(19)国家知识产权局

(12)发明专利申请

(10)申请公布号 CN 116438308 A
(43)申请公布日 2023.07.14

(21)申请号 202180056786.7
(22)申请日 2021.08.06
(30)优先权数据
63/062,268 2020.08.06 US
(85)PCT国际申请进入国家阶段日
2023.02.03
(86)PCT国际申请的申请数据
PCT/US2021/045106 2021.08.06
(87)PCT国际申请的公布数据
WO2022/032196 EN 2022.02.10
(71)申请人 磨石生物公司
地址 美国加利福尼亚州
(72)发明人 K·朱斯 R·耶冷斯凯 J·X·孙 A·R·拉帕波特 C·D·斯卡伦 L·吉特林 C·D·帕尔默 M·拉内
(74)专利代理机构 北京市金杜律师事务所 11256
专利代理师 陈文平 赵磊
(51)Int.Cl.
C12N 15/12(2006.01)
C07K 14/74(2006.01)
C12N 15/861(2006.01)
C12N 15/86(2006.01)
C12N 5/10(2006.01)
A61K 48/00(2006.01)
A61K 39/00(2006.01)
A61K 45/06(2006.01)
A61P 35/00(2006.01)
A61P 35/02(2006.01)
A61P 31/00(2006.01)

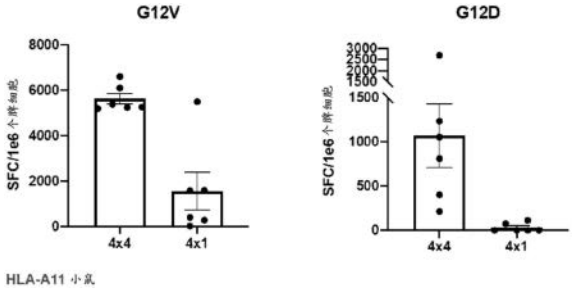
权利要求书28页 说明书98页
序列表226页 附图14页

(54)发明名称

多表位疫苗盒

(57)摘要

本文公开了包含抗原编码核酸序列的组合物,所述抗原编码核酸序列具有KRAS新表位编码序列的多次迭代和/或缺乏免疫显性表位。还公开了与所述组合物相关的核苷酸、细胞和方法,包括它们作为疫苗的用途。



CN 116438308 A

1.一种抗原编码盒或由所述盒编码的多肽序列,其中所述抗原编码盒包含至少一种由下式从5'至3'描述的抗原编码核酸序列:

$(E_x-(E^N_n)_y)_z$

其中E表示包含不同表位编码核酸序列的核苷酸序列,

n表示单独的不同表位编码核酸序列的数目并且是包括0的任何整数,

E^N表示包含对于每个相应n的单独的不同表位编码核酸序列的核苷酸序列,

对于z的每次迭代:x=0或1,对于每个n,y=0或1,和x或y中的至少一者=1,并且

z=2或更大,其中所述抗原编码核酸序列包含E、给定E^N或其组合的至少两次迭代,并且

包含所述至少两次迭代的所述不同表位编码核酸序列中的至少一者编码KRAS相关MHC I类新表位。

2.如权利要求1所述的组合物,其中所述抗原编码盒编码氨基酸序列VVVGACGVGK(SEQ ID NO:75)、VVVGADGVGK(SEQ IDNO:78)、VVGAVGVGK(SEQ ID NO:79)和ILDTAGHEEY(SEQ IDNO:82)中的每一者的至少4次迭代。

3.如权利要求1所述的组合物,其中包含所述至少两次迭代的所述不同表位编码核酸序列中的至少2个、至少3个、至少4个、至少5个、至少6个、至少7个或至少8个编码不同的KRAS相关MHC I类新表位。

4.如权利要求1所述的组合物,其中包含所述至少两次迭代的所述不同表位编码核酸序列中的每一者编码不同的KRAS相关MHC I类新表位。

5.如权利要求1-4中任一项所述的组合物,其中编码所述KRAS相关MHC I类新表位的所述表位编码核酸序列中的一者或多者包含至少2次、至少3次、至少4次、至少5次、至少6次、至少7次或至少8次迭代。

6.如权利要求1-4中任一项所述的组合物,其中编码所述KRAS相关MHC I类新表位的所述核酸序列中的每一者包含至少2次、至少3次、至少4次、至少5次、至少6次、至少7次或至少8次迭代。

7.如权利要求1-4中任一项所述的组合物,其中编码所述不同KRAS相关MHC I类新表位的所述核酸序列中的一者或多者包含至少4次迭代。

8.如权利要求1-4中任一项所述的组合物,其中编码所述不同KRAS相关MHC I类新表位的所述核酸序列中的每一者包含至少4次迭代。

9.如权利要求1-8中任一项所述的组合物,其中所述不同KRAS相关MHC I类新表位中的一者或多者独立地包含KRAS G12C突变、KRAS G12V突变、KRAS G12D突变或KRAS Q61H突变。

10.如权利要求1-9中任一项所述的组合物,其中每个E或E^N独立地包含由式$(L5_b-N_c-L3_d)$从5'至3'描述的核苷酸序列,

其中N包含与每个E或E^N相关的所述不同表位编码核酸序列,其中c=1,

L5包含5'接头序列,其中b=0或1,并且

L3包含3'接头序列,其中d=0或1。

11.如权利要求10所述的组合物,其中

每个N编码长度为7-15个氨基酸的表位,

L5是编码所述表位的原生N末端氨基酸序列的原生5'接头序列,并且其中所述5'接头序列编码长度为至少2个氨基酸的肽,并且

L3是编码所述表位的原生C末端氨基酸序列的原生3'接头序列,并且其中所述3'接头序列编码长度为至少2个氨基酸的肽。

12.如权利要求1-11中任一项所述的组合物,其中每个E和E^N编码长度为至少7个氨基酸的表位。

13.如权利要求1-11中任一项所述的组合物,其中每个E和E^N编码长度为7-15个氨基酸的表位。

14.如权利要求1-13中任一项所述的组合物,其中每个E和E^N是长度为至少21个核苷酸的核苷酸序列。

15.如权利要求1-13中任一项所述的组合物,其中每个E和E^N是长度为75个核苷酸的核苷酸序列。

16.一种用于递送抗原表达系统的组合物,所述组合物包含:

所述抗原表达系统,其中所述抗原表达系统包含一种或多种载体,

所述一种或多种载体包含:

(a)载体骨架,其中所述骨架包含:

(i)至少一种启动子核苷酸序列,和

(ii)任选地,至少一种聚腺苷酸化(聚(A))序列;和

(b)盒,其中所述盒包含:

(i)至少一种抗原编码核酸序列,所述抗原编码核酸序列包含:

(I)编码KRAS相关MHC I类新表位的表位编码核酸序列,并且

其中所述表位编码核酸序列中的每一者包含:

(A)任选地,5'接头序列,和

(B)任选地,3'接头序列;

(ii)任选地,可操作地连接到所述抗原编码核酸序列的第二启动子核苷酸序列;和

(iii)任选地,至少一种MHC II类表位编码核酸序列;

(iv)任选地,至少一种编码GPGPG氨基酸接头序列(SEQ IDNO:56)的核酸序列;和

(v)任选地,至少一种第二聚(A)序列,其中所述第二聚(A)序列是所述载体骨架的原生聚(A)序列或外源聚(A)序列,

其中如果所述第二启动子核苷酸序列不存在,则所述抗原编码核酸序列可操作地连接到所述至少一种启动子核苷酸序列,并且

其中所述至少一种抗原编码核酸序列包含编码所述KRAS相关MHC I类新表位的所述表位编码核酸序列的至少两次迭代。

17.一种用于递送抗原表达系统的组合物,所述组合物包含:

所述抗原表达系统,其中所述抗原表达系统包含一种或多种载体,

所述一种或多种载体包含:

(a)载体骨架,其中所述骨架包含:

(i)至少一种启动子核苷酸序列,和

(ii)至少一种聚腺苷酸化(聚(A))序列;和

(b)盒,其中所述盒包含:

(i)至少一种抗原编码核酸序列,所述抗原编码核酸序列包含:

(I)彼此线性连接的至少2、3、4、5、6、7、8、9或10种不同表位编码核酸序列,其中所述不同表位编码核酸序列中的至少一者编码KRAS相关MHC I类新表位,并且

其中所述表位编码核酸序列中的每一者包含:

(A)任选地,5'接头序列,和

(B)任选地,3'接头序列;

(ii)任选地,可操作地连接到所述抗原编码核酸序列的第二启动子核苷酸序列;

(iii)任选地,至少一种MHC II类表位编码核酸序列;

(iv)任选地,至少一种编码GPGPG氨基酸接头序列(SEQ IDNO:56)的核酸序列;和

(v)任选地,至少一种第二聚(A)序列,其中所述第二聚(A)序列是所述载体骨架的原生聚(A)序列或外源聚(A)序列,

其中如果所述第二启动子核苷酸序列不存在,则所述抗原编码核酸序列可操作地连接到所述至少一种启动子核苷酸序列,并且

其中所述至少一种抗原编码核酸序列包含编码所述KRAS相关MHC I类新表位的所述不同表位编码核酸序列中的至少一者的至少两条重复序列。

18.如权利要求17所述的组合物,其中所述至少一种抗原编码核酸序列包含至少两种不同表位编码核酸序列,其各自编码不同的KRAS相关MHC I类新表位。

19.一种用于递送抗原表达系统的组合物,所述组合物包含:

所述抗原表达系统,

其中所述抗原表达系统包含一种或多种载体,

所述一种或多种载体包含:

(a)载体骨架,其中所述载体骨架包含黑猩猩腺病毒载体,任选地其中所述黑猩猩腺病毒载体是ChAdV68载体,或甲病毒载体,任选地其中所述甲病毒载体是委内瑞拉马脑炎病毒载体;和

(b)盒,任选地其中所述盒整合在所述载体骨架原生的原生启动子核苷酸序列和聚(A)序列之间,任选地其中所述聚(A)序列是所述载体骨架原生的,其中所述盒包含:

(i)至少一种抗原编码核酸序列,所述抗原编码核酸序列包含:

(I)编码KRAS相关MHC I类新表位的表位编码核酸序列,任选地包含彼此线性连接的至少两种不同表位编码核酸序列,每个表位编码核酸序列任选地包含:

(A)MHC I类表位编码核酸序列,其中所述MHC I类表位编码核酸序列编码长度为7-15个氨基酸的MHC I类表位,

(B)5'接头序列,其中所述5'接头序列编码所述MHC I类表位的原生N末端氨基酸序列,并且其中所述5'接头序列编码长度为至少2个氨基酸的肽,

(C)3'接头序列,其中所述3'接头序列编码所述MHC I类表位的原生C末端酸序列,并且其中所述3'接头序列编码长度为至少2个氨基酸的肽,并且

其中所述盒可操作地连接到所述原生启动子核苷酸序列,其中所述表位编码核酸序列中的每一者编码长度为13至25个氨基酸的多肽,并且其中每个表位编码核酸序列的每个3'端连接到以下表位编码核酸序列的5'端,所述盒中的最终表位编码核酸序列除外;和

(ii)至少两种MHC II类表位编码核酸序列,所述MHC II类表位编码核酸序列包含:

(I)PADRE MHC II类序列(SEQ ID NO:48),

(II)破伤风类毒素MHC II类序列(SEQ ID NO:46)，

(III)编码连接所述PADRE MHC II类序列与所述破伤风类毒素MHC II类序列的GPGPG氨基酸接头序列(SEQ ID NO:56)的第一核酸序列，

(IV)编码连接所述至少两种MHC II类表位编码核酸序列的5'端与所述表位编码核酸序列的GPGPG氨基酸接头序列(SEQ ID NO:56)的第二核酸序列，

(V)任选地，在所述至少两种MHC II类表位编码核酸序列的3'端处编码GPGPG氨基酸接头序列(SEQ ID NO:56)的第三核酸序列；

(iii)任选地，可操作地连接到所述抗原编码核酸序列的第二启动子核苷酸序列；并且

其中如果所述第二启动子核苷酸序列不存在，则所述抗原编码核酸序列可操作地连接到所述原生启动子核苷酸序列，并且

其中所述至少一种抗原编码核酸序列包含编码所述KRAS相关MHC I类新表位的所述表位编码核酸序列的至少两次迭代。

20.如权利要求16-18中任一项所述的组合物，其中所述盒的每个元件的有序序列描述于下式，从5'至3'包含：

P_a-($L5_b$-N_c-$L3_d$)$_X$-($G5_e$-U_f)$_Y$-$G3_g$

其中，

P包含所述第二启动子核苷酸序列，其中a＝0或1，

N包含所述不同表位编码核酸序列之一，其中c＝1，

L5包含所述5'接头序列，其中b＝0或1，

L3包含所述3'接头序列，其中d＝0或1，

G5包含所述至少一种编码GPGPG氨基酸接头(SEQ ID NO:56)的核酸序列之一，其中e＝0或1，

G3包含所述至少一种编码GPGPG氨基酸接头(SEQ ID NO:56)的核酸序列之一，其中g＝0或1，

U包含所述至少一种MHC II类表位编码核酸序列之一，其中f＝1，

X＝1至400，其中对于每个X，相应的N_c是表位编码核酸序列，并且

Y＝0、1或2，其中对于每个Y，相应的U_f是MHC II类表位编码核酸序列。

21.如权利要求20所述的组合物，其中对于每个X，相应的N_c是不同表位编码核酸序列，除了对应于所述不同表位编码核酸序列的至少两条重复序列的N_c。

22.如权利要求20或21所述的组合物，其中对于每个Y，所述相应的U_f是不同的MHC II类表位编码核酸序列。

23.如权利要求20-22中任一项所述的组合物，其中

a＝0，b＝1，d＝1，e＝1，g＝1，h＝1，X＝16，Y＝2，

所述至少一种启动子核苷酸序列是所述载体骨架原生的单个原生启动子核苷酸序列，

所述至少一种聚腺苷酸化聚(A)序列是由所述载体骨架提供的至少80个连续A核苷酸的聚(A)序列，

每个N编码长度为7-15个氨基酸的表位，

L5是编码所述表位的原生N末端氨基酸序列的原生5'接头序列，并且其中所述5'接头序列编码长度为至少2个氨基酸的肽，

L3是编码所述表位的原生C末端氨基酸序列的原生3'接头序列，并且其中所述3'接头序列编码长度为至少2个氨基酸的肽，

U是PADRE II类序列和破伤风类毒素MHC II类序列中的每一者，

所述载体骨架包含黑猩猩腺病毒载体，任选地其中所述黑猩猩腺病毒载体是ChAdV68载体，或甲病毒载体，任选地其中所述甲病毒载体是委内瑞拉马脑炎病毒载体，任选地其中当所述载体骨架包含甲病毒载体时，所述原生启动子核苷酸序列是亚基因组启动子，并且

所述MHC II类表位编码核酸序列中的每一者编码长度为13至25个氨基酸的多肽。

24.如权利要求16-23中任一项所述的组合物，其中所述至少两次迭代是至少3次、至少4次、至少5次、至少6次、至少7次或至少8次迭代。

25.如权利要求16-23中任一项所述的组合物，其中所述至少两次迭代是至少至少9次、至少10次、至少11次、至少12次、至少13次、至少4次、至少15次、至少16次、至少17次、至少18次、至少19次或至少20次迭代。

26.如权利要求16-23中任一项所述的组合物，其中所述至少两次迭代是2-3、2-4、2-5、2-6、2-7条重复序列或2-8次迭代。

27.如权利要求16-23中任一项所述的组合物，其中所述至少两次迭代是7次迭代或更少、6次迭代或更少、5次迭代或更少、4次迭代或更少、或3次迭代或更少。

28.如权利要求16-27中任一项所述的组合物，其中所述至少一种抗原编码核酸序列包含至少两种不同表位编码核酸序列的至少两次迭代。

29.如权利要求16-27中任一项所述的组合物，其中所述至少一种抗原编码核酸序列包含至少3种、至少4种、至少5种、至少6种、至少7种、至少8种、至少9种或至少10种不同表位编码核酸序列的至少两次迭代。

30.如权利要求16-29中任一项所述的组合物，其中所述至少两次迭代被至少一个单独的不同表位编码核酸序列分开。

31.如权利要求16-29中任一项所述的组合物，其中所述至少两次迭代被至少2个单独的不同表位编码核酸序列分开。

32.如权利要求16-29中任一项所述的组合物，其中包括所述任选的5'接头序列和/或所述任选的3'接头序列在内的所述至少两次迭代被至少75个核苷酸分开。

33.如权利要求16-29中任一项所述的组合物，其中包括所述任选的5'接头序列和/或所述任选的3'接头序列在内的所述至少两次迭代被至少150个核苷酸、至少300个核苷酸或至少675个核苷酸分开。

34.如权利要求16-29中任一项所述的组合物，其中包括所述任选的5'接头序列和/或所述任选的3'接头序列在内的所述至少两次迭代被至少50个核苷酸、至少100个核苷酸、至少200个核苷酸、至少250个核苷酸、至少350个核苷酸、至少400个核苷酸、至少450个核苷酸、至少500个核苷酸、至少700个核苷酸、至少700个核苷酸、至少750个核苷酸、至少800个核苷酸、至少900个核苷酸或至少1000个核苷酸分开。

35.如权利要求16-29中任一项所述的组合物，其中包括所述任选的5'接头序列和/或所述任选的3'接头序列在内的所述至少两次迭代被至少10个核苷酸、至少15个核苷酸、至少20个核苷酸、至少25个核苷酸、至少30个核苷酸、至少35个核苷酸、至少40个核苷酸、至少45个核苷酸、至少50个核苷酸、至少55个核苷酸、至少60个核苷酸、至少65个核苷酸或至少

70个核苷酸分开。

36.如权利要求16-35中任一项所述的组合物,其中所述至少一种抗原编码核酸序列从5'至3'由下式描述:

$(E_x-(E^N_n)_y)_z$

其中,

E表示包含所述不同表位编码核酸序列中的至少一者的核苷酸序列,

n表示单独的不同表位编码核酸序列的数目并且是包括0的任何整数,

E^N表示包含对于每个相应n的单独的不同表位编码核酸序列的核苷酸序列,

对于z的每次迭代:x=0或1,对于每个n,y=0或1,和x或y中的至少一者=1,并且

z=2或更大,其中所述抗原编码核酸序列包含E、给定E^N或其组合的至少两次迭代。

37.如权利要求16-36中任一项所述的组合物,其中所述抗原编码盒编码氨基酸序列VVVGACGVGK(SEQ ID NO:75)、VVVGADGVGK(SEQ ID NO:78)、VVGAVGVGK(SEQ ID NO:79)和ILDTAGHEEY(SEQ ID NO:82)中的每一者的至少4次迭代。

38.如权利要求16-36中任一项所述的组合物,其中所述不同表位编码核酸序列包含至少2种、至少3种、至少4种、至少5种、至少6种、至少7种或至少8种不同表位编码核酸序列,其各自编码不同的KRAS相关MHC I类新表位。

39.如权利要求16-36中任一项所述的组合物,其中所述至少一种抗原编码核酸序列的表位编码核酸序列中的每一者编码不同的KRAS相关MHC I类新表位。

40.如权利要求16-39中任一项所述的组合物,其中编码所述不同KRAS相关MHC I类新表位的所述核酸序列中的一者或多者包含至少2次、至少3次、至少4次、至少5次、至少6次、至少7次或至少8次迭代。

41.如权利要求16-39中任一项所述的组合物,其中编码所述不同KRAS相关MHC I类新表位的所述核酸序列中的每一者包含至少2次、至少3次、至少4次、至少5次、至少6次、至少7次或至少8次迭代。

42.如权利要求16-39中任一项所述的组合物,其中编码所述不同KRAS相关MHC I类新表位的所述核酸序列中的一者或多者包含至少4次迭代。

43.如权利要求16-39中任一项所述的组合物,其中编码所述不同KRAS相关MHC I类新表位的所述核酸序列中的每一者包含至少4次迭代。

44.如权利要求16-43中任一项所述的组合物,其中所述不同KRAS相关MHC I类新表位中的一者或多者包含KRAS G12C突变、KRAS G12V突变、KRAS G12D突变或KRAS Q61H突变。

45.如权利要求1-44中任一项所述的组合物,其中所述至少两次迭代包括多次迭代,或者z包括相对于包含所述表位编码核酸序列的单次迭代的抗原编码核酸序列足以刺激更大免疫应答的数目。

46.如权利要求1-45中任一项所述的组合物,其中所述至少两次迭代包括多次迭代,或者z包括足以刺激免疫应答的数目,并且所述表位编码核酸序列的单次迭代不足以刺激所述免疫应答或不足以刺激可检测的免疫应答。

47.如权利要求45或46所述的组合物,其中所述免疫应答是用用于递送所述抗原表达系统的组合物进行体内免疫后表位特异性T细胞的扩增。

48.如权利要求45或46所述的组合物,其中所述免疫应答是在用用于递送所述抗原表

达系统的组合物进行体内免疫后增加的表位特异性T细胞激活和/或增加的由表位特异性T细胞进行的表位特异性杀伤。

49.一种用于递送抗原表达系统的组合物,所述组合物包含:

所述抗原表达系统,

其中所述抗原表达系统包含一种或多种载体,

所述一种或多种载体包含:

(a)载体骨架,其中所述骨架包含:

(i)至少一种启动子核苷酸序列,和

(ii)任选地,至少一种聚腺苷酸化(聚(A))序列;和

(b)盒,其中所述盒包含:

(i)至少一种抗原编码核酸序列,所述抗原编码核酸序列包含:

(I)至少两种不同表位编码核酸序列,任选地包含:(1)使所编码的表位序列与由野生型核酸序列编码的相应肽序列不同的至少一种改变,任选地其中所述至少一种改变是KRAS突变,或(2)编码感染性疾病生物体肽的核酸序列,所述肽选自由以下各项组成的组:病原体衍生肽、病毒衍生肽、细菌衍生肽、真菌衍生肽和寄生虫衍生肽,并且

其中所述表位编码核酸序列中的每一者包含:

(A)任选地,5'接头序列,和

(B)任选地,3'接头序列;

(ii)任选地,可操作地连接到所述抗原编码核酸序列的第二启动子核苷酸序列;和

(iii)任选地,至少一种MHC II类表位编码核酸序列;

(iv)任选地,至少一种编码GPGPG氨基酸接头序列(SEQ IDNO:56)的核酸序列;和

(v)任选地,至少一种第二聚(A)序列,其中所述第二聚(A)序列是所述载体骨架的原生聚(A)序列或外源聚(A)序列,

其中如果所述第二启动子核苷酸序列不存在,则所述抗原编码核酸序列可操作地连接到所述至少一种启动子核苷酸序列,并且

其中所述盒不编码免疫显性MHC I类表位,所述免疫显性MHC I类表位:

(1)当以疫苗组合物施用于受试者时,相对于所述盒中编码并能够刺激所述受试者中的免疫应答的另一种MHC I类表位,刺激5倍或更高的免疫应答,和/或

(2)当以疫苗组合物施用于受试者时,相对于当在不存在所述免疫显性MHC I类表位的情况下施用其他MHC I类表位时的免疫应答,降低了对所述盒中编码的另一种MHC I类表位的免疫应答,任选地其中所述免疫应答降低到检测限以下和/或其中所述免疫应答不是治疗有效应答。

50.一种用于递送抗原表达系统的组合物,所述组合物包含:

所述抗原表达系统,

其中所述抗原表达系统包含一种或多种载体,

所述一种或多种载体包含:

(a)载体骨架,其中所述骨架包含:

(i)至少一种启动子核苷酸序列,和

(ii)任选地,至少一种聚腺苷酸化(聚(A))序列;和

(b)盒,其中所述盒包含:

(i)至少一种抗原编码核酸序列,所述抗原编码核酸序列包含:

(I)彼此线性连接的至少2、3、4、5、6、7、8、9或10种不同表位编码核酸序列,

任选地包含:(1)使所编码的表位序列与由野生型核酸序列编码的相应肽序列不同的至少一种改变,任选地其中所述至少一种改变是KRAS突变,或(2)编码感染性疾病生物体肽的核酸序列,所述肽选自由以下各项组成的组:病原体衍生肽、病毒衍生肽、细菌衍生肽、真菌衍生肽和寄生虫衍生肽,并且

其中所述表位编码核酸序列中的每一者包含:

(A)任选地,5'接头序列,和

(B)任选地,3'接头序列;

(ii)任选地,可操作地连接到所述抗原编码核酸序列的第二启动子核苷酸序列;和

(iii)任选地,至少一种MHC II类表位编码核酸序列;

(iv)任选地,至少一种编码GPGPG氨基酸接头序列(SEQ IDNO:56)的核酸序列;和

(v)任选地,至少一种第二聚(A)序列,其中所述第二聚(A)序列是所述载体骨架的原生聚(A)序列或外源聚(A)序列,并且

其中如果所述第二启动子核苷酸序列不存在,则所述抗原编码核酸序列可操作地连接到所述至少一种启动子核苷酸序列,并且

其中所述盒不编码免疫显性MHC I类表位,所述免疫显性MHC I类表位:

(1)当以疫苗组合物施用于受试者时,相对于所述盒中编码并能够刺激所述受试者中的免疫应答的另一种MHC I类表位,刺激5倍或更高的免疫应答,和/或

(2)当以疫苗组合物施用于受试者时,相对于当在不存在所述免疫显性MHC I类表位的情况下施用其他MHC I类表位时的免疫应答,降低了对所述盒中编码的另一种MHC I类表位的免疫应答,任选地其中所述免疫应答降低到检测限以下和/或其中所述免疫应答不是治疗有效应答。

51.如权利要求50所述的组合物,其中所述不同表位编码核酸序列中的至少一者编码KRAS相关MHC I类新表位。

52.一种用于递送抗原表达系统的组合物,所述组合物包含:

所述抗原表达系统,

其中所述抗原表达系统包含一种或多种载体,

所述一种或多种载体包含:

(a)载体骨架,其中所述载体骨架包含黑猩猩腺病毒载体,任选地其中所述黑猩猩腺病毒载体是ChAdV68载体,或甲病毒载体,任选地其中所述甲病毒载体是委内瑞拉马脑炎病毒载体;和

(b)盒,任选地其中所述盒整合在所述载体骨架原生的原生启动子核苷酸序列和聚(A)序列之间,任选地其中所述聚(A)序列是所述载体骨架原生的,其中所述盒包含:

(i)至少一种抗原编码核酸序列,所述抗原编码核酸序列包含:

(I)至少一种表位编码核酸序列,任选地包含彼此线性连接的至少两种不同表位编码核酸序列,每个表位编码核酸序列任选地包含:

(A)MHC I类表位编码核酸序列,其中所述MHC I类表位编码核酸序列编码长度为7-15

个氨基酸的MHC I类表位，

(B)5'接头序列，其中所述5'接头序列编码所述MHC I类表位的原生N末端氨基酸序列，并且其中所述5'接头序列编码长度为至少2个氨基酸的肽，

(C)3'接头序列，其中所述3'接头序列编码所述MHC I类表位的原生C末端酸序列，并且其中所述3'接头序列编码长度为至少2个氨基酸的肽，并且

其中所述盒可操作地连接到所述原生启动子核苷酸序列，其中所述表位编码核酸序列中的每一者编码长度为13至25个氨基酸的多肽，并且其中每个表位编码核酸序列的每个3'端连接到以下表位编码核酸序列的5'端，所述盒中的最终表位编码核酸序列除外；和

(ii)至少两种MHC II类表位编码核酸序列，所述MHC II类表位编码核酸序列包含：

(I)PADRE MHC II类序列(SEQ ID NO:48)，

(II)破伤风类毒素MHC II类序列(SEQ ID NO:46)，

(III)编码连接所述PADRE MHC II类序列与所述破伤风类毒素MHC II类序列的GPGPG氨基酸接头序列(SEQ ID NO:56)的第一核酸序列，

(IV)编码连接所述至少两种MHC II类表位编码核酸序列的5'端与所述表位编码核酸序列的GPGPG氨基酸接头序列(SEQ ID NO:56)的第二核酸序列，

(V)任选地，在所述至少两种MHC II类表位编码核酸序列的3'端处编码GPGPG氨基酸接头序列(SEQ ID NO:56)的第三核酸序列，和

(iii)任选地，可操作地连接到所述抗原编码核酸序列的第二启动子核苷酸序列；并且

其中如果所述第二启动子核苷酸序列不存在，则所述抗原编码核酸序列可操作地连接到所述原生启动子核苷酸序列，并且

其中所述盒不编码免疫显性MHC I类表位，所述免疫显性MHC I类表位：

(1)当以疫苗组合物施用于受试者时，相对于所述盒中编码并能够刺激所述受试者中的免疫应答的另一种MHC I类表位，刺激5倍或更高的免疫应答，和/或

(2)当以疫苗组合物施用于受试者时，相对于当在不存在所述免疫显性MHC I类表位的情况下施用其他MHC I类表位时的免疫应答，降低了对所述盒中编码的另一种MHC I类表位的免疫应答，任选地其中所述免疫应答降低到检测限以下和/或其中所述免疫应答不是治疗有效应答。

53.如权利要求49-52中任一项所述的组合物，其中当以疫苗组合物施用于受试者时，相对于所述盒中编码并能够刺激所述受试者中的免疫应答的另一种MHC I类表位，所述免疫显性MHC I类表位刺激10倍、20倍、30倍、40倍、50倍、100倍、200倍、300倍、400倍、500倍、1000倍、2000倍、3000倍、4000倍、5000倍或10,000倍或更高的免疫应答。

54.如权利要求49-52中任一项所述的组合物，其中所述免疫显性MHC I类表位将所述其他MHC I类表位的免疫应答降低到检测限以下和/或不刺激治疗有效应答。

55.如权利要求49-54中任一项所述的组合物，其中所述受试者表达至少一种已知或预测呈递所述免疫显性MHC I类表位和所述盒中编码的所述其他MHC I类表位的HLA等位基因。

56.如权利要求1-55中任一项所述的组合物，其中所述表位编码核酸序列中的一者或多者来源于受试者的肿瘤、感染或受感染细胞。

57.如权利要求1-55中任一项所述的组合物，其中所述表位编码核酸序列中的每一者

来源于受试者的肿瘤、感染或受感染细胞。

58.如权利要求1-55中任一项所述的组合物，其中所述表位编码核酸序列中的一者或多者不是来源于受试者的肿瘤、感染或受感染细胞。

59.如权利要求1-55中任一项所述的组合物，其中所述表位编码核酸序列中的每一者不是来源于受试者的肿瘤、感染或受感染细胞。

60.如权利要求1-59中任一项所述的组合物，其中所述表位编码核酸序列编码已知或怀疑由MHC I类呈递在细胞表面上的表位，任选地其中所述细胞表面是肿瘤细胞表面或受感染细胞表面，并且任选地其中所述细胞是受试者的细胞。

61.如权利要求60所述的组合物，其中所述细胞是选自由以下各项组成的组的肿瘤细胞：肺癌、黑色素瘤、乳腺癌、卵巢癌、前列腺癌、肾癌、胃癌、结肠癌、睾丸癌、头颈癌、胰腺癌、脑癌、B细胞淋巴瘤、急性髓细胞白血病、慢性髓细胞白血病、慢性淋巴细胞白血病、T细胞淋巴细胞白血病、非小细胞肺癌和小细胞肺癌，或者

其中所述细胞是选自由以下各项组成的组的受感染细胞：病原体感染细胞、病毒感染细胞、细菌感染细胞、真菌感染细胞和寄生虫感染细胞。

62.如权利要求61所述的组合物，其中所述病毒感染细胞选自由以下组成的组：HIV感染细胞、严重急性呼吸综合征相关冠状病毒(SARS)感染细胞、严重急性呼吸综合征冠状病毒2(SARS-CoV-2)感染细胞、埃博拉感染细胞、乙型肝炎病毒(HBV)感染细胞、流感感染细胞、正粘病毒科病毒感染细胞、人乳头瘤病毒(HPV)感染细胞、巨细胞病毒(CMV)感染细胞、基孔肯雅病毒感染细胞、呼吸道合胞病毒(RSV)感染细胞、登革热病毒感染细胞和丙型肝炎病毒(HCV)感染细胞。

63.如上述权利要求中任一项所述的组合物，其中所述组合物还包含纳米微粒递送媒介物。

64.如权利要求63所述的组合物，其中所述纳米微粒递送媒介物是脂质纳米粒子(LNP)。

65.如权利要求64所述的组合物，其中所述LNP包含可电离氨基脂质。

66.如权利要求65所述的组合物，其中所述可电离氨基脂质包含MC3样(二亚油基甲基-4-二甲基氨基丁酸酯)分子。

67.如权利要求24-66中任一项所述的组合物，其中所述纳米微粒递送媒介物包封所述抗原表达系统。

68.如权利要求16-18、20-22或24-67中任一项所述的组合物，其中所述盒整合在所述至少一种启动子核苷酸序列和所述至少一种聚(A)序列之间。

69.如权利要求16-18、20-22或24-68中任一项所述的组合物，其中所述第二启动子不存在并且所述至少一种启动子核苷酸序列可操作地连接到所述抗原编码核酸序列。

70.如权利要求16-18、20-22或24-69中任一项所述的组合物，其中所述一种或多种载体包含一种或多种+-链RNA载体。

71.如权利要求70所述的组合物，其中所述一种或多种+-链RNA载体包含5'7-甲基鸟苷(m7g)帽。

72.如权利要求70或71所述的组合物，其中所述一种或多种+-链RNA载体是通过体外转录产生的。

73.如权利要求16-18、20-22或24-72中任一项所述的组合物，其中所述一种或多种载体在哺乳动物细胞内自我复制。

74.如权利要求16-18、20-22或24-73中任一项所述的组合物，其中所述骨架包含奥拉病毒、摩根堡病毒、委内瑞拉马脑炎病毒、罗斯河病毒、塞姆利基森林病毒、辛德毕斯病毒或马雅罗病毒的至少一种核苷酸序列。

75.如权利要求16-18、20-22或24-73中任一项所述的组合物，其中所述骨架包含委内瑞拉马脑炎病毒的至少一种核苷酸序列。

76.如权利要求74或75所述的组合物，其中所述骨架至少包含由所述奥拉病毒、所述摩根堡病毒、所述委内瑞拉马脑炎病毒、所述罗斯河病毒、所述塞姆利基森林病毒、所述辛德毕斯病毒或所述马雅罗病毒的核苷酸序列编码的用于非结构蛋白介导的扩增的序列、26S启动子序列、聚(A)序列、非结构蛋白1(nsP1)基因、nsP2基因、nsP3基因和nsP4基因。

77.如权利要求74或75所述的组合物，其中所述骨架至少包含由所述奥拉病毒、所述摩根堡病毒、所述委内瑞拉马脑炎病毒、所述罗斯河病毒、所述塞姆利基森林病毒、所述辛德毕斯病毒或所述马雅罗病毒的核苷酸序列编码的用于非结构蛋白介导的扩增的序列、26S启动子序列和聚(A)序列。

78.如权利要求76或77所述的组合物，其中所述用于非结构蛋白介导的扩增的序列选自由以下各项组成的组：甲病毒5'UTR、51-nt CSE、24-nt CSE、26S亚基因组启动子序列、19-nt CSE、甲病毒3'UTR或其组合。

79.如权利要求76-78中任一项所述的组合物，其中所述骨架不编码结构性病毒粒子蛋白衣壳E2和E1。

80.如权利要求79所述的组合物，其中所述盒被插入以代替所述奥拉病毒、所述摩根堡病毒、所述委内瑞拉马脑炎病毒、所述罗斯河病毒、所述塞姆利基森林病毒、所述辛德毕斯病毒或所述马雅罗病毒的核苷酸序列内的结构性病毒粒子蛋白。

81.如权利要求74或75所述的组合物，其中所述委内瑞拉马脑炎病毒包含SEQ ID NO:3或SEQ ID NO:5的序列。

82.如权利要求74或75所述的组合物，其中所述委内瑞拉马脑炎病毒包含SEQ ID NO:3或SEQ ID NO:5的序列，其还包含碱基对7544和11175之间的缺失。

83.如权利要求82所述的组合物，其中所述骨架包含以SEQ IDNO:6或SEQ ID NO:7所示的序列。

84.如权利要求82或83所述的组合物，其中所述盒被插入在位置7544以代替如SEQ ID NO:3或SEQ ID NO:5的序列中所示的碱基对7544和11175之间的缺失。

85.如权利要求80-84所述的组合物，其中所述盒的插入提供了包含所述nsP1-4基因和所述至少一种抗原编码核酸序列的多顺反子RNA的转录，其中所述nsP1-4基因和所述至少一种抗原编码核酸序列在单独的开放阅读框中。

86.如权利要求16-18、20-22或24-73中任一项所述的组合物，其中所述骨架包含黑猩猩腺病毒载体的至少一种核苷酸序列。

87.如权利要求86所述的组合物，其中所述黑猩猩腺病毒载体是ChAdV68载体，任选地其中所述ChAdV68载体包含ChAdV68载体骨架，所述ChAdV68载体骨架包含：

-以SEQ ID NO:1所示的序列；

-以SEQ ID NO:1所示的序列,除了所述序列在选自由SEQ IDNO:1所示序列的黑猩猩腺病毒E1A、E1B、E2A、E2B、E3、E4、L1、L2、L3、L4和L5基因组成的组的至少一个基因中完全缺失或功能性缺失,任选地其中所述序列在以下中完全缺失或功能性缺失:以SEQ ID NO:1所示序列的(1)E1A和E1B;(2)E1A、E1B和E3;或(3)E1A、E1B、E3和E4;

-从SEQ ID NO:1的序列获得的基因或调控序列,任选地其中所述基因选自由SEQ ID NO:1所示序列的黑猩猩腺病毒反向末端重复序列(ITR)、E1A、E1B、E2A、E2B、E3、E4、L1、L2、L3、L4和L5基因组成的组;

-部分缺失的E4基因,其包含缺失或部分缺失的E4orf2区和缺失或部分缺失的E4orf3区,以及任选地缺失或部分缺失的E4orf4区;

-以SEQ ID NO:1所示序列的至少核苷酸2至36,518,并且还包含:(1)以SEQ ID NO:1所示序列的至少核苷酸577至3403的E1缺失,(2)以SEQ ID NO:1所示序列的至少核苷酸27,125至31,825的E3缺失,和(3)以SEQ ID NO:1所示序列的至少核苷酸34,916至35,642的E4缺失;任选地其中所述抗原盒被插入在所述E1缺失内;

-以SEQ ID NO:68所示的序列,任选地其中所述抗原盒被插入在所述E1缺失内;

-在碱基对编号577和3403之间或碱基对456和3014之间的一个或多个缺失,并且任选地其中所述载体还包含在SEQ ID NO:1所示序列的碱基对27,125和31,825之间或碱基对27,816和31,333之间的一个或多个缺失;或

-在以SEQ ID NO:1所示序列的碱基对编号3957和10346之间、碱基对编号21787和23370之间以及碱基对编号33486和36193之间的一个或多个缺失,并且

任选地其中所述盒被插入在所述ChAdV载体骨架的E1区、E3区和/或允许并入所述盒的任何缺失的AdV区。

88.如权利要求16-18、20-22或24-87中任一项所述的组合物,其中所述至少一种启动子核苷酸序列是由所述骨架编码的原生26S启动子核苷酸序列。

89.如权利要求16-18、20-22或24-87中任一项所述的组合物,其中所述至少一种启动子核苷酸序列是外源RNA启动子。

90.如权利要求16-18、20-22或24-89中任一项所述的组合物,其中所述第二启动子核苷酸序列是26S启动子核苷酸序列。

91.如权利要求16-18、20-22或24-89中任一项所述的组合物,其中所述第二启动子核苷酸序列包含多个26S启动子核苷酸序列,其中每个26S启动子核苷酸序列提供了所述单独的开放阅读框中的一者或多者的转录。

92.如上述权利要求中任一项所述的组合物,其中所述一种或多种载体的大小各自为至少300nt。

93.如上述权利要求中任一项所述的组合物,其中所述一种或多种载体的大小各自为至少1kb。

94.如上述权利要求中任一项所述的组合物,其中所述一种或多种载体的大小各自为2kb。

95.如上述权利要求中任一项所述的组合物,其中所述一种或多种载体的大小各自小于5kb。

96.如上述权利要求中任一项所述的组合物,其中所述至少一种抗原编码核酸序列中

的至少一者编码由MHC I类呈递在细胞表面、任选地肿瘤细胞表面或受感染细胞表面上的多肽序列或其部分。

97.如权利要求1-18、20-22或24-96中任一项所述的组合物，其中每个表位编码核酸序列彼此直接连接。

98.如权利要求1-18、20-22或24-97中任一项所述的组合物，其中所述表位编码核酸序列中的至少一者用编码接头的核酸序列连接到不同表位编码核酸序列。

99.如权利要求98所述的组合物，其中所述接头连接两条MHC I类序列或一条MHC I类序列至一条MHC II类序列。

100.如权利要求99所述的组合物，其中所述接头选自由以下组成的组：(1)连续甘氨酸残基，长度为至少2、3、4、5、6、7、8、9或10个残基；(2)连续丙氨酸残基，长度为至少2、3、4、5、6、7、8、9或10个残基；(3)两个精氨酸残基(RR)；(4)丙氨酸、丙氨酸、酪氨酸(AAY)；(5)由哺乳动物蛋白酶体有效加工的长度为至少2、3、4、5、6、7、8、9或10个氨基酸残基的共有序列；和(6)侧接源自同源蛋白的抗原并且长度为至少2、3、4、5、6、7、8、9、10、11、12、13、14、15、16、17、18、19、20个或2-20个氨基酸残基的一种或多种原生序列。

101.如权利要求98所述的组合物，其中所述接头连接两个MHC II类序列或一条MHC II类序列至一条MHC I类序列。

102.如权利要求101所述的组合物，其中所述接头包含序列GPGPG(SEQ ID NO:56)。

103.如权利要求1-18、20-22或24-102中任一项所述的组合物，其中所述表位编码核酸序列中的至少一条序列可操作地或直接地连接到增强由其编码的表位的所述表位编码核酸序列的表达、稳定性、细胞运输、加工和呈递和/或免疫原性的单独或连续的序列。

104.如权利要求103所述的组合物，其中所述单独或连续的序列包含以下中的至少一者：泛素序列、经修饰以增加蛋白酶体靶向的泛素序列(例如，所述泛素序列在位置76处含有Gly至Ala取代)、免疫球蛋白信号序列(例如，IgK)、主要组织相容性I类序列、溶酶体相关膜蛋白(LAMP)-1、人树突状细胞溶酶体相关膜蛋白和主要组织相容性II类序列；任选地其中所述经修饰以增加蛋白酶体靶向的泛素序列是A76。

105.如上述权利要求中任一项所述的组合物，其中所述表位编码核酸序列中的至少一者编码多肽序列或其部分，其相对于经翻译的相应野生型核酸序列对其相应的MHC等位基因具有增加的结合亲和力。

106.如上述权利要求中任一项所述的组合物，其中所述表位编码核酸序列中的至少一者编码多肽序列或其部分，其相对于所述经翻译的相应野生型核酸序列对其相应的MHC等位基因具有增加的结合稳定性。

107.如上述权利要求中任一项所述的组合物，其中所述表位编码核酸序列中的至少一者编码多肽序列或其部分，其相对于所述经翻译的相应野生型核酸序列具有增加的在其相应MHC等位基因上呈递的可能性。

108.如上述权利要求中任一项所述的组合物，其中所述至少一种改变包括点突变、移码突变、非移码突变、缺失突变、插入突变、剪接变体、基因组重排或蛋白酶体产生的剪接抗原。

109.如上述权利要求中任一项所述的组合物，其中所述肿瘤选自由以下组成的组：肺癌、黑色素瘤、乳腺癌、卵巢癌、前列腺癌、肾癌、胃癌、结肠癌、睾丸癌、头颈癌、胰腺癌、膀胱

癌、脑癌、B细胞淋巴瘤、急性髓细胞白血病、成人急性淋巴母细胞白血病、慢性髓细胞白血病、慢性淋巴细胞白血病、T细胞淋巴细胞白血病、非小细胞肺癌和小细胞肺癌，

或者所述感染性疾病生物体选自由以下组成的组：严重急性呼吸综合征相关冠状病毒(SARS)、严重急性呼吸综合征冠状病毒2(SARS-CoV-2)、埃博拉、HIV、乙型肝炎病毒(HBV)、流感、丙型肝炎病毒(HCV)、人乳头瘤病毒(HPV)、巨细胞病毒(CMV)、基孔肯雅病毒、呼吸道合胞病毒(RSV)、登革热病毒、正粘病毒科病毒和肺结核。

110.如权利要求1-18、20-22或24-109中任一项所述的组合物，其中所述至少一种抗原编码核酸序列包含至少2-10、2、3、4、5、6、7、8、9或10种表位编码核酸序列。

111.如权利要求1-18、20-22或24-109中任一项所述的组合物，其中所述至少一种抗原编码核酸序列包含至少11-20、15-20、11-100、11-200、11-300、11-400、11、12、13、14、15、16、17、18、19、20或多达400种表位编码核酸序列。

112.如权利要求1-18、20-22或24-109中任一项所述的组合物，其中所述至少一种抗原编码核酸序列包含至少2-400种表位编码核酸序列，并且其中所述表位编码核酸序列中的至少两者编码由MHC I类呈递在细胞表面、任选地肿瘤细胞表面或受感染细胞表面上的多肽序列或其部分。

113.如权利要求1-18、20-22或24-112中任一项所述的组合物，其中所述至少一种抗原编码核酸序列包含至少2-10、2、3、4、5、6、7、8、9或10种抗原编码核酸序列。

114.如权利要求1-18、20-22或24-112中任一项所述的组合物，其中所述至少一种抗原编码核酸序列包含至少11-20、15-20、11-100、11-200、11-300、11-400、11、12、13、14、15、16、17、18、19、20或多达400种抗原编码核酸序列。

115.如权利要求1-18、20-22或24-112中任一项所述的组合物，其中所述至少一种抗原编码核酸序列包含至少2-400种抗原编码核酸序列，并且其中所述抗原编码核酸序列中的至少两者编码由MHC I类呈递在细胞表面、任选地肿瘤细胞表面或受感染细胞表面上的多肽序列或其部分。

116.如权利要求19或23所述的组合物，其中所述表位编码核酸序列中的至少两者编码由MHC I类呈递在细胞表面、任选地肿瘤细胞表面或受感染细胞表面上的多肽序列或其部分。

117.如上述权利要求中任一项所述的组合物，其中当施用于所述受试者并被翻译时，由所述表位编码核酸序列编码的表位中的至少一者被呈递在抗原呈递细胞上，导致靶向所述肿瘤细胞表面或所述受感染细胞表面上的抗原中的至少一者的免疫应答。

118.如上述权利要求中任一项所述的组合物，其中当将所述至少一种抗原编码核酸序列施用于所述受试者并被翻译时，所述MHC I类或II类表位中的至少一者被呈递在抗原呈递细胞上，导致靶向肿瘤细胞表面或受感染细胞表面上的表位中的至少一者的免疫应答，并且任选地其中所述至少一种抗原编码核酸序列中的每一者的表达由所述至少一种启动子核苷酸序列驱动。

119.如权利要求1-18、20-22或24-118中任一项所述的组合物，其中每个表位编码核酸序列编码长度为8至35个氨基酸，任选地长度为9-17、9-25、8、9、10、11、12、13、14、15、16、17、18、19、20、21、22、23、24、25、26、27、28、29、30、31、32、33、34或35个氨基酸的多肽序列。

120.如权利要求1-18、20-22或24-119中任一项所述的组合物，其中存在所述至少一种

MHC II类表位编码核酸序列。

121.如权利要求1-18、20-22或24-119中任一项所述的组合物，其中所述至少一种MHC II类表位编码核酸序列存在并且包含至少一种MHC II类表位编码核酸序列，所述MHC II类表位编码核酸序列包含使所编码的肽序列与由野生型核酸序列编码的相应肽序列不同的至少一种改变。

122.如权利要求1-18、20-22或24-121中任一项所述的组合物，其中所述至少一种MHC II类表位编码核酸序列的长度为12-20、12、13、14、15、16、17、18、19、20或20-40个氨基酸。

123.如权利要求1-18、20-22或24-122中任一项所述的组合物，其中所述至少一种MHC II类表位编码核酸序列存在并且包含至少一种通用MHC II类抗原编码核酸序列，任选地其中所述至少一种通用序列包含破伤风类毒素和PADRE中的至少一者。

124.如权利要求16-18、20-22或24-123中任一项所述的组合物，其中所述至少一种启动子核苷酸序列或所述第二启动子核苷酸序列是诱导型的。

125.如权利要求16-18、20-22或24-123中任一项所述的组合物，其中所述至少一种启动子核苷酸序列或所述第二启动子核苷酸序列是非诱导型的。

126.如权利要求16-18、20-22或24-125中任一项所述的组合物，其中所述至少一种聚(A)序列包含所述骨架原生的聚(A)序列。

127.如权利要求16-18、20-22或24-125中任一项所述的组合物，其中所述至少一种聚(A)序列包含所述骨架外源的聚(A)序列。

128.如权利要求16-18、20-22或24-127中任一项所述的组合物，其中所述至少一种聚(A)序列可操作地连接到所述至少一种抗原编码核酸序列中的至少一者。

129.如权利要求16-18、20-22或24-128中任一项所述的组合物，其中所述至少一种聚(A)序列是至少20、至少30、至少40、至少50、至少60、至少70、至少80、至少90个或至少100个连续A核苷酸。

130.如权利要求16-18、20-22或24-128中任一项所述的组合物，其中所述至少一种聚(A)序列是至少80个连续A核苷酸。

131.如上述权利要求中任一项所述的组合物，其中所述盒还包含以下中的至少一者：内含子序列，土拨鼠肝炎病毒转录后调控元件(WPRE)序列，内部核糖体进入序列(IRES)序列，编码2A自切割肽序列的核苷酸序列，编码弗林蛋白酶切割位点的核苷酸序列，或已知增强可操作地连接到所述至少一种抗原编码核酸序列中的至少一者的mRNA的核输出、稳定性或翻译效率的5'或3'非编码区中的序列。

132.如上述权利要求中任一项所述的组合物，其中所述盒还包含报告基因，包括但不限于绿色荧光蛋白(GFP)、GFP变体、分泌型碱性磷酸酶、荧光素酶、荧光素酶变体、或可检测的肽或表位。

133.如权利要求132所述的组合物，其中所述可检测的肽或表位选自由HA标签、Flag标签、His标签或V5标签组成的组。

134.如上述权利要求中任一项所述的组合物，其中所述一种或多种载体还包含一种或多种编码至少一种免疫调节剂的核酸序列。

135.如权利要求134所述的组合物，其中所述免疫调节剂是抗CTLA4抗体或其抗原结合片段、抗PD-1抗体或其抗原结合片段、抗PD-L1抗体或其抗原结合片段、抗4-1BB抗体或其抗

原结合片段、或抗OX-40抗体或其抗原结合片段。

136.如权利要求135所述的组合物，其中所述抗体或其抗原结合片段是Fab片段、Fab'片段、单链Fv(scFv)、呈单一特异性或连接在一起的多个特异性的单域抗体(sdAb)(例如骆驼科动物抗体结构域)，或全长单链抗体(例如，具有通过柔性接头连接的重链和轻链的全长IgG)。

137.如权利要求135或136所述的组合物，其中所述抗体的重链和轻链序列是由诸如2A或IRES的自切割序列分开的连续序列；或者所述抗体的重链和轻链序列是通过柔性接头诸如连续甘氨酸残基连接的。

138.如权利要求134所述的组合物，其中所述免疫调节剂是细胞因子。

139.如权利要求138所述的组合物，其中所述细胞因子是IL-2、IL-7、IL-12、IL-15或IL-21或其各自的变体中的至少一者。

140.如权利要求1-18、20-22或24-139中任一项所述的组合物，其中至少一种表位编码核酸序列是通过执行以下步骤来选择的：

(a)从肿瘤、受感染细胞或感染性疾病生物体获得外显子组、转录组或全基因组核苷酸测序数据中的至少一者，其中所述核苷酸测序数据用于获得代表一组抗原中的每一者的肽序列的数据；

(b)将每个抗原的肽序列输入到呈递模型中，以产生所述抗原中的每一者由MHC等位基因中的一者或多者在细胞表面、任选地肿瘤细胞表面或受感染细胞表面上呈递的数值可能性集合，所述数值可能性集合已至少基于所接收的质谱数据进行鉴定；以及

(c)基于所述数值可能性集合选择所述抗原集合的子集以产生经选择抗原集合，所述经选择抗原用于产生所述表位编码核酸序列。

141.如权利要求19或23所述的组合物，其中所述表位编码核酸序列中的每一者是通过执行以下步骤来选择：

(a)从肿瘤、受感染细胞或感染性疾病生物体获得外显子组、转录组或全基因组核苷酸测序数据中的至少一者，其中所述核苷酸测序数据用于获得代表一组抗原中的每一者的肽序列的数据；

(b)将每个抗原的肽序列输入到呈递模型中，以产生所述抗原中的每一者由MHC等位基因中的一者或多者在细胞表面、任选地肿瘤细胞表面或受感染细胞表面上呈递的数值可能性集合，所述数值可能性集合已至少基于所接收的质谱数据进行鉴定；以及

(c)基于所述数值可能性集合选择所述抗原集合的子集以产生经选择抗原集合，所述经选择抗原用于产生至少20种表位编码核酸序列。

142.如权利要求140所述的组合物，其中所述经选择抗原集合的数目为2-20。

143.如权利要求140-142所述的组合物，其中所述呈递模型表示以下两者之间的依赖性：

(a)所述MHC等位基因中的特定一者和在肽序列的特定位置处的特定氨基酸的一对的存在；和

(b)由所述对的MHC等位基因中的所述特定一者将在所述特定位置处包含所述特定氨基酸的所述肽序列呈递在细胞表面、任选地肿瘤细胞表面或受感染细胞表面上的可能性。

144.如权利要求140-143所述的组合物，其中选择所述经选择抗原集合包括基于所述

呈递模型选择相对于未选择的抗原被呈递在所述细胞表面上的可能性增加的抗原，任选地其中所述经选择抗原已被验证为由一种或多种特定的HLA等位基因呈递。

145.如权利要求140-144所述的组合物，其中选择所述经选择抗原集合包括基于所述呈递模型选择相对于未选择的抗原能够在所述受试者中诱导肿瘤特异性或感染性疾病特异性免疫应答的可能性增加的抗原。

146.如权利要求140-145所述的组合物，其中选择所述经选择抗原集合包括基于所述呈递模型选择相对于未选择的抗原能够由专职抗原呈递细胞(APC)呈递给初始T细胞的可能性增加的抗原，任选地其中所述APC是树突状细胞(DC)。

147.如权利要求140-146所述的组合物，其中选择所述经选择抗原集合包括基于所述呈递模型选择相对于未选择的抗原经由中心或外周耐受性受抑制的可能性降低的抗原。

148.如权利要求140-147所述的组合物，其中选择所述经选择抗原集合包括基于所述呈递模型选择相对于未选择的抗原能够诱导对所述受试者的正常组织的自身免疫应答的可能性降低的抗原。

149.如权利要求140-148所述的组合物，其中外显子组或转录组核苷酸测序数据是通过对肿瘤细胞或组织、受感染细胞或感染性疾病生物体进行测序而获得的。

150.如权利要求149所述的组合物，其中所述测序是下一代测序(NGS)或任何大规模并行测序方法。

151.如上述权利要求中任一项所述的组合物，其中所述盒包含由所述盒中的相邻序列形成的连接表位序列。

152.如权利要求151所述的组合物，其中至少一个或每个连接表位序列对MHC具有大于500nM的亲和力。

153.如权利要求151或152所述的组合物，其中每个连接表位序列都是非自身的。

154.如上述权利要求中任一项所述的组合物，其中预测或验证所述MHC I类表位中的每一者能够由至少5%的群体中存在的至少一个HLA等位基因呈递。

155.如上述权利要求中任一项所述的组合物，其中预测或验证所述MHC I类表位中的每一者能够由至少一个HLA等位基因呈递，其中每个抗原/HLA对在群体中具有至少0.01%的抗原/HLA流行率。

156.如上述权利要求中任一项所述的组合物，其中预测或验证所述MHC I类表位中的每一者能够由至少一个HLA等位基因呈递，其中每个抗原/HLA对在群体中具有至少0.1%的抗原/HLA流行率。

157.如上述权利要求中任一项所述的组合物，其中所述盒不编码包含经翻译的野生型核酸序列的非治疗性MHC I类或II类表位核酸序列，其中预测所述非治疗性表位将展示在所述受试者的MHC等位基因上。

158.如权利要求157所述的组合物，其中所述非治疗性预测的MHC I类或II类表位序列是由所述盒中的相邻序列形成的连接表位序列。

159.如权利要求151-158所述的组合物，其中所述预测是基于通过将所述非治疗性表位的序列输入到呈递模型中而生成的呈递可能性的。

160.如权利要求151-159中任一项所述的组合物，其中所述盒中的所述至少一种抗原编码核酸序列的顺序通过一系列包括以下的步骤确定：

(a)产生一组对应于不同顺序的所述至少一种抗原编码核酸序列的候选盒序列;

(b)对于每个候选盒序列,基于所述候选盒序列中非治疗性表位的呈递来确定呈递分数;以及

(c)选择与低于预定阈值的呈递分数相关的候选盒序列作为抗原疫苗的盒序列。

161.一种药物组合物,所述药物组合物包含如上述权利要求中任一项所述的组合物和药学上可接受的载剂。

162.如权利要求161所述的组合物,其中所述组合物还包含佐剂。

163.如权利要求161或162所述的药物组合物,其中所述组合物还包含免疫调节剂。

164.如权利要求163所述的药物组合物,其中所述免疫调节剂是抗CTLA4抗体或其抗原结合片段、抗PD-1抗体或其抗原结合片段、抗PD-L1抗体或其抗原结合片段、抗4-1BB抗体或其抗原结合片段、或抗OX-40抗体或其抗原结合片段。

165.一种经分离的核苷酸序列或分离的核苷酸序列集合,所述经分离的核苷酸序列或经分离的核苷酸序列集合包含如上述组合物权利要求中任一项所述的组合物的盒和从SEQ ID NO:3或SEQ IDNO:5的序列获得的一种或多种元件,任选地其中所述一种或多种元件选自由以下各项组成的组:非结构蛋白介导的扩增所必需的序列、26S启动子核苷酸序列、聚(A)序列以及以SEQ ID NO:3或SEQ IDNO:5所示的序列的nsP1-4基因,并且任选地其中所述核苷酸序列是cDNA。

166.如权利要求165所述的经分离的核苷酸序列,其中所述序列或经分离的核苷酸序列集合包含插入在以SEQ ID NO:6或SEQ IDNO:7所示序列的位置7544处的如上述组合物权利要求中任一项所述的盒。

167.如权利要求165或166所述的经分离的核苷酸序列,所述经分离的核苷酸序列还包含:

a)从SEQ ID NO:3或SEQ ID NO:5的序列获得的所述一种或多种元件的T7或SP6 RNA聚合酶启动子核苷酸序列5';和

b)任选地,所述聚(A)序列的一个或多个限制性位点3'。

168.如权利要求165所述的经分离的核苷酸序列,其中如上述组合物权利要求中任一项所述的盒被插入在SEQ ID NO:8或SEQ IDNO:9的位置7563处。

169.一种载体或载体集合,所述载体或载体集合包含如权利要求165-168所述的核苷酸序列。

170.一种经分离的细胞,所述经分离的细胞包含如权利要求165-169所述的核苷酸序列或经分离的核苷酸序列集合,任选地其中所述细胞是BHK-21、CHO、HEK293或其变体、911、HeLa、A549、LP-293、PER.C6或AE1-2a细胞。

171.一种试剂盒,所述试剂盒包括如上述组合物权利要求中任一项所述的组合物和使用说明书。

172.一种用于治疗患有癌症的受试者的方法,所述方法包括向所述受试者施用如上述组合物权利要求中任一项所述的组合物或如权利要求161-164中任一项所述的药物组合物。

173.如权利要求172所述的方法,其中所述表位编码核酸序列来源于所述患有癌症的受试者的肿瘤或来源于受感染受试者的细胞或样品。

174.如权利要求172所述的方法，其中所述表位编码核酸序列不是来源于所述患有癌症的受试者的肿瘤或来源于受感染受试者的细胞或样品。

175.一种用于刺激受试者中的免疫应答的方法，所述方法包括向所述受试者施用如上述组合物权利要求中任一项所述的组合物或如权利要求161-164中任一项所述的药物组合物。

176.如权利要求172-175中任一项所述的方法，其中所述受试者表达至少一种预测或已知呈递所述MHC I类表位的HLA等位基因，任选地其中所述HLA是A*03:01、A*11:01、A*02:01、C*01:02和/或A*01:01。

177.如权利要求172-176中任一项所述的方法，其中所述组合物是肌内(IM)、皮内(ID)、皮下(SC)或静脉内(IV)施用的。

178.如权利要求172-176中任一项所述的方法，其中所述组合物是肌内施用的。

179.如权利要求172-178中任一项所述的方法，所述方法还包括施用一种或多种免疫调节剂，任选地其中在施用所述组合物或药物组合物之前、同时或之后施用所述免疫调节剂。

180.如权利要求179所述的方法，其中所述一种或多种免疫调节剂选自由以下各项组成的组：抗CTLA4抗体或其抗原结合片段、抗PD-1抗体或其抗原结合片段、抗PD-L1抗体或其抗原结合片段、抗4-1BB抗体或其抗原结合片段、或抗OX-40抗体或其抗原结合片段。

181.如权利要求179或180所述的方法，其中所述免疫调节剂是静脉内(IV)、肌内(IM)、皮内(ID)或皮下(SC)施用的。

182.如权利要求181所述的方法，其中所述皮下施用在所述组合物或药物组合物施用部位附近或紧邻一个或多个载体或组合物引流淋巴结。

183.如权利要求172-182中任一项所述的方法，所述方法还包括向所述受试者施用第二疫苗组合物。

184.如权利要求183所述的方法，其中所述第二疫苗组合物是在施用如权利要求172-182中任一项所述的组合物或药物组合物之前施用的。

185.如权利要求183所述的方法，其中所述第二疫苗组合物是在施用如权利要求172-182中任一项所述的组合物或药物组合物之后施用的。

186.如权利要求184或185所述的方法，其中所述第二疫苗组合物与如权利要求172-182中任一项所述的组合物或药物组合物相同。

187.如权利要求184或185所述的方法，其中所述第二疫苗组合物与如权利要求172-182中任一项所述的组合物或药物组合物不同。

188.如权利要求187所述的方法，其中所述第二疫苗组合物包含编码至少一种抗原编码核酸序列的黑猩猩腺病毒载体。

189.如权利要求188所述的方法，其中由所述黑猩猩腺病毒载体编码的所述至少一种抗原编码核酸序列与如上述组合物权利要求中任一项所述的至少一种抗原编码核酸序列相同。

190.一种制造如上述组合物权利要求中任一项所述的一种或多种载体的方法，所述方法包括：

(a)获得包含骨架和盒的线性化DNA序列；

(b)通过将所述线性化DNA序列添加到包含将所述线性化DNA序列转录成RNA的所有必要组分的体外转录反应中,来体外转录所述线性化DNA序列,任选地还包括将m7g帽体外添加到所得RNA中;以及

(c)从所述体外转录反应中分离所述一种或多种载体。

191.如权利要求190所述的制造方法,其中所述线性化DNA序列是通过将DNA质粒序列线性化或通过使用PCR扩增而产生的。

192.如权利要求191所述的制造方法,其中所述DNA质粒序列是使用细菌重组或全基因组DNA合成或全基因组DNA合成中的一者与所合成DNA在细菌细胞中扩增产生的。

193.如权利要求190所述的制造方法,其中从所述体外转录反应中分离所述一种或多种载体涉及苯酚氯仿提取、基于硅胶柱的纯化或类似的RNA纯化方法中的一者或多者。

194.一种制造用于递送抗原表达系统的如上述组合物权利要求中任一项所述的组合物的方法,所述方法包括:

(a)提供用于纳米微粒递送媒介物的组分;

(b)提供所述抗原表达系统;以及

(c)为所述纳米微粒递送媒介物和所述抗原表达系统提供足以产生用于递送所述抗原表达系统的组合物的条件。

195.如权利要求194所述的制造方法,其中所述条件由微流体混合提供。

196.一种用于治疗患有疾病的受试者的方法,任选地其中所述疾病是癌症或感染,所述方法包括向所述受试者施用基于抗原的疫苗给所述受试者,其中所述基于抗原的疫苗包含抗原编码盒或由所述盒编码的多肽序列,其中所述抗原编码盒包含至少一种由下式从5'至3'描述的抗原编码核酸序列:

$(E_x-(E^N_n)_y)_z$

其中,

E表示核苷酸序列不同表位编码核酸序列,

n表示单独的不同表位编码核酸序列的数目并且是包括0的任何整数,

E^N表示包含对于每个相应n的单独的不同表位编码核酸序列的核苷酸序列,

对于z的每次迭代:x=0或1,对于每个n,y=0或1,和x或y中的至少一者=1,并且

z=2或更大,其中所述抗原编码核酸序列包含E、给定E^N或其组合的至少两次迭代,并且

包含所述至少两次迭代的所述不同表位编码核酸序列中的至少一者编码KRAS相关MHC I类新表位。

197.一种用于治疗患有疾病的受试者的方法,任选地其中所述疾病是癌症,所述方法包括向所述受试者施用基于抗原的疫苗给所述受试者,其中所述基于抗原的疫苗包含抗原表达系统,所述抗原表达系统包含:

所述抗原表达系统,

其中所述抗原表达系统包含一种或多种载体,

所述一种或多种载体包含:

(a)载体骨架,其中所述骨架包含:

(i)至少一种启动子核苷酸序列,和

(ii)任选地,至少一种聚腺苷酸化(聚(A))序列;和

(b)盒,其中所述盒包含:

(i)至少一种抗原编码核酸序列,所述抗原编码核酸序列包含:

(I)编码KRAS相关MHC I类新表位的表位编码核酸序列,并且

其中所述表位编码核酸序列中的每一者包含:

(A)任选地,5'接头序列,和

(B)任选地,3'接头序列;

(ii)任选地,可操作地连接到所述抗原编码核酸序列的第二启动子核苷酸序列;和

(iii)任选地,至少一种MHC II类表位编码核酸序列;

(iv)任选地,至少一种编码GPGPG氨基酸接头序列(SEQ IDNO:56)的核酸序列;和

(v)任选地,至少一种第二聚(A)序列,其中所述第二聚(A)序列是所述载体骨架的原生聚(A)序列或外源聚(A)序列,

其中如果所述第二启动子核苷酸序列不存在,则所述抗原编码核酸序列可操作地连接到所述至少一种启动子核苷酸序列,并且

其中所述至少一种抗原编码核酸序列包含编码所述KRAS相关MHC I类新表位的所述表位编码核酸序列的至少两次迭代。

198.一种用于治疗患有疾病的受试者的方法,任选地其中所述疾病是癌症,所述方法包括向所述受试者施用基于抗原的疫苗给所述受试者,其中所述基于抗原的疫苗包含抗原表达系统,所述抗原表达系统包含:

所述抗原表达系统,

其中所述抗原表达系统包含一种或多种载体,

所述一种或多种载体包含:

(a)载体骨架,其中所述骨架包含:

(i)至少一种启动子核苷酸序列,和

(ii)任选地,至少一种聚腺苷酸化(聚(A))序列;和

(b)盒,其中所述盒包含:

(i)至少一种抗原编码核酸序列,所述抗原编码核酸序列包含:

(I)至少两种不同表位编码核酸序列,任选地包含:(1)使所编码的表位序列与由野生型核酸序列编码的相应肽序列不同的至少一种改变,任选地其中所述至少一种改变是KRAS突变,或(2)编码感染性疾病生物体肽的核酸序列,所述肽选自由以下组成的组:病原体衍生肽、病毒衍生肽、细菌衍生肽、真菌衍生肽和寄生虫衍生肽,并且

其中所述表位编码核酸序列中的每一者包含:

(A)任选地,5'接头序列,和

(B)任选地,3'接头序列;

(ii)任选地,可操作地连接到所述抗原编码核酸序列的第二启动子核苷酸序列;和

(iii)任选地,至少一种MHC II类表位编码核酸序列;

(iv)任选地,至少一种编码GPGPG氨基酸接头序列(SEQ IDNO:56)的核酸序列;和

(v)任选地,至少一种第二聚(A)序列,其中所述第二聚(A)序列是所述载体骨架的原生聚(A)序列或外源聚(A)序列,

其中如果所述第二启动子核苷酸序列不存在,则所述抗原编码核酸序列可操作地连接

到所述至少一种启动子核苷酸序列,并且

其中所述盒不编码免疫显性MHC I类表位,所述免疫显性MHC I类表位:

(1)当以疫苗组合物的形式施用于受试者时,相对于所述盒中编码并能够刺激所述受试者中的免疫应答的另一种MHC I类表位,刺激5倍或更高的免疫应答,和/或

(2)当以疫苗组合物施用于受试者时,相对于当在不存在所述免疫显性MHC I类表位的情况下施用其他MHC I类表位时的免疫应答,降低了对所述盒中编码的另一种MHC I类表位的免疫应答,任选地其中所述免疫应答降低到检测限以下和/或其中所述免疫应答不是治疗有效应答。

199.如权利要求196-198中任一项所述的方法,其中所述表位编码核酸序列来源于所述患有癌症的受试者的肿瘤或来源于受感染受试者的细胞或样品。

200.如权利要求196-198中任一项所述的方法,其中所述表位编码核酸序列不是来源于所述患有癌症的受试者的肿瘤或来源于受感染受试者的细胞或样品。

201.一种用于刺激受试者中的免疫应答的方法,所述方法包括所述方法包括向所述受试者施用基于抗原的疫苗给所述受试者,其中所述基于抗原的疫苗包含抗原编码盒或由所述盒编码的多肽序列,其中所述抗原编码盒包含至少一种由下式从5'至3'描述的抗原编码核酸序列:

$(E_x-(E^N{}_n)_y)_z$

其中,

E表示包含不同表位编码核酸序列的核苷酸序列,

n表示单独的不同表位编码核酸序列的数目并且是包括0的任何整数,

E^N表示包含对于每个相应n的单独的不同表位编码核酸序列的核苷酸序列,

对于z的每次迭代:x=0或1,对于每个n,y=0或1,和x或y中的至少一者=1,并且

z=2或更大,其中所述抗原编码核酸序列包含E、给定E^N或其组合的至少两次迭代,并且

包含所述至少两次迭代的所述不同表位编码核酸序列中的至少一者编码KRAS相关MHC I类新表位。

202.一种用于刺激受试者中的免疫应答的方法,所述方法包括所述方法包括向所述受试者施用基于抗原的疫苗给所述受试者,其中所述基于抗原的疫苗包含:

抗原表达系统,所述抗原表达系统包含:

所述抗原表达系统,

其中所述抗原表达系统包含一种或多种载体,

所述一种或多种载体包含:

(a)载体骨架,其中所述骨架包含:

(i)至少一种启动子核苷酸序列,和

(ii)任选地,至少一种聚腺苷酸化(聚(A))序列;和

(b)盒,其中所述盒包含:

(i)至少一种抗原编码核酸序列,所述抗原编码核酸序列包含:

(I)编码KRAS相关MHC I类新表位的表位编码核酸序列,并且

其中所述表位编码核酸序列中的每一者包含:

(A)任选地,5'接头序列,和

(B)任选地,3'接头序列;

(ii)任选地,可操作地连接到所述抗原编码核酸序列的第二启动子核苷酸序列;和

(iii)任选地,至少一种MHC II类表位编码核酸序列;

(iv)任选地,至少一种编码GPGPG氨基酸接头序列(SEQ IDNO:56)的核酸序列;和

(v)任选地,至少一种第二聚(A)序列,其中所述第二聚(A)序列是所述载体骨架的原生聚(A)序列或外源聚(A)序列,

其中如果所述第二启动子核苷酸序列不存在,则所述抗原编码核酸序列可操作地连接到所述至少一种启动子核苷酸序列,并且

其中所述至少一种抗原编码核酸序列包含编码所述KRAS相关MHC I类新表位的所述表位编码核酸序列中的至少一者的至少两次迭代。

203.一种用于治疗患有疾病的受试者的方法,任选地其中所述疾病是感染性癌症,所述方法包括向所述受试者施用基于抗原的疫苗给所述受试者,其中所述基于抗原的疫苗包含抗原表达系统,所述抗原表达系统包含:

所述抗原表达系统,

其中所述抗原表达系统包含一种或多种载体,

所述一种或多种载体包含:

(a)载体骨架,其中所述骨架包含:

(i)至少一种启动子核苷酸序列,和

(ii)任选地,至少一种聚腺苷酸化(聚(A))序列;和

(b)盒,其中所述盒包含:

(i)至少一种抗原编码核酸序列,所述抗原编码核酸序列包含:

(I)至少两种不同表位编码核酸序列,任选地包含:(1)使所编码的表位序列与由野生型核酸序列编码的相应肽序列不同的至少一种改变,任选地其中所述至少一种改变是KRAS突变,或(2)编码感染性疾病生物体肽的核酸序列,所述肽选自由以下组成的组:病原体衍生肽、病毒衍生肽、细菌衍生肽、真菌衍生肽和寄生虫衍生肽,并且

其中所述表位编码核酸序列中的每一者包含:

(A)任选地,5'接头序列,和

(B)任选地,3'接头序列;

(ii)任选地,可操作地连接到所述抗原编码核酸序列的第二启动子核苷酸序列;和

(iii)任选地,至少一种MHC II类表位编码核酸序列;

(iv)任选地,至少一种编码GPGPG氨基酸接头序列(SEQ IDNO:56)的核酸序列;和

(v)任选地,至少一种第二聚(A)序列,其中所述第二聚(A)序列是所述载体骨架的原生聚(A)序列或外源聚(A)序列,

其中如果所述第二启动子核苷酸序列不存在,则所述抗原编码核酸序列可操作地连接到所述至少一种启动子核苷酸序列,并且

其中所述盒不编码免疫显性MHC I类表位,所述免疫显性MHC I类表位:

(1)当以疫苗组合物施用于受试者时,相对于所述盒中编码并能够刺激所述受试者中的免疫应答的另一种MHC I类表位,刺激5倍或更高的免疫应答,和/或

(2)当以疫苗组合物施用于受试者时,相对于当在不存在所述免疫显性MHC I类表位的

情况下施用其他MHC I类表位时的免疫应答,降低了对所述盒中编码的另一种MHC I类表位的免疫应答,任选地其中所述免疫应答降低到检测限以下和/或其中所述免疫应答不是治疗有效应答。

204.如权利要求196-203中任一项所述的方法,其中所述受试者表达经预测或已知呈递至少一种表位序列的至少一种HLA等位基因,任选地其中所述经预测或已知呈递的至少一种表位序列包含(1)所述KRAS相关MHC I类新表位,和/或(2)所述免疫显性MHC I类表位和所述盒中编码的所述其他MHC I类表位。

205.如权利要求196-203中任一项所述的方法,其中所述受试者表达经预测或已知呈递至少一种表位序列的至少一种HLA等位基因,并且其中所述至少一种表位序列包含已知或怀疑由MHC I类呈递在细胞表面上的表位,任选地其中所述经预测或已知呈递的至少一种表位序列包含(1)所述KRAS相关MHC I类新表位,和/或(2)所述免疫显性MHC I类表位和所述盒中编码的所述其他MHC I类表位。

206.如权利要求205所述的方法,其中所述细胞表面是肿瘤细胞表面。

207.如权利要求206所述的方法,其中所述细胞是选自由以下组成的组的肿瘤细胞:肺癌、黑色素瘤、乳腺癌、卵巢癌、前列腺癌、肾癌、胃癌、结肠癌、睾丸癌、头颈癌、胰腺癌、脑癌、B细胞淋巴瘤、急性髓细胞白血病、慢性髓细胞白血病、慢性淋巴细胞白血病、T细胞淋巴细胞白血病、非小细胞肺癌和小细胞肺癌。

208.如权利要求205所述的方法,其中所述细胞表面是受感染细胞表面。

209.如权利要求208所述的方法,其中所述细胞是选自由以下各项组成的组的受感染细胞:病原体感染细胞、病毒感染细胞、细菌感染细胞、真菌感染细胞和寄生虫感染细胞。

210.如权利要求209所述的方法,其中所述病毒感染细胞选自由以下各项组成的组:HIV感染细胞、严重急性呼吸综合征相关冠状病毒(SARS)感染细胞、严重急性呼吸综合征冠状病毒2(SARS-CoV-2)感染细胞、埃博拉感染细胞、乙型肝炎病毒(HBV)感染细胞、流感感染细胞、正粘病毒科病毒感染细胞、人乳头瘤病毒(HPV)感染细胞、巨细胞病毒(CMV)感染细胞、基孔肯雅病毒感染细胞、呼吸道合胞病毒(RSV)感染细胞、登革热病毒感染细胞和丙型肝炎病毒(HCV)感染细胞。

211.一种用于诱导受试者中的免疫应答的方法,所述方法包括向所述受试者施用基于抗原的疫苗给所述受试者,其中所述基于抗原的疫苗包含抗原编码盒或由所述盒编码的多肽序列,其中所述抗原编码盒包含至少一种由下式从5'至3'描述的抗原编码核酸序列:

$(E_x-(E^N_n)_y)_z$

其中,

E表示包含不同表位编码核酸序列的核苷酸序列,

n表示单独的不同表位编码核酸序列的数目并且是包括0的任何整数,

E^N表示包含对于每个相应n的单独的不同表位编码核酸序列的核苷酸序列,

对于z的每次迭代:x=0或1,对于每个n,y=0或1,和x或y中的至少一者=1,并且

z=2或更大,其中所述抗原编码核酸序列包含E、给定E^N或其组合的至少两次迭代,并且

包含所述至少两次迭代的所述不同表位编码核酸序列中的至少一者编码KRAS相关MHC I类新表位。

212.一种用于诱导受试者中的免疫应答的方法,所述方法包括向所述受试者施用基于

抗原的疫苗给所述受试者,其中所述基于抗原的疫苗包含:

抗原表达系统,所述抗原表达系统包含:

所述抗原表达系统,

其中所述抗原表达系统包含一种或多种载体,

所述一种或多种载体包含:

(a)载体骨架,其中所述骨架包含:

(i)至少一种启动子核苷酸序列,和

(ii)任选地,至少一种聚腺苷酸化(聚(A))序列;和

(b)盒,其中所述盒包含:

(i)至少一种抗原编码核酸序列,所述抗原编码核酸序列包含:

(I)编码KRAS相关MHC I类新表位的表位编码核酸序列,并且

其中所述表位编码核酸序列中的每一者包含:

(A)任选地,5'接头序列,和

(B)任选地,3'接头序列;

(ii)任选地,可操作地连接到所述抗原编码核酸序列的第二启动子核苷酸序列;和

(iii)任选地,至少一种MHC II类表位编码核酸序列;

(iv)任选地,至少一种编码GPGPG氨基酸接头序列(SEQ IDNO:56)的核酸序列;和

(v)任选地,至少一种第二聚(A)序列,其中所述第二聚(A)序列是所述载体骨架的原生聚(A)序列或外源聚(A)序列,

其中如果所述第二启动子核苷酸序列不存在,则所述抗原编码核酸序列可操作地连接到所述至少一种启动子核苷酸序列,

其中所述至少一种抗原编码核酸序列包含编码所述KRAS相关MHC I类新表位的所述表位编码核酸序列的至少两次迭代,并且

其中所述受试者表达经预测或已知呈递所述至少一种KRAS相关MHC I类新表位的至少一种HLA等位基因。

213.一种用于诱导受试者中的免疫应答的方法,所述方法包括向所述受试者施用基于抗原的疫苗给所述受试者,其中所述基于抗原的疫苗包含:

抗原表达系统,所述抗原表达系统包含:

所述抗原表达系统,

其中所述抗原表达系统包含一种或多种载体,

所述一种或多种载体包含:

(a)载体骨架,其中所述骨架包含:

(i)至少一种启动子核苷酸序列,和

(ii)任选地,至少一种聚腺苷酸化(聚(A))序列;和

(b)盒,其中所述盒包含:

(i)至少一种抗原编码核酸序列,所述抗原编码核酸序列包含:

(I)至少一种表位编码核酸序列,任选地包含:(1)使所编码的表位序列与由野生型核酸序列编码的相应肽序列不同的至少一种改变,任选地其中所述至少一种改变是KRAS突变,或(2)编码感染性疾病生物体肽的核酸序列,所述肽选自由以下组成的组:病原体衍生

肽、病毒衍生肽、细菌衍生肽、真菌衍生肽和寄生虫衍生肽，并且任选地其中所述表位编码核酸序列编码MHC I类表位，并且

其中所述表位编码核酸序列中的每一者包含：

(A)任选地，5'接头序列，和

(B)任选地，3'接头序列；

(ii)任选地，可操作地连接到所述抗原编码核酸序列的第二启动子核苷酸序列；和

(iii)任选地，至少一种MHC II类表位编码核酸序列；

(iv)任选地，至少一种编码GPGPG氨基酸接头序列(SEQ IDNO:56)的核酸序列；和

(v)任选地，至少一种第二聚(A)序列，其中所述第二聚(A)序列是所述载体骨架的原生聚(A)序列或外源聚(A)序列，

其中如果所述第二启动子核苷酸序列不存在，则所述抗原编码核酸序列可操作地连接到所述至少一种启动子核苷酸序列，并且

其中所述盒不编码免疫显性MHC I类表位，所述免疫显性MHC I类表位：

(1)当以疫苗组合物施用于受试者时，相对于所述盒中编码并能够刺激所述受试者中的免疫应答的另一种MHC I类表位，刺激5倍或更高的免疫应答，和/或

(2)当以疫苗组合物施用于受试者时，相对于当在不存在所述免疫显性MHC I类表位的情况下施用其他MHC I类表位时的免疫应答，降低了对所述盒中编码的另一种MHC I类表位的免疫应答，任选地其中所述免疫应答降低到检测限以下和/或其中所述免疫应答不是治疗有效应答，并且

其中所述受试者表达经预测或已知呈递所述免疫显性MHC I类表位和所述盒中编码的所述其他MHC I类表位两者的至少一种HLA等位基因。

214.如权利要求196-213中任一项所述的方法，其中所述抗原表达系统包含如权利要求1-160中任一项所述的抗原表达系统中的任一者。

215.如权利要求196-213中任一项所述的方法，其中所述基于抗原的疫苗包含如权利要求161-164中任一项所述的药物组合物中的任一者。

216.如权利要求196-215中任一项所述的方法，其中所述基于抗原的疫苗作为初免剂量施用。

217.如权利要求196-216中任一项所述的方法，其中所述基于抗原的疫苗作为一个或多个加强剂量施用。

218.如权利要求217所述的方法，其中所述加强剂量与所述初免剂量不同。

219.如权利要求218所述的方法，其中：

a)所述初免剂量包含黑猩猩腺病毒载体并且所述加强剂量包含甲病毒载体；或

b)所述初免剂量包含甲病毒载体载体并且所述加强剂量包含黑猩猩腺病毒载体。

220.如权利要求217所述的方法，其中所述加强剂量与所述初免剂量相同。

221.如权利要求217-220中任一项所述的方法，其中所述一个或多个加强剂量的注射部位尽可能靠近所述初免剂量的注射部位。

222.如上述方法权利要求中任一项所述的方法，所述方法还包括确定或已经确定所述受试者的HLA-单倍型。

223.如上述方法权利要求中任一项所述的方法，其中所述基于抗原的疫苗是肌内

(IM)、皮内(ID)、皮下(SC)或静脉内(IV)施用的。

224.如上述方法权利要求中任一项所述的方法,其中所述基于抗原的疫苗是肌内(IM)施用的。

225.如权利要求224所述的方法,其中所述IM施用是在单独的注射部位施用的。

226.如权利要求225所述的方法,其中所述单独的注射部位是在相对的三角肌中。

227.如权利要求226所述的方法,其中所述单独的注射部位是在每一侧的臀肌或股直肌部位中。

228.如上述权利要求中任一项所述的方法或组合物,其中所述KRAS相关MHC I类新表位或所述KRAS突变包含KRAS G12C突变、KRAS G12V突变、KRAS G12D突变或KRAS Q61H突变。

229.如上述权利要求中任一项所述的方法或组合物,其中所述KRAS相关MHC I类新表位或所述KRAS突变包含以SEQ ID NO:75-82所示的氨基酸序列中的任一者。

230.如上述权利要求中任一项所述的方法或组合物,其中所述抗原编码盒包含以SEQ ID NO:75-82所示的氨基酸序列中的每一者。

231.如上述权利要求中任一项所述的方法或组合物,其中所述抗原编码盒包含以SEQ ID NO:75-82所示的氨基酸序列中的每一者的两次或更多次迭代,任选地包含以SEQ ID NO:75-82所示的氨基酸序列中的每一者的4次迭代。

232.如上述权利要求中任一项所述的方法或组合物,其中所述KRAS相关MHC I类新表位或所述KRAS突变包含以SEQ ID NO:57、SEQ ID NO:58、SEQ ID NO:59或SEQ ID NO:60所示的氨基酸序列。

233.如上述权利要求中任一项所述的方法或组合物,其中所述表位编码核酸序列包含两个或更多个不同的表位编码核酸序列,所述表位编码核酸序列独立地编码不同的KRAS相关MHC I类新表位或不同的KRAS突变。

234.如上述权利要求中任一项所述的方法或组合物,其中所述表位编码核酸序列中的每一者独立地编码不同的KRAS相关MHC I类新表位或不同的KRAS突变。

235.如上述权利要求中任一项所述的方法或组合物,其中所述表位编码核酸序列包含两个或更多个不同的表位编码核酸序列,所述表位编码核酸序列独立地编码KRAS G12C突变、KRAS G12V突变、KRAS G12D突变或KRAS Q61H突变。

236.如上述权利要求中任一项所述的方法或组合物,其中所述表位编码核酸序列独立地编码KRAS G12C突变、KRAS G12V突变和KRAS G12D突变以及任选地KRAS Q61H突变中的每一者。

237.如上述权利要求中任一项所述的方法或组合物,其中所述抗原编码核酸序列编码包含以SEQ ID NO:64或SEQ ID NO:65所示的氨基酸序列的肽。

238.如上述权利要求中任一项所述的方法或组合物,其中所述盒不编码免疫显性MHC I类表位,所述免疫显性MHC I类表位:

(1)当以疫苗组合物施用于受试者时,相对于所述盒中编码并能够刺激所述受试者中的免疫应答的另一种MHC I类表位,刺激5倍或更高的免疫应答,和/或

(2)当以疫苗组合物施用于受试者时,相对于当在不存在所述免疫显性MHC I类表位的情况下施用其他MHC I类表位时的免疫应答,降低了对所述盒中编码的另一种MHC I类表位的免疫应答,任选地其中所述免疫应答降低到检测限以下和/或其中所述免疫应答不是治

疗有效应答。

239.如权利要求236所述的方法或组合物,其中所述盒不编码免疫显性MHC I类表位,所述免疫显性MHC I类表位当以疫苗组合物的形式施用于受试者时,相对于所述盒中编码并能够刺激所述受试者中的免疫应答的KRAS相关新表位,刺激5倍或更高的免疫应答。

240.如权利要求236所述的方法或组合物,其中所述盒不编码免疫显性MHC I类表位,所述免疫显性MHC I类表位当以疫苗组合物施用于受试者时,相对于当在不存在所述免疫显性MHC I类表位的情况下施用所述其他MHC I类表位时的免疫应答,降低了对所述盒中编码的另一种MHC I类表位的免疫应答。

241.如权利要求236所述的方法或组合物,其中所述盒不编码免疫显性MHC I类表位,所述免疫显性MHC I类表位当以疫苗组合物施用于受试者时,相对于当在不存在所述免疫显性MHC I类表位的情况下施用所述其他MHC I类表位时的免疫应答,将对所述盒中编码的另一个MHC I类表位的免疫应答降低至检测限以下。

242.如权利要求236所述的方法或组合物,其中所述盒不编码免疫显性MHC I类表位,所述免疫显性MHC I类表位当以疫苗组合物施用于受试者时,相对于当在不存在所述免疫显性MHC I类表位的情况下施用所述其他MHC I类表位时的免疫应答,降低了对所述盒中编码的另一个MHC I类表位的免疫应答,其中对所述其他MHC I类表位的免疫应答不是治疗有效应答。

243.如上述权利要求中任一项所述的方法或组合物,其中所述免疫显性表位是TP53相关MHC I类新表位,任选地其中所述TP53相关MHC I类新表位包含S127Y突变。

多表位疫苗盒

[0001] 相关申请的交叉引用

[0002] 本申请要求于2020年8月6日提交的美国临时申请第63/062,268号的权益,所述临时申请特此出于所有目的通过引用整体并入。

[0003] 序列表

[0004] 本申请含有序列表,所述序列表已以ASCII格式以电子方式提交并特此通过引用整体并入。所述ASCII副本创建于2021年11月17日,命名为GSO-095WO_SL.txt并且大小为456,892字节。

背景技术

[0005] 基于肿瘤特异性抗原的治疗性疫苗作为下一代个性化癌症免疫疗法具有极大的前景。[1-3]例如,具有高突变负荷的癌症,诸如非小细胞肺癌(NSCLC)和黑色素瘤,鉴于新抗原产生的可能性相对较大而为这种疗法的特别有吸引力的靶标。[4,5]早期证据表明,基于新抗原的疫苗接种可引发T细胞应答[6],并且靶向新抗原的细胞疗法可在选定患者的某些情形下引起肿瘤消退。[7]

[0006] 在癌症和感染性疾病两种情形中抗原疫苗设计的一个问题是存在的许多编码突变中的哪一个产生“最佳”治疗性抗原,例如,可引发免疫性的抗原。

[0007] 除了当前抗原预测方法的挑战之外,可用于人类抗原递送的可用载体系统也存在某些挑战,其中许多来源于人类。例如,由于先前的自然暴露,许多人对人类病毒具有预先存在的免疫性,并且这种免疫性可为使用重组人类病毒用于疫苗接种策略中的抗原递送的主要障碍,诸如在癌症治疗或针对感染性疾病的疫苗接种中。尽管在解决上述问题的疫苗接种策略上取得了一些进展,但仍需要改进,特别是对于临床应用,诸如改进的疫苗效力和功效。

发明内容

[0008] 本文公开了:一种抗原编码盒或由所述盒编码的多肽序列,其中所述抗原编码盒包含至少一种由下式从5’至3’描述的抗原编码核酸序列:

[0009] $(E_x-(E^N{}_n)_y)_z$

[0010] 其中E表示包含不同表位编码核酸序列的核苷酸序列,n表示单独的不同表位编码核酸序列的数目并且是包括0的任何整数,E^N表示包含对于每个相应n的单独的不同表位编码核酸序列的核苷酸序列,对于z的每次迭代:x=0或1,对于每个n,y=0或1,和x或y中的至少一者=1,并且z=2或更大,其中所述抗原编码核酸序列包含E、给定E^N或其组合的至少两次迭代,并且包含所述至少两次迭代的所述不同表位编码核酸序列中的至少一者编码不同的KRAS相关MHC I类新表位。

[0011] 在一些方面,所述抗原编码盒编码氨基酸序列VVVGACGVGK(SEQ ID NO:75)、VVVGADGVGK(SEQ ID NO:78)、VVGAVGVGK(SEQ ID NO:79)和ILDTAGHEEY(SEQ ID NO:82)中的每一者的至少4次迭代。在一些方面,所述KRAS相关MHC I类新表位或KRAS突变包含KRAS

G12C突变、KRAS G12V突变、KRAS G12D突变或KRAS Q61H突变。在一些方面，所述KRAS相关MHC I类新表位或KRAS突变包含以SEQ ID NO:75-82所示的氨基酸序列中的任一者。在一些方面，所述抗原编码盒包含以SEQ ID NO:75-82所示的氨基酸序列中的每一者。在一些方面，所述抗原编码盒包含以SEQ ID NO:75-82所示的氨基酸序列中的每一者的两次或更多次迭代。在一些方面，所述抗原编码盒包含以SEQ ID NO:75-82所示的氨基酸序列中的每一者的4次迭代。在一些方面，所述KRAS相关MHC I类新表位或KRAS突变包含以SEQ ID NO:57、SEQ ID NO:58、SEQ ID NO:59或SEQ ID NO:60所示的氨基酸序列。在一些方面，所述表位编码核酸序列包含两个或更多个不同的表位编码核酸序列，所述表位编码核酸序列独立地编码不同的KRAS相关MHC I类新表位或不同的KRAS突变。在一些方面，所述表位编码核酸序列中的每一者独立地编码不同的KRAS相关MHC I类新表位或不同的KRAS突变。在一些方面，所述表位编码核酸序列包含两个或更多个不同的表位编码核酸序列，所述表位编码核酸序列独立地编码KRAS G12C突变、KRAS G12V突变、KRAS G12D突变或KRAS Q61H突变。在一些方面，所述表位编码核酸序列独立地编码KRAS G12C突变、KRAS G12V突变和KRAS G12D突变以及任选地KRAS Q61H突变中的每一者。在一些方面，所述抗原编码核酸序列编码包含以SEQ ID NO:64或SEQ ID NO:65所示的氨基酸序列的肽。在一些方面，所述抗原编码核酸序列编码包含以SEQ ID NO:65所示的氨基酸序列的肽。

[0012] 在一些方面，包含至少两次迭代的所述不同表位编码核酸序列中的至少两者编码不同的KRAS相关MHC I类新表位。在一些方面，包含至少两次迭代的所述不同表位编码核酸序列中的至少3个、至少4个、至少5个、至少6个、至少7个或至少8个编码不同的KRAS相关MHC I类新表位。在一些方面，包含至少两次迭代的所述不同表位编码核酸序列中的每一者编码不同的KRAS相关MHC I类新表位。在一些方面，编码所述KRAS相关MHC I类新表位的核酸序列中的一者或多者包含至少2次、至少3次、至少4次、至少5次、至少6次、至少7次或至少8次迭代。在一些方面，编码所述KRAS相关MHC I类新表位的核酸序列中的每一者包含至少2次、至少3次、至少4次、至少5次、至少6次、至少7次或至少8次迭代。在一些方面，编码所述不同KRAS相关MHC I类新表位的核酸序列中的一者或多者包含至少4次迭代。在一些方面，编码所述不同KRAS相关MHC I类新表位的核酸序列中的每一者包含至少4次迭代。在一些方面，所述不同KRAS相关MHC I类新表位中的一者或多者独立地包含KRAS G12C突变、KRAS G12V突变、KRAS G12D突变或KRAS Q61H突变。

[0013] 在一些方面，每个E或E^N独立地包含由式($L5_b$-N_c-$L3_d$)从5'至3'描述的核苷酸序列，其中N包含与每个E或E^N相关的不同表位编码核酸序列，其中c＝1，L5包含5'接头序列，其中b＝0或1，并且L3包含3'接头序列，其中d＝0或1。在一些方面，每个N编码长度为7-15个氨基酸的表位，L5是编码所述表位的原生N末端氨基酸序列的原生5'接头序列，并且其中所述5'接头序列编码长度为至少2个氨基酸的肽，并且L3是编码所述表位的原生C末端氨基酸序列的原生3'接头序列，并且其中所述3'接头序列编码长度为至少32个氨基酸的肽。在一些方面，所述5'和/或3'接头序列编码长度为至少3个氨基酸的肽。在一些方面，所述5'和/或3'接头序列编码长度为至少4个氨基酸的肽。在一些方面，所述5'和/或3'接头序列编码长度为至少5个氨基酸的肽。在一些方面，所述5'和/或3'接头序列编码长度为至少8个氨基酸的肽。在一些方面，所述5'和/或3'接头序列编码长度为至少2-8个氨基酸的肽。在一些方面，所述5'和/或3'接头序列编码长度为至少2-10个氨基酸的肽。

[0014] 在一些方面，每个E和E^N编码长度为至少7个氨基酸的表位。在一些方面，每个E和E^N编码长度为7-15个氨基酸的表位。在一些方面，每个E和E^N是长度为至少21个核苷酸的核苷酸序列。在一些方面，每个E和E^N是长度为75个核苷酸的核苷酸序列。

[0015] 本文还公开了一种用于递送抗原表达系统的组合物，所述组合物包含：所述抗原表达系统，其中所述抗原表达系统包含一种或多种载体，所述一种或多种载体包含：(a)载体骨架，其中所述骨架包含：(i)至少一种启动子核苷酸序列，和(ii)任选地，至少一种聚腺苷酸化(聚(A))序列；和(b)盒，其中所述盒包含：(i)至少一种抗原编码核酸序列，所述抗原编码核酸序列包含：(I)编码KRAS相关MHC I类新表位的表位编码核酸序列，并且其中所述表位编码核酸序列中的每一者包含：(A)任选地，5'接头序列，和(B)任选地，3'接头序列；(ii)任选地，可操作地连接到所述抗原编码核酸序列的第二启动子核苷酸序列；和(iii)任选地，至少一种MHC II类表位编码核酸序列；(iv)任选地，至少一种编码GPGPG氨基酸接头序列(SEQ ID NO:56)的核酸序列；和(v)任选地，至少一种第二聚(A)序列，其中所述第二聚(A)序列是所述载体骨架的原生聚(A)序列或外源聚(A)序列，其中如果所述第二启动子核苷酸序列不存在，则所述抗原编码核酸序列可操作地连接到所述至少一种启动子核苷酸序列，并且其中所述至少一种抗原编码核酸序列包含编码所述KRAS相关MHC I类新表位的所述表位编码核酸序列的至少两次迭代。

[0016] 本文还公开了一种用于递送抗原表达系统的组合物，所述组合物包含：所述抗原表达系统，其中所述抗原表达系统包含一种或多种载体，所述一种或多种载体包含：(a)载体骨架，其中所述骨架包含：(i)至少一种启动子核苷酸序列，和(ii)至少一种聚腺苷酸化(聚(A))序列；和(b)盒，其中所述盒包含：(i)至少一种抗原编码核酸序列，所述抗原编码核酸序列包含：(I)彼此线性连接的至少2、3、4、5、6、7、8、9或10种不同表位编码核酸序列，其中所述不同表位编码核酸序列中的至少一者编码KRAS相关MHC I类新表位，并且其中所述表位编码核酸序列中的每一者包含：(A)任选地，5'接头序列，和(B)任选地，3'接头序列；(ii)任选地，可操作地连接到所述抗原编码核酸序列的第二启动子核苷酸序列；(iii)任选地，至少一种MHC II类表位编码核酸序列；(iv)任选地，至少一种编码GPGPG氨基酸接头序列(SEQ ID NO:56)的核酸序列；和(v)任选地，至少一种第二聚(A)序列，其中所述第二聚(A)序列是所述载体骨架的原生聚(A)序列或外源聚(A)序列，其中如果所述第二启动子核苷酸序列不存在，则所述抗原编码核酸序列可操作地连接到所述至少一种启动子核苷酸序列，并且其中所述至少一种抗原编码核酸序列包含编码所述KRAS相关MHC I类新表位的所述不同表位编码核酸序列中的至少一者的至少两次迭代。

[0017] 在一些方面，所述至少一种抗原编码核酸序列包含至少3种不同表位编码核酸序列。

[0018] 本文还公开了一种用于递送抗原表达系统的组合物，所述组合物包含：所述抗原表达系统，其中所述抗原表达系统包含一种或多种载体，所述一种或多种载体包含：(a)载体骨架，其中所述载体骨架包含黑猩猩腺病毒载体，任选地其中所述黑猩猩腺病毒载体是ChAdV68载体，或甲病毒载体，任选地其中所述甲病毒载体是委内瑞拉马脑炎病毒载体；和(b)盒，任选地其中所述盒整合在所述载体骨架原生的原生启动子核苷酸序列和聚(A)序列之间，任选地其中所述聚(A)序列是所述载体骨架原生的，其中所述盒包含：(i)至少一种抗原编码核酸序列，所述抗原编码核酸序列包含：(I)编码KRAS相关MHC I类新表位的表位编

码核酸序列,任选地包含彼此线性连接的至少两种不同表位编码核酸序列,每个表位编码核酸序列任选地包含:(A)MHC I类表位编码核酸序列,其中所述MHC I类表位编码核酸序列编码长度为7-15个氨基酸的MHC I类表位,(B)5'接头序列,其中所述5'接头序列编码所述MHC I类表位的原生N末端氨基酸序列,并且其中所述5'接头序列编码长度为至少2个氨基酸的肽,(C)3'接头序列,其中所述3'接头序列编码所述MHC I类表位的原生C末端酸序列,并且其中所述3'接头序列编码长度为至少2个氨基酸的肽,并且其中所述盒可操作地连接到所述原生启动子核苷酸序列,其中所述表位编码核酸序列中的每一者编码长度为13至25个氨基酸的多肽,并且其中每个表位编码核酸序列的每个3'端连接到以下表位编码核酸序列的5'端,所述盒中的最终表位编码核酸序列除外;和(ii)至少两种MHC II类表位编码核酸序列,所述MHC II类表位编码核酸序列包含:(I)PADRE MHC II类序列(SEQ ID NO:48),(II)破伤风类毒素MHC II类序列(SEQ ID NO:46),(III)编码连接所述PADRE MHC II类序列与所述破伤风类毒素MHC II类序列的GPGPG氨基酸接头序列(SEQ ID NO:56)的第一核酸序列,(IV)编码连接所述至少两种MHC II类表位编码核酸序列的5'端与所述表位编码核酸序列的GPGPG氨基酸接头序列(SEQ ID NO:56)的第二核酸序列,(V)任选地,在所述至少两种MHC II类表位编码核酸序列的3'端编码GPGPG氨基酸接头序列(SEQ ID NO:56)的第三核酸序列;(iii)任选地,可操作地连接到所述抗原编码核酸序列的第二启动子核苷酸序列;并且其中如果所述第二启动子核苷酸序列不存在,则所述抗原编码核酸序列可操作地连接到所述原生启动子核苷酸序列,并且其中所述至少一种抗原编码核酸序列包含编码所述KRAS相关MHC I类新表位的所述表位编码核酸序列的至少两次迭代。

[0019] 在一些方面,所述盒的每个元件的有序序列描述于下式,从5'至3'包含:

[0020] P_a-($L5_b$-N_c-$L3_d$)$_X$-($G5_e$-U_f)$_Y$-$G3_g$

[0021] 其中P包含所述第二启动子核苷酸序列,其中a=0或1,N包含所述不同表位编码核酸序列之一,其中c=1,L5包含所述5'接头序列,其中b=0或1,L3包含所述3'接头序列,其中d=0或1,G5包含至少一种编码GPGPG氨基酸接头(SEQ ID NO:56)的核酸序列之一,其中e=0或1,G3包含至少一种编码GPGPG氨基酸接头(SEQ ID NO:56)的核酸序列之一,其中g=0或1,U包含至少一种MHC II类表位编码核酸序列之一,其中f=1,X=1至400,其中对于每个X,相应的N_c是表位编码核酸序列,并且Y=0、1或2,其中对于每个Y,相应的U_f是MHC II类表位编码核酸序列。

[0022] 在一些方面,对于每个X,相应的N_c是不同表位编码核酸序列,除了对应于所述不同表位编码核酸序列的至少两次迭代的N_c。在一些方面,对于每个Y,相应的U_f是不同MHC II类表位编码核酸序列。在一些方面,a=0,b=1,d=1,e=1,g=1,h=1,X=16,Y=2,所述至少一种启动子核苷酸序列是所述载体骨架原生的单个原生启动子核苷酸序列,所述至少一种聚腺苷酸化聚(A)序列是由所述载体骨架提供的至少80个连续A核苷酸的聚(A)序列,每个N编码长度为7-15个氨基酸的表位,L5是编码所述表位的原生N末端氨基酸序列的原生5'接头序列,并且其中所述5'接头序列编码长度为至少2个氨基酸的肽,L3是编码所述表位的原生C末端氨基酸序列的原生3'接头序列,并且其中所述3'接头序列编码长度为至少2个氨基酸的肽,U是PADRE II类序列和破伤风类毒素MHC II类序列中的每一者,所述载体骨架包含黑猩猩腺病毒载体,任选地其中所述黑猩猩腺病毒载体是ChAdV68载体,或甲病毒载体,任选地其中所述甲病毒载体是委内瑞拉马脑炎病毒载体,任选地其中当所述载体骨架包含

甲病毒载体时，所述原生启动子核苷酸序列是亚基因组(例如26S)启动子，并且所述MHC II类表位编码核酸序列中的每一者编码长度为13至25个氨基酸的多肽。

[0023] 在一些方面，所述抗原编码盒编码氨基酸序列VVVGACGVGK(SEQ ID NO:75)、VVVGADGVGK(SEQ ID NO:78)、VVGAVGVGK(SEQ ID NO:79)和ILDTAGHEEY(SEQ ID NO:82)中的每一者的至少4次迭代。在一些方面，所述KRAS相关MHC I类新表位或KRAS突变包含KRAS G12C突变、KRAS G12V突变、KRAS G12D突变或KRAS Q61H突变。在一些方面，所述KRAS相关MHC I类新表位或KRAS突变包含以SEQ ID NO:75-82所示的氨基酸序列中的任一者。在一些方面，所述抗原编码盒包含以SEQ ID NO:75-82所示的氨基酸序列中的每一者。在一些方面，所述抗原编码盒包含以SEQ ID NO:75-82所示的氨基酸序列中的每一者的两次或更多次迭代。在一些方面，所述抗原编码盒包含以SEQ ID NO:75-82所示的氨基酸序列中的每一者的4次迭代。在一些方面，所述KRAS相关MHC I类新表位或KRAS突变包含以SEQ ID NO:57、SEQ ID NO:58、SEQ ID NO:59或SEQ ID NO:60所示的氨基酸序列。在一些方面，所述表位编码核酸序列包含两个或更多个不同的表位编码核酸序列，所述表位编码核酸序列独立地编码不同的KRAS相关MHC I类新表位或不同的KRAS突变。在一些方面，所述表位编码核酸序列中的每一者独立地编码不同的KRAS相关MHC I类新表位或不同的KRAS突变。在一些方面，所述表位编码核酸序列包含两个或更多个不同的表位编码核酸序列，所述表位编码核酸序列独立地编码KRAS G12C突变、KRAS G12V突变、KRAS G12D突变或KRAS Q61H突变。在一些方面，所述表位编码核酸序列独立地编码KRAS G12C突变、KRAS G12V突变和KRAS G12D突变以及任选地KRAS Q61H突变中的每一者。在一些方面，所述抗原编码核酸序列编码包含以SEQ ID NO:64或SEQ ID NO:65所示的氨基酸序列的肽。在一些方面，所述抗原编码核酸序列编码包含以SEQ ID NO:65所示的氨基酸序列的肽。

[0024] 在一些方面，所述至少两次迭代是至少3次、至少4次、至少5次、至少6次、至少7次或至少8次迭代。在一些方面，所述至少两次迭代是至少8次迭代。在一些方面，所述至少两次迭代是至少8次、至少9次、至少10次、至少11次、至少12次、至少13次、至少14次、至少15次、至少16次、至少17次、至少18次、至少19次或至少20次迭代。在一些方面，所述至少两次迭代是2-3次、2-4次、2-5次、2-6次、2-7次迭代或2-8次迭代。在一些方面，所述至少两次迭代是7次迭代或更少、6次迭代或更少、5次迭代或更少、4次迭代或更少或3次迭代或更少。

[0025] 在一些方面，所述至少一种抗原编码核酸序列包含至少两种不同表位编码核酸序列的至少两次迭代。在一些方面，所述至少一种抗原编码核酸序列包含至少3种、至少4种、至少5种、至少6种、至少7种、至少8种、至少9种或至少10种不同表位编码核酸序列的至少两次迭代。在一些方面，所述至少两次迭代被至少一个单独的不同表位编码核酸序列分开。在一些方面，所述至少两次迭代被至少2个单独的不同表位编码核酸序列分开。在一些方面，包括所述任选的5'接头序列和/或所述任选的3'接头序列在内的所述至少两次迭代被至少75个核苷酸分开。在一些方面，包括所述任选的5'接头序列和/或所述任选的3'接头序列在内的所述至少两次迭代被至少150个核苷酸、至少300个核苷酸或至少675个核苷酸分开。在一些方面，包括所述任选的5'接头序列和/或所述任选的3'接头序列在内的所述至少两次迭代被至少50个核苷酸、至少100个核苷酸、至少200个核苷酸、至少250个核苷酸、至少350个核苷酸、至少400个核苷酸、至少450个核苷酸、至少500个核苷酸、至少700个核苷酸、至少700个核苷酸、至少750个核苷酸、至少800个核苷酸、至少900个核苷酸或至少1000个核苷酸

分开。在一些方面，包括所述任选的5'接头序列和/或所述任选的3'接头序列在内的所述至少两次迭代被至少10个核苷酸、至少15个核苷酸、至少20个核苷酸、至少25个核苷酸、至少30个核苷酸、至少35个核苷酸、至少40个核苷酸、至少45个核苷酸、至少50个核苷酸、至少55个核苷酸、至少60个核苷酸、至少65个核苷酸或至少70个核苷酸分开。

[0026] 在一些方面，所述至少一种抗原编码核酸序列从5'至3'由下式描述：

[0027] $(E_x-(E^N_n)_y)_z$

[0028] 其中，E表示包含不同表位编码核酸序列中的至少一者的核苷酸序列，n表示单独的不同表位编码核酸序列的数目并且是包括0的任何整数，E^N表示包含对于每个相应n的单独的不同表位编码核酸序列的核苷酸序列，对于z的每次迭代：x＝0或1，对于每个n，y＝0或1，和x或y中的至少一者＝1，并且z＝2或更大，其中所述抗原编码核酸序列包含E、给定E^N或其组合的至少两次迭代。

[0029] 在一些方面，所述抗原编码盒编码氨基酸序列VVVGACGVGK(SEQ ID NO:75)、VVVGADGVGK(SEQ ID NO:78)、VVGAVGVGK(SEQ ID NO:79)和ILDTAGHEEY(SEQ ID NO:82)中的每一者的至少4次迭代。

[0030] 在一些方面，所述不同的表位编码核酸序列包含至少两个不同的表位编码核酸序列，其各自编码不同的KRAS相关MHC I类新表位。在一些方面，所述不同的表位编码核酸序列包含至少3个、至少4个、至少5个、至少6个、至少7个或至少8个不同的表位编码核酸序列，其各自编码不同的KRAS相关MHC I类新表位。在一些方面，所述至少一种抗原编码核酸序列的所述表位编码核酸序列中的每一者编码不同的KRAS相关MHC I类新表位。在一些方面，编码所述不同的KRAS相关MHC I类新表位的核酸序列中的一者或多者包含至少2次、至少3次、至少4次、至少5次、至少6次、至少7次或至少8次迭代。在一些方面，编码所述不同的KRAS相关MHC I类新表位的核酸序列中的每一者包含至少2次、至少3次、至少4次、至少5次、至少6次、至少7次或至少8次迭代。在一些方面，编码所述不同的KRAS相关MHC I类新表位的核酸序列中的一者或多者包含至少4次迭代。在一些方面，编码所述不同的KRAS相关MHC I类新表位的核酸序列中的每一者包含至少4次迭代。在一些方面，所述不同的KRAS相关MHC I类新表位中的一者或多者独立地包含KRAS G12C突变、KRAS G12V突变、KRAS G12D突变或KRAS Q61H突变。

[0031] 在一些方面，所述至少两次迭代包括多次迭代，或者z包括相对于包含所述表位编码核酸序列的单次迭代的抗原编码核酸序列足以刺激更大免疫应答的数目。在一些方面，所述至少两次迭代包括多次迭代，或者z包括足以刺激免疫应答的数目，并且所述表位编码核酸序列的单次迭代不足以刺激所述免疫应答或不足以刺激可检测的免疫应答。在一些方面，所述免疫应答是用所述用于递送所述抗原表达系统的组合物进行体内免疫后表位特异性T细胞的扩增。在一些方面，所述免疫应答是在用所述用于递送所述抗原表达系统的组合物进行体内免疫后增加的表位特异性T细胞的激活和/或增加的表位特异性T细胞的表位特异性杀伤。

[0032] 本文还提供了一种用于递送抗原表达系统的组合物，所述组合物包含：所述抗原表达系统，其中所述抗原表达系统包含一种或多种载体，所述一种或多种载体包含：(a)载体骨架，其中所述骨架包含：(i)至少一种启动子核苷酸序列，和(ii)任选地，至少一种聚腺苷酸化(聚(A))序列；和(b)盒，其中所述盒包含：(i)至少一种抗原编码核酸序列，所述抗原

编码核酸序列包含:(I)至少两种不同表位编码核酸序列,任选地包含:(1)使所编码的表位序列与由野生型核酸序列编码的相应肽序列不同的至少一种改变,任选地其中所述至少一种改变是KRAS突变,或(2)编码感染性疾病生物体肽的核酸序列,所述肽选自由以下组成的组:病原体衍生肽、病毒衍生肽、细菌衍生肽、真菌衍生肽和寄生虫衍生肽,并且其中所述表位编码核酸序列中的每一者包含:(A)任选地,5'接头序列,和(B)任选地,3'接头序列;(ii)任选地,可操作地连接到所述抗原编码核酸序列的第二启动子核苷酸序列;和(iii)任选地,至少一种MHC II类表位编码核酸序列;(iv)任选地,至少一种编码GPGPG氨基酸接头序列(SEQ ID NO:56)的核酸序列;和(v)任选地,至少一种第二聚(A)序列,其中所述第二聚(A)序列是所述载体骨架的原生聚(A)序列或外源聚(A)序列,其中如果所述第二启动子核苷酸序列不存在,则所述抗原编码核酸序列可操作地连接到所述至少一种启动子核苷酸序列,并且其中所述盒不编码免疫显性MHC I类表位,所述免疫显性MHC I类表位:(1)当以疫苗组合物的形式施用于受试者时,相对于所述盒中编码并能够刺激所述受试者中的免疫应答的另一种MHC I类表位,刺激5倍或更高的免疫应答,和/或(2)当以疫苗组合物的形式施用于受试者时,相对于当在不存在所述免疫显性MHC I类表位的情况下施用其他MHC I类表位时的免疫应答,降低了对所述盒中编码的另一种MHC I类表位的免疫应答,任选地其中所述免疫应答降低到检测限以下和/或其中所述免疫应答不是治疗有效应答。

[0033] 本文还提供了一种用于递送抗原表达系统的组合物,所述组合物包含:所述抗原表达系统,其中所述抗原表达系统包含一种或多种载体,所述一种或多种载体包含:(a)载体骨架,其中所述骨架包含:(i)至少一种启动子核苷酸序列,和(ii)任选地,至少一种聚腺苷酸化(聚(A))序列;和(b)盒,其中所述盒包含:(i)至少一种抗原编码核酸序列,所述抗原编码核酸序列包含:(I)彼此线性连接的至少2、3、4、5、6、7、8、9或10种不同表位编码核酸序列,任选地包含:(1)使所编码的表位序列与由野生型核酸序列编码的相应肽序列不同的至少一种改变,任选地其中所述至少一种改变是KRAS突变,或(2)编码感染性疾病生物体肽的核酸序列,所述肽选自由以下组成的组:病原体衍生肽、病毒衍生肽、细菌衍生肽、真菌衍生肽和寄生虫衍生肽,并且其中所述表位编码核酸序列中的每一者包含:(A)任选地,5'接头序列,和(B)任选地,3'接头序列;(ii)任选地,可操作地连接到所述抗原编码核酸序列的第二启动子核苷酸序列;和(iii)任选地,至少一种MHC II类表位编码核酸序列;(iv)任选地,至少一种编码GPGPG氨基酸接头序列(SEQ ID NO:56)的核酸序列;和(v)任选地,至少一种第二聚(A)序列,其中所述第二聚(A)序列是所述载体骨架的原生聚(A)序列或外源聚(A)序列,并且其中如果所述第二启动子核苷酸序列不存在,则所述抗原编码核酸序列可操作地连接到所述至少一种启动子核苷酸序列,并且其中所述盒不编码免疫显性MHC I类表位,所述免疫显性MHC I类表位:(1)当以疫苗组合物的形式施用于受试者时,相对于所述盒中编码并能够刺激所述受试者中的免疫应答的另一种MHC I类表位,刺激5倍或更高的免疫应答,和/或(2)当以疫苗组合物的形式施用于受试者时,相对于当在不存在所述免疫显性MHC I类表位的情况下施用其他MHC I类表位时的免疫应答,降低了对所述盒中编码的另一种MHC I类表位的免疫应答,任选地其中所述免疫应答降低到检测限以下和/或其中所述免疫应答不是治疗有效应答。

[0034] 在一些方面,所述不同表位编码核酸序列中的至少一者编码KRAS相关MHC I类新表位。

[0035] 本文还提供了一种用于递送抗原表达系统的组合物,所述组合物包含:所述抗原表达系统,其中所述抗原表达系统包含一种或多种载体,所述一种或多种载体包含:(a)载体骨架,其中所述载体骨架包含黑猩猩腺病毒载体,任选地其中所述黑猩猩腺病毒载体是ChAdV68载体,或甲病毒载体,任选地其中所述甲病毒载体是委内瑞拉马脑炎病毒载体;和(b)盒,任选地其中所述盒整合在所述载体骨架原生的原生启动子核苷酸序列和聚(A)序列之间,任选地其中所述聚(A)序列是所述载体骨架原生的,其中所述盒包含:(i)至少一种抗原编码核酸序列,所述抗原编码核酸序列包含:(I)至少一种表位编码核酸序列,任选地包含彼此线性连接的至少两种不同表位编码核酸序列,每个表位编码核酸序列任选地包含:(A)MHC I类表位编码核酸序列,其中所述MHC I类表位编码核酸序列编码长度为7-15个氨基酸的MHC I类表位,(B)5'接头序列,其中所述5'接头序列编码所述MHC I类表位的原生N末端氨基酸序列,并且其中所述5'接头序列编码长度为至少2个氨基酸的肽,(C)3'接头序列,其中所述3'接头序列编码所述MHC I类表位的原生C末端酸序列,并且其中所述3'接头序列编码长度为至少2个氨基酸的肽,并且其中所述盒可操作地连接到所述原生启动子核苷酸序列,其中所述表位编码核酸序列中的每一者编码长度为13至25个氨基酸的多肽,并且其中每个表位编码核酸序列的每个3'端连接到以下表位编码核酸序列的5'端,所述盒中的最终表位编码核酸序列除外;和(ii)至少两种MHC II类表位编码核酸序列,所述MHC II类表位编码核酸序列包含:(I)PADRE MHC II类序列(SEQ ID NO:48),(II)破伤风类毒素MHC II类序列(SEQ ID NO:46),(III)编码连接所述PADRE MHC II类序列与所述破伤风类毒素MHC II类序列的GPGPG氨基酸接头序列(SEQ ID NO:56)的第一核酸序列,(IV)编码连接所述至少两种MHC II类表位编码核酸序列的5'端与所述表位编码核酸序列的GPGPG氨基酸接头序列(SEQ ID NO:56)的第二核酸序列,(V)任选地,在所述至少两种MHC II类表位编码核酸序列的3'端编码GPGPG氨基酸接头序列(SEQ ID NO:56)的第三核酸序列;和(iii)任选地,可操作地连接到所述抗原编码核酸序列的第二启动子核苷酸序列;并且其中如果所述第二启动子核苷酸序列不存在,则所述抗原编码核酸序列可操作地连接到所述原生启动子核苷酸序列,并且其中所述盒不编码免疫显性MHC I类表位,所述免疫显性MHC I类表位:(1)当以疫苗组合物的形式施用于受试者时,相对于所述盒中编码并能够刺激所述受试者中的免疫应答的另一种MHC I类表位,刺激5倍或更高的免疫应答,和/或(2)当以疫苗组合物的形式施用于受试者时,相对于当在不存在所述免疫显性MHC I类表位的情况下施用其他MHC I类表位时的免疫应答,降低了对所述盒中编码的另一种MHC I类表位的免疫应答,任选地其中所述免疫应答降低到检测限以下和/或其中所述免疫应答不是治疗有效应答。

[0036] 在一些方面,当以疫苗组合物的形式施用于受试者时,相对于所述盒中编码并能够刺激所述受试者中的免疫应答的另一种MHC I类表位,所述免疫显性MHC I类表位刺激10倍、20倍、30倍、40倍、50倍、100倍、200倍、300倍、400倍、500倍、1000倍、2000倍、3000倍、4000倍、5000倍或10,000倍或更高的免疫应答。在一些方面,所述免疫显性MHC I类表位将其他MHC I类表位的免疫应答降低到检测限以下和/或不刺激治疗有效应答。在一些方面,所述受试者表达至少一种已知或预测呈递所述免疫显性MHC I类表位和所述盒中编码的其他MHC I类表位两者的HLA等位基因。

[0037] 在一些方面,所述表位编码核酸序列中的一者或多者来源于受试者的肿瘤、感染或受感染细胞。在一些方面,所述表位编码核酸序列中的每一者来源于受试者的肿瘤、感染

或受感染细胞。在一些方面，所述表位编码核酸序列中的一者或多者不是来源于受试者的肿瘤、感染或受感染细胞。在一些方面，所述表位编码核酸序列中的每一者不是来源于受试者的肿瘤、感染或受感染细胞。

[0038] 在一些方面，所述表位编码核酸序列编码已知或怀疑由MHC I类呈递在细胞表面上的表位，任选地其中所述细胞表面是肿瘤细胞表面或受感染细胞表面，并且任选地其中所述细胞是受试者的细胞。在一些方面，所述细胞是选自由以下组成的组的肿瘤细胞：肺癌、黑色素瘤、乳腺癌、卵巢癌、前列腺癌、肾癌、胃癌、结肠癌、睾丸癌、头颈癌、胰腺癌、脑癌、B细胞淋巴瘤、急性髓细胞白血病、慢性髓细胞白血病、慢性淋巴细胞白血病、T细胞淋巴细胞白血病、非小细胞肺癌和小细胞肺癌，或者其中所述细胞是选自由以下组成的组的受感染细胞：病原体感染细胞、病毒感染细胞、细菌感染细胞、真菌感染细胞和寄生虫感染细胞。在一些方面，所述病毒感染细胞选自由以下组成的组：HIV感染细胞、严重急性呼吸综合征相关冠状病毒(SARS)感染细胞、严重急性呼吸综合征冠状病毒2(SARS-CoV-2)感染细胞、埃博拉感染细胞、乙型肝炎病毒(HBV)感染细胞、流感感染细胞、正粘病毒科病毒、人乳头瘤病毒(HPV)感染细胞、巨细胞病毒(CMV)感染细胞、基孔肯雅病毒(Chikungunya virus)感染细胞、呼吸道合胞病毒(RSV)感染细胞、登革热病毒感染细胞和丙型肝炎病毒(HCV)感染细胞。

[0039] 在一些方面，所述组合物还包含纳米微粒递送媒介物。在一些方面，所述纳米微粒递送媒介物是脂质纳米粒子(LNP)。在一些方面，所述LNP包含可电离氨基脂质。在一些方面，所述可电离氨基脂质包含MC3样(二亚油基甲基-4-二甲基氨基丁酸酯)分子。在一些方面，所述纳米微粒递送媒介物包封所述抗原表达系统。

[0040] 在一些方面，所述盒整合在所述至少一种启动子核苷酸序列和所述至少一种聚(A)序列之间。在一些方面，所述第二启动子不存在并且所述至少一种启动子核苷酸序列可操作地连接到所述抗原编码核酸序列。

[0041] 在一些方面，所述一种或多种载体包含一种或多种+-链RNA载体。在一些方面，所述一种或多种+-链RNA载体包含5'7-甲基鸟苷(m7g)帽。在一些方面，所述一种或多种+-链RNA载体是通过体外转录产生的。在一些方面，所述一种或多种载体在哺乳动物细胞内自我复制。在一些方面，所述骨架包含奥拉病毒(Aura virus)、摩根堡病毒(Fort Morgan virus)、委内瑞拉马脑炎病毒、罗斯河病毒(Ross River virus)、塞姆利基森林病毒(Semiliki Forest virus)、辛德毕斯病毒(Sindbis virus)或马雅罗病毒(Mayaro virus)的至少一种核苷酸序列。在一些方面，所述骨架包含委内瑞拉马脑炎病毒的至少一种核苷酸序列。在一些方面，所述骨架至少包含由奥拉病毒、摩根堡病毒、委内瑞拉马脑炎病毒、罗斯河病毒、塞姆利基森林病毒、辛德毕斯病毒或马雅罗病毒的核苷酸序列编码的用于非结构蛋白介导的扩增的序列、26S启动子序列、聚(A)序列、非结构蛋白1(nsP1)基因、nsP2基因、nsP3基因和nsP4基因。在一些方面，所述骨架至少包含由奥拉病毒、摩根堡病毒、委内瑞拉马脑炎病毒、罗斯河病毒、塞姆利基森林病毒、辛德毕斯病毒或马雅罗病毒的核苷酸序列编码的用于非结构蛋白介导的扩增的序列、26S启动子序列和聚(A)序列。在一些方面，所述用于非结构蛋白介导的扩增的序列选自由以下组成的组：甲病毒5'UTR、51-nt CSE、24-nt CSE、26S亚基因组启动子序列、19-nt CSE、甲病毒3'UTR或其组合。在一些方面，所述骨架不编码结构性病毒粒子蛋白衣壳E2和E1。在一些方面，所述盒被插入以代替奥拉病毒、摩根堡

病毒、委内瑞拉马脑炎病毒、罗斯河病毒、塞姆利基森林病毒、辛德毕斯病毒或马雅罗病毒的核苷酸序列内的结构性病毒粒子蛋白。

[0042] 在一些方面，所述委内瑞拉马脑炎病毒包含SEQ ID NO:3或SEQ ID NO:5的序列。在一些方面，所述委内瑞拉马脑炎病毒包含SEQ ID NO:3或SEQ ID NO:5的序列，其还包含碱基对7544和11175之间的缺失。在一些方面，所述骨架包含以SEQ ID NO:6或SEQ ID NO:7所示的序列。在一些方面，所述盒被插入在位置7544以代替如SEQ ID NO:3或SEQ ID NO:5的序列中所示的碱基对7544和11175之间的缺失。

[0043] 在一些方面，所述盒的插入提供了包含所述nsP1-4基因和所述至少一种抗原编码核酸序列的多顺反子RNA的转录，其中所述nsP1-4基因和所述至少一种抗原编码核酸序列在单独的开放阅读框中。

[0044] 在一些方面，所述骨架包含黑猩猩腺病毒载体的至少一种核苷酸序列。在一些方面，所述黑猩猩腺病毒载体是ChAdV68载体。在一些方面，所述ChAdV68载体包含ChAdV68载体骨架，所述骨架包含以SEQ ID NO:1所示的序列。在一些方面，所述ChAdV68载体包含ChAdV68载体骨架，所述骨架包含以SEQ ID NO:1所示的序列，除了所述序列在选自由SEQ ID NO:1所示序列的黑猩猩腺病毒E1A、E1B、E2A、E2B、E3、E4、L1、L2、L3、L4和L5基因组成的组的至少一个基因中完全缺失或功能性缺失，任选地其中所述序列在以下中完全缺失或功能性缺失：以SEQ ID NO:1所示序列的(1)E1A和E1B；(2)E1A、E1B和E3；或(3)E1A、E1B、E3和E4。在一些方面，所述ChAdV68载体包含ChAdV68载体骨架，所述骨架包含从SEQ ID NO:1的序列获得的基因或调控序列，任选地其中所述基因选自由SEQ ID NO:1所示序列的黑猩猩腺病毒反向末端重复序列(ITR)、E1A、E1B、E2A、E2B、E3、E4、L1、L2、L3、L4和L5基因组成的组。在一些方面，所述ChAdV68载体包含ChAdV68载体骨架，所述骨架包含部分缺失的E4基因，其包含缺失或部分缺失的E4orf2区和缺失或部分缺失的E4orf3区，以及任选地缺失或部分缺失的E4orf4区。在一些方面，所述ChAdV68载体包含ChAdV68载体骨架，所述骨架包含以SEQ ID NO:1所示序列的至少核苷酸2至36,518，并且还包含：(1)以SEQ ID NO:1所示序列的至少核苷酸577至3403的E1缺失，(2)以SEQ ID NO:1所示序列的至少核苷酸27,125至31,825的E3缺失，和(3)以SEQ ID NO:1所示序列的至少核苷酸34,916至35,642的E4缺失；任选地其中所述抗原盒被插入在所述E1缺失内。在一些方面，所述ChAdV68载体包含ChAdV68载体骨架，所述骨架包含以SEQ ID NO:68所示的序列，任选地其中所述抗原盒被插入在所述E1缺失内。在一些方面，所述ChAdV68载体包含ChAdV68载体骨架，所述骨架包含在碱基对编号577和3403之间或碱基对456和3014之间的一个或多个缺失，并且任选地其中所述载体还包含在SEQ ID NO:1所示序列的碱基对27,125和31,825之间或碱基对27,816和31,333之间的一个或多个缺失。在一些方面，所述ChAdV68载体包含ChAdV68载体骨架，所述骨架包含在以SEQ ID NO:1所示序列的碱基对编号3957和10346之间、碱基对编号21787和23370之间以及碱基对编号33486和36193之间的一个或多个缺失。在一些方面，所述盒被插入在所述ChAdV载体骨架的E1区、E3区和/或允许并入所述盒的任何缺失的AdV区。

[0045] 在一些方面，所述至少一种启动子核苷酸序列是由所述骨架编码的原生26S启动子核苷酸序列。在一些方面，所述至少一种启动子核苷酸序列是外源RNA启动子。在一些方面，所述第二启动子核苷酸序列是26S启动子核苷酸序列。在一些方面，所述第二启动子核苷酸序列包含多个26S启动子核苷酸序列，其中每个26S启动子核苷酸序列提供了所述单独

的开放阅读框中的一者或多者的转录。

[0046] 在一些方面，所述一种或多种载体的大小各自为至少300nt。在一些方面，所述一种或多种载体的大小各自为至少1kb。在一些方面，所述一种或多种载体的大小各自为2kb。在一些方面，所述一种或多种载体的大小各自小于5kb。

[0047] 在一些方面，所述至少一种抗原编码核酸序列中的至少一者编码由MHC I类呈递在细胞表面、任选地肿瘤细胞表面或受感染细胞表面上的多肽序列或其部分。

[0048] 在一些方面，每个表位编码核酸序列彼此直接连接。在一些方面，所述表位编码核酸序列中的至少一者用编码接头的核酸序列连接到不同表位编码核酸序列。在一些方面，所述接头连接两个MHC I类序列或一个MHC I类序列至一个MHC II类序列。在一些方面，所述接头选自由以下组成的组：(1)连续甘氨酸残基，长度为至少2、3、4、5、6、7、8、9或10个残基；(2)连续丙氨酸残基，长度为至少2、3、4、5、6、7、8、9或10个残基；(3)两个精氨酸残基(RR)；(4)丙氨酸、丙氨酸、酪氨酸(AAY)；(5)由哺乳动物蛋白酶体有效加工的长度为至少2、3、4、5、6、7、8、9或10个氨基酸残基的共有序列；和(6)侧接源自同源蛋白的抗原并且长度为至少2、3、4、5、6、7、8、9、10、11、12、13、14、15、16、17、18、19、20个或2-20个氨基酸残基的一种或多种原生序列。在一些方面，所述接头连接两个MHC II类序列或一个MHC II类序列至一个MHC I类序列。在一些方面，所述接头包含序列GPGPG(SEQ ID NO:56)。在一些方面，所述表位编码核酸序列中的至少一个序列可操作地或直接地连接到增强由其编码的表位的所述表位编码核酸序列的表达、稳定性、细胞运输、加工和呈递和/或免疫原性的单独或连续的序列。在一些方面，所述单独或连续的序列包含以下中的至少一者：泛素序列、经修饰以增加蛋白酶体靶向的泛素序列(例如，所述泛素序列在位置76处含有Gly至Ala取代)、免疫球蛋白信号序列(例如，IgK)、主要组织相容性I类序列、溶酶体相关膜蛋白(LAMP)-1、人树突状细胞溶酶体相关膜蛋白和主要组织相容性II类序列；任选地其中所述经修饰以增加蛋白酶体靶向的泛素序列是A76。

[0049] 在一些方面，所述表位编码核酸序列中的至少一者编码多肽序列或其部分，其相对于经翻译的相应野生型核酸序列对其相应的MHC等位基因具有增加的结合亲和力。在一些方面，所述表位编码核酸序列中的至少一者编码多肽序列或其部分，其相对于经翻译的相应野生型核酸序列对其相应的MHC等位基因具有增加的结合稳定性。在一些方面，所述表位编码核酸序列中的至少一者编码多肽序列或其部分，其相对于经翻译的相应野生型核酸序列具有增加的在其相应MHC等位基因上呈递的可能性。在一些方面，所述至少一种改变包括点突变、移码突变、非移码突变、缺失突变、插入突变、剪接变体、基因组重排或蛋白酶体产生的剪接抗原。

[0050] 在一些方面，所述肿瘤选自由以下组成的组：肺癌、黑色素瘤、乳腺癌、卵巢癌、前列腺癌、肾癌、胃癌、结肠癌、睾丸癌、头颈癌、胰腺癌、膀胱癌、脑癌、B细胞淋巴瘤、急性髓细胞白血病、成人急性淋巴母细胞白血病、慢性髓细胞白血病、慢性淋巴细胞白血病、T细胞淋巴细胞白血病、非小细胞肺癌和小细胞肺癌，或者所述感染性疾病生物体选自由以下组成的组：严重急性呼吸综合征相关冠状病毒(SARS)、严重急性呼吸综合征冠状病毒2(SARS-CoV-2)、埃博拉、HIV、乙型肝炎病毒(HBV)、流感、丙型肝炎病毒(HCV)、人乳头瘤病毒(HPV)、巨细胞病毒(CMV)、基孔肯雅病毒、呼吸道合胞病毒(RSV)、登革热病毒、正粘病毒科病毒和肺结核。

[0051] 在一些方面,所述至少一种抗原编码核酸序列包含至少2-10、2、3、4、5、6、7、8、9或10种表位编码核酸序列。在一些方面,所述至少一种抗原编码核酸序列包含至少11-20、15-20、11-100、11-200、11-300、11-400、11、12、13、14、15、16、17、18、19、20或多达400种表位编码核酸序列。在一些方面,所述至少一种抗原编码核酸序列包含至少2-400种表位编码核酸序列,并且其中所述表位编码核酸序列中的至少两者编码由MHC I类呈递在细胞表面、任选地肿瘤细胞表面或受感染细胞表面上的多肽序列或其部分。在一些方面,所述至少一种抗原编码核酸序列包含至少2-10、2、3、4、5、6、7、8、9或10种抗原编码核酸序列。在一些方面,所述至少一种抗原编码核酸序列包含至少11-20、15-20、11-100、11-200、11-300、11-400、11、12、13、14、15、16、17、18、19、20或多达400种抗原编码核酸序列。在一些方面,所述至少一种抗原编码核酸序列包含至少2-400种抗原编码核酸序列,并且其中所述抗原编码核酸序列中的至少两者编码由MHC I类呈递在细胞表面、任选地肿瘤细胞表面或受感染细胞表面上的多肽序列或其部分。在一些方面,所述表位编码核酸序列中的至少两者编码由MHC I类呈递在细胞表面、任选地肿瘤细胞表面或受感染细胞表面上的多肽序列或其部分。

[0052] 在一些方面,当施用于所述受试者并被翻译时,由所述表位编码核酸序列编码的表位中的至少一者被呈递在抗原呈递细胞上,导致靶向所述肿瘤细胞表面或所述受感染细胞表面上的抗原中的至少一者的免疫应答。在一些方面,当所述至少一种抗原编码核酸序列被施用于所述受试者并被翻译时,所述MHC I类或II类表位中的至少一者被呈递在抗原呈递细胞上,导致靶向肿瘤细胞表面或受感染细胞表面上的表位中的至少一者的免疫应答,并且任选地其中所述至少一种抗原编码核酸序列中的每一者的表达由所述至少一种启动子核苷酸序列驱动。

[0053] 在一些方面,每个表位编码核酸序列编码长度为8至35个氨基酸,任选地长度为9-17、9-25、8、9、10、11、12、13、14、15、16、17、18、19、20、21、22、23、24、25、26、27、28、29、30、31、32、33、34或35个氨基酸的多肽序列。

