



(12) 发明专利申请

(10) 申请公布号 CN 113453710 A

(43) 申请公布日 2021.09.28

(21) 申请号 201980078442.9

(22) 申请日 2019.09.27

(30) 优先权数据

1815956.6 2018.09.28 GB

(85) PCT国际申请进入国家阶段日

2021.05.27

(86) PCT国际申请的申请数据

PCT/GB2019/052747 2019.09.27

(87) PCT国际申请的公布数据

W02020/065349 EN 2020.04.02

(71) 申请人 剑桥大学的校长、教师和学者

地址 英国剑桥郡

申请人 威斯敏斯特大学 雷根斯堡大学

(72) 发明人 乔纳森·卢克·希尼

西蒙·弗罗斯特 拉尔夫·瓦格纳

贝内迪克特·阿斯巴赫

丽贝卡·金斯利 爱德华·赖特

(74) 专利代理机构 北京集佳知识产权代理有限公司 11227

代理人 张福誉 韩晓帆

(51) Int.Cl.

A61K 39/12 (2006.01)

C07K 16/10 (2006.01)

A61P 31/12 (2006.01)

权利要求书9页 说明书60页

序列表42页 附图8页

(54) 发明名称

疫苗和方法

(57) 摘要

描述了用于对能够诱导对病原体的广泛中和免疫应答和相关T细胞应答的经优化抗原性病原体多肽进行鉴定的方法,以及编码这样的多肽的核酸序列。还描述了用于确定在用经优化抗原性病原体多肽或编码经优化病原体多肽的核酸免疫接种之后在对象中是否诱导了广泛中和免疫应答的方法。还描述了核酸分子、多肽、载体、细胞、融合蛋白、药物组合物,及其作为针对病原体,特别是针对新出现或再出现的病原体(特别是RNA病毒)的疫苗的用途。

1. 对能够诱导对病原体的广泛中和免疫应答的先导候选经优化抗原性病原体多肽进行鉴定的方法,其包括:

i) 提供包含多个不同候选经优化抗原性病原体多肽的多肽文库,其中每个不同候选物的氨基酸序列已经从病原体多肽的多个不同氨基酸序列中优化并且不同于所述病原体多肽中的每个不同氨基酸序列,其中所述病原体多肽中的每个不同氨基酸序列包含不同分离株之多肽的氨基酸序列,并且其中每个不同分离株是这样的病原体的分离株,所述病原体与期望诱导对其广泛中和免疫应答的病原体为同一科;

ii) 对所述多肽文库中候选经优化抗原性病原体多肽被一种或更多种广泛中和抗原结合分子的结合进行筛选,所述一种或更多种广泛中和抗原结合分子中的每一种能够结合和/或中和这样的病原体,所述病原体与期望诱导对其广泛中和免疫应答的病原体为同一科;以及

iii) 将在步骤(ii)中被一种或更多种所述抗原结合分子结合的候选经优化抗原性病原体多肽鉴定为能够诱导对所述病原体的广泛中和免疫应答的先导候选经优化抗原性病原体多肽。

2. 根据权利要求1所述的方法,其中所述一种或更多种广泛中和抗原结合分子包括获自对象的抗体或来源于获自对象之抗体的抗体,所述对象已经暴露于这样的病原体,所述病原体与期望诱导对其广泛中和免疫应答的病原体为同一科。

3. 根据权利要求1或2所述的方法,其中所述一种或更多种广泛中和抗原结合分子包括非抗体抗原结合蛋白。

4. 根据权利要求3所述的方法,其中所述一种或更多种广泛中和抗原结合分子包括经设计锚蛋白重复蛋白(DARPin)、抗运载蛋白、适配体或T细胞受体分子。

5. 根据前述权利要求中任一项所述的方法,其中所述多肽文库中的候选经优化抗原性病原体多肽被表达在哺乳动物细胞内或表面上。

6. 根据权利要求1至4中任一项所述的方法,其中所述多肽文库中的候选经优化抗原性病原体多肽已经被表达在细菌、酵母或昆虫细胞内或表面上。

7. 根据前述权利要求中任一项所述的方法,其中所述病原体是病毒,所述候选经优化抗原性病原体多肽是候选经优化抗原性病毒多肽,并且所述病原体肽是病毒多肽。

8. 根据权利要求7所述的方法,其中所述多肽文库是包含多个不同病毒假型的病毒假型文库,每个不同病毒假型包含不同候选经优化病毒多肽。

9. 根据权利要求8所述的方法,其中在步骤(ii)中,通过对所述病毒假型被一种或更多种所述抗原结合分子之结合和/或中和的筛选来对所述候选经优化抗原性病毒多肽被一种或更多种所述抗原结合分子的结合进行筛选。

10. 根据权利要求1至7中任一项所述的方法,其中通过流式细胞术测定来对所述候选经优化抗原性病原体多肽被所述一种或更多种抗原结合分子的结合进行筛选。

11. 根据前述权利要求中任一项所述的方法,其还包括产生所述多肽文库。

12. 根据权利要求11所述的方法,其中通过从包含多个不同核酸的核酸文库中表达所述不同候选经优化抗原性病原体多肽来产生所述多肽文库,每个不同核酸包含编码所述多肽文库中的不同候选经优化抗原性病原体多肽的核苷酸序列。

13. 根据权利要求12所述的方法,其中所述不同候选经优化病原体多肽被表达在哺乳

动物细胞内或表面上。

14. 根据权利要求12或13所述的方法,其中所述核酸文库中的每个不同核酸的核苷酸序列是针对在哺乳动物细胞中表达所编码多肽而经密码子优化的,任选地经基因优化的。

15. 根据权利要求12至14中任一项所述的方法,其中所述核酸文库中的每个不同核酸是用于在哺乳动物细胞中表达所述核酸的表达载体的一部分。

16. 根据权利要求12至15中任一项所述的方法,其中所述病原体是病毒,所述候选经优化抗原性病原体多肽是候选经优化抗原性病毒多肽,并且所述病原体肽是病毒多肽。

17. 根据权利要求16所述的方法,其中所述核酸文库是病毒假型载体文库,并且所述文库中的每个不同核酸是用于产生包含所编码病毒多肽的病毒假型的表达载体的一部分,并且所述多肽文库是通过从所述病毒假型载体文库中的表达载体产生病毒假型而产生的病毒假型文库,其中所述病毒假型文库包含多个不同病毒假型,每个不同病毒假型包含由所述病毒假型载体文库中的不同核酸序列编码的不同候选经优化病毒多肽。

18. 根据权利要求15至17中任一项所述的方法,其中所述表达载体也是疫苗载体。

19. 根据权利要求18所述的方法,其中疫苗载体是病毒疫苗载体、细菌疫苗载体、RNA疫苗载体或DNA疫苗载体。

20. 根据权利要求18或19所述的方法,其中所述疫苗载体基于病毒递送载体,例如基于痘病毒(例如MVA、NYVAC、AVIPOX)、疱疹病毒(例如任何宿主物种的HSV、CMV、腺病毒)、麻疹病毒(例如麻疹)、甲病毒(例如SFV、仙台)、黄病毒(例如黄热病)或弹状病毒(例如VSV)的病毒递送载体,细菌递送载体(例如沙门菌(Salmonella)、大肠杆菌(E.coli)),RNA表达载体或DNA表达载体。

21. 根据权利要求15至20中任一项所述的方法,其中所述载体是基于pEVAC的表达载体。

22. 根据权利要求12所述的方法,其中所述不同候选经优化抗原性病原体多肽被表达在细菌、酵母或昆虫细胞内或表面上。

23. 根据权利要求12至22中任一项所述的方法,其还包括通过合成多个不同核酸来产生所述核酸文库,每个不同核酸包含编码不同候选经优化抗原性病原体多肽的不同核苷酸序列。

24. 根据权利要求23所述的方法,其还包括:

i) 获得所述不同病原体分离株的病原体多肽的氨基酸序列和/或编码所述病原体多肽的核苷酸序列;以及

ii) 产生多个不同核苷酸序列,每个不同核苷酸序列编码不同候选经优化抗原性病原体多肽,其中每个不同候选经优化抗原性病原体多肽的所编码氨基酸序列由所述病原体多肽的所获得的氨基酸序列或所编码的氨基酸序列进行了优化,并且不同于每个所获得的氨基酸序列或所编码的氨基酸序列。

25. 根据权利要求24所述的方法,其中在权利要求24的步骤(ii)中产生多个不同核苷酸序列包括:

对在权利要求24的步骤(i)中获得的氨基酸或核苷酸序列进行多序列比对;

从所述多序列比对中鉴定出在所述不同病原体分离株的多肽之间高度保守的氨基酸序列或所编码的氨基酸序列;以及

产生多个不同核苷酸序列,每个不同核苷酸序列编码不同候选经优化抗原性病原体多肽,其中所述不同核苷酸序列中的一个或多个包含编码从所述多序列比对中鉴定出的高度保守的氨基酸序列或所编码氨基酸序列的序列。

26. 根据权利要求25所述的方法,其还包括:

从所述多序列比对中鉴定出作为祖先氨基酸序列的氨基酸序列或所编码的氨基酸序列;以及

在不同的所产生的核苷酸序列中的一个或多个中包含编码从所述多序列比对中鉴定出的祖先氨基酸序列的序列。

27. 根据权利要求24至26中任一项所述的方法,其包括对不同的所产生核苷酸序列的密码子进行密码子优化,任选地基因优化,以在表达系统中最佳表达所编码的候选经优化抗原性病原体多肽。

28. 根据权利要求27所述的方法,其中所述表达系统包含哺乳动物细胞。

29. 根据权利要求27所述的方法,其中所述表达系统包含酵母、细菌或昆虫细胞。

30. 根据权利要求24至29中任一项所述的方法,其包括针对所编码候选经优化抗原性病原体多肽的抗原性对所述不同核苷酸序列进行优化。

31. 根据权利要求30所述的方法,其中抗原性优化包括以下中的任一项:

对编码抑制抗病原体多肽抗体的产生和/或功能的氨基酸序列的核酸序列进行缺失或修饰(例如,黏蛋白样结构域的缺失或修饰);

区域交换,以恢复一个或多个潜在丢失的编码表位;

位点特异性突变,例如N-连接的糖基化位点的位点特异性突变;

增强稳定性的改变(例如二硫键形成,减少丝氨酸蛋白酶对所编码多肽的降解);

聚糖的去除;

核酸序列的插入,例如插入编码期望表位的核酸序列。

32. 根据前述权利要求中任一项所述的方法,其中权利要求1的步骤(ii)中所述的一种或更多种广泛中和抗原结合分子包括广泛中和抗体,优选广泛中和单克隆抗体(BNmAb)。

33. 根据前述权利要求中任一项所述的方法,其中权利要求1的步骤(ii)中所述的一种或更多种抗原结合分子包括获自对象的抗体或来源于获自对象的抗体的抗体,所述对象在这样的病原体的爆发中幸存,所述病原体与期望诱导对其广泛中和免疫应答的病原体为同一科、任选地同一亚型或类型。

34. 根据权利要求33所述的方法,其中从中获得所述抗体或所述抗体所来源的对象是人或非人哺乳动物对象。

35. 根据权利要求33或34所述的方法,其中所述一种或更多种抗原结合分子包括广泛中和单克隆抗体(BNmAb)。

36. 根据前述权利要求中任一项所述的方法,其中所述不同病原体分离株包括来自这样的病原体之爆发的不同病原体分离株,所述病原体与期望诱导对其广泛中和免疫应答的病原体为同一亚型。

37. 根据前述权利要求中任一项所述的方法,其中所述不同病原体分离株包括来自这样的病原体之爆发的不同病原体分离株,所述病原体与期望诱导对其广泛中和免疫应答的病原体为不同亚型但同一类型。

38. 根据前述权利要求中任一项所述的方法,其中所述不同病原体分离株包括来自这样的病原体之爆发的不同病原体分离株,所述病原体与期望诱导对其广泛中和免疫应答的病原体为不同组但同一科。

39. 根据前述权利要求中任一项所述的方法,其中所述不同病原体分离株包括这样的病原体的不同先前病原体分离株,所述病原体与期望诱导对其广泛中和免疫应答的病原体为同一亚型、类型或科。

40. 根据前述权利要求中任一项所述的方法,其中每个候选经优化抗原性病原体多肽包含至少20个氨基酸残基。

41. 根据前述权利要求中任一项所述的方法,其中所述病原体是病毒。

42. 根据权利要求41所述的方法,其中所述病毒是RNA病毒。

43. 根据权利要求41或42所述的方法,其中所述病毒是新出现或再出现的RNA病毒。

44. 根据权利要求41至43中任一项所述的方法,其中所述病毒是丝状病毒、沙粒病毒或正黏病毒。

45. 根据权利要求41至43中任一项所述的方法,其中所述病毒是埃博拉病毒或马尔堡病毒。

46. 根据权利要求41至43中任一项所述的方法,其中所述病毒是拉沙病毒。

47. 根据前述权利要求中任一项所述的方法,其中所述病原体多肽是病毒糖蛋白。

48. 根据前述权利要求中任一项所述的方法,其是体外方法。

49. 对编码能诱导对病原体的广泛中和免疫应答的经优化抗原性病原体多肽的核酸序列进行鉴定的方法,其包括:

i) 用核酸对人或非人动物进行免疫接种,所述核酸包含编码通过根据前述权利要求中任一项所述的方法鉴定出的先导候选经优化抗原性病原体多肽的核酸序列;

ii) 确定在步骤(i)中的免疫接种之后在所述人或非人动物中是否诱导广泛中和免疫应答;以及

iii) 如果从步骤(ii)确定在所述人或非人动物中诱导了广泛中和免疫应答,则将所述核酸序列鉴定为编码能够诱导对病原体的广泛中和免疫应答的经优化抗原性病原体多肽的核酸序列。

50. 根据权利要求49所述的方法,其包括通过确定获自所述人或非人动物的血清中的抗体是否结合和/或中和超过一种病原体亚型来确定在所述人或非人动物中是否诱导了广泛中和免疫应答。

51. 根据权利要求49或50所述的方法,其中所述非人动物是哺乳动物。

52. 根据权利要求51所述的方法,其中所述哺乳动物是豚鼠或小鼠。

53. 根据权利要求49或50所述的方法,其中所述非人动物是禽类。

54. 分离的核酸分子,其包含核酸序列或其互补序列,所述核酸序列:

i) 与SEQ ID NO:1具有至少75%、80%、85%、90%、95%、96%、97%、98%或99%同一性,或与SEQ ID NO:1相同;

ii) 与SEQ ID NO:2具有至少75%、80%、85%、90%、95%、96%、97%、98%或99%同一性,或与SEQ ID NO:2相同;

iii) 与SEQ ID NO:4具有至少75%、80%、85%、90%、95%、96%、97%、98%或99%同

一性,或与SEQ ID NO:4相同;

iv) 与SEQ ID NO:5具有至少75%、80%、85%、90%、95%、96%、97%、98%或99%同一性,或与SEQ ID NO:5相同;

v) 与SEQ ID NO:7具有至少75%、80%、85%、90%、95%、96%、97%、98%或99%同一性,或与SEQ ID NO:7相同;或者

vi) 与SEQ ID NO:8具有至少75%、80%、85%、90%、95%、96%、97%、98%或99%同一性,或与SEQ ID NO:8相同。

55. 分离的核酸分子,其包含核酸序列或其互补序列,所述核酸序列:

i) 与SEQ ID NO:10具有至少75%、80%、85%、90%、95%、96%、97%、98%或99%同一性,或与SEQ ID NO:10相同;

ii) 与SEQ ID NO:12具有至少75%、80%、85%、90%、95%、96%、97%、98%或99%同一性,或与SEQ ID NO:12相同;或者

iii) 与SEQ ID NO:14具有至少75%、80%、85%、90%、95%、96%、97%、98%或99%同一性,或与SEQ ID NO:14相同。

56. 分离的核酸分子,其包含核酸序列或其互补序列,所述核酸序列:

i) 与SEQ ID NO:19具有至少75%、80%、85%、90%、95%、96%、97%、98%或99%同一性,或与SEQ ID NO:19相同;

ii) 与SEQ ID NO:21具有至少75%、80%、85%、90%、95%、96%、97%、98%或99%同一性,或与SEQ ID NO:21相同;

iii) 与SEQ ID NO:23具有至少75%、80%、85%、90%、95%、96%、97%、98%或99%同一性,或与SEQ ID NO:23相同;

iv) 与SEQ ID NO:25具有至少75%、80%、85%、90%、95%、96%、97%、98%或99%同一性,或与SEQ ID NO:25相同;

v) 与SEQ ID NO:27具有至少75%、80%、85%、90%、95%、96%、97%、98%或99%同一性,或与SEQ ID NO:27相同;

vi) 与SEQ ID NO:29具有至少75%、80%、85%、90%、95%、96%、97%、98%或99%同一性,或与SEQ ID NO:29相同;或者

vii) 与SEQ ID NO:31具有至少75%、80%、85%、90%、95%、96%、97%、98%或99%同一性,或与SEQ ID NO:31相同。

57. 分离的多肽,其包含氨基酸序列,所述氨基酸序列:

i) 与SEQ ID NO:1编码的氨基酸序列具有至少95%、96%、97%、98%或99%同一性,或与SEQ ID NO:1编码的氨基酸序列相同;

ii) 与SEQ ID NO:2编码的氨基酸序列具有至少95%、96%、97%、98%或99%同一性,或与SEQ ID NO:2编码的氨基酸序列相同;

iii) 与SEQ ID NO:4编码的氨基酸序列具有至少95%、96%、97%、98%或99%同一性,或与SEQ ID NO:4编码的氨基酸序列相同;

iv) 与SEQ ID NO:5编码的氨基酸序列具有至少95%、96%、97%、98%或99%同一性,或与SEQ ID NO:5编码的氨基酸序列相同;

v) 与SEQ ID NO:7编码的氨基酸序列具有至少95%、96%、97%、98%或99%同一性,或

与SEQ ID NO:7编码的氨基酸序列相同；

vi) 与SEQ ID NO:8编码的氨基酸序列具有至少95%、96%、97%、98%或99%同一性，或与SEQ ID NO:8编码的氨基酸序列相同；

vii) 与SEQ ID NO:10编码的氨基酸序列具有至少95%、96%、97%、98%或99%同一性，或与SEQ ID NO:10编码的氨基酸序列相同；

viii) 与SEQ ID NO:12编码的氨基酸序列具有至少95%、96%、97%、98%或99%同一性，或与SEQ ID NO:12编码的氨基酸序列相同；

ix) 与SEQ ID NO:14编码的氨基酸序列具有至少95%、96%、97%、98%或99%同一性，或与SEQ ID NO:14编码的氨基酸序列相同。

58. 分离的多肽，其包含氨基酸序列，所述氨基酸序列：

i) 与SEQ ID NO:3具有至少95%、96%、97%、98%或99%同一性，或与SEQ ID NO:3相同；

ii) 与SEQ ID NO:6具有至少95%、96%、97%、98%或99%同一性，或与SEQ ID NO:6相同；或者

iii) 与SEQ ID NO:9具有至少95%、96%、97%、98%或99%同一性，或与SEQ ID NO:9相同；

iv) 与SEQ ID NO:11具有至少95%、96%、97%、98%或99%同一性，或与SEQ ID NO:11相同；

v) 与SEQ ID NO:13具有至少95%、96%、97%、98%或99%同一性，或与SEQ ID NO:13相同；或者

vi) 与SEQ ID NO:15具有至少95%、96%、97%、98%或99%同一性，或与SEQ ID NO:15相同。

59. 分离的多肽，其包含氨基酸序列，所述氨基酸序列：

i) 与SEQ ID NO:18具有至少95%、96%、97%、98%或99%同一性，或与SEQ ID NO:18相同；

ii) 与SEQ ID NO:20具有至少95%、96%、97%、98%或99%同一性，或与SEQ ID NO:20相同；

iii) 与SEQ ID NO:22具有至少95%、96%、97%、98%或99%同一性，或与SEQ ID NO:22相同；

iv) 与SEQ ID NO:24具有至少95%、96%、97%、98%或99%同一性，或与SEQ ID NO:24相同；

v) 与SEQ ID NO:26具有至少95%、96%、97%、98%或99%同一性，或与SEQ ID NO:26相同；

vi) 与SEQ ID NO:28具有至少95%、96%、97%、98%或99%同一性，或与SEQ ID NO:28相同；或者

vii) 与SEQ ID NO:30具有至少95%、96%、97%、98%或99%同一性，或与SEQ ID NO:30相同。

60. 分离的核酸，其编码由权利要求54、55或56所述的核酸编码的氨基酸序列，其中所述核酸针对在哺乳动物细胞中的表达进行了密码子优化，任选地基因优化。

61. 分离的核酸,其编码权利要求57、58或59所述的多肽,其中所述核酸针对在哺乳动物细胞中的表达进行了密码子优化,任选地基因优化。

62. 载体,其包含权利要求54、55、56、60或61所述的核酸。

63. 根据权利要求62所述的载体,其还包含与所述核酸可操作连接的启动子。

64. 根据权利要求63所述的载体,其中所述启动子用于在哺乳动物细胞中表达由所述核酸编码的多肽。

65. 根据权利要求63所述的载体,其中所述启动子用于在酵母或昆虫细胞中表达由所述核酸编码的多肽。

66. 根据权利要求62至65中任一项所述的载体,其是疫苗载体。

67. 根据权利要求66所述的载体,其是病毒疫苗载体、细菌疫苗载体、RNA疫苗载体或DNA疫苗载体。

68. 分离的细胞,其包含权利要求62至65中任一项所述的载体。

69. 假型病毒颗粒,其包含权利要求57、58或59所述的多肽。

70. 产生权利要求69所述的假型病毒颗粒的方法,其包括用根据权利要求62至64中任一项所述的载体转染宿主细胞。

71. 融合蛋白,其包含根据权利要求57、58或59所述的多肽。

72. 药物组合物,其包含根据权利要求54、55、56、60或61所述的核酸,以及可药用载体、赋形剂或稀释剂。

73. 药物组合物,其包含根据权利要求62至64、66或67中任一项所述的载体,以及可药用载体、赋形剂或稀释剂。

74. 药物组合物,其包含根据权利要求57、58或59所述的多肽,以及可药用载体、赋形剂或稀释剂。

75. 根据权利要求72至74中任一项所述的药物组合物,其还包含佐剂,以用于增强对象中对所述组合物中的多肽或由所述核酸编码的多肽的免疫应答。

76. 在对象中诱导对丝状病毒科(Filoviridae)病毒的免疫应答的方法,其包括向所述对象施用根据权利要求54、55、60或61中任一项所述的核酸,根据权利要求57或78所述的多肽,根据权利要求62至64、66或67中任一项所述的载体,或根据权利要求72至75中任一项所述的药物组合物。

77. 对对象进行针对丝状病毒科病毒的免疫接种的方法,其包括向所述对象施用根据权利要求54、55、60或61中任一项所述的核酸,根据权利要求57或78所述的多肽,根据权利要求62至64、66或67中任一项所述的载体,或根据权利要求72至75中任一项所述的药物组合物。

78. 在对象中诱导对沙粒病毒科(Arenaviridae)病毒的免疫应答的方法,其包括向所述对象施用根据权利要求56、60或61中任一项所述的核酸,根据权利要求59所述的多肽,根据权利要求62至64、66或67中任一项所述的载体,或根据权利要求72至75中任一项所述的药物组合物。

79. 对对象进行针对沙粒病毒科病毒的免疫接种的方法,其包括向所述对象施用根据权利要求56、60或61中任一项所述的核酸,根据权利要求59所述的多肽,根据权利要求62至64、66或67中任一项所述的载体,或根据权利要求72至75中任一项所述的药物组合物。

80. 根据权利要求76至79中任一项所述的方法,其中所述组合物肌内施用。
81. 核酸表达载体,其包含多克隆位点,所述多克隆位点包含KpnI和NotI核酸内切酶位点。
82. 根据权利要求81所述的载体,其中所述多克隆位点包含SEQ ID NO:16的核酸序列。
83. 根据权利要求81或82所述的载体,其是表达载体和病毒假型载体。
84. 根据权利要求81至83中任一项所述的载体,其是疫苗载体。
85. 根据权利要求81至84中任一项所述的载体,其从5'至3'方向包含:启动子;剪接供体位点;剪接受体位点;和终止子信号,其中所述多克隆位点位于所述剪接受体位点和所述终止子信号之间。
86. 根据权利要求85所述的载体,其中所述启动子包含CMV即早期增强子/启动子和/或所述终止子信号包含缺少KpnI限制性核酸内切酶位点的牛生长激素基因的终止子信号。
87. 根据权利要求81至86中任一项所述的载体,其还包含复制起点和编码对抗生素的抗性的核酸。
88. 根据权利要求87所述的载体,其中所述复制起点包含pUC-质粒复制起点和/或所述核酸编码对卡那霉素的抗性。
89. 根据权利要求81至88中任一项所述的载体,其包含SEQ ID NO:17的核酸序列。
90. 分离的核酸分子,其包含编码包含SEQ ID NO:6氨基酸序列的多肽和包含SEQ ID NO:9氨基酸序列的多肽的核苷酸序列。
91. 分离的核酸分子,其包含编码包含SEQ ID NO:13氨基酸序列的多肽和包含SEQ ID NO:15氨基酸序列的多肽的核苷酸序列。
92. 组合物,其包含:第一核酸,所述第一核酸包含编码包含SEQ ID NO:6氨基酸序列的多肽的核苷酸序列;以及第二核酸,所述第二核酸包含编码包含SEQ ID NO:9氨基酸序列的多肽的核苷酸序列。
93. 组合物,其包含:第一核酸,所述第一核酸包含编码包含SEQ ID NO:13氨基酸序列的多肽的核苷酸序列;以及第二核酸,所述第二核酸包含编码包含SEQ ID NO:15氨基酸序列的多肽的核苷酸序列。
94. 组合制剂,其包含:(i)第一核酸,所述第一核酸包含编码包含SEQ ID NO:6氨基酸序列的多肽的核苷酸序列;以及(ii)第二核酸,所述第二核酸包含编码包含SEQ ID NO:9氨基酸序列的多肽的核苷酸序列。
95. 组合制剂,其包含:(i)第一核酸,所述第一核酸包含编码包含SEQ ID NO:13氨基酸序列的多肽的核苷酸序列;以及(ii)第二核酸,所述第二核酸包含编码包含SEQ ID NO:15氨基酸序列的多肽的核苷酸序列。
96. 组合物,其包含:包含SEQ ID NO:6的氨基酸序列的第一多肽和包含SEQ ID NO:9的氨基酸序列的第二多肽。
97. 组合物,其包含:包含SEQ ID NO:13的氨基酸序列的第一多肽和包含SEQ ID NO:15的氨基酸序列的第二多肽。
98. 融合蛋白,其包含:包含SEQ ID NO:6的氨基酸序列的第一多肽和包含SEQ ID NO:9的氨基酸序列的第二多肽。
99. 融合蛋白,其包含:包含SEQ ID NO:13的氨基酸序列的第一多肽和包含SEQ ID NO:

15的氨基酸序列的第二多肽。

100. 组合制剂,其包含:(i) 包含SEQ ID NO:6的氨基酸序列的第一多肽;以及(ii) 包含SEQ ID NO:9的氨基酸序列的第二多肽。

101. 组合制剂,其包含:(i) 包含SEQ ID NO:13的氨基酸序列的第一多肽;以及(ii) 包含SEQ ID NO:15的氨基酸序列的第二多肽。

102. 根据权利要求54、55、60或61中任一项所述的核酸,根据权利要求57或58所述的多肽,根据权利要求62至64、66或67中任一项所述的载体或根据权利要求72至75中任一项所述的药物组合物,其用作药物。

103. 根据权利要求54、55、60或61中任一项所述的核酸,根据权利要求57或58所述的多肽,根据权利要求62至64、66或67中任一项所述的载体或根据权利要求72至75中任一项所述的药物组合物,其用于治疗病毒感染,所述病毒感染优选地是由新出现或再出现的病毒、优选地丝状病毒科病毒引起的病毒感染。

104. 根据权利要求54、55、60或61中任一项所述的核酸,根据权利要求57或58所述的多肽,根据权利要求62至64、66或67中任一项所述的载体或根据权利要求72至75中任一项所述的药物组合物在制造药物中的用途,所述药物用于治疗病毒感染,所述病毒感染优选地是由新出现或再出现的病毒、优选地丝状病毒科病毒引起的病毒感染。

105. 根据权利要求56、60或61中任一项所述的核酸,根据权利要求59所述的多肽,根据权利要求62至64、66或67中任一项所述的载体或根据权利要求72至75中任一项所述的药物组合物,其用作药物。

106. 根据权利要求56、60或61中任一项所述的核酸,根据权利要求59所述的多肽,根据权利要求62至64、66或67中任一项所述的载体或根据权利要求72至75中任一项所述的药物组合物,其用于治疗病毒感染,所述病毒感染优选地是由新出现或再出现的病毒、优选地沙粒病毒科病毒引起的病毒感染。

107. 根据权利要求56、60或61中任一项所述的核酸,根据权利要求59所述的多肽,根据权利要求62至64、66或67中任一项所述的载体或根据权利要求72至75中任一项所述的药物组合物在制造药物中的用途,所述药物用于治疗病毒感染,所述病毒感染优选地是由新出现或再出现的病毒、优选地沙粒病毒科病毒引起的病毒感染。

108. 根据权利要求90或91所述的核酸,根据权利要求92、93、96或97所述的组合物,根据权利要求94、95、100或101所述的组合制剂,或者根据权利要求98或99所述的融合蛋白,其用作药物。

109. 根据权利要求90或91所述的核酸,根据权利要求92、93、96或97所述的组合物,根据权利要求94、95、100或101所述的组合制剂,或者根据权利要求98或99所述的融合蛋白,其用于治疗病毒感染,所述病毒感染优选地是由新出现或再出现的病毒、优选地丝状病毒科病毒引起的病毒感染。

110. 以下物质在制造用于治疗病毒感染的药物中的用途:根据权利要求90或91所述的核酸,根据权利要求92、93、96或97所述的组合物,根据权利要求94、95、100或101所述的组合制剂,或者根据权利要求98或99所述的融合蛋白,所述病毒感染优选地是由新出现或再出现的病毒、优选地丝状病毒科病毒引起的病毒感染。

疫苗和方法

[0001] 本发明涉及用于鉴定能够诱导对病原体的广泛中和免疫应答的经优化抗原性病原体多肽的方法,用于鉴定编码这样的经优化抗原性病原体多肽的核酸序列的方法,以及用于确定在用经优化抗原性病原体多肽或编码经优化病原体多肽的核酸免疫接种之后在对象中是否诱导了广泛中和免疫应答的方法。本发明还涉及核酸分子、多肽、载体、细胞、融合蛋白、药物组合物,及其作为针对病原体、特别是针对新出现(emerging)或再出现(re-emerging)的病原体(特别是RNA病毒)的疫苗的用途。本发明还涉及假型病毒颗粒。

[0002] 疫苗的基本原理是使免疫系统为与病原体的接触做好准备。疫苗触发免疫系统以产生抗体和T细胞应答,从而有助于抵抗感染。从历史上看,一旦病原体被分离出并使之生长,就可以大量生产并杀伤或减毒,然后用作疫苗。后来,来自分离的病原体的重组基因被用于产生重组蛋白,将其与佐剂混合以刺激免疫应答。最近,将病原体基因克隆到载体系统(减毒细菌或病毒)中,以在体内表达和递送抗原。所有这些策略都依赖于从过去爆发中分离出的病原体,以防止将来爆发。对于那些变化不显著或缓慢的病原体,这种常规技术是有效的。然而,一些病原体容易发生突变,并且抗体并不总是能识别出不同病原体株。新出现和再出现的病原体通常会将其易受攻击的抗原针对免疫系统隐藏或掩盖起来。

[0003] 在新出现和再出现的疾病中,不成比例的数量(37%)是由核糖核酸(RNA)病毒引起的(Heeney, *Journal of Internal Medicine* 2006;260:399-408)。RNA病毒是以RNA为遗传物质的病毒。该核酸通常是单链RNA(ssRNA),但也可能是双链RNA(dsRNA)。与DNA病毒相比, RNA病毒通常具有非常高的突变率,这是因为病毒RNA聚合酶缺乏DNA聚合酶的校对能力。这就是为何很难制备有效的疫苗来预防由RNA病毒引起的疾病的原因之一。在大多数情况下,当前针对RNA病毒的疫苗候选物受限于用作疫苗插入物的病毒株,其通常是基于野生型株的可用性而不是根据知情的设计来选择的。开发用于包膜RNA病毒的疫苗的技术挑战包括:i)野生型野外分离株糖蛋白(GP)的病毒变异作为疫苗抗原提供的保护范围有限;ii)由疫苗插入物表达的疫苗抗原的选择是高度经验性的;免疫原选择是缓慢的试错过程;iii)在不断发展或无法预料的病毒流行中,开发新的疫苗候选物是耗时的,并且可能会延迟疫苗的部署。

[0004] 由RNA病毒引起的重要人疾病包括病毒性出血热(VHF),这是由多个不同病毒科引起的一组疾病。通常来说,术语“病毒性出血热”用于描述严重的多系统综合征(即,身体中的多个器官系统受到影响)。特征性地,整个血管系统受到损害,并且身体的自我调节能力受到损害。这些症状通常伴有出血(流血),尽管出血本身很少危及生命。虽然一些类型的出血热病毒可引起相对较轻的疾病,但许多病毒会导致严重的威胁生命的疾病。VHF是由至少五个不同科的病毒引起的:沙粒病毒科(Arenaviridae)、布尼亚病毒科(Bunyaviridae)、丝状病毒科(Filoviridae)、黄病毒科(Flaviviridae)和副黏病毒科(Paramyxoviridae)。这些科的病毒都是RNA病毒,并且都被脂肪(脂质)外壳覆盖或包裹。VHF的存活取决于动物或昆虫宿主(天然库)。病毒在地理上受限于其宿主物种居住的区域,并且人在与被感染的宿主接触时会被感染。在从宿主传播后,带有一些病毒的人可能相互传播病毒。由这些病毒引起的出血热的人病例或爆发偶尔和不规律发生。爆发的发生不容易预测。除少数例外,尚无

针对VHF的治愈或建立的药物治疗。

[0005] 由沙粒病毒和丝状病毒引起的VHF一起覆盖了从西非到中非的广阔地理区域,并威胁到受感染的动物库可能迁移但是尚未报道人疾病的邻近区域。丝状病毒以单链负义RNA的形式编码其基因组。该科中的两个众所周知的成员是埃博拉病毒和马尔堡病毒。埃博拉病毒是新出现和再出现的RNA病毒性疾病。爆发并不总是由完全相同的病毒引起,而是由同一病毒科的不同亲属型(relative)(类型)造成的,其中存在近亲型(close sibling)(例如埃博拉Mayinga和埃博拉Kikwit)、近堂亲型(close cousin)(塔伊森林(Tai Forest)和本迪布焦(Bundibugyo))、远堂亲型(苏丹(Sudan))和远亲型(distant relative)(马尔堡病毒)。自病毒性疾病首次被发现以来,2014年西非的埃博拉爆发是规模最大的。沙粒病毒分为两类:旧世界病毒和新世界病毒。这些群体之间的差异在地理和遗传上得到了区分。已知有至少八种沙粒病毒引起程度不一的人疾病。无菌性脑膜炎(aseptic meningitis)是一种引起覆盖脑和脊髓的炎症的严重人疾病,可由淋巴细胞性脉络丛脑膜炎病毒(LCMV)感染引起。出血热综合征源于例如以下病毒的感染:瓜纳瑞托病毒(Guanarito virus,GT0V)、胡宁病毒(Junin virus,JUNV)、拉沙病毒(Lassa virus,LASV)、卢约病毒(Lujo virus,LUJV)、马丘波病毒(Machupo virus,MACV)、萨比亚病毒(Sabia virus,SABV)或白水河病毒(Whitewater Arroyo virus,WWAV)。

[0006] 拉沙病毒(LASV)、埃博拉病毒(EBOV)和马尔堡病毒(MARV)是西非和中非最重要的出血热。拉沙热是西非地区的流行病,估计每年感染数为30万至100万,有5,000例死亡。拉沙热病毒(LASV)、埃博拉病毒(EBOV)和马尔堡病毒(MARV)都是具有人的高发病率和高死亡率的4级病原体,对于其尚无确定的治愈,目前尚无许可的疫苗用于这些病毒引起的感染。

[0007] 流感病毒是正黏病毒科(Orthomyxoviridae)的成员。共有三种类型的流感病毒,分别称为甲型流感、乙型流感和丙型流感。甲型流感病毒感染广泛多种鸟类和哺乳动物,包括人、马、海洋哺乳动物、猪、雪貂和鸡。在动物中,大多数甲型流感病毒会引起呼吸道和肠道的轻度局部感染。然而,高致病性的甲型流感病毒株(例如H5N1)会引起禽类的全身性感染,其中死亡率可达到100%。2009年,H1N1流感是人流感的最常见原因。2009年出现了新的猪源H1N1毒株,并且其被世界卫生组织宣布为大流行(pandemic)。该毒株被称为“猪流感”。H1N1甲型流感病毒也是1918年的西班牙流感大流行、1976年的迪克斯堡(Fort Dix)爆发以及1977至1978年的俄罗斯流感大流行的原因。目前,美国有两种许可的流感疫苗方法-灭活裂解疫苗和减毒活疫苗。灭活疫苗可有效地诱导体液免疫应答,但通常只能诱导较差的细胞免疫应答。由于其升高的感染风险,因此不能向免疫力低下或怀孕的患者施用活病毒疫苗。

[0008] 因此,需要提供有效的疫苗,其诱导广泛中和免疫应答,以针对新出现和再出现的疾病,尤其是由病毒例如RNA病毒引起的疾病(包括VHF和流感)提供保护。

[0009] 根据本发明,提供了用于鉴定能够诱导对病原体的广泛中和免疫应答的先导候选经优化抗原性病原体多肽的方法,其包括:

[0010] i) 提供包含多个不同候选经优化抗原性病原体多肽的多肽文库,其中每个不同候选物的氨基酸序列已经从病原体多肽的多个不同氨基酸序列中优化并且不同于所述病原体多肽的每个不同氨基酸序列,其中所述病原体多肽的每个不同氨基酸序列包含不同分离株的多肽的氨基酸序列,并且其中每个不同分离株是这样的病原体的分离株,所述病原体

与期望诱导对其广泛中和免疫应答的病原体为同一科；

[0011] ii) 对所述多肽文库中的候选经优化抗原性病原体多肽被一种或更多种广泛中和抗原结合分子的结合进行筛选,所述一种或更多种广泛中和抗原结合分子中的每一种能够结合和/或中和这样的病原体,所述病原体与期望诱导对其广泛中和免疫应答的病原体同一科;以及

[0012] iii) 将在步骤(ii)中被一种或更多种抗原结合分子结合的候选经优化抗原性病原体多肽鉴定为能够诱导对所述病原体的广泛中和免疫应答的先导候选经优化抗原性病原体多肽。

[0013] 任选地,病原体的每种不同分离株或多种不同分离株中的每一种与期望诱导对其广泛中和免疫应答的病原体是同一亚型或类型。

[0014] 任选地,病原体的每种不同分离株或多种不同分离株中的每一种与期望诱导对其广泛中和免疫应答的病原体是同一种或属。

[0015] 任选地,不同分离株包括与期望诱导对其广泛中和免疫应答的病原体同一科内的不同亚型或类型的分离株。

[0016] 任选地,不同分离株包括与期望诱导对其广泛中和免疫应答的病原体同一科内的不同种或属的分离株。

[0017] 术语“病原体”在本文中用于指代可引起疾病的任何事物,尤其是指可引起疾病的传染原,例如病毒、细菌、真菌或寄生虫。

[0018] 术语“多肽”在本文中用于指代包含多个氨基酸残基的聚合物,所述氨基酸残基通过肽键连接在一起以形成链。所有蛋白质均为多肽。术语“多肽”与术语“蛋白质”可互换使用。术语“多肽”特别地旨在包含天然存在的蛋白质以及重组或合成产生的蛋白质。任选地,多肽是经修饰的多肽,例如共翻译或翻译后修饰的多肽,例如糖基化多肽或糖基化蛋白质(“糖蛋白”)。糖蛋白是包含共价连接至氨基酸侧链的寡糖链(聚糖)的蛋白质。碳水化合物通过共翻译或翻译后糖基化而连接至蛋白质。

[0019] “病原体多肽”是指形成病原体的一部分的任何多肽。任选地,病原体多肽是病原体的结构蛋白(或其一部分)。任选地,病原体多肽是暴露在病原体表面上的结构蛋白(或其一部分)。任选地,病原体多肽是病毒蛋白(或其一部分)。任选地,病原体多肽是病毒包膜蛋白(或其一部分)。任选地,病原体多肽是糖蛋白(或其一部分)。任选地,病原体多肽是病毒糖蛋白(或其一部分)。任选地,病原体多肽是病毒包膜糖蛋白(或其一部分)。任选地,病原体多肽是外部病毒包膜糖蛋白(或其一部分)。任选地,病原体多肽包含至少20个氨基酸残基的氨基酸序列。任选地,病原体多肽包含多至1000、900、800、700或600个氨基酸残基的氨基酸序列。

[0020] 完全组装的感染性病毒被称为病毒体。最简单的病毒体由核酸(单链或双链RNA或DNA)和衣壳蛋白外壳组成。衣壳形成为单或双蛋白壳,并且由仅一种或几种结构蛋白组成。包膜病毒具有覆盖其保护性蛋白衣壳的包膜。包膜通常来源于宿主细胞膜的一些部分(磷脂和蛋白质),但包括病毒编码的糖蛋白。

[0021] 包膜表面的糖蛋白用于鉴定宿主膜上的受体位点并与之结合。病毒包膜然后与宿主膜融合,从而使衣壳和病毒基因组进入并感染宿主。病毒-细胞膜融合是所有包膜病毒(包括人病原体,例如丝状病毒、流感病毒和人免疫缺陷病毒(HIV))进入细胞并开始病毒感

染的一种手段。该膜融合过程由一种或更多种病毒包膜糖蛋白执行。融合可发生在细胞质膜或内体膜上。

[0022] 糖蛋白可帮助病毒避开宿主免疫系统。包膜病毒具有很大的适应性,并且可在短时间内改变以逃避宿主免疫系统。包膜病毒可导致持续感染。包膜RNA病毒包括例如黄病毒、披膜病毒(Togavirus)、冠状病毒、丁型肝炎、正黏病毒、副黏病毒、弹状病毒、布尼亚病毒、丝状病毒。逆转录病毒是包膜病毒。包膜DNA病毒包括疱疹病毒、痘病毒、肝炎病毒。

[0023] 大多数外部病毒包膜蛋白是糖蛋白,以膜锚定的刺突(spike)出现,通常组装成二聚体或三聚体。丝状病毒包膜上的三聚体糖蛋白(GP)刺突介导病毒进入的所有阶段,包括附着、进入和融合。细胞受体的识别位点通常位于距病毒包膜最远的结构域(远端),而近端结构域则与包膜的脂质双层相互作用。寡糖侧链(聚糖)通过N-糖苷键联或更罕见的O-糖苷键联接。由于这些是通过细胞糖基转移酶合成的,因此这些聚糖的糖组成类似于宿主细胞膜糖蛋白的糖组成。

[0024] 丝状病毒进入细胞表面已证明是由宿主细胞附着因子介导的,所述宿主细胞附着因子例如C型凝集素,包括DC-SIGN(树突状细胞特异性ICAM3捕获非整联蛋白;也称为CD209)和L-SIGN(肝和淋巴结SIGN;也称为CLEC4M)和多种细胞表面蛋白质,例如整联蛋白、T细胞免疫球蛋白和含黏蛋白结构域(TIM)的蛋白质,以及酪氨酸蛋白激酶受体3(TYR03)家族成员。与细胞表面结合后,丝状病毒通过巨胞饮过程被内化,随后通过早期和晚期内体运输。在病毒包膜与晚期内体的膜融合之后,病毒基因组然后渗透到胞质中。在胞质中,病毒基因组复制并转录,并且合成新的病毒蛋白以组装后代病毒体,后者从细胞表面出芽(bud)。

[0025] 埃博拉病毒(EBOV)的表面糖蛋白GP是许多疫苗的关键组分和中和抗体的靶标。EBOV GP是作为单个多肽合成的,随后被弗林蛋白酶样蛋白酶切割为GP1和GP2亚基,它们通过亚基间二硫键和非共价相互作用保持在一起,并在病毒表面形成GP1-GP2异二聚体的三聚体。然而,弗林蛋白酶的切割不足以引发EBOV GP。进入细胞后,病毒最终被运输到晚期内体,在那里GP被进一步引发以去除一些“帽”组分,从而触发关键膜融合事件的诱导,导致病毒渗透。EBOV GP引发由在 β 13- β 14环内切割GP1的半胱氨酸蛋白酶组织蛋白酶B和组织蛋白酶L介导。组织蛋白酶切割从GP1中去除了约60%的氨基酸,包括黏蛋白样结构域、聚糖帽和提议的受体结合区的最外 β 链,形成了GP的引发形式(称为GPc1,19kDa GP1加GP2)。与全长GP不同,引发的GPc1无法与内体膜蛋白Niemann-Pick C1(NPC1)结合,其是EBOV感染必不可少的宿主进入因子。游离NPC1-C及其与GPc1的复合物的晶体结构已经确定(Wang et al., Cell, 2016, 164, 258-268)。在埃博拉病毒感染期间,GP基因的主要产物是分泌型GP(sGP),这是缺少GP2和黏蛋白样结构域,但共有295个GP1氨基酸的可溶性二聚体。

[0026] 流感病毒体包含分段的负义RNA基因组,其编码以下蛋白质:血凝素(HA)、神经氨酸酶(NA)、基质(M1)、质子离子通道蛋白(M2)、核蛋白(NP)、聚合酶碱性蛋白1(PB1)、聚合酶碱性蛋白2(PB2)、聚合酶酸性蛋白(PA)和非结构蛋白2(NS2)。HA、NA、MI和M2是膜相关的,而NP、PB1、PB2、PA和NS2是核衣壳相关的蛋白质。MI蛋白是流感颗粒中最丰富的蛋白。HA和NA蛋白是包膜糖蛋白,负责病毒的附着和病毒颗粒向细胞内的渗透,以及病毒中和和保护性免疫的主要免疫显性表位的来源。HA和NA蛋白二者被认为是预防性流感疫苗的最重要组分。

[0027] 对于细菌或真菌,合适的病原体多肽包括对于细菌或真菌的繁殖或对于细菌或真菌在人中感染或引起疾病的能力必不可少的多肽。合适的实例包括表面表达的多肽或蛋白质(参见,例如Hu et al.,*Front.Microbiol.*8:82.doi:10.3389/fmicb.2017.00082; Santos and Levitz,*Cold Spring Harb Perspect Med.*2014;4(11):a019711)。

[0028] 本文使用的术语“抗原”是指能够在宿主生物体中诱导免疫应答的物质。免疫应答可以是体液和/或细胞免疫应答。细胞免疫应答是免疫系统细胞(例如B细胞、T细胞、巨噬细胞或多形核细胞)对刺激物(例如抗原或疫苗)的应答。免疫应答可包括身体参与宿主防御应答的任何细胞,包括例如分泌干扰素或细胞因子的上皮细胞。免疫应答包括但不限于先天免疫应答或炎症。如本文所用,保护性免疫应答是指保护对象免于感染或疾病(即,即预防感染或预防与感染相关的疾病的发展)的免疫应答。测量免疫应答的方法是本领域公知的,并且包括例如测量淋巴细胞(例如B或T细胞)的增殖和/或活性、细胞因子或趋化因子的分泌、炎症或抗体产生。

[0029] 任选地,经优化抗原性病原体多肽能够在已施用多肽(作为多肽,或例如从施用的核酸表达载体表达)的人或非人动物中诱导抗体的产生和/或T细胞应答。

[0030] 本文使用的术语“抗体”是指由具有特定氨基酸序列的B淋巴样细胞产生的免疫球蛋白分子。抗体在人或其他动物中由特定抗原(免疫原)诱发。抗体的特征在于以一些可证实的方式与抗原特异性反应,抗体和抗原各自以彼此定义。“引起抗体应答”是指抗原或其他分子诱导抗体产生的能力。

[0031] “中和”抗体或抗原结合分子不仅与病原体(例如病毒)结合,而且其以抑制(即降低)或阻断感染或感染进展的方式结合。中和抗体或抗原结合分子可阻断与受体的相互作用,或者可以以抑制基因组脱壳的方式与病毒衣壳结合。术语“中和抗体”或“中和抗原结合分子”还包括能够通过促进细胞因子应答或通过促进被免疫细胞摄取和去除来预防病原体(例如病毒)的感染的抗体或抗原结合分子。特别地,术语“中和抗体”包括能够通过抗体依赖性细胞介导的细胞毒性(ADCC)或补体依赖性细胞毒性(CDC)抑制或阻断病原体的感染(或感染的进展)的抗体(或者其片段或衍生物)。结合病毒的许多抗体中只有一小部分能够中和。

[0032] 本文使用的术语“广泛中和抗原结合分子”包括能够抑制(即降低)、中和或预防病原体的至少两种不同亚型或种(例如病毒的至少两种不同亚型或种、细菌的至少两种不同亚型或种或者真菌的至少两种不同亚型或种)的感染的抗原结合分子,例如抗体或者其片段或衍生物。任选地,广泛中和抗原结合分子能够抑制(即降低)、中和或预防病原体的大部分或全部不同亚型或种(例如病毒的大部分或全部不同亚型或种、细菌的大部分或全部不同亚型或种或者真菌的大部分或全部不同亚型或种)的感染。任选地,广泛中和抗体能够抑制(即降低)、中和或预防同一科内病原体(例如病毒、细菌或真菌)的至少两种不同类型的成员的感染。

[0033] 任选地,在本发明方法的步骤(ii)中使用多个不同广泛中和抗原结合分子。任选地,每个不同广泛中和抗原结合分子与多肽文库中的候选经优化抗原性病原体多肽的不同区域或表位结合。

[0034] 本文使用的术语“广泛中和免疫应答”意指在对象中引起的免疫应答,该免疫应答足以抑制(即降低)、中和或预防病原体的至少两种不同亚型或种(例如病毒的至少两种不

同亚型或种、细菌的至少两种不同亚型或种或者真菌的至少两种不同亚型或种)的感染和/或感染的进展。任选地,广泛中和免疫应答足以抑制、中和或预防病原体的大部分或全部不同亚型或种(例如病毒的大部分或全部不同亚型或种、细菌的大部分或全部不同亚型或种或者真菌的大部分或全部不同亚型或种)的感染和/或感染的进展。任选地,广泛中和免疫应答足以抑制、中和或预防同一科内病原体(例如病毒、细菌或真菌)的至少两种不同类型的成员的感染和/或感染的进展。任选地,广泛中和免疫应答足以抑制、中和或预防同一科内病原体(例如病毒、细菌或真菌)的至少两个不同属的成员的感染和/或感染的进展。

[0035] 数种针对病原体的广泛中和抗体是已知的。例如,已证明一些抗体能够中和整个丝状病毒科中多种亚型的病毒分离株。Saphire et al. (Cell,2018;174(4):938-952)描述了针对埃博拉病毒糖蛋白的单克隆抗体的系统分析。针对埃博拉病毒的广泛中和抗体的一个实例是免疫引起的猕猴抗体CA45,如Zhao et al.,2017 (Cell 169,891-904)所述。Bruun et al. (PLoS ONE 9(10):e109196.doi:10.1371/journal.pone.0109196)中参考了针对HIV-1包膜蛋白的广泛中和单克隆抗体。Corti et al. (Curr Opin Virol.2017 Jun;24:60-69)提供了针对甲型和乙型流感病毒的广泛反应性单克隆抗体的特异性、抗病毒和免疫学作用机理以及在临床上的开发的概述。

