(19)中华人民共和国国家知识产权局

(12)发明专利申请

(10)申请公布号 CN 110914310 A
(43)申请公布日 2020.03.24

(21)申请号 201880030290.0

(22)申请日 2018.03.09

(30)优先权数据
62/470,175 2017.03.10 US

(85)PCT国际申请进入国家阶段日
2019.11.07

(86)PCT国际申请的申请数据
PCT/US2018/021878 2018.03.09

(87)PCT国际申请的公布数据
WO2018/165629 EN 2018.09.13

(71)申请人 哈佛大学的校长及成员们
地址 美国马萨诸塞州

(72)发明人 D.R.刘 L.W.科布兰

(74)专利代理机构 北京市柳沈律师事务所 11105
代理人 余晓文 张晓飞

(51)Int.Cl.
*C07K 19/00*(2006.01)
*C12N 9/24*(2006.01)
*C12N 9/22*(2006.01)
*C12N 15/62*(2006.01)

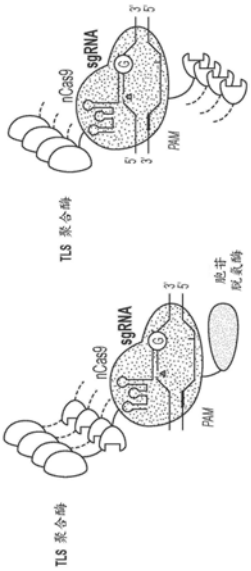
权利要求书11页 说明书97页
序列表272页 附图55页

(54)发明名称

胞嘧啶至鸟嘌呤碱基编辑器

(57)摘要

本公开的一些方面提供了可用于核酸的靶向编辑，包括编辑细胞或受试者的基因组内，例如人类基因组内的单个位点的组合物、策略、系统、试剂、方法和试剂盒。在一些实施方案中，提供了能够在核酸(例如，基因组DNA)中诱导胞嘧啶(C)改变为鸟嘌呤(G)的融合蛋白。在一些实施方案中，提供了核酸可编程DNA结合蛋白(例如，Cas9)，和核酸编辑蛋白或蛋白质域，例如脱氨酶域，聚合酶域和/或碱基切除酶的融合蛋白。在一些实施方案中，提供了用于靶向核酸编辑的方法。在一些实施方案中，提供了用于产生靶向核酸编辑蛋白，例如核酸可编程DNA结合蛋白(例如，Cas9)，和核酸编辑蛋白或域的融合蛋白的试剂和试剂盒。




CN 110914310 A

1.融合蛋白,其包含(i)核酸可编程DNA结合蛋白(napDNAbp),和(ii)胞苷脱氨酶域,以及(iii)尿嘧啶结合蛋白(UBP)。

2.权利要求1的融合蛋白,其中所述尿嘧啶结合蛋白是尿嘧啶修饰酶。

3.权利要求1或2的融合蛋白,其中所述尿嘧啶结合蛋白是尿嘧啶碱基切除酶(base excision enzyme)。

4.权利要求1-3中任一项的融合蛋白,其中所述尿嘧啶结合蛋白是尿嘧啶DNA糖基化酶(UDG)。

5.权利要求4的融合蛋白,其中所述尿嘧啶DNA糖基化酶(UDG)来自人、小鼠、大鼠、犬、猴、牛、猕猴、黑猩猩、大猩猩、海八目鳗(sea lamprey),或其具有尿嘧啶DNA糖基化酶(UDG)活性的任何变体。

6.权利要求1-5中任一项的融合蛋白,其中所述尿嘧啶结合蛋白是UdgX,其包含与SEQ ID NO:49(UdgX)的氨基酸序列至少80%、85%、90%、95%、98%或99%相同的氨基酸序列。

7.权利要求6的融合蛋白,其中所述尿嘧啶结合蛋白包含SEQ ID NO:49(UdgX)的氨基酸序列。

8.权利要求1-5中任一项的融合蛋白,其中所述尿嘧啶结合蛋白是UDG,其包含与SEQ ID NO:48(UDG)的氨基酸序列至少80%、85%、90%、95%、98%或99%相同的氨基酸序列。

9.权利要求8的融合蛋白,其中所述尿嘧啶结合蛋白包含SEQ ID NO:48(UDG)的氨基酸序列。

10.权利要求1-5中任一项的融合蛋白,其中所述尿嘧啶结合蛋白是UdgX*,其包含与SEQ ID NO:50(UdgX*)的氨基酸序列至少80%、85%、90%、95%、98%或99%相同的氨基酸序列。

11.权利要求10的融合蛋白,其中所述尿嘧啶结合蛋白包含SEQ ID NO:50(UdgX*)的氨基酸序列。

12.权利要求1-5中任一项的融合蛋白,其中所述尿嘧啶结合蛋白是UdgX_On,其包含与SEQ ID NO:51(UdgX_On)的氨基酸序列至少80%、85%、90%、95%、98%或99%相同的氨基酸序列。

13.权利要求12的融合蛋白,其中所述尿嘧啶结合蛋白包含SEQ ID NO:51(UdgX_On)的氨基酸序列。

14.权利要求1-5中任一项的融合蛋白,其中所述尿嘧啶结合蛋白是SMUG1,其包含与SEQ ID NO:53(SMUG1)的氨基酸序列至少80%、85%、90%、95%、98%或99%相同的氨基酸序列。

15.权利要求14的融合蛋白,其中尿嘧啶结合蛋白包含SEQ ID NO:53(SMUG1)的氨基酸序列。

16.权利要求1-15中任一项的融合蛋白,其中所述融合蛋白包含以下结构:

[胞苷脱氨酶]-[napDNAbp]-[UBP],其中“]-[”的每个情况包含任选的接头。

17.权利要求1-16中任一项的融合蛋白,其中所述胞苷脱氨酶和所述napDNAbp是经由接头融合的。

18.权利要求17的融合蛋白,其中所述胞苷脱氨酶和所述napDNAbp是经由包含SEQ ID NO:102-109中任一项的氨基酸序列的接头融合的。

19.权利要求1至18中任一项的融合蛋白,其中所述napDNAbp和所述UBP是经由接头融合的。

20.权利要求19的融合蛋白,其中所述napDNAbp和所述UBP是经由包含SEQ ID NO:102-109中任一项的氨基酸序列的接头融合的。

21.权利要求1-20中任一项的融合蛋白,其中所述融合蛋白进一步包含(iv)核酸聚合酶域(NAP)。

22.权利要求21的融合蛋白,其中所述核酸聚合酶域是真核核酸聚合酶域。

23.权利要求21或22的融合蛋白,其中所述核酸聚合酶域是DNA聚合酶域。

24.权利要求21至23中任一项的融合蛋白,其中所述核酸聚合酶域具有跨损伤聚合酶活性。

25.权利要求21至24中任一项的融合蛋白,其中所述核酸聚合酶域是跨损伤DNA聚合酶。

26.权利要求21-25中任一项的融合蛋白,其中所述核酸聚合酶域来自Rev7、Rev1复合物、聚合酶iota、聚合酶kappa和聚合酶eta。

27.权利要求21至25中任一项的融合蛋白,其中所述核酸聚合酶域选自真核聚合酶的组,该组由alpha、beta、gamma、delta、epsilon、gamma、eta、iota、kappa、lambda、mu、和nu组成。

28.权利要求21至27中任一项的融合蛋白,其中所述核酸聚合酶域包含与SEQ ID NO:54-64中任一项的氨基酸序列至少80%、85%、90%、95%、98%或99%相同的氨基酸序列。

29.权利要求21-28中任一项的融合蛋白,其中所述核酸聚合酶域包含SEQ ID NO:54-64中任一项的氨基酸序列。

30.权利要求21-29中任一项的融合蛋白,其中所述融合蛋白包含以下结构:

[胞苷脱氨酶]-[napDNAbp]-[UBP]-[NAP];

[胞苷脱氨酶]-[napDNAbp]-[NAP]-[UBP];

[胞苷脱氨酶]-[NAP]-[napDNAbp]-[UBP];或

[NAP]-[胞苷脱氨酶]-[napDNAbp]-[UBP];

其中"]-["的每个情况包括任选的接头。

31.权利要求21至30中任一项的融合蛋白,其中所述NAP和所述UBP是经由接头融合的。

32.权利要求31的融合蛋白,其中所述NAP和所述UBP是经由包含SEQ ID NO:102-109中任一项的氨基酸序列的接头融合的。

33.权利要求21-32中任一项的融合蛋白,其中所述NAP和所述napDNAbp是经由接头融合的。

34.权利要求33的融合蛋白,其中所述NAP和所述napDNAbp是经由包含SEQ ID NO:102-109中任一项的氨基酸序列的接头融合的。

35.权利要求21-34中任一项的融合蛋白,其中所述NAP和所述胞苷脱氨酶是经由接头融合的。

36.权利要求35的融合蛋白,其中所述NAP和胞苷脱氨酶是经由包含SEQ ID NO:102-109中任一项的氨基酸序列的接头融合的。

37.权利要求1-36中任一项的融合蛋白,其中所述napDNAbp包含Cas9域。

38.权利要求37的融合蛋白,其中所述Cas9域包含与SEQ ID NO:4-26中任一项至少85%相同的氨基酸序列。

39.权利要求37或38的融合蛋白,其中所述Cas9域包含SEQ ID NO:4-26中任一项的氨基酸序列。

40.权利要求37的融合蛋白,其中所述Cas9域是Cas9切口酶(nCas9)。

41.权利要求40的融合蛋白,其中所述Cas9切口酶(nCas9)包含与SEQ ID NO:10、13、16或21中任一项至少85%相同的氨基酸序列。

42.权利要求40或41的融合蛋白,其中所述nCas9包含SEQ ID NO:10、13、16或21中任一项的氨基酸序列。

43.权利要求37的融合蛋白,其中所述Cas9域是无核酸酶活性的Cas9(dCas9)。

44.权利要求43的融合蛋白,其中所述无核酸酶活性的Cas9(dCas9)包含与SEQ ID NO:7、8、9或22中的任一项至少85%相同的氨基酸序列。

45.权利要求43或44的融合蛋白,其中所述dCas9包含SEQ ID NO:7、8、9或22中任一项的氨基酸序列。

46.融合蛋白,其包含(i)核酸可编程DNA结合蛋白(napDNAbp),和(ii)胞苷脱氨酶域,以及(iii)核酸聚合酶(NAP)域。

47.权利要求46的融合蛋白,其中所述核酸聚合酶域是真核核酸聚合酶域。

48.权利要求46或47的融合蛋白,其中所述核酸聚合酶域是DNA聚合酶域。

49.权利要求46-48中任一项的融合蛋白,其中所述核酸聚合酶域具有跨损伤聚合酶活性。

50.权利要求46-49中任一项的融合蛋白,其中所述核酸聚合酶域是跨损伤DNA聚合酶。

51.权利要求46至50中任一项的融合蛋白,其中所述核酸聚合酶域来自Rev7、Rev1复合物、聚合酶iota、聚合酶kappa和聚合酶eta。

52.权利要求46-51中任一项的融合蛋白,其中所述核酸聚合酶域选自真核聚合酶的组,该组由alpha、beta、gamma、delta、epsilon、gamma、eta、iota、kappa、lambda、mu、和nu组成。

53.权利要求46-52中任一项的融合蛋白,其中所述核酸聚合酶域包含与SEQ ID NO:54-64中任一项的氨基酸序列至少80%、85%、90%、95%、98%或99%相同的氨基酸序列。

54.权利要求46-53中任一项的融合蛋白,其中所述核酸聚合酶域包含SEQ ID NO:54-64中任一项的氨基酸序列。

55.权利要求1-54中任一项的融合蛋白,其中所述胞苷脱氨酶域是来自载脂蛋白B mRNA编辑复合物(APOBEC)家族脱氨酶的脱氨酶。

56.权利要求55的融合蛋白,其中所述APOBEC家族脱氨酶选自下组:APOBEC1脱氨酶、APOBEC2脱氨酶、APOBEC3A脱氨酶、APOBEC3B脱氨酶、APOBEC3C脱氨酶、APOBEC3D脱氨酶、APOBEC3F脱氨酶、APOBEC3G脱氨酶、和APOBEC3H脱氨酶。

57.权利要求1-56中任一项的融合蛋白,其中所述胞苷脱氨酶域包含与SEQ ID NO:67-101中任一项的氨基酸序列至少85%相同的氨基酸序列。

58.权利要求1-57中任一项的融合蛋白,其中所述胞苷脱氨酶域包含SEQ ID NO:67-101中任一项的氨基酸序列。

59.权利要求1-57中任一项的融合蛋白，其中所述胞苷脱氨酶域是大鼠APOBEC1(rAPOBEC1)脱氨酶，其包含一个或多个选自下组的突变：SEQ ID NO:93的W90Y，R126E和R132E，或另一个APOBEC脱氨酶中的一个或多个相应突变。

60.权利要求1-57中任一项的融合蛋白，其中所述胞苷脱氨酶域是人APOBEC1(hAPOBEC1)脱氨酶，其包含一个或多个选自下组的突变：SEQ ID NO:91的W90Y，Q126E和R132E，或另一个APOBEC脱氨酶中的一个或多个相应突变。

61.权利要求1-57中任一项的融合蛋白，其中所述胞苷脱氨酶域是人APOBEC3G(hAPOBEC3G)脱氨酶，其包含一个或多个选自下组的突变：SEQ ID NO:77的W285Y，R320E和R326E，或另一个APOBEC脱氨酶中的一个或多个相应突变。

62.权利要求1-54中任一项的融合蛋白，其中所述胞苷脱氨酶域是激活诱导的脱氨酶(AID)。

63.权利要求1-54中任一项的融合蛋白，其中所述胞苷脱氨酶域是来自海七鳃鳗(Petromyzon marinus)的胞苷脱氨酶1(pmCDA1)。

64.权利要求46-63中任一项的融合蛋白，其中所述napDNAbp包含Cas9域。

65.权利要求64的融合蛋白，其中所述Cas9域包含与SEQ ID NO:4-26中任一项至少85％相同的氨基酸序列。

66.权利要求64或65的融合蛋白，其中所述Cas9域包含SEQ ID NO:4-26中任一项的氨基酸序列。

67.权利要求64的融合蛋白，其中所述Cas9域是Cas9切口酶(nCas9)。

68.权利要求67的融合蛋白，其中所述Cas9切口酶(nCas9)包含与SEQ ID NO:10、13、16或21至少85％相同的氨基酸序列。

69.权利要求67或68的融合蛋白，其中所述nCas9包含SEQ ID NO:10、13、16或21的氨基酸序列。

70.权利要求64的融合蛋白，其中所述Cas9域是无核酸酶活性的Cas9(dCas9)。

71.权利要求70的融合蛋白，其中所述无核酸酶活性的Cas9(dCas9)包含与SEQ ID NO:7、8、9或22至少85％相同的氨基酸序列。

72.权利要求70或71的融合蛋白，其中所述无核酸酶活性的Cas9(dCas9)包含SEQ ID NO:7、8、9或22的氨基酸序列。

73.权利要求46-72中任一项的融合蛋白，其中所述融合蛋白包含以下结构：

[胞苷脱氨酶]-[napDNAbp]-[NAP]，其中“]-[”的每个情况包含任选的接头。

74.权利要求46-73中任一项的融合蛋白，其中所述胞苷脱氨酶和所述napDNAbp是经由接头融合的。

75.权利要求74的融合蛋白，其中所述胞苷脱氨酶和所述napDNAbp是经由包含SEQ ID NO:102-109中任一项的氨基酸序列的接头融合的。

76.权利要求46-75中任一项的融合蛋白，其中所述napDNAbp和所述NAP是经由接头融合的。

77.权利要求76的融合蛋白，其中所述napDNAbp和所述NAP是经由包含SEQ ID NO:102-109中任一项的氨基酸序列的接头融合的。

78.融合蛋白，其包含(i)核酸可编程DNA结合蛋白(napDNAbp)，和(ii)碱基切除酶

(BEE)。

79.权利要求78的融合蛋白,其中所述碱基切除酶是胞嘧啶(C)或胸腺嘧啶(T)碱基切除酶。

80.权利要求78或79的融合蛋白,其中所述碱基切除酶是胞嘧啶(C)碱基切除酶。

81.权利要求79或80的融合蛋白,其中所述碱基切除酶包含与SEQ ID NO:48至少85%相同的氨基酸序列,并且在SEQ ID NO:48的氨基酸残基156处包含A。

82.权利要求80或81的融合蛋白,其中所述碱基切除酶包含SEQ ID NO:65的氨基酸序列。

83.权利要求78或79的融合蛋白,其中所述碱基切除酶是胸腺嘧啶(T)碱基切除酶。

84.权利要求83的融合蛋白,其中所述碱基切除酶包含与SEQ ID NO:48至少85%相同的氨基酸序列,并且在SEQ ID NO:48的氨基酸残基213处包含D。

85.权利要求B6或B7的融合蛋白,其中所述碱基切除酶包含SEQ ID NO:66的氨基酸序列。

86.权利要求78-85中任一项的融合蛋白,其中所述融合蛋白包含以下结构:

[napDNAbp]-[BEE]或[BEE]-[napDNAbp],其中"]-["的每个情况包含任选的接头。

87.权利要求78-86中任一项的融合蛋白,其中所述BEE和napDNAbp是经由接头融合的。

88.权利要求86或87的融合蛋白,其中所述BEE和所述napDNAbp是经由包含SEQ ID NO:102-109中任一项的氨基酸序列的接头融合的。

89.权利要求78-88中任一项的融合蛋白,其中所述融合蛋白进一步包含(iii)核酸聚合酶(NAP)域。

90.权利要求89的融合蛋白,其中所述核酸聚合酶域是真核核酸聚合酶域。

91.权利要求89或90的融合蛋白,其中所述核酸聚合酶域是DNA聚合酶域。

92.权利要求89-91中任一项的融合蛋白,其中所述核酸聚合酶域具有跨损伤聚合酶活性。

93.权利要求89-92中任一项的融合蛋白,其中所述核酸聚合酶域是跨损伤DNA聚合酶。

94.权利要求89-93中任一项的融合蛋白,其中所述核酸聚合酶域来自Rev7、Rev1复合物、聚合酶iota、聚合酶kappa和聚合酶eta.。

95.权利要求89-94中任一项的融合蛋白,其中所述核酸聚合酶域选自真核聚合酶的组,所述组由alpha、beta、gamma、delta、epsilon、gamma、eta、iota、kappa、lambda、mu、和nu组成。

96.权利要求89-95中任一项的融合蛋白,其中所述核酸聚合酶域包含与SEQ ID NO:54-64中任一项的氨基酸序列至少80%、85%、90%、95%、98%或99%相同的氨基酸序列。

97.权利要求89-96中任一项的融合蛋白,其中所述核酸聚合酶域包含SEQ ID NO:54-64中任一项的氨基酸序列。

98.权利要求89-97中任一项的融合蛋白,其中所述融合蛋白包含以下结构:

[napDNAbp]-[BEE]-[NAP];

[napDNAbp]-[NAP]-[BEE];

[NAP]-[napDNAbp]-[BEE];

[NAP]-[BEE]-[napDNAbp];

[BEE]-[napDNAbp]-[NAP];或

[BEE]-[NAP]-[napDNAbp]

其中"]-["的每个情况包括任选的接头。

99.权利要求89-98中任一项的融合蛋白,其中所述NAP和所述BEE是经由接头融合的。

100.权利要求99的融合蛋白,其中所述NAP和所述BEE是经由包含SEQ ID NO:102-109中任一项的氨基酸序列的接头融合的。

101.权利要求89-100中任一项的融合蛋白,其中所述BEE和napDNAbp是经由接头融合的。

102.权利要求101的融合蛋白,其中所述BEE和所述napDNAbp是经由包含SEQ ID NO:102-109中任一项的氨基酸序列的接头融合的。

103.权利要求89-102中任一项的融合蛋白,其中所述BEE和所述napDNAbp是经由接头融合的。

104.权利要求103的融合蛋白,其中所述BEE和napDNAbp是经由包含SEQ ID NO:102-109中任一项的氨基酸序列的接头融合的。

105.权利要求78-104中任一项的融合蛋白,其中所述napDNAbp包含Cas9域。

106.权利要求105的融合蛋白,其中所述Cas9域包含与SEQ ID NO:4-26中任一项至少85%相同的氨基酸序列。

107.权利要求105或106的融合蛋白,其中所述Cas9域包含SEQ ID NO:4-26中任一项的氨基酸序列。

108.权利要求105的融合蛋白,其中所述Cas9域是Cas9切口酶(nCas9)。

109.权利要求108的融合蛋白,其中所述Cas9切口酶(nCas9)包含与SEQ ID NO:10、13、16或21至少85%相同的氨基酸序列。

110.权利要求108或109的融合蛋白,其中所述nCas9包含SEQ ID NO:10、13、16或21的氨基酸序列。

111.权利要求105的融合蛋白,其中所述Cas9域是无核酸酶活性的Cas9(dCas9)。

112.权利要求111的融合蛋白,其中所述无核酸酶活性的Cas9(dCas9)包含与SEQ ID NO:7、8、9或22至少85%相同的氨基酸序列。

113.权利要求B34或B35的融合蛋白,其中所述dCas9包含SEQ ID NO:7、8、9或22的氨基酸序列。

114.融合蛋白,其包含:

(i)第一氨基酸序列,其与SEQ ID NO:4-40中任一项的氨基酸序列至少80%相同;

(ii)第二氨基酸序列,其与SEQ ID NO:67-101中任一项的氨基酸序列至少80%相同;和

(iii)第三氨基酸序列,其与SEQ ID NO:48-53中任一项的氨基酸序列至少80%相同。

115.权利要求114的融合蛋白,其进一步包含:

(iv)第四氨基酸序列,其与SEQ ID NO:54-64中任一项的氨基酸序列至少80%相同。

116.权利要求114或115的融合蛋白,其中所述第一氨基酸序列与SEQ ID NO:4-40中任一项的氨基酸序列至少85%、90%、95%、98%或99%相同。

117.权利要求114-116中任一项的融合蛋白,其中所述第二氨基酸序列与SEQ ID NO:

67-101的任一项的氨基酸序列至少85％、90％、95％、98％或99％相同。

118.权利要求114-117中任一项的融合蛋白，其中所述第三氨基酸序列与SEQ ID NO:48-53的任一项的氨基酸序列至少85％、90％、95％、98％或99％相同。

119.权利要求115-118中任一项的融合蛋白，其中所述第四氨基酸序列与SEQ ID NO:54-64的任一项的氨基酸序列至少85％、90％、95％、98％或99％相同。

120.权利要求114-119中任一项的融合蛋白，其中所述融合蛋白包含以下结构：

$NH_2$-[第二氨基酸序列]-[第一氨基酸序列]-[第三氨基酸序列]-COOH；

$NH_2$-[第二氨基酸序列]-[第三氨基酸序列]-[第一氨基酸序列]-COOH；

$NH_2$-[第一氨基酸序列]-[第二氨基酸序列]-[第三氨基酸序列]-COOH；

$NH_2$-[第一氨基酸序列]-[第三氨基酸序列]-[第二氨基酸序列]-COOH；

$NH_2$-[第三氨基酸序列]-[第一氨基酸序列]-[第二氨基酸序列]-COOH；

$NH_2$-[第三氨基酸序列]-[第二氨基酸序列]-[第一氨基酸序列]-COOH；

其中“-”的每个情况包括任选的接头。

121.权利要求115-119中任一项的融合蛋白，其中所述融合蛋白包含以下结构：

$NH_2$-[第四氨基酸序列]-[第二氨基酸序列]-[第一氨基酸序列]-[第三氨基酸序列]-COOH；

$NH_2$-[第二氨基酸序列]-[第四氨基酸序列]-[第一氨基酸序列]-[第三氨基酸序列]-COOH；

$NH_2$-[第二氨基酸序列]-[第一氨基酸序列]-[第四氨基酸序列]-[第三氨基酸序列]-COOH；

$NH_2$-[第二氨基酸序列]-[第一氨基酸序列]-[第三氨基酸序列]-[第四氨基酸序列]-COOH；

$NH_2$-[第四氨基酸序列]-[第二氨基酸序列]-[第三氨基酸序列]-[第一氨基酸序列]-COOH；

$NH_2$-[第二氨基酸序列]-[第四氨基酸序列]-[第三氨基酸序列]-[第一氨基酸序列]-COOH；

$NH_2$-[第二氨基酸序列]-[第三氨基酸序列]-[第四氨基酸序列]-[第一氨基酸序列]-COOH；

$NH_2$-[第二氨基酸序列]-[第三氨基酸序列]-[第一氨基酸序列]-[第四氨基酸序列]-COOH；

$NH_2$-[第四氨基酸序列]-[第一氨基酸序列]-[第二氨基酸序列]-[第三氨基酸序列]-COOH；

$NH_2$-[第一氨基酸序列]-[第四氨基酸序列]-[第二氨基酸序列]-[第三氨基酸序列]-COOH；

$NH_2$-[第一氨基酸序列]-[第二氨基酸序列]-[第四氨基酸序列]-[第三氨基酸序列]-COOH；

$NH_2$-[第一氨基酸序列]-[第二氨基酸序列]-[第三氨基酸序列]-[第四氨基酸序列]-COOH；

$NH_2$-[第四氨基酸序列]-[第一氨基酸序列]-[第三氨基酸序列]-[第二氨基酸序列]-

COOH；

$NH_2$-[第一氨基酸序列]-[第四氨基酸序列]-[第三氨基酸序列]-[第二氨基酸序列]-COOH；

$NH_2$-[第一氨基酸序列]-[第三氨基酸序列]-[第四氨基酸序列]-[第二氨基酸序列]-COOH；

$NH_2$-[第一氨基酸序列]-[第三氨基酸序列]-[第二氨基酸序列]-[第四氨基酸序列]-COOH；

$NH_2$-[第四氨基酸序列]-[第三氨基酸序列]-[第一氨基酸序列]-[第二氨基酸序列]-COOH；

$NH_2$-[第三氨基酸序列]-[第四氨基酸序列]-[第一氨基酸序列]-[第二氨基酸序列]-COOH；

$NH_2$-[第三氨基酸序列]-[第一氨基酸序列]-[第四氨基酸序列]-[第二氨基酸序列]-COOH；

$NH_2$-[第三氨基酸序列]-[第一氨基酸序列]-[第二氨基酸序列]-[第四氨基酸序列]-COOH；

$NH_2$-[第四氨基酸序列]-[第三氨基酸序列]-[第二氨基酸序列]-[第一氨基酸序列]-COOH；

$NH_2$-[第三氨基酸序列]-[第四氨基酸序列]-[第二氨基酸序列]-[第一氨基酸序列]-COOH；

$NH_2$-[第三氨基酸序列]-[第二氨基酸序列]-[第四氨基酸序列]-[第一氨基酸序列]-COOH；

$NH_2$-[第三氨基酸序列]-[第二氨基酸序列]-[第一氨基酸序列]-[第四氨基酸序列]-COOH；

其中“-”的每个情况包括任选的接头。

122.融合蛋白，其包含：

(i)第一氨基酸序列，其与SEQ ID NO:4-40中任一项的氨基酸序列至少80％相同；

(ii)第二氨基酸序列，其与SEQ ID NO:67-101中任一项的氨基酸序列至少80％相同；和

(iii)第三氨基酸序列，其与SEQ ID NO:54-64中任一项的氨基酸序列至少80％相同。

123.权利要求122的融合蛋白，其中所述第一氨基酸序列与SEQ ID NO:4-40中任一项的氨基酸序列至少85％、90％、95％、98％或99％相同。

124.权利要求122-123中任一项的融合蛋白，其中所述第二氨基酸序列与SEQ ID NO:67-101中任一项的氨基酸序列至少85％、90％、95％、98％或99％相同。

125.权利要求122-124中任一项的融合蛋白，其中所述第三氨基酸序列与SEQ ID NO:54-64中任一项的氨基酸序列至少85％、90％、95％、98％或99％相同。

126.权利要求122-125中任一项的融合蛋白，其中所述融合蛋白包含以下结构：

$NH_2$-[第二氨基酸序列]-[第一氨基酸序列]-[第三氨基酸序列]-COOH；

$NH_2$-[第二氨基酸序列]-[第三氨基酸序列]-[第一氨基酸序列]-COOH；

$NH_2$-[第一氨基酸序列]-[第二氨基酸序列]-[第三氨基酸序列]-COOH；

$NH_2$-[第一氨基酸序列]-[第三氨基酸序列]-[第二氨基酸序列]-COOH;

$NH_2$-[第三氨基酸序列]-[第一氨基酸序列]-[第二氨基酸序列]-COOH;

$NH_2$-[第三氨基酸序列]-[第二氨基酸序列]-[第一氨基酸序列]-COOH;

其中“-”的每个情况包括任选的接头。

127.融合蛋白,其包含:

(i)第一氨基酸序列,其与SEQ ID NO:4-40中任一项的氨基酸序列至少80%相同;

(ii)第二氨基酸序列,其与SEQ ID NO:65-66中任一项的氨基酸序列至少80%相同;和。

128.权利要求127的融合蛋白,其进一步包含:

(iii)第三氨基酸序列,其与SEQ ID NO:54-64中任一项的氨基酸序列至少80%相同。

129.权利要求127或128的融合蛋白,其中所述第一氨基酸序列与SEQ ID NO:4-40中任一项的氨基酸序列至少85%、90%、95%、98%或99%相同。

130.权利要求127-129中任一项的融合蛋白,其中所述第二氨基酸序列与SEQ ID NO:65-66中任一项的氨基酸序列至少85%、90%、95%、98%或99%相同。

131.权利要求128-130中任一项的融合蛋白,其中所述第三氨基酸序列与SEQ ID NO:54-64中任一项的氨基酸序列至少85%、90%、95%、98%或99%相同。

132.权利要求127-131中任一项的融合蛋白,其中所述融合蛋白包含以下结构:

$NH_2$-[第二氨基酸序列]-[第一氨基酸序列]-COOH;或

$NH_2$-[第一氨基酸序列]-[第二氨基酸序列]-COOH;

其中“-”的每个情况包括任选的接头。

133.权利要求128-131中任一项的融合蛋白,其中所述融合蛋白包含以下结构:

$NH_2$-[第二氨基酸序列]-[第一氨基酸序列]-[第三氨基酸序列]-COOH;

$NH_2$-[第二氨基酸序列]-[第三氨基酸序列]-[第一氨基酸序列]-COOH;

$NH_2$-[第一氨基酸序列]-[第二氨基酸序列]-[第三氨基酸序列]-COOH;

$NH_2$-[第一氨基酸序列]-[第三氨基酸序列]-[第二氨基酸序列]-COOH;

$NH_2$-[第三氨基酸序列]-[第一氨基酸序列]-[第二氨基酸序列]-COOH;

$NH_2$-[第三氨基酸序列]-[第二氨基酸序列]-[第一氨基酸序列]-COOH;

其中“-”的每个情况包括任选的接头。

134.复合物,其包含核酸分子和权利要求1-133中任一项的融合蛋白。

135.权利要求134的复合物,其中所述核酸分子是RNA。

136.权利要求134或135的复合物,其中所述核酸分子是gRNA。

137.权利要求134-136中任一项的复合物,其中所述核酸分子是sgRNA。

138.药物组合物,其包含权利要求1-133中任一项的融合蛋白。

139.药物组合物,其包含权利要求134-137中任一项的复合物。

140.权利要求138-139中任一项的药物组合物,其还包含药学上可接受的赋形剂。

141.方法,其包括使核酸分子与权利要求1-133中任一项的融合蛋白或权利要求134-137中任一项的复合物接触。

142.权利要求141的方法,其中所述核酸分子是DNA。

143.权利要求141或142的方法,其中所述核酸分子是基因组的DNA。

144.编辑双链DNA序列的核碱基对的方法,所述方法包括:

a)使所述双链DNA序列的靶区域与包含核碱基编辑器和引导核酸的复合物接触,其中所述靶区域包含靶核碱基对;

b)诱导所述靶区域的链分离;并且

c)在所述靶区域的单链中切除胞嘧啶或胸腺嘧啶。

145.权利要求144的方法,其中步骤c包括切除所述靶区域的单链中的胞嘧啶。

146.用于编辑双链DNA序列的核碱基对的方法,所述方法包括:

a)使所述双链DNA序列的靶区域与核酸可编程DNA结合蛋白接触,其中所述靶区域包含靶核碱基对;

b)切除所述靶区域的所述靶核碱基对的第一核碱基;并且

c)用不同于所述第一核碱基的第二核碱基替换所述第一核碱基。

147.包含核酸构建体的试剂盒,所述核酸构建体包含

(a)编码权利要求1-133中任一项的融合蛋白的核酸序列;和

(b)驱动(a)表达的异源启动子。

148.编码权利要求1-133中任一项的融合蛋白的多核苷酸。

149.包含权利要求148的多核苷酸的载体。

150.细胞,其包含权利要求1-133中任一项的融合蛋白。

151.细胞,其包含权利要求134-137中任一项的复合物。

152.细胞,其包含编码权利要求1-133中任一项的融合蛋白的核酸分子。

153.权利要求150-152中任一项的细胞,其中所述细胞是体外的。

154.权利要求150-152中任一项的细胞,其中所述细胞是体内的。

155.治疗患有或怀疑患有疾病或病症的受试者的方法,其包括对所述受试者施用权利要求1-133中任一项的融合蛋白,权利要求134-137中任一项的复合物,权利要求148的多核苷酸或权利要求149的载体。

156.用于编辑双链DNA序列的核碱基对的方法,所述方法包括:

a.使所述双链DNA序列的靶区域与包含核碱基编辑器和引导核酸的复合物接触,其中所述靶区域包含靶核碱基对;

b.诱导所述靶区域的链分离;

c.将所述靶区域的单链中的所述靶核碱基对的第一核碱基转化为第二核碱基;

d.从所述双链DNA序列中切除所述第二核碱基以产生无碱基位点;并且

e.切割所述靶区域的不超过一条链;

其中与所述无碱基位点相对的第三核碱基被第四核碱基替代。

157.权利要求156的方法,其中所述方法引起小于20%至少于1%的插入/缺失形成。

158.权利要求156的方法,其进一步包括将与所述第四核碱基互补的第五核碱基插入所述无碱基位点中,从而产生意图编辑的碱基对。

159.权利要求158的方法,其中产生所述意图编辑的碱基对的效率是至少5%。

160.权利要求159的方法,其中所述效率为至少10%到至少50%。

161.权利要求158的方法,其中意图编辑的碱基对与非意图编辑的碱基对的比率为2:1至10:1。

162.权利要求158的方法，其中意图编辑的碱基对与插入/缺失形成的比率为2:1至200:1。

163.权利要求156的方法，其中所述切割链与所述引导核酸杂交。

164.权利要求156的方法，其中所述切割的单链与包含所述第一核碱基的链相对。

165.权利要求156的方法，其中所述第一核碱基是胞嘧啶。

166.权利要求156的方法，其中所述第二核碱基是尿嘧啶。

167.权利要求156的方法，其中所述核碱基编辑器包含尿嘧啶结合蛋白。

168.权利要求156的方法，其中所述核碱基编辑器包含切口酶活性。

169.权利要求159的方法，其中所述意图编辑的碱基对在PAM位点的上游。

170.权利要求169的方法，其中所述意图编辑的碱基对在PAM位点上游的1、2、3、4、5、6、7、8、9、10、11、12、13、14、15、16、17、18、19或20个核苷酸。

171.权利要求158的方法，其中所述意图编辑的碱基对在PAM位点的下游。

172.权利要求171的方法，其中所述意图编辑的碱基对在PAM位点下游的1、2、3、4、5、6、7、8、9、10、11、12、13、14、15、16、17、18、19或20个核苷酸。

173.权利要求156的方法，其中所述核碱基编辑器包括接头。

174.权利要求173的方法，其中所述接头的长度为1-25个氨基酸。

175.权利要求173的方法，其中所述接头的长度为5-20个氨基酸。

176.权利要求173的方法，其中所述接头的长度为10、11、12、13、14、15、16、17、18、19或20个氨基酸。

177.权利要求156的方法，其中所述靶区域包括靶窗，其中所述靶窗包括所述靶核碱基对。

178.权利要求177的方法，其中所述靶窗包含1-10个核苷酸。

179.权利要求177的方法，其中所述靶窗的长度为1-9、1-8、1-7、1-6、1-5、1-4、1-3、1-2或1个核苷酸。

180.权利要求177的方法，其中所述靶窗的长度是1、2、3、4、5、6、7、8、9、10、11、12、13、14、15、16、17、18、19或20个核苷酸。

181.权利要求177-180的方法，其中所述靶窗包括所述意图编辑的碱基对。

# 胞嘧啶至鸟嘌呤碱基编辑器

[0001] 发明背景

[0002] 核酸序列的靶向编辑,例如对基因组DNA的靶向切割或特定修饰的靶向引入,是用于研究基因功能的高度有前景的方法,并且还具有为人类遗传疾病提供新疗法的潜力。由于许多遗传疾病原则上可以通过影响基因组中的特定位置处特定的核苷酸变化(例如,与疾病相关的基因的特定密码子中的C至G或G至C的变化)来治疗,实现此类精确的基因编辑的可编程方式的开发既代表强大的新研究工具,又代表基于基因编辑的治疗学的潜在新方法。

[0003] 发明概述

[0004] 本文中提供修饰多核苷酸(例如,DNA),例如在多核苷酸中产生胞嘧啶至鸟嘌呤突变的组合物、试剂盒和方法。如本文中更详细描述的,碱基编辑(例如,C至G编辑)通过除去核碱基(例如,胞嘧啶(C)),从而在核酸序列内产生无碱基位点来完成。然后,例如通过内源性跨损伤聚合酶将与无碱基位点相对的核碱基(例如鸟嘌呤)替换为不同的核碱基(例如胞嘧啶)。本文所述的碱基编辑融合蛋白能够在核酸(例如基因组DNA)内产生特定突变(例如C至G突变),其可用于例如治疗涉及核酸突变,例如C至G或G至C突变的疾病。

[0005] C至G碱基编辑器的一个实例包括融合蛋白,其含有核酸可编程DNA结合蛋白(例如,Cas9域)、尿嘧啶DNA糖基化酶(UDG)域和胞苷脱氨酶。不希望受任何特定理论的束缚,此类碱基编辑融合蛋白能够与特定核酸序列结合(例如,通过Cas9域),使核酸序列内的胞嘧啶脱氨基为尿苷,然后尿苷可以通过UDG从核酸分子中切除。然后可以例如通过内源性跨损伤聚合酶将与无碱基位点相对的核碱基替换为另一种碱基(例如胞嘧啶)。通常,碱基修复机制(例如,在细胞中)用胞嘧啶替换无碱基位点相对的核碱基,尽管其他碱基(例如,腺嘌呤、鸟嘌呤或胸腺嘧啶)可以替换与无碱基位点相对的核碱基。此外,发现将跨损伤聚合酶掺入碱基编辑器中可以增加与无碱基位点相对的胞嘧啶掺入。因此,对碱基编辑器进行工程化改造,以掺入各种跨损伤聚合酶,以提高碱基编辑效率。增加与无碱基位点相对的C整合的偏好的跨损伤聚合酶可以改善C至G核碱基编辑。应当理解,优先整合非C核碱基(例如腺嘌呤、鸟嘌呤和胸腺嘧啶)的其他跨损伤聚合酶可以用于产生备选突变(例如C至A突变)。

[0006] 作为另一个例子,碱基编辑融合蛋白可以包括核酸可编程DNA结合蛋白(例如,Cas9域)和除去核碱基(例如,胞嘧啶)的碱基切除酶(base excision enzyme)。如上所述,不是将胞嘧啶脱氨基成尿苷并使用UDG将尿苷切除,碱基编辑器可以包括碱基切除酶,该碱基切除酶识别并除去核碱基诸如胞嘧啶或胸腺嘧啶,而无需首先对其进行脱氨基。因此,已经通过将核酸可编程DNA结合蛋白(例如Cas9域)融合到从核酸分子中除去胞嘧啶或胸腺嘧啶的碱基切除酶来工程化改造碱基编辑器(例如,C至G碱基编辑器)。此外,与上文描述的碱基编辑器一样,将跨损伤聚合酶掺入该碱基编辑器中以增加与通过碱基编辑器的碱基切除酶产生的无碱基位点相对的胞嘧啶掺入。例如,在图1-6、33-36、40和52中可以看到示例性的碱基编辑蛋白和概述碱基编辑策略的示意图。

[0007] 在一些实施方案中,本公开内容提供了能够进行碱基编辑的融合蛋白。示例性的碱基编辑融合蛋白包括以下各项。在一些实施方案中,融合蛋白包括(i)核酸可编程DNA结

