



(21) 申请号 201780048641.6

G01N 33/50 (2006.01)

(22) 申请日 2017.06.13

(56) 对比文件

(65) 同一申请的已公布的文献号
申请公布号 CN 109641026 A

EP 3009511 A2, 2016.04.20

CN 104620107 A, 2015.05.13

US 2014273226 A1, 2014.09.18

(43) 申请公布日 2019.04.16

US 2013123129 A1, 2013.05.16

CN 103088128 A, 2013.05.08

(30) 优先权数据

62/351,172 2016.06.16 US

62/378,156 2016.08.22 US

15/467,922 2017.03.23 US

Omar O. Abudayyeh, 等. C2c2 is a single-component programmable RNA-guided RNA-targeting CRISPR effector.《Science》.2016, 第353卷 (第6299期), 1-9.

(85) PCT国际申请进入国家阶段日
2019.02.02

Omar O. Abudayyeh, 等. C2c2 is a single-component programmable RNA-guided RNA-targeting CRISPR effector.《Science》.2016, 第353卷 (第6299期), 1-9.

(86) PCT国际申请的申请数据
PCT/US2017/037308 2017.06.13

(87) PCT国际申请的公布数据
W02017/218573 EN 2017.12.21

无.NCBI Reference Sequence: WP_012985477.1.《GenBank》.2013,

(73) 专利权人 加利福尼亚大学董事会
地址 美国加利福尼亚

Sergey Shmakov, 等. Discovery and functional characterization of diverse Class 2 CRISPR-Cas systems.《Molecular cell》.2015, 第60卷 (第3期), 385-397.

(72) 发明人 J·A·多德纳 M·R·奥康奈尔
A·伊斯特-塞莱茨基 S·C·奈特
J·H·多德纳凯特

无.NCBI Reference Sequence: WP_015770004.1.《GenBank》.2013,

(74) 专利代理机构 中国贸促会专利商标事务所
有限公司 11038

无.NCBI Reference Sequence: WP_023911507.1.《GenBank》.2013,

专利代理师 韩威威

无.NCBI Reference Sequence: WP_034560163.1.《GenBank》.2014,

(51) Int.Cl.

审查员 张维捷

A61K 38/00 (2006.01)

C07K 14/47 (2006.01)

C12N 9/90 (2006.01)

权利要求书9页 说明书108页
序列表191页 附图81页

(54) 发明名称

用于检测靶RNA的方法和组合物

(57) 摘要

本公开提供了用于检测单链靶RNA的方法。本公开提供了将前体C2c2指导RNA阵列切割成两个或更多个C2c2指导RNA的方法。本公开提供了用于检测样品中的靶RNA的试剂盒。



CN 109641026 B

1. 一种检测包含多种RNA的样品中的单链靶RNA的方法,所述方法包括:
 - a) 使所述样品与以下接触:
 - (i) 经标记的检测RNA;
 - (ii) C2c2蛋白,其切割所述经标记的检测RNA;和
 - (iii) 指导RNA,其与单链靶RNA杂交,其中所述指导RNA结合所述C2c2蛋白;以及
 - b) 测量由C2c2蛋白介导的所述经标记的检测RNA切割产生的可检测信号。
2. 一种检测包含多种RNA的样品中的单链靶RNA的方法,所述方法包括:
 - (a) 使所述样品与以下接触:
 - (i) 前体C2c2指导RNA阵列,其包含两个或更多个C2c2指导RNA,每个所述C2c2指导RNA具有不同的指导序列;
 - (ii) 经标记的检测RNA;和
 - (iii) C2c2蛋白,其将所述前体C2c2指导RNA阵列切割成单独的C2c2指导RNA,并且还切割所述样品的RNA和所述经标记的检测RNA;以及
 - (b) 测量由C2c2蛋白介导的所述经标记的检测RNA切割产生的可检测信号。
3. 根据权利要求1或权利要求2所述的方法,其中所述C2c2蛋白在所述接触1小时内切割所述样品中存在的所述RNA的至少50%。
4. 根据权利要求3所述的方法,其中所述C2c2蛋白在所述接触40分钟内切割所述样品中存在的所述RNA的至少50%。
5. 根据权利要求4所述的方法,其中所述C2c2蛋白在所述接触5分钟内切割所述样品中存在的所述RNA的至少50%。
6. 根据权利要求5所述的方法,其中所述C2c2蛋白在所述接触1分钟内切割所述样品中存在的所述RNA的至少50%。
7. 根据权利要求1所述的方法,其中所述C2c2蛋白在所述接触1分钟内切割所述样品中存在的所述RNA的50%至多于90%。
8. 根据权利要求1或2所述的方法,其中能够检测到所述单链靶RNA的最小浓度在500fM至1nM的范围内。
9. 根据权利要求1或2所述的方法,其中能够在低至800fM的浓度下检测到所述单链靶RNA。
10. 根据权利要求1或2所述的方法,其中所述C2c2蛋白不是包含与SEQ ID NO:3所示的氨基酸序列具有至少80%、至少85%、至少90%、至少95%、至少98%、至少99%或100%的氨基酸序列同一性的氨基酸序列的沙氏纤毛菌 (*Leptotrichia shahii*) C2c2蛋白。
11. 根据权利要求10所述的方法,其中所述C2c2蛋白比包含与SEQ ID NO:3所示的氨基酸序列至少80%、至少85%、至少90%、至少95%、至少98%、至少99%或100%的氨基酸序列同一性的沙氏纤毛菌C2c2蛋白至少1.2倍有效地切割非靶RNA。
12. 根据权利要求11所述的方法,其中所述C2c2蛋白比包含SEQ ID NO:3所示的氨基酸序列的沙氏纤毛菌C2c2蛋白至少1.5倍有效地切割非靶RNA。
13. 根据权利要求1或2所述的方法,其中所述C2c2蛋白由SEQ ID NO:1、SEQ ID NO:2或SEQ ID NO:4至SEQ ID NO:6中任一者所示的氨基酸序列组成。
14. 根据权利要求1或2所述的方法,其中所述C2c2蛋白由SEQ ID NO:2所示的口腔纤毛

菌 (*Leptotrichia buccalis*) C2c2氨基酸序列组成。

15. 根据权利要求1或2所述的方法, 其中所述C2c2蛋白由SEQ ID NO:1所示的斯氏李斯特菌 (*Listeria seeligeri*) C2c2氨基酸序列组成。

16. 根据权利要求1或2所述的方法, 其中所述C2c2蛋白包含SEQ ID NO:1至SEQ ID NO:2和SEQ ID NO:4至SEQ ID NO:6中任一者所示的氨基酸序列。

17. 根据权利要求1所述的方法, 其中所述C2c2蛋白由SEQ ID NO:2所示的C2c2氨基酸序列以及对R472、H477、R1048和H1053中的一者或多者的取代组成。

18. 根据权利要求1所述的方法, 其中所述C2c2蛋白由SEQ ID NO:2所示的C2c2氨基酸序列以及对氨基酸R472和H477的取代组成。

19. 根据权利要求1所述的方法, 其中所述C2c2蛋白由SEQ ID NO:2所示的C2c2氨基酸序列以及对氨基酸R1048和H1053的取代组成。

20. 根据权利要求1所述的方法, 其中所述C2c2蛋白由SEQ ID NO:2所示的C2c2氨基酸序列以及对氨基酸R472、H477、R1048和H1053的取代组成。

21. 根据权利要求1或2所述的方法, 其中在所述测量之前使所述样品接触2小时或更短时间。

22. 根据权利要求21所述的方法, 其中在所述测量之前使所述样品接触60分钟或更短时间。

23. 根据权利要求22所述的方法, 其中在所述测量之前使所述样品接触30分钟或更短时间。

24. 根据权利要求23所述的方法, 其中在所述测量之前使所述样品接触10分钟或更短时间。

25. 根据权利要求24所述的方法, 其中在所述测量之前使所述样品接触1分钟或更短时间。

26. 根据权利要求1或2所述的方法, 其包括确定在所述样品中存在的靶RNA的量。

27. 根据权利要求26所述的方法, 其中所述确定包括:

测量所述可检测信号以产生测试测量值;

测量由参考样品产生的可检测信号以产生参考测量值; 以及

将所述测试测量值与所述参考测量值进行比较以确定所述样品中存在的靶RNA的量。

28. 根据权利要求26所述的方法, 其包括:

测量所述可检测信号以产生测试测量值,

测量由两个或更多个参考样品中的每一个产生的可检测信号, 其中所述两个或更多个参考样品各自包括不同量的阳性对照RNA, 以产生两个或更多个参考测量值, 以及

将所述测试测量值与所述两个或更多个参考测量值进行比较以确定所述样品中存在的靶RNA的量。

29. 根据权利要求1或2所述的方法, 其中所述样品包含5至 10^7 个序列彼此不同的RNA。

30. 根据权利要求1或2所述的方法, 其中所述样品包含10至 10^6 个序列彼此不同的RNA。

31. 根据权利要求1或2所述的方法, 其中所述样品包含来自细胞裂解物的RNA。

32. 根据权利要求1或2所述的方法, 其中测量可检测信号包括以下中的一者或多者: 基于金纳米颗粒的检测、荧光偏振、胶体相变/分散、电化学检测和基于半导体的感测。

33. 根据权利要求1或2所述的方法, 其中所述经标记的检测RNA包含荧光发射染料对, 以及所述可检测信号由所述荧光发射染料对产生。

34. 根据权利要求1或2所述的方法, 其中所述经标记的检测RNA在被切割之前产生一定量的可检测信号, 并且当所述经标记的检测RNA被切割时可检测信号的量减少。

35. 根据权利要求1或2所述的方法, 其中所述经标记的检测RNA在被切割之前产生第一可检测信号, 并且当所述经标记的检测RNA被切割时产生第二可检测信号。

36. 根据权利要求35所述的方法, 其中所述经标记的检测RNA包含荧光发射染料对。

37. 根据权利要求36所述的方法, 其中所述经标记的检测RNA包含荧光共振能量转移对。

38. 根据权利要求1或2所述的方法, 其中在所述经标记的检测RNA被切割时产生可检测信号。

39. 根据权利要求1或2所述的方法, 其中当所述经标记的检测RNA被切割时, 可检测信号的量增加。

40. 根据权利要求38所述的方法, 其中所述经标记的检测RNA包含猝灭剂/荧光剂对。

41. 根据权利要求39所述的方法, 其中所述经标记的检测RNA包含猝灭剂/荧光剂对。

42. 根据权利要求1或2所述的方法, 其中所述经标记的检测RNA包含经修饰的核碱基、经修饰的糖部分和/或经修饰的核酸键。

43. 根据权利要求1或2所述的方法, 其中所述接触在无细胞样品中进行。

44. 根据权利要求1或2所述的方法, 其中所述接触在体外、离体或体外细胞中进行。

45. 一种将前体C2c2指导RNA阵列切割成两个或更多个C2c2指导RNA的方法, 所述方法包括:

使包含两个或更多个C2c2指导RNA的前体C2c2指导RNA阵列与C2c2蛋白接触, 每个所述C2c2指导RNA具有不同的指导序列, 其中所述C2c2蛋白将所述前体C2c2指导RNA阵列切割成单独的C2c2指导RNA, 其中所述指导RNA和/或所述前体C2c2指导RNA阵列中的至少一者是可检测地标记的。

46. 根据权利要求45所述的方法, 其中所述C2c2蛋白缺乏催化活性的HEPN1结构域和/或缺乏催化活性的HEPN2结构域。

47. 根据权利要求45所述的方法, 其中所述前体C2c2指导RNA阵列包含两个或更多个C2c2指导RNA, 所述两个或更多个C2c2指导RNA靶向相同靶RNA分子内的不同靶序列。

48. 根据权利要求45-47中任一项所述的方法, 其中所述前体C2c2指导RNA阵列包含两个或更多个C2c2指导RNA, 所述两个或更多个C2c2指导RNA靶向不同靶RNA分子。

49. 根据权利要求45-47中任一项所述的方法, 其中所述接触不在细胞内发生。

50. 一种用于检测包含多种RNA的样品中的靶RNA的试剂盒, 所述试剂盒包含:

(a) 前体C2c2指导RNA阵列, 和/或编码所述前体C2c2指导RNA阵列的核酸, 其中所述前体C2c2指导RNA阵列包含两个或更多个C2c2指导RNA, 每个所述C2c2指导RNA具有不同的指导序列和/或所选择的指导序列的插入位点;

(b) C2c2蛋白, 其切割经标记的检测RNA; 以及

(c) 所述经标记的检测RNA, 其包含荧光发射染料对, 其中所述C2c2蛋白对所述经标记的检测RNA的切割产生可检测信号。

51. 根据权利要求50所述的试剂盒,其中所述C2c2蛋白缺乏催化活性的HEPN1结构域和/或缺乏催化活性的HEPN2结构域。

52. 根据权利要求50或权利要求51所述的试剂盒,其中所述前体C2c2指导RNA阵列包含两个或更多个C2c2指导RNA,所述两个或更多个C2c2指导RNA靶向相同靶RNA分子内的不同靶序列。

53. 根据权利要求50或权利要求51所述的试剂盒,其中所述前体C2c2指导RNA阵列包含两个或更多个C2c2指导RNA,所述两个或更多个C2c2指导RNA靶向不同靶RNA分子。

54. 根据权利要求50或权利要求51所述的试剂盒,其中所述指导RNA和/或所述前体C2c2指导RNA阵列中的至少一者是可检测地标记的。

55. 根据权利要求50或权利要求51所述的试剂盒,其中所述C2c2蛋白不是包含与SEQ ID NO:3所示的氨基酸序列具有至少80%、至少85%、至少90%、至少95%、至少98%、至少99%或100%的氨基酸序列同一性的氨基酸序列的沙氏纤毛菌C2c2蛋白。

56. 根据权利要求55所述的试剂盒,其中所述C2c2蛋白比包含与SEQ ID NO:3所示的氨基酸序列至少80%、至少85%、至少90%、至少95%、至少98%、至少99%或100%的氨基酸序列同一性的沙氏纤毛菌C2c2蛋白至少1.2倍有效地切割非靶RNA。

57. 根据权利要求56所述的试剂盒,其中所述C2c2蛋白比包含SEQ ID NO:3所示的氨基酸序列的沙氏纤毛菌C2c2蛋白至少1.5倍有效地切割非靶RNA。

58. 根据权利要求50或权利要求51所述的试剂盒,其中所述C2c2蛋白由SEQ ID NO:1、SEQ ID NO:2或SEQ ID NO:4至SEQ ID NO:6中任一者所示的氨基酸序列组成。

59. 根据权利要求50或权利要求51所述的试剂盒,其中所述C2c2蛋白由SEQ ID NO:2所示的口腔纤毛菌C2c2氨基酸序列组成。

60. 根据权利要求50或权利要求51所述的试剂盒,其中所述C2c2蛋白由SEQ ID NO:1所示的斯氏李斯特菌C2c2氨基酸序列组成。

61. 根据权利要求50或权利要求51所述的试剂盒,其中所述C2c2蛋白包含SEQ ID NO:1至SEQ ID NO:2和SEQ ID NO:4至SEQ ID NO:6中任一者所示的氨基酸序列。

62. 一种用于检测包含多种RNA的样品中的靶RNA的试剂盒,所述试剂盒包含:

(a) 经标记的检测RNA,其包含荧光发射染料对;以及

(b) C2c2蛋白,其切割所述经标记的检测RNA,其中所述C2c2蛋白对所述经标记的检测RNA的切割产生可检测信号。

63. 根据权利要求62所述的试剂盒,其包含阳性对照靶RNA。

64. 根据权利要求63所述的试剂盒,其中所述阳性对照靶RNA以不同的量在两个或更多个容器中的每一个中存在。

65. 根据权利要求62-64中任一项所述的试剂盒,其包含以下中的至少一者:

(c) C2c2指导RNA和/或编码所述C2c2指导RNA的核酸;

(d) 前体C2c2指导RNA和/或编码所述前体C2c2指导RNA的核酸;以及

(e) 前体C2c2指导RNA阵列,和/或编码所述前体C2c2指导RNA阵列的核酸,其中所述前体C2c2指导RNA阵列包含两个或更多个C2c2指导RNA,每个所述C2c2指导RNA具有不同的指导序列和/或所选择的指导序列的插入位点。

66. 根据权利要求62-64中任一项所述的试剂盒,其包含DNA,所述DNA包含编码具有或

不具有指导序列的C2c2指导RNA的核苷酸序列。

67. 根据权利要求66所述的试剂盒,其中所述DNA包含用于插入指导序列的插入序列。

68. 根据权利要求66所述的试剂盒,其中所述DNA是表达载体,并且所述C2c2指导RNA与启动子可操作地连接。

69. 根据权利要求68所述的试剂盒,其中所述启动子是T7启动子。

70. 根据权利要求62-64中任一项所述的试剂盒,其包含缺乏核酸酶活性的C2c2蛋白变体。

71. 根据权利要求62-64中任一项所述的试剂盒,其中所述经标记的检测RNA包含FRET对。

72. 根据权利要求62-64中任一项所述的试剂盒,其中所述经标记的检测RNA包含猝灭剂/荧光剂对。

73. 根据权利要求62-64中任一项所述的试剂盒,其中所述经标记的检测RNA包含产生第一可检测信号的FRET对和产生第二可检测信号的猝灭剂/荧光剂对。

74. 根据权利要求62-64中任一项所述的试剂盒,其中所述C2c2蛋白不是包含与SEQ ID NO:3所示的氨基酸序列具有至少80%、至少85%、至少90%、至少95%、至少98%、至少99%或100%的氨基酸序列同一性的氨基酸序列的沙氏纤毛菌C2c2蛋白。

75. 根据权利要求74所述的试剂盒,其中所述C2c2蛋白比包含与SEQ ID NO:3所示的氨基酸序列至少80%、至少85%、至少90%、至少95%、至少98%、至少99%或100%的氨基酸序列同一性的沙氏纤毛菌C2c2蛋白至少1.2倍有效地切割非靶RNA。

76. 根据权利要求75所述的试剂盒,其中所述C2c2蛋白比包含SEQ ID NO:3所示的氨基酸序列的沙氏纤毛菌C2c2蛋白至少1.5倍有效地切割非靶RNA。

77. 根据权利要求62-64中任一项所述的试剂盒,其中所述C2c2蛋白由SEQ ID NO:1、SEQ ID NO:2或SEQ ID NO:4至SEQ ID NO:6中任一者所示的氨基酸序列组成。

78. 根据权利要求62-64中任一项所述的试剂盒,其中所述C2c2蛋白由SEQ ID NO:2所示的口腔纤毛菌C2c2氨基酸序列组成。

79. 根据权利要求62-64中任一项所述的试剂盒,其中所述C2c2蛋白由SEQ ID NO:1所示的斯氏李斯特菌C2c2氨基酸序列组成。

80. 根据权利要求62-64中任一项所述的试剂盒,其中所述C2c2蛋白包含SEQ ID NO:1至SEQ ID NO:2和SEQ ID NO:4至SEQ ID NO:6中任一者所示的氨基酸序列。

81. 一种变体C2c2多肽,其包含:

a) 具有对以下的取代的SEQ ID NO:2所示的氨基酸序列:i) 氨基酸R472和H477;ii) 氨基酸R1048和H1053;或iii) 氨基酸R472、H477、R1048和H1053;

b) 具有对以下的取代的SEQ ID NO:1所示的氨基酸序列:i) 氨基酸R445和H450;ii) 氨基酸R1016和H1021;或iii) 氨基酸R445、H450、R1016和H1021;

c) 具有对以下的取代的SEQ ID NO:4所示的氨基酸序列:i) 氨基酸R464和H469;ii) 氨基酸R1052和H1057;或iii) 氨基酸R464、H469、R1052和H1057;

d) 具有对以下的取代的SEQ ID NO:5所示的氨基酸序列:i) 氨基酸R467和H472;ii) 氨基酸R1069和H1074;或iii) 氨基酸R467、H472、R1069和H1074;或者

e) 具有对以下的取代的SEQ ID NO:6所示的氨基酸序列:i) 氨基酸R472和H477;ii) 氨

氨酸R1044和H1049;或iii)氨基酸R472、H477、R1044和H1049。

82. 根据权利要求81所述的变体C2c2多肽,其中所述变体C2c2多肽具有减少的或不可检测的ssRNA切割,但保留了结合C2c2指导RNA和ssRNA的能力,并且保留了切割前体C2c2指导RNA的能力。

83. 一种核酸,其包含编码根据权利要求81或82所述的变体C2c2多肽的核苷酸序列。

84. 根据权利要求83所述的核酸,其中所述核苷酸序列与组成型启动子或可调控的启动子可操作地连接。

85. 一种重组表达载体,其包含根据权利要求83所述的核酸。

86. 一种宿主细胞,其用根据权利要求83所述的核酸或用根据权利要求85所述的重组表达载体遗传修饰。

87. 根据权利要求86所述的宿主细胞,其中所述宿主细胞是真核细胞。

88. 根据权利要求86所述的宿主细胞,其中所述宿主细胞是原核细胞。

89. 根据权利要求86-88中任一项所述的宿主细胞,其中所述宿主细胞是体外的、离体的或体内的。

90. 一种检测包含多种RNA的样品中的至少两种不同单链靶RNA的方法,所述方法包括:

a) 使所述样品与以下接触:

(i) 第一C2c2蛋白,其切割包括至少一个A的单链RNA;

(ii) 第二C2c2蛋白,其切割包括至少一个U的单链RNA;

(iii) 第一C2c2指导RNA,其包含与第一单链靶RNA杂交的第一核苷酸序列和与所述第一C2c2蛋白结合的第二核苷酸序列;

(iv) 第二C2c2指导RNA,其包含与第二单链靶RNA杂交的第一核苷酸序列和与所述第二C2c2蛋白结合的第二核苷酸序列;

(v) 第一经标记的检测RNA,其包含第一荧光发射染料对,其中所述第一C2c2蛋白切割所述第一经标记的检测RNA,并且第一可检测信号由所述第一荧光发射染料对产生;和

(vi) 第二经标记的检测RNA,其包含第二荧光发射染料对,其中所述第一C2c2蛋白切割所述第二经标记的检测RNA,并且第二可检测信号由所述第二荧光发射染料对产生;

其中所述第一C2c2蛋白不被所述第二C2c2指导RNA激活,并且其中所述第一C2c2蛋白切割包括至少一个A的单链RNA,并且

其中所述第二C2c2蛋白不被所述第一C2c2指导RNA激活,并且其中所述第二C2c2蛋白切割包括至少一个U的单链RNA;以及

b) 测量由所述第一C2c2蛋白和所述第二C2c2蛋白介导的RNA切割产生的可检测信号,其中所述第一可检测信号在所述第一C2c2蛋白被激活时产生,并且所述第二可检测信号在所述第二C2c2蛋白被激活时产生,其中检测到所述第一可检测信号表明所述第一靶单链RNA存在于所述样品中,并且其中检测到所述第二可检测信号表明所述第二靶单链RNA存在于所述样品中。

91. 根据权利要求90所述的方法,其中:

a) 所述第一C2c2蛋白由SEQ ID NO:301所示的Lba Cas13a氨基酸序列组成;并且所述第二C2c2蛋白由SEQ ID NO:6所示的Hhe Cas13a氨基酸序列组成;

b) 所述第一C2c2蛋白由SEQ ID NO:301所示的Lba Cas13a氨基酸序列组成;并且所述

第二C2c2蛋白由SEQ ID NO:4所示的Rca Cas13a氨基酸序列组成;

c) 所述第一C2c2蛋白由SEQ ID NO:301所示的Lba Cas13a氨基酸序列组成;并且所述第二C2c2蛋白由SEQ ID NO:299所示的Ppr Cas13a氨基酸序列组成;

d) 所述第一C2c2蛋白由SEQ ID NO:301所示的Lba Cas13a氨基酸序列组成;并且所述第二C2c2蛋白由SEQ ID NO:303所示的Lne Cas13a氨基酸序列组成;

e) 所述第一C2c2蛋白由SEQ ID NO:301所示的Lba Cas13a氨基酸序列组成;并且所述第二C2c2蛋白由SEQ ID NO:2所示的Lbu Cas13a氨基酸序列组成;

f) 所述第一C2c2蛋白由SEQ ID NO:301所示的Lba Cas13a氨基酸序列组成;并且所述第二C2c2蛋白由SEQ ID NO:300所示的Lwa Cas13a氨基酸序列组成;

g) 所述第一C2c2蛋白由SEQ ID NO:301所示的Lba Cas13a氨基酸序列组成;并且所述第二C2c2蛋白由SEQ ID NO:3所示的Lsh Cas13a氨基酸序列组成;

h) 所述第一C2c2蛋白由SEQ ID NO:304所示的Ere Cas13a氨基酸序列组成;并且所述第二C2c2蛋白由SEQ ID NO:6所示的Hhe Cas13a氨基酸序列组成;

i) 所述第一C2c2蛋白由SEQ ID NO:304所示的Ere Cas13a氨基酸序列组成;并且所述第二C2c2蛋白由SEQ ID NO:4所示的Rca Cas13a氨基酸序列组成;

j) 所述第一C2c2蛋白由SEQ ID NO:304所示的Ere Cas13a氨基酸序列组成;并且所述第二C2c2蛋白由SEQ ID NO:299所示的Ppr Cas13a氨基酸序列组成;

k) 所述第一C2c2蛋白由SEQ ID NO:304所示的Ere Cas13a氨基酸序列组成;并且所述第二C2c2蛋白由SEQ ID NO:303所示的Lne Cas13a氨基酸序列组成;

l) 所述第一C2c2蛋白由SEQ ID NO:304所示的Ere Cas13a氨基酸序列组成;并且所述第二C2c2蛋白由SEQ ID NO:2所示的Lbu Cas13a氨基酸序列组成;

m) 所述第一C2c2蛋白由SEQ ID NO:304所示的Ere Cas13a氨基酸序列组成;并且所述第二C2c2蛋白由SEQ ID NO:300所示的Lwa Cas13a氨基酸序列组成;

n) 所述第一C2c2蛋白由SEQ ID NO:304所示的Ere Cas13a氨基酸序列组成;并且所述第二C2c2蛋白由SEQ ID NO:3所示的Lsh Cas13a氨基酸序列组成;

o) 所述第一C2c2蛋白由SEQ ID NO:304所示的Ere Cas13a氨基酸序列组成;并且所述第二C2c2蛋白由SEQ ID NO:1所示的Lse Cas13a氨基酸序列组成;

p) 所述第一C2c2蛋白由SEQ ID NO:302所示的Cam Cas13a氨基酸序列组成;并且所述第二C2c2蛋白由SEQ ID NO:6所示的Hhe Cas13a氨基酸序列组成;

q) 所述第一C2c2蛋白由SEQ ID NO:302所示的Cam Cas13a氨基酸序列组成;并且所述第二C2c2蛋白由SEQ ID NO:4所示的Rca Cas13a氨基酸序列组成;

r) 所述第一C2c2蛋白由SEQ ID NO:302所示的Cam Cas13a氨基酸序列组成;并且所述第二C2c2蛋白由SEQ ID NO:299所示的Ppr Cas13a氨基酸序列组成;

s) 所述第一C2c2蛋白由SEQ ID NO:302所示的Cam Cas13a氨基酸序列组成;并且所述第二C2c2蛋白由SEQ ID NO:303所示的Lne Cas13a氨基酸序列组成;

t) 所述第一C2c2蛋白由SEQ ID NO:302所示的Cam Cas13a氨基酸序列组成;并且所述第二C2c2蛋白由SEQ ID NO:2所示的Lbu Cas13a氨基酸序列组成;

u) 所述第一C2c2蛋白由SEQ ID NO:302所示的Cam Cas13a氨基酸序列组成;并且所述第二C2c2蛋白由SEQ ID NO:300所示的Lwa Cas13a氨基酸序列组成;

v) 所述第一C2c2蛋白由SEQ ID NO:302所示的Cam Cas13a氨基酸序列组成;并且所述第二C2c2蛋白由SEQ ID NO:301所示的Lsh Cas13a氨基酸序列组成;或者

w) 所述第一C2c2蛋白由SEQ ID NO:302所示的Cam Cas13a氨基酸序列组成;并且所述第二C2c2蛋白由SEQ ID NO:1所示的Lse Cas13a氨基酸序列组成。

92. 根据权利要求90或91所述的方法, 其中:

i) 所述第一经标记的检测RNA包含至少一个A并且不包含U;和

ii) 所述第二经标记的检测RNA包含至少一个U并且不包含A。

93. 根据权利要求92所述的方法, 其中所述第一经标记的检测RNA包含2至15个连续的A的段和/或所述第二经标记的检测RNA包含2至15个连续的U的段。

94. 根据权利要求92所述的方法, 其中所述第一经标记的检测RNA包含4至15个连续的A的段和/或所述第二经标记的检测RNA包含4至15个连续的U的段。

95. 根据权利要求92所述的方法, 其中所述第一经标记的检测RNA包含至少3个连续的A的段和/或所述第二经标记的检测RNA包含至少3个连续的U的段。

96. 根据权利要求92所述的方法, 其中所述第一经标记的检测RNA包含至少4个连续的A的段和/或所述第二经标记的检测RNA包含至少4个连续的U的段。

97. 一种试剂盒, 其包含:

(a) 第一经标记的检测RNA, 其缺乏U并且包含至少一个A且包含第一荧光发射染料对;

(b) 第二经标记的检测RNA, 其缺乏A并且包含至少一个U且包含第二荧光发射染料对;

(c) 第一C2c2蛋白, 和/或编码所述第一C2c2蛋白的核酸, 其中所述第一C2c2蛋白能够切割所述第一经标记的检测RNA但不切割所述第二经标记的检测RNA; 以及

(d) 第二C2c2蛋白, 和/或编码所述第二C2c2蛋白的核酸, 其中所述第二C2c2蛋白能够切割所述第二经标记的检测RNA但不切割所述第一经标记的检测RNA。

98. 根据权利要求97所述的试剂盒, 其包含以下中的至少一者:

(e) 第一C2c2指导RNA和/或编码所述第一C2c2指导RNA的核酸, 其中所述第一C2c2指导RNA包含与所述第一C2c2蛋白结合的恒定区序列;

(f) 第二C2c2指导RNA和/或编码所述第二C2c2指导RNA的核酸, 其中所述第二C2c2指导RNA包含与所述第二C2c2蛋白结合的恒定区序列;

(g) 核酸, 其包含编码与所述第一C2c2蛋白结合的恒定区序列的核苷酸序列和所选择的指导序列的插入位点;

(h) 核酸, 其包含编码与所述第二C2c2蛋白结合的恒定区序列的核苷酸序列和所选择的指导序列的插入位点。

99. 根据权利要求97所述的试剂盒, 其包含核酸, 所述核酸包含编码第一C2c2指导RNA的核苷酸序列, 其中所述第一C2c2指导RNA包含与所述第一C2c2蛋白结合的恒定区序列。

100. 根据权利要求97所述的试剂盒, 其包含核酸, 所述核酸包含编码第二C2c2指导RNA的核苷酸序列, 其中所述第二C2c2指导RNA包含与所述第二C2c2蛋白结合的恒定区序列。

101. 根据权利要求97所述的试剂盒, 其包含核酸, 所述核酸包含编码与所述第一C2c2蛋白结合的恒定区序列的核苷酸序列和所选择的指导序列的插入位点。

102. 根据权利要求97所述的试剂盒, 其包含核酸, 所述核酸包含编码与所述第二C2c2蛋白结合的恒定区序列的核苷酸序列和所选择的指导序列的插入位点。

103. 根据权利要求99-102中任一项所述的试剂盒,其中所述核酸是表达载体,并且所述核苷酸序列与启动子可操作地连接。

用于检测靶RNA的方法和组合物

[0001] 交叉引用

[0002] 本申请要求2016年6月16日提交的美国临时专利申请号62/351,172、2016年8月22日提交的美国临时专利申请号62/378,156和2017年3月23日提交的美国专利申请号15/467,922的权益,这些申请以引用方式整体并入本文。

[0003] 关于联邦政府资助的研究或开发的声明

[0004] 本发明是在国家科学基金会 (National Science Foundation) 授予的1244557号在政府支持下完成的。政府对本发明享有一定权利。

[0005] 通过引用并入作为文本文件提供的序列列表

[0006] 序列列表在此提供为在2017年3月21日创建且大小为67KB的文本文件“BERK-337W0_SeqList_ST25.txt”。该文本文件的内容以引用方式整体并入本文。

[0007] 引言

[0008] 细菌适应性免疫系统采用CRISPR (成簇规律间隔短回文重复序列) 和CRISPR相关 (Cas) 蛋白进行RNA指导的核酸切割。虽然III型和VI型CRISPR系统通常靶向DNA底物,但是指导针对单链RNA (ssRNA) 底物的干扰复合物。在VI型CRISPR系统中,单亚基C2c2蛋白作为RNA指导的RNA内切核酸酶起作用。

[0009] CRISPR-Cas系统通过RNA指导的核酸干扰而赋予细菌和古细菌中的适应性免疫。在多种CRISPR类型中,仅相对罕见的VI型CRISPR系统被认为仅靶向单链RNA底物,这是由大效应蛋白C2c2赋予的活性。VI型操纵子共享其它CRISPR-Cas基因组基因座的共同特征,包括充当短病毒DNA区段的储库的CRISPR序列阵列。为了提供抗病毒免疫,将经处理的CRISPR阵列转录物 (crRNA) 与含Cas蛋白的监督复合物组装,所述监督复合物识别与crRNA的病毒来源区段 (称为间隔区) 具有序列互补性的核酸。

[0010] 免疫监督的第一步需要将由侧接病毒间隔区序列的重复序列组成的前体crRNA (pre-crRNA) 加工成单独功能性crRNA,每个所述crRNA含有单个病毒来源的序列片段。CRISPR系统采用多种机制产生成熟的crRNA,包括使用专用的内切核酸酶 (例如,I型和III型系统中的Cas6或Cas5d)、宿主内切核酸酶 (例如,RNA酶III) 与反式激活crRNA (tracrRNA, II型系统) 的偶联、或效应器酶 (例如,来自V型系统的Cpf1) 本身内源的核糖核酸酶活性。

发明内容

[0011] 本公开提供了用于检测单链靶RNA的方法。本公开提供了将前体C2c2指导RNA阵列切割成两个或更多个C2c2指导RNA的方法。本公开提供了用于检测样品中的靶RNA的试剂盒。

[0012] 提供了用于检测单链靶RNA的组合物和方法,其中所述方法包括 (i) 使具有多种RNA的样品与 (a) C2c2指导RNA和 (b) C2c2蛋白接触,所述C2c2指导RNA与单链靶RNA杂交,所述C2c2蛋白切割样品RNA;以及 (ii) 测量由切割产生的可检测信号。一旦主题C2c2蛋白由C2c2指导RNA激活,C2c2蛋白就变成切割样品RNA的内切核糖核酸酶,所述主题C2c2蛋白的激活是当样品包括指导RNA杂交的单链靶RNA (即,样品包括所靶向的单链靶RNA) 时发生的。

因此,当所靶向的单链靶RNA存在于样品中(例如,在一些情况下高于某一阈值量)时,结果是样品中RNA的切割,该切割可以使用任何合宜的检测方法检测(例如,使用经标记的检测RNA)。

[0013] 在一些情况下,可以通过使用前体C2c2指导RNA阵列来提供两个或更多个C2c2指导RNA,所述前体C2c2指导RNA阵列可以由C2c2蛋白切割成单独的指导RNA,并且这与C2c2蛋白是否具有完整的HEPN1结构域和/或HEPN2结构域无关。因此,还提供了将前体C2c2指导RNA阵列切割成两个或更多个C2c2指导RNA的方法。在一些情况下,C2c2蛋白缺乏催化活性的HEPN1结构域和/或缺乏催化活性的HEPN2结构域。

附图说明

[0014] 图1A至图1E示出了与内源C2c2基因座相关的示意图和证明重组C2c2蛋白的异源表达和纯化的实验。

[0015] 图2描绘了使用C2c2蛋白执行的切割测定的结果。

[0016] 图3描绘了显示C2c2强烈切割单链RNA的结果。

[0017] 图4A至图4B示出了使用C2c2蛋白的各种切割测定的图和结果。

[0018] 图5描绘了使用经标记的检测RNA(使用猝灭剂/荧光剂对)执行的检测由C2c2蛋白进行的RNA切割的实验的结果。

[0019] 图6描绘了证明由C2c2蛋白进行pre-crRNA加工的实验的结果。

[0020] 图7A至图7B描绘了示例性C2c2指导RNA的示意图和序列。

[0021] 图8A至图8F提供了各种C2c2多肽的氨基酸序列。

[0022] 图9A至图9C描绘了对前体crRNA转录物进行C2c2家族加工以产生成熟的crRNA。

[0023] 图10A至图10C描绘了CRISPR重复序列的结构和序列对LbuC2c2介导的crRNA生物发生的影响。

[0024] 图11A至图11C描绘了LbuC2c2对顺式和反式靶的指导依赖性ssRNA降解。

[0025] 图12A至图12D描绘了LbuC2c2的两种不同的核糖核酸酶活性。

[0026] 图13A至图13D描绘了C2c2介导的对复杂混合物中转录物的灵敏目视检测。

[0027] 图14提供了表2:在实施例8-12中描述且在图12A至图12D中示出的研究中使用的各种DNA底物。

[0028] 图15提供了表3:实施例8-12中描述的研究中使用的各种RNA底物。

[0029] 图16A至图16F描绘了数据,所述数据显示由C2c2进行的pre-crRNA加工是不依赖于间隔区序列的,可以在串联crRNA阵列上发生,受pre-cRNA的5'和/或3'侧翼区域中的突变影响,并且是不依赖于金属的。

[0030] 图17提供了pre-crRNA双突变对pre-crRNA加工活性的影响的总结。

[0031] 图18A至图18B描绘了LbuC2c2 ssRNA靶切割位点映射。

[0032] 图19A至图19C描绘了crRNA间隔区长度、反应温度和crRNA的5'末端序列对靶RNA切割效率的依赖性。

[0033] 图20A至图20C描绘了LbuC2c2与成熟crRNA和靶ssRNA的结合数据。

[0034] 图21描绘了RNA酶检测测定λ2-ssRNA的时间进程。

[0035] 图22A至图22B描绘了C2c2家族和C2c2比对的系统发生树。

[0036] 图23A至图23D描绘了C2c2的纯化和生产。

[0037] 图24A至图24C描绘了数据,所述数据显示C2c2蛋白加工前体crRNA转录物以产生成熟crRNA。a.C2c2蛋白的最大似然系统发生树。本研究使用的同系物以黄色突出显示。b.本研究使用的三种VI型CRISPR基因座的图。黑色矩形表示重复元件,黄色菱形表示间隔区序列。Cas1和Cas2仅存在于LshC2c2附近的基因组中。c.C2c2介导的对来源于LbuC2c2、LseC2c2和LshC2c2 CRISPR重复序列基因座的pre-crRNA的切割。OH:碱水解梯带;T1:RNA酶T1水解梯带;用100nM的C2c2和小于1nM的pre-crRNA执行加工切割反应。右侧描绘了切割的示意图,并且下方示出了所预测的pre-crRNA二级结构,其中箭头指示了所映射(mapped)的C2c2切割位点。

[0038] 图25A至图25C描绘了数据,所述数据显示LbuC2c2介导的crRNA生物发生取决于CRISPR重复序列的结构和序列。a.由LbuC2c2进行的对含有发夹的茎区和环区内的结构突变的pre-crRNA的代表性切割测定。下面列出的加工百分比是在60分钟处定量的(平均值±标准差,n=3)。b.条形图显示pre-crRNA加工对CRISPR重复序列的依赖性。野生型重复序列如下方所示,其中单独的条代表以红色标示的串联核苷酸突变。切割位点用卡通剪刀指示。在60分钟后测量加工百分比(平均值±标准差,n=3)。c.通过加入10-50mM的EDTA和EGTA以实现标准反应条件来测试crRNA加工反应的二价金属离子依赖性。

[0039] 图26A至图26D描绘了数据,所述数据显示LbuC2c2含有两种不同的核糖核酸酶活性。a.在37℃下由LbuC2c2执行的顺式ssRNA靶(黑色)和pre-crRNA(青色)切割的定量时间过程数据。指数拟合示出为实线(n=3),并且所计算的伪一级速率常数(pseudo-first-order rate constant) (k_{obs}) (平均值±标准差) 对于顺式ssRNA靶切割和pre-crRNA切割分别为 $9.74 \pm 1.15 \text{ 分钟}^{-1}$ 和 $0.12 \pm 0.02 \text{ 分钟}^{-1}$ 。b.LbuC2c2架构描绘了HEPN基序和加工缺陷型点突变体的位置。c、d.LbuC2c2突变体针对c中的pre-crRNA加工和d中的ssRNA靶向的核糖核酸酶活性。

[0040] 图27A至图27E示出,C2c2提供了对复杂混合物中转录物的灵敏检测。a.使用猝灭的荧光RNA报告物进行的LbuC2c2 RNA检测方法的示意图。b.在存在人总RNA的情况下,由LbuC2c2针对不同浓度的靶RNA在30分钟后产生的荧光信号的定量。RNA酶A示出为阳性RNA降解对照。(平均值±标准差,n=3)。c.对于不同量的人总RNA或细菌总RNA,由负载有β-肌动蛋白靶向crRNA的LbuC2c2在3h后产生的荧光信号的定量(作为β-肌动蛋白空阴性对照)。(平均值±标准差,n=3)。d.串联pre-crRNA加工也使得能够进行RNA检测。(平均值±标准差,n=3)。e.VI型CRISPR途径的模型突出显示了C2c2的核糖核酸酶活性。

[0041] 图28A至图28B描绘了C2c2家族和C2c2比对的完整系统发生树。a.C2c2蛋白的最大似然系统发生重建。叶包括GI蛋白数和来源生物;100个重新取样中的自展支持值被呈现用于内部分割。标尺是每个位点的取代。b.所分析的三种C2c2同系物的多重序列比对;坐标基于LbuC2c2。

[0042] 图29A至图29D描绘了与C2c2的纯化和生产有关的数据。所有C2c2同系物在大肠杆菌(E.coli)中表达为His-MBP融合物,并通过亲和性、离子交换和尺寸排阻色谱的组合进行纯化。通过与TEV蛋白酶一起孵育去除 Ni^{+} 亲和和标签。色谱级分的代表性SDS-PAGE凝胶示出在(a、b)中。c.来自Superdex 200 (16/60)柱的色谱图,表明C2c2作为单峰洗脱,缺乏核酸。d.对本稿图中使用的纯化蛋白的SDS PAGE分析。

[0043] 图30A至图30I描绘了在体外和体内通过C2c2进行的pre-crRNA加工的映射。a. 单个同源pre-crRNA阵列的LseC2c2和LshCc2c2切割的切割位点映射。OH: 碱水解梯带; T1: T1 RNA酶水解梯带。用100nM的C2c2和小于1nM的pre-crRNA进行切割反应。b-i. 对来自Shmakov等人¹⁰的LshCc2c2 (b-f) 和LseC2c2 (g-i) CRISPR阵列RNA测序实验的再分析 (分别为图S7和图5)。使用Bowtie2 (参见方法) 将所有读段 (b、g) 和经过滤的读段 (55nt或更少; 根据原始Shmakov等人的分析; c、h) 与每个CRISPR阵列进行严格比对。示出了Lsh (d-f) 和Lse (i) 的单独CRISPR重复序列间隔区的详细视图。5'末端pre-crRNA加工的差异由每个序列下方的箭头指示。可以使用分析的BAM比对文件。该映射清楚地表明, 从Lsh pre-crRNA产生的小RNA测序读段的5'末端映射到距所预测的发夹的碱基2nt的位置, 这与体外加工数据 (a) 一致。这种模式适用于从在沙氏纤毛菌 (*L. shahii*) 中的天然表达和在大肠杆菌中的异源表达两者检测到的所有成熟crRNA。不幸的是, LseC2c2 crRNA测序数据 (用于g-i中) 由于较低的读段深度而具有较少信息, 并且每个比对的crRNA表现出略微不同的5'末端, 几乎没有明显的均一性。经加工的重复序列中的一个重复序列 (重复序列间隔区2; i) 的映射与数据一致, 但由于读段深度不足而仅具有低置信度。

[0044] 图31A至图31D描绘了由C2c2进行的pre-crRNA加工是与间隔区序列无关的, 可以在串联crRNA阵列上发生, 受pre-crRNA的5'侧翼区域中的突变影响并且产生3'磷酸酯产物。a. 串联pre-crRNA阵列的LbuCc2c2切割的切割位点映射。OH: 碱水解梯带; T1: T1 RNA酶水解梯带。用100nM的LbuC2c2和小于1nM的pre-crRNA进行切割反应。右侧描绘了切割产物的示意图, 其中箭头指示了映射的C2c2切割产物。b. LbuC2c2 4聚突变体的pre-crRNA加工数据, 表明5'单链侧翼区域对于有效的pre-crRNA加工的重要性。在60分钟后测量pre-crRNA加工的百分比 (平均值±标准差, n=3)。c. 代表性的LbuC2c2 pre-crRNA切割时间过程, 表明相似的pre-crRNA加工速率与crRNA间隔区序列无关地发生, 间隔区A和间隔区λ2的伪一级速率常数 (k_{obs}) (平均值±标准差) 分别为 0.07 ± 0.04 分钟⁻¹和 0.08 ± 0.04 分钟⁻¹。d. 通过T4多核苷酸激酶 (PNK) 处理对经切割的RNA进行的端基分析。使用标准加工测定条件产生切割产物, 然后将所述切割产物与PNK一起孵育1小时以去除任何2', 3'-环状磷酸酯/3'-单磷酸酯。条带的延迟迁移指示从放射性标记的5'产物的3'末端去除了带电荷的单磷酸酯。

[0045] 图32A至图32C示出LbuC2c2催化顺式和反式靶的指导依赖性ssRNA降解。a. 两种C2c2指导依赖性ssRNA降解模式的示意图。b. 由LbuC2c2切割两种不同的放射性标记的ssRNA底物A和B。在37℃预形成100nM的C2c2与50nM的crRNA的复合物, 并在25℃下加入小于1nM的5'-标记的靶RNA时开始反应。反式切割反应包含等摩尔 (小于1nM) 浓度的放射性标记的非指导互补底物和未标记的在靶ssRNA。对于多种ssRNA底物, 观察到LbuC2c2仅在与互补crRNA结合时催化有效切割, 表明LbuC2c2:crRNA以RNA指导的方式切割ssRNA。此活性在下文中称为在靶切割或顺式靶切割。LbuC2c2介导的顺式切割导致多种产物的梯化, 其中切割优先在尿嘧啶残基之前发生, 类似于LshCc2c2⁹。在存在未标记的在靶 (与crRNA互补的) ssRNA的情况下重复非靶切割反应。与顺式执行的非靶切割实验不同, 观察到了非靶RNA的反式快速降解。对于顺式在靶切割和反式非靶切割观察到相似的RNA切割速率和几乎相同的切割产物暗示两种活性中的相同核酸酶中心。c. 在存在25mM EDTA的情况下, 测试负载有crRNA靶向间隔区A的LbuC2c2在顺式 (靶A经标记) 和反式 (在存在未标记的靶A的情况下, 靶B经标记) 切割条件下的切割活性。

[0046] 图33A至图33B示出了LbuC2c2 ssRNA靶切割位点映射。a.按照方法进行的ssRNA靶切割测定,证明LbuC2c2介导对具有相同的间隔区互补序列但不同的5'侧翼序列(具有不同长度和核苷酸组成)的若干个放射性标记的ssRNA底物的‘顺式’切割。ssRNA底物的序列示出在右侧,其中crRNA-A的间隔区互补序列以黄色突出显示。箭头指示检测到的切割位点。为了清楚起见而裁剪了凝胶。应当注意的是,在不同底物上产生的切割产物的模式(例如A.1对比A.2对比A.3)表明,切割位点选择主要由尿嘧啶偏好性驱动,并且表现为在crRNA互补靶序列内明显缺乏排他性切割机制,这与关于其它II类CRISPR单效应复合物(诸如Cas9和Cpf1)^{11,21}观察到的相反。有趣的是,针对底物A.0观察到的切割模式提示了对polyG序列的次要偏好性。b.按照方法,使用一系列覆盖ssRNA靶的长度的crRNA进行的LbuC2c2ssRNA靶切割测定。在该实验中使用的ssRNA底物的序列示出在凝胶下方,其中每个crRNA的间隔区互补序列以黄色突出显示。箭头指示所预测的切割位点。在每组泳道上方的的小图指示间隔区序列沿着靶(黄色框)的位置以及观察到(红色箭头)或不存在(黑色箭头)切割产物。同样地,应当注意的是,对于每种crRNA,切割产物长度分布是非常相似的,再次表明在crRNA结合序列内明显缺乏排他性切割。在反应的子集中不存在几种切割产物可能是由于ssRNA靶上存在结合的C2c2:crRNA,这可以通过任何顺式(分子内)或反式(分子间)LbuC2c2活性位点在空间上阻断对尿嘧啶的接取。虽然对LbuC2c2的原型间隔区侧翼位点(PFS)偏好性的适当分析超出了本研究的范围,但观察到了3'侧翼核苷酸的最小影响。预期的PFS碱基在所测试的每个指导序列旁边的图表中以红色标示。

[0047] 图34A至图34D描绘了RNA靶向对crRNA变异、温度和点突变的依赖性。a.按照方法进行的LbuC2c2 ssRNA靶切割测定,其中crRNA具有16nt、20nt或24nt的间隔区。b.按照方法在25℃和37℃处进行的LbuC2c2 ssRNA靶切割时间过程。c.按照方法进行的LbuC2c2 ssRNA靶切割时间过程,其中crRNA具有不同的5'端侧翼核苷酸突变。突变以红色突出显示。1-2个核苷酸的5'延伸可忽略不计地影响切割效率。相比之下,将侧翼区域缩短至3nt会减慢切割速率。d.点突变对C2c2在用于ssRNA靶向的HEPN基序内的保守残基突变体中的核糖核酸酶活性的影响。

[0048] 图35A至图35D描绘了LbuC2c2与成熟crRNA和靶ssRNA的结合数据。a.如在方法中所描述地进行过滤结合测定以确定成熟crRNA-A_{GG}与LbuC2c2-WT、LbuC2c2-dHEPN1、LbuC2c2-dHEPN2或LbuC2c2-dHEPN1/dHEPN2的结合亲和力。将定量的数据与标准结合等温线拟合。误差条代表三次独立实验的标准偏差。从三个独立实验测量的解离常数(平均值±标准偏差)为 27.1 ± 7.5 nM (LbuC2c2-WT)、 15.2 ± 3.2 nM (LbuC2c2-dHEPN1)、 11.5 ± 2.5 nM (LbuC2c2-dHEPN2)和 43.3 ± 11.5 nM (LbuC2c2-dHEPN1/dHEPN2)。b.如所示,关于LbuC2c2-dHEPN1/dHEPN2:crRNA-A_{GG}与‘在靶’A ssRNA或‘脱靶’B ssRNA之间的结合反应的代表性电泳迁移率变动测定。如方法中所述进行三次独立实验。为了清楚起见而裁剪了凝胶。c.将来自(b)的定量结合数据与标准结合同种型拟合。误差条代表三次独立实验的标准偏差。从三个独立实验测量的解离常数(平均值±标准偏差)对于ssRNA A为 1.62 ± 0.43 nM并且对于ssRNA B为未检测到的(>10 nM)。d.如在方法中所描述地进行过滤结合测定以确定成熟crRNA-A_{GA}与LbuC2c2-WT和LbuC2c2-R1079A的结合亲和力。将定量的数据与标准结合等温线拟合。误差条代表三次独立实验的标准偏差。从三个独立实验测量的解离常数(平均值±标准偏差)为 4.65 ± 0.6 nM (LbuC2c2-WT)和 2.52 ± 0.5 nM (LbuC2c2-R1079A)。应当注意的是,

这些结合亲和力与图a不同。这种差异被解释为在指导序列的5'端序列方面略有不同,其中图a中的指导序列以5'-GGCCA……开始并且图d中的指导序列以5'-GACCA开始。虽然天然指导序列(5'-GACCA)与LbuC2c2更紧密地结合,但是在这些指导序列变体的RNA靶向效率方面没有看到差异。

[0049] 图36A至图36B描绘了 λ 2-ssRNA的RNA酶检测测定的时间过程。a.在存在不断增加量的 λ 2ssRNA (0-1nM)的情况下,将LbuC2c2:crRNA- λ 2与RNA酶-Alert底物(Thermo-Fisher))和100ng HeLa总RNA在37℃孵育120分钟。每5分钟进行荧光测量。1nM的 λ 2ssRNA反应在可以测量第一时间点之前达到饱和。误差条代表三次独立实验的标准偏差。b.在存在或不存在在靶激活 λ 4ssRNA的情况下,将LbuC2c2:crRNA- λ 4或apo LbuC2c2在HeLa总RNA中孵育2小时。按照方法在Bioanalyzer 2100中的小RNA芯片上解析背景小RNA的降解。在apo LbuC2c2和LbuC2c2:crRNA- λ 4之间的片段谱中观察到了小差异。相比之下,在向反应中加入在靶ssRNA时,tRNA峰的急剧变宽和移位揭示了在激活LbuC2c2反式活性时反应中存在的其它结构化和非结构化RNA的广泛降解。

[0050] 图37描绘了切割实验,所述切割实验证明当蛋白在选自R1079(例如,R1079A)、R1072(例如,R1072A)和K1082(例如,K1082A)的任何氨基酸位置处包括突变时,LbuC2c2对前体指导RNA的切割(指导RNA加工)严重减少。

[0051] 图38A至图38C描绘了Cas13a家族内pre-crRNA加工的保守性。

[0052] 图39A至图39C描绘了Cas13a同系物的CRISPR基因座和crRNA重复序列架构。

[0053] 图40A至图40F描绘了对由LbuCas13a进行的pre-crRNA切割重要的残基。

[0054] 图41A至图41B描绘了Cas13a家族成员的螺旋1结构域和HEPN结构域的比对。

[0055] 图42A至图42D描绘了Cas13a家族成员对ssRNA的效率。

[0056] 图43A至图43D描绘了由Cas13a同系物进行的反式ssRNA切割。

[0057] 图44A至图44F描绘了Cas13a家族内的crRNA可交换性。

[0058] 图45A至图45C描绘了用于RNA检测的正交Cas13a亚家族的功能验证。

[0059] 图46A至图46D描绘了野生型(WT) LbuCas13a和LbuCas13a R1079A/K1080A双重突变体进行的crRNA阵列加工。

[0060] 图47A至图47C描绘了在pre-crRNA加工所涉及的区域中由LbuCas13a点突变体进行的反式切割。

[0061] 图48A至图48B描绘了LbuCas13a R1079A/K1080A双重突变体相对于野生型LbuCas13a的特征。

[0062] 图49提供表4。

[0063] 图50提供表5。

[0064] 图51提供表6。

[0065] 图52提供表7。

[0066] 图53提供表8。

[0067] 图54提供表9。

[0068] 图55A至图55B呈现VI-A型CRISPR系统功能的模型。

[0069] 图56A至图56K提供各种Cas13a多肽的氨基酸序列。

[0070] 图57提供各种Cas13a多肽的氨基酸序列的比对。

[0071] 定义

[0072] 本文可互换使用的术语“多核苷酸”和“核酸”是指任何长度的核苷酸的聚合形式，为核糖核苷酸或脱氧核糖核苷酸。因此，术语“多核苷酸”和“核酸”涵盖单链DNA；双链DNA；多链DNA；单链RNA；双链RNA；多链RNA；基因组DNA；cDNA；DNA-RNA杂交体；以及这样的聚合物，所述聚合物包含嘌呤碱基和嘧啶碱基或其它天然、经化学或生物化学修饰的、非天然的或衍生的核苷酸碱基。

[0073] 术语“寡核苷酸”是指3与100个核苷酸之间的单链或双链核酸（例如，DNA、RNA或经修饰的核酸）的多核苷酸。然而，出于本公开的目的，寡核苷酸的长度没有上限。寡核苷酸也称为“寡聚体”或“寡聚物”，并且可以从基因中分离、（体外和/或体内）转录，或化学合成。术语“多核苷酸”和“核酸”应理解为包括适用于所描述的实施方案的单链（诸如有义或反义）和双链多核苷酸。

[0074] “可杂交的”或“互补的”或“基本上互补的”是指核酸（例如RNA、DNA）包含这样的核苷酸序列，所述核苷酸序列使所述核酸能够在适当的体外和/或体内温度和溶液离子强度条件下，与另一核酸以序列特异性、反向平行方式（即，核酸特异性结合互补核酸）非共价结合（即，形成Watson-Crick碱基对和/或G/U碱基对）、“退火”或“杂交”。标准Watson-Crick碱基配对包括：腺嘌呤/腺苷（A）与胸腺嘧啶/胸苷（T）配对，A与尿嘧啶/尿苷（U）配对，以及鸟嘌呤/鸟苷（G）与胞嘧啶/胞苷（C）配对。此外，对于两个RNA分子（例如，dsRNA）之间的杂交，以及对于DNA分子与RNA分子的杂交（例如，当DNA靶核酸碱基与C2c2指导RNA配对时，等）：G也可以与U碱基配对。例如，在tRNA反密码子与mRNA中的密码子进行碱基配对的情境下，G/U碱基配对部分地负责遗传密码的简并性（即，冗余）。因此，在本公开的上下文中，G（例如，C2c2指导RNA分子的蛋白结合区段（dsRNA双链体）的G；与C2c2指导RNA碱基配对的靶核酸的G）被认为是与U和C两者互补。例如，当可以在C2c2指导RNA分子的蛋白结合区段（例如，dsRNA双链体）的给定核苷酸位置处制备G/U碱基对时，该位置不被认为是不互补的，而是替代地被认为是互补的。

[0075] 杂交和洗涤条件是众所周知的，并在Sambrook, J., Fritsch, E.F. 和 Maniatis, T. *Molecular Cloning: A Laboratory Manual*, 第二版, Cold Spring Harbor Laboratory Press, Cold Spring Harbor (1989) 中，特别是该参考文献中的第11章和表11.1中；和Sambrook, J. 和 Russell, W., *Molecular Cloning: A Laboratory Manual*, 第三版, Cold Spring Harbor Laboratory Press, Cold Spring Harbor (2001) 中例示。温度和离子强度的条件决定了杂交的“严谨性”。

[0076] 杂交需要两个核酸含有互补序列，但是碱基之间的错配是可能的。适合于两个核酸之间杂交的条件取决于核酸的长度和互补程度，核酸的长度和互补程度是本领域中众所周知的变量。两个核苷酸序列之间的互补程度越大，则具有这些序列的核酸的杂交体的解链温度（ T_m ）值越大。对于具有短链互补性（例如35个或更少个、30个或更少个、25个或更少个、22个或更少个、20个或更少个或18个或更少个核苷酸的互补性）的核酸之间的杂交，错配位置可能变得重要（参见Sambrook等人，出处同上，11.7-11.8）。通常，可杂交核酸的长度为8个核苷酸或更多个（例如，10个核苷酸或更多个、12个核苷酸或更多个、15个核苷酸或更多个、20个核苷酸或更多个、22个核苷酸或更多个、25个核苷酸或更多个或30个核苷酸或更多个）。可以根据诸如互补区域的长度和互补程度之类的因素，而根据需要调节温度和洗涤

溶液盐浓度。

[0077] 应当理解,多核苷酸的序列不需要与其可特异性杂交或可杂交的靶核酸的序列100%互补。此外,多核苷酸可以在一个或多个区段上杂交,使得插入或相邻区段不参与杂交事件(例如,环结构或发夹结构)。多核苷酸可包含与其将杂交的靶核酸序列内的靶区域60%或更多、65%或更多、70%或更多、75%或更多、80%或更多、85%或更多、90%或更多、95%或更多、98%或更多、99%或更多、99.5%或更多或100%的序列互补性。例如,这样的反义核酸将代表百分之90的互补性,在所述反义核酸中,反义化合物的20个核苷酸中的18个与靶区域互补并因此与靶区域特异性杂交。在该实例中,剩余的非互补核苷酸可以与互补核苷酸成簇或散布,并且不需要彼此邻接或与互补核苷酸邻接。可以使用任何合宜的方法确定核酸内的特定核酸序列段之间的互补百分比。示例性方法包括BLAST程序(基本局部比对搜索工具)和PowerBLAST程序(Altschul等人,J.Mol.Biol.,1990,215,403-410;Zhang和Madden,Genome Res.,1997,7,649-656),或通过使用使用默认设置的Gap程序(Wisconsin序列分析包,用于Unix的第8版,Genetics Computer Group,University Research Park,Madison Wis.),所述程序使用Smith和Waterman算法(Adv.Appl.Math.,1981,2,482-489)。

[0078] 术语“肽”、“多肽”和“蛋白”在本文中可互换使用,并且是指任何长度的氨基酸聚合形式,其可包括编码的和非编码的氨基酸、经化学修饰或生物化学修饰的或衍生化的氨基酸,以及具有经修饰的肽骨架的多肽。

[0079] 如本文所用的“结合”(例如,参照多肽的RNA结合结构域、与靶核酸结合等)是指大分子之间的非共价相互作用(例如,蛋白与核酸之间的非共价相互作用;在C2c2指导RNA复合物和靶核酸之间的非共价相互作用;等等)。当处于非共价相互作用状态时,大分子被称为“缔合的”或“相互作用”或“结合”(例如,当分子X被称为与分子Y相互作用时,则意味着分子X以非共价方式与分子Y结合)。并非结合相互作用的所有组分都需要是序列特异性的(例如,与DNA骨架中的磷酸残基接触),但结合相互作用的一些部分可以是序列特异性的。结合相互作用的特征通常在于解离常数(K_d)小于 10^{-6} M、小于 10^{-7} M、小于 10^{-8} M、小于 10^{-9} M、小于 10^{-10} M、小于 10^{-11} M、小于 10^{-12} M、小于 10^{-13} M、小于 10^{-14} M,或小于 10^{-15} M。“亲和力”是指结合强度,增加的结合亲和力与较低的 K_d 相关。

[0080] “结合结构域”是指能够与另一分子非共价结合的蛋白结构域。结合结构域可以与例如RNA分子(RNA结合结构域)和/或蛋白分子(蛋白结合结构域)结合。在具有蛋白结合结构域的蛋白的情况下,所述蛋白在一些情况下可以与其自身结合(以形成同源二聚体、同源三聚体等)和/或所述蛋白可以与一个或多个不同蛋白的一个或多个区域结合。

[0081] 术语“保守氨基酸取代”是指蛋白中具有相似侧链的氨基酸残基的可互换性。例如,一组具有脂肪族侧链的氨基酸由甘氨酸、丙氨酸、缬氨酸、亮氨酸和异亮氨酸组成;一组具有脂肪族羟基侧链的氨基酸由丝氨酸和苏氨酸组成;一组具有含酰胺侧链的氨基酸由天冬酰胺和谷氨酰胺组成;一组具有芳香族侧链的氨基酸由苯丙氨酸、酪氨酸和色氨酸组成;一组具有碱性侧链的氨基酸由赖氨酸、精氨酸和组氨酸组成;一组具有酸性侧链的氨基酸由谷氨酸和天冬氨酸组成;并且一组具有含硫侧链的氨基酸由半胱氨酸和甲硫氨酸组成。示例性保守氨基酸取代基是:缬氨酸-亮氨酸-异亮氨酸、苯丙氨酸-酪氨酸、赖氨酸-精氨酸、丙氨酸-缬氨酸-甘氨酸,以及天冬酰胺-谷氨酰胺。

[0082] 多核苷酸或多肽与另一多核苷酸或多肽具有一定百分比的“序列同一性”，意味着当比对时，在比较两个序列时，碱基或氨基酸的百分比是相同的，并且在相同的相对位置中。序列同一性可以以多种不同方式确定。为了确定序列同一性，可以使用各种方法和计算机程序（例如BLAST、T-COFFEE、MUSCLE、MAFFT、Phyre2等）比对序列，所述各种方法和计算机程序可通过万维网在包括ncbi.nlm.nih.gov/BLAST、ebi.ac.uk/Tools/msa/tcoffee/、ebi.ac.uk/Tools/msa/muscle/、mafft.cbrc.jp/alignment/software/、http://www.sbg.bio.ic.ac.uk/~phyre2/的网站处获得。参见例如Altschul等人，(1990)，J.Mol.Biol.215:403-10。

[0083] “编码”特定RNA的DNA序列是转录成RNA的DNA核酸序列。DNA多核苷酸可编码转化为蛋白的RNA (mRNA)，或者DNA多核苷酸可编码未翻译成蛋白的RNA (例如tRNA、rRNA、微RNA (miRNA)、“非编码”RNA (ncRNA)、C2c2指导RNA等)。

[0084] 本文可互换使用的术语“DNA调控序列”、“控制元件”和“调控元件”是指转录和翻译控制序列，诸如启动子、增强子、多腺苷酸化信号、终止子、蛋白降解信号等，所述转录和翻译控制序列提供和/或调控非编码序列（例如，Cas9指导RNA）或编码序列（例如，Cas9蛋白）的转录和/或调控编码多肽的翻译。

[0085] 如本文所用，“启动子序列”是能够结合RNA聚合酶并启动下游（3’方向）编码或非编码序列的转录的DNA调控区。出于本公开的目的，启动子序列在其3’末端处由转录起始位点结合并向上游（5’方向）延伸，以包含以高于背景的可检测水平起始转录所必需的最少数量的碱基或元件。在启动子序列内将发现转录起始位点，以及负责结合RNA聚合酶的蛋白结合结构域。真核启动子通常将（但不总是）包含“TATA”盒和“CAT”盒。包括诱导型启动子在内的各种启动子可用于驱动本公开的各种载体。

[0086] 如本文所用的应用于核酸、多肽、细胞或生物体的术语“天然存在的”或“未修饰的”或“野生型”是指在自然中存在的核酸、多肽、细胞或生物体。例如，存在于生物体（包括病毒）中、可以从自然中的来源分离并且尚未由人在实验室中有意修饰的多肽或多核苷酸序列是野生型（并且是天然存在的）。

[0087] 如本文所用的“重组”是指特定核酸（DNA或RNA）是克隆、限制、聚合酶链式反应（PCR）和/或连接步骤的各种组合的产物，导致具有结构编码或非编码序列的构建体可与自然系统中存在的内源核酸区分开来。编码多肽的DNA序列可以由cDNA片段或由一系列合成的寡核苷酸组装而成，以提供能够从包含在细胞中或无细胞转录和翻译系统中的重组转录单元表达的合成核酸。包含相关序列的基因组DNA也可用于形成重组基因或转录单位。非翻译DNA的序列可以存在于开放阅读框的5’端或3’端，其中此类序列不干扰编码区的操纵或表达，并且确实可以作用以通过各种机制调节所需产物的产生（参见下文的“DNA调控序列”）。或者，编码未翻译的RNA（例如，C2c2指导RNA）的DNA序列也可以被认为是重组的。因此，例如，术语“重组”核酸是指这样的核酸，所述核酸是非天然存在的，例如，是通过两个经由人工干预以其它方式分开的序列区段的人工组合而制备的。这种人工组合往往通过化学合成手段或通过人工操纵分离的核酸区段（例如通过基因工程技术）来实现。通常这样做是为了用编码相同氨基酸、保守氨基酸或非保守氨基酸的密码子替换密码子。或者，执行将所需功能的核酸区段连接在一起以产生所需的功能组合。这种人工组合往往通过化学合成手段或通过人工操纵分离的核酸区段（例如通过基因工程技术）来实现。当重组多核苷酸编码

多肽时,所编码的多肽的序列可以是天然存在的(“野生型”)或可以是天然存在序列的变体(例如,突变体)。因此,术语“重组”多肽不一定是指序列不是天然存在的多肽。相反,“重组”多肽由重组DNA序列编码,但是所述多肽的序列可以是天然存在的(“野生型”)或非天然存在的(例如,变体、突变体等)。因此,“重组”多肽是人工干预的结果,但可以是天然存在的氨基酸序列。

[0088] “载体”或“表达载体”是复制子,诸如质粒、噬菌体、病毒或粘粒,另一DNA区段(即“插入物”)可以附接到所述复制子以便引起附接片段在细胞中的复制。

[0089] “表达盒”包含与启动子可操作连接的DNA编码序列。“可操作地连接”是指并置,其中如此描述的组分处于允许它们以它们预期的方式起作用的关系中。例如,如果启动子影响编码序列的转录或表达,则所述启动子与所述编码序列可操作地连接。

[0090] 术语“重组表达载体”或“DNA构建体”在本文中可互换地使用,以指代包含载体和一个插入物的DNA分子。通常产生重组表达载体以用于表达和/或增殖插入物的目的,或用于构建其它重组核苷酸序列的目的。插入物可以或可以不与启动子序列可操作地连接,并且可以或可以不与DNA调控序列可操作地连接。

[0091] 任何给定的部件或部件组合可以是未标记的,或者可以用标记部分可检测地标记。在一些情况下,当标记两种或更多种组分时,它们可以用彼此可区分的标记部分标记。

[0092] 一般的分子和细胞生物化学方法可以在标准教科书中找到,所述标准教科书为诸如Molecular Cloning:A Laboratory Manual,第3版(Sambrook等人,Harbor Laboratory Press 2001);Short Protocols in Molecular Biology,第4版(Ausubel等人编辑,John Wiley&Sons 1999);Protein Methods(Bollag人,John Wiley&Sons 1996);Nonviral Vectors for Gene Therapy(Wagner等人编辑,Academic Press 1999);Viral Vectors(Kaplift和Loewy编辑,Academic Press 1995);Immunology Methods Manual(I.Lefkovits编辑,Academic Press 1997);以及Cell and Tissue Culture:Laboratory Procedures in Biotechnology(Doyle和Griffiths,John Wiley&Sons 1998),这些参考文献的公开内容以引用方式并入本文中。

[0093] 在进一步描述本发明之前,应当理解本发明不限于所描述的特定实施方案,因此当然可以改变。还应当理解,本文所使用的术语仅用于描述特定实施方案的目的,而非旨在为限制性的,因为本发明的范围将仅受所附权利要求书的限制。

[0094] 在提供值范围的情况下,应理解,除非上下文另有明确规定,每个居间值至该下限单位的十分之一在该范围的上限和下限之间,并且任何其它所述范围或该所述范围中的居间值被涵盖在本发明内。这些较小范围的上限和下限可以独立地包括在该等较小范围内,并且也被涵盖在本发明内,经受所述范围中任何特别排除的极限。在所述范围包括上限和下限中的一个或两个的情况下,排除了这些所包括的极限中的一者或两者的范围也被包括在本发明中。

[0095] 除非另有定义,本文所使用的所有技术和科学术语具有与本发明所属领域的普通技术人员通常理解的相同的含义。尽管与本文描述的那些方法和材料类似或等同的任何方法和材料也可用于本发明的实践或测试中,但是现在描述优选的方法和材料。本文提及的所有出版物均以引用方式并入本文,以公开和描述与所引用的出版物相关的方法和/或材料。

[0096] 必须注意,如本文和随附权利要求书中所使用,单数形式“一”、“一个”和“该”包括复数指示物,除非上下文另有明确规定。因此,例如,提及“一蛋白”包括多个此类蛋白,并且提及“该指导RNA”包括提及一个或多个此类指导RNA及本领域的技术人员已知的其等同物,等等。还应当注意,权利要求书可以撰写成排除任何任选要素。因此,本声明旨在充当使用与权利要求要素的叙述相关的诸如“单独地”、“仅”等排他性术语或使用“否定”限制的先行基础。

[0097] 应当理解,为了清楚起见而在单独实施方案的上下文中描述的本发明的某些特征也可以在单个实施方案中组合提供。相反,为简洁起见而在单个实施方案的上下文中描述的本发明的各种特征也可以单独地或以任何合适的子组合提供。与本发明有关的实施方案的所有组合由本发明具体包括,并且在本文中公开为如同每个和各个组合被单独和明确地公开一样。此外,各种实施例及其要素的所有子组合也由本发明具体包含,并且在本文中公开为如同每个此类子组合在此单独和明确地公开一样。

[0098] 本文所论述的出版物仅仅是出于它们在本申请的申请日之前的公开内容而提供的。本文中的任何内容都不应被解释为承认本发明不能凭借优先权发明而在这些出版物之前授权。此外,所提供的出版物的日期可能与实际出版日期不同,这可能需要独立确认。

具体实施方式

[0099] 本公开提供了用于检测单链靶RNA的方法。本公开提供了将前体C2c2指导RNA阵列切割成两个或更多个C2c2指导RNA的方法。本公开提供了用于检测样品中的靶RNA的试剂盒。术语“C2c2指导RNA”在本文中与“Cas13a指导RNA”可互换地使用,并且在一些情况下指导RNA被称为crRNA(例如,“Cas13a crRNA”);术语“C2c2蛋白”(或“C2c2多肽”)在本文中与“Cas13a蛋白”(或“Cas13a多肽”)可互换地使用。

[0100] 检测单链RNA的方法

[0101] 提供了用于检测单链靶RNA的组合物和方法,其中所述方法包括(i)使具有多种RNA的样品与(a)C2c2指导RNA和(b)C2c2蛋白接触,所述C2c2指导RNA与单链靶RNA杂交,所述C2c2蛋白切割样品中存在的RNA;以及(ii)测量由切割产生的可检测信号。一旦主题C2c2蛋白被C2c2指导RNA激活,C2c2蛋白就被激活并作为非特异性地切割样品中存在的RNA(包括非靶RNA)的内切核糖核酸酶起作用,所述主题C2c2蛋白的激活是当样品包括指导RNA杂交的单链靶RNA(即,样品包含所靶向的单链靶RNA)时发生的。因此,当所靶向的单链靶RNA存在于样品中(例如,在一些情况下高于某一阈值量)时,结果是样品中RNA(包括非靶RNA)的切割,该切割可以使用任何合宜的检测方法检测(例如,使用经标记的检测RNA)。接触步骤通常在包含二价金属离子的组合物中进行。接触步骤可以在无细胞环境中进行,例如在细胞外进行。接触步骤可以在细胞内进行。接触步骤可以在体外细胞中进行。接触步骤可以在离体细胞中进行。接触步骤可以在体内细胞中进行。在一些情况下,C2c2指导RNA作为RNA被提供;并且C2c2蛋白本身作为蛋白被提供。在一些情况下,C2c2指导RNA作为编码指导RNA的DNA被提供;并且C2c2蛋白本身作为蛋白被提供。在一些情况下,C2c2指导RNA作为RNA被提供;并且C2c2蛋白作为编码C2c2蛋白的RNA被提供。在一些情况下,C2c2指导RNA作为编码指导RNA的DNA被提供;并且C2c2蛋白作为编码C2c2蛋白的RNA被提供。在一些情况下,C2c2指导RNA作为RNA被提供;并且C2c2蛋白作为包含编码C2c2蛋白的核苷酸序列的DNA被提供。

在一些情况下,C2c2指导RNA作为编码指导RNA的DNA被提供;并且C2c2蛋白作为包含编码C2c2蛋白的核苷酸序列的DNA被提供。在一些情况下,本公开的方法提供了对样品中两种不同靶RNA(第一单链靶RNA和第二单链靶RNA)的基本上同时检测。

[0102] 在一些情况下,可以通过使用前体C2c2指导RNA阵列来提供两个或更多个(例如,3个或更多个、4个或更多个、5个或更多个、或6个或更多个)C2c2指导RNA,所述前体C2c2指导RNA阵列可以由C2c2蛋白切割成单独的(“成熟”)指导RNA;前体C2c2指导RNA的切割与C2c2蛋白是否具有完整的HEPN1结构域和/或HEPN2结构域无关。因此,还提供了将前体C2c2指导RNA阵列切割成两个或更多个C2c2指导RNA的方法。因此,C2c2指导RNA阵列可包括多于一个指导序列。在一些情况下,主题C2c2指导RNA可以包括来自前体crRNA的柄,但不一定必须包括多个指导序列。在一些情况下,C2c2蛋白缺乏催化活性的HEPN1结构域和/或缺乏催化活性的HEPN2结构域。接触步骤可以在无细胞环境中进行,例如在细胞外进行。接触步骤可以在细胞内进行。接触步骤可以在体外细胞中进行。接触步骤可以在离体细胞中进行。接触步骤可以在体内细胞中进行。

[0103] 在一些情况下(例如,当与C2c2指导RNA和C2c2蛋白接触时,当与前体C2c2指导RNA阵列和C2c2蛋白接触时,等等),在测量步骤之前使样品接触2小时或更短时间(例如,1.5小时或更短时间、1小时或更短时间、40分钟或更短时间、30分钟或更短时间、20分钟或更短时间、10分钟或更短时间、或5分钟或更短时间、或1分钟或更短时间)。例如,在一些情况下,在测量步骤之前使样品接触40分钟或更短时间。在一些情况下,在测量步骤之前使样品接触20分钟或更短时间。在一些情况下,在测量步骤之前使样品接触10分钟或更短时间。在一些情况下,在测量步骤之前使样品接触5分钟或更短时间。在一些情况下,在测量步骤之前使样品接触1分钟或更短时间。在一些情况下,在测量步骤之前使样品接触50秒至60秒。在一些情况下,在测量步骤之前使样品接触40秒至50秒。在一些情况下,在测量步骤之前使样品接触30秒至40秒。在一些情况下,在测量步骤之前使样品接触20秒至30秒。在一些情况下,在测量步骤之前使样品接触10秒至20秒。

[0104] 本公开提供了检测包含多种RNA(例如,包含靶RNA和多种非靶RNA)的样品中的单链RNA的方法。在一些情况下,所述方法包括:a)使样品与以下接触:(i)与单链靶RNA杂交的C2c2指导RNA,和(ii)切割样品中存在的RNA的C2c2蛋白;以及b)测量由C2c2蛋白介导的RNA切割产生的可检测信号。在一些情况下,所述方法包括:a)使样品与以下接触:i)前体C2c2指导RNA阵列,所述前体C2c2指导RNA阵列包含两个或更多个C2c2指导RNA,每个所述C2c2指导RNA具有不同的指导序列;和(ii)C2c2蛋白,所述C2c2蛋白将前体C2c2指导RNA阵列切割成单独的C2c2指导RNA,并且还切割样品的RNA;以及b)测量由C2c2蛋白介导的RNA切割产生的可检测信号。在一些情况下,本公开的方法提供了对样品中两种不同靶RNA(第一单链靶RNA和第二单链靶RNA)的基本上同时检测。

[0105] 用于检测包含多种RNA(包括单链靶RNA和多种非靶RNA)的样品中的单链RNA(单链靶RNA)的本公开的方法可以以高灵敏度检测单链靶RNA。在一些情况下,本公开的方法可用于检测包含多种RNA(包括单链靶RNA和多种非靶RNA)的样品中存在的靶单链RNA,其中靶单链RNA以每 10^7 个非靶RNA一个或多个拷贝(例如,每 10^6 个非靶RNA一个或多个拷贝、每 10^5 个非靶RNA一个或多个拷贝、每 10^4 个非靶RNA一个或多个拷贝、每 10^3 个非靶RNA一个或多个拷贝、每 10^2 个非靶RNA一个或多个拷贝、每50个非靶RNA一个或多个拷贝、每20个非靶RNA一个或多个拷贝、每10个非靶RNA一个或多个拷贝)存在。

或多个拷贝、每10个非靶RNA一个或多个拷贝、或每5个非靶RNA一个或多个拷贝)存在。

[0106] 在一些情况下,本公开的方法可以检测在包含多种RNA(包括单链靶RNA和多种非靶RNA)的样品中存在的靶单链RNA,其中所述靶单链RNA以每 10^7 个非靶RNA一个拷贝至每10个非靶RNA一个拷贝(例如,从每 10^7 个非靶RNA 1个拷贝至每 10^2 个非靶RNA 1个拷贝、从每 10^7 个非靶RNA 1个拷贝至每 10^3 个非靶RNA 1个拷贝、从每 10^7 个非靶RNA 1个拷贝至每 10^4 个非靶RNA 1个拷贝、从每 10^7 个非靶RNA 1个拷贝至每 10^5 个非靶RNA 1个拷贝、从每 10^7 个非靶RNA 1个拷贝至每 10^6 个非靶RNA 1个拷贝,从每 10^6 个非靶RNA 1个拷贝至每10个非靶RNA 1个拷贝、从每 10^6 个非靶RNA 1个拷贝至每 10^2 个非靶RNA 1个拷贝、从每 10^6 个非靶RNA 1个拷贝至每 10^3 个非靶RNA 1个拷贝、从每 10^6 个非靶RNA 1个拷贝至每 10^4 个非靶RNA 1个拷贝、从每 10^6 个非靶RNA 1个拷贝至每 10^5 个非靶RNA 1个拷贝、从每 10^5 个非靶RNA 1个拷贝至每10个非靶RNA 1个拷贝、从每 10^5 个非靶RNA 1个拷贝至每 10^2 个非靶RNA 1个拷贝、从每 10^5 个非靶RNA 1个拷贝至每 10^3 个非靶RNA 1个拷贝或从每 10^5 个非靶RNA 1个拷贝至每 10^4 个非靶RNA 1个拷贝)存在。

[0107] 在一些情况下,本公开的方法可以检测在包含多种RNA(包括单链靶RNA和多种非靶RNA)的样品中存在的靶单链RNA,其中所述靶单链RNA以每 10^7 个非靶RNA一个拷贝至每100个非靶RNA一个拷贝(例如,从每 10^7 个非靶RNA 1个拷贝至每 10^2 个非靶RNA 1个拷贝、从每 10^7 个非靶RNA 1个拷贝至每 10^3 个非靶RNA 1个拷贝、从每 10^7 个非靶RNA 1个拷贝至每 10^4 个非靶RNA 1个拷贝、从每 10^7 个非靶RNA 1个拷贝至每 10^5 个非靶RNA 1个拷贝、从每 10^7 个非靶RNA 1个拷贝至每 10^6 个非靶RNA 1个拷贝,从每 10^6 个非靶RNA 1个拷贝至每100个非靶RNA 1个拷贝、从每 10^6 个非靶RNA 1个拷贝至每 10^2 个非靶RNA 1个拷贝、从每 10^6 个非靶RNA 1个拷贝至每 10^3 个非靶RNA 1个拷贝、从每 10^6 个非靶RNA 1个拷贝至每 10^4 个非靶RNA 1个拷贝、从每 10^6 个非靶RNA 1个拷贝至每 10^5 个非靶RNA 1个拷贝、从每 10^5 个非靶RNA 1个拷贝至每100个非靶RNA 1个拷贝、从每 10^5 个非靶RNA 1个拷贝至每 10^2 个非靶RNA 1个拷贝、从每 10^5 个非靶RNA 1个拷贝至每 10^3 个非靶RNA 1个拷贝或从每 10^5 个非靶RNA 1个拷贝至每 10^4 个非靶RNA 1个拷贝)存在。

[0108] 在一些情况下,对于检测样品中单链靶RNA的本方法,检测阈值为10nM或更低。术语“检测阈值”在本文中用于描述为了进行检测样品中必须存在的靶RNA的最小量。因此,作为说明性实例,当检测阈值为10nM时,则当靶RNA以10nM或更高的浓度存在于样品中时,可以检测到信号。在一些情况下,本公开的方法具有5nM或更低的检测阈值。在一些情况下,本公开的方法具有1nM或更低的检测阈值。在一些情况下,本公开的方法具有0.5nM或更低的检测阈值。在一些情况下,本公开的方法具有0.1nM或更低的检测阈值。在一些情况下,本公开的方法具有0.05nM或更低的检测阈值。在一些情况下,本公开的方法具有0.01nM或更低的检测阈值。在一些情况下,本公开的方法具有0.005nM或更低的检测阈值。在一些情况下,本公开的方法具有0.001nM或更低的检测阈值。在一些情况下,本公开的方法具有0.0005nM或更低的检测阈值。在一些情况下,本公开的方法具有0.0001nM或更低的检测阈值。在一些情况下,本公开的方法具有0.00005nM或更低的检测阈值。在一些情况下,本公开的方法具有0.00001nM或更低的检测阈值。在一些情况下,本公开的方法具有10pM或更低的检测阈值。在一些情况下,本公开的方法具有1pM或更低的检测阈值。在一些情况下,本公开的方法具有500fM或更低的检测阈值。在一些情况下,本公开的方法具有250fM或更低的检测阈值。

在一些情况下,本公开的方法具有100fM或更低的检测阈值。在一些情况下,本公开的方法具有50fM或更低的检测阈值。

[0109] 在一些情况下,检测阈值(在本方法中用于检测单链靶RNA)在500fM至1nM(例如,500fM至500pM、500fM至200pM、500fM至100pM、500fM至10pM、500fM至1pM、800fM至1nM、800fM至500pM、800fM至200pM、800fM至100pM、800fM至10pM、800fM至1pM、1pM至1nM、1pM至500pM、1pM至200pM、1pM至100pM或1pM至10pM)的范围内(其中浓度是指可以检测至靶RNA的阈值靶RNA浓度)。在一些情况下,本公开的方法具有在800fM至100pM的范围内的检测阈值。在一些情况下,本公开的方法具有在1pM至10pM的范围内的检测阈值。在一些情况下,本公开的方法具有在10fM至500fM,例如10fM至50fM、50fM至100fM、100fM至250fM或250fM至500fM的范围内的检测阈值。

[0110] 在一些情况下,可以在样品中检测到单链靶RNA的最小浓度在500fM至1nM(例如,500fM至500pM、500fM至200pM、500fM至100pM、500fM至10pM、500fM至1pM、800fM至1nM、800fM至500pM、800fM至200pM、800fM至100pM、800fM至10pM、800fM至1pM、1pM至1nM、1pM至500pM、1pM至200pM、1pM至100pM或1pM至10pM)的范围内。在一些情况下,可以在样品中检测到单链靶RNA的最小浓度在800fM至100pM的范围内。在一些情况下,可以在样品中检测到单链靶RNA的最小浓度在1pM至10pM的范围内。

[0111] 在一些情况下,本公开的方法可以检测包含多种RNA(包括单链靶RNA和多种非靶RNA)的样品中存在的靶单链RNA,其中所述靶单链RNA以低至500fM(例如,低至800fM、低至1pM、低至10pM或低至100pM)的浓度存在。在一些情况下,本公开的方法可以检测包含多种RNA(包括单链靶RNA和多种非靶RNA)的样品中存在的靶单链RNA,其中所述靶单链RNA以低至1pM的浓度存在。

[0112] 在一些情况下,本公开的方法可以检测包含多种RNA(包括单链靶RNA和多种非靶RNA)的样品中存在的靶单链RNA,其中所述靶单链RNA以低至500fM(例如,低至800fM、低至1pM、低至10pM或低至100pM)的浓度存在,并且其中在测量步骤之前使样品接触60分钟或更短时间(例如,在一些情况下为40分钟或更短时间)。在一些情况下,本公开的方法可以检测包含多种RNA(包括单链靶RNA和多种非靶RNA)的样品中存在的靶单链RNA,其中所述靶单链RNA以低至1pM的浓度存在,并且其中在测量步骤之前使样品接触60分钟或更短时间(例如,在一些情况下为40分钟或更短时间)。

[0113] 例如,在一些情况下,本公开的方法提供用于检测以500fM或更高(例如,800fM或更高、1pM或更高、5pM或更高、10pM或更高)的浓度存在于样品中的靶RNA。在一些情况下,本公开的方法提供用于检测以1pM或更高(例如,2pM或更高、5pM或更高、或8pM或更高)的浓度存在于样品中的靶RNA。在一些情况下,本公开的方法提供用于检测以500fM或更高(例如,1pM或更高、5pM或更高、10pM或更高)的浓度存在于样品中的靶RNA,其中在测量步骤之前使样品接触60分钟或更短时间(例如,在一些情况下为40分钟或更短时间)。在一些情况下,本公开的方法提供用于检测以1pM或更高(例如,2pM或更高、5pM或更高、或8pM或更高)的浓度存在于样品中的靶RNA,其中在测量步骤之前使样品接触60分钟或更短时间(例如,在一些情况下为40分钟或更短时间)。

[0114] 在一些情况下,本公开的方法提供用于检测以10nM或更低的浓度存在于样品中的靶RNA。在一些情况下,本公开的方法提供用于检测以5nM或更低的浓度存在于样品中的靶

RNA。在一些情况下,本公开的方法提供用于检测以1nM或更低的浓度存在于样品中的靶RNA。在一些情况下,本公开的方法提供用于检测以0.5nM或更低的浓度存在于样品中的靶RNA。在一些情况下,本公开的方法提供用于检测以0.1nM或更低的浓度存在于样品中的靶RNA。在一些情况下,本公开的方法提供用于检测以0.05nM或更低的浓度存在于样品中的靶RNA。在一些情况下,本公开的方法提供用于检测以0.01nM或更低的浓度存在于样品中的靶RNA。在一些情况下,本公开的方法提供用于检测以0.005nM或更低的浓度存在于样品中的靶RNA。在一些情况下,本公开的方法提供用于检测以0.001nM或更低的浓度存在于样品中的靶RNA。在一些情况下,本公开的方法提供用于检测以0.0005nM或更低的浓度存在于样品中的靶RNA。在一些情况下,本公开的方法提供了以0.0001nM或更低的浓度检测样品中存在的靶RNA。在一些情况下,本公开的方法提供用于检测以0.00005nM或更低的浓度存在于样品中的靶RNA。在一些情况下,本公开的方法提供用于检测以0.00001nM或更低的浓度存在于样品中的靶RNA。

[0115] 在一些情况下,本公开的方法提供用于检测以 10^{-6} nM至1nM,例如 10^{-6} nM至 5×10^{-6} nM、 5×10^{-6} nM至 10^{-5} nM、 10^{-5} nM至 5×10^{-5} nM、 5×10^{-5} nM至 10^{-4} nM、 10^{-4} nM至 5×10^{-4} nM、 5×10^{-4} nM至 10^{-3} nM、 10^{-3} nM至 5×10^{-3} nM、 5×10^{-3} nM至 10^{-2} nM、 10^{-2} nM至 5×10^{-2} nM、 5×10^{-2} nM至0.1nM、0.1nM至0.5nM、0.5nM至1nM、1nM至5nM或5nM至10nM的浓度存在于样品中的靶RNA。

[0116] 在一些情况下,本公开的方法提供用于检测以低于10nM的浓度存在于样品中的靶RNA。在一些情况下,本公开的方法提供用于检测以低于5nM的浓度存在于样品中的靶RNA。在一些情况下,本公开的方法提供用于检测以低于1nM的浓度存在于样品中的靶RNA。在一些情况下,本公开的方法提供用于检测以低于0.5nM的浓度存在于样品中的靶RNA。在一些情况下,本公开的方法提供用于检测以低于0.1nM的浓度存在于样品中的靶RNA。在一些情况下,本公开的方法提供用于检测以低于0.05nM的浓度存在于样品中的靶RNA。在一些情况下,本公开的方法提供用于检测以低于0.01nM的浓度存在于样品中的靶RNA。在一些情况下,本公开的方法提供用于检测以低于0.005nM的浓度存在于样品中的靶RNA。在一些情况下,本公开的方法提供用于检测以低于0.001nM的浓度存在于样品中的靶RNA。在一些情况下,本公开的方法提供用于检测以低于0.0005nM的浓度存在于样品中的靶RNA。在一些情况下,本公开的方法提供用于检测以低于0.0001nM的浓度存在于样品中的靶RNA。在一些情况下,本公开的方法提供用于检测以低于0.00005nM的浓度存在于样品中的靶RNA。在一些情况下,本公开的方法提供用于检测以低于0.00001nM的浓度存在于样品中的靶RNA。

[0117] 在一些情况下,本公开的方法可用于确定样品(例如,包含靶RNA和多种非靶RNA的样品)中靶RNA的量。确定样品中靶RNA的量可以包括将从测试样品产生的可检测信号的量与从参考样品产生的可检测信号的量进行比较。确定样品中靶RNA的量可包括:测量可检测信号以产生测试测量值;测量由参考样品产生的可检测信号以产生参考测量值;以及将测试测量值与参考测量值进行比较以确定样品中存在的靶RNA的量。

[0118] 例如,在一些情况下,用于确定样品中靶RNA的量的本公开的方法包括:a)使样品(例如,包含靶RNA和多种非靶RNA的样品)与以下接触:(i)与单链靶RNA杂交的C2c2指导RNA,和(ii)切割样品中存在的RNA的C2c2蛋白;b)测量由C2c2蛋白介导的RNA切割产生的可检测信号,产生测试测量值;c)测量由参考样品产生的可检测信号以产生参考测量值;以及d)将测试测量值与参考测量值进行比较,以确定样品中存在的靶RNA的量。

[0119] 作为另一实例,在一些情况下,用于确定样品中靶RNA的量的本公开的方法包括:
a) 使样品(例如,包含靶RNA和多种非靶RNA的样品)与以下接触:i) 前体C2c2指导RNA阵列,所述前体C2c2指导RNA阵列包含两个或更多个C2c2指导RNA,每个所述C2c2指导RNA具有不同的指导序列;和(ii) C2c2蛋白,所述C2c2蛋白将前体C2c2指导RNA阵列切割成单独的C2c2指导RNA,并且还切割样品的RNA;b) 测量由C2c2蛋白介导的RNA切割产生的可检测信号,产生测试测量值;c) 测量由两个或更多个参考样品中的每一个产生的可检测信号,以产生两个或更多个参考测量值;以及d) 将测试测量值与参考测量值进行比较,以确定样品中存在的靶RNA的量。

[0120] 样品

[0121] 主题样品包括多种靶RNA。术语“多个”在本文中用于表示两个或更多个。因此,在一些情况下,样品包括两种或更多种(例如,3种或更多种、5种或更多种、10种或更多种、20种或更多种、50种或更多种、100种或更多种、500种或更多种、1,000种或更多种、或5,000种或更多种)RNA。本方法可以用作检测复杂的RNA混合物中存在的单链靶RNA的非常灵敏的方法。因此,在一些情况下,样品包括5种或更多种(例如,10种或更多种、20种或更多种、50种或更多种、100种或更多种、500种或更多种、1,000种或更多种、或5,000种或更多种RNA)序列彼此不同的RNA。在一些情况下,样品包括10种或更多种、20种或更多种、50种或更多种、100种或更多种、500种或更多种、 10^3 种或更多种、 5×10^3 种或更多种、 10^4 种或更多种、 5×10^4 种或更多种、 10^5 种或更多种、 5×10^5 种或更多种、 10^6 种或更多种、 5×10^6 种或更多种、或 10^7 种或更多种序列彼此不同的RNA。在一些情况下,样品包含10种至20种、20种至50种、50种至100种、100种至500种、500种至 10^3 种、 10^3 种至 5×10^3 种、 5×10^3 种至 10^4 种、 10^4 种至 5×10^4 种、 5×10^4 种至 10^5 种、 10^5 种至 5×10^5 种、 5×10^5 种至 10^6 种、 10^6 种至 5×10^6 种、或 5×10^6 种至 10^7 种、或多于 10^7 种序列彼此不同的RNA。在一些情况下,样品包含5种至 10^7 种序列彼此不同的RNA(例如,5种至 10^6 种、5种至 10^5 种、5种至50,000种、5种至30,000种、10种至 10^6 种、10种至 10^5 种、10种至50,000种、10种至30,000种、20种至 10^6 种、20种至 10^5 种、20种至50,000种、或20种至30,000种序列彼此不同的RNA)。在一些情况下,样品包含5种至50,000种序列彼此不同的RNA(例如,5种至30,000种、10种至50,000种、或10种至30,000种序列彼此不同的RNA)。在一些情况下,样品包括20种或更多种序列彼此不同的RNA。在一些情况下,样品包括来自细胞裂解物(例如,真核细胞裂解物、哺乳动物细胞裂解物、人细胞裂解物、原核细胞裂解物、植物细胞裂解物等)的RNA。例如,在一些情况下,样品包括从细胞诸如真核细胞(例如,哺乳动物细胞,诸如人细胞)表达的RNA。

[0122] 术语“样品”在本文中用于表示包括单链RNA的任何样品。样品可以来源于任何来源,例如,样品可以是纯化RNA的合成组合;样品可以是细胞裂解物、富含RNA的细胞裂解物,或从细胞裂解物中分离和/或纯化的RNA。样品可以来自患者(例如,用于诊断目的)。样品可来自经透化的细胞。样品可以来自交联细胞。样品可以在组织切片中。样品可以来自通过交联之后进行脱脂和调节以产生均匀的折射率而制备的组织。在例如Shah等人,Development (2016) 143,2862-2867doi:10.1242/dev.138560中已经描述了通过交联之后进行脱脂和调节以产生均匀的折射率而制备的组织的实例。

[0123] “样品”可包括单链靶RNA和多种非靶RNA。在一些情况下,靶单链RNA以每10个非靶RNA一个拷贝、每20个非靶RNA一个拷贝、每25个非靶RNA一个拷贝、每50个非靶RNA一个拷

贝、每100个非靶RNA一个拷贝、每500个非靶RNA一个拷贝、每 10^3 个非靶RNA一个拷贝、每 5×10^3 个非靶RNA一个拷贝、每 10^4 个非靶RNA一个拷贝、每 5×10^4 个非靶RNA一个拷贝、每 10^5 个非靶RNA一个拷贝、每 5×10^5 个非靶RNA一个拷贝、每 10^6 个非靶RNA一个拷贝、或每 10^6 个非靶RNA少于一个拷贝存在于样品中。在一些情况下,靶单链RNA以每10个非靶RNA一个拷贝至每20个非靶RNA 1个拷贝、每20个非靶RNA 1个拷贝至每50个非靶RNA 1个拷贝、每50个非靶RNA 1个拷贝至每100个非靶RNA 1个拷贝、每100个非靶RNA 1个拷贝至每500个非靶RNA 1个拷贝、每500个非靶RNA 1个拷贝至每 10^3 个非靶RNA 1个拷贝、每 10^3 个非靶RNA 1个拷贝至每 5×10^3 个非靶RNA 1个拷贝、每 5×10^3 个非靶RNA 1个拷贝至每 10^4 个非靶RNA 1个拷贝、每 10^4 个非靶RNA 1个拷贝至每 10^5 个非靶RNA 1个拷贝、每 10^5 个非靶RNA 1个拷贝至每 10^6 个非靶RNA 1个拷贝,或每 10^6 个非靶RNA 1个拷贝至每 10^7 个非靶RNA 1个拷贝存在于样品中。

[0124] 合适的样品包括但不限于血液样品、血清样品、血浆样品、尿液样品、抽吸物样品和活组织检查样品。因此,关于患者的术语“样品”涵盖血液和其它生物来源的液体样品、实体组织样品(诸如活检样本或组织培养物或来源于其的细胞及其后代)。该定义还包括在获得后已经以任何方式操纵的样品,诸如通过用试剂处理;洗涤;或富集某些细胞群,诸如癌细胞。该定义还包括已经富集了特定类型的分子(例如,RNA)的样品。术语“样品”涵盖生物样品,诸如临床样品,诸如血液、血浆、血清、抽吸物、脑脊髓液(CSF),并且还包括通过手术切除获得的组织、通过活组织检查获得的组织、培养中的细胞、细胞上清液、细胞裂解物、组织样品、器官、骨髓等。“生物样品”包括来源于其的生物流体(例如,癌细胞、感染细胞等),例如包含从此类细胞获得的RNA的样品(例如,细胞裂解物或其它包含RNA的细胞提取物)。

[0125] 样品可包括或可从多种细胞、组织、器官或无细胞流体中的任何一种获得。合适的样品来源包括真核细胞、细菌细胞和古细菌细胞。合适的样品来源包括单细胞生物和多细胞生物。合适的样品来源包括单细胞真核生物;植物或植物细胞;藻类细胞,例如布朗葡萄藻(*Botryococcus braunii*)、莱茵衣藻(*Chlamydomonas reinhardtii*)、迦得拟微绿球藻(*Nannochloropsis gaditana*)、蛋白核小球藻(*Chlorella pyrenoidosa*)、展枝马尾藻(*Sargassum patens*)、羽藻(*C. agardh*)等等;真菌细胞(例如,酵母细胞);动物细胞、组织或器官;来自无脊椎动物的细胞、组织或器官(例如,果蝇、刺胞动物、棘皮动物、线虫、昆虫、蜘蛛类动物,等);来自脊椎动物(例如,鱼类、两栖动物、爬行动物、鸟类、哺乳动物)的细胞、组织、流体或器官;来自哺乳动物(例如,人;非人灵长类动物;有蹄类动物;猫科动物;牛;绵羊;山羊;等)的细胞、组织、流体或器官。合适的样品来源包括线虫、原生动物等。合适的样品来源包括寄生虫,诸如蠕虫、疟疾寄生虫等。

[0126] 合适的样品来源包括例如以下六个界中任何一个界的细胞、组织或生物:细菌界(例如,真细菌界);古细菌界;原生生物界;真菌界;植物界;和动物界。合适的样品来源包括原生生物界的植物样成员,包括但不限于藻类(例如,绿藻、红藻、灰胞藻、蓝细菌);原生生物界的真菌样成员,例如粘液菌、水霉菌等;原生生物界的动物样成员,例如鞭毛虫类(例如,眼虫藻)、变形虫类(amoeboids)(例如,变形虫)、孢子虫类(例如,顶复门(Apicomplexa)、粘体动物门(Myxozoa)、微孢子虫纲(Microsporidia)和纤毛虫类(例如,草履虫)。合适的样品来源包括真菌界的成员,包括但不限于以下门中的任何门的成员:担子菌门(担子菌类;例如,伞菌属、鹅膏菌属、牛肝菌属、鸡油菌属等成员);子囊菌门(子囊菌类,包括例如酵母菌);菌藻门(地衣);接合菌门(接合真菌);以及不完全菌门。合适的样品

来源包括植物界的成员,包括但不限于以下门中的任何门的成员:苔藓植物门(例如,藓类)、角苔植物门(例如,角苔类)、苔类植物门(Hepaticophyta)(例如,苔类)、石松植物门(例如,石松类)、楔叶植物门(例如,木贼类)、裸蕨植物门(例如,松叶蕨类)、瓶尔小草门、蕨门(例如,蕨类)、苏铁门、银杏门、松柏门、买麻藤门和木兰门(例如,显花植物)。合适的样品来源包括动物界的成员,包括但不限于以下门中的任何门的成员:多孔动物门(海绵动物);扁盘动物门;直泳虫门(海洋无脊椎动物的寄生物);菱形虫门;刺胞动物门(珊瑚、海葵、海蜇、海笔、海肾、立方水母);栉水母门(栉水母类);扁虫动物门(扁虫类);纽形动物门(纽虫类);颚胃动物门(Ngathostomulida)(有颚蠕虫);腹毛动物门;轮虫动物门;曳鳃动物门;动吻动物门;铠甲动物门;棘头动物门;内肛亚门;线虫动物门;线形动物门;环口动物门;软体动物门(软体动物);星虫动物门(方格星虫(peanut worms));环节动物门(环节蠕虫);缓步动物门(缓步动物);有爪动物门(栉蚕);节肢动物门(包括以下亚门:有螯肢亚门、多足亚门、六足亚门和甲壳亚门,其中有螯肢亚门包括例如蛛形纲、肢口纲和海蜘蛛纲,多足亚门包括例如唇足纲(唇足类)、倍足纲(多足类)、少足纲(Paropoda)和综合纲,六足亚门包括昆虫纲,甲壳亚门包括虾、磷虾、藤壶等;帚虫动物门;外肛动物门(苔藓动物);腕足动物门;棘皮动物门(例如,海星、海维菊、毛头星、海胆、海参、海蛇尾、脆篮(brittle baskets)等);毛颚动物门(箭虫);半索动物门(玉钩虫);和脊索动物门。合适的脊索动物门成员包括以下亚门的任何成员:尾索动物亚门(海鞘纲;包括海鞘目、樽海鞘目和幼形目);头索亚门(文昌鱼);盲鳗纲(盲鳗);和脊椎动物亚门,其中脊椎动物亚门成员包括例如以下纲的成员:鳃鳗纲(七鳃类),软骨鱼纲(软骨鱼),辐鳍鱼纲(辐鳍鱼),腔棘鱼纲(腔棘鱼),肺鱼纲(肺鱼),爬行纲(爬行动物,例如蛇、短吻鳄、鳄鱼、蜥蜴等),鸟纲(鸟);和哺乳纲(哺乳动物)。合适的植物包括任何单子叶植物和任何双子叶植物。

[0127] 合适的样品来源包括取自生物;从生物体分离的特定细胞或细胞群的细胞、流体、组织或器官等等。例如,在生物是植物的情况下,合适的来源包括木质部、韧皮部、形成层、叶、根等。在生物是动物的情况下,合适的来源包括特定组织(例如,肺、肝、心脏、肾、脑、脾、皮肤、胎儿组织等)、或特定细胞类型(例如,神经元细胞、上皮细胞、内皮细胞、星形胶质细胞、巨噬细胞、神经胶质细胞、胰岛细胞、T淋巴细胞、B淋巴细胞等)。

[0128] 在一些情况下,样品的来源是患病的细胞、流体、组织或器官。在一些情况下,样品的来源是正常(非患病的)细胞、流体、组织或器官。在一些情况下,样品来源是病原体感染的细胞、组织或器官。病原体包括病毒、真菌、蠕虫、原生动物、疟疾寄生虫、疟原虫寄生虫、弓形虫寄生虫,血吸虫寄生虫等。“蠕虫”包括蛔虫、犬恶丝虫和植食性线虫(线虫纲)、吸虫(吸虫纲)、棘头虫、和绦虫(绦虫纲)。原生动物感染包括来自贾第虫属,毛滴虫属、非洲锥虫病、阿米巴痢疾、巴贝虫病、小袋虫性痢疾、查加斯病(Chaga's disease)、球虫病、疟疾和弓形体病的感染。病原体(诸如寄生/原生动物病原体)的实例包括但不限于:恶性疟原虫(*Plasmodium falciparum*)、间日疟原虫(*Plasmodium vivax*)、克氏锥虫(*Trypanosoma cruzi*)和刚地弓形虫(*Toxoplasma gondii*)。真菌病原体包括但不限于:新型隐球菌(*Cryptococcus neoformans*)、荚膜组织胞浆菌(*Histoplasma capsulatum*)、粗球孢菌(*Coccidioides immitis*)、皮炎芽生菌(*Blastomyces dermatitidis*)、沙眼衣原体(*Chlamydia trachomatis*)和白色念珠菌(*Candida albicans*)。病原性病毒包括例如免疫缺陷病毒(例如,HIV);流感病毒;登革热病毒;西尼罗河病毒;疱疹病毒;黄热病毒;肝炎病

毒C;肝炎病毒A;肝炎病毒B;乳头瘤病毒;等等。病原体包括例如HIV病毒、结核分枝杆菌(*Mycobacterium tuberculosis*)、无乳链球菌(*Streptococcus agalactiae*)、耐甲氧西林金黄色葡萄球菌(*methicillin-resistant Staphylococcus aureus*)、嗜肺军团菌(*Legionella pneumophila*)、酿脓链球菌(*Streptococcus pyogenes*)、大肠杆菌(*Escherichia coli*)、淋病奈瑟氏菌(*Neisseria gonorrhoeae*)、脑膜炎奈瑟氏菌(*Neisseria meningitidis*)、肺炎球菌(*Pneumococcus*)、新型隐球菌(*Cryptococcus neoformans*)、荚膜组织胞浆菌(*Histoplasma capsulatum*)、流感嗜血杆菌B(*Hemophilus influenzae B*)、梅毒密螺旋体(*Treponema pallidum*)、莱姆病螺旋体(*Lyme disease spirochetes*)、铜绿假单胞菌(*Pseudomonas aeruginosa*)、麻风分枝杆菌(*Mycobacterium leprae*)、流产布鲁氏菌(*Brucella abortus*)、狂犬病病毒(*rabies virus*)、流感病毒(*influenza virus*)、巨细胞病毒(*cytomegalovirus*)、单纯疱疹病毒I(*herpes simplex virus I*)、单纯疱疹病毒II(*herpes simplex virus II*)、人血清细小样病毒(*human serum parvo-like virus*)、呼吸道合胞体病毒(*respiratory syncytial virus*)、水痘-带状疱疹病毒(*varicella-zoster virus*)、乙型肝炎病毒(*hepatitis B virus*)、丙型肝炎病毒(*hepatitis C virus*)、麻疹病毒(*measles virus*)、腺病毒(*adenovirus*)、人T细胞白血病病毒(*human T-cell leukemia viruses*)、艾普斯登-巴尔病毒(*Epstein-Barr virus*)、鼠白血病病毒(*murine leukemia virus*)、腮腺炎病毒(*mumps virus*)、水疱性口炎病毒(*vesicular stomatitis virus*)、辛德毕斯病毒(*Sindbis virus*)、淋巴细胞脉络丛脑膜炎病毒(*lymphocytic choriomeningitis virus*)、疣病毒(*wart virus*)、蓝舌病毒(*blue tongue virus*)、仙台病毒(*Sendai virus*)、猫白血病病毒(*feline leukemia virus*)、呼肠孤病毒(*Reovirus*)、脊髓灰质炎病毒(*polio virus*)、猿猴病毒40(*simian virus 40*)、小鼠乳腺肿瘤病毒(*mouse mammary tumor virus*)、登革热病毒(*dengue virus*)、风疹病毒(*rubella virus*)、西尼罗河病毒(*West Nile virus*)、恶性疟原虫(*Plasmodium falciparum*)、间日疟原虫(*Plasmodium vivax*)、刚地弓形虫(*Toxoplasma gondii*)、蓝氏锥虫(*Trypanosoma rangeli*)、克氏锥虫(*Trypanosoma cruzi*)、罗氏锥虫(*Trypanosoma rhodesiense*)、布氏锥虫(*Trypanosoma brucei*)、曼氏血吸虫(*Schistosoma mansoni*)、日本血吸虫(*Schistosoma japonicum*)、巴贝斯虫(*Babesia bovis*)、柔嫩艾美球虫(*Eimeria tenella*)、盘尾丝虫(*Onchocerca volvulus*)、利什曼原虫(*Leishmania tropica*)、结核分枝杆菌(*Mycobacterium tuberculosis*)、旋毛虫(*Trichinella spiralis*)、泰勒原虫(*Theileria parva*)、胞状绦虫(*Taenia hydatigena*)、羊绦虫(*Taenia ovis*)、牛肉绦虫(*Taenia saginata*)、细粒棘球绦虫(*Echinococcus granulosus*)、柯氏中殖孔绦虫(*Mesocostoides corti*)、关节炎支原体(*Mycoplasma arthritidis*)、猪鼻支原体(*M.hyorhinis*)、口腔支原体(*M. orale*)、精氨酸支原体(*M.arginini*)、莱氏无胆甾原体(*Acholeplasma laidlawii*)、唾囊支原体(*M.salivarium*)和肺炎支原体(*M.pneumoniae*)。

[0129] 靶RNA

[0130] 靶RNA可以是任何单链RNA(ssRNA)。实例包括但不限于mRNA、rRNA、tRNA、非编码RNA(ncRNA)、长非编码RNA(lncRNA)和微RNA(miRNA)。在一些情况下,靶ssRNA是mRNA。在一些情况下,单链靶核酸是来自病毒(例如,寨卡病毒、人免疫缺陷病毒、流感病毒等)的ssRNA。在一些情况下,单链靶核酸是寄生虫的ssRNA。在一些情况下,单链靶核酸是细菌(例

如,病原细菌)的ssRNA。如上所述,靶RNA的来源可以与RNA样品的来源相同。

[0131] 测量可检测信号

[0132] 在一些情况下,本方法包括测量步骤(例如,测量由C2c2蛋白介导的RNA切割产生的可检测信号)。因为C2c2蛋白一旦被激活就切割非靶向RNA(所述激活当在存在C2c2蛋白的情况下C2c2指导RNA与靶RNA杂交时发生),所以可检测信号可以是当RNA被切割时产生的任何信号。例如,在一些情况下,测量步骤可以包括以下中的一者或多者:基于金纳米颗粒的检测(例如参见Xu等人,Angew Chem Int Ed Engl.2007;46(19):3468-70;和Xia等人,Proc Natl Acad Sci U S A.2010年6月15日;107(24):10837-41)、荧光偏振、胶体相变/分散(例如,Baksh等人,Nature.2004年1月8日;427(6970):139-41)、电化学检测、基于半导体的感测(例如,Rothberg等人,Nature.2011年7月20日;475(7356):348-52;例如,可以在RNA切割反应后使用磷酸酶以通过打开2'-3'环状磷酸酯并通过将无机磷酸盐释放到溶液中而产生pH变化),以及检测经标记的检测RNA(更多细节参见下文)。这种检测方法的读出可以是任何合宜的读出。可能的读出的实例包括但不限于:所测量的可检测荧光信号的量;对凝胶上条带(例如,表示切割产物对比未切割底物的条带)的视觉分析、基于视觉或传感器的对颜色存在与否的检测(即,颜色检测方法),以及是否存在电信号(或电信号的特定量)。

[0133] 在一些情况下,例如在检测到的信号量可以用于确定样品中存在的靶RNA的量的意义上,测量可以是定量的。在一些情况下,例如在可检测信号的存在或不存在可以指示靶RNA的存在或不存在的意义上,测量可以是定性的。在一些情况下,除非靶RNA以高于特定阈值浓度存在(例如,参见图5),否则可检测信号将不存在(例如,高于给定阈值水平)。在一些情况下,可以通过修改C2c2蛋白、指导RNA、样品体积和/或检测RNA(如果使用的话)的量来滴定检测阈值。因此,例如,如本领域的普通技术人员应当理解的,如果需要的话,可以使用许多对照以建立一个或多个反应,每个反应设置用于检测靶RNA的不同阈值水平,并且因此可以使用这样的一系列反应来确定样品中存在的靶RNA的量(例如,可以使用这样的一系列反应来确定靶RNA‘以至少X的浓度’存在于样品中)。

[0134] 经标记的检测RNA

[0135] 在一些情况下,本方法包括使样品(例如,包含靶RNA和多种非靶RNA的样品)与以下接触:i)经标记的检测RNA;ii)C2c2蛋白;以及iii)C2c2指导RNA(或前体C2c2指导RNA阵列)。例如,在一些情况下,本方法包括使样品与经标记的检测RNA接触,所述经标记的检测RNA包含荧光发射染料对;C2c2蛋白在被激活(通过在指导RNA与靶RNA杂交的背景下与C2c2指导RNA结合)后切割所述经标记的检测RNA;并且所测量到的可检测信号是由荧光发射染料对产生的。例如,在一些情况下,本方法包括使样品与经标记的检测RNA接触,所述经标记的检测RNA包含荧光共振能量转移(FRET)对或猝灭剂/荧光剂对或两者。在一些情况下,本方法包括使样品与包含FRET对的经标记的检测RNA接触。在一些情况下,本方法包括使样品与包含荧光剂/猝灭剂对的经标记的检测RNA接触。荧光发射染料对包含FRET对或猝灭剂/荧光剂对。在FRET对和猝灭剂/荧光剂对的两种情况下,染料中的一种染料的发射光谱与所述对中的另一种染料的吸收光谱的某一区域重叠。如本文所用的,术语“荧光发射染料对”是用于包括“荧光共振能量转移(FRET)对”和“猝灭剂/荧光剂对”两者的通用术语,这两个术语将在下面更详细地讨论。术语“荧光发射染料对”与短语“FRET对和/或猝灭剂/荧光剂对”可互换地使用。

[0136] 在一些情况下(例如,当检测RNA包括FRET对时),经标记的检测RNA在被切割之前产生一定量的可检测信号,并且当所述经标记的检测RNA被切割时所测量到的可检测信号的量减少。在一些情况下,经标记的检测RNA在被切割之前产生第一可检测信号(例如,来自FRET对),并且当所述经标记的检测RNA被切割时产生第二可检测信号(例如,来自猝灭剂/荧光剂对)。因此,在一些情况下,经标记的检测RNA包含FRET对和猝灭剂/荧光剂对。

[0137] 在一些情况下,经标记的检测RNA包括FRET对。FRET是这样的过程,通过所述过程,能量的无辐射转移发生为从激发态荧光团到紧邻的第二发色团。能量转移可以发生的范围限制在约10纳米(100埃),并且转移效率对荧光团之间的分开距离非常敏感。因此,如本文所使用的,术语“FRET”(“荧光共振能量转移”;也被称为“萤光共振能量转移”)是指这样的物理现象,所述物理现象涉及供体荧光团和匹配的受体荧光团被选择为使得供体的发射光谱与受体的激发光谱重叠,并且被进一步选择为使得当供体和受体彼此非常接近(通常为10nm或更短距离)时,供体的激发将引起来自受体的激发和发射,这是因为一些能量经由量子耦合效应从供体传递到受体。因此,FRET信号用作供体和受体的接近度尺度;只有当它们彼此非常接近时才会产生信号。FRET供体部分(例如,供体荧光团)和FRET受体部分(例如,受体荧光团)在本文中统称为“FRET对”。

[0138] 供体-受体对(FRET供体部分和FRET受体部分)在本文中被称为“FRET对”或“信号FRET对”。因此,在一些情况下,主题经标记的检测RNA包括两个信号配偶体(信号对),当一个信号配偶体是FRET供体部分时,另一个信号配偶体是FRET受体部分。因此,当信号配偶体非常接近时(例如,在相同的RNA分子上时),包括这种FRET对(FRET供体部分和FRET受体部分)的主题经标记的检测RNA将显示可检测信号(FRET信号),但是当所述配偶体分开时(例如,在用C2c2蛋白切割RNA分子后),信号将减少(或不存在)。

[0139] FRET供体和受体部分(FRET对)将是本领域的普通技术人员已知的,并且可以使用任何合宜的FRET对(例如,任何合宜的供体和受体部分对)。合适的FRET对的实例包括但不限于表1中所示的那些。还参见:Bajar等人,Sensors (Basel) .2016年9月14日;16(9);和Abraham等人,PLoS One.2015年8月3日;10(8):e0134436。

[0140] 表1.FRET对的实例(供体和受体FRET部分)

[0141]

供体	受体
色氨酸	丹酰
IAEDANS (1)	DDPM (2)
BFP	DsRFP
丹酰	异硫氰酸荧光素 (FITC)
丹酰	十八烷基罗丹明
青色荧光蛋白 (CFP)	绿色荧光蛋白 (GFP)
CF (3)	德克萨斯红
荧光素	四甲基罗丹明
Cy3	Cy5
GFP	黄色荧光蛋白 (YFP)
BODIPY FL(4)	BODIPY FL(4)
罗丹明 110	Cy3
罗丹明 6G	孔雀石绿
FITC	曙红氨基硫脲
B-藻红蛋白	Cy5
Cy5	Cy5.5

[0142] (1) 5-(2-碘乙酰基氨基乙基)氨基萘-1-磺酸

[0143] (2) N-(4-二甲基氨基-3,5-二硝基苯基)马来酰亚胺

[0144] (3) 羧基荧光素琥珀酰亚胺酯

[0145] (4) 4,4-二氟-4-硼杂-3a,4a-二氮杂-s-吡啶生

[0146] 在一些情况下,可检测信号在经标记的检测RNA被切割时产生(例如,在一些情况下,经标记的检测RNA包含猝灭剂/荧光剂对。信号猝灭对的一个信号配偶体产生可检测信号,而另一个信号配偶体是猝灭剂部分,所述猝灭剂部分猝灭第一信号配偶体的可检测信号(即,猝灭剂部分猝灭信号部分的信号,使得当信号配偶体彼此接近时(例如,当信号对的信号配偶体非常接近时),来自信号部分的信号减少(被猝灭)。

[0147] 例如,在一些情况下,当经标记的检测RNA被切割时,可检测信号的量增加。例如,在一些情况下,由一个信号配偶体(信号部分)表现的信号被另一个信号配偶体(猝灭剂信号部分)猝灭,例如当在由C2c2蛋白进行切割之前两者存在于相同的RNA分子上时。这种信号对在本文中称为“猝灭剂/荧光剂对”、“猝灭对”或“信号猝灭对”。例如,在一些情况下,一个信号配偶体(例如,第一信号配偶体)是产生可检测信号的信号部分,所述可检测信号由第二信号配偶体(例如,猝灭剂部分)猝灭。因此,当配偶体被分开时(例如,在由C2c2蛋白切割检测RNA之后),这种猝灭剂/荧光剂对的信号配偶体将产生可检测信号,但是当配偶体非常接近(例如,在由C2c2蛋白切割检测RNA之前)时,所述信号将被猝灭。

[0148] 猝灭剂部分可以将来自信号部分的信号(例如,在由C2c2蛋白切割检测RNA之前)猝灭至不同程度。在一些情况下,猝灭剂部分猝灭来自信号部分的信号,其中在存在猝灭剂部分的情况下(当信号配偶体彼此接近时)检测到的信号是在不存在猝灭剂部分(当信号配偶体分开时)时检测到的信号的95%或更少。例如,在一些情况下,在存在猝灭剂部分的情况下检测到的信号可以是在不存在猝灭剂部分的情况下检测到的信号的90%或更少、80%

或更少、70%或更少、60%或更少、50%或更少、40%或更少、30%或更少、20%或更少、15%或更少、10%或更少、或5%或更少。在一些情况下,在存在猝灭剂部分的情况下未检测到信号(例如,高于背景)。

[0149] 在一些情况下,在不存在猝灭剂部分的情况下(当信号配偶体分开时)检测到的信号是在存在猝灭剂部分的情况下(当信号配偶体彼此接近时)检测到的信号的至少1.2倍(例如,至少1.3倍、至少1.5倍、至少1.7倍、至少2倍、至少2.5倍、至少3倍、至少3.5倍、至少4倍、至少5倍、至少7倍、至少10倍、至少20倍或至少50倍)。

[0150] 在一些情况下,信号部分是荧光标记。在一些此类情况下,猝灭剂部分猝灭来自荧光标记的信号(光信号)(例如,通过吸收标记的发射光谱中的能量)。因此,当猝灭剂部分不与信号部分接近时,来自荧光标记的发射(信号)是可检测的,因为信号不被猝灭剂部分吸收。可以使用任何合宜的供体受体对(信号部分/猝灭剂部分对),并且许多合适的对是本领域中已知的。

[0151] 在一些情况下,猝灭剂部分从信号部分(在本文中也称为“可检测标记”)吸收能量,然后发射信号(例如,不同波长的光)。因此,在一些情况下,猝灭剂部分本身是信号部分(例如,信号部分可以是6-羧基荧光素,而猝灭剂部分可以是6-羧基-四甲基罗丹明),并且在一些此类情况下,所述对也可以是FRET对。在一些情况下,猝灭剂部分是暗猝灭剂。暗猝灭剂可以吸收激发能量并以不同方式耗散能量(例如,作为热量)。因此,暗猝灭剂本身具有极小荧光至没有荧光(不发射荧光)。暗猝灭剂的实例进一步描述于美国专利号8,822,673和8,586,718;美国专利公开20140378330、20140349295和20140194611;以及国际专利申请:W0200142505和W0200186001中,所有这些专利都据此以引用方式整体并入本文。

[0152] 荧光标记的实例包括但不限于:Alexa**Fluor**[®]染料、ATTO染料(例如,ATTO 390、ATTO 425、ATTO 465、ATTO 488、ATTO495、ATTO 514、ATTO 520、ATTO 532、ATTO Rho6G、ATTO542、ATTO 550、ATTO 565、ATTO Rho3B、ATTO Rho11、ATTO Rho12、ATTO Thio12、ATTO Rho101、ATTO 590、ATTO 594、ATTO Rho13、ATTO 610、ATTO 620、ATTO Rho14、ATTO 633、ATTO 647、ATTO 647N、ATTO 655、ATTO Oxa12、ATTO 665、ATTO 680、ATTO 700、ATTO 725、ATTO 740)、DyLight染料、花青染料(例如,Cy2、Cy3、Cy3.5、Cy3b、Cy5、Cy5.5、Cy7、Cy7.5)、FluoProbes染料、Sulfo Cy染料、Seta染料、IRIS染料、SeTau染料、SRfluor染料、方形染料(Square dye)、异硫氰酸荧光素(FITC)、四甲基罗丹明(TRITC)、德克萨斯红、俄勒冈绿、太平洋蓝、太平洋绿、太平洋橙、量子点和束缚荧光蛋白(tethered fluorescent protein)。

[0153] 在一些情况下,可检测标记是选自以下的荧光标记:Alexa**Fluor**[®]染料、ATTO染料(例如,ATTO 390、ATTO 425、ATTO465、ATTO 488、ATTO 495、ATTO 514、ATTO 520、ATTO 532、ATTO Rho6G、ATTO 542、ATTO 550、ATTO 565、ATTO Rho3B、ATTO Rho11、ATTO Rho12、ATTO Thio12、ATTO Rho101、ATTO 590、ATTO 594、ATTO Rho13、ATTO 610、ATTO620、ATTO Rho14、ATTO 633、ATTO 647、ATTO 647N、ATTO655、ATTO Oxa12、ATTO 665、ATTO 680、ATTO 700、ATTO725、ATTO 740)、DyLight染料、花青染料(例如,Cy2、Cy3、Cy3.5、Cy3b、Cy5、Cy5.5、Cy7、Cy7.5)、FluoProbes染料、Sulfo Cy染料、Seta染料、IRIS染料、SeTau染料、SRfluor染料、方形染料、异硫氰酸荧光素(FITC)、四甲基罗丹明(TRITC)、德克萨斯红、俄勒冈绿、太平洋蓝、太平洋绿和太平洋橙。

[0154] 在一些情况下,可检测标记是选自以下的荧光标记:Alexa**Fluor**[®]染料、ATTO染料

(例如, ATTO 390、ATTO 425、ATTO465、ATTO 488、ATTO 495、ATTO 514、ATTO 520、ATTO 532、ATTO Rho6G、ATTO 542、ATTO 550、ATTO 565、ATTO Rho3B、ATTO Rho11、ATTO Rho12、ATTO Thio12、ATTO Rho101、ATTO 590、ATTO 594、ATTO Rho13、ATTO 610、ATTO620、ATTO Rho14、ATTO 633、ATTO 647、ATTO 647N、ATTO655、ATTO Oxa12、ATTO 665、ATTO 680、ATTO 700、ATTO725、ATTO 740)、DyLight染料、花青染料(例如,Cy2、Cy3、Cy3.5、Cy3b、Cy5、Cy5.5、Cy7、Cy7.5)、FluoProbes染料、Sulfo Cy染料、Seta染料、IRIS染料、SeTau染料、SRfluor染料、方形染料、异硫氰酸荧光素(FITC)、四甲基罗丹明(TRITC)、德克萨斯红、俄勒冈绿、太平洋蓝、太平洋绿、太平洋橙、量子点和束缚荧光蛋白。

[0155] ATTO染料的实例包括但不限于:ATTO 390、ATTO 425、ATTO 465、ATTO 488、ATTO 495、ATTO 514、ATTO 520、ATTO 532、ATTO Rho6G、ATTO 542、ATTO 550、ATTO 565、ATTO Rho3B、ATTO Rho11、ATTO Rho12、ATTO Thio12、ATTO Rho101、ATTO 590、ATTO 594、ATTO Rho13、ATTO 610、ATTO620、ATTO Rho14、ATTO 633、ATTO 647、ATTO 647N、ATTO655、ATTO Oxa12、ATTO 665、ATTO 680、ATTO 700、ATTO725、以及ATTO 740。

[0156] AlexaFluor染料的实例包括但不限于:Alexa**Fluor**[®] 350、Alexa**Fluor**[®] 405、Alexa**Fluor**[®] 430、Alexa**Fluor**[®] 488、Alexa**Fluor**[®] 500、Alexa**Fluor**[®] 514、Alexa**Fluor**[®] 532、Alexa**Fluor**[®] 546、Alexa**Fluor**[®] 555、Alexa**Fluor**[®] 568、Alexa**Fluor**[®] 594、Alexa**Fluor**[®] 610、Alexa**Fluor**[®] 633、Alexa**Fluor**[®] 635、Alexa**Fluor**[®] 647、Alexa**Fluor**[®] 660、Alexa**Fluor**[®] 680、Alexa**Fluor**[®] 700、Alexa**Fluor**[®] 750、Alexa**Fluor**[®] 790等等。

[0157] 猝灭剂部分的实例包括但不限于:暗猝灭剂、Black Hole**Quencher**[®] (BHQ[®]) (例如,BHQ-0、BHQ-1、BHQ-2、BHQ-3)、Qx1猝灭剂、ATTO猝灭剂(例如,ATTO 540Q、ATTO 580Q和ATTO612Q)、二甲氨基偶氮苯磺酸(Dabsyl)、Iowa Black RQ、Iowa Black FQ、IRDye QC-1、QSY染料(例如,QSY 7、QSY 9、QSY 21)、AbsoluteQuencher、Eclipse和金属簇(诸如金纳米粒子)等。

[0158] 在一些情况下,猝灭剂部分选自:暗猝灭剂、Black Hole**Quencher**[®] (BHQ[®]) (例如,BHQ-0、BHQ-1、BHQ-2、BHQ-3)、Qx1猝灭剂、ATTO猝灭剂(例如,ATTO 540Q、ATTO 580Q和ATTO612Q)、二甲氨基偶氮苯磺酸(Dabsyl)、Iowa Black RQ、Iowa Black FQ、IRDye QC-1、QSY染料(例如,QSY 7、QSY 9、QSY 21)、AbsoluteQuencher、Eclipse和金属簇。

[0159] ATTO猝灭剂的实例包括但不限于:ATTO 540Q、ATTO 580Q和ATTO 612Q。Black Hole**Quencher**[®] (BHQ[®])的实例包括但不限于: BHQ-0 (493nm)、BHQ-1 (534nm)、BHQ-2 (579nm) 和BHQ-3 (672nm)。

[0160] 对于一些可检测标记(例如,荧光染料)和/或猝灭剂部分的实例,参见例如Bao等人,Annu Rev Biomed Eng.2009;11:25-47;以及美国专利号8,822,673和8,586,718;美国专利公开20140378330、20140349295、20140194611、20130323851、20130224871、20110223677、20110190486、20110172420、20060179585和20030003486;以及国际专利申请:W0200142505和W0200186001,所有这些参考文献据此以引用方式整体并入。

[0161] 在一些情况下,可以通过测量比色读出来检测对经标记的检测RNA的切割。例如,

荧光团的释放(例如,从FRET对释放、从猝灭剂/荧光剂对释放等)可导致可检测信号的波长偏移(并因此色移)。因此,在一些情况下,可以通过色移检测主题经标记的检测RNA的切割。这种偏移可以表示为一种颜色(波长)信号的量损失、另一种颜色的量的增益、一种颜色与另一种颜色的比率的变化等。

[0162] 核酸修饰

[0163] 在一些情况下,经标记的检测RNA包含一种或多种修饰,例如碱基修饰、骨架修饰、糖修饰等,以为核酸提供新的或增强的特征(例如,改善的稳定性)。如本领域中已知的,核苷是碱基-核糖组合。核苷的碱基部分通常是杂环碱基。此类杂环碱基的两种最常见的类别是嘌呤和嘧啶。核苷酸是这样的核苷,所述核苷还包括与核苷的糖部分共价连接的磷酸基团。对于那些包括戊呋喃糖基糖的核苷,磷酸基团可以与所述糖的2'、3'或5'羟基部分连接。在形成寡核苷酸时,磷酸基团将相邻的核苷彼此共价连接以形成直链聚合化合物。继而,该直链聚合化合物的各个末端可以进一步连接以形成环状化合物,然而,直链化合物通常是合适的。此外,线性化合物可以具有内部核苷酸碱基互补性,并且因此可以以产生完全或部分双链化合物的方式折叠。在寡核苷酸内,磷酸基团通常被称为形成寡核苷酸的核苷间骨架。RNA和DNA的正常键或骨架是3'至5'磷酸二酯键。

[0164] 经修饰的骨架和经修饰的核苷间键

[0165] 合适的修饰的实例包括经修饰的核酸骨架和非天然的核苷间键。具有经修饰的骨架的核酸包括在骨架中保留磷原子的核酸和在骨架中不具有磷原子的核酸。

[0166] 其中含有磷原子的合适的经修饰寡核苷酸骨架包含例如硫代磷酸酯、手性硫代磷酸酯、二硫代磷酸酯、磷酸三酯、氨基烷基磷酸三酯、甲基和其它烷基磷酸酯(包括3'-亚烷基磷酸酯、5'-亚烷基磷酸酯和手性磷酸酯)、次磷酸酯、氨基磷酸酯(包括3'-氨基氨基磷酸酯和氨基烷基氨基磷酸酯)、二氨基磷酸酯、硫代氨基磷酸酯、硫代烷基磷酸酯、硫代烷基磷酸三酯、具有正常3'-5'键的硒代磷酸酯和硼代磷酸酯、硒代磷酸酯和硼代磷酸酯的2'-5'连接的类似物,以及具有反向极性的硒代磷酸酯和硼代磷酸酯,其中一个或多个核苷酸间键是3'至3'键、5'至5'键或2'至2'键。合适的具有反向极性的寡核苷酸包含在最3'核苷酸间键处的单个3'至3'键,即单个反向核苷残基,所述单个反向核苷残基可以是碱性的(核碱基丢失或具有羟基代替其)。还包括各种盐(诸如,钾盐或钠盐)、混合盐和游离酸形式。

[0167] 在一些情况下,经标记的检测RNA包含一个或多个硫代磷酸酯和/或杂原子核苷间键,特别是 $-\text{CH}_2-\text{NH}-\text{O}-\text{CH}_2-$ 、 $-\text{CH}_2-\text{N}(\text{CH}_3)-\text{O}-\text{CH}_2-$ (称为亚甲基(甲基亚氨基)或MMI骨架)、 $-\text{CH}_2-\text{O}-\text{N}(\text{CH}_3)-\text{CH}_2-$ 、 $-\text{CH}_2-\text{N}(\text{CH}_3)-\text{N}(\text{CH}_3)-\text{CH}_2-$ 和 $-\text{O}-\text{N}(\text{CH}_3)-\text{CH}_2-\text{CH}_2-$ (其中天然磷酸二酯核苷酸间键表示为 $-\text{O}-\text{P}(=\text{O})(\text{OH})-\text{O}-\text{CH}_2-$)。MMI型核苷间键在上面引用的美国专利号5,489,677中公开。合适的酰胺核苷间键在美国专利号5,602,240中公开。

[0168] 同样合适的是如在例如美国专利号5,034,506中所述的具有吗啉代骨架结构的核酸。例如,在一些情况下,经标记的检测RNA包含6元吗啉代环代替核糖环。在一些情况下,二氨基磷酸酯或其它非磷酸二酯核苷间键取代磷酸二酯键。

[0169] 其中不含磷原子的合适的经修饰多核苷酸骨架具有由短链烷基或环烷基核苷间键、混合杂原子和烷基或环烷基核苷间键、或一个或多个短链杂原子或杂环核苷间键形成的骨架。这些骨架包括具有吗啉代键(部分地由核苷的糖部分形成)的那些骨架;硅氧烷骨架;硫化物、亚砷和砷骨架;甲酰基和硫代甲酰基骨架;亚甲基甲酰基和硫代甲酰基骨架;核

糖乙酰骨架;含有烯基的骨架;氨基磺酸酯骨架;亚甲基亚氨基和亚甲基胍基骨架;磺酸酯和磺酰胺骨架;酰胺骨架;以及其它具有混合的N、O、S和CH₂组成部分的骨架。

[0170] 模拟物

[0171] 经标记的检测RNA可以是核酸模拟物。应用于多核苷酸的术语“模拟物”旨在包括这样的多核苷酸,在所述多核苷酸中仅呋喃糖环或呋喃糖环和核苷酸间键两者被非呋喃糖基团替代,仅替代呋喃糖环也在本领域中被称作糖替代物。维持杂环碱基部分或经修饰的杂环碱基部分以与合适的靶核酸杂交。一种此类核酸,已显示具有优异杂交特性的多核苷酸模拟物,被称为肽核酸(PNA)。在PNA中,多核苷酸的糖骨架被含酰胺的骨架,特别是氨基乙基甘氨酸骨架取代。核苷酸被保留并直接或间接地结合骨架的酰胺部分的氮杂氮原子。

[0172] 一种据报道具有优异杂交特性的多核苷酸模拟物是肽核酸(PNA)。PNA化合物中的骨架是两个或更多个连接的氨基乙基甘氨酸单元,所述氨基乙基甘氨酸单元赋予PNA含酰胺的骨架。杂环碱基部分直接或间接地与骨架的酰胺部分的氮杂氮原子结合。描述PNA化合物制备的代表性美国专利包括但不限于:美国专利号5,539,082;5,714,331;和5,719,262。

[0173] 已经研究的另一类多核苷酸模拟物基于连接的吗啉代单元(吗啉代核酸),所述连接的吗啉代单元具有附接到吗啉代环的杂环碱基。已经报道了许多连接基团,所述连接基团连接吗啉代核酸中的吗啉代单体单元。一类连接基团已经被选择用于提供非离子寡聚化合物。非离子的基于吗啉代的寡聚化合物不太可能与细胞蛋白发生不希望的相互作用。基于吗啉代的多核苷酸是寡核苷酸的非离子模拟物,其不太可能与细胞蛋白形成不希望的相互作用(Dwaine A.Braasch和David R.Corey,Biochemistry,2002,41(14),4503-4510)。基于吗啉代的多核苷酸在美国专利号5,034,506中公开。已经制备了吗啉代类多核苷酸内的多种化合物,所述化合物具有各种连接单体亚基的不同连接基团。

[0174] 另一类多核苷酸模拟物被称为环己烯基核酸(CeNA)。通常存在于DNA/RNA分子中的呋喃糖环被环己烯基环替代。CeNA DMT保护的亚磷酰胺单体已被制备并用于按照经典的亚磷酰胺化学进行寡聚化合物合成。已经制备并研究了完全修饰的CeNA寡聚化合物和具有用CeNA修饰的特定位置的寡核苷酸(参见Wang等人,J.Am.Chem.Soc.,2000,122,8595-8602)。通常,将CeNA单体掺入DNA链中增加了其DNA/RNA杂交体的稳定性。CeNA寡腺苷酸以与天然复合物相似的稳定性而与RNA和DNA互补物形成复合物。通过NMR和圆二色性显示将CeNA结构掺入天然核酸结构中的研究,以进行简单的构象适应。

[0175] 另一修饰包括锁定核酸(LNA),其中2'-羟基与糖环的4'碳原子连接,从而形成2'-C,4'-C-氧亚甲基键,从而形成双环糖部分。键可以是亚甲基(-CH₂-)、桥接2'氧原子与4'碳原子的基团,其中n是1或2(Singh等人,Chem.Comm.,1998,4,455-456)。LNA和LNA类似物显示出非常高的互补DNA和RNA的双链热稳定性(T_m=+3至+10℃),朝向3'-核酸外切降解的稳定性和良好的溶解性。已经描述了含有LNA的有效和无毒的反义寡核苷酸(Wahlestedt等人,Proc.Natl.Acad.Sci.U.S.A.,2000,97,5633-5638)。

[0176] 已经描述了LNA单体腺嘌呤、胞嘧啶、鸟嘌呤、5-甲基胞嘧啶、胸腺嘧啶和尿嘧啶的合成和制备,以及它们的寡聚化和核酸识别性质(Koshkin等人,Tetrahedron,1998,54,3607-3630)。LNA及其制备还描述于WO 98/39352和WO 99/14226中。

[0177] 经修饰的糖部分

[0178] 经标记的检测RNA还可包括一个或多个经取代的糖部分。合适的多核苷酸包含选

自以下的糖取代基:OH;F;O-烷基、S-烷基或N-烷基;O-烯基、S-烯基或N-烯基;O-炔基、S-炔基或N-炔基;或O-烷基-O-烷基,其中所述烷基、烯基和炔基可以是取代或未取代的C₁至C₁₀烷基或C₂至C₁₀烯基和炔基。特别合适的是O((CH₂)_n)_mCH₃、O(CH₂)_nOCH₃、O(CH₂)_nNH₂、O(CH₂)_nCH₃、O(CH₂)_nONH₂以及O(CH₂)_nON((CH₂)_nCH₃)₂,其中n和m为1至约10。其它合适的多核苷酸包含选自以下的糖取代基:C₁至C₁₀低级烷基,取代的低级烷基、烯基、炔基,烷芳基,芳烷基,0-烷芳基或0-芳烷基,SH,SCH₃,OCN,Cl,Br,CN,CF₃,OCF₃,SOCH₃,SO₂CH₃,ONO₂,NO₂,N₃,NH₂,杂环烷基,杂环烷芳基,氨基烷基氨基,聚烷基氨基,取代的甲硅烷基,RNA切割基团,报告基团,嵌入剂,用于改善寡核苷酸的药代动力学性质的基团,或用于改善寡核苷酸的药效学性质的基团,以及其它具有相似性质的取代基。合适的修饰包括2'-甲氧基乙氧基(2'-O-CH₂CH₂OCH₃,也称为2'-O-(2-甲氧基乙基)或2'-MOE)(Martin等人,Helv.Chim.Acta,1995,78,486-504),即烷氧基烷氧基基团。另一合适的修饰包括2'-二甲基氨氧乙氧基,即O(CH₂)₂ON(CH₃)₂基团,也称为2'-DMAOE,如下文实施例中所述;以及2'-二甲氨基乙氧基乙氧基(在本领域中也称为2'-O-二甲基-氨基-乙氧基-乙基或2'-DMAEOE),即2'-O-CH₂-O-CH₂-N(CH₃)₂。

[0179] 其它合适的糖取代基包括甲氧基(-O-CH₃)、氨基丙氧基(-O-CH₂CH₂CH₂NH₂)、烯丙基(-CH₂-CH=CH₂)、-O-烯丙基(-O-CH₂-CH=CH₂)和氟(F)。2'-糖取代基可以处于阿拉伯糖(上)位置或核糖(下)位置。合适的2'-阿拉伯糖修饰是2'-F。也可以在寡聚化合物上的其它位置,特别是3'末端核苷上或2'-5'连接的寡核苷酸的糖的3'位置处,以及5'末端核苷酸的5'位置处进行类似的修饰。寡聚化合物也可以具有糖模拟物,诸如环丁基部分代替戊呋喃糖基糖。

[0180] 碱基修饰和取代

[0181] 经标记的检测RNA还可以包括核碱基(在本领域中通常简称为“碱基”)修饰或取代。如本文所用,“未修饰的”或“天然的”核碱基包括嘌呤碱基腺嘌呤(A)和鸟嘌呤(G),以及嘧啶碱基胸腺嘧啶(T)、胞嘧啶(C)和尿嘧啶(U)。经修饰的核碱基包括其它合成和天然核碱基,诸如5-甲基胞嘧啶(5-me-C)、5-羟甲基胞嘧啶、黄嘌呤、次黄嘌呤、2-氨基腺嘌呤、6-甲基腺嘌呤和鸟嘌呤以及腺嘌呤和鸟嘌呤的其它烷基衍生物、2-丙基腺嘌呤和鸟嘌呤以及腺嘌呤和鸟嘌呤的其它烷基衍生物、2-硫尿嘧啶、2-硫代胸腺嘧啶和2-硫代胞嘧啶、5-氟尿嘧啶和胞嘧啶、5-丙炔基(-C≡C-CH₃)尿嘧啶和胞嘧啶以及嘧啶碱基的其它炔基衍生物、6-偶氮尿嘧啶、胞嘧啶和胸腺嘧啶、5-尿嘧啶(假尿嘧啶)、4-硫尿嘧啶、8-卤代腺嘌呤和鸟嘌呤、8-氨基腺嘌呤和鸟嘌呤、8-巯基腺嘌呤和鸟嘌呤、8-硫代烷基腺嘌呤和鸟嘌呤、8-羟基腺嘌呤和鸟嘌呤以及其它8-取代的腺嘌呤和鸟嘌呤、5-卤代(特别是5-溴代)尿嘧啶和胞嘧啶、5-三氟甲基尿嘧啶和胞嘧啶以及其它5-取代的尿嘧啶和胞嘧啶、7-甲基鸟嘌呤和7-甲基腺嘌呤、2-F-腺嘌呤、2-氨基腺嘌呤、8-氮杂鸟嘌呤和8-氮杂腺嘌呤、7-脱氮杂鸟嘌呤和7-脱氮杂腺嘌呤,以及3-脱氮杂鸟嘌呤和3-脱氮杂腺嘌呤。其它经修饰的核碱基包括三环嘧啶,诸如吩噻嗪胞苷(1H-嘧啶并(5,4-b)(1,4)苯并噻嗪-2(3H)-酮)、吩噻嗪胞苷(1H-嘧啶(5,4-b)(1,4)苯并噻嗪-2(3H)-酮)、G-钳如取代的吩噻嗪胞苷(例如9-(2-氨基乙氧基)-H-嘧啶并(5,4-b)(1,4)苯并噻嗪-2(3H)-酮)、吡唑胞苷(2H-嘧啶并(4,5-b)吡唑-2-酮)、吡啶并吡唑胞苷(H-吡啶并(3',2':4,5)吡咯并(2,3-d)嘧啶-2-酮)。

[0182] 杂环碱基部分还可包括嘌呤碱基或嘧啶碱基被其它杂环取代的那些杂环碱基部分,例如7-脱氮杂腺嘌呤、7-脱氮杂鸟苷、2-氨基吡啶和2-吡啶酮。其它核碱基包括在美国

专利号3,687,808中公开的那些核碱基;在The Concise Encyclopedia Of Polymer Science And Engineering,第858-859页,Kroschwitz,J.I.编辑,John Wiley&Sons,1990中公开的那些核碱基;由Englisch等人,Angewandte Chemie,International Edition,1991,30,613公开的那些核碱基;以及由Sanghvi,Y.S.,第15章,Antisense Research and Applications,第289-302页,Crooke,S.T.和Lebleu,B.编辑,CRC Press,1993公开的那些核碱基。这些核碱基中的某些可用于增加寡聚化合物的结合亲和力。这些核碱基包括5-取代的嘧啶、6-氮杂嘧啶,以及N-2、N-6和O-6取代的嘌呤,包括2-氨基丙基腺嘌呤、5-丙炔基尿嘧啶和5-丙炔基胞嘧啶。5-甲基胞嘧啶取代已显示为使核酸双链体稳定性增加0.6-1.2 °C (Sanghvi等人编辑,Antisense Research and Applications,CRC Press,Boca Raton,1993,第276-278页),并且是合适的碱基取代,例如当与2'-O-甲氧基乙基糖修饰组合时。

[0183] 检测两种不同的靶RNA

[0184] 如上所述,在一些情况下,本公开的方法提供了对样品中两种不同靶RNA(第一单链靶RNA和第二单链靶RNA)的基本上同时检测。在一些情况下,所述方法包括:a)使样品(例如,包含两种不同靶RNA和多种非靶RNA的样品)与以下接触:(i)第一C2c2蛋白,其切割样品中存在的腺嘌呤⁺RNA(即,包括A的RNA,而不是缺乏A的RNA,诸如聚U RNA);(ii)第二C2c2蛋白,其切割尿嘧啶⁺RNA(即,包括U的RNA,而不是缺乏U的RNA,如聚A RNA);(iii)第一C2c2指导RNA,其包含与第一单链靶RNA杂交的第一核苷酸序列和与第一C2c2蛋白结合的第二核苷酸序列;以及(iv)第二C2c2指导RNA,其包含与第二单链靶RNA杂交的第一核苷酸序列和与第二C2c2蛋白结合的第二核苷酸序列;以及b)测量由第一和第二C2c2蛋白介导的RNA切割产生的可检测信号,其中第一可检测信号由第一C2c2蛋白产生,并且第二可检测信号由第二C2c2蛋白产生,其中所述第一可检测信号和所述第二可检测信号可彼此区分。在一些情况下,第一C2c2蛋白不被第二C2c2指导RNA激活,并且第一C2c2蛋白切割包括A的ssRNA(例如,不切割缺乏A的ssRNA);并且第二C2c2蛋白不被第一C2c2指导RNA激活,并且第二C2c2蛋白切割包括U的ssRNA(例如,不切割缺乏U的ssRNA)。在一些情况下,第一C2c2蛋白不被第二C2c2指导RNA激活,并且第一C2c2蛋白切割包括U的ssRNA(例如,不切割缺乏U的ssRNA);并且第二C2c2蛋白不被第一C2c2指导RNA激活,并且第二C2c2蛋白切割包括A的ssRNA(例如,不切割缺乏A的ssRNA)。

[0185] 在一些情况下,所述方法还包括使样品与以下接触:i)第一经标记的检测RNA,其包含第一FRET对和/或第一猝灭剂/荧光剂对(上文描述了示例性FRET对和猝灭剂/荧光剂对);以及ii)第二经标记的检测RNA,其包含第二FRET对和/或第二猝灭剂/荧光剂对(上文描述了示例性FRET对和猝灭剂/荧光剂对)。在一些情况下,第一经标记的检测RNA包含至少一个A并且不包含U;而第二经标记的检测RNA包含至少一个U并且不包含A。在一些情况下,第一经标记的检测RNA包含至少一个U并且不包含A;而第二经标记的检测RNA包含至少一个A并且不包含U。第一C2c2蛋白切割第一经标记的检测RNA,并且第一可检测信号由第一FRET对和/或第一猝灭剂/荧光剂对产生,并且第二C2c2蛋白切割第二经标记的检测RNA,并且第二可检测信号由第二FRET对和/或第二猝灭剂/荧光剂对产生。检测到第一可检测信号表明样品中存在第一靶RNA;并且检测到第二可检测信号表明样品中存在第二靶RNA。在一些情况下,检测到的第一信号和第二信号的相对量指示样品中第一靶RNA与第二靶RNA的比率。

[0186] 在一些情况下,第一经标记的检测RNA包含可与第二经标记的检测RNA的标记区分

的标记。例如,第一经标记的检测RNA可包含第一FRET对和/或第一猝灭剂/荧光剂对;并且第二经标记的检测RNA可包含第二FRET对和/或第二猝灭剂/荧光剂对。作为一个非限制性实例,第一经标记的检测RNA可包含含色氨酸的供体和含丹酰的受体;并且第二经标记的检测RNA可包含含IAEDANS的供体和含DDPM的受体。作为另一个非限制性实例,第一经标记的检测RNA包含含丹酰的供体和含FITC的受体;并且第二经标记的检测RNA包含含Cy3的供体和含Cy5的受体。在一些情况下,第一经标记的检测RNA包含5' FAM (荧光素) -3' IBFQ (Iowa **Black®**FQ) 猝灭剂/荧光剂对,并且在一些情况下,第二经标记的检测RNA包含5' FAM (荧光素) -3' IBFQ (Iowa **Black®**FQ) 猝灭剂/荧光剂对。

[0187] 在一些情况下,将第一和第二经标记的检测RNA同时加入样品中(基本上同时接触)。在一些此类情况下,由第一和第二经标记的检测RNA产生的信号同时被检测到(基本上同时接触),例如是因为在此类情况下,第一和第二经标记的检测RNA可以被区别地标记。

[0188] “基本上同时”是指在约5分钟内、约3分钟内、约2分钟内、约1分钟内、约30秒内、约15秒内、约10秒内、约5秒内,或约1秒内。

[0189] 然而,在一些情况下,由第一和第二经标记的检测RNA产生的信号不是同时被检测到的,而是替代地被顺序检测到的(一个在另一个之前)。例如,在一些情况下,第一和第二经标记的检测RNA不是同时加入到样品中的,而是替代地顺序加入(例如,可以在第一经标记的检测RNA加入之后加入第二经标记的检测RNA),并且在一些此类情况下,直到检测到由第一经标记的检测RNA产生的信号之后才添加第二经标记的检测RNA。因此,在一些情况下,不需要可区分地标记第一和第二经标记的检测RNA(例如,它们可以在一些情况下产生相同的可检测信号,例如,可以在相同波长下发荧光),这是因为信号将被顺序地检测到。

[0190] 作为说明性实例,在一些情况下:(i) 没有可区别地标记第一和第二经标记的检测RNA;(ii) 使样品与一种经标记的检测RNA接触,并且检测(例如,测量)由该经标记的检测RNA产生的信号;以及(iii) 然后使样品与另一种经标记的检测RNA接触,并检测由所加入的第二经标记的检测RNA产生的信号-因此,当样品中存在两种靶ssRNA时,加入第二经标记的检测RNA可以导致信号增强(例如,如果信号随着切割增加而增加,例如荧光剂/猝灭剂对)或者可以导致在加入第二经标记的检测RNA之后可检测的信号减少(例如,如果信号随着切割增加而减少,例如FRET对)。

[0191] 关于C2c2指导RNA结合,第一和第二C2c2蛋白可以彼此正交。在此类情况下,第一C2c2蛋白不与第二C2c2指导RNA结合;并且第二C2c2蛋白不与第一指导RNA结合。第一C2c2蛋白和第二C2c2蛋白还可以在它们的ssRNA切割偏好性方面彼此不同,使得C2c2蛋白中的一种C2c2蛋白在A处切割ssRNA而另一种C2c2蛋白在U处切割ssRNA。

[0192] 对C2c2蛋白的正交对的指导可以在图44E中找到。

[0193] 适用于本公开的方法的C2c2蛋白的正交对的非限制性实例包括下表10中描述的那些。切割偏好性在Cas13a蛋白名称后的括号中呈现。例如,“Lba (A)”是指Lba Cas13a蛋白,其在A处切割ssRNA;并且“Lbu (U)”是指Lbu Cas13a蛋白,其在U处切割ssRNA。

[0194] 表10

[0195]	C2c2 蛋白 # 1	C2c2 蛋白 # 2
	Lba (A)	Hhe (U)
	Lba (A)	Rca (U)
	Lba (A)	Ppr (U)
	Lba (A)	Lne (U)
	Lba (A)	Lbu (U)
	Lba (A)	Lwa (U)
	Lba (A)	Lsh (U)
	Ere (A)	Hhe (U)
	Ere (A)	Rca (U)
	Ere (A)	Ppr (U)
	Ere (A)	Lne (U)
	Ere (A)	Lbu (U)

[0196]	C2c2 蛋白 # 1	C2c2 蛋白 # 2
	Ere (A)	Lwa (U)
	Ere (A)	Lsh (U)
	Ere (A)	Lse (U)
	Cam (A)	Hhe (U)
	Cam (A)	Rca (U)
	Cam (A)	Ppr (U)
	Cam (A)	Lne (U)
	Cam (A)	Lbu (U)
	Cam (A)	Lwa (U)
	Cam (A)	Lsh (U)
	Cam (A)	Lse (U)