[0054] 在一些方面,存在至少一种MHC II类表位编码核酸序列。在一些方面,所述至少一种MHC II类表位编码核酸序列存在并且包含至少一种MHC II类表位编码核酸序列,所述MHC II类表位编码核酸序列包含使所编码的肽序列与由野生型核酸序列编码的相应肽序列不同的至少一种改变。在一些方面,所述至少一种MHC II类表位编码核酸序列的长度为12-20、12、13、14、15、16、17、18、19、20或20-40个氨基酸。在一些方面,所述至少一种MHC II类表位编码核酸序列存在并且包含至少一种通用MHC II类抗原编码核酸序列,任选地其中所述至少一种通用序列包含破伤风类毒素和PADRE中的至少一者。

[0055] 在一些方面,所述至少一种启动子核苷酸序列或所述第二启动子核苷酸序列是诱导型的。在一些方面,所述至少一种启动子核苷酸序列或所述第二启动子核苷酸序列是非诱导型的。

[0056] 在一些方面,所述至少一种聚(A)序列包含所述骨架原生的聚(A)序列。在一些方面,所述至少一种聚(A)序列包含所述骨架外源的聚(A)序列。在一些方面,所述至少一种聚(A)序列可操作地连接到所述至少一种抗原编码核酸序列中的至少一者。在一些方面,所述至少一种聚(A)序列是至少20、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个连续A核苷酸。在一些方面,所述至少一种聚(A)序列是至少80个连续A核苷酸。

[0057] 在一些方面,所述盒还包含以下中的至少一者:内含子序列,土拨鼠肝炎病毒转录

后调控元件(WPRE)序列,内部核糖体进入序列(IRES)序列,编码2A自切割肽序列的核苷酸序列,编码弗林蛋白酶切割位点的核苷酸序列,或已知增强可操作地连接到所述至少一种抗原编码核酸序列中的至少一者的mRNA的核输出、稳定性或翻译效率的5'或3'非编码区中的序列。在一些方面,所述盒还包含报告基因,包括但不限于绿色荧光蛋白(GFP)、GFP变体、分泌型碱性磷酸酶、荧光素酶、荧光素酶变体、或可检测的肽或表位。在一些方面,所述可检测的肽或表位选自由HA标签、Flag标签、His标签或V5标签组成的组。

[0058] 在一些方面,所述一种或多种载体还包含一种或多种编码至少一种免疫调节剂的核酸序列。在一些方面,所述免疫调节剂是抗CTLA4抗体或其抗原结合片段、抗PD-1抗体或其抗原结合片段、抗PD-L1抗体或其抗原结合片段、抗4-1BB抗体或其抗原结合片段、或抗OX-40抗体或其抗原结合片段。在一些方面,所述抗体或其抗原结合片段是Fab片段、Fab'片段、单链Fv(scFv)、呈单一特异性或连接在一起的多个特异性的单域抗体(sdAb)(例如骆驼科动物抗体结构域),或全长单链抗体(例如,具有通过柔性接头连接的重链和轻链的全长IgG)。在一些方面,所述抗体的重链和轻链序列是由例如2A或IRES的自切割序列分开的连续序列;或者所述抗体的重链和轻链序列是通过柔性接头如连续甘氨酸残基连接。在一些方面,所述免疫调节剂是细胞因子。在一些方面,所述细胞因子是IL-2、IL-7、IL-12、IL-15或IL-21或其各自的变体中的至少一者。

[0059] 在一些方面,至少一种表位编码核酸序列是通过执行以下步骤来选择:(a)从肿瘤、受感染细胞或感染性疾病生物体获得外显子组、转录组或全基因组核苷酸测序数据中的至少一者,其中所述核苷酸测序数据用于获得代表一组抗原中的每一者的肽序列的数据;(b)将每个抗原的肽序列输入到呈递模型中,以产生所述抗原中的每一者由MHC等位基因中的一者或多者在细胞表面、任选地肿瘤细胞表面或受感染细胞表面上呈递的数值可能性集合,所述数值可能性集合已至少基于所接收的质谱数据进行鉴定;以及(c)基于所述数值可能性集合选择所述抗原集合的子集以产生经选择抗原集合,所述经选择抗原用于产生所述表位编码核酸序列。

[0060] 在一些方面,所述表位编码核酸序列中的每一者是通过执行以下步骤来选择:(a)从肿瘤、受感染细胞或感染性疾病生物体获得外显子组、转录组或全基因组核苷酸测序数据中的至少一者,其中所述核苷酸测序数据用于获得代表一组抗原中的每一者的肽序列的数据;(b)将每个抗原的肽序列输入到呈递模型中,以产生所述抗原中的每一者由MHC等位基因中的一者或多者在细胞表面、任选地肿瘤细胞表面或受感染细胞表面上呈递的数值可能性集合,所述数值可能性集合已至少基于所接收的质谱数据进行鉴定;以及(c)基于所述数值可能性集合选择所述抗原集合的子集以产生经选择抗原集合,所述经选择抗原用于产生至少20种表位编码核酸序列。在一些方面,所述经选择抗原集合的数目为2-20。在一些方面,所述呈递模型表示以下两者之间的依赖性:(a)所述MHC等位基因中的特定一者和在肽序列的特定位置处的特定氨基酸的一对的存在;和(b)由所述对的所述MHC等位基因中的所述特定一者将在所述特定位置处包含所述特定氨基酸的所述肽序列呈递在细胞表面、任选地肿瘤细胞表面或受感染细胞表面上的可能性。在一些方面,选择所述经选择抗原集合包括基于所述呈递模型选择相对于未选择的抗原被呈递在细胞表面上的可能性增加的抗原,任选地其中所述经选择抗原已被验证为由一种或多种特定的HLA等位基因呈递。在一些方面,选择所述经选择抗原集合包括基于所述呈递模型选择相对于未选择的抗原能够在受试

者中诱导肿瘤特异性或感染性疾病特异性免疫应答的可能性增加的抗原。在一些方面，选择所述经选择抗原集合包括基于所述呈递模型选择相对于未选择的抗原能够由专职抗原呈递细胞(APC)呈递给初始T细胞的可能性增加的抗原，任选地其中所述APC是树突状细胞(DC)。在一些方面，选择所述经选择抗原集合包括基于所述呈递模型选择相对于未选择的抗原经由中心或外周耐受性受抑制的可能性降低的抗原。在一些方面，选择所述经选择抗原集合包括基于所述呈递模型选择相对于未选择的抗原能够诱导对受试者的正常组织的自身免疫应答的可能性降低的抗原。在一些方面，外显子组或转录组核苷酸测序数据是通过对肿瘤细胞或组织、受感染细胞或感染性疾病生物体进行测序而获得的。在一些方面，所述测序是下一代测序(NGS)或任何大规模并行测序方法。

[0061] 在一些方面，所述盒包含由所述盒中的相邻序列形成的连接表位序列。在一些方面，至少一个或每个连接表位序列对MHC具有大于500nM的亲和力。在一些方面，每个连接表位序列都是非自身的。

[0062] 在一些方面，预测或验证所述MHC I类表位中的每一者能够由至少5％的群体中存在的至少一个HLA等位基因呈递。在一些方面，预测或验证所述MHC I类表位中的每一者能够由至少一个HLA等位基因呈递，其中每个抗原/HLA对在群体中具有至少0.01％的抗原/HLA流行率。在一些方面，预测或验证所述MHC I类表位中的每一者能够由至少一个HLA等位基因呈递，其中每个抗原/HLA对在群体中具有至少0.1％的抗原/HLA流行率。

[0063] 在一些方面，所述盒不编码包含经翻译的野生型核酸序列的非治疗性MHC I类或II类表位核酸序列，其中预测所述非治疗性表位将展示在所述受试者的MHC等位基因上。在一些方面，所述非治疗性预测的MHC I类或II类表位序列是由所述盒中的相邻序列形成的连接表位序列。

[0064] 在一些方面，所述预测是基于通过将所述非治疗性表位的序列输入到呈递模型中而生成的呈递可能性。

[0065] 在一些方面，所述盒中的所述至少一种抗原编码核酸序列的顺序通过一系列包括以下的步骤确定：(a)产生一组对应于不同顺序的所述至少一种抗原编码核酸序列的候选盒序列；(b)对于每个候选盒序列，基于所述候选盒序列中非治疗性表位的呈递来确定呈递分数；以及(c)选择与低于预定阈值的呈递分数相关的候选盒序列作为抗原疫苗的盒序列。

[0066] 本文还提供了一种药物组合物，所述药物组合物包含本文所述的任何组合物和药学上可接受的载剂。在一些方面，所述组合物还包含佐剂。在一些方面，所述组合物还包含免疫调节剂。在一些方面，所述免疫调节剂是抗CTLA4抗体或其抗原结合片段、抗PD-1抗体或其抗原结合片段、抗PD-L1抗体或其抗原结合片段、抗4-1BB抗体或其抗原结合片段、或抗OX-40抗体或其抗原结合片段。

[0067] 本文还提供了一种分离的核苷酸序列或分离的核苷酸序列集合，其包含本文所述的任何组合物的盒和从SEQ ID NO:3或SEQ ID NO:5的序列获得的一种或多种元件，任选地其中所述一种或多种元件选自由以下组成的组：非结构蛋白介导的扩增所必需的序列、26S启动子核苷酸序列、聚(A)序列以及以SEQ ID NO:3或SEQ ID NO:5所示的序列的nsP1-4基因，并且任选地其中所述核苷酸序列是cDNA。在一些方面，所述序列或分离的核苷酸序列集合包含插入在以SEQ ID NO:6或SEQ ID NO:7所示序列的位置7544处的上述组合物权利要求中任一项所述的盒。在一些方面，所述组合物还包含：a)从SEQ ID NO:3或SEQ ID NO:5的

序列获得的所述一种或多种元件的T7或SP6 RNA聚合酶启动子核苷酸序列5';和b)任选地,所述聚(A)序列的一个或多个限制性位点3'。在一些方面,上述组合物权利要求中任一项所述的盒被插入在SEQ ID NO:8或SEQ ID NO:9的位置7563处。

[0068] 本文还提供了包含本文所述的任何核苷酸序列的载体或载体集合。

[0069] 本文还提供了包含本文所述的任何核苷酸序列或分离的核苷酸序列集合的分离的细胞,任选地其中所述细胞是BHK-21、CHO、HEK293或其变体、911、HeLa、A549、LP-293、PER.C6或AE1-2a细胞。

[0070] 本文还提供了一种试剂盒,所述试剂盒包括本文所述的任何组合物和使用说明书。

[0071] 在一些方面,任何上述组合物还包含纳米微粒递送媒介物。在一些方面,所述纳米微粒递送媒介物可以是脂质纳米粒子(LNP)。在一些方面,所述LNP包含可电离氨基脂质。在一些方面,所述可电离氨基脂质包含MC3样(二亚油基甲基-4-二甲基氨基丁酸酯)分子。在一些方面,所述纳米微粒递送媒介物包封抗原表达系统。

[0072] 在一些方面,任何上述组合物还包含多个LNP,其中所述LNP包含:所述抗原表达系统;阳离子脂质;非阳离子脂质;和抑制所述LNP聚集的缀合脂质,其中所述多个LNP中至少约95%的LNP:具有非层状形态;或者是电子致密的。

[0073] 在一些方面,所述非阳离子脂质是(1)磷脂和(2)胆固醇或胆固醇衍生物的混合物。

[0074] 在一些方面,所述抑制LNP聚集的缀合脂质是聚乙二醇(PEG)-脂质缀合物。在一些方面,所述PEG-脂质缀合物选自由以下组成的组:PEG-二酰基甘油(PEG-DAG)缀合物、PEG二烷氧基丙基(PEG-DAA)缀合物、PEG-磷脂缀合物、PEG-神经酰胺(PEG-Cer)缀合物及其混合物。在一些方面,所述PEG-DAA缀合物是选自由以下组成的组的成员:PEG-二癸氧基丙基(C_{10})缀合物、PEG-二月桂氧基丙基(C_{12})缀合物、PEG-二肉豆蔻酰氧基丙基(C_{14})缀合物、PEG-二棕榈酰氧基丙基(C_{16})缀合物、PEG-二硬脂酰氧基丙基(C_{18})缀合物及其混合物。

[0075] 在一些方面,所述抗原表达系统完全包封在所述LNP中。

[0076] 在一些方面,所述LNP的非层状形态包括反六角形(H_{II})或立方相结构。

[0077] 在一些方面,所述阳离子脂质占LNP中存在的总脂质的约10摩尔%至约50摩尔%。在一些方面,所述阳离子脂质占LNP中存在的总脂质的约20摩尔%至约50摩尔%。在一些方面,所述阳离子脂质占LNP中存在的总脂质的约20摩尔%至约40摩尔%。

[0078] 在一些方面,所述非阳离子脂质占LNP中存在的总脂质的约10摩尔%至约60摩尔%。在一些方面,所述非阳离子脂质占LNP中存在的总脂质的约20摩尔%至约55摩尔%。在一些方面,所述非阳离子脂质占LNP中存在的总脂质的约25摩尔%至约50摩尔%。

[0079] 在一些方面,所述缀合脂质占LNP中存在的总脂质的约0.5摩尔%至约20摩尔%。在一些方面,所述缀合脂质占LNP中存在的总脂质的约2摩尔%至约20摩尔%。在一些方面,所述缀合脂质占LNP中存在的总脂质的约1.5摩尔%至约18摩尔%。

[0080] 在一些方面,大于95%的LNP具有非层状形态。在一些方面,大于95%的LNP是电子致密的。

[0081] 在一些方面,任何上述组合物还包含多个LNP,其中所述LNP包含:阳离子脂质,其占LNP中存在的总脂质的50摩尔%至65摩尔%;抑制LNP聚集的缀合脂质,其占LNP中存在的

总脂质的0.5摩尔％至2摩尔％;和非阳离子脂质,其包含:磷脂和胆固醇或其衍生物的混合物,其中所述磷脂占LNP中存在的总脂质的4摩尔％至10摩尔％,并且所述胆固醇或其衍生物占LNP中存在的总脂质的30摩尔％至40摩尔％;磷脂和胆固醇或其衍生物的混合物,其中所述磷脂占LNP中存在的总脂质的3摩尔％至15摩尔％,并且所述胆固醇或其衍生物占LNP中存在的总脂质的30摩尔％至40摩尔％;或LNP中存在的总脂质的高达49.5摩尔％,并包含磷脂和胆固醇或其衍生物的混合物,其中所述胆固醇或其衍生物占LNP中存在的总脂质的30摩尔％至40摩尔％。

[0082] 在一些方面,任何上述组合物还包含多个LNP,其中所述LNP包含:阳离子脂质,其占LNP中存在的总脂质的50摩尔％至85摩尔％;抑制LNP聚集的缀合脂质,其占LNP中存在的总脂质的0.5摩尔％至2摩尔％;和非阳离子脂质,其占LNP中存在的总脂质的13摩尔％至49.5摩尔％。

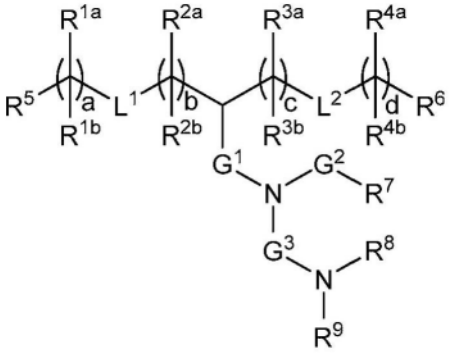
[0083] 在一些方面,所述磷脂包含二棕榈酰磷脂酰胆碱(DPPC)、二硬脂酰磷脂酰胆碱(DSPC)或其混合物。

[0084] 在一些方面,所述缀合脂质包含聚乙二醇(PEG)-脂质缀合物。在一些方面,所述PEG-脂质缀合物包含PEG-二酰基甘油(PEG-DAG)缀合物、PEG-二烷氧基丙基(PEG-DAA)缀合物或其混合物。在一些方面,所述PEG-DAA缀合物包含PEG-二肉豆蔻酰氧基丙基(PEG-DMA)缀合物、PEG-二硬脂酰氧基丙基(PEG-DSA)缀合物或其混合物。在一些方面,所述缀合物的PEG部分具有约2,000道尔顿的平均分子量。

[0085] 在一些方面,所述缀合脂质占LNP中存在的总脂质的1摩尔％至2摩尔％。

[0086] 在一些方面,所述LNP包含具有式I结构的化合物:

[0087]

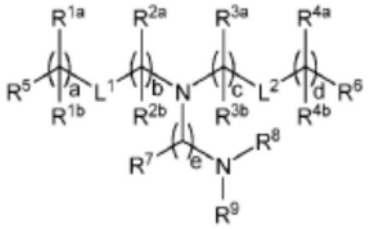


I

[0088] 或其药学上可接受的盐、互变异构体、前药或立体异构体,其中:L^1和L^2各自独立地是-O(C=O)-、-(C=O)O-、-C(=O)-、-O-、-S(O)$_x$-、-S-S-、-C(=O)S-、-SC(=O)-、-R^aC(=O)-、-C(=O)R^a-、-R^aC(=O)R^a-、-OC(=O)R^a-、-R^aC(=O)O-或直接键;G^1是Ci-C_2亚烷基、-(C=O)-、-O(C=O)-、-SC(=O)-、-R^aC(=O)-或直接键;-C(=O)-、-(C=O)O-、-C(=O)S-、-C(=O)R^a-或直接键;G是Ci-C_6亚烷基;R^a是H或C1-C12烷基;R^{1a}和R^{1b}在每次出现时独立地是:(a)H或C_1-C_{12}烷基;或(b)R^{1a}是H或C_1-C_{12}烷基,并且R^{1b}和其所结合的碳原子一起与相邻R^{1b}和其所结合的碳原子一起形成碳碳双键;R^{2a}和R^{2b}在每次出现时独立地是:(a)H或C_1-C_{12}烷基;或(b)R^{2a}是H或C_1-C_{12}烷基,并且R^{2b}和其所结合的碳原子一起与相邻R^{2b}和其所结合的碳原子一起形成碳碳双键;R^{3a}和R^{3b}在每次出现时独立地是:(a)H或C_1-C_{12}烷基;或(b)R^{3a}是H或C_1-

C_{12}烷基，并且R^{3b}和其所结合的碳原子一起与相邻R和其所结合的碳原子一起形成碳碳双键；R^{4a}和R^{4b}在每次出现时独立地是：(a)H或C1-C12烷基；或(b)R^{4a}是H或C1-C12烷基，并且R^{4b}和其所结合的碳原子一起与相邻R^{4b}和其所结合的碳原子一起形成碳碳双键；R^5和R^6各自独立地是H或甲基；R^7是C4-C20烷基；R^8和R^9各自独立地是C1-C12烷基；或者R^8和R^9与其所连接的氮原子一起形成5元、6元或7元杂环；a、b、c和d各自独立地是1至24的整数；并且x是0、1或2。

[0089] 在一些方面，所述LNP包含具有式II结构的化合物：

[0090] 

II

[0091] 或其药学上可接受的盐、互变异构体、前药或立体异构体，其中：L^1和L^2各自独立地是-O(C=O)-、-(C=O)O-或碳碳双键；R^{1a}和R^{1b}在每次出现时独立地是(a)H或C_1-C_{12}烷基，或(b)R^{1a}是H或C_1-C_{12}烷基，并且R^{1b}和其所结合的碳原子一起与相邻R^{1b}和其所结合的碳原子一起形成碳碳双键；R^{2a}和R^{2b}在每次出现时独立地是(a)H或C_1-C_{12}烷基，或(b)R^{2a}是H或C_1-C_{12}烷基，并且R^{2b}和其所结合的碳原子一起与相邻R^{2b}和其所结合的碳原子一起形成碳碳双键；R^{3a}和R^{3b}在每次出现时独立地是(a)H或C_1-C_{12}烷基，或(b)R^{3a}是H或C_1-C_{12}烷基，并且R^{3b}和其所结合的碳原子一起与相邻R^{3b}和其所结合的碳原子一起形成碳碳双键；R^{4a}和R^{4b}在每次出现时独立地是(a)H或C_1-C_{12}烷基，或(b)R^{4a}是H或C_1-C_{12}烷基，并且R^{4b}和其所结合的碳原子一起与相邻R^{4b}和其所结合的碳原子一起形成碳碳双键；R^5和R^6各自独立地是甲基或环烷基；R^7在每次出现时独立地是H或C_1-C_{12}烷基；R^8和R^9各自独立地是未被取代的C1-C12烷基；或者R^8和R^9与其所连接的氮原子一起形成包含一个氮原子的5元、6元或7元杂环；a和d各自独立地是0至24的整数；b和c各自独立地是1至24的整数；并且e是1或2，其条件是：R^{1a}、R^{2a}、R^{3a}或R^{4a}中的至少一者是C1-C12烷基，或者L^1或L^2中的至少一者是-O(C=O)-或-(C=O)O-；并且R^{1a}和R^{1b}当a为6时不是异丙基或者当a为8时不是正丁基。

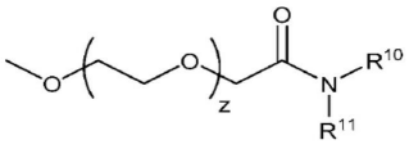
[0092] 在一些方面，任何上述组合物还包含一种或多种赋形剂，所述赋形剂包含中性脂质、类固醇和聚合物缀合脂质。在一些方面，所述中性脂质包含1,2-二硬脂酰基-sn-甘油-3-磷酸胆碱(DSPC)、1,2-二棕榈酰基-sn-甘油-3-磷酸胆碱(DPPC)、1,2-二肉豆蔻酰基-sn-甘油-3-磷酸胆碱(DMPC)、1-棕榈酰基-2-油酰基-sn-甘油-3-磷酸胆碱(POPC)、1,2-二油酰基-sn-甘油-3-磷酸胆碱(DOPC)和1,2-二油酰基-sn-甘油-3-磷酸乙醇胺(DOPE)中的至少一者。在一些方面，所述中性脂质是DSPC。

[0093] 在一些方面，所述化合物与所述中性脂质的摩尔比在约2:1至约8:1的范围内。

[0094] 在一些方面，所述类固醇是胆固醇。在一些方面，所述化合物与胆固醇的摩尔比在约2:1至1:1的范围内。

[0095] 在一些方面，所述聚合物缀合脂质是聚乙二醇化脂质。在一些方面，所述化合物与所述聚乙二醇化脂质的摩尔比在约100:1至约25:1的范围内。在一些方面，所述聚乙二醇化

脂质是PEG-DAG、PEG聚乙烯(PEG-PE)、PEG-琥珀酰基-二酰基甘油(PEG-S-DAG)、PEG-cer或PEG二烷氧基丙基氨基甲酸酯。在一些方面,所述聚乙二醇化脂质具有以下结构III:

[0096] 

III

[0097] 或其药学上可接受的盐、互变异构体或立体异构体,其中:R^{10}和R^{11}各自独立地是含有10至30个碳原子的直链或支链的饱和或不饱和的烷基链,其中所述烷基链任选地被一个或多个酯键中断;并且z具有30至60范围内的平均值。在一些方面,R^{10}和R^{11}各自独立地是具有12至16个碳原子的直链的饱和烷基链。在一些方面,平均z是约45。

[0098] 这里开始

[0099] 在一些方面,所述LNP在与聚阴离子核酸混合时自组装成非双层结构。在一些方面,所述非双层结构的直径介于60nm和120nm之间。在一些方面,所述非双层结构的直径为约70nm、约80nm、约90nm或约100nm。在一些方面,其中纳米微粒递送媒介物具有约100nm的直径。

[0100] 本文还提供了一种用于治疗患有癌症的受试者的方法,所述方法包括向所述受试者施用本文所述的任何组合物或任何药物组合物。在一些方面,所述表位编码核酸序列来源于患有癌症的受试者的肿瘤或来源于受感染受试者的细胞或样品。在一些方面,所述表位编码核酸序列不是来源于患有癌症的受试者的肿瘤或来源于受感染受试者的细胞或样品。

[0101] 本文还提供了一种用于刺激受试者中的免疫应答的方法,所述方法包括向所述受试者施用本文所述的任何组合物或任何药物组合物。

[0102] 在一些方面,所述受试者表达至少一种预测或已知呈递MHC I类表位的HLA等位基因。在一些方面,预测或已知呈递MHC I类表位的HLA等位基因是A*03:01、A*11:01、A*02:01、C*01:02和/或A*01:01。在一些方面,预测或已知呈递MHC I类表位的HLA等位基因是A*03:01。在一些方面,预测或已知呈递MHC I类表位的HLA等位基因是A*11:01。在一些方面,预测或已知呈递MHC I类表位的HLA等位基因是A*02:01。在一些方面,预测或已知呈递MHC I类表位的HLA等位基因是C*01:02。在一些方面,预测或已知呈递MHC I类表位的HLA等位基因是A*01:01。在一些方面,所述组合物是肌内(IM)、皮内(ID)、皮下(SC)或静脉内(IV)施用的。在一些方面,所述组合物是肌内施用的。在一些方面,所述方法还包括施用一种或多种免疫调节剂,任选地其中在施用所述组合物或药物组合物之前、同时或之后施用所述免疫调节剂。在一些方面,所述一种或多种免疫调节剂选自由以下组成的组:抗CTLA4抗体或其抗原结合片段、抗PD-1抗体或其抗原结合片段、抗PD-L1抗体或其抗原结合片段、抗4-1BB抗体或其抗原结合片段、或抗OX-40抗体或其抗原结合片段。在一些方面,所述免疫调节剂是静脉内(IV)、肌内(IM)、皮内(ID)或皮下(SC)施用的。在一些方面,所述皮下施用在所述组合物或药物组合物施用部位附近或紧邻一个或多个载体或组合物引流淋巴结。

[0103] 在一些方面,所述方法还包括向所述受试者施用第二疫苗组合物。在一些方面,所述第二疫苗组合物是在施用本文所述的任何组合物或药物组合物之前施用。在一些方面,

所述第二疫苗组合物是在施用本文所述的任何组合物或药物组合物之后施用。在一些方面，所述第二疫苗组合物与本文所述的任何组合物或药物组合物相同。在一些方面，所述第二疫苗组合物与本文所述的任何组合物或药物组合物不同。在一些方面，所述第二疫苗组合物包含编码至少一种抗原编码核酸序列的黑猩猩腺病毒载体。在一些方面，由所述黑猩猩腺病毒载体编码的所述至少一种抗原编码核酸序列与上述组合物权利要求中任一项所述的至少一种抗原编码核酸序列相同。

[0104] 本文还提供了一种制造如上述组合物权利要求中任一项所述的一种或多种载体的方法，所述方法包括：(a)获得包含骨架和盒的线性化DNA序列；(b)通过将所述线性化DNA序列添加到包含将所述线性化DNA序列转录成RNA的所有必要组分的体外转录反应中，来体外转录所述线性化DNA序列，任选地还包括将m7g帽体外添加到所得RNA中；以及(c)从所述体外转录反应中分离所述一种或多种载体。在一些方面，所述线性化DNA序列是通过将DNA质粒序列线性化或通过使用PCR扩增而产生的。在一些方面，所述DNA质粒序列是使用细菌重组或全基因组DNA合成或全基因组DNA合成中的一者与所合成DNA在细菌细胞中扩增产生的。在一些方面，从所述体外转录反应中分离所述一种或多种载体涉及苯酚氯仿提取、基于硅胶柱的纯化或类似的RNA纯化方法中的一者或多者。

[0105] 本文还提供了一种制造用于递送抗原表达系统的上述组合物权利要求中任一项所述的组合物的方法，所述方法包括：(a)提供用于纳米微粒递送媒介物的组分；(b)提供所述抗原表达系统；以及(c)为所述纳米微粒递送媒介物和所述抗原表达系统提供足以产生用于递送所述抗原表达系统的组合物的条件。在一些方面，所述条件由微流体混合提供。

[0106] 本文还提供了一种用于治疗患有疾病的受试者的方法，任选地其中所述疾病是癌症或感染，所述方法包括向所述受试者施用基于抗原的疫苗给所述受试者，其中所述基于抗原的疫苗包含抗原编码盒或由所述盒编码的多肽序列，其中所述抗原编码盒包含至少一种由下式从5'至3'描述的抗原编码核酸序列：

[0107] $(E_x-(E^N_n)_y)_z$

[0108] 其中E表示包含不同表位编码核酸序列的核苷酸序列，n表示单独的不同表位编码核酸序列的数目并且是包括0的任何整数，E^N表示包含对于每个相应n的单独的不同表位编码核酸序列的核苷酸序列，对于z的每次迭代：x＝0或1，对于每个n，y＝0或1，和x或y中的至少一者＝1，并且z＝2或更大，其中所述抗原编码核酸序列包含E、给定E^N或其组合的至少两次迭代，并且包含所述至少两次迭代的所述不同表位编码核酸序列中的至少一者编码不同的KRAS相关MHC I类新表位。

[0109] 本文还提供了一种用于治疗患有疾病的受试者的方法，任选地其中所述疾病是癌症，所述方法包括向所述受试者施用基于抗原的疫苗给所述受试者，其中所述基于抗原的疫苗包含抗原表达系统，所述抗原表达系统包含：抗原表达系统，其中所述抗原表达系统包含一种或多种载体，所述一种或多种载体包含：(a)载体骨架，其中所述骨架包含：(i)至少一种启动子核苷酸序列，和(ii)任选地，至少一种聚腺苷酸化(聚(A))序列；和(b)盒，其中所述盒包含：(i)至少一种抗原编码核酸序列，所述抗原编码核酸序列包含：(I)编码KRAS相关MHC I类新表位的表位编码核酸序列，并且其中所述表位编码核酸序列中的每一者包含：(A)任选地，5'接头序列，和(B)任选地，3'接头序列；(ii)任选地，可操作地连接到所述抗原编码核酸序列的第二启动子核苷酸序列；和(iii)任选地，至少一种MHC II类表位编码核酸

序列;(iv)任选地,至少一种编码GPGPG氨基酸接头序列(SEQ ID NO:56)的核酸序列;和(v)任选地,至少一种第二聚(A)序列,其中所述第二聚(A)序列是所述载体骨架的原生聚(A)序列或外源聚(A)序列,其中如果所述第二启动子核苷酸序列不存在,则所述抗原编码核酸序列可操作地连接到所述至少一种启动子核苷酸序列,并且其中所述至少一种抗原编码核酸序列包含编码所述KRAS相关MHC I类新表位的所述表位编码核酸序列的至少两次迭代。

[0110] 本文还提供了一种用于治疗患有疾病的受试者的方法,任选地其中所述疾病是癌症,所述方法包括向所述受试者施用基于抗原的疫苗给所述受试者,其中所述基于抗原的疫苗包含抗原表达系统,所述抗原表达系统包含:抗原表达系统,其中所述抗原表达系统包含一种或多种载体,所述一种或多种载体包含:(a)载体骨架,其中所述骨架包含:(i)至少一种启动子核苷酸序列,和(ii)任选地,至少一种聚腺苷酸化(聚(A))序列;和(b)盒,其中所述盒包含:(i)至少一种抗原编码核酸序列,所述抗原编码核酸序列包含:(I)至少两种不同表位编码核酸序列,任选地包含:(1)使所编码的表位序列与由野生型核酸序列编码的相应肽序列不同的至少一种改变,任选地其中所述至少一种改变是KRAS突变,或(2)编码感染性疾病生物体肽的核酸序列,所述肽选自由以下组成的组:病原体衍生肽、病毒衍生肽、细菌衍生肽、真菌衍生肽和寄生虫衍生肽,并且其中所述表位编码核酸序列中的每一者包含:(A)任选地,5'接头序列,和(B)任选地,3'接头序列;(ii)任选地,可操作地连接到所述抗原编码核酸序列的第二启动子核苷酸序列;和(iii)任选地,至少一种MHC II类表位编码核酸序列;(iv)任选地,至少一种编码GPGPG氨基酸接头序列(SEQ ID NO:56)的核酸序列;和(v)任选地,至少一种第二聚(A)序列,其中所述第二聚(A)序列是所述载体骨架的原生聚(A)序列或外源聚(A)序列,其中如果所述第二启动子核苷酸序列不存在,则所述抗原编码核酸序列可操作地连接到所述至少一种启动子核苷酸序列,并且其中所述盒不编码免疫显性MHC I类表位,所述免疫显性MHC I类表位:(1)当以疫苗组合物的形式施用于受试者时,相对于所述盒中编码并能够刺激所述受试者中的免疫应答的另一种MHC I类表位,刺激5倍或更高的免疫应答,和/或(2)当以疫苗组合物的形式施用于受试者时,相对于当在不存在所述免疫显性MHC I类表位的情况下施用其他MHC I类表位时的免疫应答,降低了对所述盒中编码的另一种MHC I类表位的免疫应答,任选地其中所述免疫应答降低到检测限以下和/或其中所述免疫应答不是治疗有效应答。

[0111] 在一些方面,所述抗原编码盒编码氨基酸序列VVVGACGVGK(SEQ ID NO:75)、VVVGADGVGK(SEQ ID NO:78)、VVGAVGVGK(SEQ ID NO:79)和ILDTAGHEEY(SEQ ID NO:82)中的每一者的至少4次迭代。

[0112] 在一些方面,所述表位编码核酸序列来源于所述患有癌症的受试者的肿瘤或来源于受感染受试者的细胞或样品。在一些方面,所述表位编码核酸序列不是来源于所述患有癌症的受试者的肿瘤或来源于受感染受试者的细胞或样品。

[0113] 本文还提供了一种用于刺激受试者中的免疫应答的方法,所述方法包括所述方法包括向所述受试者施用基于抗原的疫苗给所述受试者,其中所述基于抗原的疫苗包含抗原编码盒或由所述盒编码的多肽序列,其中所述抗原编码盒包含至少一种由下式从5'至3'描述的抗原编码核酸序列:

[0114] $(E_x\text{-}(E^N_n)_y)_z$

[0115] 其中E表示包含不同表位编码核酸序列的核苷酸序列,n表示单独的不同表位编码

核酸序列的数目并且是包括0的任何整数，E^N表示包含对于每个相应n的单独的不同表位编码核酸序列的核苷酸序列，对于z的每次迭代：x＝0或1，对于每个n，y＝0或1，和x或y中的至少一者＝1，并且z＝2或更大，其中所述抗原编码核酸序列包含E、给定E^N或其组合的至少两次迭代，并且包含所述至少两次迭代的所述不同表位编码核酸序列中的至少一者编码不同的KRAS相关MHC I类新表位。

[0116] 本文还提供了一种用于刺激受试者中的免疫应答的方法，所述方法包括所述方法包括向所述受试者施用基于抗原的疫苗给所述受试者，其中所述基于抗原的疫苗包含：抗原表达系统，所述抗原表达系统包含：抗原表达系统，其中所述抗原表达系统包含一种或多种载体，所述一种或多种载体包含：(a)载体骨架，其中所述骨架包含：(i)至少一种启动子核苷酸序列，和(ii)任选地，至少一种聚腺苷酸化(聚(A))序列；和(b)盒，其中所述盒包含：(i)至少一种抗原编码核酸序列，所述抗原编码核酸序列包含：(I)编码KRAS相关MHC I类新表位的表位编码核酸序列，并且其中所述表位编码核酸序列中的每一者包含：(A)任选地，5'接头序列，和(B)任选地，3'接头序列；(ii)任选地，可操作地连接到所述抗原编码核酸序列的第二启动子核苷酸序列；和(iii)任选地，至少一种MHC II类表位编码核酸序列；(iv)任选地，至少一种编码GPGPG氨基酸接头序列(SEQ ID NO:56)的核酸序列；和(v)任选地，至少一种第二聚(A)序列，其中所述第二聚(A)序列是所述载体骨架的原生聚(A)序列或外源聚(A)序列，其中如果所述第二启动子核苷酸序列不存在，则所述抗原编码核酸序列可操作地连接到所述至少一种启动子核苷酸序列，并且其中所述至少一种抗原编码核酸序列包含编码所述KRAS相关MHC I类新表位的所述表位编码核酸序列中的至少一者的至少两次迭代。

[0117] 本文还提供了一种用于治疗患有疾病的受试者的方法，任选地其中所述疾病是感染性癌症，所述方法包括向所述受试者施用基于抗原的疫苗给所述受试者，其中所述基于抗原的疫苗包含抗原表达系统，所述抗原表达系统包含：抗原表达系统，其中所述抗原表达系统包含一种或多种载体，所述一种或多种载体包含：(a)载体骨架，其中所述骨架包含：(i)至少一种启动子核苷酸序列，和(ii)任选地，至少一种聚腺苷酸化(聚(A))序列；和(b)盒，其中所述盒包含：(i)至少一种抗原编码核酸序列，所述抗原编码核酸序列包含：(I)至少两种不同表位编码核酸序列，任选地包含：(1)使所编码的表位序列与由野生型核酸序列编码的相应肽序列不同的至少一种改变，任选地其中所述至少一种改变是KRAS突变，或(2)编码感染性疾病生物体肽的核酸序列，所述肽选自由以下组成的组：病原体衍生肽、病毒衍生肽、细菌衍生肽、真菌衍生肽和寄生虫衍生肽，并且其中所述表位编码核酸序列中的每一者包含：(A)任选地，5'接头序列，和(B)任选地，3'接头序列；(ii)任选地，可操作地连接到所述抗原编码核酸序列的第二启动子核苷酸序列；和(iii)任选地，至少一种MHC II类表位编码核酸序列；(iv)任选地，至少一种编码GPGPG氨基酸接头序列(SEQ ID NO:56)的核酸序列；和(v)任选地，至少一种第二聚(A)序列，其中所述第二聚(A)序列是所述载体骨架的原生聚(A)序列或外源聚(A)序列，其中如果所述第二启动子核苷酸序列不存在，则所述抗原编码核酸序列可操作地连接到所述至少一种启动子核苷酸序列，并且其中所述盒不编码免疫显性MHC I类表位，所述免疫显性MHC I类表位：(1)当以疫苗组合物的形式施用于受试者时，相对于所述盒中编码并能够刺激所述受试者中的免疫应答的另一种MHC I类表位，刺激5倍或更高的免疫应答，和/或(2)当以疫苗组合物的形式施用于受试者时，相对于当在不存

在所述免疫显性MHC I类表位的情况下施用其他MHC I类表位时的免疫应答,降低了对所述盒中编码的另一种MHC I类表位的免疫应答,任选地其中所述免疫应答降低到检测限以下和/或其中所述免疫应答不是治疗有效应答。

[0118] 在一些方面,所述抗原编码盒编码氨基酸序列VVVGACGVGK(SEQ ID NO:75)、VVVGADGVGK(SEQ ID NO:78)、VVGAVGVGK(SEQ ID NO:79)和ILDTAGHEEY(SEQ ID NO:82)中的每一者的至少4次迭代。

[0119] 在一些方面,所述受试者表达经预测或已知呈递至少一种表位序列的至少一种HLA等位基因,任选地其中所述经预测或已知呈递的至少一种表位序列包含(1)所述KRAS相关MHC I类新表位,和/或(2)所述免疫显性MHC I类表位和所述盒中编码的其他MHC I类表位。在一些方面,所述受试者表达经预测或已知呈递至少一种表位序列的至少一种HLA等位基因,并且其中所述至少一种表位序列包含已知或怀疑由MHC I类呈递在细胞表面上的表位,任选地其中所述经预测或已知呈递的至少一种表位序列包含(1)所述KRAS相关MHC I类新表位,和/或(2)所述免疫显性MHC I类表位和所述盒中编码的其他MHC I类表位。在一些方面,所述细胞表面是肿瘤细胞表面。在一些方面,所述细胞是选自由以下组成的组的肿瘤细胞:肺癌、黑色素瘤、乳腺癌、卵巢癌、前列腺癌、肾癌、胃癌、结肠癌、睾丸癌、头颈癌、胰腺癌、脑癌、B细胞淋巴瘤、急性髓细胞白血病、慢性髓细胞白血病、慢性淋巴细胞白血病、T细胞淋巴细胞白血病、非小细胞肺癌和小细胞肺癌。在一些方面,所述细胞表面是受感染细胞表面。在一些方面,所述细胞是选自由以下组成的组的受感染细胞:病原体感染细胞、病毒感染细胞、细菌感染细胞、真菌感染细胞和寄生虫感染细胞。在一些方面,所述病毒感染细胞选自由以下组成的组:HIV感染细胞、严重急性呼吸综合征相关冠状病毒(SARS)感染细胞、严重急性呼吸综合征冠状病毒2(SARS-CoV-2)感染细胞、埃博拉感染细胞、乙型肝炎病毒(HBV)感染细胞、流感感染细胞、正粘病毒科病毒感染细胞、人乳头瘤病毒(HPV)感染细胞、巨细胞病毒(CMV)感染细胞、基孔肯雅病毒感染细胞、呼吸道合胞病毒(RSV)感染细胞、登革热病毒感染细胞和丙型肝炎病毒(HCV)感染细胞。

[0120] 本文还提供了一种用于诱导受试者中的免疫应答的方法,所述方法包括向所述受试者施用基于抗原的疫苗给所述受试者,其中所述基于抗原的疫苗包含抗原编码盒或由所述盒编码的多肽序列,其中所述抗原编码盒包含至少一种由下式从5'至3'描述的抗原编码核酸序列:

[0121] $(E_x\text{-}(E^N_n)_y)_z$

[0122] 其中E表示包含不同表位编码核酸序列的核苷酸序列,n表示单独的不同表位编码核酸序列的数目并且是包括0的任何整数,E^N表示包含对于每个相应n的单独的不同表位编码核酸序列的核苷酸序列,对于z的每次迭代:x=0或1,对于每个n,y=0或1,和x或y中的至少一者=1,并且z=2或更大,其中所述抗原编码核酸序列包含E、给定E^N或其组合的至少两次迭代,并且包含所述至少两次迭代的所述不同表位编码核酸序列中的至少一者编码不同的KRAS相关MHC I类新表位。

[0123] 本文还提供了一种用于诱导受试者中的免疫应答的方法,所述方法包括向所述受试者施用基于抗原的疫苗给所述受试者,其中所述基于抗原的疫苗包含:抗原表达系统,所述抗原表达系统包含:抗原表达系统,其中所述抗原表达系统包含一种或多种载体,所述一种或多种载体包含:(a)载体骨架,其中所述骨架包含:(i)至少一种启动子核苷酸序列,和

(ii)任选地,至少一种聚腺苷酸化(聚(A))序列;和(b)盒,其中所述盒包含:(i)至少一种抗原编码核酸序列,所述抗原编码核酸序列包含:(I)编码KRAS相关MHC I类新表位的表位编码核酸序列,并且其中所述表位编码核酸序列中的每一者包含:(A)任选地,5'接头序列,和(B)任选地,3'接头序列;(ii)任选地,可操作地连接到所述抗原编码核酸序列的第二启动子核苷酸序列;和(iii)任选地,至少一种MHC II类表位编码核酸序列;(iv)任选地,至少一种编码GPGPG氨基酸接头序列(SEQ ID NO:56)的核酸序列;和(v)任选地,至少一种第二聚(A)序列,其中所述第二聚(A)序列是所述载体骨架的原生聚(A)序列或外源聚(A)序列,其中如果所述第二启动子核苷酸序列不存在,则所述抗原编码核酸序列可操作地连接到所述至少一种启动子核苷酸序列,其中所述至少一种抗原编码核酸序列包含编码所述KRAS相关MHC I类新表位的表位编码核酸序列的至少两次迭代,并且其中所述受试者表达经预测或已知呈递至少一种KRAS相关MHC I类新表位的至少一种HLA等位基因。

[0124] 本文还提供了一种用于诱导受试者中的免疫应答的方法,所述方法包括向所述受试者施用基于抗原的疫苗给所述受试者,其中所述基于抗原的疫苗包含:抗原表达系统,所述抗原表达系统包含:抗原表达系统,其中所述抗原表达系统包含一种或多种载体,所述一种或多种载体包含:(a)载体骨架,其中所述骨架包含:(i)至少一种启动子核苷酸序列,和(ii)任选地,至少一种聚腺苷酸化(聚(A))序列;和(b)盒,其中所述盒包含:(i)至少一种抗原编码核酸序列,所述抗原编码核酸序列包含:(I)至少一种表位编码核酸序列,任选地包含:(1)使所编码的表位序列与由野生型核酸序列编码的相应肽序列不同的至少一种改变,任选地其中所述至少一种改变是KRAS突变,或(2)编码感染性疾病生物体肽的核酸序列,所述肽选自由以下组成的组:病原体衍生肽、病毒衍生肽、细菌衍生肽、真菌衍生肽和寄生虫衍生肽,并且任选地其中所述表位编码核酸序列编码MHC I类表位,并且其中所述表位编码核酸序列中的每一者包含:(A)任选地,5'接头序列,和(B)任选地,3'接头序列;(ii)任选地,可操作地连接到所述抗原编码核酸序列的第二启动子核苷酸序列;和(iii)任选地,至少一种MHC II类表位编码核酸序列;(iv)任选地,至少一种编码GPGPG氨基酸接头序列(SEQ ID NO:56)的核酸序列;和(v)任选地,至少一种第二聚(A)序列,其中所述第二聚(A)序列是所述载体骨架的原生聚(A)序列或外源聚(A)序列,其中如果所述第二启动子核苷酸序列不存在,则所述抗原编码核酸序列可操作地连接到所述至少一种启动子核苷酸序列,并且其中所述盒不编码免疫显性MHC I类表位,所述免疫显性MHC I类表位:(1)当以疫苗组合物的形式施用于受试者时,相对于所述盒中编码并能够刺激所述受试者中的免疫应答的另一种MHC I类表位,刺激5倍或更高的免疫应答,和/或(2)当以疫苗组合物的形式施用于受试者时,相对于当在不存在所述免疫显性MHC I类表位的情况下施用其他MHC I类表位时的免疫应答,降低了对所述盒中编码的另一种MHC I类表位的免疫应答,任选地其中所述免疫应答降低到检测限以下和/或其中所述免疫应答不是治疗有效应答,并且其中所述受试者表达经预测或已知呈递所述免疫显性MHC I类表位和所述盒中编码的其他MHC I类表位两者的至少一种HLA等位基因。

[0125] 在一些方面,所述抗原编码盒编码氨基酸序列VVVGACGVGK(SEQ ID NO:75)、VVVGADGVGK(SEQ ID NO:78)、VVGAVGVGK(SEQ ID NO:79)和ILDTAGHEEY(SEQ ID NO:82)中的每一者的至少4次迭代。在一些方面,所述KRAS相关MHC I类新表位或KRAS突变包含KRAS G12C突变、KRAS G12V突变、KRAS G12D突变或KRAS Q61H突变。在一些方面,所述KRAS相关

MHC I类新表位或KRAS突变包含以SEQ ID NO:75-82所示的氨基酸序列中的任一者。在一些方面,所述抗原编码盒包含以SEQ ID NO:75-82所示的氨基酸序列中的每一者。在一些方面,所述抗原编码盒包含以SEQ ID NO:75-82所示的氨基酸序列中的每一者的两次或更多次迭代。在一些方面,所述抗原编码盒包含以SEQ ID NO:75-82所示的氨基酸序列中的每一者的4次迭代。在一些方面,所述KRAS相关MHC I类新表位或KRAS突变包含以SEQ ID NO:57、SEQ ID NO:58、SEQ ID NO:59或SEQ ID NO:60所示的氨基酸序列。在一些方面,所述表位编码核酸序列包含两个或更多个不同的表位编码核酸序列,所述表位编码核酸序列独立地编码不同的KRAS相关MHC I类新表位或不同的KRAS突变。在一些方面,所述表位编码核酸序列中的每一者独立地编码不同的KRAS相关MHC I类新表位或不同的KRAS突变。在一些方面,所述表位编码核酸序列包含两个或更多个不同的表位编码核酸序列,所述表位编码核酸序列独立地编码KRAS G12C突变、KRAS G12V突变、KRAS G12D突变或KRAS Q61H突变。在一些方面,所述表位编码核酸序列独立地编码KRAS G12C突变、KRAS G12V突变和KRAS G12D突变以及任选地KRAS Q61H突变中的每一者。在一些方面,所述抗原编码核酸序列编码包含以SEQ ID NO:64或SEQ ID NO:65所示的氨基酸序列的肽。在一些方面,所述抗原编码核酸序列编码包含以SEQ ID NO:65所示的氨基酸序列的肽。

[0126] 在一些方面,所述盒不编码免疫显性MHC I类表位,所述免疫显性MHC I类表位:(1)当以疫苗组合物的形式施用于受试者时,相对于所述盒中编码并能够刺激所述受试者中的免疫应答的另一种MHC I类表位,刺激5倍或更高的免疫应答,和/或(2)当以疫苗组合物的形式施用于受试者时,相对于当在不存在所述免疫显性MHC I类表位的情况下施用其他MHC I类表位时的免疫应答,降低了对所述盒中编码的另一种MHC I类表位的免疫应答,任选地其中所述免疫应答降低到检测限以下和/或其中所述免疫应答不是治疗有效应答。在一些方面,所述盒不编码免疫显性MHC I类表位,所述免疫显性MHC I类表位当以疫苗组合物的形式施用于受试者时,相对于所述盒中编码并能够刺激所述受试者中的免疫应答的KRAS相关新表位,刺激5倍或更高的免疫应答。在一些方面,所述盒不编码免疫显性MHC I类表位,所述免疫显性MHC I类表位当以疫苗组合物的形式施用于受试者时,相对于当在不存在所述免疫显性MHC I类表位的情况下施用其他MHC I类表位时的免疫应答,降低了对所述盒中编码的另一种MHC I类表位的免疫应答。在一些方面,所述盒不编码免疫显性MHC I类表位,所述免疫显性MHC I类表位当以疫苗组合物的形式施用于受试者时,相对于当在不存在所述免疫显性MHC I类表位的情况下施用其他MHC I类表位时的免疫应答,将对所述盒中编码的另一种MHC I类表位的免疫应答降低到检测限以下。在一些方面,所述盒不编码免疫显性MHC I类表位,所述免疫显性MHC I类表位当以疫苗组合物的形式施用于受试者时,相对于当在不存在所述免疫显性MHC I类表位的情况下施用其他MHC I类表位时的免疫应答,降低了对所述盒中编码的另一种MHC I类表位的免疫应答,其中对所述其他MHC I类表位的免疫应答不是治疗有效应答。

[0127] 在一些方面,所述免疫显性表位是TP53相关MHC I类新表位,任选地其中所述TP53相关MHC I类新表位包含S127Y突变。

[0128] 在一些方面,所述抗原表达系统包含本文所述的抗原表达系统中的任一者。在一些方面,所述基于抗原的疫苗包含本文所述的药物组合物中的任一者。

[0129] 在一些方面,所述基于抗原的疫苗作为初免剂量施用。在一些方面,所述基于抗原

的疫苗作为一个或多个加强剂量施用。在一些方面,所述加强剂量不同于所述初免剂量。在一些方面,a)所述初免剂量包含黑猩猩腺病毒载体并且所述加强剂量包含甲病毒载体;或b)所述初免剂量包含甲病毒载体载体并且所述加强剂量包含黑猩猩腺病毒载体。在一些方面,所述加强剂量与所述初免剂量相同。在一些方面,所述一个或多个加强剂量的注射部位尽可能靠近所述初免剂量的注射部位。

[0130] 在一些方面,所述方法还包括确定或已经确定所述受试者的HLA-单倍型。

[0131] 在一些方面,所述基于抗原的疫苗是肌内(IM)、皮内(ID)、皮下(SC)或静脉内(IV)施用的。在一些方面,所述基于抗原的疫苗是肌内(IM)施用的。在一些方面,所述IM施用是在单独的注射部位施用。在一些方面,所述单独的注射部位是在相对的三角肌中。在一些方面,所述单独的注射部位是在每一侧的臀肌或股直肌部位。

[0132] 本文还公开了一种药物组合物,所述药物组合物包含本文公开的任何组合物(例如本文公开的基于甲病毒或基于ChAd的载体)和药学上可接受的载剂。在一些方面,所述药物组合物还包含佐剂。在一些方面,所述药物组合物还包含免疫调节剂。在一些方面,所述免疫调节剂是抗CTLA4抗体或其抗原结合片段、抗PD-1抗体或其抗原结合片段、抗PD-L1抗体或其抗原结合片段、抗4-1BB抗体或其抗原结合片段、或抗OX-40抗体或其抗原结合片段。

[0133] 本文还公开了一种包含本文公开的分离的核苷酸序列的载体。

[0134] 本文还公开了一种包含本文公开的载体或组合物和使用说明书的试剂盒。

[0135] 本文还公开了一种用于治疗受试者的方法,所述方法包括向所述受试者施用本文公开的载体或本文公开的药物组合物。本文还公开了一种用于诱导受试者中的免疫应答的方法,所述方法包括向所述受试者施用本文所述的任何组合物、载体或药物组合物。在一些方面,所述受试者表达经预测或已知呈递MHC I类表位的至少一种HLA等位基因。在一些方面,预测或已知呈递MHC I类表位的HLA等位基因是A*03:01、A*11:01、A*02:01、C*01:02和/或A*01:01。在一些方面,预测或已知呈递MHC I类表位的HLA等位基因是A*03:01。在一些方面,预测或已知呈递MHC I类表位的HLA等位基因是A*11:01。在一些方面,预测或已知呈递MHC I类表位的HLA等位基因是A*02:01。在一些方面,预测或已知呈递MHC I类表位的HLA等位基因是C*01:02。在一些方面,预测或已知呈递MHC I类表位的HLA等位基因是A*01:01。在一些方面,所述载体或组合物是肌内(IM)、皮内(ID)或皮下(SC)或静脉内(IV)施用的。

[0136] 本文还公开了一种制造任何上述组合物的一种或多种载体的方法,所述方法包括:获得包含骨架和抗原盒的线性化DNA序列;通过将所述线性化DNA序列添加到包含将所述线性化DNA序列转录成RNA的所有必要组分的体外转录反应中,来体外转录所述线性化DNA序列,任选地还包括将m7g帽体外添加到所得RNA中;以及从所述体外转录反应中分离所述一种或多种载体。在一些方面,所述线性化DNA序列是通过将DNA质粒序列线性化或通过使用PCR扩增而产生的。在一些方面,所述DNA质粒序列是使用细菌重组或全基因组DNA合成或全基因组DNA合成中的一者与所合成DNA在细菌细胞中扩增产生的。在一些方面,从所述体外转录反应中分离所述一种或多种载体涉及苯酚氯仿提取、基于硅胶柱的纯化或类似的RNA纯化方法中的一者或多者。

[0137] 本文还公开了一种制造本文公开的任何组合物的方法,所述方法包括:提供用于纳米微粒递送媒介物的组分;提供抗原表达系统;以及为所述纳米微粒递送媒介物和所述抗原表达系统提供足以产生用于递送所述抗原表达系统的组合物的条件。在一些方面,所

述条件由微流体混合提供。

[0138] 本文还公开了一种制造本文公开的腺病毒载体的方法,所述方法包括:获得包含至少一种启动子序列和抗原盒的质粒序列;将所述质粒序列转染到一种或多种宿主细胞中;以及从所述一种或多种宿主细胞中分离所述腺病毒载体。

[0139] 在一些方面,分离包括:裂解所述宿主细胞以获得包含所述腺病毒载体的细胞裂解物;以及从所述细胞裂解物中纯化所述腺病毒载体。

[0140] 在一些方面,所述质粒序列是使用细菌重组或全基因组DNA合成或全基因组DNA合成中的一者与所合成DNA在细菌细胞中扩增产生的。在一些方面,所述一种或多种宿主细胞是CHO、HEK293或其变体、911、HeLa、A549、LP-293、PER.C6和AE1-2a细胞中的至少一者。在一些方面,从所述细胞裂解物中纯化所述腺病毒载体涉及色谱分离、离心、病毒沉淀和过滤中的一者或多者。

附图说明

[0141] 关于以下描述和附图将更好地理解本发明的这些和其他特征、方面和优势,在所述附图中:

[0142] 图1A呈现了具有KRAS新表位G12C、G12V、G12D和Q61H的单个拷贝("KRAS 1X(20x1)";盒=SEQ ID NO:63),KRAS G12C、G12V、G12D和Q61H新表位的2条重复序列以及其他KRAS新表位的2条重复序列("KRAS 2X(8x2)";盒=SEQ ID NO:64),或KRAS新表位的4条重复序列("KRAS 4X(4x4)";盒=SEQ ID NO:65)的盒的图示。数字标识符分别指的是相对于每个盒而不是跨盒设计的表位"槽"(例如,20x1盒中的槽"3"表位与8x2盒的槽3中的表位不相同)。

[0143] 图1B展示重复表位增加了疫苗诱导的抗原特异性T细胞应答。显示的是重复新表位KRAS G12C的ELISpot结果。使用指定的ChAdV68递送载体用$8x10^{10}$个VP对经工程化以表达人类HLA-A11:01的小鼠进行免疫,并在免疫后14天分离脾细胞。在用VVVGACGVGK(SEQ ID NO:75)过夜刺激后通过IFNg ELISpot测量抗原特异性T细胞的数目。数据呈现为每只动物每$1x10^6$个脾细胞的斑点形成集落(SFC)。条形表示中位数。

[0144] 图1C展示重复表位增加了疫苗诱导的抗原特异性T细胞应答。显示的是重复新表位KRAS G12V的ELISpot结果。使用指定的ChAdV68递送载体用$8x10^{10}$个VP对经工程化以表达人类HLA-A11:01的小鼠进行免疫,并在免疫后14天分离脾细胞。在用VVVGAVGVGK(SEQ ID NO:81)过夜刺激后通过IFNg ELISpot测量抗原特异性T细胞的数目。数据呈现为每只动物每$1x10^6$个脾细胞的斑点形成集落(SFC)。条形表示中位数。虚线表示太多而无法计数的样品(TNTC)。

[0145] 图1D展示重复表位增加了疫苗诱导的抗原特异性T细胞应答。显示的是重复新表位KRAS G12D的ELISpot结果。使用指定的ChAdV68递送载体用$8x10^{10}$个VP对经工程化以表达人类HLA-A11:01的小鼠进行免疫,并在免疫后14天分离脾细胞。在用VVVGADGVGK(SEQ ID NO:78)过夜刺激后通过IFNg ELISpot测量抗原特异性T细胞的数目。数据呈现为每只动物每$1x10^6$个脾细胞的斑点形成集落(SFC)。条形表示中位数。

[0146] 图2A呈现了设计用于评估TP53表位的免疫显性的ChAdV68递送载体,特别是仅含有KRAS新表位G12C、G12V、G12D和Q61H("KRAS 4x1";盒=SEQ ID NO:66),KRAS新表位与

TP53 R213L新表位的组合(“KRAS 4x1+R213L”;盒＝SEQ ID NO:67),以及KRAS新表位与TP53 S127Y新表位的组合(“KRAS 4x1+S127Y”;盒＝SEQ ID NO:68)的载体的图示。

[0147] 图2B展示免疫显性表位的去除增加了疫苗诱导的对KRAS新表位的抗原特异性T细胞应答。显示的是新表位KRAS G12C的ELISpot结果。使用指定的ChAdV68递送载体用$5x10^{10}$个VP对经工程化以表达人类HLA-A11:01的小鼠进行免疫,并在免疫后14天分离脾细胞。在用VVVGACGVGK(SEQ ID NO:75)过夜刺激后通过IFNg ELISpot测量抗原特异性T细胞的数目。数据呈现为每只动物每$1x10^{6}$个脾细胞的斑点形成集落(SFC)。条形表示中位数。

[0148] 图2C展示免疫显性表位的去除增加了疫苗诱导的对KRAS新表位的抗原特异性T细胞应答。显示的是新表位KRAS G12D的ELISpot结果。使用指定的ChAdV68递送载体用$5x10^{10}$个VP对经工程化以表达人类HLA-A11:01的小鼠进行免疫,并在免疫后14天分离脾细胞。在用VVVGADGVGK(SEQ ID NO:78)过夜刺激后通过IFNg ELISpot测量抗原特异性T细胞的数目。数据呈现为每只动物每$1x10^{6}$个脾细胞的斑点形成集落(SFC)。条形表示中位数。

[0149] 图2D展示免疫显性表位的去除增加了疫苗诱导的对KRAS新表位的抗原特异性T细胞应答。显示的是新表位KRAS G12V的ELISpot结果。使用指定的ChAdV68递送载体用$5x10^{10}$个VP对经工程化以表达人类HLA-A11:01的小鼠进行免疫,并在免疫后14天分离脾细胞。在用VVVGAVGVGK(SEQ ID NO:81)过夜刺激后通过IFNg ELISpot测量抗原特异性T细胞的数目。数据呈现为每只动物每$1x10^{6}$个脾细胞的斑点形成集落(SFC)。条形表示中位数。

[0150] 图2E展示了免疫显性表位和相关对照表位的免疫应答。显示的是R213L和S127Y新表位的TP53新表位池的ELISpot结果。使用指定的ChAdV68递送载体用$5x10^{10}$个VP对经工程化以表达人类HLA-A11:01的小鼠进行免疫,并在免疫后14天分离脾细胞。过夜刺激后通过IFNg ELISpot测量抗原特异性T细胞的数目。数据呈现为每只动物每$1x10^{6}$个脾细胞的斑点形成集落(SFC)。条形表示中位数。虚线表示太多而无法计数的样品(TNTC)。

[0151] 图3展示重复表位增加了疫苗诱导的抗原特异性T细胞应答。显示的是重复新表位KRAS G12V(左图)或KRAS G12D(右图)的ELISpot结果。使用指定的ChAdV68递送载体用$5x10^{10}$个VP对经工程化以表达人类HLA-A11:01的小鼠进行免疫,并在免疫后14天分离脾细胞。在分别用VVVGAVGVGK(SEQ ID NO:81)或VVVGADGVGK(SEQ ID NO:78)过夜刺激后通过IFNg ELISpot测量抗原特异性T细胞的数目。数据呈现为每只动物每$1x10^{6}$个脾细胞的斑点形成集落(SFC)。条形表示中位数。

[0152] 图4展示重复表位增加了疫苗诱导的抗原特异性T细胞应答。显示的是重复新表位KRAS G12V(左图)或KRAS G12D(右图)的ELISpot结果。使用指定的ChAdV68递送载体用$7x10^{10}$个VP对经工程化以表达人类HLA-A11:01的小鼠进行免疫,并在免疫后14天分离脾细胞。在分别用VVVGAVGVGK(SEQ ID NO:81)或VVVGADGVGK(SEQ ID NO:78)过夜刺激后通过IFNg ELISpot测量抗原特异性T细胞的数目。数据呈现为每只动物每$1x10^{6}$个脾细胞的斑点形成集落(SFC)。条形表示中位数。

[0153] 图5展示重复表位增加了疫苗诱导的针对KRAS Q61H的抗原特异性T细胞应答。显示的是指定盒格式的重复新表位KRAS Q61H的ELISpot结果。使用指定的ChAdV68递送载体用$5x10^{10}$个VP对经工程化以表达人类HLA-A01:01的小鼠进行免疫,并在免疫后12天分离脾细胞。在用ILDTAGHEEY(SEQ ID NO:82)过夜刺激后通过IFNg ELISpot测量抗原特异性T细胞的数目。数据呈现为每只动物每$1x10^{6}$个脾细胞的斑点形成集落(SFC)。条形表示中位数。

虚线表示太多而无法计数的样品(TNTC)。

[0154] 图6展示重复表位增加了疫苗诱导的针对ChAdV68和SAM载体格式的抗原特异性T细胞应答。显示的是重复新表位KRAS G12V(左图)或KRAS G12D(右图)的ELISpot结果。使用指定的ChAdV68递送载体或10μg指定的SAM载体用$5x10^{10}$个VP对经工程化以表达人类HLA-A11:01的小鼠进行免疫,并在免疫后14天分离脾细胞。在用各自的肽池过夜刺激后通过IFNg ELISpot测量抗原特异性T细胞的数目,所述肽池含有跨越25聚体的所有可能的38个最小表位。数据呈现为每只动物每$1x10^{6}$个脾细胞的斑点形成集落(SFC)。条形表示中位数。从左到右的栏是ChAdV68 20x1、SAM 20x1、ChAdV68 4x4和SAM 4x4。

[0155] 图7展示重复表位增加了疫苗诱导的针对ChAdV68和SAM载体格式的抗原特异性T细胞应答。显示的是重复新表位KRAS G12V(左图)或KRAS G12D(右图)的ELISpot结果。使用指定的ChAdV68递送载体或10μg指定的SAM载体用$5x10^{10}$个VP对经工程化以表达人类HLA-A11:01的小鼠进行免疫,并在免疫后14天分离脾细胞。在用VVGAVGVGK(SEQ ID NO:79)或VVVGADGVGK(SEQ ID NO:78)过夜刺激后通过IFNg ELISpot测量抗原特异性T细胞的数目。数据呈现为每只动物每$1x10^{6}$个脾细胞的斑点形成集落(SFC)。条形表示中位数。从左到右的栏是ChAdV68 20x1、SAM 20x1、ChAdV68 4x4和SAM 4x4。

具体实施方式

[0156] I.定义

[0157] 一般而言,权利要求书和本说明书中所用的术语旨在解释为具有本领域普通技术人员所理解的普通含义。为了更清楚,某些术语定义如下。在普通含义与所提供的定义之间存在矛盾的情况下,将使用所提供的定义。

[0158] 如本文所用,术语“抗原”是刺激免疫应答的物质。抗原可以是新抗原。抗原可以是“共有抗原”,其是在特定群体(例如特定癌症患者群体)中发现的抗原。

[0159] 如本文所用,术语“新抗原”是具有至少一个使其不同于相应野生型抗原的改变的抗原,例如经由肿瘤细胞中的突变或特异性针对肿瘤细胞的翻译后修饰。新抗原可包括多肽序列或核苷酸序列。突变可包括移码或非移码插入/缺失、错义或无义取代、剪接位点改变、基因组重排或基因融合,或产生neoORF的任何基因组或表达改变。突变还可包括剪接变体。特异性针对肿瘤细胞的翻译后修饰可包括异常磷酸化。特异性针对肿瘤细胞的翻译后修饰还可包括蛋白酶体产生的剪接抗原。参见Liepe等人,A large fraction of HLA class I ligands are proteasome-generated spliced peptides;Science.2016年10月21日;354(6310):354-358。可通过使用多种诊断方法(例如下文进一步描述的患者选择方法)来鉴定受试者用于施用。

[0160] 如本文所用,术语“肿瘤抗原”是存在于受试者的肿瘤细胞或组织中但不存在于受试者的相应正常细胞或组织中或者源自已知或已发现与正常细胞或组织相比在肿瘤细胞或癌组织中具有改变表达的多肽的抗原。

[0161] 如本文所用,术语“基于抗原的疫苗”是基于一种或多种抗原(例如多种抗原)的疫苗组合物。疫苗可为基于核苷酸(例如基于病毒、基于RNA或基于DNA)的疫苗、基于蛋白质(例如基于肽)的疫苗或其组合。

[0162] 如本文所用,术语“候选抗原”是产生可代表抗原的序列的突变或其他异常。

[0163] 如本文所用,术语“编码区”是编码蛋白质的基因的一个或多个部分。

[0164] 如本文所用,术语“编码突变”是发生在编码区中的突变。

[0165] 如本文所用,术语“ORF”是指开放阅读框。

[0166] 如本文所用,术语“NEO-ORF”是由突变或其他畸变如剪接而产生的肿瘤特异性ORF。

[0167] 如本文所用,术语“错义突变”是引起从一个氨基酸取代为另一个氨基酸的突变。

[0168] 如本文所用,术语“无义突变”是引起从氨基酸到终止密码子的取代或引起典型起始密码子去除的突变。

[0169] 如本文所用,术语“移码突变”是引起蛋白质框架改变的突变。

[0170] 如本文所用,术语“插入/缺失”是一种或多种核酸的插入或缺失。

[0171] 如本文所用,在两个或更多个核酸或多肽序列的上下文中,术语百分比“同一性”是指当出于最大对应性比较和比对时,两个或更多个序列或子序列具有指定百分比的相同的核苷酸或氨基酸残基,如使用下文所描述的序列比较算法(例如BLASTP和BLASTN或技术人员可用的其他算法)中的一者或通过目视检查所测量。取决于应用,“同一性”百分比可存在于所比较的序列区域上,例如在功能结构域上,或者存在于待比较的两个序列的全长上。

[0172] 关于序列比较,通常一个序列充当与测试序列比较的参考序列。当使用序列比较算法时,将测试和参考序列输入计算机,必要时指定子序列坐标,并且指定序列算法程序参数。然后,序列比较算法根据所指定的程序参数来计算测试序列相对于参考序列的序列同一性百分比。或者,序列相似性或不相似性可通过组合存在或不存在特定核苷酸,或对于翻译序列,在所选择的序列位置(例如序列基序)处的氨基酸来建立。

[0173] 比较序列的最佳比对可例如通过以下进行:Smith和Waterman,Adv.Appl.Math.2:482(1981)的局部同源性算法;Needleman和Wunsch,J.Mol.Biol.48:443(1970)的同源性比对算法;Pearson和Lipman,Proc.Nat'l.Acad.Sci.USA 85:2444(1988)的相似性搜寻方法;这些算法的计算机化实施方式(Wisconsin Genetics软件包中的GAP、BESTFIT、FASTA和TFASTA,Genetics Computer Group,575ScienceDr.,Madison,Wis.);或目视检查(一般参见Ausubel等人,同上)。

[0174] 适合于测定序列同一性百分比和序列相似性的算法的一个实例是BLAST算法,其描述于Altschul等人,J.Mol.Biol.215:403-410(1990)中。执行BLAST分析的软件可通过美国国家生物技术信息中心(National Center for Biotechnology Information)公开获得。

[0175] 如本文所用,术语“非终止或通读”是引起天然终止密码子去除的突变。

[0176] 如本文所用,术语“表位”是通常由抗体或T细胞受体结合的抗原的特异性部分。

[0177] 如本文所用,术语“免疫原性”是例如经由T细胞、B细胞或两者刺激免疫应答的能力。

[0178] 如本文所用,术语“HLA结合亲和力”、“MHC结合亲和力”意指特异性抗原与特异性MHC等位基因之间结合的亲和力。

[0179] 如本文所用,术语“诱饵”是用于从样品中富集特定DNA或RNA序列的核酸探针。

[0180] 如本文所用,术语“变体”是受试者的核酸与用作对照的参考人类基因组之间的差异。

[0181] 如本文所用,术语“变体调用(variant call)”是通常根据测序的变体存在的算法确定。

[0182] 如本文所用,术语“多态性”是生殖系变体,即在个体的所有携带DNA的细胞中发现的变体。

[0183] 如本文所用,术语“体细胞变体”是在个体的非生殖系细胞中产生的变体。

[0184] 如本文所用,术语“等位基因”是基因的形式或基因序列的形式或蛋白质的形式。

[0185] 如本文所用,术语“HLA型”是HLA基因等位基因的补体。

[0186] 如本文所用,术语“无义介导的衰变”或“NMD”是由于过早终止密码子引起的细胞对mRNA的降解。

[0187] 如本文所用,术语“躯干突变”是起源于肿瘤发展早期且存在于大多数肿瘤细胞中的突变。

[0188] 如本文所用,术语“亚克隆突变”是起源于肿瘤发展后期且仅存在于肿瘤细胞子集中的突变。

[0189] 如本文所用,术语“外显子组”是编码蛋白质的基因组的子集。外显子组可为基因组的集合外显子。

[0190] 如本文所用,术语“逻辑回归”是来自统计的二进制数据的回归模型,其中因变量等于1的概率的逻辑被建模为因变量的线性函数。

[0191] 如本文所用,术语“神经网络”是用于分类或回归的机器学习模型,其由多层线性变换,接着是通常经由随机梯度下降和反向传播进行训练的逐元素非线性组成。

[0192] 如本文所用,术语“蛋白质组”是由细胞、细胞群或个体表达和/或翻译的全部蛋白质的集合。

[0193] 如本文所用,术语“肽组”是由MHC-I或MHC-II呈递在细胞表面上的所有肽的集合。肽组可指细胞或细胞集合的特性(例如,肿瘤肽组,意指构成肿瘤的所有细胞的肽组的联合,或感染性疾病肽组,意指被感染性疾病感染的所有细胞的肽组的联合)。

[0194] 如本文所用,术语“ELISPOT”意指酶联免疫吸附斑点测定,其是用于监测人类和动物中的免疫应答的常用方法。

[0195] 如本文所用,术语“葡聚糖肽多聚体”是在流式细胞术中用于抗原特异性T细胞染色的基于葡聚糖的肽-MHC多聚体。

[0196] 如本文所用,术语“耐受性或免疫耐受性”是对一种或多种抗原(例如自身抗原)免疫无应答的状态。

[0197] 如本文所用,术语“中心耐受性”是通过缺失自身反应性T细胞克隆或通过促进自身反应性T细胞克隆分化成免疫抑制性调控T细胞(Treg)而在胸腺中遭受的耐受性。

[0198] 如本文所用,术语“外周耐受性”是通过下调或不激活经受中心耐受性的自身反应性T细胞或促进这些T细胞分化成Treg而在外周遭受的耐受性。

[0199] 术语“样品”可包括通过包括静脉穿刺、排泄、射精、按摩、活组织检查、针抽吸、灌洗样品、刮取、手术切口或干预的手段或本领域中已知的其他手段从受试者获取的单细胞或多细胞或细胞碎片或体液等分试样。

[0200] 术语“受试者”涵盖人类或非人类,无论体内、离体或体外,雄性或雌性的细胞、组织或生物体。术语受试者包括包含人类的哺乳动物。

[0201] 术语“哺乳动物”涵盖人类和非人类，并且包括(但不限于)人类、非人类灵长类动物、犬科动物、猫科动物、鼠科动物、牛科动物、马科动物和猪科动物。

[0202] 术语“临床因素”是指受试者状况的量度，例如疾病活动或严重程度。“临床因素”涵盖受试者健康状况的所有标志物，包括非样品标志物，和/或受试者的其他特征，例如(但不限于)年龄和性别。临床因素可为可在确定条件下评价来自受试者的样品(或样品群体)或受试者而获得的评分、值或值集合。临床因素也可通过标志物和/或其他参数(例如基因表达替代物)来预测。临床因素可包括肿瘤类型、肿瘤亚型、感染类型、感染亚型和吸烟史。

[0203] 术语“源自肿瘤的抗原编码核酸序列”是指例如经由RT-PCR从肿瘤获得的核酸序列;或通过肿瘤测序获得的序列数据，然后使用测序数据例如经由本领域中已知的各种合成或基于PCR的方法合成核酸序列。衍生序列可包括核酸序列变体，例如序列优化的核酸序列变体(例如，密码子优化的和/或以其他方式针对表达优化的)，其编码与从肿瘤获得的相应原生核酸序列相同的多肽序列。

[0204] 术语“源自感染的抗原编码核酸序列”是指例如经由RT-PCR从受感染细胞或感染性疾病生物体获得的核酸序列;或通过对受感染细胞或感染性疾病生物体进行测序获得序列数据，然后使用测序数据例如经由本领域中已知的各种合成或基于PCR的方法合成核酸序列。衍生序列可包括编码与相应的原生感染性疾病生物体核酸序列相同的多肽序列的核酸序列变体，例如序列优化的核酸序列变体(例如，密码子优化的和/或以其他方式针对表达优化的)。衍生序列可包括编码经修饰的感染性疾病生物体多肽序列的核酸序列变体，所述多肽序列相对于原生感染性疾病生物体多肽序列具有一个或多个(例如，1、2、3、4或5个)突变。例如，相对于感染性疾病生物体蛋白质的原生多肽序列，经修饰的多肽序列可具有一个或多个错义突变。

[0205] 术语“甲病毒”是指披膜病毒科(Togaviridae)的成员，并且是正义单链RNA病毒。甲病毒通常分类为旧世界，例如辛德毕斯、罗斯河、马雅罗、基孔肯雅和塞姆利基森林病毒，或新世界，例如东部马脑炎、奥拉、摩根堡、或委内瑞拉马脑炎及其衍生菌株TC-83。甲病毒通常是自我复制RNA病毒。

[0206] 术语“甲病毒骨架”是指允许病毒基因组自我复制的甲病毒的最小序列。最小序列可包括用于非结构蛋白质介导的扩增的保守序列、非结构蛋白质1(nsP1)基因、nsP2基因、nsP3基因、nsP4基因和聚A序列，以及用于亚基因组病毒RNA表达的序列，包括亚基因组(例如26S)启动子元件。

[0207] 术语“用于非结构蛋白质介导的扩增的序列”包括本领域技术人员熟知的甲病毒保守序列元件(CSE)。CSE包括(但不限于)甲病毒5’UTR、51-nt CSE、24-nt CSE、亚基因组启动子序列(例如26S亚基因组启动子序列)、19-nt CSE和甲病毒3’UTR。

[0208] 术语“RNA聚合酶”包括催化由DNA模板产生RNA多核苷酸的聚合酶。RNA聚合酶包括(但不限于)源自噬菌体的聚合酶，包括T3、T7和SP6。

[0209] 术语“脂质”包括疏水性和/或两亲性分子。脂质可为阳离子、阴离子或中性的。脂质可为合成或天然来源的，并且在一些情况下是可生物降解的。脂质可包括胆固醇、磷脂、脂质缀合物，包括(但不限于)聚乙二醇(PEG)缀合物(聚乙二醇化脂质)、蜡、油、甘油酯、脂肪和脂溶性维生素。脂质还可包括二亚油基甲基-4-二甲基氨基丁酸酯(MC3)和MC3样分子。

[0210] 术语“脂质纳米粒子”或“LNP”包括使用含脂质膜围绕水性内部形成的小泡样结

构，也称为脂质体。脂质纳米粒子包括具有通过表面活性剂稳定的固体脂质核心的基于脂质的组合物。核心脂质可为脂肪酸、酰基甘油、蜡和这些表面活性剂的混合物。生物膜脂质，例如磷脂、鞘磷脂、胆汁盐(牛磺胆酸钠)和固醇(胆固醇)，可用作稳定剂。脂质纳米粒子可使用限定比率的不同脂质分子形成，包括(但不限于)限定比率的一种或多种阳离子脂质、阴离子脂质或中性脂质。脂质纳米粒子可将分子包封在外膜壳内，并且随后可与靶细胞接触以将包封的分子递送至宿主细胞细胞溶质。脂质纳米粒子可用非脂质分子修饰或官能化，包括在其表面上。脂质纳米粒子可为单层的(单层)或多层的(多层)。脂质纳米粒子可与核酸复合。单层脂质纳米粒子可与核酸复合，其中核酸在水性内部。多层脂质纳米粒子可与核酸复合，其中核酸在水性内部，或者形成或包夹在其间。

[0211] 缩写:MHC:主要组织相容性复合物;HLA:人白细胞抗原或人MHC基因座;NGS:下一代测序;PPV:阳性预测值;TSNA:肿瘤特异性新抗原;FFPE:福尔马林固定、石蜡包埋;NMD:无义介导的衰变;NSCLC:非小细胞肺癌;DC:树突状细胞。

[0212] 应注意，除非上下文另外明确规定，否则如本说明书和所附权利要求中所用，单数形式“一个”、“一种”和“所述”包括多个指示物。

[0213] 除非上下文特别陈述或以其他方式显而易见，否则如本文中所用，术语“约”应理解为在本领域的正常公差范围内，例如在平均值的2个标准偏差内。约可理解为在陈述值的10％、9％、8％、7％、6％、5％、4％、3％、2％、1％、0.5％、0.1％、0.05％或0.01％内。除非上下文另有明确说明，否则本文所提供的所有数值均通过术语约修饰。

[0214] 本文中未直接定义的任何术语应理解为具有与本发明的技术领域中所理解通常相关的含义。本文论述某些术语，以向从业者提供描述本发明的方面的组合物、装置、方法等以及如何制造或使用其的额外指导。应了解，相同事物可以超过一种方式表示。因此，替代性措辞和同义词可用于本文所论述的术语中的任何一者或多者。无论本文中是否详述或论述术语都无关紧要。提供一些同义词或可取代的方法、材料等。除非明确陈述，否则对一个或数个同义词或等效物的叙述不排除使用其他同义词或等效物。包括术语实例在内的实例的使用仅用于说明性目的，并且不在本文中限制本发明的方面的范围和含义。

[0215] 在本说明书正文内引用的所有参考文献、颁布专利和专利申请均出于所有目的特此通过引用整体并入。

[0216] II.抗原鉴定

[0217] 肿瘤和正常外显子组和转录组的NGS分析研究方法已被描述并应用于抗原鉴定领域。[6,14,15]可考虑某些优化以提高临床环境中抗原鉴定的灵敏度和特异性。这些优化可分为两个领域，与实验室过程相关的领域和与NGS数据分析相关的领域。所描述的研究方法也可应用于鉴定其他环境中的抗原，例如鉴定来自感染性疾病生物体、受试者中的感染或受试者的受感染细胞的抗原的鉴定。优化的实例对于本领域技术人员来说是已知的，例如在美国专利第10,055,540号、美国申请公开第US20200010849A1号、美国申请第16/606,577号以及国际专利申请公开WO2020181240A1、WO/2018/195357和WO/2018/208856中更详细描述的方法，所述文献各自出于所有目的通过引用整体并入本文。

[0218] 用于鉴定抗原(例如，源自肿瘤或感染性疾病生物体的抗原)的方法包括鉴定可能呈递在细胞表面上(例如，由MHC呈递在肿瘤细胞、受感染细胞或免疫细胞(包括专职抗原呈递细胞如树突状细胞)上)和/或可能具免疫原性的抗原。例如，一种这样的方法可包括以下

步骤:从肿瘤、受感染细胞或感染性疾病生物体中获得外显子组、转录组或全基因组核苷酸测序和/或表达数据中的至少一者,其中所述核苷酸测序数据和/或表达数据用于获得代表一组抗原(例如,源自肿瘤或感染性疾病生物体的抗原)中的每一者的肽序列的数据;将每种抗原的肽序列输入到一个或多个呈递模型中,以产生所述抗原中的每一者由一种或多种MHC等位基因在受试者的细胞表面、例如肿瘤细胞或受感染细胞上呈递的数值可能性集合,所述数值可能性集合已至少基于所接收的质谱数据进行鉴定;以及基于所述数值可能性集合选择所述抗原集合的子集以产生经选择抗原集合。

[0219] III.鉴定新抗原中的肿瘤特异性突变

[0220] 本文还公开了用于鉴定某些突变(例如癌细胞中存在的变体或等位基因)的方法。特别是,这些突变可存在于患有癌症的受试者的癌细胞的基因组、转录组、蛋白质组或外显子组中,而非受试者的正常组织中。用于鉴定肿瘤特异性的新抗原(包括共有新抗原)的具体方法是本领域技术人员已知的,例如在美国专利第10,055,540号、美国申请公开第US20200010849A1号以及国际专利申请公开WO/2018/195357和WO/2018/208856中更详细描述的方法,所述文献各自出于所有目的通过引用整体并入本文。对肿瘤特异的共有新抗原的实例更详细地描述于国际专利申请公开WO2019226941A1中,所述文献出于所有目的通过引用整体并入本文。共有新抗原包括但不限于KRAS相关突变(例如KRAS G12C、KRAS G12V、KRAS G12D和/或KR AS Q61H突变)。例如,KRAS相关MHC I类新表位可包括关于野生型(WT)人类KRAS,例如关于以下示例性氨基酸序列的那些突变:MTEYKLVVVGAGGVGKSALTIQLIQNHFVDEYDPTIEDSYRKQVVIDGETCLLDILDTAGQEEYSAMRDQYMRTGEGFLCVFAINNTKSFEDIHHYREQIKRVKDSEDVPMVLVGNKCDLPSRTVDTKQAQDLARSYGIPFIETSAKTRQRVEDAFYTLVREIRQYRLKKISKEEKTPGCVKIKKCIIM(SEQ ID NO:84)。

[0221] 如果肿瘤中的基因突变引起肿瘤中特有的蛋白质的氨基酸序列变化,则认为其可用于肿瘤的免疫靶向。有用的突变包括:(1)非同义突变,导致蛋白质中的氨基酸不同;(2)通读突变,其中终止密码子被修饰或缺失,导致在C末端具有新型肿瘤特异性序列的较长蛋白质的翻译;(3)剪接位点突变,导致在成熟mRNA中包含内含子并因此导致独特的肿瘤特异性蛋白质序列;(4)染色体重排,在2种蛋白质的接合处产生具有肿瘤特异性序列的嵌合蛋白(即基因融合);(5)移码突变或缺失,导致具有新型肿瘤特异性蛋白质序列的新的开放阅读框。突变还可包括非移码插入/缺失、错义或无义取代、剪接位点改变、基因组重排或基因融合、或产生neoORF的任何基因组或表达改变中的一者或多者。

[0222] 由例如肿瘤细胞中的剪接位点、移码、通读或基因融合突变产生的具有突变的肽或突变多肽可通过对肿瘤与正常细胞中的DNA、RNA或蛋白质进行测序来鉴定。

[0223] 突变还可包括先前鉴定的肿瘤特异性突变。已知肿瘤突变可见于癌症体细胞突变目录(COSMIC)数据库。

[0224] 多种方法可用于检测个体的DNA或RNA中特定突变或等位基因的存在。本领域中的进步已提供精确、容易且便宜的大规模SNP基因分型。例如,已描述数种技术,包括动态等位基因特异性杂交(DASH)、微孔板阵列对角线凝胶电泳(MADGE)、焦磷酸测序、寡核苷酸特异性连接、TaqMan系统以及各种DNA“芯片”技术,例如Affymetrix SNP芯片。这些方法利用通常通过PCR扩增靶基因区。仍有其他方法,基于通过侵入性裂解产生小信号分子,接着进行质谱法或固定化挂锁探针和滚环扩增。下文总结了本领域中已知用于检测特异性突变的数

种方法。

[0225] 基于PCR的检测手段可包括同时多重扩增多个标志物。例如,选择PCR引物以产生大小不重叠并且可同时分析的PCR产物是本领域中众所周知的。或者,可用经差异性标记且因此可各自经差异性检测的引物扩增不同的标志物。当然,基于杂交的检测手段允许样品中多个PCR产物的差异检测。本领域中已知其他技术以允许多个标志物的多重分析。

[0226] 已开发数种方法以便于基因组DNA或细胞RNA中单核苷酸多态性的分析。例如,单碱基多态性可通过使用专门的核酸外切酶抗性核苷酸来检测,如例如Mundy,C.R.(美国专利第4,656,127号)中所公开。根据所述方法,允许与紧靠着多形位点3'的等位基因序列互补的引物与获自特定动物或人的靶分子杂交。如果靶分子上的多形位点含有与所存在的特定核酸外切酶抗性核苷酸衍生物互补的核苷酸,则所述衍生物将并入到杂交引物的末端上。这种并入使得引物对核酸外切酶具有抗性,从而允许其检测。由于样品的核酸外切酶抗性衍生物的身份是已知的,因此引物已对核酸外切酶具有抗性的发现揭示靶分子的多形位点中存在的核苷酸与反应中所用的核苷酸衍生物互补。这种方法的优势在于其不需要确定大量无关序列数据。

[0227] 可使用基于溶液的方法确定多形位点的核苷酸的身份。Cohen,D.等人(法国专利2,650,840;PCT申请第WO91/02087号)。如在美国专利第4,656,127号的芒迪方法(Mundy method)中,采用与紧靠着多形位点3'的等位基因序列互补的引物。所述方法使用经标记的双脱氧核苷酸衍生物确定该位点的核苷酸的身份,如果所述核苷酸与多形位点的核苷酸互补,则将并入到引物的末端上。

[0228] 称为遗传位分析或GBA的替代方法由Goelet,P.等人(PCT申请第92/15712号)描述。Goelet,P.等人的方法使用经标记的终止子和与多形位点3'序列互补的引物的混合物。所并入的经标记的终止子因此通过所评价的靶分子的多形位点中存在的核苷酸确定且与其互补。与Cohen等人(法国专利2,650,840;PCT申请第WO91/02087号)的方法相比,Goelet,P.等人的方法可为非均相测定,其中引物或靶分子固定于固相。

[0229] 已描述数种用于测定DNA中多形位点的引物引导的核苷酸并入程序(Komher,J.S.等人,Nucl.Acids.Res.17:7779-7784(1989);Sokolov,B.P.,Nucl.Acids Res.18:3671(1990);Syvanen,A.-C.等人,Genomics 8:684-692(1990);Kuppuswamy,M.N.等人,Proc.Natl.Acad.Sci.(U.S.A.)88:1143-1147(1991);Prezant,T.R.等人,Hum.Mutat.1:159-164(1992);Ugozzoli,L.等人,GATA 9:107-112(1992);Nyren,P.等人,Anal.Biochem.208:171-175(1993))。这些方法与GBA的不同之处在于其利用并入经标记的脱氧核苷酸来区分多形位点处的碱基。在这种格式中,由于信号与并入的脱氧核苷酸的数量成比例,因此在同一核苷酸的操作中发生的多态性可产生与操作的长度成比例的信号(Syvanen,A.-C.等人,Amer.J.Hum.Genet.52:46-59(1993))。

[0230] 许多方案直接从数百万个单独的DNA或RNA分子中并行获取序列信息。实时单分子边合成边测序技术依赖于荧光核苷酸的检测,因为它们被并入到与正测序的模板互补的DNA的新生链中。在一种方法中,将长度为30-50个碱基的寡核苷酸在5'端共价锚定于玻璃盖玻片上。这些锚定链执行两种功能。首先,如果模板被配置成具有与表面结合的寡核苷酸互补的捕捉尾部,则其充当靶模板链的捕捉位点。它们还充当模板引导的引物延伸的引物,形成序列阅读的基础。捕捉引物充当固定位点以便使用多个合成、检测和化学裂解染料接

头来去除染料的循环进行序列测定。每个循环包括:添加聚合酶/经标记的核苷酸混合物,冲洗,成像和染料的裂解。在一个替代方法中,聚合酶被荧光供体分子修饰并固定在载玻片上,而每个核苷酸用连接到γ-磷酸的受体荧光部分进行颜色编码。系统检测经荧光标记的聚合酶与经荧光修饰的核苷酸之间的相互作用,因为核苷酸并入到从头链中。也存在其他边合成边测序技术。

[0231] 可使用任何合适的边合成边测序平台来鉴定突变。如上所述,目前可用四种主要边合成边测序平台:来自Roche/454Life Sciences的基因组测序仪、来自Illumina/Solexa的1G分析仪、来自Applied BioSystems的SOLiD系统和来自Helicos Biosciences的Heliscope系统。边合成边测序平台也已由Pacific BioSciences和VisiGen Biotechnologies描述。在一些实施方案中,将所测序的多个核酸分子结合于支撑物(例如固体支撑物)。为将核酸固定于支撑物上,可在模板的3'和/或5'端添加捕捉序列/通用引发位点。核酸可通过将捕捉序列与共价连接于支撑物的互补序列杂交而结合于支撑物。捕捉序列(也称为通用捕捉序列)是与连接到支撑物的序列互补的核酸序列,其可双重充当通用引物。

[0232] 作为捕捉序列的替代方案,偶联对(例如抗体/抗原、受体/配体或如例如美国专利申请第2006/0252077号中所述的抗生物素蛋白-生物素对)的一个成员可连接到每个片段,以被捕捉在用该偶联对的相应第二成员包被的表面上。

[0233] 在捕捉后,可例如通过单分子检测/测序来分析序列,例如,如实施例和美国专利第7,283,337号中所述,包括模板依赖性边合成边测序。在边合成边测序中,表面结合的分子在聚合酶存在下暴露于多个经标记的三磷酸核苷酸。模板的序列通过并入到生长链的3'端的经标记的核苷酸的顺序来确定。这可实时进行或可以分步重复模式进行。对于实时分析,可将不同的光学标记并入各核苷酸并且可利用多个激光刺激并入的核苷酸。

[0234] 测序还可包括其他大规模平行测序或下一代测序(NGS)技术和平台。大规模平行测序技术和平台的额外实例是Illumina HiSeq或MiSeq、Thermo PGM或Proton、Pac Bio RS II或Sequel、Qiagen的Gene Reader和Oxford Nanopore MinION。可使用其他类似的当前大规模平行测序技术,以及这些技术的后代。

[0235] 可利用任何细胞类型或组织来获得用于本文所述的方法的核酸样品。例如,DNA或RNA样品可获自肿瘤或体液,例如通过已知技术(例如静脉穿刺)获得的血液或唾液。或者,可对干燥样品(例如头发或皮肤)执行核酸测试。另外,可从肿瘤获得样品用于测序并且可从正常组织获得另一样品用于测序,其中正常组织具有与肿瘤相同的组织类型。可从肿瘤获得样品用于测序并且可从正常组织获得另一样品用于测序,其中正常组织相对于肿瘤具有不同的组织类型。

[0236] 肿瘤可包括肺癌、黑色素瘤、乳腺癌、卵巢癌、前列腺癌、肾癌、胃癌、结肠癌、睾丸癌、头颈癌、胰腺癌、脑癌、B细胞淋巴瘤、急性髓细胞白血病、慢性髓细胞白血病、慢性淋巴细胞白血病、和T细胞淋巴细胞白血病、非小细胞肺癌和小细胞肺癌中的一者或多者。

[0237] 或者,可使用蛋白质质谱法鉴定或验证与肿瘤细胞上的MHC蛋白质结合的突变肽的存在。肽可从肿瘤细胞或从肿瘤免疫沉淀的HLA分子酸洗脱,然后使用质谱法鉴定。

[0238] IV.抗原

[0239] 抗原可包括核苷酸或多肽。例如,抗原可为编码多肽序列的RNA序列。可用于疫苗

中的抗原可因此包括核苷酸序列或多肽序列。

[0240] 本文公开了包含通过本文公开的方法鉴定的肿瘤特异性突变的分离肽、包含已知肿瘤特异性突变的肽和通过本文公开的方法鉴定的突变多肽或其片段。新抗原肽可描述于其编码序列的上下文中，其中新抗原包括编码相关多肽序列的核苷酸序列(例如DNA或RNA)。

[0241] 具体而言，本文公开了包括KRAS相关MHC I类新表位的迭代的盒。KRAS相关MHC I类新表位包括但不限于具有KRAS G12突变和/或KRAS Q61突变的新表位。盒可包括具有KRAS G12突变的KRAS相关MHC I类新表位的迭代。盒可包括具有KRAS Q61突变的KRAS相关MHC I类新表位的迭代。盒可包括具有KRAS G12C、KRAS G12V、KRAS G12D和/或KRAS Q61H突变的KRAS相关MHC I类新表位的迭代。盒可包括具有KRAS G12C突变的KRAS相关MHC I类新表位的迭代。盒可包括具有KRAS G12V突变的KRAS相关MHC I类新表位的迭代。盒可包括具有KRAS G12D突变的KRAS相关MHC I类新表位的迭代。盒可包括具有KRAS Q61H突变的KRAS相关MHC I类新表位的迭代。盒可包括具有KRAS G12C、KRAS G12V、KRAS G12D和KRAS Q61H突变的KRAS相关MHC I类新表位中的每一者的迭代。盒可包括至少两个不同的KRAS相关MHC I类新表位的迭代，所述新表位选自由以下组成的组：KRAS G12C、KRAS G12V、KRAS G12D和KRAS Q61H突变。盒可包括至少三个不同的KRAS相关MHC I类新表位的迭代，所述新表位选自由以下组成的组：KRAS G12C、KRAS G12V、KRAS G12D和KRAS Q61H突变。盒可包括仅单个不同的KRAS相关MHC I类新表位的迭代。盒可包括具有KRAS G12C突变的仅单个不同KRAS相关MHC I类新表位的迭代。盒可包括具有KRAS G12D突变的仅单个不同KRAS相关MHC I类新表位的迭代。盒可包括具有KRAS G12V突变的仅单个不同KRAS相关MHC I类新表位的迭代。盒可包括具有KRAS Q61H突变的仅单个不同KRAS相关MHC I类新表位的迭代。

[0242] 具有KRAS G12C突变的KRAS相关MHC I类新表位包括VVVGACGVGK(SEQ ID NO:75)或KLVVVGACGV(SEQ ID NO:76)。具有KRAS G12D突变的KRAS相关MHC I类新表位包括VVGADGVGK(SEQ ID NO:77)或VVVGADGVGK(SEQ ID NO:78)，具有KRAS G12V突变的KRAS相关MHC I类新表位包括VVGAVGVGK(SEQ ID NO:79)、VVVGAVGVGK(SEQ ID NO:81)或AVGVGKSAL(SEQ ID NO:80)。

[0243] 盒可包括具有氨基酸序列VVVGACGVGK(SEQ ID NO:75)、VVVGADGVGK(SEQ ID NO:78)、VVGAVGVGK(SEQ ID NO:79)和ILDTAGHEEY(SEQ ID NO:82)的KRAS相关MHC I类新表位中的每一者的迭代。盒可包括至少两个不同的KRAS相关MHC I类新表位的迭代，所述新表位具有选自由以下组成的组的氨基酸序列：VVVGACGVGK(SEQ ID NO:75)、VVVGADGVGK(SEQ ID NO:78)、VVGAVGVGK(SEQ ID NO:79)和ILDTAGHEEY(SEQ ID NO:82)。盒可包括至少三个不同的KRAS相关MHC I类新表位的迭代，所述新表位具有选自由以下组成的组的氨基酸序列：VVVGACGVGK(SEQ ID NO:75)、VVVGADGVGK(SEQ ID NO:78)、VVGAVGVGK(SEQ ID NO:79)和ILDTAGHEEY(SEQ ID NO:82)。盒可包括具有氨基酸序列VVVGACGVGK(SEQ ID NO:75)、VVVGADGVGK(SEQ ID NO:78)、VVGAVGVGK(SEQ ID NO:79)和ILDTAGHEEY(SEQ ID NO:82)的KRAS相关MHC I类新表位中的至少一者的迭代。

[0244] KRAS相关MHC I类新表位可包括原生KRAS蛋白情形下治疗性疫苗表位的原生N末端和/或C末端侧翼序列。KRAS相关MHC I类新表位的说明性非限制性实例是25聚体MTEYKLVVVGACGVGKSALTIQLIQ(SEQ ID NO:57)(对于KRAS G12C)，

MTEYKLVVVGADGVGKSALTIQLIQ(SEQ ID NO:58)(对于KRAS G12D),MTEYKLVVVGAVGVGKSALTIQLIQ(SEQ ID NO:59)(对于KRAS G12V)和ETCLLDILDTAGHEEYSAMRDQYMR(SEQ ID NO:60)(对于KRAS Q61H)。包括原生侧翼序列的KRAS相关MHC I类新表位可连接(串接)到盒中编码的其他新表位,包括包含其各自原生侧翼序列的其他新表位(例如,其他KRAS相关MHC I类新表位)。通过其原生侧翼序列连接并且包括具有突变KRAS G12C、KRAS G12D、KRAS G12V和KRAS Q61H的KRAS新表位中的每一者的4次迭代的串接的KRAS相关MHC I类新表位的说明性非限制性盒由SEQ ID NO:65所示的氨基酸序列表示。

[0245] 编码KRAS相关MHC I类新表位的表位编码核酸序列,例如包括原生N末端和/或C末端侧翼序列的那些,可编码多个已知和/或预测的KRAS相关MHC I类新表位。作为一个说明性实例,KRAS G12V 25聚体MTEYKLVVVGAVGVGKSALTIQLIQ(SEQ ID NO:59)编码已知和/或预测的KRAS相关MHC I类新表位VVGAVGVGK(SEQ ID NO:79)、VVVGAVGVGK(SEQ ID NO:81)和AVGVGKSAL(SEQ ID NO:80)中的每一者。

[0246] 表位编码核酸序列,包括编码KRAS相关MHC I类新表位的那些,可在盒中按任何顺序排列。表位编码核酸序列,包括编码KRAS相关MHC I类新表位的那些,可按最小化连接表位的顺序排列,如本文进一步描述的。作为一个说明性非限制性实例,连接在一起以最小化连接表位的串接的KRAS相关MHC I类新表位由SEQ ID NO:65所示的氨基酸序列表示并具有以下顺序:G12C G12D Q61H G12DG12V G12C Q61H G12D G12V G12C Q61H G12D G12V Q61H G12VG12C。

[0247] 本文还公开了这样的肽,其源自已知或已发现与正常细胞或组织相比在肿瘤细胞或癌组织中具有改变表达的任何多肽,例如已知或已发现与正常细胞或组织相比在肿瘤细胞或癌组织中异常表达的任何多肽。可例如在COSMIC数据库中发现可获得抗原肽的合适的多肽。COSMIC整理了关于人类癌症体细胞突变的全面信息。所述肽含有肿瘤特异性突变。肿瘤抗原(例如,共有肿瘤抗原和肿瘤新抗原)可包括但不限于美国申请第17/058,128号中描述的那些,所述文献出于所有目的通过引用并入本文。抗原肽可在其编码序列的上下文中描述,其中抗原包括编码相关多肽序列的核苷酸序列(例如,DNA或RNA)。

[0248] 本文还公开了衍生自与感染性疾病生物体、受试者中的感染或受试者的受感染细胞相关的任何多肽的肽。抗原可来源于感染性疾病生物体的核苷酸序列或多肽序列。感染性疾病生物体的多肽序列包括但不限于病原体衍生肽、病毒衍生肽、细菌衍生肽、真菌衍生肽和/或寄生虫衍生肽。感染性疾病生物体包括但不限于严重急性呼吸综合征相关冠状病毒(SARS)、严重急性呼吸综合征冠状病毒2(SARS-CoV-2)、埃博拉、HIV、乙型肝炎病毒(HBV)、流感、丙型肝炎病毒(HCV)、人乳头瘤病毒(HPV)、巨细胞病毒(CMV)、基孔肯雅病毒、呼吸道合胞病毒(RSV)、登革热病毒、正粘病毒科病毒和肺结核。

[0249] 本文公开了包含通过本文公开的方法鉴定的感染性疾病生物体特异性抗原或表位的分离肽、包含已知感染性疾病生物体特异性抗原或表位的肽、以及通过本文公开的方法鉴定的突变多肽或其片段。抗原肽可以在其编码序列的上下文中进行描述,其中抗原包括编码相关多肽序列的核苷酸序列(例如,DNA或RNA)。

[0250] 本文所述的载体和相关组合物可用于递送来自任何生物体的抗原,包括它们的毒素或其他副产物,以预防和/或治疗与生物体或其副产物相关的感染或其他不良反应。

[0251] 可并入疫苗(例如,编码在盒中)的抗原包括可用于使人类或非人类动物免疫以抵抗病毒,例如感染人类和非人类脊椎动物的病原病毒的免疫原。抗原可选自多种病毒家族。需要免疫应答的理想病毒家族的实例包括小核糖核酸病毒家族,其包括引起约50%普通感冒病例的鼻病毒属;肠道病毒属,其包括脊髓灰质炎病毒、柯萨奇病毒(coxsackievirus)、埃可病毒(echovirus)和人类肠道病毒如甲型肝炎病毒;以及口蹄疫病毒属,其主要在非人类动物中引起口蹄疫。在小核糖核酸病毒家族中,靶抗原包括VP1、VP2、VP3、VP4和VPG。另一个病毒家族包括杯状病毒家族,其涵盖诺沃克(Norwalk)病毒群,它们是流行性胃肠炎的重要病原体。另一个可用于靶向抗原以在人类和非人类动物中刺激免疫应答的理想病毒家族是披膜病毒家族,其包括甲病毒属,其包括辛德毕斯病毒、罗斯河病毒和委内瑞拉、东方和西方马脑炎,以及风疹病毒属(rubivirus),包括风疹病毒。黄病毒科包括登革热、黄热病、日本脑炎、圣路易斯脑炎和蜱传脑炎病毒。其他靶抗原可能产生自丙型肝炎或冠状病毒家族,其包括多种非人类病毒,例如传染性支气管炎病毒(家禽)、猪传染性胃肠病毒(猪)、猪血凝性脑脊髓炎病毒(猪)、猫传染性腹膜炎病毒(猫)、猫肠道冠状病毒(猫)、犬冠状病毒(狗)和人类呼吸道冠状病毒,其可导致普通感冒和/或非甲、乙或丙型肝炎。在冠状病毒家族中,靶抗原包括E1(也称为M或基质蛋白)、E2(也称为S或刺突蛋白)、E3(也称为HE或血凝素-埃尔特糖(elterose))糖蛋白(并非在所有冠状病毒中都存在)或N(核衣壳)。还有其他抗原可靶向弹状病毒家族,其包括水疱病毒属(例如,水疱性口炎病毒)和一般的狂犬病病毒(例如,狂犬病)。在弹状病毒家族中,合适的抗原可来源于G蛋白或N蛋白。丝状病毒科(其包括出血热病毒如马尔堡病毒和埃博拉病毒)可能是合适的抗原来源。副粘病毒家族包括1型副流感病毒、3型副流感病毒、3型牛副流感病毒、风疹病毒(腮腺炎病毒)、2型副流感病毒、4型副流感病毒、新城疫病毒(鸡)、牛瘟、麻疹病毒,其包括麻疹和犬瘟热,以及肺炎病毒,其包括呼吸道合胞病毒(例如,糖(G)蛋白和融合(F)蛋白,其序列可从GenBank获得)。流感病毒属于正粘病毒科并且可以是合适的抗原来源(例如,HA蛋白、N1蛋白)。布尼亚病毒家族包括布尼亚病毒属(加利福尼亚脑炎、拉克罗斯(La Crosse))、白蛉病毒(裂谷热)、汉坦病毒(puremala是一种血红热病毒)、内罗病毒(内罗毕绵羊病)和各种未指定的布尼亚病毒(bungavirus)。沙粒病毒家族提供了抗LCM和拉沙热病毒的抗原来源。呼肠孤病毒家族包括呼肠孤病毒属、轮状病毒属(其导致儿童急性胃肠炎)、环状病毒和培养病毒(科罗拉多蜱热、利庞博(Lebombo)(人类)、马脑病、蓝舌病)。逆转录病毒家族包括涵盖人类和兽医学疾病如猫白血病病毒、HTLVI和HTLVII、慢病毒(其包括人类免疫缺陷病毒(HIV)、猿免疫缺陷病毒(SIV)、猫免疫缺陷病毒(FIV)、马传染性贫血病毒和泡沫病毒(spumavirinal))的致癌病毒亚科。在慢病毒中,已经描述了许多合适的抗原并且可容易地选择。合适的HIV和SIV抗原的实例包括但不限于gag、pol、Vif、Vpx、VPR、Env、Tat、Nef和Rev蛋白,以及它们的各种片段。例如,Env蛋白的合适片段可包括其任何亚基,例如gp120、gp160、gp41,或其更小的片段,例如,长度至少约8个氨基酸。类似地,可选择tat蛋白的片段。[参见美国专利第5,891,994号和美国专利第6,193,981号。]还参见D.H.Barouch等人,J.Virol.,75(5):2462-2467(2001年3月)和R.R.Amara等人,Science,292:69-74(2001年4月6日)中描述的HIV和SIV蛋白。在另一个实例中,HIV和/或SIV免疫原性蛋白或肽可用于形成融合蛋白或其他免疫原性分子。参见例如2001年8月2日公开的WO 01/54719和1999年4月8日公开的WO 99/16884中描述的HIV-1Tat和/或Nef融合蛋白和免疫方案。本发明不限于本文所述的HIV和/或SIV免疫

原性蛋白质或肽。此外，对这些蛋白质的多种修饰已被描述或可由本领域技术人员容易地进行。参见例如美国专利第5,972,596号中描述的经修饰的gag蛋白。此外，任何所需的HIV和/或SIV免疫原可单独或组合递送。此类组合可包括来自单个载体或来自多个载体的表达。乳多空病毒家族包括多瘤病毒亚科(BKU和JCU病毒)和乳头瘤病毒亚科(与癌症或乳头瘤的恶性进展相关)。腺病毒家族包括引起呼吸道疾病和/或肠炎的病毒(EX、AD7、ARD、O.B.)。细小病毒家族猫细小病毒(猫肠炎)、猫泛白细胞减少病毒、犬细小病毒和猪细小病毒。疱疹病毒家族包括α疱疹病毒亚科，其涵盖单纯病毒属(HSVI、HSVII)、水痘病毒(伪狂犬病、水痘带状疱疹)，和β疱疹病毒亚科，其包括巨细胞病毒属(人CMV、鼠巨细胞病毒)，和γ疱疹病毒亚科，其包括淋巴潜隐病毒属、EBV(伯基特淋巴瘤)、传染性鼻气管炎、马立克氏病(Marek's disease)病毒和细长病毒属(rhadinovirus)。痘病毒家族包括脊索动物痘病毒亚科(chordopoxyirinae)，其涵盖正痘病毒属(天花(Variola/Smallpox)和牛痘(Vaccinia/Cowpox))、副痘病毒、禽痘病毒、山羊痘病毒、兔痘病毒、猪痘病毒和昆虫痘病毒亚科。肝炎病毒病毒包括乙型肝炎病毒。一种可能是合适的抗原来源的未分类病毒是丁型肝炎病毒。其他病毒来源可包括禽传染性法氏囊病病毒以及猪呼吸和生殖综合征病毒。甲病毒家族包括马动脉炎病毒和各种脑炎病毒。

[0252] 可并入疫苗中(例如，编码在盒中)的抗原还包括可用于免疫人类或非人类动物以抵抗病原体的免疫原，所述病原体包括感染人类和非人类脊椎动物的细菌、真菌、寄生微生物或多细胞寄生虫。细菌病原体的实例包括致病性革兰氏阳性球菌，包括肺炎球菌;葡萄球菌;和链球菌。致病性革兰氏阴性球菌包括脑膜炎球菌;淋球菌。致病性肠道革兰氏阴性杆菌包括肠杆菌科(enterobacteriaceae);假单胞菌属(pseudomonas)、不动杆菌属(acinetobacteria)和艾肯氏菌属(eikenella);类鼻疽;沙门氏菌属(salmonella);志贺氏菌属(shigella);嗜血杆菌属(haemophilus)(流感嗜血杆菌(Haemophilus influenzae)、睡眠嗜血杆菌(Haemophilus somnus));莫拉菌属(moraxella);杜克雷嗜血杆菌(H.ducreyi)(其导致软下疳);布鲁氏菌属(brucella);土拉热弗朗西斯菌(Franisella tularensis)(其导致土拉菌病);耶尔森氏菌属(yersinia)(巴氏杆菌属(pasteurella));串珠状链杆菌(streptobacillus moniliformis)和螺旋菌属(spirillum)。革兰氏阳性杆菌包括单核细胞增生李斯特菌(listeria monocytogenes);红斑丹毒丝菌(erysipelothrix rhusiopathiae);白喉棒状杆菌(Corynebacterium diphtheria)(白喉);霍乱;炭疽芽胞杆菌(B.anthracis)(炭疽病);多诺瓦斯病(donovanosis)(腹股沟肉芽肿);和巴尔通体病(bartonellosis)。由致病性厌氧菌引起的疾病包括破伤风;肉毒杆菌中毒;其他梭菌;肺结核;麻风;和其他分枝杆菌。具体细菌种类的实例是但不限于肺炎链球菌(Streptococcus pneumoniae)、化脓性链球菌(Streptococcus pyogenes)、无乳链球菌(Streptococcus agalactiae)、粪链球菌(Streptococcus faecalis)、卡他莫拉菌(Moraxella catarrhalis)、幽门螺杆菌(Helicobacter pylori)、脑膜炎奈瑟菌(Neisseria meningitidis)、淋病奈瑟菌(Neisseria gonorrhoeae)、沙眼衣原体(Chlamydia trachomatis)、肺炎衣原体(Chlamydia pneumoniae)、鹦鹉热衣原体(Chlamydia psittaci)、百日咳杆菌(Bordetella pertussis)、伤寒杆菌(Salmonella typhi)、鼠伤寒沙门氏菌(Salmonella typhimurium)、霍乱沙门氏菌(Salmonella choleraesuis)、大肠杆菌(Escherichia coli)、志贺氏菌(Shigella)、霍乱弧菌(Vibrio

cholerae)、白喉棒杆菌(Corynebacterium diphtheriae)、结核分枝杆菌(Mycobacterium tuberculosis)、鸟分枝杆菌(Mycobacterium avium)、胞内分枝杆菌(Mycobacterium intracellulare)复合体、奇异变形杆菌(Proteus mirabilis)、普通变形杆菌(Proteus vulgaris)、金黄色葡萄球菌(Staphylococcus aureus)、破伤风梭菌(Clostridium tetani)、问号钩端螺旋体(Leptospira interrogans)、伯氏疏螺旋体(Borrelia burgdorferi)、溶血巴斯德菌(Pasteurella haemolytica)、多杀巴斯德菌(Pasteurella multocida)、胸膜肺炎放线杆菌(Actinobacillus pleuropneumoniae)和鸡毒支原体(Mycoplasma gallisepticum)。致病性螺旋体疾病包括梅毒;密螺旋体病:雅司病(yaws)、品他病(pinta)和地方性梅毒;和钩端螺旋体病。由高等病原体细菌和病原真菌引起的其他感染包括放线菌病;诺卡氏菌病;隐球菌病(隐球菌)、芽生菌病(芽生菌)、组织胞浆菌病(组织胞浆菌)和球孢子菌病(球孢子菌);念珠菌病(念珠菌)、曲霉病(曲霉)和毛霉菌病;孢子丝菌病;副球孢子菌病、皮氏菌病、圆球菌病、足菌肿和色霉菌病;和皮肤癣菌病。立克次体感染包括斑疹伤寒、落基山斑疹热、Q热和立克次体痘。支原体和衣原体感染的实例包括:肺炎支原体(mycoplasma pneumoniae);性病性淋巴肉芽肿;鹦鹉热;和围产期衣原体感染。致病性真核生物涵盖致病性原生动物和蠕虫,并且由此产生的感染包括:阿米巴病;疟疾;利什曼病(例如,由硕大利什曼原虫(Leishmania major)引起);锥虫病;弓形虫病(例如,由刚地弓形虫(Toxoplasma gondii)引起);卡氏肺孢子虫(Pneumocystis carinii);Trichans;刚地弓形虫;巴贝虫病;贾第鞭毛虫病(例如,由贾第鞭毛虫(Giardia)引起);旋毛虫病(例如,由滴虫(Trichomonas)引起);丝虫病;血吸虫病(例如,由血吸虫(Schistosoma)引起);线虫;吸虫(trematodes/fluke);和绦虫(cestode/tapeworm)感染。其他寄生虫感染可能由蛔虫(Ascaris)、鞭虫(Trichuris)、隐孢子虫(Cryptosporidium)和卡氏肺孢子虫等引起。

[0253] 本文还公开了衍生自与感染性疾病生物体、受试者中的感染或受试者的受感染细胞相关的任何多肽的肽。抗原可来源于感染性疾病生物体的核酸序列或多肽序列。感染性疾病生物体的多肽序列包括但不限于病原体衍生肽、病毒衍生肽、细菌衍生肽、真菌衍生肽和/或寄生虫衍生肽。感染性疾病生物体包括但不限于严重急性呼吸综合征相关冠状病毒(SARS)、严重急性呼吸综合征冠状病毒2(SARS-CoV-2)、埃博拉、HIV、乙型肝炎病毒(HBV)、流感、丙型肝炎病毒(HCV)、人乳头瘤病毒(HPV)、巨细胞病毒(CMV)、基孔肯雅病毒、呼吸道合胞病毒(RSV)、登革热病毒、正粘病毒科病毒和肺结核。

[0254] 可选择预计呈递在细胞例如肿瘤细胞、受感染细胞或免疫细胞(包括专职抗原呈递细胞如树突状细胞)的细胞表面上的抗原。可选择预计具有免疫原性的抗原。

[0255] 由抗原核苷酸序列编码的一种或多种多肽可包含以下中的至少一者:IC50值小于1000nM的与MHC的结合亲和力,对于MHC I类肽,长度为8-15、8、9、10、11、12、13、14或15个氨基酸,在肽内或肽附近存在促进蛋白酶体切割的序列基序,以及存在促进TAP转运的序列基序。对于长度为6-30、6、7、8、9、10、11、12、13、14、15、16、17、18、19、20、21、22、23、24、25、26、27、28、29或30个氨基酸的MHC II类肽,在肽内或附近存在促进通过胞外或溶酶体蛋白酶(例如组织蛋白酶)切割或HLA-DM催化的HLA结合的序列基序。

[0256] 一种或多种抗原可呈递在肿瘤的表面上。一种或多种抗原可呈递在受感染细胞的表面上。

[0257] 一种或多种抗原在患有肿瘤的受试者中可以是免疫原性的,例如,能够在受试者

中刺激T细胞应答和/或B细胞应答。一种或多种抗原在患有或怀疑患有感染的受试者中可以是免疫原性的，例如，能够在受试者中刺激T细胞应答和/或B细胞应答。一种或多种抗原在处于感染风险的受试者中可以是免疫原性的，例如，能够在受试者中刺激T细胞应答和/或B细胞应答，从而提供针对感染的免疫保护(即免疫性)，例如，刺激记忆T细胞、记忆B细胞和/或感染特异性抗体的产生。

[0258] 一种或多种抗原能够刺激B细胞应答，例如产生识别所述一种或多种抗原的抗体(例如识别肿瘤或感染性疾病抗原的抗体)。抗体可识别线性多肽序列或识别二级和三级结构。因此，B细胞抗原可包括线性多肽序列或具有二级和三级结构的多肽，包括但不限于全长蛋白质、蛋白质亚基、蛋白质结构域或已知或预测具有二级和三级结构的任何多肽序列。能够刺激B细胞对肿瘤或感染性疾病抗原的应答的抗原可为分别在肿瘤细胞或感染性疾病生物体表面发现的抗原。能够引发B细胞对肿瘤或感染性疾病抗原的应答的抗原可为分别在肿瘤或感染性疾病生物体中表达的细胞内新抗原。

[0259] 一种或多种抗原可包括能够刺激T细胞应答的抗原(例如，包括预测的T细胞表位序列的肽)和能够刺激B细胞应答的不同抗原(例如，全长蛋白质、蛋白质亚基、蛋白质结构域)的组合。

[0260] 在针对受试者产生疫苗的情况下，可排除在受试者中刺激自身免疫应答的一种或多种抗原。

[0261] 至少一种抗原肽分子(例如，表位序列)的大小可包含但不限于约5、约6、约7、约8、约9、约10、约11、约12、约13、约14、约15、约16、约17、约18、约19、约20、约21、约22、约23、约24、约25、约26、约27、约28、约29、约30、约31、约32、约33、约34、约35、约36、约37、约38、约39、约40、约41、约42、约43、约44、约45、约46、约47、约48、约49、约50、约60、约70、约80、约90、约100、约110、约120或更多个氨基分子残基，和可源自其中的任何范围。在具体实施方案中，抗原肽分子等于或小于50个氨基酸。

[0262] 抗原肽和多肽可为：对于MHC I类，15个残基或更少的长度并且通常由约8个与约11个之间的残基、特别是9或10个残基组成；对于MHC II类，6-30个残基(包括端值)。

[0263] 如果需要，可以数种方式设计较长的肽。在一种情况下，当HLA等位基因上肽的呈递可能性经预测或已知时，较长的肽可由以下任一者组成：(1)个别呈递的具有朝向每个相应基因产物的N末端和C末端延伸2-5个氨基酸的肽；(2)所呈递的肽中的一些或全部与各自的延伸序列的串接。在另一种情况下，当测序揭示肿瘤中存在长(>10个残基)新表位序列(例如归因于产生新型肽序列的移码、通读或内含子包含)时，较长的肽将由以下组成：(3)新型肿瘤特异性或感染性疾病特异性氨基酸的整个区段，从而绕过了对最强HLA呈递的较短肽的基于计算或体外测试的选择的需要。在这两种情况下，使用较长的肽允许患者细胞进行内源性加工，并且可引起更有效的抗原呈递和刺激T细胞应答。较长的肽还可包括肽的全长蛋白质、蛋白质亚基、蛋白质结构域及其组合，例如分别在肿瘤或感染性疾病生物体中表达的那些。可包括较长的肽(例如，全长蛋白质、蛋白质亚基、蛋白质结构域)及其组合以刺激B细胞应答。

[0264] 抗原肽和多肽可呈递在HLA蛋白上。在一些方面，抗原肽和多肽以比野生型肽更大的亲和力呈递在HLA蛋白上。在一些方面，抗原肽或多肽的IC50可至少小于5000nM、至少小于1000nM、至少小于500nM、至少小于250nM、至少小于200nM、至少小于150nM、至少小于

100nM、至少小于50nM或更小。

[0265] 在一些方面，抗原肽和多肽在施用至受试者时不诱发自身免疫应答和/或引起免疫耐受性。

[0266] 还提供了包含至少两种或更多种抗原肽的组合物。在一些实施方案中，所述组合物含有至少两种不同的肽。至少两种不同的肽可衍生自相同的多肽。不同的多肽是指肽因长度、氨基酸序列或两者而异。肽可包括肿瘤特异性突变。肿瘤特异性肽可衍生自已知或已发现含有肿瘤特异性突变的任何多肽，或衍生自已知或已发现与正常细胞或组织相比在肿瘤细胞或癌组织中具有改变表达的任何多肽的肽，例如已知或已发现与正常细胞或组织相比在肿瘤细胞或癌组织中异常表达的任何多肽。肽可衍生自已知或怀疑与感染性疾病生物体相关的任何多肽，或衍生自已知或已发现与正常细胞或组织相比在受感染细胞中具有改变表达的任何多肽(例如，感染性疾病多核苷酸或多肽，包括在宿主细胞中表达受限的感染性疾病多核苷酸或多肽)的肽。可衍生抗原肽的合适多肽可见于例如COSMIC数据库或AACR基因组学证据瘤形成信息交换(Genomics Evidence Neoplasia Information Exchange，GENIE)数据库。COSMIC整理了关于人类癌症体细胞突变的全面信息。AACR GENIE汇总临床级癌症基因组数据并将其与数万名癌症患者的临床结果联系起来。在一些方面，所述肿瘤特异性突变是特定癌症类型的驱动突变。肽可包括KRAS突变(例如，KRAS G12C、KRAS G12V、KRAS G12D和/或KRAS Q61H突变)。

[0267] 具有所需活性或特性的抗原肽和多肽可经修饰以提供某些所需属性，例如改善的药理学特征，同时增加或至少保留基本上所有未经修饰的肽与所需MHC分子结合和激活适当T细胞的生物活性。例如，抗原肽和多肽可进行各种变化，例如保守或非保守取代，其中此类变化可在其使用中提供某些优势，例如改进的MHC结合、稳定性或呈递。保守取代意指用生物和/或化学类似的另一氨基酸残基置换氨基酸残基(例如，一个疏水性残基置换另一个氨基酸残基，或一个极性残基置换另一氨基酸残基)。取代包括以下组合，例如Gly、Ala；Val、Ile、Leu、Met；Asp、Glu；Asn、Gln；Ser、Thr；Lys、Arg；和Phe、Tyr。单氨基酸取代的效应也可使用D-氨基酸探测。此类修饰可使用熟知的肽合成程序进行，如例如Merrifield，Science 232:341-347(1986)，Barany和Merrifield，The Peptides，Gross和Meienhofer编(N.Y.，Academic Press)，第1-284页(1979)；以及Stewart和Young，Solid Phase Peptide Synthesis，(Rockford，Ill.，Pierce)，第2版(1984)中所述。

[0268] 肽和多肽用各种氨基酸模拟物或非天然氨基酸修饰可在提高肽和多肽的体内稳定性方面特别有用。稳定性可以多种方式加以测定。例如，肽酶和各种生物介质(例如人类血浆和血清)已用于测试稳定性。参见例如Verhoef等人，Eur.J.Drug Metab Pharmacokin.11:291-302(1986)。肽的半衰期可使用25％人类血清(v/v)测定方便地确定。方案一般如下。汇集的人类血清(AB型，非热灭活)在使用之前通过离心去脂。血清然后用RPMI组织培养基稀释至25％并用于测试肽稳定性。在预定时间间隔下，移出少量反应溶液并添加至6％三氯乙酸或乙醇水溶液中。将混浊的反应样品冷却(4℃)15分钟，并且然后旋转集结沉淀的血清蛋白质。然后使用稳定性特定的色谱条件通过反相HPLC来判定肽的存在。

[0269] 肽和多肽可经修饰以提供除改进的血清半衰期以外的所需属性。例如，肽刺激CTL活性的能力可通过与含有至少一个能够刺激T辅助细胞应答的表位的序列连接来增强。免

疫原性肽/T辅助细胞缀合物可通过间隔分子连接。间隔子通常由相对较小的中性分子(例如氨基酸或氨基酸模拟物)构成,其在生理条件下基本上不带电。间隔子通常选自例如Ala、Gly或非极性氨基酸或中性极性氨基酸的其他中性间隔子。应理解,任选存在的间隔子无需由相同残基组成,并且因此可为杂寡聚物或均寡聚物。当存在时,间隔子将通常是至少一个或两个残基,更通常是三至六个残基。或者,肽可在无间隔子的情况下连接到T辅助肽。

[0270] 抗原肽可直接或经由在肽的氨基或羧基端处的间隔子连接到T辅助肽。抗原肽或T辅助肽的氨基端可被酰化。示例性T辅助肽包括破伤风类毒素830-843、流感307-319、疟疾环子孢子382-398和378-389。

[0271] 蛋白质或肽可通过本领域技术人员已知的任何技术制造,包括经由标准分子生物学技术表达蛋白质、多肽或肽;从天然来源分离蛋白质或肽;或化学合成蛋白质或肽。先前已公开对应于各种基因的核苷酸和蛋白质、多肽和肽序列,并且可见于本领域普通技术人员已知的计算机化数据库中。一种此类数据库是位于美国国家卫生研究院(National Institutes of Health)网站的美国国家生物技术信息中心(National Center for Biotechnology Information)的Genbank和GenPept数据库。已知基因的编码区可使用本文所公开或本领域普通技术人员应知晓的技术扩增和/或表达。或者,蛋白质、多肽和肽的各种市售制剂是本领域技术人员已知的。

[0272] 在另一方面,抗原包括编码抗原肽或其部分的核酸(例如多核苷酸)。多核苷酸可为例如DNA、cDNA、PNA、CNA、RNA(例如mRNA)、单链和/或双链、或原生或稳定化形式的多核苷酸,例如具有硫代磷酸骨架的多核苷酸,或其组合,并且其可含有或可不含内含子。可对编码抗原的多核苷酸序列进行序列优化以改进表达,例如通过改进转录、翻译、转录后加工和/或RNA稳定性。例如,可对编码抗原的多核苷酸序列进行密码子优化。本文中的"密码子优化"是指就给定生物体的密码子偏倚而言,用经常使用的同义密码子替换不经常使用的密码子。可对多核苷酸序列进行优化以改进转录后加工,例如优化以减少意外剪接,例如通过去除剪接基序(例如,典型和/或隐秘/非典型剪接供体、分支和/或受体序列)和/或引入外源剪接基序(例如剪接供体、分支和/或受体序列)以偏向偏好的剪接事件。外源内含子序列包括但不限于源自SV40(例如,SV40微型内含子)和源自免疫球蛋白(例如,人β-珠蛋白基因)的那些。外源内含子序列可并入在启动子/增强子序列和抗原序列之间。Callendret等人(Virology.2007年7月5日;363(2):288-302)更详细地描述了用于表达载体的外源内含子序列,所述文献出于所有目的通过引用并入本文。可对多核苷酸序列进行优化以改进转录物稳定性,例如通过去除RNA不稳定性基序(例如,富含AU的元件和3'UTR基序)和/或重复核苷酸序列。可对多核苷酸序列进行优化以改进准确转录,例如通过去除隐秘的转录起始子和/或终止子。可对多核苷酸序列进行优化以改进翻译和翻译准确性,例如通过去除隐秘的AUG起始密码子、过早的聚A序列和/或二级结构基序。可对多核苷酸序列进行优化以改进转录物的核输出,例如通过添加组成型转运元件(CTE)、RNA转运元件(RTE)或土拨鼠转录后调控元件(WPRE)。Callendret等人(Virology.2007年7月5日;363(2):288-302)更详细地描述了用于表达载体的核输出信号,所述文献出于所有目的通过引用并入本文。多核苷酸序列可关于GC含量进行优化,例如以反映给定生物体的平均GC含量。序列优化可平衡一种或多种序列特性,例如转录、翻译、转录后加工和/或RNA稳定性。序列优化可产生平衡了转录、翻译、转录后加工和RNA稳定性中的每一者的最佳序列。序列优化算法是本领域技术人员已

知的,例如GeneArt(Thermo Fisher)、Codon Optimization Tool(IDT)、Cool Tool(新加坡大学)、SGI-DNA(La Jolla California)。可分别对抗原编码蛋白的一个或多个区域进行序列优化。

[0273] 另一方面提供一种能够表达多肽或其部分的表达载体。不同细胞类型的表达载体在本领域中已熟知并且无需过度实验便可选择。一般而言,DNA以适当定向插入至表达载体(例如质粒)中并以正确阅读框进行表达。如果需要,DNA可连接到由所需宿主识别的适当转录和翻译调节控制核苷酸序列,但此类控制件一般可用于表达载体中。载体然后经由标准技术引入至宿主中。指导可见于例如Sambrook等人(1989)Molecular Cloning,A Laboratory Manual,Cold Spring Harbor Laboratory,Cold Spring Harbor,N.Y.中。

[0274] V.疫苗组合物

[0275] 本文还公开了一种免疫原性组合物,例如疫苗组合物,其能够引起特异性免疫应答,例如肿瘤特异性免疫应答或感染性疾病生物体特异性免疫应答。疫苗组合物通常包含一种或多种抗原,例如,使用本文描述的方法选择或选自病原体衍生肽、病毒衍生肽、细菌衍生肽、真菌衍生肽和/或寄生虫衍生肽。疫苗组合物也可以称为疫苗。

[0276] 疫苗可含有1至30种肽;2、3、4、5、6、7、8、9、10、11、12、13、14、15、16、17、18、19、20、21、22、23、24、25、26、27、28、29或30种不同的肽;6、7、8、9、10、11、12、13或14种不同的肽;或12、13或14种不同的肽。肽可包括翻译后修饰。疫苗可含有1至100或更多种核苷酸序列;2、3、4、5、6、7、8、9、10、11、12、13、14、15、16、17、18、19、20、21、22、23、24、25、26、27、28、29、30、31、32、33、34、35、36、37、38、39、40、41、42、43、44、45、46、47、48、49、50、51、52、53、54、55、56、57、58、59、60、61、62、63、64、65、66、67、68、69、70、71、72、73、74、75、76、77、78、79、80、81、82、83、84、85、86、87、88、89、90、91、92、93、94、95、96、97、98、99、100或更多种不同的核苷酸序列;6、7、8、9、10、11、12、13或14种不同的核苷酸序列;或12、13或14种不同的核苷酸序列。疫苗可含有1至30种抗原序列;2、3、4、5、6、7、8、9、10、11、12、13、14、15、16、17、18、19、20、21、22、23、24、25、26、27、28、29、30、31、32、33、34、35、36、37、38、39、40、41、42、43、44、45、46、47、48、49、50、51、52、53、54、55、56、57、58、59、60、61、62、63、64、65、66、67、68、69、70、71、72、73、74、75、76、77、78、79、80、81、82、83、84、85、86、87、88、89、90、91、92、93、94、95、96、97、98、99、100或更多种不同的抗原序列;6、7、8、9、10、11、12、13或14种不同的抗原序列;或12、13或14种不同的抗原序列。

[0277] 疫苗可含有1至30种抗原编码核酸序列;2、3、4、5、6、7、8、9、10、11、12、13、14、15、16、17、18、19、20、21、22、23、24、25、26、27、28、29、30、31、32、33、34、35、36、37、38、39、40、41、42、43、44、45、46、47、48、49、50、51、52、53、54、55、56、57、58、59、60、61、62、63、64、65、66、67、68、69、70、71、72、73、74、75、76、77、78、79、80、81、82、83、84、85、86、87、88、89、90、91、92、93、94、95、96、97、98、99、100或更多种不同的抗原编码核酸序列;6、7、8、9、10、11、12、13或14种不同的抗原编码核酸序列;或12、13或14种不同的抗原编码核酸序列。抗原编码核酸序列可指“抗原盒”的抗原编码部分。本文更详细地描述了抗原盒的特征。抗原编码核酸序列可含有一种或多种表位编码核酸序列(例如,编码串接的T细胞表位的抗原编码核酸序列)。

[0278] 疫苗可含有1至30种不同的表位编码核酸序列;2、3、4、5、6、7、8、9、10、11、12、13、14、15、16、17、18、19、20、21、22、23、24、25、26、27、28、29、30、31、32、33、34、35、36、37、38、

39、40、41、42、43、44、45、46、47、48、49、50、51、52、53、54、55、56、57、58、59、60、61、62、63、64、65、66、67、68、69、70、71、72、73、74、75、76、77、78、79、80、81、82、83、84、85、86、87、88、89、90、91、92、93、94、95、96、97、98、99、100或更多种不同的表位编码核酸序列;6、7、8、9、10、11、12、13或14种不同的表位编码核酸序列;或12、13或14种不同的表位编码核酸序列。表位编码核酸序列可指单个表位序列的序列,例如编码串接的T细胞表位的抗原编码核酸序列中的T细胞表位中的每一者。

[0279] 疫苗可含有表位编码核酸序列的至少两次迭代。如本文所用,“迭代”(或可互换地“重复序列”)是指抗原编码核酸序列内两个或更多个相同的核酸表位编码核酸序列(包括本文所述的任选的5’接头序列和/或任选的3’接头序列)。在一个实例中,盒的抗原编码核酸序列部分编码表位编码核酸序列的至少两次迭代。在其他非限制性实例中,盒的抗原编码核酸序列部分编码超过一个不同的表位,并且至少一个不同的表位由编码所述不同表位的核酸序列的至少两次迭代编码(即,至少两个不同的表位编码核酸序列)。在说明性的非限制性实例中,抗原编码核酸序列编码由表位编码核酸序列表位编码序列A(E_A)、表位编码序列B(E_B)和表位编码序列C(E_C)编码的表位A、B和C,并且具有至少一个不同表位的迭代的示例性抗原编码核酸序列由但不限于下式说明:

[0280] -一个不同表位的迭代(表位A的迭代):

[0281] E_A-E_B-E_C-E_A;或

[0282] E_A-E_A-E_B-E_C

[0283] -多个不同表位的迭代(表位A、B和C的迭代):

[0284] E_A-E_B-E_C-E_A-E_B-E_C;或

[0285] E_A-E_A-E_B-E_B-E_C-E_C

[0286] -多个不同表位的多次迭代(表位A、B和C的迭代):

[0287] E_A-E_B-E_C-E_A-E_B-E_C-E_A-E_B-E_C;或

[0288] E_A-E_A-E_A-E_B-E_B-E_B-E_C-E_C-E_C

[0289] 上述实例不是限制性的,并且具有至少一个不同表位的迭代的抗原编码核酸序列可以任何顺序或频率编码每个不同表位。例如,顺序和频率可以是不同表位的随机排列,例如,在具有表位A、B和C的实例中,通过式E_A-E_B-E_C-E_C-E_A-E_B-E_A-E_C-E_A-E_C-E_C-E_B。

[0290] 本文还提供了一种抗原编码盒,所述抗原编码盒具有至少一个由下式从5’至3’描述的抗原编码核酸序列:

[0291] $(E_x\text{-}(E^N_n)_y)_z$

[0292] 其中E表示包括不同表位编码核酸序列的核苷酸序列,

[0293] n表示单独的不同的表位编码核酸序列的数目并且是包括0的任何整数,

[0294] E^N表示包含对于每个相应n的单独不同的表位编码核酸序列的核苷酸序列,

[0295] 对于z的每次迭代:x=0或1,对于每个n,y=0或1,和x或y中的至少一者=1,并且

[0296] z=2或更大,其中所述抗原编码核酸序列包含E、给定E^N或其组合的至少两次迭代。在一些方面,具有至少两次迭代的所述不同的表位编码核酸序列中的至少一者编码KRAS相关MHC I类新表位。

[0297] 每个E或E^N可独立地包含本文所述的任何表位编码核酸序列(例如编码感染性疾病T细胞表位和/或新抗原表位的肽)。例如,每个E或E^N可独立地包含由式($L5_b$-N_c-$L3_d$)从5’

至3'描述的核苷酸序列,其中N包含与每个E或E^N相关的不同表位编码核酸序列,其中c=1,L5包含5'接头序列,其中b=0或1,并且L3包含3'接头序列,其中d=0或1。本文进一步描述了可使用的表位和接头。

