包含氨基酸序列,所述氨基酸序列包含位置G640处的氨基酸取代(例如,G640A),其中野生型RNA聚合酶包含SEQ ID NO:1的氨基酸序列。在一些实施方案中,位置G640处的氨基酸取代是非极性脂肪族氨基酸;极性中性氨基酸;芳族氨基酸或带电荷的氨基酸。在一些实施方案中,非极性脂肪族氨基酸选自丙氨酸(A)、甘氨酸(G)、异亮氨酸(I)、亮氨酸(L)、脯氨酸(P)和缬氨酸(V)。在一些实施方案中,位置G640处的氨基酸取代是G640A。

[0129] 在一些实施方案中,相对于野生型RNA聚合酶,RNA聚合酶变体包含氨基酸序列,所述氨基酸序列包含位置G645处的氨基酸取代(例如,G645A),其中野生型RNA聚合酶包含SEQ ID NO:1的氨基酸序列。在一些实施方案中,位置G645处的氨基酸取代是非极性脂肪族氨基酸;极性中性氨基酸;芳族氨基酸或带电荷的氨基酸。在一些实施方案中,非极性脂肪族氨基酸选自丙氨酸(A)、甘氨酸(G)、异亮氨酸(I)、亮氨酸(L)、脯氨酸(P)和缬氨酸(V)。在一些实施方案中,位置G645处的氨基酸取代是G645A。

[0130] 在一些实施方案中,相对于野生型RNA聚合酶,RNA聚合酶变体包含氨基酸序列,所述氨基酸序列包含位置Q648处的氨基酸取代(例如,Q648A),其中野生型RNA聚合酶包含SEQ ID NO:1的氨基酸序列。在一些实施方案中,位置Q648处的氨基酸取代是非极性脂肪族氨基酸;极性中性氨基酸;芳族氨基酸或带电荷的氨基酸。在一些实施方案中,非极性脂肪族氨基酸选自丙氨酸(A)、甘氨酸(G)、异亮氨酸(I)、亮氨酸(L)、脯氨酸(P)和缬氨酸(V)。在一些实施方案中,芳族氨基酸选自色氨酸(W)、酪氨酸(Y)和苯丙氨酸(F)。在一些实施方案中,带电荷的氨基酸是带正电荷的氨基酸(例如,赖氨酸(K)、组氨酸(H)或精氨酸(R))或带负电荷的氨基酸(例如,谷氨酸(E)或天冬氨酸(D))。在一些实施方案中,位置Q648处的氨基酸取代是Q648A、Q648R、Q648D、Q648E或Q648Y。

[0131] 在一些实施方案中,相对于野生型RNA聚合酶,RNA聚合酶变体包含氨基酸序列,所述氨基酸序列包含位置Q649处的氨基酸取代(例如,Q649A),其中野生型RNA聚合酶包含SEQ ID NO:1的氨基酸序列。在一些实施方案中,位置Q649处的氨基酸取代是非极性脂肪族氨基酸;极性中性氨基酸;芳族氨基酸或带电荷的氨基酸。在一些实施方案中,非极性脂肪族氨基酸选自丙氨酸(A)、甘氨酸(G)、异亮氨酸(I)、亮氨酸(L)、脯氨酸(P)和缬氨酸(V)。在一些

实施方案中,位置Q649处的氨基酸取代是Q649A。

[0132] 在一些实施方案中,相对于野生型RNA聚合酶,RNA聚合酶变体包含氨基酸序列,所述氨基酸序列包含位置E652处的氨基酸取代(例如,E652A),其中野生型RNA聚合酶包含SEQ ID NO:1的氨基酸序列。在一些实施方案中,位置E652处的氨基酸取代是非极性脂肪族氨基酸;极性中性氨基酸;芳族氨基酸或带电荷的氨基酸。在一些实施方案中,非极性脂肪族氨基酸选自丙氨酸(A)、甘氨酸(G)、异亮氨酸(I)、亮氨酸(L)、脯氨酸(P)和缬氨酸(V)。在一些实施方案中,芳族氨基酸选自色氨酸(W)、酪氨酸(Y)和苯丙氨酸(F)。在一些实施方案中,带电荷的氨基酸是带正电荷的氨基酸(例如,赖氨酸(K)、组氨酸(H)或精氨酸(R))。在一些实施方案中,位置E652处的氨基酸取代是E652R、E652A或E652Y。

[0133] 在一些实施方案中,相对于野生型RNA聚合酶,RNA聚合酶变体包含氨基酸序列,所述氨基酸序列包含位置D653处的氨基酸取代(例如,D653A),其中野生型RNA聚合酶包含SEQ ID NO:1的氨基酸序列。在一些实施方案中,位置D653处的氨基酸取代是非极性脂肪族氨基酸;极性中性氨基酸;芳族氨基酸或带电荷的氨基酸。在一些实施方案中,非极性脂肪族氨基酸选自丙氨酸(A)、甘氨酸(G)、异亮氨酸(I)、亮氨酸(L)、脯氨酸(P)和缬氨酸(V)。在一些实施方案中,芳族氨基酸选自色氨酸(W)、酪氨酸(Y)和苯丙氨酸(F)。在一些实施方案中,带电荷的氨基酸是带正电荷的氨基酸(例如,赖氨酸(K)、组氨酸(H)或精氨酸(R))。在一些实施方案中,位置D653处的氨基酸取代是D653K、D653A或D653Y。

[0134] 在一些实施方案中,相对于野生型RNA聚合酶,RNA聚合酶变体包含氨基酸序列,所述氨基酸序列包含位置Q656处的氨基酸取代(例如,Q656A),其中野生型RNA聚合酶包含SEQ ID NO:1的氨基酸序列。在一些实施方案中,位置Q656处的氨基酸取代是非极性脂肪族氨基酸;极性中性氨基酸;芳族氨基酸或带电荷的氨基酸。在一些实施方案中,非极性脂肪族氨基酸选自丙氨酸(A)、甘氨酸(G)、异亮氨酸(I)、亮氨酸(L)、脯氨酸(P)和缬氨酸(V)。在一些实施方案中,芳族氨基酸选自色氨酸(W)、酪氨酸(Y)和苯丙氨酸(F)。在一些实施方案中,带电荷的氨基酸是带正电荷的氨基酸(例如,赖氨酸(K)、组氨酸(H)或精氨酸(R))或带负电荷的氨基酸(例如,谷氨酸(E)或天冬氨酸(D))。在一些实施方案中,位置Q656处的氨基酸取代是Q656K、Q656A、Q656E或Q656Y。

[0135] 在一些实施方案中,相对于野生型RNA聚合酶,RNA聚合酶变体包含氨基酸序列,所述氨基酸序列包含位置P657处的氨基酸取代(例如,P657A),其中野生型RNA聚合酶包含SEQ ID NO:1的氨基酸序列。在一些实施方案中,位置P657处的氨基酸取代是非极性脂肪族氨基酸;极性中性氨基酸;芳族氨基酸或带电荷的氨基酸。在一些实施方案中,非极性脂肪族氨基酸选自丙氨酸(A)、甘氨酸(G)、异亮氨酸(I)、亮氨酸(L)、脯氨酸(P)和缬氨酸(V)。在一些实施方案中,芳族氨基酸选自色氨酸(W)、酪氨酸(Y)和苯丙氨酸(F)。在一些实施方案中,带电荷的氨基酸是带正电荷的氨基酸(例如,赖氨酸(K)、组氨酸(H)或精氨酸(R))或带负电荷的氨基酸(例如,谷氨酸(E)或天冬氨酸(D))。在一些实施方案中,位置P657处的氨基酸取代是P657G、P657A、P657E或P657Y。

[0136] 表1.RNA聚合酶变体

[0137]

RNA 聚合 酶变体	氨基酸序列	SEQ ID NO
E350X ¹ D351X ² K387X ³ N437X ⁴	MNTINIAKNDFSDIELAAIPFNTLADHYGERLAREQ LALEHESYEMGEARFRKMFERQLKAGEVADNAAA KPLITTLLPKMIARINDWFEEVKAKRGKRPTAFQFL QEIKPEAVAYITIKTTLACLTADNTTVQAVASAIG RAIEDEARFGRIRDLEAKHFKNVEEQLNKRVGHV YKKAQFMQVVEADMLSKGLLGGEAWSSWHKEDSI HVGVRCEIEMLIESTGMVSLHRQNAGVVGQDSETIE LAPEYAEAIATRAGALAGISPMFQPCVPPKPWTGI TGGGYWANGRRPLALVRTHSKKALMRYEDVYMP EVYKAINIAQNTAWKINKKVLAVANVITKWKHCP VX ¹ X ² IPAIEREELPMKPEDIDMNPEALTAWKRAAA AVYRX ³ DKARKSRRISLEFMLEQANKFANHKAIFW PYNMDWRGRVYAVSMFNPQG ⁴ DMTKGLLTLAK GKPIGKEGYWYWLKIHGANCAGVDKVPFPERIKFIEE NHENIMACAKSPLENTWWAEQDSPFCFLAFCFEYA GVQHHGLSYNCSLPLAFDGSCSGIQHFSAMLRDEV GGRAVNLLPSETVQDIYGIVAKKVNEILQADAINGT DNEVTVTDENTGEISEKVKLGTKALAGQWLAYG VTRSVTKRSVMTLAYGSKEFGFRQQVLEDTIQPAI DSGKGLMFTQPNQAAGYMAKLIWESVSVTVVAA VEAMNWLKSAKLLAAEVKDKKTGEILRKRCV HWVTPDGFVPVWQEYKKPIQTRLNLMFLGQFRLQP TINTNKDSEIDAHKQESGIAPNFVHSQDGSHLRKTV VWAHEKYGIESFALIHDSFGTIPADAANLFAVRET MVDTYESCDVLADFYDQFADQLHESQLDKMPALP AKGNLNLRDILESDFABA, 其中 X ¹ 是芳族氨基酸, 任选地选自 W、Y 和 F; X ² 选自非极性脂肪族氨基酸, 任选地选自 A、G、I、L、P 和 V; X ³ 是极性中性氨基 酸, 任选地选自 N、C、Q、M、S 和 T; 并且 X ⁴ 是芳 族氨基酸, 任选地选自 W、Y 和 F	2
E350X ¹ D351X ² N437X ⁴	MNTINIAKNDFSDIELAAIPFNTLADHYGERLAREQ LALEHESYEMGEARFRKMFERQLKAGEVADNAAA KPLITTLLPKMIARINDWFEEVKAKRGKRPTAFQFL QEIKPEAVAYITIKTTLACLTADNTTVQAVASAIG RAIEDEARFGRIRDLEAKHFKNVEEQLNKRVGHV YKKAQFMQVVEADMLSKGLLGGEAWSSWHKEDSI HVGVRCEIEMLIESTGMVSLHRQNAGVVGQDSETIE LAPEYAEAIATRAGALAGISPMFQPCVPPKPWTGI TGGGYWANGRRPLALVRTHSKKALMRYEDVYMP EVYKAINIAQNTAWKINKKVLAVANVITKWKHCP VX ¹ X ² IPAIEREELPMKPEDIDMNPEALTAWKRAAA AVYRKDKARKSRRISLEFMLEQANKFANHKAIFW YNMDWRGRVYAVSMFNPQG ⁴ DMTKGLLTLAKG KPIGKEGYWYWLKIHGANCAGVDKVPFPERIKFIEEN HENIMACAKSPLENTWWAEQDSPFCFLAFCFEYAG VQHHGLSYNCSLPLAFDGSCSGIQHFSAMLRDEVG GRAVNLLPSETVQDIYGIVAKKVNEILQADAINGTD	3

RNA 聚合酶变体	氨基酸序列	SEQ ID NO
	NEVVTVDENTGEISEKVKLGTKALAGQWLAYGV TRSVTKRSVMTLAYGSKEFGFRQQVLEDTIQPAIDS GKGLMFTQPNQAAGYMAKLIWESVSVTVVAAVE AMNWLKSAAKLLAAEVKDKKTGEILRKRCVHW VTPDGFPVWQEYKKPIQTRLNLMFLGQFRLQPTIN TNKDSEIDAHKQESGIAPNFVHSQDGSHLRKTVVW AHEKYGIESFALIHDSFGTIPADAANLFKAVRETMV DTYESCDVLADFYDQFADQLHESQLDKMPALPAK GNLNLRDILESDFafa, 其中 X ¹ 是芳族氨基酸, 任 选地选自 W、Y 和 F; X ² 选自非极性脂肪族氨基酸, 任 选地选自 A、G、I、L、P 和 V; 并且 X ⁴ 是芳族氨基 酸, 任选地选自 W、Y 和 F	
E350X ¹ D351X ² K387X ³	MNTINIAKNDFSDIELAAIPFNTLADHYGERLAREQ LALHEHESYEMGEARFRKMFERQLKAGEVADNAAA KPLITTLLPKMIARINDWFEEVKAKRGKRPTAFQFL QEIPEAVAYITIKTTLACLTSADNTTVQAVASAIG RAIEDEARFGRIRDLEAKHFKNVEEQLNKRVGHV YKKAQVVEADMLSKGLLGGEAWSSWHKEDSI HVGVRCEIEMLIESTGMVSLHRQNAGVVGQDSETIE LAPEYAEAIATRAGALAGISPMFQPCVPPKPTWTGI TGGGYWANGRRPLALVRTHSKKALMRYEDVYMP EVYKAINIAQNTAWKINKKVLAVANVITKWKHCP VX ¹ X ² IPAIEREELPMKPEDIDMNPEALTAWKRAAA AVYRX ³ DKARKSRISLEFMLEQANKFANHKAIF PYNMDWRGRVYAVSMFNPQGNDMTKGLLTLAKG KPIGKEGYWLKIHGANCAGVDKVPFPERIKFIEEN HENIMACAKSPLENTWWAEQDSPFCFLAFCFEYAG VQHHGLSYNCSLPLAFDGSCSGIQHFSAMLRDEVG GRAVNLLPSETVQDIYGIVAKKVNEILQADAINGTD NEVVTVDENTGEISEKVKLGTKALAGQWLAYGV TRSVTKRSVMTLAYGSKEFGFRQQVLEDTIQPAIDS GKGLMFTQPNQAAGYMAKLIWESVSVTVVAAVE AMNWLKSAAKLLAAEVKDKKTGEILRKRCVHW VTPDGFPVWQEYKKPIQTRLNLMFLGQFRLQPTIN TNKDSEIDAHKQESGIAPNFVHSQDGSHLRKTVVW AHEKYGIESFALIHDSFGTIPADAANLFKAVRETMV DTYESCDVLADFYDQFADQLHESQLDKMPALPAK GNLNLRDILESDFafa, 其中 X ¹ 是芳族氨基酸, 任 选地选自 W、Y 和 F; X ² 选自非极性脂肪族氨基酸, 任 选地选自 A、G、I、L、P 和 V; 并且 X ³ 是极性中性 氨基酸, 任选地选自 N、C、Q、M、S 和 T	4
E350X ¹ D351X ² D653X ⁵	MNTINIAKNDFSDIELAAIPFNTLADHYGERLAREQ LALHEHESYEMGEARFRKMFERQLKAGEVADNAAA KPLITTLLPKMIARINDWFEEVKAKRGKRPTAFQFL QEIPEAVAYITIKTTLACLTSADNTTVQAVASAIG RAIEDEARFGRIRDLEAKHFKNVEEQLNKRVGHV YKKAQVVEADMLSKGLLGGEAWSSWHKEDSI	5

[0138]

[0139]

RNA 聚合 酶变体	氨基酸序列	SEQ ID NO
	HVGVRCEIEMLIESTGMVSLHRQNAGVVGQDSETIE LAPEYAEAIATRAGALAGISPMFQPCVVPPKPWTGI TGGGYWANGRRPLALVRTHSKKALMRYEDVYMP EVYKAINIAQNTAWKINKKVLAVANVITKWKHCP VX ¹ X ² IPAIEREELPMKPEDIDMNPEALTAWKRAAA AVYRKDKARKSRRISLEFMLEQANKFANHKAIWFP YNMDWRGRVYAVSMFNPQGNDMTKGLLTLAKG KPIGKEGYWYWLKIHGANCAGVDKVPFPERIKFIEEN HENIMACAKSPLENTWWAEQDSPFCFLAFCFEYAG VQHHGLSYNCSLPLAFDGS CSGIQHFSAMLRDEVG GRAVNLLPSETVQDIYGIVAKKVNEILQADAINGTD NEVVTVTDENTGEISEKVKLGTKALAGQWLAYGV TRSVTKRSVMTLAYGSKEFGFRQQVLEX ⁵ TIQPAID SGKGLMFTQPNQAAGYMAKLIWESVSVTVVAAVE AMNWLKSAKLLAAEVKDKKTGEILRKRCVHW VTPDGFPVWQEYKKPIQTRLNLMFLGQFRLQPTIN TNKDSEIDAHKQESGIAPNFVHSQDGSHLRKTVVW AHEKYGIESFALIHDSFGTIPADAANLFKAVRETMV DTYESCDVLADFYDQFADQLHESQLDKMPALPAK GNLNLRDILESDFafa, 其中 X ¹ 是芳族氨基酸, 任选 地选自 W、Y 和 F; X ² 选自非极性脂肪族氨基酸, 任 选地选自 A、G、I、L、P 和 V; 并且 X ⁵ 是芳族氨基 酸, 任选地选自 W、Y 和 F	
E350W D351V K387S N437Y	MNTINIAKNDFSDIELAIPFNTLADHYGERLAREQ LALEHESYEMGEARFRKMFERQLKAGEVADNAAA KPLITTLLPKMIARINDWFEEVKA KRGRPTAFQFL QEIKPEAVAYITIKTTLACLTSADNTTVQAVASAIG RAIEDEARFGRIRDLEAKHFKNVEEQLNKRVGHV YKKA FMQVVEADMLSKGLLGGEAWSSWHKEDSI HVGVRCEIEMLIESTGMVSLHRQNAGVVGQDSETIE LAPEYAEAIATRAGALAGISPMFQPCVVPPKPWTGI TGGGYWANGRRPLALVRTHSKKALMRYEDVYMP EVYKAINIAQNTAWKINKKVLAVANVITKWKHCP VVWIPAIEREELPMKPEDIDMNPEALTAWKRAAAA VYRSDKARKSRRISLEFMLEQANKFANHKAIWFPY NMDWRGRVYAVSMFNPQGYDMTKGLLTLAKGKP IGKEGYWYWLKIHGANCAGVDKVPFPERIKFIEENHE NIMACAKSPLENTWWAEQDSPFCFLAFCFEYAGV QHHGLSYNCSLPLAFDGS CSGIQHFSAMLRDEVG RAVNLLPSETVQDIYGIVAKKVNEILQADAINGTDN EVVTVDENTGEISEKVKLGTKALAGQWLAYGVT RSVTKRSVMTLAYGSKEFGFRQQVLEDTIQPAIDSG KGLMFTQPNQAAGYMAKLIWESVSVTVVAAVEA MNWLKSAKLLAAEVKDKKTGEILRKRCVHWV TPDGFPVWQEYKKPIQTRLNLMFLGQFRLQPTINT NKDSEIDAHKQESGIAPNFVHSQDGSHLRKTVVWA HEKYGIESFALIHDSFGTIPADAANLFKAVRETMVD	6

[0140]

RNA 聚合 酶变体	氨基酸序列	SEQ ID NO
	TYESCDVLADFYDQFADQLHESQLDKMPALPAKG NLNLRDILESDFafa	
E350W D351V N437Y	MNTINIAKNDFSDIELAAIPFNTLADHYGERLAREQ LALEHESYEMGEARFRKMFERQLKAGEVADNAAA KPLITTLLPKMIARINDWFEEVKAKRGKRPTAFQFL QEIKPEAVAYITIKTTLACLTSADNTTVQAVASAIG RAIEDEARFGRIRDLEAKHFKNVEEQLNKRVGHV YKKAQFMQVVEADMLSKGLLGGEAWSSWHKEDSI HVGVRCEIEMLIESTGMVSLHRQNAGVVGGQDSETIE LAPEYAEAIATRAGALAGISPMFQPCVVPPKPWTGI TGGGYWANGRRPLALVRTHSKKALMRYEDVYMP EVYKAINIAQNTAWKINKKVLAVANVITKWKHCP VWVIPAIEREELPMKPEDIDMNPEALTAWKRAAAA VYRKDKARKSRRISLEFMLEQANKFANHKAIWFPY NMDWRGRVYAVSMFNPQGYDMTKGLLTLAKGKP IGKEGYWWLKIHGANCAGVDKVPFPERIKFIEENHE NIMACAKSPLENTWWAEQDSPFCFLAFCFEYAGV QHHGLSYNCSLPLAFDGCSCGIQHFSAMLRDEVGG RAVNLLPSETVQDIYGIVAKKVNEILQADAINGTDN EVVTVDENTGEISEKVKLGTKALAGQWLAYGVT RSVTKRSVMTLAYGSKEFGFRQQVLEDTIQPAIDSG KGLMFTQPNQAAGYMAKLIWESVSVTVVAAVEA MNWLKSAAKLLAAEVKDKKTGEILRKRCVHWV TPDGFVWQYKKPIQTRLNLMFLGQFRLQPTINT NKDSEIDAHKQESGIAPNFVHSQDGSHLRKTVVWA HEKYGIESFALIHDSFGTIPADAANLFKAVRETMDV TYESCDVLADFYDQFADQLHESQLDKMPALPAKG NLNLRDILESDFafa	7
E350W D351V K387S	MNTINIAKNDFSDIELAAIPFNTLADHYGERLAREQ LALEHESYEMGEARFRKMFERQLKAGEVADNAAA KPLITTLLPKMIARINDWFEEVKAKRGKRPTAFQFL QEIKPEAVAYITIKTTLACLTSADNTTVQAVASAIG RAIEDEARFGRIRDLEAKHFKNVEEQLNKRVGHV YKKAQFMQVVEADMLSKGLLGGEAWSSWHKEDSI HVGVRCEIEMLIESTGMVSLHRQNAGVVGGQDSETIE LAPEYAEAIATRAGALAGISPMFQPCVVPPKPWTGI TGGGYWANGRRPLALVRTHSKKALMRYEDVYMP EVYKAINIAQNTAWKINKKVLAVANVITKWKHCP VWVIPAIEREELPMKPEDIDMNPEALTAWKRAAAA VYRSDKARKSRRISLEFMLEQANKFANHKAIWFPY NMDWRGRVYAVSMFNPQGNDMTKGLLTLAKGKP IGKEGYWWLKIHGANCAGVDKVPFPERIKFIEENHE NIMACAKSPLENTWWAEQDSPFCFLAFCFEYAGV QHHGLSYNCSLPLAFDGCSCGIQHFSAMLRDEVGG RAVNLLPSETVQDIYGIVAKKVNEILQADAINGTDN EVVTVDENTGEISEKVKLGTKALAGQWLAYGVT RSVTKRSVMTLAYGSKEFGFRQQVLEDTIQPAIDSG KGLMFTQPNQAAGYMAKLIWESVSVTVVAAVEA	8