[0197] 第一和第二经标记的检测RNA可以各自独立地具有2至100个核糖核苷酸(例如,2至80个、2至60个、2至50个、2至40个、2至30个、2至20个、2至15个或2至10个核糖核苷酸)的长度。第一和第二经标记的检测RNA可以各自独立地具有2个核糖核苷酸至100个核糖核苷酸的长度,例如2个核糖核苷酸至5个核糖核苷酸、5个核糖核苷酸至7个核糖核苷酸、7个核糖核苷酸至10个核糖核苷酸、10个核糖核苷酸至15个核糖核苷酸、15个核糖核苷酸至20个核糖核苷酸、20个核糖核苷酸至25个核糖核苷酸、25个核糖核苷酸至30个核糖核苷酸、30个核糖核苷酸至35个核糖核苷酸、35个核糖核苷酸至40个核糖核苷酸、40个核糖核苷酸至45个核糖核苷酸,或45个核糖核苷酸至50个核糖核苷酸的长度。

[0198] 在一些情况下,第一经标记的检测RNA包含至少一个A(例如,至少2个、至少3个或至少4个A)并且缺乏U;并且第二经标记的检测RNA包含至少一个U(例如,至少2个、至少3个或至少4个U)并且缺乏A。

[0199] 在一些情况下,第一经标记的检测RNA缺乏U并且包括2至15个连续的A(例如,2至12个、2至10个、2至8个、2至6个、2至4个、3至15个、3至12个、3至10个、3至8个、3至6个、3至5个、4至15个、4至12个、4至10个、4至8个或4至6个连续的A)的段。在一些情况下,第一经标记的检测RNA缺乏U并且包括至少2个连续的A(例如,至少3个、至少4个或至少5个连续的A)的段。在一些情况下,第二经标记的检测RNA缺乏A并且包括2至15个连续的U(例如,2至12个、2至10个、2至8个、2至6个、2至4个、3至15个、3至12个、3至10个、3至8个、3至6个、3至5个、4至15个、4至12个、4至10个、4至8个或4至6个连续的U)的段。在一些情况下,第二经标记的检测

RNA缺乏A并且包括至少2个连续的U(例如,至少3个、至少4个或至少5个连续的U)的段。

[0200] 在一些情况下,第一经标记的检测RNA缺乏A并且包括2至15个连续的U(例如,2至12个、2至10个、2至8个、2至6个、2至4个、3至15个、3至12个、3至10个、3至8个、3至6个、3至5个、4至15个、4至12个、4至10个、4至8个或4至6个连续的U)的段。在一些情况下,第一经标记的检测RNA缺乏A并且包括至少2个连续的U(例如,至少3个、至少4个或至少5个连续的U)的段。在一些情况下,第二经标记的检测RNA缺乏U并且包括2至15个连续的A(例如,2至12个、2至10个、2至8个、2至6个、2至4个、3至15个、3至12个、3至10个、3至8个、3至6个、3至5个、4至15个、4至12个、4至10个、4至8个或4至6个连续的A)的段。在一些情况下,第二经标记的检测RNA缺乏U并且包括至少2个连续的A(例如,至少3个、至少4个或至少5个连续的A)的段。

[0201] 在一些情况下,第一经标记的检测RNA包含至少一个U并且缺乏A;并且第二经标记的检测RNA包含至少一个A并且缺乏U。

[0202] 在一些情况下,第一经标记的检测RNA包含至少一个A并且缺乏U。例如,在一些情况下,第一经标记的检测RNA是同聚腺苷聚合物(聚A RNA)。作为另一实例,第一经标记的检测RNA:i)包含至少一个A;ii)缺乏U;并且iii)包含一个或多个C和/或G。在一些情况下,第二经标记的检测RNA包含至少一个U并且缺乏A。例如,在一些情况下,第二经标记的检测RNA是同聚尿苷聚合物(聚URNA)。作为另一实例,第二经标记的检测RNA:i)包含至少一个U;ii)缺乏A;并且iii)包含一个或多个C和/或G。

[0203] 在一些情况下,第一经标记的检测RNA包含至少一个U并且缺乏A。例如,在一些情况下,第一经标记的检测RNA是同聚尿苷聚合物(聚U RNA)。作为另一实例,第二经标记的检测RNA:i)包含至少一个U;ii)缺乏A;并且iii)包含一个或多个C和/或G。在一些情况下,第二经标记的检测RNA包含至少一个A并且缺乏U。例如,在一些情况下,第二经标记的检测RNA是同聚腺苷聚合物(聚ARNA)。作为另一实例,第二经标记的检测RNA:i)包含至少一个A;ii)缺乏U;并且iii)包含一个或多个C和/或G。

[0204] 如上所述,本公开的方法可以包括使样品与以下接触:第一C2c2蛋白;第二C2c2蛋白;第一C2c2指导RNA,其包含与第一单链靶RNA杂交的第一核苷酸序列和在本文中也称为‘恒定区’或‘柄’的第二核苷酸序列,所述第二核苷酸序列与第一C2c2蛋白结合;以及第二C2c2指导RNA,其包含与第二单链靶RNA杂交的第一核苷酸序列和与第二C2c2蛋白结合的第二核苷酸序列(柄)。

[0205] 例如,在一些情况下,第一C2c2蛋白是Cas13a多肽,其包含与图56F中所示的Lba Cas13a氨基酸序列具有至少75%、至少80%、至少85%、至少90%、至少95%、至少98%、至少99%或100%的氨基酸序列同一性的氨基酸序列;并且第一C2c2指导RNA包含恒定区(‘柄’-与Cas13a多肽结合的一段核苷酸),其包含这样的核苷酸序列,所述核苷酸序列与核苷酸序列AGAUAGCCCAAGAAAGAGGGCAAUAAC (SEQ ID NO:16)的差异不超过1个核苷酸(nt)、不超过2nt、不超过3nt、不超过4nt或不超过5nt,其中crRNA的长度为约25nt、26nt、27nt、28nt、29nt或30nt。在一些情况下,crRNA具有核苷酸序列AGAUAGCCCAAGAAAGAGGGCAAUAAC (SEQ ID NO:16);并且长度为27nt。在一些情况下,第二C2c2蛋白包含与图56K中所示的Hhe Cas13a氨基酸序列具有至少75%、至少80%、至少85%、至少90%、至少95%、至少98%、至少99%或100%的氨基酸序列同一性的氨基酸序列;并且第二C2c2指导RNA包含柄(与Cas13a多肽结合的一段核苷酸),其包含这样的核苷酸序列,所述核苷酸序列与核苷酸序列

GUAACAAUCCCCGUAGACAGGGGAACUGCAAC (SEQ ID NO:17) 的差异不超过1个核苷酸(nt)、不超过2nt、不超过3nt、不超过4nt或不超过5nt。在一些情况下,第二C2c2蛋白包含与图56G中所示的Rca Cas13a氨基酸序列具有至少75%、至少80%、至少85%、至少90%、至少95%、至少98%、至少99%或100%的氨基酸序列同一性的氨基酸序列;并且第二C2c2指导RNA包含柄(与Cas13a多肽结合的一段核苷酸),其包含这样的核苷酸序列,所述核苷酸序列与核苷酸序列CAUCACCGCCAAGACGACGGCGGACUGAACC (SEQ ID NO:18) 的差异不超过1个核苷酸(nt)、不超过2nt、不超过3nt、不超过4nt或不超过5nt。在一些情况下,第二C2c2蛋白包含与图56B中所示的Ppr Cas13a氨基酸序列具有至少75%、至少80%、至少85%、至少90%、至少95%、至少98%、至少99%或100%的氨基酸序列同一性的氨基酸序列;并且第二C2c2指导RNA包含柄(与Cas13a多肽结合的一段核苷酸),其包含这样的核苷酸序列,所述核苷酸序列与核苷酸序列AAUUAUCCCCAAAAUUGAAGGGGAACUACAAC (SEQ ID NO:19) 的差异不超过1个核苷酸(nt)、不超过2nt、不超过3nt、不超过4nt或不超过5nt;其中所述柄的长度为约28nt、29nt、30nt、31nt或32nt。在一些情况下,第二C2c2指导RNA包含柄,所述柄包含核苷酸序列AAUUAUCCCCAAAAUUGAAGGGGAACUACAAC (SEQ ID NO:19);并且所述柄的长度为30nt。在一些情况下,第二C2c2蛋白包含与图56I中所示的Lne Cas13a氨基酸序列具有至少75%、至少80%、至少85%、至少90%、至少95%、至少98%、至少99%或100%的氨基酸序列同一性的氨基酸序列;并且第二C2c2指导RNA包含柄(与Cas13a多肽结合的一段核苷酸),其包含这样的核苷酸序列,所述核苷酸序列与核苷酸序列GAGUACCUCAAAACAAAAGAGGACUAAAAC (SEQ ID NO:20) 的差异不超过1个核苷酸(nt)、不超过2nt、不超过3nt、不超过4nt或不超过5nt(例如,包含与核苷酸序列GAGUACCUCAAAACAAAAGAGGACUAAAAC (SEQ ID NO:20) 仅差异1nt、2nt、3nt、4nt或5nt的核苷酸序列);其中所述柄的长度为约28nt、29nt、30nt、31nt或32nt。在一些情况下,第二指导RNA包含柄,所述柄包含核苷酸序列GAGUACCUCAAAACAAAAGAGGACUAAAAC (SEQ ID NO:20);其中所述柄的长度为30nt。在一些情况下,第二C2c2蛋白包含与图56C中所示的Lbu Cas13a氨基酸序列具有至少75%、至少80%、至少85%、至少90%、至少95%、至少98%、至少99%或100%的氨基酸序列同一性的氨基酸序列;并且第二C2c2指导RNA包含柄(与Cas13a多肽结合的一段核苷酸),其包含这样的核苷酸序列,所述核苷酸序列与核苷酸序列GACCACCCCAAAAAUGAAGGGGACUAAAACA (SEQ ID NO:9) 的差异不超过1个核苷酸(nt)、不超过2nt、不超过3nt、不超过4nt或不超过5nt(例如,包含与核苷酸序列GACCACCCCAAAAAUGAAGGGGACUAAAACA (SEQ ID NO:9) 仅差异1nt、2nt、3nt、4nt或5nt的核苷酸序列);其中所述柄的长度为约28nt、29nt、30nt、31nt或32nt。在一些情况下,第二指导RNA包含柄,所述柄包含核苷酸序列GACCACCCCAAAAAUGAAGGGGACUAAAACA (SEQ ID NO:9);其中所述柄的长度为31nt。在一些情况下,第二C2c2蛋白包含与图56E中所示的LwaCas13a氨基酸序列具有至少75%、至少80%、至少85%、至少90%、至少95%、至少98%、至少99%或100%的氨基酸序列同一性的氨基酸序列;并且第二C2c2指导RNA包含柄(与Cas13a多肽结合的一段核苷酸),其包含这样的核苷酸序列,所述核苷酸序列与核苷酸序列GACCACCCCAAUAUCGAAGGGGACUAAAACUU (SEQ ID NO:21) 的差异不超过1个核苷酸(nt)、不超过2nt、不超过3nt、不超过4nt或不超过5nt(例如,包含与核苷酸序列GACCACCCCAAUAUCGAAGGGGACUAAAACUU (SEQ ID NO:21) 仅差异1nt、2nt、3nt、4nt或5nt的核苷酸序列);其中所述柄的长度为约28nt、29nt、30nt、31nt或32nt。在一些情况下,第二指导RNA包含柄,所述柄包含核苷酸序列

GACCACCCCAAUAUCGAAGGGGACUAAAACUU (SEQ ID NO:21);其中所述柄的长度为32nt。在一些情况下,第二C2c2蛋白包含与图56D中所示的Lsh Cas13a氨基酸序列具有至少75%、至少80%、至少85%、至少90%、至少95%、至少98%、至少99%或100%的氨基酸序列同一性的氨基酸序列;并且第二C2c2指导RNA包含柄(与Cas13a多肽结合的一段核苷酸),其包含这样的核苷酸序列,所述核苷酸序列与核苷酸序列CACCCCAAUAUCGAAGGGGACUAAAAC (SEQ ID NO:22)的差异不超过1个核苷酸(nt)、不超过2nt、不超过3nt、不超过4nt或不超过5nt(例如,包含与核苷酸序列CACCCCAAUAUCGAAGGGGACUAAAAC (SEQ ID NO:22)仅差异1nt、2nt、3nt、4nt或5nt的核苷酸序列);其中所述柄的长度为25nt、26nt、27nt、28nt或29nt。在一些情况下,第二指导RNA包含柄,所述柄包含核苷酸序列CACCCCAAUAUCGAAGGGGACUAAAAC (SEQ ID NO:22);其中所述柄的长度为27nt。

[0206] 作为另一实例,在一些情况下,第一C2c2蛋白是Cas13a多肽,其包含与图56J中所示的Ere Cas13a氨基酸序列具有至少75%、至少80%、至少85%、至少90%、至少95%、至少98%、至少99%或100%的氨基酸序列同一性的氨基酸序列;并且第一C2c2指导RNA包含柄(与Cas13a多肽结合的一段核苷酸),其包含这样的核苷酸序列,所述核苷酸序列与核苷酸序列AAGUAGCCCGAUUAUAGGGGCAAUAAC (SEQ ID NO:23)的差异不超过1个核苷酸(nt)、不超过2nt、不超过3nt、不超过4nt或不超过5nt,其中所述柄的长度为约25nt、26nt、27nt、28nt、29nt或30nt。在一些情况下,所述柄具有核苷酸序列AAGUAGCCCGAUUAUAGGGGCAAUAAC (SEQ ID NO:23);并且长度为27nt。在一些情况下,第一C2c2蛋白是Cas13a多肽,其包含与图56J中所示的Ere Cas13a氨基酸序列具有至少75%、至少80%、至少85%、至少90%、至少95%、至少98%、至少99%或100%的氨基酸序列同一性的氨基酸序列;并且第一C2c2指导RNA包含柄(与Cas13a多肽结合的一段核苷酸),其包含这样的核苷酸序列,所述核苷酸序列与核苷酸序列AUACAGCUCGAUAUAGUGAGCAAUAAG (SEQ ID NO:24)的差异不超过1个核苷酸(nt)、不超过2nt、不超过3nt、不超过4nt或不超过5nt,其中所述柄的长度为约25nt、26nt、27nt、28nt、29nt或30nt。在一些情况下,所述柄具有核苷酸序列AUACAGCUCGAUAUAGUGAGCAAUAAG (SEQ ID NO:24);并且长度为27nt。在一些情况下,第二C2c2蛋白包含与图56K中所示的Hhe Cas13a氨基酸序列具有至少75%、至少80%、至少85%、至少90%、至少95%、至少98%、至少99%或100%的氨基酸序列同一性的氨基酸序列;并且第二C2c2指导RNA包含柄(与Cas13a多肽结合的一段核苷酸),其包含这样的核苷酸序列,所述核苷酸序列与核苷酸序列GUAACAAUCCCGUAGACAGGGGAACUGCAAC (SEQ ID NO:17)的差异不超过1个核苷酸(nt)、不超过2nt、不超过3nt、不超过4nt或不超过5nt;其中所述柄的长度为约30nt、31nt、32nt、33nt、34nt或35nt。在一些情况下,第二C2c2指导RNA包含柄,所述柄包含核苷酸序列GUAACAAUCCCGUAGACAGGGGAACUGCAAC (SEQ ID NO:17);并且所述柄的长度为32nt。在一些情况下,第二C2c2蛋白包含与图56G中所示的Rca Cas13a氨基酸序列具有至少75%、至少80%、至少85%、至少90%、至少95%、至少98%、至少99%或100%的氨基酸序列同一性的氨基酸序列;并且第二C2c2指导RNA包含柄(与Cas13a多肽结合的一段核苷酸),其包含这样的核苷酸序列,所述核苷酸序列与核苷酸序列UCACAUCACCGCCAAGACGACGGCGGACUGAACC (SEQ ID NO:25)的差异不超过1个核苷酸(nt)、不超过2nt、不超过3nt、不超过4nt或不超过5nt;其中所述柄的长度为约32nt、33nt、34nt、35nt、36nt或37nt。在一些情况下,第二C2c2指导RNA包含柄,所述柄包含核苷酸序列UCACAUCACCGCCAAGACGACGGCGGACUGAACC (SEQ ID NO:25);并且

所述柄的长度为34nt。在一些情况下,第二C2c2蛋白包含与图56B中所示的Ppr Cas13a氨基酸序列具有至少75%、至少80%、至少85%、至少90%、至少95%、至少98%、至少99%或100%的氨基酸序列同一性的氨基酸序列;并且第二C2c2指导RNA包含柄(与Cas13a多肽结合的一段核苷酸),其包含这样的核苷酸序列,所述核苷酸序列与核苷酸序列AAUUAUCCCCAAAAUGAAGGGAACUACAAC (SEQ ID NO:19) 的差异不超过1个核苷酸(nt)、不超过2nt、不超过3nt、不超过4nt或不超过5nt;其中所述柄的长度为约28nt、29nt、30nt、31nt或32nt。在一些情况下,第二C2c2指导RNA包含柄,所述柄包含核苷酸序列AAUUAUCCCCAAAAUGAAGGGAACUACAAC (SEQ ID NO:19);并且所述柄的长度为30nt。在一些情况下,第二C2c2蛋白包含与图56I中所示的Lne Cas13a氨基酸序列具有至少75%、至少80%、至少85%、至少90%、至少95%、至少98%、至少99%或100%的氨基酸序列同一性的氨基酸序列;并且第二C2c2指导RNA包含柄(与Cas13a多肽结合的一段核苷酸),其包含这样的核苷酸序列,所述核苷酸序列与核苷酸序列GAGUACCUCAAAAACAAAAGAGGACUAAAAC (SEQ ID NO:20) 的差异不超过1个核苷酸(nt)、不超过2nt、不超过3nt、不超过4nt或不超过5nt(例如,包含与核苷酸序列GAGUACCUCAAAAACAAAAGAGGACUAAAAC (SEQ ID NO:20) 仅差异1nt、2nt、3nt、4nt或5nt的核苷酸序列);其中所述柄的长度为约28nt、29nt、30nt、31nt或32nt。在一些情况下,第二指导RNA包含柄,所述柄包含核苷酸序列GAGUACCUCAAAAACAAAAGAGGACUAAAAC (SEQ ID NO:20);其中所述柄的长度为30nt。在一些情况下,第二C2c2蛋白包含与图56C中所示的Lbu Cas13a氨基酸序列具有至少75%、至少80%、至少85%、至少90%、至少95%、至少98%、至少99%或100%的氨基酸序列同一性的氨基酸序列;并且第二C2c2指导RNA包含柄(与Cas13a多肽结合的一段核苷酸),其包含这样的核苷酸序列,所述核苷酸序列与核苷酸序列GACCACCCCCAAAAUGAAGGGGACUAAAACA (SEQ ID NO:9) 的差异不超过1个核苷酸(nt)、不超过2nt、不超过3nt、不超过4nt或不超过5nt(例如,包含与核苷酸序列GACCACCCCCAAAAUGAAGGGGACUAAAACA (SEQ ID NO:9) 仅差异1nt、2nt、3nt、4nt或5nt的核苷酸序列);其中所述柄的长度为约28nt、29nt、30nt、31nt或32nt。在一些情况下,第二指导RNA包含柄,所述柄包含核苷酸序列GACCACCCCCAAAAUGAAGGGGACUAAAACA (SEQ ID NO:9);其中所述柄的长度为31nt。在一些情况下,第二C2c2蛋白包含与图56E中所示的Lwa Cas13a氨基酸序列具有至少75%、至少80%、至少85%、至少90%、至少95%、至少98%、至少99%或100%的氨基酸序列同一性的氨基酸序列;并且第二C2c2指导RNA包含柄(与Cas13a多肽结合的一段核苷酸),其包含这样的核苷酸序列,所述核苷酸序列与核苷酸序列GACCACCCCCAAUAUCGAAGGGGACUAAAACUU (SEQ ID NO:21) 的差异不超过1个核苷酸(nt)、不超过2nt、不超过3nt、不超过4nt或不超过5nt(例如,包含与核苷酸序列GACCACCCCCAAUAUCGAAGGGGACUAAAACUU (SEQ ID NO:21) 仅差异1nt、2nt、3nt、4nt或5nt的核苷酸序列);其中所述柄的长度为约28nt、29nt、30nt、31nt或32nt。在一些情况下,第二指导RNA包含柄,所述柄包含核苷酸序列GACCACCCCCAAUAUCGAAGGGGACUAAAACUU (SEQ ID NO:21);其中所述柄的长度为32nt。在一些情况下,第二C2c2蛋白包含与图56D中所示的Lsh Cas13a氨基酸序列具有至少75%、至少80%、至少85%、至少90%、至少95%、至少98%、至少99%或100%的氨基酸序列同一性的氨基酸序列;并且第二C2c2指导RNA包含柄(与Cas13a多肽结合的一段核苷酸),其包含这样的核苷酸序列,所述核苷酸序列与核苷酸序列CACCCCCAAUAUCGAAGGGGACUAAAAC (SEQ ID NO:22) 的差异不超过1个核苷酸(nt)、不超过2nt、不超过3nt、不超过4nt或不超过5nt(例如,包

含与核苷酸序列CACCCCAUAUCGAAGGGGACUAAAAC (SEQ ID NO:22) 仅差异1nt、2nt、3nt、4nt或5nt的核苷酸序列);其中所述柄的长度为25nt、26nt、27nt、28nt或29nt。在一些情况下,第二指导RNA包含柄,所述柄包含核苷酸序列CACCCCAUAUCGAAGGGGACUAAAAC (SEQ ID NO:22);其中所述柄的长度为27nt。在一些情况下,第二C2c2蛋白包含与图56A中所示的Lse Cas13a氨基酸序列具有至少75%、至少80%、至少85%、至少90%、至少95%、至少98%、至少99%或100%的氨基酸序列同一性的氨基酸序列;并且第二C2c2指导RNA包含柄(与Cas13a多肽结合的一段核苷酸),其包含这样的核苷酸序列,所述核苷酸序列与核苷酸序列GACUACCUCUAUAUGAAAGAGGACUAAAAC (SEQ ID NO:7) 的差异不超过1个核苷酸(nt)、不超过2nt、不超过3nt、不超过4nt或不超过5nt;其中所述柄的长度为约28nt、29nt、30nt、31nt或32nt。在一些情况下,第二C2c2指导RNA包含柄,所述柄包含核苷酸序列GACUACCUCUAUAUGAAAGAGGACUAAAAC (SEQ ID NO:7);并且所述柄的长度为30nt。

[0207] 作为另一实例,在一些情况下,第一C2c2蛋白是Cas13a多肽,其包含与图56H中所示的Cam Cas13a氨基酸序列具有至少75%、至少80%、至少85%、至少90%、至少95%、至少98%、至少99%或100%的氨基酸序列同一性的氨基酸序列;并且第一C2c2指导RNA包含柄(与Cas13a多肽结合的一段核苷酸),其包含这样的核苷酸序列,所述核苷酸序列与核苷酸序列GAACAGCCCGAUUAUAGAGGGCAAUAGAC (SEQ ID NO:26) 的差异不超过1个核苷酸(nt)、不超过2nt、不超过3nt、不超过4nt或不超过5nt,其中所述柄的长度为约26nt、27nt、28nt、29nt或30nt。在一些情况下,所述柄具有核苷酸序列GAACAGCCCGAUUAUAGAGGGCAAUAGAC (SEQ ID NO:26);并且长度为28nt。在一些情况下,第二C2c2蛋白包含与图56K中所示的Hhe Cas13a氨基酸序列具有至少75%、至少80%、至少85%、至少90%、至少95%、至少98%、至少99%或100%的氨基酸序列同一性的氨基酸序列;并且第二C2c2指导RNA包含柄(与Cas13a多肽结合的一段核苷酸),其包含这样的核苷酸序列,所述核苷酸序列与核苷酸序列GUAACAAUCCCGUAGACAGGGGAACUGCAAC (SEQ ID NO:17) 的差异不超过1个核苷酸(nt)、不超过2nt、不超过3nt、不超过4nt或不超过5nt;其中所述柄的长度为约30nt、31nt、32nt、33nt、34nt,或35nt。在一些情况下,第二C2c2指导RNA包含柄,所述柄包含核苷酸序列GUAACAAUCCCGUAGACAGGGGAACUGCAAC (SEQ ID NO:17);并且所述柄的长度为32nt。在一些情况下,第二C2c2蛋白包含与图56G中所示的Rca Cas13a氨基酸序列具有至少75%、至少80%、至少85%、至少90%、至少95%、至少98%、至少99%或100%的氨基酸序列同一性的氨基酸序列;并且第二C2c2指导RNA包含柄(与Cas13a多肽结合的一段核苷酸),其包含这样的核苷酸序列,所述核苷酸序列与核苷酸序列UCACAUCACCGCCAAGACGACGGCGGACUGAACC (SEQ ID NO:25) 的差异不超过1个核苷酸(nt)、不超过2nt、不超过3nt、不超过4nt或不超过5nt;其中所述柄的长度为约32nt、33nt、34nt、35nt、36nt或37nt。在一些情况下,第二C2c2指导RNA包含柄,所述柄包含核苷酸序列UCACAUCACCGCCAAGACGACGGCGGACUGAACC (SEQ ID NO:25);并且所述柄的长度为34nt。在一些情况下,第二C2c2蛋白包含与图56B中所示的Ppr Cas13a氨基酸序列具有至少75%、至少80%、至少85%、至少90%、至少95%、至少98%、至少99%或100%的氨基酸序列同一性的氨基酸序列;并且第二C2c2指导RNA包含柄(与Cas13a多肽结合的一段核苷酸),其包含这样的核苷酸序列,所述核苷酸序列与核苷酸序列

AAUUAUCCCAAAAUUGAAGGGAACUACAAC (SEQ ID NO:19) 的差异不超过1个核苷酸(nt)、不超过2nt、不超过3nt、不超过4nt或不超过5nt;其中所述柄的长度为约28nt、29nt、30nt、31nt或

32nt。在一些情况下,第二C2c2指导RNA包含柄,所述柄包含核苷酸序列AAUUAUCCCCAAAAUGAAGGGAACUACAAC(SEQ ID NO:19);并且所述柄的长度为30nt。在一些情况下,第二C2c2蛋白包含与图56I中所示的Lne Cas13a氨基酸序列具有至少75%、至少80%、至少85%、至少90%、至少95%、至少98%、至少99%或100%的氨基酸序列同一性的氨基酸序列;并且第二C2c2指导RNA包含柄(与Cas13a多肽结合的一段核苷酸),其包含这样的核苷酸序列,所述核苷酸序列与核苷酸序列GAGUACCUCAAAAACAAAAGAGGACUAAAAC(SEQ ID NO:20)的差异不超过1个核苷酸(nt)、不超过2nt、不超过3nt、不超过4nt或不超过5nt(例如,包含与核苷酸序列GAGUACCUCAAAAACAAAAGAGGACUAAAAC(SEQ ID NO:20)仅差异1nt、2nt、3nt、4nt或5nt的核苷酸序列);其中所述柄的长度为约28nt、29nt、30nt、31nt或32nt。在一些情况下,第二指导RNA包含柄,所述柄包含核苷酸序列GAGUACCUCAAAAACAAAAGAGGACUAAAAC(SEQ ID NO:20);其中所述柄的长度为30nt。在一些情况下,第二C2c2蛋白包含与图56C中所示的Lbu Cas13a氨基酸序列具有至少75%、至少80%、至少85%、至少90%、至少95%、至少98%、至少99%或100%的氨基酸序列同一性的氨基酸序列;并且第二C2c2指导RNA包含柄(与Cas13a多肽结合的一段核苷酸),其包含这样的核苷酸序列,所述核苷酸序列与核苷酸序列GACCACCCCCAAAAUGAAGGGGACUAAAACA(SEQ ID NO:9)的差异不超过1个核苷酸(nt)、不超过2nt、不超过3nt、不超过4nt或不超过5nt(例如,包含与核苷酸序列GACCACCCCCAAAAUGAAGGGGACUAAAACA(SEQ ID NO:9)仅差异1nt、2nt、3nt、4nt或5nt的核苷酸序列);其中所述柄的长度为约28nt、29nt、30nt、31nt或32nt。在一些情况下,第二指导RNA包含柄,所述柄包含核苷酸序列GACCACCCCCAAAAUGAAGGGGACUAAAACA(SEQ ID NO:9);其中所述柄的长度为31nt。在一些情况下,第二C2c2蛋白包含与图56E中所示的Lwa Cas13a氨基酸序列具有至少75%、至少80%、至少85%、至少90%、至少95%、至少98%、至少99%或100%的氨基酸序列同一性的氨基酸序列;并且第二C2c2指导RNA包含柄(与Cas13a多肽结合的一段核苷酸),其包含这样的核苷酸序列,所述核苷酸序列与核苷酸序列GACCACCCCAAUAUCGAAGGGGACUAAAACUU(SEQ ID NO:21)的差异不超过1个核苷酸(nt)、不超过2nt、不超过3nt、不超过4nt或不超过5nt(例如,包含与核苷酸序列GACCACCCCAAUAUCGAAGGGGACUAAAACUU(SEQ ID NO:21)仅差异1nt、2nt、3nt、4nt或5nt的核苷酸序列);其中所述柄的长度为约28nt、29nt、30nt、31nt或32nt。在一些情况下,第二指导RNA包含柄,所述柄包含核苷酸序列GACCACCCCAAUAUCGAAGGGGACUAAAACUU(SEQ ID NO:21);其中所述柄的长度为32nt。在一些情况下,第二C2c2蛋白包含与图56D中所示的Lsh Cas13a氨基酸序列具有至少75%、至少80%、至少85%、至少90%、至少95%、至少98%、至少99%或100%的氨基酸序列同一性的氨基酸序列;并且第二C2c2指导RNA包含柄(与Cas13a多肽结合的一段核苷酸),其包含这样的核苷酸序列,所述核苷酸序列与核苷酸序列CACCCCAAUAUCGAAGGGGACUAAAAC(SEQ ID NO:22)的差异不超过1个核苷酸(nt)、不超过2nt、不超过3nt、不超过4nt或不超过5nt(例如,包含与核苷酸序列CACCCCAAUAUCGAAGGGGACUAAAAC(SEQ ID NO:22)仅差异1nt、2nt、3nt、4nt或5nt的核苷酸序列);其中所述柄的长度为25nt、26nt、27nt、28nt或29nt。在一些情况下,第二指导RNA包含柄,所述柄包含核苷酸序列CACCCCAAUAUCGAAGGGGACUAAAAC(SEQ ID NO:22);其中所述柄的长度为27nt。在一些情况下,第二C2c2蛋白包含与图56A中所示的Lse Cas13a氨基酸序列具有至少75%、至少80%、至少85%、至少90%、至少95%、至少98%、至少99%或100%的氨基酸序列同一性的氨基酸序列;并且第二C2c2指导RNA包含柄(与

Cas13a多肽结合的一段核苷酸),其包含这样的核苷酸序列,所述核苷酸序列与核苷酸序列GACUACCUCUAUAUGAAAGAGGACUAAAAC (SEQ ID NO:7) 的差异不超过1个核苷酸(nt)、不超过2nt、不超过3nt、不超过4nt或不超过5nt;其中所述柄的长度为约28nt、29nt、30nt、31nt或32nt。在一些情况下,第二C2c2指导RNA包含柄,所述柄包含核苷酸序列GACUACCUCUAUAUGAAAGAGGACUAAAAC (SEQ ID NO:7);并且所述柄的长度为30nt。

[0208] 混合检测

[0209] 如上所述,在一些情况下,本公开的方法包括:a)使样品(例如,包含靶RNA和多种非靶RNA的样品)与以下接触:i)前体C2c2指导RNA阵列,其包含两个或更多个C2c2指导RNA,每个所述C2c2指导RNA具有不同的指导序列;和(ii)C2c2蛋白,其将前体C2c2指导RNA阵列切割成单独的C2c2指导RNA,并且还切割样品的RNA;以及b)测量由C2c2蛋白介导的RNA切割产生的可检测信号。

[0210] 在一些情况下,阵列(前体C2c2指导RNA阵列)上可以存在两个或更多个C2c2指导RNA。C2c2蛋白可以将前体C2c2指导RNA阵列切割成单独的C2c2指导RNA(例如,参见图4和图6)。

[0211] 在一些情况下,主题C2c2指导RNA阵列包括2个或更多个C2c2指导RNA(例如,3个或更多个、4个或更多个、5个或更多个、6个或更多个、或7个或更多个C2c2指导RNA)。给定阵列的C2c2指导RNA可以靶向相同靶RNA的不同靶位点(即,可以包括与所述不同靶位点杂交的指导序列)(例如,这可以增加检测的灵敏度)和/或可以靶向不同的靶RNA分子(例如,例如基于诸如单核苷酸多态性、单核苷酸多态性(SNP)等等之类的变异的转录物家族等可用于例如检测多种病毒株,诸如流感病毒变体、寨卡病毒变体、HIV变体等)。

[0212] C2c2蛋白

[0213] C2c2蛋白与C2c2指导RNA结合,由指导RNA(其与靶RNA杂交)引导到单链靶RNA,从而被“激活”。如果C2c2蛋白的HEPN1结构域和HEPN2结构域是完整的,则C2c2蛋白一旦被激活就切割靶RNA,但是也切割非靶RNA。

[0214] 示例性的天然存在C2c2蛋白描绘于图8中,并且如SEQ ID NO:1至SEQ ID NO:6所示。在一些情况下,主题C2c2蛋白包括与SEQ ID NO:1至SEQ ID NO:6中任一项所示的氨基酸序列具有80%或更多(例如,85%或更多、90%或更多、95%或更多、98%或更多、99%或更多、99.5%或更多、或100%的氨基酸序列同一性的氨基酸序列。在一些情况下,合适的C2c2多肽包含与SEQ ID NO:1所示的斯氏李斯特菌(*Listeria seeligeri*)C2c2氨基酸序列具有至少80%、至少85%、至少90%、至少95%、至少98%、至少99%或100%的氨基酸序列同一性的氨基酸序列。在一些情况下,合适的C2c2多肽包含与SEQ ID NO:2所示的口腔纤毛菌(*Leptotrichia buccalis*)C2c2氨基酸序列具有至少80%、至少85%、至少90%、至少95%、至少98%、至少99%或100%的氨基酸序列同一性的氨基酸序列。在一些情况下,合适的C2c2多肽包含与SEQ ID NO:4所示的荚膜红细菌(*Rhodobacter capsulatus*)C2c2氨基酸序列具有至少80%、至少85%、至少90%、至少95%、至少98%、至少99%或100%的氨基酸序列同一性的氨基酸序列。在一些情况下,合适的C2c2多肽包含与SEQ ID NO:5所示的鸡肉杆菌(*Carnobacterium gallinarum*)C2c2氨基酸序列具有至少80%、至少85%、至少90%、至少95%、至少98%、至少99%或100%的氨基酸序列同一性的氨基酸序列。在一些情况下,合适的C2c2多肽包含与SEQ ID NO:6所示的Herbinix hemicellulosilytica C2c2氨基酸

序列具有至少80%、至少85%、至少90%、至少95%、至少98%、至少99%或100%的氨基酸序列同一性的氨基酸序列。在一些情况下,C2c2蛋白包括与SEQ ID NO:2所示的口腔纤毛菌(Lbu) C2c2氨基酸序列具有80%或更多的氨基酸序列同一性的氨基酸序列。在一些情况下,C2c2蛋白是口腔纤毛菌(Lbu) C2c2蛋白(例如,参见SEQ ID NO:2)。在一些情况下,C2c2蛋白包括SEQ ID NO:1至SEQ ID NO:2和SEQ ID NO:4至SEQ ID NO:6中任一项所示的氨基酸序列。

[0215] 在一些情况下,在本公开的方法中使用的C2c2蛋白不是沙氏纤毛菌(*Leptotrichia shahii*) (Lsh) C2c2蛋白。在一些情况下,在本公开的方法中使用的C2c2蛋白不是与SEQ ID NO:3所示的Lsh C2c2多肽具有至少80%、至少85%、至少90%、至少95%、至少98%、至少99%或100%的氨基酸序列同一性的C2c2多肽。

[0216] 在一些情况下,所述C2c2蛋白比沙氏纤毛菌(Lsh) C2c2蛋白在切割未由所述方法的C2c2指导RNA靶向的RNA方面更有效1.2倍或更多倍。在一些情况下,所述C2c2蛋白比沙氏纤毛菌(Lsh) C2c2蛋白在切割未由所述方法的C2c2指导RNA靶向的RNA方面更有效1.5倍或更多倍。在一些情况下,在本公开的方法中使用的C2c2多肽当被激活时,比Lsh C2c2至少1.2倍、至少1.5倍、至少2倍、至少2.5倍、至少3倍、至少4倍、至少5倍、至少6倍、至少7倍、至少8倍、至少9倍、至少10倍、至少15倍、至少20倍、至少30倍或大于30倍有效地切割非靶RNA。

[0217] 在一些情况下,C2c2蛋白在所述接触的1小时内表现出至少50%的RNA切割效率(例如,55%或更多、60%或更多、65%或更多、70%或更多、或75%或更多的切割效率)。在一些情况下,C2c2蛋白在所述接触的40分钟内表现出至少50%的RNA切割效率(例如,55%或更多、60%或更多、65%或更多、70%或更多、或75%或更多的切割效率)。在一些情况下,C2c2蛋白在所述接触的30分钟内表现出至少50%的RNA切割效率(例如,55%或更多、60%或更多、65%或更多、70%或更多、或75%或更多的切割效率)。

[0218] 在一些情况下,适用于本公开的方法的C2c2蛋白在30秒至60分钟(例如,1分钟至60分钟、30秒至5分钟、1分钟至5分钟、1分钟至10分钟、5分钟至10分钟、10分钟至15分钟、15分钟至20分钟、20分钟至25分钟、25分钟至30分钟、30分钟至35分钟、35分钟至40分钟、40分钟至45分钟、45分钟至50分钟、50分钟至55分钟或55分钟至60分钟)的时间段内切割样品中存在的RNA的至少50%、至少60%、至少70%、至少80%、至少90%、至少95%、至少98%、至少99%或大于99%。在一些情况下,适用于本公开的方法的C2c2蛋白在30秒至5分钟的时间段内(例如,1分钟至5分钟,例如在1分钟、2分钟、3分钟、4分钟或5分钟的时间段内)切割样品中存在的RNA的至少50%、至少60%、至少70%、至少80%、至少90%、至少95%、至少98%、至少99%或大于99%。在一些情况下,适用于本公开的方法的C2c2蛋白在5分钟至10分钟的时间段内(例如,在5分钟、6分钟、7分钟、8分钟、9分钟或10分钟的时间段内)切割样品中存在的RNA的至少50%、至少60%、至少70%、至少80%、至少90%、至少95%、至少98%、至少99%或大于99%。在一些情况下,适用于本公开的方法的C2c2蛋白在10分钟至15分钟(例如,10分钟、11分钟、12分钟、13分钟、14分钟或15分钟)的时间段内切割样品中存在的RNA的至少50%、至少60%、至少70%、至少80%、至少90%、至少95%、至少98%、至少99%或大于99%。在一些情况下,适用于本公开的方法的C2c2蛋白在15分钟至20分钟的时间段内切割样品中存在的RNA的至少50%、至少60%、至少70%、至少80%、至少90%、至少95%、至少98%、至少99%或大于99%。在一些情况下,适用于本公开的方法的C2c2蛋白在

20分钟至25分钟的时间段内切割样品中存在的RNA的至少50%、至少60%、至少70%、至少80%、至少90%、至少95%、至少98%、至少99%或大于99%。在一些情况下,适用于本公开的方法的C2c2蛋白在25分钟至30分钟的时间段内切割样品中存在的RNA的至少50%、至少60%、至少70%、至少80%、至少90%、至少95%、至少98%、至少99%或大于99%。在一些情况下,适用于本公开的方法的C2c2蛋白在30分钟至35分钟的时间段内切割样品中存在的RNA的至少50%、至少60%、至少70%、至少80%、至少90%、至少95%、至少98%、至少99%或大于99%。在一些情况下,适用于本公开的方法的C2c2蛋白在35分钟至40分钟的时间段内切割样品中存在的RNA的至少50%、至少60%、至少70%、至少80%、至少90%、至少95%、至少98%、至少99%或大于99%。在一些情况下,适用于本公开的方法的C2c2蛋白在40分钟至45分钟的时间段内切割样品中存在的RNA的至少50%、至少60%、至少70%、至少80%、至少90%、至少95%、至少98%、至少99%或大于99%。在一些情况下,适用于本公开的方法的C2c2蛋白在45分钟至50分钟的时间段内切割样品中存在的RNA的至少50%、至少60%、至少70%、至少80%、至少90%、至少95%、至少98%、至少99%或大于99%。在一些情况下,适用于本公开的方法的C2c2蛋白在50分钟至55分钟的时间段内切割样品中存在的RNA的至少50%、至少60%、至少70%、至少80%、至少90%、至少95%、至少98%、至少99%或大于99%。在一些情况下,适用于本公开的方法的C2c2蛋白在55分钟至60分钟的时间段内切割样品中存在的RNA的至少50%、至少60%、至少70%、至少80%、至少90%、至少95%、至少98%、至少99%或大于99%。在一些情况下,适用于本公开的方法的C2c2蛋白在小于1分钟的时间段内(例如,在50秒至59秒、40秒至49秒、30秒至39秒或20秒至29秒的时间段内)切割样品中存在的RNA的至少50%、至少60%、至少70%、至少80%、至少90%、至少95%、至少98%、至少99%或大于99%。在一些情况下,切割在生理条件下发生。在一些情况下,切割在15℃至20℃、20℃至25℃、25℃至30℃、30℃至35℃、或35℃至40℃的温度下发生。在一些情况下,切割在约37℃发生。在一些情况下,切割在约37℃发生并且反应条件包括二价金属离子。在一些情况下,二价金属离子是 Mg^{2+} 。在一些情况下,二价金属离子是 Mn^{2+} 。在一些情况下,反应条件的pH在pH 5与pH 6之间。在一些情况下,反应条件的pH在pH 6与pH 7之间。在一些情况下,反应条件的pH在pH 6.5与pH 7.5之间。在某些情况下,反应条件的pH高于pH 7.5。

[0219] 术语“切割效率”在本文中用于指C2c2蛋白的以下能力:一旦C2c2蛋白已经通过适当的C2c2指导RNA/靶RNA杂交而被激活,就快速切割样品中的RNA。“切割效率”是指蛋白在给定时间段内可以切割的RNA的量。例如,50%的切割效率应表明50%的给定RNA在指定时间段内被切割。例如,如果样品中的RNA以100 μ M的起始浓度存在,则当已经切割了50 μ M的RNA时,已经实现了50%的切割。作为另一实例,如果样品中存在多个RNA分子,则当已经切割了50%的RNA分子时,已经实现了50%的切割;效率是切割一定百分比的总RNA所需的时间量的表示。这可以通过任何合宜的方法测量,并且许多此类方法对于本领域的普通技术人员来说将是已知的。例如,可以使用经标记的检测RNA。在一些情况下,RNA种类(切割的对比未切割的)可以在凝胶上分离,并且可以将切割的RNA的量与未切割的RNA的量进行比较,例如参见图3。

[0220] 当使用短语“其中C2c2蛋白切割样品中存在的RNA的至少X%”(例如,在指定时间段内)时,意指样品中存在的‘信号产生’RNA的X%在指定时间段内被切割。哪种RNA是‘信号

产生' RNA可以取决于所用的检测方法。例如,当使用经标记的检测RNA时,所述经标记的检测RNA可能是唯一的'信号产生RNA'。然而,经标记的检测RNA用于代表样品的RNA,并且因此关于经标记的检测RNA所观察到的情况被假定为代表样品的非靶RNA发生的情况。因此,当50%的经标记的检测RNA被切割时,这通常将被假定为表示当'样品中存在的RNA'的50%被切割时。在一些情况下,通常测量RNA切割,因此样品的所有可切割RNA都是“信号产生RNA”。因此,当提及样品中存在的RNA的被切割百分比时,可以使用任何合宜的方法测量该值,并且无论使用何种方法,该值通常在本文中意指当酶已经切割样品中的可切割靶的一半。

[0221] 在一些情况下,C2c2蛋白不是沙氏纤毛菌(Lsh) C2c2蛋白。在一些情况下,C2c2蛋白比沙氏纤毛菌(Lsh) C2c2蛋白更有效(例如,在切割非靶RNA方面)1.2倍或更高(例如,1.5倍或更高、1.7倍或更高、或2倍或更高)。因此,在一些情况下,主题C2c2蛋白比沙氏纤毛菌(Lsh) C2c2蛋白在切割未由所述方法的C2c2指导RNA靶向的RNA方面更有效1.2倍或更多倍(例如,1.5倍或更多倍、1.7倍或更多倍、或2倍或更多倍)。在一些情况下,在本公开的方法中使用的C2c2多肽当被激活时,比Lsh C2c2至少1.2倍、至少1.5倍、至少2倍、至少2.5倍、至少3倍、至少4倍、至少5倍、至少6倍、至少7倍、至少8倍、至少9倍、至少10倍、至少15倍、至少20倍、至少30倍或大于30倍有效地切割非靶RNA。

[0222] 变体C2c2多肽

[0223] 变体C2c2多肽包括SEQ ID NO:1、SEQ ID NO:2和SEQ ID NO:4至SEQ ID NO:6中任一者的变体,其中变体C2c2多肽表现出降低的(或不可检测的)核酸酶活性。例如,在一些情况下,变体C2c2蛋白缺乏催化活性的HEPN1结构域。作为另一个实例,变体C2c2蛋白缺乏催化活性的HEPN2结构域。在一些情况下,变体C2c2蛋白缺乏催化活性的HEPN1结构域并且缺乏催化活性的HEPN2结构域。

[0224] 在一些情况下,变体C2c2多肽包含对SEQ ID NO:2(口腔纤毛菌C2c2)所示的氨基酸序列的氨基酸R472、H477、R1048和H1053中的1个、2个、3个或4个氨基酸的氨基酸取代,或SEQ ID NO:1、SEQ ID NO:4、SEQ ID NO:5或SEQ ID NO:6中所示的C2c2氨基酸序列的相应氨基酸的氨基酸取代。SEQ ID NO:1、SEQ ID NO:4、SEQ ID NO:5和SEQ ID NO:6中的相应氨基酸容易地被鉴别;参见例如图22B。例如,SEQ ID NO:1(斯氏李斯特菌C2c2)中对应于SEQ ID NO:2的R472、H477、R1048和H1053的氨基酸位置分别是R445、H450、R1016和H1021。作为另一实例,SEQ ID NO:4(荚膜红细菌C2c2)中对应于SEQ ID NO:2的R472、H477、R1048和H1053的氨基酸位置分别是R464、H469、R1052和H1057。作为另一实例,SEQ ID NO:5(鸡肉杆菌C2c2)中对应于SEQ ID NO:2的R472、H477、R1048和H1053的氨基酸位置分别是R467、H472、R1069和H1074。作为另一实例,SEQ ID NO:6(Herbinix hemicellulosilytica C2c2)中对应于SEQ ID NO:2的R472、H477、R1048和H1053的氨基酸位置分别是R472、H477、R1044和H1049。

[0225] 在一些情况下,变体C2c2多肽包含对SEQ ID NO:2所示的氨基酸序列的氨基酸R472和H477的氨基酸取代,或对SEQ ID NO:1、SEQ ID NO:4、SEQ ID NO:5或SEQ ID NO:6所示的C2c2氨基酸序列的相应氨基酸的氨基酸取代。在一些情况下,变体C2c2多肽包含对SEQ ID NO:2所示的氨基酸序列的氨基酸R1048和H1053的氨基酸取代,或对SEQ ID NO:1、SEQ ID NO:4、SEQ ID NO:5或SEQ ID NO:6所示的C2c2氨基酸序列的相应氨基酸的氨基酸取代。在一些情况下,变体C2c2多肽包含对SEQ ID NO:2所示的氨基酸序列的氨基酸R472、H477、

R1048和H1053的氨基酸取代,或对SEQ ID NO:1、SEQ ID NO:4、SEQ ID NO:5或SEQ ID NO:6所示的C2c2氨基酸序列的相应氨基酸的氨基酸取代。

[0226] 在一些情况下,变体C2c2多肽包含与SEQ ID NO:2所示氨基酸序列具有至少80%、至少85%、至少90%、至少95%、至少98%或至少99%的氨基酸序列同一性的氨基酸序列,并且包含对氨基酸R472和H477的取代。在一些情况下,位置472处的氨基酸是除了Arg以外的任何氨基酸;并且位置477处的氨基酸是除了His以外的任何氨基酸。在一些情况下,取代是R472A和H477A。在一些情况下,变体C2c2多肽包含与SEQ ID NO:2所示氨基酸序列具有至少80%、至少85%、至少90%、至少95%、至少98%或至少99%的氨基酸序列同一性的氨基酸序列,并且包含对氨基酸R1048和H1053的取代。在一些情况下,位置1048处的氨基酸是除了Arg以外的任何氨基酸;并且位置1053处的氨基酸是除了His以外的任何氨基酸。在一些情况下,取代是R1048A和H1053A。在一些情况下,变体C2c2多肽包含与SEQ ID NO:2所示氨基酸序列具有至少80%、至少85%、至少90%、至少95%、至少98%或至少99%的氨基酸序列同一性的氨基酸序列,并且包含对氨基酸R472、H477、R1048和H1053的取代。在一些情况下,位置472和1048处的氨基酸是除了Arg以外的任何氨基酸;并且位置477和1053处的氨基酸是除了His以外的任何氨基酸。在一些情况下,取代是R472A、H477A、R1048A和H1053A。在一些情况下,变体C2c2多肽具有减少的或不可检测的ss RNA切割(例如,RNA指导的切割活性),并保留结合C2c2指导RNA和ss RNA的能力。在一些情况下,变体C2c2多肽保留切割前体C2c2指导RNA的能力。在一些情况下,变体C2c2多肽具有减少的或不可检测的ss RNA切割(例如,RNA指导的切割活性),但保留了结合C2c2指导RNA和ssRNA的能力,并且保留了切割前体C2c2指导RNA的能力。

[0227] 在一些情况下,变体C2c2多肽包含与SEQ ID NO:1所示氨基酸序列具有至少80%、至少85%、至少90%、至少95%、至少98%或至少99%的氨基酸序列同一性的氨基酸序列,并且包含对氨基酸R445和H450的取代。在一些情况下,位置445处的氨基酸是除了Arg以外的任何氨基酸;并且位置450处的氨基酸是除了His以外的任何氨基酸。在一些情况下,取代是R445A和H450A。在一些情况下,变体C2c2多肽包含与SEQ ID NO:1所示氨基酸序列具有至少80%、至少85%、至少90%、至少95%、至少98%或至少99%的氨基酸序列同一性的氨基酸序列,并且包含对氨基酸R1016和H1021的取代。在一些情况下,位置1016处的氨基酸是除了Arg以外的任何氨基酸;并且位置1021处的氨基酸是除了His以外的任何氨基酸。在一些情况下,取代是R1016A和H1021A。在一些情况下,变体C2c2多肽包含与SEQ ID NO:1所示氨基酸序列具有至少80%、至少85%、至少90%、至少95%、至少98%或至少99%的氨基酸序列同一性的氨基酸序列,并且包含对氨基酸R445、H450、R1016和H1021的取代。在一些情况下,位置445和1016处的氨基酸是除了Arg以外的任何氨基酸;并且位置450和1016处的氨基酸是除了His以外的任何氨基酸。在一些情况下,取代是R445A、H450A、R1016A和H1021A。在一些情况下,变体C2c2多肽具有减少的或不可检测的ss RNA切割(例如,RNA指导的切割活性),并保留结合C2c2指导RNA和ss RNA的能力。在一些情况下,变体C2c2多肽保留切割前体C2c2指导RNA的能力。在一些情况下,变体C2c2多肽具有减少的或不可检测的ss RNA切割(例如,RNA指导的切割活性),但保留了结合C2c2指导RNA和ssRNA的能力,并且保留了切割前体C2c2指导RNA的能力。

[0228] 在一些情况下,变体C2c2多肽包含与SEQ ID NO:4(荚膜红细菌C2c2)所示氨基酸

序列具有至少80%、至少85%、至少90%、至少95%、至少98%或至少99%的氨基酸序列同一性的氨基酸序列,并且包含对氨基酸R464和H469的取代。在一些情况下,位置464处的氨基酸是除了Arg以外的任何氨基酸;并且位置469处的氨基酸是除了His以外的任何氨基酸。在一些情况下,取代是R464A和H469A。在一些情况下,变体C2c2多肽包含与SEQ ID NO:4所示氨基酸序列具有至少80%、至少85%、至少90%、至少95%、至少98%或至少99%的氨基酸序列同一性的氨基酸序列,并且包含对氨基酸R1052和H1057的取代。在一些情况下,位置1052处的氨基酸是除了Arg以外的任何氨基酸;并且位置1057处的氨基酸是除了His以外的任何氨基酸。在一些情况下,取代是R1052A和H1057A。在一些情况下,变体C2c2多肽包含与SEQ ID NO:4所示氨基酸序列具有至少80%、至少85%、至少90%、至少95%、至少98%或至少99%的氨基酸序列同一性的氨基酸序列,并且包含对氨基酸R464、H469、R1052和H1057的取代。在一些情况下,位置464和1052处的氨基酸是除了Arg以外的任何氨基酸;并且位置469和1057处的氨基酸是除了His以外的任何氨基酸。在一些情况下,取代是R464A、H469A、R1052A和H1057A。在一些情况下,变体C2c2多肽具有减少的或不可检测的ss RNA切割(例如,RNA指导的切割活性),并保留结合C2c2指导RNA和ss RNA的能力。在一些情况下,变体C2c2多肽保留切割前体C2c2指导RNA的能力。在一些情况下,变体C2c2多肽具有减少的或不可检测的ss RNA切割(例如,RNA指导的切割活性),但保留了结合C2c2指导RNA和ssRNA的能力,并且保留了切割前体C2c2指导RNA的能力。

[0229] 在一些情况下,变体C2c2多肽包含与SEQ ID NO:5(鸡肉杆菌C2c2)所示氨基酸序列具有至少80%、至少85%、至少90%、至少95%、至少98%或至少99%的氨基酸序列同一性的氨基酸序列,并且包含对氨基酸R467和H472的取代。在一些情况下,位置467处的氨基酸是除了Arg以外的任何氨基酸;并且位置472处的氨基酸是除了His以外的任何氨基酸。在一些情况下,取代是R469A和H472A。在一些情况下,变体C2c2多肽包含与SEQ ID NO:5所示氨基酸序列具有至少80%、至少85%、至少90%、至少95%、至少98%或至少99%的氨基酸序列同一性的氨基酸序列,并且包含对氨基酸R1069和H1074的取代。在一些情况下,位置1069处的氨基酸是除了Arg以外的任何氨基酸;并且位置1074处的氨基酸是除了His以外的任何氨基酸。在一些情况下,取代是R1069A和H1074A。在一些情况下,变体C2c2多肽包含与SEQ ID NO:5所示氨基酸序列具有至少80%、至少85%、至少90%、至少95%、至少98%或至少99%的氨基酸序列同一性的氨基酸序列,并且包含对氨基酸R467、H472、R1069和H1074的取代。在一些情况下,位置467和1069处的氨基酸是除了Arg以外的任何氨基酸;并且位置472和1074处的氨基酸是除了His以外的任何氨基酸。在一些情况下,取代是R469A、H472A、R1069A和H1074A。在一些情况下,变体C2c2多肽具有减少的或不可检测的ss RNA切割(例如,RNA指导的切割活性),并保留结合C2c2指导RNA和ss RNA的能力。在一些情况下,变体C2c2多肽保留切割前体C2c2指导RNA的能力。在一些情况下,变体C2c2多肽具有减少的或不可检测的ss RNA切割(例如,RNA指导的切割活性),但保留了结合C2c2指导RNA和ssRNA的能力,并且保留了切割前体C2c2指导RNA的能力。

[0230] 在一些情况下,变体C2c2多肽包含与SEQ ID NO:6(*Herbinix hemicellulosilytica* C2c2)所示氨基酸序列具有至少80%、至少85%、至少90%、至少95%、至少98%或至少99%的氨基酸序列同一性的氨基酸序列,并且包含对氨基酸R472和H477的取代。在一些情况下,位置472处的氨基酸是除了Arg以外的任何氨基酸;并且位置

477处的氨基酸是除了His以外的任何氨基酸。在一些情况下,取代是R472A和H477A。在一些情况下,变体C2c2多肽包含与SEQ ID NO:6所示氨基酸序列具有至少80%、至少85%、至少90%、至少95%、至少98%或至少99%的氨基酸序列同一性的氨基酸序列,并且包含对氨基酸R1044和H1049的取代。在一些情况下,位置1044处的氨基酸是除了Arg以外的任何氨基酸;并且位置1049处的氨基酸是除了His以外的任何氨基酸。在一些情况下,取代是R1044A和H1049A。在一些情况下,变体C2c2多肽包含与SEQ ID NO:6所示氨基酸序列具有至少80%、至少85%、至少90%、至少95%、至少98%或至少99%的氨基酸序列同一性的氨基酸序列,并且包含对氨基酸R472、H477、R1044和H1049的取代。在一些情况下,位置472和1044处的氨基酸是除了Arg以外的任何氨基酸;并且位置477和1049处的氨基酸是除了His以外的任何氨基酸。在一些情况下,取代是R472A、H477A、R1044A和H1049A。在一些情况下,变体C2c2多肽具有减少的或不可检测的ss RNA切割(例如,RNA指导的切割活性),并保留结合C2c2指导RNA和ss RNA的能力。在一些情况下,变体C2c2多肽保留切割前体C2c2指导RNA的能力。在一些情况下,变体C2c2多肽具有减少的或不可检测的ss RNA切割(例如,RNA指导的切割活性),但保留了结合C2c2指导RNA和ssRNA的能力,并且保留了切割前体C2c2指导RNA的能力。

[0231] 在一些情况下,变体C2c2多肽包含与SEQ ID NO:2所示氨基酸序列具有至少80%、至少85%、至少90%、至少95%、至少98%或至少99%氨基酸序列同一性的氨基酸序列,并且包含对氨基酸R472、H477、R1048和H1053中的1个、2个、3个或4个氨基酸的取代,使得变体C2c2多肽具有减少的或不可检测的ss RNA切割(例如,RNA指导的切割活性),并且保留结合C2c2指导RNA和ss RNA的能力。在一些情况下,变体C2c2多肽保留切割前体C2c2指导RNA的能力。例如,在一些情况下,变体C2c2多肽表现出由具有SEQ ID NO:2所示氨基酸序列的C2c2多肽表现出的RNA指导的非靶RNA切割的小于50%、小于40%、小于30%、小于20%、小于10%、小于5%、小于1%或小于0.1%。在一些情况下,变体C2c2多肽具有减少的或不可检测的ss RNA切割(例如,RNA指导的切割活性),但保留了结合C2c2指导RNA和ssRNA的能力,并且保留了切割前体C2c2指导RNA的能力。

[0232] 上述变体C2c2多肽中的任一者还可以包括突变(例如,在位置R1079、R1072和K1082中的任一者处,如下文进一步详细描述),所述突变导致降低的切割前体C2c2指导RNA的能力(例如,能力丧失)。

[0233] 在一些情况下,变体C2c2多肽具有降低的切割前体C2c2指导RNA的能力(例如,参见下面的实施例和相关的图26C至图26D、图35D和图37)。例如,在一些情况下,变体C2c2多肽包含对SEQ ID NO:2(口腔纤毛菌C2c2)所示的氨基酸序列的氨基酸R1079、R1072和K1082中的1个、2个或3个氨基酸的氨基酸取代,或者对任何C2c2氨基酸序列(例如,SEQ ID NO:1、SEQ ID NO:4、SEQ ID NO:5或SEQ ID NO:6所示的C2c2氨基酸序列)的相应氨基酸的氨基酸取代。SEQ ID NO:1、SEQ ID NO:4、SEQ ID NO:5和SEQ ID NO:6中的相应氨基酸容易被鉴别。例如,SEQ ID NO:1(斯氏李斯特菌C2c2)中对应于SEQ ID NO:2的R1079、R1072和K1082和K1082的氨基酸位置分别是R1048、R1041和K1051。作为另一实例,SEQ ID NO:4(荚膜红细菌C2c2)中对应于SEQ ID NO:2的R1079、R1072和K1082的氨基酸位置分别是R1085、R1078和K1088。作为另一实例,SEQ ID NO:5(鸡肉杆菌C2c2)中对应于SEQ ID NO:2的R1079、R1072和K1082的氨基酸位置分别是R1099、R1092和K1102。作为另一实例,SEQ ID NO:6(Herbinix

hemicellulosilytica C2c2) 中对应于SEQ ID NO:2的R1079和R1072的氨基酸位置分别是R1172和R1165。

[0234] 在一些情况下,变体C2c2多肽包含对SEQ ID NO:2所示的氨基酸序列的氨基酸R1079(例如,R1079A)的氨基酸取代,或对任何C2c2氨基酸序列(例如,SEQ ID NO:1、SEQ ID NO:4、SEQ ID NO:5或SEQ ID NO:6所示的C2c2氨基酸序列)的相应氨基酸的氨基酸取代。在一些情况下,变体C2c2多肽包含对SEQ ID NO:2所示的氨基酸序列的氨基酸R1072(例如,R1072A)的氨基酸取代,或对任何C2c2氨基酸序列(例如,SEQ ID NO:1、SEQ ID NO:4、SEQ ID NO:5或SEQ ID NO:6所示的C2c2氨基酸序列)的相应氨基酸的氨基酸取代。在一些情况下,变体C2c2多肽包含对SEQ ID NO:2所示的氨基酸序列的氨基酸K1082(例如,K1082A)的氨基酸取代,或对任何C2c2氨基酸序列(例如,SEQ ID NO:1、SEQ ID NO:4、SEQ ID NO:5或SEQ ID NO:6所示的C2c2氨基酸序列)的相应氨基酸的氨基酸取代。在一些情况下,变体C2c2多肽包含在SEQ ID NO:2所示的氨基酸序列的选自R1079(例如,R1079A)、R1072(例如,R1072A)和K1082(例如,K1082A)的位置处的一个或多个(例如,两个或更多个、或全部三个)氨基酸取代,或对任何C2c2氨基酸序列(例如,SEQ ID NO:1、SEQ ID NO:4、SEQ ID NO:5或SEQ ID NO:6所示的C2c2氨基酸序列)的相应氨基酸的氨基酸取代。

[0235] 在一些情况下,变体C2c2多肽包含与SEQ ID NO:2所示的氨基酸序列具有至少80%、至少85%、至少90%、至少95%、至少98%或至少99%的氨基酸序列同一性的氨基酸序列,并且包含对SEQ ID NO:2所示的氨基酸序列的氨基酸R1079(例如,R1079A)的氨基酸取代,或对任何C2c2氨基酸序列(例如,SEQ ID NO:1、SEQ ID NO:4、SEQ ID NO:5或SEQ ID NO:6所示的C2c2氨基酸序列)的相应氨基酸的氨基酸取代。在一些情况下,变体C2c2多肽包含与SEQ ID NO:2所示的氨基酸序列具有至少80%、至少85%、至少90%、至少95%、至少98%或至少99%的氨基酸序列同一性的氨基酸序列,并且包含对SEQ ID NO:2所示的氨基酸序列的氨基酸R1072(例如,R1072A)的氨基酸取代,或对任何C2c2氨基酸序列(例如,SEQ ID NO:1、SEQ ID NO:4、SEQ ID NO:5或SEQ ID NO:6所示的C2c2氨基酸序列)的相应氨基酸的氨基酸取代。在一些情况下,变体C2c2多肽包含与SEQ ID NO:2所示的氨基酸序列具有至少80%、至少85%、至少90%、至少95%、至少98%或至少99%的氨基酸序列同一性的氨基酸序列,并且包含在SEQ ID NO:2所示的氨基酸序列的选自R1079(例如,R1079A)、R1072(例如,R1072A)和K1082(例如,K1082A)的位置处的一个或多个(例如,两个或更多个、或全部三个)氨基酸取代,或对任何C2c2氨基酸序列(例如,SEQ ID NO:1、SEQ ID NO:4、SEQ ID NO:5或SEQ ID NO:6所示的C2c2氨基酸序列)的相应氨基酸的氨基酸取代。

[0236] 在一些情况下,变体C2c2多肽包含与SEQ ID NO:1所示的氨基酸序列具有至少80%、至少85%、至少90%、至少95%、至少98%或至少99%的氨基酸序列同一性的氨基酸序列,并且包含对SEQ ID NO:1所示的氨基酸序列的氨基酸R1041(例如,R1041A)的氨基酸取代,或对任何C2c2氨基酸序列(例如,SEQ ID NO:2、SEQ ID NO:4、SEQ ID NO:5或SEQ ID

N0:6所示的C2c2氨基酸序列)的相应氨基酸的氨基酸取代。在一些情况下,变体C2c2多肽包含与SEQ ID N0:1所示的氨基酸序列具有至少80%、至少85%、至少90%、至少95%、至少98%或至少99%的氨基酸序列同一性的氨基酸序列,并且包含对SEQ ID N0:1所示的氨基酸序列的氨基酸R1048(例如,R1048A)的氨基酸取代,或对任何C2c2氨基酸序列(例如,SEQ ID N0:2、SEQ ID N0:4、SEQ ID N0:5或SEQ ID N0:6所示的C2c2氨基酸序列)的相应氨基酸的氨基酸取代。在一些情况下,变体C2c2多肽包含与SEQ ID N0:1所示的氨基酸序列具有至少80%、至少85%、至少90%、至少95%、至少98%或至少99%的氨基酸序列同一性的氨基酸序列,并且包含对SEQ ID N0:1所示的氨基酸序列的氨基酸K1051(例如,K1051A)的氨基酸取代,或对任何C2c2氨基酸序列(例如,SEQ ID N0:2、SEQ ID N0:4、SEQ ID N0:5或SEQ ID N0:6所示的C2c2氨基酸序列)的相应氨基酸的氨基酸取代。在一些情况下,变体C2c2多肽包含与SEQ ID N0:1所示的氨基酸序列具有至少80%、至少85%、至少90%、至少95%、至少98%或至少99%的氨基酸序列同一性的氨基酸序列,并且包含在SEQ ID N0:1所示的氨基酸序列的选自R1048(例如,R1048A)、R1041(例如,R1041A)和K1051(例如,K1051A)的位置处的一个或多个(例如,两个或更多个、或全部三个)氨基酸取代,或对任何C2c2氨基酸序列(例如,SEQ ID N0:2、SEQ ID N0:4、SEQ ID N0:5或SEQ ID N0:6所示的C2c2氨基酸序列)的相应氨基酸的氨基酸取代。

[0237] 在一些情况下,变体C2c2多肽包含与SEQ ID N0:4所示的氨基酸序列具有至少80%、至少85%、至少90%、至少95%、至少98%或至少99%的氨基酸序列同一性的氨基酸序列,并且包含对SEQ ID N0:4所示的氨基酸序列的氨基酸R1085(例如,R1085A)的氨基酸取代,或对任何C2c2氨基酸序列(例如,SEQ ID N0:1、SEQ ID N0:2、SEQ ID N0:5或SEQ ID N0:6所示的C2c2氨基酸序列)的相应氨基酸的氨基酸取代。在一些情况下,变体C2c2多肽包含与SEQ ID N0:4所示的氨基酸序列具有至少80%、至少85%、至少90%、至少95%、至少98%或至少99%的氨基酸序列同一性的氨基酸序列,并且包含对SEQ ID N0:4所示的氨基酸序列的氨基酸R1078(例如,R1078A)的氨基酸取代,或对任何C2c2氨基酸序列(例如,SEQ ID N0:1、SEQ ID N0:2、SEQ ID N0:5或SEQ ID N0:6所示的C2c2氨基酸序列)的相应氨基酸的氨基酸取代。在一些情况下,变体C2c2多肽包含与SEQ ID N0:4所示的氨基酸序列具有至少80%、至少85%、至少90%、至少95%、至少98%或至少99%的氨基酸序列同一性的氨基酸序列,并且包含在SEQ ID N0:4所示的氨基酸序列的选自R1085(例如,R1085A)、R1078(例如,R1078A)和K1088(例如,K1088A)的位置处的一个或多个(例如,两个或更多个、或全部三个)氨基酸取代,或对任何C2c2氨基酸序列(例如,SEQ ID N0:1、SEQ ID N0:2、SEQ ID N0:5或SEQ ID N0:6所示的C2c2氨基酸序列)的相应氨基酸的氨基酸取代。

[0238] 在一些情况下,变体C2c2多肽包含与SEQ ID N0:5所示的氨基酸序列具有至少80%、至少85%、至少90%、至少95%、至少98%或至少99%的氨基酸序列同一性的氨基酸序列,并且包含对SEQ ID N0:5所示的氨基酸序列的氨基酸R1099(例如,R1099A)的氨基酸

取代,或对任何C2c2氨基酸序列(例如,SEQ ID NO:1、SEQ ID NO:4、SEQ ID NO:2或SEQ ID NO:6所示的C2c2氨基酸序列)的相应氨基酸的氨基酸取代。在一些情况下,变体C2c2多肽包含与SEQ ID NO:5所示的氨基酸序列具有至少80%、至少85%、至少90%、至少95%、至少98%或至少99%的氨基酸序列同一性的氨基酸序列,并且包含对SEQ ID NO:5所示的氨基酸序列的氨基酸R1092(例如,R1092A)的氨基酸取代,或对任何C2c2氨基酸序列(例如,SEQ ID NO:1、SEQ ID NO:4、SEQ ID NO:2或SEQ ID NO:6所示的C2c2氨基酸序列)的相应氨基酸的氨基酸取代。在一些情况下,变体C2c2多肽包含与SEQ ID NO:5所示的氨基酸序列具有至少80%、至少85%、至少90%、至少95%、至少98%或至少99%的氨基酸序列同一性的氨基酸序列,并且包含对SEQ ID NO:5所示的氨基酸序列的氨基酸K1102(例如,K1102A)的氨基酸取代,或对任何C2c2氨基酸序列(例如,SEQ ID NO:1、SEQ ID NO:4、SEQ ID NO:2或SEQ ID NO:6所示的C2c2氨基酸序列)的相应氨基酸的氨基酸取代。在一些情况下,变体C2c2多肽包含与SEQ ID NO:5所示的氨基酸序列具有至少80%、至少85%、至少90%、至少95%、至少98%或至少99%的氨基酸序列同一性的氨基酸序列,并且包含在SEQ ID NO:5所示的氨基酸序列的选自R1099(例如,R1099A)、R1092(例如,R1092A)和K1102(例如,K1102A)的位置处的一个或多个(例如,两个或更多个、或全部三个)氨基酸取代,或对任何C2c2氨基酸序列(例如,SEQ ID NO:1、SEQ ID NO:4、SEQ ID NO:2或SEQ ID NO:6所示的C2c2氨基酸序列)的相应氨基酸的氨基酸取代。

[0239] 在一些情况下,变体C2c2多肽包含与SEQ ID NO:6所示的氨基酸序列具有至少80%、至少85%、至少90%、至少95%、至少98%或至少99%的氨基酸序列同一性的氨基酸序列,并且包含对SEQ ID NO:6所示的氨基酸序列的氨基酸R1172(例如,R1172A)的氨基酸取代,或对任何C2c2氨基酸序列(例如,SEQ ID NO:1、SEQ ID NO:4、SEQ ID NO:5或SEQ ID NO:2所示的C2c2氨基酸序列)的相应氨基酸的氨基酸取代。在一些情况下,变体C2c2多肽包含与SEQ ID NO:6所示的氨基酸序列具有至少80%、至少85%、至少90%、至少95%、至少98%或至少99%的氨基酸序列同一性的氨基酸序列,并且包含对SEQ ID NO:6所示的氨基酸序列的氨基酸R1165(例如,R1165A)的氨基酸取代,或对任何C2c2氨基酸序列(例如,SEQ ID NO:1、SEQ ID NO:4、SEQ ID NO:5或SEQ ID NO:2所示的C2c2氨基酸序列)的相应氨基酸的氨基酸取代。在一些情况下,变体C2c2多肽包含与SEQ ID NO:6所示的氨基酸序列具有至少80%、至少85%、至少90%、至少95%、至少98%或至少99%的氨基酸序列同一性的氨基酸序列,并且包含在SEQ ID NO:6所示的氨基酸序列的选自R1172(例如,R1172A)和R1165(例如,R1165A)的位置处的一个或多个(例如,两个)氨基酸取代,或对任何C2c2氨基酸序列(例如,SEQ ID NO:1、SEQ ID NO:4、SEQ ID NO:5或SEQ ID NO:2所示的C2c2氨基酸序列)的相应氨基酸的氨基酸取代。

[0240] C2c2指导RNA

[0241] 主题C2c2指导RNA(例如,C2c2 crRNA)包括指导序列和恒定区(例如,指导序列的5'端的区域)。指导序列的5'端的区域与C2c2蛋白结合(并且可以被认为是蛋白结合区),而指导序列与靶RNA的靶序列杂交。

[0242] 指导序列

[0243] 指导序列与单链靶RNA的靶序列具有互补性(与单链靶RNA的靶序列杂交)。在一些情况下,直接位于靶序列(原型间隔区)3'端的靶RNA的碱基不是G。在一些情况下,指导序列

的长度为16-28个核苷酸(nt) (例如,长度为16-26nt、16-24nt、16-22nt、16-20nt、16-18nt、17-26nt、17-24nt、17-22nt、17-20nt、17-18nt、18-26nt、18-24或18-22nt)。在一些情况下,指导序列的长度为18-24个核苷酸(nt)。在一些情况下,指导序列的长度为至少16nt (例如,长度为至少18nt、20nt或22nt)。在一些情况下,指导序列的长度为至少17nt。在一些情况下,指导序列的长度为至少18nt。在一些情况下,指导序列的长度为至少20nt。

[0244] 在一些情况下,指导序列与单链靶RNA的靶序列具有80%或更多(例如,85%或更多、90%或更多、95%或更多、或100%)的互补性。在一些情况下,指导序列与单链靶RNA的靶序列100%互补。

[0245] 恒定区

[0246] 以下3个序列各自是天然存在的C2c2指导RNA的恒定区的实例(例如,在指导序列的5'端的区域):

[0247] GACUACCUCUAUAUGAAAGAGGACUAAAAC (SEQ ID NO:7)

[0248] (斯氏李斯特菌) (“Lse”)

[0249] CCACCCCAUAUCGAAGGGGACUAAAACA (SEQ ID NO:8)

[0250] (沙氏纤毛菌) (“Lsh”)

[0251] GACCACCCCAAAAAUGAAGGGGACUAAAACA (SEQ ID NO:9)

[0252] (口腔纤毛菌) (“Lbu”)

[0253] 在一些实施方案中,主题C2c2指导RNA包括与SEQ ID NO:7至SEQ ID NO:9中任一项所示的序列具有70%或更多同一性(例如,80%或更多、85%或更多、90%或更多、95%或更多、98%或更多、99%或更多、或100%同一性)的核苷酸序列。在一些实施方案中,主题C2c2指导RNA包括与SEQ ID NO:7至SEQ ID NO:9中任一项所示的序列具有90%或更多同一性(例如,95%或更多、98%或更多、99%或更多、或100%同一性)的核苷酸序列。在一些实施方案中,主题C2c2指导RNA包含SEQ ID NO:7至SEQ ID NO:9中任一者所示的核苷酸序列。

[0254] 在一些实施方案中,主题C2c2指导RNA包括与SEQ ID NO:9所示的序列具有70%或更多同一性(例如,80%或更多、85%或更多、90%或更多、95%或更多、98%或更多、99%或更多、或100%同一性)的核苷酸序列。在一些实施方案中,主题C2c2指导RNA包括与SEQ ID NO:9中所示的序列具有90%或更多同一性(例如,95%或更多、98%或更多、99%或更多、或100%同一性)的核苷酸序列。在一些实施方案中,主题C2c2指导RNA包括SEQ ID NO:9所示的核苷酸序列。

[0255] 在一些实施方案中,主题C2c2指导RNA不包括沙氏纤毛菌(Lsh) C2c2指导RNA的核苷酸序列。例如,在一些情况下,所使用的C2c2蛋白不是来自沙氏纤毛菌的C2c2(例如,不是Lsh C2c2蛋白),并且在一些此类情况下,所使用的C2c2指导RNA也不来自沙氏纤毛菌(例如,所使用的指导RNA不包括Lsh C2c2指导RNA的恒定区)。因此,在一些情况下,主题C2c2指导RNA不包括SEQ ID NO:8所示的序列。

[0256] 在一些情况下,C2c2指导RNA包括双链RNA双链体(dsRNA双链体)。例如,参见图7A,其示出了来自Lbu的C2c2指导RNA与单链靶RNA杂交,其中所述C2c2指导RNA包括长度为4个碱基对(bp)的dsRNA双链体。在一些情况下,C2c2指导RNA包括长度为2至12bp(例如,2至10bp、2至8bp、2至6bp、2至5bp、2至4bp、3至12bp、3至10bp、3至8bp、3至6bp、3至5bp、3至4bp、4至12bp、4至10bp、4至8bp、4至6bp或4至5bp)的dsRNA双链体。在一些情况下,C2c2指导RNA

包括长度为2bp或更长(例如,长度为3bp或更长、4bp或更长、5bp或更长、6bp或更长、或7bp或更长)的dsRNA双链体。在一些情况下,C2c2指导RNA包括比相应的野生型C2c2指导RNA的dsRNA双链体更长的dsRNA双链体。例如,参见图7A,其示出了来自Lbu的C2c2指导RNA与单链靶RNA杂交,其中所述C2c2指导RNA包括长度为4个碱基对(bp)的dsRNA双链体。因此,C2c2指导RNA可以在一些情况下包括长度为5bp或更长(例如,长度为6bp或更长、7bp或更长、或8bp或更长)的dsRNA双链体。在一些情况下,C2c2指导RNA包括比相应的野生型C2c2指导RNA的dsRNA双链体更短的dsRNA双链体。因此,在一些情况下,C2c2指导RNA包括长度小于4bp的dsRNA双链体。在一些情况下,C2c2指导RNA包括长度为2或3bp的dsRNA双链体。

[0257] 在一些情况下,C2c2指导RNA的在指导序列的5'端的区域的长度为15个或更多个核苷酸(nt)(例如,长度为18nt或更长、20nt或更长、21nt或更长、22nt或更长、23nt或更长、24nt或更长、25nt或更长、26nt或更长、27nt或更长、28nt或更长、29nt或更长、30nt或更长、31nt或更长、32nt或更长、33nt或更长、34nt或更长、或35nt或更长)。在一些情况下,C2c2指导RNA的在指导序列的5'端的区域的长度为29nt或更长。

[0258] 在一些情况下,C2c2指导RNA的在指导序列的5'端的区域的长度在12至100nt(例如,12至90nt、12至80nt、12至70nt、12至60nt、12至50nt、12至40nt、15至100nt、15至90nt、15至80nt、15至70nt、15至60nt、15至50nt、15至40nt、20至100nt、20至90nt、20至80nt、20至70nt、20至60nt、20至50nt、20至40nt、25至100nt、25至90nt、25至80nt、25至70nt、25至60nt、25至50nt、25至40nt、28至100nt、28至90nt、28至80nt、28至70nt、28至60nt、28至50nt、28至40nt、29至100nt、29至90nt、29至80nt、29至70nt、29至60nt、29至50nt或29至40nt)的范围内。在一些情况下,C2c2指导RNA的在指导序列的5'端的区域的长度在28至100nt的范围内。在一些情况下,C2c2指导RNA的在指导序列的5'端的区域的长度在28至40nt的范围内。

[0259] 在一些情况下,C2c2指导RNA的在指导序列的5'端的区域相对于(短于)相应野生型C2c2指导RNA的相应区域被截短。例如,成熟的Lse C2c2指导RNA包括在指导序列的5'端的长度为30个核苷酸(nt)的区域,并且因此主题截短C2c2指导RNA(相对于Lse C2c2指导RNA)可以具有在指导序列的5'端的长度小于30nt(例如,长度小于29nt、28nt、27nt、26nt、25nt、22nt或20nt)的区域。在一些情况下,截短C2c2指导RNA包括在指导序列的5'端的长度范围为12至29nt(例如,12至28nt、12至27nt、12至26nt、12至25nt、12至22nt、12至20nt、12至18nt)的区域。在一些情况下,截短C2c2指导RNA例如相对于相应的野生型C2c2指导物被截短一个或多个nt(例如,2nt或更长、3nt或更长、4nt或更长、5nt或更长、或10nt或更长)。

[0260] 在一些情况下,C2c2指导RNA的在指导序列的5'端的区域相对于(长于)相应野生型C2c2指导RNA的相应区域被延长。例如,成熟的Lse C2c2指导RNA包括在指导序列的5'端的长度为30个核苷酸(nt)的区域,并且因此延长的C2c2指导RNA(相对于Lse C2c2指导RNA)可以具有在指导序列的5'端的长于30nt(例如,长于31nt、长于32nt、长于33nt、长于34nt或长于35nt)的区域。在一些情况下,延长的C2c2指导RNA包括在指导序列的5'端的长度范围为30至100nt(例如,30至90nt、30至80nt、30至70nt、30至60nt、30至50nt或30至40nt)的区域。在一些情况下,延伸的C2c2指导RNA包括在指导序列的5'端的(例如,相对于相应野生型C2c2指导RNA的相应区域)延伸一个或多个nt(例如,2nt或更长、3nt或更长、4nt或更长、5nt或更长、或10nt或更长)的区域。

[0261] 在一些情况下,主题C2c2指导RNA的长度为30个或更多个核苷酸(nt)(例如,长度为34nt或更长、40nt或更长、45nt或更长、50nt或更长、55nt或更长、60nt或更长、65nt或更长、70nt或更长、或80nt或更长)。在一些情况下,C2c2指导RNA的长度为35nt或更长。

[0262] 在一些情况下,主题C2c2指导RNA的长度在30至120nt(例如,30至110nt、30至100nt、30至90nt、30至80nt、30至70nt、30至60nt、35至120nt、35至110nt、35至100nt、35至90nt、35至80nt、35至70nt、35至60nt、40至120nt、40至110nt、40至100nt、40至90nt、40至80nt、40至70nt、40至60nt、50至120nt、50至110nt、50至100nt、50至90nt、50至80nt或50至70nt)的范围内。在一些情况下,C2c2指导RNA的长度在33至80nt的范围内。在一些情况下,C2c2指导RNA的长度在35至60nt的范围内。

[0263] 在一些情况下,主题C2c2指导RNA相对于(短于)相应野生型C2c2指导RNA被截短。例如,成熟的Lse C2c2指导RNA的长度可以是50个核苷酸(nt),并且因此截短的C2c2指导RNA(相对于Lse C2c2指导RNA)的长度在一些情况下可以小于50nt(例如,长度为小于49nt、48nt、47nt、46nt、45nt、42nt或40nt)。在一些情况下,截短的C2c2指导RNA的长度在30至49nt(例如,30至48nt、30至47nt、30至46nt、30至45nt、30至42nt、30至40nt、35至49nt、35至48nt、35至47nt、35至46nt、35至45nt、35至42nt、或35至40nt)的范围内。在一些情况下,截短的C2c2指导RNA例如相对于相应的野生型C2c2指导物被截短一个或多个nt(例如,2nt或更长、3nt或更长、4nt或更长、5nt或更长、或10nt或更长)。

[0264] 在一些情况下,主题C2c2指导RNA相对于(长于)相应的野生型C2c2指导RNA延长。例如,成熟的Lse C2c2指导RNA的长度可以为50个核苷酸(nt),并且因此延长的C2c2指导RNA(相对于Lse C2c2指导RNA)在一些情况下可以长于50nt(例如,长于51nt、长于52nt、长于53nt、长于54nt或长于55nt)。在一些情况下,延长的C2c2指导RNA的长度在51至100nt(例如,51至90nt、51至80nt、51至70nt、51至60nt、53至100nt、53至90nt、53至80nt、53至70nt、53至60nt、55至100nt、55至90nt、55至80nt、55至70nt或55至60nt)的范围内。在一些情况下,延长的C2c2指导RNA(例如,相对于相应野生型C2c2指导RNA)延长一个或多个nt(例如,2nt或更长、3nt或更长、4nt或更长、5nt或更长、或10nt或更长)。

[0265] 切割前体C2C2指导RNA阵列的方法

[0266] 本公开提供了将前体C2c2指导RNA阵列切割成两个或更多个C2c2指导RNA的方法。该方法包括使前体C2c2指导RNA阵列与C2c2蛋白接触。前体C2c2指导RNA阵列包含两个或更多个(例如,2个、3个、4个、5个或更多个)C2c2指导RNA,每个所述C2c2指导RNA可以具有不同的指导序列。C2c2蛋白将前体C2c2指导RNA阵列切割成单独的C2c2指导RNA。在一些情况下,C2c2指导RNA的恒定区(也称为‘柄’)包括来自前体指导RNA的核苷酸序列(例如,在切割指导RNA阵列之前通常存在的序列)。换句话说,在一些情况下,主题C2c2指导RNA的恒定区包括前体crRNA柄。

[0267] 在一些情况下,接触步骤不在细胞内(例如在活细胞内)发生。在一些情况下,接触步骤在细胞(例如,体外细胞(培养中)、离体细胞、体内细胞)内发生。任何细胞都是合适的。其中可以发生接触的细胞的实例包括但不限于:真核细胞;原核细胞(例如细菌细胞、古细菌细胞);单细胞真核生物;植物细胞;藻类细胞,例如布朗葡萄藻、莱茵衣藻、迦得拟微绿球藻、蛋白核小球藻、展枝马尾藻、羽藻等等;真菌细胞(例如,酵母细胞);动物细胞;无脊椎动物细胞(例如,果蝇、刺胞动物、棘皮动物、线虫、昆虫、蜘蛛类动物,等);脊椎动物细胞(例

如,鱼类、两栖动物、爬行动物、鸟类、哺乳动物);哺乳动物细胞(例如,人类;非人灵长类动物;有蹄类动物;猫科动物;牛;绵羊;山羊;大鼠;小鼠;啮齿动物;猪;绵羊;牛;等等);寄生虫细胞(例如,蠕虫、疟疾寄生虫等)。

[0268] C2c2蛋白

[0269] 当C2c2蛋白具有完整的HEPN结构域时,其可以在被‘激活’后切割RNA(靶RNA以及非靶RNA)。然而,C2c2蛋白也可以以不依赖HEPN的方式将前体C2c2指导RNA切割成成熟C2c2指导RNA。例如,当C2c2蛋白缺乏催化活性的HEPN1结构域并且也缺乏催化活性的HEPN2结构域时,其仍然可以将前体指导RNA切割成成熟的指导RNA。因此,当在包括前体C2c2指导RNA和/或前体C2c2指导RNA阵列的方法中使用时,C2c2蛋白可以(并且在一些情况下将)缺乏催化活性的HEPN1结构域和/或催化活性的HEPN2结构域。在一些情况下,C2c2蛋白缺乏催化活性的HEPN1结构域并且缺乏催化活性的HEPN2结构域。

[0270] 缺乏催化活性HEPN1结构域并且缺乏催化活性HEPN2结构域的C2c2蛋白在一些情况下可用于结合方法(例如,成像方法)中。例如,在一些情况下,结合(和/或成像)方法包括使样品与前体C2c2指导RNA阵列和缺乏催化活性HEPN1结构域且缺乏催化活性HEPN2结构域的C2c2蛋白接触。在此类情况下,可以检测地标记C2c2蛋白(例如,融合表位标签、与荧光团融合、与荧光蛋白(诸如绿色荧光蛋白)融合等)。

[0271] 适用于在本公开的方法中的用于切割前体C2c2指导RNA阵列的C2c2蛋白可具有完整的HEPN1和HEPN2结构域。然而,在一些情况下,C2c2蛋白缺乏催化活性的HEPN1结构域和/或缺乏催化活性的HEPN2结构域。

[0272] 在一些情况下,适用于在本公开的方法中的用于切割前体C2c2指导RNA阵列的C2c2蛋白包括与SEQ ID NO:1至SEQ ID NO:6中任一项所示的氨基酸序列具有80%或更多(例如,85%或更多、90%或更多、95%或更多、98%或更多、99%或更多、99.5%或更多、或100%的氨基酸序列同一性的氨基酸序列。在一些情况下,适用于在本公开的方法中的用于切割前体C2c2指导RNA阵列的C2c2蛋白包含与SEQ ID NO:1所示的斯氏李斯特菌C2c2氨基酸序列具有至少80%、至少85%、至少90%、至少95%、至少98%、至少99%或100%的氨基酸序列同一性的氨基酸序列。在一些情况下,适用于在本公开的方法中的用于切割前体C2c2指导RNA阵列的C2c2蛋白包含与SEQ ID NO:2所示的口腔纤毛菌C2c2氨基酸序列具有至少80%、至少85%、至少90%、至少95%、至少98%、至少99%或100%的氨基酸序列同一性的氨基酸序列。在一些情况下,适用于在本公开的方法中的用于切割前体C2c2指导RNA阵列的C2c2蛋白包含与SEQ ID NO:4所示的荚膜红细菌C2c2氨基酸序列具有至少80%、至少85%、至少90%、至少95%、至少98%、至少99%或100%的氨基酸序列同一性的氨基酸序列。在一些情况下,适用于在本公开的方法中的用于切割前体C2c2指导RNA阵列的C2c2蛋白包含与SEQ ID NO:5所示的鸡肉杆菌C2c2氨基酸序列具有至少80%、至少85%、至少90%、至少95%、至少98%、至少99%或100%的氨基酸序列同一性的氨基酸序列。在一些情况下,适用于在本公开的方法中的用于切割前体C2c2指导RNA阵列的C2c2蛋白包含与SEQ ID NO:6所示的Herbinix hemicellulosilytica C2c2氨基酸序列具有至少80%、至少85%、至少90%、至少95%、至少98%、至少99%或100%的氨基酸序列同一性的氨基酸序列。在一些情况下,适用于在本公开的方法中的用于切割前体C2c2指导RNA阵列的C2c2蛋白包含与SEQ ID NO:2所示的口腔纤毛菌(Lbu) C2c2氨基酸序列具有80%或更多的氨基酸序列同一性的

氨基酸序列。在一些情况下,适用于包括在本公开的试剂盒中的C2c2蛋白是口腔纤毛菌(Lbu) C2c2蛋白(例如,参见SEQ ID NO:2)。在一些情况下,适用于在本公开的方法中的用于切割前体C2c2指导RNA阵列的C2c2蛋白包括SEQ ID NO:1至SEQ ID NO:2和SEQ ID NO:4至SEQ ID NO:6中任一项所示的氨基酸序列。

[0273] 在一些情况下,本公开的方法中所使用的用于切割前体C2c2指导RNA阵列的C2c2蛋白不是沙氏纤毛菌(Lsh) C2c2蛋白。在一些情况下,本公开的方法中所使用的用于切割前体C2c2指导RNA阵列的C2c2蛋白不是与SEQ ID NO:3所示的Lsh C2c2多肽具有至少80%、至少85%、至少90%、至少95%、至少98%、至少99%或100%的氨基酸序列同一性的C2c2多肽。

[0274] 在一些情况下,适用于在本公开的方法中的用于切割前体C2c2指导RNA阵列的C2c2多肽是变体C2c2多肽。适用于在本公开的方法中的用于切割前体C2c2指导RNA阵列的变体C2c2多肽包括SEQ ID NO:1、SEQ ID NO:2和SEQ ID NO:4至SEQ ID NO:6中任一者的变体,其中变体C2c2多肽表现出降低的(或不可检测的)核酸酶活性。例如,在一些情况下,变体C2c2蛋白缺乏催化活性的HEPN1结构域。作为另一个实例,变体C2c2蛋白缺乏催化活性的HEPN2结构域。在一些情况下,变体C2c2蛋白缺乏催化活性的HEPN1结构域并且缺乏催化活性的HEPN2结构域。

[0275] 在一些情况下,变体C2c2多肽包含对SEQ ID NO:2(口腔纤毛菌C2c2)所示的氨基酸序列的氨基酸R472、H477、R1048和H1053中的1个、2个、3个或4个氨基酸的氨基酸取代,或SEQ ID NO:1、SEQ ID NO:4、SEQ ID NO:5或SEQ ID NO:6中所示的C2c2氨基酸序列的相应氨基酸的氨基酸取代。SEQ ID NO:1、SEQ ID NO:4、SEQ ID NO:5和SEQ ID NO:6中的相应氨基酸容易被鉴别;参见例如图22B。例如,SEQ ID NO:1(斯氏李斯特菌C2c2)中对应于SEQ ID NO:2的R472、H477、R1048和H1053的氨基酸位置分别是R445、H450、R1016和H1021。

[0276] 在一些情况下,变体C2c2多肽包含对SEQ ID NO:2所示的氨基酸序列的氨基酸R472和H477的氨基酸取代,或对SEQ ID NO:1、SEQ ID NO:4、SEQ ID NO:5或SEQ ID NO:6所示的C2c2氨基酸序列的相应氨基酸的氨基酸取代。在一些情况下,变体C2c2多肽包含对SEQ ID NO:2所示的氨基酸序列的氨基酸R1048和H1053的氨基酸取代,或对SEQ ID NO:1、SEQ ID NO:4、SEQ ID NO:5或SEQ ID NO:6所示的C2c2氨基酸序列的相应氨基酸的氨基酸取代。在一些情况下,变体C2c2多肽包含对SEQ ID NO:2所示的氨基酸序列的氨基酸R472、H477、R1048和H1053的氨基酸取代,或对SEQ ID NO:1、SEQ ID NO:4、SEQ ID NO:5或SEQ ID NO:6所示的C2c2氨基酸序列的相应氨基酸的氨基酸取代。

[0277] 在一些情况下,变体C2c2多肽包含与SEQ ID NO:2所示氨基酸序列具有至少80%、至少85%、至少90%、至少95%、至少98%或至少99%的氨基酸序列同一性的氨基酸序列,并且包含对氨基酸R472和H477的取代。在一些情况下,位置472处的氨基酸是除了Arg以外的任何氨基酸;并且位置477处的氨基酸是除了His以外的任何氨基酸。在一些情况下,取代是R472A和H477A。在一些情况下,变体C2c2多肽包含与SEQ ID NO:2所示氨基酸序列具有至少80%、至少85%、至少90%、至少95%、至少98%或至少99%的氨基酸序列同一性的氨基酸序列,并且包含对氨基酸R1048和H1053的取代。在一些情况下,位置1048处的氨基酸是除了Arg以外的任何氨基酸;并且位置1053处的氨基酸是除了His以外的任何氨基酸。在一些情况下,取代是R1048A和H1053A。在一些情况下,变体C2c2多肽包含与SEQ ID NO:2所示氨

氨基酸序列具有至少80%、至少85%、至少90%、至少95%、至少98%或至少99%的氨基酸序列同一性的氨基酸序列,并且包含对氨基酸R472、H477、R1048和H1053的取代。在一些情况下,位置472和1048处的氨基酸是除了Arg以外的任何氨基酸;并且位置477和1053处的氨基酸是除了His以外的任何氨基酸。在一些情况下,取代是R472A、H477A、R1048A和H1053A。在一些情况下,变体C2c2多肽具有减少的或不可检测的ss RNA切割(例如,RNA指导的切割活性),并保留结合C2c2指导RNA和ss RNA的能力。在一些情况下,变体C2c2多肽保留切割前体C2c2指导RNA的能力。在一些情况下,变体C2c2多肽具有减少的或不可检测的ss RNA切割(例如,RNA指导的切割活性),但保留了结合C2c2指导RNA和ssRNA的能力,并且保留了切割前体C2c2指导RNA的能力。

[0278] 在一些情况下,变体C2c2多肽包含与SEQ ID NO:1所示氨基酸序列具有至少80%、至少85%、至少90%、至少95%、至少98%或至少99%的氨基酸序列同一性的氨基酸序列,并且包含对氨基酸R445和H450的取代。在一些情况下,位置445处的氨基酸是除了Arg以外的任何氨基酸;并且位置450处的氨基酸是除了His以外的任何氨基酸。在一些情况下,取代是R445A和H450A。在一些情况下,变体C2c2多肽包含与SEQ ID NO:1所示氨基酸序列具有至少80%、至少85%、至少90%、至少95%、至少98%或至少99%的氨基酸序列同一性的氨基酸序列,并且包含对氨基酸R1016和H1021的取代。在一些情况下,位置1016处的氨基酸是除了Arg以外的任何氨基酸;并且位置1021处的氨基酸是除了His以外的任何氨基酸。在一些情况下,取代是R1016A和H1021A。在一些情况下,变体C2c2多肽包含与SEQ ID NO:1所示氨基酸序列具有至少80%、至少85%、至少90%、至少95%、至少98%或至少99%的氨基酸序列同一性的氨基酸序列,并且包含对氨基酸R445、H450、R1016和H1021的取代。在一些情况下,位置445和1016处的氨基酸是除了Arg以外的任何氨基酸;并且位置450和1016处的氨基酸是除了His以外的任何氨基酸。在一些情况下,取代是R445A、H450A、R1016A和H1021A。在一些情况下,变体C2c2多肽具有减少的或不可检测的ss RNA切割(例如,RNA指导的切割活性),并保留结合C2c2指导RNA和ss RNA的能力。在一些情况下,变体C2c2多肽保留切割前体C2c2指导RNA的能力。在一些情况下,变体C2c2多肽具有减少的或不可检测的ss RNA切割(例如,RNA指导的切割活性),但保留了结合C2c2指导RNA和ssRNA的能力,并且保留了切割前体C2c2指导RNA的能力。

[0279] 在一些情况下,变体C2c2多肽包含与SEQ ID NO:4(荚膜红细菌C2c2)所示氨基酸序列具有至少80%、至少85%、至少90%、至少95%、至少98%或至少99%的氨基酸序列同一性的氨基酸序列,并且包含对氨基酸R464和H469的取代。在一些情况下,位置464处的氨基酸是除了Arg以外的任何氨基酸;并且位置469处的氨基酸是除了His以外的任何氨基酸。在一些情况下,取代是R464A和H469A。在一些情况下,变体C2c2多肽包含与SEQ ID NO:4所示氨基酸序列具有至少80%、至少85%、至少90%、至少95%、至少98%或至少99%的氨基酸序列同一性的氨基酸序列,并且包含对氨基酸R1052和H1057的取代。在一些情况下,位置1052处的氨基酸是除了Arg以外的任何氨基酸;并且位置1057处的氨基酸是除了His以外的任何氨基酸。在一些情况下,取代是R1052A和H1057A。在一些情况下,变体C2c2多肽包含与SEQ ID NO:4所示氨基酸序列具有至少80%、至少85%、至少90%、至少95%、至少98%或至少99%的氨基酸序列同一性的氨基酸序列,并且包含对氨基酸R464、H469、R1052和H1057的取代。在一些情况下,位置464和1052处的氨基酸是除了Arg以外的任何氨基酸;并且位置

469和1057处的氨基酸是除了His以外的任何氨基酸。在一些情况下,取代是R464A、H469A、R1052A和H1057A。在一些情况下,变体C2c2多肽具有减少的或不可检测的ss RNA切割(例如,RNA指导的切割活性),并保留结合C2c2指导RNA和ss RNA的能力。在一些情况下,变体C2c2多肽保留切割前体C2c2指导RNA的能力。在一些情况下,变体C2c2多肽具有减少的或不可检测的ss RNA切割(例如,RNA指导的切割活性),但保留了结合C2c2指导RNA和ssRNA的能力,并且保留了切割前体C2c2指导RNA的能力。

[0280] 在一些情况下,变体C2c2多肽包含与SEQ ID NO:5(鸡肉杆菌C2c2)所示氨基酸序列具有至少80%、至少85%、至少90%、至少95%、至少98%或至少99%的氨基酸序列同一性的氨基酸序列,并且包含对氨基酸R467和H472的取代。在一些情况下,位置467处的氨基酸是除了Arg以外的任何氨基酸;并且位置472处的氨基酸是除了His以外的任何氨基酸。在一些情况下,取代是R469A和H472A。在一些情况下,变体C2c2多肽包含与SEQ ID NO:5所示氨基酸序列具有至少80%、至少85%、至少90%、至少95%、至少98%或至少99%的氨基酸序列同一性的氨基酸序列,并且包含对氨基酸R1069和H1074的取代。在一些情况下,位置1069处的氨基酸是除了Arg以外的任何氨基酸;并且位置1074处的氨基酸是除了His以外的任何氨基酸。在一些情况下,取代是R1069A和H1074A。在一些情况下,变体C2c2多肽包含与SEQ ID NO:5所示氨基酸序列具有至少80%、至少85%、至少90%、至少95%、至少98%或至少99%的氨基酸序列同一性的氨基酸序列,并且包含对氨基酸R467、H472、R1069和H1074的取代。在一些情况下,位置467和1069处的氨基酸是除了Arg以外的任何氨基酸;并且位置472和1074处的氨基酸是除了His以外的任何氨基酸。在一些情况下,取代是R469A、H472A、R1069A和H1074A。在一些情况下,变体C2c2多肽具有减少的或不可检测的ss RNA切割(例如,RNA指导的切割活性),并保留结合C2c2指导RNA和ss RNA的能力。在一些情况下,变体C2c2多肽保留切割前体C2c2指导RNA的能力。在一些情况下,变体C2c2多肽具有减少的或不可检测的ss RNA切割(例如,RNA指导的切割活性),但保留了结合C2c2指导RNA和ssRNA的能力,并且保留了切割前体C2c2指导RNA的能力。

[0281] 在一些情况下,变体C2c2多肽包含与SEQ ID NO:6(*Herbinix hemicellulosilytica* C2c2)所示氨基酸序列具有至少80%、至少85%、至少90%、至少95%、至少98%或至少99%的氨基酸序列同一性的氨基酸序列,并且包含对氨基酸R472和H477的取代。在一些情况下,位置472处的氨基酸是除了Arg以外的任何氨基酸;并且位置477处的氨基酸是除了His以外的任何氨基酸。在一些情况下,取代是R472A和H477A。在一些情况下,变体C2c2多肽包含与SEQ ID NO:6所示氨基酸序列具有至少80%、至少85%、至少90%、至少95%、至少98%或至少99%的氨基酸序列同一性的氨基酸序列,并且包含对氨基酸R1044和H1049的取代。在一些情况下,位置1044处的氨基酸是除了Arg以外的任何氨基酸;并且位置1049处的氨基酸是除了His以外的任何氨基酸。在一些情况下,取代是R1044A和H1049A。在一些情况下,变体C2c2多肽包含与SEQ ID NO:6所示氨基酸序列具有至少80%、至少85%、至少90%、至少95%、至少98%或至少99%的氨基酸序列同一性的氨基酸序列,并且包含对氨基酸R472、H477、R1044和H1049的取代。在一些情况下,位置472和1044处的氨基酸是除了Arg以外的任何氨基酸;并且位置477和1049处的氨基酸是除了His以外的任何氨基酸。在一些情况下,取代是R472A、H477A、R1044A和H1049A。在一些情况下,变体C2c2多肽具有减少的或不可检测的ss RNA切割(例如,RNA指导的切割活性),并保留结合

C2c2指导RNA和ss RNA的能力。在一些情况下,变体C2c2多肽保留切割前体C2c2指导RNA的能力。在一些情况下,变体C2c2多肽具有减少的或不可检测的ss RNA切割(例如,RNA指导的切割活性),但保留了结合C2c2指导RNA和ssRNA的能力,并且保留了切割前体C2c2指导RNA的能力。

[0282] 前体C2c2指导RNA阵列

[0283] 如下文工作实施例中所证明的,C2c2蛋白可以将前体C2c2指导RNA切割成成熟的指导RNA,例如通过对前体的内切核糖核酸酶切割。同样如以下工作实施例中所证明的,C2c2蛋白可以将前体C2c2指导RNA阵列(其包括串联排列的多于一个C2c2指导RNA)切割成两个或更多个单独的C2c2指导RNA。因此,在一些情况下,前体C2c2指导RNA阵列包含两个或更多个(例如,3个或更多个、4个或更多个、5个或更多个、2个、3个、4个或5个)C2c2指导RNA(例如,作为前体分子串联排列)。在一些情况下,前体C2c2指导RNA阵列的每个指导RNA具有不同的指导序列。在一些情况下,前体C2c2指导RNA阵列的两个或更多个指导RNA具有相同的指导序列。

[0284] 在一些情况下,前体C2c2指导RNA阵列包含两个或更多个C2c2指导RNA,所述两个或更多个C2c2指导RNA靶向相同靶RNA分子内的不同靶位点。例如,在一些情况下,当所述C2c2指导RNA中的任一者与靶RNA分子杂交时,这种情形可以通过激活C2c2蛋白而增加检测灵敏度。

[0285] 在一些情况下,前体C2c2指导RNA阵列包含两个或更多个C2c2指导RNA,所述两个或更多个C2c2指导RNA靶向不同的靶RNA分子。例如,当存在潜在靶RNA家族中的任何一者时,这种情况可以导致阳性信号。这种阵列可用于例如基于诸如单核苷酸多态性(SNP)之类的变异而靶向转录物家族(例如,用于诊断目的)。这也可用于检测是否存在许多不同病毒株中的任何一种(例如,流感病毒变体、寨卡病毒变体、HIV变体等)。这也可用于检测是否存在许多不同物种、菌株、分离株或细菌变体中的任何一者(例如,分枝杆菌(*Mycobacterium*)的不同物种、菌株、分离株或变体,奈瑟氏菌(*Neisseria*)的不同物种、菌株、分离株、或变体,金黄色葡萄球菌(*Staphylococcus aureus*)的不同物种、菌株、分离株、或变体;大肠杆菌的不同物种、菌株、分离株、或变体;等等)。

[0286] 变体C2C2多肽

[0287] 本公开提供了变体C2c2多肽,以及包含编码所述变体C2c2多肽的核苷酸序列的核酸(例如,重组表达载体)。

[0288] 在一些情况下,变体C2c2多肽包含与SEQ ID NO:2所示氨基酸序列具有至少80%、至少85%、至少90%、至少95%、至少98%或至少99%的氨基酸序列同一性的氨基酸序列,并且包含对氨基酸R472和H477的取代。在一些情况下,位置472处的氨基酸是除了Arg以外的任何氨基酸;并且位置477处的氨基酸是除了His以外的任何氨基酸。在一些情况下,取代是R472A和H477A。在一些情况下,变体C2c2多肽包含与SEQ ID NO:2所示氨基酸序列具有至少80%、至少85%、至少90%、至少95%、至少98%或至少99%的氨基酸序列同一性的氨基酸序列,并且包含对氨基酸R1048和H1053的取代。在一些情况下,位置1048处的氨基酸是除了Arg以外的任何氨基酸;并且位置1053处的氨基酸是除了His以外的任何氨基酸。在一些情况下,取代是R1048A和H1053A。在一些情况下,变体C2c2多肽包含与SEQ ID NO:2所示氨基酸序列具有至少80%、至少85%、至少90%、至少95%、至少98%或至少99%的氨基酸序

列同一性的氨基酸序列,并且包含对氨基酸R472、H477、R1048和H1053的取代。在一些情况下,位置472和1048处的氨基酸是除了Arg以外的任何氨基酸;并且位置477和1053处的氨基酸是除了His以外的任何氨基酸。在一些情况下,取代是R472A、H477A、R1048A和H1053A。在一些情况下,变体C2c2多肽具有减少的或不可检测的ss RNA切割(例如,RNA指导的切割活性),并保留结合C2c2指导RNA和ss RNA的能力。在一些情况下,变体C2c2多肽保留切割前体C2c2指导RNA的能力。在一些情况下,变体C2c2多肽具有减少的或不可检测的ss RNA切割(例如,RNA指导的切割活性),但保留了结合C2c2指导RNA和ssRNA的能力,并且保留了切割前体C2c2指导RNA的能力。

[0289] 在一些情况下,变体C2c2多肽包含与SEQ ID NO:1所示氨基酸序列具有至少80%、至少85%、至少90%、至少95%、至少98%或至少99%的氨基酸序列同一性的氨基酸序列,并且包含对氨基酸R445和H450的取代。在一些情况下,位置445处的氨基酸是除了Arg以外的任何氨基酸;并且位置450处的氨基酸是除了His以外的任何氨基酸。在一些情况下,取代是R445A和H450A。在一些情况下,变体C2c2多肽包含与SEQ ID NO:1所示氨基酸序列具有至少80%、至少85%、至少90%、至少95%、至少98%或至少99%的氨基酸序列同一性的氨基酸序列,并且包含对氨基酸R1016和H1021的取代。在一些情况下,位置1016处的氨基酸是除了Arg以外的任何氨基酸;并且位置1021处的氨基酸是除了His以外的任何氨基酸。在一些情况下,取代是R1016A和H1021A。在一些情况下,变体C2c2多肽包含与SEQ ID NO:1所示氨基酸序列具有至少80%、至少85%、至少90%、至少95%、至少98%或至少99%的氨基酸序列同一性的氨基酸序列,并且包含对氨基酸R445、H450、R1016和H1021的取代。在一些情况下,位置445和1016处的氨基酸是除了Arg以外的任何氨基酸;并且位置450和1016处的氨基酸是除了His以外的任何氨基酸。在一些情况下,取代是R445A、H450A、R1016A和H1021A。在一些情况下,变体C2c2多肽具有减少的或不可检测的ss RNA切割(例如,RNA指导的切割活性),并保留结合C2c2指导RNA和ss RNA的能力。在一些情况下,变体C2c2多肽保留切割前体C2c2指导RNA的能力。在一些情况下,变体C2c2多肽具有减少的或不可检测的ss RNA切割(例如,RNA指导的切割活性),但保留了结合C2c2指导RNA和ssRNA的能力,并且保留了切割前体C2c2指导RNA的能力。

[0290] 在一些情况下,变体C2c2多肽包含与SEQ ID NO:4(荚膜红细菌C2c2)所示氨基酸序列具有至少80%、至少85%、至少90%、至少95%、至少98%或至少99%的氨基酸序列同一性的氨基酸序列,并且包含对氨基酸R464和H469的取代。在一些情况下,位置464处的氨基酸是除了Arg以外的任何氨基酸;并且位置469处的氨基酸是除了His以外的任何氨基酸。在一些情况下,取代是R464A和H469A。在一些情况下,变体C2c2多肽包含与SEQ ID NO:4所示氨基酸序列具有至少80%、至少85%、至少90%、至少95%、至少98%或至少99%的氨基酸序列同一性的氨基酸序列,并且包含对氨基酸R1052和H1057的取代。在一些情况下,位置1052处的氨基酸是除了Arg以外的任何氨基酸;并且位置1057处的氨基酸是除了His以外的任何氨基酸。在一些情况下,取代是R1052A和H1057A。在一些情况下,变体C2c2多肽包含与SEQ ID NO:4所示氨基酸序列具有至少80%、至少85%、至少90%、至少95%、至少98%或至少99%的氨基酸序列同一性的氨基酸序列,并且包含对氨基酸R464、H469、R1052和H1057的取代。在一些情况下,位置464和1052处的氨基酸是除了Arg以外的任何氨基酸;并且位置469和1057处的氨基酸是除了His以外的任何氨基酸。在一些情况下,取代是R464A、H469A、

R1052A和H1057A。在一些情况下,变体C2c2多肽具有减少的或不可检测的ss RNA切割(例如,RNA指导的切割活性),并保留结合C2c2指导RNA和ss RNA的能力。在一些情况下,变体C2c2多肽保留切割前体C2c2指导RNA的能力。在一些情况下,变体C2c2多肽具有减少的或不可检测的ss RNA切割(例如,RNA指导的切割活性),但保留了结合C2c2指导RNA和ssRNA的能力,并且保留了切割前体C2c2指导RNA的能力。

[0291] 在一些情况下,变体C2c2多肽包含与SEQ ID NO:5(鸡肉杆菌C2c2)所示氨基酸序列具有至少80%、至少85%、至少90%、至少95%、至少98%或至少99%的氨基酸序列同一性的氨基酸序列,并且包含对氨基酸R467和H472的取代。在一些情况下,位置467处的氨基酸是除了Arg以外的任何氨基酸;并且位置472处的氨基酸是除了His以外的任何氨基酸。在一些情况下,取代是R469A和H472A。在一些情况下,变体C2c2多肽包含与SEQ ID NO:5所示氨基酸序列具有至少80%、至少85%、至少90%、至少95%、至少98%或至少99%的氨基酸序列同一性的氨基酸序列,并且包含对氨基酸R1069和H1074的取代。在一些情况下,位置1069处的氨基酸是除了Arg以外的任何氨基酸;并且位置1074处的氨基酸是除了His以外的任何氨基酸。在一些情况下,取代是R1069A和H1074A。在一些情况下,变体C2c2多肽包含与SEQ ID NO:5所示氨基酸序列具有至少80%、至少85%、至少90%、至少95%、至少98%或至少99%的氨基酸序列同一性的氨基酸序列,并且包含对氨基酸R467、H472、R1069和H1074的取代。在一些情况下,位置467和1069处的氨基酸是除了Arg以外的任何氨基酸;并且位置472和1074处的氨基酸是除了His以外的任何氨基酸。在一些情况下,取代是R469A、H472A、R1069A和H1074A。在一些情况下,变体C2c2多肽具有减少的或不可检测的ss RNA切割(例如,RNA指导的切割活性),并保留结合C2c2指导RNA和ss RNA的能力。在一些情况下,变体C2c2多肽保留切割前体C2c2指导RNA的能力。在一些情况下,变体C2c2多肽具有减少的或不可检测的ss RNA切割(例如,RNA指导的切割活性),但保留了结合C2c2指导RNA和ssRNA的能力,并且保留了切割前体C2c2指导RNA的能力。

[0292] 在一些情况下,变体C2c2多肽包含与SEQ ID NO:6(*Herbinix hemicellulosilytica* C2c2)所示氨基酸序列具有至少80%、至少85%、至少90%、至少95%、至少98%或至少99%的氨基酸序列同一性的氨基酸序列,并且包含对氨基酸R472和H477的取代。在一些情况下,位置472处的氨基酸是除了Arg以外的任何氨基酸;并且位置477处的氨基酸是除了His以外的任何氨基酸。在一些情况下,取代是R472A和H477A。在一些情况下,变体C2c2多肽包含与SEQ ID NO:6所示氨基酸序列具有至少80%、至少85%、至少90%、至少95%、至少98%或至少99%的氨基酸序列同一性的氨基酸序列,并且包含对氨基酸R1044和H1049的取代。在一些情况下,位置1044处的氨基酸是除了Arg以外的任何氨基酸;并且位置1049处的氨基酸是除了His以外的任何氨基酸。在一些情况下,取代是R1044A和H1049A。在一些情况下,变体C2c2多肽包含与SEQ ID NO:6所示氨基酸序列具有至少80%、至少85%、至少90%、至少95%、至少98%或至少99%的氨基酸序列同一性的氨基酸序列,并且包含对氨基酸R472、H477、R1044和H1049的取代。在一些情况下,位置472和1044处的氨基酸是除了Arg以外的任何氨基酸;并且位置477和1049处的氨基酸是除了His以外的任何氨基酸。在一些情况下,取代是R472A、H477A、R1044A和H1049A。在一些情况下,变体C2c2多肽具有减少的或不可检测的ss RNA切割(例如,RNA指导的切割活性),并保留结合C2c2指导RNA和ss RNA的能力。在一些情况下,变体C2c2多肽保留切割前体C2c2指导RNA的

能力。在一些情况下,变体C2c2多肽具有减少的或不可检测的ss RNA切割(例如,RNA指导的切割活性),但保留了结合C2c2指导RNA和ssRNA的能力,并且保留了切割前体C2c2指导RNA的能力。

[0293] 在一些情况下,变体C2c2多肽包含与SEQ ID NO:2所示氨基酸序列具有至少80%、至少85%、至少90%、至少95%、至少98%或至少99%氨基酸序列同一性的氨基酸序列,并且包含对氨基酸R472、H477、R1048和H1053中的1个、2个、3个或4个氨基酸的取代,使得变体C2c2多肽具有减少的或不可检测的ss RNA切割(例如,RNA指导的切割活性),并且保留结合C2c2指导RNA和ss RNA的能力。在一些情况下,变体C2c2多肽保留切割前体C2c2指导RNA的能力。例如,在一些情况下,变体C2c2多肽表现出由具有SEQ ID NO:2所示氨基酸序列的C2c2多肽表现出的RNA指导的非靶RNA切割的小于50%、小于40%、小于30%、小于20%、小于10%、小于5%、小于1%或小于0.1%。在一些情况下,变体C2c2多肽具有减少的或不可检测的ss RNA切割(例如,RNA指导的切割活性),但保留了结合C2c2指导RNA和ssRNA的能力,并且保留了切割前体C2c2指导RNA的能力。

[0294] 本公开提供了核酸(例如,分离的核酸),所述核酸包含编码本公开的变体C2c2多肽的核苷酸序列。在一些情况下,核苷酸序列与转录控制元件(例如,启动子)可操作地连接。在一些情况下,启动子是组成型启动子。在一些情况下,启动子是可调控的启动子。在一些情况下,启动子是诱导型启动子。在一些情况下,启动子是在真核细胞中起作用的。在一些情况下,启动子是在原核细胞中起作用的。

[0295] 本公开提供了重组表达载体,所述重组表达载体包含本公开的核酸,例如包含编码本公开的变体C2c2多肽的核苷酸序列的核酸。

[0296] 本公开提供了用本公开的核酸(例如,包含编码本公开的变体C2c2多肽的核苷酸序列的核酸)遗传修饰的宿主细胞。本公开提供了用包含本公开的核酸(例如,包含编码本公开的变体C2c2多肽的核苷酸序列的核酸)的重组表达载体遗传修饰的宿主细胞。在一些情况下,宿主细胞是原核细胞。在一些情况下,宿主细胞是真核细胞。在一些情况下,宿主细胞是体外的。在一些情况下,宿主细胞是离体的。在一些情况下,宿主细胞是体内的。在一些情况下,宿主细胞是细菌细胞。在一些情况下,宿主细胞是酵母细胞。在一些情况下,宿主细胞是植物细胞。在一些情况下,宿主细胞是哺乳动物细胞。在一些情况下,宿主细胞是人细胞。在一些情况下,宿主细胞是非人哺乳动物细胞。在一些情况下,宿主细胞是昆虫细胞。在一些情况下,宿主细胞是节肢动物细胞。在一些情况下,宿主细胞是真菌细胞。在一些情况下,宿主细胞是藻类细胞。

[0297] 试剂盒

[0298] 本公开提供了用于检测包含多种RNA的样品中的靶RNA的试剂盒。在一些情况下,所述试剂盒包含:(a)前体C2c2指导RNA阵列,其包含两个或更多个C2c2指导RNA,每个所述C2c2指导RNA具有不同的指导序列;以及(b)C2c2蛋白和/或编码所述C2c2蛋白的核酸。在一些情况下,此类试剂盒还包括经标记的检测RNA(例如,包含荧光发射染料对(即,FRET对和/或猝灭剂/荧光剂对)的经标记的检测RNA)。在一些情况下,给定前体C2c2指导RNA阵列的两个或更多个C2c2指导RNA(例如,在一些情况下,每个C2c2指导RNA)包括相同的指导序列。

[0299] 在一些情况下,主题试剂盒包含:(a)经标记的检测RNA,其包含荧光发射染料对,即FRET对和/或猝灭剂/荧光剂对;以及(b)C2c2蛋白和/或编码所述C2c2蛋白的核酸。在一

些情况下,此类试剂盒还包括(c) C2c2指导RNA (和/或编码C2c2指导RNA的核酸),和/或(d) 前体C2c2指导RNA (和/或编码前体C2c2指导RNA的核酸),和/或(e) 前体C2c2指导RNA阵列 (和/或编码前体C2c2指导RNA阵列的核酸,例如编码包括供用户插入指导序列的插入位点的前体C2c2指导RNA阵列的核酸)。

[0300] 1) 包含前体C2c2指导RNA阵列和C2c2蛋白的试剂盒

[0301] 在一些情况下,所述试剂盒包含:(a) 前体C2c2指导RNA阵列,其包含两个或更多个C2c2指导RNA,每个所述C2c2指导RNA具有不同的指导序列;以及(b) C2c2蛋白和/或编码所述C2c2蛋白的核酸。如上所述,在一些情况下,此类试剂盒还包括经标记的检测RNA (例如,包含荧光发射染料对(即,FRET对和/或猝灭剂/荧光剂对)的经标记的检测RNA)。

[0302] C2c2蛋白

[0303] 适合于包括在本公开的试剂盒中的C2c2蛋白与C2c2指导RNA结合,由指导RNA (其与靶RNA杂交) 引导到单链靶RNA,从而被“激活”。如果C2c2蛋白的HEPN1结构域和HEPN2结构域是完整的,则C2c2蛋白一旦被激活就切割靶RNA,但是也切割非靶RNA。

[0304] 在一些情况下,适合于包括在本公开的试剂盒中的C2c2蛋白包括与SEQ ID NO:1至SEQ ID NO:6中任一项所示的氨基酸序列具有80%或更多(例如,85%或更多、90%或更多、95%或更多、98%或更多、99%或更多、99.5%或更多、或100%的氨基酸序列同一性的氨基酸序列。在一些情况下,适合于包括在本公开的试剂盒中的C2c2蛋白包含与SEQ ID NO:1所示的斯氏李斯特菌C2c2氨基酸序列具有至少80%、至少85%、至少90%、至少95%、至少98%、至少99%或100%的氨基酸序列同一性的氨基酸序列。在一些情况下,适合于包括在本公开的试剂盒中的C2c2蛋白包含与SEQ ID NO:2所示的口腔纤毛菌C2c2氨基酸序列具有至少80%、至少85%、至少90%、至少95%、至少98%、至少99%或100%的氨基酸序列同一性的氨基酸序列。在一些情况下,适合于包括在本公开的试剂盒中的C2c2蛋白包含与SEQ ID NO:4所示的荚膜红细菌C2c2氨基酸序列具有至少80%、至少85%、至少90%、至少95%、至少98%、至少99%或100%的氨基酸序列同一性的氨基酸序列。在一些情况下,适合于包括在本公开的试剂盒中的C2c2蛋白包含与SEQ ID NO:5所示的鸡肉杆菌C2c2氨基酸序列具有至少80%、至少85%、至少90%、至少95%、至少98%、至少99%或100%的氨基酸序列同一性的氨基酸序列。在一些情况下,适合于包括在本公开的试剂盒中的C2c2蛋白包含与SEQ ID NO:6所示的Herbinix hemicellulosilytica C2c2氨基酸序列具有至少80%、至少85%、至少90%、至少95%、至少98%、至少99%或100%的氨基酸序列同一性的氨基酸序列。在一些情况下,适合于包括在本公开的试剂盒中的C2c2蛋白包括与SEQ ID NO:2所示的口腔纤毛菌(Lbu) C2c2氨基酸序列具有80%或更多的氨基酸序列同一性的氨基酸序列。在一些情况下,适合于包括在本公开的试剂盒中的C2c2蛋白是口腔纤毛菌(Lbu) C2c2蛋白(例如,参见SEQ ID NO:2)。在一些情况下,适合于包括在本公开的试剂盒中的C2c2蛋白包括SEQ ID NO:1至SEQ ID NO:2和SEQ ID NO:4至SEQ ID NO:6中任一项所示的氨基酸序列。

[0305] 在一些情况下,包括在本公开的试剂盒中的C2c2蛋白不是沙氏纤毛菌(Lsh) C2c2蛋白。在一些情况下,包括在本公开的试剂盒中的C2c2蛋白不是与SEQ ID NO:3所示的Lsh C2c2多肽具有至少80%、至少85%、至少90%、至少95%、至少98%、至少99%或100%的氨基酸序列同一性的C2c2多肽。

[0306] 在一些情况下,包括在本公开的试剂盒中的C2c2多肽是变体C2c2多肽。适合于包

括在本公开的试剂盒中的变体C2c2多肽包括SEQ ID NO:1、SEQ ID NO:2和SEQ ID NO:4至SEQ ID NO:6中任一者的变体,其中变体C2c2多肽表现出降低的(或不可检测的)核酸酶活性。例如,在一些情况下,变体C2c2蛋白缺乏催化活性的HEPN1结构域。作为另一个实例,变体C2c2蛋白缺乏催化活性的HEPN2结构域。在一些情况下,变体C2c2蛋白缺乏催化活性的HEPN1结构域并且缺乏催化活性的HEPN2结构域。

[0307] 在一些情况下,变体C2c2多肽包含对SEQ ID NO:2(口腔纤毛菌C2c2)所示的氨基酸序列的氨基酸R472、H477、R1048和H1053中的1个、2个、3个或4个氨基酸的氨基酸取代,或SEQ ID NO:1、SEQ ID NO:4、SEQ ID NO:5或SEQ ID NO:6中所示的C2c2氨基酸序列的相应氨基酸的氨基酸取代。SEQ ID NO:1、SEQ ID NO:4、SEQ ID NO:5和SEQ ID NO:6中的相应氨基酸容易被鉴别;参见例如图22B。例如,SEQ ID NO:1(斯氏李斯特菌C2c2)中对应于SEQ ID NO:2的R472、H477、R1048和H1053的氨基位置分别是R445、H450、R1016和H1021。

[0308] 在一些情况下,变体C2c2多肽包含对SEQ ID NO:2所示的氨基酸序列的氨基酸R472和H477的氨基酸取代,或对SEQ ID NO:1、SEQ ID NO:4、SEQ ID NO:5或SEQ ID NO:6所示的C2c2氨基酸序列的相应氨基酸的氨基酸取代。在一些情况下,变体C2c2多肽包含对SEQ ID NO:2所示的氨基酸序列的氨基酸R1048和H1053的氨基酸取代,或对SEQ ID NO:1、SEQ ID NO:4、SEQ ID NO:5或SEQ ID NO:6所示的C2c2氨基酸序列的相应氨基酸的氨基酸取代。在一些情况下,变体C2c2多肽包含对SEQ ID NO:2所示的氨基酸序列的氨基酸R472、H477、R1048和H1053的氨基酸取代,或对SEQ ID NO:1、SEQ ID NO:4、SEQ ID NO:5或SEQ ID NO:6所示的C2c2氨基酸序列的相应氨基酸的氨基酸取代。

[0309] 在一些情况下,变体C2c2多肽包含与SEQ ID NO:2所示氨基酸序列具有至少80%、至少85%、至少90%、至少95%、至少98%或至少99%的氨基酸序列同一性的氨基酸序列,并且包含对氨基酸R472和H477的取代。在一些情况下,位置472处的氨基酸是除了Arg以外的任何氨基酸;并且位置477处的氨基酸是除了His以外的任何氨基酸。在一些情况下,取代是R472A和H477A。在一些情况下,变体C2c2多肽包含与SEQ ID NO:2所示氨基酸序列具有至少80%、至少85%、至少90%、至少95%、至少98%或至少99%的氨基酸序列同一性的氨基酸序列,并且包含对氨基酸R1048和H1053的取代。在一些情况下,位置1048处的氨基酸是除了Arg以外的任何氨基酸;并且位置1053处的氨基酸是除了His以外的任何氨基酸。在一些情况下,取代是R1048A和H1053A。在一些情况下,变体C2c2多肽包含与SEQ ID NO:2所示氨基酸序列具有至少80%、至少85%、至少90%、至少95%、至少98%或至少99%的氨基酸序列同一性的氨基酸序列,并且包含对氨基酸R472、H477、R1048和H1053的取代。在一些情况下,位置472和1048处的氨基酸是除了Arg以外的任何氨基酸;并且位置477和1053处的氨基酸是除了His以外的任何氨基酸。在一些情况下,取代是R472A、H477A、R1048A和H1053A。在一些情况下,变体C2c2多肽具有减少的或不可检测的ss RNA切割(例如,RNA指导的切割活性),并保留结合C2c2指导RNA和ss RNA的能力。在一些情况下,变体C2c2多肽保留切割前体C2c2指导RNA的能力。在一些情况下,变体C2c2多肽具有减少的或不可检测的ss RNA切割(例如,RNA指导的切割活性),但保留了结合C2c2指导RNA和ssRNA的能力,并且保留了切割前体C2c2指导RNA的能力。

[0310] 在一些情况下,变体C2c2多肽包含与SEQ ID NO:1所示氨基酸序列具有至少80%、至少85%、至少90%、至少95%、至少98%或至少99%的氨基酸序列同一性的氨基酸序列,

并且包含对氨基酸R445和H450的取代。在一些情况下,位置445处的氨基酸是除了Arg以外的任何氨基酸;并且位置450处的氨基酸是除了His以外的任何氨基酸。在一些情况下,取代是R445A和H450A。在一些情况下,变体C2c2多肽包含与SEQ ID NO:1所示氨基酸序列具有至少80%、至少85%、至少90%、至少95%、至少98%或至少99%的氨基酸序列同一性的氨基酸序列,并且包含对氨基酸R1016和H1021的取代。在一些情况下,位置1016处的氨基酸是除了Arg以外的任何氨基酸;并且位置1021处的氨基酸是除了His以外的任何氨基酸。在一些情况下,取代是R1016A和H1021A。在一些情况下,变体C2c2多肽包含与SEQ ID NO:1所示氨基酸序列具有至少80%、至少85%、至少90%、至少95%、至少98%或至少99%的氨基酸序列同一性的氨基酸序列,并且包含对氨基酸R445、H450、R1016和H1021的取代。在一些情况下,位置445和1016处的氨基酸是除了Arg以外的任何氨基酸;并且位置450和1016处的氨基酸是除了His以外的任何氨基酸。在一些情况下,取代是R445A、H450A、R1016A和H1021A。在一些情况下,变体C2c2多肽具有减少的或不可检测的ss RNA切割(例如,RNA指导的切割活性),并保留结合C2c2指导RNA和ss RNA的能力。在一些情况下,变体C2c2多肽保留切割前体C2c2指导RNA的能力。在一些情况下,变体C2c2多肽具有减少的或不可检测的ss RNA切割(例如,RNA指导的切割活性),但保留了结合C2c2指导RNA和ssRNA的能力,并且保留了切割前体C2c2指导RNA的能力。

[0311] 在一些情况下,变体C2c2多肽包含与SEQ ID NO:4(荚膜红细菌C2c2)所示氨基酸序列具有至少80%、至少85%、至少90%、至少95%、至少98%或至少99%的氨基酸序列同一性的氨基酸序列,并且包含对氨基酸R464和H469的取代。在一些情况下,位置464处的氨基酸是除了Arg以外的任何氨基酸;并且位置469处的氨基酸是除了His以外的任何氨基酸。在一些情况下,取代是R464A和H469A。在一些情况下,变体C2c2多肽包含与SEQ ID NO:4所示氨基酸序列具有至少80%、至少85%、至少90%、至少95%、至少98%或至少99%的氨基酸序列同一性的氨基酸序列,并且包含对氨基酸R1052和H1057的取代。在一些情况下,位置1052处的氨基酸是除了Arg以外的任何氨基酸;并且位置1057处的氨基酸是除了His以外的任何氨基酸。在一些情况下,取代是R1052A和H1057A。在一些情况下,变体C2c2多肽包含与SEQ ID NO:4所示氨基酸序列具有至少80%、至少85%、至少90%、至少95%、至少98%或至少99%的氨基酸序列同一性的氨基酸序列,并且包含对氨基酸R464、H469、R1052和H1057的取代。在一些情况下,位置464和1052处的氨基酸是除了Arg以外的任何氨基酸;并且位置469和1057处的氨基酸是除了His以外的任何氨基酸。在一些情况下,取代是R464A、H469A、R1052A和H1057A。在一些情况下,变体C2c2多肽具有减少的或不可检测的ss RNA切割(例如,RNA指导的切割活性),并保留结合C2c2指导RNA和ss RNA的能力。在一些情况下,变体C2c2多肽保留切割前体C2c2指导RNA的能力。在一些情况下,变体C2c2多肽具有减少的或不可检测的ss RNA切割(例如,RNA指导的切割活性),但保留了结合C2c2指导RNA和ssRNA的能力,并且保留了切割前体C2c2指导RNA的能力。

[0312] 在一些情况下,变体C2c2多肽包含与SEQ ID NO:5(鸡肉杆菌C2c2)所示氨基酸序列具有至少80%、至少85%、至少90%、至少95%、至少98%或至少99%的氨基酸序列同一性的氨基酸序列,并且包含对氨基酸R467和H472的取代。在一些情况下,位置467处的氨基酸是除了Arg以外的任何氨基酸;并且位置472处的氨基酸是除了His以外的任何氨基酸。在一些情况下,取代是R469A和H472A。在一些情况下,变体C2c2多肽包含与SEQ ID NO:5所示

氨基酸序列具有至少80%、至少85%、至少90%、至少95%、至少98%或至少99%的氨基酸序列同一性的氨基酸序列,并且包含对氨基酸R1069和H1074的取代。在一些情况下,位置1069处的氨基酸是除了Arg以外的任何氨基酸;并且位置1074处的氨基酸是除了His以外的任何氨基酸。在一些情况下,取代是R1069A和H1074A。在一些情况下,变体C2c2多肽包含与SEQ ID NO:5所示氨基酸序列具有至少80%、至少85%、至少90%、至少95%、至少98%或至少99%的氨基酸序列同一性的氨基酸序列,并且包含对氨基酸R467、H472、R1069和H1074的取代。在一些情况下,位置467和1069处的氨基酸是除了Arg以外的任何氨基酸;并且位置472和1074处的氨基酸是除了His以外的任何氨基酸。在一些情况下,取代是R469A、H472A、R1069A和H1074A。在一些情况下,变体C2c2多肽具有减少的或不可检测的ss RNA切割(例如,RNA指导的切割活性),并保留结合C2c2指导RNA和ss RNA的能力。在一些情况下,变体C2c2多肽保留切割前体C2c2指导RNA的能力。在一些情况下,变体C2c2多肽具有减少的或不可检测的ss RNA切割(例如,RNA指导的切割活性),但保留了结合C2c2指导RNA和ssRNA的能力,并且保留了切割前体C2c2指导RNA的能力。

[0313] 在一些情况下,变体C2c2多肽包含与SEQ ID NO:6 (*Herbinix hemicellulosilytica* C2c2)所示氨基酸序列具有至少80%、至少85%、至少90%、至少95%、至少98%或至少99%的氨基酸序列同一性的氨基酸序列,并且包含对氨基酸R472和H477的取代。在一些情况下,位置472处的氨基酸是除了Arg以外的任何氨基酸;并且位置477处的氨基酸是除了His以外的任何氨基酸。在一些情况下,取代是R472A和H477A。在一些情况下,变体C2c2多肽包含与SEQ ID NO:6所示氨基酸序列具有至少80%、至少85%、至少90%、至少95%、至少98%或至少99%的氨基酸序列同一性的氨基酸序列,并且包含对氨基酸R1044和H1049的取代。在一些情况下,位置1044处的氨基酸是除了Arg以外的任何氨基酸;并且位置1049处的氨基酸是除了His以外的任何氨基酸。在一些情况下,取代是R1044A和H1049A。在一些情况下,变体C2c2多肽包含与SEQ ID NO:6所示氨基酸序列具有至少80%、至少85%、至少90%、至少95%、至少98%或至少99%的氨基酸序列同一性的氨基酸序列,并且包含对氨基酸R472、H477、R1044和H1049的取代。在一些情况下,位置472和1044处的氨基酸是除了Arg以外的任何氨基酸;并且位置477和1049处的氨基酸是除了His以外的任何氨基酸。在一些情况下,取代是R472A、H477A、R1044A和H1049A。在一些情况下,变体C2c2多肽具有减少的或不可检测的ss RNA切割(例如,RNA指导的切割活性),并保留结合C2c2指导RNA和ss RNA的能力。在一些情况下,变体C2c2多肽保留切割前体C2c2指导RNA的能力。在一些情况下,变体C2c2多肽具有减少的或不可检测的ss RNA切割(例如,RNA指导的切割活性),但保留了结合C2c2指导RNA和ssRNA的能力,并且保留了切割前体C2c2指导RNA的能力。

[0314] 2) 包含经标记的检测RNA和C2c2蛋白的试剂盒

[0315] 在一些情况下,所述试剂盒包含:(a) 经标记的检测RNA,其包含荧光发射染料对,即FRET对和/或猝灭剂/荧光剂对;以及(b) C2c2蛋白和/或编码所述C2c2蛋白的核酸。在一些情况下,试剂盒还包括C2c2指导RNA、前体C2c2指导RNA阵列和/或编码C2c2指导RNA的恒定区的核酸。如上所述,在一些情况下,此类试剂盒还包括(c) C2c2指导RNA(和/或编码C2c2指导RNA的核酸),和/或(d) 前体C2c2指导RNA(和/或编码前体C2c2指导RNA的核酸),和/或(e) 前体C2c2指导RNA阵列(和/或编码前体C2c2指导RNA阵列的核酸,例如编码包括供用户

插入指导序列的插入位点的前体C2c2指导RNA阵列的核酸)。

[0316] 经标记的检测RNA

[0317] 在一些情况下,本公开的试剂盒包含经标记的检测RNA,所述经标记的检测RNA包含荧光发射染料对,即FRET对和/或猝灭剂/荧光剂对。经标记的检测RNA在被切割之前产生一定量的可检测信号,并且当所述经标记的检测RNA被切割时所测量到的可检测信号的量减少。在一些情况下,经标记的检测RNA在被切割之前产生第一可检测信号(例如,来自FRET对),并且当所述经标记的检测RNA被切割时产生第二可检测信号(例如,来自猝灭剂/荧光剂对)。因此,在一些情况下,经标记的检测RNA包含FRET对和猝灭剂/荧光剂对。

[0318] 在一些情况下,经标记的检测RNA包含FRET对。FRET是这样的过程,通过所述过程,能量的无辐射转移发生为从激发态荧光团到紧邻的第二发色团。能量转移可以发生的范围限制在约10纳米(100埃),并且转移效率对荧光团之间的分开距离非常敏感。供体-受体对(FRET供体部分和FRET受体部分)在本文中被称为“FRET对”或“信号FRET对”。因此,在一些情况下,主题经标记的检测RNA包括两个信号配偶体(信号对),当一个信号配偶体是FRET供体部分时,另一个信号配偶体是FRET受体部分。因此,当信号配偶体非常接近时(例如,在相同的RNA分子上时),包括这种FRET对(FRET供体部分和FRET受体部分)的主题经标记的检测RNA将显示可检测信号(FRET信号),但是当所述配偶体分开时(例如,在用C2c2蛋白切割RNA分子后),信号将减少(或不存在)。

[0319] FRET供体和受体部分(FRET对)将是本领域的普通技术人员已知的,并且可以使用任何合宜的FRET对(例如,任何合宜的供体和受体部分对)。合适的FRET对的实例包括但不限于上表1中所示的那些。

[0320] 在一些情况下,可检测信号在经标记的检测RNA被切割时产生(例如,在一些情况下,经标记的检测RNA包含猝灭剂/荧光剂对。信号猝灭对的一个信号配偶体产生可检测信号,而另一个信号配偶体是猝灭剂部分,所述猝灭剂部分猝灭第一信号配偶体的可检测信号(即,猝灭剂部分猝灭信号部分的信号,使得当信号配偶体彼此接近时(例如,当信号对的信号配偶体非常接近时),来自信号部分的信号减少(被猝灭)。

[0321] 例如,在一些情况下,当经标记的检测RNA被切割时,可检测信号的量增加。例如,在一些情况下,由一个信号配偶体(信号部分)表现的信号被另一个信号配偶体(猝灭剂信号部分)猝灭,例如当在由C2c2蛋白进行切割之前两者存在于相同的RNA分子上时。这种信号对在本文中称为“猝灭剂/荧光剂对”、“猝灭对”或“信号猝灭对”。例如,在一些情况下,一个信号配偶体(例如,第一信号配偶体)是产生可检测信号的信号部分,所述可检测信号由第二信号配偶体(例如,猝灭剂部分)猝灭。因此,当配偶体被分开时(例如,在由C2c2蛋白切割检测RNA之后),这种猝灭剂/荧光剂对的信号配偶体将产生可检测信号,但是当配偶体非常接近(例如,在由C2c2蛋白切割检测RNA之前)时,所述信号将被猝灭。

[0322] 猝灭剂部分可以将来自信号部分的信号(例如,在由C2c2蛋白切割检测RNA之前)猝灭至不同程度。在一些情况下,猝灭剂部分猝灭来自信号部分的信号,其中在存在猝灭剂部分的情况下(当信号配偶体彼此接近时)检测到的信号是在不存在猝灭剂部分(当信号配偶体分开时)时检测到的信号的95%或更少。例如,在一些情况下,在存在猝灭剂部分的情况下检测到的信号可以是在不存在猝灭剂部分的情况下检测到的信号的90%或更少、80%或更少、70%或更少、60%或更少、50%或更少、40%或更少、30%或更少、20%或更少、15%

或更少、10%或更少、或5%或更少。在一些情况下,在存在猝灭剂部分的情况下未检测到信号(例如,高于背景)。

[0323] 在一些情况下,在不存在猝灭剂部分的情况下(当信号配偶体分开时)检测到的信号是在存在猝灭剂部分的情况下(当信号配偶体彼此接近时)检测到的信号的至少1.2倍(例如,至少1.3倍、至少1.5倍、至少1.7倍、至少2倍、至少2.5倍、至少3倍、至少3.5倍、至少4倍、至少5倍、至少7倍、至少10倍、至少20倍或至少50倍)。

[0324] 在一些情况下,信号部分是荧光标记。在一些此类情况下,猝灭剂部分猝灭来自荧光标记的信号(光信号)(例如,通过吸收标记的发射光谱中的能量)。因此,当猝灭剂部分不与信号部分接近时,来自荧光标记的发射(信号)是可检测的,因为信号不被猝灭剂部分吸收。可以使用任何合宜的供体受体对(信号部分/猝灭剂部分对),并且许多合适的对是本领域中已知的。

[0325] 在一些情况下,猝灭剂部分从信号部分(在本文中也称为“可检测标记”)吸收能量,然后发射信号(例如,不同波长的光)。因此,在一些情况下,猝灭剂部分本身是信号部分(例如,信号部分可以是6-羧基荧光素,而猝灭剂部分可以是6-羧基-四甲基罗丹明),并且在一些此类情况下,所述对也可以是FRET对。在一些情况下,猝灭剂部分是暗猝灭剂。暗猝灭剂可以吸收激发能量并以不同方式耗散能量(例如,作为热量)。因此,暗猝灭剂本身具有极小荧光至没有荧光(不发射荧光)。暗猝灭剂的实例进一步描述于美国专利号8,822,673和8,586,718;美国专利公开20140378330、20140349295和20140194611;以及国际专利申请:W0200142505和W0200186001中,所有这些专利都据此以引用方式整体并入本文。

[0326] 荧光标记的实例包括但不限于:Alexa **Fluor**[®]染料、ATTO染料(例如,ATTO 390、ATTO 425、ATTO 465、ATTO 488、ATTO495、ATTO 514、ATTO 520、ATTO 532、ATTO Rho6G、ATTO542、ATTO 550、ATTO 565、ATTO Rho3B、ATTO Rho11、ATTO Rho12、ATTO Thio12、ATTO Rho101、ATTO 590、ATTO 594、ATTO Rho13、ATTO 610、ATTO 620、ATTO Rho14、ATTO 633、ATTO 647、ATTO 647N、ATTO 655、ATTO Oxa12、ATTO 665、ATTO 680、ATTO 700、ATTO 725、ATTO 740)、DyLight染料、花青染料(例如,Cy2、Cy3、Cy3.5、Cy3b、Cy5、Cy5.5、Cy7、Cy7.5)、FluoProbes染料、Sulfo Cy染料、Seta染料、IRIS染料、SeTau染料、SRfluor染料、方形染料(Square dye)、异硫氰酸荧光素(FITC)、四甲基罗丹明(TRITC)、德克萨斯红、俄勒冈绿、太平洋蓝、太平洋绿、太平洋橙、量子点和束缚荧光蛋白(tethered fluorescent protein)。

[0327] 在一些情况下,可检测标记是选自以下的荧光标记:Alexa **Fluor**[®]染料、ATTO染料(例如,ATTO 390、ATTO 425、ATTO465、ATTO 488、ATTO 495、ATTO 514、ATTO 520、ATTO 532、ATTO Rho6G、ATTO 542、ATTO 550、ATTO 565、ATTO Rho3B、ATTO Rho11、ATTO Rho12、ATTO Thio12、ATTO Rho101、ATTO 590、ATTO 594、ATTO Rho13、ATTO 610、ATTO620、ATTO Rho14、ATTO 633、ATTO 647、ATTO 647N、ATTO655、ATTO Oxa12、ATTO 665、ATTO 680、ATTO 700、ATTO725、ATTO 740)、DyLight染料、花青染料(例如,Cy2、Cy3、Cy3.5、Cy3b、Cy5、Cy5.5、Cy7、Cy7.5)、FluoProbes染料、Sulfo Cy染料、Seta染料、IRIS染料、SeTau染料、SRfluor染料、方形染料(Square dye)、异硫氰酸荧光素(FITC)、四甲基罗丹明(TRITC)、德克萨斯红、俄勒冈绿、太平洋蓝、太平洋绿和太平洋橙。

[0328] 在一些情况下,可检测标记是选自以下的荧光标记:Alexa **Fluor**[®]染料、ATTO染料(例如,ATTO 390、ATTO 425、ATTO465、ATTO 488、ATTO 495、ATTO 514、ATTO 520、ATTO 532、

ATTO Rho6G、ATTO 542、ATTO 550、ATTO 565、ATTO Rho3B、ATTO Rho11、ATTO Rho12、ATTO Thio12、ATTO Rho101、ATTO 590、ATTO 594、ATTO Rho13、ATTO 610、ATTO620、ATTO Rho14、ATTO 633、ATTO 647、ATTO 647N、ATTO655、ATTO Oxa12、ATTO 665、ATTO 680、ATTO 700、ATTO725、ATTO 740)、DyLight染料、花青染料(例如,Cy2、Cy3、Cy3.5、Cy3b、Cy5、Cy5.5、Cy7、Cy7.5)、FluoProbes染料、Sulfo Cy染料、Seta染料、IRIS染料、SeTau染料、SRfluor染料、方形染料(Square dye)、异硫氰酸荧光素(FITC)、四甲基罗丹明(TRITC)、德克萨斯红、俄勒冈绿、太平洋蓝、太平洋绿、太平洋橙、量子点和束缚荧光蛋白。

[0329] ATTO染料的实例包括但不限于:ATTO 390、ATTO 425、ATTO 465、ATTO 488、ATTO 495、ATTO 514、ATTO 520、ATTO 532、ATTO Rho6G、ATTO 542、ATTO 550、ATTO 565、ATTO Rho3B、ATTO Rho11、ATTO Rho12、ATTO Thio12、ATTO Rho101、ATTO 590、ATTO 594、ATTO Rho13、ATTO 610、ATTO620、ATTO Rho14、ATTO 633、ATTO 647、ATTO 647N、ATTO655、ATTO Oxa12、ATTO 665、ATTO 680、ATTO 700、ATTO725以及ATTO 740。

[0330] AlexaFluor染料的实例包括但不限于:Alexa**Fluor**[®] 350、Alexa**Fluor**[®] 405、Alexa**Fluor**[®] 430、Alexa**Fluor**[®] 488、Alexa**Fluor**[®] 500、Alexa**Fluor**[®] 514、Alexa**Fluor**[®] 532、Alexa**Fluor**[®] 546、Alexa**Fluor**[®] 555、Alexa**Fluor**[®] 568、Alexa**Fluor**[®] 594、Alexa**Fluor**[®] 610、Alexa**Fluor**[®] 633、Alexa**Fluor**[®] 635、Alexa**Fluor**[®] 647、Alexa**Fluor**[®] 660、Alexa**Fluor**[®] 680、Alexa**Fluor**[®] 700、Alexa**Fluor**[®] 750、Alexa**Fluor**[®] 790等等。

[0331] 猝灭剂部分的实例包括但不限于:暗猝灭剂、Black Hole**Quencher**[®] (BHQ[®]) (例如,BHQ-0、BHQ-1、BHQ-2、BHQ-3)、Qx1猝灭剂、ATTO猝灭剂(例如,ATTO 540Q、ATTO 580Q和ATTO612Q)、二甲氨基偶氮苯磺酸(Dabsyl)、Iowa Black RQ、Iowa Black FQ、IRDye QC-1、QSY染料(例如,QSY 7、QSY 9、QSY 21)、AbsoluteQuencher、Eclipse和金属簇(诸如金纳米粒子)等。

[0332] 在一些情况下,猝灭剂部分选自:暗猝灭剂、Black Hole**Quencher**[®] (BHQ[®]) (例如,BHQ-0、BHQ-1、BHQ-2、BHQ-3)、Qx1猝灭剂、ATTO猝灭剂(例如,ATTO 540Q、ATTO 580Q和ATTO612Q)、二甲氨基偶氮苯磺酸(Dabsyl)、Iowa Black RQ、Iowa Black FQ、IRDye QC-1、QSY染料(例如,QSY 7、QSY 9、QSY 21)、AbsoluteQuencher、Eclipse和金属簇。

[0333] ATTO猝灭剂的实例包括但不限于:ATTO 540Q、ATTO 580Q和ATTO 612Q。Black Hole**Quencher**[®] (BHQ[®])的实例包括但不限于:BHQ-0 (493nm)、BHQ-1 (534nm)、BHQ-2 (579nm) 和BHQ-3 (672nm)。

[0334] 对于一些可检测标记(例如,荧光染料)和/或猝灭剂部分的实例,参见例如Bao等人,Annu Rev Biomed Eng.2009;11:25-47;以及美国专利号8,822,673和8,586,718;美国专利公开20140378330、20140349295、20140194611、20130323851、20130224871、20110223677、20110190486、20110172420、20060179585和20030003486;以及国际专利申请:W0200142505和W0200186001,所有这些参考文献据此以引用方式整体并入。

[0335] 核酸修饰

[0336] 在一些情况下,经标记的检测RNA包含一种或多种修饰,例如碱基修饰、骨架修饰、

糖修饰等,以为核酸提供新的或增强的特征(例如,改善的稳定性)。如本领域中已知的,核苷是碱基-核糖组合。核苷的碱基部分通常是杂环碱基。此类杂环碱基的两种最常见的类别是嘌呤和嘧啶。核苷酸是这样的核苷,所述核苷还包括与核苷的糖部分共价连接的磷酸基团。对于那些包括戊呋喃糖基糖的核苷,磷酸基团可以与所述糖的2'、3'或5'羟基部分连接。在形成寡核苷酸时,磷酸基团将相邻的核苷彼此共价连接以形成直链聚合化合物。继而,该直链聚合化合物的各个末端可以进一步连接以形成环状化合物;然而,直链化合物通常是合适的。此外,线性化合物可以具有内部核苷酸碱基互补性,并且因此可以以产生完全或部分双链化合物的方式折叠。在寡核苷酸内,磷酸基团通常被称为形成寡核苷酸的核苷间骨架。RNA和DNA的正常键或骨架是3'至5'磷酸二酯键。

[0337] 经修饰的骨架和经修饰的核苷间键

[0338] 合适的修饰的实例包括经修饰的核酸骨架和非天然的核苷间键。(具有经修饰的骨架的核酸包括在骨架中保留磷原子的核酸和在骨架中不具有磷原子的核酸。

[0339] 其中含有磷原子的合适的经修饰寡核苷酸骨架包括例如硫代磷酸酯、手性硫代磷酸酯、二硫代磷酸酯、磷酸三酯、氨基烷基磷酸三酯、甲基和其它烷基磷酸酯(包括3'-亚烷基磷酸酯、5'-亚烷基磷酸酯和手性磷酸酯)、次磷酸酯、氨基磷酸酯(包括3'-氨基氨基磷酸酯和氨基烷基氨基磷酸酯)、二氨基磷酸酯、硫代氨基磷酸酯、硫代烷基磷酸酯、硫代烷基磷酸三酯、具有正常3'-5'键的硒代磷酸酯和硼代磷酸酯、硒代磷酸酯和硼代磷酸酯的2'-5'连接的类似物,以及具有反向极性的硒代磷酸酯和硼代磷酸酯,其中一个或多个核苷酸间键是3'至3'键、5'至5'键或2'至2'键。合适的具有反向极性的寡核苷酸包含在最3'核苷酸间键处的单个3'至3'键,即单个反向核苷残基,所述单个反向核苷残基可以是碱性的(核碱基丢失或具有羟基代替其)。还包括各种盐(诸如,钾盐或钠盐)、混合盐和游离酸形式。

[0340] 在一些情况下,经标记的检测RNA包含一个或多个硫代磷酸酯和/或杂原子核苷间键,特别是 $-\text{CH}_2-\text{NH}-\text{O}-\text{CH}_2-$ 、 $-\text{CH}_2-\text{N}(\text{CH}_3)-\text{O}-\text{CH}_2-$ (称为亚甲基(甲基亚氨基)或MMI骨架)、 $-\text{CH}_2-\text{O}-\text{N}(\text{CH}_3)-\text{CH}_2-$ 、 $-\text{CH}_2-\text{N}(\text{CH}_3)-\text{N}(\text{CH}_3)-\text{CH}_2-$ 和 $-\text{O}-\text{N}(\text{CH}_3)-\text{CH}_2-\text{CH}_2-$ (其中天然磷酸二酯核苷酸间键表示为 $-\text{O}-\text{P}(=\text{O})(\text{OH})-\text{O}-\text{CH}_2-$)。MMI型核苷间键在上面引用的美国专利号5,489,677中公开。合适的酰胺核苷间键在美国专利号5,602,240中公开。

[0341] 同样合适的是如在例如美国专利号5,034,506中所述的具有吗啉代骨架结构的核酸。例如,在一些情况下,经标记的检测RNA包含6元吗啉代环代替核糖环。在一些情况下,二氨基磷酸酯或其它非磷酸二酯核苷间键取代磷酸二酯键。

[0342] 其中不含磷原子的合适的经修饰多核苷酸骨架具有由短链烷基或环烷基核苷间键、混合杂原子和烷基或环烷基核苷间键、或一个或多个短链杂原子或杂环核苷间键形成的骨架。这些骨架包括具有吗啉代键(部分地由核苷的糖部分形成)的那些骨架;硅氧烷骨架;硫化物、亚砷和砷骨架;甲酰基和硫代甲酰基骨架;亚甲基甲酰基和硫代甲酰基骨架;核糖乙酰骨架;含有烯基的骨架;氨基磺酸酯骨架;亚甲基亚氨基和亚甲基胍基骨架;磺酸酯和磺酰胺骨架;酰胺骨架;以及其它具有混合的N、O、S和 CH_2 组成部分的骨架。