[0036] 任选地,病原体是病毒。

[0037] 病毒主要根据表型特征(例如形态学、核酸类型、复制方式、宿主生物体及其引起的疾病类型)进行分类。一种用于病毒分类的方案是巴尔的摩分类系统(Baltimore classification system),其根据病毒的核酸(DNA或RNA)、链式(单链或双链)、有义链和复制方法的组合将病毒分为七个组之一:

[0038] • I:dsDNA病毒(例如腺病毒、疱疹病毒、痘病毒);

[0039] • II:ssDNA病毒(+链或“有义链”)DNA(例如细小病毒);

[0040] • III:dsRNA病毒(例如呼肠孤病毒);

[0041] • IV:(+) ssRNA病毒(+链或有义链)RNA(例如小RNA病毒、披膜病毒);

[0042] • V:(-) ssRNA病毒(-链或反义链)RNA(例如正黏病毒、丝状病毒、沙粒病毒、弹状病毒);

[0043] • VI:ssRNA-RT病毒(+链或有义链)RNA,在生命周期中具有DNA中间体(例如逆转录病毒));

[0044] • VII:dsDNA-RT病毒DNA,在生命周期中具有RNA中间体(例如嗜肝DNA病毒(Hepadnavirus))。

[0045] 任选地,病毒是RNA病毒。RNA病毒包括:

[0046] 组III:病毒具有双链RNA基因组。

[0047] 组IV:病毒具有正义单链RNA基因组。许多公知的病毒存在于该组中,包括小RNA病毒(其为包括公知的病毒例如甲型肝炎病毒、肠病毒、鼻病毒、脊髓灰质炎病毒和口蹄疫病毒的病毒科)、SARS病毒、丙型肝炎病毒、黄热病毒和风疹病毒;

[0048] 组V:病毒具有负义单链RNA基因组。埃博拉病毒和马尔堡病毒以及流感病毒、拉沙病毒、麻疹、流行性腮腺炎和狂犬病都是该组的公知成员。

[0049] 下表列出了巴尔的摩分类下不同RNA病毒科的分组:

[0050]

RNA 病毒科	实例 (俗名)	衣壳	衣壳对称性	核酸类型	组
呼肠孤病毒科	呼肠孤病毒、轮状病毒	裸露/包膜	二十面体	ds	III
小 RNA 病毒科	肠病毒、鼻病毒、肝病毒、 心脏病毒、口蹄疫病毒、 脊髓灰质炎病毒、副肠孤 病毒、马鼻病毒、嗜病毒、 捷申病毒、柯萨奇病毒	裸露	二十面体	ss	IV
杯状病毒科 (<i>Caliciviridae</i>)	诺沃克病毒	裸露	二十面体	ss	IV
披膜病毒科	风疹病毒、甲病毒	裸露	二十面体	ss	IV
沙粒病毒科	淋巴细胞脉络丛脑膜炎	包膜	复合体	ss(-)	V

	病毒、拉沙病毒					
	黄病毒科	登革热病毒、丙型肝炎病毒、黄热病病毒、寨卡病毒	包膜	二十面体	ss	IV
	正黏病毒科	甲型流感病毒、乙型流感病毒、丙型流感病毒、鲑传贫病毒 (isavirus)、托高土病毒 (thogotovirus)	包膜	螺旋	ss(-)	V
	副黏病毒科	麻疹病毒、流行性腮腺炎病毒、呼吸道合胞病毒、牛瘟病毒、犬瘟热病毒	包膜	螺旋	ss(-)	V
[0051]	布尼亚病毒科	加州脑炎病毒、汉坦病毒	包膜	螺旋	ss(-)	V
	弹状病毒科	狂犬病病毒	包膜	螺旋	ss(-)	V
	丝状病毒科	埃博拉病毒、马尔堡病毒	包膜	螺旋	ss(-)	V
	冠状病毒科	冠状病毒	包膜	螺旋	ss	IV
	星状病毒科 (Astroviridae)	星状病毒	包膜	二十面体	ss	IV
	博尔纳病毒科 (Bornaviridae)	博尔纳病病毒	裸露	螺旋	ss(-)	V
	动脉炎病毒科 (Arteriviridae)	动脉炎病毒、马动脉炎病毒	包膜	二十面体	ss	IV
	戊型肝炎病毒科 (Hepeviridae)	戊型肝炎病毒	包膜	二十面体	ss	IV

[0052] 任选地,病毒是新出现或再出现的RNA病毒。新出现或再出现的RNA病毒的实例包括埃博拉病毒、马尔堡病毒、拉沙病毒、流感病毒、MERS冠状病毒、亨德拉病毒、尼帕病毒。

[0053] 任选地,病毒是丝状病毒或沙粒病毒。任选地,病毒是埃博拉病毒或马尔堡病毒。任选地,病毒是拉沙病毒。任选地,病毒是流感病毒。

[0054] 任选地,病原体是DNA病毒。任选地,病原体是痘病毒科的成员,例如猴痘病毒。

[0055] DNA病毒包括:

[0056] 组I:病毒具有双链DNA。引起水痘和疱疹的病毒见于此。

[0057] 组II:病毒具有单链DNA。

[0058] 下表列出了巴尔的摩分类下不同DNA病毒科的分组：

DNA 病毒科	实例 (俗名)	衣壳: 裸露/包膜	衣壳对称性	核酸类型	组
腺病毒科	腺病毒、犬传染性肝炎病毒	裸露	二十面体	ds	I
乳多空病毒科 (<i>Papovaviridae</i>)	乳头瘤病毒、多瘤病毒科 (<i>polyomaviridae</i>)、 猴空泡病毒	裸露	二十面体	ds 环状	I
小 DNA 病毒科	细小病毒 B19、犬细小病毒	裸露	二十面体	ss	II
[0059] 疱疹病毒科	单纯疱疹病毒、水痘带状疱疹病毒、巨细胞病毒、EB 病毒	包膜	二十面体	ds	I
痘病毒科	天花病毒、牛痘病毒、绵羊痘病毒、羊口疮病毒、猴痘病毒、牛痘病毒	复合外壳	复合体	ds	I
嗜肝 DNA 病毒科 (<i>Hepadnaviridae</i>)	乙型肝炎病毒	包膜	二十面体	环状, 部分 ds	VII
非洲猪瘟病毒科 (<i>Asfarviridae</i>)	非洲猪瘟病毒	包膜	二十面体	ds	I

[0060] 任选地, 病原体是逆转录病毒。逆转录病毒包括:

[0061] 组VI: 病毒具有通过DNA中间体复制的单链RNA病毒。逆转录病毒包括在该组中, HIV是其中的一员。

[0062] 组VII: 病毒具有双链DNA基因组, 并使用逆转录酶复制。乙型肝炎病毒可见于该组中。

[0063] 术语“亚型”在本文中用于指病原体 (例如, 病毒、细菌或真菌) 的遗传变体或株。例如, 埃博拉病毒属是丝状病毒科中包括的病毒分类单元。这个属的成员被称为埃博拉病毒。六种已知的埃博拉病毒亚型以最初各自被鉴定的区域命名: 本迪布焦、雷斯顿 (Reston)、苏丹、塔伊森林、扎伊尔 (Zaire) 和邦巴利 (Bombali)。甲型流感病毒根据病毒表面的两种蛋白

质分为亚型:血凝素(HA)和神经氨酸酶(NA)。有18种已知的HA亚型和11种已知的NA亚型。HA和NA蛋白的许多不同组合是可能的。例如,“H7N2病毒”表示具有HA 7蛋白和NA 2蛋白的甲型流感病毒亚型。类似地,“H5N1”病毒具有HA 5蛋白和NA 1蛋白。

[0064] Kuhn et al. (Arch Virol. 2013 Jan; 158(1): 301-311) 讨论了丝状病毒科的自然变体的病毒命名法。根据作者, (天然) 病毒株是“给定病毒的可识别变体, 因为其具有在自然条件下保持稳定的一些独特表型特征。” 这样的“独特表型特征”是与比较的参考病毒不同的生物学特性, 例如独特的抗原特性、宿主范围或其引起的疾病的体征。具有简单“基因组序列差异的病毒变体……, 由于没有可识别的独特病毒表型, 因此没有被给予单独的毒株状态”。因此, 毒株是遗传稳定的病毒变体, 其与天然参考病毒(类型变体)的区别在于其引起明显不同的可观察到的感染表型(不同类型的疾病、感染不同类型的宿主、通过不同的方式传播等)。“遗传稳定的”意指与表型变化相关的基因组变化通过自然选择随时间推移在很大程度上得以保留。基因组序列变异的程度与将变异体分类为毒株无关, 这是因为独特的表型有时是由很少的突变产生的。“可观察到的表型”意指, 例如, 在比较动物实验中, 研究人员可在不知晓哪只动物接受了哪种病毒并且也没有关于两种病毒之间差异的任何信息的情况下区分参考对照病毒感染的动物和指称的新毒株感染的动物。将病毒变体指定为病毒株是国际专家组的责任。迄今为止, 尚未报道根据该定义的自然丝状病毒株。例如, EBOV的所有已描述的遗传变体都会在人和甚至实验动物中引起类似的出血热, 并且类似地传播。仅基于临床, 无法将已知的EBOV遗传变体与其他区分开。实际上, 它们的多样性似乎限于体外生长动力学和噬菌斑形成的细微差异或实验动物疾病持续时间的细微变化, 并且最终源自基因组序列有限但通常稳定的差异。MARV、RAVV、BDBV、RESTV和SUDV的不同遗传变体也是如此(目前, 仅有TAFV的一种分离株, 并且没有LLOV的分离株。)

[0065] 根据Kuhn等人, 天然遗传丝状病毒变体是一种天然丝状病毒, 其基因组共有序列与参考丝状病毒(特定丝状病毒物种的类型病毒)的基因组共有序列相差 $\leq 10\%$, 但与参考丝状病毒不同, 并且不会引起可观察到的不同疾病表型(丝状病毒株可以是遗传丝状病毒变体, 但如果提出毒株定义, 则大多数遗传丝状病毒变体将不是丝状病毒株)。

[0066] 病毒分类的另一种方案是国际病毒分类委员会(International Committee on Taxonomy of Viruses, ICTV) 系统。该系统与细胞生物体的分类系统共有许多功能, 例如分类单元结构。然而, 此命名系统在某些点与其他分类法代码有所不同。病毒分类从“目”级别开始, 并如下继续, 并以斜体给出分类单元后缀:

[0067] 目(-virales)

[0068] 科(-viridae)

[0069] 亚科(-virinae)

[0070] 属(-virus)

[0071] 种

[0072] 种名通常采用[疾病]病毒的形式, 特别是对于高等植物和动物。

[0073] “目”的建立基于以下推论, 其包含的病毒科最可能从共同祖先进化而来。大多数病毒科仍处于未入选(unplaced) 状态。截至2017年, ICTV定义了病毒的9个目、131个科、46个亚科、803个属和4, 853个种。目是有尾噬菌体目(Caudovirales)、疱疹病毒目(Herpesvirales)、线状病毒目(Ligamenvirales)、单股反链病毒目(Mononegavirales)、网

巢病毒目 (Nidovirales)、逆转录病毒目 (Ortervirales)、微RNA病毒目 (Picornavirales)、布尼亚病毒目 (Bunyavirales) 和芜菁黄花叶病毒目 (Tymovirales)。这些目跨越具有不同宿主范围的病毒。

[0074] • 有尾噬菌体目是有尾的dsDNA (组I) 噬菌体。

[0075] • 疱疹病毒目包含大的真核dsDNA病毒。

[0076] • 线状病毒目包含线性dsDNA (组I) 太古代病毒。

[0077] • 单股反链病毒目包括非分段 (-) 链ssRNA (组V) 植物和动物病毒。

[0078] • 网巢病毒目由具有脊椎动物宿主的 (+) 链ssRNA (组IV) 病毒构成。

[0079] • 逆转录病毒目包含通过DNA中间体复制的单链RNA和DNA病毒 (组VI和VII)。

[0080] • 微RNA病毒目包含感染多种植物、昆虫和动物宿主的小 (+) 链ssRNA病毒。

[0081] • 芜菁黄花叶病毒目包含感染植物的单分 (monopartite) (+) ssRNA病毒。

[0082] • 布尼亚病毒目包含三分 (tripartite) (-) ssRNA病毒 (组V)。

[0083] 根据ICTV, 病毒种是“病毒的单源组, 其性质可通过多种标准与其他种区分开来。”

[0084] 术语“分离株”在本文中是指从感染个体获得的纯病原体样品。仅经过一轮复制后, 病毒感染的细胞将已经包含的基因组的群体, 并且来源于这些基因组的病毒体彼此之间会略有不同。同样, 从受感染个体采集的样品将包含许多病毒体, 其中许多会略有不同, 因此, “分离株”是指群体, 而“分离株”的“序列”是分析的样品中存在的基因组群体的共有序列。病毒分离株可定义为“特定病毒的实例”。天然丝状病毒分离株是特定天然丝状病毒或特定遗传变体的实例。分离株在共有序列或个体序列上可彼此相同或略有不同。

[0085] 任选地, 一种或更多种广泛中和抗原结合分子包括获自对象的抗体或来源于获自对象的抗体的抗体, 所述对象已经暴露于这样的病原体, 所述病原体与期望诱导对其广泛中和免疫应答的病原体为同一科。

[0086] 任选地, 一种或更多种广泛中和抗原结合分子包括获自对象的抗体或来源于获自对象的抗体的抗体, 所述对象已经暴露于这样的病原体, 所述病原体与期望诱导对其广泛中和免疫应答的病原体为同一亚型或类型。

[0087] 任选地, 一种或更多种广泛中和抗原结合分子包括获自对象的抗体或来源于获自对象的抗体的抗体, 所述对象已经暴露于这样的病原体, 所述病原体与期望诱导对其广泛中和免疫应答的病原体为同一属或种。

[0088] 任选地, 一种或更多种广泛中和抗原结合分子包括非抗体抗原结合蛋白。例如, 一种或更多种广泛中和抗原结合分子可包括经设计锚蛋白重复蛋白 (DARPin)、适配体、抗运载蛋白 (anticalin) 或T细胞受体分子。

[0089] DARPin是遗传改造的抗体模拟蛋白, 其通常表现出高特异性和高亲和力的靶蛋白结合。其源自天然锚蛋白, 并包含重复的结构单元, 该结构单元形成具有大的潜在靶相互作用表面的稳定蛋白结构域。通常来说, DARPin包含四个或五个重复, 其中第一个 (N加帽重复) 和最后一个 (C加帽重复) 用于提供亲水性表面。DARPin对应于天然锚蛋白重复蛋白结构域的平均尺寸。具有少于三个重复 (即加帽重复和一个内部重复) 的蛋白质不能形成足够稳定的三级结构。DARPin的分子量取决于总重复数:

[0090]

重复	3	4	5	6	7	...
~质量 (kDa)	10	14	18	22	26	...

[0091] 可产生编码具有随机潜在靶相互作用残基的DARPin的核酸文库,其具有超过 10^{12} 种变体的多样性。从这些文库中,可使用允许共翻译分泌的信号序列使用核糖体展示或噬菌体展示选择以皮摩尔亲和力和特异性与所选的期望靶标结合的DARPin。因此,通过筛选DARPin的文库,可鉴定结合和/或中和病原体的多于一种亚型的一种或更多种DARPin。用于鉴定DARPin的基于文库的筛选描述于例如Hartmann et al. (Molecular Therapy:Methods and Clinical Development 2018 Vol.10:128-143)中。

[0092] 任选地,本发明方法的步骤(ii)中所述的一种或更多种抗原结合分子包括广泛中和抗体(或者其保留广泛中和活性的片段或衍生物),例如广泛中和单克隆抗体(BNmAb)(或者其保留广泛中和活性的片段或衍生物)。

[0093] 任选地,本发明方法的步骤(ii)中所述的一种或更多种抗原结合分子包括获自对象的抗体或来源于获自对象的抗体的抗体,所述对象在这样的病原体的爆发中幸存,所述病原体与期望诱导对其广泛中和免疫应答的病原体为同一亚型、类型或科。

[0094] 任选地,本发明方法的步骤(ii)中所述的一种或更多种抗原结合分子包括获自对象的抗体或来源于获自对象的抗体的抗体,所述对象在这样的病原体的爆发中幸存,所述病原体与期望诱导对其广泛中和免疫应答的病原体为同一种、属或科。

[0095] 本文使用的术语“爆发”是指在限定的机构(例如医院或医疗中心)、社区、地理区域或时间段内比通常可预期的更多的病例的发生。爆发可能发生在受限制的地理区域内,或可能延伸到多个国家/地区。其可能持续几天或几周,或几年。表明爆发存在的病例数将根据病原体、暴露群体的大小和类型、以前对该疾病的经历或缺乏暴露以及发生的时间和地点而有所不同。因此,爆发的状态与一年的同一季节内同一社区中同一区域内疾病的通常频率相对。可通过将当前信息与一年中同一时间内群体或社区中的先前发病率进行比较以确定所观察到的病例数是否超过预期数目来确定是否存在爆发。

[0096] 任选地,病原体的爆发可以是指在一个或多个季节中或者在一年中在一定区域(例如大陆地区)或国家中或在群体或社区中由病原体造成的比通常可预期的更多的疾病病例的发生。

[0097] 任选地,病原体(例如病毒)的爆发是在一个季节中在一定区域(例如大陆地区)中由病原体造成的比通常可预期的更多的疾病病例的发生。

[0098] 任选地,病原体(例如病毒)的爆发是在一个季节中在群体中由病原体造成的比通常可预期的更多的疾病病例的发生。

[0099] 大陆地区的实例包括非洲地区:

[0100] • 非洲北部:阿尔及利亚;加那利群岛;休达;埃及;利比亚;马德拉;梅利利亚;摩洛哥;苏丹;突尼斯;西撒哈拉;

[0101] • 非洲东部:布隆迪;科摩罗;吉布提;厄立特里亚;埃塞俄比亚;肯尼亚;马达加斯加;马拉维;毛里求斯;马约特;莫桑比克;留尼汪;卢旺达;塞舌尔;索马里;南苏丹;坦桑尼亚;乌干达;赞比亚;津巴布韦

[0102] • 非洲中部:安哥拉;喀麦隆;中非共和国;乍得;刚果民主共和国;刚果共和国;赤道几内亚;加蓬;圣多美和普林西比;

[0103] • 非洲西部:贝宁;布基纳法索;佛得角;象牙海岸;冈比亚;加纳;几内亚;几内亚比绍;利比里亚;马里;毛里塔尼亚;尼日尔;尼日利亚;圣海伦娜;塞内加尔;塞拉利昂;多

哥；

[0104] • 非洲南部:博茨瓦纳;莱索托;纳米比亚;南非;斯威士兰。

[0105] 任选地,从中获得所述抗体或所述抗体所来源的对象是人或非人哺乳动物对象。

[0106] 可使用任何合适的表达系统表达多肽文库中的候选经优化抗原性病原体多肽。合适的实例包括哺乳动物细胞、或者酵母或昆虫或细菌细胞。

[0107] 任选地,多肽文库中的候选经优化抗原性病原体多肽被表达在表达系统的细胞表面上。细胞表面表达提高了候选经优化抗原性病原体多肽正确折叠的可能性。

[0108] 任选地,通过流式细胞术对候选经优化抗原性病原体多肽被一种或更多种抗原结合分子的结合进行筛选。例如,表达候选经优化抗原性病原体多肽的细胞可用于流式细胞术测定中。

[0109] 任选地,使用第一测定(例如流式细胞术)对候选经优化抗原性病原体多肽被一种或更多种广泛中和抗原结合分子的结合进行筛选,并且使用第二测定(例如中和测定)对候选经优化抗原性病原体多肽被一种或更多种广泛中和抗原结合分子的结合进行筛选。

[0110] 任选地,病原体是病毒,候选经优化抗原性病原体多肽是候选经优化抗原性病毒多肽,并且病原体肽是病毒多肽。

[0111] 任选地,多肽文库是包括多个不同病毒假型的病毒假型文库,每个不同病毒假型包含不同候选经优化抗原性病原体多肽,例如不同候选经优化抗原性病毒多肽(例如病毒糖蛋白)。

[0112] 任选地,在步骤(ii)中,通过对病毒假型被一种或更多种抗原结合分子的结合和/或中和的筛选来对候选经优化抗原性病毒多肽被一种或更多种广泛中和抗原结合分子的结合进行筛选。

[0113] 假型化是产生与外来病毒包膜蛋白组合的病毒或病毒载体的过程。结果是假型病毒颗粒。假型颗粒不携带产生其他病毒包膜蛋白的遗传物质,因此表型变化不能传递给子代病毒颗粒。“假型”可被定义为杂交病毒颗粒,其包含包裹核酸(RNA或DNA)基因组的蛋白质核衣壳(“核芯”),核芯本身被包封在来源于宿主细胞的脂质“包膜”膜中。当核芯通过“出芽”离开细胞时获得的该包膜包含来源于其他病毒的蛋白质。这些异源包膜蛋白中的许多是宿主免疫系统的抗原性靶标。在假型中,这些包膜蛋白中的一种或更多种可能来源于研究病毒。许多假型还携带被改造到其基因组中的外来基因,称为“转移”基因。当存在易感细胞时,包膜蛋白与细胞受体结合,允许细胞进入,最终导致转移基因表达。弹状病毒(例如水泡性口炎病毒,VSV)和逆转录病毒(例如慢病毒)已被广泛用作假型化的核芯。对于逆转录病毒,其关键特征是能够将其二聚体单链RNA基因组逆转录为双链脱氧核糖核酸(dsDNA)副本,然后通过使用病毒和细胞酶将其整合到细胞基因组中。对于逆转录病毒假型,这通常导致转移/报道基因的表达,后者易于定量。报道基因的表达直接相关于病毒包膜/受体相互作用的效率,以及相反地,单个抗体应答或抗病毒剂是否会干扰天然病毒的进入和复制过程。

[0114] 病毒假型与广泛中和抗原结合分子的结合可使用技术人员已知的任何合适技术来测量,例如通过血凝抑制(HI)测定或通过酶联免疫吸附测定(ELISA)。与糖蛋白(GP)结合的抗体的ELISA分析在Saphire et al.2018(Cell 174(4):938-952)中描述,其涉及针对埃博拉病毒GP的单克隆抗体的分析。

[0115] 逆转录病毒假型的产生及其在假型中和测定和免疫原性测试中的用途在 Temperton et al.2015 (Retroviral Pseudotypes-From Scientific Tools to Clinical Utility. In: eLS. John Wiley & Sons, Ltd: Chichester. DOI:10.1002/9780470015902.a0021549.pub2) 中详细综述。

[0116] 逆转录病毒的所有七个属的代表已用于假型化研究,但迄今为止,仅广泛使用 γ 逆转录病毒或慢病毒假型。慢病毒是逆转录病毒科的一个属,与 γ 逆转录病毒不同,其可感染非增殖细胞,这使其适合于涉及包括肌肉或神经元的高度分化或静止细胞(例如在 G_0 细胞周期阶段)的基因治疗应用。用于假型化最常用慢病毒载体是1型HIV (HIV-1),尽管也使用了猿猴免疫缺陷病毒。

[0117] 逆转录病毒假型的产生是通过将一种或更多种外源包膜蛋白基因、核心逆转录病毒基因和转移基因(例如报道基因或治疗基因)的克隆版本同时引入生产细胞中实现的,生产细胞通常是高度可转染的细胞系,例如人胚胎肾(HEK)293克隆17T细胞(美国典型培养物保藏中心#CRL-11268)(Pear et al.,1993,PNAS USA 90:8392-8396)。

[0118] 1.包膜质粒。将研究病毒的一种或更多种包膜基因克隆到合适的表达质粒中。基因通常通过使用特定引物通过病毒cDNA的聚合酶链反应扩增而获得,或来自定制基因合成。一些表达载体是可商购的并且利用不同的通常强的组成型基因启动子(例如来自人巨细胞病毒(CMV)即早期基因),其可影响假型产生的效力。

[0119] 2.逆转录病毒gag-pol质粒。gag和pol基因编码多蛋白,该多蛋白随后被切割以释放见于核芯内的结构蛋白(包括基质、衣壳和核衣壳),以及参与病毒复制的蛋白质(蛋白酶、逆转录酶和整合酶),其负责处理所述结构蛋白,将ssRNA病毒基因组转化成dsDNA并确保(转移基因)整合到宿主细胞基因组中。另外,在慢病毒gag-pol构建体中包含rev基因。Rev蛋白参与病毒mRNA从细胞核输出到胞质中以进行翻译。

[0120] 3.转移/报道质粒。这是稳定整合到宿主细胞DNA中的基因,该基因从宿主细胞DNA中通过多种顺式作用转录元件表达。转移质粒包含在基因上游的包装信号,以确保在假型产生过程中将包含该基因的病毒RNA并入病毒核芯中。

[0121] 一旦细胞机器已经转录并翻译了转染的基因,就可通过包装信号将转移基因(长末端重复(LTR)之间的区域)的RNA二聚体并入到假型中。由于转移质粒是唯一被工程化以包含包装信号的质粒,因此没有其他核酸被并入到成熟的假型颗粒中。Gag的N末端的结构域将核衣壳靶向细胞质膜,在该细胞质膜中已插入了一种或更多种包膜蛋白。从细胞出芽的假型颗粒被包封在细胞膜中,其形成病毒包膜。

[0122] 假型病毒被释放到生产细胞培养基中。该上清液可滴定到靶细胞上以测量功能性颗粒的浓度。这些通过包膜蛋白-受体相互作用附着至细胞,随后是膜融合和内化。带有转移/报道基因的假型基因组被整合到宿主细胞DNA中,并从那里表达。报道基因的表达水平与活颗粒的转导水平相关。由于在假型中仅存在转移基因,因此靶细胞中不会产生病毒蛋白,因此不会发生进一步的假型产生和繁殖。与使用野生型病毒工作相比,使用假型工作提供了安全性。基于绿色荧光蛋白(GFP)的假型易于使用荧光显微术或流式细胞术进行滴定,通过发光分析(luminometry)的萤光素酶假型,以及通过显色反应的 β -半乳糖苷酶(β -gal)假型。

[0123] 许多标准血清学测定仅测量抗体结合(血凝抑制(HI)和ELISA),而不测量病毒感

染性的抑制。中和测定允许灵敏地检测功能性抗体应答。然而,对于高防控病毒(例如埃博拉),由于需要高度生物安全的实验室设施和受过专门培训的人员,因此这些检测并未广泛应用。使用利用这样的病原体的包膜假型化的逆转录病毒和慢病毒颗粒作为用于中和测定的“替代病毒”是解决此问题的一种方法。使用假型策略,仅需要病毒的一种或更多种包膜蛋白,而没有重组或天然病毒逃逸的可能性。这些假型经历了顿挫复制(abortive replication),并且无法产生具有复制能力的后代。

[0124] 假型是用于病毒中和测定的优秀血清学试剂,因为病毒体可包含报道基因,并在表面带有异源病毒包膜蛋白。这些报道基因向靶细胞的转移取决于病毒包膜蛋白的功能;因此,可通过报道基因转移和表达的减少来测量针对包膜的中和抗体的滴度。现在已经开发出针对来自多种病毒科的广泛范围的RNA病毒的PV中和测定(参见Temperton等的表1(同上))。

[0125] 已显示基于假型的流感中和测定对于测量广泛中和抗体非常有效,使其成为研究针对具有大流行潜力的多种亚型的交叉反应性应答的理想血清学工具(Corti et al., 2011, Science 333(6044):850-856)。

[0126] 在Sinn et al. 2017 (Methods Mol Biol. 2017;1628:65-78) 中描述了用丝状病毒糖蛋白假型化的慢病毒载体的产生。

[0127] 产生病毒假型的合适的一般性方法的实例如下:

[0128] 为了进行转染,将 5×10^6 个HEK-293T细胞铺板,24小时后添加包含质粒DNA和PEI的复合物,这有助于DNA转运到细胞中。将逆转录病毒gag-pol质粒和报道质粒与所需的包膜质粒同时转染。

[0129] 合适的中和测定的一个实例如下:

[0130] 在96孔板中,将约 $100 \times$ TCID₅₀假型病毒(其导致输出为 1×10^5 相对光单位(RLU))与血清稀释液在37% (5%CO₂) 下孵育1小时,然后添加 1×10^4 个靶细胞。将它们再孵育48小时,然后除去培养基,并用新鲜培养基和萤光素酶试剂的50:50混合物替换。2.5分钟后通过在光度计上读取板来检测萤光素酶活性。对于所有结果,在分析之前都会扣除背景RLU(单独病毒或DEnv)。

[0131] Sapphire等(同上)描述了涉及针对埃博拉病毒GP的单克隆抗体的分析的用于评价mAb中和的三种独立测定:

[0132] i) 生物上包含EBOV (Δ VP30) (Halfmann et al., 2008, Proc Natl Acad Sci USA. 2008;105:1129-1133); 和

[0133] ii) 在BSL-2+、BSL-3和BSL-4防控下执行的真实EBOV; 以及

[0134] iii) 带有EBOV GP的具有复制能力的水疱性口炎病毒(rVSV)。

[0135] Ebola Δ VP30-RenLuc病毒的中和

[0136] 使用其中报道基因海肾萤光素酶基因代替了病毒转录因子VP30的埃博拉病毒(Ebola Δ VP30-RenLuc病毒)来补充稳定地反式表达VP30的Vero细胞系(Vero VP30),从而允许在BSL-3进行分析(Halfmann et al., 2008)。将在最小必需培养基中的2%胎牛血清中稀释的总共 5×10^3 个病灶形成单位的Ebola Δ VP30-RenLuc病毒与50 μ g/ml的单克隆抗体在37 $^{\circ}$ C下孵育3小时。然后将感染复数(MOI)为0.001的病毒/抗体混合物添加至在前一天以 9×10^3 个细胞/孔接种在96孔板中的Vero VP30细胞,并在37 $^{\circ}$ C和5%CO₂下孵育3天。如果使

用,则将豚鼠补体 (Cedarlane) 以10%的终浓度添加至最小必需培养基。然后将活细胞荧光素酶底物EnduRen (Promega) 与细胞孵育3小时,然后使用Tecan M1000读板仪 (Tecan) 将荧光素酶值作为相对光单位 (RLU) 进行测量。一式两份进行测定,并将已知的中和 (GP 133/3.16) 和非中和单克隆抗体 (VP35 5/69.3.2) 分别用作阳性和阴性对照。将荧光素酶信号中和 $\geq 95\%$ 的抗体定义为强中和剂,而将荧光素酶信号抑制50%至94%的视为中等中和剂,并且将具有49%或更低抑制的那些归类为弱/非中和剂。

[0137] 真实EBOV的中和

[0138] 根据Holtsberg et al. (Holtsberg et al., 2015, J Virol. 2015; 90: 266-278) 中描述的方法进行评估真实EBOV中和的测定。在病毒感染之前24小时,将Vero E6细胞以 2.5×10^4 个细胞/孔接种到黑色96孔板的内部60孔中。将抗体在Vero生长培养基 (Eagle最低必需培养基,含Earle盐和L-谷氨酰胺、5%胎牛血清 (FBS) 和1%青霉素-链霉素) 中系列稀释至期望最终浓度 (50 $\mu\text{g}/\text{ml}$) 的两倍,与等体积的活EBOV混合,并在37 $^{\circ}\text{C}$ 下孵育1小时,每15分钟进行混合。然后将MOI为0.2的抗体/病毒混合物添加至Vero细胞,并在37 $^{\circ}\text{C}$ 下孵育1小时,用PBS洗涤,将单独的生长培养基添加至所有孔,并将板在37 $^{\circ}\text{C}$ 下再孵育48小时。然后将细胞用10%中性缓冲福尔马林固定,并使用与Alexa Fluor 488 (Molecular Probes) 缀合的EBOV特异性人mAb KZ52和山羊抗人IgG作为二抗通过间接免疫荧光测定确定感染细胞的百分比。在Operetta High Content成像系统 (Perkin-Elmer) 上使用20倍物镜以20视野/孔获取图像。使用自定义算法分析Operetta图像,该算法是根据Harmony软件 (Perkin-Elmer) 中提供的图像分析函数构建的。相对于用单独培养基孵育的对照细胞,确定每种抗体的抑制百分比。将感染细胞的百分比降低 $> 80\%$ 的抗体被归类为强中和剂,而将感染降低50%至79%以及小于50%的抗体分别被视为中等中和剂和弱/非中和剂。

[0139] rVSV-EBOV GP的中和

[0140] 先前描述了表达eGFP和重组表面GP (rVSV-EBOV) 来代替VSV G的重组水泡性口腔炎病毒 (VSV) (Wec et al., 2016, Science; 354: 350-354; Wong et al., 2010, Virol.; 84: 163-175)。为了进行中和测定,将Vero细胞以 6.0×10^4 个细胞/孔接种,并在补充有10%胎牛血清 (FBS) 和100 I.U./ml青霉素和100 $\mu\text{g}/\text{ml}$ 链霉素的Eagle最小必需培养基 (EMEM) 中于37 $^{\circ}\text{C}$ 和5%CO₂下培养过夜。第二天,将病毒与无血清EMEM中从330nM (~50 $\mu\text{g}/\text{ml}$) 开始的系列3倍抗体稀释液在室温下孵育1小时,然后感染96孔板中的Vero细胞单层。基于病毒储液的滴定来确定用于感染的病毒的量以在不含抗体的对照孔中实现35%至50%的最终感染 (MOI为每个细胞约0.1感染单位)。将病毒与补充有2%FBS、100 I.U./ml青霉素和100 $\mu\text{g}/\text{ml}$ 链霉素的50% v/v/EMEM中的细胞在37 $^{\circ}\text{C}$ 和5%CO₂下孵育14至16小时,然后将细胞固定,并且用Hoescht对核染色。通过使用Cellinsight CX5自动显微镜和随附软件 (Thermo Scientific) 将EGFP阳性细胞与通过核染色指示的细胞总数相比进行计数,来测量rVSV感染性。将缺乏抗体的对照孔中的感染水平设置为100%,并将一式三份测试的每种抗体稀释液的感染相对于该值进行归一化。确定平均值,并使用GraphPad Prism版本6使用完整的9点稀释曲线确定半最大抑制剂浓度IC₅₀。IC₅₀ $\leq 5\text{nM}$ 的抗体被视为强中和剂,而 $5\text{nM} < \text{IC}_{50} < 50\text{nM}$ 和 $\leq 50\text{nM}$ 的抗体分别被视为中等中和剂和弱/非中和剂。还使用测试的最高浓度 (330nM) 的抗体并测量相对于未处理的对照细胞的GFP信号来确定作为抗体效力指标的未中和级分。那些将信号降低 $\geq 98\%$ 、50%至98%、和小于50%的信号分别视为强、中等、和

弱/非中和剂。

[0141] 用于筛选多肽文库的方法描述于Bruun等(PLoS ONE 9(10):e109196中。

[0142] 任选地,本发明的方法还包括产生多肽文库。

[0143] 任选地,通过从包含多个不同核酸的核酸文库中表达不同候选经优化抗原性病原体多肽来产生多肽文库,每个不同核酸包含编码多肽文库中的不同候选经优化抗原性病原体多肽的核苷酸序列。

[0144] 任选地,不同候选经优化病原体多肽被表达在哺乳动物细胞内或表面上。合适的方法是本领域技术人员公知的。

[0145] 任选地,核酸文库中的每个不同核酸的核苷酸序列针对在哺乳动物细胞中表达所编码的多肽而进行了优化。

[0146] 任选地,核酸文库中的每个不同核酸是用于在哺乳动物细胞中表达核酸的表达载体的一部分。

[0147] 任选地,病原体是病毒,候选经优化抗原性病原体多肽是候选经优化抗原性病毒多肽,并且病原体肽是病毒多肽。

[0148] 任选地,核酸文库是病毒假型载体文库,并且文库中的每个不同核酸是用于产生包含编码的病毒多肽的病毒假型的表达载体的一部分,并且多肽文库是通过从病毒假型载体文库中的表达载体产生病毒假型而产生的病毒假型文库,其中病毒假型文库包含多个不同病毒假型,每个不同病毒假型包含由病毒假型载体文库中的不同核酸序列编码的不同候选经优化病毒多肽。

[0149] 任选地,病毒假型载体文库包含至少2、3、5、10、20、30、40、50、 10^2 、 10^3 、 10^4 、 10^5 、 10^6 、 10^7 、 10^8 或 10^9 个不同成员。

[0150] 任选地,表达载体也是疫苗载体。

[0151] 疫苗载体的实例包括病毒疫苗载体、细菌疫苗载体、RNA疫苗载体或DNA疫苗载体。

[0152] 病毒疫苗载体使用活病毒将核酸(例如DNA或RNA)携带到人或非人动物细胞中。病毒中包含的核酸编码一种或更多种抗原,其一旦在受感染的人或非人动物细胞中表达,就会引起免疫应答。病毒疫苗载体可诱导体液和细胞介导的免疫应答二者。病毒疫苗载体组合了核酸疫苗与活减毒疫苗的许多积极特性。与核酸疫苗一样,病毒疫苗载体也将核酸携带到宿主细胞中以产生抗原蛋白,可对该抗原蛋白进行定制以刺激一系列免疫应答,包括抗体、T辅助细胞(CD4⁺T细胞)和细胞毒性T淋巴细胞(CTL、CD8⁺T细胞)介导的免疫。与核酸疫苗不同,病毒疫苗载体还具有主动侵入宿主细胞并复制的潜力,非常像活减毒疫苗,像佐剂一样进一步激活免疫系统。因此,病毒疫苗载体通常包含活减毒病毒,其经过遗传改造以携带编码来自不相关生物体的蛋白质抗原的核酸(例如DNA或RNA)。尽管病毒疫苗载体通常能够产生比核酸疫苗更强的免疫应答,但是对于一些疾病,在一种称为异源初免-加强(prime-boost)的策略中,病毒载体与其他疫苗技术组合使用。在该体系中,给予一种疫苗来作为初免步骤,随后使用替代疫苗作为加强物进行疫苗接种。异源初免-加强策略旨在提供更强的整体免疫应答。病毒疫苗载体可用作初免疫苗和加强疫苗二者来作为该策略的一部分。病毒疫苗载体已由Ura等,2014(Vaccines 2014,2,624-641)以及Choi和Chang,2013(Clinical and Experimental Vaccine Research 2013;2:97-105)进行了综述。

[0153] 任选地,病毒疫苗载体基于病毒递送载体,例如基于痘病毒(例如修饰的痘苗病毒

安卡拉 (MVA)、NYVAC、AVIPOX)、疱疹病毒 (例如任何宿主物种的HSV、CMV、腺病毒)、麻疹病毒 (例如麻疹)、甲病毒 (例如SFV、仙台)、黄病毒 (例如黄热病) 或弹状病毒 (例如VSV) 的病毒递送载体, 细菌递送载体 (例如沙门菌、大肠杆菌), RNA表达载体或DNA表达载体。

[0154] 任选地, 载体是基于pEVAC的表达载体。pEVAC表达载体在下面的实施例7中有更详细的描述。

[0155] 在另一些实施方案中, 不同候选经优化抗原性病原体多肽被表达在细菌、酵母或昆虫细胞内或表面上。

[0156] 任选地, 本发明的方法还包括通过合成多个不同核酸来产生核酸文库, 每个不同核酸包含编码不同候选经优化抗原性病原体多肽的不同核苷酸序列。

[0157] 任选地, 本发明的方法还包括: i) 获得不同病原体分离株的病原体多肽的氨基酸序列和/或编码病原体多肽的核苷酸序列; 以及ii) 产生多个不同核苷酸序列, 每个不同核苷酸序列编码不同候选经优化抗原性病原体多肽, 其中每个不同候选经优化抗原性病原体多肽的编码的氨基酸序列根据病原体多肽的所获得氨基酸序列或编码的氨基酸序列进行优化, 并且不同于每个获得的氨基酸序列或编码的氨基酸序列。

[0158] 任选地, 在以上步骤(ii)中产生多个不同核苷酸序列包括: 对在以上步骤(i)中获得的氨基酸或核苷酸序列进行多序列比对; 从多序列比对中鉴定在不同病原体分离株的多肽之间高度保守的氨基酸序列或编码的氨基酸序列; 以及产生多个不同核苷酸序列, 每个不同核苷酸序列编码不同候选经优化抗原性病原体多肽, 其中不同核苷酸序列中的一个或更多个包含编码从多序列比对中鉴定的高度保守的氨基酸序列或编码的氨基酸序列的序列。

[0159] 在不同病原体分离株的多肽之间高度保守的氨基酸序列或编码的氨基酸序列的长度可为至少1、2、3、4、5、10、15、20、25、30、35、40、45、50、100、150、200、250、300、400、500、600、700或800个氨基酸残基。

[0160] 任选地, 不同病原体分离株的病原体多肽的氨基酸序列的数目或编码病原体多肽的核苷酸序列的数目为至少3、4、5、10、20、30、40、50、100、200、300、400、500、600、700、800、900、1000、 10^6 、 10^9 或 10^{12} 。通常来说, 用于多序列比对的序列数目越多越好。

[0161] 任选地, 本发明的方法还包括: 从多序列比对中鉴定出作为祖先氨基酸序列的氨基酸序列或编码的氨基酸序列; 以及在不同的所产生的核苷酸序列中的一个或更多个中包含编码从多序列比对中鉴定出的祖先氨基酸序列的序列。

[0162] 包含一个或更多个编码祖先氨基酸序列的核苷酸序列可能是有利的, 因为与现存氨基酸序列高度保守的祖先氨基酸序列被认为对于病原体的存活和/或繁殖具有结构和/或功能上的重要性。另外, 由于病原体分离株可能非常多样 (特别是例如新出现或再出现的病原体 (例如新出现或再出现的RNA病毒) 的分离株), 设计用于对一个患者病原体群体发挥作用的疫苗可能不适用于不同患者, 因为这两个病原体群体之间的进化距离可能很大。然而, 相比于两个病原体群体接近彼此, 它们最近共同祖先更接近两个病原体群体中的每一个。因此, 为共同祖先设计的疫苗可能更有可能对更大比例的流行株 (circulating strain) 有效。

[0163] 祖先序列重建 (ancestral sequence reconstruction, ASR) 在Randall等 (Nat. Commun. 7:12847 doi:10.1038/ncomms 12847 (2016)) 中进行了讨论。作者将ASR的定

义称为“在进化/系统发育背景下分析现代序列以推断在树的特定节点的祖先序列的过程”。祖先序列重建 (ASR) 用于分子进化研究。与通过水平比较系统发育树的不同分支末端的相关蛋白质同源物来研究蛋白质的传统进化方法不同, ASR以竖直方式探测树节点内统计推断的祖先蛋白(参见图1)。系统发育树是根据其物理或遗传特征的相似性和差异显示多种生物物种或其他实体之间进化关系的分支图。在有根的系统发育树中, 每个带有后代的节点代表这些后代的推断的最近共同祖先。在ASR中, 选择目的蛋白质的多种相关同源物, 并以多序列比对 (MSA) 进行比对, 构建系统发育树, 其在分支的节点处具有统计推断序列。这些序列就是所谓的“祖先”。合成相应DNA, 将其转化到细胞中并产生蛋白质的过程就是所谓的“重建”。

[0164] 祖先序列通常是通过最大似然法 (maximum likelihood) 来计算的, 然而也可执行贝叶斯法 (Bayesian method)。由于祖先是系统发育推断的, 因此系统发育的拓扑和组成在输出ASR序列中发挥主要作用。ASR并不声称要重建古蛋白质/DNA的实际序列, 而是可能与该节点处的序列相似的序列。最大似然 (ML) 法通过产生这样的序列来工作, 其中每个位置的残基通过所使用的推论方法被预测最有可能占据该位置。通常来说, 这是从现存序列计算的得分矩阵 (类似于BLAST或MSA中使用的那些)。替代方法包括最大简约法 (maximum parsimony, MP), 其基于序列进化模型构建序列, 通常最少数目的核苷酸序列变化的概念代表发生的最有效的进化途径以及最有可能的。MP经常被认为是最不可靠的重建方法, 因为其可能将进化过度简化至不适合十亿年的规模的程度。其他方法包括贝叶斯法, 其涉及对残基不确定性的考虑。这样的方法有时用于补充ML法, 但通常会产生更多模糊序列 (即, 包含无法预测明确置换的残基位置的序列)。通常在这种情况下, 会产生涵盖大多数模糊的多个ASR序列并彼此比较。

[0165] 基于Joy et al., 2016, PLOS Computational Biology 12(7):DOI:10.1371/journal.pcbi.1004763中的描述, 以下将更详细地描述用于ASR的方法和算法。

[0166] 任选地, 利用至少3、4、5、10、20、30、40、50、100、200、300、400、500、600、700、800、900、1000、 10^6 、 10^9 或 10^{12} 个不同序列进行ASR。在一些情况下, 使用的序列数越多越好。

[0167] 任选地, 用于多序列比对的每个序列是病原体分离株的病原体多肽的全长序列。

[0168] 祖先重建的任何尝试都始于系统发育。一般而言, 系统发育是关于其中种群 (称为分类单元) 通过共同祖先的谱系关联的顺序的基于树的假设。观察到的分类单元由树的尖端或末端节点表示, 这些尖端或末端节点通过分支渐进地连接到它们的共同祖先, 其由树的通常称为祖先或内部节点的分支点表示。最终, 所有谱系都收敛到整个分类单元样本的最近共同祖先。在祖先重建的背景下, 系统发育通常被视为具有已知数量 (贝叶斯法是重要的例外)。因为可能有大量的系统发育在解释数据方面几乎同样有效, 所以将数据支持的系统发育的子集简化为单个代表或点估计可能是一种方便并且有时候必要的简化假设。祖先重建可被认为是对给定的系统发育应用假设的进化模型的结果。当模型包含一个或更多个自由参数时, 总的目标是基于在观察到的来源于共同祖先的分类单元 (序列) 中测得的特征估计这些参数。简约法是此范例的重要例外。其基于特征状态 (character state) 中的变化比较稀有的启发式方法 (heuristic), 而没有尝试量化这种稀有性。

[0169] 最大简约法

[0170] 简约法是指选择最简单的竞争假设的原理。在祖先重建的背景下, 简约法试图在

给定的树中找到祖先状态的分布,以使得解释在树的尖端处观察到的状态时必需的特征状态变化的总数最小化。这种最大简约法是最早形成的用于重建祖先状态的算法之一。最大简约法可通过多种算法之一来实现。最早的实例之一是Fitch法(Fitch WM.Toward defining the course of evolution:minimum change for a specific tree topology.Systematic Biology.1971;20(4):406-16),其通过对有根的二叉树的两次遍历,通过简约法来分配祖先特征状态。第一阶段是后序遍历,其通过在其父节点之前先访问后代(子)节点而从树的尖端向根前进。最初,基于对其后代观察到的特征状态来确定第*i*个祖先的可能特征状态的集合*S_i*。每个分配都是祖先的后代的特征状态的集合交集;如果交集为空集,则为集合并集。在后一种情况下,暗示在祖先与其两个直接后代之一之间发生了特征状态变化。每个这样的事件都计入算法的成本函数,该函数可用于基于最大简约法来区分替代树。接下来,对树进行从根向尖端前进的前序遍历。然后基于其与其父节点共享的特征状态,将特征状态分配给每个后代。由于根没有父节点,因此可能需要随意选择特征状态,特别是当在根处已重建了多于一个的可能状态时。简约法具有直观吸引力和高效性,使得在一些情况下仍然可使用它们来为具有初始系统发育的ML优化算法提供开端(seed)(Stamatakis A.RAxML-VI-HPC:maximum likelihood-based phylogenetic analyses with thousands of taxa and mixed models.Bioinformatics.2006;22:2688-90.pmid:16928733)。然而,其具有多个问题:

[0171] 1. 进化速率的变化。Fitch法假设所有特征状态之间的变化以同样可能性发生;因此,对于给定的树,任何变化都将产生相同的成本。这种假设通常是不现实的,并且会限制这种方法的准确性。例如,在核酸的进化中,转换倾向于比颠换更频繁地发生。可通过为特定特征状态变化分配差异成本来放宽这一假设,从而得到加权的简约算法(Sankoff D.Minimal mutation trees of sequences.SIAM Journal on Applied Mathematics.1975;28(1):35-42)。

[0172] 2. 快速进化。这样的方法所基于的“最小进化”启发式方法的结果是,这样的方法假定变化是稀有的,因此在变化是标准而不是例外的情况下是不适当的(Schluter D, Price T, Mooers A0, Ludwig D.Likelihood of ancestor states in adaptive radiation.Evolution.1997;51(6):1699-711;Felsenstein J.Maximum likelihood and minimum-steps methods for estimating evolutionary trees from data on discrete characters.Systematic Biology.1973;22(3):240-9)。

[0173] 3. 谱系之间的时间变化。简约法隐含地假设沿着树的每个分支都经过了相同的进化时间量。因此,它们不能解释树中分支长度的变化,而分支长度通常用于量化进化或时序时间的流逝。例如,这种限制使该技术适合推断在一个很短的分支上发生一个变化,而不是在一个很长的分支上发生多个变化。通过基于模型的方法(ML和贝叶斯法二者)解决了该缺陷,所述方法在随机进化过程沿着树的每个分支展开时推断该随机进化过程(Li G, Steel M, Zhang L.More taxa are not necessarily better for the reconstruction of ancestral character states.Systematic biology.2008;57(4):647-53)。

[0174] 4. 统计依据。没有该方法所基于的统计模型,其估计值没有明确的不确定性。

[0175] 最大似然法(ML)

[0176] 祖先序列重建的ML法将树的内部节点处的特征状态作为参数,并尝试在给定假设

(与观察到的序列或分类单元相关的进化和系统发育模型)的情况下找到使数据(观察到的特征状态)的概率最大化的参数值。一些最早的ML祖先重建方法是在遗传序列进化的背景下开发的(Yang Z, Kumar S, Nei M. A new method of inference of ancestral nucleotide and amino acid sequences. *Genetics*. 1995; 141(4): 1641-50; Koshi JM, Goldstein RA. Probabilistic reconstruction of ancestral protein sequences. *Journal of Molecular Evolution*. 1996; 42(2): 313-20); 针对离散特征进化的类似情况也开发了类似的模型(Pagel M. The maximum likelihood approach to reconstructing ancestral character states of discrete characters on phylogenies. *Systematic biology*. 1999; 48(3): 612-22)。

[0177] 这些方法采用与用于推论系统发育树相同的概率框架(Felsenstein J. Evolutionary trees from DNA sequences: a maximum likelihood approach. *Journal of molecular evolution*. 1981; 17(6): 368-76)。简而言之,通过时间可逆的连续时间马尔可夫过程(Markov process)来对遗传序列的进化建模。在最简单的情况下,所有特征都随时间以恒定速率经历独立的状态转换(例如核苷酸置换)。经常扩展此基本模型,以允许在树的每个分支上具有不同速率。实际上,突变率也可能随时间而变化(例如,由于环境变化);这可通过允许速率参数沿树进化来建模,但代价是增加了参数数目。模型定义了沿着长度为 t (以进化时间为单位)的分支从状态 i 到 j 的转换概率。从对应于提出的树的分层结构的转换概率的嵌套总和来计算系统发育的似然性。在每个节点上,其后代的似然性是在该节点上所有可能的祖先特征状态上求和:

$$[0178] \quad L_x = \sum_{S_x \in \Omega} P(S_x) \left(\sum_{S_y \in \Omega} P(S_y | S_x, t_{xy}) L_y \sum_{S_z \in \Omega} P(S_z | S_x, t_{xz}) L_z \right)$$