[0298] 表位编码核酸序列的迭代(包括任选的5'接头序列和/或任选的3'接头序列)可彼此直接线性连接(例如,如上所示的E_A-E_A-...)。表位编码核酸序列的迭代可被一个或多个额外的核苷酸序列分开。一般而言,表位编码核酸序列的迭代可被适用于本文所述组合物的任何大小的核苷酸序列分开。在一个实例中,表位编码核酸序列的迭代可由单独的不同的表位编码核酸序列分开(例如,如上所示的E_A-E_B-E_C-E_A…)。在迭代被单个单独的不同表位编码核酸序列分开并且每个表位编码核酸序列(包括任选的5'接头序列和/或任选的3'接头序列)编码长度为25个氨基酸的肽的实例中,迭代可被75个核苷酸分开,例如在以E_A-E_B-E_A…表示的抗原编码核酸中,E_A被75个核苷酸分开。在一个说明性实例中,具有编码25聚体抗原Trp1(VTNTEMFVTAPDNLGYMYEVQWPGQ(SEQ ID NO:86))和Trp2(TQPQIANCSVYDFFVWLHYYSVRDT(SEQ ID NO:87))的迭代的序列VTNTEMFVTAPDNLGYMYEVQWPGQTQPQIANCSVYDFFVWLHYYSVRDTVTNTEMFVTAPDNLGYMYEVQWPGQTQPQIANCSVYDFFVWLHYYSVRDT(SEQ ID NO:85)的抗原编码核酸序列,Trp1的迭代由25聚体Trp2分开并且因此Trp1表位编码核酸序列的重复序列由75个核苷酸的Trp2表位编码核酸序列分开。在迭代由2、3、4、5、6、7、8或9个单独的不同表位编码核酸序列分开并且每个表位编码核酸序列(包括任选的5'接头序列和/或任选的3'接头序列)编码长度为25个氨基酸的肽的实例中,所述迭代可分别由150、225、300、375、450、525、600或675个核苷酸分开。

[0299] 在一个实施方案中,选择不同肽和/或多肽或编码其的核苷酸序列,使得肽和/或多肽能够与不同MHC分子(例如不同MHC I类分子和/或不同MHC II类分子)结合。在一些方面,一种疫苗组合物包含能够与最常出现的MHC I类分子和/或不同MHC II类分子缔合的肽和/或多肽的编码序列。因此,疫苗组合物可包含能够与至少2个优选的、至少3个优选的或至少4个优选的MHC I类分子和/或不同MHC II类分子缔合的不同片段。

[0300] 疫苗组合物能够刺激特异性细胞毒性T细胞应答和/或特异性辅助T细胞应答。疫苗组合物能够刺激特异性细胞毒性T细胞应答和特异性辅助T细胞应答。

[0301] 疫苗组合物能够刺激特异性B细胞应答(例如,抗体应答)。

[0302] 疫苗组合物能够刺激特异性细胞毒性T细胞应答、特异性辅助T细胞应答和/或特异性B细胞应答。疫苗组合物能够刺激特异性细胞毒性T细胞应答和特异性B细胞应答。疫苗组合物能够刺激特异性辅助T细胞应答和特异性B细胞应答。疫苗组合物能够刺激特异性细胞毒性T细胞应答、特异性辅助T细胞应答和特异性B细胞应答。

[0303] 疫苗组合物还可包含佐剂和/或载剂。有用的佐剂和载剂的实例在下文中给出。组合物可与载剂缔合,例如蛋白质或抗原呈递细胞,例如能够将肽呈递至T细胞的树突状细胞(DC)。

[0304] 佐剂是混合至疫苗组合物中增加或以其他方式修饰对抗原的免疫应答的任何物质。载剂可为支架结构,例如能够与抗原缔合的多肽或多糖。任选地,佐剂是共价或非共价缀合的。

[0305] 佐剂提高对抗原的免疫应答的能力通常显现为免疫介导性反应的显著或实质性增加或疾病症状的减少。例如,体液免疫的增强典型地显现为针对抗原所产生的抗体滴度

的显著增大,并且T细胞活性的增强典型地显现为细胞增殖或细胞性细胞毒性或细胞因子分泌的增强。佐剂还可改变免疫应答,例如通过将主要体液或Th应答变为主要细胞或Th应答。

[0306] 合适的佐剂包括(但不限于)1018ISS、明矾、铝盐、Amplivax、AS15、BCG、CP-870,893、CpG7909、CyaA、dSLIM、GM-CSF、IC30、IC31、咪喹莫特(Imiquimod)、ImuFact IMP321、IS Patch、ISS、ISCOMATRIX、JuvImmune、LipoVac、MF59、单磷酰基脂质A、Montanide IMS 1312、Montanide ISA 206、Montanide ISA 50V、Montanide ISA-51、OK-432、OM-174、OM-197-MP-EC、ONTAK、PepTel载体系统、PLG微粒、雷西莫特(resiquimod)、SRL172、病毒颗粒和其他病毒样颗粒、YF-17D、VEGF捕获剂、R848、β-葡聚糖、Pam3Cys、Aquila的源自皂素的QS21刺激子(Aquila Biotech,Worcester,Mass.,USA)、分支杆菌提取物和合成细菌细胞壁模拟物,以及其他专用佐剂,例如Ribi的Detox.Quil或Superfos。例如弗氏(Freund's)不完全或GM-CSF的佐剂是有用的。先前已描述对树突状细胞具有特异性的数种免疫佐剂(例如MF59)及其制备(Dupuis M等人,Cell Immunol.1998;186(1):18-27;Allison A C;Dev Biol Stand.1998;92:3-11)。还可使用细胞因子。数种细胞因子已直接关联于:影响树突状细胞迁移至淋巴组织(例如TNF-α)、加速树突状细胞成熟变为T-淋巴细胞的有效抗原呈递细胞(例如GM-CSF、IL-1和IL-4)(美国专利第5,849,589号,其通过引用整体明确并入本文)以及充当免疫佐剂(例如IL-12)(Gabrilovich D I等人,J Immunother Emphasis Tumor Immunol.1996(6):414-418)。

[0307] 还已报道CpG免疫刺激性寡核苷酸增强佐剂在疫苗环境中的效应。还可使用其他TLR结合分子,例如结合RNA的TLR 7、TLR 8和/或TLR 9。

[0308] 有用佐剂的其他实例包括(但不限于)经化学修饰的CpG(例如CpR、Idera)、聚(I:C)(例如聚i:CI2U)、非CpG细菌DNA或RNA以及免疫活性小分子和抗体,例如环磷酰胺、舒尼替尼(sunitinib)、贝伐单抗(bevacizumab)、西乐葆(celebrex)、NCX-4016、西地那非(sildenafil)、他达拉非(tadalafil)、伐地那非(vardenafil)、索拉菲尼(sorafinib)、XL-999、CP-547632、帕佐泮尼(pazopanib)、ZD2171、AZD2171、伊匹单抗、曲美木单抗(tremelimumab)和SC58175,其可起治疗作用和/或充当佐剂。佐剂和添加剂的量和浓度可容易由本领域技术人员确定而无需过度实验。额外佐剂包括集落刺激因子,例如颗粒细胞巨噬细胞集落刺激因子(GM-CSF,沙格司亭(sargramostim))。

[0309] 疫苗组合物可包含超过一种不同的佐剂。此外,治疗性组合物可包含任何佐剂物质,包括以上各者中的任一者或其组合。还预期,疫苗和佐剂可一起或以任何适当的顺序分开施用。

[0310] 载剂(或赋形剂)可独立于佐剂存在。载剂的功能可例如是增加特定突变体的分子量以提高活性或免疫原性、赋予稳定性、增加生物活性或增加血清半衰期。此外,载剂可辅助呈递肽至T细胞。载剂可为本领域技术人员已知的任何合适的载剂,例如蛋白质或抗原呈递细胞。载剂蛋白质可为(但不限于)匙孔螺血氰蛋白、血清蛋白质(例如转铁蛋白)、牛血清白蛋白、人类血清白蛋白、甲状腺球蛋白或卵白蛋白、免疫球蛋白或激素,例如胰岛素或棕榈酸。为了用于人类免疫接种,载剂通常是生理学上可接受的载剂,其为人类可接受的并且安全的。然而,破伤风类毒素和/或白喉类毒素是合适的载剂。或者,载剂可为葡聚糖,例如琼脂糖。

[0311] 细胞毒性T细胞(CTL)识别呈结合至MHC分子的肽形式而非完整外来抗原自身的抗原。MHC分子本身位于抗原呈递细胞的细胞表面上。因此,如果存在肽抗原、MHC分子和APC的三聚体复合物,则可能激活CTL。相应地,如果不仅将肽用于激活CTL,而且如果另外添加具有相应MHC分子的APC,则可加强免疫应答。因此,在一些实施方案中,疫苗组合物另外含有至少一种抗原呈递细胞。

[0312] 抗原还可包括于基于病毒载体的疫苗平台中,例如牛痘、禽痘、自我复制甲病毒、马拉巴病毒(marabavirus)、腺病毒(参见例如Tatsis等人,Adenoviruses,Molecular Therapy(2004)10,616-629)或慢病毒,包括(但不限于)第二、第三或杂交第二/第三代慢病毒和任一代的重组慢病毒,其设计成靶向特定细胞类型或受体(参见例如Hu等人,Immunization Delivered by Lentiviral Vectors for Cancer and Infectious Diseases,Immunol Rev.(2011)239(1):45-61;Sakuma等人,Lentiviral Vectors:basic to translational,Biochem J.(2012)443(3):603-18;Cooper等人,Rescue of splicing-mediated intron loss maximizes expression in lentiviral vectors containing the human ubiquitin C promoter,Nucl.Acids Res.(2015)43(1):682-690;Zufferey等人,Self-Inactivating Lentivirus Vector for Safe and Efficient In Vivo Gene Delivery,J.Virol.(1998)72(12):9873-9880)。取决于上述基于病毒载体的疫苗平台的包装能力,这种方法可递送编码一种或多种抗原肽的一种或多种核苷酸序列。序列可侧接非突变序列,可由接头分开或者可在前面有一种或多种靶向亚细胞区室的序列(参见例如Gros等人,Prospective identification of neoantigen-specific lymphocytes in the peripheral blood of melanoma patients,Nat Med.(2016)22(4):433-8;Stronen等人,Targeting of cancer neoantigens with donor-derived T cell receptor repertoires,Science.(2016)352(6291):1337-41;Lu等人,Efficient identification of mutated cancer antigens recognized by T cells associated with durable tumor regressions,Clin Cancer Res.(2014)20(13):3401-10)。在引入宿主中后,受感染细胞表达抗原,从而刺激针对一种或多种肽的宿主免疫(例如CTL)应答。可用于免疫方案中的牛痘载体和方法描述于例如美国专利第4,722,848号中。另一种载体是卡介苗(Bacille Calmet te Guerin,BCG)。BCG载体描述于Stover等人(Nature 351:456-460(1991))中。根据本文描述,可用于抗原的治疗性施用或免疫接种的各种其他疫苗载体,例如伤寒沙门氏菌载体等,对于本领域技术人员将是显而易见的。

[0313] V.A.抗原盒

[0314] 鉴于本文提供的教导,用于选择一种或多种抗原、“抗原盒”的克隆和构建以及将其插入病毒载体中的方法在本领域技术范围内。“抗原盒”或“盒”是指经选择的抗原或多种抗原(例如,抗原编码核酸序列)与转录抗原和表达转录产物所必需的其他调控元件的组合。经选择的抗原或多种抗原可指不同的表位序列,例如,盒中的抗原编码核酸序列可编码表位编码核酸序列(或多个表位编码核酸序列),使得表位被转录和表达。一种抗原或多种抗原可以允许转录的方式可操作地连接到调控组件。此类组件包括可驱动一种或多种抗原在用病毒载体转染的细胞中表达的常规调控元件。因此,抗原盒还可含有经选择的启动子,所述启动子与一种或多种抗原连接并且与其他任选的调控元件一起位于重组载体的经选择的病毒序列内。盒可包括一种或多种抗原,例如一种或多种病原体衍生肽、病毒衍生肽、

细菌衍生肽、真菌衍生肽、寄生虫衍生肽和/或肿瘤衍生肽。盒可具有一种或多种抗原编码核酸序列，例如含有多个抗原编码核酸序列的盒，每个抗原编码核酸序列独立地可操作地连接到单独的启动子和/或使用其他多顺反子系统例如2A核糖体跳跃序列元件(例如，E2A、P2A、F2A或T2A序列)或内部核糖体进入位点(IRES)序列元件连接在一起。接头也可具有切割位点，例如TEV或弗林蛋白酶切割位点。具有切割位点的接头可与其他元件组合使用，例如多顺反子系统中的那些。在一个非限制性说明性实例中，弗林蛋白酶切割位点可与2A核糖体跳跃序列元件结合使用，使得弗林蛋白酶切割位点被配置为促进翻译后2A序列的去除。在含有超过一个抗原编码核酸序列的盒中，每个抗原编码核酸序列可含有一种或多种表位编码核酸序列(例如，编码串接的T细胞表位的抗原编码核酸序列)。

[0315] 有用的启动子可为组成型启动子或经调控(诱导型)启动子，其将能够控制有待表达的抗原的量。例如，合乎需要的启动子是巨细胞病毒立即早期启动子/增强子的启动子[参见例如Boshart等人，Cell，41:521-530(1985)]。另一种合乎需要的启动子包括劳斯肉瘤(Rous sarcoma)病毒LTR启动子/增强子。另一种启动子/增强子序列是鸡细胞质β-肌动蛋白启动子[T.A.Kost等人，Nucl.Acids Res.，11(23):8287(1983)]。其他适合或合乎需要的启动子可由本领域技术人员选择。

[0316] 本文还公开了一种病毒载体，其包含具有至少一个有效负载序列的盒，所述有效负载序列可操作地连接到作为TET启动子系统如TET-On系统或TET-Off系统的可调控启动子。不希望受理论束缚，TET启动子系统可用于在病毒生产期间最小化盒中编码的有效负载核酸(例如疫苗盒中编码的抗原)的转录。TET启动子系统在国际专利申请公开WO2020/243719中详细描述，所述文献出于所有目的通过引用并入本文。

[0317] TET启动子系统可包括四环素(TET)阻遏蛋白(TETr)控制的启动子。因此，本文还公开了一种病毒载体，其包含具有至少一个有效负载序列的盒，所述有效负载序列可操作地连接到四环素(TET)阻遏蛋白(TETr)控制的启动子。TETr控制的启动子可包括19bp TET操纵子(TETo)序列TCCCTATCAGTGATAGAGA(SEQ ID NO:83)。TETr控制的启动子可包括2、3、4、5、6、7、8、9或10个或更多个TETo核酸序列。在TETr控制的启动子中有2个或更多个TETo核酸序列，所述TETo序列可连接在一起。在TETr控制的启动子中有2个或更多个TETo核酸序列，所述TETo序列可直接连接在一起。在TETr控制的启动子中有2个或更多个TETo核酸序列，所述TETo序列可用接头序列连接在一起，例如具有1、2、3、4、5、6、7、8、9、10、11、12、13、14、15、16、17、18、19或20个或更多个核苷酸的接头序列。一般而言，TETr控制的启动子可使用任何所需的启动子序列，例如SV40、EF-1、RSV、PGK、HSA、MCK或EBV启动子序列。TETr控制的启动子可使用CMV启动子序列。TETr控制的启动子可使用最小的CMV启动子序列。TETo序列可位于RNA聚合酶结合的启动子序列区域的上游(5')。在一个说明性实例中，7个TETo序列位于启动子序列的上游(5')。可操作地连接到至少一个有效负载核酸序列的TETr控制的启动子(在启动子序列区域的上游具有TETo序列)可具有从5'至3'以下式描述的有序序列：

[0318] $(T\text{-}L_Y)_X\text{-}P\text{-}N$

[0319] 其中N是有效负载核酸序列，P是可操作地连接到有效负载核酸序列的启动子序列的RNA聚合酶结合序列，T是包含SEQ ID NO:66所示核苷酸序列的TETo核酸序列，L是接头序列，其中对于每个X，Y＝0或1，并且其中X＝1、2、3、4、5、6、7、8、9、10、11、12、13、14、15、16、17、18、19或20。在一个说明性实例中，X＝7并且对于每个X的Y＝1描述了其中7个TETo序列

位于启动子序列的上游(5')并且每个TETo序列由接头隔开。

[0320] TETo序列可位于RNA聚合酶结合的启动子序列区域的下游(3')。在另一个说明性实例中,2个TETo序列位于启动子序列的下游(3')。可操作地连接到至少一个有效负载核酸序列的TETr控制的启动子(在启动子序列区域下游具有TETo序列)可具有从5'至3'以下式描述的有序序列:

[0321] $P-(T-L_Y)_X-N$

[0322] 其中N是有效负载核酸序列,P是可操作地连接到有效负载核酸序列的启动子序列的RNA聚合酶结合序列,T是包含SEQ ID NO:66所示核苷酸序列的TETo核酸序列,L是接头序列,其中对于每个X,Y=0或1,并且其中X=1、2、3、4、5、6、7、8、9、10、11、12、13、14、15、16、17、18、19或20。在一个说明性实例中,X=2并且对于每个X的Y=1描述了其中2个TETo序列位于启动子序列的下游(3')并且每个TETo序列由接头隔开。

[0323] 具有TETr控制的启动子的载体的病毒生产可使用任何经工程化以表达TETr序列(tTS)的病毒生产细胞系,例如293细胞系或其经工程化以表达tTS的衍生物(例如,293F细胞系)。在tTS表达细胞中具有TETr控制的启动子的载体的病毒生产可改进病毒生产。在表达tTS的细胞中具有TETr控制的启动子的载体的病毒生产可改进病毒感染性,定义为每个感染单位(IU)的病毒颗粒(VP)。在表达tTS的细胞中具有TETr控制的启动子的载体的病毒生产可相对于非tTS表达细胞中的生产将病毒生产和/或病毒感染性提高至少1.5倍、至少2倍、至少2.5倍、至少3倍、至少3.5倍、至少4倍、至少4.5倍、至少5倍、至少6倍、至少7倍、至少8倍、至少9倍或至少10倍。在tTS表达细胞中具有TETr控制的启动子的载体的病毒生产可相对于非tTS表达细胞中的生产将病毒生产和/或病毒感染性提高至少15倍、至少20倍、至少25倍、至少30倍、至少35倍、至少40倍、至少45倍、至少50倍、至少60倍、至少70倍、至少80倍、至少90倍或至少100倍。在表达tTS的细胞中具有TETr控制的启动子的载体的病毒生产可相对于不具有TETr控制的启动子的载体的生产将病毒生产和/或病毒感染性提高至少1.5倍、至少2倍、至少2.5倍、至少3倍、至少3.5倍、至少4倍、至少4.5倍、至少5倍、至少6倍、至少7倍、至少8倍、至少9倍或至少10倍。在tTS表达细胞中具有TETr控制的启动子的载体的病毒生产可相对于不具有TETr控制启动子的载体的生产将病毒生产和/或病毒感染性提高至少15倍、至少20倍、至少25倍、至少30倍、至少35倍、至少40倍、至少45倍、至少50倍、至少60倍、至少70倍、至少80倍、至少90倍或至少100倍。

[0324] 抗原盒还可包括对病毒载体序列异源的核酸序列,包括提供转录物的有效聚腺苷酸化(聚(A)、聚-A或pA)的信号的序列和具有功能性剪接供体和受体位点的内含子。本发明的示例性载体中采用的普通聚-A序列是源自乳多泡病毒SV-40的聚-A序列。聚-A序列一般可在基于抗原的序列之后和在病毒载体序列之前插入盒中。普通内含子序列也可源自SV-40,并且被称为SV-40T内含子序列。抗原盒还可含有此类内含子,位于启动子/增强子序列与抗原之间。这些和其他普通载体元件的选择是常规的[参见例如Sambrook等人,"Molecular Cloning.A Laboratory Manual.",第2版,Cold Spring Harbor Laboratory,New York(1989)和其中引用的参考文献]并且许多此类序列可从商业和工业来源以及Genbank获得。

[0325] 抗原盒可具有一种或多种抗原。例如,给定盒可包括1-10、1-20、1-30、10-20、15-25、15-20、1、2、3、4、5、6、7、8、9、10、11、12、13、14、15、16、17、18、19、20或更多种抗原。抗原

可直接彼此连接。抗原也可利用接头彼此连接。抗原相对于彼此可处于任何方向，包括N至C或C至N。

[0326] 如本文别处所述，抗原盒可定位于病毒载体中的任何经选择缺失的位点中，例如VEE骨架的缺失结构蛋白或基于ChAd的载体的E1基因区缺失或E3基因区缺失的位点，以及可选择的其他位点。

[0327] 抗原盒可使用下式描述以描述每个元件的有序序列，从5'至3'：

[0328] $(P_a\text{-}(L5_b\text{-}N_c\text{-}L3_d)_X)_Z\text{-}(P2_h\text{-}(G5_e\text{-}U_f)_Y)_W\text{-}G3_g$

[0329] 其中P和P2包含启动子核苷酸序列，N包含MHC I类表位编码核酸序列，L5包含5'接头序列，L3包含3'接头序列，G5包含编码氨基酸接头的核酸序列，G3包含编码氨基酸接头的至少一个核酸序列中的一者，U包含MHC II类抗原编码核酸序列，其中对于每个X，相应Nc是表位编码核酸序列，其中对于每个Y，相应Uf是MHC II类表位编码核酸序列(例如，通用MHC II类表位编码核酸序列)。通用序列可包含破伤风类毒素和PADRE中的至少一者。通用序列可包含破伤风类毒素肽。通用序列可包含PADRE肽。通用序列可包含破伤风类毒素和PADRE肽。所述组合物和有序序列可进一步通过选择存在元件的数目来定义，例如其中a＝0或1，其中b＝0或1，其中c＝1，其中d＝0或1，其中e＝0或1，其中f＝1，其中g＝0或1，其中h＝0或1，X＝1至400，Y＝0、1、2、3、4或5，Z＝1至400，并且W＝0、1、2、3、4或5。

[0330] 在一个实例中，存在的元件包括其中a＝0，b＝1，d＝1，e＝1，g＝1，h＝0，X＝10，Y＝2，Z＝1，和W＝1，其描述其中无额外启动子存在(例如，仅存在由载体骨架如RNA甲病毒骨架提供的启动子核苷酸序列)，存在10个MHC I类表位，每个N存在5'接头，每个N存在3'接头，存在2个MHC II类表位，存在连接两个MHC II类表位的接头，存在将两个MHC II类表位的5'端连接到最终MHC I类表位的3'接头的接头，并且存在将两个MHC II类表位的3'端连接到载体骨架(例如，RNA甲病毒骨架)的接头。将抗原盒的3'端连接到载体骨架(例如，RNA甲病毒骨架)的实例包括直接连接到由载体骨架提供的3'UTR元件(例如3'19-nt CSE)。将抗原盒的5'端连接到载体骨架(例如，RNA甲病毒骨架)的实例包括直接连接到载体骨架的启动子或5'UTR元件，例如亚基因组启动子序列(例如，26S亚基因组启动子序列)、甲病毒5'UTR、51-nt CSE或24-nt CSE。

[0331] 其他实例包括：其中a＝1，描述其中存在除载体骨架(例如，RNA甲病毒骨架)所提供的启动子核苷酸序列以外的启动子；其中a＝1并且Z大于1，其中存在除载体骨架所提供的启动子核苷酸序列以外的多个启动子，其各自驱动1个或更多个不同MHC I类表位编码核酸序列的表达；其中h＝1，描述其中存在单独启动子以驱动MHC II类表位编码核酸序列的表达；以及其中g＝0，描述MHC II类表位编码核酸序列(如果存在)直接连接到载体骨架(例如，RNA甲病毒骨架)。

[0332] 其他实例包括其中所存在的每个MHC I类表位可具有5'接头、3'接头，两者都不具有，或具有两者。在其中同一抗原盒中存在超过一种MHC I类表位的实例中，一些MHC I类表位可具有5'接头和3'接头两者，而其他MHC I类表位可具有5'接头、3'接头或两者均不具有。在其中同一抗原盒中存在超过一种MHC I类表位的其他实例中，一些MHC I类表位可具有5'接头或3'接头，而其他MHC I类表位可具有5'接头、3'接头或两者均不具有。

[0333] 在其中同一抗原盒中存在超过一种MHC II类表位的实例中，一些MHC II类表位可具有5'接头和3'接头两者，而其他MHC II类表位可具有5'接头、3'接头或两者均不具有。在

其中同一抗原盒中存在超过一种MHC II类表位的其他实例中，一些MHC II类表位可具有5'接头或3'接头，而其他MHC II类表位可具有5'接头、3'接头或两者均不具有。

[0334] 其他实例包括其中所存在的每个抗原可具有5'接头、3'接头，两者都不具有，或具有两者。在其中同一抗原盒中存在超过一种抗原的实例中，一些抗原可具有5'接头和3'接头两者，而其他抗原可具有5'接头、3'接头或两者均不具有。在其中同一抗原盒中存在超过一种抗原的其他实例中，一些抗原可具有5'接头或3'接头，而其他抗原可具有5'接头、3'接头或两者均不具有。

[0335] 启动子核苷酸序列P和/或P2可与载体骨架如RNA甲病毒骨架所提供的启动子核苷酸序列相同。例如，由载体骨架提供的启动子序列Pn和P2可各自包含亚基因组启动子序列(例如，26S亚基因组启动子序列)或CMV启动子。启动子核苷酸序列P和/或P2可不同于载体骨架(例如，RNA甲病毒骨架)所提供的启动子核苷酸序列，以及可彼此不同。

[0336] 5'接头L5可为原生序列或非天然序列。非天然序列包括但不限于AAY、RR和DPP。3'接头L3也可为原生序列或非天然序列。另外，L5和L3可两者均为原生序列，两者均为非天然序列，或者一者可为原生的并且另一者可为非天然的。对于每个X，氨基酸接头的长度可为2、3、4、5、6、7、8、9、10、11、12、13、14、15、16、17、18、19、20、21、22、23、24、25、26、27、28、29、30、31、32、33、34、35、36、37、38、39、40、41、42、43、44、45、46、47、48、49、50、51、52、53、54、55、56、57、58、59、60、61、62、63、64、65、66、67、68、69、70、71、72、73、74、75、76、77、78、79、80、81、82、83、84、85、86、87、88、89、90、91、92、93、94、95、96、97、98、99、100或更多个氨基酸。对于每个X，氨基酸接头的长度也可为至少3、至少4、至少5、至少6、至少7、至少8、至少9、至少10、至少11、至少12、至少13、至少14、至少15、至少16、至少17、至少18、至少19、至少20、至少21、至少22、至少23、至少24、至少25、至少26、至少27、至少28、至少29或至少30个氨基酸。

[0337] 对于每个Y，氨基酸接头G5的长度可为2、3、4、5、6、7、8、9、10、11、12、13、14、15、16、17、18、19、20、21、22、23、24、25、26、27、28、29、30、31、32、33、34、35、36、37、38、39、40、41、42、43、44、45、46、47、48、49、50、51、52、53、54、55、56、57、58、59、60、61、62、63、64、65、66、67、68、69、70、71、72、73、74、75、76、77、78、79、80、81、82、83、84、85、86、87、88、89、90、91、92、93、94、95、96、97、98、99、100或更多个氨基酸。对于每个Y，氨基酸接头的长度也可为至少3、至少4、至少5、至少6、至少7、至少8、至少9、至少10、至少11、至少12、至少13、至少14、至少15、至少16、至少17、至少18、至少19、至少20、至少21、至少22、至少23、至少24、至少25、至少26、至少27、至少28、至少29或至少30个氨基酸。

[0338] 氨基酸接头G3的长度可为2、3、4、5、6、7、8、9、10、11、12、13、14、15、16、17、18、19、20、21、22、23、24、25、26、27、28、29、30、31、32、33、34、35、36、37、38、39、40、41、42、43、44、45、46、47、48、49、50、51、52、53、54、55、56、57、58、59、60、61、62、63、64、65、66、67、68、69、70、71、72、73、74、75、76、77、78、79、80、81、82、83、84、85、86、87、88、89、90、91、92、93、94、95、96、97、98、99、100或更多个氨基酸。G3的长度也可为至少3、至少4、至少5、至少6、至少7、至少8、至少9、至少10、至少11、至少12、至少13、至少14、至少15、至少16、至少17、至少18、至少19、至少20、至少21、至少22、至少23、至少24、至少25、至少26、至少27、至少28、至少29或至少30个氨基酸。

[0339] 对于每个X，每个N可编码MHC I类表位、MHC II类表位、能够刺激B细胞应答的表

位/抗原,或其组合。对于每个X,每个N可编码MHC I类表位、MHC II类表位和能够刺激B细胞应答的表位/抗原的组合。对于每个X,每个N可编码MHC I类表位和MHC II类表位的组合。对于每个X,每个N可编码MHC I类表位和能够刺激B细胞应答的表位/抗原的组合。对于每个X,每个N可编码MHC II类表位和能够刺激B细胞应答的表位/抗原的组合。对于每个X,每个N可编码MHC II类表位。对于每个X,每个N可编码能够刺激B细胞应答的表位/抗原。对于每个X,每个N可编码长度为7-15个氨基酸的MHC I类表位。对于每个X,每个N也可编码长度为5、6、7、8、9、10、11、12、13、14、15、16、17、18、19、20、21、22、23、24、25、26、27、28、29或30个氨基酸的MHC I类表位。对于每个X,每个N也可编码长度为至少5、至少6、至少7、至少8、至少9、至少10、至少11、至少12、至少13、至少14、至少15、至少16、至少17、至少18、至少19、至少20、至少21、至少22、至少23、至少24、至少25、至少26、至少27、至少28、至少29或至少30个氨基酸的MHC I类表位。

[0340] 编码一种或多种抗原的盒可以是700个核苷酸或更短。编码一种或多种抗原的盒可以是700个核苷酸或更短并编码2个不同的表位编码核酸序列(例如,编码2个不同的感染性疾病或肿瘤衍生的编码免疫原性多肽的核酸序列)。编码一种或多种抗原的盒可以是700个核苷酸或更短并编码至少2个不同的表位编码核酸序列。编码一种或多种抗原的盒可以是700个核苷酸或更短并编码3个不同的表位编码核酸序列。编码一种或多种抗原的盒可以是700个核苷酸或更短并编码至少3个不同的表位编码核酸序列。编码一种或多种抗原的盒可以是700个核苷酸或更短并且包括1-10、1-5、1、2、3、4、5、6、7、8、9、10或更多种抗原。

[0341] 编码一种或多种抗原的盒的长度可在375-700个核苷酸之间。编码一种或多种抗原的盒的长度可在375-700个核苷酸之间并编码2个不同的表位编码核酸序列(例如,编码2个不同的编码免疫原性多肽的感染性疾病或肿瘤源核酸序列)。编码一种或多种抗原的盒的长度可在375-700个核苷酸之间并编码至少2个不同的表位编码核酸序列。编码一种或多种抗原的盒的长度可在375-700个核苷酸之间并编码3个不同的表位编码核酸序列。编码一种或多种抗原的盒的长度在375-700个核苷酸之间并编码至少3个不同的表位编码核酸序列。编码一种或多种抗原的盒的长度可在375-700个核苷酸之间并且包括1-10、1-5、1、2、3、4、5、6、7、8、9、10或更多种抗原。

[0342] 编码一种或多种抗原的盒的长度可以是600、500、400、300、200或100个核苷酸或更短。编码一种或多种抗原的盒的长度可以是600、500、400、300、200或100个核苷酸或更短并编码2个不同的表位编码核酸序列。编码一种或多种抗原的盒的长度可以是600、500、400、300、200或100个核苷酸或更短并编码至少2个不同的表位编码核酸序列。编码一种或多种抗原的盒的长度可以是600、500、400、300、200或100个核苷酸或更短并编码3个不同的表位编码核酸序列。编码一种或多种抗原的盒的长度可以是600、500、400、300、200或100个核苷酸或更短并编码至少3个不同的表位编码核酸序列。编码一种或多种抗原的盒的长度可以是600、500、400、300、200或100个核苷酸或更短并且包括1-10、1-5、1、2、3、4、5、6、7、8、9、10或更多种抗原。

[0343] 编码一种或多种抗原的盒的长度可为375-600、375-500或375-400个核苷酸。编码一种或多种抗原的盒的长度可为375-600、375-500或375-400个核苷酸并编码2个不同的表位编码核酸序列。编码一种或多种抗原的盒的长度可为375-600、375-500或375-400个核苷酸并编码至少2个不同的表位编码核酸序列。编码一种或多种抗原的盒的长度可为375-

600、375-500或375-400个核苷酸并编码3个不同的表位编码核酸序列。编码一种或多种抗原的盒的长度可为375-600、375-500或375-400个核苷酸并编码至少3个不同的表位编码核酸序列。编码一种或多种抗原的盒的长度可为375-600、375-500或375-400个核苷酸并且包括1-10、1-5、1、2、3、4、5、6、7、8、9、10或更多种抗原。

[0344] 在一些情况下，编码额外抗原和/或表位的盒中的抗原或表位相对于其他编码的抗原或表位可以是免疫显性表位。一般而言，免疫显性是免疫应答仅针对一种或几种特定免疫原性肽的倾斜。免疫显性可作为免疫监测方案的一部分进行评估。例如，可通过评价T细胞和/或B细胞对编码抗原的应答来评估免疫显性。

[0345] 免疫显性可评估为免疫显性抗原的存在对一种或多种其他抗原的免疫应答的影响。例如，相对于不存在免疫显性抗原时的免疫应答，免疫显性抗原及其各自的免疫应答(例如免疫显性MHC I类表位)可降低另一种抗原的免疫应答。这种降低可使得在免疫显性抗原存在下的免疫应答不被认为是治疗有效应答。例如，如果与不存在免疫显性MHC I类表位时引起的应答相比，T细胞对其他抗原的应答不再被认为是治疗有效应答，则MHC I类表位通常被认为是免疫显性的。相对于不存在免疫显性抗原下的应答，免疫应答也可降低到检测限以下或接近检测限。例如，如果与不存在免疫显性MHC I类表位时引起的应答相比，T细胞对其他抗原的应答处于或低于检测限，则MHC I类表位通常被认为是免疫显性的。一般而言，免疫显性的评估是在两种都能够刺激免疫应答的抗原之间，例如，在施用给分别具有已知或预测呈递每个表位的同源MHC等位基因的受试者的疫苗组合物中的两个T细胞表位之间。免疫显性可通过评价在存在和不存在疑似免疫显性抗原的情况下对其他抗原的相对免疫应答来评估。

[0346] 免疫显性可评估为两种或更多种抗原之间免疫应答的相对差异。免疫显性可以指特定抗原相对于同一盒中编码的另一种抗原的5倍、10倍、20倍、30倍、40倍或50倍的免疫应答。免疫显性可以指特定抗原相对于同一盒中编码的另一种抗原的100倍、200倍、300倍、400倍或500倍的免疫应答。免疫显性可以指特定抗原相对于同一盒中编码的另一种抗原的1000倍、2000倍、3000倍、4000倍或5000倍免疫应答。免疫显性可以指特定抗原相对于同一盒中编码的另一种抗原的10,000倍免疫应答。

[0347] 在一些情况下，可能需要避免含有免疫显性表位的疫苗组合物。例如，可能希望避免设计编码免疫显性表位的疫苗盒。不希望受理论束缚，将免疫显性表位与额外表位一起施用和/或编码可降低对额外表位的免疫应答，包括可能最终降低针对额外表位的疫苗功效。作为一个说明性非限制性实例，包含TP53相关新表位的疫苗组合物可具有偏向TP53相关新表位的免疫应答，例如T细胞应答，其负面影响(例如，将免疫应答降低到免疫应答并非治疗有效应答之处和/或低于检测限)对疫苗组合物中其他抗原或表位(例如，疫苗组合物中的一个或多个KRAS相关新表位，例如以SEQ ID NO.75-82所示的KRAS相关新表位中的任一者)的免疫应答。因此，疫苗组合物可设计成不含免疫显性表位，例如设计疫苗盒(例如，(新)抗原编码盒)不编码免疫显性表位。例如，当以疫苗组合物的形式施用于受试者时，相对于当在不存在免疫显性MHC I类表位的情况下施用其他表位时的免疫应答，所述盒不编码降低对盒中编码的另一个表位的免疫应答的表位。在另一个实例中，当以疫苗组合物的形式施用于受试者时，相对于当在不存在免疫显性MHC I类表位的情况下施用其他表位时的免疫应答，所述盒不编码将对盒中编码的另一个表位的免疫应答降低到检测限以下的表

位。在另一个实例中，当以疫苗组合物的形式施用于受试者时，相对于当在不存在免疫显性MHC I类表位的情况下施用其他表位时的免疫应答，所述盒不编码降低对盒中编码的另一个表位的免疫应答的表位，其中所述免疫应答不是治疗有效应答。在另一个实例中，所述盒不编码这样的表位，所述表位相对于施用给受试者的疫苗组合物中同一盒中编码的另一个表位刺激5倍、10倍、20倍、30倍、40倍或50倍或更大的免疫应答，其中每种抗原能够刺激受试者中的免疫应答。在另一个实例中，所述盒不编码这样的表位，所述表位相对于施用给受试者的疫苗组合物中同一盒中编码的另一个表位刺激100倍、200倍、300倍、400倍或500倍或更大的免疫应答，其中每种抗原能够刺激受试者中的免疫应答。在另一个实例中，所述盒不编码这样的表位，所述表位相对于施用给受试者的疫苗组合物中同一盒中编码的另一个表位刺激1000倍、2000倍、3000倍、4000倍或5000倍或更大的免疫应答，其中每种抗原能够刺激受试者中的免疫应答。在另一个实例中，所述盒不编码这样的表位，所述表位相对于施用给受试者的疫苗组合物中同一盒中编码的另一个表位导致10,000倍或更大的免疫应答，其中每种抗原能够刺激受试者中的免疫应答。

[0348] V.B.免疫调节剂

[0349] 本文所述的载体，例如本文所述的C68载体或本文所述的甲病毒载体，可包含编码至少一种抗原的核酸，并且相同或单独的载体可包含编码至少一种免疫调节剂的核酸。免疫调节剂可包括结合于并阻断免疫检查点分子的活性的结合分子(例如，抗体如scFv)。免疫调节剂可包括细胞因子，例如IL-2、IL-7、IL-12(包括IL-12p35、p40、p70和/或p70-融合构建体)、IL-15或IL-21。免疫调节剂可包括修饰的细胞因子(例如，pegIL-2)。载体可包含抗原盒和一种或多种编码免疫调节剂的核酸分子。

[0350] 可靶向用于阻断或抑制的说明性免疫检查点分子包括但不限于CTLA-4、4-1BB(CD137)、4-1BBL(CD137L)、PDL1、PDL2、PD1、B7-H3、B7-H4、BTLA、HVEM、TIM3、GAL9、LAG3、TIM3、B7H3、B7H4、VISTA、KIR、2B4(属于CD2分子家族并且在所有NK、γδ和记忆CD8+(αβ)T细胞上表达)、CD160(也称为BY55)和CGEN-15049。免疫检查点抑制剂包括抗体或其抗原结合片段或其他结合蛋白，其结合于并阻断或抑制CTLA-4、PDL1、PDL2、PD1、B7-H3、B7-H4、BTLA、HVEM、TIM3、GAL9、LAG3、TIM3、B7H3、B7H4、VISTA、KIR、2B4、CD160和CGEN-15049中的一者或多者的活性。说明性免疫检查点抑制剂包括曲美木单抗(CTLA-4阻断抗体)、抗OX40、PD-L1单克隆抗体(抗B7-H1；MEDI4736)、伊匹单抗、MK-3475(PD-1阻断剂)、纳武单抗(Nivolumamb)(抗PD1抗体)、CT-011(抗PD1抗体)、BY55单克隆抗体、AMP224(抗PDL1抗体)、BMS-936559(抗PDL1抗体)、MPLDL3280A(抗PDL1抗体)、MSB0010718C(抗PDL1抗体)和Yervoy/伊匹单抗(抗CTLA-4检查点抑制剂)。可使用本领域普通技术将抗体编码序列工程化到例如C68的载体中。一种示例性方法描述于Fang等人，Stable antibody expression at therapeutic levels using the 2A peptide.Nat Biotechnol.2005年5月;23(5):584-90.电子版2005年4月17日;出于所有目的通过引用并入本文。

[0351] V.C.疫苗设计和制造的额外考虑因素

[0352] V.C.1.确定一组涵盖所有肿瘤亚克隆的肽

[0353] 躯干肽，意指由所有或大多数肿瘤亚克隆呈递的肽，可优先排序以纳入疫苗中。任选地，如果没有预测呈递且以高概率具有免疫原性的躯干肽，或者如果预测呈递且以高概率具有免疫原性的躯干肽数目足够小，使得可在疫苗中包括额外的非躯干肽，则可通过估

计肿瘤亚克隆的数目和身份并选择肽来优先排序更多的肽,以使疫苗覆盖的肿瘤亚克隆数目最大化。

[0354] V.C.2.抗原优先排序

[0355] 在应用了所有上述抗原过滤器之后,可能仍有比疫苗技术支持的更多候选抗原可用于疫苗包涵。另外,可保留关于抗原分析的各个方面的不确定性,并且候选疫苗抗原的不同特性之间可存在折衷。因此,可考虑整合式多维模型代替选择过程的每个步骤中的预定过滤器,将候选抗原置于具有至少以下轴的空间中并使用整合方法优化选择。

[0356] 1.自身免疫或耐受性的风险(生殖系的风险)(自身免疫的风险较低通常是优选的)

[0357] 2.测序伪影的概率(伪影的概率较低通常是优选的)

[0358] 3.免疫原性的概率(免疫原性的概率较高通常是优选的)

[0359] 4.呈递的概率(呈递的概率较高通常是优选的)

[0360] 5.基因表达(较高表达通常是优选的)

[0361] 6.HLA基因的覆盖率(参与抗原集合呈递的HLA分子数目较大可降低肿瘤、感染性疾病和/或受感染细胞经由HLA分子的下调或突变逃避免疫攻击的概率)

[0362] 7.HLA类别的覆盖率(覆盖HLA-I和HLA-II两者可增加治疗应答的概率并降低肿瘤或感染性疾病逃避的概率)

[0363] 另外,任选地,如果预测抗原将由在患者的全部或部分肿瘤或受感染细胞中丢失或失活的HLA等位基因呈递,则可从疫苗接种去除所述抗原的优先排序(例如排除所述抗原)。HLA等位基因损失可通过体细胞突变、杂合性缺失或基因座的同型接合缺失发生。用于检测HLA等位基因体细胞突变的方法是本领域中所熟知,例如(Shukla等人,2015)。用于检测体细胞LOH和同型接合缺失(包括HLA基因座)的方法同样被充分描述。(Carter等人,2012;McGranahan等人,2017;Van Loo等人,2010)。如果质谱数据表明所预测的抗原未被预测的HLA等位基因呈递,则还可去除抗原的优先排序。

[0364] V.D.甲病毒

[0365] V.D.1甲病毒生物学

[0366] 甲病毒是披膜病毒科的成员,并且是正义单链RNA病毒。成员通常分类为旧世界,例如辛德毕斯、罗斯河、马雅罗、基孔肯尼亚和塞姆利基森林病毒,或新世界,例如东部马脑炎、奥拉、摩根堡、或委内瑞拉马脑炎及其衍生菌株TC-83(Strauss Microbial Review1994)。天然甲病毒基因组通常约12kb长,其中前三分之二含有编码非结构蛋白(nsP)的基因,所述非结构蛋白形成用于病毒基因组自我复制的RNA复制复合物,并且最后三分之一含有编码用于病毒粒子产生的结构蛋白的亚基因组表达盒(Frolov RNA 2001)。

[0367] 甲病毒的模型生命周期涉及数个不同步骤(Strauss Microbial Review 1994,Jose Future Microbiol 2009)。在病毒附着于宿主细胞后,病毒粒子与内饮区室内的膜融合,导致基因组RNA最终释放至细胞溶质中。以正链定向并包含5'甲基鸟苷酸帽和3'聚A尾部的基因组RNA被翻译以产生形成复制复合物的非结构蛋白nsP1-4。在感染早期,正链然后由复合物复制成负链模板。在当前模型中,复制复合物随着感染进展被进一步加工,使得所得经加工的复合物转换成将负链转录成全长正链基因组RNA以及含有结构基因的26S亚基因组正链RNA。甲病毒的数个保守序列元件(CSE)已鉴定为可能在各种RNA复制步骤中起作

用,包括:负链模板的正链RNA复制中的5'UTR的互补序列、基因组模板的负链合成复制中的51-nt CSE、负链的亚基因组RNA转录中的在nsP与26S RNA之间的接合区中的24-nt CSE、和正链模板的负链合成中的3'19-nt CSE。

[0368] 在各种RNA物种复制后,病毒颗粒然后通常在病毒的天然生命周期中组装。26S RNA被翻译并且所得蛋白质经进一步加工以产生结构蛋白,其包括衣壳蛋白、糖蛋白E1和E2以及两个小多肽E3和6K(Strauss 1994)。发生病毒RNA的衣壳化,衣壳蛋白通常仅特异性针对所包装的基因组RNA,然后病毒粒子组装并在膜表面出芽。

[0369] V.D.2.作为递送载体的甲病毒

[0370] 甲病毒(包括甲病毒序列、特征和其他元件)可用于产生基于甲病毒的递送载体(也称为甲病毒载体、甲病毒病毒载体、甲病毒疫苗载体、自我复制RNA(srRNA)载体或自扩增mRNA(SAM)载体)。甲病毒先前已被工程化以用作表达载体系统(Pushko 1997,Rheme 2004)。甲病毒提供数种优势,特别是在可能需要异源抗原表达的疫苗环境中。由于在宿主细胞溶质中自我复制的能力,因此甲病毒载体一般能够在细胞内产生高拷贝数的表达盒,从而导致高水平的异源抗原产生。另外,载体一般是瞬时的,从而使得生物安全性得以提高以及减少对载体的免疫耐受性的诱导。与其他标准病毒载体(例如人类腺病毒)相比,公众通常还缺乏对甲病毒载体预先存在的免疫性。基于甲病毒的载体也通常导致对受感染细胞的细胞毒性应答。在一定程度上,细胞毒性在疫苗环境中对于适当刺激对所表达的异源抗原的免疫应答可为重要的。然而,所需细胞毒性的程度可为平衡作用,并且因此已开发数种减毒甲病毒,包括VEE的TC-83菌株。因此,本文所述的抗原表达载体的实例可利用甲病毒骨架,其允许高水平的抗原表达、刺激对抗原的稳固免疫应答、不刺激对载体本身的免疫应答,并且可以安全方式使用。此外,抗原表达盒可经设计以经由优化载体使用的甲病毒序列(包括但不限于源自VEE或其减毒衍生物TC-83的序列)而刺激不同水平的免疫应答。

[0371] 已使用甲病毒序列对数种表达载体设计策略进行工程化(Pushko1997)。在一个策略中,甲病毒载体设计包括在结构蛋白基因下游插入26S启动子序列元件的第二拷贝,接着是异源基因(Frolov 1993)。因此,除天然非结构蛋白和结构蛋白之外,还产生表达异源蛋白的额外亚基因组RNA。在这种系统中,存在用于产生感染性病毒粒子的所有元件,并且因此可能发生在非感染细胞中表达载体的反复轮感染。

[0372] 另一种表达载体设计利用辅助病毒系统(Pushko 1997)。在这种策略中,结构蛋白由异源基因代替。因此,在由仍完整的非结构基因介导病毒RNA的自我复制之后,26S亚基因组RNA提供异源蛋白的表达。传统上,表达结构蛋白的额外载体然后例如通过细胞系的共转染以反式供应,以产生感染性病毒。系统详细描述于USPN 8,093,021中,其出于所有目的通过引用整体并入本文。辅助载体系统提供限制形成感染性颗粒的可能性的益处,因此提高生物安全性。另外,辅助载体系统减小总载体长度,潜在提高复制和表达效率。因此,本文所述的抗原表达载体的实例可利用结构蛋白由抗原盒代替的甲病毒骨架,所得载体降低生物安全问题,同时由于整体表达载体尺寸减小而促进有效表达。

[0373] V.D.3.体外甲病毒产生

[0374] 甲病毒递送载体通常是正义RNA多核苷酸。本领域中熟知的用于产生RNA的便利技术是体外转录IVT。在这种技术中,首先通过本领域技术人员熟知的技术产生所需载体的DNA模板,包括标准分子生物学技术,例如克隆、限制性消化、连接、基因合成(例如,化学和/

或酶促合成)和聚合酶链反应(PCR)。DNA模板在需要转录为RNA的序列的5'端含有一个RNA聚合酶启动子。启动子包括但不限于噬菌体聚合酶启动子,例如T3、T7或SP6。DNA模板然后与适当的RNA聚合酶、缓冲剂和核苷酸(NTP)一起孵育。所得的RNA多核苷酸可任选地进一步经修饰,包括但不限于添加5'帽结构,例如7-甲基鸟苷或相关结构,以及任选地修饰3'端以包括聚腺苷酸(聚A)尾。RNA然后可使用本领域熟知的技术如苯酚-氯仿提取或柱纯化(例如基于色谱法的纯化)进行纯化。

[0375] V.D.4.经由脂质纳米粒子递送

[0376] 在疫苗载体设计中考虑的一个重要方面是针对载体本身的免疫性(Riley 2017)。这可呈对载体本身(例如某些人类腺病毒系统)预先存在的免疫性形式,或呈在疫苗施用后对载体产生免疫性的形式。如果进行相同疫苗的多次施用(例如分开的初免和增强剂量),或者如果使用相同疫苗载体系统递送不同抗原盒,则后者是重要的考虑因素。

[0377] 在甲病毒载体的情况下,标准的传递方法是前面讨论的辅助病毒系统,其以反式提供衣壳、E1和E2蛋白以产生感染性病毒颗粒。然而,重要的是要注意E1和E2蛋白通常是中和抗体的主要靶标(Strauss1994)。因此,如果感染性颗粒被中和抗体靶向,则使用甲病毒载体将目标抗原递送至靶细胞的功效可能会降低。

[0378] 病毒颗粒介导的基因递送的替代方案是使用纳米材料递送表达载体(Riley 2017)。重要的是,纳米材料媒介物可由非免疫原性材料制成并且一般避免引发对递送载体本身的免疫性。这些材料可包括(但不限于)脂质、无机纳米材料和其他聚合材料。脂质可为阳离子、阴离子或中性的。材料可为合成或天然来源的,并且在一些情况下是可生物降解的。脂质可包括脂肪、胆固醇、磷脂、脂质缀合物,包括(但不限于)聚乙二醇(PEG)缀合物(聚乙二醇化脂质)、蜡、油、甘油酯和脂溶性维生素。

[0379] 脂质纳米粒子(LNP)是有吸引力的递送系统,因为脂质的两亲性使得能够形成膜和囊泡状结构(Riley 2017)。一般而言,这些囊泡通过吸收至靶细胞的膜中并且将核酸释放至细胞溶质中来递送表达载体。另外,LNP可经进一步修饰或官能化以有助于靶向特定细胞类型。LNP设计中的另一个考虑因素是靶向效率与细胞毒性之间的平衡。脂质组合物一般包括阳离子、中性、阴离子和两性脂质的限定混合物。在一些情况下,包括特定脂质以防止LNP聚集、防止脂质氧化、或提供有助于额外部分附着的功能性化学基团。脂质组合物可影响整体LNP大小和稳定性。在一个实例中,脂质组合物包含二亚油基甲基-4-二甲基氨基丁酸酯(MC3)和MC3样分子。MC3和MC3样脂质组合物可经配制以包括一种或多种其他脂质,例如PEG或PEG缀合的脂质、固醇或中性脂质。

[0380] 直接暴露于血清的核酸载体(例如表达载体)可具有数种不期望的结果,包括核酸由血清核酸酶降解或游离核酸对免疫系统的脱靶刺激。因此,包封甲病毒载体可用于避免降解,同时还避免潜在的脱靶影响。在某些实例中,甲病毒载体完全包封在递送媒介物内,例如在LNP的含水内部。甲病毒载体包封在LNP内可通过本领域技术人员熟知的技术来进行,例如在微流体液滴生成装置上进行的微流体混合和液滴生成。此类装置包括(但不限于)标准T形接头装置或流动聚焦装置。在一个实例中,所需脂质制剂(例如含有MC3或MC3样的组合物)与甲病毒递送载体和其他所需药剂并行提供至液滴生成装置,使得递送载体和所需药剂完全包封在基于MC3或MC3样的LNP内部。在一个实例中,液滴生成装置可控制所产生的LNP的尺寸范围和尺寸分布。例如,LNP的尺寸可在1至1000纳米直径范围内,例如1、10、

50、100、500或1000纳米。在液滴生成后，包封表达载体的递送媒介物可经进一步处理或修饰以使其准备用于施用。

[0381] V.E.黑猩猩腺病毒(ChAd)

[0382] V.E.1.用黑猩猩腺病毒进行的病毒递送

[0383] 用于递送一种或多种抗原的疫苗组合物(例如，经由抗原盒)可通过提供黑猩猩来源的腺病毒核苷酸序列、多种新型载体和表达黑猩猩腺病毒基因的细胞系来产生。黑猩猩C68腺病毒(在本文中也称为ChAdV68)的核苷酸序列可用于抗原递送的疫苗组合物中(参见SEQ ID NO:1)。源自C68腺病毒的载体的使用进一步详细描述于USPN6,083,716中，其出于所有目的通过引用整体并入本文。基于ChAdV68的载体和递送系统详细描述于美国申请公开第US20200197500A1号和国际专利申请公开WO2020243719A1，其各自出于所有目的通过引用并入本文。

[0384] 在另一方面，本文提供了包含黑猩猩腺病毒的DNA序列的重组腺病毒(例如C68)和以操作方式连接到引导其表达的调控序列的抗原盒。重组病毒能够感染哺乳动物细胞、优选人类细胞，并且能够在细胞中表达抗原盒产物。在这种载体中，原生黑猩猩E1基因和/或E3基因和/或E4基因可缺失。抗原盒可插入至这些基因缺失位点中的任一者中。抗原盒可包括抗原，针对其引发的免疫应答是所需的。

[0385] 在另一方面，本文提供了一种经黑猩猩腺病毒(例如C68)感染的哺乳动物细胞。

[0386] 在另一方面，提供了一种新型的哺乳动物细胞系，其表达黑猩猩腺病毒基因(例如来自C68)或其功能片段。

[0387] 在另一方面，本文提供了一种用于将抗原盒递送至哺乳动物细胞中的方法，其包括以下步骤：向细胞中引入有效量的已被工程化以表达抗原盒的黑猩猩腺病毒，例如C68。

[0388] 另一方面提供了一种用于在哺乳动物宿主中刺激免疫应答以治疗癌症的方法。所述方法可包括向宿主施用有效量的重组黑猩猩腺病毒，例如C68的步骤，所述重组黑猩猩腺病毒包含抗原盒，所述抗原盒编码免疫应答所针对的肿瘤的一种或多种抗原。

[0389] 另一方面提供了一种用于在哺乳动物宿主中刺激免疫应答以治疗或预防受试者疾病例如感染性疾病的方法。所述方法可包括向宿主施用有效量的重组黑猩猩腺病毒例如C68的步骤，所述重组黑猩猩腺病毒包含抗原盒，所述抗原盒编码例如来自免疫应答所针对的感染性疾病的一种或多种抗原。

[0390] 还公开了一种非猿猴哺乳动物细胞，其表达获自SEQ ID NO:1序列的黑猩猩腺病毒基因。所述基因可选自由以下组成的组：SEQ ID NO:1的腺病毒E1A、E1B、E2A、E2B、E3、E4、L1、L2、L3、L4和L5。

[0391] 还公开了一种包含黑猩猩腺病毒DNA序列的核酸分子，所述黑猩猩腺病毒DNA序列包含获自SEQ ID NO:1序列的基因。所述基因可选自由以下组成的组：SEQ ID NO:1的所述黑猩猩腺病毒E1A、E1B、E2A、E2B、E3、E4、L1、L2、L3、L4和L5基因。在一些方面，核酸分子包含SEQ ID NO:1。在一些方面，核酸分子包含SEQ ID NO:1序列，缺少选自由以下组成的组的至少一个基因：SEQ ID NO:1的E1A、E1B、E2A、E2B、E3、E4、L1、L2、L3、L4和L5基因。

[0392] 还公开了一种载体，其包含获自SEQ ID NO:1的黑猩猩腺病毒DNA序列和可操作地连接到一个或多个调控序列的抗原盒，所述一个或多个调控序列引导所述盒在异源宿主细胞中的表达，任选地其中所述黑猩猩腺病毒DNA序列包含至少用于复制和病毒粒子衣壳化

所必需的顺式元件，所述顺式元件侧接抗原盒和调控序列。在一些方面，黑猩猩腺病毒DNA序列包含选自由以下组成的组的基因：SEQ ID NO:1的E1A、E1B、E2A、E2B、E3、E4、L1、L2、L3、L4和L5基因序列。在一些方面，载体可缺乏E1A和/或E1B基因。

[0393] 本文还公开了一种腺病毒载体，其包含：部分缺失的E4基因，其包含缺失或部分缺失的E4orf2区和缺失或部分缺失的E4orf3区，以及任选地缺失或部分缺失的E4orf4区。所述部分缺失的E4可包含SEQ ID NO:1所示序列的至少核苷酸34,916至35,642的E4缺失，并且其中所述载体包含以SEQ ID NO:1所示序列的至少核苷酸2至36,518。所述部分缺失的E4可包含SEQ ID NO:1所示序列的核苷酸34,916至34,942的至少部分缺失、SEQ ID NO:1所示序列的核苷酸34,952至35,305的至少部分缺失和SEQ ID NO:1所示序列的核苷酸35,302至35,642的至少部分缺失的E4缺失，并且其中所述载体包含以SEQ ID NO:1所示序列的至少核苷酸2至36,518。所述部分缺失的E4可包含SEQ ID NO:1所示序列的至少核苷酸34,980至36,516的E4缺失，并且其中所述载体包含以SEQ ID NO:1所示序列的至少核苷酸2至36,518。所述部分缺失的E4可包含SEQ ID NO:1所示序列的至少核苷酸34,979至35,642的E4缺失，并且其中所述载体包含SEQ ID NO:1所示序列的至少核苷酸2至36,518。所述部分缺失的E4可包含至少部分缺失E4Orf2、完全缺失E4Orf3和至少部分缺失E4Orf4的E4缺失。所述部分缺失的E4可包含至少部分缺失E4Orf2、至少部分缺失E4Orf3和至少部分缺失E4Orf4的E4缺失。所述部分缺失的E4可包含至少部分缺失E4Orf1、完全缺失E4Orf2和至少部分缺失E4Orf3的E4缺失。所述部分缺失的E4可包含至少部分缺失E4Orf2和至少部分缺失E4Orf3的E4缺失。所述部分缺失的E4可包含E4Orf1的起始位点与E4Orf5的起始位点之间的E4缺失。所述部分缺失的E4可以是与E4Orf1的起始位点相邻的E4缺失。所述部分缺失的E4可以是与E4Orf2的起始位点相邻的E4缺失。所述部分缺失的E4可以是与E4Orf3的起始位点相邻的E4缺失。所述部分缺失的E4可以是与E4Orf4的起始位点相邻的E4缺失。所述E4缺失可以是至少50、至少100、至少200、至少300、至少400、至少500、至少600、至少700、至少800、至少900、至少1000、至少1100、至少1200、至少1300、至少1400、至少1500、至少1600、至少1700、至少1800、至少1900或至少2000个核苷酸。所述E4缺失可以是至少700个核苷酸。所述E4缺失可以是至少1500个核苷酸。所述E4缺失可以是50个或更少、100个或更少、200个或更少、300个或更少、400个或更少、500个或更少、600个或更少、700个或更少、800个或更少、900个或更少、1000个或更少、1100个或更少、1200个或更少、1300个或更少、1400个或更少、1500个或更少、1600个或更少、1700个或更少、1800个或更少、1900个或更少、或2000或更少的核苷酸。所述E4缺失可以是750个核苷酸或更短。所述E4缺失可以是至少1550个核苷酸或更短。

[0394] 所述部分缺失的E4基因可以是SEQ ID NO:1所示的E4基因序列，其至少缺少SEQ ID NO:1所示序列的核苷酸34,916至35,642。所述部分缺失的E4基因可以是SEQ ID NO:1所示的E4基因序列，其缺少SEQ ID NO:1所示的E4基因序列，并且其至少缺少SEQ ID NO:1所示序列的核苷酸34,916至34,942、核苷酸34,952至35,305，以及SEQ ID NO:1所示序列的核苷酸35,302至35,642。所述部分缺失的E4基因可以是SEQ ID NO:1所示的E4基因序列并且其至少缺少SEQ ID NO:1所示序列的核苷酸34,980至36,516。所述部分缺失的E4基因可以是SEQ ID NO:1所示的E4基因序列并且其至少缺少SEQ ID NO:1所示序列的核苷酸34,979至35,642。具有部分缺失的E4基因的腺病毒载体可具有盒，其中所述盒包含至少一种有效负载核酸序列，并且其中所述盒包含与所述至少一种有效负载核酸序列可操作地连接的至

少一种启动子序列。具有部分缺失的E4基因的腺病毒载体可具有SEQ ID NO:1所示的ChAdV68序列的一种或多种基因或调控序列,任选地其中所述一种或多种基因或调控序列包含黑猩猩腺病毒反向末端重复序列(ITR)、SEQ ID NO:1所示序列的E1A、E1B、E2A、E2B、E3、E4、L1、L2、L3、L4和L5基因中的至少一者。具有部分缺失的E4基因的腺病毒载体可具有SEQ ID NO:1所示序列的核苷酸2至34,916,其中所述部分缺失的E4基因是核苷酸2至34,916的3',并且任选地所述核苷酸2至34,916另外缺少SEQ ID NO:1所示序列的核苷酸577至3403,对应于E1缺失,和/或缺少SEQ ID NO:1所示序列的核苷酸27,125至31,825,对应于E3缺失。具有部分缺失的E4基因的腺病毒载体可具有SEQ ID NO:1所示序列的核苷酸35,643至36,518,并且其中所述部分缺失的E4基因是核苷酸35,643至36,518的5'。具有部分缺失的E4基因的腺病毒载体可具有SEQ ID NO:1所示序列的核苷酸2至34,916,其中所述部分缺失的E4基因是核苷酸2至34,916的3',所述核苷酸2至34,916另外缺少SEQ ID NO:1所示序列的核苷酸577至3403,对应于E1缺失,并且缺少SEQ ID NO:1所示序列的核苷酸27,125至31,825,对应于E3缺失。具有部分缺失的E4基因的腺病毒载体可具有SEQ ID NO:1所示序列的核苷酸2至34,916,其中所述部分缺失的E4基因是核苷酸2至34,916的3',所述核苷酸2至34,916另外缺少SEQ ID NO:1所示序列的核苷酸577至3403,对应于E1缺失,并且缺少SEQ ID NO:1所示序列的核苷酸27,125至31,825,对应于E3缺失,并且具有SEQ ID NO:1所示序列的核苷酸35,643至36,518,并且其中所述部分缺失的E4基因是核苷酸35,643至36,518的5'。

[0395] 所述部分缺失的E4基因可以是SEQ ID NO:1所示的E4基因序列,其至少缺少SEQ ID NO:1所示序列的核苷酸34,916至35,642、SEQ ID NO:1所示序列的核苷酸2至34,916,其中所述部分缺失的E4基因是核苷酸2至34,916的3',所述核苷酸2至34,916另外缺少SEQ ID NO:1所示序列的核苷酸577至3403,对应于E1缺失,并且缺少SEQ ID NO:1所示序列的核苷酸27,125至31,825,对应于E3缺失,并且具有SEQ ID NO:1所示序列的核苷酸35,643至36,518,并且其中所述部分缺失的E4基因是核苷酸35,643至36,518的5'。

[0396] 本文还公开了一种用本文公开的载体转染的宿主细胞,所述载体例如被工程化以表达抗原盒的C68载体。本文还公开了一种经由将本文公开的载体引入细胞中而表达其中引入的所选择基因的人类细胞。

[0397] 本文还公开了一种用于将抗原盒递送至哺乳动物细胞的方法,所述方法包括向所述细胞中引入有效量的本文公开的载体,例如被工程化以表达抗原盒的C68载体。

[0398] 本文还公开了一种用于产生抗原的方法,所述方法包括将本文公开的载体引入哺乳动物细胞中,在合适的条件下培养细胞并产生抗原。

[0399] V.E.2.表达E1的互补细胞系

[0400] 为了产生缺失本文所述的基因中的任一者的重组黑猩猩腺病毒(Ad),缺失基因区的功能如果对于病毒的复制和感染性必不可少,则可通过辅助病毒或细胞系(即互补或包装细胞系)供应至重组病毒。例如,为了产生复制缺陷型黑猩猩腺病毒载体,可使用表达人类或黑猩猩腺病毒的E1基因产物的细胞系;这种细胞系可包括HEK293或其变体。可遵循产生表达黑猩猩E1基因产物的细胞系的方案(USPN6,083,716的实施例3和4),以产生表达任何所选择的黑猩猩腺病毒基因的细胞系。

[0401] AAV增强测定可用于鉴定表达黑猩猩腺病毒E1的细胞系。这种测定可用于鉴定通

过使用例如来自其他物种的其他未表征的腺病毒的E1基因制备的细胞系中的E1功能。所述测定描述于USPN6,083,716的实施例4B中。

[0402] 所选择的黑猩猩腺病毒基因(例如E1)可在启动子的转录控制下用于在所选择的亲本细胞系中表达。诱导型或组成型启动子可用于此目的。在诱导型启动子中包括可由锌诱导的绵羊金属硫蛋白启动子,或可由糖皮质激素、特别是地塞米松(dexamethasone)诱导的小鼠乳腺肿瘤病毒(MMTV)启动子。其他诱导型启动子,例如通过引用并入本文的国际专利申请WO95/13392中所鉴定的那些诱导型启动子,还可用于产生包装细胞系。还可采用控制黑猩猩腺病毒基因表达的组成型启动子。

[0403] 亲本细胞可经选择以产生表达任何所需C68基因的新型细胞系。此类亲本细胞系可为(但不限于)HeLa[ATCC登录号CCL 2]、A549[ATCC登录号CCL 185]、KB[CCL 17]、Detroit[例如Detroit 510、CCL 72]和WI-38[CCL 75]细胞。其他合适的亲本细胞系可获自其他来源。亲本细胞系可包括CHO、HEK293或其变体、911、HeLa、A549、LP-293、PER.C6或AE1-2a。

[0404] 表达E1的细胞系可用于产生重组黑猩猩腺病毒E1缺失的载体。使用基本上相同的程序构建的表达一种或多种其他黑猩猩腺病毒基因产物的细胞系用于产生缺失编码那些产物的基因的重组黑猩猩腺病毒载体。另外,表达其他人类Ad E1基因产物的细胞系还可用于产生黑猩猩重组Ad。

[0405] V.E.3.作为载体的重组病毒颗粒

[0406] 本文公开的组合物可包含病毒载体,其将至少一种抗原递送至细胞。此类载体包含黑猩猩腺病毒DNA序列(例如C68)和可操作地连接到引导盒表达的调控序列的抗原盒。C68载体能够在经感染的哺乳动物细胞中表达盒。C68载体可功能性缺失一个或多个病毒基因。抗原盒包含至少一个在一个或多个调控序列(例如启动子)控制下的抗原。任选的辅助病毒和/或包装细胞系可向黑猩猩病毒载体供应缺失的腺病毒基因的任何必需产物。

[0407] 术语“功能性缺失”意指去除或以其他方式改变(例如通过突变或修饰)足够量的基因区,使得基因区不再能够产生一种或多种基因表达的功能性产物。可导致功能性缺失的突变或修饰包括但不限于无义突变,例如引入提前终止密码子和去除典型和非典型的起始密码子,改变mRNA剪接或其他转录加工的突变,或其组合。如果需要,可去除整个基因区。

[0408] 形成本文公开的载体的核酸序列的修饰,包括序列缺失、插入和其他突变,可使用标准分子生物学技术产生并且在本发明的范围内。

[0409] V.E.4.病毒质粒载体的构建

[0410] 可用于本发明的黑猩猩腺病毒C68载体包括重组缺陷型腺病毒,即在E1a或E1b基因中功能性缺失并且任选地携带其他突变(例如其他基因中的温度敏感性突变或缺失)的黑猩猩腺病毒序列。预期这些黑猩猩序列也可用于形成来自其他腺病毒和/或腺相关病毒序列的杂交载体。由人类腺病毒制备的同源腺病毒载体描述于公开文献中[参见例如上文所引用的Kozarsky I和II,和其中引用的参考文献,美国专利第5,240,846号]。

[0411] 在构建可用于将抗原盒递送至人类(或其他哺乳动物)细胞的有用黑猩猩腺病毒C68载体时,可将一系列腺病毒核酸序列用于载体。包含最小黑猩猩C68腺病毒序列的载体可与辅助病毒结合使用以产生感染性重组病毒颗粒。辅助病毒提供最小黑猩猩腺病毒载体的病毒感染性和繁殖所需的基本基因产物。当在另外的功能性病毒载体中仅产生黑猩猩腺

病毒基因的一个或多个所选择的缺失时，可通过在所选择的包装细胞系中繁殖病毒而在病毒载体生产过程中供应缺失的基因产物，所述包装细胞系提供反式缺失的基因功能。

[0412] V.E.5.重组最小腺病毒

[0413] 最小的黑猩猩Ad C68病毒是仅含有复制和病毒粒子衣壳化所必需的腺病毒顺式元件的病毒颗粒。也就是说，载体含有腺病毒的顺式作用5'和3'反向末端重复序列(ITR)序列(其充当复制起点)和原生5'包装/增强子结构域(其含有用于包装线性Ad基因组所必需的序列和E1启动子的增强子元件)。参见例如在国际专利申请WO96/13597中所描述并通过引用并入本文的用于制备"最小"人类Ad载体的技术。

[0414] V.E.6.其他缺陷型腺病毒

[0415] 重组复制缺陷型腺病毒也可比最小黑猩猩腺病毒序列含有更多。这些其他Ad载体可通过病毒基因区的各个部分的缺失和通过任选地使用辅助病毒和/或包装细胞系形成的感染性病毒颗粒来表征。

[0416] 作为一个实例，合适的载体可通过使C68腺病毒立即早期基因E1a和延迟早期基因E1b的全部或足够部分缺失来形成，从而消除其正常的生物功能。当在含有提供相应反式基因产物的功能性腺病毒E1a和E1b基因的黑猩猩腺病毒转化的互补细胞系上生长时，复制缺陷型E1缺失病毒能够复制并产生感染性病毒。基于已知腺病毒序列的同源性，预期正如本领域的人类重组E1缺失腺病毒，所得重组黑猩猩腺病毒能够感染许多细胞类型并且可表达抗原，但除非以极高感染倍率感染细胞，否则无法在不携带黑猩猩E1区DNA的大多数细胞中复制。

[0417] 作为另一个实例，C68腺病毒延迟早期基因E3的全部或一部分可从形成重组病毒的一部分的黑猩猩腺病毒序列消除。

[0418] 还可构建具有E4基因缺失的黑猩猩腺病毒C68载体。另一个载体可在延迟早期基因E2a中含有缺失。

[0419] 还可在黑猩猩C68腺病毒基因组的晚期基因L1至L5中的任一者中获得缺失。类似地，中间基因IX和IVa2中的缺失可用于一些目的。可在其他结构性或非结构性腺病毒基因中获得其他缺失。

[0420] 上述缺失可单独使用，即腺病毒序列可仅含有E1缺失。或者，可以任何组合使用有效破坏或降低其生物活性的完整基因或其部分的缺失。例如，在一个示例性载体中，腺病毒C68序列可缺失E1基因和E4基因，或缺失E1、E2a和E3基因，或缺失E1和E3基因，或在缺失或不缺失E3的情况下缺失E1、E2a和E4基因等等。如上文所论述，此类缺失可与其他突变(例如温度敏感性突变)组合使用，以达到所需结果。

[0421] 包含一种或多种抗原的盒任选地插入至黑猩猩C68 Ad病毒的任何缺失区中。或者，如果需要，可将盒插入至现有基因区中以破坏所述区的功能。

[0422] V.E.7.辅助病毒

[0423] 取决于用于携带抗原盒的病毒载体的黑猩猩腺病毒基因含量，可使用辅助腺病毒或非复制性病毒片段来提供足够的黑猩猩腺病毒基因序列以产生含有所述盒的感染性重组病毒颗粒。

[0424] 有用的辅助病毒含有所选择的腺病毒基因序列，其不存在于腺病毒载体构建体中和/或不由载体转染的包装细胞系表达。辅助病毒可为复制缺陷型并且除上述序列之外，还

含有多种腺病毒基因。辅助病毒可与本文所述的表达E1的细胞系组合使用。

[0425] 对于C68,“辅助”病毒可为通过用SspI剪切C68基因组的C末端形成的片段,其从病毒的左端去除约1300bp。这种经剪切的病毒然后与质粒DNA共转染至表达E1的细胞系中,由此通过与质粒中的C68序列同源重组形成重组病毒。

[0426] 辅助病毒还可形成聚阳离子缀合物,如Wu等人,J.Biol.Chem.,264:16985-16987(1989);K.J.Fisher和J.M.Wilson,Biochem.J.,299:49(1994年4月1日)中所述。辅助病毒可任选地含有报告子。许多此类报告子是本领域已知的。与腺病毒载体上的抗原盒不同,辅助病毒上报告子的存在允许独立地监测Ad载体和辅助病毒。这种第二报告子用于纯化后能够将所得重组病毒与辅助病毒分离。

[0427] V.E.8.病毒颗粒的组装和细胞系的感染

[0428] 将腺病毒的经选择DNA序列、抗原盒和其他载体元件组装于各种中间质粒和穿梭载体中,以及使用质粒和载体制造重组病毒颗粒均可使用常规技术实现。此类技术包括cDNA的常规克隆技术、体外重组技术(例如吉布森组装(Gibson assembly))、腺病毒基因组的重叠寡核苷酸序列的使用、聚合酶链反应和提供所需核苷酸序列的任何合适的方法。采用标准转染和共转染技术,例如CaPO4沉淀技术或脂质体介导的转染方法,例如脂染胺。所采用的其他常规方法包括病毒基因组的同源重组、琼脂覆层中病毒的蚀斑、测量信号产生的方法等。

[0429] 例如,在构建和组装所需含抗原盒的病毒载体之后,可在辅助病毒存在下将载体转染于包装细胞系中。同源重组发生在辅助序列与载体序列之间,其允许载体中的腺病毒-抗原序列被复制并包装至病毒粒子衣壳中,从而产生重组病毒载体颗粒。

[0430] 所得重组黑猩猩C68腺病毒可用于将抗原盒转移至所选择的细胞中。在使用包装细胞系中生长的重组病毒的体内实验中,E1缺失的重组黑猩猩腺病毒在将盒转移至非黑猩猩(优选人类)细胞中展现效用。

[0431] V.E.9.重组病毒载体的用途

[0432] 所得含有抗原盒(如上所述,通过腺病毒载体和辅助病毒或腺病毒载体和包装细胞系的合作产生)的重组黑猩猩C68腺病毒因此提供可在体内或离体将一种或多种抗原递送至受试者的有效基因转移媒介物。

[0433] 上述重组载体根据基因疗法的公开方法施用于人类。携带抗原盒的黑猩猩病毒载体可施用于患者,优选悬浮于生物相容性溶液或药学上可接受的递送媒介物中。合适的媒介物包括无菌盐水。已知为药学上可接受的载剂并且为本领域技术人员所熟知的其他水性和非水性等渗无菌注射溶液以及水性和非水性无菌悬浮液可用于此目的。

[0434] 黑猩猩腺病毒载体是以足以转导人类细胞并提供足够水平的抗原转移和表达的量施用,从而提供治疗益处而无过度不良效应或具有医学上可接受的生理效应,其可由医药领域的技术人员来确定。常规和药学上可接受的施用途径包括(但不限于)直接递送至肝脏、鼻内、静脉内、肌内、皮下、皮内、口服和其他肠胃外施用途径。如果需要,可组合施用途径。

[0435] 病毒载体的剂量将主要取决于以下因素,例如所治疗的疾患、患者的年龄、体重和健康状况,并且因此可在患者当中变化。剂量将经调节以平衡治疗益处与任何副作用,并且此类剂量可根据采用重组载体的治疗应用而变化。可监测抗原表达水平以确定剂量施用频

率。

[0436] 重组复制缺陷型腺病毒可以“药学有效量”施用,即在施用途径中有效转染所需细胞并提供所选择的基因的足够表达水平以提供疫苗益处(即一些可测量的保护性免疫水平)的重组腺病毒的量。包含抗原盒的C68载体可与佐剂共同施用。佐剂可与载体分开(例如明矾)或在载体内编码,特别是在佐剂为蛋白质时。佐剂是本领域中所熟知的。

[0437] 常规且药学上可接受的施用途径包括(但不限于)鼻内、肌内、气管内、皮下、皮内、经直肠、口服和其他肠胃外施用途径。如果需要,可组合或调整施用途径,取决于免疫原或疾病。例如,在狂犬病预防中,皮下、气管内和鼻内途径是优选的。施用途径主要将取决于所治疗的疾病的性质。

[0438] 可监测对抗原的免疫水平以确定是否需要加强剂。例如,在评估血清中的抗体滴度后,可能需要任选的加强免疫。

[0439] VI.治疗和制造方法

[0440] 还提供了一种通过向受试者施用一种或多种抗原(例如使用本文公开的方法鉴定的多种抗原)而在受试者中刺激肿瘤特异性免疫应答、针对肿瘤进行疫苗接种、治疗和/或减轻受试者中的癌症症状的方法。

[0441] 还提供了一种通过向受试者施用一种或多种抗原(例如使用本文公开的方法鉴定的多种抗原)而在受试者中刺激感染性疾病生物体特异性免疫应答、针对感染性疾病生物体进行疫苗接种、治疗和/或减轻与感染性疾病生物体相关的感染症状的方法。

[0442] 在一些方面,受试者已诊断患有癌症或具有患癌症的风险。受试者可为人、狗、猫、马或需要肿瘤特异性免疫应答的任何动物。肿瘤可为任何实体肿瘤,例如乳房肿瘤、卵巢肿瘤、前列腺肿瘤、肺肿瘤、肾脏肿瘤、胃肿瘤、结肠肿瘤、睾丸肿瘤、头颈部肿瘤、胰腺肿瘤、脑肿瘤、黑色素瘤和其他组织器官肿瘤,以及血液肿瘤,例如淋巴瘤和白血病,包括急性髓细胞白血病、慢性髓细胞白血病、慢性淋巴细胞白血病、T细胞淋巴细胞白血病和B细胞淋巴瘤。

[0443] 在一些方面,受试者已被诊断患有感染或处于感染风险中(例如年龄、地理/旅行和/或工作相关的感染风险或易感性增加,或处于季节性和/或新型疾病感染风险中)。

[0444] 可以足以刺激CTL应答的量施用抗原。可以足以刺激T细胞应答的量施用抗原。可以足以刺激B细胞应答的量施用抗原。可以足以同时刺激T细胞应答和B细胞应答的量施用抗原。

[0445] 抗原可以单独施用或与其他治疗剂组合施用。治疗剂可包括靶向感染性疾病生物体的那些,例如抗病毒剂或抗生素剂。

[0446] 此外,可进一步向受试者施用抗免疫抑制/免疫刺激剂,例如检查点抑制剂。例如,可进一步向受试者施用抗CTLA抗体或抗PD-1或抗PD-L1。抗体对CTLA-4或PD-L1的阻断可增强患者对癌细胞的免疫应答。特别是,在遵循疫苗接种方案时,CTLA-4阻断已被证明是有效的。

[0447] 可确定疫苗组合物中每种抗原的最佳量和最佳给药方案。例如,抗原或其变体可制备用于静脉内(i.v.)注射、皮下(s.c.)注射、皮内(i.d.)注射、腹膜内(i.p.)注射、肌内(i.m.)注射。注射方法包括s.c.、i.d.、i.p.、i.m.和i.v.。DNA或RNA注射方法包括i.d.、i.m.、s.c.、i.p.和i.v.。其他施用疫苗组合物的方法是本领域技术人员已知的。

[0448] 可对疫苗进行编译，使得组合物中存在的抗原的选择、数目和/或量具组织、癌症、感染性疾病和/或患者特异性。例如，肽的精确选择可由给定组织中亲本蛋白的表达模式指导，或者由患者的突变或疾病状态指导。选择可取决于特定的癌症类型、特定的感染性疾病(例如，受试者感染或有感染风险的特定感染性疾病分离株/菌株)、疾病状态、疫苗接种的目标(例如，预防性或针对正在进行的疾病)、早期治疗方案、患者的免疫状态，当然还有患者的HLA-单倍型。此外，根据特定患者的个人需要，疫苗可含有个性化组分。实施例包括根据特定患者中抗原的表达来改变抗原的选择或在第一轮或治疗方案之后调整二次治疗。

[0449] 通过使用各种诊断方法，例如下文进一步描述的患者选择方法，可鉴定患者是否施用抗原疫苗。患者选择可涉及鉴定一种或多种基因的突变或表达模式。患者选择可涉及鉴定正在进行的感染的感染性疾病。患者选择可涉及鉴定感染性疾病感染的风险。在一些情况下，患者选择涉及鉴定患者的单倍型。可并行执行各种患者选择方法，例如，测序诊断可鉴定患者的突变和单倍型。各种患者选择方法可按顺序进行，例如，一种诊断测试鉴定突变并且另一种诊断测试鉴定患者的单倍型，并且其中每个测试可以相同(例如，高通量测序)或不同(例如，一种高通量测序和另一种Sanger测序)诊断方法。

[0450] 对于用作癌症或感染性疾病疫苗的组合物，在本文所述的组合物中可避免或以低量存在在正常组织中大量表达的具有类似正常自身肽的抗原。另一方面，如果已知患者的肿瘤或受感染细胞表达大量的某种抗原，则用于治疗这种癌症或感染的相应药物组合物可以大量存在和/或可包括多于一种对这种特定抗原或这种抗原途径具特异性的抗原。

[0451] 可将包含抗原的组合物施用于已经患有癌症或感染的个体。在治疗性应用中，将组合物以足以刺激对肿瘤抗原或感染性疾病生物体抗原的有效CTL应答并治愈或至少部分阻遏症状和/或并发症的量施用于受试者。足以实现这一点的量被定义为“治疗有效剂量”。对该用途有效的量将取决于例如组合物、施用方式、所治疗疾病的阶段和严重程度、患者的体重和一般健康状况以及处方医生的判断。应记住，组合物通常可用于严重疾病状态，即危及生命或可能危及生命的情况，尤其是当癌症已转移或感染性疾病生物体已诱发器官损伤和/或其他免疫病理学时。在这些情况下，鉴于外来物质的最小化和抗原的相对无毒性质，主治医师可以并且认为需要施用实质过量的这些组合物。

[0452] 对于治疗用途，可在检测或手术切除肿瘤时或者在检测或治疗感染时开始施用。这之后可为加强剂量直到至少症状基本上减轻并且持续此后一段时间，或认为提供了免疫性(例如，记忆B细胞或T细胞群，或产生抗原特异性B细胞或抗体)。

[0453] 用于治疗性治疗的药物组合物(例如疫苗组合物)旨在用于肠胃外、局部、经鼻、口服或局部施用。药物组合物可经肠胃外施用，例如静脉内、皮下、皮内或肌内施用。组合物可施用在手术切除部位以刺激针对肿瘤的局部免疫应答。可施用组合物以靶向受试者的特定感染组织和/或细胞。本文公开了用于肠胃外施用的组合物，其包含抗原溶液，并且将疫苗组合物溶解或悬浮在可接受的载剂例如水性载剂中。可使用多种水性载剂，例如水、缓冲水、0.9％盐水、0.3％甘氨酸、透明质酸等。这些组合物可通过常规的熟知灭菌技术灭菌，或者可经无菌过滤。所得水溶液可以原样包装使用，或冻干，冻干制剂在施用之前与无菌溶液组合。组合物可根据需要含有药学上可接受的辅助物质以接近生理条件，例如pH调节剂和缓冲剂、张力调节剂、润湿剂等，例如乙酸钠、乳酸钠、氯化钠、氯化钾、氯化钙、脱水山梨糖醇单月桂酸酯、三乙醇胺油酸酯等。

[0454] 抗原也可经由脂质体施用,脂质体将它们靶向特定的细胞组织,例如淋巴组织。脂质体也可用于增加半衰期。脂质体包括乳液、发泡体、胶束、不溶性单层、液晶、磷脂分散体、层状层等。在这些制剂中,待递送的抗原作为脂质体的一部分被并入,单独或联合与结合于例如淋巴样细胞中普遍存在的受体的分子,例如结合于CD45抗原的单克隆抗体,或联合其他治疗性或免疫原性组合物。因此,可将填充有所需抗原的脂质体导向淋巴样细胞的部位,然后所述脂质体在该部位递送选定的治疗性/免疫原性组合物。脂质体可由标准的囊泡形成脂质形成,其通常包括中性和带负电荷的磷脂和固醇,例如胆固醇。脂质的选择通常通过考虑例如脂质体大小、酸不稳定性和脂质体在血流中的稳定性来指导。多种方法可用于制备脂质体,如例如Szoka等人,Ann.Rev.Biophys.Bioeng.9;467(1980),美国专利第4,235,871号、第4,501,728号、第4,501,728号、第4,837,028号和第5,019,369号中所述。

[0455] 为了靶向免疫细胞,要并入脂质体中的配体可包括例如对所需免疫系统细胞的细胞表面决定簇特异的抗体或其片段。脂质体悬浮液可以静脉内、局部等施用,其剂量尤其根据施用方式、所递送的肽和所治疗疾病的阶段而变化。

[0456] 出于治疗或免疫目的,编码肽和任选地一种或多种本文所述肽的核酸也可施用于患者。许多方法可方便地用于将核酸递送至患者。例如,核酸可作为“裸DNA”直接递送。例如在Wolff等人,Science 247:1465-1468(1990)以及美国专利第5,580,859号和第5,589,466号中描述了这种方法。核酸也可使用弹道递送来施用,如例如美国专利第5,204,253号中所述。可施用仅由DNA组成的粒子。或者,可将DNA粘附到粒子例如金粒子上。用于递送核酸序列的方法可包括带有或不带有电穿孔的病毒载体、mRNA载体和DNA载体。

[0457] 还可将核酸与阳离子化合物例如阳离子脂质复合递送。脂质介导的基因递送方法描述于例如9618372WOAWO 96/18372;9324640WOAWO 93/24640;Mannino和Gould-Fogerite,BioTechniques6(7):682-691(1988);美国专利第5,279,833号Rose美国专利第5,279,833号;9106309WOAWO 91/06309;和Felgner等人,Proc.Natl.Acad.Sci.USA 84:7413-7414(1987)。

[0458] 抗原还可包括在基于病毒载体的疫苗平台中,例如牛痘、禽痘、自我复制甲病毒、马拉巴病毒、腺病毒(参见例如Tatsis等人,Adenoviruses,Molecular Therapy(2004)10,616-629)或慢病毒,包括(但不限于)第二、第三或杂交第二/第三代慢病毒和任一代的重组慢病毒,其设计成靶向特定细胞类型或受体(参见例如Hu等人,Immunization Delivered by Lentiviral Vectors for Cancer and Infectious Diseases,Immunol Rev.(2011)239(1):45-61;Sakuma等人,Lentiviral vectors:basic to translational,Biochem J.(2012)443(3):603-18;Cooper等人,Rescue of splicing-mediated intron loss maximizes expression in lentiviral vectors containing the human ubiquitin C promoter,Nucl.Acids Res.(2015)43(1):682-690;Zufferey等人,Self-Inactivating Lentivirus Vector for Safe and Efficient In Vivo Gene Delivery,J.Virol.(1998)72(12):9873-9880)。取决于上述基于病毒载体的疫苗平台的包装能力,这种方法可递送编码一种或多种抗原肽的一种或多种核苷酸序列。序列可侧接非突变序列,可由接头分开或者可在前面有一种或多种靶向亚细胞区室的序列(参见例如Gros等人,Prospective identification of neoantigen-specific lymphocytes in the peripheral blood of melanoma patients,Nat Med.(2016)22(4):433-8;Stronen等人,Targeting of cancer

neoantigens with donor-derived T cell receptor repertoires,Science.(2016)352(6291):1337-41;Lu等人,Efficient identification of mutated cancer antigens recognized by T cells associated with durable tumor regressions,Clin Cancer Res.(2014)20(13):3401-10)。在引入宿主中后,受感染细胞表达抗原,从而刺激针对一种或多种肽的宿主免疫(例如CTL)应答。可用于免疫方案中的牛痘载体和方法描述于例如美国专利第4,722,848号中。另一种载体是卡介苗(Bacille Calmet te Guerin,BCG)。BCG载体描述于Stover等人(Nature 351:456-460(1991))中。根据本文描述,可用于抗原的治疗性施用或免疫接种的各种其他疫苗载体,例如伤寒沙门氏菌载体等,对于本领域技术人员将是显而易见的。

[0459] 一种施用核酸的方法使用编码一个或多个表位的微型基因构建体。为了创建编码所选CTL表位(微型基因)以在人类细胞中表达的DNA序列,将表位的氨基酸序列进行反向翻译。人类密码子使用表用于指导每个氨基酸的密码子选择。这些表位编码DNA序列直接相连,形成一个连续的多肽序列。为了优化表达和/或免疫原性,可在微型基因设计中并入额外的元件。可反向翻译并包括在微型基因序列中的氨基酸序列的实例包括:辅助T淋巴细胞、表位、前导(信号)序列和内质网保留信号。此外,可通过包括与CTL表位相邻的合成(例如聚丙氨酸)或天然存在的侧翼序列来改进CTL表位的MHC呈递。通过组装编码微型基因正链和负链的寡核苷酸,将微型基因序列转化为DNA。使用熟知技术在适当的条件下合成、磷酸化、纯化和退火重叠的寡核苷酸(30-100个碱基长)。使用T4 DNA连接酶连接寡核苷酸的末端。然后可将这种编码CTL表位多肽的合成微型基因克隆到所需的表达载体中。

[0460] 纯化的质粒DNA可使用多种配方制备用于注射。其中最简单的是在无菌磷酸盐缓冲盐水(PBS)中重建冻干DNA。已经描述了多种方法,并且可使用新技术。如上所述,核酸方便地与阳离子脂质一起配制。此外,糖脂、融合脂质体、肽和统称为保护性、相互作用、非缩合(PINC)的化合物也可与纯化的质粒DNA复合,以影响例如稳定性、肌内分散或运输到特定器官或细胞类型的变量。

[0461] 还公开了一种制造疫苗的方法,所述方法包括执行本文公开的方法的步骤;以及生产包含多种抗原或多种抗原子集的疫苗。

[0462] 本文公开的抗原可使用本领域已知的方法制造。例如,本文公开的产生抗原或载体(例如,包括至少一种编码一种或多种抗原的序列的载体)的方法可包括在适合表达抗原或载体的条件下培养宿主细胞,其中所述宿主细胞包含至少一种编码抗原或载体的多核苷酸,并纯化所述抗原或载体。标准纯化方法包括色谱技术、电泳、免疫学、沉淀、透析、过滤、浓缩和色谱聚焦技术。

[0463] 宿主细胞可包括中国仓鼠卵巢(CHO)细胞、NSO细胞、酵母或HEK293细胞。宿主细胞可用一种或多种多核苷酸转化,所述多核苷酸包含至少一种编码本文公开的抗原或载体的核酸序列,任选地其中经分离的多核苷酸还包含可操作地连接到至少一种编码所述抗原或载体的核酸序列的启动子序列。在某些实施方案中,经分离的多核苷酸可为cDNA。

[0464] VII.抗原使用和施用

[0465] 疫苗接种方案可用于向受试者给与一种或多种抗原。可使用初免疫苗和加强疫苗对受试者进行给药。可使用的疫苗接种方法、方案和时间表包括但不限于国际申请公开WO2021092095中描述的那些,所述文献出于所有目的通过引用并入本文。

[0466] 初免疫苗可基于C68(例如,SEQ ID NO:1或2所示的序列)或SAM(例如,SEQ ID NO:3或4所示的序列)。加强疫苗也可基于C68(例如,SEQ ID NO:1或2所示的序列)或SAM(例如,SEQ ID NO:3或4所示的序列)。

[0467] 初免/加强策略中的每个载体通常包括包含抗原的盒。盒可包括约1-50种抗原,由间隔区例如通常围绕每个抗原的天然序列或其他非天然间隔区序列如AAY分开。盒还可包括MHCII抗原,例如破伤风类毒素抗原和PADRE抗原,它们可被认为是通用II类抗原。盒还可包括靶向序列,例如泛素靶向序列。此外,每个疫苗剂量可与免疫调节剂一起(例如,同时、之前或之后)施用于受试者。每个疫苗剂量可与检查点抑制剂(CPI)一起(例如,同时、之前或之后)施用于受试者。CPI可包括抑制CTLA4、PD1和/或PDL1的那些,例如抗体或其抗原结合部分。此类抗体可包括曲美木单抗或德瓦鲁单抗(durvalumab)。每个疫苗剂量可与细胞因子如IL-2、IL-7、IL-12(包括IL-12p35、p40、p70和/或p70融合构建体)、IL-15或IL-21一起(例如,同时、之前或之后)施用于受试者。每个疫苗剂量可与修饰的细胞因子(例如pegIL-2)一起(例如,同时、之前或之后)施用于受试者。

[0468] 初免疫苗可注射(例如肌内)于受试者。可使用每剂量双侧注射。例如,可使用一次或多次ChAdV68(C68)注射(例如总剂量1×10^{12}个病毒颗粒);可使用选自0.001至1ug RNA、特别是0.1或1ug范围的低疫苗剂量的一次或多次SAM载体注射;或者可使用选自1至100ug RNA、特别是10或100ug范围的高疫苗剂量的一次或多次SAM载体注射。

[0469] 可在初免疫苗接种之后注射(例如肌内)疫苗加强剂(加强疫苗)。加强疫苗可在初免后约每1、2、3、4、5、6、7、8、9或10周,例如每4周和/或8周施用。可使用每剂量双侧注射。例如,可使用一次或多次ChAdV68(C68)注射(例如总剂量1×10^{12}个病毒颗粒);可使用选自0.001至1ug RNA、特别是0.1或1ug范围的低疫苗剂量的一次或多次SAM载体注射;或者可使用选自1至100ug RNA、特别是10或100ug范围的高疫苗剂量的一次或多次SAM载体注射。

[0470] 也可向受试者施用抗CTLA-4(例如,曲美木单抗)。例如,抗CTLA4可在肌内疫苗注射(ChAdV68初免或SAM低剂量)部位附近皮下施用,以确保引流到同一淋巴结。曲美木单抗是CTLA-4的选择性人类IgG2 mAb抑制剂。靶向抗CTLA-4(曲美木单抗)皮下剂量通常为70-75mg(特别是75mg),剂量范围为例如1-100mg或5-420mg。

[0471] 在某些情况下,可使用抗PD-L1抗体,例如德瓦鲁单抗(MEDI4736)。德瓦鲁单抗是一种选择性、高亲和力的人类IgG1 mAb,其阻断PD-L1与PD-1和CD80的结合。德瓦鲁单抗通常每4周以20mg/kg i.v.施用。

[0472] 免疫监测可在疫苗施用之前、期间和/或之后进行。这种监测可告知安全性和功效以及其他参数。

[0473] 为了进行免疫监测,通常使用PBMC。PBMC可在初免疫苗接种之前和在初免疫苗接种之后(例如4周和8周)分离。PBMC可在即将加强疫苗接种之前和在每次加强疫苗接种之后(例如4周和8周)收集。

[0474] 免疫应答,例如T细胞应答和B细胞应答,可作为免疫监测方案的一部分进行评估。例如,可监测和/或评估本文所述的疫苗组合物刺激免疫应答的能力。如本文所用,“刺激免疫应答”是指免疫应答的任何增加,例如引发免疫应答(例如,在未接受过治疗的受试者中刺激引发免疫应答的初免疫苗)或增强免疫应答(例如,在对抗原具有预先存在的免疫应答,例如由初免疫苗引发的预先存在的免疫应答的受试者中刺激免疫应答增强的加强疫

苗)。可使用本领域中已知的一种或多种方法测量T细胞应答,例如ELISpot、细胞内细胞因子染色、细胞因子分泌和细胞表面捕捉、T细胞增殖、MHC多聚体染色或通过细胞毒性测定。针对疫苗中编码的表位的T细胞应答可通过使用ELISpot测定来测量细胞因子(例如IFN-γ)的诱导而从PBMC监测。针对疫苗中编码的表位的特异性CD4或CD8 T细胞应答可通过使用流式细胞术测量胞内或胞外捕捉的细胞因子(例如IFN-γ)的诱导而从PBMC监测。针对疫苗中编码的表位的特异性CD4或CD8 T细胞应答可通过使用MHC多聚体染色测量表达特异性针对表位/MHC I类复合物的T细胞受体的T细胞群体而从PBMC监测。针对疫苗中编码的表位的特异性CD4或CD8 T细胞应答可通过在3H-胸苷、溴脱氧尿苷和羧基荧光素-二乙酸酯-琥珀酰亚胺酯(CFSE)并入后测量T细胞群的离体扩增而从PBMC监测。特异性针对疫苗中编码的表位的源自PBMC的T细胞的抗原识别能力和溶解活性可通过铬释放测定或替代性比色细胞毒性测定来功能性评估。

[0475] B细胞应答可使用本领域已知的一种或多种方法测量,例如用于确定B细胞分化(例如,分化为浆细胞)、B细胞或浆细胞增殖、B细胞或浆细胞激活(例如,共刺激标志物如CD80或CD86的上调)、抗体类别转换和/或抗体产生(例如,ELISA)的测定法。也可评估抗体的功能,例如评估中和能力。

[0476] 可在施用本文所述的任何疫苗组合物之后监测受试者的疾病状态。例如,可使用从受试者分离的无细胞DNA(cfDNA)来监测疾病状态。此外,可使用从受试者分离的cfDNA监测疫苗疗法的功效。cfDNA监测可包括以下步骤:a.从受试者中分离出或已经分离出cfDNA;b.对分离的cfDNA进行测序或已经进行测序;c.确定或已经确定cfDNA中相对于受试者的野生型种系核酸序列的一个或多个突变的频率,以及d.从步骤(c)评估或已经评估受试者的疾病状态。所述方法还可包括,在上述步骤(c)之后,d.对给定受试者执行步骤(a)-(c)的多于一次迭代,并比较在多于一次迭代中确定的一个或多个突变的频率;以及f.从步骤(d)评估或已经评估受试者的疾病状态。可在不同的时间点进行多于一次的迭代,例如在施用疫苗组合物之前进行步骤(a)-(c)的第一次迭代并且在施用疫苗组合物之后进行步骤(a)-(c)的第二次迭代。步骤(c)可包括比较:在多于一次迭代中确定的一个或多个突变的频率,或在第一次迭代中确定的一个或多个突变的频率与在第二次迭代中确定的一个或多个突变的频率。在后续迭代或第二次迭代中确定的一个或多个突变频率的增加可被评估为疾病进展。在后续迭代或第二次迭代中确定的一个或多个突变的频率降低可被评估为应答。在一些方面,应答是完全应答(CR)或部分应答(PR)。可在评估步骤之后对受试者施用疗法,例如当评估cfDNA中的一个或多个突变的频率表明受试者患有疾病时。cfDNA分离步骤可使用离心将cfDNA从细胞或细胞碎片中分离出来。cfDNA可从全血中分离出来,例如通过分离血浆层、血沉棕黄层和红血。cfDNA测序可使用下一代测序(NGS)、Sanger测序、双重测序、全外显子组测序、全基因组测序、从头测序、定相测序、靶向扩增子测序、鸟枪法测序或其组合,并且可包括在测序前针对一种或多种目标多核苷酸区域(例如,已知或怀疑编码一个或多个突变的多核苷酸、编码区域和/或肿瘤外显子组多核苷酸)富集cfDNA。富集cfDNA可包括将一种或多种可被修饰(例如,生物素化)的多核苷酸探针与一种或多种目标多核苷酸区域杂交。一般而言,可同时或并行监测任何数目的突变。

[0477] VIII.HLA肽的分离与检测

[0478] 在组织样品裂解和溶解后,使用经典的免疫沉淀(IP)方法进行HLA-肽分子的分离

(55-58)。澄清的裂解物用于HLA特异性IP。

[0479] 免疫沉淀是使用与珠粒偶联的抗体来进行,其中抗体特异性针对HLA分子。对于泛I类HLA免疫沉淀,使用泛I类CR抗体;对于II类HLA-DR,使用HLA-DR抗体。在过夜孵育期间,抗体共价连接到NHS-琼脂糖珠粒。在共价连接后,将珠粒洗涤并等分用于IP。(59,60)免疫沉淀也可用未共价连接到珠粒的抗体来进行。这通常使用包被有蛋白A和/或蛋白G的琼脂糖或磁性珠粒来完成,所述珠粒将抗体固定于柱。下文列出可用于选择性地富集MHC/肽复合物的一些抗体。

[0480]

抗体名称	特异性
W6/32	I类HLA-A、B、C
L243	II类—HLA-DR
Tu36	II类—HLA-DR
LN3	II类—HLA-DR
Tu39	II类—HLA-DR、DP、DQ

[0481] 将澄清的组织裂解物添加至抗体珠粒中进行免疫沉淀。在免疫沉淀后,从裂解物去除珠粒并且将裂解物储存用于额外实验,包括额外IP。洗涤IP珠粒以去除非特异性结合并且使用标准技术从珠粒洗脱HLA/肽复合物。使用分子量旋转柱或C18分级分离从肽去除蛋白质组分。所得肽通过SpeedVac蒸发变干并且在一些情况下在MS分析之前储存在-20C下。

[0482] 干燥的肽在适于反相色谱的HPLC缓冲液中复溶并装载于C-18微毛细管HPLC柱上,以便在Fusion Lumos质谱仪(Thermo)中梯度洗脱。在Orbitrap检测器中以高分辨率收集肽质量/电荷(m/z)的MS1谱,接着在所选择的离子的HCD片段化后,在离子阱检测器中收集MS2低分辨率扫描。另外,可使用CID或ETD片段化方法或三种技术的任何组合来获得MS2谱,以达到肽的更大的氨基酸覆盖率。MS2谱也可在Orbitrap检测器中使用称为平行反应监测的靶向方法以高分辨率质量精度测量。在靶向PRM中,特定的肽前体离子在Orbitrap检测器中被分离,并且所有产生的HCD碎裂离子在肽峰的洗脱过程中被扫描。这使得在共注射的稳定同位素标记的肽标准物存在的情况下能够对内源性肽进行肽鉴定和定量。

[0483] 使用Comet对来自每个分析的MS2谱进行蛋白质数据库搜索(61,62),并使用Percolator对肽鉴定进行评分(63-65)。使用PEAKS studio(Bioinformatics Solutions Inc.)来进行额外测序,并且可使用其他搜索引擎或测序方法,包括频谱匹配和从头测序(97)。靶向MS1和MS2谱通过Skyline(104)进行处理。

[0484] VIII.B.1.支持综合HLA肽测序的MS检测限值研究

[0485] 使用肽YVYVADVAAK(SEQ ID NO:88),使用装载于LC柱上的不同量的肽确定检测限值。所测试的肽的量为1pmol、100fmol、10fmol、1fmol和100amol。(表1)这些结果表明,最低检测限值(LoD)在埃摩尔范围(10^{-18})中,动态范围跨越五个数量级,并且信噪比足以在低飞摩尔范围(10^{-15})测序。

[0486] 表1

[0487]

肽m/z	装载于柱上	1e9个细胞中的拷贝数/细胞
566.830	1pmol	600
562.823	100fmol	60

559.816	10fmol	6
556.810	1fmol	0.6
553.802	100amol	0.06

[0488] IX.呈递模型

[0489] 呈递模型可用于鉴定在患者中肽呈递的可能性。本领域技术人员已知各种呈递模型，例如在美国专利第10,055,540号、美国申请公开第US20200010849A1号和第US20110293637号以及国际专利申请公开WO/2018/195357、WO/2018/208856和WO2016187508中更详细描述的呈递模型，所述文献出于所有目的各自通过引用整体并入本文。

[0490] X.训练模块

[0491] 训练模块可用于基于训练数据集构建一个或多个呈递模型，所述模型产生肽序列是否会被与所述肽序列相关联的MHC等位基因呈递的可能性。各种训练模块是本领域技术人员已知的，例如更详细地描述于美国专利第10,055,540号、美国申请公开第US20200010849A1号和国际专利申请公开WO/2018/195357和WO/2018/208856中的呈递模型，所述文献出于所有目的各自通过引用整体并入本文。训练模块可在独立等位基因(per-allele)基础上构建呈递模型以预测肽的呈递可能性。训练模块也可在存在两种或更多种MHC等位基因的多等位基因环境中构建呈递模型以预测肽的呈递可能性。

[0492] XI.预测模块

[0493] 预测模块可用于接收序列数据并使用呈递模型在序列数据中选择候选抗原。具体地，序列数据可以是从患者的肿瘤组织细胞、患者的受感染细胞或感染性疾病生物体本身中提取的DNA序列、RNA序列和/或蛋白质序列。预测模块可通过比较从患者的正常组织细胞中提取的序列数据与从患者的肿瘤组织细胞中提取的序列数据以鉴定含有一个或多个突变的部分，从而鉴定作为突变肽序列的候选新抗原。预测模块可例如通过比较从患者的正常组织细胞中提取的序列数据与从患者的受感染细胞中提取的序列数据以鉴定含有一种或多种感染性疾病生物体相关抗原的部分，从而鉴定作为病原体衍生肽、病毒衍生肽、细菌衍生肽、真菌衍生肽和寄生虫衍生肽的候选抗原。预测模块可通过将从患者的正常组织细胞中提取的序列数据与从患者的肿瘤组织细胞中提取的序列数据进行比较以鉴定不当表达的候选抗原，从而鉴定与正常细胞或组织相比在肿瘤细胞或癌组织中的表达有所改变的候选抗原。预测模块可通过比较从患者的正常组织细胞中提取的序列数据与从患者的受感染组织细胞中提取的序列数据以鉴定表达的候选抗原(例如，鉴定感染性疾病特异的表达的多核苷酸和/或多肽)，从而鉴定与正常细胞或组织相比在受感染细胞或受感染组织中表达的候选抗原。

[0494] 呈递模块可将一种或多种呈递模型应用于处理过的肽序列以估计肽序列的呈递可能性。具体地，预测模块可通过将呈递模型应用于候选抗原来选择一种或多种候选抗原肽序列，所述候选抗原肽序列可能在肿瘤HLA分子或受感染细胞HLA分子上呈递。在一种实现方式中，呈递模块选择估计呈递可能性高于预定阈值的候选抗原序列。在另一个实现方式中，呈递模型选择具有最高估计呈递可能性的N个候选抗原序列(其中N通常是可在疫苗中递送的表位的最大数目)。可将包含针对给定患者选择的候选抗原的疫苗注射到受试者体内以刺激免疫应答。

[0495] XI.B.盒设计模块

[0496] XI.B.1概述

[0497] 盒设计模块可用于基于所选择的候选肽产生疫苗盒序列以注射到患者体内。各种盒设计模块是本领域技术人员已知的，例如更详细地描述于美国专利第10,055,540号、美国申请公开第US20200010849A1号以及国际专利申请公开WO/2018/195357和WO/2018/208856中的盒设计模块，所述文献各自出于所有目的通过引用整体并入本文。

[0498] 可基于由预测模块确定的与超过预定阈值的呈递可能性相关联的所选肽产生治疗性表位集合，其中所述呈递可能性是由呈递模型确定。然而，应了解，在其他实施方案中，可基于多种方法中的任一种或多种(单独或组合形式)，例如基于针对患者的HLA I类或II类等位基因的结合亲和力或预测的结合亲和力、针对患者的HLA I类或II类等位基因的结合稳定性或预测的结合稳定性、随机取样等产生治疗性表位的集合。

[0499] 治疗性表位可对应于自身选择的肽。除所选肽外，治疗性表位还可包括C末端和/或N末端侧接序列。N末端和C末端侧接序列可为治疗性疫苗表位在其源蛋白的背景下的原生N末端和C末端侧接序列。治疗性表位可表示固定长度的表位。治疗性表位可表示可变长度的表位，其中表位的长度可根据例如C-侧接序列或N-侧接序列的长度而改变。例如，C末端侧接序列和N末端侧接序列可各自具有2-5个残基的变化长度，由此产生16种可能的表位选择。

[0500] 盒设计模块也可通过考虑横跨所述盒中一对治疗性表位之间的接合点的接合点表位的呈递来产生盒序列。接合点表位是由于在所述盒中串接治疗性表位和接头序列的过程而在所述盒中产生的新型非自身但不相关的表位序列。接合点表位的新型序列不同于所述盒的治疗性表位本身。

[0501] 盒设计模块可产生降低在患者中呈递接合点表位的可能性的盒序列。具体来说，当将盒注射至患者体内时，接合点表位有可能被患者的I类HLA或II类HLA等位基因呈递，并且分别刺激CD8或CD4T细胞应答。由于T细胞与接合点表位的反应没有治疗益处，并且可能因抗原竞争而减弱针对所述盒中所选治疗性表位的免疫应答，因此此类反应常常是不合需要的。[76]

[0502] 盒设计模块可迭代一个或多个候选盒，并确定与盒序列相关联的接合点表位的呈递评分低于数字阈值的盒序列。接合点表位呈递评分是与所述盒中接合点表位的呈递可能性相关联的量，并且较高的接合点表位呈递评分值指示所述盒的接合点表位将由I类HLA或II类HLA或两者呈递的可能性较高。

[0503] 在一个实施方案中，盒设计模块可确定候选盒序列中与最低接合点表位呈递评分相关联的盒序列。

[0504] 盒设计模块可迭代一个或多个候选盒序列，确定候选盒的接合点表位呈递评分，并且鉴定与低于阈值的接合点表位呈递评分相关联的最佳盒序列。

[0505] 盒设计模块可进一步检查所述一个或多个候选盒序列以鉴定候选盒序列中的接合点表位中的任一个是否是设计使用所述疫苗的给定患者的自身表位。为了实现此目的，盒设计模块针对已知数据库，例如BLAST检查接合点表位。在一个实施方案中，盒设计模块可被配置以设计避免接合点自身表位的盒。

[0506] 盒设计模块可执行蛮力方法并且迭代所有或大部分可能的候选盒序列以选择具有最小接合点表位呈递评分的序列。然而，由于疫苗容量增加，此类候选盒的数量可能极

大。例如,对于20个表位的疫苗容量,盒设计模块必须迭代约10^{18}个可能的候选盒,才能确定具有最低接合点表位呈递评分的盒。对于盒设计模块在合理的时间量内完成以产生用于患者的疫苗而言,这种确定在计算上可能较为繁琐(就所需的计算处理资源而言)并且有时难以处理。另外,考虑到每个候选盒的可能接合点表位,甚至可能更繁琐。因此,盒设计模块可基于迭代明显少于蛮力方法中的候选盒序列数量的候选盒数量来选择盒序列。

[0507] 盒设计模块可产生随机产生或至少伪随机产生的候选盒子集,并且选择与低于预定阈值的接合点表位呈递评分相关联的候选盒作为盒序列。另外,盒设计模块可从所述子集中选择具有最低接合点表位呈递评分的候选盒作为盒序列。例如,盒设计模块可针对20个所选表位的集合产生约1百万个候选盒的子集,并且选出具有最小接合点表位呈递评分的候选盒。尽管产生随机盒序列的子集并从所述子集中选出具有低接合点表位呈递评分的盒序列可能不如蛮力方法好,但其需要明显较少的计算资源,由此使其实施在技术上是可行的。另外,相对于这种更高效的技术,执行蛮力法可能仅引起接合点表位呈递评分的微小或甚至可忽略的改进,因此,从资源分配的观点看,蛮力法是不值得实施的。盒设计模块可通过将盒的表位序列用非对称旅行商问题(TSP)公式表示来确定改进的盒配置。根据节点列表和每对节点之间的距离,TSP确定与最短总距离相关联的节点的序列以访问每个节点恰好一次并返回至原始节点。例如,鉴于彼此之间距离已知的城市A、B和C,TSP的解决方案产生一个闭合的城市序列,对于所述序列,访问每个城市恰好一次所行进的总距离是可能途径当中最短的。TSP的非对称形式确定当一对节点之间的距离不对称时节点的最佳序列。例如,从节点A行进至节点B的“距离”可以不同于从节点B行进至节点A的“距离”。通过使用非对称TSP解决改进的最佳盒,盒设计模块可找到使所述盒的表位之间的接合点的呈递评分降低的盒序列。非对称TSP解决方案指示对应于应当在盒中串接表位以使所述盒的所有接合点的接合点表位呈递评分减到最小的次序的治疗性表位序列。相较于随机取样方法,通过这种方法确定的盒序列可产生具有明显较少接合点表位呈递的序列,同时可能需要明显较少的计算资源,尤其是在所产生的候选盒序列数量很大时。不同计算方法和用于优化盒设计的比较的说明性实例更详细地描述于美国专利第10,055,540号、美国申请公开第US20200010849A1号以及国际专利申请公开WO/2018/195357和WO/2018/208856中,所述文献出于所有目的各自通过引用整体并入本文。

[0508] 通过其原生侧翼序列连接的串接KRAS相关MHC I类新表位的说明性非限制性盒包括具有突变KRAS G12C、KRAS G12D、KRAS G12V和KRAS Q61H的KRAS新表位中的每一者的4次迭代,并且已被排序以最小化潜在的连接表位,由SEQ ID NO:65所示的氨基酸序列表示并具有以下KRAS相关新表位的顺序:G12C G12D Q61HG12D G12V G12C Q61H G12D G12V G12C Q61H G12D G12V Q61HG12V G12C。

[0509] 本领域技术人员可选择用于包括在共有抗原疫苗中的共有(新)抗原序列和利用这种疫苗治疗的适当患者,例如,如出于所有目的通过引用并入本文的美国申请第17/058,128号中所述。候选共有(新)抗原的质谱(MS)验证可作为选择过程的一部分进行。

[0510] XIII.示例计算机

[0511] 计算机可用于计算本文所述的方法中的任一者。本领域技术人员将认识到计算机可具有不同架构。计算机的实例是本领域技术人员已知的,例如更详细地描述于美国专利第10,055,540号、美国申请公开第US20200010849A1号以及国际专利申请公开WO/2018/

195357和WO/2018/208856中的计算机,所述文献出于所有目的各自通过引用整体并入本文。

[0512] XIV.实施例

[0513] 下文是用于执行本发明的特定实施方案的实施例。所述实施例仅出于说明性目的提供,并且不旨在以任何方式限制本发明的范围。已作出努力以确保所使用数字的精确性(例如,量、温度等),但一些实验性误差和偏差当然应是允许的。

[0514] 除非另外指明,否则本发明的实践将采用本领域的技能范围内的蛋白质化学、生物化学、重组DNA技术和药理学常规方法。在文献中充分解释此类技术。参见例如T.E.Creighton,Proteins:Structures and Molecular Properties(W.H.Freeman and Company,1993);A.L.Lehninger,Biochemistry(Worth Publishers,Inc.,现行版);Sambrook等人,Molecular Cloning:A Laboratory Manual(第2版,1989);Methods In Enzymology(S.Colowick和N.Kaplan编,Academic Press,Inc.);Remington's Pharmaceutical Sciences,第18版(Easton,Pennsylvania:Mack Publishing Company,1990);Carey和Sundberg Advanced Organic Chemistry第3版(Plenum Press)第A卷和第B卷(1992)。

[0515] XIV.A.具有重复新表位和/或去除免疫显性新表位的盒的评价概述

[0516] 通过疫苗接种,可递送刺激一种或多种相应细胞免疫应答的多种I类MHC限制性新抗原。若干疫苗盒被工程化而以单一基因产物形式编码多个新表位,特别是多个不同的KRAS相关新表位,其中所述表位嵌入其天然的周围肽序列内。各种盒设计的特征在于一个或多个新表位的多个拷贝(即迭代)。各种盒设计的特征还在于去除免疫显性新表位。

[0517] XIV.B.重复新表位和/或去除免疫显性新表位评价材料和方法

[0518] 抗原

[0519] 以下实施例中使用的含有表位的25聚体氨基酸序列(即,侧接有原生N和C末端氨基酸接头的表位)呈现于表2A中。如下所述使用。用于各种构建体的盒的抗原编码序列是通过按照以下实施例中描述的顺序和编号直接彼此连接每个25聚体序列(即,连续25聚体序列之间无额外氨基酸)来构建,例如,参见图1A、图2A、图3A和图4A。包含含有连接在一起的多个不同表位的全长抗原编码序列以及通用MHC II类抗原破伤风类毒素和PADRE(粗体序列)的盒呈现于表2B中。插入载体中的完整外源核苷酸从5'至3'包括:Kozak序列(GCCACC)、编码三个氨基酸MAG的核苷酸(ATGGCCGGG)、表2B的盒序列之一和两个终止密码子(TAATGA)。用个别的KRAS突变G12V、G12C和G12D(即,SEQ ID NO 57、58或59的单独多联体)的4次迭代生成更多的盒。

[0520] 表2A-含有25聚体新表位的序列

[0521]

抗原名称	25聚体序列	SEQ ID NO:
KRAS G12C	MTEYKLVVVGACGVGKSALTIQLIQ	57
KRAS G12D	MTEYKLVVVGADGVGKSALTIQLIQ	58
KRAS G12V	MTEYKLVVVGAVGVGKSALTIQLIQ	59
KRAS Q61H	ETCLLDILDTAGHEEYSAMRDQYMR	60
TP53 R213L	LRVEYLDDRNTFLHSVVVPYEPPEV	61
TP53 S127Y	HSGTAKSVTCTYYPALNKMFCQLAK	62

[0522] 表2B-盒中的全长多新表位序列

1X (20x1)盒(SEQ ID NO: 63)
MTEYKLVVVGAGDVGKSALTIQLIQETCLLDILDTAGKEEYSAMR DQYMRMCNSSCMGGMNRMPILTIITLEDSSSGIHSGATTTAPPLSG KGNPEEEDVSGIHSGATTTAPFLSGKGNPEEEDVEILDEAYVMAYV MAGVGSPYVSRLLMTEYKLVVVGADGVGKSALTIQLIQETCLLDI LDTAGREEYSAMRDQYMRSYLDSGIHSGATATAPSLSGKGNPEKS VTCTYSPALNNMFCQLAKTCPVQMTEYKLVVVGAAGVGKSALTI QLIQETCLLDILDTAGLEEYSAMRDQYMRLRVEYLDDRNTFLHSV VVPYEPPEVDGQITVGQRIGSVSFGTVYKGKWHGMTEYKLVVVG AVGVGKSALTIQLIQETCLLDILDTAGHEEYSAMRDQYMRWQQQS YLDSGIHFGATTTAPSLSGKHSGTAKSVTCTYYPALNKMFCQLAK KSVTCTYSPALNEMFCQLAKTCPVQMTEYKLVVVGACGVGKSAL TIQLIQ**GPGPGAKFVAAWTLKAAAGPGPGQYIKANSKFIGITEL GPGPG**
KRAS 2X (8x2)盒(SEQ ID NO: 64)
MTEYKLVVVGACGVGKSALTIQLIQETCLLDILDTAGHEEYSAMR DQYMRMTEYKLVVVGAGDVGKSALTIQLIQETCLLDILDTAGREE YSAMRDQYMRMTEYKLVVVGADGVGKSALTIQLIQETCLLDILD TAGLEEYSAMRDQYMRMTEYKLVVVGAVGVGKSALTIQLIQETC LLDILDTAGKEEYSAMRDQYMRMTEYKLVVVGACGVGKSALTIQ LIQETCLLDILDTAGKEEYSAMRDQYMRMTEYKLVVVGAGDVGK SALTIQLIQETCLLDILDTAGHEEYSAMRDQYMRMTEYKLVVVGA VGVGKSALTIQLIQETCLLDILDTAGREEYSAMRDQYMRMTEYKL VVVGADGVGKSALTIQLIQETCLLDILDTAGLEEYSAMRDQYMR **GPGPGAKFVAAWTLKAAAGPGPGQYIKANSKFIGITELGPGPG**
KRAS 4X (4x4)盒(SEQ ID NO: 65)
MTEYKLVVVGACGVGKSALTIQLIQMTEYKLVVVGADGVGKSAL TIQLIQETCLLDILDTAGHEEYSAMRDQYMRMTEYKLVVVGADG VGKSALTIQLIQMTEYKLVVVGAVGVGKSALTIQLIQMTEYKLVV VGACGVGKSALTIQLIQETCLLDILDTAGHEEYSAMRDQYMRMTE YKLVVVGADGVGKSALTIQLIQMTEYKLVVVGAVGVGKSALTIQL IQMTEYKLVVVGACGVGKSALTIQLIQETCLLDILDTAGHEEYSA MRDQYMRMTEYKLVVVGADGVGKSALTIQLIQMTEYKLVVVGA VGVGKSALTIQLIQETCLLDILDTAGHEEYSAMRDQYMRMTEYKL VVVGAVGVGKSALTIQLIQMTEYKLVVVGACGVGKSALTIQLIQ**G PGPGAKFVAAWTLKAAAGPGPGQYIKANSKFIGITELGPGPG**
KRAS 4x1 盒(SEQ ID NO: 66)
MTEYKLVVVGACGVGKSALTIQLIQMTEYKLVVVGADGVGKSAL TIQLIQETCLLDILDTAGHEEYSAMRDQYMRMTEYKLVVVGAVG VGKSALTIQLIQ**GPGPGAKFVAAWTLKAAAGPGPGQYIKANSK FIGITELGPGPG**
KRAS 4x1 + R213L 盒(SEQ ID NO: 67)

[0523]

MTEYKLVVVGACGVGKSALTIQLIQMTEYKLVVVGADGVGKSAL TIQLIQETCLLDILDTAGHEEYSAMRDQYMRMTEYKLVVVGAVG VGKSALTIQLIQLRVEYLDDRNTFLHSVVVPYEPPEV**GPGPGAKF** **VAAWTLKAAAGPGPGQYIKANSKFIGITELGPGPG**
KRAS 4x1 + S127Y 盒(SEQ ID NO: 68)
MTEYKLVVVGACGVGKSALTIQLIQMTEYKLVVVGADGVGKSAL TIQLIQETCLLDILDTAGHEEYSAMRDQYMRMTEYKLVVVGAVG VGKSALTIQLIQHSGTAKSVTCTYYPALNKMFCQLAK**GPGPGAKF** **VAAWTLKAAAGPGPGQYIKANSKFIGITELGPGPG**

[0524]

[0525] 腺病毒载体

[0526] 用于抗原表达系统的经修饰ChAdV68载体是基于AC_000011.1生成的,其中E1(nt 577至3403)和E3(nt 27,125-31,825)序列缺失并且相应的ATCC VR-594(独立测序的全长VR-594C68 SEQ ID NO:10)核苷酸在五个位置处被取代。在五个位置处被相应ATCC VR-594核苷酸取代的全长ChAdVC68 AC_000011.1序列被称为“ChAdV68.5WTnt”(SEQ ID NO:1)。插入在CMV启动子/增强子控制下的抗原盒代替缺失的E1序列。在抗原盒中含有20个模型抗原的代表性ChAdV68载体是“ChAdV68.5WTnt.MAG25聚体”(SEQ ID NO:2)。以下文描述的具有MAG25聚体聚体盒的抗原盒为特征的载体可被上文(例如在表2B中)描述的抗原盒替换。

[0527] 293F细胞中的腺病毒生产

[0528] ChAdV68病毒生产在293F细胞中进行,所述细胞在293FreeStyle™(ThermoFisher)培养基中的8%CO_2的孵育箱中生长。在感染当天,细胞被稀释至每毫升10^6个细胞,具有98%的存活率并且每次生产操作在1L摇瓶(Corning)中使用400mL。每次感染使用4mL的目标MOI>3.3的三级病毒原液。将细胞孵育48-72小时,直到如通过台盼蓝测量的存活率<70%。然后通过离心、全速台式离心机收获受感染细胞并在1XPBS中洗涤,重新离心,然后重悬在20mL的10mM Tris pH7.4中。通过冻融3次来裂解细胞团块并通过在4,300Xg下离心5分钟来澄清。

[0529] 通过CsCl离心进行腺病毒纯化

[0530] 通过CsCl离心使病毒DNA纯化。执行两次不连续梯度操作。第一次是从细胞组分中纯化出病毒并且第二次是从细胞组分进一步优化分离并且将缺陷性粒子与感染性粒子分离。

[0531] 将10mL的1.2(26.8g CsCl溶解于92mL的10mM Tris pH 8.0中)CsCl添加至异质同晶聚合物管中。然后,小心地添加8mL的1.4CsCl(53g CsCl溶解于87mL的10mM Tris pH 8.0中),使用移液管递送至管底部。将澄清的病毒小心地铺在所述1.2层的顶部上。需要时,再添加10mM Tris以使各管平衡。然后将所述管置放于SW-32Ti旋转器中并在10℃下离心2小时30分钟。然后将所述管移至层流柜中并且使用18号针和10mL注射器抽吸病毒带。注意避免取出污染性宿主细胞DNA和蛋白质。然后用10mM Tris pH 8.0将所述带稀释至少2倍并如前所述铺在如上文所描述的不连续梯度上。如前所述进行操作,然而此时进行所述操作过夜。第二天,小心抽吸带以避免抽吸出任何缺陷性粒子带。然后使用Slide-a-Lyzer™盒(Pierce)针对ARM缓冲液(20mM Tris pH 8.0、25mM NaCl、2.5%甘油)透析病毒。进行此操作3次,每次更换缓冲液保持1小时。然后将病毒等分以在-80℃下储存。

[0532] 腺病毒病毒测定

[0533] 基于1.1×10^{12}个病毒颗粒(VP)的消光系数相当于在OD260 nm下的吸光度值1,通过使用OD 260测定来确定VP浓度。在病毒裂解缓冲液(0.1%SDS、10mM Tris pH 7.4、1mM EDTA)中制备腺病毒的两种稀释液(1:5和1:10)。一式两份测量所述两种稀释液的OD并通过用OD260值乘以稀释因子乘以1.1×10^{12}VP来测量VP浓度/mL。

[0534] 通过病毒原液的限制性稀释测定来计算感染单位(IU)滴度。病毒起初在DMEM/5%NS/1X PS中稀释100倍并且随后,使用10倍稀释法稀释至1×10^{-7}。然后,将100μL这些稀释液添加至在之前至少一小时以3e5个细胞/孔接种于24孔板中的293A细胞中。一式两份地执行此操作。将板在37℃下在CO2(5%)孵育箱中孵育48小时。然后用1XPBS洗涤细胞,并且然后用100%冷甲醇(-20℃)固定。然后将板在-20℃下孵育最少20分钟。用1XPBS洗涤各孔,然后在室温下在1XPBS/0.1%BSA中封闭1小时。以的1:8,000稀释度添加兔抗Ad抗体(Abcam,Cambridge,MA)于封闭缓冲液中(每孔0.25ml)并在室温下孵育1小时。用每孔0.5mL PBS洗涤各孔4次。每孔添加1000倍稀释的HRP缀合的山羊抗兔抗体(Bethyl Labs,Montgomery Texas)并孵育1小时,然后进行最后一轮洗涤。进行5次PBS洗涤并使用于含0.01%H_2O_2的Tris缓冲盐水中的二氨基联苯胺四盐酸盐(Diaminobenzidine tetrahydrochloride,DAB)底物(0.67mg/mL DAB于50mM Tris pH 7.5中,150mM NaCl)使所述板显色。使各孔显色5分钟,然后计数。使用产生每个视野4-40个经染色细胞的稀释液,在10X物镜下对细胞计数。所用视野是0.32mm^2栅格,其中相当于在24孔板上每个视野有625个栅格。可通过每个栅格中经染色细胞的数目乘以每个视野的栅格数目乘以稀释因子10来确定感染性病毒数目/mL。类似地,当用GFP表达细胞操作时,可使用荧光而非衣壳染色来确定每毫升中GFP表达病毒粒子的数目。

[0535] SAM载体

[0536] 用于抗原表达系统的RNA甲病毒骨架由自我复制的委内瑞拉马脑炎(VEE)病毒产生(Kinney,1986,Virology 152:400-413),其通过缺失位于26S亚基因组启动子的3'的VEE的结构蛋白(缺失了VEE序列7544至11,175;编号基于Kinney等人,1986;SEQ ID NO:6)。为了生成自扩增mRNA("SAM")载体,将缺失的序列替换为抗原序列。含有20种模型抗原的代表性SAM载体是"VEE-MAG25聚体"(SEQ ID NO:4)。以具有MAG25聚体盒的所述抗原盒为特征的载体可用本文所述的SARS-CoV-2盒和/或全长蛋白替换。

[0537] 体外转录以生成SAM

[0538] 对于体内研究:SAM载体生成为"AU-SAM"载体。将缺少典型3'二核苷酸GG的经修饰的T7 RNA聚合酶启动子(TAATACGACTCACTATA(SEQ ID NO: 89))添加到SAM载体的5'端以生成体外转录模板DNA(例如,以SEQ ID NO:6所示的序列;具有7544至11,175缺失而无插入抗原盒的SAM)。反应条件描述如下:

[0539] -使用最终浓度为1x T7 RNA聚合酶混合物(E2040S)的1x转录缓冲液(40mM Tris-HCL[pH7.9]、10mM二硫苏糖醇、2mM亚精胺、0.002%Triton X-100和27mM氯化镁);0.025mg/mL DNA转录模板(通过限制性消化来线性化);8mM CleanCap试剂AU(目录号N-7114)和各10mM的ATP、三磷酸胞苷(CTP)、GTP和三磷酸尿苷(UTP)

[0540] -将转录反应物在37℃下孵育2小时并在37℃下用DNA酶I缓冲液中的最终2U DNA酶I(AM2239)/0.001mg DNA转录模板处理1小时。

[0541] -SAM由RNeasy Maxi(QIAGEN,75162)纯化

[0542] 除了共转录添加5'帽结构之外,可在转录后使用含有mRNA2'-O-甲基转移酶和S-腺苷甲硫氨酸的牛痘加帽系统(NEB目录号M2080)酶促添加7-甲基鸟苷或相关的5'帽结构。

[0543] 免疫

[0544] 对于ChAdV68疫苗,向表达嵌合HLA-A11:01的转基因小鼠(Taconic型号9660-[CB6F1-Tg(HLA-A*1101/H2-Kb)A11.01])注射$8x10^{10}$个病毒颗粒(VP)或$5x10^{10}$个VP,如所示,体积为100μL,双侧肌内注射(每条腿50μL)。

[0545] 对于SAM疫苗,100μL体积中的10μg RNA-LNP复合物作为双侧肌内注射(每条腿50μL)进行施用。

[0546] 脾细胞解离

[0547] 免疫后14天分离脾细胞。将每只小鼠的脾脏汇集于3mL完全RPMI(RPMI、10%FBS、青霉素/链霉素)中。使用gentleMACS解离器(Miltenyi Biotec),遵循制造商的方案进行机械解离。通过40微米过滤器过滤经解离的细胞并用ACK裂解缓冲液(150mM NH_4Cl、10mM $KHCO_3$、0.1mM Na_2EDTA)裂解红细胞。再次经由30微米过滤器过滤细胞,然后使其重悬于完全RPMI中。在Cytoflex LX(Beckman Coulter)上使用碘化丙锭染色对细胞计数以排除死亡和细胞凋亡的细胞。然后将细胞调整至适当活细胞浓度以供后续分析。

[0548] 离体酶联免疫斑点(ELISpot)分析

[0549] ELISPOT分析是根据ELISPOT统一准则{DOI:10.1038/nprot.2015.068},利用小鼠IFNg ELISpotPLUS试剂盒(MABTECH)进行。将1×10^5个脾细胞与10uM指定肽一起在包被有IFNg抗体的96孔板中孵育16小时。使用碱性磷酸酶使斑点显色。对反应定时10分钟并通过用自来水流过板终止反应。使用AID vSpot读取器谱图对斑点计数。对于ELISPOT分析,将饱和度>50%的孔记录为"太多而无法计数"。将重复孔的偏差>10%的样品从分析中排除。然后,使用下式,针对孔汇合校正斑点计数:斑点计数+2x(斑点计数x汇合%/[100%-汇合%])。通过用抗原刺激的孔减去阴性肽刺激孔中的斑点计数来校正阴性背景。最后,将标记为太多而无法计数的孔设定成最高观测校正值,四舍五入至最接近的百分数。

[0550] XIV.C.重复新表位评价结果

[0551] 将具有重复的(即迭代的)新表位的盒的疫苗功效与仅具有给定新表位的单个拷贝的盒进行比较。使用下文所述的ChAdV68递送载体用$8x10^{10}$个VP对经工程化以表达人类HLA-A11:01的小鼠进行免疫,并通过IFNγELISpot评估功效。

[0552] 如图1A所示,设计了一系列ChAdV68递送载体以评估特征为KRAS新表位G12C、G12V、G12D和Q61H的单个拷贝("KRAS 1X(20x1)";盒=SEQ ID NO:63),KRAS G12C、G12V、G12D和Q61H新表位的2条重复序列以及其他KRAS新表位的2条重复序列("KRAS 2X(8x2)";盒=SEQ ID NO:64),或KRAS新表位的4条重复序列("KRAS 4X(4x4)";盒=SEQ ID NO:65)的盒的功效。先前预测HLA-A11:01会呈递KRAS G12C、G12V和G12D新表位(数据未示出)。

[0553] 如图1B所示以及表3所呈现,用特征为新表位的多条重复序列的载体进行疫苗接种展示斑点形成集落(SFC)增加,表明盒中表位编码序列的重复导致针对KRAS新表位G12C的抗原特异性免疫应答增加。然而,在进一步观察到对具有G12V 1x4盒(仅表达G12V)的G12C肽存在应答而对G12C 1x4盒(仅表达G12C)没有应答时,观察到的对具有4x4盒的G12C肽的应答很可能由G12V表位驱动。质谱分析证实G12C肽被G12V肽污染。

[0554] 如图1C和图1D所示以及表3A所呈现,使用ChAdV68递送系统,用特征为新表位的多

条重复序列的载体进行疫苗接种展示斑点形成集落(SFC)增加,表明盒中表位编码序列的重复导致针对KRAS新表位、特别是G12V和G12D新表位的抗原特异性免疫应答增加。

[0555] 表3-KRAS新表位G12C、G12V和G12D+/-重复的ELISpot数据

[0556]

抗原盒	G12C		G12V		G12D	
	平均值	+/- SE	平均值	+/- SE	平均值	+/- SE
20x1	17	7	3	2	5	3
4x4	2277+	580	6169*	31	970	864
8x2	66	31	1563	293	97	45

[0557] *5/6的样品因太多而无法计数(最大值6200)

[0558] +可能是由于G12V污染

[0559] XIV.D.免疫显性表位去除评价结果

[0560] 将具有能够刺激免疫应答的潜在免疫显性TP53新表位的盒的疫苗功效与不具有能够刺激免疫应答的TP53新表位的盒进行比较。使用下文所述的ChAdV68递送载体用$5x10^{10}$个VP对经工程化以表达人类HLA-A11:01的小鼠进行免疫,并通过ELISpot评估功效。先前预测HLA-A11:01会呈递KRAS G12C、G12V、G12D和TP53 S127Y新表位(数据未示出)。

[0561] 如图2A所示,设计了一系列ChAdV68递送载体来评估TP53表位的免疫显性。仅含有KRAS新表位G12C、G12V、G12D和Q61H(“KRAS 4x1”;盒=SEQ ID NO:66),KRAS新表位与TP53 R213L新表位的组合的载体被认为无法在HLA-A11:01小鼠模型中刺激免疫应答,因为它被预测由人类HLA-A02:01而不是HLA-A11:01(“KRAS4x1+R213L”;盒=SEQ ID NO:67)呈递,或者KRAS新表位与TP53S127Y新表位的组合预测由HLA-A11:01(“KRAS 4x1+S127Y”;盒=SEQ ID NO:68)呈递。

[0562] 如图2B所示以及表4所呈现,相对于不包括TP53新表位(“4x1”)或包括非免疫原性TP53新表位(“4x1+R213L”)的载体,用包括免疫原性TP53 S127Y表位的载体进行疫苗接种展示KRAS表位G12C的斑点形成集落(SFC)减少,表明疫苗盒中TP53 S127Y新表位的存在充当免疫显性新表位,从而降低了KRAS特异性T细胞应答。然而,在进一步观察到对具有G12V 1x4盒(仅表达G12V)的G12C肽有应答而对G12C 1x4盒(仅表达G12C)没有应答时,观察到的对具有4x4盒的G12C肽的应答可能由G12V表位驱动。质谱分析证实G12C肽被G12V肽污染。