[0343] 模拟物

[0344] 经标记的检测RNA可以是核酸模拟物。应用于多核苷酸的术语“模拟物”旨在包括这样的多核苷酸,在所述多核苷酸中仅呋喃糖环或呋喃糖环和核苷酸间键两者被非呋喃糖基团替代,仅替代呋喃糖环也在本领域中被称作糖替代物。维持杂环碱基部分或经修饰的

杂环碱基部分以与合适的靶核酸杂交。一种此类核酸,已显示具有优异杂交特性的多核苷酸模拟物,被称为肽核酸(PNA)。在PNA中,多核苷酸的糖骨架被含酰胺的骨架,特别是氨基乙基甘氨酸骨架取代。核苷酸被保留并直接或间接地结合骨架的酰胺部分的氮杂氮原子。

[0345] 一种据报道具有优异杂交特性的多核苷酸模拟物是肽核酸(PNA)。PNA化合物中的骨架是两个或更多个连接的氨基乙基甘氨酸单元,所述氨基乙基甘氨酸单元赋予PNA含酰胺的骨架。杂环碱基部分直接或间接地与骨架的酰胺部分的氮杂氮原子结合。描述PNA化合物制备的代表性美国专利包括但不限于:美国专利号5,539,082;5,714,331;和5,719,262。

[0346] 已经研究的另一类多核苷酸模拟物基于连接的吗啉代单元(吗啉代核酸),所述连接的吗啉代单元具有附接到吗啉代环的杂环碱基。已经报道了许多连接基团,所述连接基团连接吗啉代核酸中的吗啉代单体单元。一类连接基团已经被选择用于提供非离子寡聚化合物。非离子的基于吗啉代的寡聚化合物不太可能与细胞蛋白发生不希望的相互作用。基于吗啉代的多核苷酸是寡核苷酸的非离子模拟物,其不太可能与细胞蛋白形成不希望的相互作用(Dwayne A. Braasch和David R. Corey, Biochemistry, 2002, 41 (14), 4503-4510)。基于吗啉代的多核苷酸在美国专利号5,034,506中公开。已经制备了吗啉代类多核苷酸内的多种化合物,所述化合物具有各种连接单体亚基的不同连接基团。

[0347] 另一类多核苷酸模拟物被称为环己烯基核酸(CeNA)。通常存在于DNA/RNA分子中的呋喃糖环被环己烯基环替代。CeNA DMT保护的亚磷酰胺单体已被制备并用于按照经典的亚磷酰胺化学进行寡聚化合物合成。已经制备并研究了完全修饰的CeNA寡聚化合物和具有用CeNA修饰的特定位置的寡核苷酸(参见Wang等人, J. Am. Chem. Soc., 2000, 122, 8595-8602)。通常,将CeNA单体掺入DNA链中增加了其DNA/RNA杂交体的稳定性。CeNA寡腺苷酸以与天然复合物相似的稳定性而与RNA和DNA互补物形成复合物。通过NMR和圆二色性显示将CeNA结构掺入天然核酸结构中的研究,以进行简单的构象适应。

[0348] 另一修饰包括锁定核酸(LNA),其中2'-羟基与糖环的4'碳原子连接,从而形成2'-C,4'-C-氧亚甲基键,从而形成双环糖部分。键可以是亚甲基(-CH₂-)、桥接2'氧原子与4'碳原子的基团,其中n是1或2(Singh等人, Chem. Commun., 1998, 4, 455-456)。LNA和LNA类似物显示出非常高的互补DNA和RNA的双链热稳定性(T_m=+3至+10°C),朝向3'-核酸外切降解的稳定性和良好的溶解性。已经描述了含有LNA的有效和无毒的反义寡核苷酸(Wahlestedt等人, Proc. Natl. Acad. Sci. U.S.A., 2000, 97, 5633-5638)。

[0349] 已经描述了LNA单体腺嘌呤、胞嘧啶、鸟嘌呤、5-甲基胞嘧啶、胸腺嘧啶和尿嘧啶的合成和制备,以及它们的寡聚化和核酸识别性质(Koshkin等人, Tetrahedron, 1998, 54, 3607-3630)。LNA及其制备还描述于WO 98/39352和WO 99/14226中。

[0350] 经修饰的糖部分

[0351] 经标记的检测RNA还可包括一个或多个经取代的糖部分。合适的多核苷酸包含选自以下的糖取代基:OH;F;O-烷基、S-烷基或N-烷基;O-烯基、S-烯基或N-烯基;O-炔基、S-炔基或N-炔基;或O-烷基-O-烷基,其中所述烷基、烯基和炔基可以是取代或未取代的C₁至C₁₀烷基或C₂至C₁₀烯基和炔基。特别合适的是O((CH₂)_n)_mCH₃、O(CH₂)_nOCH₃、O(CH₂)_nNH₂、O(CH₂)_nCH₃、O(CH₂)_nONH₂、以及O(CH₂)_nON((CH₂)_nCH₃)₂,其中n和m为1至约10。其它合适的多核苷酸包含选自以下的糖取代基:C₁至C₁₀低级烷基,取代的低级烷基、烯基、炔基,烷芳基,芳烷基,0-烷芳基或O-芳烷基,SH,SCH₃,OCN,Cl,Br,CN,CF₃,OCF₃,SOCH₃,SO₂CH₃,ON₂,NO₂,N₃,NH₂,杂环

烷基,杂环烷基,氨基烷基,聚烷基氨基,取代的甲硅烷基,RNA切割基团,报告基团,嵌入剂,用于改善寡核苷酸的药代动力学性质的基团,或用于改善寡核苷酸的药效学性质的基团,以及其它具有相似性质的取代基。合适的修饰包括2'-甲氧基乙氧基(2'-O-CH₂CH₂OCH₃,也称为2'-O-(2-甲氧基乙基)或2'-MOE)(Martin等人,Helv.Chim.Acta,1995,78,486-504),即烷氧基烷氧基基团。另一合适的修饰包括2'-二甲基氨氧乙氧基,即O(CH₂)₂ON(CH₃)₂基团,也称为2'-DMAOE,如下文实施例中所述;以及2'-二甲氨基乙氧基乙氧基(在本领域中也称为2'-O-二甲基-氨基-乙氧基-乙基或2'-DMAEOE),即2'-O-CH₂-O-CH₂-N(CH₃)₂。[0352] 其它合适的糖取代基包括甲氧基(-O-CH₃)、氨基丙氧基(-O-CH₂CH₂CH₂NH₂)、烯丙基(-CH₂-CH=CH₂)、-O-烯丙基(-O-CH₂-CH=CH₂)和氟(F)。2'-糖取代基可以处于阿拉伯糖(上)位置或核糖(下)位置。合适的2'-阿拉伯糖修饰是2'-F。也可以在寡聚化合物上的其它位置,特别是3'末端核苷上或2'-5'连接的寡核苷酸的糖的3'位置处,以及5'末端核苷酸的5'位置处进行类似的修饰。寡聚化合物也可以具有糖模拟物,诸如环丁基部分代替戊呋喃糖基糖。

[0353] 碱基修饰和取代

[0354] 经标记的检测RNA还可以包含核碱基(在本领域中通常简称为“碱基”)修饰或取代。如本文所用,“未修饰的”或“天然的”核碱基包括嘌呤碱基腺嘌呤(A)和鸟嘌呤(G),以及嘧啶碱基胸腺嘧啶(T)、胞嘧啶(C)和尿嘧啶(U)。经修饰的核碱基包括其它合成和天然核碱基,诸如5-甲基胞嘧啶(5-me-C)、5-羟甲基胞嘧啶、黄嘌呤、次黄嘌呤、2-氨基腺嘌呤、6-甲基腺嘌呤和鸟嘌呤以及腺嘌呤和鸟嘌呤的其它烷基衍生物、2-丙基腺嘌呤和鸟嘌呤以及腺嘌呤和鸟嘌呤的其它烷基衍生物、2-硫尿嘧啶、2-硫代胸腺嘧啶和2-硫代胞嘧啶、5-氟尿嘧啶和胞嘧啶、5-丙炔基(-C≡C-CH₃)尿嘧啶和胞嘧啶以及嘧啶碱基的其它炔基衍生物、6-偶氮尿嘧啶、胞嘧啶和胸腺嘧啶、5-尿嘧啶(假尿嘧啶)、4-硫尿嘧啶、8-卤代腺嘌呤和鸟嘌呤、8-氨基腺嘌呤和鸟嘌呤、8-巯基腺嘌呤和鸟嘌呤、8-硫代烷基腺嘌呤和鸟嘌呤、8-羟基腺嘌呤和鸟嘌呤以及其它8-取代的腺嘌呤和鸟嘌呤、5-卤代(特别是5-溴代)尿嘧啶和胞嘧啶、5-三氟甲基尿嘧啶和胞嘧啶以及其它5-取代的尿嘧啶和胞嘧啶、7-甲基鸟嘌呤和7-甲基腺嘌呤、2-F-腺嘌呤、2-氨基腺嘌呤、8-氮杂鸟嘌呤和8-氮杂腺嘌呤、7-脱氮杂鸟嘌呤和7-脱氮杂腺嘌呤,以及3-脱氮杂鸟嘌呤和3-脱氮杂腺嘌呤。其它经修饰的核碱基包括三环嘧啶,诸如吩噻嗪胞苷(1H-嘧啶并(5,4-b)(1,4)苯并噻嗪-2(3H)-酮)、吩噻嗪胞苷(1H-嘧啶(5,4-b)(1,4)苯并噻嗪-2(3H)-酮)、G-钳如取代的吩噻嗪胞苷(例如9-(2-氨基乙氧基)-H-嘧啶并(5,4-b)(1,4)苯并噻嗪-2(3H)-酮)、吡唑胞苷(2H-嘧啶并(4,5-b)吡唑-2-酮)、吡啶并吡唑胞苷(H-吡啶并(3',2':4,5)吡咯并(2,3-d)嘧啶-2-酮)。

[0355] 杂环碱基部分还可包括嘌呤碱基或嘧啶碱基被其它杂环取代的那些杂环碱基部分,例如7-脱氮杂腺嘌呤、7-脱氮杂鸟苷、2-氨基吡啶和2-吡啶酮。其它核碱基包括在美国专利号3,687,808中公开的那些核碱基;在The Concise Encyclopedia Of Polymer Science And Engineering,第858-859页,Kroschwitz,J.I.编辑,John Wiley&Sons,1990中公开的那些核碱基;由Englisch等人,Angewandte Chemie,International Edition,1991,30,613公开的那些核碱基;以及由Sanghvi,Y.S.,第15章,Antisense Research and Applications,第289-302页,Crooke,S.T.和Lebleu,B.编辑,CRC Press,1993公开的那些核碱基。这些核碱基中的某些可用于增加寡聚化合物的结合亲和力。这些核碱基包括5-取

代的嘧啶、6-氮杂嘧啶,以及N-2、N-6和O-6取代的嘌呤,包括2-氨基丙基腺嘌呤、5-丙炔基尿嘧啶和5-丙炔基胞嘧啶。5-甲基胞嘧啶取代已显示为使核酸双链体稳定性增加0.6-1.2℃。(Sanghvi等人编辑,Antisense Research and Applications,CRC Press,Boca Raton,1993,第276-278页),并且是合适的碱基取代,例如当与2'-O-甲氧基乙基糖修饰组合时。

[0356] 3) 包含两种不同C2c2蛋白和两种不同经标记的检测RNA的试剂盒

[0357] 在一些情况下,主题试剂盒包含:(a) 第一经标记的检测RNA,其缺乏U但包含至少一个A(例如,至少2个、至少3个、或至少4个A)并且包含第一FRET对和/或第一猝灭剂/荧光剂对;(b) 第二经标记的检测RNA,其缺乏A但包含至少一个U(例如,至少2个、至少3个或至少4个U)并且包含第二FRET对和/或第二猝灭剂/荧光剂对;(c) 第一C2c2蛋白和/或编码所述第一C2c2蛋白的核酸,其中所述第一C2c2蛋白在被激活时切割腺嘌呤⁺RNA(包括A的RNA)而不切割缺乏A的RNA(例如,聚U RNA)[例如,第一C2c2蛋白可以切割第一经标记的检测RNA而不切割第二经标记的检测RNA];以及(d) 第二C2c2蛋白和/或编码所述第二C2c2蛋白的核酸,其中所述第二C2c2蛋白在被激活时切割尿嘧啶⁺RNA(包括U的RNA)而不切割缺乏U的RNA(例如,聚A RNA)(例如,第二C2c2蛋白可以切割第二经标记的检测RNA而不切割第一经标记的检测RNA)。

[0358] 在一些情况下,第一经标记的检测RNA缺乏U并且包括2至15个连续的A(例如,2至12个、2至10个、2至8个、2至6个、2至4个、3至15个、3至12个、3至10个、3至8个、3至6个、3至5个、4至15个、4至12个、4至10个、4至8个或4至6个连续的A)的段。在一些情况下,第一经标记的检测RNA缺乏U并且包括至少2个连续的A(例如,至少3个、至少4个或至少5个连续的A)的段。在一些情况下,第二经标记的检测RNA缺乏A并且包括2至15个连续的U(例如,2至12个、2至10个、2至8个、2至6个、2至4个、3至15个、3至12个、3至10个、3至8个、3至6个、3至5个、4至15个、4至12个、4至10个、4至8个或4至6个连续的U)的段。在一些情况下,第二经标记的检测RNA缺乏A并且包括至少2个连续的U(例如,至少3个、至少4个或至少5个连续的U)的段。

[0359] 在一些情况下,此类试剂盒还包括:(e) 第一C2c2指导RNA(和/或编码所述第一C2c2指导RNA的核酸),例如包含编码所述第一C2c2指导RNA的核苷酸序列的核酸,其中所述核酸包括供用户插入指导序列(例如,与靶RNA杂交的核苷酸序列)的序列插入位点;以及(f) 第二C2c2指导RNA(和/或编码所述第二C2c2指导RNA的核酸),例如包含编码所述第二C2c2指导RNA的核苷酸序列的核酸,其中所述核酸包括供用户插入指导序列(例如,与靶RNA杂交的核苷酸序列)的序列插入位点)。第一C2c2指导RNA包含与第一单链靶RNA杂交的第一核苷酸序列和与第一C2c2蛋白结合的第二核苷酸序列。第二C2c2指导RNA包含与第二单链靶RNA杂交的第一核苷酸序列和与第二C2c2蛋白结合的第二核苷酸序列。第一C2c2蛋白不被第二C2c2指导RNA激活,并且第一C2c2蛋白切割包括至少一个A的ssRNA(例如,不切割缺乏A的ssRNA)。第二个C2c2蛋白不被第一C2c2指导RNA激活,并且第二C2c2蛋白切割包括至少一个U的ssRNA(例如,不切割缺乏U的ssRNA)。

[0360] 以下是适合于包括在本公开的试剂盒中的第一和第二C2c2蛋白的非限制性实例(下文列出为a)至w)):

[0361] a) 第一C2c2蛋白包含与图56F所示的Lba Cas13a氨基酸序列具有至少75%、至少80%、至少85%、至少90%、至少95%、至少98%、至少99%或100%的氨基酸序列同一性的氨基酸序列;并且第二C2c2蛋白包含与图56K所示的Hhe Cas13a氨基酸序列具有至少75%、

至少80%、至少85%、至少90%、至少95%、至少98%、至少99%或100%的氨基酸序列同一性的氨基酸序列；

[0362] b) 第一C2c2蛋白包含与图56F所示的Lba Cas13a氨基酸序列具有至少75%、至少80%、至少85%、至少90%、至少95%、至少98%、至少99%或100%的氨基酸序列同一性的氨基酸序列；并且第二C2c2蛋白包含与图56G所示的Rca Cas13a氨基酸序列具有至少75%、至少80%、至少85%、至少90%、至少95%、至少98%、至少99%或100%的氨基酸序列同一性的氨基酸序列；

[0363] c) 第一C2c2蛋白包含与图56F所示的Lba Cas13a氨基酸序列具有至少75%、至少80%、至少85%、至少90%、至少95%、至少98%、至少99%或100%的氨基酸序列同一性的氨基酸序列；并且第二C2c2蛋白包含与图56B所示的Ppr Cas13a氨基酸序列具有至少75%、至少80%、至少85%、至少90%、至少95%、至少98%、至少99%或100%的氨基酸序列同一性的氨基酸序列；

[0364] d) 第一C2c2蛋白包含与图56F所示的Lba Cas13a氨基酸序列具有至少75%、至少80%、至少85%、至少90%、至少95%、至少98%、至少99%或100%的氨基酸序列同一性的氨基酸序列；并且第二C2c2蛋白包含与图56I所示的Lne Cas13a氨基酸序列具有至少75%、至少80%、至少85%、至少90%、至少95%、至少98%、至少99%或100%的氨基酸序列同一性的氨基酸序列；

[0365] e) 第一C2c2蛋白包含与图56F所示的Lba Cas13a氨基酸序列具有至少75%、至少80%、至少85%、至少90%、至少95%、至少98%、至少99%或100%的氨基酸序列同一性的氨基酸序列；并且第二C2c2蛋白包含与图56C所示的Lbu Cas13a氨基酸序列具有至少75%、至少80%、至少85%、至少90%、至少95%、至少98%、至少99%或100%的氨基酸序列同一性的氨基酸序列；

[0366] f) 第一C2c2蛋白包含与图56F所示的Lba Cas13a氨基酸序列具有至少75%、至少80%、至少85%、至少90%、至少95%、至少98%、至少99%或100%的氨基酸序列同一性的氨基酸序列；并且第二C2c2蛋白包含与图56E所示的Lwa Cas13a氨基酸序列具有至少75%、至少80%、至少85%、至少90%、至少95%、至少98%、至少99%或100%的氨基酸序列同一性的氨基酸序列；

[0367] g) 第一C2c2蛋白包含与图56F所示的Lba Cas13a氨基酸序列具有至少75%、至少80%、至少85%、至少90%、至少95%、至少98%、至少99%或100%的氨基酸序列同一性的氨基酸序列；并且第二C2c2蛋白包含与图56D所示的Lsh Cas13a氨基酸序列具有至少75%、至少80%、至少85%、至少90%、至少95%、至少98%、至少99%或100%的氨基酸序列同一性的氨基酸序列；

[0368] h) 第一C2c2蛋白包含与图56J所示的Ere Cas13a氨基酸序列具有至少75%、至少80%、至少85%、至少90%、至少95%、至少98%、至少99%或100%的氨基酸序列同一性的氨基酸序列；并且第二C2c2蛋白包含与图56K所示的Hhe Cas13a氨基酸序列具有至少75%、至少80%、至少85%、至少90%、至少95%、至少98%、至少99%或100%的氨基酸序列同一性的氨基酸序列；

[0369] i) 第一C2c2蛋白包含与图56J所示的Ere Cas13a氨基酸序列具有至少75%、至少80%、至少85%、至少90%、至少95%、至少98%、至少99%或100%的氨基酸序列同一性的

氨基酸序列;并且第二C2c2蛋白包含与图56G所示的Rca Cas13a氨基酸序列具有至少75%、至少80%、至少85%、至少90%、至少95%、至少98%、至少99%或100%的氨基酸序列同一性的氨基酸序列;

[0370] j) 第一C2c2蛋白包含与图56J所示的Ere Cas13a氨基酸序列具有至少75%、至少80%、至少85%、至少90%、至少95%、至少98%、至少99%或100%的氨基酸序列同一性的氨基酸序列;并且第二C2c2蛋白包含与图56B所示的Ppr Cas13a氨基酸序列具有至少75%、至少80%、至少85%、至少90%、至少95%、至少98%、至少99%或100%的氨基酸序列同一性的氨基酸序列;

[0371] k) 第一C2c2蛋白包含与图56J所示的Ere Cas13a氨基酸序列具有至少75%、至少80%、至少85%、至少90%、至少95%、至少98%、至少99%或100%的氨基酸序列同一性的氨基酸序列;并且第二C2c2蛋白包含与图56I所示的Lne Cas13a氨基酸序列具有至少75%、至少80%、至少85%、至少90%、至少95%、至少98%、至少99%或100%的氨基酸序列同一性的氨基酸序列;

[0372] l) 第一C2c2蛋白包含与图56J所示的Ere Cas13a氨基酸序列具有至少75%、至少80%、至少85%、至少90%、至少95%、至少98%、至少99%或100%的氨基酸序列同一性的氨基酸序列;并且第二C2c2蛋白包含与图56C所示的Lbu Cas13a氨基酸序列具有至少75%、至少80%、至少85%、至少90%、至少95%、至少98%、至少99%或100%的氨基酸序列同一性的氨基酸序列;

[0373] m) 第一C2c2蛋白包含与图56J所示的Ere Cas13a氨基酸序列具有至少75%、至少80%、至少85%、至少90%、至少95%、至少98%、至少99%或100%的氨基酸序列同一性的氨基酸序列;并且第二C2c2蛋白包含与图56E所示的Lwa Cas13a氨基酸序列具有至少75%、至少80%、至少85%、至少90%、至少95%、至少98%、至少99%或100%的氨基酸序列同一性的氨基酸序列;

[0374] n) 第一C2c2蛋白包含与图56J所示的Ere Cas13a氨基酸序列具有至少75%、至少80%、至少85%、至少90%、至少95%、至少98%、至少99%或100%的氨基酸序列同一性的氨基酸序列;并且第二C2c2蛋白包含与图56D所示的Lsh Cas13a氨基酸序列具有至少75%、至少80%、至少85%、至少90%、至少95%、至少98%、至少99%或100%的氨基酸序列同一性的氨基酸序列;

[0375] o) 第一C2c2蛋白包含与图56J所示的Ere Cas13a氨基酸序列具有至少75%、至少80%、至少85%、至少90%、至少95%、至少98%、至少99%或100%的氨基酸序列同一性的氨基酸序列;并且第二C2c2蛋白包含与图56A所示的Lse Cas13a氨基酸序列具有至少75%、至少80%、至少85%、至少90%、至少95%、至少98%、至少99%或100%的氨基酸序列同一性的氨基酸序列;

[0376] p) 第一C2c2蛋白包含与图56H所示的Cam Cas13a氨基酸序列具有至少75%、至少80%、至少85%、至少90%、至少95%、至少98%、至少99%或100%的氨基酸序列同一性的氨基酸序列;并且第二C2c2蛋白包含与图56K所示的Hhe Cas13a氨基酸序列具有至少75%、至少80%、至少85%、至少90%、至少95%、至少98%、至少99%或100%的氨基酸序列同一性的氨基酸序列;

[0377] q) 第一C2c2蛋白包含与图56H所示的Cam Cas13a氨基酸序列具有至少75%、至少

80%、至少85%、至少90%、至少95%、至少98%、至少99%或100%的氨基酸序列同一性的氨基酸序列；并且第二C2c2蛋白包含与图56G所示的Rca Cas13a氨基酸序列具有至少75%、至少80%、至少85%、至少90%、至少95%、至少98%、至少99%或100%的氨基酸序列同一性的氨基酸序列；

[0378] r) 第一C2c2蛋白包含与图56H所示的Cam Cas13a氨基酸序列具有至少75%、至少80%、至少85%、至少90%、至少95%、至少98%、至少99%或100%的氨基酸序列同一性的氨基酸序列；并且第二C2c2蛋白包含与图56B所示的Ppr Cas13a氨基酸序列具有至少75%、至少80%、至少85%、至少90%、至少95%、至少98%、至少99%或100%的氨基酸序列同一性的氨基酸序列；

[0379] s) 第一C2c2蛋白包含与图56H所示的Cam Cas13a氨基酸序列具有至少75%、至少80%、至少85%、至少90%、至少95%、至少98%、至少99%或100%的氨基酸序列同一性的氨基酸序列；并且第二C2c2蛋白包含与图56I所示的Lne Cas13a氨基酸序列具有至少75%、至少80%、至少85%、至少90%、至少95%、至少98%、至少99%或100%的氨基酸序列同一性的氨基酸序列；

[0380] t) 第一C2c2蛋白包含与图56H所示的Cam Cas13a氨基酸序列具有至少75%、至少80%、至少85%、至少90%、至少95%、至少98%、至少99%或100%的氨基酸序列同一性的氨基酸序列；并且第二C2c2蛋白包含与图56C所示的Lbu Cas13a氨基酸序列具有至少75%、至少80%、至少85%、至少90%、至少95%、至少98%、至少99%或100%的氨基酸序列同一性的氨基酸序列；

[0381] u) 第一C2c2蛋白包含与图56H所示的Cam Cas13a氨基酸序列具有至少75%、至少80%、至少85%、至少90%、至少95%、至少98%、至少99%或100%的氨基酸序列同一性的氨基酸序列；并且第二C2c2蛋白包含与图56E所示的Lwa Cas13a氨基酸序列具有至少75%、至少80%、至少85%、至少90%、至少95%、至少98%、至少99%或100%的氨基酸序列同一性的氨基酸序列；

[0382] v) 第一C2c2蛋白包含与图56H所示的Cam Cas13a氨基酸序列具有至少75%、至少80%、至少85%、至少90%、至少95%、至少98%、至少99%或100%的氨基酸序列同一性的氨基酸序列；并且第二C2c2蛋白包含与图56F所示的Lsh Cas13a氨基酸序列具有至少75%、至少80%、至少85%、至少90%、至少95%、至少98%、至少99%或100%的氨基酸序列同一性的氨基酸序列；或者

[0383] w) 第一C2c2蛋白包含与图56H所示的Cam Cas13a氨基酸序列具有至少75%、至少80%、至少85%、至少90%、至少95%、至少98%、至少99%或100%的氨基酸序列同一性的氨基酸序列；并且第二C2c2蛋白包含与图56A所示的Lse Cas13a氨基酸序列具有至少75%、至少80%、至少85%、至少90%、至少95%、至少98%、至少99%或100%的氨基酸序列同一性的氨基酸序列。

[0384] C2c2蛋白

[0385] 适合于包括在本公开的试剂盒中的C2c2蛋白与C2c2指导RNA结合，由指导RNA（其与靶RNA杂交）引导到单链靶RNA，从而被“激活”。如果C2c2蛋白的HEPN1结构域和HEPN2结构域是完整的，则C2c2蛋白一旦被激活就切割靶RNA，但是也切割非靶RNA。

[0386] 在一些情况下，适合于包括在本公开的试剂盒中的C2c2蛋白包括与SEQ ID NO:1

至SEQ ID NO:6中任一项所示的氨基酸序列具有80%或更多(例如,85%或更多、90%或更多、95%或更多、98%或更多、99%或更多、99.5%或更多、或100%的氨基酸序列同一性的氨基酸序列。在一些情况下,适合于包括在本公开的试剂盒中的C2c2蛋白包含与SEQ ID NO:1所示的斯氏李斯特菌C2c2氨基酸序列具有至少80%、至少85%、至少90%、至少95%、至少98%、至少99%或100%的氨基酸序列同一性的氨基酸序列。在一些情况下,适合于包括在本公开的试剂盒中的C2c2蛋白包含与SEQ ID NO:2所示的口腔纤毛菌C2c2氨基酸序列具有至少80%、至少85%、至少90%、至少95%、至少98%、至少99%或100%的氨基酸序列同一性的氨基酸序列。在一些情况下,适合于包括在本公开的试剂盒中的C2c2蛋白包含与SEQ ID NO:4所示的荚膜红细菌C2c2氨基酸序列具有至少80%、至少85%、至少90%、至少95%、至少98%、至少99%或100%的氨基酸序列同一性的氨基酸序列。在一些情况下,适合于包括在本公开的试剂盒中的C2c2蛋白包含与SEQ ID NO:5所示的鸡肉杆菌C2c2氨基酸序列具有至少80%、至少85%、至少90%、至少95%、至少98%、至少99%或100%的氨基酸序列同一性的氨基酸序列。在一些情况下,适合于包括在本公开的试剂盒中的C2c2蛋白包含与SEQ ID NO:6所示的Herbinix hemicellulosilytica C2c2氨基酸序列具有至少80%、至少85%、至少90%、至少95%、至少98%、至少99%或100%的氨基酸序列同一性的氨基酸序列。在一些情况下,适合于包括在本公开的试剂盒中的C2c2蛋白包括与SEQ ID NO:2所示的口腔纤毛菌(Lbu) C2c2氨基酸序列具有80%或更多的氨基酸序列同一性的氨基酸序列。在一些情况下,适合于包括在本公开的试剂盒中的C2c2蛋白是口腔纤毛菌(Lbu) C2c2蛋白(例如,参见SEQ ID NO:2)。在一些情况下,适合于包括在本公开的试剂盒中的C2c2蛋白包括SEQ ID NO:1至SEQ ID NO:2和SEQ ID NO:4至SEQ ID NO:6中任一项所示的氨基酸序列。

[0387] 在一些情况下,包括在本公开的试剂盒中的C2c2蛋白不是沙氏纤毛菌(Lsh) C2c2蛋白。在一些情况下,包括在本公开的试剂盒中的C2c2蛋白不是与SEQ ID NO:3所示的Lsh C2c2多肽具有至少80%、至少85%、至少90%、至少95%、至少98%、至少99%或100%的氨基酸序列同一性的C2c2多肽。

[0388] 在一些情况下,适合于包括在本公开的试剂盒中的C2c2蛋白比沙氏纤毛菌(Lsh) C2c2蛋白在切割未由所述方法的C2c2指导RNA靶向的RNA方面更有效1.2倍或更多倍。在一些情况下,所述C2c2蛋白比沙氏纤毛菌(Lsh) C2c2蛋白在切割未由所述方法的C2c2指导RNA靶向的RNA方面更有效1.5倍或更多倍。在一些情况下,在本公开的方法中使用的C2c2多肽当被激活时,比Lsh C2c2至少1.2倍、至少1.5倍、至少2倍、至少2.5倍、至少3倍、至少4倍、至少5倍、至少6倍、至少7倍、至少8倍、至少9倍、至少10倍、至少15倍、至少20倍、至少30倍、或大于30倍有效地切割非靶RNA。

[0389] 在一些情况下,适合于包括在本公开的试剂盒中的C2c2蛋白在所述接触的1小时内表现出至少50%的RNA切割效率(例如,55%或更多、60%或更多、65%或更多、70%或更多、或75%或更多的切割效率)。在一些情况下,适合于包括在本公开的试剂盒中的C2c2蛋白在所述接触的40分钟内表现出至少50%的RNA切割效率(例如,55%或更多、60%或更多、65%或更多、70%或更多、或75%或更多的切割效率)。在一些情况下,适合于包括在本公开的试剂盒中的C2c2蛋白在所述接触的30分钟内表现出至少50%的RNA切割效率(例如,55%或更多、60%或更多、65%或更多、70%或更多、或75%或更多的切割效率)。

[0390] 在一些情况下,适合于包括在本公开的试剂盒中的C2c2蛋白在30秒至60分钟(例

如,1分钟至60分钟、30秒至5分钟、1分钟至5分钟、1分钟至10分钟、5分钟至10分钟、10分钟至15分钟、15分钟至20分钟、20分钟至25分钟、25分钟至30分钟、30分钟至35分钟、35分钟至40分钟、40分钟至45分钟、45分钟至50分钟、50分钟至55分钟或55分钟至60分钟)的时间段内切割样品中存在的RNA的至少50%、至少60%、至少70%、至少80%、至少90%、至少95%、至少98%、至少99%或大于99%。在一些情况下,适合于包括在本公开的试剂盒中的C2c2蛋白在30秒至5分钟的时间段内(例如,1分钟至5分钟,例如在1分钟、2分钟、3分钟、4分钟或5分钟的时间段内)切割样品中存在的RNA的至少50%、至少60%、至少70%、至少80%、至少90%、至少95%、至少98%、至少99%或大于99%。在一些情况下,适合于包括在本公开的试剂盒中的C2c2蛋白在5分钟至10分钟的时间段内(例如,在5分钟、6分钟、7分钟、8分钟、9分钟或10分钟的时间段内)切割样品中存在的RNA的至少50%、至少60%、至少70%、至少80%、至少90%、至少95%、至少98%、至少99%或大于99%。在一些情况下,适合于包括在本公开的试剂盒中的C2c2蛋白在10分钟至15分钟(例如,10分钟、11分钟、12分钟、13分钟、14分钟或15分钟)的时间段内切割样品中存在的RNA的至少50%、至少60%、至少70%、至少80%、至少90%、至少95%、至少98%、至少99%或大于99%。在一些情况下,适合于包括在本公开的试剂盒中的C2c2蛋白在15分钟至20分钟的时间段内切割样品中存在的RNA的至少50%、至少60%、至少70%、至少80%、至少90%、至少95%、至少98%、至少99%或大于99%。在一些情况下,适合于包括在本公开的试剂盒中的C2c2蛋白在20分钟至25分钟的时间段内切割样品中存在的RNA的至少50%、至少60%、至少70%、至少80%、至少90%、至少95%、至少98%、至少99%或大于99%。在一些情况下,适合于包括在本公开的试剂盒中的C2c2蛋白在25分钟至30分钟的时间段内切割样品中存在的RNA的至少50%、至少60%、至少70%、至少80%、至少90%、至少95%、至少98%、至少99%或大于99%。在一些情况下,适合于包括在本公开的试剂盒中的C2c2蛋白在30分钟至35分钟的时间段内切割样品中存在的RNA的至少50%、至少60%、至少70%、至少80%、至少90%、至少95%、至少98%、至少99%或大于99%。在一些情况下,适合于包括在本公开的试剂盒中的C2c2蛋白在35分钟至40分钟的时间段内切割样品中存在的RNA的至少50%、至少60%、至少70%、至少80%、至少90%、至少95%、至少98%、至少99%或大于99%。在一些情况下,适合于包括在本公开的试剂盒中的C2c2蛋白在40分钟至45分钟的时间段内切割样品中存在的RNA的至少50%、至少60%、至少70%、至少80%、至少90%、至少95%、至少98%、至少99%或大于99%。在一些情况下,适合于包括在本公开的试剂盒中的C2c2蛋白在45分钟至50分钟的时间段内切割样品中存在的RNA的至少50%、至少60%、至少70%、至少80%、至少90%、至少95%、至少98%、至少99%或大于99%。在一些情况下,适合于包括在本公开的试剂盒中的C2c2蛋白在50分钟至55分钟的时间段内切割样品中存在的RNA的至少50%、至少60%、至少70%、至少80%、至少90%、至少95%、至少98%、至少99%或大于99%。在一些情况下,适合于包括在本公开的试剂盒中的C2c2蛋白在55分钟至60分钟的时间段内切割样品中存在的RNA的至少50%、至少60%、至少70%、至少80%、至少90%、至少95%、至少98%、至少99%或大于99%。

[0391] 在一些情况下,包括在本公开的试剂盒中的C2c2蛋白不是沙氏纤毛菌(Lsh) C2c2蛋白。在一些情况下,C2c2蛋白比沙氏纤毛菌(Lsh) C2c2蛋白更有效(例如,在切割非靶RNA方面)1.2倍或更高(例如,1.5倍或更高、1.7倍或更高、或2倍或更高)。因此,在一些情况下,适合于包括在本公开的试剂盒中的C2c2蛋白比沙氏纤毛菌(Lsh) C2c2蛋白在切割未由所述

方法的C2c2指导RNA靶向的RNA方面更有效1.2倍或更多倍(例如,1.5倍或更多倍、1.7倍或更多倍、或2倍或更多倍)。在一些情况下,适合于包括在本公开的试剂盒中的C2c2蛋白当被激活时,比Lsh C2c2至少1.2倍、至少1.5倍、至少2倍、至少2.5倍、至少3倍、至少4倍、至少5倍、至少6倍、至少7倍、至少8倍、至少9倍、至少10倍、至少15倍、至少20倍、至少30倍或大于30倍有效地切割非靶RNA。

[0392] 阳性对照

[0393] 包含经标记的检测RNA和C2c2多肽的本公开的试剂盒还可以包括阳性对照靶RNA。在一些情况下,试剂盒还包括阳性对照指导RNA,所述阳性对照指导RNA包含与对照靶RNA杂交的核苷酸序列。在一些情况下,阳性对照靶RNA以不同的量在分开的容器中提供。在一些情况下,阳性对照靶RNA以各种已知浓度在分开的容器中与对照非靶RNA一起提供。

[0394] 编码C2c2指导RNA和/或前体C2c2指导RNA阵列和/或C2c2蛋白的核酸

[0395] 虽然本公开的RNA(例如,C2c2指导RNA和前体C2c2指导RNA阵列)可以使用任何合适的方法合成(例如,使用RNA聚合酶(例如,T7聚合酶、T3聚合酶、SP6聚合酶)在体外化学合成,等等),但是也设想了编码C2c2指导RNA和/或前体C2c2指导RNA阵列的核酸。另外,虽然可以以蛋白形式提供本公开的C2c2蛋白(例如,作为试剂盒的部分),但是也可以提供编码所述C2c2蛋白的核酸(诸如mRNA和/或DNA)。

[0396] 例如,在一些实施方案中,本公开的试剂盒包含核酸(例如,DNA,例如重组表达载体),所述核酸包含编码C2c2指导RNA的核苷酸序列。在一些情况下,核苷酸序列编码没有指导序列的C2c2指导RNA。例如,在一些情况下,核酸包含编码C2c2指导RNA(没有指导序列的C2c2指导RNA)的恒定区的核苷酸序列,并且包含编码指导序列的核酸的插入位点。在一些实施方案中,本公开的试剂盒包含核酸(例如,mRNA、DNA,例如重组表达载体),所述核酸包含编码C2c2蛋白的核苷酸序列。

[0397] 在一些实施方案中,本公开的试剂盒包含核酸(例如,DNA,例如,重组表达载体),所述核酸包含编码前体C2c2指导RNA阵列的核苷酸序列(例如,在一些情况下,其中阵列的每个指导RNA具有不同的指导序列)。在一些情况下,阵列的经编码的指导RNA中的一者或者多者不具有指导序列,例如,核酸可以包括所述阵列的一个或多个指导RNA的指导序列的插入位点。在一些情况下,主题C2c2指导RNA可以包括来自前体crRNA的柄,但不一定必须包括多个指导序列。

[0398] 在一些情况下,编码C2c2指导RNA的核苷酸序列(和/或编码前体C2c2指导RNA阵列的核苷酸序列)与启动子可操作地连接,所述启动子为例如在原核细胞中起作用的启动子、在真核细胞中起作用的启动子、在哺乳动物细胞中起作用的启动子、在人细胞中起作用的启动子等。在一些情况下,编码C2c2蛋白的核苷酸序列与启动子可操作地连接,所述启动子为例如在原核细胞中起作用的启动子、在真核细胞中起作用的启动子、在哺乳动物细胞中起作用的启动子、在人细胞中起作用的启动子、细胞型特异性启动子、可调控的启动子、组织特异性启动子等。

[0399] 本公开的非限制性方面的实例

[0400] 上述本主题的各方面(包括实施方案)可以单独或与一个或多个其它方面或实施方案一起为有益的。在不限前述描述的情况下,下面提供了编号1-90的本公开的某些非限制性方面。对于本领域的技术人员在阅读本公开时将显而易见的是,每个单独编号的方

面可以使用或与前面或后面的单独编号的方面中的任一者一起组合。这旨在为各方面的所有此类组合提供支持,并且不限于以下明确提供的各方面的组合:

[0401] 方面1.一种检测包含多种RNA的样品中的单链靶RNA的方法,所述方法包括:

[0402] a) 使所述样品与以下接触:(i) C2c2指导RNA,其与所述单链靶RNA杂交;和(ii) C2c2蛋白,其切割所述样品中存在的RNA;以及

[0403] b) 测量由C2c2蛋白介导的RNA切割产生的可检测信号。

[0404] 方面2.一种检测包含多种RNA的样品中的单链靶RNA的方法,所述方法包括:

[0405] (a) 使样品与以下接触:i) 前体C2c2指导RNA阵列,其包含两个或更多个C2c2指导RNA,每个所述C2c2指导RNA具有不同的指导序列;和(ii) C2c2蛋白,其将所述前体C2c2指导RNA阵列切割成单独的C2c2指导RNA,并且还切割所述样品的RNA;以及

[0406] (b) 测量由C2c2蛋白介导的RNA切割产生的可检测信号。

[0407] 方面3.根据方面1或方面2所述的方法,其中所述C2c2蛋白在所述接触1小时内切割所述样品中存在的所述RNA的至少50%。

[0408] 方面4.根据方面3所述的方法,其中所述C2c2蛋白在所述接触40分钟内切割所述样品中存在的所述RNA的至少50%。

[0409] 方面5.根据方面4所述的方法,其中所述C2c2蛋白在所述接触5分钟内切割所述样品中存在的所述RNA的至少50%。

[0410] 方面6.根据方面5所述的方法,其中所述C2c2蛋白在所述接触1分钟内切割所述样品中存在的所述RNA的至少50%。

[0411] 方面7.根据方面1所述的方法,其中所述C2c2蛋白在所述接触1分钟内切割所述样品中存在的所述RNA的50%至多于90%。

[0412] 方面8.根据方面1-7中任一方面所述的方法,其中可以检测到所述单链靶RNA的最小浓度在500fM至1nM的范围内。

[0413] 方面9.根据方面1-7中任一方面所述的方法,其中可以在低至800fM的浓度下检测到所述单链靶RNA。

[0414] 方面10.根据方面1-9中任一方面所述的方法,其中所述C2c2蛋白不是包含与SEQ ID NO:3所示的氨基酸序列具有至少80%、至少85%、至少90%、至少95%、至少98%、至少99%或100%的氨基酸序列同一性的氨基酸序列的沙氏纤毛菌(Lsh) C2c2蛋白。

[0415] 方面11.根据方面10所述的方法,其中所述C2c2蛋白比包含与SEQ ID NO:3所示的氨基酸序列至少80%、至少85%、至少90%、至少95%、至少98%、至少99%或100%的氨基酸序列同一性的沙氏纤毛菌(Lsh) C2c2蛋白至少1.2倍更有效地切割非靶RNA。

[0416] 方面12.根据方面11所述的方法,其中所述C2c2蛋白比包含SEQ ID NO:3所示的氨基酸序列的沙氏纤毛菌(Lsh) C2c2蛋白至少1.5倍更有效地切割非靶RNA。

[0417] 方面13.根据方面1-9中任一方面所述的方法,其中所述C2c2蛋白包含与SEQ ID NO:1、SEQ ID NO:2或SEQ ID NO:4至SEQ ID NO:6中任一者所示的氨基酸序列具有80%或更多氨基酸序列同一性的氨基酸序列。

[0418] 方面14.根据方面1-9中任一方面所述的方法,其中所述C2c2蛋白包含与SEQ ID NO:2所示的口腔纤毛菌(Lbu) C2c2氨基酸序列具有80%或更多氨基酸序列同一性的氨基酸序列。

[0419] 方面15.根据方面1-9中任一方面所述的方法,其中所述C2c2蛋白包含与SEQ ID NO:1所示的斯氏李斯特菌C2c2氨基酸序列具有80%或更多氨基酸序列同一性的氨基酸序列。

[0420] 方面16.根据方面1-9中任一方面所述的方法,其中所述C2c2蛋白包含SEQ ID NO:1至SEQ ID NO:2和SEQ ID NO:4至SEQ ID NO:6中任一者所示的氨基酸序列。

[0421] 方面17.根据方面1所述的方法,其中所述C2c2蛋白包含与SEQ ID NO:2所示的C2c2氨基酸序列具有至少80%氨基酸序列同一性的氨基酸序列,并且包含对R472、H477、R1048和H1053中的一者或多者的取代。

[0422] 方面18.根据方面1所述的方法,其中所述C2c2蛋白包含与SEQ ID NO:2所示的C2c2氨基酸序列具有至少80%氨基酸序列同一性的氨基酸序列,并且包含对氨基酸R472和H477的取代。

[0423] 方面19.根据方面1所述的方法,其中所述C2c2蛋白包含与SEQ ID NO:2所示的C2c2氨基酸序列具有至少80%氨基酸序列同一性的氨基酸序列,并且包含对氨基酸R1048和H1053的取代。

[0424] 方面20.根据方面1所述的方法,其中所述C2c2蛋白包含与SEQ ID NO:2所示的C2c2氨基酸序列具有至少80%氨基酸序列同一性的氨基酸序列,并且包含对氨基酸R472、H477、R1048和H1053的取代。

[0425] 方面21.根据方面1-20中任一方面所述的方法,其中在所述测量之前使所述样品接触2小时或更短时间。

[0426] 方面22.根据方面21所述的方法,其中在所述测量之前使所述样品接触60分钟或更短时间。

[0427] 方面23.根据方面22所述的方法,其中在所述测量之前使所述样品接触30分钟或更短时间。

[0428] 方面24.根据方面23所述的方法,其中在所述测量之前使所述样品接触10分钟或更短时间。

[0429] 方面25.根据方面24所述的方法,其中在所述测量之前使所述样品接触1分钟或更短时间。

[0430] 方面26.根据方面1-25中任一方面所述的方法,其包括确定在所述样品中存在的靶RNA的量。

[0431] 方面27.根据方面26所述的方法,其中所述确定包括:

[0432] 测量所述可检测信号以产生测试测量值;

[0433] 测量由参考样品产生的可检测信号以产生参考测量值;以及

[0434] 将所述测试测量值与所述参考测量值进行比较以确定所述样品中存在的靶RNA的量。

[0435] 方面28.根据方面26所述的方法,其包括:

[0436] 测量所述可检测信号以产生测试测量值,

[0437] 测量由两个或更多个参考样品中的每一个产生的可检测信号,其中所述两个或更多个参考样品各自包括不同量的阳性对照RNA,以产生两个或更多个参考测量值,以及

[0438] 将所述测试测量值与所述两个或更多个参考测量值进行比较以确定所述样品中

存在的靶RNA的量。

[0439] 方面29.根据方面1-28中任一方面所述的方法,其中所述样品包含5至 10^7 个序列彼此不同的RNA。

[0440] 方面30.根据方面1-28中任一方面所述的方法,其中所述样品包含10至 10^6 个序列彼此不同的RNA。

[0441] 方面31.根据方面1-30中任一方面所述的方法,其中所述样品包含来自细胞裂解物的RNA。

[0442] 方面32.根据方面1-31中任一方面所述的方法,其中测量可检测信号包括以下中的一者或多者:基于金纳米颗粒的检测、荧光偏振、胶体相变/分散、电化学检测和基于半导体的感测。

[0443] 方面33.根据方面1-32中任一方面所述的方法,其中(i)所述方法包括使所述样品与包含荧光发射染料对(即,荧光共振能量转移(FRET)对和/或猝灭剂/荧光剂对)的经标记的检测RNA接触,(ii)所述C2c2蛋白切割所述经标记的检测RNA,以及(iii)所述可检测信号由所述FRET对和/或所述猝灭剂/荧光剂对产生。

[0444] 方面34.根据方面33所述的方法,其中所述经标记的检测RNA在被切割之前产生一定量的可检测信号,并且当所述经标记的检测RNA被切割时可检测信号的量减少。

[0445] 方面35.根据方面33所述的方法,其中所述经标记的检测RNA在被切割之前产生第一可检测信号,并且当所述经标记的检测RNA被切割时产生第二可检测信号。

[0446] 方面36.根据方面35所述的方法,其中所述经标记的检测RNA包含FRET对和猝灭剂/荧光剂对。

[0447] 方面37.根据方面33-36中任一方面所述的方法,其中所述经标记的检测RNA包含FRET对。

[0448] 方面38.根据方面33所述的方法,其中在所述经标记的检测RNA被切割时产生可检测信号。

[0449] 方面39.根据方面33所述的方法,其中当所述经标记的检测RNA被切割时,可检测信号的量增加。

[0450] 方面40.根据方面38或方面39所述的方法,其中所述经标记的检测RNA包含猝灭剂/荧光剂对。

[0451] 方面41.根据方面33-40中任一方面所述的方法,其中所述经标记的检测RNA包含经修饰的核碱基、经修饰的糖部分和/或经修饰的核酸键。

[0452] 方面42.根据方面1-41中任一方面所述的方法,其中所述接触在无细胞样品中进行。

[0453] 方面43.根据方面1-41中任一方面所述的方法,其中所述接触在体外、离体或体外细胞中进行。

[0454] 方面44.一种将前体C2c2指导RNA阵列切割成两个或更多个C2c2指导RNA的方法,所述方法包括:

[0455] 使包含两个或更多个C2c2指导RNA的前体C2c2指导RNA阵列与C2c2蛋白接触,每个所述C2c2指导RNA阵列具有不同的指导序列,其中所述C2c2蛋白将所述前体C2c2指导RNA阵列切割成单独的C2c2指导RNA。

[0456] 方面45.根据方面44所述的方法,其中所述C2c2蛋白缺乏催化活性的HEPN1结构域和/或缺乏催化活性的HEPN2结构域。

[0457] 方面46.根据方面44或方面45所述的方法,其中所述前体C2c2指导RNA阵列包含两个或更多个C2c2指导RNA,所述两个或更多个C2c2指导RNA靶向相同靶RNA分子内的不同靶序列。

[0458] 方面47.根据方面44-46中任一方面所述的方法,其中所述前体C2c2指导RNA阵列包含两个或更多个C2c2指导RNA,所述两个或更多个C2c2指导RNA靶向不同靶RNA分子。

[0459] 方面48.根据方面44-47中任一方面所述的方法,其中所述接触不在细胞内发生。

[0460] 方面49.根据方面44-48中任一方面所述的方法,其中所述指导RNA和/或所述前体C2c2指导RNA阵列中的至少一者是可检测地标记的。

[0461] 方面50.一种用于检测包含多种RNA的样品中的靶RNA的试剂盒,所述试剂盒包含:

[0462] (a) 前体C2c2指导RNA阵列,和/或编码所述前体C2c2指导RNA阵列的核酸,其中所述前体C2c2指导RNA阵列包含两个或更多个C2c2指导RNA,每个所述C2c2指导RNA具有不同的指导序列和/或所选择的指导序列的插入位点;以及

[0463] (b) C2c2蛋白。

[0464] 方面51.根据方面50所述的试剂盒,其中所述C2c2蛋白缺乏催化活性的HEPN1结构域和/或缺乏催化活性的HEPN2结构域。

[0465] 方面52.根据方面50或方面51所述的试剂盒,其中所述前体C2c2指导RNA阵列包含两个或更多个C2c2指导RNA,所述两个或更多个C2c2指导RNA靶向相同靶RNA分子内的不同靶序列。

[0466] 方面53.根据方面50-52中任一方面所述的试剂盒,其中所述前体C2c2指导RNA阵列包含两个或更多个C2c2指导RNA,所述两个或更多个C2c2指导RNA靶向不同靶RNA分子。

[0467] 方面54.根据方面50-53中任一方面所述的试剂盒,其中所述指导RNA和/或所述前体C2c2指导RNA阵列中的至少一者是可检测地标记的。

[0468] 方面55.根据方面50-54中任一方面所述的试剂盒,其还包含经标记的检测RNA,所述经标记的检测RNA包含荧光发射染料对(即,FRET对和/或猝灭剂/荧光剂对)。

[0469] 方面56.一种用于检测包含多种RNA的样品中的靶RNA的试剂盒,所述试剂盒包含:

[0470] (a) 经标记的检测RNA,其包含荧光发射染料对(即,FRET对和/或猝灭剂/荧光剂对);以及

[0471] (b) C2c2蛋白。

[0472] 方面57.根据方面56所述的试剂盒,其包含阳性对照靶RNA。

[0473] 方面58.根据方面57所述的试剂盒,其中所述阳性对照靶RNA以不同的量在两个或更多个容器中的每一个中存在。

[0474] 方面59.根据方面56-58中任一方面所述的试剂盒,其包含以下中的至少一者:

[0475] (c) C2c2指导RNA和/或编码所述C2c2指导RNA的核酸;

[0476] (d) 前体C2c2指导RNA和/或编码所述前体C2c2指导RNA的核酸;以及

[0477] (e) 前体C2c2指导RNA阵列,和/或编码所述前体C2c2指导RNA阵列的核酸,其中所述前体C2c2指导RNA阵列包含两个或更多个C2c2指导RNA,每个所述C2c2指导RNA具有不同的指导序列和/或所选择的指导序列的插入位点。

[0478] 方面60.根据方面56-59中任一方面所述的试剂盒,其包含DNA,所述DNA包含编码具有或不具有指导序列的C2c2指导RNA的核苷酸序列。

[0479] 方面61.根据方面60所述的试剂盒,其中所述DNA包含用于插入指导序列的插入序列。

[0480] 方面62.根据方面60或方面61所述的试剂盒,其中所述DNA是表达载体,并且所述C2c2指导RNA与启动子可操作地连接。

[0481] 方面63.根据方面62所述的试剂盒,其中所述启动子是T7启动子。

[0482] 方面64.根据方面56-63中任一方面所述的试剂盒,其包含缺乏核酸酶活性的C2c2内切核糖核酸酶变体。

[0483] 方面65.根据方面56-64中任一方面所述的试剂盒,其中所述经标记的检测RNA包含FRET对。

[0484] 方面66.根据方面56-65中任一方面所述的试剂盒,其中所述经标记的检测RNA包含猝灭剂/荧光剂对。

[0485] 方面67.根据方面56-66中任一方面所述的试剂盒,其中所述经标记的检测RNA包含产生第一可检测信号的FRET对和产生第二可检测信号的猝灭剂/荧光剂对。

[0486] 方面68.一种变体C2c2多肽,其包含:

[0487] a) 与SEQ ID NO:2所示氨基酸序列具有至少80%、至少85%、至少90%、至少95%、至少98%或至少99%的氨基酸序列同一性的氨基酸序列,并且包含对以下的取代:i) 氨基酸R472和H477;ii) 氨基酸R1048和H1053;或iii) 氨基酸R472、H477、R1048和H1053;

[0488] b) 与SEQ ID NO:1所示氨基酸序列具有至少80%、至少85%、至少90%、至少95%、至少98%或至少99%的氨基酸序列同一性的氨基酸序列,并且包含对以下的取代:i) 氨基酸R445和H450;ii) 氨基酸R1016和H1021;或iii) 氨基酸R445、H450、R1016和H1021;

[0489] c) 与SEQ ID NO:4所示氨基酸序列具有至少80%、至少85%、至少90%、至少95%、至少98%或至少99%的氨基酸序列同一性的氨基酸序列,并且包含对以下的取代:i) 氨基酸R464和H469;ii) 氨基酸R1052和H1057;或iii) 氨基酸R464、H469、R1052和H1057;

[0490] d) 与SEQ ID NO:5所示氨基酸序列具有至少80%、至少85%、至少90%、至少95%、至少98%或至少99%的氨基酸序列同一性的氨基酸序列,并且包含对以下的取代:i) 氨基酸R467和H472;ii) 氨基酸R1069和H1074;或iii) 氨基酸R467、H472、R1069和H1074;或者

[0491] e) 与SEQ ID NO:6所示氨基酸序列具有至少80%、至少85%、至少90%、至少95%、至少98%或至少99%的氨基酸序列同一性的氨基酸序列,并且包含对以下的取代:i) 氨基酸R472和H477;ii) 氨基酸R1044和H1049;iii) 或氨基酸R472、H477、R1044和H1049。

[0492] 方面69.根据68所述的变体C2c2多肽,其中所述变体C2c2多肽具有减少的或不可检测的ss RNA切割(例如,RNA指导的切割活性),但保留了结合C2c2指导RNA和ssRNA的能力,并且保留了切割前体C2c2指导RNA的能力。

[0493] 方面70.一种核酸,其包含编码根据方面68或方面69所述的变体C2c2多肽的核苷酸序列。

[0494] 方面71.根据方面70所述的核酸,其中所述核苷酸序列与组成型启动子或可调控的启动子可操作地连接。

[0495] 方面72.一种重组表达载体,其包含根据方面70或方面71所述的核酸。

[0496] 方面73.一种宿主细胞,用根据方面70或方面71所述的核酸或用根据方面72所述的重组表达载体遗传修饰。

[0497] 方面74.根据方面73所述的宿主细胞,其中所述宿主细胞是真核细胞。

[0498] 方面75.根据方面73所述的宿主细胞,其中所述宿主细胞是原核细胞。

[0499] 方面76.根据方面73-75中任一方面所述的宿主细胞,其中所述宿主细胞是体外、离体或体内的。

[0500] 方面77.一种检测包含多种RNA的样品中的至少两种不同单链靶RNA的方法,所述方法包括:

[0501] a) 使所述样品与以下接触:

[0502] (i) 第一C2c2蛋白,其切割包括至少一个A的单链RNA(ssRNA);

[0503] (ii) 第二C2c2蛋白,其切割包括至少一个U的ssRNA;

[0504] (iii) 第一C2c2指导RNA,其包含与所述第一单链靶RNA杂交的第一核苷酸序列和与所述第一C2c2蛋白结合的第二核苷酸序列;以及

[0505] (iv) 第二C2c2指导RNA,其包含与所述第二单链靶RNA杂交的第一核苷酸序列和与所述第二C2c2蛋白结合的第二核苷酸序列;

[0506] 其中所述第一C2c2蛋白不被所述第二C2c2指导RNA激活,并且其中所述第一C2c2蛋白切割包括至少一个A的ssRNA,并且

[0507] 其中所述第二C2c2蛋白不被所述第一C2c2指导RNA激活,并且其中所述第二C2c2蛋白切割包括至少一个U的ssRNA;以及

[0508] b) 测量由所述第一C2c2蛋白和所述第二C2c2蛋白介导的RNA切割产生的可检测信号,其中第一可检测信号在所述第一C2c2蛋白被激活时产生,并且第二可检测信号在所述第二C2c2蛋白被激活时产生,其中检测到所述第一信号表明所述第一靶ssRNA存在于所述样品中,并且其中检测到所述第二信号表明所述第二靶ssRNA存在于所述样品中。

[0509] 方面78.根据方面77所述的方法,其中:

[0510] a) 所述第一C2c2蛋白包含与图56F所示的Lba Cas13a氨基酸序列具有至少75%、至少80%、至少85%、至少90%、至少95%、至少98%、至少99%或100%的氨基酸序列同一性的氨基酸序列;并且所述第二C2c2蛋白包含与图56K所示的Hhe Cas13a氨基酸序列具有至少75%、至少80%、至少85%、至少90%、至少95%、至少98%、至少99%或100%的氨基酸序列同一性的氨基酸序列;

[0511] b) 所述第一C2c2蛋白包含与图56F所示的Lba Cas13a氨基酸序列具有至少75%、至少80%、至少85%、至少90%、至少95%、至少98%、至少99%或100%的氨基酸序列同一性的氨基酸序列;并且所述第二C2c2蛋白包含与图56G所示的Rca Cas13a氨基酸序列具有至少75%、至少80%、至少85%、至少90%、至少95%、至少98%、至少99%或100%的氨基酸序列同一性的氨基酸序列;

[0512] c) 所述第一C2c2蛋白包含与图56F所示的Lba Cas13a氨基酸序列具有至少75%、至少80%、至少85%、至少90%、至少95%、至少98%、至少99%或100%的氨基酸序列同一性的氨基酸序列;并且所述第二C2c2蛋白包含与图56B所示的Ppr Cas13a氨基酸序列具有至少75%、至少80%、至少85%、至少90%、至少95%、至少98%、至少99%或100%的氨基酸序列同一性的氨基酸序列;

[0513] d) 所述第一C2c2蛋白包含与图56F所示的Lba Cas13a氨基酸序列具有至少75%、至少80%、至少85%、至少90%、至少95%、至少98%、至少99%或100%的氨基酸序列同一性的氨基酸序列;并且所述第二C2c2蛋白包含与图56I所示的Lne Cas13a氨基酸序列具有至少75%、至少80%、至少85%、至少90%、至少95%、至少98%、至少99%或100%的氨基酸序列同一性的氨基酸序列;

[0514] e) 所述第一C2c2蛋白包含与图56F所示的Lba Cas13a氨基酸序列具有至少75%、至少80%、至少85%、至少90%、至少95%、至少98%、至少99%或100%的氨基酸序列同一性的氨基酸序列；并且所述第二C2c2蛋白包含与图56C所示的Lbu Cas13a氨基酸序列具有至少75%、至少80%、至少85%、至少90%、至少95%、至少98%、至少99%或100%的氨基酸序列同一性的氨基酸序列；

[0515] f) 所述第一C2c2蛋白包含与图56F所示的Lba Cas13a氨基酸序列具有至少75%、至少80%、至少85%、至少90%、至少95%、至少98%、至少99%或100%的氨基酸序列同一性的氨基酸序列；并且所述第二C2c2蛋白包含与图56E所示的Lwa Cas13a氨基酸序列具有至少75%、至少80%、至少85%、至少90%、至少95%、至少98%、至少99%或100%的氨基酸序列同一性的氨基酸序列；

[0516] g) 所述第一C2c2蛋白包含与图56F所示的Lba Cas13a氨基酸序列具有至少75%、至少80%、至少85%、至少90%、至少95%、至少98%、至少99%或100%的氨基酸序列同一性的氨基酸序列；并且所述第二C2c2蛋白包含与图56D所示的Lsh Cas13a氨基酸序列具有至少75%、至少80%、至少85%、至少90%、至少95%、至少98%、至少99%或100%的氨基酸序列同一性的氨基酸序列；

[0517] h) 所述第一C2c2蛋白包含与图56J所示的Ere Cas13a氨基酸序列具有至少75%、至少80%、至少85%、至少90%、至少95%、至少98%、至少99%或100%的氨基酸序列同一性的氨基酸序列；并且所述第二C2c2蛋白包含与图56K所示的Hhe Cas13a氨基酸序列具有至少75%、至少80%、至少85%、至少90%、至少95%、至少98%、至少99%或100%的氨基酸序列同一性的氨基酸序列；

[0518] i) 所述第一C2c2蛋白包含与图56J所示的Ere Cas13a氨基酸序列具有至少75%、至少80%、至少85%、至少90%、至少95%、至少98%、至少99%或100%的氨基酸序列同一性的氨基酸序列；并且所述第二C2c2蛋白包含与图56G所示的Rca Cas13a氨基酸序列具有至少75%、至少80%、至少85%、至少90%、至少95%、至少98%、至少99%或100%的氨基酸序列同一性的氨基酸序列；

[0519] j) 所述第一C2c2蛋白包含与图56J所示的Ere Cas13a氨基酸序列具有至少75%、至少80%、至少85%、至少90%、至少95%、至少98%、至少99%或100%的氨基酸序列同一性的氨基酸序列；并且所述第二C2c2蛋白包含与图56B所示的Ppr Cas13a氨基酸序列具有至少75%、至少80%、至少85%、至少90%、至少95%、至少98%、至少99%或100%的氨基酸序列同一性的氨基酸序列；

[0520] k) 所述第一C2c2蛋白包含与图56J所示的Ere Cas13a氨基酸序列具有至少75%、至少80%、至少85%、至少90%、至少95%、至少98%、至少99%或100%的氨基酸序列同一性的氨基酸序列；并且所述第二C2c2蛋白包含与图56I所示的Lne Cas13a氨基酸序列具有至少75%、至少80%、至少85%、至少90%、至少95%、至少98%、至少99%或100%的氨基酸

序列同一性的氨基酸序列；

[0521] 1) 所述第一C2c2蛋白包含与图56J所示的Ere Cas13a氨基酸序列具有至少75%、至少80%、至少85%、至少90%、至少95%、至少98%、至少99%或100%的氨基酸序列同一性的氨基酸序列；并且所述第二C2c2蛋白包含与图56C所示的Lbu Cas13a氨基酸序列具有至少75%、至少80%、至少85%、至少90%、至少95%、至少98%、至少99%或100%的氨基酸序列同一性的氨基酸序列；

[0522] m) 所述第一C2c2蛋白包含与图56J所示的Ere Cas13a氨基酸序列具有至少75%、至少80%、至少85%、至少90%、至少95%、至少98%、至少99%或100%的氨基酸序列同一性的氨基酸序列；并且所述第二C2c2蛋白包含与图56E所示的Lwa Cas13a氨基酸序列具有至少75%、至少80%、至少85%、至少90%、至少95%、至少98%、至少99%或100%的氨基酸序列同一性的氨基酸序列；

[0523] n) 所述第一C2c2蛋白包含与图56J所示的Ere Cas13a氨基酸序列具有至少75%、至少80%、至少85%、至少90%、至少95%、至少98%、至少99%或100%的氨基酸序列同一性的氨基酸序列；并且所述第二C2c2蛋白包含与图56D所示的Lsh Cas13a氨基酸序列具有至少75%、至少80%、至少85%、至少90%、至少95%、至少98%、至少99%或100%的氨基酸序列同一性的氨基酸序列；

[0524] o) 所述第一C2c2蛋白包含与图56J所示的Ere Cas13a氨基酸序列具有至少75%、至少80%、至少85%、至少90%、至少95%、至少98%、至少99%或100%的氨基酸序列同一性的氨基酸序列；并且所述第二C2c2蛋白包含与图56A所示的Lse Cas13a氨基酸序列具有至少75%、至少80%、至少85%、至少90%、至少95%、至少98%、至少99%或100%的氨基酸序列同一性的氨基酸序列；

[0525] p) 所述第一C2c2蛋白包含与图56H所示的Cam Cas13a氨基酸序列具有至少75%、至少80%、至少85%、至少90%、至少95%、至少98%、至少99%或100%的氨基酸序列同一性的氨基酸序列；并且所述第二C2c2蛋白包含与图56K所示的Hhe Cas13a氨基酸序列具有至少75%、至少80%、至少85%、至少90%、至少95%、至少98%、至少99%或100%的氨基酸序列同一性的氨基酸序列；

[0526] q) 所述第一C2c2蛋白包含与图56H所示的Cam Cas13a氨基酸序列具有至少75%、至少80%、至少85%、至少90%、至少95%、至少98%、至少99%或100%的氨基酸序列同一性的氨基酸序列；并且所述第二C2c2蛋白包含与图56G所示的Rca Cas13a氨基酸序列具有至少75%、至少80%、至少85%、至少90%、至少95%、至少98%、至少99%或100%的氨基酸序列同一性的氨基酸序列；

[0527] r) 所述第一C2c2蛋白包含与图56H所示的Cam Cas13a氨基酸序列具有至少75%、至少80%、至少85%、至少90%、至少95%、至少98%、至少99%或100%的氨基酸序列同一性的氨基酸序列；并且所述第二C2c2蛋白包含与图56B所示的Ppr Cas13a氨基酸序列具有至少75%、至少80%、至少85%、至少90%、至少95%、至少98%、至少99%或100%的氨基酸序列同一性的氨基酸序列；

[0528] s) 所述第一C2c2蛋白包含与图56H所示的Cam Cas13a氨基酸序列具有至少75%、至少80%、至少85%、至少90%、至少95%、至少98%、至少99%或100%的氨基酸序列同一性的氨基酸序列；并且所述第二C2c2蛋白包含与图56I所示的Lne Cas13a氨基酸序列具有

至少75%、至少80%、至少85%、至少90%、至少95%、至少98%、至少99%或100%的氨基酸序列同一性的氨基酸序列；

[0529] t) 所述第一C2c2蛋白包含与图56H所示的Cam Cas13a氨基酸序列具有至少75%、至少80%、至少85%、至少90%、至少95%、至少98%、至少99%或100%的氨基酸序列同一性的氨基酸序列；并且所述第二C2c2蛋白包含与图56C所示的Lbu Cas13a氨基酸序列具有至少75%、至少80%、至少85%、至少90%、至少95%、至少98%、至少99%或100%的氨基酸序列同一性的氨基酸序列；

[0530] u) 所述第一C2c2蛋白包含与图56H所示的Cam Cas13a氨基酸序列具有至少75%、至少80%、至少85%、至少90%、至少95%、至少98%、至少99%或100%的氨基酸序列同一性的氨基酸序列；并且所述第二C2c2蛋白包含与图56E所示的Lwa Cas13a氨基酸序列具有至少75%、至少80%、至少85%、至少90%、至少95%、至少98%、至少99%或100%的氨基酸序列同一性的氨基酸序列；

[0531] v) 所述第一C2c2蛋白包含与图56H所示的Cam Cas13a氨基酸序列具有至少75%、至少80%、至少85%、至少90%、至少95%、至少98%、至少99%或100%的氨基酸序列同一性的氨基酸序列；并且所述第二C2c2蛋白包含与图56F所示的Lsh Cas13a氨基酸序列具有至少75%、至少80%、至少85%、至少90%、至少95%、至少98%、至少99%或100%的氨基酸序列同一性的氨基酸序列；或者

[0532] w) 所述第一C2c2蛋白包含与图56H所示的Cam Cas13a氨基酸序列具有至少75%、至少80%、至少85%、至少90%、至少95%、至少98%、至少99%或100%的氨基酸序列同一性的氨基酸序列；并且所述第二C2c2蛋白包含与图56A所示的Lse Cas13a氨基酸序列具有至少75%、至少80%、至少85%、至少90%、至少95%、至少98%、至少99%或100%的氨基酸序列同一性的氨基酸序列。

[0533] 方面79. 根据方面77或方面78所述的方法，其中所述方法包括使所述样品与以下接触：

[0534] i) 第一经标记的检测RNA，其包含第一荧光共振能量转移 (FRET) 对和/或第一猝灭剂/荧光剂对，其中所述第一经标记的检测RNA包含至少一个A并且不包含U；和

[0535] ii) 第二经标记的检测RNA，其包含第二FRET对和/或第二猝灭剂/荧光剂对，其中所述第二经标记的检测RNA包含至少一个U并且不包含A，

[0536] 其中所述第一C2c2蛋白切割所述第一经标记的检测RNA，并且所述第一可检测信号由所述第一FRET对和/或所述第一猝灭剂/荧光剂对产生，并且

[0537] 其中所述第一C2c2蛋白切割所述第二经标记的检测RNA，并且所述第二可检测信号由所述第二FRET对和/或所述第二猝灭剂/荧光剂对产生。

[0538] 方面80. 根据方面79所述的方法，其中所述第一经标记的检测RNA包含2至15个连续的A的段和/或所述第二经标记的检测RNA包含2至15个连续的U的段。

[0539] 方面81. 根据方面79所述的方法，其中所述第一经标记的检测RNA包含4至15个连续的A的段和/或所述第二经标记的检测RNA包含4至15个连续的U的段。

[0540] 方面82. 根据方面79所述的方法，其中所述第一经标记的检测RNA包含至少3个连续的A的段和/或所述第二经标记的检测RNA包含至少3个连续的U的段。

[0541] 方面83. 根据方面79所述的方法，其中所述第一经标记的检测RNA包含至少4个连

续的A的段和/或所述第二经标记的检测RNA包含至少4个连续的U的段。

[0542] 方面84. 一种试剂盒,其包含:

[0543] (a) 第一经标记的检测RNA,其缺乏U并且包含至少一个A且包含第一荧光发射染料对;

[0544] (b) 第二经标记的检测RNA,其缺乏A并且包含至少一个U且包含第二荧光发射染料对;

[0545] (c) 第一C2c2蛋白,和/或编码所述第一C2c2蛋白的核酸,其中所述第一C2c2蛋白可切割所述第一经标记的检测RNA但不切割所述第二经标记的检测RNA(例如,所述第一C2c2蛋白当被激活时切割腺嘌呤+RNA并且不切割缺乏A的RNA);以及

[0546] (d) 第二C2c2蛋白,和/或编码所述第二C2c2蛋白的核酸,其中所述第二C2c2蛋白可切割所述第二经标记的检测RNA但不切割所述第一经标记的检测RNA(例如,所述第二C2c2蛋白当被激活时切割尿嘧啶+RNA但不切割缺乏U的RNA)。

[0547] 方面85. 根据方面84所述的试剂盒,其包含以下中的至少一者:

[0548] (e) 第一C2c2指导RNA和/或编码所述第一C2c2指导RNA的核酸,其中所述第一C2c2指导RNA包含与所述第一C2c2蛋白结合的恒定区序列;

[0549] (f) 第二C2c2指导RNA和/或编码所述第二C2c2指导RNA的核酸,其中所述第二C2c2指导RNA包含与所述第二C2c2蛋白结合的恒定区序列;

[0550] (g) 核酸,其包含编码与所述第一C2c2蛋白结合的恒定区序列的核苷酸序列和所选择的指导序列的插入位点;

[0551] (h) 核酸,其包含编码与所述第二C2c2蛋白结合的恒定区序列的核苷酸序列和所选择的指导序列的插入位点。

[0552] 方面86. 根据方面84所述的试剂盒,其包含核酸,所述核酸包含编码第一C2c2指导RNA的核苷酸序列,其中所述第一C2c2指导RNA包含与所述第一C2c2蛋白结合的恒定区序列。

[0553] 方面87. 根据方面84所述的试剂盒,其包含核酸,所述核酸包含编码第二C2c2指导RNA的核苷酸序列,其中所述第二C2c2指导RNA包含与所述第二C2c2蛋白结合的恒定区序列。

[0554] 方面88. 根据方面84所述的试剂盒,其包含核酸,所述核酸包含编码与所述第一C2c2蛋白结合的恒定区序列的核苷酸序列和所选择的指导序列的插入位点。

[0555] 方面89. 根据方面84所述的试剂盒,其包含核酸,所述核酸包含编码与所述第二C2c2蛋白结合的恒定区序列的核苷酸序列和所选择的指导序列的插入位点。

[0556] 方面90. 根据方面86-89中任一方面所述的试剂盒,其中所述核酸是表达载体,并且所述核苷酸序列与启动子可操作地连接。

[0557] 实施例

[0558] 提出以下实施例是为了向本领域的普通技术人员提供如何制备和使用本发明的完整公开和描述,而非旨在限制本发明人认为是其发明的范围,也非旨在表示以下实验是所执行的全部或唯一实验。已经努力确保关于所使用的数值(例如量、温度等)的准确性,但是应该考虑一些实验误差和偏差。除非另有说明,否则份数是重量份,分子量是重均分子量,温度以摄氏度计,并且压力处于大气压或接近大气压。可以使用标准缩写,例如bp表示

碱基对;kb表示千碱基;pI表示皮升;s或sec表示秒;min表示分钟;h或hr表示小时;aa表示氨基酸;kb表示千碱基;bp表示碱基对;nt表示核苷酸;i.m.表示肌肉(肌内地);i.p.表示腹膜内(腹膜内地);s.c.表示皮下(皮下地);等等。

[0559] 本文描述的数据显示,C2c2具有负责CRISPR RNA加工和ssRNA降解的两种不同核糖核酸酶活性。前体CRISPR RNA(pre-crRNA)的成熟是相对缓慢的、高度特异性的催化事件。相比之下,当与具有与crRNA的指导区段的序列互补性的靶RNA结合时,C2c2被激活为强大的通用RNA酶,所述RNA酶通过引发尿嘧啶核苷酸处的链断裂而顺式和反式地切割RNA。数据显示,这种反式切割活性可用于复杂混合物内的超灵敏RNA检测。

[0560] 实施例1:C2c2重组蛋白纯化

[0561] 图1A至图1E.对内源C2c2基因座以及重组C2c2蛋白的异源表达和纯化的概述。图1A.口腔纤毛菌C2c2基因座的示意图。所预测的HEPN活性位点以黄色指示,其中活性位点残基被标示。图1B.来自该VI型CRISPR系统的成熟crRNA的示意图。图1C.对阳离子交换色谱级分的十二烷基硫酸钠-聚丙烯酰胺凝胶电泳(SDS-PAGE)分析。收获表达His-麦芽糖结合蛋白(MBP)-C2c2的大肠杆菌细胞并通过超声处理进行裂解。通过离心澄清裂解物,然后经由金属离子亲和色谱分离His-MBP-C2c2。将蛋白洗脱液与TEV蛋白酶一起孵育以切除His-MBP标签。将切割的蛋白上样到HiTrap肝素柱上,并以线性KCl梯度洗脱。将洗脱级分上样到10%SDS-PAGE凝胶上进行分析。图1D.将阳离子交换色谱级分汇集、浓缩,然后上样到S200尺寸排阻色谱柱上。经由SDS-PAGE分析来自尺寸交换色谱的级分。图1E.C2c2酶促活性的工作模型。

[0562] 实施例2:C2c2是可编程的内切核糖核酸酶

[0563] 纯化的C2c2以crRNA指导的方式切割单链RNA(ssRNA),所述crRNA指导的方式依赖于催化活性HEPN结构域的存在(图2)

[0564] 图2.在20mM Tris-HCl pH 6.8、50mM KCl、5mM MgCl₂、5%甘油中执行切割测定。将100nM C2c2蛋白与crRNA以2:1的比率在37℃下预孵育10分钟以促进复合物组装。约1nM的5' ³²P标记的ssRNA的加入引发反应,并且通过用甲酰胺上样缓冲液猝灭来获取时间点。将切割产物形成通过15%变性尿素-PAGE进行解析,并使用磷光成像仪可视化。灭活蛋白(由dC2c2表示法表示)是四点突变体(R472A、H477A、R1048A、H1053A),其在两个预测的活性位点中均从R-X₄₋₆-H HEPN基序消除精氨酸和组氨酸两者。在这种凝胶中,所关注的泳道用红色框突出显示。使用脱靶ssRNA时或当HEPN结构域失活时未检测到切割活性。

[0565] 实施例3:C2c2切割活性对单链RNA有特异性

[0566] C2c2表现为不切割双链RNA或DNA,其中与ssRNA相比,以大大降低的速率有限地切割单链DNA(图3)。

[0567] 图3.C2c2强烈切割单链RNA并且不切割双链底物。切割测定条件如图2所示。将切割产物在15%变性尿素-PAGE凝胶上解析。在5分钟、10分钟、30分钟、75分钟时取样,其中所有对照泳道平行孵育75分钟。具有预期尺寸范围的切割产物可仅在ssRNA反应中检测到,并且在单链DNA条件下以更小程度检测到。对单链RNA靶的3'末端标记表明,切割发生在靶-间隔区序列杂交区外部。图中顶部的标签指示是添加了蛋白("Lbu C2c2")还是C2c2指导RNA("crGAPDH1")。对照包括(i)蛋白但没有指导RNA,(ii)指导RNA但没有蛋白,以及(iii)没有蛋白和没有指导RNA。

[0568] 实施例4:C2c2将前体crRNA加工成成熟crRNA

[0569] C2c2以不依赖于HEPN结构域的方式将前体crRNA加工成成熟crRNA(图4A至图4B),表明存在另外的内切核酸酶结构域。

[0570] 图4A.由C2c2催化的pre-crRNA加工反应的图。图4B.在用甲酰胺上样缓冲液进行猝灭之前,将100nM Lbu C2c2与约5nM 5' ³²P标记的ssRNA一起在20mM的HEPES pH 6.8、50mM KCl、5mM MgCl₂、5%甘油中孵育60分钟。将反应产物在15%变性尿素-PAGE凝胶上解析,并使用磷光成像仪(phosphoimager)可视化。灭活蛋白(由dC2c2表示法表示)是四点突变体(R472A、H477A、R1048A、H1053A),其在两个预测的活性位点中均从R-X₄₋₆-H HEPN基序消除精氨酸和组氨酸两者。测试具有可变间隔区序列的三种不同pre-crRNA的C2c2加工能力。

[0571] 实施例5:复杂混合物中的转录物的灵敏检测

[0572] 使用C2c2检测复杂混合物中的靶转录物(图5)。

[0573] 图5.在存在不断增加量的λ2ssRNA(0-10nM)的情况下,将50nM的靶向C2c2:crRNA的λ2ssRNA与185nM的RNA酶-Alert底物(小型荧光-猝灭剂RNA寡核苷酸-用猝灭剂/荧光剂对标记的经标记的检测RNA)和100ng的HEK293T总RNA一起在37C下孵育30分钟。当‘经激活的’C2c2(C2c2:crRNA:λ2ssRNA)切割RNA酶-Alert底物而从猝灭剂部分释放荧光染料时,观察到荧光增加。

[0574] 实施例6:由C2c2蛋白进行前体crRNA(pre-crRNA)加工

[0575] C2c2切割(加工)前体crRNA(pre-crRNA),并且该切割不依赖于HEPN结构域(图6)。

[0576] 图6.将300nM的C2c2与1nM的5' -³²P标记的pre-crRNA一起在37C下孵育0-60分钟。将RNA产物在15%尿素-PAGE上分离,并使用磷光成像仪成像。通过较小的‘经切割的RNA产物’的存在来确定pre-crRNA的成功加工。

[0577] 实施例7

[0578] 材料和方法

[0579] 使用以下材料和方法,并且以下材料和方法适用于实施例8-12。

[0580] C2c2系统发生和候选选择。使用具有PROTGAMMALG进化模型和100个自展抽样(bootstrap sampling)的RAxML计算C2c2最大似然系统发生。通过MAFFT使用‘einsi’方法比对序列。Niewoehner, O.和Jinek, M., Structural basis for the endoribonuclease activity of the type III-A CRISPR-associated protein Csm6. RNA 22, 318-329 (2016)。选择候选同系物来对蛋白家族的主要分支进行取样。

[0581] C2c2蛋白的生产和纯化。使用从Integrated DNA Technologies订购的合成gBlock组装用于蛋白纯化的表达载体。将密码子优化的C2c2基因组序列用His₆-MBP-TEV切割位点进行N末端标记,其中表达由T7启动子驱动(根据要求可用的映射(map available upon request))。经由野生型C2c2构建体的圆角化(round-the-horn)、定点诱变来克隆突变蛋白。将表达载体转化到于37°C下在2xYT肉汤中生长的Rosetta2大肠杆菌细胞中。将大肠杆菌细胞在对数期期间用0.5M的异丙基β-D-1-硫代半乳糖苷(ITPG)诱导,并将温度降至16°C以过夜表达His-MBP-C2c2。随后收获细胞,重悬于裂解缓冲液(50mM Tris-HCl pH 7.0、500mM NaCl、5%甘油、1mM三(2-羧乙基)膦(TCEP)、0.5mM苯甲基磺酰氟(PMSF)以及不含乙二胺四乙酸(EDTA)的蛋白酶抑制剂(Roche)中,并通过超声裂解,并且将裂解物通过离心澄清。通过金属离子亲和色谱分离出可溶性His-MBP-C2c2,将含蛋白的洗脱物与烟草蚀

纹病毒 (TEV) 蛋白酶一起在4℃下孵育过夜,同时透析到离子交换缓冲液 (50mM Tris-HCl pH 7.0、250mM KCl、5% 甘油、1mM TCEP) 中以切除His₆-MBP标签。将切割的蛋白上样到 HiTrap SP柱上,并以线性KCl (0.25-1.5M) 梯度洗脱。将阳离子交换色谱级分汇集,并用 30kD的截留浓缩管 (Thermo Fisher) 浓缩。将C2c2蛋白经由在S200柱上进行尺寸排阻色谱而进一步纯化,并储存在凝胶过滤缓冲液 (20mM Tris-HCl pH 7.0、200mM KCl、5% 甘油、1mM TCEP) 中以供后续酶测定。

[0582] 生成ssRNA。除了crRNA AES461是订购合成的 (Integrated DNA Technologies) 之外,本研究中使用的RNA都是在体外转录的[参见图15和图19]。如前所述进行体外转录反应,具有以下改变:T7聚合酶浓度降至10μg/mL,并且UTP浓度降至2.5mM。将转录在37℃下孵育1-2小时以减少核苷酸的非模板加入。使用15%变性PAGE凝胶纯化所有转录反应。将所有RNA重悬于切割缓冲液 (20mM HEPES pH 6.8、50mM KCl、5mM MgCl₂和5% 甘油) 中。对于放射性实验,在放射性标记之前用小牛肠磷酸盐 (New England Biolabs) 去除5'三磷酸酯,然后如前所述使用T4多核苷酸激酶 (New England Biolabs) 和[γ-³²P]-ATP (Perkin Elmer) 对ssRNA底物进行5'末端标记。Sternberg, S.H.、Haurwitz, R.E. 和Doudna, J.A. Mechanism of substrate selection by a highly specific CRISPR endoribonuclease. RNA 18, 661-672 (2012)。

[0583] 在该研究中使用的DNA底物列于图14的表2中。

[0584] 在该研究中使用的RNA底物列于图15的表3中。

[0585] pre-crRNA加工测定。在37℃下在加工缓冲液 (20mM HEPES pH 6.8、50mM KCl、5mM MgCl₂、10μg/mL牛血清白蛋白 (BSA)、10μg/mL tRNA、0.05% Igepal CA-630和5% 甘油) 以及相对于5'-标记的pre-crRNA 100倍摩尔过量的C2c2 (最终浓度分别为100nM和大于1nM) 中执行Pre-crRNA切割测定。除非另有说明,否则在60分钟后用1.5X RNA上样染料 (100% 甲酰胺、0.025w/v% 溴酚蓝和200μg mL⁻¹ 肝素) 猝灭反应。猝灭后,将反应在95℃变性5分钟,然后通过12%或15%变性PAGE (0.5X TBE缓冲液) 解析。通过向反应缓冲液中加入浓度范围为10-100mM的EDTA或EGTA来测试反应的金属依赖性。将条带通过磷光成像可视化,并用 ImageQuant (GE Healthcare) 定量。切割百分比被确定为产物条带强度与产物条带和未切割的pre-crRNA条带两者的总强度的比率,并使用ImageQuant TL软件 (GE Healthcare) 针对各测量的底物内的背景进行归一化,并使用Prism (GraphPad) 与单相指数关联进行拟合。

[0586] 产物尺寸映射。通过使用得自Ambion的RNA酶T1试剂盒比较产物条带到碱水解梯带和RNA酶T1消化梯带的凝胶迁移,而生物化学地确定切割产物的长度。对于水解梯带,将15nM的全长RNA底物在95℃下于1X碱水解缓冲液 (Ambion) 中孵育5分钟。用1.5X RNA上样缓冲液猝灭反应,并冷却至-20℃以立即停止水解。对于RNA酶T1梯带,将15nM的全长RNA底物在65℃下在1X RNA测序缓冲液 (Ambion) 中解折叠。将反应冷却至环境温度,然后向反应中加入1μl的RNA酶T1。15分钟后,通过苯酚-氯仿提取终止反应,并加入1.5X RNA上样缓冲液进行储存。将水解条带与切割样品平行地在15%变性PAGE上解析,并通过磷光成像可视化。

[0587] 靶切割测定。在25℃和37℃下在切割缓冲液 (20mM HEPES pH6.8、50mM KCl、5mM MgCl₂和5% 甘油) 中执行靶切割测定。通过加热至65℃达5分钟,以及然后在切割缓冲液中缓慢冷却至环境温度而将crRNA指导物预折叠。在加入5'末端标记的靶和/或其它非放射性标记的RNA靶底物之前,通常以2:1的蛋白与crRNA摩尔比在切割缓冲液中于37℃下执行RNP

复合物形成达10分钟。除非另有说明,否则对于所有反应,蛋白、指导物和靶的最终浓度分别为100nM、50nM和大于1nM。用1.5X RNA上样染料猝灭反应,并通过15%变性PAGE (0.5X TBE缓冲液)解析。将条带通过磷光成像可视化,并用ImageQuant (GE Healthcare) 定量。切割百分比被确定为所有较短产物的总条带化强度相对于未切割条带的比率,并使用ImageQuant TL软件 (GE Healthcare) 针对各测量的底物内的背景进行归一化,并使用Prism (GraphPad) 与单相指数关联进行拟合。

[0588] crRNA过滤结合测定。在RNA加工缓冲液 (20mM HEPES pH6.8、50mM KCl、5mM MgCl₂、10μg/mL BSA、10μg/mL酵母tRNA、0.01% Igepal CA-630和5%甘油) 中进行过滤结合测定。将LbuC2c2与放射性标记的crRNA (小于0.1nM) 一起在37℃下孵育1小时。将Tufryn、Protran和Hybond-N+以上面列出的顺序组装到点印迹设备上。将膜用50μL的平衡缓冲液 (20mM HEPES pH 6.8、50mM KCl、5mM MgCl₂和5%甘油) 洗涤两次,然后将样品施加到所述膜上。将膜再次用50μL的平衡缓冲液洗涤,干燥并通过磷光成像可视化。将数据用ImageQuant TL软件 (GE Healthcare) 定量,并使用Prism (GraphPad Software) 与结合等温线拟合。所有实验一式三份地进行。图例中报告了解离常数和相关误差。

[0589] 电泳迁移率变动测定。为了避免在ssRNA结合实验期间解离低浓度的LbuC2c2-dHEPN1/dHEPN2:crRNA复合物,结合反应含有恒定过量的LbuC2c2-dHEPN1/dHEPN2 (200nM) 以及不断增加浓度的crRNA-A和小于0.1nM的靶ssRNA。在C2c2电泳迁移率变动测定 (EMSA) 缓冲液 (20mM HEPES pH 6.8、50mM KCl、10μg/mL BSA、100μg/mL酵母tRNA、0.01% Igepal CA-630和5%甘油) 中进行测定。如上所述在37℃下预形成LbuC2c2-crRNA-A复合物10分钟,然后加入5'-放射性标记的ssRNA底物,并在37℃下进一步孵育45分钟。然后将样品通过4℃下的8%天然PAGE (0.5X TBE缓冲液) 进行解析。将凝胶通过磷光成像而成像,使用ImageQuant TL软件 (GE Healthcare) 定量,并使用Prism (GraphPad Software) 与结合等温线拟合。所有实验一式三份地进行。图例中报告了解离常数和相关误差。

[0590] 荧光RNA检测测定。通过将1μM的Lbu-C2c2:C2c2与500nM的crRNA在37℃下孵育而预组装LbuC2c2:crRNA复合物。然后在存在185nM的RNA酶-Alert底物 (Thermo-Fisher)、100ng的HeLa总RNA和不断增加量的ssRNA (0-1nM) 的情况下,将这些复合物稀释至在RNA加工缓冲液 (20mM HEPES pH 6.8、50mM KCl、5mM MgCl₂、10μg/mL BSA、10μg/mL酵母tRNA、0.01% Igepal CA-630和5%甘油) 中100nM的LbuC2c2:50nM的crRNA-λ₂。将这些反应在荧光平板读数器中于37℃下孵育长达120分钟,其中每5分钟进行一次荧光测量 (λ_{ex}:485nm;λ_{em}:535nm)。通过减去在不存在靶ssRNA的情况下进行的反应获得的荧光值来获得经背景校正的荧光值。通过将50nM的RNA酶A与185nM的RNA酶-Alert底物一起孵育来测量最大荧光。对于耦合的pre-crRNA加工和RNA检测测定,通过将1μM的Lbu-C2c2:C2c2与500nM的pre-crRNA-A-λ₂一起在37℃下孵育20分钟而预组装LbuCas9-sgRNA复合物,并在存在不断增加量的ssRNA A和ssRNAλ₂ (各自0-1nM) 的情况下如上所述地进行反应。在每种情况下,误差条代表三次独立实验的标准偏差。

[0591] 实施例8:C2c2家族加工前体crRNA转录物以产生成熟crRNA

[0592] VI型CRISPR基因座缺乏明显的Cas6或Cas5d样内切核酸酶或tracrRNA。问题是C2c2本身是否具有pre-crRNA加工活性。为了测试此点,将来自口腔纤毛菌 (Lbu)、沙氏纤毛菌 (Lsh) 和斯氏李斯特菌 (Lse) 的重组C2c2同系物 (所述重组C2c2同系物来源于C2c2蛋白家

族的三个不同分支)在大肠杆菌中表达和纯化。图9A至图9B。

[0593] 所有三种蛋白在60分钟内切割经5'末端放射性标记的pre-crRNA底物,所述底物由全长共有重复序列和20个核苷酸(nt)的间隔区序列组成。图9C.使用T1 RNA酶梯带和氢氧化物水解梯带映射所有pre-crRNA同系物的切割位点;结果表明,加工取决于同系物而发生在所预测的发夹结构上游2nt或5nt的位置处。图9C。

[0594] 细胞因子可能进一步加工crRNA,因为来自大肠杆菌中LseC2c2操纵子和LshC2c2操纵子的异源表达的测序数据发现是相似但不相同的加工事件。用LbuC2c2和含有串联发夹-重复序列阵列的pre-crRNA进行切割测定产生了两种对应于两种连续切割事件的产物(图16A至图16F),这与C2c2在加工初级crRNA转录物中的作用相符。这些结果表明,在VI型CRISPR系统中,C2c2同系物催化其相关crRNA的成熟。

[0595] 图9A至图9C.C2c2家族加工前体crRNA转录物以产生成熟crRNA。图9A.C2c2蛋白的最大似然系统发生树。图22A中提供了包括种质、生物名称和自展值的完整细节。本研究使用的同系物以黄色突出显示。图9B,三种不同的VI型CRISPR基因座的图。黑色矩形表示重复元件,而黄色菱形表示间隔区序列。Cas1和Cas2仅存在于Lsh C2c2附近的基因组中。图9C,CC2c2介导的对来源于LbuC2c2、LseC2c2和LshC2c2的pre-crRNA的切割。OH:碱水解梯带;T1:T1RNA酶水解梯带;用100nM的C2c2和大于1nM的pre-crRNA执行加工切割反应。右侧描绘了切割的示意图,并且下方示出了所预测的pre-crRNA二级结构,其中箭头指示了所映射的C2c2切割位点。

[0596] 图22A至图22B.C2c2家族的完整系统发生树以及C2c2比对。图22A,C2c2蛋白的最大似然系统发生重建。叶包括GI蛋白数和来源生物;100个重新取样中的自展支持值被呈现用于内部分割。标尺是每个位点的取代。图22B,三种所分析的C2c2的同系物的多重序列比对,坐标基于LbuC2c2。

[0597] 图23A至图23D.C2c2的纯化和生产。所有C2c2同系物在大肠杆菌(E.coli)中表达为His-MBP融合物,并通过亲和性、离子交换和尺寸排阻色谱的组合进行纯化。通过与TEV蛋白酶一起孵育去除Ni⁺亲和和标签。色谱级分的代表性SDS-PAGE凝胶如上文在(图23A至图23B)中所示。图23C,来自Superdex 200(16/60)柱的色谱图表明,C2c2作为单峰洗脱、缺乏核酸。图23D,本稿图中使用的所有纯化蛋白的SDS PAGE分析。

[0598] 图16A至图16F.由C2c2进行的pre-crRNA加工是不依赖于间隔区序列的,可以在串联crRNA阵列上发生,受pre-crRNA的5'侧翼区域中的突变影响并且是不依赖于金属的。a,pre-crRNA切割测定的代表性时间过程表明LbuC2c2、LshC2c2和LseC2c2进行的pre-crRNA加工具有相似速率。b,单个同源pre-crRNA阵列的LseC2c2切割和LshCC2c2切割的切割位点映射。OH:碱水解梯带;T1:T1 RNA酶水解梯带。用100nM的Lbu C2c2和小于1nM的pre-crRNA进行切割反应。描绘了切割产物的示意图,其中箭头指示所映射的C2c2切割产物。c,串联pre-crRNA阵列的LbuCC2c2切割的切割位点映射。OH:碱水解梯带;T1:T1 RNA酶水解梯带。用100nM的Lbu C2c2和小于1nM的pre-crRNA进行切割反应。右侧描绘了切割产物的示意图,其中箭头指示映射的C2c2切割产物。d,代表性的LbuC2c2pre-crRNA切割时间过程表明,相似的pre-crRNA加工速率不依赖于crRNA间隔区序列地发生。e,LbuC2c2 4聚突变体pre-crRNA加工数据表明了5'单链侧翼区域对于有效的pre-crRNA加工的重要性。60分钟后测量pre-crRNA加工的百分比(平均值±标准差,n=3)。f,变性凝胶示出了在一系列不同EDTA和

EGTA浓度下的加工活性。高浓度的螯合剂改变产物的预期迁移模式

[0599] **实施例9:** LbuC2c2介导的crRNA生物发生取决于CRISPR重复序列的结构和序列。

[0600] 其它pre-crRNA加工酶, 诸如Cas6和Cas5d, 以高度特异性的方式识别它们各自相应的CRISPR重复序列的发夹。确定LbuC2c2指导RNA加工的序列和结构要求。用在共有重复序列的茎环或单链侧翼区域中具有突变的pre-crRNA执行切割测定(图10A至图10C)。在改变重复区域中的茎长度后, pre-crRNA切割活性显著减弱; 而重复共有的4碱基对(bp)茎在60分钟后导致 $85 \pm 0.7\%$ 的切割, -1bp和+1bp的茎分别导致仅 $54 \pm 1.3\%$ 和 $6 \pm 3.8\%$ 的切割(图10a)。类似地, 茎环的倒置或环长度的缩短降低了加工活性。对包括或靠近易断裂键的连续4nt突变的初步研究完全消除了LbuC2c2加工pre-crRNA的能力(图16D)。对完整crRNA重复序列的更广泛突变分析揭示发夹两侧的两个不同区域对碱基变化具有显著的敏感性(图10B; 图17)。加工活性完全不受二价金属离子螯合剂EDTA或EGTA的存在的影响(图10C; 图16E)。总之, 这些数据表明, C2c2 pre-crRNA切割是一种不依赖于二价金属离子的过程, 所述过程由对前体CRISPR转录物内的重复序列来源的发夹的结构特异性识别和序列特异性识别的组控制。

[0601] 图10A至图10C. LbuC2c2介导的crRNA生物发生取决于CRISPR重复序列的结构和序列。图10A, 在发夹的茎区域和环区域内含有结构突变的pre-crRNA的LbuC2c2加工的代表性凝胶。下面列出的加工百分比是在60分钟处针对每种条件定量的(平均值 $\pm$ 标准差, $n=3$)。图10B, 图形映射显示了pre-crRNA加工对CRISPR重复序列的依赖性。野生型重复序列如下方所示, 其中单独的条代表上文以红色标示的2个核苷酸突变。切割位点用卡通剪刀指示。在60分钟后测量加工百分比(平均值 $\pm$ 标准差, $n=3$)。所测试的突变体的图示发夹可以在图16A至图16F和图17中找到。图10C, 通过向标准加工反应条件中加入10-100mM的EDTA来激发加工反应的二价金属依赖性。示出了60分钟处的反应终点。

[0602] 图17. pre-crRNA双突变对pre-crRNA加工活性的影响的详细总结。根据方法, 在60分钟后测量pre-crRNA加工百分比(平均值 $\pm$ 标准差, $n=3$)。突变的核苷酸以黄色突出显示。关于图形表示, 参见图10B。

[0603] **实施例10:** LbuC2c2催化顺式和反式靶上的指导依赖性ssRNA降解

[0604] 成熟后, crRNA通常以高亲和力与Cas效应蛋白结合, 以产生能够进行序列特异性核酸识别的RNA指导的监督复合物。为了测试经C2c2加工的crRNA的功能, 使用LbuC2c2蛋白, 其在初始切割实验中表现出最强的活性。对于多种ssRNA底物, 观察到LbuC2c2仅在与互补crRNA结合时有效地切割, 表明LbuC2c2: crRNA以RNA指导的方式切割ssRNA(图11b)。此活性在下文中称为在靶或顺式靶切割(图11a)。LbuC2c2介导的顺式切割导致多种产物的梯化, 其中切割优先在尿嘧啶残基之前发生, 类似于LshC2c2(图18A至图18B)⁹, 这种活性不同于仅在与crRNA的碱基配对所指定的限定区域内催化靶核酸切割的其它II类CRISPR效应子(诸如Cas9和Cpf1)。测试了LbuC2c2反式地充当经crRNA激活的非特异性RNA内切核酸酶(即, C2c2催化的对未由crRNA靶向的RNA分子的切割)的能力(图11A)。通过在存在未标记的在靶(与crRNA互补的)ssRNA的情况下重复非靶切割反应来研究该能力。观察到在这些反式切割条件下5'-标记的非靶RNA快速降解(图11B)。该结果表明, 识别与crRNA的间隔区段互补的靶ssRNA激活C2c2以进行快速和非特异性的RNA反式降解。对于顺式在靶切割和反式非靶切割观察到相似的RNA切割速率和几乎相同的切割产物暗示两种活性中的相同核酸酶中

心。此外,这些结果表明,观察为“顺式”活性的情况仅是当只有一种底物可用时发生的反式活性(图18A至图18B)。

[0605] 图11A-11C.LbuC2c2催化顺式和反式靶上的指导依赖性ssRNA降解。图11A,C2c2指导依赖性ssRNA降解的两种模式的示意图。图11B,由LbuC2c2切割两种不同的放射性标记的ssRNA底物A和B。在37℃预形成100nM C2c2与50nM crRNA的复合物,并在25℃下加入大于1nM的5'-标记的靶RNA时开始反应。反式切割反应含有等摩尔(大于1nM)浓度的放射性标记的非指导互补底物和未标记的在靶ssRNA。图11C,在存在EDTA的情况下,激发负载有crRNA以及间隔区A的LbuC2c2以获得其在顺式(靶A经标记)和反式(在存在未标记的靶A的情况下,靶B经标记)切割条件下的切割活性。

[0606] 图18A至图18B.LbuC2c2 ssRNA靶切割位点映射a,按照方法进行的ssRNA靶切割测定,证明LbuC2c2介导对具有相同的间隔区互补序列但不同的5'侧翼序列(具有不同长度和核苷酸组成)的若干个放射性标记的ssRNA底物的‘顺式’切割。ssRNA底物的序列示出在凝胶的右侧和底部,其中crRNA-A和crRNA-B的间隔互补序列分别以黄色和红色突出显示。箭头指示所预测的切割位点。为了清楚起见而裁剪了凝胶。应当注意的是,在不同底物上产生的切割产物的模式(例如A.1对比A.2对比A.3)表明,切割位点选择主要由尿嘧啶偏好性驱动,并且表现为明显缺乏‘基于crRNA标尺的’切割机制,这与关于其它II类CRISPR单效应复合物(诸如Cas9和Cpf1)观察到的相反。有趣的是,针对底物A.0观察到的切割模式提示了对polyG序列的次要偏好性,所述次要偏好性可能可以设想地形成G-四链体或类似的二级结构。b,按照方法,使用一系列覆盖(tile)ssRNA靶的长度的crRNA进行的LbuC2c2 ssRNA靶切割测定。在该实验中使用的ssRNA底物的序列示出在凝胶下方,其中每个crRNA的间隔区互补序列以黄色突出显示。箭头指示所预测的切割位点。同样地,应当注意的是,对于每种crRNA,切割产物长度分布是非常相似的,再次表明明显缺乏‘基于crRNA标尺的’切割机制。在反应的子集中不存在几种切割产物可以通过ssRNA靶上存在结合的C2c2:crRNA,从而通过任何顺式(分子内)或反式(分子间)LbuC2c2活性位点在空间上阻断了对尿嘧啶的可及来解释。虽然对LbuC2c2的PFS偏好性的适当分析超出了本研究的范围,但观察到了3'侧翼序列的最小影响。预测的PFS碱基在所测试的每个指导序列旁边的图表中以红色标示。

[0607] 实施例11:LbuC2c2含有两种不同的核糖核酸酶活性

[0608] pre-crRNA加工活性以比crRNA介导的对靶RNA或非互补反式ssRNA切割慢80倍的表现速率发生(图12a)。在饱和酶条件下,1h后pre-crRNA加工仅达到约85%,而反式底物的RNA指导切割的速率由于反应在37℃下于1分钟内完成而几乎是不可测量的(图19A至图19C)。还发现,与pre-crRNA加工相对,在存在EDTA的情况下RNA指导的切割被完全消除,表明该活性是二价金属离子依赖性的(图11c)。鉴于pre-crRNA加工与RNA指导的切割之间在动力学和金属离子依赖性方面的明显差异,有理由认为C2c2可能代表具有两种正交RNA切割活性的新型CRISPR效应蛋白:一种用于crRNA成熟,并且另一种用于crRNA指导的非特异性RNA降解。为了测试该假设,将LbuC2c2的保守HEPN结构域内的几个残基(预测所述残基含有RNA切割活性位点)进行系统突变;并且评定突变体的pre-crRNA加工和RNA指导的RNA酶活性(图12b)。保守HEPN残基(R472A、R477A、R1048A和R1053)的双重突变体和四重突变体保留了强pre-crRNA切割活性(图12c)。相比之下,所有HEPN突变体都消除了RNA指导的切割活性,同时不影响与crRNA或ssRNA结合的能力(图20A至图20C)。切割速率差异和对点突变和

EDTA的差别敏感性表明,pre-crRNA加工和RNA指导的RNA切割是由C2c2蛋白内的不同催化中心介导的,前者是不依赖于金属的,而后者需要二价金属依赖性HEPN结构域。

[0609] 图12A至图12D.LbuC2c2含有两种不同的核糖核酸酶活性。图4A,由LbuC2c2在37°C下执行的对顺式ssRNA靶(黑色)和pre-crRNA(青色)切割的定量时间过程数据。指数拟合示出为实线($n=3$),并且所计算的一级速率常数(k_{obs})对于顺式靶切割和pre-crRNA切割分别为 9.74 ± 1.15 和 0.12 ± 0.02 。图12B,LbuC2c2的结构域架构描绘了HEPN突变的位置。图12C和图12D,双重和四重HEPN结构域突变体对图12C中的pre-crRNA加工和图12D中的ssRNA靶向的核糖核酸酶活性。对于两种活性,利用crRNA靶向间隔物A,如图11c所示执行顺式和反式切割反应。

[0610] 图19A至图19C crRNA间隔区长度、反应温度和crRNA的5'末端序列对靶RNA切割效率的依赖性。图19A,按照方法进行LbuC2c2 ssRNA靶切割测定,其中crRNA具有16nt、20nt或24nt的间隔区。图19B,按照方法在25°C或37°C下进行的LbuC2c2 ssRNA靶切割的时间过程。图19C,按照方法进行LbuC2c2 ssRNA靶切割时间过程,其中crRNA具有不同的5'-侧翼核苷酸突变。突变以红色突出显示。1-2个核苷酸的5'延伸可忽略不计地影响切割效率。相比之下,将侧翼区域缩短至3nt会减慢切割速率。关于代表性平行测定,指数拟合示出为实线。

[0611] 图20A-20C.LbuC2c2与成熟crRNA和靶ssRNA的结合数据。图20A,如在方法中所描述地进行过滤结合测定以确定成熟crRNA-A与LbuC2c2-WT、LbuC2c2-dHEPN1、LbuC2c2-dHEPN2或LbuC2c2-dHEPN1/dHEPN2的结合亲和力。将定量的数据与标准结合同种型拟合。误差条代表三次独立实验的标准偏差。来自三个独立实验的测量的解离常数(平均值 $\pm$ 标准偏差)为 27.1 ± 7.5 nM(LbuC2c2-WT)、 15.2 ± 3.2 nM(LbuC2c2-dHEPN1)、 11.5 ± 2.5 nM(LbuC2c2-dHEPN2),以及 43.3 ± 11.5 nM(LbuC2c2-dHEPN1/dHEPN2)。图20B,如所示,关于LbuC2c2-dHEPN1/dHEPN2:crRNA-A与‘在靶’A ssRNA或‘脱靶’B ssRNA之间的结合反应的代表性电泳迁移率变动测定。如方法中所述进行三次独立实验。为了清楚起见而裁剪了凝胶。图20C,将来自(b)的定量结合数据与标准结合同种型拟合。误差条代表三次独立实验的标准偏差。从三个独立实验测量的解离常数(平均值 $\pm$ 标准偏差)对于ssRNA A为 1.62 ± 0.43 nM并且对于ssRNA B为未检测到的(>10 nM)。

[0612] 实施例12:C2c2提供对复杂混合物中转录物的灵敏视觉检测。

[0613] C2c2对反式底物的强RNA刺激切割可用作检测转录物集合内的特定RNA的手段。迄今为止,最灵敏的RNA检测策略需要多轮基于聚合酶和/或逆转录酶的扩增。作为替代方案,测试了C2c2的RNA指导的反式内切核酸酶活性是否可用于切割用荧光团-猝灭剂标记的报告RNA底物,从而导致在靶RNA触发的RNA酶激活后荧光增加(图13a)。简而言之,使LbuC2c2负载有靶向细菌噬菌体 λ 的crRNA,并测试其检测掺入HeLa细胞总RNA的相应 λ ssRNA靶的能力。如果LbuC2c2成功检测到互补靶,则反式切割活性应被激活以切割报告物,从而从猝灭剂中释放荧光团。加入少至1-10pM的互补 λ 靶RNA后,在30分钟后观察到显著的crRNA特异性荧光增加(图13b和图21)。单独使用C2c2:crRNA复合物或在存在crRNA和非互补靶RNA的情况下的对照实验导致相对于RNA酶A阳性对照的可忽略不计的荧光增加(图13b和图21)。在10pM的 λ 靶RNA浓度下,预测仅约0.02%的C2c2:crRNA复合物处于活性状态,但观察到的荧光信号反映了取决于RNA靶而对报告RNA底物的约25-50%切割,表明了LbuC2c2具有强大的

多转换酶促活性。因此,即使在极低的活性蛋白水平下,crRNA指导的反式切割也是有效并且可检测的。

[0614] 鉴于C2c2加工其自身的pre-crRNA,测试了pre-crRNA加工和RNA检测是否可以在单一反应中组合。为了测试该想法,设计了含有与靶RNA A和 λ 2互补的间隔区的串联crRNA重复序列,并测试了它们检测掺入HeLa总RNA中的不断减少量的A和 λ 2RNA的能力。观察到了显著的荧光增加,这在量值和灵敏度方面与使用成熟crRNA的实验相似(图13b、图13c),表明串联的pre-crRNA可以由C2c2成功加工并用于RNA靶向。这些数据突出了在一个测定中使用单一串联pre-crRNA转录物对单个和/或多个RNA分子进行多重RNA检测的可能性。

[0615] 不受理论束缚,提出当经由与crRNA的碱基配对而在宿主细胞内检测到侵入性转录物时,C2c2被激活以反式地进行混杂的RNA切割(图13d)。

[0616] 图13A至图13D.C2c2提供对复杂混合物中转录物的灵敏视觉检测。图13A,使用猝灭的荧光RNA报告物通过C2c2进行的RNA检测方法的示意图。在互补的靶RNA结合后,C2c2催化报告RNA的降解,从而导致荧光信号的积累。图13B,在存在100ng总RNA的情况下,在30分钟后通过C2c2对不同浓度的靶RNA的荧光信号进行的定量。RNA酶A示出作为阳性RNA降解对照。数据显示为n=3个独立实验的平均值 $\pm$ 标准差。图13C,串联pre-crRNA加工还使得能够进行RNA检测。数据显示为n=3个独立实验的平均值 $\pm$ 标准差。图13D,VI型CRISPR途径的模型突显了C2c2的两种核糖核酸酶活性。

[0617] 图21.RNA酶检测测定 λ 2-ssRNA的时程。在存在不断增加量的 λ 2ssRNA(0-1nM)的情况下,将LbuC2c2:crRNA- λ 2与RNA酶-Alert底物(Thermo-Fisher))和100ng HeLa总RNA在37℃孵育120分钟。每5分钟进行荧光测量。1nM的 λ 2ssRNA反应在可以测量第一个时间点之前达到饱和。误差条代表三次独立实验的标准偏差。

[0618] 实施例13

[0619] 材料和方法

[0620] MEC2c2系统发生和候选选择。使用具有PROTGAMMALG进化模型和100个自展抽样的RAxML计算C2c2最大似然系统发生。通过MAFFT使用'einsi'方法比对序列。

[0621] C2c2蛋白的生产和纯化.使用从Integrated DNA Technologies订购的合成gBlock组装用于蛋白纯化的表达载体。将密码子优化的C2c2基因组序列用His₆-MBP-TEV切割位点进行N末端标记,其中表达由T7启动子驱动。经由野生型C2c2构建体的定点诱变来克隆突变蛋白。将表达载体转化到于37℃下在2xYT肉汤中生长的Rosetta2大肠杆菌细胞中。将大肠杆菌细胞在对数期期间用0.5M的ITPG诱导,并将温度降至16℃以过夜表达His-MBP-C2c2。随后收获细胞,重悬于裂解缓冲液(50mM Tris-HCl pH 7.0、500mM NaCl、5%甘油、1mM TCEP、0.5mM PMSF、以及不含EDTA的蛋白酶抑制剂(Roche))中,并通过超声裂解,并且将裂解物通过离心澄清。通过金属离子亲和色谱分离出可溶性His-MBP-C2c2,将含蛋白的洗脱物与TEV蛋白酶一起在4℃下孵育过夜,同时透析到离子交换缓冲液(50mM Tris-HCl pH 7.0、250mM KCl、5%甘油、1mM TCEP)中以切除His₆-MBP标签。将切割的蛋白上样到HiTrap SP柱上,并以线性KCl(0.25-1.5M)梯度洗脱。将阳离子交换色谱级分汇集,并用30kD的截留浓缩管(Thermo Fisher)浓缩。将C2c2蛋白经由在S200柱上进行尺寸排阻色谱而进一步纯化,并储存在凝胶过滤缓冲液(20mM Tris-HCl pH 7.0、200mM KCl、5%甘油、1mM TCEP)中以供后续酶测定。将表达质粒由Addgene保藏。

[0622] 产生RNA。除了crRNA AES461是订购合成的(Integrated DNA Technologies)之外,本研究中使用的RNA都是在体外转录的[参见图33]。如前所述进行体外转录反应,具有以下改变:T7聚合酶浓度降至10 μ g/mL,并且UTP浓度降至2.5mM。将转录在37 $^{\circ}$ C下孵育1-2小时以减少核苷酸的非模板加入,并经由用DNA酶I在37 $^{\circ}$ C下处理0.5-1小时来进行猝灭。将转录反应通过15%变性聚丙烯酰胺凝胶电泳(PAGE)纯化,并将所有RNA重悬于切割缓冲液(20mM HEPES pH 6.8、50mM KCl、5mM MgCl₂和5%甘油)中。对于放射性实验,在放射性标记之前用小牛肠磷酸盐(New England Biolabs)去除5'三磷酸酯,然后如前所述使用T4多核苷酸激酶(New England Biolabs)和[γ -³²P]-ATP(Perkin Elmer)对ssRNA底物进行5'末端标记。

[0623] pre-crRNA加工测定。在37 $^{\circ}$ C下在RNA加工缓冲液(20mM HEPES pH 6.8、50mM KCl、5mM MgCl₂、10 μ g/mL BSA、10 μ g/mL tRNA、0.05% Igepal CA-630和5%甘油)以及相对于5'-标记的pre-crRNA 100倍摩尔过量的C2c2(最终浓度分别为100nM和小于1nM)中执行Pre-crRNA切割测定。除非另有说明,否则在60分钟后用1.5X RNA上样染料(100%甲酰胺、0.025w/v%溴酚蓝和200 μ g mL⁻¹的肝素)猝灭反应。猝灭后,将反应在95 $^{\circ}$ C变性5分钟,然后通过12%或15%变性PAGE(0.5X TBE缓冲液)解析。通过向反应缓冲液中加入浓度范围为10-100mM的EDTA或EGTA来测试反应的金属依赖性。将条带通过磷光成像可视化,并用ImageQuant(GE Healthcare)定量。切割百分比被确定为产物条带强度与产物条带和未切割的pre-crRNA条带两者的总强度的比率,并使用ImageQuant TL软件(GE Healthcare)针对各测量的底物内的背景进行归一化,并使用Prism(GraphPad)与单相指数关联进行拟合。

[0624] 产物尺寸映射和3'末端部分鉴别。通过使用得自Ambion的RNA酶T1试剂盒比较产物条带到碱水解梯带和RNA酶T1消化梯带的凝胶迁移,而生物化学地确定切割产物的长度。对于水解梯带,将15nM的全长RNA底物在95 $^{\circ}$ C下于1X碱水解缓冲液(Ambion)中孵育5分钟。用1.5X RNA上样缓冲液猝灭反应,并冷却至-20 $^{\circ}$ C以立即停止水解。对于RNA酶T1梯带,将15nM的全长RNA底物在65 $^{\circ}$ C下在1X RNA测序缓冲液(Ambion)中解折叠。将反应冷却至环境温度,然后向反应中加入1U的RNA酶T1(Ambion)。15分钟后,通过苯酚-氯仿提取终止反应,并加入1.5X RNA上样缓冲液进行储存。将水解条带与切割样品平行地在15%变性PAGE上解析,并通过磷光成像可视化。对于3'末端部分鉴别,将来自加工反应的产物与10U的T4多核苷酸激酶(New England Biolabs)一起在37 $^{\circ}$ C下于加工缓冲液中孵育1小时。用1.5X RNA上样缓冲液猝灭反应,在20%变性PAGE上解析,并通过磷光成像可视化。

[0625] 小RNA测序分析。从SRA运行SRR3713697、SRR3713948和SRR3713950下载来自Smakov等人的RNA读段。使用具有以下选项的Bowtie2将配对末端读段局部映射到参考序列:“--重排--非常快速地定位--定位”。然后使用具有“-m 0-p两者”选项的mapped.py过滤映射以仅保留不包含错配的比对。所得映射的BAM文件是可用的。使用Geneious可视化读段覆盖,并使用Prism(GraphPad)绘图。

[0626] 靶切割测定。在25 $^{\circ}$ C或37 $^{\circ}$ C下在切割缓冲液(20mM HEPES pH6.8、50mM KCl、5mM MgCl₂和5%甘油)中执行靶切割测定。通过加热至65 $^{\circ}$ C达5分钟,以及然后在切割缓冲液中缓慢冷却至环境温度而将crRNA指导物预折叠。在加入5'末端标记的靶和/或其它非放射性标记的RNA靶底物之前,通常以2:1的蛋白与crRNA摩尔比在切割缓冲液中于37 $^{\circ}$ C下执行C2c2:crRNA复合物形成达10分钟。除非另有说明,否则对于所有反应,蛋白、指导物和靶的

最终浓度分别为100nM、50nM和小于1nM。用1.5X RNA上样染料猝灭反应,并通过15%变性PAGE (0.5X TBE缓冲液) 解析。将条带通过磷光成像可视化,并用ImageQuant (GE Healthcare) 定量。切割百分比被确定为所有较短产物的总条带化强度相对于未切割条带的比率,并使用ImageQuant TL软件 (GE Healthcare) 针对各测量的底物内的背景进行归一化,并使用Prism (GraphPad) 与单相指数关联进行拟合。

[0627] crRNA过滤结合测定。在RNA加工缓冲液 (20mM HEPES pH6.8、50mM KCl、5mM MgCl₂、10μg/mL BSA、10μg/mL酵母tRNA、0.01% Igepal CA-630和5%甘油) 中进行过滤结合测定。将LbuC2c2与放射性标记的crRNA (小于0.1nM) 一起在37℃下孵育1小时。将Tufryn、Protran和Hybond-N+以上面列出的顺序组装到点印迹设备上。将膜用50μL的平衡缓冲液 (20mM HEPES pH 6.8、50mM KCl、5mM MgCl₂和5%甘油) 洗涤两次,然后将样品施加到所述膜上。将膜再次用50μL的平衡缓冲液洗涤,干燥并通过磷光成像可视化。将数据用ImageQuant TL软件 (GE Healthcare) 定量,并使用Prism (GraphPad Software) 与结合等温线拟合。所有实验一式三份地进行。图例中报告了解离常数和相关误差。

[0628] 电泳迁移率变动测定。为了避免在ssRNA结合实验期间解离低浓度的LbuC2c2-dHEPN1/dHEPN2:crRNA复合物,结合反应含有恒定过量的LbuC2c2-dHEPN1/dHEPN2 (200nM),以及不断增加浓度的crRNA-A和小于0.1nM的靶ssRNA。在C2c2 EMSA缓冲液 (20mM HEPES pH 6.8、50mM KCl、10μg/mL BSA、100μg/mL酵母tRNA、0.01% Igepal CA-630和5%甘油) 中进行测定。如上所述在37℃下预形成LbuC2c2-crRNA-A复合物10分钟,然后加入5'-放射性标记的ssRNA底物,并在37℃下进一步孵育45分钟。然后将样品通过4℃下的8%天然PAGE (0.5X TBE缓冲液) 进行解析。将凝胶通过磷光成像而成像,使用ImageQuant TL软件 (GE Healthcare) 定量,并使用Prism (GraphPad Software) 与结合等温线拟合。所有实验一式三份地进行。图例中报告了解离常数和相关误差。

[0629] 荧光RNA检测测定。通过将1μM的Lbu-C2c2:C2c2与500nM的crRNA在37℃下孵育10分钟而预组装LbuC2c2:crRNA复合物。然后在存在185nM的RNA酶-Alert底物 (Thermo-Fisher)、100ng的HeLa总RNA和不断增加量的靶60nt ssRNA (0-1nM) 的情况下,将这些复合物稀释至在RNA加工缓冲液 (20mM HEPES pH 6.8、50mM KCl、5mM MgCl₂、10μg/mL BSA、10μg/mL酵母tRNA、0.01% Igepal CA-630和5%甘油) 中100nM的LbuC2c2:50nM的crRNA-λ2。将这些反应在荧光平板读数器中于37℃下孵育长达120分钟,其中每5分钟进行一次荧光测量 (λ_{ex} :485nm; λ_{em} :535nm)。通过减去在不存在靶ssRNA的情况下进行的反应获得的荧光值来获得经背景校正的荧光值。通过将50nM的RNA酶A与185nM的RNA酶-Alert底物一起孵育来测量最大荧光。对于用人总RNA中的β-肌动蛋白mRNA测量crRNA-ACTB介导的LbuC2c2激活,通过将1μM的LbuC2c2与500nM的crRNA-ACTB一起在37℃下孵育10分钟而预组装LbuCas9:crRNA复合物,并在上述条件下在存在不断增加量 (0-1μg) 的HeLa细胞总RNA或大肠杆菌总RNA (作为阴性对照) 的情况下进行反应。将这些反应在荧光平板读数器中于37℃下孵育长达180分钟,其中每5分钟进行一次荧光测量 (λ_{ex} :485nm; λ_{em} :535nm)。通过减去在不存在靶ssRNA的情况下进行的反应获得的荧光值来获得经背景校正的荧光值。对于耦合的pre-crRNA加工和RNA检测测定,通过将1μM的LbuC2c2与500nM的pre-crRNA-A-λ2一起在37℃下孵育20分钟而预组装LbuCas9-crRNA复合物,并在存在不断增加量的ssRNA A和ssRNAλ2 (各自0-1nM) 的情况下如上所述地进行反应。在每种情况下,误差条代表三次独立实验的标准

偏差。

[0630] 总RNA中的背景切割。如先前关于荧光RNA检测测定所述地组装LbuC2c2:crRNA λ 4复合物。在存在含有和不含10nM的 λ 4ssRNA靶的3 μ g总RNA的情况下,将复合物在RNA加工缓冲液中孵育。2小时后,通过trizol提取和乙醇沉淀分离RNA。使用制造商的方案,使用小RNA分析试剂盒(Small RNA Analysis Kit) (Agilent)在Bioanalyzer 2100 (Agilent)上解析重悬样品的RNA片段尺寸分布。将荧光强度曲线在Prism(GraphPad Software)中针对曲线覆盖归一化。

[0631] 结果

[0632] VI型CRISPR基因座缺乏明显的Cas6或Cas5d样内切核酸酶或tracrRNA。通过使用来自C2c2蛋白家族三个不同分支的纯化的重组C2c2蛋白同系物(图24a至图24b和图28和图29),在此呈现的数据显示所有三种C2c2酶切割经5'末端放射性标记的pre-crRNA底物,所述底物由全长共有重复序列和20个核苷酸(nt)的间隔区序列组成(图24c)。映射每个pre-crRNA:C2c2同源对的切割位点,揭示加工取决于C2c2同系物而发生在所预测的重复序列发夹结构上游2个或5个核苷酸的位置处(图24c、图30A)。令人惊讶的是,在生物化学上映射的5'-切割位点与之前报道的沙氏纤毛菌(LshC2c2)或斯氏李斯特菌(LseC2c2)pre-crRNA的切割位点不一致。对Shmakov等人的RNA测序数据集表的重新分析表明体内切割位点与此处报道的体外位点一致(图30b至图30i)。此外,使用来自口腔纤毛菌(LbuC2c2)的C2c2和包含串联发夹-重复序列阵列的较大pre-crRNA的切割测定产生了来源于两种单独的切割事件的两种产物(图31a),这与C2c2在加工从VI型CRISPR基因座产生的前体crRNA转录物中的作用一致。

[0633] 为了理解C2c2指导RNA加工的底物要求和机制,产生在共有重复序列的茎环或单链侧翼区域中具有突变的pre-crRNA,并测试其由LbuC2c2进行加工的能力(图25)。数据显示,在改变重复区域中的茎长度时,C2c2催化的切割减弱(图25a)。茎环的倒置或环长度的缩短也降低了C2c2的加工活性,而包括或靠近易断裂键的连续4nt突变完全消除了所述加工活性(图31b)。对全长crRNA重复序列的更广泛突变分析揭示发夹任一侧的两个不同区域对碱基变化具有显著的敏感性(图25b)。相比之下,加工动力学不存在对间隔区序列的依赖性(图31c)。对发夹的两个侧翼区域的这种敏感性使人联想到许多Cas6酶和Cas5d酶所需的序列和结构基序。相比之下,Cpf1不具有对3'发夹侧翼区域的任何依赖性,因为可变间隔区域邻接发夹茎。

[0634] LbuC2c2的机理研究揭示,加工活性不受二价金属离子螯合剂EDTA或EGTA存在的影响(图25c),表明了不依赖于金属离子的RNA水解机制。不依赖于金属离子的RNA水解的典型特征是分别在crRNA切割产物的5'和3'半部上形成2',3'-环磷酸酯和5'-氢氧化物。为了确定C2c2加工的底物的末端基团化学特性,将5'侧翼产物与T4多核苷酸激酶一起进一步孵育,此举去除2',3'-环磷酸酯以留下3'-羟基。在激酶处理后观察到了5'侧翼产物的改变的变性凝胶迁移,这与3'磷酸基团的去除一致(图31d)。C2c2的pre-crRNA加工活性的二价金属离子非依赖性与Cpf1的二价金属离子依赖性形成鲜明对比,Cpf1是唯一显示为执行指导加工的其它单蛋白CRISPR效应物。总的来说,这些数据表明,C2c2催化的pre-crRNA切割是一种不依赖于二价金属离子的过程,该过程可能使用一般的酸碱催化机制。

[0635] 成熟后,crRNA通常以高亲和力与Cas效应蛋白结合,以产生能够进行序列特异性

核酸识别的RNA指导的监督复合物。与先前使用LshC2c2的工作一致,LbuC2c2仅在此类底物可与crRNA中的互补序列碱基配对时催化有效的靶RNA切割(图32至图34)。鉴于观察到的C2c2混杂切割模式(图33),测试了LbuC2c2反式地充当经crRNA激活的非特异性RNA内切核酸酶的能力(图32b)。与顺式执行的非靶切割实验形成鲜明对比并且与先前对LshC2c2的观察一致,观察到了非靶RNA的反式快速降解(图32b)。该结果显示,靶识别激活C2c2进行RNA的一般非特异性降解。重要的是,对于相同RNA底物的顺式在靶切割和反式非靶切割观察到的相似的RNA切割速率和几乎相同的切割产物暗示两种活性中的相同核酸酶中心(图32b)。

[0636] 值得注意的是,crRNA介导的靶ssRNA切割以比pre-crRNA加工快约80倍的速率发生(图26a),并且与pre-crRNA加工不同,在存在EDTA的情况下RNA指导的靶切割严重减少,表明该活性是二价金属离子依赖性的(图26a, (图32c、图34)。鉴于这些明显的差异,有理由认为C2c2可能具有两种正交RNA切割活性:一种用于crRNA成熟,并且另一种用于crRNA指导的非特异性RNA降解。为了测试该假设,将LbuC2c2的保守HEPN基序内的几个残基进行系统突变;并且评定突变体的pre-crRNA加工和RNA指导的RNA酶活性(图26、图34d)。保守HEPN残基(R472A、R477A、R1048A和R1053)的双重突变体和四重突变体保留了强pre-crRNA切割活性(图26c)。相比之下,所有HEPN突变体都消除了RNA指导的切割活性,同时不影响与crRNA或ssRNA结合的能力(图34d、图35)。

[0637] 接下来,寻找可以在不破坏靶RNA切割的情况下消除pre-crRNA加工活性的突变。鉴于不能预测超出HEPN基序的任何其它潜在的RNA酶基序,并且C2c2蛋白与Cpf1没有同源性,所以整个LbuC2c2中的荷电残基被系统地突变。精氨酸残基(R1079A)被鉴别为在突变后导致严重减弱的pre-crRNA加工活性(图26c)。该C2c2突变酶保留了crRNA结合能力以及RNA靶切割活性(图35d、图26d)。总的来说,结果显示C2c2蛋白内的不同活性位点催化pre-crRNA加工和RNA指导的RNA切割。

[0638] 接下来测试C2c2对反式底物的强RNA刺激切割是否可用作检测转录物集合内的特定RNA的手段。虽然已经开发了许多基于聚合酶的方法用于RNA扩增和随后的检测,但是基本上没有方法能够直接检测靶RNA而无需对每个新RNA靶进行大量工程化或严格的设计约束。作为一种易于编程的替代方案,测试了C2c2的RNA指导的反式内切核酸酶活性是否可用于切割用荧光团-猝灭剂标记的报告RNA底物,从而导致在靶RNA触发的RNA酶激活后荧光增加(图27a)。使LbuC2c2负载有靶向细菌噬菌体 λ 的crRNA,并测试其检测掺入HeLa细胞总RNA的相应 λ ssRNA靶的能力。加入少至1-10pM的互补 λ 靶RNA后,在30分钟内发生了显著的crRNA特异性荧光增加(图27b和图36a)。单独使用C2c2:crRNA复合物或在存在crRNA和非互补靶RNA的情况下的对照实验导致相对于RNA酶A阳性对照的可忽略不计的荧光增加(图27b和图36a)。应当注意的是,在10pM的 λ 靶RNA浓度下,预测仅约0.02%的C2c2:crRNA复合物处于活性状态,但观察到的荧光信号反映了取决于RNA靶而对报告RNA底物的约25-50%切割。这些反应中背景RNA的片段尺寸解析揭示了显著的降解,即使是对高度结构化的tRNA也是如此(图36b)。由于报告RNA切割在存在大量过量的未标记RNA的情况下发生,因此推断LbuC2c2是一种强多转换酶,其能够识别每靶RNA至少 10^4 个转换。因此,与先前的观察结果不同,即使在极低的激活蛋白水平下,crRNA指导的反式切割也是有效并且可检测的。

[0639] 为了扩展该LbuC2c2 RNA检测系统,将crRNA设计为靶向内源性 β -肌动蛋白mRNA。相对于大肠杆菌总RNA,在存在人总RNA的情况下观察到了可测量的荧光增加,证明了该方

法的特异性(图27c)。此外,鉴于C2c2加工其自身的指导物,所以通过设计含有与靶RNA A和 λ 2互补的间接区的串联crRNA重复序列而将pre-crRNA加工和RNA检测组合在单一反应中。在检测测定中将LbuC2c2与该未加工的串联指导RNA一起孵育产生了显著的荧光增加,这在量值和灵敏度方面与使用成熟crRNA的实验相似(图27b、图27d)。总的来说,这些数据突显了利用C2c2的两种不同RNA酶活性进行一系列生物技术的令人兴奋的机会(图27e)。

[0640] 在细菌中,C2c2可能作为病毒RNA的前哨工作。不受理论束缚,提出当经由与crRNA的碱基配对而在宿主细胞内检测到侵入性转录物时,C2c2被激活以反式地进行混杂的RNA切割(图27e)。作为防御机制,这与真核生物中的RNA酶L和半胱天冬酶系统具有惊人的相似性,由此细胞信号分别在宿主细胞内引发混杂的核糖核酸降解或蛋白水解性降解,从而导致细胞凋亡。虽然III型CRISPR系统的RNA靶向机制通常导致原型间隔区-指导双链体内的RNA切割,但相关核酸酶Csx1和Csm6的最近实例提供了在VI型系统与多组分III型推断复合物之间令人信服的相似性。

[0641] 在此描述的数据显示,CRISPR-C2c2蛋白代表一类新的酶,其能够具有两种单独的RNA识别和切割活性。有效的pre-crRNA加工需要CRISPR重复序列内的序列和结构基序,这阻止了非内源性crRNA负载并有助于降低该有效RNA酶的潜在毒性。C2c2和V型CRISPR效应蛋白Cpf1的完全不同的pre-crRNA加工机制表明,每个蛋白家族已经会聚于独立的活性,所述独立的活性涵盖它们各自相应的CRISPR途径的加工和干扰功能。此外,C2c2的两种不同催化能力可以一起用于RNA检测,因为激活C2c2以将每个检测到的靶RNA切割成数千个反式RNA使得能够实现有效的信号放大。C2c2从阵列加工其自身的指导RNA的能力允许使用组织特异性Pol II启动子来指导表达,以及用于广泛应用的多靶点混合检测(target multiplexing)。C2c2酶在细菌适应性免疫中因其双RNA酶活性而是独特的,并突显了利用CRISPR蛋白在细胞和无细胞系统中进行精确核酸操纵的实用性。

[0642] 实施例14

[0643] 在LbuC2c2中鉴别出了另外的氨基酸位置,所述另外的氨基酸位置在突变时导致严重减弱的pre-crRNA加工活性(图37)。氨基酸位置包括R1079(例如,R1079A)、R1072(例如,R1072A)和K1082(例如,K1082A)。

[0644] 实施例15

[0645] Cas13a酶包含两种不同的官能团,这两种不同的官能团识别正交的crRNA组并具有不同的ssRNA切割特异性。Cas13a pre-crRNA加工对于ssRNA切割不是必需的,但是其增强了ssRNA对在CRISPR阵列内部编码的crRNA的靶向。定义了两种Cas13a蛋白亚家族,这些亚家族可以并行操作,以用于在细菌中进行RNA检测和破坏两者和用于诊断应用。

[0646] 材料和方法

[0647] Cas13a系统发生和重复保守分析。使用具有PROTGAMMALG进化模型和100个自展值的RAxML (Stamatakis, Bioinformatics. 2014年5月1日; 30 (9): 1312-3) 计算Cas13a最大似然系统发生。蛋白进化枝 α 和 β 被定义为自展值大于90的分支点,表明在具有共同祖先方面的高置信度。其余的蛋白被标记为模糊的祖先,因为它们之间的系统发生关系是低置信度的,反映为自展值小于90。通过MAFFT使用'einsi'方法 (Kato和Standley, Mol Biol Evol. 2013年4月; 30 (4): 772-80) 比对序列。选择候选者以代表Cas13a蛋白树的每个主要分支。对所有非冗余同系物(图41)和候选蛋白进行比对。使用Needleman-Wunsch算法,经由

EMBL-EBI上的Needle工具计算逐对相似性得分而进行CRISPR-RNA (crRNA) 重复序列的比较 (McWilliam等人, *Nucleic Acids Res.* 2013年7月; 41 (Web Server issue) :W597-600)。使用相似性得分矩阵在R中进行CRISPR crRNA的分层聚类。

[0648] Cas13a蛋白的表达和纯化。使用从Integrated DNA Technologies (IDT) 订购的合成gBlock组装用于蛋白纯化的表达载体。将密码子优化的Cas13a基因组序列用His₆-MBP-TEV切割位点序列进行N末端标记, 其中表达由T7启动子驱动。经由野生型Cas13a构建体的定点诱变来克隆突变蛋白。如上文实施例7和13中所述地进行所有同系物的纯化。简而言之, 将表达载体转化到于37℃下在2xYT肉汤中生长的Rosetta2 DE3或BL21大肠杆菌细胞中, 在对数中期用0.5mM的ITPG诱导, 然后将温度转变至16℃以进行过夜表达。将细胞沉淀物重悬于裂解缓冲液 (50mM Tris-Cl pH 7.0、500mM NaCl、5%甘油、1mM TCEP、0.5mM PMSF、以及不含EDTA的蛋白酶抑制剂 (Roche)) 中, 通过超声裂解, 并且通过以15,000g离心澄清。通过金属离子亲和色谱分离出可溶性His₆-MBP-TEV-Cas13a, 并且为了切除His₆-MBP标签, 将含蛋白的洗脱物与TEV蛋白酶一起在4℃下孵育过夜, 同时透析到离子交换缓冲液 (50mM Tris-Cl pH 7.0、250mM KCl、5%甘油、1mM TCEP) 中。将切割的蛋白上样到HiTrap SP柱 (GE Healthcare) 上, 并以线性KCl (0.25-1.5M) 梯度洗脱。将含有Cas13a的级分汇集、浓缩, 并在S200柱 (GE Healthcare) 上于凝胶过滤缓冲液 (20mM Tris-Cl pH 7.0、200mM KCl、5%甘油、1mM TCEP) 中经由尺寸排阻色谱进一步纯化, 随后在-80℃下储存。使用该方案纯化所有同系物, 不同之处在于LwaCas13a结合到HiTrap肝素柱而不是SP柱, 并且尺寸排阻色谱步骤由于样品在离子交换后的足够纯度而被省略。将所有表达质粒由Addgene保藏。

[0649] 体外RNA转录。使用先前描述的方法 (Sternberg等人, *RNA*. 2012年4月; 18 (4) :661-72) 并且如在以上实施例中所述地在体外转录所有pre-crRNA、成熟crRNA和靶。简而言之, 所有底物都从单链DNA寡核苷酸模板 (IDT) 转录, 不同之处在于成熟crRNA需要非GR的5'末端。对于这些成熟crRNA, 使用重叠PCR产生紧接在成熟crRNA序列上游的含有锤头型核酶序列的T7聚合酶模板, 然后纯化以用作T7转录的模板 (关于序列, 参见图50 (表5))。Lbu六聚CRISPR阵列体外转录模板由GeneArt (ThermoFisher) 合成以作为质粒。将T7启动子-CRISPR阵列区域在用作T7转录的模板之前进行PCR扩增和纯化。使用15%的尿素-PAGE纯化所有转录的RNA, 不同之处在于使用6%的尿素-PAGE纯化阵列。随后用小牛碱性磷酸酶处理所有RNA以去除5'磷酸酯。使用先前描述的方法 (Sternberg等人, *RNA*. 2012年4月; 18 (4) :661-72) 并且如在以上实施例中所述地执行放射性标记。由IDT合成A、C、G和U均聚物, 以及用于反式ssRNA切割的荧光标记的RNA报道基因。在放射性标记后使用25%的尿素-PAGE纯化均聚物, 以降低底物异质性。

[0650] 放射性标记的ssRNA核酸酶测定。在37℃下在RNA加工缓冲液 (20mM HEPES-Na pH 6.8、50mM KCl、5mM MgCl₂、10μg/mL BSA、10μg/mL tRNA、0.05% Igepal CA-630和5%甘油) 以及相对于5'-标记的pre-crRNA 100倍摩尔过量的Cas13a (最终浓度分别为100nM和小于1nM) 中执行Pre-crRNA加工测定。除非另有说明, 否则在60分钟后用1.5X RNA上样染料 (100%甲酰胺、0.025%溴酚蓝和200μg mL⁻¹肝素) 猝灭反应。猝灭后, 将反应在95℃变性5分钟, 然后通过15%变性PAGE (0.5x TBE缓冲液) 解析。在37℃下在切割缓冲液 (20mM HEPES-Na pH 6.8、50mM KCl、5mM MgCl₂和5%甘油) 中执行靶切割测定。通常, 在加入5'标记的靶和/或其它非放射性标记的RNA靶底物之前, 以2:1的蛋白与crRNA摩尔比在切割缓冲液中于

37℃下执行Cas13a:crRNA复合物形成达60分钟。除非另有说明,否则最终浓度为100nM Cas13a、50nM crRNA或pre-crRNA、50nM的与crRNA互补的靶ssRNA(此后称为“激活物”)和小于1nM的反式ssRNA靶。将所有条带通过磷光成像(Typhoon, GE Healthcare)可视化,并用ImageQuant (GE Healthcare)定量。对于pre-crRNA加工,切割百分比被确定为产物条带强度与产物条带和未切割的pre-crRNA条带两者的总强度的比率,并针对各测量的底物内的背景进行归一化。对于反式ssRNA切割反应,切割百分比被确定为小于靶的所有片段与泳道内总强度的比率,并针对每个底物内的背景进行归一化。随后使用Prism7 (GraphPad)将这些数据与单指数衰减拟合,并且在图例中报告切割率。

[0651] 荧光ssRNA核酸酶测定。如上所述,在切割缓冲液中组装Cas13a:crRNA复合物。加入150nM的RNA酶Alert报告物(IDT)和各种最终浓度(0-1μM)的ssRNA激活物以引发反应。值得注意的是,这些反应是在没有竞争tRNA或总RNA的情况下,以更准确地测量反式切割活性。将这些反应在荧光平板读数器(Tecan Infinite Pro F2000)中于37℃下孵育长达120分钟,其中每5分钟进行一次荧光测量(λ_{ex} :485nm; λ_{em} :535nm)。通过减去在不存在靶ssRNA激活物的情况下进行的反应获得的荧光值来获得经背景校正的荧光值。为了确定同系物敏感度和阵列加工效果,使用Prism7 (GraphPad)将经背景校正的值与单指数衰减进行拟合,并且将所计算的速率和它们与n=3的相关标准偏差一起进行绘图。为了比较非同源crRNA指导的反式ssRNA切割,替代地计算初始反应速率,这是由于整个数据集的荧光平台值差异。然后将速率相对于同源crRNA进行缩放,以使速率针对同系物归一化。参见图52、图53和图54,分别给出了关于归一化值的表7、表8和表9。对于荧光均聚物ssRNA报告物研究,使用标准条件在37℃下预孵育Cas13a:crRNA复合物60分钟。在将反应放入平板读数器之前,立即加入激活物ssRNA和200nM的荧光ssRNA报告物以引发反应。对于具有荧光均聚物ssRNA报告物的含Lbu-Cas13a和Lba-Cas13a的样品,分别使用10pM和1nM的激活物。对于pre-crRNA阵列实验,首先将300nM的Cas13a与50nM的pre-crRNA阵列在切割缓冲液中孵育1小时,以实现阵列的结合和加工。对于每种间隔区序列,以生物一式三份的方式将100pM的每种ssRNA激活物与150nM的RNA酶Alert报告物(IDT)一起加入以引发反应。使用单指数衰减、使用Prism7 (GraphPad)来计算表观速率,并将所计算的速率与它们的相关标准偏差一起进行绘图。

[0652] crRNA过滤结合测定。如上文实施例中所述地进行过滤结合测定。简而言之,将Cas13a和放射性标记的crRNA在37℃下在RNA加工缓冲液(20mM HEPES-Na pH 6.8、50mM KCl、5mM MgCl₂、10μg/mL BSA、10μg/mL酵母tRNA、0.01% Igepal CA-630和5%甘油)中孵育1小时。将Tufryn、Protran和Hybond-N+以上面列出的顺序组装到点印迹设备上。将膜用50μL的平衡缓冲液(20mM HEPES-Na pH 6.8、50mM的KCl、5mM MgCl₂和5%甘油)洗涤两次,然后将样品施加到所述膜上。将膜再次用50μL的平衡缓冲液洗涤,干燥并通过磷光成像可视化。将数据用ImageQuant TL软件(GE Healthcare)定量,并使用Prism (GraphPad Software)与结合等温线拟合。所有实验一式三份地进行。图例中报告了解离常数和相关误差。

[0653] 结果

[0654] 大多数Cas13a同系物具有pre-crRNA加工活性

[0655] 为了探索Cas13a蛋白的功能多样性,对于它们从同源pre-crRNA产生成熟crRNA的能力,比较来自蛋白家族树的10种同系物的pre-crRNA加工活性(图38A;图39)。与上面讨论

的三种同系物类似,另外七种Cas13a酶具有crRNA成熟活性(图38B和图38C)。迄今为止测试的十一种Cas13a蛋白中只有一种在广泛的测定条件下未表现出对其同源pre-crRNA的可检测的切割(HheCas13a)(图38;图39A)。在加工其天然crRNA的同系物中,除了LshCas13a之外全部都在保守crRNA-重复序列发夹上游四个核苷酸的磷酸二酯键处进行切割(图38B和图38C)。这些结果显示了在VI-A型CRISPR系统中Cas13a介导的crRNA生物发生活性的强保守性。

[0656] 大多数Cas13a酶中的保守crRNA成熟中心

[0657] 先前的研究已经暗示Cas13a的两个不同区域负责pre-crRNA加工。由LshCas13a晶体结构建立的一般Cas13a蛋白架构由N末端结构域和由两个螺旋结构域分开的两个HEPN(高等真核生物和原核生物核苷酸结合)结构域组成(图40A)(Liu等人,Cell 2017年1月12日;168(1-2):121-134.e12)。对于LbuCas13a,HEPN2结构域内单个残基(R1079)的突变足以显著降低pre-crRNA加工活性。相比之下,位于螺旋1结构域中的两个位置R438和K441处的突变显示为减少由LshCas13a进行的pre-crRNA切割(Liu等人,Cell 2017年1月12日;168(1-2):121-134.e12)。虽然这两个区域都可能参与pre-crRNA加工,但尚不清楚由LbuCas13a进行的加工是否被螺旋1结构域突变抑制,以及哪个结构域主要负责跨Cas13a蛋白家族的crRNA成熟。

[0658] 跨19个Cas13a同系物检查了螺旋1结构域和HEPN2结构域两者的保守性。先前报道的比对(Liu等人,Cell 2017年1月12日;168(1-2):121-134.e12;Shmakov等人,Mol Cell.2015年11月5日;60(3):385-97)在螺旋1域区域中有冲突,表明了同系物之间关系在该结构域方面的高度模糊性。观察到了参与pre-crRNA加工的螺旋1结构域的比对中的最小保守;相比之下,在HEPN2结构域内存在一致的保守(图40B至图40C;图41)。唯一的pre-crRNA加工缺陷型同系物HheCas13a,在两个结构域中保持大部分保守的荷电残基,表明蛋白或重复序列的其它部分可阻止pre-crRNA切割。在该研究中使用的同系物中,LshCas13a在HEPN2结构域中是最不同的,从而潜在地解释了该同系物的替代催化结构域和由该同系物进行的非典型切割位点选择。

[0659] 测试了对LshCas13a pre-crRNA加工至关重要的螺旋1结构域残基对LbuCas13a的pre-crRNA加工活性的影响。测试了四种残基(E299、K310、R311和N314)在pre-crRNA切割中的作用(图40B和图40C)(Liu等人,Cell 2017年1月12日;168(1-2):121-134.e12;Shmakov等人,Mol Cell.2015年11月5日;60(3):385-97)。这些残基突变为丙氨酸揭示了对pre-crRNA加工效率的以下一系列影响:N314A显著降低切割速率,R311A活性受损最小,并且E299A和K310对pre-crRNA加工没有影响(图40D)。并行地执行LbuCas13a的HEPN2结构域内的诱变。R1072和K1082处的丙氨酸取代显著减少了pre-crRNA切割,而同一区域中的其它突变(D1078A、K1080A和K1087A)具有对pre-crRNA加工的最小影响(图40E)。这些结果表明,HEPN2和(较小程度地)螺旋1结构域在LbuCas13a的crRNA生物发生中起重要作用,但是直接负责催化水解的残基仍然未知。LbuCas13a和LshCas13a在参与pre-crRNA加工的区域之间的差异(Liu等人,Cell 2017年1月12日;168(1-2):121-134.e12)并不一定相互矛盾,因为crRNA的5'末端保持在LshCas13a结构中的HEPN2结构域与螺旋1结构域之间(Liu等人,Cell 2017年1月12日;168(1-2):121-134.e12)。来自这两个结构域的残基可能通过使底物结合稳定化、促进适当的底物取向和/或催化水解而在pre-crRNA加工中起关键作用。

[0660] LshCas13a的HEPN2结构域内缺乏保守、其推定的pre-crRNA加工活性位点以及该酶的非典型pre-crRNA切割位点使我们假设,LshCas13a可利用螺旋1结构域内的不同区域来催化pre-crRNA加工。在缺乏与pre-crRNA结合的Cas13a同系物的三维结构的情况下,通过将LshCas13a切割位点映射到非同源pre-crRNA上来测试该假设(图40F)。LshCas13a能够加工来自LwaCas13a和LbuCas13a的pre-crRNA,从而产生相对于所预测的发夹碱基移位一个核苷酸的切割位点。与该观察结果一致,LwaCas13a和LbuCas13a对Lsh pre-crRNA的加工发生在距重复序列茎标准四核苷酸间隔处,不同于同源LshCas13a位点。这支持以下观察结果:不同的LshCas13a加工位点取决于蛋白架构而不是pre-crRNA序列,并且LshCas13a是Cas13a树内关于pre-crRNA加工的异常蛋白。

[0661] Cas13a酶在尿嘧啶或腺苷处引发ssRNA切割

[0662] 先前的研究证实,Cas13a:crRNA复合物识别并结合互补的ssRNA靶,由此被称为ssRNA激活物,以触发在暴露的尿苷残基处的一般RNA酶活性(图42A)。测试同源组是否保留了由LbuCas13a和LshCas13a所表现出的非特异性降解活性。还测试HEPN活性位点内的尿苷偏好性是否在该家族中是普遍的。为了系统地测试一般RNA酶活性,监测包含Cas13a:crRNA以及结合的ssRNA激活物的三元复合物降解反式ssRNA靶的能力(图42A)。在一小时的过程中检测到十种同系物中的八种具有反式ssRNA切割活性(图42B;图43A)。最值得注意的是,具有反式ssRNA切割活性的Cas13a同系物是HheCas13a,其具有不可检测的pre-crRNA加工活性,但是催化由其同源全长pre-crRNA指导的底物的完全降解。虽然先前表征的同系物LshCas13a (Abudayyeh等人,Science.2016年8月5日;353(6299):aaf5573) 和LbuCas13a都表现出对易断裂键的5'端尿苷的偏好性,但是由其它同系物产生的不同长度的产物表明,该蛋白家族中可能存在不同的活性位点核苷酸偏好性(图43B)。

[0663] 为了进一步探测Cas13a同系物的反式ssRNA切割核苷酸偏好性,使用A、C、G和U的5聚体均聚物作为底物来测量这些酶的反式切割能力。六种同系物能够切割这些短底物(图42C;图43C;图50(表5))。虽然同系物中的四种LbuCas13a、LwaCas13a、HheCas13a和PprCas13a表现出对同型尿苷底物的偏好性切割,但是观察到了具有最高活性的同系物的次要偏好性(图42C;图43D)。相比之下,LbaCas13a和EreCas13a偏好性同型腺苷,这与较长靶上的生物化学映射的切割位点一致(图43B)。CamCas13a从这些长底物产生相同产物与Cas13a家族树的该进化枝的腺苷偏好性一致(图42、图39;图43)。

[0664] 这些酶之间的一个显著区别是在等摩尔的ssRNA激活物和Cas13a:crRNA干扰复合物的测试条件下,反式ssRNA切割达到饱和的速率。这些酶差异可能对Cas13a的生物学作用具有巨大影响。定量Cas13a同源家族内在反式ssRNA切割方面的方差。开发考虑到ssRNA激活物结合和反式ssRNA切割两者(有助于总酶促输出的Cas13a酶的两种核心属性),利用短荧光ssRNA报告物进行的RNA切割高通量筛选。为了查询每种Cas13a同系物的灵敏度,加入不断减少量的互补ssRNA激活物以引发反应,并且根据所得时间过程中的每一者计算荧光ssRNA报告物切割的表观速率。虽然所计算的速率是所述酶中的每种酶的ssRNA激活物结合亲和力和催化转换速率的卷积,但是它们给出了同系物中相当的切割活性的相对量度。

[0665] 五种同系物(LbuCas13a、LwaCas13a、LbaCas13a、HheCas13a和PprCas13a)在该测定中显示出足够用于可再现分析的可检测切割活性。在这五种同系物中,LbuCas13a表现出最高的灵敏度,在存在仅10fM互补激活物的情况下具有可检测的报告物切割(图42D)。只有

两种同系物LwaCas13a和PprCas13a显示出足够用于在皮摩尔范围内检测激活物的活性,其中灵敏度分别为10pM和100pM。LbaCas13a和HheCas13a的灵敏度要低得多,仅在纳摩尔水平的报告物时变得有活性,该纳摩尔水平相对于Cas13a复合物接近等摩尔。由于该测定依赖于大量的反式切割事件以产生可检测的荧光,因此可以假设虽然具有相似的切割位点偏好性但不能产生高于背景的可检测信号的三种同系物(EreCas13a、CamCas13a和LshCas13a)具有灵敏性甚至更低的互补靶灵敏度。灵敏度范围非常广(约 10^7 倍)表明Cas13a酶具有保护宿主生物免受外来RNA影响的多样性能力。

[0666] CRISPR重复序列确定非同源pre-crRNA加工

[0667] Cas13a的双重活性提供了研究pre-crRNA加工与不同VI-A型CRISPR操纵子之间靶向的相互依赖性的机会。为了确定这两种活性的底物需求,测试了不同同系物可以识别非同源crRNA以进行指导加工和靶向的程度。最初,尝试经由对Cas13a家族和crRNA相似性的系统发生分析而生物信息学地预测可能的crRNA可交换性。该分析表明存在两种进化枝的同系物,为了清楚起见称为 α 和 β (Stamatakis, Bioinformatics. 2014年5月1日; 30(9): 1312-3)(图44A)。所述纯化的同系物中的五种存在于这些进化枝之外,具有模糊的祖先关系;有疑问的是pre-crRNA(CRISPR重复序列)序列是否可能决定功能正交性。由于CRISPR重复序列的短且结构化的内容物,所以使用逐对序列比对评分矩阵在CRISPR重复序列之间建立分层聚类关系以对整个家族的变异进行评分(Burstein等人, Nat Commun. 2016年2月3日; 7:10613)。令人惊讶的是,该分析指出存在两种crRNA聚类,该两种crRNA聚类重叠但不同于由氨基酸序列确定的蛋白进化枝(图44B和图44C)。虽然聚类1crRNA与 α -进化枝蛋白的子集良好相关,并且所有 β -进化枝相关的crRNA在聚类2内,但是具有模糊系统发生关系的同系物分散在两个聚类中(图44C)。

[0668] 不能单独使用生物信息学分析容易地预测Cas13a:crRNA正交性,所以测试了非同源crRNA之间在由所述Cas13a同系物中的每一种进行加工和反式ssRNA靶切割两者方面功能可交换的程度(图44D和图44F)。对于pre-crRNA加工,发现由逐对序列比较定义的crRNA聚类预测了它们被它们的相关Cas13a蛋白加工的能力(图44D)。例如,来自聚类1的pre-crRNA仅由进化枝 α 的蛋白加工,并且对于聚类2和进化枝 β 反之亦然。相比之下,蛋白分类的预测性较差,因为大多数模糊分类的蛋白可加工来自重复聚类2的序列,而与它们的重复序列聚类的位置无关。