[0179] 其中计算出以具有直接后代 y 和 z 的节点 x 为根的子树的似然性,其中 S_i 表示第 i 个节点的特征状态, $t_{i,j}$ 是节点 i 和 j 之间的分支长度(进化时间),并且 Ω 是所有可能的特征状态(例如,核苷酸A、C、G和T)的集合。因此,祖先重建的目的是找到所有 x 个内部节点对 S_x 的分配,以使给定树的观测数据的似然性最大化。

[0180] 祖先重建的问题不是为了计算替代树的总体似然性,而是要找到每个具有最高边际ML的祖先节点处的特征状态的组合。一般来说,针对此问题有两种方法。首先,可从树的后代开始向上工作,以在仅考虑其直系后代的情况下向每个祖先渐进地分配最可能的特征状态。这种方法称为边际重建(marginal reconstruction)。类似于贪婪算法(greedy algorithm),该算法在优化问题的每个阶段都做出局部最优选择。尽管其可能非常高效,但不能保证获得针对问题的全局最佳解决方案。其次,替代地,可尝试找到整个树中祖先特征状态的联合组合(joint combination),从而共同使数据的似然性最大化。因此,这种方法称为联合重建(joint reconstruction)。尽管其不如边际重建那么迅速,但也不太可能陷入现代优化方法和启发式方法被设计要避免的非凸目标函数的局部最优中。在祖先重建的情况下,这意味着边际重建可将特征状态分配给直接祖先,其是局部最优的,但会使祖先特征状态的联合分布偏向远离全局最优。联合重建比边际重建在计算方面更加复杂。然而,已经开发了用于联合重建的有效算法,其时间复杂度通常与观察到的分类单元或序列的数目成线性关系。

[0181] 在特征之间(或基因组中的位点之间)存在进化速率的变化的情况下,基于ML的祖

先重建方法倾向于比MP法具有更高的准确性。然而,这些方法尚不能适应进化速率随时间的变化,也称为异质速度 (heterotachy)。如果特定特征的进化速度在系统发育的分支上加快,则对于给定的分支长度并假设该特征的进化速率恒定,该分支上发生的进化量将被低估。除此之外,很难将异质速度与进化速率的特征之间的变化区分开。

[0182] 由于ML(不同于最大简约法)要求研究者指定进化模型,因此其准确性可能会受到使用严重错误的模型(模型错误指定)的影响。此外,ML可能仅提供特征状态的单一重建(其通常称为“点估计”) - 当似然表面高度非凸,包括多个峰(局部最优值)时,则单点估计不能提供足够的表示,而贝叶斯法可能更合适。

[0183] 贝叶斯推断

[0184] 贝叶斯推断使用观测数据的似然性来更新研究者的信念或先验分布,以得出后验分布。在祖先重建的背景下,目的是推断给定树的每个内部节点处祖先特征状态的后验概率。而且,可将这些概率在进化模型的参数和所有可能树的空间上的后验分布上进行积分。这可表达为贝叶斯定理的应用:

$$[0185] \quad P(S|D, \theta) = \frac{P(D|S, \theta) P(S|\theta)}{P(D|\theta)}$$

[0186] $\alpha P(D|S, \theta) P(S|\theta) P(\theta)$,

[0187] 其中S代表祖先状态,D对应于观测数据,并且 θ 代表进化模型和系统发育树二者。 $P(D|S, \theta)$ 是观测数据的似然性,其可通过如上给出的Felsenstein修剪算法计算得出。 $P(S|\theta)$ 是给定模型和树的祖先状态的先验概率。最后, $P(D|\theta)$ 是给定模型和树的数据概率,其在所有可能祖先状态上积分。给出了两种表述来强调贝叶斯定理的两种不同应用,在以下讨论。

[0188] Yang及同事开发了贝叶斯祖先序列重建方法的第一实现中之一,其中分别使用进化模型和树的ML估计来定义先验分布。因此,他们的方法是计算祖先特征状态的后验概率的经验贝叶斯法的一个实例;该方法首先在软件包PAML中实现(Yang Z. PAML 4: phylogenetic analysis by maximum likelihood. Molecular biology and evolution. 2007;24(8):1586-91)。根据上述贝叶斯规则公式,经验贝叶斯法修正了从数据中获得的模型和树的经验估计,从而有效地从公式的后验似然性和先验条件中剔除掉。此外,Yang及同事(Yang Z, Kumar S, Nei M. A new method of inference of ancestral nucleotide and amino acid sequences. Genetics. 1995;141(4):1641-50)在分母中在其观察到的核苷酸序列的比对中使用位点模式的经验分布(即,将核苷酸分配给树的尖端),而不是在S的所有可能的值上在给出 θ 的情况下详尽地计算P(D)。在计算上,经验贝叶斯法类似于祖先状态的ML重建,不同之处在于,不是基于在每个内部节点处其各自概率分布来搜索状态的ML分配,而是直接报告概率分布本身。

[0189] 用于祖先重建的经验贝叶斯法要求研究者无误地假设进化模型参数和树是已知的。当数据的大小或复杂性使之成为不切实际的假设时,采用完全分层贝叶斯法并推断祖先特征状态、模型和树上的联合后验分布可能更为谨慎(Huelsenbeck JP, Bollback JP. Empirical and hierarchical Bayesian estimation of ancestral states. Systematic Biology. 2001;50(3):351-66)。Huelsenbeck和Bollback首先提出了通过使用马尔科夫链蒙特卡洛(Markov chain Monte Carlo, MCMC)法从此联合后验分布中

对祖先序列采样的层次贝叶斯法来重建祖先。相似的方法也被用来重建藻类在真菌物种中共生的进化(地衣化) (Lutzoni F, Pagel M, Reeb V. Major fungal lineages are derived from lichen symbiotic ancestors. *Nature*. 2001; 411 (6840) : 937-40)。例如, 用于MCMC的Metropolis-Hastings算法通过基于后验概率的比率接受或拒绝参数分配来探索联合后验分布。

[0190] 因此, 经验贝叶斯法计算特定树和进化模型的多种祖先状态的概率。通过将祖先状态的重建表示为概率的集合, 可直接量化将任何特定状态分配给祖先的不确定性。另一方面, 分层贝叶斯法在考虑到已观察到的数据的情况下在所有可能的树和进化模型在将这些概率平均化, 与这些树和模型的可能性成正比。

[0191] 完全贝叶斯法仅限于分析相对较少数量的序列或分类单元, 因为所有可能的树的空间会迅速变得太大, 从而使得链样本在合理的时间内收敛在计算上是不可行的。

[0192] 病原体(尤其是新出现或再出现的病原体, 例如新出现或再出现的RNA病毒) 以比哺乳动物或鸟类快几个数量级的极快速率进化。对于这些生物, 可在短得多的时间范围内应用祖先重建, 例如, 重建跨越数十年而不是数百万年的流行的全球或区域祖先。已经提出将这种重建的毒株用作疫苗设计工作的靶标, 这与当今从患者中分离的序列相反 (Gaschen et al., *Science*. 2002; 296 (5577) : 2354-60)。

[0193] 根据本发明方法的一些实施方案, ARS的任何合适的方法都可用于从多序列比对中鉴定出作为祖先氨基酸序列的氨基酸序列或编码的氨基酸序列。

[0194] 任选地, 从多序列比对中鉴定出祖先氨基酸序列包括进行最大简约法祖先序列重建(MP-ASR)。

[0195] 任选地, 从多序列比对中鉴定出祖先氨基酸序列包括进行最大似然法祖先序列重建(ML-ASR)。

[0196] 任选地, 从多序列比对中鉴定出祖先氨基酸序列包括进行贝叶斯推断祖先序列重建(BI-ASR)。

[0197] 有许多可用于执行祖先序列重建的软件包。下表(摘自Joy et al., 2016, *PLoS Computational Biology* 12 (7) : DOI:10.1371/journal.pcbi.1004763) 提供了广泛多种软件包的代表性样本, 其以不同强度和特征执行祖先重建方法:

[0198]

名称	方法	平台	支持的输入格式	特征类型	连续(C)或离散(D)特征	软件许可
PAML	ML	Unix、 Mac、Win	PHYLIP、NEXUS、 FASTA	核苷酸、蛋白质	D	所有权
BEAST2	贝叶斯	Unix、 Mac、Win	NEXUS、BEAST、XML	核苷酸、蛋白质、 地理	C、D	GNU 较宽松通用公 共许可证
APE	ML	Unix、 Mac、Win	NEXUS、FASTA、 CLUSTAL	核苷酸、蛋白质	C、D	GNU 通用公共许可 证
Diversitree	ML	Unix、 Mac、Win	NEXUS	定性和定量性状、 地理	C、D	GNU 通用公共许可 证, 版本 2
HyPhy	ML	Unix、 Mac、Win	MEGA、NEXUS、 FASTA、PHYLIP	核苷酸、蛋白质(可 定制)	D	GNU 自由文档许可 证 1.3
BayesTraits	贝叶斯	Unix、 Mac、Win	TSV 或空格分隔的表。行是 物种, 并且列是性状	定性和定量性状	C、D	知识共享署名许可 证
Lagrange	ML	Unix、 Mac、Win	物种区域的 TSV/CSV。行 是物种, 并且列是地理区域	地理	-	GNU 通用公共许可 证, 版本 2
Mesquite	简约法、 ML	Unix、 Mac、Win	Fasta、NBRF、 Genbank、PHYLIP、 CLUSTAL、TSV	核苷酸、蛋白质、 地理	C、D	知识共享署名 3.0 许 可证
Phylomapper	ML、贝叶斯	Unix、	NEXUS	地理、生态位	C、D	-

[0199]

	(版本 2)	Mac、Win				
Ancestors	ML	Web	Fasta	核苷酸(插入缺失)	D	-
Phyrex	最大简约法	Linux	Fasta	基因表达	C、D	所有权
SIMMAP	SM	Mac	类似于 XML 的格式	核苷酸、定性性状	D	所有权
MrBayes	贝叶斯	Unix、 Mac、Win	NEXUS	核苷酸、蛋白质	D	GNU 通用公共许可证
PARANA	最大简约法	Unix、 Mac、Win	Newick	生物网络	D	Apache 许可证
PHAST (PREQUEL)	ML	Unix、 Mac、Win	多比对	核苷酸	D	BSD 许可证
RASP	ML、贝叶斯	Unix、 Mac、Win	Newick	地理	D	-
VIP	最大简约法	Linux、Win	Newick	地理	D (网络)	GPL 知识共享
FastML	ML	Web、Unix	Fasta	核苷酸、蛋白质	D	版权
MLGO	ML	Web	自定义	基因顺序排列	D	GNU
BADGER	贝叶斯	Unix、 Mac、Win	自定义	基因顺序排列	D	GNU GPL 版本 2
COUNT	最大简约法	Unix、 Mac、Win	制表符分隔的文本文件，行为分类单元，列中为计数数据	计数数据 (同源物家族大小)	D	BSD
MEGA	最大简约法、ML	Mac、Win	MEGA	核苷酸、蛋白质	D	所有权
ANGES	局部、简约法	Unix	自定义	基因组图谱	D	GNU 通用公共许可证, 版本 3
EREM	ML	Win、 Unix、 Matlab 模块	模型参数、树、观察到的特征值的自定义文本格式	二元	D	未指定, 尽管站点指示该软件可免费获得

[0200] 这些软件包中的大多数设计用于分析遗传序列数据。例如, PAML (Yang Z. PAML 4: phylogenetic analysis by maximum likelihood. Molecular biology and evolution. 2007; 24 (8) : 1586-91) 是用于通过 ML 对 DNA 和蛋白质序列比对进行系统发育分

析的程序的集合。祖先重建可使用codeml程序进行。HyPhy、Mesquite和MEGA也是用于序列数据系统发育分析的软件包,但它们被设计地具有更高模块化和可定制性。HyPhy (Pond SLK, Muse SV. HyPhy: hypothesis testing using phylogenies. *Statistical methods in molecular evolution*: Springer; 2005. p. 125-81) 执行祖先序列重建的联合ML法 (Pupko T, Pe I, Shamir R, Graur D. A fast algorithm for joint reconstruction of ancestral amino acid sequences. *Molecular Biology and Evolution*. 2000; 17 (6) : 890-6), 其可通过以其批处理语言指定自定义模型, 容易地适于重建更广泛范围的离散祖先特征状态, 例如地理位置。Mesquite (Maddison W, Maddison D. Mesquite: a modular system for evolutionary analysis. 2.75 ed 2011) 提供了使用最大简约法和ML法二者对于离散和连续特征二者的祖先状态重建的方法。其还提供了多种可视化工具来解释祖先重建的结果。MEGA (Tamura K, Dudley J, Nei M, Kumar S. MEGA4: molecular evolutionary genetics analysis (MEGA) software version 4.0. *Molecular biology and evolution*. 2007; 24 (8) : 1596-9) 也是一个模块化系统, 但比自定义分析更注重易用性。从版本5开始, MEGA允许用户使用最大简约法、ML和经验贝叶斯法来重建祖先状态。

[0201] 遗传序列的贝叶斯分析可赋予模型错误指定更大的稳健性。MrBayes (Huelsenbeck JP, Ronquist F. MRBAYES: Bayesian inference of phylogenetic trees. *Bioinformatics*. 2001; 17 (8) : 754-5) 允许使用完全分层贝叶斯法推断祖先节点的祖先状态。PHAST软件包中分发的PREQUEL程序使用祖先序列重建来执行比较进化基因组学 (Hubisz MJ, Pollard KS, Siepel A. PHAST and RPHAST: phylogenetic analysis with space/time models. *Briefings in bioinformatics*. 2011; 12 (1) : 41-51)。SIMMAP随机地将突变映射到系统发育上 (Bollback JP. SIMMAP: stochastic character mapping of discrete traits on phylogenies. *BMC bioinformatics*. 2006; 7 (1) : 88)。BayesTraits (Pagel M. The maximum likelihood approach to reconstructing ancestral character states of discrete characters on phylogenies. *Systematic biology*. 1999; 48 (3) : 612-22) 分析贝叶斯框架中的离散或连续特征, 以评价进化模型, 重建模型祖先状态, 并检测性状对之间的相关进化。

[0202] 另一些软件包更面向定性和定量性状 (表型) 的分析。例如, 统计计算环境R中的ape包 (Paradis E. *Analysis of phylogenetics and evolution with R*. New York: Springer; 2006) 还提供了通过ace函数 (包括ML) 对离散和连续特征二者进行祖先状态重建的方法。请注意, ace执行通过计算缩放的 (scaled) 条件似然而不是其他基于ML的方法进行祖先重建所使用的边际似然或联合似然来进行重建, 这可能会对根以外节点处的重建准确性产生不利影响。除了用于重建祖先遗传序列的ML法 (通过包装PAML中的baseml函数) 之外, Phyrex还实现了基于最大简约法的算法来重建祖先基因表达谱 (Rossnes R, Eidhammer I, Liberles DA. Phylogenetic reconstruction of ancestral character states for gene expression and mRNA splicing data. *BMC bioinformatics*. 2005; 6 (1) : 127)。

[0203] 多种软件包还重建系统发生生物地理学 (phylogeography)。BEAST (Bayesian Evolutionary Analysis by Sampling Trees (基于抽样树的贝叶斯进化分析) (Bouckaert R, Heled J, Kühnert D, Vaughan T, Wu C-H, Xie D, et al. BEAST 2: a software platform for Bayesian evolutionary analysis. *PLoS Comput Biol*. 2014; 10 (4) : e1003537)) 提供

了使用贝叶斯MCMC采样方法从注释有位置数据的观测序列中重建祖先地理位置的工具。Diversitree (FitzJohn RG. Diversitree: comparative phylogenetic analyses of diversification in R. *Methods in Ecology and Evolution*. 2012; 3 (6) : 1084-92) 是一个R包, 其提供了在Mk2 (二元特征进化的连续时间马尔可夫模型 (Pagel M. Detecting Correlated Evolution on Phylogenies—a General-Method for the Comparative-Analysis of Discrete Characters. *Proceedings of the Royal Society of London Series B-Biological Sciences*. 1994; 255 (1342) : 37-45)) 和BiSSE模型下进行祖先状态重建的方法。Lagrange在系统发育树上对地理范围进化的重建进行分析 (Ree RH, Smith SA. Maximum likelihood inference of geographic range evolution by dispersal, local extinction, and cladogenesis. *Systematic Biology*. 2008; 57 (1) : 4-14)。Phylomapper (Lemmon AR, Lemmon EM. A likelihood framework for estimating phylogeographic history on a continuous landscape. *Systematic Biology*. 2008; 57 (4) : 544-61) 是一种统计框架, 用于估算基因流和祖先地理位置的历史模式。RASP (Yu Y, Harris AJ, Blair C, He X. RASP (Reconstruct Ancestral State in Phylogenies): a tool for historical biogeography. *Molecular Phylogenetics and Evolution*. 2015; 87: 46-9) 使用统计DIVA、Lagrange、Bayes-Lagrange、BayArea和BBM法推断祖先状态。VIP (Arias JS, Szumik CA, Goloboff PA. Spatial analysis of vicariance: a method for using direct geographical information in historical biogeography. *Cladistics*. 2011; 27 (6) : 617-28) 通过检查不连续的地理分布来推断历史生物地理学。

[0204] 基因组重排为物种之间的比较基因组学提供了有价值的信息。ANGES (Jones BR, Rajaraman A, Tannier E, Chauve C. ANGES: reconstructing ANcestral GENome S maps. *Bioinformatics*. 2012; 28 (18) : 2388-90) 通过遗传标志物的祖先重建比较了现存相关基因组。BADGER (Larget B, Kadane JB, Simon DL. A Bayesian approach to the estimation of ancestral genome arrangements. *Molecular phylogenetics and evolution*. 2005; 36 (2) : 214-23) 使用贝叶斯法检查基因重排的历史。Count (Csűös M. Count: evolutionary analysis of phylogenetic profiles with parsimony and likelihood. *Bioinformatics*. 2010; 26 (15) : 1910-2) 重建了基因家族大小的进化。EREM (Afire L, Thompson JD, Debussche M. Genetic structure of continental and island populations of the Mediterranean endemic *Cyclamen balearicum* (Primulaceae). *American Journal of Botany*. 1997; 84 (4) : 437-51) 分析了由二元特征编码的遗传特征的获得和丢失。PARANA (Patro R, Sefer E, Malin J, Marçais G, Navlakha S, Kingsford C. Parsimonious reconstruction of network evolution. *Algorithms for Molecular Biology*. 2012; 7 (1) : 1) 对代表基因丢失和重复的祖先生物网络进行基于简约法的推断。

[0205] 还有一些基于Web服务器的应用程序, 其允许研究人员使用ML法进行不同特征类型的祖先重建, 而无需安装任何软件。例如, Ancestors (Diallo AB, Makarenkov V, Blanchette M. Ancestors 1.0: a web server for ancestral sequence reconstruction. *Bioinformatics*. 2010; 26 (1) : 130-1) 是用于通过识别和排列同义区域来重建祖先基因组的Web服务器。FastML (Ashkenazy H, Penn O, Doron-Faigenboim A, Cohen O, Cannarozzi

G,Zomer O,et al.FastML:a web server for probabilistic reconstruction of ancestral sequences.Nucleic acids research.2012;40 (W1) :W580-W4) 是用于通过ML对祖先序列进行概率重建的Web服务器,其使用间隙特征模型来重建插入缺失变异。MLG0 (Hu F,Lin Y,Tang J.MLG0:phylogeny reconstruction and ancestral inference from gene-order data.BMC bioinformatics.2014;15 (1) :1) 是用于ML基因顺序分析的Web服务器。

[0206] 多肽文库中的候选经优化抗原性病原体多肽可包含已经通过ARS鉴定出的氨基酸序列的一个或更多个区域。任选地,对于候选经优化抗原性病原体多肽,祖先氨基酸序列的所述或每个区域的长度为至少1、2、3、5、10、15、20、25、30、35、40、45或50个氨基酸残基。任选地,对于候选经优化抗原性病原体多肽,祖先氨基酸序列的所述或每个区域的长度为多至5、10、15、20、25、30、40、50、100、150、200、250、300、350、400、450、或500、600、700或800个氨基酸残基。

[0207] 任选地,多肽文库中的候选经优化抗原性病原体多肽包含氨基酸序列,其在其整个长度上与从中优化所述候选经优化抗原性病原体多肽的一个或更多个不同分离株的病原体多肽的氨基酸序列具有至少20%、30%、40%、50%、60%、70%、80%、90%、95%、96%、97%、98%或99%氨基酸同一性。

[0208] 任选地,本发明的方法包括对不同的所产生的核苷酸序列的密码子进行优化,以在表达系统中最佳表达所编码的候选经优化抗原性病原体多肽。密码子优化利用了遗传密码的简并性,并且不会改变编码多肽的氨基酸序列。由于简并性,一种蛋白质可由许多替代核酸序列编码。每个生物体中的密码子偏好性(密码子使用偏倚)不同,并且这会给在异源表达系统中表达重组蛋白带来挑战,导致表达水平低下和不可靠。

[0209] 可使用任何合适的表达系统。多个合适的实例是技术人员公知的,包括在哺乳动物、酵母、昆虫或细菌细胞中表达。任选地,表达系统包含哺乳动物细胞。任选地,表达系统包含酵母、昆虫或细菌细胞。

[0210] 密码子优化方法是本领域普通技术人员公知的。密码子优化算法可用于设计编码氨基酸序列的密码子优化核苷酸序列。这样的算法旨在提供使多肽或蛋白质在期望表达系统中的表达最大化的密码子优化序列。合适的密码子优化算法的实例包括GeneOptimizerTM算法(ThermoFisher)、OptimumGeneTM算法(GenScript)和GeneGPS[®](ATUM)。

[0211] 任选地,本发明的方法还包括其他序列优化以使期望表达系统中的蛋白质表达最大化。这样的基因优化考虑了密码子使用偏倚以及基因表达(例如转录、剪接、翻译和mRNA降解)中涉及的其他序列相关参数。这样的序列相关参数的实例如下(这些参数在以下分类为影响转录效率、翻译效率或蛋白质重折叠,但是多种参数可能影响这些步骤中的多于一个步骤):

[0212] 转录效力:

[0213] • GC含量

• SD序列

[0214] • CpG二核苷酸含量

• TATA盒

[0215] • 隐蔽剪接位点

• 终止信号

[0216] • 负CpG岛

• 人工重组位点

[0217] 翻译效率:

- [0218] • 密码子使用偏倚 • RNA不稳定基序 (ARE)
- [0219] • GC含量 • mRNA稳定的自由能
- [0220] • mRNA二级结构 • 内部chi位点和核糖体结合位点
- [0221] • 过早的PolyA位点 • 重复序列
- [0222] 蛋白质重折叠:
- [0223] • 密码子使用偏倚 • 密码子环境
- [0224] • 密码子和反密码子的相互作用 • RNA二级结构
- [0225] 基因优化算法例如GeneOptimizer™和OptimumGene™考虑了这些参数中的多种。
- [0226] Maertens等(Protein Science 2010 Vol.19:1312-1326)讨论了用于人蛋白质在大肠杆菌中表达的基因优化。
- [0227] 任选地,本发明的方法包括针对编码的候选物经优化抗原性病原体多肽的抗原性来对不同核苷酸序列进行优化。
- [0228] 抗原优化可能包括以下任何一项:
- [0229] (a) 编码被认为抑制抗病原体多肽抗体的产生和/或功能的氨基酸序列的核酸序列的缺失或修饰(例如,黏蛋白样结构域的缺失或修饰-参见Reynard et al., Journal of Virology, 2009, 9596-9601);
- [0230] (b) 区域交换,以恢复一个或更多个潜在丢失的编码表位;
- [0231] (c) 位点特异性突变,例如N-连接的糖基化位点。通常将位点特异性突变设计为缺失N-连接的糖基化位点,尽管在某些情况下可能期望引入另外的位点,例如以掩盖引发非中和抗体的表位。对于流感病毒,已经证明糖基化在空间上阻断抗体与HA结合并由此提供针对宿主免疫应答的保护作用的能力。Sun等(Journal of Virology, 2013, 87(15):8756-8766)证明,通过具有较高糖基化位点数目的病毒诱导的抗体比通过具有较少糖基化位点的病毒诱导的抗体具有更宽的中和活性;
- [0232] (d) 增强稳定性的改变(例如二硫键形成,减少丝氨酸蛋白酶对编码的多肽的降解);
- [0233] (e) 聚糖的去除(改善B细胞的进入);
- [0234] (f) 核酸序列的插入,例如插入编码期望表位的核酸序列。
- [0235] Joyce等(J Virol. 2013 Feb; 87(4):2294-306)描述了HIV-1 gp120的外部结构域的抗原优化。
- [0236] 任选地,不同病原体分离株包括来自这样的病原体之爆发的不同病原体分离株,所述病原体与期望诱导对其广泛中和免疫应答的病原体为同一亚型。
- [0237] 任选地,不同病原体分离株包括来自这样的病原体之爆发的不同病原体分离株,所述病原体与期望诱导对其广泛中和免疫应答的病原体为不同亚型但同一类型。
- [0238] 任选地,不同病原体分离株包括来自这样的病原体之爆发的不同病原体分离株,所述病原体与期望诱导对其广泛中和免疫应答的病原体为不同类型但同一科。
- [0239] 任选地,不同病原体分离株包括这样的病原体的不同先前病原体分离株,所述病原体与期望诱导对其广泛中和免疫应答的病原体为同一亚型、类型或科。
- [0240] 任选地,不同病原体分离株包括这样的病原体的不同先前病原体分离株,所述病原体与期望诱导对其广泛中和免疫应答的病原体为同一种、属或科。

[0241] 任选地,用于鉴定能够诱导对病原体的广泛中和免疫应答的先导候选经优化抗原性病原体多肽的本发明方法是体外方法。

[0242] 根据本发明,还提供了鉴定编码能诱导对病原体的广泛中和免疫应答的经优化抗原性病原体多肽的核酸序列的方法,其包括:

[0243] i) 用核酸对人或非人动物进行免疫接种,所述核酸包含编码通过根据本发明的方法鉴定出的先导候选经优化抗原性病原体多肽的核酸序列;

[0244] ii) 确定在步骤(i)中的免疫接种之后在人或非人动物中是否诱导广泛中和免疫应答;以及

[0245] iii) 如果从步骤(ii)确定在人或非人动物中诱导了广泛中和免疫应答,则将所述核酸序列鉴定为编码能够诱导对病原体的广泛中和免疫应答的经优化抗原性病原体多肽的核酸序列。

[0246] 任选地,通过确定获自人或非人动物的血清中的抗体是否与和期望对其广泛中和免疫应答的病原体同一科内的多于一种病原体亚型结合来确定在人或非人动物中是否诱导广泛中和免疫应答。

[0247] 任选地,通过确定获自人或非人动物的血清中的抗体是否与和期望对其广泛中和免疫应答的病原体同一科内的多于一种病原体类型结合来确定在人或非人动物中是否诱导广泛中和免疫应答。

[0248] 可使用任何合适的非人动物。任选地,非人动物是哺乳动物。任选地,哺乳动物是豚鼠或小鼠。任选地,非人动物是禽类。

[0249] 根据本发明还提供了分离的核酸分子,所述分离的核酸分子包含核酸序列或其互补序列,所述核酸序列:

[0250] i) 与SEQ ID NO:1具有至少75%、80%、85%、90%、95%、96%、97%、98%或99%同一性,或与SEQ ID NO:1相同;

[0251] ii) 与SEQ ID NO:2具有至少75%、80%、85%、90%、95%、96%、97%、98%或99%同一性,或与SEQ ID NO:2相同;

[0252] iii) 与SEQ ID NO:4具有至少75%、80%、85%、90%、95%、96%、97%、98%或99%同一性,或与SEQ ID NO:4相同;

[0253] iv) 与SEQ ID NO:5具有至少75%、80%、85%、90%、95%、96%、97%、98%或99%同一性,或与SEQ ID NO:5相同;

[0254] v) 与SEQ ID NO:7具有至少75%、80%、85%、90%、95%、96%、97%、98%或99%同一性,或与SEQ ID NO:7相同;或者

[0255] vi) 与SEQ ID NO:8具有至少75%、80%、85%、90%、95%、96%、97%、98%或99%同一性,或与SEQ ID NO:8相同。

[0256] 根据本发明还提供了分离的核酸分子,所述分离的核酸分子包含核酸序列或其互补序列,所述核酸序列:

[0257] i) 与SEQ ID NO:10具有至少75%、80%、85%、90%、95%、96%、97%、98%或99%同一性,或与SEQ ID NO:10相同;

[0258] ii) 与SEQ ID NO:12具有至少75%、80%、85%、90%、95%、96%、97%、98%或99%同一性,或与SEQ ID NO:12相同;或者

[0259] iii) 与SEQ ID NO:14具有至少75%、80%、85%、90%、95%、96%、97%、98%或99%同一性,或与SEQ ID NO:14相同。

[0260] 根据本发明还提供了分离的核酸分子,所述分离的核酸分子包含核酸序列或其互补序列,所述核酸序列:

[0261] i) 与SEQ ID NO:19具有至少75%、80%、85%、90%、95%、96%、97%、98%或99%同一性,或与SEQ ID NO:19相同;

[0262] ii) 与SEQ ID NO:21具有至少75%、80%、85%、90%、95%、96%、97%、98%或99%同一性,或与SEQ ID NO:21相同;

[0263] iii) 与SEQ ID NO:23具有至少75%、80%、85%、90%、95%、96%、97%、98%或99%同一性,或与SEQ ID NO:23相同;

[0264] iv) 与SEQ ID NO:25具有至少75%、80%、85%、90%、95%、96%、97%、98%或99%同一性,或与SEQ ID NO:25相同;

[0265] v) 与SEQ ID NO:27具有至少75%、80%、85%、90%、95%、96%、97%、98%或99%同一性,或与SEQ ID NO:27相同;

[0266] vi) 与SEQ ID NO:29具有至少75%、80%、85%、90%、95%、96%、97%、98%或99%同一性,或与SEQ ID NO:29相同;或者

[0267] vii) 与SEQ ID NO:31具有至少75%、80%、85%、90%、95%、96%、97%、98%或99%同一性,或与SEQ ID NO:31相同。

[0268] 根据本发明还提供了分离的多肽,其包含氨基酸序列,所述氨基酸序列:

[0269] i) 与SEQ ID NO:1编码的氨基酸序列具有至少95%、96%、97%、98%或99%同一性,或与SEQ ID NO:1编码的氨基酸序列相同;

[0270] ii) 与SEQ ID NO:2编码的氨基酸序列具有至少95%、96%、97%、98%或99%同一性,或与SEQ ID NO:2编码的氨基酸序列相同;

[0271] iii) 与SEQ ID NO:4编码的氨基酸序列具有至少95%、96%、97%、98%或99%同一性,或与SEQ ID NO:4编码的氨基酸序列相同;

[0272] iv) 与SEQ ID NO:5编码的氨基酸序列具有至少95%、96%、97%、98%或99%同一性,或与SEQ ID NO:5编码的氨基酸序列相同;

[0273] v) 与SEQ ID NO:7编码的氨基酸序列具有至少95%、96%、97%、98%或99%同一性,或与SEQ ID NO:7编码的氨基酸序列相同;

[0274] vi) 与SEQ ID NO:8编码的氨基酸序列具有至少95%、96%、97%、98%或99%同一性,或与SEQ ID NO:8编码的氨基酸序列相同;

[0275] vii) 与SEQ ID NO:10编码的氨基酸序列具有至少95%、96%、97%、98%或99%同一性,或与SEQ ID NO:10编码的氨基酸序列相同;

[0276] viii) 与SEQ ID NO:12编码的氨基酸序列具有至少95%、96%、97%、98%或99%同一性,或与SEQ ID NO:12编码的氨基酸序列相同;或者

[0277] ix) 与SEQ ID NO:14编码的氨基酸序列具有至少95%、96%、97%、98%或99%同一性,或与SEQ ID NO:14编码的氨基酸序列相同。

[0278] 根据本发明还提供了分离的多肽,其包含氨基酸序列,所述氨基酸序列:

[0279] i) 与SEQ ID NO:3具有至少95%、96%、97%、98%或99%同一性,或与SEQ ID NO:

3相同;

[0280] ii) 与SEQ ID NO:6具有至少95%、96%、97%、98%或99%同一性,或与SEQ ID NO:6相同;

[0281] iii) 与SEQ ID NO:9具有至少95%、96%、97%、98%或99%同一性,或与SEQ ID NO:9相同;

[0282] iv) 与SEQ ID NO:11具有至少95%、96%、97%、98%或99%同一性,或与SEQ ID NO:11相同;

[0283] v) 与SEQ ID NO:13具有至少95%、96%、97%、98%或99%同一性,或与SEQ ID NO:13相同;或者

[0284] vi) 与SEQ ID NO:15具有至少95%、96%、97%、98%或99%同一性,或与SEQ ID NO:15相同。

[0285] 根据本发明还提供了分离的多肽,其包含氨基酸序列,所述氨基酸序列:

[0286] i) 与SEQ ID NO:18具有至少95%、96%、97%、98%或99%同一性,或与SEQ ID NO:18相同;

[0287] ii) 与SEQ ID NO:20具有至少95%、96%、97%、98%或99%同一性,或与SEQ ID NO:20相同;

[0288] iii) 与SEQ ID NO:22具有至少95%、96%、97%、98%或99%同一性,或与SEQ ID NO:22相同;

[0289] iv) 与SEQ ID NO:24具有至少95%、96%、97%、98%或99%同一性,或与SEQ ID NO:24相同;

[0290] v) 与SEQ ID NO:26具有至少95%、96%、97%、98%或99%同一性,或与SEQ ID NO:26相同;

[0291] vi) 与SEQ ID NO:28具有至少95%、96%、97%、98%或99%同一性,或与SEQ ID NO:28相同;或者

[0292] vii) 与SEQ ID NO:30具有至少95%、96%、97%、98%或99%同一性,或与SEQ ID NO:30相同。

[0293] 氨基酸或核酸序列之间的相似性以序列之间的相似性表示,也称为序列同一性。序列同一性经常根据百分比同一性(或相似性或同源性)来衡量;百分比越高,两个序列越相似。当使用标准方法比对时,给定基因或蛋白质的同源物或变体将具有相对较高程度的序列同一性。用于比较的序列比对方法是本领域公知的。多种程序和比对算法描述在以下中:Smith and Waterman,Adv.Appl.Math.2:482,1981;Needleman and Wunsch,J.Mol.Biol.48:443,1970;Pearson and Lipman,Proc.Natl.Acad.Sci.U.S.A.85:2444,1988;Higgins and Sharp,Gene 73:237-244,1988;Higgins and Sharp,CABIOS 5:151-153,1989;Corpet et al.,Nucleic Acids' Research 16:10881-10890,1988;and Pearson and Lipman,Proc.Natl.Acad.Sci.U.S.A.85:2444,1988.Altschul et al.,Nature Genet.6:119-129,1994.NCBI基本局部比对搜索工具(NCBI Basic Local Alignment Search Tool)(BLASTM)(Altschul et al.,J.Mol.Biol.215:403-410,1990)可从多种来源获得,包括美国国家生物技术信息中心(NCBI,Bethesda,MD)和互联网上,以与序列分析程序blastp、blastn、blastx、tblastn和tblastx结合使用。

[0294] 核酸序列之间或氨基酸序列之间的序列同一性可通过比较序列的比对来确定。当所比较序列中的等效位置被相同的核苷酸或氨基酸占据时,则分子在该位置是相同的。将比对评分为同一性百分比是所比较序列共有的位置上相同核苷酸或氨基酸数目的函数。当比较序列时,最佳比对可能需要将空位引入到一个或多个序列中,以考虑序列中可能的插入和缺失。序列比较方法可能会采用空位罚分,以便对于所比较的序列中相同数目的相同分子,具有尽可能少的空位(反映两个比较序列之间更高关联性)的序列比对将比具有很多空位的序列比对获得更高的评分。最大百分比同一性的计算涉及在考虑空位罚分的情况下产生最佳比对。

[0295] 用于进行序列比较的合适的计算机程序在商业和公共部门中广泛可获得。实例包括MatGat(Campanella et al.,2003,BMC Bioinformatics 4:29;程序可获自<http://bitincka.com/ledion/matgat>)、Gap(Needleman&Wunsch,1970,J.Mol.Biol.48:443-453)、FASTA(Altschul et al.,1990,J.Mol.Biol.215:403-410;程序可获自<http://www.ebi.ac.uk/fasta>)、Clustal W 2.0和X 2.0(Larkin et al.,2007,Bioinformatics 23:2947-2948;程序可获自<http://www.ebi.ac.uk/tools/clustalw2>)和EMBOSS成对比对算法(Needleman&Wunsch,1970,同上;Kruskal,1983,在Time warps,string edits and macromolecules:the theory and practice of sequence comparison中,Sankoff&Kruskal(eds),pp 1-44,Addison Wesley;程序可获自<http://www.ebi.ac.uk/tools/emboss/align>)。所有程序都可使用默认参数运行。

[0296] 例如,可使用EMBOSS成对比对算法的“needle”法进行序列比较,该方法在两个序列的全长上考虑时确定两个序列的最佳比对(包括空位)并提供同一性百分比评分。氨基酸序列比较的默认参数(“Protein Molecule”选项)可以是:空位延伸罚分:0.5,空位开放罚分:10.0,矩阵:Blosum 62。

[0297] 可在参考序列的全长上进行序列比较。

[0298] 根据本发明,还提供了分离的核酸分子,其包含编码包含SEQ ID NO:6氨基酸序列的多肽和包含SEQ ID NO:9氨基酸序列的多肽的核苷酸序列。

[0299] 根据本发明,还提供了分离的核酸分子,其包含编码包含SEQ ID NO:13氨基酸序列的多肽和包含SEQ ID NO:15氨基酸序列的多肽的核苷酸序列。

[0300] 根据本发明,还提供了组合物,其包含:第一核酸,所述第一核酸包含编码包含SEQ ID NO:6氨基酸序列的多肽的核苷酸序列;以及第二核酸,所述第二核酸包含编码包含SEQ ID NO:9氨基酸序列的多肽的核苷酸序列。

[0301] 根据本发明,还提供了组合物,其包含:第一核酸,所述第一核酸包含编码包含SEQ ID NO:13氨基酸序列的多肽的核苷酸序列;以及第二核酸,所述第二核酸包含编码包含SEQ ID NO:15氨基酸序列的多肽的核苷酸序列。

[0302] 根据本发明,还提供了组合制剂,其包含:(i)第一核酸,所述第一核酸包含编码包含SEQ ID NO:6氨基酸序列的多肽的核苷酸序列;以及(ii)第二核酸,所述第二核酸包含编码包含SEQ ID NO:9氨基酸序列的多肽的核苷酸序列。

[0303] 根据本发明,还提供了组合制剂,其包含:(i)第一核酸,所述第一核酸包含编码包含SEQ ID NO:13氨基酸序列的多肽的核苷酸序列;以及(ii)第二核酸,所述第二核酸包含编码包含SEQ ID NO:15氨基酸序列的多肽的核苷酸序列。

[0304] 根据本发明,还提供了组合物,其包含:包含SEQ ID NO:6的氨基酸序列的第一多肽和包含SEQ ID NO:9的氨基酸序列的第二多肽。

[0305] 根据本发明,还提供了组合物,其包含:包含SEQ ID NO:13的氨基酸序列的第一多肽和包含SEQ ID NO:15的氨基酸序列的第二多肽。

[0306] 根据本发明,还提供了融合蛋白,其包含:包含SEQ ID NO:6的氨基酸序列的第一多肽和包含SEQ ID NO:9的氨基酸序列的第二多肽。

[0307] 根据本发明,还提供了融合蛋白,其包含:包含SEQ ID NO:13的氨基酸序列的第一多肽和包含SEQ ID NO:15的氨基酸序列的第二多肽。

[0308] 根据本发明,还提供了组合制剂,其包含:(i) 包含SEQ ID NO:6的氨基酸序列的第一多肽;以及(ii) 包含SEQ ID NO:9的氨基酸序列的第二多肽。

[0309] 根据本发明,还提供了组合制剂,其包含:(i) 包含SEQ ID NO:13的氨基酸序列的第一多肽;以及(ii) 包含SEQ ID NO:15的氨基酸序列的第二多肽。

[0310] 如本文所用,术语“组合制剂”是指以下含义中的“套件(kit of parts)”:如上所定义的组合组分(i)和(ii)可独立地给药或通过使用具有独特量的组合组分(i)和(ii)的不同固定组合来给药。所述组分可同时施用或者一种接一种地施用。如果将组分一种接一种地施用,则优选地选择施用之间的时间间隔,以使组合使用这些组分的治疗效果大于仅使用组合组分(i)和(ii)中的任一种所获得的效果。

[0311] 组合制剂的组分可以以一种组合的单位剂型存在,或作为组分(i)的第一单位剂型和组分(ii)的单独的第二单位剂型存在。可改变在组合制剂中施用的组合组分(i)与组合组分(ii)的总量之比,例如,以应对待治疗的患者亚群的需求或单个患者的需求,这可能是由于例如患者的特定疾病、年龄、性别或体重。

[0312] 优选地,存在至少一种有益作用,例如增强组分(i)或组分(ii)的作用,或共同增强组合组分(i)和(ii)的作用,例如与组合组分(i)和(ii)中的一种或两种的有效剂量相比,超过累加作用,另外的有益效果,更少的副作用,更少的毒性或组合的治疗效果,并且非常优选地,组合组分(i)和(ii)的协同作用。

[0313] 本发明的组合制剂可以以用于向哺乳动物,优选人施用的药物组合制剂提供。组分(i)可任选地与可药用载体、赋形剂或稀释剂一起提供,和/或组分(ii)可任选地与可药用载体、赋形剂或稀释剂一起提供。

[0314] 根据本发明,还提供了编码由本发明的核酸编码的氨基酸序列的分离的核酸分子。

[0315] 根据本发明,还提供了编码由本发明的核酸编码的氨基酸序列的分离的核酸分子,其中所述核酸针对在哺乳动物细胞中的表达进行了密码子优化。

[0316] 根据本发明,还提供了编码由本发明的核酸编码的氨基酸序列的分离的核酸分子,其中所述核酸针对在哺乳动物细胞中的表达进行了基因优化。

[0317] 根据本发明,还提供了编码本发明多肽的分离的核酸分子。

[0318] 根据本发明,还提供了编码本发明多肽的分离的核酸分子,其中所述核酸针对在哺乳动物细胞中的表达进行了密码子优化。

[0319] 根据本发明,还提供了编码本发明多肽的分离的核酸分子,其中所述核酸针对在哺乳动物细胞中的表达进行了基因优化。

- [0320] 根据本发明,还提供了包含本发明的核酸的载体。
- [0321] 任选地,载体还包含与核酸可操作地连接的启动子。
- [0322] 任选地,启动子用于在哺乳动物细胞中表达由核酸编码的多肽。
- [0323] 任选地,启动子用于在酵母、细菌或昆虫细胞中表达由核酸编码的多肽。
- [0324] 任选地,载体是疫苗载体。任选地,疫苗载体是病毒疫苗载体、细菌疫苗载体或核酸载体(例如RNA疫苗载体或DNA疫苗载体)。
- [0325] 本发明的核酸分子可包含DNA或RNA分子。对于其中核酸分子包含RNA分子的实施方案,应当理解,分子可包含与SEQ ID NO:1、2、4、5、7、8、10、12、14、19、21、23、25、27、29或31中的任一个具有至少75%、80%、85%、90%、95%、96%、97%、98%或99%同一性或者相同并且其中每个“T”核苷酸都被“U”代替的RNA序列,或其互补序列。
- [0326] 例如,将理解的是,在提供包含本发明的核酸的RNA疫苗载体的情况下,本发明的核酸的核酸序列将是RNA序列,因此可包含例如与SEQ ID NO:1、2、4、5、7、8、10、12、14、19、21、23、25、27、29或31中的任一个具有至少75%、80%、85%、90%、95%、96%、97%、98%或99%同一性或者相同并且其中每个“T”核苷酸都被“U”代替的RNA序列,或其互补序列。
- [0327] 根据本发明,还提供了包含本发明的载体或用本发明的载体转染的分离的细胞。
- [0328] 根据本发明,还提供了包含本发明多肽的病毒假型颗粒。
- [0329] 根据本发明,还提供了产生病毒假型颗粒的方法,其包括用包含本发明的核酸的载体转染宿主细胞。
- [0330] 根据本发明,还提供了包含本发明的多肽的融合蛋白。
- [0331] 根据本发明,还提供了包含本发明的核酸以及可药用载体、赋形剂或稀释剂的药物组合物。
- [0332] 根据本发明,还提供了包含本发明的载体以及可药用载体、赋形剂或稀释剂的药物组合物。
- [0333] 根据本发明,还提供了包含本发明的多肽以及可药用载体、赋形剂或稀释剂的药物组合物。
- [0334] 任选地,本发明的药物组合物还包含佐剂,以用于增强对象中对组合物的多肽或由核酸编码的多肽的免疫应答。
- [0335] 根据本发明,还提供了在对象中诱导对病原体的免疫应答的方法,其包括向对象施用本发明的核酸、本发明的多肽、本发明的载体或本发明的药物组合物。
- [0336] 任选地,病原体是病毒。任选地,病毒是丝状病毒科、沙粒病毒科或正黏病毒科的成员。
- [0337] 根据本发明,还提供了在对象中诱导对丝状病毒科或沙粒病毒科病毒的免疫应答的方法,其包括向对象施用本发明的核酸、本发明的多肽、本发明的载体或本发明的药物组合物。
- [0338] 根据本发明,还提供了对对象进行针对病原体的免疫接种的方法,其包括向对象施用本发明的核酸、本发明的多肽、本发明的载体或本发明的药物组合物。
- [0339] 任选地,病原体是病毒。任选地,病毒是丝状病毒科、沙粒病毒科或正黏病毒科的成员。
- [0340] 根据本发明,还提供了对对象进行针对丝状病毒科病毒的免疫接种的方法,其包

括向对象施用本发明的核酸、本发明的多肽、本发明的载体或本发明的药物组合物。

[0341] 根据本发明,还提供了在对象中诱导对丝状病毒科病毒的免疫应答的方法,其包括向对象施用本发明的核酸、本发明的多肽、本发明的载体或本发明的药物组合物。

[0342] 任选地,本发明的核酸、载体或药物组合物包含核酸,所述核酸包含与SEQ ID NO: 1、2、4、5、7、8、10、12或14中的任一个具有至少75%、80%、85%、90%、95%、96%、97%、98%或99%同一性或者相同的序列,或者包含核酸,所述核酸编码由包含与SEQ ID NO:1、2、4、5、7、8、10、12或14中的任一个具有至少75%、80%、85%、90%、95%、96%、97%、98%或99%同一性或者相同的序列的核酸编码的氨基酸序列。

[0343] 任选地,本发明的多肽、载体或药物组合物包含多肽,所述多肽包含与由SEQ ID NO:1、2、4、5、7、8、10、12或14中的任一个编码的氨基酸序列具有至少95%、96%、97%、98%或99%同一性或者相同的氨基酸序列,或者包含多肽,所述多肽包含与SEQ ID NO:3、6、9、11、13或15中的任一个具有至少95%、96%、97%、98%或99%同一性或者相同的氨基酸序列。

[0344] 根据本发明,还提供了对对象进行针对沙粒病毒科病毒的免疫接种的方法,其包括向对象施用本发明的核酸、本发明的多肽、本发明的载体或本发明的药物组合物。

[0345] 根据本发明,还提供了在对象中诱导对沙粒病毒科病毒的免疫应答的方法,其包括向对象施用本发明的核酸、本发明的多肽、本发明的载体或本发明的药物组合物。

[0346] 任选地,本发明的核酸、载体或药物组合物包含核酸,所述核酸包含与SEQ ID NO: 19、21、23、25、27、29或31中的任一个具有至少75%、80%、85%、90%、95%、96%、97%、98%或99%同一性或者相同的序列,或者包含核酸,所述核酸编码由包含与SEQ ID NO:19、21、23、25、27、29或31中的任一个具有至少75%、80%、85%、90%、95%、96%、97%、98%或99%同一性或者相同的序列的核酸编码的氨基酸序列。

[0347] 任选地,本发明的多肽、载体或药物组合物包含多肽,所述多肽包含与由SEQ ID NO:19、21、23、25、27、29或31中的任一个编码的氨基酸序列具有至少95%、96%、97%、98%或99%同一性或者相同的氨基酸序列,或者包含多肽,所述多肽包含与SEQ ID NO:18、20、22、24、26、28或30中的任一个具有至少95%、96%、97%、98%或99%同一性或者相同的氨基酸序列。

[0348] 可使用任何合适的施用途径。施用方法包括但不限于皮内、肌内、腹膜内、肠胃外、静脉内、皮下、经阴道、经直肠、鼻内、吸入或经口。肠胃外施用(例如皮下、静脉内或肌内施用)通常通过注射来实现。注射剂可以以常规形式制备,作为液体溶液剂或混悬剂,或者适合于注射前的液体中的溶液剂或混悬剂的固体形式,或作为乳剂。注射溶液剂和混悬剂可由上述种类的无菌散剂、颗粒剂和片剂制备。施用可以是全身的或局部的。

[0349] 组合物可以以任何合适的方式施用,例如与可药用载体一起施用。可药用载体部分地由所施用的特定组合物以及用于施用组合物的特定方法来决定。用于肠胃外施用的制剂包括无菌水性或非水性溶液剂、混悬剂和乳剂。非水性溶剂的实例是丙二醇、聚乙二醇、植物油例如橄榄油,以及可注射有机酯例如油酸乙酯。水性载体包括水、醇/水性溶液剂、乳剂或混悬剂,包含盐水和缓冲介质。肠胃外载剂包括氯化钠溶液、林格氏右旋糖、右旋糖和氯化钠、乳酸林格氏液或不挥发性油。静脉内载剂包括流体和营养补充剂、电解质补充剂(例如基于林格氏右旋糖的补充剂)等。也可存在防腐剂和其他添加剂,例如抗微生物剂、

抗氧化剂、螯合剂和惰性气体等。

[0350] 一些组合物可以以可药用酸或碱加成盐施用,所述盐是通过与无机酸(例如盐酸、氢溴酸、高氯酸、硝酸、硫氰酸、硫酸和磷酸)和有机酸(例如甲酸、乙酸、丙酸、乙醇酸、乳酸、丙酮酸、草酸、丙二酸、琥珀酸、马来酸和富马酸)反应形成或通过与无机碱(例如氢氧化钠、氢氧化铵、氢氧化钾)和有机碱(例如单、二、三烷基和芳基胺和取代的乙醇胺)反应形成。

[0351] 施用可通过单剂量或多剂量来完成。在本公开内容的上下文中,向对象施用的剂量应足以随着时间在对象中诱导有益的治疗反应,或者抑制或预防感染。取决于对象的种类、年龄、体重和一般状况,所治疗的感染的严重程度,所使用的特定组合物及其施用方式,在对象之间所需的剂量将有所不同。适当的剂量可由本领域普通技术人员仅使用常规实验来确定。

[0352] 可药用载体包括但不限于盐水、缓冲盐水、右旋糖、水、甘油、乙醇及其组合。载体和组合物可以是无菌的,并且制剂适合于施用方式。组合物还可包含少量的润湿剂或乳化剂或pH缓冲剂。组合物可以是液体溶液剂、混悬剂、乳剂、片剂、丸剂、胶囊剂、缓释制剂或散剂。组合物可用传统的黏合剂和载体(例如甘油三酸酯)配制成栓剂。经口制剂可包含标准载体,例如药用级的甘露醇、乳糖、淀粉、硬脂酸镁、糖精钠、纤维素和碳酸镁。可使用任何常见的药用载体,例如无菌盐溶液或芝麻油。介质还可包含常规的药物辅助材料,例如如调节渗透压的可药用盐、缓冲剂、防腐剂等。可与本文提供的组合物和方法一起使用的其他介质是生理盐水和芝麻油。