[0141]

RNA 聚合酶变体	氨基酸序列	SEQ ID NO
	MNWLKSAAKLLAAEVKDKKTGEILRKRCVHWV TPDGFPVWQEYKKPIQTRLNLMFLGQFRLQPTINT NKDSEIDAHKQESGIAPNFVHSQDGSHLRKTVVWA HEKYGIESFALIHDSFGTIPADAANLFKAVRETMVD TYESCDVLADFYDQFADQLHESQLDKMPALPAKG NLNLRDILESDFafa	
E350W D351V D653W	MNTINIAKNDFSDIELAAIPFNTLADHYGERLAREQ LALEHESYEMGEARFRKMFERQLKAGEVADNAAA KPLITTLLPKMIARINDWFEEVKAARGKRPTAFQFL QEIKPEAVAYITIKTTLACLTSADNTTVQAVASAIG RAIEDEARFGRIRDLEAKHFKNVEEQLNKRVGHV YKKAQMVFVEADMLSKGLLGGEAWSSWHKEDSI HVGVRCEMLIESTGMVSLHRQNAGVVGQDSETIE LAPEYAEAIATRAGALAGISPMFQPCVPPKPWTGI TGGGYWANGRRPLALVRTHSKKALMRYEDVYMP EVYKAINIAQNTAWKINKKVLAVANVITKWKHCP VWVIPAIEREELPMKPEDIDMNPEALTAWKRAAAA VYRKDKARKSRRISLEFMLEQANKFANHKAIWFPY NMDWRGRVYAVSMFNPQGNDMTKGLLTLAKGKP IGKEGYYYWLKIHGANCAGVDKVPFPERIKFIEENHE NIMACAKSPLENTWWAEQDSPFCFLAFCFEYAGV QHHGLSYNCSLPLAFDGSCSGIQHFSAMLRDEVGG RAVNLLPSETVQDIYGIVAKKVNEILQADAINGTDN EVVTVTDENTGEISEKVKLGTKALAGQWLAYGVT RSVTKRSVMTLAYGSKEFGFRQQVLEWTIQPAIDS GKGLMFTQPNQAAGYMAKLIWESVSVTVVAAVE AMNWLKSAAKLLAAEVKDKKTGEILRKRCVHWV VTPDGFPVWQEYKKPIQTRLNLMFLGQFRLQPTIN TNKDSEIDAHKQESGIAPNFVHSQDGSHLRKTVVW AHEKYGIESFALIHDSFGTIPADAANLFKAVRETMV DTYESCDVLADFYDQFADQLHESQLDKMPALPAK GNLNLRDILESDFafa	9

[0142] 在一些实施方案中, RNA聚合酶变体进一步包含一个或多个纯化标签。举例来说, RNA聚合酶变体可包含组氨酸纯化标签(例如, -HHHHHH-的氨基酸序列(SEQ ID NO:14))或可用于纯化的任何其他氨基酸序列。组氨酸纯化标签或类似带电荷的氨基酸序列能够结合Ni²⁺树脂。在一些实施方案中, 组氨酸纯化标签包含-HHHHHHV-的氨基酸序列(SEQ ID NO:15)。在一些实施方案中, 纯化标签是共价附接至RNA聚合酶变体的N末端的N末端纯化标签。在一些实施方案中, 纯化标签是共价附接至RNA聚合酶变体的C末端的C末端纯化标签。在一些实施方案中, 蛋白质纯化标签是FLAG标签(例如, -DYKDDDK-的氨基酸序列(SEQ ID NO:16))或血凝素标签。在一些实施方案中, 包含N末端His标签的RNA聚合酶变体包含SEQ ID NO:10-13中的任一个。

[0143] 表2. 包含N末端His标签的RNA聚合酶变体

[0144]

RNA 聚合酶变体	氨基酸序列	SEQ ID NO
E350W D351V K387S N437Y	MHHHHHHVNSNTINIAKNDFSDIELAAIPFNTLADHY GERLAREQLALEHESYEMGEARFRKMFERQLKAGE VADNAAAKPLITTLLPKMIARINDWFEEVKAKRGKR PTAFQFLQEIKPEAVAYITIKTTLACLTSADNTTVQAV ASAIGRAIEDEARFGRIRDLEAKHFKNVEEQLNKR GHVYKKAFMQVVEADMLSKGLLGGEAWSSWHKED SIHVGVRCEIEMLIESTGMVSLHRQNAGVVGQDSETIE LAPEYAEAIATRAGALAGISPMFQPCVVPPKPWTGIT GGGYWANGRRPLALVRTHSKKALMRYEDVYMPEV YKAINIAQNTAWKINKKVLAVANVITKWKHCPVWV IPAIEREELPMKPEDIDMNPEALTAWKRAAAAVYRS DKARKSRRISLEFMLEQANKFANHKAIWFPYNMDW RGRVYAVSMFNPQGYDMTKGLLTLAKGKPIGKEGY YWLKIHGANCAGVDKVPFPERIKFIEENHENIMACAK SPLENTWWAEQDSPFCFLAFCFEYAGVQHHGLSYNC SLPLAFDGSCSGIQHFSAMLRDEVGGRAVNLLPSETV QDIYGIVAKKVNEILQADAINGTDNEVVTVTDENTGE ISEKVKLGTKALAGQWLAYGVTRSVTKRSVMTLAY GSKEFGFRQQVLEDTIQPAIDSGKGLMFTQPNQAAG YMAKLIWESVSVTVVAAVEAMNWLKSAKLLAAE VKDKKTGEILRKRCVHWVTPDGFPVWQEYKKPIQT RLNLMFLGQFRLQPTINTNKDSEIDAHKQESGIAPNF VHSQDGSHLRKTVVWAHEKYGIESFALIHDSFGTIPA DAANLFKAVRETMDVDTYESCDVLADFYDQFADQLH ESQLDKMPALPAKGNLNLRDILESDFAF	10
N437Y E350W D351V	MHHHHHHVNSNTINIAKNDFSDIELAAIPFNTLADHY GERLAREQLALEHESYEMGEARFRKMFERQLKAGE VADNAAAKPLITTLLPKMIARINDWFEEVKAKRGKR PTAFQFLQEIKPEAVAYITIKTTLACLTSADNTTVQAV ASAIGRAIEDEARFGRIRDLEAKHFKNVEEQLNKR GHVYKKAFMQVVEADMLSKGLLGGEAWSSWHKED SIHVGVRCEIEMLIESTGMVSLHRQNAGVVGQDSETIE LAPEYAEAIATRAGALAGISPMFQPCVVPPKPWTGIT GGGYWANGRRPLALVRTHSKKALMRYEDVYMPEV YKAINIAQNTAWKINKKVLAVANVITKWKHCPVWV IPAIEREELPMKPEDIDMNPEALTAWKRAAAAVYRK DKARKSRRISLEFMLEQANKFANHKAIWFPYNMDW RGRVYAVSMFNPQGYDMTKGLLTLAKGKPIGKEGY YWLKIHGANCAGVDKVPFPERIKFIEENHENIMACAK	11

RNA 聚合酶变体	氨基酸序列	SEQ ID NO
	SPLENTWWAEQDSPFCFLAFCFEYAGVQHHGLSYNC SLPLAFDGS CSGIQHFSAMLRDEVGGRAVNLLPSETV QDIYGIVAKKVNEILQADAINGTDNEVVTVTDENTGE ISEKVKLGTKALAGQWLAYGVTRSVTKRSVMTLAY GSKEFGFRQQVLEDTIQPAIDSGKGLMFTQPNQAAG YMAKLIWESVSVTVVAAVEAMNWLKSAAKLLAAE VKDKKTGEILRKRCVHWVTPDGFPVWQEYKKPIQT RLNLMFLGQFRLQPTINTNKDSEIDAHKQESGIAPNF VHSQDGSHLRKTVVWAHEKYGIESFALIHDSFGTIPA DAANLFKAVRETMVDTYESCDVLADFYDQFADQLH ESQLDKMPALPAKGNNLRDILESDFafa	
K387S E350W D351V	MHHHHHHVNSNTINIAKNDFSDIELAAIPFNTLADHY GERLAREQLALEHESYEMGEARFRKMFERQLKAGE VADNAAAKPLITTL PKMIARINDWFEEVKA KRGR PTAFLQFLQEIKPEAVAYITIKTTLACLTSADNTTVQAV ASAIGRAIEDEARFGRIRDLEAKHFKNVEEQLNKR GHVYKKA FMQVVEADMLSKGLLGGEAWSSWHKED SIHVGVRCEMLIESTGMVSLHRQNAGVVGQDSETIE LAPEYAEAIATRAGALAGISPMFQPCVPPKPWTGIT GGGYWANGRRPLALVRTHSKKALMRYEDVYMPEV YKAINIAQNTAWKINKKVLAVANVITKWKHCPVWV IPAIEREELPMKPEDIDMNPEALTAWKRAAAAVYRS DKARKSRRISLEFMLEQANKFANHKAIWFPYNMDW RGRVYAVSMFNPQGNMTKGLLTLAKGKPIGKEGY YWLKIHGANCAGVDKVPFPERIKFIEENHENIMACAK SPLENTWWAEQDSPFCFLAFCFEYAGVQHHGLSYNC SLPLAFDGS CSGIQHFSAMLRDEVGGRAVNLLPSETV QDIYGIVAKKVNEILQADAINGTDNEVVTVTDENTGE ISEKVKLGTKALAGQWLAYGVTRSVTKRSVMTLAY GSKEFGFRQQVLEDTIQPAIDSGKGLMFTQPNQAAG YMAKLIWESVSVTVVAAVEAMNWLKSAAKLLAAE VKDKKTGEILRKRCVHWVTPDGFPVWQEYKKPIQT RLNLMFLGQFRLQPTINTNKDSEIDAHKQESGIAPNF VHSQDGSHLRKTVVWAHEKYGIESFALIHDSFGTIPA DAANLFKAVRETMVDTYESCDVLADFYDQFADQLH ESQLDKMPALPAKGNNLRDILESDFafa	12
D653W E350W D351V	MHHHHHHVNSNTINIAKNDFSDIELAAIPFNTLADHY GERLAREQLALEHESYEMGEARFRKMFERQLKAGE VADNAAAKPLITTL PKMIARINDWFEEVKA KRGR PTAFLQFLQEIKPEAVAYITIKTTLACLTSADNTTVQAV ASAIGRAIEDEARFGRIRDLEAKHFKNVEEQLNKR GHVYKKA FMQVVEADMLSKGLLGGEAWSSWHKED SIHVGVRCEMLIESTGMVSLHRQNAGVVGQDSETIE LAPEYAEAIATRAGALAGISPMFQPCVPPKPWTGIT GGGYWANGRRPLALVRTHSKKALMRYEDVYMPEV YKAINIAQNTAWKINKKVLAVANVITKWKHCPVWV IPAIEREELPMKPEDIDMNPEALTAWKRAAAAVYRK DKARKSRRISLEFMLEQANKFANHKAIWFPYNMDW	13

[0145]

RNA 聚合酶变体	氨基酸序列	SEQ ID NO
[0146]	RGRVYAVSMFNPQGNDMTKGLLTLAKGKPIGKEGY YWLKIHGANCAGVDKVPFPERIKFIEENHENIMACAK SPLENTWWAEQDSPFCFLAFCFEYAGVQHHGLSYNC SLPLAFDGCSCGSIQHFSAMLRDEVGGRAVNLLPSETV QDIYGIVAKKVNEILQADAINGTDNEVVTVTDENTGE ISEKVKLGTKALAGQWLAYGVTRSVTKRSVMTLAY GSKEFGFRQQVLEWTIQPAIDSGKGLMFTQPNQAAG YMAKLIWESVSVTVVAAVEAMNWLKSAAKLLAAE VKDKKTGEILRKRCVHWVTPDGFVWQEYKKPIQT RLNLMFLGQFRLQPTINTNKDSEIDAHKQESGIAPNF VHSQDGSHLRKTVVWAHEKYGIESFALIHDSFGTIPA DAANLFKAVRETMVDTYESCDVLADFYDQFADQLH ESQLDKMPALPAKGNLNLRDILESDFafa	

[0147] 在一些实施方案中, RNA聚合酶变体与包含SEQ ID NO:2-13中任一个的氨基酸序列的RNA聚合酶具有至少90%、至少91%、至少92%、至少93%、至少94%、至少95%、至少96%、至少97%、至少98%或至少99%同一性。在一些实施方案中, RNA聚合酶变体可与包含SEQ ID NO:1的氨基酸序列的RNA聚合酶共享至少90%、至少91%、至少92%、至少93%、至少94%、至少95%同一性。

[0148] 术语“同一性”是指两个或更多个多肽(例如酶)或多核苷酸(核酸)的序列之间的关系, 如通过比较序列所确定的。同一性还是指序列之间的序列相关性程度, 如通过两个或更多个氨基酸残基或核酸残基的串之间的匹配数量所确定的。同一性测量由特定数学模型或计算机程序(例如, “算法”)处理的具有缺口比对(如果有的话)的两个或更多个序列中较小序列之间的同一性匹配的百分比。相关蛋白质或核酸的同一性可以通过已知方法容易地计算。当应用于多肽或多核苷酸序列时, “百分比(%)同一性”被定义为在比对序列并引入缺口(如果需要)以实现最大百分比同一性之后, 候选氨基酸或核酸序列中与第二序列的氨基酸序列或核酸序列中的残基同一的残基(氨基酸残基或核酸残基)百分比。用于比对的方法和计算机程序是本领域熟知的。可以理解, 同一性取决于百分比同一性的计算, 但由于计算中引入的缺口和罚分, 其值可能会有所不同。通常, 特定多核苷酸或多肽(例如抗原)的变体与特定参考多核苷酸或多肽具有至少40%、45%、50%、55%、60%、65%、70%、75%、80%、85%、90%、91%、92%、93%、94%、95%、96%、97%、98%、99%但小于100%序列同一性, 如通过本文所述的本领域技术人员已知的序列比对程序和参数确定的。用于比对的此类工具包括BLAST套件的那些工具(Stephen F. Altschul, 等人(1997), “Gapped BLAST and PSI-BLAST: a new generation of protein database search programs”, *Nucleic Acids Res.* 25:3389-3402)。另一种流行的局部比对技术是基于Smith-Waterman算法(Smith, T.F. 和Waterman, M.S. (1981) “Identification of common molecular subsequences.” *J. Mol. Biol.* 147:195-197)。基于动态编程的通用全局比对技术是Needleman-wunsch算法(Needleman, S.B. & Wunsch, C.D. (1970) “A general method applicable to the search for similarities in the amino acid sequences of two proteins.” *J. Mol. Biol.* 48:443-453)。最近, 已开发了快速最佳全局序列比对算法(FOGSAA), 据称所述算法比其他最佳全局比对方法(包括Needleman-Wunsch算法)更快地产

生核苷酸和蛋白质序列的全局比对。

[0149] RNA加帽

[0150] 多价RNA组合物可包含具有编码蛋白质或肽的开放阅读框的一种或多种mRNA。这些mRNA中的每一者都可具有5'帽。可在共IVT反应(例如,转录共加帽)期间或在IVT反应之后添加5'帽。

[0151] 一些方面还包括包含5'帽和多核苷酸(例如,包含编码待表达的多肽的核苷酸序列的多核苷酸)两者的多核苷酸。

[0152] 天然mRNA的5'帽结构参与核输出,增加mRNA稳定性并结合mRNA帽结合蛋白(CBP),例如eIF4E,其通过CBP与多聚(A)结合蛋白的缔合而负责细胞中的mRNA稳定性和翻译能力以形成成熟环状mRNA种类。帽进一步有助于在mRNA剪接期间去除5'近端内含子。

[0153] 内源mRNA分子可以在5'端加帽,从而在mRNA分子的末端鸟苷帽残基与5'-末端转录的有义核苷酸之间产生5'-ppp-5'-三磷酸键。然后可以将这种5'-鸟苷酸帽甲基化以产生N7-甲基-鸟苷酸残基。mRNA的5'端的末端和/或末端前转录核苷酸的核糖也可以任选地被2'-O-甲基化。通过鸟苷酸帽结构的水解和裂解使5'-脱帽可以靶向核酸分子,诸如mRNA分子,以进行降解。

[0154] 在一些实施方案中,将多核苷酸(例如,包含编码多肽的核苷酸序列的多核苷酸)并入帽部分。

[0155] 在一些实施方案中,多核苷酸包含防止脱帽并因此增加mRNA半衰期的不可水解帽结构。因为帽结构水解需要5'-ppp-5'磷酸二酯键的裂解,所以可以在加帽反应期间使用修饰的核苷酸。举例来说,来自New England Biolabs(Ipswich,MA)的牛痘加帽酶可以根据制造商的说明书与 α -硫代-鸟苷核苷酸一起使用,以在5'-ppp-5'帽中产生硫代磷酸酯键。可以使用额外修饰的鸟苷核苷酸,诸如 α -甲基-磷酸和硒代-磷酸核苷酸。

[0156] 额外的修饰包括但不限于多核苷酸(如上文所提及)的5'-末端和/或5'-末端前核苷酸的核糖在糖环的2'-羟基上的2'-O-甲基化。多个不同的5'-帽结构可以用于产生核酸分子的5'-帽,诸如用作mRNA分子的多核苷酸。帽类似物,在本文中也称为合成帽类似物、化学帽、化学帽类似物或结构或功能性帽类似物,在其化学结构上与天然(即,内源、野生型或生理)5'-帽不同,同时保留帽功能。帽类似物可以是化学(即,非酶促)或酶促合成的和/或连接至多核苷酸的。

[0157] 举例来说,抗反向帽类似物(ARCA)帽含有由5'-5'-三磷酸基团连接的两个鸟嘌呤,其中一个鸟嘌呤含有N7甲基以及3'-O-甲基(即,N7,3'-O-二甲基-鸟苷-5'-三磷酸-5'-鸟苷($m^7G-3'mppp-G$;其可等效地命名为3'-O-Me- $m^7G(5')ppp(5')G$)。另一个未修饰的鸟嘌呤的3'-O原子变成连接至加帽多核苷酸的5'-末端核苷酸。N7-和3'-O-甲基化鸟嘌呤提供了加帽多核苷酸的末端部分。

[0158] 另一个示例性帽是mCAP,它类似于ARCA,但在鸟苷上具有2'-O-甲基(即,N7,2'-O-二甲基-鸟苷-5'-三磷酸-5'-鸟苷, $m^7Gm-ppp-G$)。

[0159] 另一个示例性帽是 $m^7G-ppp-Gm-AG$ (即,N7,鸟苷-5'-三磷酸-2'-O-二甲基-鸟苷-腺苷-鸟苷)。

[0160] 在一些实施方案中,帽是二核苷酸帽类似物。作为一个非限制性实例,二核苷酸帽类似物可以在不同的磷酸位置用硼代磷酸基团或硒代磷酸基团进行修饰,诸如美国专利号

US 8,519,110中所描述的二核苷酸帽类似物,所述专利的内容以全文引用的方式并入本文中。

[0161] 在另一个实施方案中,帽是帽类似物,是本领域中已知和/或本文所描述的帽类似物的N7-(4-氯苯氧基乙基)取代的二核苷酸形式。帽类似物的N7-(4-氯苯氧基乙基)取代的二核苷酸形式的非限制性实例包括N7-(4-氯苯氧基乙基)-G(5')ppp(5')G和N7-(4-氯苯氧基乙基)-m^{3',-0}G(5')ppp(5')G帽类似物(参见例如,Kore等人*Bioorganic&Medicinal Chemistry* 2013 21:4570-4574中所述的各种帽类似物和合成帽类似物的方法;所述文献的内容以全文引用的方式并入本文中)。在另一个实施方案中,帽类似物是4-氯/溴苯氧基乙基类似物。