[0669] 三种同系物(PprCas13a、HheCas13a和Rca13a)是关于它们在crRNA聚类内的位置方面异常的pre-crRNA加工同系物。HheCas13a不能切割任何非同源的pre-crRNA,并且反之,没有同系物(包括HheCas13a在内)加工Hhe pre-crRNA。这表明,HheCas13a不能加工其同源pre-crRNA不仅反映了不同的重复序列(图38;图39),还反映了该蛋白中pre-crRNA加工活性的丧失。相比之下,PprCas13a和RcaCas13a的偏离活性通过crRNA重复序列差异来解释。PprCas13a和RcaCas13a加工它们自己的crRNA,但是两者也加工crRNA重复聚类2非同源序列。PprCas13a crRNA重复序列与其它聚类2crRNA重复序列在5'侧翼区域上的切割位点方面不同,表明由PprCas13a进行的pre-crRNA加工具有更大的底物灵活性。类似地,RcaCas13a crRNA的不同序列特征是相对于存在于该家族其余部分的crRNA中的标准五碱基对茎环延伸的六碱基对茎环。值得注意的是,每个crRNA重复序列内对Cas13a pre-crRNA加工的位置取代耐受性与上文讨论的对LbuCas13a:crRNA复合物的突变研究,以及所获得

的对LshCas13a:crRNA复合物的最近结构见解一致(Liu等人,Cell 2017年1月12日;168(1-2):121-134.e12)。总的来说,这些结果表明,VI-A型CRISPR重复序列的序列决定了其被Cas13a家族进行pre-crRNA加工的能力。然而,在存在不同重复序列的情况下进化的同系物(PprCas13a和RcaCas13a)保留了加工其它聚类2序列的能力。

[0670] 功能正交Cas13a酶的两个亚家族

[0671] 对在图44D中定义的pre-crRNA加工可交换性聚类是否胜任于指导由非同源Cas13a同系物进行的反式ssRNA切割进行了研究。为了研究由非同源pre-crRNA和成熟crRNA指导的Cas13a介导反式ssRNA切割,对上述实施例中所描述的荧光测定进行修改,并且分析仅限于在ssRNA切割实验中表现出显著切割活性的五种同系物(参见图42C)。概括地说,这些结果反映了pre-crRNA加工结果,其中crRNA重复聚类身份确定了功能组,但是其中与加工和靶向一致的一些鲜明对比是与酶促活性无关的(图44E和图44F)。例如,Ppr pre-crRNA和成熟crRNA可以指导由非同源蛋白LwaCas13a和LbuCas13a进行的ssRNA切割,尽管它们不能加工这些pre-crRNA。另一惊喜是HheCas13a(其由所有的聚类2的pre-crRNA和成熟crRNA指导进行反式ssRNA切割)的混杂性,尽管其缺乏与对这些指导物中的任一者的pre-crRNA加工活性。这表明,crRNA成熟不是反式ssRNA切割所需要的,这是与以下发现相符的观察结果:LbuCas13a pre-crRNA加工缺陷型突变体具有不受影响的反式ssRNA切割能力。

[0672] 关于pre-crRNA加工和反式ssRNA切割的crRNA可交换性比较限定了Cas13a蛋白家族中的两种功能正交的亚家族。第一组(例如,LbuCas13a)具有跨该蛋白家族的关于由crRNA指导的pre-crRNA加工和反式ssRNA切割两方面的广泛混杂性。此多源组也显示出对反式ssRNA切割活性位点内的尿苷的偏好性。由crRNA和蛋白序列两者限定的第二组(例如,LbaCas13a),具有不同的crRNA可交换特征并且在反式ssRNA靶切割期间优先在腺苷处进行切割。为了测试这些Cas13a亚家族的正交反应性,验证LbuCas13a和LbaCas13a分别切割同型AssRNA报告物或同型U ssRNA报告物(图45A和图45B)。虽然不同的探针产生相似量的荧光信号,但应当注意的是,由于同系物的不同灵敏度(对于LbuCas13a和LbaCas13a分别为10pM和1nM),所以加入了实质上不同量的ssRNA激活物。为了验证正交性,测试了一组具有在crRNA、激活物和报告物之间的所有可能非同源组合的对照反应,其中除了同源组合外没有检测到实质信号(图45C)。总的来说,这些结果限定了可以在相同系统内并行起作用的不同Cas13a同系物。

[0673] pre-crRNA加工增强CRISPR阵列情境中的靶向效率

[0674] 这项研究的一个令人费解的发现是,触发随后由Cas13a进行的反式ssRNA靶切割反应缺乏对成熟crRNA的严格需要。此外,LbuCas13a的加工缺陷型突变体保持相似的反式ssRNA切割效率,甚至当由pre-crRNA代替成熟crRNA进行指导时也是如此(图46;图47)。这导致了以下假设:pre-crRNA加工在VI型CRISPR基因座中的作用不一定是用于有效的ssRNA靶向,而是替代地用于从长CRISPR阵列转录物的范围中释放每个crRNA。有疑问的是pre-crRNA加工是否可以缓解在crRNA负载和/或ssRNA靶向期间的RNA折叠约束以及相邻Cas13a:crRNA-间隔区物质的潜在空间位阻。为了测试此点,比较使用野生型LbuCas13a或无pre-crRNA加工活性的突变体进行的CRISPR阵列指导的ssRNA切割的效率。

[0675] 由于LbuCas13a的所有加工缺陷型单点突变体(图40)保留低水平的pre-crRNA加

工活性,所以建立了不具有可检测的加工活性,但保留了与野生型LbuCas13a相似或更高的反式ssRNA切割效率的双重突变体(R1079A/K1080A)(图46A和图46B)。在存在靶RNA和由六个不同的重复间隔区单元组成的小CRISPR阵列转录物的情况下测试该双重突变体和野生型LbuCas13a酶的ssRNA切割(图46C和图46D)。对于该阵列内的所有间隔区序列,由无crRNA加工活性的突变体进行的反式ssRNA切割的速率显著下降(单侧t检验;对于所有对 $p < 0.001$)。与野生型LbuCas13a相比活性降低对于该阵列内的各个连续间隔区更明显,其中最后一个间隔区以仅为野生型酶催化的速率的15%的速率指导切割。该发现表明,虽然pre-crRNA加工对于靶向不是必需的,但是其通过从CRISPR阵列释放crRNA来增强活性,导致改进的VI型CRISPR系统模型(图55)。

[0676] 图38A至图38C.Pre-crRNA加工在Cas13a蛋白家族中是广泛保守的。(A) Cas13a催化的crRNA生物发生途径的示意图。由Cas13a切割pre-crRNA转录物以产生成熟crRNA。下面是pre-crRNA的示意图,突出显示了重要的功能特征。(B) 来自所研究的VI型CRISPR系统的CRISPR重复序列的5'部分的比对,突出显示了pre-crRNA切割位点。所映射的切割的切割位点示出为红色条。以黑色文本标示了与Lbu crRNA重复序列的偏差。小写g是转录目的所需要的,并且不是天然crRNA重复序列的部分。完整的CRISPR重复序列示于图39B中。(C) 与5'-放射性标记的pre-crRNA底物一起孵育60分钟后,由9种Cas13a同系物进行的Cas13a介导pre-crRNA切割的代表性凝胶。

[0677] 图39A至图39C.关于本研究中使用的Cas13a同系物的CRISPR基因座和crRNA重复结构(A) Cas13a蛋白的最大似然系统发生树,其中VI-A型基因座的图改编自(Shmakov等人,Mol Cell.2015年11月5日;60(3):385-97)。Cas13a ORF以青色示出。CRISPR阵列描绘为黑色框(重复序列)、黄色菱形(间隔区)和针对上面提到的较大阵列的间隔区阵列尺寸。基因座周围的所关注的ORF以以下缩写标示:T-毒素、AT-抗毒素,以及TP-转座酶。(B) 对来自所用同系物的CRISPR重复序列的人工比对。pre-crRNA加工切割位点以红线标示。以黑色文本标示了与Lbu crRNA重复序列的偏差。小写g是转录目的所需要的,并且不是天然crRNA重复序列的部分。测试了两种单独的Hhe crRNA序列,第一种含有天然序列,而第二种具有四核苷酸延伸部以延长非典型短的天然重复序列。在任何研究条件下,HheCas13a都不切割crRNA重复序列。(C) 在可变的盐和pH条件下,使用HheCas13a对天然crRNA重复序列进行的pre-crRNA加工测定。没有观察到切割产物。

[0678] 图40A至图40F.对于由LbuCas13a进行的pre-crRNA切割重要的残基的鉴别(A) LbuCas13a结构域的组构示意图,其中结构域基于LshCas13a晶体结构注释(Liu等人,Cell 2017年1月12日;168(1-2):121-134.e12)。(B) 通过对LshCas13a的研究,对Cas13a的螺旋1结构域中参与pre-crRNA加工的局部区域,以及(C) 通过对LbuCas13a的研究,对Cas13a的HEPN2结构域内参与pre-crRNA加工的区域进行多重序列氨基酸比对。关于这些区域的完整家族树比对参见图41,并且关于完整的蛋白比对参见图57。突变严重影响pre-crRNA加工的残基以黄色菱形标记,并且突变最低限度地影响pre-crRNA加工的残基以青色菱形标记。LbuCas13a序列上方的符号对应于对LbuCas13a进行的突变,并且LshCas13a序列下方的符号对应于由Liu等人,Cell 2017年1月12日;168(1-2):121-134.e12对LshCas13a进行的突变。矩阵比对的着色表示使用ClustalX方案的残基保守,其中较暗的色调指示保守的强度。测量突变体的(D) 螺旋1结构域和(E) HEPN2结构域中在单一转换条件下的pre-crRNA加工。

将定量数据与单指数衰减拟合,其中所计算的伪一级速率常数(kob)(平均值±标准差,n=3)如下:Lbu WT为 0.074 ± 0.003 分钟⁻¹,E299A为 0.071 ± 0.005 分钟⁻¹,K310A为 0.071 ± 0.003 分钟⁻¹,R311A为 0.054 ± 0.007 分钟⁻¹,N314A为 0.029 ± 0.008 分钟⁻¹,R1079A为 0.009 ± 0.007 分钟⁻¹,D1078A为 0.023 ± 0.002 分钟⁻¹,K1080A为 0.016 ± 0.004 分钟⁻¹,并且K1087A为 0.076 ± 0.007 分钟⁻¹,而R1072A和K1082A无法进行拟合。(F)使用标准条件,通过LbuCas13a、LshCas13a和LwaCas13a蛋白对LshCas13a、LwaCas13a和LbuCas13a pre-crRNA进行pre-crRNA加工的代表性凝胶。最右边泳道中的水解梯带允许相对尺寸比较,尽管三种pre-crRNA之间的细微序列差异将改变这些小片段的迁移。

[0679] 图41A至图41B.螺旋1结构域和HEPN结构域的完全比对(A-B)跨(A)螺旋1结构域和(B)HEPN2结构域对19种Cas13a家族成员的多重序列比对。GI登录号列在每个物种名称之前。下面列出了基于比对方列出的LshCas13a序列的坐标和LbuCas13a序列的坐标。在该研究中测试的突变在比对方以黄色和青色菱形标示,所述黄色和青色菱形分别对应于不利地影响pre-crRNA加工反应的丙氨酸取代或对pre-crRNA加工具有最小影响的丙氨酸取代。由(Liu等人,Cell 2017年1月12日;168(1-2):121-134.e12)测试的突变描绘于LshCas13a序列下方。通过Jalview,使用AMAS方法计算保守评分。

[0680] 图42A至图42D.Cas13a蛋白家族的成员以一系列效率切割ssRNA(A)由Cas13a进行的ssRNA靶向的示意图。为简单起见,反式ssRNA切割是研究的重点。(B)在孵育60分钟后,由所有十种同系物进行的Cas13a介导反式ssRNA切割的代表性凝胶。如方法中所述,使用成熟crRNA产物形成Cas13a:crRNA复合物,其中最终RNP复合物浓度为50nM。在存在和不存在50nM未标记的与crRNA互补的ssRNA激活物的情况下,加入小于1nM的放射性标记的反式ssRNA靶以引发反应。使用LshCas13a观察到了弱的反式ssRNA切割活性,其中产物条带在右侧用双箭头指示。(C)热图,报告了对于六种不同的Cas13a,各5聚体均聚物ssRNA底物的Cas13a催化的反式ssRNA切割百分比。测定条件与部分(B)相同,不同的是LbuCas13a和LwaCas13a,它们孵育5分钟而不是60分钟。(n=3,具有相关误差的值在图51(表6)中示出。(D)在一系列ssRNA激活物浓度下,五种同系物对荧光ssRNA报告物的表观切割速率。将Cas13a:crRNA复合物分别以2:1的比率预孵育,其中最终活性复合物浓度为50nM。加入互补ssRNA激活物和荧光ssRNA切割报告物以引发反应。将归一化的报告物信号曲线的时间过程使用单指数衰减进行拟合,并且将表观速率绘图(n=3)。一些条件在第一测量时间点之前达到平台,因此它们的速率被最低限度地假设为0.5分钟⁻¹,并在图表中用*标记。

[0681] 图43A至图43D.由Cas13a同系物进行的反式ssRNA切割(A)对在存在和不存在ssRNA激活物的情况下由十种不同的Cas13a同系物进行的反式ssRNA切割的时间过程分析。时间点取在1分钟、10分钟和60分钟处。标示了LbuCas13a、LwaCas13a、LbaCas13a、EreCas13a、HheCas13a、PprCas13a、CamCas13a和LshCas13a的ssRNA激活物特异性切割产物。虚线表示单独的PAGE凝胶之间的边界。(B)针对四种Cas13a同系物的多个切割反应映射的反式ssRNA切割位点。不同的切割模式以红色箭头标示。似乎LbaCas13a和EreCas13a可具有腺苷偏好性,而LwaCas13a似乎在核苷酸偏好性方面更加混杂。(C-D)由5种Cas13a同系物进行的对均聚ssRNA底物的代表性反式ssRNA切割凝胶。

[0682] 图44A至图44F.Cas13a家族中的crRNA可交换性(A)Cas13a蛋白的最大似然系统发生树。本研究使用的同系物以粗体表示,并且进化枝被突出显示。自展值位于图39中。(B)

关于来自本研究中使用同系物的CRISPR重复序列的对称相似性得分矩阵。行和列是按照CRISPR重复序列聚类排序的。(C)关于来自本研究中使用同系物的CRISPR重复序列的不对称相似性得分矩阵。在(B)中呈现了相同的逐对评分,不同之处为行被重新排序以对应于Cas13a系统发生树。(D-F) (D)由非同源蛋白进行的pre-crRNA加工、(E)由pre-crRNA指导的反式ssRNA切割,以及(F)由成熟crRNA指导的反式ssRNA切割的功能活性矩阵。使用标准条件执行加工测定,并分析60分钟的反应终点。使用荧光ssRNA报告物测定,以拟合的初始速率执行反式ssRNA切割测定。ssRNA激活物的浓度如下:LbuCas13a:100pM、LwaCas13a:100pM、PprCas13a:100nM、LbaCas13a:10nM以及HheCas13a:100nM。对三次平行测定的初始速率进行拟合以计算荧光平台值的差异,并针对每种Cas13a:crRNA同源对进行归一化。关于数值和相关误差($n=3$),参见图52、图53和图54(分别为表7、表8和表9)。

[0683] 图45A至图45C.对用于RNA检测的正交Cas13a亚家族的功能验证(A)RNA检测测定的示意图,所述RNA检测测定经改进以使用荧光均聚ssRNA报告底物来测定由LbuCas13a或LbaCas13a进行的反式ssRNA切割激活。(B)由与LbuCas13a:Lbu-crRNA:10pM ssRNA激活物或LbaCas13a:Lba-crRNA:1nM ssRNA激活物一起孵育的均聚报告物产生的原始荧光测量值的时间过程(平均值 $\pm$ 标准差, $n=3$)。(C)对于一组crRNA、ssRNA激活物和Cas13a蛋白组合,由荧光均聚ssRNA报告物产生的原始荧光测量值(平均值 $\pm$ 标准差, $n=3$)。

[0684] 图46A至图46D.对LbuCas13a的crDNA阵列加工作用的解释(A)相较于野生型LbuCas13a的R1079A/K1080A突变体的pre-crRNA加工测定的定量时间过程数据。将定量数据与单指数衰减拟合,并且LbuCas13a WT的伪一级速率常数(k_{ob}) (平均值 $\pm$ 标准差, $n=3$)为 0.074 ± 0.003 分钟 $^{-1}$,而R1079A/K1080A突变体不能以足够的置信度拟合产生速率常数。(B)由pre-crRNA和成熟crRNA指导的LbuCas13a野生型和无R1079A/K1080A加工活性的突变体的荧光报告物表观比率。将Cas13a:RNA复合物以1:1的比率预孵育60分钟,然后加入10pM的激活物和150nM的报告物以引发反应。(平均值 $\pm$ 标准差, $n=3$) (C)如由50nM的含有六个crRNA重复间隔区的CRISPR阵列指导的,由300nM野生型LbuCas13a或无R1079A/K1080A pre-crRNA加工活性的突变体进行的荧光ssRNA报告物切割的表观速率。每个条组代表加入100pM的不同ssRNA激活物序列,所述不同ssRNA激活物序列与在每个条组下方指示的CRISPR阵列内的示意位置(schematized position)互补。将每个速率根据来自三次生物平行测定的数据进行拟合,并描绘速率的标准偏差。对于所有间隔区位置,突变蛋白的速率与野生型LbuCas13a在统计学上不同(单侧t检验: $p<0.001$)。(D)来自(C)的数据被描述为野生型LbuCas13a活性的百分比,证明了由无pre-crRNA加工活性的突变体降低反式ssRNA靶向效率的位置效应。

[0685] 图47A至图47C.由LbuCas13a点突变体在参与pre-crRNA加工的区域中进行的反式切割。(A-C)由LbuCas13a的各种‘参与pre-crRNA加工的点突变体进行的反式ssRNA切割。如上所述,在25℃或37℃用0.2nM或1nM的ssRNA激活物使用50nM的Cas13a:crRNA执行切割反应。使用较低的温度和激活物浓度以减慢反应动力学,以进行更准确的测量。拟合的曲线是单指数衰减,其中所计算的伪一级速率常数(平均值 $\pm$ 标准差, $n=3$)如下:25℃条件:Lbu WT为 1.76 ± 0.24 分钟 $^{-1}$,K299A为 1.72 ± 0.65 分钟 $^{-1}$,K310A为 3.19 ± 0.27 分钟 $^{-1}$,R1072A为 2.95 ± 0.53 分钟 $^{-1}$,R1079A为 0.93 ± 0.28 分钟 $^{-1}$,并且K1082A为 2.64 ± 0.63 分钟 $^{-1}$,并且37℃条件:R311A为 0.39 ± 0.09 分钟 $^{-1}$,并且N314A为 0.54 ± 0.11 分钟 $^{-1}$,而野生型和无加工活

性的突变体 (R1079A/K1080A) 过快地达到平台而无法进行准确的速率测量:WT为 2.95 ± 1.97 分钟⁻¹,并且R1079A/K1080A为 4.86 ± 4.99 分钟⁻¹。

[0686] 图48A至图48B.无crRNA加工活性的突变体R1079A/K1080A保留相似的crRNA结合亲和力并且不加工pre-cRNA阵列(A)如方法中所述地进行过滤结合测定以确定成熟crRNA与LbuCas13a WT和LbuCas13a R1079A/K1080A的结合亲和力。将定量的数据与标准结合等温线拟合。从三个独立实验测量的解离常数(平均值 $\pm$ 标准偏差)为 1.21 ± 0.57 nM (LbuCas13a WT)和 3.11 ± 0.89 nM (LbuCas13a R1079A/K1080A)。(B)使用六聚体CRISPR阵列作为底物、以及LbuCas13a和LbuCas13a R1079A/K1080A突变体以及各种尺寸的标记物进行的pre-cRNA加工测定。产物标识描绘在凝胶的右侧。由于偶尔未加工的另外前导区,如所标示的偶尔会存在其它尺寸的产物。

[0687] 图49(表4).CRISPR重复共有序列:针对在含有Cas13a的操纵子(RcaCas13a、LbaCas13a和EreCas13a)的10kb内具有多个CRISPR基因座的同系物,或具有重复变异(PprCas13a)的长阵列,列出了多种crRNA重复序列,其中突变以红色文本突出显示。对于PprCas13a,选择阵列前导侧的第一crRNA重复序列用于该研究。对于RcaCas13a、LbaCas13a和EreCas13a,分析crRNA重复序列的以下两个因素以选择用于该研究的代表性crRNA(1):阵列的长度,以及(2)由同源Cas13a蛋白指导反式ssRNA切割的能力。以主要文本使用的序列在最后一列中注明。

[0688] 图50(表5). *寡聚物编号-用于保持本研究中使用的RNA底物的一致性的指数。 **来源缩写:SS-单链DNA寡核苷酸模板用于体外转录,HH PCR-体外转录模板是包含锤头型核酶模板序列的重叠寡核苷酸的PCR产物,IDT-由IDT合成的,PCR-从质粒扩增的体外转录模板。

[0689] 图51表示表6。图52表示表7。图53表示表8。图54表示表9。

[0690] 图55.VI-A型CRISPR系统功能的改进模型(A)本研究中的关键发现的图形概要。本研究使用的同系物用粗体缩写表示,其中无反式ssRNA切割活性的同系物用灰色绘示。彩色圆圈突出显示两个正交的Cas13a酶组,所述两个正交的Cas13a酶组如由它们的广义crRNA可交换性和进行反式ssRNA切割的底物核苷酸偏好性所限定。(B)描绘了VI-A型CRISPR系统功能的改进模型的示意图。

[0691] 虽然已经参考本发明的特定实施方案描述了本发明,但是本领域的技术人员应当理解,在不脱离本发明的真实精神和范围的情况下,可以进行各种改变并且可以替换等同物。另外,可以进行许多修改以使具体情况、材料、物质组成、过程、一个或多个过程步骤适应本发明的目的、精神和范围。所有此类修改都在所附权利要求书的范围内。

- [0001] 序列表(修改)
- [0002] <110> 加利福尼亚大学董事会(The Regents Of The University Of California)
- [0003] Doudna, Jennifer A
- [0004] O'Connell, Mitchell
- [0005] Seletsky, Alexandra E
- [0006] Knight, Spencer C
- [0007] Cate, Jamie
- [0008] <120> 用于检测靶RNA的方法和组合物
- [0009] <130> BERK-337CN
- [0010] <140> 17813959.8
- [0011] <141> 2018-11-26
- [0012] <150> US 62/351,172
- [0013] <151> 2016-06-16
- [0014] <150> US 62/378,156
- [0015] <151> 2016-08-22
- [0016] <150> PCT/US2017/037308
- [0017] <151> 2017-06-17
- [0018] <160> 323
- [0019] <170> PatentIn version 3.5
- [0020] <210> 1
- [0021] <211> 1120
- [0022] <212> PRT
- [0023] <213> 斯氏李斯特菌(*Listeria seeligeri*)
- [0024] <400> 1
- [0025] Met Trp Ile Ser Ile Lys Thr Leu Ile His His Leu Gly Val Leu Phe
- [0026] 1 5 10 15
- [0027] Phe Cys Asp Tyr Met Tyr Asn Arg Arg Glu Lys Lys Ile Ile Glu Val
- [0028] 20 25 30
- [0029] Lys Thr Met Arg Ile Thr Lys Val Glu Val Asp Arg Lys Lys Val Leu
- [0030] 35 40 45
- [0031] Ile Ser Arg Asp Lys Asn Gly Gly Lys Leu Val Tyr Glu Asn Glu Met
- [0032] 50 55 60
- [0033] Gln Asp Asn Thr Glu Gln Ile Met His His Lys Lys Ser Ser Phe Tyr
- [0034] 65 70 75 80
- [0035] Lys Ser Val Val Asn Lys Thr Ile Cys Arg Pro Glu Gln Lys Gln Met
- [0036] 85 90 95
- [0037] Lys Lys Leu Val His Gly Leu Leu Gln Glu Asn Ser Gln Glu Lys Ile

[0038]	100	105	110
[0039]	Lys Val Ser Asp Val Thr Lys Leu Asn Ile Ser Asn Phe Leu Asn His		
[0040]	115	120	125
[0041]	Arg Phe Lys Lys Ser Leu Tyr Tyr Phe Pro Glu Asn Ser Pro Asp Lys		
[0042]	130	135	140
[0043]	Ser Glu Glu Tyr Arg Ile Glu Ile Asn Leu Ser Gln Leu Leu Glu Asp		
[0044]	145	150	155
[0045]	Ser Leu Lys Lys Gln Gln Gly Thr Phe Ile Cys Trp Glu Ser Phe Ser		
[0046]	165	170	175
[0047]	Lys Asp Met Glu Leu Tyr Ile Asn Trp Ala Glu Asn Tyr Ile Ser Ser		
[0048]	180	185	190
[0049]	Lys Thr Lys Leu Ile Lys Lys Ser Ile Arg Asn Asn Arg Ile Gln Ser		
[0050]	195	200	205
[0051]	Thr Glu Ser Arg Ser Gly Gln Leu Met Asp Arg Tyr Met Lys Asp Ile		
[0052]	210	215	220
[0053]	Leu Asn Lys Asn Lys Pro Phe Asp Ile Gln Ser Val Ser Glu Lys Tyr		
[0054]	225	230	235
[0055]	Gln Leu Glu Lys Leu Thr Ser Ala Leu Lys Ala Thr Phe Lys Glu Ala		
[0056]	245	250	255
[0057]	Lys Lys Asn Asp Lys Glu Ile Asn Tyr Lys Leu Lys Ser Thr Leu Gln		
[0058]	260	265	270
[0059]	Asn His Glu Arg Gln Ile Ile Glu Glu Leu Lys Glu Asn Ser Glu Leu		
[0060]	275	280	285
[0061]	Asn Gln Phe Asn Ile Glu Ile Arg Lys His Leu Glu Thr Tyr Phe Pro		
[0062]	290	295	300
[0063]	Ile Lys Lys Thr Asn Arg Lys Val Gly Asp Ile Arg Asn Leu Glu Ile		
[0064]	305	310	315
[0065]	Gly Glu Ile Gln Lys Ile Val Asn His Arg Leu Lys Asn Lys Ile Val		
[0066]	325	330	335
[0067]	Gln Arg Ile Leu Gln Glu Gly Lys Leu Ala Ser Tyr Glu Ile Glu Ser		
[0068]	340	345	350
[0069]	Thr Val Asn Ser Asn Ser Leu Gln Lys Ile Lys Ile Glu Glu Ala Phe		
[0070]	355	360	365
[0071]	Ala Leu Lys Phe Ile Asn Ala Cys Leu Phe Ala Ser Asn Asn Leu Arg		
[0072]	370	375	380
[0073]	Asn Met Val Tyr Pro Val Cys Lys Lys Asp Ile Leu Met Ile Gly Glu		
[0074]	385	390	395
[0075]	Phe Lys Asn Ser Phe Lys Glu Ile Lys His Lys Lys Phe Ile Arg Gln		
[0076]	405	410	415

[0077]	Trp Ser Gln Phe Phe Ser Gln Glu Ile Thr Val Asp Asp Ile Glu Leu
[0078]	420 425 430
[0079]	Ala Ser Trp Gly Leu Arg Gly Ala Ile Ala Pro Ile Arg Asn Glu Ile
[0080]	435 440 445
[0081]	Ile His Leu Lys Lys His Ser Trp Lys Lys Phe Phe Asn Asn Pro Thr
[0082]	450 455 460
[0083]	Phe Lys Val Lys Lys Ser Lys Ile Ile Asn Gly Lys Thr Lys Asp Val
[0084]	465 470 475 480
[0085]	Thr Ser Glu Phe Leu Tyr Lys Glu Thr Leu Phe Lys Asp Tyr Phe Tyr
[0086]	485 490 495
[0087]	Ser Glu Leu Asp Ser Val Pro Glu Leu Ile Ile Asn Lys Met Glu Ser
[0088]	500 505 510
[0089]	Ser Lys Ile Leu Asp Tyr Tyr Ser Ser Asp Gln Leu Asn Gln Val Phe
[0090]	515 520 525
[0091]	Thr Ile Pro Asn Phe Glu Leu Ser Leu Leu Thr Ser Ala Val Pro Phe
[0092]	530 535 540
[0093]	Ala Pro Ser Phe Lys Arg Val Tyr Leu Lys Gly Phe Asp Tyr Gln Asn
[0094]	545 550 555 560
[0095]	Gln Asp Glu Ala Gln Pro Asp Tyr Asn Leu Lys Leu Asn Ile Tyr Asn
[0096]	565 570 575
[0097]	Glu Lys Ala Phe Asn Ser Glu Ala Phe Gln Ala Gln Tyr Ser Leu Phe
[0098]	580 585 590
[0099]	Lys Met Val Tyr Tyr Gln Val Phe Leu Pro Gln Phe Thr Thr Asn Asn
[0100]	595 600 605
[0101]	Asp Leu Phe Lys Ser Ser Val Asp Phe Ile Leu Thr Leu Asn Lys Glu
[0102]	610 615 620
[0103]	Arg Lys Gly Tyr Ala Lys Ala Phe Gln Asp Ile Arg Lys Met Asn Lys
[0104]	625 630 635 640
[0105]	Asp Glu Lys Pro Ser Glu Tyr Met Ser Tyr Ile Gln Ser Gln Leu Met
[0106]	645 650 655
[0107]	Leu Tyr Gln Lys Lys Gln Glu Glu Lys Glu Lys Ile Asn His Phe Glu
[0108]	660 665 670
[0109]	Lys Phe Ile Asn Gln Val Phe Ile Lys Gly Phe Asn Ser Phe Ile Glu
[0110]	675 680 685
[0111]	Lys Asn Arg Leu Thr Tyr Ile Cys His Pro Thr Lys Asn Thr Val Pro
[0112]	690 695 700
[0113]	Glu Asn Asp Asn Ile Glu Ile Pro Phe His Thr Asp Met Asp Asp Ser
[0114]	705 710 715 720
[0115]	Asn Ile Ala Phe Trp Leu Met Cys Lys Leu Leu Asp Ala Lys Gln Leu

[0116]		725		730		735
[0117]	Ser Glu Leu Arg Asn Glu Met Ile Lys Phe Ser Cys Ser Leu Gln Ser					
[0118]		740		745		750
[0119]	Thr Glu Glu Ile Ser Thr Phe Thr Lys Ala Arg Glu Val Ile Gly Leu					
[0120]		755		760		765
[0121]	Ala Leu Leu Asn Gly Glu Lys Gly Cys Asn Asp Trp Lys Glu Leu Phe					
[0122]		770		775		780
[0123]	Asp Asp Lys Glu Ala Trp Lys Lys Asn Met Ser Leu Tyr Val Ser Glu					
[0124]	785		790		795	800
[0125]	Glu Leu Leu Gln Ser Leu Pro Tyr Thr Gln Glu Asp Gly Gln Thr Pro					
[0126]		805		810		815
[0127]	Val Ile Asn Arg Ser Ile Asp Leu Val Lys Lys Tyr Gly Thr Glu Thr					
[0128]		820		825		830
[0129]	Ile Leu Glu Lys Leu Phe Ser Ser Ser Asp Asp Tyr Lys Val Ser Ala					
[0130]		835		840		845
[0131]	Lys Asp Ile Ala Lys Leu His Glu Tyr Asp Val Thr Glu Lys Ile Ala					
[0132]		850		855		860
[0133]	Gln Gln Glu Ser Leu His Lys Gln Trp Ile Glu Lys Pro Gly Leu Ala					
[0134]	865		870		875	880
[0135]	Arg Asp Ser Ala Trp Thr Lys Lys Tyr Gln Asn Val Ile Asn Asp Ile					
[0136]		885		890		895
[0137]	Ser Asn Tyr Gln Trp Ala Lys Thr Lys Val Glu Leu Thr Gln Val Arg					
[0138]		900		905		910
[0139]	His Leu His Gln Leu Thr Ile Asp Leu Leu Ser Arg Leu Ala Gly Tyr					
[0140]		915		920		925
[0141]	Met Ser Ile Ala Asp Arg Asp Phe Gln Phe Ser Ser Asn Tyr Ile Leu					
[0142]		930		935		940
[0143]	Glu Arg Glu Asn Ser Glu Tyr Arg Val Thr Ser Trp Ile Leu Leu Ser					
[0144]	945		950		955	960
[0145]	Glu Asn Lys Asn Lys Asn Lys Tyr Asn Asp Tyr Glu Leu Tyr Asn Leu					
[0146]		965		970		975
[0147]	Lys Asn Ala Ser Ile Lys Val Ser Ser Lys Asn Asp Pro Gln Leu Lys					
[0148]		980		985		990
[0149]	Val Asp Leu Lys Gln Leu Arg Leu Thr Leu Glu Tyr Leu Glu Leu Phe					
[0150]		995		1000		1005
[0151]	Asp Asn Arg Leu Lys Glu Lys Arg Asn Asn Ile Ser His Phe Asn					
[0152]		1010		1015		1020
[0153]	Tyr Leu Asn Gly Gln Leu Gly Asn Ser Ile Leu Glu Leu Phe Asp					
[0154]		1025		1030		1035

[0155]	Asp	Ala	Arg	Asp	Val	Leu	Ser	Tyr	Asp	Arg	Lys	Leu	Lys	Asn	Ala	
[0156]	1040						1045					1050				
[0157]	Val	Ser	Lys	Ser	Leu	Lys	Glu	Ile	Leu	Ser	Ser	His	Gly	Met	Glu	
[0158]	1055						1060					1065				
[0159]	Val	Thr	Phe	Lys	Pro	Leu	Tyr	Gln	Thr	Asn	His	His	Leu	Lys	Ile	
[0160]	1070						1075					1080				
[0161]	Asp	Lys	Leu	Gln	Pro	Lys	Lys	Ile	His	His	Leu	Gly	Glu	Lys	Ser	
[0162]	1085						1090					1095				
[0163]	Thr	Val	Ser	Ser	Asn	Gln	Val	Ser	Asn	Glu	Tyr	Cys	Gln	Leu	Val	
[0164]	1100						1105					1110				
[0165]	Arg	Thr	Leu	Leu	Thr	Met	Lys									
[0166]	1115						1120									
[0167]	<210>	2														
[0168]	<211>	1159														
[0169]	<212>	PRT														
[0170]	<213>	口腔纤毛菌(Leptotrichia buccalis)														
[0171]	<400>	2														
[0172]	Met	Lys	Val	Thr	Lys	Val	Gly	Gly	Ile	Ser	His	Lys	Lys	Tyr	Thr	Ser
[0173]	1				5					10					15	
[0174]	Glu	Gly	Arg	Leu	Val	Lys	Ser	Glu	Ser	Glu	Glu	Asn	Arg	Thr	Asp	Glu
[0175]					20					25					30	
[0176]	Arg	Leu	Ser	Ala	Leu	Leu	Asn	Met	Arg	Leu	Asp	Met	Tyr	Ile	Lys	Asn
[0177]					35					40					45	
[0178]	Pro	Ser	Ser	Thr	Glu	Thr	Lys	Glu	Asn	Gln	Lys	Arg	Ile	Gly	Lys	Leu
[0179]		50							55					60		
[0180]	Lys	Lys	Phe	Phe	Ser	Asn	Lys	Met	Val	Tyr	Leu	Lys	Asp	Asn	Thr	Leu
[0181]	65						70					75				80
[0182]	Ser	Leu	Lys	Asn	Gly	Lys	Lys	Glu	Asn	Ile	Asp	Arg	Glu	Tyr	Ser	Glu
[0183]						85					90					95
[0184]	Thr	Asp	Ile	Leu	Glu	Ser	Asp	Val	Arg	Asp	Lys	Lys	Asn	Phe	Ala	Val
[0185]						100					105				110	
[0186]	Leu	Lys	Lys	Ile	Tyr	Leu	Asn	Glu	Asn	Val	Asn	Ser	Glu	Glu	Leu	Glu
[0187]						115								120		125
[0188]	Val	Phe	Arg	Asn	Asp	Ile	Lys	Lys	Lys	Leu	Asn	Lys	Ile	Asn	Ser	Leu
[0189]						130								135		140
[0190]	Lys	Tyr	Ser	Phe	Glu	Lys	Asn	Lys	Ala	Asn	Tyr	Gln	Lys	Ile	Asn	Glu
[0191]	145							150					155			160
[0192]	Asn	Asn	Ile	Glu	Lys	Val	Glu	Gly	Lys	Ser	Lys	Arg	Asn	Ile	Ile	Tyr
[0193]							165						170			175

[0194]	Asp	Tyr	Tyr	Arg	Glu	Ser	Ala	Lys	Arg	Asp	Ala	Tyr	Val	Ser	Asn	Val
[0195]				180					185				190			
[0196]	Lys	Glu	Ala	Phe	Asp	Lys	Leu	Tyr	Lys	Glu	Glu	Asp	Ile	Ala	Lys	Leu
[0197]				195				200					205			
[0198]	Val	Leu	Glu	Ile	Glu	Asn	Leu	Thr	Lys	Leu	Glu	Lys	Tyr	Lys	Ile	Arg
[0199]		210					215					220				
[0200]	Glu	Phe	Tyr	His	Glu	Ile	Ile	Gly	Arg	Lys	Asn	Asp	Lys	Glu	Asn	Phe
[0201]	225					230					235				240	
[0202]	Ala	Lys	Ile	Ile	Tyr	Glu	Glu	Ile	Gln	Asn	Val	Asn	Asn	Met	Lys	Glu
[0203]					245				250						255	
[0204]	Leu	Ile	Glu	Lys	Val	Pro	Asp	Met	Ser	Glu	Leu	Lys	Lys	Ser	Gln	Val
[0205]				260					265					270		
[0206]	Phe	Tyr	Lys	Tyr	Tyr	Leu	Asp	Lys	Glu	Glu	Leu	Asn	Asp	Lys	Asn	Ile
[0207]			275					280					285			
[0208]	Lys	Tyr	Ala	Phe	Cys	His	Phe	Val	Glu	Ile	Glu	Met	Ser	Gln	Leu	Leu
[0209]		290					295					300				
[0210]	Lys	Asn	Tyr	Val	Tyr	Lys	Arg	Leu	Ser	Asn	Ile	Ser	Asn	Asp	Lys	Ile
[0211]	305					310					315				320	
[0212]	Lys	Arg	Ile	Phe	Glu	Tyr	Gln	Asn	Leu	Lys	Lys	Leu	Ile	Glu	Asn	Lys
[0213]				325					330						335	
[0214]	Leu	Leu	Asn	Lys	Leu	Asp	Thr	Tyr	Val	Arg	Asn	Cys	Gly	Lys	Tyr	Asn
[0215]				340					345					350		
[0216]	Tyr	Tyr	Leu	Gln	Asp	Gly	Glu	Ile	Ala	Thr	Ser	Asp	Phe	Ile	Ala	Arg
[0217]			355					360					365			
[0218]	Asn	Arg	Gln	Asn	Glu	Ala	Phe	Leu	Arg	Asn	Ile	Ile	Gly	Val	Ser	Ser
[0219]		370					375					380				
[0220]	Val	Ala	Tyr	Phe	Ser	Leu	Arg	Asn	Ile	Leu	Glu	Thr	Glu	Asn	Glu	Asn
[0221]	385					390					395				400	
[0222]	Asp	Ile	Thr	Gly	Arg	Met	Arg	Gly	Lys	Thr	Val	Lys	Asn	Asn	Lys	Gly
[0223]				405					410						415	
[0224]	Glu	Glu	Lys	Tyr	Val	Ser	Gly	Glu	Val	Asp	Lys	Ile	Tyr	Asn	Glu	Asn
[0225]				420					425				430			
[0226]	Lys	Lys	Asn	Glu	Val	Lys	Glu	Asn	Leu	Lys	Met	Phe	Tyr	Ser	Tyr	Asp
[0227]			435					440					445			
[0228]	Phe	Asn	Met	Asp	Asn	Lys	Asn	Glu	Ile	Glu	Asp	Phe	Phe	Ala	Asn	Ile
[0229]		450						455				460				
[0230]	Asp	Glu	Ala	Ile	Ser	Ser	Ile	Arg	His	Gly	Ile	Val	His	Phe	Asn	Leu
[0231]	465					470					475				480	
[0232]	Glu	Leu	Glu	Gly	Lys	Asp	Ile	Phe	Ala	Phe	Lys	Asn	Ile	Ala	Pro	Ser

[0233]		485		490		495
[0234]	Glu Ile Ser Lys Lys Met Phe Gln Asn Glu Ile Asn Glu Lys Lys Leu					
[0235]		500		505		510
[0236]	Lys Leu Lys Ile Phe Arg Gln Leu Asn Ser Ala Asn Val Phe Arg Tyr					
[0237]		515		520		525
[0238]	Leu Glu Lys Tyr Lys Ile Leu Asn Tyr Leu Lys Arg Thr Arg Phe Glu					
[0239]		530		535		540
[0240]	Phe Val Asn Lys Asn Ile Pro Phe Val Pro Ser Phe Thr Lys Leu Tyr					
[0241]		545		550		555
[0242]	Ser Arg Ile Asp Asp Leu Lys Asn Ser Leu Gly Ile Tyr Trp Lys Thr					
[0243]		565		570		575
[0244]	Pro Lys Thr Asn Asp Asp Asn Lys Thr Lys Glu Ile Ile Asp Ala Gln					
[0245]		580		585		590
[0246]	Ile Tyr Leu Leu Lys Asn Ile Tyr Tyr Gly Glu Phe Leu Asn Tyr Phe					
[0247]		595		600		605
[0248]	Met Ser Asn Asn Gly Asn Phe Phe Glu Ile Ser Lys Glu Ile Ile Glu					
[0249]		610		615		620
[0250]	Leu Asn Lys Asn Asp Lys Arg Asn Leu Lys Thr Gly Phe Tyr Lys Leu					
[0251]		625		630		635
[0252]	Gln Lys Phe Glu Asp Ile Gln Glu Lys Ile Pro Lys Glu Tyr Leu Ala					
[0253]		645		650		655
[0254]	Asn Ile Gln Ser Leu Tyr Met Ile Asn Ala Gly Asn Gln Asp Glu Glu					
[0255]		660		665		670
[0256]	Glu Lys Asp Thr Tyr Ile Asp Phe Ile Gln Lys Ile Phe Leu Lys Gly					
[0257]		675		680		685
[0258]	Phe Met Thr Tyr Leu Ala Asn Asn Gly Arg Leu Ser Leu Ile Tyr Ile					
[0259]		690		695		700
[0260]	Gly Ser Asp Glu Glu Thr Asn Thr Ser Leu Ala Glu Lys Lys Gln Glu					
[0261]		705		710		715
[0262]	Phe Asp Lys Phe Leu Lys Lys Tyr Glu Gln Asn Asn Asn Ile Lys Ile					
[0263]		725		730		735
[0264]	Pro Tyr Glu Ile Asn Glu Phe Leu Arg Glu Ile Lys Leu Gly Asn Ile					
[0265]		740		745		750
[0266]	Leu Lys Tyr Thr Glu Arg Leu Asn Met Phe Tyr Leu Ile Leu Lys Leu					
[0267]		755		760		765
[0268]	Leu Asn His Lys Glu Leu Thr Asn Leu Lys Gly Ser Leu Glu Lys Tyr					
[0269]		770		775		780
[0270]	Gln Ser Ala Asn Lys Glu Glu Ala Phe Ser Asp Gln Leu Glu Leu Ile					
[0271]		785		790		795
						800

[0272]	Asn	Leu	Leu	Asn	Leu	Asp	Asn	Asn	Arg	Val	Thr	Glu	Asp	Phe	Glu	Leu
[0273]						805				810					815	
[0274]	Glu	Ala	Asp	Glu	Ile	Gly	Lys	Phe	Leu	Asp	Phe	Asn	Gly	Asn	Lys	Val
[0275]						820				825					830	
[0276]	Lys	Asp	Asn	Lys	Glu	Leu	Lys	Lys	Phe	Asp	Thr	Asn	Lys	Ile	Tyr	Phe
[0277]						835				840					845	
[0278]	Asp	Gly	Glu	Asn	Ile	Ile	Lys	His	Arg	Ala	Phe	Tyr	Asn	Ile	Lys	Lys
[0279]						850				855					860	
[0280]	Tyr	Gly	Met	Leu	Asn	Leu	Leu	Glu	Lys	Ile	Ala	Asp	Lys	Ala	Gly	Tyr
[0281]	865									870					875	
[0282]	Lys	Ile	Ser	Ile	Glu	Glu	Leu	Lys	Lys	Tyr	Ser	Asn	Lys	Lys	Asn	Glu
[0283]						885				890					895	
[0284]	Ile	Glu	Lys	Asn	His	Lys	Met	Gln	Glu	Asn	Leu	His	Arg	Lys	Tyr	Ala
[0285]						900				905					910	
[0286]	Arg	Pro	Arg	Lys	Asp	Glu	Lys	Phe	Thr	Asp	Glu	Asp	Tyr	Glu	Ser	Tyr
[0287]						915				920					925	
[0288]	Lys	Gln	Ala	Ile	Glu	Asn	Ile	Glu	Glu	Tyr	Thr	His	Leu	Lys	Asn	Lys
[0289]						930				935					940	
[0290]	Val	Glu	Phe	Asn	Glu	Leu	Asn	Leu	Leu	Gln	Gly	Leu	Leu	Leu	Arg	Ile
[0291]	945									950					955	
[0292]	Leu	His	Arg	Leu	Val	Gly	Tyr	Thr	Ser	Ile	Trp	Glu	Arg	Asp	Leu	Arg
[0293]						965				970					975	
[0294]	Phe	Arg	Leu	Lys	Gly	Glu	Phe	Pro	Glu	Asn	Gln	Tyr	Ile	Glu	Glu	Ile
[0295]						980				985					990	
[0296]	Phe	Asn	Phe	Glu	Asn	Lys	Lys	Asn	Val	Lys	Tyr	Lys	Gly	Gly	Gln	Ile
[0297]						995				1000					1005	
[0298]	Val	Glu	Lys	Tyr	Ile	Lys	Phe	Tyr	Lys	Glu	Leu	His	Gln	Asn	Asp	
[0299]						1010				1015					1020	
[0300]	Glu	Val	Lys	Ile	Asn	Lys	Tyr	Ser	Ser	Ala	Asn	Ile	Lys	Val	Leu	
[0301]						1025				1030					1035	
[0302]	Lys	Gln	Glu	Lys	Lys	Asp	Leu	Tyr	Ile	Arg	Asn	Tyr	Ile	Ala	His	
[0303]						1040				1045					1050	
[0304]	Phe	Asn	Tyr	Ile	Pro	His	Ala	Glu	Ile	Ser	Leu	Leu	Glu	Val	Leu	
[0305]						1055				1060					1065	
[0306]	Glu	Asn	Leu	Arg	Lys	Leu	Leu	Ser	Tyr	Asp	Arg	Lys	Leu	Lys	Asn	
[0307]						1070				1075					1080	
[0308]	Ala	Val	Met	Lys	Ser	Val	Val	Asp	Ile	Leu	Lys	Glu	Tyr	Gly	Phe	
[0309]						1085				1090					1095	
[0310]	Val	Ala	Thr	Phe	Lys	Ile	Gly	Ala	Asp	Lys	Lys	Ile	Gly	Ile	Gln	

[0311]	1100	1105	1110
[0312]	Thr Leu Glu Ser Glu Lys Ile Val His Leu Lys Asn Leu Lys Lys		
[0313]	1115	1120	1125
[0314]	Lys Lys Leu Met Thr Asp Arg Asn Ser Glu Glu Leu Cys Lys Leu		
[0315]	1130	1135	1140
[0316]	Val Lys Ile Met Phe Glu Tyr Lys Met Glu Glu Lys Lys Ser Glu		
[0317]	1145	1150	1155
[0318]	Asn		
[0319]	<210> 3		
[0320]	<211> 1389		
[0321]	<212> PRT		
[0322]	<213> 沙氏纤毛菌(Leptotrichia shahii)		
[0323]	<400> 3		
[0324]	Met Gly Asn Leu Phe Gly His Lys Arg Trp Tyr Glu Val Arg Asp Lys		
[0325]	1 5 10 15		
[0326]	Lys Asp Phe Lys Ile Lys Arg Lys Val Lys Val Lys Arg Asn Tyr Asp		
[0327]	20 25 30		
[0328]	Gly Asn Lys Tyr Ile Leu Asn Ile Asn Glu Asn Asn Asn Lys Glu Lys		
[0329]	35 40 45		
[0330]	Ile Asp Asn Asn Lys Phe Ile Arg Lys Tyr Ile Asn Tyr Lys Lys Asn		
[0331]	50 55 60		
[0332]	Asp Asn Ile Leu Lys Glu Phe Thr Arg Lys Phe His Ala Gly Asn Ile		
[0333]	65 70 75 80		
[0334]	Leu Phe Lys Leu Lys Gly Lys Glu Gly Ile Ile Arg Ile Glu Asn Asn		
[0335]	85 90 95		
[0336]	Asp Asp Phe Leu Glu Thr Glu Glu Val Val Leu Tyr Ile Glu Ala Tyr		
[0337]	100 105 110		
[0338]	Gly Lys Ser Glu Lys Leu Lys Ala Leu Gly Ile Thr Lys Lys Lys Ile		
[0339]	115 120 125		
[0340]	Ile Asp Glu Ala Ile Arg Gln Gly Ile Thr Lys Asp Asp Lys Lys Ile		
[0341]	130 135 140		
[0342]	Glu Ile Lys Arg Gln Glu Asn Glu Glu Glu Ile Glu Ile Asp Ile Arg		
[0343]	145 150 155 160		
[0344]	Asp Glu Tyr Thr Asn Lys Thr Leu Asn Asp Cys Ser Ile Ile Leu Arg		
[0345]	165 170 175		
[0346]	Ile Ile Glu Asn Asp Glu Leu Glu Thr Lys Lys Ser Ile Tyr Glu Ile		
[0347]	180 185 190		
[0348]	Phe Lys Asn Ile Asn Met Ser Leu Tyr Lys Ile Ile Glu Lys Ile Ile		
[0349]	195 200 205		

[0350]	Glu Asn Glu Thr Glu Lys Val Phe Glu Asn Arg Tyr Tyr Glu Glu His		
[0351]	210	215	220
[0352]	Leu Arg Glu Lys Leu Leu Lys Asp Asp Lys Ile Asp Val Ile Leu Thr		
[0353]	225	230	235 240
[0354]	Asn Phe Met Glu Ile Arg Glu Lys Ile Lys Ser Asn Leu Glu Ile Leu		
[0355]	245	250	255
[0356]	Gly Phe Val Lys Phe Tyr Leu Asn Val Gly Gly Asp Lys Lys Lys Ser		
[0357]	260	265	270
[0358]	Lys Asn Lys Lys Met Leu Val Glu Lys Ile Leu Asn Ile Asn Val Asp		
[0359]	275	280	285
[0360]	Leu Thr Val Glu Asp Ile Ala Asp Phe Val Ile Lys Glu Leu Glu Phe		
[0361]	290	295	300
[0362]	Trp Asn Ile Thr Lys Arg Ile Glu Lys Val Lys Lys Val Asn Asn Glu		
[0363]	305	310	315 320
[0364]	Phe Leu Glu Lys Arg Arg Asn Arg Thr Tyr Ile Lys Ser Tyr Val Leu		
[0365]	325	330	335
[0366]	Leu Asp Lys His Glu Lys Phe Lys Ile Glu Arg Glu Asn Lys Lys Asp		
[0367]	340	345	350
[0368]	Lys Ile Val Lys Phe Phe Val Glu Asn Ile Lys Asn Asn Ser Ile Lys		
[0369]	355	360	365
[0370]	Glu Lys Ile Glu Lys Ile Leu Ala Glu Phe Lys Ile Asp Glu Leu Ile		
[0371]	370	375	380
[0372]	Lys Lys Leu Glu Lys Glu Leu Lys Lys Gly Asn Cys Asp Thr Glu Ile		
[0373]	385	390	395 400
[0374]	Phe Gly Ile Phe Lys Lys His Tyr Lys Val Asn Phe Asp Ser Lys Lys		
[0375]	405	410	415
[0376]	Phe Ser Lys Lys Ser Asp Glu Glu Lys Glu Leu Tyr Lys Ile Ile Tyr		
[0377]	420	425	430
[0378]	Arg Tyr Leu Lys Gly Arg Ile Glu Lys Ile Leu Val Asn Glu Gln Lys		
[0379]	435	440	445
[0380]	Val Arg Leu Lys Lys Met Glu Lys Ile Glu Ile Glu Lys Ile Leu Asn		
[0381]	450	455	460
[0382]	Glu Ser Ile Leu Ser Glu Lys Ile Leu Lys Arg Val Lys Gln Tyr Thr		
[0383]	465	470	475 480
[0384]	Leu Glu His Ile Met Tyr Leu Gly Lys Leu Arg His Asn Asp Ile Asp		
[0385]	485	490	495
[0386]	Met Thr Thr Val Asn Thr Asp Asp Phe Ser Arg Leu His Ala Lys Glu		
[0387]	500	505	510
[0388]	Glu Leu Asp Leu Glu Leu Ile Thr Phe Phe Ala Ser Thr Asn Met Glu		

[0389]	515	520	525
[0390]	Leu Asn Lys Ile Phe Ser Arg Glu Asn Ile Asn Asn Asp Glu Asn Ile		
[0391]	530	535	540
[0392]	Asp Phe Phe Gly Gly Asp Arg Glu Lys Asn Tyr Val Leu Asp Lys Lys		
[0393]	545	550	555
[0394]	Ile Leu Asn Ser Lys Ile Lys Ile Ile Arg Asp Leu Asp Phe Ile Asp		
[0395]	565	570	575
[0396]	Asn Lys Asn Asn Ile Thr Asn Asn Phe Ile Arg Lys Phe Thr Lys Ile		
[0397]	580	585	590
[0398]	Gly Thr Asn Glu Arg Asn Arg Ile Leu His Ala Ile Ser Lys Glu Arg		
[0399]	595	600	605
[0400]	Asp Leu Gln Gly Thr Gln Asp Asp Tyr Asn Lys Val Ile Asn Ile Ile		
[0401]	610	615	620
[0402]	Gln Asn Leu Lys Ile Ser Asp Glu Glu Val Ser Lys Ala Leu Asn Leu		
[0403]	625	630	635
[0404]	Asp Val Val Phe Lys Asp Lys Lys Asn Ile Ile Thr Lys Ile Asn Asp		
[0405]	645	650	655
[0406]	Ile Lys Ile Ser Glu Glu Asn Asn Asn Asp Ile Lys Tyr Leu Pro Ser		
[0407]	660	665	670
[0408]	Phe Ser Lys Val Leu Pro Glu Ile Leu Asn Leu Tyr Arg Asn Asn Pro		
[0409]	675	680	685
[0410]	Lys Asn Glu Pro Phe Asp Thr Ile Glu Thr Glu Lys Ile Val Leu Asn		
[0411]	690	695	700
[0412]	Ala Leu Ile Tyr Val Asn Lys Glu Leu Tyr Lys Lys Leu Ile Leu Glu		
[0413]	705	710	715
[0414]	Asp Asp Leu Glu Glu Asn Glu Ser Lys Asn Ile Phe Leu Gln Glu Leu		
[0415]	725	730	735
[0416]	Lys Lys Thr Leu Gly Asn Ile Asp Glu Ile Asp Glu Asn Ile Ile Glu		
[0417]	740	745	750
[0418]	Asn Tyr Tyr Lys Asn Ala Gln Ile Ser Ala Ser Lys Gly Asn Asn Lys		
[0419]	755	760	765
[0420]	Ala Ile Lys Lys Tyr Gln Lys Lys Val Ile Glu Cys Tyr Ile Gly Tyr		
[0421]	770	775	780
[0422]	Leu Arg Lys Asn Tyr Glu Glu Leu Phe Asp Phe Ser Asp Phe Lys Met		
[0423]	785	790	795
[0424]	Asn Ile Gln Glu Ile Lys Lys Gln Ile Lys Asp Ile Asn Asp Asn Lys		
[0425]	805	810	815
[0426]	Thr Tyr Glu Arg Ile Thr Val Lys Thr Ser Asp Lys Thr Ile Val Ile		
[0427]	820	825	830

[0428]	Asn Asp Asp Phe Glu Tyr Ile Ile Ser Ile Phe Ala Leu Leu Asn Ser
[0429]	835 840 845
[0430]	Asn Ala Val Ile Asn Lys Ile Arg Asn Arg Phe Phe Ala Thr Ser Val
[0431]	850 855 860
[0432]	Trp Leu Asn Thr Ser Glu Tyr Gln Asn Ile Ile Asp Ile Leu Asp Glu
[0433]	865 870 875 880
[0434]	Ile Met Gln Leu Asn Thr Leu Arg Asn Glu Cys Ile Thr Glu Asn Trp
[0435]	885 890 895
[0436]	Asn Leu Asn Leu Glu Glu Phe Ile Gln Lys Met Lys Glu Ile Glu Lys
[0437]	900 905 910
[0438]	Asp Phe Asp Asp Phe Lys Ile Gln Thr Lys Lys Glu Ile Phe Asn Asn
[0439]	915 920 925
[0440]	Tyr Tyr Glu Asp Ile Lys Asn Asn Ile Leu Thr Glu Phe Lys Asp Asp
[0441]	930 935 940
[0442]	Ile Asn Gly Cys Asp Val Leu Glu Lys Lys Leu Glu Lys Ile Val Ile
[0443]	945 950 955 960
[0444]	Phe Asp Asp Glu Thr Lys Phe Glu Ile Asp Lys Lys Ser Asn Ile Leu
[0445]	965 970 975
[0446]	Gln Asp Glu Gln Arg Lys Leu Ser Asn Ile Asn Lys Lys Asp Leu Lys
[0447]	980 985 990
[0448]	Lys Lys Val Asp Gln Tyr Ile Lys Asp Lys Asp Gln Glu Ile Lys Ser
[0449]	995 1000 1005
[0450]	Lys Ile Leu Cys Arg Ile Ile Phe Asn Ser Asp Phe Leu Lys Lys
[0451]	1010 1015 1020
[0452]	Tyr Lys Lys Glu Ile Asp Asn Leu Ile Glu Asp Met Glu Ser Glu
[0453]	1025 1030 1035
[0454]	Asn Glu Asn Lys Phe Gln Glu Ile Tyr Tyr Pro Lys Glu Arg Lys
[0455]	1040 1045 1050
[0456]	Asn Glu Leu Tyr Ile Tyr Lys Lys Asn Leu Phe Leu Asn Ile Gly
[0457]	1055 1060 1065
[0458]	Asn Pro Asn Phe Asp Lys Ile Tyr Gly Leu Ile Ser Asn Asp Ile
[0459]	1070 1075 1080
[0460]	Lys Met Ala Asp Ala Lys Phe Leu Phe Asn Ile Asp Gly Lys Asn
[0461]	1085 1090 1095
[0462]	Ile Arg Lys Asn Lys Ile Ser Glu Ile Asp Ala Ile Leu Lys Asn
[0463]	1100 1105 1110
[0464]	Leu Asn Asp Lys Leu Asn Gly Tyr Ser Lys Glu Tyr Lys Glu Lys
[0465]	1115 1120 1125
[0466]	Tyr Ile Lys Lys Leu Lys Glu Asn Asp Asp Phe Phe Ala Lys Asn

[0467]	1130	1135	1140
[0468]	Ile Gln Asn Lys Asn Tyr Lys Ser Phe Glu Lys Asp Tyr Asn Arg		
[0469]	1145	1150	1155
[0470]	Val Ser Glu Tyr Lys Lys Ile Arg Asp Leu Val Glu Phe Asn Tyr		
[0471]	1160	1165	1170
[0472]	Leu Asn Lys Ile Glu Ser Tyr Leu Ile Asp Ile Asn Trp Lys Leu		
[0473]	1175	1180	1185
[0474]	Ala Ile Gln Met Ala Arg Phe Glu Arg Asp Met His Tyr Ile Val		
[0475]	1190	1195	1200
[0476]	Asn Gly Leu Arg Glu Leu Gly Ile Ile Lys Leu Ser Gly Tyr Asn		
[0477]	1205	1210	1215
[0478]	Thr Gly Ile Ser Arg Ala Tyr Pro Lys Arg Asn Gly Ser Asp Gly		
[0479]	1220	1225	1230
[0480]	Phe Tyr Thr Thr Thr Ala Tyr Tyr Lys Phe Phe Asp Glu Glu Ser		
[0481]	1235	1240	1245
[0482]	Tyr Lys Lys Phe Glu Lys Ile Cys Tyr Gly Phe Gly Ile Asp Leu		
[0483]	1250	1255	1260
[0484]	Ser Glu Asn Ser Glu Ile Asn Lys Pro Glu Asn Glu Ser Ile Arg		
[0485]	1265	1270	1275
[0486]	Asn Tyr Ile Ser His Phe Tyr Ile Val Arg Asn Pro Phe Ala Asp		
[0487]	1280	1285	1290
[0488]	Tyr Ser Ile Ala Glu Gln Ile Asp Arg Val Ser Asn Leu Leu Ser		
[0489]	1295	1300	1305
[0490]	Tyr Ser Thr Arg Tyr Asn Asn Ser Thr Tyr Ala Ser Val Phe Glu		
[0491]	1310	1315	1320
[0492]	Val Phe Lys Lys Asp Val Asn Leu Asp Tyr Asp Glu Leu Lys Lys		
[0493]	1325	1330	1335
[0494]	Lys Phe Lys Leu Ile Gly Asn Asn Asp Ile Leu Glu Arg Leu Met		
[0495]	1340	1345	1350
[0496]	Lys Pro Lys Lys Val Ser Val Leu Glu Leu Glu Ser Tyr Asn Ser		
[0497]	1355	1360	1365
[0498]	Asp Tyr Ile Lys Asn Leu Ile Ile Glu Leu Leu Thr Lys Ile Glu		
[0499]	1370	1375	1380
[0500]	Asn Thr Asn Asp Thr Leu		
[0501]	1385		
[0502]	<210> 4		
[0503]	<211> 1285		
[0504]	<212> PRT		
[0505]	<213> 荚膜红细菌 (Rhodobacter capsulatus) R121		

[0506]	<400> 4															
[0507]	Met	Gln	Ile	Gly	Lys	Val	Gln	Gly	Arg	Thr	Ile	Ser	Glu	Phe	Gly	Asp
[0508]	1				5				10				15			
[0509]	Pro	Ala	Gly	Gly	Leu	Lys	Arg	Lys	Ile	Ser	Thr	Asp	Gly	Lys	Asn	Arg
[0510]	20				25				30							
[0511]	Lys	Glu	Leu	Pro	Ala	His	Leu	Ser	Ser	Asp	Pro	Lys	Ala	Leu	Ile	Gly
[0512]	35				40				45							
[0513]	Gln	Trp	Ile	Ser	Gly	Ile	Asp	Lys	Ile	Tyr	Arg	Lys	Pro	Asp	Ser	Arg
[0514]	50				55				60							
[0515]	Lys	Ser	Asp	Gly	Lys	Ala	Ile	His	Ser	Pro	Thr	Pro	Ser	Lys	Met	Gln
[0516]	65				70				75				80			
[0517]	Phe	Asp	Ala	Arg	Asp	Asp	Leu	Gly	Glu	Ala	Phe	Trp	Lys	Leu	Val	Ser
[0518]	85				90				95							
[0519]	Glu	Ala	Gly	Leu	Ala	Gln	Asp	Ser	Asp	Tyr	Asp	Gln	Phe	Lys	Arg	Arg
[0520]	100				105				110							
[0521]	Leu	His	Pro	Tyr	Gly	Asp	Lys	Phe	Gln	Pro	Ala	Asp	Ser	Gly	Ala	Lys
[0522]	115				120				125							
[0523]	Leu	Lys	Phe	Glu	Ala	Asp	Pro	Pro	Glu	Pro	Gln	Ala	Phe	His	Gly	Arg
[0524]	130				135				140							
[0525]	Trp	Tyr	Gly	Ala	Met	Ser	Lys	Arg	Gly	Asn	Asp	Ala	Lys	Glu	Leu	Ala
[0526]	145				150				155				160			
[0527]	Ala	Ala	Leu	Tyr	Glu	His	Leu	His	Val	Asp	Glu	Lys	Arg	Ile	Asp	Gly
[0528]	165				170				175							
[0529]	Gln	Pro	Lys	Arg	Asn	Pro	Lys	Thr	Asp	Lys	Phe	Ala	Pro	Gly	Leu	Val
[0530]	180				185				190							
[0531]	Val	Ala	Arg	Ala	Leu	Gly	Ile	Glu	Ser	Ser	Val	Leu	Pro	Arg	Gly	Met
[0532]	195				200				205							
[0533]	Ala	Arg	Leu	Ala	Arg	Asn	Trp	Gly	Glu	Glu	Glu	Ile	Gln	Thr	Tyr	Phe
[0534]	210				215				220							
[0535]	Val	Val	Asp	Val	Ala	Ala	Ser	Val	Lys	Glu	Val	Ala	Lys	Ala	Ala	Val
[0536]	225				230				235				240			
[0537]	Ser	Ala	Ala	Gln	Ala	Phe	Asp	Pro	Pro	Arg	Gln	Val	Ser	Gly	Arg	Ser
[0538]	245				250				255							
[0539]	Leu	Ser	Pro	Lys	Val	Gly	Phe	Ala	Leu	Ala	Glu	His	Leu	Glu	Arg	Val
[0540]	260				265				270							
[0541]	Thr	Gly	Ser	Lys	Arg	Cys	Ser	Phe	Asp	Pro	Ala	Ala	Gly	Pro	Ser	Val
[0542]	275				280				285							
[0543]	Leu	Ala	Leu	His	Asp	Glu	Val	Lys	Lys	Thr	Tyr	Lys	Arg	Leu	Cys	Ala
[0544]	290				295				300							

[0545]	Arg Gly Lys Asn Ala Ala Arg Ala Phe Pro Ala Asp Lys Thr Glu Leu		
[0546]	305	310	315 320
[0547]	Leu Ala Leu Met Arg His Thr His Glu Asn Arg Val Arg Asn Gln Met		
[0548]		325 330	335
[0549]	Val Arg Met Gly Arg Val Ser Glu Tyr Arg Gly Gln Gln Ala Gly Asp		
[0550]		340 345	350
[0551]	Leu Ala Gln Ser His Tyr Trp Thr Ser Ala Gly Gln Thr Glu Ile Lys		
[0552]		355 360	365
[0553]	Glu Ser Glu Ile Phe Val Arg Leu Trp Val Gly Ala Phe Ala Leu Ala		
[0554]		370 375	380
[0555]	Gly Arg Ser Met Lys Ala Trp Ile Asp Pro Met Gly Lys Ile Val Asn		
[0556]	385	390	395 400
[0557]	Thr Glu Lys Asn Asp Arg Asp Leu Thr Ala Ala Val Asn Ile Arg Gln		
[0558]		405 410	415
[0559]	Val Ile Ser Asn Lys Glu Met Val Ala Glu Ala Met Ala Arg Arg Gly		
[0560]		420 425	430
[0561]	Ile Tyr Phe Gly Glu Thr Pro Glu Leu Asp Arg Leu Gly Ala Glu Gly		
[0562]		435 440	445
[0563]	Asn Glu Gly Phe Val Phe Ala Leu Leu Arg Tyr Leu Arg Gly Cys Arg		
[0564]		450 455	460
[0565]	Asn Gln Thr Phe His Leu Gly Ala Arg Ala Gly Phe Leu Lys Glu Ile		
[0566]	465	470	475 480
[0567]	Arg Lys Glu Leu Glu Lys Thr Arg Trp Gly Lys Ala Lys Glu Ala Glu		
[0568]		485 490	495
[0569]	His Val Val Leu Thr Asp Lys Thr Val Ala Ala Ile Arg Ala Ile Ile		
[0570]		500 505	510
[0571]	Asp Asn Asp Ala Lys Ala Leu Gly Ala Arg Leu Leu Ala Asp Leu Ser		
[0572]		515 520	525
[0573]	Gly Ala Phe Val Ala His Tyr Ala Ser Lys Glu His Phe Ser Thr Leu		
[0574]		530 535	540
[0575]	Tyr Ser Glu Ile Val Lys Ala Val Lys Asp Ala Pro Glu Val Ser Ser		
[0576]	545	550	555 560
[0577]	Gly Leu Pro Arg Leu Lys Leu Leu Leu Lys Arg Ala Asp Gly Val Arg		
[0578]		565 570	575
[0579]	Gly Tyr Val His Gly Leu Arg Asp Thr Arg Lys His Ala Phe Ala Thr		
[0580]		580 585	590
[0581]	Lys Leu Pro Pro Pro Pro Ala Pro Arg Glu Leu Asp Asp Pro Ala Thr		
[0582]		595 600	605
[0583]	Lys Ala Arg Tyr Ile Ala Leu Leu Arg Leu Tyr Asp Gly Pro Phe Arg		

[0584]	610	615	620
[0585]	Ala Tyr Ala Ser Gly Ile Thr Gly Thr Ala Leu Ala Gly Pro Ala Ala		
[0586]	625	630	635 640
[0587]	Arg Ala Lys Glu Ala Ala Thr Ala Leu Ala Gln Ser Val Asn Val Thr		
[0588]	645	650	655
[0589]	Lys Ala Tyr Ser Asp Val Met Glu Gly Arg Ser Ser Arg Leu Arg Pro		
[0590]	660	665	670
[0591]	Pro Asn Asp Gly Glu Thr Leu Arg Glu Tyr Leu Ser Ala Leu Thr Gly		
[0592]	675	680	685
[0593]	Glu Thr Ala Thr Glu Phe Arg Val Gln Ile Gly Tyr Glu Ser Asp Ser		
[0594]	690	695	700
[0595]	Glu Asn Ala Arg Lys Gln Ala Glu Phe Ile Glu Asn Tyr Arg Arg Asp		
[0596]	705	710	715 720
[0597]	Met Leu Ala Phe Met Phe Glu Asp Tyr Ile Arg Ala Lys Gly Phe Asp		
[0598]	725	730	735
[0599]	Trp Ile Leu Lys Ile Glu Pro Gly Ala Thr Ala Met Thr Arg Ala Pro		
[0600]	740	745	750
[0601]	Val Leu Pro Glu Pro Ile Asp Thr Arg Gly Gln Tyr Glu His Trp Gln		
[0602]	755	760	765
[0603]	Ala Ala Leu Tyr Leu Val Met His Phe Val Pro Ala Ser Asp Val Ser		
[0604]	770	775	780
[0605]	Asn Leu Leu His Gln Leu Arg Lys Trp Glu Ala Leu Gln Gly Lys Tyr		
[0606]	785	790	795 800
[0607]	Glu Leu Val Gln Asp Gly Asp Ala Thr Asp Gln Ala Asp Ala Arg Arg		
[0608]	805	810	815
[0609]	Glu Ala Leu Asp Leu Val Lys Arg Phe Arg Asp Val Leu Val Leu Phe		
[0610]	820	825	830
[0611]	Leu Lys Thr Gly Glu Ala Arg Phe Glu Gly Arg Ala Ala Pro Phe Asp		
[0612]	835	840	845
[0613]	Leu Lys Pro Phe Arg Ala Leu Phe Ala Asn Pro Ala Thr Phe Asp Arg		
[0614]	850	855	860
[0615]	Leu Phe Met Ala Thr Pro Thr Thr Ala Arg Pro Ala Glu Asp Asp Pro		
[0616]	865	870	875 880
[0617]	Glu Gly Asp Gly Ala Ser Glu Pro Glu Leu Arg Val Ala Arg Thr Leu		
[0618]	885	890	895
[0619]	Arg Gly Leu Arg Gln Ile Ala Arg Tyr Asn His Met Ala Val Leu Ser		
[0620]	900	905	910
[0621]	Asp Leu Phe Ala Lys His Lys Val Arg Asp Glu Glu Val Ala Arg Leu		
[0622]	915	920	925

[0623]	Ala Glu Ile Glu Asp Glu Thr Gln Glu Lys Ser Gln Ile Val Ala Ala
[0624]	930 935 940
[0625]	Gln Glu Leu Arg Thr Asp Leu His Asp Lys Val Met Lys Cys His Pro
[0626]	945 950 955 960
[0627]	Lys Thr Ile Ser Pro Glu Glu Arg Gln Ser Tyr Ala Ala Ala Ile Lys
[0628]	965 970 975
[0629]	Thr Ile Glu Glu His Arg Phe Leu Val Gly Arg Val Tyr Leu Gly Asp
[0630]	980 985 990
[0631]	His Leu Arg Leu His Arg Leu Met Met Asp Val Ile Gly Arg Leu Ile
[0632]	995 1000 1005
[0633]	Asp Tyr Ala Gly Ala Tyr Glu Arg Asp Thr Gly Thr Phe Leu Ile
[0634]	1010 1015 1020
[0635]	Asn Ala Ser Lys Gln Leu Gly Ala Gly Ala Asp Trp Ala Val Thr
[0636]	1025 1030 1035
[0637]	Ile Ala Gly Ala Ala Asn Thr Asp Ala Arg Thr Gln Thr Arg Lys
[0638]	1040 1045 1050
[0639]	Asp Leu Ala His Phe Asn Val Leu Asp Arg Ala Asp Gly Thr Pro
[0640]	1055 1060 1065
[0641]	Asp Leu Thr Ala Leu Val Asn Arg Ala Arg Glu Met Met Ala Tyr
[0642]	1070 1075 1080
[0643]	Asp Arg Lys Arg Lys Asn Ala Val Pro Arg Ser Ile Leu Asp Met
[0644]	1085 1090 1095
[0645]	Leu Ala Arg Leu Gly Leu Thr Leu Lys Trp Gln Met Lys Asp His
[0646]	1100 1105 1110
[0647]	Leu Leu Gln Asp Ala Thr Ile Thr Gln Ala Ala Ile Lys His Leu
[0648]	1115 1120 1125
[0649]	Asp Lys Val Arg Leu Thr Val Gly Gly Pro Ala Ala Val Thr Glu
[0650]	1130 1135 1140
[0651]	Ala Arg Phe Ser Gln Asp Tyr Leu Gln Met Val Ala Ala Val Phe
[0652]	1145 1150 1155
[0653]	Asn Gly Ser Val Gln Asn Pro Lys Pro Arg Arg Arg Asp Asp Gly
[0654]	1160 1165 1170
[0655]	Asp Ala Trp His Lys Pro Pro Lys Pro Ala Thr Ala Gln Ser Gln
[0656]	1175 1180 1185
[0657]	Pro Asp Gln Lys Pro Pro Asn Lys Ala Pro Ser Ala Gly Ser Arg
[0658]	1190 1195 1200
[0659]	Leu Pro Pro Pro Gln Val Gly Glu Val Tyr Glu Gly Val Val Val
[0660]	1205 1210 1215
[0661]	Lys Val Ile Asp Thr Gly Ser Leu Gly Phe Leu Ala Val Glu Gly

[0662]	1220	1225	1230
[0663]	Val Ala Gly Asn Ile Gly Leu His Ile Ser Arg Leu Arg Arg Ile		
[0664]	1235	1240	1245
[0665]	Arg Glu Asp Ala Ile Ile Val Gly Arg Arg Tyr Arg Phe Arg Val		
[0666]	1250	1255	1260
[0667]	Glu Ile Tyr Val Pro Pro Lys Ser Asn Thr Ser Lys Leu Asn Ala		
[0668]	1265	1270	1275
[0669]	Ala Asp Leu Val Arg Ile Asp		
[0670]	1280	1285	
[0671]	<210> 5		
[0672]	<211> 1175		
[0673]	<212> PRT		
[0674]	<213> 鸡肉杆菌 (Carnobacterium gallinarum)		
[0675]	<400> 5		
[0676]	Met Arg Ile Thr Lys Val Lys Ile Lys Leu Asp Asn Lys Leu Tyr Gln		
[0677]	1 5 10 15		
[0678]	Val Thr Met Gln Lys Glu Glu Lys Tyr Gly Thr Leu Lys Leu Asn Glu		
[0679]	20 25 30		
[0680]	Glu Ser Arg Lys Ser Thr Ala Glu Ile Leu Arg Leu Lys Lys Ala Ser		
[0681]	35 40 45		
[0682]	Phe Asn Lys Ser Phe His Ser Lys Thr Ile Asn Ser Gln Lys Glu Asn		
[0683]	50 55 60		
[0684]	Lys Asn Ala Thr Ile Lys Lys Asn Gly Asp Tyr Ile Ser Gln Ile Phe		
[0685]	65 70 75 80		
[0686]	Glu Lys Leu Val Gly Val Asp Thr Asn Lys Asn Ile Arg Lys Pro Lys		
[0687]	85 90 95		
[0688]	Met Ser Leu Thr Asp Leu Lys Asp Leu Pro Lys Lys Asp Leu Ala Leu		
[0689]	100 105 110		
[0690]	Phe Ile Lys Arg Lys Phe Lys Asn Asp Asp Ile Val Glu Ile Lys Asn		
[0691]	115 120 125		
[0692]	Leu Asp Leu Ile Ser Leu Phe Tyr Asn Ala Leu Gln Lys Val Pro Gly		
[0693]	130 135 140		
[0694]	Glu His Phe Thr Asp Glu Ser Trp Ala Asp Phe Cys Gln Glu Met Met		
[0695]	145 150 155 160		
[0696]	Pro Tyr Arg Glu Tyr Lys Asn Lys Phe Ile Glu Arg Lys Ile Ile Leu		
[0697]	165 170 175		
[0698]	Leu Ala Asn Ser Ile Glu Gln Asn Lys Gly Phe Ser Ile Asn Pro Glu		
[0699]	180 185 190		
[0700]	Thr Phe Ser Lys Arg Lys Arg Val Leu His Gln Trp Ala Ile Glu Val		

[0701]	195	200	205
[0702]	Gln Glu Arg Gly Asp Phe Ser Ile Leu Asp Glu Lys Leu Ser Lys Leu		
[0703]	210	215	220
[0704]	Ala Glu Ile Tyr Asn Phe Lys Lys Met Cys Lys Arg Val Gln Asp Glu		
[0705]	225	230	235
[0706]	Leu Asn Asp Leu Glu Lys Ser Met Lys Lys Gly Lys Asn Pro Glu Lys		
[0707]	245	250	255
[0708]	Glu Lys Glu Ala Tyr Lys Lys Gln Lys Asn Phe Lys Ile Lys Thr Ile		
[0709]	260	265	270
[0710]	Trp Lys Asp Tyr Pro Tyr Lys Thr His Ile Gly Leu Ile Glu Lys Ile		
[0711]	275	280	285
[0712]	Lys Glu Asn Glu Glu Leu Asn Gln Phe Asn Ile Glu Ile Gly Lys Tyr		
[0713]	290	295	300
[0714]	Phe Glu His Tyr Phe Pro Ile Lys Lys Glu Arg Cys Thr Glu Asp Glu		
[0715]	305	310	315
[0716]	Pro Tyr Tyr Leu Asn Ser Glu Thr Ile Ala Thr Thr Val Asn Tyr Gln		
[0717]	325	330	335
[0718]	Leu Lys Asn Ala Leu Ile Ser Tyr Leu Met Gln Ile Gly Lys Tyr Lys		
[0719]	340	345	350
[0720]	Gln Phe Gly Leu Glu Asn Gln Val Leu Asp Ser Lys Lys Leu Gln Glu		
[0721]	355	360	365
[0722]	Ile Gly Ile Tyr Glu Gly Phe Gln Thr Lys Phe Met Asp Ala Cys Val		
[0723]	370	375	380
[0724]	Phe Ala Thr Ser Ser Leu Lys Asn Ile Ile Glu Pro Met Arg Ser Gly		
[0725]	385	390	395
[0726]	Asp Ile Leu Gly Lys Arg Glu Phe Lys Glu Ala Ile Ala Thr Ser Ser		
[0727]	405	410	415
[0728]	Phe Val Asn Tyr His His Phe Phe Pro Tyr Phe Pro Phe Glu Leu Lys		
[0729]	420	425	430
[0730]	Gly Met Lys Asp Arg Glu Ser Glu Leu Ile Pro Phe Gly Glu Gln Thr		
[0731]	435	440	445
[0732]	Glu Ala Lys Gln Met Gln Asn Ile Trp Ala Leu Arg Gly Ser Val Gln		
[0733]	450	455	460
[0734]	Gln Ile Arg Asn Glu Ile Phe His Ser Phe Asp Lys Asn Gln Lys Phe		
[0735]	465	470	475
[0736]	Asn Leu Pro Gln Leu Asp Lys Ser Asn Phe Glu Phe Asp Ala Ser Glu		
[0737]	485	490	495
[0738]	Asn Ser Thr Gly Lys Ser Gln Ser Tyr Ile Glu Thr Asp Tyr Lys Phe		
[0739]	500	505	510

[0740]	Leu	Phe	Glu	Ala	Glu	Lys	Asn	Gln	Leu	Glu	Gln	Phe	Phe	Ile	Glu	Arg
[0741]			515					520					525			
[0742]	Ile	Lys	Ser	Ser	Gly	Ala	Leu	Glu	Tyr	Tyr	Pro	Leu	Lys	Ser	Leu	Glu
[0743]			530					535					540			
[0744]	Lys	Leu	Phe	Ala	Lys	Lys	Glu	Met	Lys	Phe	Ser	Leu	Gly	Ser	Gln	Val
[0745]	545						550				555					560
[0746]	Val	Ala	Phe	Ala	Pro	Ser	Tyr	Lys	Lys	Leu	Val	Lys	Lys	Gly	His	Ser
[0747]					565					570					575	
[0748]	Tyr	Gln	Thr	Ala	Thr	Glu	Gly	Thr	Ala	Asn	Tyr	Leu	Gly	Leu	Ser	Tyr
[0749]				580					585					590		
[0750]	Tyr	Asn	Arg	Tyr	Glu	Leu	Lys	Glu	Glu	Ser	Phe	Gln	Ala	Gln	Tyr	Tyr
[0751]			595					600					605			
[0752]	Leu	Leu	Lys	Leu	Ile	Tyr	Gln	Tyr	Val	Phe	Leu	Pro	Asn	Phe	Ser	Gln
[0753]		610						615					620			
[0754]	Gly	Asn	Ser	Pro	Ala	Phe	Arg	Glu	Thr	Val	Lys	Ala	Ile	Leu	Arg	Ile
[0755]	625						630					635				640
[0756]	Asn	Lys	Asp	Glu	Ala	Arg	Lys	Lys	Met	Lys	Lys	Asn	Lys	Lys	Phe	Leu
[0757]					645					650					655	
[0758]	Arg	Lys	Tyr	Ala	Phe	Glu	Gln	Val	Arg	Glu	Met	Glu	Phe	Lys	Glu	Thr
[0759]				660						665					670	
[0760]	Pro	Asp	Gln	Tyr	Met	Ser	Tyr	Leu	Gln	Ser	Glu	Met	Arg	Glu	Glu	Lys
[0761]			675						680					685		
[0762]	Val	Arg	Lys	Ala	Glu	Lys	Asn	Asp	Lys	Gly	Phe	Glu	Lys	Asn	Ile	Thr
[0763]		690						695					700			
[0764]	Met	Asn	Phe	Glu	Lys	Leu	Leu	Met	Gln	Ile	Phe	Val	Lys	Gly	Phe	Asp
[0765]	705						710					715				720
[0766]	Val	Phe	Leu	Thr	Thr	Phe	Ala	Gly	Lys	Glu	Leu	Leu	Leu	Ser	Ser	Glu
[0767]				725						730					735	
[0768]	Glu	Lys	Val	Ile	Lys	Glu	Thr	Glu	Ile	Ser	Leu	Ser	Lys	Lys	Ile	Asn
[0769]				740						745				750		
[0770]	Glu	Arg	Glu	Lys	Thr	Leu	Lys	Ala	Ser	Ile	Gln	Val	Glu	His	Gln	Leu
[0771]			755						760					765		
[0772]	Val	Ala	Thr	Asn	Ser	Ala	Ile	Ser	Tyr	Trp	Leu	Phe	Cys	Lys	Leu	Leu
[0773]		770							775					780		
[0774]	Asp	Ser	Arg	His	Leu	Asn	Glu	Leu	Arg	Asn	Glu	Met	Ile	Lys	Phe	Lys
[0775]	785						790					795				800
[0776]	Gln	Ser	Arg	Ile	Lys	Phe	Asn	His	Thr	Gln	His	Ala	Glu	Leu	Ile	Gln
[0777]					805					810					815	
[0778]	Asn	Leu	Leu	Pro	Ile	Val	Glu	Leu	Thr	Ile	Leu	Ser	Asn	Asp	Tyr	Asp

[0779]	820	825	830
[0780]	Glu Lys Asn Asp Ser Gln Asn Val Asp Val Ser Ala Tyr Phe Glu Asp		
[0781]	835	840	845
[0782]	Lys Ser Leu Tyr Glu Thr Ala Pro Tyr Val Gln Thr Asp Asp Arg Thr		
[0783]	850	855	860
[0784]	Arg Val Ser Phe Arg Pro Ile Leu Lys Leu Glu Lys Tyr His Thr Lys		
[0785]	865	870	875
[0786]	Ser Leu Ile Glu Ala Leu Leu Lys Asp Asn Pro Gln Phe Arg Val Ala		
[0787]	885	890	895
[0788]	Ala Thr Asp Ile Gln Glu Trp Met His Lys Arg Glu Glu Ile Gly Glu		
[0789]	900	905	910
[0790]	Leu Val Glu Lys Arg Lys Asn Leu His Thr Glu Trp Ala Glu Gly Gln		
[0791]	915	920	925
[0792]	Gln Thr Leu Gly Ala Glu Lys Arg Glu Glu Tyr Arg Asp Tyr Cys Lys		
[0793]	930	935	940
[0794]	Lys Ile Asp Arg Phe Asn Trp Lys Ala Asn Lys Val Thr Leu Thr Tyr		
[0795]	945	950	955
[0796]	Leu Ser Gln Leu His Tyr Leu Ile Thr Asp Leu Leu Gly Arg Met Val		
[0797]	965	970	975
[0798]	Gly Phe Ser Ala Leu Phe Glu Arg Asp Leu Val Tyr Phe Ser Arg Ser		
[0799]	980	985	990
[0800]	Phe Ser Glu Leu Gly Gly Glu Thr Tyr His Ile Ser Asp Tyr Lys Asn		
[0801]	995	1000	1005
[0802]	Leu Ser Gly Val Leu Arg Leu Asn Ala Glu Val Lys Pro Ile Lys		
[0803]	1010	1015	1020
[0804]	Ile Lys Asn Ile Lys Val Ile Asp Asn Glu Glu Asn Pro Tyr Lys		
[0805]	1025	1030	1035
[0806]	Gly Asn Glu Pro Glu Val Lys Pro Phe Leu Asp Arg Leu His Ala		
[0807]	1040	1045	1050
[0808]	Tyr Leu Glu Asn Val Ile Gly Ile Lys Ala Val His Gly Lys Ile		
[0809]	1055	1060	1065
[0810]	Arg Asn Gln Thr Ala His Leu Ser Val Leu Gln Leu Glu Leu Ser		
[0811]	1070	1075	1080
[0812]	Met Ile Glu Ser Met Asn Asn Leu Arg Asp Leu Met Ala Tyr Asp		
[0813]	1085	1090	1095
[0814]	Arg Lys Leu Lys Asn Ala Val Thr Lys Ser Met Ile Lys Ile Leu		
[0815]	1100	1105	1110
[0816]	Asp Lys His Gly Met Ile Leu Lys Leu Lys Ile Asp Glu Asn His		
[0817]	1115	1120	1125

[0818]	Lys Asn Phe Glu Ile Glu Ser Leu Ile Pro Lys Glu Ile Ile His
[0819]	1130 1135 1140
[0820]	Leu Lys Asp Lys Ala Ile Lys Thr Asn Gln Val Ser Glu Glu Tyr
[0821]	1145 1150 1155
[0822]	Cys Gln Leu Val Leu Ala Leu Leu Thr Thr Asn Pro Gly Asn Gln
[0823]	1160 1165 1170
[0824]	Leu Asn
[0825]	1175
[0826]	<210> 6
[0827]	<211> 1285
[0828]	<212> PRT
[0829]	<213> Herbinix hemicellulosilytica
[0830]	<400> 6
[0831]	Met Lys Leu Thr Arg Arg Arg Ile Ser Gly Asn Ser Val Asp Gln Lys
[0832]	1 5 10 15
[0833]	Ile Thr Ala Ala Phe Tyr Arg Asp Met Ser Gln Gly Leu Leu Tyr Tyr
[0834]	20 25 30
[0835]	Asp Ser Glu Asp Asn Asp Cys Thr Asp Lys Val Ile Glu Ser Met Asp
[0836]	35 40 45
[0837]	Phe Glu Arg Ser Trp Arg Gly Arg Ile Leu Lys Asn Gly Glu Asp Asp
[0838]	50 55 60
[0839]	Lys Asn Pro Phe Tyr Met Phe Val Lys Gly Leu Val Gly Ser Asn Asp
[0840]	65 70 75 80
[0841]	Lys Ile Val Cys Glu Pro Ile Asp Val Asp Ser Asp Pro Asp Asn Leu
[0842]	85 90 95
[0843]	Asp Ile Leu Ile Asn Lys Asn Leu Thr Gly Phe Gly Arg Asn Leu Lys
[0844]	100 105 110
[0845]	Ala Pro Asp Ser Asn Asp Thr Leu Glu Asn Leu Ile Arg Lys Ile Gln
[0846]	115 120 125
[0847]	Ala Gly Ile Pro Glu Glu Glu Val Leu Pro Glu Leu Lys Lys Ile Lys
[0848]	130 135 140
[0849]	Glu Met Ile Gln Lys Asp Ile Val Asn Arg Lys Glu Gln Leu Leu Lys
[0850]	145 150 155 160
[0851]	Ser Ile Lys Asn Asn Arg Ile Pro Phe Ser Leu Glu Gly Ser Lys Leu
[0852]	165 170 175
[0853]	Val Pro Ser Thr Lys Lys Met Lys Trp Leu Phe Lys Leu Ile Asp Val
[0854]	180 185 190
[0855]	Pro Asn Lys Thr Phe Asn Glu Lys Met Leu Glu Lys Tyr Trp Glu Ile
[0856]	195 200 205

[0857]	Tyr Asp Tyr Asp Lys Leu Lys Ala Asn Ile Thr Asn Arg Leu Asp Lys		
[0858]	210	215	220
[0859]	Thr Asp Lys Lys Ala Arg Ser Ile Ser Arg Ala Val Ser Glu Glu Leu		
[0860]	225	230	235 240
[0861]	Arg Glu Tyr His Lys Asn Leu Arg Thr Asn Tyr Asn Arg Phe Val Ser		
[0862]	245	250	255
[0863]	Gly Asp Arg Pro Ala Ala Gly Leu Asp Asn Gly Gly Ser Ala Lys Tyr		
[0864]	260	265	270
[0865]	Asn Pro Asp Lys Glu Glu Phe Leu Leu Phe Leu Lys Glu Val Glu Gln		
[0866]	275	280	285
[0867]	Tyr Phe Lys Lys Tyr Phe Pro Val Lys Ser Lys His Ser Asn Lys Ser		
[0868]	290	295	300
[0869]	Lys Asp Lys Ser Leu Val Asp Lys Tyr Lys Asn Tyr Cys Ser Tyr Lys		
[0870]	305	310	315 320
[0871]	Val Val Lys Lys Glu Val Asn Arg Ser Ile Ile Asn Gln Leu Val Ala		
[0872]	325	330	335
[0873]	Gly Leu Ile Gln Gln Gly Lys Leu Leu Tyr Tyr Phe Tyr Tyr Asn Asp		
[0874]	340	345	350
[0875]	Thr Trp Gln Glu Asp Phe Leu Asn Ser Tyr Gly Leu Ser Tyr Ile Gln		
[0876]	355	360	365
[0877]	Val Glu Glu Ala Phe Lys Lys Ser Val Met Thr Ser Leu Ser Trp Gly		
[0878]	370	375	380
[0879]	Ile Asn Arg Leu Thr Ser Phe Phe Ile Asp Asp Ser Asn Thr Val Lys		
[0880]	385	390	395 400
[0881]	Phe Asp Asp Ile Thr Thr Lys Lys Ala Lys Glu Ala Ile Glu Ser Asn		
[0882]	405	410	415
[0883]	Tyr Phe Asn Lys Leu Arg Thr Cys Ser Arg Met Gln Asp His Phe Lys		
[0884]	420	425	430
[0885]	Glu Lys Leu Ala Phe Phe Tyr Pro Val Tyr Val Lys Asp Lys Lys Asp		
[0886]	435	440	445
[0887]	Arg Pro Asp Asp Asp Ile Glu Asn Leu Ile Val Leu Val Lys Asn Ala		
[0888]	450	455	460
[0889]	Ile Glu Ser Val Ser Tyr Leu Arg Asn Arg Thr Phe His Phe Lys Glu		
[0890]	465	470	475 480
[0891]	Ser Ser Leu Leu Glu Leu Leu Lys Glu Leu Asp Asp Lys Asn Ser Gly		
[0892]	485	490	495
[0893]	Gln Asn Lys Ile Asp Tyr Ser Val Ala Ala Glu Phe Ile Lys Arg Asp		
[0894]	500	505	510
[0895]	Ile Glu Asn Leu Tyr Asp Val Phe Arg Glu Gln Ile Arg Ser Leu Gly		

[0896]	515	520	525
[0897]	Ile Ala Glu Tyr Tyr Lys Ala Asp Met Ile Ser Asp Cys Phe Lys Thr		
[0898]	530	535	540
[0899]	Cys Gly Leu Glu Phe Ala Leu Tyr Ser Pro Lys Asn Ser Leu Met Pro		
[0900]	545	550	555
[0901]	Ala Phe Lys Asn Val Tyr Lys Arg Gly Ala Asn Leu Asn Lys Ala Tyr		
[0902]	565	570	575
[0903]	Ile Arg Asp Lys Gly Pro Lys Glu Thr Gly Asp Gln Gly Gln Asn Ser		
[0904]	580	585	590
[0905]	Tyr Lys Ala Leu Glu Glu Tyr Arg Glu Leu Thr Trp Tyr Ile Glu Val		
[0906]	595	600	605
[0907]	Lys Asn Asn Asp Gln Ser Tyr Asn Ala Tyr Lys Asn Leu Leu Gln Leu		
[0908]	610	615	620
[0909]	Ile Tyr Tyr His Ala Phe Leu Pro Glu Val Arg Glu Asn Glu Ala Leu		
[0910]	625	630	635
[0911]	Ile Thr Asp Phe Ile Asn Arg Thr Lys Glu Trp Asn Arg Lys Glu Thr		
[0912]	645	650	655
[0913]	Glu Glu Arg Leu Asn Thr Lys Asn Asn Lys Lys His Lys Asn Phe Asp		
[0914]	660	665	670
[0915]	Glu Asn Asp Asp Ile Thr Val Asn Thr Tyr Arg Tyr Glu Ser Ile Pro		
[0916]	675	680	685
[0917]	Asp Tyr Gln Gly Glu Ser Leu Asp Asp Tyr Leu Lys Val Leu Gln Arg		
[0918]	690	695	700
[0919]	Lys Gln Met Ala Arg Ala Lys Glu Val Asn Glu Lys Glu Glu Gly Asn		
[0920]	705	710	715
[0921]	Asn Asn Tyr Ile Gln Phe Ile Arg Asp Val Val Val Trp Ala Phe Gly		
[0922]	725	730	735
[0923]	Ala Tyr Leu Glu Asn Lys Leu Lys Asn Tyr Lys Asn Glu Leu Gln Pro		
[0924]	740	745	750
[0925]	Pro Leu Ser Lys Glu Asn Ile Gly Leu Asn Asp Thr Leu Lys Glu Leu		
[0926]	755	760	765
[0927]	Phe Pro Glu Glu Lys Val Lys Ser Pro Phe Asn Ile Lys Cys Arg Phe		
[0928]	770	775	780
[0929]	Ser Ile Ser Thr Phe Ile Asp Asn Lys Gly Lys Ser Thr Asp Asn Thr		
[0930]	785	790	795
[0931]	Ser Ala Glu Ala Val Lys Thr Asp Gly Lys Glu Asp Glu Lys Asp Lys		
[0932]	805	810	815
[0933]	Lys Asn Ile Lys Arg Lys Asp Leu Leu Cys Phe Tyr Leu Phe Leu Arg		
[0934]	820	825	830

[0935]	Leu	Leu	Asp	Glu	Asn	Glu	Ile	Cys	Lys	Leu	Gln	His	Gln	Phe	Ile	Lys
[0936]			835					840					845			
[0937]	Tyr	Arg	Cys	Ser	Leu	Lys	Glu	Arg	Arg	Phe	Pro	Gly	Asn	Arg	Thr	Lys
[0938]			850					855					860			
[0939]	Leu	Glu	Lys	Glu	Thr	Glu	Leu	Leu	Ala	Glu	Leu	Glu	Glu	Leu	Met	Glu
[0940]	865						870					875				880
[0941]	Leu	Val	Arg	Phe	Thr	Met	Pro	Ser	Ile	Pro	Glu	Ile	Ser	Ala	Lys	Ala
[0942]					885					890						895
[0943]	Glu	Ser	Gly	Tyr	Asp	Thr	Met	Ile	Lys	Lys	Tyr	Phe	Lys	Asp	Phe	Ile
[0944]				900						905					910	
[0945]	Glu	Lys	Lys	Val	Phe	Lys	Asn	Pro	Lys	Thr	Ser	Asn	Leu	Tyr	Tyr	His
[0946]				915						920					925	
[0947]	Ser	Asp	Ser	Lys	Thr	Pro	Val	Thr	Arg	Lys	Tyr	Met	Ala	Leu	Leu	Met
[0948]		930							935					940		
[0949]	Arg	Ser	Ala	Pro	Leu	His	Leu	Tyr	Lys	Asp	Ile	Phe	Lys	Gly	Tyr	Tyr
[0950]	945					950						955				960
[0951]	Leu	Ile	Thr	Lys	Lys	Glu	Cys	Leu	Glu	Tyr	Ile	Lys	Leu	Ser	Asn	Ile
[0952]				965							970					975
[0953]	Ile	Lys	Asp	Tyr	Gln	Asn	Ser	Leu	Asn	Glu	Leu	His	Glu	Gln	Leu	Glu
[0954]				980							985					990
[0955]	Arg	Ile	Lys	Leu	Lys	Ser	Glu	Lys	Gln	Asn	Gly	Lys	Asp	Ser	Leu	Tyr
[0956]			995							1000						1005
[0957]	Leu	Asp	Lys	Lys	Asp	Phe	Tyr	Lys	Val	Lys	Glu	Tyr	Val	Glu	Asn	
[0958]		1010							1015					1020		
[0959]	Leu	Glu	Gln	Val	Ala	Arg	Tyr	Lys	His	Leu	Gln	His	Lys	Ile	Asn	
[0960]		1025							1030					1035		
[0961]	Phe	Glu	Ser	Leu	Tyr	Arg	Ile	Phe	Arg	Ile	His	Val	Asp	Ile	Ala	
[0962]		1040							1045					1050		
[0963]	Ala	Arg	Met	Val	Gly	Tyr	Thr	Gln	Asp	Trp	Glu	Arg	Asp	Met	His	
[0964]		1055							1060					1065		
[0965]	Phe	Leu	Phe	Lys	Ala	Leu	Val	Tyr	Asn	Gly	Val	Leu	Glu	Glu	Arg	
[0966]		1070							1075					1080		
[0967]	Arg	Phe	Glu	Ala	Ile	Phe	Asn	Asn	Asn	Asp	Asp	Asn	Asn	Asp	Gly	
[0968]		1085							1090					1095		
[0969]	Arg	Ile	Val	Lys	Lys	Ile	Gln	Asn	Asn	Leu	Asn	Asn	Lys	Asn	Arg	
[0970]		1100							1105					1110		
[0971]	Glu	Leu	Val	Ser	Met	Leu	Cys	Trp	Asn	Lys	Lys	Leu	Asn	Lys	Asn	
[0972]		1115							1120					1125		
[0973]	Glu	Phe	Gly	Ala	Ile	Ile	Trp	Lys	Arg	Asn	Pro	Ile	Ala	His	Leu	

[0974]	1130	1135	1140	
[0975]	Asn His Phe Thr Gln Thr Glu Gln Asn Ser Lys Ser Ser Leu Glu			
[0976]	1145	1150	1155	
[0977]	Ser Leu Ile Asn Ser Leu Arg Ile Leu Leu Ala Tyr Asp Arg Lys			
[0978]	1160	1165	1170	
[0979]	Arg Gln Asn Ala Val Thr Lys Thr Ile Asn Asp Leu Leu Leu Asn			
[0980]	1175	1180	1185	
[0981]	Asp Tyr His Ile Arg Ile Lys Trp Glu Gly Arg Val Asp Glu Gly			
[0982]	1190	1195	1200	
[0983]	Gln Ile Tyr Phe Asn Ile Lys Glu Lys Glu Asp Ile Glu Asn Glu			
[0984]	1205	1210	1215	
[0985]	Pro Ile Ile His Leu Lys His Leu His Lys Lys Asp Cys Tyr Ile			
[0986]	1220	1225	1230	
[0987]	Tyr Lys Asn Ser Tyr Met Phe Asp Lys Gln Lys Glu Trp Ile Cys			
[0988]	1235	1240	1245	
[0989]	Asn Gly Ile Lys Glu Glu Val Tyr Asp Lys Ser Ile Leu Lys Cys			
[0990]	1250	1255	1260	
[0991]	Ile Gly Asn Leu Phe Lys Phe Asp Tyr Glu Asp Lys Asn Lys Ser			
[0992]	1265	1270	1275	
[0993]	Ser Ala Asn Pro Lys His Thr			
[0994]	1280	1285		
[0995]	<210> 7			
[0996]	<211> 30			
[0997]	<212> RNA			
[0998]	<213> 斯氏李斯特菌			
[0999]	<400> 7			
[1000]	gacuaccucu auaugaaaga ggacuaaaac			30
[1001]	<210> 8			
[1002]	<211> 29			
[1003]	<212> RNA			
[1004]	<213> 沙氏纤毛菌			
[1005]	<400> 8			
[1006]	ccaccccaau aucgaagggg acuaaaaca			29
[1007]	<210> 9			
[1008]	<211> 31			
[1009]	<212> RNA			
[1010]	<213> 口腔纤毛菌			
[1011]	<400> 9			
[1012]	gaccaccca aaaaugaagg ggacuaaaac a			31

[1013]	<210>	10	
[1014]	<211>	50	
[1015]	<212>	RNA	
[1016]	<213>	斯氏李斯特菌	
[1017]	<400>	10	
[1018]		gacuaccucu auaugaaaga ggacuaaaac caaacaugau cugggucauc	50
[1019]	<210>	11	
[1020]	<211>	55	
[1021]	<212>	RNA	
[1022]	<213>	斯氏李斯特菌	
[1023]	<400>	11	
[1024]		uaagagacua ccucuauaug aaagaggacu aaaaccaaac augaucuggg ucauc	55
[1025]	<210>	12	
[1026]	<211>	49	
[1027]	<212>	RNA	
[1028]	<213>	沙氏纤毛菌	
[1029]	<400>	12	
[1030]		ccaccccaau aucgaagggg acuaaaacag gggcagagau gaugacccu	49
[1031]	<210>	13	
[1032]	<211>	58	
[1033]	<212>	RNA	
[1034]	<213>	沙氏纤毛菌	
[1035]	<400>	13	
[1036]		ggauuuagac caccccaaua ucgaagggga cuaaaacagg ggcagagaug augacccu	58
[1037]	<210>	14	
[1038]	<211>	51	
[1039]	<212>	RNA	
[1040]	<213>	口腔纤毛菌	
[1041]	<400>	14	
[1042]		gaccacccca aaaaugaagg ggacuaaaac aggggcagag augaugacc u	51
[1043]	<210>	15	
[1044]	<211>	56	
[1045]	<212>	RNA	
[1046]	<213>	口腔纤毛菌	
[1047]	<400>	15	
[1048]		auuuagacca ccccaaaaau gaaggggacu aaaacagggg cagagaugau gacccu	56
[1049]	<210>	16	
[1050]	<211>	27	
[1051]	<212>	RNA	

[1052]	<213>	人工序列	
[1053]	<220>		
[1054]	<223>	合成序列	
[1055]	<400>	16	
[1056]	agauagccca agaaagaggg caauaac		27
[1057]	<210>	17	
[1058]	<211>	32	
[1059]	<212>	RNA	
[1060]	<213>	人工序列	
[1061]	<220>		
[1062]	<223>	合成序列	
[1063]	<400>	17	
[1064]	guaacaaucc ccguagacag gggaacugca ac		32
[1065]	<210>	18	
[1066]	<211>	31	
[1067]	<212>	RNA	
[1068]	<213>	人工序列	
[1069]	<220>		
[1070]	<223>	合成序列	
[1071]	<400>	18	
[1072]	caucaccgcc aagacgacgg cggacugaac c		31
[1073]	<210>	19	
[1074]	<211>	30	
[1075]	<212>	RNA	
[1076]	<213>	人工序列	
[1077]	<220>		
[1078]	<223>	合成序列	
[1079]	<400>	19	
[1080]	aaauauccca aaauugaagg gaacuacaac		30
[1081]	<210>	20	
[1082]	<211>	30	
[1083]	<212>	RNA	
[1084]	<213>	人工序列	
[1085]	<220>		
[1086]	<223>	合成序列	
[1087]	<400>	20	
[1088]	gaguaccuca aaacaaaaga ggacuaaaac		30
[1089]	<210>	21	
[1090]	<211>	32	

[1091]	<212>	RNA	
[1092]	<213>	人工序列	
[1093]	<220>		
[1094]	<223>	合成序列	
[1095]	<400>	21	
[1096]		gaccacccca auaucgaagg ggacuaaaac uu	32
[1097]	<210>	22	
[1098]	<211>	27	
[1099]	<212>	RNA	
[1100]	<213>	人工序列	
[1101]	<220>		
[1102]	<223>	合成序列	
[1103]	<400>	22	
[1104]		caccccaaua ucgaagggga cuaaaac	27
[1105]	<210>	23	
[1106]	<211>	27	
[1107]	<212>	RNA	
[1108]	<213>	人工序列	
[1109]	<220>		
[1110]	<223>	合成序列	
[1111]	<400>	23	
[1112]		aaguagcccg auauagaggg caauaac	27
[1113]	<210>	24	
[1114]	<211>	27	
[1115]	<212>	RNA	
[1116]	<213>	人工序列	
[1117]	<220>		
[1118]	<223>	合成序列	
[1119]	<400>	24	
[1120]		auacagcucg auauagugag caauaag	27
[1121]	<210>	25	
[1122]	<211>	34	
[1123]	<212>	RNA	
[1124]	<213>	人工序列	
[1125]	<220>		
[1126]	<223>	合成序列	
[1127]	<400>	25	
[1128]		ucacaucacc gccaaagacga cggcggacug aacc	34
[1129]	<210>	26	

[1130]	<211>	28	
[1131]	<212>	RNA	
[1132]	<213>	人工序列	
[1133]	<220>		
[1134]	<223>	合成序列	
[1135]	<400>	26	
[1136]		gaacagcccg auauagaggg caauagac	28
[1137]	<210>	27	
[1138]	<211>	60	
[1139]	<212>	RNA	
[1140]	<213>	人工序列	
[1141]	<220>		
[1142]	<223>	合成序列	
[1143]	<400>	27	
[1144]		ggcucauuug caggggggag ccaaaagggg caucaucucu gccccucug cugaugcccc	60
[1145]	<210>	28	
[1146]	<211>	51	
[1147]	<212>	RNA	
[1148]	<213>	人工序列	
[1149]	<220>		
[1150]	<223>	合成序列	
[1151]	<400>	28	
[1152]		gaccacccca aaaaugaagg ggacuaaaac aggggcagag augaugaccc u	51
[1153]	<210>	29	
[1154]	<211>	50	
[1155]	<212>	RNA	
[1156]	<213>	人工序列	
[1157]	<220>		
[1158]	<223>	合成序列	
[1159]	<400>	29	
[1160]		gacuaccucu auaugaaaga ggacuaaaac caaacaugau cugggucauc	50
[1161]	<210>	30	
[1162]	<211>	55	
[1163]	<212>	RNA	
[1164]	<213>	人工序列	
[1165]	<220>		
[1166]	<223>	合成序列	
[1167]	<400>	30	
[1168]		uaagagacua ccucuauaug aaagaggacu aaaaccaaac augaucuggg ucauc	55

[1169]	<210>	31	
[1170]	<211>	49	
[1171]	<212>	RNA	
[1172]	<213>	人工序列	
[1173]	<220>		
[1174]	<223>	合成序列	
[1175]	<400>	31	
[1176]		ccaccccaau aucgaagggg acuaaaacag gggcagagau gaugacccu	49
[1177]	<210>	32	
[1178]	<211>	58	
[1179]	<212>	RNA	
[1180]	<213>	人工序列	
[1181]	<220>		
[1182]	<223>	合成序列	
[1183]	<400>	32	
[1184]		ggauuuagac caccccaaua ucgaagggga cuaaaacagg ggcagagaug augacccu	58
[1185]	<210>	33	
[1186]	<211>	51	
[1187]	<212>	RNA	
[1188]	<213>	人工序列	
[1189]	<220>		
[1190]	<223>	合成序列	
[1191]	<400>	33	
[1192]		gaccaccca aaaaugaagg ggacuaaaac aggggcagag augaugaccc u	51
[1193]	<210>	34	
[1194]	<211>	56	
[1195]	<212>	RNA	
[1196]	<213>	人工序列	
[1197]	<220>		
[1198]	<223>	合成序列	
[1199]	<400>	34	
[1200]		auuuagacca ccccaaaaau gaaggggacu aaaacagggg cagagaugau gacccu	56
[1201]	<210>	35	
[1202]	<211>	30	
[1203]	<212>	RNA	
[1204]	<213>	人工序列	
[1205]	<220>		
[1206]	<223>	合成序列	
[1207]	<400>	35	

[1208]	ggauuuagac caccceaaaa augaagggga	30
[1209]	<210> 36	
[1210]	<211> 30	
[1211]	<212> RNA	
[1212]	<213> 人工序列	
[1213]	<220>	
[1214]	<223> 合成序列	
[1215]	<400> 36	
[1216]	gguaagagac uaccucuaua ugaaagagga	30
[1217]	<210> 37	
[1218]	<211> 30	
[1219]	<212> RNA	
[1220]	<213> 人工序列	
[1221]	<220>	
[1222]	<223> 合成序列	
[1223]	<400> 37	
[1224]	ggauauagac caccceaaaua ucgaagggga	30
[1225]	<210> 38	
[1226]	<211> 37	
[1227]	<212> RNA	
[1228]	<213> 人工序列	
[1229]	<220>	
[1230]	<223> 合成序列	
[1231]	<400> 38	
[1232]	ggauuuagac caccceaaaa augaagggga cuaaaac	37
[1233]	<210> 39	
[1234]	<211> 14	
[1235]	<212> RNA	
[1236]	<213> 人工序列	
[1237]	<220>	
[1238]	<223> 合成序列	
[1239]	<400> 39	
[1240]	aaacgaccac aagg	14
[1241]	<210> 40	
[1242]	<211> 19	
[1243]	<212> RNA	
[1244]	<213> 人工序列	
[1245]	<220>	
[1246]	<223> 合成序列	

[1247]	<400>	40	
[1248]	ccaagaaacg accacaagg		19
[1249]	<210>	41	
[1250]	<211>	19	
[1251]	<212>	RNA	
[1252]	<213>	人工序列	
[1253]	<220>		
[1254]	<223>	合成序列	
[1255]	<400>	41	
[1256]	aaagcaaaaa aggccccgc		19
[1257]	<210>	42	
[1258]	<211>	23	
[1259]	<212>	RNA	
[1260]	<213>	人工序列	
[1261]	<220>		
[1262]	<223>	合成序列	
[1263]	<400>	42	
[1264]	cacgccgcac aaaaccacg ccg		23
[1265]	<210>	43	
[1266]	<211>	25	
[1267]	<212>	RNA	
[1268]	<213>	人工序列	
[1269]	<220>		
[1270]	<223>	合成序列	
[1271]	<400>	43	
[1272]	ccggagcgag gaagaaacgc ccgcg		25
[1273]	<210>	44	
[1274]	<211>	25	
[1275]	<212>	RNA	
[1276]	<213>	人工序列	
[1277]	<220>		
[1278]	<223>	合成序列	
[1279]	<400>	44	
[1280]	agcgcccaac ggaaagaagg aagga		25
[1281]	<210>	45	
[1282]	<211>	58	
[1283]	<212>	RNA	
[1284]	<213>	人工序列	
[1285]	<220>		

[1286]	<223>	合成序列	
[1287]	<400>	45	
[1288]	ggauuuagac caccaccaaa augaagggga cuaaaacagg ggcagagaug augacccu	58	
[1289]	<210>	46	
[1290]	<211>	57	
[1291]	<212>	RNA	
[1292]	<213>	人工序列	
[1293]	<220>		
[1294]	<223>	合成序列	
[1295]	<400>	46	
[1296]	gguaagagac uaccucuaua ugaaagagga cuaaaaccaa acaugaucug ggucauc	57	
[1297]	<210>	47	
[1298]	<211>	58	
[1299]	<212>	RNA	
[1300]	<213>	人工序列	
[1301]	<220>		
[1302]	<223>	合成序列	
[1303]	<400>	47	
[1304]	ggauuuagac caccaccaua ucgaagggga cuaaaacagg ggcagagaug augacccu	58	
[1305]	<210>	48	
[1306]	<211>	58	
[1307]	<212>	RNA	
[1308]	<213>	人工序列	
[1309]	<220>		
[1310]	<223>	合成序列	
[1311]	<400>	48	
[1312]	ggauuuagac cagggaagu aaaaaccca cuaaaacagg ggcagagaug augacccu	58	
[1313]	<210>	49	
[1314]	<211>	60	
[1315]	<212>	RNA	
[1316]	<213>	人工序列	
[1317]	<220>		
[1318]	<223>	合成序列	
[1319]	<400>	49	
[1320]	ggauuuagac caccaccaaa aaugaagggg gacuaaaaca ggggcagaga ugaugacccu	60	
[1321]	<210>	50	
[1322]	<211>	54	
[1323]	<212>	RNA	
[1324]	<213>	人工序列	

[1325]	<220>	
[1326]	<223>	合成序列
[1327]	<400>	50
[1328]	ggauuuagac caccccauga agggggacuaa aacaggggca gagaugauga ccu	54
[1329]	<210>	51
[1330]	<211>	54
[1331]	<212>	RNA
[1332]	<213>	人工序列
[1333]	<220>	
[1334]	<223>	合成序列
[1335]	<400>	51
[1336]	ggauuuagac caccccauga agggggacuaa aacaggggca gagaugauga ccu	54
[1337]	<210>	52
[1338]	<211>	58
[1339]	<212>	RNA
[1340]	<213>	人工序列
[1341]	<220>	
[1342]	<223>	合成序列
[1343]	<400>	52
[1344]	ggauuuagac cacccca aaa augaagggca cuaaaacagg ggcagagaug augacccu	58
[1345]	<210>	53
[1346]	<211>	58
[1347]	<212>	RNA
[1348]	<213>	人工序列
[1349]	<220>	
[1350]	<223>	合成序列
[1351]	<400>	53
[1352]	ggcguuagac cacccca aaa augaagggga cuaaaacagg ggcagagaug augacccu	58
[1353]	<210>	54
[1354]	<211>	58
[1355]	<212>	RNA
[1356]	<213>	人工序列
[1357]	<220>	
[1358]	<223>	合成序列
[1359]	<400>	54
[1360]	ggagcuagac cacccca aaa augaagggga cuaaaacagg ggcagagaug augacccu	58
[1361]	<210>	55
[1362]	<211>	58
[1363]	<212>	RNA

[1364]	<213>	人工序列	
[1365]	<220>		
[1366]	<223>	合成序列	
[1367]	<400>	55	
[1368]	ggauccagac cacccecaaaa augaagggga cuaaaacagg ggcagagaug augacccu	58	
[1369]	<210>	56	
[1370]	<211>	58	
[1371]	<212>	RNA	
[1372]	<213>	人工序列	
[1373]	<220>		
[1374]	<223>	合成序列	
[1375]	<400>	56	
[1376]	ggauucggac cacccecaaaa augaagggga cuaaaacagg ggcagagaug augacccu	58	
[1377]	<210>	57	
[1378]	<211>	58	
[1379]	<212>	RNA	
[1380]	<213>	人工序列	
[1381]	<220>		
[1382]	<223>	合成序列	
[1383]	<400>	57	
[1384]	ggauuuaccc cacccecaaaa augaagggga cuaaaacagg ggcagagaug augacccu	58	
[1385]	<210>	58	
[1386]	<211>	58	
[1387]	<212>	RNA	
[1388]	<213>	人工序列	
[1389]	<220>		
[1390]	<223>	合成序列	
[1391]	<400>	58	
[1392]	ggauuuuauuc cacccecaaaa augaagggga cuaaaacagg ggcagagaug augacccu	58	
[1393]	<210>	59	
[1394]	<211>	58	
[1395]	<212>	RNA	
[1396]	<213>	人工序列	
[1397]	<220>		
[1398]	<223>	合成序列	
[1399]	<400>	59	
[1400]	ggauuuagaa aacccecaaaa augaagggga cuaaaacagg ggcagagaug augacccu	58	
[1401]	<210>	60	
[1402]	<211>	58	

[1403]	<212>	RNA	
[1404]	<213>	人工序列	
[1405]	<220>		
[1406]	<223>	合成序列	
[1407]	<400>	60	
[1408]	ggauuuagac	cgccccaaaa	augaagggga cuaaaacagg ggcagagaug augacccu 58
[1409]	<210>	61	
[1410]	<211>	58	
[1411]	<212>	RNA	
[1412]	<213>	人工序列	
[1413]	<220>		
[1414]	<223>	合成序列	
[1415]	<400>	61	
[1416]	ggauuuagac	cagccccaaaa	augaagggca cuaaaacagg ggcagagaug augacccu 58
[1417]	<210>	62	
[1418]	<211>	58	
[1419]	<212>	RNA	
[1420]	<213>	人工序列	
[1421]	<220>		
[1422]	<223>	合成序列	
[1423]	<400>	62	
[1424]	ggauuuagac	caccgcaaaa	augaagcgga cuaaaacagg ggcagagaug augacccu 58
[1425]	<210>	63	
[1426]	<211>	58	
[1427]	<212>	RNA	
[1428]	<213>	人工序列	
[1429]	<220>		
[1430]	<223>	合成序列	
[1431]	<400>	63	
[1432]	ggauuuagac	cacaccaaaa	augaagguga cuaaaacagg ggcagagaug augacccu 58
[1433]	<210>	64	
[1434]	<211>	58	
[1435]	<212>	RNA	
[1436]	<213>	人工序列	
[1437]	<220>		
[1438]	<223>	合成序列	
[1439]	<400>	64	
[1440]	ggauuuagac	caccacaaaa	augaagugga cuaaaacagg ggcagagaug augacccu 58
[1441]	<210>	65	

[1442]	<211>	58	
[1443]	<212>	RNA	
[1444]	<213>	人工序列	
[1445]	<220>		
[1446]	<223>	合成序列	
[1447]	<400>	65	
[1448]	ggauuuagac cacuccaaaa augaaggaga cuaaaacagg ggcagagaug augacccu	58	
[1449]	<210>	66	
[1450]	<211>	58	
[1451]	<212>	RNA	
[1452]	<213>	人工序列	
[1453]	<220>		
[1454]	<223>	合成序列	
[1455]	<400>	66	
[1456]	ggauuuagac caccceaaaa augaaggggc auaaaacagg ggcagagaug augacccu	58	
[1457]	<210>	67	
[1458]	<211>	58	
[1459]	<212>	RNA	
[1460]	<213>	人工序列	
[1461]	<220>		
[1462]	<223>	合成序列	
[1463]	<400>	67	
[1464]	ggauuuagac caccceaaaa augaagggga cgcaaacagg ggcagagaug augacccu	58	
[1465]	<210>	68	
[1466]	<211>	58	
[1467]	<212>	RNA	
[1468]	<213>	人工序列	
[1469]	<220>		
[1470]	<223>	合成序列	
[1471]	<400>	68	
[1472]	ggauuuagac caccceaaaa augaagggga cuagcacagg ggcagagaug augacccu	58	
[1473]	<210>	69	
[1474]	<211>	58	
[1475]	<212>	RNA	
[1476]	<213>	人工序列	
[1477]	<220>		
[1478]	<223>	合成序列	
[1479]	<400>	69	
[1480]	ggauuuagac caccceaaaa augaagggga cuaaaguagg ggcagagaug augacccu	58	

[1481]	<210>	70	
[1482]	<211>	51	
[1483]	<212>	RNA	
[1484]	<213>	人工序列	
[1485]	<220>		
[1486]	<223>	合成序列	
[1487]	<400>	70	
[1488]	ggccacccca	aaaaugaagg ggacuaaaac aggggcagag augaugaccc u	51
[1489]	<210>	71	
[1490]	<211>	51	
[1491]	<212>	RNA	
[1492]	<213>	人工序列	
[1493]	<220>		
[1494]	<223>	合成序列	
[1495]	<400>	71	
[1496]	ggccacccca	aaaaugaagg ggacuaaaac acaaacauga ucugggucau c	51
[1497]	<210>	72	
[1498]	<211>	60	
[1499]	<212>	RNA	
[1500]	<213>	人工序列	
[1501]	<220>		
[1502]	<223>	合成序列	
[1503]	<400>	72	
[1504]	ggcacacccg	caggggggag ccaaaagggg caucaucucu gccccacag cagaagcccc	60
[1505]	<210>	73	
[1506]	<211>	62	
[1507]	<212>	RNA	
[1508]	<213>	人工序列	
[1509]	<220>		
[1510]	<223>	合成序列	
[1511]	<400>	73	
[1512]	gggaacccca	aggccaaccg cgagaagaug acccagauca uguuugagac cuucaacacc	60
[1513]	cc		62
[1514]	<210>	74	
[1515]	<211>	51	
[1516]	<212>	RNA	
[1517]	<213>	人工序列	
[1518]	<220>		
[1519]	<223>	合成序列	

[1520]	<400>	74	
[1521]	ggccacccca	aaaaugaagg	ggacuaaaac agugauaagu ggaaugccau g 51
[1522]	<210>	75	
[1523]	<211>	51	
[1524]	<212>	RNA	
[1525]	<213>	人工序列	
[1526]	<220>		
[1527]	<223>	合成序列	
[1528]	<400>	75	
[1529]	ggccacccca	aaaaugaagg	ggacuaaaac acuggugaac uuccgauagu g 51
[1530]	<210>	76	
[1531]	<211>	51	
[1532]	<212>	RNA	
[1533]	<213>	人工序列	
[1534]	<220>		
[1535]	<223>	合成序列	
[1536]	<400>	76	
[1537]	ggccacccca	aaaaugaagg	ggacuaaaac acagauauag ccuggugguu c 51
[1538]	<210>	77	
[1539]	<211>	56	
[1540]	<212>	RNA	
[1541]	<213>	人工序列	
[1542]	<220>		
[1543]	<223>	合成序列	
[1544]	<400>	77	
[1545]	ggcucaauuu	ugacagcggu	cauggcauuc cacuuauac uggaucuuu ccacuc 56
[1546]	<210>	78	
[1547]	<211>	55	
[1548]	<212>	RNA	
[1549]	<213>	人工序列	
[1550]	<220>		
[1551]	<223>	合成序列	
[1552]	<400>	78	
[1553]	ggaaaucuu	caacacccgc	acuaucggaa guucaccagc cagccgcagc acguu 55
[1554]	<210>	79	
[1555]	<211>	57	
[1556]	<212>	RNA	
[1557]	<213>	人工序列	
[1558]	<220>		

[1559]	<223>	合成序列	
[1560]	<400>	79	
[1561]	ggcaauaaaa	augcgccgcc	ugaaccacca
	ggcuauaucu	gccacucauu	guuguga
[1562]	<210>	80	57
[1563]	<211>	114	
[1564]	<212>	RNA	
[1565]	<213>	人工序列	
[1566]	<220>		
[1567]	<223>	合成序列	
[1568]	<400>	80	
[1569]	ggauuuagac	cacccccaaaa	augaagggga
	cuaaaacagg	ggcagagaug	augacccuau
[1570]	uuagaccacc	ccaaaaauga	aggggacuaa
	aacagugaua	aguggaaugc	caug
[1571]	<210>	81	60
[1572]	<211>	58	
[1573]	<212>	RNA	
[1574]	<213>	人工序列	
[1575]	<220>		
[1576]	<223>	合成序列	
[1577]	<400>	81	
[1578]	ggauuuagac	cacccccaaaa	augaagggga
	cuaaaacagu	gauaagugga	augccaug
[1579]	<210>	82	58
[1580]	<211>	58	
[1581]	<212>	RNA	
[1582]	<213>	人工序列	
[1583]	<220>		
[1584]	<223>	合成序列	
[1585]	<400>	82	
[1586]	ggauuuauua	aacccccaaaa	augaagggga
	cuaaaacagg	ggcagagaug	augacccu
[1587]	<210>	83	58
[1588]	<211>	58	
[1589]	<212>	RNA	
[1590]	<213>	人工序列	
[1591]	<220>		
[1592]	<223>	合成序列	
[1593]	<400>	83	
[1594]	ggauucgauc	cacccccaaaa	augaagggga
	cuaaaacagg	ggcagagaug	augacccu
[1595]	<210>	84	58
[1596]	<211>	58	
[1597]	<212>	RNA	

[1598]	<213>	人工序列	
[1599]	<220>		
[1600]	<223>	合成序列	
[1601]	<400>	84	
[1602]		ggauuuagga agccccaaaa augaaggga cuaaaacagg ggcagagaug augacccu	58
[1603]	<210>	85	
[1604]	<211>	58	
[1605]	<212>	RNA	
[1606]	<213>	人工序列	
[1607]	<220>		
[1608]	<223>	合成序列	
[1609]	<400>	85	
[1610]		ggauuuagac caggccaaaa augaaggcca cuaaaacagg ggcagagaug augacccu	58
[1611]	<210>	86	
[1612]	<211>	55	
[1613]	<212>	RNA	
[1614]	<213>	人工序列	
[1615]	<220>		
[1616]	<223>	合成序列	
[1617]	<400>	86	
[1618]		ggccacccca aaaaugaagg ggacuaaaac auuuuggcuc cccccugcaa augag	55
[1619]	<210>	87	
[1620]	<211>	55	
[1621]	<212>	RNA	
[1622]	<213>	人工序列	
[1623]	<220>		
[1624]	<223>	合成序列	
[1625]	<400>	87	
[1626]		ggccacccca aaaaugaagg ggacuaaaac aaccuuuuug gcucuuuuuu gcaaa	55
[1627]	<210>	88	
[1628]	<211>	55	
[1629]	<212>	RNA	
[1630]	<213>	人工序列	
[1631]	<220>		
[1632]	<223>	合成序列	
[1633]	<400>	88	
[1634]		ggccacccca aaaaugaagg ggacuaaaac agaagacccu uuuggcucuu ccuug	55
[1635]	<210>	89	
[1636]	<211>	55	

[1637]	<212>	RNA	
[1638]	<213>	人工序列	
[1639]	<220>		
[1640]	<223>	合成序列	
[1641]	<400>	89	
[1642]		ggccacccca aaaaugaagg ggacuaaaac aagaugauga ccuuuuggc ucccc	55
[1643]	<210>	90	
[1644]	<211>	55	
[1645]	<212>	RNA	
[1646]	<213>	人工序列	
[1647]	<220>		
[1648]	<223>	合成序列	
[1649]	<400>	90	
[1650]		ggccacccca aaaaugaagg ggacuaaaac agcagagaug augacccuuu ugguu	55
[1651]	<210>	91	
[1652]	<211>	55	
[1653]	<212>	RNA	
[1654]	<213>	人工序列	
[1655]	<220>		
[1656]	<223>	合成序列	
[1657]	<400>	91	
[1658]		ggccacccca aaaaugaagg ggacuaaaac agggggcaga gaugaugacc cuuuu	55
[1659]	<210>	92	
[1660]	<211>	55	
[1661]	<212>	RNA	
[1662]	<213>	人工序列	
[1663]	<220>		
[1664]	<223>	合成序列	
[1665]	<400>	92	
[1666]		ggccacccca aaaaugaagg ggacuaaaac acagaggggg cagagaugau gaccc	55
[1667]	<210>	93	
[1668]	<211>	55	
[1669]	<212>	RNA	
[1670]	<213>	人工序列	
[1671]	<220>		
[1672]	<223>	合成序列	
[1673]	<400>	93	
[1674]		ggccacccca aaaaugaagg ggacuaaaac aucagcagag ggggcagaga ugaug	55
[1675]	<210>	94	

[1676]	<211>	60	
[1677]	<212>	RNA	
[1678]	<213>	人工序列	
[1679]	<220>		
[1680]	<223>	合成序列	
[1681]	<400>	94	
[1682]	ggcucuuug caggggggag ccaaaagggg	caucaucucu gccccucug cugaugcccc	60
[1683]	<210>	95	
[1684]	<211>	70	
[1685]	<212>	RNA	
[1686]	<213>	人工序列	
[1687]	<220>		
[1688]	<223>	合成序列	
[1689]	<400>	95	
[1690]	ggccugacug cucucuuug caguugggag ccaaaagggg	caucaucucu gccccucug	60
[1691]	cugaugcccc		70
[1692]	<210>	96	
[1693]	<211>	80	
[1694]	<212>	RNA	
[1695]	<213>	人工序列	
[1696]	<220>		
[1697]	<223>	合成序列	
[1698]	<400>	96	
[1699]	ggaccuguga auccugacug cucucuuug caguugggag ccaaaagggg	caucaucucu	60
[1700]	gccccucug cugaugcccc		80
[1701]	<210>	97	
[1702]	<211>	60	
[1703]	<212>	RNA	
[1704]	<213>	人工序列	
[1705]	<220>		
[1706]	<223>	合成序列	
[1707]	<400>	97	
[1708]	ggcacacccg caggguuuag ccaaaagggg	caucaucucu gccccucug cugaugcccc	60
[1709]	<210>	98	
[1710]	<211>	47	
[1711]	<212>	RNA	
[1712]	<213>	人工序列	
[1713]	<220>		
[1714]	<223>	合成序列	

[1715]	<400>	98	
[1716]	ggccacccca	aaaaugaagg	ggacuaaaac acagagauga ugacccu 47
[1717]	<210>	99	
[1718]	<211>	55	
[1719]	<212>	RNA	
[1720]	<213>	人工序列	
[1721]	<220>		
[1722]	<223>	合成序列	
[1723]	<400>	99	
[1724]	ggccacccca	aaaaugaagg	ggacuaaaac aagagggggc agagaugaug acccu 55
[1725]	<210>	100	
[1726]	<211>	51	
[1727]	<212>	RNA	
[1728]	<213>	人工序列	
[1729]	<220>		
[1730]	<223>	合成序列	
[1731]	<400>	100	
[1732]	gaccacccca	aaaaugaagg	ggacuaaaac aggggcagag augaugaccc u 51
[1733]	<210>	101	
[1734]	<211>	53	
[1735]	<212>	RNA	
[1736]	<213>	人工序列	
[1737]	<220>		
[1738]	<223>	合成序列	
[1739]	<400>	101	
[1740]	gggaccaccc	caaaaugaa	ggggacuaaa acaggggcag agaugaugac ccu 53
[1741]	<210>	102	
[1742]	<211>	49	
[1743]	<212>	RNA	
[1744]	<213>	人工序列	
[1745]	<220>		
[1746]	<223>	合成序列	
[1747]	<400>	102	
[1748]	ccacccccaa	aaugaagggg	acuaaaacag gggcagagau gaugacccu 49
[1749]	<210>	103	
[1750]	<211>	38	
[1751]	<212>	RNA	
[1752]	<213>	人工序列	
[1753]	<220>		

[1754]	<223>	合成序列	
[1755]	<400>	103	
[1756]	ggauucgauc	cacccccaaaa	augaagggga cuaaaaca 38
[1757]	<210>	104	
[1758]	<211>	38	
[1759]	<212>	RNA	
[1760]	<213>	人工序列	
[1761]	<220>		
[1762]	<223>	合成序列	
[1763]	<400>	104	
[1764]	ggauuuuaaua	aacccccaaaa	augaagggga cuaaaaca 38
[1765]	<210>	105	
[1766]	<211>	38	
[1767]	<212>	RNA	
[1768]	<213>	人工序列	
[1769]	<220>		
[1770]	<223>	合成序列	
[1771]	<400>	105	
[1772]	ggauuuagga	agccccaaaa	augaagggga cuaaaaca 38
[1773]	<210>	106	
[1774]	<211>	38	
[1775]	<212>	RNA	
[1776]	<213>	人工序列	
[1777]	<220>		
[1778]	<223>	合成序列	
[1779]	<400>	106	
[1780]	ggauuuagac	caggccaaaa	augaaggcca cuaaaaca 38
[1781]	<210>	107	
[1782]	<211>	38	
[1783]	<212>	RNA	
[1784]	<213>	人工序列	
[1785]	<220>		
[1786]	<223>	合成序列	
[1787]	<400>	107	
[1788]	ggauuuagac	cacccccaaaa	augaagggga cuaaaaca 38
[1789]	<210>	108	
[1790]	<211>	38	
[1791]	<212>	RNA	
[1792]	<213>	人工序列	

[1793]	<220>	
[1794]	<223>	合成序列
[1795]	<400>	108
[1796]	ggcguuagac cacccecaaaa augaagggga cuaaaaca	38
[1797]	<210>	109
[1798]	<211>	38
[1799]	<212>	RNA
[1800]	<213>	人工序列
[1801]	<220>	
[1802]	<223>	合成序列
[1803]	<400>	109
[1804]	ggagcuagac cacccecaaaa augaagggga cuaaaaca	38
[1805]	<210>	110
[1806]	<211>	38
[1807]	<212>	RNA
[1808]	<213>	人工序列
[1809]	<220>	
[1810]	<223>	合成序列
[1811]	<400>	110
[1812]	ggauccagac cacccecaaaa augaagggga cuaaaaca	38
[1813]	<210>	111
[1814]	<211>	38
[1815]	<212>	RNA
[1816]	<213>	人工序列
[1817]	<220>	
[1818]	<223>	合成序列
[1819]	<400>	111
[1820]	ggauucggac cacccecaaaa augaagggga cuaaaaca	38
[1821]	<210>	112
[1822]	<211>	38
[1823]	<212>	RNA
[1824]	<213>	人工序列
[1825]	<220>	
[1826]	<223>	合成序列
[1827]	<400>	112
[1828]	ggauuuacc cacccecaaaa augaagggga cuaaaaca	38
[1829]	<210>	113
[1830]	<211>	38
[1831]	<212>	RNA

[1832]	<213>	人工序列	
[1833]	<220>		
[1834]	<223>	合成序列	
[1835]	<400>	113	
[1836]	ggauuuuauuc caccaccaaaa augaagggga cuaaaaca		38
[1837]	<210>	114	
[1838]	<211>	38	
[1839]	<212>	RNA	
[1840]	<213>	人工序列	
[1841]	<220>		
[1842]	<223>	合成序列	
[1843]	<400>	114	
[1844]	ggauuuuagaa aaccaccaaaa augaagggga cuaaaaca		38
[1845]	<210>	115	
[1846]	<211>	38	
[1847]	<212>	RNA	
[1848]	<213>	人工序列	
[1849]	<220>		
[1850]	<223>	合成序列	
[1851]	<400>	115	
[1852]	ggauuuuagac cgccaccaaaa augaagggga cuaaaaca		38
[1853]	<210>	116	
[1854]	<211>	38	
[1855]	<212>	RNA	
[1856]	<213>	人工序列	
[1857]	<220>		
[1858]	<223>	合成序列	
[1859]	<400>	116	
[1860]	ggauuuuagac caccgcaaaa augaagggca cuaaaaca		38
[1861]	<210>	117	
[1862]	<211>	38	
[1863]	<212>	RNA	
[1864]	<213>	人工序列	
[1865]	<220>		
[1866]	<223>	合成序列	
[1867]	<400>	117	
[1868]	ggauuuuagac cagccaccaaaa augaagggca cuaaaaca		38
[1869]	<210>	118	
[1870]	<211>	38	

[1871]	<212>	RNA	
[1872]	<213>	人工序列	
[1873]	<220>		
[1874]	<223>	合成序列	
[1875]	<400>	118	
[1876]	ggauuuagac caccacaaaa augaagugga cuaaaaca		38
[1877]	<210>	119	
[1878]	<211>	38	
[1879]	<212>	RNA	
[1880]	<213>	人工序列	
[1881]	<220>		
[1882]	<223>	合成序列	
[1883]	<400>	119	
[1884]	ggauuuagac cacaccaaaa augaagguga cuaaaaca		38
[1885]	<210>	120	
[1886]	<211>	38	
[1887]	<212>	RNA	
[1888]	<213>	人工序列	
[1889]	<220>		
[1890]	<223>	合成序列	
[1891]	<400>	120	
[1892]	ggauuuagac cacuccaaaa augaaggaga cuaaaaca		38
[1893]	<210>	121	
[1894]	<211>	38	
[1895]	<212>	RNA	
[1896]	<213>	人工序列	
[1897]	<220>		
[1898]	<223>	合成序列	
[1899]	<400>	121	
[1900]	ggauuuagac caccaccaaaa augaaggggc auaaaaca		38
[1901]	<210>	122	
[1902]	<211>	38	
[1903]	<212>	RNA	
[1904]	<213>	人工序列	
[1905]	<220>		
[1906]	<223>	合成序列	
[1907]	<400>	122	
[1908]	ggauuuagac caccaccaaaa augaagggga cgcaaaca		38
[1909]	<210>	123	

[1910]	<211>	38	
[1911]	<212>	RNA	
[1912]	<213>	人工序列	
[1913]	<220>		
[1914]	<223>	合成序列	
[1915]	<400>	123	
[1916]	ggauuuagac cacccecaaaa augaaggga cuagcaca		38
[1917]	<210>	124	
[1918]	<211>	38	
[1919]	<212>	RNA	
[1920]	<213>	人工序列	
[1921]	<220>		
[1922]	<223>	合成序列	
[1923]	<400>	124	
[1924]	ggauuuagac cacccecaaaa augaaggga cuaaagua		38
[1925]	<210>	125	
[1926]	<211>	60	
[1927]	<212>	RNA	
[1928]	<213>	人工序列	
[1929]	<220>		
[1930]	<223>	合成序列	
[1931]	<400>	125	
[1932]	ggcucauuug caggguuuag ccaaaagggu caucaucucu gccccucug cugaugcccc		60
[1933]	<210>	126	
[1934]	<211>	60	
[1935]	<212>	RNA	
[1936]	<213>	人工序列	
[1937]	<220>		
[1938]	<223>	合成序列	
[1939]	<400>	126	
[1940]	ggcacacccg caggguuuag ccaaaagggu caucaucucu gccccucug cugaugcccc		60
[1941]	<210>	127	
[1942]	<211>	60	
[1943]	<212>	RNA	
[1944]	<213>	人工序列	
[1945]	<220>		
[1946]	<223>	合成序列	
[1947]	<400>	127	
[1948]	ggcacacccg caggggggag ccaaaagggu caucaucucu gccccucug cugaugcccc		60

[1949]	<210>	128	
[1950]	<211>	62	
[1951]	<212>	RNA	
[1952]	<213>	人工序列	
[1953]	<220>		
[1954]	<223>	合成序列	
[1955]	<400>	128	
[1956]	gggaacccca aggccaaccg cgagaagaug acccagauca uguuugagac cuucaacacc	60	
[1957]	cc	62	
[1958]	<210>	129	
[1959]	<211>	70	
[1960]	<212>	RNA	
[1961]	<213>	人工序列	
[1962]	<220>		
[1963]	<223>	合成序列	
[1964]	<400>	129	
[1965]	ggccugacug cucucauuug caguugggag ccaaaagggg caucaucucu gccccucug	60	
[1966]	cugaugcccc	70	
[1967]	<210>	130	
[1968]	<211>	80	
[1969]	<212>	RNA	
[1970]	<213>	人工序列	
[1971]	<220>		
[1972]	<223>	合成序列	
[1973]	<400>	130	
[1974]	ggaccuguga auccugacug cucucauuug caguugggag ccaaaagggg caucaucucu	60	
[1975]	gccccucug cugaugcccc	80	
[1976]	<210>	131	
[1977]	<211>	60	
[1978]	<212>	RNA	
[1979]	<213>	人工序列	
[1980]	<220>		
[1981]	<223>	合成序列	
[1982]	<400>	131	
[1983]	ggcucauuug caggguuuag ccaaaagggg caucaucucu gccccucug cugaugcccc	60	
[1984]	<210>	132	
[1985]	<211>	31	
[1986]	<212>	RNA	
[1987]	<213>	人工序列	

[1988]	<220>	
[1989]	<223>	合成序列
[1990]	<400>	132
[1991]	gaccacccca aaaaugaagg ggacuaaaac a	31
[1992]	<210>	133
[1993]	<211>	31
[1994]	<212>	RNA
[1995]	<213>	人工序列
[1996]	<220>	
[1997]	<223>	合成序列
[1998]	<400>	133
[1999]	ggccacccca aaaaugaagg ggacuaaaac a	31
[2000]	<210>	134
[2001]	<211>	33
[2002]	<212>	RNA
[2003]	<213>	人工序列
[2004]	<220>	
[2005]	<223>	合成序列
[2006]	<400>	134
[2007]	gggaccaccc caaaaugaa ggggacuaaa aca	33
[2008]	<210>	135
[2009]	<211>	29
[2010]	<212>	RNA
[2011]	<213>	人工序列
[2012]	<220>	
[2013]	<223>	合成序列
[2014]	<400>	135
[2015]	ccaccccaaa aaugaagggg acuaaaaca	29
[2016]	<210>	136
[2017]	<211>	30
[2018]	<212>	RNA
[2019]	<213>	人工序列
[2020]	<220>	
[2021]	<223>	合成序列
[2022]	<400>	136
[2023]	ggauuuagac caccceaaaa augaagggga	30
[2024]	<210>	137
[2025]	<211>	30
[2026]	<212>	RNA

[2027]	<213>	人工序列	
[2028]	<220>		
[2029]	<223>	合成序列	
[2030]	<400>	137	
[2031]	gguaagagac uaccucuaua ugaaagagga		30
[2032]	<210>	138	
[2033]	<211>	30	
[2034]	<212>	RNA	
[2035]	<213>	人工序列	
[2036]	<220>		
[2037]	<223>	合成序列	
[2038]	<400>	138	
[2039]	ggauauagac caccceaaua ucgaagggga		30
[2040]	<210>	139	
[2041]	<211>	68	
[2042]	<212>	DNA	
[2043]	<213>	人工序列	
[2044]	<220>		
[2045]	<223>	合成序列	
[2046]	<400>	139	
[2047]	atagaccacc ccaatatcga aggggactaa aactagattg ctgttctacc aagtaatcca		60
[2048]	tattgata		68
[2049]	<210>	140	
[2050]	<211>	69	
[2051]	<212>	DNA	
[2052]	<213>	人工序列	
[2053]	<220>		
[2054]	<223>	合成序列	
[2055]	<400>	140	
[2056]	atagaccacc ccaatatcga aggggactaa aacgacaaat ctatctgaat aaactcttct		60
[2057]	ttttcgata		69
[2058]	<210>	141	
[2059]	<211>	69	
[2060]	<212>	DNA	
[2061]	<213>	人工序列	
[2062]	<220>		
[2063]	<223>	合成序列	
[2064]	<400>	141	
[2065]	atagaccacc ccaatatcga aggggactaa aactctaaag aattatctat tctgtctttt		60

[2066]	aaattgata	69
[2067]	<210> 142	
[2068]	<211> 57	
[2069]	<212> DNA	
[2070]	<213> 人工序列	
[2071]	<220>	
[2072]	<223> 合成序列	
[2073]	<400> 142	
[2074]	gtaagagact acctctatat gaaagaggac taaaactttt aacagtggcc ttattaa	57
[2075]	<210> 143	
[2076]	<211> 38	
[2077]	<212> RNA	
[2078]	<213> 人工序列	
[2079]	<220>	
[2080]	<223> 合成序列	
[2081]	<400> 143	
[2082]	ggauucgauc caccceaaaa augaagggga cuaaaaca	38
[2083]	<210> 144	
[2084]	<211> 38	
[2085]	<212> RNA	
[2086]	<213> 人工序列	
[2087]	<220>	
[2088]	<223> 合成序列	
[2089]	<400> 144	
[2090]	ggauuuuaua aaccceaaaa augaagggga cuaaaaca	38
[2091]	<210> 145	
[2092]	<211> 38	
[2093]	<212> RNA	
[2094]	<213> 人工序列	
[2095]	<220>	
[2096]	<223> 合成序列	
[2097]	<400> 145	
[2098]	ggauuuagga agccceaaaa augaagggga cuaaaaca	38
[2099]	<210> 146	
[2100]	<211> 38	
[2101]	<212> RNA	
[2102]	<213> 人工序列	
[2103]	<220>	
[2104]	<223> 合成序列	

[2105]	<400> 146	
[2106]	ggauuuagac caggccaaaa augaaggcca cuaaaaca	38
[2107]	<210> 147	
[2108]	<211> 31	
[2109]	<212> RNA	
[2110]	<213> 人工序列	
[2111]	<220>	
[2112]	<223> 合成序列	
[2113]	<400> 147	
[2114]	ggccacccca aaaaugaagg ggacuaaaac a	31
[2115]	<210> 148	
[2116]	<211> 33	
[2117]	<212> RNA	
[2118]	<213> 人工序列	
[2119]	<220>	
[2120]	<223> 合成序列	
[2121]	<400> 148	
[2122]	gggaccaccc caaaaugaa ggggacuaaa aca	33
[2123]	<210> 149	
[2124]	<211> 29	
[2125]	<212> RNA	
[2126]	<213> 人工序列	
[2127]	<220>	
[2128]	<223> 合成序列	
[2129]	<400> 149	
[2130]	ccaccccaaa aaugaagggg acuaaaaca	29
[2131]	<210> 150	
[2132]	<211> 16	
[2133]	<212> RNA	
[2134]	<213> 人工序列	
[2135]	<220>	
[2136]	<223> 合成序列	
[2137]	<400> 150	
[2138]	ggauuuagac caccce	16
[2139]	<210> 151	
[2140]	<211> 17	
[2141]	<212> RNA	
[2142]	<213> 人工序列	
[2143]	<220>	

[2144]	<223>	合成序列	
[2145]	<400>	151	
[2146]	gguaagagac	uaccucu	17
[2147]	<210>	152	
[2148]	<211>	17	
[2149]	<212>	RNA	
[2150]	<213>	人工序列	
[2151]	<220>		
[2152]	<223>	合成序列	
[2153]	<400>	152	
[2154]	ggcuugugaa	uuaucce	17
[2155]	<210>	153	
[2156]	<211>	15	
[2157]	<212>	RNA	
[2158]	<213>	人工序列	
[2159]	<220>		
[2160]	<223>	合成序列	
[2161]	<400>	153	
[2162]	ggucacauca	ccgcc	15
[2163]	<210>	154	
[2164]	<211>	16	
[2165]	<212>	RNA	
[2166]	<213>	人工序列	
[2167]	<220>		
[2168]	<223>	合成序列	
[2169]	<400>	154	
[2170]	ggauauagac	cacccc	16
[2171]	<210>	155	
[2172]	<211>	16	
[2173]	<212>	RNA	
[2174]	<213>	人工序列	
[2175]	<220>		
[2176]	<223>	合成序列	
[2177]	<400>	155	
[2178]	ggauuuagag	uaccuc	16
[2179]	<210>	156	
[2180]	<211>	19	
[2181]	<212>	RNA	
[2182]	<213>	人工序列	

[2183]	<220>	
[2184]	<223>	合成序列
[2185]	<400>	156
[2186]	gggcuggaga agauagccc	19
[2187]	<210>	157
[2188]	<211>	15
[2189]	<212>	RNA
[2190]	<213>	人工序列
[2191]	<220>	
[2192]	<223>	合成序列
[2193]	<400>	157
[2194]	gggugaauac agcuc	15
[2195]	<210>	158
[2196]	<211>	18
[2197]	<212>	RNA
[2198]	<213>	人工序列
[2199]	<220>	
[2200]	<223>	合成序列
[2201]	<400>	158
[2202]	ggguuuggag aacagccc	18
[2203]	<210>	159
[2204]	<211>	16
[2205]	<212>	RNA
[2206]	<213>	人工序列
[2207]	<220>	
[2208]	<223>	合成序列
[2209]	<400>	159
[2210]	ggauauagac cacccc	16
[2211]	<210>	160
[2212]	<211>	13
[2213]	<212>	RNA
[2214]	<213>	人工序列
[2215]	<220>	
[2216]	<223>	合成序列
[2217]	<400>	160
[2218]	gguaacaauccccc	13
[2219]	<210>	161
[2220]	<211>	38
[2221]	<212>	RNA

[2222]	<213>	人工序列	
[2223]	<220>		
[2224]	<223>	合成序列	
[2225]	<400>	161	
[2226]	ggauuuagac caccceaaaa augaaggga cuaaaaca		38
[2227]	<210>	162	
[2228]	<211>	37	
[2229]	<212>	RNA	
[2230]	<213>	人工序列	
[2231]	<220>		
[2232]	<223>	合成序列	
[2233]	<400>	162	
[2234]	gguaagagac uaccucuaua ugaaagagga cuaaaac		37
[2235]	<210>	163	
[2236]	<211>	38	
[2237]	<212>	RNA	
[2238]	<213>	人工序列	
[2239]	<220>		
[2240]	<223>	合成序列	
[2241]	<400>	163	
[2242]	ggcuugugaa uuaucccaaa auugaaggga acuacaac		38
[2243]	<210>	164	
[2244]	<211>	36	
[2245]	<212>	RNA	
[2246]	<213>	人工序列	
[2247]	<220>		
[2248]	<223>	合成序列	
[2249]	<400>	164	
[2250]	ggucacauca ccgccaagac gacggcggac ugaacc		36
[2251]	<210>	165	
[2252]	<211>	39	
[2253]	<212>	RNA	
[2254]	<213>	人工序列	
[2255]	<220>		
[2256]	<223>	合成序列	
[2257]	<400>	165	
[2258]	ggauauagac caccceaaaua ucgaaggga cuaaaacuu		39
[2259]	<210>	166	
[2260]	<211>	37	

[2261]	<212>	RNA	
[2262]	<213>	人工序列	
[2263]	<220>		
[2264]	<223>	合成序列	
[2265]	<400>	166	
[2266]		ggauuuagag uaccucaaaa caaaagagga cuaaaac	37
[2267]	<210>	167	
[2268]	<211>	37	
[2269]	<212>	RNA	
[2270]	<213>	人工序列	
[2271]	<220>		
[2272]	<223>	合成序列	
[2273]	<400>	167	
[2274]		gggcuggaga agauagccca agaaagaggg caauaac	37
[2275]	<210>	168	
[2276]	<211>	33	
[2277]	<212>	RNA	
[2278]	<213>	人工序列	
[2279]	<220>		
[2280]	<223>	合成序列	
[2281]	<400>	168	
[2282]		gggugaauac agcucgauau agugagcaau aag	33
[2283]	<210>	169	
[2284]	<211>	37	
[2285]	<212>	RNA	
[2286]	<213>	人工序列	
[2287]	<220>		
[2288]	<223>	合成序列	
[2289]	<400>	169	
[2290]		ggguuuggag aacagcccga uauagagggc aaugac	37
[2291]	<210>	170	
[2292]	<211>	37	
[2293]	<212>	RNA	
[2294]	<213>	人工序列	
[2295]	<220>		
[2296]	<223>	合成序列	
[2297]	<400>	170	
[2298]		ggauauagac caccccaaua ucgaagggga cuaaaac	37
[2299]	<210>	171	

[2300]	<211>	31	
[2301]	<212>	RNA	
[2302]	<213>	人工序列	
[2303]	<220>		
[2304]	<223>	合成序列	
[2305]	<400>	171	
[2306]	gguaacaau	cccguagaca	ggggaacugu g 31
[2307]	<210>	172	
[2308]	<211>	34	
[2309]	<212>	RNA	
[2310]	<213>	人工序列	
[2311]	<220>		
[2312]	<223>	合成序列	
[2313]	<400>	172	
[2314]	ggagguaaca	aucccccguag	acaggggaac ugug 34
[2315]	<210>	173	
[2316]	<211>	25	
[2317]	<212>	PRT	
[2318]	<213>	人工序列	
[2319]	<220>		
[2320]	<223>	合成序列	
[2321]	<400>	173	
[2322]	Ala Phe Cys His Phe Val Glu Ile Glu Met Ser Gln Leu Leu Lys Asn		
[2323]	1	5	10 15
[2324]	Tyr Val Tyr Lys Arg Leu Ser Asn Ile		
[2325]		20	25
[2326]	<210>	174	
[2327]	<211>	26	
[2328]	<212>	PRT	
[2329]	<213>	人工序列	
[2330]	<220>		
[2331]	<223>	合成序列	
[2332]	<400>	174	
[2333]	Ile Ile Tyr Arg Tyr Leu Lys Gly Arg Ile Glu Lys Ile Leu Val Asn		
[2334]	1	5	10 15
[2335]	Glu Gln Lys Val Arg Leu Lys Lys Met Glu		
[2336]		20	25
[2337]	<210>	175	
[2338]	<211>	22	