[0563] 如图2C-E所示以及表4所呈现,相对于不包括TP53新表位(“4x1”)或包括非免疫原性TP53新表位(“4x1+R213L”)的载体,用包括免疫原性TP53 S127Y表位的载体进行疫苗接种展示KRAS表位G12V和G12D的斑点形成集落(SFC)减少,表明疫苗盒中TP53 S127Y新表位的存在充当免疫显性新表位,从而降低了KRAS特异性T细胞应答。

[0564] 表4-KRAS和TP53新表位的ELISpot数据

[0565]

抗原盒	G12C		G12D		G12V		R213L		S127Y	
	平均值	+/-SE	平均值	+/-SE	平均值	+/-SE	平均值	+/-SE	平均值	+/-SE
4x1	92	41	32	21	1570	833	ND	ND	ND	ND
4x1+R213L	86	74	62	46	773	378	5	3	ND	ND
4x1+S127Y	2	1	7	5	0	0	ND	ND	6118*	264

[0566] ND＝未测定;*3/6样品因太多而无法计数(最大值6200)

[0567] XIV.E.通过质谱评价KRAS新表位的呈递

[0568] 使用靶向质谱方法对候选KRAS表位进行质谱(MS)验证。评估了供体肿瘤切除(多等位基因;来自EDGE预测的等位基因分配)或经工程化以表达指定盒的单个等位基因K562系。Gillete等人(Nat Methods.2013年1月;10(1):28-34)更详细地描述了质谱分析方法,所述文献出于所有目的通过引用整体并入本文。国际申请公开WO2021092095中描述了靶向质谱分析,包括使用K562单等位基因体外系统,所述文献出于所有目的通过引用并入本文。结果显示在表5中,证实了各种KRAS新表位对HLA特异性呈递的确认。

[0569] 表5-KRAS新表位质谱

[0570]

突变	HLA	序列	SEQ ID NO:	肿瘤	20x1 K562细胞系	4x4 K562细胞系	8x2 K562细胞系
G12C	A*03:01	VVVGACGVGK	75	否	是	NA	NA
G12C	A*11:01	VVVGACGVGK	75	否	是	是	否
G12D	A*03:01	VVGADGVGK	77	否	是	NA	NA
G12D	A*03:01	VVVGADGVGK	78	否	是	NA	NA
G12D	A*11:01	VVGADGVGK	77	是	是	是	否
G12D	A*11:01	VVVGADGVGK	78	是	是	是	是

[0571]

G12V	A*02:01	KLVVVGAVGV	90	否	否	是	否
G12V	A*03:01	VVGAVGVGK	79	否	是	NA	NA
G12V	A*03:01	VVVGAVGVGK	81	是	是	NA	NA
G12V	A*11:01	VVGAVGVGK	79	是	是	是	是
G12V	A*11:01	VVVGAVGVGK	81	是	是	是	是
G12V	C*01:02	AVGVGKSAL	80	是	NA	NA	NA
Q61H	A*01:01	ILDTAGHEEY	82	NA	是	是	是

[0572] NA＝不适用(肿瘤或细胞系不可用/无法生成以进行测试)

[0573] XIV.F.迭代的KRAS新表位盒评价结果

[0574] 评价了具有重复的(即迭代的)新表位或单个拷贝的给定新表位的各种盒的疫苗功效。使用下文所述的ChAdV68递送载体用$5x10^{10}$个VP对经工程化以表达KRAS G12突变的人类HLA-A11:01或KRAS Q61H的人类HLA-A01:01的小鼠进行免疫,并通过IFNγELISpot评估功效。

[0575] 设计了一系列ChAdV68递送载体以评估特征为KRAS新表位G12C、G12V、G12D和Q61H的单个拷贝(“KRAS 4x1”;盒＝SEQ ID NO:66),KRAS新表位G12C、G12V、G12D和Q61H中的每一者的4次迭代(“KRAS 4X(4x4)”;盒＝SEQ ID NO:65),或KRAS新表位G12C、G12V、G12D或Q61H中的仅一者的4次迭代(“KRAS 1x4”)的盒的功效。

[0576] 比较了特征为单个拷贝的KRAS新表位(“4x1”)或KRAS新表位中的每一者的4次迭代(“4x4”)的盒。如图3所示并在表6中量化,使用ChAdV68递送系统,用特征为新表位多次迭

代的载体进行疫苗接种展示斑点形成集落(SFC)增加,表明盒中表位编码序列的重复导致针对KRAS新表位、特别是G12V和G12D新表位的抗原特异性免疫应答增加。

[0577] 比较了特征为KRAS新表位中的每一者的4次迭代(“4x4”)或KRAS中仅一者的4次迭代(“1x4”)的盒。如图4所示并在表7中量化,使用ChAdV68递送系统,用特征为4x4或1x4格式的新表位的多次迭代的载体进行疫苗接种通过斑点形成集落(SFC)产生免疫应答,并且在格式之间具有可比性,表明盒中表位编码序列的重复导致针对KRAS新表位、特别是G12V和G12D新表位的抗原特异性免疫应答增加。

[0578] 比较了KRAS Q61H盒格式的免疫应答,包括特征为KRAS Q61H的单个拷贝与19个其他不同新表位(“20x1”)、KRAS Q61H的2次迭代(“8x2”)、KRAS Q61H的单个拷贝和KRAS G12新表位的单个拷贝(“4x1”)、KRAS Q61H的4次迭代和KRAS G12新表位迭代(“4x4”);或仅KRAS Q61H的4次迭代(“1x4”)的盒。如图5所示并在表8中量化,使用ChAdV68递送系统,虽然用KRAS Q61H的单个拷贝和KRAS G12新表位的单个拷贝的载体进行疫苗接种展示相比于较大盒的小改进(4x1对20x1或8x2),但用特征为4x4或1x4格式的新表位多次迭代的载体进行疫苗接种通过斑点形成集落(SFC)产生免疫应答,其中1x4格式展示出最稳健的应答,表明盒中表位编码序列的重复以及较短的格式和/或去除额外的不同新表位导致针对KRAS新表位Q61H的抗原特异性免疫应答增加。

[0579] 我们比较了使用ChAdV68递送系统或SAM递送系统进行免疫后的应答。评估了特征为KRAS新表位的单个拷贝以及其他新表位(“20x1”)或KRAS新表位中的每一者的4次迭代(“4x4”)的盒。如图6和图7所示并在表9中量化,使用ChAdV68或SAM递送系统,用特征为新表位多次迭代的载体进行疫苗接种展示,ChAdV68和SAM格式的斑点形成集落(SFC)增加,表明盒中表位编码序列的重复导致针对KRAS新表位、特别是G12V和G12D新表位的抗原特异性免疫应答增加,表明使用特征为新表位迭代的盒的应答改进与所使用的载体平台无关。

[0580] 表6-KRAS新表位G12V和G12D+/-重复的ELISpot数据

[0581]

抗原盒	G12V		G12D	
	(VVVGAVGGK(SEQ ID NO: 91)))		(VVVGADGVGK(SEQ ID NO: 78)))	
	平均值	+/- SE	平均值	+/- SE
4x4	5632	237.9	1069	361.4
4x1	1570	833.2	32	20.6

[0582]

[0583] 表7-KRAS新表位G12V和G12D+/-重复的ELISpot数据

[0584]

抗原盒	G12V (VVVGAVGGK(SEQ ID NO: 91))		G12D (VVVGADGVGK(SEQ ID NO: 78))*	
	平均值	+/- SE	平均值	+/- SE
4x4	3700	404.9	536	316.1
1x4	3358	310.8	1726	784.7

[0585] *结果在格式之间没有显著差异(p=0.1896)

[0586] 表8-KRAS新表位Q61H+/-重复的ELISpot数据

[0587]

抗原盒	Q61H (ILDTAGHEEY(SEQ ID NO: 82))	
	平均值	+/- SE
20x1	3	1.3
4x4	554	225
8x2	20	10.8
4x1	262	112.2
1x4	1289	451.7
初始	1	0.6

[0588] 表9-呈Chad/SAM格式的KRAS G12V和G12D+/-重复的ELISpot

[0589]

	ChAdV68 20x1			SAM 20x1			ChAdV68 4x4			SAM 4x4		
	平均值	SEM	N	平均值	SEM	N	平均值	SEM	N	平均值	SEM	N
G12V 池	13	6	6	33	6	6	4483	767	6	6939	1149	6
G12D 池	5	2	6	6	3	6	297	227	6	560	163	6

[0590]

G12V (VVGAVGVGK(SEQ ID NO: 79))	4	2	6	12	3	6	3167	846	6	6878	1216	6
G12D (VVVGADGVGK(SEQ ID NO: 78))	4	3	6	8	3	6	356	288	6	564	148	6

[0591] 某些序列

[0592] 本文提及的载体、盒和抗体描述于下文并且由SEQ ID NO提及。

[0593]

曲美木单抗 VL (SEQ ID NO:16)
曲美木单抗 VH (SEQ ID NO:17)
曲美木单抗 VH CDR1 (SEQ ID NO:18)
曲美木单抗 VH CDR2 (SEQ ID NO:19)
曲美木单抗 VH CDR3 (SEQ ID NO:20)
曲美木单抗 VL CDR1 (SEQ ID NO:21)
曲美木单抗 VL CDR2 (SEQ ID NO:22)
曲美木单抗 VL CDR3 (SEQ ID NO:23)
德瓦鲁单抗(MEDI4736) VL (SEQ ID NO:24)
MEDI4736 VH (SEQ ID NO:25)
MEDI4736 VH CDR1 (SEQ ID NO:26)
MEDI4736 VH CDR2 (SEQ ID NO:27)
MEDI4736 VH CDR3 (SEQ ID NO:28)
MEDI4736 VL CDR1 (SEQ ID NO:29)
MEDI4736 VL CDR2 (SEQ ID NO:30)
MEDI4736 VL CDR3 (SEQ ID NO:31)
UbA76-25 聚体 PDTT 核苷酸(SEQ ID NO:32)
UbA76-25 聚体 PDTT 多肽(SEQ ID NO:33)
MAG-25 聚体 PDTT 核苷酸(SEQ ID NO:34)
MAG-25 聚体 PDTT 多肽(SEQ ID NO:35)
Ub7625 聚体 PDTT_NoSFL 核苷酸(SEQ ID NO:36)
Ub7625 聚体 PDTT_NoSFL 多肽(SEQ ID NO:37)
ChAdV68.5WTnt.MAG25 聚体(SEQ ID NO:2) 具有 E1 (nt 577 至 3403) 和 E3 (nt 27,125- 31,825)序列缺失的 AC_000011.1; 在五个位置处被取代的相应 ATCC VR-594 核苷酸; 被插入以代替缺失 E1 的 CMV 启动子/增强子控制下的模型新抗原盒; 盒的 SV40 聚 A 3'
委内瑞拉马脑炎病毒[VEE] (SEQ ID NO:3) GenBank: L01442.2

[0594]

VEE-MAG25 聚体(SEQ ID NO:4) 含有 MAG-25 聚体 PDTT 核苷酸(碱基 30-1755)
委内瑞拉马脑炎病毒株 TC-83 [TC-83](SEQ ID NO:5) GenBank: L01443.1
VEE 递送载体(SEQ ID NO:6); 具有核苷酸 7544-11175 缺失的 VEE 基因组[甲病毒结构蛋白被去除]
TC-83 递送载体(SEQ ID NO:7); 具有核苷酸 7544-11175 缺失的 TC-83 基因组[甲病毒结构蛋白被去除]
VEE 生产载体(SEQ ID NO:8); 具有核苷酸 7544-11175 缺失的 VEE 基因组, 加 5' T7-启动子, 加 3'限制性位点
TC-83 生产载体(SEQ ID NO:9); 具有核苷酸 7544-11175 缺失的 TC-83 基因组, 加 5' T7-启动子, 加 3'限制性位点
VEE-UbAAY (SEQ ID NO:14); 具有 MHC I 类小鼠肿瘤表位 SIINFEKL(SEQ ID NO: 92)和 AH1-A5 插入的 VEE 递送载体
VEE-荧光素酶(SEQ ID NO: 15); 具有插入在 7545 处的荧光素酶基因的 VEE 递送载体
泛素(SEQ ID NO:38)>UbG76 0-228
泛素 A76 (SEQ ID NO:39)>UbA76 0-228
HLA-A2 (MHC I 类)信号肽(SEQ ID NO:40)>MHC 信号肽 0-78
HLA-A2 (MHC I 类)跨膜结构域(SEQ ID NO:41)>HLA A2 TM 结构域 0-201
IgK 前导序列(SEQ ID NO:42)>IgK 前导序列 0-60
人类 DC-Lamp (SEQ ID NO:43)>人类 DCLAMP 0-3178
小鼠 LAMP1 (SEQ ID NO:44)>小鼠 Lamp1 0-1858
人类 Lamp1 cDNA (SEQ ID NO:45)>人类 Lamp1 0-2339
破伤风类毒素核酸序列(SEQ ID NO:46)
破伤风类毒素氨基酸序列(SEQ ID NO:47)
PADRE 核苷酸序列(SEQ ID NO:48)
PADRE 氨基酸序列(SEQ ID NO:49)
WPRE (SEQ ID NO:50)>WPRE 0-593
IRES (SEQ ID NO:51)>eGFP_IRES_SEAP_插入 1746-2335
GFP (SEQ ID NO:52)
SEAP (SEQ ID NO:53)
萤火虫荧光素酶(SEQ ID NO:54)
FMDV 2A (SEQ ID NO:55)
GPGPG 接头(SEQ ID NO:56)
1X (20x1)盒核苷酸序列(SEQ ID NO: 69)
ATGGCTGGCATGACCGAGTACAAACTTGTGGTGGTTGGTGCTG

[0595]

GAGACGTGGGGAAAAGCGCCCTGACAATTCAGCTTATTCAGGA AACTTGCTTATTAGACATATTAGATACTGCCGGAAAAGAAGAGT ATTCTGCTATGAGAGATCAGTACATGCGGATGTGCAACAGCAGC TGCATGGGCGGCATGAACCGCATGCCCATCTTGACTATTATCACC CTGGAGGATTCCTCCAGTGGAATACATTCGGGAGCCACCACCA CTGCTCCTCCCCTGTCAGGTAAGGGAAATCCTGAAGAAGAAGA TGTTAGCGGCATTCATTCCGGTGCTACGACAACTGCACCATTTC TGTCTGGTAAAGGTAATCCAGAGGAGGAGGATGTGGAAATTCT GGATGAGGCCTACGTCATGGCCTATGTGATGGCTGGCGTTGGAA GCCCATACGTGAGTAGACTGCTGATGACAGAATATAAATTAGTG GTAGTAGGAGCAGATGGGGTAGGAAAGAGTGCCTTAACAATCC AACTCATTCAGGAGACCTGTTTGTTGGATATTCTGGATACAGCT GGCCGGGAAGAGTATTCGGCAATGAGGGACCAATATATGCGATC TTACCTCGACAGCGGAATCCACTCCGGTGCCACTGCAACAGCG CCGAGCCTGAGTGGTAAGGGGAACCCTGAGAAGTCAGTAACGT GTACATACAGCCCTGCTTTAAACAACATGTTTTGCCAACTTGCA AAAACATGTCCTGTACAGATGACAGAGTATAAACTTGTAGTGGT GGGGGCAGCTGGTGTAGGTAAAAGTGCACTTACCATCCAGCTG ATTCAAGAGACATGTCTGTTAGACATCTTGGACACAGCCGGGCT CGAGGAATATAGCGCCATGAGAGACCAGTACATGCGACTGCGC GTGGAGTATCTGGATGACAGGAACACCTTCCTGCATTCTGTTGT GGTGCCCTATGAGCCCCCGGAGGTGGATGGACAGATCACAGTG GGCCAGAGAATTGGCAGCGTGTCCTTCGGGACTGTGTACAAGG GCAAGTGGCACGGAATGACGGAATATAAGCTGGTGGTGGTAGG AGCAGTAGGTGTCGGTAAATCAGCCCTCACCATTCAGCTTATAC AGGAGACTTGTTTGCTGGACATCCTAGACACTGCTGGCCATGA GGAGTACTCCGCCATGCGCGATCAATACATGAGGTGGCAGCAG CAGAGCTACCTGGACTCAGGGATCCACTTTGGAGCTACCACAA CTGCGCCATCACTAAGTGGGAAACACAGCGGCACCGCCAAATC TGTCACGTGTACTTACTACCCAGCCTTGAACAAGATGTTCTGTC AATTAGCAAAGAAAAGTGTCACCTGCACATATTCACCGGCACTG AATGAGATGTTCTGCCAGCTGGCTAAGACCTGTCCTGTGCAGAT GACTGAGTACAAACTGGTCGTTGTGGGGGCCTGTGGGGTTGGG AAGAGCGCTCTGACCATCCAGCTCATCCAGGGACCCGGACCAG GCGCCAAATTTGTTGCTGCTTGGACACTGAAAGCTGCTGCTGG GCCCGGACCAGGCCAGTACATCAAGGCCAACTCTAAGTTTATCG GCATCACCGAATTGGGACCTGGACCCGGCTAGTAGTGA
KRAS 2X (8x2)盒核苷酸序列(SEQ ID NO: 70)
ATGGCTGGCATGACTGAATACAAATTAGTTGTCGTCGGTGCATG CGGGGTAGGCAAGTCGGCCTTAACGATACAACTTATTCAAGAG ACATGTCTACTGGATATCCTTGACACTGCCGGTCACGAAGAATAT AGTGCAATGAGGGACCAGTACATGCGTATGACGGAATATAAGTT AGTAGTAGTTGGCGCAGGCGACGTAGGGAAATCCGCTCTAACTA

[0596]

TCCAACTAATTCAAGAAACGTGCCTACTAGATATATTAGACACA GCGGGACGTGAGGAGTATAGCGCCATGCGTGACCAGTATATGCG GATGACCGAGTATAAGCTAGTCGTAGTCGGCGCGGATGGGGTTG GTAAGAGCGCCTTGACCATCCAGCTCATACAAGAAACTTGCCTT CTGGACATCCTAGACACCGCGGGTCTCGAAGAGTACAGTGCGA TGAGAGATCAGTACATGAGGATGACAGAATACAAGCTCGTTGTT GTTGGTGCGGTCGGTGTTGGAAAGAGTGCGCTAACCATTCAGC TTATCCAGGAAACCTGTCTGTTAGACATCTTAGATACCGCAGGT AAAGAGGAATATTCGGCCATGAGGGATCAATATATGCGAATGAC AGAGTATAAATTGGTAGTGGTAGGGGCTTGCGGAGTGGGGAAA AGCGCATTGACTATACAATTGATTCAGGAAACATGCCTATTGGA CATACTCGACACGGCCGGGAAAGAAGAGTATTCCGCGATGCGA GATCAATACATGCGCATGACCGAATATAAACTTGTCGTTGTCGG AGCGGGTGATGTAGGTAAATCGGCGCTCACAATCCAATTAATCC AAGAGACGTGCTTGCTAGACATTCTGGATACAGCTGGGCACGA GGAGTACTCAGCTATGCGCGATCAGTATATGAGGATGACTGAGT ACAAGTTGGTCGTCGTAGGAGCGGTTGGTGTCGGAAAATCTGC GTTGACAATTCAACTGATACAAGAGACTTGTTTGTTAGATATTCT CGATACTGCGGGTCGGGAAGAATACTCGGCTATGAGAGACCAAT ATATGAGAATGACGGAGTACAAACTCGTAGTTGTAGGTGCGGAC GGTGTAGGAAAGTCTGCGCTTACGATTCAGTTGATACAGGAGA CCTGTTTGCTCGATATCTTGGATACGGCGGGTTTGGAGGAATAC AGCGCAATGCGGGACCAATACATGAGAGGACCCGGACCAGGCG CCAAATTTGTTGCTGCTTGGACACTGAAAGCTGCTGCTGGGCCC GGACCAGGCCAGTACATCAAGGCCAACTCTAAGTTTATCGGCAT CACCGAATTGGGACCTGGACCCGGCTAG
KRAS 4X (4x4)盒核苷酸序列(SEQ ID NO: 71)
ATGGCTGGCATGACCGAGTATAAACTAGTAGTTGTGGGAGCGTG TGGTGTAGGCAAGTCGGCACTTACAATTCAGTTGATACAAATGA CGGAATATAAGCTCGTAGTAGTCGGAGCAGACGGCGTGGGGAA ATCAGCGTTGACTATCCAGTTAATACAGGAAACTTGCCTATTAG ACATCTTGGATACGGCAGGTCATGAGGAATATTCCGCTATGAGA GATCAGTATATGCGCATGACGGAGTATAAGCTTGTGGTTGTCGG GGCCGACGGGGTAGGTAAGTCAGCGCTCACGATACAATTAATTC AAATGACCGAATACAAGTTGGTCGTGGTGGGGGCAGTTGGGGT CGGTAAATCCGCGTTAACGATCCAACTTATCCAAATGACAGAAT ATAAACTCGTTGTTGTAGGTGCATGTGGCGTAGGAAAAAGCGCA TTGACCATCCAGCTAATTCAGGAGACGTGTCTCCTTGATATCCTA GACACGGCGGGGCACGAAGAATACTCGGCTATGCGCGACCAGT ACATGAGAATGACGGAATACAAACTTGTTGTCGTGGGTGCGGA TGGAGTAGGGAAAAGTGCTCTAACAATACAACTCATTCAGATG ACAGAGTACAAATTGGTAGTCGTCGGTGCGGTAGGAGTTGGGA AGTCTGCACTAACTATTCAGCTCATACAGATGACCGAGTACAAG

[0597]

CTGGTGGTGGTAGGCGCTTGCGGTGTGGGTAAGAGTGCATTAA CCATACAGCTTATACAAGAGACATGTCTGCTAGATATATTAGATA CCGCCGGGCATGAAGAGTACTCTGCCATGCGAGACCAATACATG CGTATGACAGAGTATAAATTAGTAGTGGTTGGGGCGGACGGTGT TGGCAAGAGCGCCTTAACTATACAGTTGATCCAGATGACGGAGT ACAAACTGGTCGTCGTTGGTGCAGTGGGAGTGGGAAAATCTGC GCTGACGATTCAACTAATCCAAGAAACATGTTTACTTGACATCC TCGACACTGCGGGTCACGAGGAGTATTCGGCGATGCGTGATCA ATATATGAGGATGACTGAGTATAAGTTAGTCGTAGTTGGAGCGG TCGGTGTCGGAAAGTCCGCGCTAACCATTCAATTGATTCAAATG ACTGAATACAAGCTAGTGGTAGTAGGAGCATGCGGCGTCGGCA AATCGGCTTTAACAATCCAACTGATACAGGGACCCGGACCAGG CGCCAAATTTGTTGCTGCTTGGACACTGAAAGCTGCTGCTGGG CCCGGACCAGGCCAGTACATCAAGGCCAACTCTAAGTTTATCGG CATCACCGAATTGGGACCTGGACCCGGCTAG
KRAS 4x1 盒核苷酸序列(SEQ ID NO: 72)
ATGGCTGGCATGACCGAGTATAAACTAGTAGTTGTGGGAGCGTG TGGTGTAGGCAAGTCGGCACTTACAATTCAGTTGATACAAATGA CGGAATATAAGCTCGTAGTAGTCGGAGCAGACGGCGTGGGGAA ATCAGCGTTGACTATCCAGTTAATACAGGAAACTTGCCTATTAG ACATCTTGGATACGGCAGGTCATGAGGAATATTCCGCTATGAGA GATCAGTATATGCGCATGACCGAATACAAGTTGGTCGTGGTGGG GGCAGTTGGGGTCGGTAAATCCGCGTTAACGATCCAACTTATCC AAGGACCCGGACCAGGCGCCAAATTTGTTGCTGCTTGGACACT GAAAGCTGCTGCTGGGCCCGGACCAGGCCAGTACATCAAGGCC AACTCTAAGTTTATCGGCATCACCGAATTGGGACCTGGACCCGG CTAG
KRAS 4x1 + R213L 盒核苷酸序列(SEQ ID NO: 73)
ATGGCTGGCATGACCGAGTATAAACTAGTAGTTGTGGGAGCGTG TGGTGTAGGCAAGTCGGCACTTACAATTCAGTTGATACAAATGA CGGAATATAAGCTCGTAGTAGTCGGAGCAGACGGCGTGGGGAA ATCAGCGTTGACTATCCAGTTAATACAGGAAACTTGCCTATTAG ACATCTTGGATACGGCAGGTCATGAGGAATATTCCGCTATGAGA GATCAGTATATGCGCATGACCGAATACAAGTTGGTCGTGGTGGG GGCAGTTGGGGTCGGTAAATCCGCGTTAACGATCCAACTTATCC AACTGCGCGTGGAGTATCTGGATGACAGGAACACCTTCCTGCAT TCTGTTGTGGTGCCCTATGAGCCCCCGGAGGTGGGACCCGGAC CAGGCGCCAAATTTGTTGCTGCTTGGACACTGAAAGCTGCTGC TGGGCCCGGACCAGGCCAGTACATCAAGGCCAACTCTAAGTTT ATCGGCATCACCGAATTGGGACCTGGACCCGGCTAG
KRAS 4x1 + S127Y 盒核苷酸序列(SEQ ID NO: 74)
ATGGCTGGCATGACCGAGTATAAACTAGTAGTTGTGGGAGCGTG

[0598]

TGGTGTAGGCAAGTCGGCACTTACAATTCAGTTGATACAAATGA CGGAATATAAGCTCGTAGTAGTCGGAGCAGACGGCGTGGGGAA ATCAGCGTTGACTATCCAGTTAATACAGGAAACTTGCCTATTAG ACATCTTGGATACGGCAGGTCATGAGGAATATTCCGCTATGAGA GATCAGTATATGCGCATGACCGAATACAAGTTGGTCGTGGTGGG GGCAGTTGGGGTCGGTAAATCCGCGTTAACGATCCAACTTATCC AACACAGCGGCACCGCCAAATCTGTCACGTGTACTTACTACCCA GCCTTGAACAAGATGTTCTGTCAATTAGCAAAGGGACCCGGAC CAGGCGCCAAATTTGTTGCTGCTTGGACACTGAAAGCTGCTGC TGGGCCCGGACCAGGCCAGTACATCAAGGCCAACTCTAAGTTT ATCGGCATCACCGAATTGGGACCTGGACCCGGCTAG
KRAS G12C 新表位(SEQ ID NO: 75)
VVVGACGVGK
KRAS G12C 新表位(SEQ ID NO: 76)
KLVVVGACGV
KRAS G12D 新表位(SEQ ID NO: 77)
VVGADGVGK
KRAS G12D 新表位(SEQ ID NO: 78)
VVVGADGVGK
KRAS G12V 新表位(SEQ ID NO: 79)
VVGAVGVGK
KRAS G12V 新表位(SEQ ID NO: 80)
AVGVGKSAL
KRAS G12V 新表位(SEQ ID NO: 81)
VVVGAVGVGK
KRAS Q61H 新表位(SEQ ID NO: 82)
ILDTAGHEEY

[0599] 参考文献

[0600] 1.Desrichard,A.,Snyder,A.&Chan,T.A.Cancer Neoantigens and Applications for Immunotherapy.Clin.Cancer Res.off.J.Am.Assoc.Cancer Res.(2015).doi:10.1158/1078-0432.CCR-14-3175

[0601] 2.Schumacher,T.N.&Schreiber,R.D.Neoantigens in cancer immunotherapy.Science 348,69-74(2015).

[0602] 3.Gubin,M.M.,Artyomov,M.N.,Mardis,E.R.&Schreiber,R.D.Tumor neoantigens:building a framework for personalized cancer immunotherapy.J.Clin.Invest.125,3413-3421(2015).

[0603] 4.Rizvi,N.A.et al.Cancer immunology.Mutational landscape determines sensitivity to PD-1 blockade in non-small cell lung cancer.Science 348,124-128(2015).

[0604] 5.Snyder,A.et al.Genetic basis for clinical response to CTLA-4 blockade in melanoma.N.Engl.J.Med.371,2189-2199(2014).

[0605] 6.Carreno,B.M.et al.Cancer immunotherapy.A dendritic cell vaccine increases the breadth and diversity of melanoma neoantigen-specific T cells.Science 348,803-808(2015).

[0606] 7.Tran,E.et al.Cancer immunotherapy based on mutation-specific CD4+ T cells in a patient with epithelial cancer.Science 344,641-645(2014).

[0607] 8.Hacohen,N.&Wu,C.J.-Y.United States Patent Application:20110293637-COMPOSITIONS AND METHODS OF IDENTIFYING TUMOR SPECIFIC NEOANTIGENS.(A1).at<http://appft1.uspto.gov/netacgi/nph-Parser?Sect1＝PTO1&Sect2＝HITOFF&d＝PG01&p＝1&u＝/netahtml/PTO/srchnum.html&r＝1&f＝G&l＝50&s1＝20110293637.PGNR.＞

[0608] 9.Lundegaard,C.,Hoof,I.,Lund,O.&Nielsen,M.State of the art and challenges in sequence based T-cell epitope prediction.Immunome Res.6 Suppl 2,S3(2010).

[0609] 10.Yadav,M.et al.Predicting immunogenic tumour mutations by combining mass spectrometry and exome sequencing.Nature 515,572-576(2014).

[0610] 11.Bassani-Sternberg,M.,Pletscher-Frankild,S.,Jensen,L.J.&Mann,M.Massspectrometry of human leukocyte antigen class I peptidomes reveals strong effects of protein abundance and turnover on antigen presentation.Mol.Cell.Proteomics MCP 14,658-673(2015).

[0611] 12.Van Allen,E.M.et al.Genomic correlates of response to CTLA-4 blockade in metastatic melanoma.Science 350,207-211(2015).

[0612] 13.Yoshida,K.&Ogawa,S.Splicing factor mutations and cancer.Wiley Interdiscip.Rev.RNA 5,445-459(2014).

[0613] 14.Cancer Genome Atlas Research Network.Comprehensive molecular profiling of lung adenocarcinoma.Nature 511,543-550(2014).

[0614] 15.Rajasagi,M.et al.Systematic identification of personal tumor-specific neoantigens in chronic lymphocytic leukemia.Blood 124,453-462(2014).

[0615] 16.Downing,S.R.et al.United States Patent Application:0120208706-OPTIMIZATION OF MULTIGENE ANALYSIS OF TUMOR SAMPLES.(A1).at<http://appft1.uspto.gov/netacgi/nph-Parser?Sect1＝PTO1&Sect2＝HITOFF&d＝PG01&p＝1&u＝/netahtml/PTO/srchnum.html&r＝1&f＝G&l＝50&s1＝20120208706.PGNR.＞

[0616] 17.Target Capture for NextGen Sequencing-IDT.at<http://www.idtdna.com/pages/products/nextgen/target-capture>

[0617] 18.Shukla,S.A.et al.Comprehensive analysis of cancer-associated somatic mutations in class I HLA genes.Nat.Biotechnol.33,1152-1158(2015).

[0618] 19.Cieslik,M.et al.The use of exome capture RNA-seq for highly degraded RNA with application to clinical cancer sequencing.Genome Res.25,1372-1381(2015).

[0619] 20.Bodini,M.et al.The hidden genomic landscape of acute myeloid leukemia:subclonal structure revealed by undetected mutations.Blood 125,600-

605(2015).

[0620] 21.Saunders,C.T.et al.Strelka:accurate somatic small-variant calling from sequenced tumor-normal sample pairs.Bioinforma.Oxf.Engl.28,1811-1817(2012).

[0621] 22.Cibulskis,K.et al.Sensitive detection of somatic point mutations in impure and heterogeneous cancer samples.Nat.Biotechnol.31,213-219(2013).

[0622] 23.Wilkerson,M.D.et al.Integrated RNA and DNA sequencing improves mutation detection in low purity rumors.Nucleic Acids Res.42,e107(2014).

[0623] 24.Mose,L.E.,Wilkerson,M.D.,Hayes,D.N.,Perou,C.M.&Parker,J.S.ABRA:improved coding indel detection via assembly-based realignment.Bioinforma.Oxf.Engl.30,2813-2815(2014).

[0624] 25.Ye,K.,Schulz,M.H.,Long,Q.,Apweiler,R.&Ning,Z.Pindel:a pattern growth approach to detect break points of large deletions and medium sized insertions from paired-end short reads.Bioinforma.Oxf.Engl.25,2865-2871(2009).

[0625] 26.Lam,H.Y.K.et al.Nucleotide-resolution analysis of structural variants using BreakSeq and a breakpoint library.Nat.Biotechnol.28,47-55(2010).

[0626] 27.Frampton,G.M.et al.Development and validation of a clinical cancer genomic profiling test based on massively parallel DNA sequencing.Nat.Biotechnol.31,1023-1031(2013).

[0627] 28.Boegel,S.et al.HLA typing from RNA-Seq sequence reads.Genome Med.4,102(2012).

[0628] 29.Liu,C.et al.ATHLATES:accurate typing of human leukocyte antigen through exome sequencing.Nucleic Acids Res.41,e 142(2013).

[0629] 30.Mayor,N.P.et al.HLA Typing for the Next Generation.PloS One 10,e0127153(2015).

[0630] 31.Roy,C.K.,Olson,S.,Graveley,B.R.,Zamore,P.D.&Moore,M.J.Assessing long-distance RNA sequence connectivity via RNA-templated DNA-DNA ligation.eLife 4,(2015).

[0631] 32.Song,L.&Florea,L.CLASS:constrained transcript assembly of RNA-seq reads.BMC Bioinformatics 14 Suppl 5,S14(2013).

[0632] 33.Maretty,L.,Sibbesen,J.A.&Krogh,A.Bayesian transcriptome assembly.Genome Biol.15,501(2014).

[0633] 34.Pertea,M.et al.StringTie enables improved reconstruction of a transcriptome from RNA-seq reads.Nat.Biotechnol.33,290-295(2015).

[0634] 35.Roberts,A.,Pimentel,H.,Trapnell,C.&Pachter,L.Identification of novel transcripts in annotated genomes using RNA-Seq.Bioinforma.Oxf.Engl.(2011).doi:10.1093/bioinformatics/btr355

[0635] 36.Vitting-Seerup,K.,Porse,B.T.,Sandelin,A.&Waage,J.spliceR:an R package for classification of alternative splicing and prediction of coding potential from RNA-seq data.BMC Bioinformatics 15,81(2014).

[0636] 37.Rivas,M.A.et al.Human genomics.Effect of predicted protein-truncating genetic variants on the human transcriptome.Science 348,666-669(2015).

[0637] 38.Skelly,D.A.,Johansson,M.,Madeoy,J.,Wakefield,J.&Akey,J.M.A powerful and flexible statistical framework for testing hypotheses of allele-specific gene expression from RNA-seq data.Genome Res.21,1728-1737(2011).

[0638] 39.Anders,S.,Pyl,P.T.&Huber,W.HTSeq--a Python framework to work with high-throughput sequencing data.Bioinforma.Oxf.Engl.31,166-169(2015).

[0639] 40.Furney,S.J.et al.SF3B1 mutations are associated with alternative splicing in uveal melanoma.Cancer Discov.(2013).doi:10.1158/2159-8290.CD-13-0330

[0640] 41.Zhou,Q.et al.A chemical genetics approach for the functional assessment of novel cancer genes.Cancer Res.(2015).doi:10.1158/0008-5472.CAN-14-2930

[0641] 42.Maguire,S.L.et al.SF3B1 mutations constitute a novel therapeutic target in breast cancer.J.Pathol.235,571-580(2015).

[0642] 43.Carithers,L.J.et al.A Novel Approach to High-Quality Postmortem Tissue Procurement:The GTEx Project.Biopreservation Biobanking 13,311-319(2015).

[0643] 44.Xu,G.et al.RNA CoMPASS:a dual approach for pathogen and host transcriptome analysis of RNA-seq datasets.PloS One 9,e89445(2014).

[0644] 45.Andreatta,M.&Nielsen,M.Gapped sequence alignment using artificial neural networks:application to the MHC class I system.Bioinforma.Oxf.Engl.(2015).doi:10.1093/bioinformatics/btv639

[0645] 46.Jørgensen,K.W.,Rasmussen,M.,Buus,S.&Nielsen,M.NetMHCstab-predicting stability of peptide-MHC-I complexes;impacts for cytotoxic T lymphocyte epitope discovery.Immunology 141,18-26(2014).

[0646] 47.Larsen,M.V.et al.An integrative approach to CTL epitope prediction:a combined algorithm integrating MHC class I binding,TAP transport efficiency,and proteasomal cleavage predictions.Eur.J.Immunol.35,2295-2303(2005).

[0647] 48.Nielsen,M.,Lundegaard,C.,Lund,O.&Keşmir,C.The role of the proteasome in generating cytotoxic T-cell epitopes:insights obtained from improved predictions of proteasomal cleavage,Immunogenetics 57,33-41(2005).

[0648] 49.Boisvert,F.-M.et al.A Quantitative Spatial Proteomics Analysis of Proteome Turnover in Human Cells.Mol.Cell.Proteomics 11,M111.011429-

M111.011429(2012).

[0649] 50.Duan,F.et al.Genomic and bioinformatic profiling of mutational neoepitopes reveals new rules to predict anticancer immunogenicity.J.Exp.Med.211,2231-2248(2014).

[0650] 51.Janeway's Immunobiology:9780815345312:Medicine&Health Science Books@Amazon.com.at<http://www.amazon.com/Janeways-Immunobiology-Kenneth-Murphy/dp/0815345313>

[0651] 52.Calis,J.J.A.et al.Properties of MHC Class I Presented Peptides That Enhance Immunogenicity.PLoS Comput.Biol.9,e 1003266(2013).

[0652] 53.Zhang,J.et al.Intratumor heterogeneity in localized lung adenocarcinomas delineated by multiregion sequencing.Science 346,256-259(2014)

[0653] 54.Walter,M.J.et al.Clonal architecture of secondary acute myeloid leukemia.N.Engl.J.Med.366,1090-1098(2012).

[0654] 55.Hunt DF,Henderson RA,Shabanowitz J,Sakaguchi K,Michel H,Sevilir N,Cox AL,Appella E,Engelhard VH.Characterization of peptides bound to the class I MHC molecule HLA-A2.1 by mass spectrometry.Science 1992.255:1261-1263.

[0655] 56.Zarling AL,Polefrone JM,Evans AM,Mikesh LM,Shabanowitz J,Lewis ST,Engelhard VH,Hunt DF.Identification of class I MHC-associated phosphopeptides as targets for cancer immunotherapy.Proc Natl Acad Sci U S A.2006 Oct 3;103(40):14889-94.

[0656] 57.Bassani-Sternberg M,Pletscher-Frankild S,Jensen LJ,Mann M.Mass spectrometry of human leukocyte antigen class I peptidomes reveals strong effects of protein abundance and turnover on antigen presentation.Mol Cell Proteomics.2015 Mar;14(3):658-73.doi:10.1074/mcp.M114.042812.

[0657] 58.Abelin JG,Trantham PD,Penny SA,Paterson AM,Ward ST,Hildebrand WH,Cobbold M,Bai DL,ShabanowitzJ,Hunt DF.Complementary IMAC enrichment methods for HLA-associated phosphopeptide identification by mass spectrometry.Nat Protoc.2015 Sep;10(9):1308-18.doi:10.1038/nprot.2015.086.Epub 2015 Aug 6

[0658] 59.Barnstable CJ,Bodmer WF,Brown G,Galfre G,Milstein C,Williams AF,Ziegler A.Production of monoclonal antibodies to group A erythrocytes,HLA and other human cell surface antigens-new tools for genetic analvsis.Cell.1978 May;14(1):9-20.

[0659] 60.Goldman JM,Hibbin J,Kearney L,Orchard K,Th'ng KH.HLA-DR monoclonal antibodies inhibit the proliferation of norml and chronic granulocytic leukemia myeloid progenitor cells.Br J Haematol.1982 Nov;52(3):411-20.

[0660] 61.Eng JK,JahanTA,Hoopmann MR.Comet:an open-source MS/MS sequence database search tool.Proteomics.2013 Jan;13(1):22-4.doi:10.1002/pmic.201200439.Epub 2012 Dec 4.

[0661] 62.Eng JK,Hoopmann MR,Jahan TA,Egertson JD,Noble WS,MacCoss MJ.A deeper look into Comet--implementation and features.J Am Soc Mass Spectrom.2015 Nov;26(11):1865-74.doi:10.1007/s13361-015-1179-x.Epub 2015 Jun 27.

[0662] 63.Lukas Käll,Jesse Canterbury,Jason Weston,William Stafford Noble and Michael J.MacCoss.Semi-supervised learning for peptide identification from shotgun proteomics datasets.Nature Methods 4:923-925,November 2007

[0663] 64.Lukas Käll,John D.Storey,Michael J.MacCoss and William Stafford Noble.Assigning confidence measures to peptides identified by tandem mass spectrometry.Journal of Proteome Research,7(1):29-34,January 2008

[0664] 65.Lukas Käll,John D.Storey and William Stafford Noble.Nonparametric estimation of posterior error probabilities associated with peptides identified by tandem mass spectrometry.Bioinformatics,24(16):i42-i48,August 2008

[0665] 66.Kinney RM,BJ Johnson,VL Brown,DW Trent.Nucleotide Sequence of the 26 S mRNA of the Virulent Trinidad Donkey Strain of Venezuelan Equine Encephalitis Virus and Deduced Sequence of the Encoded Structural Proteins.Virology 152(2),400-413.1986 Jul 30.

[0666] 67.Jill E Slansky,Frédérique M Rattis,Lisa F Boyd,Tarek Fahmy,Elizabeth M Jaffee,Jonathan P Schneck,David H Margulies,Drew M Pardoll.Enhanced Antigen-Specific Antitumor Immunity with Altered Peptide Ligands that Stabilize the MHC-Peptide-TCR Complex.Immunity,Volume 13,Issue 4,1 October 2000,Pages 529-538.

[0667] 68.A Y Huang,P H Gulden,A S Woods,M C Thomas,C D Tong,W Wang,V H Engelhard,GPasternack,R Cotter,D Hunt,D M Pardoll,and E M Jaffee.The immunodominant major histocompatibility complex class I-restricted antigen of a murine colon tumor derives from an endogenous retroviral gene product.Proc Natl Acad Sci U S A.;93(18):9730-9735,1996 Sep 3.

[0668] 69.JOHNSON,BARBARA J.B.,RICHARD M.KINNEY,CRYSTLE L.KOST AND DENNIS W.TRENT.Molecular Determinants of Alphavirus Neurovirulence:Nucleotide and Deduced Protein Sequence Changes during Attenuation of Venezuelan Equine Encephalitis Virus.J Gen Virol 67:1951-1960,1986.

[0669] 70.Aarnoudse,C.A.,Krüse,M.,Konopitzky,R.,Brouwenstijn,N.,and Schrier,P.I.(2002).TCR reconstitution in Jurkat reporter cells facilitates the identification of novel tumor antigens by cDNA expression cloning.Int J Cancer 99,7-13.

[0670] 71.Alexander,J.,Sidney,J.,Southwood,S.,Ruppert,J.,Oseroff,C.,Maewal,A.,Snoke,K.,Serra,H.M.,Kubo,R.T.,and Sette,A.(1994).Development of high potency universal DR-restricted helper epitopes by modification of high

affinity DR-blocking peptides.Immunity 1,751-761.

[0671] 72.Banu,N.,Chia,A.,Ho,Z.Z.,Gareia,A.T.,Paravasivam,K.,Grotenbreg,G.M.,Bertoletti,A.,and Gehring,A.J.(2014).Building and optimizing a virus-specific T cell receptor library for targeted immunotherapy in viral infections.Scientific Reports 4,4166.

[0672] 73.Cornet,S.,Miconnet,I.,Menez,J.,Lemonnier,F.,and Kosmatopoulos,K.(2006).Optimal organization of a polypeptide-based candidate cancer vaccine composed of cryptic tumor peptides with enhanced immunogenicity.Vaccine 24,2102-2109.

[0673] 74.Depla,E.,van der Aa,A.,Livingston,B.D.,Crimi,C.,Allosery,K.,de Brabandere,V.,Krakover,J.,Murthy,S.,Huang,M.,Power,S.,et al.(2008).Rational design ofa multiepitope vaccine encoding T-lymphocyte epitopes for treatment of chronic hepatitis B virus infections.Journal of Virology 82,435-450.

[0674] 75.Ishioka,G.Y.,Fikes,J.,Hermanson,G.,Livingston,B.,Crimi,C.,Qin,M.,del Guercio,M.F.,Oseroff,C.,Dahlberg,C.,Alexander,J.,et al.(1999).Utilization of MHC class I transgenic mice for development of minigene DNA vaccines encoding multiple HLA-restricted CTL epitopes.J Immunol 162,3915-3925.

[0675] 76.Janetzki,S.,Price,L.,Schroeder,H.,Britten,C.M.,Welters,M.J.P.,and Hoos,A.(2015).Guidelines for the automated evaluation of Elispot assays.Nat Protoc 10,1098-1115.

[0676] 77.Lyons,G.E.,Moore,T.,Brasic,N.,Li,M.,Roszkowski,J.J.,and Nishimura,M.I.(2006).Influence of human CD8 on antigen recognition by T-cell receptor-transduced cells.Cancer Res 66,11455-11461.

[0677] 78.Nagai,K.,Ochi,T.,Fujiwara,H.,An,J.,Shirakata,T.,Mineno,J.,Kuzushima,K.,Shiku,H.,Melenhorst,J.J.,Gostick,E.,et al.(2012).Aurora kinase A-specific T-cell receptor gene transfer redirects T lymphocytes to display effective antileukemia reactivity.Blood 119,368-376.

[0678] 79.Panina-Bordignon,P.,Tan,A.,Termijtelen,A.,Demotz,S.,Corradin,G.,and Lanzavecchia,A.(1989).Universally immunogenic T cell epitopes:promiscuous binding to human MHC class II and promiscuous recognition by T cells.Eur J Immunol 19,2237-2242.

[0679] 80.Vitiello,A.,Marchesini,D.,Furze,J.,Sherman,L.A.,and Chesnut,R.W.(1991).Analysis of the HLA-restricted influenza-specific cytotoxic T lymphocyte response in transgenic mice carrying a chimeric human-mouse class I major histocompatibility complex.J Exp Med 173,1007-1015.

[0680] 81.Yachi,P.P.,Ampudia,J.,Zal,T.,and Gascoigne,N.R.J.(2006).Altered peptide ligands induce delayed CD8-T cell receptor interaction--a role for CD8 in distinguishing antigen quality.Immunity 25,203-211.

[0681] 82.Pushko P,Parker M,Ludwig GV,Davis NL,Johnston RE,Smith

JF.Replicon-helper systems from attenuated Venezuelan equine encephalitis virus:expression of heterologous genes in vitro and immunization against heterologous pathogens in vivo.Virology.1997 Dec 22;239(2):389-401.

[0682] 83.Strauss,JH and E G Strauss.The alphaviruses:gene expression, replication,and evolution.Microbiol Rev.1994 Sep;58(3):491-562.

[0683] 84.Rhême C,Ehrengruber MU,Grandgirard D.Alphaviral cytotoxicity and its implication in vector development.Exp Physiol.2005 Jan;90(1):45-52.Epub 2004 Nov 12.

[0684] 85.Riley,Michael K.II,and Wilfred Vermerris.Recent Advances in Nanomaterials for Gene Delivery-A Review.Nanomaterials 2017,7(5),94.

[0685] 86.Frolov I,Hardy R,Rice CM.Cis-acting RNA elements at the 5′end of Sindbis vires genome RNA regulate minus-and plus-strand RNA synthesis.RNA.2001 Nov;7(11):1638-51.

[0686] 87.Jose J,Snyder JE,Kuhn RJ.A structural and functional perspective of alphavirus replication and assembly.Future Microbiol.2009 Sep;4(7):837-56.

[0687] 88.Bo Li and C.olin N.Dewey.RSEM:accurate transcript quantification from RNA-Seq data with or without a referenfe genome.BMC Bioinformatics,12: 323,August 2011

[0688] 89.Hillary Pearson,Tariq Daouda,Diana Paola Granados,Chantal Durette, Eric Bonneil,Mathieu Courcelles,Anja Rodenbrock,Jean-Philippe Laverdure, Caroline Côté,Sylvie Mader,Sébastien Lemieux,Pierre Thibault,and Claude Perreault.MHC class I-associated peptides derive from selective regions of the human genome.The Journal of Clinical Investigation,2016,

[0689] 90.Juliane Liepe,Fabio Marino,John Sidney,Anita Jeko,Daniel E.Bunting,Alessandro Sette,Peter M.Kloetzel,Michael P.H.Stumpf,Albert J.R.Heck,Michele Mishto.A large fraction of HLA class I ligands are proteasome-generated spliced peptides.Science,21,October 2016.

[0690] 91.Mommen GP.,Marino,F.,Meiring HD.,Poelen,MC.,van Gaans-van den Brink,JA.,Mohammed S.,Heck AJ.,and van Els CA.Sampling From the Proteome to the Human Leukocyte Antigen-DR(HLA-DR)Ligandome Proceeds Via High Specificity.Mol Cell Proteomics 15(4):1412-1423,April 2016.

[0691] 92.Sebastian Kreiter,Mathias Vormehr,Niels van de Roemer,Mustafa Diken,Martin Löwer,Jan Diekmann,Sebastian Boegel,Barbara Schrörs,Fulvia Vascotto, John C.Castle,Arbel D.Tadmor,Stephen P.Schoenberger,Christoph Huber,Özlem Türeci,and Ugur Sahin.Mutant MHC class II epitopes drive therapeutic immune responses to caner.Nature 520,692-696,April 2015.

[0692] 93.Tran E.,Turcotte S.,Gros A.,Robbins P.F.,Lu Y.C.,Dudley M.E., Wunderlich J.R.,Somerville R.P.,Hogan K.,Hinrichs C.S.,Parkhurst M.R.,Yang J.C.,Rosenberg S.A.Cancer immunotherapy based on mutation-specific CD4+ T

cells in a patient with epithelial cancer.Science 344(6184)641-645,May 2014.

[0693] 94.Andreatta M.,Karosiene E.,Rasmussen M.,Stryhn A.,Buus S.,Nielsen M.Accurate pan-specific prediction of peptide-MHC class II binding affinity with improved binding core identification.Immunogenetics 67(11-12)641-650,November 2015.

[0694] 95.Nielsen,M.,Lund,O.NN-align.An artificial neural network-based alignment algorithm for MHC class II peptide binding prediction.BMC Bioinformatics 10:296,September 2009.

[0695] 96.Nielsen,M.,Lundegaard,C.,Lund,O.Prediction of MHC class II binding affinity using SMM-align,a novel stabilization matrix alignment method.BMC Bioinformatics 8:238,July 2007.

[0696] 97.Zhang,J.,et al.PEAKS DB:de novo sequencing assisted database search for sensitive and accurate peptide identification.Molecular&Cellular Proteomics.11(4):1-8.1/2/2012.

[0697] 98.Jensen,Kamilla Kjaergaard,et al.“Improved Methods for Prediting Peptide Binding Affinity to MHC Class II Molecules.”Immunology,2018,doi:10.1111/imm.12889.

[0698] 99.Carter,S.L.,Cibulskis,K.,Helman,E.,McKenna,A.,Shen,H.,Zack,T.,Laird,P.W.,Onofrio,R.C.,Winckler,W.,Weir,B.A.,et al.(2012).Absolute quantification of somatic DNA alterations in human cancer.Nat.Biotechnol.30,413-421

[0699] 100.McGranahan,N.,Rosenthal,R.,Hiley,C.T.,Rowan,A.J.,Watkins,T.B.K.,Wilson,G.A.,Birkbak,N.J.,Veeriah,S.,Van Loo,P.,Herrero,J.,et al.(2017).Allele-Specific HLA Loss and Immune Escape in Lung Cancer Evolution.Cell 171,1259-1271.e11.

[0700] 101.Shukla,S.A.,Rooney,M.S.,Rajasagi,M.,Tiao,G.,Dixon,P.M.,Lawrence,M.S.,Stevens,J.,Lane,W.J.,Dellagatta,J.L.,Steelman,S.,et al.(2015).Comprehensive analysis ofcancer-associated somatic mutations in class I HLA genes.Nat.Biotechnol.33,1152-1158.

[0701] 102.Van Loo,P.,Nordgard,S.H.,Lingjærde,O.C.,Russnes,H.G.,Rye,I.H.,Sun,W.,Weigman,V.J.,Marynen,P.,Zetterberg,A.,Naume,B.,et al.(2010).Allele-specific copy number analysis of tumors.Proc.Natl.Acad.Sci.U.S.A.107,16910-16915.

[0702] 103.Van Loo,P.,Nordgard,S.H.,Lingjærde,O.C.,Russnes,H.G.,Rye,I.H.,Sun,W.,Weigman,V.J.,Marynen,P.,Zetterberg,A.,Naume,B.,et al.(2010).Allele-specific copy number analysis of tumors.Proc.Natl.Acad.Sci.U.S.A.107,16910-16915.

[0703] 104.MacLean,B.,Tomazela,D.M.,Shulman,N.,Chambers,M.,Finney,G.L.,Frewen,B.,Kern,R.,Tabb,D.L.,Liebler,D.C.,MacCoss,M.J.,(2010).Skyline:an open

source document editor for creating and analyzing targeted proteomics experiments.Bioinformatics 26(7),966-8.

序列表

<110> 磨石肿瘤生物技术公司(GRITSTONE ONCOLOGY, INC.)

<120>多表位疫苗盒

<130> GSO-095

<140>
<141>

<150>
<151>

<160> 56

<170> PatentIn version 3.5

<210> 1
<211> 36519
<212> DNA
<213> 人工序列(Artificial Sequence)

<220>
<223> 人工序列的描述：合成多核苷酸

[0001]

<400> 1
ccatcttcaa taatatacct caaacttttt gtgcgcgtta atatgcaaat gaggcgtttg 60

aatttgggga ggaagggcgg tgattggtcg agggatgagc gaccgttagg ggcggggcga 120

gtgacgtttt gatgacgtgg ttgcgaggag gagccagttt gcaagttctc gtgggaaaag 180

tgacgtcaaa cgaggtgtgg tttgaacacg gaaatactca attttcccgc gctctctgac 240

aggaaatgag gtgtttctgg gcggatgcaa gtgaaaacgg gccattttcg cgcgaaaact 300

gaatgaggaa gtgaaaatct gagtaatttc gcgtttatgg cagggaggag tatttgccga 360

gggccgagta gactttgacc gattacgtgg gggtttcgat taccgtgttt ttcacctaaa 420

tttccgcgta cggtgtcaaa gtccggtgtt tttacgtagg tgtcagctga tcgccagggt 480

atttaaacct gcgctctcca gtcaagaggc cactcttgag tgccagcgag aagagttttc 540

tcctccgcgc cgcgagtcag atctacactt tgaaagatga ggcacctgag agacctgccc 600

gatgagaaaa tcatcatcgc ttccgggaac gagattctgg aactggtggt aaatgccatg 660

atgggcgacg accctccgga gccccccacc ccatttgaga caccttcgct gcacgatttg 720

tatgatctgg aggtggatgt gcccgaggac gatcccaatg aggaggcggt aaatgatttt 780

tttagcgatg ccgcgctgct agctgccgag gaggcttcga gctctagctc agacagcgac 840

tcttcactgc ataccctag acccggcaga ggtgagaaaa agatccccga gcttaaaggg 900

gaagagatgg acttgcgctg ctatgaggaa tgcttgcccc cgagcgatga tgaggacgag 960

caggcgatcc agaacgcagc gagccaggga gtgcaagccg ccagcgagag ctttgcgctg 1020

gactgcccgc ctctgcccgg acacggctgt aagtcttgtg aatttcatcg catgaatact 1080

ggagataaag ctgtgttgtg tgcactttgc tatatgagag cttacaacca ttgtgtttac 1140

agtaagtgtg attaagttga actttagagg gaggcagaga gcagggtgac tgggcgatga 1200

ctggtttatt tatgtatata tgttctttat ataggtcccg tctctgacgc agatgatgag 1260

acccccacta caaagtccac ttcgtcaccc ccagaaattg gcacatctcc acctgagaat 1320

attgttagac cagttcctgt tagagccact gggaggagag cagctgtgga atgtttggat 1380

gacttgctac agggtggggt tgaacctttg gacttgtgta cccggaaacg ccccaggcac 1440

[0002] taagtgccac acatgtgtgt ttacttgagg tgatgtcagt atttataggg tgtggagtgc 1500

aataaaaaat gtgttgactt taagtgcgtg gtttatgact caggggtggg gactgtgagt 1560

atataagcag gtgcagacct gtgtggttag ctcagagcgg catggagatt tggacggtct 1620

tggaagactt tcacaagact agacagctgc tagagaacgc ctcgaacgga gtctcttacc 1680

tgtggagatt ctgcttcggt ggcgacctag ctaggctagt ctacagggcc aaacaggatt 1740

atagtgaaca atttgaggtt attttgagag agtgttctgg tctttttgac gctcttaact 1800

tgggccatca gtctcacttt aaccagagga tttcgagagc ccttgatttt actactcctg 1860

gcagaaccac tgcagcagta gccttttttg cttttattct tgacaaatgg agtcaagaaa 1920

cccatttcag cagggattac cagctggatt tcttagcagt agctttgtgg agaacatgga 1980

agtgccagcg cctgaatgca atctccggct acttgccggt acagccgcta gacactctga 2040

ggatcctgaa tctccaggag agtcccaggg cacgccaacg tcgccagcag cagcagcagg 2100

aggaggatca agaagagaac ccgagagccg gcctggaccc tccggcggag gaggaggagt 2160

agctgacctg tttcctgaac tgcgccgggt gctgactagg tcttcgagtg gtcgggagag 2220

ggggattaag cgggagaggc atgatgagac taatcacaga actgaactga ctgtgggtct 2280

gatgagtcgc aagcgcccag aaacagtgtg gtggcatgag gtgcagtcga ctggcacaga 2340

tgaggtgtcg gtgatgcatg agaggttttc tctagaacaa gtcaagactt gttggttaga 2400

gcctgaggat gattgggagg tagccatcag gaattatgcc aagctggctc tgaggccaga 2460

caagaagtac aagattacta agctgataaa tatcagaaat gcctgctaca tctcagggaa 2520

tggggctgaa gtggagatct gtctccagga aagggtggct ttcagatgct gcatgatgaa 2580

tatgtacccg ggagtggtgg gcatggatgg ggttaccttt atgaacatga ggttcagggg 2640

agatgggtat aatggcacgg tctttatggc caataccaag ctgacagtcc atggctgctc 2700

cttctttggg tttaataaca cctgcatcga ggcctggggt caggtcggtg tgaggggctg 2760

cagtttttca gccaactgga tgggggtcgt gggcaggacc aagagtatgc tgtccgtgaa 2820

gaaatgcttg tttgagaggt gccacctggg ggtgatgagc gagggcgaag ccagaatccg 2880

ccactgcgcc tctaccgaga cgggctgctt tgtgctgtgc aagggcaatg ctaagatcaa 2940

[0003] gcataatatg atctgtggag cctcggacga gcgcggctac cagatgctga cctgcgccgg 3000

cgggaacagc catatgctgg ccaccgtaca tgtggcttcc catgctcgca agccctggcc 3060

cgagttcgag cacaatgtca tgaccaggtg caatatgcat ctggggtccc gccgaggcat 3120

gttcatgccc taccagtgca acctgaatta tgtgaaggtg ctgctggagc ccgatgccat 3180

gtccagagtg agcctgacgg gggtgtttga catgaatgtg gaggtgtgga agattctgag 3240

atatgatgaa tccaagacca ggtgccgagc ctgcgagtgc ggagggaagc atgccaggtt 3300

ccagcccgtg tgtgtggatg tgacggagga cctgcgaccc gatcatttgg tgttgccctg 3360

caccgggacg gagttcggtt ccagcgggga agaatctgac tagagtgagt agtgttctgg 3420

ggcgggggag gacctgcatg agggccagaa taactgaaat ctgtgctttt ctgtgtgttg 3480

cagcagcatg agcggaagcg gctcctttga gggaggggta ttcagccctt atctgacggg 3540

gcgtctcccc tcctgggcgg gagtgcgtca gaatgtgatg ggatccacgg tggacggccg 3600

gcccgtgcag cccgcgaact cttcaaccct gacctatgca accctgagct cttcgtcgtt 3660

ggacgcagct gccgccgcag ctgctgcatc tgccgccagc gccgtgcgcg gaatggccat 3720

gggcgccggc tactacggca ctctggtggc caactcgagt tccaccaata atcccgccag 3780

cctgaacgag gagaagctgt tgctgctgat ggcccagctc gaggccttga cccagcgcct 3840

gggcgagctg acccagcagg tggctcagct gcaggagcag acgcgggccg cggttgccac 3900

ggtgaaatcc aaataaaaaa tgaatcaata aataaacgga gacggttgtt gattttaaca 3960

cagagtctga atctttattt gatttttcgc gcgcggtagg ccctggacca ccggtctcga 4020

tcattgagca cccggtggat cttttccagg acccggtaga ggtgggcttg gatgttgagg 4080

tacatgggca tgagcccgtc ccgggggtgg aggtagctcc attgcagggc ctcgtgctcg 4140

ggggtggtgt tgtaaatcac ccagtcatag caggggcgca gggcatggtg ttgcacaata 4200

tctttgagga ggagactgat ggccacgggc agcccttttgg tgtaggtgtt tacaaatctg 4260

ttgagctggg agggatgcat gcgggggggag atgaggtgca tcttggcctg gatcttgaga 4320

ttggcgatgt taccgcccag atcccgcctg gggttcatgt tgtgcaggac caccagcacg 4380

gtgtatccgg tgcacttggg gaatttatca tgcaacttgg aagggaaggc gtgaaagaat 4440

[0004] ttggcgacgc ctttgtgccc gcccaggttt tccatgcact catccatgat gatggcgatg 4500

ggcccgtggg cggcggcctg ggcaaagacg tttcgggggt cggacacatc atagttgtgg 4560

tcctgggtga ggtcatcata ggccatttta atgaatttgg ggcggagggt gccggactgg 4620

gggacaaagg taccctcgat cccgggggcg tagttcccct cacagatctg catctcccag 4680

gctttgagct cggaggggggg gatcatgtcc acctgcgggg cgataaagaa cacggtttcc 4740

ggggcggggg agatgagctg ggccgaaagc aagttccgga gcagctggga cttgccgcag 4800

ccggtggggc cgtagatgac cccgatgacc ggctgcaggt ggtagttgag ggagagacag 4860

ctgccgtcct cccggaggag gggggccacc tcgttcatca tctcgcgcac gtgcatgttc 4920

tcgcgcacca gttccgccag gaggcgctct ccccccaggg ataggagctc ctggagcgag 4980

gcgaagtttt tcagcggctt gagtccgtcg gccatgggca ttttggagag ggtttgttgc 5040

aagagttcca ggcggtccca gagctcggtg atgtgctcta cggcatctcg atccagcaga 5100

cctcctcgtt tcgcgggttg ggacggctgc gggagtaggg caccagacga tgggcgtcca 5160

gcgcagccag ggtccggtcc ttccagggtc gcagcgtccg cgtcagggtg gtctccgtca 5220

cggtgaaggg gtgcgcgccg ggctgggcgc ttgcgagggt gcgcttcagg ctcatccggc 5280

tggtcgaaaa ccgctcccga tcggcgccct gcgcgtcggc caggtagcaa ttgaccatga 5340

gttcgtagtt gagcgcctcg gccgcgtggc ctttggcgcg gagcttacct ttggaagtct 5400

gcccgcaggc gggacagagg agggacttga gggcgtagag cttgggggcg aggaagacgg 5460

actcgggggc gtaggcgtcc gcgccgcagt gggcgcagac ggtctcgcac tccacgagcc 5520

aggtgaggtc gggctggtcg gggtcaaaaa ccagtttccc gccgttcttt ttgatgcgtt 5580

tcttaccttt ggtctccatg agctcgtgtc cccgctgggt gacaaagagg ctgtccgtgt 5640

ccccgtagac cgactttatg ggccggtcct cgagcggtgt gccgcggtcc tcctcgtaga 5700

ggaaccccgc ccactccgag acgaaagccc gggtccaggc cagcacgaag gaggccacgt 5760

gggacgggta gcggtcgttg tccaccagcg ggtccacctt ttccagggta tgcaaacaca 5820

tgtccccctc gtccacatcc aggaaggtga ttggcttgta agtgtaggcc acgtgaccgg 5880

gggtcccggc cgggggggta taaaagggtg cgggtccctg ctcgtcctca ctgtcttccg 5940

[0005] gatcgctgtc caggagcgcc agctgttggg gtaggtattc cctctcgaag gcgggcatga 6000

cctcggcact caggttgtca gtttctagaa acgaggagga tttgatattg acggtgccgg 6060

cggagatgcc tttcaagagc ccctcgtcca tctggtcaga aaagacgatc tttttgttgt 6120

cgagcttggt ggcgaaggag ccgtagaggg cgttggagag gagcttggcg atggagcgca 6180

tggtctggtt tttttccttg tcggcgcgct ccttggcggc gatgttgagc tgcacgtact 6240

cgcgcgccac gcacttccat tcggggaaga cggtggtcag ctcgtcgggc acgattctga 6300

cctgccagcc ccgattatgc agggtgatga ggtccacact ggtggccacc tcgccgcgca 6360

ggggctcatt agtccagcag aggcgtccgc ccttgcgcga gcagaagggg ggcaggggtg 6420

ccagcatgac ctcgtcgggg gggtcggcat cgatggtgaa gatgccgggc aggaggtcgg 6480

ggtcaaagta gctgatggaa gtggccagat cgtccagggc agcttgccat tcgcgcacgg 6540

ccagcgcgcg ctcgtaggga ctgaggggcg tgccccaggg catgggatgg gtaagcgcgg 6600

aggcgtacat gccgcagatg tcgtagacgt agaggggctc ctcgaggatg ccgatgtagg 6660

tggggtagca gcgccccccg cggatgctgg cgcgcacgta gtcatacagc tcgtgcgagg 6720

gggcgaggag ccccgggccc aggttggtgc gactgggctt ttcggcgcgg tagacgatct 6780

ggcggaaaat ggcatgcgag ttggaggaga tggtgggcct ttggaagatg ttgaagtggg 6840

cgtggggcag tccgaccgag tcgcggatga agtgggcgta ggagtcttgc agcttggcga 6900

cgagctcggc ggtgactagg acgtccagag cgcagtagtc gagggtctcc tggatgatgt 6960

catacttgag ctgtcccttt tgtttccaca gctcgcggtt gagaaggaac tcttcgcggt 7020

ccttccagta ctcttcgagg gggaacccgt cctgatctgc acggtaagag cctagcatgt 7080

agaactggtt gacggccttg taggcgcagc agcccttctc cacggggagg gcgtaggcct 7140

gggcggcctt gcgcagggag gtgtgcgtga gggcgaaagt gtccctgacc atgaccttga 7200

ggaactggtg cttgaagtcg atatcgtcgc agccccccctg ctcccagagc tggaagtccg 7260

tgcgcttctt gtaggcgggg ttgggcaaag cgaaagtaac atcgttgaag aggatcttgc 7320

ccgcgcgggg cataaagttg cgagtgatgc ggaaaggttg gggcacctcg gcccggttgt 7380

tgatgacctg ggcggcgagc acgatctcgt cgaagccgtt gatgttgtgg cccacgatgt 7440

[0006] agagttccac gaatcgcgga cggcccttga cgtggggcag tttcttgagc tcctcgtagg 7500

tgagctcgtc ggggtcgctg agcccgtgct gctcgagcgc ccagtcggcg agatgggggt 7560

tggcgcggag gaaggaagtc cagagatcca cggccagggc ggtttgcaga cggtcccggt 7620

actgacggaa ctgctgcccg acggccattt tttcgggggt gacgcagtag aaggtgcggg 7680

ggtccccgtg ccagcgatcc catttgagct ggagggcgag atcgagggcg agctcgacga 7740

gccggtcgtc cccggagagt ttcatgacca gcatgaaggg gacgagctgc ttgccgaagg 7800

accccatcca ggtgtaggtt tccacatcgt aggtgaggaa gagcctttcg gtgcgaggat 7860

gcgagccgat ggggaagaac tggatctcct gccaccaatt ggaggaatgg ctgttgatgt 7920

gatggaagta gaaatgccga cggcgcgccg aacactcgtg cttgtgttta tacaagcggc 7980

cacagtgctc gcaacgctgc acgggatgca cgtgctgcac gagctgtacc tgagttcctt 8040

tgacgaggaa tttcagtggg aagtggagtc gtggcgcctg catctcgtgc tgtactacgt 8100

cgtggtggtc ggcctggccc tcttctgcct cgatggtggt catgctgacg agcccgcgcg 8160

ggaggcaggt ccagacctcg gcgcgagcgg gtcggagagc gaggacgagg gcgcgcaggc 8220

cggagctgtc cagggtcctg agacgctgcg gagtcaggtc agtgggcagc ggcggcgcgc 8280

ggttgacttg caggagtttt tccagggcgc gcgggaggtc cagatggtac ttgatctcca 8340

ccgcgccatt ggtggcgacg tcgatggctt gcagggtccc gtgcccctgg ggtgtgacca 8400

ccgtcccccg tttcttcttg ggcggctggg gcgacggggg cggtgcctct tccatggtta 8460

gaagcggcgg cgaggacgcg cgccgggcgg caggggcggc tcggggcccg gaggcagggg 8520

cggcaggggc acgtcggcgc cgcgcgcggg taggttctgg tactgcgccc ggagaagact 8580

ggcgtgagcg acgacgcgac ggttgacgtc ctggatctga cgcctctggg tgaaggccac 8640

gggacccgtg agtttgaacc tgaaagagag ttcgacagaa tcaatctcgg tatcgttgac 8700

ggcggcctgc cgcaggatct cttgcacgtc gcccgagttg tcctggtagg cgatctcggt 8760

catgaactgc tcgatctcct cctcttgaag gtctccgcgg ccggcgcgct ccacggtggc 8820

cgcgaggtcg ttggagatgc ggcccatgag ctgcgagaag gcgttcatgc ccgcctcgtt 8880

ccagacgcgg ctgtagacca cgacgccctc gggatcgcgg gcgcgcatga ccacctgggc 8940

[0007] gaggttgagc tccacgtggc gcgtgaagac cgcgtagttg cagaggcgct ggtagaggta 9000

gttgagcgtg gtggcgatgt gctcggtgac gaagaaatac atgatccagc ggcggagcgg 9060

catctcgctg acgtcgccca gcgcctccaa acgttccatg gcctcgtaaa agtccacggc 9120

gaagttgaaa aactgggagt tgcgcgccga gacggtcaac tcctcctcca gaagacggat 9180

gagctcggcg atggtggcgc gcacctcgcg ctcgaaggcc cccgggagtt cctccacttc 9240

ctcttcttcc tcctccacta acatctcttc tacttcctcc tcaggcggca gtggtggcgg 9300

gggaggggcc ctgcgtcgcc ggcggcgcac gggcagacgg tcgatgaagc gctcgatggt 9360

ctcgccgcgc cggcgtcgca tggtctcggt gacggcgcgc ccgtcctcgc ggggccgcag 9420

cgtgaagacg ccgccgcgca tctccaggtg gccgggggcg tccccgttgg gcagggagag 9480

ggcgctgacg atgcatctta tcaattgccc cgtagggact ccgcgcaagg acctgagcgt 9540

ctcgagatcc acgggatctg aaaaccgctg aacgaaggct tcgagccagt cgcagtcgca 9600

aggtaggctg agcacggttt cttctggcgg gtcatgttgg ttgggagcgg ggcgggcgat 9660

gctgctggtg atgaagttga ataggcggt tctgagacgg cggatggtgg cgaggagcac 9720

caggtctttg ggcccggctt gctggatgcg cagacggtcg gccatgcccc aggcgtggtc 9780

ctgacacctg gccaggtcct tgtagtagtc ctgcatgagc cgctccacgg gcacctcctc 9840

ctcgcccgcg cggccgtgca tgcgcgtgag cccgaagccg cgctggggct ggacgagcgc 9900

caggtcggcg acgacgcgct cggcgaggat ggcttgctgg atctgggtga gggtggtctg 9960

gaagtcatca aagtcgacga agcggtggta ggctccggtg ttgatggtgt aggagcagtt 10020

ggccatgacg gaccagttga cggtctggtg gccccggacgc acgagctcgt ggtacttgag 10080

gcgcgagtag gcgcgcgtgt cgaagatgta gtcgttgcag gtgcgcacca ggtactggta 10140

gccgatgagg aagtgcggcg gcggctggcg gtagagcggc catcgctcgg tggcgggggc 10200

gccgggcgcg aggtcctcga gcatggtgcg gtggtagccg tagatgtacc tggacatcca 10260

ggtgatgccg gcggcggtgg tggaggcgcg cgggaactcg cggacgcggt tccagatgtt 10320

gcgcagcggc aggaagtagt tcatggtggg cacggtctgg cccgtgaggc gcgcgcagtc 10380

gtggatgctc tatacgggca aaaacgaaag cggtcagcgg ctcgactccg tggcctggag 10440

[0008] gctaagcgaa cgggttgggc tgcgcgtgta ccccggttcg aatctcgaat caggctggag 10500

ccgcagctaa cgtggtattg gcactcccgt ctcgacccaa gcctgcacca accctccagg 10560

atacggaggc gggtcgtttt gcaacttttt tttggaggcc ggatgagact agtaagcgcg 10620

gaaagcggcc gaccgcgatg gctcgctgcc gtagtctgga gaagaatcgc cagggttgcg 10680

ttgcggtgtg ccccggttcg aggccggccg gattccgcgg ctaacgaggg cgtggctgcc 10740

ccgtcgtttc caagacccca tagccagccg acttctccag ttacggagcg agcccctctt 10800

ttgttttgtt tgtttttgcc agatgcatcc cgtactgcgg cagatgcgcc cccaccaccc 10860

tccaccgcaa caacagcccc ctccacagcc ggcgcttctg cccccgcccc agcagcaact 10920

tccagccacg accgccgcgg ccgccgtgag cggggctgga cagagttatg atcaccagct 10980

ggccttggaa gagggcgagg ggctggcgcg cctgggggcg tcgtcgccgg agcggcaccc 11040

gcgcgtgcag atgaaaaggg acgctcgcga ggcctacgtg cccaagcaga acctgttcag 11100

agacaggagc ggcgaggagc ccgaggagat gcgcgcggcc cggttccacg cggggcggga 11160

gctgcggcgc ggcctggacc gaaagagggt gctgagggac gaggatttcg aggcggacga 11220

gctgacgggg atcagccccg cgcgcgcgca cgtggccgcg gccaacctgg tcacggcgta 11280

cgagcagacc gtgaaggagg agagcaactt ccaaaaatcc ttcaacaacc acgtgcgcac 11340

cctgatcgcg cgcgaggagg tgaccctggg cctgatgcac ctgtgggacc tgctggaggc 11400

catcgtgcag aaccccacca gcaagccgct gacggcgcag ctgttcctgg tggtgcagca 11460

tagtcgggac aacgaagcgt tcagggaggc gctgctgaat atcaccgagc ccgagggccg 11520

ctggctcctg gacctggtga acattctgca gagcatcgtg gtgcaggagc gcgggctgcc 11580

gctgtccgag aagctggcgg ccatcaactt ctcggtgctg agtttgggca agtactacgc 11640

taggaagatc tacaagaccc cgtacgtgcc catagacaag gaggtgaaga tcgacgggtt 11700

ttacatgcgc atgaccctga aagtgctgac cctgagcgac gatctggggg tgtaccgcaa 11760

cgacaggatg caccgtgcgg tgagcgccag caggcggcgc gagctgagcg accaggagct 11820

gatgcatagt ctgcagcggg ccctgaccgg ggccgggacc gaggggggaga gctactttga 11880

catgggcgcg gacctgcact ggcagcccag ccgccgggcc ttggaggcgg cggcaggacc 11940

[0009] ctacgtagaa gaggtggacg atgaggtgga cgaggagggc gagtacctgg aagactgatg 12000

gcgcgaccgt atttttgcta gatgcaacaa caacagccac ctcctgatcc cgcgatgcgg 12060

gcggcgctgc agagccagcc gtccggcatt aactcctcgg acgattggac ccaggccatg 12120

caacgcatca tggcgctgac gacccgcaac cccgaagcct ttagacagca gccccaggcc 12180

aaccggctct cggccatcct ggaggccgtg gtgccctcgc gctccaaccc cacgcacgag 12240

aaggtcctgg ccatcgtgaa cgcgctggtg gagaacaagg ccatccgcgg cgacgaggcc 12300

ggcctggtgt acaacgcgct gctggagcgc gtggcccgct acaacagcac caacgtgcag 12360

accaacctgg accgcatggt gaccgacgtg cgcgaggccg tggcccagcg cgagcggttc 12420

caccgcgagt ccaacctggg atccatggtg gcgctgaacg ccttcctcag cacccagccc 12480

gccaacgtgc cccggggcca ggaggactac accaacttca tcagcgccct gcgcctgatg 12540

gtgaccgagg tgccccagag cgaggtgtac cagtccgggc cggactactt cttccagacc 12600

agtcgccagg gcttgcagac cgtgaacctg agccaggctt tcaagaactt gcagggcctg 12660

tggggcgtgc aggccccggt cggggaccgc gcgacggtgt cgagcctgct gacgccgaac 12720

tcgcgcctgc tgctgctgct ggtggccccc ttcacggaca gcggcagcat caaccgcaac 12780

tcgtacctgg gctacctgat taacctgtac cgcgaggcca tcggccaggc gcacgtggac 12840

gagcagacct accaggagat cacccacgtg agccgcgccc tgggccagga cgacccgggc 12900

aacctggaag ccaccctgaa ctttttgctg accaaccggt cgcagaagat cccgccccag 12960

tacgcgctca gcaccgagga ggagcgcatc ctgcgttacg tgcagcagag cgtgggcctg 13020

ttcctgatgc aggagggggc cacccccagc gccgcgctcg acatgaccgc gcgcaacatg 13080

gagcccagca tgtacgccag caaccgcccg ttcatcaata aactgatgga ctacttgcat 13140

cgggcggccg ccatgaactc tgactatttc accaacgcca tcctgaatcc ccactggctc 13200

ccgccgccgg ggttctacac gggcgagtac gacatgcccg accccaatga cgggttcctg 13260

tgggacgatg tggacagcag cgtgttctcc ccccgaccgg gtgctaacga gcgccccttg 13320

tggaagaagg aaggcagcga ccgacgcccg tcctcggcgc tgtccggccg cgagggtgct 13380

gccgcggcgg tgcccgaggc cgccagtcct ttcccgagct tgcccttctc gctgaacagt 13440

[0010] atccgcagca gcgagctggg caggatcacg cgcccgcgct tgctgggcga agaggagtac 13500

ttgaatgact cgctgttgag acccgagcgg gagaagaact tccccaataa cgggatagaa 13560

agcctggtgg acaagatgag ccgctggaag acgtatgcgc aggagcacag ggacgatccc 13620

cgggcgtcgc agggggccac gagccggggc agcgccgccc gtaaacgccg gtggcacgac 13680

aggcagcggg gacagatgtg ggacgatgag gactccgccg acgacagcag cgtgttggac 13740

ttgggtggga gtggtaaccc gttcgctcac ctgcgccccc gtatcgggcg catgatgtaa 13800

gagaaaccga aaataaatga tactcaccaa ggccatggcg accagcgtgc gttcgtttct 13860

tctctgttgt tgttgtatct agtatgatga ggcgtgcgta cccggagggt cctcctccct 13920

cgtacgagag cgtgatgcag caggcgatgg cggcggcggc gatgcagccc ccgctggagg 13980

ctccttacgt gcccccgcgg tacctggcgc ctacggaggg gcggaacagc attcgttact 14040

cggagctggc acccttgtac gataccaccc ggttgtacct ggtggacaac aagtcggcgg 14100

acatcgcctc gctgaactac cagaacgacc acagcaactt cctgaccacc gtggtgcaga 14160

acaatgactt cacccccacg gaggccagca cccagaccat caactttgac gagcgctcgc 14220

ggtggggcgg ccagctgaaa accatcatgc acaccaacat gcccaacgtg aacgagttca 14280

tgtacagcaa caagttcaag gcgcgggtga tggtctcccg caagaccccc aatggggtga 14340

cagtgacaga ggattatgat ggtagtcagg atgagctgaa gtatgaatgg gtggaatttg 14400

agctgcccga aggcaacttc tcggtgacca tgaccatcga cctgatgaac aacgccatca 14460

tcgacaatta cttggcggtg gggcggcaga acggggtgct ggagagcgac atcggcgtga 14520

agttcgacac taggaacttc aggctgggct gggaccccgt gaccgagctg gtcatgcccg 14580

gggtgtacac caacgaggct ttccatcccg atattgtctt gctgcccggc tgcggggtgg 14640

acttcaccga gagccgcctc agcaacctgc tgggcattcg caagaggcag cccttccagg 14700

aaggcttcca gatcatgtac gaggatctgg agggggggcaa catccccgcg ctcctggatg 14760

tcgacgccta tgagaaaagc aaggaggatg cagcagctga agcaactgca gccgtagcta 14820

ccgcctctac cgaggtcagg ggcgataatt ttgcaagcgc cgcagcagtg gcagcggccg 14880

aggcggctga aaccgaaagt aagatagtca ttcagccggt ggagaaggat agcaagaaca 14940

[0011] ggagctacaa cgtactaccg gacaagataa acaccgccta ccgcagctgg tacctagcct 15000

acaactatgg cgaccccgag aagggcgtgc gctcctggac gctgctcacc acctcggacg 15060

tcacctgcgg cgtggagcaa gtctactggt cgctgcccga catgatgcaa gacccggtca 15120

ccttccgctc cacgcgtcaa gttagcaact acccggtggt gggcgccgag ctcctgcccg 15180

tctactccaa gagcttcttc aacgagcagg ccgtctactc gcagcagctg cgcgccttca 15240

cctcgcttac gcacgtcttc aaccgcttcc ccgagaacca gatcctcgtc cgcccgcccg 15300

cgcccaccat taccaccgtc agtgaaaacg ttcctgctct cacagatcac gggaccctgc 15360

cgctgcgcag cagtatccgg ggagtccagc gcgtgaccgt tactgacgcc agacgccgca 15420

cctgccccta cgtctacaag gccctgggca tagtcgcgcc gcgcgtcctc tcgagccgca 15480

ccttctaaat gtccattctc atctcgccca gtaataacac cggttggggc ctgcgcgcgc 15540

ccagcaagat gtacggaggc gctcgccaac gctccacgca acaccccgtg cgcgtgcgcg 15600

ggcacttccg cgctccctgg ggcgccctca agggccgcgt gcggtcgcgc accaccgtcg 15660

acgacgtgat cgaccaggtg gtggccgacg cgcgcaacta cacccccgcc gccgcgcccg 15720

tctccaccgt ggacgccgtc atcgacagcg tggtggccga cgcgcgccgg tacgcccgcg 15780

ccaagagccg gcggcggcgc atcgcccggc ggcaccggag cacccccgcc atgcgcgcgg 15840

cgcgagcctt gctgcgcagg gccaggcgca cgggacgcag ggccatgctc agggcggcca 15900

gacgcgcggc ttcaggcgcc agcgccggca ggacccggag acgcgcggcc acggcggcgg 15960

cagcggccat cgccagcatg tcccgcccgc ggcgagggaa cgtgtactgg gtgcgcgacg 16020

ccgccaccgg tgtgcgcgtg cccgtgcgca cccgcccccc tcgcacttga agatgttcac 16080

ttcgcgatgt tgatgtgtcc cagcggcgag gaggatgtcc aagcgcaaat tcaaggaaga 16140

gatgctccag gtcatcgcgc ctgagatcta cggccctgcg gtggtgaagg aggaaagaaa 16200

gccccgcaaa atcaagcggg tcaaaaagga caaaaaggaa gaagaaagtg atgtggacgg 16260

attggtggag tttgtgcgcg agttcgcccc ccggcggcgc gtgcagtggc gcgggcggaa 16320

ggtgcaaccg gtgctgagac ccggcaccac cgtggtcttc acgcccggcg agcgctccgg 16380

caccgcttcc aagcgctcct acgacgaggt gtacggggat gatgatattc tggagcaggc 16440

[0012] ggccgagcgc ctgggcgagt ttgcttacgg caagcgcagc cgttccgcac cgaaggaaga 16500

ggcggtgtcc atcccgctgg accacggcaa ccccacgccg agcctcaagc ccgtgacctt 16560

gcagcaggtg ctgccgaccg cggcgccgcg ccgggggttc aagcgcgagg gcgaggatct 16620

gtaccccacc atgcagctga tggtgcccaa gcgccagaag ctggaagacg tgctggagac 16680

catgaaggtg gacccggacg tgcagcccga ggtcaaggtg cggcccatca agcaggtggc 16740

cccgggcctg ggcgtgcaga ccgtggacat caagattccc acggagccca tggaaacgca 16800

gaccgagccc atgatcaagc ccagcaccag caccatggag gtgcagacgg atccctggat 16860

gccatcggct cctagtcgaa gaccccggcg caagtacggc gcggccagcc tgctgatgcc 16920

caactacgcg ctgcatcctt ccatcatccc cacgccgggc taccgcggca cgcgcttcta 16980

ccgcggtcat accagcagcc gccgccgcaa gaccaccact cgccgccgcc gtcgccgcac 17040

cgccgctgca accacccctg ccgccctggt gcggagagtg taccgccgcg gccgcgcacc 17100

tctgaccctg ccgcgcgcgc gctaccaccc gagcatcgcc atttaaactt tcgcctgctt 17160

tgcagatcaa tggccctcac atgccgcctt cgcgttccca ttacgggcta ccgaggaaga 17220

aaaccgcgcc gtagaaggct ggcggggaac gggatgcgtc gccaccacca ccggcggcgg 17280

cgcgccatca gcaagcggtt gggggggaggc ttcctgcccg cgctgatccc catcatcgcc 17340

gcggcgatcg gggcgatccc cggcattgct tccgtggcgg tgcaggcctc tcagcgccac 17400

tgagacacac ttggaaacat cttgtaataa accaatggac tctgacgctc ctggtcctgt 17460

gatgtgtttt cgtagacaga tggaagacat caatttttcg tccctggctc cgcgacacgg 17520

cacgcggccg ttcatgggca cctggagcga catcggcacc agccaactga acgggggcgc 17580

cttcaattgg agcagtctct ggagcgggct taagaatttc gggtccacgc ttaaaaccta 17640

tggcagcaag gcgtggaaca gcaccacagg gcaggcgctg agggataagc tgaaagagca 17700

gaacttccag cagaaggtgg tcgatgggct cgcctcgggc atcaacgggg tggtggacct 17760

ggccaaccag gccgtgcagc ggcagatcaa cagccgcctg gacccggtgc cgcccgccgg 17820

ctccgtggag atgccgcagg tggaggagga gctgcctccc ctggacaagc ggggcgagaa 17880

gcgacccccgc cccgatgcgg aggagacgct gctgacgcac acggacgagc cgcccccgta 17940

[0013] cgaggaggcg gtgaaactgg gtctgcccac cacgcggccc atcgcgcccc tggccaccgg 18000

ggtgctgaaa cccgaaaagc ccgcgaccct ggacttgcct cctccccagc cttcccgccc 18060

ctctacagtg gctaagcccc tgccgccggt ggccgtggcc cgcgcgcgac ccgggggcac 18120

cgcccgccct catgcgaact ggcagagcac tctgaacagc atcgtgggtc tgggagtgca 18180

gagtgtgaag cgccgccgct gctattaaac ctaccgtagc gcttaacttg cttgtctgtg 18240

tgtgtatgta ttatgtcgcc gccgccgctg tccaccagaa ggaggagtga agaggcgcgt 18300

cgccgagttg caagatggcc accccatcga tgctgcccca gtgggcgtac atgcacatcg 18360

ccggacagga cgcttcggag tacctgagtc cgggtctggt gcagtttgcc cgcgccacag 18420

acacctactt cagtctgggg aacaagttta ggaaccccac ggtggcgccc acgcacgatg 18480

tgaccaccga ccgcagccag cggctgacgc tgcgcttcgt gcccgtggac cgcgaggaca 18540

acacctactc gtacaaagtg cgctacacgc tggccgtggg cgacaaccgc gtgctggaca 18600

tggccagcac ctactttgac atccgcggcg tgctggatcg gggccctagc ttcaaaccct 18660

actccggcac cgcctacaac agtctggccc ccaagggagc acccaacact tgtcagtgga 18720

catataaagc cgatggtgaa actgccacag aaaaaaccta tacatatgga aatgcacccg 18780

tgcagggcat taacatcaca aaagatggta ttcaacttgg aactgacacc gatgatcagc 18840

caatctacgc agataaaacc tatcagcctg aacctcaagt gggtgatgct gaatggcatg 18900

acatcactgg tactgatgaa aagtatggag gcagagctct taagcctgat accaaaatga 18960

agccttgtta tggttctttt gccaagccta ctaataaaga aggaggtcag gcaaatgtga 19020

aaacaggaac aggcactact aaagaatatg acatagacat ggctttcttt gacaacagaa 19080

gtgcggctgc tgctggccta gctccagaaa ttgttttgta tactgaaaat gtggatttgg 19140

aaactccaga tacccatatt gtatacaaag caggcacaga tgacagcagc tcttctatta 19200

atttgggtca gcaagccatg cccaacagac ctaactacat tggtttcaga gacaacttta 19260

tcgggctcat gtactacaac agcactggca atatgggggt gctggccggt caggcttctc 19320

agctgaatgc tgtggttgac ttgcaagaca gaaacaccga gctgtcctac cagctcttgc 19380

ttgactctct gggtgacaga acccggtatt tcagtatgtg gaatcaggcg gtggacagct 19440

[0014] atgatcctga tgtgcgcatt attgaaaatc atggtgtgga ggatgaactt cccaactatt 19500

gtttccctct ggatgctgtt ggcagaacag atacttatca gggaattaag gctaatggaa 19560

ctgatcaaac cacatggacc aaagatgaca gtgtcaatga tgctaatgag ataggcaagg 19620

gtaatccatt cgccatggaa atcaacatcc aagccaacct gtggaggaac ttcctctacg 19680

ccaacgtggc cctgtacctg cccgactctt acaagtacac gccggccaat gttaccctgc 19740

ccaccaacac caacacctac gattacatga acggccgggt ggtggcgccc tcgctggtgg 19800

actcctacat caacatcggg gcgcgctggt cgctggatcc catggacaac gtgaacccct 19860

tcaaccacca ccgcaatgcg gggctgcgct accgctccat gctcctgggc aacgggcgct 19920

acgtgccctt ccacatccag gtgccccaga aatttttcgc catcaagagc ctcctgctcc 19980

tgcccgggtc ctacacctac gagtggaact tccgcaagga cgtcaacatg atcctgcaga 20040

gctccctcgg caacgacctg cgcacggacg gggcctccat ctccttcacc agcatcaacc 20100

tctacgccac cttcttcccc atggcgcaca acacggcctc cacgctcgag gccatgctgc 20160

gcaacgacac caacgaccag tccttcaacg actacctctc ggcggccaac atgctctacc 20220

ccatcccggc caacgccacc aacgtgccca tctccatccc ctcgcgcaac tgggccgcct 20280

tccgcggctg gtccttcacg cgtctcaaga ccaaggagac gccctcgctg ggctccgggt 20340

tcgaccccta cttcgtctac tcgggctcca tcccctacct cgacggcacc ttctacctca 20400

accacacctt caagaaggtc tccatcacct tcgactcctc cgtcagctgg cccggcaacg 20460

accggctcct gacgcccaac gagttcgaaa tcaagcgcac cgtcgacggc gagggctaca 20520

acgtggccca gtgcaacatg accaaggact ggttcctggt ccagatgctg gcccactaca 20580

acatcggcta ccagggcttc tacgtgcccg agggctacaa ggaccgcatg tactccttct 20640

tccgcaactt ccagcccatg agccgccagg tggtggacga ggtcaactac aaggactacc 20700

aggccgtcac cctggcctac cagcacaaca actcgggctt cgtcggctac ctcgcgccca 20760

ccatgcgcca gggccagccc taccccgcca actaccccta cccgctcatc ggcaagagcg 20820

ccgtcaccag cgtcacccag aaaaagttcc tctgcgacag ggtcatgtgg cgcatccccct 20880

tctccagcaa cttcatgtcc atgggcgcgc tcaccgacct cggccagaac atgctctatg 20940

[0015] ccaactccgc ccacgcgcta gacatgaatt tcgaagtcga ccccatggat gagtccaccc 21000

ttctctatgt tgtcttcgaa gtcttcgacg tcgtccgagt gcaccagccc caccgcggcg 21060

tcatcgaggc cgtctacctg cgcaccccct tctcggccgg taacgccacc acctaagctc 21120

ttgcttcttg caagccatgg ccgcgggctc cggcgagcag gagctcaggg ccatcatccg 21180

cgacctgggc tgcgggccct acttcctggg caccttcgat aagcgcttcc cgggattcat 21240

ggccccgcac aagctggcct gcgccatcgt caacacggcc ggccgcgaga ccgggggcga 21300

gcactggctg gccttcgcct ggaacccgcg ctcgaacacc tgctacctct tcgacccctt 21360

cgggttctcg gacgagcgcc tcaagcagat ctaccagttc gagtacgagg gcctgctgcg 21420

ccgcagcgcc ctggccaccg aggaccgctg cgtcaccctg gaaaagtcca cccagaccgt 21480

gcagggtccg cgctcggccg cctgcgggct cttctgctgc atgttcctgc acgccttcgt 21540

gcactggccc gaccgcccca tggacaagaa ccccaccatg aacttgctga cgggggtgcc 21600

caacggcatg ctccagtcgc cccaggtgga acccaccctg cgccgcaacc aggaggcgct 21660

ctaccgcttc ctcaactccc actccgccta ctttcgctcc caccgcgcgc gcatcgagaa 21720

ggccaccgcc ttcgaccgca tgaatcaaga catgtaaacc gtgtgtgtat gttaaatgtc 21780

tttaataaac agcactttca tgttacacat gcatctgaga tgatttattt agaaatcgaa 21840

agggttctgc cgggtctcgg catggcccgc gggcagggac acgttgcgga actggtactt 21900

ggccagccac ttgaactcgg ggatcagcag tttgggcagc ggggtgtcgg ggaaggagtc 21960

ggtccacagc ttccgcgtca gttgcagggc gcccagcagg tcgggcgcgg agatcttgaa 22020

atcgcagttg ggacccgcgt tctgcgcgcg ggagttgcgg tacacggggt tgcagcactg 22080

gaacaccatc agggccgggt gcttcacgct cgccagcacc gtcgcgtcgg tgatgctctc 22140

cacgtcgagg tcctcggcgt tggccatccc gaaggggGtc atcttgcagg tctgccttcc 22200

catggtgggc acgcacccgg gcttgtggtt gcaatcgcag tgcaggggga tcagcatcat 22260

ctgggcctgg tcggcgttca tccccgggta catggccttc atgaaagcct ccaattgcct 22320

gaacgcctgc tgggccttgg ctccctcggt gaagaagacc ccgcaggact tgctagagaa 22380

ctggttggtg gcgcacccgg cgtcgtgcac gcagcagcgc gcgtcgttgt tggccagctg 22440

[0016] caccacgctg cgcccccagc ggttctgggt gatcttggcc cggtcggggt tctccttcag 22500

cgcgcgctgc ccgttctcgc tcgccacatc catctcgatc atgtgctcct tctggatcat 22560

ggtggtcccg tgcaggcacc gcagcttgcc ctcggcctcg gtgcacccgt gcagccacag 22620

cgcgcacccg gtgcactccc agttcttgtg ggcgatctgg gaatgcgcgt gcacgaagcc 22680

ctgcaggaag cggcccatca tggtggtcag ggtcttgttg ctagtgaagg tcagcggaat 22740

gccgcggtgc tcctcgttga tgtacaggtg gcagatgcgg cggtacacct cgccctgctc 22800

gggcatcagc tggaagttgg ctttcaggtc ggtctccacg cggtagcggt ccatcagcat 22860

agtcatgatt tccataccct tctcccaggc cgagacgatg ggcaggctca tagggttctt 22920

caccatcatc ttagcgctag cagccgcggc caggggGtcg ctctcgtcca gggtctcaaa 22980

gctccgcttg ccgtccttct cggtgatccg caccgggggg tagctgaagc ccacggccgc 23040

cagctcctcc tcggcctgtc tttcgtcctc gctgtcctgg ctgacgtcct gcaggaccac 23100

atgcttggtc ttgcggggtt tcttcttggg cggcagcggc ggcggagatg ttggagatgg 23160

cgaggggGag cgcgagttct cgctcaccac tactatctct tcctcttctt ggtccgaggc 23220

cacgcggcgg taggtatgtc tcttcggggg cagaggcgga ggcgacgggc tctcgccgcc 23280

gcgacttggc ggatggctgg cagagcccct tccgcgttcg ggggtgcgct cccggcggcg 23340

ctctgactga cttcctccgc ggccggccat tgtgttctcc tagggaggaa caacaagcat 23400

ggagactcag ccatcgccaa cctcgccatc tgcccccacc gccgacgaga agcagcagca 23460

gcagaatgaa agcttaaccg ccccgccgcc cagccccgcc acctccgacg cggccgtccc 23520

agacatgcaa gagatggagg aatccatcga gattgacctg ggctatgtga cgcccgcgga 23580

gcacgaggag gagctggcag tgcgcttttc acaagaagag atacaccaag aacagccaga 23640

gcaggaagca gagaatgagc agagtcaggc tgggctcgag catgacggcg actacctcca 23700

cctgagcggg ggggaggacg cgctcatcaa gcatctggcc cggcaggcca ccatcgtcaa 23760

ggatgcgctg ctcgaccgca ccgaggtgcc cctcagcgtg gaggagctca gccgcgccta 23820

cgagttgaac ctcttctcgc cgcgcgtgcc ccccaagcgc cagcccaatg gcacctgcga 23880

gcccaacccg cgcctcaact tctacccggt cttcgcggtg cccgaggccc tggccaccta 23940

[0017] ccacatcttt ttcaagaacc aaaagatccc cgtctcctgc cgcgccaacc gcacccgcgc 24000

cgacgccctt ttcaacctgg gtcccggcgc ccgcctacct gatatcgcct ccttggaaga 24060

ggttcccaag atcttcgagg gtctgggcag cgacgagact cgggccgcga acgctctgca 24120

aggagaagga ggagagcatg agcaccacag cgccctggtc gagttggaag gcgacaacgc 24180

gcggctggcg gtgctcaaac gcacggtcga gctgacccat ttcgcctacc cggctctgaa 24240

cctgcccccc aaagtcatga gcgcggtcat ggaccaggtg ctcatcaagc gcgcgtcgcc 24300

catctccgag gacgagggca tgcaagactc cgaggagggc aagcccgtgg tcagcgacga 24360

gcagctggcc cggtggctgg gtcctaatgc tagtccccag agtttggaag agcggcgcaa 24420

actcatgatg gccgtggtcc tggtgaccgt ggagctggag tgcctgcgcc gcttcttcgc 24480

cgacgcggag accctgcgca aggtcgagga gaacctgcac tacctcttca ggcacgggtt 24540

cgtgcgccag gcctgcaaga tctccaacgt ggagctgacc aacctggtct cctacatggg 24600

catcttgcac gagaaccgcc tggggcagaa cgtgctgcac accaccctgc gcggggaggc 24660

ccggcgcgac tacatccgcg actgcgtcta cctctacctc tgccacacct ggcagacggg 24720

catgggcgtg tggcagcagt gtctggagga gcagaacctg aaagagctct gcaagctcct 24780

gcagaagaac ctcaagggtc tgtggaccgg gttcgacgag cgcaccaccg cctcggacct 24840

ggccgacctc attttccccg agcgcctcag gctgacgctg cgcaacggcc tgcccgactt 24900

tatgagccaa agcatgttgc aaaactttcg ctctttcatc ctcgaacgct ccggaatcct 24960

gcccgccacc tgctccgcgc tgccctcgga cttcgtgccg ctgaccttcc gcgagtgccc 25020

cccgccgctg tggagccact gctacctgct gcgcctggcc aactacctgg cctaccactc 25080

ggacgtgatc gaggacgtca gcggcgaggg cctgctcgag tgccactgcc gctgcaacct 25140

ctgcacgccg caccgctccc tggcctgcaa cccccagctg ctgagcgaga cccagatcat 25200

cggcaccttc gagttgcaag ggcccagcga aggcgagggt tcagccgcca agggggtct 25260

gaaactcacc ccggggctgt ggacctcggc ctacttgcgc aagttcgtgc ccgaggacta 25320

ccatcccttc gagatcaggt tctacgagga ccaatcccat ccgcccaagg ccgagctgtc 25380

ggcctgcgtc atcacccagg gggcgatcct ggcccaattg caagccatcc agaaatcccg 25440

[0018] ccaagaattc ttgctgaaaa agggccgcgg ggtctacctc gacccccaga ccggtgagga 25500

gctcaacccc ggcttccccc aggatgcccc gaggaaacaa gaagctgaaa gtggagctgc 25560

cgcccgtgga ggatttggag gaagactggg agaacagcag tcaggcagag gaggaggaga 25620

tggaggaaga ctgggacagc actcaggcag aggaggacag cctgcaagac agtctggagg 25680

aagacgagga ggaggcagag gaggaggtgg aagaagcagc cgccgccaga ccgtcgtcct 25740

cggcggggga gaaagcaagc agcacggata ccatctccgc tccgggtcgg ggtcccgctc 25800

gaccacacag tagatgggac gagaccggac gattcccgaa ccccaccacc cagaccggta 25860

agaaggagcg gcagggatac aagtcctggc gggggcacaa aaacgccatc gtctcctgct 25920

tgcaggcctg cgggggcaac atctccttca cccggcgcta cctgctcttc caccgcgggg 25980

tgaactttcc ccgcaacatc ttgcattact accgtcacct ccacagcccc tactacttcc 26040

aagaagaggc agcagcagca gaaaaagacc agcagaaaac cagcagctag aaaatccaca 26100

gcggcggcag caggtggact gaggatcgcg gcgaacgagc cggcgcaaac ccgggagctg 26160

aggaaccgga tctttcccac cctctatgcc atcttccagc agagtcgggg gcaggagcag 26220

gaactgaaag tcaagaaccg ttctctgcgc tcgctcaccc gcagttgtct gtatcacaag 26280

agcgaagacc aacttcagcg cactctcgag gacgccgagg ctctcttcaa caagtactgc 26340

gcgctcactc ttaaagagta gcccgcgccc gcccagtcgc agaaaaaggc gggaattacg 26400

tcacctgtgc ccttcgccct agccgcctcc acccatcatc atgagcaaag agattcccac 26460

gccttacatg tggagctacc agccccagat gggcctggcc gccggtgccg cccaggacta 26520

ctccacccgc atgaattggc tcagcgccgg gcccgcgatg atctcacggg tgaatgacat 26580

ccgcgcccac cgaaaccaga tactcctaga acagtcagcg ctcaccgcca cgccccgcaa 26640

tcacctcaat ccgcgtaatt ggcccgccgc cctggtgtac caggaaattc cccagcccac 26700

gaccgtacta cttccgcgag acgcccaggc cgaagtccag ctgactaact caggtgtcca 26760

gctggcgggc ggcgccaccc tgtgtcgtca ccgccccgct cagggtataa agcggctggt 26820

gatccggggc agaggcacac agctcaacga cgaggtggtg agctcttcgc tgggtctgcg 26880

acctgacgga gtcttccaac tcgccggatc ggggagatct tccttcacgc ctcgtcaggc 26940

[0019] cgtcctgact ttggagagtt cgtcctcgca gccccgctcg ggtggcatcg gcactctcca 27000

gttcgtggag gagttcactc cctcggtcta cttcaacccc ttctccggct cccccggcca 27060

ctacccggac gagttcatcc cgaacttcga cgccatcagc gagtcggtgg acggctacga 27120

ttgaatgtcc catggtggcg cagctgacct agctcggctt cgacacctgg accactgccg 27180

ccgcttccgc tgcttcgctc gggatctcgc cgagtttgcc tactttgagc tgcccgagga 27240

gcaccctcag ggcccggccc acggagtgcg gatcgtcgtc gaaggggccc tcgactccca 27300

cctgcttcgg atcttcagcc agcgtccgat cctggtcgag cgcgagcaag gacagaccct 27360

tctgactctg tactgcatct gcaaccaccc cggcctgcat gaaagtcttt gttgtctgct 27420

gtgtactgag tataataaaa gctgagatca gcgactactc cggacttccg tgtgttcctg 27480

aatccatcaa ccagtctttg ttcttcaccg ggaacgagac cgagctccag ctccagtgta 27540

agccccacaa gaagtacctc acctggctgt tccagggctc cccgatcgcc gttgtcaacc 27600

actgcgacaa cgacggagtc ctgctgagcg gccctgccaa ccttactttt tccacccgca 27660

gaagcaagct ccagctcttc caacccttcc tccccgggac ctatcagtgc gtctcgggac 27720

cctgccatca caccttccac ctgatcccga ataccacagc gtcgctcccc gctactaaca 27780

accaaactaa cctccaccaa cgccaccgtc gcgacctttc tgaatctaat actaccaccc 27840

acaccggagg tgagctccga ggtcaaccaa cctctgggat ttactacggc ccctgggagg 27900

tggttgggtt aatagcgcta ggcctagttg cgggtgggct tttggttctc tgctacctat 27960

acctcccttg ctgttcgtac ttagtggtgc tgtgttgctg gtttaagaaa tggggaagat 28020

caccctagtg agctgcggtg cgctggtggc ggtgttgctt tcgattgtgg gactgggcgg 28080

tgcggctgta gtgaaggaga aggccgatcc ctgcttgcat ttcaatccca acaaatgcca 28140

gctgagtttt cagcccgatg gcaatcggtg cgcggtactg atcaagtgcg gatgggaatg 28200

cgagaacgtg agaatcgagt acaataacaa gactcggaac aatactctcg cgtccgtgtg 28260

gcagcccggg gaccccgagt ggtacaccgt ctctgtcccc ggtgctgacg gctccccgcg 28320

caccgtgaat aatactttca tttttgcgca catgtgcgac acggtcatgt ggatgagcaa 28380

gcagtacgat atgtggcccc ccacgaagga gaacatcgtg gtcttctcca tcgcttacag 28440

[0020] cctgtgcacg gcgctaatca ccgctatcgt gtgcctgagc attcacatgc tcatcgctat 28500

tcgccccaga aataatgccg aaaaagaaaa acagccataa cgtttttttt cacacctttt 28560

tcagaccatg gcctctgtta aatttttgct tttatttgcc agtctcattg ccgtcattca 28620

tggaatgagt aatgagaaaa ttactattta cactggcact aatcacacat tgaaaggtcc 28680

agaaaaagcc acagaagttt catggtattg ttattttaat gaatcagatg tatctactga 28740

actctgtgga aacaataaca aaaaaaatga gagcattact ctcatcaagt ttcaatgtgg 28800

atctgactta accctaatta acatcactag agactatgta ggtatgtatt atggaactac 28860

agcaggcatt tcggacatgg aattttatca agtttctgtg tctgaaccca ccacgcctag 28920

aatgaccaca accacaaaaa ctacacctgt taccactatg cagctcacta ccaataacat 28980

ttttgccatg cgtcaaatgg tcaacaatag cactcaaccc accccaccca gtgaggaaat 29040

tcccaaatcc atgattggca ttattgttgc tgtagtggtg tgcatgttga tcatcgcctt 29100

gtgcatggtg tactatgcct tctgctacag aaagcacaga ctgaacgaca agctggaaca 29160

cttactaagt gttgaatttt aatttttttag aaccatgaag atcctaggcc ttttaatttt 29220

ttctatcatt acctctgctc tatgcaattc tgacaatgag gacgttactg tcgttgtcgg 29280

atcaaattat acactgaaag gtccagcgaa gggtatgctt tcgtggtatt gctattttgg 29340

atctgacact acagaaactg aattatgcaa tcttaagaat ggcaaaattc aaaattctaa 29400

aattaacaat tatatatgca atggtactga tctgatactc ctcaatatca cgaaatcata 29460

tgctggcagt tacacctgcc ctggagatga tgctgacagt atgatttttt acaaagtaac 29520

tgttgttgat cccactactc cacctccacc caccacaact actcacacca cacacacaga 29580

tcaaaccgca gcagaggagg cagcaaagtt agccttgcag gtccaagaca gttcatttgt 29640

tggcattacc cctacacctg atcagcggtg tccggggctg ctagtcagcg gcattgtcgg 29700

tgtgctttcg ggattagcag tcataatcat ctgcatgttc atttttgctt gctgctatag 29760

aaggctttac cgacaaaaat cagacccact gctgaacctc tatgtttaat tttttccaga 29820

gtcatgaagg cagttagcgc tctagttttt tgttctttga ttggcattgt tttttgcaat 29880

cctattccta aagttagctt tattaaagat gtgaatgtta ctgaggggggg caatgtgaca 29940

[0021] ctggtaggtg tagagggtgc tgaaaacacc acctggacaa aataccacct caatgggtgg 30000

aaagatattt gcaattggag tgtattagtt tatacatgtg agggagttaa tcttaccatt 30060

gtcaatgcca cctcagctca aaatggtaga attcaaggac aaagtgtcag tgtatctaat 30120

gggtatttta cccaacatac ttttatctat gacgttaaag tcataccact gcctacgcct 30180

agcccaccta gcactaccac acagacaacc cacactacac agacaaccac atacagtaca 30240

ttaaatcagc ctaccaccac tacagcagca gaggttgcca gctcgtctgg ggtccgagtg 30300

gcatttttga tgtgggcccc atctagcagt cccactgcta gtaccaatga gcagactact 30360

gaatttttgt ccactgtcga gagccacacc acagctacct ccagtgcctt ctctagcacc 30420

gccaatctct cctcgctttc ctctacacca atcagtcccg ctactactcc tagccccgct 30480

cctcttccca ctcccctgaa gcaaacagac ggcggcatgc aatggcagat caccctgctc 30540

attgtgatcg ggttggtcat cctggccgtg ttgctctact acatcttctg ccgccgcatt 30600

cccaacgcgc accgcaagcc ggtctacaag cccatcattg tcgggcagcc ggagccgctt 30660

caggtggaag ggggtctaag gaatcttctc ttctctttta cagtatggtg attgaactat 30720

gattcctaga caattcttga tcactattct tatctgcctc ctccaagtct gtgccaccct 30780

cgctctggtg gccaacgcca gtccagactg tattgggccc ttcgcctcct acgtgctctt 30840

tgccttcacc acctgcatct gctgctgtag catagtctgc ctgcttatca ccttcttcca 30900

gttcattgac tggatctttg tgcgcatcgc ctacctgcgc caccaccccc agtaccgcga 30960

ccagcgagtg gcgcggctgc tcaggctcct ctgataagca tgcgggctct gctacttctc 31020

gcgcttctgc tgttagtgct cccccgtccc gtcgaccccc ggtcccccac ccagtccccc 31080

gaggaggtcc gcaaatgcaa attccaagaa ccctggaaat tcctcaaatg ctaccgccaa 31140

aaatcagaca tgcatcccag ctggatcatg atcattggga tcgtgaacat tctggcctgc 31200

accctcatct cctttgtgat ttacccctgc tttgactttg gttggaactc gccagaggcg 31260

ctctatctcc cgcctgaacc tgacacacca ccacagcaac ctcaggcaca cgcactacca 31320

ccactacagc ctaggccaca atacatgccc atattagact atgaggccga gccacagcga 31380

cccatgctcc ccgctattag ttacttcaat ctaaccggcg gagatgactg acccactggc 31440

[0022] caacaacaac gtcaacgacc ttctcctgga catggacggc cgcgcctcgg agcagcgact 31500

cgcccaactt cgcattcgcc agcagcagga gagagccgtc aaggagctgc aggatgcggt 31560

ggccatccac cagtgcaaga gaggcatctt ctgcctggtg aaacaggcca agatctccta 31620

cgaggtcact ccaaacgacc atcgcctctc ctacgagctc ctgcagcagc gccagaagtt 31680

cacctgcctg gtcggagtca accccatcgt catcacccag cagtctggcg ataccaaggg 31740

gtgcatccac tgctcctgcg actcccccga ctgcgtccac actctgatca agaccctctg 31800

cggcctccgc gacctcctcc ccatgaacta atcacccccc tatccagtga aataaagatc 31860

atattgatga tgattttaca gaaataaaaa ataatcattt gatttgaaat aaagatacaa 31920

tcatattgat gatttgagtt taacaaaaaa ataaagaatc acttacttga aatctgatac 31980

caggtctctg tccatgtttt ctgccaacac cacttcactc ccctcttccc agctctggta 32040

ctgcaggccc cggcgggctg caaacttcct ccacacgctg aaggggatgt caaattcctc 32100

ctgtccctca atcttcattt tatcttctat cagatgtcca aaaagcgcgt ccgggtggat 32160

gatgacttcg accccgtcta cccctacgat gcagacaacg caccgaccgt gcccttcatc 32220

aaccccccct tcgtctcttc agatggattc caagagaagc ccctgggggt gttgtccctg 32280

cgactggccg accccgtcac caccaagaac ggggaaatca ccctcaagct gggagagggg 32340

gtggacctcg attcctcggg aaaactcatc tccaacacgg ccaccaaggc cgccgccccc 32400

ctcagttttt ccaacaacac catttccctt aacatggatc acccctttta cactaaagat 32460

ggaaaattat ccttacaagt ttctccacca ttaaatatac tgagaacaag cattctaaac 32520

acactagctt taggttttgg atcaggttta ggactccgtg gctctgcctt ggcagtacag 32580

ttagtctctc cacttacatt tgatactgat ggaaacataa agcttacctt agacagaggt 32640

ttgcatgtta caacaggaga tgcaattgaa agcaacataa gctgggctaa aggtttaaaa 32700

tttgaagatg gagccatagc aaccaacatt ggaaatgggt tagagtttgg aagcagtagt 32760

acagaaacag gtgttgatga tgcttaccca atccaagtta aacttggatc tggccttagc 32820

tttgacagta caggagccat aatggctggt aacaaagaag acgataaact cactttgtgg 32880

acaacacctg atccatcacc aaactgtcaa atactcgcag aaaatgatgc aaaactaaca 32940

[0023] ctttgcttga ctaaatgtgg tagtcaaata ctggccactg tgtcagtctt agttgtagga 33000

agtggaaacc taaaccccat tactggcacc gtaagcagtg ctcaggtgtt tctacgtttt 33060

gatgcaaacg gtgttctttt aacagaacat tctacactaa aaaaatactg ggggtatagg 33120

cagggagata gcatagatgg cactccatat accaatgctg taggattcat gcccaattta 33180

aaagcttatc caaagtcaca aagttctact actaaaaata atatagtagg gcaagtatac 33240

atgaatggag atgtttcaaa acctatgctt ctcactataa ccctcaatgg tactgatgac 33300

agcaacagta catattcaat gtcattttca tacacctgga ctaatggaag ctatgttgga 33360

gcaacatttg gggctaactc ttataccttc tcatacatcg cccaagaatg aacactgtat 33420

cccaccctgc atgccaaccc ttcccacccc actctgtgga acaaactctg aaacacaaaa 33480

taaaataaag ttcaagtgtt ttattgattc aacagtttta caggattcga gcagttattt 33540

ttcctccacc ctcccaggac atggaataca ccaccctctc cccccgcaca gccttgaaca 33600

tctgaatgcc attggtgatg gacatgcttt tggtctccac gttccacaca gtttcagagc 33660

gagccagtct cgggtcggtc agggagatga aaccctccgg gcactcccgc atctgcacct 33720

cacagctcaa cagctgagga ttgtcctcgg tggtcgggat cacggttatc tggaagaagc 33780

agaagagcgg cggtgggaat catagtccgc gaacgggatc ggccggtggt gtcgcatcag 33840

gccccgcagc agtcgctgcc gccgccgctc cgtcaagctg ctgctcaggg ggtccgggtc 33900

cagggactcc ctcagcatga tgcccacggc cctcagcatc agtcgtctgg tgcggcgggc 33960

gcagcagcgc atgcggatct cgctcaggtc gctgcagtac gtgcaacaca gaaccaccag 34020

gttgttcaac agtccatagt tcaacacgct ccagccgaaa ctcatcgcgg gaaggatgct 34080

acccacgtgg ccgtcgtacc agatcctcag gtaaatcaag tggtgccccc tccagaacac 34140

gctgcccacg tacatgatct ccttgggcat gtggcggttc accacctccc ggtaccacat 34200

caccctctgg ttgaacatgc agccccggat gatcctgcgg aaccacaggg ccagcaccgc 34260

cccgcccgcc atgcagcgaa gagaccccgg gtcccggcaa tggcaatgga ggacccaccg 34320

ctcgtacccg tggatcatct gggagctgaa caagtctatg ttggcacagc acaggcatat 34380

gctcatgcat ctcttcagca ctctcaactc ctcgggggtc aaaaccatat cccagggcac 34440

[0024] ggggaactct tgcaggacag cgaaccccgc agaacagggc aatcctcgca cagaacttac 34500

attgtgcatg gacagggtat cgcaatcagg cagcaccggg tgatcctcca ccagagaagc 34560

gcgggtctcg gtctcctcac agcgtggtaa gggggccggc cgatacgggt gatggcggga 34620

cgcggctgat cgtgttcgcg accgtgtcat gatgcagttg ctttcggaca ttttcgtact 34680

tgctgtagca gaacctggtc cgggcgctgc acaccgatcg ccggcggcgg tctcggcgct 34740

tggaacgctc ggtgttgaaa ttgtaaaaca gccactctct cagaccgtgc agcagatcta 34800

gggcctcagg agtgatgaag atcccatcat gcctgatggc tctgatcaca tcgaccaccg 34860

tggaatgggc cagacccagc cagatgatgc aattttgttg ggtttcggtg acggcggggg 34920

agggaagaac aggaagaacc atgattaact tttaatccaa acggtctcgg agtacttcaa 34980

aatgaagatc gcggagatgg cacctctcgc ccccgctgtg ttggtggaaa ataacagcca 35040

ggtcaaaggt gatacggttc tcgagatgtt ccacggtggc ttccagcaaa gcctccacgc 35100

gcacatccag aaacaagaca atagcgaaag cgggagggtt ctctaattcc tcaatcatca 35160

tgttacactc ctgcaccatc cccagataat tttcattttt ccagccttga atgattcgaa 35220

ctagttcctg aggtaaatcc aagccagcca tgataaagag ctcgcgcaga gcgccctcca 35280

ccggcattct taagcacacc ctcataattc caagatattc tgctcctggt tcacctgcag 35340

cagattgaca agcggaatat caaaatctct gccgcgatcc ctgagctcct ccctcagcaa 35400

taactgtaag tactctttca tatcctctcc gaaattttta gccataggac caccaggaat 35460

aagattaggg caagccacag tacagataaa ccgaagtcct ccccagtgag cattgccaaa 35520

tgcaagactg ctataagcat gctggctaga cccggtgata tcttccagat aactggacag 35580

aaaatcgccc aggcaatttt taagaaaatc aacaaaagaa aaatcctcca ggtggacgtt 35640

tagagcctcg ggaacaacga tgaagtaaat gcaagcggtg cgttccagca tggttagtta 35700

gctgatctgt agaaaaaaca aaaatgaaca ttaaaccatg ctagcctggc gaacaggtgg 35760

gtaaatcgtt ctctccagca ccaggcaggc cacggggtct ccggcgcgac cctcgtaaaa 35820

attgtcgcta tgattgaaaa ccatcacaga gagacgttcc cggtggccgg cgtgaatgat 35880

tcgacaagat gaatacaccc ccggaacatt ggcgtccgcg agtgaaaaaa agcgcccgag 35940

[0025] gaagcaataa ggcactacaa tgctcagtct caagtccagc aaagcgatgc catgcggatg 36000

aagcacaaaa ttctcaggtg cgtacaaaat gtaattactc ccctcctgca caggcagcaa 36060

agcccccgat ccctccaggt acacatacaa agcctcagcg tccatagctt accgagcagc 36120

agcacacaac aggcgcaaga gtcagagaaa ggctgagctc taacctgtcc acccgctctc 36180

tgctcaatat atagcccaga tctacactga cgtaaaggcc aaagtctaaa aatacccgcc 36240

aaataatcac acacgcccag cacacgccca gaaaccggtg acacactcaa aaaaatacgc 36300

gcacttcctc aaacgcccaa aactgccgtc atttccgggt tcccacgcta cgtcatcaaa 36360

acacgacttt caaattccgt cgaccgttaa aaacgtcacc cgccccgccc ctaacggtcg 36420

cccgtctctc agccaatcag cgccccgcat ccccaaattc aaacacctca tttgcatatt 36480

aacgcgcaca aaaagtttga ggtatattat tgatgatgg 36519

<210> 2
<211> 31588
<212> DNA
<213> 人工序列(Artificial Sequence)