[0162] 多核苷酸也可以在制造后(无论是IVT还是化学合成)使用酶加帽,以产生更真实的5'-帽结构。如本文所用,短语“更真实”是指在结构上或功能上密切反映或模仿内源或野生型特征的特征。也就是说,“更真实”的特征与现有技术的合成特征或类似物等相比更能代表内源、野生型、天然或生理细胞功能和/或结构,或在一个或多个方面优于对应的内源、野生型、天然或生理特征。更真实的5'帽结构的非限制性实例是那些帽结构:除了其他之外,它们与本领域中已知的合成5'帽结构(或野生型、天然或生理5'帽结构)相比具有增强的帽结合蛋白结合、增加的半衰期、减少的对5'核酸内切酶的敏感性和/或减少的5'脱帽。举例来说,重组牛痘病毒加帽酶和重组2'-O-甲基转移酶可以在多核苷酸的5'-末端核苷酸与鸟嘌呤帽核苷酸之间产生规范的5'-5'-三磷酸键,其中帽鸟嘌呤含有N7甲基化并且mRNA的5'-末端核苷酸含有2'-O-甲基。此种结构称为帽1结构。与例如本领域中已知的其他5'帽类似物结构相比,此帽导致更高的翻译能力和细胞稳定性以及细胞促炎细胞因子的活化减少。帽结构包括但不限于7mG(5')ppp(5')N, pN2p(帽0)、7mG(5')ppp(5')N1mpNp(帽1)和7mG(5')-ppp(5')N1mpN2mp(帽2)。

[0163] 作为一个非限制性实例,制造后对嵌合多核苷酸加帽可能更有效,因为几乎100%的嵌合多核苷酸可以被加帽。这与当在体外转录反应过程中帽类似物连接至嵌合多核苷酸时约80%成对比。

[0164] 在一些实施方案中,5'末端帽可包括内源帽或帽类似物。在一些实施方案中,5'末端帽可包括鸟嘌呤类似物。可用的鸟嘌呤类似物包括但不限于肌苷、N1-甲基-鸟苷、2'氟-鸟苷、7-脱氮-鸟苷、8-氧代-鸟苷、2-氨基-鸟苷、LNA-鸟苷和2-叠氮基-鸟苷。

[0165] 还描述了示例性帽,所述帽包括可以用于核糖核酸(RNA)合成的共转录加帽方法中的那些帽,所述合成使用RNA聚合酶,例如野生型RNA聚合酶或其变体,例如像所描述的那些变体。在一个实施方案中,当在“一锅(one-pot)”反应中产生RNA时可以添加帽,而不需要单独的加帽反应。因此,在一些实施方案中,所述方法包括在体外转录反应条件下使多核苷酸模板与RNA聚合酶变体、核苷三磷酸和帽类似物反应以产生RNA转录物。

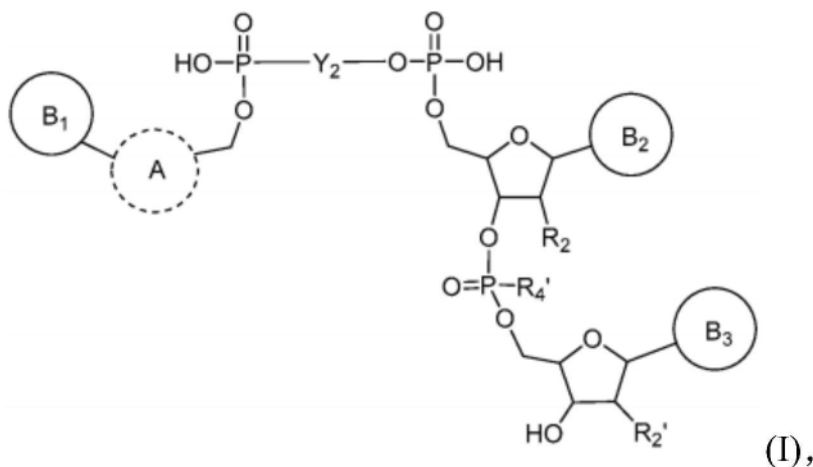
[0166] 在一些实施方案中,帽类似物结合包含启动子区的多核苷酸模板,所述启动子区包含具有在核苷酸位置+1处的第一核苷酸、在核苷酸位置+2处的第二核苷酸和在核苷酸位置+3处的第三核苷酸的转录起始位点。在一些实施方案中,帽类似物至少在多核苷酸位置+1处,诸如在+1和+2位置处或在+1、+2和+3位置处与多核苷酸模板杂交。

[0167] 举例来说,帽类似物可以是二核苷酸帽、三核苷酸帽或四核苷酸帽。在一些实施方案中,帽类似物是二核苷酸帽。在一些实施方案中,帽类似物是三核苷酸帽。在一些实施方

案中,帽类似物是四核苷酸帽。如本文所用,术语“帽”包含倒置G核苷酸并且可以包含倒置G的3'的额外核苷酸,例如,倒置G的3'和5'UTR的5'的1、2或更多个核苷酸。

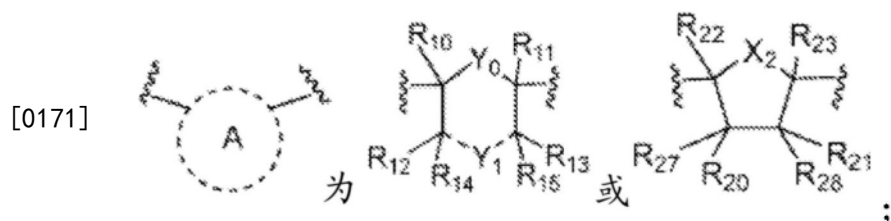
[0168] 示例性帽包含序列GG、GA或GGA,其中帶下划线的斜体G是倒置G。

[0169] 在一些实施方案中,核苷酸帽(例如,三核苷酸帽或四核苷酸帽)包含式(I)



的化合物或其立体异

构体、互变异构体或盐,其中



[0172] 环B₁是修饰或未修饰的鸟嘌呤；

[0173] 环B₁和环B₂各自独立地是核碱基或修饰的核碱基;

[0174] X_2 是 O 、 $S(O)_p$ 、 NR_{24} 或 $CR_{25}R_{26}$,其中 p 是0、1或2;

[0175] Y_0 是 0 或 CR_6R_7 ;

[0176] Y1是O、S(O)_n、CR₆R₇或NR₈,其中n是0、1或2;

[0177] 每个---是单键或不存在,其中当每个---是单键时, Y_i 是O、S(O)_n、 CR_6R_7 或 NR_8 ;并且当每个---不存在时, Y_i 是缺失的;

[0178] Y_2 是 $(OP(O)R_4)_m$,其中m是0、1或2,或 $-O-(CR_{40}R_{41})_u-Q_0-(CR_{42}R_{43})_v-$,其中 Q_0 是键、 O 、 $S(O)_r$ 、 NR_{44} 或 $CR_{45}R_{46}$,r是0、1或2,并且u和v中的每一者独立地是1、2、3或4;

[0179] 每个 R_9 和 R_9' 独立地是卤基、LNA或 OR_3 ;

[0180] 每个R₃独立地是H、C₁-C₆烷基、C₂-C₆烯基或C₂-C₆炔基,并且当R₃是C₁-C₆烷基、C₂-C₆烯基或C₂-C₆炔基时任选地被以下一者或多者取代:卤基、OH和任选地被一个或多个OH或OC(O)-C₁-C₆烷基取代的C₁-C₆烷氧基;

[0181] 每个 R_4 和 R_4' 独立地是H、卤基、 C_1 - C_6 烷基、OH、SH、SeH或 BH_3^- ;

[0182] R_6 、 R_7 和 R_8 中的每一者独立地是 $-Q_1-T_1$,其中 Q_1 是键或任选地被卤基、氰基、OH和 C_1-C_6 烷氧基中的一者或多者取代的 C_1-C_3 烷基连接子,并且 T_1 是H、卤基、OH、COOH、氰基或 R_{s1} ,其中 R_{s1} 是 C_1-C_3 烷基、 C_2-C_6 烯基、 C_2-C_6 炔基、 C_1-C_6 烷氧基、 $C(=O)O-C_1-C_6$ 烷基、 C_3-C_8 环烷基、 C_6-C_{10} 芳基、 $NR_{31}R_{32}$ 、 $(NR_{31}R_{32}R_{33})^+$ 、4元至12元杂环烷基或5元或6元杂芳基,并且 R_{s1} 任选地被一个

或多个选自以下组成的组的取代基取代:卤基、OH、氧代、 C_1-C_6 烷基、 $COOH$ 、 $C(O)O-C_1-C_6$ 烷基、氰基、 C_1-C_6 烷氧基、 $NR_{31}R_{32}$ 、 $(NR_{31}R_{32}R_{33})^+$ 、 C_3-C_8 环烷基、 C_6-C_{10} 芳基、4元至12元杂环烷基和5元或6元杂芳基;

[0183] R_{10} 、 R_{11} 、 R_{12} 、 R_{13} 、 R_{14} 和 R_{15} 中的每一者独立地是 $-Q_2-T_2$,其中 Q_2 是键或视情况被卤基、氰基、OH和 C_1-C_6 烷氧基中的一者或多者取代的 C_1-C_3 烷基连接子,并且 T_2 是H、卤基、OH、 NH_2 、氰基、 NO_2 、 N_3 、 R_{s2} 或 OR_{s2} ,其中 R_{s2} 是 C_1-C_6 烷基、 C_2-C_6 烯基、 C_2-C_6 炔基、 C_3-C_8 环烷基、 C_6-C_{10} 芳基、 $NHC(O)-C_1-C_6$ 烷基、 $NR_{31}R_{32}$ 、 $(NR_{31}R_{32}R_{33})^+$ 、4元至12元杂环烷基或5元或6元杂芳基,并且 R_{s2} 任选地被一个或多个选自以下组成的组的取代基取代:卤基、OH、氧代、 C_1-C_6 烷基、 $COOH$ 、 $C(O)O-C_1-C_6$ 烷基、氰基、 C_1-C_6 烷氧基、 $NR_{31}R_{32}$ 、 $(NR_{31}R_{32}R_{33})^+$ 、 C_3-C_8 环烷基、 C_6-C_{10} 芳基、4元至12元杂环烷基和5元或6元杂芳基;或替代地, R_{12} 与 R_{14} 一起是氧代,或 R_{13} 与 R_{15} 一起是氧代,

[0184] R_{20} 、 R_{21} 、 R_{22} 和 R_{23} 中的每一者独立地是 $-Q_3-T_3$,其中 Q_3 是键或任选地被卤基、氰基、OH和 C_1-C_6 烷氧基中的一者或多者取代的 C_1-C_3 烷基连接子,并且 T_3 是H、卤基、OH、 NH_2 、氰基、 NO_2 、 N_3 、 R_{s3} 或 OR_{s3} ,其中 R_{s3} 是 C_1-C_6 烷基、 C_2-C_6 烯基、 C_2-C_6 炔基、 C_3-C_8 环烷基、 C_6-C_{10} 芳基、 $NHC(O)-C_1-C_6$ 烷基、单 C_1-C_6 烷基氨基、二 C_1-C_6 烷基氨基、4元至12元杂环烷基或5元或6元杂芳基,并且 R_{s3} 任选地被一个或多个选自以下组成的组的取代基取代:卤基、OH、氧代、 C_1-C_6 烷基、 $COOH$ 、 $C(O)O-C_1-C_6$ 烷基、氰基、 C_1-C_6 烷氧基、氨基、单 C_1-C_6 烷基氨基、二 C_1-C_6 烷基氨基、 C_3-C_8 环烷基、 C_6-C_{10} 芳基、4元至12元杂环烷基和5元或6元杂芳基; R_{24} 、 R_{25} 和 R_{26} 中的每一者独立地是H或 C_1-C_6 烷基;

[0185] R_{27} 和 R_{28} 中的每一者独立地是H或 OR_{29} ;或 R_{27} 与 R_{28} 一起形成 $OR_{30}-O$;每个 R_{29} 独立地是H、 C_1-C_6 烷基、 C_2-C_6 烯基或 C_2-C_6 炔基,并且当 R_{29} 是 C_1-C_6 烷基、 C_2-C_6 烯基或 C_2-C_6 炔基时任选地被以下一者或多者取代:卤基、OH和任选地被一个或多个OH或 $OC(O)-C_1-C_6$ 烷基取代的 C_1-C_6 烷氧基;

[0186] R_{30} 是任选地被卤基、OH和 C_1-C_6 烷氧基中的一者或多者取代的 C_1-C_6 亚烷基;

[0187] R_{31} 、 R_{32} 和 R_{33} 中的每一者独立地是H、 C_1-C_6 烷基、 C_3-C_8 环烷基、 C_6-C_{10} 芳基、4元至12元杂环烷基或5元或6元杂芳基;

[0188] R_{40} 、 R_{41} 、 R_{42} 和 R_{43} 中的每一者独立地是H、卤基、OH、氰基、 N_3 、 $OP(O)R_{47}R_{48}$ 或任选地被一个或多个 $OP(O)R_{47}R_{48}$ 取代的 C_1-C_6 烷基,或一个 R_{41} 和一个 R_{43} 与它们所连接的碳原子和 Q_0 一起形成 C_4-C_{10} 环烷基、4元至14元杂环烷基、 C_6-C_{10} 芳基或5元至14元杂芳基,并且所述环烷基、杂环烷基、苯基或5元至6元杂芳基中的每一者任选地被以下一者或多者取代:OH、卤基、氰基、 N_3 、氧代、 $OP(O)R_{47}R_{48}$ 、 C_1-C_6 烷基、 C_1-C_6 卤烷基、 $COOH$ 、 $C(O)O-C_1-C_6$ 烷基、 C_1-C_6 烷氧基、 C_1-C_6 卤烷氧基、氨基、单 C_1-C_6 烷基氨基和二 C_1-C_6 烷基氨基;

[0189] R_{44} 是H、 C_1-C_6 烷基或胺保护基;

[0190] R_{45} 和 R_{46} 中的每一者独立地是H、 $OP(O)R_{47}R_{48}$ 或任选地被一个或多个 $OP(O)R_{47}R_{48}$ 取代的 C_1-C_6 烷基,并且

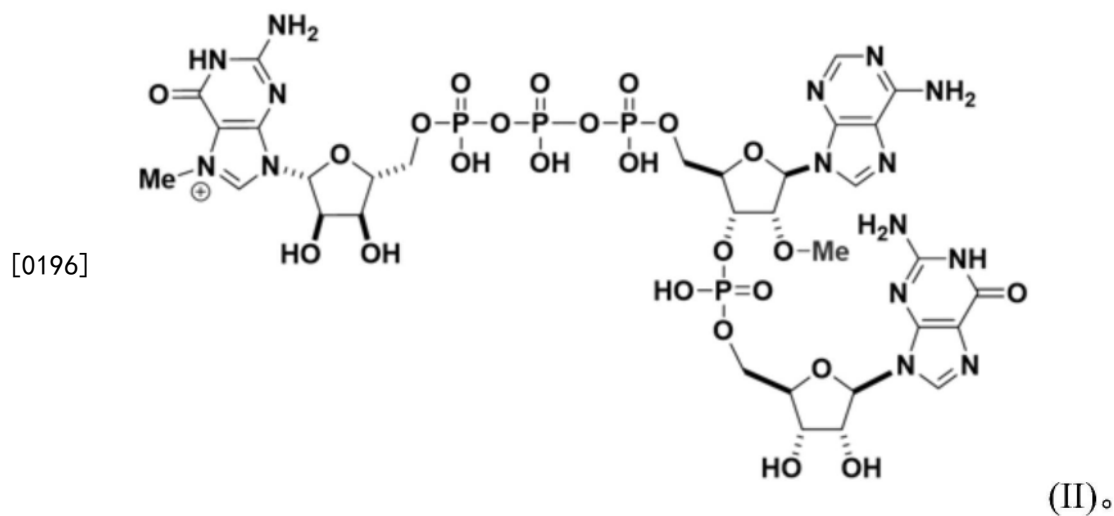
[0191] R_{47} 和 R_{48} 中的每一者独立地是H、卤基、 C_1-C_6 烷基、OH、SH、SeH或 BH_3^- 。

[0192] 应理解,帽类似物可以包括2017年4月20日公布的国际公布WO 2017/066797中所述的帽类似物中的任一者,所述公布以全文引用的方式并入本文中。

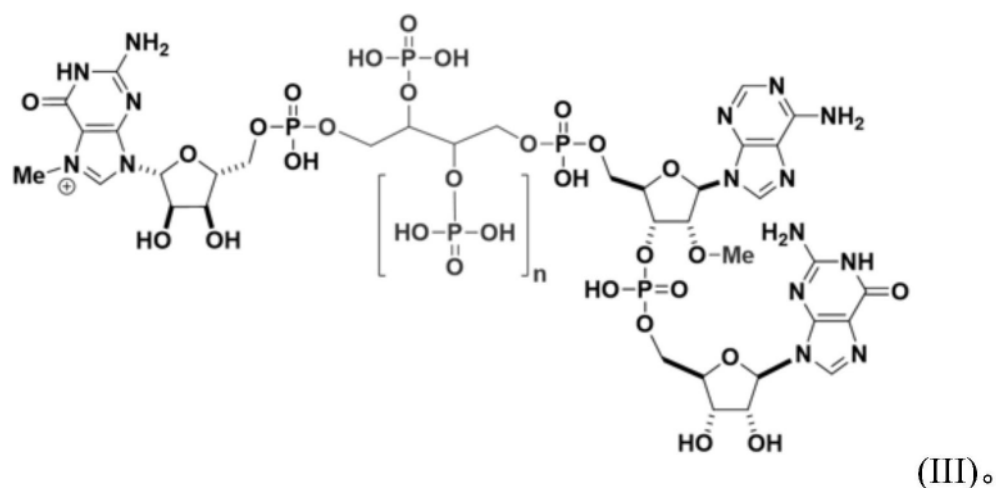
[0193] 在一些实施方案中, B_2 中间位置可以是非核糖分子,诸如阿拉伯糖。

[0194] 在一些实施方案中, R_2 是基于乙基的。

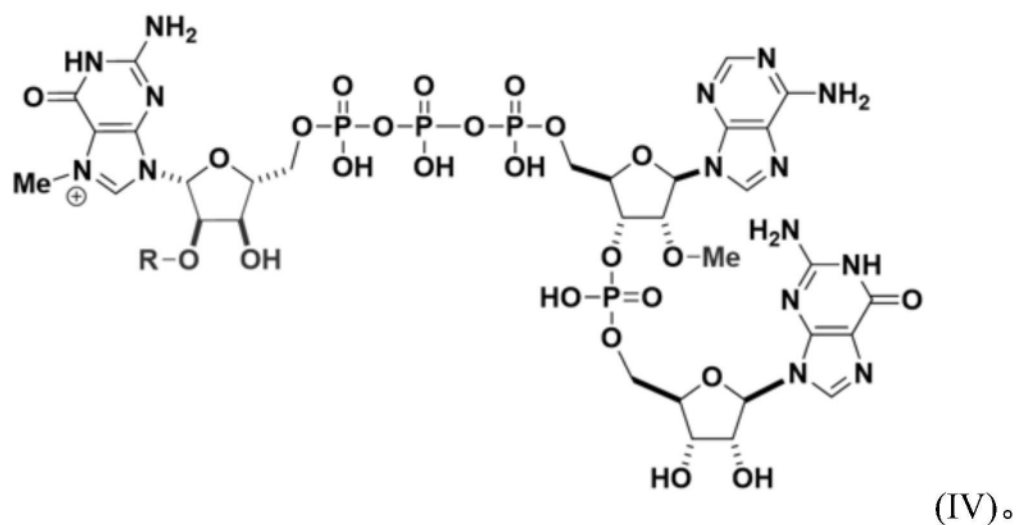
[0195] 因此, 在一些实施方案中, 三核苷酸帽包含以下结构:



[0197] 在其他实施方案中, 三核苷酸帽包含以下结构:

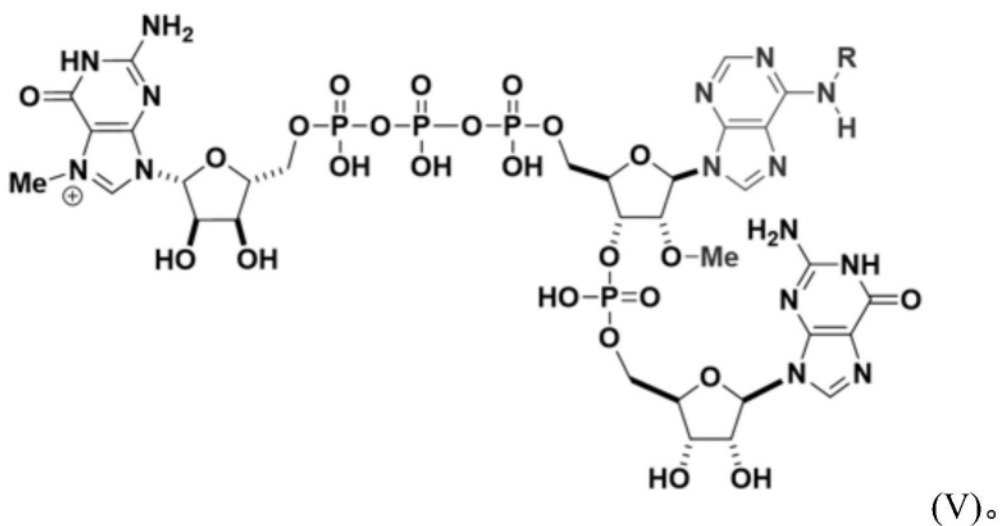


[0198] 在其他实施方案中, 三核苷酸帽包含以下结构:

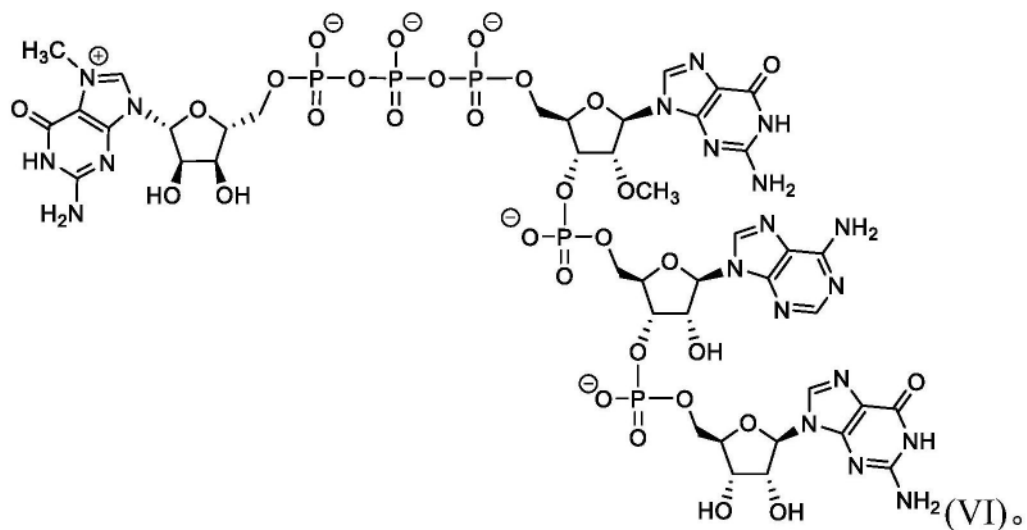


[0199] 在其他实施方案中, 三核苷酸帽包含以下结构:

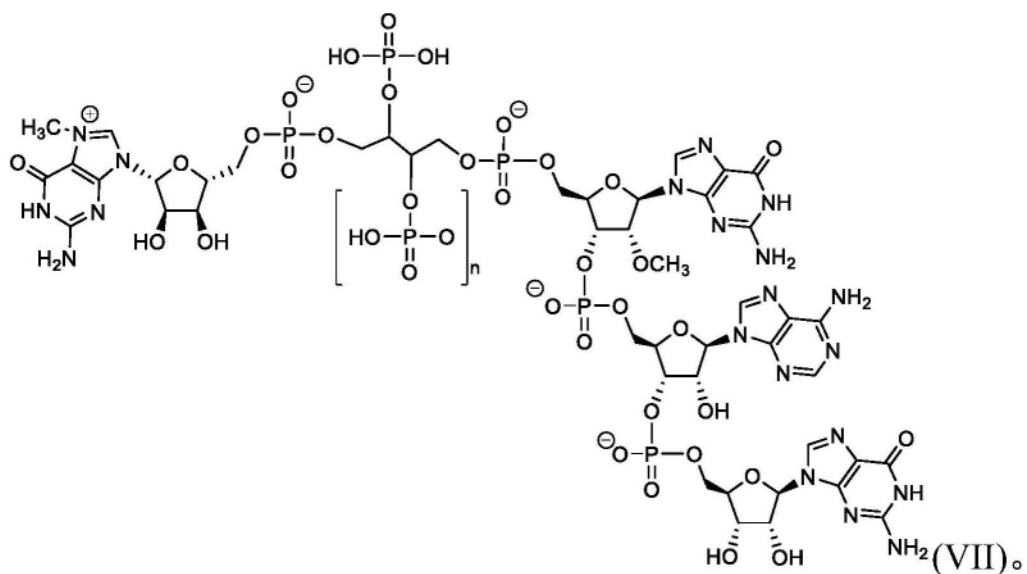
[0200]



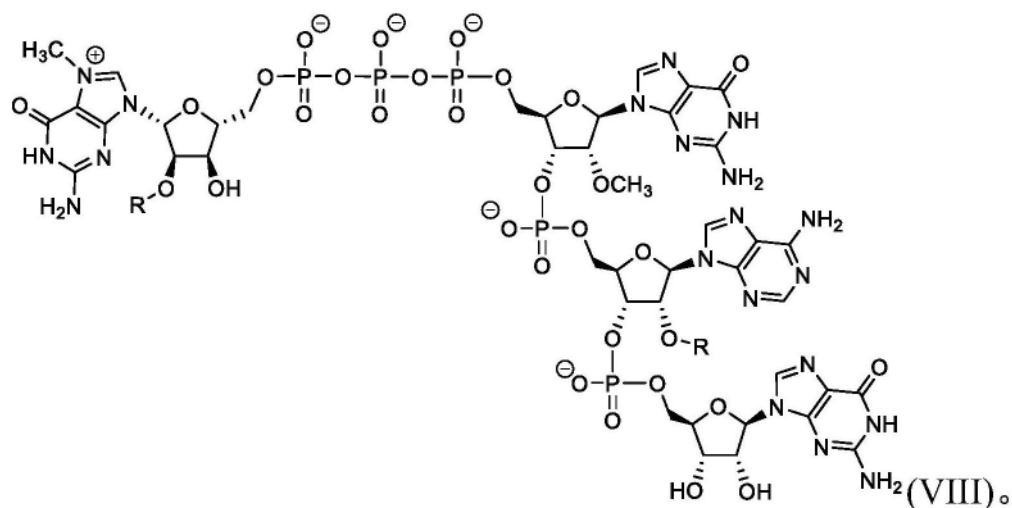
[0201] 因此，在一些实施方案中，四核苷酸帽包含以下结构：



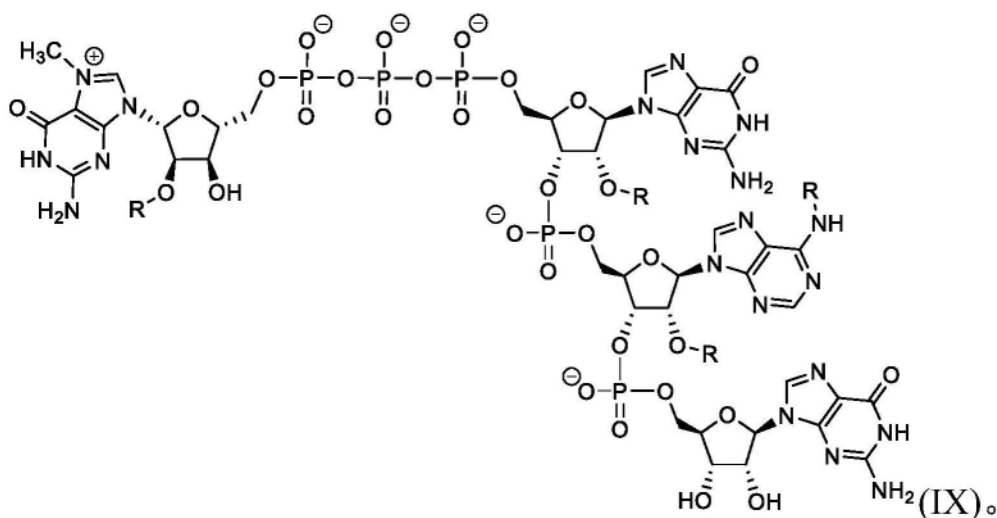
[0202] 在其他实施方案中，四核苷酸帽包含以下结构：



[0203] 在其他实施方案中，四核苷酸帽包含以下结构：



[0204] 在其他实施方案中,四核苷酸帽包含以下结构:



[0206] 在一些实施方案中,R是烷基(例如, C_1 - C_6 烷基)。在一些实施方案中,R是甲基(例如, C_1 烷基)。在一些实施方案中,R是乙基(例如, C_2 烷基)。在一些实施方案中,R是氢。

[0207] 在一些实施方案中,三核苷酸帽包含选自以下序列的序列:GAA、GAC、GAG、GAU、GCA、GCC、GCG、GCU、GGA、GGC、GGG、GGU、GUA、GUC、GUG和GUU。在一些实施方案中,三核苷酸帽包含GAA。在一些实施方案中,三核苷酸帽包含GAC。在一些实施方案中,三核苷酸帽包含GAG。在一些实施方案中,三核苷酸帽包含GAU。在一些实施方案中,三核苷酸帽包含GCA。在一些实施方案中,三核苷酸帽包含GCC。在一些实施方案中,三核苷酸帽包含GCG。在一些实施方案中,三核苷酸帽包含GCU。在一些实施方案中,三核苷酸帽包含GGA。在一些实施方案中,三核苷酸帽包含GGC。在一些实施方案中,三核苷酸帽包含GGG。在一些实施方案中,三核苷酸帽包含GGU。在一些实施方案中,三核苷酸帽包含GUA。在一些实施方案中,三核苷酸帽包含GUC。在一些实施方案中,三核苷酸帽包含GUG。在一些实施方案中,三核苷酸帽包含GUU。

[0208] 在一些实施方案中,三核苷酸帽包含选自以下序列的序列: m^7 GpppApA、 m^7 GpppApC、 m^7 GpppApG、 m^7 GpppApU、 m^7 GpppCpA、 m^7 GpppCpC、 m^7 GpppCpG、 m^7 GpppCpU、 m^7 GpppGpA、 m^7 GpppGpC、 m^7 GpppGpG、 m^7 GpppGpU、 m^7 GpppUpA、 m^7 GpppUpC、 m^7 GpppUpG和 m^7 GpppUpU。

[0209] 在一些实施方案中,三核苷酸帽包含 m^7 GpppApA。在一些实施方案中,三核苷酸帽

包含 $m^7GpppApC$ 。在一些实施方案中,三核苷酸帽包含 $m^7GpppApG$ 。在一些实施方案中,三核苷酸帽包含 $m^7GpppApU$ 。在一些实施方案中,三核苷酸帽包含 $m^7GpppCpA$ 。在一些实施方案中,三核苷酸帽包含 $m^7GpppCpC$ 。在一些实施方案中,三核苷酸帽包含 $m^7GpppCpG$ 。在一些实施方案中,三核苷酸帽包含 $m^7GpppCpU$ 。在一些实施方案中,三核苷酸帽包含 $m^7GpppGpA$ 。在一些实施方案中,三核苷酸帽包含 $m^7GpppGpC$ 。在一些实施方案中,三核苷酸帽包含 $m^7GpppGpG$ 。在一些实施方案中,三核苷酸帽包含 $m^7GpppGpU$ 。在一些实施方案中,三核苷酸帽包含 $m^7GpppUpA$ 。在一些实施方案中,三核苷酸帽包含 $m^7GpppUpC$ 。在一些实施方案中,三核苷酸帽包含 $m^7GpppUpG$ 。在一些实施方案中,三核苷酸帽包含 $m^7GpppUpU$ 。

[0210] 在一些实施方案中,三核苷酸帽包含选自以下序列的序列: m^7G_3, OMe pppApA、 m^7G_3, OMe pppApC、 m^7G_3, OMe pppApG、 m^7G_3, OMe pppApU、 m^7G_3, OMe pppCpA、 m^7G_3, OMe pppCpC、 m^7G_3, OMe pppCpG、 m^7G_3, OMe pppCpU、 m^7G_3, OMe pppGpA、 m^7G_3, OMe pppGpC、 m^7G_3, OMe pppGpG、 m^7G_3, OMe pppGpU、 m^7G_3, OMe pppUpA、 m^7G_3, OMe pppUpC、 m^7G_3, OMe pppUpG和 m^7G_3, OMe pppUpU。

[0211] 在一些实施方案中,三核苷酸帽包含 $m^7G_3,_{OMe}$ pppApA。在一些实施方案中,三核苷酸帽包含 $m^7G_3,_{OMe}$ pppApC。在一些实施方案中,三核苷酸帽包含 $m^7G_3,_{OMe}$ pppApG。在一些实施方案中,三核苷酸帽包含 $m^7G_3,_{OMe}$ pppApU。在一些实施方案中,三核苷酸帽包含 $m^7G_3,_{OMe}$ pppCpA。在一些实施方案中,三核苷酸帽包含 $m^7G_3,_{OMe}$ pppCpC。在一些实施方案中,三核苷酸帽包含 $m^7G_3,_{OMe}$ pppCpG。在一些实施方案中,三核苷酸帽包含 $m^7G_3,_{OMe}$ pppCpU。在一些实施方案中,三核苷酸帽包含 $m^7G_3,_{OMe}$ pppGpA。在一些实施方案中,三核苷酸帽包含 $m^7G_3,_{OMe}$ pppGpC。在一些实施方案中,三核苷酸帽包含 $m^7G_3,_{OMe}$ pppGpG。在一些实施方案中,三核苷酸帽包含 $m^7G_3,_{OMe}$ pppGpU。在一些实施方案中,三核苷酸帽包含 $m^7G_3,_{OMe}$ pppUpA。在一些实施方案中,三核苷酸帽包含 $m^7G_3,_{OMe}$ pppUpC。在一些实施方案中,三核苷酸帽包含 $m^7G_3,_{OMe}$ pppUpG。在一些实施方案中,三核苷酸帽包含 $m^7G_3,_{OMe}$ pppUpU。

[0212] 在其他实施方案中,三核苷酸帽包含选自以下序列的序列: $m^7G_{3,OMe}pppA_{2,OMe}pA$ 、 $m^7G_{3,OMe}pppA_{2,OMe}pC$ 、 $m^7G_{3,OMe}pppA_{2,OMe}pG$ 、 $m^7G_{3,OMe}pppA_{2,OMe}pU$ 、 $m^7G_{3,OMe}pppC_{2,OMe}pA$ 、 $m^7G_{3,OMe}pppC_{2,OMe}pC$ 、 $m^7G_{3,OMe}pppC_{2,OMe}pG$ 、 $m^7G_{3,OMe}pppC_{2,OMe}pU$ 、 $m^7G_{3,OMe}pppG_{2,OMe}pA$ 、 $m^7G_{3,OMe}pppG_{2,OMe}pC$ 、 $m^7G_{3,OMe}pppG_{2,OMe}pG$ 、 $m^7G_{3,OMe}pppG_{2,OMe}pU$ 、 $m^7G_{3,OMe}pppU_{2,OMe}pA$ 、 $m^7G_{3,OMe}pppU_{2,OMe}pC$ 、 $m^7G_{3,OMe}pppU_{2,OMe}pG$ 和 $m^7G_{3,OMe}pppU_{2,OMe}pU$ 。

[0213] 在一些实施方案中,三核苷酸帽包含 $m^7G_{3',OMe}pppA_{2',OMe}pA$ 。在一些实施方案中,三核苷酸帽包含 $m^7G_{3',OMe}pppA_{2',OMe}pC$ 。在一些实施方案中,三核苷酸帽包含 $m^7G_{3',OMe}pppA_{2',OMe}pG$ 。在一些实施方案中,三核苷酸帽包含 $m^7G_{3',OMe}pppA_{2',OMe}pU$ 。在一些实施方案中,三核苷酸帽包含 $m^7G_{3',OMe}pppC_{2',OMe}pA$ 。在一些实施方案中,三核苷酸帽包含 $m^7G_{3',OMe}pppC_{2',OMe}pC$ 。在一些实施方案中,三核苷酸帽包含 $m^7G_{3',OMe}pppC_{2',OMe}pG$ 。在一些实施方案中,三核苷酸帽包含 $m^7G_{3',OMe}pppC_{2',OMe}pU$ 。在一些实施方案中,三核苷酸帽包含 $m^7G_{3',OMe}pppG_{2',OMe}pA$ 。在一些实施方案中,三核苷酸帽包含 $m^7G_{3',OMe}pppG_{2',OMe}pC$ 。在一些实施方案中,三核苷酸帽包含 $m^7G_{3',OMe}pppG_{2',OMe}pG$ 。在一些实施方案中,三核苷酸帽包含 $m^7G_{3',OMe}pppG_{2',OMe}pU$ 。在一些实施方案中,三核苷酸帽包含 $m^7G_{3',OMe}pppU_{2',OMe}pA$ 。在一些实施方案中,三核苷酸帽包含 $m^7G_{3',OMe}pppU_{2',OMe}pC$ 。在一些实施方案中,三核苷酸帽包含 $m^7G_{3',OMe}pppU_{2',OMe}pG$ 。在一些实施方案中,三核苷酸帽包含 $m^7G_{3',OMe}pppU_{2',OMe}pU$ 。

[0214] 在其他实施方案中,三核苷酸帽包含选自以下序列的序列: m^7GpppA_2, m^7GpppA 、

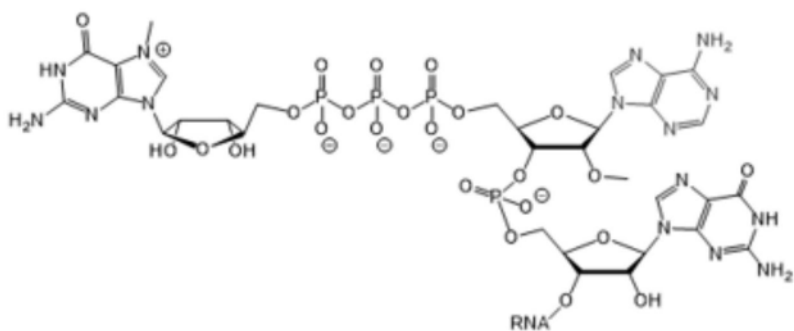
$m^7GpppA_{2,OMe}pC$ 、 $m^7GpppA_{2,OMe}pG$ 、 $m^7GpppA_{2,OMe}pU$ 、 $m^7GpppC_{2,OMe}pA$ 、 $m^7GpppC_{2,OMe}pC$ 、 $m^7GpppC_{2,OMe}pG$ 、 $m^7GpppC_{2,OMe}pU$ 、 $m^7GpppG_{2,OMe}pA$ 、 $m^7GpppG_{2,OMe}pC$ 、 $m^7GpppG_{2,OMe}pG$ 、 $m^7GpppG_{2,OMe}pU$ 、 $m^7GpppU_{2,OMe}pA$ 、 $m^7GpppU_{2,OMe}pC$ 、 $m^7GpppU_{2,OMe}pG$ 和 $m^7GpppU_{2,OMe}pU$ 。

[0215] 在一些实施方案中,三核苷酸帽包含 $m^7GpppA_{2,OMe}pA$ 。在一些实施方案中,三核苷酸帽包含 $m^7GpppA_{2,OMe}pC$ 。在一些实施方案中,三核苷酸帽包含 $m^7GpppA_{2,OMe}pG$ 。在一些实施方案中,三核苷酸帽包含 $m^7GpppA_{2,OMe}pU$ 。在一些实施方案中,三核苷酸帽包含 $m^7GpppC_{2,OMe}pA$ 。在一些实施方案中,三核苷酸帽包含 $m^7GpppC_{2,OMe}pC$ 。在一些实施方案中,三核苷酸帽包含 $m^7GpppC_{2,OMe}pG$ 。在一些实施方案中,三核苷酸帽包含 $m^7GpppC_{2,OMe}pU$ 。在一些实施方案中,三核苷酸帽包含 $m^7GpppG_{2,OMe}pA$ 。在一些实施方案中,三核苷酸帽包含 $m^7GpppG_{2,OMe}pC$ 。在一些实施方案中,三核苷酸帽包含 $m^7GpppG_{2,OMe}pG$ 。在一些实施方案中,三核苷酸帽包含 $m^7GpppG_{2,OMe}pU$ 。在一些实施方案中,三核苷酸帽包含 $m^7GpppU_{2,OMe}pA$ 。在一些实施方案中,三核苷酸帽包含 $m^7GpppU_{2,OMe}pC$ 。在一些实施方案中,三核苷酸帽包含 $m^7GpppU_{2,OMe}pG$ 。在一些实施方案中,三核苷酸帽包含 $m^7GpppU_{2,OMe}pU$ 。

[0216] 在一些实施方案中,三核苷酸帽包含 $m^7Gpppm^6A_{2,OMe}pG$ 。在一些实施方案中,三核苷酸帽包含 $m^7Gpppm^6A_{2,OMe}pG$ 。

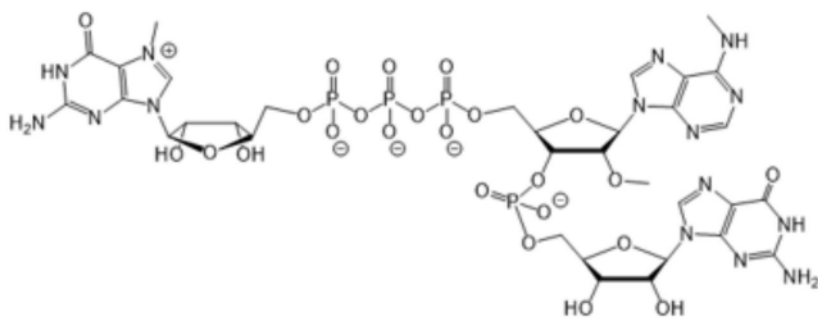
[0217] 在一些实施方案中,三核苷酸帽包含GAG。在一些实施方案中,三核苷酸帽包含GCG。在一些实施方案中,三核苷酸帽包含GUG。在一些实施方案中,三核苷酸帽包含GGG。

[0218] 在一些实施方案中,三核苷酸帽包含以下结构中的任一者:

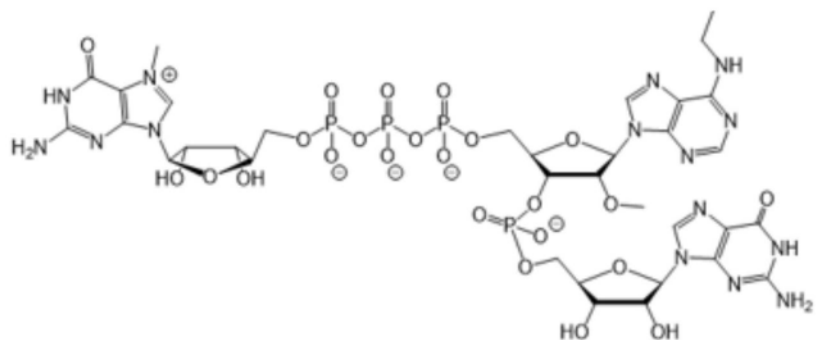


(X);