[2339]	<212>	PRT
[2340]	<213>	人工序列
[2341]	<220>	
[2342]	<223>	合成序列
[2343]	<400>	175
[2344]	Leu Phe Leu Lys Glu Val Glu Gln Tyr Phe Lys Lys Tyr Phe Pro Val	
[2345]	1 5 10 15	
[2346]	Lys Ser Lys His Ser Asn	
[2347]	20	
[2348]	<210>	176
[2349]	<211>	19
[2350]	<212>	PRT
[2351]	<213>	人工序列
[2352]	<220>	
[2353]	<223>	合成序列
[2354]	<400>	176
[2355]	Tyr Trp Ile His His Ile Glu Asn Ala Val Glu Arg Ile Leu Lys Asn	
[2356]	1 5 10 15	
[2357]	Cys Lys Ala	
[2358]	<210>	177
[2359]	<211>	22
[2360]	<212>	PRT
[2361]	<213>	人工序列
[2362]	<220>	
[2363]	<223>	合成序列
[2364]	<400>	177
[2365]	Gln Phe Asn Ile Glu Ile Arg Lys His Leu Glu Thr Tyr Phe Pro Ile	
[2366]	1 5 10 15	
[2367]	Lys Lys Thr Asn Arg Lys	
[2368]	20	
[2369]	<210>	178
[2370]	<211>	22
[2371]	<212>	PRT
[2372]	<213>	人工序列
[2373]	<220>	
[2374]	<223>	合成序列
[2375]	<400>	178
[2376]	Ile Tyr His Leu Glu Val Val Lys Tyr Leu Glu His Tyr Phe Pro Ile	
[2377]	1 5 10 15	

[2378]	Lys Thr Ser Lys Arg Arg
[2379]	20
[2380]	<210> 179
[2381]	<211> 22
[2382]	<212> PRT
[2383]	<213> 人工序列
[2384]	<220>
[2385]	<223> 合成序列
[2386]	<400> 179
[2387]	Ile Tyr Asn Leu Glu Leu Lys Lys Tyr Ile Glu Asn Asn Phe Ser Tyr
[2388]	1 5 10 15
[2389]	Lys Lys Gln Lys Ser Lys
[2390]	20
[2391]	<210> 180
[2392]	<211> 19
[2393]	<212> PRT
[2394]	<213> 人工序列
[2395]	<220>
[2396]	<223> 合成序列
[2397]	<400> 180
[2398]	Phe Phe Ile His Arg Ile Glu Tyr Gly Val Glu Lys Ile Tyr Ala Asn
[2399]	1 5 10 15
[2400]	Leu Lys Gln
[2401]	<210> 181
[2402]	<211> 16
[2403]	<212> PRT
[2404]	<213> 人工序列
[2405]	<220>
[2406]	<223> 合成序列
[2407]	<400> 181
[2408]	Ala Leu His Asp Glu Val Lys Lys Thr Tyr Lys Arg Leu Cys Ala Arg
[2409]	1 5 10 15
[2410]	<210> 182
[2411]	<211> 19
[2412]	<212> PRT
[2413]	<213> 人工序列
[2414]	<220>
[2415]	<223> 合成序列
[2416]	<400> 182

[2417]	Phe Trp Ile His His Ile Glu Ser Ala Val Glu Arg Ile Leu Ala Ser
[2418]	1 5 10 15
[2419]	Ile Asn Pro
[2420]	<210> 183
[2421]	<211> 22
[2422]	<212> PRT
[2423]	<213> 人工序列
[2424]	<220>
[2425]	<223> 合成序列
[2426]	<400> 183
[2427]	Gln Tyr Gly Ile Glu Val Lys Lys Tyr Ile Gln Arg Tyr Phe Pro Ile
[2428]	1 5 10 15
[2429]	Lys Arg Ala Pro Asn Arg
[2430]	20
[2431]	<210> 184
[2432]	<211> 26
[2433]	<212> PRT
[2434]	<213> 人工序列
[2435]	<220>
[2436]	<223> 合成序列
[2437]	<400> 184
[2438]	Glu Asn Leu Arg Lys Leu Leu Ser Tyr Asp Arg Lys Leu Lys Asn Ala
[2439]	1 5 10 15
[2440]	Val Met Lys Ser Val Val Asp Ile Leu Lys
[2441]	20 25
[2442]	<210> 185
[2443]	<211> 25
[2444]	<212> PRT
[2445]	<213> 人工序列
[2446]	<220>
[2447]	<223> 合成序列
[2448]	<400> 185
[2449]	Asp Arg Val Ser Asn Leu Leu Ser Tyr Ser Thr Arg Tyr Asn Asn Ser
[2450]	1 5 10 15
[2451]	Thr Tyr Ala Ser Val Phe Glu Val Phe
[2452]	20 25
[2453]	<210> 186
[2454]	<211> 26
[2455]	<212> PRT

[2456]	<213>	人工序列
[2457]	<220>	
[2458]	<223>	合成序列
[2459]	<400>	186
[2460]	Asn Ser Leu Arg Ile Leu Leu Ala Tyr Asp Arg Lys Arg Gln Asn Ala	
[2461]	1	5 10 15
[2462]	Val Thr Lys Thr Ile Asn Asp Leu Leu Leu	
[2463]		20 25
[2464]	<210>	187
[2465]	<211>	27
[2466]	<212>	PRT
[2467]	<213>	人工序列
[2468]	<220>	
[2469]	<223>	合成序列
[2470]	<400>	187
[2471]	Ser Glu Val Phe Asp Arg Phe Phe Thr Tyr Asp Ile Lys Tyr Gln Lys	
[2472]	1	5 10 15
[2473]	Asn Val Leu Asn Leu Leu Gln Asn Ile Leu Leu	
[2474]		20 25
[2475]	<210>	188
[2476]	<211>	26
[2477]	<212>	PRT
[2478]	<213>	人工序列
[2479]	<220>	
[2480]	<223>	合成序列
[2481]	<400>	188
[2482]	Asp Asp Ala Arg Asp Val Leu Ser Tyr Asp Arg Lys Leu Lys Asn Ala	
[2483]	1	5 10 15
[2484]	Val Ser Lys Ser Leu Lys Glu Ile Leu Ser	
[2485]		20 25
[2486]	<210>	189
[2487]	<211>	26
[2488]	<212>	PRT
[2489]	<213>	人工序列
[2490]	<220>	
[2491]	<223>	合成序列
[2492]	<400>	189
[2493]	Asn Glu Leu Arg Glu Leu Leu His Tyr Asp Arg Lys Leu Lys Asn Ala	
[2494]	1	5 10 15

[2495]	Val Ser Lys Ala Phe Ile Asp Leu Phe Asp
[2496]	20 25
[2497]	<210> 190
[2498]	<211> 26
[2499]	<212> PRT
[2500]	<213> 人工序列
[2501]	<220>
[2502]	<223> 合成序列
[2503]	<400> 190
[2504]	Asn Leu Leu Ile Lys Leu Phe Ser Tyr Asp Lys Lys Val Gln Asn His
[2505]	1 5 10 15
[2506]	Ile Leu Lys Ser Thr Lys Thr Leu Leu Glu
[2507]	20 25
[2508]	<210> 191
[2509]	<211> 27
[2510]	<212> PRT
[2511]	<213> 人工序列
[2512]	<220>
[2513]	<223> 合成序列
[2514]	<400> 191
[2515]	Ser Glu Val Phe Asp Arg Phe Phe Thr Tyr Asp Leu Lys Tyr Arg Lys
[2516]	1 5 10 15
[2517]	Asn Val Pro Thr Ile Leu Tyr Asn Ile Leu Leu
[2518]	20 25
[2519]	<210> 192
[2520]	<211> 26
[2521]	<212> PRT
[2522]	<213> 人工序列
[2523]	<220>
[2524]	<223> 合成序列
[2525]	<400> 192
[2526]	Asn Arg Ala Arg Glu Met Met Ala Tyr Asp Arg Lys Arg Lys Asn Ala
[2527]	1 5 10 15
[2528]	Val Pro Arg Ser Ile Leu Asp Met Leu Ala
[2529]	20 25
[2530]	<210> 193
[2531]	<211> 27
[2532]	<212> PRT
[2533]	<213> 人工序列

[2534]	<220>	
[2535]	<223>	合成序列
[2536]	<400>	193
[2537]	Ser Glu Ile His Asp Arg Phe Phe Thr Tyr Asp Met Lys Tyr Gln Lys	
[2538]	1	5 10 15
[2539]	Asn Val Ala Asn Met Leu Glu Asn Ile Leu Leu	
[2540]		20 25
[2541]	<210>	194
[2542]	<211>	26
[2543]	<212>	PRT
[2544]	<213>	人工序列
[2545]	<220>	
[2546]	<223>	合成序列
[2547]	<400>	194
[2548]	Glu Arg Leu Arg Glu Ile Phe Lys Tyr Asp Arg Lys Leu Lys Asn Ala	
[2549]	1	5 10 15
[2550]	Val Ser Lys Ser Leu Ile Asp Ile Leu Asp	
[2551]		20 25
[2552]	<210>	195
[2553]	<211>	29
[2554]	<212>	PRT
[2555]	<213>	人工序列
[2556]	<220>	
[2557]	<223>	合成序列
[2558]	<400>	195
[2559]	Tyr Ala Phe Cys His Phe Val Glu Ile Glu Met Ser Gln Leu Leu Lys	
[2560]	1	5 10 15
[2561]	Asn Tyr Val Tyr Lys Arg Leu Ser Asn Ile Ser Asn Asp	
[2562]		20 25
[2563]	<210>	196
[2564]	<211>	27
[2565]	<212>	PRT
[2566]	<213>	人工序列
[2567]	<220>	
[2568]	<223>	合成序列
[2569]	<400>	196
[2570]	Ile Ile Tyr Arg Tyr Leu Lys Gly Arg Ile Glu Lys Ile Leu Val Asn	
[2571]	1	5 10 15
[2572]	Glu Gln Lys Val Arg Leu Lys Lys Met Glu Lys	

[illegible]

[2612]	<400>	200
[2613]	Ile Tyr His Leu Glu Val Val Lys Tyr Leu Glu His Tyr Phe Pro Ile	
[2614]	1	5 10 15
[2615]	Lys Thr Ser Lys Arg Arg Asn Thr Ala	
[2616]	20	25
[2617]	<210>	201
[2618]	<211>	25
[2619]	<212>	PRT
[2620]	<213>	人工序列
[2621]	<220>	
[2622]	<223>	合成序列
[2623]	<400>	201
[2624]	Ile Tyr Asn Leu Glu Leu Lys Lys Tyr Ile Glu Asn Asn Phe Ser Tyr	
[2625]	1	5 10 15
[2626]	Lys Lys Gln Lys Ser Lys Ser Lys Asn	
[2627]	20	25
[2628]	<210>	202
[2629]	<211>	27
[2630]	<212>	PRT
[2631]	<213>	人工序列
[2632]	<220>	
[2633]	<223>	合成序列
[2634]	<400>	202
[2635]	Ile Ile Tyr Arg Tyr Leu Lys Gly Arg Ile Glu Lys Ile Leu Val Asn	
[2636]	1	5 10 15
[2637]	Glu Gln Lys Val Arg Leu Lys Lys Met Glu Lys	
[2638]	20	25
[2639]	<210>	203
[2640]	<211>	29
[2641]	<212>	PRT
[2642]	<213>	人工序列
[2643]	<220>	
[2644]	<223>	合成序列
[2645]	<400>	203
[2646]	Tyr Ala Phe Cys His Phe Val Glu Ile Glu Met Ser Gln Leu Leu Lys	
[2647]	1	5 10 15
[2648]	Asn Tyr Val Tyr Lys Arg Leu Ser Asn Ile Ser Asn Asp	
[2649]	20	25
[2650]	<210>	204

[2651]	<211>	31
[2652]	<212>	PRT
[2653]	<213>	人工序列
[2654]	<220>	
[2655]	<223>	合成序列
[2656]	<400>	204
[2657]	Phe Ala Phe Cys Tyr Phe Val Glu Ile Glu Val Asn Asn Leu Leu Lys	
[2658]	1	5 10 15
[2659]	Glu Asn Val Tyr Lys Ile Lys Arg Phe Asn Glu Ser Asn Lys Lys	
[2660]	20	25 30
[2661]	<210>	205
[2662]	<211>	19
[2663]	<212>	PRT
[2664]	<213>	人工序列
[2665]	<220>	
[2666]	<223>	合成序列
[2667]	<400>	205
[2668]	Phe Phe Ile His Arg Ile Glu Tyr Gly Val Glu Lys Ile Tyr Ala Asn	
[2669]	1	5 10 15
[2670]	Leu Lys Gln	
[2671]	<210>	206
[2672]	<211>	15
[2673]	<212>	PRT
[2674]	<213>	人工序列
[2675]	<220>	
[2676]	<223>	合成序列
[2677]	<400>	206
[2678]	Ala Leu His Asp Glu Val Lys Lys Thr Tyr Lys Arg Leu Cys Ala	
[2679]	1	5 10 15
[2680]	<210>	207
[2681]	<211>	19
[2682]	<212>	PRT
[2683]	<213>	人工序列
[2684]	<220>	
[2685]	<223>	合成序列
[2686]	<400>	207
[2687]	Phe Trp Ile His Arg Ile Glu Asn Ala Val Glu Arg Ile Leu Gly Ser	
[2688]	1	5 10 15
[2689]	Ile Asn Asp	

[2690]	<210>	208
[2691]	<211>	19
[2692]	<212>	PRT
[2693]	<213>	人工序列
[2694]	<220>	
[2695]	<223>	合成序列
[2696]	<400>	208
[2697]	Phe Trp Ile His His Ile Glu Ser Ala Val Glu Arg Ile Leu Ala Ser	
[2698]	1 5 10 15	
[2699]	Ile Asn Pro	
[2700]	<210>	209
[2701]	<211>	23
[2702]	<212>	PRT
[2703]	<213>	人工序列
[2704]	<220>	
[2705]	<223>	合成序列
[2706]	<400>	209
[2707]	Gln Phe Asn Ile Glu Ile Gly Lys Tyr Phe Glu His Tyr Phe Pro Ile	
[2708]	1 5 10 15	
[2709]	Lys Lys Glu Arg Cys Thr Glu	
[2710]	20	
[2711]	<210>	210
[2712]	<211>	22
[2713]	<212>	PRT
[2714]	<213>	人工序列
[2715]	<220>	
[2716]	<223>	合成序列
[2717]	<400>	210
[2718]	Gln Phe Asn Ile Glu Leu Gly Lys Tyr Phe Glu His Tyr Phe Pro Lys	
[2719]	1 5 10 15	
[2720]	Thr Gly Lys Lys Glu Ser	
[2721]	20	
[2722]	<210>	211
[2723]	<211>	25
[2724]	<212>	PRT
[2725]	<213>	人工序列
[2726]	<220>	
[2727]	<223>	合成序列
[2728]	<400>	211

[2729]	Gln Tyr Gly Ile Glu Val Lys Lys Tyr Ile Gln Arg Tyr Phe Pro Ile
[2730]	1 5 10 15
[2731]	Lys Arg Ala Pro Asn Arg Ser Lys His
[2732]	20 25
[2733]	<210> 212
[2734]	<211> 19
[2735]	<212> PRT
[2736]	<213> 人工序列
[2737]	<220>
[2738]	<223> 合成序列
[2739]	<400> 212
[2740]	Phe Trp Ile His Leu Ile Glu Asn Glu Val Glu Arg Leu Tyr Asn Gly
[2741]	1 5 10 15
[2742]	Ile Glu Asn
[2743]	<210> 213
[2744]	<211> 29
[2745]	<212> PRT
[2746]	<213> 人工序列
[2747]	<220>
[2748]	<223> 合成序列
[2749]	<400> 213
[2750]	Tyr Val Phe Cys His Phe Val Glu Ile Glu Met Ser Lys Leu Leu Lys
[2751]	1 5 10 15
[2752]	Asn Tyr Val Tyr Lys Lys Pro Ser Asn Ile Ser Asn Asp
[2753]	20 25
[2754]	<210> 214
[2755]	<211> 26
[2756]	<212> PRT
[2757]	<213> 人工序列
[2758]	<220>
[2759]	<223> 合成序列
[2760]	<400> 214
[2761]	Glu Asn Leu Arg Lys Leu Leu Ser Tyr Asp Arg Lys Leu Lys Asn Ala
[2762]	1 5 10 15
[2763]	Val Met Lys Ser Val Val Asp Ile Leu Lys
[2764]	20 25
[2765]	<210> 215
[2766]	<211> 25
[2767]	<212> PRT

[2768]	<213>	人工序列
[2769]	<220>	
[2770]	<223>	合成序列
[2771]	<400>	215
[2772]	Asp Arg Val Ser Asn Leu Leu Ser Tyr Ser Thr Arg Tyr Asn Asn Ser	
[2773]	1 5 10 15	
[2774]	Thr Tyr Ala Ser Val Phe Glu Val Phe	
[2775]	20 25	
[2776]	<210>	216
[2777]	<211>	26
[2778]	<212>	PRT
[2779]	<213>	人工序列
[2780]	<220>	
[2781]	<223>	合成序列
[2782]	<400>	216
[2783]	Asn Ser Leu Arg Ile Leu Leu Ala Tyr Asp Arg Lys Arg Gln Asn Ala	
[2784]	1 5 10 15	
[2785]	Val Thr Lys Thr Ile Asn Asp Leu Leu Leu	
[2786]	20 25	
[2787]	<210>	217
[2788]	<211>	27
[2789]	<212>	PRT
[2790]	<213>	人工序列
[2791]	<220>	
[2792]	<223>	合成序列
[2793]	<400>	217
[2794]	Ser Glu Val Phe Asp Arg Phe Phe Thr Tyr Asp Ile Lys Tyr Gln Lys	
[2795]	1 5 10 15	
[2796]	Asn Val Leu Asn Leu Leu Gln Asn Ile Leu Leu	
[2797]	20 25	
[2798]	<210>	218
[2799]	<211>	26
[2800]	<212>	PRT
[2801]	<213>	人工序列
[2802]	<220>	
[2803]	<223>	合成序列
[2804]	<400>	218
[2805]	Asp Asp Ala Arg Asp Val Leu Ser Tyr Asp Arg Lys Leu Lys Asn Ala	
[2806]	1 5 10 15	

[2807]	Val Ser Lys Ser Leu Lys Glu Ile Leu Ser
[2808]	20 25
[2809]	<210> 219
[2810]	<211> 26
[2811]	<212> PRT
[2812]	<213> 人工序列
[2813]	<220>
[2814]	<223> 合成序列
[2815]	<400> 219
[2816]	Asn Glu Leu Arg Glu Leu Leu His Tyr Asp Arg Lys Leu Lys Asn Ala
[2817]	1 5 10 15
[2818]	Val Ser Lys Ala Phe Ile Asp Leu Phe Asp
[2819]	20 25
[2820]	<210> 220
[2821]	<211> 26
[2822]	<212> PRT
[2823]	<213> 人工序列
[2824]	<220>
[2825]	<223> 合成序列
[2826]	<400> 220
[2827]	Asn Leu Leu Ile Lys Leu Phe Ser Tyr Asp Lys Lys Val Gln Asn His
[2828]	1 5 10 15
[2829]	Ile Leu Lys Ser Thr Lys Thr Leu Leu Glu
[2830]	20 25
[2831]	<210> 221
[2832]	<211> 25
[2833]	<212> PRT
[2834]	<213> 人工序列
[2835]	<220>
[2836]	<223> 合成序列
[2837]	<400> 221
[2838]	Asp Arg Val Ser Asn Leu Leu Ser Tyr Ser Thr Arg Tyr Asn Asn Ser
[2839]	1 5 10 15
[2840]	Thr Tyr Ala Ser Val Phe Glu Val Phe
[2841]	20 25
[2842]	<210> 222
[2843]	<211> 26
[2844]	<212> PRT
[2845]	<213> 人工序列

[2846]	<220>	
[2847]	<223>	合成序列
[2848]	<400>	222
[2849]	Glu Asn Leu Arg Lys Leu Leu Ser Tyr Asp Arg Lys Leu Lys Asn Ala	
[2850]	1	5 10 15
[2851]	Ile Met Lys Ser Ile Val Asp Ile Leu Lys	
[2852]	20	25
[2853]	<210>	223
[2854]	<211>	26
[2855]	<212>	PRT
[2856]	<213>	人工序列
[2857]	<220>	
[2858]	<223>	合成序列
[2859]	<400>	223
[2860]	Glu Glu Leu Arg Asn Leu Leu Lys Tyr Asp Arg Lys Leu Lys Asn Ala	
[2861]	1	5 10 15
[2862]	Val Met Lys Ser Ile Lys Asp Ile Phe Lys	
[2863]	20	25
[2864]	<210>	224
[2865]	<211>	27
[2866]	<212>	PRT
[2867]	<213>	人工序列
[2868]	<220>	
[2869]	<223>	合成序列
[2870]	<400>	224
[2871]	Ser Glu Val Phe Asp Arg Phe Phe Thr Tyr Asp Leu Lys Tyr Arg Lys	
[2872]	1	5 10 15
[2873]	Asn Val Pro Thr Ile Leu Tyr Asn Ile Leu Leu	
[2874]	20	25
[2875]	<210>	225
[2876]	<211>	26
[2877]	<212>	PRT
[2878]	<213>	人工序列
[2879]	<220>	
[2880]	<223>	合成序列
[2881]	<400>	225
[2882]	Asn Arg Ala Arg Glu Met Met Ala Tyr Asp Arg Lys Arg Lys Asn Ala	
[2883]	1	5 10 15
[2884]	Val Pro Arg Ser Ile Leu Asp Met Leu Ala	

[2885]		20		25
[2886]	<210>	226		
[2887]	<211>	27		
[2888]	<212>	PRT		
[2889]	<213>	人工序列		
[2890]	<220>			
[2891]	<223>	合成序列		
[2892]	<400>	226		
[2893]	Ser Glu Phe Phe Asp Arg Phe Phe Thr Tyr Asp Met Lys Tyr Gln Lys			
[2894]	1	5	10	15
[2895]	Asn Val Val Asn Val Leu Glu Asn Ile Leu Leu			
[2896]		20	25	
[2897]	<210>	227		
[2898]	<211>	27		
[2899]	<212>	PRT		
[2900]	<213>	人工序列		
[2901]	<220>			
[2902]	<223>	合成序列		
[2903]	<400>	227		
[2904]	Ser Glu Ile His Asp Arg Phe Phe Thr Tyr Asp Met Lys Tyr Gln Lys			
[2905]	1	5	10	15
[2906]	Asn Val Ala Asn Met Leu Glu Asn Ile Leu Leu			
[2907]		20	25	
[2908]	<210>	228		
[2909]	<211>	26		
[2910]	<212>	PRT		
[2911]	<213>	人工序列		
[2912]	<220>			
[2913]	<223>	合成序列		
[2914]	<400>	228		
[2915]	Asn Asn Leu Arg Asp Leu Met Ala Tyr Asp Arg Lys Leu Lys Asn Ala			
[2916]	1	5	10	15
[2917]	Val Thr Lys Ser Met Ile Lys Ile Leu Asp			
[2918]		20	25	
[2919]	<210>	229		
[2920]	<211>	26		
[2921]	<212>	PRT		
[2922]	<213>	人工序列		
[2923]	<220>			

[2924]	<223>	合成序列
[2925]	<400>	229
[2926]	Asn Asp Leu Arg Lys Leu Met Ser Tyr Asp Arg Lys Leu Lys Asn Ala	
[2927]	1	5 10 15
[2928]	Val Thr Lys Ala Ile Ile Lys Ile Leu Asp	
[2929]	20	25
[2930]	<210>	230
[2931]	<211>	26
[2932]	<212>	PRT
[2933]	<213>	人工序列
[2934]	<220>	
[2935]	<223>	合成序列
[2936]	<400>	230
[2937]	Glu Arg Leu Arg Glu Ile Phe Lys Tyr Asp Arg Lys Leu Lys Asn Ala	
[2938]	1	5 10 15
[2939]	Val Ser Lys Ser Leu Ile Asp Ile Leu Asp	
[2940]	20	25
[2941]	<210>	231
[2942]	<211>	27
[2943]	<212>	PRT
[2944]	<213>	人工序列
[2945]	<220>	
[2946]	<223>	合成序列
[2947]	<400>	231
[2948]	Ser Glu Val Phe Asp Arg Phe Phe Thr Tyr Asp Met Lys Leu Arg Lys	
[2949]	1	5 10 15
[2950]	Asn Val Val Asn Met Leu Tyr Asn Ile Leu Leu	
[2951]	20	25
[2952]	<210>	232
[2953]	<211>	26
[2954]	<212>	PRT
[2955]	<213>	人工序列
[2956]	<220>	
[2957]	<223>	合成序列
[2958]	<400>	232
[2959]	Glu Asn Leu Arg Lys Leu Leu Ser Tyr Asp Arg Lys Leu Lys Asn Ala	
[2960]	1	5 10 15
[2961]	Ile Met Lys Ser Ile Val Asp Ile Leu Lys	
[2962]	20	25

[2963]	<210>	233	
[2964]	<211>	24	
[2965]	<212>	RNA	
[2966]	<213>	人工序列	
[2967]	<220>		
[2968]	<223>	合成序列	
[2969]	<400>	233	
[2970]		ggcaugacua gucguacugc aucg	24
[2971]	<210>	234	
[2972]	<211>	31	
[2973]	<212>	RNA	
[2974]	<213>	人工序列	
[2975]	<220>		
[2976]	<223>	合成序列	
[2977]	<400>	234	
[2978]		gguccaaggg ucaucaucuc ugccccaca g	31
[2979]	<210>	235	
[2980]	<211>	38	
[2981]	<212>	DNA	
[2982]	<213>	人工序列	
[2983]	<220>		
[2984]	<223>	合成序列	
[2985]	<400>	235	
[2986]		gatatagacc accccaatat cgaaggggac taaaactt	38
[2987]	<210>	236	
[2988]	<211>	34	
[2989]	<212>	DNA	
[2990]	<213>	人工序列	
[2991]	<220>		
[2992]	<223>	合成序列	
[2993]	<400>	236	
[2994]		tcacatcacc gccaaagacga cggcggactg aacc	34
[2995]	<210>	237	
[2996]	<211>	38	
[2997]	<212>	DNA	
[2998]	<213>	人工序列	
[2999]	<220>		
[3000]	<223>	合成序列	
[3001]	<400>	237	

[3002]	ggagcacatc accgccaaaa tgacggcgga ctgaacct	38
[3003]	<210> 238	
[3004]	<211> 32	
[3005]	<212> DNA	
[3006]	<213> 人工序列	
[3007]	<220>	
[3008]	<223> 合成序列	
[3009]	<400> 238	
[3010]	gtaacaatcc ccgtagacag gggaactgca ac	32
[3011]	<210> 239	
[3012]	<211> 37	
[3013]	<212> DNA	
[3014]	<213> 人工序列	
[3015]	<220>	
[3016]	<223> 合成序列	
[3017]	<400> 239	
[3018]	gatttagacc accccaaaaa tgaaggggac taaaaca	37
[3019]	<210> 240	
[3020]	<211> 36	
[3021]	<212> DNA	
[3022]	<213> 人工序列	
[3023]	<220>	
[3024]	<223> 合成序列	
[3025]	<400> 240	
[3026]	gtaagagact acctctatat gaaagaggac taaaac	36
[3027]	<210> 241	
[3028]	<211> 36	
[3029]	<212> DNA	
[3030]	<213> 人工序列	
[3031]	<220>	
[3032]	<223> 合成序列	
[3033]	<400> 241	
[3034]	cttgtgaatt atcccaaaat tgaagggaac tacaac	36
[3035]	<210> 242	
[3036]	<211> 36	
[3037]	<212> DNA	
[3038]	<213> 人工序列	
[3039]	<220>	
[3040]	<223> 合成序列	

[3041]	<400> 242	
[3042]	cttgtggatt atcccaatat tgaagggaac tacaac	36
[3043]	<210> 243	
[3044]	<211> 36	
[3045]	<212> DNA	
[3046]	<213> 人工序列	
[3047]	<220>	
[3048]	<223> 合成序列	
[3049]	<400> 243	
[3050]	cttgtggatt atcccaaaat tgaagggaac taaaac	36
[3051]	<210> 244	
[3052]	<211> 35	
[3053]	<212> DNA	
[3054]	<213> 人工序列	
[3055]	<220>	
[3056]	<223> 合成序列	
[3057]	<400> 244	
[3058]	gctggagaag atagcccaag aaagagggaac ataac	35
[3059]	<210> 245	
[3060]	<211> 35	
[3061]	<212> DNA	
[3062]	<213> 人工序列	
[3063]	<220>	
[3064]	<223> 合成序列	
[3065]	<400> 245	
[3066]	gttgatgaga agagcccaag atagagggaac ataac	35
[3067]	<210> 246	
[3068]	<211> 31	
[3069]	<212> DNA	
[3070]	<213> 人工序列	
[3071]	<220>	
[3072]	<223> 合成序列	
[3073]	<400> 246	
[3074]	gtgaaagtag cccgatatag agggcaataa c	31
[3075]	<210> 247	
[3076]	<211> 31	
[3077]	<212> DNA	
[3078]	<213> 人工序列	
[3079]	<220>	

[3080]	<223>	合成序列	
[3081]	<400>	247	
[3082]		gtgaatacag ctcgatatag tgagcaataa g	31
[3083]	<210>	248	
[3084]	<211>	36	
[3085]	<212>	DNA	
[3086]	<213>	人工序列	
[3087]	<220>		
[3088]	<223>	合成序列	
[3089]	<400>	248	
[3090]		gatttagagt acctcaaaac aaaagaggac taaaac	36
[3091]	<210>	249	
[3092]	<211>	35	
[3093]	<212>	DNA	
[3094]	<213>	人工序列	
[3095]	<220>		
[3096]	<223>	合成序列	
[3097]	<400>	249	
[3098]		gtttggagaa cagcccgata tagagggcaa tagac	35
[3099]	<210>	250	
[3100]	<211>	36	
[3101]	<212>	DNA	
[3102]	<213>	人工序列	
[3103]	<220>		
[3104]	<223>	合成序列	
[3105]	<400>	250	
[3106]		gatatagacc accccaatat cgaaggggac taaaac	36
[3107]	<210>	251	
[3108]	<211>	32	
[3109]	<212>	RNA	
[3110]	<213>	人工序列	
[3111]	<220>		
[3112]	<223>	合成序列	
[3113]	<400>	251	
[3114]		gaccacccca auaucgaagg ggacuaaaac uu	32
[3115]	<210>	252	
[3116]	<211>	31	
[3117]	<212>	RNA	
[3118]	<213>	人工序列	

[3119]	<220>	
[3120]	<223>	合成序列
[3121]	<400>	252
[3122]	caucaccgcc aagacgacgg cggacugaac c	31
[3123]	<210>	253
[3124]	<211>	31
[3125]	<212>	RNA
[3126]	<213>	人工序列
[3127]	<220>	
[3128]	<223>	合成序列
[3129]	<400>	253
[3130]	gaccacccca aaaaugaagg ggacuaaaac a	31
[3131]	<210>	254
[3132]	<211>	30
[3133]	<212>	RNA
[3134]	<213>	人工序列
[3135]	<220>	
[3136]	<223>	合成序列
[3137]	<400>	254
[3138]	gacuaccucu auaugaaaga ggacuaaaac	30
[3139]	<210>	255
[3140]	<211>	30
[3141]	<212>	RNA
[3142]	<213>	人工序列
[3143]	<220>	
[3144]	<223>	合成序列
[3145]	<400>	255
[3146]	aaauauccca aaauugaagg gaacuacaac	30
[3147]	<210>	256
[3148]	<211>	27
[3149]	<212>	RNA
[3150]	<213>	人工序列
[3151]	<220>	
[3152]	<223>	合成序列
[3153]	<400>	256
[3154]	agauagccca agaaagaggg caauaac	27
[3155]	<210>	257
[3156]	<211>	27
[3157]	<212>	RNA

[3158]	<213>	人工序列	
[3159]	<220>		
[3160]	<223>	合成序列	
[3161]	<400>	257	
[3162]	gaagagccca agauagaggg caauaac		27
[3163]	<210>	258	
[3164]	<211>	27	
[3165]	<212>	RNA	
[3166]	<213>	人工序列	
[3167]	<220>		
[3168]	<223>	合成序列	
[3169]	<400>	258	
[3170]	aaguagcccg auauagaggg caauaac		27
[3171]	<210>	259	
[3172]	<211>	27	
[3173]	<212>	RNA	
[3174]	<213>	人工序列	
[3175]	<220>		
[3176]	<223>	合成序列	
[3177]	<400>	259	
[3178]	auacagcucg auauagugag caauaag		27
[3179]	<210>	260	
[3180]	<211>	30	
[3181]	<212>	RNA	
[3182]	<213>	人工序列	
[3183]	<220>		
[3184]	<223>	合成序列	
[3185]	<400>	260	
[3186]	gaguaccuca aaacaaaaga ggacuaaaac		30
[3187]	<210>	261	
[3188]	<211>	28	
[3189]	<212>	RNA	
[3190]	<213>	人工序列	
[3191]	<220>		
[3192]	<223>	合成序列	
[3193]	<400>	261	
[3194]	gaacagcccg auauagaggg caauagac		28
[3195]	<210>	262	
[3196]	<211>	27	

[3197]	<212>	RNA	
[3198]	<213>	人工序列	
[3199]	<220>		
[3200]	<223>	合成序列	
[3201]	<400>	262	
[3202]		caccccaua ucgaagggga cuaaaac	27
[3203]	<210>	263	
[3204]	<211>	61	
[3205]	<212>	RNA	
[3206]	<213>	人工序列	
[3207]	<220>		
[3208]	<223>	合成序列	
[3209]	<400>	263	
[3210]		ggguuuggag aacagcccga uauagagggc aaugaccag auauagccug gugguucagg	60
[3211]		c	61
[3212]	<210>	264	
[3213]	<211>	57	
[3214]	<212>	RNA	
[3215]	<213>	人工序列	
[3216]	<220>		
[3217]	<223>	合成序列	
[3218]	<400>	264	
[3219]		gggugaauac agcucgauau agugagcaau aagcagauau agccuggugg uucaggc	57
[3220]	<210>	265	
[3221]	<211>	61	
[3222]	<212>	RNA	
[3223]	<213>	人工序列	
[3224]	<220>		
[3225]	<223>	合成序列	
[3226]	<400>	265	
[3227]		gguauguaac aaucgccgua gacaggggaa cugcaaccag auauagccug gugguucagg	60
[3228]		c	61
[3229]	<210>	266	
[3230]	<211>	61	
[3231]	<212>	RNA	
[3232]	<213>	人工序列	
[3233]	<220>		
[3234]	<223>	合成序列	
[3235]	<400>	266	

[3236]	gggcuggaga agauagccca agaaagaggg caauaaccag auauagccug gugguucagg	60
[3237]	c	61
[3238]	<210> 267	
[3239]	<211> 62	
[3240]	<212> RNA	
[3241]	<213> 人工序列	
[3242]	<220>	
[3243]	<223> 合成序列	
[3244]	<400> 267	
[3245]	ggauuuagac caccceaaaa augaagggga cuaaaacaca gauauagccu ggugguucag	60
[3246]	gc	62
[3247]	<210> 268	
[3248]	<211> 61	
[3249]	<212> RNA	
[3250]	<213> 人工序列	
[3251]	<220>	
[3252]	<223> 合成序列	
[3253]	<400> 268	
[3254]	ggauuuagag uaccucaaaa caaaagagga cuaaaaccag auauagccug gugguucagg	60
[3255]	c	61
[3256]	<210> 269	
[3257]	<211> 61	
[3258]	<212> RNA	
[3259]	<213> 人工序列	
[3260]	<220>	
[3261]	<223> 合成序列	
[3262]	<400> 269	
[3263]	gguaagagac uaccucuaua ugaaagagga cuaaaaccag auauagccug gugguucagg	60
[3264]	c	61
[3265]	<210> 270	
[3266]	<211> 61	
[3267]	<212> RNA	
[3268]	<213> 人工序列	
[3269]	<220>	
[3270]	<223> 合成序列	
[3271]	<400> 270	
[3272]	ggauauagac caccceaaau ucgaagggga cuaaaaccag auauagccug gugguucagg	60
[3273]	c	61
[3274]	<210> 271	

[3275]	<211>	64	
[3276]	<212>	RNA	
[3277]	<213>	人工序列	
[3278]	<220>		
[3279]	<223>	合成序列	
[3280]	<400>	271	
[3281]	gggauauaga ccacccaau aucgaagggg acuaaaacu cagauauagc cuggugguuc	60	
[3282]	aggc		64
[3283]	<210>	272	
[3284]	<211>	62	
[3285]	<212>	RNA	
[3286]	<213>	人工序列	
[3287]	<220>		
[3288]	<223>	合成序列	
[3289]	<400>	272	
[3290]	ggcuugugaa uuaucccaa auugaaggga acuacaacca gauauagccu ggugguucag	60	
[3291]	gc		62
[3292]	<210>	273	
[3293]	<211>	60	
[3294]	<212>	RNA	
[3295]	<213>	人工序列	
[3296]	<220>		
[3297]	<223>	合成序列	
[3298]	<400>	273	
[3299]	ggucacauca ccgccaagac gacggcggac ugaaccacaga uauagccugg ugguucaggc	60	
[3300]	<210>	274	
[3301]	<211>	98	
[3302]	<212>	RNA	
[3303]	<213>	人工序列	
[3304]	<220>		
[3305]	<223>	合成序列	
[3306]	<400>	274	
[3307]	gguguucucu gaugaggccu ucgggccgaa acggugaaag ccguaagaac agccccgauau	60	
[3308]	agagggcaau agaccagaua uagccuggug guucaggc		98
[3309]	<210>	275	
[3310]	<211>	52	
[3311]	<212>	RNA	
[3312]	<213>	人工序列	
[3313]	<220>		

[3314]	<223>	合成序列	
[3315]	<400>	275	
[3316]	gaacagcccg auauagaggg caauagacca gauauagccu ggugguucag gc		52
[3317]	<210>	276	
[3318]	<211>	97	
[3319]	<212>	RNA	
[3320]	<213>	人工序列	
[3321]	<220>		
[3322]	<223>	合成序列	
[3323]	<400>	276	
[3324]	gguguauucu gaugaggccu ucgggccgaa acggugaaag ccguaaauac agcucgauau		60
[3325]	agugagcaau aagcagauau agccuggugg uucaggc		97
[3326]	<210>	277	
[3327]	<211>	51	
[3328]	<212>	RNA	
[3329]	<213>	人工序列	
[3330]	<220>		
[3331]	<223>	合成序列	
[3332]	<400>	277	
[3333]	aaguagcccg auauagaggg caauaaccag auauagccug gugguucagg c		51
[3334]	<210>	278	
[3335]	<211>	97	
[3336]	<212>	RNA	
[3337]	<213>	人工序列	
[3338]	<220>		
[3339]	<223>	合成序列	
[3340]	<400>	278	
[3341]	gguaucuucu gaugaggccu ucgggccgaa acggugaaag ccguaaagau agcccaagaa		60
[3342]	agagggcaau aaccagauau agccuggugg uucaggc		97
[3343]	<210>	279	
[3344]	<211>	51	
[3345]	<212>	RNA	
[3346]	<213>	人工序列	
[3347]	<220>		
[3348]	<223>	合成序列	
[3349]	<400>	279	
[3350]	agauagccca agaaagaggg caauaaccag auauagccug gugguucagg c		51
[3351]	<210>	280	
[3352]	<211>	51	

[3353]	<212>	RNA	
[3354]	<213>	人工序列	
[3355]	<220>		
[3356]	<223>	合成序列	
[3357]	<400>	280	
[3358]	ggccacccca	aaaaugaagg ggacuaaaac acuggugaac uuccgauagu g	51
[3359]	<210>	281	
[3360]	<211>	51	
[3361]	<212>	RNA	
[3362]	<213>	人工序列	
[3363]	<220>		
[3364]	<223>	合成序列	
[3365]	<400>	281	
[3366]	ggccacccca	aaaaugaagg ggacuaaaac acagauauag ccuggugguu c	51
[3367]	<210>	282	
[3368]	<211>	51	
[3369]	<212>	RNA	
[3370]	<213>	人工序列	
[3371]	<220>		
[3372]	<223>	合成序列	
[3373]	<400>	282	
[3374]	gaccacccca	aaaaugaagg ggacuaaaac aggggcagag augaugaccc u	51
[3375]	<210>	283	
[3376]	<211>	54	
[3377]	<212>	RNA	
[3378]	<213>	人工序列	
[3379]	<220>		
[3380]	<223>	合成序列	
[3381]	<400>	283	
[3382]	gaguaccuca	aaacaaaaga ggacuaaaac cagauauagc cuggugguuc aggc	54
[3383]	<210>	284	
[3384]	<211>	55	
[3385]	<212>	RNA	
[3386]	<213>	人工序列	
[3387]	<220>		
[3388]	<223>	合成序列	
[3389]	<400>	284	
[3390]	gggccacccc	aauaucgaag gggacuaaaa ccagauauag ccuggugguu caggc	55
[3391]	<210>	285	

[3392]	<211>	56	
[3393]	<212>	RNA	
[3394]	<213>	人工序列	
[3395]	<220>		
[3396]	<223>	合成序列	
[3397]	<400>	285	
[3398]		gaccacccca auaucgaagg ggacuaaaac uucagauaua gccugguggu ucaggc	56
[3399]	<210>	286	
[3400]	<211>	99	
[3401]	<212>	RNA	
[3402]	<213>	人工序列	
[3403]	<220>		
[3404]	<223>	合成序列	
[3405]	<400>	286	
[3406]		ggauaaauucu gaugaggccu ucgggccgaa acggugaaag ccguaaaaua ucccatauuu	60
[3407]		gaagggaacu acaaccagau auagccuggu gguucaggc	99
[3408]	<210>	287	
[3409]	<211>	54	
[3410]	<212>	RNA	
[3411]	<213>	人工序列	
[3412]	<220>		
[3413]	<223>	合成序列	
[3414]	<400>	287	
[3415]		aaauauccca aaauugaagg gaacuacaac cagauauagc cuggugguuc aggc	54
[3416]	<210>	288	
[3417]	<211>	99	
[3418]	<212>	RNA	
[3419]	<213>	人工序列	
[3420]	<220>		
[3421]	<223>	合成序列	
[3422]	<400>	288	
[3423]		gggugaugcu gaugaggccu ucgggccgaa acggugaaag ccguacauca ccgccaagac	60
[3424]		gacggcggac ugaaccagau auagccuggu gguucaggc	99
[3425]	<210>	289	
[3426]	<211>	46	
[3427]	<212>	RNA	
[3428]	<213>	人工序列	
[3429]	<220>		
[3430]	<223>	合成序列	

[3431]	<400>	289	
[3432]	cacaccgcca agacgacggc ggacgaacca gaaagccggg gcaggc		46
[3433]	<210>	290	
[3434]	<211>	24	
[3435]	<212>	RNA	
[3436]	<213>	人工序列	
[3437]	<220>		
[3438]	<223>	合成序列	
[3439]	<400>	290	
[3440]	ggcaugacua gucguacugc aucg		24
[3441]	<210>	291	
[3442]	<211>	55	
[3443]	<212>	RNA	
[3444]	<213>	人工序列	
[3445]	<220>		
[3446]	<223>	合成序列	
[3447]	<400>	291	
[3448]	ggaaaucauu caacacccgc acuaucggaa guucaccagc cagccgcagc acguu		55
[3449]	<210>	292	
[3450]	<211>	57	
[3451]	<212>	RNA	
[3452]	<213>	人工序列	
[3453]	<220>		
[3454]	<223>	合成序列	
[3455]	<400>	292	
[3456]	ggcaauaaaa augcgccgcc ugaaccacca ggcuauaucu gccacucauu guuguga		57
[3457]	<210>	293	
[3458]	<211>	31	
[3459]	<212>	RNA	
[3460]	<213>	人工序列	
[3461]	<220>		
[3462]	<223>	合成序列	
[3463]	<400>	293	
[3464]	gguccaaggg ucaucaucuc ugccccccaca g		31
[3465]	<210>	294	
[3466]	<211>	60	
[3467]	<212>	RNA	
[3468]	<213>	人工序列	
[3469]	<220>		

[3470]	<223>	合成序列	
[3471]	<400>	294	
[3472]	ggcacacccg caggggggag ccaaaagggg	caucaucucu gcccccacag cagaagcccc	60
[3473]	<210>	295	
[3474]	<211>	43	
[3475]	<212>	RNA	
[3476]	<213>	人工序列	
[3477]	<220>		
[3478]	<223>	合成序列	
[3479]	<400>	295	
[3480]	ggaccauaua ucgaaaguua agcuagaau	ugucauaugg cag	43
[3481]	<210>	296	
[3482]	<211>	62	
[3483]	<212>	RNA	
[3484]	<213>	人工序列	
[3485]	<220>		
[3486]	<223>	合成序列	
[3487]	<400>	296	
[3488]	gggaacccca aggccaaccg cgagaagaug	acccagauca uguuugagac cuucaacacc	60
[3489]	cc		62
[3490]	<210>	297	
[3491]	<211>	56	
[3492]	<212>	RNA	
[3493]	<213>	人工序列	
[3494]	<220>		
[3495]	<223>	合成序列	
[3496]	<400>	297	
[3497]	ggcucauuuu ugacagcggu cauggcauuc	cacuuaucac uggcauccuu ccacuc	56
[3498]	<210>	298	
[3499]	<211>	470	
[3500]	<212>	RNA	
[3501]	<213>	人工序列	
[3502]	<220>		
[3503]	<223>	合成序列	
[3504]	<400>	298	
[3505]	gggcgaauug aaggaaggcc gucaaggccg	caugccauua auacgacuca cuauaggauu	60
[3506]	uagaccaccc caaaaugaa ggggacuaaa	acacugguga acuccgaua gugcgggugu	120
[3507]	ugaaugauuu agaccacccc aaaaugaag	gggacuaaaa cacagauua gccugguggu	180
[3508]	ucaggcggcg cauugauuuu gaccacccca	aaaugaagg ggacuaaaac aggggcagag	240

```

[3509]   augaugaccc uuuuggcucc cccgauuuag accaccccaa aaaugaaggg gacuaaaaca 300
[3510]   auaugacaca uucuagcuua acuuucgaua uagauuuaga ccacccccaaa aaugaagggg 360
[3511]   acuaaaacac aaacaugauc ugguucaucu ucucgcgguu ggauuuagac ccccccaaaa 420
[3512]   augaagggga cuaaaacaau gccagugaua aguggaaugc caugaccgcu          470
[3513]   <210>   299
[3514]   <211>   1154
[3515]   <212>   PRT
[3516]   <213>   Paludibacter propionicigenes
[3517]   <400>   299
[3518]   Met Arg Val Ser Lys Val Lys Val Lys Asp Gly Gly Lys Asp Lys Met
[3519]   1           5           10           15
[3520]   Val Leu Val His Arg Lys Thr Thr Gly Ala Gln Leu Val Tyr Ser Gly
[3521]           20           25           30
[3522]   Gln Pro Val Ser Asn Glu Thr Ser Asn Ile Leu Pro Glu Lys Lys Arg
[3523]           35           40           45
[3524]   Gln Ser Phe Asp Leu Ser Thr Leu Asn Lys Thr Ile Ile Lys Phe Asp
[3525]           50           55           60
[3526]   Thr Ala Lys Lys Gln Lys Leu Asn Val Asp Gln Tyr Lys Ile Val Glu
[3527]   65           70           75           80
[3528]   Lys Ile Phe Lys Tyr Pro Lys Gln Glu Leu Pro Lys Gln Ile Lys Ala
[3529]           85           90           95
[3530]   Glu Glu Ile Leu Pro Phe Leu Asn His Lys Phe Gln Glu Pro Val Lys
[3531]           100          105          110
[3532]   Tyr Trp Lys Asn Gly Lys Glu Glu Ser Phe Asn Leu Thr Leu Leu Ile
[3533]           115          120          125
[3534]   Val Glu Ala Val Gln Ala Gln Asp Lys Arg Lys Leu Gln Pro Tyr Tyr
[3535]           130          135          140
[3536]   Asp Trp Lys Thr Trp Tyr Ile Gln Thr Lys Ser Asp Leu Leu Lys Lys
[3537]   145          150          155          160
[3538]   Ser Ile Glu Asn Asn Arg Ile Asp Leu Thr Glu Asn Leu Ser Lys Arg
[3539]           165          170          175
[3540]   Lys Lys Ala Leu Leu Ala Trp Glu Thr Glu Phe Thr Ala Ser Gly Ser
[3541]           180          185          190
[3542]   Ile Asp Leu Thr His Tyr His Lys Val Tyr Met Thr Asp Val Leu Cys
[3543]           195          200          205
[3544]   Lys Met Leu Gln Asp Val Lys Pro Leu Thr Asp Asp Lys Gly Lys Ile
[3545]           210          215          220
[3546]   Asn Thr Asn Ala Tyr His Arg Gly Leu Lys Lys Ala Leu Gln Asn His
[3547]   225          230          235          240

```

[3548]	Gln	Pro	Ala	Ile	Phe	Gly	Thr	Arg	Glu	Val	Pro	Asn	Glu	Ala	Asn	Arg
[3549]						245					250				255	
[3550]	Ala	Asp	Asn	Gln	Leu	Ser	Ile	Tyr	His	Leu	Glu	Val	Val	Lys	Tyr	Leu
[3551]						260					265				270	
[3552]	Glu	His	Tyr	Phe	Pro	Ile	Lys	Thr	Ser	Lys	Arg	Arg	Asn	Thr	Ala	Asp
[3553]						275					280				285	
[3554]	Asp	Ile	Ala	His	Tyr	Leu	Lys	Ala	Gln	Thr	Leu	Lys	Thr	Thr	Ile	Glu
[3555]			290							295					300	
[3556]	Lys	Gln	Leu	Val	Asn	Ala	Ile	Arg	Ala	Asn	Ile	Ile	Gln	Gln	Gly	Lys
[3557]	305						310					315				320
[3558]	Thr	Asn	His	His	Glu	Leu	Lys	Ala	Asp	Thr	Thr	Ser	Asn	Asp	Leu	Ile
[3559]						325					330				335	
[3560]	Arg	Ile	Lys	Thr	Asn	Glu	Ala	Phe	Val	Leu	Asn	Leu	Thr	Gly	Thr	Cys
[3561]						340					345				350	
[3562]	Ala	Phe	Ala	Ala	Asn	Asn	Ile	Arg	Asn	Met	Val	Asp	Asn	Glu	Gln	Thr
[3563]						355								360		365
[3564]	Asn	Asp	Ile	Leu	Gly	Lys	Gly	Asp	Phe	Ile	Lys	Ser	Leu	Leu	Lys	Asp
[3565]						370								375		380
[3566]	Asn	Thr	Asn	Ser	Gln	Leu	Tyr	Ser	Phe	Phe	Phe	Gly	Glu	Gly	Leu	Ser
[3567]	385						390					395				400
[3568]	Thr	Asn	Lys	Ala	Glu	Lys	Glu	Thr	Gln	Leu	Trp	Gly	Ile	Arg	Gly	Ala
[3569]						405					410				415	
[3570]	Val	Gln	Gln	Ile	Arg	Asn	Asn	Val	Asn	His	Tyr	Lys	Lys	Asp	Ala	Leu
[3571]						420					425				430	
[3572]	Lys	Thr	Val	Phe	Asn	Ile	Ser	Asn	Phe	Glu	Asn	Pro	Thr	Ile	Thr	Asp
[3573]						435					440				445	
[3574]	Pro	Lys	Gln	Gln	Thr	Asn	Tyr	Ala	Asp	Thr	Ile	Tyr	Lys	Ala	Arg	Phe
[3575]						450					455				460	
[3576]	Ile	Asn	Glu	Leu	Glu	Lys	Ile	Pro	Glu	Ala	Phe	Ala	Gln	Gln	Leu	Lys
[3577]	465						470				475					480
[3578]	Thr	Gly	Gly	Ala	Val	Ser	Tyr	Tyr	Thr	Ile	Glu	Asn	Leu	Lys	Ser	Leu
[3579]						485					490				495	
[3580]	Leu	Thr	Thr	Phe	Gln	Phe	Ser	Leu	Cys	Arg	Ser	Thr	Ile	Pro	Phe	Ala
[3581]						500					505				510	
[3582]	Pro	Gly	Phe	Lys	Lys	Val	Phe	Asn	Gly	Gly	Ile	Asn	Tyr	Gln	Asn	Ala
[3583]						515					520				525	
[3584]	Lys	Gln	Asp	Glu	Ser	Phe	Tyr	Glu	Leu	Met	Leu	Glu	Gln	Tyr	Leu	Arg
[3585]						530					535				540	
[3586]	Lys	Glu	Asn	Phe	Ala	Glu	Glu	Ser	Tyr	Asn	Ala	Arg	Tyr	Phe	Met	Leu

[3587]	545	550	555	560
[3588]	Lys Leu Ile Tyr Asn Asn Leu Phe Leu Pro Gly Phe Thr Thr Asp Arg			
[3589]	565	570	575	
[3590]	Lys Ala Phe Ala Asp Ser Val Gly Phe Val Gln Met Gln Asn Lys Lys			
[3591]	580	585	590	
[3592]	Gln Ala Glu Lys Val Asn Pro Arg Lys Lys Glu Ala Tyr Ala Phe Glu			
[3593]	595	600	605	
[3594]	Ala Val Arg Pro Met Thr Ala Ala Asp Ser Ile Ala Asp Tyr Met Ala			
[3595]	610	615	620	
[3596]	Tyr Val Gln Ser Glu Leu Met Gln Glu Gln Asn Lys Lys Glu Glu Lys			
[3597]	625	630	635	640
[3598]	Val Ala Glu Glu Thr Arg Ile Asn Phe Glu Lys Phe Val Leu Gln Val			
[3599]	645	650	655	
[3600]	Phe Ile Lys Gly Phe Asp Ser Phe Leu Arg Ala Lys Glu Phe Asp Phe			
[3601]	660	665	670	
[3602]	Val Gln Met Pro Gln Pro Gln Leu Thr Ala Thr Ala Ser Asn Gln Gln			
[3603]	675	680	685	
[3604]	Lys Ala Asp Lys Leu Asn Gln Leu Glu Ala Ser Ile Thr Ala Asp Cys			
[3605]	690	695	700	
[3606]	Lys Leu Thr Pro Gln Tyr Ala Lys Ala Asp Asp Ala Thr His Ile Ala			
[3607]	705	710	715	720
[3608]	Phe Tyr Val Phe Cys Lys Leu Leu Asp Ala Ala His Leu Ser Asn Leu			
[3609]	725	730	735	
[3610]	Arg Asn Glu Leu Ile Lys Phe Arg Glu Ser Val Asn Glu Phe Lys Phe			
[3611]	740	745	750	
[3612]	His His Leu Leu Glu Ile Ile Glu Ile Cys Leu Leu Ser Ala Asp Val			
[3613]	755	760	765	
[3614]	Val Pro Thr Asp Tyr Arg Asp Leu Tyr Ser Ser Glu Ala Asp Cys Leu			
[3615]	770	775	780	
[3616]	Ala Arg Leu Arg Pro Phe Ile Glu Gln Gly Ala Asp Ile Thr Asn Trp			
[3617]	785	790	795	800
[3618]	Ser Asp Leu Phe Val Gln Ser Asp Lys His Ser Pro Val Ile His Ala			
[3619]	805	810	815	
[3620]	Asn Ile Glu Leu Ser Val Lys Tyr Gly Thr Thr Lys Leu Leu Glu Gln			
[3621]	820	825	830	
[3622]	Ile Ile Asn Lys Asp Thr Gln Phe Lys Thr Thr Glu Ala Asn Phe Thr			
[3623]	835	840	845	
[3624]	Ala Trp Asn Thr Ala Gln Lys Ser Ile Glu Gln Leu Ile Lys Gln Arg			
[3625]	850	855	860	

[3626]	Glu Asp His His Glu Gln Trp Val Lys Ala Lys Asn Ala Asp Asp Lys		
[3627]	865	870	875 880
[3628]	Glu Lys Gln Glu Arg Lys Arg Glu Lys Ser Asn Phe Ala Gln Lys Phe		
[3629]		885	890 895
[3630]	Ile Glu Lys His Gly Asp Asp Tyr Leu Asp Ile Cys Asp Tyr Ile Asn		
[3631]		900	905 910
[3632]	Thr Tyr Asn Trp Leu Asp Asn Lys Met His Phe Val His Leu Asn Arg		
[3633]		915	920 925
[3634]	Leu His Gly Leu Thr Ile Glu Leu Leu Gly Arg Met Ala Gly Phe Val		
[3635]		930	935 940
[3636]	Ala Leu Phe Asp Arg Asp Phe Gln Phe Phe Asp Glu Gln Gln Ile Ala		
[3637]	945	950	955 960
[3638]	Asp Glu Phe Lys Leu His Gly Phe Val Asn Leu His Ser Ile Asp Lys		
[3639]		965	970 975
[3640]	Lys Leu Asn Glu Val Pro Thr Lys Lys Ile Lys Glu Ile Tyr Asp Ile		
[3641]		980	985 990
[3642]	Arg Asn Lys Ile Ile Gln Ile Asn Gly Asn Lys Ile Asn Glu Ser Val		
[3643]		995	1000 1005
[3644]	Arg Ala Asn Leu Ile Gln Phe Ile Ser Ser Lys Arg Asn Tyr Tyr		
[3645]		1010	1015 1020
[3646]	Asn Asn Ala Phe Leu His Val Ser Asn Asp Glu Ile Lys Glu Lys		
[3647]		1025	1030 1035
[3648]	Gln Met Tyr Asp Ile Arg Asn His Ile Ala His Phe Asn Tyr Leu		
[3649]		1040	1045 1050
[3650]	Thr Lys Asp Ala Ala Asp Phe Ser Leu Ile Asp Leu Ile Asn Glu		
[3651]		1055	1060 1065
[3652]	Leu Arg Glu Leu Leu His Tyr Asp Arg Lys Leu Lys Asn Ala Val		
[3653]		1070	1075 1080
[3654]	Ser Lys Ala Phe Ile Asp Leu Phe Asp Lys His Gly Met Ile Leu		
[3655]		1085	1090 1095
[3656]	Lys Leu Lys Leu Asn Ala Asp His Lys Leu Lys Val Glu Ser Leu		
[3657]		1100	1105 1110
[3658]	Glu Pro Lys Lys Ile Tyr His Leu Gly Ser Ser Ala Lys Asp Lys		
[3659]		1115	1120 1125
[3660]	Pro Glu Tyr Gln Tyr Cys Thr Asn Gln Val Met Met Ala Tyr Cys		
[3661]		1130	1135 1140
[3662]	Asn Met Cys Arg Ser Leu Leu Glu Met Lys Lys		
[3663]		1145	1150
[3664]	<210>	300	

[3665]	<211>	1182														
[3666]	<212>	PRT														
[3667]	<213>	韦德纤毛菌(Leptotrichia wadei)														
[3668]	<400>	300														
[3669]	Met	Tyr	Met	Lys	Ile	Thr	Lys	Ile	Asp	Gly	Val	Ser	His	Tyr	Lys	Lys
[3670]	1				5					10					15	
[3671]	Gln	Asp	Lys	Gly	Ile	Leu	Lys	Lys	Lys	Trp	Lys	Asp	Leu	Asp	Glu	Arg
[3672]				20					25					30		
[3673]	Lys	Gln	Arg	Glu	Lys	Ile	Glu	Ala	Arg	Tyr	Asn	Lys	Gln	Ile	Glu	Ser
[3674]			35					40					45			
[3675]	Lys	Ile	Tyr	Lys	Glu	Phe	Phe	Arg	Leu	Lys	Asn	Lys	Lys	Arg	Ile	Glu
[3676]		50					55					60				
[3677]	Lys	Glu	Glu	Asp	Gln	Asn	Ile	Lys	Ser	Leu	Tyr	Phe	Phe	Ile	Lys	Glu
[3678]	65					70					75				80	
[3679]	Leu	Tyr	Leu	Asn	Glu	Lys	Asn	Glu	Glu	Trp	Glu	Leu	Lys	Asn	Ile	Asn
[3680]				85					90						95	
[3681]	Leu	Glu	Ile	Leu	Asp	Asp	Lys	Glu	Arg	Val	Ile	Lys	Gly	Tyr	Lys	Phe
[3682]				100					105					110		
[3683]	Lys	Glu	Asp	Val	Tyr	Phe	Phe	Lys	Glu	Gly	Tyr	Lys	Glu	Tyr	Tyr	Leu
[3684]			115					120					125			
[3685]	Arg	Ile	Leu	Phe	Asn	Asn	Leu	Ile	Glu	Lys	Val	Gln	Asn	Glu	Asn	Arg
[3686]		130					135					140				
[3687]	Glu	Lys	Val	Arg	Lys	Asn	Lys	Glu	Phe	Leu	Asp	Leu	Lys	Glu	Ile	Phe
[3688]	145					150					155					160
[3689]	Lys	Lys	Tyr	Lys	Asn	Arg	Lys	Ile	Asp	Leu	Leu	Leu	Lys	Ser	Ile	Asn
[3690]				165					170						175	
[3691]	Asn	Asn	Lys	Ile	Asn	Leu	Glu	Tyr	Lys	Lys	Glu	Asn	Val	Asn	Glu	Glu
[3692]				180					185					190		
[3693]	Ile	Tyr	Gly	Ile	Asn	Pro	Thr	Asn	Asp	Arg	Glu	Met	Thr	Phe	Tyr	Glu
[3694]			195					200					205			
[3695]	Leu	Leu	Lys	Glu	Ile	Ile	Glu	Lys	Lys	Asp	Glu	Gln	Lys	Ser	Ile	Leu
[3696]		210					215					220				
[3697]	Glu	Glu	Lys	Leu	Asp	Asn	Phe	Asp	Ile	Thr	Asn	Phe	Leu	Glu	Asn	Ile
[3698]	225					230					235					240
[3699]	Glu	Lys	Ile	Phe	Asn	Glu	Glu	Thr	Glu	Ile	Asn	Ile	Ile	Lys	Gly	Lys
[3700]				245					250						255	
[3701]	Val	Leu	Asn	Glu	Leu	Arg	Glu	Tyr	Ile	Lys	Glu	Lys	Glu	Glu	Asn	Asn
[3702]				260					265					270		
[3703]	Ser	Asp	Asn	Lys	Leu	Lys	Gln	Ile	Tyr	Asn	Leu	Glu	Leu	Lys	Lys	Tyr

[3704]	275	280	285
[3705]	Ile Glu Asn Asn Phe Ser Tyr Lys Lys Gln Lys Ser Lys Ser Lys Asn		
[3706]	290	295	300
[3707]	Gly Lys Asn Asp Tyr Leu Tyr Leu Asn Phe Leu Lys Lys Ile Met Phe		
[3708]	305	310	315
[3709]	Ile Glu Glu Val Asp Glu Lys Lys Glu Ile Asn Lys Glu Lys Phe Lys		
[3710]	325	330	335
[3711]	Asn Lys Ile Asn Ser Asn Phe Lys Asn Leu Phe Val Gln His Ile Leu		
[3712]	340	345	350
[3713]	Asp Tyr Gly Lys Leu Leu Tyr Tyr Lys Glu Asn Asp Glu Tyr Ile Lys		
[3714]	355	360	365
[3715]	Asn Thr Gly Gln Leu Glu Thr Lys Asp Leu Glu Tyr Ile Lys Thr Lys		
[3716]	370	375	380
[3717]	Glu Thr Leu Ile Arg Lys Met Ala Val Leu Val Ser Phe Ala Ala Asn		
[3718]	385	390	395
[3719]	Ser Tyr Tyr Asn Leu Phe Gly Arg Val Ser Gly Asp Ile Leu Gly Thr		
[3720]	405	410	415
[3721]	Glu Val Val Lys Ser Ser Lys Thr Asn Val Ile Lys Val Gly Ser His		
[3722]	420	425	430
[3723]	Ile Phe Lys Glu Lys Met Leu Asn Tyr Phe Phe Asp Phe Glu Ile Phe		
[3724]	435	440	445
[3725]	Asp Ala Asn Lys Ile Val Glu Ile Leu Glu Ser Ile Ser Tyr Ser Ile		
[3726]	450	455	460
[3727]	Tyr Asn Val Arg Asn Gly Val Gly His Phe Asn Lys Leu Ile Leu Gly		
[3728]	465	470	475
[3729]	Lys Tyr Lys Lys Lys Asp Ile Asn Thr Asn Lys Arg Ile Glu Glu Asp		
[3730]	485	490	495
[3731]	Leu Asn Asn Asn Glu Glu Ile Lys Gly Tyr Phe Ile Lys Lys Arg Gly		
[3732]	500	505	510
[3733]	Glu Ile Glu Arg Lys Val Lys Glu Lys Phe Leu Ser Asn Asn Leu Gln		
[3734]	515	520	525
[3735]	Tyr Tyr Tyr Ser Lys Glu Lys Ile Glu Asn Tyr Phe Glu Val Tyr Glu		
[3736]	530	535	540
[3737]	Phe Glu Ile Leu Lys Arg Lys Ile Pro Phe Ala Pro Asn Phe Lys Arg		
[3738]	545	550	555
[3739]	Ile Ile Lys Lys Gly Glu Asp Leu Phe Asn Asn Lys Asn Asn Lys Lys		
[3740]	565	570	575
[3741]	Tyr Glu Tyr Phe Lys Asn Phe Asp Lys Asn Ser Ala Glu Glu Lys Lys		
[3742]	580	585	590

[3743]	Glu Phe Leu Lys Thr Arg Asn Phe Leu Leu Lys Glu Leu Tyr Tyr Asn
[3744]	595 600 605
[3745]	Asn Phe Tyr Lys Glu Phe Leu Ser Lys Lys Glu Glu Phe Glu Lys Ile
[3746]	610 615 620
[3747]	Val Leu Glu Val Lys Glu Glu Lys Lys Ser Arg Gly Asn Ile Asn Asn
[3748]	625 630 635 640
[3749]	Lys Lys Ser Gly Val Ser Phe Gln Ser Ile Asp Asp Tyr Asp Thr Lys
[3750]	645 650 655
[3751]	Ile Asn Ile Ser Asp Tyr Ile Ala Ser Ile His Lys Lys Glu Met Glu
[3752]	660 665 670
[3753]	Arg Val Glu Lys Tyr Asn Glu Glu Lys Gln Lys Asp Thr Ala Lys Tyr
[3754]	675 680 685
[3755]	Ile Arg Asp Phe Val Glu Glu Ile Phe Leu Thr Gly Phe Ile Asn Tyr
[3756]	690 695 700
[3757]	Leu Glu Lys Asp Lys Arg Leu His Phe Leu Lys Glu Glu Phe Ser Ile
[3758]	705 710 715 720
[3759]	Leu Cys Asn Asn Asn Asn Asn Val Val Asp Phe Asn Ile Asn Ile Asn
[3760]	725 730 735
[3761]	Glu Glu Lys Ile Lys Glu Phe Leu Lys Glu Asn Asp Ser Lys Thr Leu
[3762]	740 745 750
[3763]	Asn Leu Tyr Leu Phe Phe Asn Met Ile Asp Ser Lys Arg Ile Ser Glu
[3764]	755 760 765
[3765]	Phe Arg Asn Glu Leu Val Lys Tyr Lys Gln Phe Thr Lys Lys Arg Leu
[3766]	770 775 780
[3767]	Asp Glu Glu Lys Glu Phe Leu Gly Ile Lys Ile Glu Leu Tyr Glu Thr
[3768]	785 790 795 800
[3769]	Leu Ile Glu Phe Val Ile Leu Thr Arg Glu Lys Leu Asp Thr Lys Lys
[3770]	805 810 815
[3771]	Ser Glu Glu Ile Asp Ala Trp Leu Val Asp Lys Leu Tyr Val Lys Asp
[3772]	820 825 830
[3773]	Ser Asn Glu Tyr Lys Glu Tyr Glu Glu Ile Leu Lys Leu Phe Val Asp
[3774]	835 840 845
[3775]	Glu Lys Ile Leu Ser Ser Lys Glu Ala Pro Tyr Tyr Ala Thr Asp Asn
[3776]	850 855 860
[3777]	Lys Thr Pro Ile Leu Leu Ser Asn Phe Glu Lys Thr Arg Lys Tyr Gly
[3778]	865 870 875 880
[3779]	Thr Gln Ser Phe Leu Ser Glu Ile Gln Ser Asn Tyr Lys Tyr Ser Lys
[3780]	885 890 895
[3781]	Val Glu Lys Glu Asn Ile Glu Asp Tyr Asn Lys Lys Glu Glu Ile Glu

[3782]	900	905	910
[3783]	Gln Lys Lys Lys Ser Asn Ile Glu Lys Leu Gln Asp Leu Lys Val Glu		
[3784]	915	920	925
[3785]	Leu His Lys Lys Trp Glu Gln Asn Lys Ile Thr Glu Lys Glu Ile Glu		
[3786]	930	935	940
[3787]	Lys Tyr Asn Asn Thr Thr Arg Lys Ile Asn Glu Tyr Asn Tyr Leu Lys		
[3788]	945	950	955
[3789]	Asn Lys Glu Glu Leu Gln Asn Val Tyr Leu Leu His Glu Met Leu Ser		
[3790]	965	970	975
[3791]	Asp Leu Leu Ala Arg Asn Val Ala Phe Phe Asn Lys Trp Glu Arg Asp		
[3792]	980	985	990
[3793]	Phe Lys Phe Ile Val Ile Ala Ile Lys Gln Phe Leu Arg Glu Asn Asp		
[3794]	995	1000	1005
[3795]	Lys Glu Lys Val Asn Glu Phe Leu Asn Pro Pro Asp Asn Ser Lys		
[3796]	1010	1015	1020
[3797]	Gly Lys Lys Val Tyr Phe Ser Val Ser Lys Tyr Lys Asn Thr Val		
[3798]	1025	1030	1035
[3799]	Glu Asn Ile Asp Gly Ile His Lys Asn Phe Met Asn Leu Ile Phe		
[3800]	1040	1045	1050
[3801]	Leu Asn Asn Lys Phe Met Asn Arg Lys Ile Asp Lys Met Asn Cys		
[3802]	1055	1060	1065
[3803]	Ala Ile Trp Val Tyr Phe Arg Asn Tyr Ile Ala His Phe Leu His		
[3804]	1070	1075	1080
[3805]	Leu His Thr Lys Asn Glu Lys Ile Ser Leu Ile Ser Gln Met Asn		
[3806]	1085	1090	1095
[3807]	Leu Leu Ile Lys Leu Phe Ser Tyr Asp Lys Lys Val Gln Asn His		
[3808]	1100	1105	1110
[3809]	Ile Leu Lys Ser Thr Lys Thr Leu Leu Glu Lys Tyr Asn Ile Gln		
[3810]	1115	1120	1125
[3811]	Ile Asn Phe Glu Ile Ser Asn Asp Lys Asn Glu Val Phe Lys Tyr		
[3812]	1130	1135	1140
[3813]	Lys Ile Lys Asn Arg Leu Tyr Ser Lys Lys Gly Lys Met Leu Gly		
[3814]	1145	1150	1155
[3815]	Lys Asn Asn Lys Phe Glu Ile Leu Glu Asn Glu Phe Leu Glu Asn		
[3816]	1160	1165	1170
[3817]	Val Lys Ala Met Leu Glu Tyr Ser Glu		
[3818]	1175	1180	
[3819]	<210> 301		
[3820]	<211> 1437		

[3821]	<212>	PRT
[3822]	<213>	未知的
[3823]	<220>	
[3824]	<223>	毛螺菌科(Lachnospiraceae)细菌
[3825]	<400>	301
[3826]	Met Lys Ile Ser Lys Val Arg Glu Glu Asn Arg Gly Ala Lys Leu Thr	
[3827]	1	5 10 15
[3828]	Val Asn Ala Lys Thr Ala Val Val Ser Glu Asn Arg Ser Gln Glu Gly	
[3829]	20	25 30
[3830]	Ile Leu Tyr Asn Asp Pro Ser Arg Tyr Gly Lys Ser Arg Lys Asn Asp	
[3831]	35	40 45
[3832]	Glu Asp Arg Asp Arg Tyr Ile Glu Ser Arg Leu Lys Ser Ser Gly Lys	
[3833]	50	55 60
[3834]	Leu Tyr Arg Ile Phe Asn Glu Asp Lys Asn Lys Arg Glu Thr Asp Glu	
[3835]	65	70 75 80
[3836]	Leu Gln Trp Phe Leu Ser Glu Ile Val Lys Lys Ile Asn Arg Arg Asn	
[3837]	85	90 95
[3838]	Gly Leu Val Leu Ser Asp Met Leu Ser Val Asp Asp Arg Ala Phe Glu	
[3839]	100	105 110
[3840]	Lys Ala Phe Glu Lys Tyr Ala Glu Leu Ser Tyr Thr Asn Arg Arg Asn	
[3841]	115	120 125
[3842]	Lys Val Ser Gly Ser Pro Ala Phe Glu Thr Cys Gly Val Asp Ala Ala	
[3843]	130	135 140
[3844]	Thr Ala Glu Arg Leu Lys Gly Ile Ile Ser Glu Thr Asn Phe Ile Asn	
[3845]	145	150 155 160
[3846]	Arg Ile Lys Asn Asn Ile Asp Asn Lys Val Ser Glu Asp Ile Ile Asp	
[3847]	165	170 175
[3848]	Arg Ile Ile Ala Lys Tyr Leu Lys Lys Ser Leu Cys Arg Glu Arg Val	
[3849]	180	185 190
[3850]	Lys Arg Gly Leu Lys Lys Leu Leu Met Asn Ala Phe Asp Leu Pro Tyr	
[3851]	195	200 205
[3852]	Ser Asp Pro Asp Ile Asp Val Gln Arg Asp Phe Ile Asp Tyr Val Leu	
[3853]	210	215 220
[3854]	Glu Asp Phe Tyr His Val Arg Ala Lys Ser Gln Val Ser Arg Ser Ile	
[3855]	225	230 235 240
[3856]	Lys Asn Met Asn Met Pro Val Gln Pro Glu Gly Asp Gly Lys Phe Ala	
[3857]	245	250 255
[3858]	Ile Thr Val Ser Lys Gly Gly Thr Glu Ser Gly Asn Lys Arg Ser Ala	
[3859]	260	265 270

[3860]	Glu Lys Glu Ala Phe Lys Lys Phe Leu Ser Asp Tyr Ala Ser Leu Asp		
[3861]	275	280	285
[3862]	Glu Arg Val Arg Asp Asp Met Leu Arg Arg Met Arg Arg Leu Val Val		
[3863]	290	295	300
[3864]	Leu Tyr Phe Tyr Gly Ser Asp Asp Ser Lys Leu Ser Asp Val Asn Glu		
[3865]	305	310	315 320
[3866]	Lys Phe Asp Val Trp Glu Asp His Ala Ala Arg Arg Val Asp Asn Arg		
[3867]	325	330	335
[3868]	Glu Phe Ile Lys Leu Pro Leu Glu Asn Lys Leu Ala Asn Gly Lys Thr		
[3869]	340	345	350
[3870]	Asp Lys Asp Ala Glu Arg Ile Arg Lys Asn Thr Val Lys Glu Leu Tyr		
[3871]	355	360	365
[3872]	Arg Asn Gln Asn Ile Gly Cys Tyr Arg Gln Ala Val Lys Ala Val Glu		
[3873]	370	375	380
[3874]	Glu Asp Asn Asn Gly Arg Tyr Phe Asp Asp Lys Met Leu Asn Met Phe		
[3875]	385	390	395 400
[3876]	Phe Ile His Arg Ile Glu Tyr Gly Val Glu Lys Ile Tyr Ala Asn Leu		
[3877]	405	410	415
[3878]	Lys Gln Val Thr Glu Phe Lys Ala Arg Thr Gly Tyr Leu Ser Glu Lys		
[3879]	420	425	430
[3880]	Ile Trp Lys Asp Leu Ile Asn Tyr Ile Ser Ile Lys Tyr Ile Ala Met		
[3881]	435	440	445
[3882]	Gly Lys Ala Val Tyr Asn Tyr Ala Met Asp Glu Leu Asn Ala Ser Asp		
[3883]	450	455	460
[3884]	Lys Lys Glu Ile Glu Leu Gly Lys Ile Ser Glu Glu Tyr Leu Ser Gly		
[3885]	465	470	475 480
[3886]	Ile Ser Ser Phe Asp Tyr Glu Leu Ile Lys Ala Glu Glu Met Leu Gln		
[3887]	485	490	495
[3888]	Arg Glu Thr Ala Val Tyr Val Ala Phe Ala Ala Arg His Leu Ser Ser		
[3889]	500	505	510
[3890]	Gln Thr Val Glu Leu Asp Ser Glu Asn Ser Asp Phe Leu Leu Leu Lys		
[3891]	515	520	525
[3892]	Pro Lys Gly Thr Met Asp Lys Asn Asp Lys Asn Lys Leu Ala Ser Asn		
[3893]	530	535	540
[3894]	Asn Ile Leu Asn Phe Leu Lys Asp Lys Glu Thr Leu Arg Asp Thr Ile		
[3895]	545	550	555 560
[3896]	Leu Gln Tyr Phe Gly Gly His Ser Leu Trp Thr Asp Phe Pro Phe Asp		
[3897]	565	570	575
[3898]	Lys Tyr Leu Ala Gly Gly Lys Asp Asp Val Asp Phe Leu Thr Asp Leu		

[3899]	580							585							590													
[3900]	Lys	Asp	Val	Ile	Tyr	Ser	Met	Arg	Asn	Asp	Ser	Phe	His	Tyr	Ala	Thr												
[3901]	595							600							605													
[3902]	Glu	Asn	His	Asn	Asn	Gly	Lys	Trp	Asn	Lys	Glu	Leu	Ile	Ser	Ala	Met												
[3903]	610							615							620													
[3904]	Phe	Glu	His	Glu	Thr	Glu	Arg	Met	Thr	Val	Val	Met	Lys	Asp	Lys	Phe												
[3905]	625							630							635							640						
[3906]	Tyr	Ser	Asn	Asn	Leu	Pro	Met	Phe	Tyr	Lys	Asn	Asp	Asp	Leu	Lys	Lys												
[3907]	645							650							655													
[3908]	Leu	Leu	Ile	Asp	Leu	Tyr	Lys	Asp	Asn	Val	Glu	Arg	Ala	Ser	Gln	Val												
[3909]	660							665							670													
[3910]	Pro	Ser	Phe	Asn	Lys	Val	Phe	Val	Arg	Lys	Asn	Phe	Pro	Ala	Leu	Val												
[3911]	675							680							685													
[3912]	Arg	Asp	Lys	Asp	Asn	Leu	Gly	Ile	Glu	Leu	Asp	Leu	Lys	Ala	Asp	Ala												
[3913]	690							695							700													
[3914]	Asp	Lys	Gly	Glu	Asn	Glu	Leu	Lys	Phe	Tyr	Asn	Ala	Leu	Tyr	Tyr	Met												
[3915]	705							710							715							720						
[3916]	Phe	Lys	Glu	Ile	Tyr	Tyr	Asn	Ala	Phe	Leu	Asn	Asp	Lys	Asn	Val	Arg												
[3917]	725							730							735													
[3918]	Glu	Arg	Phe	Ile	Thr	Lys	Ala	Thr	Lys	Val	Ala	Asp	Asn	Tyr	Asp	Arg												
[3919]	740							745							750													
[3920]	Asn	Lys	Glu	Arg	Asn	Leu	Lys	Asp	Arg	Ile	Lys	Ser	Ala	Gly	Ser	Asp												
[3921]	755							760							765													
[3922]	Glu	Lys	Lys	Lys	Leu	Arg	Glu	Gln	Leu	Gln	Asn	Tyr	Ile	Ala	Glu	Asn												
[3923]	770							775							780													
[3924]	Asp	Phe	Gly	Gln	Arg	Ile	Lys	Asn	Ile	Val	Gln	Val	Asn	Pro	Asp	Tyr												
[3925]	785							790							795							800						
[3926]	Thr	Leu	Ala	Gln	Ile	Cys	Gln	Leu	Ile	Met	Thr	Glu	Tyr	Asn	Gln	Gln												
[3927]	805							810							815													
[3928]	Asn	Asn	Gly	Cys	Met	Gln	Lys	Lys	Ser	Ala	Ala	Arg	Lys	Asp	Ile	Asn												
[3929]	820							825							830													
[3930]	Lys	Asp	Ser	Tyr	Gln	His	Tyr	Lys	Met	Leu	Leu	Leu	Val	Asn	Leu	Arg												
[3931]	835							840							845													
[3932]	Lys	Ala	Phe	Leu	Glu	Phe	Ile	Lys	Glu	Asn	Tyr	Ala	Phe	Val	Leu	Lys												
[3933]	850							855							860													
[3934]	Pro	Tyr	Lys	His	Asp	Leu	Cys	Asp	Lys	Ala	Asp	Phe	Val	Pro	Asp	Phe												
[3935]	865							870							875							880						
[3936]	Ala	Lys	Tyr	Val	Lys	Pro	Tyr	Ala	Gly	Leu	Ile	Ser	Arg	Val	Ala	Gly												
[3937]	885							890							895													

[3938]	Ser Ser Glu Leu Gln Lys Trp Tyr Ile Val Ser Arg Phe Leu Ser Pro
[3939]	900 905 910
[3940]	Ala Gln Ala Asn His Met Leu Gly Phe Leu His Ser Tyr Lys Gln Tyr
[3941]	915 920 925
[3942]	Val Trp Asp Ile Tyr Arg Arg Ala Ser Glu Thr Gly Thr Glu Ile Asn
[3943]	930 935 940
[3944]	His Ser Ile Ala Glu Asp Lys Ile Ala Gly Val Asp Ile Thr Asp Val
[3945]	945 950 955 960
[3946]	Asp Ala Val Ile Asp Leu Ser Val Lys Leu Cys Gly Thr Ile Ser Ser
[3947]	965 970 975
[3948]	Glu Ile Ser Asp Tyr Phe Lys Asp Asp Glu Val Tyr Ala Glu Tyr Ile
[3949]	980 985 990
[3950]	Ser Ser Tyr Leu Asp Phe Glu Tyr Asp Gly Gly Asn Tyr Lys Asp Ser
[3951]	995 1000 1005
[3952]	Leu Asn Arg Phe Cys Asn Ser Asp Ala Val Asn Asp Gln Lys Val
[3953]	1010 1015 1020
[3954]	Ala Leu Tyr Tyr Asp Gly Glu His Pro Lys Leu Asn Arg Asn Ile
[3955]	1025 1030 1035
[3956]	Ile Leu Ser Lys Leu Tyr Gly Glu Arg Arg Phe Leu Glu Lys Ile
[3957]	1040 1045 1050
[3958]	Thr Asp Arg Val Ser Arg Ser Asp Ile Val Glu Tyr Tyr Lys Leu
[3959]	1055 1060 1065
[3960]	Lys Lys Glu Thr Ser Gln Tyr Gln Thr Lys Gly Ile Phe Asp Ser
[3961]	1070 1075 1080
[3962]	Glu Asp Glu Gln Lys Asn Ile Lys Lys Phe Gln Glu Met Lys Asn
[3963]	1085 1090 1095
[3964]	Ile Val Glu Phe Arg Asp Leu Met Asp Tyr Ser Glu Ile Ala Asp
[3965]	1100 1105 1110
[3966]	Glu Leu Gln Gly Gln Leu Ile Asn Trp Ile Tyr Leu Arg Glu Arg
[3967]	1115 1120 1125
[3968]	Asp Leu Met Asn Phe Gln Leu Gly Tyr His Tyr Ala Cys Leu Asn
[3969]	1130 1135 1140
[3970]	Asn Asp Ser Asn Lys Gln Ala Thr Tyr Val Thr Leu Asp Tyr Gln
[3971]	1145 1150 1155
[3972]	Gly Lys Lys Asn Arg Lys Ile Asn Gly Ala Ile Leu Tyr Gln Ile
[3973]	1160 1165 1170
[3974]	Cys Ala Met Tyr Ile Asn Gly Leu Pro Leu Tyr Tyr Val Asp Lys
[3975]	1175 1180 1185
[3976]	Asp Ser Ser Glu Trp Thr Val Ser Asp Gly Lys Glu Ser Thr Gly

[3977]	1190	1195	1200
[3978]	Ala Lys Ile Gly Glu Phe Tyr Arg Tyr Ala Lys Ser Phe Glu Asn		
[3979]	1205	1210	1215
[3980]	Thr Ser Asp Cys Tyr Ala Ser Gly Leu Glu Ile Phe Glu Asn Ile		
[3981]	1220	1225	1230
[3982]	Ser Glu His Asp Asn Ile Thr Glu Leu Arg Asn Tyr Ile Glu His		
[3983]	1235	1240	1245
[3984]	Phe Arg Tyr Tyr Ser Ser Phe Asp Arg Ser Phe Leu Gly Ile Tyr		
[3985]	1250	1255	1260
[3986]	Ser Glu Val Phe Asp Arg Phe Phe Thr Tyr Asp Leu Lys Tyr Arg		
[3987]	1265	1270	1275
[3988]	Lys Asn Val Pro Thr Ile Leu Tyr Asn Ile Leu Leu Gln His Phe		
[3989]	1280	1285	1290
[3990]	Val Asn Val Arg Phe Glu Phe Val Ser Gly Lys Lys Met Ile Gly		
[3991]	1295	1300	1305
[3992]	Ile Asp Lys Lys Asp Arg Lys Ile Ala Lys Glu Lys Glu Cys Ala		
[3993]	1310	1315	1320
[3994]	Arg Ile Thr Ile Arg Glu Lys Asn Gly Val Tyr Ser Glu Gln Phe		
[3995]	1325	1330	1335
[3996]	Thr Tyr Lys Leu Lys Asn Gly Thr Val Tyr Val Asp Ala Arg Asp		
[3997]	1340	1345	1350
[3998]	Lys Arg Tyr Leu Gln Ser Ile Ile Arg Leu Leu Phe Tyr Pro Glu		
[3999]	1355	1360	1365
[4000]	Lys Val Asn Met Asp Glu Met Ile Glu Val Lys Glu Lys Lys Lys		
[4001]	1370	1375	1380
[4002]	Pro Ser Asp Asn Asn Thr Gly Lys Gly Tyr Ser Lys Arg Asp Arg		
[4003]	1385	1390	1395
[4004]	Gln Gln Asp Arg Lys Glu Tyr Asp Lys Tyr Lys Glu Lys Lys Lys		
[4005]	1400	1405	1410
[4006]	Lys Glu Gly Asn Phe Leu Ser Gly Met Gly Gly Asn Ile Asn Trp		
[4007]	1415	1420	1425
[4008]	Asp Glu Ile Asn Ala Gln Leu Lys Asn		
[4009]	1430	1435	
[4010]	<210> 302		
[4011]	<211> 1385		
[4012]	<212> PRT		
[4013]	<213> Clostridium aminophilum		
[4014]	<400> 302		
[4015]	Met Lys Phe Ser Lys Val Asp His Thr Arg Ser Ala Val Gly Ile Gln		

[4016]	1	5	10	15
[4017]	Lys Ala Thr Asp Ser Val His Gly Met Leu Tyr Thr Asp Pro Lys Lys			
[4018]	20	25	30	
[4019]	Gln Glu Val Asn Asp Leu Asp Lys Arg Phe Asp Gln Leu Asn Val Lys			
[4020]	35	40	45	
[4021]	Ala Lys Arg Leu Tyr Asn Val Phe Asn Gln Ser Lys Ala Glu Glu Asp			
[4022]	50	55	60	
[4023]	Asp Asp Glu Lys Arg Phe Gly Lys Val Val Lys Lys Leu Asn Arg Glu			
[4024]	65	70	75	80
[4025]	Leu Lys Asp Leu Leu Phe His Arg Glu Val Ser Arg Tyr Asn Ser Ile			
[4026]	85	90	95	
[4027]	Gly Asn Ala Lys Tyr Asn Tyr Tyr Gly Ile Lys Ser Asn Pro Glu Glu			
[4028]	100	105	110	
[4029]	Ile Val Ser Asn Leu Gly Met Val Glu Ser Leu Lys Gly Glu Arg Asp			
[4030]	115	120	125	
[4031]	Pro Gln Lys Val Ile Ser Lys Leu Leu Leu Tyr Tyr Leu Arg Lys Gly			
[4032]	130	135	140	
[4033]	Leu Lys Pro Gly Thr Asp Gly Leu Arg Met Ile Leu Glu Ala Ser Cys			
[4034]	145	150	155	160
[4035]	Gly Leu Arg Lys Leu Ser Gly Asp Glu Lys Glu Leu Lys Val Phe Leu			
[4036]	165	170	175	
[4037]	Gln Thr Leu Asp Glu Asp Phe Glu Lys Lys Thr Phe Lys Lys Asn Leu			
[4038]	180	185	190	
[4039]	Ile Arg Ser Ile Glu Asn Gln Asn Met Ala Val Gln Pro Ser Asn Glu			
[4040]	195	200	205	
[4041]	Gly Asp Pro Ile Ile Gly Ile Thr Gln Gly Arg Phe Asn Ser Gln Lys			
[4042]	210	215	220	
[4043]	Asn Glu Glu Lys Ser Ala Ile Glu Arg Met Met Ser Met Tyr Ala Asp			
[4044]	225	230	235	240
[4045]	Leu Asn Glu Asp His Arg Glu Asp Val Leu Arg Lys Leu Arg Arg Leu			
[4046]	245	250	255	
[4047]	Asn Val Leu Tyr Phe Asn Val Asp Thr Glu Lys Thr Glu Glu Pro Thr			
[4048]	260	265	270	
[4049]	Leu Pro Gly Glu Val Asp Thr Asn Pro Val Phe Glu Val Trp His Asp			
[4050]	275	280	285	
[4051]	His Glu Lys Gly Lys Glu Asn Asp Arg Gln Phe Ala Thr Phe Ala Lys			
[4052]	290	295	300	
[4053]	Ile Leu Thr Glu Asp Arg Glu Thr Arg Lys Lys Glu Lys Leu Ala Val			
[4054]	305	310	315	320

[4055]	Lys	Glu	Ala	Leu	Asn	Asp	Leu	Lys	Ser	Ala	Ile	Arg	Asp	His	Asn	Ile
[4056]					325					330					335	
[4057]	Met	Ala	Tyr	Arg	Cys	Ser	Ile	Lys	Val	Thr	Glu	Gln	Asp	Lys	Asp	Gly
[4058]				340					345						350	
[4059]	Leu	Phe	Phe	Glu	Asp	Gln	Arg	Ile	Asn	Arg	Phe	Trp	Ile	His	His	Ile
[4060]				355					360						365	
[4061]	Glu	Ser	Ala	Val	Glu	Arg	Ile	Leu	Ala	Ser	Ile	Asn	Pro	Glu	Lys	Leu
[4062]			370						375					380		
[4063]	Tyr	Lys	Leu	Arg	Ile	Gly	Tyr	Leu	Gly	Glu	Lys	Val	Trp	Lys	Asp	Leu
[4064]	385					390					395					400
[4065]	Leu	Asn	Tyr	Leu	Ser	Ile	Lys	Tyr	Ile	Ala	Val	Gly	Lys	Ala	Val	Phe
[4066]					405					410						415
[4067]	His	Phe	Ala	Met	Glu	Asp	Leu	Gly	Lys	Thr	Gly	Gln	Asp	Ile	Glu	Leu
[4068]				420						425					430	
[4069]	Gly	Lys	Leu	Ser	Asn	Ser	Val	Ser	Gly	Gly	Leu	Thr	Ser	Phe	Asp	Tyr
[4070]			435						440						445	
[4071]	Glu	Gln	Ile	Arg	Ala	Asp	Glu	Thr	Leu	Gln	Arg	Gln	Leu	Ser	Val	Glu
[4072]			450						455						460	
[4073]	Val	Ala	Phe	Ala	Ala	Asn	Asn	Leu	Phe	Arg	Ala	Val	Val	Gly	Gln	Thr
[4074]	465					470					475					480
[4075]	Gly	Lys	Lys	Ile	Glu	Gln	Ser	Lys	Ser	Glu	Glu	Asn	Glu	Glu	Asp	Phe
[4076]					485					490						495
[4077]	Leu	Leu	Trp	Lys	Ala	Glu	Lys	Ile	Ala	Glu	Ser	Ile	Lys	Lys	Glu	Gly
[4078]				500						505					510	
[4079]	Glu	Gly	Asn	Thr	Leu	Lys	Ser	Ile	Leu	Gln	Phe	Phe	Gly	Gly	Ala	Ser
[4080]			515							520					525	
[4081]	Ser	Trp	Asp	Leu	Asn	His	Phe	Cys	Ala	Ala	Tyr	Gly	Asn	Glu	Ser	Ser
[4082]			530							535					540	
[4083]	Ala	Leu	Gly	Tyr	Glu	Thr	Lys	Phe	Ala	Asp	Asp	Leu	Arg	Lys	Ala	Ile
[4084]	545						550				555					560
[4085]	Tyr	Ser	Leu	Arg	Asn	Glu	Thr	Phe	His	Phe	Thr	Thr	Leu	Asn	Lys	Gly
[4086]					565					570						575
[4087]	Ser	Phe	Asp	Trp	Asn	Ala	Lys	Leu	Ile	Gly	Asp	Met	Phe	Ser	His	Glu
[4088]				580						585					590	
[4089]	Ala	Ala	Thr	Gly	Ile	Ala	Val	Glu	Arg	Thr	Arg	Phe	Tyr	Ser	Asn	Asn
[4090]				595						600					605	
[4091]	Leu	Pro	Met	Phe	Tyr	Arg	Glu	Ser	Asp	Leu	Lys	Arg	Ile	Met	Asp	His
[4092]			610							615					620	
[4093]	Leu	Tyr	Asn	Thr	Tyr	His	Pro	Arg	Ala	Ser	Gln	Val	Pro	Ser	Phe	Asn

[4094]	625	630	635	640
[4095]	Ser Val Phe Val Arg Lys Asn Phe Arg Leu Phe Leu Ser Asn Thr Leu			
[4096]		645	650	655
[4097]	Asn Thr Asn Thr Ser Phe Asp Thr Glu Val Tyr Gln Lys Trp Glu Ser			
[4098]		660	665	670
[4099]	Gly Val Tyr Tyr Leu Phe Lys Glu Ile Tyr Tyr Asn Ser Phe Leu Pro			
[4100]		675	680	685
[4101]	Ser Gly Asp Ala His His Leu Phe Phe Glu Gly Leu Arg Arg Ile Arg			
[4102]		690	695	700
[4103]	Lys Glu Ala Asp Asn Leu Pro Ile Val Gly Lys Glu Ala Lys Lys Arg			
[4104]	705	710	715	720
[4105]	Asn Ala Val Gln Asp Phe Gly Arg Arg Cys Asp Glu Leu Lys Asn Leu			
[4106]		725	730	735
[4107]	Ser Leu Ser Ala Ile Cys Gln Met Ile Met Thr Glu Tyr Asn Glu Gln			
[4108]		740	745	750
[4109]	Asn Asn Gly Asn Arg Lys Val Lys Ser Thr Arg Glu Asp Lys Arg Lys			
[4110]		755	760	765
[4111]	Pro Asp Ile Phe Gln His Tyr Lys Met Leu Leu Leu Arg Thr Leu Gln			
[4112]		770	775	780
[4113]	Glu Ala Phe Ala Ile Tyr Ile Arg Arg Glu Glu Phe Lys Phe Ile Phe			
[4114]	785	790	795	800
[4115]	Asp Leu Pro Lys Thr Leu Tyr Val Met Lys Pro Val Glu Glu Phe Leu			
[4116]		805	810	815
[4117]	Pro Asn Trp Lys Ser Gly Met Phe Asp Ser Leu Val Glu Arg Val Lys			
[4118]		820	825	830
[4119]	Gln Ser Pro Asp Leu Gln Arg Trp Tyr Val Leu Cys Lys Phe Leu Asn			
[4120]		835	840	845
[4121]	Gly Arg Leu Leu Asn Gln Leu Ser Gly Val Ile Arg Ser Tyr Ile Gln			
[4122]		850	855	860
[4123]	Phe Ala Gly Asp Ile Gln Arg Arg Ala Lys Ala Asn His Asn Arg Leu			
[4124]	865	870	875	880
[4125]	Tyr Met Asp Asn Thr Gln Arg Val Glu Tyr Tyr Ser Asn Val Leu Glu			
[4126]		885	890	895
[4127]	Val Val Asp Phe Cys Ile Lys Gly Thr Ser Arg Phe Ser Asn Val Phe			
[4128]		900	905	910
[4129]	Ser Asp Tyr Phe Arg Asp Glu Asp Ala Tyr Ala Asp Tyr Leu Asp Asn			
[4130]		915	920	925
[4131]	Tyr Leu Gln Phe Lys Asp Glu Lys Ile Ala Glu Val Ser Ser Phe Ala			
[4132]		930	935	940

[4133]	Ala	Leu	Lys	Thr	Phe	Cys	Asn	Glu	Glu	Glu	Val	Lys	Ala	Gly	Ile	Tyr
[4134]	945						950					955				960
[4135]	Met	Asp	Gly	Glu	Asn	Pro	Val	Met	Gln	Arg	Asn	Ile	Val	Met	Ala	Lys
[4136]							965					970				975
[4137]	Leu	Phe	Gly	Pro	Asp	Glu	Val	Leu	Lys	Asn	Val	Val	Pro	Lys	Val	Thr
[4138]							980					985				990
[4139]	Arg	Glu	Glu	Ile	Glu	Glu	Tyr	Tyr	Gln	Leu	Glu	Lys	Gln	Ile	Ala	Pro
[4140]							995					1000				1005
[4141]	Tyr	Arg	Gln	Asn	Gly	Tyr	Cys	Lys	Ser	Glu	Glu	Asp	Gln	Lys	Lys	
[4142]							1010					1015				1020
[4143]	Leu	Leu	Arg	Phe	Gln	Arg	Ile	Lys	Asn	Arg	Val	Glu	Phe	Gln	Thr	
[4144]							1025					1030				1035
[4145]	Ile	Thr	Glu	Phe	Ser	Glu	Ile	Ile	Asn	Glu	Leu	Leu	Gly	Gln	Leu	
[4146]							1040					1045				1050
[4147]	Ile	Ser	Trp	Ser	Phe	Leu	Arg	Glu	Arg	Asp	Leu	Leu	Tyr	Phe	Gln	
[4148]							1055					1060				1065
[4149]	Leu	Gly	Phe	His	Tyr	Leu	Cys	Leu	His	Asn	Asp	Thr	Glu	Lys	Pro	
[4150]							1070					1075				1080
[4151]	Ala	Glu	Tyr	Lys	Glu	Ile	Ser	Arg	Glu	Asp	Gly	Thr	Val	Ile	Arg	
[4152]							1085					1090				1095
[4153]	Asn	Ala	Ile	Leu	His	Gln	Val	Ala	Ala	Met	Tyr	Val	Gly	Gly	Leu	
[4154]							1100					1105				1110
[4155]	Pro	Val	Tyr	Thr	Leu	Ala	Asp	Lys	Lys	Leu	Ala	Ala	Phe	Glu	Lys	
[4156]							1115					1120				1125
[4157]	Gly	Glu	Ala	Asp	Cys	Lys	Leu	Ser	Ile	Ser	Lys	Asp	Thr	Ala	Gly	
[4158]							1130					1135				1140
[4159]	Ala	Gly	Lys	Lys	Ile	Lys	Asp	Phe	Phe	Arg	Tyr	Ser	Lys	Tyr	Val	
[4160]							1145					1150				1155
[4161]	Leu	Ile	Lys	Asp	Arg	Met	Leu	Thr	Asp	Gln	Asn	Gln	Lys	Tyr	Thr	
[4162]							1160					1165				1170
[4163]	Ile	Tyr	Leu	Ala	Gly	Leu	Glu	Leu	Phe	Glu	Asn	Thr	Asp	Glu	His	
[4164]							1175					1180				1185
[4165]	Asp	Asn	Ile	Thr	Asp	Val	Arg	Lys	Tyr	Val	Asp	His	Phe	Lys	Tyr	
[4166]							1190					1195				1200
[4167]	Tyr	Ala	Thr	Ser	Asp	Glu	Asn	Ala	Met	Ser	Ile	Leu	Asp	Leu	Tyr	
[4168]							1205					1210				1215
[4169]	Ser	Glu	Ile	His	Asp	Arg	Phe	Phe	Thr	Tyr	Asp	Met	Lys	Tyr	Gln	
[4170]							1220					1225				1230
[4171]	Lys	Asn	Val	Ala	Asn	Met	Leu	Glu	Asn	Ile	Leu	Leu	Arg	His	Phe	

[4172]	1235	1240	1245
[4173]	Val Leu Ile Arg Pro Glu Phe	Phe Thr Gly Ser Lys	Lys Val Gly
[4174]	1250	1255	1260
[4175]	Glu Gly Lys Lys Ile Thr Cys	Lys Ala Arg Ala Gln	Ile Glu Ile
[4176]	1265	1270	1275
[4177]	Ala Glu Asn Gly Met Arg Ser	Glu Asp Phe Thr Tyr	Lys Leu Ser
[4178]	1280	1285	1290
[4179]	Asp Gly Lys Lys Asn Ile Ser	Thr Cys Met Ile Ala	Ala Arg Asp
[4180]	1295	1300	1305
[4181]	Gln Lys Tyr Leu Asn Thr Val	Ala Arg Leu Leu Tyr	Tyr Pro His
[4182]	1310	1315	1320
[4183]	Glu Ala Lys Lys Ser Ile Val	Asp Thr Arg Glu Lys	Lys Asn Asn
[4184]	1325	1330	1335
[4185]	Lys Lys Thr Asn Arg Gly Asp	Gly Thr Phe Asn Lys	Gln Lys Gly
[4186]	1340	1345	1350
[4187]	Thr Ala Arg Lys Glu Lys Asp	Asn Gly Pro Arg Glu	Phe Asn Asp
[4188]	1355	1360	1365
[4189]	Thr Gly Phe Ser Asn Thr Pro	Phe Ala Gly Phe Asp	Pro Phe Arg
[4190]	1370	1375	1380
[4191]	Asn Ser		
[4192]	1385		
[4193]	<210>	303	
[4194]	<211>	1051	
[4195]	<212>	PRT	
[4196]	<213>	Listeria newyorkensis	
[4197]	<400>	303	
[4198]	Met Lys Ile Thr Lys Met Arg Val	Asp Gly Arg Thr Ile Val	Met Glu
[4199]	1	5	10 15
[4200]	Arg Thr Ser Lys Glu Gly Gln Leu	Gly Tyr Glu Gly Ile Asp	Gly Asn
[4201]	20	25	30
[4202]	Lys Thr Thr Glu Ile Ile Phe	Asp Lys Lys Lys Glu Ser	Phe Tyr Lys
[4203]	35	40	45
[4204]	Ser Ile Leu Asn Lys Thr Val	Arg Lys Pro Asp Glu Lys	Glu Lys Asn
[4205]	50	55	60
[4206]	Arg Arg Lys Gln Ala Ile Asn	Lys Ala Ile Asn Lys Glu	Ile Thr Glu
[4207]	65	70	75 80
[4208]	Leu Met Leu Ala Val Leu His	Gln Glu Val Pro Ser Gln	Lys Leu His
[4209]	85	90	95
[4210]	Asn Leu Lys Ser Leu Asn Thr	Glu Ser Leu Thr Lys Leu	Phe Lys Pro

[4211]	100	105	110
[4212]	Lys Phe Gln Asn Met Ile Ser Tyr Pro Pro Ser Lys Gly Ala Glu His		
[4213]	115	120	125
[4214]	Val Gln Phe Cys Leu Thr Asp Ile Ala Val Pro Ala Ile Arg Asp Leu		
[4215]	130	135	140
[4216]	Asp Glu Ile Lys Pro Asp Trp Gly Ile Phe Phe Glu Lys Leu Lys Pro		
[4217]	145	150	155
[4218]	Tyr Thr Asp Trp Ala Glu Ser Tyr Ile His Tyr Lys Gln Thr Thr Ile		
[4219]	165	170	175
[4220]	Gln Lys Ser Ile Glu Gln Asn Lys Ile Gln Ser Pro Asp Ser Pro Arg		
[4221]	180	185	190
[4222]	Lys Leu Val Leu Gln Lys Tyr Val Thr Ala Phe Leu Asn Gly Glu Pro		
[4223]	195	200	205
[4224]	Leu Gly Leu Asp Leu Val Ala Lys Lys Tyr Lys Leu Ala Asp Leu Ala		
[4225]	210	215	220
[4226]	Glu Ser Phe Lys Leu Val Asp Leu Asn Glu Asp Lys Ser Ala Asn Tyr		
[4227]	225	230	235
[4228]	Lys Ile Lys Ala Cys Leu Gln Gln His Gln Arg Asn Ile Leu Asp Glu		
[4229]	245	250	255
[4230]	Leu Lys Glu Asp Pro Glu Leu Asn Gln Tyr Gly Ile Glu Val Lys Lys		
[4231]	260	265	270
[4232]	Tyr Ile Gln Arg Tyr Phe Pro Ile Lys Arg Ala Pro Asn Arg Ser Lys		
[4233]	275	280	285
[4234]	His Ala Arg Ala Asp Phe Leu Lys Lys Glu Leu Ile Glu Ser Thr Val		
[4235]	290	295	300
[4236]	Glu Gln Gln Phe Lys Asn Ala Val Tyr His Tyr Val Leu Glu Gln Gly		
[4237]	305	310	315
[4238]	Lys Met Glu Ala Tyr Glu Leu Thr Asp Pro Lys Thr Lys Asp Leu Gln		
[4239]	325	330	335
[4240]	Asp Ile Arg Ser Gly Glu Ala Phe Ser Phe Lys Phe Ile Asn Ala Cys		
[4241]	340	345	350
[4242]	Ala Phe Ala Ser Asn Asn Leu Lys Met Ile Leu Asn Pro Glu Cys Glu		
[4243]	355	360	365
[4244]	Lys Asp Ile Leu Gly Lys Gly Asn Phe Lys Lys Asn Leu Pro Asn Ser		
[4245]	370	375	380
[4246]	Thr Thr Arg Ser Asp Val Val Lys Lys Met Ile Pro Phe Phe Ser Asp		
[4247]	385	390	395
[4248]	Glu Leu Gln Asn Val Asn Phe Asp Glu Ala Ile Trp Ala Ile Arg Gly		
[4249]	405	410	415

[4250]	Ser	Ile	Gln	Gln	Ile	Arg	Asn	Glu	Val	Tyr	His	Cys	Lys	Lys	His	Ser
[4251]				420					425					430		
[4252]	Trp	Lys	Ser	Ile	Leu	Lys	Ile	Lys	Gly	Phe	Glu	Phe	Glu	Pro	Asn	Asn
[4253]				435					440					445		
[4254]	Met	Lys	Tyr	Ala	Asp	Ser	Asp	Met	Gln	Lys	Leu	Met	Asp	Lys	Asp	Ile
[4255]				450					455					460		
[4256]	Ala	Lys	Ile	Pro	Glu	Phe	Ile	Glu	Glu	Lys	Leu	Lys	Ser	Ser	Gly	Val
[4257]	465						470					475				480
[4258]	Val	Arg	Phe	Tyr	Arg	His	Asp	Glu	Leu	Gln	Ser	Ile	Trp	Glu	Met	Lys
[4259]					485					490					495	
[4260]	Gln	Gly	Phe	Ser	Leu	Leu	Thr	Thr	Asn	Ala	Pro	Phe	Val	Pro	Ser	Phe
[4261]				500						505					510	
[4262]	Lys	Arg	Val	Tyr	Ala	Lys	Gly	His	Asp	Tyr	Gln	Thr	Ser	Lys	Asn	Arg
[4263]				515						520					525	
[4264]	Tyr	Tyr	Asn	Leu	Asp	Leu	Thr	Thr	Phe	Asp	Ile	Leu	Glu	Tyr	Gly	Glu
[4265]				530						535				540		
[4266]	Glu	Asp	Phe	Arg	Ala	Arg	Tyr	Phe	Leu	Thr	Lys	Leu	Val	Tyr	Tyr	Gln
[4267]	545						550					555				560
[4268]	Gln	Phe	Met	Pro	Trp	Phe	Thr	Ala	Asp	Asn	Asn	Ala	Phe	Arg	Asp	Ala
[4269]					565						570				575	
[4270]	Ala	Asn	Phe	Val	Leu	Arg	Leu	Asn	Lys	Asn	Arg	Gln	Gln	Asp	Ala	Lys
[4271]				580							585				590	
[4272]	Ala	Phe	Ile	Asn	Ile	Arg	Glu	Val	Glu	Glu	Gly	Glu	Met	Pro	Arg	Asp
[4273]				595						600				605		
[4274]	Tyr	Met	Gly	Tyr	Val	Gln	Gly	Gln	Ile	Ala	Ile	His	Glu	Asp	Ser	Ile
[4275]				610						615				620		
[4276]	Glu	Asp	Thr	Pro	Asn	His	Phe	Glu	Lys	Phe	Ile	Ser	Gln	Val	Phe	Ile
[4277]	625						630					635				640
[4278]	Lys	Gly	Phe	Asp	Arg	His	Met	Arg	Ser	Ala	Asn	Leu	Lys	Phe	Ile	Lys
[4279]					645						650				655	
[4280]	Asn	Pro	Arg	Asn	Gln	Gly	Leu	Glu	Gln	Ser	Glu	Ile	Glu	Glu	Met	Ser
[4281]				660							665				670	
[4282]	Phe	Asp	Ile	Lys	Val	Glu	Pro	Ser	Phe	Leu	Lys	Asn	Lys	Asp	Asp	Tyr
[4283]				675						680				685		
[4284]	Ile	Ala	Phe	Trp	Ile	Phe	Cys	Lys	Met	Leu	Asp	Ala	Arg	His	Leu	Ser
[4285]				690						695				700		
[4286]	Glu	Leu	Arg	Asn	Glu	Met	Ile	Lys	Tyr	Asp	Gly	His	Leu	Thr	Gly	Glu
[4287]	705						710					715				720
[4288]	Gln	Glu	Ile	Ile	Gly	Leu	Ala	Leu	Leu	Gly	Val	Asp	Ser	Arg	Glu	Asn

[4289]	725																730				735							
[4290]	Asp	Trp	Lys	Gln	Phe	Phe	Ser	Ser	Glu	Arg	Glu	Tyr	Glu	Lys	Ile	Met												
[4291]	740																745				750							
[4292]	Lys	Gly	Tyr	Val	Val	Glu	Glu	Leu	Tyr	Gln	Arg	Glu	Pro	Tyr	Arg	Gln												
[4293]	755																760				765							
[4294]	Ser	Asp	Gly	Lys	Thr	Pro	Ile	Leu	Phe	Arg	Gly	Val	Glu	Gln	Ala	Arg												
[4295]	770																775				780							
[4296]	Lys	Tyr	Gly	Thr	Glu	Thr	Val	Ile	Gln	Arg	Leu	Phe	Asp	Ala	Asn	Pro												
[4297]	785																790				795				800			
[4298]	Glu	Phe	Lys	Val	Ser	Lys	Cys	Asn	Leu	Ala	Glu	Trp	Glu	Arg	Gln	Lys												
[4299]	805																810				815							
[4300]	Glu	Thr	Ile	Glu	Glu	Thr	Ile	Lys	Arg	Arg	Lys	Glu	Leu	His	Asn	Glu												
[4301]	820																825				830							
[4302]	Trp	Ala	Lys	Asn	Pro	Lys	Lys	Pro	Gln	Asn	Asn	Ala	Phe	Phe	Lys	Glu												
[4303]	835																840				845							
[4304]	Tyr	Lys	Glu	Cys	Cys	Asp	Ala	Ile	Asp	Ala	Tyr	Asn	Trp	His	Lys	Asn												
[4305]	850																855				860							
[4306]	Lys	Thr	Thr	Leu	Ala	Tyr	Val	Asn	Glu	Leu	His	His	Leu	Leu	Ile	Glu												
[4307]	865																870				875				880			
[4308]	Ile	Leu	Gly	Arg	Tyr	Val	Gly	Tyr	Val	Ala	Ile	Ala	Asp	Arg	Asp	Phe												
[4309]	885																890				895							
[4310]	Gln	Cys	Met	Ala	Asn	Gln	Tyr	Phe	Lys	His	Ser	Gly	Ile	Thr	Glu	Arg												
[4311]	900																905				910							
[4312]	Val	Glu	Tyr	Trp	Gly	Asp	Asn	Arg	Leu	Lys	Ser	Ile	Lys	Lys	Leu	Asp												
[4313]	915																920				925							
[4314]	Thr	Phe	Leu	Lys	Lys	Glu	Gly	Leu	Phe	Val	Ser	Glu	Lys	Asn	Ala	Arg												
[4315]	930																935				940							
[4316]	Asn	His	Ile	Ala	His	Leu	Asn	Tyr	Leu	Ser	Leu	Lys	Ser	Glu	Cys	Thr												
[4317]	945																950				955				960			
[4318]	Leu	Leu	Tyr	Leu	Ser	Glu	Arg	Leu	Arg	Glu	Ile	Phe	Lys	Tyr	Asp	Arg												
[4319]	965																970				975							
[4320]	Lys	Leu	Lys	Asn	Ala	Val	Ser	Lys	Ser	Leu	Ile	Asp	Ile	Leu	Asp	Arg												
[4321]	980																985				990							
[4322]	His	Gly	Met	Ser	Val	Val	Phe	Ala	Asn	Leu	Lys	Glu	Asn	Lys	His	Arg												
[4323]	995																1000				1005							
[4324]	Leu	Val	Ile	Lys	Ser	Leu	Glu	Pro	Lys	Lys	Leu	Arg	His	Leu	Gly													
[4325]	1010																1015				1020							
[4326]	Gly	Lys	Lys	Ile	Asp	Gly	Gly	Tyr	Ile	Glu	Thr	Asn	Gln	Val	Ser													
[4327]	1025																1030				1035							

[4328]	Glu Glu Tyr Cys Gly Ile Val Lys Arg Leu Leu Glu Met
[4329]	1040 1045 1050
[4330]	<210> 304
[4331]	<211> 1344
[4332]	<212> PRT
[4333]	<213> 未知的
[4334]	<220>
[4335]	<223> Eubacterium rectale genome assembly T1815
[4336]	<400> 304
[4337]	Met Leu Arg Arg Asp Lys Glu Val Lys Lys Leu Tyr Asn Val Phe Asn
[4338]	1 5 10 15
[4339]	Gln Ile Gln Val Gly Thr Lys Pro Lys Lys Trp Asn Asn Asp Glu Lys
[4340]	20 25 30
[4341]	Leu Ser Pro Glu Glu Asn Glu Arg Arg Ala Gln Gln Lys Asn Ile Lys
[4342]	35 40 45
[4343]	Met Lys Asn Tyr Lys Trp Arg Glu Ala Cys Ser Lys Tyr Val Glu Ser
[4344]	50 55 60
[4345]	Ser Gln Arg Ile Ile Asn Asp Val Ile Phe Tyr Ser Tyr Arg Lys Ala
[4346]	65 70 75 80
[4347]	Lys Asn Lys Leu Arg Tyr Met Arg Lys Asn Glu Asp Ile Leu Lys Lys
[4348]	85 90 95
[4349]	Met Gln Glu Ala Glu Lys Leu Ser Lys Phe Ser Gly Gly Lys Leu Glu
[4350]	100 105 110
[4351]	Asp Phe Val Ala Tyr Thr Leu Arg Lys Ser Leu Val Val Ser Lys Tyr
[4352]	115 120 125
[4353]	Asp Thr Gln Glu Phe Asp Ser Leu Ala Ala Met Val Val Phe Leu Glu
[4354]	130 135 140
[4355]	Cys Ile Gly Lys Asn Asn Ile Ser Asp His Glu Arg Glu Ile Val Cys
[4356]	145 150 155 160
[4357]	Lys Leu Leu Glu Leu Ile Arg Lys Asp Phe Ser Lys Leu Asp Pro Asn
[4358]	165 170 175
[4359]	Val Lys Gly Ser Gln Gly Ala Asn Ile Val Arg Ser Val Arg Asn Gln
[4360]	180 185 190
[4361]	Asn Met Ile Val Gln Pro Gln Gly Asp Arg Phe Leu Phe Pro Gln Val
[4362]	195 200 205
[4363]	Tyr Ala Lys Glu Asn Glu Thr Val Thr Asn Lys Asn Val Glu Lys Glu
[4364]	210 215 220
[4365]	Gly Leu Asn Glu Phe Leu Leu Asn Tyr Ala Asn Leu Asp Asp Glu Lys
[4366]	225 230 235 240

[4367]	Arg	Ala	Glu	Ser	Leu	Arg	Lys	Leu	Arg	Arg	Ile	Leu	Asp	Val	Tyr	Phe
[4368]					245					250					255	
[4369]	Ser	Ala	Pro	Asn	His	Tyr	Glu	Lys	Asp	Met	Asp	Ile	Thr	Leu	Ser	Asp
[4370]					260					265					270	
[4371]	Asn	Ile	Glu	Lys	Glu	Lys	Phe	Asn	Val	Trp	Glu	Lys	His	Glu	Cys	Gly
[4372]					275					280					285	
[4373]	Lys	Lys	Glu	Thr	Gly	Leu	Phe	Val	Asp	Ile	Pro	Asp	Val	Leu	Met	Glu
[4374]			290					295					300			
[4375]	Ala	Glu	Ala	Glu	Asn	Ile	Lys	Leu	Asp	Ala	Val	Val	Glu	Lys	Arg	Glu
[4376]	305					310					315				320	
[4377]	Arg	Lys	Val	Leu	Asn	Asp	Arg	Val	Arg	Lys	Gln	Asn	Ile	Ile	Cys	Tyr
[4378]					325					330					335	
[4379]	Arg	Tyr	Thr	Arg	Ala	Val	Val	Glu	Lys	Tyr	Asn	Ser	Asn	Glu	Pro	Leu
[4380]					340					345					350	
[4381]	Phe	Phe	Glu	Asn	Asn	Ala	Ile	Asn	Gln	Tyr	Trp	Ile	His	His	Ile	Glu
[4382]					355					360					365	
[4383]	Asn	Ala	Val	Glu	Arg	Ile	Leu	Lys	Asn	Cys	Lys	Ala	Gly	Lys	Leu	Phe
[4384]					370					375					380	
[4385]	Lys	Leu	Arg	Lys	Gly	Tyr	Leu	Ala	Glu	Lys	Val	Trp	Lys	Asp	Ala	Ile
[4386]	385					390					395				400	
[4387]	Asn	Leu	Ile	Ser	Ile	Lys	Tyr	Ile	Ala	Leu	Gly	Lys	Ala	Val	Tyr	Asn
[4388]					405					410					415	
[4389]	Phe	Ala	Leu	Asp	Asp	Ile	Trp	Lys	Asp	Lys	Lys	Asn	Lys	Glu	Leu	Gly
[4390]					420					425					430	
[4391]	Ile	Val	Asp	Glu	Arg	Ile	Arg	Asn	Gly	Ile	Thr	Ser	Phe	Asp	Tyr	Glu
[4392]					435					440					445	
[4393]	Met	Ile	Lys	Ala	His	Glu	Asn	Leu	Gln	Arg	Glu	Leu	Ala	Val	Asp	Ile
[4394]					450					455					460	
[4395]	Ala	Phe	Ser	Val	Asn	Asn	Leu	Ala	Arg	Ala	Val	Cys	Asp	Met	Ser	Asn
[4396]	465					470					475				480	
[4397]	Leu	Gly	Asn	Lys	Glu	Ser	Asp	Phe	Leu	Leu	Trp	Lys	Arg	Asn	Asp	Ile
[4398]					485					490					495	
[4399]	Ala	Asp	Lys	Leu	Lys	Asn	Lys	Asp	Asp	Met	Ala	Ser	Val	Ser	Ala	Val
[4400]					500					505					510	
[4401]	Leu	Gln	Phe	Phe	Gly	Gly	Lys	Ser	Ser	Trp	Asp	Ile	Asn	Ile	Phe	Lys
[4402]					515					520					525	
[4403]	Asp	Ala	Tyr	Lys	Gly	Lys	Lys	Lys	Tyr	Asn	Tyr	Glu	Val	Arg	Phe	Ile
[4404]					530					535					540	
[4405]	Asp	Asp	Leu	Arg	Lys	Ala	Ile	Tyr	Cys	Ala	Arg	Asn	Glu	Asn	Phe	His

[4406]	545	550	555	560
[4407]	Phe Lys Thr Ala Leu Val Asn Asp Glu Lys Trp Asn Thr Glu Leu Phe			
[4408]		565	570	575
[4409]	Gly Lys Ile Phe Glu Arg Glu Thr Glu Phe Cys Leu Asn Val Glu Lys			
[4410]		580	585	590
[4411]	Asp Arg Phe Tyr Ser Asn Asn Leu Tyr Met Phe Tyr Gln Val Ser Glu			
[4412]		595	600	605
[4413]	Leu Arg Asn Met Leu Asp His Leu Tyr Ser Arg Ser Val Ser Arg Ala			
[4414]		610	615	620
[4415]	Ala Gln Val Pro Ser Tyr Asn Ser Val Ile Val Arg Thr Ala Phe Pro			
[4416]	625	630	635	640
[4417]	Glu Tyr Ile Thr Asn Val Leu Gly Tyr Gln Lys Pro Ser Tyr Asp Ala			
[4418]		645	650	655
[4419]	Asp Thr Leu Gly Lys Trp Tyr Ser Ala Cys Tyr Tyr Leu Leu Lys Glu			
[4420]		660	665	670
[4421]	Ile Tyr Tyr Asn Ser Phe Leu Gln Ser Asp Arg Ala Leu Gln Leu Phe			
[4422]		675	680	685
[4423]	Glu Lys Ser Val Lys Thr Leu Ser Trp Asp Asp Lys Lys Gln Gln Arg			
[4424]		690	695	700
[4425]	Ala Val Asp Asn Phe Lys Asp His Phe Ser Asp Ile Lys Ser Ala Cys			
[4426]	705	710	715	720
[4427]	Thr Ser Leu Ala Gln Val Cys Gln Ile Tyr Met Thr Glu Tyr Asn Gln			
[4428]		725	730	735
[4429]	Gln Asn Asn Gln Ile Lys Lys Val Arg Ser Ser Asn Asp Ser Ile Phe			
[4430]		740	745	750
[4431]	Asp Gln Pro Val Tyr Gln His Tyr Lys Val Leu Leu Lys Lys Ala Ile			
[4432]		755	760	765
[4433]	Ala Asn Ala Phe Ala Asp Tyr Leu Lys Asn Asn Lys Asp Leu Phe Gly			
[4434]		770	775	780
[4435]	Phe Ile Gly Lys Pro Phe Lys Ala Asn Glu Ile Arg Glu Ile Asp Lys			
[4436]	785	790	795	800
[4437]	Glu Gln Phe Leu Pro Asp Trp Thr Ser Arg Lys Tyr Glu Ala Leu Cys			
[4438]		805	810	815
[4439]	Ile Glu Val Ser Gly Ser Gln Glu Leu Gln Lys Trp Tyr Ile Val Gly			
[4440]		820	825	830
[4441]	Lys Phe Leu Asn Ala Arg Ser Leu Asn Leu Met Val Gly Ser Met Arg			
[4442]		835	840	845
[4443]	Ser Tyr Ile Gln Tyr Val Thr Asp Ile Lys Arg Arg Ala Ala Ser Ile			
[4444]		850	855	860

[4445]	Gly Asn Glu Leu His Val Ser Val His Asp Val Glu Lys Val Glu Lys
[4446]	865 870 875 880
[4447]	Trp Val Gln Val Ile Glu Val Cys Ser Leu Leu Ala Ser Arg Thr Ser
[4448]	885 890 895
[4449]	Asn Gln Phe Glu Asp Tyr Phe Asn Asp Lys Asp Asp Tyr Ala Arg Tyr
[4450]	900 905 910
[4451]	Leu Lys Ser Tyr Val Asp Phe Ser Asn Val Asp Met Pro Ser Glu Tyr
[4452]	915 920 925
[4453]	Ser Ala Leu Val Asp Phe Ser Asn Glu Glu Gln Ser Asp Leu Tyr Val
[4454]	930 935 940
[4455]	Asp Pro Lys Asn Pro Lys Val Asn Arg Asn Ile Val His Ser Lys Leu
[4456]	945 950 955 960
[4457]	Phe Ala Ala Asp His Ile Leu Arg Asp Ile Val Glu Pro Val Ser Lys
[4458]	965 970 975
[4459]	Asp Asn Ile Glu Glu Phe Tyr Ser Gln Lys Ala Glu Ile Ala Tyr Cys
[4460]	980 985 990
[4461]	Lys Ile Lys Gly Lys Glu Ile Thr Ala Glu Glu Gln Lys Ala Val Leu
[4462]	995 1000 1005
[4463]	Lys Tyr Gln Lys Leu Lys Asn Arg Val Glu Leu Arg Asp Ile Val
[4464]	1010 1015 1020
[4465]	Glu Tyr Gly Glu Ile Ile Asn Glu Leu Leu Gly Gln Leu Ile Asn
[4466]	1025 1030 1035
[4467]	Trp Ser Phe Met Arg Glu Arg Asp Leu Leu Tyr Phe Gln Leu Gly
[4468]	1040 1045 1050
[4469]	Phe His Tyr Asp Cys Leu Arg Asn Asp Ser Lys Lys Pro Glu Gly
[4470]	1055 1060 1065
[4471]	Tyr Lys Asn Ile Lys Val Asp Glu Asn Ser Ile Lys Asp Ala Ile
[4472]	1070 1075 1080
[4473]	Leu Tyr Gln Ile Ile Gly Met Tyr Val Asn Gly Val Thr Val Tyr
[4474]	1085 1090 1095
[4475]	Ala Pro Glu Lys Asp Gly Asp Lys Leu Lys Glu Gln Cys Val Lys
[4476]	1100 1105 1110
[4477]	Gly Gly Val Gly Val Lys Val Ser Ala Phe His Arg Tyr Ser Lys
[4478]	1115 1120 1125
[4479]	Tyr Leu Gly Leu Asn Glu Lys Thr Leu Tyr Asn Ala Gly Leu Glu
[4480]	1130 1135 1140
[4481]	Ile Phe Glu Val Val Ala Glu His Glu Asp Ile Ile Asn Leu Arg
[4482]	1145 1150 1155
[4483]	Asn Gly Ile Asp His Phe Lys Tyr Tyr Leu Gly Asp Tyr Arg Ser

[4484]	1160	1165	1170
[4485]	Met Leu Ser Ile Tyr Ser Glu Val Phe Asp Arg Phe Phe Thr Tyr		
[4486]	1175	1180	1185
[4487]	Asp Ile Lys Tyr Gln Lys Asn Val Leu Asn Leu Leu Gln Asn Ile		
[4488]	1190	1195	1200
[4489]	Leu Leu Arg His Asn Val Ile Val Glu Pro Ile Leu Glu Ser Gly		
[4490]	1205	1210	1215
[4491]	Phe Lys Thr Ile Gly Glu Gln Thr Lys Pro Gly Ala Lys Leu Ser		
[4492]	1220	1225	1230
[4493]	Ile Arg Ser Ile Lys Ser Asp Thr Phe Gln Tyr Lys Val Lys Gly		
[4494]	1235	1240	1245
[4495]	Gly Thr Leu Ile Thr Asp Ala Lys Asp Glu Arg Tyr Leu Glu Thr		
[4496]	1250	1255	1260
[4497]	Ile Arg Lys Ile Leu Tyr Tyr Ala Glu Asn Glu Glu Asp Asn Leu		
[4498]	1265	1270	1275
[4499]	Lys Lys Ser Val Val Val Thr Asn Ala Asp Lys Tyr Glu Lys Asn		
[4500]	1280	1285	1290
[4501]	Lys Glu Ser Asp Asp Gln Asn Lys Gln Lys Glu Lys Lys Asn Lys		
[4502]	1295	1300	1305
[4503]	Asp Asn Lys Gly Lys Lys Asn Glu Glu Thr Lys Ser Asp Ala Glu		
[4504]	1310	1315	1320
[4505]	Lys Asn Asn Asn Glu Arg Leu Ser Tyr Asn Pro Phe Ala Asn Leu		
[4506]	1325	1330	1335
[4507]	Asn Phe Lys Leu Ser Asn		
[4508]	1340		
[4509]	<210> 305		
[4510]	<211> 1285		
[4511]	<212> PRT		
[4512]	<213> Herbinix hemicellulosilytica		
[4513]	<400> 305		
[4514]	Met Lys Leu Thr Arg Arg Arg Ile Ser Gly Asn Ser Val Asp Gln Lys		
[4515]	1 5 10 15		
[4516]	Ile Thr Ala Ala Phe Tyr Arg Asp Met Ser Gln Gly Leu Leu Tyr Tyr		
[4517]	20 25 30		
[4518]	Asp Ser Glu Asp Asn Asp Cys Thr Asp Lys Val Ile Glu Ser Met Asp		
[4519]	35 40 45		
[4520]	Phe Glu Arg Ser Trp Arg Gly Arg Ile Leu Lys Asn Gly Glu Asp Asp		
[4521]	50 55 60		
[4522]	Lys Asn Pro Phe Tyr Met Phe Val Lys Gly Leu Val Gly Ser Asn Asp		

[4523]	65	70	75	80
[4524]	Lys Ile Val Cys Glu Pro Ile Asp Val Asp Ser Asp Pro Asp Asn Leu			
[4525]	85	90	95	
[4526]	Asp Ile Leu Ile Asn Lys Asn Leu Thr Gly Phe Gly Arg Asn Leu Lys			
[4527]	100	105	110	
[4528]	Ala Pro Asp Ser Asn Asp Thr Leu Glu Asn Leu Ile Arg Lys Ile Gln			
[4529]	115	120	125	
[4530]	Ala Gly Ile Pro Glu Glu Glu Val Leu Pro Glu Leu Lys Lys Ile Lys			
[4531]	130	135	140	
[4532]	Glu Met Ile Gln Lys Asp Ile Val Asn Arg Lys Glu Gln Leu Leu Lys			
[4533]	145	150	155	160
[4534]	Ser Ile Lys Asn Asn Arg Ile Pro Phe Ser Leu Glu Gly Ser Lys Leu			
[4535]	165	170	175	
[4536]	Val Pro Ser Thr Lys Lys Met Lys Trp Leu Phe Lys Leu Ile Asp Val			
[4537]	180	185	190	
[4538]	Pro Asn Lys Thr Phe Asn Glu Lys Met Leu Glu Lys Tyr Trp Glu Ile			
[4539]	195	200	205	
[4540]	Tyr Asp Tyr Asp Lys Leu Lys Ala Asn Ile Thr Asn Arg Leu Asp Lys			
[4541]	210	215	220	
[4542]	Thr Asp Lys Lys Ala Arg Ser Ile Ser Arg Ala Val Ser Glu Glu Leu			
[4543]	225	230	235	240
[4544]	Arg Glu Tyr His Lys Asn Leu Arg Thr Asn Tyr Asn Arg Phe Val Ser			
[4545]	245	250	255	
[4546]	Gly Asp Arg Pro Ala Ala Gly Leu Asp Asn Gly Gly Ser Ala Lys Tyr			
[4547]	260	265	270	
[4548]	Asn Pro Asp Lys Glu Glu Phe Leu Leu Phe Leu Lys Glu Val Glu Gln			
[4549]	275	280	285	
[4550]	Tyr Phe Lys Lys Tyr Phe Pro Val Lys Ser Lys His Ser Asn Lys Ser			
[4551]	290	295	300	
[4552]	Lys Asp Lys Ser Leu Val Asp Lys Tyr Lys Asn Tyr Cys Ser Tyr Lys			
[4553]	305	310	315	320
[4554]	Val Val Lys Lys Glu Val Asn Arg Ser Ile Ile Asn Gln Leu Val Ala			
[4555]	325	330	335	
[4556]	Gly Leu Ile Gln Gln Gly Lys Leu Leu Tyr Tyr Phe Tyr Tyr Asn Asp			
[4557]	340	345	350	
[4558]	Thr Trp Gln Glu Asp Phe Leu Asn Ser Tyr Gly Leu Ser Tyr Ile Gln			
[4559]	355	360	365	
[4560]	Val Glu Glu Ala Phe Lys Lys Ser Val Met Thr Ser Leu Ser Trp Gly			
[4561]	370	375	380	

[4562]	Ile Asn Arg Leu Thr Ser Phe Phe Ile Asp Asp Ser Asn Thr Val Lys
[4563]	385 390 395 400
[4564]	Phe Asp Asp Ile Thr Thr Lys Lys Ala Lys Glu Ala Ile Glu Ser Asn
[4565]	405 410 415
[4566]	Tyr Phe Asn Lys Leu Arg Thr Cys Ser Arg Met Gln Asp His Phe Lys
[4567]	420 425 430
[4568]	Glu Lys Leu Ala Phe Phe Tyr Pro Val Tyr Val Lys Asp Lys Lys Asp
[4569]	435 440 445
[4570]	Arg Pro Asp Asp Asp Ile Glu Asn Leu Ile Val Leu Val Lys Asn Ala
[4571]	450 455 460
[4572]	Ile Glu Ser Val Ser Tyr Leu Arg Asn Arg Thr Phe His Phe Lys Glu
[4573]	465 470 475 480
[4574]	Ser Ser Leu Leu Glu Leu Leu Lys Glu Leu Asp Asp Lys Asn Ser Gly
[4575]	485 490 495
[4576]	Gln Asn Lys Ile Asp Tyr Ser Val Ala Ala Glu Phe Ile Lys Arg Asp
[4577]	500 505 510
[4578]	Ile Glu Asn Leu Tyr Asp Val Phe Arg Glu Gln Ile Arg Ser Leu Gly
[4579]	515 520 525
[4580]	Ile Ala Glu Tyr Tyr Lys Ala Asp Met Ile Ser Asp Cys Phe Lys Thr
[4581]	530 535 540
[4582]	Cys Gly Leu Glu Phe Ala Leu Tyr Ser Pro Lys Asn Ser Leu Met Pro
[4583]	545 550 555 560
[4584]	Ala Phe Lys Asn Val Tyr Lys Arg Gly Ala Asn Leu Asn Lys Ala Tyr
[4585]	565 570 575
[4586]	Ile Arg Asp Lys Gly Pro Lys Glu Thr Gly Asp Gln Gly Gln Asn Ser
[4587]	580 585 590
[4588]	Tyr Lys Ala Leu Glu Glu Tyr Arg Glu Leu Thr Trp Tyr Ile Glu Val
[4589]	595 600 605
[4590]	Lys Asn Asn Asp Gln Ser Tyr Asn Ala Tyr Lys Asn Leu Leu Gln Leu
[4591]	610 615 620
[4592]	Ile Tyr Tyr His Ala Phe Leu Pro Glu Val Arg Glu Asn Glu Ala Leu
[4593]	625 630 635 640
[4594]	Ile Thr Asp Phe Ile Asn Arg Thr Lys Glu Trp Asn Arg Lys Glu Thr
[4595]	645 650 655
[4596]	Glu Glu Arg Leu Asn Thr Lys Asn Asn Lys Lys His Lys Asn Phe Asp
[4597]	660 665 670
[4598]	Glu Asn Asp Asp Ile Thr Val Asn Thr Tyr Arg Tyr Glu Ser Ile Pro
[4599]	675 680 685
[4600]	Asp Tyr Gln Gly Glu Ser Leu Asp Asp Tyr Leu Lys Val Leu Gln Arg

[4601]	690	695	700
[4602]	Lys Gln Met Ala Arg Ala Lys Glu Val Asn Glu Lys Glu Glu Gly Asn		
[4603]	705	710	715
[4604]	Asn Asn Tyr Ile Gln Phe Ile Arg Asp Val Val Val Trp Ala Phe Gly		
[4605]	725	730	735
[4606]	Ala Tyr Leu Glu Asn Lys Leu Lys Asn Tyr Lys Asn Glu Leu Gln Pro		
[4607]	740	745	750
[4608]	Pro Leu Ser Lys Glu Asn Ile Gly Leu Asn Asp Thr Leu Lys Glu Leu		
[4609]	755	760	765
[4610]	Phe Pro Glu Glu Lys Val Lys Ser Pro Phe Asn Ile Lys Cys Arg Phe		
[4611]	770	775	780
[4612]	Ser Ile Ser Thr Phe Ile Asp Asn Lys Gly Lys Ser Thr Asp Asn Thr		
[4613]	785	790	795
[4614]	Ser Ala Glu Ala Val Lys Thr Asp Gly Lys Glu Asp Glu Lys Asp Lys		
[4615]	805	810	815
[4616]	Lys Asn Ile Lys Arg Lys Asp Leu Leu Cys Phe Tyr Leu Phe Leu Arg		
[4617]	820	825	830
[4618]	Leu Leu Asp Glu Asn Glu Ile Cys Lys Leu Gln His Gln Phe Ile Lys		
[4619]	835	840	845
[4620]	Tyr Arg Cys Ser Leu Lys Glu Arg Arg Phe Pro Gly Asn Arg Thr Lys		
[4621]	850	855	860
[4622]	Leu Glu Lys Glu Thr Glu Leu Leu Ala Glu Leu Glu Glu Leu Met Glu		
[4623]	865	870	875
[4624]	Leu Val Arg Phe Thr Met Pro Ser Ile Pro Glu Ile Ser Ala Lys Ala		
[4625]	885	890	895
[4626]	Glu Ser Gly Tyr Asp Thr Met Ile Lys Lys Tyr Phe Lys Asp Phe Ile		
[4627]	900	905	910
[4628]	Glu Lys Lys Val Phe Lys Asn Pro Lys Thr Ser Asn Leu Tyr Tyr His		
[4629]	915	920	925
[4630]	Ser Asp Ser Lys Thr Pro Val Thr Arg Lys Tyr Met Ala Leu Leu Met		
[4631]	930	935	940
[4632]	Arg Ser Ala Pro Leu His Leu Tyr Lys Asp Ile Phe Lys Gly Tyr Tyr		
[4633]	945	950	955
[4634]	Leu Ile Thr Lys Lys Glu Cys Leu Glu Tyr Ile Lys Leu Ser Asn Ile		
[4635]	965	970	975
[4636]	Ile Lys Asp Tyr Gln Asn Ser Leu Asn Glu Leu His Glu Gln Leu Glu		
[4637]	980	985	990
[4638]	Arg Ile Lys Leu Lys Ser Glu Lys Gln Asn Gly Lys Asp Ser Leu Tyr		
[4639]	995	1000	1005

[4640]	Leu	Asp	Lys	Lys	Asp	Phe	Tyr	Lys	Val	Lys	Glu	Tyr	Val	Glu	Asn
[4641]		1010					1015					1020			
[4642]	Leu	Glu	Gln	Val	Ala	Arg	Tyr	Lys	His	Leu	Gln	His	Lys	Ile	Asn
[4643]		1025					1030					1035			
[4644]	Phe	Glu	Ser	Leu	Tyr	Arg	Ile	Phe	Arg	Ile	His	Val	Asp	Ile	Ala
[4645]		1040					1045					1050			
[4646]	Ala	Arg	Met	Val	Gly	Tyr	Thr	Gln	Asp	Trp	Glu	Arg	Asp	Met	His
[4647]		1055					1060					1065			
[4648]	Phe	Leu	Phe	Lys	Ala	Leu	Val	Tyr	Asn	Gly	Val	Leu	Glu	Glu	Arg
[4649]		1070					1075					1080			
[4650]	Arg	Phe	Glu	Ala	Ile	Phe	Asn	Asn	Asn	Asp	Asp	Asn	Asn	Asp	Gly
[4651]		1085					1090					1095			
[4652]	Arg	Ile	Val	Lys	Lys	Ile	Gln	Asn	Asn	Leu	Asn	Asn	Lys	Asn	Arg
[4653]		1100					1105					1110			
[4654]	Glu	Leu	Val	Ser	Met	Leu	Cys	Trp	Asn	Lys	Lys	Leu	Asn	Lys	Asn
[4655]		1115					1120					1125			
[4656]	Glu	Phe	Gly	Ala	Ile	Ile	Trp	Lys	Arg	Asn	Pro	Ile	Ala	His	Leu
[4657]		1130					1135					1140			
[4658]	Asn	His	Phe	Thr	Gln	Thr	Glu	Gln	Asn	Ser	Lys	Ser	Ser	Leu	Glu
[4659]		1145					1150					1155			
[4660]	Ser	Leu	Ile	Asn	Ser	Leu	Arg	Ile	Leu	Leu	Ala	Tyr	Asp	Arg	Lys
[4661]		1160					1165					1170			
[4662]	Arg	Gln	Asn	Ala	Val	Thr	Lys	Thr	Ile	Asn	Asp	Leu	Leu	Leu	Asn
[4663]		1175					1180					1185			
[4664]	Asp	Tyr	His	Ile	Arg	Ile	Lys	Trp	Glu	Gly	Arg	Val	Asp	Glu	Gly
[4665]		1190					1195					1200			
[4666]	Gln	Ile	Tyr	Phe	Asn	Ile	Lys	Glu	Lys	Glu	Asp	Ile	Glu	Asn	Glu
[4667]		1205					1210					1215			
[4668]	Pro	Ile	Ile	His	Leu	Lys	His	Leu	His	Lys	Lys	Asp	Cys	Tyr	Ile
[4669]		1220					1225					1230			
[4670]	Tyr	Lys	Asn	Ser	Tyr	Met	Phe	Asp	Lys	Gln	Lys	Glu	Trp	Ile	Cys
[4671]		1235					1240					1245			
[4672]	Asn	Gly	Ile	Lys	Glu	Glu	Val	Tyr	Asp	Lys	Ser	Ile	Leu	Lys	Cys
[4673]		1250					1255					1260			
[4674]	Ile	Gly	Asn	Leu	Phe	Lys	Phe	Asp	Tyr	Glu	Asp	Lys	Asn	Lys	Ser
[4675]		1265					1270					1275			
[4676]	Ser	Ala	Asn	Pro	Lys	His	Thr								
[4677]		1280					1285								
[4678]	<210>	306													

[4679]	<211>	1120	
[4680]	<212>	PRT	
[4681]	<213>	斯氏李斯特菌	
[4682]	<400>	306	
[4683]	Met Trp Ile Ser Ile Lys Thr Leu Ile His His Leu Gly Val Leu Phe		
[4684]	1	5	10 15
[4685]	Phe Cys Asp Tyr Met Tyr Asn Arg Arg Glu Lys Lys Ile Ile Glu Val		
[4686]	20	25	30
[4687]	Lys Thr Met Arg Ile Thr Lys Val Glu Val Asp Arg Lys Lys Val Leu		
[4688]	35	40	45
[4689]	Ile Ser Arg Asp Lys Asn Gly Gly Lys Leu Val Tyr Glu Asn Glu Met		
[4690]	50	55	60
[4691]	Gln Asp Asn Thr Glu Gln Ile Met His His Lys Lys Ser Ser Phe Tyr		
[4692]	65	70	75 80
[4693]	Lys Ser Val Val Asn Lys Thr Ile Cys Arg Pro Glu Gln Lys Gln Met		
[4694]	85	90	95
[4695]	Lys Lys Leu Val His Gly Leu Leu Gln Glu Asn Ser Gln Glu Lys Ile		
[4696]	100	105	110
[4697]	Lys Val Ser Asp Val Thr Lys Leu Asn Ile Ser Asn Phe Leu Asn His		
[4698]	115	120	125
[4699]	Arg Phe Lys Lys Ser Leu Tyr Tyr Phe Pro Glu Asn Ser Pro Asp Lys		
[4700]	130	135	140
[4701]	Ser Glu Glu Tyr Arg Ile Glu Ile Asn Leu Ser Gln Leu Leu Glu Asp		
[4702]	145	150	155 160
[4703]	Ser Leu Lys Lys Gln Gln Gly Thr Phe Ile Cys Trp Glu Ser Phe Ser		
[4704]	165	170	175
[4705]	Lys Asp Met Glu Leu Tyr Ile Asn Trp Ala Glu Asn Tyr Ile Ser Ser		
[4706]	180	185	190
[4707]	Lys Thr Lys Leu Ile Lys Lys Ser Ile Arg Asn Asn Arg Ile Gln Ser		
[4708]	195	200	205
[4709]	Thr Glu Ser Arg Ser Gly Gln Leu Met Asp Arg Tyr Met Lys Asp Ile		
[4710]	210	215	220
[4711]	Leu Asn Lys Asn Lys Pro Phe Asp Ile Gln Ser Val Ser Glu Lys Tyr		
[4712]	225	230	235 240
[4713]	Gln Leu Glu Lys Leu Thr Ser Ala Leu Lys Ala Thr Phe Lys Glu Ala		
[4714]	245	250	255
[4715]	Lys Lys Asn Asp Lys Glu Ile Asn Tyr Lys Leu Lys Ser Thr Leu Gln		
[4716]	260	265	270
[4717]	Asn His Glu Arg Gln Ile Ile Glu Glu Leu Lys Glu Asn Ser Glu Leu		

[4718]	275	280	285
[4719]	Asn Gln Phe Asn Ile Glu Ile Arg Lys His Leu Glu Thr Tyr Phe Pro		
[4720]	290	295	300
[4721]	Ile Lys Lys Thr Asn Arg Lys Val Gly Asp Ile Arg Asn Leu Glu Ile		
[4722]	305	310	315
[4723]	Gly Glu Ile Gln Lys Ile Val Asn His Arg Leu Lys Asn Lys Ile Val		
[4724]	325	330	335
[4725]	Gln Arg Ile Leu Gln Glu Gly Lys Leu Ala Ser Tyr Glu Ile Glu Ser		
[4726]	340	345	350
[4727]	Thr Val Asn Ser Asn Ser Leu Gln Lys Ile Lys Ile Glu Glu Ala Phe		
[4728]	355	360	365
[4729]	Ala Leu Lys Phe Ile Asn Ala Cys Leu Phe Ala Ser Asn Asn Leu Arg		
[4730]	370	375	380
[4731]	Asn Met Val Tyr Pro Val Cys Lys Lys Asp Ile Leu Met Ile Gly Glu		
[4732]	385	390	395
[4733]	Phe Lys Asn Ser Phe Lys Glu Ile Lys His Lys Lys Phe Ile Arg Gln		
[4734]	405	410	415
[4735]	Trp Ser Gln Phe Phe Ser Gln Glu Ile Thr Val Asp Asp Ile Glu Leu		
[4736]	420	425	430
[4737]	Ala Ser Trp Gly Leu Arg Gly Ala Ile Ala Pro Ile Arg Asn Glu Ile		
[4738]	435	440	445
[4739]	Ile His Leu Lys Lys His Ser Trp Lys Lys Phe Phe Asn Asn Pro Thr		
[4740]	450	455	460
[4741]	Phe Lys Val Lys Lys Ser Lys Ile Ile Asn Gly Lys Thr Lys Asp Val		
[4742]	465	470	475
[4743]	Thr Ser Glu Phe Leu Tyr Lys Glu Thr Leu Phe Lys Asp Tyr Phe Tyr		
[4744]	485	490	495
[4745]	Ser Glu Leu Asp Ser Val Pro Glu Leu Ile Ile Asn Lys Met Glu Ser		
[4746]	500	505	510
[4747]	Ser Lys Ile Leu Asp Tyr Tyr Ser Ser Asp Gln Leu Asn Gln Val Phe		
[4748]	515	520	525
[4749]	Thr Ile Pro Asn Phe Glu Leu Ser Leu Leu Thr Ser Ala Val Pro Phe		
[4750]	530	535	540
[4751]	Ala Pro Ser Phe Lys Arg Val Tyr Leu Lys Gly Phe Asp Tyr Gln Asn		
[4752]	545	550	555
[4753]	Gln Asp Glu Ala Gln Pro Asp Tyr Asn Leu Lys Leu Asn Ile Tyr Asn		
[4754]	565	570	575
[4755]	Glu Lys Ala Phe Asn Ser Glu Ala Phe Gln Ala Gln Tyr Ser Leu Phe		
[4756]	580	585	590

[4757]	Lys Met Val Tyr Tyr Gln Val Phe Leu Pro Gln Phe Thr Thr Asn Asn
[4758]	595 600 605
[4759]	Asp Leu Phe Lys Ser Ser Val Asp Phe Ile Leu Thr Leu Asn Lys Glu
[4760]	610 615 620
[4761]	Arg Lys Gly Tyr Ala Lys Ala Phe Gln Asp Ile Arg Lys Met Asn Lys
[4762]	625 630 635 640
[4763]	Asp Glu Lys Pro Ser Glu Tyr Met Ser Tyr Ile Gln Ser Gln Leu Met
[4764]	645 650 655
[4765]	Leu Tyr Gln Lys Lys Gln Glu Glu Lys Glu Lys Ile Asn His Phe Glu
[4766]	660 665 670
[4767]	Lys Phe Ile Asn Gln Val Phe Ile Lys Gly Phe Asn Ser Phe Ile Glu
[4768]	675 680 685
[4769]	Lys Asn Arg Leu Thr Tyr Ile Cys His Pro Thr Lys Asn Thr Val Pro
[4770]	690 695 700
[4771]	Glu Asn Asp Asn Ile Glu Ile Pro Phe His Thr Asp Met Asp Asp Ser
[4772]	705 710 715 720
[4773]	Asn Ile Ala Phe Trp Leu Met Cys Lys Leu Leu Asp Ala Lys Gln Leu
[4774]	725 730 735
[4775]	Ser Glu Leu Arg Asn Glu Met Ile Lys Phe Ser Cys Ser Leu Gln Ser
[4776]	740 745 750
[4777]	Thr Glu Glu Ile Ser Thr Phe Thr Lys Ala Arg Glu Val Ile Gly Leu
[4778]	755 760 765
[4779]	Ala Leu Leu Asn Gly Glu Lys Gly Cys Asn Asp Trp Lys Glu Leu Phe
[4780]	770 775 780
[4781]	Asp Asp Lys Glu Ala Trp Lys Lys Asn Met Ser Leu Tyr Val Ser Glu
[4782]	785 790 795 800
[4783]	Glu Leu Leu Gln Ser Leu Pro Tyr Thr Gln Glu Asp Gly Gln Thr Pro
[4784]	805 810 815
[4785]	Val Ile Asn Arg Ser Ile Asp Leu Val Lys Lys Tyr Gly Thr Glu Thr
[4786]	820 825 830
[4787]	Ile Leu Glu Lys Leu Phe Ser Ser Ser Asp Asp Tyr Lys Val Ser Ala
[4788]	835 840 845
[4789]	Lys Asp Ile Ala Lys Leu His Glu Tyr Asp Val Thr Glu Lys Ile Ala
[4790]	850 855 860
[4791]	Gln Gln Glu Ser Leu His Lys Gln Trp Ile Glu Lys Pro Gly Leu Ala
[4792]	865 870 875 880
[4793]	Arg Asp Ser Ala Trp Thr Lys Lys Tyr Gln Asn Val Ile Asn Asp Ile
[4794]	885 890 895
[4795]	Ser Asn Tyr Gln Trp Ala Lys Thr Lys Val Glu Leu Thr Gln Val Arg

[4796]	900	905	910
[4797]	His Leu His Gln Leu Thr Ile Asp Leu Leu Ser Arg Leu Ala Gly Tyr		
[4798]	915	920	925
[4799]	Met Ser Ile Ala Asp Arg Asp Phe Gln Phe Ser Ser Asn Tyr Ile Leu		
[4800]	930	935	940
[4801]	Glu Arg Glu Asn Ser Glu Tyr Arg Val Thr Ser Trp Ile Leu Leu Ser		
[4802]	945	950	955
[4803]	Glu Asn Lys Asn Lys Asn Lys Tyr Asn Asp Tyr Glu Leu Tyr Asn Leu		
[4804]	965	970	975
[4805]	Lys Asn Ala Ser Ile Lys Val Ser Ser Lys Asn Asp Pro Gln Leu Lys		
[4806]	980	985	990
[4807]	Val Asp Leu Lys Gln Leu Arg Leu Thr Leu Glu Tyr Leu Glu Leu Phe		
[4808]	995	1000	1005
[4809]	Asp Asn Arg Leu Lys Glu Lys Arg Asn Asn Ile Ser His Phe Asn		
[4810]	1010	1015	1020
[4811]	Tyr Leu Asn Gly Gln Leu Gly Asn Ser Ile Leu Glu Leu Phe Asp		
[4812]	1025	1030	1035
[4813]	Asp Ala Arg Asp Val Leu Ser Tyr Asp Arg Lys Leu Lys Asn Ala		
[4814]	1040	1045	1050
[4815]	Val Ser Lys Ser Leu Lys Glu Ile Leu Ser Ser His Gly Met Glu		
[4816]	1055	1060	1065
[4817]	Val Thr Phe Lys Pro Leu Tyr Gln Thr Asn His His Leu Lys Ile		
[4818]	1070	1075	1080
[4819]	Asp Lys Leu Gln Pro Lys Lys Ile His His Leu Gly Glu Lys Ser		
[4820]	1085	1090	1095
[4821]	Thr Val Ser Ser Asn Gln Val Ser Asn Glu Tyr Cys Gln Leu Val		
[4822]	1100	1105	1110
[4823]	Arg Thr Leu Leu Thr Met Lys		
[4824]	1115	1120	
[4825]	<210> 307		
[4826]	<211> 1154		
[4827]	<212> PRT		
[4828]	<213> Paludibacter propionicigenes		
[4829]	<400> 307		
[4830]	Met Arg Val Ser Lys Val Lys Val Lys Asp Gly Gly Lys Asp Lys Met		
[4831]	1	5	10
[4832]	Val Leu Val His Arg Lys Thr Thr Gly Ala Gln Leu Val Tyr Ser Gly		
[4833]	20	25	30
[4834]	Gln Pro Val Ser Asn Glu Thr Ser Asn Ile Leu Pro Glu Lys Lys Arg		