<220>
<223> 人工序列的描述：合成多核苷酸

<400> 2

ccatcttcaa taatatacct caaacttttt gtgcgcgtta atatgcaaat gaggcgtttg 60

aatttgggga ggaagggcgg tgattggtcg agggatgagc gaccgttagg ggcggggcga 120

gtgacgtttt gatgacgtgg ttgcgaggag gagccagttt gcaagttctc gtgggaaaag 180

tgacgtcaaa cgaggtgtgg tttgaacacg gaaatactca attttcccgc gctctctgac 240

aggaaatgag gtgtttctgg gcggatgcaa gtgaaaacgg gccattttcg cgcgaaaact 300

gaatgaggaa gtgaaaatct gagtaatttc gcgtttatgg cagggaggag tatttgccga 360

gggccgagta gactttgacc gattacgtgg gggtttcgat taccgtgttt ttcacctaaa 420

tttccgcgta cggtgtcaaa gtccggtgtt tttacgtagg tgtcagctga tcgccagggt 480

atttaaacct gcgctctcca gtcaagaggc cactcttgag tgccagcgag aagagttttc 540

tcctccgcgc cgcgagtcag atctacactt tgaaagtagg gataacaggg taatgacatt 600

[0026] gattattgac tagttgttaa tagtaatcaa ttacggggtc attagttcat agcccatata 660

tggagttccg cgttacataa cttacggtaa atggcccgcc tggctgaccg cccaacgacc 720

cccgcccatt gacgtcaata atgacgtatg ttcccatagt aacgccaata gggactttcc 780

attgacgtca atgggtggag tatttacggt aaactgccca cttggcagta catcaagtgt 840

atcatatgcc aagtccgccc cctattgacg tcaatgacgg taaatggccc gcctggcatt 900

atgcccagta catgacctta cgggactttc ctacttggca gtacatctac gtattagtca 960

tcgctattac catggtgatg cggttttggc agtacaccaa tgggcgtgga tagcggtttg 1020

actcacgggg atttccaagt ctccacccca ttgacgtcaa tgggagtttg ttttggcacc 1080

aaaatcaacg ggactttcca aaatgtcgta ataaccccgc cccgttgacg caaatgggcg 1140

gtaggcgtgt acggtgggag gtctatataa gcagagctcg tttagtgaac cgtcagatcg 1200

cctggaacgc catccacgct gttttgacct ccatagaaga cagcgatcgc gccaccatgg 1260

ccgggatgtt ccaggcactg tccgaaggct gcacacccta tgatattaac cagatgctga 1320

atgtcctggg agaccaccag gtctctggcc tggagcagct ggagagcatc atcaacttcg 1380

agaagctgac cgagtggaca agctccaatg tgatgcctat cctgtcccca ctgaccaagg 1440

gcatcctggg cttcgtgttt accctgacag tgccttctga gcggggcctg tcttgcatca 1500

gcgaggcaga cgcaaccaca ccagagtccg ccaatctggg cgaggagatc ctgtctcagc 1560

tgtacctgtg gccccgggtg acatatcact ccccttctta cgcctatcac cagttcgagc 1620

ggagagccaa gtacaagaga cacttcccag gctttggcca gtctctgctg ttcggctacc 1680

ccgtgtacgt gttcggcgat tgcgtgcagg gcgactggga tgccatccgg tttagatact 1740

gcgcaccacc tggatatgca ctgctgaggt gtaacgacac caattattcc gccctgctgg 1800

cagtgggcgc cctggagggc cctcgcaatc aggattggct gggcgtgcca aggcagctgg 1860

tgacacgcat gcaggccatc cagaacgcag gcctgtgcac cctggtggca atgctggagg 1920

agacaatctt ctggctgcag gcctttctga tggccctgac cgacagcggc cccaagacaa 1980

acatcatcgt ggattcccag tacgtgatgg gcatctccaa gccttctttc caggagtttg 2040

tggactggga gaacgtgagc ccagagctga attccaccga tcagccattc tggcaggcag 2100

[0027] gaatcctggc aaggaacctg gtgcctatgg tggccacagt gcagggccag aatctgaagt 2160

accagggcca gagcctggtc atcagcgcct ccatcatcgt gtttaacctg ctggagctgg 2220

agggcgacta tcgggacgat ggcaacgtgt gggtgcacac cccactgagc cccagaacac 2280

tgaacgcctg ggtgaaggcc gtggaggaga agaagggcat cccagtgcac ctggagctgg 2340

cctccatgac caatatggag ctgatgtcta gcatcgtgca ccagcaggtg aggacatacg 2400

gacccgtgtt catgtgcctg ggaggcctgc tgaccatggt ggcaggagcc gtgtggctga 2460

cagtgcgggt gctggagctg ttcagagccg cccagctggc caacgatgtg gtgctgcaga 2520

tcatggagct gtgcggagca gcctttcgcc aggtgtgcca caccacagtg ccatggccca 2580

atgcctccct gacccccaag tggaacaatg agacaacaca gcctcagatc gccaactgta 2640

gcgtgtacga cttcttcgtg tggctgcact actatagcgt gagggatacc ctgtggcccc 2700

gcgtgacata ccacatgaat aagtacgcct atcacatgct ggagaggcgc gccaagtata 2760

agagaggccc tggcccaggc gcaaagtttg tggcagcatg gaccctgaag gccgccgccg 2820

gccccggccc cggccagtat atcaaggcta acagtaagtt cattggaatc acagagctgg 2880

gacccggacc tggataatga gtttaaactc ccatttaaat gtgagggtta atgcttcgag 2940

cagacatgat aagatacatt gatgagtttg gacaaaccac aactagaatg cagtgaaaaa 3000

aatgctttat ttgtgaaatt tgtgatgcta ttgctttatt tgtaaccatt ataagctgca 3060

ataaacaagt taacaacaac aattgcattc attttatgtt tcaggttcag ggggagatgt 3120

gggaggtttt ttaaagcaag taaaacctct acaaatgtgg taaaataact ataacggtcc 3180

taaggtagcg agtgagtagt gttctggggc gggggaggac ctgcatgagg gccagaataa 3240

ctgaaatctg tgcttttctg tgtgttgcag cagcatgagc ggaagcggct cctttgaggg 3300

aggggtattc agcccttatc tgacggggcg tctcccctcc tgggcgggag tgcgtcagaa 3360

tgtgatggga tccacggtgg acggccggcc cgtgcagccc gcgaactctt caaccctgac 3420

ctatgcaacc ctgagctctt cgtcgttgga cgcagctgcc gccgcagctg ctgcatctgc 3480

cgccagcgcc gtgcgcggaa tggccatggg cgccggctac tacggcactc tggtggccaa 3540

ctcgagttcc accaataatc ccgccagcct gaacgaggag aagctgttgc tgctgatggc 3600

[0028] ccagctcgag gccttgaccc agcgcctggg cgagctgacc cagcaggtgg ctcagctgca 3660

ggagcagacg cgggccgcgg ttgccacggt gaaatccaaa taaaaaatga atcaataaat 3720

aaacggagac ggttgttgat tttaacacag agtctgaatc tttatttgat ttttcgcgcg 3780

cggtaggccc tggaccaccg gtctcgatca ttgagcaccc ggtggatctt ttccaggacc 3840

cggtagaggt gggcttggat gttgaggtac atgggcatga gcccgtcccg ggggtggagg 3900

tagctccatt gcagggcctc gtgctcgggg gtggtgttgt aaatcaccca gtcatagcag 3960

gggcgcaggg catggtgttg cacaatatct ttgaggagga gactgatggc cacgggcagc 4020

cctttggtgt aggtgtttac aaatctgttg agctgggagg gatgcatgcg gggggagatg 4080

aggtgcatct tggcctggat cttgagattg gcgatgttac cgcccagatc ccgcctgggg 4140

ttcatgttgt gcaggaccac cagcacggtg tatccggtgc acttggggaa tttatcatgc 4200

aacttggaag ggaaggcgtg aaagaatttg gcgacgcctt tgtgcccgcc caggttttcc 4260

atgcactcat ccatgatgat ggcgatgggc ccgtgggcgg cggcctgggc aaagacgttt 4320

cgggggtcgg acacatcata gttgtggtcc tgggtgaggt catcataggc cattttaatg 4380

aatttggggc ggagggtgcc ggactggggg acaaaggtac cctcgatccc gggggcgtag 4440

ttcccctcac agatctgcat ctcccaggct ttgagctcgg agggggggat catgtccacc 4500

tgcggggcga taaagaacac ggtttccggg gcgggggaga tgagctgggc cgaaagcaag 4560

ttccggagca gctgggactt gccgcagccg gtggggccgt agatgacccc gatgaccggc 4620

tgcaggtggt agttgaggga gagacagctg ccgtcctccc ggaggagggg ggccacctcg 4680

ttcatcatct cgcgcacgtg catgttctcg cgcaccagtt ccgccaggag gcgctctccc 4740

cccagggata ggagctcctg gagcgaggcg aagtttttca gcggcttgag tccgtcggcc 4800

atgggcattt tggagagggt ttgttgcaag agttccaggc ggtcccagag ctcggtgatg 4860

tgctctacgg catctcgatc cagcagacct cctcgtttcg cgggttggga cggctgcggg 4920

agtagggcac cagacgatgg gcgtccagcg cagccagggt ccggtccttc cagggtcgca 4980

gcgtccgcgt cagggtggtc tccgtcacgg tgaaggggtg cgcgccgggc tgggcgcttg 5040

cgagggtgcg cttcaggctc atccggctgg tcgaaaaccg ctcccgatcg gcgccctgcg 5100

[0029] cgtcggccag gtagcaattg accatgagtt cgtagttgag cgcctcggcc gcgtggcctt 5160

tggcgcggag cttacctttg gaagtctgcc cgcaggcggg acagaggagg gacttgaggg 5220

cgtagagctt gggggcgagg aagacggact cgggggcgta ggcgtccgcg ccgcagtggg 5280

cgcagacggt ctcgcactcc acgagccagg tgaggtcggg ctggtcgggg tcaaaaacca 5340

gtttcccgcc gttctttttg atgcgtttct tacctttggt ctccatgagc tcgtgtcccc 5400

gctgggtgac aaagaggctg tccgtgtccc cgtagaccga ctttatgggc cggtcctcga 5460

gcggtgtgcc gcggtcctcc tcgtagagga accccgccca ctccgagacg aaagcccggg 5520

tccaggccag cacgaaggag gccacgtggg acgggtagcg gtcgttgtcc accagcgggt 5580

ccacctttтc cagggtatgc aaacacatgt ccccctcgtc cacatccagg aaggtgattg 5640

gcttgtaagt gtaggccacg tgaccggggg tcccggccgg gggggtataa aagggtgcgg 5700

gtccctgctc gtcctcactg tcttccggat cgctgtccag gagcgccagc tgttggggta 5760

ggtattccct ctcgaaggcg ggcatgacct cggcactcag gttgtcagtt tctagaaacg 5820

aggaggattt gatattgacg gtgccggcgg agatgccttt caagagcccc tcgtccatct 5880

ggtcagaaaa gacgatcttt ttgttgtcga gcttggtggc gaaggagccg tagagggcgt 5940

tggagaggag cttggcgatg gagcgcatgg tctggttttt ttccttgtcg gcgcgctcct 6000

tggcggcgat gttgagctgc acgtactcgc gcgccacgca cttccattcg gggaagacgg 6060

tggtcagctc gtcgggcacg attctgacct gccagccccg attatgcagg gtgatgaggt 6120

ccacactggt ggccacctcg ccgcgcaggg gctcattagt ccagcagagg cgtccgccct 6180

tgcgcgagca gaaggggggc agggggtcca gcatgacctc gtcgggggggg tcggcatcga 6240

tggtgaagat gccgggcagg aggtcggggt caaagtagct gatggaagtg gccagatcgt 6300

ccagggcagc ttgccattcg cgcacggcca gcgcgcgctc gtagggactg aggggcgtgc 6360

cccagggcat gggatgggta agcgcggagg cgtacatgcc gcagatgtcg tagacgtaga 6420

ggggctcctc gaggatgccg atgtaggtgg ggtagcagcg ccccccgcgg atgctggcgc 6480

gcacgtagtc atacagctcg tgcgaggggg cgaggagccc cgggcccagg ttggtgcgac 6540

tgggcttttc ggcgcggtag acgatctggc ggaaaatggc atgcgagttg gaggagatgg 6600

[0030] tgggcctttg gaagatgttg aagtgggcgt ggggcagtcc gaccgagtcg cggatgaagt 6660

gggcgtagga gtcttgcagc ttggcgacga gctcggcggt gactaggacg tccagagcgc 6720

agtagtcgag ggtctcctgg atgatgtcat acttgagctg tccctttgt ttccacagct 6780

cgcggttgag aaggaactct tcgcggtcct tccagtactc ttcgaggggg aacccgtcct 6840

gatctgcacg gtaagagcct agcatgtaga actggttgac ggccttgtag gcgcagcagc 6900

ccttctccac ggggagggcg taggcctggg cggccttgcg cagggaggtg tgcgtgaggg 6960

cgaaagtgtc cctgaccatg accttgagga actggtgctt gaagtcgata tcgtcgcagc 7020

ccccctgctc ccagagctgg aagtccgtgc gcttcttgta ggcggggttg ggcaaagcga 7080

aagtaacatc gttgaagagg atcttgcccg cgcggggcat aaagttgcga gtgatgcgga 7140

aaggttgggg cacctcggcc cggttgttga tgacctgggc ggcgagcacg atctcgtcga 7200

agccgttgat gttgtggccc acgatgtaga gttccacgaa tcgcggacgg cccttgacgt 7260

ggggcagttt cttgagctcc tcgtaggtga gctcgtcggg gtcgctgagc ccgtgctgct 7320

cgagcgccca gtcggcgaga tgggggttgg cgcggaggaa ggaagtccag agatccacgg 7380

ccagggcggt ttgcagacgg tcccggtact gacggaactg ctgcccgacg gccatttttt 7440

cgggggtgac gcagtagaag gtgcgggggt ccccgtgcca gcgatcccat ttgagctgga 7500

gggcgagatc gagggcgagc tcgacgagcc ggtcgtcccc ggagagtttc atgaccagca 7560

tgaaggggac gagctgcttg ccgaaggacc ccatccaggt gtaggtttcc acatcgtagg 7620

tgaggaagag cctttcggtg cgaggatgcg agccgatggg gaagaactgg atctcctgcc 7680

accaattgga ggaatggctg ttgatgtgat ggaagtagaa atgccgacgg cgcgccgaac 7740

actcgtgctt gtgtttatac aagcggccac agtgctcgca acgctgcacg ggatgcacgt 7800

gctgcacgag ctgtacctga gttcctttga cgaggaattt cagtgggaag tggagtcgtg 7860

gcgcctgcat ctcgtgctgt actacgtcgt ggtggtcggc ctggccctct tctgcctcga 7920

tggtggtcat gctgacgagc ccgcgcggga ggcaggtcca gacctcggcg cgagcgggtc 7980

ggagagcgag gacgagggcg cgcaggccgg agctgtccag ggtcctgaga cgctgcggag 8040

tcaggtcagt gggcagcggc ggcgcgcggt tgacttgcag gagtttttcc agggcgcgcg 8100

[0031] ggaggtccag atggtacttg atctccaccg cgccattggt ggcgacgtcg atggcttgca 8160

gggtcccgtg cccctggggt gtgaccaccg tcccccgttt cttcttgggc ggctggggcg 8220

acgggggcgg tgcctcttcc atggttagaa gcggcggcga ggacgcgcgc cgggcggcag 8280

gggcggctcg gggcccggag gcaggggcgg caggggcacg tcggcgccgc gcgcgggtag 8340

gttctggtac tgcgcccgga gaagactggc gtgagcgacg acgcgacggt tgacgtcctg 8400

gatctgacgc ctctgggtga aggccacggg acccgtgagt ttgaacctga aagagagttc 8460

gacagaatca atctcggtat cgttgacggc ggcctgccgc aggatctctt gcacgtcgcc 8520

cgagttgtcc tggtaggcga tctcggtcat gaactgctcg atctcctcct cttgaaggtc 8580

tccgcggccg gcgcgctcca cggtggccgc gaggtcgttg gagatgcggc ccatgagctg 8640

cgagaaggcg ttcatgcccg cctcgttcca gacgcggctg tagaccacga cgccctcggg 8700

atcgcgggcg cgcatgacca cctgggcgag gttgagctcc acgtggcgcg tgaagaccgc 8760

gtagttgcag aggcgctggt agaggtagtt gagcgtggtg gcgatgtgct cggtgacgaa 8820

gaaatacatg atccagcggc ggagcggcat ctcgctgacg tcgcccagcg cctccaaacg 8880

ttccatggcc tcgtaaaagt ccacggcgaa gttgaaaaac tgggagttgc gcgccgagac 8940

ggtcaactcc tcctccagaa gacggatgag ctcggcgatg gtggcgcgca cctcgcgctc 9000

gaaggccccc gggagttcct ccacttcctc ttcttcctcc tccactaaca tctcttctac 9060

ttcctcctca ggcggcagtg gtggcggggg agggggcctg cgtcgccggc ggcgcacggg 9120

cagacggtcg atgaagcgct cgatggtctc gccgcgccgg cgtcgcatgg tctcggtgac 9180

ggcgcgcccg tcctcgcggg gccgcagcgt gaagacgccg ccgcgcatct ccaggtggcc 9240

gggggggtcc ccgttgggca gggagagggc gctgacgatg catcttatca attgccccgt 9300

agggactccg cgcaaggacc tgagcgtctc gagatccacg ggatctgaaa accgctgaac 9360

gaaggcttcg agccagtcgc agtcgcaagg taggctgagc acggtttctt ctggcgggtc 9420

atgttggttg ggagcggggc gggcgatgct gctggtgatg aagttgaaat aggcggttct 9480

gagacggcgg atggtggcga ggagcaccag gtctttgggc ccggcttgct ggatgcgcag 9540

acggtcggcc atgccccagg cgtggtcctg acacctggcc aggtccttgt agtagtcctg 9600

[0032] catgagccgc tccacgggca cctcctcctc gcccgcgcgg ccgtgcatgc gcgtgagccc 9660

gaagccgcgc tggggctgga cgagcgccag gtcggcgacg acgcgctcgg cgaggatggc 9720

ttgctggatc tgggtgaggg tggtctggaa gtcatcaaag tcgacgaagc ggtggtaggc 9780

tccggtgttg atggtgtagg agcagttggc catgacggac cagttgacgg tctggtggcc 9840

cggacgcacg agctcgtggt acttgaggcg cgagtaggcg cgcgtgtcga agatgtagtc 9900

gttgcaggtg cgcaccaggt actggtagcc gatgaggaag tgcggcggcg gctggcggta 9960

gagcggccat cgctcggtgg cgggggcgcc gggcgcgagg tcctcgagca tggtgcggtg 10020

gtagccgtag atgtacctgg acatccaggt gatgccggcg gcggtggtgg aggcgcgcgg 10080

gaactcgcgg acgcggttcc agatgttgcg cagcggcagg aagtagttca tggtgggcac 10140

ggtctggccc gtgaggcgcg cgcagtcgtg gatgctctat acgggcaaaa acgaaagcgg 10200

tcagcggctc gactccgtgg cctggaggct aagcgaacgg gttgggctgc gcgtgtaccc 10260

cggttcgaat ctcgaatcag gctggagccg cagctaacgt ggtattggca ctcccgtctc 10320

gacccaagcc tgcaccaacc ctccaggata cggaggcggg tcgttttgca actttttttt 10380

ggaggccgga tgagactagt aagcgcggaa agcggccgac cgcgatggct cgctgccgta 10440

gtctggagaa gaatcgccag ggttgcgttg cggtgtgccc cggttcgagg ccggccggat 10500

tccgcggcta acgagggcgt ggctgccccg tcgtttccaa gaccccatag ccagccgact 10560

tctccagtta cggagcgagc ccctcttttg ttttgtttgt ttttgccaga tgcatcccgt 10620

actgcggcag atgcgccccc accaccctcc accgcaacaa cagccccctc cacagccggc 10680

gcttctgccc ccgccccagc agcaacttcc agccacgacc gccgcggccg ccgtgagcgg 10740

ggctggacag agttatgatc accagctggc cttggaagag ggcgaggggc tggcgcgcct 10800

gggggcgtcg tcgccggagc ggcacccgcg cgtgcagatg aaaagggacg ctcgcgaggc 10860

ctacgtgccc aagcagaacc tgttcagaga caggagcggc gaggagcccg aggagatgcg 10920

cgcggcccgg ttccacgcgg ggcgggagct gcggcgcggc ctggaccgaa agagggtgct 10980

gagggacgag gatttcgagg cggacgagct gacggggatc agccccgcgc gcgcgcacgt 11040

ggccgcggcc aacctggtca cggcgtacga gcagaccgtg aaggaggaga gcaacttcca 11100

[0033] aaaatccttc aacaaccacg tgcgcaccct gatcgcgcgc gaggaggtga ccctgggcct 11160

gatgcacctg tgggacctgc tggaggccat cgtgcagaac cccaccagca agccgctgac 11220

ggcgcagctg ttcctggtgg tgcagcatag tcgggacaac gaagcgttca gggaggcgct 11280

gctgaatatc accgagcccg agggccgctg gctcctggac ctggtgaaca ttctgcagag 11340

catcgtggtg caggagcgcg ggctgccgct gtccgagaag ctggcggcca tcaacttctc 11400

ggtgctgagt ttgggcaagt actacgctag gaagatctac aagaccccgt acgtgcccat 11460

agacaaggag gtgaagatcg acgggtttta catgcgcatg accctgaaag tgctgaccct 11520

gagcgacgat ctgggggtgt accgcaacga caggatgcac cgtgcggtga gcgccagcag 11580

gcggcgcgag ctgagcgacc aggagctgat gcatagtctg cagcgggccc tgaccggggc 11640

cgggaccgag ggggagagct actttgacat gggcgcggac ctgcactggc agcccagccg 11700

ccgggccttg gaggcggcgg caggacccta cgtagaagag gtggacgatg aggtggacga 11760

ggagggcgag tacctggaag actgatggcg cgaccgtatt tttgctagat gcaacaacaa 11820

cagccacctc ctgatcccgc gatgcgggcg gcgctgcaga gccagccgtc cggcattaac 11880

tcctcggacg attggaccca ggccatgcaa cgcatcatgg cgctgacgac ccgcaacccc 11940

gaagccttta gacagcagcc ccaggccaac cggctctcgg ccatcctgga ggccgtggtg 12000

ccctcgcgct ccaaccccac gcacgagaag gtcctggcca tcgtgaacgc gctggtggag 12060

aacaaggcca tccgcggcga cgaggccggc ctggtgtaca acgcgctgct ggagcgcgtg 12120

gcccgctaca acagcaccaa cgtgcagacc aacctggacc gcatggtgac cgacgtgcgc 12180

gaggccgtgg cccagcgcga gcggttccac cgcgagtcca acctgggatc catggtggcg 12240

ctgaacgcct tcctcagcac ccagcccgcc aacgtgcccc ggggccagga ggactacacc 12300

aacttcatca gcgccctgcg cctgatggtg accgaggtgc cccagagcga ggtgtaccag 12360

tccgggccgg actacttctt ccagaccagt cgccagggct tgcagaccgt gaacctgagc 12420

caggctttca agaacttgca gggcctgtgg ggcgtgcagg ccccggtcgg ggaccgcgcg 12480

acggtgtcga gcctgctgac gccgaactcg cgcctgctgc tgctgctggt ggcccccttc 12540

acggacagcg gcagcatcaa ccgcaactcg tacctgggct acctgattaa cctgtaccgc 12600

[0034] gaggccatcg gccaggcgca cgtggacgag cagacctacc aggagatcac ccacgtgagc 12660

cgcgccctgg gccaggacga cccgggcaac ctggaagcca ccctgaactt tttgctgacc 12720

aaccggtcgc agaagatccc gccccagtac gcgctcagca ccgaggagga gcgcatcctg 12780

cgttacgtgc agcagagcgt gggcctgttc ctgatgcagg agggggccac ccccagcgcc 12840

gcgctcgaca tgaccgcgcg caacatggag cccagcatgt acgccagcaa ccgcccgttc 12900

atcaataaac tgatggacta cttgcatcgg gcggccgcca tgaactctga ctatttcacc 12960

aacgccatcc tgaatcccca ctggctcccg ccgccggggt tctacacggg cgagtacgac 13020

atgcccgacc ccaatgacgg gttcctgtgg gacgatgtgg acagcagcgt gttctccccc 13080

cgaccgggtg ctaacgagcg ccccttgtgg aagaaggaag gcagcgaccg acgcccgtcc 13140

tcggcgctgt ccggccgcga gggtgctgcc gcggcggtgc ccgaggccgc cagtcctttc 13200

ccgagcttgc ccttctcgct gaacagtatc cgcagcagcg agctgggcag gatcacgcgc 13260

ccgcgcttgc tgggcgaaga ggagtacttg aatgactcgc tgttgagacc cgagcgggag 13320

aagaacttcc ccaataacgg gatagaaagc ctggtggaca agatgagccg ctggaagacg 13380

tatgcgcagg agcacaggga cgatccccgg gcgtcgcagg gggccacgag ccggggcagc 13440

gccgcccgta aacgccggtg gcacgacagg cagcggggac agatgtggga cgatgaggac 13500

tccgccgacg acagcagcgt gttggacttg ggtgggagtg gtaacccgtt cgctcacctg 13560

cgcccccgta tcgggcgcat gatgtaagag aaaccgaaaa taaatgatac tcaccaaggc 13620

catggcgacc agcgtgcgtt cgtttcttct ctgttgttgt tgtatctagt atgatgaggc 13680

gtgcgtaccc ggagggtcct cctccctcgt acgagagcgt gatgcagcag gcgatggcgg 13740

cggcggcgat gcagcccccg ctggaggctc cttacgtgcc cccgcggtac ctggcgccta 13800

cggaggggcg gaacagcatt cgttactcgg agctggcacc cttgtacgat accacccggt 13860

tgtacctggt ggacaacaag tcggcggaca tcgcctcgct gaactaccag aacgaccaca 13920

gcaacttcct gaccaccgtg gtgcagaaca atgacttcac ccccacggag gccagcaccc 13980

agaccatcaa ctttgacgag cgctcgcggt ggggcggcca gctgaaaacc atcatgcaca 14040

ccaacatgcc caacgtgaac gagttcatgt acagcaacaa gttcaaggcg cgggtgatgg 14100

[0035] tctcccgcaa gacccccaat ggggtgacag tgacagagga ttatgatggt agtcaggatg 14160

agctgaagta tgaatgggtg gaatttgagc tgcccgaagg caacttctcg gtgaccatga 14220

ccatcgacct gatgaacaac gccatcatcg acaattactt ggcggtgggg cggcagaacg 14280

gggtgctgga gagcgacatc ggcgtgaagt tcgacactag gaacttcagg ctgggctggg 14340

accccgtgac cgagctggtc atgcccgggg tgtacaccaa cgaggctttc catcccgata 14400

ttgtcttgct gcccggctgc ggggtggact tcaccgagag ccgcctcagc aacctgctgg 14460

gcattcgcaa gaggcagccc ttccaggaag gcttccagat catgtacgag gatctggagg 14520

ggggcaacat ccccgcgctc ctggatgtcg acgcctatga gaaaagcaag gaggatgcag 14580

cagctgaagc aactgcagcc gtagctaccg cctctaccga ggtcaggggc gataattttg 14640

caagcgccgc agcagtggca gcggccgagg cggctgaaac cgaaagtaag atagtcattc 14700

agccggtgga gaaggatagc aagaacagga gctacaacgt actaccggac aagataaaca 14760

ccgcctaccg cagctggtac ctagcctaca actatggcga ccccgagaag ggcgtgcgct 14820

cctggacgct gctcaccacc tcggacgtca cctgcggcgt ggagcaagtc tactggtcgc 14880

tgcccgacat gatgcaagac ccggtcacct tccgctccac gcgtcaagtt agcaactacc 14940

cggtggtggg cgccgagctc ctgcccgtct actccaagag cttcttcaac gagcaggccg 15000

tctactcgca gcagctgcgc gccttcacct cgcttacgca cgtcttcaac cgcttccccg 15060

agaaccagat cctcgtccgc ccgcccgcgc ccaccattac caccgtcagt gaaaacgttc 15120

ctgctctcac agatcacggg accctgccgc tgcgcagcag tatccgggga gtccagcgcg 15180

tgaccgttac tgacgccaga cgccgcacct gcccctacgt ctacaaggcc ctgggcatag 15240

tcgcgccgcg cgtcctctcg agccgcacct tctaaatgtc cattctcatc tcgcccagta 15300

ataacaccgg ttggggcctg cgcgcgccca gcaagatgta cggaggcgct cgccaacgct 15360

ccacgcaaca ccccgtgcgc gtgcgcgggc acttccgcgc tccctggggc gccctcaagg 15420

gccgcgtgcg gtcgcgcacc accgtcgacg acgtgatcga ccaggtggtg gccgacgcgc 15480

gcaactacac ccccgccgcc gcgcccgtct ccaccgtgga cgccgtcatc gacagcgtgg 15540

tggccgacgc gcgccggtac gcccgcgcca agagccggcg gcggcgcatc gcccggcggc 15600

[0036] accggagcac ccccgccatg cgcgcggcgc gagccttgct gcgcagggcc aggcgcacgg 15660

gacgcagggc catgctcagg gcggccagac gcgcggcttc aggcgccagc gccggcagga 15720

cccggagacg cgcggccacg gcggcggcag cggccatcgc cagcatgtcc cgcccgcggc 15780

gagggaacgt gtactgggtg cgcgacgccg ccaccggtgt gcgcgtgccc gtgcgcaccc 15840

gcccccctcg cacttgaaga tgttcacttc gcgatgttga tgtgtcccag cggcgaggag 15900

gatgtccaag cgcaaattca aggaagagat gctccaggtc atcgcgcctg agatctacgg 15960

ccctgcggtg gtgaaggagg aaagaaagcc ccgcaaaatc aagcgggtca aaaaggacaa 16020

aaaggaagaa gaaagtgatg tggacggatt ggtggagttt gtgcgcgagt tcgccccccg 16080

gcggcgcgtg cagtggcgcg ggcggaaggt gcaaccggtg ctgagacccg gcaccaccgt 16140

ggtcttcacg cccggcgagc gctccggcac cgcttccaag cgctcctacg acgaggtgta 16200

cggggatgat gatattctgg agcaggcggc cgagcgcctg ggcgagtttg cttacggcaa 16260

gcgcagccgt tccgcaccga aggaagaggc ggtgtccatc ccgctggacc acggcaaccc 16320

cacgccgagc ctcaagcccg tgaccttgca gcaggtgctg ccgaccgcgg cgccgcgccg 16380

ggggttcaag cgcgagggcg aggatctgta ccccaccatg cagctgatgg tgcccaagcg 16440

ccagaagctg gaagacgtgc tggagaccat gaaggtggac ccggacgtgc agcccgaggt 16500

caaggtgcgg cccatcaagc aggtggcccc gggcctgggc gtgcagaccg tggacatcaa 16560

gattcccacg gagcccatgg aaacgcagac cgagcccatg atcaagccca gcaccagcac 16620

catggaggtg cagacggatc cctggatgcc atcggctcct agtcgaagac cccggcgcaa 16680

gtacggcgcg gccagcctgc tgatgcccaa ctacgcgctg catccttcca tcatccccac 16740

gccgggctac cgcggcacgc gcttctaccg cggtcatacc agcagccgcc gccgcaagac 16800

caccactcgc cgccgccgtc gccgcaccgc cgctgcaacc acccctgccg ccctggtgcg 16860

gagagtgtac cgccgcggcc gcgcacctct gaccctgccg cgcgcgcgct accacccgag 16920

catcgccatt taaactttcg cctgctttgc agatcaatgg ccctcacatg ccgccttcgc 16980

gttcccatta cgggctaccg aggaagaaaa ccgcgccgta gaaggctggc ggggaacggg 17040

atgcgtcgcc accaccaccg gcggcggcgc gccatcagca agcggttggg gggaggcttc 17100

[0037] ctgcccgcgc tgatccccat catcgccgcg gcgatcgggg cgatccccgg cattgcttcc 17160

gtggcggtgc aggcctctca gcgccactga gacacacttg gaaacatctt gtaataaacc 17220

aatggactct gacgctcctg gtcctgtgat gtgttttcgt agacagatgg aagacatcaa 17280

tttttcgtcc ctggctccgc gacacggcac gcggccgttc atgggcacct ggagcgacat 17340

cggcaccagc caactgaacg ggggcgcctt caattggagc agtctctgga gcgggcttaa 17400

gaatttcggg tccacgctta aaacctatgg cagcaaggcg tggaacagca ccacagggca 17460

ggcgctgagg gataagctga aagagcagaa cttccagcag aaggtggtcg atgggctcgc 17520

ctcgggcatc aacggggtgg tggacctggc caaccaggcc gtgcagcggc agatcaacag 17580

ccgcctggac ccggtgccgc ccgccggctc cgtggagatg ccgcaggtgg aggaggagct 17640

gcctcccctg gacaagcggg gcgagaagcg accccgcccc gatgcggagg agacgctgct 17700

gacgcacacg gacgagccgc ccccgtacga ggaggcggtg aaactgggtc tgcccaccac 17760

gcggcccatc gcgcccctgg ccaccggggt gctgaaaccc gaaaagcccg cgaccctgga 17820

cttgcctcct ccccagcctt cccgcccctc tacagtggct aagcccctgc cgccggtggc 17880

cgtggcccgc gcgcgacccg ggggcaccgc ccgccctcat gcgaactggc agagcactct 17940

gaacagcatc gtgggtctgg gagtgcagag tgtgaagcgc cgccgctgct attaaaccta 18000

ccgtagcgct taacttgctt gtctgtgtgt gtatgtatta tgtcgccgcc gccgctgtcc 18060

accagaagga ggagtgaaga ggcgcgtcgc cgagttgcaa gatggccacc ccatcgatgc 18120

tgccccagtg ggcgtacatg cacatcgccg gacaggacgc ttcggagtac ctgagtccgg 18180

gtctggtgca gtttgcccgc gccacagaca cctacttcag tctggggaac aagtttagga 18240

accccacggt ggcgcccacg cacgatgtga ccaccgaccg cagccagcgg ctgacgctgc 18300

gcttcgtgcc cgtggaccgc gaggacaaca cctactcgta caaagtgcgc tacacgctgg 18360

ccgtgggcga caaccgcgtg ctggacatgg ccagcaccta ctttgacatc cgcggcgtgc 18420

tggatcgggg ccctagcttc aaaccctact ccggcaccgc ctacaacagt ctggccccca 18480

agggagcacc caacacttgt cagtggacat ataaagccga tggtgaaact gccacagaaa 18540

aaacctatac atatggaaat gcacccgtgc agggcattaa catcacaaaa gatggtattc 18600

[0038] aacttggaac tgacaccgat gatcagccaa tctacgcaga taaaacctat cagcctgaac 18660

ctcaagtggg tgatgctgaa tggcatgaca tcactggtac tgatgaaaag tatggaggca 18720

gagctcttaa gcctgatacc aaaatgaagc cttgttatgg ttcttttgcc aagcctacta 18780

ataaagaagg aggtcaggca aatgtgaaaa caggaacagg cactactaaa gaatatgaca 18840

tagacatggc tttctttgac aacagaagtg cggctgctgc tggcctagct ccagaaattg 18900

ttttgtatac tgaaaatgtg gatttggaaa ctccagatac ccatattgta tacaaagcag 18960

gcacagatga cagcagctct tctattaatt tgggtcagca agccatgccc aacagaccta 19020

actacattgg tttcagagac aactttatcg ggctcatgta ctacaacagc actggcaata 19080

tgggggtgct ggccggtcag gcttctcagc tgaatgctgt ggttgacttg caagacagaa 19140

acaccgagct gtcctaccag ctcttgcttg actctctggg tgacagaacc cggtatttca 19200

gtatgtggaa tcaggcggtg gacagctatg atcctgatgt gcgcattatt gaaaatcatg 19260

gtgtggagga tgaacttccc aactattgtt tccctctgga tgctgttggc agaacagata 19320

cttatcaggg aattaaggct aatggaactg atcaaaccac atggaccaaa gatgacagtg 19380

tcaatgatgc taatgagata ggcaagggta atccattcgc catggaaatc aacatccaag 19440

ccaacctgtg gaggaacttc ctctacgcca acgtggccct gtacctgccc gactcttaca 19500

agtacacgcc ggccaatgtt accctgccca ccaacaccaa cacctacgat tacatgaacg 19560

gccgggtggt ggcgccctcg ctggtggact cctacatcaa catcggggcg cgctggtcgc 19620

tggatcccat ggacaacgtg aaccccttca accaccaccg caatgcgggg ctgcgctacc 19680

gctccatgct cctgggcaac gggcgctacg tgcccttcca catccaggtg ccccagaaat 19740

ttttcgccat caagagcctc ctgctcctgc ccgggtccta cacctacgag tggaacttcc 19800

gcaaggacgt caacatgatc ctgcagagct ccctcggcaa cgacctgcgc acggacgggg 19860

cctccatctc cttcaccagc atcaacctct acgccacctt cttccccatg gcgcacaaca 19920

cggcctccac gctcgaggcc atgctgcgca acgacaccaa cgaccagtcc ttcaacgact 19980

acctctcggc ggccaacatg ctctacccca tcccggccaa cgccaccaac gtgcccatct 20040

ccatcccctc gcgcaactgg gccgccttcc gcggctggtc cttcacgcgt ctcaagacca 20100

[0039] aggagacgcc ctcgctgggc tccgggttcg acccctactt cgtctactcg ggctccatcc 20160

cctacctcga cggcaccttc tacctcaacc acaccttcaa gaaggtctcc atcaccttcg 20220

actcctccgt cagctggccc ggcaacgacc ggctcctgac gcccaacgag ttcgaaatca 20280

agcgcaccgt cgacggcgag ggctacaacg tggcccagtg caacatgacc aaggactggt 20340

tcctggtcca gatgctggcc cactacaaca tcggctacca gggcttctac gtgcccgagg 20400

gctacaagga ccgcatgtac tccttcttcc gcaacttcca gcccatgagc cgccaggtgg 20460

tggacgaggt caactacaag gactaccagg ccgtcaccct ggcctaccag cacaacaact 20520

cgggcttcgt cggctacctc gcgcccacca tgcgccaggg ccagccctac cccgccaact 20580

acccctaccc gctcatcggc aagagcgccg tcaccagcgt cacccagaaa aagttcctct 20640

gcgacagggt catgtggcgc atccccttct ccagcaactt catgtccatg ggcgcgctca 20700

ccgacctcgg ccagaacatg ctctatgcca actccgccca cgcgctagac atgaatttcg 20760

aagtcgaccc catggatgag tccacccttc tctatgttgt cttcgaagtc ttcgacgtcg 20820

tccgagtgca ccagccccac cgcggcgtca tcgaggccgt ctacctgcgc acccccttct 20880

cggccggtaa cgccaccacc taagctcttg cttcttgcaa gccatggccg cgggctccgg 20940

cgagcaggag ctcagggcca tcatccgcga cctgggctgc gggccctact tcctgggcac 21000

cttcgataag cgcttcccgg gattcatggc cccgcacaag ctggcctgcg ccatcgtcaa 21060

cacggccggc cgcgagaccg ggggcgagca ctggctggcc ttcgcctgga acccgcgctc 21120

gaacacctgc tacctcttcg acccccttcgg gttctcggac gagcgcctca agcagatcta 21180

ccagttcgag tacgagggcc tgctgcgccg cagcgccctg gccaccgagg accgctgcgt 21240

caccctggaa aagtccaccc agaccgtgca gggtccgcgc tcggccgcct gcgggctctt 21300

ctgctgcatg ttcctgcacg ccttcgtgca ctggcccgac cgccccatgg acaagaaccc 21360

caccatgaac ttgctgacgg gggtgcccaa cggcatgctc cagtcgcccc aggtggaacc 21420

caccctgcgc cgcaaccagg aggcgctcta ccgcttcctc aactcccact ccgcctactt 21480

tcgctcccac cgcgcgcgca tcgagaaggc caccgccttc gaccgcatga atcaagacat 21540

gtaaaccgtg tgtgtatgtt aaatgtcttt aataaacagc actttcatgt tacacatgca 21600

[0040] tctgagatga tttatttaga aatcgaaagg gttctgccgg gtctcggcat ggcccgcggg 21660

cagggacacg ttgcggaact ggtacttggc cagccacttg aactcgggga tcagcagttt 21720

gggcagcggg gtgtcgggga aggagtcggt ccacagcttc cgcgtcagtt gcagggcgcc 21780

cagcaggtcg ggcgcggaga tcttgaaatc gcagttggga cccgcgttct gcgcgcggga 21840

gttgcggtac acggggttgc agcactggaa caccatcagg gccgggtgct tcacgctcgc 21900

cagcaccgtc gcgtcggtga tgctctccac gtcgaggtcc tcggcgttgg ccatcccgaa 21960

gggggtcatc ttgcaggtct gccttcccat ggtgggcacg cacccgggct tgtggttgca 22020

atcgcagtgc agggggatca gcatcatctg ggcctggtcg gcgttcatcc ccgggtacat 22080

ggccttcatg aaagcctcca attgcctgaa cgcctgctgg gccttggctc cctcggtgaa 22140

gaagaccccg caggacttgc tagagaactg gttggtggcg cacccggcgt cgtgcacgca 22200

gcagcgcgcg tcgttgttgg ccagctgcac cacgctgcgc ccccagcggt tctgggtgat 22260

cttggcccgg tcggggttct ccttcagcgc gcgctgcccg ttctcgctcg ccacatccat 22320

ctcgatcatg tgctccttct ggatcatggt ggtcccgtgc aggcaccgca gcttgccctc 22380

ggcctcggtg cacccgtgca gccacagcgc gcacccggtg cactcccagt tcttgtgggc 22440

gatctgggaa tgcgcgtgca cgaagccctg caggaagcgg cccatcatgg tggtcagggt 22500

cttgttgcta gtgaaggtca gcggaatgcc gcggtgctcc tcgttgatgt acaggtggca 22560

gatgcggcgg tacacctcgc cctgctcggg catcagctgg aagttggctt tcaggtcggt 22620

ctccacgcgg tagcggtcca tcagcatagt catgatttcc ataccttct cccaggccga 22680

gacgatgggc aggctcatag ggttcttcac catcatctta gcgctagcag ccgcggccag 22740

ggggtcgctc tcgtccaggg tctcaaagct ccgcttgccg tccttctcgg tgatccgcac 22800

cgggggggtag ctgaagccca cggccgccag ctcctcctcg gcctgtcttt cgtcctcgct 22860

gtcctggctg acgtcctgca ggaccacatg cttggtcttg cggggtttct tcttgggcgg 22920

cagcggcggc ggagatgttg gagatggcga gggggagcgc gagttctcgc tcaccactac 22980

tatctcttcc tcttcttggt ccgaggccac gcggcggtag gtatgtctct tcgggggcag 23040

aggcggaggc gacgggctct cgccgccgcg acttggcgga tggctggcag agcccccttcc 23100

[0041] gcgttcgggg gtgcgctccc ggcggcgctc tgactgactt cctccgcggc cggccattgt 23160

gttctcctag ggaggaacaa caagcatgga gactcagcca tcgccaacct cgccatctgc 23220

ccccaccgcc gacgagaagc agcagcagca gaatgaaagc ttaaccgccc cgccgcccag 23280

ccccgccacc tccgacgcgg ccgtcccaga catgcaagag atggaggaat ccatcgagat 23340

tgacctgggc tatgtgacgc ccgcggagca cgaggaggag ctggcagtgc gcttttcaca 23400

agaagagata caccaagaac agccagagca ggaagcagag aatgagcaga gtcaggctgg 23460

gctcgagcat gacggcgact acctccacct gagcgggggg gaggacgcgc tcatcaagca 23520

tctggcccgg caggccacca tcgtcaagga tgcgctgctc gaccgcaccg aggtgcccct 23580

cagcgtggag gagctcagcc gcgcctacga gttgaacctc ttctcgccgc gcgtgccccc 23640

caagcgccag cccaatggca cctgcgagcc caacccgcgc ctcaacttct acccggtctt 23700

cgcggtgccc gaggccctgg ccacctacca catctttttc aagaaccaaa agatccccgt 23760

ctcctgccgc gccaaccgca cccgcgccga cgccctttc aacctgggtc ccggcgcccg 23820

cctacctgat atcgcctcct tggaagaggt tcccaagatc ttcgagggtc tgggcagcga 23880

cgagactcgg gccgcgaacg ctctgcaagg agaaggagga gagcatgagc accacagcgc 23940

cctggtcgag ttggaaggcg acaacgcgcg gctggcggtg ctcaaacgca cggtcgagct 24000

gacccatttc gcctacccgg ctctgaacct gccccccaaa gtcatgagcg cggtcatgga 24060

ccaggtgctc atcaagcgcg cgtcgcccat ctccgaggac gagggcatgc aagactccga 24120

ggagggcaag cccgtggtca gcgacgagca gctggcccgg tggctgggtc ctaatgctag 24180

tccccagagt ttggaagagc ggcgcaaact catgatggcc gtggtcctgg tgaccgtgga 24240

gctggagtgc ctgcgccgct tcttcgccga cgcggagacc ctgcgcaagg tcgaggagaa 24300

cctgcactac ctcttcaggc acgggttcgt gcgccaggcc tgcaagatct ccaacgtgga 24360

gctgaccaac ctggtctcct acatgggcat cttgcacgag aaccgcctgg ggcagaacgt 24420

gctgcacacc accctgcgcg gggaggcccg gcgcgactac atccgcgact gcgtctacct 24480

ctacctctgc cacacctggc agacgggcat gggcgtgtgg cagcagtgtc tggaggagca 24540

gaacctgaaa gagctctgca agctcctgca gaagaacctc aagggtctgt ggaccgggtt 24600

[0042] cgacgagcgc accaccgcct cggacctggc cgacctcatt ttccccgagc gcctcaggct 24660

gacgctgcgc aacggcctgc ccgactttat gagccaaagc atgttgcaaa actttcgctc 24720

tttcatcctc gaacgctccg gaatcctgcc cgccacctgc tccgcgctgc cctcggactt 24780

cgtgccgctg accttccgcg agtgcccccc gccgctgtgg agccactgct acctgctgcg 24840

cctggccaac tacctggcct accactcgga cgtgatcgag gacgtcagcg gcgagggcct 24900

gctcgagtgc cactgccgct gcaacctctg cacgccgcac cgctccctgg cctgcaaccc 24960

ccagctgctg agcgagaccc agatcatcgg caccttcgag ttgcaagggc ccagcgaagg 25020

cgagggttca gccgccaagg ggggtctgaa actcaccccg gggctgtgga cctcggccta 25080

cttgcgcaag ttcgtgcccg aggactacca tcccttcgag atcaggttct acgaggacca 25140

atcccatccg cccaaggccg agctgtcggc ctgcgtcatc acccaggggg cgatcctggc 25200

ccaattgcaa gccatccaga aatcccgcca agaattcttg ctgaaaaagg gccgcggggt 25260

ctacctcgac ccccagaccg gtgaggagct caaccccggc ttccccccagg atgccccgag 25320

gaaacaagaa gctgaaagtg gagctgccgc ccgtggagga tttggaggaa gactgggaga 25380

acagcagtca ggcagaggag gaggagatgg aggaagactg ggacagcact caggcagagg 25440

aggacagcct gcaagacagt ctggaggaag acgaggagga ggcagaggag gaggtggaag 25500

aagcagccgc cgccagaccg tcgtcctcgg cgggggagaa agcaagcagc acggatacca 25560

tctccgctcc gggtcggggt cccgctcgac cacacagtag atgggacgag accggacgat 25620

tcccgaaccc caccacccag accggtaaga aggagcggca gggatacaag tcctggcggg 25680

ggcacaaaaa cgccatcgtc tcctgcttgc aggcctgcgg gggcaacatc tccttcaccc 25740

ggcgctacct gctcttccac cgcggggtga actttccccg caacatcttg cattactacc 25800

gtcacctcca cagcccctac tacttccaag aagaggcagc agcagcagaa aaagaccagc 25860

agaaaaccag cagctagaaa atccacagcg gcggcagcag gtggactgag gatcgcggcg 25920

aacgagccgg cgcaaacccg ggagctgagg aaccggatct ttcccaccct ctatgccatc 25980

ttccagcaga gtcgggggca ggagcaggaa ctgaaagtca agaaccgttc tctgcgctcg 26040

ctcacccgca gttgtctgta tcacaagagc gaagaccaac ttcagcgcac tctcgaggac 26100

[0043] gccgaggctc tcttcaacaa gtactgcgcg ctcactctta aagagtagcc cgcgcccgcc 26160

cagtcgcaga aaaaggcggg aattacgtca cctgtgccct tcgccctagc cgcctccacc 26220

catcatcatg agcaaagaga ttcccacgcc ttacatgtgg agctaccagc cccagatggg 26280

cctggccgcc ggtgccgccc aggactactc cacccgcatg aattggctca gcgccgggcc 26340

cgcgatgatc tcacgggtga atgacatccg cgcccaccga aaccagatac tcctagaaca 26400

gtcagcgctc accgccacgc cccgcaatca cctcaatccg cgtaattggc ccgccgccct 26460

ggtgtaccag gaaattcccc agcccacgac cgtactactt ccgcgagacg cccaggccga 26520

agtccagctg actaactcag gtgtccagct ggcgggcggc gccaccctgt gtcgtcaccg 26580

ccccgctcag ggtataaagc ggctggtgat ccggggcaga ggcacacagc tcaacgacga 26640

ggtggtgagc tcttcgctgg gtctgcgacc tgacggagtc ttccaactcg ccggatcggg 26700

gagatcttcc ttcacgcctc gtcaggccgt cctgactttg gagagttcgt cctcgcagcc 26760

ccgctcgggt ggcatcggca ctctccagtt cgtggaggag ttcactccct cggtctactt 26820

caaccccttc tccggctccc ccggccacta cccggacgag ttcatcccga acttcgacgc 26880

catcagcgag tcggtggacg gctacgattg aaactaatca ccccccttatc cagtgaaata 26940

aagatcatat tgatgatgat tttacagaaa taaaaaaataa tcatttgatt tgaaataaag 27000

atacaatcat attgatgatt tgagtttaac aaaaaaaataa agaatcactt acttgaaatc 27060

tgataccagg tctctgtcca tgttttctgc caacaccact tcactcccct cttcccagct 27120

ctggtactgc aggccccggc gggctgcaaa cttcctccac acgctgaagg ggatgtcaaa 27180

ttcctcctgt ccctcaatct tcattttatc ttctatcaga tgtccaaaaa gcgcgtccgg 27240

gtggatgatg acttcgaccc cgtctacccc tacgatgcag acaacgcacc gaccgtgccc 27300

ttcatcaacc ccccttcgt ctcttcagat ggattccaag agaagcccct gggggtgttg 27360

tccctgcgac tggccgaccc cgtcaccacc aagaacgggg aaatcaccct caagctggga 27420

gagggggtgg acctcgattc ctcgggaaaa ctcatctcca acacggccac caaggccgcc 27480

gcccctctca gtttttccaa caacaccatt tcccttaaca tggatcaccc cttttacact 27540

aaagatggaa aattatcctt acaagtttct ccaccattaa atatactgag aacaagcatt 27600

[0044] ctaaacacac tagctttagg ttttggatca ggtttaggac tccgtggctc tgccttggca 27660

gtacagttag tctctccact tacatttgat actgatggaa acataaagct taccttagac 27720

agaggtttgc atgttacaac aggagatgca attgaaagca acataagctg ggctaaaggt 27780

ttaaaatttg aagatggagc catagcaacc aacattggaa atgggttaga gtttggaagc 27840

agtagtacag aaacaggtgt tgatgatgct tacccaatcc aagttaaact tggatctggc 27900

cttagctttg acagtacagg agccataatg gctggtaaca aagaagacga taaactcact 27960

ttgtggacaa cacctgatcc atcaccaaac tgtcaaatac tcgcagaaaa tgatgcaaaa 28020

ctaacacttt gcttgactaa atgtggtagt caaatactgg ccactgtgtc agtcttagtt 28080

gtaggaagtg gaaacctaaa ccccattact ggcaccgtaa gcagtgctca ggtgtttcta 28140

cgttttgatg caaacggtgt tcttttaaca gaacattcta cactaaaaaa atactggggg 28200

tataggcagg gagatagcat agatggcact ccatatacca atgctgtagg attcatgccc 28260

aatttaaaag cttatccaaa gtcacaaagt tctactacta aaaataatat agtagggcaa 28320

gtatacatga atggagatgt ttcaaaacct atgcttctca ctataaccct caatggtact 28380

gatgacagca acagtacata ttcaatgtca ttttcataca cctggactaa tggaagctat 28440

gttggagcaa catttggggc taactcttat accttctcat acatcgccca agaatgaaca 28500

ctgtatccca ccctgcatgc caacccttcc caccccactc tgtggaacaa actctgaaac 28560

acaaaataaa ataaagttca agtgttttat tgattcaaca gttttacagg attcgagcag 28620

ttatttttcc tccaccctcc caggacatgg aatacaccac cctctccccc cgcacagcct 28680

tgaacatctg aatgccattg gtgatggaca tgcttttggt ctccacgttc cacacagttt 28740

cagagcgagc cagtctcggg tcggtcaggg agatgaaacc ctccgggcac tcccgcatct 28800

gcacctcaca gctcaacagc tgaggattgt cctcggtggt cgggatcacg gttatctgga 28860

agaagcagaa gagcggcggt gggaatcata gtccgcgaac gggatcggcc ggtggtgtcg 28920

catcaggccc cgcagcagtc gctgccgccg ccgctccgtc aagctgctgc tcaggggGtc 28980

cgggtccagg gactccctca gcatgatgcc cacggccctc agcatcagtc gtctggtgcg 29040

gcgggcgcag cagcgcatgc ggatctcgct caggtcgctg cagtacgtgc aacacagaac 29100

[0045] caccaggttg ttcaacagtc catagttcaa cacgctccag ccgaaactca tcgcgggaag 29160

gatgctaccc acgtggccgt cgtaccagat cctcaggtaa atcaagtggt gccccctcca 29220

gaacacgctg cccacgtaca tgatctcctt gggcatgtgg cggttcacca cctcccggta 29280

ccacatcacc ctctggttga acatgcagcc ccggatgatc ctgcggaacc acagggccag 29340

caccgccccg cccgccatgc agcgaagaga ccccgggtcc cggcaatggc aatggaggac 29400

ccaccgctcg tacccgtgga tcatctggga gctgaacaag tctatgttgg cacagcacag 29460

gcatatgctc atgcatctct tcagcactct caactcctcg ggggtcaaaa ccatatccca 29520

gggcacgggg aactcttgca ggacagcgaa ccccgcagaa cagggcaatc ctcgcacaga 29580

acttacattg tgcatggaca gggtatcgca atcaggcagc accgggtgat cctccaccag 29640

agaagcgcgg gtctcggtct cctcacagcg tggtaagggg gccggccgat acgggtgatg 29700

gcgggacgcg gctgatcgtg ttcgcgaccg tgtcatgatg cagttgcttt cggacatttt 29760

cgtacttgct gtagcagaac ctggtccggg cgctgcacac cgatcgccgg cggcggtctc 29820

ggcgcttgga acgctcggtg ttgaaattgt aaaacagcca ctctctcaga ccgtgcagca 29880

gatctagggc ctcaggagtg atgaagatcc catcatgcct gatggctctg atcacatcga 29940

ccaccgtgga atgggccaga cccagccaga tgatgcaatt ttgttgggtt tcggtgacgg 30000

cgggggaggg aagaacagga agaaccatga ttaactttta atccaaacgg tctcggagta 30060

cttcaaaatg aagatcgcgg agatggcacc tctcgccccc gctgtgttgg tggaaaataa 30120

cagccaggtc aaaggtgata cggttctcga gatgttccac ggtggcttcc agcaaagcct 30180

ccacgcgcac atccagaaac aagacaatag cgaaagcggg agggttctct aattcctcaa 30240

tcatcatgtt acactcctgc accatcccca gataattttc atttttccag ccttgaatga 30300

ttcgaactag ttcctgaggt aaatccaagc cagccatgat aaagagctcg cgcagagcgc 30360

cctccaccgg cattcttaag cacaccctca taattccaag atattctgct cctggttcac 30420

ctgcagcaga ttgacaagcg gaatatcaaa atctctgccg cgatccctga gctcctccct 30480

cagcaataac tgtaagtact ctttcatatc ctctccgaaa tttttagcca taggaccacc 30540

aggaataaga ttagggcaag ccacagtaca gataaaccga agtcctcccc agtgagcatt 30600

[0046] gccaaatgca agactgctat aagcatgctg gctagacccg gtgatatctt ccagataact 30660

ggacagaaaa tcgcccaggc aatttttaag aaaatcaaca aaagaaaaat cctccaggtg 30720

gacgtttaga gcctcgggaa caacgatgaa gtaaatgcaa gcggtgcgtt ccagcatggt 30780

tagttagctg atctgtagaa aaaacaaaaa tgaacattaa accatgctag cctggcgaac 30840

aggtgggtaa atcgttctct ccagcaccag gcaggccacg gggtctccgg cgcgaccctc 30900

gtaaaaattg tcgctatgat tgaaaaccat cacagagaga cgttcccggt ggccggcgtg 30960

aatgattcga caagatgaat acacccccgg aacattggcg tccgcgagtg aaaaaaagcg 31020

cccgaggaag caataaggca ctacaatgct cagtctcaag tccagcaaag cgatgccatg 31080

cggatgaagc acaaaattct caggtgcgta caaaatgtaa ttactcccct cctgcacagg 31140

cagcaaagcc cccgatccct ccaggtacac atacaaagcc tcagcgtcca tagcttaccg 31200

agcagcagca cacaacaggc gcaagagtca gagaaaggct gagctctaac ctgtccaccc 31260

gctctctgct caatatatag cccagatcta cactgacgta aaggccaaag tctaaaaata 31320

cccgccaaat aatcacacac gcccagcaca cgcccagaaa ccggtgacac actcaaaaaa 31380

atacgcgcac ttcctcaaac gcccaaaact gccgtcattt ccgggttccc acgctacgtc 31440

atcaaaacac gactttcaaa ttccgtcgac cgttaaaaac gtcacccgcc ccgcccctaa 31500

cggtcgcccg tctctcagcc aatcagcgcc ccgcatcccc aaattcaaac acctcatttg 31560

catattaacg cgcacaaaaa gtttgagg 31588

<210> 3
<211> 11447
<212> DNA
<213> 委内瑞拉马脑炎病毒(Venezuelan equine encephalitis virus)

<400> 3
atgggcggcg catgagagaa gcccagacca attacctacc caaaatggag aaagttcacg 60

ttgacatcga ggaagacagc ccattcctca gagctttgca gcggagcttc ccgcagtttg 120

aggtagaagc caagcaggtc actgataatg accatgctaa tgccagagcg ttttcgcatc 180

tggcttcaaa actgatcgaa acggaggtgg acccatccga cacgatcctt gacattggaa 240

[0047] gtgcgcccgc ccgcagaatg tattctaagc acaagtatca ttgtatctgt ccgatgagat 300

gtgcggaaga tccggacaga ttgtataagt atgcaactaa gctgaagaaa aactgtaagg 360

aaataactga taaggaattg gacaagaaaa tgaaggagct cgccgccgtc atgagcgacc 420

ctgacctgga aactgagact atgtgcctcc acgacgacga gtcgtgtcgc tacgaagggc 480

aagtcgctgt ttaccaggat gtatacgcgg ttgacggacc gacaagtctc tatcaccaag 540

ccaataaggg agttagagtc gcctactgga taggctttga caccacccct tttatgttta 600

agaacttggc tggagcatat ccatcatact ctaccaactg ggccgacgaa accgtgttaa 660

cggctcgtaa cataggccta tgcagctctg acgttatgga gcggtcacgt agagggatgt 720

ccattcttag aaagaagtat ttgaaaccat ccaacaatgt tctattctct gttggctcga 780

ccatctacca cgagaagagg gacttactga ggagctggca cctgccgtct gtatttcact 840

tacgtggcaa gcaaaattac acatgtcggt gtgagactat agttagttgc gacgggtacg 900

tcgttaaaag aatagctatc agtccaggcc tgtatgggaa gccttcaggc tatgctgcta 960

cgatgcaccg cgagggattc ttgtgctgca aagtgacaga cacattgaac ggggagaggg 1020

tctcttttcc cgtgtgcacg tatgtgccag ctacattgtg tgaccaaatg actggcatac 1080

tggcaacaga tgtcagtgcg gacgacgcgc aaaaactgct ggttgggctc aaccagcgta 1140

tagtcgtcaa cggtcgcacc cagagaaaca ccaataccat gaaaaattac cttttgcccg 1200

tagtggccca ggcatttgct aggtgggcaa aggaatataa ggaagatcaa gaagatgaaa 1260

ggccactagg actacgagat agacagttag tcatggggtg ttgttgggct tttagaaggc 1320

acaagataac atctatttat aagcgcccgg atacccaaac catcatcaaa gtgaacagcg 1380

atttccactc attcgtgctg cccaggatag gcagtaacac attggagatc gggctgagaa 1440

caagaatcag gaaaatgtta gaggagcaca aggagccgtc acctctcatt accgccgagg 1500

acgtacaaga agctaagtgc gcagccgatg aggctaagga ggtgcgtgaa gccgaggagt 1560

tgcgcgcagc tctaccacct ttggcagctg atgttgagga gcccactctg gaagccgatg 1620

tcgacttgat gttacaagag gctggggccg gctcagtgga gacacctcgt ggcttgataa 1680

aggttaccag ctacgctggc gaggacaaga tcggctctta cgctgtgctt tctccgcagg 1740

[0048] ctgtactcaa gagtgaaaaa ttatcttgca tccaccctct cgctgaacaa gtcatagtga 1800

taacacactc tggccgaaaa gggcgttatg ccgtggaacc ataccatggt aaagtagtgg 1860

tgccagaggg acatgcaata cccgtccagg actttcaagc tctgagtgaa agtgccacca 1920

ttgtgtacaa cgaacgtgag ttcgtaaaca ggtacctgca ccatattgcc acacatggag 1980

gagcgctgaa cactgatgaa gaatattaca aaactgtcaa gcccagcgag cacgacggcg 2040

aatacctgta cgacatcgac aggaaacagt gcgtcaagaa agaactagtc actgggctag 2100

ggctcacagg cgagctggtg gatcctccct tccatgaatt cgcctacgag agtctgagaa 2160

cacgaccagc cgctccttac caagtaccaa ccataggggt gtatggcgtg ccaggatcag 2220

gcaagtctgg catcattaaa agcgcagtca ccaaaaaaga tctagtggtg agcgccaaga 2280

aagaaaactg tgcagaaatt ataagggacg tcaagaaaat gaaagggctg gacgtcaatg 2340

ccagaactgt ggactcagtg ctcttgaatg gatgcaaaca ccccgtagag accctgtata 2400

ttgacgaagc ttttgcttgt catgcaggta ctctcagagc gctcatagcc attataagac 2460

ctaaaaaggc agtgctctgc ggggatccca aacagtgcgg ttttttaac atgatgtgcc 2520

tgaaagtgca ttttaaccac gagatttgca cacaagtctt ccacaaaagc atctctcgcc 2580

gttgcactaa atctgtgact tcggtcgtct caaccttgtt ttacgacaaa aaaatgagaa 2640

cgacgaatcc gaaagagact aagattgtga ttgacactac cggcagtacc aaacctaagc 2700

aggacgatct cattctcact tgtttcagag ggtgggtgaa gcagttgcaa atagattaca 2760

aaggcaacga aataatgacg gcagctgcct ctcaagggct gacccgtaaa ggtgtgtatg 2820

ccgttcggta caaggtgaat gaaaatcctc tgtacgcacc cacctcagaa catgtgaacg 2880

tcctactgac ccgcacggag gaccgcatcg tgtggaaaac actagccggc gacccatgga 2940

taaaaacact gactgccaag taccctggga atttcactgc cacgatagag gagtggcaag 3000

cagagcatga tgccatcatg aggcacatct tggagagacc ggaccctacc gacgtcttcc 3060

agaataaggc aaacgtgtgt tgggccaagg ctttagtgcc ggtgctgaag accgctggca 3120

tagacatgac cactgaacaa tggaacactg tggattattt tgaaacggac aaagctcact 3180

cagcagagat agtattgaac caactatgcg tgaggttctt tggactcgat ctggactccg 3240

[0049] gtctattttc tgcacccact gttccgttat ccattaggaa taatcactgg gataactccc 3300

cgtcgcctaa catgtacggg ctgaataaag aagtggtccg tcagctctct cgcaggtacc 3360

cacaactgcc tcgggcagtt gccactggaa gagtctatga catgaacact ggtacactgc 3420

gcaattatga tccgcgcata aacctagtac ctgtaaacag aagactgcct catgctttag 3480

tcctccacca taatgaacac ccacagagtg acttttcttc attcgtcagc aaattgaagg 3540

gcagaactgt cctggtggtc ggggaaaagt tgtccgtccc aggcaaaatg gttgactggt 3600

tgtcagaccg gcctgaggct accttcagag ctcggctgga tttaggcatc ccaggtgatg 3660

tgcccaaata tgacataata tttgttaatg tgaggacccc atataaatac catcactatc 3720

agcagtgtga agaccatgcc attaagctta gcatgttgac caagaaagct tgtctgcatc 3780

tgaatcccgg cggaacctgt gtcagcatag gttatggtta cgctgacagg gccagcgaaa 3840

gcatcattgg tgctatagcg cggcagttca agttttcccg ggtatgcaaa ccgaaatcct 3900

cacttgaaga gacggaagtt ctgtttgtat tcattgggta cgatcgcaag gcccgtacgc 3960

acaatcctta caagctttca tcaaccttga ccaacattta tacaggttcc agactccacg 4020

aagccggatg tgcaccctca tatcatgtgg tgcgagggga tattgccacg gccaccgaag 4080

gagtgattat aaatgctgct aacagcaaag gacaacctgg cggaggggtg tgcggagcgc 4140

tgtataagaa attcccggaa agcttcgatt tacagccgat cgaagtagga aaagcgcgac 4200

tggtcaaagg tgcagctaaa catatcattc atgccgtagg accaaacttc aacaaagttt 4260

cggaggttga aggtgacaaa cagttggcag aggcttatga gtccatcgct aagattgtca 4320

acgataacaa ttacaagtca gtagcgattc cactgttgtc caccggcatc ttttccggga 4380

acaaagatcg actaacccaa tcattgaacc atttgctgac agctttagac accactgatg 4440

cagatgtagc catatactgc agggacaaga aatgggaaat gactctcaag gaagcagtgg 4500

ctaggagaga agcagtggag gagatatgca tatccgacga ctcttcagtg acagaacctg 4560

atgcagagct ggtgagggtg catccgaaga gttctttggc tggaaggaag ggctacagca 4620

caagcgatgg caaaactttc tcatatttgg aagggaccaa gtttcaccag gcggccaagg 4680

atatagcaga aattaatgcc atgtggcccg ttgcaacgga ggccaatgag caggtatgca 4740

[0050] tgtatatcct cggagaaagc atgagcagta ttaggtcgaa atgccccgtc gaagagtcgg 4800

aagcctccac accacctagc acgctgcctt gcttgtgcat ccatgccatg actccagaaa 4860

gagtacagcg cctaaaagcc tcacgtccag aacaaattac tgtgtgctca tcctttccat 4920

tgccgaagta tagaatcact ggtgtgcaga agatccaatg ctcccagcct atattgttct 4980

caccgaaagt gcctgcgtat attcatccaa ggaagtatct cgtggaaaca ccaccggtag 5040

acgagactcc ggagccatcg gcagagaacc aatccacaga ggggacacct gaacaaccac 5100

cacttataac cgaggatgag accaggacta gaacgcctga gccgatcatc atcgaagagg 5160

aagaagagga tagcataagt ttgctgtcag atggcccgac ccaccaggtg ctgcaagtcg 5220

aggcagacat tcacgggccg ccctctgtat ctagctcatc ctggtccatt cctcatgcat 5280

ccgactttga tgtggacagt ttatccatac ttgacaccct ggagggagct agcgtgacca 5340

gcggggcaac gtcagccgag actaactctt acttcgcaaa gagtatggag tttctggcgc 5400

gaccggtgcc tgcgcctcga acagtattca ggaaccctcc acatcccgct ccgcgcacaa 5460

gaacaccgtc acttgcaccc agcagggcct gctcgagaac cagcctagtt tccaccccgc 5520

caggcgtgaa tagggtgatc actagagagg agctcgaggc gcttaccccg tcacgcactc 5580

ctagcaggtc ggtctcgaga accagcctgg tctccaaccc gccaggcgta aatagggtga 5640

ttacaagaga ggagtttgag gcgttcgtag cacaacaaca atgacggttt gatgcgggtg 5700

catacatctt ttcctccgac accggtcaag ggcatttaca acaaaaatca gtaaggcaaa 5760

cggtgctatc cgaagtggtg ttggagagga ccgaattgga gatttcgtat gccccgcgcc 5820

tcgaccaaga aaaagaagaa ttactacgca agaaattaca gttaaatccc acacctgcta 5880

acagaagcag ataccagtcc aggaaggtgg agaacatgaa agccataaca gctagacgta 5940

ttctgcaagg cctagggcat tatttgaagg cagaaggaaa agtggagtgc taccgaaccc 6000

tgcatcctgt tcctttgtat tcatctagtg tgaaccgtgc cttttcaagc cccaaggtcg 6060

cagtggaagc ctgtaacgcc atgttgaaag agaactttcc gactgtggct tcttactgta 6120

ttattccaga gtacgatgcc tatttggaca tggttgacgg agcttcatgc tgcttagaca 6180

ctgccagttt ttgccctgca aagctgcgca gctttccaaa gaaacactcc tatttggaac 6240

[0051] ccacaatacg atcggcagtg ccttcagcga tccagaacac gctccagaac gtcctggcag 6300

ctgccacaaa aagaaattgc aatgtcacgc aaatgagaga attgcccgta ttggattcgg 6360

cggcctttaa tgtggaatgc ttcaagaaat atgcgtgtaa taatgaatat tgggaaacgt 6420

ttaaagaaaa ccccatcagg cttactgaag aaaacgtggt aaattacatt accaaattaa 6480

aaggaccaaa agctgctgct ctttttgcga agacacataa tttgaatatg ttgcaggaca 6540

taccaatgga caggtttgta atggacttaa agagagacgt gaaagtgact ccaggaacaa 6600

aacatactga agaacggccc aaggtacagg tgatccaggc tgccgatccg ctagcaacag 6660

cgtatctgtg cggaatccac cgagagctgg ttaggagatt aaatgcggtc ctgcttccga 6720

acattcatac actgtttgat atgtcggctg aagactttga cgctattata gccgagcact 6780

tccagcctgg ggattgtgtt ctggaaactg acatcgcgtc gtttgataaa agtgaggacg 6840

acgccatggc tctgaccgcg ttaatgattc tggaagactt aggtgtggac gcagagctgt 6900

tgacgctgat tgaggcggct ttcggcgaaa tttcatcaat acatttgccc actaaaacta 6960

aatttaaatt cggagccatg atgaaatctg gaatgttcct cacactgttt gtgaacacag 7020

tcattaacat tgtaatcgca agcagagtgt tgagagaacg gctaaccgga tcaccatgtg 7080

cagcattcat tggagatgac aatatcgtga aaggagtcaa atcggacaaa ttaatggcag 7140

acaggtgcgc cacctggttg aatatggaag tcaagattat agatgctgtg gtgggcgaga 7200

aagcgcctta tttctgtgga gggtttattt tgtgtgactc cgtgaccggc acagcgtgcc 7260

gtgtggcaga ccccctaaaa aggctgttta agcttggcaa acctctggca gcagacgatg 7320

aacatgatga tgacaggaga agggcattgc atgaagagtc aacacgctgg aaccgagtgg 7380

gtattctttc agagctgtgc aaggcagtag aatcaaggta tgaaaccgta ggaacttcca 7440

tcatagttat ggccatgact actctagcta gcagtgttaa atcattcagc tacctgagag 7500

gggcccctat aactctctac ggctaacctg aatggactac gacatagtct agtccgccaa 7560

gatgttcccg ttccagccaa tgtatccgat gcagccaatg ccctatcgca acccgttcgc 7620

ggccccgcgc aggccctggt tccccagaac cgacccttttt ctggcgatgc aggtgcagga 7680

attaacccgc tcgatggcta acctgacgtt caagcaacgc cgggacgcgc cacctgaggg 7740

[0052] gccatccgct aagaaaccga agaaggaggc ctcgcaaaaa cagaaagggg gaggccaagg 7800

gaagaagaag aagaaccaag ggaagaagaa ggctaagaca gggccgccta atccgaaggc 7860

acagaatgga aacaagaaga agaccaacaa gaaaccaggc aagagacagc gcatggtcat 7920

gaaattggaa tctgacaaga cgttcccaat catgttggaa gggaagataa acggctacgc 7980

ttgtgtggtc ggagggaagt tattcaggcc gatgcatgtg gaaggcaaga tcgacaacga 8040

cgttctggcc gcgcttaaga cgaagaaagc atccaaatac gatcttgagt atgcagatgt 8100

gccacagaac atgcgggccg atacattcaa atacacccat gagaaacccc aaggctatta 8160

cagctggcat catggagcag tccaatatga aaatgggcgt ttcacggtgc cgaaaggagt 8220

tggggccaag ggagacagcg gacgacccat tctggataac cagggacggg tggtcgctat 8280

tgtgctggga ggtgtgaatg aaggatctag gacagccctt tcagtcgtca tgtggaacga 8340

gaagggagtt accgtgaagt atactccgga gaactgcgag caatggtcac tagtgaccac 8400

catgtgtctg ctcgccaatg tgacgttccc atgtgctcaa ccaccaattt gctacgacag 8460

aaaaccagca gagactttgg ccatgctcag cgttaacgtt gacaacccgg gctacgatga 8520

gctgctggaa gcagctgtta agtgccccgg aaggaaaagg agatccaccg aggagctgtt 8580

taaggagtat aagctaacgc gcccttacat ggccagatgc atcagatgtg cagttgggag 8640

ctgccatagt ccaatagcaa tcgaggcagt aaagagcgac gggcacgacg gttatgttag 8700

acttcagact tcctcgcagt atggcctgga ttcctccggc aacttaaagg gcaggaccat 8760

gcggtatgac atgcacggga ccattaaaga gataccacta catcaagtgt cactccatac 8820

atctcgcccg tgtcacattg tggatgggca cggttatttc ctgcttgcca ggtgcccggc 8880

aggggactcc atcaccatgg aatttaagaa agattccgtc acacactcct gctcggtgcc 8940

gtatgaagtg aaatttaatc ctgtaggcag agaactctat actcatcccc cagaacacgg 9000

agtagagcaa gcgtgccaag tctacgcaca tgatgcacag aacagaggag cttatgtcga 9060

gatgcacctc ccgggctcag aagtggacag cagtttggtt tccttgagcg gcagttcagt 9120

caccgtgaca cctcctgttg ggactagcgc cctggtggaa tgcgagtgtg gcggcacaaa 9180

gatctccgag accatcaaca agacaaaaca gttcagccag tgcacaaaga aggagcagtg 9240

[0053] cagagcatat cggctgcaga acgataagtg ggtgtataat tctgacaaac tgcccaaagc 9300

agcgggagcc accttaaaag gaaaactgca tgtcccattc ttgctggcag acggcaaatg 9360

caccgtgcct ctagcaccag aacctatgat aacctttggt ttcagatcag tgtcactgaa 9420

actgcaccct aagaatccca catatctaac cacccgccaa cttgctgatg agcctcacta 9480

cacgcacgag ctcatatctg aaccagctgt taggaatttt accgtcaccg aaaaagggtg 9540

ggagtttgta tggggaaacc acccgccgaa aaggttttgg gcacaggaaa cagcacccgg 9600

aaatccacat gggctaccgc acgaggtgat aactcattat taccacagat accctatgtc 9660

caccatcctg ggtttgtcaa tttgtgccgc cattgcaacc gtttccgttg cagcgtctac 9720

ctggctgttt tgcagatcta gagttgcgtg cctaactcct taccggctaa cacctaacgc 9780

taggatacca ttttgtctgg ctgtgctttg ctgcgcccgc actgcccggg ccgagaccac 9840

ctgggagtcc ttggatcacc tatggaacaa taaccaacag atgttctgga ttcaattgct 9900

gatccctctg gccgccttga tcgtagtgac tcgcctgctc aggtgcgtgt gctgtgtcgt 9960

gccttttta gtcatggccg gcgccgcagg cgccggcgcc tacgagcacg cgaccacgat 10020

gccgagccaa gcgggaatct cgtataacac tatagtcaac agagcaggct acgcaccact 10080

ccctatcagc ataacaccaa caaagatcaa gctgatacct acagtgaact tggagtacgt 10140

cacctgccac tacaaaacag gaatggattc accagccatc aaatgctgcg gatctcagga 10200

atgcactcca acttacaggc ctgatgaaca gtgcaaagtc ttcacagggg tttacccgtt 10260

catgtggggt ggtgcatatt gcttttgcga cactgagaac acccaagtca gcaaggccta 10320

cgtaatgaaa tctgacgact gccttgcgga tcatgctgaa gcatataaag cgcacacagc 10380

ctcagtgcag gcgttcctca acatcacagt gggagaacac tctattgtga ctaccgtgta 10440

tgtgaatgga gaaactcctg tgaatttcaa tggggtcaaa ttaactgcag gtccgctttc 10500

cacagcttgg acacccttg atcgcaaaat cgtgcagtat gccggggaga tctataatta 10560

tgattttcct gagtatgggg caggacaacc aggagcattt ggagatatac aatccagaac 10620

agtctcaagc tcagatctgt atgccaatac caacctagtg ctgcagagac ccaaagcagg 10680

agcgatccac gtgccataca ctcaggcacc ttcgggtttt gagcaatgga agaaagataa 10740

[0054] agctccatca ttgaaattta ccgccccttt cggatgcgaa atatatacaa accccattcg 10800

cgccgaaaac tgtgctgtag ggtcaattcc attagccttt gacattcccg acgccttgtt 10860

caccagggtg tcagaaacac cgacactttc agcggccgaa tgcactctta acgagtgcgt 10920

gtattcttcc gactttggtg ggatcgccac ggtcaagtac tcggccagca agtcaggcaa 10980

gtgcgcagtc catgtgccat cagggactgc taccctaaaa gaagcagcag tcgagctaac 11040

cgagcaaggg tcggcgacta tccatttctc gaccgcaaat atccacccgg agttcaggct 11100

ccaaatatgc acatcatatg ttacgtgcaa aggtgattgt caccccccga aagaccatat 11160

tgtgacacac cctcagtatc acgcccaaac atttacagcc gcggtgtcaa aaaccgcgtg 11220

gacgtggtta acatccctgc tgggaggatc agccgtaatt attataattg gcttggtgct 11280

ggctactatt gtggccatgt acgtgctgac caaccagaaa cataattgaa tacagcagca 11340

attggcaagc tgcttacata gaactcgcgg cgattggcat gccgccttaa aatttttatt 11400

ttatttttc ttttcttttc cgaatcggat tttgttttta atatttc 11447

<210> 4

<211> 9577
<212> DNA
<213> 人工序列(Artificial Sequence)

<220>
<223> 人工序列的描述：合成多核苷酸

<400> 4

[0055]

atgggcggcg catgagagaa gcccagacca attacctacc caaaatggag aaagttcacg 60

ttgacatcga ggaagacagc ccattcctca gagctttgca gcggagcttc ccgcagtttg 120

aggtagaagc caagcaggtc actgataatg accatgctaa tgccagagcg ttttcgcatc 180

tggcttcaaa actgatcgaa acggaggtgg acccatccga cacgatcctt gacattggaa 240

gtgcgcccgc ccgcagaatg tattctaagc acaagtatca ttgtatctgt ccgatgagat 300

gtgcggaaga tccggacaga ttgtataagt atgcaactaa gctgaagaaa aactgtaagg 360

aaataactga taaggaattg gacaagaaaa tgaaggagct cgccgccgtc atgagcgacc 420

ctgacctgga aactgagact atgtgcctcc acgacgacga gtcgtgtcgc tacgaagggc 480

aagtcgctgt ttaccaggat gtatacgcgg ttgacggacc gacaagtctc tatcaccaag 540

ccaataaggg agttagagtc gcctactgga taggctttga caccacccct tttatgttta 600

agaacttggc tggagcatat ccatcatact ctaccaactg ggccgacgaa accgtgttaa 660

cggctcgtaa cataggccta tgcagctctg acgttatgga gcggtcacgt agagggatgt 720

ccattcttag aaagaagtat ttgaaaccat ccaacaatgt tctattctct gttggctcga 780

ccatctacca cgagaagagg gacttactga ggagctggca cctgccgtct gtatttcact 840

tacgtggcaa gcaaaattac acatgtcggt gtgagactat agttagttgc gacgggtacg 900

tcgttaaaag aatagctatc agtccaggcc tgtatgggaa gccttcaggc tatgctgcta 960

cgatgcaccg cgagggattc ttgtgctgca aagtgacaga cacattgaac ggggagaggg 1020

tctcttttcc cgtgtgcacg tatgtgccag ctacattgtg tgaccaaatg actggcatac 1080

tggcaacaga tgtcagtgcg gacgacgcgc aaaaactgct ggttgggctc aaccagcgta 1140

tagtcgtcaa cggtcgcacc cagagaaaca ccaataccat gaaaaattac cttttgcccg 1200

tagtggccca ggcatttgct aggtgggcaa aggaatataa ggaagatcaa gaagatgaaa 1260

ggccactagg actacgagat agacagttag tcatggggtg ttgttgggct tttagaaggc 1320

acaagataac atctatttat aagcgcccgg atacccaaac catcatcaaa gtgaacagcg 1380

atttccactc attcgtgctg cccaggatag gcagtaacac attggagatc gggctgagaa 1440

caagaatcag gaaaatgtta gaggagcaca aggagccgtc acctctcatt accgccgagg 1500

acgtacaaga agctaagtgc gcagccgatg aggctaagga ggtgcgtgaa gccgaggagt 1560

tgcgcgcagc tctaccacct ttggcagctg atgttgagga gcccactctg gaagccgatg 1620

tcgacttgat gttacaagag gctggggccg gctcagtgga gacacctcgt ggcttgataa 1680

aggttaccag ctacgctggc gaggacaaga tcggctctta cgctgtgctt tctccgcagg 1740

ctgtactcaa gagtgaaaaa ttatcttgca tccaccctct cgctgaacaa gtcatagtga 1800

taacacactc tggccgaaaa gggcgttatg ccgtggaacc ataccatggt aaagtagtgg 1860

tgccagaggg acatgcaata cccgtccagg actttcaagc tctgagtgaa agtgccacca 1920

ttgtgtacaa cgaacgtgag ttcgtaaaca ggtacctgca ccatattgcc acacatggag 1980

[0056] gagcgctgaa cactgatgaa gaatattaca aaactgtcaa gcccagcgag cacgacggcg 2040

aatacctgta cgacatcgac aggaaacagt gcgtcaagaa agaactagtc actgggctag 2100

ggctcacagg cgagctggtg gatcctccct tccatgaatt cgcctacgag agtctgagaa 2160

cacgaccagc cgctccttac caagtaccaa ccataggggt gtatggcgtg ccaggatcag 2220

gcaagtctgg catcattaaa agcgcagtca ccaaaaaaga tctagtggtg agcgccaaga 2280

aagaaaactg tgcagaaatt ataagggacg tcaagaaaat gaaagggctg gacgtcaatg 2340

ccagaactgt ggactcagtg ctcttgaatg gatgcaaaca ccccgtagag accctgtata 2400

ttgacgaagc ttttgcttgt catgcaggta ctctcagagc gctcatagcc attataagac 2460

ctaaaaaggc agtgctctgc ggggatccca aacagtgcgg ttttttaac atgatgtgcc 2520

tgaaagtgca ttttaaccac gagatttgca cacaagtctt ccacaaaagc atctctcgcc 2580

gttgcactaa atctgtgact tcggtcgtct caaccttgtt ttacgacaaa aaaatgagaa 2640

cgacgaatcc gaaagagact aagattgtga ttgacactac cggcagtacc aaacctaagc 2700

aggacgatct cattctcact tgtttcagag ggtgggtgaa gcagttgcaa atagattaca 2760

aaggcaacga aataatgacg gcagctgcct ctcaagggct gacccgtaaa ggtgtgtatg 2820

ccgttcggta caaggtgaat gaaaatcctc tgtacgcacc cacctcagaa catgtgaacg 2880

tcctactgac ccgcacggag gaccgcatcg tgtggaaaac actagccggc gacccatgga 2940

taaaaacact gactgccaag taccctggga atttcactgc cacgatagag gagtggcaag 3000

cagagcatga tgccatcatg aggcacatct tggagagacc ggaccctacc gacgtcttcc 3060

agaataaggc aaacgtgtgt tgggccaagg ctttagtgcc ggtgctgaag accgctggca 3120

tagacatgac cactgaacaa tggaacactg tggattattt tgaaacggac aaagctcact 3180

cagcagagat agtattgaac caactatgcg tgaggttctt tggactcgat ctggactccg 3240

gtctattttc tgcacccact gttccgttat ccattaggaa taatcactgg gataactccc 3300

cgtcgcctaa catgtacggg ctgaataaag aagtggtccg tcagctctct cgcaggtacc 3360

cacaactgcc tcgggcagtt gccactggaa gagtctatga catgaacact ggtacactgc 3420

gcaattatga tccgcgcata aacctagtac ctgtaaacag aagactgcct catgctttag 3480

[0057] tcctccacca taatgaacac ccacagagtg acttttcttc attcgtcagc aaattgaagg 3540

gcagaactgt cctggtggtc ggggaaaagt tgtccgtccc aggcaaaatg gttgactggt 3600

tgtcagaccg gcctgaggct accttcagag ctcggctgga tttaggcatc ccaggtgatg 3660

tgcccaaata tgacataata tttgttaatg tgaggacccc atataaatac catcactatc 3720

agcagtgtga agaccatgcc attaagctta gcatgttgac caagaaagct tgtctgcatc 3780

tgaatcccgg cggaacctgt gtcagcatag gttatggtta cgctgacagg gccagcgaaa 3840

gcatcattgg tgctatagcg cggcagttca agttttcccg ggtatgcaaa ccgaaatcct 3900

cacttgaaga gacggaagtt ctgtttgtat tcattgggta cgatcgcaag gcccgtacgc 3960

acaatcctta caagctttca tcaaccttga ccaacattta tacaggttcc agactccacg 4020

aagccggatg tgcaccctca tatcatgtgg tgcgagggga tattgccacg gccaccgaag 4080

gagtgattat aaatgctgct aacagcaaag gacaacctgg cggaggggtg tgcggagcgc 4140

tgtataagaa attcccggaa agcttcgatt tacagccgat cgaagtagga aaagcgcgac 4200

tggtcaaagg tgcagctaaa catatcattc atgccgtagg accaaacttc aacaaagttt 4260

cggaggttga aggtgacaaa cagttggcag aggcttatga gtccatcgct aagattgtca 4320

acgataacaa ttacaagtca gtagcgattc cactgttgtc caccggcatc ttttccggga 4380

acaaagatcg actaacccaa tcattgaacc atttgctgac agctttagac accactgatg 4440

cagatgtagc catatactgc agggacaaga aatgggaaat gactctcaag gaagcagtgg 4500

ctaggagaga agcagtggag gagatatgca tatccgacga ctcttcagtg acagaacctg 4560

atgcagagct ggtgagggtg catccgaaga gttctttggc tggaaggaag ggctacagca 4620

caagcgatgg caaaactttc tcatatttgg aagggaccaa gtttcaccag gcggccaagg 4680

atatagcaga aattaatgcc atgtggcccg ttgcaacgga ggccaatgag caggtatgca 4740

tgtatatcct cggagaaagc atgagcagta ttaggtcgaa atgccccgtc gaagagtcgg 4800

aagcctccac accacctagc acgctgcctt gcttgtgcat ccatgccatg actccagaaa 4860

gagtacagcg cctaaaagcc tcacgtccag aacaaattac tgtgtgctca tcctttccat 4920

tgccgaagta tagaatcact ggtgtgcaga agatccaatg ctcccagcct atattgttct 4980

[0058] caccgaaagt gcctgcgtat attcatccaa ggaagtatct cgtggaaaca ccaccggtag 5040

acgagactcc ggagccatcg gcagagaacc aatccacaga ggggacacct gaacaaccac 5100

cacttataac cgaggatgag accaggacta gaacgcctga gccgatcatc atcgaagagg 5160

aagaagagga tagcataagt ttgctgtcag atggcccgac ccaccaggtg ctgcaagtcg 5220

aggcagacat tcacgggccg ccctctgtat ctagctcatc ctggtccatt cctcatgcat 5280

ccgactttga tgtggacagt ttatccatac ttgacaccct ggagggagct agcgtgacca 5340

gcggggcaac gtcagccgag actaactctt acttcgcaaa gagtatggag tttctggcgc 5400

gaccggtgcc tgcgcctcga acagtattca ggaaccctcc acatcccgct ccgcgcacaa 5460

gaacaccgtc acttgcaccc agcagggcct gctcgagaac cagcctagtt tccaccccgc 5520

caggcgtgaa tagggtgatc actagagagg agctcgaggc gcttaccccg tcacgcactc 5580

ctagcaggtc ggtctcgaga accagcctgg tctccaaccc gccaggcgta aatagggtga 5640

ttacaagaga ggagtttgag gcgttcgtag cacaacaaca atgacggttt gatgcgggtg 5700

catacatctt ttcctccgac accggtcaag ggcatttaca acaaaaatca gtaaggcaaa 5760

cggtgctatc cgaagtggtg ttggagagga ccgaattgga gatttcgtat gccccgcgcc 5820

tcgaccaaga aaaagaagaa ttactacgca agaaattaca gttaaatccc acacctgcta 5880

acagaagcag ataccagtcc aggaaggtgg agaacatgaa agccataaca gctagacgta 5940

ttctgcaagg cctagggcat tatttgaagg cagaaggaaa agtggagtgc taccgaaccc 6000

tgcatcctgt tcctttgtat tcatctagtg tgaaccgtgc cttttcaagc cccaaggtcg 6060

cagtggaagc ctgtaacgcc atgttgaaag agaactttcc gactgtggct tcttactgta 6120

ttattccaga gtacgatgcc tatttggaca tggttgacgg agcttcatgc tgcttagaca 6180

ctgccagttt ttgccctgca aagctgcgca gctttccaaa gaaacactcc tatttggaac 6240

ccacaatacg atcggcagtg ccttcagcga tccagaacac gctccagaac gtcctggcag 6300

ctgccacaaa aagaaattgc aatgtcacgc aaatgagaga attgcccgta ttggattcgg 6360

cggcctttaa tgtggaatgc ttcaagaaat atgcgtgtaa taatgaatat tgggaaacgt 6420

ttaaagaaaa ccccatcagg cttactgaag aaaacgtggt aaattacatt accaaattaa 6480

[0059] aaggaccaaa agctgctgct ctttttgcga agacacataa tttgaatatg ttgcaggaca 6540

taccaatgga caggtttgta atggacttaa agagagacgt gaaagtgact ccaggaacaa 6600

aacatactga agaacggccc aaggtacagg tgatccaggc tgccgatccg ctagcaacag 6660

cgtatctgtg cggaatccac cgagagctgg ttaggagatt aaatgcggtc ctgcttccga 6720

acattcatac actgtttgat atgtcggctg aagactttga cgctattata gccgagcact 6780

tccagcctgg ggattgtgtt ctggaaactg acatcgcgtc gtttgataaa agtgaggacg 6840

acgccatggc tctgaccgcg ttaatgattc tggaagactt aggtgtggac gcagagctgt 6900

tgacgctgat tgaggcggct ttcggcgaaa tttcatcaat acatttgccc actaaaacta 6960

aatttaaatt cggagccatg atgaaatctg gaatgttcct cacactgttt gtgaacacag 7020

tcattaacat tgtaatcgca agcagagtgt tgagagaacg gctaaccgga tcaccatgtg 7080

cagcattcat tggagatgac aatatcgtga aaggagtcaa atcggacaaa ttaatggcag 7140

acaggtgcgc cacctggttg aatatggaag tcaagattat agatgctgtg gtgggcgaga 7200

aagcgcctta tttctgtgga gggtttattt tgtgtgactc cgtgaccggc acagcgtgcc 7260

gtgtggcaga ccccctaaaa aggctgttta agcttggcaa acctctggca gcagacgatg 7320

aacatgatga tgacaggaga agggcattgc atgaagagtc aacacgctgg aaccgagtgg 7380

gtattctttc agagctgtgc aaggcagtag aatcaaggta tgaaaccgta ggaacttcca 7440

tcatagttat ggccatgact actctagcta gcagtgttaa atcattcagc tacctgagag 7500

gggcccctat aactctctac ggctaacctg aatggactac gactctagaa tagtctttaa 7560

ttaagccacc atggcaggca tgtttcaggc gctgagcgaa ggctgcaccc cgtatgatat 7620

taaccagatg ctgaacgtgc tgggcgatca tcaggtctca ggccttgagc agcttgagag 7680

tataatcaac tttgaaaaac tgactgaatg gaccagttct aatgttatgc ctatcctgtc 7740

tcctctgaca aagggcatcc tgggcttcgt gtttaccctg accgtgcctt ctgagagagg 7800

acttagctgc attagcgaag cggatgcgac caccccggaa agcgcgaacc tgggcgaaga 7860

aattctgagc cagctgtatc tttggccaag ggtgacctac cattccccta gttatgctta 7920

ccaccaattt gaaagacgag ccaaatataa aagacacttc cccggctttg gccagagcct 7980

[0060] gctgtttggc taccctgtgt acgtgttcgg cgattgcgtg cagggcgatt gggatgcgat 8040

tcgctttcgc tattgcgcgc cgccgggcta tgcgctgctg cgctgcaacg ataccaacta 8100

tagcgctctg ctggctgtgg gggccctaga aggacccagg aatcaggact ggcttggtgt 8160

cccaagacaa cttgtaactc ggatgcaggc tattcagaat gccggcctgt gtaccctggt 8220

ggccatgctg gaagagacaa tcttctggct gcaagcgttt ctgatggcgc tgaccgatag 8280

cggcccgaaa accaacatta ttgtggatag ccagtatgtg atgggcatta gcaaaccgag 8340

ctttcaggaa tttgtggatt gggaaaacgt gagcccggaa ctgaacagca ccgatcagcc 8400

gttttggcaa gccggaatcc tggccagaaa tctggtgcct atggtggcca cagtgcaggg 8460

ccagaacctg aagtaccagg gtcagtcact agtcatctct gcttctatca ttgtcttcaa 8520

cctgctggaa ctggaaggtg attatcgaga tgatggcaac gtgtgggtgc ataccccgct 8580

gagcccgcgc accctgaacg cgtgggtgaa agcggtggaa gaaaaaaaag gtattccagt 8640

tcacctagag ctggccagta tgaccaacat ggagctcatg agcagtattg tgcatcagca 8700

ggtcagaaca tacggccccg tgttcatgtg tctcggcgga ctgcttacaa tggtggctgg 8760

tgctgtgtgg ctgacagtgc gagtgctcga gctgttccgg gccgcgcagc tggccaacga 8820

cgtggtcctc cagatcatgg agctttgtgg tgcagcgttt cgccaggtgt gccataccac 8880

cgtgccgtgg ccgaacgcga gcctgacccc gaaatggaac aacgaaacca cccagcccca 8940

gatcgccaac tgcagcgtgt atgacttttt tgtgtggctc cattattatt ctgttcgaga 9000

cacactttgg ccaagggtga cctaccatat gaacaaatat gcgtatcata tgctggaaag 9060

acgagccaaa tataaaagag gaccaggacc tggcgctaaa tttgtggccg cctggacact 9120

gaaagccgct gctggtcctg gacctggcca gtacatcaag gccaacagca agttcatcgg 9180

catcaccgaa ctcggacccg gaccaggctg atgattcgaa cggccgtatc acgcccaaac 9240

atttacagcc gcggtgtcaa aaaccgcgtg gacgtggtta acatccctgc tgggaggatc 9300

agccgtaatt attataattg gcttggtgct ggctactatt gtggccatgt acgtgctgac 9360

caaccagaaa cataattgaa tacagcagca attggcaagc tgcttacata gaactcgcgg 9420

cgattggcat gccgccttaa aatttttatt ttatttttttc ttttcttttc cgaatcggat 9480

[0061] tttgttttta atatttcaaa aaaaaaaaaa aaaaaaaaaa aaaaaaaaaa aaaaaaaaaa 9540

aaaaaaaaaa aaaaaaaaaa aaaaaaaaaa aaaaaaa 9577

<210> 5
<211> 11447
<212> DNA
<213> 委内瑞拉马脑炎病毒(Venezuelan equine encephalitis virus)

<400> 5
atgggcggcg catgagagaa gcccagacca attacctacc caaaatggag aaagttcacg 60

ttgacatcga ggaagacagc ccattcctca gagctttgca gcggagcttc ccgcagtttg 120

aggtagaagc caagcaggtc actgataatg accatgctaa tgccagagcg ttttcgcatc 180

tggcttcaaa actgatcgaa acggaggtgg acccatccga cacgatcctt gacattggaa 240

gtgcgcccgc ccgcagaatg tattctaagc acaagtatca ttgtatctgt ccgatgagat 300

gtgcggaaga tccggacaga ttgtataagt atgcaactaa gctgaagaaa aactgtaagg 360

aaataactga taaggaattg gacaagaaaa tgaaggagct cgccgccgtc atgagcgacc 420

ctgacctgga aactgagact atgtgcctcc acgacgacga gtcgtgtcgc tacgaagggc 480

aagtcgctgt ttaccaggat gtatacgcgg ttgacggacc gacaagtctc tatcaccaag 540

ccaataaggg agttagagtc gcctactgga taggctttga caccacccct tttatgttta 600

agaacttggc tggagcatat ccatcatact ctaccaactg ggccgacgaa accgtgttaa 660

cggctcgtaa cataggccta tgcagctctg acgttatgga gcggtcacgt agagggatgt 720

ccattcttag aaagaagtat ttgaaaccat ccaacaatgt tctattctct gttggctcga 780

ccatctacca cgagaagagg gacttactga ggagctggca cctgccgtct gtatttcact 840

tacgtggcaa gcaaaattac acatgtcggt gtgagactat agttagttgc gacgggtacg 900

tcgttaaaag aatagctatc agtccaggcc tgtatgggaa gccttcaggc tatgctgcta 960

cgatgcaccg cgagggattc ttgtgctgca aagtgacaga cacattgaac ggggagaggg 1020

tctcttttcc cgtgtgcacg tatgtgccag ctacattgtg tgaccaaatg actggcatac 1080

tggcaacaga tgtcagtgcg gacgacgcgc aaaaactgct ggttgggctc aaccagcgta 1140

tagtcgtcaa cggtcgcacc cagagaaaca ccaataccat gaaaaattac cttttgcccg 1200

[0062] tagtggccca ggcatttgct aggtgggcaa aggaatataa ggaagatcaa gaagatgaaa 1260

ggccactagg actacgagat agacagttag tcatggggtg ttgttgggct tttagaaggc 1320

acaagataac atctatttat aagcgcccgg atacccaaac catcatcaaa gtgaacagcg 1380

atttccactc attcgtgctg cccaggatag gcagtaacac attggagatc gggctgagaa 1440

caagaatcag gaaaatgtta gaggagcaca aggagccgtc acctctcatt accgccgagg 1500

acgtacaaga agctaagtgc gcagccgatg aggctaagga ggtgcgtgaa gccgaggagt 1560

tgcgcgcagc tctaccacct ttggcagctg atgttgagga gcccactctg gaagccgatg 1620

tcgacttgat gttacaagag gctggggccg gctcagtgga gacacctcgt ggcttgataa 1680

aggttaccag ctacgctggc gaggacaaga tcggctctta cgctgtgctt tctccgcagg 1740

ctgtactcaa gagtgaaaaa ttatcttgca tccaccctct cgctgaacaa gtcatagtga 1800

taacacactc tggccgaaaa gggcgttatg ccgtggaacc ataccatggt aaagtagtgg 1860

tgccagaggg acatgcaata cccgtccagg actttcaagc tctgagtgaa agtgccacca 1920

ttgtgtacaa cgaacgtgag ttcgtaaaca ggtacctgca ccatattgcc acacatggag 1980

gagcgctgaa cactgatgaa gaatattaca aaactgtcaa gcccagcgag cacgacggcg 2040

aatacctgta cgacatcgac aggaaacagt gcgtcaagaa agaactagtc actgggctag 2100

ggctcacagg cgagctggtg gatcctccct tccatgaatt cgcctacgag agtctgagaa 2160

cacgaccagc cgctccttac caagtaccaa ccataggggt gtatggcgtg ccaggatcag 2220

gcaagtctgg catcattaaa agcgcagtca ccaaaaaaga tctagtggtg agcgccaaga 2280

aagaaaactg tgcagaaatt ataagggacg tcaagaaaat gaaagggctg gacgtcaatg 2340

ccagaactgt ggactcagtg ctcttgaatg gatgcaaaca ccccgtagag accctgtata 2400

ttgacgaagc ttttgcttgt catgcaggta ctctcagagc gctcatagcc attataagac 2460

ctaaaaaggc agtgctctgc ggggatccca aacagtgcgg ttttttaac atgatgtgcc 2520

tgaaagtgca ttttaaccac gagatttgca cacaagtctt ccacaaaagc atctctcgcc 2580

gttgcactaa atctgtgact tcggtcgtct caaccttgtt ttacgacaaa aaaatgagaa 2640

cgacgaatcc gaaagagact aagattgtga ttgacactac cggcagtacc aaacctaagc 2700

[0063] aggacgatct cattctcact tgtttcagag ggtgggtgaa gcagttgcaa atagattaca 2760

aaggcaacga aataatgacg gcagctgcct ctcaagggct gacccgtaaa ggtgtgtatg 2820

ccgttcggta caaggtgaat gaaaatcctc tgtacgcacc cacctcagaa catgtgaacg 2880

tcctactgac ccgcacggag gaccgcatcg tgtggaaaac actagccggc gacccatgga 2940

taaaaacact gactgccaag taccctggga atttcactgc cacgatagag gagtggcaag 3000

cagagcatga tgccatcatg aggcacatct tggagagacc ggaccctacc gacgtcttcc 3060

agaataaggc aaacgtgtgt tgggccaagg ctttagtgcc ggtgctgaag accgctggca 3120

tagacatgac cactgaacaa tggaacactg tggattattt tgaaacggac aaagctcact 3180

cagcagagat agtattgaac caactatgcg tgaggttctt tggactcgat ctggactccg 3240

gtctattttc tgcacccact gttccgttat ccattaggaa taatcactgg gataactccc 3300

cgtcgcctaa catgtacggg ctgaataaag aagtggtccg tcagctctct cgcaggtacc 3360

cacaactgcc tcgggcagtt gccactggaa gagtctatga catgaacact ggtacactgc 3420

gcaattatga tccgcgcata aacctagtac ctgtaaacag aagactgcct catgctttag 3480

tcctccacca taatgaacac ccacagagtg acttttcttc attcgtcagc aaattgaagg 3540

gcagaactgt cctggtggtc ggggaaaagt tgtccgtccc aggcaaaatg gttgactggt 3600

tgtcagaccg gcctgaggct accttcagag ctcggctgga tttaggcatc ccaggtgatg 3660

tgcccaaata tgacataata tttgttaatg tgaggacccc atataaatac catcactatc 3720

agcagtgtga agaccatgcc attaagctta gcatgttgac caagaaagct tgtctgcatc 3780

tgaatcccgg cggaacctgt gtcagcatag gttatggtta cgctgacagg gccagcgaaa 3840

gcatcattgg tgctatagcg cggcagttca agttttcccg ggtatgcaaa ccgaaatcct 3900

cacttgaaga gacggaagtt ctgtttgtat tcattgggta cgatcgcaag gcccgtacgc 3960

acaatcctta caagctttca tcaaccttga ccaacattta tacaggttcc agactccacg 4020

aagccggatg tgcaccctca tatcatgtgg tgcgagggga tattgccacg gccaccgaag 4080

gagtgattat aaatgctgct aacagcaaag gacaacctgg cggaggggtg tgcggagcgc 4140

tgtataagaa attcccggaa agcttcgatt tacagccgat cgaagtagga aaagcgcgac 4200

[0064] tggtcaaagg tgcagctaaa catatcattc atgccgtagg accaaacttc aacaaagttt 4260

cggaggttga aggtgacaaa cagttggcag aggcttatga gtccatcgct aagattgtca 4320

acgataacaa ttacaagtca gtagcgattc cactgttgtc caccggcatc ttttccggga 4380

acaaagatcg actaacccaa tcattgaacc atttgctgac agctttagac accactgatg 4440

cagatgtagc catatactgc agggacaaga aatgggaaat gactctcaag gaagcagtgg 4500

ctaggagaga agcagtggag gagatatgca tatccgacga ctcttcagtg acagaacctg 4560

atgcagagct ggtgagggtg catccgaaga gttctttggc tggaaggaag ggctacagca 4620

caagcgatgg caaaactttc tcatatttgg aagggaccaa gtttcaccag gcggccaagg 4680

atatagcaga aattaatgcc atgtggcccg ttgcaacgga ggccaatgag caggtatgca 4740

tgtatatcct cggagaaagc atgagcagta ttaggtcgaa atgccccgtc gaagagtcgg 4800

aagcctccac accacctagc acgctgcctt gcttgtgcat ccatgccatg actccagaaa 4860

gagtacagcg cctaaaagcc tcacgtccag aacaaattac tgtgtgctca tcctttccat 4920

tgccgaagta tagaatcact ggtgtgcaga agatccaatg ctcccagcct atattgttct 4980

caccgaaagt gcctgcgtat attcatccaa ggaagtatct cgtggaaaca ccaccggtag 5040

acgagactcc ggagccatcg gcagagaacc aatccacaga ggggacacct gaacaaccac 5100

cacttataac cgaggatgag accaggacta gaacgcctga gccgatcatc atcgaagagg 5160

aagaagagga tagcataagt ttgctgtcag atggcccgac ccaccaggtg ctgcaagtcg 5220

aggcagacat tcacgggccg ccctctgtat ctagctcatc ctggtccatt cctcatgcat 5280

ccgactttga tgtggacagt ttatccatac ttgacaccct ggagggagct agcgtgacca 5340

gcggggcaac gtcagccgag actaactctt acttcgcaaa gagtatggag tttctggcgc 5400

gaccggtgcc tgcgcctcga acagtattca ggaaccctcc acatcccgct ccgcgcacaa 5460

gaacaccgtc acttgcaccc agcagggcct gctcgagaac cagcctagtt tccaccccgc 5520

caggcgtgaa tagggtgatc actagagagg agctcgaggc gcttaccccg tcacgcactc 5580

ctagcaggtc ggtctcgaga accagcctgg tctccaaccc gccaggcgta aatagggtga 5640

ttacaagaga ggagtttgag gcgttcgtag cacaacaaca atgacggttt gatgcgggtg 5700

[0065] catacatctt ttcctccgac accggtcaag ggcatttaca acaaaaatca gtaaggcaaa 5760

cggtgctatc cgaagtggtg ttggagagga ccgaattgga gatttcgtat gccccgcgcc 5820

tcgaccaaga aaaagaagaa ttactacgca agaaattaca gttaaatccc acacctgcta 5880

acagaagcag ataccagtcc aggaaggtgg agaacatgaa agccataaca gctagacgta 5940

ttctgcaagg cctagggcat tatttgaagg cagaaggaaa agtggagtgc taccgaaccc 6000

tgcatcctgt tcctttgtat tcatctagtg tgaaccgtgc cttttcaagc cccaaggtcg 6060

cagtggaagc ctgtaacgcc atgttgaaag agaactttcc gactgtggct tcttactgta 6120

ttattccaga gtacgatgcc tatttggaca tggttgacgg agcttcatgc tgcttagaca 6180

ctgccagttt ttgccctgca aagctgcgca gctttccaaa gaaacactcc tatttggaac 6240

ccacaatacg atcggcagtg ccttcagcga tccagaacac gctccagaac gtcctggcag 6300

ctgccacaaa aagaaattgc aatgtcacgc aaatgagaga attgcccgta ttggattcgg 6360

cggcctttaa tgtggaatgc ttcaagaaat atgcgtgtaa taatgaatat tgggaaacgt 6420

ttaaagaaaa ccccatcagg cttactgaag aaaacgtggt aaattacatt accaaattaa 6480

aaggaccaaa agctgctgct ctttttgcga agacacataa tttgaatatg ttgcaggaca 6540

taccaatgga caggtttgta atggacttaa agagagacgt gaaagtgact ccaggaacaa 6600

aacatactga agaacggccc aaggtacagg tgatccaggc tgccgatccg ctagcaacag 6660

cgtatctgtg cggaatccac cgagagctgg ttaggagatt aaatgcggtc ctgcttccga 6720

acattcatac actgtttgat atgtcggctg aagactttga cgctattata gccgagcact 6780

tccagcctgg ggattgtgtt ctggaaactg acatcgcgtc gtttgataaa agtgaggacg 6840

acgccatggc tctgaccgcg ttaatgattc tggaagactt aggtgtggac gcagagctgt 6900

tgacgctgat tgaggcggct ttcggcgaaa tttcatcaat acatttgccc actaaaacta 6960

aatttaaatt cggagccatg atgaaatctg gaatgttcct cacactgttt gtgaacacag 7020

tcattaacat tgtaatcgca agcagagtgt tgagagaacg gctaaccgga tcaccatgtg 7080

cagcattcat tggagatgac aatatcgtga aaggagtcaa atcggacaaa ttaatggcag 7140

acaggtgcgc cacctggttg aatatggaag tcaagattat agatgctgtg gtgggcgaga 7200

[0066] aagcgcctta tttctgtgga gggtttattt tgtgtgactc cgtgaccggc acagcgtgcc 7260

gtgtggcaga ccccctaaaa aggctgttta agcttggcaa acctctggca gcagacgatg 7320

aacatgatga tgacaggaga agggcattgc atgaagagtc aacacgctgg aaccgagtgg 7380

gtattctttc agagctgtgc aaggcagtag aatcaaggta tgaaaccgta ggaacttcca 7440

tcatagttat ggccatgact actctagcta gcagtgttaa atcattcagc tacctgagag 7500

gggcccctat aactctctac ggctaacctg aatggactac gacatagtct agtccgccaa 7560

gatgttcccg ttccagccaa tgtatccgat gcagccaatg ccctatcgca acccgttcgc 7620

ggccccgcgc aggccctggt tccccagaac cgacccttt ctggcgatgc aggtgcagga 7680

attaacccgc tcgatggcta acctgacgtt caagcaacgc cgggacgcgc cacctgaggg 7740

gccatccgct aagaaaccga agaaggaggc ctcgcaaaaa cagaaagggg gaggccaagg 7800

gaagaagaag aagaaccaag ggaagaagaa ggctaagaca gggccgccta atccgaaggc 7860

acagaatgga aacaagaaga agaccaacaa gaaaccaggc aagagacagc gcatggtcat 7920

gaaattggaa tctgacaaga cgttcccaat catgttggaa gggaagataa acggctacgc 7980

ttgtgtggtc ggagggaagt tattcaggcc gatgcatgtg gaaggcaaga tcgacaacga 8040

cgttctggcc gcgcttaaga cgaagaaagc atccaaatac gatcttgagt atgcagatgt 8100

gccacagaac atgcgggccg atacattcaa atacacccat gagaaacccc aaggctatta 8160

cagctggcat catggagcag tccaatatga aaatgggcgt ttcacggtgc cgaaaggagt 8220

tggggccaag ggagacagcg gacgacccat tctggataac cagggacggg tggtcgctat 8280

tgtgctggga ggtgtgaatg aaggatctag gacagccctt tcagtcgtca tgtggaacga 8340

gaagggagtt accgtgaagt atactccgga gaactgcgag caatggtcac tagtgaccac 8400

catgtgtctg ctcgccaatg tgacgttccc atgtgctcaa ccaccaattt gctacgacag 8460

aaaaccagca gagactttgg ccatgctcag cgttaacgtt gacaacccgg gctacgatga 8520

gctgctggaa gcagctgtta agtgccccgg aaggaaaagg agatccaccg aggagctgtt 8580

taaggagtat aagctaacgc gcccttacat ggccagatgc atcagatgtg cagttgggag 8640

ctgccatagt ccaatagcaa tcgaggcagt aaagagcgac gggcacgacg gttatgttag 8700

[0067] acttcagact tcctcgcagt atggcctgga ttcctccggc aacttaaagg gcaggaccat 8760

gcggtatgac atgcacggga ccattaaaga gataccacta catcaagtgt cactccatac 8820

atctcgcccg tgtcacattg tggatgggca cggttatttc ctgcttgcca ggtgcccggc 8880

aggggactcc atcaccatgg aatttaagaa agattccgtc acacactcct gctcggtgcc 8940

gtatgaagtg aaatttaatc ctgtaggcag agaactctat actcatcccc cagaacacgg 9000

agtagagcaa gcgtgccaag tctacgcaca tgatgcacag aacagaggag cttatgtcga 9060

gatgcacctc ccgggctcag aagtggacag cagtttggtt tccttgagcg gcagttcagt 9120

caccgtgaca cctcctgttg ggactagcgc cctggtggaa tgcgagtgtg gcggcacaaa 9180

gatctccgag accatcaaca agacaaaaca gttcagccag tgcacaaaga aggagcagtg 9240

cagagcatat cggctgcaga acgataagtg ggtgtataat tctgacaaac tgcccaaagc 9300

agcgggagcc accttaaaag gaaaactgca tgtcccattc ttgctggcag acggcaaatg 9360

caccgtgcct ctagcaccag aacctatgat aacctttggt ttcagatcag tgtcactgaa 9420

actgcaccct aagaatccca catatctaac cacccgccaa cttgctgatg agcctcacta 9480

cacgcacgag ctcatatctg aaccagctgt taggaatttt accgtcaccg aaaaagggtg 9540

ggagtttgta tggggaaacc acccgccgaa aaggttttgg gcacaggaaa cagcacccgg 9600

aaatccacat gggctaccgc acgaggtgat aactcattat taccacagat accctatgtc 9660

caccatcctg ggtttgtcaa tttgtgccgc cattgcaacc gtttccgttg cagcgtctac 9720

ctggctgttt tgcagatcta gagttgcgtg cctaactcct taccggctaa cacctaacgc 9780

taggatacca ttttgtctgg ctgtgctttg ctgcgcccgc actgcccggg ccgagaccac 9840

ctgggagtcc ttggatcacc tatggaacaa taaccaacag atgttctgga ttcaattgct 9900

gatccctctg gccgccttga tcgtagtgac tcgcctgctc aggtgcgtgt gctgtgtcgt 9960

gccttttta gtcatggccg gcgccgcagg cgccggcgcc tacgagcacg cgaccacgat 10020

gccgagccaa gcgggaatct cgtataacac tatagtcaac agagcaggct acgcaccact 10080

ccctatcagc ataacaccaa caaagatcaa gctgatacct acagtgaact tggagtacgt 10140

cacctgccac tacaaaacag gaatggattc accagccatc aaatgctgcg gatctcagga 10200

[0068] atgcactcca acttacaggc ctgatgaaca gtgcaaagtc ttcacagggg tttacccgtt 10260

catgtggggt ggtgcatatt gcttttgcga cactgagaac acccaagtca gcaaggccta 10320

cgtaatgaaa tctgacgact gccttgcgga tcatgctgaa gcatataaag cgcacacagc 10380

ctcagtgcag gcgttcctca acatcacagt gggagaacac tctattgtga ctaccgtgta 10440

tgtgaatgga gaaactcctg tgaatttcaa tggggtcaaa ttaactgcag gtccgctttc 10500

cacagcttgg acacccttg atcgcaaaat cgtgcagtat gccggggaga tctataatta 10560

tgattttcct gagtatgggg caggacaacc aggagcattt ggagatatac aatccagaac 10620

agtctcaagc tcagatctgt atgccaatac caacctagtg ctgcagagac ccaaagcagg 10680

agcgatccac gtgccataca ctcaggcacc ttcgggtttt gagcaatgga agaaagataa 10740

agctccatca ttgaaattta ccgccccttt cggatgcgaa atatatacaa accccattcg 10800

cgccgaaaac tgtgctgtag ggtcaattcc attagccttt gacattcccg acgccttgtt 10860

caccagggtg tcagaaacac cgacactttc agcggccgaa tgcactctta acgagtgcgt 10920

gtattcttcc gactttggtg ggatcgccac ggtcaagtac tcggccagca agtcaggcaa 10980

gtgcgcagtc catgtgccat cagggactgc taccctaaaa gaagcagcag tcgagctaac 11040

cgagcaaggg tcggcgacta tccatttctc gaccgcaaat atccacccgg agttcaggct 11100

ccaaatatgc acatcatatg ttacgtgcaa aggtgattgt caccccccga aagaccatat 11160

tgtgacacac cctcagtatc acgcccaaac atttacagcc gcggtgtcaa aaaccgcgtg 11220

gacgtggtta acatccctgc tgggaggatc agccgtaatt attataattg gcttggtgct 11280

ggctactatt gtggccatgt acgtgctgac caaccagaaa cataattgaa tacagcagca 11340

attggcaagc tgcttacata gaactcgcgg cgattggcat gccgccttaa aatttttatt 11400

ttatttttc ttttcttttc cgaatcggat tttgttttta atatttc 11447

<210> 6
<211> 7894
<212> DNA
<213> 人工序列(Artificial Sequence)

<220>
<223> 人工序列的描述：合成多核苷酸

[0069]

<400> 6
atgggcggcg catgagagaa gcccagacca attacctacc caaaatggag aaagttcacg 60

ttgacatcga ggaagacagc ccattcctca gagctttgca gcggagcttc ccgcagtttg 120

aggtagaagc caagcaggtc actgataatg accatgctaa tgccagagcg ttttcgcatc 180

tggcttcaaa actgatcgaa acggaggtgg acccatccga cacgatcctt gacattggaa 240

gtgcgcccgc ccgcagaatg tattctaagc acaagtatca ttgtatctgt ccgatgagat 300

gtgcggaaga tccggacaga ttgtataagt atgcaactaa gctgaagaaa aactgtaagg 360

aaataactga taaggaattg gacaagaaaa tgaaggagct cgccgccgtc atgagcgacc 420

ctgacctgga aactgagact atgtgcctcc acgacgacga gtcgtgtcgc tacgaagggc 480

aagtcgctgt ttaccaggat gtatacgcgg ttgacggacc gacaagtctc tatcaccaag 540

ccaataaggg agttagagtc gcctactgga taggctttga caccacccct tttatgttta 600

agaacttggc tggagcatat ccatcatact ctaccaactg ggccgacgaa accgtgttaa 660

cggctcgtaa cataggccta tgcagctctg acgttatgga gcggtcacgt agagggatgt 720

ccattcttag aaagaagtat ttgaaaccat ccaacaatgt tctattctct gttggctcga 780

ccatctacca cgagaagagg gacttactga ggagctggca cctgccgtct gtatttcact 840

tacgtggcaa gcaaaattac acatgtcggt gtgagactat agttagttgc gacgggtacg 900

tcgttaaaag aatagctatc agtccaggcc tgtatgggaa gccttcaggc tatgctgcta 960

cgatgcaccg cgagggattc ttgtgctgca aagtgacaga cacattgaac ggggagaggg 1020

tctcttttcc cgtgtgcacg tatgtgccag ctacattgtg tgaccaaatg actggcatac 1080

tggcaacaga tgtcagtgcg gacgacgcgc aaaaactgct ggttgggctc aaccagcgta 1140

tagtcgtcaa cggtcgcacc cagagaaaca ccaataccat gaaaaattac cttttgcccg 1200

tagtggccca ggcatttgct aggtgggcaa aggaatataa ggaagatcaa gaagatgaaa 1260

ggccactagg actacgagat agacagttag tcatggggtg ttgttgggct tttagaaggc 1320

acaagataac atctatttat aagcgcccgg atacccaaac catcatcaaa gtgaacagcg 1380

atttccactc attcgtgctg cccaggatag gcagtaacac attggagatc gggctgagaa 1440

[0070] caagaatcag gaaaatgtta gaggagcaca aggagccgtc acctctcatt accgccgagg 1500

acgtacaaga agctaagtgc gcagccgatg aggctaagga ggtgcgtgaa gccgaggagt 1560

tgcgcgcagc tctaccacct ttggcagctg atgttgagga gcccactctg gaagccgatg 1620

tcgacttgat gttacaagag gctggggccg gctcagtgga gacacctcgt ggcttgataa 1680

aggttaccag ctacgctggc gaggacaaga tcggctctta cgctgtgctt tctccgcagg 1740

ctgtactcaa gagtgaaaaa ttatcttgca tccaccctct cgctgaacaa gtcatagtga 1800

taacacactc tggccgaaaa gggcgttatg ccgtggaacc ataccatggt aaagtagtgg 1860

tgccagaggg acatgcaata cccgtccagg actttcaagc tctgagtgaa agtgccacca 1920

ttgtgtacaa cgaacgtgag ttcgtaaaca ggtacctgca ccatattgcc acacatggag 1980

gagcgctgaa cactgatgaa gaatattaca aaactgtcaa gcccagcgag cacgacggcg 2040

aatacctgta cgacatcgac aggaaacagt gcgtcaagaa agaactagtc actgggctag 2100

ggctcacagg cgagctggtg gatcctccct tccatgaatt cgcctacgag agtctgagaa 2160

cacgaccagc cgctccttac caagtaccaa ccataggggt gtatggcgtg ccaggatcag 2220

gcaagtctgg catcattaaa agcgcagtca ccaaaaaaga tctagtggtg agcgccaaga 2280

aagaaaactg tgcagaaatt ataagggacg tcaagaaaat gaaagggctg gacgtcaatg 2340

ccagaactgt ggactcagtg ctcttgaatg gatgcaaaca ccccgtagag accctgtata 2400

ttgacgaagc ttttgcttgt catgcaggta ctctcagagc gctcatagcc attataagac 2460

ctaaaaaggc agtgctctgc ggggatccca aacagtgcgg ttttttaac atgatgtgcc 2520

tgaaagtgca ttttaaccac gagatttgca cacaagtctt ccacaaaagc atctctcgcc 2580

gttgcactaa atctgtgact tcggtcgtct caaccttgtt ttacgacaaa aaaatgagaa 2640

cgacgaatcc gaaagagact aagattgtga ttgacactac cggcagtacc aaacctaagc 2700

aggacgatct cattctcact tgtttcagag ggtgggtgaa gcagttgcaa atagattaca 2760

aaggcaacga aataatgacg gcagctgcct ctcaagggct gacccgtaaa ggtgtgtatg 2820

ccgttcggta caaggtgaat gaaaatcctc tgtacgcacc cacctcagaa catgtgaacg 2880

tcctactgac ccgcacggag gaccgcatcg tgtggaaaac actagccggc gacccatgga 2940

[0071] taaaaacact gactgccaag taccctggga atttcactgc cacgatagag gagtggcaag 3000

cagagcatga tgccatcatg aggcacatct tggagagacc ggaccctacc gacgtcttcc 3060

agaataaggc aaacgtgtgt tgggccaagg ctttagtgcc ggtgctgaag accgctggca 3120

tagacatgac cactgaacaa tggaacactg tggattattt tgaaacggac aaagctcact 3180

cagcagagat agtattgaac caactatgcg tgaggttctt tggactcgat ctggactccg 3240

gtctattttc tgcacccact gttccgttat ccattaggaa taatcactgg gataactccc 3300

cgtcgcctaa catgtacggg ctgaataaag aagtggtccg tcagctctct cgcaggtacc 3360

cacaactgcc tcgggcagtt gccactggaa gagtctatga catgaacact ggtacactgc 3420

gcaattatga tccgcgcata aacctagtac ctgtaaacag aagactgcct catgctttag 3480

tcctccacca taatgaacac ccacagagtg acttttcttc attcgtcagc aaattgaagg 3540

gcagaactgt cctggtggtc ggggaaaagt tgtccgtccc aggcaaaatg gttgactggt 3600

tgtcagaccg gcctgaggct accttcagag ctcggctgga tttaggcatc ccaggtgatg 3660

tgcccaaata tgacataata tttgttaatg tgaggacccc atataaatac catcactatc 3720

agcagtgtga agaccatgcc attaagctta gcatgttgac caagaaagct tgtctgcatc 3780

tgaatcccgg cggaacctgt gtcagcatag gttatggtta cgctgacagg gccagcgaaa 3840

gcatcattgg tgctatagcg cggcagttca agttttcccg ggtatgcaaa ccgaaatcct 3900

cacttgaaga gacggaagtt ctgtttgtat tcattgggta cgatcgcaag gcccgtacgc 3960

acaatcctta caagctttca tcaaccttga ccaacattta tacaggttcc agactccacg 4020

aagccggatg tgcaccctca tatcatgtgg tgcgagggga tattgccacg gccaccgaag 4080

gagtgattat aaatgctgct aacagcaaag gacaacctgg cggaggggtg tgcggagcgc 4140

tgtataagaa attcccggaa agcttcgatt tacagccgat cgaagtagga aaagcgcgac 4200

tggtcaaagg tgcagctaaa catatcattc atgccgtagg accaaacttc aacaaagttt 4260

cggaggttga aggtgacaaa cagttggcag aggcttatga gtccatcgct aagattgtca 4320

acgataacaa ttacaagtca gtagcgattc cactgttgtc caccggcatc ttttccggga 4380

acaaagatcg actaacccaa tcattgaacc atttgctgac agctttagac accactgatg 4440

[0072] cagatgtagc catatactgc agggacaaga aatgggaaat gactctcaag gaagcagtgg 4500

ctaggagaga agcagtggag gagatatgca tatccgacga ctcttcagtg acagaacctg 4560

atgcagagct ggtgagggtg catccgaaga gttctttggc tggaaggaag ggctacagca 4620

caagcgatgg caaaactttc tcatatttgg aagggaccaa gtttcaccag gcggccaagg 4680

atatagcaga aattaatgcc atgtggcccg ttgcaacgga ggccaatgag caggtatgca 4740

tgtatatcct cggagaaagc atgagcagta ttaggtcgaa atgccccgtc gaagagtcgg 4800

aagcctccac accacctagc acgctgcctt gcttgtgcat ccatgccatg actccagaaa 4860

gagtacagcg cctaaaagcc tcacgtccag aacaaattac tgtgtgctca tcctttccat 4920

tgccgaagta tagaatcact ggtgtgcaga agatccaatg ctcccagcct atattgttct 4980

caccgaaagt gcctgcgtat attcatccaa ggaagtatct cgtggaaaca ccaccggtag 5040

acgagactcc ggagccatcg gcagagaacc aatccacaga ggggacacct gaacaaccac 5100

cacttataac cgaggatgag accaggacta gaacgcctga gccgatcatc atcgaagagg 5160

aagaagagga tagcataagt ttgctgtcag atggcccgac ccaccaggtg ctgcaagtcg 5220

aggcagacat tcacgggccg ccctctgtat ctagctcatc ctggtccatt cctcatgcat 5280

ccgactttga tgtggacagt ttatccatac ttgacaccct ggagggagct agcgtgacca 5340

gcggggcaac gtcagccgag actaactctt acttcgcaaa gagtatggag tttctggcgc 5400

gaccggtgcc tgcgcctcga acagtattca ggaaccctcc acatcccgct ccgcgcacaa 5460

gaacaccgtc acttgcaccc agcagggcct gctcgagaac cagcctagtt tccaccccgc 5520

caggcgtgaa tagggtgatc actagagagg agctcgaggc gcttaccccg tcacgcactc 5580

ctagcaggtc ggtctcgaga accagcctgg tctccaaccc gccaggcgta aatagggtga 5640

ttacaagaga ggagtttgag gcgttcgtag cacaacaaca atgacggttt gatgcgggtg 5700

catacatctt ttcctccgac accggtcaag ggcatttaca acaaaaatca gtaaggcaaa 5760

cggtgctatc cgaagtggtg ttggagagga ccgaattgga gatttcgtat gccccgcgcc 5820

tcgaccaaga aaaagaagaa ttactacgca agaaattaca gttaaatccc acacctgcta 5880

acagaagcag ataccagtcc aggaaggtgg agaacatgaa agccataaca gctagacgta 5940

[0073] ttctgcaagg cctagggcat tatttgaagg cagaaggaaa agtggagtgc taccgaaccc 6000

tgcatcctgt tcctttgtat tcatctagtg tgaaccgtgc cttttcaagc cccaaggtcg 6060

cagtggaagc ctgtaacgcc atgttgaaag agaactttcc gactgtggct tcttactgta 6120

ttattccaga gtacgatgcc tatttggaca tggttgacgg agcttcatgc tgcttagaca 6180

ctgccagttt ttgccctgca aagctgcgca gctttccaaa gaaacactcc tatttggaac 6240

ccacaatacg atcggcagtg ccttcagcga tccagaacac gctccagaac gtcctggcag 6300

ctgccacaaa aagaaattgc aatgtcacgc aaatgagaga attgcccgta ttggattcgg 6360

cggcctttaa tgtggaatgc ttcaagaaat atgcgtgtaa taatgaatat tgggaaacgt 6420

ttaaagaaaa ccccatcagg cttactgaag aaaacgtggt aaattacatt accaaattaa 6480

aaggaccaaa agctgctgct ctttttgcga agacacataa tttgaatatg ttgcaggaca 6540

taccaatgga caggtttgta atggacttaa agagagacgt gaaagtgact ccaggaacaa 6600

aacatactga agaacggccc aaggtacagg tgatccaggc tgccgatccg ctagcaacag 6660

cgtatctgtg cggaatccac cgagagctgg ttaggagatt aaatgcggtc ctgcttccga 6720

acattcatac actgtttgat atgtcggctg aagactttga cgctattata gccgagcact 6780

tccagcctgg ggattgtgtt ctggaaactg acatcgcgtc gtttgataaa agtgaggacg 6840

acgccatggc tctgaccgcg ttaatgattc tggaagactt aggtgtggac gcagagctgt 6900

tgacgctgat tgaggcggct ttcggcgaaa tttcatcaat acatttgccc actaaaacta 6960

aatttaaatt cggagccatg atgaaatctg gaatgttcct cacactgttt gtgaacacag 7020

tcattaacat tgtaatcgca agcagagtgt tgagagaacg gctaaccgga tcaccatgtg 7080

cagcattcat tggagatgac aatatcgtga aaggagtcaa atcggacaaa ttaatggcag 7140

acaggtgcgc cacctggttg aatatggaag tcaagattat agatgctgtg gtgggcgaga 7200

aagcgcctta tttctgtgga gggtttattt tgtgtgactc cgtgaccggc acagcgtgcc 7260

gtgtggcaga ccccctaaaa aggctgttta agcttggcaa acctctggca gcagacgatg 7320

aacatgatga tgacaggaga agggcattgc atgaagagtc aacacgctgg aaccgagtgg 7380

gtattctttc agagctgtgc aaggcagtag aatcaaggta tgaaaccgta ggaacttcca 7440

[0074]

tcatagttat ggccatgact actctagcta gcagtgttaa atcattcagc tacctgagag 7500

gggcccctat aactctctac ggctaacctg aatggactac gactatcacg cccaaacatt 7560

tacagccgcg gtgtcaaaaa ccgcgtggac gtggttaaca tccctgctgg gaggatcagc 7620

cgtaattatt ataattggct tggtgctggc tactattgtg gccatgtacg tgctgaccaa 7680

ccagaaacat aattgaatac agcagcaatt ggcaagctgc ttacatagaa ctcgcggcga 7740

ttggcatgcc gccttaaaat ttttatttta ttttttcttt tcttttccga atcggatttt 7800

gtttttaata tttcaaaaaa aaaaaaaaaa aaaaaaaaaa aaaaaaaaaa aaaaaaaaaa 7860

aaaaaaaaaa aaaaaaaaaa aaaaaaaaaa aaaa 7894

<210> 7
<211> 7893
<212> DNA
<213> 人工序列(Artificial Sequence)

<220>
<223> 人工序列的描述：合成多核苷酸

<400> 7

[0075]

ataggcggcg catgagagaa gcccagacca attacctacc caaaatggag aaagttcacg 60

ttgacatcga ggaagacagc ccattcctca gagctttgca gcggagcttc ccgcagtttg 120

aggtagaagc caagcaggtc actgataatg accatgctaa tgccagagcg ttttcgcatc 180

tggcttcaaa actgatcgaa acggaggtgg acccatccga cacgatcctt gacattggaa 240

gtgcgcccgc ccgcagaatg tattctaagc acaagtatca ttgtatctgt ccgatgagat 300

gtgcggaaga tccggacaga ttgtataagt atgcaactaa gctgaagaaa aactgtaagg 360

aaataactga taaggaattg gacaagaaaa tgaaggagct cgccgccgtc atgagcgacc 420

ctgacctgga aactgagact atgtgcctcc acgacgacga gtcgtgtcgc tacgaagggc 480

aagtcgctgt ttaccaggat gtatacgcgg ttgacggacc gacaagtctc tatcaccaag 540

ccaataaggg agttagagtc gcctactgga taggctttga caccacccct tttatgttta 600

agaacttggc tggagcatat ccatcatact ctaccaactg ggccgacgaa accgtgttaa 660

cggctcgtaa cataggccta tgcagctctg acgttatgga gcggtcacgt agagggatgt 720

ccattcttag aaagaagtat ttgaaaccat ccaacaatgt tctattctct gttggctcga 780

ccatctacca cgagaagagg gacttactga ggagctggca cctgccgtct gtatttcact 840

tacgtggcaa gcaaaattac acatgtcggt gtgagactat agttagttgc gacgggtacg 900

tcgttaaaag aatagctatc agtccaggcc tgtatgggaa gccttcaggc tatgctgcta 960

cgatgcaccg cgagggattc ttgtgctgca aagtgacaga cacattgaac ggggagaggg 1020

tctcttttcc cgtgtgcacg tatgtgccag ctacattgtg tgaccaaatg actggcatac 1080

tggcaacaga tgtcagtgcg gacgacgcgc aaaaactgct ggttgggctc aaccagcgta 1140

tagtcgtcaa cggtcgcacc cagagaaaca ccaataccat gaaaaattac cttttgcccg 1200

tagtggccca ggcatttgct aggtgggcaa aggaatataa ggaagatcaa gaagatgaaa 1260

ggccactagg actacgagat agacagttag tcatggggtg ttgttgggct tttagaaggc 1320

acaagataac atctatttat aagcgcccgg atacccaaac catcatcaaa gtgaacagcg 1380

atttccactc attcgtgctg cccaggatag gcagtaacac attggagatc gggctgagaa 1440

caagaatcag gaaaatgtta gaggagcaca aggagccgtc acctctcatt accgccgagg 1500

acgtacaaga agctaagtgc gcagccgatg aggctaagga ggtgcgtgaa gccgaggagt 1560

tgcgcgcagc tctaccacct ttggcagctg atgttgagga gcccactctg gaagccgatg 1620

tcgacttgat gttacaagag gctggggccg gctcagtgga gacacctcgt ggcttgataa 1680

aggttaccag ctacgatggc gaggacaaga tcggctctta cgctgtgctt tctccgcagg 1740

ctgtactcaa gagtgaaaaa ttatcttgca tccaccctct cgctgaacaa gtcatagtga 1800

taacacactc tggccgaaaa gggcgttatg ccgtggaacc ataccatggt aaagtagtgg 1860

tgccagaggg acatgcaata cccgtccagg actttcaagc tctgagtgaa agtgccacca 1920

ttgtgtacaa cgaacgtgag ttcgtaaaca ggtacctgca ccatattgcc acacatggag 1980

gagcgctgaa cactgatgaa gaatattaca aaactgtcaa gcccagcgag cacgacggcg 2040

aatacctgta cgacatcgac aggaaacagt gcgtcaagaa agaactagtc actgggctag 2100

ggctcacagg cgagctggtg gatcctccct tccatgaatt cgcctacgag agtctgagaa 2160

cacgaccagc cgctccttac caagtaccaa ccataggggt gtatggcgtg ccaggatcag 2220

[0076] gcaagtctgg catcattaaa agcgcagtca ccaaaaaaga tctagtggtg agcgccaaga 2280

aagaaaactg tgcagaaatt ataagggacg tcaagaaaat gaaagggctg gacgtcaatg 2340

ccagaactgt ggactcagtg ctcttgaatg gatgcaaaca ccccgtagag accctgtata 2400

ttgacgaagc ttttgcttgt catgcaggta ctctcagagc gctcatagcc attataagac 2460

ctaaaaaggc agtgctctgc ggggatccca aacagtgcgg ttttttaac atgatgtgcc 2520

tgaaagtgca ttttaaccac gagatttgca cacaagtctt ccacaaaagc atctctcgcc 2580

gttgcactaa atctgtgact tcggtcgtct caaccttgtt ttacgacaaa aaaatgagaa 2640

cgacgaatcc gaaagagact aagattgtga ttgacactac cggcagtacc aaacctaagc 2700

aggacgatct cattctcact tgtttcagag ggtgggtgaa gcagttgcaa atagattaca 2760

aaggcaacga aataatgacg gcagctgcct ctcaagggct gacccgtaaa ggtgtgtatg 2820

ccgttcggta caaggtgaat gaaaatcctc tgtacgcacc cacctcagaa catgtgaacg 2880

tcctactgac ccgcacggag gaccgcatcg tgtggaaaac actagccggc gacccatgga 2940

taaaaacact gactgccaag taccctggga atttcactgc cacgatagag gagtggcaag 3000

cagagcatga tgccatcatg aggcacatct tggagagacc ggaccctacc gacgtcttcc 3060

agaataaggc aaacgtgtgt tgggccaagg ctttagtgcc ggtgctgaag accgctggca 3120

tagacatgac cactgaacaa tggaacactg tggattattt tgaaacggac aaagctcact 3180

cagcagagat agtattgaac caactatgcg tgaggttctt tggactcgat ctggactccg 3240

gtctattttc tgcacccact gttccgttat ccattaggaa taatcactgg gataactccc 3300

cgtcgcctaa catgtacggg ctgaataaag aagtggtccg tcagctctct cgcaggtacc 3360

cacaactgcc tcgggcagtt gccactggaa gagtctatga catgaacact ggtacactgc 3420

gcaattatga tccgcgcata aacctagtac ctgtaaacag aagactgcct catgctttag 3480

tcctccacca taatgaacac ccacagagtg acttttcttc attcgtcagc aaattgaagg 3540

gcagaactgt cctggtggtc ggggaaaagt tgtccgtccc aggcaaaatg gttgactggt 3600

tgtcagaccg gcctgaggct accttcagag ctcggctgga tttaggcatc ccaggtgatg 3660

tgcccaaata tgacataata tttgttaatg tgaggacccc atataaatac catcactatc 3720

[0077] agcagtgtga agaccatgcc attaagctta gcatgttgac caagaaagct tgtctgcatc 3780

tgaatcccgg cggaacctgt gtcagcatag gttatggtta cgctgacagg gccagcgaaa 3840

gcatcattgg tgctatagcg cggcagttca agttttcccg ggtatgcaaa ccgaaatcct 3900

cacttgaaga gacggaagtt ctgtttgtat tcattgggta cgatcgcaag gcccgtacgc 3960

acaatcctta caagctttca tcaaccttga ccaacattta tacaggttcc agactccacg 4020

aagccggatg tgcaccctca tatcatgtgg tgcgagggga tattgccacg gccaccgaag 4080

gagtgattat aaatgctgct aacagcaaag gacaacctgg cggaggggtg tgcggagcgc 4140

tgtataagaa attcccggaa agcttcgatt tacagccgat cgaagtagga aaagcgcgac 4200

tggtcaaagg tgcagctaaa catatcattc atgccgtagg accaaacttc aacaaagttt 4260

cggaggttga aggtgacaaa cagttggcag aggcttatga gtccatcgct aagattgtca 4320

acgataacaa ttacaagtca gtagcgattc cactgttgtc caccggcatc ttttccggga 4380

acaaagatcg actaacccaa tcattgaacc atttgctgac agctttagac accactgatg 4440

cagatgtagc catatactgc agggacaaga aatgggaaat gactctcaag gaagcagtgg 4500

ctaggagaga agcagtggag gagatatgca tatccgacga ctcttcagtg acagaacctg 4560

atgcagagct ggtgagggtg catccgaaga gttctttggc tggaaggaag ggctacagca 4620

caagcgatgg caaaactttc tcatatttgg aagggaccaa gtttcaccag gcggccaagg 4680

atatagcaga aattaatgcc atgtggcccg ttgcaacgga ggccaatgag caggtatgca 4740

tgtatatcct cggagaaagc atgagcagta ttaggtcgaa atgccccgtc gaagagtcgg 4800

aagcctccac accacctagc acgctgcctt gcttgtgcat ccatgccatg actccagaaa 4860

gagtacagcg cctaaaagcc tcacgtccag aacaaattac tgtgtgctca tcctttccat 4920

tgccgaagta tagaatcact ggtgtgcaga agatccaatg ctcccagcct atattgttct 4980

caccgaaagt gcctgcgtat attcatccaa ggaagtatct cgtggaaaca ccaccggtag 5040

acgagactcc ggagccatcg gcagagaacc aatccacaga ggggacacct gaacaaccac 5100

cacttataac cgaggatgag accaggacta gaacgcctga gccgatcatc atcgaagagg 5160

aagaagagga tagcataagt ttgctgtcag atggcccgac ccaccaggtg ctgcaagtcg 5220

[0078] aggcagacat tcacgggccg ccctctgtat ctagctcatc ctggtccatt cctcatgcat 5280

ccgactttga tgtggacagt ttatccatac ttgacaccct ggagggagct agcgtgacca 5340

gcggggcaac gtcagccgag actaactctt acttcgcaaa gagtatggag tttctggcgc 5400

gaccggtgcc tgcgcctcga acagtattca ggaaccctcc acatcccgct ccgcgcacaa 5460

gaacaccgtc acttgcaccc agcagggcct gctcgagaac cagcctagtt tccacccccgc 5520

caggcgtgaa tagggtgatc actagagagg agctcgaggc gcttaccccg tcacgcactc 5580

ctagcaggtc ggtctcgaga accagcctgg tctccaaccc gccaggcgta aatagggtga 5640

ttacaagaga ggagtttgag gcgttcgtag cacaacaaca atgacggttt gatgcgggtg 5700

catacatctt ttcctccgac accggtcaag ggcatttaca acaaaaatca gtaaggcaaa 5760

cggtgctatc cgaagtggtg ttggagagga ccgaattgga gatttcgtat gccccgcgcc 5820

tcgaccaaga aaaagaagaa ttactacgca agaaattaca gttaaatccc acacctgcta 5880

acagaagcag ataccagtcc aggaaggtgg agaacatgaa agccataaca gctagacgta 5940

ttctgcaagg cctagggcat tatttgaagg cagaaggaaa agtggagtgc taccgaaccc 6000

tgcatcctgt tcctttgtat tcatctagtg tgaaccgtgc cttttcaagc cccaaggtcg 6060

cagtggaagc ctgtaacgcc atgttgaaag agaactttcc gactgtggct tcttactgta 6120

ttattccaga gtacgatgcc tatttggaca tggttgacgg agcttcatgc tgcttagaca 6180

ctgccagttt ttgccctgca aagctgcgca gctttccaaa gaaacactcc tatttggaac 6240

ccacaatacg atcggcagtg ccttcagcga tccagaacac gctccagaac gtcctggcag 6300

ctgccacaaa aagaaattgc aatgtcacgc aaatgagaga attgcccgta ttggattcgg 6360

cggcctttaa tgtggaatgc ttcaagaaat atgcgtgtaa taatgaatat tgggaaacgt 6420

ttaaagaaaa ccccatcagg cttactgaag aaaacgtggt aaattacatt accaaattaa 6480

aaggaccaaa agctgctgct ctttttgcga agacacataa tttgaatatg ttgcaggaca 6540

taccaatgga caggtttgta atggacttaa agagagacgt gaaagtgact ccaggaacaa 6600

aacatactga agaacggccc aaggtacagg tgatccaggc tgccgatccg ctagcaacag 6660

cgtatctgtg cggaatccac cgagagctgg ttaggagatt aaatgcggtc ctgcttccga 6720

[0079] acattcatac actgtttgat atgtcggctg aagactttga cgctattata gccgagcact 6780

tccagcctgg ggattgtgtt ctggaaactg acatcgcgtc gtttgataaa agtgaggacg 6840

acgccatggc tctgaccgcg ttaatgattc tggaagactt aggtgtggac gcagagctgt 6900

tgacgctgat tgaggcggct ttcggcgaaa tttcatcaat acatttgccc actaaaacta 6960

aatttaaatt cggagccatg atgaaatctg gaatgttcct cacactgttt gtgaacacag 7020

tcattaacat tgtaatcgca agcagagtgt tgagagaacg gctaaccgga tcaccatgtg 7080

cagcattcat tggagatgac aatatcgtga aaggagtcaa atcggacaaa ttaatggcag 7140

acaggtgcgc cacctggttg aatatggaag tcaagattat agatgctgtg gtgggcgaga 7200

aagcgcctta tttctgtgga gggtttattt tgtgtgactc cgtgaccggc acagcgtgcc 7260

gtgtggcaga ccccctaaaa aggctgttta agcttggcaa acctctggca gcagacgatg 7320

aacatgatga tgacaggaga agggcattgc atgaagagtc aacacgctgg aaccgagtgg 7380

gtattctttc agagctgtgc aaggcagtag aatcaaggta tgaaaccgta ggaacttcca 7440

tcatagttat ggccatgact actctagcta gcagtgttaa atcattcagc tacctgagag 7500

gggcccctat aactctctac ggctaacctg aatggactac gactatcacg cccaaacatt 7560

tacagccgcg gtgtcaaaaa ccgcgtggac gtggttaaca tccctgctgg gaggatcagc 7620

cgtaattatt ataattggct tggtgctggc tactattgtg gccatgtacg tgctgaccaa 7680

ccagaaacat aattgaatac agcagcaatt ggcaagctgc ttacatagaa ctcgcggcga 7740

ttggcatgcc gccttaaaat ttttatttta ttttctttt cttttccgaa tcggattttg 7800

tttttaatat ttcaaaaaaa aaaaaaaaaa aaaaaaaaaa aaaaaaaaaa aaaaaaaaaa 7860

aaaaaaaaaa aaaaaaaaaa aaaaaaaaaa aaa 7893

<210> 8
<211> 7927
<212> DNA
<213> 人工序列(Artificial Sequence)