合蛋白(napDNAbp),(ii)胞苷脱氨酶域和(iii)尿嘧啶结合蛋白(UBP)。在一些实施方案中,融合蛋白进一步包含(iv)核酸聚合酶域(NAP)。作为另一个例子,融合蛋白可包含(i)核酸可编程DNA结合蛋白(napDNAbp),(ii)胞苷脱氨酶域和(iii)核酸聚合酶(NAP)域。作为另一个例子,融合蛋白可包含(i)核酸可编程DNA结合蛋白(napDNAbp),和(ii)碱基切除酶(BEE)。在一些实施方案中,融合蛋白还包括(iii)核酸聚合酶(NAP)域。在下面进一步详细描述碱基编辑器和使用碱基编辑器的方法。

[0008] 附图简述

[0009] 图1显示了一般示意图,其显示了C至T和C至G碱基编辑。已知某些DNA聚合酶(例如,跨损伤聚合酶)用G替换无碱基位点相对的碱基。实现C至G碱基编辑的一种策略是诱导无碱基位点的创建,然后募集或拴系此类聚合酶来用C替换替换无碱基位点的G。

[0010] 图2显示了一般示意图,其显示了通过无碱基位点生成的碱基编辑和用于C至G编辑的碱基特异性修复。

[0011] 图3显示了显示来自图1的方案1的示意图,其中形成无碱基位点,用于C至G碱基编辑。若有效生成无碱基,则这可以增加经由C至G编辑途径的总通量。

[0012] 图4显示了显示用于C至G碱基编辑的方法1的示意图,其中使用无碱基位点形成的增加。若通过例如使用UDG域和跨损伤聚合酶有效地产生无碱基,则这可以增加经由C至G编辑途径的总通量。

[0013] 图5显示了显示UdgX对碱基编辑的影响的示意图。UdgX(一种经鉴定与具有最小尿嘧啶切割活性的尿嘧啶紧密结合的UDG直向同源物)增加C至G编辑的量。在1.)中,UdgX*是通过体外测定法确定缺乏尿嘧啶结合活性的UDG变体。在2.)中,UdgX_On是通过体外测定法显示增加尿嘧啶切割的变体。在3.)中,UDG指导融合物切割尿嘧啶。

[0014] 图6显示了示意图(左图),其显示了示例性的C至T碱基编辑器(例如,BE3),其包含尿嘧啶糖基化酶抑制剂(UGI)、Cas9域(例如,nCas9)和胞苷脱氨酶。右边是显示C至G碱基编辑器的示意图,所述C至G碱基编辑器包含尿嘧啶DNA糖基化酶(UDG)(或其变体)、Cas9域(例如nCas9)和胞苷脱氨酶。

[0015] 图7显示了使用七种碱基编辑器(BE3;BE3_UdgX;BE3_UdgX*;BE2_UdgX_On;BE3_UdgX_On;BE2_UDG;和BE3_UDG)在WT Hap1细胞中HEK2位点处的总编辑百分比。原始编辑值显示在左图中。右图显示了对与含有经编辑的C的链相对的DNA链进行测序时原始编辑值的图示,其中C至G碱基编辑通过点柱形(G)移至实心柱形(C)以图形显示。

[0016] 图8显示了在WT Hap1细胞中用其它C至G碱基编辑器(BE3;BE3_UdgX;BE3_REV7;和SMUG1,其中从图4重复BE3和BE3_UdgX)在HEK2位点处的总编辑百分比。顶部图显示原始编辑值。底部图显示了对与含有经编辑的C的链相对的DNA链进行测序时原始编辑值的图示,其中C至G的碱基编辑通过点柱形(G)移至实心柱形(C)以图形显示。

[0017] 图9显示了WT Hap1细胞中用各种C至G碱基编辑器(BE3;BE3_UdgX;BE3_UdgX*;BE3_REV7;BE2_UDG;BE3_UDG BE2_UdgX_On;BE3_UdgX_On;和SMUG1)在HEK2位点处的编辑特异性比。顶部图显示了编辑物的总百分比以及已经从G至A、C或T产生的编辑物的比率。底部图是特异性比率值的图示。

[0018] 图10显示了使用七种碱基编辑器(BE3;BE3_UdgX;BE3_UdgX*;BE2_UdgX_On;BE3_UdgX_On;BE2_UDG;和BE3_UDG)在WT Hap1细胞中RNF2位点处的总编辑百分比。原始编辑值

显示在左图中。右边的图显示了对与含有经编辑的C的链相对的DNA链进行测序时原始编辑值的图示，其中C至G的碱基编辑通过点柱形(G)移至实心柱形(C)以图形显示。

[0019] 图11显示了在WT Hap1细胞中，RNF2位点上具有其他C至G碱基编辑器(BE3;BE3_UdgX;BE3_REV7;和SMUG1，其中从图7重复了BE3和BE3_UdgX)的总编辑百分比。顶部图显示原始编辑值。底部图显示了对与含有经编辑的C的链相对的DNA链进行测序时原始编辑值的图示，其中C至G的碱基编辑通过点柱形(G)移至实心柱形(C)以图形显示。

[0020] 图12显示了在WT Hap1细胞中用各种C至G碱基编辑器(BE3;BE3_UdgX;BE3_UdgX*;BE3_REV7;BE2_UDG;BE3_UDG BE2_UdgX_On;BE3_UdgX_On;和SMUG1)在RNF2位点处的编辑特异性比率。顶部图显示了编辑物的总百分比以及已经从G至A、C或T产生的编辑物比率。底部图是特异性比率值的图示。

[0021] 图13显示了使用七种碱基编辑器(BE3;BE3_UdgX;BE3_UdgX*;BE2_UdgX_On;BE3_UdgX_On;BE2_UDG;和BE3_UDG)在WT Hap1细胞中FANCF位点处的总编辑百分比。原始编辑值显示在左图中。右侧的图显示了原始编辑值的图示，其中C至G的碱基编辑通过实心柱形(C)移至点柱形(G)以图形显示。

[0022] 图14显示了WT Hap1细胞中用其他C至G碱基编辑器(BE3;BE3_UdgX;BE3_REV7;和SMUG1，其中BE3和BE3_UdgX从图10重复)在FANCF位点处的总编辑百分比。顶部图显示原始编辑值。底部图显示了原始编辑值的图示，其中C至G的碱基编辑通过通过实心柱形(C)移至点柱形(G)以图形显示。

[0023] 图15显示了WT Hap细胞中用各种C至G碱基编辑器(BE3;BE3_UdgX;BE3_UdgX*;BE3_REV7;BE2_UDG;BE3_UDG BE2_UdgX_On;BE3_UdgX_On;和SMUG1)在FANCF位点处的编辑特异性比率。顶部图显示了编辑物的总百分比和已经从C至A、G或T产生的编辑物的比率。底部图是特异性比率值的图示。

[0024] 图16显示了使用七种碱基编辑器(BE3;BE3_UdgX;BE3_UdgX*;BE2_UdgX_On;BE3_UdgX_On;BE2_UDG;和BE3_UDG)在$UDG^{-/-}$Hap1细胞中HEK2位点处的总编辑百分比。原始编辑值显示在左图中。右边的图显示了对与含有经编辑的C的链相对的DNA链进行测序时原始编辑值的图示，其中C至G的碱基编辑通过点柱形(G)移至实心柱形(C)以图形显示。

[0025] 图17显示了在$UDG^{-/-}$Hap1细胞中用其他C至G碱基编辑器(BE3;BE3_UdgX;BE3_REV7;和SMUG1，其中BE3和BE3_UdgX从图13重复)在HEK2处的总编辑百分比。顶部图显示原始编辑值。底部图显示了对与含有经编辑的C的链相对的DNA链进行测序时原始编辑值的图示，其中C至G的碱基编辑通过点柱形(G)移至实心柱形(C)以图形显示。

[0026] 图18显示了在$UDG^{-/-}$Hap1细胞中用各种C至G碱基编辑器(BE3;BE3_UdgX;BE3_UdgX*;BE3_REV7;BE2_UDG;BE3_UDG BE2_UdgX_On;BE3_UdgX_On;和SMUG1)在HEK2位点处的编辑特异性比率。顶部图显示了编辑物的总百分比以及已经从G至A、C或T产生的编辑物比率。底部图是特异性比率值的图示。

[0027] 图19显示了使用七种碱基编辑器(BE3;BE3_UdgX;BE3_UdgX*;BE2_UdgX_On;BE3_UdgX_On;BE2_UDG;和BE3_UDG)在$UDG^{-/-}$Hap1细胞中RNF2位点处的总编辑百分比。原始编辑值显示在左图中。右边的图显示了对与含有经编辑的C的链相对的DNA链进行测序时原始编辑值的图示，其中C至G的碱基编辑通过点柱形(G)移至实心柱形(C)以图形显示。

[0028] 图20显示了$UDG^{-/-}$Hap1细胞中用其他C至G碱基编辑器(BE3;BE3_UdgX;BE3_REV7;

和SMUG1,其中BE3和BE3_UdgX从图16重复)在RNF2位点处的总编辑百分比。顶部图显示原始编辑值。底部图显示了对与含有经编辑的C的链相对的DNA链进行测序时原始编辑值的图示,其中C至G的碱基编辑通过点柱形(G)移至实心柱形(C)以图形显示。

[0029] 图21显示了UDG$^{-/-}$Hap1细胞中用各种C至G碱基编辑器(BE3;BE3_UdgX;BE3_UdgX*;BE3_REV7;BE2_UDG;BE3_UDG BE2_UdgX_On;BE3_UdgX_On;和SMUG1)在RNF2位点处的编辑特异性比率。顶部图显示了编辑物的总百分比以及已经从G至A、C或T产生的编辑物比率。底部图是特异性比率值的图示。

[0030] 图22显示了使用七种碱基编辑器(BE3;BE3_UdgX;BE3_UdgX*;BE2_UdgX_On;BE3_UdgX_On;BE2_UDG;和BE3_UDG)在UDG$^{-/-}$Hap1细胞中FANCF位点处的总编辑百分比。原始编辑值显示在左图中。右侧的图显示了原始编辑值的图示,其中C至G的碱基编辑通过通过实心柱形(C)移至点柱形(G)以图形显示。

[0031] 图23显示了在UDG$^{-/-}$Hap1细胞中用其他C至G碱基编辑器(BE3;BE3_UdgX;BE3_REV7;和SMUG1,其中BE3和BE3_UdgX从图19重复)在FANCF位点处的总编辑百分比。顶部图显示原始编辑值。底部图显示了原始编辑值的图示,其中C至G的碱基编辑通过通过实心柱形(C)移至点柱形(G)以图形显示。

[0032] 图24显示了在UDG$^{-/-}$Hap1细胞上用各种C至G碱基编辑器(BE3;BE3_UdgX;BE3_UdgX*;BE3_REV7;BE2_UDG;BE3_UDG BE2_UdgX_On;BE3_UdgX_On;和SMUG1)在FANCF位点处的编辑特异性比率。顶部图显示了总编辑物的百分比和已经从C至A、G或T产生的编辑物比率。底部图是特异性比率值的图示。

[0033] 图25显示了REV1$^{-/-}$Hap1细胞中用各种C至G碱基编辑器(BE3;BE3_UdgX;BE2_UNG;BE3_UNG;BE2UdgX_On;BE3UdgX_On;和SMUG1)在HEK2位点处的总编辑百分比。顶部图显示原始编辑值。底部图显示了对与含有经编辑的C的链相对的DNA链进行测序时原始编辑值的图示,其中C至G的碱基编辑通过点柱形(G)移至实心柱形(C)以图形显示。

[0034] 图26显示了在REV1$^{-/-}$Hap1细胞中用各种C至G碱基编辑器(BE3;BE3_UdgX;BE2_UNG;BE3_UNG;BE2UdgX_On;BE3UdgX_On;和SMUG1)在HEK2位点处的编辑特异性比率。顶部图显示了总编辑物百分比以及从G至A、C或T产生的编辑物比率。底部图是特异性比率值的图示。

[0035] 图27显示了REV1$^{-/-}$Hap1细胞中用各种C至G碱基编辑器(BE3;BE3_UdgX;BE2_UNG;BE3_UNG;BE2UdgX_On;BE3UdgX_On;和SMUG1)在RNF2位点处的总编辑百分比。顶部图显示原始编辑值。底部图显示了对与含有经编辑的C的链相对的DNA链进行测序时原始编辑值的图示,其中C至G的碱基编辑通过点柱形(G)移至实心柱形(C)以图形显示。

[0036] 图28显示了REV1$^{-/-}$Hap1细胞中用各种C至G碱基编辑器(BE3;BE3_UdgX;BE2_UNG;BE3_UNG;BE2UdgX_On;BE3UdgX_On;和SMUG1)在RNF2位点处的编辑特异性比率。顶部图显示了总编辑物百分比以及从G至A、C或T产生的编辑物比率。底部图是特异性比率值的图示。

[0037] 图29显示了REV1$^{-/-}$Hap1细胞中用各种C至G碱基编辑器(BE3;BE3_UdgX;BE2_UNG;BE3_UNG;BE2UdgX_On;BE3UdgX_On;和SMUG1)在FANCF位点处的总编辑百分比。顶部图显示原始编辑值。底部图显示了原始编辑值的图示,其中C至G的碱基编辑通过通过实心柱形(C)移至点柱形(G)以图形显示。

[0038] 图30显示了REV1$^{-/-}$Hap1细胞中用各种C至G碱基编辑器(BE3;BE3_UdgX;BE2_UNG;

BE3_UNG;BE2UdgX_On;BE3UdgX_On;和SMUG1)在FANCF位点处的编辑特异性比率。顶部图显示了总编辑物百分比以及从G至A、C或T产生的编辑物比率。底部图是特异性比率值的图示。

[0039] 图31显示了使用指定的C至G碱基编辑器在HEK2、RNF2和FANCF位点处总编辑百分比的原始编辑值的图示。

[0040] 图32显示了在HEK2、RNF2和FANCF位点处的总编辑百分比的特异性比率的图示。

[0041] 图33显示了示意图,该示意图显示了与无碱基位点相对的C的掺入增加的方法,用于C至G碱基编辑。若例如通过使用聚合酶(例如,跨损伤聚合酶)增加与无碱基位点相对的C整合的偏好,则总的C至G碱基编辑也将增加。

[0042] 图34显示了示意图,该示意图显示了与无碱基位点相对的C的掺入增加的方法,用于C至G碱基编辑。若例如通过将跨损伤聚合酶掺入碱基编辑器中增加与无碱基位点相对的C整合的偏好,则总的C至G碱基编辑也可增加。

[0043] 图35显示了示意图,其显示了可以在图33和34的C至G碱基编辑方法中使用的不同聚合酶。

[0044] 图36显示了示意图(左图),其显示了示例性的C至T碱基编辑器(例如,BE3),其包含尿嘧啶糖基化酶抑制剂(UGI)、Cas9域(例如,nCas9)和胞苷脱氨酶。右边是显示C至G碱基编辑器的示意图,该C至G碱基编辑器含有跨损伤聚合酶、Cas9域(例如,nCas9)和胞苷脱氨酶。

[0045] 图37显示了使用拴系于REV1、Pol Kappa、Pol Eta、和Pol Iota的碱基编辑器在WT细胞中HEK2位点处的碱基编辑。C至G的编辑在右图的示意图中通过点柱形(G)移到实心柱形(C)以图形显示。Pol Kappa栓系大大提高C至G编辑的效率。原始编辑值显示在左图上。

[0046] 图38显示了使用拴系到REV1、Pol Kappa、Pol Eta和Pol Iota的碱基编辑器在WT细胞中RNF2位点处的碱基编辑。C至G的编辑在右图的示意图中通过点柱形(G)移到实心柱形(C)以图形显示。Pol Kappa栓系大大提高C至G编辑的效率。原始编辑值显示在左图上。

[0047] 图39显示了使用拴系到REV1、Pol Kappa、Pol Eta和Pol Iota的碱基编辑器在WT细胞中FANCF位点处的碱基编辑。C至G的编辑在右图的示意图中通过点柱形(C)移到实心柱形(G)以图形显示。Pol Kappa栓系大大提高C至G编辑的效率。原始编辑值显示在左图上。

[0048] 图40显示了示意图(左图),其显示了示例性的C至G碱基编辑器,其包含尿嘧啶DNA糖基化酶(UDG)、跨损伤聚合酶、Cas9域(例如nCas9)和胞苷脱氨酶。右侧是显示C至G碱基编辑器的示意图,该编辑器包含跨损伤聚合酶、Cas9域(例如nCas9)和碱基切除酶(例如能够切割C或T残基的UDG变体)。

[0049] 图41显示了使用Pol Kappa或Pol Iota拴系构建体,在HEK2、RNF2和FANCF位点处使用图40的左图中所示的碱基编辑器(包含尿嘧啶DNA糖基化酶(UDG)、跨损伤聚合酶、Cas9域和胞苷脱氨酶的碱基编辑器)的C至G碱基编辑。C至G的编辑对于HEK2和RNF2通过点柱形(G)移到实心柱形(C)以及对于FANCF通过点柱形(G)移到实心柱形(G)以图形显示。

[0050] 图42显示了使用与Pol Kappa、Pol Eta、Pol Iota和REV1拴系的碱基编辑器在WT细胞中的HEK2位点处的碱基编辑,所述碱基编辑器在图40的右图中显示(含有跨损伤聚合酶、Cas9域和切割T的碱基切除酶(UDG 147)的碱基编辑器)。C至G的量在HEK2位点中的特定残基处图示。UDG 147是直接除去T的UDG变体。

[0051] 图43显示了使用与Pol Kappa,Pol Eta,Pol Iota和REV1拴系的碱基编辑器在WT

细胞中的RNF2位点处的碱基编辑，所述碱基编辑器在图40的右图中显示(含有跨损伤聚合酶、Cas9域和切割T的碱基切除酶(UDG 147)的碱基编辑器)。C至G的量在HEK2位点中的特定残基处图示。UDG 147是直接除去T的UDG变体。

[0052] 图44显示了使用与Pol Kappa,Pol Eta,Pol Iota和REV1拴系的碱基编辑器在WT细胞中的FANCF位点处的碱基编辑，所述碱基编辑器在图40的右图中显示(含有跨损伤聚合酶、Cas9域和切割T的碱基切除酶(UDG 147)的碱基编辑器)。C至G的量在HEK2位点中的特定残基处图示。UDG 147是直接除去T的UDG变体。

[0053] 图45显示了使用与Pol Kappa,Pol Eta,Pol Iota和REV1拴系的碱基编辑器在WT细胞中的HEK2位点处的碱基编辑，所述碱基编辑器在图40的右图中显示(含有跨损伤聚合酶、Cas9域和切割C的碱基切除酶(UDG 204)的碱基编辑器)。C至G的量在HEK2位点中的特定残基处图示。UDG 204是直接除去C的UDG变体。

[0054] 图46显示了使用与Pol Kappa,Pol Eta,Pol Iota和REV1拴系的碱基编辑器在WT细胞中的RNF2位点处的碱基编辑，其在图40的右图中显示(含有跨损伤聚合酶、Cas9域和切割C的碱基切除酶(UDG 204)的碱基编辑器)。C至G的量在HEK2位点中的特定残基处图示。UDG 204是直接除去C的UDG变体。

[0055] 图47显示了使用与Pol Kappa,Pol Eta,Pol Iota和REV1拴系的碱基编辑器在WT细胞中的FANCF位点处的碱基编辑，其在图40的右图中显示(含有跨损伤聚合酶、Cas9域和切割C的碱基切除酶(UDG 204)的碱基编辑器)。C至G的量在HEK2位点中的特定残基处图示。UDG 204是直接除去C的UDG变体。

[0056] 图48显示了示意图，其显示了MSH2在碱基修复中的作用，其中MSH2可以促进DNA中尿嘧啶(U)向胞嘧啶(C)的转化。

[0057] 图49显示了使用六个碱基编辑器(BE3;BE3_UdgX;BE3_UdgX*;BE2_UdgX_On;BE3_UdgX_On;和BE3_UDG)在MSH2-/-细胞中HEK2位点处的碱基编辑。原始编辑值显示在左图中。右边的图显示了原始编辑值的图示，其中C至G的碱基编辑通过点柱形(G)移至实心柱形(C)以图形显示。

[0058] 图50显示了使用六个碱基编辑器(BE3;BE3_UdgX;BE3_UdgX*;BE2_UdgX_On;BE3_UdgX_On;和BE3_UDG)在MSH2-/-细胞中RNF2位点处的碱基编辑。原始编辑值显示在左图中。右边的图显示了原始编辑值的图示，其中C至G的碱基编辑通过点柱形(G)移至实心柱形(C)以图形显示。

[0059] 图51显示了使用六个碱基编辑器(BE3;BE3_UdgX;BE3_UdgX*;BE2_UdgX_On;BE3_UdgX_On;和BE3_UNG)在MSH2-/-细胞中的FANCF位点处的碱基编辑。原始编辑值显示在左图中。右侧的图显示了原始编辑值的图示，其中C至G的碱基编辑通过通过实心柱形(C)移至点柱形(G)以图形显示。

[0060] 图52显示了示意图，其显示了碱基编辑方法，其中含有UDG(或UDG变体)、Cas9(例如nCas9)域和胞苷脱氨酶的C至G碱基编辑器与跨损伤聚合酶一起反式表达。

[0061] 图53显示了使用与各种聚合酶(Pol Kappa,Pol Eta,Pol Iota,REV1,Pol Beta,和Pol Delta)一起反式表达的五个碱基编辑器(BE3;BE3_UdgX;BE3_UdgX*;BE2_UdgX_On;和BE3_UDG)在HEK293细胞中HEK2位点处的碱基编辑。C至G的碱基编辑用点柱形(G)移至实心柱形(C)以图形显示。

[0062] 图54显示了使用与各种聚合酶(Pol Kappa,Pol Eta,Pol Iota,REV1,Pol Beta,和Pol Delta)一起反式表达的五个碱基编辑器(BE3;BE3_UdgX;BE3_UdgX*;BE2_UdgX_On;和BE3_UDG)在HEK293细胞中RNF2位点处的碱基编辑。C至G的碱基编辑用点柱形(G)移至实心柱形(C)以图形显示。

[0063] 图55显示了使用与各种聚合酶(Pol Kappa,Pol Eta,Pol Iota,REV1,Pol Beta,和Pol Delta)一起反式表达的五个碱基编辑器(BE3;BE3_UdgX;BE3_UdgX*;BE2_UdgX_On;和BE3_UDG)在HEK293细胞中FANCF位点处的碱基编辑。C至G的碱基编辑用点柱形(G)移至实心柱形(C)以图形显示。

[0064] 定义

[0065] 如本文和权利要求书中所用,除非上下文另外明确指出,否则单数形式“一种”、“一个”和“该/所述”包括单数和复数。因此,例如,提及“试剂”包括单一试剂和多个此类试剂。

[0066] 如本文所用,术语“脱氨酶”或“脱氨酶域”是指催化脱氨基反应的蛋白质或酶。在一些实施方案中,脱氨酶或脱氨酶域是胞苷脱氨酶,其分别催化胞苷或脱氧胞苷水解为尿苷或脱氧尿苷。在一些实施方案中,脱氨酶或脱氨酶域是胞苷脱氨酶域,其催化胞嘧啶水解脱氨基为尿嘧啶。在一些实施方案中,脱氨酶或脱氨酶域是来自生物体,例如人,黑猩猩,大猩猩,猴,牛,狗,大鼠或小鼠的天然存在的脱氨酶。在一些实施方案中,脱氨酶或脱氨酶域是自然界中不存在的来自生物体的天然存在的脱氨酶的变体。例如,在一些实施方案中,脱氨酶或脱氨酶域与来自生物体的天然存在的脱氨酶是至少50%,至少55%,至少60%,至少65%,至少70%,至少75%,至少80%,至少85%,至少90%,至少95%,至少96%,至少97%,至少98%,至少99%或至少99.5%相同的。

[0067] 术语“碱基编辑器(BE)”或“核碱基编辑器(NBE)”是指包含能够对核酸序列(例如DNA或RNA)内碱基(例如,A,T,C,G或U)进行修饰的多肽的试剂。在一些实施方案中,碱基编辑器能够使核酸内的碱基脱氨基。在一些实施方案中,碱基编辑器能够使DNA分子内的碱基脱氨基。在一些实施方案中,碱基编辑器能够使DNA中的胞嘧啶(C)脱氨基。在一些实施方案中,碱基编辑器能够切割DNA分子内的碱基。在一些实施方案中,碱基编辑器能够切割核酸(例如DNA或RNA)分子内的腺嘌呤,鸟嘌呤,胞嘧啶,胸腺嘧啶或尿嘧啶。在一些实施方案中,碱基编辑器是蛋白质(例如,融合蛋白),其包含与胞苷脱氨酶融合的核酸可编程DNA结合蛋白(napDNAbp)。在一些实施方案中,碱基编辑物与尿嘧啶结合蛋白(UBP),例如尿嘧啶DNA糖基化酶(UDG)融合。在一些实施方案中,碱基编辑物与核酸聚合酶(NAP)域融合。在一些实施方案中,NAP域是跨损伤DNA聚合酶。在一些实施方案中,碱基编辑器包含napDNAbp,胞苷脱氨酶和UBP(例如UDG)。在一些实施方案中,碱基编辑物包含napDNAbp,胞苷脱氨酶和核酸聚合酶(例如,跨损伤DNA聚合酶)。在一些实施方案中,碱基编辑物包含napDNAbp,胞苷脱氨酶,UBP(例如,UDG),以及核酸聚合酶(例如,跨损伤DNA聚合酶)。

[0068] 在一些实施方案中,碱基编辑器的napDNAbp是Cas9域。在一些实施方案中,碱基编辑器包含与胞苷脱氨酶融合的Cas9蛋白。在一些实施方案中,碱基编辑器包含与胞苷脱氨酶融合的Cas9切口酶(nCas9)。在一些实施方案中,Cas9切口酶包含D10A突变并包含在SEQ ID NO:6的残基840处的组氨酸,或在本文提供的任何Cas9,例如SEQ ID NO:4-26中的任何一项中的相应突变,其使Cas9仅能切割核酸双链体的一条链。在一些实施方案中,碱基编辑

器包含与胞苷脱氨酶融合的无核酸酶活性的Cas9(dCas9)。在一些实施方案中,dCas9域包含SEQ ID NO:6的D10A和H840A突变,或本文提供的任何Cas9(例如SEQ ID NO:4-26中的任何一项)中的相应突变,其使Cas9蛋白的核酸酶活性失活。

[0069] 如本文所用,术语"接头"是指连接两个分子或部分,例如融合蛋白的两个域,诸如例如无核酸酶活性的Cas9域和核酸编辑域(例如胞苷脱氨酶域)的键(例如共价键)、化学基团或分子。在一些实施方案中,接头接合RNA可编程核酸酶的gRNA结合域,包括Cas9核酸酶域和核酸编辑蛋白的催化域。在一些实施方案中,接头接合dCas9和核酸编辑蛋白。通常,接头位于两个基团、分子或其他部分之间或侧翼有两个基团、分子或其他部分,并且经由共价键与每一个连接,从而连接两者。在一些实施方案中,接头是一个氨基酸或多个氨基酸(例如肽或蛋白质)。在一些实施方案中,接头是有机分子、基团、聚合物或化学部分。在一些实施方案中,接头的长度为5-100个氨基酸,例如长度为5、6、7、8、9、10、11、12、13、14、15、16、17、18、19、20、21、22、23、24、25、26、27、28、29、30、30-35、35-40、40-45、45-50、50-60、60-70、70-80、80-90、90-100、100-150或150-200个氨基酸。也考虑了更长或更短的接头。在一些实施方案中,接头包含氨基酸序列SGSETPGTSESATPES(SEQ ID NO:10),其也可以称为XTEN接头。在一些实施方案中,接头包含氨基酸序列SGGS(SEQ ID NO:37)。在一些实施方案中,接头包含$(SGGS)_n$(SEQ ID NO:37)、$(GGGS)_n$(SEQ ID NO:38)、$(GGGGS)_n$(SEQ ID NO:39)、$(G)_n$、$(EAAAK)_n$(SEQ ID NO:40)、$(GGS)_n$、SGSETPGTSESATPES(SEQ ID NO:10)或$(XP)_n$基序,或这些中任何的组合,其中n独立地是1和30之间的整数,并且其中X是任何氨基酸。在一些实施方案中,n是1、2、3、4、5、6、7、8、9、10、11、12、13、14或15。

[0070] 如本文所用,术语"突变"是指序列(例如核酸或氨基酸序列)内的残基用另一个残基取代或序列内一个或多个残基的缺失或插入。本文通常通过鉴定初始残基,随后是序列内残基的位置和新取代的残基的身份来描述突变。用于产生本文提供的氨基酸取代(突变)的各种方法在本领域中是熟知的,并且由例如Green and Sambrook,Molecular Cloning:A Laboratory Manual(第4版,Cold Spring Harbor Laboratory Press,Cold Spring Harbor,N.Y.(2012))提供。

[0071] 如本文所用,术语"尿嘧啶结合蛋白"或"UBP"是指能够结合尿嘧啶的蛋白质。在一些实施方案中,尿嘧啶结合蛋白是尿嘧啶修饰酶。在一些实施方案中,尿嘧啶结合蛋白是尿嘧啶碱基切除酶。在一些实施方案中,尿嘧啶结合蛋白是尿嘧啶DNA糖基化酶(UDG)。在一些实施方案中,尿嘧啶结合蛋白以野生型UDG(例如人UDG)结合尿嘧啶的亲和力的至少1%,2%,3%,5%,10%,15%,20%,30%,40%,50%,60%,70%,80%,90%或至少95%的亲和力结合尿嘧啶。

[0072] 如本文所用,术语"碱基切除酶"或"BEE"是指能够从核酸分子(例如DNA或RNA)中除去碱基(例如,A,T,C,G或U)的蛋白质。在一些实施方案中,BEE能够从DNA除去胞嘧啶。在一些实施方案中,BEE能够从DNA除去胸腺嘧啶。示例性的BEE包括但不限于UDG Tyr147Ala和UDG Asn204Asp,如Sang et al.,"A Unique Uracil-DNA binding protein of the uracil DNA glycosylase superfamily,"Nucleic Acids Research,Vol.43,No.17 2015所述;其全部内容通过引用并入本文。

[0073] 术语"核酸聚合酶"或"NAP"是指从核苷酸(例如脱氧核糖核苷酸和核糖核苷酸)合成核酸分子(例如DNA和RNA)的酶。在一些实施方案中,NAP是DNA聚合酶。在一些实施方案

中,NAP是跨损伤聚合酶。跨损伤聚合酶在诱变中起作用,例如,通过重启复制叉或填补由于DNA损伤的存在而保留在基因组中的缺口。示例性的跨损伤聚合酶包括但不限于Pol Beta,Pol Lambda,Pol Eta,Pol Mu,Pol Iota,Pol Kappa,Pol Alpha,Pol Delta,Pol Gamma,和Pol Nu。

[0074] 术语“核定位序列”或“NLS”是指促进蛋白质输入细胞核(例如通过核转运)的氨基酸序列。核定位序列是本领域已知的,并且对于熟练技术人员是显而易见的。在一些实施方案中,NLS是单分NLS。在一些实施方案中,NLS是二分NLS。二分NLS由相对较短的间隔物序列隔开(例如,2-20个氨基酸,5-15个氨基酸或8-12个氨基酸)。例如,NLS序列描述于2001年11月23日提交的Plank等人的国际PCT申请PCT/EP2000/011690,2001年5月31日公布为WO/2001/038547,和Kethar,K.M.V.,et al.,“Applicationof bioinformatics-coupled experimental analysis reveals a new transport-competent nuclear localization signal in the nucleoptotein of Influenza A virus strain”BMC Cell Biol,2008,9:22;每篇的内容通过引用并入本文,用于其对示例性的核定位序列的公开内容。在一些实施方案中,NLS包含氨基酸序列PKKKRKV(SEQ ID NO:41),MDSLLMNRRKFLYQFKNVRWAKGRRETYLC(SEQ ID NO:42),KRTADGSEFESPKKKRKV(SEQ ID NO:43),KRGINDRNFWRGENGRKTR(SEQ ID NO:44),KKTGGPIYRRVDGKWRR(SEQ ID NO:45),RRELILYDKEEIRRIWR(SEQ ID NO:46),或AVSRKRKA(SEQ ID NO:47)。

[0075] 术语“核酸可编程DNA结合蛋白”或“napDNAbp”是指与核酸(例如DNA或RNA),例如引导核酸结合的蛋白质,所述核酸将napDNAbp引导至特定核酸序列。例如,Cas9蛋白可以与引导RNA结合,所述引导RNA将Cas9蛋白引导至与引导RNA互补的特定DNA序列。在一些实施方案中,napDNAbp是2类微生物CRISPR-Cas效应物。在一些实施方案中,napDNAbp是Cas9域,例如核酸酶活性Cas9、Cas9切口酶(nCas9)或无核酸酶活性Cas9(dCas9)。核酸可编程DNA结合蛋白的实例包括但不限于Cas9(例如dCas9和nCas9)、CasX、CasY、Cpf1、C2c1、C2c2、C2C3和Argonaute。然而,应当理解,核酸可编程DNA结合蛋白还包括结合RNA的核酸可编程蛋白。例如,napDNAbp可以与将napDNAbp引导至RNA的核酸结合。其他核酸可编程DNA结合蛋白也在本公开内容的范围内,但它们可以未在本公开内容中具体列出。

[0076] 术语“Cas9”或“Cas9域”是指包含Cas9蛋白或其片段的RNA引导的核酸酶(例如,包含Cas9的活性、无活性或部分活性的DNA切割域,和/或Cas9的gRNA结合域的蛋白质)。Cas9核酸酶有时也称为casn1核酸酶或CRISPR(聚簇规则间隔短回文重复)相关核酸酶。CRISPR是适应性免疫系统,其提供针对移动遗传元件(病毒、可转座元件和接合质粒)的保护。CRISPR簇含有间隔区,与先前的移动元件互补的序列,并靶向侵入核酸。CRISPR簇得以转录并加工成CRISPR RNA(crRNA)。在II型CRISPR系统中,对pre-crRNA的正确加工需要反式编码的小RNA(tracrRNA)、内源性核糖核酸酶3(rnc)和Cas9蛋白。tracrRNA充当用于pre-crRNA的核糖核酸酶3辅助加工的引导。随后,Cas9/crRNA/tracrRNA以内切核水解方式切割与间隔区互补的线性或环状dsDNA靶标。首先以内切核水解方式切割不与crRNA互补的靶链,然后以3’-5’外切核水解方式修剪(trim)。在自然界中,DNA结合和切割通常需要蛋白质和这两种RNA。然而,单一引导RNA(“sgRNA”或简称“gNRA”)可以经工程化以将crRNA和tracrRNA两者的方面并入单一RNA种类中。参见例如Jinek M.,Chylinski K.,Fonfara I.,Hauer M.,Doudna J.A.,Charpentier E.Science337:816-821(2012),其全部内容通过引

用并入本文。Cas9识别CRISPR重复序列中的短基序(PAM或前间隔区相邻基序),以帮助区分自我与非自我。Cas9核酸酶序列和结构是本领域技术人员熟知的(参见例如“Complete genome sequence of an M1 strain of Streptococcus pyogenes.”Ferretti et al., J.J.,McShan W.M.,Ajdic D.J.,Savic D.J.,Savic G.,Lyon K.,Primeaux C.,Sezate S.,Suvorov A.N.,Kenton S.,Lai H.S.,Lin S.P.,Qian Y.,Jia H.G.,Najar F.Z.,Ren Q.,Zhu H.,Song L.,White J.,Yuan X.,Clifton S.W.,Roe B.A.,McLaughlin R.E., Proc.Natl.Acad.Sci.U.S.A.98:4658-4663(2001);“CRISPR RNA maturation by trans-encoded small RNA and host factor RNase III.”Deltcheva E.,Chylinski K.,Sharma C.M.,Gonzales K.,Chao Y.,Pirzada Z.A.,Eckert M.R.,Vogel J.,Charpentier E., Nature 471:602-607(2011);和“A programmable dual-RNA-guided DNA endonuclease in adaptive bacterial immunity.”Jinek M.,Chylinski K.,Fonfara I.,Hauer M., Doudna J.A.,Charpentier E.Science337:816-821(2012),其各自的全部内容通过引用并入本文)。已经在各种物种中描述了Cas9直系同源物,包括但不限于酿脓链球菌(S.pyogenes)和嗜热链球菌(S.thermophilus)。基于本公开内容,其他合适的Cas9核酸酶和序列对于本领域技术人员将是显而易见的,并且此类Cas9核酸酶和序列包括来自Chylinski,Rhun,and Charpentier,“The tracrRNA and Cas9 families of type II CRISPR-Cas immunity systems”(2013)RNA Biology 10:5,726-737中公开的生物体和基因座的Cas9序列;其全部内容通过引用并入本文。在一些实施方案中,Cas9核酸酶具有无活性的(例如失活的)DNA切割域,也就是说,Cas9是切口的酶。

[0077] 核酸酶失活的Cas9蛋白可以互换地称为“dCas9”蛋白(代表核酸酶-“死亡的”Cas9)。用于生成具有无活性的DNA切割域的Cas9蛋白(或其片段)的方法是已知的(参见例如Jinek et al.,Science.337:816-821(2012);Qi et al.,“Repurposing CRISPR as an RNA-Guided Platform for Sequence-Specific Control of Gene Expression”(2013) Cell.28;152(5):1173-83,其各自的全部内容通过引用并入本文)。例如,已知Cas9的DNA切割域包括两个亚域,即HNH核酸酶亚域和RuvC1亚域。HNH亚域切割与gRNA互补的链,而RuvC1亚域切割非互补链。这些亚域内的突变可以沉默Cas9的核酸酶活性。例如,突变D10A和H840A完全使酿脓链球菌Cas9的核酸酶活性失活(Jinek et al.,Science.337:816-821 (2012);Qi et al.,Cell.28;152(5):1173-83(2013))。在一些实施方案中,提供了包含Cas9的片段的蛋白质。例如,在一些实施方案中,蛋白质包含两个Cas9域的一个:(1)Cas9的gRNA结合域;或(2)Cas9的DNA切割域。在一些实施方案中,包含Cas9或其片段的蛋白质称为“Cas9变体”。Cas9变体与Cas9或其片段共享同源性。例如,Cas9变体与野生型Cas9至少约70%相同、至少约80%相同、至少约90%相同、至少约95%相同、至少约96%相同、至少约97%相同、至少约98%相同、至少约99%相同、至少约99.5%相同或至少约99.9%相同。在一些实施方案中,Cas9变体与野生型Cas9相比,可以具有1、2、3、4、5、6、7、8、9、10、11、12、13、14、15、16、17、18、19、20、21、22、21、24、25、26、27、28、29、30、31、32、33、34、35、36、37、38、39、40、41、42、43、44、45、46、47、48、49、50个或更多个氨基酸变化。在一些实施方案中,Cas9变体包含Cas9的片段(例如,gRNA结合域或DNA切割域),使得该片段与野生型Cas9的相应的片段至少约70%相同、至少约80%相同、至少约90%相同、至少约95%相同、至少约96%相同、至少约97%相同、至少约98%相同、至少约99%相同、至少约99.5%相同或至少

约99.9％相同。在一些实施方案中，片段是相应的野生型Cas9的氨基酸长度的至少30％、至少35％、至少40％、至少45％、至少50％、至少55％、至少60％、至少65％、至少70％、至少75％、至少80％、至少85％、至少90％、至少95％相同、至少96％、至少97％、至少98％、至少99％或至少99.5％。

[0078] 在一些实施方案中，片段的长度为至少100个氨基酸。在一些实施方案中，片段的长度为至少100、150、200、250、300、350、400、450、500、550、600、650、700、750、800、850、900、950、1000、1050、1100、1150、1200、1250或1300个氨基酸。在一些实施方案中，野生型Cas9对应与来自酿脓链球菌(Streptococcus pyogenes)的Cas9(NCBI参考序列:NC_017053.1,SEQ ID NO:1(核苷酸);SEQ ID NO:4(氨基酸))。