[4835]	35	40	45
[4836]	Gln Ser Phe Asp Leu Ser Thr Leu Asn Lys Thr Ile Ile Lys Phe Asp		
[4837]	50	55	60
[4838]	Thr Ala Lys Lys Gln Lys Leu Asn Val Asp Gln Tyr Lys Ile Val Glu		
[4839]	65	70	75
[4840]	Lys Ile Phe Lys Tyr Pro Lys Gln Glu Leu Pro Lys Gln Ile Lys Ala		
[4841]	85	90	95
[4842]	Glu Glu Ile Leu Pro Phe Leu Asn His Lys Phe Gln Glu Pro Val Lys		
[4843]	100	105	110
[4844]	Tyr Trp Lys Asn Gly Lys Glu Glu Ser Phe Asn Leu Thr Leu Leu Ile		
[4845]	115	120	125
[4846]	Val Glu Ala Val Gln Ala Gln Asp Lys Arg Lys Leu Gln Pro Tyr Tyr		
[4847]	130	135	140
[4848]	Asp Trp Lys Thr Trp Tyr Ile Gln Thr Lys Ser Asp Leu Leu Lys Lys		
[4849]	145	150	155
[4850]	Ser Ile Glu Asn Asn Arg Ile Asp Leu Thr Glu Asn Leu Ser Lys Arg		
[4851]	165	170	175
[4852]	Lys Lys Ala Leu Leu Ala Trp Glu Thr Glu Phe Thr Ala Ser Gly Ser		
[4853]	180	185	190
[4854]	Ile Asp Leu Thr His Tyr His Lys Val Tyr Met Thr Asp Val Leu Cys		
[4855]	195	200	205
[4856]	Lys Met Leu Gln Asp Val Lys Pro Leu Thr Asp Asp Lys Gly Lys Ile		
[4857]	210	215	220
[4858]	Asn Thr Asn Ala Tyr His Arg Gly Leu Lys Lys Ala Leu Gln Asn His		
[4859]	225	230	235
[4860]	Gln Pro Ala Ile Phe Gly Thr Arg Glu Val Pro Asn Glu Ala Asn Arg		
[4861]	245	250	255
[4862]	Ala Asp Asn Gln Leu Ser Ile Tyr His Leu Glu Val Val Lys Tyr Leu		
[4863]	260	265	270
[4864]	Glu His Tyr Phe Pro Ile Lys Thr Ser Lys Arg Arg Asn Thr Ala Asp		
[4865]	275	280	285
[4866]	Asp Ile Ala His Tyr Leu Lys Ala Gln Thr Leu Lys Thr Thr Ile Glu		
[4867]	290	295	300
[4868]	Lys Gln Leu Val Asn Ala Ile Arg Ala Asn Ile Ile Gln Gln Gly Lys		
[4869]	305	310	315
[4870]	Thr Asn His His Glu Leu Lys Ala Asp Thr Thr Ser Asn Asp Leu Ile		
[4871]	325	330	335
[4872]	Arg Ile Lys Thr Asn Glu Ala Phe Val Leu Asn Leu Thr Gly Thr Cys		
[4873]	340	345	350

[4874]	Ala	Phe	Ala	Ala	Asn	Asn	Ile	Arg	Asn	Met	Val	Asp	Asn	Glu	Gln	Thr
[4875]				355				360					365			
[4876]	Asn	Asp	Ile	Leu	Gly	Lys	Gly	Asp	Phe	Ile	Lys	Ser	Leu	Leu	Lys	Asp
[4877]			370				375					380				
[4878]	Asn	Thr	Asn	Ser	Gln	Leu	Tyr	Ser	Phe	Phe	Phe	Gly	Glu	Gly	Leu	Ser
[4879]	385					390					395				400	
[4880]	Thr	Asn	Lys	Ala	Glu	Lys	Glu	Thr	Gln	Leu	Trp	Gly	Ile	Arg	Gly	Ala
[4881]				405					410					415		
[4882]	Val	Gln	Gln	Ile	Arg	Asn	Asn	Val	Asn	His	Tyr	Lys	Lys	Asp	Ala	Leu
[4883]				420					425					430		
[4884]	Lys	Thr	Val	Phe	Asn	Ile	Ser	Asn	Phe	Glu	Asn	Pro	Thr	Ile	Thr	Asp
[4885]			435					440					445			
[4886]	Pro	Lys	Gln	Gln	Thr	Asn	Tyr	Ala	Asp	Thr	Ile	Tyr	Lys	Ala	Arg	Phe
[4887]		450					455					460				
[4888]	Ile	Asn	Glu	Leu	Glu	Lys	Ile	Pro	Glu	Ala	Phe	Ala	Gln	Gln	Leu	Lys
[4889]	465					470					475				480	
[4890]	Thr	Gly	Gly	Ala	Val	Ser	Tyr	Tyr	Thr	Ile	Glu	Asn	Leu	Lys	Ser	Leu
[4891]				485						490					495	
[4892]	Leu	Thr	Thr	Phe	Gln	Phe	Ser	Leu	Cys	Arg	Ser	Thr	Ile	Pro	Phe	Ala
[4893]				500					505					510		
[4894]	Pro	Gly	Phe	Lys	Lys	Val	Phe	Asn	Gly	Gly	Ile	Asn	Tyr	Gln	Asn	Ala
[4895]			515					520					525			
[4896]	Lys	Gln	Asp	Glu	Ser	Phe	Tyr	Glu	Leu	Met	Leu	Glu	Gln	Tyr	Leu	Arg
[4897]		530					535					540				
[4898]	Lys	Glu	Asn	Phe	Ala	Glu	Glu	Ser	Tyr	Asn	Ala	Arg	Tyr	Phe	Met	Leu
[4899]	545					550					555				560	
[4900]	Lys	Leu	Ile	Tyr	Asn	Asn	Leu	Phe	Leu	Pro	Gly	Phe	Thr	Thr	Asp	Arg
[4901]				565					570					575		
[4902]	Lys	Ala	Phe	Ala	Asp	Ser	Val	Gly	Phe	Val	Gln	Met	Gln	Asn	Lys	Lys
[4903]				580					585					590		
[4904]	Gln	Ala	Glu	Lys	Val	Asn	Pro	Arg	Lys	Lys	Glu	Ala	Tyr	Ala	Phe	Glu
[4905]			595					600					605			
[4906]	Ala	Val	Arg	Pro	Met	Thr	Ala	Ala	Asp	Ser	Ile	Ala	Asp	Tyr	Met	Ala
[4907]		610						615					620			
[4908]	Tyr	Val	Gln	Ser	Glu	Leu	Met	Gln	Glu	Gln	Asn	Lys	Lys	Glu	Glu	Lys
[4909]	625					630					635				640	
[4910]	Val	Ala	Glu	Glu	Thr	Arg	Ile	Asn	Phe	Glu	Lys	Phe	Val	Leu	Gln	Val
[4911]				645					650					655		
[4912]	Phe	Ile	Lys	Gly	Phe	Asp	Ser	Phe	Leu	Arg	Ala	Lys	Glu	Phe	Asp	Phe

[4913]	660							665							670													
[4914]	Val	Gln	Met	Pro	Gln	Pro	Gln	Leu	Thr	Ala	Thr	Ala	Ser	Asn	Gln	Gln												
[4915]	675							680							685													
[4916]	Lys	Ala	Asp	Lys	Leu	Asn	Gln	Leu	Glu	Ala	Ser	Ile	Thr	Ala	Asp	Cys												
[4917]	690							695							700													
[4918]	Lys	Leu	Thr	Pro	Gln	Tyr	Ala	Lys	Ala	Asp	Asp	Ala	Thr	His	Ile	Ala												
[4919]	705							710							715							720						
[4920]	Phe	Tyr	Val	Phe	Cys	Lys	Leu	Leu	Asp	Ala	Ala	His	Leu	Ser	Asn	Leu												
[4921]	725							730							735													
[4922]	Arg	Asn	Glu	Leu	Ile	Lys	Phe	Arg	Glu	Ser	Val	Asn	Glu	Phe	Lys	Phe												
[4923]	740							745							750													
[4924]	His	His	Leu	Leu	Glu	Ile	Ile	Glu	Ile	Cys	Leu	Leu	Ser	Ala	Asp	Val												
[4925]	755							760							765													
[4926]	Val	Pro	Thr	Asp	Tyr	Arg	Asp	Leu	Tyr	Ser	Ser	Glu	Ala	Asp	Cys	Leu												
[4927]	770							775							780													
[4928]	Ala	Arg	Leu	Arg	Pro	Phe	Ile	Glu	Gln	Gly	Ala	Asp	Ile	Thr	Asn	Trp												
[4929]	785							790							795							800						
[4930]	Ser	Asp	Leu	Phe	Val	Gln	Ser	Asp	Lys	His	Ser	Pro	Val	Ile	His	Ala												
[4931]	805							810							815													
[4932]	Asn	Ile	Glu	Leu	Ser	Val	Lys	Tyr	Gly	Thr	Thr	Lys	Leu	Leu	Glu	Gln												
[4933]	820							825							830													
[4934]	Ile	Ile	Asn	Lys	Asp	Thr	Gln	Phe	Lys	Thr	Thr	Glu	Ala	Asn	Phe	Thr												
[4935]	835							840							845													
[4936]	Ala	Trp	Asn	Thr	Ala	Gln	Lys	Ser	Ile	Glu	Gln	Leu	Ile	Lys	Gln	Arg												
[4937]	850							855							860													
[4938]	Glu	Asp	His	His	Glu	Gln	Trp	Val	Lys	Ala	Lys	Asn	Ala	Asp	Asp	Lys												
[4939]	865							870							875							880						
[4940]	Glu	Lys	Gln	Glu	Arg	Lys	Arg	Glu	Lys	Ser	Asn	Phe	Ala	Gln	Lys	Phe												
[4941]	885							890							895													
[4942]	Ile	Glu	Lys	His	Gly	Asp	Asp	Tyr	Leu	Asp	Ile	Cys	Asp	Tyr	Ile	Asn												
[4943]	900							905							910													
[4944]	Thr	Tyr	Asn	Trp	Leu	Asp	Asn	Lys	Met	His	Phe	Val	His	Leu	Asn	Arg												
[4945]	915							920							925													
[4946]	Leu	His	Gly	Leu	Thr	Ile	Glu	Leu	Leu	Gly	Arg	Met	Ala	Gly	Phe	Val												
[4947]	930							935							940													
[4948]	Ala	Leu	Phe	Asp	Arg	Asp	Phe	Gln	Phe	Phe	Asp	Glu	Gln	Gln	Ile	Ala												
[4949]	945							950							955							960						
[4950]	Asp	Glu	Phe	Lys	Leu	His	Gly	Phe	Val	Asn	Leu	His	Ser	Ile	Asp	Lys												
[4951]	965							970							975													

[4952]	Lys	Leu	Asn	Glu	Val	Pro	Thr	Lys	Lys	Ile	Lys	Glu	Ile	Tyr	Asp	Ile
[4953]				980					985					990		
[4954]	Arg	Asn	Lys	Ile	Ile	Gln	Ile	Asn	Gly	Asn	Lys	Ile	Asn	Glu	Ser	Val
[4955]				995					1000					1005		
[4956]	Arg	Ala	Asn	Leu	Ile	Gln	Phe	Ile	Ser	Ser	Lys	Arg	Asn	Tyr	Tyr	
[4957]		1010						1015						1020		
[4958]	Asn	Asn	Ala	Phe	Leu	His	Val	Ser	Asn	Asp	Glu	Ile	Lys	Glu	Lys	
[4959]		1025						1030						1035		
[4960]	Gln	Met	Tyr	Asp	Ile	Arg	Asn	His	Ile	Ala	His	Phe	Asn	Tyr	Leu	
[4961]		1040						1045						1050		
[4962]	Thr	Lys	Asp	Ala	Ala	Asp	Phe	Ser	Leu	Ile	Asp	Leu	Ile	Asn	Glu	
[4963]		1055						1060						1065		
[4964]	Leu	Arg	Glu	Leu	Leu	His	Tyr	Asp	Arg	Lys	Leu	Lys	Asn	Ala	Val	
[4965]		1070						1075						1080		
[4966]	Ser	Lys	Ala	Phe	Ile	Asp	Leu	Phe	Asp	Lys	His	Gly	Met	Ile	Leu	
[4967]		1085						1090						1095		
[4968]	Lys	Leu	Lys	Leu	Asn	Ala	Asp	His	Lys	Leu	Lys	Val	Glu	Ser	Leu	
[4969]		1100						1105						1110		
[4970]	Glu	Pro	Lys	Lys	Ile	Tyr	His	Leu	Gly	Ser	Ser	Ala	Lys	Asp	Lys	
[4971]		1115						1120						1125		
[4972]	Pro	Glu	Tyr	Gln	Tyr	Cys	Thr	Asn	Gln	Val	Met	Met	Ala	Tyr	Cys	
[4973]		1130						1135						1140		
[4974]	Asn	Met	Cys	Arg	Ser	Leu	Leu	Glu	Met	Lys	Lys					
[4975]		1145						1150								
[4976]	<210>	308														
[4977]	<211>	1159														
[4978]	<212>	PRT														
[4979]	<213>	口腔纤毛菌														
[4980]	<400>	308														
[4981]	Met	Lys	Val	Thr	Lys	Val	Gly	Gly	Ile	Ser	His	Lys	Lys	Tyr	Thr	Ser
[4982]	1				5					10					15	
[4983]	Glu	Gly	Arg	Leu	Val	Lys	Ser	Glu	Ser	Glu	Glu	Asn	Arg	Thr	Asp	Glu
[4984]				20						25					30	
[4985]	Arg	Leu	Ser	Ala	Leu	Leu	Asn	Met	Arg	Leu	Asp	Met	Tyr	Ile	Lys	Asn
[4986]			35						40					45		
[4987]	Pro	Ser	Ser	Thr	Glu	Thr	Lys	Glu	Asn	Gln	Lys	Arg	Ile	Gly	Lys	Leu
[4988]		50							55					60		
[4989]	Lys	Lys	Phe	Phe	Ser	Asn	Lys	Met	Val	Tyr	Leu	Lys	Asp	Asn	Thr	Leu
[4990]	65						70					75				80

[4991]	Ser	Leu	Lys	Asn	Gly	Lys	Lys	Glu	Asn	Ile	Asp	Arg	Glu	Tyr	Ser	Glu
[4992]					85				90						95	
[4993]	Thr	Asp	Ile	Leu	Glu	Ser	Asp	Val	Arg	Asp	Lys	Lys	Asn	Phe	Ala	Val
[4994]					100				105					110		
[4995]	Leu	Lys	Lys	Ile	Tyr	Leu	Asn	Glu	Asn	Val	Asn	Ser	Glu	Glu	Leu	Glu
[4996]					115				120					125		
[4997]	Val	Phe	Arg	Asn	Asp	Ile	Lys	Lys	Lys	Leu	Asn	Lys	Ile	Asn	Ser	Leu
[4998]					130				135					140		
[4999]	Lys	Tyr	Ser	Phe	Glu	Lys	Asn	Lys	Ala	Asn	Tyr	Gln	Lys	Ile	Asn	Glu
[5000]	145						150					155				160
[5001]	Asn	Asn	Ile	Glu	Lys	Val	Glu	Gly	Lys	Ser	Lys	Arg	Asn	Ile	Ile	Tyr
[5002]					165					170						175
[5003]	Asp	Tyr	Tyr	Arg	Glu	Ser	Ala	Lys	Arg	Asp	Ala	Tyr	Val	Ser	Asn	Val
[5004]					180					185						190
[5005]	Lys	Glu	Ala	Phe	Asp	Lys	Leu	Tyr	Lys	Glu	Glu	Asp	Ile	Ala	Lys	Leu
[5006]					195					200						205
[5007]	Val	Leu	Glu	Ile	Glu	Asn	Leu	Thr	Lys	Leu	Glu	Lys	Tyr	Lys	Ile	Arg
[5008]					210					215						220
[5009]	Glu	Phe	Tyr	His	Glu	Ile	Ile	Gly	Arg	Lys	Asn	Asp	Lys	Glu	Asn	Phe
[5010]	225						230					235				240
[5011]	Ala	Lys	Ile	Ile	Tyr	Glu	Glu	Ile	Gln	Asn	Val	Asn	Asn	Met	Lys	Glu
[5012]					245						250					255
[5013]	Leu	Ile	Glu	Lys	Val	Pro	Asp	Met	Ser	Glu	Leu	Lys	Lys	Ser	Gln	Val
[5014]					260					265						270
[5015]	Phe	Tyr	Lys	Tyr	Tyr	Leu	Asp	Lys	Glu	Glu	Leu	Asn	Asp	Lys	Asn	Ile
[5016]					275					280						285
[5017]	Lys	Tyr	Ala	Phe	Cys	His	Phe	Val	Glu	Ile	Glu	Met	Ser	Gln	Leu	Leu
[5018]					290					295						300
[5019]	Lys	Asn	Tyr	Val	Tyr	Lys	Arg	Leu	Ser	Asn	Ile	Ser	Asn	Asp	Lys	Ile
[5020]	305						310					315				320
[5021]	Lys	Arg	Ile	Phe	Glu	Tyr	Gln	Asn	Leu	Lys	Lys	Leu	Ile	Glu	Asn	Lys
[5022]					325						330					335
[5023]	Leu	Leu	Asn	Lys	Leu	Asp	Thr	Tyr	Val	Arg	Asn	Cys	Gly	Lys	Tyr	Asn
[5024]					340					345						350
[5025]	Tyr	Tyr	Leu	Gln	Asp	Gly	Glu	Ile	Ala	Thr	Ser	Asp	Phe	Ile	Ala	Arg
[5026]					355					360						365
[5027]	Asn	Arg	Gln	Asn	Glu	Ala	Phe	Leu	Arg	Asn	Ile	Ile	Gly	Val	Ser	Ser
[5028]					370						375					380
[5029]	Val	Ala	Tyr	Phe	Ser	Leu	Arg	Asn	Ile	Leu	Glu	Thr	Glu	Asn	Glu	Asn

[5030]	385	390	395	400
[5031]	Asp Ile Thr Gly Arg Met Arg Gly Lys Thr Val Lys Asn Asn Lys Gly			
[5032]		405	410	415
[5033]	Glu Glu Lys Tyr Val Ser Gly Glu Val Asp Lys Ile Tyr Asn Glu Asn			
[5034]		420	425	430
[5035]	Lys Lys Asn Glu Val Lys Glu Asn Leu Lys Met Phe Tyr Ser Tyr Asp			
[5036]		435	440	445
[5037]	Phe Asn Met Asp Asn Lys Asn Glu Ile Glu Asp Phe Phe Ala Asn Ile			
[5038]		450	455	460
[5039]	Asp Glu Ala Ile Ser Ser Ile Arg His Gly Ile Val His Phe Asn Leu			
[5040]	465	470	475	480
[5041]	Glu Leu Glu Gly Lys Asp Ile Phe Ala Phe Lys Asn Ile Ala Pro Ser			
[5042]		485	490	495
[5043]	Glu Ile Ser Lys Lys Met Phe Gln Asn Glu Ile Asn Glu Lys Lys Leu			
[5044]		500	505	510
[5045]	Lys Leu Lys Ile Phe Arg Gln Leu Asn Ser Ala Asn Val Phe Arg Tyr			
[5046]		515	520	525
[5047]	Leu Glu Lys Tyr Lys Ile Leu Asn Tyr Leu Lys Arg Thr Arg Phe Glu			
[5048]		530	535	540
[5049]	Phe Val Asn Lys Asn Ile Pro Phe Val Pro Ser Phe Thr Lys Leu Tyr			
[5050]	545	550	555	560
[5051]	Ser Arg Ile Asp Asp Leu Lys Asn Ser Leu Gly Ile Tyr Trp Lys Thr			
[5052]		565	570	575
[5053]	Pro Lys Thr Asn Asp Asp Asn Lys Thr Lys Glu Ile Ile Asp Ala Gln			
[5054]		580	585	590
[5055]	Ile Tyr Leu Leu Lys Asn Ile Tyr Tyr Gly Glu Phe Leu Asn Tyr Phe			
[5056]		595	600	605
[5057]	Met Ser Asn Asn Gly Asn Phe Phe Glu Ile Ser Lys Glu Ile Ile Glu			
[5058]		610	615	620
[5059]	Leu Asn Lys Asn Asp Lys Arg Asn Leu Lys Thr Gly Phe Tyr Lys Leu			
[5060]	625	630	635	640
[5061]	Gln Lys Phe Glu Asp Ile Gln Glu Lys Ile Pro Lys Glu Tyr Leu Ala			
[5062]		645	650	655
[5063]	Asn Ile Gln Ser Leu Tyr Met Ile Asn Ala Gly Asn Gln Asp Glu Glu			
[5064]		660	665	670
[5065]	Glu Lys Asp Thr Tyr Ile Asp Phe Ile Gln Lys Ile Phe Leu Lys Gly			
[5066]		675	680	685
[5067]	Phe Met Thr Tyr Leu Ala Asn Asn Gly Arg Leu Ser Leu Ile Tyr Ile			
[5068]		690	695	700

[5069]	Gly Ser Asp Glu Glu Thr Asn Thr Ser Leu Ala Glu Lys Lys Gln Glu
[5070]	705 710 715 720
[5071]	Phe Asp Lys Phe Leu Lys Lys Tyr Glu Gln Asn Asn Asn Ile Lys Ile
[5072]	725 730 735
[5073]	Pro Tyr Glu Ile Asn Glu Phe Leu Arg Glu Ile Lys Leu Gly Asn Ile
[5074]	740 745 750
[5075]	Leu Lys Tyr Thr Glu Arg Leu Asn Met Phe Tyr Leu Ile Leu Lys Leu
[5076]	755 760 765
[5077]	Leu Asn His Lys Glu Leu Thr Asn Leu Lys Gly Ser Leu Glu Lys Tyr
[5078]	770 775 780
[5079]	Gln Ser Ala Asn Lys Glu Glu Ala Phe Ser Asp Gln Leu Glu Leu Ile
[5080]	785 790 795 800
[5081]	Asn Leu Leu Asn Leu Asp Asn Asn Arg Val Thr Glu Asp Phe Glu Leu
[5082]	805 810 815
[5083]	Glu Ala Asp Glu Ile Gly Lys Phe Leu Asp Phe Asn Gly Asn Lys Val
[5084]	820 825 830
[5085]	Lys Asp Asn Lys Glu Leu Lys Lys Phe Asp Thr Asn Lys Ile Tyr Phe
[5086]	835 840 845
[5087]	Asp Gly Glu Asn Ile Ile Lys His Arg Ala Phe Tyr Asn Ile Lys Lys
[5088]	850 855 860
[5089]	Tyr Gly Met Leu Asn Leu Leu Glu Lys Ile Ala Asp Lys Ala Gly Tyr
[5090]	865 870 875 880
[5091]	Lys Ile Ser Ile Glu Glu Leu Lys Lys Tyr Ser Asn Lys Lys Asn Glu
[5092]	885 890 895
[5093]	Ile Glu Lys Asn His Lys Met Gln Glu Asn Leu His Arg Lys Tyr Ala
[5094]	900 905 910
[5095]	Arg Pro Arg Lys Asp Glu Lys Phe Thr Asp Glu Asp Tyr Glu Ser Tyr
[5096]	915 920 925
[5097]	Lys Gln Ala Ile Glu Asn Ile Glu Glu Tyr Thr His Leu Lys Asn Lys
[5098]	930 935 940
[5099]	Val Glu Phe Asn Glu Leu Asn Leu Leu Gln Gly Leu Leu Leu Arg Ile
[5100]	945 950 955 960
[5101]	Leu His Arg Leu Val Gly Tyr Thr Ser Ile Trp Glu Arg Asp Leu Arg
[5102]	965 970 975
[5103]	Phe Arg Leu Lys Gly Glu Phe Pro Glu Asn Gln Tyr Ile Glu Glu Ile
[5104]	980 985 990
[5105]	Phe Asn Phe Glu Asn Lys Lys Asn Val Lys Tyr Lys Gly Gly Gln Ile
[5106]	995 1000 1005
[5107]	Val Glu Lys Tyr Ile Lys Phe Tyr Lys Glu Leu His Gln Asn Asp

[5108]	1010	1015	1020
[5109]	Glu Val Lys Ile Asn Lys Tyr Ser Ser Ala Asn Ile Lys Val Leu		
[5110]	1025	1030	1035
[5111]	Lys Gln Glu Lys Lys Asp Leu Tyr Ile Arg Asn Tyr Ile Ala His		
[5112]	1040	1045	1050
[5113]	Phe Asn Tyr Ile Pro His Ala Glu Ile Ser Leu Leu Glu Val Leu		
[5114]	1055	1060	1065
[5115]	Glu Asn Leu Arg Lys Leu Leu Ser Tyr Asp Arg Lys Leu Lys Asn		
[5116]	1070	1075	1080
[5117]	Ala Val Met Lys Ser Val Val Asp Ile Leu Lys Glu Tyr Gly Phe		
[5118]	1085	1090	1095
[5119]	Val Ala Thr Phe Lys Ile Gly Ala Asp Lys Lys Ile Gly Ile Gln		
[5120]	1100	1105	1110
[5121]	Thr Leu Glu Ser Glu Lys Ile Val His Leu Lys Asn Leu Lys Lys		
[5122]	1115	1120	1125
[5123]	Lys Lys Leu Met Thr Asp Arg Asn Ser Glu Glu Leu Cys Lys Leu		
[5124]	1130	1135	1140
[5125]	Val Lys Ile Met Phe Glu Tyr Lys Met Glu Glu Lys Lys Ser Glu		
[5126]	1145	1150	1155
[5127]	Asn		
[5128]	<210> 309		
[5129]	<211> 1389		
[5130]	<212> PRT		
[5131]	<213> 沙氏纤毛菌		
[5132]	<400> 309		
[5133]	Met Gly Asn Leu Phe Gly His Lys Arg Trp Tyr Glu Val Arg Asp Lys		
[5134]	1 5 10 15		
[5135]	Lys Asp Phe Lys Ile Lys Arg Lys Val Lys Val Lys Arg Asn Tyr Asp		
[5136]	20 25 30		
[5137]	Gly Asn Lys Tyr Ile Leu Asn Ile Asn Glu Asn Asn Asn Lys Glu Lys		
[5138]	35 40 45		
[5139]	Ile Asp Asn Asn Lys Phe Ile Arg Lys Tyr Ile Asn Tyr Lys Lys Asn		
[5140]	50 55 60		
[5141]	Asp Asn Ile Leu Lys Glu Phe Thr Arg Lys Phe His Ala Gly Asn Ile		
[5142]	65 70 75 80		
[5143]	Leu Phe Lys Leu Lys Gly Lys Glu Gly Ile Ile Arg Ile Glu Asn Asn		
[5144]	85 90 95		
[5145]	Asp Asp Phe Leu Glu Thr Glu Glu Val Val Leu Tyr Ile Glu Ala Tyr		
[5146]	100 105 110		

[5147]	Gly	Lys	Ser	Glu	Lys	Leu	Lys	Ala	Leu	Gly	Ile	Thr	Lys	Lys	Lys	Ile
[5148]			115					120					125			
[5149]	Ile	Asp	Glu	Ala	Ile	Arg	Gln	Gly	Ile	Thr	Lys	Asp	Asp	Lys	Lys	Ile
[5150]			130					135					140			
[5151]	Glu	Ile	Lys	Arg	Gln	Glu	Asn	Glu	Glu	Glu	Ile	Glu	Ile	Asp	Ile	Arg
[5152]	145						150					155				160
[5153]	Asp	Glu	Tyr	Thr	Asn	Lys	Thr	Leu	Asn	Asp	Cys	Ser	Ile	Ile	Leu	Arg
[5154]					165					170						175
[5155]	Ile	Ile	Glu	Asn	Asp	Glu	Leu	Glu	Thr	Lys	Lys	Ser	Ile	Tyr	Glu	Ile
[5156]				180						185					190	
[5157]	Phe	Lys	Asn	Ile	Asn	Met	Ser	Leu	Tyr	Lys	Ile	Ile	Glu	Lys	Ile	Ile
[5158]			195					200					205			
[5159]	Glu	Asn	Glu	Thr	Glu	Lys	Val	Phe	Glu	Asn	Arg	Tyr	Tyr	Glu	Glu	His
[5160]		210						215					220			
[5161]	Leu	Arg	Glu	Lys	Leu	Leu	Lys	Asp	Asp	Lys	Ile	Asp	Val	Ile	Leu	Thr
[5162]	225						230					235				240
[5163]	Asn	Phe	Met	Glu	Ile	Arg	Glu	Lys	Ile	Lys	Ser	Asn	Leu	Glu	Ile	Leu
[5164]					245					250						255
[5165]	Gly	Phe	Val	Lys	Phe	Tyr	Leu	Asn	Val	Gly	Gly	Asp	Lys	Lys	Lys	Ser
[5166]				260						265					270	
[5167]	Lys	Asn	Lys	Lys	Met	Leu	Val	Glu	Lys	Ile	Leu	Asn	Ile	Asn	Val	Asp
[5168]			275						280					285		
[5169]	Leu	Thr	Val	Glu	Asp	Ile	Ala	Asp	Phe	Val	Ile	Lys	Glu	Leu	Glu	Phe
[5170]		290						295					300			
[5171]	Trp	Asn	Ile	Thr	Lys	Arg	Ile	Glu	Lys	Val	Lys	Lys	Val	Asn	Asn	Glu
[5172]	305					310						315				320
[5173]	Phe	Leu	Glu	Lys	Arg	Arg	Asn	Arg	Thr	Tyr	Ile	Lys	Ser	Tyr	Val	Leu
[5174]					325					330						335
[5175]	Leu	Asp	Lys	His	Glu	Lys	Phe	Lys	Ile	Glu	Arg	Glu	Asn	Lys	Lys	Asp
[5176]				340					345					350		
[5177]	Lys	Ile	Val	Lys	Phe	Phe	Val	Glu	Asn	Ile	Lys	Asn	Asn	Ser	Ile	Lys
[5178]			355						360					365		
[5179]	Glu	Lys	Ile	Glu	Lys	Ile	Leu	Ala	Glu	Phe	Lys	Ile	Asp	Glu	Leu	Ile
[5180]		370						375					380			
[5181]	Lys	Lys	Leu	Glu	Lys	Glu	Leu	Lys	Lys	Gly	Asn	Cys	Asp	Thr	Glu	Ile
[5182]	385						390					395				400
[5183]	Phe	Gly	Ile	Phe	Lys	Lys	His	Tyr	Lys	Val	Asn	Phe	Asp	Ser	Lys	Lys
[5184]					405					410					415	
[5185]	Phe	Ser	Lys	Lys	Ser	Asp	Glu	Glu	Lys	Glu	Leu	Tyr	Lys	Ile	Ile	Tyr

[5186]	420	425	430
[5187]	Arg Tyr Leu Lys Gly Arg Ile Glu Lys Ile Leu Val Asn Glu Gln Lys		
[5188]	435	440	445
[5189]	Val Arg Leu Lys Lys Met Glu Lys Ile Glu Ile Glu Lys Ile Leu Asn		
[5190]	450	455	460
[5191]	Glu Ser Ile Leu Ser Glu Lys Ile Leu Lys Arg Val Lys Gln Tyr Thr		
[5192]	465	470	475
[5193]	Leu Glu His Ile Met Tyr Leu Gly Lys Leu Arg His Asn Asp Ile Asp		
[5194]	485	490	495
[5195]	Met Thr Thr Val Asn Thr Asp Asp Phe Ser Arg Leu His Ala Lys Glu		
[5196]	500	505	510
[5197]	Glu Leu Asp Leu Glu Leu Ile Thr Phe Phe Ala Ser Thr Asn Met Glu		
[5198]	515	520	525
[5199]	Leu Asn Lys Ile Phe Ser Arg Glu Asn Ile Asn Asn Asp Glu Asn Ile		
[5200]	530	535	540
[5201]	Asp Phe Phe Gly Gly Asp Arg Glu Lys Asn Tyr Val Leu Asp Lys Lys		
[5202]	545	550	555
[5203]	Ile Leu Asn Ser Lys Ile Lys Ile Ile Arg Asp Leu Asp Phe Ile Asp		
[5204]	565	570	575
[5205]	Asn Lys Asn Asn Ile Thr Asn Asn Phe Ile Arg Lys Phe Thr Lys Ile		
[5206]	580	585	590
[5207]	Gly Thr Asn Glu Arg Asn Arg Ile Leu His Ala Ile Ser Lys Glu Arg		
[5208]	595	600	605
[5209]	Asp Leu Gln Gly Thr Gln Asp Asp Tyr Asn Lys Val Ile Asn Ile Ile		
[5210]	610	615	620
[5211]	Gln Asn Leu Lys Ile Ser Asp Glu Glu Val Ser Lys Ala Leu Asn Leu		
[5212]	625	630	635
[5213]	Asp Val Val Phe Lys Asp Lys Lys Asn Ile Ile Thr Lys Ile Asn Asp		
[5214]	645	650	655
[5215]	Ile Lys Ile Ser Glu Glu Asn Asn Asn Asp Ile Lys Tyr Leu Pro Ser		
[5216]	660	665	670
[5217]	Phe Ser Lys Val Leu Pro Glu Ile Leu Asn Leu Tyr Arg Asn Asn Pro		
[5218]	675	680	685
[5219]	Lys Asn Glu Pro Phe Asp Thr Ile Glu Thr Glu Lys Ile Val Leu Asn		
[5220]	690	695	700
[5221]	Ala Leu Ile Tyr Val Asn Lys Glu Leu Tyr Lys Lys Leu Ile Leu Glu		
[5222]	705	710	715
[5223]	Asp Asp Leu Glu Glu Asn Glu Ser Lys Asn Ile Phe Leu Gln Glu Leu		
[5224]	725	730	735

[5225]	Lys Lys Thr Leu Gly Asn Ile Asp Glu Ile Asp Glu Asn Ile Ile Glu
[5226]	740 745 750
[5227]	Asn Tyr Tyr Lys Asn Ala Gln Ile Ser Ala Ser Lys Gly Asn Asn Lys
[5228]	755 760 765
[5229]	Ala Ile Lys Lys Tyr Gln Lys Lys Val Ile Glu Cys Tyr Ile Gly Tyr
[5230]	770 775 780
[5231]	Leu Arg Lys Asn Tyr Glu Glu Leu Phe Asp Phe Ser Asp Phe Lys Met
[5232]	785 790 795 800
[5233]	Asn Ile Gln Glu Ile Lys Lys Gln Ile Lys Asp Ile Asn Asp Asn Lys
[5234]	805 810 815
[5235]	Thr Tyr Glu Arg Ile Thr Val Lys Thr Ser Asp Lys Thr Ile Val Ile
[5236]	820 825 830
[5237]	Asn Asp Asp Phe Glu Tyr Ile Ile Ser Ile Phe Ala Leu Leu Asn Ser
[5238]	835 840 845
[5239]	Asn Ala Val Ile Asn Lys Ile Arg Asn Arg Phe Phe Ala Thr Ser Val
[5240]	850 855 860
[5241]	Trp Leu Asn Thr Ser Glu Tyr Gln Asn Ile Ile Asp Ile Leu Asp Glu
[5242]	865 870 875 880
[5243]	Ile Met Gln Leu Asn Thr Leu Arg Asn Glu Cys Ile Thr Glu Asn Trp
[5244]	885 890 895
[5245]	Asn Leu Asn Leu Glu Glu Phe Ile Gln Lys Met Lys Glu Ile Glu Lys
[5246]	900 905 910
[5247]	Asp Phe Asp Asp Phe Lys Ile Gln Thr Lys Lys Glu Ile Phe Asn Asn
[5248]	915 920 925
[5249]	Tyr Tyr Glu Asp Ile Lys Asn Asn Ile Leu Thr Glu Phe Lys Asp Asp
[5250]	930 935 940
[5251]	Ile Asn Gly Cys Asp Val Leu Glu Lys Lys Leu Glu Lys Ile Val Ile
[5252]	945 950 955 960
[5253]	Phe Asp Asp Glu Thr Lys Phe Glu Ile Asp Lys Lys Ser Asn Ile Leu
[5254]	965 970 975
[5255]	Gln Asp Glu Gln Arg Lys Leu Ser Asn Ile Asn Lys Lys Asp Leu Lys
[5256]	980 985 990
[5257]	Lys Lys Val Asp Gln Tyr Ile Lys Asp Lys Asp Gln Glu Ile Lys Ser
[5258]	995 1000 1005
[5259]	Lys Ile Leu Cys Arg Ile Ile Phe Asn Ser Asp Phe Leu Lys Lys
[5260]	1010 1015 1020
[5261]	Tyr Lys Lys Glu Ile Asp Asn Leu Ile Glu Asp Met Glu Ser Glu
[5262]	1025 1030 1035
[5263]	Asn Glu Asn Lys Phe Gln Glu Ile Tyr Tyr Pro Lys Glu Arg Lys

[5264]	1040	1045	1050
[5265]	Asn Glu Leu Tyr Ile Tyr Lys Lys Asn Leu Phe Leu Asn Ile Gly		
[5266]	1055	1060	1065
[5267]	Asn Pro Asn Phe Asp Lys Ile Tyr Gly Leu Ile Ser Asn Asp Ile		
[5268]	1070	1075	1080
[5269]	Lys Met Ala Asp Ala Lys Phe Leu Phe Asn Ile Asp Gly Lys Asn		
[5270]	1085	1090	1095
[5271]	Ile Arg Lys Asn Lys Ile Ser Glu Ile Asp Ala Ile Leu Lys Asn		
[5272]	1100	1105	1110
[5273]	Leu Asn Asp Lys Leu Asn Gly Tyr Ser Lys Glu Tyr Lys Glu Lys		
[5274]	1115	1120	1125
[5275]	Tyr Ile Lys Lys Leu Lys Glu Asn Asp Asp Phe Phe Ala Lys Asn		
[5276]	1130	1135	1140
[5277]	Ile Gln Asn Lys Asn Tyr Lys Ser Phe Glu Lys Asp Tyr Asn Arg		
[5278]	1145	1150	1155
[5279]	Val Ser Glu Tyr Lys Lys Ile Arg Asp Leu Val Glu Phe Asn Tyr		
[5280]	1160	1165	1170
[5281]	Leu Asn Lys Ile Glu Ser Tyr Leu Ile Asp Ile Asn Trp Lys Leu		
[5282]	1175	1180	1185
[5283]	Ala Ile Gln Met Ala Arg Phe Glu Arg Asp Met His Tyr Ile Val		
[5284]	1190	1195	1200
[5285]	Asn Gly Leu Arg Glu Leu Gly Ile Ile Lys Leu Ser Gly Tyr Asn		
[5286]	1205	1210	1215
[5287]	Thr Gly Ile Ser Arg Ala Tyr Pro Lys Arg Asn Gly Ser Asp Gly		
[5288]	1220	1225	1230
[5289]	Phe Tyr Thr Thr Thr Ala Tyr Tyr Lys Phe Phe Asp Glu Glu Ser		
[5290]	1235	1240	1245
[5291]	Tyr Lys Lys Phe Glu Lys Ile Cys Tyr Gly Phe Gly Ile Asp Leu		
[5292]	1250	1255	1260
[5293]	Ser Glu Asn Ser Glu Ile Asn Lys Pro Glu Asn Glu Ser Ile Arg		
[5294]	1265	1270	1275
[5295]	Asn Tyr Ile Ser His Phe Tyr Ile Val Arg Asn Pro Phe Ala Asp		
[5296]	1280	1285	1290
[5297]	Tyr Ser Ile Ala Glu Gln Ile Asp Arg Val Ser Asn Leu Leu Ser		
[5298]	1295	1300	1305
[5299]	Tyr Ser Thr Arg Tyr Asn Asn Ser Thr Tyr Ala Ser Val Phe Glu		
[5300]	1310	1315	1320
[5301]	Val Phe Lys Lys Asp Val Asn Leu Asp Tyr Asp Glu Leu Lys Lys		
[5302]	1325	1330	1335

[5303]	Lys Phe	Lys Leu	Ile Gly	Asn Asn	Asp Ile	Leu Glu	Arg Leu	Met	
[5304]	1340			1345			1350		
[5305]	Lys Pro	Lys Lys	Val Ser	Val Leu	Glu Leu	Glu Ser	Tyr Asn	Ser	
[5306]	1355			1360			1365		
[5307]	Asp Tyr	Ile Lys	Asn Leu	Ile Ile	Glu Leu	Leu Thr	Lys Ile	Glu	
[5308]	1370			1375			1380		
[5309]	Asn Thr	Asn Asp	Thr Leu						
[5310]	1385								
[5311]	<210>	310							
[5312]	<211>	1182							
[5313]	<212>	PRT							
[5314]	<213>	韦德纤毛菌							
[5315]	<400>	310							
[5316]	Met Tyr	Met Lys	Ile Thr	Lys Ile	Asp Gly	Val Ser	His Tyr	Lys Lys	
[5317]	1		5			10		15	
[5318]	Gln Asp	Lys Gly	Ile Leu	Lys Lys	Lys Trp	Lys Asp	Leu Asp	Glu Arg	
[5319]		20			25			30	
[5320]	Lys Gln	Arg Glu	Lys Ile	Glu Ala	Arg Tyr	Asn Lys	Gln Ile	Glu Ser	
[5321]		35			40			45	
[5322]	Lys Ile	Tyr Lys	Glu Phe	Phe Arg	Leu Lys	Asn Lys	Lys Arg	Ile Glu	
[5323]	50			55			60		
[5324]	Lys Glu	Glu Asp	Gln Asn	Ile Lys	Ser Leu	Tyr Phe	Phe Ile	Lys Glu	
[5325]	65			70			75		80
[5326]	Leu Tyr	Leu Asn	Glu Lys	Asn Glu	Glu Trp	Glu Leu	Lys Asn	Ile Asn	
[5327]			85			90		95	
[5328]	Leu Glu	Ile Leu	Asp Asp	Lys Glu	Arg Val	Ile Lys	Gly Tyr	Lys Phe	
[5329]			100			105		110	
[5330]	Lys Glu	Asp Val	Tyr Phe	Phe Lys	Glu Gly	Tyr Lys	Glu Tyr	Tyr Leu	
[5331]			115			120		125	
[5332]	Arg Ile	Leu Phe	Asn Asn	Leu Ile	Glu Lys	Val Gln	Asn Glu	Asn Arg	
[5333]			130			135		140	
[5334]	Glu Lys	Val Arg	Lys Asn	Lys Glu	Phe Leu	Asp Leu	Lys Glu	Ile Phe	
[5335]	145			150			155		160
[5336]	Lys Lys	Tyr Lys	Asn Arg	Lys Ile	Asp Leu	Leu Leu	Lys Ser	Ile Asn	
[5337]			165			170		175	
[5338]	Asn Asn	Lys Ile	Asn Leu	Glu Tyr	Lys Lys	Glu Asn	Val Asn	Glu Glu	
[5339]			180			185		190	
[5340]	Ile Tyr	Gly Ile	Asn Pro	Thr Asn	Asp Arg	Glu Met	Thr Phe	Tyr Glu	
[5341]			195			200		205	

[5342]	Leu	Leu	Lys	Glu	Ile	Ile	Glu	Lys	Lys	Asp	Glu	Gln	Lys	Ser	Ile	Leu
[5343]	210						215					220				
[5344]	Glu	Glu	Lys	Leu	Asp	Asn	Phe	Asp	Ile	Thr	Asn	Phe	Leu	Glu	Asn	Ile
[5345]	225					230					235				240	
[5346]	Glu	Lys	Ile	Phe	Asn	Glu	Glu	Thr	Glu	Ile	Asn	Ile	Ile	Lys	Gly	Lys
[5347]					245					250				255		
[5348]	Val	Leu	Asn	Glu	Leu	Arg	Glu	Tyr	Ile	Lys	Glu	Lys	Glu	Glu	Asn	Asn
[5349]			260						265					270		
[5350]	Ser	Asp	Asn	Lys	Leu	Lys	Gln	Ile	Tyr	Asn	Leu	Glu	Leu	Lys	Lys	Tyr
[5351]		275						280					285			
[5352]	Ile	Glu	Asn	Asn	Phe	Ser	Tyr	Lys	Lys	Gln	Lys	Ser	Lys	Ser	Lys	Asn
[5353]	290						295					300				
[5354]	Gly	Lys	Asn	Asp	Tyr	Leu	Tyr	Leu	Asn	Phe	Leu	Lys	Lys	Ile	Met	Phe
[5355]	305					310					315				320	
[5356]	Ile	Glu	Glu	Val	Asp	Glu	Lys	Lys	Glu	Ile	Asn	Lys	Glu	Lys	Phe	Lys
[5357]					325					330					335	
[5358]	Asn	Lys	Ile	Asn	Ser	Asn	Phe	Lys	Asn	Leu	Phe	Val	Gln	His	Ile	Leu
[5359]			340					345						350		
[5360]	Asp	Tyr	Gly	Lys	Leu	Leu	Tyr	Tyr	Lys	Glu	Asn	Asp	Glu	Tyr	Ile	Lys
[5361]		355						360					365			
[5362]	Asn	Thr	Gly	Gln	Leu	Glu	Thr	Lys	Asp	Leu	Glu	Tyr	Ile	Lys	Thr	Lys
[5363]	370							375					380			
[5364]	Glu	Thr	Leu	Ile	Arg	Lys	Met	Ala	Val	Leu	Val	Ser	Phe	Ala	Ala	Asn
[5365]	385					390					395					400
[5366]	Ser	Tyr	Tyr	Asn	Leu	Phe	Gly	Arg	Val	Ser	Gly	Asp	Ile	Leu	Gly	Thr
[5367]				405					410					415		
[5368]	Glu	Val	Val	Lys	Ser	Ser	Lys	Thr	Asn	Val	Ile	Lys	Val	Gly	Ser	His
[5369]				420					425					430		
[5370]	Ile	Phe	Lys	Glu	Lys	Met	Leu	Asn	Tyr	Phe	Phe	Asp	Phe	Glu	Ile	Phe
[5371]		435						440					445			
[5372]	Asp	Ala	Asn	Lys	Ile	Val	Glu	Ile	Leu	Glu	Ser	Ile	Ser	Tyr	Ser	Ile
[5373]		450						455					460			
[5374]	Tyr	Asn	Val	Arg	Asn	Gly	Val	Gly	His	Phe	Asn	Lys	Leu	Ile	Leu	Gly
[5375]	465					470					475				480	
[5376]	Lys	Tyr	Lys	Lys	Lys	Asp	Ile	Asn	Thr	Asn	Lys	Arg	Ile	Glu	Glu	Asp
[5377]					485					490				495		
[5378]	Leu	Asn	Asn	Asn	Glu	Glu	Ile	Lys	Gly	Tyr	Phe	Ile	Lys	Lys	Arg	Gly
[5379]				500					505				510			
[5380]	Glu	Ile	Glu	Arg	Lys	Val	Lys	Glu	Lys	Phe	Leu	Ser	Asn	Asn	Leu	Gln

[5381]	515	520	525
[5382]	Tyr Tyr Tyr Ser Lys Glu Lys Ile Glu Asn Tyr Phe Glu Val Tyr Glu		
[5383]	530	535	540
[5384]	Phe Glu Ile Leu Lys Arg Lys Ile Pro Phe Ala Pro Asn Phe Lys Arg		
[5385]	545	550	555
[5386]	Ile Ile Lys Lys Gly Glu Asp Leu Phe Asn Asn Lys Asn Asn Lys Lys		
[5387]	565	570	575
[5388]	Tyr Glu Tyr Phe Lys Asn Phe Asp Lys Asn Ser Ala Glu Glu Lys Lys		
[5389]	580	585	590
[5390]	Glu Phe Leu Lys Thr Arg Asn Phe Leu Leu Lys Glu Leu Tyr Tyr Asn		
[5391]	595	600	605
[5392]	Asn Phe Tyr Lys Glu Phe Leu Ser Lys Lys Glu Glu Phe Glu Lys Ile		
[5393]	610	615	620
[5394]	Val Leu Glu Val Lys Glu Glu Lys Lys Ser Arg Gly Asn Ile Asn Asn		
[5395]	625	630	635
[5396]	Lys Lys Ser Gly Val Ser Phe Gln Ser Ile Asp Asp Tyr Asp Thr Lys		
[5397]	645	650	655
[5398]	Ile Asn Ile Ser Asp Tyr Ile Ala Ser Ile His Lys Lys Glu Met Glu		
[5399]	660	665	670
[5400]	Arg Val Glu Lys Tyr Asn Glu Glu Lys Gln Lys Asp Thr Ala Lys Tyr		
[5401]	675	680	685
[5402]	Ile Arg Asp Phe Val Glu Glu Ile Phe Leu Thr Gly Phe Ile Asn Tyr		
[5403]	690	695	700
[5404]	Leu Glu Lys Asp Lys Arg Leu His Phe Leu Lys Glu Glu Phe Ser Ile		
[5405]	705	710	715
[5406]	Leu Cys Asn Asn Asn Asn Asn Val Val Asp Phe Asn Ile Asn Ile Asn		
[5407]	725	730	735
[5408]	Glu Glu Lys Ile Lys Glu Phe Leu Lys Glu Asn Asp Ser Lys Thr Leu		
[5409]	740	745	750
[5410]	Asn Leu Tyr Leu Phe Phe Asn Met Ile Asp Ser Lys Arg Ile Ser Glu		
[5411]	755	760	765
[5412]	Phe Arg Asn Glu Leu Val Lys Tyr Lys Gln Phe Thr Lys Lys Arg Leu		
[5413]	770	775	780
[5414]	Asp Glu Glu Lys Glu Phe Leu Gly Ile Lys Ile Glu Leu Tyr Glu Thr		
[5415]	785	790	795
[5416]	Leu Ile Glu Phe Val Ile Leu Thr Arg Glu Lys Leu Asp Thr Lys Lys		
[5417]	805	810	815
[5418]	Ser Glu Glu Ile Asp Ala Trp Leu Val Asp Lys Leu Tyr Val Lys Asp		
[5419]	820	825	830

[5420]	Ser Asn Glu Tyr Lys Glu Tyr Glu Glu Ile Leu Lys Leu Phe Val Asp
[5421]	835 840 845
[5422]	Glu Lys Ile Leu Ser Ser Lys Glu Ala Pro Tyr Tyr Ala Thr Asp Asn
[5423]	850 855 860
[5424]	Lys Thr Pro Ile Leu Leu Ser Asn Phe Glu Lys Thr Arg Lys Tyr Gly
[5425]	865 870 875 880
[5426]	Thr Gln Ser Phe Leu Ser Glu Ile Gln Ser Asn Tyr Lys Tyr Ser Lys
[5427]	885 890 895
[5428]	Val Glu Lys Glu Asn Ile Glu Asp Tyr Asn Lys Lys Glu Glu Ile Glu
[5429]	900 905 910
[5430]	Gln Lys Lys Lys Ser Asn Ile Glu Lys Leu Gln Asp Leu Lys Val Glu
[5431]	915 920 925
[5432]	Leu His Lys Lys Trp Glu Gln Asn Lys Ile Thr Glu Lys Glu Ile Glu
[5433]	930 935 940
[5434]	Lys Tyr Asn Asn Thr Thr Arg Lys Ile Asn Glu Tyr Asn Tyr Leu Lys
[5435]	945 950 955 960
[5436]	Asn Lys Glu Glu Leu Gln Asn Val Tyr Leu Leu His Glu Met Leu Ser
[5437]	965 970 975
[5438]	Asp Leu Leu Ala Arg Asn Val Ala Phe Phe Asn Lys Trp Glu Arg Asp
[5439]	980 985 990
[5440]	Phe Lys Phe Ile Val Ile Ala Ile Lys Gln Phe Leu Arg Glu Asn Asp
[5441]	995 1000 1005
[5442]	Lys Glu Lys Val Asn Glu Phe Leu Asn Pro Pro Asp Asn Ser Lys
[5443]	1010 1015 1020
[5444]	Gly Lys Lys Val Tyr Phe Ser Val Ser Lys Tyr Lys Asn Thr Val
[5445]	1025 1030 1035
[5446]	Glu Asn Ile Asp Gly Ile His Lys Asn Phe Met Asn Leu Ile Phe
[5447]	1040 1045 1050
[5448]	Leu Asn Asn Lys Phe Met Asn Arg Lys Ile Asp Lys Met Asn Cys
[5449]	1055 1060 1065
[5450]	Ala Ile Trp Val Tyr Phe Arg Asn Tyr Ile Ala His Phe Leu His
[5451]	1070 1075 1080
[5452]	Leu His Thr Lys Asn Glu Lys Ile Ser Leu Ile Ser Gln Met Asn
[5453]	1085 1090 1095
[5454]	Leu Leu Ile Lys Leu Phe Ser Tyr Asp Lys Lys Val Gln Asn His
[5455]	1100 1105 1110
[5456]	Ile Leu Lys Ser Thr Lys Thr Leu Leu Glu Lys Tyr Asn Ile Gln
[5457]	1115 1120 1125
[5458]	Ile Asn Phe Glu Ile Ser Asn Asp Lys Asn Glu Val Phe Lys Tyr

[5459]	1130	1135	1140
[5460]	Lys Ile Lys Asn Arg Leu Tyr Ser Lys Lys Gly Lys Met Leu Gly		
[5461]	1145	1150	1155
[5462]	Lys Asn Asn Lys Phe Glu Ile Leu Glu Asn Glu Phe Leu Glu Asn		
[5463]	1160	1165	1170
[5464]	Val Lys Ala Met Leu Glu Tyr Ser Glu		
[5465]	1175	1180	
[5466]	<210> 311		
[5467]	<211> 1385		
[5468]	<212> PRT		
[5469]	<213> 未知的		
[5470]	<220>		
[5471]	<223> 纤毛菌物种口腔分类879		
[5472]	<400> 311		
[5473]	Met Gly Asn Leu Phe Gly His Lys Arg Trp Tyr Glu Val Arg Asp Lys		
[5474]	1 5 10 15		
[5475]	Lys Asp Phe Lys Ile Lys Arg Lys Val Lys Val Lys Arg Asn Tyr Asp		
[5476]	20 25 30		
[5477]	Gly Asn Lys Tyr Ile Leu Asn Ile Asn Glu Asn Asn Asn Lys Glu Lys		
[5478]	35 40 45		
[5479]	Ile Asp Asn Asn Lys Phe Ile Gly Glu Phe Val Asn Tyr Lys Lys Asn		
[5480]	50 55 60		
[5481]	Asn Asn Val Leu Lys Glu Phe Lys Arg Lys Phe His Ala Gly Asn Ile		
[5482]	65 70 75 80		
[5483]	Leu Phe Lys Leu Lys Gly Lys Glu Glu Ile Ile Arg Ile Glu Asn Asn		
[5484]	85 90 95		
[5485]	Asp Asp Phe Leu Glu Thr Glu Glu Val Val Leu Tyr Ile Glu Val Tyr		
[5486]	100 105 110		
[5487]	Gly Lys Ser Glu Lys Leu Lys Ala Leu Glu Ile Thr Lys Lys Lys Ile		
[5488]	115 120 125		
[5489]	Ile Asp Glu Ala Ile Arg Gln Gly Ile Thr Lys Asp Asp Lys Lys Ile		
[5490]	130 135 140		
[5491]	Glu Ile Lys Arg Gln Glu Asn Glu Glu Glu Ile Glu Ile Asp Ile Arg		
[5492]	145 150 155 160		
[5493]	Asp Glu Tyr Thr Asn Lys Thr Leu Asn Asp Cys Ser Ile Ile Leu Arg		
[5494]	165 170 175		
[5495]	Ile Ile Glu Asn Asp Glu Leu Glu Thr Lys Lys Ser Ile Tyr Glu Ile		
[5496]	180 185 190		
[5497]	Phe Lys Asn Ile Asn Met Ser Leu Tyr Lys Ile Ile Glu Lys Ile Ile		

[5498]	195	200	205
[5499]	Glu Asn Glu Thr Glu Lys Val Phe Glu Asn Arg Tyr Tyr Glu Glu His		
[5500]	210	215	220
[5501]	Leu Arg Glu Lys Leu Leu Lys Asp Asn Lys Ile Asp Val Ile Leu Thr		
[5502]	225	230	235
[5503]	Asn Phe Met Glu Ile Arg Glu Lys Ile Lys Ser Asn Leu Glu Ile Met		
[5504]	245	250	255
[5505]	Gly Phe Val Lys Phe Tyr Leu Asn Val Ser Gly Asp Lys Lys Lys Ser		
[5506]	260	265	270
[5507]	Glu Asn Lys Lys Met Phe Val Glu Lys Ile Leu Asn Thr Asn Val Asp		
[5508]	275	280	285
[5509]	Leu Thr Val Glu Asp Ile Val Asp Phe Ile Val Lys Glu Leu Lys Phe		
[5510]	290	295	300
[5511]	Trp Asn Ile Thr Lys Arg Ile Glu Lys Val Lys Lys Phe Asn Asn Glu		
[5512]	305	310	315
[5513]	Phe Leu Glu Asn Arg Arg Asn Arg Thr Tyr Ile Lys Ser Tyr Val Leu		
[5514]	325	330	335
[5515]	Leu Asp Lys His Glu Lys Phe Lys Ile Glu Arg Glu Asn Lys Lys Asp		
[5516]	340	345	350
[5517]	Lys Ile Val Lys Phe Phe Val Glu Asn Ile Lys Asn Asn Ser Ile Lys		
[5518]	355	360	365
[5519]	Glu Lys Ile Glu Lys Ile Leu Ala Glu Phe Lys Ile Asn Glu Leu Ile		
[5520]	370	375	380
[5521]	Lys Lys Leu Glu Lys Glu Leu Lys Lys Gly Asn Cys Asp Thr Glu Ile		
[5522]	385	390	395
[5523]	Phe Gly Ile Phe Lys Lys His Tyr Lys Val Asn Phe Asp Ser Lys Lys		
[5524]	405	410	415
[5525]	Phe Ser Asn Lys Ser Asp Glu Glu Lys Glu Leu Tyr Lys Ile Ile Tyr		
[5526]	420	425	430
[5527]	Arg Tyr Leu Lys Gly Arg Ile Glu Lys Ile Leu Val Asn Glu Gln Lys		
[5528]	435	440	445
[5529]	Val Arg Leu Lys Lys Met Glu Lys Ile Glu Ile Glu Lys Ile Leu Asn		
[5530]	450	455	460
[5531]	Glu Ser Ile Leu Ser Glu Lys Ile Leu Lys Arg Val Lys Gln Tyr Thr		
[5532]	465	470	475
[5533]	Leu Glu His Ile Met Tyr Leu Gly Lys Leu Arg His Asn Asp Ile Val		
[5534]	485	490	495
[5535]	Lys Met Thr Val Asn Thr Asp Asp Phe Ser Arg Leu His Ala Lys Glu		
[5536]	500	505	510

[5537]	Glu Leu Asp Leu Glu Leu Ile Thr Phe Phe Ala Ser Thr Asn Met Glu
[5538]	515 520 525
[5539]	Leu Asn Lys Ile Phe Asn Gly Lys Glu Lys Val Thr Asp Phe Phe Gly
[5540]	530 535 540
[5541]	Phe Asn Leu Asn Gly Gln Lys Ile Thr Leu Lys Glu Lys Val Pro Ser
[5542]	545 550 555 560
[5543]	Phe Lys Leu Asn Ile Leu Lys Lys Leu Asn Phe Ile Asn Asn Glu Asn
[5544]	565 570 575
[5545]	Asn Ile Asp Glu Lys Leu Ser His Phe Tyr Ser Phe Gln Lys Glu Gly
[5546]	580 585 590
[5547]	Tyr Leu Leu Arg Asn Lys Ile Leu His Asn Ser Tyr Gly Asn Ile Gln
[5548]	595 600 605
[5549]	Glu Thr Lys Asn Leu Lys Gly Glu Tyr Glu Asn Val Glu Lys Leu Ile
[5550]	610 615 620
[5551]	Lys Glu Leu Lys Val Ser Asp Glu Glu Ile Ser Lys Ser Leu Ser Leu
[5552]	625 630 635 640
[5553]	Asp Val Ile Phe Glu Gly Lys Val Asp Ile Ile Asn Lys Ile Asn Ser
[5554]	645 650 655
[5555]	Leu Lys Ile Gly Glu Tyr Lys Asp Lys Lys Tyr Leu Pro Ser Phe Ser
[5556]	660 665 670
[5557]	Lys Ile Val Leu Glu Ile Thr Arg Lys Phe Arg Glu Ile Asn Lys Asp
[5558]	675 680 685
[5559]	Lys Leu Phe Asp Ile Glu Ser Glu Lys Ile Ile Leu Asn Ala Val Lys
[5560]	690 695 700
[5561]	Tyr Val Asn Lys Ile Leu Tyr Glu Lys Ile Thr Ser Asn Glu Glu Asn
[5562]	705 710 715 720
[5563]	Glu Phe Leu Lys Thr Leu Pro Asp Lys Leu Val Lys Lys Ser Asn Asn
[5564]	725 730 735
[5565]	Lys Lys Glu Asn Lys Asn Leu Leu Ser Ile Glu Glu Tyr Tyr Lys Asn
[5566]	740 745 750
[5567]	Ala Gln Val Ser Ser Ser Lys Gly Asp Lys Lys Ala Ile Lys Lys Tyr
[5568]	755 760 765
[5569]	Gln Asn Lys Val Thr Asn Ala Tyr Leu Glu Tyr Leu Glu Asn Thr Phe
[5570]	770 775 780
[5571]	Thr Glu Ile Ile Asp Phe Ser Lys Phe Asn Leu Asn Tyr Asp Glu Ile
[5572]	785 790 795 800
[5573]	Lys Thr Lys Ile Glu Glu Arg Lys Asp Asn Lys Ser Lys Ile Ile Ile
[5574]	805 810 815
[5575]	Asp Ser Ile Ser Thr Asn Ile Asn Ile Thr Asn Asp Ile Glu Tyr Ile

[5576]	820	825	830
[5577]	Ile Ser Ile Phe Ala Leu Leu Asn Ser Asn Thr Tyr Ile Asn Lys Ile		
[5578]	835	840	845
[5579]	Arg Asn Arg Phe Phe Ala Thr Ser Val Trp Leu Glu Lys Gln Asn Gly		
[5580]	850	855	860
[5581]	Thr Lys Glu Tyr Asp Tyr Glu Asn Ile Ile Ser Ile Leu Asp Glu Val		
[5582]	865	870	875
[5583]	Leu Leu Ile Asn Leu Leu Arg Glu Asn Asn Ile Thr Asp Ile Leu Asp		
[5584]	885	890	895
[5585]	Leu Lys Asn Ala Ile Ile Asp Ala Lys Ile Val Glu Asn Asp Glu Thr		
[5586]	900	905	910
[5587]	Tyr Ile Lys Asn Tyr Ile Phe Glu Ser Asn Glu Glu Lys Leu Lys Lys		
[5588]	915	920	925
[5589]	Arg Leu Phe Cys Glu Glu Leu Val Asp Lys Glu Asp Ile Arg Lys Ile		
[5590]	930	935	940
[5591]	Phe Glu Asp Glu Asn Phe Lys Phe Lys Ser Phe Ile Lys Lys Asn Glu		
[5592]	945	950	955
[5593]	Ile Gly Asn Phe Lys Ile Asn Phe Gly Ile Leu Ser Asn Leu Glu Cys		
[5594]	965	970	975
[5595]	Asn Ser Glu Val Glu Ala Lys Lys Ile Ile Gly Lys Asn Ser Lys Lys		
[5596]	980	985	990
[5597]	Leu Glu Ser Phe Ile Gln Asn Ile Ile Asp Glu Tyr Lys Ser Asn Ile		
[5598]	995	1000	1005
[5599]	Arg Thr Leu Phe Ser Ser Glu Phe Leu Glu Lys Tyr Lys Glu Glu		
[5600]	1010	1015	1020
[5601]	Ile Asp Asn Leu Val Glu Asp Thr Glu Ser Glu Asn Lys Asn Lys		
[5602]	1025	1030	1035
[5603]	Phe Glu Lys Ile Tyr Tyr Pro Lys Glu His Lys Asn Glu Leu Tyr		
[5604]	1040	1045	1050
[5605]	Ile Tyr Lys Lys Asn Leu Phe Leu Asn Ile Gly Asn Pro Asn Phe		
[5606]	1055	1060	1065
[5607]	Asp Lys Ile Tyr Gly Leu Ile Ser Lys Asp Ile Lys Asn Val Asp		
[5608]	1070	1075	1080
[5609]	Thr Lys Ile Leu Phe Asp Asp Asp Ile Lys Lys Asn Lys Ile Ser		
[5610]	1085	1090	1095
[5611]	Glu Ile Asp Ala Ile Leu Lys Asn Leu Asn Asp Lys Leu Asn Gly		
[5612]	1100	1105	1110
[5613]	Tyr Ser Asn Asp Tyr Lys Ala Lys Tyr Val Asn Lys Leu Lys Glu		
[5614]	1115	1120	1125

[5615]	Asn Asp Asp Phe Phe Ala Lys Asn Ile Gln Asn Glu Asn Tyr Ser
[5616]	1130 1135 1140
[5617]	Ser Phe Gly Glu Phe Glu Lys Asp Tyr Asn Lys Val Ser Glu Tyr
[5618]	1145 1150 1155
[5619]	Lys Lys Ile Arg Asp Leu Val Glu Phe Asn Tyr Leu Asn Lys Ile
[5620]	1160 1165 1170
[5621]	Glu Ser Tyr Leu Ile Asp Ile Asn Trp Lys Leu Ala Ile Gln Met
[5622]	1175 1180 1185
[5623]	Ala Arg Phe Glu Arg Asp Met His Tyr Ile Val Asn Gly Leu Arg
[5624]	1190 1195 1200
[5625]	Glu Leu Gly Ile Ile Lys Leu Ser Gly Tyr Asn Thr Gly Ile Ser
[5626]	1205 1210 1215
[5627]	Arg Ala Tyr Pro Lys Arg Asn Gly Ser Asp Gly Phe Tyr Thr Thr
[5628]	1220 1225 1230
[5629]	Thr Ala Tyr Tyr Lys Phe Phe Asp Glu Glu Ser Tyr Lys Lys Phe
[5630]	1235 1240 1245
[5631]	Glu Lys Ile Cys Tyr Gly Phe Gly Ile Asp Leu Ser Glu Asn Ser
[5632]	1250 1255 1260
[5633]	Glu Ile Asn Lys Pro Glu Asn Glu Ser Ile Arg Asn Tyr Ile Ser
[5634]	1265 1270 1275
[5635]	His Phe Tyr Ile Val Arg Asn Pro Phe Ala Asp Tyr Ser Ile Ala
[5636]	1280 1285 1290
[5637]	Glu Gln Ile Asp Arg Val Ser Asn Leu Leu Ser Tyr Ser Thr Arg
[5638]	1295 1300 1305
[5639]	Tyr Asn Asn Ser Thr Tyr Ala Ser Val Phe Glu Val Phe Lys Lys
[5640]	1310 1315 1320
[5641]	Asp Val Asn Leu Asp Tyr Asp Glu Leu Lys Lys Lys Phe Arg Leu
[5642]	1325 1330 1335
[5643]	Ile Gly Asn Asn Asp Ile Leu Glu Arg Leu Met Lys Pro Lys Lys
[5644]	1340 1345 1350
[5645]	Val Ser Val Leu Glu Leu Glu Ser Tyr Asn Ser Asp Tyr Ile Lys
[5646]	1355 1360 1365
[5647]	Asn Leu Ile Ile Glu Leu Leu Thr Lys Ile Glu Asn Thr Asn Asp
[5648]	1370 1375 1380
[5649]	Thr Leu
[5650]	1385
[5651]	<210> 312
[5652]	<211> 1152
[5653]	<212> PRT

[5654]	<213>	韦德纤毛菌															
[5655]	<400>	312															
[5656]	Met Lys Val Thr Lys Val Asp Gly Ile Ser His Lys Lys Tyr Ile Glu																
[5657]	1	5				10				15							
[5658]	Glu Gly Lys Leu Val Lys Ser Thr Ser Glu Glu Asn Arg Thr Ser Glu																
[5659]		20				25				30							
[5660]	Arg Leu Ser Glu Leu Leu Ser Ile Arg Leu Asp Ile Tyr Ile Lys Asn																
[5661]		35				40				45							
[5662]	Pro Asp Asn Ala Ser Glu Glu Glu Asn Arg Ile Arg Arg Glu Asn Leu																
[5663]		50				55				60							
[5664]	Lys Lys Phe Phe Ser Asn Lys Val Leu His Leu Lys Asp Ser Val Leu																
[5665]	65	70				75				80							
[5666]	Tyr Leu Lys Asn Arg Lys Glu Lys Asn Ala Val Gln Asp Lys Asn Tyr																
[5667]		85				90				95							
[5668]	Ser Glu Glu Asp Ile Ser Glu Tyr Asp Leu Lys Asn Lys Asn Ser Phe																
[5669]		100				105				110							
[5670]	Ser Val Leu Lys Lys Ile Leu Leu Asn Glu Asp Val Asn Ser Glu Glu																
[5671]		115				120				125							
[5672]	Leu Glu Ile Phe Arg Lys Asp Val Glu Ala Lys Leu Asn Lys Ile Asn																
[5673]		130				135				140							
[5674]	Ser Leu Lys Tyr Ser Phe Glu Glu Asn Lys Ala Asn Tyr Gln Lys Ile																
[5675]	145	150				155				160							
[5676]	Asn Glu Asn Asn Val Glu Lys Val Gly Gly Lys Ser Lys Arg Asn Ile																
[5677]		165				170				175							
[5678]	Ile Tyr Asp Tyr Tyr Arg Glu Ser Ala Lys Arg Asn Asp Tyr Ile Asn																
[5679]		180				185				190							
[5680]	Asn Val Gln Glu Ala Phe Asp Lys Leu Tyr Lys Lys Glu Asp Ile Glu																
[5681]		195				200				205							
[5682]	Lys Leu Phe Phe Leu Ile Glu Asn Ser Lys Lys His Glu Lys Tyr Lys																
[5683]		210				215				220							
[5684]	Ile Arg Glu Tyr Tyr His Lys Ile Ile Gly Arg Lys Asn Asp Lys Glu																
[5685]	225	230				235				240							
[5686]	Asn Phe Ala Lys Ile Ile Tyr Glu Glu Ile Gln Asn Val Asn Asn Ile																
[5687]		245				250				255							
[5688]	Lys Glu Leu Ile Glu Lys Ile Pro Asp Met Ser Glu Leu Lys Lys Ser																
[5689]		260				265				270							
[5690]	Gln Val Phe Tyr Lys Tyr Tyr Leu Asp Lys Glu Glu Leu Asn Asp Lys																
[5691]		275				280				285							
[5692]	Asn Ile Lys Tyr Ala Phe Cys His Phe Val Glu Ile Glu Met Ser Glu																

[5693]	290	295	300
[5694]	Leu Leu Lys Asn Tyr Val Tyr Lys Arg Leu Ser Asn Ile Ser Asn Asp		
[5695]	305	310	315
[5696]	Lys Ile Lys Arg Ile Phe Glu Tyr Gln Asn Leu Lys Lys Leu Ile Glu		
[5697]	325	330	335
[5698]	Asn Lys Leu Leu Asn Lys Leu Asp Thr Tyr Val Arg Asn Cys Gly Lys		
[5699]	340	345	350
[5700]	Tyr Asn Tyr Tyr Leu Gln Val Gly Glu Ile Ala Thr Ser Asp Phe Ile		
[5701]	355	360	365
[5702]	Ala Arg Asn Arg Gln Asn Glu Ala Phe Leu Arg Asn Ile Ile Gly Val		
[5703]	370	375	380
[5704]	Ser Ser Val Ala Tyr Phe Ser Leu Arg Asn Ile Leu Glu Thr Glu Asn		
[5705]	385	390	395
[5706]	Glu Asn Asp Ile Thr Gly Arg Met Arg Gly Lys Thr Val Lys Asn Asn		
[5707]	405	410	415
[5708]	Lys Gly Glu Glu Lys Tyr Val Ser Gly Glu Val Asp Lys Ile Tyr Asn		
[5709]	420	425	430
[5710]	Glu Asn Lys Gln Asn Glu Val Lys Glu Asn Leu Lys Met Phe Tyr Ser		
[5711]	435	440	445
[5712]	Tyr Asp Phe Asn Met Asp Asn Lys Asn Glu Ile Glu Asp Phe Phe Ala		
[5713]	450	455	460
[5714]	Asn Ile Asp Glu Ala Ile Ser Ser Ile Arg His Gly Ile Val His Phe		
[5715]	465	470	475
[5716]	Asn Leu Glu Leu Glu Gly Lys Asp Ile Phe Ala Phe Lys Asn Ile Ala		
[5717]	485	490	495
[5718]	Pro Ser Glu Ile Ser Lys Lys Met Phe Gln Asn Glu Ile Asn Glu Lys		
[5719]	500	505	510
[5720]	Lys Leu Lys Leu Lys Ile Phe Lys Gln Leu Asn Ser Ala Asn Val Phe		
[5721]	515	520	525
[5722]	Asn Tyr Tyr Glu Lys Asp Val Ile Ile Lys Tyr Leu Lys Asn Thr Lys		
[5723]	530	535	540
[5724]	Phe Asn Phe Val Asn Lys Asn Ile Pro Phe Val Pro Ser Phe Thr Lys		
[5725]	545	550	555
[5726]	Leu Tyr Asn Lys Ile Glu Asp Leu Arg Asn Thr Leu Lys Phe Phe Trp		
[5727]	565	570	575
[5728]	Ser Val Pro Lys Asp Lys Glu Glu Lys Asp Ala Gln Ile Tyr Leu Leu		
[5729]	580	585	590
[5730]	Lys Asn Ile Tyr Tyr Gly Glu Phe Leu Asn Lys Phe Val Lys Asn Ser		
[5731]	595	600	605

[5732]	Lys Val Phe Phe Lys Ile Thr Asn Glu Val Ile Lys Ile Asn Lys Gln
[5733]	610 615 620
[5734]	Arg Asn Gln Lys Thr Gly His Tyr Lys Tyr Gln Lys Phe Glu Asn Ile
[5735]	625 630 635 640
[5736]	Glu Lys Thr Val Pro Val Glu Tyr Leu Ala Ile Ile Gln Ser Arg Glu
[5737]	645 650 655
[5738]	Met Ile Asn Asn Gln Asp Lys Glu Glu Lys Asn Thr Tyr Ile Asp Phe
[5739]	660 665 670
[5740]	Ile Gln Gln Ile Phe Leu Lys Gly Phe Ile Asp Tyr Leu Asn Lys Asn
[5741]	675 680 685
[5742]	Asn Leu Lys Tyr Ile Glu Ser Asn Asn Asn Asn Asp Asn Asn Asp Ile
[5743]	690 695 700
[5744]	Phe Ser Lys Ile Lys Ile Lys Lys Asp Asn Lys Glu Lys Tyr Asp Lys
[5745]	705 710 715 720
[5746]	Ile Leu Lys Asn Tyr Glu Lys His Asn Arg Asn Lys Glu Ile Pro His
[5747]	725 730 735
[5748]	Glu Ile Asn Glu Phe Val Arg Glu Ile Lys Leu Gly Lys Ile Leu Lys
[5749]	740 745 750
[5750]	Tyr Thr Glu Asn Leu Asn Met Phe Tyr Leu Ile Leu Lys Leu Leu Asn
[5751]	755 760 765
[5752]	His Lys Glu Leu Thr Asn Leu Lys Gly Ser Leu Glu Lys Tyr Gln Ser
[5753]	770 775 780
[5754]	Ala Asn Lys Glu Glu Thr Phe Ser Asp Glu Leu Glu Leu Ile Asn Leu
[5755]	785 790 795 800
[5756]	Leu Asn Leu Asp Asn Asn Arg Val Thr Glu Asp Phe Glu Leu Glu Ala
[5757]	805 810 815
[5758]	Asn Glu Ile Gly Lys Phe Leu Asp Phe Asn Glu Asn Lys Ile Lys Asp
[5759]	820 825 830
[5760]	Arg Lys Glu Leu Lys Lys Phe Asp Thr Asn Lys Ile Tyr Phe Asp Gly
[5761]	835 840 845
[5762]	Glu Asn Ile Ile Lys His Arg Ala Phe Tyr Asn Ile Lys Lys Tyr Gly
[5763]	850 855 860
[5764]	Met Leu Asn Leu Leu Glu Lys Ile Ala Asp Lys Ala Lys Tyr Lys Ile
[5765]	865 870 875 880
[5766]	Ser Leu Lys Glu Leu Lys Glu Tyr Ser Asn Lys Lys Asn Glu Ile Glu
[5767]	885 890 895
[5768]	Lys Asn Tyr Thr Met Gln Gln Asn Leu His Arg Lys Tyr Ala Arg Pro
[5769]	900 905 910
[5770]	Lys Lys Asp Glu Lys Phe Asn Asp Glu Asp Tyr Lys Glu Tyr Glu Lys

[5771]	915	920	925
[5772]	Ala Ile Gly Asn Ile Gln Lys Tyr Thr His Leu Lys Asn Lys Val Glu		
[5773]	930	935	940
[5774]	Phe Asn Glu Leu Asn Leu Leu Gln Gly Leu Leu Leu Lys Ile Leu His		
[5775]	945	950	955
[5776]	Arg Leu Val Gly Tyr Thr Ser Ile Trp Glu Arg Asp Leu Arg Phe Arg		
[5777]	965	970	975
[5778]	Leu Lys Gly Glu Phe Pro Glu Asn His Tyr Ile Glu Glu Ile Phe Asn		
[5779]	980	985	990
[5780]	Phe Asp Asn Ser Lys Asn Val Lys Tyr Lys Ser Gly Gln Ile Val Glu		
[5781]	995	1000	1005
[5782]	Lys Tyr Ile Asn Phe Tyr Lys Glu Leu Tyr Lys Asp Asn Val Glu		
[5783]	1010	1015	1020
[5784]	Lys Arg Ser Ile Tyr Ser Asp Lys Lys Val Lys Lys Leu Lys Gln		
[5785]	1025	1030	1035
[5786]	Glu Lys Lys Asp Leu Tyr Ile Arg Asn Tyr Ile Ala His Phe Asn		
[5787]	1040	1045	1050
[5788]	Tyr Ile Pro His Ala Glu Ile Ser Leu Leu Glu Val Leu Glu Asn		
[5789]	1055	1060	1065
[5790]	Leu Arg Lys Leu Leu Ser Tyr Asp Arg Lys Leu Lys Asn Ala Ile		
[5791]	1070	1075	1080
[5792]	Met Lys Ser Ile Val Asp Ile Leu Lys Glu Tyr Gly Phe Val Ala		
[5793]	1085	1090	1095
[5794]	Thr Phe Lys Ile Gly Ala Asp Lys Lys Ile Glu Ile Gln Thr Leu		
[5795]	1100	1105	1110
[5796]	Glu Ser Glu Lys Ile Val His Leu Lys Asn Leu Lys Lys Lys Lys		
[5797]	1115	1120	1125
[5798]	Leu Met Thr Asp Arg Asn Ser Glu Glu Leu Cys Glu Leu Val Lys		
[5799]	1130	1135	1140
[5800]	Val Met Phe Glu Tyr Lys Ala Leu Glu		
[5801]	1145	1150	
[5802]	<210> 313		
[5803]	<211> 1197		
[5804]	<212> PRT		
[5805]	<213> 韦德纤毛菌		
[5806]	<400> 313		
[5807]	Met Lys Val Thr Lys Ile Asp Gly Leu Ser His Lys Lys Phe Glu Asp		
[5808]	1	5	10
[5809]	Glu Gly Lys Leu Val Lys Phe Arg Asn Asn Lys Asn Ile Asn Glu Ile		15

[5810]	20	25	30
[5811]	Lys Glu Arg Leu Lys Lys Leu Lys Glu Leu Lys Leu Asp Asn Tyr Ile		
[5812]	35	40	45
[5813]	Lys Asn Pro Glu Asn Val Lys Asn Lys Asp Lys Asp Ala Glu Lys Glu		
[5814]	50	55	60
[5815]	Thr Lys Ile Arg Arg Thr Asn Leu Lys Lys Tyr Phe Ser Glu Ile Ile		
[5816]	65	70	75
[5817]	Leu Arg Lys Glu Asp Glu Lys Tyr Ile Leu Lys Lys Thr Lys Lys Phe		
[5818]	85	90	95
[5819]	Lys Asp Ile Asn Gln Glu Ile Asp Tyr Tyr Asp Val Lys Ser Lys Lys		
[5820]	100	105	110
[5821]	Asn Gln Gln Glu Ile Phe Asp Val Leu Lys Glu Ile Leu Glu Leu Lys		
[5822]	115	120	125
[5823]	Ile Lys Glu Thr Glu Lys Glu Glu Ile Ile Thr Phe Asp Ser Glu Lys		
[5824]	130	135	140
[5825]	Leu Lys Lys Val Phe Gly Glu Asp Phe Val Lys Lys Glu Ala Lys Ile		
[5826]	145	150	155
[5827]	Lys Ala Ile Glu Lys Ser Leu Lys Ile Asn Lys Ala Asn Tyr Lys Lys		
[5828]	165	170	175
[5829]	Asp Ser Ile Lys Ile Gly Asp Asp Lys Tyr Ser Asn Val Lys Gly Glu		
[5830]	180	185	190
[5831]	Asn Lys Arg Ser Arg Ile Tyr Glu Tyr Tyr Lys Lys Ser Glu Asn Leu		
[5832]	195	200	205
[5833]	Lys Lys Phe Glu Glu Asn Ile Arg Glu Ala Phe Glu Lys Leu Tyr Thr		
[5834]	210	215	220
[5835]	Glu Glu Asn Ile Lys Glu Leu Tyr Ser Lys Ile Glu Glu Ile Leu Lys		
[5836]	225	230	235
[5837]	Lys Thr His Leu Lys Ser Ile Val Arg Glu Phe Tyr Gln Asn Glu Ile		
[5838]	245	250	255
[5839]	Ile Gly Glu Ser Glu Phe Ser Lys Lys Asn Gly Asp Gly Ile Ser Ile		
[5840]	260	265	270
[5841]	Leu Tyr Asn Gln Ile Lys Asp Ser Ile Lys Lys Glu Glu Asn Phe Ile		
[5842]	275	280	285
[5843]	Glu Phe Ile Glu Asn Thr Gly Asn Leu Glu Leu Lys Glu Leu Thr Lys		
[5844]	290	295	300
[5845]	Ser Gln Ile Phe Tyr Lys Tyr Phe Leu Glu Asn Glu Glu Leu Asn Asp		
[5846]	305	310	315
[5847]	Glu Asn Ile Lys Phe Ala Phe Cys Tyr Phe Val Glu Ile Glu Val Asn		
[5848]	325	330	335

[5849]	Asn	Leu	Leu	Lys	Glu	Asn	Val	Tyr	Lys	Ile	Lys	Arg	Phe	Asn	Glu	Ser
[5850]				340					345					350		
[5851]	Asn	Lys	Lys	Arg	Ile	Glu	Asn	Ile	Phe	Glu	Tyr	Gly	Lys	Leu	Lys	Lys
[5852]				355					360					365		
[5853]	Leu	Ile	Val	Tyr	Lys	Leu	Glu	Asn	Lys	Leu	Asn	Asn	Tyr	Val	Arg	Asn
[5854]			370					375					380			
[5855]	Cys	Gly	Lys	Tyr	Asn	Tyr	His	Met	Glu	Asn	Gly	Asp	Ile	Ala	Thr	Ser
[5856]	385					390					395					400
[5857]	Asp	Ile	Asn	Met	Arg	Asn	Arg	Gln	Thr	Glu	Ala	Phe	Leu	Arg	Ser	Ile
[5858]					405					410					415	
[5859]	Ile	Gly	Val	Ser	Ser	Phe	Gly	Tyr	Phe	Ser	Leu	Arg	Asn	Ile	Leu	Gly
[5860]				420					425					430		
[5861]	Val	Asn	Asp	Asp	Asp	Phe	Tyr	Glu	Thr	Glu	Glu	Asp	Leu	Thr	Lys	Lys
[5862]				435					440					445		
[5863]	Glu	Arg	Arg	Asn	Leu	Glu	Lys	Ala	Lys	Glu	Asp	Ile	Thr	Ile	Lys	Asn
[5864]		450					455					460				
[5865]	Thr	Phe	Asp	Glu	Val	Val	Val	Lys	Ser	Phe	Gln	Lys	Lys	Gly	Ile	Tyr
[5866]	465					470					475					480
[5867]	Asn	Ile	Lys	Glu	Asn	Leu	Lys	Met	Phe	Tyr	Gly	Asp	Ser	Phe	Asp	Asn
[5868]					485					490					495	
[5869]	Ala	Asp	Lys	Asp	Glu	Leu	Lys	Gln	Phe	Phe	Val	Asn	Met	Leu	Asn	Ala
[5870]				500					505					510		
[5871]	Ile	Thr	Ser	Ile	Arg	His	Arg	Val	Val	His	Tyr	Asn	Met	Asn	Thr	Asn
[5872]			515					520					525			
[5873]	Ser	Glu	Asn	Ile	Phe	Asn	Phe	Ser	Gly	Ile	Glu	Val	Ser	Lys	Leu	Leu
[5874]		530					535					540				
[5875]	Lys	Ser	Ile	Phe	Glu	Lys	Glu	Thr	Asp	Lys	Arg	Glu	Leu	Lys	Leu	Lys
[5876]	545					550					555					560
[5877]	Ile	Phe	Arg	Gln	Leu	Asn	Ser	Ala	Gly	Val	Phe	Asp	Tyr	Trp	Glu	Asn
[5878]				565					570				575			
[5879]	Arg	Lys	Ile	Asp	Lys	Tyr	Leu	Glu	Asn	Ile	Glu	Phe	Lys	Phe	Val	Asn
[5880]				580					585				590			
[5881]	Lys	Asn	Ile	Pro	Phe	Val	Pro	Ser	Phe	Thr	Lys	Leu	Tyr	Asn	Arg	Ile
[5882]			595					600					605			
[5883]	Asp	Asn	Leu	Lys	Gly	Asn	Asn	Ala	Leu	Asn	Leu	Gly	Tyr	Ile	Asn	Ile
[5884]		610						615				620				
[5885]	Pro	Lys	Arg	Lys	Glu	Ala	Arg	Asp	Ser	Gln	Ile	Tyr	Leu	Leu	Lys	Asn
[5886]	625					630					635					640
[5887]	Ile	Tyr	Tyr	Gly	Glu	Phe	Val	Glu	Lys	Phe	Val	Asn	Asn	Asn	Asp	Asn

[5888]	645							650							655				
[5889]	Phe	Glu	Lys	Ile	Phe	Arg	Glu	Ile	Ile	Glu	Ile	Asn	Lys	Lys	Asp	Gly			
[5890]	660							665							670				
[5891]	Thr	Asn	Thr	Lys	Thr	Lys	Phe	Tyr	Lys	Leu	Glu	Lys	Phe	Glu	Thr	Leu			
[5892]	675							680							685				
[5893]	Lys	Ala	Asn	Ala	Pro	Ile	Glu	Tyr	Leu	Glu	Lys	Leu	Gln	Ser	Leu	His			
[5894]	690							695							700				
[5895]	Gln	Ile	Asn	Tyr	Asn	Arg	Glu	Lys	Val	Glu	Glu	Asp	Lys	Asp	Ile	Tyr			
[5896]	705							710							715			720	
[5897]	Val	Asp	Phe	Val	Gln	Lys	Ile	Phe	Leu	Lys	Gly	Phe	Ile	Asn	Tyr	Leu			
[5898]	725							730							735				
[5899]	Gln	Gly	Ser	Asp	Leu	Leu	Lys	Ser	Leu	Asn	Leu	Leu	Asn	Leu	Lys	Lys			
[5900]	740							745							750				
[5901]	Asp	Glu	Ala	Ile	Ala	Asn	Lys	Lys	Ser	Phe	Tyr	Asp	Glu	Lys	Leu	Lys			
[5902]	755							760							765				
[5903]	Leu	Trp	Gln	Asn	Asn	Gly	Ser	Asn	Leu	Ser	Lys	Met	Pro	Glu	Glu	Ile			
[5904]	770							775							780				
[5905]	Tyr	Asp	Tyr	Ile	Lys	Lys	Ile	Lys	Ile	Asn	Lys	Ile	Asn	Tyr	Ser	Asp			
[5906]	785							790							795			800	
[5907]	Arg	Met	Ser	Ile	Phe	Tyr	Leu	Leu	Leu	Lys	Leu	Ile	Asp	His	Lys	Glu			
[5908]	805							810							815				
[5909]	Leu	Thr	Asn	Leu	Arg	Gly	Asn	Leu	Glu	Lys	Tyr	Val	Ser	Met	Asn	Lys			
[5910]	820							825							830				
[5911]	Asn	Lys	Ile	Tyr	Ser	Glu	Glu	Leu	Asn	Ile	Val	Asn	Leu	Val	Ser	Leu			
[5912]	835							840							845				
[5913]	Asp	Asn	Asn	Lys	Val	Arg	Ala	Asn	Phe	Asn	Leu	Lys	Pro	Glu	Asp	Ile			
[5914]	850							855							860				
[5915]	Gly	Lys	Phe	Leu	Lys	Thr	Glu	Thr	Ser	Ile	Arg	Asn	Ile	Asn	Gln	Leu			
[5916]	865							870							875			880	
[5917]	Asn	Asn	Phe	Ser	Glu	Ile	Phe	Ala	Asp	Gly	Glu	Asn	Val	Ile	Lys	His			
[5918]	885							890							895				
[5919]	Arg	Ser	Phe	Tyr	Asn	Ile	Lys	Lys	Tyr	Gly	Ile	Leu	Asp	Leu	Leu	Glu			
[5920]	900							905							910				
[5921]	Lys	Ile	Val	Asp	Lys	Ala	Asp	Leu	Lys	Ile	Thr	Lys	Glu	Glu	Ile	Lys			
[5922]	915							920							925				
[5923]	Lys	Tyr	Glu	Asn	Leu	Gln	Asn	Glu	Leu	Lys	Arg	Asn	Asp	Phe	Tyr	Lys			
[5924]	930							935							940				
[5925]	Ile	Gln	Glu	Arg	Ile	His	Arg	Asn	Tyr	Asn	Gln	Lys	Pro	Phe	Leu	Ile			
[5926]	945							950							955			960	

[5927]	Lys Asn Asn Glu Lys Asp Phe Asn Asp Tyr Lys Lys Ala Ile Glu Asn
[5928]	965 970 975
[5929]	Ile Gln Asn Tyr Thr Gln Leu Lys Asn Lys Ile Glu Phe Asn Asp Leu
[5930]	980 985 990
[5931]	Asn Leu Leu Gln Ser Leu Leu Phe Arg Ile Leu His Arg Leu Ala Gly
[5932]	995 1000 1005
[5933]	Tyr Thr Ser Leu Trp Glu Arg Asp Leu Gln Phe Lys Leu Lys Gly
[5934]	1010 1015 1020
[5935]	Glu Tyr Pro Glu Asn Lys Tyr Ile Asp Glu Ile Phe Asn Phe Asp
[5936]	1025 1030 1035
[5937]	Asn Ser Lys Asn Lys Ile Tyr Asn Glu Lys Asn Glu Arg Gly Gly
[5938]	1040 1045 1050
[5939]	Ser Val Val Ser Lys Tyr Gly Tyr Phe Leu Val Glu Lys Asp Gly
[5940]	1055 1060 1065
[5941]	Glu Ile Gln Arg Lys Asn Ala Arg Asp Lys Lys Lys Asn Lys Ile
[5942]	1070 1075 1080
[5943]	Ile Lys Lys Glu Gly Leu Glu Ile Arg Asn Tyr Ile Ala His Phe
[5944]	1085 1090 1095
[5945]	Asn Tyr Ile Pro Asp Ala Thr Lys Ser Ile Leu Glu Ile Leu Glu
[5946]	1100 1105 1110
[5947]	Glu Leu Arg Asn Leu Leu Lys Tyr Asp Arg Lys Leu Lys Asn Ala
[5948]	1115 1120 1125
[5949]	Val Met Lys Ser Ile Lys Asp Ile Phe Lys Glu Tyr Gly Leu Ile
[5950]	1130 1135 1140
[5951]	Ile Glu Phe Lys Ile Ser His Val Asn Asn Ser Glu Lys Ile Glu
[5952]	1145 1150 1155
[5953]	Val Leu Asn Val Asp Ser Glu Lys Ile Lys His Leu Lys Asn Asn
[5954]	1160 1165 1170
[5955]	Gly Leu Val Thr Thr Arg Asn Ser Glu Asp Leu Cys Glu Leu Ile
[5956]	1175 1180 1185
[5957]	Lys Met Met Leu Glu Tyr Lys Lys Ser
[5958]	1190 1195
[5959]	<210> 314
[5960]	<211> 1437
[5961]	<212> PRT
[5962]	<213> 未知的
[5963]	<220>
[5964]	<223> 毛螺菌科细菌NK4A179
[5965]	<400> 314

[5966]	Met	Lys	Ile	Ser	Lys	Val	Arg	Glu	Glu	Asn	Arg	Gly	Ala	Lys	Leu	Thr
[5967]	1				5					10					15	
[5968]	Val	Asn	Ala	Lys	Thr	Ala	Val	Val	Ser	Glu	Asn	Arg	Ser	Gln	Glu	Gly
[5969]					20					25					30	
[5970]	Ile	Leu	Tyr	Asn	Asp	Pro	Ser	Arg	Tyr	Gly	Lys	Ser	Arg	Lys	Asn	Asp
[5971]					35					40					45	
[5972]	Glu	Asp	Arg	Asp	Arg	Tyr	Ile	Glu	Ser	Arg	Leu	Lys	Ser	Ser	Gly	Lys
[5973]					50					55					60	
[5974]	Leu	Tyr	Arg	Ile	Phe	Asn	Glu	Asp	Lys	Asn	Lys	Arg	Glu	Thr	Asp	Glu
[5975]	65					70						75				80
[5976]	Leu	Gln	Trp	Phe	Leu	Ser	Glu	Ile	Val	Lys	Lys	Ile	Asn	Arg	Arg	Asn
[5977]					85					90						95
[5978]	Gly	Leu	Val	Leu	Ser	Asp	Met	Leu	Ser	Val	Asp	Asp	Arg	Ala	Phe	Glu
[5979]					100					105					110	
[5980]	Lys	Ala	Phe	Glu	Lys	Tyr	Ala	Glu	Leu	Ser	Tyr	Thr	Asn	Arg	Arg	Asn
[5981]					115					120					125	
[5982]	Lys	Val	Ser	Gly	Ser	Pro	Ala	Phe	Glu	Thr	Cys	Gly	Val	Asp	Ala	Ala
[5983]					130					135					140	
[5984]	Thr	Ala	Glu	Arg	Leu	Lys	Gly	Ile	Ile	Ser	Glu	Thr	Asn	Phe	Ile	Asn
[5985]	145					150						155				160
[5986]	Arg	Ile	Lys	Asn	Asn	Ile	Asp	Asn	Lys	Val	Ser	Glu	Asp	Ile	Ile	Asp
[5987]					165					170						175
[5988]	Arg	Ile	Ile	Ala	Lys	Tyr	Leu	Lys	Lys	Ser	Leu	Cys	Arg	Glu	Arg	Val
[5989]					180					185						190
[5990]	Lys	Arg	Gly	Leu	Lys	Lys	Leu	Leu	Met	Asn	Ala	Phe	Asp	Leu	Pro	Tyr
[5991]					195					200					205	
[5992]	Ser	Asp	Pro	Asp	Ile	Asp	Val	Gln	Arg	Asp	Phe	Ile	Asp	Tyr	Val	Leu
[5993]					210					215					220	
[5994]	Glu	Asp	Phe	Tyr	His	Val	Arg	Ala	Lys	Ser	Gln	Val	Ser	Arg	Ser	Ile
[5995]	225					230						235				240
[5996]	Lys	Asn	Met	Asn	Met	Pro	Val	Gln	Pro	Glu	Gly	Asp	Gly	Lys	Phe	Ala
[5997]					245					250						255
[5998]	Ile	Thr	Val	Ser	Lys	Gly	Gly	Thr	Glu	Ser	Gly	Asn	Lys	Arg	Ser	Ala
[5999]					260					265						270
[6000]	Glu	Lys	Glu	Ala	Phe	Lys	Lys	Phe	Leu	Ser	Asp	Tyr	Ala	Ser	Leu	Asp
[6001]					275					280					285	
[6002]	Glu	Arg	Val	Arg	Asp	Asp	Met	Leu	Arg	Arg	Met	Arg	Arg	Leu	Val	Val
[6003]					290					295					300	
[6004]	Leu	Tyr	Phe	Tyr	Gly	Ser	Asp	Asp	Ser	Lys	Leu	Ser	Asp	Val	Asn	Glu

[6005]	305	310	315	320
[6006]	Lys Phe Asp Val Trp Glu Asp His Ala Ala Arg Arg Val Asp Asn Arg			
[6007]		325	330	335
[6008]	Glu Phe Ile Lys Leu Pro Leu Glu Asn Lys Leu Ala Asn Gly Lys Thr			
[6009]		340	345	350
[6010]	Asp Lys Asp Ala Glu Arg Ile Arg Lys Asn Thr Val Lys Glu Leu Tyr			
[6011]		355	360	365
[6012]	Arg Asn Gln Asn Ile Gly Cys Tyr Arg Gln Ala Val Lys Ala Val Glu			
[6013]		370	375	380
[6014]	Glu Asp Asn Asn Gly Arg Tyr Phe Asp Asp Lys Met Leu Asn Met Phe			
[6015]	385	390	395	400
[6016]	Phe Ile His Arg Ile Glu Tyr Gly Val Glu Lys Ile Tyr Ala Asn Leu			
[6017]		405	410	415
[6018]	Lys Gln Val Thr Glu Phe Lys Ala Arg Thr Gly Tyr Leu Ser Glu Lys			
[6019]		420	425	430
[6020]	Ile Trp Lys Asp Leu Ile Asn Tyr Ile Ser Ile Lys Tyr Ile Ala Met			
[6021]		435	440	445
[6022]	Gly Lys Ala Val Tyr Asn Tyr Ala Met Asp Glu Leu Asn Ala Ser Asp			
[6023]		450	455	460
[6024]	Lys Lys Glu Ile Glu Leu Gly Lys Ile Ser Glu Glu Tyr Leu Ser Gly			
[6025]	465	470	475	480
[6026]	Ile Ser Ser Phe Asp Tyr Glu Leu Ile Lys Ala Glu Glu Met Leu Gln			
[6027]		485	490	495
[6028]	Arg Glu Thr Ala Val Tyr Val Ala Phe Ala Ala Arg His Leu Ser Ser			
[6029]		500	505	510
[6030]	Gln Thr Val Glu Leu Asp Ser Glu Asn Ser Asp Phe Leu Leu Leu Lys			
[6031]		515	520	525
[6032]	Pro Lys Gly Thr Met Asp Lys Asn Asp Lys Asn Lys Leu Ala Ser Asn			
[6033]		530	535	540
[6034]	Asn Ile Leu Asn Phe Leu Lys Asp Lys Glu Thr Leu Arg Asp Thr Ile			
[6035]	545	550	555	560
[6036]	Leu Gln Tyr Phe Gly Gly His Ser Leu Trp Thr Asp Phe Pro Phe Asp			
[6037]		565	570	575
[6038]	Lys Tyr Leu Ala Gly Gly Lys Asp Asp Val Asp Phe Leu Thr Asp Leu			
[6039]		580	585	590
[6040]	Lys Asp Val Ile Tyr Ser Met Arg Asn Asp Ser Phe His Tyr Ala Thr			
[6041]		595	600	605
[6042]	Glu Asn His Asn Asn Gly Lys Trp Asn Lys Glu Leu Ile Ser Ala Met			
[6043]		610	615	620

[6044]	Phe	Glu	His	Glu	Thr	Glu	Arg	Met	Thr	Val	Val	Met	Lys	Asp	Lys	Phe
[6045]	625					630				635						640
[6046]	Tyr	Ser	Asn	Asn	Leu	Pro	Met	Phe	Tyr	Lys	Asn	Asp	Asp	Leu	Lys	Lys
[6047]					645					650						655
[6048]	Leu	Leu	Ile	Asp	Leu	Tyr	Lys	Asp	Asn	Val	Glu	Arg	Ala	Ser	Gln	Val
[6049]				660					665						670	
[6050]	Pro	Ser	Phe	Asn	Lys	Val	Phe	Val	Arg	Lys	Asn	Phe	Pro	Ala	Leu	Val
[6051]			675					680					685			
[6052]	Arg	Asp	Lys	Asp	Asn	Leu	Gly	Ile	Glu	Leu	Asp	Leu	Lys	Ala	Asp	Ala
[6053]		690					695					700				
[6054]	Asp	Lys	Gly	Glu	Asn	Glu	Leu	Lys	Phe	Tyr	Asn	Ala	Leu	Tyr	Tyr	Met
[6055]	705					710				715						720
[6056]	Phe	Lys	Glu	Ile	Tyr	Tyr	Asn	Ala	Phe	Leu	Asn	Asp	Lys	Asn	Val	Arg
[6057]				725						730						735
[6058]	Glu	Arg	Phe	Ile	Thr	Lys	Ala	Thr	Lys	Val	Ala	Asp	Asn	Tyr	Asp	Arg
[6059]				740						745					750	
[6060]	Asn	Lys	Glu	Arg	Asn	Leu	Lys	Asp	Arg	Ile	Lys	Ser	Ala	Gly	Ser	Asp
[6061]			755					760					765			
[6062]	Glu	Lys	Lys	Lys	Leu	Arg	Glu	Gln	Leu	Gln	Asn	Tyr	Ile	Ala	Glu	Asn
[6063]		770					775					780				
[6064]	Asp	Phe	Gly	Gln	Arg	Ile	Lys	Asn	Ile	Val	Gln	Val	Asn	Pro	Asp	Tyr
[6065]	785					790					795					800
[6066]	Thr	Leu	Ala	Gln	Ile	Cys	Gln	Leu	Ile	Met	Thr	Glu	Tyr	Asn	Gln	Gln
[6067]				805						810						815
[6068]	Asn	Asn	Gly	Cys	Met	Gln	Lys	Lys	Ser	Ala	Ala	Arg	Lys	Asp	Ile	Asn
[6069]			820						825					830		
[6070]	Lys	Asp	Ser	Tyr	Gln	His	Tyr	Lys	Met	Leu	Leu	Leu	Val	Asn	Leu	Arg
[6071]			835					840					845			
[6072]	Lys	Ala	Phe	Leu	Glu	Phe	Ile	Lys	Glu	Asn	Tyr	Ala	Phe	Val	Leu	Lys
[6073]		850					855					860				
[6074]	Pro	Tyr	Lys	His	Asp	Leu	Cys	Asp	Lys	Ala	Asp	Phe	Val	Pro	Asp	Phe
[6075]	865					870					875					880
[6076]	Ala	Lys	Tyr	Val	Lys	Pro	Tyr	Ala	Gly	Leu	Ile	Ser	Arg	Val	Ala	Gly
[6077]				885						890						895
[6078]	Ser	Ser	Glu	Leu	Gln	Lys	Trp	Tyr	Ile	Val	Ser	Arg	Phe	Leu	Ser	Pro
[6079]				900					905					910		
[6080]	Ala	Gln	Ala	Asn	His	Met	Leu	Gly	Phe	Leu	His	Ser	Tyr	Lys	Gln	Tyr
[6081]			915					920					925			
[6082]	Val	Trp	Asp	Ile	Tyr	Arg	Arg	Ala	Ser	Glu	Thr	Gly	Thr	Glu	Ile	Asn

[6083]	930	935	940
[6084]	His Ser Ile Ala Glu Asp Lys Ile Ala Gly Val Asp Ile Thr Asp Val		
[6085]	945	950	955 960
[6086]	Asp Ala Val Ile Asp Leu Ser Val Lys Leu Cys Gly Thr Ile Ser Ser		
[6087]	965	970	975
[6088]	Glu Ile Ser Asp Tyr Phe Lys Asp Asp Glu Val Tyr Ala Glu Tyr Ile		
[6089]	980	985	990
[6090]	Ser Ser Tyr Leu Asp Phe Glu Tyr Asp Gly Gly Asn Tyr Lys Asp Ser		
[6091]	995	1000	1005
[6092]	Leu Asn Arg Phe Cys Asn Ser Asp Ala Val Asn Asp Gln Lys Val		
[6093]	1010	1015	1020
[6094]	Ala Leu Tyr Tyr Asp Gly Glu His Pro Lys Leu Asn Arg Asn Ile		
[6095]	1025	1030	1035
[6096]	Ile Leu Ser Lys Leu Tyr Gly Glu Arg Arg Phe Leu Glu Lys Ile		
[6097]	1040	1045	1050
[6098]	Thr Asp Arg Val Ser Arg Ser Asp Ile Val Glu Tyr Tyr Lys Leu		
[6099]	1055	1060	1065
[6100]	Lys Lys Glu Thr Ser Gln Tyr Gln Thr Lys Gly Ile Phe Asp Ser		
[6101]	1070	1075	1080
[6102]	Glu Asp Glu Gln Lys Asn Ile Lys Lys Phe Gln Glu Met Lys Asn		
[6103]	1085	1090	1095
[6104]	Ile Val Glu Phe Arg Asp Leu Met Asp Tyr Ser Glu Ile Ala Asp		
[6105]	1100	1105	1110
[6106]	Glu Leu Gln Gly Gln Leu Ile Asn Trp Ile Tyr Leu Arg Glu Arg		
[6107]	1115	1120	1125
[6108]	Asp Leu Met Asn Phe Gln Leu Gly Tyr His Tyr Ala Cys Leu Asn		
[6109]	1130	1135	1140
[6110]	Asn Asp Ser Asn Lys Gln Ala Thr Tyr Val Thr Leu Asp Tyr Gln		
[6111]	1145	1150	1155
[6112]	Gly Lys Lys Asn Arg Lys Ile Asn Gly Ala Ile Leu Tyr Gln Ile		
[6113]	1160	1165	1170
[6114]	Cys Ala Met Tyr Ile Asn Gly Leu Pro Leu Tyr Tyr Val Asp Lys		
[6115]	1175	1180	1185
[6116]	Asp Ser Ser Glu Trp Thr Val Ser Asp Gly Lys Glu Ser Thr Gly		
[6117]	1190	1195	1200
[6118]	Ala Lys Ile Gly Glu Phe Tyr Arg Tyr Ala Lys Ser Phe Glu Asn		
[6119]	1205	1210	1215
[6120]	Thr Ser Asp Cys Tyr Ala Ser Gly Leu Glu Ile Phe Glu Asn Ile		
[6121]	1220	1225	1230

[6122]	Ser Glu His Asp Asn Ile Thr Glu Leu Arg Asn Tyr Ile Glu His
[6123]	1235 1240 1245
[6124]	Phe Arg Tyr Tyr Ser Ser Phe Asp Arg Ser Phe Leu Gly Ile Tyr
[6125]	1250 1255 1260
[6126]	Ser Glu Val Phe Asp Arg Phe Phe Thr Tyr Asp Leu Lys Tyr Arg
[6127]	1265 1270 1275
[6128]	Lys Asn Val Pro Thr Ile Leu Tyr Asn Ile Leu Leu Gln His Phe
[6129]	1280 1285 1290
[6130]	Val Asn Val Arg Phe Glu Phe Val Ser Gly Lys Lys Met Ile Gly
[6131]	1295 1300 1305
[6132]	Ile Asp Lys Lys Asp Arg Lys Ile Ala Lys Glu Lys Glu Cys Ala
[6133]	1310 1315 1320
[6134]	Arg Ile Thr Ile Arg Glu Lys Asn Gly Val Tyr Ser Glu Gln Phe
[6135]	1325 1330 1335
[6136]	Thr Tyr Lys Leu Lys Asn Gly Thr Val Tyr Val Asp Ala Arg Asp
[6137]	1340 1345 1350
[6138]	Lys Arg Tyr Leu Gln Ser Ile Ile Arg Leu Leu Phe Tyr Pro Glu
[6139]	1355 1360 1365
[6140]	Lys Val Asn Met Asp Glu Met Ile Glu Val Lys Glu Lys Lys Lys
[6141]	1370 1375 1380
[6142]	Pro Ser Asp Asn Asn Thr Gly Lys Gly Tyr Ser Lys Arg Asp Arg
[6143]	1385 1390 1395
[6144]	Gln Gln Asp Arg Lys Glu Tyr Asp Lys Tyr Lys Glu Lys Lys Lys
[6145]	1400 1405 1410
[6146]	Lys Glu Gly Asn Phe Leu Ser Gly Met Gly Gly Asn Ile Asn Trp
[6147]	1415 1420 1425
[6148]	Asp Glu Ile Asn Ala Gln Leu Lys Asn
[6149]	1430 1435
[6150]	<210> 315
[6151]	<211> 1285
[6152]	<212> PRT
[6153]	<213> 荚膜红细菌R121
[6154]	<400> 315
[6155]	Met Gln Ile Gly Lys Val Gln Gly Arg Thr Ile Ser Glu Phe Gly Asp
[6156]	1 5 10 15
[6157]	Pro Ala Gly Gly Leu Lys Arg Lys Ile Ser Thr Asp Gly Lys Asn Arg
[6158]	20 25 30
[6159]	Lys Glu Leu Pro Ala His Leu Ser Ser Asp Pro Lys Ala Leu Ile Gly
[6160]	35 40 45

[6161]	Gln Trp Ile Ser Gly Ile Asp Lys Ile Tyr Arg Lys Pro Asp Ser Arg
[6162]	50 55 60
[6163]	Lys Ser Asp Gly Lys Ala Ile His Ser Pro Thr Pro Ser Lys Met Gln
[6164]	65 70 75 80
[6165]	Phe Asp Ala Arg Asp Asp Leu Gly Glu Ala Phe Trp Lys Leu Val Ser
[6166]	85 90 95
[6167]	Glu Ala Gly Leu Ala Gln Asp Ser Asp Tyr Asp Gln Phe Lys Arg Arg
[6168]	100 105 110
[6169]	Leu His Pro Tyr Gly Asp Lys Phe Gln Pro Ala Asp Ser Gly Ala Lys
[6170]	115 120 125
[6171]	Leu Lys Phe Glu Ala Asp Pro Pro Glu Pro Gln Ala Phe His Gly Arg
[6172]	130 135 140
[6173]	Trp Tyr Gly Ala Met Ser Lys Arg Gly Asn Asp Ala Lys Glu Leu Ala
[6174]	145 150 155 160
[6175]	Ala Ala Leu Tyr Glu His Leu His Val Asp Glu Lys Arg Ile Asp Gly
[6176]	165 170 175
[6177]	Gln Pro Lys Arg Asn Pro Lys Thr Asp Lys Phe Ala Pro Gly Leu Val
[6178]	180 185 190
[6179]	Val Ala Arg Ala Leu Gly Ile Glu Ser Ser Val Leu Pro Arg Gly Met
[6180]	195 200 205
[6181]	Ala Arg Leu Ala Arg Asn Trp Gly Glu Glu Glu Ile Gln Thr Tyr Phe
[6182]	210 215 220
[6183]	Val Val Asp Val Ala Ala Ser Val Lys Glu Val Ala Lys Ala Ala Val
[6184]	225 230 235 240
[6185]	Ser Ala Ala Gln Ala Phe Asp Pro Pro Arg Gln Val Ser Gly Arg Ser
[6186]	245 250 255
[6187]	Leu Ser Pro Lys Val Gly Phe Ala Leu Ala Glu His Leu Glu Arg Val
[6188]	260 265 270
[6189]	Thr Gly Ser Lys Arg Cys Ser Phe Asp Pro Ala Ala Gly Pro Ser Val
[6190]	275 280 285
[6191]	Leu Ala Leu His Asp Glu Val Lys Lys Thr Tyr Lys Arg Leu Cys Ala
[6192]	290 295 300
[6193]	Arg Gly Lys Asn Ala Ala Arg Ala Phe Pro Ala Asp Lys Thr Glu Leu
[6194]	305 310 315 320
[6195]	Leu Ala Leu Met Arg His Thr His Glu Asn Arg Val Arg Asn Gln Met
[6196]	325 330 335
[6197]	Val Arg Met Gly Arg Val Ser Glu Tyr Arg Gly Gln Gln Ala Gly Asp
[6198]	340 345 350
[6199]	Leu Ala Gln Ser His Tyr Trp Thr Ser Ala Gly Gln Thr Glu Ile Lys

[6200]	355	360	365
[6201]	Glu Ser Glu Ile Phe Val Arg Leu Trp Val Gly Ala Phe Ala Leu Ala		
[6202]	370	375	380
[6203]	Gly Arg Ser Met Lys Ala Trp Ile Asp Pro Met Gly Lys Ile Val Asn		
[6204]	385	390	395
[6205]	Thr Glu Lys Asn Asp Arg Asp Leu Thr Ala Ala Val Asn Ile Arg Gln		
[6206]	405	410	415
[6207]	Val Ile Ser Asn Lys Glu Met Val Ala Glu Ala Met Ala Arg Arg Gly		
[6208]	420	425	430
[6209]	Ile Tyr Phe Gly Glu Thr Pro Glu Leu Asp Arg Leu Gly Ala Glu Gly		
[6210]	435	440	445
[6211]	Asn Glu Gly Phe Val Phe Ala Leu Leu Arg Tyr Leu Arg Gly Cys Arg		
[6212]	450	455	460
[6213]	Asn Gln Thr Phe His Leu Gly Ala Arg Ala Gly Phe Leu Lys Glu Ile		
[6214]	465	470	475
[6215]	Arg Lys Glu Leu Glu Lys Thr Arg Trp Gly Lys Ala Lys Glu Ala Glu		
[6216]	485	490	495
[6217]	His Val Val Leu Thr Asp Lys Thr Val Ala Ala Ile Arg Ala Ile Ile		
[6218]	500	505	510
[6219]	Asp Asn Asp Ala Lys Ala Leu Gly Ala Arg Leu Leu Ala Asp Leu Ser		
[6220]	515	520	525
[6221]	Gly Ala Phe Val Ala His Tyr Ala Ser Lys Glu His Phe Ser Thr Leu		
[6222]	530	535	540
[6223]	Tyr Ser Glu Ile Val Lys Ala Val Lys Asp Ala Pro Glu Val Ser Ser		
[6224]	545	550	555
[6225]	Gly Leu Pro Arg Leu Lys Leu Leu Leu Lys Arg Ala Asp Gly Val Arg		
[6226]	565	570	575
[6227]	Gly Tyr Val His Gly Leu Arg Asp Thr Arg Lys His Ala Phe Ala Thr		
[6228]	580	585	590
[6229]	Lys Leu Pro Pro Pro Pro Ala Pro Arg Glu Leu Asp Asp Pro Ala Thr		
[6230]	595	600	605
[6231]	Lys Ala Arg Tyr Ile Ala Leu Leu Arg Leu Tyr Asp Gly Pro Phe Arg		
[6232]	610	615	620
[6233]	Ala Tyr Ala Ser Gly Ile Thr Gly Thr Ala Leu Ala Gly Pro Ala Ala		
[6234]	625	630	635
[6235]	Arg Ala Lys Glu Ala Ala Thr Ala Leu Ala Gln Ser Val Asn Val Thr		
[6236]	645	650	655
[6237]	Lys Ala Tyr Ser Asp Val Met Glu Gly Arg Ser Ser Arg Leu Arg Pro		
[6238]	660	665	670

[6239]	Pro	Asn	Asp	Gly	Glu	Thr	Leu	Arg	Glu	Tyr	Leu	Ser	Ala	Leu	Thr	Gly
[6240]			675					680					685			
[6241]	Glu	Thr	Ala	Thr	Glu	Phe	Arg	Val	Gln	Ile	Gly	Tyr	Glu	Ser	Asp	Ser
[6242]			690					695					700			
[6243]	Glu	Asn	Ala	Arg	Lys	Gln	Ala	Glu	Phe	Ile	Glu	Asn	Tyr	Arg	Arg	Asp
[6244]	705						710					715				720
[6245]	Met	Leu	Ala	Phe	Met	Phe	Glu	Asp	Tyr	Ile	Arg	Ala	Lys	Gly	Phe	Asp
[6246]						725					730				735	
[6247]	Trp	Ile	Leu	Lys	Ile	Glu	Pro	Gly	Ala	Thr	Ala	Met	Thr	Arg	Ala	Pro
[6248]						740					745				750	
[6249]	Val	Leu	Pro	Glu	Pro	Ile	Asp	Thr	Arg	Gly	Gln	Tyr	Glu	His	Trp	Gln
[6250]						755							760		765	
[6251]	Ala	Ala	Leu	Tyr	Leu	Val	Met	His	Phe	Val	Pro	Ala	Ser	Asp	Val	Ser
[6252]						770							775		780	
[6253]	Asn	Leu	Leu	His	Gln	Leu	Arg	Lys	Trp	Glu	Ala	Leu	Gln	Gly	Lys	Tyr
[6254]	785						790						795			800
[6255]	Glu	Leu	Val	Gln	Asp	Gly	Asp	Ala	Thr	Asp	Gln	Ala	Asp	Ala	Arg	Arg
[6256]						805					810				815	
[6257]	Glu	Ala	Leu	Asp	Leu	Val	Lys	Arg	Phe	Arg	Asp	Val	Leu	Val	Leu	Phe
[6258]						820					825				830	
[6259]	Leu	Lys	Thr	Gly	Glu	Ala	Arg	Phe	Glu	Gly	Arg	Ala	Ala	Pro	Phe	Asp
[6260]						835					840				845	
[6261]	Leu	Lys	Pro	Phe	Arg	Ala	Leu	Phe	Ala	Asn	Pro	Ala	Thr	Phe	Asp	Arg
[6262]						850					855				860	
[6263]	Leu	Phe	Met	Ala	Thr	Pro	Thr	Thr	Ala	Arg	Pro	Ala	Glu	Asp	Asp	Pro
[6264]	865						870					875				880
[6265]	Glu	Gly	Asp	Gly	Ala	Ser	Glu	Pro	Glu	Leu	Arg	Val	Ala	Arg	Thr	Leu
[6266]						885					890				895	
[6267]	Arg	Gly	Leu	Arg	Gln	Ile	Ala	Arg	Tyr	Asn	His	Met	Ala	Val	Leu	Ser
[6268]						900					905				910	
[6269]	Asp	Leu	Phe	Ala	Lys	His	Lys	Val	Arg	Asp	Glu	Glu	Val	Ala	Arg	Leu
[6270]						915					920				925	
[6271]	Ala	Glu	Ile	Glu	Asp	Glu	Thr	Gln	Glu	Lys	Ser	Gln	Ile	Val	Ala	Ala
[6272]						930						935			940	
[6273]	Gln	Glu	Leu	Arg	Thr	Asp	Leu	His	Asp	Lys	Val	Met	Lys	Cys	His	Pro
[6274]	945						950					955				960
[6275]	Lys	Thr	Ile	Ser	Pro	Glu	Glu	Arg	Gln	Ser	Tyr	Ala	Ala	Ala	Ile	Lys
[6276]						965					970				975	
[6277]	Thr	Ile	Glu	Glu	His	Arg	Phe	Leu	Val	Gly	Arg	Val	Tyr	Leu	Gly	Asp

[6278]	980						985						990					
[6279]	His	Leu	Arg	Leu	His	Arg	Leu	Met	Met	Asp	Val	Ile	Gly	Arg	Leu	Ile		
[6280]	995						1000						1005					
[6281]	Asp	Tyr	Ala	Gly	Ala	Tyr	Glu	Arg	Asp	Thr	Gly	Thr	Phe	Leu	Ile			
[6282]	1010						1015						1020					
[6283]	Asn	Ala	Ser	Lys	Gln	Leu	Gly	Ala	Gly	Ala	Asp	Trp	Ala	Val	Thr			
[6284]	1025						1030						1035					
[6285]	Ile	Ala	Gly	Ala	Ala	Asn	Thr	Asp	Ala	Arg	Thr	Gln	Thr	Arg	Lys			
[6286]	1040						1045						1050					
[6287]	Asp	Leu	Ala	His	Phe	Asn	Val	Leu	Asp	Arg	Ala	Asp	Gly	Thr	Pro			
[6288]	1055						1060						1065					
[6289]	Asp	Leu	Thr	Ala	Leu	Val	Asn	Arg	Ala	Arg	Glu	Met	Met	Ala	Tyr			
[6290]	1070						1075						1080					
[6291]	Asp	Arg	Lys	Arg	Lys	Asn	Ala	Val	Pro	Arg	Ser	Ile	Leu	Asp	Met			
[6292]	1085						1090						1095					
[6293]	Leu	Ala	Arg	Leu	Gly	Leu	Thr	Leu	Lys	Trp	Gln	Met	Lys	Asp	His			
[6294]	1100						1105						1110					
[6295]	Leu	Leu	Gln	Asp	Ala	Thr	Ile	Thr	Gln	Ala	Ala	Ile	Lys	His	Leu			
[6296]	1115						1120						1125					
[6297]	Asp	Lys	Val	Arg	Leu	Thr	Val	Gly	Gly	Pro	Ala	Ala	Val	Thr	Glu			
[6298]	1130						1135						1140					
[6299]	Ala	Arg	Phe	Ser	Gln	Asp	Tyr	Leu	Gln	Met	Val	Ala	Ala	Val	Phe			
[6300]	1145						1150						1155					
[6301]	Asn	Gly	Ser	Val	Gln	Asn	Pro	Lys	Pro	Arg	Arg	Arg	Asp	Asp	Gly			
[6302]	1160						1165						1170					
[6303]	Asp	Ala	Trp	His	Lys	Pro	Pro	Lys	Pro	Ala	Thr	Ala	Gln	Ser	Gln			
[6304]	1175						1180						1185					
[6305]	Pro	Asp	Gln	Lys	Pro	Pro	Asn	Lys	Ala	Pro	Ser	Ala	Gly	Ser	Arg			
[6306]	1190						1195						1200					
[6307]	Leu	Pro	Pro	Pro	Gln	Val	Gly	Glu	Val	Tyr	Glu	Gly	Val	Val	Val			
[6308]	1205						1210						1215					
[6309]	Lys	Val	Ile	Asp	Thr	Gly	Ser	Leu	Gly	Phe	Leu	Ala	Val	Glu	Gly			
[6310]	1220						1225						1230					
[6311]	Val	Ala	Gly	Asn	Ile	Gly	Leu	His	Ile	Ser	Arg	Leu	Arg	Arg	Ile			
[6312]	1235						1240						1245					
[6313]	Arg	Glu	Asp	Ala	Ile	Ile	Val	Gly	Arg	Arg	Tyr	Arg	Phe	Arg	Val			
[6314]	1250						1255						1260					
[6315]	Glu	Ile	Tyr	Val	Pro	Pro	Lys	Ser	Asn	Thr	Ser	Lys	Leu	Asn	Ala			
[6316]	1265						1270						1275					

[6317]	Ala Asp Leu Val Arg Ile Asp		
[6318]	1280	1285	
[6319]	<210> 316		
[6320]	<211> 1334		
[6321]	<212> PRT		
[6322]	<213> 未知的		
[6323]	<220>		
[6324]	<223> 毛螺菌科细菌NK4A144		
[6325]	<400> 316		
[6326]	Met Lys Ile Ser Lys Val Asp His Thr Arg Met Ala Val Ala Lys Gly		
[6327]	1	5	10 15
[6328]	Asn Gln His Arg Arg Asp Glu Ile Ser Gly Ile Leu Tyr Lys Asp Pro		
[6329]	20	25	30
[6330]	Thr Lys Thr Gly Ser Ile Asp Phe Asp Glu Arg Phe Lys Lys Leu Asn		
[6331]	35	40	45
[6332]	Cys Ser Ala Lys Ile Leu Tyr His Val Phe Asn Gly Ile Ala Glu Gly		
[6333]	50	55	60
[6334]	Ser Asn Lys Tyr Lys Asn Ile Val Asp Lys Val Asn Asn Asn Leu Asp		
[6335]	65	70	75 80
[6336]	Arg Val Leu Phe Thr Gly Lys Ser Tyr Asp Arg Lys Ser Ile Ile Asp		
[6337]	85	90	95
[6338]	Ile Asp Thr Val Leu Arg Asn Val Glu Lys Ile Asn Ala Phe Asp Arg		
[6339]	100	105	110
[6340]	Ile Ser Thr Glu Glu Arg Glu Gln Ile Ile Asp Asp Leu Leu Glu Ile		
[6341]	115	120	125
[6342]	Gln Leu Arg Lys Gly Leu Arg Lys Gly Lys Ala Gly Leu Arg Glu Val		
[6343]	130	135	140
[6344]	Leu Leu Ile Gly Ala Gly Val Ile Val Arg Thr Asp Lys Lys Gln Glu		
[6345]	145	150	155 160
[6346]	Ile Ala Asp Phe Leu Glu Ile Leu Asp Glu Asp Phe Asn Lys Thr Asn		
[6347]	165	170	175
[6348]	Gln Ala Lys Asn Ile Lys Leu Ser Ile Glu Asn Gln Gly Leu Val Val		
[6349]	180	185	190
[6350]	Ser Pro Val Ser Arg Gly Glu Glu Arg Ile Phe Asp Val Ser Gly Ala		
[6351]	195	200	205
[6352]	Gln Lys Gly Lys Ser Ser Lys Lys Ala Gln Glu Lys Glu Ala Leu Ser		
[6353]	210	215	220
[6354]	Ala Phe Leu Leu Asp Tyr Ala Asp Leu Asp Lys Asn Val Arg Phe Glu		
[6355]	225	230	235 240

[6356]	Tyr	Leu	Arg	Lys	Ile	Arg	Arg	Leu	Ile	Asn	Leu	Tyr	Phe	Tyr	Val	Lys
[6357]					245					250					255	
[6358]	Asn	Asp	Asp	Val	Met	Ser	Leu	Thr	Glu	Ile	Pro	Ala	Glu	Val	Asn	Leu
[6359]					260					265					270	
[6360]	Glu	Lys	Asp	Phe	Asp	Ile	Trp	Arg	Asp	His	Glu	Gln	Arg	Lys	Glu	Glu
[6361]					275					280					285	
[6362]	Asn	Gly	Asp	Phe	Val	Gly	Cys	Pro	Asp	Ile	Leu	Leu	Ala	Asp	Arg	Asp
[6363]					290					295					300	
[6364]	Val	Lys	Lys	Ser	Asn	Ser	Lys	Gln	Val	Lys	Ile	Ala	Glu	Arg	Gln	Leu
[6365]	305							310				315				320
[6366]	Arg	Glu	Ser	Ile	Arg	Glu	Lys	Asn	Ile	Lys	Arg	Tyr	Arg	Phe	Ser	Ile
[6367]					325					330					335	
[6368]	Lys	Thr	Ile	Glu	Lys	Asp	Asp	Gly	Thr	Tyr	Phe	Phe	Ala	Asn	Lys	Gln
[6369]					340					345					350	
[6370]	Ile	Ser	Val	Phe	Trp	Ile	His	Arg	Ile	Glu	Asn	Ala	Val	Glu	Arg	Ile
[6371]					355					360					365	
[6372]	Leu	Gly	Ser	Ile	Asn	Asp	Lys	Lys	Leu	Tyr	Arg	Leu	Arg	Leu	Gly	Tyr
[6373]					370					375				380		
[6374]	Leu	Gly	Glu	Lys	Val	Trp	Lys	Asp	Ile	Leu	Asn	Phe	Leu	Ser	Ile	Lys
[6375]	385							390				395				400
[6376]	Tyr	Ile	Ala	Val	Gly	Lys	Ala	Val	Phe	Asn	Phe	Ala	Met	Asp	Asp	Leu
[6377]					405					410					415	
[6378]	Gln	Glu	Lys	Asp	Arg	Asp	Ile	Glu	Pro	Gly	Lys	Ile	Ser	Glu	Asn	Ala
[6379]					420					425					430	
[6380]	Val	Asn	Gly	Leu	Thr	Ser	Phe	Asp	Tyr	Glu	Gln	Ile	Lys	Ala	Asp	Glu
[6381]					435					440					445	
[6382]	Met	Leu	Gln	Arg	Glu	Val	Ala	Val	Asn	Val	Ala	Phe	Ala	Ala	Asn	Asn
[6383]					450					455					460	
[6384]	Leu	Ala	Arg	Val	Thr	Val	Asp	Ile	Pro	Gln	Asn	Gly	Glu	Lys	Glu	Asp
[6385]	465							470				475				480
[6386]	Ile	Leu	Leu	Trp	Asn	Lys	Ser	Asp	Ile	Lys	Lys	Tyr	Lys	Lys	Asn	Ser
[6387]					485					490					495	
[6388]	Lys	Lys	Gly	Ile	Leu	Lys	Ser	Ile	Leu	Gln	Phe	Phe	Gly	Gly	Ala	Ser
[6389]					500					505					510	
[6390]	Thr	Trp	Asn	Met	Lys	Met	Phe	Glu	Ile	Ala	Tyr	His	Asp	Gln	Pro	Gly
[6391]					515					520				525		
[6392]	Asp	Tyr	Glu	Glu	Asn	Tyr	Leu	Tyr	Asp	Ile	Ile	Gln	Ile	Ile	Tyr	Ser
[6393]					530					535				540		
[6394]	Leu	Arg	Asn	Lys	Ser	Phe	His	Phe	Lys	Thr	Tyr	Asp	His	Gly	Asp	Lys

[6395]	545	550	555	560
[6396]	Asn Trp Asn Arg Glu Leu Ile Gly Lys Met Ile Glu His Asp Ala Glu			
[6397]	565	570	575	
[6398]	Arg Val Ile Ser Val Glu Arg Glu Lys Phe His Ser Asn Asn Leu Pro			
[6399]	580	585	590	
[6400]	Met Phe Tyr Lys Asp Ala Asp Leu Lys Lys Ile Leu Asp Leu Leu Tyr			
[6401]	595	600	605	
[6402]	Ser Asp Tyr Ala Gly Arg Ala Ser Gln Val Pro Ala Phe Asn Thr Val			
[6403]	610	615	620	
[6404]	Leu Val Arg Lys Asn Phe Pro Glu Phe Leu Arg Lys Asp Met Gly Tyr			
[6405]	625	630	635	640
[6406]	Lys Val His Phe Asn Asn Pro Glu Val Glu Asn Gln Trp His Ser Ala			
[6407]	645	650	655	
[6408]	Val Tyr Tyr Leu Tyr Lys Glu Ile Tyr Tyr Asn Leu Phe Leu Arg Asp			
[6409]	660	665	670	
[6410]	Lys Glu Val Lys Asn Leu Phe Tyr Thr Ser Leu Lys Asn Ile Arg Ser			
[6411]	675	680	685	
[6412]	Glu Val Ser Asp Lys Lys Gln Lys Leu Ala Ser Asp Asp Phe Ala Ser			
[6413]	690	695	700	
[6414]	Arg Cys Glu Glu Ile Glu Asp Arg Ser Leu Pro Glu Ile Cys Gln Ile			
[6415]	705	710	715	720
[6416]	Ile Met Thr Glu Tyr Asn Ala Gln Asn Phe Gly Asn Arg Lys Val Lys			
[6417]	725	730	735	
[6418]	Ser Gln Arg Val Ile Glu Lys Asn Lys Asp Ile Phe Arg His Tyr Lys			
[6419]	740	745	750	
[6420]	Met Leu Leu Ile Lys Thr Leu Ala Gly Ala Phe Ser Leu Tyr Leu Lys			
[6421]	755	760	765	
[6422]	Gln Glu Arg Phe Ala Phe Ile Gly Lys Ala Thr Pro Ile Pro Tyr Glu			
[6423]	770	775	780	
[6424]	Thr Thr Asp Val Lys Asn Phe Leu Pro Glu Trp Lys Ser Gly Met Tyr			
[6425]	785	790	795	800
[6426]	Ala Ser Phe Val Glu Glu Ile Lys Asn Asn Leu Asp Leu Gln Glu Trp			
[6427]	805	810	815	
[6428]	Tyr Ile Val Gly Arg Phe Leu Asn Gly Arg Met Leu Asn Gln Leu Ala			
[6429]	820	825	830	
[6430]	Gly Ser Leu Arg Ser Tyr Ile Gln Tyr Ala Glu Asp Ile Glu Arg Arg			
[6431]	835	840	845	
[6432]	Ala Ala Glu Asn Arg Asn Lys Leu Phe Ser Lys Pro Asp Glu Lys Ile			
[6433]	850	855	860	

[6434]	Glu Ala Cys Lys Lys Ala Val Arg Val Leu Asp Leu Cys Ile Lys Ile		
[6435]	865	870	875 880
[6436]	Ser Thr Arg Ile Ser Ala Glu Phe Thr Asp Tyr Phe Asp Ser Glu Asp		
[6437]		885	890 895
[6438]	Asp Tyr Ala Asp Tyr Leu Glu Lys Tyr Leu Lys Tyr Gln Asp Asp Ala		
[6439]		900	905 910
[6440]	Ile Lys Glu Leu Ser Gly Ser Ser Tyr Ala Ala Leu Asp His Phe Cys		
[6441]		915	920 925
[6442]	Asn Lys Asp Asp Leu Lys Phe Asp Ile Tyr Val Asn Ala Gly Gln Lys		
[6443]		930	935 940
[6444]	Pro Ile Leu Gln Arg Asn Ile Val Met Ala Lys Leu Phe Gly Pro Asp		
[6445]	945	950	955 960
[6446]	Asn Ile Leu Ser Glu Val Met Glu Lys Val Thr Glu Ser Ala Ile Arg		
[6447]		965	970 975
[6448]	Glu Tyr Tyr Asp Tyr Leu Lys Lys Val Ser Gly Tyr Arg Val Arg Gly		
[6449]		980	985 990
[6450]	Lys Cys Ser Thr Glu Lys Glu Gln Glu Asp Leu Leu Lys Phe Gln Arg		
[6451]		995	1000 1005
[6452]	Leu Lys Asn Ala Val Glu Phe Arg Asp Val Thr Glu Tyr Ala Glu		
[6453]		1010	1015 1020
[6454]	Val Ile Asn Glu Leu Leu Gly Gln Leu Ile Ser Trp Ser Tyr Leu		
[6455]		1025	1030 1035
[6456]	Arg Glu Arg Asp Leu Leu Tyr Phe Gln Leu Gly Phe His Tyr Met		
[6457]		1040	1045 1050
[6458]	Cys Leu Lys Asn Lys Ser Phe Lys Pro Ala Glu Tyr Val Asp Ile		
[6459]		1055	1060 1065
[6460]	Arg Arg Asn Asn Gly Thr Ile Ile His Asn Ala Ile Leu Tyr Gln		
[6461]		1070	1075 1080
[6462]	Ile Val Ser Met Tyr Ile Asn Gly Leu Asp Phe Tyr Ser Cys Asp		
[6463]		1085	1090 1095
[6464]	Lys Glu Gly Lys Thr Leu Lys Pro Ile Glu Thr Gly Lys Gly Val		
[6465]		1100	1105 1110
[6466]	Gly Ser Lys Ile Gly Gln Phe Ile Lys Tyr Ser Gln Tyr Leu Tyr		
[6467]		1115	1120 1125
[6468]	Asn Asp Pro Ser Tyr Lys Leu Glu Ile Tyr Asn Ala Gly Leu Glu		
[6469]		1130	1135 1140
[6470]	Val Phe Glu Asn Ile Asp Glu His Asp Asn Ile Thr Asp Leu Arg		
[6471]		1145	1150 1155
[6472]	Lys Tyr Val Asp His Phe Lys Tyr Tyr Ala Tyr Gly Asn Lys Met		

[6473]	1160	1165	1170
[6474]	Ser Leu Leu Asp Leu Tyr Ser Glu Phe Phe Asp Arg Phe Phe Thr		
[6475]	1175	1180	1185
[6476]	Tyr Asp Met Lys Tyr Gln Lys Asn Val Val Asn Val Leu Glu Asn		
[6477]	1190	1195	1200
[6478]	Ile Leu Leu Arg His Phe Val Ile Phe Tyr Pro Lys Phe Gly Ser		
[6479]	1205	1210	1215
[6480]	Gly Lys Lys Asp Val Gly Ile Arg Asp Cys Lys Lys Glu Arg Ala		
[6481]	1220	1225	1230
[6482]	Gln Ile Glu Ile Ser Glu Gln Ser Leu Thr Ser Glu Asp Phe Met		
[6483]	1235	1240	1245
[6484]	Phe Lys Leu Asp Asp Lys Ala Gly Glu Glu Ala Lys Lys Phe Pro		
[6485]	1250	1255	1260
[6486]	Ala Arg Asp Glu Arg Tyr Leu Gln Thr Ile Ala Lys Leu Leu Tyr		
[6487]	1265	1270	1275
[6488]	Tyr Pro Asn Glu Ile Glu Asp Met Asn Arg Phe Met Lys Lys Gly		
[6489]	1280	1285	1290
[6490]	Glu Thr Ile Asn Lys Lys Val Gln Phe Asn Arg Lys Lys Lys Ile		
[6491]	1295	1300	1305
[6492]	Thr Arg Lys Gln Lys Asn Asn Ser Ser Asn Glu Val Leu Ser Ser		
[6493]	1310	1315	1320
[6494]	Thr Met Gly Tyr Leu Phe Lys Asn Ile Lys Leu		
[6495]	1325	1330	
[6496]	<210> 317		
[6497]	<211> 1385		
[6498]	<212> PRT		
[6499]	<213> Clostridium aminophilum		
[6500]	<400> 317		
[6501]	Met Lys Phe Ser Lys Val Asp His Thr Arg Ser Ala Val Gly Ile Gln		
[6502]	1 5 10 15		
[6503]	Lys Ala Thr Asp Ser Val His Gly Met Leu Tyr Thr Asp Pro Lys Lys		
[6504]	20 25 30		
[6505]	Gln Glu Val Asn Asp Leu Asp Lys Arg Phe Asp Gln Leu Asn Val Lys		
[6506]	35 40 45		
[6507]	Ala Lys Arg Leu Tyr Asn Val Phe Asn Gln Ser Lys Ala Glu Glu Asp		
[6508]	50 55 60		
[6509]	Asp Asp Glu Lys Arg Phe Gly Lys Val Val Lys Lys Leu Asn Arg Glu		
[6510]	65 70 75 80		
[6511]	Leu Lys Asp Leu Leu Phe His Arg Glu Val Ser Arg Tyr Asn Ser Ile		

[6512]					85					90					95		
[6513]	Gly	Asn	Ala	Lys	Tyr	Asn	Tyr	Tyr	Gly	Ile	Lys	Ser	Asn	Pro	Glu	Glu	
[6514]					100					105					110		
[6515]	Ile	Val	Ser	Asn	Leu	Gly	Met	Val	Glu	Ser	Leu	Lys	Gly	Glu	Arg	Asp	
[6516]					115					120					125		
[6517]	Pro	Gln	Lys	Val	Ile	Ser	Lys	Leu	Leu	Leu	Tyr	Tyr	Leu	Arg	Lys	Gly	
[6518]					130					135					140		
[6519]	Leu	Lys	Pro	Gly	Thr	Asp	Gly	Leu	Arg	Met	Ile	Leu	Glu	Ala	Ser	Cys	
[6520]	145									150					155		160
[6521]	Gly	Leu	Arg	Lys	Leu	Ser	Gly	Asp	Glu	Lys	Glu	Leu	Lys	Val	Phe	Leu	
[6522]					165					170					175		
[6523]	Gln	Thr	Leu	Asp	Glu	Asp	Phe	Glu	Lys	Lys	Thr	Phe	Lys	Lys	Asn	Leu	
[6524]					180					185					190		
[6525]	Ile	Arg	Ser	Ile	Glu	Asn	Gln	Asn	Met	Ala	Val	Gln	Pro	Ser	Asn	Glu	
[6526]					195					200					205		
[6527]	Gly	Asp	Pro	Ile	Ile	Gly	Ile	Thr	Gln	Gly	Arg	Phe	Asn	Ser	Gln	Lys	
[6528]					210					215					220		
[6529]	Asn	Glu	Glu	Lys	Ser	Ala	Ile	Glu	Arg	Met	Met	Ser	Met	Tyr	Ala	Asp	
[6530]	225									230					235		240
[6531]	Leu	Asn	Glu	Asp	His	Arg	Glu	Asp	Val	Leu	Arg	Lys	Leu	Arg	Arg	Leu	
[6532]					245					250					255		
[6533]	Asn	Val	Leu	Tyr	Phe	Asn	Val	Asp	Thr	Glu	Lys	Thr	Glu	Glu	Pro	Thr	
[6534]					260					265					270		
[6535]	Leu	Pro	Gly	Glu	Val	Asp	Thr	Asn	Pro	Val	Phe	Glu	Val	Trp	His	Asp	
[6536]					275					280					285		
[6537]	His	Glu	Lys	Gly	Lys	Glu	Asn	Asp	Arg	Gln	Phe	Ala	Thr	Phe	Ala	Lys	
[6538]					290					295					300		
[6539]	Ile	Leu	Thr	Glu	Asp	Arg	Glu	Thr	Arg	Lys	Lys	Glu	Lys	Leu	Ala	Val	
[6540]	305									310					315		320
[6541]	Lys	Glu	Ala	Leu	Asn	Asp	Leu	Lys	Ser	Ala	Ile	Arg	Asp	His	Asn	Ile	
[6542]					325					330					335		
[6543]	Met	Ala	Tyr	Arg	Cys	Ser	Ile	Lys	Val	Thr	Glu	Gln	Asp	Lys	Asp	Gly	
[6544]					340					345					350		
[6545]	Leu	Phe	Phe	Glu	Asp	Gln	Arg	Ile	Asn	Arg	Phe	Trp	Ile	His	His	Ile	
[6546]					355					360					365		
[6547]	Glu	Ser	Ala	Val	Glu	Arg	Ile	Leu	Ala	Ser	Ile	Asn	Pro	Glu	Lys	Leu	
[6548]					370					375					380		
[6549]	Tyr	Lys	Leu	Arg	Ile	Gly	Tyr	Leu	Gly	Glu	Lys	Val	Trp	Lys	Asp	Leu	
[6550]	385									390					395		400

[6551]	Leu	Asn	Tyr	Leu	Ser	Ile	Lys	Tyr	Ile	Ala	Val	Gly	Lys	Ala	Val	Phe
[6552]					405					410					415	
[6553]	His	Phe	Ala	Met	Glu	Asp	Leu	Gly	Lys	Thr	Gly	Gln	Asp	Ile	Glu	Leu
[6554]				420					425						430	
[6555]	Gly	Lys	Leu	Ser	Asn	Ser	Val	Ser	Gly	Gly	Leu	Thr	Ser	Phe	Asp	Tyr
[6556]				435					440						445	
[6557]	Glu	Gln	Ile	Arg	Ala	Asp	Glu	Thr	Leu	Gln	Arg	Gln	Leu	Ser	Val	Glu
[6558]			450					455					460			
[6559]	Val	Ala	Phe	Ala	Ala	Asn	Asn	Leu	Phe	Arg	Ala	Val	Val	Gly	Gln	Thr
[6560]	465					470					475					480
[6561]	Gly	Lys	Lys	Ile	Glu	Gln	Ser	Lys	Ser	Glu	Glu	Asn	Glu	Glu	Asp	Phe
[6562]				485					490							495
[6563]	Leu	Leu	Trp	Lys	Ala	Glu	Lys	Ile	Ala	Glu	Ser	Ile	Lys	Lys	Glu	Gly
[6564]				500					505						510	
[6565]	Glu	Gly	Asn	Thr	Leu	Lys	Ser	Ile	Leu	Gln	Phe	Phe	Gly	Gly	Ala	Ser
[6566]			515					520					525			
[6567]	Ser	Trp	Asp	Leu	Asn	His	Phe	Cys	Ala	Ala	Tyr	Gly	Asn	Glu	Ser	Ser
[6568]		530					535					540				
[6569]	Ala	Leu	Gly	Tyr	Glu	Thr	Lys	Phe	Ala	Asp	Asp	Leu	Arg	Lys	Ala	Ile
[6570]	545					550				555						560
[6571]	Tyr	Ser	Leu	Arg	Asn	Glu	Thr	Phe	His	Phe	Thr	Thr	Leu	Asn	Lys	Gly
[6572]				565					570							575
[6573]	Ser	Phe	Asp	Trp	Asn	Ala	Lys	Leu	Ile	Gly	Asp	Met	Phe	Ser	His	Glu
[6574]			580					585					590			
[6575]	Ala	Ala	Thr	Gly	Ile	Ala	Val	Glu	Arg	Thr	Arg	Phe	Tyr	Ser	Asn	Asn
[6576]			595					600					605			
[6577]	Leu	Pro	Met	Phe	Tyr	Arg	Glu	Ser	Asp	Leu	Lys	Arg	Ile	Met	Asp	His
[6578]		610					615					620				
[6579]	Leu	Tyr	Asn	Thr	Tyr	His	Pro	Arg	Ala	Ser	Gln	Val	Pro	Ser	Phe	Asn
[6580]	625					630				635						640
[6581]	Ser	Val	Phe	Val	Arg	Lys	Asn	Phe	Arg	Leu	Phe	Leu	Ser	Asn	Thr	Leu
[6582]				645					650						655	
[6583]	Asn	Thr	Asn	Thr	Ser	Phe	Asp	Thr	Glu	Val	Tyr	Gln	Lys	Trp	Glu	Ser
[6584]				660					665						670	
[6585]	Gly	Val	Tyr	Tyr	Leu	Phe	Lys	Glu	Ile	Tyr	Tyr	Asn	Ser	Phe	Leu	Pro
[6586]			675					680					685			
[6587]	Ser	Gly	Asp	Ala	His	His	Leu	Phe	Phe	Glu	Gly	Leu	Arg	Arg	Ile	Arg
[6588]		690					695					700				
[6589]	Lys	Glu	Ala	Asp	Asn	Leu	Pro	Ile	Val	Gly	Lys	Glu	Ala	Lys	Lys	Arg

[6590]	705	710	715	720
[6591]	Asn Ala Val Gln Asp Phe Gly Arg Arg Cys Asp Glu Leu Lys Asn Leu			
[6592]		725	730	735
[6593]	Ser Leu Ser Ala Ile Cys Gln Met Ile Met Thr Glu Tyr Asn Glu Gln			
[6594]		740	745	750
[6595]	Asn Asn Gly Asn Arg Lys Val Lys Ser Thr Arg Glu Asp Lys Arg Lys			
[6596]		755	760	765
[6597]	Pro Asp Ile Phe Gln His Tyr Lys Met Leu Leu Leu Arg Thr Leu Gln			
[6598]		770	775	780
[6599]	Glu Ala Phe Ala Ile Tyr Ile Arg Arg Glu Glu Phe Lys Phe Ile Phe			
[6600]	785	790	795	800
[6601]	Asp Leu Pro Lys Thr Leu Tyr Val Met Lys Pro Val Glu Glu Phe Leu			
[6602]		805	810	815
[6603]	Pro Asn Trp Lys Ser Gly Met Phe Asp Ser Leu Val Glu Arg Val Lys			
[6604]		820	825	830
[6605]	Gln Ser Pro Asp Leu Gln Arg Trp Tyr Val Leu Cys Lys Phe Leu Asn			
[6606]		835	840	845
[6607]	Gly Arg Leu Leu Asn Gln Leu Ser Gly Val Ile Arg Ser Tyr Ile Gln			
[6608]		850	855	860
[6609]	Phe Ala Gly Asp Ile Gln Arg Arg Ala Lys Ala Asn His Asn Arg Leu			
[6610]	865	870	875	880
[6611]	Tyr Met Asp Asn Thr Gln Arg Val Glu Tyr Tyr Ser Asn Val Leu Glu			
[6612]		885	890	895
[6613]	Val Val Asp Phe Cys Ile Lys Gly Thr Ser Arg Phe Ser Asn Val Phe			
[6614]		900	905	910
[6615]	Ser Asp Tyr Phe Arg Asp Glu Asp Ala Tyr Ala Asp Tyr Leu Asp Asn			
[6616]		915	920	925
[6617]	Tyr Leu Gln Phe Lys Asp Glu Lys Ile Ala Glu Val Ser Ser Phe Ala			
[6618]		930	935	940
[6619]	Ala Leu Lys Thr Phe Cys Asn Glu Glu Glu Val Lys Ala Gly Ile Tyr			
[6620]	945	950	955	960
[6621]	Met Asp Gly Glu Asn Pro Val Met Gln Arg Asn Ile Val Met Ala Lys			
[6622]		965	970	975
[6623]	Leu Phe Gly Pro Asp Glu Val Leu Lys Asn Val Val Pro Lys Val Thr			
[6624]		980	985	990
[6625]	Arg Glu Glu Ile Glu Glu Tyr Tyr Gln Leu Glu Lys Gln Ile Ala Pro			
[6626]		995	1000	1005
[6627]	Tyr Arg Gln Asn Gly Tyr Cys Lys Ser Glu Glu Asp Gln Lys Lys			
[6628]		1010	1015	1020

[6629]	Leu	Leu	Arg	Phe	Gln	Arg	Ile	Lys	Asn	Arg	Val	Glu	Phe	Gln	Thr
[6630]		1025					1030					1035			
[6631]	Ile	Thr	Glu	Phe	Ser	Glu	Ile	Ile	Asn	Glu	Leu	Leu	Gly	Gln	Leu
[6632]		1040					1045					1050			
[6633]	Ile	Ser	Trp	Ser	Phe	Leu	Arg	Glu	Arg	Asp	Leu	Leu	Tyr	Phe	Gln
[6634]		1055					1060					1065			
[6635]	Leu	Gly	Phe	His	Tyr	Leu	Cys	Leu	His	Asn	Asp	Thr	Glu	Lys	Pro
[6636]		1070					1075					1080			
[6637]	Ala	Glu	Tyr	Lys	Glu	Ile	Ser	Arg	Glu	Asp	Gly	Thr	Val	Ile	Arg
[6638]		1085					1090					1095			
[6639]	Asn	Ala	Ile	Leu	His	Gln	Val	Ala	Ala	Met	Tyr	Val	Gly	Gly	Leu
[6640]		1100					1105					1110			
[6641]	Pro	Val	Tyr	Thr	Leu	Ala	Asp	Lys	Lys	Leu	Ala	Ala	Phe	Glu	Lys
[6642]		1115					1120					1125			
[6643]	Gly	Glu	Ala	Asp	Cys	Lys	Leu	Ser	Ile	Ser	Lys	Asp	Thr	Ala	Gly
[6644]		1130					1135					1140			
[6645]	Ala	Gly	Lys	Lys	Ile	Lys	Asp	Phe	Phe	Arg	Tyr	Ser	Lys	Tyr	Val
[6646]		1145					1150					1155			
[6647]	Leu	Ile	Lys	Asp	Arg	Met	Leu	Thr	Asp	Gln	Asn	Gln	Lys	Tyr	Thr
[6648]		1160					1165					1170			
[6649]	Ile	Tyr	Leu	Ala	Gly	Leu	Glu	Leu	Phe	Glu	Asn	Thr	Asp	Glu	His
[6650]		1175					1180					1185			
[6651]	Asp	Asn	Ile	Thr	Asp	Val	Arg	Lys	Tyr	Val	Asp	His	Phe	Lys	Tyr
[6652]		1190					1195					1200			
[6653]	Tyr	Ala	Thr	Ser	Asp	Glu	Asn	Ala	Met	Ser	Ile	Leu	Asp	Leu	Tyr
[6654]		1205					1210					1215			
[6655]	Ser	Glu	Ile	His	Asp	Arg	Phe	Phe	Thr	Tyr	Asp	Met	Lys	Tyr	Gln
[6656]		1220					1225					1230			
[6657]	Lys	Asn	Val	Ala	Asn	Met	Leu	Glu	Asn	Ile	Leu	Leu	Arg	His	Phe
[6658]		1235					1240					1245			
[6659]	Val	Leu	Ile	Arg	Pro	Glu	Phe	Phe	Thr	Gly	Ser	Lys	Lys	Val	Gly
[6660]		1250					1255					1260			
[6661]	Glu	Gly	Lys	Lys	Ile	Thr	Cys	Lys	Ala	Arg	Ala	Gln	Ile	Glu	Ile
[6662]		1265					1270					1275			
[6663]	Ala	Glu	Asn	Gly	Met	Arg	Ser	Glu	Asp	Phe	Thr	Tyr	Lys	Leu	Ser
[6664]		1280					1285					1290			
[6665]	Asp	Gly	Lys	Lys	Asn	Ile	Ser	Thr	Cys	Met	Ile	Ala	Ala	Arg	Asp
[6666]		1295					1300					1305			
[6667]	Gln	Lys	Tyr	Leu	Asn	Thr	Val	Ala	Arg	Leu	Leu	Tyr	Tyr	Pro	His

[6668]	1310	1315	1320
[6669]	Glu Ala Lys Lys Ser Ile Val	Asp Thr Arg Glu Lys	Lys Asn Asn
[6670]	1325	1330	1335
[6671]	Lys Lys Thr Asn Arg Gly Asp	Gly Thr Phe Asn Lys	Gln Lys Gly
[6672]	1340	1345	1350
[6673]	Thr Ala Arg Lys Glu Lys Asp	Asn Gly Pro Arg Glu	Phe Asn Asp
[6674]	1355	1360	1365
[6675]	Thr Gly Phe Ser Asn Thr Pro	Phe Ala Gly Phe Asp	Pro Phe Arg
[6676]	1370	1375	1380
[6677]	Asn Ser		
[6678]	1385		
[6679]	<210> 318		
[6680]	<211> 1175		
[6681]	<212> PRT		
[6682]	<213> 鸡肉杆菌		
[6683]	<400> 318		
[6684]	Met Arg Ile Thr Lys Val Lys Ile Lys Leu Asp Asn Lys Leu Tyr Gln		
[6685]	1	5	10
[6686]	Val Thr Met Gln Lys Glu Glu Lys Tyr Gly Thr Leu Lys Leu Asn Glu		
[6687]	20	25	30
[6688]	Glu Ser Arg Lys Ser Thr Ala Glu Ile Leu Arg Leu Lys Lys Ala Ser		
[6689]	35	40	45
[6690]	Phe Asn Lys Ser Phe His Ser Lys Thr Ile Asn Ser Gln Lys Glu Asn		
[6691]	50	55	60
[6692]	Lys Asn Ala Thr Ile Lys Lys Asn Gly Asp Tyr Ile Ser Gln Ile Phe		
[6693]	65	70	75
[6694]	Glu Lys Leu Val Gly Val Asp Thr Asn Lys Asn Ile Arg Lys Pro Lys		
[6695]	85	90	95
[6696]	Met Ser Leu Thr Asp Leu Lys Asp Leu Pro Lys Lys Asp Leu Ala Leu		
[6697]	100	105	110
[6698]	Phe Ile Lys Arg Lys Phe Lys Asn Asp Asp Ile Val Glu Ile Lys Asn		
[6699]	115	120	125
[6700]	Leu Asp Leu Ile Ser Leu Phe Tyr Asn Ala Leu Gln Lys Val Pro Gly		
[6701]	130	135	140
[6702]	Glu His Phe Thr Asp Glu Ser Trp Ala Asp Phe Cys Gln Glu Met Met		
[6703]	145	150	155
[6704]	Pro Tyr Arg Glu Tyr Lys Asn Lys Phe Ile Glu Arg Lys Ile Ile Leu		
[6705]	165	170	175
[6706]	Leu Ala Asn Ser Ile Glu Gln Asn Lys Gly Phe Ser Ile Asn Pro Glu		