<220>
<223> 人工序列的描述：合成多核苷酸

<400> 8
[0080] taatacgact cactatagga tgggcggcgc atgagagaag cccagaccaa ttacctaccc 60

aaaatggaga aagttcacgt tgacatcgag gaagacagcc cattcctcag agctttgcag 120

cggagcttcc cgcagtttga ggtagaagcc aagcaggtca ctgataatga ccatgctaat 180

gccagagcgt tttcgcatct ggcttcaaaa ctgatcgaaa cggaggtgga cccatccgac 240

acgatccttg acattggaag tgcgcccgcc cgcagaatgt attctaagca caagtatcat 300

tgtatctgtc cgatgagatg tgcggaagat ccggacagat tgtataagta tgcaactaag 360

ctgaagaaaa actgtaagga aataactgat aaggaattgg acaagaaaat gaaggagctc 420

gccgccgtca tgagcgaccc tgacctggaa actgagacta tgtgcctcca cgacgacgag 480

tcgtgtcgct acgaagggca agtcgctgtt taccaggatg tatacgcggt tgacggaccg 540

acaagtctct atcaccaagc caataaggga gttagagtcg cctactggat aggctttgac 600

accacccctt ttatgtttaa gaacttggct ggagcatatc catcatactc taccaactgg 660

gccgacgaaa ccgtgttaac ggctcgtaac ataggcctat gcagctctga cgttatggag 720

cggtcacgta gagggatgtc cattcttaga aagaagtatt tgaaaccatc caacaatgtt 780

ctattctctg ttggctcgac catctaccac gagaagaggg acttactgag gagctggcac 840

ctgccgtctg tatttcactt acgtggcaag caaaattaca catgtcggtg tgagactata 900

gttagttgcg acgggtacgt cgttaaaaga atagctatca gtccaggcct gtatgggaag 960

ccttcaggct atgctgctac gatgcaccgc gagggattct tgtgctgcaa agtgacagac 1020

acattgaacg gggagagggt ctcttttccc gtgtgcacgt atgtgccagc tacattgtgt 1080

gaccaaatga ctggcatact ggcaacagat gtcagtgcgg acgacgcgca aaaactgctg 1140

gttgggctca accagcgtat agtcgtcaac ggtcgcaccc agagaaacac caataccatg 1200

aaaaattacc ttttgcccgt agtggcccag gcatttgcta ggtgggcaaa ggaatataag 1260

gaagatcaag aagatgaaag gccactagga ctacgagata gacagttagt catggggtgt 1320

tgttgggctt ttagaaggca caagataaca tctatttata agcgcccgga tacccaaacc 1380

atcatcaaag tgaacagcga tttccactca ttcgtgctgc ccaggatagg cagtaacaca 1440

ttggagatcg ggctgagaac aagaatcagg aaaatgttag aggagcacaa ggagccgtca 1500

[0081] cctctcatta ccgccgagga cgtacaagaa gctaagtgcg cagccgatga ggctaaggag 1560

gtgcgtgaag ccgaggagtt gcgcgcagct ctaccacctt tggcagctga tgttgaggag 1620

cccactctgg aagccgatgt cgacttgatg ttacaagagg ctggggccgg ctcagtggag 1680

acacctcgtg gcttgataaa ggttaccagc tacgctggcg aggacaagat cggctcttac 1740

gctgtgcttt ctccgcaggc tgtactcaag agtgaaaaat tatcttgcat ccaccctctc 1800

gctgaacaag tcatagtgat aacacactct ggccgaaaag ggcgttatgc cgtggaacca 1860

taccatggta aagtagtggt gccagaggga catgcaatac ccgtccagga ctttcaagct 1920

ctgagtgaaa gtgccaccat tgtgtacaac gaacgtgagt tcgtaaacag gtacctgcac 1980

catattgcca cacatggagg agcgctgaac actgatgaag aatattacaa aactgtcaag 2040

cccagcgagc acgacggcga atacctgtac gacatcgaca ggaaacagtg cgtcaagaaa 2100

gaactagtca ctgggctagg gctcacaggc gagctggtgg atcctccctt ccatgaattc 2160

gcctacgaga gtctgagaac acgaccagcc gctccttacc aagtaccaac catggggtg 2220

tatggcgtgc caggatcagg caagtctggc atcattaaaa gcgcagtcac caaaaaagat 2280

ctagtggtga gcgccaagaa agaaaactgt gcagaaatta taagggacgt caagaaaatg 2340

aaagggctgg acgtcaatgc cagaactgtg gactcagtgc tcttgaatgg atgcaaacac 2400

cccgtagaga ccctgtatat tgacgaagct tttgcttgtc atgcaggtac tctcagagcg 2460

ctcatagcca ttataagacc taaaaaggca gtgctctgcg gggatcccaa acagtgcggt 2520

ttttttaaca tgatgtgcct gaaagtgcat tttaaccacg agatttgcac acaagtcttc 2580

cacaaaagca tctctcgccg ttgcactaaa tctgtgactt cggtcgtctc aaccttgttt 2640

tacgacaaaa aaatgagaac gacgaatccg aaagagacta agattgtgat tgacactacc 2700

ggcagtacca aacctaagca ggacgatctc attctcactt gtttcagagg gtgggtgaag 2760

cagttgcaaa tagattacaa aggcaacgaa ataatgacgg cagctgcctc tcaagggctg 2820

acccgtaaag gtgtgtatgc cgttcggtac aaggtgaatg aaaatcctct gtacgcaccc 2880

acctcagaac atgtgaacgt cctactgacc cgcacggagg accgcatcgt gtggaaaaca 2940

ctagccggcg acccatggat aaaaacactg actgccaagt accctgggaa tttcactgcc 3000

[0082] acgatagagg agtggcaagc agagcatgat gccatcatga ggcacatctt ggagagaccg 3060

gaccctaccg acgtcttcca gaataaggca aacgtgtgtt gggccaaggc tttagtgccg 3120

gtgctgaaga ccgctggcat agacatgacc actgaacaat ggaacactgt ggattatttt 3180

gaaacggaca aagctcactc agcagagata gtattgaacc aactatgcgt gaggttcttt 3240

ggactcgatc tggactccgg tctattttct gcacccactg ttccgttatc cattaggaat 3300

aatcactggg ataactcccc gtcgcctaac atgtacgggc tgaataaaga agtggtccgt 3360

cagctctctc gcaggtaccc acaactgcct cgggcagttg ccactggaag agtctatgac 3420

atgaacactg gtacactgcg caattatgat ccgcgcataa acctagtacc tgtaaacaga 3480

agactgcctc atgctttagt cctccaccat aatgaacacc cacagagtga cttttcttca 3540

ttcgtcagca aattgaaggg cagaactgtc ctggtggtcg gggaaaagtt gtccgtccca 3600

ggcaaaatgg ttgactggtt gtcagaccgg cctgaggcta ccttcagagc tcggctggat 3660

ttaggcatcc caggtgatgt gcccaaatat gacataatat ttgttaatgt gaggacccca 3720

tataaatacc atcactatca gcagtgtgaa gaccatgcca ttaagcttag catgttgacc 3780

aagaaagctt gtctgcatct gaatcccggc ggaacctgtg tcagcatagg ttatggttac 3840

gctgacaggg ccagcgaaag catcattggt gctatagcgc ggcagttcaa gttttcccgg 3900

gtatgcaaac cgaaatcctc acttgaagag acggaagttc tgtttgtatt cattgggtac 3960

gatcgcaagg cccgtacgca caatccttac aagctttcat caaccttgac caacatttat 4020

acaggttcca gactccacga agccggatgt gcaccctcat atcatgtggt gcgaggggat 4080

attgccacgg ccaccgaagg agtgattata aatgctgcta acagcaaagg acaacctggc 4140

ggaggggtgt gcggagcgct gtataagaaa ttcccggaaa gcttcgattt acagccgatc 4200

gaagtaggaa aagcgcgact ggtcaaaggt gcagctaaac atatcattca tgccgtagga 4260

ccaaacttca acaaagtttc ggaggttgaa ggtgacaaac agttggcaga ggcttatgag 4320

tccatcgcta agattgtcaa cgataacaat tacaagtcag tagcgattcc actgttgtcc 4380

accggcatct tttccgggaa caaagatcga ctaacccaat cattgaacca tttgctgaca 4440

gctttagaca ccactgatgc agatgtagcc atatactgca gggacaagaa atgggaaatg 4500

[0083] actctcaagg aagcagtggc taggagagaa gcagtggagg agatatgcat atccgacgac 4560

tcttcagtga cagaacctga tgcagagctg gtgagggtgc atccgaagag ttctttggct 4620

ggaaggaagg gctacagcac aagcgatggc aaaactttct catatttgga agggaccaag 4680

tttcaccagg cggccaagga tatagcagaa attaatgcca tgtggcccgt tgcaacggag 4740

gccaatgagc aggtatgcat gtatatcctc ggagaaagca tgagcagtat taggtcgaaa 4800

tgccccgtcg aagagtcgga agcctccaca ccacctagca cgctgccttg cttgtgcatc 4860

catgccatga ctccagaaag agtacagcgc ctaaaagcct cacgtccaga acaaattact 4920

gtgtgctcat cctttccatt gccgaagtat agaatcactg gtgtgcagaa gatccaatgc 4980

tcccagccta tattgttctc accgaaagtg cctgcgtata ttcatccaag gaagtatctc 5040

gtggaaacac caccggtaga cgagactccg gagccatcgg cagagaacca atccacagag 5100

gggacacctg aacaaccacc acttataacc gaggatgaga ccaggactag aacgcctgag 5160

ccgatcatca tcgaagagga agaagaggat agcataagtt tgctgtcaga tggcccgacc 5220

caccaggtgc tgcaagtcga ggcagacatt cacgggccgc cctctgtatc tagctcatcc 5280

tggtccattc ctcatgcatc cgactttgat gtggacagtt tatccatact tgacaccctg 5340

gagggagcta gcgtgaccag cggggcaacg tcagccgaga ctaactctta cttcgcaaag 5400

agtatggagt ttctggcgcg accggtgcct gcgcctcgaa cagtattcag gaaccctcca 5460

catcccgctc cgcgcacaag aacaccgtca cttgcaccca gcagggcctg ctcgagaacc 5520

agcctagttt ccaccccgcc aggcgtgaat agggtgatca ctagagagga gctcgaggcg 5580

cttaccccgt cacgcactcc tagcaggtcg gtctcgagaa ccagcctggt ctccaacccg 5640

ccaggcgtaa atagggtgat tacaagagag gagtttgagg cgttcgtagc acaacaacaa 5700

tgacggtttg atgcgggtgc atacatcttt tcctccgaca ccggtcaagg gcatttacaa 5760

caaaaatcag taaggcaaac ggtgctatcc gaagtggtgt tggagaggac cgaattggag 5820

atttcgtatg ccccgcgcct cgaccaagaa aaagaagaat tactacgcaa gaaattacag 5880

ttaaatccca cacctgctaa cagaagcaga taccagtcca ggaaggtgga gaacatgaaa 5940

gccataacag ctagacgtat tctgcaaggc ctagggcatt atttgaaggc agaaggaaaa 6000

[0084] gtggagtgct accgaaccct gcatcctgtt cctttgtatt catctagtgt gaaccgtgcc 6060

ttttcaagcc ccaaggtcgc agtggaagcc tgtaacgcca tgttgaaaga gaactttccg 6120

actgtggctt cttactgtat tattccagag tacgatgcct atttggacat ggttgacgga 6180

gcttcatgct gcttagacac tgccagtttt tgccctgcaa agctgcgcag ctttccaaag 6240

aaacactcct atttggaacc cacaatacga tcggcagtgc cttcagcgat ccagaacacg 6300

ctccagaacg tcctggcagc tgccacaaaa agaaattgca atgtcacgca aatgagagaa 6360

ttgcccgtat tggattcggc ggcctttaat gtggaatgct tcaagaaata tgcgtgtaat 6420

aatgaatatt gggaaacgtt taaagaaaac cccatcaggc ttactgaaga aaacgtggta 6480

aattacatta ccaaattaaa aggaccaaaa gctgctgctc tttttgcgaa gacacataat 6540

ttgaatatgt tgcaggacat accaatggac aggtttgtaa tggacttaaa gagagacgtg 6600

aaagtgactc caggaacaaa acatactgaa gaacggccca aggtacaggt gatccaggct 6660

gccgatccgc tagcaacagc gtatctgtgc ggaatccacc gagagctggt taggagatta 6720

aatgcggtcc tgcttccgaa cattcataca ctgtttgata tgtcggctga agactttgac 6780

gctattatag ccgagcactt ccagcctggg gattgtgttc tggaaactga catcgcgtcg 6840

tttgataaaa gtgaggacga cgccatggct ctgaccgcgt taatgattct ggaagactta 6900

ggtgtggacg cagagctgtt gacgctgatt gaggcggctt tcggcgaaat ttcatcaata 6960

catttgccca ctaaaactaa atttaaattc ggagccatga tgaaatctgg aatgttcctc 7020

acactgtttg tgaacacagt cattaacatt gtaatcgcaa gcagagtgtt gagagaacgg 7080

ctaaccggat caccatgtgc agcattcatt ggagatgaca atatcgtgaa aggagtcaaa 7140

tcggacaaat taatggcaga caggtgcgcc acctggttga atatggaagt caagattata 7200

gatgctgtgg tgggcgagaa agcgccttat ttctgtggag ggtttatttt gtgtgactcc 7260

gtgaccggca cagcgtgccg tgtggcagac cccctaaaaa ggctgtttaa gcttggcaaa 7320

cctctggcag cagacgatga acatgatgat gacaggagaa gggcattgca tgaagagtca 7380

acacgctgga accgagtggg tattctttca gagctgtgca aggcagtaga atcaaggtat 7440

gaaaccgtag gaacttccat catagttatg gccatgacta ctctagctag cagtgttaaa 7500

[0085] tcattcagct acctgagagg ggcccctata actctctacg gctaacctga atggactacg 7560

actatcacgc ccaaacattt acagccgcgg tgtcaaaaac cgcgtggacg tggttaacat 7620

ccctgctggg aggatcagcc gtaattatta taattggctt ggtgctggct actattgtgg 7680

ccatgtacgt gctgaccaac cagaaacata attgaataca gcagcaattg gcaagctgct 7740

tacatagaac tcgcggcgat tggcatgccg ccttaaaatt tttattttat tttttctttt 7800

cttttccgaa tcggattttg tttttaatat ttcaaaaaaa aaaaaaaaaa aaaaaaaaaa 7860

aaaaaaaaaa aaaaaaaaaa aaaaaaaaaa aaaaaaaaaa aaaaaaaaaa aaatacgtag 7920

tttaaac 7927

<210> 9
<211> 7926
<212> DNA
<213> 人工序列(Artificial Sequence)

<220>
<223> 人工序列的描述：合成多核苷酸

<400> 9

[0086]

taatacgact cactatagga taggcggcgc atgagagaag cccagaccaa ttacctaccc 60

aaaatggaga aagttcacgt tgacatcgag gaagacagcc cattcctcag agctttgcag 120

cggagcttcc cgcagtttga ggtagaagcc aagcaggtca ctgataatga ccatgctaat 180

gccagagcgt tttcgcatct ggcttcaaaa ctgatcgaaa cggaggtgga cccatccgac 240

acgatccttg acattggaag tgcgcccgcc cgcagaatgt attctaagca caagtatcat 300

tgtatctgtc cgatgagatg tgcggaagat ccggacagat tgtataagta tgcaactaag 360

ctgaagaaaa actgtaagga aataactgat aaggaattgg acaagaaaat gaaggagctc 420

gccgccgtca tgagcgaccc tgacctggaa actgagacta tgtgcctcca cgacgacgag 480

tcgtgtcgct acgaagggca agtcgctgtt taccaggatg tatacgcggt tgacggaccg 540

acaagtctct atcaccaagc caataaggga gttagagtcg cctactggat aggctttgac 600

accacccctt ttatgtttaa gaacttggct ggagcatatc catcatactc taccaactgg 660

gccgacgaaa ccgtgttaac ggctcgtaac ataggcctat gcagctctga cgttatggag 720

cggtcacgta gagggatgtc cattcttaga aagaagtatt tgaaaccatc caacaatgtt 780

ctattctctg ttggctcgac catctaccac gagaagaggg acttactgag gagctggcac 840

ctgccgtctg tatttcactt acgtggcaag caaaattaca catgtcggtg tgagactata 900

gttagttgcg acgggtacgt cgttaaaaga atagctatca gtccaggcct gtatgggaag 960

ccttcaggct atgctgctac gatgcaccgc gagggattct tgtgctgcaa agtgacagac 1020

acattgaacg gggagagggt ctcttttccc gtgtgcacgt atgtgccagc tacattgtgt 1080

gaccaaatga ctggcatact ggcaacagat gtcagtgcgg acgacgcgca aaaactgctg 1140

gttgggctca accagcgtat agtcgtcaac ggtcgcaccc agagaaacac caataccatg 1200

aaaaattacc ttttgcccgt agtggcccag gcatttgcta ggtgggcaaa ggaatataag 1260

gaagatcaag aagatgaaag gccactagga ctacgagata gacagttagt catggggtgt 1320

tgttgggctt ttagaaggca caagataaca tctatttata agcgcccgga tacccaaacc 1380

atcatcaaag tgaacagcga tttccactca ttcgtgctgc ccaggatagg cagtaacaca 1440

ttggagatcg ggctgagaac aagaatcagg aaaatgttag aggagcacaa ggagccgtca 1500

cctctcatta ccgccgagga cgtacaagaa gctaagtgcg cagccgatga ggctaaggag 1560

gtgcgtgaag ccgaggagtt gcgcgcagct ctaccacctt tggcagctga tgttgaggag 1620

cccactctgg aagccgatgt cgacttgatg ttacaagagg ctggggccgg ctcagtggag 1680

acacctcgtg gcttgataaa ggttaccagc tacgatggcg aggacaagat cggctcttac 1740

gctgtgcttt ctccgcaggc tgtactcaag agtgaaaaat tatcttgcat ccaccctctc 1800

gctgaacaag tcatagtgat aacacactct ggccgaaaag ggcgttatgc cgtggaacca 1860

taccatggta aagtagtggt gccagaggga catgcaatac ccgtccagga ctttcaagct 1920

ctgagtgaaa gtgccaccat tgtgtacaac gaacgtgagt tcgtaaacag gtacctgcac 1980

catattgcca cacatggagg agcgctgaac actgatgaag aatattacaa aactgtcaag 2040

cccagcgagc acgacggcga atacctgtac gacatcgaca ggaaacagtg cgtcaagaaa 2100

gaactagtca ctgggctagg gctcacaggc gagctggtgg atcctccctt ccatgaattc 2160

gcctacgaga gtctgagaac acgaccagcc gctccttacc aagtaccaac cataggggtg 2220

[0087] tatggcgtgc caggatcagg caagtctggc atcattaaaa gcgcagtcac caaaaaagat 2280

ctagtggtga gcgccaagaa agaaaactgt gcagaaatta taagggacgt caagaaaatg 2340

aaagggctgg acgtcaatgc cagaactgtg gactcagtgc tcttgaatgg atgcaaacac 2400

cccgtagaga ccctgtatat tgacgaagct tttgcttgtc atgcaggtac tctcagagcg 2460

ctcatagcca ttataagacc taaaaaggca gtgctctgcg gggatcccaa acagtgcggt 2520

ttttttaaca tgatgtgcct gaaagtgcat tttaaccacg agatttgcac acaagtcttc 2580

cacaaaagca tctctcgccg ttgcactaaa tctgtgactt cggtcgtctc aaccttgttt 2640

tacgacaaaa aaatgagaac gacgaatccg aaagagacta agattgtgat tgacactacc 2700

ggcagtacca aacctaagca ggacgatctc attctcactt gtttcagagg gtgggtgaag 2760

cagttgcaaa tagattacaa aggcaacgaa ataatgacgg cagctgcctc tcaagggctg 2820

acccgtaaag gtgtgtatgc cgttcggtac aaggtgaatg aaaatcctct gtacgcaccc 2880

acctcagaac atgtgaacgt cctactgacc cgcacggagg accgcatcgt gtggaaaaca 2940

ctagccggcg acccatggat aaaaacactg actgccaagt accctgggaa tttcactgcc 3000

acgatagagg agtggcaagc agagcatgat gccatcatga ggcacatctt ggagagaccg 3060

gaccctaccg acgtcttcca gaataaggca aacgtgtgtt gggccaaggc tttagtgccg 3120

gtgctgaaga ccgctggcat agacatgacc actgaacaat ggaacactgt ggattatttt 3180

gaaacggaca aagctcactc agcagagata gtattgaacc aactatgcgt gaggttcttt 3240

ggactcgatc tggactccgg tctattttct gcacccactg ttccgttatc cattaggaat 3300

aatcactggg ataactcccc gtcgcctaac atgtacgggc tgaataaaga agtggtccgt 3360

cagctctctc gcaggtaccc acaactgcct cgggcagttg ccactggaag agtctatgac 3420

atgaacactg gtacactgcg caattatgat ccgcgcataa acctagtacc tgtaaacaga 3480

agactgcctc atgctttagt cctccaccat aatgaacacc cacagagtga cttttcttca 3540

ttcgtcagca aattgaaggg cagaactgtc ctggtggtcg gggaaaagtt gtccgtccca 3600

ggcaaaatgg ttgactggtt gtcagaccgg cctgaggcta ccttcagagc tcggctggat 3660

ttaggcatcc caggtgatgt gcccaaatat gacataatat ttgttaatgt gaggacccca 3720

[0088] tataaatacc atcactatca gcagtgtgaa gaccatgcca ttaagcttag catgttgacc 3780

aagaaagctt gtctgcatct gaatcccggc ggaacctgtg tcagcatagg ttatggttac 3840

gctgacaggg ccagcgaaag catcattggt gctatagcgc ggcagttcaa gttttcccgg 3900

gtatgcaaac cgaaatcctc acttgaagag acggaagttc tgtttgtatt cattgggtac 3960

gatcgcaagg cccgtacgca caatccttac aagctttcat caaccttgac caacatttat 4020

acaggttcca gactccacga agccggatgt gcaccctcat atcatgtggt gcgaggggat 4080

attgccacgg ccaccgaagg agtgattata aatgctgcta acagcaaagg acaacctggc 4140

ggaggggtgt gcggagcgct gtataagaaa ttcccggaaa gcttcgattt acagccgatc 4200

gaagtaggaa aagcgcgact ggtcaaaggt gcagctaaac atatcattca tgccgtagga 4260

ccaaacttca acaaagtttc ggaggttgaa ggtgacaaac agttggcaga ggcttatgag 4320

tccatcgcta agattgtcaa cgataacaat tacaagtcag tagcgattcc actgttgtcc 4380

accggcatct tttccgggaa caaagatcga ctaacccaat cattgaacca tttgctgaca 4440

gctttagaca ccactgatgc agatgtagcc atatactgca gggacaagaa atgggaaatg 4500

actctcaagg aagcagtggc taggagagaa gcagtggagg agatatgcat atccgacgac 4560

tcttcagtga cagaacctga tgcagagctg gtgagggtgc atccgaagag ttctttggct 4620

ggaaggaagg gctacagcac aagcgatggc aaaactttct catatttgga agggaccaag 4680

tttcaccagg cggccaagga tatagcagaa attaatgcca tgtggcccgt tgcaacggag 4740

gccaatgagc aggtatgcat gtatatcctc ggagaaagca tgagcagtat taggtcgaaa 4800

tgccccgtcg aagagtcgga agcctccaca ccacctagca cgctgccttg cttgtgcatc 4860

catgccatga ctccagaaag agtacagcgc ctaaaagcct cacgtccaga acaaattact 4920

gtgtgctcat cctttccatt gccgaagtat agaatcactg gtgtgcagaa gatccaatgc 4980

tcccagccta tattgttctc accgaaagtg cctgcgtata ttcatccaag gaagtatctc 5040

gtggaaacac caccggtaga cgagactccg gagccatcgg cagagaacca atccacagag 5100

gggacacctg aacaaccacc acttataacc gaggatgaga ccaggactag aacgcctgag 5160

ccgatcatca tcgaagagga agaagaggat agcataagtt tgctgtcaga tggcccgacc 5220

[0089] caccaggtgc tgcaagtcga ggcagacatt cacgggccgc cctctgtatc tagctcatcc 5280

tggtccattc ctcatgcatc cgactttgat gtggacagtt tatccatact tgacaccctg 5340

gagggagcta gcgtgaccag cggggcaacg tcagccgaga ctaactctta cttcgcaaag 5400

agtatggagt ttctggcgcg accggtgcct gcgcctcgaa cagtattcag gaaccctcca 5460

catcccgctc cgcgcacaag aacaccgtca cttgcaccca gcagggcctg ctcgagaacc 5520

agcctagttt ccaccccgcc aggcgtgaat agggtgatca ctagagagga gctcgaggcg 5580

cttaccccgt cacgcactcc tagcaggtcg gtctcgagaa ccagcctggt ctccaacccg 5640

ccaggcgtaa atagggtgat tacaagagag gagtttgagg cgttcgtagc acaacaacaa 5700

tgacggtttg atgcgggtgc atacatcttt tcctccgaca ccggtcaagg gcatttacaa 5760

caaaaatcag taaggcaaac ggtgctatcc gaagtggtgt tggagaggac cgaattggag 5820

atttcgtatg ccccgcgcct cgaccaagaa aaagaagaat tactacgcaa gaaattacag 5880

ttaaatccca cacctgctaa cagaagcaga taccagtcca ggaaggtgga gaacatgaaa 5940

gccataacag ctagacgtat tctgcaaggc ctagggcatt atttgaaggc agaaggaaaa 6000

gtggagtgct accgaaccct gcatcctgtt cctttgtatt catctagtgt gaaccgtgcc 6060

ttttcaagcc ccaaggtcgc agtggaagcc tgtaacgcca tgttgaaaga gaactttccg 6120

actgtggctt cttactgtat tattccagag tacgatgcct atttggacat ggttgacgga 6180

gcttcatgct gcttagacac tgccagtttt tgccctgcaa agctgcgcag ctttccaaag 6240

aaacactcct atttggaacc cacaatacga tcggcagtgc cttcagcgat ccagaacacg 6300

ctccagaacg tcctggcagc tgccacaaaa agaaattgca atgtcacgca aatgagagaa 6360

ttgcccgtat tggattcggc ggcctttaat gtggaatgct tcaagaaata tgcgtgtaat 6420

aatgaatatt gggaaacgtt taaagaaaac cccatcaggc ttactgaaga aaacgtggta 6480

aattacatta ccaaattaaa aggaccaaaa gctgctgctc tttttgcgaa gacacataat 6540

ttgaatatgt tgcaggacat accaatggac aggtttgtaa tggacttaaa gagagacgtg 6600

aaagtgactc caggaacaaa acatactgaa gaacggccca aggtacaggt gatccaggct 6660

gccgatccgc tagcaacagc gtatctgtgc ggaatccacc gagagctggt taggagatta 6720

[0090] aatgcggtcc tgcttccgaa cattcataca ctgtttgata tgtcggctga agactttgac 6780

gctattatag ccgagcactt ccagcctggg gattgtgttc tggaaactga catcgcgtcg 6840

tttgataaaa gtgaggacga cgccatggct ctgaccgcgt taatgattct ggaagactta 6900

ggtgtggacg cagagctgtt gacgctgatt gaggcggctt tcggcgaaat ttcatcaata 6960

catttgccca ctaaaactaa atttaaattc ggagccatga tgaaatctgg aatgttcctc 7020

acactgtttg tgaacacagt cattaacatt gtaatcgcaa gcagagtgtt gagagaacgg 7080

ctaaccggat caccatgtgc agcattcatt ggagatgaca atatcgtgaa aggagtcaaa 7140

tcggacaaat taatggcaga caggtgcgcc acctggttga atatggaagt caagattata 7200

gatgctgtgg tgggcgagaa agcgccttat ttctgtggag ggtttatttt gtgtgactcc 7260

gtgaccggca cagcgtgccg tgtggcagac cccctaaaaa ggctgtttaa gcttggcaaa 7320

cctctggcag cagacgatga acatgatgat gacaggagaa gggcattgca tgaagagtca 7380

acacgctgga accgagtggg tattctttca gagctgtgca aggcagtaga atcaaggtat 7440

gaaaccgtag gaacttccat catagttatg gccatgacta ctctagctag cagtgttaaa 7500

tcattcagct acctgagagg ggcccctata actctctacg gctaacctga atggactacg 7560

actatcacgc ccaaacattt acagccgcgg tgtcaaaaac cgcgtggacg tggttaacat 7620

ccctgctggg aggatcagcc gtaattatta taattggctt ggtgctggct actattgtgg 7680

ccatgtacgt gctgaccaac cagaaacata attgaataca gcagcaattg gcaagctgct 7740

tacatagaac tcgcggcgat tggcatgccg ccttaaaatt tttattttat ttttcttttc 7800

ttttccgaat cggattttgt ttttaatatt tcaaaaaaaa aaaaaaaaaa aaaaaaaaaa 7860

aaaaaaaaaa aaaaaaaaaa aaaaaaaaaa aaaaaaaaaa aaaaaaaaaa aatacgtagt 7920

ttaaac 7926

<210> 10
<211> 36519
<212> DNA
<213> 人工序列(Artificial Sequence)

<220>
<223> 人工序列的描述：合成多核苷酸

[0091]

<400> 10
ccatcttcaa taatatacct caaacttttt gtgcgcgtta atatgcaaat gaggcgtttg 60

aatttgggga ggaagggcgg tgattggtcg agggatgagc gaccgttagg ggcggggcga 120

gtgacgtttt gatgacgtgg ttgcgaggag gagccagttt gcaagttctc gtgggaaaag 180

tgacgtcaaa cgaggtgtgg tttgaacacg gaaatactca attttcccgc gctctctgac 240

aggaaatgag gtgtttctgg gcggatgcaa gtgaaaacgg gccattttcg cgcgaaaact 300

gaatgaggaa gtgaaaatct gagtaatttc gcgtttatgg cagggaggag tatttgccga 360

gggccgagta gactttgacc gattacgtgg gggtttcgat taccgtgttt ttcacctaaa 420

tttccgcgta cggtgtcaaa gtccggtgtt tttacgtagg tgtcagctga tcgccagggt 480

atttaaacct gcgctctcca gtcaagaggc cactcttgag tgccagcgag aagagttttc 540

tcctccgcgc cgcgagtcag atctacactt tgaaagatga ggcacctgag agacctgccc 600

gatgagaaaa tcatcatcgc ttccgggaac gagattctgg aactggtggt aaatgccatg 660

atgggcgacg accctccgga gccccccacc ccatttgaga caccttcgct gcacgatttg 720

tatgatctgg aggtggatgt gcccgaggac gatcccaatg aggaggcggt aaatgatttt 780

tttagcgatg ccgcgctgct agctgccgag gaggcttcga gctctagctc agacagcgac 840

tcttcactgc ataccctag acccggcaga ggtgagaaaa agatccccga gcttaaaggg 900

gaagagatgg acttgcgctg ctatgaggaa tgcttgcccc cgagcgatga tgaggacgag 960

caggcgatcc agaacgcagc gagccaggga gtgcaagccg ccagcgagag ctttgcgctg 1020

gactgcccgc ctctgcccgg acacggctgt aagtcttgtg aatttcatcg catgaatact 1080

ggagataaag ctgtgttgtg tgcactttgc tatatgagag cttacaacca ttgtgtttac 1140

agtaagtgtg attaagttga actttagagg gaggcagaga gcagggtgac tgggcgatga 1200

ctggtttatt tatgtatata tgttctttat ataggtcccg tctctgacgc agatgatgag 1260

acccccacta caaagtccac ttcgtcaccc ccagaaattg gcacatctcc acctgagaat 1320

attgttagac cagttcctgt tagagccact gggaggagag cagctgtgga atgtttggat 1380

gacttgctac agggtggggt tgaacctttg gacttgtgta cccggaaacg ccccaggcac 1440

[0092] taagtgccac acatgtgtgt ttacttgagg tgatgtcagt atttataggg tgtggagtgc 1500

aataaaaaat gtgttgactt taagtgcgtg gtttatgact caggggtggg gactgtgagt 1560

atataagcag gtgcagacct gtgtggttag ctcagagcgg catggagatt tggacggtct 1620

tggaagactt tcacaagact agacagctgc tagagaacgc ctcgaacgga gtctcttacc 1680

tgtggagatt ctgcttcggt ggcgacctag ctaggctagt ctacagggcc aaacaggatt 1740

atagtgaaca atttgaggtt attttgagag agtgttctgg tctttttgac gctcttaact 1800

tgggccatca gtctcacttt aaccagagga tttcgagagc ccttgatttt actactcctg 1860

gcagaaccac tgcagcagta gccttttttg cttttattct tgacaaatgg agtcaagaaa 1920

cccatttcag cagggattac cagctggatt tcttagcagt agctttgtgg agaacatgga 1980

agtgccagcg cctgaatgca atctccggct acttgccggt acagccgcta gacactctga 2040

ggatcctgaa tctccaggag agtcccaggg cacgccaacg tcgccagcag cagcagcagg 2100

aggaggatca agaagagaac ccgagagccg gcctggaccc tccggcggag gaggaggagt 2160

agctgacctg tttcctgaac tgcgccgggt gctgactagg tcttcgagtg gtcgggagag 2220

ggggattaag cgggagaggc atgatgagac taatcacaga actgaactga ctgtgggtct 2280

gatgagtcgc aagcgcccag aaacagtgtg gtggcatgag gtgcagtcga ctggcacaga 2340

tgaggtgtcg gtgatgcatg agaggttttc tctagaacaa gtcaagactt gttggttaga 2400

gcctgaggat gattgggagg tagccatcag gaattatgcc aagctggctc tgaggccaga 2460

caagaagtac aagattacta agctgataaa tatcagaaat gcctgctaca tctcagggaa 2520

tggggctgaa gtggagatct gtctccagga aagggtggct ttcagatgct gcatgatgaa 2580

tatgtacccg ggagtggtgg gcatggatgg ggttaccttt atgaacatga ggttcagggg 2640

agatgggtat aatggcacgg tctttatggc caataccaag ctgacagtcc atggctgctc 2700

cttctttggg tttaataaca cctgcatcga ggcctggggt caggtcggtg tgaggggctg 2760

cagtttttca gccaactgga tgggggtcgt gggcaggacc aagagtatgc tgtccgtgaa 2820

gaaatgcttg tttgagaggt gccacctggg ggtgatgagc gagggcgaag ccagaatccg 2880

ccactgcgcc tctaccgaga cgggctgctt tgtgctgtgc aagggcaatg ctaagatcaa 2940

[0093] gcataatatg atctgtggag cctcggacga gcgcggctac cagatgctga cctgcgccgg 3000

cgggaacagc catatgctgg ccaccgtaca tgtggcttcc catgctcgca agccctggcc 3060

cgagttcgag cacaatgtca tgaccaggtg caatatgcat ctggggtccc gccgaggcat 3120

gttcatgccc taccagtgca acctgaatta tgtgaaggtg ctgctggagc ccgatgccat 3180

gtccagagtg agcctgacgg gggtgtttga catgaatgtg gaggtgtgga agattctgag 3240

atatgatgaa tccaagacca ggtgccgagc ctgcgagtgc ggagggaagc atgccaggtt 3300

ccagcccgtg tgtgtggatg tgacggagga cctgcgaccc gatcatttgg tgttgccctg 3360

caccgggacg gagttcggtt ccagcgggga agaatctgac tagagtgagt agtgttctgg 3420

ggcgggggag gacctgcatg agggccagaa taactgaaat ctgtgctttt ctgtgtgttg 3480

cagcagcatg agcggaagcg gctcctttga gggaggggta ttcagccctt atctgacggg 3540

gcgtctcccc tcctgggcgg gagtgcgtca gaatgtgatg ggatccacgg tggacggccg 3600

gcccgtgcag cccgcgaact cttcaaccct gacctatgca accctgagct cttcgtcgtt 3660

ggacgcagct gccgccgcag ctgctgcatc tgccgccagc gccgtgcgcg gaatggccat 3720

gggcgccggc tactacggca ctctggtggc caactcgagt tccaccaata atcccgccag 3780

cctgaacgag gagaagctgt tgctgctgat ggcccagctc gaggccttga cccagcgcct 3840

gggcgagctg acccagcagg tggctcagct gcaggagcag acgcgggccg cggttgccac 3900

ggtgaaatcc aaataaaaaa tgaatcaata aataaacgga gacggttgtt gattttaaca 3960

cagagtctga atctttattt gatttttcgc gcgcggtagg ccctggacca ccggtctcga 4020

tcattgagca cccggtggat cttttccagg acccggtaga ggtgggcttg gatgttgagg 4080

tacatgggca tgagcccgtc ccgggggtgg aggtagctcc attgcagggc ctcgtgctcg 4140

ggggtggtgt tgtaaatcac ccagtcatag caggggcgca gggcatggtg ttgcacaata 4200

tctttgagga ggagactgat ggccacgggc agcccttttgg tgtaggtgtt tacaaatctg 4260

ttgagctggg agggatgcat gcgggggggag atgaggtgca tcttggcctg gatcttgaga 4320

ttggcgatgt taccgcccag atcccgcctg gggttcatgt tgtgcaggac caccagcacg 4380

gtgtatccgg tgcacttggg gaatttatca tgcaacttgg aagggaaggc gtgaaagaat 4440

[0094] ttggcgacgc ctttgtgccc gcccaggttt tccatgcact catccatgat gatggcgatg 4500

ggcccgtggg cggcggcctg ggcaaagacg tttcgggggt cggacacatc atagttgtgg 4560

tcctgggtga ggtcatcata ggccatttta atgaatttgg ggcggagggt gccggactgg 4620

gggacaaagg taccctcgat cccgggggcg tagttcccct cacagatctg catctcccag 4680

gctttgagct cggaggggggg gatcatgtcc acctgcgggg cgataaagaa cacggtttcc 4740

ggggcggggg agatgagctg ggccgaaagc aagttccgga gcagctggga cttgccgcag 4800

ccggtggggc cgtagatgac cccgatgacc ggctgcaggt ggtagttgag ggagagacag 4860

ctgccgtcct cccggaggag gggggccacc tcgttcatca tctcgcgcac gtgcatgttc 4920

tcgcgcacca gttccgccag gaggcgctct ccccccaggg ataggagctc ctggagcgag 4980

gcgaagtttt tcagcggctt gagtccgtcg gccatgggca ttttggagag ggtttgttgc 5040

aagagttcca ggcggtccca gagctcggtg atgtgctcta cggcatctcg atccagcaga 5100

cctcctcgtt tcgcgggttg ggacggctgc gggagtaggg caccagacga tgggcgtcca 5160

gcgcagccag ggtccggtcc ttccagggtc gcagcgtccg cgtcagggtg gtctccgtca 5220

cggtgaaggg gtgcgcgccg ggctgggcgc ttgcgagggt gcgcttcagg ctcatccggc 5280

tggtcgaaaa ccgctcccga tcggcgccct gcgcgtcggc caggtagcaa ttgaccatga 5340

gttcgtagtt gagcgcctcg gccgcgtggc ctttggcgcg gagcttacct ttggaagtct 5400

gcccgcaggc gggacagagg agggacttga gggcgtagag cttgggggcg aggaagacgg 5460

actcgggggc gtaggcgtcc gcgccgcagt gggcgcagac ggtctcgcac tccacgagcc 5520

aggtgaggtc gggctggtcg gggtcaaaaa ccagtttccc gccgttcttt ttgatgcgtt 5580

tcttaccttt ggtctccatg agctcgtgtc cccgctgggt gacaaagagg ctgtccgtgt 5640

ccccgtagac cgactttatg ggccggtcct cgagcggtgt gccgcggtcc tcctcgtaga 5700

ggaaccccgc ccactccgag acgaaagccc gggtccaggc cagcacgaag gaggccacgt 5760

gggacgggta gcggtcgttg tccaccagcg ggtccacctt ttccagggta tgcaaacaca 5820

tgtccccctc gtccacatcc aggaaggtga ttggcttgta agtgtaggcc acgtgaccgg 5880

gggtcccggc cgggggggta taaaagggtg cgggtccctg ctcgtcctca ctgtcttccg 5940

[0095] gatcgctgtc caggagcgcc agctgttggg gtaggtattc cctctcgaag gcgggcatga 6000

cctcggcact caggttgtca gtttctagaa acgaggagga tttgatattg acggtgccgg 6060

cggagatgcc tttcaagagc ccctcgtcca tctggtcaga aaagacgatc tttttgttgt 6120

cgagcttggt ggcgaaggag ccgtagaggg cgttggagag gagcttggcg atggagcgca 6180

tggtctggtt tttttccttg tcggcgcgct ccttggcggc gatgttgagc tgcacgtact 6240

cgcgcgccac gcacttccat tcggggaaga cggtggtcag ctcgtcgggc acgattctga 6300

cctgccagcc ccgattatgc agggtgatga ggtccacact ggtggccacc tcgccgcgca 6360

ggggctcatt agtccagcag aggcgtccgc ccttgcgcga gcagaagggg ggcagggggt 6420

ccagcatgac ctcgtcgggg gggtcggcat cgatggtgaa gatgccgggc aggaggtcgg 6480

ggtcaaagta gctgatggaa gtggccagat cgtccagggc agcttgccat tcgcgcacgg 6540

ccagcgcgcg ctcgtaggga ctgaggggcg tgccccaggg catgggatgg gtaagcgcgg 6600

aggcgtacat gccgcagatg tcgtagacgt agaggggctc ctcgaggatg ccgatgtagg 6660

tggggtagca gcgccccccg cggatgctgg cgcgcacgta gtcatacagc tcgtgcgagg 6720

gggcgaggag ccccgggccc aggttggtgc gactgggctt ttcggcgcgg tagacgatct 6780

ggcggaaaat ggcatgcgag ttggaggaga tggtgggcct ttggaagatg ttgaagtggg 6840

cgtggggcag tccgaccgag tcgcggatga agtgggcgta ggagtcttgc agcttggcga 6900

cgagctcggc ggtgactagg acgtccagag cgcagtagtc gagggtctcc tggatgatgt 6960

catacttgag ctgtcccttt tgtttccaca gctcgcggtt gagaaggaac tcttcgcggt 7020

ccttccagta ctcttcgagg gggaacccgt cctgatctgc acggtaagag cctagcatgt 7080

agaactggtt gacggccttg taggcgcagc agcccttctc cacggggagg gcgtaggcct 7140

gggcggcctt gcgcagggag gtgtgcgtga gggcgaaagt gtccctgacc atgaccttga 7200

ggaactggtg cttgaagtcg atatcgtcgc agccccccctg ctcccagagc tggaagtccg 7260

tgcgcttctt gtaggcgggg ttgggcaaag cgaaagtaac atcgttgaag aggatcttgc 7320

ccgcgcgggg cataaagttg cgagtgatgc ggaaaggttg gggcacctcg gcccggttgt 7380

tgatgacctg ggcggcgagc acgatctcgt cgaagccgtt gatgttgtgg cccacgatgt 7440

[0096] agagttccac gaatcgcgga cggcccttga cgtggggcag tttcttgagc tcctcgtagg 7500

tgagctcgtc ggggtcgctg agcccgtgct gctcgagcgc ccagtcggcg agatgggggt 7560

tggcgcggag gaaggaagtc cagagatcca cggccagggc ggtttgcaga cggtcccggt 7620

actgacggaa ctgctgcccg acggccattt tttcgggggt gacgcagtag aaggtgcggg 7680

ggtccccgtg ccagcgatcc catttgagct ggagggcgag atcgagggcg agctcgacga 7740

gccggtcgtc cccggagagt ttcatgacca gcatgaaggg gacgagctgc ttgccgaagg 7800

accccatcca ggtgtaggtt tccacatcgt aggtgaggaa gagcctttcg gtgcgaggat 7860

gcgagccgat ggggaagaac tggatctcct gccaccaatt ggaggaatgg ctgttgatgt 7920

gatggaagta gaaatgccga cggcgcgccg aacactcgtg cttgtgttta tacaagcggc 7980

cacagtgctc gcaacgctgc acgggatgca cgtgctgcac gagctgtacc tgagttcctt 8040

tgacgaggaa tttcagtggg aagtggagtc gtggcgcctg catctcgtgc tgtactacgt 8100

cgtggtggtc ggcctggccc tcttctgcct cgatggtggt catgctgacg agcccgcgcg 8160

ggaggcaggt ccagacctcg gcgcgagcgg gtcggagagc gaggacgagg gcgcgcaggc 8220

cggagctgtc cagggtcctg agacgctgcg gagtcaggtc agtgggcagc ggcggcgcgc 8280

ggttgacttg caggagtttt tccagggcgc gcgggaggtc cagatggtac ttgatctcca 8340

ccgcgccatt ggtggcgacg tcgatggctt gcagggtccc gtgcccctgg ggtgtgacca 8400

ccgtcccccg tttcttcttg ggcggctggg gcgacggggg cggtgcctct tccatggtta 8460

gaagcggcgg cgaggacgcg cgccgggcgg caggggcggc tcggggcccg gaggcagggg 8520

cggcaggggc acgtcggcgc cgcgcgcggg taggttctgg tactgcgccc ggagaagact 8580

ggcgtgagcg acgacgcgac ggttgacgtc ctggatctga cgcctctggg tgaaggccac 8640

gggacccgtg agtttgaacc tgaaagagag ttcgacagaa tcaatctcgg tatcgttgac 8700

ggcggcctgc cgcaggatct cttgcacgtc gcccgagttg tcctggtagg cgatctcggt 8760

catgaactgc tcgatctcct cctcttgaag gtctccgcgg ccggcgcgct ccacggtggc 8820

cgcgaggtcg ttggagatgc ggcccatgag ctgcgagaag gcgttcatgc ccgcctcgtt 8880

ccagacgcgg ctgtagacca cgacgccctc gggatcgcgg gcgcgcatga ccacctgggc 8940

[0097] gaggttgagc tccacgtggc gcgtgaagac cgcgtagttg cagaggcgct ggtagaggta 9000

gttgagcgtg gtggcgatgt gctcggtgac gaagaaatac atgatccagc ggcggagcgg 9060

catctcgctg acgtcgccca gcgcctccaa acgttccatg gcctcgtaaa agtccacggc 9120

gaagttgaaa aactgggagt tgcgcgccga gacggtcaac tcctcctcca gaagacggat 9180

gagctcggcg atggtggcgc gcacctcgcg ctcgaaggcc cccgggagtt cctccacttc 9240

ctcttcttcc tcctccacta acatctcttc tacttcctcc tcaggcggca gtggtggcgg 9300

gggaggggcc ctgcgtcgcc ggcggcgcac gggcagacgg tcgatgaagc gctcgatggt 9360

ctcgccgcgc cggcgtcgca tggtctcggt gacggcgcgc ccgtcctcgc ggggccgcag 9420

cgtgaagacg ccgccgcgca tctccaggtg gccgggcggg tccccgttgg gcagggagag 9480

ggcgctgacg atgcatctta tcaattgccc cgtagggact ccgcgcaagg acctgagcgt 9540

ctcgagatcc acgggatctg aaaaccgctg aacgaaggct tcgagccagt cgcagtcgca 9600

aggtaggctg agcacggttt cttctggcgg gtcatgttgg ttgggagcgg ggcgggcgat 9660

gctgctggtg atgaagttga aataggcggt tctgagacgg cggatggtgg cgaggagcac 9720

caggtctttg ggcccggctt gctggatgcg cagacggtcg gccatgcccc aggcgtggtc 9780

ctgacacctg gccaggtcct tgtagtagtc ctgcatgagc cgctccacgg gcacctcctc 9840

ctcgcccgcg cggccgtgca tgcgcgtgag cccgaagccg cgctggggct ggacgagcgc 9900

caggtcggcg acgacgcgct cggcgaggat ggcttgctgg atctgggtga gggtggtctg 9960

gaagtcatca aagtcgacga agcggtggta ggctccggtg ttgatggtgt aggagcagtt 10020

ggccatgacg gaccagttga cggtctggtg gccccggacgc acgagctcgt ggtacttgag 10080

gcgcgagtag gcgcgcgtgt cgaagatgta gtcgttgcag gtgcgcacca ggtactggta 10140

gccgatgagg aagtgcggcg gcggctggcg gtagagcggc catcgctcgg tggcgggggc 10200

gccgggcgcg aggtcctcga gcatggtgcg gtggtagccg tagatgtacc tggacatcca 10260

ggtgatgccg gcggcggtgg tggaggcgcg cgggaactcg cggacgcggt tccagatgtt 10320

gcgcagcggc aggaagtagt tcatggtggg cacggtctgg cccgtgaggc gcgcgcagtc 10380

gtggatgctc tatacgggca aaaacgaaag cggtcagcgg ctcgactccg tggcctggag 10440

[0098] gctaagcgaa cgggttgggc tgcgcgtgta ccccggttcg aatctcgaat caggctggag 10500

ccgcagctaa cgtggtattg gcactcccgt ctcgacccaa gcctgcacca accctccagg 10560

atacggaggc gggtcgtttt gcaacttttt tttggaggcc ggatgagact agtaagcgcg 10620

gaaagcggcc gaccgcgatg gctcgctgcc gtagtctgga gaagaatcgc cagggttgcg 10680

ttgcggtgtg ccccggttcg aggccggccg gattccgcgg ctaacgaggg cgtggctgcc 10740

ccgtcgtttc caagacccca tagccagccg acttctccag ttacggagcg agcccctctt 10800

ttgttttgtt tgtttttgcc agatgcatcc cgtactgcgg cagatgcgcc cccaccaccc 10860

tccaccgcaa caacagcccc ctccacagcc ggcgcttctg cccccgcccc agcagcaact 10920

tccagccacg accgccgcgg ccgccgtgag cggggctgga cagagttatg atcaccagct 10980

ggccttggaa gagggcgagg ggctggcgcg cctgggggcg tcgtcgccgg agcggcaccc 11040

gcgcgtgcag atgaaaaggg acgctcgcga ggcctacgtg cccaagcaga acctgttcag 11100

agacaggagc ggcgaggagc ccgaggagat gcgcgcggcc cggttccacg cggggcggga 11160

gctgcggcgc ggcctggacc gaaagagggt gctgagggac gaggatttcg aggcggacga 11220

gctgacgggg atcagccccg cgcgcgcgca cgtggccgcg gccaacctgg tcacggcgta 11280

cgagcagacc gtgaaggagg agagcaactt ccaaaaatcc ttcaacaacc acgtgcgcac 11340

cctgatcgcg cgcgaggagg tgaccctggg cctgatgcac ctgtgggacc tgctggaggc 11400

catcgtgcag aaccccacca gcaagccgct gacggcgcag ctgttcctgg tggtgcagca 11460

tagtcgggac aacgaagcgt tcagggaggc gctgctgaat atcaccgagc ccgagggccg 11520

ctggctcctg gacctggtga acattctgca gagcatcgtg gtgcaggagc gcgggctgcc 11580

gctgtccgag aagctggcgg ccatcaactt ctcggtgctg agtttgggca agtactacgc 11640

taggaagatc tacaagaccc cgtacgtgcc catagacaag gaggtgaaga tcgacgggtt 11700

ttacatgcgc atgaccctga aagtgctgac cctgagcgac gatctggggg tgtaccgcaa 11760

cgacaggatg caccgtgcgg tgagcgccag caggcggcgc gagctgagcg accaggagct 11820

gatgcatagt ctgcagcggg ccctgaccgg ggccgggacc gaggggggaga gctactttga 11880

catgggcgcg gacctgcact ggcagcccag ccgccgggcc ttggaggcgg cggcaggacc 11940

[0099] ctacgtagaa gaggtggacg atgaggtgga cgaggagggc gagtacctgg aagactgatg 12000

gcgcgaccgt atttttgcta gatgcaacaa caacagccac ctcctgatcc cgcgatgcgg 12060

gcggcgctgc agagccagcc gtccggcatt aactcctcgg acgattggac ccaggccatg 12120

caacgcatca tggcgctgac gacccgcaac cccgaagcct ttagacagca gccccaggcc 12180

aaccggctct cggccatcct ggaggccgtg gtgccctcgc gctccaaccc cacgcacgag 12240

aaggtcctgg ccatcgtgaa cgcgctggtg gagaacaagg ccatccgcgg cgacgaggcc 12300

ggcctggtgt acaacgcgct gctggagcgc gtggcccgct acaacagcac caacgtgcag 12360

accaacctgg accgcatggt gaccgacgtg cgcgaggccg tggcccagcg cgagcggttc 12420

caccgcgagt ccaacctggg atccatggtg gcgctgaacg ccttcctcag cacccagccc 12480

gccaacgtgc cccggggcca ggaggactac accaacttca tcagcgccct gcgcctgatg 12540

gtgaccgagg tgccccagag cgaggtgtac cagtccgggc cggactactt cttccagacc 12600

agtcgccagg gcttgcagac cgtgaacctg agccaggctt tcaagaactt gcagggcctg 12660

tggggcgtgc aggccccggt cggggaccgc gcgacggtgt cgagcctgct gacgccgaac 12720

tcgcgcctgc tgctgctgct ggtggccccc ttcacggaca gcggcagcat caaccgcaac 12780

tcgtacctgg gctacctgat taacctgtac cgcgaggcca tcggccaggc gcacgtggac 12840

gagcagacct accaggagat cacccacgtg agccgcgccc tgggccagga cgacccgggc 12900

aacctggaag ccaccctgaa ctttttgctg accaaccggt cgcagaagat cccgccccag 12960

tacgcgctca gcaccgagga ggagcgcatc ctgcgttacg tgcagcagag cgtgggcctg 13020

ttcctgatgc aggagggggc cacccccagc gccgcgctcg acatgaccgc gcgcaacatg 13080

gagcccagca tgtacgccag caaccgcccg ttcatcaata aactgatgga ctacttgcat 13140

cgggcggccg ccatgaactc tgactatttc accaacgcca tcctgaatcc ccactggctc 13200

ccgccgccgg ggttctacac gggcgagtac gacatgcccg accccaatga cgggttcctg 13260

tgggacgatg tggacagcag cgtgttctcc ccccgaccgg gtgctaacga gcgccccttg 13320

tggaagaagg aaggcagcga ccgacgcccg tcctcggcgc tgtccggccg cgagggtgct 13380

gccgcggcgg tgcccgaggc cgccagtcct ttcccgagct tgcccttctc gctgaacagt 13440

[0100] atccgcagca gcgagctggg caggatcacg cgcccgcgct tgctgggcga agaggagtac 13500

ttgaatgact cgctgttgag acccgagcgg gagaagaact tccccaataa cgggatagaa 13560

agcctggtgg acaagatgag ccgctggaag acgtatgcgc aggagcacag ggacgatccc 13620

cgggcgtcgc agggggccac gagccggggc agcgccgccc gtaaacgccg gtggcacgac 13680

aggcagcggg gacagatgtg ggacgatgag gactccgccg acgacagcag cgtgttggac 13740

ttgggtggga gtggtaaccc gttcgctcac ctgcgccccc gtatcgggcg catgatgtaa 13800

gagaaaccga aaataaatga tactcaccaa ggccatggcg accagcgtgc gttcgtttct 13860

tctctgttgt tgttgtatct agtatgatga ggcgtgcgta cccggagggt cctcctccct 13920

cgtacgagag cgtgatgcag caggcgatgg cggcggcggc gatgcagccc ccgctggagg 13980

ctccttacgt gcccccgcgg tacctggcgc ctacggaggg gcggaacagc attcgttact 14040

cggagctggc acccttgtac gataccaccc ggttgtacct ggtggacaac aagtcggcgg 14100

acatcgcctc gctgaactac cagaacgacc acagcaactt cctgaccacc gtggtgcaga 14160

acaatgactt cacccccacg gaggccagca cccagaccat caactttgac gagcgctcgc 14220

ggtggggcgg ccagctgaaa accatcatgc acaccaacat gcccaacgtg aacgagttca 14280

tgtacagcaa caagttcaag gcgcgggtga tggtctcccg caagaccccc aatggggtga 14340

cagtgacaga ggattatgat ggtagtcagg atgagctgaa gtatgaatgg gtggaatttg 14400

agctgcccga aggcaacttc tcggtgacca tgaccatcga cctgatgaac aacgccatca 14460

tcgacaatta cttggcggtg gggcggcaga acggggtgct ggagagcgac atcggcgtga 14520

agttcgacac taggaacttc aggctgggct gggaccccgt gaccgagctg gtcatgcccg 14580

gggtgtacac caacgaggct ttccatcccg atattgtctt gctgcccggc tgcggggtgg 14640

acttcaccga gagccgcctc agcaacctgc tgggcattcg caagaggcag cccttccagg 14700

aaggcttcca gatcatgtac gaggatctgg agggggggcaa catccccgcg ctcctggatg 14760

tcgacgccta tgagaaaagc aaggaggatg cagcagctga agcaactgca gccgtagcta 14820

ccgcctctac cgaggtcagg ggcgataatt ttgcaagcgc cgcagcagtg gcagcggccg 14880

aggcggctga aaccgaaagt aagatagtca ttcagccggt ggagaaggat agcaagaaca 14940

[0101] ggagctacaa cgtactaccg gacaagataa acaccgccta ccgcagctgg tacctagcct 15000

acaactatgg cgaccccgag aagggcgtgc gctcctggac gctgctcacc acctcggacg 15060

tcacctgcgg cgtggagcaa gtctactggt cgctgcccga catgatgcaa gacccggtca 15120

ccttccgctc cacgcgtcaa gttagcaact acccggtggt gggcgccgag ctcctgcccg 15180

tctactccaa gagcttcttc aacgagcagg ccgtctactc gcagcagctg cgcgccttca 15240

cctcgcttac gcacgtcttc aaccgcttcc ccgagaacca gatcctcgtc cgcccgcccg 15300

cgcccaccat taccaccgtc agtgaaaacg ttcctgctct cacagatcac gggaccctgc 15360

cgctgcgcag cagtatccgg ggagtccagc gcgtgaccgt tactgacgcc agacgccgca 15420

cctgccccta cgtctacaag gccctgggca tagtcgcgcc gcgcgtcctc tcgagccgca 15480

ccttctaaat gtccattctc atctcgccca gtaataacac cggttggggc ctgcgcgcgc 15540

ccagcaagat gtacggaggc gctcgccaac gctccacgca acaccccgtg cgcgtgcgcg 15600

ggcacttccg cgctccctgg ggcgccctca agggccgcgt gcggtcgcgc accaccgtcg 15660

acgacgtgat cgaccaggtg gtggccgacg cgcgcaacta cacccccgcc gccgcgcccg 15720

tctccaccgt ggacgccgtc atcgacagcg tggtggccga cgcgcgccgg tacgcccgcg 15780

ccaagagccg gcggcggcgc atcgcccggc ggcaccggag cacccccgcc atgcgcgcgg 15840

cgcgagcctt gctgcgcagg gccaggcgca cgggacgcag ggccatgctc agggcggcca 15900

gacgcgcggc ttcaggcgcc agcgccggca ggacccggag acgcgcggcc acggcggcgg 15960

cagcggccat cgccagcatg tcccgcccgc ggcgagggaa cgtgtactgg gtgcgcgacg 16020

ccgccaccgg tgtgcgcgtg cccgtgcgca cccgcccccc tcgcacttga agatgttcac 16080

ttcgcgatgt tgatgtgtcc cagcggcgag gaggatgtcc aagcgcaaat tcaaggaaga 16140

gatgctccag gtcatcgcgc ctgagatcta cggccctgcg gtggtgaagg aggaaagaaa 16200

gccccgcaaa atcaagcggg tcaaaaagga caaaaaggaa gaagaaagtg atgtggacgg 16260

attggtggag tttgtgcgcg agttcgcccc ccggcggcgc gtgcagtggc gcgggcggaa 16320

ggtgcaaccg gtgctgagac ccggcaccac cgtggtcttc acgcccggcg agcgctccgg 16380

caccgcttcc aagcgctcct acgacgaggt gtacggggat gatgatattc tggagcaggc 16440

[0102] ggccgagcgc ctgggcgagt ttgcttacgg caagcgcagc cgttccgcac cgaaggaaga 16500

ggcggtgtcc atcccgctgg accacggcaa ccccacgccg agcctcaagc ccgtgacctt 16560

gcagcaggtg ctgccgaccg cggcgccgcg ccgggggttc aagcgcgagg gcgaggatct 16620

gtaccccacc atgcagctga tggtgcccaa gcgccagaag ctggaagacg tgctggagac 16680

catgaaggtg gacccggacg tgcagcccga ggtcaaggtg cggcccatca agcaggtggc 16740

cccgggcctg ggcgtgcaga ccgtggacat caagattccc acggagccca tggaaacgca 16800

gaccgagccc atgatcaagc ccagcaccag caccatggag gtgcagacgg atccctggat 16860

gccatcggct cctagtcgaa gaccccggcg caagtacggc gcggccagcc tgctgatgcc 16920

caactacgcg ctgcatcctt ccatcatccc cacgccgggc taccgcggca cgcgcttcta 16980

ccgcggtcat accagcagcc gccgccgcaa gaccaccact cgccgccgcc gtcgccgcac 17040

cgccgctgca accacccctg ccgccctggt gcggagagtg taccgccgcg gccgcgcacc 17100

tctgaccctg ccgcgcgcgc gctaccaccc gagcatcgcc atttaaactt tcgcctgctt 17160

tgcagatcaa tggccctcac atgccgcctt cgcgttccca ttacgggcta ccgaggaaga 17220

aaaccgcgcc gtagaaggct ggcggggaac gggatgcgtc gccaccacca ccggcggcgg 17280

cgcgccatca gcaagcggtt gggggggaggc ttcctgcccg cgctgatccc catcatcgcc 17340

gcggcgatcg gggcgatccc cggcattgct tccgtggcgg tgcaggcctc tcagcgccac 17400

tgagacacac ttggaaacat cttgtaataa accaatggac tctgacgctc ctggtcctgt 17460

gatgtgtttt cgtagacaga tggaagacat caatttttcg tccctggctc cgcgacacgg 17520

cacgcggccg ttcatgggca cctggagcga catcggcacc agccaactga acgggggcgc 17580

cttcaattgg agcagtctct ggagcgggct taagaatttc gggtccacgc ttaaaaccta 17640

tggcagcaag gcgtggaaca gcaccacagg gcaggcgctg agggataagc tgaaagagca 17700

gaacttccag cagaaggtgg tcgatgggct cgcctcgggc atcaacgggg tggtggacct 17760

ggccaaccag gccgtgcagc ggcagatcaa cagccgcctg gacccggtgc cgcccgccgg 17820

ctccgtggag atgccgcagg tggaggagga gctgcctccc ctggacaagc ggggcgagaa 17880

gcgacccccgc cccgatgcgg aggagacgct gctgacgcac acggacgagc cgcccccgta 17940

[0103] cgaggaggcg gtgaaactgg gtctgcccac cacgcggccc atcgcgcccc tggccaccgg 18000

ggtgctgaaa cccgaaaagc ccgcgaccct ggacttgcct cctccccagc cttcccgccc 18060

ctctacagtg gctaagcccc tgccgccggt ggccgtggcc cgcgcgcgac ccgggggcac 18120

cgcccgccct catgcgaact ggcagagcac tctgaacagc atcgtgggtc tgggagtgca 18180

gagtgtgaag cgccgccgct gctattaaac ctaccgtagc gcttaacttg cttgtctgtg 18240

tgtgtatgta ttatgtcgcc gccgccgctg tccaccagaa ggaggagtga agaggcgcgt 18300

cgccgagttg caagatggcc accccatcga tgctgcccca gtgggcgtac atgcacatcg 18360

ccggacagga cgcttcggag tacctgagtc cgggtctggt gcagtttgcc cgcgccacag 18420

acacctactt cagtctgggg aacaagttta ggaaccccac ggtggcgccc acgcacgatg 18480

tgaccaccga ccgcagccag cggctgacgc tgcgcttcgt gcccgtggac cgcgaggaca 18540

acacctactc gtacaaagtg cgctacacgc tggccgtggg cgacaaccgc gtgctggaca 18600

tggccagcac ctactttgac atccgcggcg tgctggatcg gggccctagc ttcaaaccct 18660

actccggcac cgcctacaac agtctggccc ccaagggagc acccaacact tgtcagtgga 18720

catataaagc cgatggtgaa actgccacag aaaaaaccta tacatatgga aatgcacccg 18780

tgcagggcat taacatcaca aaagatggta ttcaacttgg aactgacacc gatgatcagc 18840

caatctacgc agataaaacc tatcagcctg aacctcaagt gggtgatgct gaatggcatg 18900

acatcactgg tactgatgaa aagtatggag gcagagctct taagcctgat accaaaatga 18960

agccttgtta tggttctttt gccaagccta ctaataaaga aggaggtcag gcaaatgtga 19020

aaacaggaac aggcactact aaagaatatg acatagacat ggctttcttt gacaacagaa 19080

gtgcggctgc tgctggccta gctccagaaa ttgttttgta tactgaaaat gtggatttgg 19140

aaactccaga tacccatatt gtatacaaag caggcacaga tgacagcagc tcttctatta 19200

atttgggtca gcaagccatg cccaacagac ctaactacat tggtttcaga gacaacttta 19260

tcgggctcat gtactacaac agcactggca atatgggggt gctggccggt caggcttctc 19320

agctgaatgc tgtggttgac ttgcaagaca gaaacaccga gctgtcctac cagctcttgc 19380

ttgactctct gggtgacaga acccggtatt tcagtatgtg gaatcaggcg gtggacagct 19440

[0104] atgatcctga tgtgcgcatt attgaaaatc atggtgtgga ggatgaactt cccaactatt 19500

gtttccctct ggatgctgtt ggcagaacag atacttatca gggaattaag gctaatggaa 19560

ctgatcaaac cacatggacc aaagatgaca gtgtcaatga tgctaatgag ataggcaagg 19620

gtaatccatt cgccatggaa atcaacatcc aagccaacct gtggaggaac ttcctctacg 19680

ccaacgtggc cctgtacctg cccgactctt acaagtacac gccggccaat gttaccctgc 19740

ccaccaacac caacacctac gattacatga acggccgggt ggtggcgccc tcgctggtgg 19800

actcctacat caacatcggg gcgcgctggt cgctggatcc catggacaac gtgaacccct 19860

tcaaccacca ccgcaatgcg gggctgcgct accgctccat gctcctgggc aacgggcgct 19920

acgtgccctt ccacatccag gtgccccaga aatttttcgc catcaagagc ctcctgctcc 19980

tgcccgggtc ctacacctac gagtggaact tccgcaagga cgtcaacatg atcctgcaga 20040

gctccctcgg caacgacctg cgcacggacg gggcctccat ctccttcacc agcatcaacc 20100

tctacgccac cttcttcccc atggcgcaca acacggcctc cacgctcgag gccatgctgc 20160

gcaacgacac caacgaccag tccttcaacg actacctctc ggcggccaac atgctctacc 20220

ccatcccggc caacgccacc aacgtgccca tctccatccc ctcgcgcaac tgggccgcct 20280

tccgcggctg gtccttcacg cgtctcaaga ccaaggagac gccctcgctg ggctccgggt 20340

tcgaccccta cttcgtctac tcgggctcca tcccctacct cgacggcacc ttctacctca 20400

accacacctt caagaaggtc tccatcacct tcgactcctc cgtcagctgg cccggcaacg 20460

accggctcct gacgcccaac gagttcgaaa tcaagcgcac cgtcgacggc gagggctaca 20520

acgtggccca gtgcaacatg accaaggact ggttcctggt ccagatgctg gcccactaca 20580

acatcggcta ccagggcttc tacgtgcccg agggctacaa ggaccgcatg tactccttct 20640

tccgcaactt ccagcccatg agccgccagg tggtggacga ggtcaactac aaggactacc 20700

aggccgtcac cctggcctac cagcacaaca actcgggctt cgtcggctac ctcgcgccca 20760

ccatgcgcca gggccagccc taccccgcca actaccccta cccgctcatc ggcaagagcg 20820

ccgtcaccag cgtcacccag aaaaagttcc tctgcgacag ggtcatgtgg cgcatccccct 20880

tctccagcaa cttcatgtcc atgggcgcgc tcaccgacct cggccagaac atgctctatg 20940

[0105] ccaactccgc ccacgcgcta gacatgaatt tcgaagtcga ccccatggat gagtccaccc 21000

ttctctatgt tgtcttcgaa gtcttcgacg tcgtccgagt gcaccagccc caccgcggcg 21060

tcatcgaggc cgtctacctg cgcacccccct tctcggccgg taacgccacc acctaagctc 21120

ttgcttcttg caagccatgg ccgcgggctc cggcgagcag gagctcaggg ccatcatccg 21180

cgacctgggc tgcgggccct acttcctggg caccttcgat aagcgcttcc cgggattcat 21240

ggccccgcac aagctggcct gcgccatcgt caacacggcc ggccgcgaga ccgggggcga 21300

gcactggctg gccttcgcct ggaacccgcg ctcgaacacc tgctacctct tcgacccctt 21360

cgggttctcg gacgagcgcc tcaagcagat ctaccagttc gagtacgagg gcctgctgcg 21420

ccgcagcgcc ctggccaccg aggaccgctg cgtcaccctg gaaaagtcca cccagaccgt 21480

gcagggtccg cgctcggccg cctgcgggct cttctgctgc atgttcctgc acgccttcgt 21540

gcactggccc gaccgcccca tggacaagaa ccccaccatg aacttgctga cgggggtgcc 21600

caacggcatg ctccagtcgc cccaggtgga acccaccctg cgccgcaacc aggaggcgct 21660

ctaccgcttc ctcaactccc actccgccta ctttcgctcc caccgcgcgc gcatcgagaa 21720

ggccaccgcc ttcgaccgca tgaatcaaga catgtaaacc gtgtgtgtat gttaaatgtc 21780

tttaataaac agcactttca tgttacacat gcatctgaga tgatttattt agaaatcgaa 21840

agggttctgc cgggtctcgg catggcccgc gggcagggac acgttgcgga actggtactt 21900

ggccagccac ttgaactcgg ggatcagcag tttgggcagc ggggtgtcgg ggaaggagtc 21960

ggtccacagc ttccgcgtca gttgcagggc gcccagcagg tcgggcgcgg agatcttgaa 22020

atcgcagttg ggacccgcgt tctgcgcgcg ggagttgcgg tacacggggt tgcagcactg 22080

gaacaccatc agggccgggt gcttcacgct cgccagcacc gtcgcgtcgg tgatgctctc 22140

cacgtcgagg tcctcggcgt tggccatccc gaaggggtc atcttgcagg tctgccttcc 22200

catggtgggc acgcacccgg gcttgtggtt gcaatcgcag tgcaggggga tcagcatcat 22260

ctgggcctgg tcggcgttca tccccgggta catggccttc atgaaagcct ccaattgcct 22320

gaacgcctgc tgggccttgg ctccctcggt gaagaagacc ccgcaggact tgctagagaa 22380

ctggttggtg gcgcacccgg cgtcgtgcac gcagcagcgc gcgtcgttgt tggccagctg 22440

[0106] caccacgctg cgcccccagc ggttctgggt gatcttggcc cggtcggggt tctccttcag 22500

cgcgcgctgc ccgttctcgc tcgccacatc catctcgatc atgtgctcct tctggatcat 22560

ggtggtcccg tgcaggcacc gcagcttgcc ctcggcctcg gtgcacccgt gcagccacag 22620

cgcgcacccg gtgcactccc agttcttgtg ggcgatctgg gaatgcgcgt gcacgaagcc 22680

ctgcaggaag cggcccatca tggtggtcag ggtcttgttg ctagtgaagg tcagcggaat 22740

gccgcggtgc tcctcgttga tgtacaggtg gcagatgcgg cggtacacct cgccctgctc 22800

gggcatcagc tggaagttgg ctttcaggtc ggtctccacg cggtagcggt ccatcagcat 22860

agtcatgatt tccataccct tctcccaggc cgagacgatg ggcaggctca tagggttctt 22920

caccatcatc ttagcgctag cagccgcggc caggggtcg ctctcgtcca gggtctcaaa 22980

gctccgcttg ccgtccttct cggtgatccg caccgggggg tagctgaagc ccacggccgc 23040

cagctcctcc tcggcctgtc tttcgtcctc gctgtcctgg ctgacgtcct gcaggaccac 23100

atgcttggtc ttgcggggtt tcttcttggg cggcagcggc ggcggagatg ttggagatgg 23160

cgaggggggag cgcgagttct cgctcaccac tactatctct tcctcttctt ggtccgaggc 23220

cacgcggcgg taggtatgtc tcttcggggg cagaggcgga ggcgacgggc tctcgccgcc 23280

gcgacttggc ggatggctgg cagagcccct tccgcgttcg ggggtgcgct cccggcggcg 23340

ctctgactga cttcctccgc ggccggccat tgtgttctcc tagggaggaa caacaagcat 23400

ggagactcag ccatcgccaa cctcgccatc tgcccccacc gccgacgaga agcagcagca 23460

gcagaatgaa agcttaaccg ccccgccgcc cagccccgcc acctccgacg cggccgtccc 23520

agacatgcaa gagatggagg aatccatcga gattgacctg ggctatgtga cgcccgcgga 23580

gcacgaggag gagctggcag tgcgcttttc acaagaagag atacaccaag aacagccaga 23640

gcaggaagca gagaatgagc agagtcaggc tgggctcgag catgacggcg actacctcca 23700

cctgagcggg ggggaggacg cgctcatcaa gcatctggcc cggcaggcca ccatcgtcaa 23760

ggatgcgctg ctcgaccgca ccgaggtgcc cctcagcgtg gaggagctca gccgcgccta 23820

cgagttgaac ctcttctcgc cgcgcgtgcc ccccaagcgc cagcccaatg gcacctgcga 23880

gcccaacccg cgcctcaact tctacccggt cttcgcggtg cccgaggccc tggccaccta 23940

[0107] ccacatcttt ttcaagaacc aaaagatccc cgtctcctgc cgcgccaacc gcacccgcgc 24000

cgacgccctt ttcaacctgg gtcccggcgc ccgcctacct gatatcgcct ccttggaaga 24060

ggttcccaag atcttcgagg gtctgggcag cgacgagact cgggccgcga acgctctgca 24120

aggagaagga ggagagcatg agcaccacag cgccctggtc gagttggaag gcgacaacgc 24180

gcggctggcg gtgctcaaac gcacggtcga gctgacccat ttcgcctacc cggctctgaa 24240

cctgcccccc aaagtcatga gcgcggtcat ggaccaggtg ctcatcaagc gcgcgtcgcc 24300

catctccgag gacgagggca tgcaagactc cgaggagggc aagcccgtgg tcagcgacga 24360

gcagctggcc cggtggctgg gtcctaatgc tagtccccag agtttggaag agcggcgcaa 24420

actcatgatg gccgtggtcc tggtgaccgt ggagctggag tgcctgcgcc gcttcttcgc 24480

cgacgcggag accctgcgca aggtcgagga gaacctgcac tacctcttca ggcacgggtt 24540

cgtgcgccag gcctgcaaga tctccaacgt ggagctgacc aacctggtct cctacatggg 24600

catcttgcac gagaaccgcc tggggcagaa cgtgctgcac accaccctgc gcggggaggc 24660

ccggcgcgac tacatccgcg actgcgtcta cctctacctc tgccacacct ggcagacggg 24720

catgggcgtg tggcagcagt gtctggagga gcagaacctg aaagagctct gcaagctcct 24780

gcagaagaac ctcaagggtc tgtggaccgg gttcgacgag cgcaccaccg cctcggacct 24840

ggccgacctc attttccccg agcgcctcag gctgacgctg cgcaacggcc tgcccgactt 24900

tatgagccaa agcatgttgc aaaactttcg ctctttcatc ctcgaacgct ccggaatcct 24960

gcccgccacc tgctccgcgc tgccctcgga cttcgtgccg ctgaccttcc gcgagtgccc 25020

cccgccgctg tggagccact gctacctgct gcgcctggcc aactacctgg cctaccactc 25080

ggacgtgatc gaggacgtca gcggcgaggg cctgctcgag tgccactgcc gctgcaacct 25140

ctgcacgccg caccgctccc tggcctgcaa cccccagctg ctgagcgaga cccagatcat 25200

cggcaccttc gagttgcaag ggcccagcga aggcgagggt tcagccgcca agggggtct 25260

gaaactcacc ccggggctgt ggacctcggc ctacttgcgc aagttcgtgc ccgaggacta 25320

ccatcccttc gagatcaggt tctacgagga ccaatcccat ccgcccaagg ccgagctgtc 25380

ggcctgcgtc atcacccagg gggcgatcct ggcccaattg caagccatcc agaaatcccg 25440

[0108] ccaagaattc ttgctgaaaa agggccgcgg ggtctacctc gacccccaga ccggtgagga 25500

gctcaacccc ggcttccccc aggatgcccc gaggaaacaa gaagctgaaa gtggagctgc 25560

cgcccgtgga ggatttggag gaagactggg agaacagcag tcaggcagag gaggaggaga 25620

tggaggaaga ctgggacagc actcaggcag aggaggacag cctgcaagac agtctggagg 25680

aagacgagga ggaggcagag gaggaggtgg aagaagcagc cgccgccaga ccgtcgtcct 25740

cggcggggga gaaagcaagc agcacggata ccatctccgc tccgggtcgg ggtcccgctc 25800

gaccacacag tagatgggac gagaccggac gattcccgaa ccccaccacc cagaccggta 25860

agaaggagcg gcagggatac aagtcctggc gggggcacaa aaacgccatc gtctcctgct 25920

tgcaggcctg cgggggcaac atctccttca cccggcgcta cctgctcttc caccgcgggg 25980

tgaactttcc ccgcaacatc ttgcattact accgtcacct ccacagcccc tactacttcc 26040

aagaagaggc agcagcagca gaaaaagacc agcagaaaac cagcagctag aaaatccaca 26100

gcggcggcag caggtggact gaggatcgcg gcgaacgagc cggcgcaaac ccgggagctg 26160

aggaaccgga tctttcccac cctctatgcc atcttccagc agagtcgggg gcaggagcag 26220

gaactgaaag tcaagaaccg ttctctgcgc tcgctcaccc gcagttgtct gtatcacaag 26280

agcgaagacc aacttcagcg cactctcgag gacgccgagg ctctcttcaa caagtactgc 26340

gcgctcactc ttaaagagta gcccgcgccc gcccagtcgc agaaaaaggc gggaattacg 26400

tcacctgtgc ccttcgccct agccgcctcc acccatcatc atgagcaaag agattcccac 26460

gccttacatg tggagctacc agccccagat gggcctggcc gccggtgccg cccaggacta 26520

ctccacccgc atgaattggc tcagcgccgg gcccgcgatg atctcacggg tgaatgacat 26580

ccgcgcccac cgaaaccaga tactcctaga acagtcagcg ctcaccgcca cgccccgcaa 26640

tcacctcaat ccgcgtaatt ggcccgccgc cctggtgtac caggaaattc cccagcccac 26700

gaccgtacta cttccgcgag acgcccaggc cgaagtccag ctgactaact caggtgtcca 26760

gctggcgggc ggcgccaccc tgtgtcgtca ccgccccgct cagggtataa agcggctggt 26820

gatccggggc agaggcacac agctcaacga cgaggtggtg agctcttcgc tgggtctgcg 26880

acctgacgga gtcttccaac tcgccggatc ggggagatct tccttcacgc ctcgtcaggc 26940

[0109] cgtcctgact ttggagagtt cgtcctcgca gccccgctcg ggtggcatcg gcactctcca 27000

gttcgtggag gagttcactc cctcggtcta cttcaacccc ttctccggct cccccggcca 27060

ctacccggac gagttcatcc cgaacttcga cgccatcagc gagtcggtgg acggctacga 27120

ttgaatgtcc catggtggcg cagctgacct agctcggctt cgacacctgg accactgccg 27180

ccgcttccgc tgcttcgctc gggatctcgc cgagtttgcc tactttgagc tgcccgagga 27240

gcaccctcag ggcccggccc acggagtgcg gatcgtcgtc gaaggggccc tcgactccca 27300

cctgcttcgg atcttcagcc agcgtccgat cctggtcgag cgcgagcaag gacagaccct 27360

tctgactctg tactgcatct gcaaccaccc cggcctgcat gaaagtcttt gttgtctgct 27420

gtgtactgag tataataaaa gctgagatca gcgactactc cggacttccg tgtgttcctg 27480

aatccatcaa ccagtctttg ttcttcaccg ggaacgagac cgagctccag ctccagtgta 27540

agccccacaa gaagtacctc acctggctgt tccagggctc cccgatcgcc gttgtcaacc 27600

actgcgacaa cgacggagtc ctgctgagcg gccctgccaa ccttactttt tccacccgca 27660

gaagcaagct ccagctcttc caacccttcc tccccgggac ctatcagtgc gtctcgggac 27720

cctgccatca caccttccac ctgatcccga ataccacagc gtcgctcccc gctactaaca 27780

accaaactaa cctccaccaa cgccaccgtc gcgacctttc tgaatctaat actaccaccc 27840

acaccggagg tgagctccga ggtcaaccaa cctctgggat ttactacggc ccctgggagg 27900

tggttgggtt aatagcgcta ggcctagttg cgggtgggct tttggttctc tgctacctat 27960

acctcccttg ctgttcgtac ttagtggtgc tgtgttgctg gtttaagaaa tggggaagat 28020

caccctagtg agctgcggtg cgctggtggc ggtgttgctt tcgattgtgg gactgggcgg 28080

tgcggctgta gtgaaggaga aggccgatcc ctgcttgcat ttcaatccca acaaatgcca 28140

gctgagtttt cagcccgatg gcaatcggtg cgcggtactg atcaagtgcg gatgggaatg 28200

cgagaacgtg agaatcgagt acaataacaa gactcggaac aatactctcg cgtccgtgtg 28260

gcagcccggg gaccccgagt ggtacaccgt ctctgtcccc ggtgctgacg gctccccgcg 28320

caccgtgaat aatactttca tttttgcgca catgtgcgac acggtcatgt ggatgagcaa 28380

gcagtacgat atgtggcccc ccacgaagga gaacatcgtg gtcttctcca tcgcttacag 28440

[0110] cctgtgcacg gcgctaatca ccgctatcgt gtgcctgagc attcacatgc tcatcgctat 28500

tcgccccaga aataatgccg aaaaagaaaa acagccataa cgtttttttt cacacctttt 28560

tcagaccatg gcctctgtta aatttttgct tttatttgcc agtctcattg ccgtcattca 28620

tggaatgagt aatgagaaaa ttactattta cactggcact aatcacacat tgaaaggtcc 28680

agaaaaagcc acagaagttt catggtattg ttattttaat gaatcagatg tatctactga 28740

actctgtgga aacaataaca aaaaaaatga gagcattact ctcatcaagt ttcaatgtgg 28800

atctgactta accctaatta acatcactag agactatgta ggtatgtatt atggaactac 28860

agcaggcatt tcggacatgg aattttatca agtttctgtg tctgaaccca ccacgcctag 28920

aatgaccaca accacaaaaa ctacacctgt taccactatg cagctcacta ccaataacat 28980

ttttgccatg cgtcaaatgg tcaacaatag cactcaaccc accccaccca gtgaggaaat 29040

tcccaaatcc atgattggca ttattgttgc tgtagtggtg tgcatgttga tcatcgcctt 29100

gtgcatggtg tactatgcct tctgctacag aaagcacaga ctgaacgaca agctggaaca 29160

cttactaagt gttgaatttt aatttttttag aaccatgaag atcctaggcc ttttaatttt 29220

ttctatcatt acctctgctc tatgcaattc tgacaatgag gacgttactg tcgttgtcgg 29280

atcaaattat acactgaaag gtccagcgaa gggtatgctt tcgtggtatt gctattttgg 29340

atctgacact acagaaactg aattatgcaa tcttaagaat ggcaaaattc aaaattctaa 29400

aattaacaat tatatatgca atggtactga tctgatactc ctcaatatca cgaaatcata 29460

tgctggcagt tacacctgcc ctggagatga tgctgacagt atgatttttt acaaagtaac 29520

tgttgttgat cccactactc cacctccacc caccacaact actcacacca cacacacaga 29580

tcaaaccgca gcagaggagg cagcaaagtt agccttgcag gtccaagaca gttcatttgt 29640

tggcattacc cctacacctg atcagcggtg tccggggctg ctagtcagcg gcattgtcgg 29700

tgtgctttcg ggattagcag tcataatcat ctgcatgttc atttttgctt gctgctatag 29760

aaggctttac cgacaaaaat cagacccact gctgaacctc tatgtttaat tttttccaga 29820

gtcatgaagg cagttagcgc tctagttttt tgttctttga ttggcattgt tttttgcaat 29880

cctattccta aagttagctt tattaaagat gtgaatgtta ctgaggggggg caatgtgaca 29940

[0111] ctggtaggtg tagagggtgc tgaaaacacc acctggacaa aataccacct caatgggtgg 30000

aaagatattt gcaattggag tgtattagtt tatacatgtg agggagttaa tcttaccatt 30060

gtcaatgcca cctcagctca aaatggtaga attcaaggac aaagtgtcag tgtatctaat 30120

gggtatttta cccaacatac ttttatctat gacgttaaag tcataccact gcctacgcct 30180

agcccaccta gcactaccac acagacaacc cacactacac agacaaccac atacagtaca 30240

ttaaatcagc ctaccaccac tacagcagca gaggttgcca gctcgtctgg ggtccgagtg 30300

gcatttttga tgttggcccc atctagcagt cccactgcta gtaccaatga gcagactact 30360

gaatttttgt ccactgtcga gagccacacc acagctacct ccagtgcctt ctctagcacc 30420

gccaatctct cctcgctttc ctctacacca atcagtcccg ctactactcc tagccccgct 30480

cctcttccca ctcccctgaa gcaaacagac ggcggcatgc aatggcagat caccctgctc 30540

attgtgatcg ggttggtcat cctggccgtg ttgctctact acatcttctg ccgccgcatt 30600

cccaacgcgc accgcaagcc ggtctacaag cccatcattg tcgggcagcc ggagccgctt 30660

caggtggaag ggggtctaag gaatcttctc ttctctttta cagtatggtg attgaactat 30720

gattcctaga caattcttga tcactattct tatctgcctc ctccaagtct gtgccaccct 30780

cgctctggtg gccaacgcca gtccagactg tattgggccc ttcgcctcct acgtgctctt 30840

tgccttcacc acctgcatct gctgctgtag catagtctgc ctgcttatca ccttcttcca 30900

gttcattgac tggatctttg tgcgcatcgc ctacctgcgc caccaccccc agtaccgcga 30960

ccagcgagtg gcgcggctgc tcaggctcct ctgataagca tgcgggctct gctacttctc 31020

gcgcttctgc tgttagtgct cccccgtccc gtcgaccccc ggtcccccac ccagtccccc 31080

gaggaggtcc gcaaatgcaa attccaagaa ccctggaaat tcctcaaatg ctaccgccaa 31140

aaatcagaca tgcatcccag ctggatcatg atcattggga tcgtgaacat tctggcctgc 31200

accctcatct cctttgtgat ttacccctgc tttgactttg gttggaactc gccagaggcg 31260

ctctatctcc cgcctgaacc tgacacacca ccacagcaac ctcaggcaca cgcactacca 31320

ccactacagc ctaggccaca atacatgccc atattagact atgaggccga gccacagcga 31380

cccatgctcc ccgctattag ttacttcaat ctaaccggcg gagatgactg acccactggc 31440

[0112] caacaacaac gtcaacgacc ttctcctgga catggacggc cgcgcctcgg agcagcgact 31500

cgcccaactt cgcattcgcc agcagcagga gagagccgtc aaggagctgc aggatgcggt 31560

ggccatccac cagtgcaaga gaggcatctt ctgcctggtg aaacaggcca agatctccta 31620

cgaggtcact ccaaacgacc atcgcctctc ctacgagctc ctgcagcagc gccagaagtt 31680

cacctgcctg gtcggagtca accccatcgt catcacccag cagtctggcg ataccaaggg 31740

gtgcatccac tgctcctgcg actcccccga ctgcgtccac actctgatca agaccctctg 31800

cggcctccgc gacctcctcc ccatgaacta atcacccccc tatccagtga aataaagatc 31860

atattgatga tgattttaca gaaataaaaa ataatcattt gatttgaaat aaagatacaa 31920

tcatattgat gatttgagtt taacaaaaaa ataaagaatc acttacttga aatctgatac 31980

caggtctctg tccatgtttt ctgccaacac cacttcactc ccctcttccc agctctggta 32040

ctgcaggccc cggcgggctg caaacttcct ccacacgctg aaggggatgt caaattcctc 32100

ctgtccctca atcttcattt tatcttctat cagatgtcca aaaagcgcgt ccgggtggat 32160

gatgacttcg accccgtcta cccctacgat gcagacaacg caccgaccgt gcccttcatc 32220

aaccccccct tcgtctcttc agatggattc caagagaagc ccctgggggt gttgtccctg 32280

cgactggccg accccgtcac caccaagaac ggggaaatca ccctcaagct gggagagggg 32340

gtggacctcg attcctcggg aaaactcatc tccaacacgg ccaccaaggc cgccgccccc 32400

ctcagttttt ccaacaacac catttccctt aacatggatc acccctttta cactaaagat 32460

ggaaaattat ccttacaagt ttctccacca ttaaatatac tgagaacaag cattctaaac 32520

acactagctt taggttttgg atcaggttta ggactccgtg gctctgcctt ggcagtacag 32580

ttagtctctc cacttacatt tgatactgat ggaaacataa agcttacctt agacagaggt 32640

ttgcatgtta caacaggaga tgcaattgaa agcaacataa gctgggctaa aggtttaaaa 32700

tttgaagatg gagccatagc aaccaacatt ggaaatgggt tagagtttgg aagcagtagt 32760

acagaaacag gtgttgatga tgcttaccca atccaagtta aacttggatc tggccttagc 32820

tttgacagta caggagccat aatggctggt aacaaagaag acgataaact cactttgtgg 32880

acaacacctg atccatcacc aaactgtcaa atactcgcag aaaatgatgc aaaactaaca 32940

[0113] ctttgcttga ctaaatgtgg tagtcaaata ctggccactg tgtcagtctt agttgtagga 33000

agtggaaacc taaaccccat tactggcacc gtaagcagtg ctcaggtgtt tctacgtttt 33060

gatgcaaacg gtgttctttt aacagaacat tctacactaa aaaaatactg ggggtatagg 33120

cagggagata gcatagatgg cactccatat accaatgctg taggattcat gcccaattta 33180

aaagcttatc caaagtcaca aagttctact actaaaaata atatagtagg gcaagtatac 33240

atgaatggag atgtttcaaa acctatgctt ctcactataa ccctcaatgg tactgatgac 33300

agcaacagta catattcaat gtcattttca tacacctgga ctaatggaag ctatgttgga 33360

gcaacatttg gggctaactc ttataccttc tcatacatcg cccaagaatg aacactgtat 33420

cccaccctgc atgccaaccc ttcccacccc actctgtgga acaaactctg aaacacaaaa 33480

taaaataaag ttcaagtgtt ttattgattc aacagtttta caggattcga gcagttattt 33540

ttcctccacc ctcccaggac atggaataca ccaccctctc cccccgcaca gccttgaaca 33600

tctgaatgcc attggtgatg gacatgcttt tggtctccac gttccacaca gtttcagagc 33660

gagccagtct cgggtcggtc agggagatga aaccctccgg gcactcccgc atctgcacct 33720

cacagctcaa cagctgagga ttgtcctcgg tggtcgggat cacggttatc tggaagaagc 33780

agaagagcgg cggtgggaat catagtccgc gaacgggatc ggccggtggt gtcgcatcag 33840

gccccgcagc agtcgctgcc gccgccgctc cgtcaagctg ctgctcaggg ggtccgggtc 33900

cagggactcc ctcagcatga tgcccacggc cctcagcatc agtcgtctgg tgcggcgggc 33960

gcagcagcgc atgcggatct cgctcaggtc gctgcagtac gtgcaacaca gaaccaccag 34020

gttgttcaac agtccatagt tcaacacgct ccagccgaaa ctcatcgcgg gaaggatgct 34080

acccacgtgg ccgtcgtacc agatcctcag gtaaatcaag tggtgccccc tccagaacac 34140

gctgcccacg tacatgatct ccttgggcat gtggcggttc accacctccc ggtaccacat 34200

caccctctgg ttgaacatgc agccccggat gatcctgcgg aaccacaggg ccagcaccgc 34260

cccgcccgcc atgcagcgaa gagaccccgg gtcccggcaa tggcaatgga ggacccaccg 34320

ctcgtacccg tggatcatct gggagctgaa caagtctatg ttggcacagc acaggcatat 34380

gctcatgcat ctcttcagca ctctcaactc ctcgggggtc aaaaccatat cccagggcac 34440

[0114] ggggaactct tgcaggacag cgaaccccgc agaacagggc aatcctcgca cagaacttac 34500

attgtgcatg gacagggtat cgcaatcagg cagcaccggg tgatcctcca ccagagaagc 34560

gcgggtctcg gtctcctcac agcgtggtaa gggggccggc cgatacgggt gatggcggga 34620

cgcggctgat cgtgttcgcg accgtgtcat gatgcagttg ctttcggaca ttttcgtact 34680

tgctgtagca gaacctggtc cgggcgctgc acaccgatcg ccggcggcgg tctcggcgct 34740

tggaacgctc ggtgttgaaa ttgtaaaaca gccactctct cagaccgtgc agcagatcta 34800

gggcctcagg agtgatgaag atcccatcat gcctgatggc tctgatcaca tcgaccaccg 34860

tggaatgggc cagacccagc cagatgatgc aattttgttg ggtttcggtg acggcggggg 34920

agggaagaac aggaagaacc atgattaact tttaatccaa acggtctcgg agtacttcaa 34980

aatgaagatc gcggagatgg cacctctcgc ccccgctgtg ttggtggaaa ataacagcca 35040

ggtcaaaggt gatacggttc tcgagatgtt ccacggtggc ttccagcaaa gcctccacgc 35100

gcacatccag aaacaagaca atagcgaaag cgggagggtt ctctaattcc tcaatcatca 35160

tgttacactc ctgcaccatc cccagataat tttcattttt ccagccttga atgattcgaa 35220

ctagttcctg aggtaaatcc aagccagcca tgataaagag ctcgcgcaga gcgccctcca 35280

ccggcattct taagcacacc ctcataattc caagatattc tgctcctggt tcacctgcag 35340

cagattgaca agcggaatat caaaatctct gccgcgatcc ctgagctcct ccctcagcaa 35400

taactgtaag tactctttca tatcctctcc gaaattttta gccataggac caccaggaat 35460

aagattaggg caagccacag tacagataaa ccgaagtcct ccccagtgag cattgccaaa 35520

tgcaagactg ctataagcat gctggctaga cccggtgata tcttccagat aactggacag 35580

aaaatcgccc aggcaatttt taagaaaatc aacaaaagaa aaatcctcca ggtggacgtt 35640

tagagcctcg ggaacaacga tgaagtaaat gcaagcggtg cgttccagca tggttagtta 35700

gctgatctgt agaaaaaaca aaaatgaaca ttaaaccatg ctagcctggc gaacaggtgg 35760

gtaaatcgtt ctctccagca ccaggcaggc cacggggtct ccggcgcgac cctcgtaaaa 35820

attgtcgcta tgattgaaaa ccatcacaga gagacgttcc cggtggccgg cgtgaatgat 35880

tcgacaagat gaatacaccc ccggaacatt ggcgtccgcg agtgaaaaaa agcgcccgag 35940

[0115] gaagcaataa ggcactacaa tgctcagtct caagtccagc aaagcgatgc catgcggatg 36000

aagcacaaaa ttctcaggtg cgtacaaaat gtaattactc ccctcctgca caggcagcaa 36060

agcccccgat ccctccaggt acacatacaa agcctcagcg tccatagctt accgagcagc 36120

agcacacaac aggcgcaaga gtcagagaaa ggctgagctc taacctgtcc acccgctctc 36180

tgctcaatat atagcccaga tctacactga cgtaaaggcc aaagtctaaa aatacccgcc 36240

aaataatcac acacgcccag cacacgccca gaaaccggtg acacactcaa aaaaatacgc 36300

gcacttcctc aaacgcccaa aactgccgtc atttccgggt tcccacgcta cgtcatcaaa 36360

acacgacttt caaattccgt cgaccgttaa aaacgtcacc cgccccgccc ctaacggtcg 36420

cccgtctctc agccaatcag cgccccgcat ccccaaattc aaacacctca tttgcatatt 36480

aacgcgcaca aaaagtttga ggtatattat tgatgatgg 36519

<210> 11
<211> 31867
<212> DNA
<213> 人工序列(Artificial Sequence)

<220>
<223> 人工序列的描述：合成多核苷酸

<400> 11

ccatcttcaa taatatacct caaacttttt gtgcgcgtta atatgcaaat gaggcgtttg 60

aatttgggga ggaagggcgg tgattggtcg agggatgagc gaccgttagg ggcggggcga 120

gtgacgtttt gatgacgtgg ttgcgaggag gagccagttt gcaagttctc gtgggaaaag 180

tgacgtcaaa cgaggtgtgg tttgaacacg gaaatactca attttcccgc gctctctgac 240

aggaaatgag gtgtttctgg gcggatgcaa gtgaaaacgg gccattttcg cgcgaaaact 300

gaatgaggaa gtgaaaatct gagtaatttc gcgtttatgg cagggaggag tatttgccga 360

gggccgagta gactttgacc gattacgtgg gggtttcgat taccgtgttt ttcacctaaa 420

tttccgcgta cggtgtcaaa gtccggtgtt tttacgtagg tgtcagctga tcgccagggt 480

atttaaacct gcgctctcca gtcaagaggc cactcttgag tgccagcgag aagagttttc 540

tcctccgcgc cgcgagtcag atctacactt tgaaagtagg gataacaggg taatgacatt 600

[0116] gattattgac tagttgttaa tagtaatcaa ttacggggtc attagttcat agcccatata 660

tggagttccg cgttacataa cttacggtaa atggcccgcc tggctgaccg cccaacgacc 720

cccgcccatt gacgtcaata atgacgtatg ttcccatagt aacgccaata gggactttcc 780

attgacgtca atgggtggag tatttacggt aaactgccca cttggcagta catcaagtgt 840

atcatatgcc aagtccgccc cctattgacg tcaatgacgg taaatggccc gcctggcatt 900

atgcccagta catgacctta cgggactttc ctacttggca gtacatctac gtattagtca 960

tcgctattac catggtgatg cggttttggc agtacaccaa tgggcgtgga tagcggtttg 1020

actcacgggg atttccaagt ctccacccca ttgacgtcaa tgggagtttg ttttggcacc 1080

aaaatcaacg ggactttcca aaatgtcgta ataaccccgc cccgttgacg caaatgggcg 1140

gtaggcgtgt acggtgggag gtctatataa gcagagctcg tttagtgaac cgtcagatcg 1200

cctggaacgc catccacgct gttttgacct ccatagaaga cagcgatcgc gccaccatgg 1260

tgagcaaggg cgaggagctg ttcaccgggg tggtgcccat cctggtcgag ctggacggcg 1320

acgtaaacgg ccacaagttc agcgtgtccg gcgagggcga gggcgatgcc acctacggca 1380

agctgaccct gaagttcatc tgcaccaccg gcaagctgcc cgtgccctgg cccaccctcg 1440

tgaccaccct gacctacggc gtgcagtgct tcagccgcta ccccgaccac atgaagcagc 1500

acgacttctt caagtccgcc atgcccgaag gctacgtcca ggagcgcacc atcttcttca 1560

aggacgacgg caactacaag acccgcgccg aggtgaagtt cgagggcgac accctggtga 1620

accgcatcga gctgaagggc atcgacttca aggaggacgg caacatcctg gggcacaagc 1680

tggagtacaa ctacaacagc cacaacgtct atatcatggc cgacaagcag aagaacggca 1740

tcaaggtgaa cttcaagatc cgccacaaca tcgaggacgg cagcgtgcag ctcgccgacc 1800

actaccagca gaacaccccc atcggcgacg gccccgtgct gctgcccgac aaccactacc 1860

tgagcaccca gtccgccctg agcaaagacc ccaacgagaa gcgcgatcac atggtcctgc 1920

tggagttcgt gaccgccgcc gggatcactc tcggcatgga cgagctttac aagtagtgag 1980

tttaaactcc catttaaatg tgagggttaa tgcttcgagc agacatgata agatacattg 2040

atgagtttgg acaaaccaca actagaatgc agtgaaaaaa atgctttatt tgtgaaattt 2100

[0117] gtgatgctat tgctttattt gtaaccatta taagctgcaa taaacaagtt aacaacaaca 2160

attgcattca ttttatgttt caggttcagg gggagatgtg ggaggttttt taaagcaagt 2220

aaaacctcta caaatgtggt aaaataacta taacggtcct aaggtagcga gtgagtagtg 2280

ttctggggcg ggggaggacc tgcatgaggg ccagaataac tgaaatctgt gcttttctgt 2340

gtgttgcagc agcatgagcg gaagcggctc ctttgaggga ggggtattca gcccttatct 2400

gacggggcgt ctcccctcct gggcgggagt gcgtcagaat gtgatgggat ccacggtgga 2460

cggccggccc gtgcagcccg cgaactcttc aaccctgacc tatgcaaccc tgagctcttc 2520

gtcgttggac gcagctgccg ccgcagctgc tgcatctgcc gccagcgccg tgcgcggaat 2580

ggccatgggc gccggctact acggcactct ggtggccaac tcgagttcca ccaataatcc 2640

cgccagcctg aacgaggaga agctgttgct gctgatggcc cagctcgagg ccttgaccca 2700

gcgcctgggc gagctgaccc agcaggtggc tcagctgcag gagcagacgc gggccgcggt 2760

tgccacggtg aaatccaaat aaaaaatgaa tcaataaata aacggagacg gttgttgatt 2820

ttaacacaga gtctgaatct ttatttgatt tttcgcgcgc ggtaggccct ggaccaccgg 2880

tctcgatcat tgagcacccg gtggatcttt tccaggaccc ggtagaggtg ggcttggatg 2940

ttgaggtaca tgggcatgag cccgtcccgg gggtggaggt agctccattg cagggcctcg 3000

tgctcggggg tggtgttgta aatcacccag tcatagcagg ggcgcagggc atggtgttgc 3060

acaatatctt tgaggaggag actgatggcc acgggcagcc ctttggtgta ggtgtttaca 3120

aatctgttga gctgggaggg atgcatgcgg ggggagatga ggtgcatctt ggcctggatc 3180

ttgagattgg cgatgttacc gcccagatcc cgcctggggt tcatgttgtg caggaccacc 3240

agcacggtgt atccggtgca cttggggaat ttatcatgca acttggaagg gaaggcgtga 3300

aagaatttgg cgacgccttt gtgcccgccc aggttttcca tgcactcatc catgatgatg 3360

gcgatgggcc cgtgggcggc ggcctgggca aagacgtttc gggggtcgga cacatcatag 3420

ttgtggtcct gggtgaggtc atcataggcc attttaatga atttggggcg gagggtgccg 3480

gactggggga caaaggtacc ctcgatcccg ggggcgtagt tcccctcaca gatctgcatc 3540

tcccaggctt tgagctcgga gggggggatc atgtccacct gcggggcgat aaagaacacg 3600

[0118] gtttccgggg cgggggagat gagctgggcc gaaagcaagt tccggagcag ctgggacttg 3660

ccgcagccgg tggggccgta gatgaccccg atgaccggct gcaggtggta gttgagggag 3720

agacagctgc cgtcctcccg gaggaggggg gccacctcgt tcatcatctc gcgcacgtgc 3780

atgttctcgc gcaccagttc cgccaggagg cgctctcccc ccagggatag gagctcctgg 3840

agcgaggcga agtttttcag cggcttgagt ccgtcggcca tgggcatttt ggagagggtt 3900

tgttgcaaga gttccaggcg gtcccagagc tcggtgatgt gctctacggc atctcgatcc 3960

agcagacctc ctcgtttcgc gggttgggac ggctgcggga gtagggcacc agacgatggg 4020

cgtccagcgc agccagggtc cggtccttcc agggtcgcag cgtccgcgtc agggtggtct 4080

ccgtcacggt gaaggggtgc gcgccgggct gggcgcttgc gagggtgcgc ttcaggctca 4140

tccggctggt cgaaaaccgc tcccgatcgg cgccctgcgc gtcggccagg tagcaattga 4200

ccatgagttc gtagttgagc gcctcggccg cgtggccttt ggcgcggagc ttacctttgg 4260

aagtctgccc gcaggcggga cagaggaggg acttgagggc gtagagcttg ggggcgagga 4320

agacggactc gggggcgtag gcgtccgcgc cgcagtgggc gcagacggtc tcgcactcca 4380

cgagccaggt gaggtcgggc tggtcggggt caaaaaccag tttcccgccg ttctttttga 4440

tgcgtttctt acctttggtc tccatgagct cgtgtccccg ctgggtgaca aagaggctgt 4500

ccgtgtcccc gtagaccgac tttatgggcc ggtcctcgag cggtgtgccg cggtcctcct 4560

cgtagaggaa ccccgcccac tccgagacga aagcccgggt ccaggccagc acgaaggagg 4620

ccacgtggga cgggtagcgg tcgttgtcca ccagcgggtc cacctttcc agggtatgca 4680

aacacatgtc cccctcgtcc acatccagga aggtgattgg cttgtaagtg taggccacgt 4740

gaccgggggt cccggccggg ggggtataaa agggtgcggg tccctgctcg tcctcactgt 4800

cttccggatc gctgtccagg agcgccagct gttggggtag gtattccctc tcgaaggcgg 4860

gcatgacctc ggcactcagg ttgtcagttt ctagaaacga ggaggatttg atattgacgg 4920

tgccggcgga gatgcctttc aagagcccct cgtccatctg gtcagaaaag acgatctttt 4980

tgttgtcgag cttggtggcg aaggagccgt agagggcgtt ggagaggagc ttggcgatgg 5040

agcgcatggt ctggtttttt tccttgtcgg cgcgctcctt ggcggcgatg ttgagctgca 5100

[0119] cgtactcgcg cgccacgcac ttccattcgg ggaagacggt ggtcagctcg tcgggcacga 5160

ttctgacctg ccagccccga ttatgcaggg tgatgaggtc cacactggtg gccacctcgc 5220

cgcgcagggg ctcattagtc cagcagaggc gtccgccctt gcgcgagcag aaggggggca 5280

gggggtccag catgacctcg tcgggggggt cggcatcgat ggtgaagatg ccgggcagga 5340

ggtcggggtc aaagtagctg atggaagtgg ccagatcgtc cagggcagct tgccattcgc 5400

gcacggccag cgcgctctcg tagggactga ggggcgtgcc ccagggcatg ggatgggtaa 5460

gcgcggaggc gtacatgccg cagatgtcgt agacgtagag gggctcctcg aggatgccga 5520

tgtaggtggg gtagcagcgc cccccgcgga tgctggcgcg cacgtagtca tacagctcgt 5580

gcgagggggc gaggagcccc gggcccaggt tggtgcgact gggcttttcg gcgcggtaga 5640

cgatctggcg gaaaatggca tgcgagttgg aggagatggt gggcctttgg aagatgttga 5700

agtgggcgtg gggcagtccg accgagtcgc ggatgaagtg ggcgtaggag tcttgcagct 5760

tggcgacgag ctcggcggtg actaggacgt ccagagcgca gtagtcgagg gtctcctgga 5820

tgatgtcata cttgagctgt ccctttttgtt tccacagctc gcggttgaga aggaactctt 5880

cgcggtcctt ccagtactct tcgaggggga acccgtcctg atctgcacgg taagagccta 5940

gcatgtagaa ctggttgacg gccttgtagg cgcagcagcc cttctccacg gggagggcgt 6000

aggcctgggc ggccttgcgc agggaggtgt gcgtgagggc gaaagtgtcc ctgaccatga 6060

ccttgaggaa ctggtgcttg aagtcgatat cgtcgcagcc cccctgctcc cagagctgga 6120

agtccgtgcg cttcttgtag gcggggttgg gcaaagcgaa agtaacatcg ttgaagagga 6180

tcttgcccgc gcggggcata aagttgcgag tgatgcggaa aggttggggc acctcggccc 6240

ggttgttgat gacctgggcg gcgagcacga tctcgtcgaa gccgttgatg ttgtggccca 6300

cgatgtagag ttccacgaat cgcggacggc ccttgacgtg gggcagtttc ttgagctcct 6360

cgtaggtgag ctcgtcgggg tcgctgagcc cgtgctgctc gagcgcccag tcggcgagat 6420

gggggttggc gcggaggaag gaagtccaga gatccacggc cagggcggtt tgcagacggt 6480

cccggtactg acggaactgc tgcccgacgg ccatttttttc gggggtgacg cagtagaagg 6540

tgcgggggtc cccgtgccag cgatcccatt tgagctggag ggcgagatcg agggcgagct 6600

[0120] cgacgagccg gtcgtccccg gagagtttca tgaccagcat gaaggggacg agctgcttgc 6660

cgaaggaccc catccaggtg taggtttcca catcgtaggt gaggaagagc ctttcggtgc 6720

gaggatgcga gccgatgggg aagaactgga tctcctgcca ccaattggag gaatggctgt 6780

tgatgtgatg gaagtagaaa tgccgacggc gcgccgaaca ctcgtgcttg tgtttataca 6840

agcggccaca gtgctcgcaa cgctgcacgg gatgcacgtg ctgcacgagc tgtacctgag 6900

ttcctttgac gaggaatttc agtgggaagt ggagtcgtgg cgcctgcatc tcgtgctgta 6960

ctacgtcgtg gtggtcggcc tggccctctt ctgcctcgat ggtggtcatg ctgacgagcc 7020

cgcgcgggag gcaggtccag acctcggcgc gagcgggtcg gagagcgagg acgagggcgc 7080

gcaggccgga gctgtccagg gtcctgagac gctgcggagt caggtcagtg ggcagcggcg 7140

gcgcgcggtt gacttgcagg agttttcca gggcgcgcgg gaggtccaga tggtacttga 7200

tctccaccgc gccattggtg gcgacgtcga tggcttgcag ggtcccgtgc ccctggggtg 7260

tgaccaccgt cccccgtttc ttcttgggcg gctggggcga cgggggcggt gcctcttcca 7320

tggttagaag cggcggcgag gacgcgcgcc gggcggcagg ggcggctcgg ggcccggagg 7380

caggggcggc aggggcacgt cggcgccgcg cgcgggtagg ttctggtact gcgcccggag 7440

aagactggcg tgagcgacga cgcgacggtt gacgtcctgg atctgacgcc tctgggtgaa 7500

ggccacggga cccgtgagtt tgaacctgaa agagagttcg acagaatcaa tctcggtatc 7560

gttgacggcg gcctgccgca ggatctcttg cacgtcgccc gagttgtcct ggtaggcgat 7620

ctcggtcatg aactgctcga tctcctcctc ttgaaggtct ccgcggccgg cgcgctccac 7680

ggtggccgcg aggtcgttgg agatgcggcc catgagctgc gagaaggcgt tcatgcccgc 7740

ctcgttccag acgcggctgt agaccacgac gccctcggga tcgcgggcgc gcatgaccac 7800

ctgggcgagg ttgagctcca cgtggcgcgt gaagaccgcg tagttgcaga ggcgctggta 7860

gaggtagttg agcgtggtgg cgatgtgctc ggtgacgaag aaatacatga tccagcggcg 7920

gagcggcatc tcgctgacgt cgcccagcgc ctccaaacgt tccatggcct cgtaaaagtc 7980

cacggcgaag ttgaaaaact gggagttgcg cgccgagacg gtcaactcct cctccagaag 8040

acggatgagc tcggcgatgg tggcgcgcac ctcgcgctcg aaggcccccg ggagttcctc 8100

[0121] cacttcctct tcttcctcct ccactaacat ctcttctact tcctcctcag gcggcagtgg 8160

tggcggggga gggggcctgc gtcgccggcg gcgcacgggc agacggtcga tgaagcgctc 8220

gatggtctcg ccgcgccggc gtcgcatggt ctcggtgacg gcgcgcccgt cctcgcgggg 8280

ccgcagcgtg aagacgccgc cgcgcatctc caggtggccg ggggggtccc cgttgggcag 8340

ggagagggcg ctgacgatgc atcttatcaa ttgccccgta gggactccgc gcaaggacct 8400

gagcgtctcg agatccacgg gatctgaaaa ccgctgaacg aaggcttcga gccagtcgca 8460

gtcgcaaggt aggctgagca cggtttcttc tggcgggtca tgttggttgg gagcggggcg 8520

ggcgatgctg ctggtgatga agttgaaata ggcggttctg agacggcgga tggtggcgag 8580

gagcaccagg tctttgggcc cggcttgctg gatgcgcaga cggtcggcca tgccccaggc 8640

gtggtcctga cacctggcca ggtccttgta gtagtcctgc atgagccgct ccacgggcac 8700

ctcctcctcg cccgcgcggc cgtgcatgcg cgtgagcccg aagccgcgct ggggctggac 8760

gagcgccagg tcggcgacga cgcgctcggc gaggatggct tgctggatct gggtgagggt 8820

ggtctggaag tcatcaaagt cgacgaagcg gtggtaggct ccggtgttga tggtgtagga 8880

gcagttggcc atgacggacc agttgacggt ctggtggccc ggacgcacga gctcgtggta 8940

cttgaggcgc gagtaggcgc gcgtgtcgaa gatgtagtcg ttgcaggtgc gcaccaggta 9000

ctggtagccg atgaggaagt gcggcggcgg ctggcggtag agcggccatc gctcggtggc 9060

gggggcgccg ggcgcgaggt cctcgagcat ggtgcggtgg tagccgtaga tgtacctgga 9120

catccaggtg atgccggcgg cggtggtgga ggcgcgcggg aactcgcgga cgcggttcca 9180

gatgttgcgc agcggcagga agtagttcat ggtgggcacg gtctggcccg tgaggcgcgc 9240

gcagtcgtgg atgctctata cgggcaaaaa cgaaagcggt cagcggctcg actccgtggc 9300

ctggaggcta agcgaacggg ttgggctgcg cgtgtacccc ggttcgaatc tcgaatcagg 9360

ctggagccgc agctaacgtg gtattggcac tcccgtctcg acccaagcct gcaccaaccc 9420

tccaggatac ggaggcgggt cgttttgcaa ctttttttg gaggccggat gagactagta 9480

agcgcggaaa gcggccgacc gcgatggctc gctgccgtag tctggagaag aatcgccagg 9540

gttgcgttgc ggtgtgcccc ggttcgaggc cggccggatt ccgcggctaa cgagggcgtg 9600

[0122] gctgccccgt cgtttccaag accccatagc cagccgactt ctccagttac ggagcgagcc 9660

cctcttttgt tttgtttgtt tttgccagat gcatcccgta ctgcggcaga tgcgccccca 9720

ccaccctcca ccgcaacaac agccccctcc acagccggcg cttctgcccc cgccccagca 9780

gcaacttcca gccacgaccg ccgcggccgc cgtgagcggg gctggacaga gttatgatca 9840

ccagctggcc ttggaagagg gcgaggggct ggcgcgcctg ggggcgtcgt cgccggagcg 9900

gcacccgcgc gtgcagatga aaagggacgc tcgcgaggcc tacgtgccca agcagaacct 9960

gttcagagac aggagcggcg aggagcccga ggagatgcgc gcggcccggt tccacgcggg 10020

gcgggagctg cggcgcggcc tggaccgaaa gagggtgctg agggacgagg atttcgaggc 10080

ggacgagctg acggggatca gccccgcgcg cgcgcacgtg gccgcggcca acctggtcac 10140

ggcgtacgag cagaccgtga aggaggagag caacttccaa aaatccttca acaaccacgt 10200

gcgcaccctg atcgcgcgcg aggaggtgac cctgggcctg atgcacctgt gggacctgct 10260

ggaggccatc gtgcagaacc ccaccagcaa gccgctgacg gcgcagctgt tcctggtggt 10320

gcagcatagt cgggacaacg aagcgttcag ggaggcgctg ctgaatatca ccgagcccga 10380

gggccgctgg ctcctggacc tggtgaacat tctgcagagc atcgtggtgc aggagcgcgg 10440

gctgccgctg tccgagaagc tggcggccat caacttctcg gtgctgagtt tgggcaagta 10500

ctacgctagg aagatctaca agaccccgta cgtgcccata gacaaggagg tgaagatcga 10560

cgggttttac atgcgcatga ccctgaaagt gctgaccctg agcgacgatc tgggggtgta 10620

ccgcaacgac aggatgcacc gtgcggtgag cgccagcagg cggcgcgagc tgagcgacca 10680

ggagctgatg catagtctgc agcgggccct gaccggggcc gggaccgagg gggagagcta 10740

ctttgacatg ggcgcggacc tgcactggca gcccagccgc cgggccttgg aggcggcggc 10800

aggaccctac gtagaagagg tggacgatga ggtggacgag gagggcgagt acctggaaga 10860

ctgatggcgc gaccgtattt ttgctagatg caacaacaac agccacctcc tgatcccgcg 10920

atgcgggcgg cgctgcagag ccagccgtcc ggcattaact cctcggacga ttggacccag 10980

gccatgcaac gcatcatggc gctgacgacc cgcaaccccg aagcctttag acagcagccc 11040

caggccaacc ggctctcggc catcctggag gccgtggtgc cctcgcgctc caaccccacg 11100

[0123] cacgagaagg tcctggccat cgtgaacgcg ctggtggaga acaaggccat ccgcggcgac 11160

gaggccggcc tggtgtacaa cgcgctgctg gagcgcgtgg cccgctacaa cagcaccaac 11220

gtgcagacca acctggaccg catggtgacc gacgtgcgcg aggccgtggc ccagcgcgag 11280

cggttccacc gcgagtccaa cctgggatcc atggtggcgc tgaacgcctt cctcagcacc 11340

cagcccgcca acgtgccccg gggccaggag gactacacca acttcatcag cgccctgcgc 11400

ctgatggtga ccgaggtgcc ccagagcgag gtgtaccagt ccgggccgga ctacttcttc 11460

cagaccagtc gccagggctt gcagaccgtg aacctgagcc aggctttcaa gaacttgcag 11520

ggcctgtggg gcgtgcaggc cccggtcggg gaccgcgcga cggtgtcgag cctgctgacg 11580

ccgaactcgc gcctgctgct gctgctggtg gccccccttca cggacagcgg cagcatcaac 11640

cgcaactcgt acctgggcta cctgattaac ctgtaccgcg aggccatcgg ccaggcgcac 11700

gtggacgagc agacctacca ggagatcacc cacgtgagcc gcgccctggg ccaggacgac 11760

ccgggcaacc tggaagccac cctgaacttt ttgctgacca accggtcgca gaagatcccg 11820

ccccagtacg cgctcagcac cgaggaggag cgcatcctgc gttacgtgca gcagagcgtg 11880

ggcctgttcc tgatgcagga gggggccacc cccagcgccg cgctcgacat gaccgcgcgc 11940

aacatggagc ccagcatgta cgccagcaac cgcccgttca tcaataaact gatggactac 12000

ttgcatcggg cggccgccat gaactctgac tatttcacca acgccatcct gaatccccac 12060

tggctcccgc cgccggggtt ctacacgggc gagtacgaca tgcccgaccc caatgacggg 12120

ttcctgtggg acgatgtgga cagcagcgtg ttctcccccc gaccgggtgc taacgagcgc 12180

cccttgtgga agaaggaagg cagcgaccga cgcccgtcct cggcgctgtc cggccgcgag 12240

ggtgctgccg cggcggtgcc cgaggccgcc agtcctttcc cgagcttgcc cttctcgctg 12300

aacagtatcc gcagcagcga gctgggcagg atcacgcgcc cgcgcttgct gggcgaagag 12360

gagtacttga atgactcgct gttgagaccc gagcgggaga agaacttccc caataacggg 12420

atagaaagcc tggtggacaa gatgagccgc tggaagacgt atgcgcagga gcacagggac 12480

gatccccggg cgtcgcaggg ggccacgagc cggggcagcg ccgcccgtaa acgccggtgg 12540

cacgacaggc agcggggaca gatgtgggac gatgaggact ccgccgacga cagcagcgtg 12600

[0124] ttggacttgg gtgggagtgg taacccgttc gctcacctgc gcccccgtat cgggcgcatg 12660

atgtaagaga aaccgaaaat aaatgatact caccaaggcc atggcgacca gcgtgcgttc 12720

gtttcttctc tgttgttgtt gtatctagta tgatgaggcg tgcgtacccg gagggtcctc 12780

ctccctcgta cgagagcgtg atgcagcagg cgatggcggc ggcggcgatg cagcccccgc 12840

tggaggctcc ttacgtgccc ccgcggtacc tggcgcctac ggaggggcgg aacagcattc 12900

gttactcgga gctggcaccc ttgtacgata ccacccggtt gtacctggtg gacaacaagt 12960

cggcggacat cgcctcgctg aactaccaga acgaccacag caacttcctg accaccgtgg 13020

tgcagaacaa tgacttcacc cccacggagg ccagcaccca gaccatcaac tttgacgagc 13080

gctcgcggtg gggcggccag ctgaaaacca tcatgcacac caacatgccc aacgtgaacg 13140

agttcatgta cagcaacaag ttcaaggcgc gggtgatggt ctcccgcaag acccccaatg 13200

gggtgacagt gacagaggat tatgatggta gtcaggatga gctgaagtat gaatgggtgg 13260

aatttgagct gcccgaaggc aacttctcgg tgaccatgac catcgacctg atgaacaacg 13320

ccatcatcga caattacttg gcggtggggc ggcagaacgg ggtgctggag agcgacatcg 13380

gcgtgaagtt cgacactagg aacttcaggc tgggctggga ccccgtgacc gagctggtca 13440

tgcccggggt gtacaccaac gaggctttcc atcccgatat tgtcttgctg cccggctgcg 13500

gggtggactt caccgagagc cgcctcagca acctgctggg cattcgcaag aggcagccct 13560

tccaggaagg cttccagatc atgtacgagg atctggaggg gggcaacatc cccgcgctcc 13620

tggatgtcga cgcctatgag aaaagcaagg aggatgcagc agctgaagca actgcagccg 13680

tagctaccgc ctctaccgag gtcaggggcg ataattttgc aagcgccgca gcagtggcag 13740

cggccgaggc ggctgaaacc gaaagtaaga tagtcattca gccggtggag aaggatagca 13800

agaacaggag ctacaacgta ctaccggaca agataaacac cgcctaccgc agctggtacc 13860

tagcctacaa ctatggcgac cccgagaagg gcgtgcgctc ctggacgctg ctcaccacct 13920

cggacgtcac ctgcggcgtg gagcaagtct actggtcgct gcccgacatg atgcaagacc 13980

cggtcacctt ccgctccacg cgtcaagtta gcaactaccc ggtggtgggc gccgagctcc 14040

tgcccgtcta ctccaagagc ttcttcaacg agcaggccgt ctactcgcag cagctgcgcg 14100

[0125] ccttcacctc gcttacgcac gtcttcaacc gcttccccga gaaccagatc ctcgtccgcc 14160

cgcccgcgcc caccattacc accgtcagtg aaaacgttcc tgctctcaca gatcacggga 14220

ccctgccgct gcgcagcagt atccggggag tccagcgcgt gaccgttact gacgccagac 14280

gccgcacctg cccctacgtc tacaaggccc tgggcatagt cgcgccgcgc gtcctctcga 14340

gccgcacctt ctaaatgtcc attctcatct cgcccagtaa taacaccggt tggggcctgc 14400

gcgcgcccag caagatgtac ggaggcgctc gccaacgctc cacgcaacac cccgtgcgcg 14460

tgcgcgggca cttccgcgct ccctggggcg ccctcaaggg ccgcgtgcgg tcgcgcacca 14520

ccgtcgacga cgtgatcgac caggtggtgg ccgacgcgcg caactacacc cccgccgccg 14580

cgcccgtctc caccgtggac gccgtcatcg acagcgtggt ggccgacgcg cgccggtacg 14640

cccgcgccaa gagccggcgg cggcgcatcg cccggcggca ccggagcacc cccgccatgc 14700

gcgcggcgcg agccttgctg cgcagggcca ggcgcacggg acgcagggcc atgctcaggg 14760

cggccagacg cgcggcttca ggcgccagcg ccggcaggac ccggagacgc gcggccacgg 14820

cggcggcagc ggccatcgcc agcatgtccc gcccgcggcg agggaacgtg tactgggtgc 14880

gcgacgccgc caccggtgtg cgcgtgcccg tgcgcacccg ccccccctcgc acttgaagat 14940

gttcacttcg cgatgttgat gtgtcccagc ggcgaggagg atgtccaagc gcaaattcaa 15000

ggaagagatg ctccaggtca tcgcgcctga gatctacggc cctgcggtgg tgaaggagga 15060

aagaaagccc cgcaaaatca agcgggtcaa aaaggacaaa aaggaagaag aaagtgatgt 15120

ggacggattg gtggagtttg tgcgcgagtt cgcccccccgg cggcgcgtgc agtggcgcgg 15180

gcggaaggtg caaccggtgc tgagacccgg caccaccgtg gtcttcacgc ccggcgagcg 15240

ctccggcacc gcttccaagc gctcctacga cgaggtgtac ggggatgatg atattctgga 15300

gcaggcggcc gagcgcctgg gcgagtttgc ttacggcaag cgcagccgtt ccgcaccgaa 15360

ggaagaggcg gtgtccatcc cgctggacca cggcaacccc acgccgagcc tcaagcccgt 15420

gaccttgcag caggtgctgc cgaccgcggc gccgcgccgg gggttcaagc gcgagggcga 15480

ggatctgtac cccaccatgc agctgatggt gcccaagcgc cagaagctgg aagacgtgct 15540

ggagaccatg aaggtggacc cggacgtgca gcccgaggtc aaggtgcggc ccatcaagca 15600

[0126] ggtggccccg ggcctgggcg tgcagaccgt ggacatcaag attcccacgg agcccatgga 15660

aacgcagacc gagcccatga tcaagcccag caccagcacc atggaggtgc agacggatcc 15720

ctggatgcca tcggctccta gtcgaagacc ccggcgcaag tacggcgcgg ccagcctgct 15780

gatgcccaac tacgcgctgc atccttccat catccccacg ccgggctacc gcggcacgcg 15840

cttctaccgc ggtcatacca gcagccgccg ccgcaagacc accactcgcc gccgccgtcg 15900

ccgcaccgcc gctgcaacca cccctgccgc cctggtgcgg agagtgtacc gccgcggccg 15960

cgcacctctg accctgccgc gcgcgcgcta ccacccgagc atcgccattt aaactttcgc 16020

ctgctttgca gatcaatggc cctcacatgc cgccttcgcg ttcccattac gggctaccga 16080

ggaagaaaac cgcgccgtag aaggctggcg gggaacggga tgcgtcgcca ccaccaccgg 16140

cggcggcgcg ccatcagcaa gcggttgggg ggaggcttcc tgcccgcgct gatccccatc 16200

atcgccgcgg cgatcggggc gatccccggc attgcttccg tggcggtgca ggcctctcag 16260

cgccactgag acacacttgg aaacatcttg taataaacca atggactctg acgctcctgg 16320

tcctgtgatg tgttttcgta gacagatgga agacatcaat ttttcgtccc tggctccgcg 16380

acacggcacg cggccgttca tgggcacctg gagcgacatc ggcaccagcc aactgaacgg 16440

gggcgccttc aattggagca gtctctggag cgggcttaag aatttcgggt ccacgcttaa 16500

aacctatggc agcaaggcgt ggaacagcac cacagggcag gcgctgaggg ataagctgaa 16560

agagcagaac ttccagcaga aggtggtcga tgggctcgcc tcgggcatca acggggtggt 16620

ggacctggcc aaccaggccg tgcagcggca gatcaacagc cgcctggacc cggtgccgcc 16680

cgccggctcc gtggagatgc cgcaggtgga ggaggagctg cctcccctgg acaagcgggg 16740

cgagaagcga ccccgccccg atgcggagga gacgctgctg acgcacacgg acgagccgcc 16800

cccgtacgag gaggcggtga aactgggtct gcccaccacg cggcccatcg cgcccctggc 16860

caccggggtg ctgaaacccg aaaagcccgc gaccctggac ttgcctcctc cccagccttc 16920

ccgcccctct acagtggcta agcccctgcc gccggtggcc gtggcccgcg cgcgacccgg 16980

gggcaccgcc cgccctcatg cgaactggca gagcactctg aacagcatcg tgggtctggg 17040

agtgcagagt gtgaagcgcc gccgctgcta ttaaacctac cgtagcgctt aacttgcttg 17100

[0127] tctgtgtgtg tatgtattat gtcgccgccg ccgctgtcca ccagaaggag gagtgaagag 17160

gcgcgtcgcc gagttgcaag atggccaccc catcgatgct gccccagtgg gcgtacatgc 17220

acatcgccgg acaggacgct tcggagtacc tgagtccggg tctggtgcag tttgcccgcg 17280

ccacagacac ctacttcagt ctggggaaca agtttaggaa ccccacggtg gcgcccacgc 17340

acgatgtgac caccgaccgc agccagcggc tgacgctgcg cttcgtgccc gtggaccgcg 17400

aggacaacac ctactcgtac aaagtgcgct acacgctggc cgtgggcgac aaccgcgtgc 17460

tggacatggc cagcacctac tttgacatcc gcggcgtgct ggatcggggc cctagcttca 17520

aaccctactc cggcaccgcc tacaacagtc tggcccccaa gggagcaccc aacacttgtc 17580

agtggacata taaagccgat ggtgaaactg ccacagaaaa aacctataca tatggaaatg 17640

cacccgtgca gggcattaac atcacaaaag atggtattca acttggaact gacaccgatg 17700

atcagccaat ctacgcagat aaaacctatc agcctgaacc tcaagtgggt gatgctgaat 17760

ggcatgacat cactggtact gatgaaaagt atggaggcag agctcttaag cctgatacca 17820

aaatgaagcc ttgttatggt tcttttgcca agcctactaa taaagaagga ggtcaggcaa 17880

atgtgaaaac aggaacaggc actactaaag aatatgacat agacatggct ttctttgaca 17940

acagaagtgc ggctgctgct ggcctagctc cagaaattgt tttgtatact gaaaatgtgg 18000

atttggaaac tccagatacc catattgtat acaaagcagg cacagatgac agcagctctt 18060

ctattaattt gggtcagcaa gccatgccca acagacctaa ctacattggt ttcagagaca 18120

actttatcgg gctcatgtac tacaacagca ctggcaatat gggggtgctg gccggtcagg 18180

cttctcagct gaatgctgtg gttgacttgc aagacagaaa caccgagctg tcctaccagc 18240

tcttgcttga ctctctgggt gacagaaccc ggtatttcag tatgtggaat caggcggtgg 18300

acagctatga tcctgatgtg cgcattattg aaaatcatgg tgtggaggat gaacttccca 18360

actattgttt ccctctggat gctgttggca gaacagatac ttatcaggga attaaggcta 18420

atggaactga tcaaaccaca tggaccaaag atgacagtgt caatgatgct aatgagatag 18480

gcaagggtaa tccattcgcc atggaaatca acatccaagc caacctgtgg aggaacttcc 18540

tctacgccaa cgtggccctg tacctgcccg actcttacaa gtacacgccg gccaatgtta 18600

[0128] ccctgcccac caacaccaac acctacgatt acatgaacgg ccgggtggtg gcgccctcgc 18660

tggtggactc ctacatcaac atcggggcgc gctggtcgct ggatcccatg gacaacgtga 18720

acccccttcaa ccaccaccgc aatgcggggc tgcgctaccg ctccatgctc ctgggcaacg 18780

ggcgctacgt gcccttccac atccaggtgc cccagaaatt tttcgccatc aagagcctcc 18840

tgctcctgcc cgggtcctac acctacgagt ggaacttccg caaggacgtc aacatgatcc 18900

tgcagagctc cctcggcaac gacctgcgca cggacggggc ctccatctcc ttcaccagca 18960

tcaacctcta cgccaccttc ttccccatgg cgcacaacac ggcctccacg ctcgaggcca 19020

tgctgcgcaa cgacaccaac gaccagtcct tcaacgacta cctctcggcg gccaacatgc 19080

tctaccccat cccggccaac gccaccaacg tgcccatctc catcccctcg cgcaactggg 19140

ccgccttccg cggctggtcc ttcacgcgtc tcaagaccaa ggagacgccc tcgctgggct 19200

ccgggttcga cccctacttc gtctactcgg gctccatccc ctacctcgac ggcaccttct 19260

acctcaacca caccttcaag aaggtctcca tcaccttcga ctcctccgtc agctggcccg 19320

gcaacgaccg gctcctgacg cccaacgagt tcgaaatcaa gcgcaccgtc gacggcgagg 19380

gctacaacgt ggcccagtgc aacatgacca aggactggtt cctggtccag atgctggccc 19440

actacaacat cggctaccag ggcttctacg tgcccgaggg ctacaaggac cgcatgtact 19500

ccttcttccg caacttccag cccatgagcc gccaggtggt ggacgaggtc aactacaagg 19560

actaccaggc cgtcaccctg gcctaccagc acaacaactc gggcttcgtc ggctacctcg 19620

cgcccaccat gcgccagggc cagccctacc ccgccaacta cccctacccg ctcatcggca 19680

agagcgccgt caccagcgtc acccagaaaa agttcctctg cgacagggtc atgtggcgca 19740

tccccttctc cagcaacttc atgtccatgg gcgcgctcac cgacctcggc cagaacatgc 19800

tctatgccaa ctccgcccac gcgctagaca tgaatttcga agtcgacccc atggatgagt 19860

ccacccttct ctatgttgtc ttcgaagtct tcgacgtcgt ccgagtgcac cagccccacc 19920

gcggcgtcat cgaggccgtc tacctgcgca ccccttctc ggccggtaac gccaccacct 19980

aagctcttgc ttcttgcaag ccatggccgc gggctccggc gagcaggagc tcagggccat 20040

catccgcgac ctgggctgcg ggccctactt cctgggcacc ttcgataagc gcttcccggg 20100

[0129] attcatggcc ccgcacaagc tggcctgcgc catcgtcaac acggccggcc gcgagaccgg 20160

gggcgagcac tggctggcct tcgcctggaa cccgcgctcg aacacctgct acctcttcga 20220

ccccttcggg ttctcggacg agcgcctcaa gcagatctac cagttcgagt acgagggcct 20280

gctgcgccgc agcgccctgg ccaccgagga ccgctgcgtc accctggaaa agtccaccca 20340

gaccgtgcag ggtccgcgct cggccgcctg cgggctcttc tgctgcatgt tcctgcacgc 20400

cttcgtgcac tggcccgacc gccccatgga caagaacccc accatgaact tgctgacggg 20460

ggtgcccaac ggcatgctcc agtcgcccca ggtggaaccc accctgcgcc gcaaccagga 20520

ggcgctctac cgcttcctca actcccactc cgcctacttt cgctcccacc gcgcgcgcat 20580

cgagaaggcc accgccttcg accgcatgaa tcaagacatg taaaccgtgt gtgtatgtta 20640

aatgtcttta ataaacagca ctttcatgtt acacatgcat ctgagatgat ttatttagaa 20700

atcgaaaggg ttctgccggg tctcggcatg gcccgcgggc agggacacgt tgcggaactg 20760

gtacttggcc agccacttga actcggggat cagcagtttg ggcagcgggg tgtcggggaa 20820

ggagtcggtc cacagcttcc gcgtcagttg cagggcgccc agcaggtcgg gcgcggagat 20880

cttgaaatcg cagttgggac ccgcgttctg cgcgcgggag ttgcggtaca cggggttgca 20940

gcactggaac accatcaggg ccgggtgctt cacgctcgcc agcaccgtcg cgtcggtgat 21000

gctctccacg tcgaggtcct cggcgttggc catcccgaag ggggtcatct tgcaggtctg 21060

ccttcccatg gtgggcacgc acccgggctt gtggttgcaa tcgcagtgca gggggatcag 21120

catcatctgg gcctggtcgg cgttcatccc cgggtacatg gccttcatga aagcctccaa 21180

ttgcctgaac gcctgctggg ccttggctcc ctcggtgaag aagaccccgc aggacttgct 21240

agagaactgg ttggtggcgc acccggcgtc gtgcacgcag cagcgcgcgt cgttgttggc 21300

cagctgcacc acgctgcgcc cccagcggtt ctgggtgatc ttggcccggt cggggttctc 21360

cttcagcgcg cgctgcccgt tctcgctcgc cacatccatc tcgatcatgt gctccttctg 21420

gatcatggtg gtcccgtgca ggcaccgcag cttgccctcg gcctcggtgc acccgtgcag 21480

ccacagcgcg cacccggtgc actcccagtt cttgtgggcg atctgggaat gcgcgtgcac 21540

gaagccctgc aggaagcggc ccatcatggt ggtcagggtc ttgttgctag tgaaggtcag 21600

[0130] cggaatgccg cggtgctcct cgttgatgta caggtggcag atgcggcggt acacctcgcc 21660

ctgctcgggc atcagctgga agttggcttt caggtcggtc tccacgcggt agcggtccat 21720

cagcatagtc atgatttcca taccttctc ccaggccgag acgatgggca ggctcatagg 21780

gttcttcacc atcatcttag cgctagcagc cgcggccagg gggtcgctct cgtccagggt 21840

ctcaaagctc cgcttgccgt ccttctcggt gatccgcacc ggggggtagc tgaagcccac 21900

ggccgccagc tcctcctcgg cctgtctttc gtcctcgctg tcctggctga cgtcctgcag 21960

gaccacatgc ttggtcttgc ggggtttctt cttgggcggc agcggcggcg gagatgttgg 22020

agatggcgag ggggagcgcg agttctcgct caccactact atctcttcct cttcttggtc 22080

cgaggccacg cggcggtagg tatgtctctt cgggggcaga ggcggaggcg acgggctctc 22140

gccgccgcga cttggcggat ggctggcaga gcccccttccg cgttcggggg tgcgctcccg 22200

gcggcgctct gactgacttc ctccgcggcc ggccattgtg ttctcctagg gaggaacaac 22260

aagcatggag actcagccat cgccaacctc gccatctgcc cccaccgccg acgagaagca 22320

gcagcagcag aatgaaagct taaccgcccc gccgcccagc cccgccacct ccgacgcggc 22380

cgtcccagac atgcaagaga tggaggaatc catcgagatt gacctgggct atgtgacgcc 22440

cgcggagcac gaggaggagc tggcagtgcg cttttcacaa gaagagatac accaagaaca 22500

gccagagcag gaagcagaga atgagcagag tcaggctggg ctcgagcatg acggcgacta 22560

cctccacctg agcgggggggg aggacgcgct catcaagcat ctggcccggc aggccaccat 22620

cgtcaaggat gcgctgctcg accgcaccga ggtgcccctc agcgtggagg agctcagccg 22680

cgcctacgag ttgaacctct tctcgccgcg cgtgcccccc aagcgccagc ccaatggcac 22740

ctgcgagccc aacccgcgcc tcaacttcta cccggtcttc gcggtgcccg aggccctggc 22800

cacctaccac atctttttca agaaccaaaa gatccccgtc tcctgccgcg ccaaccgcac 22860

ccgcgccgac gccctttttca acctgggtcc cggcgcccgc ctacctgata tcgcctcctt 22920

ggaagaggtt cccaagatct tcgagggtct gggcagcgac gagactcggg ccgcgaacgc 22980

tctgcaagga gaaggaggag agcatgagca ccacagcgcc ctggtcgagt tggaaggcga 23040

caacgcgcgg ctggcggtgc tcaaacgcac ggtcgagctg acccatttcg cctacccggc 23100

[0131] tctgaacctg ccccccaaag tcatgagcgc ggtcatggac caggtgctca tcaagcgcgc 23160

gtcgcccatc tccgaggacg agggcatgca agactccgag gagggcaagc ccgtggtcag 23220

cgacgagcag ctggcccggt ggctgggtcc taatgctagt ccccagagtt tggaagagcg 23280

gcgcaaactc atgatggccg tggtcctggt gaccgtggag ctggagtgcc tgcgccgctt 23340

cttcgccgac gcggagaccc tgcgcaaggt cgaggagaac ctgcactacc tcttcaggca 23400

cgggttcgtg cgccaggcct gcaagatctc caacgtggag ctgaccaacc tggtctccta 23460

catgggcatc ttgcacgaga accgcctggg gcagaacgtg ctgcacacca ccctgcgcgg 23520

ggaggcccgg cgcgactaca tccgcgactg cgtctacctc tacctctgcc acacctggca 23580

gacgggcatg ggcgtgtggc agcagtgtct ggaggagcag aacctgaaag agctctgcaa 23640

gctcctgcag aagaacctca agggtctgtg gaccgggttc gacgagcgca ccaccgcctc 23700

ggacctggcc gacctcattt tccccgagcg cctcaggctg acgctgcgca acggcctgcc 23760

cgactttatg agccaaagca tgttgcaaaa ctttcgctct ttcatcctcg aacgctccgg 23820

aatcctgccc gccacctgct ccgcgctgcc ctcggacttc gtgccgctga ccttccgcga 23880

gtgccccccg ccgctgtgga gccactgcta cctgctgcgc ctggccaact acctggccta 23940

ccactcggac gtgatcgagg acgtcagcgg cgagggcctg ctcgagtgcc actgccgctg 24000

caacctctgc acgccgcacc gctccctggc ctgcaacccc cagctgctga gcgagaccca 24060

gatcatcggc accttcgagt tgcaagggcc cagcgaaggc gagggttcag ccgccaaggg 24120

gggtctgaaa ctcaccccgg ggctgtggac ctcggcctac ttgcgcaagt tcgtgcccga 24180

ggactaccat cccttcgaga tcaggttcta cgaggaccaa tcccatccgc ccaaggccga 24240

gctgtcggcc tgcgtcatca cccagggggc gatcctggcc caattgcaag ccatccagaa 24300

atcccgccaa gaattcttgc tgaaaaaggg ccgcggggtc tacctcgacc cccagaccgg 24360

tgaggagctc aaccccggct tcccccagga tgccccgagg aaacaagaag ctgaaagtgg 24420

agctgccgcc cgtggaggat ttggaggaag actgggagaa cagcagtcag gcagaggagg 24480

aggagatgga ggaagactgg gacagcactc aggcagagga ggacagcctg caagacagtc 24540

tggaggaaga cgaggaggag gcagaggagg aggtggaaga agcagccgcc gccagaccgt 24600

[0132] cgtcctcggc gggggagaaa gcaagcagca cggataccat ctccgctccg ggtcggggtc 24660

ccgctcgacc acacagtaga tgggacgaga ccggacgatt cccgaacccc accacccaga 24720

ccggtaagaa ggagcggcag ggatacaagt cctggcgggg gcacaaaaac gccatcgtct 24780

cctgcttgca ggcctgcggg ggcaacatct ccttcacccg gcgctacctg ctcttccacc 24840

gcggggtgaa ctttccccgc aacatcttgc attactaccg tcacctccac agcccctact 24900

acttccaaga agaggcagca gcagcagaaa aagaccagca gaaaaccagc agctagaaaa 24960

tccacagcgg cggcagcagg tggactgagg atcgcggcga acgagccggc gcaaacccgg 25020

gagctgagga accggatctt tcccaccctc tatgccatct tccagcagag tcgggggcag 25080

gagcaggaac tgaaagtcaa gaaccgttct ctgcgctcgc tcacccgcag ttgtctgtat 25140

cacaagagcg aagaccaact tcagcgcact ctcgaggacg ccgaggctct cttcaacaag 25200

tactgcgcgc tcactcttaa agagtagccc gcgcccgccc agtcgcagaa aaaggcggga 25260

attacgtcac ctgtgccctt cgccctagcc gcctccaccc atcatcatga gcaaagagat 25320

tcccacgcct tacatgtgga gctaccagcc ccagatgggc ctggccgccg gtgccgccca 25380

ggactactcc acccgcatga attggctcag cgccgggccc gcgatgatct cacgggtgaa 25440

tgacatccgc gcccaccgaa accagatact cctagaacag tcagcgctca ccgccacgcc 25500

ccgcaatcac ctcaatccgc gtaattggcc cgccgccctg gtgtaccagg aaattcccca 25560

gcccacgacc gtactacttc cgcgagacgc ccaggccgaa gtccagctga ctaactcagg 25620

tgtccagctg gcgggcggcg ccaccctgtg tcgtcaccgc cccgctcagg gtataaagcg 25680

gctggtgatc cggggcagag gcacacagct caacgacgag gtggtgagct cttcgctggg 25740

tctgcgacct gacggagtct tccaactcgc cggatcgggg agatcttcct tcacgcctcg 25800

tcaggccgtc ctgactttgg agagttcgtc ctcgcagccc cgctcgggtg gcatcggcac 25860

tctccagttc gtggaggagt tcactccctc ggtctacttc aacccccttct ccggctcccc 25920

cggccactac ccggacgagt tcatcccgaa cttcgacgcc atcagcgagt cggtggacgg 25980

ctacgattga atgtcccatg gtggcgcagc tgacctagct cggcttcgac acctggacca 26040

ctgccgccgc ttccgctgct tcgctcggga tctcgccgag tttgcctact ttgagctgcc 26100

[0133] cgaggagcac cctcagggcc cggcccacgg agtgcggatc gtcgtcgaag ggggcctcga 26160

ctcccacctg cttcggatct tcagccagcg tccgatcctg gtcgagcgcg agcaaggaca 26220

gacccttctg actctgtact gcatctgcaa ccaccccggc ctgcatgaaa gtctttgttg 26280

tctgctgtgt actgagtata ataaaagctg agatcagcga ctactccgga cttccgtgtg 26340

ttcctgaatc catcaaccag tctttgttct tcaccgggaa cgagaccgag ctccagctcc 26400

agtgtaagcc ccacaagaag tacctcacct ggctgttcca gggctccccg atcgccgttg 26460

tcaaccactg cgacaacgac ggagtcctgc tgagcggccc tgccaacctt acttttttcca 26520

cccgcagaag caagctccag ctcttccaac ccttcctccc cgggacctat cagtgcgtct 26580

cgggaccctg ccatcacacc ttccacctga tcccgaatac cacagcgtcg ctccccgcta 26640

ctaacaacca aactaacctc caccaacgcc accgtcgcga cggccacaat acatgcccat 26700

attagactat gaggccgagc cacagcgacc catgctcccc gctattagtt acttcaatct 26760

aaccggcgga gatgactgac ccactggcca acaacaacgt caacgacctt ctcctggaca 26820

tggacggccg cgcctcggag cagcgactcg cccaacttcg cattcgccag cagcaggaga 26880

gagccgtcaa ggagctgcag gatgcggtgg ccatccacca gtgcaagaga ggcatcttct 26940

gcctggtgaa acaggccaag atctcctacg aggtcactcc aaacgaccat cgcctctcct 27000

acgagctcct gcagcagcgc cagaagttca cctgcctggt cggagtcaac cccatcgtca 27060

tcacccagca gtctggcgat accaaggggt gcatccactg ctcctgcgac tcccccgact 27120

gcgtccacac tctgatcaag accctctgcg gcctccgcga cctcctcccc atgaactaat 27180

caccccctta tccagtgaaa taaagatcat attgatgatg attttacaga aataaaaaat 27240

aatcatttga tttgaaataa agatacaatc atattgatga tttgagttta acaaaaaaat 27300

aaagaatcac ttacttgaaa tctgatacca ggtctctgtc catgttttct gccaacacca 27360

cttcactccc ctcttcccag ctctggtact gcaggccccg gcgggctgca aacttcctcc 27420

acacgctgaa ggggatgtca aattcctcct gtccctcaat cttcatttta tcttctatca 27480

gatgtccaaa aagcgcgtcc gggtggatga tgacttcgac cccgtctacc cctacgatgc 27540

agacaacgca ccgaccgtgc ccttcatcaa ccccccccttc gtctcttcag atggattcca 27600

[0134] agagaagccc ctgggggtgt tgtccctgcg actggccgac cccgtcacca ccaagaacgg 27660

ggaaatcacc ctcaagctgg gagaggggggt ggacctcgat tcctcgggaa aactcatctc 27720

caacacggcc accaaggccg ccgcccctct cagtttttcc aacaacacca tttcccttaa 27780

catggatcac ccctttttaca ctaaagatgg aaaattatcc ttacaagttt ctccaccatt 27840

aaatatactg agaacaagca ttctaaacac actagcttta ggttttggat caggtttagg 27900

actccgtggc tctgccttgg cagtacagtt agtctctcca cttacatttg atactgatgg 27960

aaacataaag cttaccttag acagaggttt gcatgttaca acaggagatg caattgaaag 28020

caacataagc tgggctaaag gtttaaaatt tgaagatgga gccatagcaa ccaacattgg 28080

aaatgggtta gagtttggaa gcagtagtac agaaacaggt gttgatgatg cttacccaat 28140

ccaagttaaa cttggatctg gccttagctt tgacagtaca ggagccataa tggctggtaa 28200

caaagaagac gataaactca ctttgtggac aacacctgat ccatcaccaa actgtcaaat 28260

actcgcagaa aatgatgcaa aactaacact ttgcttgact aaatgtggta gtcaaatact 28320

ggccactgtg tcagtcttag ttgtaggaag tggaaaccta aaccccatta ctggcaccgt 28380

aagcagtgct caggtgtttc tacgttttga tgcaaacggt gttcttttaa cagaacattc 28440

tacactaaaa aaatactggg ggtataggca gggagatagc atagatggca ctccatatac 28500

caatgctgta ggattcatgc ccaatttaaa agcttatcca aagtcacaaa gttctactac 28560

taaaaataat atagtagggc aagtatacat gaatggagat gtttcaaaac ctatgcttct 28620

cactataacc ctcaatggta ctgatgacag caacagtaca tattcaatgt cattttcata 28680

cacctggact aatggaagct atgttggagc aacatttggg gctaactctt ataccttctc 28740

atacatcgcc caagaatgaa cactgtatcc caccctgcat gccaaccctt cccaccccac 28800

tctgtggaac aaactctgaa acacaaaata aaataaagtt caagtgtttt attgattcaa 28860

cagttttaca ggattcgagc agttattttt cctccaccct cccaggacat ggaatacacc 28920

accctctccc cccgcacagc cttgaacatc tgaatgccat tggtgatgga catgcttttg 28980

gtctccacgt tccacacagt ttcagagcga gccagtctcg ggtcggtcag ggagatgaaa 29040

ccctccgggc actcccgcat ctgcacctca cagctcaaca gctgaggatt gtcctcggtg 29100

[0135] gtcgggatca cggttatctg gaagaagcag aagagcggcg gtgggaatca tagtccgcga 29160

acgggatcgg ccggtggtgt cgcatcaggc cccgcagcag tcgctgccgc cgccgctccg 29220

tcaagctgct gctcaggggg tccgggtcca gggactccct cagcatgatg cccacggccc 29280

tcagcatcag tcgtctggtg cggcgggcgc agcagcgcat gcggatctcg ctcaggtcgc 29340

tgcagtacgt gcaacacaga accaccaggt tgttcaacag tccatagttc aacacgctcc 29400

agccgaaact catcgcggga aggatgctac ccacgtggcc gtcgtaccag atcctcaggt 29460

aaatcaagtg gtgccccctc cagaacacgc tgcccacgta catgatctcc ttgggcatgt 29520

ggcggttcac cacctcccgg taccacatca ccctctggtt gaacatgcag ccccggatga 29580

tcctgcggaa ccacagggcc agcaccgccc cgcccgccat gcagcgaaga gaccccgggt 29640

cccggcaatg gcaatggagg acccaccgct cgtacccgtg gatcatctgg gagctgaaca 29700

agtctatgtt ggcacagcac aggcatatgc tcatgcatct cttcagcact ctcaactcct 29760

cgggggtcaa aaccatatcc cagggcacgg ggaactcttg caggacagcg aaccccgcag 29820

aacagggcaa tcctcgcaca gaacttacat tgtgcatgga cagggtatcg caatcaggca 29880

gcaccgggtg atcctccacc agagaagcgc gggtctcggt ctcctcacag cgtggtaagg 29940

gggccggccg atacgggtga tggcgggacg cggctgatcg tgttcgcgac cgtgtcatga 30000

tgcagttgct ttcggacatt ttcgtacttg ctgtagcaga acctggtccg ggcgctgcac 30060

accgatcgcc ggcggcggtc tcggcgcttg gaacgctcgg tgttgaaatt gtaaaacagc 30120

cactctctca gaccgtgcag cagatctagg gcctcaggag tgatgaagat cccatcatgc 30180

ctgatggctc tgatcacatc gaccaccgtg gaatgggcca gacccagcca gatgatgcaa 30240

ttttgttggg tttcggtgac ggcgggggag ggaagaacag gaagaaccat gattaacttt 30300

taatccaaac ggtctcggag tacttcaaaa tgaagatcgc ggagatggca cctctcgccc 30360

ccgctgtgtt ggtggaaaat aacagccagg tcaaaggtga tacggttctc gagatgttcc 30420

acggtggctt ccagcaaagc ctccacgcgc acatccagaa acaagacaat agcgaaagcg 30480

ggagggttct ctaattcctc aatcatcatg ttacactcct gcaccatccc cagataattt 30540

tcatttttcc agccttgaat gattcgaact agttcgtgag gtaaatccaa gccagccatg 30600

[0136] ataaagagct cgcgcagagc gccctccacc ggcattctta agcacaccct cataattcca 30660

agatattctg ctcctggttc acctgcagca gattgacaag cggaatatca aaatctctgc 30720

cgcgatccct gagctcctcc ctcagcaata actgtaagta ctctttcata tcctctccga 30780

aatttttagc cataggacca ccaggaataa gattagggca agccacagta cagataaacc 30840

gaagtcctcc ccagtgagca ttgccaaatg caagactgct ataagcatgc tggctagacc 30900

cggtgatatc ttccagataa ctggacagaa aatcgcccag gcaattttta agaaaatcaa 30960

caaaagaaaa atcctccagg tggacgttta gagcctcggg aacaacgatg aagtaaatgc 31020

aagcggtgcg ttccagcatg gttagttagc tgatctgtag aaaaaacaaa aatgaacatt 31080

aaaccatgct agcctggcga acaggtgggt aaatcgttct ctccagcacc aggcaggcca 31140

cggggtctcc ggcgcgaccc tcgtaaaaat tgtcgctatg attgaaaacc atcacagaga 31200

gacgttcccg gtggccggcg tgaatgattc gacaagatga atacaccccc ggaacattgg 31260

cgtccgcgag tgaaaaaaag cgcccgagga agcaataagg cactacaatg ctcagtctca 31320

agtccagcaa agcgatgcca tgcggatgaa gcacaaaatt ctcaggtgcg tacaaaatgt 31380

aattactccc ctcctgcaca ggcagcaaag cccccgatcc ctccaggtac acatacaaag 31440

cctcagcgtc catagcttac cgagcagcag cacacaacag gcgcaagagt cagagaaagg 31500

ctgagctcta acctgtccac ccgctctctg ctcaatatat agcccagatc tacactgacg 31560

taaaggccaa agtctaaaaa tacccgccaa ataatcacac acgcccagca cacgcccaga 31620

aaccggtgac acactcaaaa aaatacgcgc acttcctcaa acgcccaaaa ctgccgtcat 31680

ttccgggttc ccacgctacg tcatcaaaac acgactttca aattccgtcg accgttaaaa 31740

acgtcacccg ccccgcccct aacggtcgcc cgtctctcag ccaatcagcg ccccgcatcc 31800

ccaaattcaa acacctcatt tgcatattaa cgcgcacaaa aagtttgagg tatattattg 31860

atgatgg 31867

<210> 12
<211> 32788
<212> DNA
<213> 人工序列(Artificial Sequence)