[0353] 在一些实施方案中,组合物包含可药用载体和/或佐剂。例如,佐剂可以是明矾、弗氏完全佐剂、生物佐剂或免疫刺激性寡核苷酸(例如CpG寡核苷酸)。

[0354] 可用于本公开内容中的可药用载体(载剂)是常规的。E.W.Martin的Remington's Pharmaceutical Sciences,Mack Publishing Co.,Easton,PA,15thEdition(1975)描述了适用于药物递送一种或更多种治疗组合物(例如一种或更多种流感疫苗和另外的药剂)的组合物和制剂。

[0355] 通常来说,载体的性质将取决于所采用的特定施用方式。例如,肠胃外制剂通常包含注射液,所述注射液包含可药用和生理学上可接受的流体例如水、生理盐水、平衡盐溶液、右旋糖水溶液、甘油等作为载剂。对于固体组合物(例如散剂、丸剂、片剂或胶囊剂形式),常规的无毒固体载体可包括例如药用级的甘露醇、乳糖、淀粉或硬脂酸镁。除生物中性载体外,待施用的药物组合物还可包含少量的无毒辅助物质,例如润湿剂或乳化剂、防腐剂和pH缓冲剂等,例如乙酸钠或山梨醇单月桂酸酯。

[0356] 任选地,本发明的组合物肌内施用。

[0357] 任选地,组合物通过针肌内、皮内、皮下施用,或者通过基因枪或电穿孔施用。

[0358] 根据本发明,还提供了核酸表达载体,其包含含有KpnI和NotI核酸内切酶位点的多克隆位点。

[0359] 任选地,多克隆位点包含SEQ ID NO:16的核酸序列。

[0360] 任选地,核酸表达载体是核酸表达载体和病毒假型载体。

[0361] 任选地,核酸表达载体是疫苗载体。

[0362] 任选地,核酸表达载体从5'至3'方向包含:启动子;剪接供体位点(SD);剪接受体位点(SA);和终止子信号,其中多克隆位点位于所述剪接受体位点和所述终止子信号

之间。

[0363] 任选地,启动子包含CMV即早期增强子/启动子(CMV-IE-E/P)和/或终止子信号包含缺少KpnI限制性核酸内切酶位点的牛生长激素基因(Tbgh)终止子信号。

[0364] 任选地,核酸表达载体还包含复制起点和编码对抗生素的抗性的核酸。任选地,复制起点包含pUC-质粒复制起点和/或核酸编码对卡那霉素的抗性。

[0365] 任选地,核酸表达载体包含SEQ ID NO:17的核酸序列(pEVAC)。

[0366] 本发明的多肽可包含一个或多个保守氨基酸置换。保守氨基酸置换是这些置换,其在进行时,最少地干扰原始蛋白质的性质,即,蛋白质的结构并且尤其是功能被保留并且不会因这样的置换而显著改变。保守置换的实例如下所示:

	<u>原始残基</u>	<u>保守置换</u>
	Ala	Ser
	Arg	Lys
	Asn	Gln、 His
[0367]	Asp	Glu
	Cys	Ser
	Gln	Asn
	Glu	Asp
	His	Asn; Gln
	Ile	Leu、 Val
	Leu	Ile; Val
	Lys	Arg; Gln;
	Met	Leu; Ile
[0368]	Phe	Met; Leu; Tyr
	Ser	Thr
	Thr	Ser
	Trp	Tyr
	Tyr	Trp; Phe
	Val	Ile; Leu

[0369] 保守置换通常维持 (a) 置换区域中多肽骨架的结构,例如片或螺旋构型, (b) 分子在靶位点的电荷或疏水性,或 (c) 侧链的体积。

[0370] 通常预期会产生蛋白质性质最大变化的置换将是非保守的,例如这样的变化:(a) 亲水性残基(例如丝氨酸或苏氨酸)置换疏水性残基(例如亮氨酸、异亮氨酸、苯丙氨酸、缬氨酸或丙氨酸)(或被其置换);(b) 半胱氨酸或脯氨酸置换任何其他残基(或被其置换);(c) 具有电正性侧链的残基(例如赖氨酸、精氨酸或组氨酸)置换电负性残基(例如谷氨酸或天冬氨酸)(或被其置换);或(d) 具有庞大侧链的残基(例如苯丙氨酸)置换没有侧链的残基(例如甘氨酸)(或被其置换)。

[0371] 在本发明的特定实施方案中,序列比对和祖先序列重建(ASR)用于鉴定病原体不能改变且在该病毒科的未来爆发中始终存在(即使是在大部分高度可变的RNA病毒中也是如此)的高度保守的免疫靶标。合成基因技术用于产生计算机生成的病毒基因,从而使其高度表达并可容易地克隆到表达载体,例如pEVAC载体(已被证明是用于产生病毒假型以及动物和/或人的直接DNA疫苗接种的高度通用的表达载体)中。

[0372] 可使用pEVAC载体产生基因的大组,以便快速产生病毒假型。这允许用一大组单克隆抗体探测病毒假型文库,每个病毒假型都有自己独特的病毒蛋白。该过程确保插入物产生构象正确的病毒表面蛋白,以向病毒要害(Achilles heel)提供最容易接近的靶标。通过假型形成和mAb结合对候选物的向下选择,提供了最佳候选基因的候选列表以通过疫苗接种测试。这可在豚鼠中完成,其中递送了pEVAC-疫苗插入物的精简过程。如果需要的话,这使得基于先进设计的便捷克隆位点,疫苗插入物能够从DNA pEVAC载体穿梭到多种病毒载体中。由于黑猩猩腺载体(ChAd)被广泛用于评价西非爆发中第一阶段的大多数埃博拉病毒疫苗候选物,因此我们选择进行比较以将同一载体用于人的面对面(head to head)比较。为了在豚鼠中进行筛选,我们使用了用pEVAC-疫苗插入物随后是ChAd-疫苗插入物进行的DNA引发。

[0373] 在本发明的特定实施方案中:

[0374] 1) 高通量“深度”测序技术提供了来自当前和过去爆发的病毒变异数据。通过分析该数据,可鉴定出可用作设计最佳疫苗插入物的骨架并保留了已知的B和T细胞表位的结构高度保守的区域。

[0375] 2) 人单克隆抗体(mAb)技术允许针对疫苗靶标(例如病毒包膜蛋白)产生抗病毒mAb,其识别广泛中和单克隆抗体(BNmAb)靶向的富表位区域。

[0376] 3) 最佳基因设计和合成结合了(1)中鉴定出的基因的数字建模保守骨架,以在这些骨架上包含最广泛的NmAb表位(BN表位可能不会在自然保守的GP中最佳展示)。

[0377] 4) 在设计和合成作为疫苗插入物的RNA和密码子优化合成基因时,考虑符合疫苗和假型载体要求的便捷克隆位点的下游知识,以使得能够快速高效的克隆和穿梭至不同筛选(即慢病毒假型;PV)载体和疫苗(即(MVA、ChAd、VSV、DNA等)载体)。

[0378] 5) 通过转导和感染研究在体外筛选由数字化设计的插入物产生的病毒假型(慢病毒)的功能性。除此之外,使用一组BNmAb和患者血清进行中和测定,以确保保留已知的表位。

[0379] 6) 使用快速DNA引发(并且如果需要的话)和腺病毒加强(产生高且可重现的滴度的方法)确认多种合成疫苗插入物到同类最佳(best-in-class)疫苗插入物的向下选择在豚鼠中的免疫原性。体内筛选证实了哪些具有最高的免疫原性和最大的中和广度。

[0380] 病毒糖蛋白在细胞附着、融合和脱壳中的核心作用使其成为用于病毒疫苗和单克

隆抗体治疗的关键抗原性靶标,这些疫苗和抗体在西非埃博拉爆发期间得到了开创。EBOV物种之间的GP序列分析显示了在核苷酸和氨基酸水平上的高度多样性(仅约60%至65%nt同一性)。对于当前的常规丝状病毒疫苗方法,靶向GP的疫苗必须是多价的,编码每个物种特异性的GP,这些GP更为保守(在GP核苷酸序列中约97%至98%的同一性)。尽管提出使用EBOV的较旧毒株的疫苗(rVSV.ZEBOV=Kikwit)可提供交叉保护作用(Henao-Restrepo AM, Lancet 2015),但人们担心这种疫苗对未来其他多种高度致病性丝状病毒的爆发的效力有限。

[0381] 我们可基于序列数据(任选地包括例如爆发序列数据)在针对新病毒变体的疫苗效力方面取得显著改善,从而产生合成的优化疫苗插入物,从而为未来的可变RNA病毒的爆发提供最广泛的可能的疫苗保护。在特定的实施方案中,我们的新疫苗技术组合了:

[0382] (1) 爆发病原体的序列

[0383] (2) 来自爆发幸存者的广泛抗病毒中和单克隆抗体(BNmAb)

[0384] (3) 计算建模方法

[0385] (4) 合成基因技术和抗原展示技术

[0386] (5) 高通量病毒结合和中和筛选

[0387] (6) 体内免疫选择和疫苗功效读数

[0388] 最终产品是用于触发最广泛的保护性免疫应答谱的新免疫原。我们已经提供了概念证明,即下一代单一疫苗插入物确实诱导对埃博拉病毒属(扎伊尔、苏丹、本迪布焦)的广泛中和谱,另外还靶向更远的丝状病毒马尔堡病毒。

[0389] 在下面的实施例中,参考附图仅以举例说明的方式描述本发明的实施方案,其中:

[0390] 图1示出了系统发育树及其与祖先序列重建的关系的图示;

[0391] 图2示出了比较埃博拉病毒和马尔堡病毒的系统发育树。数字表示分支的百分比置信度;

[0392] 图3示出了pEVAC的质粒图;

[0393] 图4示出了埃博拉攻击模型的攻击研究结果。埃博拉攻击模型对未经过疫苗接种的豚鼠(组1,下排)是致命的,而所有经疫苗接种的豚鼠(组2,上排)均受到保护(左)并继续增重(右);

[0394] 图5示出了假型病毒中和测定的结果,其举例说明了对代表所有埃博拉病毒物种和马尔堡病毒的假型病毒的表面上表达的靶抗原的中和抗体应答的强度。中和的强度由热图指示,其中红色(最深的阴影)是非常强的中和,从橙色到黄色(逐渐变浅的阴影)降低,并且没有中和/等于阴性对照值为白色。T2-4和T2-6是编码先导候选经优化抗原性埃博拉多肽的核酸疫苗,其在用取自经免疫豚鼠的血清样品进行测试的临床前阶段中与马尔堡候选T2-11组合;

[0395] 图6示出了确定编码使用本发明方法的实施方案鉴定出的不同先导候选经优化抗原性病原体多肽的核酸疫苗的有效性的研究结果。通过将其表面上带有两种不同的组1甲型流感糖蛋白(H1大流行和季节性)的两组细胞与合并的小鼠血清一起孵育来测量抗体结合。然后,通过二抗检测任何结合的抗体,并使用流式细胞仪记录结果。在用所有构建体进行疫苗接种之前和之后,结合显著提高,但是在用PBS(对照)进行疫苗接种后没有。总体而言,在这两种情况下,疫苗候选物的表现优于来自COBRA的那些(*);

[0396] 图7示出了通过来自用COBRA或DIOS HA基因抗原免疫接种的动物的鼠血清确定在其细胞表面上表达两种不同的组1甲型流感糖蛋白(季节性H1N1和大流行起源H1N1)的细胞的结合的研究结果;以及

[0397] 图8示出了通过来自DIOS或COBRA DNA免疫接种的鼠的血清的H7N9(A/Shanghai2/2013)的跨HA组结合(左图)和假型中和(右)的结果。在右图中,最上面的曲线是CR9114的,从曲线图左侧的最低两个起始点开始下降的两条曲线是H1N1s的,并且其余两条曲线是H1N1pdm的。

[0398] 下文给出了未优化埃博拉和马尔堡病毒祖先核酸序列(即尚未进行密码子优化或基因优化的序列)以及编码候选抗原性病原体多肽的基因优化核酸序列的实例。

[0399] 方法

[0400] 对于给定的病毒种类,从例如GenBank(以及从其他任何可用来源,例如爆发数据)下载候选原始序列,并进行过滤以去除相同的序列,不跨越目的蛋白的序列,以及具有大量模糊核苷酸的序列。产生经过滤序列的多序列比对(通常使用MAFFT),并手动检查以确保序列处于正确的开放阅读框中。使用具有自动模型选择的IQTREE产生最大似然系统发育,并使用多种方法之一进行生根:外类群序列(outgroup sequence)、中点生根(midpoint rooting)、树中心(centre-of-the-tree)或使根部到尖端的距离(root-to-tip distance)和采样时间之间的关联最大化的树。使用HyPhy假设密码子置换的MG94 by F3x4模型来产生祖先序列,并进行检查以确保已保留了已知的表位。使用IQTREE产生具有初级序列和祖先序列二者的系统发育树,以检查祖先株的位置。然后以多种方式修饰祖先序列:区域的缺失(例如,去除黏蛋白样结构域);区域交换(以恢复潜在丢失的表位);特定位点的突变(例如在丝状病毒的融合结构域中),包括编辑N-连接的糖基化位点和引入突变以增强稳定性。

[0401] 实施例1

[0402] 埃博拉苏丹祖先(T2-4)

[0403] 未优化的(SEQ ID NO:1)

```
ATGGGGGGTCTTAGCCTACTCCAATTGCCAGGGACAAATTTCCGAAAAGCTCTTTCTTTGTTTGGG
TCATCATCTTATFCCAAAAGGCCTTTTCCATGCCTTTGGGTGTTGTGACTAACAGCACTTTAGAAGT
AACAGAGATTGACCAGCTAGTCTGCAAGGATCATCTTGCATCCACTGACCAGCTGAAATCAGTTGGT
CTCAACCTCGAGGGGAGCGGAGTATCTACTGATATCCCATCTGCAACAAAGCGTTGGGGCTTCAGAT
CTGGTGTTCCTCCAAGGTGGTCAGCTATGAAGCGGGAGAATGGGCTGAAAATTGCTACAATCTTGA
AATAAAGAAGCCGGACGGGAGCGAATGCTTACCCCCACCGCCAGATGGTGTGAGAGGCTTTCCAAGG
TGCCGCTATGTTTCAAAAGCCCAAGGAACCGGGCCCTGCCAGGTGACTACGCCTTTCAAAAGGATG
GAGCTTTCTTCTCTATGACAGGCTGGCTTCAACTGTAATTTACAGAGGAGTCAATTTTGTGAGGG
GGTAATTGCATCTTGATATTGGCTAAACCAAAGAAAACGTTCCCTTCAGTCACCCCCCATTTCGAGAG
GCAGTAAACTACACTGAAAATACATCAAGTTATTATGCCACATCCTACTTGGAGTATGAAAATCGAAA
ATTTTGGTGCTCAACACTCCACGACCCTTTTCAAATTTGACAATAATACTTTTGTTCGTCTGGACAG
GCCCCACACGCCTCAGTTCTTTTCCAGCTGAATGATACCATTACCTTCACCAACAGTTGAGCAAC
ACAACCTGGGAGACTAATTTGGACACTAGATGCTAATATCAATGCTGATATTGGTGAATGGGCTTTTT
```

[0404]

```
GGGAAAATAAAAAAATCTCTCCGAACAACTACGTGGAGAAGAGCTGTCTTTTCGAAGCTTTATCGCT
CACAAACAGCGGTTAAAACCTGTCTTGCCACAGGAGTCCACAAGCAACGGTCTAATAACTTCAACAGTA
ACAGGGATTCTTGGGAGTCTTGGGCTTCGAAAACGCAGCAGAAGACAAGTTAACACCAAAGCCACGG
GTAATGCAATCCCAACTTACACTACTGGACTGCACAAGAACAACATAATGCTGCTGGGATTGCCTG
GATCCCGTACTTTGGACCGGGTGCAGGAGCATATACACTGAAGGCCCTGATGCATAACAAAATGCC
TTAGTCTGTGGACTTAGGCAACTTGCAATGAAACAACTCAAGCTCTGCAGCTTTTCTTAAGAGCCA
CAACGGAGCTGCGGACATATACCATACTCAATAGGAAGGCCATAGATTTCCCTTCTGCGACGATGGGG
CGGGACATGCAGGATCCTGGGACCAGATTGTTGCATTGAGCCACATGATTGGACAAAAAACATCACT
GATAAAAATCAACCAATCATCCATGATTTTCATCGACAACCCCTTACCTAATCAGGATAATGATGATA
ATTGGTGGACGGGCTGGAGACAGTGGATCCCTGCAGGAATAGGCATTACTGGAATTATTATTGCAAT
TATTGCTCTTCTTTGCGTTTGAAGCTGCTTTGCTAG
```

[0405] 经基因优化的 (SEQ ID NO:2)

ATGGGAGGACTGTCTCTGCTGCAACTGCCCGGGACAAGTTCGGGAAGTCCAGCTTCTTCGTGTGGG
TCATCATCTGTTCAGAAAGCCTTCAGCATGCCCTGGGCGTCGTGACCAATAGCACACTGGAAGT
GACCGAGATCGACCAGCTCGTGTGCAAGGATCACCTGGCCAGCACCGATCAGCTGAAGTCTGTGGGA
CTGAATCTGGAAGGCAGCGGCGTGTCCACAGATATCCCTAGCGCCACCAAGAGATGGGGCTTTAGAA
GCGGAGTGCCTCCTAAGGTGGTGTCTTATGAAGCCGGCGAGTGGGCCGAGAAGTCTACAACCTGGA
AATCAAGAAGCCCGACGGCAGCGAGTGTCTGCCTCCTCCACCTGATGGCGTCAGAGGCTTCCCTAGA
TGCAGATACGTGCACAAGGCCAAGGCACAGGACCTGTCTGGCGATTACGCCTTTCACAAGGACG
GGCCTTTTTCTGTACGATCGGCTGGCCTCCACCGTGATCTACAGAGGGCTTAACTTTGCCGAGGG
CGTGTATGCCTTCCTGATCCTGGCCAAGCCTAAAGAGACATTCCTGCAAAGCCCTCCAATCCGCGAG
GCCGTGAACACACAGAGAACCACAGCAGTACTACGCCACAGCTACCTGGAATACGAGATCGAGA
ATTTGGCGGCCAGCACAGCACCACACTGTTCAAGATCGACAACAACACCTTCGTGCGGCTGGACAG
ACCCACACACCTCAGTTTCTGTTCCAGCTGAACGACACCATCCATCTGCATCAGCAGCTGAGCAAC
ACCACCGGACAGCTGATTTGGACCCTGGACGCCAACATCAACGCCGACATTGGAGAGTGGGCCTTTT
GGGAGAACAAAGAAACCTGAGCGAACAGCTGAGAGGCGAGGAAGTGGAGCTTTGAGGCCCTGTCTCT
GACCACCGCCGTGAAAACAGTGTCTGCCTCAAGAGTCCACCAGCAACGGCCTGATCACAAGCACAGTG
ACAGGATCCTGGGCAGCCTGGGCCTGAGAAAAAGTCCAGACGGCAAGTGAATACCAAGGCCACCG
GCAAGTGAACACCCCAACCTGCACATTTGGACAGCCCAAGAGCAGCAACAATGCCCGGAATCGCCTG
GATTCCTTATTTTGGACCTGGCGCCGAGGGCATCTATACCGAGGGACTGATGCACAACCCAGAAGCC
CTCGTGTGTGGACTGAGACAGCTGGCCAATGAGACAACACAGGCCCTCCAGCTGTTTCTGAGAGCCA
CCACCGAGCTGAGAACCTACACCATCCTGAACCGGAAGGCCATCGACTTTCTGCTGAGAAGATGGGG
CGGCACCTGTAGAACTCCTGGGACCTGATTGCTGCATCGAGCCCCACGACTGGACCAAGAACATCACC
GACAAGATCAACCAGATCATCCAGACTTCATCGACAACCTCTGCCTAACCAGGACAACGACGACA
ATTGGTGGACAGGCTGGCGGCAGTGGATTCTGCAGGAATTGGCATCACCGGCATCATCATTGCCAT
TATCGCCCTGCTGTGTGTGTGCAAGCTGCTGTGTTGA

[0406]

[0407] 由未优化和经基因优化的序列编码的氨基酸序列 (SEQ ID NO:3) :

MGGLSLLQLPRDKFRKSSFFVWVIILFQKAFSMLPLGVVTNSTLEVTEIDQLVCKDHLASTDQLKSVGLNLEGG
VSTDIPSATKRWGRFSGVPPKVVS YEAGEWAENCYNLEIKKPDGSECLPPPDPGVRGFPRCRYVHKAQGTGPCP
GDYAFHKDGAFFLYDRLASTVIYRGNFAEGVIAFLILAKPKETFLOSPPIREAVNYTENTSSYYATSYLEYEI
ENFGAQSHTLFLKIDNNTFVRLDRPHTPQFLFQLNDTIHLHQQLSNTTGRLIWTL DANINADIGEWAFWENKKN
LSEQLRGEELSFEALSLLTAVKTVLPQESTSNGLITSTVTGILGSLGLRKRSRQVNTKATGKCNPNLHYWTAQ
EQHNAAGIAWIPYFGPGAEGIYTEGLMHNQNALVGLRQLANETTQALQLFLRATTELRTYTIILNRKAIDFLLR
RWGGTCRIILGPDCCIEPHDWTKNITDKINQIIHDFIDNPLPNQDNDNDNWWTGWRQWIPAGIGITGIIIAIALL
CVCKLLC

[0408]

[0409] 实施例2

[0410] 埃博拉病毒全球祖先 (T2-6)

[0411] 未优化的 (SEQ ID NO:4)

ATGGGGGGTGGATCCAGACTTCTCCAATTGCCCGGGAAACGCTTTTCGAAAAACCTCATTCTTTGTTT
GGGTAATCATCCTATTCCAAAAGCCTTTTCCATGCCATTGGGTGTTGTAACCAACAGCACTCTAAA
AGTAACAGAAATTGACCAATTGGTTTGGCCGGACAAACTTTTCATCCACAAGTCAGCTGAAATCAGTT
GGGCTGAATCTGGAAGGGAATGGAGTTGCAACTGATGTCCCATCAGCAACAAAACGATGGGGCTTCC
GATCTGGTGTTCCTCCCAAGGTGGTCAGCTATGAAGCTGGAGAATGGGCTGAAAATTGCTACAATCT
GGAAATCAAGAAGCCAGACGGGAGTGAATGCCTACCCTCCACCGCCAGACGGTGTAAAGAGGCTTCCCC
AGGTGCCGCTATGTCCACAAAGTTCAAGGAACAGGGCCGTGTCTGGTGACTTCGCCCTTCCACAAAG
ATGGAGCTTTCTTCCGTATGATAGACTGGCTTCACTGTCAATTTACCGAGGGACAACCTTTTGCTGA
AGGTGTGCTGCATTTTTGATCCTGCCCAAACCTAAAAAGGACTTTTTCCAATCACCCCAATACGT
GAGCCGGTAAACACCACAGAAGATCCATCAAGTTACTACACCACATCAACACTTAGCTATGAGATTG
ACAATTTTGGGGCCAATAAACTAAAACCTTTTTCAAAGTTGACAATCACACTTATGTGCAACTAGA
CCGACCACACACCACAGTTCTTGTCCAGCTCAATGAAACCATTACATAAAAATAACCGTCTAAGC
AACACCACAGGGAGACTAATTTGGACATTAGATCCTAAAATTGATACCGACATTGGTGAGTGGGCCT
TCTGGGAAAAATAAAAAAATTTCTCCAAACAACCTTCGTGGAGAAGAGTTGTCTTCAAAGCTCTATC
AACAAAACTGGAGCTAACGCAGTAGACACTGACGAATCAAGCAAACCTGGCCTAATTACCAACACA
GTAAGAGGGGTTGCTGATTTACTGAGCCCTTGGAGAAGAAAAAGAAGACAAGTCAACCCAAACACAA
CAAATAAATGCAACCCAAACCTACACTATTGGACAGCCCAAGATGAAGGTGCTGCCGTTGGATTAGC
CTGGATCCCATACTTCGGACCAGCAGCAGAAGGCATTTACACTGAAGGAATAATGCATAATCAAAT
GGGTAAATCTGTGGGCTGAGGCAGCTGGCCAATGAAACGACTCAAGCTCTTCAATTATTCTTGAGGG
CCACAACGGAGCTGCGGACTTACTCTATACTCAATAGAAAAGCCATTGATTTCTTCTCCAACGATG
GGGAGGAACATGCCGCATCTTAGGACCAGATTGTTGCATTGAGCCACATGATTGGACAAAAACATT
ACTGATAAAATTAACCAATCATACATGATTTTATTGACAACCCCTTACCAGATCAGGACGATGATG
ACAATTGGTGGACAGGCTGGAGACAATGGATCCCTGCTGGAATTGGAATTACTGGAGTTATAATTGC
AATTATAGCTCTACTTTGTATTTGCAAGTTTCTGTGTTAG

[0412]

[0413] 经基因优化的 (SEQ ID NO:5)

ATGGGCGGAGGATCTAGACTGCTGCAACTGCCAGAGAGCGGTTTCAGAAAGACCAGCTTCTTCGTGT
GGGTCAATCATCCTGTTCCAGAAAGCCTTCAGCATGCCCTGGGGCGTCGTGACCAATAGCACCCCTGAA
AGTGACCCGAGATCGACCAGCTCGTGTGCAGAGATAAGCTGAGCAGCACCAGCCAGCTGAAGTCCGTG
GGACTGAATCTGGAAGGCAATGGCGTGGCCACAGATGTGCCCTAGCGCCACCAAAAGATGGGGCTTTA
GAAGCGGCGTGCCACCCTAAGGTGGTGTCTTATGAAGCCGGCGAGTGGGCCGAGAACTGCTACAACCT
GAAATCAAGAAGCCCGACGGCAGCGAGTGTCTGCCCTCCACCTGATGGCGTTCAGAGGCTTCCCT
AGATGCAGATACGTGCACAAGGTGCAAGGCACAGGCCCTGTCCTGGCGATTTCGCCTTTACAAGG
ACGGCGCCTTTTTCTGTACGATCGGCTGGCCTCCACCGTATCTACAGAGGCACAACATTTGCCGA
AGGCGTGGTGGCTTCTGATCCTGCCTAAGCCTAAGAAGGACTTCTTTTCAGAGCCCTCCTATCCGC
GAGCCTGTGAACACAACAGAGGACCCAGCAGCTACTACACCACCAGCACACTGAGCTACGAGATCG
ATAACTTCGGCGCCAACAAGACCAAGACACTGTTCAAGGTGGACAACCACACCTACGTGCAGCTGGA
CAGACCCACACACCTCAGTTTCTGGTGCAGCTGAACGAGACAATCCACACCAACAACAGACTGAGC
AACACCACCGGCAGGCTGATCTGGACCTGGATCCTAAGATCGACACCGACATCGGAGAGTGGGCCT
TTTGGGAGAACAAGAAGAACTTCAGCAAGCAGCTGAGAGGCGAGGAACTGAGCTTTAAGGCCCTGAG
CACCAAGACAGGCGCCAACGCTGTGGATAACCGATGAGTCTAGCAAGCCCGGCTGATCACCAACACA
GTTAGAGGCGTTGCCGACCTGCTGAGCCCTTGGAGAAGAAAGCGGAGACAAGTGAACCCCAATACCA
CCAACAAGTGCAACCCCTAACCTGCACTACTGGACAGCCAGGATGAAGGGCGCTGCTGTTGGACTGGC
CTGGATTCTTATTTTGGACCTGCCGCCGAGGGCATCTACACAGAGGGAATCATGCACAACCAGAAT
GGCCTGATCTGCGCCCTGAGACAGCTGGCCAATGAGACAACACAGGCCCTCCAGCTGTTTCTGAGAG
CCACCACCGAGCTGAGAACCCTACAGCATCCTGAACCGAAGGCCATCGACTTTCTGCTGCAAAGATG
GGGAGCACCTGTAGAATCCTGGGACCTGATTGCTGCATCGAGCCCACGACTGGACCAAGAACATC
ACCGACAAGATCAACCAGATCATCCACGACTTCATCGACAACCCTCTGCCTGACCAGGACGACGACG
ATAATTGGTGGACAGGATGGCGGCAGTGGATTCTGCCGGAATCGGAATCACAGGCGTGATCATTGC
CATTATCGCCCTGCTGTGCATCTGCAAGTTTCTGTGCTGA

[0414]

[0415] 由未优化和经基因优化的序列编码的氨基酸序列 (SEQ ID NO:6) :

MGGGSRLQLPRERFRKTSFFVWVILFQKAFSMP LGVVTNSTLKVTEIDQLVCRDKLSSTSOLKSV
GLNLEGNVATDVPSATKRWGFRRSGVPPKVVSYEAGEWAENCYNLEIKKPDGSECLP PPDGVRGFP
RCRYVHKVQGTGFCP GDFAFHKDGAFFLYDRLASTVIYRGTTFAEGVVAFLILPKPKKDFQSPPIR
EPVNTTEDPSSYYTSTLSYEIDNFGANKTKTLFKVDNHTYVQLDRPHTPQFLVQLNETIHTNRLS
NTTGRLIWTLPKIDTDIGEWAFWENKKNFSKQLRGEELSFKALSTKTGANAVDTDESSKPLITNT
VRGVADLLSPWRRKRQVNPNTTNKCNPNLHYWTAQDEGAAVGLAWIPYFGPAAEGIYTEGIMHNQN
GLICGLRQLANETQALQLFLRATTELRTYSILNRKAIDFLLQRWGGTCRILGPDCCIEPHDWTKNI
TDKINQI IHDF IDNPLPDQDDDDNWWTGWRQWIPAGIGITGVIIAIIALLCICKE LC

[0416]

[0417] 实施例3

[0418] 马尔堡病毒祖先 (T2-11)

[0419] 未优化的 (SEQ ID NO:7)

ATGAAGACCATATATTTTCTGATTAGTCTCATTTTAATCCAAAGTATAAAAACTCTCCCTGTTTTAG
AAATTGCTAGTAACAGCCAACCTCAAGATGTAGATTCAGTGTGCTCCGGAACCCCTCCAAAAGACAGA
AGATGTTTCATCTGATGGGATTTACTACTGAGTGGGCAAAAAGTTGCTGATTCCCTTTGGAAGCATCT
AAACGATGGGCTTTCAGGACAGGTGTTCCCTCCCAAGAAGCTTGAGTATACGGAAGGAGAAGAAGCCA
AAACATGTTACAATAAAGTGTAAACAGACCTTCTGGAAAATCCTTGCTGCTGGATCCTCCAGTAA
TATCCGCGATTACCCATAAATGTAAAACCTGTTTCATCATATTTCAAGGTCAAACCCCTCATGCACAGGGG
ATTGCCCTCCATTTGTGGGGGGCATTTTTCTGTATGATCGCATTGCCTCCACAACAATGTACCGAG
GCAAAGTCTTCACTGAAGGGAACATAGCAGCTATGATTGTCAATAAGACAGTGCACAAAATGATTTT
CTCGAGGCAAGGACAAGGGTACCCTCACATGAATCTGACTTCTACTAATAAATATGGACAAGTAGC
AACGGAACGCAAACGAATGACACTGGATGCTTCGGTGTCTTCAAGAATACAATCTACGAAGAACC
AAACATGTGCTCCGTCCAAAATACCTCCACCCTGCCCACAGCCCGTCCGGAGATCAAACCCACAAG
CACCCCAACTGATGCCACCAAACCTCAACACCACAGACCCAAACAGTGTATGATGAGGACCTCACAACA
TCCGGCTCAGGTTCCGGAGAACAGGAACCCCTACACAACCTTCTGATGCGGTACTAAGCAAGGGCTTT
CATCAACAATGCCACCCACTCCCTCACCACAACCAAGCACGCCACAGCAAGGAGGAAACACACAAA
CCATTCCCAAGGTGCTGTGACTGAACCCGACAAAACCAACACAACCTGCACAACCGTCCATGCCCCCC
[0420] CACAACACTACTACAATCTCTACTAACAACACCTCCAAGCACAACCTTTCAGCACTCTCTCTGCACCAC
TACAAAACACCACCAATTACAACACACAGAGCAGCGCCACTGAAAATGAGCAAACCAAGTGCCTCCCTC
GAAAACAACCCGTCCCTCCAACAGGAAATCCTACCACAGCAAAGAGCACCAACAGCACAAAAGGCCCC
ACCACAACCGCACCAATACGACAAATGGGCAATTCACCACTCCCTCCCCACCCCACTCGACTA
CACACAATCTGTATATTTTCAAGGAAACGAAGTATCCTCTGGAGGGAAGGCACATGTTCCCTTT
TTTAGATGGGTTAATAAATACTGAAATTTGATTTTGAATCCAATCCCAAACACAGAAACAATCTTTGAT
GAATCCCCCAGCTTTAATACTTCAACTAATGAGGAACAACACACTCCCCGAATATCAGTTTAACTT
TCTCTTATTTTCTGATAAAAATGGAGATACTGCCACTCTGGGAAAACGAGAATGATTGTGATGC
AGAGTTGAGGATTTGGAGTGTGCAGGAGGACGATTTGGCGGCAGGGCTTAGCTGGATACCATTTTTT
GGCCCTGGAATCGAAGGACTCTATACTGCCGTTAATCAAAAATCAGAACAATTTAGTTTGTAGGT
TGAGGCGCTTAGCTAATCAAACCTGCTAAATCCTTGGAGCTCTTGTAAAGGGTACAACCCGAGGAAAG
GACATTTTCTTAATCAATAGGCATGCAATTTGACTTTTGCCTTACGAGGTGGGGCGGAACATGCAAG
GTGCTAGGACCTGATTGTTGCATAGGAATAGAAGTCTATCTAAAAATATCTCAGAACAAATTGACA
AAATCAGAAAGGATGAACAAAAGGAGGAAACTGGCTGGGGTCTAGGTGGCAAATGGTGGACATCTGA
CTGGGGTGTCTCACCAATTTGGGCATCCTGCTACTATTATCTATAGCTGTTCTGATTGCTCTGTCC
TGTATCTGTCTATCTTCACTAAATATATCGGATAG

[0421] 经基因优化的 (SEQ ID NO:8)

ATGAAGACCATCTACTTTCTGATCAGCCTGATCCTGATCCAGAGCATCAAGACCCCTGCCTGTGCTGG
AAATCGCCAGCAACAGTCAGCCCAGGATGTGGATAGCGTGTGTAGCGGCACCCCTCCAGAAAACCGA
GGATGTGCACCTGATGGGCTTTACCCTGAGCGGCCAGAAAGTGGCCGATTCTCCACTGGAAGCCAGC
AAGAGATGGCCTTTAGAACCAGCGTCCACCTAAGAAGCTCGAGTACACAGAGGGCGAAGAGGCCA
AGACCTGCTACAACATCAGCGTGACCGATCCTAGCGGCAAGAGCCTGCTGCTGGACCCCTCCTAGCAA
CATCAGAGACTACCCCAAGTGAAGACCGTGCACCACATCCAGGGACAGAATCCCATGCTCAGGGA
ATTGCCCTGCACCTGTGGGGCGCCTTTTTCTGTATGATCGGATCGCCTCCACCACCATGTACAGAG
GCAAAGTGTTCACCGAGGGCAATATCGCCGCCATGATCGTGAACAAGACAGTGCACAAGATGATCTT
CACCCGGCAAGCCAGGGCTACAGACATGAAATCTGACCAGCACCAACAAGTACTGGACAGCAGC
AACGGCACCCAGACCAATGATACAGGCTGCTTTGGCGCCCTGCAAGAGTACAACAGCACCAGAATC
AGACATGCGCCCTAGCAAGATCCCTCCTCCACTGCCTACTGCCAGACCTGAGATCAAGCCTACCAG
CACACCTACCGACGCCACCAAGCTGAACACCACCGATCCAAACAGCGACGACGAGGATCTGACAACA
AGCGGATCTGGCTCTGGCGAGCAAGAGCCATACACCACCTCTGATGCGGTGACAAAGCAGGGCCTGA
GCAGCACAATGCCCTCAACACCTTCTCCACAGCCTAGCACACCTCAGCAAGGCGGCAACAACACAAA
[0422] TCACTCTCAGGGCGCCGTGACCGAGCCTGACAAGACAAATACCACAGCTCAGCCCAGCATGCCCTCT
CACAACACCACCACAATCTCCACCAACAACACCAGCAAGCACAACCTTTCAGCACACTGAGCGCCCTC
TCCAGAATACCACCAACTACAATACCAGAGCACCGCCACCAGAGAACGAGCAGACATCTGCCCTTC
TAAGACCACACTGCCACCTACCGCAATCCTACCACCGCCAAGAGCACCATAAGCACAAGGGCCCT
ACCACCACCGCTCCTAACACCACAATGGCCACTTCACAAGCCCAAGTCTTACACCTAACAGCACAA
CCCAGCACCTGGTGTACTTCAGACGGAAGCGGAGCATCCTTTGGCGCGAGGGCGATATGTTCCCTTT
CCTGGACGGCTGATCAACACCGAGATCGACTTCGACCCCATTCAAAACACCGAACCATCTTCGCAC
GAGAGCCCCAGCTTCAACACCTCCACCAATGAGGAACAGCACACCCCTCCAAAACATCTCCCTGACCT
TCAGTACTTTCCCGACAAGAACGGCGATACAGCCTACAGCGCGAGAATGAGAATGACTGCCAGCG
CGAGCTGCGGATTTGGAGCGTTCAAGAGGATGATCTGGCTGCCGGCCTGAGCTGGATCCCTTTTTTT
GGACCTGGCATCGAGGGCCTGTACACCGCCGACTGATCAAGAACCAGAACAACCTCGTGTGCAGAC
TGCAGGAGACTGGCCAATCAGACCGCAAGTCTCTGGAACCTGCTGCTGCGCGTGACCACCGAGGAAAG
AACCTTCTCTCTGATCAACCGGCACGCCATCGATTTCTGCTGACCAGATGGGGCGGCACCTGTAAA
GTTCTGGGCCCTGATTGCTGCATCGGAATCGAGGACCTGAGCAAGAACATCTCCGAGCAGATCGACA
AGATCCGCAAGGACGAGCAGAAAGAGGAAACAGGCTGGGGACTCGGCGCAAGTGGTGGACATCTGA
TTGGGGCGTGTGACCAATCTGGGAATCTGCTGCTCTATCGCCGTGCTGATCGCCCTGAGC
TGCATCTGCCGGATCTTCAACCAAGTACATCGGCTGA

[0423] 由未优化和经基因优化的序列编码的氨基酸序列 (SEQ ID NO:9) :

MKTIYFLISLILIQSIKTLPLVLEIASNSQFQDVDSVCSGTLQKTEDVHLMGFSLSGQKVADSPLEASKRWAFRT
 GVPPKNVEYTEGEEAKTCYNI SVTDP SGKSLLLDPPSNIRDYPCKCTVHHIQGONPHAQGIALLHWGAFFLYDR
 IASTIMYRGKVFTEGNIAAMIVNKTVHKMIFSRQGGYRHMNLTSTNKYWTSSNGTQTNDTGCFGALQEQYNSTK
 NQTCAPSKIPPLPTARPEIKPTSTPDTATKLNITDPNSDDEDLTSSGSGSGEQEPYTTSDAVTKQGLSSSTMPF
 TPSPQPSPTPQQGNNTNHSQGAVTEPKTNTTAQPSMPPHNTTITSTNNTSKHNFSTLSAPLQNTTNYNTQSTA
 TENEQTSAPSKTTLPTGNPTTAKSTNSTKGP TTTAPNTTNGHFTSP SPTPNSTTQHLVYFRKRKRSILWREGDM
 FPFLDGLINTEIDFDPINTEIFDESPSFNTSTNNEEQHTPPNISLTF SYFPDKNGDTAYS GENENDCDAELRI
 WSVQEDDLAAGLSWIPFFGPGIEGLYTAGLIKNNLVCRLRLRLANQTAKSLELLLRVTTEERTFSLINRHAID
 FLLTRWGGTCKVLGPDCCIGIEDLSKNI SEQIDKIRKDEQKEETGWGLGGKWWTSDWGVLTNLGILLLSIAVL
 IALSCICRIFTKYIG

[0424] 实施例4

[0426] Tier 2-4 (SUDV anc-MLD)

[0427] 具有缺失的(负“-”)黏蛋白样结构域的苏丹埃博拉病毒祖先序列

[0428] 核苷酸序列 (SEQ ID NO:10):

```

atgggaggac tgtctctgct gcaactgccc cgggacaagt tccggaagtc cagcttcttc      60
gtgtgggtca tcatcctggt ccagaaagcc ttcagcatgc ccctggcgt cgtgaccaat      120
agcacactgg aagtgaccga gatcgaccag ctctgtgca aggatcacct ggccagcacc      180
gatcagctga agtctgtggg actgaatctg gaaggcagcg gcgtgtccac agatccct      240
agcccaacca agagatgggg ctttagaagc ggagtgcctc ctaagtggt gtcttatgaa      300
gcccgcgagt gggccgagaa ctgctacaac ctggaatca agaagcccga cggcagcgag      360
tgtctgcctc ctccacctga tggcgtcaga ggctcccta gatgcagata cgtgcacaag      420
gcccaggca caggacctg tcctggcgat tacgccttcc acaaggacgg cgccttttcc      480
ctgtacgatc ggctggcctc caccgtgatc tacagaggcg ttaactttgc cgagggcgtg      540
atcgccctcc tgatcctggc caagcctaaa gagacattcc tgcaaagccc tccaatccgc      600
gaggccgtga actacacaga gaacaccagc agctactagc ccaccagcta cctggaatac      660
gagatcgaga atttcggcgc ccagcacagc accacactgt tcaagatcga caacaacacc      720
ttcgtcgggc tggacagacc ccacacacct cagttctgt tccagctgaa cgacaccatc      780
[0429] catctgcctc agcagctgag caacaccacc ggcagactga tttggacct ggacgccaac      840
atcaacgccc acattggaga gtgggccttt tgggagaaca agaagacct gagcgaacag      900
ctgagaggcg aggaactgag ctttgaggcc ctgtctctga ccaccgccgt gaaaacagtg      960
ctgcctcaag agtccaccag caacggcctg atcacaagca cagtgcacag catcctgggc      1020
agcctgggccc tgagaaaaag gtcacagcgg caagtgaata ccaaggccac cggcaagtgc      1080
aaccacaacc tgcactattg gacagcccaa gagcagcaca atgcccggg aatcgctgg      1140
attccttatt ttggacctgg cgcgagggc atctataccg agggactgat gcacaaccag      1200
aacgccctcg tgtgtggact gagacagctg gccaatgaga caacacaggc cctccagctg      1260
tttctgagag ccaccaccga gctgagaacc tacaccatcc tgaaccggaa ggccatcgac      1320
tttctgctga gaagatgggg cggcacctgt agaatcctgg gacctgattg ctgcatcgag      1380
ccccacgact ggaccaagaa catcaccgac aagatcaacc agatcatcca cgacttcac      1440
gacaaccctc gcctaacca ggacaacgac gacaattggt ggacaggctg gcggcagtgg      1500
attcctgccc gaattggcat caccggcatc atcattgcca ttatcgccct gctgtgtgtg      1560
tgcaagctgc tgtgttga      1578

```

[0430] 氨基酸序列 (SEQ ID NO:11):

```

MGGLSLLQLPRDKFRKSSFFVWVILFQKAFSMP LGVVTNSTLEVTEIDQ 50
LVCKDHLASTDQLKSVGLNLEGGVSTDI PSATKRWGFRSGVPPKVVSYE 100
AGEWAENCYNLEIKKPDGSECLPPPPDGVRGFPRCRYVHKAQGTGPCPGD 150
YAFHKDGAFFLYDRLASTVIYRGVNF AEGVIAFLILAKPKETFLQSPPIR 200
EAVNYTENTSSYYATSYLEYIEIENFGAQHSTTLFKIDNNTFVRLDRPHTP 250
[0431] LRFNLNDTIHLHQQLSNTTGRLIWTL DANINADIGEWAFWENKKNLSEQ 300
LRGEELSFEALSLTTAVKTVLPQESTS NGLITSTVTGILGSLGLRKRSR 350
QVNTKATGKCNPNLHYWTAQE QHNAAGIAWIPYFGPGAEGIYTEGLMHNQ 400
NALVCGLRQLANETTQALQLFLRATTE LRITYTILNRKAIDFLLRRWGGTC 450
RILGPDCCIEPHDWTKNITDKINQI IHDFIDNPLPNQDNDNWWTGWRQW 500
IPAGIGITGIIIAIIALLCVCKLLC*

```

[0432] 实施例5

[0433] Tier 2-6 (SUDV EBOV-TAFV-BDBV anc-MLD)

[0434] 苏丹、扎伊尔、塔伊森林和本迪布焦埃博拉病毒四种物种的祖先序列,其中黏蛋白样结构域缺失。

[0435] 核苷酸序列 (SEQ ID NO:12):

	atggggcggag gatctagact gctgcaactg cccagagagc ggttcagaaa gaccagcttc	60
	ttcgtgtggg tcatcatcct gttccagaaa gccttcagca tgcccctggg cgtcgtgacc	120
[0436]	aatagcacc tgaagtgtac cgagatcgac cagctcgtgt gcagagataa gctgagcagc	180
	accagccagc tgaagtccgt gggactgaat ctggaaggca atggcgtggc cacagatgtg	240
	cctagcgcca ccaaaagatg gggctttaga agcggcgtgc cacctaaggt ggtgtcttat	300
	gaagccggcg agtgggccga gaactgctac aacctggaaa tcaagaagcc cgacggcagc	360
	gagtgtctgc ctctccacc tgatggcgtc agaggcttcc ctatagtcag atactgtcac	420
	aaggtgcaag gcacaggccc ctgtcctggc gatttcgcct ttcacaagga cggcgccttt	480
	ttcctgtacg atcggctggc ctccaccgtg atctacagag gcacaacatt tgccgaaggc	540
	gtggtggcct tctgatcct gcctaagcct aagaaggact tctttcagag ccctcctatc	600
	cgcgagcctg tgaacacaac agaggacccc agcagctact acaccaccag cacactgagc	660
	tacgagatcg ataacttcgg cgccaacaag accaagacac tgttcaaggt ggacaaccac	720
	acctacgtgc agctggacag accccacaca cctcagtttc tgggtcagct gaacgagaca	780
	atccacacca acaacagact gagcaacacc accggcaggc tgatctggac cctggatcct	840
	aagatcgaca ccgacatcgg agagtgggcc ttttgggaga acaagaagaa cttcagcaag	900
[0437]	cagctgagag gcgaggaaact gagctttaag gccttgagca ccaagacagg cgccaacgct	960
	gtggataccg atgagtctag caagcccggc ctgatcacca acacagttag aggcgttgcc	1020
	gacctgctga gcccttggag aagaaagcgg agacaagtga accccaatac caccaacaag	1080
	tgcaacccta acctgcacta ctggacagcc caggatgaag gcgctgctgt tggactggcc	1140
	tggattcctt attttggacc tgccgccgag ggcattaca cagagggaat catgcacaac	1200
	cagaatggcc tgatctggcg cctgagacag ctggccaatg agacaacaca ggccctccag	1260
	ctgtttctga gagccaccac cgagctgaga acctacagca tcctgaaccg gaaggccatc	1320
	gactttctgc tgcaaagatg gggaggcacc tgtagaatcc tgggacctga ttgctgcatc	1380
	gagccccacg actggaccaa gaacatcacc gacaagatca accagatcat ccacgacttc	1440
	atcgacaacc ctctgctga ccaggacgac gacgataatt ggtggacagg atggcggcag	1500
	tggattcctg ccggaatcgg aatcacaggc gtgatcattg ccattatcgc cctgctgtgc	1560
	atctgcaagt ttctgtgctg a	1581

[0438] 氨基酸序列 (SEQ ID NO:13) :

MGGGSRLQLPRERFRKTSFFVWVIILFQKAFSMLPLGVVTNSTLKVTEID 50
QLVCRDKLSSTSSQLKSVGLNLEGNVATDVP SATKRWGF RSGVPPKVVS Y 100
EAGEWAENCYNLEIKKPDGSECLPPPPDGVRGFP RCRYVHKVQGTGPCPG 150
DFAFHKDGAFFLYDRLASTVIYRGTTF AEGVVAFLILPKPKDFFQSPP I 200
REPVNTTEDPSSYYTSTLSYEIDNFGANKTKTLFKVDNHTYVQLDRPHT 250
PQFLVQLNETIHTNNRLSNTTGRLIWTLPKIDTDIGEWAFWENKKNFSK 300
QLRGEELSFKALSTKTGANAVDTDESSKPLITNTVRGVADLLSPWRRKR 350
RQVNPNTTNKCNPNLHYWTAQDEGA AVGLAWIPYFGPAAEGIYTEGIMHN 400
QNGLICGLRQLANETTQALQLFLRATTELRTYSILNRKAIDFLLQRWGGT 450
CRILGPDCCIEPHDWTKNITDKINQI IHDFIDNPLPDQDDDDNWWTGWRQ 500
WIPAGIGITGVIIAIIALLCICKFLC*

[0440] 实施例6

[0441] Tier 2-11 (RAVV MARV anc)

[0442] 毒株马尔堡病毒和Ravn病毒的祖先序列

[0443] 核苷酸序列 (SEQ ID NO:14) :

[0444] atgaagacca tctactttct gatcagcctg atcctgatcc agagcatcaa gaccctgcct 60
 gtgctggaaa tcgccagcaa cagtcagccc caggatgtgg atagcgtgtg tagcggcacc 120
 ctccagaaaa ccgaggatgt gcacctgatg ggctttacc tgagcggcca gaaagtggcc 180
 gattctccac tggaaagccag caagagatgg gcctttagaa ccggcgtgcc acctaaagac 240
 gtcgagtaca cagagggcga agagggcaag acctgctaca acatcagcgt gaccgatcct 300
 agcggcaaga gcctgctgct ggaccctcct agcaacatca gagactacc caagtgaag 360
 accgtgcacc acatccaggg acagaatccc catgctcagg gaattgcoct gcacctgtgg 420
 gggcctttt tctgtatga tcggatcgcc tccaccacca tgtacagagg caaagtgttc 480
 accgagggca atatcgccgc catgatcgtg aacaagacag tgcacaagat gatcttcagc 540
 cggcaaggcc agggctacag acacatgaat ctgaccagca ccaacaagta ctggaccagc 600
 agcaacggca cccagaccaa tgatacaggc tgctttggcg ccctgcaaga gtacaacagc 660
 accaagaatc agacatggc ccctagcaag atccctcctc cactgcctac tgccagacct 720
 gagatcaagc ctaccagcac acctaccgac gccaccaagc tgaaccaccac cgatccaac 780
 agcgacgacg aggatctgac aacaagcggg tctggtctct gcgagcaaga gccatacacc 840
 acctctgatg ccgtgacaaa gcagggcctg agcagcacia tgctccaac accttctcca 900
 cagcctagca cacctcagca aggcggcaac aacacaaatc actctcaggg cgccgtgacc 960
 gagcctgaca agacaaatc cacagctcag cccagcatgc ctctcacia caccaccaca 1020
 atctccacca acaacaccag caagcacaac ttcagcacac tgagcgcctc tctccagaat 1080
 accaccaact acaataccca gagcaccgcc accgagaaag agcagacatc tgcccttct 1140

aagaccacac tgccacctac cggcaatcct accaccgcca agagcaccaa tagcaciaag 1200
 ggccctacca ccaccgctcc taaccaccaca aatggcact tcacaagccc aagtctaca 1260
 cctaacagca caaccagca cctggtgtac ttcagacgga agcggagcat cctttggcgc 1320
 gagggcgata tgttcccttt cctggacggc ctgatcaaca ccgagatcga cttcgacccc 1380
 attccaaca ccgaaaccat cttcgacgag agcccagct tcaaacctc caccatgag 1440
 gaacagcaca cccctcaaaa catctccctg acctcagct acctccccga caagaacggc 1500
 gatcacgctc acagcggcga gaatgagaat gactgagcag ccgagctgog gatttgagc 1560
 gttcaagagg atgatctggc tgcggcctg agctggatcc cttttttgg acctggcatc 1620
 gagggcctgt acaccgcccg actgatcaag aaccagaaca acctcgtgtg cagactgcgg 1680
 agactggcca atcagaccgc caagtctctg gaactgctgc tgcgcgtgac caccgaggaa 1740
 agaacttct ctctgatcaa ccggcacgcc atcgattttc tgctgaccag atggggcggc 1800
 acctgtaaag ttctgggccc tgattgctgc atcggaatcg aggacctgag caagaacatc 1860
 tccgagcaga tcgacaagat ccgcaaggac gagcagaaag aggaaacagg ctggggactc 1920
 ggcggcaagt ggtggacatc tgattggggc gtgctgacca atctgggaat cctgctgctc 1980
 ctgctatcg ccgtgctgat cgccctgagc tgcatctgcc ggatcttcac caagtacatc 2040
 ggctga 2046