[0079] ATGGATAAGAAATACTCAATAGGCTTAGATATCGGCACAAATAGCGTCGGATGGGCGGTGATCACTGA
TGATTATAAGGTTCCGTCTAAAAAGTTCAAGGTTCTGGGAAATACAGACCGCCACAGTATCAAAAAAAATCTTATA
GGGGCTCTTTTATTTGGCAGTGGAGAGACAGCGGAAGCGACTCGTCTCAAACGGACAGCTCGTAGAAGGTATACAC
GTCGGAAGAATCGTATTTGTTATCTACAGGAGATTTTTTCAAATGAGATGGCGAAAGTAGATGATAGTTTCTTTCA
TCGACTTGAAGAGTCTTTTTTGGTGGAAGAAGACAAGAAGCATGAACGTCATCCTATTTTTGGAAATATAGTAGAT
GAAGTTGCTTATCATGAGAAATATCCAACTATCTATCATCTGCGAAAAAAATTGGCAGATTCTACTGATAAAGCGG
ATTTGCGCTTAATCTATTTGGCCTTAGCGCATATGATTAAGTTTCGTGGTCATTTTTTGATTGAGGGAGATTTAAA
TCCTGATAATAGTGATGTGGACAAACTATTTATCCAGTTGGTACAAATCTACAATCAATTATTTGAAGAAAACCCT
ATTAACGCAAGTAGAGTAGATGCTAAAGCGATTCTTTCTGCACGATTGAGTAAATCAAGACGATTAGAAAATCTCA
TTGCTCAGCTCCCCGGTGAGAAGAGAAATGGCTTGTTTGGGAATCTCATTGCTTTGTCATTGGGATTGACCCCTAA
TTTTAAATCAAATTTTGATTTGGCAGAAGATGCTAAATTACAGCTTTCAAAAGATACTTACGATGATGATTTAGAT
AATTTATTGGCGCAAATTGGAGATCAATATGCTGATTTGTTTTTGGCAGCTAAGAATTTATCAGATGCTATTTTAC
TTTCAGATATCCTAAGAGTAAATAGTGAAATAACTAAGGCTCCCCTATCAGCTTCAATGATTAAGCGCTACGATGA
ACATCATCAAGACTTGACTCTTTTAAAAGCTTTAGTTCGACAACAACTTCCAGAAAAGTATAAAGAAATCTTTTTT
GATCAATCAAAAAACGGATATGCAGGTTATATTGATGGGGGAGCTAGCCAAGAAGAATTTTATAAATTTATCAAAC
CAATTTTAGAAAAAATGGATGGTACTGAGGAATTATTGGTGAAACTAAATCGTGAAGATTTGCTGCGCAAGCAACG
GACCTTTGACAACGGCTCTATTCCCCATCAAATTCACTTGGGTGAGCTGCATGCTATTTTGAGAAGACAAGAAGAC
TTTTATCCATTTTTAAAAGACAATCGTGAGAAGATTGAAAAAATCTTGACTTTTCGAATTCCTTATTATGTTGGTC
CATTGGCGCGTGGCAATAGTCGTTTTGCATGGATGACTCGGAAGTCTGAAGAAACAATTACCCCATGGAATTTTGA
AGAAGTTGTCGATAAAGGTGCTTCAGCTCAATCATTTATTGAACGCATGACAAACTTTGATAAAAATCTTCCAAAT
GAAAAAGTACTACCAAAACATAGTTTGCTTTATGAGTATTTTACGGTTTATAACGAATTGACAAAGGTCAAATATG
TTACTGAGGGAATGCGAAAACCAGCATTTCTTTCAGGTGAACAGAAGAAAGCCATTGTTGATTTACTCTTCAAAAC
AAATCGAAAAGTAACCGTTAAGCAATTAAAAGAAGATTATTTCAAAAAAATAGAATGTTTTGATAGTGTTGAAATT
TCAGGAGTTGAAGATAGATTTAATGCTTCATTAGGCGCCTACCATGATTTGCTAAAAATTATTAAAGATAAAGATT
TTTTGGATAATGAAGAAAATGAAGATATCTTAGAGGATATTGTTTTAACATTGACCTTATTTGAAGATAGGGGGAT
GATTGAGGAAAGACTTAAAACATATGCTCACCTCTTTGATGATAAGGTGATGAAACAGCTTAAACGTCGCCGTTAT
ACTGGTTGGGGACGTTTGTCTCGAAAATTGATTAATGGTATTAGGGATAAGCAATCTGGCAAAACAATATTAGATT
TTTTGAAATCAGATGGTTTTGCCAATCGCAATTTTATGCAGCTGATCCATGATGATAGTTTGACATTTAAAGAAGA
TATTCAAAAAGCACAGGTGTCTGGACAAGGCCATAGTTTACATGAACAGATTGCTAACTTAGCTGGCAGTCCTGCT
ATTAAAAAAGGTATTTTACAGACTGTAAAAATTGTTGATGAACTGGTCAAAGTAATGGGGCATAAGCCAGAAAATA

TCGTTATTGAAATGGCACGTGAAAATCAGACAACTCAAAAGGGCCAGAAAAATTCGCGAGAGCGTATGAAACGAAT
CGAAGAAGGTATCAAAGAATTAGGAAGTCAGATTCTTAAAGAGCATCCTGTTGAAAATACTCAATTGCAAAATGAA
AAGCTCTATCTCTATTATCTACAAAATGGAAGAGACATGTATGTGGACCAAGAATTAGATATTAATCGTTTAAGTG
ATTATGATGTCGATCACATTGTTCCACAAAGTTTCATTAAAGACGATTCAATAGACAATAAGGTACTAACGCGTTC
TGATAAAAATCGTGGTAAATCGGATAACGTTCCAAGTGAAGAAGTAGTCAAAAAGATGAAAAACTATTGGAGACAA
CTTCTAAACGCCAAGTTAATCACTCAACGTAAGTTTGATAATTTAACGAAAGCTGAACGTGGAGGTTTGAGTGAAC
TTGATAAAGCTGGTTTTATCAAACGCCAATTGGTTGAAACTCGCCAAATCACTAAGCATGTGGCACAAATTTTGGA
TAGTCGCATGAATACTAAATACGATGAAAATGATAAACTTATTCGAGAGGTTAAAGTGATTACCTTAAAATCTAAA
TTAGTTTCTGACTTCCGAAAAGATTTCCAATTCTATAAAGTACGTGAGATTAACAATTACCATCATGCCCATGATG
CGTATCTAAATGCCGTCGTTGGAACTGCTTTGATTAAGAAATATCCAAAACTTGAATCGGAGTTTGTCTATGGTGA
TTATAAAGTTTATGATGTTCGTAAAATGATTGCTAAGTCTGAGCAAGAAATAGGCAAAGCAACCGCAAAATATTTC
TTTTACTCTAATATCATGAACTTCTTCAAAACAGAAATTACACTTGCAAATGGAGAGATTCGCAAACGCCCTCTAA
TCGAAACTAATGGGGAAACTGGAGAAATTGTCTGGGATAAAGGGCGAGATTTTGCCACAGTGCGCAAAGTATTGTC
CATGCCCCAAGTCAATATTGTCAAGAAAACAGAAGTACAGACAGGCGGATTCTCCAAGGAGTCAATTTTACCAAAA
AGAAATTCGGACAAGCTTATTGCTCGTAAAAAAGACTGGGATCCAAAAAAATATGGTGGTTTTGATAGTCCAACGG
TAGCTTATTCAGTCCTAGTGGTTGCTAAGGTGGAAAAAGGGAAATCGAAGAAGTTAAAATCCGTTAAAGAGTTACT
AGGGATCACAATTATGGAAAGAAGTTCCTTTGAAAAAAATCCGATTGACTTTTTAGAAGCTAAAGGATATAAGGAA
GTTAAAAAAGACTTAATCATTAAACTACCTAAATATAGTCTTTTTGAGTTAGAAAACGGTCGTAAACGGATGCTGG
CTAGTGCCGGAGAATTACAAAAAGGAAATGAGCTGGCTCTGCCAAGCAAATATGTGAATTTTTTATATTTAGCTAG
TCATTATGAAAAGTTGAAGGGTAGTCCAGAAGATAACGAACAAAAACAATTGTTTGTGGAGCAGCATAAGCATTAT
TTAGATGAGATTATTGAGCAAATCAGTGAATTTTCTAAGCGTGTTATTTTAGCAGATGCCAATTTAGATAAAGTTC
TTAGTGCATATAACAAACATAGAGACAAACCAATACGTGAACAAGCAGAAAATATTATTCATTTATTTACGTTGAC
GAATCTTGGAGCTCCCGCTGCTTTTAAATATTTTGATACAACAATTGATCGTAAACGATATACGTCTACAAAAGAA
GTTTTAGATGCCACTCTTATCCATCAATCCATCACTGGTCTTTATGAAACACGCATTGATTTGAGTCAGCTAGGAG
GTGACTGA(SEQ ID NO:1)

[0080]

MDKKYSIGLDIGTNSVGWAVITDDYKVPSKKFKVLGNTDRHSIKKNLIGALLFGSGETAEATRLKRT
ARRRYTRRKNRICYLQEIFSNEMAKVDDSFFHRLEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTI
YHLRKKLADSTDKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQIYNQLFEENPINA
SRVDAKAILSARLSKSRRLENLIAQLPGEKRNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTY
DDDLDNLLAQIGDQYADLFLAAKNLSDAILLSDILRVNSEITKAPLSASMIKRYDEHHQDLTLLKAL
VRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLRKQRTF
DNGSIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPYYVGPLARGNSRFAWMTRKSEETITP
WNFEEVVDKGASAQSFIERMTNFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYVTEGMRKPAFL
SGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGAYHDLLKIIKDKDFL
DNEENEDILEDIVLTLTLFEDRGMIEERLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQ
SGKTILDFLKSDGFANRNFMQLIHDDSLTFKEDIQKAQVSGQGHSLHEQIANLAGSPAIKKGILQTVK
IVDELVKVMGHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHPVENTQLQNEK
LYLYYLQNGRDMYVDQELDINRLSDYDVDHIVPQSFIKDDSIDNKVLTRSDKNRGKSDNVPSEEVV
KKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKAGFIKRQLVETRQITKHVAQILDSRMNT
KYDENDKLIREVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYLNAVVGTALIKKYPKLESE
FVYGDYKVYDVRKMIAKSEQEIGKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWD
KGRDFATVRKVLSMPQVNIVKKTEVQTGGFSKESILPKRNSDKLIARKKDWDPKKYGGFDSPTVAY
SVLVVAKVEKGKSKKLKSVKELLGITIMERSSFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFELENG
RKRMLASAGELQKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQHKHYLDEIIEQISEF

SKRVILADANLDKVLSAYNKHRDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVLD

ATLIHQSITGLYETRIDLSQLGGD (SEQ ID NO:4)(单下划线:HNH域;双下划线:RuvC域)

[0081] 在一些实施方案中,野生型Cas9对应于或包含SEQ ID NO:2(核苷酸)和/或SEQ ID NO:5(氨基酸):

[0082] ATGGATAAAAAGTATTCTATTGGTTTAGACATCGGCACTAATTCCGTTGGATGGGCTGTCATAACCGA
TGAATACAAAGTACCTTCAAAGAAATTTAAGGTGTTGGGGAACACAGACCGTCATTCGATTAAAAAGAATCTTATC
GGTGCCCTCCTATTCGATAGTGGCGAAACGGCAGAGGCGACTCGCCTGAAACGAACCGCTCGGAGAAGGTATACAC
GTCGCAAGAACCGAATATGTTACTTACAAGAAATTTTTAGCAATGAGATGGCCAAAGTTGACGATTCTTTCTTTCA
CCGTTTGGAAGAGTCCTTCCTTGTCGAAGAGGACAAGAAACATGAACGGCACCCCATCTTTGGAAACATAGTAGAT
GAGGTGGCATATCATGAAAAGTACCCAACGATTTATCACCTCAGAAAAAAGCTAGTTGACTCAACTGATAAAGCGG
ACCTGAGGTTAATCTACTTGGCTCTTGCCCATATGATAAAGTTCCGTGGGCACTTTCTCATTGAGGGTGATCTAAA
TCCGGACAACTCGGATGTCGACAAACTGTTCATCCAGTTAGTACAAACCTATAATCAGTTGTTTGAAGAGAACCCT
ATAAATGCAAGTGGCGTGGATGCGAAGGCTATTCTTAGCGCCCGCCTCTCTAAATCCCGACGGCTAGAAAACCTGA
TCGCACAATTACCCGGAGAGAAGAAAAATGGGTTGTTCGGTAACCTTATAGCGCTCTCACTAGGCCTGACACCAAA
TTTTAAGTCGAACTTCGACTTAGCTGAAGATGCCAAATTGCAGCTTAGTAAGGACACGTACGATGACGATCTCGAC
AATCTACTGGCACAAATTGGAGATCAGTATGCGGACTTATTTTTGGCTGCCAAAAACCTTAGCGATGCAATCCTCC
TATCTGACATACTGAGAGTTAATACTGAGATTACCAAGGCGCCGTTATCCGCTTCAATGATCAAAAGGTACGATGA
ACATCACCAAGACTTGACACTTCTCAAGGCCCTAGTCCGTCAGCAACTGCCTGAGAAATATAAGGAAATATTCTTT
GATCAGTCGAAAAACGGGTACGCAGGTTATATTGACGGCGGAGCGAGTCAAGAGGAATTCTACAAGTTTATCAAAC
CCATATTAGAGAAGATGGATGGGACGGAAGAGTTGCTTGTAAAACTCAATCGCGAAGATCTACTGCGAAAGCAGCG
GACTTTCGACAACGGTAGCATTCCACATCAAATCCACTTAGGCGAATTGCATGCTATACTTAGAAGGCAGGAGGAT
TTTTATCCGTTCCTCAAAGACAATCGTGAAAAGATTGAGAAAATCCTAACCTTTCGCATACCTTACTATGTGGGAC
CCCTGGCCCGAGGGAACTCTCGGTTCGCATGGATGACAAGAAAGTCCGAAGAAACGATTACTCCATGGAATTTTGA
GGAAGTTGTCGATAAAGGTGCGTCAGCTCAATCGTTCATCGAGAGGATGACCAACTTTGACAAGAATTTACCGAAC
GAAAAAGTATTGCCTAAGCACAGTTTACTTTACGAGTATTTCACAGTGTACAATGAACTCACGAAAGTTAAGTATG
TCACTGAGGGCATGCGTAAACCCGCCTTTCTAAGCGGAGAACAGAAGAAAGCAATAGTAGATCTGTTATTCAAGAC
CAACCGCAAAGTGACAGTTAAGCAATTGAAAGAGGACTACTTTAAGAAAATTGAATGCTTCGATTCTGTCGAGATC
TCCGGGGTAGAAGATCGATTTAATGCGTCACTTGGTACGTATCATGACCTCCTAAAGATAATTAAAGATAAGGACT
TCCTGGATAACGAAGAGAATGAAGATATCTTAGAAGATATAGTGTTGACTCTTACCCTCTTTGAAGATCGGGAAAT
GATTGAGGAAAGACTAAAAACATACGCTCACCTGTTCGACGATAAGGTTATGAAACAGTTAAAGAGGCGTCGCTAT
ACGGGCTGGGGACGATTGTCGCGGAAACTTATCAACGGGATAAGAGACAAGCAAAGTGGTAAAACTATTCTCGATT
TTCTAAAGAGCGACGGCTTCGCCAATAGGAACTTTATGCAGCTGATCCATGATGACTCTTTAACCTTCAAAGAGGA
TATACAAAAGGCACAGGTTTCCGGACAAGGGGACTCATTGCACGAACATATTGCGAATCTTGCTGGTTCGCCAGCC
ATCAAAAAGGGCATACTCCAGACAGTCAAAGTAGTGGATGAGCTAGTTAAGGTCATGGGACGTCACAAACCGGAAA
ACATTGTAATCGAGATGGCACGCGAAAATCAAACGACTCAGAAGGGGCAAAAAAACAGTCGAGAGCGGATGAAGAG
AATAGAAGAGGGTATTAAAGAACTGGGCAGCCAGATCTTAAAGGAGCATCCTGTGGAAAATACCCAATTGCAGAAC
GAGAAACTTTACCTCTATTACCTACAAAATGGAAGGGACATGTATGTTGATCAGGAACTGGACATAAACCGTTTAT
CTGATTACGACGTCGATCACATTGTACCCCAATCCTTTTTGAAGGACGATTCAATCGACAATAAAGTGCTTACACG
CTCGGATAAGAACCGAGGGAAAAGTGACAATGTTCCAAGCGAGGAAGTCGTAAAGAAAATGAAGAACTATTGGCGG

CAGCTCCTAAATGCGAAACTGATAACGCAAAGAAAGTTCGATAACTTAACTAAAGCTGAGAGGGGTGGCTTGTCTG
AACTTGACAAGGCCGGATTTATTAAACGTCAGCTCGTGGAAACCCGCCAAATCACAAAGCATGTTGCACAGATACT
AGATTCCCGAATGAATACGAAATACGACGAGAACGATAAGCTGATTCGGGAAGTCAAAGTAATCACTTTAAAGTCA
AAATTGGTGTCGGACTTCAGAAAGGATTTTCAATTCTATAAAGTTAGGGAGATAAATAACTACCACCATGCGCACG
ACGCTTATCTTAATGCCGTCGTAGGGACCGCACTCATTAAGAAATACCCGAAGCTAGAAAGTGAGTTTGTGTATGG
TGATTACAAAGTTTATGACGTCCGTAAGATGATCGCGAAAAGCGAACAGGAGATAGGCAAGGCTACAGCCAAATAC
TTCTTTTATTCTAACATTATGAATTTCTTTAAGACGGAAATCACTCTGGCAAACGGAGAGATACGCAAACGACCTT
TAATTGAAACCAATGGGGAGACAGGTGAAATCGTATGGGATAAGGGCCGGGACTTCGCGACGGTGAGAAAAGTTTT
GTCCATGCCCCAAGTCAACATAGTAAAGAAAACTGAGGTGCAGACCGGAGGGTTTTCAAAGGAATCGATTCTTCCA
AAAAGGAATAGTGATAAGCTCATCGCTCGTAAAAAGGACTGGGACCCGAAAAAGTACGGTGGCTTCGATAGCCCTA
CAGTTGCCTATTCTGTCCTAGTAGTGGCAAAAGTTGAGAAGGGAAAATCCAAGAAACTGAAGTCAGTCAAAGAATT
ATTGGGGATAACGATTATGGAGCGCTCGTCTTTTGAAAAGAACCCCATCGACTTCCTTGAGGCGAAAGGTTACAAG
GAAGTAAAAAAGGATCTCATAATTAAACTACCAAAGTATAGTCTGTTTGAGTTAGAAAATGGCCGAAAACGGATGT
TGGCTAGCGCCGGAGAGCTTCAAAAGGGGAACGAACTCGCACTACCGTCTAAATACGTGAATTTCCTGTATTTAGC
GTCCCATTACGAGAAGTTGAAAGGTTCACCTGAAGATAACGAACAGAAGCAACTTTTTGTTGAGCAGCACAAACAT
TATCTCGACGAAATCATAGAGCAAATTTCGGAATTCAGTAAGAGAGTCATCCTAGCTGATGCCAATCTGGACAAAG
TATTAAGCGCATACAACAAGCACAGGGATAAACCCATACGTGAGCAGGCGGAAAATATTATCCATTTGTTTACTCT
TACCAACCTCGGCGCTCCAGCCGCATTCAAGTATTTTGACACAACGATAGATCGCAAACGATACACTTCTACCAAG
GAGGTGCTAGACGCGACACTGATTCACCAATCCATCACGGGATTATATGAAACTCGGATAGATTTGTCACAGCTTG
GGGGTGACGGATCCCCCAAGAAGAAGAGGAAAGTCTCGAGCGACTACAAAGACCATGACGGTGATTATAAAGATCA
TGACATCGATTACAAGGATGACGATGACAAGGCTGCAGGA(SEQ ID NO:2)

[0083] MDKKYSIGLAIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETAEATRLKRT
ARRRYTRRKNRICYLQEIFSNEMAKVDDSFFHRLEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTI
YHLRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFEENPIN
ASGVDAKAILSARLSKSRRLENLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQLSKDT
YDDDLDNLLAQIGDQYADLFLAAKNLSDAILLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKA
LVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLRKQRT
FDNGSIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPYYVGPLARGNSRFAWMTRKSEETIT
PWNFEEVVDKGASAQSFIERMTNFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYVTEGMRKPAF
LSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKIIKDKDFL
DNEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQ
SGKTILDFLKSDGFANRNFMQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVK
VVDELVKVMGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHPVENTQLQN
EKLYLYYLQNGRDMYVDQELDINRLSDYDVDHIVPQSFLKDDSIDNKVLTRSDKNRGKSDNVPSEE
VVKKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKAGFIKRQLVETRQITKHVAQILDSRM
NTKYDENDKLIREVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYLNAVVGTALIKKYPKLE
SEFVYGDYKVYDVRKMIAKSEQEIGKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGEIV
WDKGRDFATVRKVLSMPQVNIVKKTEVQTGGFSKESILPKRNSDKLIARKKDWDPKKYGGFDSPT
VAYSVLVVAKVEKGKSKKLKSVKELLGITIMERSSFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFEL
ENGRKRMLASAGELQKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQHKHYLDEIIEQ
ISEFSKRVILADANLDKVLSAYNKHRDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKE
VLDATLIHQSITGLYETRIDLSQLGGD(SEQ ID NO:5)(单下划线:HNH域;双下划线:RuvC域)

[0084] 在一些实施方案中,野生型Cas9对应于来自酿脓链球菌的Cas9(NCBI参考序列:NC_002737.2,SEQ ID NO:3(核苷酸);和Uniport参考序列:Q99ZW2,SEQ ID NO:6(氨基酸)。

[0085] ATGGATAAGAAATACTCAATAGGCTTAGATATCGGCACAAATAGCGTCGGATGGGCGGTGATCACTGA

TGAATATAAGGTTCCGTCTAAAAAGTTCAAGGTTCTGGGAAATACAGACCGCCACAGTATCAAAAAAAATCTTATA
GGGGCTCTTTTATTTGACAGTGGAGAGACAGCGGAAGCGACTCGTCTCAAACGGACAGCTCGTAGAAGGTATACAC
GTCGGAAGAATCGTATTTGTTATCTACAGGAGATTTTTTCAAATGAGATGGCGAAAGTAGATGATAGTTTCTTTCA
TCGACTTGAAGAGTCTTTTTTGGTGGAAGAAGACAAGAAGCATGAACGTCATCCTATTTTTGGAAATATAGTAGAT
GAAGTTGCTTATCATGAGAAATATCCAACTATCTATCATCTGCGAAAAAAAATTGGTAGATTCTACTGATAAAGCGG
ATTTGCGCTTAATCTATTTGGCCTTAGCGCATATGATTAAGTTTCGTGGTCATTTTTTGATTGAGGGAGATTTAAA
TCCTGATAATAGTGATGTGGACAAACTATTTATCCAGTTGGTACAAACCTACAATCAATTATTTGAAGAAAACCCT
ATTAACGCAAGTGGAGTAGATGCTAAAGCGATTCTTTCTGCACGATTGAGTAAATCAAGACGATTAGAAAATCTCA
TTGCTCAGCTCCCCGGTGAGAAGAAAAATGGCTTATTTGGGAATCTCATTGCTTTGTCATTGGGTTTGACCCCTAA
TTTTAAATCAAATTTTGATTTGGCAGAAGATGCTAAATTACAGCTTTCAAAAGATACTTACGATGATGATTTAGAT
AATTTATTGGCGCAAATTGGAGATCAATATGCTGATTTGTTTTTGGCAGCTAAGAATTTATCAGATGCTATTTTAC
TTTCAGATATCCTAAGAGTAAATACTGAAATAACTAAGGCTCCCCTATCAGCTTCAATGATTAAACGCTACGATGA
ACATCATCAAGACTTGACTCTTTTAAAAGCTTTAGTTCGACAACAACTTCCAGAAAAGTATAAAGAAATCTTTTTT
GATCAATCAAAAAACGGATATGCAGGTTATATTGATGGGGGAGCTAGCCAAGAAGAATTTTATAAATTTATCAAAC
CAATTTTAGAAAAAATGGATGGTACTGAGGAATTATTGGTGAAACTAAATCGTGAAGATTTGCTGCGCAAGCAACG
GACCTTTGACAACGGCTCTATTCCCCATCAAATTCACTTGGGTGAGCTGCATGCTATTTTGAGAAGACAAGAAGAC
TTTTATCCATTTTTAAAAGACAATCGTGAGAAGATTGAAAAAATCTTGACTTTTCGAATTCCTTATTATGTTGGTC
CATTGGCGCGTGGCAATAGTCGTTTTGCATGGATGACTCGGAAGTCTGAAGAAACAATTACCCCATGGAATTTTGA
AGAAGTTGTCGATAAAGGTGCTTCAGCTCAATCATTTATTGAACGCATGACAAACTTTGATAAAAATCTTCCAAAT
GAAAAAGTACTACCAAAACATAGTTTGCTTTATGAGTATTTTACGGTTTATAACGAATTGACAAAGGTCAAATATG
TTACTGAAGGAATGCGAAAACCAGCATTTCTTTCAGGTGAACAGAAGAAAGCCATTGTTGATTTACTCTTCAAAAC
AAATCGAAAAGTAACCGTTAAGCAATTAAAAGAAGATTATTTCAAAAAAATAGAATGTTTTGATAGTGTTGAAATT
TCAGGAGTTGAAGATAGATTTAATGCTTCATTAGGTACCTACCATGATTTGCTAAAAATTATTAAAGATAAAGATT
TTTTGGATAATGAAGAAAATGAAGATATCTTAGAGGATATTGTTTTAACATTGACCTTATTTGAAGATAGGGAGAT
GATTGAGGAAAGACTTAAAACATATGCTCACCTCTTTGATGATAAGGTGATGAAACAGCTTAAACGTCGCCGTTAT
ACTGGTTGGGGACGTTTGTCTCGAAAATTGATTAATGGTATTAGGGATAAGCAATCTGGCAAAACAATATTAGATT
TTTTGAAATCAGATGGTTTTGCCAATCGCAATTTTATGCAGCTGATCCATGATGATAGTTTGACATTTAAAGAAGA
CATTCAAAAAGCACAAGTGTCTGGACAAGGCGATAGTTTACATGAACATATTGCAAATTTAGCTGGTAGCCCTGCT
ATTAAAAAAGGTATTTTACAGACTGTAAAAGTTGTTGATGAATTGGTCAAAGTAATGGGGCGGCATAAGCCAGAAA
ATATCGTTATTGAAATGGCACGTGAAAATCAGACAACTCAAAAGGGCCAGAAAAATTCGCGAGAGCGTATGAAACG
AATCGAAGAAGGTATCAAAGAATTAGGAAGTCAGATTCTTAAAGAGCATCCTGTTGAAAATACTCAATTGCAAAAT
GAAAAGCTCTATCTCTATTATCTCCAAAATGGAAGAGACATGTATGTGGACCAAGAATTAGATATTAATCGTTTAA
GTGATTATGATGTCGATCACATTGTTCCACAAAGTTTCCTTAAAGACGATTCAATAGACAATAAGGTCTTAACGCG
TTCTGATAAAAATCGTGGTAAATCGGATAACGTTCCAAGTGAAGAAGTAGTCAAAAAGATGAAAAACTATTGGAGA
CAACTTCTAAACGCCAAGTTAATCACTCAACGTAAGTTTGATAATTTAACGAAAGCTGAACGTGGAGGTTTGAGTG
AACTTGATAAAGCTGGTTTTATCAAACGCCAATTGGTTGAAACTCGCCAAATCACTAAGCATGTGGCACAAATTTT
GGATAGTCGCATGAATACTAAATACGATGAAAATGATAAACTTATTCGAGAGGTTAAAGTGATTACCTTAAAATCT
AAATTAGTTTCTGACTTCCGAAAAGATTTCCAATTCTATAAAGTACGTGAGATTAACAATTACCATCATGCCCATG
ATGCGTATCTAAATGCCGTCGTTGGAACTGCTTTGATTAAGAAATATCCAAAACTTGAATCGGAGTTTGTCTATGG

TGATTATAAAGTTTATGATGTTCGTAAAATGATTGCTAAGTCTGAGCAAGAAATAGGCAAAGCAACCGCAAAATAT
TTCTTTTACTCTAATATCATGAACTTCTTCAAAACAGAAATTACACTTGCAAATGGAGAGATTCGCAAACGCCCTC
TAATCGAAACTAATGGGGAAACTGGAGAAATTGTCTGGGATAAAGGGCGAGATTTTGCCACAGTGCGCAAAGTATT
GTCCATGCCCCAAGTCAATATTGTCAAGAAAACAGAAGTACAGACAGGCGGATTCTCCAAGGAGTCAATTTTACCA
AAAAGAAATTCGGACAAGCTTATTGCTCGTAAAAAAGACTGGGATCCAAAAAAATATGGTGGTTTTGATAGTCCAA
CGGTAGCTTATTCAGTCCTAGTGGTTGCTAAGGTGGAAAAAGGGAAATCGAAGAAGTTAAAATCCGTTAAAGAGTT
ACTAGGGATCACAATTATGGAAAGAAGTTCCTTTGAAAAAAATCCGATTGACTTTTTAGAAGCTAAAGGATATAAG
GAAGTTAAAAAAGACTTAATCATTAAACTACCTAAATATAGTCTTTTTGAGTTAGAAAACGGTCGTAAACGGATGC
TGGCTAGTGCCGGAGAATTACAAAAAGGAAATGAGCTGGCTCTGCCAAGCAAATATGTGAATTTTTTATATTTAGC
TAGTCATTATGAAAAGTTGAAGGGTAGTCCAGAAGATAACGAACAAAAACAATTGTTTGTGGAGCAGCATAAGCAT
TATTTAGATGAGATTATTGAGCAAATCAGTGAATTTTCTAAGCGTGTTATTTTAGCAGATGCCAATTTAGATAAAG
TTCTTAGTGCATATAACAAACATAGAGACAAACCAATACGTGAACAAGCAGAAAATATTATTCATTTATTTACGTT
GACGAATCTTGGAGCTCCCGCTGCTTTTAAATATTTTGATACAACAATTGATCGTAAACGATATACGTCTACAAAA
GAAGTTTTAGATGCCACTCTTATCCATCAATCCATCACTGGTCTTTATGAAACACGCATTGATTTGAGTCAGCTAG
GAGGTGACTGA(SEQ ID NO:3)

[0086] MDKKYSIGLDIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETAEATRLKRT
ARRRYTRRKNRICYLQEIFSNEMAKVDDSFFHRLEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTI
YHLRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFEENPIN
ASGVDAKAILSARLSKSRRLENLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQLSKDT
YDDDLDNLLAQIGDQYADLFLAAKNLSDAILLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKA
LVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLRKQRT
FDNGSIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPYYVGPLARGNSRFAWMTRKSEETIT
PWNFEEVVDKGASAQSFIERMTNFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYVTEGMRKPAF
LSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKIIKDKDFL
DNEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQ
SGKTILDFLKSDGFANRNFMQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVK
VVDELVKVMGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHPVENTQLQN
EKLYLYYLQNGRDMYVDQELDINRLSDYDVDHIVPQSFLKDDSIDNKVLTRSDKNRGKSDNVPSEE
VVKKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKAGFIKRQLVETRQITKHVAQILDSRM
NTKYDENDKLIREVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYLNAVVGTALIKKYPKLE
SEFVYGDYKVYDVRKMIAKSEQEIGKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGEIV
WDKGRDFATVRKVLSMPQVNIVKKTEVQTGGFSKESILPKRNSDKLIARKKDWDPKKYGGFDSPT
VAYSVLVVAKVEKGKSKKLKSVKELLGITIMERSSFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFEL
ENGRKRMLASAGELQKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQHKHYLDEIIEQ
ISEFSKRVILADANLDKVLSAYNKHRDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKE
VLDATLIHQSITGLYETRIDLSQLGGD(SEQ ID NO:6)(单下划线:HNH域;双下划线:RuvC域)

[0087] 在一些实施方案中,Cas9是指来自以下的Cas9:溃疡棒杆菌(Corynebacterium ulcerans)(NCBI Ref:NC_015683.1、NC_017317.1);白喉棒杆菌(Corynebacterium diphtheria)(NCBI Ref:NC_016782.1、NC_016786.1);Spiroplasma syrphidicola(NCBI Ref:NC_021284.1);间型普雷沃氏菌(Prevotella intermedia)(NCBI Ref:NC_017861.1);台湾螺原体(Spiroplasma taiwanense(NCBI Ref:NC_021846.1);海豚链球菌(Streptococcus iniae)(NCBI Ref:NC_021314.1);波罗的海贝尔氏菌(Belliella baltica)(NCBI Ref:NC_018010.1);Psychroflexus torquisI(NCBI Ref:NC_018721.1);嗜热链球菌(Streptococcus thermophilus)(NCBI Ref:YP_820832.1)、无害李斯特氏菌(Listeria innocua)(NCBI Ref:NP_472073.1)、空肠弯曲杆菌(Campylobacter jejuni)

(NCBI Ref:YP_002344900.1)或脑膜炎奈瑟氏球菌(Neisseria.meningitides)(NCBI Ref:YP_002342100.1)或者是指来自任何其他生物体的Cas9。

[0088] 在一些实施方案中,dCas9对应于,或包含部分或全部的Cas9氨基酸序列,所述Cas9氨基酸序列具有一个或多个使Cas9核酸酶活性失活的突变。例如,在一些实施方案中,dCas9域包含SEQ ID NO:6的D10A和H840A突变,或另一种Cas9中的相应的突变。在一些实施方案中,dCas9包含SEQ ID NO:7的氨基酸序列

[0089] dCas9(D10A和H840A):

[0090] MDKKYSIGLAIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETAEATRLKRT
ARRRYTRRKNRICYLQEIFSNEMAKVDDSFFHRLEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTI
YHLRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFEENPIN
ASGVDAKAILSARLSKSRRLENLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQLSKDT
YDDDLDNLLAQIGDQYADLFLAAKNLSDAILLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKA
LVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLRKQRT
FDNGSIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPYYVGPLARGNSRFAWMTRKSEETIT
PWNFEEVVDKGASAQSFIERMTNFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYVTEGMRKPAF
LSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKIIKDKDFL
DNEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQ
SGKTILDFLKSDGFANRNFMQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVK
VVDELVKVMGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHPVENTQLQN
EKLYLYYLQNGRDMYVDQELDINRLSDYDVDAIVPQSFLKDDSIDNKVLTRSDKNRGKSDNVPSEE
VVKKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKAGFIKRQLVETRQITKHVAQILDSRM
NTKYDENDKLIREVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYLNAVVGTALIKKYPKLE
SEFVYGDYKVYDVRKMIAKSEQEIGKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGEIV
WDKGRDFATVRKVLSMPQVNIVKKTEVQTGGFSKESILPKRNSDKLIARKKDWDPKKYGGFDSPT
VAYSVLVVAKVEKGKSKKLKSVKELLGITIMERSSFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFEL
ENGRKRMLASAGELQKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQHKHYLDEIIEQ
ISEFSKRVILADANLDKVLSAYNKHRDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKE
VLDATLIHQSITGLYETRIDLSQLGGD (SEQ ID NO: 7) (单下划线:HNH域;双下划线:RuvC域)。

[0091] 在一些实施方案中,Cas9域包含D10A突变,而在SEQ ID NO:6中提供的氨基酸序列中的位置840处,或在另一种Cas9,诸如SEQ ID NO:4-26中提供的任何氨基酸序列中所示的Cas9的相应的位置处的残基仍然是组氨酸。不希望受任何特定理论的束缚,催化残基H840的存在维持Cas9的活性以切割含有与靶定的A相对的T的非编辑的(例如,非脱氨基的)链。H840的回复(例如,从dCas9的A840)不导致含有A的靶链的切割。此类Cas9变体能够基于gRNA定义的靶序列在特定位置处产生单链DNA断裂(切口),导致非编辑的链的修复,最终导致非编辑的链上的T至C变化。

[0092] 在其他实施方案中,提供了具有除D10A和H840A之外的突变的dCas9变体,其例如导致核酸酶失活的Cas9(dCas9)。举例来说,此类突变包括D10和H840处的其他氨基酸取代,或Cas9的核酸酶域内的其他取代(例如,HNH核酸酶亚域和/或RuvC1亚域中的取代)。在一些实施方案中,提供了dCas9的变体或同源物(例如SEQ ID NO:6,7,8,9,或22的变体),其与SEQ ID NO:6,7,8,9,或22至少约70%相同、至少约80%相同、至少约90%相同、至少约95%相同、至少约98%相同、至少约99%相同、至少约99.5%相同或至少约99.9%相同。在一些实施方案中,提供了dCas9的变体(例如SEQ ID NO:6,7,8,9,或22的变体),其具有比SEQ ID NO:7,8,9,或22短或长约5个氨基酸、约10个氨基酸、约15个氨基酸、约20个氨基酸、约25个氨基酸、约30个氨基酸、约40个氨基酸、约50个氨基酸、约75个氨基酸、约100个氨基酸或更多的氨基酸序列。

[0093] 在一些实施方案中，如本文提供的Cas9融合蛋白包含Cas9蛋白的全长氨基酸序列，例如本文提供的Cas9序列之一。然而，在其他实施方案中，如本文提供的融合蛋白不包含全长Cas9序列，而仅包含其片段。例如，在一些实施方案中，本文提供的Cas9融合蛋白包含Cas9片段，其中所述片段结合crRNA和tracrRNA或sgRNA，但不包含功能性核酸酶域，例如其中其仅包含截短形式的核酸酶域或根本没有核酸酶域。

[0094] 本文提供了合适的Cas9域和Cas9片段的示例性氨基酸序列，并且对于本领域技术人员来说，Cas9域和片段的另外合适的序列将是显而易见的。

[0095] 在一些实施方案中，Cas9是指来自以下的Cas9：溃疡棒杆菌(Corynebacterium ulcerans)(NCBI Ref:NC_015683.1、NC_017317.1)；白喉棒杆菌(Corynebacterium diphtheria)(NCBI Ref:NC_016782.1、NC_016786.1)；Spiroplasma syrphidicola(NCBI Ref:NC_021284.1)；间型普雷沃氏菌(Prevotella intermedia)(NCBI Ref:NC_017861.1)；台湾螺原体(Spiroplasma taiwanense(NCBI Ref:NC_021846.1)；海豚链球菌(Streptococcus iniae)(NCBI Ref:NC_021314.1)；波罗的海贝尔氏菌(Belliella baltica)(NCBI Ref:NC_018010.1)；Psychroflexus torquisI(NCBI Ref:NC_018721.1)；嗜热链球菌(Streptococcus thermophilus)(NCBI Ref:YP_820832.1)；无害李斯特氏菌(Listeria innocua)(NCBI Ref:NP_472073.1)；空肠弯曲杆菌(Campylobacter jejuni)(NCBI Ref:YP_002344900.1)；或脑膜炎奈瑟氏球菌(Neisseria.meningitides)(NCBI Ref:YP_002342100.1)。

[0096] 应当理解，另外的Cas9蛋白(例如，核酸酶死亡Cas9(dCas9)、Cas9切口酶(nCas9)或核酸酶活性Cas9)，包括其变体和同源物，都在本公开内容的范围内。示例性Cas9蛋白包括但不限于下文提供的那些。在一些实施方案中，Cas9蛋白是核酸酶死亡Cas9(dCas9)。在一些实施方案中，dCas9包含氨基酸序列(SEQ ID NO:7,8,9,或22)。在一些实施方案中，Cas9蛋白是Cas9切口酶(nCas9)。在一些实施方案中，nCas9包含氨基酸序列(SEQ ID NO:10,13,16,或21)。在一些实施方案中，Cas9蛋白是核酸酶活性Cas9。在一些实施方案中，核酸酶活性Cas9包含氨基酸序列(SEQ ID NO:4,5,6,11,12,14,15,16,17,18,19,20,23,24,25,或26)。

[0097] 示例性的催化惰性Cas9(dCas9)：

[0098] DKKYSIGLAIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETAEATRLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFFHRLEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFEENPINASGVDAKAILSARLSKSRRLENLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAILLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLRKQRTFDNGSIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPYYVGPLARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQSFIERMTNFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKIIKDKDFLDNEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFANRNFMQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKVVDELVKVMGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVDAIVPQSFLKDDSIDNKVLTRSDKNRGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNLTKAER

GGLSELDKAGFIKRQLVETRQITKHVAQILDSRMNTKYDENDKLIREVKVITLKSKLVSDFRKDFQFYKVREINNY
HHAHDAYLNAVVGTALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQEIGKATAKYFFYSNIMNFFKTEITLANGEI
RKRPLIETNGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEVQTGGFSKESILPKRNSDKLIARKKDWDPKKYGG
FDSPTVAYSVLVVAKVEKGKSKKLKSVKELLGITIMERSSFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFELENG
RKRMLASAGELQKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQHKHYLDEIIEQISEFSKRVILADA
NLDKVLSAYNKHRDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVLDATLIHQSITGLYETRID
LSQLGGD(SEQ ID NO:8)

[0099] 示例性Cas9切口酶(nCas9):

[0100] DKKYSIGLAIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETAEATRLKRTAR
RRYTRRKNRICYLQEIFSNEMAKVDDSFFHRLEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDS
TDKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFEENPINASGVDAKAILSARLSKSRR
LENLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLS
DAILLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFY
KFIKPILEKMDGTEELLVKLNREDLLRKQRTFDNGSIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIP
YYVGPLARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQSFIERMTNFDKNLPNEKVLPKHSLLYEYFTVYNELT
KVKYVTEGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKII
KDKDFLDNEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGK
TILDFLKSDGFANRNFMQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKVVDELVKVMGR
HKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQELD
INRLSDYDVDHIVPQSFLKDDSIDNKVLTRSDKNRGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNLTKAER
GGLSELDKAGFIKRQLVETRQITKHVAQILDSRMNTKYDENDKLIREVKVITLKSKLVSDFRKDFQFYKVREINNY
HHAHDAYLNAVVGTALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQEIGKATAKYFFYSNIMNFFKTEITLANGEI
RKRPLIETNGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEVQTGGFSKESILPKRNSDKLIARKKDWDPKKYGG
FDSPTVAYSVLVVAKVEKGKSKKLKSVKELLGITIMERSSFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFELENG
RKRMLASAGELQKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQHKHYLDEIIEQISEFSKRVILADA
NLDKVLSAYNKHRDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVLDATLIHQSITGLYETRID
LSQLGGD(SEQ ID NO:10)

[0101] 示例性催化活性的Cas9:

[0102] DKKYSIGLDIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETAEATRLKRTAR
RRYTRRKNRICYLQEIFSNEMAKVDDSFFHRLEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDS
TDKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFEENPINASGVDAKAILSARLSKSRR
LENLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLS
DAILLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFY
KFIKPILEKMDGTEELLVKLNREDLLRKQRTFDNGSIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIP
YYVGPLARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQSFIERMTNFDKNLPNEKVLPKHSLLYEYFTVYNELT
KVKYVTEGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKII
KDKDFLDNEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGK
TILDFLKSDGFANRNFMQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKVVDELVKVMGR
HKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQELD

INRLSDYDVDHIVPQSFLKDDSIDNKVLTRSDKNRGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNLTKAER
GGLSELDKAGFIKRQLVETRQITKHVAQILDSRMNTKYDENDKLIREVKVITLKSKLVSDFRKDFQFYKVREINNY
HHAHDAYLNAVVGTALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQEIGKATAKYFFYSNIMNFFKTEITLANGEI
RKRPLIETNGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEVQTGGFSKESILPKRNSDKLIARKKDWDPKKYGG
FDSPTVAYSVLVVAKVEKGKSKKLKSVKELLGITIMERSSFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFELENG
RKRMLASAGELQKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQHKHYLDEIIEQISEFSKRVILADA
NLDKVLSAYNKHRDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVLDATLIHQSITGLYETRID
LSQLGGD(SEQ ID NO:11).