[0137] <220>
<223> 人工序列的描述：合成多核苷酸

<400> 12
ccatcttcaa taatatacct caaacttttt gtgcgcgtta atatgcaaat gaggcgtttg 60

aatttgggga ggaagggcgg tgattggtcg agggatgagc gaccgttagg ggcggggcga 120

gtgacgtttt gatgacgtgg ttgcgaggag gagccagttt gcaagttctc gtgggaaaag 180

tgacgtcaaa cgaggtgtgg tttgaacacg gaaatactca attttcccgc gctctctgac 240

aggaaatgag gtgtttctgg gcggatgcaa gtgaaaacgg gccattttcg cgcgaaaact 300

gaatgaggaa gtgaaaatct gagtaatttc gcgtttatgg cagggaggag tatttgccga 360

gggccgagta gactttgacc gattacgtgg gggtttcgat taccgtgttt ttcacctaaa 420

tttccgcgta cggtgtcaaa gtccggtgtt tttacgtagg tgtcagctga tcgccagggt 480

atttaaacct gcgctctcca gtcaagaggc cactcttgag tgccagcgag aagagttttc 540

tcctccgcgc cgcgagtcag atctacactt tgaaagtagg gataacaggg taatgacatt 600

gattattgac tagttgttaa tagtaatcaa ttacggggtc attagttcat agcccatata 660

tggagttccg cgttacataa cttacggtaa atggcccgcc tggctgaccg cccaacgacc 720

cccgcccatt gacgtcaata atgacgtatg ttcccatagt aacgccaata gggactttcc 780

attgacgtca atgggtggag tatttacggt aaactgccca cttggcagta catcaagtgt 840

atcatatgcc aagtccgccc cctattgacg tcaatgacgg taaatggccc gcctggcatt 900

atgcccagta catgacctta cgggactttc ctacttggca gtacatctac gtattagtca 960

tcgctattac catggtgatg cggttttggc agtacaccaa tgggcgtgga tagcggtttg 1020

actcacgggg atttccaagt ctccacccca ttgacgtcaa tgggagtttg ttttggcacc 1080

aaaatcaacg ggactttcca aaatgtcgta ataaccccgc cccgttgacg caaatgggcg 1140

gtaggcgtgt acggtgggag gtctatataa gcagagctcg tttagtgaac cgtcagatcg 1200

cctggaacgc catccacgct gttttgacct ccatagaaga cagcgatcgc gccaccatgg 1260

ccgggatgtt ccaggcactg tccgaaggct gcacacccta tgatattaac cagatgctga 1320

atgtcctggg agaccaccag gtctctggcc tggagcagct ggagagcatc atcaacttcg 1380

[0138] agaagctgac cgagtggaca agctccaatg tgatgcctat cctgtcccca ctgaccaagg 1440

gcatcctggg cttcgtgttt accctgacag tgccttctga gcggggcctg tcttgcatca 1500

gcgaggcaga cgcaaccaca ccagagtccg ccaatctggg cgaggagatc ctgtctcagc 1560

tgtacctgtg gccccgggtg acatatcact ccccttctta cgcctatcac cagttcgagc 1620

ggagagccaa gtacaagaga cacttcccag gctttggcca gtctctgctg ttcggctacc 1680

ccgtgtacgt gttcggcgat tgcgtgcagg gcgactggga tgccatccgg tttagatact 1740

gcgcaccacc tggatatgca ctgctgaggt gtaacgacac caattattcc gccctgctgg 1800

cagtgggcgc cctggagggc cctcgcaatc aggattggct gggcgtgcca aggcagctgg 1860

tgacacgcat gcaggccatc cagaacgcag gcctgtgcac cctggtggca atgctggagg 1920

agacaatctt ctggctgcag gcctttctga tggccctgac cgacagcggc cccaagacaa 1980

acatcatcgt ggattcccag tacgtgatgg gcatctccaa gccttctttc caggagtttg 2040

tggactggga gaacgtgagc ccagagctga attccaccga tcagccattc tggcaggcag 2100

gaatcctggc aaggaacctg gtgcctatgg tggccacagt gcagggccag aatctgaagt 2160

accagggcca gagcctggtc atcagcgcct ccatcatcgt gtttaacctg ctggagctgg 2220

agggcgacta tcgggacgat ggcaacgtgt gggtgcacac cccactgagc cccagaacac 2280

tgaacgcctg ggtgaaggcc gtggaggaga agaagggcat cccagtgcac ctggagctgg 2340

cctccatgac caatatggag ctgatgtcta gcatcgtgca ccagcaggtg aggacatacg 2400

gacccgtgtt catgtgcctg ggaggcctgc tgaccatggt ggcaggagcc gtgtggctga 2460

cagtgcgggt gctggagctg ttcagagccg cccagctggc caacgatgtg gtgctgcaga 2520

tcatggagct gtgcggagca gcctttcgcc aggtgtgcca caccacagtg ccatggccca 2580

atgcctccct gacccccaag tggaacaatg agacaacaca gcctcagatc gccaactgta 2640

gcgtgtacga cttcttcgtg tggctgcact actatagcgt gagggatacc ctgtggcccc 2700

gcgtgacata ccacatgaat aagtacgcct atcacatgct ggagaggcgc gccaagtata 2760

agagaggccc tggcccaggc gcaaagtttg tggcagcatg gaccctgaag gccgccgccg 2820

gccccggccc cggccagtat atcaaggcta acagtaagtt cattggaatc acagagctgg 2880

[0139] gacccggacc tggataatga gtttaaactc ccatttaaat gtgagggtta atgcttcgag 2940

cagacatgat aagatacatt gatgagtttg gacaaaccac aactagaatg cagtgaaaaa 3000

aatgctttat ttgtgaaatt tgtgatgcta ttgctttatt tgtaaccatt ataagctgca 3060

ataaacaagt taacaacaac aattgcattc attttatgtt tcaggttcag ggggagatgt 3120

gggaggtttt ttaaagcaag taaaacctct acaaatgtgg taaaataact ataacggtcc 3180

taaggtagcg agtgagtagt gttctggggc gggggaggac ctgcatgagg gccagaataa 3240

ctgaaatctg tgcttttctg tgtgttgcag cagcatgagc ggaagcggct cctttgaggg 3300

aggggtattc agcccttatc tgacggggcg tctcccctcc tgggcgggag tgcgtcagaa 3360

tgtgatggga tccacggtgg acggccggcc cgtgcagccc gcgaactctt caaccctgac 3420

ctatgcaacc ctgagctctt cgtcgttgga cgcagctgcc gccgcagctg ctgcatctgc 3480

cgccagcgcc gtgcgcggaa tggccatggg cgccggctac tacggcactc tggtggccaa 3540

ctcgagttcc accaataatc ccgccagcct gaacgaggag aagctgttgc tgctgatggc 3600

ccagctcgag gccttgaccc agcgcctggg cgagctgacc cagcaggtgg ctcagctgca 3660

ggagcagacg cgggccgcgg ttgccacggt gaaatccaaa taaaaaatga atcaataaat 3720

aaacggagac ggttgttgat tttaacacag agtctgaatc tttatttgat ttttcgcgcg 3780

cggtaggccc tggaccaccg gtctcgatca ttgagcaccc ggtggatctt ttccaggacc 3840

cggtagaggt gggcttggat gttgaggtac atgggcatga gcccgtcccg ggggtggagg 3900

tagctccatt gcagggcctc gtgctcgggg gtggtgttgt aaatcaccca gtcatagcag 3960

gggcgcaggg catggtgttg cacaatatct ttgaggagga gactgatggc cacgggcagc 4020

cctttggtgt aggtgtttac aaatctgttg agctgggagg gatgcatgcg gggggagatg 4080

aggtgcatct tggcctggat cttgagattg gcgatgttac cgcccagatc ccgcctgggg 4140

ttcatgttgt gcaggaccac cagcacggtg tatccggtgc acttggggaa tttatcatgc 4200

aacttggaag ggaaggcgtg aaagaatttg gcgacgcctt tgtgcccgcc caggttttcc 4260

atgcactcat ccatgatgat ggcgatgggc ccgtgggcgg cggcctgggc aaagacgttt 4320

cgggggtcgg acacatcata gttgtggtcc tgggtgaggt catcataggc cattttaatg 4380

[0140] aatttggggc ggagggtgcc ggactggggg acaaaggtac cctcgatccc gggggcgtag 4440

ttcccctcac agatctgcat ctcccaggct ttgagctcgg agggggggat catgtccacc 4500

tgcggggcga taaagaacac ggtttccggg gcgggggaga tgagctgggc cgaaagcaag 4560

ttccggagca gctgggactt gccgcagccg gtggggccgt agatgacccc gatgaccggc 4620

tgcaggtggt agttgaggga gagacagctg ccgtcctccc ggaggagggg ggccacctcg 4680

ttcatcatct cgcgcacgtg catgttctcg cgcaccagtt ccgccaggag gcgctctccc 4740

cccagggata ggagctcctg gagcgaggcg aagtttttca gcggcttgag tccgtcggcc 4800

atgggcattt tggagagggt ttgttgcaag agttccaggc ggtcccagag ctcggtgatg 4860

tgctctacgg catctcgatc cagcagacct cctcgtttcg cgggttggga cggctgcggg 4920

agtagggcac cagacgatgg gcgtccagcg cagccagggt ccggtccttc cagggtcgca 4980

gcgtccgcgt cagggtggtc tccgtcacgg tgaaggggtg cgcgccgggc tgggcgcttg 5040

cgagggtgcg cttcaggctc atccggctgg tcgaaaaccg ctcccgatcg gcgccctgcg 5100

cgtcggccag gtagcaattg accatgagtt cgtagttgag cgcctcggcc gcgtggcctt 5160

tggcgcggag cttacctttg gaagtctgcc cgcaggcggg acagaggagg gacttgaggg 5220

cgtagagctt gggggcgagg aagacggact cgggggcgta ggcgtccgcg ccgcagtggg 5280

cgcagacggt ctcgcactcc acgagccagg tgaggtcggg ctggtcgggg tcaaaaacca 5340

gtttcccgcc gttctttttg atgcgtttct tacctttggt ctccatgagc tcgtgtcccc 5400

gctgggtgac aaagaggctg tccgtgtccc cgtagaccga ctttatgggc cggtcctcga 5460

gcggtgtgcc gcggtcctcc tcgtagagga accccgccca ctccgagacg aaagcccggg 5520

tccaggccag cacgaaggag gccacgtggg acgggtagcg gtcgttgtcc accagcgggt 5580

ccacctttc cagggtatgc aaacacatgt ccccctcgtc cacatccagg aaggtgattg 5640

gcttgtaagt gtaggccacg tgaccggggg tcccggccgg gggggtataa aagggtgcgg 5700

gtccctgctc gtcctcactg tcttccggat cgctgtccag gagcgccagc tgttggggta 5760

ggtattccct ctcgaaggcg ggcatgacct cggcactcag gttgtcagtt tctagaaacg 5820

aggaggattt gatattgacg gtgccggcgg agatgccttt caagagcccc tcgtccatct 5880

[0141] ggtcagaaaa gacgatcttt ttgttgtcga gcttggtggc gaaggagccg tagagggcgt 5940

tggagaggag cttggcgatg gagcgcatgg tctggttttt ttccttgtcg gcgcgctcct 6000

tggcggcgat gttgagctgc acgtactcgc gcgccacgca cttccattcg gggaagacgg 6060

tggtcagctc gtcgggcacg attctgacct gccagccccg attatgcagg gtgatgaggt 6120

ccacactggt ggccacctcg ccgcgcaggg gctcattagt ccagcagagg cgtccgccct 6180

tgcgcgagca gaaggggggc aggggtcca gcatgacctc gtcgggggg tcggcatcga 6240

tggtgaagat gccgggcagg aggtcggggt caaagtagct gatggaagtg gccagatcgt 6300

ccagggcagc ttgccattcg cgcacggcca gcgcgctctc gtagggactg aggggcgtgc 6360

cccagggcat gggatgggta agcgcggagg cgtacatgcc gcagatgtcg tagacgtaga 6420

ggggctcctc gaggatgccg atgtaggtgg ggtagcagcg ccccccgcgg atgctggcgc 6480

gcacgtagtc atacagctcg tgcgaggggg cgaggagccc cgggcccagg ttggtgcgac 6540

tgggcttttc ggcgcggtag acgatctggc ggaaaatggc atgcgagttg gaggagatgg 6600

tgggcctttg gaagatgttg aagtgggcgt ggggcagtcc gaccgagtcg cggatgaagt 6660

gggcgtagga gtcttgcagc ttggcgacga gctcggcggt gactaggacg tccagagcgc 6720

agtagtcgag ggtctcctgg atgatgtcat acttgagctg tccctttgt ttccacagct 6780

cgcggttgag aaggaactct tcgcggtcct tccagtactc ttcgaggggg aacccgtcct 6840

gatctgcacg gtaagagcct agcatgtaga actggttgac ggccttgtag gcgcagcagc 6900

ccttctccac ggggagggcg taggcctggg cggccttgcg cagggaggtg tgcgtgaggg 6960

cgaaagtgtc cctgaccatg accttgagga actggtgctt gaagtcgata tcgtcgcagc 7020

ccccctgctc ccagagctgg aagtccgtgc gcttcttgta ggcggggttg ggcaaagcga 7080

aagtaacatc gttgaagagg atcttgcccg cgcggggcat aaagttgcga gtgatgcgga 7140

aaggttgggg cacctcggcc cggttgttga tgacctgggc ggcgagcacg atctcgtcga 7200

agccgttgat gttgtggccc acgatgtaga gttccacgaa tcgcggacgg cccttgacgt 7260

ggggcagttt cttgagctcc tcgtaggtga gctcgtcggg gtcgctgagc ccgtgctgct 7320

cgagcgccca gtcggcgaga tgggggttgg cgcggaggaa ggaagtccag agatccacgg 7380

[0142] ccagggcggt ttgcagacgg tcccggtact gacggaactg ctgcccgacg gccatttttt 7440

cgggggtgac gcagtagaag gtgcgggggt ccccgtgcca gcgatcccat ttgagctgga 7500

gggcgagatc gagggcgagc tcgacgagcc ggtcgtcccc ggagagtttc atgaccagca 7560

tgaaggggac gagctgcttg ccgaaggacc ccatccaggt gtaggtttcc acatcgtagg 7620

tgaggaagag cctttcggtg cgaggatgcg agccgatggg gaagaactgg atctcctgcc 7680

accaattgga ggaatggctg ttgatgtgat ggaagtagaa atgccgacgg cgcgccgaac 7740

actcgtgctt gtgtttatac aagcggccac agtgctcgca acgctgcacg ggatgcacgt 7800

gctgcacgag ctgtacctga gttcctttga cgaggaattt cagtgggaag tggagtcgtg 7860

gcgcctgcat ctcgtgctgt actacgtcgt ggtggtcggc ctggccctct tctgcctcga 7920

tggtggtcat gctgacgagc ccgcgcggga ggcaggtcca gacctcggcg cgagcgggtc 7980

ggagagcgag gacgagggcg cgcaggccgg agctgtccag ggtcctgaga cgctgcggag 8040

tcaggtcagt gggcagcggc ggcgcgcggt tgacttgcag gagtttttcc agggcgcgcg 8100

ggaggtccag atggtacttg atctccaccg cgccattggt ggcgacgtcg atggcttgca 8160

gggtcccgtg cccctggggt gtgaccaccg tcccccgttt cttcttgggc ggctggggcg 8220

acgggggcgg tgcctcttcc atggttagaa gcggcggcga ggacgcgcgc cgggcggcag 8280

gggcggctcg gggcccggag gcaggggcgg caggggcacg tcggcgccgc gcgcgggtag 8340

gttctggtac tgcgcccgga gaagactggc gtgagcgacg acgcgacggt tgacgtcctg 8400

gatctgacgc ctctgggtga aggccacggg acccgtgagt ttgaacctga aagagagttc 8460

gacagaatca atctcggtat cgttgacggc ggcctgccgc aggatctctt gcacgtcgcc 8520

cgagttgtcc tggtaggcga tctcggtcat gaactgctcg atctcctcct cttgaaggtc 8580

tccgcggccg gcgcgctcca cggtggccgc gaggtcgttg gagatgcggc ccatgagctg 8640

cgagaaggcg ttcatgcccg cctcgttcca gacgcggctg tagaccacga cgccctcggg 8700

atcgcgggcg cgcatgacca cctgggcgag gttgagctcc acgtggcgcg tgaagaccgc 8760

gtagttgcag aggcgctggt agaggtagtt gagcgtggtg gcgatgtgct cggtgacgaa 8820

gaaatacatg atccagcggc ggagcggcat ctcgctgacg tcgcccagcg cctccaaacg 8880

[0143] ttccatggcc tcgtaaaagt ccacggcgaa gttgaaaaac tgggagttgc gcgccgagac 8940

ggtcaactcc tcctccagaa gacggatgag ctcggcgatg gtggcgcgca cctcgcgctc 9000

gaaggccccc gggagttcct ccacttcctc ttcttcctcc tccactaaca tctcttctac 9060

ttcctcctca ggcggcagtg gtggcggggg agggggcctg cgtcgccggc ggcgcacggg 9120

cagacggtcg atgaagcgct cgatggtctc gccgcgccgg cgtcgcatgg tctcggtgac 9180

ggcgcgcccg tcctcgcggg gccgcagcgt gaagacgccg ccgcgcatct ccaggtggcc 9240

gggggggtcc ccgttgggca gggagagggc gctgacgatg catcttatca attgccccgt 9300

agggactccg cgcaaggacc tgagcgtctc gagatccacg ggatctgaaa accgctgaac 9360

gaaggcttcg agccagtcgc agtcgcaagg taggctgagc acggtttctt ctggcgggtc 9420

atgttggttg ggagcggggc gggcgatgct gctggtgatg aagttgaaat aggcggttct 9480

gagacggcgg atggtggcga ggagcaccag gtctttgggc ccggcttgct ggatgcgcag 9540

acggtcggcc atgccccagg cgtggtcctg acacctggcc aggtccttgt agtagtcctg 9600

catgagccgc tccacgggca cctcctcctc gcccgcgcgg ccgtgcatgc gcgtgagccc 9660

gaagccgcgc tggggctgga cgagcgccag gtcggcgacg acgcgctcgg cgaggatggc 9720

ttgctggatc tgggtgaggg tggtctggaa gtcatcaaag tcgacgaagc ggtggtaggc 9780

tccggtgttg atggtgtagg agcagttggc catgacggac cagttgacgg tctggtggcc 9840

cggacgcacg agctcgtggt acttgaggcg cgagtaggcg cgcgtgtcga agatgtagtc 9900

gttgcaggtg cgcaccaggt actggtagcc gatgaggaag tgcggcggcg gctggcggta 9960

gagcggccat cgctcggtgg cgggggcgcc gggcgcgagg tcctcgagca tggtgcggtg 10020

gtagccgtag atgtacctgg acatccaggt gatgccggcg gcggtggtgg aggcgcgcgg 10080

gaactcgcgg acgcggttcc agatgttgcg cagcggcagg aagtagttca tggtgggcac 10140

ggtctggccc gtgaggcgcg cgcagtcgtg gatgctctat acgggcaaaa acgaaagcgg 10200

tcagcggctc gactccgtgg cctggaggct aagcgaacgg gttgggctgc gcgtgtaccc 10260

cggttcgaat ctcgaatcag gctggagccg cagctaacgt ggtattggca ctcccgtctc 10320

gacccaagcc tgcaccaacc ctccaggata cggaggcggg tcgttttgca actttttttt 10380

[0144] ggaggccgga tgagactagt aagcgcggaa agcggccgac cgcgatggct cgctgccgta 10440

gtctggagaa gaatcgccag ggttgcgttg cggtgtgccc cggttcgagg ccggccggat 10500

tccgcggcta acgagggcgt ggctgccccg tcgtttccaa gaccccatag ccagccgact 10560

tctccagtta cggagcgagc ccctcttttg ttttgtttgt ttttgccaga tgcatcccgt 10620

actgcggcag atgcgccccc accaccctcc accgcaacaa cagccccctc cacagccggc 10680

gcttctgccc ccgccccagc agcaacttcc agccacgacc gccgcggccg ccgtgagcgg 10740

ggctggacag agttatgatc accagctggc cttggaagag ggcgaggggc tggcgcgcct 10800

gggggcgtcg tcgccggagc ggcacccgcg cgtgcagatg aaaagggacg ctcgcgaggc 10860

ctacgtgccc aagcagaacc tgttcagaga caggagcggc gaggagcccg aggagatgcg 10920

cgcggcccgg ttccacgcgg ggcgggagct gcggcgcggc ctggaccgaa agagggtgct 10980

gagggacgag gatttcgagg cggacgagct gacgggggatc agccccgcgc gcgcgcacgt 11040

ggccgcggcc aacctggtca cggcgtacga gcagaccgtg aaggaggaga gcaacttcca 11100

aaaatccttc aacaaccacg tgcgcaccct gatcgcgcgc gaggaggtga ccctgggcct 11160

gatgcacctg tgggacctgc tggaggccat cgtgcagaac cccaccagca agccgctgac 11220

ggcgcagctg ttcctggtgg tgcagcatag tcgggacaac gaagcgttca gggaggcgct 11280

gctgaatatc accgagcccg agggccgctg gctcctggac ctggtgaaca ttctgcagag 11340

catcgtggtg caggagcgcg ggctgccgct gtccgagaag ctggcggcca tcaacttctc 11400

ggtgctgagt ttgggcaagt actacgctag gaagatctac aagaccccgt acgtgcccat 11460

agacaaggag gtgaagatcg acgggtttta catgcgcatg accctgaaag tgctgaccct 11520

gagcgacgat ctgggggtgt accgcaacga caggatgcac cgtgcggtga gcgccagcag 11580

gcggcgcgag ctgagcgacc aggagctgat gcatagtctg cagcgggccc tgaccggggc 11640

cgggaccgag ggggagagct actttgacat gggcgcggac ctgcactggc agcccagccg 11700

ccgggccttg gaggcggcgg caggacccta cgtagaagag gtggacgatg aggtggacga 11760

ggagggcgag tacctggaag actgatggcg cgaccgtatt tttgctagat gcaacaacaa 11820

cagccacctc ctgatcccgc gatgcgggcg gcgctgcaga gccagccgtc cggcattaac 11880

[0145] tcctcggacg attggaccca ggccatgcaa cgcatcatgg cgctgacgac ccgcaacccc 11940

gaagccttta gacagcagcc ccaggccaac cggctctcgg ccatcctgga ggccgtggtg 12000

ccctcgcgct ccaaccccac gcacgagaag gtcctggcca tcgtgaacgc gctggtggag 12060

aacaaggcca tccgcggcga cgaggccggc ctggtgtaca acgcgctgct ggagcgcgtg 12120

gcccgctaca acagcaccaa cgtgcagacc aacctggacc gcatggtgac cgacgtgcgc 12180

gaggccgtgg cccagcgcga gcggttccac cgcgagtcca acctgggatc catggtggcg 12240

ctgaacgcct tcctcagcac ccagcccgcc aacgtgcccc ggggccagga ggactacacc 12300

aacttcatca gcgccctgcg cctgatggtg accgaggtgc cccagagcga ggtgtaccag 12360

tccgggccgg actacttctt ccagaccagt cgccagggct tgcagaccgt gaacctgagc 12420

caggctttca agaacttgca gggcctgtgg ggcgtgcagg ccccggtcgg ggaccgcgcg 12480

acggtgtcga gcctgctgac gccgaactcg cgcctgctgc tgctgctggt ggcccccttc 12540

acggacagcg gcagcatcaa ccgcaactcg tacctgggct acctgattaa cctgtaccgc 12600

gaggccatcg gccaggcgca cgtggacgag cagacctacc aggagatcac ccacgtgagc 12660

cgcgccctgg gccaggacga cccgggcaac ctggaagcca ccctgaactt tttgctgacc 12720

aaccggtcgc agaagatccc gccccagtac gcgctcagca ccgaggagga gcgcatcctg 12780

cgttacgtgc agcagagcgt gggcctgttc ctgatgcagg agggggccac ccccagcgcc 12840

gcgctcgaca tgaccgcgcg caacatggag cccagcatgt acgccagcaa ccgcccgttc 12900

atcaataaac tgatggacta cttgcatcgg gcggccgcca tgaactctga ctatttcacc 12960

aacgccatcc tgaatcccca ctggctcccg ccgccggggt tctacacggg cgagtacgac 13020

atgcccgacc ccaatgacgg gttcctgtgg gacgatgtgg acagcagcgt gttctccccc 13080

cgaccgggtg ctaacgagcg ccccttgtgg aagaaggaag gcagcgaccg acgcccgtcc 13140

tcggcgctgt ccggccgcga gggtgctgcc gcggcggtgc ccgaggccgc cagtcctttc 13200

ccgagcttgc ccttctcgct gaacagtatc cgcagcagcg agctgggcag gatcacgcgc 13260

ccgcgcttgc tgggcgaaga ggagtacttg aatgactcgc tgttgagacc cgagcgggag 13320

aagaacttcc ccaataacgg gatagaaagc ctggtggaca agatgagccg ctggaagacg 13380

[0146] tatgcgcagg agcacaggga cgatccccgg gcgtcgcagg gggccacgag ccggggcagc 13440

gccgcccgta aacgccggtg gcacgacagg cagcggggac agatgtggga cgatgaggac 13500

tccgccgacg acagcagcgt gttggacttg ggtgggagtg gtaacccgtt cgctcacctg 13560

cgcccccgta tcgggcgcat gatgtaagag aaaccgaaaa taaatgatac tcaccaaggc 13620

catggcgacc agcgtgcgtt cgtttcttct ctgttgttgt tgtatctagt atgatgaggc 13680

gtgcgtaccc ggagggtcct cctccctcgt acgagagcgt gatgcagcag gcgatggcgg 13740

cggcggcgat gcagcccccg ctggaggctc cttacgtgcc cccgcggtac ctggcgccta 13800

cggaggggcg gaacagcatt cgttactcgg agctggcacc cttgtacgat accacccggt 13860

tgtacctggt ggacaacaag tcggcggaca tcgcctcgct gaactaccag aacgaccaca 13920

gcaacttcct gaccaccgtg gtgcagaaca atgacttcac ccccacggag gccagcaccc 13980

agaccatcaa ctttgacgag cgctcgcggt ggggcggcca gctgaaaacc atcatgcaca 14040

ccaacatgcc caacgtgaac gagttcatgt acagcaacaa gttcaaggcg cgggtgatgg 14100

tctcccgcaa gacccccaat ggggtgacag tgacagagga ttatgatggt agtcaggatg 14160

agctgaagta tgaatgggtg gaatttgagc tgcccgaagg caacttctcg gtgaccatga 14220

ccatcgacct gatgaacaac gccatcatcg acaattactt ggcggtgggg cggcagaacg 14280

gggtgctgga gagcgacatc ggcgtgaagt tcgacactag gaacttcagg ctgggctggg 14340

accccgtgac cgagctggtc atgcccgggg tgtacaccaa cgaggctttc catcccgata 14400

ttgtcttgct gcccggctgc ggggtggact tcaccgagag ccgcctcagc aacctgctgg 14460

gcattcgcaa gaggcagccc ttccaggaag gcttccagat catgtacgag gatctggagg 14520

ggggcaacat ccccgcgctc ctggatgtcg acgcctatga gaaaagcaag gaggatgcag 14580

cagctgaagc aactgcagcc gtagctaccg cctctaccga ggtcaggggc gataattttg 14640

caagcgccgc agcagtggca gcggccgagg cggctgaaac cgaaagtaag atagtcattc 14700

agccggtgga gaaggatagc aagaacagga gctacaacgt actaccggac aagataaaca 14760

ccgcctaccg cagctggtac ctagcctaca actatggcga ccccgagaag ggcgtgcgct 14820

cctggacgct gctcaccacc tcggacgtca cctgcggcgt ggagcaagtc tactggtcgc 14880

[0147] tgcccgacat gatgcaagac ccggtcacct tccgctccac gcgtcaagtt agcaactacc 14940

cggtggtggg cgccgagctc ctgcccgtct actccaagag cttcttcaac gagcaggccg 15000

tctactcgca gcagctgcgc gccttcacct cgcttacgca cgtcttcaac cgcttccccg 15060

agaaccagat cctcgtccgc ccgcccgcgc ccaccattac caccgtcagt gaaaacgttc 15120

ctgctctcac agatcacggg accctgccgc tgcgcagcag tatccgggga gtccagcgcg 15180

tgaccgttac tgacgccaga cgccgcacct gcccctacgt ctacaaggcc ctgggcatag 15240

tcgcgccgcg cgtcctctcg agccgcacct tctaaatgtc cattctcatc tcgcccagta 15300

ataacaccgg ttggggcctg cgcgcgccca gcaagatgta cggaggcgct cgccaacgct 15360

ccacgcaaca ccccgtgcgc gtgcgcgggc acttccgcgc tccctggggc gccctcaagg 15420

gccgcgtgcg gtcgcgcacc accgtcgacg acgtgatcga ccaggtggtg gccgacgcgc 15480

gcaactacac ccccgccgcc gcgcccgtct ccaccgtgga cgccgtcatc gacagcgtgg 15540

tggccgacgc gcgccggtac gcccgcgcca agagccggcg gcggcgcatc gcccggcggc 15600

accggagcac ccccgccatg cgcgcggcgc gagccttgct gcgcagggcc aggcgcacgg 15660

gacgcagggc catgctcagg gcggccagac gcgcggcttc aggcgccagc gccggcagga 15720

cccggagacg cgcggccacg gcggcggcag cggccatcgc cagcatgtcc cgcccgcggc 15780

gagggaacgt gtactgggtg cgcgacgccg ccaccggtgt gcgcgtgccc gtgcgcaccc 15840

gccccccctcg cacttgaaga tgttcacttc gcgatgttga tgtgtcccag cggcgaggag 15900

gatgtccaag cgcaaattca aggaagagat gctccaggtc atcgcgcctg agatctacgg 15960

ccctgcggtg gtgaaggagg aaagaaagcc ccgcaaaatc aagcgggtca aaaaggacaa 16020

aaaggaagaa gaaagtgatg tggacggatt ggtggagttt gtgcgcgagt tcgccccccg 16080

gcggcgcgtg cagtggcgcg ggcggaaggt gcaaccggtg ctgagacccg gcaccaccgt 16140

ggtcttcacg cccggcgagc gctccggcac cgcttccaag cgctcctacg acgaggtgta 16200

cggggatgat gatattctgg agcaggcggc cgagcgcctg ggcgagtttg cttacggcaa 16260

gcgcagccgt tccgcaccga aggaagaggc ggtgtccatc ccgctggacc acggcaaccc 16320

cacgccgagc ctcaagcccg tgaccttgca gcaggtgctg ccgaccgcgg cgccgcgccg 16380

[0148] ggggttcaag cgcgagggcg aggatctgta ccccaccatg cagctgatgg tgcccaagcg 16440

ccagaagctg gaagacgtgc tggagaccat gaaggtggac ccggacgtgc agcccgaggt 16500

caaggtgcgg cccatcaagc aggtggcccc gggcctgggc gtgcagaccg tggacatcaa 16560

gattcccacg gagcccatgg aaacgcagac cgagcccatg atcaagccca gcaccagcac 16620

catggaggtg cagacggatc cctggatgcc atcggctcct agtcgaagac cccggcgcaa 16680

gtacggcgcg gccagcctgc tgatgcccaa ctacgcgctg catccttcca tcatccccac 16740

gccgggctac cgcggcacgc gcttctaccg cggtcatacc agcagccgcc gccgcaagac 16800

caccactcgc cgccgccgtc gccgcaccgc cgctgcaacc acccctgccg ccctggtgcg 16860

gagagtgtac cgccgcggcc gcgcacctct gaccctgccg cgcgcgcgct accacccgag 16920

catcgccatt taaactttcg cctgctttgc agatcaatgg ccctcacatg ccgccttcgc 16980

gttcccatta cgggctaccg aggaagaaaa ccgcgccgta gaaggctggc ggggaacggg 17040

atgcgtcgcc accaccaccg gcggcggcgc gccatcagca agcggttggg gggaggcttc 17100

ctgcccgcgc tgatccccat catcgccgcg gcgatcgggg cgatccccgg cattgcttcc 17160

gtggcggtgc aggcctctca gcgccactga gacacacttg gaaacatctt gtaataaacc 17220

aatggactct gacgctcctg gtcctgtgat gtgttttcgt agacagatgg aagacatcaa 17280

tttttcgtcc ctggctccgc gacacggcac gcggccgttc atgggcacct ggagcgacat 17340

cggcaccagc caactgaacg ggggcgcctt caattggagc agtctctgga gcgggcttaa 17400

gaatttcggg tccacgctta aaacctatgg cagcaaggcg tggaacagca ccacagggca 17460

ggcgctgagg gataagctga aagagcagaa cttccagcag aaggtggtcg atgggctcgc 17520

ctcgggcatc aacggggtgg tggacctggc caaccaggcc gtgcagcggc agatcaacag 17580

ccgcctggac ccggtgccgc ccgccggctc cgtggagatg ccgcaggtgg aggaggagct 17640

gcctcccctg gacaagcggg gcgagaagcg accccgcccc gatgcggagg agacgctgct 17700

gacgcacacg gacgagccgc ccccgtacga ggaggcggtg aaactgggtc tgcccaccac 17760

gcggcccatc gcgcccctgg ccaccggggt gctgaaaccc gaaaagcccg cgaccctgga 17820

cttgcctcct ccccagcctt cccgccctc tacagtggct aagcccctgc cgccggtggc 17880

[0149] cgtggcccgc gcgcgacccg ggggcaccgc ccgccctcat gcgaactggc agagcactct 17940

gaacagcatc gtgggtctgg gagtgcagag tgtgaagcgc cgccgctgct attaaaccta 18000

ccgtagcgct taacttgctt gtctgtgtgt gtatgtatta tgtcgccgcc gccgctgtcc 18060

accagaagga ggagtgaaga ggcgcgtcgc cgagttgcaa gatggccacc ccatcgatgc 18120

tgccccagtg ggcgtacatg cacatcgccg gacaggacgc ttcggagtac ctgagtccgg 18180

gtctggtgca gtttgcccgc gccacagaca cctacttcag tctggggaac aagtttagga 18240

accccacggt ggcgcccacg cacgatgtga ccaccgaccg cagccagcgg ctgacgctgc 18300

gcttcgtgcc cgtggaccgc gaggacaaca cctactcgta caaagtgcgc tacacgctgg 18360

ccgtgggcga caaccgcgtg ctggacatgg ccagcaccta ctttgacatc cgcggcgtgc 18420

tggatcgggg ccctagcttc aaaccctact ccggcaccgc ctacaacagt ctggccccca 18480

agggagcacc caacacttgt cagtggacat ataaagccga tggtgaaact gccacagaaa 18540

aaacctatac atatggaaat gcacccgtgc agggcattaa catcacaaaa gatggtattc 18600

aacttggaac tgacaccgat gatcagccaa tctacgcaga taaaacctat cagcctgaac 18660

ctcaagtggg tgatgctgaa tggcatgaca tcactggtac tgatgaaaag tatggaggca 18720

gagctcttaa gcctgatacc aaaatgaagc cttgttatgg ttcttttgcc aagcctacta 18780

ataaagaagg aggtcaggca aatgtgaaaa caggaacagg cactactaaa gaatatgaca 18840

tagacatggc tttctttgac aacagaagtg cggctgctgc tggcctagct ccagaaattg 18900

ttttgtatac tgaaaatgtg gatttggaaa ctccagatac ccatattgta tacaaagcag 18960

gcacagatga cagcagctct tctattaatt tgggtcagca agccatgccc aacagaccta 19020

actacattgg tttcagagac aactttatcg ggctcatgta ctacaacagc actggcaata 19080

tgggggtgct ggccggtcag gcttctcagc tgaatgctgt ggttgacttg caagacagaa 19140

acaccgagct gtcctaccag ctcttgcttg actctctggg tgacagaacc cggtatttca 19200

gtatgtggaa tcaggcggtg gacagctatg atcctgatgt gcgcattatt gaaaatcatg 19260

gtgtggagga tgaacttccc aactattgtt tccctctgga tgctgttggc agaacagata 19320

cttatcaggg aattaaggct aatggaactg atcaaaccac atggaccaaa gatgacagtg 19380

[0150] tcaatgatgc taatgagata ggcaagggta atccattcgc catggaaatc aacatccaag 19440

ccaacctgtg gaggaacttc ctctacgcca acgtggccct gtacctgccc gactcttaca 19500

agtacacgcc ggccaatgtt accctgccca ccaacaccaa cacctacgat tacatgaacg 19560

gccgggtggt ggcgccctcg ctggtggact cctacatcaa catcggggcg cgctggtcgc 19620

tggatcccat ggacaacgtg aaccccttca accaccaccg caatgcgggg ctgcgctacc 19680

gctccatgct cctgggcaac gggcgctacg tgcccttcca catccaggtg ccccagaaat 19740

ttttcgccat caagagcctc ctgctcctgc ccgggtccta cacctacgag tggaacttcc 19800

gcaaggacgt caacatgatc ctgcagagct ccctcggcaa cgacctgcgc acggacgggg 19860

cctccatctc cttcaccagc atcaacctct acgccacctt cttccccatg gcgcacaaca 19920

cggcctccac gctcgaggcc atgctgcgca acgacaccaa cgaccagtcc ttcaacgact 19980

acctctcggc ggccaacatg ctctacccca tcccggccaa cgccaccaac gtgcccatct 20040

ccatcccctc gcgcaactgg gccgccttcc gcgggctggtc cttcacgcgt ctcaagacca 20100

aggagacgcc ctcgctgggc tccgggttcg acccctactt cgtctactcg ggctccatcc 20160

cctacctcga cggcaccttc tacctcaacc acaccttcaa gaaggtctcc atcaccttcg 20220

actcctccgt cagctggccc ggcaacgacc ggctcctgac gcccaacgag ttcgaaatca 20280

agcgcaccgt cgacggcgag ggctacaacg tggcccagtg caacatgacc aaggactggt 20340

tcctggtcca gatgctggcc cactacaaca tcggctacca gggcttctac gtgcccgagg 20400

gctacaagga ccgcatgtac tccttcttcc gcaacttcca gcccatgagc cgccaggtgg 20460

tggacgaggt caactacaag gactaccagg ccgtcaccct ggcctaccag cacaacaact 20520

cgggcttcgt cggctacctc gcgcccacca tgcgccaggg ccagccctac cccgccaact 20580

acccctaccc gctcatcggc aagagcgccg tcaccagcgt cacccagaaa aagttcctct 20640

gcgacagggt catgtggcgc atccccttct ccagcaactt catgtccatg ggcgcgctca 20700

ccgacctcgg ccagaacatg ctctatgcca actccgccca cgcgctagac atgaatttcg 20760

aagtcgaccc catggatgag tccacccttc tctatgttgt cttcgaagtc ttcgacgtcg 20820

tccgagtgca ccagccccac cgcggcgtca tcgaggccgt ctacctgcgc acccccttct 20880

[0151] cggccggtaa cgccaccacc taagctcttg cttcttgcaa gccatggccg cgggctccgg 20940

cgagcaggag ctcagggcca tcatccgcga cctgggctgc gggccctact tcctgggcac 21000

cttcgataag cgcttcccgg gattcatggc cccgcacaag ctggcctgcg ccatcgtcaa 21060

cacggccggc cgcgagaccg ggggcgagca ctggctggcc ttcgcctgga acccgcgctc 21120

gaacacctgc tacctcttcg acccccttcgg gttctcggac gagcgcctca agcagatcta 21180

ccagttcgag tacgagggcc tgctgcgccg cagcgccctg gccaccgagg accgctgcgt 21240

caccctggaa aagtccaccc agaccgtgca gggtccgcgc tcggccgcct gcgggctctt 21300

ctgctgcatg ttcctgcacg ccttcgtgca ctggcccgac cgccccatgg acaagaaccc 21360

caccatgaac ttgctgacgg gggtgcccaa cggcatgctc cagtcgcccc aggtggaacc 21420

caccctgcgc cgcaaccagg aggcgctcta ccgcttcctc aactcccact ccgcctactt 21480

tcgctcccac cgcgcgcgca tcgagaaggc caccgccttc gaccgcatga atcaagacat 21540

gtaaaccgtg tgtgtatgtt aaatgtcttt aataaacagc actttcatgt tacacatgca 21600

tctgagatga tttatttaga aatcgaaagg gttctgccgg gtctcggcat ggcccgcggg 21660

cagggacacg ttgcggaact ggtacttggc cagccacttg aactcgggga tcagcagttt 21720

gggcagcggg gtgtcgggga aggagtcggt ccacagcttc cgcgtcagtt gcagggcgcc 21780

cagcaggtcg ggcgcggaga tcttgaaatc gcagttggga cccgcgttct gcgcgcggga 21840

gttgcggtac acggggttgc agcactggaa caccatcagg gccgggtgct tcacgctcgc 21900

cagcaccgtc gcgtcggtga tgctctccac gtcgaggtcc tcggcgttgg ccatcccgaa 21960

gggggtcatc ttgcaggtct gccttcccat ggtgggcacg cacccgggct tgtggttgca 22020

atcgcagtgc agggggatca gcatcatctg ggcctggtcg gcgttcatcc ccgggtacat 22080

ggccttcatg aaagcctcca attgcctgaa cgcctgctgg gccttggctc cctcggtgaa 22140

gaagaccccg caggacttgc tagagaactg gttggtggcg cacccggcgt cgtgcacgca 22200

gcagcgcgcg tcgttgttgg ccagctgcac cacgctgcgc ccccagcggt tctgggtgat 22260

cttggcccgg tcggggttct ccttcagcgc gcgctgcccg ttctcgctcg ccacatccat 22320

ctcgatcatg tgctccttct ggatcatggt ggtcccgtgc aggcaccgca gcttgccctc 22380

[0152] ggcctcggtg cacccgtgca gccacagcgc gcacccggtg cactcccagt tcttgtgggc 22440

gatctgggaa tgcgcgtgca cgaagccctg caggaagcgg cccatcatgg tggtcagggt 22500

cttgttgcta gtgaaggtca gcggaatgcc gcggtgctcc tcgttgatgt acaggtggca 22560

gatgcggcgg tacacctcgc cctgctcggg catcagctgg aagttggctt tcaggtcggt 22620

ctccacgcgg tagcggtcca tcagcatagt catgatttcc atacccttct cccaggccga 22680

gacgatgggc aggctcatag ggttcttcac catcatctta gcgctagcag ccgcggccag 22740

ggggtcgctc tcgtccaggg tctcaaagct ccgcttgccg tccttctcgg tgatccgcac 22800

cgggggggtag ctgaagccca cggccgccag ctcctcctcg gcctgtcttt cgtcctcgct 22860

gtcctggctg acgtcctgca ggaccacatg cttggtcttg cggggtttct tcttgggcgg 22920

cagcggcggc ggagatgttg gagatggcga gggggagcgc gagttctcgc tcaccactac 22980

tatctcttcc tcttcttggt ccgaggccac gcggcggtag gtatgtctct tcgggggcag 23040

aggcggaggc gacgggctct cgccgccgcg acttggcgga tggctggcag agccccttcc 23100

gcgttcgggg gtgcgctccc ggcggcgctc tgactgactt cctccgcggc cggccattgt 23160

gttctcctag ggaggaacaa caagcatgga gactcagcca tcgccaacct cgccatctgc 23220

ccccaccgcc gacgagaagc agcagcagca gaatgaaagc ttaaccgccc cgccgcccag 23280

ccccgccacc tccgacgcgg ccgtcccaga catgcaagag atggaggaat ccatcgagat 23340

tgacctgggc tatgtgacgc ccgcggagca cgaggaggag ctggcagtgc gcttttcaca 23400

agaagagata caccaagaac agccagagca ggaagcagag aatgagcaga gtcaggctgg 23460

gctcgagcat gacggcgact acctccacct gagcgggggg gaggacgcgc tcatcaagca 23520

tctggcccgg caggccacca tcgtcaagga tgcgctgctc gaccgcaccg aggtgccccc 23580

cagcgtggag gagctcagcc gcgcctacga gttgaacctc ttctcgccgc gcgtgccccc 23640

caagcgccag cccaatggca cctgcgagcc caacccgcgc ctcaacttct acccggtctt 23700

cgcggtgccc gaggccctgg ccacctacca catcttttttc aagaaccaaa agatccccgt 23760

ctcctgccgc gccaaccgca cccgcgccga cgccctttc aacctgggtc ccggcgcccg 23820

cctacctgat atcgcctcct tggaagaggt tcccaagatc ttcgagggtc tgggcagcga 23880

[0153] cgagactcgg gccgcgaacg ctctgcaagg agaaggagga gagcatgagc accacagcgc 23940

cctggtcgag ttggaaggcg acaacgcgcg gctggcggtg ctcaaacgca cggtcgagct 24000

gacccatttc gcctacccgg ctctgaacct gccccccaaa gtcatgagcg cggtcatgga 24060

ccaggtgctc atcaagcgcg cgtcgcccat ctccgaggac gagggcatgc aagactccga 24120

ggagggcaag cccgtggtca gcgacgagca gctggcccgg tggctgggtc ctaatgctag 24180

tccccagagt ttggaagagc ggcgcaaact catgatggcc gtggtcctgg tgaccgtgga 24240

gctggagtgc ctgcgccgct tcttcgccga cgcggagacc ctgcgcaagg tcgaggagaa 24300

cctgcactac ctcttcaggc acgggttcgt gcgccaggcc tgcaagatct ccaacgtgga 24360

gctgaccaac ctggtctcct acatgggcat cttgcacgag aaccgcctgg ggcagaacgt 24420

gctgcacacc accctgcgcg gggaggcccg gcgcgactac atccgcgact gcgtctacct 24480

ctacctctgc cacacctggc agacgggcat gggcgtgtgg cagcagtgtc tggaggagca 24540

gaacctgaaa gagctctgca agctcctgca gaagaacctc aagggtctgt ggaccgggtt 24600

cgacgagcgc accaccgcct cggacctggc cgacctcatt ttccccgagc gcctcaggct 24660

gacgctgcgc aacggcctgc ccgactttat gagccaaagc atgttgcaaa actttcgctc 24720

tttcatcctc gaacgctccg gaatcctgcc cgccacctgc tccgcgctgc cctcggactt 24780

cgtgccgctg accttccgcg agtgcccccc gccgctgtgg agccactgct acctgctgcg 24840

cctggccaac tacctggcct accactcgga cgtgatcgag gacgtcagcg gcgagggcct 24900

gctcgagtgc cactgccgct gcaacctctg cacgccgcac cgctccctgg cctgcaaccc 24960

ccagctgctg agcgagaccc agatcatcgg caccttcgag ttgcaagggc ccagcgaagg 25020

cgagggttca gccgccaagg ggggtctgaa actcaccccg gggctgtgga cctcggccta 25080

cttgcgcaag ttcgtgcccg aggactacca tcccttcgag atcaggttct acgaggacca 25140

atcccatccg cccaaggccg agctgtcggc ctgcgtcatc acccaggggg cgatcctggc 25200

ccaattgcaa gccatccaga aatcccgcca agaattcttg ctgaaaaagg gccgcggggt 25260

ctacctcgac ccccagaccg gtgaggagct caaccccggc ttcccccagg atgccccgag 25320

gaaacaagaa gctgaaagtg gagctgccgc ccgtggagga tttggaggaa gactgggaga 25380

[0154] acagcagtca ggcagaggag gaggagatgg aggaagactg ggacagcact caggcagagg 25440

aggacagcct gcaagacagt ctggaggaag acgaggagga ggcagaggag gaggtggaag 25500

aagcagccgc cgccagaccg tcgtcctcgg cgggggagaa agcaagcagc acggatacca 25560

tctccgctcc gggtcggggt cccgctcgac cacacagtag atgggacgag accggacgat 25620

tcccgaaccc caccacccag accggtaaga aggagcggca gggatacaag tcctggcggg 25680

ggcacaaaaa cgccatcgtc tcctgcttgc aggcctgcgg gggcaacatc tccttcaccc 25740

ggcgctacct gctcttccac cgcggggtga actttccccg caacatcttg cattactacc 25800

gtcacctcca cagcccctac tacttccaag aagaggcagc agcagcagaa aaagaccagc 25860

agaaaaccag cagctagaaa atccacagcg gcggcagcag gtggactgag gatcgcggcg 25920

aacgagccgg cgcaaacccg ggagctgagg aaccggatct ttcccaccct ctatgccatc 25980

ttccagcaga gtcgggggca ggagcaggaa ctgaaagtca agaaccgttc tctgcgctcg 26040

ctcacccgca gttgtctgta tcacaagagc gaagaccaac ttcagcgcac tctcgaggac 26100

gccgaggctc tcttcaacaa gtactgcgcg ctcactctta aagagtagcc cgcgcccgcc 26160

cagtcgcaga aaaaggcggg aattacgtca cctgtgccct tcgccctagc cgcctccacc 26220

catcatcatg agcaaagaga ttcccacgcc ttacatgtgg agctaccagc cccagatggg 26280

cctggccgcc ggtgccgccc aggactactc cacccgcatg aattggctca gcgccgggcc 26340

cgcgatgatc tcacgggtga atgacatccg cgcccaccga aaccagatac tcctagaaca 26400

gtcagcgctc accgccacgc cccgcaatca cctcaatccg cgtaattggc ccgccgccct 26460

ggtgtaccag gaaattcccc agcccacgac cgtactactt ccgcgagacg cccaggccga 26520

agtccagctg actaactcag gtgtccagct ggcgggcggc gccaccctgt gtcgtcaccg 26580

ccccgctcag ggtataaagc ggctggtgat ccggggcaga ggcacacagc tcaacgacga 26640

ggtggtgagc tcttcgctgg gtctgcgacc tgacggagtc ttccaactcg ccggatcggg 26700

gagatcttcc ttcacgcctc gtcaggccgt cctgactttg gagagttcgt cctcgcagcc 26760

ccgctcgggt ggcatcggca ctctccagtt cgtggaggag ttcactccct cggtctactt 26820

caaccccttc tccggctccc ccggccacta cccggacgag ttcatcccga acttcgacgc 26880

[0155] catcagcgag tcggtggacg gctacgattg aatgtcccat ggtggcgcag ctgacctagc 26940

tcggcttcga cacctggacc actgccgccg cttccgctgc ttcgctcggg atctcgccga 27000

gtttgcctac tttgagctgc ccgaggagca ccctcagggc ccggcccacg gagtgcggat 27060

cgtcgtcgaa gggggcctcg actcccacct gcttcggatc ttcagccagc gtccgatcct 27120

ggtcgagcgc gagcaaggac agacccttct gactctgtac tgcatctgca accaccccgg 27180

cctgcatgaa agtctttgtt gtctgctgtg tactgagtat aataaaagct gagatcagcg 27240

actactccgg acttccgtgt gttcctgaat ccatcaacca gtctttgttc ttcaccggga 27300

acgagaccga gctccagctc cagtgtaagc cccacaagaa gtacctcacc tggctgttcc 27360

agggctcccc gatcgccgtt gtcaaccact gcgacaacga cggagtcctg ctgagcggcc 27420

ctgccaacct tacttttttcc acccgcagaa gcaagctcca gctcttccaa cccttcctcc 27480

ccgggaccta tcagtgcgtc tcgggaccct gccatcacac cttccacctg atcccgaata 27540

ccacagcgtc gctccccgct actaacaacc aaactaacct ccaccaacgc caccgtcgcg 27600

acggccacaa tacatgccca tattagacta tgaggccgag ccacagcgac ccatgctccc 27660

cgctattagt tacttcaatc taaccggcgg agatgactga cccactggcc aacaacaacg 27720

tcaacgacct tctcctggac atggacggcc gcgcctcgga gcagcgactc gcccaacttc 27780

gcattcgcca gcagcaggag agagccgtca aggagctgca ggatgcggtg gccatccacc 27840

agtgcaagag aggcatcttc tgcctggtga aacaggccaa gatctcctac gaggtcactc 27900

caaacgacca tcgcctctcc tacgagctcc tgcagcagcg ccagaagttc acctgcctgg 27960

tcggagtcaa ccccatcgtc atcacccagc agtctggcga taccaagggg tgcatccact 28020

gctcctgcga ctccccgac tgcgtccaca ctctgatcaa gaccctctgc ggcctccgcg 28080

acctcctccc catgaactaa tcacccccctt atccagtgaa ataaagatca tattgatgat 28140

gattttacag aaataaaaaa taatcatttg atttgaaata aagatacaat catattgatg 28200

atttgagttt aacaaaaaaa taaagaatca cttacttgaa atctgatacc aggtctctgt 28260

ccatgttttc tgccaacacc acttcactcc cctcttccca gctctggtac tgcaggcccc 28320

ggcgggctgc aaacttcctc cacacgctga aggggatgtc aaattcctcc tgtccctcaa 28380

[0156] tcttcatttt atcttctatc agatgtccaa aaagcgcgtc cgggtggatg atgacttcga 28440

ccccgtctac ccctacgatg cagacaacgc accgaccgtg cccttcatca acccccccctt 28500

cgtctcttca gatggattcc aagagaagcc cctgggggtg ttgtccctgc gactggccga 28560

ccccgtcacc accaagaacg gggaaatcac cctcaagctg ggagaggggg tggacctcga 28620

ttcctcggga aaactcatct ccaacacggc caccaaggcc gccgcccctc tcagtttttc 28680

caacaacacc atttccctta acatggatca ccccttttac actaaagatg gaaaattatc 28740

cttacaagtt tctccaccat taaatatact gagaacaagc attctaaaca cactagcttt 28800

aggttttgga tcaggtttag gactccgtgg ctctgccttg gcagtacagt tagtctctcc 28860

acttacattt gatactgatg gaaacataaa gcttacctta gacagaggtt tgcatgttac 28920

aacaggagat gcaattgaaa gcaacataag ctgggctaaa ggtttaaaat ttgaagatgg 28980

agccatagca accaacattg gaaatgggtt agagtttgga agcagtagta cagaaacagg 29040

tgttgatgat gcttacccaa tccaagttaa acttggatct ggccttagct ttgacagtac 29100

aggagccata atggctggta acaaagaaga cgataaactc actttgtgga caacacctga 29160

tccatcacca aactgtcaaa tactcgcaga aaatgatgca aaactaacac tttgcttgac 29220

taaatgtggt agtcaaatac tggccactgt gtcagtctta gttgtaggaa gtggaaacct 29280

aaaccccatt actggcaccg taagcagtgc tcaggtgttt ctacgttttg atgcaaacgg 29340

tgttctttta acagaacatt ctacactaaa aaaatactgg gggtataggc agggagatag 29400

catagatggc actccatata ccaatgctgt aggattcatg cccaatttaa aagcttatcc 29460

aaagtcacaa agttctacta ctaaaaataa tatagtaggg caagtataca tgaatggaga 29520

tgtttcaaaa cctatgcttc tcactataac cctcaatggt actgatgaca gcaacagtac 29580

atattcaatg tcattttcat acacctggac taatggaagc tatgttggag caacatttgg 29640

ggctaactct tataccttct catacatcgc ccaagaatga acactgtatc ccaccctgca 29700

tgccaaccct tcccacccca ctctgtggaa caaactctga aacacaaaat aaaataaagt 29760

tcaagtgttt tattgattca acagttttac aggattcgag cagttatttt tcctccaccc 29820

tcccaggaca tggaatacac caccctctcc ccccgcacag ccttgaacat ctgaatgcca 29880

[0157] ttggtgatgg acatgctttt ggtctccacg ttccacacag tttcagagcg agccagtctc 29940

gggtcggtca gggagatgaa accctccggg cactcccgca tctgcacctc acagctcaac 30000

agctgaggat tgtcctcggt ggtcgggatc acggttatct ggaagaagca gaagagcggc 30060

ggtgggaatc atagtccgcg aacgggatcg gccggtggtg tcgcatcagg ccccgcagca 30120

gtcgctgccg ccgccgctcc gtcaagctgc tgctcagggg gtccgggtcc agggactccc 30180

tcagcatgat gcccacggcc ctcagcatca gtcgtctggt gcggcgggcg cagcagcgca 30240

tgcggatctc gctcaggtcg ctgcagtacg tgcaacacag aaccaccagg ttgttcaaca 30300

gtccatagtt caacacgctc cagccgaaac tcatcgcggg aaggatgcta cccacgtggc 30360

cgtcgtacca gatcctcagg taaatcaagt ggtgcccccc ccagaacacg ctgcccacgt 30420

acatgatctc cttgggcatg tggcggttca ccacctcccg gtaccacatc accctctggt 30480

tgaacatgca gccccggatg atcctgcgga accacagggc cagcaccgcc ccgcccgcca 30540

tgcagcgaag agacccccggg tcccggcaat ggcaatggag gacccaccgc tcgtacccgt 30600

ggatcatctg ggagctgaac aagtctatgt tggcacagca caggcatatg ctcatgcatc 30660

tcttcagcac tctcaactcc tcgggggtca aaaccatatc ccagggcacg gggaactctt 30720

gcaggacagc gaaccccgca gaacagggca atcctcgcac agaacttaca ttgtgcatgg 30780

acagggtatc gcaatcaggc agcaccgggt gatcctccac cagagaagcg cgggtctcgg 30840

tctcctcaca gcgtggtaag ggggccggcc gatacgggtg atggcgggac gcggctgatc 30900

gtgttcgcga ccgtgtcatg atgcagttgc tttcggacat tttcgtactt gctgtagcag 30960

aacctggtcc gggcgctgca caccgatcgc cggcggcggt ctcggcgctt ggaacgctcg 31020

gtgttgaaat tgtaaaacag ccactctctc agaccgtgca gcagatctag ggcctcagga 31080

gtgatgaaga tcccatcatg cctgatggct ctgatcacat cgaccaccgt ggaatgggcc 31140

agacccagcc agatgatgca attttgttgg gtttcggtga cggcgggga gggaagaaca 31200

ggaagaacca tgattaactt ttaatccaaa cggtctcgga gtacttcaaa atgaagatcg 31260

cggagatggc acctctcgcc cccgctgtgt tggtggaaaa taacagccag gtcaaaggtg 31320

atacggttct cgagatgttc cacggtggct tccagcaaag cctccacgcg cacatccaga 31380

[0158] aacaagacaa tagcgaaagc gggagggttc tctaattcct caatcatcat gttacactcc 31440

tgcaccatcc ccagataatt ttcatttttc cagccttgaa tgattcgaac tagttcgtga 31500

ggtaaatcca agccagccat gataaagagc tcgcgcagag cgccctccac cggcattctt 31560

aagcacaccc tcataattcc aagatattct gctcctggtt cacctgcagc agattgacaa 31620

gcggaatatc aaaatctctg ccgcgatccc tgagctcctc cctcagcaat aactgtaagt 31680

actctttcat atcctctccg aaatttttag ccataggacc accaggaata agattagggc 31740

aagccacagt acagataaac cgaagtcctc cccagtgagc attgccaaat gcaagactgc 31800

tataagcatg ctggctagac ccggtgatat cttccagata actggacaga aaatcgccca 31860

ggcaattttt aagaaaatca acaaaagaaa aatcctccag gtggacgttt agagcctcgg 31920

gaacaacgat gaagtaaatg caagcggtgc gttccagcat ggttagttag ctgatctgta 31980

gaaaaaacaa aaatgaacat taaaccatgc tagcctggcg aacaggtggg taaatcgttc 32040

tctccagcac caggcaggcc acggggtctc cggcgcgacc ctcgtaaaaa ttgtcgctat 32100

gattgaaaac catcacagag agacgttccc ggtggccggc gtgaatgatt cgacaagatg 32160

aatacacccc cggaacattg gcgtccgcga gtgaaaaaaa gcgcccgagg aagcaataag 32220

gcactacaat gctcagtctc aagtccagca aagcgatgcc atgcggatga agcacaaaat 32280

tctcaggtgc gtacaaaatg taattactcc cctcctgcac aggcagcaaa gcccccgatc 32340

cctccaggta cacatacaaa gcctcagcgt ccatagctta ccgagcagca gcacacaaca 32400

ggcgcaagag tcagagaaag gctgagctct aacctgtcca cccgctctct gctcaatata 32460

tagcccagat ctacactgac gtaaaggcca aagtctaaaa atacccgcca aataatcaca 32520

cacgcccagc acacgcccag aaaccggtga cacactcaaa aaaatacgcg cacttcctca 32580

aacgcccaaa actgccgtca tttccgggtt cccacgctac gtcatcaaaa cacgactttc 32640

aaattccgtc gaccgttaaa aacgtcaccc gccccgcccc taacggtcgc ccgtctctca 32700

gccaatcagc gccccgcatc cccaaattca aacacctcat ttgcatatta acgcgcacaa 32760

aaagtttgag gtatattatt gatgatgg 32788

<210> 13

[0159] <211> 30684

<212> DNA

<213> 人工序列(Artificial Sequence)

<220>

<223> 人工序列的描述：合成多核苷酸

<400> 13

ccatcttcaa taatatacct caaacttttt gtgcgcgtta atatgcaaat gaggcgtttg 60

aatttgggga ggaagggcgg tgattggtcg agggatgagc gaccgttagg ggcggggcga 120

gtgacgtttt gatgacgtgg ttgcgaggag gagccagttt gcaagttctc gtgggaaaag 180

tgacgtcaaa cgaggtgtgg tttgaacacg gaaatactca attttcccgc gctctctgac 240

aggaaatgag gtgtttctgg gcggatgcaa gtgaaaacgg gccattttcg cgcgaaaact 300

gaatgaggaa gtgaaaatct gagtaatttc gcgtttatgg cagggaggag tatttgccga 360

gggccgagta gactttgacc gattacgtgg gggtttcgat taccgtgttt ttcacctaaa 420

tttccgcgta cggtgtcaaa gtccggtgtt tttacgtagg tgtcagctga tcgccagggt 480

atttaaacct gcgctctcca gtcaagaggc cactcttgag tgccagcgag aagagttttc 540

tcctccgcgc cgcgagtcag atctacactt tgaaagtagg gataacaggg taatgacatt 600

gattattgac tagttgttaa tagtaatcaa ttacggggtc attagttcat agcccatata 660

tggagttccg cgttacataa cttacggtaa atggcccgcc tggctgaccg cccaacgacc 720

cccgcccatt gacgtcaata atgacgtatg ttcccatagt aacgccaata gggactttcc 780

attgacgtca atgggtggag tatttacggt aaactgccca cttggcagta catcaagtgt 840

atcatatgcc aagtccgccc cctattgacg tcaatgacgg taaatggccc gcctggcatt 900

atgcccagta catgacctta cgggactttc ctacttggca gtacatctac gtattagtca 960

tcgctattac catggtgatg cggttttggc agtacaccaa tgggcgtgga tagcggtttg 1020

actcacgggg atttccaagt ctccacccca ttgacgtcaa tgggagtttg ttttggcacc 1080

aaaatcaacg ggactttcca aaatgtcgta ataaccccgc cccgttgacg caaatgggcg 1140

gtaggcgtgt acggtgggag gtctatataa gcagagctcg tttagtgaac cgtcagatcg 1200

cctggaacgc catccacgct gttttgacct ccatagaaga cagcgatcgc gccaccatgg 1260

[0160] tgagcaaggg cgaggagctg ttcaccgggg tggtgcccat cctggtcgag ctggacggcg 1320

acgtaaacgg ccacaagttc agcgtgtccg gcgagggcga gggcgatgcc acctacggca 1380

agctgaccct gaagttcatc tgcaccaccg gcaagctgcc cgtgccctgg cccaccctcg 1440

tgaccaccct gacctacggc gtgcagtgct tcagccgcta ccccgaccac atgaagcagc 1500

acgacttctt caagtccgcc atgcccgaag gctacgtcca ggagcgcacc atcttcttca 1560

aggacgacgg caactacaag acccgcgccg aggtgaagtt cgagggcgac accctggtga 1620

accgcatcga gctgaagggc atcgacttca aggaggacgg caacatcctg gggcacaagc 1680

tggagtacaa ctacaacagc cacaacgtct atatcatggc cgacaagcag aagaacggca 1740

tcaaggtgaa cttcaagatc cgccacaaca tcgaggacgg cagcgtgcag ctcgccgacc 1800

actaccagca gaacaccccc atcggcgacg gccccgtgct gctgcccgac aaccactacc 1860

tgagcaccca gtccgccctg agcaaagacc ccaacgagaa gcgcgatcac atggtcctgc 1920

tggagttcgt gaccgccgcc gggatcactc tcggcatgga cgagctttac aagtagtgag 1980

tttaaactcc catttaaatg tgagggttaa tgcttcgagc agacatgata agatacattg 2040

atgagtttgg acaaaccaca actagaatgc agtgaaaaaa atgctttatt tgtgaaattt 2100

gtgatgctat tgctttattt gtaaccatta taagctgcaa taaacaagtt aacaacaaca 2160

attgcattca ttttatgttt caggttcagg gggagatgtg ggaggttttt taaagcaagt 2220

aaaacctcta caaatgtggt aaaataacta taacggtcct aaggtagcga gtgagtagtg 2280

ttctggggcg ggggaggacc tgcatgaggg ccagaataac tgaaatctgt gcttttctgt 2340

gtgttgcagc agcatgagcg gaagcggctc ctttgaggga ggggtattca gcccttatct 2400

gacggggcgt ctcccctcct gggcgggagt gcgtcagaat gtgatgggat ccacggtgga 2460

cggccggccc gtgcagcccg cgaactcttc aaccctgacc tatgcaaccc tgagctcttc 2520

gtcgttggac gcagctgccg ccgcagctgc tgcatctgcc gccagcgccg tgcgcggaat 2580

ggccatgggc gccggctact acggcactct ggtggccaac tcgagttcca ccaataatcc 2640

cgccagcctg aacgaggaga agctgttgct gctgatggcc cagctcgagg ccttgaccca 2700

gcgcctgggc gagctgaccc agcaggtggc tcagctgcag gagcagacgc gggccgcggt 2760

[0161] tgccacggtg aaatccaaat aaaaaatgaa tcaataaata aacggagacg gttgttgatt 2820

ttaacacaga gtctgaatct ttatttgatt tttcgcgcgc ggtaggccct ggaccaccgg 2880

tctcgatcat tgagcacccg gtggatcttt tccaggaccc ggtagaggtg ggcttggatg 2940

ttgaggtaca tgggcatgag cccgtcccgg gggtggaggt agctccattg cagggcctcg 3000

tgctcggggg tggtgttgta aatcacccag tcatagcagg ggcgcagggc atggtgttgc 3060

acaatatctt tgaggaggag actgatggcc acgggcagcc ctttggtgta ggtgtttaca 3120

aatctgttga gctgggaggg atgcatgcgg ggggagatga ggtgcatctt ggcctggatc 3180

ttgagattgg cgatgttacc gcccagatcc cgcctggggt tcatgttgtg caggaccacc 3240

agcacggtgt atccggtgca cttggggaat ttatcatgca acttggaagg gaaggcgtga 3300

aagaatttgg cgacgccttt gtgcccgccc aggttttcca tgcactcatc catgatgatg 3360

gcgatgggcc cgtgggcggc ggcctgggca aagacgtttc gggggtcgga cacatcatag 3420

ttgtggtcct gggtgaggtc atcataggcc attttaatga atttggggcg gagggtgccg 3480

gactggggga caaaggtacc ctcgatcccg ggggcgtagt tcccctcaca gatctgcatc 3540

tcccaggctt tgagctcgga ggggggggatc atgtccacct gcggggcgat aaagaacacg 3600

gtttccgggg cgggggagat gagctgggcc gaaagcaagt tccggagcag ctgggacttg 3660

ccgcagccgg tggggccgta gatgaccccg atgaccggct gcaggtggta gttgagggag 3720

agacagctgc cgtcctcccg gaggaggggg gccacctcgt tcatcatctc gcgcacgtgc 3780

atgttctcgc gcaccagttc cgccaggagg cgctctcccc ccagggatag gagctcctgg 3840

agcgaggcga agtttttcag cggcttgagt ccgtcggcca tgggcatttt ggagagggtt 3900

tgttgcaaga gttccaggcg gtcccagagc tcggtgatgt gctctacggc atctcgatcc 3960

agcagacctc ctcgtttcgc gggttgggac ggctgcggga gtagggcacc agacgatggg 4020

cgtccagcgc agccagggtc cggtccttcc agggtcgcag cgtccgcgtc agggtggtct 4080

ccgtcacggt gaaggggtgc gcgccgggct gggcgcttgc gagggtgcgc ttcaggctca 4140

tccggctggt cgaaaaccgc tcccgatcgg cgccctgcgc gtcggccagg tagcaattga 4200

ccatgagttc gtagttgagc gcctcggccg cgtggccttt ggcgcggagc ttacctttgg 4260

[0162] aagtctgccc gcaggcggga cagaggaggg acttgagggc gtagagcttg ggggcgagga 4320

agacggactc gggggcgtag gcgtccgcgc cgcagtgggc gcagacggtc tcgcactcca 4380

cgagccaggt gaggtcgggc tggtcggggt caaaaaccag tttcccgccg ttctttttga 4440

tgcgtttctt acctttggtc tccatgagct cgtgtccccg ctgggtgaca aagaggctgt 4500

ccgtgtcccc gtagaccgac tttatgggcc ggtcctcgag cggtgtgccg cggtcctcct 4560

cgtagaggaa ccccgcccac tccgagacga aagcccgggt ccaggccagc acgaaggagg 4620

ccacgtggga cgggtagcgg tcgttgtcca ccagcgggtc cacctttcc agggtatgca 4680

aacacatgtc cccctcgtcc acatccagga aggtgattgg cttgtaagtg taggccacgt 4740

gaccgggggt cccggccggg ggggtataaa agggtgcggg tccctgctcg tcctcactgt 4800

cttccggatc gctgtccagg agcgccagct gttggggtag gtattccctc tcgaaggcgg 4860

gcatgacctc ggcactcagg ttgtcagttt ctagaaacga ggaggatttg atattgacgg 4920

tgccggcgga gatgcctttc aagagcccct cgtccatctg gtcagaaaag acgatctttt 4980

tgttgtcgag cttggtggcg aaggagccgt agagggcgtt ggagaggagc ttggcgatgg 5040

agcgcatggt ctggtttttt tccttgtcgg cgcgctcctt ggcggcgatg ttgagctgca 5100

cgtactcgcg cgccacgcac ttccattcgg ggaagacggt ggtcagctcg tcgggcacga 5160

ttctgacctg ccagccccga ttatgcaggg tgatgaggtc cacactggtg gccacctcgc 5220

cgcgcagggg ctcattagtc cagcagaggc gtccgcccttgcgcgagcag aaggggggca 5280

gggggtccag catgacctcg tcgggggggt cggcatcgat ggtgaagatg ccgggcagga 5340

ggtcggggtc aaagtagctg atggaagtgg ccagatcgtc cagggcagct tgccattcgc 5400

gcacggccag cgcgcgctcg tagggactga ggggcgtgcc ccagggcatg ggatgggtaa 5460

gcgcggaggc gtacatgccg cagatgtcgt agacgtagag gggctcctcg aggatgccga 5520

tgtaggtggg gtagcagcgc cccccgcgga tgctggcgcg cacgtagtca tacagctcgt 5580

gcgaggggggc gaggagcccc gggcccaggt tggtgcgact gggcttttcg gcgcggtaga 5640

cgatctggcg gaaaatggca tgcgagttgg aggagatggt gggcctttgg aagatgttga 5700

agtgggcgtg gggcagtccg accgagtcgc ggatgaagtg ggcgtaggag tcttgcagct 5760

[0163] tggcgacgag ctcggcggtg actaggacgt ccagagcgca gtagtcgagg gtctcctgga 5820

tgatgtcata cttgagctgt ccctttttgtt tccacagctc gcggttgaga aggaactctt 5880

cgcggtcctt ccagtactct tcgaggggga acccgtcctg atctgcacgg taagagccta 5940

gcatgtagaa ctggttgacg gccttgtagg cgcagcagcc cttctccacg gggagggcgt 6000

aggcctgggc ggccttgcgc agggaggtgt gcgtgagggc gaaagtgtcc ctgaccatga 6060

ccttgaggaa ctggtgcttg aagtcgatat cgtcgcagcc cccctgctcc cagagctgga 6120

agtccgtgcg cttcttgtag gcggggttgg gcaaagcgaa agtaacatcg ttgaagagga 6180

tcttgcccgc gcggggcata aagttgcgag tgatgcggaa aggttggggc acctcggccc 6240

ggttgttgat gacctgggcg gcgagcacga tctcgtcgaa gccgttgatg ttgtggccca 6300

cgatgtagag ttccacgaat cgcggacggc ccttgacgtg gggcagtttc ttgagctcct 6360

cgtaggtgag ctcgtcgggg tcgctgagcc cgtgctgctc gagcgcccag tcggcgagat 6420

gggggttggc gcggaggaag gaagtccaga gatccacggc cagggcggtt tgcagacggt 6480

cccggtactg acggaactgc tgcccgacgg ccatttttttc gggggtgacg cagtagaagg 6540

tgcgggggtc cccgtgccag cgatcccatt tgagctggag ggcgagatcg agggcgagct 6600

cgacgagccg gtcgtccccg gagagtttca tgaccagcat gaaggggacg agctgcttgc 6660

cgaaggaccc catccaggtg taggtttcca catcgtaggt gaggaagagc ctttcggtgc 6720

gaggatgcga gccgatgggg aagaactgga tctcctgcca ccaattggag gaatggctgt 6780

tgatgtgatg gaagtagaaa tgccgacggc gcgccgaaca ctcgtgcttg tgtttataca 6840

agcggccaca gtgctcgcaa cgctgcacgg gatgcacgtg ctgcacgagc tgtacctgag 6900

ttcctttgac gaggaatttc agtgggaagt ggagtcgtgg cgcctgcatc tcgtgctgta 6960

ctacgtcgtg gtggtcggcc tggccctctt ctgcctcgat ggtggtcatg ctgacgagcc 7020

cgcgcgggag gcaggtccag acctcggcgc gagcgggtcg gagagcgagg acgagggcgc 7080

gcaggccgga gctgtccagg gtcctgagac gctgcggagt caggtcagtg ggcagcggcg 7140

gcgcgcggtt gacttgcagg agtttttcca gggcgcgcgg gaggtccaga tggtacttga 7200

tctccaccgc gccattggtg gcgacgtcga tggcttgcag ggtcccgtgc ccctggggtg 7260

[0164] tgaccaccgt cccccgtttc ttcttgggcg gctggggcga cgggggcggt gcctcttcca 7320

tggttagaag cggcggcgag gacgcgcgcc gggcggcagg ggcggctcgg ggcccggagg 7380

caggggcggc aggggcacgt cggcgccgcg cgcgggtagg ttctggtact gcgcccggag 7440

aagactggcg tgagcgacga cgcgacggtt gacgtcctgg atctgacgcc tctgggtgaa 7500

ggccacggga cccgtgagtt tgaacctgaa agagagttcg acagaatcaa tctcggtatc 7560

gttgacggcg gcctgccgca ggatctcttg cacgtcgccc gagttgtcct ggtaggcgat 7620

ctcggtcatg aactgctcga tctcctcctc ttgaaggtct ccgcggccgg cgcgctccac 7680

ggtggccgcg aggtcgttgg agatgcggcc catgagctgc gagaaggcgt tcatgcccgc 7740

ctcgttccag acgcggctgt agaccacgac gccctcggga tcgcgggcgc gcatgaccac 7800

ctgggcgagg ttgagctcca cgtggcgcgt gaagaccgcg tagttgcaga ggcgctggta 7860

gaggtagttg agcgtggtgg cgatgtgctc ggtgacgaag aaatacatga tccagcggcg 7920

gagcggcatc tcgctgacgt cgcccagcgc ctccaaacgt tccatggcct cgtaaaagtc 7980

cacggcgaag ttgaaaaact gggagttgcg cgccgagacg gtcaactcct cctccagaag 8040

acggatgagc tcggcgatgg tggcgcgcac ctcgcgctcg aaggcccccg ggagttcctc 8100

cacttcctct tcttcctcct ccactaacat ctcttctact tcctcctcag gcggcagtgg 8160

tggcggggga gggggcctgc gtcgccggcg gcgcacgggc agacggtcga tgaagcgctc 8220

gatggtctcg ccgcgccggc gtcgcatggt ctcggtgacg gcgcgcccgt cctcgcgggg 8280

ccgcagcgtg aagacgccgc cgcgcatctc caggtggccg gggggtccc cgttgggcag 8340

ggagagggcg ctgacgatgc atcttatcaa ttgccccgta gggactccgc gcaaggacct 8400

gagcgtctcg agatccacgg gatctgaaaa ccgctgaacg aaggcttcga gccagtcgca 8460

gtcgcaaggt aggctgagca cggtttcttc tggcgggtca tgttggttgg gagcggggcg 8520

ggcgatgctg ctggtgatga agttgaaata ggcggttctg agacggcgga tggtggcgag 8580

gagcaccagg tctttgggcc cggcttgctg gatgcgcaga cggtcggcca tgccccaggc 8640

gtggtcctga cacctggcca ggtccttgta gtagtcctgc atgagccgct ccacgggcac 8700

ctcctcctcg cccgcgcggc cgtgcatgcg cgtgagcccg aagccgcgct ggggctggac 8760

[0165] gagcgccagg tcggcgacga cgcgctcggc gaggatggct tgctggatct gggtgagggt 8820

ggtctggaag tcatcaaagt cgacgaagcg gtggtaggct ccggtgttga tggtgtagga 8880

gcagttggcc atgacggacc agttgacggt ctggtggccc ggacgcacga gctcgtggta 8940

cttgaggcgc gagtaggcgc gcgtgtcgaa gatgtagtcg ttgcaggtgc gcaccaggta 9000

ctggtagccg atgaggaagt gcggcggcgg ctggcggtag agcggccatc gctcggtggc 9060

gggggcgccg ggcgcgaggt cctcgagcat ggtgcggtgg tagccgtaga tgtacctgga 9120

catccaggtg atgccggcgg cggtggtgga ggcgcgcggg aactcgcgga cgcggttcca 9180

gatgttgcgc agcggcagga agtagttcat ggtgggcacg gtctggcccg tgaggcgcgc 9240

gcagtcgtgg atgctctata cgggcaaaaa cgaaagcggt cagcggctcg actccgtggc 9300

ctggaggcta agcgaacggg ttgggctgcg cgtgtacccc ggttcgaatc tcgaatcagg 9360

ctggagccgc agctaacgtg gtattggcac tcccgtctcg acccaagcct gcaccaaccc 9420

tccaggatac ggaggcgggt cgttttgcaa ctttttttg gaggccggat gagactagta 9480

agcgcggaaa gcggccgacc gcgatggctc gctgccgtag tctggagaag aatcgccagg 9540

gttgcgttgc ggtgtgcccc ggttcgaggc cggccggatt ccgcggctaa cgagggcgtg 9600

gctgccccgt cgtttccaag accccatagc cagccgactt ctccagttac ggagcgagcc 9660

cctcttttgt tttgtttgtt tttgccagat gcatcccgta ctgcggcaga tgcgccccca 9720

ccaccctcca ccgcaacaac agccccctcc acagccggcg cttctgcccc cgccccagca 9780

gcaacttcca gccacgaccg ccgcggccgc cgtgagcggg gctggacaga gttatgatca 9840

ccagctggcc ttggaagagg gcgaggggct ggcgcgcctg ggggcgtcgt cgccggagcg 9900

gcacccgcgc gtgcagatga aaagggacgc tcgcgaggcc tacgtgccca agcagaacct 9960

gttcagagac aggagcggcg aggagcccga ggagatgcgc gcggcccggt tccacgcggg 10020

gcgggagctg cggcgcggcc tggaccgaaa gagggtgctg agggacgagg atttcgaggc 10080

ggacgagctg acggggatca gccccgcgcg cgcgcacgtg gccgcggcca acctggtcac 10140

ggcgtacgag cagaccgtga aggaggagag caacttccaa aaatccttca acaaccacgt 10200

gcgcaccctg atcgcgcgcg aggaggtgac cctgggcctg atgcacctgt gggacctgct 10260

[0166] ggaggccatc gtgcagaacc ccaccagcaa gccgctgacg gcgcagctgt tcctggtggt 10320

gcagcatagt cgggacaacg aagcgttcag ggaggcgctg ctgaatatca ccgagcccga 10380

gggccgctgg ctcctggacc tggtgaacat tctgcagagc atcgtggtgc aggagcgcgg 10440

gctgccgctg tccgagaagc tggcggccat caacttctcg gtgctgagtt tgggcaagta 10500

ctacgctagg aagatctaca agaccccgta cgtgcccata gacaaggagg tgaagatcga 10560

cgggttttac atgcgcatga ccctgaaagt gctgaccctg agcgacgatc tgggggtgta 10620

ccgcaacgac aggatgcacc gtgcggtgag cgccagcagg cggcgcgagc tgagcgacca 10680

ggagctgatg catagtctgc agcgggccct gaccggggcc gggaccgagg gggagagcta 10740

ctttgacatg ggcgcggacc tgcactggca gcccagccgc cgggccttgg aggcggcggc 10800

aggaccctac gtagaagagg tggacgatga ggtggacgag gagggcgagt acctggaaga 10860

ctgatggcgc gaccgtattt ttgctagatg caacaacaac agccacctcc tgatcccgcg 10920

atgcgggcgg cgctgcagag ccagccgtcc ggcattaact cctcggacga ttggacccag 10980

gccatgcaac gcatcatggc gctgacgacc cgcaaccccg aagcctttag acagcagccc 11040

caggccaacc ggctctcggc catcctggag gccgtggtgc cctcgcgctc caaccccacg 11100

cacgagaagg tcctggccat cgtgaacgcg ctggtggaga acaaggccat ccgcggcgac 11160

gaggccggcc tggtgtacaa cgcgctgctg gagcgcgtgg cccgctacaa cagcaccaac 11220

gtgcagacca acctggaccg catggtgacc gacgtgcgcg aggccgtggc ccagcgcgag 11280

cggttccacc gcgagtccaa cctgggatcc atggtggcgc tgaacgcctt cctcagcacc 11340

cagcccgcca acgtgccccg gggccaggag gactacacca acttcatcag cgccctgcgc 11400

ctgatggtga ccgaggtgcc ccagagcgag gtgtaccagt ccgggccgga ctacttcttc 11460

cagaccagtc gccagggctt gcagaccgtg aacctgagcc aggctttcaa gaacttgcag 11520

ggcctgtggg gcgtgcaggc cccggtcggg gaccgcgcga cggtgtcgag cctgctgacg 11580

ccgaactcgc gcctgctgct gctgctggtg gccccccttca cggacagcgg cagcatcaac 11640

cgcaactcgt acctgggcta cctgattaac ctgtaccgcg aggccatcgg ccaggcgcac 11700

gtggacgagc agacctacca ggagatcacc cacgtgagcc gcgccctggg ccaggacgac 11760

[0167] ccgggcaacc tggaagccac cctgaacttt ttgctgacca accggtcgca gaagatcccg 11820

ccccagtacg cgctcagcac cgaggaggag cgcatcctgc gttacgtgca gcagagcgtg 11880

ggcctgttcc tgatgcagga gggggccacc cccagcgccg cgctcgacat gaccgcgcgc 11940

aacatggagc ccagcatgta cgccagcaac cgcccgttca tcaataaact gatggactac 12000

ttgcatcggg cggccgccat gaactctgac tatttcacca acgccatcct gaatccccac 12060

tggctcccgc cgccggggtt ctacacgggc gagtacgaca tgcccgaccc caatgacggg 12120

ttcctgtggg acgatgtgga cagcagcgtg ttctcccccc gaccgggtgc taacgagcgc 12180

cccttgtgga agaaggaagg cagcgaccga cgcccgtcct cggcgctgtc cggccgcgag 12240

ggtgctgccg cggcggtgcc cgaggccgcc agtcctttcc cgagcttgcc cttctcgctg 12300

aacagtatcc gcagcagcga gctgggcagg atcacgcgcc cgcgcttgct gggcgaagag 12360

gagtacttga atgactcgct gttgagaccc gagcgggaga agaacttccc caataacggg 12420

atagaaagcc tggtggacaa gatgagccgc tggaagacgt atgcgcagga gcacagggac 12480

gatccccggg cgtcgcaggg ggccacgagc cggggcagcg ccgcccgtaa acgccggtgg 12540

cacgacaggc agcggggaca gatgtgggac gatgaggact ccgccgacga cagcagcgtg 12600

ttggacttgg gtgggagtgg taacccgttc gctcacctgc gcccccgtat cgggcgcatg 12660

atgtaagaga aaccgaaaat aaatgatact caccaaggcc atggcgacca gcgtgcgttc 12720

gtttcttctc tgttgttgtt gtatctagta tgatgaggcg tgcgtacccg gagggtcctc 12780

ctccctcgta cgagagcgtg atgcagcagg cgatggcggc ggcggcgatg cagcccccgc 12840

tggaggctcc ttacgtgccc ccgcggtacc tggcgcctac ggaggggcgg aacagcattc 12900

gttactcgga gctggcaccc ttgtacgata ccacccggtt gtacctggtg gacaacaagt 12960

cggcggacat cgcctcgctg aactaccaga acgaccacag caacttcctg accaccgtgg 13020

tgcagaacaa tgacttcacc cccacggagg ccagcaccca gaccatcaac tttgacgagc 13080

gctcgcggtg gggcggccag ctgaaaacca tcatgcacac caacatgccc aacgtgaacg 13140

agttcatgta cagcaacaag ttcaaggcgc gggtgatggt ctcccgcaag acccccaatg 13200

gggtgacagt gacagaggat tatgatggta gtcaggatga gctgaagtat gaatgggtgg 13260

[0168] aatttgagct gcccgaaggc aacttctcgg tgaccatgac catcgacctg atgaacaacg 13320

ccatcatcga caattacttg gcggtggggc ggcagaacgg ggtgctggag agcgacatcg 13380

gcgtgaagtt cgacactagg aacttcaggc tgggctggga ccccgtgacc gagctggtca 13440

tgcccggggt gtacaccaac gaggctttcc atcccgatat tgtcttgctg cccggctgcg 13500

gggtggactt caccgagagc cgcctcagca acctgctggg cattcgcaag aggcagccct 13560

tccaggaagg cttccagatc atgtacgagg atctggaggg gggcaacatc cccgcgctcc 13620

tggatgtcga cgcctatgag aaaagcaagg aggatgcagc agctgaagca actgcagccg 13680

tagctaccgc ctctaccgag gtcaggggcg ataattttgc aagcgccgca gcagtggcag 13740

cggccgaggc ggctgaaacc gaaagtaaga tagtcattca gccggtggag aaggatagca 13800

agaacaggag ctacaacgta ctaccggaca agataaacac cgcctaccgc agctggtacc 13860

tagcctacaa ctatggcgac cccgagaagg gcgtgcgctc ctggacgctg ctcaccacct 13920

cggacgtcac ctgcggcgtg gagcaagtct actggtcgct gcccgacatg atgcaagacc 13980

cggtcacctt ccgctccacg cgtcaagtta gcaactaccc ggtggtgggc gccgagctcc 14040

tgcccgtcta ctccaagagc ttcttcaacg agcaggccgt ctactcgcag cagctgcgcg 14100

ccttcacctc gcttacgcac gtcttcaacc gcttccccga gaaccagatc ctcgtccgcc 14160

cgcccgcgcc caccattacc accgtcagtg aaaacgttcc tgctctcaca gatcacggga 14220

ccctgccgct gcgcagcagt atccggggag tccagcgcgt gaccgttact gacgccagac 14280

gccgcacctg cccctacgtc tacaaggccc tgggcatagt cgcgccgcgc gtcctctcga 14340

gccgcacctt ctaaatgtcc attctcatct cgcccagtaa taacaccggt tggggcctgc 14400

gcgcgcccag caagatgtac ggaggcgctc gccaacgctc cacgcaacac cccgtgcgcg 14460

tgcgcgggca cttccgcgct ccctggggcg ccctcaaggg ccgcgtgcgg tcgcgcacca 14520

ccgtcgacga cgtgatcgac caggtggtgg ccgacgcgcg caactacacc cccgccgccg 14580

cgcccgtctc caccgtggac gccgtcatcg acagcgtggt ggccgacgcg cgccggtacg 14640

cccgcgccaa gagccggcgg cggcgcatcg cccggcggca ccggagcacc cccgccatgc 14700

gcgcggcgcg agccttgctg cgcagggcca ggcgcacggg acgcagggcc atgctcaggg 14760

[0169] cggccagacg cgcggcttca ggcgccagcg ccggcaggac ccggagacgc gcggccacgg 14820

cggcggcagc ggccatcgcc agcatgtccc gcccgcggcg agggaacgtg tactgggtgc 14880

gcgacgccgc caccggtgtg cgcgtgcccg tgcgcacccg ccccccctcgc acttgaagat 14940

gttcacttcg cgatgttgat gtgtcccagc ggcgaggagg atgtccaagc gcaaattcaa 15000

ggaagagatg ctccaggtca tcgcgcctga gatctacggc cctgcggtgg tgaaggagga 15060

aagaaagccc cgcaaaatca agcgggtcaa aaaggacaaa aaggaagaag aaagtgatgt 15120

ggacggattg gtggagtttg tgcgcgagtt cgccccccgg cggcgcgtgc agtggcgcgg 15180

gcggaaggtg caaccggtgc tgagacccgg caccaccgtg gtcttcacgc ccggcgagcg 15240

ctccggcacc gcttccaagc gctcctacga cgaggtgtac ggggatgatg atattctgga 15300

gcaggcggcc gagcgcctgg gcgagtttgc ttacggcaag cgcagccgtt ccgcaccgaa 15360

ggaagaggcg gtgtccatcc cgctggacca cggcaacccc acgccgagcc tcaagcccgt 15420

gaccttgcag caggtgctgc cgaccgcggc gccgcgccgg gggttcaagc gcgagggcga 15480

ggatctgtac cccaccatgc agctgatggt gcccaagcgc cagaagctgg aagacgtgct 15540

ggagaccatg aaggtggacc cggacgtgca gcccgaggtc aaggtgcggc ccatcaagca 15600

ggtggccccg ggcctgggcg tgcagaccgt ggacatcaag attcccacgg agcccatgga 15660

aacgcagacc gagcccatga tcaagcccag caccagcacc atggaggtgc agacggatcc 15720

ctggatgcca tcggctccta gtcgaagacc ccggcgcaag tacggcgcgg ccagcctgct 15780

gatgcccaac tacgcgctgc atccttccat catccccacg ccgggctacc gcggcacgcg 15840

cttctaccgc ggtcatacca gcagccgccg ccgcaagacc accactcgcc gccgccgtcg 15900

ccgcaccgcc gctgcaacca cccctgccgc cctggtgcgg agagtgtacc gccgcggccg 15960

cgcacctctg accctgccgc gcgcgcgcta ccacccgagc atcgccattt aaactttcgc 16020

ctgctttgca gatcaatggc cctcacatgc cgccttcgcg ttcccattac gggctaccga 16080

ggaagaaaac cgcgccgtag aaggctggcg gggaacggga tgcgtcgcca ccaccaccgg 16140

cggcggcgcg ccatcagcaa gcggttgggg ggaggcttcc tgcccgcgct gatccccatc 16200

atcgccgcgg cgatcggggc gatccccggc attgcttccg tggcggtgca ggcctctcag 16260

[0170] cgccactgag acacacttgg aaacatcttg taataaacca atggactctg acgctcctgg 16320

tcctgtgatg tgttttcgta gacagatgga agacatcaat ttttcgtccc tggctccgcg 16380

acacggcacg cggccgttca tgggcacctg gagcgacatc ggcaccagcc aactgaacgg 16440

gggcgccttc aattggagca gtctctggag cgggcttaag aatttcgggt ccacgcttaa 16500

aacctatggc agcaaggcgt ggaacagcac cacagggcag gcgctgaggg ataagctgaa 16560

agagcagaac ttccagcaga aggtggtcga tgggctcgcc tcgggcatca acggggtggt 16620

ggacctggcc aaccaggccg tgcagcggca gatcaacagc cgcctggacc cggtgccgcc 16680

cgccggctcc gtggagatgc cgcaggtgga ggaggagctg cctcccctgg acaagcgggg 16740

cgagaagcga ccccgccccg atgcggagga gacgctgctg acgcacacgg acgagccgcc 16800

cccgtacgag gaggcggtga aactgggtct gcccaccacg cggcccatcg cgcccctggc 16860

caccggggtg ctgaaacccg aaaagcccgc gaccctggac ttgcctcctc cccagccttc 16920

ccgcccctct acagtggcta agcccctgcc gccggtggcc gtggcccgcg cgcgacccgg 16980

gggcaccgcc cgccctcatg cgaactggca gagcactctg aacagcatcg tgggtctggg 17040

agtgcagagt gtgaagcgcc gccgctgcta ttaaacctac cgtagcgctt aacttgcttg 17100

tctgtgtgtg tatgtattat gtcgccgccg ccgctgtcca ccagaaggag gagtgaagag 17160

gcgcgtcgcc gagttgcaag atggccaccc catcgatgct gccccagtgg gcgtacatgc 17220

acatcgccgg acaggacgct tcggagtacc tgagtccggg tctggtgcag tttgcccgcg 17280

ccacagacac ctacttcagt ctggggaaca agtttaggaa ccccacggtg gcgcccacgc 17340

acgatgtgac caccgaccgc agccagcggc tgacgctgcg cttcgtgccc gtggaccgcg 17400

aggacaacac ctactcgtac aaagtgcgct acacgctggc cgtgggcgac aaccgcgtgc 17460

tggacatggc cagcacctac tttgacatcc gcggcgtgct ggatcggggc cctagcttca 17520

aaccctactc cggcaccgcc tacaacagtc tggcccccaa gggagcaccc aacacttgtc 17580

agtggacata taaagccgat ggtgaaactg ccacagaaaa aacctataca tatggaaatg 17640

cacccgtgca gggcattaac atcacaaaag atggtattca acttggaact gacaccgatg 17700

atcagccaat ctacgcagat aaaacctatc agcctgaacc tcaagtgggt gatgctgaat 17760

[0171] ggcatgacat cactggtact gatgaaaagt atggaggcag agctcttaag cctgatacca 17820

aaatgaagcc ttgttatggt tcttttgcca agcctactaa taaagaagga ggtcaggcaa 17880

atgtgaaaac aggaacaggc actactaaag aatatgacat agacatggct ttctttgaca 17940

acagaagtgc ggctgctgct ggcctagctc cagaaattgt tttgtatact gaaaatgtgg 18000

atttggaaac tccagatacc catattgtat acaaagcagg cacagatgac agcagctctt 18060

ctattaattt gggtcagcaa gccatgccca acagacctaa ctacattggt ttcagagaca 18120

actttatcgg gctcatgtac tacaacagca ctggcaatat gggggtgctg gccggtcagg 18180

cttctcagct gaatgctgtg gttgacttgc aagacagaaa caccgagctg tcctaccagc 18240

tcttgcttga ctctctgggt gacagaaccc ggtatttcag tatgtggaat caggcggtgg 18300

acagctatga tcctgatgtg cgcattattg aaaatcatgg tgtggaggat gaacttccca 18360

actattgttt ccctctggat gctgttggca gaacagatac ttatcaggga attaaggcta 18420

atggaactga tcaaaccaca tggaccaaag atgacagtgt caatgatgct aatgagatag 18480

gcaagggtaa tccattcgcc atggaaatca acatccaagc caacctgtgg aggaacttcc 18540

tctacgccaa cgtggccctg tacctgcccg actcttacaa gtacacgccg gccaatgtta 18600

ccctgcccac caacaccaac acctacgatt acatgaacgg ccgggtggtg gcgccctcgc 18660

tggtggactc ctacatcaac atcggggcgc gctggtcgct ggatcccatg gacaacgtga 18720

acccccttcaa ccaccaccgc aatgcggggc tgcgctaccg ctccatgctc ctgggcaacg 18780

ggcgctacgt gcccttccac atccaggtgc cccagaaatt tttcgccatc aagagcctcc 18840

tgctcctgcc cgggtcctac acctacgagt ggaacttccg caaggacgtc aacatgatcc 18900

tgcagagctc cctcggcaac gacctgcgca cggacggggc ctccatctcc ttcaccagca 18960

tcaacctcta cgccaccttc ttccccatgg cgcacaacac ggcctccacg ctcgaggcca 19020

tgctgcgcaa cgacaccaac gaccagtcct tcaacgacta cctctcggcg gccaacatgc 19080

tctaccccat cccggccaac gccaccaacg tgcccatctc catcccctcg cgcaactggg 19140

ccgccttccg cggctggtcc ttcacgcgtc tcaagaccaa ggagacgccc tcgctgggct 19200

ccgggttcga cccctacttc gtctactcgg gctccatccc ctacctcgac ggcaccttct 19260

[0172] acctcaacca caccttcaag aaggtctcca tcaccttcga ctcctccgtc agctggcccg 19320

gcaacgaccg gctcctgacg cccaacgagt tcgaaatcaa gcgcaccgtc gacggcgagg 19380

gctacaacgt ggcccagtgc aacatgacca aggactggtt cctggtccag atgctggccc 19440

actacaacat cggctaccag ggcttctacg tgcccgaggg ctacaaggac cgcatgtact 19500

ccttcttccg caacttccag cccatgagcc gccaggtggt ggacgaggtc aactacaagg 19560

actaccaggc cgtcaccctg gcctaccagc acaacaactc gggcttcgtc ggctacctcg 19620

cgcccaccat gcgccagggc cagccctacc ccgccaacta cccctacccg ctcatcggca 19680

agagcgccgt caccagcgtc acccagaaaa agttcctctg cgacagggtc atgtggcgca 19740

tccccttctc cagcaacttc atgtccatgg gcgcgctcac cgacctcggc cagaacatgc 19800

tctatgccaa ctccgcccac gcgctagaca tgaatttcga agtcgacccc atggatgagt 19860

ccacccttct ctatgttgtc ttcgaagtct tcgacgtcgt ccgagtgcac cagccccacc 19920

gcggcgtcat cgaggccgtc tacctgcgca ccccccttctc ggccggtaac gccaccacct 19980

aagctcttgc ttcttgcaag ccatggccgc gggctccggc gagcaggagc tcagggccat 20040

catccgcgac ctgggctgcg ggccctactt cctgggcacc ttcgataagc gcttcccggg 20100

attcatggcc ccgcacaagc tggcctgcgc catcgtcaac acggccggcc gcgagaccgg 20160

gggcgagcac tggctggcct tcgcctggaa cccgcgctcg aacacctgct acctcttcga 20220

ccccttcggg ttctcggacg agcgcctcaa gcagatctac cagttcgagt acgagggcct 20280

gctgcgccgc agcgccctgg ccaccgagga ccgctgcgtc accctggaaa agtccaccca 20340

gaccgtgcag ggtccgcgct cggccgcctg cgggctcttc tgctgcatgt tcctgcacgc 20400

cttcgtgcac tggcccgacc gccccatgga caagaacccc accatgaact tgctgacggg 20460

ggtgcccaac ggcatgctcc agtcgcccca ggtggaaccc accctgcgcc gcaaccagga 20520

ggcgctctac cgcttcctca actcccactc cgcctacttt cgctcccacc gcgcgcgcat 20580

cgagaaggcc accgccttcg accgcatgaa tcaagacatg taaaccgtgt gtgtatgtta 20640

aatgtcttta ataaacagca ctttcatgtt acacatgcat ctgagatgat ttatttagaa 20700

atcgaaaggg ttctgccggg tctcggcatg gcccgcgggc agggacacgt tgcggaactg 20760

[0173] gtacttggcc agccacttga actcggggat cagcagtttg ggcagcgggg tgtcggggaa 20820

ggagtcggtc cacagcttcc gcgtcagttg cagggcgccc agcaggtcgg gcgcggagat 20880

cttgaaatcg cagttgggac ccgcgttctg cgcgcgggag ttgcggtaca cggggttgca 20940

gcactggaac accatcaggg ccgggtgctt cacgctcgcc agcaccgtcg cgtcggtgat 21000

gctctccacg tcgaggtcct cggcgttggc catcccgaag ggggtcatct tgcaggtctg 21060

ccttcccatg gtgggcacgc acccgggctt gtggttgcaa tcgcagtgca gggggatcag 21120

catcatctgg gcctggtcgg cgttcatccc cgggtacatg gccttcatga aagcctccaa 21180

ttgcctgaac gcctgctggg ccttggctcc ctcggtgaag aagaccccgc aggacttgct 21240

agagaactgg ttggtggcgc acccggcgtc gtgcacgcag cagcgcgcgt cgttgttggc 21300

cagctgcacc acgctgcgcc cccagcggtt ctgggtgatc ttggcccggt cggggttctc 21360

cttcagcgcg cgctgcccgt tctcgctcgc cacatccatc tcgatcatgt gctccttctg 21420

gatcatggtg gtcccgtgca ggcaccgcag cttgccctcg gcctcggtgc acccgtgcag 21480

ccacagcgcg cacccggtgc actcccagtt cttgtgggcg atctgggaat gcgcgtgcac 21540

gaagccctgc aggaagcggc ccatcatggt ggtcagggtc ttgttgctag tgaaggtcag 21600

cggaatgccg cggtgctcct cgttgatgta caggtggcag atgcggcggt acacctcgcc 21660

ctgctcgggc atcagctgga agttggcttt caggtcggtc tccacgcggt agcggtccat 21720

cagcatagtc atgatttcca tacccttctc ccaggccgag acgatgggca ggctcatagg 21780

gttcttcacc atcatcttag cgctagcagc cgcggccagg gggtcgctct cgtccagggt 21840

ctcaaagctc cgcttgccgt ccttctcggt gatccgcacc ggggggtagc tgaagcccac 21900

ggccgccagc tcctcctcgg cctgtctttc gtcctcgctg tcctggctga cgtcctgcag 21960

gaccacatgc ttggtcttgc ggggtttctt cttgggcggc agcggcggcg gagatgttgg 22020

agatggcgag ggggagcgcg agttctcgct caccactact atctcttcct cttcttggtc 22080

cgaggccacg cggcggtagg tatgtctctt cgggggcaga ggcggaggcg acgggctctc 22140

gccgccgcga cttggcggat ggctggcaga gcccccttccg cgttcggggg tgcgctcccg 22200

gcggcgctct gactgacttc ctccgcggcc ggccattgtg ttctcctagg gaggaacaac 22260

[0174] aagcatggag actcagccat cgccaacctc gccatctgcc cccaccgccg acgagaagca 22320

gcagcagcag aatgaaagct taaccgcccc gccgcccagc cccgccacct ccgacgcggc 22380

cgtcccagac atgcaagaga tggaggaatc catcgagatt gacctgggct atgtgacgcc 22440

cgcggagcac gaggaggagc tggcagtgcg cttttcacaa gaagagatac accaagaaca 22500

gccagagcag gaagcagaga atgagcagag tcaggctggg ctcgagcatg acggcgacta 22560

cctccacctg agcgggggggg aggacgcgct catcaagcat ctggcccggc aggccaccat 22620

cgtcaaggat gcgctgctcg accgcaccga ggtgcccctc agcgtggagg agctcagccg 22680

cgcctacgag ttgaacctct tctcgccgcg cgtgcccccc aagcgccagc ccaatggcac 22740

ctgcgagccc aacccgcgcc tcaacttcta cccggtcttc gcggtgcccg aggccctggc 22800

cacctaccac atctttttca agaaccaaaa gatccccgtc tcctgccgcg ccaaccgcac 22860

ccgcgccgac gccctttttca acctgggtcc cggcgcccgc ctacctgata tcgcctcctt 22920

ggaagaggtt cccaagatct tcgagggtct gggcagcgac gagactcggg ccgcgaacgc 22980

tctgcaagga gaaggaggag agcatgagca ccacagcgcc ctggtcgagt tggaaggcga 23040

caacgcgcgg ctggcggtgc tcaaacgcac ggtcgagctg acccatttcg cctacccggc 23100

tctgaacctg ccccccaaag tcatgagcgc ggtcatggac caggtgctca tcaagcgcgc 23160

gtcgcccatc tccgaggacg agggcatgca agactccgag gagggcaagc ccgtggtcag 23220

cgacgagcag ctggcccggt ggctgggtcc taatgctagt ccccagagtt tggaagagcg 23280

gcgcaaactc atgatggccg tggtcctggt gaccgtggag ctggagtgcc tgcgccgctt 23340

cttcgccgac gcggagaccc tgcgcaaggt cgaggagaac ctgcactacc tcttcaggca 23400

cgggttcgtg cgccaggcct gcaagatctc caacgtggag ctgaccaacc tggtctccta 23460

catgggcatc ttgcacgaga accgcctggg gcagaacgtg ctgcacacca ccctgcgcgg 23520

ggaggcccgg cgcgactaca tccgcgactg cgtctacctc tacctctgcc acacctggca 23580

gacgggcatg ggcgtgtggc agcagtgtct ggaggagcag aacctgaaag agctctgcaa 23640

gctcctgcag aagaacctca agggtctgtg gaccgggttc gacgagcgca ccaccgcctc 23700

ggacctggcc gacctcattt tccccgagcg cctcaggctg acgctgcgca acggcctgcc 23760

[0175] cgactttatg agccaaagca tgttgcaaaa ctttcgctct ttcatcctcg aacgctccgg 23820

aatcctgccc gccacctgct ccgcgctgcc ctcggacttc gtgccgctga ccttccgcga 23880

gtgccccccg ccgctgtgga gccactgcta cctgctgcgc ctggccaact acctggccta 23940

ccactcggac gtgatcgagg acgtcagcgg cgagggcctg ctcgagtgcc actgccgctg 24000

caacctctgc acgccgcacc gctccctggc ctgcaacccc cagctgctga gcgagaccca 24060

gatcatcggc accttcgagt tgcaagggcc cagcgaaggc gagggttcag ccgccaaggg 24120

gggtctgaaa ctcaccccgg ggctgtggac ctcggcctac ttgcgcaagt tcgtgcccga 24180

ggactaccat cccttcgaga tcaggttcta cgaggaccaa tcccatccgc ccaaggccga 24240

gctgtcggcc tgcgtcatca cccaggggGC gatcctggcc caattgcaag ccatccagaa 24300

atcccgccaa gaattcttgc tgaaaaaggg ccgcggggtc tacctcgacc cccagaccgg 24360

tgaggagctc aaccccggct tcccccagga tgccccgagg aaacaagaag ctgaaagtgg 24420

agctgccgcc cgtggaggat ttggaggaag actgggagaa cagcagtcag gcagaggagg 24480

aggagatgga ggaagactgg gacagcactc aggcagagga ggacagcctg caagacagtc 24540

tggaggaaga cgaggaggag gcagaggagg aggtggaaga agcagccgcc gccagaccgt 24600

cgtcctcggc gggggagaaa gcaagcagca cggataccat ctccgctccg ggtcggggtc 24660

ccgctcgacc acacagtaga tgggacgaga ccggacgatt cccgaacccc accacccaga 24720

ccggtaagaa ggagcggcag ggatacaagt cctggcgggg gcacaaaaac gccatcgtct 24780

cctgcttgca ggcctgcggg ggcaacatct ccttcacccg gcgctacctg ctcttccacc 24840

gcggggtgaa ctttccccgc aacatcttgc attactaccg tcacctccac agcccctact 24900

acttccaaga agaggcagca gcagcagaaa aagaccagca gaaaaccagc agctagaaaa 24960

tccacagcgg cggcagcagg tggactgagg atcgcggcga acgagccggc gcaaacccgg 25020

gagctgagga accggatctt tcccaccctc tatgccatct tccagcagag tcgggggcag 25080

gagcaggaac tgaaagtcaa gaaccgttct ctgcgctcgc tcacccgcag ttgtctgtat 25140

cacaagagcg aagaccaact tcagcgcact ctcgaggacg ccgaggctct cttcaacaag 25200

tactgcgcgc tcactcttaa agagtagccc gcgcccgccc agtcgcagaa aaaggcggga 25260

[0176] attacgtcac ctgtgccctt cgccctagcc gcctccaccc atcatcatga gcaaagagat 25320

tcccacgcct tacatgtgga gctaccagcc ccagatgggc ctggccgccg gtgccgccca 25380

ggactactcc acccgcatga attggctcag cgccgggccc gcgatgatct cacgggtgaa 25440

tgacatccgc gcccaccgaa accagatact cctagaacag tcagcgctca ccgccacgcc 25500

ccgcaatcac ctcaatccgc gtaattggcc cgccgccctg gtgtaccagg aaattcccca 25560

gcccacgacc gtactacttc cgcgagacgc ccaggccgaa gtccagctga ctaactcagg 25620

tgtccagctg gcgggcggcg ccaccctgtg tcgtcaccgc cccgctcagg gtataaagcg 25680

gctggtgatc cggggcagag gcacacagct caacgacgag gtggtgagct cttcgctggg 25740

tctgcgacct gacggagtct tccaactcgc cggatcgggg agatcttcct tcacgcctcg 25800

tcaggccgtc ctgactttgg agagttcgtc ctcgcagccc cgctcgggtg gcatcggcac 25860

tctccagttc gtggaggagt tcactccctc ggtctacttc aacccccttct ccggctcccc 25920

cggccactac ccggacgagt tcatcccgaa cttcgacgcc atcagcgagt cggtggacgg 25980

ctacgattga aactaatcac ccccttatcc agtgaaataa agatcatatt gatgatgatt 26040

ttacagaaat aaaaaataat catttgattt gaaataaaga tacaatcata ttgatgattt 26100

gagtttaaca aaaaaataaa gaatcactta cttgaaatct gataccaggt ctctgtccat 26160

gttttctgcc aacaccactt cactcccctc ttcccagctc tggtactgca ggccccggcg 26220

ggctgcaaac ttcctccaca cgctgaaggg gatgtcaaat tcctcctgtc cctcaatctt 26280

cattttatct tctatcagat gtccaaaaag cgcgtccggg tggatgatga cttcgacccc 26340

gtctaccccct acgatgcaga caacgcaccg accgtgccct tcatcaaccc ccccttcgtc 26400

tcttcagatg gattccaaga gaagcccctg ggggtgttgt ccctgcgact ggccgacccc 26460

gtcaccacca agaacgggga aatcaccctc aagctgggag agggggtgga cctcgattcc 26520

tcgggaaaac tcatctccaa cacggccacc aaggccgccg cccctctcag tttttccaac 26580

aacaccattt cccttaacat ggatcacccc ttttacacta aagatggaaa attatcctta 26640

caagtttctc caccattaaa tatactgaga acaagcattc taaacacact agctttaggt 26700

tttggatcag gtttaggact ccgtggctct gccttggcag tacagttagt ctctccactt 26760

[0177] acatttgata ctgatggaaa cataaagctt accttagaca gaggtttgca tgttacaaca 26820

ggagatgcaa ttgaaagcaa cataagctgg gctaaaggtt taaaatttga agatggagcc 26880

atagcaacca acattggaaa tgggttagag tttggaagca gtagtacaga aacaggtgtt 26940

gatgatgctt acccaatcca agttaaactt ggatctggcc ttagctttga cagtacagga 27000

gccataatgg ctggtaacaa agaagacgat aaactcactt tgtggacaac acctgatcca 27060

tcaccaaact gtcaaatact cgcagaaaat gatgcaaaac taacactttg cttgactaaa 27120

tgtggtagtc aaatactggc cactgtgtca gtcttagttg taggaagtgg aaacctaaac 27180

cccattactg gcaccgtaag cagtgctcag gtgtttctac gttttgatgc aaacggtgtt 27240

cttttaacag aacattctac actaaaaaaa tactgggggt ataggcaggg agatagcata 27300

gatggcactc catataccaa tgctgtagga ttcatgccca atttaaaagc ttatccaaag 27360

tcacaaagtt ctactactaa aaataatata gtagggcaag tatacatgaa tggagatgtt 27420

tcaaaaccta tgcttctcac tataaccctc aatggtactg atgacagcaa cagtacatat 27480

tcaatgtcat tttcatacac ctggactaat ggaagctatg ttggagcaac atttggggct 27540

aactcttata ccttctcata catcgcccaa gaatgaacac tgtatcccac cctgcatgcc 27600

aacccttccc accccactct gtggaacaaa ctctgaaaca caaaataaaa taaagttcaa 27660

gtgttttatt gattcaacag ttttacagga ttcgagcagt tatttttcct ccaccctccc 27720

aggacatgga atacaccacc ctctcccccc gcacagcctt gaacatctga atgccattgg 27780

tgatggacat gcttttggtc tccacgttcc acacagtttc agagcgagcc agtctcgggt 27840

cggtcaggga gatgaaaccc tccgggcact cccgcatctg cacctcacag ctcaacagct 27900

gaggattgtc ctcggtggtc gggatcacgg ttatctggaa gaagcagaag agcggcggtg 27960

ggaatcatag tccgcgaacg ggatcggccg gtggtgtcgc atcaggcccc gcagcagtcg 28020

ctgccgccgc cgctccgtca agctgctgct caggggggtcc gggtccaggg actccctcag 28080

catgatgccc acggccctca gcatcagtcg tctggtgcgg cgggcgcagc agcgcatgcg 28140

gatctcgctc aggtcgctgc agtacgtgca acacagaacc accaggttgt tcaacagtcc 28200

atagttcaac acgctccagc cgaaactcat cgcgggaagg atgctaccca cgtggccgtc 28260

[0178] gtaccagatc ctcaggtaaa tcaagtggtg ccccctccag aacacgctgc ccacgtacat 28320

gatctccttg ggcatgtggc ggttcaccac ctcccggtac cacatcaccc tctggttgaa 28380

catgcagccc cggatgatcc tgcggaacca cagggccagc accgccccgc ccgccatgca 28440

gcgaagagac cccgggtccc ggcaatggca atggaggacc caccgctcgt acccgtggat 28500

catctgggag ctgaacaagt ctatgttggc acagcacagg catatgctca tgcatctctt 28560

cagcactctc aactcctcgg gggtcaaaac catatcccag ggcacgggga actcttgcag 28620

gacagcgaac cccgcagaac agggcaatcc tcgcacagaa cttacattgt gcatggacag 28680

ggtatcgcaa tcaggcagca ccgggtgatc ctccaccaga gaagcgcggg tctcggtctc 28740

ctcacagcgt ggtaaggggg ccggccgata cgggtgatgg cgggacgcgg ctgatcgtgt 28800

tcgcgaccgt gtcatgatgc agttgctttc ggacattttc gtacttgctg tagcagaacc 28860

tggtccgggc gctgcacacc gatcgccggc ggcggtctcg gcgcttggaa cgctcggtgt 28920

tgaaattgta aaacagccac tctctcagac cgtgcagcag atctagggcc tcaggagtga 28980

tgaagatccc atcatgcctg atggctctga tcacatcgac caccgtggaa tgggccagac 29040

ccagccagat gatgcaattt tgttgggttt cggtgacggc gggggaggga agaacaggaa 29100

gaaccatgat taacttttaa tccaaacggt ctcggagtac ttcaaaatga agatcgcgga 29160

gatggcacct ctcgcccccg ctgtgttggt ggaaaataac agccaggtca aaggtgatac 29220

ggttctcgag atgttccacg gtggcttcca gcaaagcctc cacgcgcaca tccagaaaca 29280

agacaatagc gaaagcggga gggttctcta attcctcaat catcatgtta cactcctgca 29340

ccatccccag ataattttca tttttccagc cttgaatgat tcgaactagt tcctgaggta 29400

aatccaagcc agccatgata aagagctcgc gcagagcgcc ctccaccggc attcttaagc 29460

acaccctcat aattccaaga tattctgctc ctggttcacc tgcagcagat tgacaagcgg 29520

aatatcaaaa tctctgccgc gatccctgag ctcctccctc agcaataact gtaagtactc 29580

tttcatatcc tctccgaaat ttttagccat aggaccacca ggaataagat tagggcaagc 29640

cacagtacag ataaaccgaa gtcctcccca gtgagcattg ccaaatgcaa gactgctata 29700

agcatgctgg ctagacccgg tgatatcttc cagataactg gacagaaaat cgcccaggca 29760

[0179] atttttaaga aaatcaacaa aagaaaaatc ctccaggtgg acgtttagag cctcgggaac 29820

aacgatgaag taaatgcaag cggtgcgttc cagcatggtt agttagctga tctgtagaaa 29880

aaacaaaaat gaacattaaa ccatgctagc ctggcgaaca ggtgggtaaa tcgttctctc 29940

cagcaccagg caggccacgg ggtctccggc gcgaccctcg taaaaattgt cgctatgatt 30000

gaaaaccatc acagagagac gttcccggtg gccggcgtga atgattcgac aagatgaata 30060

caccccgga acattggcgt ccgcgagtga aaaaaagcgc ccgaggaagc aataaggcac 30120

tacaatgctc agtctcaagt ccagcaaagc gatgccatgc ggatgaagca caaaattctc 30180

aggtgcgtac aaaatgtaat tactcccctc ctgcacaggc agcaaagccc ccgatccctc 30240

caggtacaca tacaaagcct cagcgtccat agcttaccga gcagcagcac acaacaggcg 30300

caagagtcag agaaaggctg agctctaacc tgtccacccg ctctctgctc aatatatagc 30360

ccagatctac actgacgtaa aggccaaagt ctaaaaatac ccgccaaata atcacacacg 30420

cccagcacac gcccagaaac cggtgacaca ctcaaaaaaa tacgcgcact tcctcaaacg 30480

cccaaaactg ccgtcatttc cgggttccca cgctacgtca tcaaaacacg actttcaaat 30540

tccgtcgacc gttaaaaacg tcacccgccc cgcccctaac ggtcgcccgt ctctcagcca 30600

atcagcgccc cgcatcccca aattcaaaca cctcatttgc atattaacgc gcacaaaaag 30660

tttgaggtat attattgatg atgg 30684

<210> 14
<211> 8602
<212> DNA
<213> 人工序列(Artificial Sequence)