[0445] 氨基酸序列 (SEQ ID NO:15):

MKTIYFLISLILIQSIKTLPLVLEIASNSQPQDVDSVCSGTLQKTEDVHLM 50
 GFTLSGQKVADSPLEASKRWFRTGVPPKNVEYTEGEEAKTCYNISVTD 100
 SGKSLLLDPPSNIRDYPKCKTVHHIQGNPHAQGIALHLWGAFFLYDRIA 150
 STTMYRGKVFTEGNIAAMIVNKTVHKMIFSRQGGYRHMNLTSTNKYWTS 200
 SNGTQTNDTGCFFALQEYNSTKNQTCAPSKIPPLPTARPEIKPTSTPTD 250
 ATKLNTPDPSDDDLTTSGSGSGEQEPYTTSDAVTKQLSSTMPPTPSP 300
 QPSTPQQGNNNTNHSQAVTEPDKTNNTAQP SMPPHNTTTISTNNTSKHN 350
 FSTLSAPLQNTTNYNTQSTATENEQTSAPSKTTLPTGNPTTAKSTNSTK 400
 GPSTTAPNTTNGHFTSPSPTPNSTTQHLVYFRRKRSILWREGDMFFLDG 450
 LINTEIDFDPIPNTETIFDESPSFNTSTNEEQHTPPNISLTFSYFPDKNG 500
 DTAYSGENENDCDAELRIWSVQEDDLAAGLSWIPFFPGGIEGLYTAGLIK 550
 NQNNLVCRLRLRLANQTAKSLELLLRVTTEERTFSLINRHAIDFLLTRWGG 600
 TCKVLGPDCIGIEDLSKNISEQIDKIRKDEQKEETGWGLGGKWWTSWDG 650
 VLTNLGILLLLSIAVLIALSCICRIFTKYIG*

[0447] 实施例7

[0448] pEVAC表达载体

[0449] 图3示出了pEVAC表达载体的图。载体的多克隆位点的序列在下面给出,随后是其完整核苷酸序列。

[0450] pEVAC多克隆位点 (MCS) 的序列 (SEQ ID NO:16):

```

                                PstI      KpnI  Sall
[0451] pEVAC 1301 ACAGACTGTT CCTTCCATG GGTCTTTTCT GCAGTCACCG TCGCTACCGT
                                BclI  XbaI  BamHI NotI  BglII
[0452] pEVAC 1351 CGACACGTGT GATCATCTAG AGGATCCGCG GCCCAGATC T
pEVAC的整个序列 (SEQ ID NO:17) :
CMV-IE-E/P: 248 - 989 CMV即早期1增强子/启动子

KanR: 3445 - 4098 卡那霉素抗性

SD: 990 - 1220 剪接供体
[0453] SA: 1221 - 1343 剪接接受体

Tbgh: 1392 - 1942 牛生长激素的终止子信号

pUC-ori: 2096 - 2769 pUC-质粒复制起点
  1 TCGCGCGTTT CGGTGATGAC GGTGAAAACC TCTGACACAT GCAGCTCCCG
  5 GAGACGGTCA CAGCTTGTCT GTAAGCGGAT GCCGGGAGCA GACAAGCCCG
[0454] 10 TCAGGGCGCG TCAGCGGGTG TTGGCGGGTG TCGGGGCTGG CTTAACTATG
  15 CGGCATCAGA GCAGATTGTA CTGAGAGTGC ACCATATGCG GTGTGAAATA
  20 CCGCACAGAT GCGTAAGGAG AAAATACCGC ATCAGATTGG CTATTGGCCA

```

251 TTGCATACGT TGTATCCATA TCATAATATG TACATTTATA TTGGCTCATG
 301 TCCAACATTA CCGCCATGTT GACATTGATT ATTGACTAGT TATTAATAGT
 351 AATCAATTAC GGGGTCATTA GTTCATAGCC CATATATGGA GTTCCGCGTT
 401 ACATAACTTA CGGTAAATGG CCCGCCTGGC TGACCGCCCA ACGACCCCGG
 451 CCCATTGACG TCAATAATGA CGTATGTTCC CATAGTAACG CCAATAGGGA
 501 CTTTCCATTG ACGTCAATGG GTGGAGTATT TACGGTAAAC TGCCACTTG
 551 GCAGTACATC AAGTGTATCA TATGCCAAGT ACGCCCCCTA TTGACGTCAA
 601 TGACGGTAAA TGGCCCCGCT GGCATTATGC CCAGTACATG ACCTTATGGG
 651 ACTTTCCTAC TTGGCAGTAC ATCTAOGTAT TAGTCATCGC TATTACCATG
 701 GTGATGCGGT TTTGGCAGTA CATCAATGGG CGTGGATAGC GGTTTGACTC
 751 ACGGGGATTT CCAAGTCTCC ACCCCATTGA CGTCAATGGG AGTTTGTTTT
 801 GGCACCAAAA TCAACGGGAC TTTCCAAAAT GTCGTACAA CTCGGCCCA
 851 TTGACGCAAA TGGGCGGTAG CCGTGTACGG TGGGAGGTCT ATATAAGCAG
 901 AGCTCGTTTA GTGAACCGTC AGATCGCCTG GAGACGCCAT CCACGCTGTT
 951 TTGACCTCCA TAGAAGACAC CGGGACCGAT CCAGCCTCCA TCGGCTCGCA
 1001 TCTCTCCTTC ACGCGCCCGC CGCCCTACCT GAGGCGGCCA TCCACGCCGG
 1051 TTGAGTCGGC TTCTGCCGCC TCCCGCCTGT GGTGCCCTCT GAACTGCGTC
 1101 CGCCGTCTAG GTAAGTTTAA AGCTCAGGTC GAGACCGGGC CTTTGTCCGG
 1151 CGCTCCCTTG GAGCCTACCT AGACTCAGCC GGCTCTCCAC GCTTTGCCCTG
 1201 ACCCTGCTTG CTCAACTCTA GTTAACGGTG GAGGGCAGTG TAGTCTGAGC
 1251 AGTACTCGTT GCTGCCGCGC GCGCCACCAG ACATAATAGC TGACAGACTA
 1301 ACAGACTGTT CCTTTCCTAG GGTCTTTTCT GCAGTCACCG TCGGTACCCT
 1351 CGACACGTGT GATCATCTAG AGGATCCCGG GCCGCAGATC TGCTGTGCCT
 1401 TCTAGTTGCC AGCCATCTGT TGTTTGCCCC TCCCCGTCG CTTCTTGAC
 1451 CCTGGAAGGT GCCACTCCA CTGTCCTTC CTAATAAAAT GAGGAAATG
 1501 CATCGCATG TCTGAGTAGG TGTCATTCTA TTCTGGGGGG TGGGGTGGGG
 1551 CAGGACAGCA AGGGGGAGGA TTGGGAAGAC AATAGCAGGC ATGCTGGGGA
 1601 TGCGGTGGGC TCTATGGCTA CCCAGGTGCT GAAGAATTGA CCCGTTCCCT
 1651 CCTGGGCCAG AAAGAAGCAG GCACATCCCC TTCTCTGTGA CACACCCTGT
 1701 CCACGCCCTT GGTTCCTAGT TCCAGCCCCA CTCATAGGAC ACTCATAGCT
 1751 CAGGAGGGCT CCGCCTTCAA TCCCACCCG TAAAGTACTT GGAGCGGTCT
 1801 CTCCCTCCCT CATCAGCCCA CCAAACCAA CCTAGCCTCC AAGAGTGGGA
 1851 AGAAATTTAA GCAAGATAGG CTATTAAGTG CAGAGGGAGA GAAATGCCT
 1901 CCAACATGTG AGGAAGTAAT GAGAGAAATC ATAGAATTTT AAGGCCATGA
 1951 TTTAAGGCCA TCAATGGCCTT AATCTTCCGC TTCTCTGCTC ACTGACTCGC
 2001 TGCGCTCGGT CGTTCGGCTG CCGCGAGCGG TATCAGCTCA CTCAAAGGCG
 2051 GTAATACGGT TATCCACAGA ATCAGGGGAT AACGCAGGAA AGAACATGTG
 2101 AGCAAAAGC CAGCAAAAGC CCAGGAACCG TAAAAGGCC GCGTTGCTGG
 2151 CTTTTTTCCA TAGGCTCCGC CCCCTGACG AGCATCACAA AAATGCACGC
 2201 TCAAGTCAGA GGTGGCGAAA CCCGACAGGA CTATAAAGAT ACCAGGCGTT
 2251 TCCCCCTGGA AGCTCCCTCG TGCGCTCTCC TGTTCGACC CTGCCGCTTA
 2301 CCGGATACCT GTCCGCTTTT CTCCCTTCGG GAAGCGTGGC GCTTTCAT
 2351 AGCTCACGCT GTAGGTATCT CAGTTCGGTG TAGGTCGTTT GCTCCAAGCT
 2401 GGGCTGTGTG CACGAACCC CCGTTCAGCC CGACCGCTGC GCCTTATCCG
 2451 GTAACATCG TCTTGAGTCC AACCCGGTAA GACACGACTT ATCGCCACTG
 2501 GCAGCAGCCA CTGGTAACAG GATTAGCAGA GCGAGGTATG TAGGCGGTGC
 2551 TACAGAGTTC TTGAAGTGGT GGCCTAAC TA CGGCTACACT AGAAGAACAG
 2601 TATTTGGTAT CTGCGCTCTG CTGAAGCCAG TTACCTTCGG AAAAAGAGTT
 2651 GGTAGCTCTT GATCCGGCAA ACAAACCACC GCTGGTAGCG GTGGTTTTTT
 2701 TGTTTGCAAG CAGCAGATTA CGCGCAGAAA AAAAGGATCT CAAGAAGATC
 2751 CTTTGATCTT TTCTACGGG TCTGACGCTC AGTGGAACGA AAACACAGT
 2801 TAAGGGATTT TGGTCAAGT ATTAFCAAA AGGATCTTCA CCTAGATCCT
 2851 TTTAAATTA AAATGAAGTT TTAATCAAT CTAAAGTATA TATGAGTAAA
 2901 CTTGGTCTGA CAGTTACCAA TGCTTAATCA GTGAGGCACC TATCTCAGCG
 2951 ATCTGTCTAT TTCGTTTATC CATAGTTGCC TGACTCGGGG GGGGGGGCG
 3001 CTGAGGTCTG CCTCGTGAAG AAGGTGTTGC TGACTCATAC CAGGCCTGAA
 3051 TCGCCCCATC ATCCAGCCAG AAAGTGAGGG AGCCACGGTT GATGAGAGCT
 3101 TTGTTGTAGG TGGACCAGTT GGTGATTTTG AACTTTTGCT TTGCCACGGA
 3151 ACGGTCTGCG TTGTCGGGAA GATGCGTGAT CTGATCCTTC AACTCAGCAA
 3201 AAGTTCGATT TATTCAACAA AGCCGCCGTC CCGTCAAGTC AGCGTAATGC

[0455]

3251 TCTGCCAGTG TTACAACCAA TTAACCAATT CTGATTAGAA AAATCATCG
 3301 AGCATCAAAT GAAACTGCAA TTTATTTCATA TCAGGATTAT CAATACCATA
 3351 TTTTTGAAAA AGCCGTTTCT GTAATGAAGG AGAAAACCTCA CCGAGGCAGT
 3401 TCCATAGGAT GGCAAGATCC TGGTATCGGT CTGCGATTCC GACTCGTCCA
 3451 ACATCAATAC AACCTATTAA TTTCCCCTCG TCAAAAATAA GGTTATCAAG
 3501 TGAGAAATCA CCATGAGTGA CGACTGAATC CGGTGAGAAT GGCAAAAGCT
 3551 TATGCATTTT TTTCCAGACT TGTTC AACAG GCCAGCCATT ACGCTCGTCA
 3601 TCAAAATCAC TCGCATCAAC CAAACCGTTA TTCATTCTGT ATTGCGCCTG
 3651 AGCGAGACGA AATACGCGAT CGCTGTTAAA AGGACAATTA CAAACAGGAA
 3701 TCGAATGCAA CCGGCGCAGG AACACTGCCA GCGCATCAAC AATATTTTCA
 3751 CCTGAATCAG GATATTCTTC TAATACCTGG AATGCTGTTT TCCC GGGGAT
 [0456] 3801 CGCAGTGGTG AGTAACCATG CATCATCAGG AGTACGGATA AAATGCTTGA
 3851 TGGTCGGAAG AGGCATAAAT TCCGTCAGCC AGTTTAGTCT GACCATCTCA
 3901 TCTGTAACAT CATTGGCAAC GCTACCTTTG CCATGTTTCA GAAACAATC
 3951 TGGCGCATCG GGCTTCCCAT ACAATCGATA GATTGTCGCA CCTGATTGCC
 4001 CGACATTATC GCGAGCCCAT TTATACCCAT ATAAATCAGC ATCCATGTTG
 4051 GAATTTAATC GCGGCCTCGA GCAAGACGTT TCCC GTTGAA TATGGCTCAT
 4101 AACACCCCTT GTATTACTGT TTATGTAAGC AGACAGTTTT ATTGTT CATG
 4151 ATGATATATT TTTATCTTGT GCAATGTAAC ATCAGAGATT TTGAGACACA
 4201 ACGTGGCTTT CCCCCCCCCC CCATTATTGA AGCATTATC AGGGTTATTG
 4251 TCTCATGAGC GGATACATAT TTGAATGTAT TTAGAAAAAT AAACAAATAG
 4301 GGGTTCCGCG CACATTTCCC CGAAAAGTGC CACCTGACGT CTAAGAAACC
 4351 ATTATTATCA TGACATTAAC CTATAAAAAAT AGGCGTATCA CGAGGCCCTT
 4401 TCGTC

[0457] 实施例8

[0458] 能够诱导广泛中和抗体应答的先导候选经优化抗原性埃博拉多肽

[0459] 在2014年西非爆发之后,研发针对埃博拉的疫苗引起了极大的兴趣。目前为止,目前临床开发中的计划已采用“经典”方法,使用来自在病毒载体骨架中表达的一至三种毒株的埃博拉和/或马尔堡病毒表面糖蛋白(GP)进行疫苗开发。抗原特异性仅来自于所包含的EBOV毒株:例如Merck使用来自Kikwit的GP;GSK使用Mayinga EBOV和Gulu SUDV毒株;Crucell和Profectus Biosciences二者使用马尔堡病毒以及扎伊尔和苏丹埃博拉毒株;Novavax方法独特地使用2014Makona EBOV毒株。

[0460] 下表1示出了流式细胞术测定结果,说明了抗体与代表所有埃博拉病毒物种(亚型)和马尔堡病毒的靶抗原结合的强度。结合强度由热图指示,其中红色(当以灰度等级观察时,最深的阴影)是非常强的结合,从橙色到黄色逐渐降低(当以灰度等级观察时,逐渐变浅的阴影),并且没有结合/等于阴性对照值为白色的。血清样品1至22取自用其他埃博拉病毒疫苗候选物免疫接种的个体。T2-4和T2-6是编码先导候选经优化抗原性埃博拉多肽的核酸疫苗,在使用取自经免疫接种豚鼠的血清样品的临床前阶段测试中与马尔堡候选T2-11组合。

[0461] 表1

[0462]

	埃博拉 Mayinga T1-1	埃博拉 苏丹 T1-2	埃博拉 本迪布焦 T1-3	埃博拉 塔伊森林 T1-4	埃博拉 雷斯顿 T1-5	马尔堡 T1-6	马尔堡 Ravn T0-1	埃博拉 Kikwit Kikwit	GFP con GFP con
样品 1	40,010	7,303	9,103	12,483	9,997	6,351	5,106	28,086	10,040
样品 2	28,510	7,139	6,406	10,790	6,728	5,447	4,103	27,544	7,705
样品 3	17,334	4,308	4,068	6,412	4,169	2,896	2,570	20,704	4,993
样品 4	38,712	7,183	6,245	8,809	6,621	4,239	3,629	35,045	8,031
样品 5	1,092,362	8,675	11,082	12,339	6,556	5,327	4,960	1,215,877	9,983
样品 6	42,743	8,351	8,319	10,038	4,622	2,803	2,488	41,380	5,212
样品 7	112,021	21,100	18,551	23,329	15,623	8,557	6,373	107,072	10,203
样品 8	102,788	42,134	28,650	54,319	13,858	6,584	4,878	58,113	7,517
样品 9	165,516	23,701	25,515	59,485	12,529	4,405	2,584	153,091	4,617
样品 10	131,842	21,760	14,868	21,924	10,612	7,508	3,682	95,488	8,455
样品 11	452,017	17,926	25,222	38,485	12,897	5,714	3,411	356,139	6,650
样品 12	73,588	12,226	11,916	28,503	7,268	3,415	2,626	58,328	4,556
样品 13	102,051	12,899	15,227	24,098	11,086	8,645	7,866	97,909	12,811
样品 14	113,028	19,731	24,986	37,834	15,667	12,576	9,206	142,277	19,049
样品 15	97,896	19,349	21,949	33,013	16,803	12,244	10,926	112,780	26,297
样品 16	57,387	8,085	9,380	14,060	5,866	3,545	3,882	70,301	5,107
样品 17	30,537	3,942	5,098	6,354	4,138	2,721	2,329	16,354	3,641
样品 18	22,323	7,173	8,845	8,560	11,392	4,424	4,007	18,922	6,802
样品 19	395,818	69,158	51,226	72,452	61,557	46,423	36,302	217,982	111,316
样品 20	17,428	6,195	5,918	10,031	5,386	3,589	3,386	12,405	5,454
样品 21	40,444	6,791	8,479	14,824	6,580	4,636	3,774	45,013	5,558
样品 22	15,314	3,510	4,424	5,254	4,206	2,868	2,725	14,592	3,654
T2-4 1	152,841	2,180,488	125,057	229,144	66,656	3,612	2,771	180,762	5,828
T2-4 2	547,668	2,616,185	285,405	491,530	402,143	6,334	5,728	664,060	9,198
T2-4 3	167,604	2,002,488	318,211	179,231	58,615	2,958	2,007	194,433	3,701
T2-4 4	156,012	1,781,655	136,079	241,383	70,518	8,964	6,316	179,352	10,413
T2-4 5	69,069	2,276,901	76,009	207,765	43,840	6,328	5,096	72,613	8,829
T2-4 6	202,425	2,489,959	179,014	321,938	119,511	9,847	6,712	168,017	11,929
T2-6 T2-11 1	236,142	213,846	148,494	634,665	111,867	1,463,752	2,233,784	261,750	9,654
T2-6 T2-11 2	220,176	454,052	194,068	353,923	134,736	1,122,815	1,100,621	204,227	10,497
T2-6 T2-11 3	442,525	347,609	418,041	558,831	127,288	1,571,834	1,278,907	522,335	10,925
T2-6 T2-11 4	427,268	239,099	186,990	881,740	81,274	1,294,796	836,995	400,394	10,945
T2-6 T2-11 5	414,571	378,611	296,297	1,030,040	86,352	1,111,935	1,624,647	473,564	10,033
T2-6 T2-11 6	36,653	50,512	42,979	239,226	53,948	1,498,011	789,981	52,957	7,050
无血清 hu	1,389	1,478	1,323	2,276	1,459	1,112	1,008	1,310	1,320
无血清 gp	1,637	1,395	1,078	1,402	1,724	851	1,006	894	1,131

[0463] 实施例9

[0464] 在埃博拉攻击模型中通过三价拉沙病毒、埃博拉病毒和马尔堡病毒疫苗 (Tri-LEMvac) 实现的保护

[0465] 我们已经开发了三价疫苗 (Tri-LEMvac), 其可产生针对未来出血热拉沙病毒、埃博拉病毒和马尔堡病毒变体的爆发的组合疫苗效力。

[0466] 我们已经从来自所有可获得的沙粒病毒和丝状病毒数据库的LASV (L) 以及EBOV (E) 和MARV (M) 的GPC开放阅读框中通过生物信息学设计了合成糖蛋白序列。这些保守序列由每种这些病毒的中和抗体和T细胞富含的表位组成。为了确保这些合成设计的LASV、EBOV 和MARV包膜具有功能性和抗原性, 将其表达为假型, 并且针对与一组广泛中和抗体的结合和中和二者进行了质量控制。本文中, 我们选择了疫苗来源的载体修饰的痘苗安卡拉 (MVA) 来构建三聚体LEM疫苗。

[0467] 修饰的痘苗安卡拉 (MVA) 疫苗平台是一种非复制株 (即在人细胞中不复制), 第三代天花疫苗和人临床试验中最先进的重组痘病毒疫苗载体之一 (Cottingham&Carroll, Vaccine, 2013, 31 (39): 4247-51)。MVA是一种稳健的载体系统, 能够共表达多至四个转基因, 以促进强效的启动子和稳定的插入位点 (Orubu et al, Pone, 2012, 7 (6) e0040167)。选

择MVA是因为:1) 其通过具有强且及时调节的启动子的兼容表达盒稳定表达多个独立ORF, 以在一个经济有效的疫苗批次中进行三价LEM疫苗接种的能力;2) 其在动物和人中诱导稳健的B和T细胞免疫应答的能力, 尤其是在用DNA或RNA载体进行初免或加强免疫时;以及3) 疫苗批次可进行热稳定化, 以用于在发展中国家在没有冷链的情况下进行存储和运输 (Frey et al, Vaccine, 2015, 33 (39) :5225-34)。该三价疫苗候选物的原理证明已通过以下方式证明:i) 用于独立的L、E和M GPC表达和表位呈递的盒式验证;以及ii) 通过丝状病毒攻击的临床前效力。攻击研究结果如图4所示。埃博拉攻击模型对未经过疫苗接种的豚鼠 (组1, 下排) 是致命的, 而所有经疫苗接种的豚鼠 (组2, 上排) 均受到保护 (左) 并继续增重 (右)。

[0468] 实施例10

[0469] 假型病毒中和测定

[0470] 图5示出了假型病毒中和测定的结果, 其举例说明了对代表所有埃博拉病毒物种和马尔堡病毒的假型病毒表面上表达的靶抗原的中和抗体应答的强度。中和的强度由热图指示, 其中红色 (当以灰度等级观察时, 最深的阴影) 是非常强的中和, 从橙色到黄色 (当以灰度等级观察时, 逐渐变浅的阴影) 降低, 并且没有中和/等于阴性对照值为白色。

[0471] T2-4和T2-6是各自编码先导候选经优化抗原性埃博拉多肽的核酸疫苗, 其在用取自经免疫接种豚鼠的血清样品进行测试的临床前阶段中与T2-11马尔堡候选组合。

[0472] 结果表明, 施用T2-6与T2-11疫苗插入物的组合导致免疫应答的宽度协同增加。

[0473] 实施例11

[0474] 抗体结合测定

[0475] 图6示出了抗体结合测定的结果。通过将其表面带有两种不同的组1甲型流感病毒糖蛋白 (H1大流行和季节性) 的两组细胞与合并的小鼠血清一起孵育来测量抗体结合。然后, 通过二抗检测任何结合的抗体, 并使用流式细胞仪记录结果。在用所有构建体进行疫苗接种之前和之后, 结合显著提高, 但是在用PBS (对照) 进行的疫苗接种之后则没有。总体而言, 在这两种情况下, DIOS疫苗候选物的表现优于来自COBRA的那些 (*);

[0476] 实施例12

[0477] 通过两种不同计算方法诱导的免疫应答的比较

[0478] 用25 μ g的编码HA基因抗原的四种独立的pEVAC质粒以两周的间隔对四组小鼠 (每组6只) 进行免疫接种5次, 所述pEVAC质粒通过根据本发明实施方案的方法 (DIOS) 或通过常规方法 (COBRA) 设计。

[0479] 对在其细胞表面上表达两种不同的组1甲型流感糖蛋白 (季节性H1N1和大流行起源H1N1) 的细胞进行基于抗体的FACS。这些用于测试来自用COBRA或DIOS HA基因抗原免疫接种的动物的小鼠血清。结果如图7所示。

[0480] 总体而言, DIOS HA基因抗原与COBRA HA基因抗原相当或明显优于后者 (** $p < 0.01$, *** $p < 0.001$)。

[0481] 实施例13

[0482] H7N9 (A/Shanghai/2/2013) 的跨HA组结合和假型中和

[0483] 我们测试了实施例12的DIOS-H1N1pdm疫苗 (与H1N1-COBRA相比, 产生更高水平的与大流行H1HA抗原的抗体结合) 是否可引起识别并结合相异的组2病毒HA (例如来自大流行潜力H7N9毒株A/Shanghai/2/2013的那些) 的抗体。

[0484] 图8示出了通过来自DIOS或COBRA DNA免疫接种的小鼠的血清的H7N9 (A/Shanghai2/2013) 的跨HA组结合(左图)和假型中和(右)的结果。通过假型中和数据(右)证实的H7结合数据(左)表明,与其他组相比,H1N1pdm疫苗接种的小鼠显示出最高的中和作用。与其他测试组相比,由DIOS-H1N1pdm疫苗引起的结合明显更多,并且与阳性对照广泛中和单克隆抗体F16 (Corti et al., 2011, 同上) 和CR9114 (Dreyfus et al, Science, 2012; 337 (6100): 1343-1348) 相当。

[0485] 这些结果支持以下结论:DIOS-H1N1pdm免疫原交叉中和H7,并且用通过本发明方法产生的疫苗可能实现跨HA组免疫保护。

[0486] 实施例14

[0487] 拉沙病毒糖蛋白

[0488] 该实施例描述了使用根据本发明的实施方案的方法产生的拉沙病毒糖蛋白祖先序列,以及对祖先序列的修饰以通过使结构稳定来改善其免疫原性。

[0489] 拉沙热是由被称为拉沙病毒 (LASV) 的旧世界 (OW) 沙粒病毒引起的出血性疾病。该病毒于1969年在尼日利亚首次分离,目前在西非流行。由于与拉沙出血热相关的高发病率和死亡率,因此LASV被归类为A类病原体。

[0490] 拉沙病毒是具有双分段基因组的包膜双义RNA病毒。病毒颗粒被介导病毒进入的成熟糖蛋白 (GP) 三聚体刺突覆盖。像其他1类病毒融合蛋白一样,包膜糖蛋白前体 (GPC) 被翻译为单个多肽,并被蛋白水解切割成三个亚基。加工首先通过细胞信号肽酶在内质网 (ER) 中发生。然后将GPC运输至顺面高尔基体,并通过细胞前蛋白转化酶枯草杆菌蛋白酶 kexin同工酶-1/位点1蛋白酶 (SKI-1/S1P) 进行加工,以产生非共价稳定信号肽 (SSP) /GP1/GP2异源三聚体。与其他I类融合蛋白不同,相对较长的GPC信号肽不会被降解;其具有正确运输和加工GP所必需的分子伴侣样功能。SSP与GP2的胞质结构域相互作用,并参与pH传感。GP1负责与细胞受体结合,而GP2在病毒进入过程中介导膜融合。

[0491] 使用根据本发明的一个实施方案的方法产生了针对谱系III和IV的拉沙病毒糖蛋白祖先序列 (L-10) (构建体1)。然后将修饰独立地引入亲代祖先序列 (构建体1) 中以提供:(A) SOSEP (构建体2); 和 (B) FLEP (构建体4), 以及与称为NtoK的聚糖敲除组合 (以提供构建体3和5), 以使在其他情况下柔性的异源三聚体稳定并阻止糖蛋白外部结构域从非共价连接的跨膜结构域解离。

[0492] (A) 在第207和360位引入两个半胱氨酸残基,以允许在GP的外部结构和跨膜结构域之间形成二硫桥 (SOS)。为了促进这两个结构域的完全切割,将第256至259位处的弗林蛋白酶切割位点从RRLL修饰为RRRR。在第329位处谷氨酸突变为脯氨酸 (EP) 阻止了结构重排,从而使蛋白质的柔性降低。

[0493] (B) 用具有序列256-GGGSGGGGS-265的柔性接头替换在外部结构域的C末端与跨膜结构域的N末端之间的弗林蛋白酶切割位点 (256-RRLL-259)。此外,在位置335处引入了如(A)中的EP突变。

[0494] 产生了两种设计的变体,分别在SOSEP-NtoK或FLEP-NtoK的第272或278位处另外包含天冬酰胺至赖氨酸突变,以使糖基化基序失活。在此位置的聚糖可能会阻止某些中和抗体(例如37.7H)的进入。

[0495] 构建体1:

[0496] 针对谱系III和IV的拉沙病毒糖蛋白祖先序列

[0497] (L-10=LASV III IV anc)

[0498] 氨基酸序列 (SEQ ID NO:18) :

MGQIVTFFQEVPHVIEEVMNIVLIALSLLAILKGLYNVATCGLIGLVTF L 50
LLCGRSCSTTLYKGVYELQ TLELNMETLNMTMPLSCTKNNSHHYIRVGNE 100
TGLELTLTNTS I INHKFCNLSDAHKKNLVDHALMSII STFHL SIPNFNQY 150
EAMSCDFNGGKISVQYNLSHSYAVDAANHC GTVANGVLQTFMRMAWGGSY 200
IALDSGRGNWDCIMTSYQYLI IQNTTWEHDHCQFSRPSPIGYLG LLSQRTR 250
DIYISRLLGTF TWTLS DSEGNETPGGYCLTRWMLIEAELKCFGNTAVAK 300
CNEKHDEEFC DMLR LFD FNKQAIRRLKAEAQMSIQLINKAVNALINDQLI 350
MKNHLR DIMGIPYCNYSKYWYLNHTITGKTS LPKWLVSNGSYLNETHFS 400
DDIEQQADNMITEMLQKEYMDRQ GKTP LGLVDL FV FSTSFY LISIFLHLV 450
KIPTHR HIVGKPCPKPHRLNHMGICSCGLYKQPGVPVRWKR*

[0500] DNA序列 (SEQ ID NO:19) :

1 ATGGGCCAGA TCGTGACATT CTTCCAAGAG GTGCCCCACG TGATCGAAGA
51 AGTGATGAAC ATCGTCCTGA TCGCCCTGAG CCTGCTGGCC ATCCTGAAGG
101 GCCTGTATAA TGTGGCCACC TGTGGCCTGA TCGGCCTGGT CACATTTCTG
151 CTGCTGTGCG GCAGAAGCTG CTCCACCACA CTGTATAAGG GCGTGTACGA
201 GCTGCAAACC CTGGAAGTGA ACATGGAAAC CCTGAACATG ACCATGCCTC
251 TGAGCTGCAC CAAGAACAAC AGCCACC ACT ACATCAGAGT GGGCAACGAG
301 ACAGGCCTCG AGCTGACCC T GACCAACACC AGCATCATCA ACCACAAGTT
351 CTGCAACCTG AGCGACGCC ACAAGAAGAA CCTGTACGAT CACGCCCTGA
401 TGAGCATCAT CTCCACCTTC CACCTGAGCA TCCCCA ACTT CAACCAGTAC
451 GAGGCCATGA GCTGCGACTT CAACGGCGGA AAGATCAGCG TGCAGTACAA
[0501] 501 TCTGAGCCAC AGCTATGCCG TGGACGCCGC CAATCATTGT GGAACAGTGG
551 CCAATGGCGT GCTCCAGACC TTCATGAGAA TGGCCTGGGG CGGCAGCTAT
601 ATCGCCCTGG ATTCTGGCAG AGGCAACTGG GACTGCATCA TGACCAGCTA
651 CCAGTACCTG ATCATCCAGA ACACCACCTG GGAAGATCAC TGCCAGTTCA
701 GCAGACCCTC TCCTATCGGA TACCTGGGCC TGCTGTCCCA GAGAACCCGG
751 GACATCTACA TCTCTAGACG GCTGCTGGGC ACCTTCACCT GGACACTGTC
801 TGATAGCGAG GGCAATGAGA CACCTGGCGG CTACTGTCTG ACCCGGTGGA
851 TGCTGATTGA GGCCGAGCTG AAGTGCTTCG GAAATACCGC CGTGGCCAAG
901 TGCAACGAGA AGCACGACGA GGAATTC TGC GACATGCTGC GGCTGTT CGA
951 TTTCAACAAG CAGGCCATCA GACGGCTGAA GGCCGAGGCT CAGATGTCCA
1001 TCCAGCTGAT CAACAAGGCC GTGAATGCC TGATTAACGA CCAGCTCATC
1051 ATGAAGAACC ACCTCAGGGA CATCATGGGC ATCCCTTACT GCAACTACAG
1101 CAAGTACTGG TATCTGAACC ACACCATCAC CGGCAAGACC AGCCTGCCTA
1151 AGTGCTGGCT GGTGTCCAAC GGCAGCTACC TGAACGAGAC AACTTCAGC
1201 GACGACATCG AGCAGCAGGC CGACAACATG ATCACCAGAA TGCTCCAGAA
[0502] 1251 AGAGTACATG GACCGGCAGG GCAAGACACC TCTGGGCCTT GTGGATCTGT
1301 TCGTGTT CAG CACCAGCTTC TACCTGATCT CTATCTTCCT GCACCTGGTC
1351 AAGATCCCCA CACACAGACA CATCGTGGGC AAGCCCTGTC CTAAGCCTCA
1401 CAGACTGAAC CATATGGGCA TCTGTAGCTG CGGCCTGTAC AACAGCCTG
1451 GCGTGCCAGT GCGGTGGAAG AGATAA

[0503] 构建体2:

[0504] 构建体1的SOSEP变体 (L-10-SQSEP)

[0505] 氨基酸序列 (SEQ ID NO:20) :

MGQIVTFFQEVPHVIEEVMNIVLIALSLLAILKGLYNVATCGLIGLVTF 50
 LLCGRSCSTTLYKGVYELQTLLELNMETLNMTMPLSCTKNNSHHYIRVNE 100
 TGLELTLTNTSIINHKFCNLSDAHKKNLYDHALMSIISTFHLSIPNFNQY 150
 EAMSCDFNGGKISVQYNLSHSYAVDAANHCCTVANGVLQTFMRMAWGGSY 200
 [0506] IALDSGCGNWDICIMTSYQYLI IQNTTWEDEHCQFSRPSPIGYLGLLSQRT 250
 DIYISR~~RR~~RRTFTWTLSDSEGNETPGGYCLTRWMLIEAELKCFGNTAVAK 300
 CNEKHDEEFCDMLR~~LF~~DFNKQAIRRLKAPAQMSIQLINKAVNALINDQLI 350
 MKNHLRDIMCIPYCNYSKYWYLNHTITGKTS~~LP~~KCWLVSNGSYLNETHFS 400
 DDIEQQADNMITEMLQKEYMDRQKTP~~LG~~LVDLFV~~F~~STSFYLISIFLHLV 450
 KIP~~TR~~HIVGKPCPKPHRLNHMGICSCGLYKQPGVPV~~R~~WKR*

[0507] DNA序列 (SEQ ID NO:21) :

1 ATGGGCCAGA TCGTGACATT CTTCCAAGAG GTGCCCCACG TGATCGAAGA
 51 AGTGATGAAC ATCGTCTGGA TCGCCCTGAG CCTGCTGGCC ATCCTGAAGG
 101 GCCTGTATAA TGTGGCCACC TGTGGCCTGA TCGGCCTGGT CACATTTCTG
 151 CTGCTGTGCG GCAGAAGCTG CTCCACCACA CTGTATAAGG GCGTGTACGA
 201 GCTGCAAACC CTGGA~~ACT~~G~~A~~ ACATGGAAAC CCTGAACATG ACCATGCCTC
 251 TGAGCTGCAC CAAGAACAAC AGCCACCACT ACATCAGAGT GGGCAACGAG
 301 ACAGGCCTCG AGCTGACCCT GACCAACACC AGCATCATCA ACCACAAGTT
 351 CTGCAACCTG AGCGACGCC ACAAGAAGAA CCTGTACGAT CACGCCCTGA
 401 TGAGCATCAT CTCCACCTC CACCTGAGCA TCCCCA~~ACT~~T CAACCAGTAC
 451 GAGGCCATGA GCTGCGACTT CAACGGCGGA AAGATCAGCG TGCAGTACAA
 501 TCTGAGCCAC AGCTATGCCG TGGACGCCGC CAATCATTGT GGAACAGTGG
 551 CCAATGGCGT GCTCCAGACC TTCATGAGAA TGGCCTGGGG CGGCAGCTAT
 601 ATCGCCCTGG ATTCTGGCTG TGGCAACTGG GACTGCATCA TGACCAGCTA
 651 CCAGTACCTG ATCATCCAGA ACACCACCTG GGAAGATCAC TGCCAGTTCA
 701 GCAGACCCTC TCCTATCGGA TACCTGGGCC TGCTGTCCCA GAGAACCCGG
 [0508] 751 GACATCTACA TCTCTCGCG GAGAAGAGGC ACCTTCACCT GGACACTGTC
 801 TGATAGCGAG GGCAATGAGA CACCTGGCGG C~~T~~ACTGTCTG ACCCGGTGGA
 851 TGCTGATTGA GGCCGAGCTG AAGTGCTTCG GAAATACCGC CGTGGCCAAG
 901 TGCAACGAGA AGCAGCAGCA GGAATTCCTG GACATGCTGC GGCTGTTCGA
 951 TTTCAACAAG CAGGCCATCA GACGGCTGAA GGCCCTGCT CAGATGTCCA
 1001 TCCAGCTGAT CAACAAGGCC GTGAATGCC TGATTAACGA CCAGCTCATC
 1051 ATGAAGAACC ACCTCAGGGA CATCATGTGC ATCCCTTACT GCAACTACAG
 1101 CAAGTACTGG TATCTGAACC ACACCATCAC CGGCAAGACC AGCCTGCCTA
 1151 AGTGCTGGCT GGTGTCCAAC GGCAGCTACC TGAACGAGAC A~~C~~ACTTCAGC
 1201 GACGACATCG AGCAGCAGGC CGACAACATG ATCACCAGAA TGCTCCAGAA
 1251 AGAGTACATG GACCGGCAGG GCAAGACACC TCTGGGCCTT GTGGATCTGT

 1301 TCGTGTTT~~C~~AG CACCAGCTTC TACCTGATCT CTATCTTCCT GCACCTGGTC
 1351 AAGATCCCCA CACACAGACA CATCGTGGGC AAGCCCTGTC CTAAGCCTCA
 1401 CAGACTGAAC CATATGGGCA TCTGTAGCTG CGGCCTGTAC AAACAGCCTG
 1451 GCGTGCCAGT GCGGTGGAAG AGATAA

[0509] 构建体3:

[0510] 具有N至K突变的构建体1的SOSEP变体(L-10-SQSEP-NtoK)

[0511] 氨基酸序列 (SEQ ID NO:22) :

MGQIVTFFQEVPHVIEEVMNIVLIALSLLAILKGLYNVATCGLIGLVTF 50
 LLCGRSCSTTLYKGVYELQTLLELNMETLNMTMPLSCTKNNSHHYIRVGNE 100
 TGLELTLTNTSIINHKFCNLSDAHKKNLYDHALMSIISTFHLSIPNFNQY 150
 EAMSCDFNGGKISVQYNLSHSYAVDAANHCCTVANGVLQTFMRMAWGGSY 200
 [0512] IALDSGCGNWDICIMTSYQYLIQNTTWEHDHCQFSRPSPIGYLGLLSQRT 250
 DIYISR^{RR}RGTTFTWTLSDSEGKETPGGYCLTRWMLIEAELKCFGNTAVAK 300
 CNEKHDEEFCDMLR^LFD^NKNQ^AIRRLKAPAQMSIQLINKAVNALINDQLI 350
 MKNHLRDIMCIPYCNYSKYWYLNHTITGKTS^LPKCWLVSNGSYLNETHFS 400
 DDIEQQADNMITEMLQKEYMDRQ^GKTP^LGLVD^LFV^FST^SFY^LISIFLHLV 450
 KIP^THR^HIV^GK^PCP^KPH^RLN^HMG^ICSC^GLY^KQ^PGP^VVR^WK^R*

[0513] DNA序列 (SEQ ID NO:23) :

1 ATGGGCCAGA TCGTGACATT CTTCCAAGAG GTGCCCCACG TGATCGAAGA
 51 AGTGATGAAC ATCGTCTGA TCGCCCTGAG CCTGCTGGCC ATCCTGAAGG
 101 GCCTGTATAA TGTGGCCACC TGTGGCTGA TCGGCCCTGGT CACATTTCTG
 151 CTGCTGTGCG GCAGAAGCTG CTCCACCACA CTGTATAAGG GCGTGTACGA
 201 GCTGCAAACC CTGGA^ACTGA ACATGGAAAC CCTGAACATG ACCATGCCTC
 251 TGAGCTGCAC CAAGAACAAC AGCCACCACT ACATCAGAGT GGGCAACGAG
 301 ACAGGCCTCG AGCTGACCCT GACCAACACC AGCATCATCA ACCACAAGTT
 351 CTGCAACCTG AGCGACGCC ACAAGAAGAA CCTGTACGAT CACGCCCTGA
 401 TGAGCATCAT CTCCACCTC CACCTGAGCA TCCCCA^ACTT CAACCAGTAC
 451 GAGGCCATGA GCTGCGACTT CAACGGCGGA AAGATCAGCG TGCAGTACAA
 501 TCTGAGCCAC AGCTATGCCG TGGACGCCGC CAATCATTGT GGAACAGTGG
 551 CCAATGGCGT GCTCCAGACC TTCATGAGAA TGGCCTGGGG CGGCAGCTAT
 601 ATCGCCCTGG ATTCTGGCTG TGGCAACTGG GACTGCATCA TGACCAGCTA
 651 CCAGTACCTG ATCATCCAGA ACACCACCTG GGAAGATCAC TGCCAGTTCA
 [0514] 701 GCAGACCCTC TCCTATCGGA TACCTGGGCC TGCTGTCCCA GAGAACCCGG
 751 GACATCTACA TCTCTCGCG GAGAAGAGGC ACCTTCACCT GGACACTGTC
 801 TGATAGCGAG GGCAAAGAGA CACCTGGCGG C^TACTGTCTG ACCCGGTGGA
 851 TGCTGATTGA GGCCGAGCTG AAGTGCTTCG GAAATACCGC CGTGGCCAAG
 901 TGCAACGAGA AGCACGACGA GGAAT^TCTGC GACATGCTGC GGCTGTTCGA
 951 TTTCAACAAG CAGGCCATCA GACGGCTGAA GGCCCTGCT CAGATGTCCA
 1001 TCCAGCTGAT CAACAAGGCC GTGAATGCC TGATTAACGA CCAGCTCATC
 1051 ATGAAGAACC ACCTCAGGGA CATCATGTGC ATCCCTTACT GCAACTACAG
 1101 CAAGTACTGG TATCTGAACC ACACCATCAC CGGCAAGACC AGCCTGCCTA
 1151 AGTGCTGGCT GGTGTCCAAC GGCAGCTACC TGAACGAGAC A^CACTTCAGC
 1201 GACGACATCG AGCAGCAGGC CGACAACATG ATCACCAGAA TGCTCCAGAA
 1251 AGAGTACATG GACCGGCAGG GCAAGACACC TCTGGGCCTT GTGGATCTGT
 1301 TCGTGTTT^CAG CACCAGCTTC TACCTGATCT CTATCTTCTT GCACCTGGTC
 1351 AAGATCCCCA CACACAGACA CATCGTGGGC AAGCCCTGTC CTAAGCCTCA
 1401 CAGACTGAAC CATATGGGCA TCTGTAGCTG CGGCCTGTAC AAACAGCCTG
 1451 GCGTGCCAGT GCGGTGGAAG AGATAA

[0515] 构建体4:

[0516] 构建体1的FLEP变体 (L-10-FLEP)

[0517] 氨基酸序列 (SEQ ID NO:24) :

MGQIVTFFQEVPHVIEEVMNIVLIALSLLAILKGLYNVATCGLIGLVTF 50
 LLCGRSCSTTLYKGVYELQTLLELNMETLNMTMPLSCTKNNSHHYIRVNE 100
 TGLELTLTNTSIINHKFCNLSDAHKKNLYDHALMSIISTFHLSIPNFNQY 150
 EAMSCDFNGGKISVQYNLSHSYAVDAANHCCTVANGVLQTFMRMAWGGSY 200
 IALDSGRGNWDCIMTSYQYLIQNTTWEHDHCQFSRPSPIGYLGLLSQRT 250
 [0518] DIYISGGGGSGGGSGTFTWTLSDSENETPGGYCLTRWMLIEAELKCFG 300
 NTAVAKCNEKHDEEFCDMLRFLDFNKQAIRRLKAPAQMSIQLINKAVNAL 350
 INDQLIMKNHLRDI MGIPYCNYSKYWYLNHTITGKTS LPKCWLVSNGSYL 400
 NETHFSDDIEQQADNMITEMLOKEYMDROGKTPLGLVDLDFVSTSFYLIS 450
 IFLHLVKIPTHRHI V GKPCPKPHRLNHMGICSCGLYKQPGVPVRWKR*

[0519] DNA序列 (SEQ ID NO:25) :

1 ATGGGCCAGA TCGTGACATT CTTCCAAGAG GTGCCCCACG TGATCGAAGA
 51 AGTGATGAAC ATCGTCTCTGA TCGCCCTGAG CCTGCTGGCC ATCCTGAAGG
 101 GCCTGTATAA TGTGGCCACC TGTGGCCTGA TCGCCCTGGT CACATTTCTG
 151 CTGCTGTGCG GCAGAAGCTG CTCCACCACA CTGTATAAGG GCGTGTACGA
 201 GCTGCAAACC CTGGAAGTGA ACATGGAAAC CCTGAACATG ACCATGCCTC
 251 TGAGCTGCAC CAAGAACAAC AGCCACCACT ACATCAGAGT GGGCAACGAG
 301 ACAGGCCTCG AGCTGACCC T GACCAACACC AGCATCATCA ACCACAAGTT
 351 CTGCAACCTG AGCGACGCC ACAAGAAGAA CCTGTACGAT CACGCCCTGA
 401 TGAGCATCAT CTCCACCTTC CACCTGAGCA TCCCCA ACTT CAACCAGTAC
 451 GAGGCCATGA GCTGCGACTT CAACGGCGGA AAGATCAGCG TGCAGTACAA
 501 TCTGAGCCAC AGCTATGCCG TGGACGCCGC CAATCATTGT GGAACAGTGG
 551 CCAATGGCGT GCTCCAGACC TTCATGAGAA TGGCCTGGGG CGGCAGCTAT
 601 ATCGCCCTGG ATTCTGGCAG AGGCAACTGG GACTGCATCA TGACCAGCTA
 651 CCAGTACCTG ATCATCCAGA ACACCACCTG GGAAGATCAC TGCCAGTTCA
 [0520] 701 GCAGACCCTC TCCTATCGGA TACCTGGGCC TGCTGTCCCA GAGAACCCGG
 751 GACATCTACA TCTCTGGCGG CGGAGGATCT GGCGGAGGTG GAAGTGGCAC
 801 CTTACCTGG AACTGTCTG ATAGCGAGGG CAATGAGACA CCTGGCGGCT
 851 ACTGTCTGAC CCGGTGGATG CTGATTGAGG CCGAGCTGAA GTGCTTCGGA
 901 AATACCGCCG TGGCCAAGTG CAACGAGAAG CACGACGAGG AATTCTGCGA
 951 CATGCTGCGG CTGTTGATT TCAACAAGCA GGCCATCAGA CGGCTGAAGG
 1001 CCCCTGCTCA GATGTCCATC CAGCTGATCA ACAAGGCCGT GAATGCCCTG
 1051 ATTAACGACC AGCTCATCAT GAAGAACCAC CTCAGGGACA TCATGGGCAT
 1101 CCCTTACTGC AACTACAGCA AGTACTGGTA TCTGAACCAC ACCATCACCG
 1151 GCAAGACCAG CCTGCCTAAG TGCTGGCTGG TGTCCAACGG CAGCTACCTG
 1201 AACGAGACAC ACTTCAGCGA CGACATCGAG CAGCAGGCCG ACAACATGAT
 1251 CACCGAGATG CTCCAGAAAG AGTACATGGA CCGGCAGGGC AAGACACCTC
 1301 TGGGCCTTGT GGATCTGTTT GTGTTGAGCA CCAGCTTCTA CCTGATCTCT
 1351 ATCTTCTGAC ACCTGGTCAA GATCCCCACA CACAGACACA TCGTGGGCAA
 1401 GCCCTGTCTT AAGCCTCACA GACTGAACCA TATGGGCATC TGTAGCTGCG
 1451 GCCTGTACAA ACAGCCTGGC GTGCCAGTGC GGTGGAAGAG ATAA

[0521] 构建体5:

[0522] 具有N至K突变的构建体1的FLEP变体 (L-10-FLEP-NtoK)

[0523] 氨基酸序列 (SEQ ID NO:26) :

MGQIVTFFQEVPHVIEEVMNIVLIALSLLAILKGLYNVATCGLIGLVTF 50
 LLCGRSCSTTLYKGVYELQTLLELNMETLNMTMPLSCTKNNSHHYIRVNE 100
 TGLELTLTNTSIINHKFCNLSDAHKKNLYDHALMSIISTFHLSIPNFNQY 150
 EAMSCDFNGGKISVQYNLSHSYAVDAANHCCTVANGVLQTFMRMAWGGSY 200
 IALDSGRGNWDCIMTSYQYLIQNTTWEDEHCQFSRPSPIGYLGLLSQRT 250
 DIYISGGGGSGGGSGTFTTWTLSDEGKETPGGYCLTRWMLIEAELKCFG 300
 NTAVAKCNEKHDEEFCDMLRFLDFNKQAIRRLKAPAQMSIQLINKAVNAL 350
 INDQLIMKNHLRDI MGIPYCNYSKYWYLNHETITGKTSPLKCWLVSNGSYL 400
 NETHESDDIEQQADNMITEMLOKEYMDROGKTPGLGLVDLFFVSTSFYLIS 450
 IFLHLVKIPTHRHI V GKPCPKPHRLNHMGICSCGLYKQPGVPVRWKR*

[0524]

[0525] DNA序列 (SEQ ID NO:27) :