[0103] 如本文所用,术语“Cas9切口酶”是指能够切割仅双链体核酸分子(例如,双链体DNA分子)的一条链的Cas9蛋白。在一些实施方案中,Cas9切口酶包含D10A突变并且具有在SEQ ID NO:6的H840位置处的组氨酸,或者在所提供的任何Cas9,例如SEQ ID NO:4-26中的任何一项中的相应的突变。例如,Cas9切口酶可以包含如SEQ ID NO:10、13、16或21所示的氨基酸序列。此类Cas9切口酶具有活性的HNH核酸酶域,并且能够切割DNA的非靶定链,即由gRNA结合的链。此外,此类Cas9切口酶具有无活性的RuvC核酸酶域,并且不能切割DNA的靶定链,即期望进行碱基编辑的链。

[0104] 在一些实施方案中,Cas9是指来自构成单细胞原核微生物的域和界的古生菌(arehaea)(例如纳古生菌(nanoarchaea))的Cas9。在一些实施方案中,Cas9是指CasX或CasY,其已经描述于例如Burstein et al.,“New CRISPR-Cas systems from uncultivated microbes.”Cell Res.2017Feb 21.doi:10.1038/cr.2017.21,其全部内容通过引用并入本文。使用基因组分辨的宏基因组学,鉴定了许多CRISPR-Cas系统,包括在生命的古生菌域中首次报道的Cas9。这种趋异的Cas9蛋白在研究很少的纳古生菌中作为活性CRISPR-Cas系统的一部分发现。在细菌中,发现了两个以前未知的系统,CRISPR-CasX和CRISPR-CasY,它们是迄今发现的最紧凑的系统之一。在一些实施方案中,Cas9是指CasX或CasX的变体。在一些实施方案中,Cas9是指CasY或CasY的变体。应当理解,其他RNA引导的DNA结合蛋白可以用作核酸可编程DNA结合蛋白(napDNAbp),并且在本公开内容的范围内。

[0105] 在一些实施方案中,本文提供的任何融合蛋白的核酸可编程DNA结合蛋白(napDNAbp)可以是CasX或CasY蛋白。在一些实施方案中,napDNAbp是CasX蛋白。在一些实施方案中,CasX蛋白是无核酸酶活性的CasX蛋白(dCasX)、CasX切口酶(CasXn)或核酸酶活性CasX。在一些实施方案中,napDNAbp是CasY蛋白。在一些实施方案中,CasY蛋白是无核酸酶活性的CasY蛋白(dCasY)、CasY切口酶(CasYn)或核酸酶活性CasY。在一些实施方案中,napDNAbp包含与天然存在的CasX或CasY蛋白至少85%、至少90%、至少91%、至少92%、至少93%、至少94%、至少95%、至少96%、至少97%、至少98%、至少99%或至少99.5%相同的氨基酸序列。在一些实施方案中,napDNAbp是天然存在的CasX或CasY蛋白。在一些实施方案中,napDNAbp包含与SEQ ID NO:27-29中的任一个至少85%、至少90%、至少91%、至少92%、至少93%、至少94%、至少95%、至少96%、至少97%、至少98%、至少99%或至少99.5%相同的氨基酸序列。在一些实施方案中,napDNAbp包含SEQ ID NO:27-29中的任一个的氨基酸序列。应当理解,根据本公开内容也可以使用来自其他细菌物种的CasX和CasY。

[0106] CasX(uniprot.org/uniprot/F0NN87;http://www.uniprot.org/uniprot/F0NH53)>tr|F0NN87|F0NN87_SULIH CRISPR相关的Casx蛋白质OS=冰岛硫化叶菌

(Sulfolobus islandicus)(菌株HVE10/4)GN＝SiH_0402PE＝4SV＝1

[0107] MEVPLYNIFGDNYIIQVATEAENSTIYNNKVEIDDEELRNVLNLAYKIAKNNEDAAAERRGKAKKKKG EEGETTTSNIILPLSGNDKNPWTETLKCYNFPTTVALSEVFKNFSQVKECEEVSAPSFVKPEFYEFGRSPGMVERT RRVKLEVEPHYLIIAAAGWVLTRLGKAKVSEGDYVGVNVFTPTRGILYSLIQNVNGIVPGIKPETAFGLWIARKVV SSVTNPNVSVVRIYTISDAVGQNPTTINGGFSIDLTKLLEKRYLLSERLEAIARNALSISSNMRERYIVLANYIYE YLTGSKRLEDLLYFANRDLIMNLNSDDGKVRDLKLISAYVNGELIRGEG(SEQ ID NO:27)

[0108] >tr|F0NH53|F0NH53_SULIR CRISPR相关的蛋白质,Casx OS＝冰岛硫化叶菌 (Sulfolobus islandicus)(菌株REY15A)GN＝SiRe_0771PE＝4SV＝1

[0109] MEVPLYNIFGDNYIIQVATEAENSTIYNNKVEIDDEELRNVLNLAYKIAKNNEDAAAERRGKAKKKKG EEGETTTSNIILPLSGNDKNPWTETLKCYNFPTTVALSEVFKNFSQVKECEEVSAPSFVKPEFYKFGRSPGMVERT RRVKLEVEPHYLIMAAAGWVLTRLGKAKVSEGDYVGVNVFTPTRGILYSLIQNVNGIVPGIKPETAFGLWIARKVV SSVTNPNVSVVSIYTISDAVGQNPTTINGGFSIDLTKLLEKRDLLSERLEAIARNALSISSNMRERYIVLANYIYE YLTGSKRLEDLLYFANRDLIMNLNSDDGKVRDLKLISAYVNGELIRGEG(SEQ ID NO:28)

[0110] CasY(ncbi.nlm.nih.gov/protein/APG80656.1)

[0111] >APG80656.1 CRISPR相关的蛋白质CasY[未培养的Parcubacteria组细菌]

[0112] MSKRHPRISGVKGYRLHAQRLEYTGKSGAMRTIKYPLYSSPSGGRTVPREIVSAINDDYVGLYGLSNF DDLYNAEKRNEEKVYSVLDFWYDCVQYGAVFSYTAPGLLKNVAEVRGGSYELTKTLKGSHLYDELQIDKVIKFLNK KEISRANGSLDKLKKDIIDCFKAEYRERHKDQCNKLADDIKNAKKDAGASLGERQKKLFRDFFGISEQSENDKPSF TNPLNLTCCLLPFDTVNNNRNRGEVLFNKLKEYAQKLDKNEGSLEMWEYIGIGNSGTAFSNFLGEGFLGRLRENKI TELKKAMMDITDAWRGQEQEEELEKRLRILAALTIKLREPKFDNHWGGYRSDINGKLSSWLQNYINQTVKIKEDLK GHKKDLKKAKEMINRFGESDTKEEAVVSSLLESIEKIVPDDSADDEKPDIPAIAIYRRFLSDGRLTLNRFVQREDV QEALIKERLEAEKKKKPKKRKKKSDAEDEKETIDFKELFPHLAKPLKLVPNFYGDSKRELYKKYKNAAIYTDALWK AVEKIYKSAFSSSLKNSFFDTDFDKDFFIKRLQKIFSVYRRFNTDKWKPIVKNSFAPYCDIVSLAENEVLYKPKQS RSRKSAAIDKNRVRLPSTENIAKAGIALARELSVAGFDWKDLLKKEEHEEYIDLIELHKTALALLLAVTETQLDIS ALDFVENGTVKDFMKTRDGNLVLEGRFLEMFSQSIVFSELRGLAGLMSRKEFITRSAIQTMNGKQAELLYIPHEFQ SAKITTPKEMSRAFLDLAPAEFATSLEPESLSEKSLLKLKQMRYYPHYFGYELTRTGQGIDGGVAENALRLEKSPV KKREIKCKQYKTLGRGQNKIVLYVRSSYYQTQFLEWFLHRPKNVQTDVAVSGSFLIDEKKVKTRWNYDALTVALEP VSGSERVFVSQPFTIFPEKSAEEEGQRYLGIDIGEYGIAYTALEITGDSAKILDQNFISDPQLKTLREEVKGLKLD QRRGTFAMPSTKIARIRESLVHSLRNRIHHLALKHKAKIVYELEVSRFEEGKQKIKKVYATLKKADVYSEIDADKN LQTTVWGKLAVASEISASYTSQFCGACKKLWRAEMQVDETITTQELIGTVRVIKGGTLIDAIKDFMRPPIFDENDT PFPKYRDFCDKHHISKKMRGNSCLFICPFCRANADADIQASQTIALLRYVKEEKKVEDYFERFRKLKNIKVLGQMK KI(SEQ ID NO:29)

[0113] 如本文所用,术语“有效量”是指足以引起期望的生物学反应的生物活性剂的量。例如,在一些实施方案中,核碱基编辑器的有效量可以指足以诱导由核碱基编辑器突变的特异性结合的靶位点的突变的核碱基编辑器的量。在一些实施方案中,本文提供的融合蛋白,例如包含核酸可编程DNA结合蛋白和脱氨酶域(例如胞苷脱氨酶域)的融合蛋白的有效量可以指足以诱导融合蛋白特异性结合和编辑的靶位点的编辑的融合蛋白的量。如熟练技术人员将理解的,试剂,例如融合蛋白、核碱基编辑器、脱氨酶、杂合蛋白、蛋白质二聚体、蛋白质(或蛋白质二聚体)和多核苷酸的复合物,或多核苷酸的有效量可以随各种因素而变

化，诸如例如随期望的生物学反应，例如随待编辑的特定等位基因、基因组或靶位点，随靶定的细胞或组织和使用的试剂而变化。

[0114] 如本文所用，术语“核酸”和“核酸分子”是指包含核碱基和酸性部分(例如核苷、核苷酸或核苷酸的聚合物)的化合物。通常，聚合核酸，例如包含三个或更多个核苷酸的核酸分子是线性分子，其中相邻的核苷酸经由磷酸二酯连接彼此连接。在一些实施方案中，“核酸”是指个别的核酸残基(例如核苷酸和/或核苷)。在一些实施方案中，“核酸”是指包含三个或更多个个别核苷酸残基的寡核苷酸链。如本文所用，术语“寡核苷酸”和“多核苷酸”可以可互换地使用以指核苷酸的聚合物(例如，至少三个核苷酸的串)。在一些实施方案中，“核酸”涵盖RNA以及单链和/或双链DNA。核酸可以是天然存在的，例如在基因组、转录物、mRNA、tRNA、rRNA、siRNA、snRNA、质粒、粘粒、染色体、染色单体或其他天然存在的核酸分子的背景下。另一方面，核酸分子可以是非天然存在的分子，例如重组DNA或RNA、人工染色体、工程化的基因组或其片段，或合成DNA、RNA、DNA/RNA杂交体，或包括非天然存在的核苷酸或核苷。此外，术语“核酸”、“DNA”、“RNA”和/或类似术语包括核酸类似物，例如具有除磷酸二酯主链之外的类似物。核酸可以从天然来源纯化，使用重组表达系统产生并任选地纯化、化学合成等。在适当的情况下，例如在化学合成分子的情况下，核酸可以包含核苷类似物，例如具有化学修饰的碱基或糖、和主链修饰的类似物。除非另有说明，核酸序列以5’至3’方向呈现。在一些实施方案中，核酸是或包含天然核苷(例如腺苷、胸苷、鸟苷、胞苷、尿苷、脱氧腺苷、脱氧胸苷、脱氧鸟苷和脱氧胞苷)；核苷类似物(例如2-氨基腺苷、2-硫代胸苷、肌苷、吡咯并嘧啶、3-甲基腺苷、5-甲基胞苷、2-氨基腺苷、C5-溴尿苷、C5-氟尿苷、C5-碘尿苷、C5-丙炔基-尿苷、C5-丙炔基-胞苷、C5-甲基胞苷、2-氨基腺苷、7-脱氮腺苷、7-脱氮鸟苷、8-氧代腺苷、8-氧代鸟苷、O(6)-甲基鸟嘌呤和2-硫代胞苷)；化学修饰的碱基；生物修饰的碱基(例如甲基化碱基)；插入的碱基；修饰的糖(例如2’-氟核糖、核糖、2’-脱氧核糖、阿拉伯糖和己糖)；和/或修饰的磷酸基团(例如硫代磷酸酯和5’-N-亚磷酰胺连接)。

[0115] 如本文所用，术语“增殖性疾病”是指其中细胞或组织稳态受到干扰，使得细胞或细胞群表现出异常升高的增殖速率的任何疾病。增殖性疾病包括过度增殖性疾病，如肿瘤前期增生性状况和肿瘤性疾病。新生性疾病的特征是细胞的异常增殖，并包括良性和恶性新生物两者。恶性新生物也称为癌症。

[0116] 术语“蛋白质”、“肽”和“多肽”在本文中可互换使用，并且是指通过肽(酰胺)键连接在一起的氨基酸残基的聚合物。该术语是指具有任何大小、结构或功能的蛋白质、肽或多肽。通常，蛋白质、肽或多肽将是至少三个氨基酸长。蛋白质、肽或多肽可以指个别的蛋白质或蛋白质的集合。蛋白质、肽或多肽中的一个或多个氨基酸可以被修饰，例如通过添加化学实体如碳水化合物基团、羟基、磷酸基团、法呢基、异法呢基、脂肪酸基团，用于缀合、官能化或其他修饰的接头等。蛋白质、肽或多肽也可以是单个分子或者可以是多分子复合物。蛋白质、肽或多肽可以仅仅是天然存在的蛋白质或肽的片段。蛋白质、肽或多肽可以是天然存在的、重组的或合成的，或其任何组合。如本文所用的术语“融合蛋白”是指包含来自至少两种不同蛋白质的蛋白质域的杂合多肽。一种蛋白质可以位于融合蛋白的氨基端(N端)部分或羧基端(C端)蛋白质，从而分别形成“氨基端融合蛋白”或“羧基端融合蛋白”。蛋白质可以包含不同的域，例如核酸结合域(例如指导蛋白质与靶位点结合的Cas9的gRNA结合域)和核酸编辑蛋白的核酸切割域或催化域。在一些实施方案中，蛋白质包含蛋白质性部分，例如构成

核酸结合域的氨基酸序列,和有机化合物,例如可以起核酸切割试剂作用的化合物。在一些实施方案中,蛋白质与核酸例如RNA复合或缔合。本文提供的任何蛋白质可以通过本领域已知的任何方法产生。例如,本文提供的蛋白质可以经由重组蛋白质表达和纯化产生,其特别适用于包含肽接头的融合蛋白。用于重组蛋白质表达和纯化的方法是熟知的,并且包括Green and Sambrook,Molecular Cloning:A Laboratory Manual(第4版,Cold Spring Harbor Laboratory Press,Cold Spring Harbor,N.Y.(2012))描述的那些,其全部内容通过引用并入本文。

[0117] 术语"RNA可编程核酸酶"和"RNA引导的核酸酶"在本文中可互换使用,并且是指与一个或多个不是切割靶标的RNA形成复合物(例如,结合或缔合)的核酸酶。在一些实施方案中,当与RNA形成复合物时,RNA可编程核酸酶可以称为核酸酶:RNA复合物。通常,结合的RNA称为引导RNA(gRNA)。gRNA可以作为两个或更多个RNA的复合物或者作为单个RNA分子存在。作为单个RNA分子存在的gRNA可以称为单引导RNA(sgRNA),尽管"gRNA"可互换使用以指作为单个分子或作为两个或更多个分子的复合物存在的引导RNA。通常,作为单一RNA种类存在的gRNA包含两个域:(1)与靶核酸共享同源性(例如,并指导Cas9复合物与靶物的结合)的域;和(2)结合Cas9蛋白的域。在一些实施方案中,域(2)对应已知为tracrRNA的序列,并且包含茎-环结构。例如,在一些实施方案中,域(2)与Jineket al.,Science 337:816-821(2012)中提供的tracrRNA相同或同源,其全部内容通过引用并入本文。gRNA的其他实例(例如包括域2的那些)可以在2013年9月6日提交的题为"Switchable Cas9 Nucleases and Uses Thereof"的美国临时专利申请U.S.S.N.61/874,682和2013年9月6号提交的题为"Delivery System For Functional Nucleases"的美国临时专利申请U.S.S.N.61/874,746中找到,每篇的全部内容通过引用以其整体并入本文。在一些实施方案中,gRNA包含域(1)和(2)中的两个或更多个,并且可以称为"延伸的gRNA"。例如,延伸的gRNA将例如结合两个或更多个Cas9蛋白并在两个或更多个不同区域处结合靶核酸,如本文所述。gRNA包含互补靶位点的核苷酸序列,其介导核酸酶/RNA复合物与所述靶位点的结合,提供了核酸酶:RNA复合物的序列特异性。在一些实施方案中,RNA可编程核酸酶是(CRISPR相关系统)(CRISPR-associated system)Cas9内切核酸酶,例如来自酿脓链球菌的Cas9(Csn1)(参见例如"Complete genome sequence of an M1 strain of Streptococcus pyogenes."Ferretti J.J.,McShan W.M.,Ajdic D.J.,Savic D.J.,Savic G.,Lyon K.,Primeaux C.,Sezate S.,Suvorov A.N.,Kenton S.,Lai H.S.,Lin S.P.,Qian Y.,Jia H.G.,Najar F.Z.,Ren Q.,Zhu H.,Song L.,White J.,Yuan X.,Clifton S.W.,Roe B.A.,McLaughlin R.E.,Proc.Natl.Acad.Sci.U.S.A.98:4658-4663(2001);"CRISPR RNA maturation by trans-encoded small RNA and host factor RNase III."Deltcheva E.,Chylinski K.,Sharma C.M.,Gonzales K.,Chao Y.,Pirzada Z.A.,Eckert M.R.,Vogel J.,Charpentier E.,Nature 471:602-607(2011);和"A programmable dual-RNA-guided DNA endonuclease in adaptive bacterial immunity."Jinek M.,Chylinski K.,Fonfara I.,Hauer M.,Doudna J.A.,Charpentier E.Science 337:816-821(2012),每篇的全部内容通过引用并入本文。

[0118] 因为RNA可编程核酸酶(例如Cas9)使用RNA:DNA杂交来靶向DNA切割位点,所以这些蛋白质原则上能够被靶向到由引导RNA规定的任何序列。使用RNA可编程核酸酶例如Cas9

进行位点特异性切割(例如,以修饰基因组)的方法是本领域已知的(参见例如Cong,L.et al.,Multiplex genome engineering using CRISPR/Cas systems.Science 339,819-823(2013);Mali,P.et al.,RNA-guided human genome engineering via Cas9.Science 339,823-826(2013);

[0119] Hwang,W.Y.et al.,Efficient genome editing in zebrafish using a CRISPR-Cas system.Nature biotechnology 31,227-229(2013);Jinek,M.et al.,RNA-programmed genome editing in human cells.eLife 2,e00471(2013);

[0120] Dicarlo,J.E.et al.,Genome engineering in Saccharomyces cerevisiae using CRISPR-Cas systems.Nucleic acids research(2013);Jiang,W.et al.RNA-guided editing of bacterial genomes using CRISPR-Cas systems.Nature biotechnology 31,233-239(2013);每篇的全部内容通过引用并入本文)。

[0121] 如本文所用,术语“受试者”是指个体生物体,例如个体哺乳动物。在一些实施方案中,受试者是人。在一些实施方案中,受试者是非人哺乳动物。在一些实施方案中,受试者是非人灵长类动物。在一些实施方案中,受试者是啮齿动物。在一些实施方案中,受试者是绵羊、山羊、牛、猫或狗。在一些实施方案中,受试者是脊椎动物、两栖动物、爬行动物、鱼、昆虫、苍蝇或线虫。在一些实施方案中,受试者是研究动物。在一些实施方案中,受试者是经遗传工程化的,例如基因遗传化的非人受试者。受试者可以是任何一个性别和处于任何发展阶段的。

[0122] 术语“靶位点”是指由碱基编辑器,诸如包含胞苷脱氨酶的融合蛋白(例如本文提供的dCas9-胞苷脱氨酶融合蛋白)修饰的核酸分子内的序列。

[0123] 术语“治疗/处理”是指如本文所述旨在逆转、缓解疾病或病症或其一种或多种症状、延迟疾病或病症或其一种或多种症状的发作或抑制疾病或病症或其一种或多种症状进展的临床干预。如本文所用,术语“治疗/处理”是指如本文所述旨在逆转、缓解疾病或病症或其一种或多种症状、延迟疾病或病症或其一种或多种症状的发作或抑制疾病或病症或其一种或多种症状进展的临床干预。在一些实施方案中,可以在一种或多种症状已经得以形成之后和/或疾病已经得到诊断之后施用治疗。在其他实施方案中,可以在没有症状的情况下施用治疗,例如用于预防或延迟症状的发作或抑制疾病的发作或进展。例如,可以在症状发作之前(例如,鉴于症状的历史和/或鉴于遗传或其他易感性因素)施用治疗于易感个体。治疗也可以在症状消退后继续进行,例如以预防或延迟其复发。

[0124] 如本文中在蛋白质或核酸的背景中使用,术语“重组”是指自然界中不存在,但是作为人工程化的产物的蛋白质或核酸。例如,在一些实施方案中,重组蛋白质或核酸分子包含氨基酸或核苷酸序列,其相比于任何天然存在的序列包含至少一个、至少两个、至少三个、至少四个、至少五个、至少六个或至少七个突变。

[0125] 发明详述

[0126] 核酸可编程DNA结合蛋白(napDNAbp)

[0127] 本公开内容的一些方面提供了核酸可编程的DNA结合蛋白,其可以用于将蛋白质(例如碱基编辑器)引导至特定的核酸(例如DNA或RNA)序列。核酸可编程DNA结合蛋白包括但不限于Cas9(例如,dCas9和nCas9),CasX,CasY,Cpf1,C2c1,C2c2,C2C3和Argonaute。具有不同于Cas9的PAM特异性的核酸可编程DNA结合蛋白的一个例子是来自Prevotella和弗朗

西丝菌属(Francisella)的成簇规则间隔短回文重复序列1(Cpf1)。与Cas9相似,Cpf1也是2类CRISPR效应器。已经显示Cpf1介导具有不同于Cas9的特征的稳健的DNA干扰。Cpf1是一种缺少tracrRNA的单一RNA指导的内切核酸酶,并且它利用富含T的原间隔物邻近基序(TTN,TTTN或YTN)。此外,Cpf1通过交错的DNA双链断裂切割DNA。在16种Cpf1家族蛋白中,显示来自氨基酸球菌属(Acidaminococcus)和毛螺菌科(Lachnospiraceae)的两种酶在人细胞中具有有效的基因组编辑活性。Cpf1蛋白在本领域中是已知的,并且先前已经记载于Yamano et al.,"Crystal structure of Cpf1 in complex with guide RNA and target DNA." Cell(165)2016,p.949-962;其全部内容通过引用并入本文。

[0128] 在本组合物和方法中也有用的是无核酸酶活性的Cpf1(dCpf1)变体,其可用作引导核苷酸序列可编程DNA结合蛋白域。Cpf1蛋白具有类似于Cas9的RuvC域的RuvC样内切核酸酶域,但不具有HNH内切核酸酶域,并且Cpf1的N末端不具有Cas9的α-螺旋识别叶。Zetsche et al.,Cell,163,759-771,2015(通过引用并入本文)显示了Cpf1的RuvC样域负责切割两条DNA链并且RuvC样域的失活使Cpf1核酸酶活性失活。例如,与新凶手弗朗西斯菌Cpf1(SEQ ID NO:30)中的D917A,E1006A或D1255A相对应的突变使Cpf1核酸酶活性失活。在一些实施方案中,本公开内容的dCpf1包含对应于SEQ ID NO:30中的D917A,E1006A,D1255A,D917A/E1006A,D917A/D1255A,E1006A/D1255A,或D917A/E1006A/D1255A或另一个Cpf1中的相应突变的突变。应当理解,根据本公开内容,可以使用使Cpf1的RuvC域失活的任何突变,例如取代突变,缺失或插入。

[0129] 在一些实施方案中,本文提供的任何融合蛋白的核酸可编程DNA结合蛋白(napDNAbp)可以是Cpf1蛋白。在一些实施方案中,Cpf1蛋白是Cpf1切口酶(nCpf1)。在一些实施方案中,Cpf1蛋白是无核酸酶活性的Cpf1(dCpf1)。在一些实施方案中,Cpf1,nCpf1或dCpf1包含与SEQ ID NO:30-37中的任何一项至少85%,至少90%,至少91%,至少92%,至少93%,至少94%,至少95%,至少96%,至少97%,至少98%,至少99%或至少99.5%相同的氨基酸序列。在一些实施方案中,dCpf1包含与SEQ ID NO:30-37中的任何一项至少85%,至少90%,至少91%,至少92%,至少93%,至少94%,至少95%,至少96%,至少97%,至少98%,至少99%或至少99.5%相同的氨基酸序列,并且包含对应于SEQ ID NO:30中的D917A,E1006A,D1255A,D917A/E1006A,D917A/D1255A,E1006A/D1255A,和或D917A/E1006A/D1255A或另一个Cpf1中的相应突变的突变。在一些实施方案中,dCpf1包含SEQ ID NO:30-37中任何一项的氨基酸序列。应当理解,根据本公开内容,也可以使用来自其他细菌物种的Cpf1。

[0130] 野生型新凶手弗朗西斯菌Cpf1(SEQ ID NO:30)(D917,E1006和D1255是粗体且加下划线的)

[0131] MSIYQEFVNKYSLSKTLRFELIPQGKTLENIKARGLILDDEKRAKDYKKAKQIIDKYHQFFIEEILSSVC
ISEDLLQNYSDVYFKLKKSDDDNLQKDFKSAKDTIKKQISEYIKDSEKFKNLFNQNLIDAKKGQESD
LILWLKQSKDNGIELFKANSDITDIDEALEIIKSFKGWTTYFKGFHENRKNVYSSNDIPTSIIYRIVDDN
LPKFLENKAKYESLKDKAPEAINYEQIKKDLAEELTFDIDYKTSEVNQRVFSLDEVFEIANFNNYLNQ
SGITKFNTIIGGKFVNGENTKRKGINEYINLYSQQINDKTLKKYKMSVLFKQILSDTESKSFVIDKLED
DSDVVTTMQSFYEQIAAFKTVEEKSIKETLSLLFDDLKAQKLDLSKIYFKNDKSLTDLSQQVFDDYS
VIGTAVLEYITQQIAPKNLDNPSKKEQELIAKKTEKAKYLSLETIKLALEEFNKHRDIDKQCRFEEILA
NFAAIPMIFDEIAQNKDNLAQISIKYQNQGKKDLLQASAEDDVKAIKDLLDQTNNLLHKLKIFHISQS
EDKANILDKDEHFYLVFEECYFELANIVPLYNKIRNYITQKPYSDEKFKLNFENSTLANGWDKNKEP
DNTAILFIKDDKYYLGVMNKKNNKIFDDKAIKENKGEGYKKIVYKLLPGANKMLPKVFFSAKSIKF
YNPSEDILRIRNHSTHTKNGSPQKGYEKFEFNIEDCRKFIDFYKQSISKHPEWKDFGFRFSDTQRYNSI
DEFYREVENQGYKLTFENISESYIDSVVNQGKLYLFQIYNKDFSAYSKGRPNLHTLYWKALFDERNL
QDVVYKLNGEAELFYRKQSIPKKITHPAKEAIANKNKDNPKKESVFEYDLIKDKRFTEDKFFFHCPITI
NFKSSGANKFNDEINLLLKEKANDVHILSI**D**RGERHLAYYTLVDGKGNIIKQDTFNIIGNDRMKTNY
HDKLAAIEKDRDSARKDWKKINNIKEMKEGYLSQVVHEIAKLVIEYNAIVVF**E**DLNFGFKRGRFKV
EKQVYQKLEKMLIEKLNYLVFKDNEFDKTGGVLRAYQLTAPFETFKKMGKQTGIIYYVPAGFTSKI
CPVTGFVNQLYPKYESVSKSQEFFSKFDKICYNLDKGYFEFSFDYKNFGDKAAKGKWTIASFGSRLI
NFRNSDKNHNWDTREVYPTKELEKLLKDYSIEYGHGECIKAAICGESDKKFFAKLTSVLNTILQMRN
SKTGTELDYLISPVADVNGNFFDSRQAPKNMPQDA**D**ANGAYHIGLKGLMLLGRIKNNQEGKKLNL
VIKNEEYFEFVQNRNN (SEQ ID NO: 30)

[0132] 新凶手弗朗西斯菌Cpf1 D917A(SEQ ID NO:31)(A917,E1006和D1255是粗体且加下划线的)

[0133] MSIYQEFVNKYSLSKTLRFELIPQGKTLENIKARGLILDDEKRAKDYKKAKQIIDKYHQFFIEEILSSVC
ISEDLLQNYSDVYFKLKKSDDDNLQKDFKSAKDTIKKQISEYIKDSEKFKNLFNQNLIDAKKGQESD
LILWLKQSKDNGIELFKANSDITDIDEALEIIKSFKGWTTYFKGFHENRKNVYSSNDIPTSIIYRIVDDN
LPKFLENKAKYESLKDKAPEAINYEQIKKDLAEELTFDIDYKTSEVNQRVFSLDEVFEIANFNNYLNQ
SGITKFNTIIGGKFVNGENTKRKGINEYINLYSQQINDKTLKKYKMSVLFKQILSDTESKSFVIDKLED
DSDVVTTMQSFYEQIAAFKTVEEKSIKETLSLLFDDLKAQKLDLSKIYFKNDKSLTDLSQQVFDDYS
VIGTAVLEYITQQIAPKNLDNPSKKEQELIAKKTEKAKYLSLETIKLALEEFNKHRDIDKQCRFEEILA
NFAAIPMIFDEIAQNKDNLAQISIKYQNQGKKDLLQASAEDDVKAIKDLLDQTNNLLHKLKIFHISQS
EDKANILDKDEHFYLVFEECYFELANIVPLYNKIRNYITQKPYSDEKFKLNFENSTLANGWDKNKEP
DNTAILFIKDDKYYLGVMNKKNNKIFDDKAIKENKGEGYKKIVYKLLPGANKMLPKVFFSAKSIKF
YNPSEDILRIRNHSTHTKNGSPQKGYEKFEFNIEDCRKFIDFYKQSISKHPEWKDFGFRFSDTQRYNSI
DEFYREVENQGYKLTFENISESYIDSVVNQGKLYLFQIYNKDFSAYSKGRPNLHTLYWKALFDERNL
QDVVYKLNGEAELFYRKQSIPKKITHPAKEAIANKNKDNPKKESVFEYDLIKDKRFTEDKFFFHCPITI
NFKSSGANKFNDEINLLLKEKANDVHILSI**A**RGERHLAYYTLVDGKGNIIKQDTFNIIGNDRMKTNY
HDKLAAIEKDRDSARKDWKKINNIKEMKEGYLSQVVHEIAKLVIEYNAIVVF**E**DLNFGFKRGRFKV
EKQVYQKLEKMLIEKLNYLVFKDNEFDKTGGVLRAYQLTAPFETFKKMGKQTGIIYYVPAGFTSKI
CPVTGFVNQLYPKYESVSKSQEFFSKFDKICYNLDKGYFEFSFDYKNFGDKAAKGKWTIASFGSRLI
NFRNSDKNHNWDTREVYPTKELEKLLKDYSIEYGHGECIKAAICGESDKKFFAKLTSVLNTILQMRN
SKTGTELDYLISPVADVNGNFFDSRQAPKNMPQDA**D**ANGAYHIGLKGLMLLGRIKNNQEGKKLNL
VIKNEEYFEFVQNRNN (SEQ ID NO: 31)

[0134] 新凶手弗朗西斯菌Cpf1 E1006A(SEQ ID NO:32)(D917,A1006和D1255是粗体且加下划线的)

[0135] MSIYQEFVNKYSLSKTLRFELIPQGKTLENIKARGLILDDEKRAKDYKKAKQIIDKYHQFFIEEILSSVC
ISEDLLQNYSDVYFKLKKSDDDNLQKDFKSAKDTIKKQISEYIKDSEKFKNLFNQNLIDAKKGQESD
LILWLKQSKDNGIELFKANSDITDIDEALEIIKSFKGWTTYFKGFHENRKNVYSSNDIPTSIIYRIVDDN
LPKFLENKAKYESLKDKAPEAINYEQIKKDLAEELTFDIDYKTSEVNQRVFSLDEVFEIANFNNYLNQ
SGITKFNTIIGGKFVNGENTKRKGINEYINLYSQQINDKTLKKYKMSVLFKQILSDTESKSFVIDKLED
DSDVVTTMQSFYEQIAAFKTVEEKSIKETLSLLFDDLKAQKLDLSKIYFKNDKSLTDLSQQVFDDYS
VIGTAVLEYITQQIAPKNLDNPSKKEQELIAKKTEKAKYLSLETIKLALEEFNKHRDIDKQCRFEEILA
NFAAIPMIFDEIAQNKDNLAQISIKYQNQGKKDLLQASAEDDVKAIKDLLDQTNNLLHKLKIFHISQS
EDKANILDKDEHFYLVFEECYFELANIVPLYNKIRNYITQKPYSDEKFKLNFENSTLANGWDKNKEP
DNTAILFIKDDKYYLGVMNKKNNKIFDDKAIKENKGEGYKKIVYKLLPGANKMLPKVFFSAKSIKF
YNPSEDILRIRNHSTHTKNGSPQKGYEKFEFNIEDCRKFIDFYKQSISKHPEWKDFGFRFSDTQRYNSI
DEFYREVENQGYKLTFENISESYIDSVVNQGKLYLFQIYNKDFSAYSKGRPNLHTLYWKALFDERNL
QDVVYKLNGEAELFYRKQSIPKKITHPAKEAIANKNKDNPKKESVFEYDLIKDKRFTEDKFFFHCPITI
NFKSSGANKFNDEINLLLKEKANDVHILSI**D**RGERHLAYYTLVDGKGNIIKQDTFNIIGNDRMKTNY
HDKLAAIEKDRDSARKDWKKINNIKEMKEGYLSQVVHEIAKLVIEYNAIVVF**A**DLNFGFKRGRFKV
EKQVYQKLEKMLIEKLNYLVFKDNEFDKTGGVLRAYQLTAPFETFKKMGKQTGIIYYVPAGFTSKI
CPVTGFVNQLYPKYESVSKSQEFFSKFDKICYNLDKGYFEFSFDYKNFGDKAAKGKWTIASFGSRLI
NFRNSDKNHNWDTREVYPTKELEKLLKDYSIEYGHGECIKAAICGESDKKFFAKLTSVLNTILQMRN
SKTGTELDYLISPVADVNGNFFDSRQAPKNMPQDA**D**ANGAYHIGLKGLMLLGRIKNNQEGKKLNL
VIKNEEYFEFVQNRNN (SEQ ID NO: 32)

[0136] 新凶手弗朗西斯菌Cpf1 D1255A(SEQ ID NO:33)(D917,E1006和A1255是粗体且加下划线的)

[0137] MSIYQEFVNKYSLSKTLRFELIPQGKTLENIKARGLILDDEKRAKDYKKAKQIIDKYHQFFIEEILSSVC
ISEDLLQNYSDVYFKLKKSDDDNLQKDFKSAKDTIKKQISEYIKDSEKFKNLFNQNLIDAKKGQESD
LILWLKQSKDNGIELFKANSDITDIDEALEIIKSFKGWTTYFKGFHENRKNVYSSNDIPTSIIYRIVDDN
LPKFLENKAKYESLKDKAPEAINYEQIKKDLAEELTFDIDYKTSEVNQRVFSLDEVFEIANFNNYLNQ
SGITKFNTIIGGKFVNGENTKRKGINEYINLYSQQINDKTLKKYKMSVLFKQILSDTESKSFVIDKLED
DSDVVTTMQSFYEQIAAFKTVEEKSIKETLSLLFDDLKAQKLDLSKIYFKNDKSLTDLSQQVFDDYS
VIGTAVLEYITQQIAPKNLDNPSKKEQELIAKKTEKAKYLSLETIKLALEEFNKHRDIDKQCRFEEILA
NFAAIPMIFDEIAQNKDNLAQISIKYQNQGKKDLLQASAEDDVKAIKDLLDQTNNLLHKLKIFHISQS
EDKANILDKDEHFYLVFEECYFELANIVPLYNKIRNYITQKPYSDEKFKLNFENSTLANGWDKNKEP
DNTAILFIKDDKYYLGVMNKKNNKIFDDKAIKENKGEGYKKIVYKLLPGANKMLPKVFFSAKSIKF
YNPSEDILRIRNHSTHTKNGSPQKGYEKFEFNIEDCRKFIDFYKQSISKHPEWKDFGFRFSDTQRYNSI
DEFYREVENQGYKLTFENISESYIDSVVNQGKLYLFQIYNKDFSAYSKGRPNLHTLYWKALFDERNL
QDVVYKLNGEAELFYRKQSIPKKITHPAKEAIANKNKDNPKKESVFEYDLIKDKRFTEDKFFFHCPITI
NFKSSGANKFNDEINLLLKEKANDVHILSI**D**RGERHLAYYTLVDGKGNIIKQDTFNIIGNDRMKTNY
HDKLAAIEKDRDSARKDWKKINNIKEMKEGYLSQVVHEIAKLVIEYNAIVVF**E**DLNFGFKRGRFKV
EKQVYQKLEKMLIEKLNYLVFKDNEFDKTGGVLRAYQLTAPFETFKKMGKQTGIIYYVPAGFTSKI
CPVTGFVNQLYPKYESVSKSQEFFSKFDKICYNLDKGYFEFSFDYKNFGDKAAKGKWTIASFGSRLI
NFRNSDKNHNWDTREVYPTKELEKLLKDYSIEYGHGECIKAAICGESDKKFFAKLTSVLNTILQMRN
SKTGTELDYLISPVADVNGNFFDSRQAPKNMPQDA**A**ANGAYHIGLKGLMLLGRIKNNQEGKKLNL
VIKNEEYFEFVQNRNN (SEQ ID NO: 33)