[6707]		180		185		190
[6708]	Thr Phe Ser Lys Arg Lys Arg Val Leu His Gln Trp Ala Ile Glu Val					
[6709]		195		200		205
[6710]	Gln Glu Arg Gly Asp Phe Ser Ile Leu Asp Glu Lys Leu Ser Lys Leu					
[6711]		210		215		220
[6712]	Ala Glu Ile Tyr Asn Phe Lys Lys Met Cys Lys Arg Val Gln Asp Glu					
[6713]	225		230		235	240
[6714]	Leu Asn Asp Leu Glu Lys Ser Met Lys Lys Gly Lys Asn Pro Glu Lys					
[6715]		245		250		255
[6716]	Glu Lys Glu Ala Tyr Lys Lys Gln Lys Asn Phe Lys Ile Lys Thr Ile					
[6717]		260		265		270
[6718]	Trp Lys Asp Tyr Pro Tyr Lys Thr His Ile Gly Leu Ile Glu Lys Ile					
[6719]		275		280		285
[6720]	Lys Glu Asn Glu Glu Leu Asn Gln Phe Asn Ile Glu Ile Gly Lys Tyr					
[6721]		290		295		300
[6722]	Phe Glu His Tyr Phe Pro Ile Lys Lys Glu Arg Cys Thr Glu Asp Glu					
[6723]	305		310		315	320
[6724]	Pro Tyr Tyr Leu Asn Ser Glu Thr Ile Ala Thr Thr Val Asn Tyr Gln					
[6725]		325		330		335
[6726]	Leu Lys Asn Ala Leu Ile Ser Tyr Leu Met Gln Ile Gly Lys Tyr Lys					
[6727]		340		345		350
[6728]	Gln Phe Gly Leu Glu Asn Gln Val Leu Asp Ser Lys Lys Leu Gln Glu					
[6729]		355		360		365
[6730]	Ile Gly Ile Tyr Glu Gly Phe Gln Thr Lys Phe Met Asp Ala Cys Val					
[6731]		370		375		380
[6732]	Phe Ala Thr Ser Ser Leu Lys Asn Ile Ile Glu Pro Met Arg Ser Gly					
[6733]	385		390		395	400
[6734]	Asp Ile Leu Gly Lys Arg Glu Phe Lys Glu Ala Ile Ala Thr Ser Ser					
[6735]		405		410		415
[6736]	Phe Val Asn Tyr His His Phe Phe Pro Tyr Phe Pro Phe Glu Leu Lys					
[6737]		420		425		430
[6738]	Gly Met Lys Asp Arg Glu Ser Glu Leu Ile Pro Phe Gly Glu Gln Thr					
[6739]		435		440		445
[6740]	Glu Ala Lys Gln Met Gln Asn Ile Trp Ala Leu Arg Gly Ser Val Gln					
[6741]		450		455		460
[6742]	Gln Ile Arg Asn Glu Ile Phe His Ser Phe Asp Lys Asn Gln Lys Phe					
[6743]	465		470		475	480
[6744]	Asn Leu Pro Gln Leu Asp Lys Ser Asn Phe Glu Phe Asp Ala Ser Glu					
[6745]		485		490		495

[6746]	Asn Ser Thr Gly Lys Ser Gln Ser Tyr Ile Glu Thr Asp Tyr Lys Phe		
[6747]		500	505
[6748]	Leu Phe Glu Ala Glu Lys Asn Gln Leu Glu Gln Phe Phe Ile Glu Arg		
[6749]		515	520
[6750]	Ile Lys Ser Ser Gly Ala Leu Glu Tyr Tyr Pro Leu Lys Ser Leu Glu		
[6751]		530	535
[6752]	Lys Leu Phe Ala Lys Lys Glu Met Lys Phe Ser Leu Gly Ser Gln Val		
[6753]	545	550	555
[6754]	Val Ala Phe Ala Pro Ser Tyr Lys Lys Leu Val Lys Lys Gly His Ser		
[6755]		565	570
[6756]	Tyr Gln Thr Ala Thr Glu Gly Thr Ala Asn Tyr Leu Gly Leu Ser Tyr		
[6757]		580	585
[6758]	Tyr Asn Arg Tyr Glu Leu Lys Glu Glu Ser Phe Gln Ala Gln Tyr Tyr		
[6759]		595	600
[6760]	Leu Leu Lys Leu Ile Tyr Gln Tyr Val Phe Leu Pro Asn Phe Ser Gln		
[6761]		610	615
[6762]	Gly Asn Ser Pro Ala Phe Arg Glu Thr Val Lys Ala Ile Leu Arg Ile		
[6763]	625	630	635
[6764]	Asn Lys Asp Glu Ala Arg Lys Lys Met Lys Lys Asn Lys Lys Phe Leu		
[6765]		645	650
[6766]	Arg Lys Tyr Ala Phe Glu Gln Val Arg Glu Met Glu Phe Lys Glu Thr		
[6767]		660	665
[6768]	Pro Asp Gln Tyr Met Ser Tyr Leu Gln Ser Glu Met Arg Glu Glu Lys		
[6769]		675	680
[6770]	Val Arg Lys Ala Glu Lys Asn Asp Lys Gly Phe Glu Lys Asn Ile Thr		
[6771]		690	695
[6772]	Met Asn Phe Glu Lys Leu Leu Met Gln Ile Phe Val Lys Gly Phe Asp		
[6773]	705	710	715
[6774]	Val Phe Leu Thr Thr Phe Ala Gly Lys Glu Leu Leu Leu Ser Ser Glu		
[6775]		725	730
[6776]	Glu Lys Val Ile Lys Glu Thr Glu Ile Ser Leu Ser Lys Lys Ile Asn		
[6777]		740	745
[6778]	Glu Arg Glu Lys Thr Leu Lys Ala Ser Ile Gln Val Glu His Gln Leu		
[6779]		755	760
[6780]	Val Ala Thr Asn Ser Ala Ile Ser Tyr Trp Leu Phe Cys Lys Leu Leu		
[6781]		770	775
[6782]	Asp Ser Arg His Leu Asn Glu Leu Arg Asn Glu Met Ile Lys Phe Lys		
[6783]	785	790	795
[6784]	Gln Ser Arg Ile Lys Phe Asn His Thr Gln His Ala Glu Leu Ile Gln		

[6785]		805		810		815
[6786]	Asn Leu Leu Pro Ile Val Glu Leu Thr Ile Leu Ser Asn Asp Tyr Asp					
[6787]		820		825		830
[6788]	Glu Lys Asn Asp Ser Gln Asn Val Asp Val Ser Ala Tyr Phe Glu Asp					
[6789]		835		840		845
[6790]	Lys Ser Leu Tyr Glu Thr Ala Pro Tyr Val Gln Thr Asp Asp Arg Thr					
[6791]		850		855		860
[6792]	Arg Val Ser Phe Arg Pro Ile Leu Lys Leu Glu Lys Tyr His Thr Lys					
[6793]		865		870		875
[6794]	Ser Leu Ile Glu Ala Leu Leu Lys Asp Asn Pro Gln Phe Arg Val Ala					
[6795]		885		890		895
[6796]	Ala Thr Asp Ile Gln Glu Trp Met His Lys Arg Glu Glu Ile Gly Glu					
[6797]		900		905		910
[6798]	Leu Val Glu Lys Arg Lys Asn Leu His Thr Glu Trp Ala Glu Gly Gln					
[6799]		915		920		925
[6800]	Gln Thr Leu Gly Ala Glu Lys Arg Glu Glu Tyr Arg Asp Tyr Cys Lys					
[6801]		930		935		940
[6802]	Lys Ile Asp Arg Phe Asn Trp Lys Ala Asn Lys Val Thr Leu Thr Tyr					
[6803]		945		950		955
[6804]	Leu Ser Gln Leu His Tyr Leu Ile Thr Asp Leu Leu Gly Arg Met Val					
[6805]		965		970		975
[6806]	Gly Phe Ser Ala Leu Phe Glu Arg Asp Leu Val Tyr Phe Ser Arg Ser					
[6807]		980		985		990
[6808]	Phe Ser Glu Leu Gly Gly Glu Thr Tyr His Ile Ser Asp Tyr Lys Asn					
[6809]		995		1000		1005
[6810]	Leu Ser Gly Val Leu Arg Leu Asn Ala Glu Val Lys Pro Ile Lys					
[6811]		1010		1015		1020
[6812]	Ile Lys Asn Ile Lys Val Ile Asp Asn Glu Glu Asn Pro Tyr Lys					
[6813]		1025		1030		1035
[6814]	Gly Asn Glu Pro Glu Val Lys Pro Phe Leu Asp Arg Leu His Ala					
[6815]		1040		1045		1050
[6816]	Tyr Leu Glu Asn Val Ile Gly Ile Lys Ala Val His Gly Lys Ile					
[6817]		1055		1060		1065
[6818]	Arg Asn Gln Thr Ala His Leu Ser Val Leu Gln Leu Glu Leu Ser					
[6819]		1070		1075		1080
[6820]	Met Ile Glu Ser Met Asn Asn Leu Arg Asp Leu Met Ala Tyr Asp					
[6821]		1085		1090		1095
[6822]	Arg Lys Leu Lys Asn Ala Val Thr Lys Ser Met Ile Lys Ile Leu					
[6823]		1100		1105		1110

[6824]	Asp Lys His Gly Met Ile Leu Lys Leu Lys Ile Asp Glu Asn His
[6825]	1115 1120 1125
[6826]	Lys Asn Phe Glu Ile Glu Ser Leu Ile Pro Lys Glu Ile Ile His
[6827]	1130 1135 1140
[6828]	Leu Lys Asp Lys Ala Ile Lys Thr Asn Gln Val Ser Glu Glu Tyr
[6829]	1145 1150 1155
[6830]	Cys Gln Leu Val Leu Ala Leu Leu Thr Thr Asn Pro Gly Asn Gln
[6831]	1160 1165 1170
[6832]	Leu Asn
[6833]	1175
[6834]	<210> 319
[6835]	<211> 1164
[6836]	<212> PRT
[6837]	<213> 鸡肉杆菌
[6838]	<400> 319
[6839]	Met Arg Met Thr Lys Val Lys Ile Asn Gly Ser Pro Val Ser Met Asn
[6840]	1 5 10 15
[6841]	Arg Ser Lys Leu Asn Gly His Leu Val Trp Asn Gly Thr Thr Asn Thr
[6842]	20 25 30
[6843]	Val Asn Ile Leu Thr Lys Lys Glu Gln Ser Phe Ala Ala Ser Phe Leu
[6844]	35 40 45
[6845]	Asn Lys Thr Leu Val Lys Ala Asp Gln Val Lys Gly Tyr Lys Val Leu
[6846]	50 55 60
[6847]	Ala Glu Asn Ile Phe Ile Ile Phe Glu Gln Leu Glu Lys Ser Asn Ser
[6848]	65 70 75 80
[6849]	Glu Lys Pro Ser Val Tyr Leu Asn Asn Ile Arg Arg Leu Lys Glu Ala
[6850]	85 90 95
[6851]	Gly Leu Lys Arg Phe Phe Lys Ser Lys Tyr His Glu Glu Ile Lys Tyr
[6852]	100 105 110
[6853]	Thr Ser Glu Lys Asn Gln Ser Val Pro Thr Lys Leu Asn Leu Ile Pro
[6854]	115 120 125
[6855]	Leu Phe Phe Asn Ala Val Asp Arg Ile Gln Glu Asp Lys Phe Asp Glu
[6856]	130 135 140
[6857]	Lys Asn Trp Ser Tyr Phe Cys Lys Glu Met Ser Pro Tyr Leu Asp Tyr
[6858]	145 150 155 160
[6859]	Lys Lys Ser Tyr Leu Asn Arg Lys Lys Glu Ile Leu Ala Asn Ser Ile
[6860]	165 170 175
[6861]	Gln Gln Asn Arg Gly Phe Ser Met Pro Thr Ala Glu Glu Pro Asn Leu
[6862]	180 185 190

[6863]	Leu Ser Lys Arg Lys Gln Leu Phe Gln Gln Trp Ala Met Lys Phe Gln		
[6864]	195	200	205
[6865]	Glu Ser Pro Leu Ile Gln Gln Asn Asn Phe Ala Val Glu Gln Phe Asn		
[6866]	210	215	220
[6867]	Lys Glu Phe Ala Asn Lys Ile Asn Glu Leu Ala Ala Val Tyr Asn Val		
[6868]	225	230	235 240
[6869]	Asp Glu Leu Cys Thr Ala Ile Thr Glu Lys Leu Met Asn Phe Asp Lys		
[6870]	245	250	255
[6871]	Asp Lys Ser Asn Lys Thr Arg Asn Phe Glu Ile Lys Lys Leu Trp Lys		
[6872]	260	265	270
[6873]	Gln His Pro His Asn Lys Asp Lys Ala Leu Ile Lys Leu Phe Asn Gln		
[6874]	275	280	285
[6875]	Glu Gly Asn Glu Ala Leu Asn Gln Phe Asn Ile Glu Leu Gly Lys Tyr		
[6876]	290	295	300
[6877]	Phe Glu His Tyr Phe Pro Lys Thr Gly Lys Lys Glu Ser Ala Glu Ser		
[6878]	305	310	315 320
[6879]	Tyr Tyr Leu Asn Pro Gln Thr Ile Ile Lys Thr Val Gly Tyr Gln Leu		
[6880]	325	330	335
[6881]	Arg Asn Ala Phe Val Gln Tyr Leu Leu Gln Val Gly Lys Leu His Gln		
[6882]	340	345	350
[6883]	Tyr Asn Lys Gly Val Leu Asp Ser Gln Thr Leu Gln Glu Ile Gly Met		
[6884]	355	360	365
[6885]	Tyr Glu Gly Phe Gln Thr Lys Phe Met Asp Ala Cys Val Phe Ala Ser		
[6886]	370	375	380
[6887]	Ser Ser Leu Arg Asn Ile Ile Gln Ala Thr Thr Asn Glu Asp Ile Leu		
[6888]	385	390	395 400
[6889]	Thr Arg Glu Lys Phe Lys Lys Glu Leu Glu Lys Asn Val Glu Leu Lys		
[6890]	405	410	415
[6891]	His Asp Leu Phe Phe Lys Thr Glu Ile Val Glu Glu Arg Asp Glu Asn		
[6892]	420	425	430
[6893]	Pro Ala Lys Lys Ile Ala Met Thr Pro Asn Glu Leu Asp Leu Trp Ala		
[6894]	435	440	445
[6895]	Ile Arg Gly Ala Val Gln Arg Val Arg Asn Gln Ile Phe His Gln Gln		
[6896]	450	455	460
[6897]	Ile Asn Lys Arg His Glu Pro Asn Gln Leu Lys Val Gly Ser Phe Glu		
[6898]	465	470	475 480
[6899]	Asn Gly Asp Leu Gly Asn Val Ser Tyr Gln Lys Thr Ile Tyr Gln Lys		
[6900]	485	490	495
[6901]	Leu Phe Asp Ala Glu Ile Lys Asp Ile Glu Ile Tyr Phe Ala Glu Lys		

[6902]	500	505	510
[6903]	Ile Lys Ser Ser Gly Ala Leu Glu Gln Tyr Ser Met Lys Asp Leu Glu		
[6904]	515	520	525
[6905]	Lys Leu Phe Ser Asn Lys Glu Leu Thr Leu Ser Leu Gly Gly Gln Val		
[6906]	530	535	540
[6907]	Val Ala Phe Ala Pro Ser Tyr Lys Lys Leu Tyr Lys Gln Gly Tyr Phe		
[6908]	545	550	555
[6909]	Tyr Gln Asn Glu Lys Thr Ile Glu Leu Glu Gln Phe Thr Asp Tyr Asp		
[6910]	565	570	575
[6911]	Phe Ser Asn Asp Val Phe Lys Ala Asn Tyr Tyr Leu Ile Lys Leu Ile		
[6912]	580	585	590
[6913]	Tyr His Tyr Val Phe Leu Pro Gln Phe Ser Gln Ala Asn Asn Lys Leu		
[6914]	595	600	605
[6915]	Phe Lys Asp Thr Val His Tyr Val Ile Gln Gln Asn Lys Glu Leu Asn		
[6916]	610	615	620
[6917]	Thr Thr Glu Lys Asp Lys Lys Asn Asn Lys Lys Ile Arg Lys Tyr Ala		
[6918]	625	630	635
[6919]	Phe Glu Gln Val Lys Leu Met Lys Asn Glu Ser Pro Glu Lys Tyr Met		
[6920]	645	650	655
[6921]	Gln Tyr Leu Gln Arg Glu Met Gln Glu Glu Arg Thr Ile Lys Glu Ala		
[6922]	660	665	670
[6923]	Lys Lys Thr Asn Glu Glu Lys Pro Asn Tyr Asn Phe Glu Lys Leu Leu		
[6924]	675	680	685
[6925]	Ile Gln Ile Phe Ile Lys Gly Phe Asp Thr Phe Leu Arg Asn Phe Asp		
[6926]	690	695	700
[6927]	Leu Asn Leu Asn Pro Ala Glu Glu Leu Val Gly Thr Val Lys Glu Lys		
[6928]	705	710	715
[6929]	Ala Glu Gly Leu Arg Lys Arg Lys Glu Arg Ile Ala Lys Ile Leu Asn		
[6930]	725	730	735
[6931]	Val Asp Glu Gln Ile Lys Thr Gly Asp Glu Glu Ile Ala Phe Trp Ile		
[6932]	740	745	750
[6933]	Phe Ala Lys Leu Leu Asp Ala Arg His Leu Ser Glu Leu Arg Asn Glu		
[6934]	755	760	765
[6935]	Met Ile Lys Phe Lys Gln Ser Ser Val Lys Lys Gly Leu Ile Lys Asn		
[6936]	770	775	780
[6937]	Gly Asp Leu Ile Glu Gln Met Gln Pro Ile Leu Glu Leu Cys Ile Leu		
[6938]	785	790	795
[6939]	Ser Asn Asp Ser Glu Ser Met Glu Lys Glu Ser Phe Asp Lys Ile Glu		
[6940]	805	810	815

[6941]	Val Phe Leu Glu Lys Val Glu Leu Ala Lys Asn Glu Pro Tyr Met Gln
[6942]	820 825 830
[6943]	Glu Asp Lys Leu Thr Pro Val Lys Phe Arg Phe Met Lys Gln Leu Glu
[6944]	835 840 845
[6945]	Lys Tyr Gln Thr Arg Asn Phe Ile Glu Asn Leu Val Ile Glu Asn Pro
[6946]	850 855 860
[6947]	Glu Phe Lys Val Ser Glu Lys Ile Val Leu Asn Trp His Glu Glu Lys
[6948]	865 870 875 880
[6949]	Glu Lys Ile Ala Asp Leu Val Asp Lys Arg Thr Lys Leu His Glu Glu
[6950]	885 890 895
[6951]	Trp Ala Ser Lys Ala Arg Glu Ile Glu Glu Tyr Asn Glu Lys Ile Lys
[6952]	900 905 910
[6953]	Lys Asn Lys Ser Lys Lys Leu Asp Lys Pro Ala Glu Phe Ala Lys Phe
[6954]	915 920 925
[6955]	Ala Glu Tyr Lys Ile Ile Cys Glu Ala Ile Glu Asn Phe Asn Arg Leu
[6956]	930 935 940
[6957]	Asp His Lys Val Arg Leu Thr Tyr Leu Lys Asn Leu His Tyr Leu Met
[6958]	945 950 955 960
[6959]	Ile Asp Leu Met Gly Arg Met Val Gly Phe Ser Val Leu Phe Glu Arg
[6960]	965 970 975
[6961]	Asp Phe Val Tyr Met Gly Arg Ser Tyr Ser Ala Leu Lys Lys Gln Ser
[6962]	980 985 990
[6963]	Ile Tyr Leu Asn Asp Tyr Asp Thr Phe Ala Asn Ile Arg Asp Trp Glu
[6964]	995 1000 1005
[6965]	Val Asn Glu Asn Lys His Leu Phe Gly Thr Ser Ser Ser Asp Leu
[6966]	1010 1015 1020
[6967]	Thr Phe Gln Glu Thr Ala Glu Phe Lys Asn Leu Lys Lys Pro Met
[6968]	1025 1030 1035
[6969]	Glu Asn Gln Leu Lys Ala Leu Leu Gly Val Thr Asn His Ser Phe
[6970]	1040 1045 1050
[6971]	Glu Ile Arg Asn Asn Ile Ala His Leu His Val Leu Arg Asn Asp
[6972]	1055 1060 1065
[6973]	Gly Lys Gly Glu Gly Val Ser Leu Leu Ser Cys Met Asn Asp Leu
[6974]	1070 1075 1080
[6975]	Arg Lys Leu Met Ser Tyr Asp Arg Lys Leu Lys Asn Ala Val Thr
[6976]	1085 1090 1095
[6977]	Lys Ala Ile Ile Lys Ile Leu Asp Lys His Gly Met Ile Leu Lys
[6978]	1100 1105 1110
[6979]	Leu Thr Asn Asn Asp His Thr Lys Pro Phe Glu Ile Glu Ser Leu

[6980]	1115	1120	1125
[6981]	Lys Pro Lys Lys Ile Ile His	Leu Glu Lys Ser Asn	His Ser Phe
[6982]	1130	1135	1140
[6983]	Pro Met Asp Gln Val Ser Gln	Glu Tyr Cys Asp Leu	Val Lys Lys
[6984]	1145	1150	1155
[6985]	Met Leu Val Phe Thr Asn		
[6986]	1160		
[6987]	<210> 320		
[6988]	<211> 1051		
[6989]	<212> PRT		
[6990]	<213> <i>Listeria newyorkensis</i>		
[6991]	<400> 320		
[6992]	Met Lys Ile Thr Lys Met Arg Val Asp Gly Arg Thr Ile Val Met Glu		
[6993]	1 5 10 15		
[6994]	Arg Thr Ser Lys Glu Gly Gln Leu Gly Tyr Glu Gly Ile Asp Gly Asn		
[6995]	20 25 30		
[6996]	Lys Thr Thr Glu Ile Ile Phe Asp Lys Lys Lys Glu Ser Phe Tyr Lys		
[6997]	35 40 45		
[6998]	Ser Ile Leu Asn Lys Thr Val Arg Lys Pro Asp Glu Lys Glu Lys Asn		
[6999]	50 55 60		
[7000]	Arg Arg Lys Gln Ala Ile Asn Lys Ala Ile Asn Lys Glu Ile Thr Glu		
[7001]	65 70 75 80		
[7002]	Leu Met Leu Ala Val Leu His Gln Glu Val Pro Ser Gln Lys Leu His		
[7003]	85 90 95		
[7004]	Asn Leu Lys Ser Leu Asn Thr Glu Ser Leu Thr Lys Leu Phe Lys Pro		
[7005]	100 105 110		
[7006]	Lys Phe Gln Asn Met Ile Ser Tyr Pro Pro Ser Lys Gly Ala Glu His		
[7007]	115 120 125		
[7008]	Val Gln Phe Cys Leu Thr Asp Ile Ala Val Pro Ala Ile Arg Asp Leu		
[7009]	130 135 140		
[7010]	Asp Glu Ile Lys Pro Asp Trp Gly Ile Phe Phe Glu Lys Leu Lys Pro		
[7011]	145 150 155 160		
[7012]	Tyr Thr Asp Trp Ala Glu Ser Tyr Ile His Tyr Lys Gln Thr Thr Ile		
[7013]	165 170 175		
[7014]	Gln Lys Ser Ile Glu Gln Asn Lys Ile Gln Ser Pro Asp Ser Pro Arg		
[7015]	180 185 190		
[7016]	Lys Leu Val Leu Gln Lys Tyr Val Thr Ala Phe Leu Asn Gly Glu Pro		
[7017]	195 200 205		
[7018]	Leu Gly Leu Asp Leu Val Ala Lys Lys Tyr Lys Leu Ala Asp Leu Ala		

[7019]	210	215	220
[7020]	Glu Ser Phe Lys Leu Val Asp Leu Asn Glu Asp Lys Ser Ala Asn Tyr		
[7021]	225	230	235 240
[7022]	Lys Ile Lys Ala Cys Leu Gln Gln His Gln Arg Asn Ile Leu Asp Glu		
[7023]	245	250	255
[7024]	Leu Lys Glu Asp Pro Glu Leu Asn Gln Tyr Gly Ile Glu Val Lys Lys		
[7025]	260	265	270
[7026]	Tyr Ile Gln Arg Tyr Phe Pro Ile Lys Arg Ala Pro Asn Arg Ser Lys		
[7027]	275	280	285
[7028]	His Ala Arg Ala Asp Phe Leu Lys Lys Glu Leu Ile Glu Ser Thr Val		
[7029]	290	295	300
[7030]	Glu Gln Gln Phe Lys Asn Ala Val Tyr His Tyr Val Leu Glu Gln Gly		
[7031]	305	310	315 320
[7032]	Lys Met Glu Ala Tyr Glu Leu Thr Asp Pro Lys Thr Lys Asp Leu Gln		
[7033]	325	330	335
[7034]	Asp Ile Arg Ser Gly Glu Ala Phe Ser Phe Lys Phe Ile Asn Ala Cys		
[7035]	340	345	350
[7036]	Ala Phe Ala Ser Asn Asn Leu Lys Met Ile Leu Asn Pro Glu Cys Glu		
[7037]	355	360	365
[7038]	Lys Asp Ile Leu Gly Lys Gly Asn Phe Lys Lys Asn Leu Pro Asn Ser		
[7039]	370	375	380
[7040]	Thr Thr Arg Ser Asp Val Val Lys Lys Met Ile Pro Phe Phe Ser Asp		
[7041]	385	390	395 400
[7042]	Glu Leu Gln Asn Val Asn Phe Asp Glu Ala Ile Trp Ala Ile Arg Gly		
[7043]	405	410	415
[7044]	Ser Ile Gln Gln Ile Arg Asn Glu Val Tyr His Cys Lys Lys His Ser		
[7045]	420	425	430
[7046]	Trp Lys Ser Ile Leu Lys Ile Lys Gly Phe Glu Phe Glu Pro Asn Asn		
[7047]	435	440	445
[7048]	Met Lys Tyr Ala Asp Ser Asp Met Gln Lys Leu Met Asp Lys Asp Ile		
[7049]	450	455	460
[7050]	Ala Lys Ile Pro Glu Phe Ile Glu Glu Lys Leu Lys Ser Ser Gly Val		
[7051]	465	470	475 480
[7052]	Val Arg Phe Tyr Arg His Asp Glu Leu Gln Ser Ile Trp Glu Met Lys		
[7053]	485	490	495
[7054]	Gln Gly Phe Ser Leu Leu Thr Thr Asn Ala Pro Phe Val Pro Ser Phe		
[7055]	500	505	510
[7056]	Lys Arg Val Tyr Ala Lys Gly His Asp Tyr Gln Thr Ser Lys Asn Arg		
[7057]	515	520	525

[7058]	Tyr	Tyr	Asn	Leu	Asp	Leu	Thr	Thr	Phe	Asp	Ile	Leu	Glu	Tyr	Gly	Glu
[7059]		530					535					540				
[7060]	Glu	Asp	Phe	Arg	Ala	Arg	Tyr	Phe	Leu	Thr	Lys	Leu	Val	Tyr	Tyr	Gln
[7061]	545					550					555					560
[7062]	Gln	Phe	Met	Pro	Trp	Phe	Thr	Ala	Asp	Asn	Asn	Ala	Phe	Arg	Asp	Ala
[7063]					565					570					575	
[7064]	Ala	Asn	Phe	Val	Leu	Arg	Leu	Asn	Lys	Asn	Arg	Gln	Gln	Asp	Ala	Lys
[7065]				580					585					590		
[7066]	Ala	Phe	Ile	Asn	Ile	Arg	Glu	Val	Glu	Glu	Gly	Glu	Met	Pro	Arg	Asp
[7067]			595					600						605		
[7068]	Tyr	Met	Gly	Tyr	Val	Gln	Gly	Gln	Ile	Ala	Ile	His	Glu	Asp	Ser	Ile
[7069]		610					615					620				
[7070]	Glu	Asp	Thr	Pro	Asn	His	Phe	Glu	Lys	Phe	Ile	Ser	Gln	Val	Phe	Ile
[7071]	625					630					635					640
[7072]	Lys	Gly	Phe	Asp	Arg	His	Met	Arg	Ser	Ala	Asn	Leu	Lys	Phe	Ile	Lys
[7073]					645					650					655	
[7074]	Asn	Pro	Arg	Asn	Gln	Gly	Leu	Glu	Gln	Ser	Glu	Ile	Glu	Glu	Met	Ser
[7075]				660					665						670	
[7076]	Phe	Asp	Ile	Lys	Val	Glu	Pro	Ser	Phe	Leu	Lys	Asn	Lys	Asp	Asp	Tyr
[7077]			675					680					685			
[7078]	Ile	Ala	Phe	Trp	Ile	Phe	Cys	Lys	Met	Leu	Asp	Ala	Arg	His	Leu	Ser
[7079]		690					695					700				
[7080]	Glu	Leu	Arg	Asn	Glu	Met	Ile	Lys	Tyr	Asp	Gly	His	Leu	Thr	Gly	Glu
[7081]	705					710					715					720
[7082]	Gln	Glu	Ile	Ile	Gly	Leu	Ala	Leu	Leu	Gly	Val	Asp	Ser	Arg	Glu	Asn
[7083]				725						730					735	
[7084]	Asp	Trp	Lys	Gln	Phe	Phe	Ser	Ser	Glu	Arg	Glu	Tyr	Glu	Lys	Ile	Met
[7085]				740					745					750		
[7086]	Lys	Gly	Tyr	Val	Val	Glu	Glu	Leu	Tyr	Gln	Arg	Glu	Pro	Tyr	Arg	Gln
[7087]			755					760					765			
[7088]	Ser	Asp	Gly	Lys	Thr	Pro	Ile	Leu	Phe	Arg	Gly	Val	Glu	Gln	Ala	Arg
[7089]		770					775					780				
[7090]	Lys	Tyr	Gly	Thr	Glu	Thr	Val	Ile	Gln	Arg	Leu	Phe	Asp	Ala	Asn	Pro
[7091]	785					790					795					800
[7092]	Glu	Phe	Lys	Val	Ser	Lys	Cys	Asn	Leu	Ala	Glu	Trp	Glu	Arg	Gln	Lys
[7093]				805						810					815	
[7094]	Glu	Thr	Ile	Glu	Glu	Thr	Ile	Lys	Arg	Arg	Lys	Glu	Leu	His	Asn	Glu
[7095]				820					825					830		
[7096]	Trp	Ala	Lys	Asn	Pro	Lys	Lys	Pro	Gln	Asn	Asn	Ala	Phe	Phe	Lys	Glu

[7097]	835	840	845
[7098]	Tyr Lys Glu Cys Cys Asp Ala Ile Asp Ala Tyr Asn Trp His Lys Asn		
[7099]	850	855	860
[7100]	Lys Thr Thr Leu Ala Tyr Val Asn Glu Leu His His Leu Leu Ile Glu		
[7101]	865	870	875
[7102]	Ile Leu Gly Arg Tyr Val Gly Tyr Val Ala Ile Ala Asp Arg Asp Phe		
[7103]	885	890	895
[7104]	Gln Cys Met Ala Asn Gln Tyr Phe Lys His Ser Gly Ile Thr Glu Arg		
[7105]	900	905	910
[7106]	Val Glu Tyr Trp Gly Asp Asn Arg Leu Lys Ser Ile Lys Lys Leu Asp		
[7107]	915	920	925
[7108]	Thr Phe Leu Lys Lys Glu Gly Leu Phe Val Ser Glu Lys Asn Ala Arg		
[7109]	930	935	940
[7110]	Asn His Ile Ala His Leu Asn Tyr Leu Ser Leu Lys Ser Glu Cys Thr		
[7111]	945	950	955
[7112]	Leu Leu Tyr Leu Ser Glu Arg Leu Arg Glu Ile Phe Lys Tyr Asp Arg		
[7113]	965	970	975
[7114]	Lys Leu Lys Asn Ala Val Ser Lys Ser Leu Ile Asp Ile Leu Asp Arg		
[7115]	980	985	990
[7116]	His Gly Met Ser Val Val Phe Ala Asn Leu Lys Glu Asn Lys His Arg		
[7117]	995	1000	1005
[7118]	Leu Val Ile Lys Ser Leu Glu Pro Lys Lys Leu Arg His Leu Gly		
[7119]	1010	1015	1020
[7120]	Gly Lys Lys Ile Asp Gly Gly Tyr Ile Glu Thr Asn Gln Val Ser		
[7121]	1025	1030	1035
[7122]	Glu Glu Tyr Cys Gly Ile Val Lys Arg Leu Leu Glu Met		
[7123]	1040	1045	1050
[7124]	<210> 321		
[7125]	<211> 1340		
[7126]	<212> PRT		
[7127]	<213> 未知的		
[7128]	<220>		
[7129]	<223> 毛螺菌科 (Lachnospiraceae) 细菌MA2020		
[7130]	<400> 321		
[7131]	Met Gln Ile Ser Lys Val Asn His Lys His Val Ala Val Gly Gln Lys		
[7132]	1	5	10
[7133]	Asp Arg Glu Arg Ile Thr Gly Phe Ile Tyr Asn Asp Pro Val Gly Asp		
[7134]	20	25	30
[7135]	Glu Lys Ser Leu Glu Asp Val Val Ala Lys Arg Ala Asn Asp Thr Lys		

[7136]	35	40	45
[7137]	Val Leu Phe Asn Val Phe Asn Thr Lys Asp Leu Tyr Asp Ser Gln Glu		
[7138]	50	55	60
[7139]	Ser Asp Lys Ser Glu Lys Asp Lys Glu Ile Ile Ser Lys Gly Ala Lys		
[7140]	65	70	75
[7141]	Phe Val Ala Lys Ser Phe Asn Ser Ala Ile Thr Ile Leu Lys Lys Gln		
[7142]	85	90	95
[7143]	Asn Lys Ile Tyr Ser Thr Leu Thr Ser Gln Gln Val Ile Lys Glu Leu		
[7144]	100	105	110
[7145]	Lys Asp Lys Phe Gly Gly Ala Arg Ile Tyr Asp Asp Asp Ile Glu Glu		
[7146]	115	120	125
[7147]	Ala Leu Thr Glu Thr Leu Lys Lys Ser Phe Arg Lys Glu Asn Val Arg		
[7148]	130	135	140
[7149]	Asn Ser Ile Lys Val Leu Ile Glu Asn Ala Ala Gly Ile Arg Ser Ser		
[7150]	145	150	155
[7151]	Leu Ser Lys Asp Glu Glu Glu Leu Ile Gln Glu Tyr Phe Val Lys Gln		
[7152]	165	170	175
[7153]	Leu Val Glu Glu Tyr Thr Lys Thr Lys Leu Gln Lys Asn Val Val Lys		
[7154]	180	185	190
[7155]	Ser Ile Lys Asn Gln Asn Met Val Ile Gln Pro Asp Ser Asp Ser Gln		
[7156]	195	200	205
[7157]	Val Leu Ser Leu Ser Glu Ser Arg Arg Glu Lys Gln Ser Ser Ala Val		
[7158]	210	215	220
[7159]	Ser Ser Asp Thr Leu Val Asn Cys Lys Glu Lys Asp Val Leu Lys Ala		
[7160]	225	230	235
[7161]	Phe Leu Thr Asp Tyr Ala Val Leu Asp Glu Asp Glu Arg Asn Ser Leu		
[7162]	245	250	255
[7163]	Leu Trp Lys Leu Arg Asn Leu Val Asn Leu Tyr Phe Tyr Gly Ser Glu		
[7164]	260	265	270
[7165]	Ser Ile Arg Asp Tyr Ser Tyr Thr Lys Glu Lys Ser Val Trp Lys Glu		
[7166]	275	280	285
[7167]	His Asp Glu Gln Lys Ala Asn Lys Thr Leu Phe Ile Asp Glu Ile Cys		
[7168]	290	295	300
[7169]	His Ile Thr Lys Ile Gly Lys Asn Gly Lys Glu Gln Lys Val Leu Asp		
[7170]	305	310	315
[7171]	Tyr Glu Glu Asn Arg Ser Arg Cys Arg Lys Gln Asn Ile Asn Tyr Tyr		
[7172]	325	330	335
[7173]	Arg Ser Ala Leu Asn Tyr Ala Lys Asn Asn Thr Ser Gly Ile Phe Glu		
[7174]	340	345	350

[7175]	Asn	Glu	Asp	Ser	Asn	His	Phe	Trp	Ile	His	Leu	Ile	Glu	Asn	Glu	Val
[7176]				355				360					365			
[7177]	Glu	Arg	Leu	Tyr	Asn	Gly	Ile	Glu	Asn	Gly	Glu	Glu	Phe	Lys	Phe	Glu
[7178]			370				375					380				
[7179]	Thr	Gly	Tyr	Ile	Ser	Glu	Lys	Val	Trp	Lys	Ala	Val	Ile	Asn	His	Leu
[7180]	385					390				395						400
[7181]	Ser	Ile	Lys	Tyr	Ile	Ala	Leu	Gly	Lys	Ala	Val	Tyr	Asn	Tyr	Ala	Met
[7182]				405					410					415		
[7183]	Lys	Glu	Leu	Ser	Ser	Pro	Gly	Asp	Ile	Glu	Pro	Gly	Lys	Ile	Asp	Asp
[7184]			420					425					430			
[7185]	Ser	Tyr	Ile	Asn	Gly	Ile	Thr	Ser	Phe	Asp	Tyr	Glu	Ile	Ile	Lys	Ala
[7186]			435				440					445				
[7187]	Glu	Glu	Ser	Leu	Gln	Arg	Asp	Ile	Ser	Met	Asn	Val	Val	Phe	Ala	Thr
[7188]		450				455					460					
[7189]	Asn	Tyr	Leu	Ala	Cys	Ala	Thr	Val	Asp	Thr	Asp	Lys	Asp	Phe	Leu	Leu
[7190]	465				470				475							480
[7191]	Phe	Ser	Lys	Glu	Asp	Ile	Arg	Ser	Cys	Thr	Lys	Lys	Asp	Gly	Asn	Leu
[7192]			485					490					495			
[7193]	Cys	Lys	Asn	Ile	Met	Gln	Phe	Trp	Gly	Gly	Tyr	Ser	Thr	Trp	Lys	Asn
[7194]			500					505					510			
[7195]	Phe	Cys	Glu	Glu	Tyr	Leu	Lys	Asp	Asp	Lys	Asp	Ala	Leu	Glu	Leu	Leu
[7196]			515				520					525				
[7197]	Tyr	Ser	Leu	Lys	Ser	Met	Leu	Tyr	Ser	Met	Arg	Asn	Ser	Ser	Phe	His
[7198]		530				535					540					
[7199]	Phe	Ser	Thr	Glu	Asn	Val	Asp	Asn	Gly	Ser	Trp	Asp	Thr	Glu	Leu	Ile
[7200]	545				550					555						560
[7201]	Gly	Lys	Leu	Phe	Glu	Glu	Asp	Cys	Asn	Arg	Ala	Ala	Arg	Ile	Glu	Lys
[7202]			565					570					575			
[7203]	Glu	Lys	Phe	Tyr	Asn	Asn	Asn	Leu	His	Met	Phe	Tyr	Ser	Ser	Ser	Leu
[7204]			580					585					590			
[7205]	Leu	Glu	Lys	Val	Leu	Glu	Arg	Leu	Tyr	Ser	Ser	His	His	Glu	Arg	Ala
[7206]			595				600					605				
[7207]	Ser	Gln	Val	Pro	Ser	Phe	Asn	Arg	Val	Phe	Val	Arg	Lys	Asn	Phe	Pro
[7208]		610				615					620					
[7209]	Ser	Ser	Leu	Ser	Glu	Gln	Arg	Ile	Thr	Pro	Lys	Phe	Thr	Asp	Ser	Lys
[7210]	625				630					635						640
[7211]	Asp	Glu	Gln	Ile	Trp	Gln	Ser	Ala	Val	Tyr	Tyr	Leu	Cys	Lys	Glu	Ile
[7212]			645					650				655				
[7213]	Tyr	Tyr	Asn	Asp	Phe	Leu	Gln	Ser	Lys	Glu	Ala	Tyr	Lys	Leu	Phe	Arg

[7214]		660		665		670
[7215]	Glu Gly Val Lys Asn Leu Asp Lys Asn Asp Ile Asn Asn Gln Lys Ala					
[7216]		675		680		685
[7217]	Ala Asp Ser Phe Lys Gln Ala Val Val Tyr Tyr Gly Lys Ala Ile Gly					
[7218]		690		695		700
[7219]	Asn Ala Thr Leu Ser Gln Val Cys Gln Ala Ile Met Thr Glu Tyr Asn					
[7220]	705		710		715	720
[7221]	Arg Gln Asn Asn Asp Gly Leu Lys Lys Lys Ser Ala Tyr Ala Glu Lys					
[7222]		725		730		735
[7223]	Gln Asn Ser Asn Lys Tyr Lys His Tyr Pro Leu Phe Leu Lys Gln Val					
[7224]		740		745		750
[7225]	Leu Gln Ser Ala Phe Trp Glu Tyr Leu Asp Glu Asn Lys Glu Ile Tyr					
[7226]		755		760		765
[7227]	Gly Phe Ile Ser Ala Gln Ile His Lys Ser Asn Val Glu Ile Lys Ala					
[7228]		770		775		780
[7229]	Glu Asp Phe Ile Ala Asn Tyr Ser Ser Gln Gln Tyr Lys Lys Leu Val					
[7230]	785		790		795	800
[7231]	Asp Lys Val Lys Lys Thr Pro Glu Leu Gln Lys Trp Tyr Thr Leu Gly					
[7232]		805		810		815
[7233]	Arg Leu Ile Asn Pro Arg Gln Ala Asn Gln Phe Leu Gly Ser Ile Arg					
[7234]		820		825		830
[7235]	Asn Tyr Val Gln Phe Val Lys Asp Ile Gln Arg Arg Ala Lys Glu Asn					
[7236]		835		840		845
[7237]	Gly Asn Pro Ile Arg Asn Tyr Tyr Glu Val Leu Glu Ser Asp Ser Ile					
[7238]		850		855		860
[7239]	Ile Lys Ile Leu Glu Met Cys Thr Lys Leu Asn Gly Thr Thr Ser Asn					
[7240]	865		870		875	880
[7241]	Asp Ile His Asp Tyr Phe Arg Asp Glu Asp Glu Tyr Ala Glu Tyr Ile					
[7242]		885		890		895
[7243]	Ser Gln Phe Val Asn Phe Gly Asp Val His Ser Gly Ala Ala Leu Asn					
[7244]		900		905		910
[7245]	Ala Phe Cys Asn Ser Glu Ser Glu Gly Lys Lys Asn Gly Ile Tyr Tyr					
[7246]		915		920		925
[7247]	Asp Gly Ile Asn Pro Ile Val Asn Arg Asn Trp Val Leu Cys Lys Leu					
[7248]		930		935		940
[7249]	Tyr Gly Ser Pro Asp Leu Ile Ser Lys Ile Ile Ser Arg Val Asn Glu					
[7250]	945		950		955	960
[7251]	Asn Met Ile His Asp Phe His Lys Gln Glu Asp Leu Ile Arg Glu Tyr					
[7252]		965		970		975

[7253]	Gln Ile Lys Gly Ile Cys Ser Asn Lys Lys Glu Gln Gln Asp Leu Arg		
[7254]		980	985 990
[7255]	Thr Phe Gln Val Leu Lys Asn Arg Val Glu Leu Arg Asp Ile Val Glu		
[7256]		995	1000 1005
[7257]	Tyr Ser Glu Ile Ile Asn Glu Leu Tyr Gly Gln Leu Ile Lys Trp		
[7258]		1010	1015 1020
[7259]	Cys Tyr Leu Arg Glu Arg Asp Leu Met Tyr Phe Gln Leu Gly Phe		
[7260]		1025	1030 1035
[7261]	His Tyr Leu Cys Leu Asn Asn Ala Ser Ser Lys Glu Ala Asp Tyr		
[7262]		1040	1045 1050
[7263]	Ile Lys Ile Asn Val Asp Asp Arg Asn Ile Ser Gly Ala Ile Leu		
[7264]		1055	1060 1065
[7265]	Tyr Gln Ile Ala Ala Met Tyr Ile Asn Gly Leu Pro Val Tyr Tyr		
[7266]		1070	1075 1080
[7267]	Lys Lys Asp Asp Met Tyr Val Ala Leu Lys Ser Gly Lys Lys Ala		
[7268]		1085	1090 1095
[7269]	Ser Asp Glu Leu Asn Ser Asn Glu Gln Thr Ser Lys Lys Ile Asn		
[7270]		1100	1105 1110
[7271]	Tyr Phe Leu Lys Tyr Gly Asn Asn Ile Leu Gly Asp Lys Lys Asp		
[7272]		1115	1120 1125
[7273]	Gln Leu Tyr Leu Ala Gly Leu Glu Leu Phe Glu Asn Val Ala Glu		
[7274]		1130	1135 1140
[7275]	His Glu Asn Ile Ile Ile Phe Arg Asn Glu Ile Asp His Phe His		
[7276]		1145	1150 1155
[7277]	Tyr Phe Tyr Asp Arg Asp Arg Ser Met Leu Asp Leu Tyr Ser Glu		
[7278]		1160	1165 1170
[7279]	Val Phe Asp Arg Phe Phe Thr Tyr Asp Met Lys Leu Arg Lys Asn		
[7280]		1175	1180 1185
[7281]	Val Val Asn Met Leu Tyr Asn Ile Leu Leu Asp His Asn Ile Val		
[7282]		1190	1195 1200
[7283]	Ser Ser Phe Val Phe Glu Thr Gly Glu Lys Lys Val Gly Arg Gly		
[7284]		1205	1210 1215
[7285]	Asp Ser Glu Val Ile Lys Pro Ser Ala Lys Ile Arg Leu Arg Ala		
[7286]		1220	1225 1230
[7287]	Asn Asn Gly Val Ser Ser Asp Val Phe Thr Tyr Lys Val Gly Ser		
[7288]		1235	1240 1245
[7289]	Lys Asp Glu Leu Lys Ile Ala Thr Leu Pro Ala Lys Asn Glu Glu		
[7290]		1250	1255 1260
[7291]	Phe Leu Leu Asn Val Ala Arg Leu Ile Tyr Tyr Pro Asp Met Glu		

[7292]	1265	1270	1275
[7293]	Ala Val Ser Glu Asn Met Val Arg Glu Gly Val Val Lys Val Glu		
[7294]	1280	1285	1290
[7295]	Lys Ser Asn Asp Lys Lys Gly Lys Ile Ser Arg Gly Ser Asn Thr		
[7296]	1295	1300	1305
[7297]	Arg Ser Ser Asn Gln Ser Lys Tyr Asn Asn Lys Ser Lys Asn Arg		
[7298]	1310	1315	1320
[7299]	Met Asn Tyr Ser Met Gly Ser Ile Phe Glu Lys Met Asp Leu Lys		
[7300]	1325	1330	1335
[7301]	Phe Asp		
[7302]	1340		
[7303]	<210> 322		
[7304]	<211> 913		
[7305]	<212> PRT		
[7306]	<213> 未知的		
[7307]	<220>		
[7308]	<223> 纤毛菌物种(Leptotrichia sp.)口腔分类225		
[7309]	<400> 322		
[7310]	Met Lys Glu Leu Ile Glu Lys Val Pro Asn Val Ser Glu Leu Lys Lys		
[7311]	1 5 10 15		
[7312]	Ser Gln Val Phe Tyr Lys Tyr Tyr Leu Asn Lys Glu Lys Leu Asn Asp		
[7313]	20 25 30		
[7314]	Glu Asn Ile Lys Tyr Val Phe Cys His Phe Val Glu Ile Glu Met Ser		
[7315]	35 40 45		
[7316]	Lys Leu Leu Lys Asn Tyr Val Tyr Lys Lys Pro Ser Asn Ile Ser Asn		
[7317]	50 55 60		
[7318]	Asp Lys Val Lys Arg Ile Phe Glu Tyr Gln Ser Leu Lys Lys Leu Ile		
[7319]	65 70 75 80		
[7320]	Glu Asn Lys Leu Leu Asn Lys Leu Asp Thr Tyr Ile Arg Asn Cys Gly		
[7321]	85 90 95		
[7322]	Lys Tyr Ser Phe Tyr Leu Gln Asp Gly Glu Ile Ala Thr Ser Asp Phe		
[7323]	100 105 110		
[7324]	Ile Val Gly Asn Arg Gln Asn Glu Ala Phe Leu Arg Asn Ile Ile Gly		
[7325]	115 120 125		
[7326]	Val Ser Ser Ala Ala Tyr Phe Ser Leu Arg Asn Ile Leu Glu Thr Glu		
[7327]	130 135 140		
[7328]	Asn Glu Asn Asp Ile Thr Gly Lys Met Arg Gly Lys Thr Val Lys Asn		
[7329]	145 150 155 160		
[7330]	Lys Lys Gly Glu Glu Lys Tyr Ile Ser Gly Glu Ile Asp Lys Leu Tyr		

[7331]	165										170				175			
[7332]	Asp	Asn	Asn	Lys	Gln	Asn	Glu	Val	Lys	Lys	Asn	Leu	Lys	Met	Phe	Tyr		
[7333]	180										185				190			
[7334]	Ser	Tyr	Asp	Phe	Asn	Met	Asn	Ser	Lys	Lys	Glu	Ile	Glu	Asp	Phe	Phe		
[7335]	195										200				205			
[7336]	Ser	Asn	Ile	Asp	Glu	Ala	Ile	Ser	Ser	Ile	Arg	His	Gly	Ile	Val	His		
[7337]	210										215				220			
[7338]	Phe	Asn	Leu	Glu	Leu	Glu	Gly	Lys	Asp	Ile	Phe	Thr	Phe	Lys	Asn	Ile		
[7339]	225	230										235				240		
[7340]	Val	Pro	Ser	Gln	Ile	Ser	Lys	Lys	Met	Phe	His	Asp	Glu	Ile	Asn	Glu		
[7341]	245										250				255			
[7342]	Lys	Lys	Leu	Lys	Leu	Lys	Ile	Phe	Lys	Gln	Leu	Asn	Ser	Ala	Asn	Val		
[7343]	260										265				270			
[7344]	Phe	Arg	Tyr	Leu	Glu	Lys	Tyr	Lys	Ile	Leu	Asn	Tyr	Leu	Asn	Arg	Thr		
[7345]	275										280				285			
[7346]	Arg	Phe	Glu	Phe	Val	Asn	Lys	Asn	Ile	Pro	Phe	Val	Pro	Ser	Phe	Thr		
[7347]	290										295				300			
[7348]	Lys	Leu	Tyr	Ser	Arg	Ile	Asp	Asp	Leu	Lys	Asn	Ser	Leu	Cys	Ile	Tyr		
[7349]	305	310										315				320		
[7350]	Trp	Lys	Ile	Pro	Lys	Ala	Asn	Asp	Asn	Asn	Lys	Thr	Lys	Glu	Ile	Thr		
[7351]	325										330				335			
[7352]	Asp	Ala	Gln	Ile	Tyr	Leu	Leu	Lys	Asn	Ile	Tyr	Tyr	Ser	Glu	Phe	Leu		
[7353]	340										345				350			
[7354]	Asn	Tyr	Phe	Met	Ser	Asn	Asn	Gly	Asn	Phe	Phe	Glu	Ile	Ile	Lys	Glu		
[7355]	355										360				365			
[7356]	Ile	Ile	Glu	Leu	Asn	Lys	Asn	Asp	Lys	Arg	Asn	Leu	Lys	Thr	Gly	Phe		
[7357]	370										375				380			
[7358]	Tyr	Lys	Leu	Gln	Lys	Phe	Glu	Asn	Leu	Gln	Glu	Lys	Thr	Pro	Lys	Glu		
[7359]	385	390										395				400		
[7360]	Tyr	Leu	Ala	Asn	Ile	Gln	Ser	Phe	Tyr	Met	Ile	Asp	Ala	Gly	Asn	Lys		
[7361]	405										410				415			
[7362]	Asp	Glu	Glu	Glu	Lys	Asp	Ala	Tyr	Ile	Asp	Phe	Ile	Gln	Lys	Ile	Phe		
[7363]	420										425				430			
[7364]	Leu	Lys	Gly	Phe	Met	Thr	Tyr	Leu	Ala	Asn	Asn	Gly	Arg	Leu	Ser	Leu		
[7365]	435										440				445			
[7366]	Met	Tyr	Ile	Gly	Asn	Asp	Glu	Gln	Ile	Asn	Thr	Ser	Leu	Ala	Glu	Lys		
[7367]	450										455				460			
[7368]	Lys	Gln	Glu	Phe	Asp	Lys	Phe	Leu	Lys	Lys	Tyr	Glu	Gln	Asn	Asn	Asn		
[7369]	465	470										475				480		

[7370]	Ile Lys Ile Pro Tyr Glu Ile Asn Glu Phe Leu Arg Glu Ile Lys Leu
[7371]	485 490 495
[7372]	Gly Asn Ile Leu Lys Tyr Thr Glu Arg Leu Asn Met Phe Tyr Leu Ile
[7373]	500 505 510
[7374]	Leu Lys Leu Leu Asn His Lys Glu Leu Thr Asn Leu Lys Gly Ser Leu
[7375]	515 520 525
[7376]	Glu Lys Tyr Gln Ser Ala Asn Lys Glu Glu Ala Phe Ser Asp Gln Leu
[7377]	530 535 540
[7378]	Glu Leu Ile Asn Leu Leu Asn Leu Asp Asn Asn Arg Val Thr Glu Asp
[7379]	545 550 555 560
[7380]	Phe Glu Leu Glu Ala Asp Glu Ile Gly Lys Phe Leu Asp Phe Asn Gly
[7381]	565 570 575
[7382]	Asn Lys Val Lys Asp Asn Lys Glu Leu Lys Lys Phe Asp Thr Asn Lys
[7383]	580 585 590
[7384]	Ile Tyr Phe Asp Gly Glu Asn Ile Ile Lys His Arg Ala Phe Tyr Asn
[7385]	595 600 605
[7386]	Ile Lys Lys Tyr Gly Met Leu Asn Leu Leu Glu Lys Ile Ser Asp Glu
[7387]	610 615 620
[7388]	Ala Lys Tyr Lys Ile Ser Ile Glu Glu Leu Lys Asn Tyr Ser Asn Lys
[7389]	625 630 635 640
[7390]	Lys Asn Glu Ile Glu Lys Asn His Thr Asn Gln Glu Asn Leu His Arg
[7391]	645 650 655
[7392]	Lys Tyr Ala Arg Pro Arg Lys Asp Glu Lys Phe Asn Asp Glu Asp Tyr
[7393]	660 665 670
[7394]	Lys Lys Tyr Glu Lys Ala Ile Arg Asn Ile Gln Gln Tyr Thr His Leu
[7395]	675 680 685
[7396]	Lys Asn Lys Val Glu Phe Asn Glu Leu Asn Leu Leu Gln Ser Leu Leu
[7397]	690 695 700
[7398]	Leu Arg Ile Leu His Arg Leu Val Gly Tyr Thr Ser Ile Trp Glu Arg
[7399]	705 710 715 720
[7400]	Asp Leu Arg Phe Arg Leu Lys Gly Glu Phe Pro Glu Asn Gln Tyr Ile
[7401]	725 730 735
[7402]	Glu Glu Ile Phe Asn Phe Asn Asn Ser Lys Asn Val Lys Tyr Lys Asn
[7403]	740 745 750
[7404]	Gly Gln Ile Val Glu Lys Tyr Ile Ser Phe Tyr Lys Glu Leu Tyr Lys
[7405]	755 760 765
[7406]	Asp Asp Thr Glu Lys Ile Ser Ile Tyr Ser Asp Lys Lys Val Lys Glu
[7407]	770 775 780
[7408]	Leu Lys Lys Glu Lys Lys Asp Leu Tyr Ile Arg Asn Tyr Ile Ala His

[7409]	785	790	795	800
[7410]	Phe Asn Tyr Ile Pro Asn Ala Glu Ile Ser Leu Leu Glu Val Leu Glu			
[7411]		805	810	815
[7412]	Asn Leu Arg Lys Leu Leu Ser Tyr Asp Arg Lys Leu Lys Asn Ala Ile			
[7413]		820	825	830
[7414]	Met Lys Ser Ile Val Asp Ile Leu Lys Glu Tyr Gly Phe Val Val Thr			
[7415]		835	840	845
[7416]	Phe Lys Ile Glu Lys Asp Lys Lys Ile Arg Ile Glu Ser Leu Lys Ser			
[7417]		850	855	860
[7418]	Glu Glu Val Val His Leu Lys Lys Leu Lys Leu Lys Asp Asn Asp Lys			
[7419]	865	870	875	880
[7420]	Lys Lys Glu Pro Ile Lys Thr Tyr Arg Asn Ser Lys Glu Leu Cys Glu			
[7421]		885	890	895
[7422]	Leu Val Lys Val Met Phe Glu Tyr Lys Met Lys Glu Lys Lys Ser Glu			
[7423]		900	905	910
[7424]	Asn			
[7425]	<210>	323		
[7426]	<211>	22		
[7427]	<212>	RNA		
[7428]	<213>	人工序列		
[7429]	<220>			
[7430]	<223>	合成序列		
[7431]	<400>	323		
[7432]	gccgccaagc uucacagcgc gu			

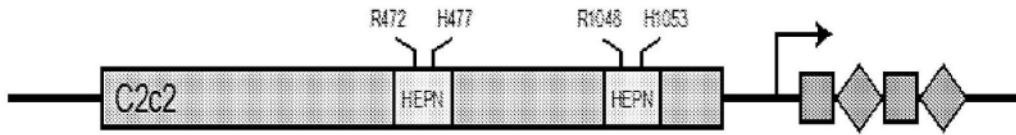


图1A

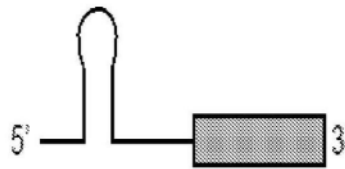


图1B

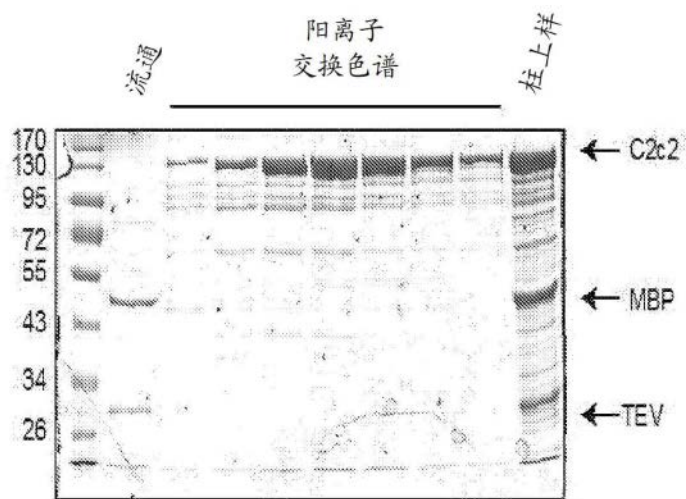


图1C

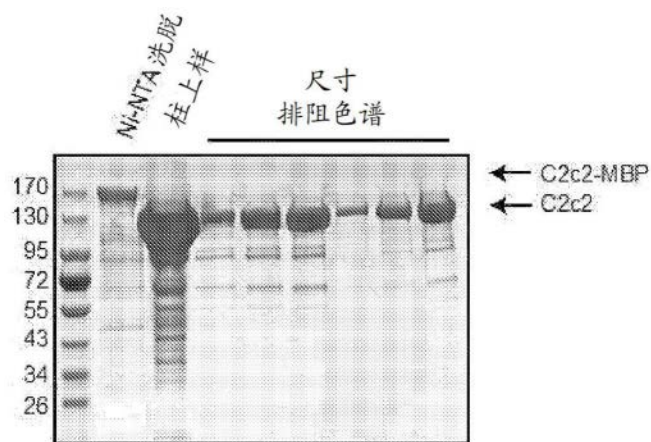


图1D

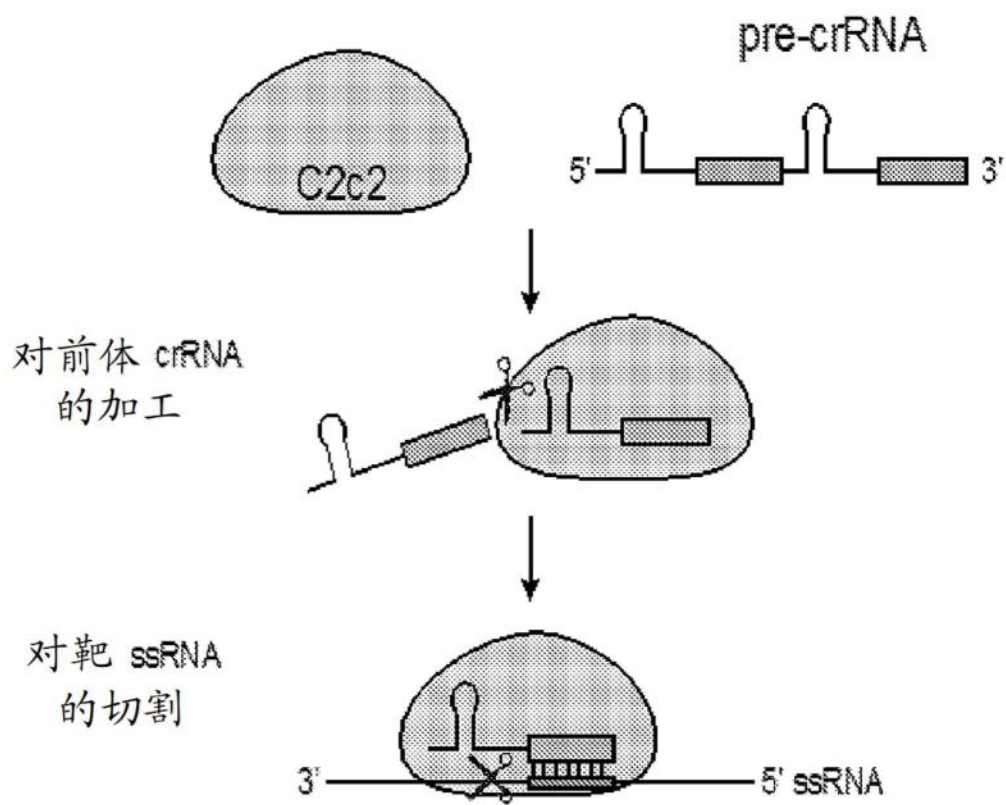


图1E

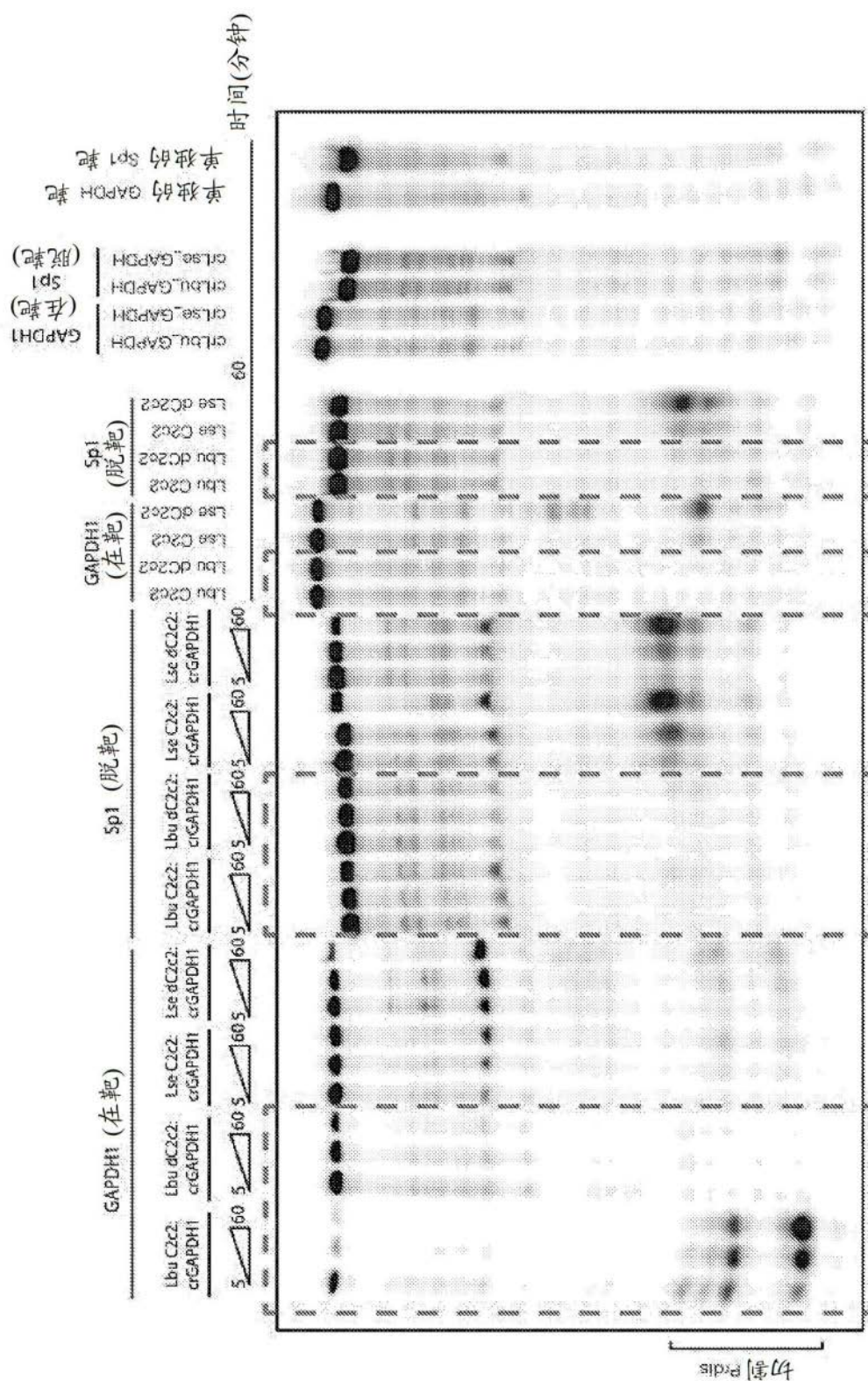


图2

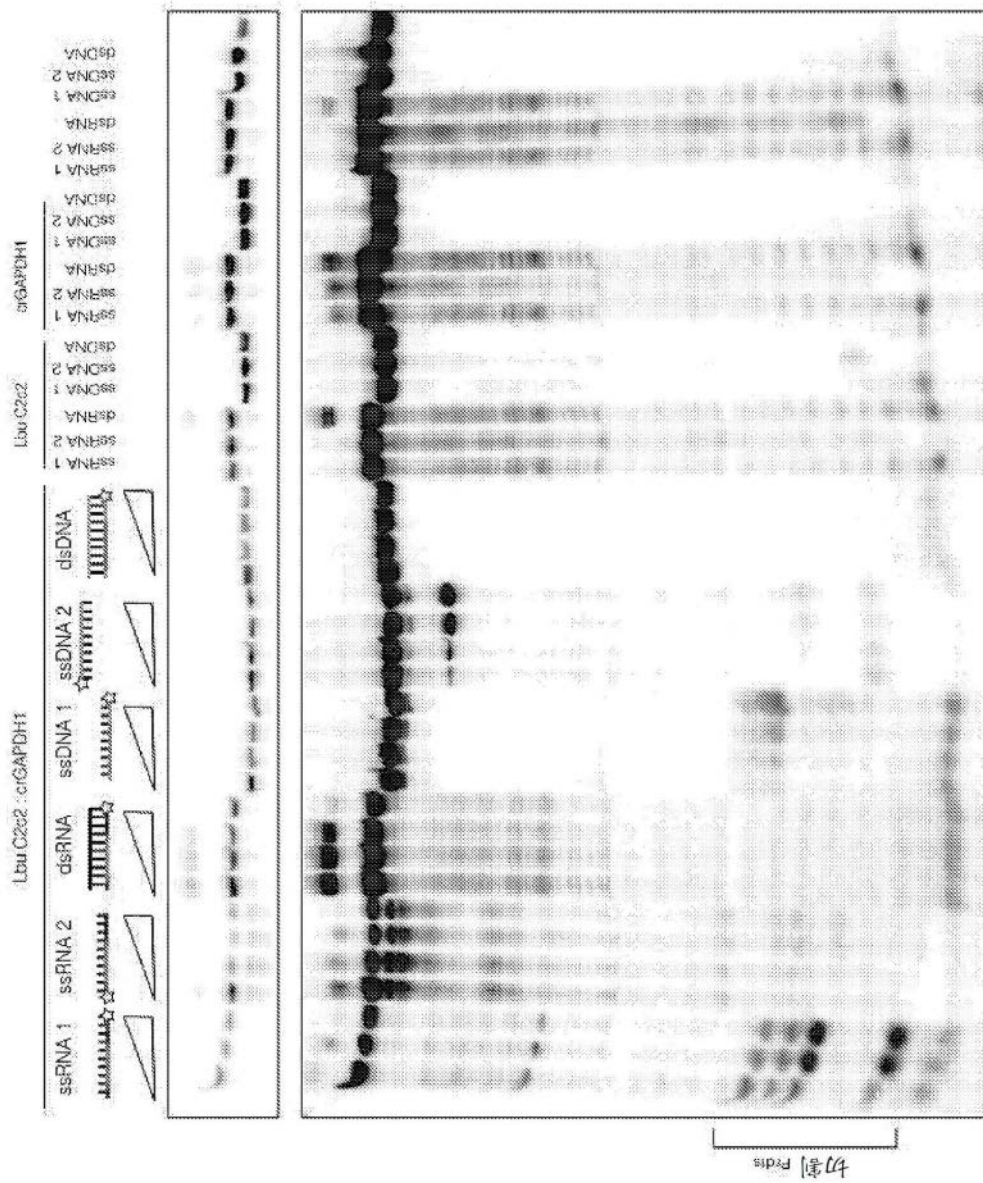


图3

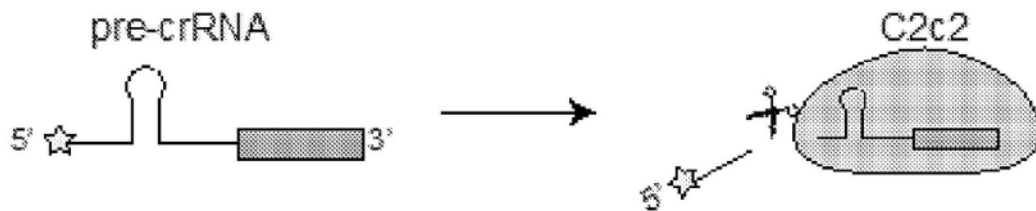


图4A

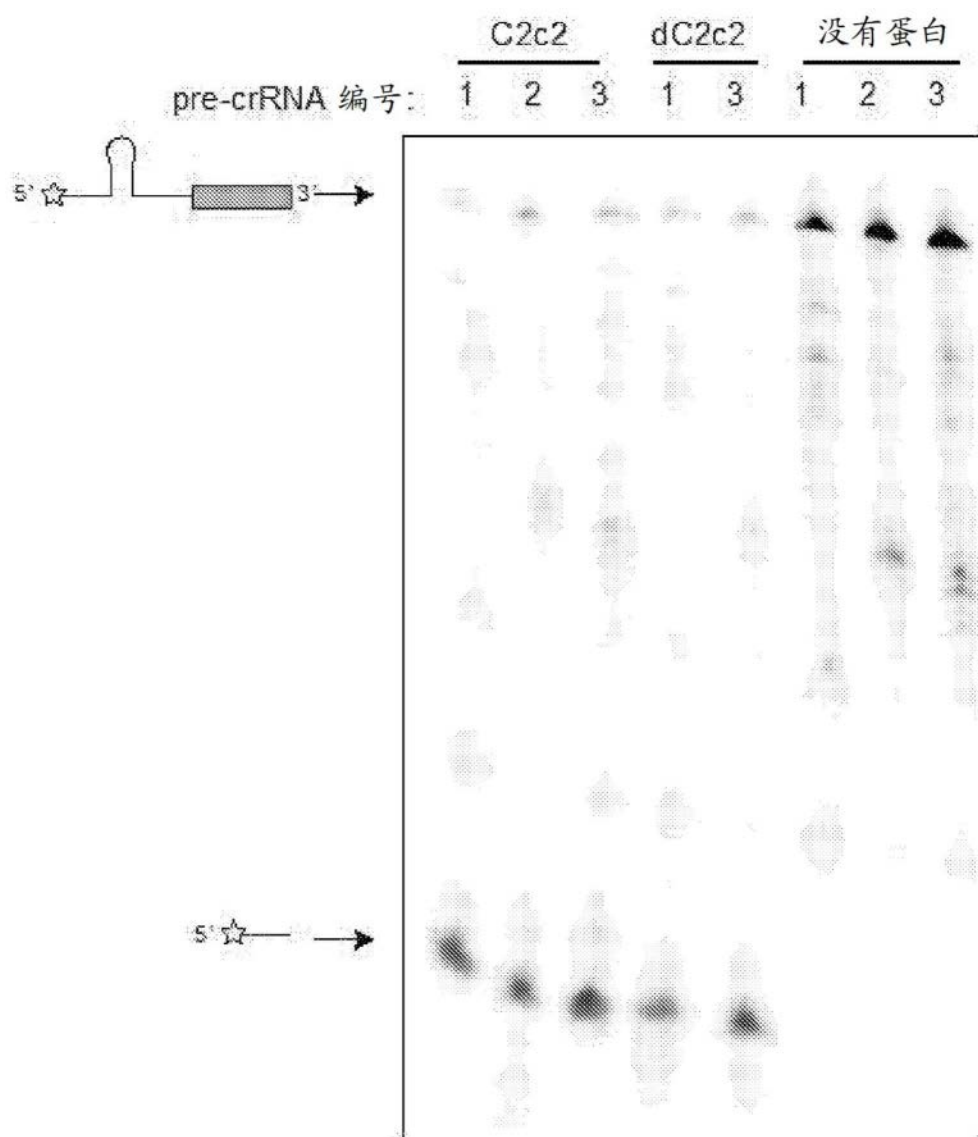


图4B

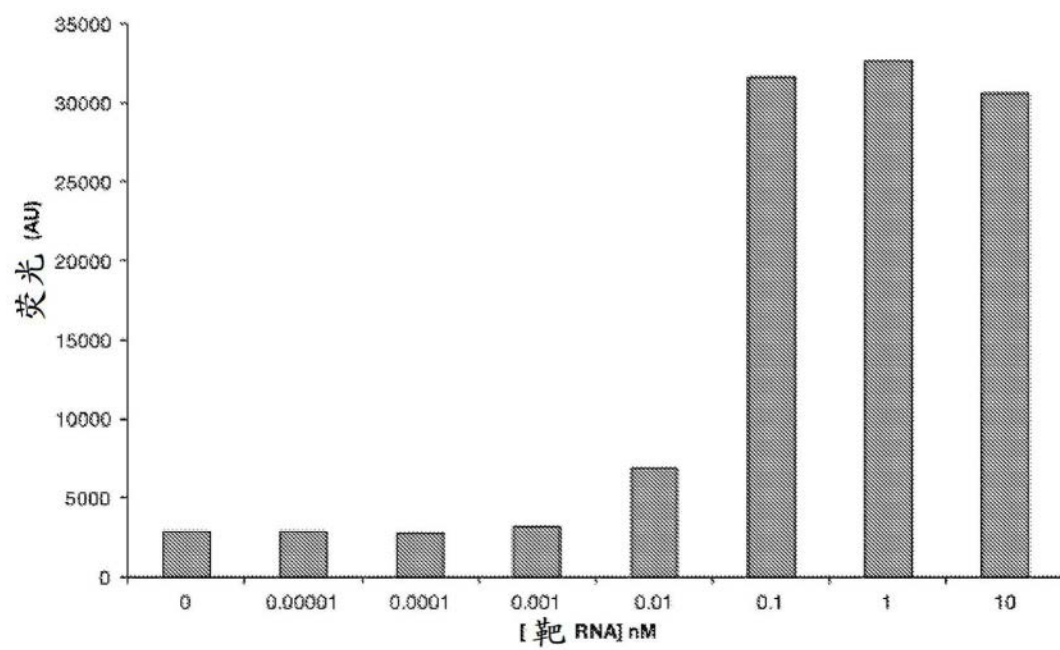


图5

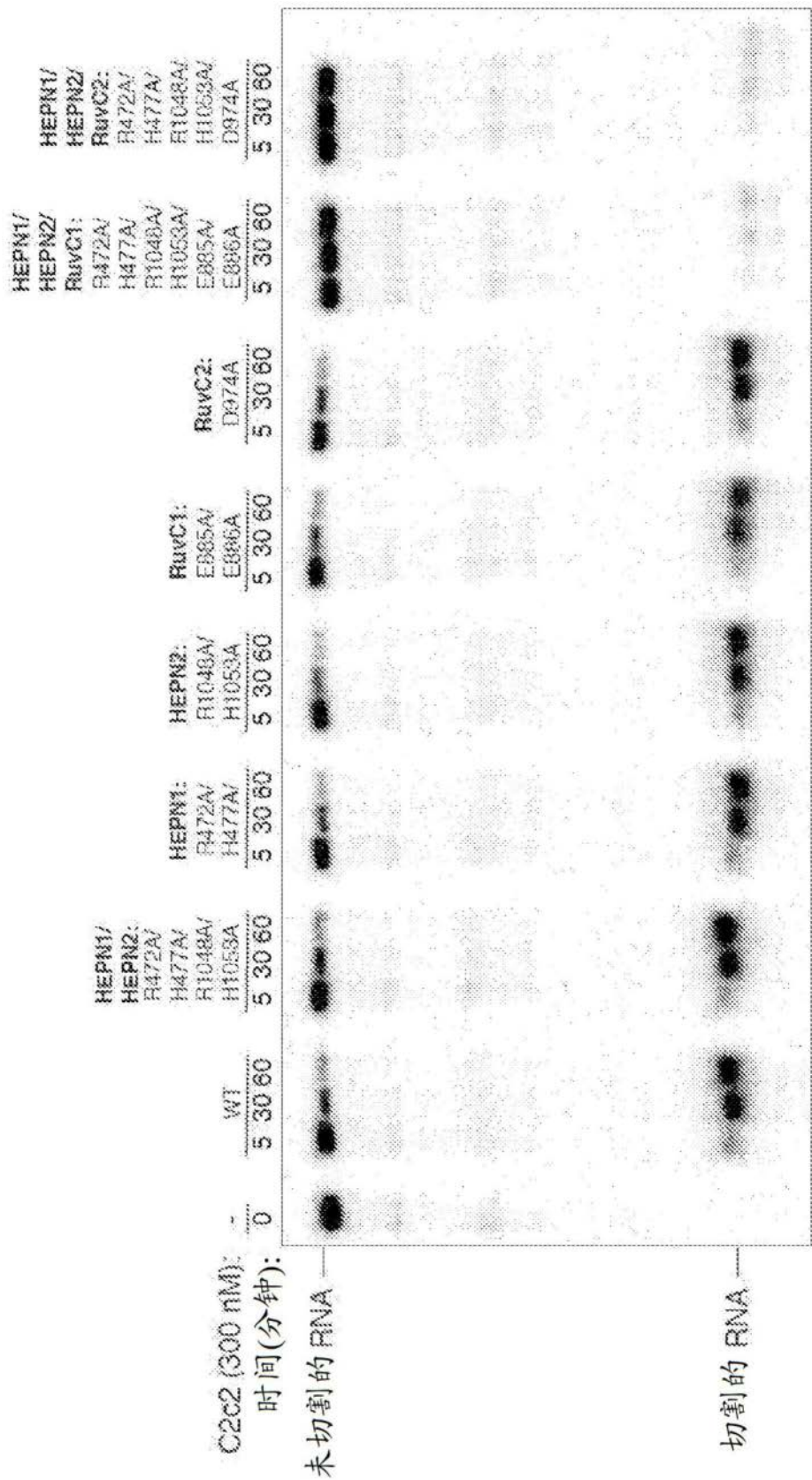


图6

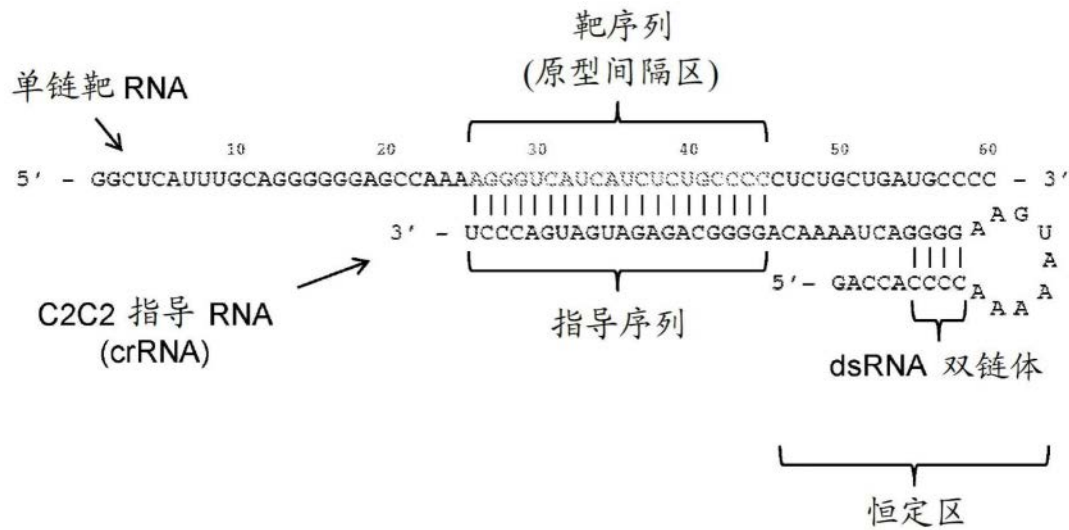


图7A

成熟 crRNA Lse: GACUACCUCUAUAUGAAAGAGGACUAAAACCAAACAUGAUCUGGGUCAUC
pre-crRNA Lse: UAAGAGACUACCUCUAUAUGAAAGAGGACUAAAACCAAACAUGAUCUGGGUCAUC

成熟 crRNA Lsh: CCACCCCAUAUUCGAAGGGGACUAAAACAGGGGCAGAGAUGAUGACCCU
pre-crRNA Lsh: GGAUUUAGACCACCCCAUAUUCGAAGGGGACUAAAACAGGGGCAGAGAUGAUGACCCU

成熟 crRNA Lbu: GACCACCCCAAAAAUGAAGGGGACUAAAACAGGGGCAGAGAUGAUGACCCU
pre-crRNA Lbu: AUUUAGACCACCCCAAAAAUGAAGGGGACUAAAACAGGGGCAGAGAUGAUGACCCU

图7B

> 斯氏李斯特菌 (WP_012985477.1)
 MWISIKTLIHHLGVLFCDYMYNRRREKKIEVKTMRITKVEVDRKKVLISRDKNKGKLVY
 ENEMQDNTEQIMHHKKSSFYKSVVNKTI CRPEQKQMKLVHGLLQENSQEKIKVSDVTKL
 NISNFLNHRFKKSLYFPENSPDKSEERYIEINLSQLLEDSLKKQQTFCWESFSKDME
 LYINWAENYISSKTKLIKKSIRNNRIQSTESRSGQLMDRYMKDILNKNKPFDIQSVSEKY
 QLEKLTSAKATFKEAKKNDKEINYLKSTLQNHQRQIEELKENSELNQNFIEIRKHL
 TYFPIKKTNRKVGDIRNLEIGEIQKIVNHRKKNKIVQRILQEGKLASYEIESTVNSNSLQ
 KIKIEEAFALKFINACLFASNNLRNMVYPVCKKDI LMIGEFKNSFKEIKHKKFIRQWSQF
 FSQEITVDDIELASWGLRGAIAPIRNEIIHLKKHSWKKFFNNPTFKVKKSKIINGKTKDV
 TSEFLYKETL FKDYFYSELDSPPEL I INKMESSKILDYSSDQLNQVFTIPNFELSLLTS
 AVPFAPSFKRVYLGFDYQNQDEAQPDYNLKLNIYNEKAFNSEAFQAQYSLFKMVYYQVF
 LPQFTTNNDL FKSSVDFILTLNKRKGYAKAFQDIRKMNKDEKPSYMSYIQSOLMLYQK
 KQEEKEKINHFEKFINQVFIKGFNSFIEKNRLTYICHPTKNTVPENDNIEIPFHTDMDD
 NIAFWLMCKLLDAKQSELNEMIKFSCSLQSTEEISTFTKAREVIGLALLNGEKGENDW
 KELFDDKEAWKKNMSLYVSEELLQSLPYTQEDGQTPVINRSIDL VKKYGTETILEKLFSS
 SDDYKVSADIAKLHEYDVTEKIAQQESLHKQWIEKPGGLARDSAWTKKYQNVINDISNYQ
 WAKTKVELTQVRHLHQLTIDLLSRLAGYMSIADRDQFSSNYILERENSEYRVTSWILLS
 ENKNKNYNDYELYNLKNASIKVSSKNDPQLKVDLKQLRLTLEYLELFDNRLKEKRNIS
 HFNYLNGQLGNSILELFD DARDVLSYDRKLNKNAVSKSLKEILSSHGMEVTFKPLYQTNNH
 LKIDKLQPKKIHHLGEKSTVSSNQVSNEYCQLVRTLLTMK (SEQ ID NO: 1)

图8A

> 口腔纤毛菌 (WP_015770004.1)

MKVTKVGGISHKKYTSEGRLVKSEENRTDERLSALLNMRDMYIKNPSSSTETKENQKR
 IGKLLKFFSNKVMVYLKDNLTSLKNGKKENIDREYSETDILESVDVRDKKNFAVLKKIYLN
 NVNSEELVFRNDIKKKLNKINSLSKYSEKKNANYQKINENNIEKVEGSKRNIIDYYR
 ESAKRDAYVSNVKEAFDKLYKEEDIAKLVLIEIENLTKLEKYKIREFYHEIIGRKNDKENF
 AKIIYEEIQNVNNMKELIEKVPDMSELKKSQVFYKYQLDKEELNDKNIKYAFCHFVEIEM
 SQLLKNVYKRLSNISNDKIKRIFEFYQNLKLIENKLLNKLDYVVRNCGYNYLQDGEI
 ATSDFIARNRQNEAFLRNIIGVSSVAYFSLRNIILETENENDITGRMRGKTVKNNKGEEKY
 VSGEVDKIYNENKKNEVKENLKMFSYDFNMDNKNEIEDFFANIDEAIISSIRHGIVHFNL
 ELEGKDIFAFKNIAPSEISKMFQNEINEKKLKLKIFRQLNSANVFYRYLEKYKIILNYLKR
 TRFEFVNKNIPFVPSFTKLYSRIDDLKNSLGIYWKTPKTNDNDNKTKEIIDAQIYLLKNIY
 YGEFLNYFMSNNGNFFEISKEIIELNKNDKRNLTGTGYKLQKFEDIQEKIPKEYLANIQS
 LYMINAGNQDEEEKDTYIDFIQKIFLKGFTYLANNGRSLIYIGSDEETNTSLAEKKQE
 FDKFLKKYEQNNNIKIPYEINEFLREIKLGNILKYTERLNMFYLIKLNLHKLTLNLKGS
 LEKYQSANKEEAFSDQLELINLNDNNRVTEDFELEADEIGKFLDFNGNKVKDNKELKK
 FDTNKIYFDGENIKHRAFYNIKKYGMNLNLEKIAADKAGYKISIEELKKYSNKKNEIEKN
 HKMQENLHRKYARPRKDEKFTDEDYESYKQAIENIEEYTHLKNKVEFNELNLQGLLRI
 LHRLVGYTSIWERDLRFLKGEPENQYIEEINFENKNVYKGGQIVEKYIKFYKELH
 QNDEVKINKYSSANIKVLKQEKKDLYIRNYIAHFNYIPHAIEISLLEVLNLRKLLSYDRK
 LKNVAMKSVVDILKEYGFVATFKIGADKKIGIQTLESEKIVHLKNLKKKLTDRNSEEL
 CKLVKIMFEYKMEKKSEN (SEQ ID NO: 2)

图8B

> 沙氏纤毛菌 (WP_018451595.1)

MGNLFGHKRWYEVDRDKDFKIKRKVKVKNYDGNKYILNINENNKEKIDNNKFIRKYIN
YKNDNIIKEFTRKFHAGNIIFLKLGKEGIIRIENDDFLETEEVVLYIEAYGKSEKLKA
LGITKKKIIIDEAIRQGITKDDKIEIKRQENEEIEIDIRDEYTNKTLNDCSIIIRIEN
DELETKKSIIYEIFKNINMSLYKIIIEKIENETEKVFENRYEEHLREKLLKDDKIDVILT
NMEIREKIKSNLEILGFVKFYLNVGGDKKSKNKMMLVEKILNINVDLTVEDIADFVIK
ELEFWNITKRIEKVKVKNNEFLEKRRNRTYIKSYVLLDKHEKFIERENKKDKIVKFFVE
NIKNNSIKEKIEKILAEFKIDEIKLEKELKKGNCDTEIFGIFKKHYKVNFDKKFFSKK
SDEEKELYKIIYRYLKGRIEKILVNEQVKVRLKKMEKIEIEKILNESILSEKILKRVKQYT
LEHIMYLGKLRHNDIDMTTVNTDDFSRLHAKKEELDLELITFFASTNMELNKIFSRENINN
DENIDFFGGDREKNYVLDKKILNSKIKIIRDLDVIDNKNNTNNFIRKFTKIGTNERNRI
LHAISKERDLQGTQDDYNKVINIQNLKISDEEVSKALNDVVFDDKKNITKINDIKIS
EENNDIKYLPFSKVLPEIILNLYRNNPKNEPFDTIETEKIVLNALIYVNKELYKKLILE
DDLEENESKNIFLQELKKTGLNIDEIDENIIENYKNAQISASKGNNKAIKKYQKKVIEC
YIGYLRKNYEELFDFSDFKMNIQEIKKQIKDINDNKTYERITVKTSDKTIVINDDFEYII
SIFALLNSNAVINKIRNRFATSVWLTSEYQNIIDILDEIMQLNTRLNECITENWNLNL
EEFIQMKKEIEKDFDFDKIQTKKEIFNNYEDIKNNILTEFKDDINGCDVLEKKLEKIVI
FDDETKFEIDKKSNIQDEQRKLSNINKKDLKKVDQYIKDKDQEIKSKILCRIIFNSDF
LKKYKKEIDNLIEDMESENENKFQEIYYPKERKNELYIYKKNLFLNIGNPNFDKIYGLIS
NDIKMADAKFLFNIDGKNIRKNKISEIDAILKNLNDKLNKNGYSKEYKEYIKKLEKENDDF
AKNIQKNKYKSEKFDYNRVSEYKKIRDLVEFNLYLNKIESYLIDINWKLAIQMARFERDMH
YIVNGLRELGIKLSGYNTGISRAYPKRNGSDGFYTTTAYKFFDEESYKKFEKICYGFG
IDLSENSEINKPENESIRNYISHFYIVRNPFAADYSIAEQIDRVSNLLSYSTRYNNSTYAS
VFEVFKDVLNDYDELKKKFKLIGNNDILERLMKPKVSVLELESYNSDYIKNLIIELLT
KIENTNDTL (SEQ ID NO: 3)

图8C

> 荚膜红细菌 (R121) (ETD76934.1) (U717_11515)
 MQIGKVQGRITISEFGDPAGGLKRKISTDGKNRKELPAHLSSDPKALIGQWISGIDKIYRK
 PDSRKSDGKAIHSPTPSKMQFDARDDLGEAFWKLVSEAGLAQDSYDQFKRRLHPYGDKF
 QPADSGAKLKFEADPPPEPQAFHGRWYGAMSKRGNDAKELAAALYEHLHVDEKRIDGQPKR
 NPKTDKFAPGLVVARALGIESSVLPRGMARLARNWGEETIQTYFVVDVAASVKEVAKAAV
 SAAQAFDPPRQVSGRSLSPKVGFFALAEHLERVTGSKRCSFDPAAGPSVLALHDEVKKTYK
 RLCARGKNAARAFPADKTELLALMRHTHENRVRNQVMRGRVSEYRGQAGDLAQSHYWT
 SAGQTEIKESEIFVRLWVGAFALAGRSMKAWIDPMGKI VNTTEKNDRLDTAAVNIRQVISN
 KEMVAEAMARRGIYFGETPELDRLGAE~~NEG~~VFALLRYLRGCRNQTFHLGARAGFLKEI
 RKELEKTRWGKAKEAEHVLTDKTVAAIRAIIDNDAKALGARLLADLSGAFVAHYASKEH
 FSTLYSEIVKAVKDAPEVSSGLPRLKLLKRADGVRGYVHGLRDTTRKHAFATKLPPPPAP
 RELDDPATKARYIALRLYDGPFRAYASGITGTALAGPAARAKEAATALAQSNNVTKAYS
 DVMEGRSSRLRPPNDGETLREYLSALTGETATEFRVQIGYESDSENARKQAEFIENYRRD
 MLAFMFEDYIRAKGFDWILKIEPGATAMTRAPVLPPEPIDTRGQYEHWQAAALYLVMHFVPA
 SDVSNLLHQLRKWEALQKYEVLVQDGDATDQADARREALDLVKRFRDVLVLFLKTGEARF
 EGRAAPFDLKPFRALFANPATFDRLFMATPTTARPAEDDDPEGDGASEPELRVARTLRGLR
 QIARYNHMAVLSDLFAKHKVRDEEVARLAEIEDETQEKSQIVAAQELRTDLHDKVMKCHP
 KTISPPEERQSYAAAIAKTIEEHRFLVGRVYLGDLRLHRLMMDVIGRLIDYAGAYERDTGT
 FLINASKQLGAGADWAVTIAGAANTDARTQT~~R~~KDLA~~H~~FNVLDRADGTPDALTALVNRAREM
 MAYDRKRKNVPRSI~~LD~~MLARLGLTLKWQMKDHL~~LQ~~DATITQAAIKHL~~DK~~VLTVGGPAA
 VTEARFSQDYLQMVAAVFN~~GS~~VQNP~~PR~~RRDDGD~~A~~WHKPPK~~P~~ATAQSQPDQKPPN~~K~~APSA
 GSRLPP~~Q~~QVGEVYEGVVVKVIDT~~GS~~LGLAVEGVAGNIGLHISRLR~~RI~~RED~~AI~~IVGRRYR
 FRVEIYVPPKSN~~TS~~KLNAADLV~~RI~~D (SEQ ID NO: 4)

图8D

>WP_034560163.1 WP_034560163.1 假想蛋白质 [鸡肉杆菌]
 MRITKVKIKLDNKL YQVTMQEEKYGT LKLNESRKSTAEILRLKKA SFNKS FHSKTINS
 QKENKNATIKKNGDYISQIF EKLGVDTNKNIRKPKMSLTDLKDLPKKDLALFIKRKFKN
 DDIVEIKNLDLISLFYNALQKVPGEHFTDESWADFCQEMMPYREYKNKFIERKIIILLANS
 IEQNKGFSINPETFSKRKRVLHQWAIEVQERGDFSILDEKLSKLAEIYNFKKMCKRVQDE
 LNDLEKSMKKGK NPEKEEAYKKQKNFKIKTIWKDYPYKTHIGLIEKIKENEFEELNQFNIE
 IGKYFEHYFP IKKERCTEDEPYYNSETIATTVNYQLKNALISYLMQIGKYKQFGLENQV
 LDSKKLQEI GIEYEGFQTKFMDACVFATSSLNKNIIEPMRSGDILGKREFKEAIA TSSFVNY
 HHFFPYFPFELKGMKDRESELIPFGEQTEAKQMNIWALRGSVQQI RNEI FHSFDKNQKF
 NLPQLDKSNFEFDASENSTGKSQSYIETDYKFLFEAEKNQLEQFFIERIKSSGALEYYPPL
 KSLEKLFAKKEMKFSLSQVVAFAFAPSYKKLVKKGHSYQTATEGTANYLGLSYYNRYELKE
 ESFQAQYYLLKLIYQYVFLPNFSQGNSPA FRET VKAILRINKDEARKKMKKKFLRKYA
 FEQVREMEFKE TPDDQYMSYLQSEMRREEKVRKA EKNDKGFEKNI TMNFEKLLMQIFVKGFD
 VFLTTFAGKELLLSSEEKVIKETEISLSKKINEREKTLKASI QVEHQLVATNSAISYWL
 CKLLDSRHLNELRNEMIKFKQSR IKFNHTQHAELIQNLLPIVELTILSNDYDEKND SQNV
 DVSAYFEDKSLYETAPYVQTD DRTRVSRPILKLEKYHTKSLIEALLKDN PQFRVAATDI
 QEWMHKREEIGELVEKRKNLHTEWAEGQQT LGAEKREERYDYCKIDRFNWKANKVTLTY
 LSQHLHYLITDLLGRMVGFSA LFERDLVYFSRSFSELGGETYHISDYKNLSGVLRLNAEVK
 PIKIKNIKVIDNEENPYKGNPEVKPFLDR LHAYLENVIGIKAVHGKI RNQTAHLSVLQ
 ELSMIESMNNLRDL MAYDRKLKNAVTKSMIKILDKHGMILKLDENHKNFIE SLIPKE
 I IHLKDKA IKTNQVSEECYQLVLALLTTNPGNQLN (SEQ ID NO: 5)

图8E

>CRZ35554.1 Herbinix hemicellulosilytica 基因组组件 TUM3/55, 毗连序列群
 02_T3/55T_毗连序列群 26, 全基因组鸟枪序列

MKLTTRRRISGNSVDQKITAAFYRDMSQGLLYYDSEDNDCTDKVIESMDFERSWRGRIILKN
 GEDDKNPFYMFVKGLVGSNDKIVCEPIDVDSDDPDNLDILINKNLTGFRNLKAPDSNDTL
 ENLIRKIQAGIPEEVLPCLKIKEMIQKDIVNRKEQLLSIKNNRIPFSLEGSKLVPST
 KMKWLFKLIDVPNKTFNEKMLEKYWEIYDYDKLKANI TNRLDKTDKKARSISRAVSEEL
 REYHKNLRTNYNRFVSGDRPAAGLDNGGSAKYNPDKEEFLFLKEVEQYFKKYFPVKSKH
 SNKSKDSLVDKYKNYCSYKVVKKEVNRSI INQLVAGLIQQGKLLYFYFYNQDQWEDFLN
 SYGLSYIQVEEAFKKSVMTSLSWGINRLTSFFIDDSNTVKFDDITTKKAKEAIESNYFNK
 LRTC SRMQDHFKELAFFYPVYVVKDKDRPDDDIENLI VLVKNAIESVSYLRNRRTFHFKE
 SSLLELLKELDDKNSGQNKIDYSVAAEFIKRDIENLYDVFREQIRSLGIAEYKADMISSD
 CFKTCGLEFALYSPKNSLMPAFKNVYKRGANLNKAYIRDKGPKETGDQGGQNSYKALEEYR
 ELTWYIEVKNNDQSYNAYKNLLQLIYYHAFLEPVRENEALITDFINRTKEWNRKETEERL
 NTKNNKKHKNFDENDITVNTYRYESIPDYQGESLDDYLKVLQRKQMARAKEVNEKEEGN
 NNYIQFIRDVVVWAFGAYLENKLNKYKNELQPPLSKENIGLNDTLKELFPEEKVKSPFNI
 KCRFSISTFIDNKGKSTDNTSAEAVKTDGKEDEKDKKNIKRKDLLCFYLFRLLDENEIC
 KLQHQFIKYRCSLKERRFPGNRTKLEKETELLAELEELMELVRFMTMPSIPEISAKAESGY
 DTMIKKYFKDFIEKKVFKNPKTSNLYYHSDSKTPVTRKYMALLMRSAPLHLYKDIFKGY
 LITKKECLEYIKLSNI IKDYQNSLNEHQLERIKLSEKQNGKDSLYLDKKDFYKVKEY
 VENLEQVARYKHLQH KINFESLYIRIHVVDIAARMVGYTQDWERDMHFLFKALVYNGVL
 EERRFEAIFNNDDNDGRIVKKIQNNLNKNRELVSMLCWNNKKLNKNEFGAI IWKRNP
 AHLNHFTQTEQNSKSSLESLSRLRILLAYDRKRQNAVTKTINDLLNDYHIRIKWEGRV
 DEGQIYFNIKEKEDIENEPI IHLKHLHKKDCYIYKNSYMFDKQKEWICNGIKEEVYDKSI
 LKICIGNLFKFDYEDKNKSSANPKHT (SEQ ID NO: 6)

图8F

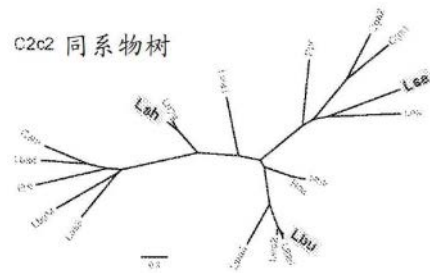


图9A

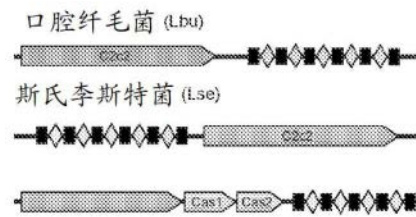


图9B

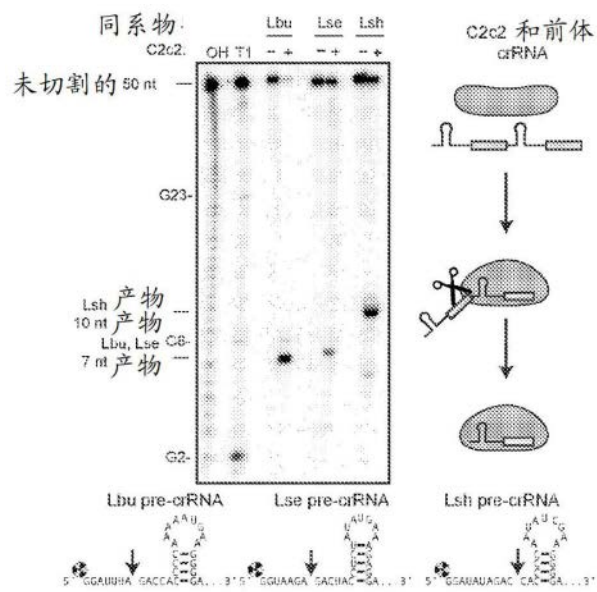


图9C

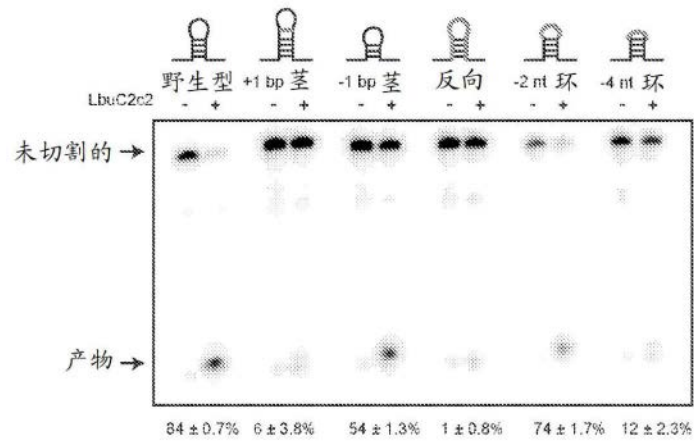


图10A

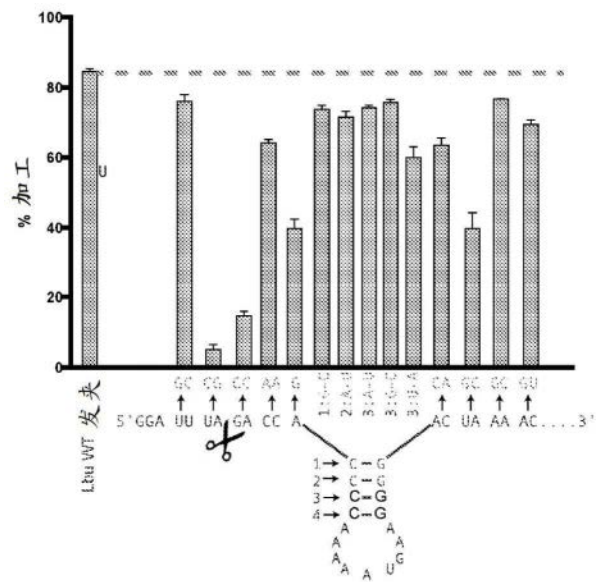


图10B

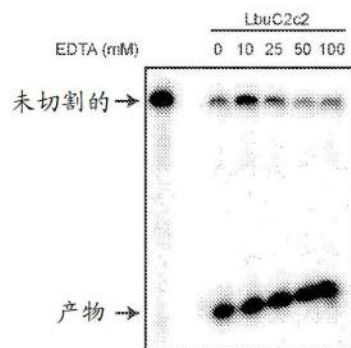


图10C

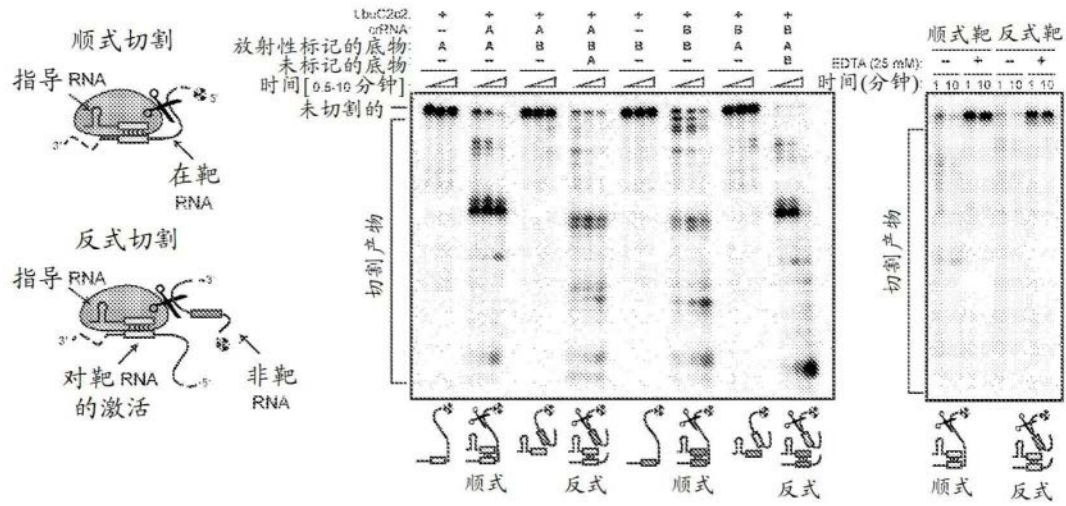


图 11A

图 11B

图 11C

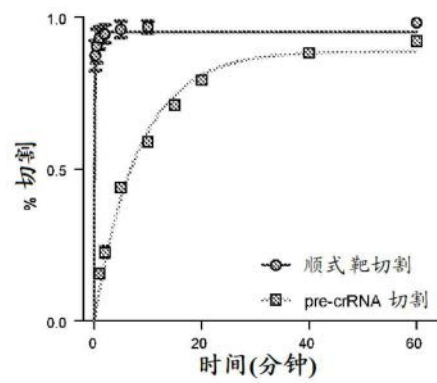


图12A

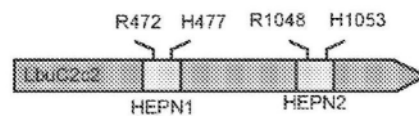


图12B

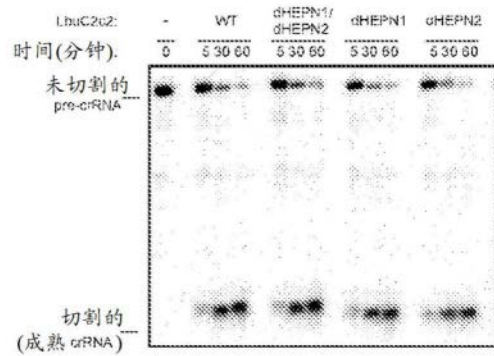


图12C

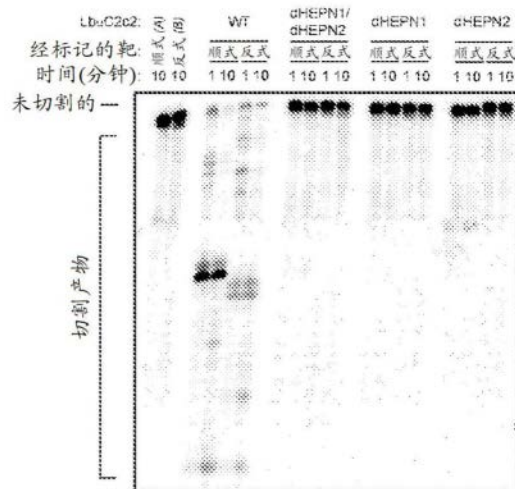


图12D

对荧光 RNA 寡核苷酸的反式切割

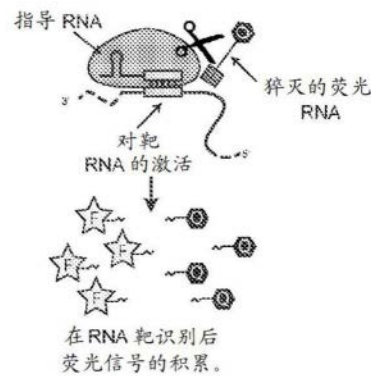


图13A

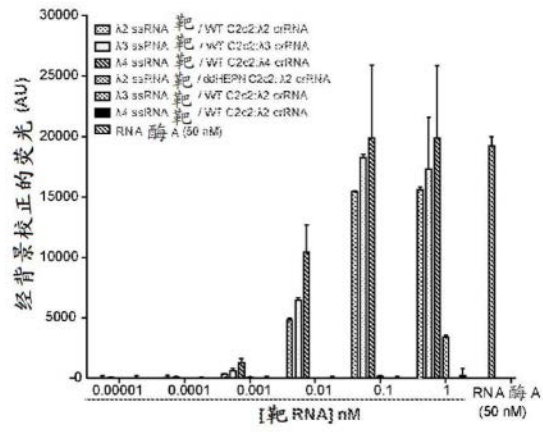


图13B

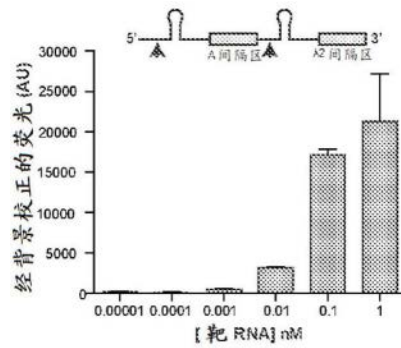


图13C

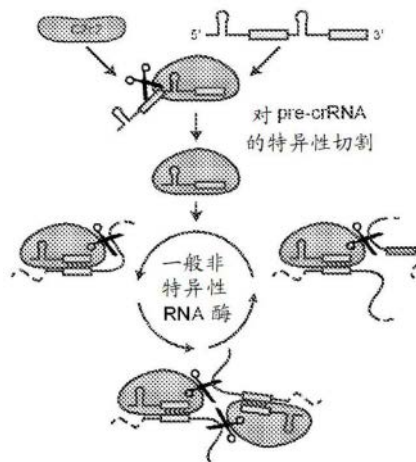


图13D

表2: 在本研究中使用的DNA寡聚核苷酸

名称	序列
T7_forward_SCK337	5'-TAATACGACTCACTATAGG-3'
T7_forward_long_AES428	5'-CCAAGTAATACGACTCACTATAGG-3'
Lbu_R1048A_AES441	5'-TAATTAGCAATATAAAGGTCCYTCYTTTCTTGC-3'
Lbu_H1053A_AES442	5'-CATCGCCGCATTCAATTATATYCCYCACGCCG-3'
Lbu_R472A_AES443	5'-CCGTGAGCGATGGAAGAAATCGCCTCGTCG-3'
Lbu_H477A_AES444	5'-TATYGTCCCTTCAACTTGGAAATTAGAAGGTAAGGA-3'

图14

表3: 在本研究使用的RNA

名称	序列 ^a	用于
Lbu_pre-crRNA_A_SCK14	5'-GGAUUJAGACCCACCCAAAAUJGAAAGGAGACUAAAACAGGGGCAGAGAUGAUGACCU-3'	图 9C, 10A-10C, 12A, C, 16
Lso_pre-crRNA_B_AES484	5'-GGUAAAGACTUACUCUUAUJAGAAAGAACUAAAACCAACUJGUJGUGUUAUC-3'	图 9C, 16
Lsh_pre-crRNA_A_SCK339	5'-GGAUUJAGACACCACCCAAUJAGUCAAAGGAGCTUAAACAGGGGCAGAGAUGAUGACCU-3'	图 9C, 16
Lbu_pre-crRNA_Invert_SCK321	5'-GGAUUJAGACCCAGGGGAAGUAAAACCCCACUAAAACAGGGGCAGAGAUGAUGACCU-3'	图 10A
Lbu_pre-crRNA_5stem_SCK331	5'-GGAUUJAGACACCACCCCAAAAUJGAAAGGAGACUAAAACAGGGGCAGAGAUGAUGACCU-3'	图 10A
Lbu_pre-crRNA_7bubble_SCK334	5'-GGAUUJAGACACCACCCUUAAGGCGGACUAAAACAGGGGCAGAGAUGAUGAUGACCU-3'	图 10A
Lbu_pre-crRNA_5bubble_SCK335	5'-GGAUUJAGACACCACCCUUAAGGAGGACUAAAACAGGGGCAGAGAUGAUGAUGACCU-3'	图 10A
Lbu_pre-crRNA_3stem_SCK342	5'-GGAUUJAGACACCACCCAAAAAUJGAAAGGAGACUAAAACAGGGGCAGAGAUGAUGACCU-3'	图 10A
Lbu_pre-cr_5'_mut1_AES497	5'-GGCGUJAGACACCACCCAAAAAUJGAAAGGAGACUAAAACAGGGGCAGAGAUGAUGACCU-3'	图 10B, 17
Lbu_pre-cr_5'_mut2_AES496	5'-GGAGUJAGACACCACCCAAAAAUJGAAAGGAGACUAAAACAGGGGCAGAGAUGAUGACCU-3'	图 10B, 17
Lbu_pre-cr_5'_mut3_AES495	5'-GGAUUJAGACACCACCCAAAAAUJGAAAGGAGACUAAAACAGGGGCAGAGAUGAUGACCU-3'	图 10B, 17
Lbu_pre-cr_5'_mut4_AES497	5'-GGAUUJAGACACCACCCAAAAAUJGAAAGGAGACUAAAACAGGGGCAGAGAUGAUGACCU-3'	图 10B, 17
Lbu_pre-cr_5'_mut5_AES482	5'-GGAUUJAGACACCACCCAAAAAUJGAAAGGAGACUAAAACAGGGGCAGAGAUGAUGACCU-3'	图 10B, 17
Lbu_pre-cr_5'_mut6_AES478	5'-GGAUUJAGUACACCACCCAAAAAUJGAAAGGAGACUAAAACAGGGGCAGAGAUGAUGACCU-3'	图 10B, 17
Lbu_pre-cr_5'_mut7_AES480	5'-GGAUUJAGUAAACACCACCCAAAAAUJGAAAGGAGACUAAAACAGGGGCAGAGAUGAUGACCU-3'	图 10B, 17
Lbu_pre_cr_5'_mut8_AES488	5'-GGAUUJAGACACGCCACCCAAAAAUJGAAAGGAGACUAAAACAGGGGCAGAGAUGAUGACCU-3'	图 10B, 17
Lbu_pre_stem_mut1_AES502	5'-GGAUUJAGACACGCCCAAAAUJGAAAGGAGACUAAAACAGGGGCAGAGAUGAUGACCU-3'	图 10B, 17
Lbu_pre_stem_mut2_AES501	5'-GGAUUJAGACACGCCCAAAAUJGAAAGGAGACUAAAACAGGGGCAGAGAUGAUGACCU-3'	图 10B, 17
Lbu_pre_stem_mut3_AES500	5'-GGAUUJAGACACACCAAAAUJGAAAGGAGACUAAAACAGGGGCAGAGAUGAUGACCU-3'	图 10B, 17
Lbu_pre_stem_mut4_AES499	5'-GGAUUJAGACACACCAAAAUJGAAAGGAGACUAAAACAGGGGCAGAGAUGAUGACCU-3'	图 10B, 17
Lbu_pre_stem_mut5_AES504	5'-GGAUUJAGACACUCCAAAAAUJGAAAGGAGACUAAAACAGGGGCAGAGAUGAUGACCU-3'	图 10B, 17
Lbu_pre_cr_3'_mut1_AES505	5'-GGAUUJAGACACCACCCAAAAAUJGAAAGGAGACUAAAACAGGGGCAGAGAUGAUGACCU-3'	图 10B, 17
Lbu_pre_cr_3'_mut2_AES506	5'-GGAUUJAGACACCACCCAAAAAUJGAAAGGAGACUAAAACAGGGGCAGAGAUGAUGACCU-3'	图 10B, 17
Lbu_pre_cr_3'_mut3_AES507	5'-GGAUUJAGACACCACCCAAAAAUJGAAAGGAGACUAAAACAGGGGCAGAGAUGAUGACCU-3'	图 10B, 17
Lbu_pre_cr_3'_mut4_AES508	5'-GGAUUJAGACACCACCCAAAAAUJGAAAGGAGACUAAAACAGGGGCAGAGAUGAUGACCU-3'	图 10B, 17
crLbu_A_IG_AES432	5'-GGCCACCCCAAAAUJGAAAGGAGACUAAAACAGGGGCAGAGAUGAUGACCU-3'	图 11, 12, 19-20
crLbu_B_AES451	5'-GGCCACCCCAAAAUJGAAAGGAGACUAAAACAGGGGCAGAGAUGAUGAUGACCU-3'	图 11
A0_target_AES450	5'-GGCACACCCAGAGGGGGAGCCAAAGGGUUAUUAUCUGCCCCACAGCAGAAAGCC-3'	图 11, 12, 20
B_target_AES452	5'-GGGAACCCCAAGGCCACCCGCGAGAGAGAGACCCAGAUUAUGUAGAGACCUUACACACCC-3'	图 11, 12, 20
crLbu_Lambda2_AES453	5'-GGCCACCCCAAAAUJGAAAGGAGACUAAAACAGAGAUAAGUAGAUJGCCAU-3'	图 13B-13C
crLbu_Lambda3_MOC410	5'-GGCCACCCCAAAAUJGAAAGGAGACUAAAACAGAGUUAUCCGAGUAGH-3'	图 13B
crLbu_Lambda4_MOC411	5'-GGCCACCCCAAAAUJGAAAGGAGACUAAAACAGAGAUAAGCUGUGUAGJC-3'	图 13B
Lambda2_target_MOC28	5'-GGCUCAUUUJGACACGGUUAUGGCAUCCACUUAUACUGGCAUCCUCCACAC-3'	图 13B-C
Lambda3_target_MOC36	5'-GGAAAUCAUUAACACCCGCAUUAUUGGAGUUAACAGCCAGCCAGCAGAGU-3'	图 13B
Lambda4_target_MOC37	5'-GGCAUAAUAAUUGCGCGCGUAGAACACAGGCUUAUUGCCACUUAUUGUAGA-3'	图 13B
pre-crLbu_dimer_SCK324	5'-GGAUUJAGACACCACCCAAAAAUJGAAAGGAGACUAAAACAGGGGCAGAGAUGAUGAUGACCUUUAUAGACCCCAAAAUJGAAAGGAGACUAAAACAGGGGCAGAGAUGAUGAUGACCU-3'	图 13C, 16A-16F
crLbu_Lambda2_SCK315	5'-GGAUUJAGACACCACCCAAAAAUJGAAAGGAGACUAAAACAGAGAUAAGUGGUAUGCCU-3'	图 16A-16F
Lbu_pre_cr_5'_4mer1_AES481	5'-GGAUUJAGUUAACACCACCCAAAAAUJGAAAGGAGACUAAAACAGGGGCAGAGAUGAUGAUGACCU-3'	图 16A-16F
Lbu_pre_cr_5'_4mer2_AES479	5'-GGAUUJAGUUAACACCACCCAAAAAUJGAAAGGAGACUAAAACAGGGGCAGAGAUGAUGAUGACCU-3'	图 16A-16F
Lbu_pre_cr_5'_4mer3_SCK343	5'-GGAUUJAGUUAACACCACCCAAAAAUJGAAAGGAGACUAAAACAGGGGCAGAGAUGAUGAUGACCU-3'	图 16A-16F
Lbu_pre_cr_5'_4mer4_AES503	5'-GGAUUJAGUUAACACCACCCAAAAAUJGAAAGGAGACUAAAACAGGGGCAGAGAUGAUGAUGACCU-3'	图 16A-16F
crLbu_GuideWalk1_SCK302	5'-GGCCACCCCAAAAUJGAAAGGAGACUAAAACAGUUAUUGGCUCCUCCUCCUAGAA-3'	图 18A-18B
crLbu_GuideWalk2_SCK303	5'-GGCCACCCCAAAAUJGAAAGGAGACUAAAACAGUUAUUGGCUCCUCCUCCUAGAA-3'	图 18A-18B
crLbu_GuideWalk3_SCK304	5'-GGCCACCCCAAAAUJGAAAGGAGACUAAAACAGUUAUUGGCUCCUCCUCCUAGAA-3'	图 18A-18B
crLbu_GuideWalk4_SCK305	5'-GGCCACCCCAAAAUJGAAAGGAGACUAAAACAGUUAUUGGCUCCUCCUCCUAGAA-3'	图 18A-18B
crLbu_GuideWalk5_SCK306	5'-GGCCACCCCAAAAUJGAAAGGAGACUAAAACAGUUAUUGGCUCCUCCUCCUAGAA-3'	图 18A-18B
crLbu_GuideWalk6_SCK307	5'-GGCCACCCCAAAAUJGAAAGGAGACUAAAACAGUUAUUGGCUCCUCCUCCUAGAA-3'	图 18A-18B
crLbu_GuideWalk7_SCK308	5'-GGCCACCCCAAAAUJGAAAGGAGACUAAAACAGUUAUUGGCUCCUCCUCCUAGAA-3'	图 18A-18B
crLbu_GuideWalk8_SCK309	5'-GGCCACCCCAAAAUJGAAAGGAGACUAAAACAGUUAUUGGCUCCUCCUCCUAGAA-3'	图 18A-18B
A1_target_U_MOC279	5'-GGCUCAUUAUGAGGGGGAGCCAAAGGGUUAUUAUUCUCCUCCUCCUCCUAGAA-3'	图 18A-18B
A2_target_70nt_AES447	5'-GGCUCAUUAUGAGGGGGAGCCAAAGGGUUAUUAUUCUCCUCCUCCUCCUAGAA-3'	图 18A-18B
A3_target_80nt_AES448	5'-GGCUCAUUAUGAGGGGGAGCCAAAGGGUUAUUAUUCUCCUCCUCCUCCUAGAA-3'	图 18A-18B
A4_5'_ts_shif_AES449	5'-GGCUCAUUAUGAGGGGGAGCCAAAGGGUUAUUAUUCUCCUCCUCCUCCUAGAA-3'	图 18A-18B
crLbu_A_16nt_trunc_SCK282	5'-GGCCACCCCAAAAUJGAAAGGAGACUAAAACAGAGAUAUAGCCU-3'	图 19A-19C
crLbu_A_24nt_ext_SCK283	5'-GGCCACCCCAAAAUJGAAAGGAGACUAAAACAGAGAUAUAGCCU-3'	图 19A-19C
crLbu_A_mature_GA_SCK340	5'-GACCAACCCCAAAAUJGAAAGGAGACUAAAACAGAGAUAUAGCCU-3'	图 19A-19C
crLbu_A_mature_GGG_SCK341	5'-GGGACACCCCAAAAUJGAAAGGAGACUAAAACAGAGAUAUAGCCU-3'	图 19A-19C
crLbu_A_mature_CCA_AES461	5'-GACACCCCAAAAUJGAAAGGAGACUAAAACAGAGAUAUAGCCU-3'	图 19A-19C

³所有RNA是订购合成的 (Integrated DNA technologies), 或者是如在方法中所描述的从互补模板转录的

图15

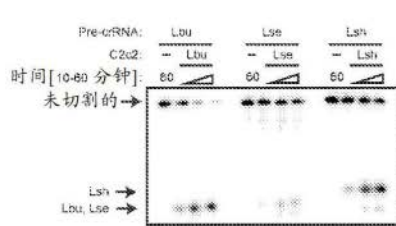


图 16A

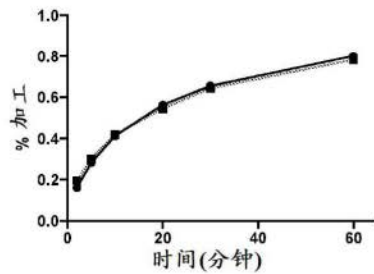


图 16D

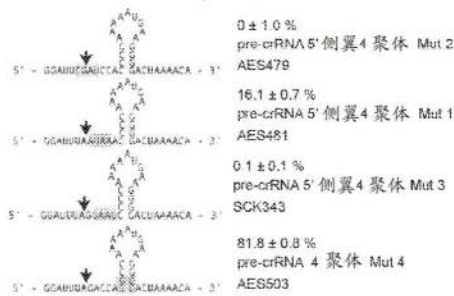
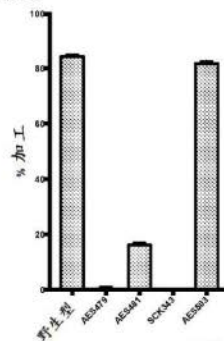


图 16E

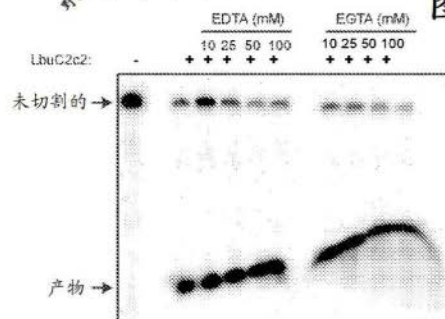


图 16F

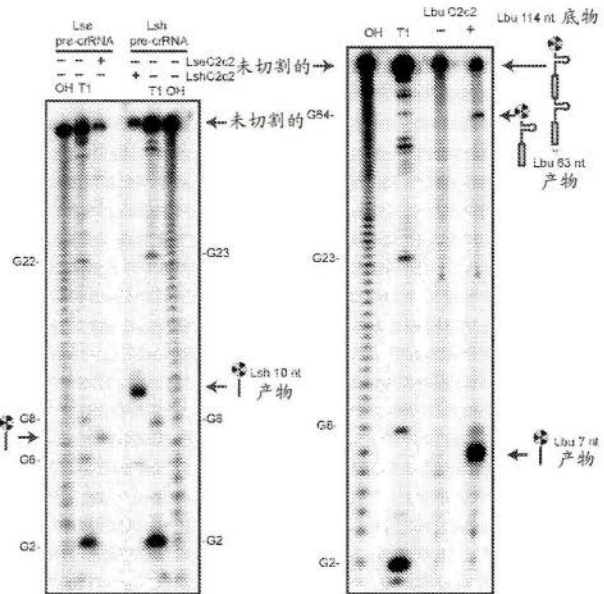


图 16B

图 16C

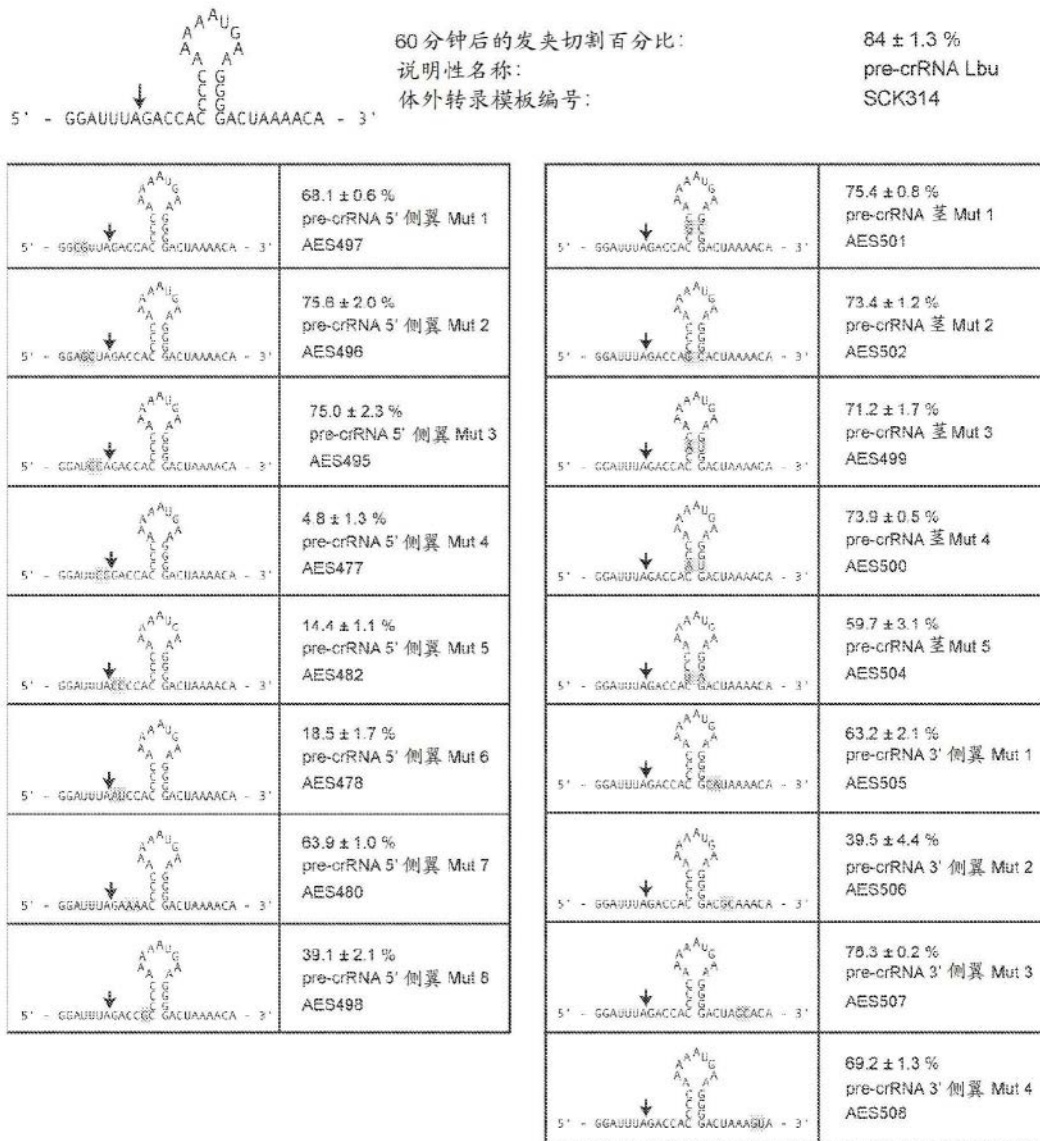


图17

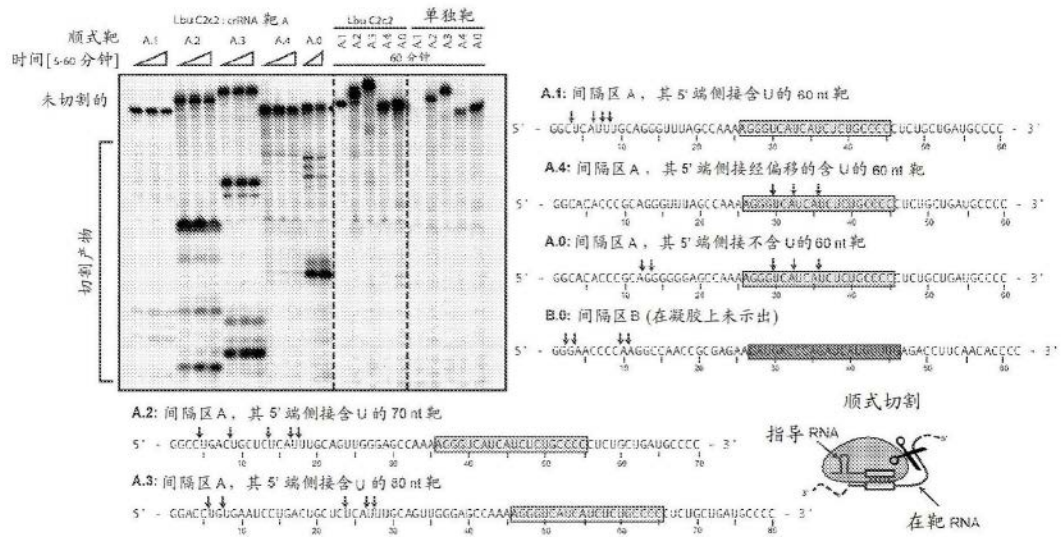


图18A

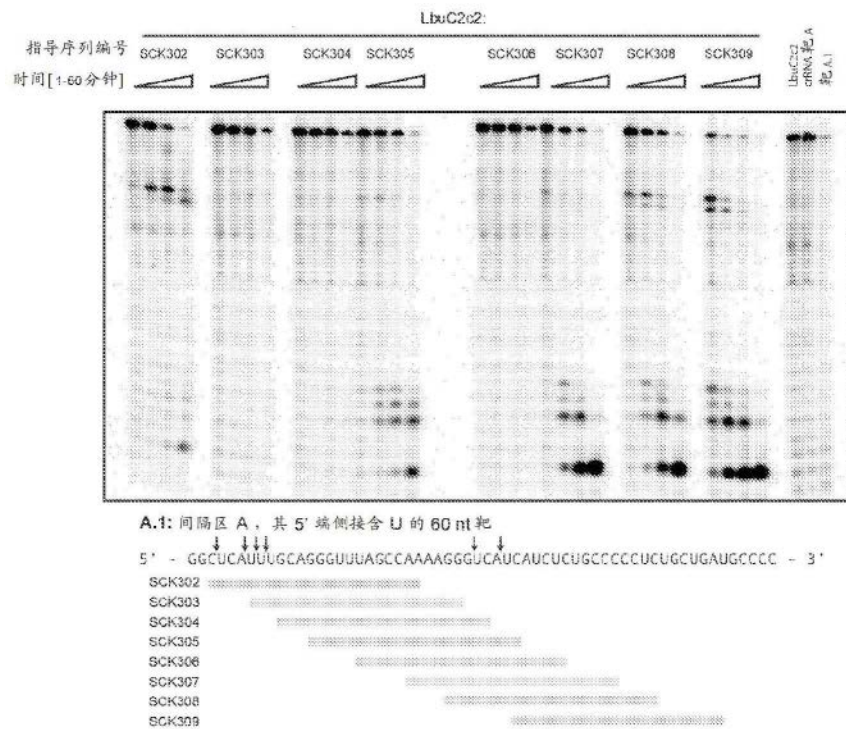


图18B

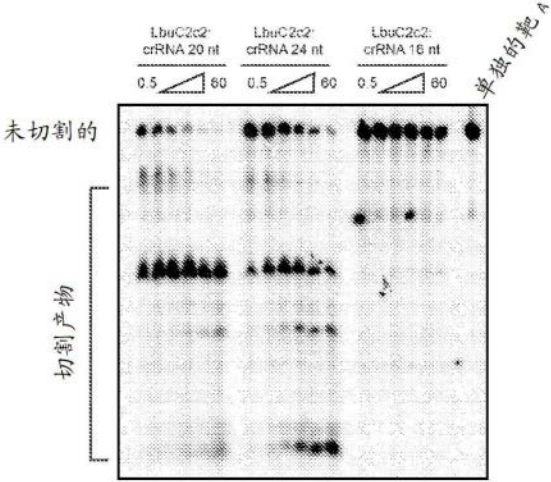


图19A

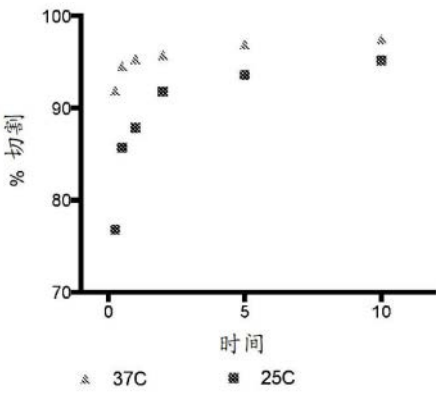


图19B

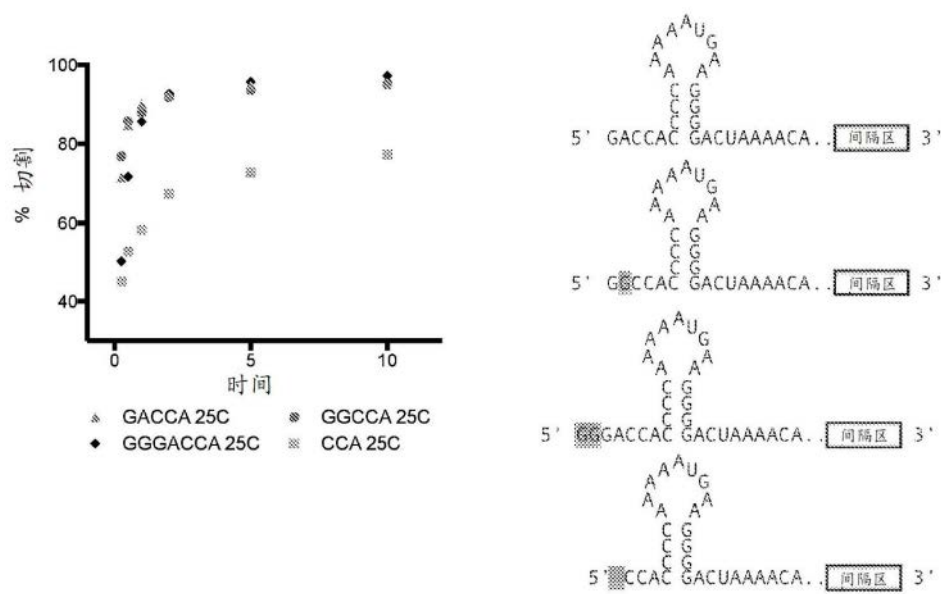


图19C

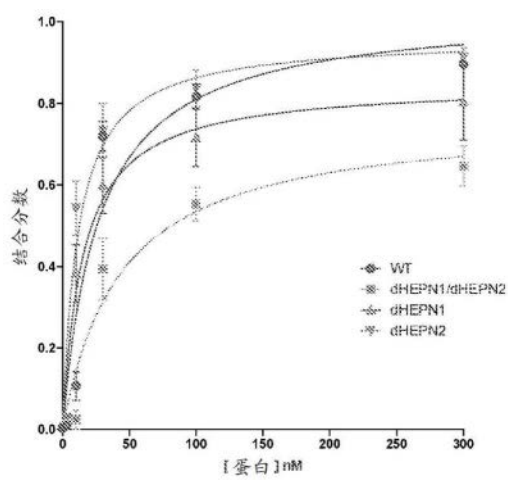


图20A

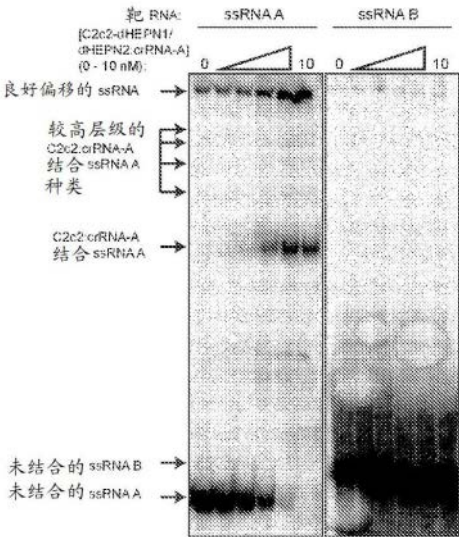


图20B

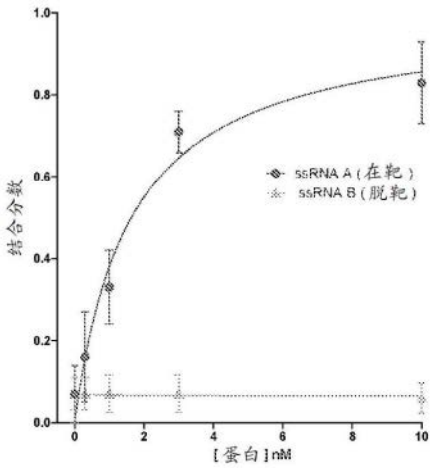


图20C

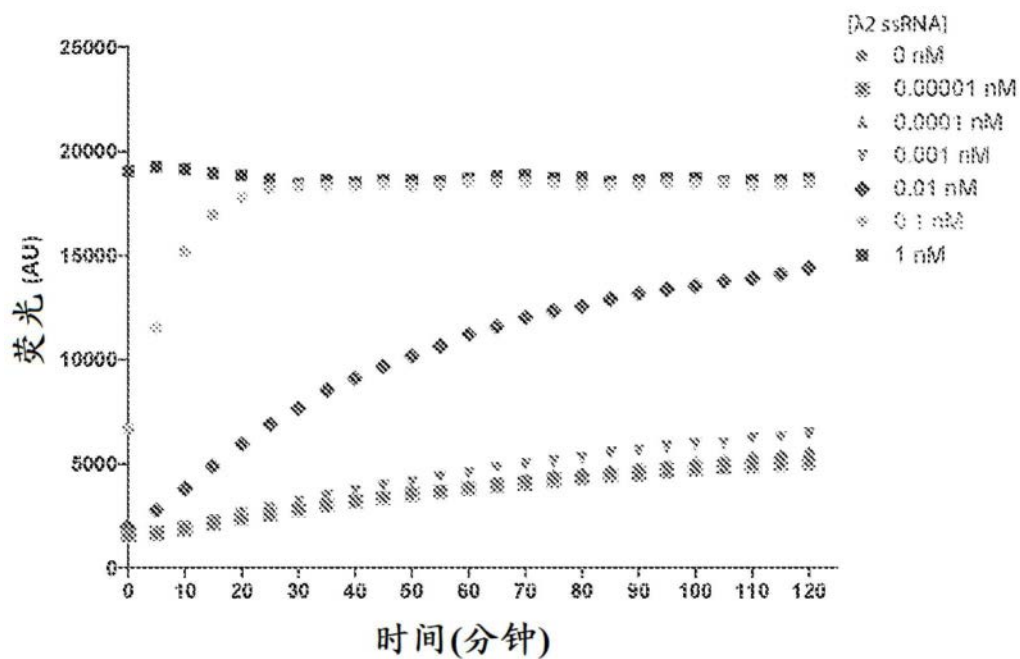


图21

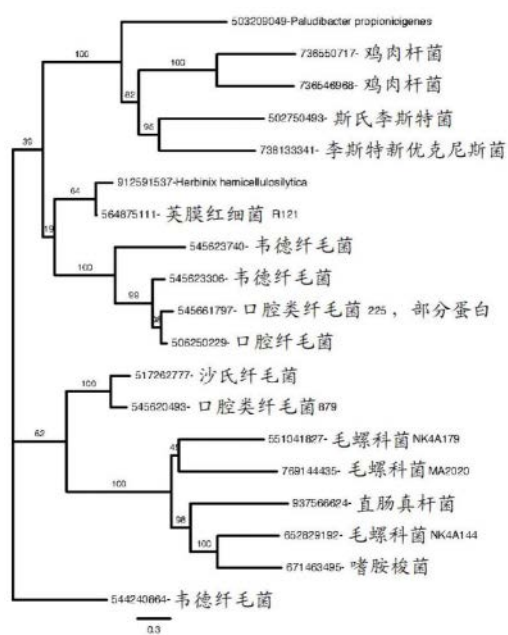


图22A

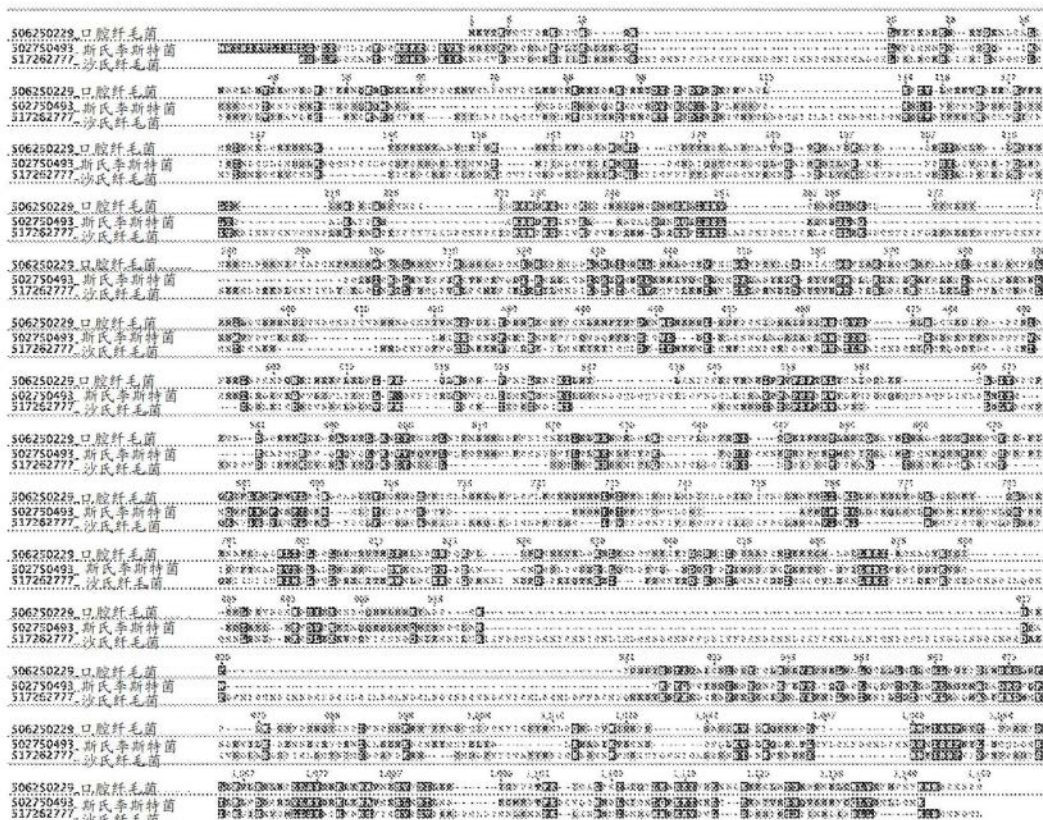


图22B

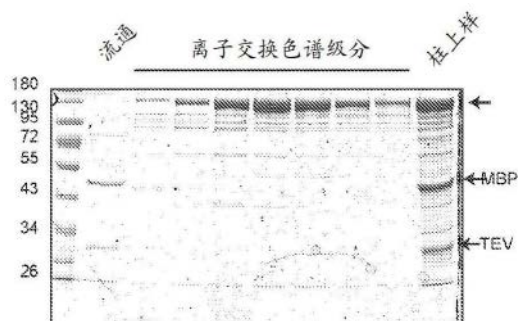


图23A

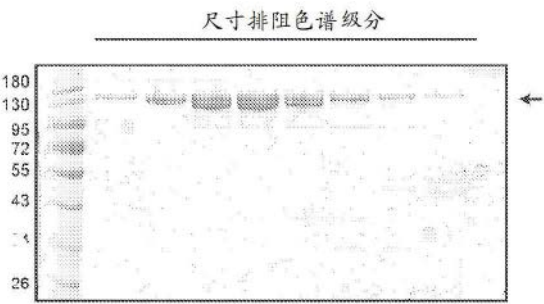


图23B

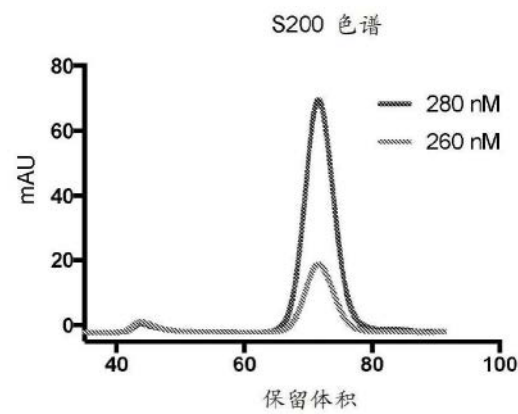


图23C

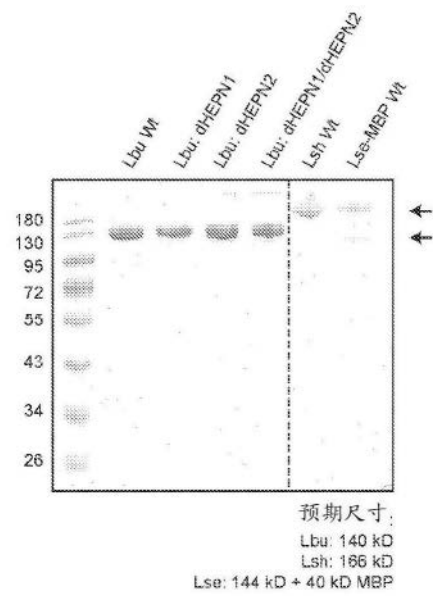


图23D

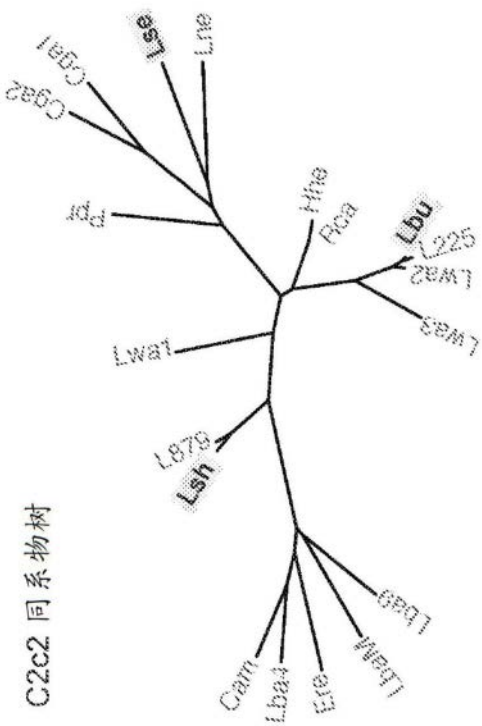


图24A

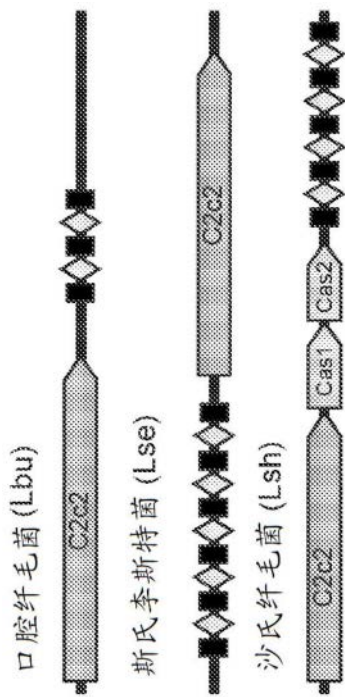


图24B

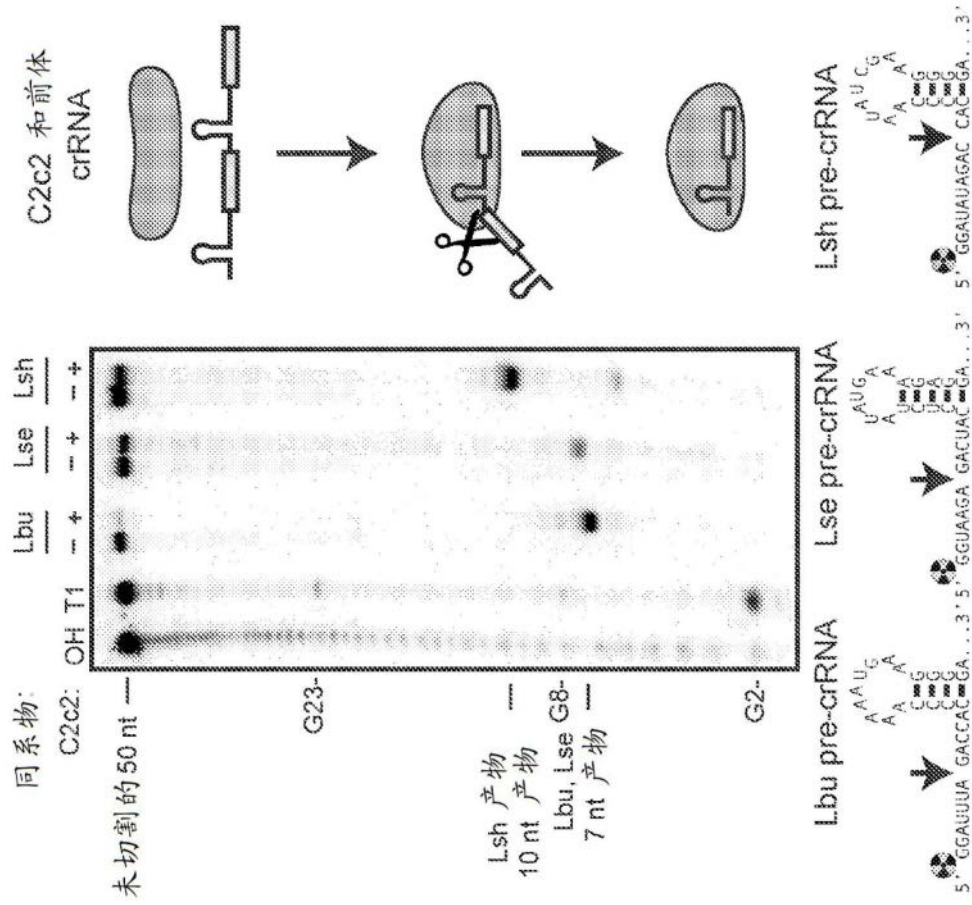


图24C

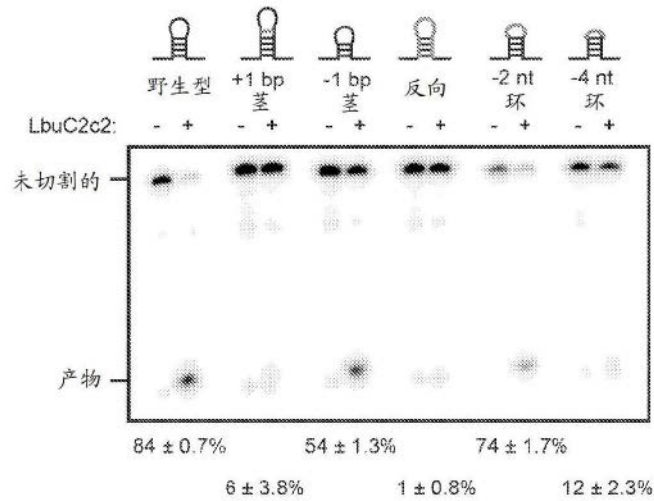


图25A

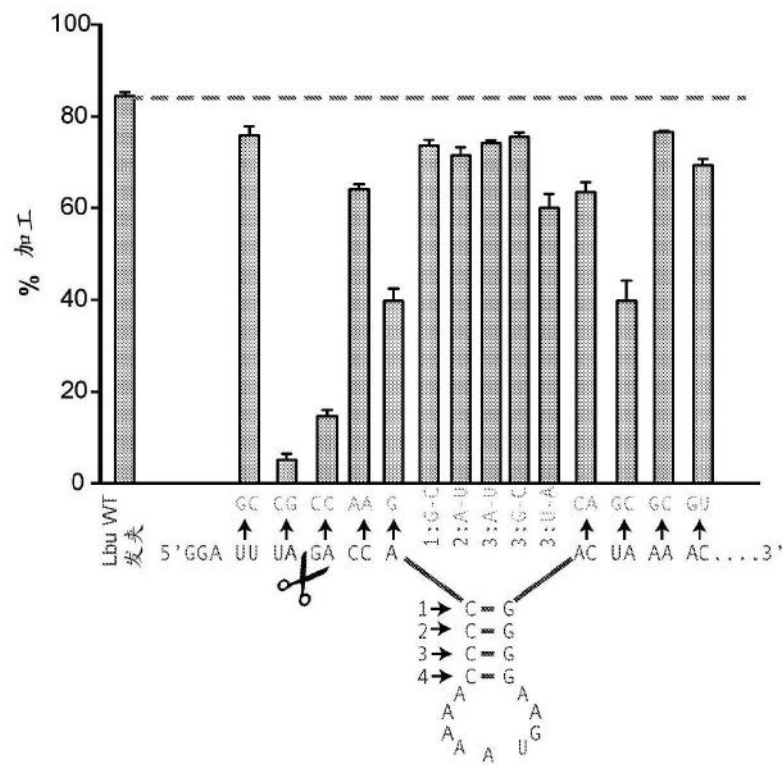


图25B

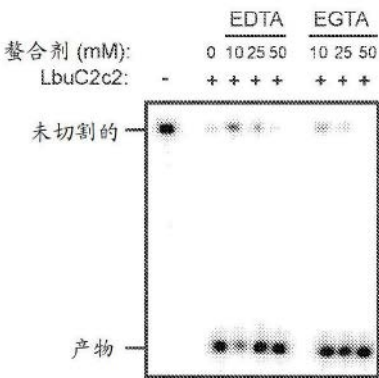


图25C

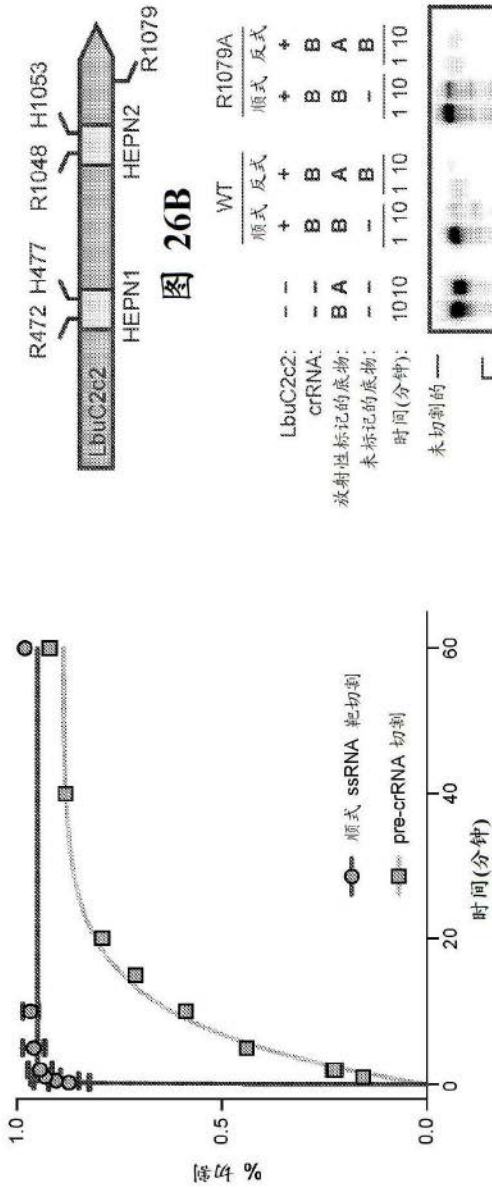
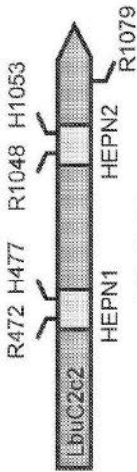