1 ATGGGCCAGA TCGTGACATT CTTCCAAGAG GTGCCCCACG TGATCGAAGA
 51 AGTGATGAAC ATCGTCTCTGA TCGCCCTGAG CCTGCTGGCC ATCCTGAAGG
 101 GCCTGTATAA TGTGGCCACC TGTGGCCTGA TCGCCCTGGT CACATTTCTG
 151 CTGCTGTGCG GCAGAAGCTG CTCCACCACA CTGTATAAGG GCGTGTACGA
 201 GCTGCAAACC CTGGAAGTGA ACATGGAAAC CCTGAACATG ACCATGCCTC
 251 TGAGCTGCAC CAAGAACAAC AGCCACCACT ACATCAGAGT GGGCAACGAG
 301 ACAGGCCTCG AGCTGACCC T GACCAACACC AGCATCATCA ACCACAAGTT
 351 CTGCAACCTG AGCGACGCC ACAAGAAGAA CCTGTACGAT CACGCCCTGA
 401 TGAGCATCAT CTCCACCTC CACCTGAGCA TCCCCA ACTT CAACCAGTAC
 451 GAGGCCATGA GCTGCGACTT CAACGGCGGA AAGATCAGCG TGCAGTACAA
 501 TCTGAGCCAC AGCTATGCCG TGGACGCCGC CAATCATTGT GGAACAGTGG
 551 CCAATGGCGT GCTCCAGACC TTCATGAGAA TGGCCTGGGG CGGCAGCTAT
 601 ATCGCCCTGG ATTCTGGCAG AGGCAACTGG GACTGCATCA TGACCAGCTA
 651 CCAGTACCTG ATCATCCAGA ACACCACCTG GGAAGATCAC TGCCAGTTCA
 701 GCAGACCCTC TCCTATCGGA TACCTGGGCC TGCTGTCCA GAGAACCCGG
 751 GACATCTACA TCTCTGGCGG CGGAGGATCT GGCGGAGGTG GAAGTGGCAC
 801 CTTACCTGG AACTGTCTG ATAGCGAGGG CAAAGAGACA CCTGGCGGCT
 851 ACTGTCTGAC CCGGTGGATG CTGATTGAGG CCGAGCTGAA GTGCTTCGGA
 901 AATACCGCCG TGGCCAAGTG CAACGAGAAG CACGACGAGG AATTCTGCGA
 951 CATGCTGCGG CTGTTGATT TCAACAAGCA GGCCATCAGA CGGCTGAAGG
 1001 CCCCTGCTCA GATGTCCATC CAGCTGATCA ACAAGGCCGT GAATGCCCTG
 1051 ATTAACGACC AGCTCATCAT GAAGAACCAC CTCAGGGACA TCATGGGCAT
 1101 CCCTTACTGC AACTACAGCA AGTACTGGTA TCTGAACCAC ACCATCACCG
 1151 GCAAGACCAG CCTGCCTAAG TGCTGGCTGG TGTCCAACGG CAGCTACCTG
 1201 AACGAGACAC ACTTCAGCGA CGACATCGAG CAGCAGGCCG ACAACATGAT
 1251 CACCGAGATG CTCCAGAAAG AGTACATGGA CCGGCAGGGC AAGACACCTC
 1301 TGGGCCTTGT GGATCTGTTT GTGTTGAGCA CCAGCTTCTA CCTGATCTCT
 1351 ATCTTCTGTC ACCTGGTCAA GATCCCCACA CACAGACACA TCGTGGGCAA
 1401 GCCCTGTCTT AAGCCTCACA GACTGAACCA TATGGGCATC TG TAGCTGCG
 1451 GCCTGTACAA ACAGCCTGGC GTGCCAGTGC GGTGGAAGAG ATAA

[0526]

[0527] 实施例15

[0528] 拉沙病毒核蛋白

[0529] 该实施例描述了使用根据本发明的实施方案的方法产生的拉沙病毒核蛋白祖先序列。

[0530] 构建体6:

[0531] 尼日利亚拉沙分离株的拉沙病毒核蛋白祖先序列

[0532] (L-NP-1=L-NP-CovAnc-1 N)

[0533] 氨基酸序列 (SEQ ID NO:28) :

MSASKEVKSFLLWTQSLRRELSGYCSNIKLOVVKDAQALLHGLDFSEVSNV 50
 ORLMRQKRDDSDLKRLRDLNQAVNNLVELKSTQOKSILRVGTLTSDDLL 100
 TLAADLEKLKSKVIRTERPLSSGVYMGNLSTQOLEQRRALLNMIGMVGGA 150
 QGTQFGRDGVVRVWDVKNPDLNNOFGTMSLTLACLTKQGQVDLNDVAVL 200
 ALTDLGLIYTAKYPNSSDLDRLSQSHPIILNMVDTKKSSLNISGYNFSLGA 250
 [0534] AVKAGACMLDGGNMLETIKVTPTMDGILKSILKVKKSIGMFVSDTPGER 300
 NPYENILYKICLSGDGWPYIASRTSIVGRAWENTTVDLESDGKPKQKVGTA 350
 GSNKSLQSAGFPTGLTYSQLMTLKDMMQLDPSAKTWIDIEGRPEDPVEI 400
 ALYQPMSCYIHHFFREPTDLKQFKQDAKYSHGIDVADLFPAQPGLTSAVI 450
 EALPRNMVLTQCGSDDIKRLLDSSQRRDIKLIIDIALSKADSRRFENAVWD 500
 QCKDLCHMHTGVVVEKKRGGKEEITPHCALMDCIMYDAAVSGGLNIPVL 550
 RAVLPRDMVFRITSSPKVVL*

[0535] DNA序列 (SEQ ID NO:29) :

1 ATGAGCGCCA GCAAAGAAGT GAAAAGCTTC CTCTGGACCC AGAGCCTGCG
 51 GAGAGAGCTG TCTGGCTACT GCTCCAACAT CAAGCTCCAG GTGGTCAAGG
 101 ACGCCAGGC TCTGCTGCAT GGCCTGGATT TCAGCGAGGT GTCCAACGTG
 151 CAGCGGCTGA TGAGAAAGCA GAAGCGGGAC GACAGCGACC TGAAGAGACT
 201 GAGGGATCTG AACCAGGCCG TGAACAACCT GGTGGAAGT AAGTCTACCC
 251 AGCAGAAATC CATCCTGAGA GTGGGCACCC TGACCAGCGA CGATCTGCTG
 301 AACTGGCCG CCGATCTGGA AAAGCTGAAG TCCAAAGTGA TCCGGACCGA
 351 GAGGCCACTG TCTAGCGGAG TGTACATGGG CAACCTGAGC ACCCAGCAGC
 401 TGGAACAGAG AAGGGCCCTG CTGAACATGA TCGGCATGGT TGGAGGCGCC
 451 CAGGGAACAC AGCCTGGAAG AGATGGTGTG GTCAGAGTGT GGGACGTGAA
 501 GAACCCCGAC CTGCTCAACA ACCAGTTCGG CACCATGCCT TCTCTGACCC
 551 TGGCCTGCCT GACAAAGCAG GGCCAAGTGG ACCTGAACGA TGCCGTGCTG
 601 GCTCTGACTG ATCTGGGCTT GATCTACACC GCCAAGTATC CCAACAGCTC
 651 CGACCTGGAC AGGCTGAGCC AGTCTCACCC CATCCTGAAC ATGGTGGACA
 701 CCAAGAAGTC CAGCCTGAAC ATCAGCGGCT ACAACTTCTC TCTGGGCGCT
 751 GCCGTGAAAG CCGGCGCTTG TATGCTTGAC GCGGCAACA TGCTGGAAC
 801 CATCAAAGTG ACCCCTCAGA CCATGGACGG CATCCTGAAA AGTATCCTGA
 851 AAGTGAAGAA ATCCCTGGGC ATGTTCTGTG CCGACACACC CGGCGAGAGA
 [0536] 901 AACCCCTACG AGAACATCCT GTACAAGATT TGCCTGAGCG GCGACGGCTG
 951 GCCCTATATC GCCAGCAGAA CATCTATCGT GGGCAGAGCT TGGGAGAACA
 1001 CCACCGTGGA CCTGGAATCC GATGGCAAGC CTCAGAAAGT GGGCACAGCC
 1051 GGCAGCAACA AGAGCCTCCA GTCTGCCGGA TTTCTACCG GCCTGACATA
 1101 CAGCCAGCTG ATGACCCTGA AGGACAGCAT GATGCAGCTG GACCCTAGCG
 1151 CCAAGACCTG GATCGACATT GAGGGCAGAC CCGAGGATCC CGTGGAAATC
 1201 GCTCTGTACC AGCCTATGAG CGGCTGCTAT ATCCACTTCT TCAGAGAGCC

1251 CACCGATCTG AAGCAGTTCA AGCAGGACGC CAAGTACAGC CACGGAATCG
 1301 ACGTGCCGA TCTGTTCCCA GCTCAGCCAG GACTGACATC CGCCGTGATT
 1351 GAAGCCCTGC CTAGAAACAT GGTGCTGACC TGTCAGGGCA GCGACGACAT
 1401 CAAGAGACTG CTGGACAGCC AGGGCAGAAG AGATATCAAG CTGATCGATA
 1451 TCGCCCTGAG CAAGGCCGAC TCTCGGAGAT TCGAAAACGC CGTGTGGGAC
 1501 CAGTGCAAGG ACCTGTGTCA CATGCACACA GGCGTGGTGG TGGAAGAA
 1551 GAAGCGCGGA GGCAAAGAGG AAATCACCCC TCACTGCGCC CTGATGGACT
 1601 GCATTATGTA TGACGCCGCC GTGTCTGGCG GCCTGAATAT CCCTGTTCTG
 1651 AGAGCCGTGC TGCCCCGCGA CATGGTGTGTT AGAACAAGCA GCCCAAGGT
 1701 GGTGCTCTGA

[0537] 实施例16

[0538] 拉沙病毒核蛋白

[0539] 该实施例描述了使用根据本发明的实施方案的方法产生的拉沙病毒核蛋白祖先

序列。

[0540] 构建体7:

[0541] 塞拉利昂分离株的拉沙病毒核蛋白祖先序列

[0542] (L-NP-1=L-NP-CovAnc-2 SL)

[0543] 氨基酸序列 (SEQ ID NO:30):

MSASKEIKSFLWTQSLRRELSGYCSNIKLQVVKDAQALLHGLDFSEVSNV 50
 QRLMRKERRDDNDLKRRLRDLNQAANNLVELKSTQOKSILRVGTLTSDDLL 100
 ILAADLEKLSKVTRTERPLSAGVYMGNLSSQQLDORRALLNMIGMSGGN 150
 QGARAGRDRGVVRVWDVKNAELLNQFGTMPSLTLACLTKQGQVDLNDVQ 200
 ALTDLGLIYTAKYPNTSDLRLTQSHPIILNMIDTKKSSLNISGYNFSLGA 250
 AVKAGACMLDGGNMLETIKVPQTMGILKSILKVKKALGMFISDTPGER 300
 NPYENILYKICLSGDGPYIASRTSITGRAWENTVVDLESDGKPKQAGSN 350
 NSNKSLSQAGFTAGLTYSQLMTLKDAMLQLDPNAKTWMDIEGRPEPVEI 400
 ALYQPSGCIYHFFREPTDLKQFKQDAKYSHGIDVTDLFAAQPLTSAVI 450
 DALPRNMVITCQGSDDIRKLLSQGRKDIKLIDIALSKTDSRKYENAVWD 500
 QYKDLCHMHTGVVVEKKRGGKEEITPHCALMDCIMFDAAVSGGLNTSVL 550
 RAVLPRDMVFRTSTPRVVL*

[0545] DNA序列 (SEQ ID NO:31):

1 ATGAGCGCCA GCAAAGAGAT CAAGAGCTTC CTGTGGACCC AGAGCCTGCG
 51 GAGAGAGCTG TCTGGCTACT GCTCCAACAT CAAGCTCCAG GTGGTCAAGG
 101 ACGCCCAGGC TCTGCTGCAT GGCCTGGATT TCAGCGAGGT GTCCAACGTG
 151 CAGCGGCTGA TGCGGAAAGA GAGAAGGGAC GACAACGACC TGAAGCGGCT
 201 GAGGGATCTG AACCAGGCCG TGAACAACCT GGTGGAAGT AAGTCTACCC
 251 AGCAGAAATC CATCCTGAGA GTGGGCACCC TGACCAGCGA CGATCTGCTG
 301 ATTCTGGCCG CCGACCTGGA AAAGCTGAAG TCCAAAGTGA CCCGGACCGA
 351 GAGGCCACTG TCTGCTGGTG TCTACATGGG CAACCTGAGC AGCCAGCAGC
 401 TGGATCAGAG AAGGGCCCTG CTGAACATGA TCGGCATGAG CGGCGGAAAT
 451 CAGGGCGCTA GAGCTGGCAG AGATGGCCTC GTCAGAGTGT GGGACGTGAA
 501 GAATGCCGAG CTGCTCAACA ACCAGTTCGG CACCATGCCT AGCCTGACAC
 551 TGGCCTGCCT GACAAAGCAG GGCCAAGTGG ACCTGAACGA TGCTGTGCAG
 601 GCCCTGACTG ATCTGGGCCT GATCTACACC GCCAAGTATC CCAACACCAG
 651 CGACCTGGAC AGACTGACCC AGTCTCACC CATCCTGAAT ATGATCGACA
 701 CCAAGAAGTC CAGCCTGAAC ATCAGCGGCT ACAACTTCTC TCTGGGCGCT
 751 GCCGTGAAAG CCGGCGCTTG TATGCTTGAC GCGGCAACA TGCTGGAAAC
 801 CATCAAGGTG TCCCCACAGA CCATGGACGG CATCCTGAAA AGTATCCTGA
 851 AAGTGAAGAA AGCCCTGGGC ATGTTTCATCA GCGACACCCC TGCGGAGAGA
 901 AACCCCTACG AGAACATCCT GTACAAGATT TGCCCTGAGCG GCGACGGCTG
 951 GCCCTATATC GCCAGCAGAA CCAGCATTAC CGGCAGAGCT TGGGAGAACA
 1001 CCGTGGTGGA TCTGGAAAGC GACGGCAAGC CTCAGAAGGC CGGCAGCAAC
 1051 AACTCCAACA AGAGCCTCCA GTCCGCCGGC TTCACAGCCG GCCTGACATA
 1101 TAGCCAGCTG ATGACCCTGA AGGACGCCAT GCTGCAACTG GACCCCAATG
 1151 CCAAGACCTG GATGGACATC GAGGGCAGAC CTGAGGACCC TGTGGAAATC
 1201 GCCCTGTACC AGCCTAGCTC CGGCTGCTAT ATCCACTTCT TCAGAGAGCC
 1251 CACCGATCTG AAGCAGTTCA AGCAGGACGC CAAGTACAGC CACGGCATCG
 1301 ACGTGACCGA TCTGTTTGCT GCTCAGCCCG GACTGACCTC CGCCGTGATT
 1351 GATGCCCTGC CTCGGAACAT GGTCATCACC TGTCAGGGCA GCGACGACAT
 1401 CCGGAAGCTG CTGGAATCTC AGGGCAGAAA GGATATCAAG CTGATCGATA
 1451 TCGCCCTGAG CAAGACCGAC AGCCGGAAGT ACGAAAACGC CGTGTGGGAC
 1501 CAGTACAAGG ACCTGTGCCA CATGCACACA GCGGTGGTGG TGGAAGAAAGAA
 1551 GAAGCGCGGA GGCAAAGAGG AAATCACCCC TCACTGCGCT CTGATGGACT
 1601 GCATCATGTT TGACGCCGCC GTGTCTGGCG GCCTGAATAC CTCTGTTCTG
 1651 AGAGCCGTGC TGCCAGAGA CATGGTGTTT AGAACAAGCA CCCCTAGAGT
 1701 GGTGCTCTGA

- <110> The Chancellor, Masters and Scholars of the University of Cambridge; University of Westminster; Universitat Regensburg
- <120> 疫苗和方法
- <130> P/79311.W001
- <150> GB 1815956.6
- <151> 2018-09-28
- <160> 31
- <170> PatentIn version 3.5
- <210> 1
- <211> 1578
- <212> DNA
- <213> 人工序列
- <220>
- <223> 埃博拉苏丹祖先 (T2-4), 未优化

[0001]

```

<400> 1
atggggggtc ttagcctact ccaattgcc agggacaaat ttcggaaaag ctctttcttt      60
gtttgggtca tcatcttatt ccaaaaggcc tttccatgc ctttgggtgt tgtgactaac      120
agcactttag aagtaacaga gattgaccag ctagtctgca aggatcatct tgcattccact      180
gaccagctga aatcagttgg tctcaacctc gaggggagcg gagtatctac tgatatccca      240
tctgcaacaa agcgttgggg cttcagatct ggtgttcctc ccaaggtggt cagctatgaa      300
gcggggagaat gggctgaaaa ttgctacaat cttgaaataa agaagccgga cgggagcgaa      360
tgcttaccct caccgcaga tgggtgcaga ggctttccaa ggtgccgcta gttcacaaa      420
gccaagga cgggcccctg cccaggtgac tacgccttc acaaggatgg agctttcttc      480
ctctatgaca ggctggcttc aactgtaatt tacagaggag tcaattttgc tgaggggta      540
attgcattct tgatattggc taaacaaaa gaaacgttc ttcagtcacc cccattcga      600
gaggcagtaa actacactga aaatacatca agtattatg ccacatccta ctggagat      660
gaaatcgaaa attttgggtc tcaacactcc acgaccttt tcaaaattga caataaact      720
ttgttcgctc tggacagccc ccacacgcct cagttccttt tccagctgaa tgataccatt      780
caccttcacc aacagttgag caacacaact gggagactaa tttggacact agatgctaat      840
atcaatgctg atattgggtg atgggctttt tgggaaaata aaaaaatct ctccgaacaa      900
ctacgtggag aagagctgtc tttcgaagct ttatcgtca caacagcggg taaaactgtc      960
ttgccacagg agtccacaag caacggctca ataactcaa cagtaacagg gattcttggg      1020
agtcttgggc ttcgaaaacg cagcagaaga caagttaaca ccaaagccac gggtaaatgc      1080
aatcccaact tacactactg gactgcacaa gaacaacata atgctgctgg gattgcctgg      1140
atcccgtact ttggaccggg tgcggaaggc atatactg aaggcctgat gcataacaa      1200
aatgccttag tctgtggact taggcaactt gcaaatgaaa caactcaagc tctgcagctt      1260
ttcttaagag ccacaacgga gctgcggaca tataccatac tcaataggaa ggccatagat      1320
ttcttctgac gacgatgggg cgggacatgc aggatcctgg gaccagattg ttgattgag      1380
ccacatgatt ggacaaaaaa catcactgat aaaatcaacc aatcatcca tgatttcac      1440
gacaaccctt tacctaatac ggataatgat gataattggt ggacgggctg gagacagtg      1500
atccctgcag gaataggcat tactggaatt attattgcaa ttattgctct tctttgogtt      1560
    
```

	tgcaagctgc tttgctag	1578
	<210> 2	
	<211> 1578	
	<212> DNA	
	<213> 人工序列	
	<220>	
	<223> 埃博拉苏丹祖先 (T2-4); 经基因优化	
	<400> 2	
	atgggaggac tgtctctgct gcaactgccc cgggacaagt tccggaagtc cagcttcttc	60
	gtgtgggtca tcacctgtt ccagaaagcc ttcagcatgc ccctgggcgt cgtgaccaat	120
	agcacactgg aagtgaccga gatcgaccag ctctgtgtca aggatcacct ggccagcacc	180
	gatcagctga agtctgtggg actgaatctg gaaggcagcg gcgtgtccac agatatccct	240
	agcgcacca agagatgggg ctttagaagc ggagtgcctc ctaagggtgt gtcttatgaa	300
	gccggcgagt gggccgagaa ctgctacaac ctggaaatca agaagcccca cggcagcgag	360
	tgtctgcctc ctccacctga tggcgtcaga ggcttccta gatgcagata cgtgcacaag	420
	gccaaggca caggaccctg tcctggcgat tacgccttc acaaggacgg cgcctttttc	480
	ctgtacgac ggctggcctc caccgtgatc tacagaggcg ttaactttgc cgagggcgtg	540
	atcgccttcc tgatcctggc caagcctaaa gagacattcc tgcaaagccc tccaatccgc	600
	gaggccgtga actacacaga gaacaccagc agctactacg ccaccagcta cctggaatac	660
	gagatcgaga atttcggcgc ccagcacagc accacactgt tcaagatcga caacaacacc	720
[0002]	ttcgtgcggc tggacagacc ccacacacct cagtttctgt tccagctgaa cgacaccatc	780
	catctgcatc agcagctgag caacaccacc ggcagactga tttggaccct ggacgccaac	840
	atcaacgccg acattggaga gtgggccttt tgggagaaca agaagaacct gagcgaacag	900
	ctgagaggcg aggaactgag ctttgaggcc ctgtctctga ccaccgccgt gaaaacagtg	960
	ctgcctcaag agtccaccag caacggcctg atcacaagca cagtacagc catcctgggc	1020
	agcctgggcc tgagaaaaag gtccagacgg caagtgaata ccaaggccac cggcaagtgc	1080
	aacccaacc tgcactattg gacagcccaa gagcagcaca atgccgccgg aatcgcctgg	1140
	attccttatt ttggacctgg cggcgagggc atctataccg agggactgat gcacaaccag	1200
	aacgccctcg tgtgtggact gagacagctg gccaatgaga caacacagc cctccagctg	1260
	tttctgagag ccaccaccga gctgagaacc tacacatcc tgaaccgga ggccatcgac	1320
	tttctgtga gaagatgggg cggcacctgt agaatcctgg gacctgatg ctgcatcgag	1380
	ccccacgact ggaccaagaa catcaccgac aagatcaacc agatcatcca cgacttcac	1440
	gacaaccctc tgcctaacca ggacaacgac gacaattggt ggacaggctg cggcagtg	1500
	attcctgccg gaattggcat caccggcatc atcattgcca ttatcgccct gctgtgtgtg	1560
	tgcaagctgc tgtgttga	1578
	<210> 3	
	<211> 525	
	<212> PRT	
	<213> 人工序列	
	<220>	
	<223> 埃博拉苏丹祖先 (T2-4)	

<400> 3

Met Gly Gly Leu Ser Leu Leu Gln Leu Pro Arg Asp Lys Phe Arg Lys
1 5 10 15

Ser Ser Phe Phe Val Trp Val Ile Ile Leu Phe Gln Lys Ala Phe Ser
20 25 30

Met Pro Leu Gly Val Val Thr Asn Ser Thr Leu Glu Val Thr Glu Ile
35 40 45

Asp Gln Leu Val Cys Lys Asp His Leu Ala Ser Thr Asp Gln Leu Lys
50 55 60

Ser Val Gly Leu Asn Leu Glu Gly Ser Gly Val Ser Thr Asp Ile Pro
65 70 75 80

Ser Ala Thr Lys Arg Trp Gly Phe Arg Ser Gly Val Pro Pro Lys Val
85 90 95

Val Ser Tyr Glu Ala Gly Glu Trp Ala Glu Asn Cys Tyr Asn Leu Glu
100 105 110

Ile Lys Lys Pro Asp Gly Ser Glu Cys Leu Pro Pro Pro Asp Gly
115 120 125

Val Arg Gly Phe Pro Arg Cys Arg Tyr Val His Lys Ala Gln Gly Thr
130 135 140

[0003]

Gly Pro Cys Pro Gly Asp Tyr Ala Phe His Lys Asp Gly Ala Phe Phe
145 150 155 160

Leu Tyr Asp Arg Leu Ala Ser Thr Val Ile Tyr Arg Gly Val Asn Phe
165 170 175

Ala Glu Gly Val Ile Ala Phe Leu Ile Leu Ala Lys Pro Lys Glu Thr
180 185 190

Phe Leu Gln Ser Pro Pro Ile Arg Glu Ala Val Asn Tyr Thr Glu Asn
195 200 205

Thr Ser Ser Tyr Tyr Ala Thr Ser Tyr Leu Glu Tyr Glu Ile Glu Asn
210 215 220

Phe Gly Ala Gln His Ser Thr Thr Leu Phe Lys Ile Asp Asn Asn Thr
225 230 235 240

Phe Val Arg Leu Asp Arg Pro His Thr Pro Gln Phe Leu Phe Gln Leu
245 250 255

Asn Asp Thr Ile His Leu His Gln Gln Leu Ser Asn Thr Thr Gly Arg
260 265 270

Leu Ile Trp Thr Leu Asp Ala Asn Ile Asn Ala Asp Ile Gly Glu Trp
275 280 285

Ala Phe Trp Glu Asn Lys Lys Asn Leu Ser Glu Gln Leu Arg Gly Glu
290 295 300

Glu Leu Ser Phe Glu Ala Leu Ser Leu Thr Thr Ala Val Lys Thr Val
 305 310 315 320

Leu Pro Gln Glu Ser Thr Ser Asn Gly Leu Ile Thr Ser Thr Val Thr
 325 330 335

Gly Ile Leu Gly Ser Leu Gly Leu Arg Lys Arg Ser Arg Arg Gln Val
 340 345 350

Asn Thr Lys Ala Thr Gly Lys Cys Asn Pro Asn Leu His Tyr Trp Thr
 355 360 365

Ala Gln Glu Gln His Asn Ala Ala Gly Ile Ala Trp Ile Pro Tyr Phe
 370 375 380

Gly Pro Gly Ala Glu Gly Ile Tyr Thr Glu Gly Leu Met His Asn Gln
 385 390 395 400

Asn Ala Leu Val Cys Gly Leu Arg Gln Leu Ala Asn Glu Thr Thr Gln
 405 410 415

Ala Leu Gln Leu Phe Leu Arg Ala Thr Thr Glu Leu Arg Thr Tyr Thr
 420 425 430

Ile Leu Asn Arg Lys Ala Ile Asp Phe Leu Leu Arg Arg Trp Gly Gly
 435 440 445

[0004] Thr Cys Arg Ile Leu Gly Pro Asp Cys Cys Ile Glu Pro His Asp Trp
 450 455 460

Thr Lys Asn Ile Thr Asp Lys Ile Asn Gln Ile Ile His Asp Phe Ile
 465 470 475 480

Asp Asn Pro Leu Pro Asn Gln Asp Asn Asp Asp Asn Trp Trp Thr Gly
 485 490 495

Trp Arg Gln Trp Ile Pro Ala Gly Ile Gly Ile Thr Gly Ile Ile Ile
 500 505 510

Ala Ile Ile Ala Leu Leu Cys Val Cys Lys Leu Leu Cys
 515 520 525

<210> 4
 <211> 1581
 <212> DNA
 <213> 人工序列

<220>
 <223> 埃博拉病毒全球祖先 (T2-6) 未优化

<400> 4
 atggggggtg gatccagact tctccaattg ccccggaac gctttcggaa aacctcattc 60
 tttgtttggg taatcatcct attcAAAAA gccttttcca tgccattggg tgttgtaacc 120
 aacagcactc taaaagtaac agaaattgac caattggttt gccgggacaa actttcatcc 180
 acaagtgcag tgaatcagt tgggctgaat ctggaaggga atggagtgc aactgatgctc 240
 ccatcagcaa caaacgatg gggcttccga tctggtgttc ctccaaggt ggtcagctat 300

	gaagctggag aatgggctga aaattgctac aatctggaaa tcaagaagcc agacgggagt	360
	gaatgcctac ctccaccgcc agacgggtga agaggcttcc ccagggtccg ctatgtccac	420
	aaagtccaag gaacagggcc gtgtcctggt gacttcgcct tccacaaaga tggagcttc	480
	ttcctgtatg atagactggc ticaactgtc atttaccgag ggacaacttt tgctgaaggt	540
	gtcgttgc at tttgatcct gcccaaacct aaaaaggact tttccaate accccaata	600
	ctgagccgg taaacaccac agaagatcca tcaagtact acaccacatc aacacttagc	660
	tatgagattg acaattttgg ggccaataaa actaaaactc tttcaaagt tgacaatcac	720
	acttatgtgc aactagaccg accacacaca ccacagttcc ttgtccagct caatgaaacc	780
	attcatacaa ataaccgtct aagcaacacc acaggagac taatttggac attagatcct	840
	aaaattgata ccgacattgg tgagtgggcc ttctgggaaa ataaaaaaaa cttctccaaa	900
	caacttcgtg gagaagagtt gtctttcaaa gctctatcaa caaaaactgg agtaacgca	960
	gtagacactg acgaatcaag caaacctggc ctaattacca acacagtaag aggggttgc	1020
	gatttactga gcccttggag aagaaaaaga agacaagtca acccaaacac acaataaaa	1080
	tgcaacccaa acctacacta ttggacagcc caagatgaag gtgctgccgt tggattagcc	1140
	tggatcccat acttcggacc agcagcagaa ggcaattaca ctgaaggaat aatgcataat	1200
	caaatgggt taatctgtgg gctgaggcag ctggccaatg aaacgactca agctcttcaa	1260
	ttattcttga gggccacaac ggagctgcgg acttactcta tactcaatag aaaagccatt	1320
	gatttcttc tccaacgatg gggaggaaca tgccgcatct taggaccaga ttgttcatt	1380
	gagccacatg attggacaaa aaacattact gataaaatta accaaatcat acatgatttt	1440
[0005]	attgacaacc ctctaccaga tcaggacgat gatgacaatt ggtggacagg ctggagacaa	1500
	tggatccctg ctggaattgg aattactgga gttataattg caattatagc tctactttgt	1560
	atttgaagt ttctgtgta g	1581

<210> 5
 <211> 1581
 <212> DNA
 <213> 人工序列

<220>
 <223> 埃博拉病毒全球祖先 (T2-6) 经基因优化

<400>	5	
	atggcggag gatctagact gctgcaactg cccagagagc ggttcagaaa gaccagcttc	60
	ttcgtgtggg tcatcatcct gttccagaaa gccctcagca tgcccctggg cgtcgtgacc	120
	aatagcacc tgaagtgc cgagatcgac cagctcgtgt gcagagataa gctgagcagc	180
	accagccagc tgaagtccgt gggactgaat ctggaaggca atggcgtggc cacagatgtg	240
	cttagccca ccaaagatg gggctttaga agcggcgtgc cacctaaggt ggtgtcttat	300
	gaagccggcg agtgggccga gaactgctac aacctggaaa tcaagaagcc cgacggcagc	360
	gagtgtctgc ctctccacc tgatggcgtc agaggcttcc ctgatgcag atacgtcac	420
	aaggtgcaag gcacaggccc ctgtcctggc gatttcgcct tcacaagga cggcgccttt	480
	ttcctgtacg atcggctggc ctccaccgtg atctacagag gcacaacatt tgccgaaggc	540
	gtgggtggcct tcctgatcct gcctaagcct aagaaggact tctttcagag ccctcctatc	600
	cgcgagcctg tgaacacaac agaggacccc agcagctact acaccaccag cacactgagc	660

tacgagatcg ataacttcgg cgccaacaag accaagacac tgttcaaggt ggacaaccac 720
 acctacgtgc agctggacag accccacaca cctcagtttc tggatgcagct gaacgagaca 780
 atccacacca acaacagact gagcaacacc accggcaggc tgatctggac cctggatcct 840
 aagatcgaca ccgacatcgg agagtgggccc ttttgggaga acaagaagaa cttcagcaag 900
 cagctgagag gcgaggaact gagctttaag gccctgagca ccaagacagg cgccaacgct 960
 gtggataccg atgagtctag caagcccggc ctgatcacca acacagttag aggcgttgcc 1020
 gacctgtga gcccttggag aagaaagcgg agacaagtga accccaatac caccaacaag 1080
 tgcaacccta acctgcaacta ctggacagcc caggatgaag gcgctgctgt tggactggcc 1140
 tggattcctt attttggacc tgccgccgag ggcatctaca cagagggaaat catgcacaac 1200
 cagaatggcc tgatctgcgg cctgagacag ctggccaatg agacaacaca ggccctccag 1260
 ctgtttctga gagccaccac cgagctgaga acctacagca tcctgaaccg gaaggccatc 1320
 gactttctgc tgcaaagatg gggaggcacc tgtagaatcc tgggacctga ttgetgcatc 1380
 gagccccacg actggaccaa gaacatcacc gacaagatca accagatcat ccacgacttc 1440
 atcgacaacc ctctgcctga ccaggacgac gacgataatt ggtggacagg atggcggcag 1500
 tggattcctg ccggaatcgg aatcacaggc gtgatcattg ccattatcgc cctgctgtgc 1560
 atctgcaagt ttctgtgctg a 1581

[0006]

<210> 6
 <211> 526
 <212> PRT
 <213> 人工序列
 <220>
 <223> 埃博拉病毒全球祖先 (T2-6)
 <400> 6

Met Gly Gly Gly Ser Arg Leu Leu Gln Leu Pro Arg Glu Arg Phe Arg
 1 5 10 15
 Lys Thr Ser Phe Phe Val Trp Val Ile Ile Leu Phe Gln Lys Ala Phe
 20 25 30
 Ser Met Pro Leu Gly Val Val Thr Asn Ser Thr Leu Lys Val Thr Glu
 35 40 45
 Ile Asp Gln Leu Val Cys Arg Asp Lys Leu Ser Ser Thr Ser Gln Leu
 50 55 60
 Lys Ser Val Gly Leu Asn Leu Glu Gly Asn Gly Val Ala Thr Asp Val
 65 70 75 80
 Pro Ser Ala Thr Lys Arg Trp Gly Phe Arg Ser Gly Val Pro Pro Lys
 85 90 95
 Val Val Ser Tyr Glu Ala Gly Glu Trp Ala Glu Asn Cys Tyr Asn Leu
 100 105 110
 Glu Ile Lys Lys Pro Asp Gly Ser Glu Cys Leu Pro Pro Pro Asp
 115 120 125

Gly Val Arg Gly Phe Pro Arg Cys Arg Tyr Val His Lys Val Gln Gly
 130 135 140

Thr Gly Pro Cys Pro Gly Asp Phe Ala Phe His Lys Asp Gly Ala Phe
 145 150 155 160

Phe Leu Tyr Asp Arg Leu Ala Ser Thr Val Ile Tyr Arg Gly Thr Thr
 165 170 175

Phe Ala Glu Gly Val Val Ala Phe Leu Ile Leu Pro Lys Pro Lys Lys
 180 185 190

Asp Phe Phe Gln Ser Pro Pro Ile Arg Glu Pro Val Asn Thr Thr Glu
 195 200 205

Asp Pro Ser Ser Tyr Tyr Thr Thr Ser Thr Leu Ser Tyr Glu Ile Asp
 210 215 220

Asn Phe Gly Ala Asn Lys Thr Lys Thr Leu Phe Lys Val Asp Asn His
 225 230 235 240

Thr Tyr Val Gln Leu Asp Arg Pro His Thr Pro Gln Phe Leu Val Gln
 245 250 255

Leu Asn Glu Thr Ile His Thr Asn Asn Arg Leu Ser Asn Thr Thr Gly
 260 265 270

[0007] Arg Leu Ile Trp Thr Leu Asp Pro Lys Ile Asp Thr Asp Ile Gly Glu
 275 280 285

Trp Ala Phe Trp Glu Asn Lys Lys Asn Phe Ser Lys Gln Leu Arg Gly
 290 295 300

Glu Glu Leu Ser Phe Lys Ala Leu Ser Thr Lys Thr Gly Ala Asn Ala
 305 310 315 320

Val Asp Thr Asp Glu Ser Ser Lys Pro Gly Leu Ile Thr Asn Thr Val
 325 330 335

Arg Gly Val Ala Asp Leu Leu Ser Pro Trp Arg Arg Lys Arg Arg Gln
 340 345 350

Val Asn Pro Asn Thr Thr Asn Lys Cys Asn Pro Asn Leu His Tyr Trp
 355 360 365

Thr Ala Gln Asp Glu Gly Ala Ala Val Gly Leu Ala Trp Ile Pro Tyr
 370 375 380

Phe Gly Pro Ala Ala Glu Gly Ile Tyr Thr Glu Gly Ile Met His Asn
 385 390 395 400

Gln Asn Gly Leu Ile Cys Gly Leu Arg Gln Leu Ala Asn Glu Thr Thr
 405 410 415

Gln Ala Leu Gln Leu Phe Leu Arg Ala Thr Thr Glu Leu Arg Thr Tyr
 420 425 430

Ser Ile Leu Asn Arg Lys Ala Ile Asp Phe Leu Leu Gln Arg Trp Gly
 435 440 445

Gly Thr Cys Arg Ile Leu Gly Pro Asp Cys Cys Ile Glu Pro His Asp
 450 455 460

Trp Thr Lys Asn Ile Thr Asp Lys Ile Asn Gln Ile Ile His Asp Phe
 465 470 475 480

Ile Asp Asn Pro Leu Pro Asp Gln Asp Asp Asp Asp Asn Trp Trp Thr
 485 490 495

Gly Trp Arg Gln Trp Ile Pro Ala Gly Ile Gly Ile Thr Gly Val Ile
 500 505 510

Ile Ala Ile Ile Ala Leu Leu Cys Ile Cys Lys Phe Leu Cys
 515 520 525

<210> 7
 <211> 2046
 <212> DNA
 <213> 人工序列

<220>
 <223> 马尔堡病毒祖先 (T2-11) 未优化

[0008]

<400> 7
 atgaagacca tatatcttct gattagtctc attttaatcc aaagtataaa aactctcct 60
 gttttagaaa ttgctagtaa cagccaacct caagatgtag attcagtgtg ctccggaacc 120
 ctccaaaaga cagaagatgt tcatctgatg ggatttacac tgagtgggca aaaagttgct 180
 gattcccctt tggaagcatc taaacgatgg gctttcagga cagggtgtcc tccaagaac 240
 gttgagtata cggaaggaga agaagccaaa acatgttaca atataagtgt aacagaccct 300
 tctggaaaat ccttgctgct ggatcctccc agtaatatcc gcgattacc taaatgtaa 360
 actgttcac atattcaagg tcaaacacct catgcacagg ggattgccct ccatttgtgg 420
 ggggcatttt tctgtatga tcgattgcc tccacaaca tgtaccgagg caaagtcttc 480
 actgaaggga acatagcagc tatgattgtc aataagacag tgcacaaaat gattttctcg 540
 aggcaaggac aagggtaccg tcacatgaat ctgacttcta ctaataaata ttggacaagt 600
 agcaacggaa cgcaaacgaa tgacactgga tgcttcggtg ctcttcaaga atacaattct 660
 acgaagaacc aaacatgtgc tccgtccaaa atacctccac cactgcccac agcccgtccg 720
 gagatcaaac ccacaagcac cccaactgat gccaccaaac tcaacaccac agacccaaac 780
 agtgatgatg aggacctcac aacatccggc tcagggtccg gagaacagga accctacaca 840
 acttctgatg cggctactaa gcaaggcctt tcatcaaca tgccaccac tccctcacca 900
 caaccaagca cgccacagca aggaggaaac aacacaaacc attccaaggg tgctgtgact 960
 gaaccgaca aaaccaacac aactgcaca cctccatgc cccccacaa cactactaca 1020
 atcttacta acaacaccte caagcacaac ttcagcactc tctctgcacc actacaaaac 1080
 accaccaatt acaacacaca gagcacggcc actgaaaatg agcaaaccag tgccccctcg 1140
 aaaacaacc tgctccaac aggaaatcct accacagcaa agagaccaa cagcacaana 1200
 ggccccacca caacggcacc aaatacgaca aatgggcatt tcaccagtcc ctccccacc 1260
 cccaactcga ctacacaaca tcttgatat ttcagaagga aacgaagtat cctctggagg 1320

gaagggcaca tgttccttt tttagatggg ttaataaata ctgaaattga ttttgatcca 1380
 atcccaaca cagaacaat ctttgatgaa tccccagct ttaatacttc aactaatgag 1440
 gaacaacaca ctccccgaa tatcagttta actttctctt attttctga taaaaatgga 1500
 gatactgcct actctgggga aaacgagaat gatttgtatg cagagttgag gatttggagt 1560
 gtgcaggagg acgatttggc ggcagggtt agctggatac cattttttgg ccctggaatc 1620
 gaaggactct atactgccgg tttaatcaaa aatcagaaca atttagttg taggttgagg 1680
 cgcttagcta atcaactgc taaatccttg gagctcttgt taagggtcac aaccgaggaa 1740
 aggacatfff ccttaatcaa taggcatgca attgactttt tgcttacgag gtggggcgga 1800
 acatgcaagg tgctaggacc tgattgttgc ataggaatag aagatctatc taaaaatc 1860
 tcagaacaaa ttgacaaaa cagaaaggat gaacaaaagg aggaaactgg ctggggctca 1920
 ggtggcaaat ggtggacatc tgactggggt gttctacca atttgggcat cctgctacta 1980
 ttatctatag ctgttctgat tgctctgtcc tgtatctgtc gtatcttcac taaatatac 2040
 ggatag 2046

<210> 8
 <211> 2046
 <212> DNA
 <213> 人工序列

<220>
 <223> 马尔堡病毒祖先 (T2-11) 经基因优化

<400> 8
 [0009] atgaagacca tctactttct gatcagcctg atcctgatcc agagatcaa gacctgcct 60
 gtgctggaaa tcgccagcaa cagtcagccc caggatgtgg atagcgtgtg tagcggcacc 120
 ctccagaaaa ccgaggatgt gcacctgatg ggctttacc tgagcggcca gaaagtggcc 180
 gattctccac tggaaaccag caagagatgg gccttagaa ccggcgtgcc acctaaagaac 240
 gtcgagtaca cagagggcga agagccaag acctgctaca acatcagcgt gaccgatcct 300
 agcggcaaga gcctgtgtct ggacctcct agcaacatca gagactacc caagtcaag 360
 accgtgcacc acatccaggg acagaatccc catgctcagg gaattgcct gcacctgtgg 420
 ggcgcctttt tcctgtatga tcggatcgcc tccaccacca tgtacagagg caaagtgtc 480
 accgagggca atatcgccgc catgatcgtg aacaagacag tgcacaagat gatcttcagc 540
 cggcaaggcc agggctacag acacatgaat ctgaccagca ccaacaagta ctggaccagc 600
 agcaacggca cccagaccaa tgatacaggc tgctttggcg cctgcaaga gtacaacagc 660
 accaagaatc agacatgcgc ccctagcaag atcccctc cactgcctac tgccagacct 720
 gagatcaagc ctaccagcac acctaccgac gccaccaagc tgaacaccac cgatccaac 780
 agcgacgacg aggatctgac aacaagcgga tetggtctg gcgagcaaga gccatacacc 840
 acctctgatg ccgtgacaaa gcagggcctg agcagcaca tgcctccaac accttctcca 900
 cagcctagca cacctcagca aggcggcaac aacacaaatc actctcagg cgccgtgacc 960
 gagcctgaca agacaaatc cacagctcag cccagcatgc ctctcaca caccaccaca 1020
 atctccacca acaaccagg caagcacac ttcagcacac tgagcgcgcc tctccagaat 1080
 accaccaact acaataccga gagcaccgcc accgagaacg agcagacatc tgccccttct 1140
 aagaccacac tgccacctac cggcaatcct accaccgcca agagcaccaa tagcacaag 1200

ggcctacca ccaccgtcc taacaccaca aatggccact tcacaagccc aagtccata 1260
 cctaacagca caaccagca cctggtgtac ttcagacgga agcggagcat cctttggcgc 1320
 gagggcgata tgttccctt cctggacggc ctgatcaaca ccgagatcga cttcgacccc 1380
 attccaaaca ccgaaacct cttcgacgag agccccagct tcaacacctc caccaatgag 1440
 gaacagcaca cccctcaaaa catctcctg accttcagct acttccccga caagaacggc 1500
 gatacagcct acagcggcga gaatgagaat gactgcgacg ccgagctgcg gatttgagc 1560
 gttcaagagg atgatctggc tgccggcctg agctggatcc ctttttttgg acctggcatc 1620
 gagggcctgt acaccgccgg actgatcaag aaccagaaca acctcgtgtg cagactgcgg 1680
 agactggcca atcagaccgc caagtctctg gaactgctgc tgcgcgtgac caccgaggaa 1740
 agaaccttct ctctgatcaa ccggcacgcc atcgattttc tgctgaccag atggggcggc 1800
 acctgtaaag ttctggggcc tgattgctgc atcggaatcg aggacctgag caagaacatc 1860
 tccgagcaga tcgacaagat ccgcaaggac gagcagaaaag aggaaacagg ctggggactc 1920
 ggcggcaagt ggtggacatc tgattggggc gtgctgacca atctgggaat cctgctgctc 1980
 ctgtctatcg ccgtgctgat cgccctgagc tgcactctgc ggatcttcac caagtacatc 2040
 ggctga 2046

<210> 9
 <211> 681
 <212> PRT
 <213> 人工序列

[0010]

<220>
 <223> 马尔堡病毒祖先 (T2-11)
 <400> 9

Met Lys Thr Ile Tyr Phe Leu Ile Ser Leu Ile Leu Ile Gln Ser Ile
 1 5 10 15
 Lys Thr Leu Pro Val Leu Glu Ile Ala Ser Asn Ser Gln Pro Gln Asp
 20 25 30
 Val Asp Ser Val Cys Ser Gly Thr Leu Gln Lys Thr Glu Asp Val His
 35 40 45
 Leu Met Gly Phe Thr Leu Ser Gly Gln Lys Val Ala Asp Ser Pro Leu
 50 55 60
 Glu Ala Ser Lys Arg Trp Ala Phe Arg Thr Gly Val Pro Pro Lys Asn
 65 70 75 80
 Val Glu Tyr Thr Glu Gly Glu Glu Ala Lys Thr Cys Tyr Asn Ile Ser
 85 90 95
 Val Thr Asp Pro Ser Gly Lys Ser Leu Leu Leu Asp Pro Pro Ser Asn
 100 105 110
 Ile Arg Asp Tyr Pro Lys Cys Lys Thr Val His His Ile Gln Gly Gln
 115 120 125
 Asn Pro His Ala Gln Gly Ile Ala Leu His Leu Trp Gly Ala Phe Phe
 130 135 140

Leu Tyr Asp Arg Ile Ala Ser Thr Thr Met Tyr Arg Gly Lys Val Phe
 145 150 155 160

Thr Glu Gly Asn Ile Ala Ala Met Ile Val Asn Lys Thr Val His Lys
 165 170 175

Met Ile Phe Ser Arg Gln Gly Gln Gly Tyr Arg His Met Asn Leu Thr
 180 185 190

Ser Thr Asn Lys Tyr Trp Thr Ser Ser Asn Gly Thr Gln Thr Asn Asp
 195 200 205

Thr Gly Cys Phe Gly Ala Leu Gln Glu Tyr Asn Ser Thr Lys Asn Gln
 210 215 220

Thr Cys Ala Pro Ser Lys Ile Pro Pro Pro Leu Pro Thr Ala Arg Pro
 225 230 235 240

Glu Ile Lys Pro Thr Ser Thr Pro Thr Asp Ala Thr Lys Leu Asn Thr
 245 250 255

Thr Asp Pro Asn Ser Asp Asp Glu Asp Leu Thr Thr Ser Gly Ser Gly
 260 265 270

Ser Gly Glu Gln Glu Pro Tyr Thr Thr Ser Asp Ala Val Thr Lys Gln
 275 280 285

[0011] Gly Leu Ser Ser Thr Met Pro Pro Thr Pro Ser Pro Gln Pro Ser Thr
 290 295 300

Pro Gln Gln Gly Gly Asn Asn Thr Asn His Ser Gln Gly Ala Val Thr
 305 310 315 320

Glu Pro Asp Lys Thr Asn Thr Thr Ala Gln Pro Ser Met Pro Pro His
 325 330 335

Asn Thr Thr Thr Ile Ser Thr Asn Asn Thr Ser Lys His Asn Phe Ser
 340 345 350

Thr Leu Ser Ala Pro Leu Gln Asn Thr Thr Asn Tyr Asn Thr Gln Ser
 355 360 365

Thr Ala Thr Glu Asn Glu Gln Thr Ser Ala Pro Ser Lys Thr Thr Leu
 370 375 380

Pro Pro Thr Gly Asn Pro Thr Thr Ala Lys Ser Thr Asn Ser Thr Lys
 385 390 395 400

Gly Pro Thr Thr Thr Ala Pro Asn Thr Thr Asn Gly His Phe Thr Ser
 405 410 415

Pro Ser Pro Thr Pro Asn Ser Thr Thr Gln His Leu Val Tyr Phe Arg
 420 425 430

Arg Lys Arg Ser Ile Leu Trp Arg Glu Gly Asp Met Phe Pro Phe Leu
 435 440 445

Asp Gly Leu Ile Asn Thr Glu Ile Asp Phe Asp Pro Ile Pro Asn Thr
 450 455 460

Glu Thr Ile Phe Asp Glu Ser Pro Ser Phe Asn Thr Ser Thr Asn Glu
 465 470 475 480

Glu Gln His Thr Pro Pro Asn Ile Ser Leu Thr Phe Ser Tyr Phe Pro
 485 490 495

Asp Lys Asn Gly Asp Thr Ala Tyr Ser Gly Glu Asn Glu Asn Asp Cys
 500 505 510

Asp Ala Glu Leu Arg Ile Trp Ser Val Gln Glu Asp Asp Leu Ala Ala
 515 520 525

Gly Leu Ser Trp Ile Pro Phe Phe Gly Pro Gly Ile Glu Gly Leu Tyr
 530 535 540

Thr Ala Gly Leu Ile Lys Asn Gln Asn Asn Leu Val Cys Arg Leu Arg
 545 550 555 560

Arg Leu Ala Asn Gln Thr Ala Lys Ser Leu Glu Leu Leu Leu Arg Val
 565 570 575

Thr Thr Glu Glu Arg Thr Phe Ser Leu Ile Asn Arg His Ala Ile Asp
 580 585 590

[0012] Phe Leu Leu Thr Arg Trp Gly Gly Thr Cys Lys Val Leu Gly Pro Asp
 595 600 605

Cys Cys Ile Gly Ile Glu Asp Leu Ser Lys Asn Ile Ser Glu Gln Ile
 610 615 620

Asp Lys Ile Arg Lys Asp Glu Gln Lys Glu Glu Thr Gly Trp Gly Leu
 625 630 635 640

Gly Gly Lys Trp Trp Thr Ser Asp Trp Gly Val Leu Thr Asn Leu Gly
 645 650 655

Ile Leu Leu Leu Leu Ser Ile Ala Val Leu Ile Ala Leu Ser Cys Ile
 660 665 670

Cys Arg Ile Phe Thr Lys Tyr Ile Gly
 675 680

<210> 10

<211> 1578

<212> DNA

<213> 人工序列

<220>

<223> Tier 2-4 (SUDV_anc_-MLD)

<400> 10

atgggaggac tgtctctgct gcaactgcc cgggacaagt tccggaagtc cagcttcttc 60

gtgtgggtca tcatcctggt ccagaaagcc ttcagcatgc ccctgggcgt cgtgaccaat 120

agcacttg g aagtgaccga gatcgaccag ctctgtgtgca aggatcacct ggccagcacc 180

gatcagctga agtctgtggg actgaatctg gaaggcagcg gcgtgtccac agatatccct	240
agcgccacca agagatgggg ctttagaagc ggagtcctc ctaaggtagt gtcttatgaa	300
gccggcgagt gggccgagaa ctgctacaac ctggaaatca agaagcccga cggcagcgag	360
tgtctgcctc ctccacctga tggcgtcaga ggcttcccta gatgcagata cgtgcacaag	420
gcccaggca caggacctg tectggcgat tacgccttcc acaaggacgg cgcctttttc	480
ctgtacgata ggctggcctc caccgtgata tacagaggcg ttaactttgc cgagggcgtg	540
atcgcttcc tgatcctggc caagcctaaa gagacattcc tgcaaagccc tccaatccgc	600
gaggccgtga actacacaga gaacaccagc agctactacg ccaccagcta cctggaatac	660
gagatcgaga atttcggcgc ccagcacagc accacactgt tcaagatcga caacaacacc	720
ttcgtgcggc tggacagacc ccacacacct cagtttctgt tccagctgaa cgacaccatc	780
catctgcatc agcagctgag caacaccacc ggacagactga tttggaccct ggacgccaac	840
atcaacgccc acattggaga gtgggccttt tgggagaaca agaagaacct gagcgaacag	900
ctgagaggcg aggaactgag ctttgaggcc ctgtctctga ccaccgccgt gaaaacagtg	960
ctgcctcaag agtccaccag caacggcctg atcacaagca cagtgcagcag catcctgggc	1020
agcctgggcc tgagaaaaag gtcagacagc caagtgaata ccaaggccac cggcaagtgc	1080
aacccaacc tgcactattg gacagcccaa gagcagcaca atgccgcccg aatgccttgg	1140
attccttatt ttggacctgg cggcagggc atctataccg agggactgat gcacaaccag	1200
aacgccctcg tgtgtggact gagacagctg gccaatgaga caacacaggc cctccagctg	1260
[0013] tttctgagag ccaccaccga gctgagaacc tacaccatcc tgaaccggaa ggccatcgac	1320
tttctgctga gaagatgggg cggcacctgt agaatcctgg gacctgattg ctgcatcgag	1380
ccccagact ggaccaagaa catcaccgac aagatcaacc agatcatcca cgacttcac	1440
gacaaccctc tgcctaacca ggacaacgac gacaattggt ggacaggctg gcggcagttg	1500
attcctgccg gaattggcat caccggcacc atcattgcca ttatgccct gctgtgtgtg	1560
tgcaagctgc tgtgttga	1578