[0219]



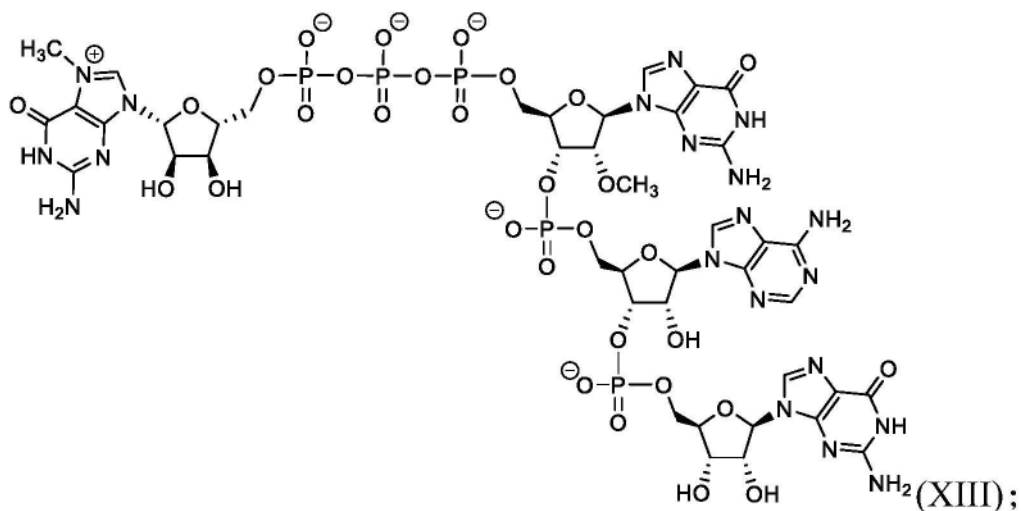
(XI); 或



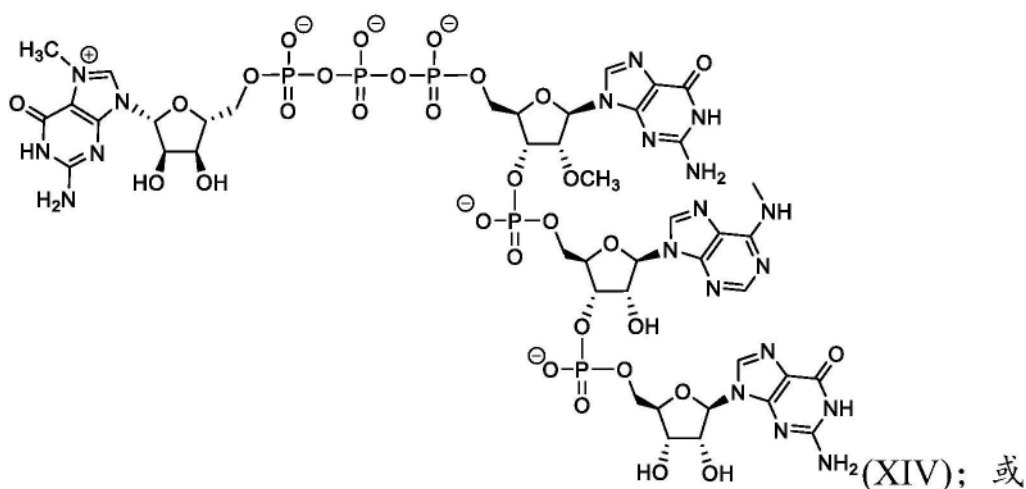
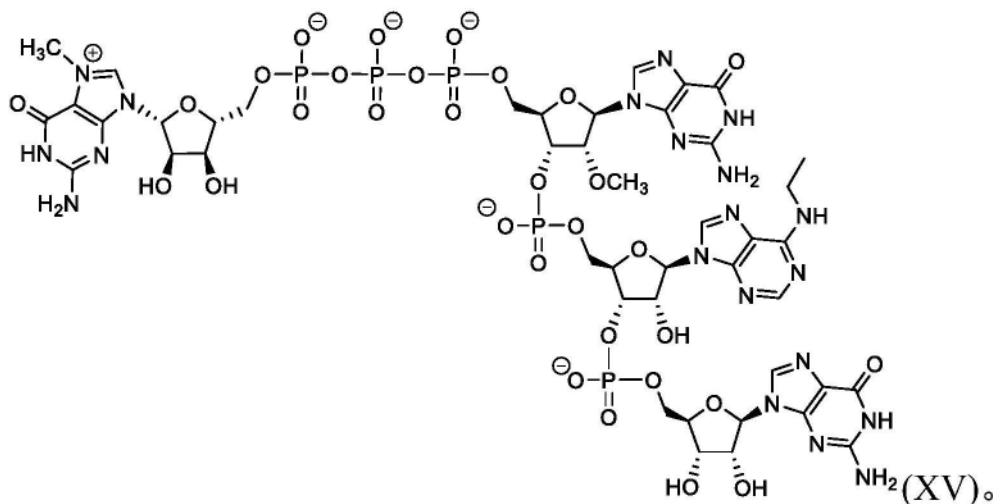
(XII)。

[0220] 在一些实施方案中,帽类似物包括四核苷酸帽。在一些实施方案中,帽类似物包含 GGAG。

[0221] 在一些实施方案中,四核苷酸帽包含以下结构中的任一者:



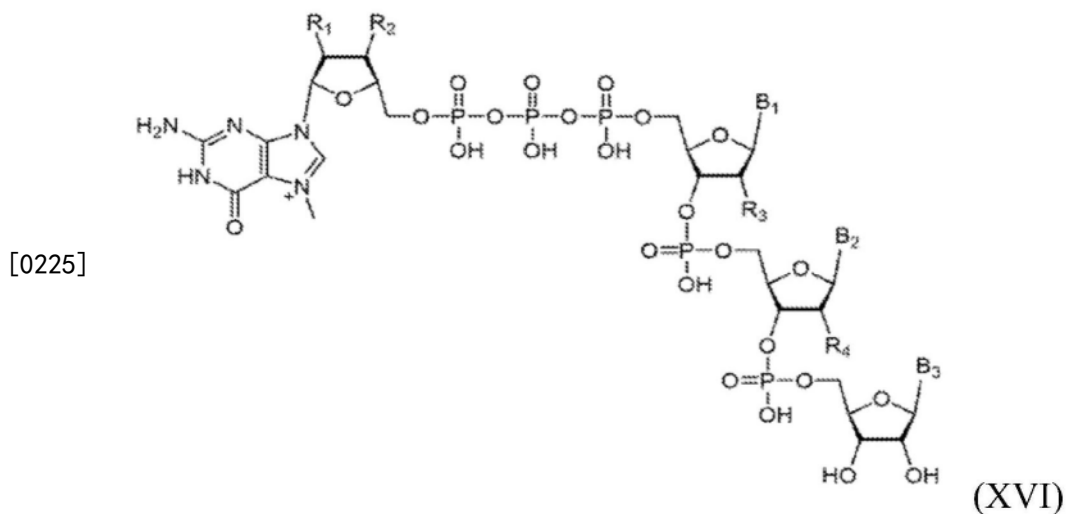
[0222]



[0223] 在一些实施方案中,四核苷酸帽包含如上所示的三核苷酸。在一些实施方案中,四核苷酸帽包含^{m7}GpppN₁N₂N₃,其中N₁、N₂和N₃是任选的(即,可以不存在或一者或者多者可以存在)并且独立地是天然的、修饰的或非天然的核苷碱基。在一些实施方案中,^{m7}G被进一步甲基化,例如,在3'位置处。在一些实施方案中,^{m7}G在3'位置处包含O-甲基。在一些实施方案中,N₁、N₂和N₃,如果存在的话,任选地独立地是腺嘌呤、尿嘧啶、胍、胸腺嘧啶或胞嘧啶。在一

些实施方案中, N_1 、 N_2 和 N_3 中的一者或多者(或全部), 如果存在的话, 例如在2'位置处被甲基化。在一些实施方案中, N_1 、 N_2 和 N_3 中的一者或多者(或全部), 如果存在的话, 在2'位置处具有O-甲基。

[0224] 在一些实施方案中, 四核苷酸帽包含以下结构:



[0226] 其中 B_1 、 B_2 和 B_3 独立地是天然的、修饰的或非天然的核苷碱基; 并且 R_1 、 R_2 、 R_3 和 R_4 独立地是OH或O-甲基。在一些实施方案中, R_3 是O-甲基并且 R_4 是OH。在一些实施方案中, R_3 和 R_4 是O-甲基。在一些实施方案中, R_4 是O-甲基。在一些实施方案中, R_1 是OH, R_2 是OH, R_3 是O-甲基, 并且 R_4 是OH。在一些实施方案中, R_1 是OH, R_2 是OH, R_3 是O-甲基, 并且 R_4 是O-甲基。在一些实施方案中, R_1 和 R_2 中的至少一者是O-甲基, R_3 是O-甲基, 并且 R_4 是OH。在一些实施方案中, R_1 和 R_2 中的至少一者是O-甲基, R_3 是O-甲基, 并且 R_4 是O-甲基。

[0227] 在一些实施方案中, B_1 、 B_2 和 B_3 是天然核苷碱基。在一些实施方案中, B_1 、 B_2 和 B_3 中的至少一者是修饰的或非天然的碱基。在一些实施方案中, B_1 、 B_2 和 B_3 中的至少一者是N6-甲基腺嘌呤。在一些实施方案中, B_1 是腺嘌呤、胞嘧啶、胸腺嘧啶或尿嘧啶。在一些实施方案中, B_1 是腺嘌呤, B_2 是尿嘧啶, 并且 B_3 是腺嘌呤。在一些实施方案中, R_1 和 R_2 是OH, R_3 和 R_4 是O-甲基, B_1 是腺嘌呤, B_2 是尿嘧啶, 并且 B_3 是腺嘌呤。

[0228] 在一些实施方案中, 四核苷酸帽包含选自以下序列的序列: GAAA、GACA、GAGA、GAUA、GCAA、GCCA、GCGA、GCUA、GGAA、GGCA、GGGA、GGUA、GUCA和GUUA。在一些实施方案中, 四核苷酸帽包含选自以下序列的序列: GAAG、GACG、GAGG、GAUG、GCAG、GCCG、GCGG、GCUG、GGAG、GGCG、GGGG、GGUG、GUCG、GUGG和GUUG。在一些实施方案中, 四核苷酸帽包含选自以下序列的序列: GAAU、GACU、GAGU、GAUU、GCAU、GCCU、GCGU、GCUU、GGAU、GGCU、GGGU、GGUU、GUAU、GUCU、GUGU和GUUU。在一些实施方案中, 四核苷酸帽包含选自以下序列的序列: GAAC、GACC、GAGC、GAUC、GCAC、GCCC、GCGC、GCUC、GGAC、GGCC、GGGC、GGUC、GUAC、GUCC、GUGC和GUUC。

[0229] 在一些实施方案中, 四核苷酸帽包含选自以下序列的序列: $m^7G_{3', OMe}pppApApN$ 、 $m^7G_{3', OMe}pppApCpN$ 、 $m^7G_{3', OMe}pppApGpN$ 、 $m^7G_{3', OMe}pppApUpN$ 、 $m^7G_{3', OMe}pppCpApN$ 、 $m^7G_{3', OMe}pppCpCpN$ 、 $m^7G_{3', OMe}pppCpGpN$ 、 $m^7G_{3', OMe}pppCpUpN$ 、 $m^7G_{3', OMe}pppGpApN$ 、 $m^7G_{3', OMe}pppGpCpN$ 、 $m^7G_{3', OMe}pppGpGpN$ 、 $m^7G_{3', OMe}pppGpUpN$ 、 $m^7G_{3', OMe}pppUpApN$ 、 $m^7G_{3', OMe}pppUpCpN$ 、 $m^7G_{3', OMe}pppUpGpN$ 和 $m^7G_{3', OMe}pppUpUpN$, 其中N是天然的、修饰的或非天然的核苷碱基。

[0230] 在其他实施方案中, 四核苷酸帽包含选自以下序列的序列: $m^7G_{3', OMe}pppA_{2', OMe}pApN$ 、

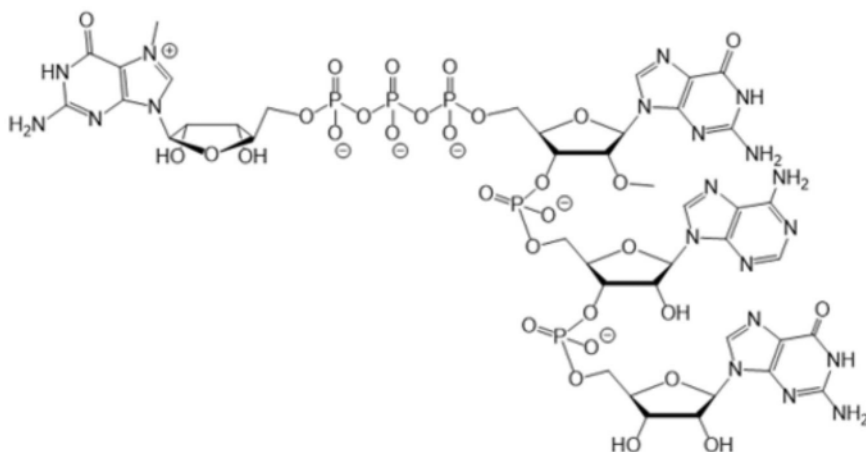
$m^7G_{3',OMe}pppA_{2',OMe}pCpN$ 、 $m^7G_{3',OMe}pppA_{2',OMe}pGpN$ 、 $m^7G_{3',OMe}pppA_{2',OMe}pUpN$ 、 $m^7G_{3',OMe}pppC_{2',OMe}pApN$ 、 $m^7G_{3',OMe}pppC_{2',OMe}pCpN$ 、 $m^7G_{3',OMe}pppC_{2',OMe}pGpN$ 、 $m^7G_{3',OMe}pppC_{2',OMe}pUpN$ 、 $m^7G_{3',OMe}pppG_{2',OMe}pApN$ 、 $m^7G_{3',OMe}pppG_{2',OMe}pCpN$ 、 $m^7G_{3',OMe}pppG_{2',OMe}pGpN$ 、 $m^7G_{3',OMe}pppG_{2',OMe}pUpN$ 、 $m^7G_{3',OMe}pppU_{2',OMe}pApN$ 、 $m^7G_{3',OMe}pppU_{2',OMe}pCpN$ 、 $m^7G_{3',OMe}pppU_{2',OMe}pGpN$ 和 $m^7G_{3',OMe}pppU_{2',OMe}pUpN$ ，其中N是天然的、修饰的或非天然的核苷碱基。

[0231] 在其他实施方案中，四核苷酸帽包含选自以下序列的序列： $m^7GpppA_{2',OMe}pApN$ 、 $m^7GpppA_{2',OMe}pCpN$ 、 $m^7GpppA_{2',OMe}pGpN$ 、 $m^7GpppA_{2',OMe}pUpN$ 、 $m^7GpppC_{2',OMe}pApN$ 、 $m^7GpppC_{2',OMe}pCpN$ 、 $m^7GpppC_{2',OMe}pGpN$ 、 $m^7GpppC_{2',OMe}pUpN$ 、 $m^7GpppG_{2',OMe}pApN$ 、 $m^7GpppG_{2',OMe}pCpN$ 、 $m^7GpppG_{2',OMe}pGpN$ 、 $m^7GpppG_{2',OMe}pUpN$ 、 $m^7GpppU_{2',OMe}pApN$ 、 $m^7GpppU_{2',OMe}pCpN$ 、 $m^7GpppU_{2',OMe}pGpN$ 和 $m^7GpppU_{2',OMe}pUpN$ ，其中N是天然的、修饰的或非天然的核苷碱基。

[0232] 在其他实施方案中，四核苷酸帽包含选自以下序列的序列： $m^7G_{3',OMe}pppA_{2',OMe}pA_{2',OMe}pN$ 、 $m^7G_{3',OMe}pppA_{2',OMe}pC_{2',OMe}pN$ 、 $m^7G_{3',OMe}pppA_{2',OMe}pG_{2',OMe}pN$ 、 $m^7G_{3',OMe}pppA_{2',OMe}pU_{2',OMe}pN$ 、 $m^7G_{3',OMe}pppC_{2',OMe}pA_{2',OMe}pN$ 、 $m^7G_{3',OMe}pppC_{2',OMe}pC_{2',OMe}pN$ 、 $m^7G_{3',OMe}pppC_{2',OMe}pG_{2',OMe}pN$ 、 $m^7G_{3',OMe}pppC_{2',OMe}pU_{2',OMe}pN$ 、 $m^7G_{3',OMe}pppG_{2',OMe}pA_{2',OMe}pN$ 、 $m^7G_{3',OMe}pppG_{2',OMe}pC_{2',OMe}pN$ 、 $m^7G_{3',OMe}pppG_{2',OMe}pG_{2',OMe}pN$ 、 $m^7G_{3',OMe}pppG_{2',OMe}pU_{2',OMe}pN$ 、 $m^7G_{3',OMe}pppU_{2',OMe}pA_{2',OMe}pN$ 、 $m^7G_{3',OMe}pppU_{2',OMe}pC_{2',OMe}pN$ 、 $m^7G_{3',OMe}pppU_{2',OMe}pG_{2',OMe}pN$ 和 $m^7G_{3',OMe}pppU_{2',OMe}pU_{2',OMe}pN$ ，其中N是天然的、修饰的或非天然的核苷碱基。

[0233] 在其他实施方案中，四核苷酸帽包含选自以下序列的序列： $m^7GpppA_{2',OMe}pA_{2',OMe}pN$ 、 $m^7GpppA_{2',OMe}pC_{2',OMe}pN$ 、 $m^7GpppA_{2',OMe}pG_{2',OMe}pN$ 、 $m^7GpppA_{2',OMe}pU_{2',OMe}pN$ 、 $m^7GpppC_{2',OMe}pA_{2',OMe}pN$ 、 $m^7GpppC_{2',OMe}pC_{2',OMe}pN$ 、 $m^7GpppC_{2',OMe}pG_{2',OMe}pN$ 、 $m^7GpppC_{2',OMe}pU_{2',OMe}pN$ 、 $m^7GpppG_{2',OMe}pA_{2',OMe}pN$ 、 $m^7GpppG_{2',OMe}pC_{2',OMe}pN$ 、 $m^7GpppG_{2',OMe}pG_{2',OMe}pN$ 、 $m^7GpppG_{2',OMe}pU_{2',OMe}pN$ 、 $m^7GpppU_{2',OMe}pA_{2',OMe}pN$ 、 $m^7GpppU_{2',OMe}pC_{2',OMe}pN$ 、 $m^7GpppU_{2',OMe}pG_{2',OMe}pN$ 和 $m^7GpppU_{2',OMe}pU_{2',OMe}pN$ ，其中N是天然的、修饰的或非天然的核苷碱基。

[0234] 在一些实施方案中，四核苷酸帽包含GGAG。在一些实施方案中，四核苷酸帽包含以下结构：



[0235]

[0236] 转录后或共转录加帽反应的加帽效率可变化。术语“加帽效率”可指包含帽结构的mRNA相对于混合物中的总mRNA的量(例如,以百分比表示)(例如,翻译后加帽反应或共转录调用反应(capping reaction))。在一些实施方案中,加帽反应的加帽效率为至少60%、70%、80%、90%、95%、99%或99.9%(例如,在加帽反应后至少60%、70%、80%、90%、

95%、99%或99.9%的输入mRNA包含帽)。在一些实施方案中,多价共IVT反应不影响由IVT反应产生的mRNA的加帽效率。

[0237] 体外转录方法

[0238] 一些方面涉及产生(例如,合成)RNA转录物(例如,mRNA转录物)的方法,所述方法包括在导致RNA转录物产生的条件下使DNA模板(例如,第一输入DNA和第二输入DNA)与RNA聚合酶(例如,T7 RNA聚合酶、T7 RNA聚合酶变体等)接触。

[0239] 一些方面涉及进行IVT反应的方法,所述方法包括在核苷三磷酸和缓冲液存在下在导致RNA转录物产生的条件下使DNA模板与RNA聚合酶(例如,T7 RNA聚合酶,诸如T7 RNA聚合酶变体)接触。

[0240] 其他方面提供了共转录加帽方法,所述方法包括在体外转录反应条件下使多核苷酸模板与T7 RNA聚合酶变体、核苷三磷酸和帽类似物反应以产生RNA转录物。

[0241] 在一些实施方案中,用于RNA合成的共转录加帽方法包括在体外转录反应条件下使多核苷酸模板与以下反应:(a)相对于野生型RNA聚合酶,包含至少一个氨基酸取代的T7 RNA聚合酶变体(例如,相对于SEQ ID NO:1,包含位置437、387、350和351处的氨基酸取代的T7聚合酶变体)、(b)核苷三磷酸和(c)帽类似物(例如,包含序列GpppA_{2'}·_{0me}pG的三核苷酸帽),以产生RNA转录物,其中多核苷酸模板包含在模板位置+1处的2'-脱氧胸苷残基。

[0242] IVT条件通常需要纯化的线性DNA模板,所述模板含有启动子、核苷三磷酸、包含二硫苏糖醇(DTT)和镁离子的缓冲系统以及RNA聚合酶。转录反应中使用的确切条件取决于特定应用所需的RNA的量。典型的IVT反应可以通过在转录缓冲液中温育DNA模板与RNA聚合酶和核苷三磷酸(包括GTP、ATP、CTP、和UTP(或核苷酸类似物))来进行。由此反应产生具有5'末端鸟苷三磷酸的RNA转录物。