图 26B



LbuC2c2:	WT		R1079A	
	顺式	反式	顺式	反式
crRNA:	+	+	+	+
放射性标记的底物:	B	B	B	B
未标记的底物:	B	A	B	A
时间(分钟):	-	-	-	B
未切割的	1010	1 10 1 10	1 10 1 10	1 10 1 10

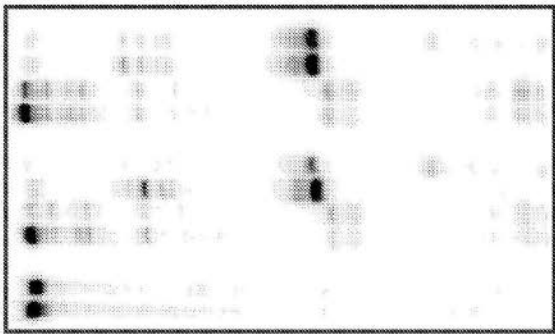


图 26D

图 26A

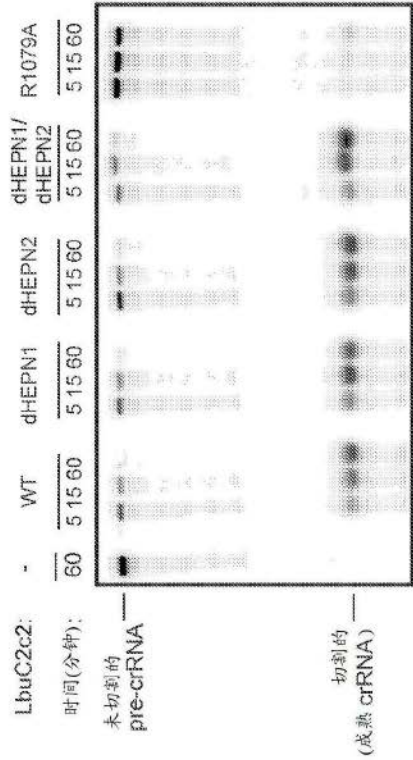


图 26C

对荧光 RNA 寡核苷酸的反式切割

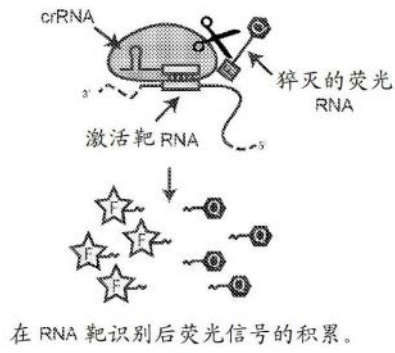


图27A

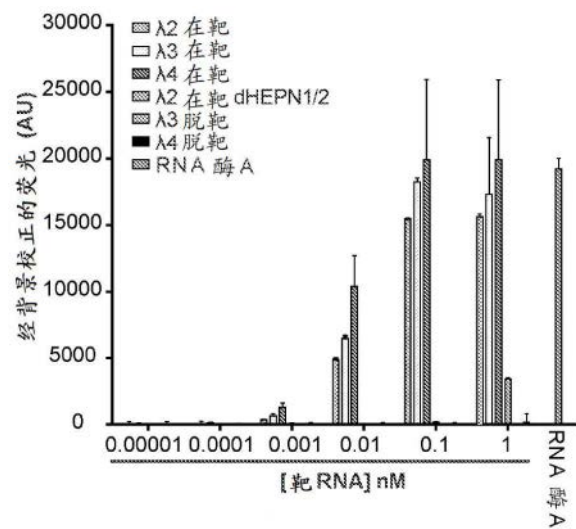


图27B

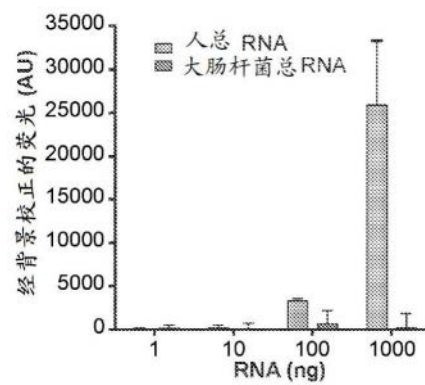


图27C

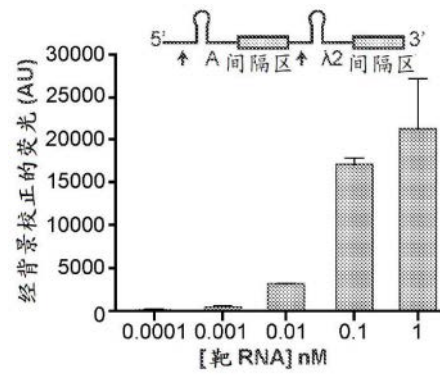


图27D

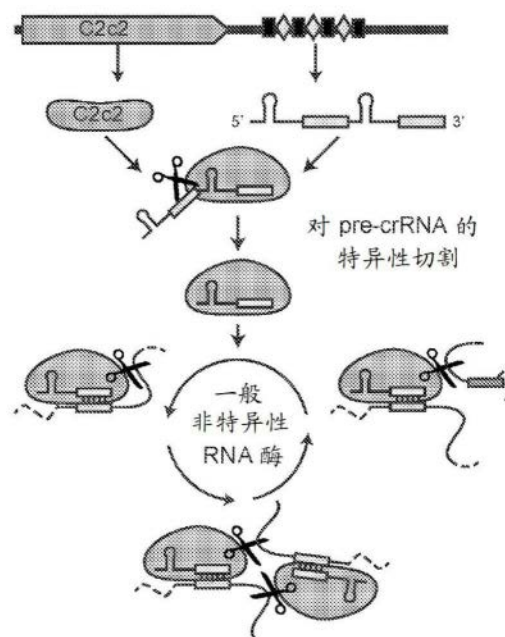


图27E



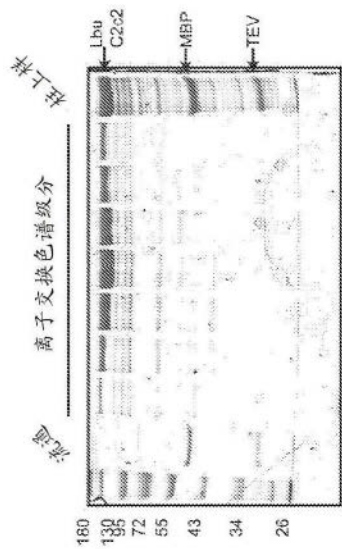


图29A

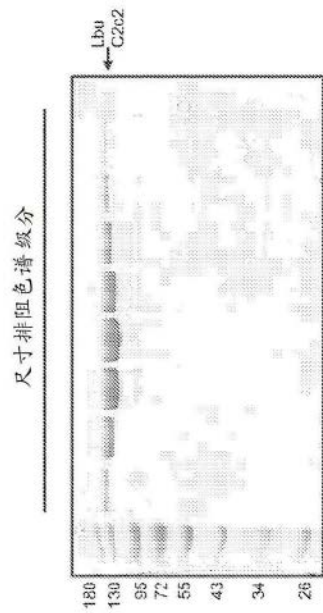


图29B

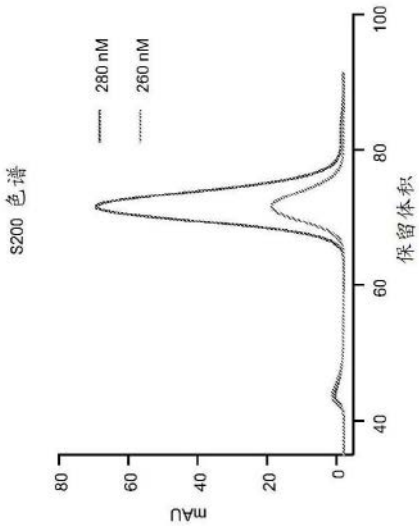


图29C

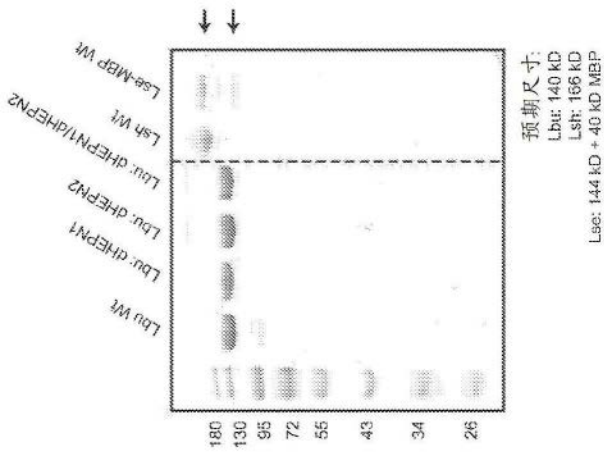
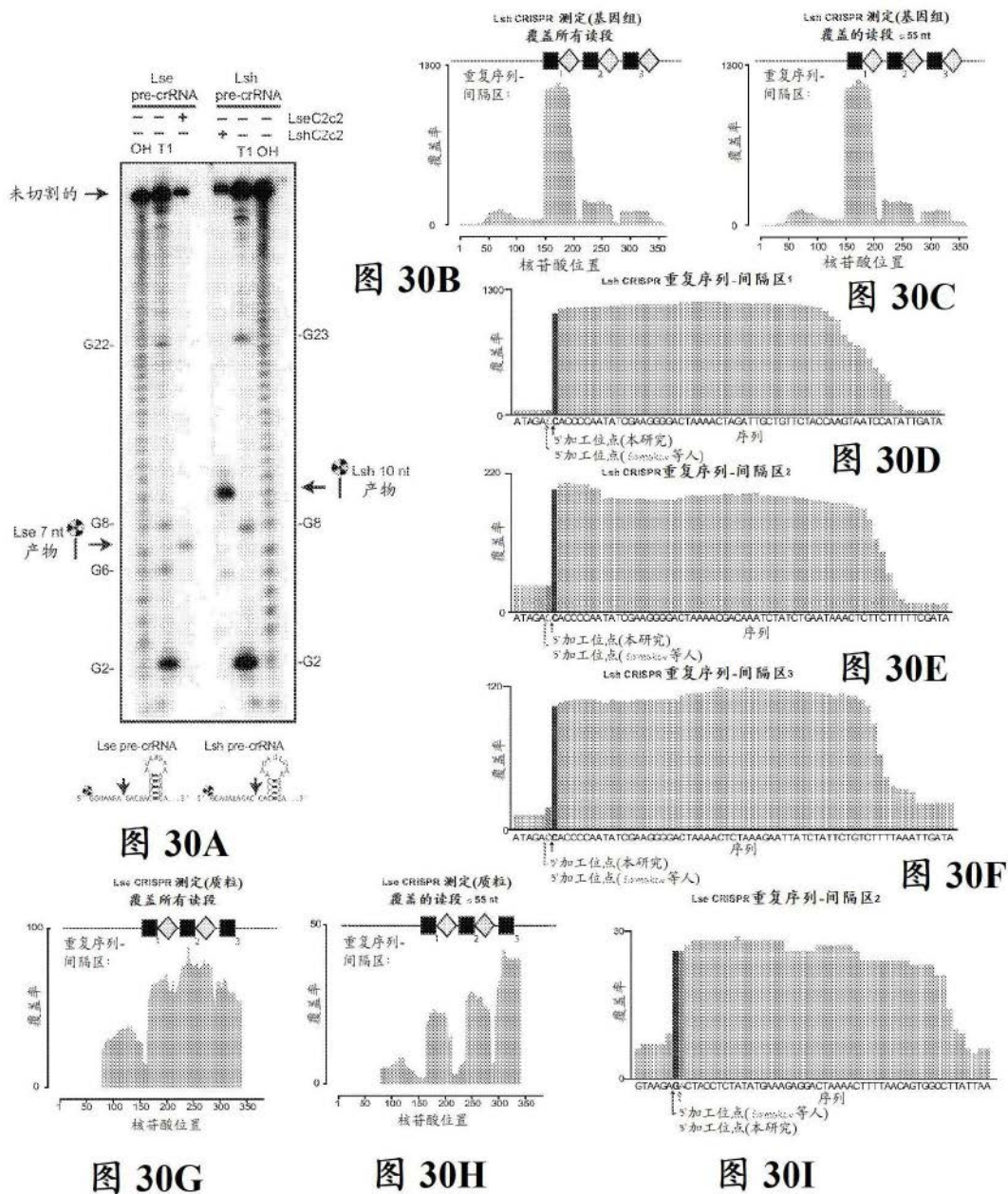


图29D



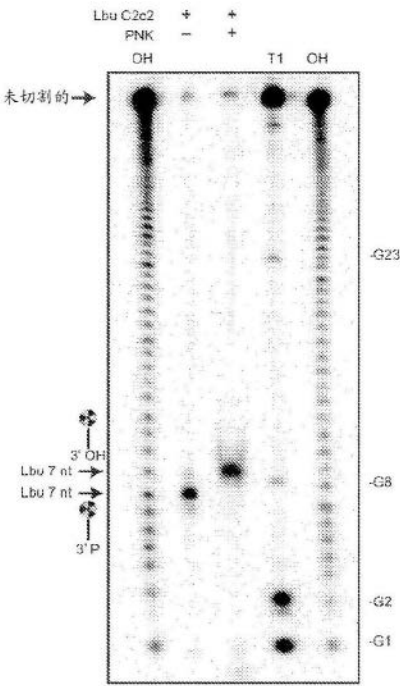


图31D

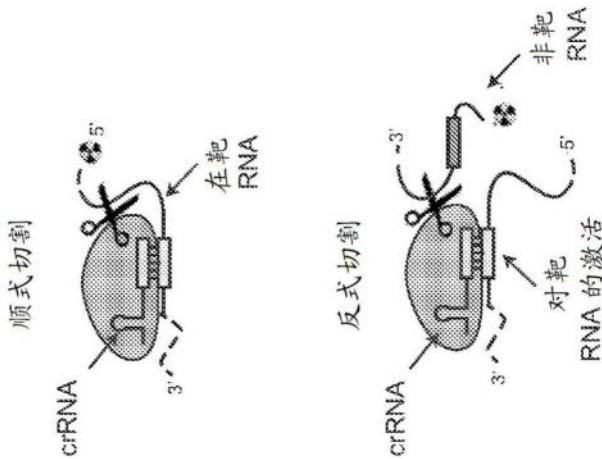


图32A

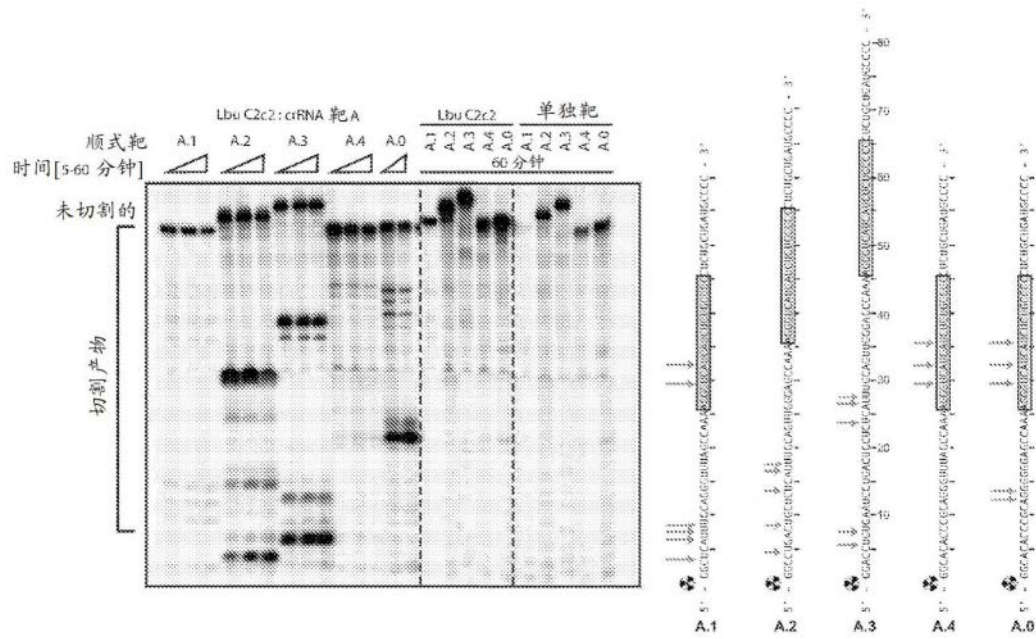


图33A

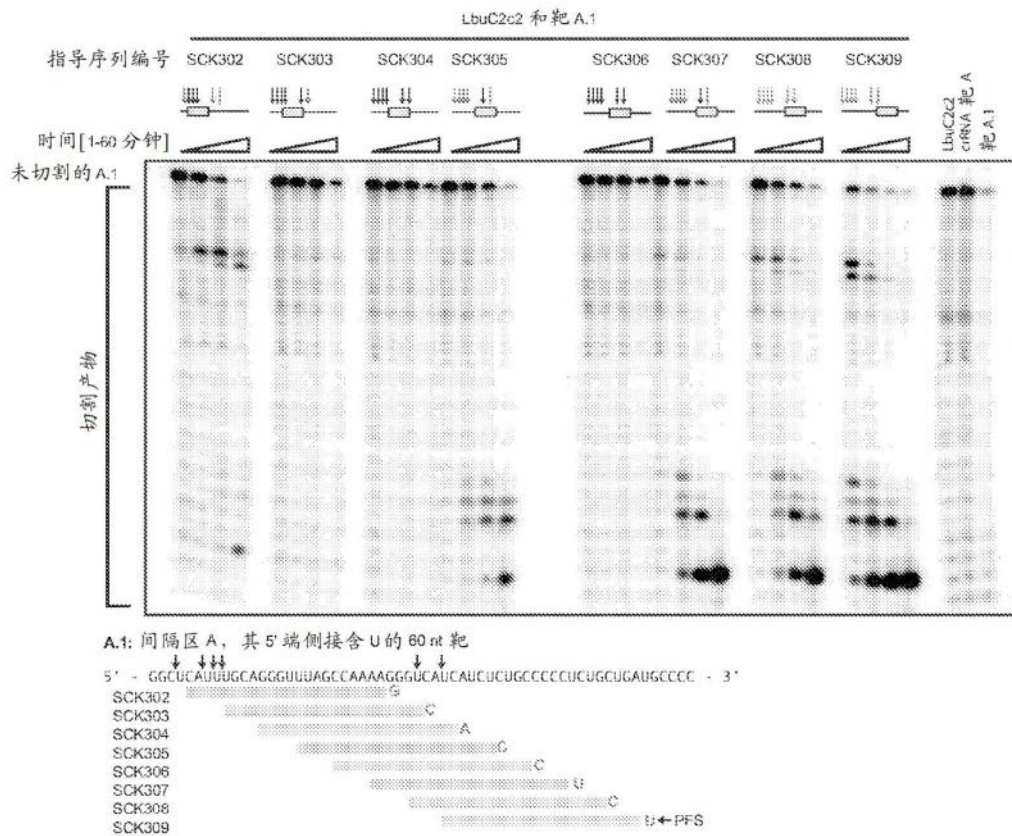


图33B

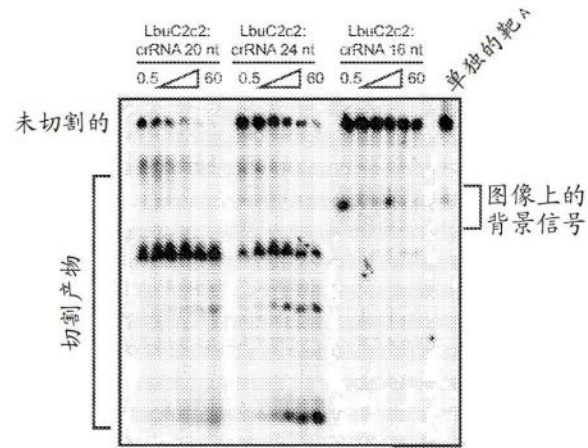


图34A

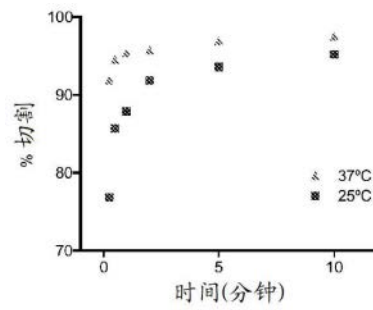


图34B

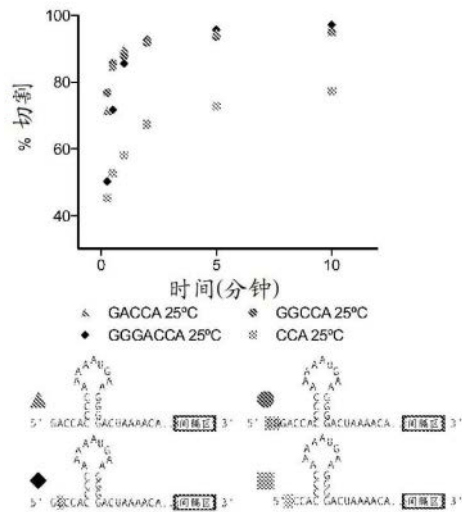


图34C

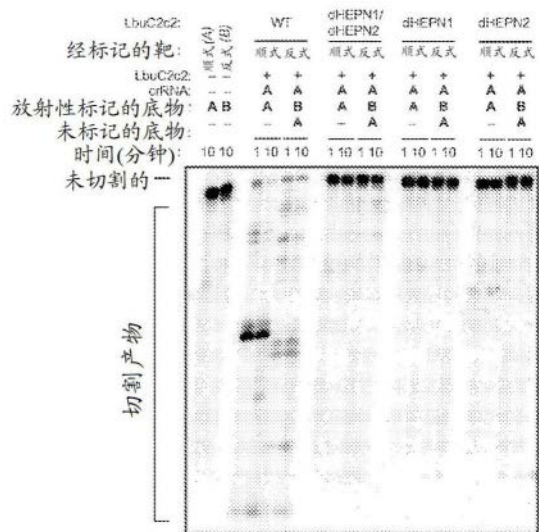


图34D

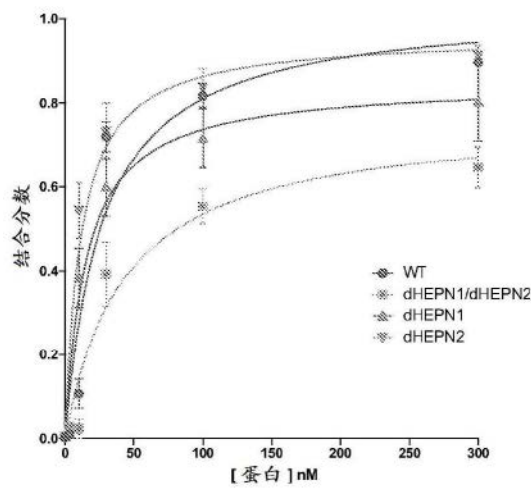


图35A

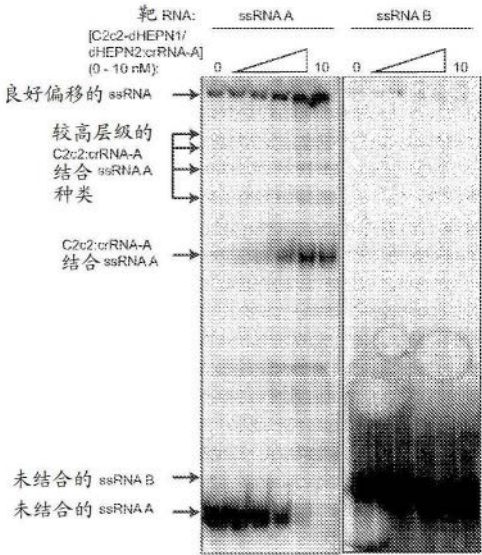


图35B

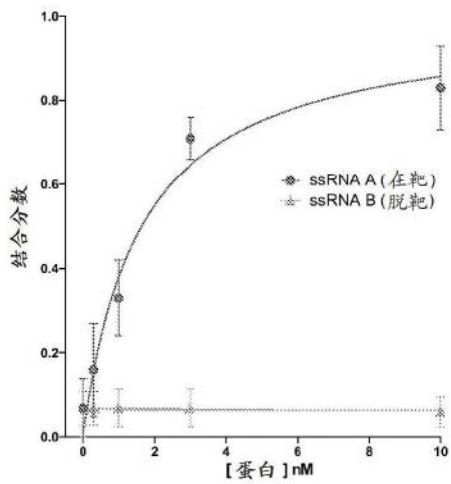


图35C

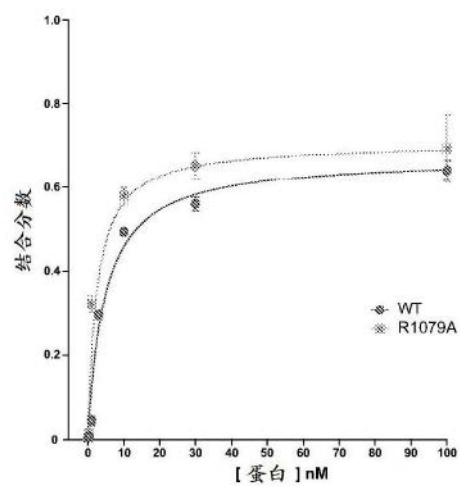


图35D

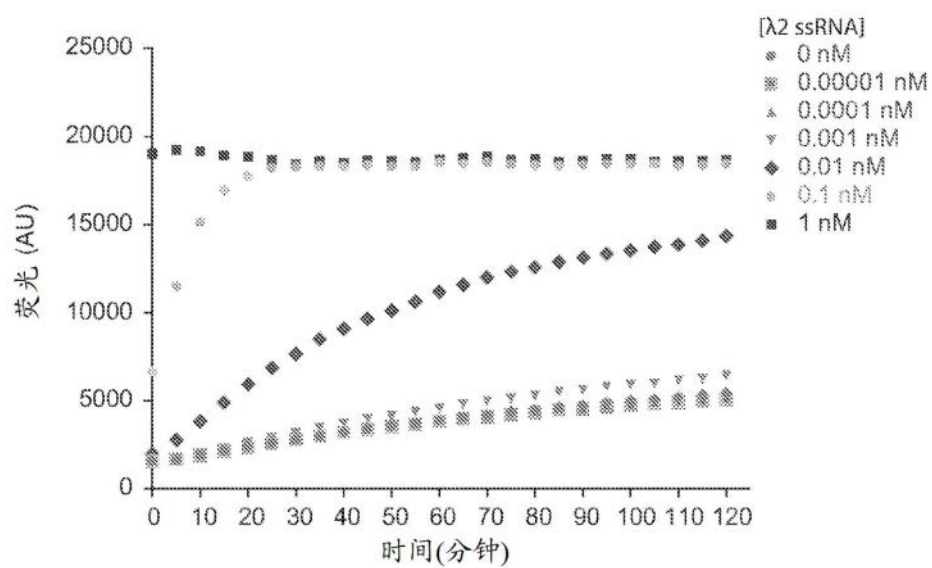


图36A

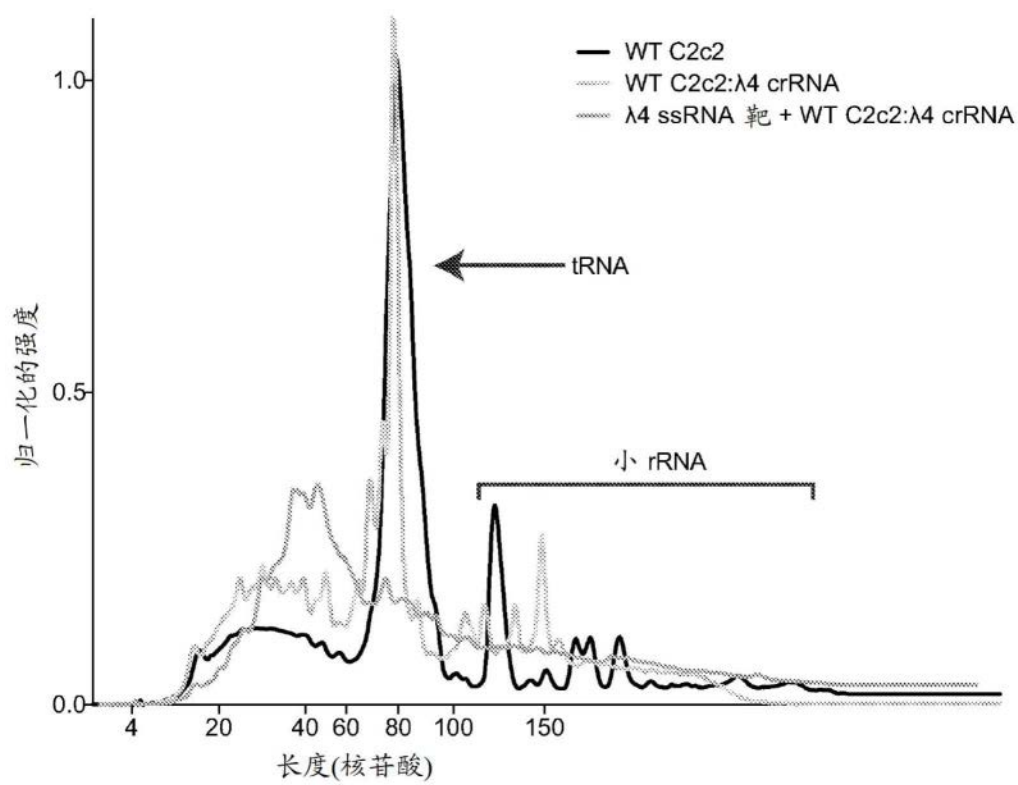


图36B

C2c2 加工 025

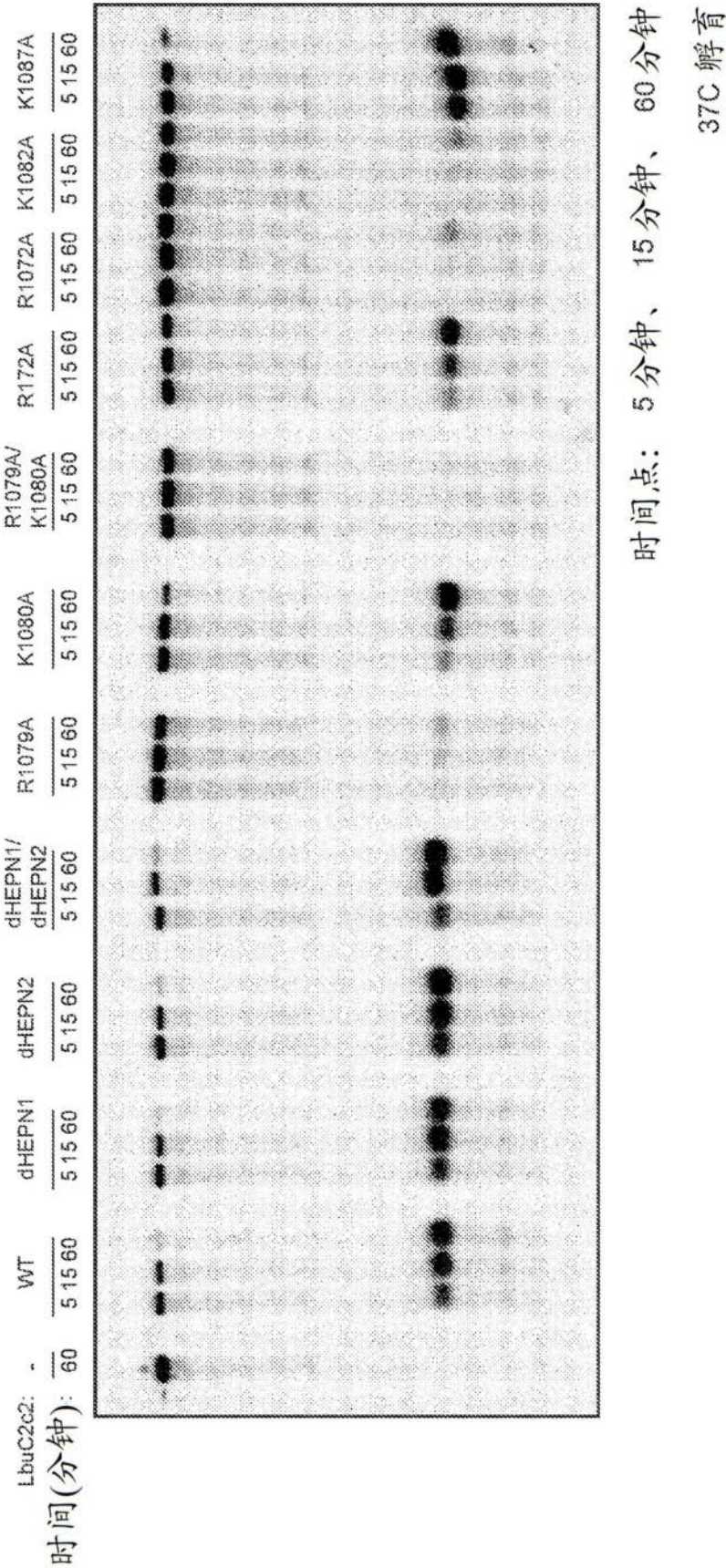


图37

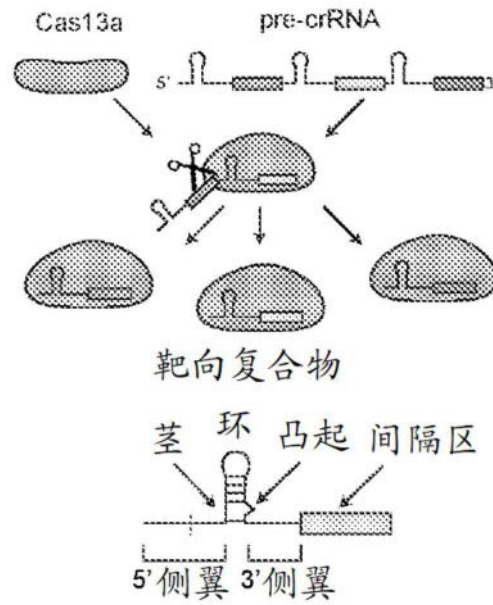


图38A

	5' 侧翼	茎
Lbu:	gGAUUUA GACC	ACCCC
Lse:	gGUAAGA CACU	ACCUCU
Ppr:	ggCUUGUG AAUU	AUCCC
Rca:	gGUCA CAUC	ACCGCC
Lwa:	gGAUAUA GACC	ACCCC
Lne:	gGAUUUA GAGU	ACCUC
Lba:	ggGCUGGAG AGAU	AGCCC
Ere:	ggGUGA AUAC	AGCUC
Cam:	ggGUUUGG CAAC	AGCCC
Lsh:	ggAUADAGAC C	ACCCC
Hhe:	gGUA -ACA	AUCCCC

图38B

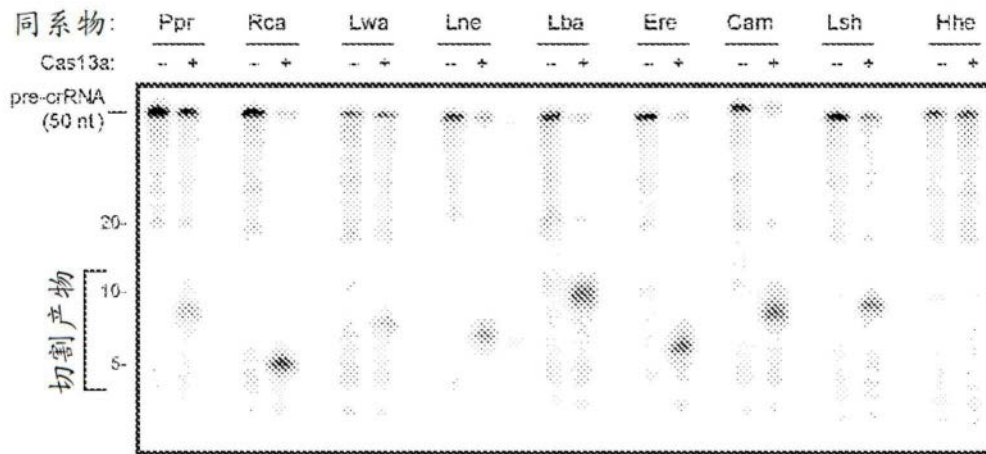


图38C

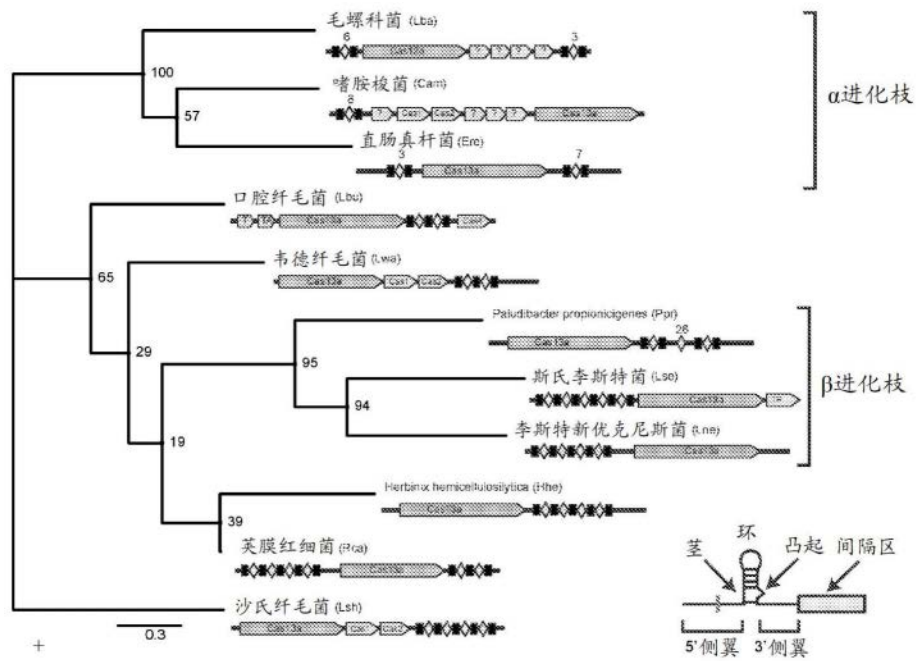


图39A

	5'侧翼	茎	环	茎+凸起	3'侧翼
Lbu:	GAUUUA\$GACC	ACCCC	AAAAAUGAA	GGGG	ACU AAAACA
Lse:	GAUAAG\$GACU	ACCUCC	AUA-UGAA	ACAGG	ACU AAAAC
Ppr:	GCUGUG\$AAUU	AUCUCC	AAAABUGAA	GGGA	ACU ACAAC
Rca:	GCUC\$CAUC	ACCGCC	AGA-CCAC	GCCGG	ACU G\$ACU
Lwa:	GAUAUA\$GACC	ACCCC	AAAUCCGAA	GGGG	ACU AAAACUU
Lne:	GAUUA\$GAGU	ACCUCC	AAAACAA\$A	GAGG	ACU AAAAC
Lba:	GCUGGAG\$AGAU	AGUCC	AAGAAACA	GGGC	AAU AAC
Ere:	GGGGA\$AUAC	AGUCC	GAUAUACU	GAGC	AAU AAG
Cam:	GUUUGG\$GAAC	AGUCC	GAUAUACA	GAGC	AAU AG\$C
Lsh:	GAUAUA\$GACU	ACCCC	AAAUCCGAA	GGGG	ACU AAAAC
Hhe:	GA -ACA	AUCCCC	GUAGAC-A	GGGCA	ACU GUG
	ggatagGA -ACA	AUCCCC	GUAGAC-A	GGGCA	ACU GUG

图39B

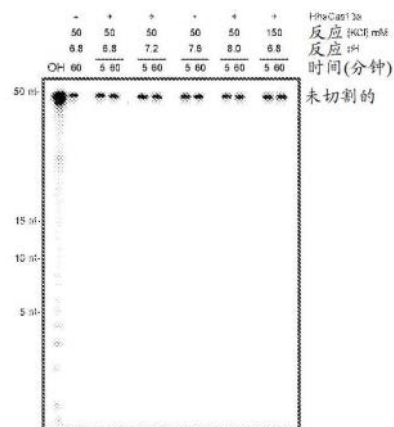


图39C

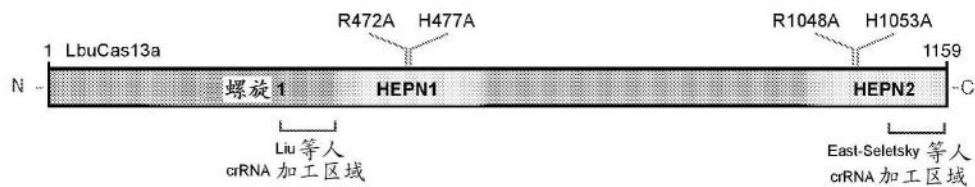


图40A

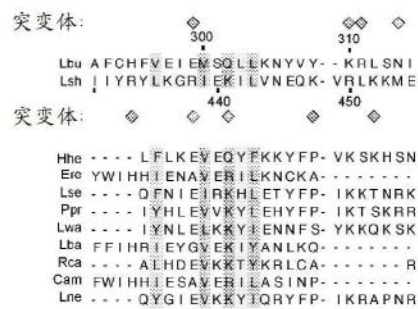


图40B



图40C

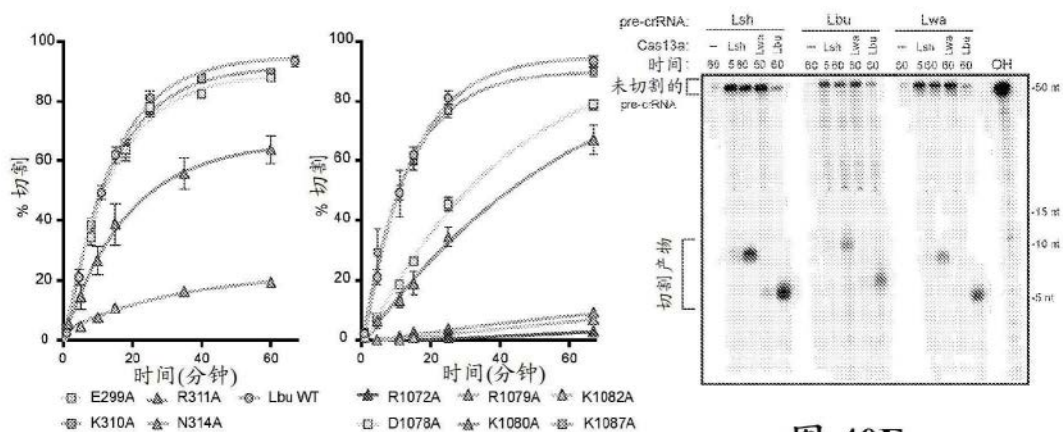
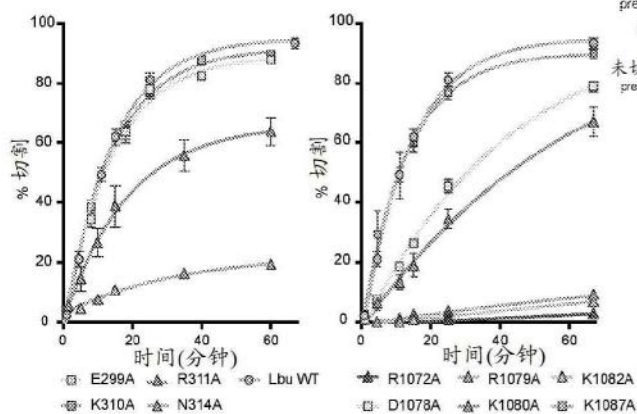


图 40F



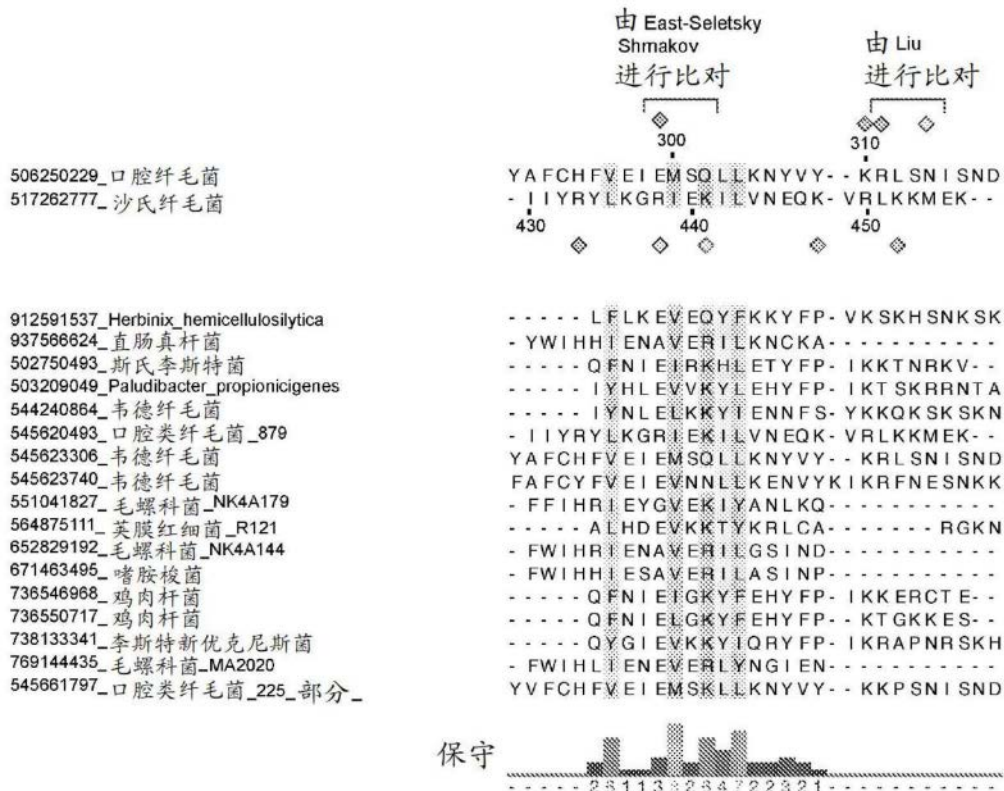


图41A



图41B

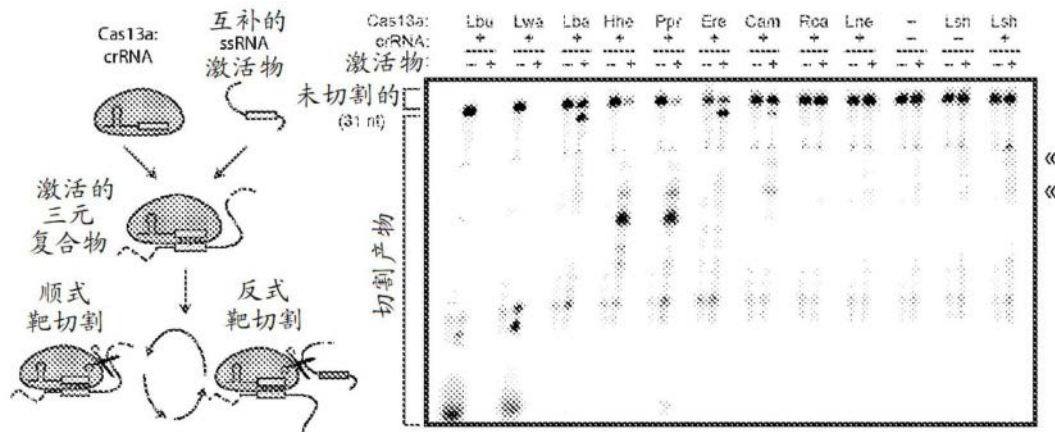


图 42A

图 42B

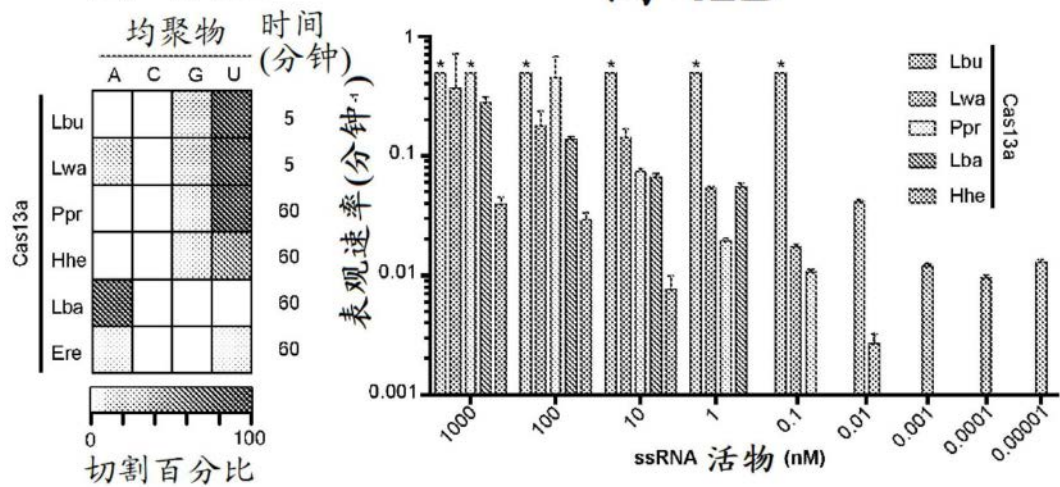


图 42C

图 42D

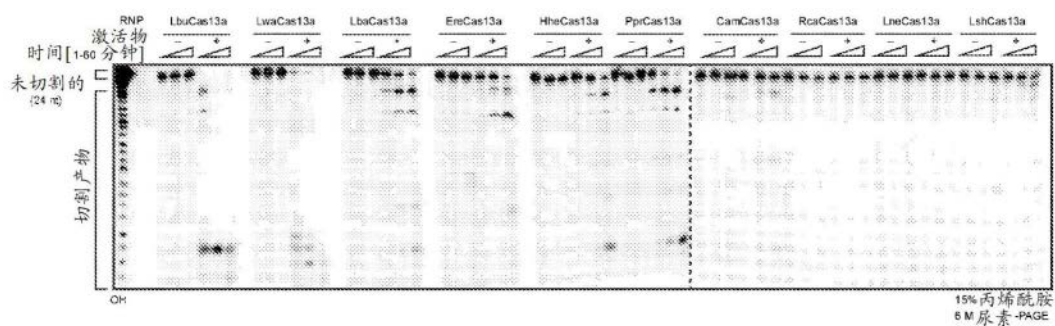


图43A

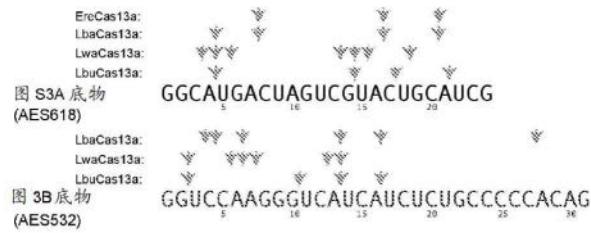
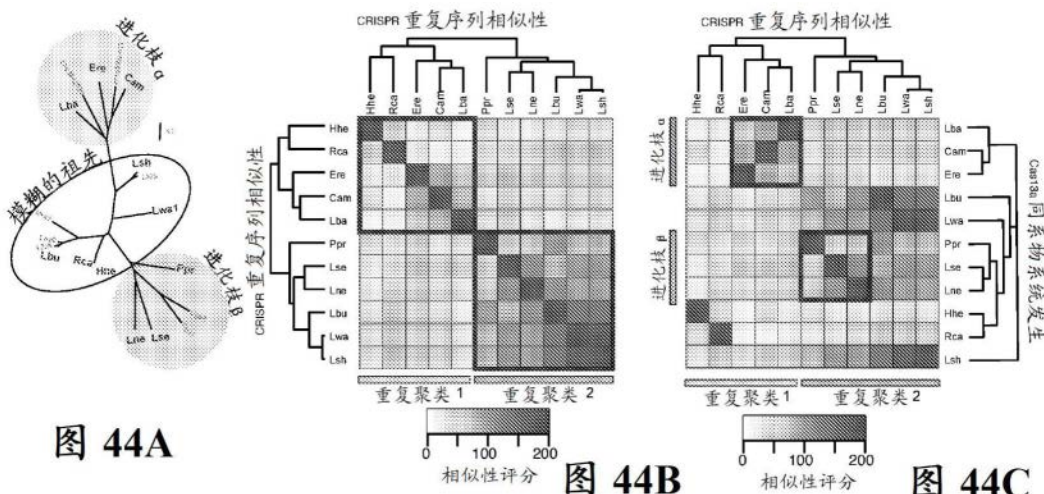
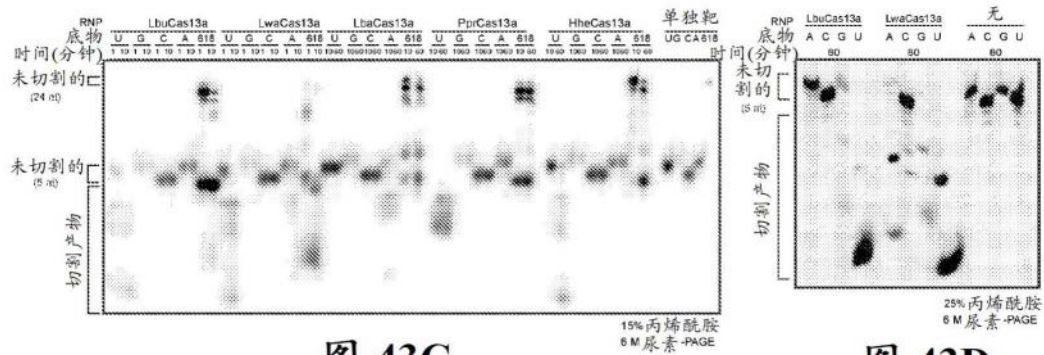


图43B



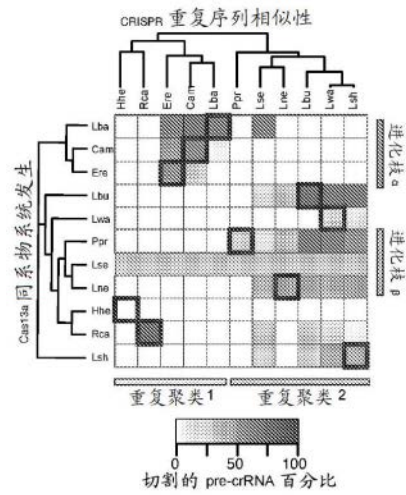


图44D

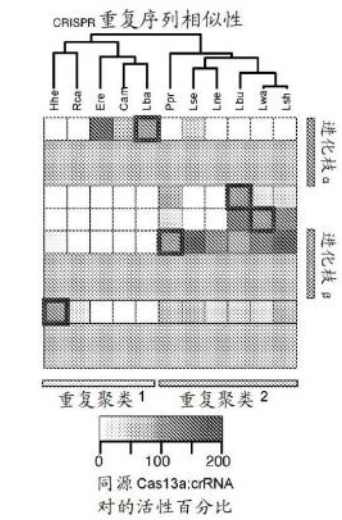


图44E

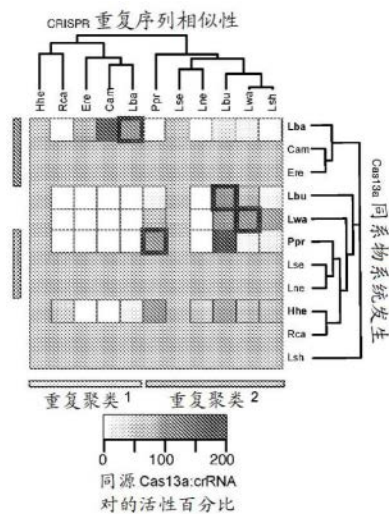


图44F

总体反应

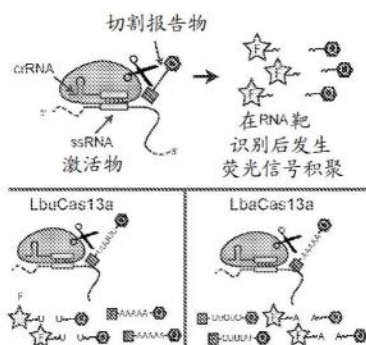


图45A

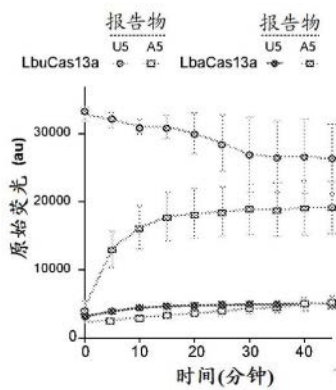


图45B

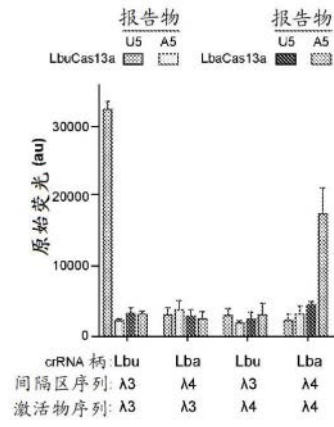


图45C

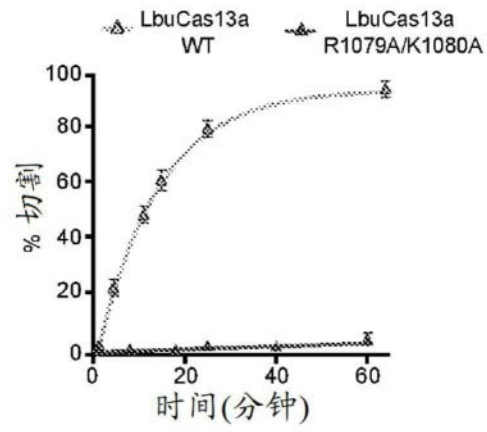


图46A

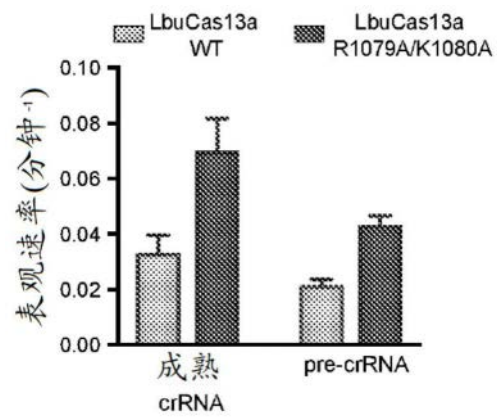


图46B

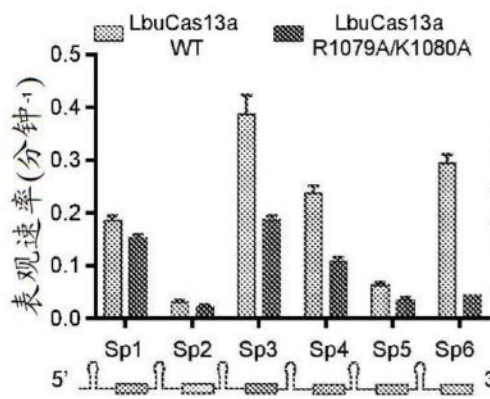


图 46C

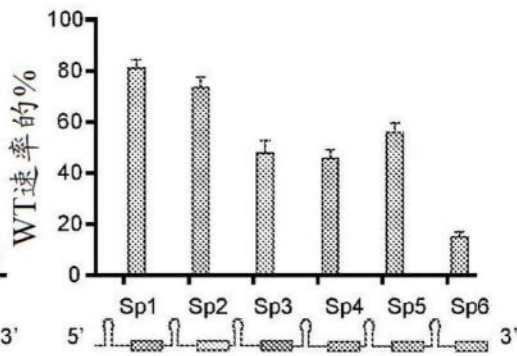


图 46D

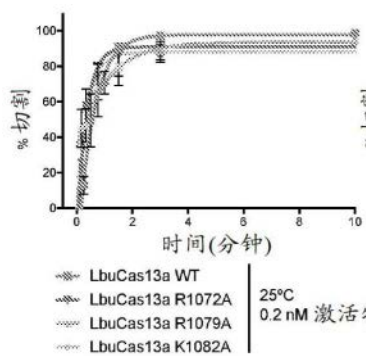


图 47A

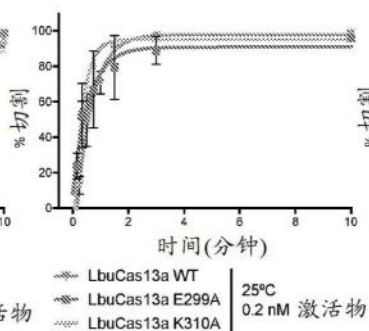


图 47B

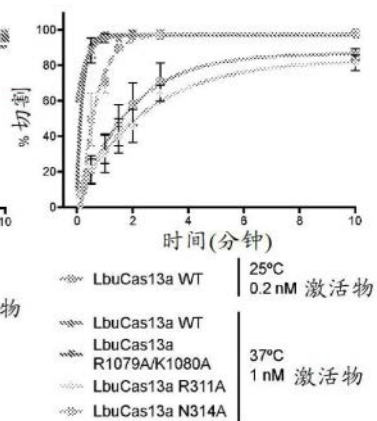


图 47C

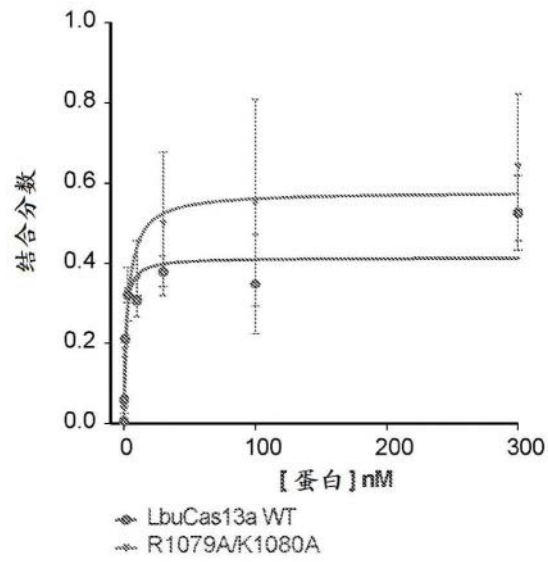


图48A

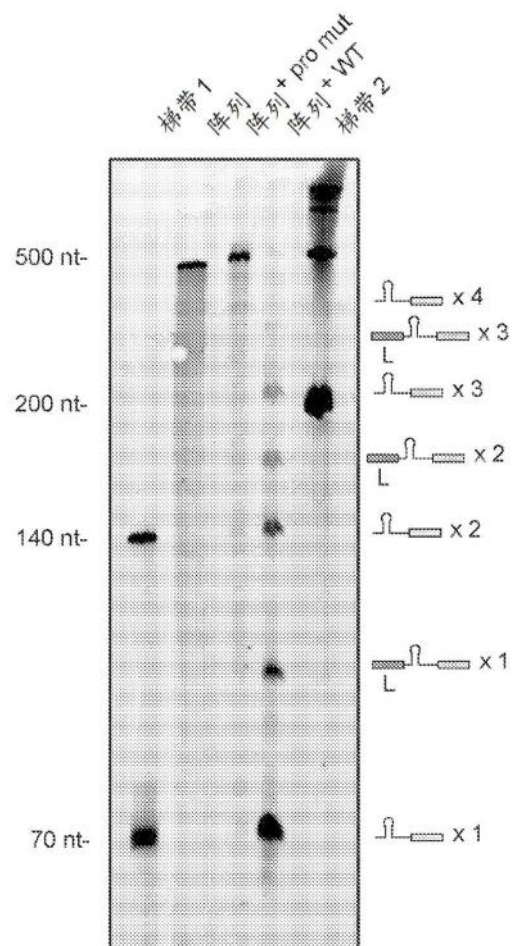


图48B

表 5

菌株/核酸编号*	类型	来源**	说明	序列
AES559	pre-crRNA	SS	Cam pre-crRNA L4	GGGUUUGGAGAACAGCCCGAUUAGAGGGCAUAGACCAGAUUAGCCUGGUGGUUCAGGC
AES557	pre-crRNA	SS	Ere pre-crRNA L4	GGGUGAAUACAGCUCGAUUAUGAGGCAUUAAGCAGAUUAGCCUGGUGGUUCAGGC
AES520	pre-crRNA	SS	Hhe pre-crRNA L4	GGUUAUGUAACAUCGCCGUAGACAGGGGAACUGCAACCAGAUUAGCCUGGUGGUUCAGGC
AES556	pre-crRNA	SS	Lba pre-crRNA L4	GGGUGGAGAGAAUAGCCCAAGAAAGAGGGCAUUAACCAGAUUAGCCUGGUGGUUCAGGC
AES534	pre-crRNA	SS	Lbu pre-crRNA L4	GGAUUUAGACCAACCCCAAAUUGAAGGGGACUAAACACAGAUUAGCCUGGUGGUUCAGGC
AES619	pre-crRNA	SS	Lne pre-crRNA L4	GGAUUUAGAGUACCUCAAAACAAAAGAGGACUAAACACAGAUUAGCCUGGUGGUUCAGGC
AES625	pre-crRNA	SS	Lse pre-crRNA L4	GGUUAAGAGACUACCUUUAUUGAAGAGGACUAAACACAGAUUAGCCUGGUGGUUCAGGC
AES586	pre-crRNA	SS	Lsh pre-crRNA L4	GGAUUAGACCAACCCCAUUAUCGAAGGGGACUAAACACAGAUUAGCCUGGUGGUUCAGGC
AES553	pre-crRNA	SS	Lwa pre-crRNA L4	GGGAUUAAGACCAACCCCAUUAUCGAAGGGGACUAAACACUUCAGAUUAGCCUGGUGGUUCAGGC
AES597	pre-crRNA	SS	Ppr pre-crRNA L4	GGCUUGUGAAUUAUCCAAAUAUGAAGGGAAUACAACAGAUUAGCCUGGUGGUUCAGGC
AES588	pre-crRNA	SS	Rca pre-crRNA L4	GGUCACAUACCCGCAAGACGACGGCGGACUGAACCCAGAUUAGCCUGGUGGUUCAGGC
AES614	crRNA	HH PCR	未切割的HH-Cam crRNA	GGUUGUCUGAUGAGGCCUUCGGGCCGAAACGGUGAAGCCGUAAAGAACGCCCGAUUAGAGGG CAUAGACACAGAUUAGCCUGGUGGUUCAGGC
			Cam crRNA L4	GAACAGCCCGAUUAGAGGGCAUUAAGACCAGAUUAGCCUGGUGGUUCAGGC
AES610	crRNA	HH PCR	未切割的HH-Ere crRNA	GGUGUAUUCUGAUGAGGCCUUCGGGCCGAAACGGUGAAGCCGUAAAUACAGCUCGAUUAUGAG CAUUAAGCAGAUUAGCCUGGUGGUUCAGGC
			Ere crRNA L4	AAGUAGCCCGAUUAGAGGGCAUUAACCAGAUUAGCCUGGUGGUUCAGGC
AES612	crRNA	HH PCR	未切割的HH-Lba crRNA	GGUAUCUUCUGAUGAGGCCUUCGGGCCGAAACGGUGAAGCCGUAAAGAUAGCCCAAGAAAGAGGG CAUUAACCAGAUUAGCCUGGUGGUUCAGGC
			Lba crRNA L4	AGAUAGCCCAAGAAAGAGGGCAUUAACCAGAUUAGCCUGGUGGUUCAGGC
MOC410	crRNA	SS	Lbu crRNA L3	GGCCACCCCAAAAUAGAGGGGACUAAACACUGGUAACUUCGGAUAGUG
MOC411	crRNA	SS	Lbu crRNA L4	GGCCACCCCAAAAUAGAGGGGACUAAACACAGAUUAGCCUGGUGGUUC
SCK340	crRNA	SS	Lbu crRNA Sp3	GACCACCCCAAAAUAGAGGGGACUAAACAGGGGACAGAUUAGACCU
AES641	crRNA	SS	Lne crRNA L4	GAGUACCUCAAAACAAAGAGGACUAAACACAGAUUAGCCUGGUGGUUCAGGC
AES646	crRNA	SS	Lsh crRNA (gggc) L4	GGGCCACCCCAUUAUCGAAGGGGACUAAACACAGAUUAGCCUGGUGGUUCAGGC
AES647	crRNA	SS	Lwa crRNA L4	GACCACCCCAUUAUCGAAGGGGACUAAACUUCAGAUUAGCCUGGUGGUUCAGGC
AES639	crRNA	HH PCR	未切割的HH-Ppr crRNA	GGAUAAUUCUGAUGAGGCCUUCGGGCCGAAACGGUGAAGCCGUAAAUUAUCCCAAAUUAAGGG AACUACAACCCAGAUUAGCCUGGUGGUUCAGGC
			Ppr crRNA L4	AAUUAUCCCAAAUUAAGGGGAACUACAACCAGAUUAGCCUGGUGGUUCAGGC
AES642	crRNA	HH PCR	未切割的HH-Rca crRNA	GGGUGAUGCUGAUGAGGCCUUCGGGCCGAAACGGUGAAGCCGUACAUCACCCGCAAGACGACGGC GGACUGAACCCAGAUUAGCCUGGUGGUUCAGGC
			Rca-crRNA L4	CATCACCAGCAAGACGACGGCGACTGAACCATATAGCTGGTGTTCAGGC
AES618	靶	SS	18 聚体随机序列	GGCAUGACUAGUGCAUCUGCAUCG
MOC36	靶	SS	Sp1/L3 靶	GGAAUUAUUAACACCCGACUUAUCCGAAGUUCACAGCCAGCCGACGACGUU
MOC37	靶	SS	Sp2/L4 靶	GGCAUUAUUAUAGCGCCGCGGACCAACACAGGCUUAUUAUCGCCACUUAUUGUUGA
AES532	靶	SS	Sp3 靶-31 聚体	GGUCCAAGGGU/CAUCAUCUCGCCCCACAG
AES450	靶	SS	Sp3 靶-60 聚体	GGCACACCCGAGGGGGAGGCAAAAGGGUACAUCUCUGCCCCACAGCAGAAGCCCC
AES653	靶	SS	Sp4 靶	GGACCAUUAUUAUAGGUAAGUAGAUUGU/CAUAUUGGCAG
AES452	靶	SS	Sp5 靶	GGGAACCCCAAGGCCAACCGGAGAGAGUACCCAGAUCAUGUUUGAGACCUUAACACCCC
MOC28	靶	SS	Sp6/L2 靶	GGCUCAAUUUUGACAGCGGCAUUGGCAUUAUCCAUUAUACUUGGCAUCCUCCACUC
AES652	靶	IDT	FQ A5	/56-FAM/rArArArArA/3'ABkFQ/
MOC533	靶	IDT	FQ U5	/56-FAM/rUrUrUrUrU/3'ABkFQ/
AES626	靶	IDT	均聚 A	rArArArArA
AES627	靶	IDT	均聚 C	rCrCrCrCrC
AES628	靶	IDT	均聚 G	rGrGrGrGrG
AES629	靶	IDT	均聚 U	rUrUrUrUrU
Lbu-CRISPR 阵列	pre-crRNA 阵列	PCR	六聚体阵列前导区 -Sp1-Sp2-Sp3-Sp4-Sp5-Sp6	GGGCGAAUUGAAGGAAGGCCGUAAGGCCGUAUUAUUAACGACUACUUAAGGAUUAUAGACC ACCCCAAAAUAGAGGGGACUAAACACUGGUAACUUAUAGGUGGUGUUAAGAUUAUAG ACCAACCCCAAAAUAGAGGGGACUAAACACAGAUUAGCCUGGUGGUUACAGCGGCCGAUUAUUA AGACCAACCCCAAAAUAGAGGGGACUAAACAGGGGACAGAUUAGGACCUUUAUUGGCUCCCGAU UUAAGACCAACCCCAAAAUAGAGGGGACUAAACAAUUAUGACACAUUUAUUAUUAUUAUUAUUA AUUUAAGACCAACCCCAAAAUAGAGGGGACUAAACAAUUAUUAUUAUUAUUAUUAUUAUUAUUA GGAUUAUUAAGACCAACCCCAAAAUAGAGGGGACUAAACAAUUAUUAUUAUUAUUAUUAUUAUUA GCU

图50

表 6

Cast3a 同系物	rArArArArA		rCrCrCrCrC		rGrGrGrGrG		rUrUrUrUrU	
	Avg	SD	Avg	SD	Avg	SD	Avg	SD
Lbu	2.68	0.15	0.98	0.06	17.83	0.79	92.69	2.10
Lwa	13.67	0.92	0.87	0.10	18.97	1.31	90.87	1.78
Lba	82.34	5.35	3.01	1.16	2.23	0.75	1.73	0.70
Ppr	1.19	0.28	0.50	0.14	10.21	6.63	93.20	4.78
Hhe	3.56	0.51	1.01	0.12	10.34	8.55	48.04	9.07
Ere	6.32	0.18	2.42	0.03	2.37	1.58	4.25	1.99

图51

表 7

说明性编号	RNA 编号	Lbu WT		CamCas13a		EreCas13a		LbaCas13a		LwaCas13a		RcaCas13a		HheCas13a		PprCas13a		LshCas13a		LnyCas13a	
		Avg	SD	Avg	SD	Avg	SD	Avg	SD	Avg	SD	Avg	SD	Avg	SD	Avg	SD	Avg	SD	Avg	SD
pre-crLbu	AES834	82.0	3.1	0.0	3.2	0.3	0.7	0.2	0.7	3.6	2.1	44.0	29.7	4.7	3.1	76.5	5.5	40.8	9.7	58.8	38.5
pre-crLse	AES625	29.1	11.9	0.8	1.3	0.3	1.7	74.7	7.6	0.9	0.9	49.3	28.4	-0.5	2.2	27.0	4.7	39.9	7.8	55.9	21.1
pre-crLba1	AES555	1.7	2.7	15.3	6.8	2.5	1.8	75.9	7.0	2.5	3.1	2.5	2.8	0.8	2.0	1.3	3.0	-0.2	1.8	2.8	2.2
pre-crLba2	AES556	0.1	1.9	7.6	2.5	-0.1	3.7	75.4	7.1	0.4	3.4	-0.7	4.1	-1.9	5.0	-2.2	5.5	-2.5	4.1	0.3	7.3
pre-crEre2	AES557	1.6	1.4	63.9	6.0	66.7	1.7	76.2	6.4	0.7	1.4	-1.4	3.8	-2.2	3.7	-1.4	3.5	-2.7	8.3	-1.8	6.3
pre-crEre1	AES558	-0.9	2.3	68.5	9.5	66.4	6.8	74.5	9.7	-0.3	2.7	0.4	3.4	-1.7	5.1	0.2	4.9	-0.5	7.1	2.5	3.2
pre-crCam	AES559	-0.1	1.2	72.1	4.4	33.5	5.0	77.7	3.4	0.7	0.9	-0.2	0.4	-1.2	0.6	-1.3	1.1	-0.8	0.7	0.6	0.3
pre-crLwa	AES553	78.4	9.4	2.8	1.8	1.1	2.6	2.8	2.0	32.7	10.1	29.4	8.6	3.5	2.3	74.4	9.3	59.9	9.5	56.6	5.8
pre-crLsh	AES566	79.0	7.0	1.3	2.7	0.8	1.7	1.9	2.4	24.0	9.4	6.1	1.8	5.4	8.1	69.0	11.5	59.9	9.6	80.9	16.8
pre-crRca	AES568	1.6	1.4	0.8	2.0	1.5	2.8	2.1	2.3	4.6	2.7	80.7	3.4	5.4	3.6	4.4	2.0	5.4	5.0	4.4	3.5
pre-crPpr	AES567	-1.0	2.2	-0.8	1.2	-1.0	1.8	2.1	2.3	-1.0	2.2	-0.2	2.8	-1.5	2.3	41.3	4.9	-1.0	2.5	1.5	1.9
pre-crHhe	AES820	5.8	4.5	3.9	14.3	1.9	10.4	-1.0	6.9	0.5	4.4	-2.8	9.7	-6.6	5.2	3.5	9.4	-6.5	6.6	3.8	10.8
pre-crLne	AES619	48.6	1.7	0.7	2.3	-0.5	1.3	-0.1	0.3	-0.9	3.2	5.6	3.6	-2.4	1.5	50.6	4.4	1.3	2.9	61.4	1.1

图52

表 8

指导物	LwaCas13a		LbeCas13a		LbuCas13a		PprCas13a		HheCas13a	
	斜率	SE	斜率	SE	斜率	SE	斜率	SE	斜率	SE
Lwa pre-crRNA	100.0	4.2	124.5	37.4	35.6	1.9	31.5	6.4	1.5	0.2
Lba pre-crRNA	-0.3	0.0	0.0	0.0	1.8	0.1	-0.8	-0.1	100.0	15.8
Ere pre-crRNA	-0.2	0.0	-0.1	0.0	3.2	0.2	0.0	0.0	152.5	27.3
Cam pre-crRNA	-0.3	0.0	0.0	0.0	2.7	0.2	0.0	0.0	77.8	10.0
Lsh pre-crRNA	120.7	7.7	166.0	49.8	71.7	4.8	30.7	8.3	2.4	0.3
Rca pre-crRNA	0.0	0.0	0.0	-0.7	26.6	1.4	0.1	0.0	0.8	0.1
Ppr pre-crRNA	37.4	1.4	100.0	42.4	66.4	4.3	86.7	22.8	1.5	0.2
Lne pre-crRNA	-0.2	0.0	104.6	31.4	29.8	1.6	7.1	1.6	1.3	0.2
Hhe pre-crRNA	0.0	0.0	16.2	4.9	100.0	7.3	-0.1	0.0	0.8	0.1
Lse pre-crRNA	1.5	0.0	170.2	51.1	63.7	3.6	6.8	1.8	51.5	8.0
Lbu pre-crRNA	140.0	12.0	90.0	27.0	42.6	2.2	100.0	28.0	7.8	0.9

图53

表 9

指导物	LwaCas13a		LbeCas13a		LbuCas13a		PprCas13a		HheCas13a	
	斜率	SE	斜率	SE	斜率	SE	斜率	SE	斜率	SE
Cam crRNA	0.0	0.1	174.8	50.9	-0.2	0.0	-0.9	-0.1	7.4	0.4
Lba crRNA	-0.4	-0.1	100.0	22.3	0.0	0.0	-1.2	-0.1	12.6	0.7
Ere crRNA	-0.5	-0.1	81.7	23.5	0.1	0.0	-0.8	-0.1	6.5	0.4
Ppr crRNA	77.8	14.3	1.5	0.2	11.6	0.7	100.0	2.8	100.0	7.5
Lne crRNA	0.1	0.0	1.8	0.3	0.1	0.0	-0.2	-0.1	54.7	3.0
Rca crRNA	-0.1	-0.1	3.9	0.6	-0.8	0.0	-0.1	-0.1	50.4	2.7
Lsh crRNA	93.3	21.5	0.9	0.2	10.0	0.5	27.2	1.1	75.9	4.5
Lwa crRNA	100.0	24.3	10.6	1.7	75.7	5.5	12.8	0.4	60.4	3.7
Lbu crRNA	98.2	18.2	21.1	3.4	100.0	6.1	398.3	20.8	92.6	7.0

图54

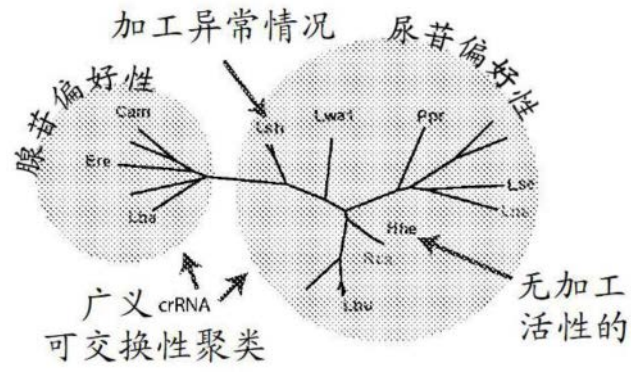


图55A

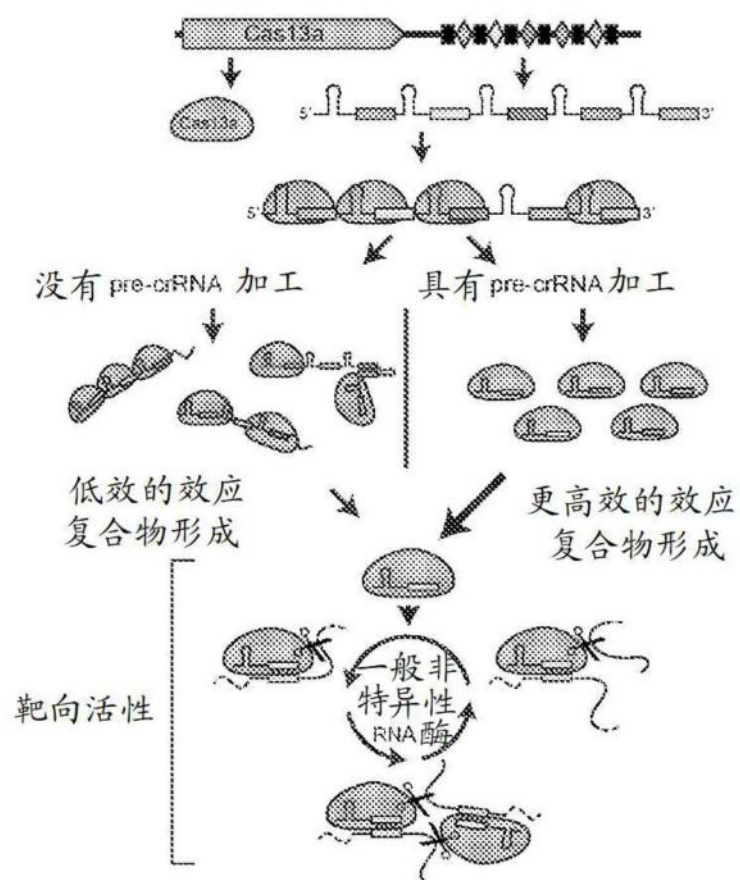


图55B

>WP_012985477.1 WP_012985477.1 假想蛋白[斯氏李斯特菌]
 MWISIKTLIHHLGVLFCDYMYNRREKKIIEVKTMRTKVEVDRKKVLISRDKNNGGKLVY
 ENEMQDNTREQIMHHKSSFYKSVVNKTICRPEQKQMKKLHGLLQENSQEKIKVSDVTKL
 NISNLFNHRFKSLYYFPENSPDKSEERYIEINLSQILLEDLSKKQQGTFCWESFSKME
 LYINWAENYISSKTKLIKKSIRNNRIQSTESRSGQLMDRYMKDILNKNKPFDIQSVSEKY
 QLEKLT SALKATFKEAKNDKEINYKLKSTLQNHHERQII EELKENSELNQFNIEIRKHLE
 TYFPIKKTNRKVGDINLEIGEIQKIVNHRLLKNKIVQRILQEGKLASYEIESTVNSNSLQ
 KIKIEEAFALKFINACLFASNNLRNMVYPVCKKDILMIGEFKNSFKEIKHKKFIRQWSQF
 FSQEITVDDIELASWGLRGAIAPIRNEIIHLKKHSHWKKFFNNPTFKVKKSKIINGKT KDV
 TSEFLYKETLFDKYFYSELDSVPELIINKMESSKILDYSSDQLNQVFTIPNFELSLTS
 AVPFAPSFKRVLKGFYQNDQDEAQP DYNLKLNINYEKAFNSEAFQAQYSLFKMVYYQVF
 LPQFTTNNDLFKSSVDFILT LNKERKGYAKAFQDIRKMNKDEKPSYMSYIQSQLMLYQK
 KQEEKEKINHFEKFINQVFIKGFNSFIEKNRLTYICHPTKNTVPENDNIEIPFHTDMDDS
 NIAFWLMCKLLDAKQLSELNEMIKFSCSLQSTEEISTFTKAREVIGLALLNGEKGCDNW
 KELFDDKEAWKKNMSLYVSEELLQSLPYTQEDGQTPVINRSIDL VKKYGTETILEKLFSS
 SDDYKVSADIAKLHEYDVTEKIAQQESLHKQWIEKPLGLARDSAWTKKYQNVINDISNYQ
 WAKTKVELTQVRHLHQLTIDL SRLAGYMSIADRDFQSSNYILERENSEYRVTSWILLS
 ENKNKNKYNDYELYNLKNASIKVSSKNDPQLKVDLKQLRLTLEYLELFDNRLKEKRNNIS
 HFNYLNGQLGNSILELFD DARDVLSYDRKLKNAVSKSLKEILSSHGMEVTFKPLYQTNHH
 LKIDKLQPKKIHLGKSTVSSNQVSNEYCQLVRTLLTMK

图56A

>WP_013443710.1 WP_013443710.1 假想蛋白 [Paludibacter
 propionicigenes]
 MRVSKVKVDGGKDKMVLVHRKTTGAQLVYSGQPVSNETSNILPEKKRQSF DLSTLNKTI
 IKFDTAKKQKLNVDQYKIVEKIFKYPKQELPKQIKAEIILPFLNHKFQEPVKYWKNGKEE
 SFNLTL LIVEAVQAQDKRKLQPYDWTWYIQTKSDLLKKSIIENNRIDL TENLSKRKKAL
 LAWETEFTASGSIDLTHYHKVYMTDVLCKMLQDVKPLTDDKGKINTNAYHRGLKKALQNH
 QPAIFGTREVPNEANRADNQLSIYHLEVVKYLEHYFPIKTSKRNTADDIAHYLKAQTLK
 TTIEKQLVNAIRANIIQQGKTNNHELKADTTSDNLI RIKTNEAFVLNLTGTCAFAANNIR
 NMVDNEQTNDILGKGDFIKSLLKDNTNSQLYSFFFGEGLSTNKA EKETQLWGIRGAVQQI
 RNNVNHYKKDALKTVFNISNFENPTITDPKQQTNYADTIYKARFINELEKIPEAFAQQLK
 TGGAVSYTTIENLKSLLTTFQFSLCRSTIPFAPGFKKVFNGGINYQNAKQDES FYELMLE
 QYLRKENFAEESYNARYFMLKLIYNNLFLPGFTTDRKAFADSVGFVQM QNKKQAEKVNPR
 KKEAYAFEAVRPMTAADS IADY MAYVQSELMQE QNKKEEKVAEETRINF EKEFVLQVFIKG
 FDSFLRAKEFDFVQMPQPQLTATASNQQKADKLNQLEASITADCKLTPQYAKADDATHIA
 FYVFCKLLDAAHLSNLRNELIKFRESVNEFKFHHLLEII EICLLSADVVP TDYRDLYSSE
 ADCLARLPFIEQGADITNWSDLFVQSDKHSPVIHANIELSVKYGTTKLLEQIINKDTQF
 KTTEANFTAWNTAQKSIEQLIKQREDHHEQWVKAKNADDKEKQERKREKS NFAQKFIEKH
 GDDYLDICDYINTYNWLDNKMHFVHLNRLHGLTIELLGRMAGFVALFDRDFQFFDEQQIA
 DEFKLHG FVNLSIDKKLNEVPTKKIKEIYDIRNKIIQINGNKINESVRANLIQFISSKR
 NYYNNAFLHVSND EIKEKQMYDIRNHIAHFNYLT KDAADFSLIDLINELRELLHYDRKLK
 NAVSKAFIDLFDKHGMILKLKLNADHKLKVESLEPKKIYHLGSSAKDKPEYQYCTNQVMM
 AYCNCMSLLEMKK

图56B

>WP_015770004.1 WP_015770004.1 假想蛋白[口腔纤毛菌]
 MKVTKVGGISHKYTSSEGRVLKSESEENRTDERLSALLNMRLDMYIKNPSSTETKENQKR
 IGKLLKFFSNKMVYLKDNLTSLKNGKKENIDREYSETDILESVDKKNFAVLKKIYLN
 NVNSEEELEVRNDIKKKLNKINSLKYSFEKNKANYQKINENNIEKVEGKSKRNIIYDYR
 ESAKRDAYVSNVKEAFDKLYKEEDIAKLVLIEIENLTKEKYKIREFYHEIIGRKNDKENF
 AKIIYEEIQNVNMMKELIEKVPDMSELKKSQVFYKYLDKEELNDKNIKYAFCHFVEIEM
 SQLLKNYVYKRLSNISNDKIKRIFEQNLKKLIENKLLNKLDITYVRNCGKYNYYLQDGEI
 ATSDFIARNRQNEAFLRNIIGVSSVAYFSLRNILETENENDITGRMRGKTVKNNKGEEKY
 VSGEVDKIYNENKKNEVKENLKMFSYDFNMDNKNEIEDFFANIDEAIISSIRHGIVHFNL
 ELEGKDI FAFKNIAPSEISKMMFQNEINEKKLKLKIFRQLNSANVFRYLEKYKILNLYLKR
 TRFEFVNKNIPFVPSFTKLYSRIDDLKNSLGIYWKTPKTNDNDNKTKEIIDAQIYLLKNIY
 YGEFLNYFMSNNGNFFEISKEIIELNKNDKRNLTGFKYKLQKFEDIQEKIPKEYLANIQS
 LYMINAGNQDEEEKDITYIDFIQKIFLKGFMITYLANNGRLSLIYIGSDEETNTSLAEKKQE
 FDKFLKKEEQNNNIKIPYEINEFLREIKLGNILKYTERLNMFYLLKLLNHKELTNLKG
 LEKYQSANKEEAFSDQLELINLLNLDNRRVTEDEFELEADEIGKFLDFNGNKVKDNKELKK
 FDTNKIYFDGENI IKHRAFYNIKKYGMLNLLLEKIADKAGYKISIEELKKYSNKKNEIEKN
 HKMQENLHRKYARPRKDEKFTDEDEYESYQAIENIEEYTHLKNKVEFNELNLLQGLLLRI
 LHRLVGYTSIWERDLRFRLKGEFPENQYIEEIEFNFNKKNVKYKGGQIVEKYIKFYKELH
 QNDEVKINKYSSANIKVLKQEKDLYIRNYIAHFNYIPHAIEISLLEVLENLRKLLSYDRK
 LKNAVMSVVDILKEYGFVATFKIGADKKIGIQTLESEKIVHLKNLKKKLMTDRNSEEL
 CKLVKIMFEYKMEEKSEN

图56C

>WP_018451595.1 WP_018451595.1 假想蛋白[沙氏纤毛菌]
 MGNLFGHKRWYEVDRDKDFKIKRKVKVRNYDGNKYILNINENNNKEKIDNNKFIRKYIN
 YKNDNLIKFEFTRKFHAGNIFKLKGKEGIIRIENNDDFLETEEVVLYIEAYGKSEKLKA
 LGITKKKIIDEAIRQGITKDDKKIEIKRQNEEEIEIDIRDEYTNKTLNDCSIILRIEN
 DELETKKSIYEIFKNINMSLYKIEKIIENETEKVFENRYEEHLREKLLKDDKIDVILT
 NFMEIREKIKSNLEILGFVKFYLNVGDKKKSKNKKMLVEKILNINVDLTVEDIADFIK
 ELEFWNITKRIEKVKVNNFELEKRRNRTYIKSYVLLDKHEKFKIERENKKDKIVKFFVE
 NIKNSIKEKIEKILAEFKIDELIKKLEKELKKGNCDTEIFGIFKKHYKVNFDSSKFSKK
 SDEEKELYKIYRYLKGRIEKILVNEQVRLKKMEKIEIEKILNESILSEKILKRVQYT
 LEHIMYLGKLRHNDIDMTTVNTDDFSRLHAKELDLELITFFASTNMELNKIFSRENINN
 DENIDFFGGDREKNYVLDKKILNSKIKIIRDLDFIDNKNITNNFIRKFTKIGTNERNRI
 LHAISKERDLQGTQDDYNKVINIIQNLKISDEEVSKALNLDVVFDDKKNIIITKINDIKIS
 EENNDIKYLPFSFVKLPEILNLYRNNPKNEPFDTIETEKIVLNALIYVNKELYKKLILE
 DDLEENESKNIFLQELKKTGLNIDEIDENIENYKNAQISASKGNNAIKKYQKKVIEC
 YIGYLRKNYEELFDFDFKMNIQEIKKQIKDINDNKTYERITVKTSDKTIVINDDFEYII
 SIFALLNSNAVINKIRNRFFATSVWLNTSEYQNIIDILDEIMQLNLTNRNENITENWNLNL
 EEFIQKMEIEKDFDDFKIQTKKEIFNNYEDIKNNILTEFKDDINGCDVLEKKLEKIVI
 FDDETKFEIDKKSNIQDEQRKLSNINKKDLKKKVDQYIKDKDQEI KSKILCRIIFNSDF
 LKKYKKEIDNLIEDMESENENKFQEIYYPKERKNELYIYKKNLFLNIGNPNFDKIYGLIS
 NDIKMAKFLFNIDGKNIRKNKISEIDAILKNLNDKLNKYSKEYKEYIKKLKENDDDFF
 AKNIQNKYKSFEDYNRVSEYKKIRDLVEFNLYLNKIESYLIDINWKLAIQMARFERDMH
 YIVNGLRELGIKLSGYNTGISRAYPKRNGSDGFYTTTAYYKFFDEESYKKFEKICYGFG
 IDLSENSEINKPENESIRNYISHFYIVRNPFADYSIAEQIDRVSNLLSYSTRYNNSTYAS
 VFEVFKKDVNLDYDELKKKFKLIGNNDILERLMKPKKVSVLELESYNSDYIKNLIIELLT
 KIENTNDTL

图56D

>ERK53440.1 ERK53440.1 假想蛋白 HMPREF9015_00520
 [韦德纤毛菌 F0279]
 MYMKITKIDGVSHYKKQDKGILKKKWKDLDERKQREKIEARYNKQIESKIYKEFFRLKNK
 KRIEKEEDQNIKSLYFFIKELYLNEKNEEWELKNINLEILDDKERVIKGYKFEDVYFFK
 EGYKEYYLRLFNFLIEKVQENREKVRKNKEFLDLKEIFKKYKNRKIDLLLSINNKI
 NLEYKKENVNEEIIYGINPTNDREMTFYELLKEIEKKDEQKSILEEKLDNFDITNFLENI
 EKIFNEETEINI IKGVLNELREYIKEKEENNSDNKQIYNLELKKYIENNFSYKKQKS
 KSKNGKNDYLYLNLFLKKIMFIEEVDEKKEINKEKFKNKINSNFKNLFVQHILDYGKLLYY
 KENDEYIKNTGQLETKDLEYIKTKETLIRKMAVLVSFAANSYYNLFGRVSGDILGTEVVK
 SSKTNVIKVGSHIFKEKMLNYFFDFEIFDANKIVEILESISYSIYNVRNGVGHFNKLILG
 KYKKKDINTNKRIEEDLNNNEEIKGYFIKKRGEIERKVKEKFLSNLQYYSKEKIENYF
 EVYEFELKRKIPFAPNFKRIIKKGEDLFNNKNNKKYEFKNFDKNSAEKKEFLKTRNF
 LLKELYNNNFYKEFLSKKEEFKIVLEVKEEKKSRGNINNKKSGVSFQSIDDYDTKINIS
 DYIASIHKKEMERVEKYNEEKQKDTAKYIRDFVEEIIFLTGFINYLEKDRLHFLKEEFSI
 LCNNNNNVDFNININEEKIKEFLKENDSKTLNLYLFFNMIDSKRISEFRNELVKYKQFT
 KKRLDEEKEFLGIKIELYETLIEFVILTREKLDTKKSEEIDAWLVDKLYVKDSNEYKEYE
 EILKLFVDEKILSSKEAPYYATDNKTPILLSNFEKTRKYGTQSFLSEIQSNYKYSKVEKE
 NIEDYNKKEEIEQKKKSNIKLQDLKVELHKKWEQNKITEKEIEKYNNTTRKINEYNYLK
 NKEELQNVYLLHEMLS DLLARNVAFFNKWERDFKFVIAIKQFLRENDKEKVNEFLNPPD
 NSKGKKVYFSVSKYKNTVENIDGIHKNFMNLI FLNNKFMNRKIDKMNCAI WVYFRNYIAH
 FLHLHTKNEKISLISQMNLLIKLFSYDKKVQNHILKSTKTLLKYNIQINFEISNDKNEV
 FKYKIKNRLYSKKGKMLGKNNKFEILENEFLENVKAMLEYSE

图56E

>WP_022785443.1 WP_022785443.1 假想蛋白
 [毛螺科菌 NK4A179]
 MKISKVREENRGAKLTVNAKTAVVSENRSQEGILYNDPSRYGKSRKNDIEDRDRIIESRLK
 SSGKLYRIFNEDKNKRETDDELQWFLSEIVKKINRRNGLVLSMDSVDDRAFEKAFKEYAE
 LSYTNRRNKVSGSPAFETCGVDAATAERLKGIISETNFINRIKNNIDNKVSEDIIDRIIA
 KYLKSLCRERVKRGLKLLMNAFDLPYSDPDIDVQRDFIDYVLEDFYHVRASQVRSRI
 KNMNMVPQPEGDGKFAITVSKGGTESGNKRSAEKEAFKKFLSDYASLDERVRDDMLRRMR
 RLVVLYFYGSDDSKLSDVNEKFDVWEDHAARRVDNREFIKLPLENKLANKTDDKAERIR
 KNTVKELYRNQNI GCYRQAVKAVEEDNNGRYFDDKMLNMFFIHRIEYGVKEKIYANLKQVT
 EFKARTGYLSEKIWKDLINYSIKYIAMGKAVVNYAMDELNASDKKEIELGKISEEYLSG
 ISSFDYELIKAEEMLQRETAVYVAFARHLSSQTVELDSENSDFLLLPKGTMDKNDKNK
 LASNNILNFLKDKETLRDTILQYFGGHS LWTDFPFDKYL AGGKDDVDFTDLKDVIIY SMR
 NDSFHYATENHNNGKWNKELISAMFEHETERMTVMKDKFY SNNLPMFYKNDDLKLLID
 LYKDNVERASQVPSFNKV FVRKNFPALVRDKDNLGIELDLKADADKGENELKFYNALYYM
 FKEIYYNAFLNDKNVRERFITKATKVADNYDRNKERNLKDRIKSAGSDEKKKLREQQLQNY
 IAENDFGQRIKNIVQVNPDYTLAQICQLIMTEYNQQNNGCMQKKSAAKDKINKDSYQHYK
 MLLLVNLRKAFLEFIKENYAFVLKPYKHDLCADKADFVPDFAKYVKPYAGLISRVAGSSEL
 QKWYIVSRFLSPAQANHMLGFLHSYKQYVWDIYRRASETGTEINHSIAEDKIAGVDITDV
 DAVIDLSVKLCGTISSEISDYFKDEVYAEYISSYLD FEYDGGNYKDSLNRFCNSDAVND
 QKVALYYDGEHPKLN RNIILSKLYGERRFLEKITDRVSRSDIVEYYKLKETSQYQTKGI
 FDSEDEQKNIKKFQEMKNIVEFRDLMDYSEIADQLQGQLINWIYLRERDL MNFQLGYHYA
 CLNNSDNKQATYVTLDYQGKKNRKINGAILYQICAMYINGLPLYVVDKDSSEWTVSDGKE
 STGAKIGEFYRYAKSFENTS DCYASGLEIFENISEHDNITELRNYIEHFRYYSSFDRSFL
 GIYSEVFDRFFTYDLKYRKNVPTILYNILLQH FVNVRFEFVSGKKMIGIDKKDRKIAKEK
 ECARITIREKNGVYSEQFTYKLKNGTVYVDARDKRYLQSIIRLLFYPEKVNMDemieVKE
 KKKPSDNN TGKGYSKRDRQQDRKEYDKYKEKKKKEGNFLSGMGGNINWDEINAQLKN

图56F

>ETD76934.1 ETD76934.1 假想蛋白 U717_11515 [荚膜红细菌 R121]
MQIGKVQGRITISEFGDPAGGLKRKISTDGKNRKELPAHLSSDPKALIGQWISGIDKIYRK
PDSRKSDGKAIHSPTPSKMQFDARDDLGEAFWKLVSEAGLAQDSYDQFKRRLHPYGDKE
QPADSGAKLKFEADPPEPQAFHGRWYGAMSKRGNDAKELAAALYEHVDEKRIDGQPKR
NPKTDKFAPGLVVARALGIESSVLPRGMARLARNWGEEIIQTYFVVDVAASVKEVAKAAV
SAAQAFDPPRQVSGRSLSPKVGFALEHLERVTGSKRCSFDPAAGPSVLALHDEVKKT
RLCARGKNAARAFPADKTELLALMRHTHENRVRNQVMRMGRVSEYRGQQAGDLAQSHYWT
SAGQTEIKESEIFVRLWVGAFALAGRSMKAWIDPMGKIVNTEKNDRDLTA AVNIRQVISN
KEMVAEAMARRGIYFGETPELDRLGAEGNEGFVFALLRYLRGCRNQTFHLGARAGFLKEI
RKELEKTRWGKAKEAEHVVLTDKTVAAIRAIIDNDAKALGARLLADLSGAFVAHYASKEH
FSTLYSEIVKAVKDAPEVSSGLPRLKLLKLRADGVRGYVHGLRDTRKHAFATKLPPPPAP
RELDDPATKARYIALRLYDGPFRAYASGITGTALAGPAARAKEAATALAQSVNVTKAYS
DVMEGRSSRLRPPNDGETLREYLSALTGETATEFRVQIGYESDSENARKQAEFTIENYRRD
MLAFMFEDIIRAKGFDWILKIEPGATAMTRAPVLPEPIDTRGQYEHWQAALYLVMHFVPA
SDVSNLLHQLRKWEALQGYELVQDGDATDQADARREALDLVKRFRDVLVFLKTGEARF
EGRAAPFDLKPFRALFANPATFDRLFMATPTTARPAEDDPEGDGASEPELRVARTLRGLR
QIARYNHMAVLSDLEFAKHKVRDEEVARLAEIEDETQEKSQIVAAQELRTDLHDKVMKCHP
KTISPEERQSYAAAIKTIEHRFLVGRVYLGDLHLRLHRLMMDVIGRLIDYAGAYERDTGT
FLINASKQLGAGADWAVTIAGAANTDARTQTRKDLAHFNVLD RADGTPDLTALVNRAREM
MAYDRKRKNAVPRSILDMLARLGLTLKWQMKDHLQDATITQAAIKHLDKVRLTVGGPAA
VTEARFSQDYLMVA AVFNGSVQNPKRPRRRDDGDAWHKPPKPATAQSQPDQKPPNKAPSA
GSRLPPPQVGEVYEGVVVKVIDTGSLGFLAVEGVAGNIGLHISRLRRIREDAIIVGRRYR
FRVEIYVPPKSNTSKLNAADLVRIID

图56G

>WP_031473346.1 WP_031473346.1 假想蛋白[嗜胺梭菌]
MKFSKVDHTRSAVGIQKATDSVHGMLYTDPPKKQEVNDLDRFDQINVKAKRLYNVFNQSK
AEEDDDDEKRFQKVVVKLNRELKDLLFHREVSRYNSIGNAKYNYGIIKSNPEEIVSNLGMV
ESLKGGERDPQKVISKLLLYLRKGLKPGTDGLRMILEASCGLRKLSGDEKELKVFLQTL
EDFEKKTFFKNLIRSIENQNMVQPSNEGDP IIGITQGRFNSQKNEEKSAIERMMSMYAD
LNEDHREDVLRKLRLNVLVFNVDTEKTEEPTLPGEVDNPNVFEVWHDHEKGKENDRQFA
TFAKILTEDRETRKKEKLAVKEALNDLKS AIRDHNIMAYRCSIKVTEQDKDGLFFEDQRI
NRFWIHIESAVERILASINPEKLYKLRIGYLGKVKWDLNLYLSIKYIAGKAVFHFAM
EDLGKTGQDIELGKLSNSVSGGLTSFDYEQIRADETLQRQLSVEVAFANLFRVVGQT
GKKIEQSKSEENEEFLLWKAEKIAESIKKEGEGNTLKSILQFFGGASSWDLNHFCAAYG
NESSALGYETKFADDLRKAIYSLRNETFHFTTLNKGSFDWNAKLIGDMFSHEAATGIAVE
RTRFYSNNLPMFYRESDLKRIMDHLTYHPRASQVPSFNSVFVRKNFRLFLSNTLNTNT
SFDTEVYQKWESGVYYLFKEIYYNSFLPSGDAHHLFFEGLRRIRKEADNLPVIGKEAKKR
NAVQDFGRRCDELKNLSLSAICQMIMTEYNEQNNGNRKVKSTREDKRKPDIFQHYKMLLL
RTLQEAFAIYIRREEFKFIFDLPKTLYVMKPVVEEFLPNWKS GMFDSLVERVKQSPDLQRW
YVLCKFLNGRLNQLSGVIRS YIQFAGDIQRRAKANHNRLYMDNTQRVEYYSNVLEVVD
CIKGTSRFSNVFSDYFRDEDAYADYLDNYLQFKDEKIAEVSSFAALKTFCNEEEVKAGIY
MDGENPVMQRNIVMAKLFPGPEVLKNVVPKVTREEIEEYYQLEKQIAPYRQNGYCKSEED
QKKLLRFQRIKNRVEFQTITEFSEIINELLGQLISWSFLRERDLYFQLGFHYLCLHNDT
EKPAEYKEISREDGTVIRNAILHQVAAMYVGGLPVYTLADKKLA AFEKGEADCKLSISKD
TAGAGKKIKDFFRYSKYVLIKDRMLTDQNQKYTIYLAGLELFENTDEHDNITDVRKYVDH
FKYYATSDENAMSILDLYSEIHDRFFTYDMKYQKNVANMLENILLRHFVLRPEFFTGSK
KVGEKKITCKARAQIEIAENGMRSEDFTYKLSDGKKNISTCMIAARDQKYLNTVARLLY
YPHEAKKSIVDTREKKNKKTNRGDGT FNKQKGTARKEKDNGPREFNDTGFSTNPFAGFD
PFRNS

图56H

>WP_036091002.1 WP_036091002.1 假想蛋白[李斯特新优克尼斯菌]
MKITKMRVDGRTIVMERTSKEGQLGYEGIDGNKTTEIIFDKKESFYKSILNKTVRKPD
KEKNRRKQAINKAINKEITELMLAVLHQEVPSQKLHNLKSLNTESLTKLFKPKFQNMISY
PPSKGAEHVQFCLTDIAVPAIRDLD EIKPDWGIFFEKLPYTDWAESYIHYKQTTIQKSI
EQNKIQSPDSPRKLVLQKYVTAFLNGEPLGLDLVAKKYKLADLAESFKLVDLNEKDSANY
KIKACLQHQRNILDELKEDPELNQYGIEVKKYIQRYFPIKRAPNRSKHARADFLKKELI
ESTVEQQFKNVYHYVLEQQGMEAYELTDPKTKDLQDIRSGEAFSFKFINACAFASNNLK
MILNPECEKDILGKGNFKKNLPNSTTRSDVVKMIPFFSDELQNVNFDEAIWAIRGSIQQ
IRNEVYHCKKHWSKLSILKIGFEFEPNNMKYADSDMQKLMKDIAKIPEFIEEKLKSSGV
VRFYRHDELQSIWEMKQGFSLTTNAPFVPSFKRVYAKGHYQTSKNRYNLDLTTFDIL
EYGEEDFRARYFLTCLVYYQQFMPWFTADNNAFRDAANFVLRNLNKNRQQDAKAFINIREV
ELEGEMPRDYMGYVQGQIAIHEDSIEDTPNHFEKFISQVFIKGFDRHMRSANLKFIKNPRN
QGLEQSEIEEMSFDIKVEPSFLKNKDDYIAFWIFCKMLDARHLSELNEMIKYDGHLTGE
QEIIIGLALLGVDSRENDWKQFFSSEREY EKIMKGYVVEELYQREPYRQSDGKTPILFRGV
EQARKYGTETVIQRLFDANPEFKVSKCNLAEWERQKETIETIKRRKELHNEWAKNPKKP
QNNAFFKEYKECCDAIDAYNWHKNKTTLAYVNELHLLIEILGRYVGYVAIADRDFQCMA
NQYFKHSGITERVEYWGDNRLKSIKKLDTFLKKEGLFVSEKNARNHIAHLNLYSLKSECT
LLYLSERLREIFKYDRKLKNAVSKSLIDILDRHGMSVVFANLKENKHRLVIKSLEPKKL
HLGGKKIDGGYIETNQVSEYCGIVKRLLM

图56I

>WP_055061018.1 直肠真杆菌基因组组件 T1815, 毗连序列群
T1815_16, 全基因组鸟枪序列
MLRRDKEVKKLYNVFNQIQVGTKPKKWNNDEKLSPEENERRAQQKNIMKKNYKWREACSK
YVESSQRIINDVIFYSYRKAKNKLRMRKNEDILKKMQEAEKLSKFSGGKLEDFVAYTLR
KSLVVS KYDTQEFDSLAA MVVFLECIGKNNISDHEREIVCKLLELIRKDFSKLDPNVKGS
QGANIVRSVRNQNMIVQPQGD RFLFPQVYAKENETVTNKNVEKEGLNEFLLNYANLDDEK
RAESLRKLRRILDVYFSAPNHYEKDMDITLSDNIEKEKFNWVEKHECGKKETGLFVDIPD
VLMEAEAEENIKLDAVVEKRERKVLNDRVRKQNIICYRYTRAVVEKYNSNEPLFFENNAIN
QYWIHHIENAVERILKNCKAGKLFKLKGYLA EK VWKDAINLISIKYIALGKAVYNFALD
DIWKDKKNKELGIVDERIRNGITSFDYEMIKAHENLQRELAVDIAFSVNNLARAVCDMSN
LGNKESDFLLWKRNDIADKLKNKDDMASVSAVLQFFGGKSSWDINIFKDAYKGKKKYNIE
VRFIDDLRKAIYCARNENFHFKTALVNDEKWNTELF GKIFERETEFCLNVEKDRFYSNNL
YMFYQVSELNMLDHLYSRVSRAAQVPSYNSVIVRTAFPEYITNVLG YQKPSYDADTLG
KWYSACYLLKEIYYNSFLQSDRALQLFEKSVKTL SWDDKKQ QRAVDNFKDHFSDIKSAC
TSLAQVCQIYMTEYNQONNQIKKVRSSNDSIFDQPVYQHYKVLLKKA IANAFADYLKNNK
DLFGFIGKPFKANEIREIDKEQFLPDWTSRKYEALCIEVSGSQELQKWYIVGKFLNARSL
NLMVGS MRSYIQYVTDIKRRAASIGNELHVSVDVEKVEKWVQVIEVCSLLASRTSNQFE
DYFNDKDDYARYLKS YVDFSNVDM PSEYSALVDFSNEEQSDLYVDPKNPKVNRNIVH SKL
FAADHILRDIVEPVSKDNIEEFYSQKAEIAYCKIKGKEITAE EQAVLKYQKLKNRVELR
DIVEYGEIINELLGQLINWSFMRE RDLLYFQLGFHYDCLRND SKKPEGYKNIKVDENS IK
DAILYQIIGMYVNGVT VYAPEKDGDKLKEQCVKGGVGKVSAFHRYSKYLGLNEKTL YNA
GLEIFEVVAEHEDIINLRNGIDHFKYYLGDYRSMLSIYSEVFD RFFTYDIKYQKNVNL LL
QNILLRHN VIVEPILES GFKTIGEQT KPGAKLSIRS IKS DTFQYKVKG GTLITDAKDER Y
LETIRKILYYAENEEDNLKKS VVVTNADKY EKNKESDDQNKQKEKKNKDNKGKKNEETKS
DAEKNNNERLSYNPFANLNFKLSN

图56J

>CRZ35554.1 Herbinix hemicellulosilytica 基因组组件 TUM3/55,
毗连序列群 02_T3/55T_contig26, 全基因组鸟枪序列
MKLTRRRISGNSVDQKITAIFYRDMSQGLLYDSEDNDCTDKVIESMDFERSWRGRILKN
GEDDKNPFYMFVKGLVGSNDKIVCEPIDVSDPDNLDILINKNLTGFGRNLKAPDSNDTL
ENLIRKIQAGIPEEEVLPELKKIKEMIQKDIVNRKEQLLKSIIKNNRIPFSLEGSKLVPST
KKMKWLFKLIDVPNKTFNEKMLEKYWEIYDYDKLKANITNRLDKTDKKARSISRASEEL
REYHKNLRTNYNRFVSGDRPAAGLDNGGSAKYNPDKEEFLLFLKEVEQYFKKYFPVKSKH
SNKSKDKSLVDKYKNYCSYKVVKKEVNRSIIINQLVAGLIQQGKLLYFYNDTWQEDFLN
SYGLSYIQVEEAFKKSVMTSLSWGINRLTSFFIDDSNTVKFDDITTKAKEAIESNYFNK
LRTC SRMQDHFEKELAFFYPVYVKDKDRPDDDIENLIVLVKNAIESVSYLRNRTFHFKE
SSLLELLKELDDKNSGQNKIDYSVAAEFIKRDIENLYDVFREQIRSLGIAEYYKADMISD
CFKTCGLEFALYSPKNSLMPAFKNVYKRGANLNKAYIRDKGPKETGDQGQNSYKALEEYR
ELTWYIEVKNNDDQSYNAYKNLLQLIYYHAFLPEVRENEALITDFINRTKEWNRKETEEERL
NTKNNKKHKNFEDNDITVNTYRYESIPDYQGESLDDYLKVLQRKQMARAKEVNEKEEGN
NNYIQFIRDVVVWAFGAYLENKLKNYKNELQPPLSKENIGLNDTLKELFPEEKVKSPFNI
KCRFSISTFIDNKGKSTDNTSAEAVKTDGKEDEKDKKNIKRKDLLCFYLFRLLDENEIC
KLQHQFIKYRCSLKERFPGNRTKLEKETELLAELEELMELVRFTMPISPEISAKAESGY
DTMIKKYFKDFIEKKVFKNPKTSNLYYHSDSKTPVTRKYMALLMRSAPLHLYKDIFKGY
LITKKECLEYIKLSNIIKDYQNSLNELEHEQLERIKLKSEKQNGKDSLYLDKKDFYKVKEY
VENLEQVARYKHLQHKINFESLYRIFRIHVDIAARMVGYTQDWERDMHFLFKALVYNGVL
EERRFEAIFNNNDNDNDGRIVKKIQNNLNNKNRELVSMLCWNKKLNKNEFGAIIWKRNP
AHLNHFTQTEQNSKSSLESLSLIRILLAYDRKRQNAVTKTINDLLNDYHIRIKWEGRV
DEGQIYFNIKEKEDIENEPIIHLKHLHKDCYIYKNSYMFQKQKEWICNGIKEEVYDKSI
LKCIGNLFKFDYEDKNKSSANPKHT

图56K

参考序列(1): 937566624_直肠真杆菌
通过比对长度归一化的身份
着色是根据: 身份+属性

1	937566624_直肠真杆菌	100.0%	1		80
2	912591537_Herbinix_hemicellulosilytica	10.6%			
3	502750493_斯氏李斯特菌	11.1%			
4	503209049_Paludibacter_propionigenes	11.3%			
5	506250229_口腔纤毛菌	11.6%			
6	517262777_沙氏纤毛菌	9.2%			
7	544240864_韦德氏毛菌	11.8%			
8	545620493_口腔类纤毛菌_879	8.8%			
9	545623306_韦德氏毛菌	12.3%			
10	545623740_韦德氏毛菌	11.0%			
11	551041827_毛螺科菌_NK4A179	32.0%			
12	564875111_荚膜红细菌_R121	7.2%			
13	652829192_毛螺科菌_NK4A144	36.5%			
14	671463495_嗜胺梭菌	35.3%			
15	736546968_鸡肉杆菌	9.7%			
16	736550717_鸡肉杆菌	10.3%			
17	738133341_李斯特新优克尼斯菌	10.9%			
18	769144435_毛螺科菌_MA2020	33.7%			
19	545661797_口腔纤毛菌_225_部分_	13.5%			
共有序列/100%					
共有序列/90%					
共有序列/80%					
共有序列/70%					
1	937566624_直肠真杆菌	100.0%	81		160
2	912591537_Herbinix_hemicellulosilytica	10.6%			
3	502750493_斯氏李斯特菌	11.1%			
4	503209049_Paludibacter_propionigenes	11.3%			
5	506250229_口腔纤毛菌	11.6%			
6	517262777_沙氏纤毛菌	9.2%			
7	544240864_韦德氏毛菌	11.8%			
8	545620493_口腔类纤毛菌_879	8.8%			
9	545623306_韦德氏毛菌	12.3%			
10	545623740_韦德氏毛菌	11.0%			
11	551041827_毛螺科菌_NK4A179	32.0%			
12	564875111_荚膜红细菌_R121	7.2%			
13	652829192_毛螺科菌_NK4A144	36.5%			
14	671463495_嗜胺梭菌	35.3%			
15	736546968_鸡肉杆菌	9.7%			
16	736550717_鸡肉杆菌	10.3%			
17	738133341_李斯特新优克尼斯菌	10.9%			
18	769144435_毛螺科菌_MA2020	33.7%			
19	545661797_口腔纤毛菌_225_部分_	13.5%			
共有序列/100%					
共有序列/90%					
共有序列/80%					
共有序列/70%					
1	937566624_直肠真杆菌	100.0%	161		240
2	912591537_Herbinix_hemicellulosilytica	10.6%			
3	502750493_斯氏李斯特菌	11.1%			
4	503209049_Paludibacter_propionigenes	11.3%			
5	506250229_口腔纤毛菌	11.6%			
6	517262777_沙氏纤毛菌	9.2%			
7	544240864_韦德氏毛菌	11.8%			
8	545620493_口腔类纤毛菌_879	8.8%			
9	545623306_韦德氏毛菌	12.3%			
10	545623740_韦德氏毛菌	11.0%			
11	551041827_毛螺科菌_NK4A179	32.0%			
12	564875111_荚膜红细菌_R121	7.2%			
13	652829192_毛螺科菌_NK4A144	36.5%			
14	671463495_嗜胺梭菌	35.3%			
15	736546968_鸡肉杆菌	9.7%			
16	736550717_鸡肉杆菌	10.3%			
17	738133341_李斯特新优克尼斯菌	10.9%			
18	769144435_毛螺科菌_MA2020	33.7%			
19	545661797_口腔纤毛菌_225_部分_	13.5%			
共有序列/100%					
共有序列/90%					
共有序列/80%					
共有序列/70%					
1	937566624_直肠真杆菌	100.0%	241		320
2	912591537_Herbinix_hemicellulosilytica	10.6%			
3	502750493_斯氏李斯特菌	11.1%			
4	503209049_Paludibacter_propionigenes	11.3%			
5	506250229_口腔纤毛菌	11.6%			
6	517262777_沙氏纤毛菌	9.2%			
7	544240864_韦德氏毛菌	11.8%			
8	545620493_口腔类纤毛菌_879	8.8%			
9	545623306_韦德氏毛菌	12.3%			
10	545623740_韦德氏毛菌	11.0%			
11	551041827_毛螺科菌_NK4A179	32.0%			
12	564875111_荚膜红细菌_R121	7.2%			
13	652829192_毛螺科菌_NK4A144	36.5%			

图57

14	671463495_嗜酸梭菌	35.3%	VSPHGVSESLVSE....ADPGVSESLVSLVSLGKLP....-GTGLRMFLASCPH..
15	736546968_鸡肉杆菌	9.7%	SLTDLRELPL....-KSL-LALFELKRFKMDIVEL....-KMLDLTFYFNA..
16	736550717_鸡肉杆菌	10.3%	VLPRLKSLPL....-EAG-LKRFKSTVSE....-EKYYEKKHVSPTKLLPLFENAV..
17	738133341_李斯特新优克尼菌	10.9%	-KLPRLKSLN....-EGL-LKLPKRFKMDIVEL....-MYSVPKSGM-VGLTDLVAPV..
18	769144435_毛螺杆菌_MA2020	33.7%	-FLTSQVQVSEHAKRFGGG....-RYVDSGLSEALTEFLKRFKREN....-VHSEFVLEHNAAGINSS
19	545651797_口腔纤毛菌_225_部分_	13.5%	
共有序列/100%			
共有序列/98%			
共有序列/80%			
共有序列/70%			
1	937566624_直肠真杆菌	100.0%	321.....4480
2	912591537_Herbinix_hemicellulosilytica	10.6%	
3	502750493_斯氏李斯特菌	11.1%	
4	503209049_Paludibacter_propioniciigenes	11.3%	
5	506250229_口腔纤毛菌	11.6%	
6	517262777_沙氏纤毛菌	9.2%	
7	544240864_韦德氏毛菌	11.8%	
8	545620493_口腔纤毛菌_879	8.8%	
9	545623306_韦德氏毛菌	12.3%	
10	545623740_韦德氏毛菌	11.0%	
11	551041827_毛螺杆菌_NK4A179	32.0%	
12	564875111_荚膜红细菌_R121	7.2%	
13	652829192_毛螺杆菌_NK4A144	36.5%	
14	671463495_嗜酸梭菌	35.3%	
15	736546968_鸡肉杆菌	9.7%	
16	736550717_鸡肉杆菌	10.3%	
17	738133341_李斯特新优克尼菌	10.9%	
18	769144435_毛螺杆菌_MA2020	33.7%	
19	545651797_口腔纤毛菌_225_部分_	13.5%	
共有序列/100%			
共有序列/98%			
共有序列/80%			
共有序列/70%			
1	937566624_直肠真杆菌	100.0%	481.....480
2	912591537_Herbinix_hemicellulosilytica	10.6%	
3	502750493_斯氏李斯特菌	11.1%	
4	503209049_Paludibacter_propioniciigenes	11.3%	
5	506250229_口腔纤毛菌	11.6%	
6	517262777_沙氏纤毛菌	9.2%	
7	544240864_韦德氏毛菌	11.8%	
8	545620493_口腔纤毛菌_879	8.8%	
9	545623306_韦德氏毛菌	12.3%	
10	545623740_韦德氏毛菌	11.0%	
11	551041827_毛螺杆菌_NK4A179	32.0%	
12	564875111_荚膜红细菌_R121	7.2%	
13	652829192_毛螺杆菌_NK4A144	36.5%	
14	671463495_嗜酸梭菌	35.3%	
15	736546968_鸡肉杆菌	9.7%	
16	736550717_鸡肉杆菌	10.3%	
17	738133341_李斯特新优克尼菌	10.9%	
18	769144435_毛螺杆菌_MA2020	33.7%	
19	545651797_口腔纤毛菌_225_部分_	13.5%	
共有序列/100%			
共有序列/98%			
共有序列/80%			
共有序列/70%			
1	937566624_直肠真杆菌	100.0%	481.....560
2	912591537_Herbinix_hemicellulosilytica	10.6%	
3	502750493_斯氏李斯特菌	11.1%	
4	503209049_Paludibacter_propioniciigenes	11.3%	
5	506250229_口腔纤毛菌	11.6%	
6	517262777_沙氏纤毛菌	9.2%	
7	544240864_韦德氏毛菌	11.8%	
8	545620493_口腔纤毛菌_879	8.8%	
9	545623306_韦德氏毛菌	12.3%	
10	545623740_韦德氏毛菌	11.0%	
11	551041827_毛螺杆菌_NK4A179	32.0%	
12	564875111_荚膜红细菌_R121	7.2%	
13	652829192_毛螺杆菌_NK4A144	36.5%	
14	671463495_嗜酸梭菌	35.3%	
15	736546968_鸡肉杆菌	9.7%	
16	736550717_鸡肉杆菌	10.3%	
17	738133341_李斯特新优克尼菌	10.9%	
18	769144435_毛螺杆菌_MA2020	33.7%	
19	545651797_口腔纤毛菌_225_部分_	13.5%	
共有序列/100%			
共有序列/98%			
共有序列/80%			
共有序列/70%			
1	937566624_直肠真杆菌	100.0%	561.....648
2	912591537_Herbinix_hemicellulosilytica	10.6%	
3	502750493_斯氏李斯特菌	11.1%	
4	503209049_Paludibacter_propioniciigenes	11.3%	
5	506250229_口腔纤毛菌	11.6%	
6	517262777_沙氏纤毛菌	9.2%	
7	544240864_韦德氏毛菌	11.8%	

图57(续1)

8	545620493_口腔类杆菌_879	8.8%	641	937566624_直肠真杆菌	100.0%
9	545623306_韦德氏菌	12.3%	721	912591537_Herbinix_hemicellulosilytica	10.6%
10	545623740_韦德氏菌	11.0%	801	502750493_斯氏李斯特菌	11.1%
11	551041827_毛螺杆菌_NK4A179	32.0%	881	503209049_Paludibacter_propionicigenes	11.3%
12	564875111_荚膜红细菌_R121	7.2%	961	506250229_口腔杆菌	11.6%
13	652829192_毛螺杆菌_NK4A144	36.5%	1041	517262777_沙氏氏菌	9.2%
14	671463495_嗜酸梭菌	35.3%	1121	544240864_韦德氏菌	11.8%
15	736546958_鸡肉杆菌	9.7%	1201	545620493_口腔类杆菌_879	8.8%
16	736550717_鸡肉杆菌	10.3%	1281	545623306_韦德氏菌	12.3%
17	738133341_李斯特新优克尼菌	10.9%	1361	545623740_韦德氏菌	11.0%
18	769144435_毛螺杆菌_MA2020	33.7%	1441	551041827_毛螺杆菌_NK4A179	32.0%
19	545661797_口腔杆菌_225_部分_	13.5%	1521	564875111_荚膜红细菌_R121	7.2%
共有序列/100%			1601	652829192_毛螺杆菌_NK4A144	36.5%
共有序列/90%			1681	671463495_嗜酸梭菌	35.3%
共有序列/80%			1761	736546958_鸡肉杆菌	9.7%
共有序列/70%			1841	736550717_鸡肉杆菌	10.3%
			1921	738133341_李斯特新优克尼菌	10.9%
			2001	769144435_毛螺杆菌_MA2020	33.7%
			2081	545661797_口腔杆菌_225_部分_	13.5%
			2161	共有序列/100%	
			2241	共有序列/90%	
			2321	共有序列/80%	
			2401	共有序列/70%	
			2481		
			2561		
			2641		
			2721		
			2801		
			2881		
			2961		
			3041		
			3121		
			3201		
			3281		
			3361		
			3441		
			3521		
			3601		
			3681		
			3761		
			3841		
			3921		
			4001		
			4081		
			4161		
			4241		
			4321		
			4401		
			4481		
			4561		
			4641		
			4721		
			4801		
			4881		
			4961		
			5041		
			5121		
			5201		
			5281		
			5361		
			5441		
			5521		
			5601		
			5681		
			5761		
			5841		
			5921		
			6001		
			6081		
			6161		
			6241		
			6321		
			6401		
			6481		
			6561		
			6641		
			6721		
			6801		
			6881		
			6961		
			7041		
			7121		
			7201		
			7281		
			7361		
			7441		
			7521		
			7601		
			7681		
			7761		
			7841		
			7921		
			8001		
			8081		
			8161		
			8241		
			8321		
			8401		
			8481		
			8561		
			8641		
			8721		
			8801		
			8881		
			8961		
			9041		
			9121		
			9201		
			9281		
			9361		
			9441		
			9521		
			9601		
			9681		
			9761		
			9841		
			9921		
			10001		

图57(续2)

1 937566624_直肠真杆菌	100.0%	1 100.0%	1 100.0%
2 912591537_Herbindex_hemicellulosilytica	10.6%	10.6%	10.6%
3 502750493_斯氏李斯特菌	11.1%	11.1%	11.1%
4 503209049_Paludibacter_propionigenes	11.3%	11.3%	11.3%
5 506250229_口腔纤毛菌	11.6%	11.6%	11.6%
6 517262777_沙氏纤毛菌	9.2%	9.2%	9.2%
7 544240864_韦德氏菌	11.8%	11.8%	11.8%
8 545620493_口腔类杆菌_879	8.8%	8.8%	8.8%
9 545623306_韦德氏菌	12.3%	12.3%	12.3%
10 545623740_韦德氏菌	11.0%	11.0%	11.0%
11 551041827_毛螺杆菌_NK4A179	32.0%	32.0%	32.0%
12 564875111_类膜红细菌_R121	7.2%	7.2%	7.2%
13 652829192_毛螺杆菌_NK4A144	35.5%	35.5%	35.5%
14 671463495_嗜酸梭菌	35.3%	35.3%	35.3%
15 736546968_鸡肉杆菌	9.7%	9.7%	9.7%
16 736550717_鸡肉杆菌	10.3%	10.3%	10.3%
17 738133341_李斯特新优克尼菌	10.9%	10.9%	10.9%
18 769144435_毛螺杆菌_MA2020	33.7%	33.7%	33.7%
19 545661797_口腔纤毛菌_225_部分_	13.5%	13.5%	13.5%
共有序列/100%			
共有序列/90%			
共有序列/80%			
共有序列/70%			
1 937566624_直肠真杆菌	100.0%	100.0%	100.0%
2 912591537_Herbindex_hemicellulosilytica	10.6%	10.6%	10.6%
3 502750493_斯氏李斯特菌	11.1%	11.1%	11.1%
4 503209049_Paludibacter_propionigenes	11.3%	11.3%	11.3%
5 506250229_口腔纤毛菌	11.6%	11.6%	11.6%
6 517262777_沙氏纤毛菌	9.2%	9.2%	9.2%
7 544240864_韦德氏菌	11.8%	11.8%	11.8%
8 545620493_口腔类杆菌_879	8.8%	8.8%	8.8%
9 545623306_韦德氏菌	12.3%	12.3%	12.3%
10 545623740_韦德氏菌	11.0%	11.0%	11.0%
11 551041827_毛螺杆菌_NK4A179	32.0%	32.0%	32.0%
12 564875111_类膜红细菌_R121	7.2%	7.2%	7.2%
13 652829192_毛螺杆菌_NK4A144	35.5%	35.5%	35.5%
14 671463495_嗜酸梭菌	35.3%	35.3%	35.3%
15 736546968_鸡肉杆菌	9.7%	9.7%	9.7%
16 736550717_鸡肉杆菌	10.3%	10.3%	10.3%
17 738133341_李斯特新优克尼菌	10.9%	10.9%	10.9%
18 769144435_毛螺杆菌_MA2020	33.7%	33.7%	33.7%
19 545661797_口腔纤毛菌_225_部分_	13.5%	13.5%	13.5%
共有序列/100%			
共有序列/90%			
共有序列/80%			
共有序列/70%			
1 937566624_直肠真杆菌	100.0%	100.0%	100.0%
2 912591537_Herbindex_hemicellulosilytica	10.6%	10.6%	10.6%
3 502750493_斯氏李斯特菌	11.1%	11.1%	11.1%
4 503209049_Paludibacter_propionigenes	11.3%	11.3%	11.3%
5 506250229_口腔纤毛菌	11.6%	11.6%	11.6%
6 517262777_沙氏纤毛菌	9.2%	9.2%	9.2%
7 544240864_韦德氏菌	11.8%	11.8%	11.8%
8 545620493_口腔类杆菌_879	8.8%	8.8%	8.8%
9 545623306_韦德氏菌	12.3%	12.3%	12.3%
10 545623740_韦德氏菌	11.0%	11.0%	11.0%
11 551041827_毛螺杆菌_NK4A179	32.0%	32.0%	32.0%
12 564875111_类膜红细菌_R121	7.2%	7.2%	7.2%
13 652829192_毛螺杆菌_NK4A144	35.5%	35.5%	35.5%
14 671463495_嗜酸梭菌	35.3%	35.3%	35.3%
15 736546968_鸡肉杆菌	9.7%	9.7%	9.7%
16 736550717_鸡肉杆菌	10.3%	10.3%	10.3%
17 738133341_李斯特新优克尼菌	10.9%	10.9%	10.9%
18 769144435_毛螺杆菌_MA2020	33.7%	33.7%	33.7%
19 545661797_口腔纤毛菌_225_部分_	13.5%	13.5%	13.5%
共有序列/100%			
共有序列/90%			
共有序列/80%			
共有序列/70%			

图57(续3)

共有序列/100%					
共有序列/90%					
共有序列/80%					
共有序列/70%					
1	937566624_直肠真杆菌	100.0%	1281	1280	
2	912591537_Herbinix_hemicellulosilytica	10.6%	1281	1280	
3	502750493_斯氏李斯特菌	11.1%	1281	1280	
4	503209049_Paludibacter_propionigenes	11.3%	1281	1280	
5	506250229_口腔纤毛菌	11.6%	1281	1280	
6	517262777_沙氏纤毛菌	9.2%	1281	1280	
7	544240864_韦德氏毛菌	11.8%	1281	1280	
8	545620493_口腔类纤毛菌_879	8.8%	1281	1280	
9	545623306_韦德氏毛菌	12.3%	1281	1280	
10	545623740_韦德氏毛菌	11.0%	1281	1280	
11	551041827_毛螺科菌_NK4A179	32.0%	1281	1280	
12	564875111_荚膜红细菌_R121	7.2%	1281	1280	
13	652829192_毛螺科菌_NK4A144	36.5%	1281	1280	
14	671463495_嗜酸梭菌	35.3%	1281	1280	
15	736546968_鸡肉杆菌	9.7%	1281	1280	
16	7365509717_鸡肉杆菌	10.3%	1281	1280	
17	738133341_李斯特新优克尼斯菌	10.9%	1281	1280	
18	769144435_毛螺科菌_MA2020	33.7%	1281	1280	
19	545661797_口腔纤毛菌_225_部分_	13.5%	1281	1280	
共有序列/100%					
共有序列/90%					
共有序列/80%					
共有序列/70%					
1	937566624_直肠真杆菌	100.0%	1281	1360	
2	912591537_Herbinix_hemicellulosilytica	10.6%	1281	1360	
3	502750493_斯氏李斯特菌	11.1%	1281	1360	
4	503209049_Paludibacter_propionigenes	11.3%	1281	1360	
5	506250229_口腔纤毛菌	11.6%	1281	1360	
6	517262777_沙氏纤毛菌	9.2%	1281	1360	
7	544240864_韦德氏毛菌	11.8%	1281	1360	
8	545620493_口腔类纤毛菌_879	8.8%	1281	1360	
9	545623306_韦德氏毛菌	12.3%	1281	1360	
10	545623740_韦德氏毛菌	11.0%	1281	1360	
11	551041827_毛螺科菌_NK4A179	32.0%	1281	1360	
12	564875111_荚膜红细菌_R121	7.2%	1281	1360	
13	652829192_毛螺科菌_NK4A144	36.5%	1281	1360	
14	671463495_嗜酸梭菌	35.3%	1281	1360	
15	736546968_鸡肉杆菌	9.7%	1281	1360	
16	7365509717_鸡肉杆菌	10.3%	1281	1360	
17	738133341_李斯特新优克尼斯菌	10.9%	1281	1360	
18	769144435_毛螺科菌_MA2020	33.7%	1281	1360	
19	545661797_口腔纤毛菌_225_部分_	13.5%	1281	1360	
共有序列/100%					
共有序列/90%					
共有序列/80%					
共有序列/70%					
1	937566624_直肠真杆菌	100.0%	1361	1440	
2	912591537_Herbinix_hemicellulosilytica	10.6%	1361	1440	
3	502750493_斯氏李斯特菌	11.1%	1361	1440	
4	503209049_Paludibacter_propionigenes	11.3%	1361	1440	
5	506250229_口腔纤毛菌	11.6%	1361	1440	
6	517262777_沙氏纤毛菌	9.2%	1361	1440	
7	544240864_韦德氏毛菌	11.8%	1361	1440	
8	545620493_口腔类纤毛菌_879	8.8%	1361	1440	
9	545623306_韦德氏毛菌	12.3%	1361	1440	
10	545623740_韦德氏毛菌	11.0%	1361	1440	
11	551041827_毛螺科菌_NK4A179	32.0%	1361	1440	
12	564875111_荚膜红细菌_R121	7.2%	1361	1440	
13	652829192_毛螺科菌_NK4A144	36.5%	1361	1440	
14	671463495_嗜酸梭菌	35.3%	1361	1440	
15	736546968_鸡肉杆菌	9.7%	1361	1440	
16	7365509717_鸡肉杆菌	10.3%	1361	1440	
17	738133341_李斯特新优克尼斯菌	10.9%	1361	1440	
18	769144435_毛螺科菌_MA2020	33.7%	1361	1440	
19	545661797_口腔纤毛菌_225_部分_	13.5%	1361	1440	
共有序列/100%					
共有序列/90%					
共有序列/80%					
共有序列/70%					
1	937566624_直肠真杆菌	100.0%	1441	1520	
2	912591537_Herbinix_hemicellulosilytica	10.6%	1441	1520	
3	502750493_斯氏李斯特菌	11.1%	1441	1520	
4	503209049_Paludibacter_propionigenes	11.3%	1441	1520	
5	506250229_口腔纤毛菌	11.6%	1441	1520	
6	517262777_沙氏纤毛菌	9.2%	1441	1520	
7	544240864_韦德氏毛菌	11.8%	1441	1520	
8	545620493_口腔类纤毛菌_879	8.8%	1441	1520	
9	545623306_韦德氏毛菌	12.3%	1441	1520	
10	545623740_韦德氏毛菌	11.0%	1441	1520	
11	551041827_毛螺科菌_NK4A179	32.0%	1441	1520	
12	564875111_荚膜红细菌_R121	7.2%	1441	1520	
13	652829192_毛螺科菌_NK4A144	36.5%	1441	1520	
14	671463495_嗜酸梭菌	35.3%	1441	1520	

图57(续4)

图57 (续5)

8	545620493_口腔类杆菌_879	8.8%	8	545620493_口腔类杆菌_879	8.8%
9	545623306_韦德氏杆菌	12.3%	9	545623306_韦德氏杆菌	12.3%
10	545623740_韦德氏杆菌	11.0%	10	545623740_韦德氏杆菌	11.0%
11	551041827_毛螺杆菌_NK4A179	32.0%	11	551041827_毛螺杆菌_NK4A179	32.0%
12	564875111_类膜红细菌_R121	7.2%	12	564875111_类膜红细菌_R121	7.2%
13	652829192_毛螺杆菌_NK4A144	36.5%	13	652829192_毛螺杆菌_NK4A144	36.5%
14	671463495_嗜酸梭菌	35.3%	14	671463495_嗜酸梭菌	35.3%
15	736546968_鸡肉杆菌	9.7%	15	736546968_鸡肉杆菌	9.7%
16	736550717_鸡肉杆菌	10.3%	16	736550717_鸡肉杆菌	10.3%
17	738133341_李斯特新优克尼斯菌	10.9%	17	738133341_李斯特新优克尼斯菌	10.9%
18	769144435_毛螺杆菌_MA2020	33.7%	18	769144435_毛螺杆菌_MA2020	33.7%
19	545661797_口腔链球菌_225_部分_	13.5%	19	545661797_口腔链球菌_225_部分_	13.5%
共有序列/100%			共有序列/100%		
共有序列/90%			共有序列/90%		
共有序列/80%			共有序列/80%		
共有序列/70%			共有序列/70%		
1	937566624_直肠真杆菌	100.0%	1	937566624_直肠真杆菌	100.0%
2	912591537_Herbinix_hemicellulosilytica	10.6%	2	912591537_Herbinix_hemicellulosilytica	10.6%
3	502750493_斯氏李斯特菌	11.1%	3	502750493_斯氏李斯特菌	11.1%
4	503209049_Paludibacter_propionicigenes	11.3%	4	503209049_Paludibacter_propionicigenes	11.3%
5	506250229_口腔链球菌	11.6%	5	506250229_口腔链球菌	11.6%
6	517262777_沙氏杆菌	9.2%	6	517262777_沙氏杆菌	9.2%
7	544240864_韦德氏杆菌	11.8%	7	544240864_韦德氏杆菌	11.8%
8	545620493_口腔类杆菌_879	8.8%	8	545620493_口腔类杆菌_879	8.8%
9	545623306_韦德氏杆菌	12.3%	9	545623306_韦德氏杆菌	12.3%
10	545623740_韦德氏杆菌	11.0%	10	545623740_韦德氏杆菌	11.0%
11	551041827_毛螺杆菌_NK4A179	32.0%	11	551041827_毛螺杆菌_NK4A179	32.0%
12	564875111_类膜红细菌_R121	7.2%	12	564875111_类膜红细菌_R121	7.2%
13	652829192_毛螺杆菌_NK4A144	36.5%	13	652829192_毛螺杆菌_NK4A144	36.5%
14	671463495_嗜酸梭菌	35.3%	14	671463495_嗜酸梭菌	35.3%
15	736546968_鸡肉杆菌	9.7%	15	736546968_鸡肉杆菌	9.7%
16	736550717_鸡肉杆菌	10.3%	16	736550717_鸡肉杆菌	10.3%
17	738133341_李斯特新优克尼斯菌	10.9%	17	738133341_李斯特新优克尼斯菌	10.9%
18	769144435_毛螺杆菌_MA2020	33.7%	18	769144435_毛螺杆菌_MA2020	33.7%
19	545661797_口腔链球菌_225_部分_	13.5%	19	545661797_口腔链球菌_225_部分_	13.5%
共有序列/100%			共有序列/100%		
共有序列/90%			共有序列/90%		
共有序列/80%			共有序列/80%		
共有序列/70%			共有序列/70%		
1	937566624_直肠真杆菌	100.0%	1	937566624_直肠真杆菌	100.0%
2	912591537_Herbinix_hemicellulosilytica	10.6%	2	912591537_Herbinix_hemicellulosilytica	10.6%
3	502750493_斯氏李斯特菌	11.1%	3	502750493_斯氏李斯特菌	11.1%
4	503209049_Paludibacter_propionicigenes	11.3%	4	503209049_Paludibacter_propionicigenes	11.3%
5	506250229_口腔链球菌	11.6%	5	506250229_口腔链球菌	11.6%
6	517262777_沙氏杆菌	9.2%	6	517262777_沙氏杆菌	9.2%
7	544240864_韦德氏杆菌	11.8%	7	544240864_韦德氏杆菌	11.8%
8	545620493_口腔类杆菌_879	8.8%	8	545620493_口腔类杆菌_879	8.8%
9	545623306_韦德氏杆菌	12.3%	9	545623306_韦德氏杆菌	12.3%
10	545623740_韦德氏杆菌	11.0%	10	545623740_韦德氏杆菌	11.0%
11	551041827_毛螺杆菌_NK4A179	32.0%	11	551041827_毛螺杆菌_NK4A179	32.0%
12	564875111_类膜红细菌_R121	7.2%	12	564875111_类膜红细菌_R121	7.2%
13	652829192_毛螺杆菌_NK4A144	36.5%	13	652829192_毛螺杆菌_NK4A144	36.5%
14	671463495_嗜酸梭菌	35.3%	14	671463495_嗜酸梭菌	35.3%
15	736546968_鸡肉杆菌	9.7%	15	736546968_鸡肉杆菌	9.7%
16	736550717_鸡肉杆菌	10.3%	16	736550717_鸡肉杆菌	10.3%
17	738133341_李斯特新优克尼斯菌	10.9%	17	738133341_李斯特新优克尼斯菌	10.9%
18	769144435_毛螺杆菌_MA2020	33.7%	18	769144435_毛螺杆菌_MA2020	33.7%
19	545661797_口腔链球菌_225_部分_	13.5%	19	545661797_口腔链球菌_225_部分_	13.5%
共有序列/100%			共有序列/100%		
共有序列/90%			共有序列/90%		
共有序列/80%			共有序列/80%		
共有序列/70%			共有序列/70%		
1	937566624_直肠真杆菌	100.0%	1	937566624_直肠真杆菌	100.0%
2	912591537_Herbinix_hemicellulosilytica	10.6%	2	912591537_Herbinix_hemicellulosilytica	10.6%
3	502750493_斯氏李斯特菌	11.1%	3	502750493_斯氏李斯特菌	11.1%
4	503209049_Paludibacter_propionicigenes	11.3%	4	503209049_Paludibacter_propionicigenes	11.3%
5	506250229_口腔链球菌	11.6%	5	506250229_口腔链球菌	11.6%
6	517262777_沙氏杆菌	9.2%	6	517262777_沙氏杆菌	9.2%
7	544240864_韦德氏杆菌	11.8%	7	544240864_韦德氏杆菌	11.8%
8	545620493_口腔类杆菌_879	8.8%	8	545620493_口腔类杆菌_879	8.8%
9	545623306_韦德氏杆菌	12.3%	9	545623306_韦德氏杆菌	12.3%
10	545623740_韦德氏杆菌	11.0%	10	545623740_韦德氏杆菌	11.0%
11	551041827_毛螺杆菌_NK4A179	32.0%	11	551041827_毛螺杆菌_NK4A179	32.0%
12	564875111_类膜红细菌_R121	7.2%	12	564875111_类膜红细菌_R121	7.2%
13	652829192_毛螺杆菌_NK4A144	36.5%	13	652829192_毛螺杆菌_NK4A144	36.5%
14	671463495_嗜酸梭菌	35.3%	14	671463495_嗜酸梭菌	35.3%
15	736546968_鸡肉杆菌	9.7%	15	736546968_鸡肉杆菌	9.7%
16	736550717_鸡肉杆菌	10.3%	16	736550717_鸡肉杆菌	10.3%
17	738133341_李斯特新优克尼斯菌	10.9%	17	738133341_李斯特新优克尼斯菌	10.9%
18	769144435_毛螺杆菌_MA2020	33.7%	18	769144435_毛螺杆菌_MA2020	33.7%
19	545661797_口腔链球菌_225_部分_	13.5%	19	545661797_口腔链球菌_225_部分_	13.5%
共有序列/100%			共有序列/100%		
共有序列/90%			共有序列/90%		
共有序列/80%			共有序列/80%		
共有序列/70%			共有序列/70%		

图57(续6)

2 512591537_Herbinix_hemicellulosilytica	10.6%	
3 502750493_斯氏李斯特菌	11.1%	
4 503209049_Paludibacter_propionigenes	11.3%	
5 506250229_口腔纤毛菌	11.6%	
6 517262777_沙氏纤毛菌	9.2%	
7 544240864_韦德纤毛菌	11.8%	
8 545620493_口腔类纤毛菌_879	8.8%	
9 545623306_韦德纤毛菌	12.3%	
10 545623740_韦德纤毛菌	11.0%	
11 551041827_毛螺科菌_NK4A179	32.0%	
12 564875111_荚膜红细菌_R121	7.2%	
13 652829192_毛螺科菌_NK4A144	36.5%	
14 671463495_嗜酸梭菌	35.3%	
15 736546968_鸡肉杆菌	9.7%	
16 736550717_鸡肉杆菌	10.3%	
17 738133341_李斯特新优克尼斯菌	10.9%	
18 769144435_毛螺科菌_MA2020	33.7%	
19 545661797_口腔纤毛菌_225_部分_	13.5%	
共有序列/100%		
共有序列/90%		
共有序列/80%		
共有序列/70%		
1 937566624_直肠真杆菌	100.0%	
2 912591537_Herbinix_hemicellulosilytica	10.6%	
3 502750493_斯氏李斯特菌	11.1%	
4 503209049_Paludibacter_propionigenes	11.3%	
5 506250229_口腔纤毛菌	11.6%	
6 517262777_沙氏纤毛菌	9.2%	
7 544240864_韦德纤毛菌	11.8%	
8 545620493_口腔类纤毛菌_879	8.8%	
9 545623306_韦德纤毛菌	12.3%	
10 545623740_韦德纤毛菌	11.0%	
11 551041827_毛螺科菌_NK4A179	32.0%	
12 564875111_荚膜红细菌_R121	7.2%	
13 652829192_毛螺科菌_NK4A144	36.5%	
14 671463495_嗜酸梭菌	35.3%	
15 736546968_鸡肉杆菌	9.7%	
16 736550717_鸡肉杆菌	10.3%	
17 738133341_李斯特新优克尼斯菌	10.9%	
18 769144435_毛螺科菌_MA2020	33.7%	
19 545661797_口腔纤毛菌_225_部分_	13.5%	
共有序列/100%		
共有序列/90%		
共有序列/80%		
共有序列/70%		

图57 (续7)