- <210> 11
- <211> 525
- <212> PRT
- <213> 人工序列

<220>
 <223> Tier 2-4 (SUDV_anc_-MLD)

<400> 11

Met Gly Gly Leu Ser Leu Leu Gln Leu Pro Arg Asp Lys Phe Arg Lys
 1 5 10 15

Ser Ser Phe Phe Val Trp Val Ile Ile Leu Phe Gln Lys Ala Phe Ser
 20 25 30

Met Pro Leu Gly Val Val Thr Asn Ser Thr Leu Glu Val Thr Glu Ile
 35 40 45

Asp Gln Leu Val Cys Lys Asp His Leu Ala Ser Thr Asp Gln Leu Lys
 50 55 60

Ser Val Gly Leu Asn Leu Glu Gly Ser Gly Val Ser Thr Asp Ile Pro
 65 70 75 80

Ser Ala Thr Lys Arg Trp Gly Phe Arg Ser Gly Val Pro Pro Lys Val
 85 90 95

Val Ser Tyr Glu Ala Gly Glu Trp Ala Glu Asn Cys Tyr Asn Leu Glu
 100 105 110

Ile Lys Lys Pro Asp Gly Ser Glu Cys Leu Pro Pro Pro Pro Asp Gly
 115 120 125

Val Arg Gly Phe Pro Arg Cys Arg Tyr Val His Lys Ala Gln Gly Thr
 130 135 140

Gly Pro Cys Pro Gly Asp Tyr Ala Phe His Lys Asp Gly Ala Phe Phe
 145 150 155 160

Leu Tyr Asp Arg Leu Ala Ser Thr Val Ile Tyr Arg Gly Val Asn Phe
 165 170 175

Ala Glu Gly Val Ile Ala Phe Leu Ile Leu Ala Lys Pro Lys Glu Thr
 180 185 190

Phe Leu Gln Ser Pro Pro Ile Arg Glu Ala Val Asn Tyr Thr Glu Asn
 195 200 205

[0014] Thr Ser Ser Tyr Tyr Ala Thr Ser Tyr Leu Glu Tyr Glu Ile Glu Asn
 210 215 220

Phe Gly Ala Gln His Ser Thr Thr Leu Phe Lys Ile Asp Asn Asn Thr
 225 230 235 240

Phe Val Arg Leu Asp Arg Pro His Thr Pro Gln Phe Leu Phe Gln Leu
 245 250 255

Asn Asp Thr Ile His Leu His Gln Gln Leu Ser Asn Thr Thr Gly Arg
 260 265 270

Leu Ile Trp Thr Leu Asp Ala Asn Ile Asn Ala Asp Ile Gly Glu Trp
 275 280 285

Ala Phe Trp Glu Asn Lys Lys Asn Leu Ser Glu Gln Leu Arg Gly Glu
 290 295 300

Glu Leu Ser Phe Glu Ala Leu Ser Leu Thr Thr Ala Val Lys Thr Val
 305 310 315 320

Leu Pro Gln Glu Ser Thr Ser Asn Gly Leu Ile Thr Ser Thr Val Thr
 325 330 335

Gly Ile Leu Gly Ser Leu Gly Leu Arg Lys Arg Ser Arg Arg Gln Val
 340 345 350

Asn Thr Lys Ala Thr Gly Lys Cys Asn Pro Asn Leu His Tyr Trp Thr
 355 360 365

Ala Gln Glu Gln His Asn Ala Ala Gly Ile Ala Trp Ile Pro Tyr Phe
 370 375 380

Gly Pro Gly Ala Glu Gly Ile Tyr Thr Glu Gly Leu Met His Asn Gln
 385 390 395 400

Asn Ala Leu Val Cys Gly Leu Arg Gln Leu Ala Asn Glu Thr Thr Gln
 405 410 415

Ala Leu Gln Leu Phe Leu Arg Ala Thr Thr Glu Leu Arg Thr Tyr Thr
 420 425 430

Ile Leu Asn Arg Lys Ala Ile Asp Phe Leu Leu Arg Arg Trp Gly Gly
 435 440 445

Thr Cys Arg Ile Leu Gly Pro Asp Cys Cys Ile Glu Pro His Asp Trp
 450 455 460

Thr Lys Asn Ile Thr Asp Lys Ile Asn Gln Ile Ile His Asp Phe Ile
 465 470 475 480

Asp Asn Pro Leu Pro Asn Gln Asp Asn Asp Asp Asn Trp Trp Thr Gly
 485 490 495

Trp Arg Gln Trp Ile Pro Ala Gly Ile Gly Ile Thr Gly Ile Ile Ile
 500 505 510

Ala Ile Ile Ala Leu Leu Cys Val Cys Lys Leu Leu Cys
 515 520 525

[0015]

<210> 12
 <211> 1581
 <212> DNA
 <213> 人工序列

<220>
 <223> Tier 2-6 (SUDV_EBOV-TAFV-BDBV_anc_-MLD)

<400> 12
 atgggcggag gatctagact gctgcaactg cccagagagc ggttcagaaa gaccagcttc 60
 ttcgtgtggg tcatcatcct gttccagaaa gccttcagca tgcccctggg cgtegtgacc 120
 aatagcacc tgaagtgc cgagatcgac cagctcgtgt gcagagataa gctgagcagc 180
 accagccagc tgaagtcctt gggactgaat ctggaaggca atggcgtggc cacagatgtg 240
 cctagcgcca caaaagatg gggctttaga agcggcgtgc cacctaaggt ggtgtcttat 300
 gaagccggcg agtgggccga gaactgtac aacctggaaa tcaagaagcc cgacggcagc 360
 gagtgtctgc ctctccacc tgatggcgtc agaggcttcc ctatagtcag atactgcac 420
 aaggtgcaag gcacaggccc ctgtcctggc gatttcgctt ttcacaagga cggegccttt 480
 ttctgtacg atcggtggc ctccaccgtg atctacagag gcacaacatt tgccgaaggc 540
 gtgtggcct tctgtatcct gcctaagcct aagaaggact tctttcagag cctcctatc 600
 cgcgagcctg tgaacacaac agaggacccc agcagctact acaccaccag cacactgagc 660
 tacgagatcg ataactcgg cgccaacaag accaagacac tgttcaaggt ggacaaccac 720
 acctactgtc agctggacag accccacaca cctcagtttc tgggtcagct gaacgagaca 780
 atccacacca acaacagact gagcaacacc accggcaggc tgatctggac cctggatcct 840

aagatcgaca cgcacatcgg agagtgggcc ttttgggaga acaagaagaa cttcagcaag 900
cagctgagag gcgaggaact gagctttaag gccctgagca ccaagacagg cgccaacgct 960
gtggataccg atgagtctag caagcccggc ctgatcacca acacagttag aggcgttgcc 1020
gacctgctga gcccttggag aagaaagcgg agacaagtga accccaatac caccaacaag 1080
tgcaacccta acctgcaact ctggacagcc caggatgaag gcgctgctgt tggactggcc 1140
tggattcctt attttggacc tgccgccgag ggcattctaca cagagggaaat catgcacaac 1200
cagaatggcc tgatctgagg cctgagacag ctggccaatg agacaacaca ggccctccag 1260
ctgtttctga gagccaccac cgagctgaga acctacagca tcctgaaccg gaaggccatc 1320
gactttctgc tgcaaagatg gggaggcacc ttagaatcc tgggacctga ttgctgcatc 1380
gagccccacg actggaccaa gaacatcacc gacaagatca accagatcat ccacgacttc 1440
atcgacaacc ctctgcctga ccaggacgac gacgataatt ggtggacagg atggcggcag 1500
tggattcctg ccggaatcgg aatcacaggc gtgatcattg ccattatcgc cctgctgtgc 1560
atctgcaagt ttctgtgctg a 1581

<210> 13
<211> 526
<212> PRT
<213> 人工序列

<220>
<223> Tier 2-6 (SUDV_EBOV-TAFV-BDBV_anc_-MLD)

<400> 13

[0016]

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

Phe Leu Tyr Asp Arg Leu Ala Ser Thr Val Ile Tyr Arg Gly Thr Thr
 165 170 175

Phe Ala Glu Gly Val Val Ala Phe Leu Ile Leu Pro Lys Pro Lys Lys
 180 185 190

Asp Phe Phe Gln Ser Pro Pro Ile Arg Glu Pro Val Asn Thr Thr Glu
 195 200 205

Asp Pro Ser Ser Tyr Tyr Thr Thr Ser Thr Leu Ser Tyr Glu Ile Asp
 210 215 220

Asn Phe Gly Ala Asn Lys Thr Lys Thr Leu Phe Lys Val Asp Asn His
 225 230 235 240

Thr Tyr Val Gln Leu Asp Arg Pro His Thr Pro Gln Phe Leu Val Gln
 245 250 255

Leu Asn Glu Thr Ile His Thr Asn Asn Arg Leu Ser Asn Thr Thr Gly
 260 265 270

Arg Leu Ile Trp Thr Leu Asp Pro Lys Ile Asp Thr Asp Ile Gly Glu
 275 280 285

Trp Ala Phe Trp Glu Asn Lys Lys Asn Phe Ser Lys Gln Leu Arg Gly
 290 295 300

[0017] Glu Glu Leu Ser Phe Lys Ala Leu Ser Thr Lys Thr Gly Ala Asn Ala
 305 310 315 320

Val Asp Thr Asp Glu Ser Ser Lys Pro Gly Leu Ile Thr Asn Thr Val
 325 330 335

Arg Gly Val Ala Asp Leu Leu Ser Pro Trp Arg Arg Lys Arg Arg Gln
 340 345 350

Val Asn Pro Asn Thr Thr Asn Lys Cys Asn Pro Asn Leu His Tyr Trp
 355 360 365

Thr Ala Gln Asp Glu Gly Ala Ala Val Gly Leu Ala Trp Ile Pro Tyr
 370 375 380

Phe Gly Pro Ala Ala Glu Gly Ile Tyr Thr Glu Gly Ile Met His Asn
 385 390 395 400

Gln Asn Gly Leu Ile Cys Gly Leu Arg Gln Leu Ala Asn Glu Thr Thr
 405 410 415

Gln Ala Leu Gln Leu Phe Leu Arg Ala Thr Thr Glu Leu Arg Thr Tyr
 420 425 430

Ser Ile Leu Asn Arg Lys Ala Ile Asp Phe Leu Leu Gln Arg Trp Gly
 435 440 445

Gly Thr Cys Arg Ile Leu Gly Pro Asp Cys Cys Ile Glu Pro His Asp
 450 455 460

Trp Thr Lys Asn Ile Thr Asp Lys Ile Asn Gln Ile Ile His Asp Phe
 465 470 475 480

Ile Asp Asn Pro Leu Pro Asp Gln Asp Asp Asp Asp Asn Trp Trp Thr
 485 490 495

Gly Trp Arg Gln Trp Ile Pro Ala Gly Ile Gly Ile Thr Gly Val Ile
 500 505 510

Ile Ala Ile Ile Ala Leu Leu Cys Ile Cys Lys Phe Leu Cys
 515 520 525

<210> 14
 <211> 2046
 <212> DNA
 <213> 人工序列

<220>
 <223> Tier 2-11 (RAVV_MARV_anc)

<400> 14
 atgaagacca tctactttct gatcagcctg atcctgatcc agagcatcaa gaccctgctt 60
 gtgtctggaaa tcgccagcaa cagtcagccc caggatgtgg atagcgtgtg tagcggcacc 120
 ctccagaaaa ccgaggatgt gcacctgatg ggctttacc tgagcggcca gaaagtggcc 180
 gattctccac tggaagccag caagagatgg gcctttagaa cggcgtgcc acctaagaac 240
 gtcgagtaca cagagggcga agaggccaag acctgctaca acatcagcgt gaccgatcct 300
 agcggcaaga gcctgctgct ggacctcct agcaacatca gagactacc caagtgaag 360
 [0018] accgtgcacc acatccaggg acagaatccc catgctcagg gaattgcct gcacctgtgg 420
 ggcgcctttt tcctgtatga tcggatcgcc tccaccacca tgtacagagg caaagtgttc 480
 accgagggca atatgccgc catgatcgtg aacaagacag tgcacaagat gatcttcagc 540
 cggcaaggcc agggctacag acacatgaat ctgaccagca ccaacaagta ctggaccagc 600
 agcaacggca cccagaccaa tgatacagc tgctttggcg ccctgcaaga gtacaacagc 660
 accaagaatc agacatgccc ccttagcaag atccctcctc cactgcctac tgccagacct 720
 gagatcaagc ctaccagcac acctaccgac gccaccaagc tgaacaccac cgateccaac 780
 agcgcagcagc aggatctgac aacaagcgga tctggctctg gcgagcaaga gccatacacc 840
 acctctgatg ccgtgacaaa gcagggcctg agcagcacia tgccctcaac accttctcca 900
 cagcctagca cacctcagca aggcggcaac aacacaaatc actctcaggg cgcctgacc 960
 gagcctgaca agacaaatac cacagctcag cccagcatgc ctctcacia caccaccaca 1020
 atctccacca acaacaccag caagcacaac ttcagcacac tgagcgcctc tctccagaat 1080
 accaccaact acaatacca gagcaccgcc accgagaacg agcagacatc tgccccttct 1140
 aagaccacac tgccacctac cggcaatcct accaccgcca agagcaccaa tagcacaag 1200
 ggcctacca ccaccgtcc taacaccaca aatggccact tcacaagccc aagtcctaca 1260
 cctaacagca caaccagca cctggtgtac ttcagacgga agcggagcat cctttggcgc 1320
 gagggcgata tgttcccttt cctggacggc ctgatcaaca ccgagatcga cttcgacccc 1380
 attcacaaca ccgaaccat ctctcagcag agccccagct tcaacacctc caccaatgag 1440
 gaacagcaca cccctcaaaa catctccctg accttcagct acttccccga caagaacggc 1500

gatacacgct acagcggcga gaatgagaat gactgcgacg ccgagctcgg gatttggagc 1560
 gttcaagagg atgatctggc tgccggcctg agctggatcc ctttttttgg acctggcatc 1620
 gagggcctgt acaccgccgg actgatcaag aaccagaaca acctcgtgtg cagactgcgg 1680
 agactggcca atcagaccgc caagtctctg gaactgctgc tgcgcgtgac caccgaggaa 1740
 agaaccttct ctctgatcaa ccggcacgcc atcgattttc tgetgaccag atggggcggc 1800
 acctgtaaag ttctgggccc tgattgctgc atcggaatcg aggacctgag caagaacatc 1860
 tccgagcaga tcgacaagat ccgcaaggac gagcagaaaag aggaaacagg ctgggggactc 1920
 ggcggcaagt ggtggacatc tgattggggc gtgctgacca atctgggaat cctgctgctc 1980
 ctgtctatcg ccgtgctgat cgccttgagc tgcattctgc ggatcttcac caagtacatc 2040
 ggctga 2046

<210> 15
 <211> 681
 <212> PRT
 <213> 人工序列

<220>
 <223> Tier 2-11 (RAVV_MARV_anc)

<400> 15

Met Lys Thr Ile Tyr Phe Leu Ile Ser Leu Ile Leu Ile Gln Ser Ile
 1 5 10 15

[0019]

Lys Thr Leu Pro Val Leu Glu Ile Ala Ser Asn Ser Gln Pro Gln Asp
 20 25 30

Val Asp Ser Val Cys Ser Gly Thr Leu Gln Lys Thr Glu Asp Val His
 35 40 45

Leu Met Gly Phe Thr Leu Ser Gly Gln Lys Val Ala Asp Ser Pro Leu
 50 55 60

Glu Ala Ser Lys Arg Trp Ala Phe Arg Thr Gly Val Pro Pro Lys Asn
 65 70 75 80

Val Glu Tyr Thr Glu Gly Glu Glu Ala Lys Thr Cys Tyr Asn Ile Ser
 85 90 95

Val Thr Asp Pro Ser Gly Lys Ser Leu Leu Leu Asp Pro Pro Ser Asn
 100 105 110

Ile Arg Asp Tyr Pro Lys Cys Lys Thr Val His His Ile Gln Gly Gln
 115 120 125

Asn Pro His Ala Gln Gly Ile Ala Leu His Leu Trp Gly Ala Phe Phe
 130 135 140

Leu Tyr Asp Arg Ile Ala Ser Thr Thr Met Tyr Arg Gly Lys Val Phe
 145 150 155 160

Thr Glu Gly Asn Ile Ala Ala Met Ile Val Asn Lys Thr Val His Lys
 165 170 175

Met Ile Phe Ser Arg Gln Gly Gln Gly Tyr Arg His Met Asn Leu Thr
180 185 190

Ser Thr Asn Lys Tyr Trp Thr Ser Ser Asn Gly Thr Gln Thr Asn Asp
195 200 205

Thr Gly Cys Phe Gly Ala Leu Gln Glu Tyr Asn Ser Thr Lys Asn Gln
210 215 220

Thr Cys Ala Pro Ser Lys Ile Pro Pro Pro Leu Pro Thr Ala Arg Pro
225 230 235 240

Glu Ile Lys Pro Thr Ser Thr Pro Thr Asp Ala Thr Lys Leu Asn Thr
245 250 255

Thr Asp Pro Asn Ser Asp Asp Glu Asp Leu Thr Thr Ser Gly Ser Gly
260 265 270

Ser Gly Glu Gln Glu Pro Tyr Thr Thr Ser Asp Ala Val Thr Lys Gln
275 280 285

Gly Leu Ser Ser Thr Met Pro Pro Thr Pro Ser Pro Gln Pro Ser Thr
290 295 300

Pro Gln Gln Gly Gly Asn Asn Thr Asn His Ser Gln Gly Ala Val Thr
305 310 315 320

[0020] Glu Pro Asp Lys Thr Asn Thr Thr Ala Gln Pro Ser Met Pro Pro His
325 330 335

Asn Thr Thr Thr Ile Ser Thr Asn Asn Thr Ser Lys His Asn Phe Ser
340 345 350

Thr Leu Ser Ala Pro Leu Gln Asn Thr Thr Asn Tyr Asn Thr Gln Ser
355 360 365

Thr Ala Thr Glu Asn Glu Gln Thr Ser Ala Pro Ser Lys Thr Thr Leu
370 375 380

Pro Pro Thr Gly Asn Pro Thr Thr Ala Lys Ser Thr Asn Ser Thr Lys
385 390 395 400

Gly Pro Thr Thr Thr Ala Pro Asn Thr Thr Asn Gly His Phe Thr Ser
405 410 415

Pro Ser Pro Thr Pro Asn Ser Thr Thr Gln His Leu Val Tyr Phe Arg
420 425 430

Arg Lys Arg Ser Ile Leu Trp Arg Glu Gly Asp Met Phe Pro Phe Leu
435 440 445

Asp Gly Leu Ile Asn Thr Glu Ile Asp Phe Asp Pro Ile Pro Asn Thr
450 455 460

Glu Thr Ile Phe Asp Glu Ser Pro Ser Phe Asn Thr Ser Thr Asn Glu
465 470 475 480

Glu Gln His Thr Pro Pro Asn Ile Ser Leu Thr Phe Ser Tyr Phe Pro
 485 490 495

Asp Lys Asn Gly Asp Thr Ala Tyr Ser Gly Glu Asn Glu Asn Asp Cys
 500 505 510

Asp Ala Glu Leu Arg Ile Trp Ser Val Gln Glu Asp Asp Leu Ala Ala
 515 520 525

Gly Leu Ser Trp Ile Pro Phe Phe Gly Pro Gly Ile Glu Gly Leu Tyr
 530 535 540

Thr Ala Gly Leu Ile Lys Asn Gln Asn Asn Leu Val Cys Arg Leu Arg
 545 550 555 560

Arg Leu Ala Asn Gln Thr Ala Lys Ser Leu Glu Leu Leu Leu Arg Val
 565 570 575

Thr Thr Glu Glu Arg Thr Phe Ser Leu Ile Asn Arg His Ala Ile Asp
 580 585 590

Phe Leu Leu Thr Arg Trp Gly Gly Thr Cys Lys Val Leu Gly Pro Asp
 595 600 605

Cys Cys Ile Gly Ile Glu Asp Leu Ser Lys Asn Ile Ser Glu Gln Ile
 610 615 620

Asp Lys Ile Arg Lys Asp Glu Gln Lys Glu Glu Thr Gly Trp Gly Leu
 625 630 635 640

Gly Gly Lys Trp Trp Thr Ser Asp Trp Gly Val Leu Thr Asn Leu Gly
 645 650 655

Ile Leu Leu Leu Leu Ser Ile Ala Val Leu Ile Ala Leu Ser Cys Ile
 660 665 670

Cys Arg Ile Phe Thr Lys Tyr Ile Gly
 675 680

[0021]

<210> 16
 <211> 91
 <212> DNA
 <213> 人工序列

<220>
 <223> pEVAC多克隆位点

<400> 16
 acagactgtt cctttccatg ggtcttttct gcagtcaccg tcggtaccgt cgacacgtgt 60
 gatcatctag aggatccgcg gccgcagatc t 91

<210> 17
 <211> 4405
 <212> DNA
 <213> 人工序列

<220>
 <223> pEVAC的整个序列

<400> 17
 tcgcgcgttt cggtgatgac ggtgaaaacc tctgacacat gcagctccc gagacggtea 60

cagcttgtct gtaagcggat gccgggagca gacaagcccc tcagggcgcg tcagcgggtg	120
ttggcgggtg tcggggctgg cttaactatg cggcatcaga gcagattgta ctgagagtgc	180
accatatgcg gtgtgaaata ccgcacagat gcgtaaggag aaaataccgc atcagattgg	240
ctattggcca ttgcatacgt tgtatccata tcataatatg tacatttata ttggctcatg	300
tccaacatta ccgcatggt gacattgatt attgactagt tattaatagt aatcaattac	360
ggggtcatta gttcatagcc catatatgga gitcccggtt acataactta cggtaaatgg	420
cccgctggc tgaccgccc acgaccccc cccattgacg tcaataatga cgtatgttc	480
catagtaacg ccaatagga ctttcattg acgtcaatgg gtggagtatt tacggtaac	540
tgcccacttg gcagtacatc aagtgtatca tatgccaagt acgcccccta ttgacgtcaa	600
tgacggtaaa tggcccgcct ggcatatgc ccagtacatg acctatggg actttcctac	660
ttggcagtac atctacgtat tagtcatcgc tattaccatg gtgatgcggt tttggcagta	720
catcaatggg cgtggatagc ggtttgactc acggggattt ccaagtctcc accccattga	780
cgtcaatggg agtttgttt ggcacaaaa tcaacgggac ttccaaaat gtcgtaacaa	840
ctccgcccc ttgacgcaa tgggcggtag gcgtgtacgg tgggaggtct atataagcag	900
agctcgttta gtgaaccgtc agatcgctg gagacccat ccacgtggt ttgacctcca	960
tagaagacac cgggaccgat ccagcctcca tcggctcgca tctctcctc acgcgcccgc	1020
cgcctacct gaggcgcca tccacgccg ttgagtcgcg tctgcccgc tcccgcctgt	1080
ggtgcctcct gaactgcgtc cgccgtctag gtaagttta agctcaggtc gagaccgggc	1140
ctttgtccg cgctccctg gagcctacct agactcagcc ggctctccac gctttgcctg	1200
[0022] accctgcttg ctcaactcta gttaacggtg gagggcagtg tagtctgagc agtactcgtt	1260
gctgccgcg gcgccaccag acataatagc tgacagacta acagactggt cctttccatg	1320
ggtctttct gcagtcaccg tcggtaccgt cgacacgtgt gatcatctag aggatccgcg	1380
gccgcagatc tgctgtgct tctagtggc agccatctgt tgtttgccc tccccgtgc	1440
ctccttgac cctggaaggt gccactcca ctgtccttc ctaataaaat gaggaattg	1500
catcgcattg tctgagtagg tgtcattcta ttctgggggg tgggtgggg caggacagca	1560
agggggagga ttggaagac aatagcaggc atgctgggga tgcggtgggc tctatggcta	1620
cccagggtct gaagaattga cccggttct cctgggccag aaagaagcag gcacatcccc	1680
ttctctgta cacacctgt ccacgccct gtttcttagt tccagcccc ctcataggac	1740
actcatagct caggagggt ccgccttcaa tcccaccgc taaagtactt ggagcggct	1800
ctccctcct catcagccc ccaaacaaa cctagcctc aagagtggga agaaattaa	1860
gcaagatagg ctattaagt cagagggaga gaaaatgcct ccaacatgtg aggaagtaat	1920
gagagaaatc atagaatttt aaggccatga tttaaggcca tcatggcctt aatcttccgc	1980
ttctcgtc actgactgc tgcgtcggc cgttcggctg cggcgagcgg taccagctca	2040
ctcaaaggcg gtaatacggc tatccacaga atcaggggat aacgcaggaa agaacatgtg	2100
agcaaaaggc cagcaaaagg ccaggaaccg taaaaaggcc gcgttgctgg cgttttcca	2160
taggctccg ccccctgac agcatcaca aaatcgacgc tcaagtacga ggtggcga	2220
cccagacaga ctataaagat accaggcgtt tccccctgga agctccctcg tgcgtctcc	2280
gttccgacc ctgcccgtta ccgatacct gtccccttt cctccctcgg gaagcgtggc	2340

gctttctcat agctcacgct gtaggtatct cagttcggtg taggtcgttc gctccaagct	2400
gggctgtgtg cacgaacccc ccgttcagcc cgaccgctgc gccttatccg gtaactatcg	2460
tcttgagtcc aaccggtaa gacacgactt atcgccactg gcagcagcca ctggtaacag	2520
gattagcaga gcgaggtatg taggcgggtc tacagagttc ttgaagtggg ggcctaacta	2580
cggtacact agaagaacag tatttggat ctgcgctctg ctgaagccag ttaccttcgg	2640
aaaaagagtt gtagctctt gatccggcaa acaaacacc gctggtagcg gtggtttttt	2700
tgtttgcaag cagcagatta cgcgagaaa aaaaggatct caagaagatc ctttgatctt	2760
ttctacgggg tctgacgctc agtggaacga aaactcacgt taagggattt tggcatgag	2820
attatcaaaa aggatcttca cctagatcct ttaaatata aaatgaagtt ttaaatcaat	2880
ctaaagtata tatgagtaaa cttggcttga cagttaccaa tgcttaatca gtgaggcacc	2940
tatctcagcg atctgtctat ttcgttcac catagttgcc tgactcgggg gggggggcg	3000
ctgaggctctg cctcgtgaag aagggttgc tgactcatac caggcctgaa tcgccccatc	3060
atccagccag aaagtgagg agccacggtt gatgagagct ttgtttagg tggaccagtt	3120
gggtattttg aacttttctt ttgccacgga acggtctcgc ttgtcgggaa gatgcgtgat	3180
ctgatcctc aactcagcaa aagttcgatt tattcaaaa agccgccgct ccgtcaagtc	3240
agcgtaatgc tctgccagtg ttacaaccaa ttaaccaatt ctgattagaa aaactcatcg	3300
agcatcaaat gaaactgcaa tttattcata tcaggattat caataccata ttttgaaaa	3360
agccgtttct gtaatgaagg agaaaactca ccgaggcagt tccataggat ggcaagatcc	3420
tggtatcggg ctgcgattcc gactcgtcca acatcaatac aaactataa tttcccctcg	3480
[0023] tcaaaaaataa ggttatcaag tgagaaatca ccatgagtga cgactgaatc cggtgagaat	3540
ggcaaaagct tatgcatttc tttccagact tgttcaacag gccagccatt acgctcgtca	3600
tcaaaatcac tcgcatcaac caaacggtt ttcatctgtg attgcgctg agcgagacga	3660
aatacgcgat cgctgttaaa aggacaatta caaacaggaa tcgaatgcaa ccggcgcagg	3720
aacactgcca gcgcatcaac aatattttca cctgaatcag gatattcttc taatacctgg	3780
aatgctgttt tcccgggat cgcagtggtg agtaaccatg catcatcagg agtacggata	3840
aaatgcttga tggtcggaag aggcataaat tccgtcagcc agtttagtct gaccatctca	3900
tctgtaacat cattggcaac gctacctttg ccatgtttca gaaacaactc tggcgcacatg	3960
ggcttccat acaatcgata gattgtcgca cctgattgcc cgacattatc gcgagcccat	4020
ttataccat ataaatcagc atccatgttg gaatttaac gcggcctcga gcaagacgtt	4080
tcccggtgaa tatggctcat aacaccctt gtattactgt ttatgtaagc agacagtttt	4140
attgttcatg atgatatatt tttatcttgt gcaatgtaac atcagagatt ttgagacaca	4200
acgtggcttt cccccccc ccattattga agcatttate agggttattg tctcatgagc	4260
ggatacatat ttgaatgtat ttagaaaaat aaacaaatag gggttccgcg cacatttccc	4320
cgaaaagtgc cacctgacgt ctaagaaacc attattatca tgacattaac ctataaaaat	4380
aggcgtatca cgaggccctt tcgtc	4405
<210> 18	
<211> 491	
<212> PRT	
<213> 人工序列	

<220>
 <223> L-10 = LASV_III_IV_anc
 <400> 18
 Met Gly Gln Ile Val Thr Phe Phe Gln Glu Val Pro His Val Ile Glu
 1 5 10 15
 Glu Val Met Asn Ile Val Leu Ile Ala Leu Ser Leu Leu Ala Ile Leu
 20 25 30
 Lys Gly Leu Tyr Asn Val Ala Thr Cys Gly Leu Ile Gly Leu Val Thr
 35 40 45
 Phe Leu Leu Leu Cys Gly Arg Ser Cys Ser Thr Thr Leu Tyr Lys Gly
 50 55 60
 Val Tyr Glu Leu Gln Thr Leu Glu Leu Asn Met Glu Thr Leu Asn Met
 65 70 75 80
 Thr Met Pro Leu Ser Cys Thr Lys Asn Asn Ser His His Tyr Ile Arg
 85 90 95
 Val Gly Asn Glu Thr Gly Leu Glu Leu Thr Leu Thr Asn Thr Ser Ile
 100 105 110
 Ile Asn His Lys Phe Cys Asn Leu Ser Asp Ala His Lys Lys Asn Leu
 115 120 125
 [0024] Tyr Asp His Ala Leu Met Ser Ile Ile Ser Thr Phe His Leu Ser Ile
 130 135 140
 Pro Asn Phe Asn Gln Tyr Glu Ala Met Ser Cys Asp Phe Asn Gly Gly
 145 150 155 160
 Lys Ile Ser Val Gln Tyr Asn Leu Ser His Ser Tyr Ala Val Asp Ala
 165 170 175
 Ala Asn His Cys Gly Thr Val Ala Asn Gly Val Leu Gln Thr Phe Met
 180 185 190
 Arg Met Ala Trp Gly Gly Ser Tyr Ile Ala Leu Asp Ser Gly Arg Gly
 195 200 205
 Asn Trp Asp Cys Ile Met Thr Ser Tyr Gln Tyr Leu Ile Ile Gln Asn
 210 215 220
 Thr Thr Trp Glu Asp His Cys Gln Phe Ser Arg Pro Ser Pro Ile Gly
 225 230 235 240
 Tyr Leu Gly Leu Leu Ser Gln Arg Thr Arg Asp Ile Tyr Ile Ser Arg
 245 250 255
 Arg Leu Leu Gly Thr Phe Thr Trp Thr Leu Ser Asp Ser Glu Gly Asn
 260 265 270
 Glu Thr Pro Gly Gly Tyr Cys Leu Thr Arg Trp Met Leu Ile Glu Ala
 275 280 285

Glu Leu Lys Cys Phe Gly Asn Thr Ala Val Ala Lys Cys Asn Glu Lys
 290 295 300

His Asp Glu Glu Phe Cys Asp Met Leu Arg Leu Phe Asp Phe Asn Lys
 305 310 315 320

Gln Ala Ile Arg Arg Leu Lys Ala Glu Ala Gln Met Ser Ile Gln Leu
 325 330 335

Ile Asn Lys Ala Val Asn Ala Leu Ile Asn Asp Gln Leu Ile Met Lys
 340 345 350

Asn His Leu Arg Asp Ile Met Gly Ile Pro Tyr Cys Asn Tyr Ser Lys
 355 360 365

Tyr Trp Tyr Leu Asn His Thr Ile Thr Gly Lys Thr Ser Leu Pro Lys
 370 375 380

Cys Trp Leu Val Ser Asn Gly Ser Tyr Leu Asn Glu Thr His Phe Ser
 385 390 395 400

Asp Asp Ile Glu Gln Gln Ala Asp Asn Met Ile Thr Glu Met Leu Gln
 405 410 415

Lys Glu Tyr Met Asp Arg Gln Gly Lys Thr Pro Leu Gly Leu Val Asp
 420 425 430

[0025] Leu Phe Val Phe Ser Thr Ser Phe Tyr Leu Ile Ser Ile Phe Leu His
 435 440 445

Leu Val Lys Ile Pro Thr His Arg His Ile Val Gly Lys Pro Cys Pro
 450 455 460

Lys Pro His Arg Leu Asn His Met Gly Ile Cys Ser Cys Gly Leu Tyr
 465 470 475 480

Lys Gln Pro Gly Val Pro Val Arg Trp Lys Arg
 485 490

<210> 19

<211> 1476

<212> DNA

<213> 人工序列

<220>

<223> L-10 = LASV_III_IV_anc

<400> 19

atgggccaga tcgtgacatt cttccaagag gtgccccacg tgatcgaaga agtgatgaac 60

atcgtcctga tcgcctgag cctgctggcc atcctgaagg gcctgtataa tgtggccacc 120

tgtggcctga tcggcctggt cacatttctg ctgctgtgcg gcagaagctg ctccaccaca 180

ctgtataagg gcgtgtacga gctgcaaacc ctggaactga acatggaaac cctgaacatg 240

accatgcctc tgagctgcac caagaacaac agccaccact acatcagagt gggcaacgag 300

acaggcctcg agctgaccct gaccaacacc agcatcatca accacaagtt ctgcaacctg 360

agcgacgccc acaagaagaa cctgtacgat cacgccctga tgagcatcat ctccaccttc 420

cacctgagca tccccaactt caaccagtac gaggccatga gctgcgactt caacggcgga	480
aagatcagcg tgcagtacaa tctgagccac agctatgccg tggacgccgc caatcattgt	540
ggaacagtgg ccaatggcgt gctccagacc ttcattgagaa tggcctgggg cggcagctat	600
atcgccctgg attctggcag aggcaactgg gactgcatca tgaccagcta ccagtacctg	660
atcatccaga acaccacctg ggaagatcac tgccagtcca gcagaccctc tcctatcgga	720
tacctgggcc tgctgtccca gagaaccggg gacatctaca tctctagacg gctgctgggc	780
accttcacct ggacactgtc tgatagcgag ggcaatgaga cacctggcgg ctactgtctg	840
accgggtgga tgctgattga ggccgagctg aagtgtctcg gaaataccgc cgtggccaag	900
tgcaacgaga agcacgacga ggaattctgc gacatgctgc ggctgttcga tttcaacaag	960
caggccatca gacggctgaa ggccgaggct cagatgtcca tccagctgat caacaaggcc	1020
gtgaatgccc tgattaacga ccagctcacc atgaagaacc acctcagga catcatgggc	1080
atcccttact gcaactacag caagtactgg tatctgaacc acaccatcac cggcaagacc	1140
agcctgccta agtgtggct ggtgtccaac ggcagctacc tgaacgagac acacttcage	1200
gacgacatcg agcagcaggc cgacaacatg atcaccgaga tgctccagaa agagtacatg	1260
gaccggcagg gcaagacacc tctgggcctt gtggatctgt tcgtgttcag caccagcttc	1320
tacctgatct ctatcttctt gcacctggtc aagatcccca cacacagaca catcgtgggc	1380
aagccctgtc ctaagcctca cagactgaac catatgggca tctgtagctg cggcctgtac	1440
aaacagcctg gcgtgccagt gcggtggaag agataa	1476

[0026]

<210> 20
<211> 491
<212> PRT
<213> 人工序列
<220>
<223> L-10-SOSEP
<400> 20
Met Gly Gln Ile Val Thr Phe Phe Gln Glu Val Pro His Val Ile Glu
1 5 10 15
Glu Val Met Asn Ile Val Leu Ile Ala Leu Ser Leu Leu Ala Ile Leu
20 25 30
Lys Gly Leu Tyr Asn Val Ala Thr Cys Gly Leu Ile Gly Leu Val Thr
35 40 45
Phe Leu Leu Leu Cys Gly Arg Ser Cys Ser Thr Thr Leu Tyr Lys Gly
50 55 60
Val Tyr Glu Leu Gln Thr Leu Glu Leu Asn Met Glu Thr Leu Asn Met
65 70 75 80
Thr Met Pro Leu Ser Cys Thr Lys Asn Asn Ser His His Tyr Ile Arg
85 90 95
Val Gly Asn Glu Thr Gly Leu Glu Leu Thr Leu Thr Asn Thr Ser Ile
100 105 110

Ile Asn His Lys Phe Cys Asn Leu Ser Asp Ala His Lys Lys Asn Leu
 115 120 125

Tyr Asp His Ala Leu Met Ser Ile Ile Ser Thr Phe His Leu Ser Ile
 130 135 140

Pro Asn Phe Asn Gln Tyr Glu Ala Met Ser Cys Asp Phe Asn Gly Gly
 145 150 155 160

Lys Ile Ser Val Gln Tyr Asn Leu Ser His Ser Tyr Ala Val Asp Ala
 165 170 175

Ala Asn His Cys Gly Thr Val Ala Asn Gly Val Leu Gln Thr Phe Met
 180 185 190

Arg Met Ala Trp Gly Gly Ser Tyr Ile Ala Leu Asp Ser Gly Cys Gly
 195 200 205

Asn Trp Asp Cys Ile Met Thr Ser Tyr Gln Tyr Leu Ile Ile Gln Asn
 210 215 220

Thr Thr Trp Glu Asp His Cys Gln Phe Ser Arg Pro Ser Pro Ile Gly
 225 230 235 240

Tyr Leu Gly Leu Leu Ser Gln Arg Thr Arg Asp Ile Tyr Ile Ser Arg
 245 250 255

[0027] Arg Arg Arg Gly Thr Phe Thr Trp Thr Leu Ser Asp Ser Glu Gly Asn
 260 265 270

Glu Thr Pro Gly Gly Tyr Cys Leu Thr Arg Trp Met Leu Ile Glu Ala
 275 280 285

Glu Leu Lys Cys Phe Gly Asn Thr Ala Val Ala Lys Cys Asn Glu Lys
 290 295 300

His Asp Glu Glu Phe Cys Asp Met Leu Arg Leu Phe Asp Phe Asn Lys
 305 310 315 320

Gln Ala Ile Arg Arg Leu Lys Ala Pro Ala Gln Met Ser Ile Gln Leu
 325 330 335

Ile Asn Lys Ala Val Asn Ala Leu Ile Asn Asp Gln Leu Ile Met Lys
 340 345 350

Asn His Leu Arg Asp Ile Met Cys Ile Pro Tyr Cys Asn Tyr Ser Lys
 355 360 365

Tyr Trp Tyr Leu Asn His Thr Ile Thr Gly Lys Thr Ser Leu Pro Lys
 370 375 380

Cys Trp Leu Val Ser Asn Gly Ser Tyr Leu Asn Glu Thr His Phe Ser
 385 390 395 400

Asp Asp Ile Glu Gln Gln Ala Asp Asn Met Ile Thr Glu Met Leu Gln
 405 410 415

Lys Glu Tyr Met Asp Arg Gln Gly Lys Thr Pro Leu Gly Leu Val Asp
 420 425 430
 Leu Phe Val Phe Ser Thr Ser Phe Tyr Leu Ile Ser Ile Phe Leu His
 435 440 445
 Leu Val Lys Ile Pro Thr His Arg His Ile Val Gly Lys Pro Cys Pro
 450 455 460
 Lys Pro His Arg Leu Asn His Met Gly Ile Cys Ser Cys Gly Leu Tyr
 465 470 475 480
 Lys Gln Pro Gly Val Pro Val Arg Trp Lys Arg
 485 490

<210> 21
 <211> 1476
 <212> DNA
 <213> 人工序列

<220>
 <223> L-10-SOSEP

[0028]

<400> 21
 atgggccaga tcgtgacatt cttccaagag gtgccccacg tgatcgaaga agtgatgaac 60
 atcgtcctga tcgccctgag cctgctggcc atcctgaagg gcctgtataa tgtggccacc 120
 tgtggcctga tcggcctggt cacatttctg ctgctgtgcg gcagaagctg ctccaccaca 180
 ctgtataagg gcgtgtacga gctgcaaacc ctggaactga acatggaaac cctgaacatg 240
 accatgcctc tgagctgcac caagaacaac agccaccact acatcagagt gggcaacgag 300
 acaggcctcg agctgacct gaccaacacc agcatcatca accacaagtt ctgcaacctg 360
 agcgacgcc acaagaagaa cctgtacgat cacgccctga tgagcatcat ctccaccttc 420
 cacctgagca tccccactt caaccagtac gaggccatga gctgcgactt caacggcgga 480
 aagatcagcg tgcaagtacaa tctgagccac agctatgccg tggacgccgc caatcattgt 540
 ggaacagtgg ccaatggcgt gctccagacc ttcattgagaa tggcctgggg cggcagctat 600
 atcgccctgg attctggctg tggcaactgg gactgcatca tgaccagcta ccagtacctg 660
 atcatccaga acaccacctg ggaagatcac tgccagtcca gcagaccctc tectatcgga 720
 tacctgggcc tgctgtccca gagaaccggg gacatctaca tctctcggcg gagaagaggc 780
 accttcacct ggacactgtc tgatagcgag ggcaatgaga cacctggcgg ctactgtctg 840
 acccggtgga tgctgattga ggccgagctg aagtgcctcg gaaataccgc cgtggccaag 900
 tgcaacgaga agcacgacga ggaattctgc gacatgctgc ggctgttcga ttcaacaag 960
 caggccatca gacggctgaa ggcccctgct cagatgtcca tccagctgat caacaaggcc 1020
 gtgaatgcc tgattaacga ccagctcatc atgaagaacc acctcagga catcatgtgc 1080
 atcccttact gcaactacag caagtactgg tatctgaacc acaccatcac cggcaagacc 1140
 agcctgccta agtgctggct ggtgtccaac ggcagctacc tgaacgagac acacttcagc 1200
 gacgacatcg agcagcaggc cgacaacatg atcaccgaga tgctccagaa agagtacatg 1260
 gaccggcagg gcaagacacc tctgggcctt ttggatctgt tcgtgttcag caccagcttc 1320
 tacctgatct ctatcttctt gcacctggtc aagatcccca cacacagaca catcgtgggc 1380
 aagccctgtc ctaagcctca cagactgaac catatgggca tctgtagctg cggcctgtac 1440

aaacagcctg gcgtgccagt gcggtggaag agataa

1476

<210> 22
 <211> 491
 <212> PRT
 <213> 人工序列

<220>
 <223> L-10-S0SEP-NtoK

<400> 22

Met Gly Gln Ile Val Thr Phe Phe Gln Glu Val Pro His Val Ile Glu
 1 5 10 15

Glu Val Met Asn Ile Val Leu Ile Ala Leu Ser Leu Leu Ala Ile Leu
 20 25 30

Lys Gly Leu Tyr Asn Val Ala Thr Cys Gly Leu Ile Gly Leu Val Thr
 35 40 45

Phe Leu Leu Leu Cys Gly Arg Ser Cys Ser Thr Thr Leu Tyr Lys Gly
 50 55 60

Val Tyr Glu Leu Gln Thr Leu Glu Leu Asn Met Glu Thr Leu Asn Met
 65 70 75 80

Thr Met Pro Leu Ser Cys Thr Lys Asn Asn Ser His His Tyr Ile Arg
 85 90 95

[0029] Val Gly Asn Glu Thr Gly Leu Glu Leu Thr Leu Thr Asn Thr Ser Ile
 100 105 110

Ile Asn His Lys Phe Cys Asn Leu Ser Asp Ala His Lys Lys Asn Leu
 115 120 125

Tyr Asp His Ala Leu Met Ser Ile Ile Ser Thr Phe His Leu Ser Ile
 130 135 140

Pro Asn Phe Asn Gln Tyr Glu Ala Met Ser Cys Asp Phe Asn Gly Gly
 145 150 155 160

Lys Ile Ser Val Gln Tyr Asn Leu Ser His Ser Tyr Ala Val Asp Ala
 165 170 175

Ala Asn His Cys Gly Thr Val Ala Asn Gly Val Leu Gln Thr Phe Met
 180 185 190

Arg Met Ala Trp Gly Gly Ser Tyr Ile Ala Leu Asp Ser Gly Cys Gly
 195 200 205

Asn Trp Asp Cys Ile Met Thr Ser Tyr Gln Tyr Leu Ile Ile Gln Asn
 210 215 220

Thr Thr Trp Glu Asp His Cys Gln Phe Ser Arg Pro Ser Pro Ile Gly
 225 230 235 240

Tyr Leu Gly Leu Leu Ser Gln Arg Thr Arg Asp Ile Tyr Ile Ser Arg
 245 250 255

Arg Arg Arg Gly Thr Phe Thr Trp Thr Leu Ser Asp Ser Glu Gly Lys
 260 265 270

Glu Thr Pro Gly Gly Tyr Cys Leu Thr Arg Trp Met Leu Ile Glu Ala
 275 280 285

Glu Leu Lys Cys Phe Gly Asn Thr Ala Val Ala Lys Cys Asn Glu Lys
 290 295 300

His Asp Glu Glu Phe Cys Asp Met Leu Arg Leu Phe Asp Phe Asn Lys
 305 310 315 320

Gln Ala Ile Arg Arg Leu Lys Ala Pro Ala Gln Met Ser Ile Gln Leu
 325 330 335

Ile Asn Lys Ala Val Asn Ala Leu Ile Asn Asp Gln Leu Ile Met Lys
 340 345 350

Asn His Leu Arg Asp Ile Met Cys Ile Pro Tyr Cys Asn Tyr Ser Lys
 355 360 365

Tyr Trp Tyr Leu Asn His Thr Ile Thr Gly Lys Thr Ser Leu Pro Lys
 370 375 380

Cys Trp Leu Val Ser Asn Gly Ser Tyr Leu Asn Glu Thr His Phe Ser
 385 390 395 400

[0030] Asp Asp Ile Glu Gln Gln Ala Asp Asn Met Ile Thr Glu Met Leu Gln
 405 410 415

Lys Glu Tyr Met Asp Arg Gln Gly Lys Thr Pro Leu Gly Leu Val Asp
 420 425 430

Leu Phe Val Phe Ser Thr Ser Phe Tyr Leu Ile Ser Ile Phe Leu His
 435 440 445

Leu Val Lys Ile Pro Thr His Arg His Ile Val Gly Lys Pro Cys Pro
 450 455 460

Lys Pro His Arg Leu Asn His Met Gly Ile Cys Ser Cys Gly Leu Tyr
 465 470 475 480

Lys Gln Pro Gly Val Pro Val Arg Trp Lys Arg
 485 490

<210> 23
 <211> 1476
 <212> DNA
 <213> 人工序列

<220>
 <223> L-10-S0SEP-NtoK

<400> 23
 atgggccaga tcgtgacatt cttccaagag gtgccccacg tgatcgaaga agtgatgaac 60
 atcgtcctga tcgcctgag cctgctggcc atcctgaagg gcctgtataa tgtggccacc 120
 tgtggcctga tcggcctggt cacatttctg ctgctgtgcg gcagaagctg ctccaccaca 180

ctgtataagg gcgtgtacga gctgcaaacc ctggaactga acatggaaac cctgaacatg	240
accatgcctc tgagctgcac caagaacaac agccaccact acatcagagt gggcaacgag	300
acaggcctcg agctgacct gaccaacacc agcatcatca accacaagtt ctgcaacctg	360
agcgacgccc acaagaagaa cctgtacgat cacgccctga tgagcatcat ctccaccttc	420
cacctgagca tccccaaact caaccagtac gaggccatga gctgcgactt caacggcgga	480
aagatcagcg tgacgtacaa tctgagccac agctatgccg tggacgccgc caatcattgt	540
ggaacagtgg ccaatggcgt gctccagacc ttcattgagaa tggcctgggg cggcagctat	600
atcgccctgg attctggctg tggcaactgg gactgcatca tgaccagcta ccagtacctg	660
atcatccaga acaccacctg ggaagatcac tgccagttca gcagaccctc tcctatcgga	720
tacctgggcc tgctgtcca gagaaccggg gacatctaca tctctcggcg gagaagaggc	780
accttcacct ggacactgtc tgatagcgag ggcaaagaga cacctggcgg ctactgtctg	840
accgggtgga tgctgattga ggccgagctg aagtgtctcg gaaataccgc cgtggccaag	900
tgcaacgaga agcagcagca ggaattctgc gacatgctgc ggctgttcga tttcaacaag	960
caggccatca gacggctgaa ggcccctgct cagatgtcca tccagctgat caacaaggcc	1020
gtgaatgccc tgattaacga ccagctcatc atgaagaacc acctcagga catcatgtgc	1080
atcccttact gcaactacag caagtactgg tatctgaacc acaccatcac cggcaagacc	1140
agcctgccta agtgtggct ggtgtccaac ggcagctacc tgaacgagac acacttcagc	1200
gacgacatcg agcagcaggc cgacaacatg atcaccgaga tgctccagaa agagtacatg	1260
[0031] gaccggcagg gcaagacacc tctgggcctt gtggatctgt tcgtgttcag caccagcttc	1320
tacctgatct ctatcttct gcacctggtc aagatcccca cacacagaca catcgtgggc	1380
aagccctgtc ctaagcctca cagactgaac catatgggca tctgtagctg cggcctgtac	1440
aaacagcctg gcgtgccagt gcggtggaag agataa	1476