[0138] 新凶手弗朗西斯菌Cpf1 D917A/E1006A(SEQ ID NO:34)(A917,A1006,和D1255是粗体且加下划线的)

[0139] MSIYQEFVNKYSLSKTLRFELIPQGKTLENIKARGLILDDEKRAKDYKKAKQIIDKYHQFFIEEILSSVC
ISEDLLQNYSDVYFKLKKSDDDNLQKDFKSAKDTIKKQISEYIKDSEKFKNLFNQNLIDAKKGQESD
LILWLKQSKDNGIELFKANSDITDIDEALEIIKSFKGWTTYFKGFHENRKNVYSSNDIPTSIIYRIVDDN
LPKFLENKAKYESLKDKAPEAINYEQIKKDLAEELTFDIDYKTSEVNQRVFSLDEVFEIANFNNYLNQ
SGITKFNTIIGGKFVNGENTKRKGINEYINLYSQQINDKTLKKYKMSVLFKQILSDTESKSFVIDKLED

[0140] DSDVVTTMQSFYEQIAAFKTVEEKSIKETLSLLFDDLKAQKLDLSKIYFKNDKSLTDLSQQVFDDYS
VIGTAVLEYITQQIAPKNLDNPSKKEQELIAKKTEKAKYLSLETIKLALEEFNKHRDIDKQCRFEEILA
NFAAIPMIFDEIAQNKDNLAQISIKYQNQGKKDLLQASAEDDVKAIKDLLDQTNNLLHKLKIFHISQS
EDKANILDKDEHFYLVFEECYFELANIVPLYNKIRNYITQKPYSDEKFKLNFENSTLANGWDKNKEP
DNTAILFIKDDKYYLGVMNKKNNKIFDDKAIKENKGEGYKKIVYKLLPGANKMLPKVFFSAKSIKF
YNPSEDILRIRNHSTHTKNGSPQKGYEKFEFNIEDCRKFIDFYKQSISKHPEWKDFGFRFSDTQRYNSI
DEFYREVENQGYKLTFENISESYIDSVVNQGKLYLFQIYNKDFSAYSKGRPNLHTLYWKALFDERNL
QDVVYKLNGEAELFYRKQSIPKKITHPAKEAIANKNKDNPKKESVFEYDLIKDKRFTEDKFFFHCPITI
NFKSSGANKFNDEINLLLKEKANDVHILSIARGERHLAYYTLVDGKGNIIKQDTFNIIGNDRMKTNY
HDKLAAIEKDRDSARKDWKKINNIKEMKEGYLSQVVHEIAKLVIEYNAIVVFADLNFGFKRGRFKV
EKQVYQKLEKMLIEKLNYLVFKDNEFDKTGGVLRAYQLTAPFETFKKMGKQTGIIYYVPAGFTSKI
CPVTGFVNQLYPKYESVSKSQEFFSKFDKICYNLDKGYFEFSFDYKNFGDKAAKGKWTIASFGSRLI
NFRNSDKNHNWDTREVYPTKELEKLLKDYSIEYGHGECIKAAICGESDKKFFAKLTSVLNTILQMRN
SKTGTELDYLISPVADVNGNFFDSRQAPKNMPQDADANGAYHIGLKGLMLLGRIKNNQEGKKLNL
VIKNEEYFEFVQNRNN (SEQ ID NO: 34)

[0141] 新凶手弗朗西斯菌Cpf1 D917A/D1255A(SEQ ID NO:35)(A917,E1006,和A1255是粗体且加下划线的)

[0142] MSIYQEFVNKYSLSKTLRFELIPQGKTLENIKARGLILDDEKRAKDYKKAKQIIDKYHQFFIEEILSSVC
ISEDLLQNYSDVYFKLKKSDDDNLQKDFKSAKDTIKKQISEYIKDSEKFKNLFNQNLIDAKKGQESD
LILWLKQSKDNGIELFKANSDITDIDEALEIIKSFKGWTTYFKGFHENRKNVYSSNDIPTSIIYRIVDDN
LPKFLENKAKYESLKDKAPEAINYEQIKKDLAEELTFDIDYKTSEVNQRVFSLDEVFEIANFNNYLNQ
SGITKFNTIIGGKFVNGENTKRKGINEYINLYSQQINDKTLKKYKMSVLFKQILSDTESKSFVIDKLED
DSDVVTTMQSFYEQIAAFKTVEEKSIKETLSLLFDDLKAQKLDLSKIYFKNDKSLTDLSQQVFDDYS
VIGTAVLEYITQQIAPKNLDNPSKKEQELIAKKTEKAKYLSLETIKLALEEFNKHRDIDKQCRFEEILA
NFAAIPMIFDEIAQNKDNLAQISIKYQNQGKKDLLQASAEDDVKAIKDLLDQTNNLLHKLKIFHISQS
EDKANILDKDEHFYLVFEECYFELANIVPLYNKIRNYITQKPYSDEKFKLNFENSTLANGWDKNKEP
DNTAILFIKDDKYYLGVMNKKNNKIFDDKAIKENKGEGYKKIVYKLLPGANKMLPKVFFSAKSIKF
YNPSEDILRIRNHSTHTKNGSPQKGYEKFEFNIEDCRKFIDFYKQSISKHPEWKDFGFRFSDTQRYNSI
DEFYREVENQGYKLTFENISESYIDSVVNQGKLYLFQIYNKDFSAYSKGRPNLHTLYWKALFDERNL
QDVVYKLNGEAELFYRKQSIPKKITHPAKEAIANKNKDNPKKESVFEYDLIKDKRFTEDKFFFHCPITI
NFKSSGANKFNDEINLLLKEKANDVHILSIARGERHLAYYTLVDGKGNIIKQDTFNIIGNDRMKTNY
HDKLAAIEKDRDSARKDWKKINNIKEMKEGYLSQVVHEIAKLVIEYNAIVVFEDLNFGFKRGRFKV
EKQVYQKLEKMLIEKLNYLVFKDNEFDKTGGVLRAYQLTAPFETFKKMGKQTGIIYYVPAGFTSKI
CPVTGFVNQLYPKYESVSKSQEFFSKFDKICYNLDKGYFEFSFDYKNFGDKAAKGKWTIASFGSRLI
NFRNSDKNHNWDTREVYPTKELEKLLKDYSIEYGHGECIKAAICGESDKKFFAKLTSVLNTILQMRN
SKTGTELDYLISPVADVNGNFFDSRQAPKNMPQDAAANGAYHIGLKGLMLLGRIKNNQEGKKLNL
VIKNEEYFEFVQNRNN (SEQ ID NO: 35)

[0143] 新凶手弗朗西斯菌Cpf1 E1006A/D1255A(SEQ ID NO:36)(D917,A1006和A1255是粗体且加下划线的)

[0144] MSIYQEFVNKYSLSKTLRFELIPQGKTLENIKARGLILDDEKRAKDYKKAKQIIDKY
HQFFIEEILSSVCISEDLLQNYSDVYFKLKKSDDDNLQKDFKSAKDTIKKQISEYIKD
SEKFKNLFNQNLIDAKKGQESDLILWLKQSKDNGIELFKANSDITDIDEALEIIKSFK
GWTTYFKGFHENRKNVYSSNDIPTSIIYRIVDDNLPKFLENKAKYESLKDKAPEAIN
YEQIKKDLAEELTFDIDYKTSEVNQRVFSLDEVFEIANFNNYLNQSGITKFNTIIGGKF
VNGENTKRKGINEYINLYSQQINDKTLKKYKMSVLFKQILSDTESKSFVIDKLEDDS
DVVTTMQSFYEQIAAFKTVEEKSIKETLSLLFDDLKAQKLDLSKIYFKNDKSLTDLS
QQVFDDYSVIGTAVLEYITQQIAPKNLDNPSKKEQELIAKKTEKAKYLSLETIKLALE

[0145] EFNKHRDIDKQCRFEEILANFAAIPMIFDEIAQNKDNLAQISIKYQNQGKKDLLQAS
AEDDVKAIKDLLDQTNNLLHKLKIFHISQSEDKANILDKDEHFYLVFEECYFELANI
VPLYNKIRNYITQKPYSDEKFKLNFENSTLANGWDKNKEPDNTAILFIKDDKYYLG
VMNKKNNKIFDDKAIKENKGEGYKKIVYKLLPGANKMLPKVFFSAKSIKFYNPSE
DILRIRNHSTHTKNGSPQKGYEKFEFNIEDCRKFIDFYKQSISKHPEWKDFGFRFSDT
QRYNSIDEFYREVENQGYKLTFENISESYIDSVVNQGKLYLFQIYNKDFSAYSKGRP
NLHTLYWKALFDERNLQDVVYKLNGEAELFYRKQSIPKKITHPAKEAIANKNKDN
PKKESVFEYDLIKDKRFTEDKFFFHCPITINFKSSGANKFNDEINLLLKEKANDVHIL
SIDRGERHLAYYTLVDGKGNIIKQDTFNIIGNDRMKTNYHDKLAAIEKDRDSARKD
WKKINNIKEMKEGYLSQVVHEIAKLVIEYNAIVVFADLNFGFKRGRFKVEKQVYQ
KLEKMLIEKLNYLVFKDNEFDKTGGVLRAYQLTAPFETFKKMGKQTGIIYYVPAGF
TSKICPVTGFVNQLYPKYESVSKSQEFFSKFDKICYNLDKGYFEFSFDYKNFGDKAA
KGKWTIASFGSRLINFRNSDKNHNWDTREVYPTKELEKLLKDYSIEYGHGECIKAA
ICGESDKKFFAKLTSVLNTILQMRNSKTGTELDYLISPVADVNGNFFDSRQAPKNMP
QDAAANGAYHIGLKGLMLLGRIKNNQEGKKLNLVIKNEEYFEFVQNRNN (SEQ
ID NO: 36)

[0146] 新凶手弗朗西斯菌Cpf1 D917A/E1006A/D1255A(SEQ ID NO:37)(A917,A1006,和A1255是粗体且加下划线的)

[0147] MSIYQEFVNKYSLSKTLRFELIPQGKTLENIKARGLILDDEKRAKDYKKAKQIIDKYHQFFIEEILSSVC
ISEDLLQNYSDVYFKLKKSDDDNLQKDFKSAKDTIKKQISEYIKDSEKFKNLFNQNLIDAKKGQESD
LILWLKQSKDNGIELFKANSDITDIDEALEIIKSFKGWTTYFKGFHENRKNVYSSNDIPTSIIYRIVDDN
LPKFLENKAKYESLKDKAPEAINYEQIKKDLAEELTFDIDYKTSEVNQRVFSLDEVFEIANFNNYLNQ
SGITKFNTIIGGKFVNGENTKRKGINEYINLYSQQINDKTLKKYKMSVLFKQILSDTESKSFVIDKLED
DSDVVTTMQSFYEQIAAFKTVEEKSIKETLSLLFDDLKAQKLDLSKIYFKNDKSLTDLSQQVFDDYS
VIGTAVLEYITQQIAPKNLDNPSKKEQELIAKKTEKAKYLSLETIKLALEEFNKHRDIDKQCRFEEILA
NFAAIPMIFDEIAQNKDNLAQISIKYQNQGKKDLLQASAEDDVKAIKDLLDQTNNLLHKLKIFHISQS
EDKANILDKDEHFYLVFEECYFELANIVPLYNKIRNYITQKPYSDEKFKLNFENSTLANGWDKNKEP
DNTAILFIKDDKYYLGVMNKKNNKIFDDKAIKENKGEGYKKIVYKLLPGANKMLPKVFFSAKSIKF
YNPSEDILRIRNHSTHTKNGSPQKGYEKFEFNIEDCRKFIDFYKQSISKHPEWKDFGFRFSDTQRYNSI
DEFYREVENQGYKLTFENISESYIDSVVNQGKLYLFQIYNKDFSAYSKGRPNLHTLYWKALFDERNL
QDVVYKLNGEAELFYRKQSIPKKITHPAKEAIANKNKDNPKKESVFEYDLIKDKRFTEDKFFFHCPITI
NFKSSGANKFNDEINLLLKEKANDVHILSIARGERHLAYYTLVDGKGNIIKQDTFNIIGNDRMKTNY
HDKLAAIEKDRDSARKDWKKINNIKEMKEGYLSQVVHEIAKLVIEYNAIVVFADLNFGFKRGRFKV
EKQVYQKLEKMLIEKLNYLVFKDNEFDKTGGVLRAYQLTAPFETFKKMGKQTGIIYYVPAGFTSKI
CPVTGFVNQLYPKYESVSKSQEFFSKFDKICYNLDKGYFEFSFDYKNFGDKAAKGKWTIASFGSRLI
NFRNSDKNHNWDTREVYPTKELEKLLKDYSIEYGHGECIKAAICGESDKKFFAKLTSVLNTILQMRN
SKTGTELDYLISPVADVNGNFFDSRQAPKNMPQDAAANGAYHIGLKGLMLLGRIKNNQEGKKLNL
VIKNEEYFEFVQNRNN (SEQ ID NO: 37)

[0148] 在一些实施方案中,核酸可编程DNA结合蛋白(napDNAbp)是不需要规范(NGG)PAM序列的核酸可编程DNA结合蛋白。在一些实施方案中,napDNAbp是argonaute蛋白。此类核酸可编程DNA结合蛋白的一个例子是来自格氏嗜盐碱杆菌(Natronobacterium gregoryi)的Argonaute蛋白(NgAgo)。NgAgo是ssDNA引导的内切核酸酶。NgAgo结合约24个核苷酸的5'磷酸化ssDNA(gDNA)以将其引导至其靶位点,并在gDNA位点处使DNA双链断裂。与Cas9相比,NgAgo-gDNA系统不需要原间隔物邻近基序(PAM)。使用无核酸酶活性的NgAgo(dNgAgo)可以大大扩展可以靶向的碱基。NgAgo的表征和用途已经记载于Gao et al.,Nat Biotechnol.,2016 Jul;34(7):768-73.PubMed PMID:27136078;Swarts et al.,Nature.507(7491)(2014):258-61;和Swarts et al.,Nucleic Acids Res.43(10)(2015):5120-9,每篇通过

引用并入本文。SEQ ID NO:38中提供了格氏嗜盐碱杆菌Argonaute的序列。

[0149] 野生型格氏嗜盐碱杆菌Argonaute(SEQ ID NO:38)

[0150] MTVIDLDSTTTADELTSGHTYDISVTLTGVYDNTDEQHPRMSLAFEQDNGERRYITLWKNTTPKDVFT
YDYATGSTYIFTNIDYEVKDGYENLTATYQTTVENATAQEVGTTDEDETFAGGEPLDHHLDDALNETPDDAETESD
SGHVMTSFASRDQLPEWTLHTYTLTATDGAKTDTEYARRTLAYTVRQELYTDHDAAPVATDGLMLLTPEPLGETPL
DLDCGVRVEADETRTLDYTTAKDRLLARELVEEGLKRSLWDDYLVRGIDEVLSKEPVLTCDEFDLHERYDLSVEVG
HSGRAYLHINFRHRFVPKLTLADIDDDNIYPGLRVKTTYRPRRGHIVWGLRDECATDSLNTLGNQSVVAYHRNNQT
PINTDLLDAIEAADRRVVETRRQGHGDDAVSFPQELLAVEPNTHQIKQFASDGFHQQARSKTRLSASRCSEKAQAF
AERLDPVRLNGSTVEFSSEFFTGNNEQQLRLLYENGESVLTFRDGARGAHPDETFSKGIVNPPESFEVAVVLPEQQ
ADTCKAQWDTMADLLNQAGAPPTRSETVQYDAFSSPESISLNVAGAIDPSEVDAAFVVLPPDQEGFADLASPTETY
DELKKALANMGIYSQMAYFDRFRDAKIFYTRNVALGLLAAAGGVAFTTEHAMPGDADMFIGIDVSRSYPEDGASGQ
INIAATATAVYKDGTILGHSSTRPQLGEKLQSTDVRDIMKNAILGYQQVTGESPTHIVIHRDGFMNEDLDPATEFL
NEQGVEYDIVEIRKQPQTRLLAVSDVQYDTPVKSIAAINQNEPRATVATFGAPEYLATRDGGGLPRPIQIERVAGE
TDIETLTRQVYLLSQSHIQVHNSTARLPITTAYADQASTHATKGYLVQTGAFESNVGFL(SEQ ID NO:38)

[0151] 在一些实施方案中,napDNAbp是Argonaute蛋白的原核同源物。Argonaute蛋白的原核同源物是已知的,并且已经记载于例如Makarova K.,et al.,“Prokaryotic homologs of Argonaute proteins are predicted to function as key components of a novel system of defense against mobile genetic elements”,Biol Direct.2009 Aug 25;4:29.doi:10.1186/1745-6150-4-29,其全部内容通过引用并入本文。在一些实施方案中,napDNAbp是Marinitoga piezophila Argunaute(MpAgo)蛋白。CRISPR相关的Marinitoga piezophila Argunaute(MpAgo)蛋白使用5’磷酸化的引导切割单链靶序列。所有已知的Argonautes都使用5’引导。MpAgo-RNA复合物的晶体结构显示了一个引导链结合位点,该位点包含阻断5’磷酸相互作用的残基。此数据表明对5’-羟基化引导具有非规范特异性的Argonaute亚类的进化。参见例如Kaya et al.,“A bacterial Argonaute with noncanonical guide RNA specificity”,Proc Natl Acad Sci U S A.2016 Apr 12;113(15):4057-62,其全部内容通过引用并入本文。应当理解,可以使用其他argonaute蛋白,并且它们在本公开内容的范围内。

[0152] 在一些实施方案中,核酸可编程DNA结合蛋白(napDNAbp)是微生物CRISPR-Cas系统的单一效应器。微生物CRISPR-Cas系统的单一效应器包括但不限于Cas9,Cpf1,C2c1,C2c2和C2c3。通常,微生物CRISPR-Cas系统分为1类和2类系统。1类系统具有多亚基效应器复合物,而2类系统具有单一蛋白质效应器。例如,Cas9和Cpf1是2类效应器。除Cas9和Cpf1外,Shmakov et al.,“Discovery and Functional Characterization of Diverse Class 2CRISPR Cas Systems”,Mol.Cell,2015Nov 5;60(3):385-397(其全部内容通过引用并入本文)已经描述了三个不同的2类CRISPR-Cas系统(C2c1,C2c2和C2c3)。系统中的两种(C2c1和C2c3)的效应器含有与Cpf1相关的RuvC样内切核酸酶域。第三个系统C2c2包含具有两个预测的HEPN RNA酶域的效应器。与C2c1产生CRISPR RNA不同,成熟的CRISPR RNA的产生不依赖于tracrRNA。C2c1依赖于CRISPR RNA和tracrRNA两者进行DNA切割。已经显示了细菌C2c2拥有用于CRISPR RNA成熟的独特的RNA酶活性,不同于其RNA激活的单链RNA降解活性。这些RNA酶功能彼此不同并且与Cpf1的CRISPR RNA加工行为不同。参见例如East-

Seletsky,et al.,“Two distinct RNase activities of CRISPR-C2c2 enable guide-RNA processing and RNA detection”,Nature,2016Oct 13;538(7624):270-273,其全部内容通过引用并入本文。Leptotrichia shahii中C2c2的体外生化分析显示了C2c2由单个CRISPR RNA引导,并可被编程为切割携带互补原间隔物的ssRNA靶物。两个保守的HEPN域中的催化残基介导切割。催化残基中的突变产生无催化活性的RNA结合蛋白。参见例如Abudayyeh et al.,“C2c2 is a single-component programmable RNA-guided RNA-targeting CRISPR effector”,Science,2016Aug 5;353(6299),其全部内容通过引用并入本文。

[0153] 已经报道了与嵌合单分子引导RNA(sgRNA)复合的酸土脂环酸芽孢杆菌(Alicyclobaccillus acidoterrastris)C2c1(AacC2c1)的晶体结构。参见例如“C2c1-sgRNA Complex Structure Reveals RNA-Guided DNA Cleavage Mechanism”,Mol.Cell,2017Jan 19;65(2):310-322,其全部内容在此通过引用并入。还已经报道了在与靶DNA结合的酸土脂环酸芽孢杆菌C2c1中作为三元复合物的晶体结构。参见例如Yang et al.,“PAM-dependent Target DNA Recognition and Cleavage by C2C1 CRISPR-Cas endonuclease”,Cell,2016Dec 15;167(7):1814-1828,其全部内容在此通过引用并入。具有靶DNA链和非靶DNA链两者的AacC2c1的催化能力构象已被独立地捕获,定位在单一RuvC催化口袋内,C2c1介导的切割导致靶DNA的交错的七核苷酸断裂。C2c1三元复合物与先前鉴定的Cas9和Cpf1对应物之间的结构比较证明了CRISPR-Cas9系统使用的机制的多样性。

[0154] 在一些实施方案中,本文提供的任何融合蛋白的核酸可编程DNA结合蛋白(napDNAbp)可以是C2c1,C2c2或C2c3蛋白。在一些实施方案中,napDNAbp是C2c1蛋白。在一些实施方案中,napDNAbp是C2c2蛋白。在一些实施方案中,napDNAbp是C2c3蛋白。在一些实施方案中,napDNAbp包含与天然存在的C2c1,C2c2或C2c3蛋白至少85%,至少90%,至少91%,至少92%,至少93%,至少94%,至少95%,至少96%,至少97%,至少98%,至少99%或至少99.5%相同的氨基酸序列。在一些实施方案中,napDNAbp是天然存在的C2c1,C2c2或C2c3蛋白。在一些实施方案中,napDNAbp包含与SEQ ID NO:39-40中的任何一项至少85%,至少90%,至少91%,至少92%,至少93%,至少94%,至少95%,至少96%,至少97%,至少98%,至少99%或至少99.5%相同的氨基酸序列。应当理解,根据本公开内容,也可以使用来自其他细菌物种的C2c1,C2c2或C2c3。

[0155] C2c1(uniprot.org/uniprot/T0D7A2#)

[0156] sp|T0D7A2|C2C1_ALIAG CRISPR相关内切核酸酶C2c1 OS=酸土脂环酸芽孢杆菌(菌株ATCC 49025/DSM 3922/CIP 106132/NCIMB 13137/GD3B)GN=c2c1 PE=1SV=1

[0157] MAVKSIKVKLRLDDMPEIRAGLWKLHKEVNAGVRYYTEWLSLLRQENLYRRSPNGDGEQECDKTAEEC
KAELLERLRARQVENGHRGPAGSDDELLQLARQLYELLVPQAIGAKGDAQQIARKFLSPLADKDAVGGLGIAKAGN
KPRWVRMREAGEPGWEEEKEKAETRKSADRTADVLRALADFGLKPLMRVYTDSEMSSVEWKPLRKGQAVRTWDRDM
FQQAIERMMSWESWNQRVGQEYAKLVEQKNRFEQKNFVGQEHLVHLVNQLQQDMKEASPGLESKEQTAHYVTGRAL
RGSDKVFEKWGKLAPDAPFDLYDAEIKNVQRRNTRRFGSHDLFAKLAEPEYQALWREDASFLTRYAVYNSILRKLN
HAKMFATFTLPDATAHPIWTRFDKLGGNLHQYTFLFNEFGERRHAIRFHKLLKVENGVAREVDDVTVPISMSEQLD
NLLPRDPNEPIALYFRDYGAEQHFTGEFGGAKIQCRRDQLAHMHRRRGARDVYLNVSVRVQSQSEARGERRPPYAA
VFRLVGDNHRAFVHFDKLSDYLAEHPDDGKLGSEGLLSGLRVMSVDLGLRTSASISVFRVARKDELKPNSKGRVPF

FFPIKGNDNLVAVHERSQLLKLPGETESKDLRAIREERQRTLRQLRTQLAYLRLLVRCGSEDVGRRERSWAKLIEQ
PVDAANHMTPDWREAFENELQKLKSLHGICSDKEWMDAVYESVRRVWRHMGKQVRDWRKDVRSGERPKIRGYAKDV
VGGNSIEQIEYLERQYKFLKSWSFFGKVSGQVIRAEKGSRFAITLREHIDHAKEDRLKKLADRIIMEALGYVYALD
ERGKGKWVAKYPPCQLILLEELSEYQFNNDRPPSENNQLMQWSHRGVFQELINQAQVHDLLVGTMYAAFSSRFDAR
TGAPGIRCRRVPARCTQEHNPEPFPWWLNKFVVEHTLDACPLRADDLIPTGEGEIFVSPFSAEEGDFHQIHADLNA
AQNLQQRLWSDFDISQIRLRCDWGEVDGELVLIPRLTGKRTADSYSNKVFYTNTGVTYYERERGKKRRKVFAQEKL
SEEEAELLVEADEAREKSVVLMRDPSGIINRGNWTRQKEFWSMVNQRIEGYLVKQIRSRVPLQDSACENTGDI(SEQ
ID NO:39)

[0158] C2c2(uniprot.org/uniprot/P0DOC6)

[0159] >sp|P0DOC6|C2C2_LEPSD CRISPR-相关内切核糖核酸酶C2c2OS＝Leptotrichia shahii(菌株DSM 19757/CCUG 47503/CIP 107916/JCM16776/LB37)GN＝c2c2 PE＝1SV＝1

[0160] MGNLFGHKRWYEVRDKKDFKIKRKVKVKRNYDGNKYILNINENNNKEKIDNNKFIRKYINYKKNDNIL
KEFTRKFHAGNILFKLKGKEGIIRIENNDDFLETEEVVLYIEAYGKSEKLKALGITKKKIIDEAIRQGITKDDKKI
EIKRQENEEEIEIDIRDEYTNKTLNDCSIILRIIENDELETKKSIYEIFKNINMSLYKIIEKIIENETEKVFENRY
YEEHLREKLLKDDKIDVILTNFMEIREKIKSNLEILGFVKFYLNVGGDKKKSKNKKMLVEKILNINVDLTVEDIAD
FVIKELEFWNITKRIEKVKKVNNEFLEKRRNRTYIKSYVLLDKHEKFKIERENKKDKIVKFFVENIKNNSIKEKIE
KILAEFKIDELIKKLEKELKKGNCDTEIFGIFKKHYKVNFDSKKFSKKSDEEKELYKIIYRYLKGRIEKILVNEQK
VRLKKMEKIEIEKILNESILSEKILKRVKQYTLEHIMYLGKLRHNDIDMTTVNTDDFSRLHAKEELDLELITFFAS
TNMELNKIFSRENINNDENIDFFGGDREKNYVLDKKILNSKIKIIRDLDFIDNKNNITNNFIRKFTKIGTNERNRI
LHAISKERDLQGTQDDYNKVINIIQNLKISDEEVSKALNLDVVFKDKKNIITKINDIKISEENNNDIKYLPSFSKV
LPEILNLYRNNPKNEPFDTIETEKIVLNALIYVNKELYKKLILEDDLEENESKNIFLQELKKTLGNIDEIDENIIE
NYYKNAQISASKGNNKAIKKYQKKVIECYIGYLRKNYEELFDFSDFKMNIQEIKKQIKDINDNKTYERITVKTSDK
TIVINDDFEYIISIFALLNSNAVINKIRNRFFATSVWLNTSEYQNIIDILDEIMQLNTLRNECITENWNLNLEEFI
QKMKEIEKDFDDFKIQTKKEIFNNYYEDIKNNILTEFKDDINGCDVLEKKLEKIVIFDDETKFEIDKKSNILQDEQ
RKLSNINKKDLKKKVDQYIKDKDQEIKSKILCRIIFNSDFLKKYKKEIDNLIEDMESENENKFQEIYYPKERKNEL
YIYKKNLFLNIGNPNFDKIYGLISNDIKMADAKFLFNIDGKNIRKNKISEIDAILKNLNDKLNGYSKEYKEKYIKK
LKENDDFFAKNIQNKNYKSFEKDYNRVSEYKKIRDLVEFNYLNKIESYLIDINWKLAIQMARFERDMHYIVNGLRE
LGIIKLSGYNTGISRAYPKRNGSDGFYTTTAYYKFFDEESYKKFEKICYGFGIDLSENSEINKPENESIRNYISHF
YIVRNPFADYSIAEQIDRVSNLLSYSTRYNNSTYASVFEVFKKDVNLDYDELKKKFKLIGNNDILERLMKPKKVSV
LELESYNSDYIKNLIIELLTKIENTNDTL(SEQ ID NO:40)

[0161] 核碱基编辑器的Cas9域

[0162] 在一些方面,核酸可编程DNA结合蛋白(napDNAbp)是Cas9域。本文提供了非限制性的示例性Cas9域。Cas9域可以是核酸酶活性Cas9域、无核酸酶活性Cas9域或Cas9切口酶。在一些实施方案中,Cas9域是核酸酶活性域。例如,Cas9域可以是切割双链核酸的两条链(例如,双链DNA分子的两条链)的Cas9域。在一些实施方案中,Cas9域包含如SEQ ID NO:4-29中所示的氨基酸序列的任一个。在一些实施方案中Cas9域包含与本文中提供的Cas9或与SEQ ID NO:4-29中所示的氨基酸序列的任一个至少60%、至少65%、至少70%、至少75%、至少80%、至少85%、至少90%、至少95%、至少96%、至少97%、至少98%、至少99%或至少99.5%相同的氨基酸序列。在一些实施方案中,Cas9域包含与本文中提供的Cas9或与SEQ

ID NO:4-29中所示的任一个氨基酸序列相比具有1、2、3、4、5、6、7、8、9、10、11、12、13、14、15、16、17、18、19、20、21、22、21、24、25、26、27、28、29、30、31、32、33、34、35、36、37、38、39、40、41、42、43、44、45、46、47、48、49、50个或更多个突变的氨基酸序列。在一些实施方案中，Cas9域包含具有与本文中提供的Cas9或与SEQ ID NO:4-29中所示的任一个氨基酸序列相比具有至少10、至少15、至少20、至少30、至少40、至少50、至少60、至少70、至少80、至少90、至少100、至少150、至少200、至少250、至少300、至少350、至少400、至少500、至少600、至少700、至少800、至少900、至少1000、至少1100、或至少1200个相同的连续氨基酸残基的氨基酸序列。

[0163] 在一些实施方案中，Cas9域是无核酸酶活性的Cas9域(dCas9)。例如，dCas9域可以结合双链核酸分子(例如，经由gRNA分子)而不切割双链核酸分子的任一条链。在一些实施方案中，无核酸酶活性dCas9域包含SEQ ID NO:6中所示的氨基酸序列的D10X突变和H840X突变，或本文中提供的任何Cas9，诸如SEQ ID NO:4-26中所提供的氨基酸序列之一中的相应的突变，其中X是任何氨基酸变化。在一些实施方案中，无核酸酶活性dCas9域包含SEQ ID NO:6中所示的氨基酸序列的D10A突变和H840A突变，或本文中提供的任何Cas9，诸如SEQ ID NO:4-26中所提供的任一项氨基酸序列中的相应的突变。作为一个实例，无核酸酶活性的Cas9域包含SEQ ID NO:9中所示的氨基酸序列(克隆载体pPlatTET-gRNA2，登录号BAV54124)。

[0164] MDKKYSIGLAIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETAEATRLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFFHRLEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFEENPINASGVDAKAILSARLSKSRRLENLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAILLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLRKQRTFDNGSIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPYYVGPLARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQSFIERMTNFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKIIKDKDFLDNEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFANRNFMQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKVVDELVKVMGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVDAIVPQSFLKDDSIDNKVLTRSDKNRGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKAGFIKRQLVETRQITKHVAQILDSRMNTKYDENDKLIREVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYLNAVVGTALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQEIGKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEVQTGGFSKESILPKRNSDKLIARKKDWDPKKYGGFDSPTVAYSVLVVAKVEKGKSKKLKSVKELLGITIMERSSFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFELENGRKRMLASAGELQKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQHKHYLDEIIEQISEFSKRVILADANLDKVLSAYNKHRDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVLDATLIHQSITGLYETRIDLSQLGGD(SEQ ID NO:9;参见例如Qi et al.,

[0165] “Repurposing CRISPR as an RNA-guided platform for sequence-specific control of gene expression.”Cell.2013;152(5):1173-83,其全部内容通过引用并入本文)。

[0166] 基于本公开内容和本领域的知识，另外的合适的无核酸酶活性dCas9域对于本领域技术人员将是显而易见的，并且在本公开内容的范围内。此类另外的示例性合适的无核酸酶活性Cas9域包括但不限于D10A/H840A、D10A/D839A/H840A和D10A/D839A/H840A/N863A突变体域(参见例如Prashant et al.,CAS9 transcriptional activators for target specificity screening and paired nickases for cooperative genome engineering.Nature Biotechnology.2013;31(9):833-838，其全部内容通过引用并入本文)。在一些实施方案中，dCas9域包含与本文提供的任一个dCas9域至少60％、至少65％、至少70％、至少75％、至少80％、至少85％、至少90％、至少95％、至少96％、至少97％、至少98％、至少99％或至少99.5％相同的氨基酸序列。在一些实施方案中，Cas9域包含与SEQ ID NO:7,8,9,或22中所示的任一个氨基酸序列相比，具有1、2、3、4、5、6、7、8、9、10、11、12、13、14、15、16、17、18、19、20、21、22、21、24、25、26、27、28、29、30、31、32、33、34、35、36、37、38、39、40、41、42、43、44、45、46、47、48、49、50个或更多个突变的氨基酸序列。在一些实施方案中，Cas9域包含与SEQ ID NO:7,8,9,或22中所示的任一个氨基酸序列相比，具有至少10个、至少15个、至少20个、至少30个、至少40个、至少50个、至少60个、至少70个、至少80个、至少90个、至少100个、至少150个、至少200个、至少250个、至少300个、至少350个、至少400个、至少500个、至少600个、至少700个、至少800个、至少900个、至少1000个、至少1100个、或至少1200个相同的连续氨基酸残基的氨基酸序列。

[0167] 在一些实施方案中，Cas9域是Cas9切口酶。Cas9切口酶可以是能够仅切割双链核酸分子(例如双链DNA分子)的一条链的Cas9蛋白。在一些实施方案中，Cas9切口酶切割双链核酸分子的靶链，这意味着Cas9切口酶切割与结合到Cas9的gRNA(例如，sgRNA)碱基配对(互补)的链。在一些实施方案中，Cas9切口酶包含D10A突变并且具有SEQ ID NO:6的位置840处的组氨酸，或在本文提供的任何Cas9，诸如SEQ ID NO:4-26中的任一个中的突变。例如，Cas9切口酶可以包含如SEQ ID NO:10,13,16,或21中所示的氨基酸序列。在一些实施方案中，Cas9切口酶切割双链核酸分子的非靶标、非碱基编辑链，这意味着Cas9切口酶切割不与结合到Cas9的gRNA(例如，sgRNA)碱基配对的链。在一些实施方案中，Cas9切口酶包含H840A突变并且具有在SEQ ID NO:6的位置10处的天冬氨酸残基或在本文提供的任何Cas9，诸如SEQ ID NO:4-26中的任一个中的相应突变。在一些实施方案中，Cas9切口酶包含与本文提供的任一个Cas9切口酶至少60％、至少65％、至少70％、至少75％、至少80％、至少85％、至少90％、至少95％、至少96％、至少97％、至少98％、至少99％或至少99.5％相同的氨基酸序列。基于本公开内容和本领域的知识，另外的合适的Cas9切口酶对于本领域技术人员将是显而易见的，并且在本公开内容的范围内。

[0168] 具有降低的PAM排他性(exclusivity)的Cas9域

[0169] 本公开内容的一些方面提供了具有不同PAM特异性的Cas9域。通常，Cas9蛋白，例如来自酿脓链球菌的Cas9(spCas9)，需要规范的NGG PAM序列来结合特定的核酸区域，其中“NGG”中的“N”是腺嘌呤(A)、胸腺嘧啶(T)、鸟嘌呤(G)或胞嘧啶(C)，并且G是鸟嘌呤。这可以限制在基因组内编辑期望的碱基的能力。在一些实施方案中，本文提供的碱基编辑融合蛋白需要定位于精确的位置处，例如，其中靶碱基在4碱基区域(例如“脱氨基窗”)内，其在PAM的上游的约15个碱基。参见Komor,A.C.,et al.,“Programmable editing of a target base in genomic DNA without double-stranded DNA cleavage”Nature 533,420-424

(2016),其全部内容在此通过引用并入。在一些实施方案中,脱氨基窗在2、3、4、5、6、7、8、9或10碱基区域内。在一些实施方案中,脱氨基窗在PAM上游的5、6、7、8、9、10、11、12、13、14、15、16、17、18、19、20、21、22、23、24或25个碱基。因此,在一些实施方案中,本文提供的任何融合蛋白可以含有能够结合不含规范的(例如,NGG)PAM序列的核苷酸序列的Cas9域。本领域已经描述了结合非规范PAM序列的Cas9域,并且其对于熟练技术人员而言是显而易见的。例如,结合非规范PAM序列的Cas9域已经描述于Kleinstiver,B.P.,et al.,"Engineered CRISPR-Cas9 nucleases with altered PAM specificities"Nature 523,481-485(2015);和Kleinstiver,B.P.,et al.,"Broadening the targeting range of Staphylococcus aureus CRISPR-Cas9 by modifying PAM recognition"Nature Biotechnology 33,1293-1298(2015);每篇的全部内容在此通过引用并入。

[0170] 在一些实施方案中,Cas9域是来自金黄色葡萄球菌的Cas9域(SaCas9)。在一些实施方案中,SaCas9域是核酸酶活性的SaCas9、无核酸酶活性的SaCas9(SaCas9d)或SaCas9切口酶(SaCas9n)。在一些实施方案中,SaCas9包含氨基酸序列SEQ ID NO:12。在一些实施方案中,SaCas9包含SEQ ID NO:12的N579X突变或SEQ ID NO:13-14中提供的任何氨基酸序列中的相应的突变,其中X是除N之外的任何氨基酸。在一些实施方案中,SaCas9包含SEQ ID NO:12的N579A突变或SEQ ID NO:13-14中提供的任何氨基酸序列中的相应的突变。

[0171] 在一些实施方案中,SaCas9域、SaCas9d域或SaCas9n域可以结合具有非规范PAM的核酸序列。在一些实施方案中,SaCas9域、SaCas9d域或SaCas9n域可以结合具有NNGRRT PAM序列的核酸序列,其中N=A、T、C或G,并且R=A或G。在一些实施方案中,SaCas9域包含SEQ ID NO:12的E781X、N967X和R1014X突变中的一个或多个,或SEQ ID NO:13-14中提供的任何氨基酸序列中的相应的突变,其中X是任何氨基酸。在一些实施方案中,SaCas9域包含SEQ ID NO:12中的E781K、N967K和R1014H突变,或者在SEQ ID NO:13-14中提供的任何氨基酸序列中的一个或多个相应的突变中的一个或多个。在一些实施方案中,SaCas9域包含SEQ ID NO:12的E781K、N967K或R1014H突变或在SEQ ID NO:13-14中提供的任何氨基酸序列中的相应的突变。

[0172] 在一些实施方案中,本文提供的任何融合蛋白的Cas9域包含与SEQ ID NO:12-14中的任一个至少60%、至少65%、至少70%、至少75%、至少80%、至少85%、至少90%、至少95%、至少96%、至少97%、至少98%、至少99%或至少99.5%相同的氨基酸序列。在一些实施方案中,本文提供的任何融合蛋白的Cas9域包含SEQ ID NO:12-14中任一个的氨基酸序列。在一些实施方案中,本文提供的任何融合蛋白的Cas9域由SEQ ID NO:12-14中任一个的氨基酸序列组成。