[0243] 两个序列(例如,核酸或氨基酸)的“百分比同一性”、“序列同一性”、“%同一性”或“%序列同一性”(如它们可在本文中互换使用)是指两个序列(例如,核酸或氨基酸)之间的类似性的定量测量。百分比同一性可使用Karlin和Altschul,Proc.Natl.Acad.Sci.USA 87:2264-68,1990的算法来确定,按照Karlin和Altschul,Proc.Natl.Acad.Sci.USA90:5873-77,1993进行修改。此类算法被并入Altschul等人,J.Mol.Biol.215:403-10,1990的NBLAST和XBLAST程序(版本2.0)中。可用XBLAST程序进行BLAST蛋白质搜索,分数=50,字长=3,以获得与感兴趣的蛋白质分子同源的氨基酸序列。在两个序列之间存在缺口的情况下,可以如Altschul等人,Nucleic Acids Res.25(17):3389-3402,1997中所述利用Gapped BLAST。当利用BLAST和Gapped BLAST程序时,可以使用相应程序(例如,XBLAST和NBLAST)的默认参数。除非另有说明,否则当规定百分比同一性或其范围(例如,至少、超过等)时,端点应是包含性的并且范围(例如,至少70%同一性)应包括引用范围内的所有范围。

[0244] 输入脱氧核糖核酸(DNA)充当RNA聚合酶的核酸模板。DNA模板可以包含编码感兴趣的多肽(例如,抗原性多肽)的多核苷酸。在一些实施方案中,DNA模板包含RNA聚合酶启动子(例如,T7 RNA聚合酶启动子),所述启动子位于编码感兴趣的多肽的多核苷酸的5'并且可操作地连接至所述多核苷酸。DNA模板还可以包含编码位于感兴趣的基因的3'端的多聚腺苷酸化(多聚A)尾的核苷酸序列。在一些实施方案中,输入DNA包含质粒DNA(pDNA)。术语“质粒DNA”或“pDNA”可指在细胞中与染色体DNA物理分离并且可以独立复制的染色体外DNA分子。在一些实施方案中,从细胞中分离质粒DNA(例如,作为质粒DNA制备物)。在一些实施

方案中,质粒DNA包含复制起点,其可含有一种或多种异源核酸,例如编码可用作RNA聚合酶模板的治疗性蛋白的核酸。质粒DNA可以是环化的或线性的(例如,已通过限制酶消化而线性化的质粒DNA)。

[0245] 在一些实施方案中,共IVT反应中的每个输入DNA(例如,输入DNA分子群体)获自不同来源(例如,例如在不同细胞或细胞群体中单独合成)。在一些实施方案中,各输入DNA(例如,输入DNA群体)获自不同的细菌细胞或细菌细胞群体。举例来说,在具有输入DNA的三个群体的共IVT反应中,第一输入DNA在细菌细胞群体A中产生,第二输入DNA在细菌细胞群体B中产生,并且第三输入DNA在细菌群体C中产生,其中A、B和C中的每一者不是相同的细菌培养物(例如,在相同容器或板中共培养)。在另一个实例中,获自不同来源的两个输入DNA i) 在单独的合成反应中化学合成,或ii) 通过单独的扩增(例如,聚合酶链反应(PCR反应))产生。获得输入DNA群体(例如,质粒DNA)的方法是已知的,例如由 Sambrook, Joseph. *Molecular Cloning: a Laboratory Manual*. Cold Spring Harbor, N.Y.: Cold Spring Harbor Laboratory Press, 2001所述。

[0246] 一些方面包括将多价共IVT反应中使用的DNA的量归一化。在一些实施方案中,归一化是基于输入DNA的摩尔质量。在一些实施方案中,归一化是基于输入DNA的降解率。在一些实施方案中,归一化是基于所得mRNA的降解率(例如,基于反应混合物中存在的多聚A变体,或T7聚合酶无效转录物或截短的转录物来测量)。在一些实施方案中,归一化是基于输入DNA的核苷酸含量(例如,A、G、C、U或其任何组合的量)。在一些实施方案中,归一化是基于输入DNA的纯度。在一些实施方案中,归一化是基于输入DNA的多聚A加尾效率。在一些实施方案中,归一化是基于输入DNA的长度。

[0247] 在一些实施方案中,归一化是基于输入DNA中存在的最低水平(例如,最低摩尔质量、降解率(例如,输入DNA和/或输出RNA的降解率)、核苷酸含量、纯度和/或多聚A加尾效率)。在一些实施方案中,归一化是基于输入DNA中存在的最高水平(例如,最高摩尔质量、降解率(例如,输入DNA和/或输出RNA的降解率)、核苷酸含量、纯度和/或多聚A加尾效率)。在一些实施方案中,归一化是基于输入DNA的RNA产生速率(例如,反应混合物中输入DNA的最高RNA产生速率或输入DNA的最低RNA产生速率)。

[0248] 一些方面涉及IVT方法,其中调整或归一化输入DNA(例如,第一DNA或第二DNA)的量,以提高具有预定义的mRNA组分比率的多价RNA组合物的产生。本公开部分基于以下发现:影响多价RNA组合物纯度的某些因素,诸如在IVT期间输入DNA之间大小的巨大差异(例如,长度超过100、200、500、1000或更多个核苷酸的差异)和/或给定DNA的多聚A加尾效率,可以在IVT之前通过基于这些因素中的一者或多者将输入DNA的量归一化得到解决。举例来说,在一些实施方案中,两个输入DNA的量基于从输入DNA转录的第一RNA与第二RNA的所期望的摩尔比来计算。在一些实施方案中,计算包括基于输入DNA所期望的摩尔比确定质粒质量比。在一些实施方案中,在IVT期间基于输入DNA的最高多聚A加尾效率来归一化输入DNA的量。

[0249] IVT反应中使用的输入DNA(例如,输入DNA分子群体)的数量可以变化,取决于多价RNA组合物中所期望包含的不同RNA分子的数量。在一些实施方案中,IVT反应混合物包含2个或更多个不同的输入DNA,例如2、3、4、5、6、7、8、9、10或更多个不同的输入DNA。在一些实施方案中,IVT反应包含超过10个不同的输入DNA。术语“不同的输入DNA”涵盖编码不同的

RNA例如,具有以下的不同RNA的输入DNA:i)不同的长度(无论RNA在两个长度中较短者的整个长度上是否同一)、ii)不同的核苷酸序列、iii)不同的化学修饰模式,或iv)前述任何组合。

[0250] DNA分子群体中的每一者的浓度也可变化。在一些实施方案中,IVT反应中的每个DNA分子群体的浓度在约0.005mg/mL至约0.5mg/ml的范围内。在一些实施方案中,IVT反应中每个DNA分子群体的浓度在约0.02mg/ml至约0.05mg/ml、0.02至约0.15mg/ml、约0.05mg/ml至约0.20mg/ml、约0.175至约0.3mg/ml、约0.2mg/ml至约0.5mg/ml、约0.3mg/ml至约0.6mg/ml、约0.5mg/ml至约0.75mg/ml、约0.5mg/ml至约1.0mg/ml、约0.75mg/ml至约0.9mg/ml、约0.75mg/ml至约1.5mg/ml、约0.8mg/ml至约1.2mg/ml、约1.0mg/ml至约1.5mg/ml、约1.0mg/ml至约2.5mg/ml、约1.5mg/ml至约3.0mg/ml、约2.0mg/ml至约4.0mg/ml或约2.5mg/ml至约5.0mg/ml的范围内。

[0251] 在一些实施方案中,添加至IVT反应中的输入DNA是预定义的DNA比率,其可以包括2、3、4、5、6、7、8、9、10或更多个不同输入DNA之间的比率(例如,取决于组合中不同RNA的数量)。在一些实施方案中,预定义的输入DNA比率包括超过10个输入DNA之间的比率。术语“预定义的输入DNA比率”可指IVT反应中DNA分子所期望的最终比率。输入DNA所期望的最终比率可取决于由输入DNA编码的RNA编码的最终肽或多肽产物。在一些实施方案中,输入DNA可具有期望的比率,其可包含2至8个输入DNA(例如,a:b、a:b:c、a:b:c:d、a:b:c:d:e、a:b:c:d:e:f、a:b:c:d:e:f:g、a:b:c:d:e:f:g:h等,其中a-h中的每一者是1至10之间的数)。在一些实施方案中,预定义的输入DNA比率与预定义的mRNA比率不同。

[0252] 两个或更多个输入DNA(例如,两个或更多个不同输入DNA群体中的DNA)的大小可变化。在一些实施方案中,输入DNA包括约15至约8,000个碱基对(例如,15至50、15至100、15至200、15至300、15至400、15至500、15至600、15至700、15至800、15至900、15至1000、15至1200、15至1400、15至1500、15至1800、15至2000、15至2500、15至3000、50至100、50至200、50至300、50至400、50至500、50至600、50至700、50至800、50至900、50至1000、50至1200、50至1400、50至1500、50至1800、50至2000、50至2500、50至3000、100至200、100至300、100至400、100至500、100至600、100至700、100至800、100至900、100至1000、100至1200、100至1400、100至1500、100至1800、100至2000、100至2500、100至3000、200至300、200至400、200至500、200至600、200至700、200至800、200至900、200至1000、200至1500、200至3000、500至1000、500至1500、500至2000、500至2500、500至3000、1000至1500、1000至2000、1000至2500、1000至3000、1500至3000、2500至3000、2000至3000、2500至4000、3000至5000、3500至6500、5000至7500或6500至8000个碱基对)。

[0253] IVT反应中每个输入DNA分子群体的质量可变化。在一些实施方案中,每个输入DNA群体的质量基于IVT反应混合物的总体积变化。在一些实施方案中,IVT混合物中每个输入DNA分子的每个群体的质量个别地从IVT反应混合物中存在的总输入DNA的约0.5%至约99.9%变化。在一些实施方案中,IVT反应中每个输入DNA分子群体的摩尔比可变化。

[0254] 在一些实施方案中,IVT反应中使用的两个或更多个输入DNA分子具有不同长度(例如,包含不同数量的核苷酸)。在一些实施方案中,IVT反应混合物中的两个或更多个(例如,3、4、5、6、7、8、9、10或更多个)不同输入DNA分子之间的长度差异大于70个碱基对、80个碱基对、90个碱基对或100个碱基对(例如,组合物中的两个输入DNA彼此的长度不在70、80、

90或100个碱基对内)。在一些实施方案中,两个或更多个(例如,3、4、5、6、7、8、9、10或更多个)不同输入DNA分子之间的长度差异超过100个碱基对,例如500个碱基对、1000个碱基对、1500个碱基对、2000个碱基对、3000个碱基对、4000个碱基对、5000个碱基对、6000个碱基对、7000个碱基对、8000个碱基对或更多。

[0255] 在一些实施方案中,IVT反应中使用的两个或更多个输入DNA分子编码具有不同长度(例如,包含不同数量的核苷酸)的mRNA分子。在一些实施方案中,IVT反应混合物中由不同输入DNA分子编码的两个或更多个mRNA分子之间的长度差异大于70个核苷酸、80个核苷酸、90个核苷酸或100个核苷酸(例如,组合物中的两个输入DNA编码彼此长度不在70个、80个、90个或100个核苷酸以内的mRNA分子)。在一些实施方案中,由不同输入DNA分子编码的两个或更多个mRNA分子之间的长度差异超过100个核苷酸,例如500个核苷酸、1000个核苷酸、1500个核苷酸、2000个核苷酸、3000个核苷酸、4000个核苷酸或更多。

[0256] 在一些实施方案中,多价IVT包括以A:B:C:D:E:F:G:H:I:J的比率共转录至少2个不同的输入DNA(例如,DNA A、B、C、D、E、F、G、H、I、J等中的至少2者),其中如果DNA A被归一化为1,则DNA B、C、D、E、F、G、H、I、J等中的一者或多者可各自独立地以A的量(例如,浓度)的0.01至100倍的量(例如,浓度)存在,诸如A的量的0.05倍至20倍、A的量的0.1倍至10倍、A的量的0.2倍至5倍、A的量的0.3倍至3倍、A的量的0.5倍至2倍、A的量的0.75倍至1.4倍、A的量的0.8倍至1.25倍或A的量的0.9倍至1.15倍。DNA B、C、D、E、F、G、H、I或J中的一者或多者也可不存在。

[0257] 在一些实施方案中,通过组合来自单独来源的RNA转录物(例如,mRNA)来产生多价RNA组合物。在一些实施方案中,通过在单独的IVT反应中单独转录两个或更多个DNA模板并组合转录RNA来产生多价RNA组合物。在一些实施方案中,RNA转录物通过IVT产生,然后添加至一种或多种其他RNA中。RNA可以任何所期望的量进行组合,以产生包含特定的比率的两个或更多个RNA的多价RNA组合物。

[0258] 在一些实施方案中,RNA转录物是IVT反应的产物。在一些实施方案中,RNA转录物是信使RNA(mRNA),其包含编码连接至多聚A尾的感兴趣的多肽(例如,治疗性蛋白质或治疗性肽)的核苷酸序列。在一些实施方案中,mRNA是修饰的mRNA(mmRNA),其包含至少一个修饰的核苷酸。

[0259] 核苷三磷酸(NTP)可以包含未修饰或修饰的ATP、修饰或未修饰的UTP、修饰或未修饰的GTP和/或修饰或未修饰的CTP。在一些实施方案中,IVT反应的NTP包含未修饰的ATP。在一些实施方案中,IVT反应的NTP包含修饰的ATP。在一些实施方案中,IVT反应的NTP包含未修饰的UTP。在一些实施方案中,IVT反应的NTP包含修饰的UTP。在一些实施方案中,IVT反应的NTP包含未修饰的GTP。在一些实施方案中,IVT反应的NTP包含修饰的GTP。在一些实施方案中,IVT反应的NTP包含未修饰的CTP。在一些实施方案中,IVT反应的NTP包含修饰的CTP。

[0260] IVT反应中NTP的组成也可变化。在一些实施方案中,IVT反应中的每个NTP以等摩尔量存在。在一些实施方案中,IVT反应中的每个NTP以非等摩尔量存在。举例来说,可以使用超过GTP、CTP和UTP的ATP。作为非限制性实例,IVT反应可以包含7.5毫摩尔GTP、7.5毫摩尔CTP、7.5毫摩尔UTP和3.75毫摩尔ATP。在一些实施方案中,G:C:U:A的摩尔比为2:1:0.5:1。在一些实施方案中,G:C:U:A的摩尔比为1:1:0.7:1。在一些实施方案中,G:C:A:U的摩尔比为1:1:1:1。相同的IVT反应可包含3.75毫摩尔帽类似物(例如,三核苷酸帽或四核苷酸

帽)。在一些实施方案中,G:C:U:A:帽的摩尔比为1:1:1:0.5:0.5。在一些实施方案中,G:C:U:A:帽的摩尔比为1:1:0.5:1:0.5。在一些实施方案中,G:C:U:A:帽的摩尔比为1:0.5:1:1:0.5。在一些实施方案中,G:C:U:A:帽的摩尔比为0.5:1:1:1:0.5。在一些实施方案中,共IVT反应中NTP的量根据经验计算。举例来说,IVT反应中每个NTP的消耗率可针对每个单个输入DNA根据经验确定,然后基于那些单个NTP消耗率的NTP的平衡比率可添加至包含多个输入DNA的共IVT中。

[0261] 在一些实施方案中,IVT反应混合物进一步包含帽类似物。存在于IVT反应中的核苷三磷酸和帽类似物的浓度可变化。在一些实施方案中,NTP和帽类似物以等摩尔浓度存在于反应中。在一些实施方案中,反应中帽类似物(例如,三核苷酸帽或四核苷酸帽)与核苷三磷酸的摩尔比大于1:1。举例来说,反应中帽类似物与核苷三磷酸的摩尔比可以为2:1、3:1、4:1、5:1、6:1、7:1、8:1、9:1、10:1、15:1、20:1、25:1、50:1或100:1。在一些实施方案中,反应中帽类似物(例如,三核苷酸帽或四核苷酸帽)与核苷三磷酸的摩尔比小于1:1。举例来说,反应中帽类似物(例如,三核苷酸帽或四核苷酸帽)与核苷三磷酸的摩尔比可以为1:2、1:3、1:4、1:5、1:6、1:7、1:8、1:9、1:10、1:15、1:20、1:25、1:50或1:100。

[0262] 在一些实施方案中,RNA转录物(例如,mRNA转录物)包含选自以下的修饰的核碱基:假尿苷(ψ)、1-甲基假尿苷($m^1\psi$)、5-甲氧基尿苷(mo^5U)、5-甲基胞苷(m^5C)、 α -硫代-鸟苷和 α -硫代-腺苷。在一些实施方案中,RNA转录物(例如,mRNA转录物)包括至少两种(例如,2种、3种、4种或更多种)前述修饰的核碱基的组合。

[0263] 在一些实施方案中,RNA转录物(例如,mRNA转录物)包括假尿苷(ψ)。在一些实施方案中,RNA转录物(例如,mRNA转录物)包括1-甲基假尿苷($m^1\psi$)。在一些实施方案中,RNA转录物(例如,mRNA转录物)包括5-甲氧基尿苷(mo^5U)。在一些实施方案中,RNA转录物(例如,mRNA转录物)包括5-甲基胞苷(m^5C)。在一些实施方案中,RNA转录物(例如,mRNA转录物)包括 α -硫-鸟苷。在一些实施方案中,RNA转录物(例如,mRNA转录物)包括 α -硫-腺苷。

[0264] 在一些实施方案中,多核苷酸(例如RNA多核苷酸,诸如mRNA多核苷酸)被统一修饰(例如,完全修饰,遍及整个序列进行修饰)以用于特定修饰。举例来说,多核苷酸可以统一被1-甲基假尿苷($m^1\psi$)修饰,意指mRNA序列中的所有尿苷残基均被1-甲基假尿苷($m^1\psi$)置换。类似地,多核苷酸可以通过用修饰的残基(诸如上文所阐述的那些残基中的任一者)置换而针对存在于序列中的任何类型的核苷残基进行统一修饰。替代地,多核苷酸(例如,RNA多核苷酸,诸如mRNA多核苷酸)可以不被统一修饰(例如,部分修饰,序列的一部分被修饰)。每种可能性代表单独的实施方案。

[0265] IVT反应混合物的缓冲系统可变化。在一些实施方案中,缓冲系统含有tris。举例来说,IVT反应中使用的tris的浓度可以为至少10mM、至少20mM、至少30mM、至少40mM、至少50mM、至少60mM、至少70mM、至少80mM、至少90mM、至少100mM或至少110mM磷酸盐。在一些实施方案中,磷酸盐的浓度为20-60mM或10-100mM。

[0266] 在一些实施方案中,缓冲系统含有二硫苏糖醇(DTT)。举例来说,IVT反应中使用的DTT的浓度可以为至少1mM、至少5mM或至少50mM。在一些实施方案中,IVT反应中使用的DTT的浓度为1-50mM或5-50mM。在一些实施方案中,IVT反应中使用的DTT的浓度为5mM。

[0267] 在一些实施方案中,缓冲系统含有镁。在一些实施方案中,存在于IVT反应中的NTP与镁离子(Mg^{2+} ;例如, $MgCl_2$)的摩尔比为1:1至1:5。举例来说,NTP与镁离子的摩尔比可以为

1:0.25、1:0.5、1:1、1:2、1:3、1:4或1:5。

[0268] 在一些实施方案中,存在于IVT反应中的NTP加上帽类似物(例如,三核苷酸帽,诸如GAG)与镁离子(Mg^{2+} ;例如, $MgCl_2$)的摩尔比为1:1至1:5。举例来说,NTP+三核苷酸帽(例如,GAG)与镁离子的摩尔比可以为1:1、1:2、1:3、1:4或1:5。

[0269] 在一些实施方案中,缓冲系统含有Tris-HCl、亚精胺(例如,浓度为1-30mM)、TRITON[®]X-100(聚乙二醇对(1,1,3,3-四甲基丁基)-苯基醚)和/或聚乙二醇(PEG)。

[0270] 在一些实施方案中,IVT方法进一步包括将体外转录产物(例如,mRNA)与其他反应组分分离(例如,纯化)的步骤。在一些实施方案中,分离包括对IVT反应混合物进行色谱法。在一些实施方案中,色谱法包括基于大小(例如,基于长度)的色谱法。在一些实施方案中,色谱法包括寡dT色谱法。

[0271] 将核苷三磷酸(NTP)添加至生长RNA链的3'端是通过聚合酶,诸如T7 RNA聚合酶,例如,T7 RNA聚合酶变体(例如,包含D653W/E350W/D351V取代的RNA聚合酶)催化的。在一些实施方案中,RNA聚合酶(例如,T7 RNA聚合酶变体)以0.01mg/ml至1mg/ml的浓度存在于反应(例如,IVT反应)中。举例来说,RNA聚合酶可以0.01mg/mL、0.05mg/ml、0.1mg/ml、0.5mg/ml或1.0mg/ml的浓度存在于反应中。

[0272] 令人惊讶的是,体外转录反应中T7 RNA聚合酶变体(例如,包含D653W/E350W/D351V取代的RNA聚合酶)与帽类似物(例如, $GpppA_{2,0me}pG$)的组合的使用,例如,导致RNA转录物的产生,其中大于80%的产生的RNA转录物包含功能性帽。在一些实施方案中,大于85%的产生的RNA转录物包含功能性帽。在一些实施方案中,大于90%的产生的RNA转录物包含功能性帽。在一些实施方案中,大于95%的产生的RNA转录物包含功能性帽。在一些实施方案中,大于96%的产生的RNA转录物包含功能性帽。在一些实施方案中,大于97%的产生的RNA转录物包含功能性帽。在一些实施方案中,大于98%的产生的RNA转录物包含功能性帽。在一些实施方案中,大于99%的产生的RNA转录物包含功能性帽。