<210> 24
 <211> 497
 <212> PRT
 <213> 人工序列

<220>
 <223> L-10-FLEP

<400> 24

Met Gly Gln Ile Val Thr Phe Phe Gln Glu Val Pro His Val Ile Glu
 1 5 10 15

Glu Val Met Asn Ile Val Leu Ile Ala Leu Ser Leu Leu Ala Ile Leu
 20 25 30

Lys Gly Leu Tyr Asn Val Ala Thr Cys Gly Leu Ile Gly Leu Val Thr
 35 40 45

Phe Leu Leu Leu Cys Gly Arg Ser Cys Ser Thr Thr Leu Tyr Lys Gly
 50 55 60

Val Tyr Glu Leu Gln Thr Leu Glu Leu Asn Met Glu Thr Leu Asn Met
 65 70 75 80

Thr Met Pro Leu Ser Cys Thr Lys Asn Asn Ser His His Tyr Ile Arg
 85 90 95

Val Gly Asn Glu Thr Gly Leu Glu Leu Thr Leu Thr Asn Thr Ser Ile
 100 105 110

Ile Asn His Lys Phe Cys Asn Leu Ser Asp Ala His Lys Lys Asn Leu
 115 120 125

Tyr Asp His Ala Leu Met Ser Ile Ile Ser Thr Phe His Leu Ser Ile
 130 135 140

Pro Asn Phe Asn Gln Tyr Glu Ala Met Ser Cys Asp Phe Asn Gly Gly
 145 150 155 160

Lys Ile Ser Val Gln Tyr Asn Leu Ser His Ser Tyr Ala Val Asp Ala
 165 170 175

Ala Asn His Cys Gly Thr Val Ala Asn Gly Val Leu Gln Thr Phe Met
 180 185 190

Arg Met Ala Trp Gly Gly Ser Tyr Ile Ala Leu Asp Ser Gly Arg Gly
 195 200 205

Asn Trp Asp Cys Ile Met Thr Ser Tyr Gln Tyr Leu Ile Ile Gln Asn
 210 215 220

[0032] Thr Thr Trp Glu Asp His Cys Gln Phe Ser Arg Pro Ser Pro Ile Gly
 225 230 235 240

Tyr Leu Gly Leu Leu Ser Gln Arg Thr Arg Asp Ile Tyr Ile Ser Gly
 245 250 255

Gly Gly Gly Ser Gly Gly Gly Gly Ser Gly Thr Phe Thr Trp Thr Leu
 260 265 270

Ser Asp Ser Glu Gly Asn Glu Thr Pro Gly Gly Tyr Cys Leu Thr Arg
 275 280 285

Trp Met Leu Ile Glu Ala Glu Leu Lys Cys Phe Gly Asn Thr Ala Val
 290 295 300

Ala Lys Cys Asn Glu Lys His Asp Glu Glu Phe Cys Asp Met Leu Arg
 305 310 315 320

Leu Phe Asp Phe Asn Lys Gln Ala Ile Arg Arg Leu Lys Ala Pro Ala
 325 330 335

Gln Met Ser Ile Gln Leu Ile Asn Lys Ala Val Asn Ala Leu Ile Asn
 340 345 350

Asp Gln Leu Ile Met Lys Asn His Leu Arg Asp Ile Met Gly Ile Pro
 355 360 365

Tyr Cys Asn Tyr Ser Lys Tyr Trp Tyr Leu Asn His Thr Ile Thr Gly
 370 375 380

Lys Thr Ser Leu Pro Lys Cys Trp Leu Val Ser Asn Gly Ser Tyr Leu
385 390 395 400

Asn Glu Thr His Phe Ser Asp Asp Ile Glu Gln Gln Ala Asp Asn Met
405 410 415

Ile Thr Glu Met Leu Gln Lys Glu Tyr Met Asp Arg Gln Gly Lys Thr
420 425 430

Pro Leu Gly Leu Val Asp Leu Phe Val Phe Ser Thr Ser Phe Tyr Leu
435 440 445

Ile Ser Ile Phe Leu His Leu Val Lys Ile Pro Thr His Arg His Ile
450 455 460

Val Gly Lys Pro Cys Pro Lys Pro His Arg Leu Asn His Met Gly Ile
465 470 475 480

Cys Ser Cys Gly Leu Tyr Lys Gln Pro Gly Val Pro Val Arg Trp Lys
485 490 495

Arg

<210> 25
<211> 1494
<212> DNA
<213> 人工序列

[0033]

<220>
<223> L-10-FLEP

<400> 25
atgggccaga tcgtgacatt cttccaagag gtgccccacg tgatcgaaga agt gatgaac 60
atcgtcctga tcgccctgag cctgctggcc atcctgaagg gcctgtataa tgtggccacc 120
tgtggcctga tcggcctggt cacatttctg ctgctgtgcg gcagaagctg ctccaccaca 180
ctgtataagg gcgtgtacga gctgcaaacc ctggaactga acatggaaac cctgaacatg 240
accatgcctc tgagctgcac caagaacaac agccaccact acatcagagt gggcaacgag 300
acaggcctcg agctgacct gaccaacacc agcatcatca accacaagtt ctgcaacctg 360
agcgagcccc acaagaagaa cctgtacgat cagccccgta tgagcatcat ctccaccttc 420
cacctgagca tccccactt caaccagtac gaggccatga gctgcgactt caacggcgga 480
aagatcagcg tcagtagaaa tctgagccac agctatgccg tggacgccc caatcattgt 540
ggaacagtgg ccaatggcgt gctccagacc tcatgagaa tggcctgggg cggcagctat 600
atcgccctgg attctggcag aggcaactgg gactgcatca tgaccagcta ccagtacctg 660
atcatccaga acaccacctg ggaagatcac tgccagtcca gcagaccctc tcctatcgga 720
tacctgggcc tgcgtgccca gagaaccggg gacatctaca tctctggcgg cggaggatct 780
ggcggaggtg gaagtggcac cttcacctgg aactgtctg atagcgaggg caatgagaca 840
cctggcgget actgtctgac ccggtggatg ctgattgagg ccgagctgaa gtgcttcgga 900
aataccgccg tgccaagtg caacgagaag cagcagagg aattctgca catgctgcgg 960
ctgttcgatt tcaacaagca ggccatcaga cggctgaagg cccctgctca gatgtccatc 1020
cagctgatca acaaggcctg gaatgcctg attaacgacc agctcatcat gaagaaccac 1080

ctcagggaca tcatgggcat cccttactgc aactacagca agtactggta tctgaaccac 1140
 accatcacccg gcaagaccag cctgcctaag tgctggctgg tgtccaacgg cagctacctg 1200
 aacgagacac acttcagcga cgacatcgag cagcaggccg acaacatgat caccgagatg 1260
 ctccagaaag agtacatgga cgggcagggc aagacacctc tgggccttgt ggatctgttc 1320
 gtgttcagca ccagcttcta cctgatctct atcttctctg acctgggtcaa gatccccaca 1380
 cacagacaca tcgtgggcaa gcctgtcct aagcctcaca gactgaacca tatgggcatc 1440
 tgtagctgcg gcctgtacaa acagcctggc gtgccagtgc ggtggaagag ataa 1494

<210> 26
 <211> 497
 <212> PRT
 <213> 人工序列

<220>
 <223> L-10-FLEP-NtoK

<400> 26

Met Gly Gln Ile Val Thr Phe Phe Gln Glu Val Pro His Val Ile Glu
 1 5 10 15

Glu Val Met Asn Ile Val Leu Ile Ala Leu Ser Leu Leu Ala Ile Leu
 20 25 30

Lys Gly Leu Tyr Asn Val Ala Thr Cys Gly Leu Ile Gly Leu Val Thr
 35 40 45

[0034] Phe Leu Leu Leu Cys Gly Arg Ser Cys Ser Thr Thr Leu Tyr Lys Gly
 50 55 60

Val Tyr Glu Leu Gln Thr Leu Glu Leu Asn Met Glu Thr Leu Asn Met
 65 70 75 80

Thr Met Pro Leu Ser Cys Thr Lys Asn Asn Ser His His Tyr Ile Arg
 85 90 95

Val Gly Asn Glu Thr Gly Leu Glu Leu Thr Leu Thr Asn Thr Ser Ile
 100 105 110

Ile Asn His Lys Phe Cys Asn Leu Ser Asp Ala His Lys Lys Asn Leu
 115 120 125

Tyr Asp His Ala Leu Met Ser Ile Ile Ser Thr Phe His Leu Ser Ile
 130 135 140

Pro Asn Phe Asn Gln Tyr Glu Ala Met Ser Cys Asp Phe Asn Gly Gly
 145 150 155 160

Lys Ile Ser Val Gln Tyr Asn Leu Ser His Ser Tyr Ala Val Asp Ala
 165 170 175

Ala Asn His Cys Gly Thr Val Ala Asn Gly Val Leu Gln Thr Phe Met
 180 185 190

Arg Met Ala Trp Gly Gly Ser Tyr Ile Ala Leu Asp Ser Gly Arg Gly
 195 200 205

Asn Trp Asp Cys Ile Met Thr Ser Tyr Gln Tyr Leu Ile Ile Gln Asn
 210 215 220

Thr Thr Trp Glu Asp His Cys Gln Phe Ser Arg Pro Ser Pro Ile Gly
 225 230 235 240

Tyr Leu Gly Leu Leu Ser Gln Arg Thr Arg Asp Ile Tyr Ile Ser Gly
 245 250 255

Gly Gly Gly Ser Gly Gly Gly Gly Ser Gly Thr Phe Thr Trp Thr Leu
 260 265 270

Ser Asp Ser Glu Gly Lys Glu Thr Pro Gly Gly Tyr Cys Leu Thr Arg
 275 280 285

Trp Met Leu Ile Glu Ala Glu Leu Lys Cys Phe Gly Asn Thr Ala Val
 290 295 300

Ala Lys Cys Asn Glu Lys His Asp Glu Glu Phe Cys Asp Met Leu Arg
 305 310 315 320

Leu Phe Asp Phe Asn Lys Gln Ala Ile Arg Arg Leu Lys Ala Pro Ala
 325 330 335

Gln Met Ser Ile Gln Leu Ile Asn Lys Ala Val Asn Ala Leu Ile Asn
 340 345 350

[0035] Asp Gln Leu Ile Met Lys Asn His Leu Arg Asp Ile Met Gly Ile Pro
 355 360 365

Tyr Cys Asn Tyr Ser Lys Tyr Trp Tyr Leu Asn His Thr Ile Thr Gly
 370 375 380

Lys Thr Ser Leu Pro Lys Cys Trp Leu Val Ser Asn Gly Ser Tyr Leu
 385 390 395 400

Asn Glu Thr His Phe Ser Asp Asp Ile Glu Gln Gln Ala Asp Asn Met
 405 410 415

Ile Thr Glu Met Leu Gln Lys Glu Tyr Met Asp Arg Gln Gly Lys Thr
 420 425 430

Pro Leu Gly Leu Val Asp Leu Phe Val Phe Ser Thr Ser Phe Tyr Leu
 435 440 445

Ile Ser Ile Phe Leu His Leu Val Lys Ile Pro Thr His Arg His Ile
 450 455 460

Val Gly Lys Pro Cys Pro Lys Pro His Arg Leu Asn His Met Gly Ile
 465 470 475 480

Cys Ser Cys Gly Leu Tyr Lys Gln Pro Gly Val Pro Val Arg Trp Lys
 485 490 495

Arg

<210> 27
 <211> 1494
 <212> DNA
 <213> 人工序列

<220>
 <223> L-10-FLEP-NtoK

<400> 27
 atgggccaga tcgtgacatt cttccaagag gtgccccacg tgatcgaaga agtgatgaac 60
 atcgtcctga tcgccctgag cctgctggcc atcctgaagg gcctgtataa tgtggccacc 120
 tgtggcctga tcggcctggt cacatttctg ctgctgtgcg gcagaagctg ctccaccaca 180
 ctgtataagg gcgtgtacga gctgcaaacc ctggaactga acatggaaac cctgaacatg 240
 accatgcctc tgagctgcac caagaacaac agccaccact acatcagagt gggcaacgag 300
 acaggcctcg agctgaccct gaccaacacc agcatcatca accacaagtt ctgcaacctg 360
 agcgacgccc acaagaagaa cctgtacgat cacgccctga tgagcatcat ctccaccttc 420
 cacctgagca tccccaactt caaccagtac gaggccatga gctgcgactt caacggcgga 480
 aagatcagcg tgacgtacaa tctgagccac agctatgccg tggacgccgc caatcattgt 540
 ggaacagtgg ccaatggcgt gctccagacc ttcattgagaa tggcctgggg cggcagctat 600
 atgccctgg attctggcag aggcaactgg gactgcatca tgaccagcta ccagtacctg 660
 atcatccaga acaccacctg ggaagatcac tgccagttca gcagaccctc tcctatcgga 720
 tacctgggcc tgctgtccca gagaaccgag gacatctaca tctctggcgg cggaggatct 780
 [0036] ggcggagggt gaagtggcac cttcacctgg aactgtctg atagcgaggg caaagagaca 840
 cctggcggct actgtctgac ccgggtgatg ctgattgagg ccgagctgaa gtgcttcgga 900
 aataccgccg tggccaagtg caacgagaag cacgacgagg aattctgcga catgctgcgg 960
 ctgttcgatt tcaacaagca ggccatcaga cggctgaagg cccctgctca gatgtccatc 1020
 cagctgatca acaaggcctg gaatgccctg attaacgacc agctcatcat gaagaaccac 1080
 ctcagggaca tcatgggcat cccttactgc aactacagca agtactggtg tctgaaccac 1140
 accatcaccg gcaagaccag cctgcctaag tgctggctgg tgtccaacgg cagctacctg 1200
 aacgagacac acttcagcga cgacatcgag cagcaggccg acaacatgat caccgagatg 1260
 ctccagaaag agtatactga ccggcagggc aagacacctc tgggccttgt ggatctgttc 1320
 gtgttcagca ccagcttcta cctgatctct atcttcttgc acctggtcaa gatccccaca 1380
 cacagacaca tcgtgggcaa gcctgtcct aagcctcaca gactgaacca tatgggcatc 1440
 ttagctgctg gcctgtacaa acagcctggc gtgccagtgc ggtggaagag ataa 1494

<210> 28
 <211> 569
 <212> PRT
 <213> 人工序列

<220>
 <223> L-NP-1 = L-NP-CovAnc-1_N

<400> 28
 Met Ser Ala Ser Lys Glu Val Lys Ser Phe Leu Trp Thr Gln Ser Leu
 1 5 10 15

Arg Arg Glu Leu Ser Gly Tyr Cys Ser Asn Ile Lys Leu Gln Val Val
 20 25 30

Lys Asp Ala Gln Ala Leu Leu His Gly Leu Asp Phe Ser Glu Val Ser
 35 40 45

Asn Val Gln Arg Leu Met Arg Lys Gln Lys Arg Asp Asp Ser Asp Leu
 50 55 60

Lys Arg Leu Arg Asp Leu Asn Gln Ala Val Asn Asn Leu Val Glu Leu
 65 70 75 80

Lys Ser Thr Gln Gln Lys Ser Ile Leu Arg Val Gly Thr Leu Thr Ser
 85 90 95

Asp Asp Leu Leu Thr Leu Ala Ala Asp Leu Glu Lys Leu Lys Ser Lys
 100 105 110

Val Ile Arg Thr Glu Arg Pro Leu Ser Ser Gly Val Tyr Met Gly Asn
 115 120 125

Leu Ser Thr Gln Gln Leu Glu Gln Arg Arg Ala Leu Leu Asn Met Ile
 130 135 140

Gly Met Val Gly Gly Ala Gln Gly Thr Gln Pro Gly Arg Asp Gly Val
 145 150 155 160

[0037] Val Arg Val Trp Asp Val Lys Asn Pro Asp Leu Leu Asn Asn Gln Phe
 165 170 175

Gly Thr Met Pro Ser Leu Thr Leu Ala Cys Leu Thr Lys Gln Gly Gln
 180 185 190

Val Asp Leu Asn Asp Ala Val Leu Ala Leu Thr Asp Leu Gly Leu Ile
 195 200 205

Tyr Thr Ala Lys Tyr Pro Asn Ser Ser Asp Leu Asp Arg Leu Ser Gln
 210 215 220

Ser His Pro Ile Leu Asn Met Val Asp Thr Lys Lys Ser Ser Leu Asn
 225 230 235 240

Ile Ser Gly Tyr Asn Phe Ser Leu Gly Ala Ala Val Lys Ala Gly Ala
 245 250 255

Cys Met Leu Asp Gly Gly Asn Met Leu Glu Thr Ile Lys Val Thr Pro
 260 265 270

Gln Thr Met Asp Gly Ile Leu Lys Ser Ile Leu Lys Val Lys Lys Ser
 275 280 285

Leu Gly Met Phe Val Ser Asp Thr Pro Gly Glu Arg Asn Pro Tyr Glu
 290 295 300

Asn Ile Leu Tyr Lys Ile Cys Leu Ser Gly Asp Gly Trp Pro Tyr Ile
 305 310 315 320

Ala Ser Arg Thr Ser Ile Val Gly Arg Ala Trp Glu Asn Thr Thr Val
 325 330 335

Asp Leu Glu Ser Asp Gly Lys Pro Gln Lys Val Gly Thr Ala Gly Ser
 340 345 350

Asn Lys Ser Leu Gln Ser Ala Gly Phe Pro Thr Gly Leu Thr Tyr Ser
 355 360 365

Gln Leu Met Thr Leu Lys Asp Ser Met Met Gln Leu Asp Pro Ser Ala
 370 375 380

Lys Thr Trp Ile Asp Ile Glu Gly Arg Pro Glu Asp Pro Val Glu Ile
 385 390 395 400

Ala Leu Tyr Gln Pro Met Ser Gly Cys Tyr Ile His Phe Phe Arg Glu
 405 410 415

Pro Thr Asp Leu Lys Gln Phe Lys Gln Asp Ala Lys Tyr Ser His Gly
 420 425 430

Ile Asp Val Ala Asp Leu Phe Pro Ala Gln Pro Gly Leu Thr Ser Ala
 435 440 445

Val Ile Glu Ala Leu Pro Arg Asn Met Val Leu Thr Cys Gln Gly Ser
 450 455 460

Asp Asp Ile Lys Arg Leu Leu Asp Ser Gln Gly Arg Arg Asp Ile Lys
 465 470 475 480

Leu Ile Asp Ile Ala Leu Ser Lys Ala Asp Ser Arg Arg Phe Glu Asn
 485 490 495

Ala Val Trp Asp Gln Cys Lys Asp Leu Cys His Met His Thr Gly Val
 500 505 510

Val Val Glu Lys Lys Lys Arg Gly Gly Lys Glu Glu Ile Thr Pro His
 515 520 525

Cys Ala Leu Met Asp Cys Ile Met Tyr Asp Ala Ala Val Ser Gly Gly
 530 535 540

Leu Asn Ile Pro Val Leu Arg Ala Val Leu Pro Arg Asp Met Val Phe
 545 550 555 560

Arg Thr Ser Ser Pro Lys Val Val Leu
 565

[0038]

<210> 29
 <211> 1710
 <212> DNA
 <213> 人工序列

<220>
 <223> L-NP-1 = L-NP-CovAnc-1_N

<400> 29
 atgagcgcca gcaaagaagt gaaaagcttc ctctggaccc agagcctgcg gagagagctg 60
 tctggctact gctccaacat caagctccag gtggtaagg acgcccaggc tctgctgcat 120

ggcctggatt tcagcgaggt gtccaacgtg cagcggctga tgagaaagca gaagcgggac	180
gacagcgacc tgaagagact gagggatctg aaccaggccg tgaacaacct ggtggaactg	240
aagtctacce agcagaaaac catcctgaga gtgggcaccc tgaccagcga cgatctgctg	300
acactggccg ccgatctgga aaagctgaag tccaaagtga tccggaccga gaggccactg	360
tctagcggag tgtacatggg caacctgagc acccagcagc tggaaacagag aagggccctg	420
ctgaacatga tcggcatggt tggaggcgcc cagggaacac agcctggaag agatggtgtc	480
gtcagagtgt gggacgtgaa gaaccccgac ctgctcaaca accagtctgg caccatgcct	540
tctctgacct tggcctgcct gacaaagcag ggccaagtgg acctgaacga tgccgtgctg	600
gctctgactg atctgggctt gatctacacc gccaagtatc ccaacagctc cgacctggac	660
aggctgagcc agtctcacc catcctgaac atggtggaca ccaagaagtc cagcctgaac	720
atcagcggct acaacttctc tctgggcgct gccgtgaaag cggcgcttg tatgcttgac	780
ggcggaaca tgctggaac catcaaagt acccctcaga ccatggacgg catcctgaaa	840
agtatcctga aagtgaagaa atccctgggc atgttcgtgt ccgacacacc cggcgagaga	900
aaccctacg agaacatcct gtacaagatt tgcctgagcg gcgacggctg gccctatatc	960
gccagcagaa catctatcgt gggcagagct tgggagaaca ccaccgtgga cctggaatcc	1020
gatggcaagc ctcagaaagt gggcacagcc ggcagcaaca agagcctcca gtctgccgga	1080
tttctaccg gcctgacata cagccagctg atgaccctga aggacagcat gatgcagctg	1140
gaccctagcg ccaagacctg gatcgacatt gagggcagac ccgaggatcc cgtggaaatc	1200
gctctgtacc agcctatgag cggctgctat atccacttct tcagagagcc caccgatctg	1260
aagcagtca agcaggacgc caagtacagc cacggaatcg acgtggccga tctgttccca	1320
gctcagccag gactgacatc cgccgtgatt gaagccctgc ctagaaacat ggtgctgacc	1380
gtcagggca gcgacgacat caagagactg ctggacagcc agggcagaag agatatcaag	1440
ctgatcgata tcgcccagc caagccgac tctcggagat tcgaaaacgc cgtgtgggac	1500
cagtcaagg acctgtgtca catgcacaca ggcgtggtgg tggaaaagaa gaagcgggga	1560
ggcaaagagg aaatcaccoc tcaactgcgc ctgatggact gcattatgta tgacccgcc	1620
gtgtctggcg gcctgaatat ccctgttctg agagccctgc tgccccgca catggtgttt	1680
agaacaagca gccccaaggt ggtgctctga	1710

[0039]

- <210> 30
- <211> 569
- <212> PRT
- <213> 人工序列

<220>
 <223> L-NP-1 = L-NP-CovAnc-2_SL

<400> 30

Met Ser Ala Ser Lys Glu Ile Lys Ser Phe Leu Trp Thr Gln Ser Leu
 1 5 10 15

Arg Arg Glu Leu Ser Gly Tyr Cys Ser Asn Ile Lys Leu Gln Val Val
 20 25 30

Lys Asp Ala Gln Ala Leu Leu His Gly Leu Asp Phe Ser Glu Val Ser
 35 40 45

Asn Val Gln Arg Leu Met Arg Lys Glu Arg Arg Asp Asp Asn Asp Leu
 50 55 60

Lys Arg Leu Arg Asp Leu Asn Gln Ala Val Asn Asn Leu Val Glu Leu
 65 70 75 80

Lys Ser Thr Gln Gln Lys Ser Ile Leu Arg Val Gly Thr Leu Thr Ser
 85 90 95

Asp Asp Leu Leu Ile Leu Ala Ala Asp Leu Glu Lys Leu Lys Ser Lys
 100 105 110

Val Thr Arg Thr Glu Arg Pro Leu Ser Ala Gly Val Tyr Met Gly Asn
 115 120 125

Leu Ser Ser Gln Gln Leu Asp Gln Arg Arg Ala Leu Leu Asn Met Ile
 130 135 140

Gly Met Ser Gly Gly Asn Gln Gly Ala Arg Ala Gly Arg Asp Gly Val
 145 150 155 160

Val Arg Val Trp Asp Val Lys Asn Ala Glu Leu Leu Asn Asn Gln Phe
 165 170 175

Gly Thr Met Pro Ser Leu Thr Leu Ala Cys Leu Thr Lys Gln Gly Gln
 180 185 190

[0040] Val Asp Leu Asn Asp Ala Val Gln Ala Leu Thr Asp Leu Gly Leu Ile
 195 200 205

Tyr Thr Ala Lys Tyr Pro Asn Thr Ser Asp Leu Asp Arg Leu Thr Gln
 210 215 220

Ser His Pro Ile Leu Asn Met Ile Asp Thr Lys Lys Ser Ser Leu Asn
 225 230 235 240

Ile Ser Gly Tyr Asn Phe Ser Leu Gly Ala Ala Val Lys Ala Gly Ala
 245 250 255

Cys Met Leu Asp Gly Gly Asn Met Leu Glu Thr Ile Lys Val Ser Pro
 260 265 270

Gln Thr Met Asp Gly Ile Leu Lys Ser Ile Leu Lys Val Lys Lys Ala
 275 280 285

Leu Gly Met Phe Ile Ser Asp Thr Pro Gly Glu Arg Asn Pro Tyr Glu
 290 295 300

Asn Ile Leu Tyr Lys Ile Cys Leu Ser Gly Asp Gly Trp Pro Tyr Ile
 305 310 315 320

Ala Ser Arg Thr Ser Ile Thr Gly Arg Ala Trp Glu Asn Thr Val Val
 325 330 335

Asp Leu Glu Ser Asp Gly Lys Pro Gln Lys Ala Gly Ser Asn Asn Ser
 340 345 350

Asn Lys Ser Leu Gln Ser Ala Gly Phe Thr Ala Gly Leu Thr Tyr Ser
 355 360 365

Gln Leu Met Thr Leu Lys Asp Ala Met Leu Gln Leu Asp Pro Asn Ala
 370 375 380

Lys Thr Trp Met Asp Ile Glu Gly Arg Pro Glu Asp Pro Val Glu Ile
 385 390 395 400

Ala Leu Tyr Gln Pro Ser Ser Gly Cys Tyr Ile His Phe Phe Arg Glu
 405 410 415

Pro Thr Asp Leu Lys Gln Phe Lys Gln Asp Ala Lys Tyr Ser His Gly
 420 425 430

Ile Asp Val Thr Asp Leu Phe Ala Ala Gln Pro Gly Leu Thr Ser Ala
 435 440 445

Val Ile Asp Ala Leu Pro Arg Asn Met Val Ile Thr Cys Gln Gly Ser
 450 455 460

Asp Asp Ile Arg Lys Leu Leu Glu Ser Gln Gly Arg Lys Asp Ile Lys
 465 470 475 480

Leu Ile Asp Ile Ala Leu Ser Lys Thr Asp Ser Arg Lys Tyr Glu Asn
 485 490 495

[0041] Ala Val Trp Asp Gln Tyr Lys Asp Leu Cys His Met His Thr Gly Val
 500 505 510

Vai Val Glu Lys Lys Lys Arg Gly Gly Lys Glu Glu Ile Thr Pro His
 515 520 525

Cys Ala Leu Met Asp Cys Ile Met Phe Asp Ala Ala Val Ser Gly Gly
 530 535 540

Leu Asn Thr Ser Val Leu Arg Ala Val Leu Pro Arg Asp Met Val Phe
 545 550 555 560

Arg Thr Ser Thr Pro Arg Val Val Leu
 565

<210> 31

<211> 1710

<212> DNA

<213> 人工序列

<220>

<223> L-NP-1 = L-NP-CovAnc-2_SL

<400> 31

atgagcgcca gcaaaagat caagagcttc ctgtggaccc agagcctgcg gagagagctg 60

tctggctact gctccaacat caagctccag gtggctcaagg acgccaggc tctgctgcat 120

ggcctggatt tcagcgaggt gtccaactg cagcggctga tgcggaaaga gagaaggac 180

gacaacgacc tgaagcggct gagggatctg aaccaggccg tgaacaacct ggtggaactg 240

aagtctaccc agcagaaatc catcctgaga gtgggcaccc tgaccagcga cgatctgctg 300

	attctggccg cgcacctgga aaagctgaag tccaaagtga cccggaccga gaggccactg	360
	tctgctgggt tctacatggg caacctgagc agccagcagc tggatcagag aagggccctg	420
	ctgaacatga tcggcatgag cggcggaaat cagggcgcta gagctggcag agatggcgtc	480
	gtcagagtgt gggacgtgaa gaatgccgag ctgctcaaca accagttcgg caccatgcct	540
	agcctgacac tggcctgcct gacaaagcag ggccaagtgg acctgaacga tgctgtgcag	600
	gccctgactg atctgggcct gatctacacc gccaaagtac ccaacaccag cgacctggac	660
	agactgacce agtctcacc cctcctgaat atgatcgaca ccaagaagtc cagcctgaac	720
	atcagcggct acaacttctc tctgggcgct gccgtgaaag cgggcgcttg tatgcttgac	780
	ggcggcaaca tgctggaaac catcaagggt tccccacaga ccatggacgg catcctgaaa	840
	agtatcctga aagtgaagaa agccctgggc atgttcatca gcgacacccc tggcgagaga	900
	aaccctacg agaacatcct gtacaagatt tgccctgagcg gcgacggctg gccctatatc	960
[0042]	gccagcagaa ccagcattac cggcagagct tgggagaaca cctgtgtgga tctggaaagc	1020
	gacggcaagc ctcagaaggc cggcagcaac aactccaaca agagcctcca gtccgccggc	1080
	ttcacagccg gcctgacata tagccagctg atgaccctga aggacgcat gctgcaactg	1140
	gaccccaatg ccaagacctg gatggacatc gagggcagac ctgaggacct tgtggaaatc	1200
	gccctgtacc agcctagctc cggctgctat atccattctc tcagagagcc caccgatctg	1260
	aagcagttca agcaggacgc caagtacgc cacggcatcg acgtgaccga tctgtttgct	1320
	gctcagcccg gactgacctc cggcgtgatt gatgccctgc ctcggaacat ggtcatcacc	1380
	tgtcagggca gcgacgacat ccggaagctg ctggaatctc agggcagaaa ggatatcaag	1440
	ctgatcgata tcgccctgag caagaccgac agccggaagt acgaaaacgc cgtgtgggac	1500
	cagtacaagg acctgtgcca catgcacaca ggcgtgggtg tgaaaagaa gaagcgcgga	1560
	ggcaaagagg aaatcacccc tcaactgcct ctgatggact gcatcatggt tgacgccgcc	1620
	gtgtctggcg gcctgaatac ctctgttctg agagccgtgc tgcccagaga catggtgttc	1680
	agaacaagca cccctagagt ggtgctctga	1710

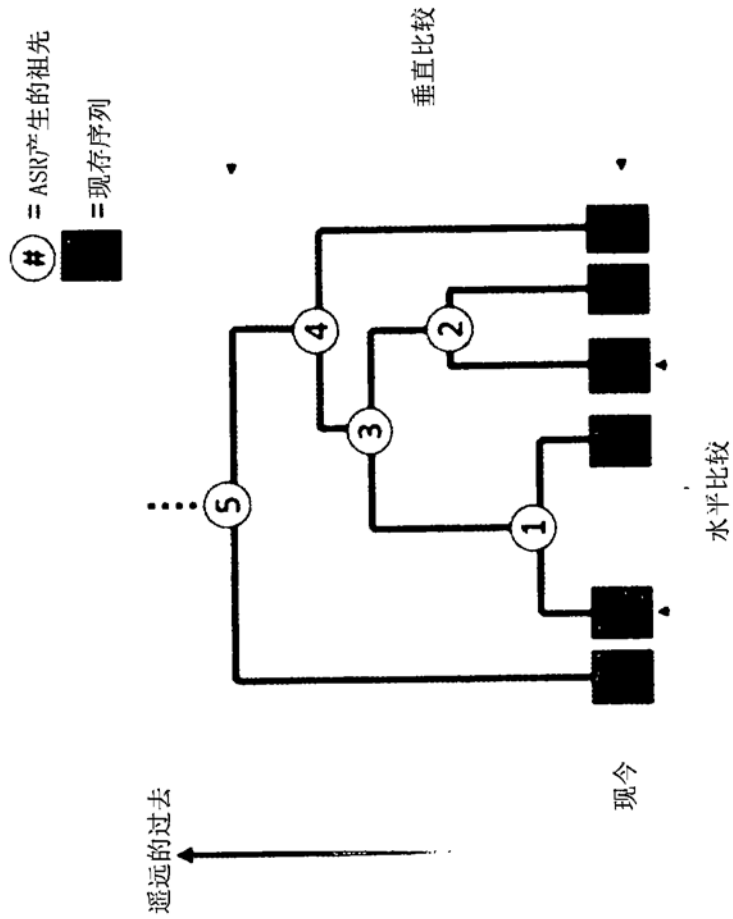


图1

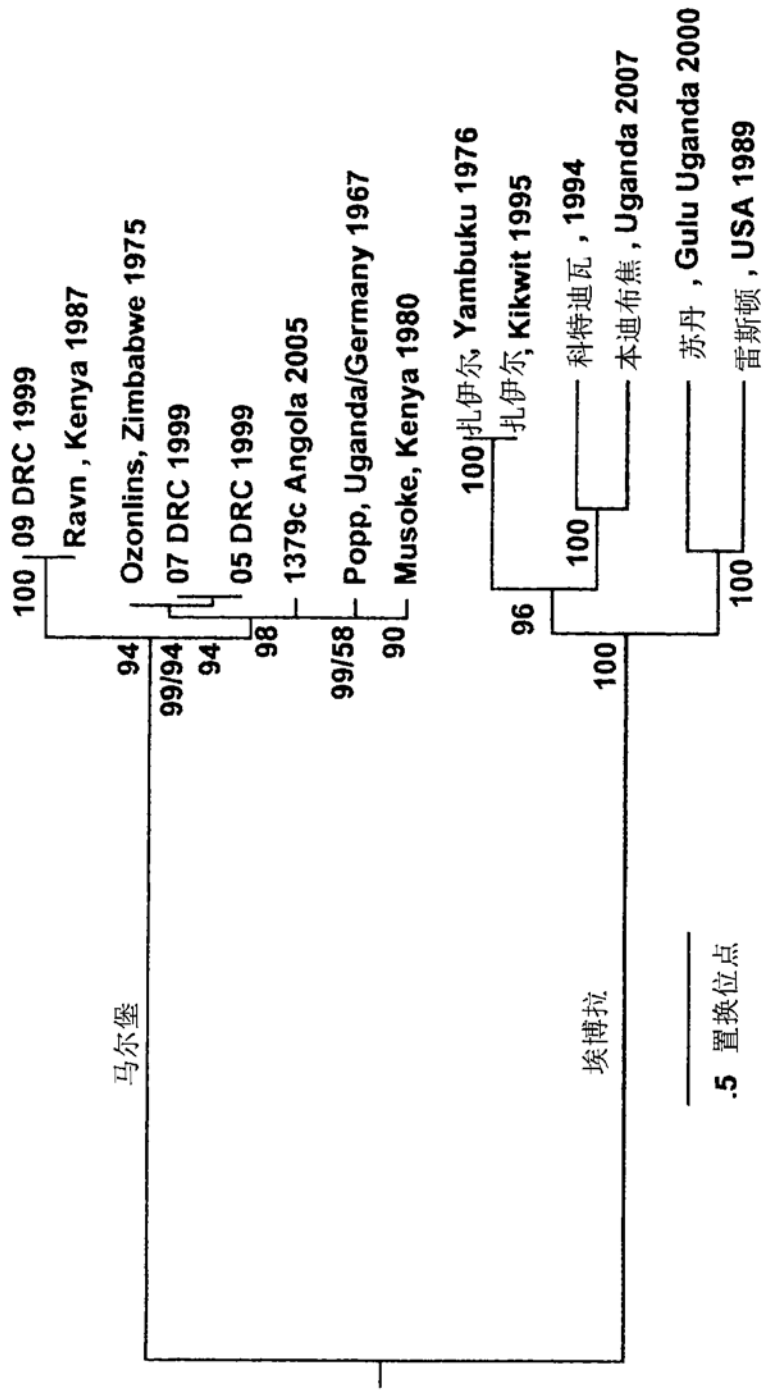


图2

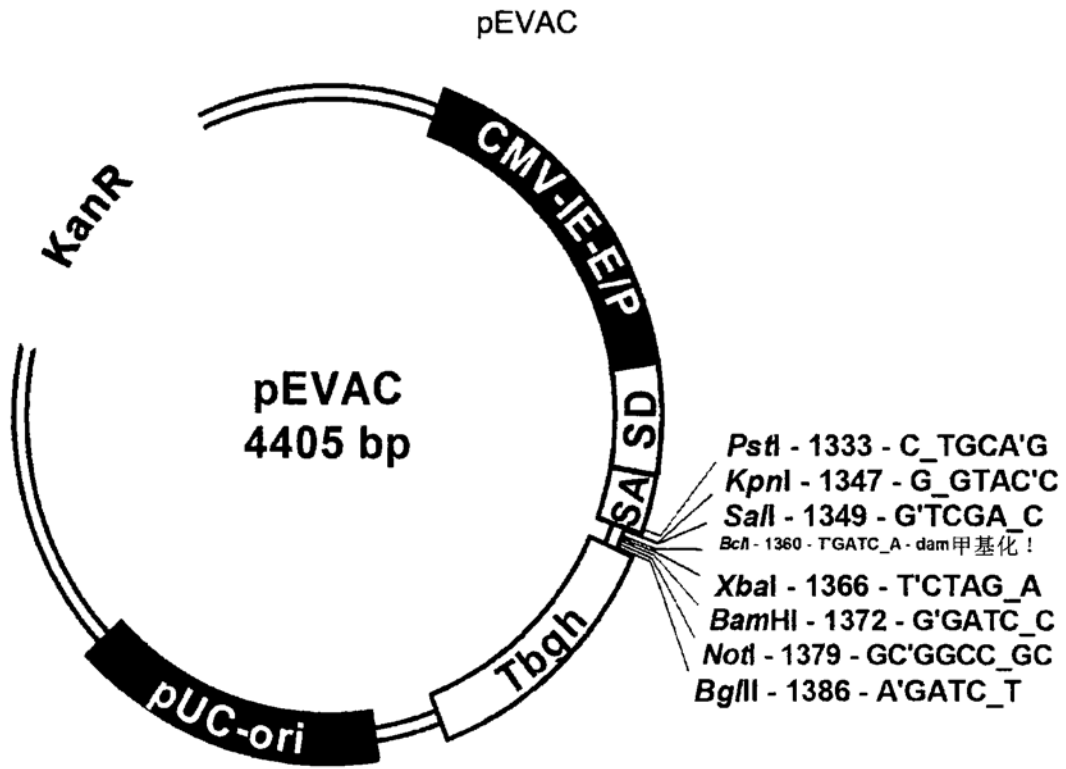


图3

结果

埃博拉效力

● 在动物模型中实现的保护

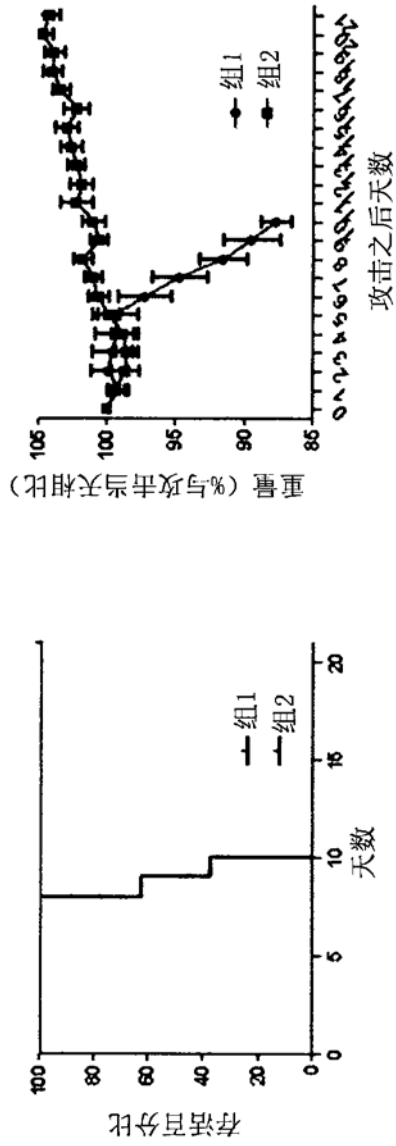


图4

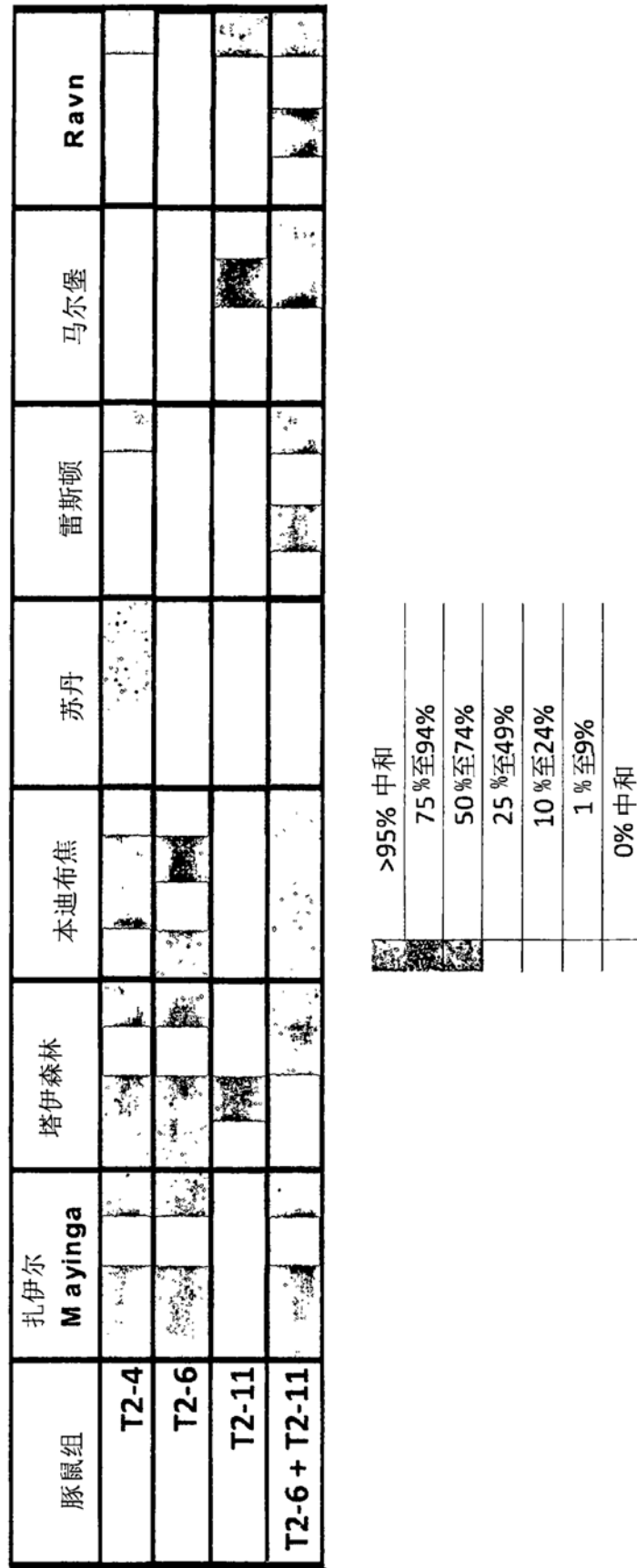


图5

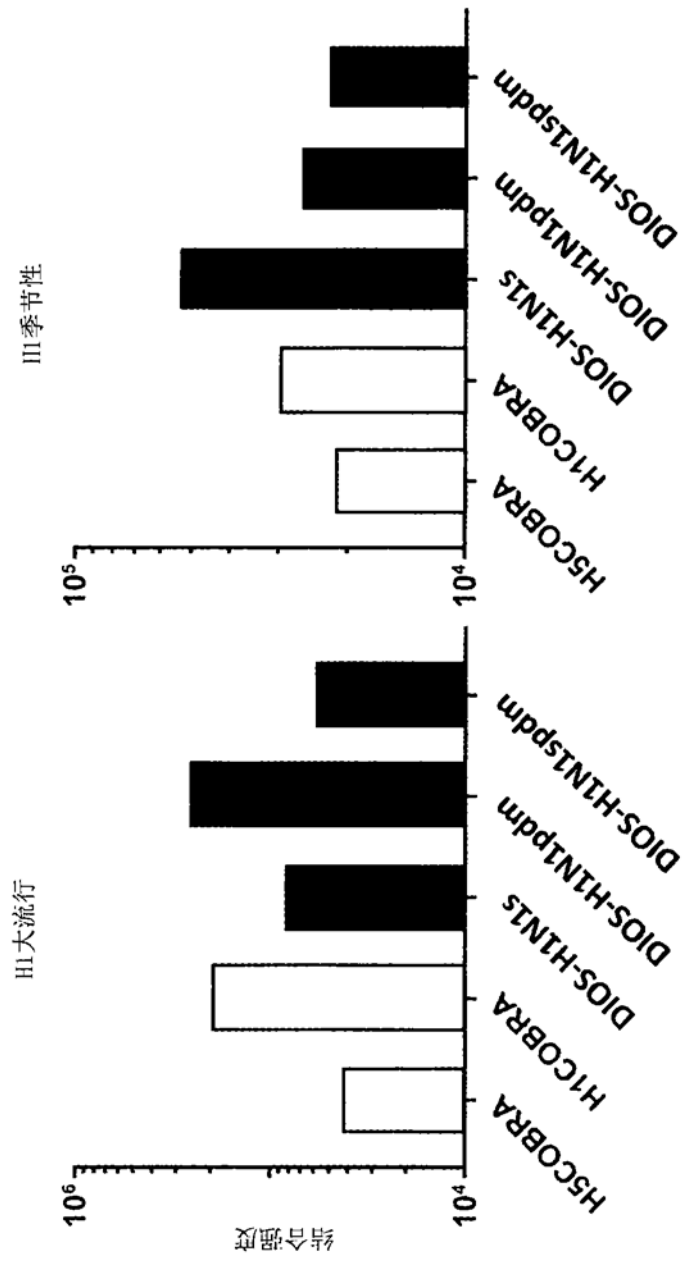


图6

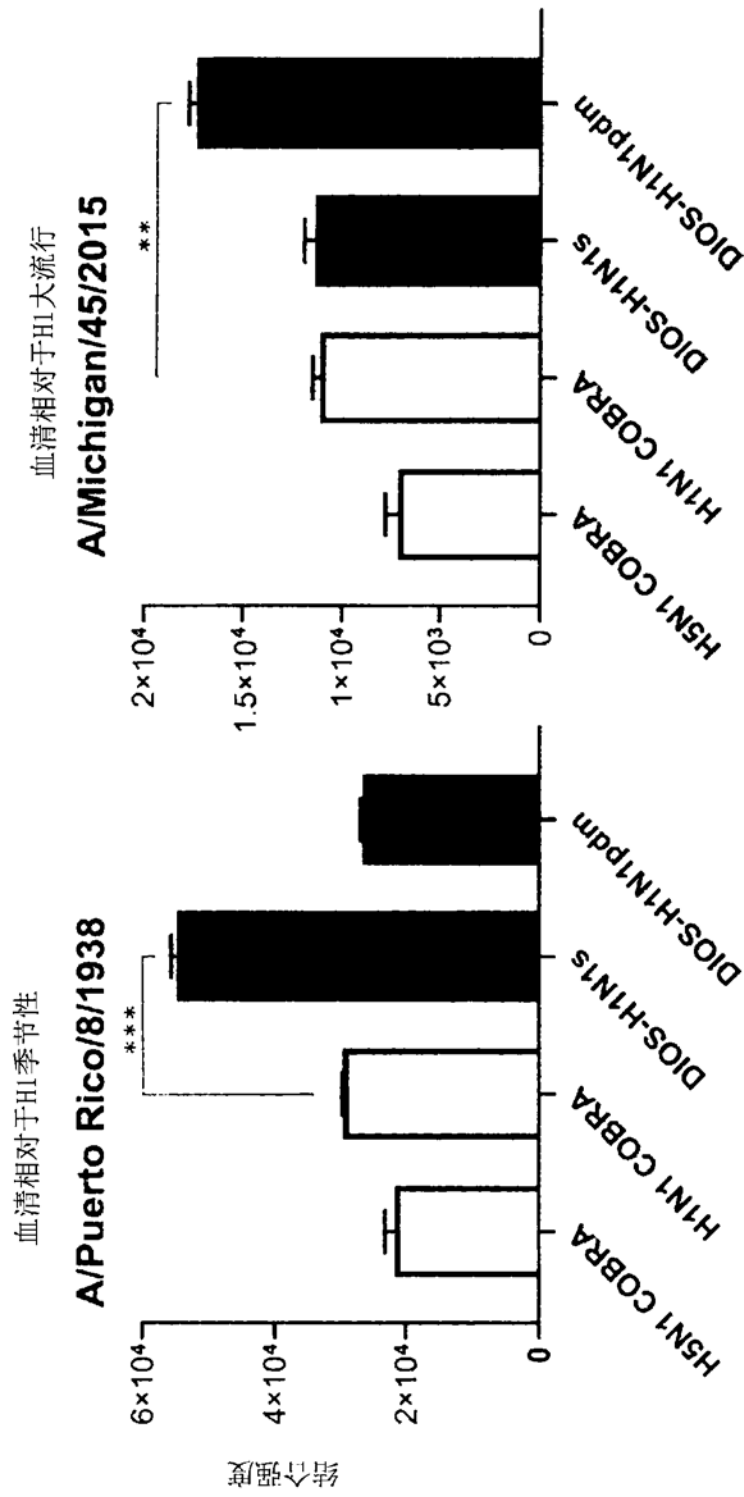


图7

血清相对于H7N9 (A/Shanghai/2/2013)

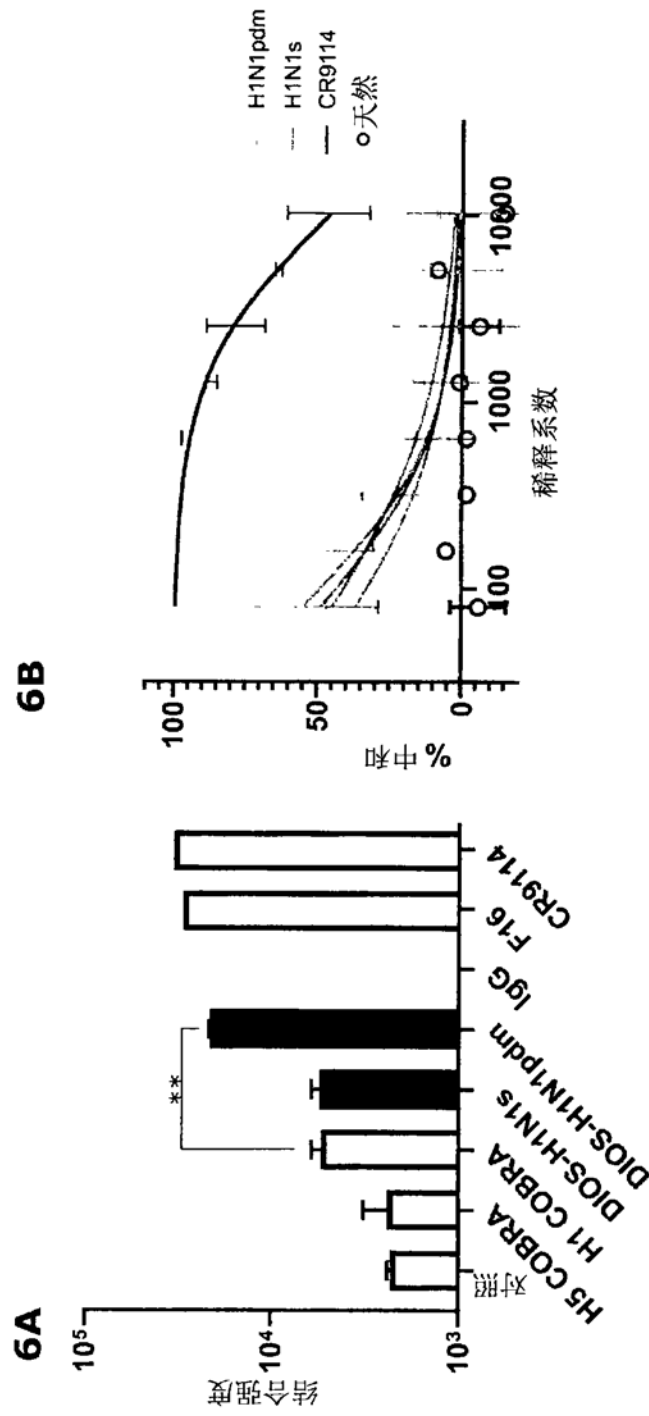


图8