[0173] 示例性SaCas9序列

[0174] KRNYILGLDIGITSVGYGIIDYETRDVIDAGVRLFKEANVENNEGRRSKRGARRLKRRRRHRIQRVK
KLLFDYNLLTDHSELSGINPYEARVKGLSQKLSEEEFSAALLHLAKRRGVHNVNEVEEDTGNELSTK
EQISRNSKALEEKYVAELQLERLKKDGEVRGSINRFKTSDYVKEAKQLLKVQKAYHQLDQSFIDTYI
DLLETRRTYYEGPGEGSPFGWKDIKEWYEMLMGHCTYFPEELRSVKYAYNADLYNALNDLNNLVI
TRDENEKLEYYEKFQIIENVFKQKKKPTLKQIAKEILVNEEDIKGYRVTSTGKPEFTNLKVYHDIKDIT
ARKEIIENAELLDQIAKILTIYQSSEDIQEELTNLNSELTQEEIEQISNLKGYTGTHNLSLKAINLILDEL
WHTNDNQIAIFNRLKLVPKKVDLSQQKEIPTTLVDDFILSPVVKRSFIQSIKVINAIIKKYGLPNDIIIEL
AREKNSKDAQKMINEMQKRNRQTNERIEEIIRTTGKENAKYLIEKIKLHDMQEGKCLYSLEAIPLED
LLNNPFNYEVDHIIPRSVSFDNSFNNKVLVKQEENSKKGNRTPFQYLSSSDSKISYETFKKHILNLAKG
KGRISKTKKEYLLEERDINRFSVQKDFINRNLVDTRYATRGLMNLLRSYFRVNNLDVKVKSINGGFT
SFLRRKWKFKKERNKGYKHHAEDALIIANADFIFKEWKKLDKAKKVMENQMFEEKQAESMPEIET
EQEYKEIFITPHQIKHIKDFKDYKYSHRVDKKPNRELINDTLYSTRKDDKGNTLIVNNLNGLYDKDN
DKLKKLINKSPEKLLMYHHDPQTYQKLKLIMEQYGDEKNPLYKYYEETGNYLTKYSKKDNGPVIK
[0175] KIKYYGNKLNAHLDITDDYPNSRNKVVKLSLKPYRFDVYLDNGVYKFVTVKNLDVIKKENYYEVN
SKCYEEAKKLKKISNQAEFIASFYNNDLIKINGELYRVIGVNNDLLNRIEVNMIDITYREYLENMNDK
RPPRIIKTIASKTQSIKKYSTDILGNLYEVKSKKHPQIIKKG (SEQ ID NO: 12)

[0176] 可以将SEQ ID NO:12的残基N579(其是加下划线且粗体的)突变(例如突变为A579)以产生SaCas9切口酶。

[0177] 示例性SaCas9n序列

[0178] KRNYILGLDIGITSVGYGIIDYETRDVIDAGVRLFKEANVENNEGRRSKRGARRLKRRRRHRIQRVK
KLLFDYNLLTDHSELSGINPYEARVKGLSQKLSEEEFSAALLHLAKRRGVHNVNEVEEDTGNELSTK
EQISRNSKALEEKYVAELQLERLKKDGEVRGSINRFKTSDYVKEAKQLLKVQKAYHQLDQSFIDTYI
DLLETRRTYYEGPGEGSPFGWKDIKEWYEMLMGHCTYFPEELRSVKYAYNADLYNALNDLNNLVI
TRDENEKLEYYEKFQIIENVFKQKKKPTLKQIAKEILVNEEDIKGYRVTSTGKPEFTNLKVYHDIKDIT
ARKEIIENAELLDQIAKILTIYQSSEDIQEELTNLNSELTQEEIEQISNLKGYTGTHNLSLKAINLILDEL
WHTNDNQIAIFNRLKLVPKKVDLSQQKEIPTTLVDDFILSPVVKRSFIQSIKVINAIIKKYGLPNDIIIEL
AREKNSKDAQKMINEMQKRNRQTNERIEEIIRTTGKENAKYLIEKIKLHDMQEGKCLYSLEAIPLED
LLNNPFNYEVDHIIPRSVSFDNSFNNKVLVKQEEASKKGNRTPFQYLSSSDSKISYETFKKHILNLAKG
KGRISKTKKEYLLEERDINRFSVQKDFINRNLVDTRYATRGLMNLLRSYFRVNNLDVKVKSINGGFT
SFLRRKWKFKKERNKGYKHHAEDALIIANADFIFKEWKKLDKAKKVMENQMFEEKQAESMPEIET
EQEYKEIFITPHQIKHIKDFKDYKYSHRVDKKPNRELINDTLYSTRKDDKGNTLIVNNLNGLYDKDN
DKLKKLINKSPEKLLMYHHDPQTYQKLKLIMEQYGDEKNPLYKYYEETGNYLTKYSKKDNGPVIK
KIKYYGNKLNAHLDITDDYPNSRNKVVKLSLKPYRFDVYLDNGVYKFVTVKNLDVIKKENYYEVN
SKCYEEAKKLKKISNQAEFIASFYNNDLIKINGELYRVIGVNNDLLNRIEVNMIDITYREYLENMNDK
RPPRIIKTIASKTQSIKKYSTDILGNLYEVKSKKHPQIIKKG (SEQ ID NO: 13)。

[0179] SEQ ID NO:13的残基A579(其可以从SEQ ID NO:12的N579突变以产生SaCas9切口酶)是加下划线且粗体的。

[0180] 示例性SaKKH Cas9

[0181] KRNYILGLDIGITSVGYGIIDYETRDVIDAGVRLFKEANVENNEGRRSKRGARRLKRRRRHRIQRVK
KLLFDYNLLTDHSELSGINPYEARVKGLSQKLSEEEFSAALLHLAKRRGVHNVNEVEEDTGNELSTK
EQISRNSKALEEKYVAELQLERLKKDGEVRGSINRFKTSDYVKEAKQLLKVQKAYHQLDQSFIDTYI
DLLETRRTYYEGPGEGSPFGWKDIKEWYEMLMGHCTYFPEELRSVKYAYNADLYNALNDLNNLVI
TRDENEKLEYYEKFQIIENVFKQKKKPTLKQIAKEILVNEEDIKGYRVTSTGKPEFTNLKVYHDIKDIT
ARKEIIENAELLDQIAKILTIYQSSEDIQEELTNLNSELTQEEIEQISNLKGYTGTHNLSLKAINLILDEL
WHTNDNQIAIFNRLKLVPKKVDLSQQKEIPTTLVDDFILSPVVKRSFIQSIKVINAIIKKYGLPNDIIIEL
AREKNSKDAQKMINEMQKRNRQTNERIEEIIRTTGKENAKYLIEKIKLHDMQEGKCLYSLEAIPLED
LLNNPFNYEVDHIIPRSVSFDNSFNNKVLVKQEEASKKGNRTPFQYLSSSDSKISYETFKKHILNLAKG
KGRISKTKKEYLLEERDINRFSVQKDFINRNLVDTRYATRGLMNLLRSYFRVNNLDVKVKSINGGFT
SFLRRKWKFKKERNKGYKHHAEDALIIANADFIFKEWKKLDKAKKVMENQMFEEKQAESMPEIET
EQEYKEIFITPHQIKHIKDFKDYKYSHRVDKKPNRKLINDTLYSTRKDDKGNTLIVNNLNGLYDKDN
DKLKKLINKSPEKLLMYHHDPQTYQKLKLIMEQYGDEKNPLYKYYEETGNYLTKYSKKDNGPVIK
KIKYYGNKLNAHLDITDDYPNSRNKVVKLSLKPYRFDVYLDNGVYKFVTVKNLDVIKKENYYEVN
SKCYEEAKKLKKISNQAEFIASFYKNDLIKINGELYRVIGVNNDLLNRIEVNMIDITYREYLENMNDK
RPPHIIKTIASKTQSIKKYSTDILGNLYEVKSKKHPQIIKKG (SEQ ID NO: 14)。

[0182] SEQ ID NO:14的残基A579(其可以从SEQ ID NO:12的N579突变以产生SaCas9切口酶)是加下划线且粗体的。SEQ ID NO:14的残基K781、K967和H1014(其可以从SEQ ID NO:12的E781、N967和R1014突变以产生SaKKH Cas9)是加下划线且斜体的。

[0183] 在一些实施方案中,Cas9域是来自酿脓链球菌的Cas9域(SpCas9)。在一些实施方案中,SpCas9域是核酸酶活性的SpCas9、无核酸酶活性的SpCas9(SpCas9d)或SpCas9切口酶(SpCas9n)。在一些实施方案中,SpCas9包含氨基酸序列SEQ ID NO:15。在一些实施方案中,SpCas9包含SEQ ID NO:15的D9X突变或在任何Cas9,诸如SEQ ID NO:4-26中提供的任何氨基酸序列中的相应的突变,其中X是除D之外的任何氨基酸。在一些实施方案中,SpCas9包含SEQ ID NO:15的D9A突变或在本文提供的任何Cas9,诸如SEQ ID NO:4-26中提供的任何氨基酸序列中的相应的突变。在一些实施方案中,SpCas9域、SpCas9d域或SpCas9n域可以结合具有非规范PAM的核酸序列。在一些实施方案中,SpCas9域、SpCas9d域或SpCas9n域可以结合具有NGG、NGA或NGCG PAM序列的核酸序列。在一些实施方案中,SpCas9域包含SEQ ID NO:15的D1134X、R1334X和T1336X突变或SEQ ID NO:108-35中提供的任何氨基酸序列中的相应的突变中的一个或多个,其中X是任何氨基酸。在一些实施方案中,SpCas9域包含SEQ ID NO:15的D1134E、R1334Q和T1336R突变,或在本文提供的任何Cas9,诸如SEQ ID NO:4-26中提供的任何氨基酸序列中的相应的突变中的一个或多个。在一些实施方案中,SpCas9域包含SEQ ID NO:15的D1134E、R1334Q和T1336R突变或在本文提供的任何Cas9,诸如SEQ ID NO:4-26中提供的任何氨基酸序列中的相应的突变。在一些实施方案中,SpCas9域包含SEQ ID NO:15的D1134X、R1334X和T1336X突变或SEQ ID NO:108-35中提供的任何氨基酸序列中的相应的突变中的一个或多个,其中X是任何氨基酸。在一些实施方案中,SpCas9域包含SEQ ID NO:15的D1134V、R1334Q和T1336R突变,或在本文提供的任何Cas9,诸如SEQ ID NO:4-26中提供的任何氨基酸序列中的相应的突变中的一个或多个。在一些实施方案中,SpCas9域包含SEQ ID NO:15的D1134V、R1334Q和T1336R突变,或在本文提供的任何Cas9,诸如SEQ ID NO:4-26中提供的任何氨基酸序列中的相应的突变。在一些实施方案中,SpCas9域包含SEQ ID NO:15的D1134X、G1217X、R1334X和T1336X突变,或SEQ ID NO:108-35中提供的任何氨基酸序列中的相应的突变中的一个或多个,其中X是任何氨基酸。在一些实施方案中,SpCas9域包含SEQ ID NO:15的D1134V、G1217R、R1334Q和T1336R突变,或在本文提供的任何Cas9,

诸如SEQ ID NO:4-26中提供的任何氨基酸序列中的相应的突变中的一个或多个。在一些实施方案中,SpCas9域包含SEQ ID NO:15的D1134V、G1217R、R1334Q和T1336R突变或在本文提供的任何Cas9,诸如SEQ ID NO:4-26中提供的任何氨基酸序列中的相应的突变。

[0184] 在一些实施方案中,本文提供的任何融合蛋白的Cas9域包含与SEQ ID NO:15-19至少60%、至少65%、至少70%、至少75%、至少80%、至少85%、至少90%、至少95%、至少96%、至少97%、至少98%、至少99%或至少99.5%相同的氨基酸序列。在一些实施方案中,本文提供的任何融合蛋白的Cas9域包含SEQ ID NO:15-19的任一个的氨基酸序列。在一些实施方案中,本文提供的任何融合蛋白的Cas9域由SEQ ID NO:15-19的任一个的氨基酸序列组成。

[0185] 示例性SpCas9

[0186] DKKYSIGLDIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETAEATRLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFFHRLEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFEENPINASGVDAKAILSARLSKSRRLENLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAILLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLRKQRTFDNGSIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPYYVGPLARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQSFIERMTNFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKIIKDKDFLDNEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFANRNFMQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKVVDELVKVMGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVDHIVPQSFLKDDSIDNKVLTRSDKNRGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKAGFIKRQLVETRQITKHVAQILDSRMNTKYDENDKLIREVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYLNAVVGTALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQEIGKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEVQTGGFSKESILPKRNSDKLIARKKDWDPKKYGGFDSPTVAYSVLVVAKVEKGKSKKLKSVKELLGITIMERSSFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFELENGRKRMLASAGELQKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQHKHYLDEIIEQISEFSKRVILADANLDKVLSAYNKHRDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVLDATLIHQSITGLYETRIDLSQLGGD(SEQ ID NO:15)

[0187] 示例性SpCas9n

[0188] DKKYSIGLAIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETAEATRLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFFHRLEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFEENPINASGVDAKAILSARLSKSRRLENLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAILLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLRKQRTFDNGSIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPYYVGPLARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQSFIERMTNFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKIIKDKDFLDNEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGK

TILDFLKSDGFANRNFMQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKVVDELVKVMGR
HKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQELD
INRLSDYDVDHIVPQSFLKDDSIDNKVLTRSDKNRGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNLTKAER
GGLSELDKAGFIKRQLVETRQITKHVAQILDSRMNTKYDENDKLIREVKVITLKSKLVSDFRKDFQFYKVREINNY
HHAHDAYLNAVVGTALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQEIGKATAKYFFYSNIMNFFKTEITLANGEI
RKRPLIETNGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEVQTGGFSKESILPKRNSDKLIARKKDWDPKKYGG
FDSPTVAYSVLVVAKVEKGKSKKLKSVKELLGITIMERSSFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFELENG
RKRMLASAGELQKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQHKHYLDEIIEQISEFSKRVILADA
NLDKVLSAYNKHRDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVLDATLIHQSITGLYETRID
LSQLGGD(SEQ ID NO:16)

[0189] 示例性SpEQR Cas9

[0190] DKKYSIGLAIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETAE
ATRLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFFHRLEESFLVEEDKKHERHPIF
GNIVDEVAYHEKYPTIYHLRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIEGDLNPD
NSDVDKLFIQLVQTYNQLFEENPINASGVDAKAILSARLSKSRRLENLIAQLPGEKKNG
LFGNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAK
NLSDAILLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQ
SKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLRKQRTFDNGSIPHQ
IHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPYYVGPLARGNSRFAWMTRKSEETI
TPWNFEEVVDKGASAQSFIERMTNFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYV
TEGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFN
ASLGTYHDLLKIIKDKDFLDNEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKV
MKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFANRNFMQLIHDDSLTFK
EDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKVVDELVKVMGRHKPENIVIEM
ARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHPVENTQLQNEKLYLYYLQNGRD
MYVDQELDINRLSDYDVDHIVPQSFLKDDSIDNKVLTRSDKNRGKSDNVPSEEVVKK
MKNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKAGFIKRQLVETRQITKHVAQIL
DSRMNTKYDENDKLIREVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYLNAV
VGTALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQEIGKATAKYFFYSNIMNFFKTEIT
LANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEVQTGGFSKESI
LPKRNSDKLIARKKDWDPKKYGGF**E**SPTVAYSVLVVAKVEKGKSKKLKSVKELLGITI
MERSSFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFELENGRKRMLASAGELQKGNEL
ALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQHKHYLDEIIEQISEFSKRVILADA
NLDKVLSAYNKHRDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRK**Q**Y**R**STKEVLD
ATLIHQSITGLYETRIDLSQLGGD (SEQ ID NO: 17)

[0191] SEQ ID NO:17的残基E1134、Q1334和R1336(其可以从SEQ ID NO:15的D1134、R1334和T1336突变以产生SpEQR Cas9)是加下划线且粗体的。

[0192] 示例性SpVQR Cas9

[0193] DKKYSIGLAIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETAE
ATRLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFFHRLEESFLVEEDKKHERHPIF
GNIVDEVAYHEKYPTIYHLRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIEGDLNPD
NSDVDKLFIQLVQTYNQLFEENPINASGVDAKAILSARLSKSRRLENLIAQLPGEKKNG
LFGNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAK
NLSDAILLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQ
SKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLRKQRTFDNGSIPHQ
IHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPYYVGPLARGNSRFAWMTRKSEETI
TPWNFEEVVDKGASAQSFIERMTNFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYV

[0194] TEGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFN
ASLGTYHDLLKIIKDKDFLDNEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKV
MKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFANRNFMQLIHDDSLTFK
EDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKVVDELVKVMGRHKPENIVIEM
ARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHPVENTQLQNEKLYLYYLQNGRD
MYVDQELDINRLSDYDVDHIVPQSFLKDDSIDNKVLTRSDKNRGKSDNVPSEEVVKK
MKNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKAGFIKRQLVETRQITKHVAQIL
DSRMNTKYDENDKLIREVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYLNAV
VGTALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQEIGKATAKYFFYSNIMNFFKTEIT
LANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEVQTGGFSKESI
LPKRNSDKLIARKKDWDPKKYGGF**V**SPTVAYSVLVVAKVEKGKSKKLKSVKELLGITI
MERSSFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFELENGRKRMLASAGELQKGNEL
ALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQHKHYLDEIIEQISEFSKRVILADA
NLDKVLSAYNKHRDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRK**Q**Y**R**STKEVLD
ATLIHQSITGLYETRIDLSQLGGD (SEQ ID NO: 18)

[0195] SEQ ID NO:18的残基V1134、Q1334和R1336(其可以从SEQ ID NO:15的D1134、R1334和T1336突变以产生SpVQR Cas9)是加下划线且粗体的。

[0196] 示例性SpVRER Cas9

[0197] DKKYSIGLAIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETAE
ATRLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFFHRLEESFLVEEDKKHERHPIF
GNIVDEVAYHEKYPTIYHLRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIEGDLNPD
NSDVDKLFIQLVQTYNQLFEENPINASGVDAKAILSARLSKSRRLENLIAQLPGEKKNG
LFGNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAK
NLSDAILLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQ
SKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLRKQRTFDNGSIPHQ
IHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPYYVGPLARGNSRFAWMTRKSEETI
TPWNFEEVVDKGASAQSFIERMTNFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYV
TEGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFN
ASLGTYHDLLKIIKDKDFLDNEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKV
MKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFANRNFMQLIHDDSLTFK
EDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKVVDELVKVMGRHKPENIVIEM
ARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHPVENTQLQNEKLYLYYLQNGRD
MYVDQELDINRLSDYDVDHIVPQSFLKDDSIDNKVLTRSDKNRGKSDNVPSEEVVKK
MKNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKAGFIKRQLVETRQITKHVAQIL
DSRMNTKYDENDKLIREVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYLNAV
VGTALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQEIGKATAKYFFYSNIMNFFKTEIT
LANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEVQTGGFSKESI
LPKRNSDKLIARKKDWDPKKYGGF**V**SPTVAYSVLVVAKVEKGKSKKLKSVKELLGITI
MERSSFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFELENGRKRMLASA**R**ELQKGNEL
ALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQHKHYLDEIIEQISEFSKRVILADA
NLDKVLSAYNKHRDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRK**E**Y**R**STKEVLD
ATLIHQSITGLYETRIDLSQLGGD (SEQ ID NO: 19)

[0198] SEQ ID NO:19的残基V1134、R1217、Q1334和R1336(其可以从SEQ ID NO:15的D1134、G1217、R1334和T1336突变以产生SpVRER Cas9)是加下划线且粗体的。

[0199] 高保真性Cas9域

[0200] 本公开内容的一些方面提供了本文提供的核碱基编辑器的高保真性Cas9域。在一些实施方案中,与相应的野生型Cas9域相比,高保真性Cas9域是包含一个或多个突变的经工程化的Cas9域,所述突变降低Cas9域和DNA的糖-磷酸主链之间的静电相互作用。不希望受任何特定理论的束缚,具有降低的与DNA的糖-磷酸主链的静电相互作用的高保真性Cas9域可以具有较少的脱靶效应。在一些实施方案中,Cas9域(例如,野生型Cas9域)包含一个或

多个降低Cas9域与DNA的糖-磷酸主链之间的缔合的突变。在一些实施方案中，Cas9域包含一个或多个将Cas9域与DNA的糖-磷酸主链之间的缔合降低至少1％、至少2％、至少3％、至少4％、至少5％、至少10％、至少15％、至少20％、至少25％、至少30％、至少35％、至少40％、至少45％、至少50％、至少55％、至少60％、至少65％、至少70％或更多的突变。

[0201] 在一些实施方案中，本文提供的任何Cas9融合蛋白包含SEQ ID NO:6中提供的氨基酸序列的N497X、R661X、Q695X和/或Q926X突变，或在本文中提供的任何Cas9，诸如SEQ ID NO:4-26中提供的任何氨基酸序列中的相应的突变中的一个或多个，其中X是任何氨基酸。在一些实施方案中，本文提供的任何Cas9融合蛋白包含SEQ ID NO:6中提供的氨基酸序列的N497A、R661A、Q695A和/或Q926A突变，或在本文中提供的任何Cas9，诸如SEQ ID NO:4-26中提供的任何氨基酸序列中的相应的突变中的一个或多个。在一些实施方案中，Cas9域包含SEQ ID NO:6中提供的氨基酸序列的D10A突变，或在本文中提供的任何Cas9，诸如SEQ ID NO:4-26中提供的任何氨基酸序列中的相应的突变。在一些实施方案中，Cas9域(例如本文提供的任何融合蛋白的)包含如SEQ ID NO:20中所示的氨基酸序列。在一些实施方案中，Cas9域包含与SEQ ID NO:20至少60％，至少65％，至少70％，至少75％，至少80％，至少85％，至少90％，至少95％，至少96％，至少97％，至少98％，至少99％，或至少99.5％相同的氨基酸序列。具有高保真度的Cas9域是本领域已知的，并且对于本领域技术人员而言是显而易见的。例如，具有高保真度的Cas9域已经描述于Kleinstiver,B.P.,et al.“High-fidelity CRISPR-Cas9 nucleases with no detectable genome-wide off-target effects.”Nature 529,490-495(2016)；和Slaymaker,I.M.,et al.“Rationally engineered Cas9 nucleases with improved specificity.”Science 351,84-88(2015)；每篇的全部内容通过引用并入本文。

[0202] 应当理解，本文提供的任何碱基编辑器，例如，本文提供的任何C至G碱基编辑器，可以通过如本文所述修饰Cas9域而转换成高保真碱基编辑器以产生高保真碱基编辑器，例如，高保真C至G碱基编辑器。在一些实施方案中，高保真性Cas9域是dCas9域。在一些实施方案中，高保真性Cas9域是nCas9域。

[0203] 高保真性Cas9域，其中相对于SEQ ID NO:6的Cas9的突变以粗体和下划线显示

[0204] DKKYSIGLAIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETAE
ATRLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFFHRLEESFLVEEDKKHERHPIF
GNIVDEVAYHEKYPTIYHLRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIEGDLNPD
NSDVDKLFIQLVQTYNQLFEENPINASGVDAKAILSARLSKSRRLENLIAQLPGEKKNG
LFGNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAK
NLSDAILLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQ
SKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLRKQRTFDNGSIPHQ
IHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPYYVGPLARGNSRFAWMTRKSEETI
TPWNFEEVVDKGASAQSFIERMTAFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYV
TEGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFN
ASLGTYHDLLKIIKDKDFLDNEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKV
MKQLKRRRYTGWGALSRKLINGIRDKQSGKTILDFLKSDGFANRNFMALIHDDSLTFK
EDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKVVDELVKVMGRHKPENIVIEM
ARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHPVENTQLQNEKLYLYYLQNGRD
MYVDQELDINRLSDYDVDHIVPQSFLKDDSIDNKVLTRSDKNRGKSDNVPSEEVVKK
MKNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKAGFIKRQLVETRAITKHVAQIL
DSRMNTKYDENDKLIREVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYLNAV
VGTALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQEIGKATAKYFFYSNIMNFFKTEIT
LANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEVQTGGFSKESI
LPKRNSDKLIARKKDWDPKKYGGFDSPTVAYSVLVVAKVEKGKSKKLKSVKELLGITI
MERSSFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFELENGRKRMLASAGELQKGNEL
ALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQHKHYLDEIIEQISEFSKRVILADA
NLDKVLSAYNKHRDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVLD
ATLIHQSITGLYETRIDLSQLGGD (SEQ ID NO: 20)

[0205] 本公开内容还提供了napDNAbps的片段，诸如本文提供的任何napDNAbps的截短。在一些实施方案中，napDNAbp是N端截短，其中napDNAbp的N端不存在一个或多个氨基酸。在一些实施方案中，napDNAbp不存在napDNAbp的N端起的1、2、3、4、5、6、7、8、9、10、11、12、13、14、15、16、17、18、19、20、21、22、21、24、25、26、27、28、29、30、31、32、33、34、35、36、37、38、39、40、41、42、43、44、45、46、47、48、49或50个氨基酸。例如，napDNAbp的N端截短可以是本文提供的任何napDNAbp的N端截短，例如SEQ ID NO:4-40中任一项中提供的任一种napDNAbp。在一些实施方案中，napDNAbp是C端截短，其中napDNAbp的C端不存在一个或多个氨基酸。在一些实施方案中，napDNAbp不存在napDNAbp C端起的1、2、3、4、5、6、7、8、9、10、11、12、13、14、15、16、17、18、19、20、21，22、21、24、25、26、27、28、29、30、31、32、33、34、35、36、37、38、39、40、41、42、43、44、45、46、47、48、49或50个氨基酸。例如，napDNAbp的C端截短可以是本文提供的任何napDNAbp的C端截短，例如SEQ ID NO:4-40中任一项提供的任一种NAP。

[0206] 在一些实施方案中，与本文提供的任何napDNAbp，例如SEQ ID NO:4-40中提供的任一种napDNAbp相比，本文提供的任何napDNAbp具有1,2,3,4,5,6,7,8,9,10,11,12,13,14,15,16,17,18,19,20,21,22,21,24,25,26,27,28,29,30,31,32,33,34,35,36,37,38,39,40,41,42,43,44,45,46,47,48,49,50或更多个氨基酸变化。

[0207] 尿嘧啶结合蛋白(UBP)

[0208] 尿嘧啶结合蛋白或UBP是指能够结合尿嘧啶的蛋白。在一些实施方案中，尿嘧啶结合蛋白是尿嘧啶修饰酶。在一些实施方案中，尿嘧啶结合蛋白是尿嘧啶碱基切除酶。在一些实施方案中，尿嘧啶结合蛋白是尿嘧啶DNA糖基化酶(UDG)。在一些实施方案中，尿嘧啶结合蛋白以野生型UDG(例如人UDG)结合尿嘧啶的亲和力的至少1％，2％，3％，5％，10％，15％，20％，30％，40％，50％，60％，70％，80％，90％或至少95％的亲和力结合尿嘧啶。在一些实施方案中，与野生型尿嘧啶结合蛋白，例如结合尿嘧啶的野生型UDG(例如人UDG)相比，尿嘧

啶结合蛋白可具有1,2,3,4,5,6,7,8,9,10,11,12,13,14,15,16,17,18,19,20,21,22,21,24,25,26,27,28,29,30,31,32,33,34,35,36,37,38,39,40,41,42,43,44,45,46,47,48,49,50或更多个氨基酸变化。

[0209] 在一些实施方案中,UBP是尿嘧啶修饰酶。在一些实施方案中,UBP是尿嘧啶碱基切除酶。在一些实施方案中,UBP是尿嘧啶DNA糖基化酶。在一些实施方案中,UBP是本文提供的任何尿嘧啶结合蛋白。例如,UBP可以是UDG,UdgX,UdgX*,UdgX_On或SMUG1。在一些实施方案中,UBP包含与尿嘧啶结合蛋白,尿嘧啶碱基切除酶或尿嘧啶DNA糖基化酶(UDG)至少75%,80%,85%,90%,95%,96%,97%,98%,99%或99.5%相同的氨基酸序列。在一些实施方案中,UBP包含与本文提供的任何尿嘧啶结合蛋白,例如下文提供的任何UBP和UBP变体至少75%,80%,85%,90%,95%,96%,97%,98%,99%或99.5%相同的氨基酸序列。在一些实施方案中,UBP包含与SEQ ID NO:48-53的任一项至少75%,80%,85%,90%,95%,96%,97%,98%,99%或99.5%相同的氨基酸序列。在一些实施方案中,UBP包含SEQ ID NO:48-53中任一项的氨基酸序列。在一些实施方案中,与本文提供的任何UBP(例如SEQ ID NO:48-53中的任一项)相比,尿嘧啶结合蛋白具有1,2,3,4,5,6,7,8,9,10,11,12,13,14,15,16,17,18,19,20,21,22,21,24,25,26,27,28,29,30,31,32,33,34,35,36,37,38,39,40,41,42,43,44,45,46,47,48,49,50或更多个氨基酸变化。

[0210] 本公开内容还提供了UBP的片段,例如本文提供的任何UBP的截短。在一些实施方案中,UBP是N端截短,其中UBP的N端不存在一个或多个氨基酸。在一些实施方案中,UBP不存在从UBP N端起的1,2,3,4,5,6,7,8,9,10,11,12,13,14,15,16,17,18,19,20,21,22,21,24,25,26,27,28,29,30,31,32,33,34,35,36,37,38,39,40,41,42,43,44,45,46,47,48,49,或50个氨基酸。例如,UBP的N端截短可以是本文提供的任何UBP的N端截短,例如在SEQ ID NO:48-53中的任一项中提供的任一种UBP。在一些实施方案中,UBP是C端截短,其中UBP的C端不存在一个或多个氨基酸。在一些实施方案中,UBP不存在自UBP C端起的1,2,3,4,5,6,7,8,9,10,11,12,13,14,15,16,17,18,19,20,21,22,21,24,25,26,27,28,29,30,31,32,33,34,35,36,37,38,39,40,41,42,43,44,45,46,47,48,49或50个氨基酸。例如,UBP的C端截短可以是本文提供的任何UBP的C端截短,例如在SEQ ID NO:48-53中的任一项中提供的任一种UBP。

[0211] 应当理解,其他UBP对于熟练技术人员将是显而易见的,并且在本公开内容的范围内。例如,UBP先前已经记载于Sang et al.,“A Unique Uracil-DNA binding protein of the uracil DNA glycosylase superfamily,”Nucleic Acids Research,Vol.43,No.17 2015;其全部内容通过引用并入本文。

[0212] UDGMIGQKTLYSFFSPSPARKRHAPSPEPAVQGTGVAGVPEESGDAAAIPAKKAPAGQEEPGTPPSSPLSAEQLDRIQRNKAAALLRLAARNVPVGFGESWKKHLSGEFGKPYFIKLMGFVAEERKHYTVYPPPHQVFTWTQMCDIKDVKVVILGQDPYHGPNQAHGLCFSVQRPVPPPPSLENIYKELSTDIEDFVHPGHGDLSGWAKQGVLLLNAVLTVRAHQANSHKERGWEQFTDAVVSWLNQNSNGLVFLLWGSYAQKKGSAIDRKRHHVLQTAHPSPLSVYRGFFGCRHFSKTNELLQKSGKKPIDWKEL(SEQ ID NO:48)

[0213] UdgXMAGAQDFVPHTADLAELAAAAGECRGCGLYRDATQAVFGAGGRSARIMMIGEQPGDKEDLAGLPFVGPAGRLLDRALEAADIDRDALYVTNAVKHFKFTRAAGGKRRIHKTPSRTEVVACRPWLIAEMTSVEPDVVVLLGATAAKALLGNDFRVTQHRGEVLHVDDVPGDPALVATVHPSSLLRGPKEERESAFAGLVDDLRVAADVRP(SEQ ID

NO:49)

[0214] UdgX*(R107S)

[0215] MAGAQDFVPHTADLAELAAAAGECRGCGLYRDATQAVFGAGGRSARIMMIGEQPGDKEDLAGLPFVGP AGRLLDRALEAADIDRDALYVTNAVKHFKFTRAAGGKRSIHKTPSRTEVVACRPWLIAEMTSVEPDVVVLLGATAA KALLGNDFRVTQHRGEVLHVDDVPGDPALVATVHPSSLLRGPKEERESAFAGLVDDLRVAADVRP(SEQ ID NO: 50)

[0216] UdgX_On(H109S)

[0217] MAGAQDFVPHTADLAELAAAAGECRGCGLYRDATQAVFGAGGRSARIMMIGEQPGDKEDLAGLPFVGP AGRLLDRALEAADIDRDALYVTNAVKHFKFTRAAGGKRRISKTPSRTEVVACRPWLIAEMTSVEPDVVVLLGATAA KALLGNDFRVTQHRGEVLHVDDVPGDPALVATVHPSSLLRGPKEERESAFAGLVDDLRVAADVRP(SEQ ID NO: 51)

[0218] Rev7

[0219] MTTLTRQDLNFGQVVADVLCEFLEVAVHLILYVREVYPVGIFQKRKKYNVPVQMSCHPELNQYIQDTL HCVKPLLEKNDVEKVVVVILDKEHRPVEKFVFEITQPPLLSISSDSLLSHVEQLLRAFILKISVCDAVLDHNPPGC TFTVLVHTREAATRNMEKIQVIKDFPWILADEQDVHMHDPRLIPLKTMTSDILKMQLYVEERAHKGS(SEQ ID NO:52)

[0220] Smug1

[0221] MPQAFLLGSIHEPAGALMEPQPCPGSLAESFLEEELRLNAELSQLQFSEPVGIIYNPVEYAWEPHRNY VTRYCQGPKEVLFLGMNPGPFGMAQTGVPFGEVSMVRDWLGIVGPVLTPPQEHPKRPVLGLECPQSEVSGARFWGF FRNLCGQPEVFFHHCFVHNLCPLLFLAPSGRNLTPAELPAKQREQLLGICDAALCRQVQLLGVRLVVGVGRLAEQR ARRALAGLMPEVQVEGLLHPSPRNPQANKGWEAVAKERLNELGLLPLLLK(SEQ ID NO:53)

[0222] 核酸聚合酶(NAP)

[0223] 核酸聚合酶或NAP是指从核苷酸(例如脱氧核糖核苷酸和核糖核苷酸)合成核酸分子(例如DNA和RNA)的酶。在一些实施方案中,NAP是DNA聚合酶。在一些实施方案中,NAP是跨损伤聚合酶。跨损伤聚合酶在诱变中起作用,例如,通过重启复制叉或填补由于DNA损伤的存在而保留在基因组中的缺口。示例性的跨损伤聚合酶包括但不限于Pol Beta,Pol Lambda,Pol Eta,Pol Mu,Pol Iota,Pol Kappa,Pol Alpha,Pol Delta,Pol Gamma,和Pol Nu。

[0224] 在一些实施方案中,NAP是真核核酸聚合酶。在一些实施方案中,NAP是DNA聚合酶。在一些实施方案中,NAP具有跨损伤聚合酶活性。在一些实施方案中,NAP是跨损伤DNA聚合酶。在一些实施方案中,NAP是Rev7、Rev1复合物、聚合酶iota、聚合酶kappa或聚合酶eta。在一些实施方案中,NAP是真核聚合酶alpha,beta,gamma,delta,epsilon,gamma,eta,iota,kappa,lambda,mu,或nu。在一些实施方案中,NAP包含与天然存在的核酸聚合酶(例如,跨损伤DNA聚合酶)至少75%,80%,85%,90%,95%,96%,97%,98%,99%或99.5%相同的氨基酸序列。在一些实施方案中,NAP包含与本文(例如下文)提供的任何核酸聚合酶至少75%,80%,85%,90%,95%,96%,97%,98%,99%或99.5%相同的氨基酸序列。例如,NAP可以包含与SEQ ID NOs:54-64中的任一项至少75%,80%,85%,90%,95%,96%,97%,98%,99%或99.5%相同的氨基酸序列。在一些实施方案中,NAP包含SEQ ID NO:54-64中任一项的氨基酸序列。应当理解,其他NAP对于熟练技术人员而言将是显而易见的,并且在本

公开内容的范围内。在一些实施方案中,NAP包含SEQ ID NO:54-64中任一项的氨基酸序列。在一些实施方案中,与本文提供的任何NAP(例如SEQ ID NO:54-64中的任一项)相比,核酸聚合酶具有1,2,3,4,5,6,7,8,9,10,11,12,13,14,15,16,17,18,19,20,21,22,21,24,25,26,27,28,29,30,31,32,33,34,35,36,37,38,39,40,41,42,43,44,45,46,47,48,49,50或更多个氨基酸变化。

[0225] 本公开内容还提供了NAP的片段,例如本文提供的任何NAP的截短。在一些实施方案中,NAP是N端截短,其中NAP的N端不存在一个或多个氨基酸。在一些实施方案中,NAP不存在从NAP N端起的1,2,3,4,5,6,7,8,9,10,11,12,13,14,15,16,17,18,19,20,21,22,21,24,25,26,27,28,29,30,31,32,33,34,35,36,37,38,39,40,41,42,43,44,45,46,47,48,49或50个氨基酸。例如,NAP的N端截短可以是本文提供的任何NAP,例如在SEQ ID NO:54-64中的任一项中提供的任一种NAP的N端截短。在一些实施方案中,NAP是C端截短,其中NAP的C端不存在一个或多个氨基酸。在一些实施方案中,NAP不存在从NAP C末端起的1,2,3,4,5,6,7,8,9,10,11,12,13,14,15,16,17,18,19,20,21,22,21,24,25,26,27,28,29,30,31,32,33,34,35,36,37,38,39,40,41,42,43,44,45,46,47,48,49或50个氨基酸。例如,NAP的C端截短可以是本文提供的任何NAP,例如在SEQ ID NO:54-64中的任一项中提供的任一种NAP的C端截短。

[0226] Pol Beta

[0227] MSKRKAPQETLNGGITDMLTELANFEKNVSQAIHKYNAYRKAASVIAKYPHKIKSGAEAKKLPGVGTKIAEKIDEFLATGKLRKLEKIRQDDTSSSINFLTRVSGIGPSAARKFVDEGIKTLEDLRKNEDKLNHHQRIGLKYFGDFEKRIPREEMLQMQDIVLNEVKKVDSEYIATVCGSFRRGAESSGDMDVLLTHPSFTSESTKQPKLLHQVVEQLQKVHFITDTLSKGETKFMGVCQLPSKNDEKEYPHRRIDIRLIPKDQYYCGVLYFTGSDIFNKNMRAHALEKGFTINEYTIRPLGVTGVAGEPLPVDSEKDIFDYIQWKYREPKDRSE(SEQ ID NO:54)

[0228] Pol Lambda

[0229] MDPRGILKAFPKRQKIHADASSKVLAKIPRREEGEEAEEWLSSLRAHVVRTGIGRARAELFEKQIVQHGGQLCPAQGPGVTHIVVDEGMDYERALRLLRLPQLPPGAQLVKSAWLSLCLQERRLVDVAGFSIFIPSRYLDHPQPSKAEQDASIPPGTHEALLQTALSPPPPPTRPVSPPQKAKEAPNTQAQPISDDEASDGEETQVSAADLEALISGHYPTSLEGDCEPSPAPAVLDKWVCAQPSSQKATNHNLHITEKLEVLAKAYSVQGDKWRALGYAKAINALKSFHKPVTSYQEACSIPGIGKRMAEKIIEILESGHLRKLDHISESVPVLELFSNIWGAGTKTAQMWYQQGFRSLEDIRSQASLTTQQAIGLKHYSDFLERMPREEATEIEQTVQKAAQAFNSGLLCVACGSYRRGKATCGDVDVLITHPDGRSHRGIFSRLLDSLRQEGFLTDDLVSQEENGQQQKYLGVCRLPGPGRRHRRLDIIVVPYSEFACALLYFTGSAHFNRSMRALAKTKGMSLSEHALSTAVVRNTHGCKVGPGRVLPTPTEKDVFRLLGLPYREPAERDW(SEQ ID NO:55)

[0230] Pol Eta

[0231] MATGQDRVVALVDMDCFFVQVEQRQNPHLRNKPCAVVQYKSWKGGGIIAVSYEARAFGVTRSMWADDAKKLCPDLLLAQVRESRGKANLTKYREASVEVMEIMSRFAVIERASIDEAYVDLTSAVQERLQKLQGQPISADLLPSTYIEGLPQGPTTAEETVQKEGMRKQGLFQWLDSLQIDNLTSPDLQLTVGAVIVEEMRAAIERETGFQCSAGISHNKVLAKLACGLNKPNRQTLVSHGSVPQLFSQMPIRKIRSLGGKLGASVIEILGIEYMGELTQFTESQLQSHFGEKNGSWLYAMCRGIEHDPVKPRQLPKTIGCSKNFPGKTALATREQVQWWLLQLAQELEERLTKDRNDNDRVATQLVVSIRVQGDKRLSSLRRCCALTRYDAHKMSHDAFTVIKNCNTSGIQTEWSPPLTMLFLCATKFSASAPSSSTDITSFLSSDPSSLPKVPVTSSEAKTQGSGPAVTATKKATTSLESFFQKAAERQKVKEASLSSLTAPTQAPMSNSPSKPSLPFQTSQ

STGTEPFFKQKSLLLKQKQLNNSSVSSPQQNPWSNCKALPNSLPTEYPGCVPVCEGVSKLEESSKATPAEMDLAHN SQSMHASSASKSVLEVTQKATPNPSLLAAEDQVPCEKCGSLVPVWDMPEHMDYHFALELQKSFLQPHSSNPQVVSA VSHQGKRNPKSPLACTNKRPRPEGMQTLESFFKPLTH(SEQ ID NO:56)