[0273] 还令人惊讶的是,发现包含在模板位置+1处的2'-脱氧胸苷残基或2'-脱氧胞苷残基的多核苷酸模板的使用导致RNA转录物的产生,其中大于80%(例如,大于85%、大于90%或大于95%)的产生的RNA转录物包含功能性帽。因此,在一些实施方案中,使用的多核苷酸(例如,DNA)模板,例如在IVT反应中,包含在模板位置+1处的2'-脱氧胸苷残基。在其他实施方案中,使用的多核苷酸(例如,DNA)模板,例如在IVT反应中,包含在模板位置+1处的2'-脱氧胞苷残基。

[0274] 应用

[0275] 使用RNA聚合酶变体产生的RNA转录物可包括mRNA(包括修饰的mRNA和/或未修饰的RNA)、lncRNA、自我复制RNA、环状RNA、CRISPR引导RNA等。在实施方案中,RNA是编码多肽(例如,治疗性多肽)的RNA(例如,mRNA或自我复制RNA)。因此,使用RNA聚合酶变体产生的RNA转录物可用于无数应用中。

[0276] 举例来说,RNA转录物可用于产生感兴趣的多肽,例如,治疗性蛋白质、疫苗抗原等。在一些实施方案中,RNA转录物是治疗性RNA。治疗性mRNA是编码治疗性蛋白质(术语‘蛋白质’涵盖肽)的mRNA。治疗性蛋白质在宿主细胞或受试者中介导各种效应以治疗疾病或改善疾病的体征和症状。举例来说,治疗性蛋白质可以替代有缺陷或异常的蛋白质,增强内源蛋白质的功能,为细胞提供新功能(例如,抑制或激活内源细胞活性,或充当另一种治疗性

化合物(例如,抗体-药物缀合物)的递送剂。治疗性mRNA可用于以下疾病和病状的治疗:细菌感染、病毒感染、寄生虫感染、细胞增殖病症、遗传病症和自身免疫病症。其他疾病和病状涵盖在本文中。

[0277] 使用RNA聚合酶变体产生的RNA转录物可以编码一种或多种生物制剂。生物制剂是基于多肽的分子,其可用于治疗、治愈、减轻、预防或诊断严重或危及生命的疾病或医学状况。生物制剂包括但不限于过敏提取物(例如用于过敏注射和测试)、血液组分、基因疗法产品、用于移植的人组织或细胞产品、疫苗、单克隆抗体、细胞因子、生长因子、酶、溶栓剂和免疫调节剂等。

[0278] 当前在销售或开发中的一种或多种生物制剂可以由RNA聚合酶变体产生的RNA编码。尽管不希望受理论的束缚,但据信将已知生物制剂的编码多核苷酸并入RNA中将会导致改善的治疗功效,这至少部分是由于构建体设计的特异性、纯度和/或选择性。

[0279] 使用RNA聚合酶变体产生的RNA转录物可以编码一种或多种抗体。术语“抗体”包括单克隆抗体(包括具有免疫球蛋白Fc区的全长抗体)、具有多表位特异性的抗体组合物、多特异性抗体(例如,双特异性抗体、双抗体和单链分子)以及抗体片段。术语“免疫球蛋白”(Ig)在本文中与“抗体”可互换使用。单克隆抗体是获自基本上同质抗体的群体的抗体,即,包含所述群体的单个抗体是同一的,除了可以微量存在的可能的天然存在的突变和/或翻译后修饰(例如,异构化、酰胺化)。单克隆抗体具有高度特异性(针对单一抗原位点)。

[0280] 单克隆抗体具体包括嵌合抗体(免疫球蛋白),以及此类抗体的片段,其中重链和/或轻链的一部分与源自特定物种或属于特定抗体类或亚类的抗体中的对应序列同一或同源,而链的其余部分与源自其他物种或属于其他抗体类或亚类的抗体中的对应序列同一或同源,只要其表现出所期望的生物学活性。嵌合抗体包括但不限于“灵长类化(primatized)”抗体,其包含源自非人灵长类动物(例如,旧世纪猴(Old World Monkey)、猿等)的可变结构域抗原结合序列和人恒定区序列。

[0281] 使用RNA聚合酶变体产生的RNA转录物可以编码一种或多种疫苗抗原。疫苗抗原是提高对特定疾病或感染剂的免疫力的生物学制剂。当前在销售或开发中的一种或多种疫苗抗原可由RNA编码。RNA中编码的疫苗抗原可用于治疗许多治疗领域的病状或疾病,诸如但不限于癌症、过敏和传染病。在一些实施方案中,癌症疫苗可以是多联体或编码肽表位的单个RNA或其组合形式的个性化癌症疫苗。

[0282] 使用RNA聚合酶变体产生的RNA转录物可被设计成编码一种或多种抗微生物肽(AMP)或抗病毒肽(AVP)。已经从广泛多种动物(诸如但不限于微生物、无脊椎动物、植物、两栖动物、鸟类、鱼类和哺乳动物)中分离和描述了AMP和AVP。

[0283] 在一些实施方案中, RNA转录物用作放射性标记的RNA探针。在一些实施方案中, RNA转录物用于非同位素RNA标记。在一些实施方案中, RNA转录物用作基因靶向的引导RNA(gRNA)。在一些实施方案中, RNA转录物(例如, mRNA)用于体外翻译和微注射。在一些实施方案中, RNA转录物用于RNA结构、加工和催化研究。在一些实施方案中, RNA转录物用于RNA扩增。在一些实施方案中, RNA转录物用作基因表达实验的反义RNA。

[0284] 野生型T7 RNA聚合酶

[0285] MNTINIAKNDFSDIELAAIPFNLTADHYGERLAREQLALEHESYEMGEARFRKMFERQLKAGEVADNA
AAKPLITTLTPKMIARINDWFEEVKAKRGKRPTAFQFLQEIKPEAVAYITIKTTLACLTSADNTTVQAVASAI GRA

IEDEARFGRIRDLEAKHFKNVEEQLNKRVGHV YKKAQMVEADMLSKGLLGGEAWSSWHKEDSIHVGVRCIEM
LIESTGMVSLHRQNAGVVGQDSETIELAPEYAEAIATRAGALAGISPMFQPCVVPPKPWTGITGGGYWANGRRPLA
LVRTHSKKALMRYEDVYMPEVYKAINIAQNTAWKINKKVLAVANVITKWKHCPVEDIPAIEREELPMKPEDIDMNP
EALTAWKRAAAAVYRKDKARKSRRISLEFMLEQANKFANHKAIWFPYNMDWRGRVYAVSMFNPQGNMTKGLLTLA
KGKPIGKEGYWLKIHGANCAGVDKVPFPERIKFIEENHENIMACAKSPLENTWWAEQDSPFCFLAFCFEYAGVQH
HGLSYNCSLPLAFDGSCSGIQHFSAMLRDEVGGRVNLNLPSETVQDIYGIVAKKVNEILQADAINGTDNEVVTVT
ENTGEISEKVKLGTKALAGQWLAYGVTRSVTKRSVMTLAYGSKEFGFRQQVLEDTIQPAIDSGKGLMFTQPNQAAG
YMAKLIWESVSVTVVAAVEAMNWLKSAKLLAAEVKDKKTGEILRKRCVHWVTPDGFVPVWQEYKKPIQTRLNLMF
LGQFRLQPTINTNKDSEIDAHKQESGIAPNFVHSQDGSHLRKTVVWAHEKYGIESFALIHDSFGTIPADAANLFKA
VRETMVDTYESCDVLADFYDQFADQLHESQLDKMPALPAKGNLNLRDILESDFABA (SEQ ID NO: 1)

[0286] 额外的实施方案

[0287] 额外的实施方案由以下带编号的段落1-71涵盖：

[0288] 1. 一种核糖核酸 (RNA) 聚合酶变体, 所述RNA聚合酶变体包含: 氨基酸序列, 相对于包含SEQ ID NO: 1的氨基酸序列的野生型T7RNA聚合酶, 所述氨基酸序列包含 (i) 位置E350处的氨基酸取代, (ii) 位置D351处的氨基酸取代, 和 (iii) 位置K387、位置N437处或位置K387和位置N437处的氨基酸取代。

[0289] 2. 如段落1所述的RNA聚合酶变体, 其中所述变体的所述氨基酸序列包含位置K387处的氨基酸取代。

[0290] 3. 如段落1所述的RNA聚合酶变体, 其中所述变体的所述氨基酸序列包含位置N437处的氨基酸取代。

[0291] 4. 如段落1所述的RNA聚合酶变体, 其中所述变体的所述氨基酸序列包含位置K387处和位置N437处的氨基酸取代。

[0292] 5. 如段落1-4中任一项所述的RNA聚合酶变体, 其中位置K387处的所述氨基酸取代是极性中性氨基酸。

[0293] 6. 如段落5所述的RNA聚合酶变体, 其中所述极性中性氨基酸选自天冬酰胺 (N)、半胱氨酸 (C)、谷氨酰胺 (Q)、蛋氨酸 (M)、丝氨酸 (S) 和苏氨酸 (T)。

[0294] 7. 如段落6所述的RNA聚合酶变体, 其中所述极性中性氨基酸是天冬酰胺 (K387N)。

[0295] 8. 如段落6所述的RNA聚合酶变体, 其中所述极性中性氨基酸是半胱氨酸 (K387C)。

[0296] 9. 如段落6所述的RNA聚合酶变体, 其中所述极性中性氨基酸是谷氨酰胺 (K387Q)。

[0297] 10. 如段落6所述的RNA聚合酶变体, 其中所述极性中性氨基酸是蛋氨酸 (K387M)。

[0298] 11. 如段落6所述的RNA聚合酶变体, 其中所述极性中性氨基酸是丝氨酸 (K387S)。

[0299] 12. 如段落6所述的RNA聚合酶变体, 其中所述极性中性氨基酸是苏氨酸 (K387T)。

[0300] 13. 如段落1-12中任一项所述的RNA聚合酶变体, 其中位置N437处的所述氨基酸取代是芳族氨基酸。

[0301] 14. 如段落13所述的RNA聚合酶变体, 其中所述芳族氨基酸选自色氨酸 (W)、酪氨酸 (Y) 和苯丙氨酸 (F)。

[0302] 15. 如段落14所述的RNA聚合酶变体, 其中所述芳族氨基酸是色氨酸 (N437W)。

[0303] 16. 如段落14所述的RNA聚合酶变体, 其中所述芳族氨基酸是酪氨酸 (N437Y)。

[0304] 17. 如段落14所述的RNA聚合酶变体, 其中所述芳族氨基酸是苯丙氨酸 (N437F)。

[0305] 18.一种核糖核酸(RNA)聚合酶变体,所述RNA聚合酶变体包含氨基酸序列,相对于包含SEQ ID NO 1的氨基酸序列的野生型T7RNA聚合酶,所述氨基酸序列包含(i)位置E350处的氨基酸取代,(ii)D351处的氨基酸取代,和(iii)位置D653处的氨基酸取代。

[0306] 19.如段落18所述的RNA聚合酶变体,其中位置D653处的所述氨基酸取代是芳族氨基酸。

[0307] 20.如段落19所述的RNA聚合酶变体,其中所述芳族氨基酸选自色氨酸(W)、酪氨酸(Y)和苯丙氨酸(F)。

[0308] 21.如段落20所述的RNA聚合酶变体,其中所述芳族氨基酸是色氨酸(D653W)。

[0309] 22.如段落20所述的RNA聚合酶变体,其中所述芳族氨基酸是酪氨酸(D653Y)。

[0310] 23.如段落20所述的RNA聚合酶变体,其中所述芳族氨基酸是苯丙氨酸(D653F)。

[0311] 24.如段落1-23中任一项所述的RNA聚合酶变体,其中位置E350处的所述氨基酸取代是芳族氨基酸。

[0312] 25.如段落24所述的RNA聚合酶变体,其中所述芳族氨基酸选自色氨酸(W)、酪氨酸(Y)和苯丙氨酸(F)。

[0313] 26.如段落25所述的RNA聚合酶变体,其中所述芳族氨基酸是色氨酸(E350W)。

[0314] 27.如段落25所述的RNA聚合酶变体,其中所述芳族氨基酸是酪氨酸(E350Y)。

[0315] 28.如段落25所述的RNA聚合酶变体,其中所述芳族氨基酸是苯丙氨酸(E350F)。

[0316] 29.如段落1-28所述的RNA聚合酶变体,其中位置D351处的所述氨基酸取代是非极性脂肪族氨基酸。

[0317] 30.如段落29所述的RNA聚合酶变体,其中所述非极性脂肪族氨基酸选自丙氨酸(A)、甘氨酸(G)、异亮氨酸(I)、亮氨酸(L)、脯氨酸(P)和缬氨酸(V)。

[0318] 31.如段落30所述的RNA聚合酶变体,其中所述非极性脂肪族氨基酸是丙氨酸(D351A)。

[0319] 32.如段落30所述的RNA聚合酶变体,其中所述非极性脂肪族氨基酸是甘氨酸(D351G)。

[0320] 33.如段落30所述的RNA聚合酶变体,其中所述非极性脂肪族氨基酸是异亮氨酸(D351I)。

[0321] 34.如段落30所述的RNA聚合酶变体,其中所述非极性脂肪族氨基酸是亮氨酸(D351L)。

[0322] 35.如段落30所述的RNA聚合酶变体,其中所述非极性脂肪族氨基酸是脯氨酸(D351P)。

[0323] 36.如段落30所述的RNA聚合酶变体,其中所述非极性脂肪族氨基酸是缬氨酸(D351V)。

[0324] 37.一种RNA聚合酶变体,所述RNA聚合酶变体包含:与SEQ ID NO:1的氨基酸序列具有至少70%同一性的氨基酸序列,其中相对于包含SEQ ID NO:1的氨基酸序列的野生型T7 RNA聚合酶,所述变体的所述氨基酸序列包含(i)位置E350处的氨基酸取代,(ii)D351处的氨基酸取代,和(iii)位置K387、位置N437处或位置K387和位置N437处的氨基酸取代。

[0325] 38.如段落37所述的RNA聚合酶变体,其中所述氨基酸序列与SEQ ID NO:1的氨基酸序列具有至少75%、至少80%、至少85%、至少95%或至少98%同一性。

[0326] 39.如段落37或38所述的RNA聚合酶变体,其中所述变体的所述氨基酸序列包含位置K387处的氨基酸取代。

[0327] 40.如段落37或38所述的RNA聚合酶变体,其中所述变体的所述氨基酸序列包含位置N437处的氨基酸取代。

[0328] 41.如段落37或38所述的RNA聚合酶变体,其中所述变体的所述氨基酸序列包含位置K387处和位置N437处的氨基酸取代。

[0329] 42.如段落37-41中任一项所述的RNA聚合酶变体,其中位置K387处的所述氨基酸取代是极性中性氨基酸。

[0330] 43.如段落42所述的RNA聚合酶变体,其中所述极性中性氨基酸选自天冬酰胺(K387N)、半胱氨酸(K387C)、谷氨酰胺(K387Q)、蛋氨酸(K387M)、丝氨酸(K387S)和苏氨酸(K387T)。

[0331] 44.如段落37-42中任一项所述的RNA聚合酶变体,其中位置N437处的所述氨基酸取代是芳族氨基酸。

[0332] 45.如段落44所述的RNA聚合酶变体,其中所述芳族氨基酸选自色氨酸(N437W)、酪氨酸(N437Y)和苯丙氨酸(N437F)。

[0333] 46.一种RNA聚合酶变体,所述RNA聚合酶变体包含:与SEQ ID NO:1的氨基酸序列具有至少70%同一性的氨基酸序列,其中相对于包含SEQ ID NO:1的氨基酸序列的野生型T7 RNA聚合酶,所述变体的所述氨基酸序列包含(i)位置E350处的氨基酸取代,(ii)D351处的氨基酸取代,和(iii)位置D653处的氨基酸取代。

[0334] 47.如段落46所述的RNA聚合酶变体,其中所述氨基酸序列与SEQ ID NO:1的氨基酸序列具有至少75%、至少80%、至少85%、至少95%或至少98%同一性。

[0335] 48.如段落46或47所述的RNA聚合酶变体,其中位置D653处的所述氨基酸取代是芳族氨基酸。

[0336] 49.如段落48所述的RNA聚合酶变体,其中所述芳族氨基酸选自色氨酸(D653W)、酪氨酸(D653Y)和苯丙氨酸(D653F)。

[0337] 50.如段落37-49中任一项所述的RNA聚合酶变体,其中位置E350处的所述氨基酸取代是芳族氨基酸。

[0338] 51.如段落50所述的RNA聚合酶变体,其中所述芳族氨基酸选自色氨酸(E350W)、酪氨酸(E350Y)和苯丙氨酸(E350F)。

[0339] 52.如段落37-51中任一项所述的RNA聚合酶变体,其中位置D351处的所述氨基酸取代是非极性脂肪族氨基酸。

[0340] 53.如段落52所述的RNA聚合酶变体,其中所述非极性脂肪族氨基酸选自丙氨酸(D351A)、甘氨酸(D351G)、异亮氨酸(D351I)、亮氨酸(D351L)、脯氨酸(D351P)和缬氨酸(D351V)。

[0341] 54.一种核糖核酸(RNA)聚合酶变体,所述RNA聚合酶变体包含SEQ ID NO:2的氨基酸序列,其中 X^1 是芳族氨基酸,任选地选自W、Y、和F; X^2 选自非极性脂肪族氨基酸,任选地选自A、G、I、L、P和V; X^3 是极性中性氨基酸,任选地选自N、C、Q、M、S和T;并且 X^4 是芳族氨基酸,任选地选自W、Y和F。

[0342] 55.一种核糖核酸(RNA)聚合酶变体,所述RNA聚合酶变体包含SEQ ID NO:6的氨基

酸序列。

[0343] 56. 一种核糖核酸 (RNA) 聚合酶变体, 所述RNA聚合酶变体包含SEQ ID NO:3的氨基酸序列, 其中 X^1 是芳族氨基酸, 任选地选自W、Y、和F; X^2 选自非极性脂肪族氨基酸, 任选地选自A、G、I、L、P和V; 并且 X^4 是芳族氨基酸, 任选地选自W、Y和F。

[0344] 57. 一种核糖核酸 (RNA) 聚合酶变体, 所述RNA聚合酶变体包含SEQ ID NO:7的氨基酸序列。

[0345] 58. 一种核糖核酸 (RNA) 聚合酶变体, 所述RNA聚合酶变体包含SEQ ID NO:4的氨基酸序列, 其中 X^1 是芳族氨基酸, 任选地选自W、Y、和F; X^2 选自非极性脂肪族氨基酸, 任选地选自A、G、I、L、P和V; 并且 X^3 是极性中性氨基酸, 任选地选自N、C、Q、M、S和T。

[0346] 59. 一种核糖核酸 (RNA) 聚合酶变体, 所述RNA聚合酶变体包含SEQ ID NO:8的氨基酸序列。

[0347] 60. 一种核糖核酸 (RNA) 聚合酶变体, 所述RNA聚合酶变体包含SEQ ID NO:5的氨基酸序列, 其中 X^1 是芳族氨基酸, 任选地选自W、Y、和F; X^2 选自非极性脂肪族氨基酸, 任选地选自A、G、I、L、P和V; 并且 X^5 是芳族氨基酸, 任选地选自W、Y和F。

[0348] 61. 一种核糖核酸 (RNA) 聚合酶变体, 所述RNA聚合酶变体包含SEQ ID NO:9的氨基酸序列。

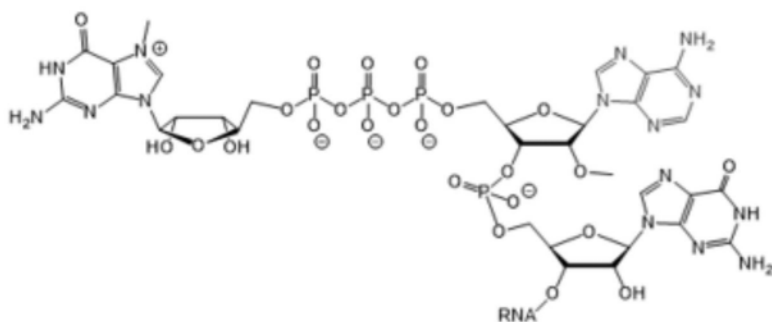
[0349] 62. 一种方法, 所述方法包括: 在体外转录反应中产生信使RNA (mRNA), 所述转录反应包括DNA、核苷三磷酸、如段落1-53中任一项所述的RNA聚合酶变体和任选的帽类似物。

[0350] 63. 如段落62所述的方法, 其中所述反应包括帽类似物。

[0351] 64. 如段落62或63所述的方法, 其中所述帽类似物是二核苷酸帽类似物、三核苷酸帽类似物或四核苷酸帽类似物。

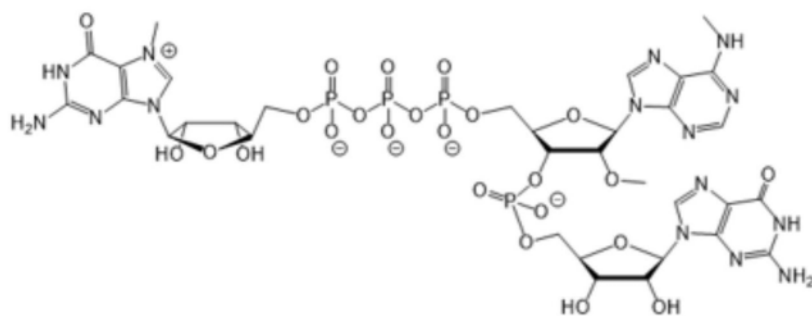
[0352] 65. 如段落64所述的方法, 其中所述帽类似物是包含GAG序列的三核苷酸帽类似物。