<220>
<223> 人工序列的描述：合成多核苷酸

<400> 14
atgggcggcg catgagagaa gcccagacca attacctacc caaaatggag aaagttcacg 60

ttgacatcga ggaagacagc ccattcctca gagctttgca gcggagcttc ccgcagtttg 120

aggtagaagc caagcaggtc actgataatg accatgctaa tgccagagcg ttttcgcatc 180

tggcttcaaa actgatcgaa acggaggtgg acccatccga cacgatcctt gacattggaa 240

[0180] gtgcgcccgc ccgcagaatg tattctaagc acaagtatca ttgtatctgt ccgatgagat 300

gtgcggaaga tccggacaga ttgtataagt atgcaactaa gctgaagaaa aactgtaagg 360

aaataactga taaggaattg gacaagaaaa tgaaggagct cgccgccgtc atgagcgacc 420

ctgacctgga aactgagact atgtgcctcc acgacgacga gtcgtgtcgc tacgaagggc 480

aagtcgctgt ttaccaggat gtatacgcgg ttgacggacc gacaagtctc tatcaccaag 540

ccaataaggg agttagagtc gcctactgga taggctttga caccaccct tttatgttta 600

agaacttggc tggagcatat ccatcatact ctaccaactg ggccgacgaa accgtgttaa 660

cggctcgtaa cataggccta tgcagctctg acgttatgga gcggtcacgt agagggatgt 720

ccattcttag aaagaagtat ttgaaaccat ccaacaatgt tctattctct gttggctcga 780

ccatctacca cgagaagagg gacttactga ggagctggca cctgccgtct gtatttcact 840

tacgtggcaa gcaaaattac acatgtcggt gtgagactat agttagttgc gacgggtacg 900

tcgttaaaag aatagctatc agtccaggcc tgtatgggaa gccttcaggc tatgctgcta 960

cgatgcaccg cgagggattc ttgtgctgca aagtgacaga cacattgaac ggggagaggg 1020

tctcttttcc cgtgtgcacg tatgtgccag ctacattgtg tgaccaaatg actggcatac 1080

tggcaacaga tgtcagtgcg gacgacgcgc aaaaactgct ggttgggctc aaccagcgta 1140

tagtcgtcaa cggtcgcacc cagagaaaca ccaataccat gaaaaattac cttttgcccg 1200

tagtggccca ggcatttgct aggtgggcaa aggaatataa ggaagatcaa gaagatgaaa 1260

ggccactagg actacgagat agacagttag tcatggggtg ttgttgggct tttagaaggc 1320

acaagataac atctatttat aagcgcccgg atacccaaac catcatcaaa gtgaacagcg 1380

atttccactc attcgtgctg cccaggatag gcagtaacac attggagatc gggctgagaa 1440

caagaatcag gaaaatgtta gaggagcaca aggagccgtc acctctcatt accgccgagg 1500

acgtacaaga agctaagtgc gcagccgatg aggctaagga ggtgcgtgaa gccgaggagt 1560

tgcgcgcagc tctaccacct ttggcagctg atgttgagga gcccactctg gaagccgatg 1620

tcgacttgat gttacaagag gctggggccg gctcagtgga gacacctcgt ggcttgataa 1680

aggttaccag ctacgctggc gaggacaaga tcggctctta cgctgtgctt tctccgcagg 1740

[0181] ctgtactcaa gagtgaaaaa ttatcttgca tccaccctct cgctgaacaa gtcatagtga 1800

taacacactc tggccgaaaa gggcgttatg ccgtggaacc ataccatggt aaagtagtgg 1860

tgccagaggg acatgcaata cccgtccagg actttcaagc tctgagtgaa agtgccacca 1920

ttgtgtacaa cgaacgtgag ttcgtaaaca ggtacctgca ccatattgcc acacatggag 1980

gagcgctgaa cactgatgaa gaatattaca aaactgtcaa gcccagcgag cacgacggcg 2040

aatacctgta cgacatcgac aggaaacagt gcgtcaagaa agaactagtc actgggctag 2100

ggctcacagg cgagctggtg gatcctccct tccatgaatt cgcctacgag agtctgagaa 2160

cacgaccagc cgctccttac caagtaccaa ccataggggt gtatggcgtg ccaggatcag 2220

gcaagtctgg catcattaaa agcgcagtca ccaaaaaaga tctagtggtg agcgccaaga 2280

aagaaaactg tgcagaaatt ataagggacg tcaagaaaat gaaagggctg gacgtcaatg 2340

ccagaactgt ggactcagtg ctcttgaatg gatgcaaaca ccccgtagag accctgtata 2400

ttgacgaagc ttttgcttgt catgcaggta ctctcagagc gctcatagcc attataagac 2460

ctaaaaaggc agtgctctgc ggggatccca aacagtgcgg ttttttaac atgatgtgcc 2520

tgaaagtgca ttttaaccac gagatttgca cacaagtctt ccacaaaagc atctctcgcc 2580

gttgcactaa atctgtgact tcggtcgtct caaccttgtt ttacgacaaa aaaatgagaa 2640

cgacgaatcc gaaagagact aagattgtga ttgacactac cggcagtacc aaacctaagc 2700

aggacgatct cattctcact tgtttcagag ggtgggtgaa gcagttgcaa atagattaca 2760

aaggcaacga aataatgacg gcagctgcct ctcaagggct gacccgtaaa ggtgtgtatg 2820

ccgttcggta caaggtgaat gaaaatcctc tgtacgcacc cacctcagaa catgtgaacg 2880

tcctactgac ccgcacggag gaccgcatcg tgtggaaaac actagccggc gacccatgga 2940

taaaaacact gactgccaag taccctggga atttcactgc cacgatagag gagtggcaag 3000

cagagcatga tgccatcatg aggcacatct tggagagacc ggaccctacc gacgtcttcc 3060

agaataaggc aaacgtgtgt tgggccaagg ctttagtgcc ggtgctgaag accgctggca 3120

tagacatgac cactgaacaa tggaacactg tggattattt tgaaacggac aaagctcact 3180

cagcagagat agtattgaac caactatgcg tgaggttctt tggactcgat ctggactccg 3240

[0182] gtctattttc tgcacccact gttccgttat ccattaggaa taatcactgg gataactccc 3300

cgtcgcctaa catgtacggg ctgaataaag aagtggtccg tcagctctct cgcaggtacc 3360

cacaactgcc tcgggcagtt gccactggaa gagtctatga catgaacact ggtacactgc 3420

gcaattatga tccgcgcata aacctagtac ctgtaaacag aagactgcct catgctttag 3480

tcctccacca taatgaacac ccacagagtg acttttcttc attcgtcagc aaattgaagg 3540

gcagaactgt cctggtggtc ggggaaaagt tgtccgtccc aggcaaaatg gttgactggt 3600

tgtcagaccg gcctgaggct accttcagag ctcggctgga tttaggcatc ccaggtgatg 3660

tgcccaaata tgacataata tttgttaatg tgaggacccc atataaatac catcactatc 3720

agcagtgtga agaccatgcc attaagctta gcatgttgac caagaaagct tgtctgcatc 3780

tgaatcccgg cggaacctgt gtcagcatag gttatggtta cgctgacagg gccagcgaaa 3840

gcatcattgg tgctatagcg cggcagttca agttttcccg ggtatgcaaa ccgaaatcct 3900

cacttgaaga gacggaagtt ctgtttgtat tcattgggta cgatcgcaag gcccgtacgc 3960

acaatcctta caagctttca tcaaccttga ccaacattta tacaggttcc agactccacg 4020

aagccggatg tgcaccctca tatcatgtgg tgcgagggga tattgccacg gccaccgaag 4080

gagtgattat aaatgctgct aacagcaaag gacaacctgg cggaggggtg tgcggagcgc 4140

tgtataagaa attcccggaa agcttcgatt tacagccgat cgaagtagga aaagcgcgac 4200

tggtcaaagg tgcagctaaa catatcattc atgccgtagg accaaacttc aacaaagttt 4260

cggaggttga aggtgacaaa cagttggcag aggcttatga gtccatcgct aagattgtca 4320

acgataacaa ttacaagtca gtagcgattc cactgttgtc caccggcatc ttttccggga 4380

acaaagatcg actaacccaa tcattgaacc atttgctgac agctttagac accactgatg 4440

cagatgtagc catatactgc agggacaaga aatgggaaat gactctcaag gaagcagtgg 4500

ctaggagaga agcagtggag gagatatgca tatccgacga ctcttcagtg acagaacctg 4560

atgcagagct ggtgagggtg catccgaaga gttctttggc tggaaggaag ggctacagca 4620

caagcgatgg caaaactttc tcatatttgg aagggaccaa gtttcaccag gcggccaagg 4680

atatagcaga aattaatgcc atgtggcccg ttgcaacgga ggccaatgag caggtatgca 4740

[0183] tgtatatcct cggagaaagc atgagcagta ttaggtcgaa atgccccgtc gaagagtcgg 4800

aagcctccac accacctagc acgctgcctt gcttgtgcat ccatgccatg actccagaaa 4860

gagtacagcg cctaaaagcc tcacgtccag aacaaattac tgtgtgctca tcctttccat 4920

tgccgaagta tagaatcact ggtgtgcaga agatccaatg ctcccagcct atattgttct 4980

caccgaaagt gcctgcgtat attcatccaa ggaagtatct cgtggaaaca ccaccggtag 5040

acgagactcc ggagccatcg gcagagaacc aatccacaga ggggacacct gaacaaccac 5100

cacttataac cgaggatgag accaggacta gaacgcctga gccgatcatc atcgaagagg 5160

aagaagagga tagcataagt ttgctgtcag atggcccgac ccaccaggtg ctgcaagtcg 5220

aggcagacat tcacgggccg ccctctgtat ctagctcatc ctggtccatt cctcatgcat 5280

ccgactttga tgtggacagt ttatccatac ttgacaccct ggagggagct agcgtgacca 5340

gcggggcaac gtcagccgag actaactctt acttcgcaaa gagtatggag tttctggcgc 5400

gaccggtgcc tgcgcctcga acagtattca ggaaccctcc acatcccgct ccgcgcacaa 5460

gaacaccgtc acttgcaccc agcagggcct gctcgagaac cagcctagtt tccaccccgc 5520

caggcgtgaa tagggtgatc actagagagg agctcgaggc gcttaccccg tcacgcactc 5580

ctagcaggtc ggtctcgaga accagcctgg tctccaaccc gccaggcgta aatagggtga 5640

ttacaagaga ggagtttgag gcgttcgtag cacaacaaca atgacggttt gatgcgggtg 5700

catacatctt ttcctccgac accggtcaag ggcatttaca acaaaaatca gtaaggcaaa 5760

cggtgctatc cgaagtggtg ttggagagga ccgaattgga gatttcgtat gccccgcgcc 5820

tcgaccaaga aaaagaagaa ttactacgca agaaattaca gttaaatccc acacctgcta 5880

acagaagcag ataccagtcc aggaaggtgg agaacatgaa agccataaca gctagacgta 5940

ttctgcaagg cctagggcat tatttgaagg cagaaggaaa agtggagtgc taccgaaccc 6000

tgcatcctgt tcctttgtat tcatctagtg tgaaccgtgc cttttcaagc cccaaggtcg 6060

cagtggaagc ctgtaacgcc atgttgaaag agaactttcc gactgtggct tcttactgta 6120

ttattccaga gtacgatgcc tatttggaca tggttgacgg agcttcatgc tgcttagaca 6180

ctgccagttt ttgccctgca aagctgcgca gctttccaaa gaaacactcc tatttggaac 6240

[0184] ccacaatacg atcggcagtg ccttcagcga tccagaacac gctccagaac gtcctggcag 6300

ctgccacaaa aagaaattgc aatgtcacgc aaatgagaga attgcccgta ttggattcgg 6360

cggcctttaa tgtggaatgc ttcaagaaat atgcgtgtaa taatgaatat tgggaaacgt 6420

ttaaagaaaa ccccatcagg cttactgaag aaaacgtggt aaattacatt accaaattaa 6480

aaggaccaaa agctgctgct ctttttgcga agacacataa tttgaatatg ttgcaggaca 6540

taccaatgga caggtttgta atggacttaa agagagacgt gaaagtgact ccaggaacaa 6600

aacatactga agaacggccc aaggtacagg tgatccaggc tgccgatccg ctagcaacag 6660

cgtatctgtg cggaatccac cgagagctgg ttaggagatt aaatgcggtc ctgcttccga 6720

acattcatac actgtttgat atgtcggctg aagactttga cgctattata gccgagcact 6780

tccagcctgg ggattgtgtt ctggaaactg acatcgcgtc gtttgataaa agtgaggacg 6840

acgccatggc tctgaccgcg ttaatgattc tggaagactt aggtgtggac gcagagctgt 6900

tgacgctgat tgaggcggct ttcggcgaaa tttcatcaat acatttgccc actaaaacta 6960

aatttaaatt cggagccatg atgaaatctg gaatgttcct cacactgttt gtgaacacag 7020

tcattaacat tgtaatcgca agcagagtgt tgagagaacg gctaaccgga tcaccatgtg 7080

cagcattcat tggagatgac aatatcgtga aaggagtcaa atcggacaaa ttaatggcag 7140

acaggtgcgc cacctggttg aatatggaag tcaagattat agatgctgtg gtgggcgaga 7200

aagcgcctta tttctgtgga gggtttattt tgtgtgactc cgtgaccggc acagcgtgcc 7260

gtgtggcaga ccccctaaaa aggctgttta agcttggcaa acctctggca gcagacgatg 7320

aacatgatga tgacaggaga agggcattgc atgaagagtc aacacgctgg aaccgagtgg 7380

gtattctttc agagctgtgc aaggcagtag aatcaaggta tgaaaccgta ggaacttcca 7440

tcatagttat ggccatgact actctagcta gcagtgttaa atcattcagc tacctgagag 7500

gggcccctat aactctctac ggctaacctg aatggactac gactctagaa tagtctttaa 7560

ttaaagtccg ccatatgagg ccaccatgca gatcttcgtg aagaccctga ccggcaagac 7620

catcacccta gaggtggagc ccagtgacac catcgagaac gtgaaggcca agatccagga 7680

taaagagggc atccccccctg accagcagag gctgatcttt gccggcaagc agctggaaga 7740

[0185] tggccgcacc ctctctgatt acaacatcca gaaggagtca accctgcacc tggtccttcg 7800

cctgagaggt ggcgctgctt acagtataat caactttgaa aaactggctg cttacggcat 7860

cctgggcttt gtgtttacac tggctgccta cctgctgttt ggctatcctg tgtacgtggc 7920

cgcttatgga ctgtgtaccc tggtggccat gctggctgct tacaatctgg tgcctatggt 7980

ggccacagtg gccgcctatt gtcttggcgg actgctgaca atggtggcag cctacagccc 8040

gagctatgcg tatcatcagt ttgcagccta cggcccagga ccaggcgcta aatttgtggc 8100

tgcctggaca ctgaaagccg ccgctggacc aggtcctgga cagtacatca aggccaacag 8160

caagttcatc ggcatcaccg aactcggccc aggaccaggc tatccctacg atgtgcctga 8220

ttacgcctga tagtgatgat tcgaacggcc gtatcacgcc caaacattta cagccgcggt 8280

gtcaaaaacc gcgtggacgt ggttaacatc cctgctggga ggatcagccg taattattat 8340

aattggcttg gtgctggcta ctattgtggc catgtacgtg ctgaccaacc agaaacataa 8400

ttgaatacag cagcaattgg caagctgctt acatagaact cgcggcgatt ggcatgccgc 8460

cttaaaattt ttattttatt ttttcttttc ttttccgaat cggattttgt ttttaatatt 8520

tcaaaaaaaa aaaaaaaaaa aaaaaaaaaa aaaaaaaaaa aaaaaaaaaa aaaaaaaaaa 8580

aaaaaaaaaa aaaaaaaaaa aa 8602

<210> 15
<211> 9595
<212> DNA
<213> 人工序列(Artificial Sequence)

<220>
<223> 人工序列的描述：合成多核苷酸

<400> 15
atgggcggcg catgagagaa gcccagacca attacctacc caaaatggag aaagttcacg 60

ttgacatcga ggaagacagc ccattcctca gagctttgca gcggagcttc ccgcagtttg 120

aggtagaagc caagcaggtc actgataatg accatgctaa tgccagagcg ttttcgcatc 180

tggcttcaaa actgatcgaa acggaggtgg acccatccga cacgatcctt gacattggaa 240

gtgcgcccgc ccgcagaatg tattctaagc acaagtatca ttgtatctgt ccgatgagat 300

[0186] gtgcggaaga tccggacaga ttgtataagt atgcaactaa gctgaagaaa aactgtaagg 360

aaataactga taaggaattg gacaagaaaa tgaaggagct cgccgccgtc atgagcgacc 420

ctgacctgga aactgagact atgtgcctcc acgacgacga gtcgtgtcgc tacgaagggc 480

aagtcgctgt ttaccaggat gtatacgcgg ttgacggacc gacaagtctc tatcaccaag 540

ccaataaggg agttagagtc gcctactgga taggctttga caccacccct tttatgttta 600

agaacttggc tggagcatat ccatcatact ctaccaactg ggccgacgaa accgtgttaa 660

cggctcgtaa cataggccta tgcagctctg acgttatgga gcggtcacgt agagggatgt 720

ccattcttag aaagaagtat ttgaaaccat ccaacaatgt tctattctct gttggctcga 780

ccatctacca cgagaagagg gacttactga ggagctggca cctgccgtct gtatttcact 840

tacgtggcaa gcaaaattac acatgtcggt gtgagactat agttagttgc gacgggtacg 900

tcgttaaaag aatagctatc agtccaggcc tgtatgggaa gccttcaggc tatgctgcta 960

cgatgcaccg cgagggattc ttgtgctgca aagtgacaga cacattgaac ggggagaggg 1020

tctcttttcc cgtgtgcacg tatgtgccag ctacattgtg tgaccaaatg actggcatac 1080

tggcaacaga tgtcagtgcg gacgacgcgc aaaaactgct ggttgggctc aaccagcgta 1140

tagtcgtcaa cggtcgcacc cagagaaaca ccaataccat gaaaaattac cttttgcccg 1200

tagtggccca ggcatttgct aggtgggcaa aggaatataa ggaagatcaa gaagatgaaa 1260

ggccactagg actacgagat agacagttag tcatggggtg ttgttgggct tttagaaggc 1320

acaagataac atctatttat aagcgcccgg atacccaaac catcatcaaa gtgaacagcg 1380

atttccactc attcgtgctg cccaggatag gcagtaacac attggagatc gggctgagaa 1440

caagaatcag gaaaatgtta gaggagcaca aggagccgtc acctctcatt accgccgagg 1500

acgtacaaga agctaagtgc gcagccgatg aggctaagga ggtgcgtgaa gccgaggagt 1560

tgcgcgcagc tctaccacct ttggcagctg atgttgagga gcccactctg gaagccgatg 1620

tcgacttgat gttacaagag gctggggccg gctcagtgga gacacctcgt ggcttgataa 1680

aggttaccag ctacgctggc gaggacaaga tcggctctta cgctgtgctt tctccgcagg 1740

ctgtactcaa gagtgaaaaa ttatcttgca tccaccctct cgctgaacaa gtcatagtga 1800

[0187] taacacactc tggccgaaaa gggcgttatg ccgtggaacc ataccatggt aaagtagtgg 1860

tgccagaggg acatgcaata cccgtccagg actttcaagc tctgagtgaa agtgccacca 1920

ttgtgtacaa cgaacgtgag ttcgtaaaca ggtacctgca ccatattgcc acacatggag 1980

gagcgctgaa cactgatgaa gaatattaca aaactgtcaa gcccagcgag cacgacggcg 2040

aatacctgta cgacatcgac aggaaacagt gcgtcaagaa agaactagtc actgggctag 2100

ggctcacagg cgagctggtg gatcctccct tccatgaatt cgcctacgag agtctgagaa 2160

cacgaccagc cgctccttac caagtaccaa ccataggggt gtatggcgtg ccaggatcag 2220

gcaagtctgg catcattaaa agcgcagtca ccaaaaaaga tctagtggtg agcgccaaga 2280

aagaaaactg tgcagaaatt ataagggacg tcaagaaaat gaaagggctg gacgtcaatg 2340

ccagaactgt ggactcagtg ctcttgaatg gatgcaaaca ccccgtagag accctgtata 2400

ttgacgaagc ttttgcttgt catgcaggta ctctcagagc gctcatagcc attataagac 2460

ctaaaaaggc agtgctctgc ggggatccca aacagtgcgg ttttttttaac atgatgtgcc 2520

tgaaagtgca ttttaaccac gagatttgca cacaagtctt ccacaaaagc atctctcgcc 2580

gttgcactaa atctgtgact tcggtcgtct caaccttgtt ttacgacaaa aaaatgagaa 2640

cgacgaatcc gaaagagact aagattgtga ttgacactac cggcagtacc aaacctaagc 2700

aggacgatct cattctcact tgtttcagag ggtgggtgaa gcagttgcaa atagattaca 2760

aaggcaacga aataatgacg gcagctgcct ctcaagggct gacccgtaaa ggtgtgtatg 2820

ccgttcggta caaggtgaat gaaaatcctc tgtacgcacc cacctcagaa catgtgaacg 2880

tcctactgac ccgcacggag gaccgcatcg tgtggaaaac actagccggc gacccatgga 2940

taaaaacact gactgccaag taccctggga atttcactgc cacgatagag gagtggcaag 3000

cagagcatga tgccatcatg aggcacatct tggagagacc ggaccctacc gacgtcttcc 3060

agaataaggc aaacgtgtgt tgggccaagg ctttagtgcc ggtgctgaag accgctggca 3120

tagacatgac cactgaacaa tggaacactg tggattattt tgaaacggac aaagctcact 3180

cagcagagat agtattgaac caactatgcg tgaggttctt tggactcgat ctggactccg 3240

gtctattttc tgcacccact gttccgttat ccattaggaa taatcactgg gataactccc 3300

[0188] cgtcgcctaa catgtacggg ctgaataaag aagtggtccg tcagctctct cgcaggtacc 3360

cacaactgcc tcgggcagtt gccactggaa gagtctatga catgaacact ggtacactgc 3420

gcaattatga tccgcgcata aacctagtac ctgtaaacag aagactgcct catgctttag 3480

tcctccacca taatgaacac ccacagagtg acttttcttc attcgtcagc aaattgaagg 3540

gcagaactgt cctggtggtc ggggaaaagt tgtccgtccc aggcaaaatg gttgactggt 3600

tgtcagaccg gcctgaggct accttcagag ctcggctgga tttaggcatc ccaggtgatg 3660

tgcccaaata tgacataata tttgttaatg tgaggacccc atataaatac catcactatc 3720

agcagtgtga agaccatgcc attaagctta gcatgttgac caagaaagct tgtctgcatc 3780

tgaatcccgg cggaacctgt gtcagcatag gttatggtta cgctgacagg gccagcgaaa 3840

gcatcattgg tgctatagcg cggcagttca agttttcccg ggtatgcaaa ccgaaatcct 3900

cacttgaaga gacggaagtt ctgtttgtat tcattgggta cgatcgcaag gcccgtacgc 3960

acaatcctta caagctttca tcaaccttga ccaacattta tacaggttcc agactccacg 4020

aagccggatg tgcaccctca tatcatgtgg tgcgagggga tattgccacg gccaccgaag 4080

gagtgattat aaatgctgct aacagcaaag gacaacctgg cggaggggtg tgcggagcgc 4140

tgtataagaa attcccggaa agcttcgatt tacagccgat cgaagtagga aaagcgcgac 4200

tggtcaaagg tgcagctaaa catatcattc atgccgtagg accaaacttc aacaaagttt 4260

cggaggttga aggtgacaaa cagttggcag aggcttatga gtccatcgct aagattgtca 4320

acgataacaa ttacaagtca gtagcgattc cactgttgtc caccggcatc ttttccggga 4380

acaaagatcg actaacccaa tcattgaacc atttgctgac agctttagac accactgatg 4440

cagatgtagc catatactgc agggacaaga aatgggaaat gactctcaag gaagcagtgg 4500

ctaggagaga agcagtggag gagatatgca tatccgacga ctcttcagtg acagaacctg 4560

atgcagagct ggtgagggtg catccgaaga gttctttggc tggaaggaag ggctacagca 4620

caagcgatgg caaaactttc tcatatttgg aagggaccaa gtttcaccag gcggccaagg 4680

atatagcaga aattaatgcc atgtggcccg ttgcaacgga ggccaatgag caggtatgca 4740

tgtatatcct cggagaaagc atgagcagta ttaggtcgaa atgccccgtc gaagagtcgg 4800

[0189] aagcctccac accacctagc acgctgcctt gcttgtgcat ccatgccatg actccagaaa 4860

gagtacagcg cctaaaagcc tcacgtccag aacaaattac tgtgtgctca tcctttccat 4920

tgccgaagta tagaatcact ggtgtgcaga agatccaatg ctcccagcct atattgttct 4980

caccgaaagt gcctgcgtat attcatccaa ggaagtatct cgtggaaaca ccaccggtag 5040

acgagactcc ggagccatcg gcagagaacc aatccacaga ggggacacct gaacaaccac 5100

cacttataac cgaggatgag accaggacta gaacgcctga gccgatcatc atcgaagagg 5160

aagaagagga tagcataagt ttgctgtcag atggcccgac ccaccaggtg ctgcaagtcg 5220

aggcagacat tcacgggccg ccctctgtat ctagctcatc ctggtccatt cctcatgcat 5280

ccgactttga tgtggacagt ttatccatac ttgacaccct ggagggagct agcgtgacca 5340

gcggggcaac gtcagccgag actaactctt acttcgcaaa gagtatggag tttctggcgc 5400

gaccggtgcc tgcgcctcga acagtattca ggaaccctcc acatcccgct ccgcgcacaa 5460

gaacaccgtc acttgcaccc agcagggcct gctcgagaac cagcctagtt tccaccccgc 5520

caggcgtgaa tagggtgatc actagagagg agctcgaggc gcttaccccg tcacgcactc 5580

ctagcaggtc ggtctcgaga accagcctgg tctccaaccc gccaggcgta aatagggtga 5640

ttacaagaga ggagtttgag gcgttcgtag cacaacaaca atgacggttt gatgcgggtg 5700

catacatctt ttcctccgac accggtcaag ggcatttaca acaaaaatca gtaaggcaaa 5760

cggtgctatc cgaagtggtg ttggagagga ccgaattgga gatttcgtat gccccgcgcc 5820

tcgaccaaga aaaagaagaa ttactacgca agaaattaca gttaaatccc acacctgcta 5880

acagaagcag ataccagtcc aggaaggtgg agaacatgaa agccataaca gctagacgta 5940

ttctgcaagg cctagggcat tatttgaagg cagaaggaaa agtggagtgc taccgaaccc 6000

tgcatcctgt tcctttgtat tcatctagtg tgaaccgtgc cttttcaagc cccaaggtcg 6060

cagtggaagc ctgtaacgcc atgttgaaag agaactttcc gactgtggct tcttactgta 6120

ttattccaga gtacgatgcc tatttggaca tggttgacgg agcttcatgc tgcttagaca 6180

ctgccagttt ttgccctgca aagctgcgca gctttccaaa gaaacactcc tatttggaac 6240

ccacaatacg atcggcagtg ccttcagcga tccagaacac gctccagaac gtcctggcag 6300

[0190] ctgccacaaa aagaaattgc aatgtcacgc aaatgagaga attgcccgta ttggattcgg 6360

cggcctttaa tgtggaatgc ttcaagaaat atgcgtgtaa taatgaatat tgggaaacgt 6420

ttaaagaaaa ccccatcagg cttactgaag aaaacgtggt aaattacatt accaaattaa 6480

aaggaccaaa agctgctgct ctttttgcga agacacataa tttgaatatg ttgcaggaca 6540

taccaatgga caggtttgta atggacttaa agagagacgt gaaagtgact ccaggaacaa 6600

aacatactga agaacggccc aaggtacagg tgatccaggc tgccgatccg ctagcaacag 6660

cgtatctgtg cggaatccac cgagagctgg ttaggagatt aaatgcggtc ctgcttccga 6720

acattcatac actgtttgat atgtcggctg aagactttga cgctattata gccgagcact 6780

tccagcctgg ggattgtgtt ctggaaactg acatcgcgtc gtttgataaa agtgaggacg 6840

acgccatggc tctgaccgcg ttaatgattc tggaagactt aggtgtggac gcagagctgt 6900

tgacgctgat tgaggcggct ttcggcgaaa tttcatcaat acatttgccc actaaaacta 6960

aatttaaatt cggagccatg atgaaatctg gaatgttcct cacactgttt gtgaacacag 7020

tcattaacat tgtaatcgca agcagagtgt tgagagaacg gctaaccgga tcaccatgtg 7080

cagcattcat tggagatgac aatatcgtga aaggagtcaa atcggacaaa ttaatggcag 7140

acaggtgcgc cacctggttg aatatggaag tcaagattat agatgctgtg gtgggcgaga 7200

aagcgcctta tttctgtgga gggtttattt tgtgtgactc cgtgaccggc acagcgtgcc 7260

gtgtggcaga ccccctaaaa aggctgttta agcttggcaa acctctggca gcagacgatg 7320

aacatgatga tgacaggaga agggcattgc atgaagagtc aacacgctgg aaccgagtgg 7380

gtattctttc agagctgtgc aaggcagtag aatcaaggta tgaaaccgta ggaacttcca 7440

tcatagttat ggccatgact actctagcta gcagtgttaa atcattcagc tacctgagag 7500

gggcccctat aactctctac ggctaacctg aatggactac gactctagaa tagtctttaa 7560

ttaaagtccg ccatatgaga tggaagatgc caaaaacatt aagaagggcc cagcgccatt 7620

ctacccactc gaagacggga ccgccggcga gcagctgcac aaagccatga agcgctacgc 7680

cctggtgccc ggcaccatcg cctttaccga cgcacatatc gaggtggaca ttacctacgc 7740

cgagtacttc gagatgagcg ttcggctggc agaagctatg aagcgctatg ggctgaatac 7800

[0191] aaaccatcgg atcgtggtgt gcagcgagaa tagcttgcag ttcttcatgc ccgtgttggg 7860

tgccctgttc atcggtgtgg ctgtggcccc agctaacgac atctacaacg agcgcgagct 7920

gctgaacagc atgggcatca gccagcccac cgtcgtattc gtgagcaaga aagggctgca 7980

aaagatcctc aacgtgcaaa agaagctacc gatcatacaa aagatcatca tcatggatag 8040

caagaccgac taccagggct tccaaagcat gtacaccttc gtgacttccc atttgccacc 8100

cggcttcaac gagtacgact tcgtgcccga gagcttcgac cgggacaaaa ccatcgccct 8160

gatcatgaac agtagtggca gtaccggatt gcccaagggc gtagccctac cgcaccgcac 8220

cgcttgtgtc cgattcagtc atgcccgcga ccccatcttc ggcaaccaga tcatccccga 8280

caccgctatc ctcagcgtgg tgccatttca ccacggcttc ggcatgttca ccacgctggg 8340

ctacttgatc tgcggctttc gggtcgtgct catgtaccgc ttcgaggagg agctattctt 8400

gcgcagcttg caagactata agattcaatc tgccctgctg gtgcccacac tatttagctt 8460

cttcgctaag agcactctca tcgacaagta cgacctaagc aacttgcacg agatcgccag 8520

cggcggggcg ccgctcagca aggaggtagg tgaggccgtg gccaaacgct tccacctacc 8580

aggcatccgc cagggctacg gcctgacaga aacaaccagc gccattctga tcacccccga 8640

aggggacgac aagcctggcg cagtaggcaa ggtggtgccc ttcttcgagg ctaaggtggt 8700

ggacttggac accggtaaga cactgggtgt gaaccagcgc ggcgagctgt gcgtccgtgg 8760

ccccatgatc atgagcggct acgttaacaa ccccgaggct acaaacgctc tcatcgacaa 8820

ggacggctgg ctgcacagcg gcgacatcgc ctactgggac gaggacgagc acttcttcat 8880

cgtggaccgg ctgaagagcc tgatcaaata caagggctac caggtagccc cagccgaact 8940

ggagagcatc ctgctgcaac accccaacat cttcgacgcc ggggtcgccg gcctgcccga 9000

cgacgatgcc ggcgagctgc ccgccgcagt cgtcgtgctg gaacacggta aaaccatgac 9060

cgagaaggag atcgtggact atgtggccag ccaggttaca accgccaaga agctgcgcgg 9120

tggtgttgtg ttcgtggacg aggtgcctaa aggactgacc ggcaagttgg acgcccgcaa 9180

gatccgcgag attctcatta aggccaagaa gggcggcaag atcgccgtgt aattcgaacg 9240

gccgtatcac gcccaaacat ttacagccgc ggtgtcaaaa accgcgtgga cgtggttaac 9300

[0192] atccctgctg ggaggatcag ccgtaattat tataattggc ttggtgctgg ctactattgt 9360

ggccatgtac gtgctgacca accagaaaca taattgaata cagcagcaat tggcaagctg 9420

cttacataga actcgcggcg attggcatgc cgccttaaaa tttttatttt atttttttctt 9480

ttcttttccg aatcggattt tgtttttaat atttcaaaaa aaaaaaaaaa aaaaaaaaaa 9540

aaaaaaaaaa aaaaaaaaaa aaaaaaaaaa aaaaaaaaaa aaaaaaaaaa aaaaa 9595

<210> 16
<211> 139
<212> PRT
<213> 人工序列(Artificial Sequence)

<220>
<223> 人工序列的描述：合成多肽

<400> 16
Pro Ser Ser Leu Ser Ala Ser Val Gly Asp Arg Val Thr Ile Thr Cys
1 5 10 15

Arg Ala Ser Gln Ser Ile Asn Ser Tyr Leu Asp Trp Tyr Gln Gln Lys

20 25 30

Pro Gly Lys Ala Pro Lys Leu Leu Ile Tyr Ala Ala Ser Ser Leu Gln
35 40 45

Ser Gly Val Pro Ser Arg Phe Ser Gly Ser Gly Ser Gly Thr Asp Phe
50 55 60

Thr Leu Thr Ile Ser Ser Leu Gln Pro Glu Asp Phe Ala Thr Tyr Tyr
65 70 75 80

Cys Gln Gln Tyr Tyr Ser Thr Pro Phe Thr Phe Gly Pro Gly Thr Lys
85 90 95

Val Glu Ile Lys Arg Thr Val Ala Ala Pro Ser Val Phe Ile Phe Pro
100 105 110

Pro Ser Asp Glu Gln Leu Lys Ser Gly Thr Ala Ser Val Val Cys Leu
115 120 125

[0193]

Leu Asn Asn Phe Tyr Pro Arg Glu Ala Lys Val
130 135

<210> 17
<211> 167
<212> PRT
<213> 人工序列(Artificial Sequence)

<220>
<223> 人工序列的描述：合成多肽

<400> 17
Gly Val Val Gln Pro Gly Arg Ser Leu Arg Leu Ser Cys Ala Ala Ser
1 5 10 15

Gly Phe Thr Phe Ser Ser Tyr Gly Met His Trp Val Arg Gln Ala Pro
20 25 30

Gly Lys Gly Leu Glu Trp Val Ala Val Ile Trp Tyr Asp Gly Ser Asn
35 40 45

Lys Tyr Tyr Ala Asp Ser Val Lys Gly Arg Phe Thr Ile Ser Arg Asp
50 55 60

Asn Ser Lys Asn Thr Leu Tyr Leu Gln Met Asn Ser Leu Arg Ala Glu
65 70 75 80

Asp Thr Ala Val Tyr Tyr Cys Ala Arg Asp Pro Arg Gly Ala Thr Leu
85 90 95

Tyr Tyr Tyr Tyr Tyr Gly Met Asp Val Trp Gly Gln Gly Thr Thr Val
100 105 110

Thr Val Ser Ser Ala Ser Thr Lys Gly Pro Ser Val Phe Pro Leu Ala
115 120 125

Pro Cys Ser Arg Ser Thr Ser Glu Ser Thr Ala Ala Leu Gly Cys Leu
130 135 140

[0194]

Val Lys Asp Tyr Phe Pro Glu Pro Val Thr Val Ser Trp Asn Ser Gly
145 150 155 160

Ala Leu Thr Ser Gly Val His
165

<210> 18
<211> 10
<212> PRT
<213> 人工序列(Artificial Sequence)

<220>
<223> 人工序列的描述：合成肽

<400> 18
Gly Phe Thr Phe Ser Ser Tyr Gly Met His
1 5 10

<210> 19
<211> 15
<212> PRT

<213> 人工序列(Artificial Sequence)

<220>
<223> 人工序列的描述：合成肽

<400> 19
Val Ile Trp Tyr Asp Gly Ser Asn Lys Tyr Tyr Ala Asp Ser Val
1 5 10 15

<210> 20
<211> 16
<212> PRT
<213> 人工序列(Artificial Sequence)

<220>
<223> 人工序列的描述：合成肽

<400> 20
Asp Pro Arg Gly Ala Thr Leu Tyr Tyr Tyr Tyr Tyr Gly Met Asp Val
1 5 10 15

[0195]

<210> 21
<211> 11
<212> PRT
<213> 人工序列(Artificial Sequence)

<220>
<223> 人工序列的描述：合成肽

<400> 21
Arg Ala Ser Gln Ser Ile Asn Ser Tyr Leu Asp
1 5 10

<210> 22
<211> 7
<212> PRT
<213> 人工序列(Artificial Sequence)

<220>
<223> 人工序列的描述：合成肽

<400> 22
Ala Ala Ser Ser Leu Gln Ser
1 5

<210> 23

<211> 9
<212> PRT
<213> 人工序列(Artificial Sequence)

<220>
<223> 人工序列的描述：合成肽

<400> 23
Gln Gln Tyr Tyr Ser Thr Pro Phe Thr
1 5

<210> 24
<211> 108
<212> PRT
<213> 人工序列(Artificial Sequence)

<220>
<223> 人工序列的描述：合成多肽

<400> 24
Glu Ile Val Leu Thr Gln Ser Pro Gly Thr Leu Ser Leu Ser Pro Gly
1 5 10 15

[0196]

Glu Arg Ala Thr Leu Ser Cys Arg Ala Ser Gln Arg Val Ser Ser Ser
20 25 30

Tyr Leu Ala Trp Tyr Gln Gln Lys Pro Gly Gln Ala Pro Arg Leu Leu
35 40 45

Ile Tyr Asp Ala Ser Ser Arg Ala Thr Gly Ile Pro Asp Arg Phe Ser
50 55 60

Gly Ser Gly Ser Gly Thr Asp Phe Thr Leu Thr Ile Ser Arg Leu Glu
65 70 75 80

Pro Glu Asp Phe Ala Val Tyr Tyr Cys Gln Gln Tyr Gly Ser Leu Pro
85 90 95

Trp Thr Phe Gly Gln Gly Thr Lys Val Glu Ile Lys
100 105

<210> 25

<211> 121
<212> PRT
<213> 人工序列(Artificial Sequence)

<220>
<223> 人工序列的描述：合成多肽

<400> 25
Glu Val Gln Leu Val Glu Ser Gly Gly Gly Leu Val Gln Pro Gly Gly
1 5 10 15

Ser Leu Arg Leu Ser Cys Ala Ala Ser Gly Phe Thr Phe Ser Arg Tyr
20 25 30

Trp Met Ser Trp Val Arg Gln Ala Pro Gly Lys Gly Leu Glu Trp Val
35 40 45

Ala Asn Ile Lys Gln Asp Gly Ser Glu Lys Tyr Tyr Val Asp Ser Val
50 55 60

[0197] Lys Gly Arg Phe Thr Ile Ser Arg Asp Asn Ala Lys Asn Ser Leu Tyr
65 70 75 80

Leu Gln Met Asn Ser Leu Arg Ala Glu Asp Thr Ala Val Tyr Tyr Cys
85 90 95

Ala Arg Glu Gly Gly Trp Phe Gly Glu Leu Ala Phe Asp Tyr Trp Gly
100 105 110

Gln Gly Thr Leu Val Thr Val Ser Ser
115 120

<210> 26
<211> 5
<212> PRT
<213> 人工序列(Artificial Sequence)

<220>
<223> 人工序列的描述：合成肽

<400> 26
Arg Tyr Trp Met Ser

1 5

<210> 27
<211> 17
<212> PRT
<213> 人工序列(Artificial Sequence)

<220>
<223> 人工序列的描述：合成肽

<400> 27
Asn Ile Lys Gln Asp Gly Ser Glu Lys Tyr Tyr Val Asp Ser Val Lys
1 5 10 15

Gly

<210> 28
<211> 12
<212> PRT
<213> 人工序列(Artificial Sequence)

[0198]

<220>
<223> 人工序列的描述：合成肽

<400> 28
Glu Gly Gly Trp Phe Gly Glu Leu Ala Phe Asp Tyr
1 5 10

<210> 29
<211> 12
<212> PRT
<213> 人工序列(Artificial Sequence)

<220>
<223> 人工序列的描述：合成肽

<400> 29
Arg Ala Ser Gln Arg Val Ser Ser Ser Tyr Leu Ala
1 5 10

<210> 30
<211> 7
<212> PRT
<213> 人工序列(Artificial Sequence)

<220>
<223> 人工序列的描述：合成肽

<400> 30
Asp Ala Ser Ser Arg Ala Thr
1 5

<210> 31
<211> 9
<212> PRT
<213> 人工序列(Artificial Sequence)

<220>
<223> 人工序列的描述：合成肽

<400> 31
Gln Gln Tyr Gly Ser Leu Pro Trp Thr
1 5

<210> 32
<211> 2019
<212> DNA
[0199] <213> 人工序列(Artificial Sequence)

<220>
<223> 人工序列的描述：合成多核苷酸

<400> 32
gcccgggcat ttaaatgcga tcgcatcgat tacgactcta gaatagtcta gtccgcaggc 60

caccatgcag atcttcgtga agaccctgac cggcaagacc atcaccctag aggtggagcc 120

cagtgacacc atcgagaacg tgaaggccaa gatccaggat aaagagggca tccccccctga 180

ccagcagagg ctgatctttg ccggcaagca gctggaagat ggccgcaccc tctctgatta 240

caacatccag aaggagtcaa ccctgcacct ggtccttcgc ctgagaggtg ccatgtttca 300

ggcgctgagc gaaggctgca ccccgtatga tattaaccag atgctgaacg tgctgggcga 360

tcatcaggtc tcaggccttg agcagcttga gagtataatc aactttgaaa aactgactga 420

atggaccagt tctaatgtta tgcctatcct gtctcctctg acaaagggca tcctgggctt 480

cgtgtttacc ctgaccgtgc cttctgagag aggacttagc tgcattagcg aagcggatgc 540

gaccacccccg gaaagcgcga acctgggcga agaaattctg agccagctgt atctttggcc 600

aagggtgacc taccattccc ctagttatgc ttaccaccaa tttgaaagac gagccaaata 660

taaaagacac ttccccggct ttggccagag cctgctgttt ggctaccctg tgtacgtgtt 720

cggcgattgc gtgcagggcg attgggatgc gattcgcttt cgctattgcg cgccgccggg 780

ctatgcgctg ctgcgctgca acgataccaa ctatagcgct ctgctggctg tgggggccct 840

agaaggaccc aggaatcagg actggcttgg tgtcccaaga caacttgtaa ctcggatgca 900

ggctattcag aatgccggcc tgtgtaccct ggtggccatg ctggaagaga caatcttctg 960

gctgcaagcg tttctgatgg cgctgaccga tagcggcccg aaaaccaaca ttattgtgga 1020

tagccagtat gtgatgggca ttagcaaacc gagctttcag gaatttgtgg attgggaaaa 1080

cgtgagcccg gaactgaaca gcaccgatca gccgttttgg caagccggaa tcctggccag 1140

aaatctggtg cctatggtgg ccacagtgca gggccagaac ctgaagtacc agggtcagtc 1200

actagtcatc tctgcttcta tcattgtctt caacctgctg gaactggaag gtgattatcg 1260

agatgatggc aacgtgtggg tgcatacccc gctgagcccg cgcaccctga acgcgtgggt 1320

[0200]

gaaagcggtg gaagaaaaaa aaggtattcc agttcaccta gagctggcca gtatgaccaa 1380

catggagctc atgagcagta ttgtgcatca gcaggtcaga acatacggcc ccgtgttcat 1440

gtgtctcggc ggactgctta caatggtggc tggtgctgtg tggctgacag tgcgagtgct 1500

cgagctgttc cgggccgcgc agctggccaa cgacgtggtc ctccagatca tggagctttg 1560

tggtgcagcg tttcgccagg tgtgccatac caccgtgccg tggccgaacg cgagcctgac 1620

cccgaaatgg aacaacgaaa ccacccagcc ccagatcgcc aactgcagcg tgtatgactt 1680

ttttgtgtgg ctccattatt attctgttcg agacacactt tggccaaggg tgacctacca 1740

tatgaacaaa tatgcgtatc atatgctgga aagacgagcc aaatataaaa gaggaccagg 1800

acctggcgct aaatttgtgg ccgcctggac actgaaagcc gctgctggtc ctggacctgg 1860

ccagtacatc aaggccaaca gcaagttcat cggcatcacc gaactcggac ccggaccagg 1920

ctgatgattt cgaaatttaa ataagcttgc ggccgctagg gataacaggg taattatcac 1980

gcccaaacat ttacagccgc ggtgtcaaaa accgcgtgg 2019

<210> 33
<211> 619
<212> PRT
<213> 人工序列(Artificial Sequence)

<220>
<223> 人工序列的描述：合成多肽

<400> 33

```
Met Gln Ile Phe Val Lys Thr Leu Thr Gly Lys Thr Ile Thr Leu Glu
1               5                  10                  15

Val Glu Pro Ser Asp Thr Ile Glu Asn Val Lys Ala Lys Ile Gln Asp
            20                  25                  30

Lys Glu Gly Ile Pro Pro Asp Gln Gln Arg Leu Ile Phe Ala Gly Lys
        35                  40                  45

Gln Leu Glu Asp Gly Arg Thr Leu Ser Asp Tyr Asn Ile Gln Lys Glu
    50                  55                  60
```

[0201]

```
Ser Thr Leu His Leu Val Leu Arg Leu Arg Gly Ala Met Phe Gln Ala
65                  70                  75                  80

Leu Ser Glu Gly Cys Thr Pro Tyr Asp Ile Asn Gln Met Leu Asn Val
                85                  90                  95

Leu Gly Asp His Gln Val Ser Gly Leu Glu Gln Leu Glu Ser Ile Ile
            100                 105                 110

Asn Phe Glu Lys Leu Thr Glu Trp Thr Ser Ser Asn Val Met Pro Ile
        115                 120                 125

Leu Ser Pro Leu Thr Lys Gly Ile Leu Gly Phe Val Phe Thr Leu Thr
    130                 135                 140

Val Pro Ser Glu Arg Gly Leu Ser Cys Ile Ser Glu Ala Asp Ala Thr
145                 150                 155                 160

Thr Pro Glu Ser Ala Asn Leu Gly Glu Glu Ile Leu Ser Gln Leu Tyr
```

165 170 175

Leu Trp Pro Arg Val Thr Tyr His Ser Pro Ser Tyr Ala Tyr His Gln
180 185 190

Phe Glu Arg Arg Ala Lys Tyr Lys Arg His Phe Pro Gly Phe Gly Gln
195 200 205

Ser Leu Leu Phe Gly Tyr Pro Val Tyr Val Phe Gly Asp Cys Val Gln
210 215 220

Gly Asp Trp Asp Ala Ile Arg Phe Arg Tyr Cys Ala Pro Pro Gly Tyr
225 230 235 240

Ala Leu Leu Arg Cys Asn Asp Thr Asn Tyr Ser Ala Leu Leu Ala Val
245 250 255

Gly Ala Leu Glu Gly Pro Arg Asn Gln Asp Trp Leu Gly Val Pro Arg
[0202] 260 265 270

Gln Leu Val Thr Arg Met Gln Ala Ile Gln Asn Ala Gly Leu Cys Thr
275 280 285

Leu Val Ala Met Leu Glu Glu Thr Ile Phe Trp Leu Gln Ala Phe Leu
290 295 300

Met Ala Leu Thr Asp Ser Gly Pro Lys Thr Asn Ile Ile Val Asp Ser
305 310 315 320

Gln Tyr Val Met Gly Ile Ser Lys Pro Ser Phe Gln Glu Phe Val Asp
325 330 335

Trp Glu Asn Val Ser Pro Glu Leu Asn Ser Thr Asp Gln Pro Phe Trp
340 345 350

Gln Ala Gly Ile Leu Ala Arg Asn Leu Val Pro Met Val Ala Thr Val
355 360 365

Gln Gly Gln Asn Leu Lys Tyr Gln Gly Gln Ser Leu Val Ile Ser Ala
370 375 380

Ser Ile Ile Val Phe Asn Leu Leu Glu Leu Glu Gly Asp Tyr Arg Asp
385 390 395 400

Asp Gly Asn Val Trp Val His Thr Pro Leu Ser Pro Arg Thr Leu Asn
405 410 415

Ala Trp Val Lys Ala Val Glu Glu Lys Lys Gly Ile Pro Val His Leu
420 425 430

Glu Leu Ala Ser Met Thr Asn Met Glu Leu Met Ser Ser Ile Val His
435 440 445

Gln Gln Val Arg Thr Tyr Gly Pro Val Phe Met Cys Leu Gly Gly Leu
450 455 460

[0203] Leu Thr Met Val Ala Gly Ala Val Trp Leu Thr Val Arg Val Leu Glu
465 470 475 480

Leu Phe Arg Ala Ala Gln Leu Ala Asn Asp Val Val Leu Gln Ile Met
485 490 495

Glu Leu Cys Gly Ala Ala Phe Arg Gln Val Cys His Thr Thr Val Pro
500 505 510

Trp Pro Asn Ala Ser Leu Thr Pro Lys Trp Asn Asn Glu Thr Thr Gln
515 520 525

Pro Gln Ile Ala Asn Cys Ser Val Tyr Asp Phe Phe Val Trp Leu His
530 535 540

Tyr Tyr Ser Val Arg Asp Thr Leu Trp Pro Arg Val Thr Tyr His Met
545 550 555 560

Asn Lys Tyr Ala Tyr His Met Leu Glu Arg Arg Ala Lys Tyr Lys Arg

565 570 575

Gly Pro Gly Pro Gly Ala Lys Phe Val Ala Ala Trp Thr Leu Lys Ala
580 585 590

Ala Ala Gly Pro Gly Pro Gly Gln Tyr Ile Lys Ala Asn Ser Lys Phe
595 600 605

Ile Gly Ile Thr Glu Leu Gly Pro Gly Pro Gly
610 615

<210> 34
<211> 1638
<212> DNA
<213> 人工序列(Artificial Sequence)

<220>
<223> 人工序列的描述：合成多核苷酸

<400> 34
[0204] atggccggga tgttccaggc actgtccgaa ggctgcacac cctatgatat taaccagatg 60

ctgaatgtcc tgggagacca ccaggtctct ggcctggagc agctggagag catcatcaac 120

ttcgagaagc tgaccgagtg gacaagctcc aatgtgatgc ctatcctgtc cccactgacc 180

aagggcatcc tgggcttcgt gtttaccctg acagtgcctt ctgagcgggg cctgtcttgc 240

atcagcgagg cagacgcaac cacaccagag tccgccaatc tgggcgagga gatcctgtct 300

cagctgtacc tgtggccccg ggtgacatat cactcccctt cttacgccta tcaccagttc 360

gagcggagag ccaagtacaa gagacacttc ccaggctttg gccagtctct gctgttcggc 420

taccccgtgt acgtgttcgg cgattgcgtg cagggcgact gggatgccat ccggtttaga 480

tactgcgcac cacctggata tgcactgctg aggtgtaacg acaccaatta ttccgccctg 540

ctggcagtgg gcgccctgga gggccctcgc aatcaggatt ggctgggcgt gccaaggcag 600

ctggtgacac gcatgcaggc catccagaac gcaggcctgt gcaccctggt ggcaatgctg 660

gaggagacaa tcttctggct gcaggccttt ctgatggccc tgaccgacag cggccccaag 720

acaaacatca tcgtggattc ccagtacgtg atgggcatct ccaagccttc tttccaggag 780

tttgtggact gggagaacgt gagcccagag ctgaattcca ccgatcagcc attctggcag 840

gcaggaatcc tggcaaggaa cctggtgcct atggtggcca cagtgcaggg ccagaatctg 900

aagtaccagg gccagagcct ggtcatcagc gcctccatca tcgtgtttaa cctgctggag 960

ctggagggcg actatcggga cgatggcaac gtgtgggtgc acaccccact gagccccaga 1020

acactgaacg cctgggtgaa ggccgtggag gagaagaagg gcatcccagt gcacctggag 1080

ctggcctcca tgaccaatat ggagctgatg tctagcatcg tgcaccagca ggtgaggaca 1140

tacggacccg tgttcatgtg cctgggaggc ctgctgacca tggtggcagg agccgtgtgg 1200

ctgacagtgc gggtgctgga gctgttcaga gccgcccagc tggccaacga tgtggtgctg 1260

cagatcatgg agctgtgcgg agcagccttt cgccaggtgt gccacaccac agtgccatgg 1320

cccaatgcct ccctgacccc caagtggaac aatgagacaa cacagcctca gatcgccaac 1380

tgtagcgtgt acgacttctt cgtgtggctg cactactata gcgtgaggga taccctgtgg 1440

ccccgcgtga cataccacat gaataagtac gcctatcaca tgctggagag gcgcgccaag 1500

[0205] tataagagag gccctggccc aggcgcaaag tttgtggcag catggaccct gaaggccgcc 1560

gccggccccg gccccggcca gtatatcaag gctaacagta agttcattgg aatcacagag 1620

ctgggacccg gacctgga 1638

<210> 35
<211> 546
<212> PRT
<213> 人工序列(Artificial Sequence)

<220>
<223> 人工序列的描述：合成多肽

<400> 35

Met	Ala	Gly	Met	Phe	Gln	Ala	Leu	Ser	Glu	Gly	Cys	Thr	Pro	Tyr	Asp
1				5				10					15		

Ile	Asn	Gln	Met	Leu	Asn	Val	Leu	Gly	Asp	His	Gln	Val	Ser	Gly	Leu
			20					25					30		

Glu	Gln	Leu	Glu	Ser	Ile	Ile	Asn	Phe	Glu	Lys	Leu	Thr	Glu	Trp	Thr
		35					40					45			

Ser Ser Asn Val Met Pro Ile Leu Ser Pro Leu Thr Lys Gly Ile Leu
50 55 60

Gly Phe Val Phe Thr Leu Thr Val Pro Ser Glu Arg Gly Leu Ser Cys
65 70 75 80

Ile Ser Glu Ala Asp Ala Thr Thr Pro Glu Ser Ala Asn Leu Gly Glu
85 90 95

Glu Ile Leu Ser Gln Leu Tyr Leu Trp Pro Arg Val Thr Tyr His Ser
100 105 110

Pro Ser Tyr Ala Tyr His Gln Phe Glu Arg Arg Ala Lys Tyr Lys Arg
115 120 125

His Phe Pro Gly Phe Gly Gln Ser Leu Leu Phe Gly Tyr Pro Val Tyr
130 135 140

[0206]

Val Phe Gly Asp Cys Val Gln Gly Asp Trp Asp Ala Ile Arg Phe Arg
145 150 155 160

Tyr Cys Ala Pro Pro Gly Tyr Ala Leu Leu Arg Cys Asn Asp Thr Asn
165 170 175

Tyr Ser Ala Leu Leu Ala Val Gly Ala Leu Glu Gly Pro Arg Asn Gln
180 185 190

Asp Trp Leu Gly Val Pro Arg Gln Leu Val Thr Arg Met Gln Ala Ile
195 200 205

Gln Asn Ala Gly Leu Cys Thr Leu Val Ala Met Leu Glu Glu Thr Ile
210 215 220

Phe Trp Leu Gln Ala Phe Leu Met Ala Leu Thr Asp Ser Gly Pro Lys
225 230 235 240

Thr Asn Ile Ile Val Asp Ser Gln Tyr Val Met Gly Ile Ser Lys Pro
245 250 255

Ser Phe Gln Glu Phe Val Asp Trp Glu Asn Val Ser Pro Glu Leu Asn
260 265 270

Ser Thr Asp Gln Pro Phe Trp Gln Ala Gly Ile Leu Ala Arg Asn Leu
275 280 285

Val Pro Met Val Ala Thr Val Gln Gly Gln Asn Leu Lys Tyr Gln Gly
290 295 300

Gln Ser Leu Val Ile Ser Ala Ser Ile Ile Val Phe Asn Leu Leu Glu
305 310 315 320

Leu Glu Gly Asp Tyr Arg Asp Asp Gly Asn Val Trp Val His Thr Pro
325 330 335

[0207]

Leu Ser Pro Arg Thr Leu Asn Ala Trp Val Lys Ala Val Glu Glu Lys
340 345 350

Lys Gly Ile Pro Val His Leu Glu Leu Ala Ser Met Thr Asn Met Glu
355 360 365

Leu Met Ser Ser Ile Val His Gln Gln Val Arg Thr Tyr Gly Pro Val
370 375 380

Phe Met Cys Leu Gly Gly Leu Leu Thr Met Val Ala Gly Ala Val Trp
385 390 395 400

Leu Thr Val Arg Val Leu Glu Leu Phe Arg Ala Ala Gln Leu Ala Asn
405 410 415

Asp Val Val Leu Gln Ile Met Glu Leu Cys Gly Ala Ala Phe Arg Gln
420 425 430

Val Cys His Thr Thr Val Pro Trp Pro Asn Ala Ser Leu Thr Pro Lys
435 440 445

Trp Asn Asn Glu Thr Thr Gln Pro Gln Ile Ala Asn Cys Ser Val Tyr
450 455 460

Asp Phe Phe Val Trp Leu His Tyr Tyr Ser Val Arg Asp Thr Leu Trp
465 470 475 480

Pro Arg Val Thr Tyr His Met Asn Lys Tyr Ala Tyr His Met Leu Glu
485 490 495

Arg Arg Ala Lys Tyr Lys Arg Gly Pro Gly Pro Gly Ala Lys Phe Val
500 505 510

Ala Ala Trp Thr Leu Lys Ala Ala Ala Gly Pro Gly Pro Gly Gln Tyr
515 520 525

Ile Lys Ala Asn Ser Lys Phe Ile Gly Ile Thr Glu Leu Gly Pro Gly
530 535 540

[0208]

Pro Gly
545

<210> 36
<211> 2019
<212> DNA
<213> 人工序列(Artificial Sequence)

<220>
<223> 人工序列的描述：合成多核苷酸

<400> 36
gcccgggcat ttaaatgcga tcgcatcgat tacgactcta gaatagtcta gtccgcaggc 60

caccatgcag atcttcgtga agaccctgac cggcaagacc atcaccctag aggtggagcc 120

cagtgacacc atcgagaacg tgaaggccaa gatccaggat aaagagggca tccccccctga 180

ccagcagagg ctgatctttg ccggcaagca gctggaagat ggccgcaccc tctctgatta 240

caacatccag aaggagtcaa ccctgcacct ggtccttcgc ctgagaggtg ccatgtttca 300

ggcgctgagc gaaggctgca ccccgtatga tattaaccag atgctgaacg tgctgggcga 360

tcatcagttt aagcacatca aagcctttga ccggacattt gctaacaacc caggtcccat 420

ggttgtgttt gccacacctg ggcctatcct gtctcctctg acaaagggca tcctgggctt 480

cgtgtttacc ctgaccgtgc cttctgagag aggacttagc tgcattagcg aagcggatgc 540

gaccaccccg gaaagcgcga acctgggcga agaaattctg agccagctgt atctttggcc 600

aagggtgacc taccattccc ctagttatgc ttaccaccaa tttgaaagac gagccaaata 660

taaaagacac ttccccggct ttggccagag cctgctgttt ggctaccctg tgtacgtgtt 720

cggcgattgc gtgcagggcg attgggatgc gattcgcttt cgctattgcg cgccgccggg 780

ctatgcgctg ctgcgctgca acgataccaa ctatagcgct ctgctggctg tgggggccct 840

agaaggaccc aggaatcagg actggcttgg tgtcccaaga caacttgtaa ctcggatgca 900

ggctattcag aatgccggcc tgtgtaccct ggtggccatg ctggaagaga caatcttctg 960

gctgcaagcg tttctgatgg cgctgaccga tagcggcccg aaaaccaaca ttattgtgga 1020

tagccagtat gtgatgggca ttagcaaacc gagctttcag gaatttgtgg attgggaaaa 1080

[0209] cgtgagcccg gaactgaaca gcaccgatca gccgttttgg caagccggaa tcctggccag 1140

aaatctggtg cctatggtgg ccacagtgca gggccagaac ctgaagtacc agggtcagtc 1200

actagtcatc tctgcttcta tcattgtctt caacctgctg gaactggaag gtgattatcg 1260

agatgatggc aacgtgtggg tgcatacccc gctgagcccg cgcaccctga acgcgtgggt 1320

gaaagcggtg gaagaaaaaa aaggtattcc agttcaccta gagctggcca gtatgaccaa 1380

catggagctc atgagcagta ttgtgcatca gcaggtcaga acatacggcc ccgtgttcat 1440

gtgtctcggc ggactgctta caatggtggc tggtgctgtg tggctgacag tgcgagtgct 1500

cgagctgttc cgggccgcgc agctggccaa cgacgtggtc ctccagatca tggagctttg 1560

tggtgcagcg tttcgccagg tgtgccatac caccgtgccg tggccgaacg cgagcctgac 1620

cccgaaatgg aacaacgaaa ccacccagcc ccagatcgcc aactgcagcg tgtatgactt 1680

ttttgtgtgg ctccattatt attctgttcg agacacactt tggccaaggg tgacctacca 1740

tatgaacaaa tatgcgtatc atatgctgga aagacgagcc aaatataaaa gaggaccagg 1800

acctggcgct aaatttgtgg ccgcctggac actgaaagcc gctgctggtc ctggacctgg 1860

ccagtacatc aaggccaaca gcaagttcat cggcatcacc gaactcggac ccggaccagg 1920

ctgatgattt cgaaatttaa ataagcttgc ggccgctagg gataacaggg taattatcac 1980

gcccaaacat ttacagccgc ggtgtcaaaa accgcgtgg 2019

<210> 37
<211> 619
<212> PRT
<213> 人工序列(Artificial Sequence)

<220>
<223> 人工序列的描述：合成多肽

<400> 37
Met Gln Ile Phe Val Lys Thr Leu Thr Gly Lys Thr Ile Thr Leu Glu
1 5 10 15

Val Glu Pro Ser Asp Thr Ile Glu Asn Val Lys Ala Lys Ile Gln Asp
20 25 30

[0210] Lys Glu Gly Ile Pro Pro Asp Gln Gln Arg Leu Ile Phe Ala Gly Lys
35 40 45

Gln Leu Glu Asp Gly Arg Thr Leu Ser Asp Tyr Asn Ile Gln Lys Glu
50 55 60

Ser Thr Leu His Leu Val Leu Arg Leu Arg Gly Ala Met Phe Gln Ala
65 70 75 80

Leu Ser Glu Gly Cys Thr Pro Tyr Asp Ile Asn Gln Met Leu Asn Val
85 90 95

Leu Gly Asp His Gln Phe Lys His Ile Lys Ala Phe Asp Arg Thr Phe
100 105 110

Ala Asn Asn Pro Gly Pro Met Val Val Phe Ala Thr Pro Gly Pro Ile
115 120 125

Leu Ser Pro Leu Thr Lys Gly Ile Leu Gly Phe Val Phe Thr Leu Thr

130 135 140

Val Pro Ser Glu Arg Gly Leu Ser Cys Ile Ser Glu Ala Asp Ala Thr
145 150 155 160

Thr Pro Glu Ser Ala Asn Leu Gly Glu Glu Ile Leu Ser Gln Leu Tyr
165 170 175

Leu Trp Pro Arg Val Thr Tyr His Ser Pro Ser Tyr Ala Tyr His Gln
180 185 190

Phe Glu Arg Arg Ala Lys Tyr Lys Arg His Phe Pro Gly Phe Gly Gln
195 200 205

Ser Leu Leu Phe Gly Tyr Pro Val Tyr Val Phe Gly Asp Cys Val Gln
210 215 220

Gly Asp Trp Asp Ala Ile Arg Phe Arg Tyr Cys Ala Pro Pro Gly Tyr
[0211] 225 230 235 240

Ala Leu Leu Arg Cys Asn Asp Thr Asn Tyr Ser Ala Leu Leu Ala Val
245 250 255

Gly Ala Leu Glu Gly Pro Arg Asn Gln Asp Trp Leu Gly Val Pro Arg
260 265 270

Gln Leu Val Thr Arg Met Gln Ala Ile Gln Asn Ala Gly Leu Cys Thr
275 280 285

Leu Val Ala Met Leu Glu Glu Thr Ile Phe Trp Leu Gln Ala Phe Leu
290 295 300

Met Ala Leu Thr Asp Ser Gly Pro Lys Thr Asn Ile Ile Val Asp Ser
305 310 315 320

Gln Tyr Val Met Gly Ile Ser Lys Pro Ser Phe Gln Glu Phe Val Asp
325 330 335

Trp Glu Asn Val Ser Pro Glu Leu Asn Ser Thr Asp Gln Pro Phe Trp
340 345 350

Gln Ala Gly Ile Leu Ala Arg Asn Leu Val Pro Met Val Ala Thr Val
355 360 365

Gln Gly Gln Asn Leu Lys Tyr Gln Gly Gln Ser Leu Val Ile Ser Ala
370 375 380

Ser Ile Ile Val Phe Asn Leu Leu Glu Leu Glu Gly Asp Tyr Arg Asp
385 390 395 400

Asp Gly Asn Val Trp Val His Thr Pro Leu Ser Pro Arg Thr Leu Asn
405 410 415

Ala Trp Val Lys Ala Val Glu Glu Lys Lys Gly Ile Pro Val His Leu
420 425 430

[0212] Glu Leu Ala Ser Met Thr Asn Met Glu Leu Met Ser Ser Ile Val His
435 440 445

Gln Gln Val Arg Thr Tyr Gly Pro Val Phe Met Cys Leu Gly Gly Leu
450 455 460

Leu Thr Met Val Ala Gly Ala Val Trp Leu Thr Val Arg Val Leu Glu
465 470 475 480

Leu Phe Arg Ala Ala Gln Leu Ala Asn Asp Val Val Leu Gln Ile Met
485 490 495

Glu Leu Cys Gly Ala Ala Phe Arg Gln Val Cys His Thr Thr Val Pro
500 505 510

Trp Pro Asn Ala Ser Leu Thr Pro Lys Trp Asn Asn Glu Thr Thr Gln
515 520 525

Pro Gln Ile Ala Asn Cys Ser Val Tyr Asp Phe Phe Val Trp Leu His

530 535 540

Tyr Tyr Ser Val Arg Asp Thr Leu Trp Pro Arg Val Thr Tyr His Met
545 550 555 560

Asn Lys Tyr Ala Tyr His Met Leu Glu Arg Arg Ala Lys Tyr Lys Arg
565 570 575

Gly Pro Gly Pro Gly Ala Lys Phe Val Ala Ala Trp Thr Leu Lys Ala
580 585 590

Ala Ala Gly Pro Gly Pro Gly Gln Tyr Ile Lys Ala Asn Ser Lys Phe
595 600 605

Ile Gly Ile Thr Glu Leu Gly Pro Gly Pro Gly
610 615

[0213]

<210> 38
<211> 228
<212> DNA
<213> 人工序列(Artificial Sequence)

<220>
<223> 人工序列的描述：合成多核苷酸

<400> 38
atgcagatct tcgtgaagac cctgaccggc aagaccatca ccctagaggt ggagcccagt 60

gacaccatcg agaacgtgaa ggccaagatc caggataaag agggcatccc ccctgaccag 120

cagaggctga tctttgccgg caagcagctg gaagatggcc gcaccctctc tgattacaac 180

atccagaagg agtcaaccct gcacctggtc cttcgcctga gaggtggc 228

<210> 39
<211> 228
<212> DNA
<213> 人工序列(Artificial Sequence)

<220>
<223> 人工序列的描述：合成多核苷酸

<400> 39

atgcagatct tcgtgaagac cctgaccggc aagaccatca ccctagaggt ggagcccagt 60

gacaccatcg agaacgtgaa ggccaagatc caggataaag agggcatccc ccctgaccag 120

cagaggctga tctttgccgg caagcagctg gaagatggcc gcaccctctc tgattacaac 180

atccagaagg agtcaaccct gcacctggtc cttcgcctga gaggtgcc 228

<210> 40
<211> 78
<212> DNA
<213> 智人(Homo sapiens)

<400> 40
atggccgtca tggcgccccg aaccctcgtc ctgctactct cgggggctct ggccctgacc 60

cagacctggg cgggctct 78

<210> 41
<211> 201
<212> DNA
<213> 智人(Homo sapiens)

[0214]

<400> 41
ccgtcttccc agcccaccat ccccatcgtg ggcatcattg ctggcctggt tctctttgga 60

gctgtgatca ctggagctgt ggtcgctgct gtgatgtgga ggaggaagag ctcagataga 120

aaaggaggga gctactctca ggctgcaagc agtgacagtg cccagggctc tgatgtgtct 180

ctcacagctt gtaaagtgtg a 201

<210> 42
<211> 60
<212> DNA
<213> 人工序列(Artificial Sequence)

<220>
<223> 人工序列的描述：合成寡核苷酸

<400> 42
atggagaccg atacactgct gctgtgggtg ctgctcctgt gggtgccagg aagcacaggc 60

<210> 43
<211> 3178
<212> DNA

<213> 智人(Homo sapiens)

<400> 43

[0215]

ggcaccgatt cggggcctgc ccggacttcg ccgcacgctg cagaacctcg cccagcgccc 60

accatgcccc ggcagctcag cgcggcggcc gcgctcttcg cgtccctggc cgtaattttg 120

cacgatggca gtcaaatgag agcaaaagca tttccagaaa ccagagatta ttctcaacct 180

actgcagcag caacagtaca ggacataaaa aaacctgtcc agcaaccagc taagcaagca 240

cctcaccaaa ctttagcagc aagattcatg gatggtcata tcacctttca aacagcggcc 300

acagtaaaaa ttccaacaac taccccagca actacaaaaa acactgcaac caccagccca 360

attacctaca ccctggtcac aacccaggcc acacccaaca actcacacac agctcctcca 420

gttactgaag ttacagtcgg ccctagctta gcccccttatt cactgccacc caccatcacc 480

ccaccagctc atacagctgg aaccagttca tcaaccgtca gccacacaac tgggaacacc 540

actcaaccca gtaaccagac cacccttcca gcaactttat cgatagcact gcacaaaagc 600

acaaccggtc agaagcctga tcaacccacc catgccccag gaacaacggc agctgcccac 660

aataccaccc gcacagctgc acctgcctcc acggttcctg ggcccaccct tgcacctcag 720

ccatcgtcag tcaagactgg aatttatcag gttctaaacg gaagcagact ctgtataaaa 780

gcagagatgg ggatacagct gattgttcaa gacaaggagt cggttttttc acctcggaga 840

tacttcaaca tcgaccccaa cgcaacgcaa gcctctggga actgtggcac ccgaaaatcc 900

aaccttctgt tgaattttca gggcggattt gtgaatctca catttaccaa ggatgaagaa 960

tcatattata tcagtgaagt gggagcctat ttgaccgtct cagatccaga gacagtttac 1020

caaggaatca aacatgcggt ggtgatgttc cagacagcag tcgggcattc cttcaagtgc 1080

gtgagtgaac agagcctcca gttgtcagcc cacctgcagg tgaaaacaac cgatgtccaa 1140

cttcaagcct ttgattttga agatgaccac tttggaaatg tggatgagtg ctcgtctgac 1200

tacacaattg tgcttcctgt gattggggcc atcgtggttg gtctctgcct tatgggtatg 1260

ggtgtctata aaatccgcct aaggtgtcaa tcatctggat accagagaat ctaattgttg 1320

cccgggggga atgaaaataa tggaatttag agaactcttt catcccttcc aggatggatg 1380

ttgggaaatt ccctcagagt gtgggtcctt caaacaatgt aaaccaccat cttctattca 1440

aatgaagtga gtcatgtgtg atttaagttc aggcagcaca tcaatttcta aatacttttt 1500

gtttatttta tgaaagatat agtgagctgt ttattttcta gtttccttta gaatatttta 1560

gccactcaaa gtcaacattt gagatatgtt gaattaacat aatatatgta aagtagaata 1620

agccttcaaa ttataaacca agggtcaatt gtaactaata ctactgtgtg tgcattgaag 1680

attttatttt acccttgatc ttaacaaagc ctttgctttg ttatcaaatg gactttcagt 1740

gcttttacta tctgtgtttt atggtttcat gtaacataca tattcctggt gtagcactta 1800

actccttttc cactttaaat ttgtttttgt tttttgagac ggagtttcac tcttgtcacc 1860

caggctggag tacagtggca cgatctcggc ttatggcaac ctccgcctcc cgggttcaag 1920

tgattctcct gcttcagctt cccgagtagc tgggattaca ggcacacact accacgcctg 1980

gctaattttt gtatttttat tatagacggg tttcaccatg ttggccagac tggtcttgaa 2040

ctcttgacct caggtgatcc acccacctca gcctcccaaa gtgctgggat tacaggcatg 2100

agccattgcg cccggcctta aatgtttttt ttaatcatca aaaagaacaa catatctcag 2160

[0216] gttgtctaag tgtttttatg taaaaccaac aaaaagaaca aatcagctta tatttttat 2220

cttgatgact cctgctccag aattgctaga ctaagaatta ggtggctaca gatggtagaa 2280

ctaaacaata agcaagagac aataataatg gcccttaatt attaacaaag tgccagagtc 2340

taggctaagc actttatcta tatctcattt cattctcaca acttataagt gaatgagtaa 2400

actgagactt aagggaactg aatcacttaa atgtcacctg gctaactgat ggcagagcca 2460

gagcttgaat tcatgttggt ctgacatcaa ggtctttggt cttctcccta caccaagtta 2520

cctacaagaa caatgacacc acactctgcc tgaaggctca cacctcatac cagcatacgc 2580

tcaccttaca gggaaatggg tttatccagg atcatgagac attagggtag atgaaaggag 2640

agctttgcag ataacaaaat agcctatcct taataaatcc tccactctct ggaaggagac 2700

tgaggggctt tgtaaaacat tagtcagttg ctcattttta tgggattgct tagctgggct 2760

gtaaagatga aggcatcaaa taaactcaaa gtatttttaa atttttttga taatagagaa 2820

acttcgctaa ccaactgttc tttcttgagt gtatagcccc atcttgtggt aacttgctgc 2880

ttctgcactt catatccata tttcctattg ttcactttat tctgtagagc agcctgccaa 2940

gaattttatt tctgctgttt tttttgctgc taaagaaagg aactaagtca ggatgttaac 3000

agaaaagtcc acataaccct agaattctta gtcaaggaat aattcaagtc agcctagaga 3060

ccatgttgac tttcctcatg tgtttcctta tgactcagta agttggcaag gtcctgactt 3120

tagtcttaat aaaacattga attgtagtaa aggtttttgc aataaaaact tactttgg 3178

<210> 44
<211> 1858
<212> DNA
<213> 小鼠属某种(Mus sp.)

<400> 44
attccggagg tgaaaaacaa tggcacaacg tgtataatgg ccagcttctc tgcctccttt 60

ctgaccacct acgagactgc gaatggttct cagatcgtga acatttccct gccagcctct 120

gcagaagtac tgaaaaatgg cagttcttgt ggtaaagaaa atgtttctga ccccagcctc 180

acaattactt ttggaagagg atatttactg acactcaact tcacaaaaaa tacaacacgt 240

[0217] tacagtgtcc agcatatgta ttttacatat aacttgtcag atacagaaca ttttcccaat 300

gccatcagca aagagatcta caccatggat tccacaactg acatcaaggc agacatcaac 360

aaagcatacc ggtgtgtcag tgatatccgg gtctacatga agaatgtgac cgttgtgctc 420

cgggatgcca ctatccaggc ctacctgtcg agtggcaact tcagcaagga agagacacac 480

tgcacacagg atggaccttc cccaaccact gggccaccca gcccctcacc accacttgtg 540

cccacaaacc ccactgtatc caagtacaat gttactggta acaacggaac ctgcctgctg 600

gcctctatgg cactgcaact gaatatcacc tacctgaaaa aggacaacaa gacggtgacc 660

agagcgttca acatcagccc aaatgacaca tctagtggga gttgcggtat caacttggtg 720

accctgaaag tggagaacaa gaacagagcc ctggaattgc agtttgggat gaatgccagc 780

tctagcctgt ttttcttgca aggagtgcgc ttgaatatga ctcttcctga tgccctagtg 840

cccacattca gcatctccaa ccattcactg aaagctcttc aggccactgt gggaaactca 900

tacaagtgca acactgagga acacatcttt gtcagcaaga tgctctccct caatgtcttc 960

agtgtgcagg tccaggcttt caaggtggac agtgacaggt ttgggtctgt ggaagagtgt 1020

gttcaggatg gtaacaacat gttgatcccc attgctgtgg gcggtgccct ggcagggctg 1080

atcctcatcg tcctcattgc ctacctcatt ggcaggaaga ggagtcacgc cggctatcag 1140

accatctagc ctggtgggca ggtgcaccag agatgcacag gggcctgttc tcacatcccc 1200

aagcttagat aggtgtggaa gggaggcaca ctttctggca aactgtttta aaatctgctt 1260

tatcaaatgt gaagttcatc ttgcaacatt tactatgcac aaaggaataa ctattgaaat 1320

gacggtgtta attttgctaa ctgggttaaa tattgatgag aaggctccac tgatttgact 1380

tttaagactt ggtgtttggt tcttcattct tttactcaga tttaagccta tcaaagggat 1440

actctggtcc agaccttggc ctggcaaggg tggctgatgg ttaggctgca cacacttaag 1500

aagcaacggg agcagggaag gcttgcacac aggcacgcac agggtcaacc tctggacact 1560

tggcttgggc tacctggcct tgggggggct gaactctggc atctggctgg gtacacaccc 1620

ccccaatttc tgtgctctgc cacccgtgag ctgccacttt cctaaataga aaatggcatt 1680

atttttattt actttttgt aaagtgattt ccagtcttgt gttggcgttc agggtggccc 1740

[0218] tgtctctgca ctgtgtacaa taatagattc acactgctga cgtgtcttgc agcgtaggtg 1800

ggttgtacac tgggcatcag ctcacgtaat gcattgcctg taacgatgct aataaaaa 1858

<210> 45
<211> 2339
<212> DNA
<213> 智人(Homo sapiens)

<400> 45
ggcccaaccg ccgcccgcgc ccccgctctc cgcaccgtac ccggccgcct cgcgccatgg 60

cggcccccgg cagcgcccgg cgacccctgc tgctgctact gctgttgctg ctgctcggcc 120

tcatgcattg tgcgtcagca gcaatgttta tggtgaaaaa tggcaacggg accgcgtgca 180

taatggccaa cttctctgct gccttctcag tgaactacga caccaagagt ggccctaaga 240

acatgacctt tgacctgcca tcagatgcca cagtggtgct caaccgcagc tcctgtggaa 300

aagagaacac ttctgacccc agtctcgtga ttgcttttgg aagaggacat acactcactc 360

tcaatttcac gagaaatgca acacgttaca gcgtccagct catgagtttt gtttataact 420

tgtcagacac acacctttc cccaatgcga gctccaaaga aatcaagact gtggaatcta 480

taactgacat cagggcagat atagataaaa aatacagatg tgttagtggc acccaggtcc 540

acatgaacaa cgtgaccgta acgctccatg atgccaccat ccaggcgtac ctttccaaca 600

gcagcttcag caggggagag acacgctgtg aacaagacag gccttcccca accacagcgc 660

cccctgcgcc acccagcccc tcgccctcac ccgtgcccaa gagcccctct gtggacaagt 720

acaacgtgag cggcaccaac gggacctgcc tgctggccag catggggctg cagctgaacc 780

tcacctatga gaggaaggac aacacgacgg tgacaaggct tctcaacatc aaccccaaca 840

agacctcggc cagcgggagc tgcggcgccc acctggtgac tctggagctg cacagcgagg 900

gcaccaccgt cctgctcttc cagttcggga tgaatgcaag ttctagccgg tttttcctac 960

aaggaatcca gttgaataca attcttcctg acgccagaga ccctgccttt aaagctgcca 1020

acggctccct gcgagcgctg caggccacag tcggcaattc ctacaagtgc aacgcggagg 1080

agcacgtccg tgtcacgaag gcgttttcag tcaatatatt caaagtgtgg gtccaggctt 1140

tcaaggtgga aggtggccag tttggctctg tggaggagtg tctgctggac gagaacagca 1200

[0219] tgctgatccc catcgctgtg ggtggtgccc tggcggggct ggtcctcatc gtcctcatcg 1260

cctacctcgt cggcaggaag aggagtcacg caggctacca gactatctag cctggtgcac 1320

gcaggcacag cagctgcagg ggcctctgtt cctttctctg ggcttagggt cctgtcgaag 1380

gggaggcaca ctttctggca aacgtttctc aaatctgctt catccaatgt gaagttcatc 1440

ttgcagcatt tactatgcac aacagagtaa ctatcgaaat gacggtgtta attttgctaa 1500

ctgggttaaa tattttgcta actggttaaa cattaatatt taccaaagta ggattttgag 1560

ggtgggggtg ctctctctga gggggtgggg gtgccgctgt ctctgagggg tgggggtgcc 1620

gctgtctctg aggggtgggg gtgccgctct ctctgagggg gtgggggtgc cgctttctct 1680

gagggggtgg gggtgccgct ctctctgagg gggtgggggt gctgctctct ccgaggggtg 1740

gaatgccgct gtctctgagg ggtgggggtg ccgctctaaa ttggctccat atcatttgag 1800

tttagggttc tggtgtttgg tttcttcatt ctttactgca ctcagattta agccttacaa 1860

agggaaagcc tctggccgtc acacgtagga cgcatgaagg tcactcgtgg tgaggctgac 1920

atgctcacac attacaacag tagagaggga aaatcctaag acagaggaac tccagagatg 1980

agtgtctgga gcgcttcagt tcagctttaa aggccaggac gggccacacg tggctggcgg 2040

cctcgttcca gtggcggcac gtccttgggc gtctctaatg tctgcagctc aagggctggc 2100

actttttaa atataaaaat gggtgttatt tttatttttt tttgtaaagt gatttttggt 2160

cttctgttga cattcggggt gatcctgttc tgcgctgtgt acaatgtgag atcggtgcgt 2220

tctcctgatg ttttgccgtg gcttggggat tgtacacggg accagctcac gtaatgcatt 2280

gcctgtaaca atgtaataaa aagcctcttt cttttaaaaa aaaaaaaaaa aaaaaaaaa 2339

<210> 46
<211> 45
<212> DNA
<213> 人工序列(Artificial Sequence)

<220>
<223> 人工序列的描述：合成寡核苷酸

<400> 46
cagtacatca aggccaacag caagttcatc ggcatcaccg aactc 45

[0220]

<210> 47
<211> 15
<212> PRT
<213> 人工序列(Artificial Sequence)

<220>
<223> 人工序列的描述：合成肽

<400> 47
Gln Tyr Ile Lys Ala Asn Ser Lys Phe Ile Gly Ile Thr Glu Leu
1 5 10 15

<210> 48
<211> 39
<212> DNA
<213> 人工序列(Artificial Sequence)

<220>
<223> 人工序列的描述：合成寡核苷酸

<400> 48
gctaaatttg tggctgcctg gacactgaaa gccgccgct 39

<210> 49
<211> 13
<212> PRT
<213> 人工序列(Artificial Sequence)

<220>
<223> 人工序列的描述：合成肽

<400> 49
Ala Lys Phe Val Ala Ala Trp Thr Leu Lys Ala Ala Ala
1 5 10

<210> 50
<211> 593
<212> DNA
<213> 土拨鼠肝炎病毒(Woodchuck hepatitis virus)

<400> 50
aatcaacctc tggattacaa aatttgtgaa agattgactg gtattcttaa ctatgttgct 60

ccttttacgc tatgtggata cgctgcttta atgcctttgt atcatgctat tgcttcccgt 120

atggctttca ttttctcctc cttgtataaa tcctggttgc tgtctcttta tgaggagttg 180

[0221]

tggcccgttg tcaggcaacg tggcgtggtg tgcactgtgt ttgctgacgc aacccccact 240

ggttggggca ttgccaccac ctgtcagctc ctttccggga ctttcgcttt ccccctccct 300

attgccacgg cggaactcat cgccgcctgc cttgcccgct gctggacagg ggctcggctg 360

ttgggcactg acaattccgt ggtgttgtcg gggaagctga cgtcctttcc atggctgctc 420

gcctgtgttg ccacctggat tctgcgcggg acgtccttct gctacgtccc ttcggccctc 480

aatccagcgg accttccttc ccgcggcctg ctgccggctc tgcggcctct tccgcgtctt 540

cgccttcgcc ctcagacgag tcggatctcc ctttgggccg cctccccgcc tgt 593

<210> 51
<211> 589
<212> DNA
<213> 人工序列(Artificial Sequence)

<220>
<223> 人工序列的描述：合成多核苷酸

<400> 51

tctccccccc ccccctctcc ctcccccccc cctaacgtta ctggccgaag ccgcttggaa 60

taaggccggt gtgcgtttgt ctatatgtta ttttccacca tattgccgtc ttttggcaat 120

gtgagggccc ggaaacctgg ccctgtcttc ttgacgagca ttcctagggg tctttccct 180

ctcgccaaag gaatgcaagg tctgttgaat gtcgtgaagg aagcagttcc tctggaagct 240

tcttgaagac aaacaacgtc tgtagcgacc ctttgcaggc agcggaaccc cccacctggc 300

gacaggtgcc tctgcggcca aaagccacgt gtataagata cacctgcaaa ggcggcacaa 360

ccccagtgcc acgttgtgag ttggatagtt gtggaaagag tcaaatggct ctcctcaagc 420

gtattcaaca aggggctgaa ggatgcccag aaggtacccc attgtatggg atctgatctg 480

gggcctcggt gcacatgctt tacatgtgtt tagtcgaggt taaaaaaacg tctaggcccc 540

ccgaaccacg gggacgtggt tttcctttga aaaacacgat gataatatg 589

<210> 52
<211> 720
<212> DNA
[0222] <213> 人工序列(Artificial Sequence)

<220>
<223> 人工序列的描述：合成多核苷酸

<400> 52
atggtgagca agggcgagga gctgttcacc ggggtggtgc ccatcctggt cgagctggac 60

ggcgacgtaa acggccacaa gttcagcgtg tccggcgagg gcgagggcga tgccacctac 120

ggcaagctga ccctgaagtt catctgcacc accggcaagc tgcccgtgcc ctggcccacc 180

ctcgtgacca ccctgaccta cggcgtgcag tgcttcagcc gctaccccga ccacatgaag 240

cagcacgact tcttcaagtc cgccatgccc gaaggctacg tccaggagcg caccatcttc 300

ttcaaggacg acggcaacta caagacccgc gccgaggtga agttcgaggg cgacaccctg 360

gtgaaccgca tcgagctgaa gggcatcgac ttcaaggagg acggcaacat cctggggcac 420

aagctggagt acaactacaa cagccacaac gtctatatca tggccgacaa gcagaagaac 480

ggcatcaagg tgaacttcaa gatccgccac aacatcgagg acggcagcgt gcagctcgcc 540

gaccactacc agcagaacac ccccatcggc gacggccccg tgctgctgcc cgacaaccac 600

tacctgagca cccagtccgc cctgagcaaa gaccccaacg agaagcgcga tcacatggtc 660

ctgctggagt tcgtgaccgc cgccgggatc actctcggca tggacgagct gtacaagtag 720

<210> 53
<211> 1563
<212> DNA
<213> 人工序列(Artificial Sequence)

<220>
<223> 人工序列的描述：合成多核苷酸

<400> 53
atgctgctgc tgctgctgct gctgggcctg aggctacagc tctccctggg catcatccca 60

gttgaggagg agaacccgga cttctggaac cgcgaggcag ccgaggccct gggtgccgcc 120

aagaagctgc agcctgcaca gacagccgcc aagaacctca tcatcttcct gggcgatggg 180

atgggggtgt ctacggtgac agctgccagg atcctaaaag ggcagaagaa ggacaaactg 240

gggcctgaga taccctggc catggaccgc ttcccatatg tggctctgtc caagacatac 300

[0223] aatgtagaca aacatgtgcc agacagtgga gccacagcca cggcctacct gtgcggggtc 360

aagggcaact tccagaccat tggcttgagt gcagccgccc gctttaacca gtgcaacacg 420

acacgcggca acgaggtcat ctccgtgatg aatcgggcca agaaagcagg gaagtcagtg 480

ggagtggtaa ccaccacacg agtgcagcac gcctcgccag ccggcaccta cgcccacacg 540

gtgaaccgca actggtactc ggacgccgac gtgcctgcct cggcccgcca ggaggggtgc 600

caggacatcg ctacgcagct catctccaac atggacattg acgtgatcct aggtggaggc 660

cgaaagtaca tgtttcgcat gggaacccca gaccctgagt acccagatga ctacagccaa 720

ggtgggacca ggctggacgg gaagaatctg gtgcaggaat ggctggcgaa gcgccagggt 780

gcccggtatg tgtggaaccg cactgagctc atgcaggctt ccctggaccc gtctgtgacc 840

catctcatgg gtctctttga gcctggagac atgaaatacg agatccaccg agactccaca 900

ctggacccct ccctgatgga gatgacagag gctgccctgc gcctgctgag caggaacccc 960

cgcggcttct tcctcttcgt ggagggtggt cgcatcgacc atggtcatca tgaaagcagg 1020

gcttaccggg cactgactga gacgatcatg ttcgacgacg ccattgagag ggcgggccag 1080

ctcaccagcg aggaggacac gctgagcctc gtcactgccg accactccca cgtcttctcc 1140

ttcggaggct acccccctgcg agggagctcc atcttcgggc tggcccctgg caaggcccgg 1200

gacaggaagg cctacacggt cctcctatac ggaaacggtc caggctatgt gctcaaggac 1260

ggcgcccggc cggatgttac cgagagcgag agcgggagcc ccgagtatcg gcagcagtca 1320

gcagtgcccc tggacgaaga gacccacgca ggcgaggacg tggcggtgtt cgcgcgcggc 1380

ccgcaggcgc acctggttca cggcgtgcag gagcagacct tcatagcgca cgtcatggcc 1440

ttcgccgcct gcctggagcc ctacaccgcc tgcgacctgg cgcccccccgc cggcaccacc 1500

gacgccgcgc acccgggtta ctctagagtc ggggcggccg gccgcttcga gcagacatga 1560

taa 1563

<210> 54
<211> 1653
<212> DNA
<213> 人工序列(Artificial Sequence)

[0224] <220>
<223> 人工序列的描述：合成多核苷酸

<400> 54
atggaagatg ccaaaaacat taagaagggc ccagcgccat tctacccact cgaagacggg 60

accgccggcg agcagctgca caaagccatg aagcgctacg ccctggtgcc cggcaccatc 120

gcctttaccg acgcacatat cgaggtggac attacctacg ccgagtactt cgagatgagc 180

gttcggctgg cagaagctat gaagcgctat gggctgaata caaaccatcg gatcgtggtg 240

tgcagcgaga atagcttgca gttcttcatg cccgtgttgg gtgccctgtt catcggtgtg 300

gctgtggccc cagctaacga catctacaac gagcgcgagc tgctgaacag catgggcatc 360

agccagccca ccgtcgtatt cgtgagcaag aaagggctgc aaaagatcct caacgtgcaa 420

aagaagctac cgatcataca aaagatcatc atcatggata gcaagaccga ctaccagggc 480

ttccaaagca tgtacacctt cgtgacttcc catttgccac ccggcttcaa cgagtacgac 540

ttcgtgcccg agagcttcga ccgggacaaa accatcgccc tgatcatgaa cagtagtggc 600

agtaccggat tgcccaaggg cgtagccctа ccgcaccgca ccgcttgtgt ccgattcagt 660

catgcccgcg accccatctt cggcaaccag atcatccccg acaccgctat cctcagcgtg 720

gtgccatttc accacggctt cggcatgttc accacgctgg gctacttgat ctgcggcttt 780

cgggtcgtgc tcatgtaccg cttcgaggag gagctattct tgcgcagctt gcaagactat 840

aagattcaat ctgccctgct ggtgcccaca ctatttagct tcttcgctaa gagcactctc 900

atcgacaagt acgacctaag caacttgcac gagatcgcca gcggcggggc gccgctcagc 960

aaggaggtag gtgaggccgt ggccaaacgc ttccacctac caggcatccg ccagggctac 1020

ggcctgacag aaacaaccag cgccattctg atcacccccg aaggggacga caagcctggc 1080

gcagtaggca aggtggtgcc cttcttcgag gctaaggtgg tggacttgga caccggtaag 1140

acactgggtg tgaaccagcg cggcgagctg tgcgtccgtg gccccatgat catgagcggc 1200

tacgttaaca accccgaggc tacaaacgct ctcatcgaca aggacggctg gctgcacagc 1260

ggcgacatcg cctactggga cgaggacgag cacttcttca tcgtggaccg gctgaagagc 1320

ctgatcaaat acaagggcta ccaggtagcc ccagccgaac tggagagcat cctgctgcaa 1380

[0225] caccccaaca tcttcgacgc cggggtcgcc ggcctgcccg acgacgatgc cggcgagctg 1440

cccgccgcag tcgtcgtgct ggaacacggt aaaaccatga ccgagaagga gatcgtggac 1500

tatgtggcca gccaggttac aaccgccaag aagctgcgcg gtggtgttgt gttcgtggac 1560

gaggtgccta aaggactgac cggcaagttg gacgcccgca agatccgcga gattctcatt 1620

aaggccaaga agggcggcaa gatcgccgtg taa 1653

<210> 55
<211> 66
<212> DNA
<213> 口蹄疫病毒(Foot-and-mouth disease virus)

<400> 55
gtaaagcaaa cactgaactt tgaccttctc aagttggctg gagacgttga gtccaatcct 60

gggccc 66

<210> 56
<211> 5
<212> PRT
<213> 人工序列(Artificial Sequence)

<220>
<223> 人工序列的描述：合成肽

[0226]

<400> 56
Gly Pro Gly Pro Gly
1 5

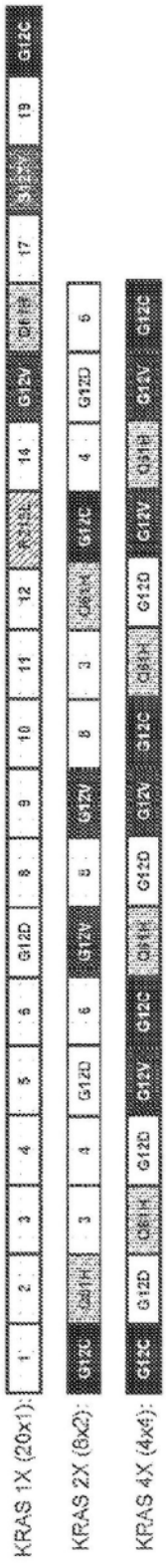


图1A

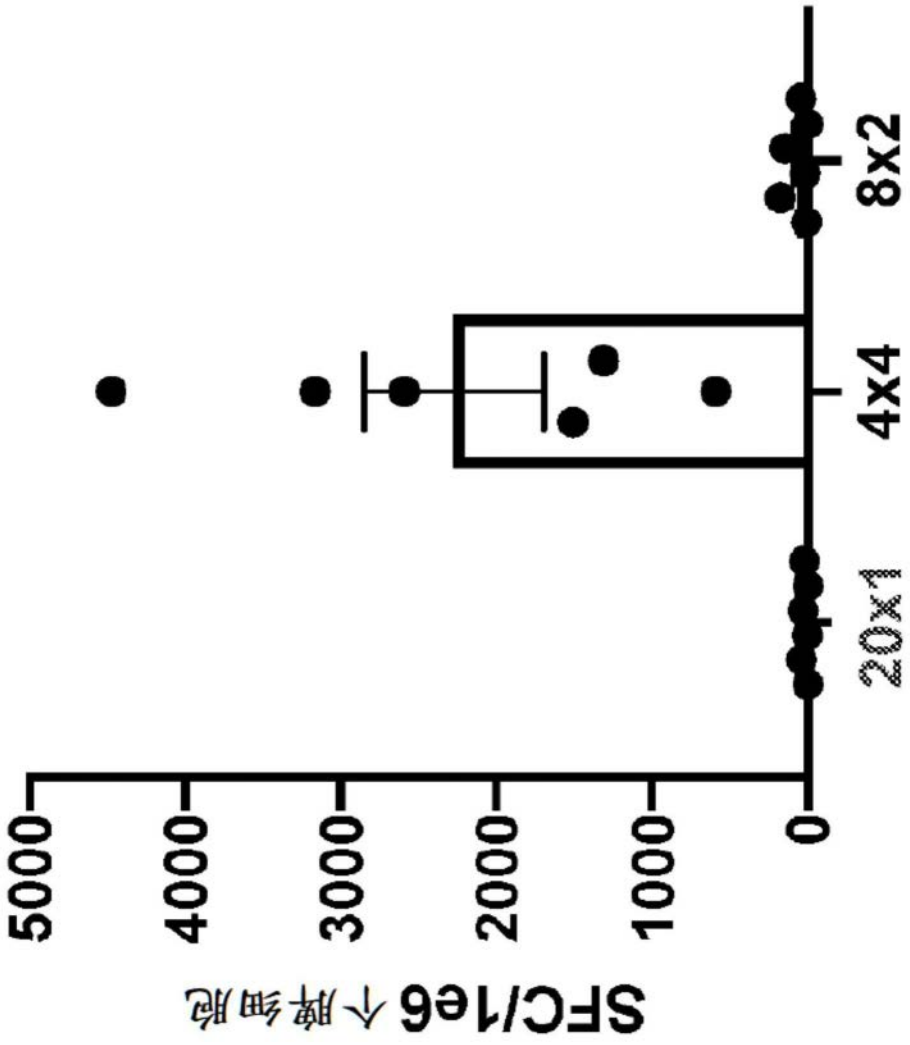


图1B

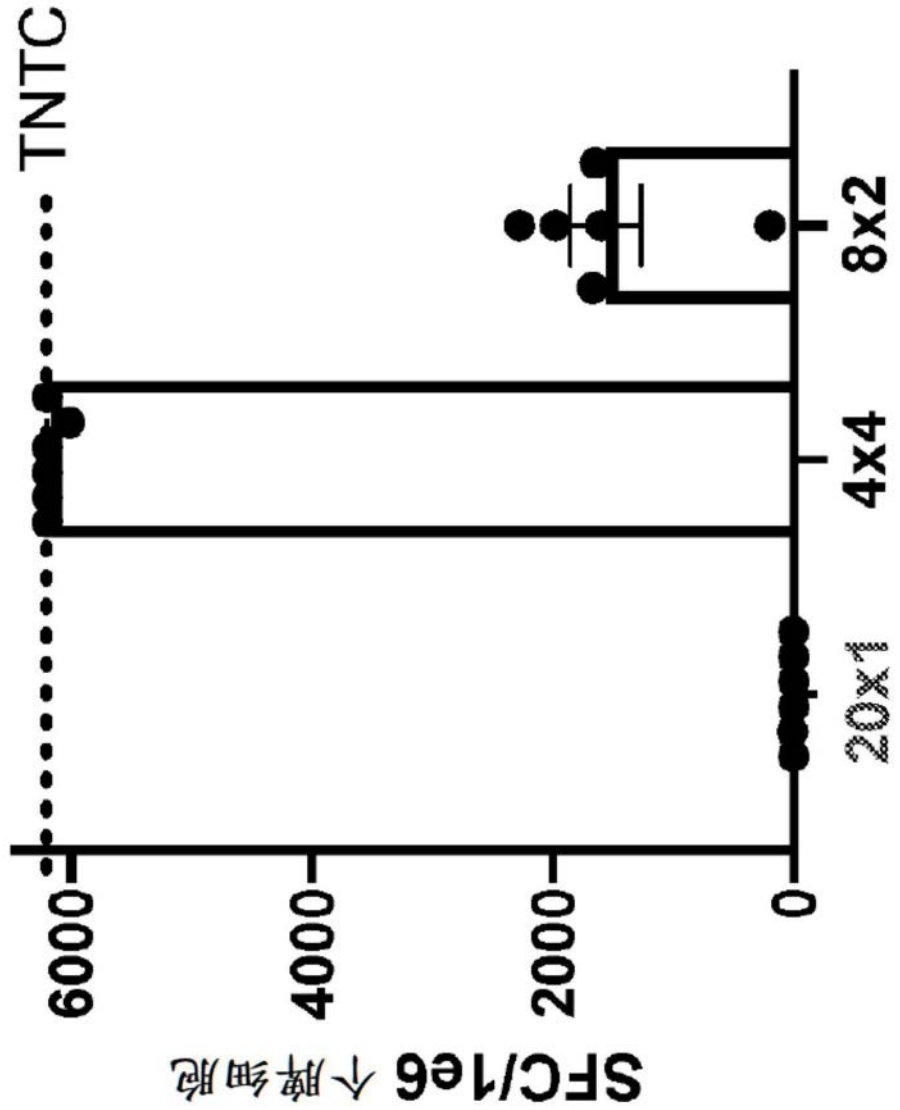


图1C

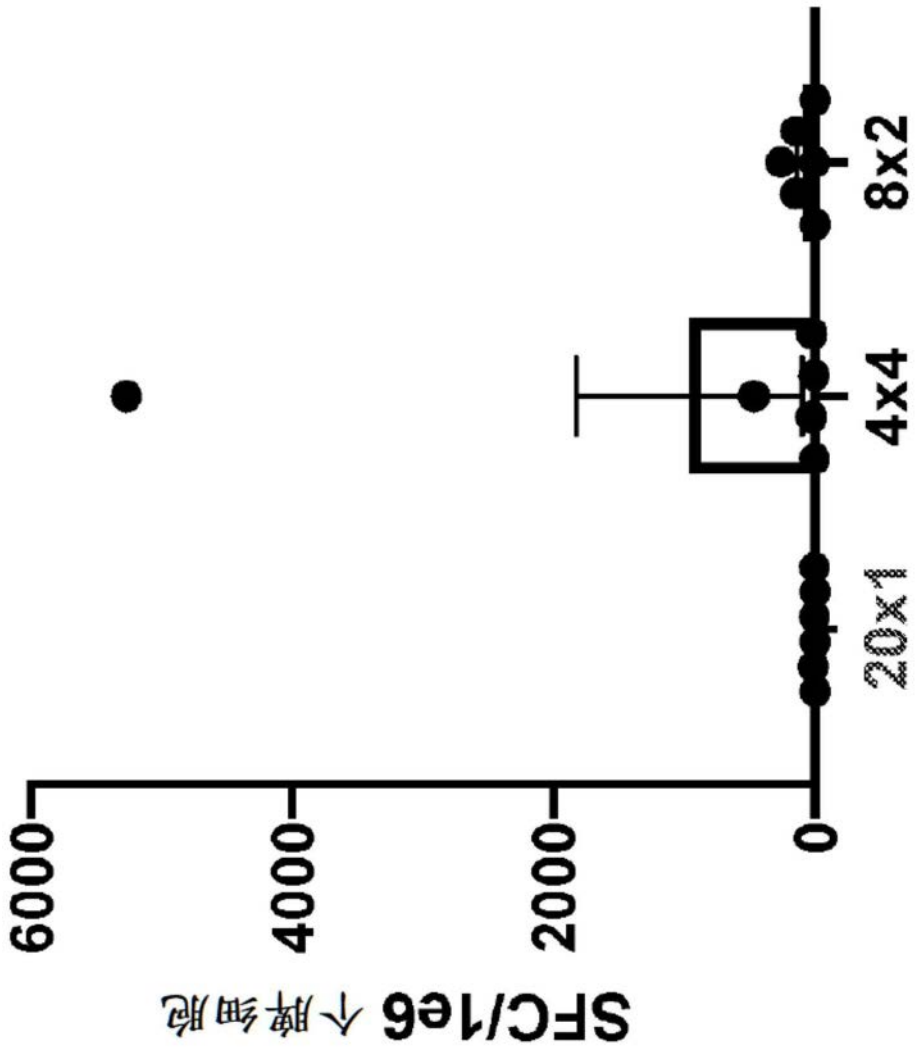


图1D

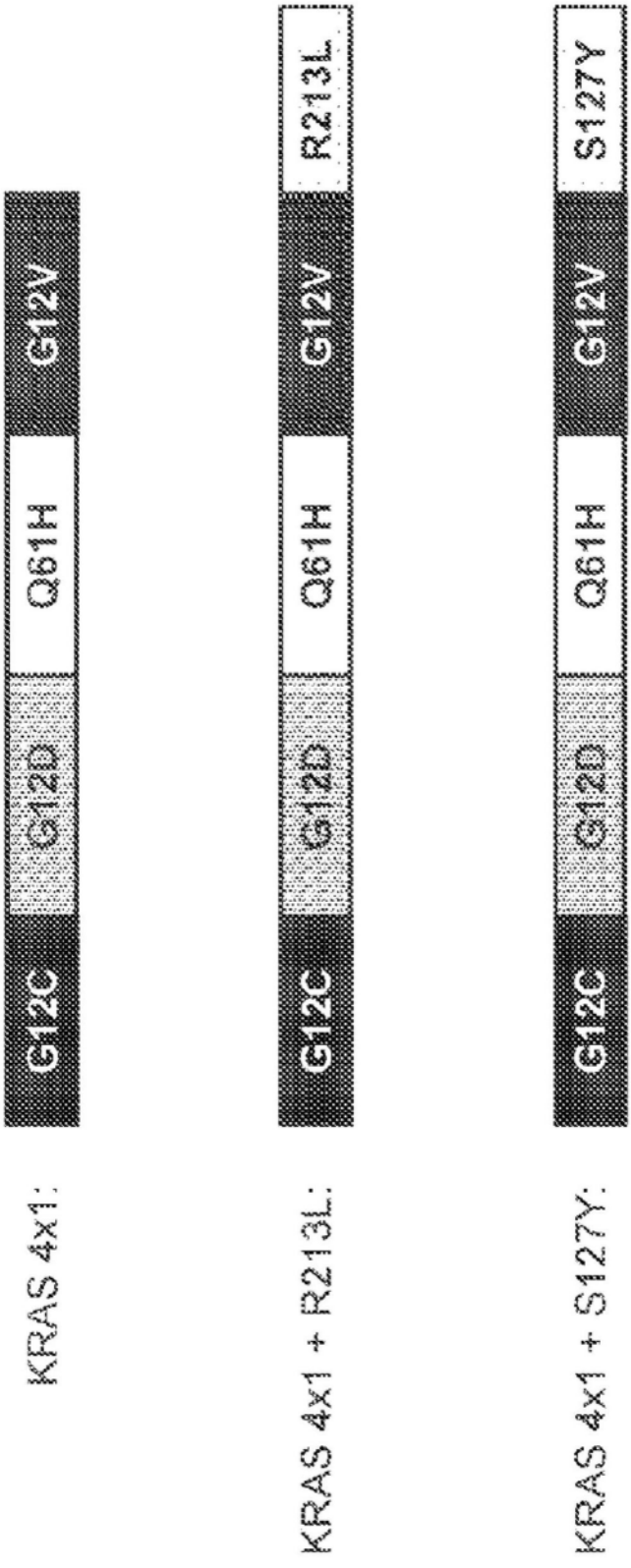

KRAS 4x1:
G12C
G12D
Q61H
G12V
KRAS 4x1 + R213L:
G12C
G12D
Q61H
G12V
R213L
KRAS 4x1 + S127Y:
G12C
G12D
Q61H
G12V
S127Y

图2A

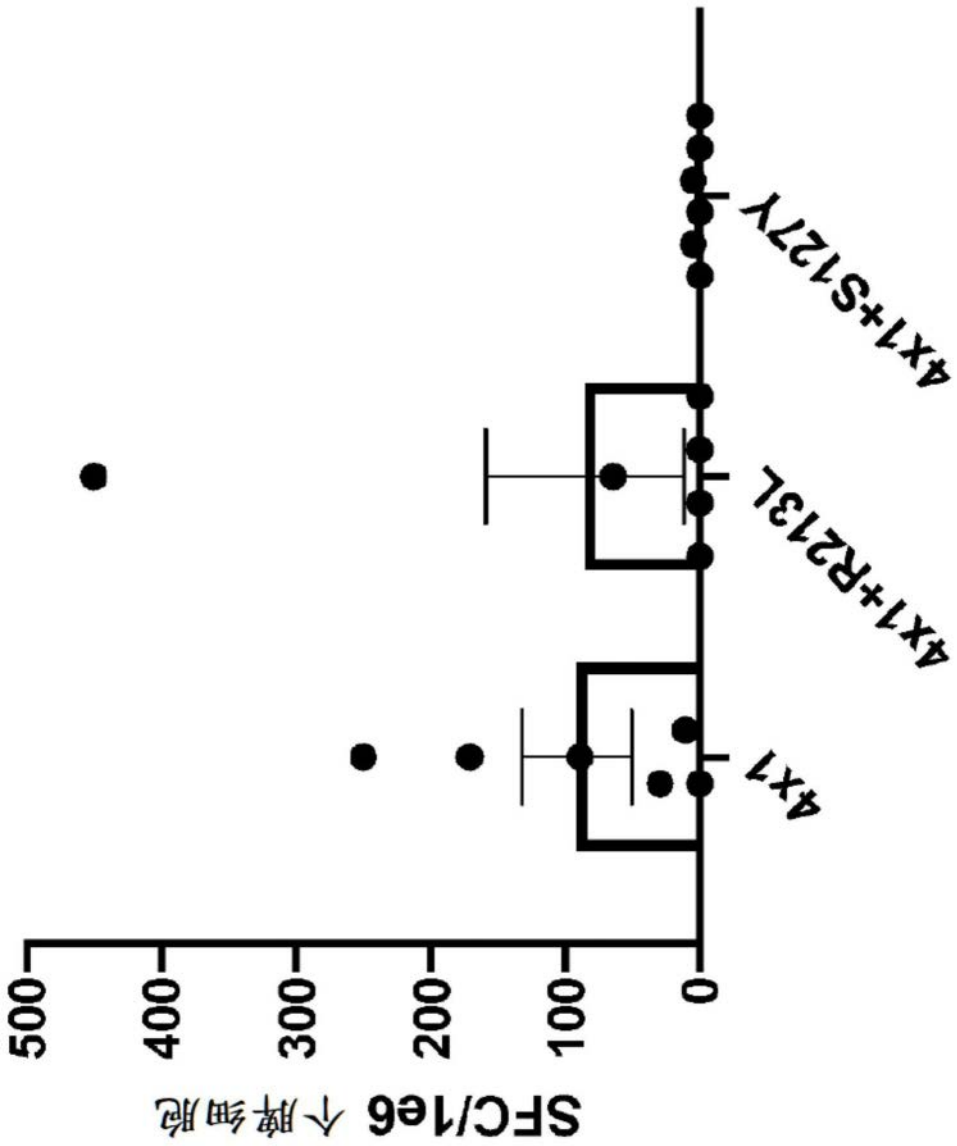


图2B

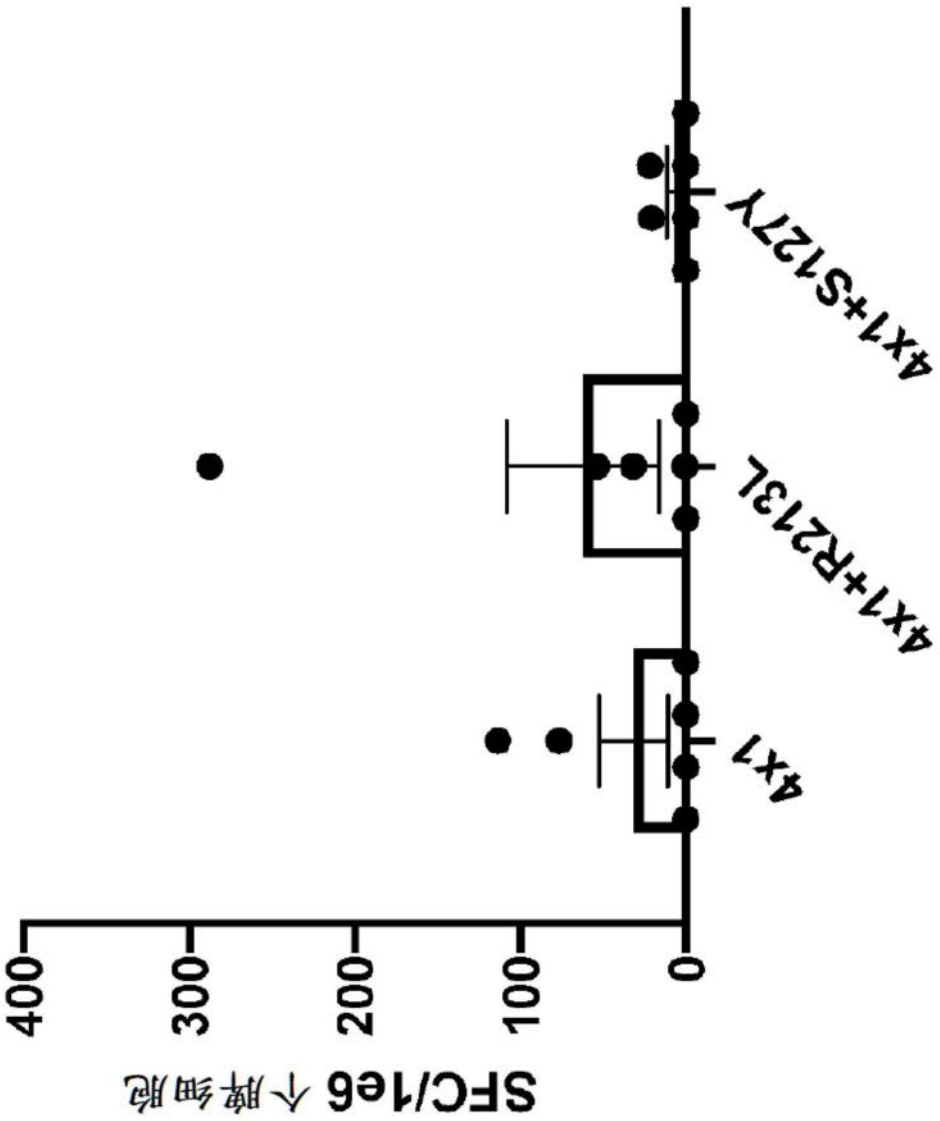


图2C

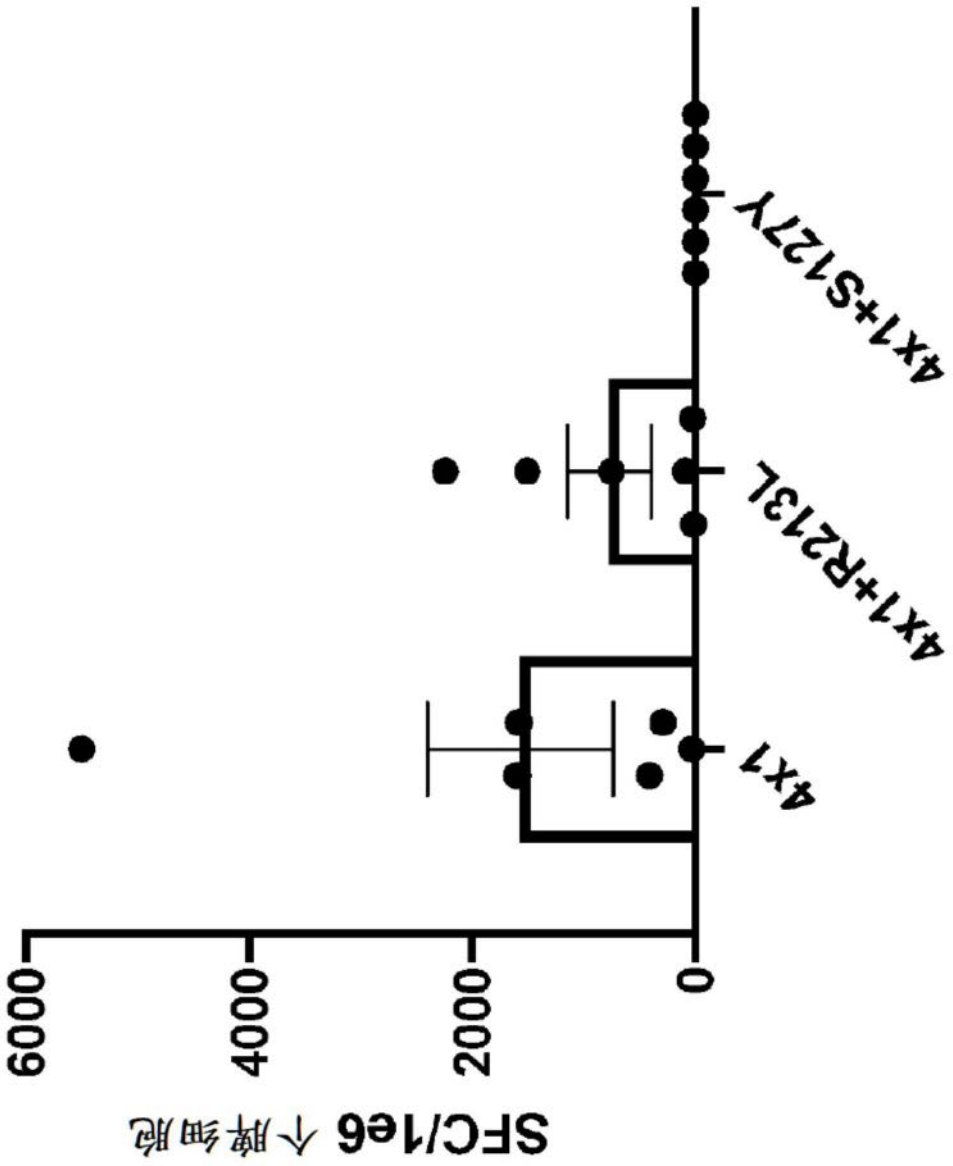


图2D

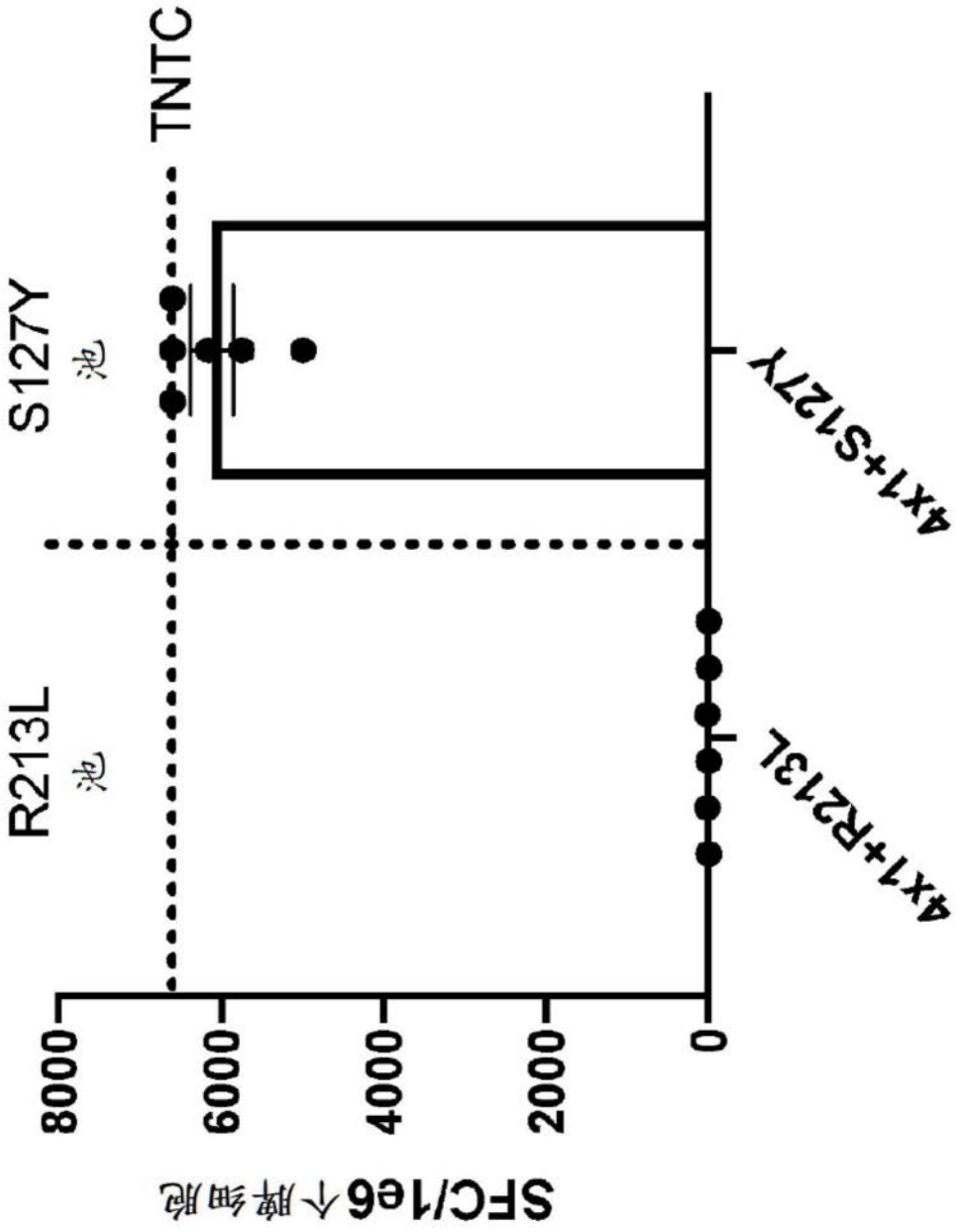

SFC/1e6个脾细胞
8000
6000
4000
2000
0
R213L
池
S127Y
池
TNTC
4X1+R213L
4X1+S127Y

图2E

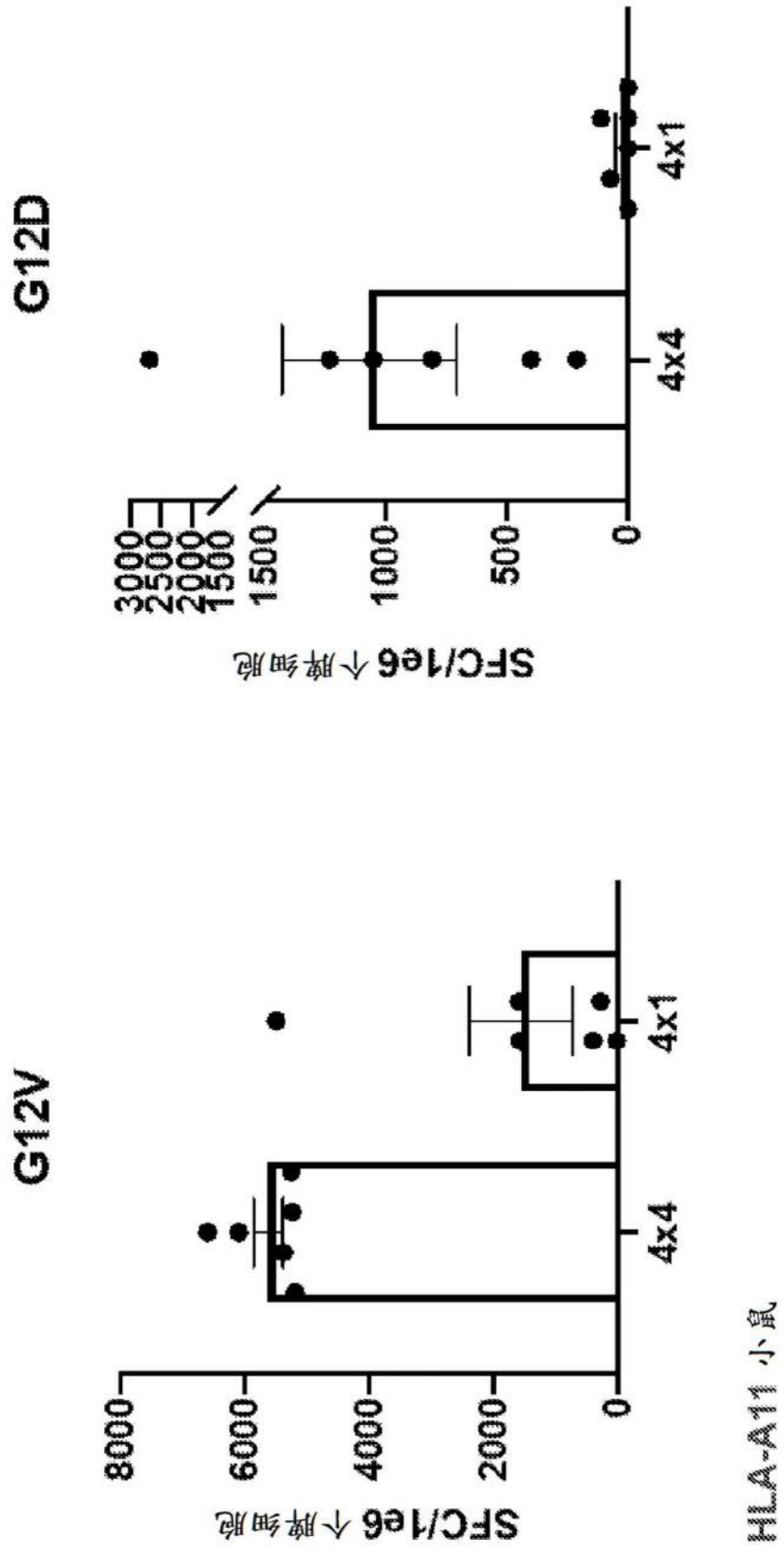

G12D
3000
2500
2000
1500
1000
500
0
SFC/1e6个脾细胞
4x4
4x1
G12V
8000
6000
4000
2000
0
SFC/1e6个脾细胞
4x4
4x1
HLA-A11小鼠

图3

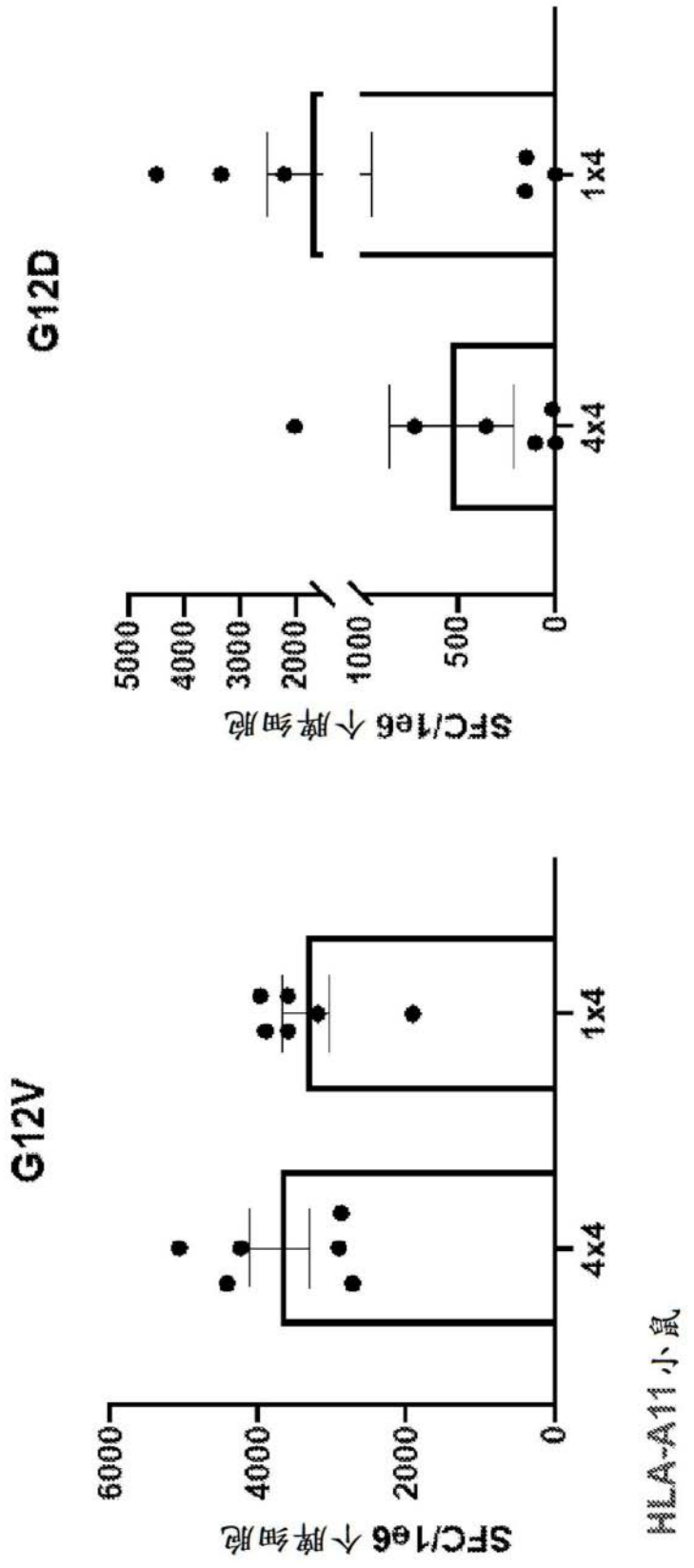

G12D
SFC/1e6 个脾细胞
5000
4000
3000
2000
1000
500
0
4x4
1x4
G12V
SFC/1e6 个脾细胞
6000
4000
2000
0
4x4
1x4
HLA-A11小鼠

图4

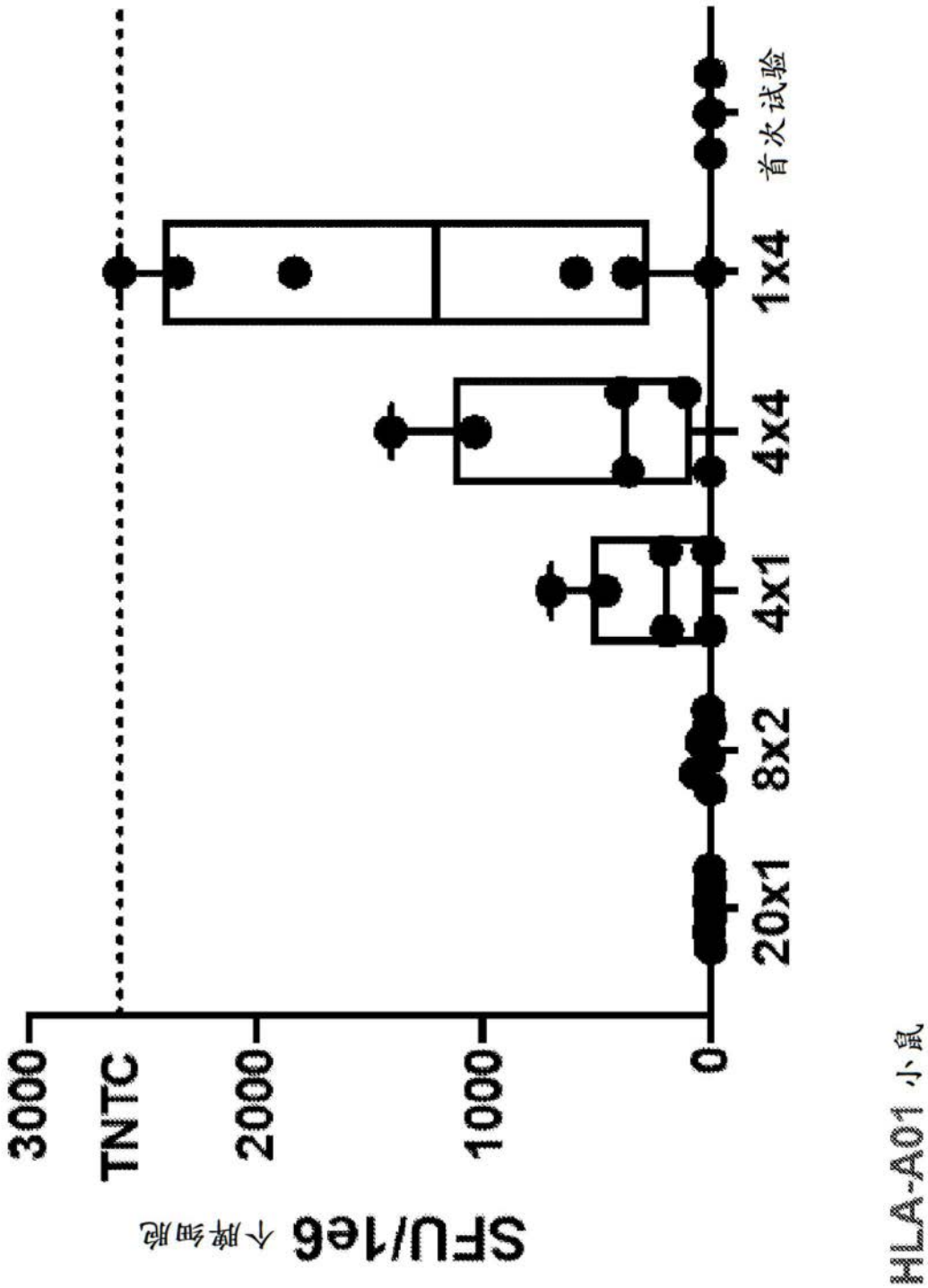

3000
TNTC
2000
1000
0
SFU/1e6 个脾细胞
20x1
8x2
4x1
4x4
1x4
首次试验
HLA-A01 小鼠

图5

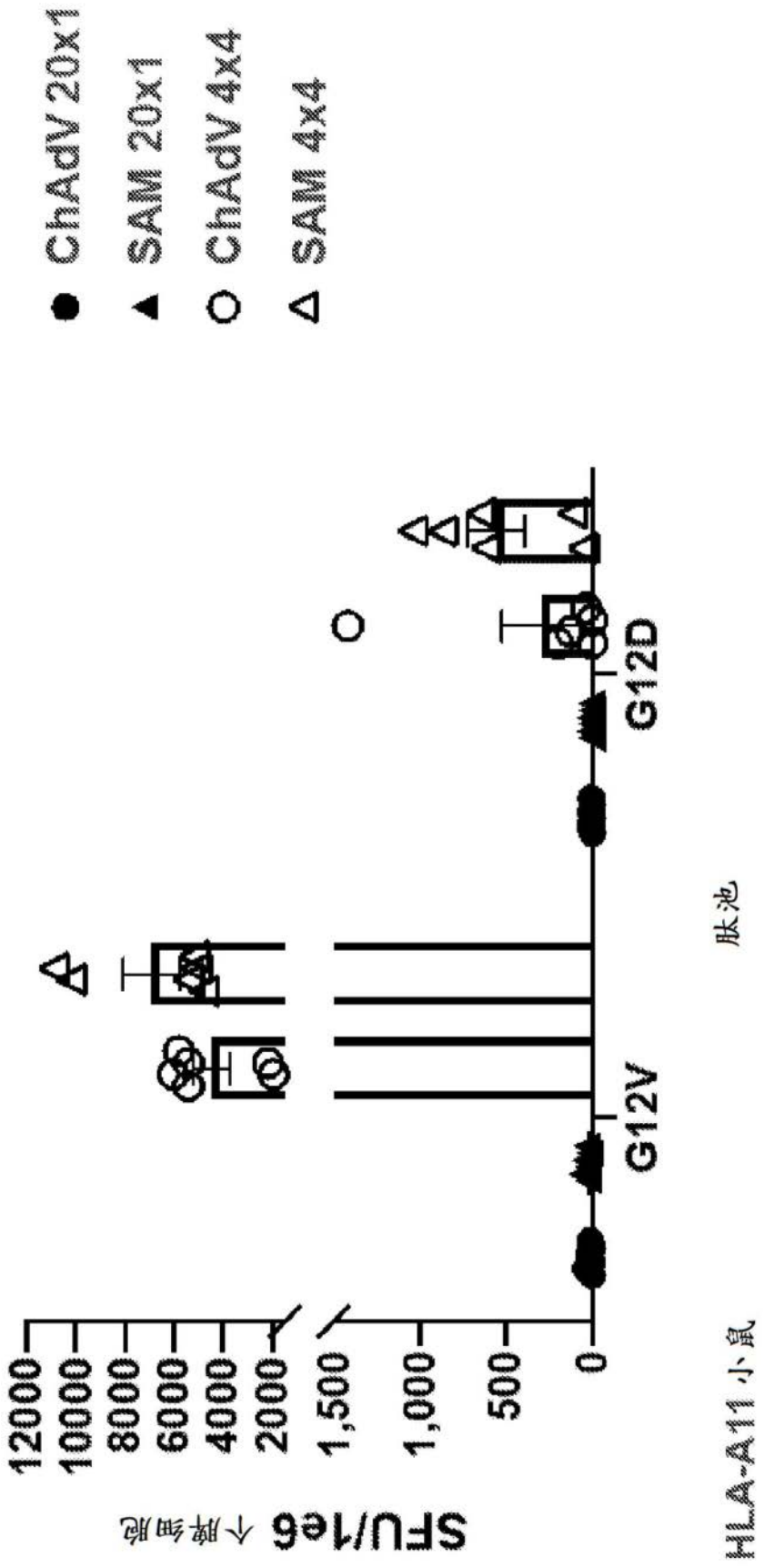

ChAdV 20x1
SAM 20x1
ChAdV 4x4
SAM 4x4
SFU/1e6 个脾细胞
12000
10000
8000
6000
4000
2000
1,500
1,000
500
0
G12V
G12D
肽池
HLA-A11 小鼠

图6

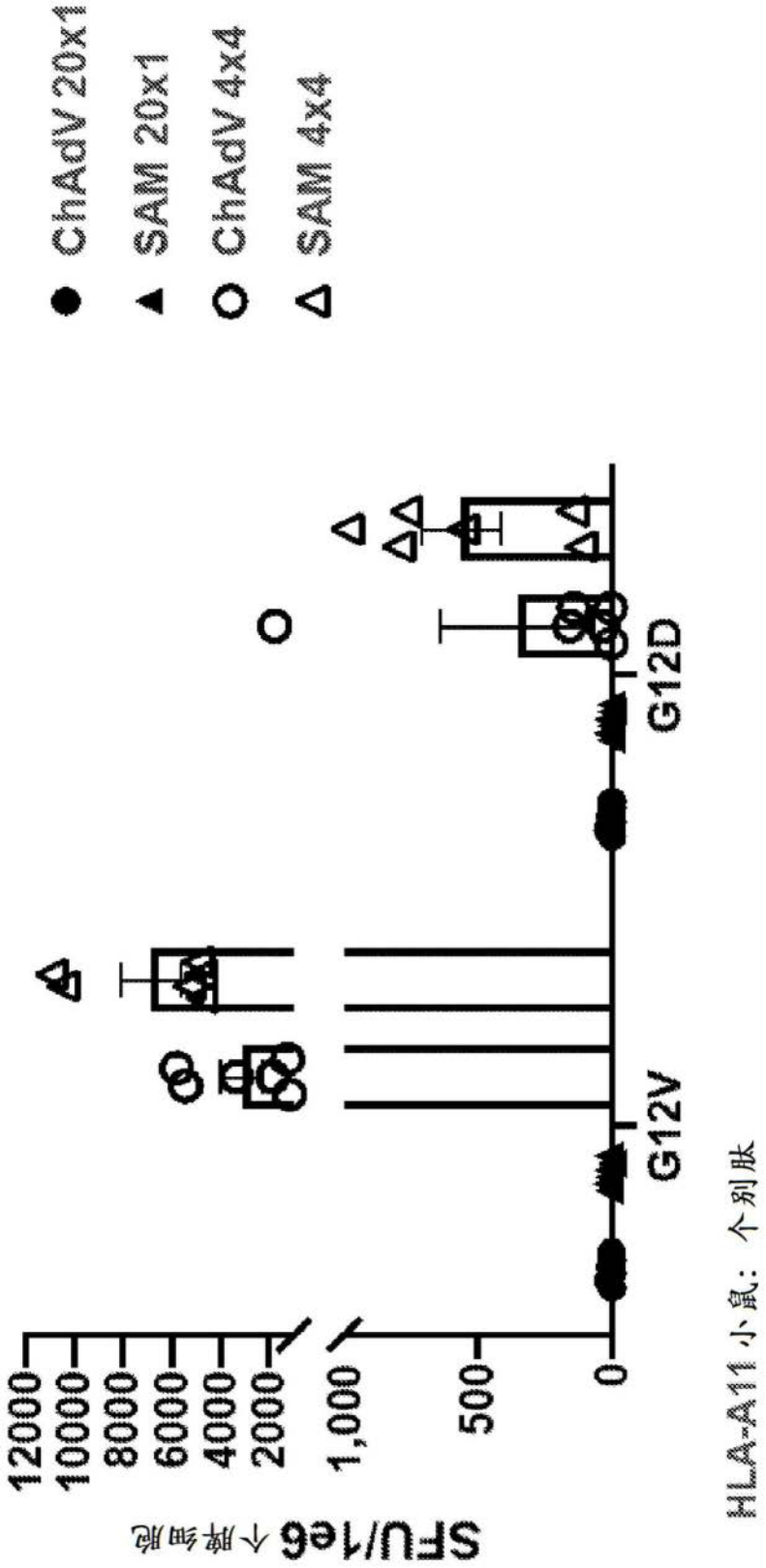

ChAdV 20x1
SAM 20x1
ChAdV 4x4
SAM 4x4
SFU/1e6个脾细胞
12000
10000
8000
6000
4000
2000
1,000
500
0
G12V
G12D
HLA-A11小鼠：个别肽

图7