[0232] Pol Mu

[0233] MLPKRRRARVGSPSGDAASSTPPSTRFPGVAIYLVEPRMGRSRRAFLTGLARSKGFRVLDACSSEATH VVMEETSAEEAVSWQERRMAAAPPGCTPPALLDISWLTESLGAGQPVPVECRHRLEVAGPRKGPLSPAWMPAYACQ RPTPLTHHNTGLSEALEILAEAAGFEGSEGRLLTFCRAASVLKALPSPVTTLSQLQGLPHFGEHSSRVVQELLEHG VCEEVERVRRSERYQTMKLFTQIFGVGVKTADRWYREGLRTLDDLREQPQKLTQQQKAGLQHHQDLSTPVLRSDVD ALQQVVEEAVGQALPGATVTLTGGFRRGKLQGHDVDFLITHPKEGQEAGLLPRVMCRLQDQGLILYHQHQHSCCES PTRLAQQSHMDAFERSFCIFRLPQPPGAAVGGSTRPCPSWKAVRVDLVVAPVSQFPFALLGWTGSKLFQRELRRFS RKEKGLWLNSHGLFDPEQKTFFQAASEEDIFRHLGLEYLPPEQRNA(SEQ ID NO:57)

[0234] Pol Iota

[0235] MEKLGVEPEEEGGGDDDEEDAEAWAMELADVGAAASSQGVHDQVLPTPNASSRVIVHVDLDCFYAQVE MISNPELKDKPLGVQQKYLVVTCNYEARKLGVKKLMNVRDAKEKCPQLVLVNGEDLTRYREMSYKVTELLEEFSPV VERLGFDENFVDLTEMVEKRLQQLQSDELSAVTVSGHVYNNQSINLLDVLHIRLLVGSQIAAEMREAMYNQLGLTG CAGVASNKLLAKLVSGVFKPNQQTVLLPESCQHLIHSLNHIKEIPGIGYKTAKCLEALGINSVRDLQTFSPKILEK ELGISVAQRIQKLSFGEDNSPVILSGPPQSFSEEDSFKKCSSEVEAKNKIEELLASLLNRVCQDGRKPHTVRLIIR RYSSEKHYGRESRQCPIPSHVIQKLGTGNYDVMTPMVDILMKLFRNMVNVKMPFHLTLLSVCFCNLKALNTAKKGL IDYYLMPSLSTTSRSGKHSFKMKDTHMEDFPKDKETNRDFLPSGRIESTRTRESPLDTTNFSKEKDINEFPLCSLP EGVDQEVFKQLPVDIQEEILSGKSREKFQGKGSVSCPLHASRGVLSFFSKKQMQDIPINPRDHLSSSKQVSSVSPC EPGTSGFNSSSSSYMSSQKDYSYYLDNRLKDERISQGPKEPQGFHFTNSNPAVSAFHSFPNLQSEQLFSRNHTTDS HKQTVATDSHEGLTENREPDSVDEKITFPSDIDPQVFYELPEAVQKELLAEWKRAGSDFHIGHK(SEQ ID NO: 58)

[0236] Pol Kappa

[0237] MDSTKEKCDSYKDDLLLRMGLNDNKAGMEGLDKEKINKIIMEATKGSRFYGNELKKEKQVNQRIENMM QQKAQITSQQLRKAQLQVDRFAMELEQSRNLSNTIVHIDMDAFYAAVEMRDNPELKDKPIAVGSMSMLSTSNYHAR RFGVRAAMPGFIAKRLCPQLIIVPPNFDKYRAVSKEVKEILADYDPNFMAMSLDEAYLNITKHLEERQNWPEDKRR YFIKMGSSVENDNPGKEVNKLSEHERSISPLLFEESPSDVQPPGDPFQVNFEEQNNPQILQNSVVFGTSAQEVVKE IRFRIEQKTTLTASAGIAPNTMLAKVCSDKNKPNGQYQILPNRQAVMDFIKDLPIRKVSGIGKVTEKMLKALGIIT CTELYQQRALLSLLFSETSWHYFLHISLGLGSTHLTRDGERKSMSVERTFSEINKAEEQYSLCQELCSELAQDLQK ERLKGRTVTIKLKNVNFEVKTRASTVSSVVSTAEEIFAIAKELLKTEIDADFPHPLRLRLMGVRISSFPNEEDRKH QQRSIIGFLQAGNQALSATECTLEKTDKDKFVKPLEMSHKKSFFDKKRSERKWSHQDTFKCEAVNKQSFQTSQPFQ VLKKKMNENLEISENSDDCQILTCPVCFRAQGCISLEALNKHVDECLDGPSISENFKMFSCSHVSATKVNKKENVP ASSLCEKQDYEAHPKIKEISSVDCIALVDTIDNSSKAESIDALSNKHSKEECSSLPSKSFNIEHCHQNSSSTVSLE NEDVGSFRQEYRQPYLCEVKTGQALVCPVCNVEQKTSDLTLFNVHVDVCLNKSFIQELRKDKFNPVNQPKESSRST GSSSGVQKAVTRTKRPGLMTKYSTSKKIKPNNPKHTLDIFFK(SEQ ID NO:59)

[0238] Pol Alpha

[0239] MAPVHGDDCEIGASALSDSGSFVSSRARREKKSKKGRQEALERLKKAKAGEKYKYEVEDFTGVYEEVD EEQYSKLVQARQDDDWIVDDDGIGYVEDGREIFDDDLEDDALDADEKGKDGKARNKDKRNVKKLAVTKPNNIKSMF

IACAGKKTADKAVDLSKDGLLGDILQDLNTETPQITPPPVMILKKKRSIGASPNPFSVHTATAVPSGKIASPVSRK
EPPLTPVPLKRAEFAGDDVQVESTEEEQESGAMEFEDGDFDEPMEVEEVDLEPMAAKAWDKESEPAEEVKQEADSG
KGTVSYLGSFLPDVSCWDIDQEGDSSFSVQEVQVDSSHLPLVKGADEEQVFHFYWLDAYEDQYNQPGVVFLFGKVW
IESAETHVSCCVMVKNIERTLYFLPREMKIDLNTGKETGTPISMKDVYEEFDEKIATKYKIMKFKSKPVEKNYAFE
IPDVPEKSEYLEVKYSAEMPQLPQDLKGETFSHVFGTNTSSLELFLMNRKIKGPCWLEVKSPQLLNQPVSWCKVEA
MALKPDLVNVIKDVSPPPLVVMAFSMKTMQNAKNHQNEIIAMAALVHHSFALDKAAPKPPFQSHFCVVSKPKDCIF
PYAFKEVIEKKNVKVEVAATERTLLGFFLAKVHKIDPDIIVGHNIYGFELEVLLQRINVCKAPHWSKIGRLKRSNM
PKLGGRSGFGERNATCGRMICDVEISAKELIRCKSYHLSELVQQILKTERVVIPMENIQNMYSESSQLLYLLEHTW
KDAKFILQIMCELNVLPLALQITNIAGNIMSRTLMGGRSERNEFLLLHAFYENNYIVPDKQIFRKPQQKLGDEDEE
IDGDTNKYKKGRKKAAYAGGLVLDPKVGFYDKFILLLDFNSLYPSIIQEFNICFTTVQRVASEAQKVTEDGEQEQI
PELPDPSLEMGILPREIRKLVERRKQVKQLMKQQDLNPDLILQYDIRQKALKLTANSMYGCLGFSYSRFYAKPLAA
LVTYKGREILMHTKEMVQKMNLEVIYGDTDSIMINTNSTNLEEVFKLGNKVKSEVNKLYKLLEIDIDGVFKSLLLL
KKKKYAALVVEPTSDGNYVTKQELKGLDIVRRDWCDLAKDTGNFVIGQILSDQSRDTIVENIQKRLIEIGENVLNG
SVPVSQFEINKALTKDPQDYPDKKSLPHVHVALWINSQGGRKVKAGDTVSYVICQDGSNLTASQRAYAPEQLQKQD
NLTIDTQYYLAQQIHPVVARICEPIDGIDAVLIATWLGLDPTQFRVHHYHKDEENDALLGGPAQLTDEEKYRDCER
FKCPCPTCGTENIYDNVFDGSGTDMEPSLYRCSNIDCKASPLTFTVQLSNKLIMDIRRFIKKYYDGWLICEEPTCR
NRTRHLPLQFSRTGPLCPACMKATLQPEYSDKSLYTQLCFYRYIFDAECALEKLTTDHEKDKLKKQFFTPKVLQDY
RKLKNTAEQFLSRSGYSEVNLSKLFAGCAVKS(SEQ ID NO:60)

[0240] Pol Delta

[0241] MDGKRRPGPGPGVPPKRARGGLWDDDDAPRPSQFEEDLALMEEMEAEHRLQEQEEEELQSVLEGVADG
QVPPSAIDPRWLRPTPPALDPQTEPLIFQQLEIDHYVGPAQPVPGGPPPSHGSVPVLRAFGVTDEGFSVCCHIHGF
APYFYTPAPPGFGPEHMGDLQRELNLAISRDSRGGRELTGPAVLAVELCSRESMFGYHGHGPSPFLRITVALPRLV
APARRLLEQGIRVAGLGTPSFAPYEANVDFEIRFMVDTDIVGCNWLELPAGKYALRLKEKATQCQLEADVLWSDVV
SHPPEGPWQRIAPLRVLSFDIECAGRKGIFPEPERDPVIQICSLGLRWGEPEPFLRLALTLRPCAPILGAKVQSYE
KEEDLLQAWSTFIRIMDPDVITGYNIQNFDLPYLISRAQTLKVQTFPFLGRVAGLCSNIRDSSFQSKQTGRRDTKV
VSMVGRVQMDMLQVLLREYKLRSYTLNAVSFHFLGEQKEDVQHSIITDLQNGNDQTRRRLAVYCLKDAYLPLRLLE
RLMVLVNAVEMARVTGVPLSYLLSRGQQVKVVSQLLRQAMHEGLLMPVVKSEGGEDYTGATVIEPLKGYYDVPIAT
LDFSSLYPSIMMAHNLCYTTLLRPGTAQKLGLTEDQFIRTPTGDEFVKTSVRKGLLPQILENLLSARKRAKAELAK
ETDPLRRQVLDGRQLALKVSANSVYGFTGAQVGKLPCLEISQSVTGFGRQMIEKTKQLVESKYTVENGYSTSAKVV
YGDTDSVMCRFGVSSVAEAMALGREAADWVSGHFPSPIRLEFEKVYFPYLLISKKRYAGLLFSSRPDAHDRMDCKG
LEAVRRDNCPLVANLVTASLRRLLIDRDPEGAVAHAQDVISDLLCNRIDISQLVITKELTRAASDYAGKQAHVELA
ERMRKRDPGSAPSLGDRVPYVIISAAKGVAAYMKSEDPLFVLEHSLPIDTQYYLEQQLAKPLLRIFEPILGEGRAE
AVLLRGDHTRCKTVLTGKVGGLLAFAKRRNCCIGCRTVLSHQGAVCEFCQPRESELYQKEVSHLNALEERFSRLWT
QCQRCQGSLHEDVICTSRDCPIFYMRKKVRKDLEDQEQLLRRFGPPGPEAW(SEQ ID NO:61)

[0242] Pol Gamma

[0243] MSRLLWRKVAGATVGPGPVPAPGRWVSSSVPASDPSDGQRRRQQQQQQQQQQQQPQQPQVLSSEGGQ
LRHNPLDIQMLSRGLHEQIFGQGGEMPGEAAVRRSVEHLQKHGLWGQPAVPLPDVELRLPPLYGDNLDQHFRLLAQ
KQSLPYLEAANLLLQAQLPPKPPAWAWAEGWTRYGPEGEAVPVAIPEERALVFDVEVCLAEGTCPTLAVAISPSAW
YSWCSQRLVEERYSWTSQLSPADLIPLEVPTGASSPTQRDWQEQLVVGHNVSFDRAHIREQYLIQGSRMRFLDTMS

MHMAISGLSSFQRSLWIAAKQGKHKVQPPTKQGQKSQRKARRGPAISSWDWLDISSVNSLAEVHRLYVGGPPLEKE
PRELFVKGTMKDIRENFQDLMQYCAQDVWATHEVFQQQLPLFLERCPHPVTLAGMLEMGVSYLPVNQNWERYLAEA
QGTYEELQREMKKSLMDLANDACQLLSGERYKEDPWLWDLEWDLQEFKQKKAKKVKKEPATASKLPIEGAGAPGDP
MDQEDLGPCSEEEEFQQDVMARACLQKLKGTTELLPKRPQHLPGHPGWYRKLCPRLDDPAWTPGPSLLSLQMRVTP
KLMALTWDGFPLHYSERHGWGYLVPGRRDNLAKLPTGTTLESAGVVCPYRAIESLYRKHCLEQGKQQLMPQEAGLA
EEFLLTDNSAIWQTVEELDYLEVEAEAKMENLRAAVPGQPLALTARGGPKDTQPSYHHGNGPYNDVDIPGCWFFKL
PHKDGNSCNVGSPFAKDFLPKMEDGTLQAGPGGASGPRALEINKMISFWRNAHKRISSQMVVWLPRSALPRAVIRH
PDYDEEGLYGAILPQVVTAGTITRRAVEPTWLTASNARPDRVGSELKAMVQAPPGYTLVGADVDSQELWIAAVLGD
AHFAGMHGCTAFGWMTLQGRKSRGTDLHSKTATTVGISREHAKIFNYGRIYGAGQPFAERLLMQFNHRLTQQEAAE
KAQQMYAATKGLRWYRLSDEGEWLVRELNLPVDRTEGGWISLQDLRKVQRETARKSQWKKWEVVAERAWKGGTESE
MFNKLESIATSDIPRTPVLGCCISRALEPSAVQEEFMTSRVNWVVQSSAVDYLHLMLVAMKWLFEEFAIDGRFCIS
IHDEVRYLVREEDRYRAALALQITNLLTRCMFAYKLGLNDLPQSVAFFSAVDIDRCLRKEVTMDCKTPSNPTGMER
RYGIPQGEALDIYQIIELTKGSLEKRSQPGP(SEQ ID NO:62)

[0244] Pol Nu

[0245] MENYEALVGFDLCNTPLSSVAQKIMSAMHSGDLVDSKTWGKSTETMEVINKSSVKYSVQLEDRKTQSP
EKKDLKSLRSQTSRGSAKLSPQSFSVRLTDQLSADQKQKSISSLTLSSCLIPQYNQEASVLQKKGHKRKHFLMENI
NNENKGSINLKRKHITYNNLSEKTSKQMALEEDTDDAEGYLNSGNSGALKKHFCDIRHLDDWAKSQLIEMLKQAAA
LVITVMYTDGSTQLGADQTPVSSVRGIVVLVKRQAEGGHGCPDAPACGPVLEGFVSDDPCIYIQIEHSAIWDQEQE
AHQQFARNVLFQTMKCKCPVICFNAKDFVRIVLQFFGNDGSWKHVADFIGLDPRIAAWLIDPSDATPSFEDLVEKY
CEKSITVKVNSTYGNSSRNIVNQNVRENLKTLYRLTMDLCSKLKDYGLWQLFRTLELPLIPILAVMESHAIQVNKE
EMEKTSALLGARLKELEQEAHFVAGERFLITSNNQLREILFGKLKLHLLSQRNSLPRTGLQKYPSTSEAVLNALRD
LHPLPKIILEYRQVHKIKSTFVDGLLACMKKGSISSTWNQTGTVTGRLSAKHPNIQGISKHPIQITTPKNFKGKED
KILTISPRAMFVSSKGHTFLAADFSQIELRILTHLSGDPELLKLFQESERDDVFSTLTSQWKDVPVEQVTHADREQ
TKKVVYAVVYGAGKERLAACLGVPIQEAAQFLESFLQKYKKIKDFARAAIAQCHQTGCVVSIMGRRRPLPRIHAHD
QQLRAQAERQAVNFVVQGSAADLCKLAMIHVFTAVAASHTLTARLVAQIHDELLFEVEDPQIPECAALVRRTMESL
EQVQALELQLQVPLKVSLSAGRSWGHLVPLQEAWGPPPGPCRTESPSNSLAAPGSPASTQPPPLHFSPSFCL(SEQ
ID NO:63)

[0246] Rev1

[0247] MRRGGWRKRAENDGWETWGGYMAAKVQKLEEQFRSDAAMQKDGTSSTIFSGVAIYVNGYTDPSAEELR
KLMMLHGGQYHVYYSRSKTTHIIATNLPNAKIKELKGEKVIRPEWIVESIKAGRLLSYIPYQLYTKQSSVQKGLSF
NPVCRPEDPLPGPSNIAKQLNNRVNHIVKKIETENEVKVNGMNSWNEEDENNDFSFVDLEQTSPGRKQNGIPHPRG
STAIFNGHTPSSNGALKTQDCLVPMVNSVASRLSPAFSQEEDKAEKSSTDFRDCTLQQLQQSTRNTDALRNPHRTN
SFSLSPLHSNTKINGAHHSTVQGPSSTKSTSSVSTFSKAAPSVPSKPSDCNFISNFYSHSRLHHISMWKCELTEFV
NTLQRQSNGIFPGREKLKKMKTGRSALVVTDTGDMSVLNSPRHQSCIMHVDMDCFFVSVGIRNRPDLKGKPVAVTS
NRGTGRAPLRPGANPQLEWQYYQNKILKGKAADIPDSSLWENPDSAQANGIDSVLSRAEIASCSYEARQLGIKNGM
FFGHAKQLCPNLQAVPYDFHAYKEVAQTLYETLASYTHNIEAVSCDEALVDITEILAETKLTPDEFANAVRMEIKD
QTKCAASVGIGSNILLARMATRKAKPDGQYHLKPEEVDDFIRGQLVTNLPGVGHSMESKLASLGIKTCGDLQYMTM
AKLQKEFGPKTGQMLYRFCRGLDDRPVRTEKERKSVSAEINYGIRFTQPKEAEAFLLSLSEEIQRRLEATGMKGKR
LTLKIMVRKPGAPVETAKFGGHGICDNIARTVTLDQATDNAKIIGKAMLNMFHTMKLNISDMRGVGIHVNQLVPTN

LNPSTCPSRPSVQSSHFPSGSYSVRDVFQVQKAKKSTEEEHKEVFRAAVDLEISSASRTCTFLPPFPAHLPTSPDT NKAESSGKWNGLHTPVSVQSRLNLSIEVPSPSQLDQSVLEALPPDLREQVEQVCAVQQAESHGDKKKEPVNGCNTG ILPQPVGTVLLQIPEPQESNSDAGINLIALPAFSQVDPEVFAALPAELQRELKAAYDQRQRQGENSTHQQSASASV PKNPLLHLKAAVKEKKRNKKKKTIGSPKRIQSPLNNKLLNSPAKTLPGACGSPQKLIDGFLKHEGPPAEKPLEELS ASTSGVPGLSSLQSDPAGCVRPPAPNLAGAVEFNDVKTLLREWITTISDPMEEDILQVVKYCTDLIEEKDLEKLDL VIKYMKRLMQQSVESVWNMAFDFILDNVQVVLQQTYGSTLKVT(SEQ ID NO:64)

[0248] 碱基切除酶(BEE)

[0249] 碱基切除酶或BEE是指能够从核酸分子(例如DNA或RNA)除去碱基(例如A,T,C,G或U)的蛋白质。在一些实施方案中,BEE能够从DNA除去胞嘧啶。在一些实施方案中,BEE能够从DNA除去胸腺嘧啶。示例性的BEE包括但不限于UDG Tyr147Ala和UDG Asn204Asp,如记载于Sang et al.,“A Unique Uracil-DNA binding protein of the uracil DNA glycosylase superfamily,”Nucleic Acids Research,Vol.43,No.17 2015;其全部内容通过引用并入本文。

[0250] 在一些实施方案中,碱基切除酶(BEE)是胞嘧啶,胸腺嘧啶,腺嘌呤,鸟嘌呤或尿嘧啶碱基切除酶。在一些实施方案中,碱基切除酶(BEE)是胞嘧啶碱基切除酶。在一些实施方案中,BEE是胸腺嘧啶碱基切除酶。在一些实施方案中,碱基切除酶包含与天然存在的BEE至少75%,80%,85%,90%,95%,96%,97%,98%,99%或99.5%相同的氨基酸序列。在一些实施方案中,碱基切除酶包含与本文提供的任何BEE,例如下面的UDG(Tyr147Ala)或UDG(Asn204Asp)至少75%,80%,85%,90%,95%,96%,97%,98%,99%或99.5%相同的氨基酸序列。在一些实施方案中,碱基切除酶包含与SEQ ID NOs:65-66的任一项至少75%,80%,85%,90%,95%,96%,97%,98%,99%或99.5%相同的氨基酸序列。在一些实施方案中,碱基切除酶包含SEQ ID NO:65-66中任一项的氨基酸序列。在一些实施方案中,与本文提供的任何BEE,例如SEQ ID NO:65-66中的任一项相比,碱基切除酶具有1,2,3,4,5,6,7,8,9,10,11,12,13,14,15,16,17,18,19,20,21,22,21,24,25,26,27,28,29,30,31,32,33,34,35,36,37,38,39,40,41,42,43,44,45,46,47,48,49,50或更多个氨基酸变化。

[0251] 本公开内容还提供了BEE的片段,例如本文提供的任何BEE的截短。在一些实施方案中,BEE是N端截短,其中BEE的N端不存在一个或多个氨基酸。在一些实施方案中,BEE不存在从BEE N端起的1,2,3,4,5,6,7,8,9,10,11,12,13,14,15,16,17,18,19,20,21,22,21,24,25,26,27,28,29,30,31,32,33,34,35,36,37,38,39,40,41,42,43,44,45,46,47,48,49或50个氨基酸。例如,BEE的N端截短可以是本文提供的任何BEE,例如在SEQ ID NO:65-66中的任一项中提供的任一种BEE的N端截短。在一些实施方案中,BEE是C端截短,其中在BEE的C端不存在一个或多个氨基酸。在一些实施方案中,BEE不存在从BEE C端起的1,2,3,4,5,6,7,8,9,10,11,12,13,14,15,16,17,18,19,20,21,22,21,24,25,26,27,28,29,30,31,32,33,34,35,36,37,38,39,40,41,42,43,44,45,46,47,48,49或50个氨基酸。例如,BEE的C端截短可以是本文提供的任何BEE,例如在SEQ ID NO:65-66中的任一项中提供的任一种BEE的C端截短。

[0252] 应当理解,其他BEE对熟练技术人员而言是显而易见的,并且在本公开内容的范围内。例如,BEE先前已经记载于Sang et al.,“A Unique Uracil-DNA binding protein of the uracil DNA glycosylase superfamily,”Nucleic Acids Research,Vol.43,No.17

2015;其全部内容通过引用并入本文。

[0253] UDG(Tyr147Ala)-突变残基以粗体且加下划线指示。

[0254] MIGQKTLYSFFSPSPARKRHAPSPEPAVQGTGVAGVPEESGDAAAIPAKKA PAGQEEPGTPPSSPLSAEQLDRIQRNKAAALLRLAARNVPVGFGESWKKH LSGEFGKPYFIKLMGFVAEERKHYTVYPPPHQVFTWTQMCDIKDVKVVIL GQDP**A**HGPNQAHGLCFSVQRPVPPPPSLENIYKELSTDIEDFVHPGHGDLS GWAKQGVLLLNAVLTVRAHQANSHKERGWEQFTDAVVSWLNQNSNGL VFLLWGSYAQKKGSAIDRKRHHVLQTAHPSPLSVYRGFFGCRHFSKTNEL

[0255] LQKSGKKPIDWKEL (SEQ ID NO: 65)

[0256] UDG(Asn204Asp)-突变残基以粗体且加下划线指示。

[0257] MIGQKTLYSFFSPSPARKRHAPSPEPAVQGTGVAGVPEESGDAAAIPAKKA PAGQEEPGTPPSSPLSAEQLDRIQRNKAAALLRLAARNVPVGFGESWKKH LSGEFGKPYFIKLMGFVAEERKHYTVYPPPHQVFTWTQMCDIKDVKVVIL GQDPYHGPNQAHGLCFSVQRPVPPPPSLENIYKELSTDIEDFVHPGHGDLS GWAKQGVLLL**D**AVLTVRAHQANSHKERGWEQFTDAVVSWLNQNSNGL VFLLWGSYAQKKGSAIDRKRHHVLQTAHPSPLSVYRGFFGCRHFSKTNEL LQKSGKKPIDWKEL (SEQ ID NO: 66)

[0258] 脱氨酶域

[0259] 在一些实施方案中,本文提供的任何融合蛋白或碱基编辑器包含胞苷脱氨酶域。在一些实施方案中,胞苷脱氨酶域可以催化C至U碱基变化。在一些实施方案中,胞苷脱氨酶域是载脂蛋白B mRNA编辑复合体(APOBEC)家族脱氨酶。在一些实施方案中,胞苷脱氨酶域是APOBEC1脱氨酶。在一些实施方案中,胞苷脱氨酶域是APOBEC2脱氨酶。在一些实施方案中,胞苷脱氨酶域是APOBEC3脱氨酶。在一些实施方案中,胞苷脱氨酶域是APOBEC3A脱氨酶。在一些实施方案中,胞苷脱氨酶域是APOBEC3B脱氨酶。在一些实施方案中,胞苷脱氨酶域是APOBEC3C脱氨酶。在一些实施方案中,胞苷脱氨酶域是APOBEC3D脱氨酶。在一些实施方案中,胞苷脱氨酶域是APOBEC3E脱氨酶。在一些实施方案中,胞苷脱氨酶域是APOBEC3F脱氨酶。在一些实施方案中,胞苷脱氨酶域是APOBEC3G脱氨酶。在一些实施方案中,胞苷脱氨酶域是APOBEC3H脱氨酶。在一些实施方案中,胞苷脱氨酶域是APOBEC4脱氨酶。在一些实施方案中,胞苷脱氨酶域是激活诱导的脱氨酶(AID)。在一些实施方案中,胞苷脱氨酶域是脊椎动物脱氨酶。在一些实施方案中,胞苷脱氨酶域是无脊椎动物脱氨酶。在一些实施方案中,胞苷脱氨酶域是人,黑猩猩,大猩猩,猴,牛,犬,大鼠或小鼠脱氨酶。在一些实施方案中,胞苷脱氨酶域是人脱氨酶。在一些实施方案中,胞苷脱氨酶域是大鼠脱氨酶,例如rAPOBEC1。在一些实施方案中,胞苷脱氨酶域是海七鳃鳗(Petromyzon marinus)胞苷脱氨酶1(pmCDA1)。在一些实施方案中,胞苷脱氨酶域是人APOBEC3G(SEQ ID NO:77)。在一些实施方案中,胞苷脱氨酶域是人APOBEC3G的片段(SEQ ID NO:100)。在一些实施方案中,胞苷脱氨酶域是包含D316R_D317R突变(SEQ ID NO:99)的人APOBEC3G变体。在一些实施方案中,胞苷脱氨酶域是人APOBEC3G的片段,并且包含对应于SEQID NO:77中的D316R_D317R突变的突变(SEQ ID NO:101)。

[0260] 在一些实施方案中,胞苷脱氨酶域与天然存在的胞苷脱氨酶是至少80%,至少85%,至少90%,至少92%,至少95%,至少96%,至少97%,至少98%相同,至少99%或至少

99.5％相同的。在一些实施方案中，胞苷脱氨酶域与本文提供的任何胞苷脱氨酶是至少80％，至少85％，至少90％，至少92％，至少95％，至少96％，至少97％，至少98％，至少99％或至少99.5％相同的。在一些实施方案中，胞苷脱氨酶域与SEQ ID NO:67-101中任一项的脱氨酶域是至少80％，至少85％，至少90％，至少92％，至少95％，至少96％，至少97％，至少98％，至少99％或至少99.5％相同的。在一些实施方案中，核酸编辑域包含SEQ ID NO:67-101中任一项的氨基酸序列。在一些实施方案中，与本文提供的任何胞苷脱氨酶域，例如SEQ ID NO:67-101中的任一项相比，胞苷脱氨酶域具1,2,3,4,5,6,7,8,9,10,11,12,13,14,15,16,17,18,19,20,21,22,21,24,25,26,27,28,29,30,31,32,33,34,35,36,37,38,39,40,41,42,43,44,45,46,47,48,49,50或更多个氨基酸变化。

[0261] 本公开内容还提供了胞苷脱氨酶域的片段，例如本文提供的任何胞苷脱氨酶域的截短。在一些实施方案中，胞苷脱氨酶域是N端截短，其中胞苷脱氨酶域的N端不存在一个或多个氨基酸。在一些实施方案中，胞苷脱氨酶域不存在从胞苷脱氨酶域N端起的1,2,3,4,5,6,7,8,9,10,11,12,13,14,15,16,17,18,19,20,21,22,21,24,25,26,27,28,29,30,31,32,33,34,35,36,37,38,39,40,41,42,43,44,45,46,47,48,49或50个氨基酸。例如，胞苷脱氨酶域的N端截短可以是本文提供的任何胞苷脱氨酶域，例如SEQ ID NO:67-101中任一项提供的任一种胞苷脱氨酶域的N端截短。在一些实施方案中，胞苷脱氨酶域是C端截短，其中胞苷脱氨酶域的C端不存在一个或多个氨基酸。在一些实施方案中，胞苷脱氨酶域不存在从胞苷脱氨酶域C端起的1,2,3,4,5,6,7,8,9,10,11,12,13,14,15,16,17,18,19,20,21,22,21,24,25,26,27,28,29,30,31,32,33,34,35,36,37,38,39,40,41,42,43,44,45,46,47,48,49或50个氨基酸。例如，胞苷脱氨酶域的C端截短可以是本文提供的任何胞苷脱氨酶域，例如SEQ ID NO:67-101中任一项中提供的任一种胞苷脱氨酶域的C端截短。

[0262] 一些示例性的胞苷脱氨酶域包括但不限于下文提供的那些。应当理解，在一些实施方案中，可以使用相应序列的活性域，例如没有定位信号(核定位序列，没有核输出信号，胞质定位信号)的域。

[0263] 人AID:

[0264] MDSLLMNRRKFLYQFKNVRWAKGRRETYLCYVVKRRDSATSFSLDFGYLRNKNGCHVELLFLRYISDWDLDPGRCYRVTWFTSWSPCYDCARHVADFLRGNPNLSLRIFTARLYFCEDRKAEPEGLRRLHRAGVQIAIMTFKDYFYCWNTFVENHERTFKAWEGLHENSVRLSRQLRRILLPLYEVDDLRDAFRTLGL (SEQ ID NO: 67)

[0265] (下划线:核定位信号;双下划线:核输出信号)

[0266] 小鼠AID:

[0267] MDSLLMKQKKFLYHFKNVRWAKGRHETYLCYVVKRRDSATSCSLDFGHLRNKSGCHVELLFLRYISDWDLDPGRCYRVTWFTSWSPCYDCARHVAEFLRWNPNLSLRIFTARLYFCEDRKAEPEGLRRLHRAGVQIGIMTFKDYFYCWNTFVENRERTFKAWEGLHENSVRLTRQLRRILLPLYEVDDLRDAFRMLGF (SEQ ID NO: 68)

[0268] (下划线:核定位信号;双下划线:核输出信号)

[0269] 犬AID:

[0270] MDSLLMKQRKFLYHFKNVRWAKGRHETYLCYVVKRRDSATSFSLDFGHLRNKSGCHVELLFLRYISDWDLDPGRCYRVTWFTSWSPCYDCARHVADFLRGYPNLSLRIFAARLYFCEDRKAEPEGLRRLHRAGVQIAIMTFKDYFYCWNTFVENREKTFKAWEGLHENSVRLSRQLRRILLPLYEVDDLRDAFRTLGL (SEQ ID NO: 69)

[0271] (下划线:核定位信号;双下划线:核输出信号)

[0272] 牛AID:

[0273] MDSLLKKQRQFLYQFKNVRWAKGRHETYLCYVVKRRDSPTSFSLDFGHLRNKAGCH VELLFLRYISDWDLDPGRCYRVTWFTSWSPCYDCARHVADFLRGYPNLSLRIFTARLY FCDKERKAEPEGLRRLHRAGVQIAIMTFKDYFYCWNTFVENHERTFKAWEGLHENSV RLSRQLRRILLPLYEVDDLRDAFRTLGL (SEQ ID NO: 70)

[0274] (下划线:核定位信号;双下划线:核输出信号)

[0275] 大鼠AID

[0276] MAVGSKPKAALVGPHWERERIWCFLCSTGLGTQQTGQTSRWLRPAATQDPVSPPRSL LMKQRKFLYHFKNVRWAKGRHETYLCYVVKRRDSATSFSLDFGYLRNKSGCHVELL FLRYISDWDLDPGRCYRVTWFTSWSPCYDCARHVADFLRGNPNLSLRIFTARLTGWG ALPAGLMSPARPSDYFYCWNTFVENHERTFKAWEGLHENSVRLSRRLRRILLPLYEVD DLRDAFRTLGL (SEQ ID NO: 71)

[0277] (下划线:核定位信号;双下划线:核输出信号)

[0278] 小鼠APOBEC-3:

[0279] MGPFCLGCSHRKCYSPIRNLISQETFKFHFKNLGYAKGRKDTFLCYEVTRKDCDSPVSL HHGVFKNKDNI*HAEICFLYWFHDKVLKVLSPREEFKITWYMSWSPCFEC*AEQIVRFLATH HNLSLDIFSSRLYNVQDPETQQNLCRLVQEGAQVAAMDLYEFKKCWKKFVDNGGRR FRPWKRLLTNFRYQDSKLQEILRPCYIPVPSSSSSTLSNICLTKGLPETRFCVEGRRMDP LSEEEFYSQFYNQRVKHLCYYHRMKPYLCYQLEQFNGQAPLKGCLLSEKGKQ*HAEILF LDKIRSMELSQVTITCYLTWSPCPNC*AWQLAAFKRDRPDLILHIYTSRLYFHWKRPFQKG LCSLWQSGILVDVMDLPQFTDCWTNFVNPKRPFWPWKGLEIISRRTQRRLRRIKESWG LQDLVNDFGNLQLGPPMS (SEQ ID NO: 72)

[0280] (斜体:核酸编辑域)

[0281] 大鼠APOBEC-3:

[0282] MGPFCLGCSHRKCYSPIRNLISQETFKFHFKNLRYAIDRKDTFLCYEVTRKDCDSPVSL HHGVFKNKDNI*HAEICFLYWFHDKVLKVLSPREEFKITWYMSWSPCFEC*AEQVLRFLATH HNLSLDIFSSRLYNIRDPENQQNLCRLVQEGAQVAAMDLYEFKKCWKKFVDNGGRRF RPWKKLLTNFRYQDSKLQEILRPCYIPVPSSSSSTLSNICLTKGLPETRFCVERRRVHLL SEEEFYSQFYNQRVKHLCYYHGVKPYLCYQLEQFNGQAPLKGCLLSEKGKQ*HAEILFL DKIRSMELSQVIITCYLTWSPCPNC*AWQLAAFKRDRPDLILHIYTSRLYFHWKRPFQKGL

[0283] CSLWQSGILVDVMDLPQFTDCWTNFVNPKRPFWPWKGLEIISRRTQRRLHRIKESWGL QDLVNDFGNLQLGPPMS (SEQ ID NO: 73)

[0284] (斜体:核酸编辑域)

[0285] 猕猴APOBEC-3G:

[0286] MVEPMDPRTFVSNFNNRPILSGLNTVWLCCEVKTKDPSGPPLDAKIFQGKVYSKAKY *HPEMRFLRWFHKWRQLHHDQEYKVTWYVSWSPCTRC*ANSVATFLAKDPKVTLTIFVAR LYYFWKPDYQQALRILCQKRGGPHATMKIMNYNEFQDCWNKFVDGRGKPFKPRNNL PKHYTLLQATLGELLRHLMDPGTFTSNFNNKPWVSGQHETYLCYKVERLHNDTWVP LNQHRGFLRNQAPNIHGFPKGR*HAELCFLDLIPFWKLDGQQYRVTCFTSWSPCFSC*AQE MAKFISNNEHVSLCIFAARIYDDQGRYQEGLRALHRDGAKIAMMNYSEFEYCWDTFV DRQGRPFQPWDGLDEHSQALSGRLRAI (SEQ ID NO: 74)

[0287] (斜体:核酸编辑域;下划线:胞质定位信号)

[0288] 黑猩猩APOBEC-3G:

[0289] MKPHFRNPVERMYQDTFSDNFYNRPILSHRNTVWLCYEVKTKGPSRPPLDAKIFRGQV YSKLKY*HPEMRFFHWFSKWRKLHRDQEYEVTWYISWSPCTKC*TRDVATFLAEDPKVTLT IFVARLYYFWDPDYQEALRSLCQKRDGPRATMKIMNYDEFQHCWSKFVYSQRELFEP WNNLPKYYILLHIMLGEILRHSMDPPTFTSNFNNELWVRGRHETYLCYEVERLHNDT WVLLNQRRGFLCNQAPHKHGFLEGR*HAELCFLDVIPFWKLDLHQDYRVTCFTSWSPCFS C*AQEMAKFISNNKHVSLCIFAARIYDDQGRCQEGLRTLAKAGAKISIMTYSEFKHCWD TFVDHQGCPFQPWDGLEEHSQALSGRLRAILQNQGN (SEQ ID NO: 75)

[0290] (斜体:核酸编辑域;下划线:胞质定位信号)

[0291] 绿猴APOBEC-3G:

[0292] MNPQIRNMVEQMEPDIFVYYFNNRPILSGRNTVWLCYEVKTKDPSGPPLDANIFQGKL YPEAKD*HPEMKFLHWFRKWRQLHRDQEYEVTWYVSWSPCTRC*ANSVATFLAEDPKVTL TIFVARLYYFWKPDYQQALRILCQERGGPHATMKIMNYNEFQHCWNEFVDGQGKPF KPRKNLPKHYTLLHATLGELLRHVMDPGTFTSNFNNKPWVSGQRETYLCYKVERSHN DTWVLLNQHRGFLRNQAPDRHGFPKGR*HAELCFLDLIPFWKLDDQQYRVTCFTSWSPC FSC*AQKMAKFISNNKHVSLCIFAARIYDDQGRCQEGLRTLHRDGAKIAVMNYSEFEYC WDTFVDRQGRPFQPWDGLDEHSQALSGRLRAI (SEQ ID NO: 76)

[0293] (斜体:核酸编辑域;下划线:胞质定位信号)

[0294] 人APOBEC-3G:

[0295] MKPHFRNTVERMYRDTFSYNFYNRPILSRRNTVWLCYEVKTKGPSRPPLDAKIFRGQV YSELKY*HPEMRFFHWFSKWRKLHRDQEYEVTWYISWSPCTKC*TRDMATFLAEDPKVTL TIFVARLYYFWDPDYQEALRSLCQKRDGPRATMKIMNYDEFQHCWSKFVYSQRELFE PWNNLPKYYILLHIMLGEILRHSMDPPTFTFNFNNEPWVRGRHETYLCYEVERMHNDT WVLLNQRRGFLCNQAPHKHGFLEGR*HAELCFLDVIPFWKLDLDQDYRVTCFTSWSPCFS C*AQEMAKFISKNKHVSLCIFTARIYDDQGRCQEGLRTLAEAGAKISIMTYSEFKHCWD TFVDHQGCPFQPWDGLDEHSQDLSGRLRAILQNQEN (SEQ ID NO: 77)

[0296] (斜体:核酸编辑域;下划线:胞质定位信号)

[0297] 人APOBEC-3F:

[0298] MKPHFRNTVERMYRDTFSYNFYNRPILSRRNTVWLCYEVKTKGPSRPRLDAKIFRGQ VYSQPEH*HAEMCFLSWFCGNQLPAYKCFQITWFVSWTPCPDC*VAKLAEFLAEHPNVTLT ISAARLYYYWERDYRRALCRLSQAGARVKIMDDEEFAYCWENFVYSEGQPFMPWYK FDDNYAFLHRTLKEILRNPMEAMYPHIFYFHFKNLRKAYGRNESWLCFTMEVVKHHS PVSWKRGVFRNQVDPETHC*HAERCFLSWFCDDILSPNTNYEVTWYTSWSPCPEC*AGEVA EFLARHSNVNLTIFTARLYYFWDTDYQEGLRSLSQEGASVEIMGYKDFKYCWENFVY NDDEPFKPWKGLKYNFLFLDSKLQEILE (SEQ ID NO: 78)

[0299] (斜体:核酸编辑域)