[0353] 66. 如段落65所述的方法, 其中所述GAG帽类似物包含选自以下的化合物:

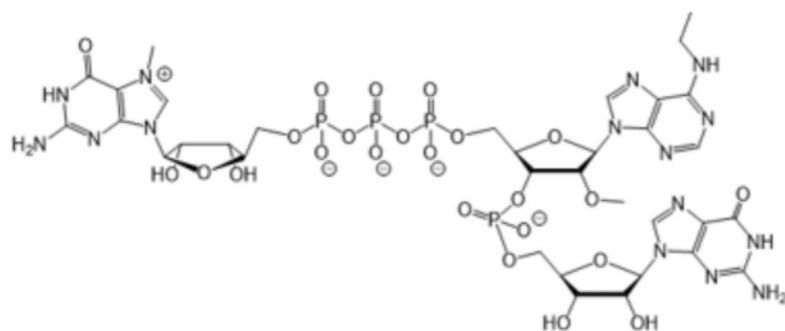


(X);

[0354]



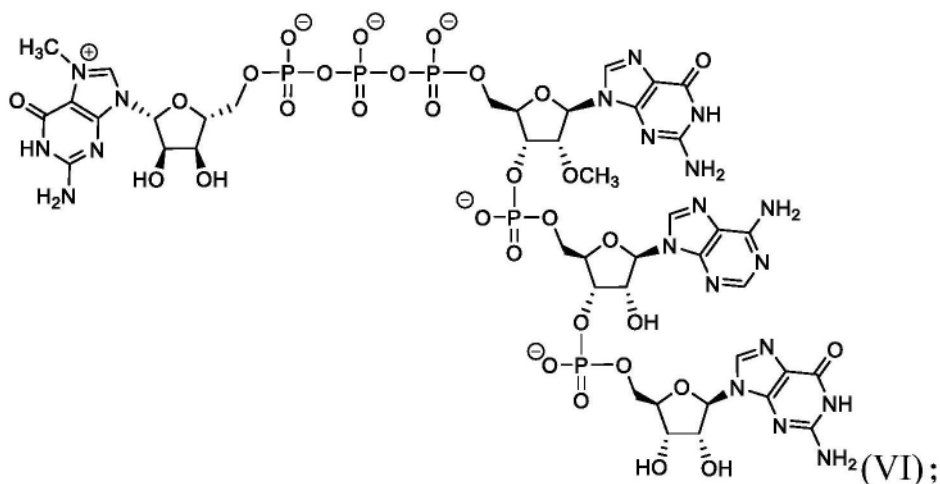
(XI); 或



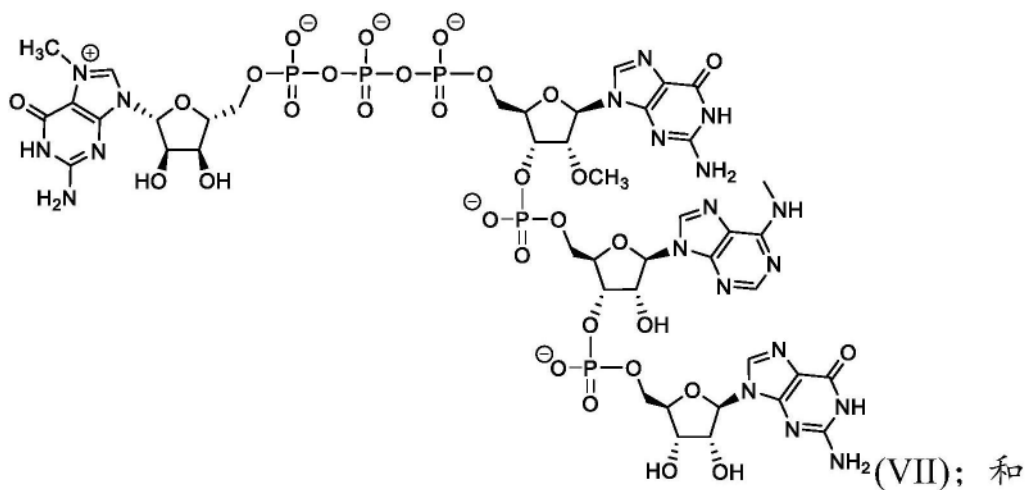
(XII)。

[0355] 67. 如段落64所述的方法,其中所述四核苷酸帽类似物包含GGAG序列。

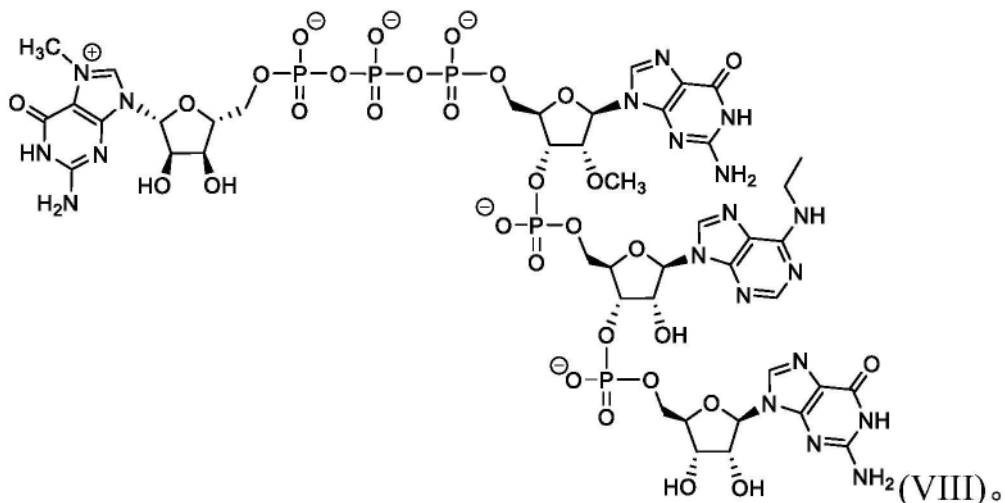
[0356] 68. 如段落67所述的方法,其中所述四核苷酸帽类似物包含选自以下的化合物:



[0357]



[0358]



[0359] 69. 如段落62-68中任一项所述的方法,其中所述DNA包含位置+1处的2'-脱氧胸苷残基或2'-脱氧胞苷残基。

[0360] 70. 一种组合物或试剂盒,所述组合物或试剂盒包括如段落1-61中任一项所述的RNA聚合酶变体和选自由DNA、核苷三磷酸和帽类似物组成的组的体外转录 (IVT) 试剂。

[0361] 71. 一种核酸,所述核酸编码如段落1-61中任一项所述的RNA聚合酶变体。

[0362] 实施例

[0363] 实施例1.使用RNA聚合酶变体的IVT反应

[0364] 使用DNA模板、GGAG帽类似物和如表1中所提供的选择的单个RNA聚合酶变体进行体外转录 (IVT) 反应。具体地,在本实施例中测试了包含N437Y、K387S、E350W和D351V取代 (SEQ ID NO:6;“变体A”);N437Y、E350W和D351V取代 (SEQ ID NO:7;“变体B”);K387S、E350W和D351V取代 (SEQ ID NO:8;“变体C”);以及D653W、E350W和D351V取代 (SEQ ID NO:9;“变体D”)的RNA聚合酶变体。还使用对照T7 RNA聚合酶 (SEQ ID NO:1) 进行反应。在IVT反应之后,对来自每个反应的转录RNA产物进行表征,以解决所述RNA产物的质量,包括总RNA产率、加帽效率 (包含GGAG帽的总RNA的百分比)、dsRNA污染和尾纯度。

[0365] 寡dT纯化后,通过UV吸收测量总RNA的总产率。通过LC-MS分析总RNA产物以确定加帽效率 (即,包含GGAG帽的转录RNA的百分比)。在本实施例中,在IVT反应之后使用标准ELISA来评估dsRNA污染物 (例如,长于40个核苷酸碱基对的dsRNA)。使用Tris RP (反相) 方法来评估加尾RNA的百分比 (即,包含多聚A尾的转录RNA的百分比)。

[0366] 每个测试的RNA聚合酶变体在IVT反应中产生具有至少80%加帽RNA (包含GGAG帽的总RNA的百分比) 和至少约80%加尾RNA (即,包含多聚A尾的转录RNA的百分比) 的RNA。包含N437Y、E350W和D351V取代 (SEQ ID NO:7);以及K387S、E350W和D351V取代 (SEQ ID NO:8) 的RNA聚合酶变体产生小于0.007%dsRNA (重量:重量)。此外,每个测试的RNA聚合酶变体的总RNA产率 (大于8mg/mL) 与对照T7 RNA聚合酶相当。

[0367] 在每个测试特性上,每个测试的RNA聚合酶变体与对照T7 RNA聚合酶表现相当或表现比其更好。具体地,相对于对照T7 RNA聚合酶,N437Y+K387S+E350W+D351V提供了具有更高加帽效率 (约85%加帽RNA)、类似产率和类似加尾纯度的RNA。相对于对照T7RNA聚合酶,N437Y+E350W+D351V提供了具有更高的加帽效率 (约80%加帽RNA)、类似的产率、类似的加尾纯度和类似的dsRNA污染的RNA。相对于对照T7 RNA聚合酶,K387S+E350W+D351V提供了具有更高的加帽效率 (约83%加帽RNA)、类似的产率、更高的加尾纯度 (约85%加尾RNA) 和更少的dsRNA污染 (0.00327dsRNA重量:重量) 的RNA。相对于对照T7 RNA聚合酶,D653W+E350W+D351V提供了具有更高的加帽效率 (约95%加帽RNA) 的RNA。

[0368] 每个测试的RNA聚合酶的数据提供于表3和图1A-1D中。

[0369] 表3.由实施例1中使用的RNA聚合酶产生的RNA的特性

[0370]		产率(mg/ml)	加帽百分比	加尾百分比	dsRNA 百分比 (重量:重量)
	对照 RNAP	9.34	69.51	83.17	0.00624
	N437Y + K387S + E350W + D351V	11.14	84.95	79.51	0.01583
	N437Y + E350W + D351V	9.56	80.19	80.87	0.006955
	K387S + E350W + D351V	9.20	82.88	85.00	0.00327
	D653W + E350W + D351V	8.87	94.47	80.30	0.02815

[0371] 实施例2.RNA聚合酶变体在低浓度的GGAG帽类似物下产生具有高水平的加帽效率的RNA产物

[0372] 使用DNA模板、等摩尔NTP、可变量的GGAG四核苷酸帽类似物(0.25mM、0.5mM、0.75mM、1mM、1.25mM、1.5mM、3mM)和T7 RNA聚合酶进行体外转录反应。在本实施例中测试了包含N437Y、K387S、E350W和D351V取代(SEQ ID NO:6);N437Y、E350W和D351V取代(SEQ ID NO:7);K387S、E350W和D351V取代(SEQ ID NO:8);以及D653W、E350W和D351V取代(SEQ ID NO:9)的RNA聚合酶变体。还使用对照T7 RNA聚合酶进行反应。

[0373] 在IVT反应之后,mRNA产物经过寡dT纯化,然后通过LC-MS进行分析以确定%加帽RNA(即,包含帽的转录RNA的百分比),通过HPLC以确定反应的RNA产率,并通过Tris RP(反相)方法以确定加尾RNA百分比。

[0374] 在GGAG帽类似物的存在下,每个测试的RNA聚合酶变体比对照聚合酶变体产生具有更高水平的加帽RNA百分比的RNA,无论GGAG类似物的浓度如何(图2A)。即使在最低测试浓度的GGAG帽类似物(0.25mM)下,所有测试的变体产生至少50%加帽RNA,远高于由对照聚合酶变体产生的约25%加帽RNA。在1.5mM GGAG帽类似物下,所有测试的变体产生约80-95%加帽RNA。

[0375] 相对于对照聚合酶变体,每个测试的变体产生具有相当的产率(图2B)和加尾RNA百分比(图2C)的RNA。

[0376] 这些数据证明每个测试的RNA聚合酶变体(即,N437Y+K387S+E350W+D351V;N437Y+E350W+D351V;K387S+E350W+D351V;和D653W+E350W+D351V)能够产生比对照T7 RNA聚合酶具有更高的加帽效率的RNA,而不放弃任何产率或加尾含量。

[0377] 等效物和范围

[0378] 本文公开的所有参考文献、专利和专利申请均以引用的方式并入,就其各自所引用的主题而言,其在一些情况下可涵盖整个文献。

[0379] 在说明书和权利要求中使用的不定冠词“一个”和“一种”,除非明确表明相反,应理解为意指“至少一个/一种”。

[0380] 还应当理解,除非明确表明相反,在本文要求保护的包括一个以上步骤或动作的任何方法中,所述方法的步骤或动作的顺序不一定限于引用所述方法的步骤或动作的顺序。

[0381] 在权利要求以及以上说明书中,所有过渡性短语诸如“包含”、“包括”、“携带”、“具有”、“含有”、“涉及”、“持有”、“由…构成”等应理解为开放式的,即意指包括但不限于。仅过渡性短语“由…组成”和“基本上由…组成”应分别是如美国专利局专利审查程序手册第2111.03节所述的封闭式或半封闭式过渡性短语。

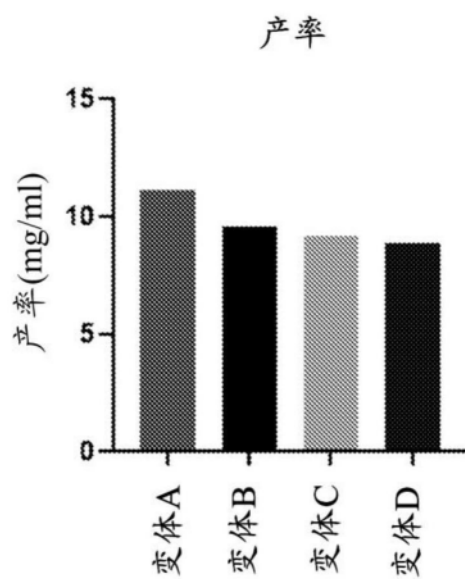


图1A

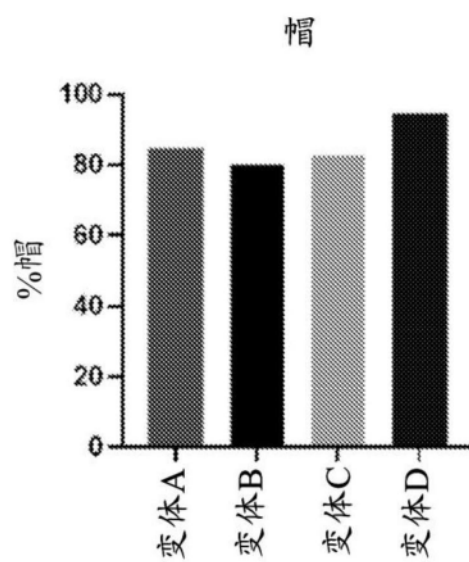


图1B

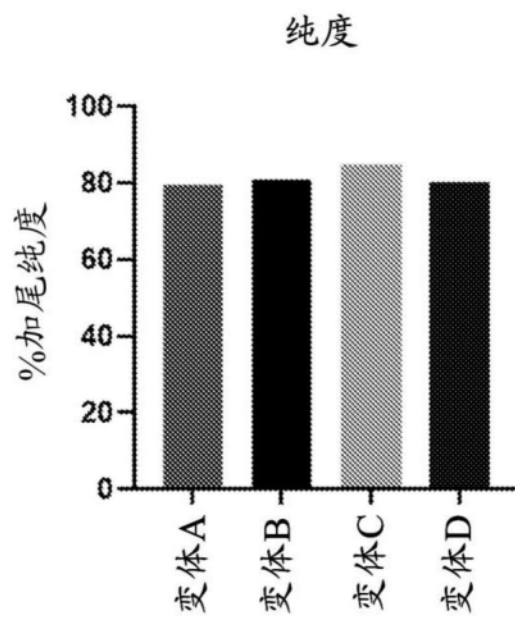


图1C

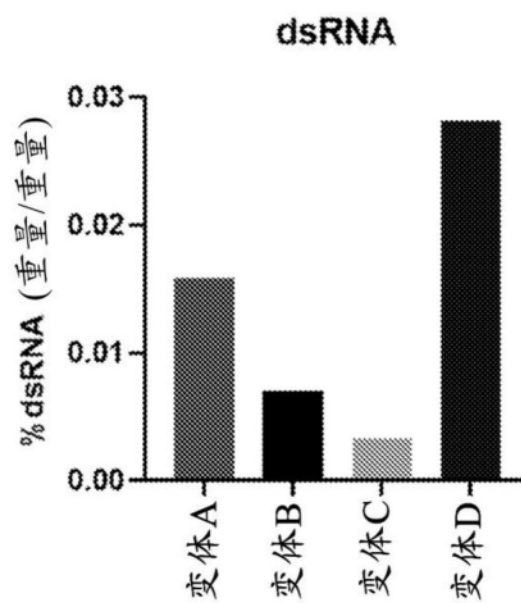


图1D

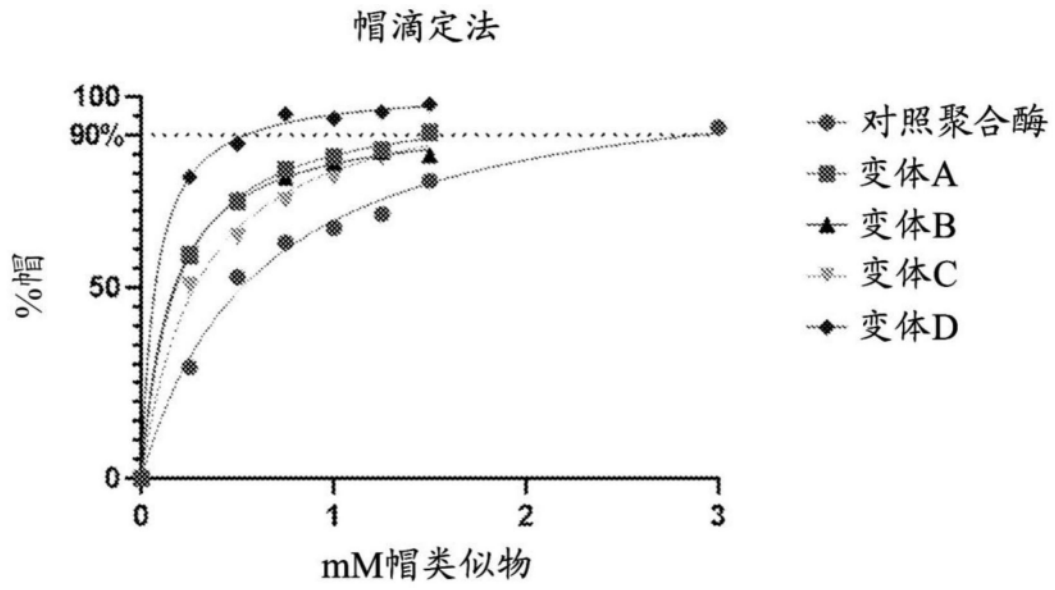


图2A

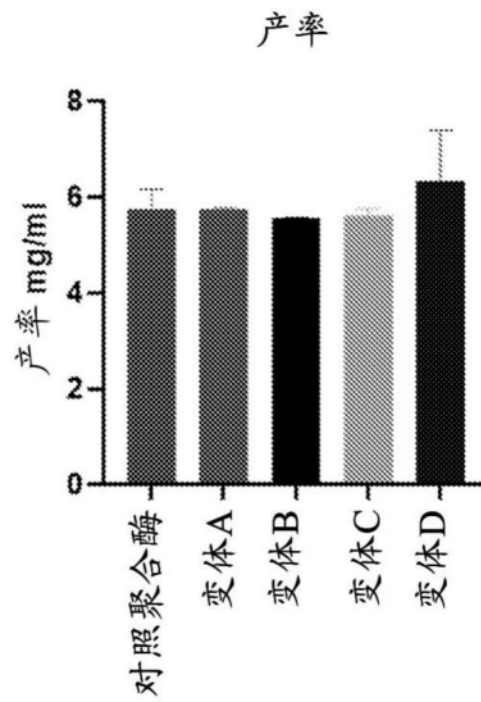


图2B

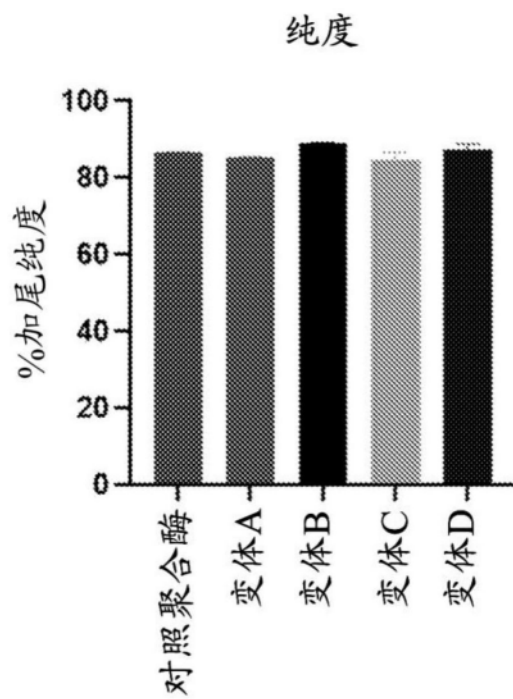


图2C