[0300] 人APOBEC-3B:

[0301] MNPQIRNPMERMYRDTFYDNFENEPILYGRSYTWLCYEVKIKRGRSNLLWDTGVFRG QVYFKPQY*HAEMCFLSWFCGNQLPAYKCFQITWFVSWTPCPDC*VAKLAEFLSEHPNVT LTISAARLYYYWERDYRRALCRLSQAGARVTIMDYEEFAYCWENFVYNEGQQFMPW YKFDENYAFLHRTLKEILRYLMDPDTFTFNFNNDPLVLRRRQTYLCYEVERLDNGTW VLMDQHMGFLCNEAKNLLCGFY*GRHAELRFLDLVPSLQLDPAQIYRVTWFISWSPCFSW GC*AGEVRAFLQENTHVRLRIFAARIYDYDPLYKEALQMLRDAGAQVSIMTYDEFEYC WDTFVYRQGCPFQPWDGLEEHSQALSGRLRAILQNQGN (SEQ ID NO: 79)

[0302] (斜体:核酸编辑域)

[0303] 大鼠APOBEC3:

[0304] MQPQGLGPNAGMGPVCLGCSHRRPYSPIRNPLKKLYQQTFYFHFKNVRYAWGRKNNFLCYEVNGMDCA LPVPLRQGVFRKQGHIHAELCFIYWFHDKVLRVLSPMEEFKVTWYMSWSPCSKCAEQVARFLAAHRNLSLAIFSSR LYYYLRNPNYQQKLCRLIQEGVHVAAMDLPEFKKCWNKFVDNDGQPFRPWMRLRINFSFYDCKLQEIFSRMNLLRE

DVFYLQFNNSHRVKPVQNRYYRRKSYLCYQLERANGQEPLKGYLLYKKGEQHVEILFLEKMRSMELSQVRITCYLT WSPCPNCARQLAAFKKDHPDLILRIYTSRLYFYWRKKFQKGLCTLWRSGIHVDVMDLPQFADCWTNFVNPQRPFRP WNELEKNSWRIQRRLRRIKESWGL(SEQ ID NO:80)

[0305] 牛APOBEC-3B:

[0306] DGWEVAFRSGTVLKAGVLGVSMTEGWAGSGHPGQGACVWTPGTRNTMNLLREVLFKQQFGNQPRVPAP YYRRKTYLCYQLKQRNDLTLDRGCFRNKKQRHAEIRFIDKINSLDLNPSQSYKIICYITWSPCPNCANELVNFITR NNHLKLEIFASRLYFHWIKSFKMGLQDLQNAGISVAVMTHTEFEDCWEQFVDNQSRPFQPWDKLEQYSASIRRRLQ RILTAPI(SEQ ID NO:81)

[0307] 黑猩猩APOBEC-3B:

[0308] MNPQIRNPMEWMYQRTFYYNFENEPILYGRSYTWLCYEVKIRRGHSNLLWDTGVFRGQMYSQPEHHAE MCFLSWFCGNQLSAYKCFQITWFVSWTPCPDCVAKLAKFLAEHPNVTLTISAARLYYYWERDYRRALCRLSQAGAR VKIMDDEEFAYCWENFVYNEGQPFMPWYKFDDNYAFLHRTLKEIIRHLMDPDTFTFNFNNDPLVLRRHQTYLCYEV ERLDNGTWVLMDQHMGFLCNEAKNLLCGFYGRHAELRFLDLVPSLQLDPAQIYRVTWFISWSPCFSWGCAGQVRAF LQENTHVRLRIFAARIYDYDPLYKEALQMLRDAGAQVSIMTYDEFEYCWDTFVYRQGCPFQPWDGLEEHSQALSGR LRAILQVRASSLCMVPHRPPPPPQSPGPCLPLCSEPPLGSLLPTGRPAPSLPFLLTASFSFPPPASLPPLPSLSLS PGHLPVPSFHSLTSCSIQPPCSSRIRETEGWASVSKEGRDLG(SEQ ID NO:82)

[0309] 人APOBEC-3C:

[0310] MNPQIRNPMKAMYPGTFYFQFKNLWEANDRNETWLCFTVEGIKRRSVVSWKTGVFR NQVDSETH*CHAERCFLSWFCDDILSPNTKYQVTWYTSWSPCPDC*AGEVAEFLARHSNVN LTIFTARLYYFQYPCYQEGLRSLSQEGVAVEIMDYEDFKYCWENFVYNDNEPFKPWK GLKTNFRLLKRRLRESLQ (SEQ ID NO: 83)

[0311] (斜体:核酸编辑域)

[0312] 大猩猩APOBEC3C

[0313] MNPQIRNPMKAMYPGTFYFQFKNLWEANDRNETWLCFTVEGIKRRSVVSWKTGVFR NQVDSETH*CHAERCFLSWFCDDILSPNTNYQVTWYTSWSPCPEC*AGEVAEFLARHSNVN LTIFTARLYYFQDTDYQEGLRSLSQEGVAVKIMDYKDFKYCWENFVYNDDEPFKPWK GLKYNFRFLKRRLQEILE (SEQ ID NO: 84)

[0314] (斜体:核酸编辑域)

[0315] 人APOBEC-3A:

[0316] MEASPASGPRHLMDPHIFTSNFNNGIGRHKTYLCYEVERLDNGTSVKMDQHRGFLHN QAKNLLCGFYGR*HAELRFLDLVPSLQLDPAQIYRVTWFISWSPCFSWGC*AGEVRAFLQEN THVRLRIFAARIYDYDPLYKEALQMLRDAGAQVSIMTYDEFKHCWDTFVDHQGCPFQ PWDGLDEHSQALSGRLRAILQNQGN (SEQ ID NO: 85)

[0317] (斜体:核酸编辑域)

[0318] 猕猴APOBEC-3A:

[0319] MDGSPASRPRHLMDPNTFTFNFNNDLSVRGRHQTYLCYEVERLDNGTWVPMDERRG FLCNKAKNVPCGDYGC*HVELRFLCEVPSWQLDPAQTYRVTWFISWSPC*FRRGCAGQVR VFLQENKHVRLRIFAARIYDYDPLYQEALRTLRDAGAQVSIMTYEEFKHCWDTFVDR QGRPFQPWDGLDEHSQALSGRLRAILQNQGN (SEQ ID NO: 86)

[0320] (斜体:核酸编辑域)

[0321] 牛APOBEC-3A:

[0322] MDEYTFTENFNNQGWPSKTYLCYEMERLDGDATIPLDEYKGFVRNKGLDQPEKPC*HAELYFLGKIHSWNLDRNQHYRLTCFISWSPC*YDCAQKLTTFLKENHHISLHILASRIYTHNRFGCHQSGLCELQAAGARITIMTFEDFKHCWETFVDHKGKPFQPWEGLNVKSQALCTELQAILKTQQN (SEQ ID NO: 87)

[0323] (斜体:核酸编辑域)

[0324] 人APOBEC-3H:

[0325] MALLTAETFRLQFNNKRRLRRPYYPRKALLCYQLTPQNGSTPTRGYFENKKKC*HAEICFINEIKSMGLDETQCYQVTCYLTWSPCSSC*AWELVDFIKAHDHLNLGIFASRLYYHWCKPQQKGLRLLCGSQVPVEVMGFPKFADCWENFVDHEKPLSFNPYKMLEELDKNSRAIKRRLERIKIPGVRAQGRYMDILCDAEV (SEQ ID NO: 88)

[0326] (斜体:核酸编辑域)

[0327] 猕猴APOBEC-3H:

[0328] MALLTAKTFSLQFNNKRRVNKPYYPRKALLCYQLTPQNGSTPTRGHLKNKKKDHAEIRFINKIKSMGLDETQCYQVTCYLTWSPCPSCAGELVDFIKAHRHLNLRIFASRLYYHWRPNYQEGLLLLCGSQVPVEVMGLPEFTDCWENFVDHKEPPSFNPSEKLEELDKNSQAIKRRLERIKSRSVDVLENGLRSLQLGPVTPSSSIRNSR(SEQ ID NO: 89)

[0329] 人APOBEC-3D:

[0330] MNPQIRNPMERMYRDTFYDNFENEPILYGRSYTWLCYEVKIKRGRSNLLWDTGVFRGPVLPKRQSNHRQEVYFRFEN*HAEMCFLSWFCGNRLPANRRFQITWFVSWNPCLPC*VVKVTKFLAEHPNVTLTISAARLYYYRDRDWRWVLLRLHKAGARVKIMDYEDFAYCWENFVCNEGQPFMPWYKFDDNYASLHRTLKEILRNPMEAMYPHIFYFHFKNLLKACGRNESWLCFTMEVTKHHSAVFRKRGVFRNQVDPETHC*HAERCFLSWFCDDILSPNTNYEVTW YTSWSPCPEC*AGEVAEFLARHSNVNLTIFTARLCYFWDTDYQEGLCSLSQEGASVKIMGYKDFVSCWKNFVYSDDEPFKPWKGLQTNFRLLKRRLREILQ (SEQ ID NO: 90)

[0331] (斜体:核酸编辑域)

[0332] 人APOBEC-1:

[0333] MTSEKGPSTGDPTLRRRIEPWEFDVFYDPRELRKEACLLYEIKWGMSRKIWRSSGKNTTNHVEVNFIKKFTSERDFHPSMSCSITWFLSWSPCWECSQAIREFLSRHPGVTLVIYVARLFWHMDQQNRQGLRDLVNSGVTIQIMRASEYYHCWRNFVNYPPGDEAHWPQYPPLWMMLYALELHCIILSLPPCLKISRRWQNHLTFFRLHLQNCHYQTIPPHILLATGLIHPSVAWR(SEQ ID NO:91)

[0334] 小鼠APOBEC-1:

[0335] MSSETGPVAVDPTLRRRIEPHEFEVFFDPRELRKETCLLYEINWGGRHSVWRHTSQNTSNHVEVNFLEKFTTERYFRPNTRCSITWFLSWSPCGECSRAITEFLSRHPYVTLFIYIARLYHHTDQRNRQGLRDLISSGVTIQIMTEQEYCYCWRNFVNYPPSNEAYWPRYPHLWVKLYVLELYCIILGLPPCLKILRRKQPQLTFFTITLQTCHYQRIPPHLLWATGLK(SEQ ID NO:92)

[0336] 大鼠APOBEC-1:

[0337] MSSETGPVAVDPTLRRRIEPHEFEVFFDPRELRKETCLLYEINWGGRHSIWRHTSQNTNKHVEVNFIEKFTTERYFCPNTRCSITWFLSWSPCGECSRAITEFLSRYPHVTLFIYIARLYHHADPRNRQGLRDLISSGVTIQIMTEQESGYCWRNFVNYSPSNEAHWPRYPHLWVRLYVLELYCIILGLPPCLNILRRKQPQLTFFTIALQSCHYQRLPPHILWATGLK(SEQ ID NO:93)

[0338] 人APOBEC-2:

[0339] MAQKEEAAVATEAASQNGEDLENLDDPEKLKELIELPPFEIVTGERLPANFFKFQFRNVEYSSGRNKTFLCYVVEAQGKGGQVQASRGYLEDEHAAAHAEEAFFNTILPAFDPALRYNVTWYVSSSPCAACADRIIKTLSKTKN

LRLLILVGRLFMWEEPEIQAALKKLKEAGCKLRIMKPQDFEYVWQNFVEQEEGESKAFQPWEDIQENFLYYEEKLA DILK(SEQ ID NO:94)

[0340] 小鼠APOBEC-2:

[0341] MAQKEEAAEAAAPASQNGDDLENLEDPEKLKELIDLPPFEIVTGVRLPVNFFKFQFRNVEYSSGRNKT FLCYVVEVQSKGGQAQATQGYLEDEHAGAHAEEAFFNTILPAFDPALKYNVTWYVSSSPCAACADRILKTLSKTKN LRLLILVSRLFMWEEPEVQAALKKLKEAGCKLRIMKPQDFEYIWQNFVEQEEGESKAFEPWEDIQENFLYYEEKLA DILK(SEQ ID NO:95)

[0342] 大鼠APOBEC-2:

[0343] MAQKEEAAEAAAPASQNGDDLENLEDPEKLKELIDLPPFEIVTGVRLPVNFFKFQFRNVEYSSGRNKT FLCYVVEAQSKGGQVQATQGYLEDEHAGAHAEEAFFNTILPAFDPALKYNVTWYVSSSPCAACADRILKTLSKTKN LRLLILVSRLFMWEEPEVQAALKKLKEAGCKLRIMKPQDFEYLWQNFVEQEEGESKAFEPWEDIQENFLYYEEKLA DILK(SEQ ID NO:96)

[0344] 牛APOBEC-2:

[0345] MAQKEEAAAAAEPASQNGEEVENLEDPEKLKELIELPPFEIVTGERLPAHYFKFQFRNVEYSSGRNKT FLCYVVEAQSKGGQVQASRGYLEDEHATNHAEEAFFNSIMPTFDPALRYMVTWYVSSSPCAACADRIVKTLNKTKN LRLLILVGRLFMWEEPEIQAALRKLKEAGCRLRIMKPQDFEYIWQNFVEQEEGESKAFEPWEDIQENFLYYEEKLA DILK(SEQ ID NO:97)

[0346] 海七鳃鳗CDA1(pmCDA1)

[0347] MTDAEYVRIHEKLDIYTFKKQFFNNKKSVSHRCYVLFELKRRGERRACFWGYAVNKPQSGTERGIHAE IFSIRKVEEYLRDNPGQFTINWYSSWSPCADCAEKILEWYNQELRGNGHTLKIWACKLYYEKNARNQIGLWNLRDN GVGLNVMVSEHYQCCRKIFIQSSHNQLNENRWLEKTLKRAEKRRSELSIMIQVKILHTTKSPAV(SEQ ID NO: 98)

[0348] 人APOBEC3G D316R_D317R

[0349] MKPHFRNTVERMYRDTFSYNFYNRPILSRRNTVWLCYEVKTKGPSRPPLDAKIFRGQVYSELKYHPEM RFFHWFSKWRKLHRDQEYEVTWYISWSPCTKCTRDMATFLAEDPKVTLTIFVARLYYFWDPDYQEALRSLCQKRDG PRATMKIMNYDEFQHCWSKFVYSQRELFEPWNNLPKYYILLHIMLGEILRHSMDPPTFTFNFNNEPWVRGRHETYL CYEVERMHNDTWVLLNQRRGFLCNQAPHKHGFLEGRHAELCFLDVIPFWKLDLDQDYRVTCFTSWSPCFSCAQEMA KFISKNKHVSLCIFTARIYRRQGRCQEGLRTLAEAGAKISIMTYSEFKHCWDTFVDHQGCPFQPWDGLDEHSQDLS GRLRAILQNQEN(SEQ ID NO:99)

[0350] 人APOBEC3G A链

[0351] MDPPTFTFNFNNEPWVRGRHETYLCYEVERMHNDTWVLLNQRRGFLCNQAPHKHGFLEGRHAELCFLD VIPFWKLDLDQDYRVTCFTSWSPCFSCAQEMAKFISKNKHVSLCIFTARIYDDQGRCQEGLRTLAEAGAKISIMTY SEFKHCWDTFVDHQGCPFQPWDGLDEHSQDLSGRLRAILQ(SEQ ID NO:100)

[0352] 人APOBEC3G A链D120R_D121R

[0353] MDPPTFTFNFNNEPWVRGRHETYLCYEVERMHNDTWVLLNQRRGFLCNQAPHKHGFLEGRHAELCFLD VIPFWKLDLDQDYRVTCFTSWSPCFSCAQEMAKFISKNKHVSLCIFTARIYRRQGRCQEGLRTLAEAGAKISIMTY SEFKHCWDTFVDHQGCPFQPWDGLDEHSQDLSGRLRAILQ(SEQ ID NO:101)

[0354] 调节碱基编辑器的编辑窗的脱氨酶域

[0355] 本公开内容的一些方面基于以下认识:调节本文中提供的任何融合蛋白的脱氨酶

域催化活性(例如通过在脱氨酶域中进行点突变)来影响融合蛋白(例如,碱基编辑器)的持续合成能力。例如,减少但不消除碱基编辑融合蛋白内脱氨酶域催化活性的突变可以使得脱氨酶域不太可能会催化邻近靶残基的残基脱氨基,从而缩小脱氨基窗。缩小脱氨基窗的能力可以防止与特定靶残基相邻的残基的不想要的脱氨基,这可以减少或防止脱靶效应。

[0356] 在一些实施方案中,本文提供的任何融合蛋白包含具有降低的催化脱氨酶活性的脱氨酶域(例如胞苷脱氨酶域)。在一些实施方案中,本文提供的任何融合蛋白包含脱氨酶域(例如,胞苷脱氨酶域),其与合适的对照相比具有降低的催化脱氨酶活性。例如,合适的对照可以是在将一种或多种突变引入脱氨酶之前的脱氨酶的脱氨酶活性。在其他实施方案中,合适的对照可以是野生型脱氨酶。在一些实施方案中,合适的对照是野生型载脂蛋白B mRNA编辑复合物(APOBEC)家族脱氨酶。在一些实施方案中,合适的对照是APOBEC1脱氨酶,APOBEC2脱氨酶,APOBEC3A脱氨酶,APOBEC3B脱氨酶,APOBEC3C脱氨酶,APOBEC3D脱氨酶,APOBEC3F脱氨酶,APOBEC3G脱氨酶或APOBEC3G脱氨酶。在一些实施方案中,合适的对照是激活诱导的脱氨酶(AID)。在一些实施方案中,合适的对照是来自海七鳃鳗的胞苷脱氨酶1(pmCDA1)。在一些实施方案中,脱氨酶域可以是与合适的对照相比具有小至少1%,至少5%,至少15%,至少20%,至少25%,至少30%,至少40%,至少50%,至少60%,至少70%,至少80%,至少90%或至少95%的催化脱氨酶活性的脱氨酶域。

[0357] 在一些实施方案中,本文提供的任何融合蛋白包含APOBEC脱氨酶,其包含一个或多个选自下组的突变:rAPOBEC1(SEQ ID NO:93)的H121X,H122X,R126X,R126X,R118X,W90X,W90X和R132X,或另一种APOBEC脱氨酶中的一个或多个相应突变,其中X是任何氨基酸。在一些实施方案中,本文提供的任何融合蛋白包含APOBEC脱氨酶,其包含一个或多个选自下组的突变:rAPOBEC1(SEQ ID NO:93)的H121R,H122R,R126A,R126E,R118A,W90A,W90Y和R132E,或另一种APOBEC脱氨酶中的一个或多个相应突变。

[0358] 在一些实施方案中,本文提供的任何融合蛋白包含APOBEC脱氨酶,其包含一个或多个选自下组的突变:hAPOBEC3G(SEQ ID NO:77)的D316X,D317X,R320X,R320X,R313X,W285X,W285X,R326X,或另一种APOBEC脱氨酶中的一个或多个相应突变,其中X是任何氨基酸。在一些实施方案中,本文提供的任何融合蛋白包含APOBEC脱氨酶,其包含一个或多个选自下组的突变:hAPOBEC3G(SEQ ID NO:77)的D316R,D317R,R320A,R320E,R313A,W285A,W285Y,R326E,或另一种APOBEC脱氨酶中的一个或多个相应突变。

[0359] 在一些实施方案中,本文提供的任何融合蛋白包含APOBEC脱氨酶,其包含rAPOBEC1(SEQ ID NO:93)的H121R和H122R突变,或另一种APOBEC脱氨酶中的一个或多个相应突变。在一些实施方案中,本文提供的任何融合蛋白包含APOBEC脱氨酶,其包含rAPOBEC1(SEQ ID NO:93)的R126A突变,或另一种APOBEC脱氨酶中的一个或多个相应突变。在一些实施方案中,本文提供的任何融合蛋白包含APOBEC脱氨酶,其包含rAPOBEC1(SEQ ID NO:93)的R126E突变,或另一种APOBEC脱氨酶中的一个或多个相应突变。在一些实施方案中,本文提供的任何融合蛋白包含APOBEC脱氨酶,其包含rAPOBEC1(SEQ ID NO:93)的R118A突变,或另一种APOBEC脱氨酶中的一个或多个相应突变。在一些实施方案中,本文提供的任何融合蛋白包含APOBEC脱氨酶,其包含rAPOBEC1(SEQ ID NO:93)的W90A突变,或另一种APOBEC脱氨酶中的一个或多个相应突变。在一些实施方案中,本文提供的任何融合蛋白包含APOBEC脱氨酶,其包含rAPOBEC1(SEQ ID NO:93)的W90Y突变,或另一种APOBEC脱氨酶中的一个或

多个相应突变。在一些实施方案中，本文提供的任何融合蛋白包含APOBEC脱氨酶，其包含rAPOBEC1(SEQ ID NO:93)的R132E突变，或另一种APOBEC脱氨酶中的一个或多个相应突变。在一些实施方案中，本文提供的任何融合蛋白包含APOBEC脱氨酶，其包含rAPOBEC1(SEQ ID NO:93)的W90Y和R126E突变，或另一种APOBEC脱氨酶中的一个或多个相应突变。在一些实施方案中，本文提供的任何融合蛋白包含APOBEC脱氨酶，其包含rAPOBEC1(SEQ ID NO:93)的R126E和R132E突变，或另一种APOBEC脱氨酶中的一个或多个相应突变。在一些实施方案中，本文提供的任何融合蛋白包含APOBEC脱氨酶，其包含rAPOBEC1(SEQ ID NO:93)的W90Y和R132E突变，或另一种APOBEC脱氨酶中的一个或多个相应突变。在一些实施方案中，本文提供的任何融合蛋白包含APOBEC脱氨酶，其包含rAPOBEC1(SEQ ID NO:93)的W90Y，R126E和R132E突变，或另一种APOBEC脱氨酶中的一个或多个相应突变。

[0360] 在一些实施方案中，本文提供的任何融合蛋白包含APOBEC脱氨酶，其包含hAPOBEC3G(SEQ ID NO:77)的D316R和D317R突变，或另一APOBEC脱氨酶中的一个或多个相应突变。在一些实施方案中，本文提供的任何融合蛋白包含APOBEC脱氨酶，其包含hAPOBEC3G(SEQ ID NO:77)的R320A突变，或另一种APOBEC脱氨酶中的一个或多个相应突变。在一些实施方案中，本文提供的任何融合蛋白包含APOBEC脱氨酶，其包含hAPOBEC3G(SEQ ID NO:77)的R320E突变，或另一种APOBEC脱氨酶中的一个或多个相应突变。在一些实施方案中，本文提供的任何融合蛋白包含APOBEC脱氨酶，其包含hAPOBEC3G(SEQ ID NO:77)的R313A突变，或另一APOBEC脱氨酶中的一个或多个相应突变。在一些实施方案中，本文提供的任何融合蛋白包含APOBEC脱氨酶，其包含hAPOBEC3G(SEQ ID NO:77)的W285A突变，或另一种APOBEC脱氨酶中的一个或多个相应突变。在一些实施方案中，本文提供的任何融合蛋白均包含APOBEC脱氨酶，其包含hAPOBEC3G(SEQ ID NO:77)的W285Y突变，或另一种APOBEC脱氨酶中的一个或多个相应突变。本文中的本文包含APOBEC脱氨酶，其包含hAPOBEC3G(SEQ ID NO:77)的R326E突变，或另一APOBEC脱氨酶中的一个或多个相应突变。在一些实施方案中，本文提供的任何融合蛋白包含APOBEC脱氨酶，其包含hAPOBEC3G(SEQ ID NO:77)的W285Y和R320E突变，或另一种APOBEC脱氨酶中的一个或多个相应突变。在一些实施方案中，本文提供的任何融合蛋白包含APOBEC脱氨酶，其包含hAPOBEC3G(SEQ ID NO:77)的R320E和R326E突变，或另一种APOBEC脱氨酶中的一个或多个相应突变。在一些实施方案中，本文提供的任何融合蛋白包含APOBEC脱氨酶，其包含hAPOBEC3G(SEQ ID NO:77)的W285Y和R326E突变，或另一种APOBEC脱氨酶中的一个或多个相应突变。在一些实施方案中，本文提供的任何融合蛋白包含APOBEC脱氨酶，其包含hAPOBEC3G(SEQ ID NO:77)的W285Y，R320E和R326E突变，或另一种APOBEC脱氨酶中的一个或多个相应突变。

[0361] 融合蛋白，包含核酸酶可编程DNA结合蛋白(napDNAbp)、胞苷脱氨酶和尿嘧啶结合蛋白(UBP)

[0362] 本公开内容的一些方面提供了融合蛋白，其包含核酸可编程DNA结合蛋白(napDNAbp)、胞苷脱氨酶和尿嘧啶结合蛋白(UBP)。在一些实施方案中，本文提供的任何融合蛋白是碱基编辑器。在一些实施方案中，UBP是尿嘧啶修饰酶。在一些实施方案中，UBP是尿嘧啶碱基切除酶。在一些实施方案中，UBP是尿嘧啶DNA糖基化酶。在一些实施方案中，UBP是本文提供的任何尿嘧啶结合蛋白。例如，UBP可以是UDG，UdgX，UdgX*，UdgX_On或SMUG1。在一些实施方案中，UBP包含与尿嘧啶结合蛋白，尿嘧啶碱基切除酶或尿嘧啶DNA糖基化酶

(UDG)至少75％,80％,85％,90％,95％,96％,97％,98％,99％或99.5％相同的氨基酸序列。在一些实施方案中,UBP包含与本文提供的任何尿嘧啶结合蛋白至少75％,80％,85％,90％,95％,96％,97％,98％,99％或99.5％相同的氨基酸序列。例如,UBP可以包含与SEQ ID NO:48-53中的任一项至少75％,80％,85％,90％,95％,96％,97％,98％,99％或99.5％相同的氨基酸序列。在一些实施方案中,UBP包含SEQ ID NO:48-53中任一项的氨基酸序列。

[0363] 在一些实施方案中,napDNAbp是Cas9域,Cpf1域,CasX域,CasY域,C2c1域,C2c2域,aC2c3域或Argonaute域。在一些实施方案中,napDNAbp是本文提供的任何napDNAbp。在一些实施方案中,本文提供的任何融合蛋白的napDNAbp是Cas9域。Cas9域可以是本文提供的任何Cas9域或Cas9蛋白(例如,dCas9或nCas9)。在一些实施方案中,本文提供的任何Cas9域或Cas9蛋白(例如,dCas9或nCas9)可以与本文提供的任何胞苷脱氨酶融合。在一些实施方案中,融合蛋白包含以下结构:

[0364] $NH_2$-[胞苷脱氨酶]-[napDNAbp]-[UBP]-COOH;

[0365] $NH_2$-[胞苷脱氨酶]-[UBP]-[napDNAbp]-COOH;

[0366] $NH_2$-[UBP]-[胞苷脱氨酶]-[napDNAbp]-COOH;

[0367] $NH_2$-[UBP]-[napDNAbp]-[胞苷脱氨酶]-COOH;

[0368] $NH_2$-[napDNAbp]-[UBP]-[胞苷脱氨酶]-COOH;或

[0369] $NH_2$-[napDNAbp]-[胞苷脱氨酶]-[UBP]-COOH。

[0370] 在一些实施方案中,包含胞苷脱氨酶,napDNAbp(例如Cas9域)和UBP的融合蛋白不包括接头序列。在一些实施方案中,在胞苷脱氨酶域和napDNAbp之间存在接头。在一些实施方案中,在胞苷脱氨酶域和UBP之间存在接头。在一些实施方案中,在napDNAbp和UBP之间存在接头。在一些实施例中,以上通用结构中使用的“-”表示任选的接头的存在。在一些实施方案中,胞苷脱氨酶和napDNAbp,胞苷脱氨酶和UBP和/或napDNAbp和UBP经由本文提供的任何接头融合。例如,在一些实施方案中,胞苷脱氨酶和napDNAbp,胞苷脱氨酶和UBP,和/或napDNAbp和UBP经由下面标题为“接头”的部分中提供的任何接头融合。在一些实施方案中,胞苷脱氨酶和napDNAbp,胞苷脱氨酶和UBP和/或napDNAbp和UBP经由包含1至200个氨基酸的接头融合。在一些实施方案中,胞苷脱氨酶和napDNAbp,胞苷脱氨酶和UBP,和/或napDNAbp和UBP经由包含长度1至5、1至10、1至20、1至30、1至40、1至50、1至60、1至80、1至100、1至150、1至200、5至10、5至20、5至30、5至40、5至60、5至80、5至100、5至150、5至200、10至20、10至30、10至40、10至50、10至60、10至80、10至100、10至150、10至200、20至30、20至40、20至50、20至60、20至80、20至100、20至150、20至200、30至40、30至50、30至60、30至80、30至100、30至150、30至200、40至50、40至60、40至80、40至100、40至150、40至200、50至6050至80、50至100、50至150、50至200、60至80、60至100、60至150、60至200、80至100、80至150、80至200、100至150、100至200、或150至200个氨基酸的接头融合。在一些实施方案中,胞苷脱氨酶和napDNAbp,胞苷脱氨酶和UBP和/或napDNAbp和UBP经由长度为4、16、24、32、91或104个氨基酸的接头融合。在一些实施方案中,胞苷脱氨酶和napDNAbp,胞苷脱氨酶和UBP和/或napDNAbp和UBP经由包含SGSETPGTSESATPES(SEQ ID NO:102),SGGS(SEQ ID NO:103),SGGSSGSETPGTSESATPESSGGS(SEQ ID NO:107),SGGSSGGSSGSETPGTSESATPESSGGSSGGS(SEQ ID NO:108),GGSGGSPGSPAGSPTSTEEGTSESATPESGPGTSTEPSEGSAPGSPAGSPTSTEEGTSTEPSEGSAPGTSTEPSEGSAPGTSESATPESGPGSEPATSGGSGGS(SEQ ID NO:109),或SGGSGGSGGS(SEQ

ID NO:120)的氨基酸序列的接头融合。在一些实施方案中,胞苷脱氨酶和napDNAbp,胞苷脱氨酶和UBP和/或napDNAbp和UBP经由包含氨基酸序列SGSETPGTSESATPES(SEQ ID NO:102)的接头(其也可以称为XTEN接头)融合。

[0371] 融合蛋白,其包含核酸酶可编程DNA结合蛋白(napDNAbp)、胞苷脱氨酶和核酸聚合酶(NAP)域

[0372] 本公开内容的一些方面提供了融合蛋白,其包含核酸可编程DNA结合蛋白(napDNAbp),胞苷脱氨酶和核酸聚合酶(NAP)域。在一些实施方案中,本文提供的任何融合蛋白是碱基编辑器。在一些实施方案中,NAP是真核核酸聚合酶。在一些实施方案中,NAP是DNA聚合酶。在一些实施方案中,NAP具有跨损伤聚合酶活性。在一些实施方案中,NAP是跨损伤DNA聚合酶。在一些实施方案中,NAP是Rev7,Rev1复合物,聚合酶iota,聚合酶kappa或聚合酶eta。在一些实施方案中,NAP是真核聚合酶alpha,beta,gamma,delta,epsilon,gamma,eta,iota,kappa,lambda,mu,或nu。在一些实施方案中,NAP包含与核酸聚合酶(例如,跨损伤DNA聚合酶)至少75%,80%,85%,90%,95%,96%,97%,98%,99%或99.5%相同的氨基酸序列。在一些实施方案中,NAP包含与本文提供的任何核酸聚合酶至少75%,80%,85%,90%,95%,96%,97%,98%,99%或99.5%相同的氨基酸序列。例如,NAP可以包含与SEQ ID NO:54-64中任一项至少75%,80%,85%,90%,95%,96%,97%,98%,99%或99.5%相同的氨基酸序列。在一些实施方案中,NAP包含SEQ ID NO:54-64中任一项的氨基酸序列。

[0373] 在一些实施方案中,napDNAbp是Cas9域,Cpf1域,CasX域,CasY域,C2c1域,C2c2域,C2c3域或Argonaute域。在一些实施方案中,napDNAbp是本文提供的任何napDNAbp。在一些实施方案中,本文提供的任何融合蛋白的napDNAbp是Cas9域。Cas9域可以是本文提供的任何Cas9域或Cas9蛋白(例如,dCas9或nCas9)。在一些实施方案中,本文提供的任何Cas9域或Cas9蛋白(例如,dCas9或nCas9)可以与本文提供的任何胞苷脱氨酶融合。在一些实施方案中,融合蛋白包含以下结构:

[0374] $NH_2$-[胞苷脱氨酶]-[napDNAbp]-[NAP]-COOH;

[0375] $NH_2$-[胞苷脱氨酶]-[NAP]-[napDNAbp]-COOH;

[0376] $NH_2$-[NAP]-[胞苷脱氨酶]-[napDNAbp]-COOH;

[0377] $NH_2$-[NAP]-[napDNAbp]-[胞苷脱氨酶]-COOH;

[0378] $NH_2$-[napDNAbp]-[NAP]-[胞苷脱氨酶]-COOH;或

[0379] $NH_2$-[napDNAbp]-[胞苷脱氨酶]-[NAP]-COOH。

[0380] 在一些实施方案中,包含胞苷脱氨酶,napDNAbp(例如,Cas9域)和NAP的融合蛋白不包括接头序列。在一些实施方案中,在胞苷脱氨酶域和napDNAbp之间存在接头。在一些实施方案中,在胞苷脱氨酶域和NAP之间存在接头。在一些实施方案中,在napDNAbp和NAP之间存在接头。在一些实施例中,以上通用结构中使用的“-”表示任选的接头的存在。在一些实施方案中,胞苷脱氨酶和napDNAbp,胞苷脱氨酶和NAP和/或napDNAbp和NAP经由本文提供的任何接头融合。例如,在一些实施方案中,胞苷脱氨酶和napDNAbp,胞苷脱氨酶和NAP,和/或napDNAbp和NAP经由下文标题为“接头”的部分中提供的任何接头融合。在一些实施方案中,胞苷脱氨酶和napDNAbp,胞苷脱氨酶和NAP和/或napDNAbp和NAP经由包含1至200个氨基酸的接头融合。在一些实施方案中,胞苷脱氨酶和napDNAbp,胞苷脱氨酶和NAP和/或napDNAbp

和NAP经由包含长度为1至5、1至10、1至20、1至30、1至40、1至50、1至60、1至80、1至100、1至150、1至200、5至10、5至20、5至30、5至40、5至60、5至80、5至100、5至150、5至200、10至20、10至30、10至40、10至50、10至60、10至80、10至100、10至150、10至200、20至30、20至40、20至50、20至60、20至80、20至100、20至150、20至200、30至40、30至50、30至60、30至80、30至100、30至150、30至200、40至50、40至60、40至80、40至100、40至150、40至200、50至6050至80、50至100、50至150、50至200、60至80、60至100、60至150、60至200、80至100、80至150、80至200、100至150、100至200、或150至200个氨基酸的接头融合。在一些实施方案中，胞苷脱氨酶和napDNAbp，胞苷脱氨酶和NAP，和/或napDNAbp和NAP经由长度为4、16、32或104个氨基酸的接头融合。在一些实施方案中，胞苷脱氨酶和napDNAbp，胞苷脱氨酶和NAP，和/或napDNAbp和NAP经由包含SGSETPGTSESATPES(SEQ ID NO:102)，SGGS(SEQ ID NO:103)，SGGSSGSETPGTSESATPESSGGS(SEQ ID NO:107)，SGGSSGGSSGSETPGTSESATPESSGGSSGGS(SEQ ID NO:108)，GGSGGSPGSPAGSPTSTEEGTSESATPESGPGTSTEPSEGSAPGSPAGSPTSTEEGTSTEPSEGSAPGTSTEPSEGSAPGTSESATPESGPGSEPATSGGSGGS(SEQ ID NO:109)，或SGGSGGSGGS(SEQ ID NO:120)的氨基酸序列的接头融合。在一些实施方案中，胞苷脱氨酶和napDNAbp，胞苷脱氨酶和NAP和/或napDNAbp和NAP经由包含氨基酸序列SGSETPGTSESATPES(SEQ ID NO:102)的接头(其也可以称为XTEN接头)融合。

[0381] 融合蛋白，其包含核酸酶可编程DNA结合蛋白(napDNAbp)、胞苷脱氨酶、尿嘧啶结合蛋白(UBP)和核酸聚合酶(NAP)域

[0382] 本公开内容的一些方面提供融合蛋白，其包含核酸可编程DNA结合蛋白(napDNAbp)，胞苷脱氨酶，尿嘧啶结合蛋白(UBP)和核酸聚合酶(NAP)域。在一些实施方案中，本文提供的任何融合蛋白是碱基编辑器。在一些实施方案中，NAP是真核核酸聚合酶。在一些实施方案中，NAP是DNA聚合酶。在一些实施方案中，NAP具有跨损伤聚合酶活性。在一些实施方案中，NAP是跨损伤DNA聚合酶。在一些实施方案中，NAP是Rev7，Rev1复合物，聚合酶iota，聚合酶kappa或聚合酶eta。在一些实施方案中，NAP是真核聚合酶alpha，beta，gamma，delta，epsilon，gamma，eta，iota，kappa，lambda，mu，或nu。在一些实施方案中，NAP包含与核酸聚合酶(例如，跨损伤DNA聚合酶)至少75％，80％，85％，90％，95％，96％，97％，98％，99％或99.5％相同的氨基酸序列。在一些实施方案中，NAP包含与本文提供的任何核酸聚合酶至少75％，80％，85％，90％，95％，96％，97％，98％，99％或99.5％相同的氨基酸序列。例如，NAP可以包含与SEQ ID NO:54-64中任一项至少75％，80％，85％，90％，95％，96％，97％，98％，99％或99.5％相同的氨基酸序列。在一些实施方案中，NAP包含SEQ ID NO:54-64中任一项的氨基酸序列。

[0383] 在一些实施方案中，UBP是尿嘧啶修饰酶。在一些实施方案中，UBP是尿嘧啶碱基切除酶。在一些实施方案中，UBP是尿嘧啶DNA糖基化酶。在一些实施方案中，UBP是本文提供的任何尿嘧啶结合蛋白。例如，UBP可以是UDG，UdgX，UdgX*，UdgX_On或SMUG1。在一些实施方案中，UBP包含与尿嘧啶结合蛋白，尿嘧啶碱基切除酶或尿嘧啶DNA糖基化酶(UDG)至少75％，80％，85％，90％，95％，96％，97％，98％，99％或99.5％相同的氨基酸序列。在一些实施方案中，UBP包含与本文提供的任何尿嘧啶结合蛋白至少75％，80％，85％，90％，95％，96％，97％，98％，99％或99.5％相同的氨基酸序列。例如，UBP可包含与SEQ ID NO:48-53中任一项至少75％，80％，85％，90％，95％，96％，97％，98％，99％或99.5％相同的氨基酸序列。在

一些实施方案中,UBP包含SEQ ID NO:48-53中任一项的氨基酸序列。

[0384] 在一些实施方案中,napDNAbp是Cas9域,Cpf1域,CasX域,CasY域,C2c1域,C2c2域,aC2c3域或Argonaute域。在一些实施方案中,napDNAbp是本文提供的任何napDNAbp。在一些实施方案中,本文提供的任何融合蛋白的napDNAbp是Cas9域。Cas9域可以是本文提供的任何Cas9域或Cas9蛋白(例如,dCas9或nCas9)。在一些实施方案中,本文提供的任何Cas9域或Cas9蛋白(例如,dCas9或nCas9)可以与本文提供的任何胞苷脱氨酶融合。在一些实施方案中,融合蛋白包含以下结构:

[0385] $NH_2$-[NAP]-[胞苷脱氨酶]-[napDNAbp]-[UBP]-COOH;

[0386] $NH_2$-[胞苷脱氨酶]-[NAP]-[napDNAbp]-[UBP]-COOH;

[0387] $NH_2$-[胞苷脱氨酶]-[napDNAbp]-[NAP]-[UBP]-COOH;

[0388] $NH_2$-[胞苷脱氨酶]-[napDNAbp]-[UBP]-[NAP]-COOH;

[0389] $NH_2$-[NAP]-[胞苷脱氨酶]-[UBP]-[napDNAbp]-COOH;

[0390] $NH_2$-[胞苷脱氨酶]-[NAP]-[UBP]-[napDNAbp]-COOH;

[0391] $NH_2$-[胞苷脱氨酶]-[UBP]-[NAP]-[napDNAbp]-COOH;

[0392] $NH_2$-[胞苷脱氨酶]-[UBP]-[napDNAbp]-[NAP]-COOH;

[0393] $NH_2$-[NAP]-[UBP]-[胞苷脱氨酶]-[napDNAbp]-COOH;

[0394] $NH_2$-[UBP]-[NAP]-[胞苷脱氨酶]-[napDNAbp]-COOH;

[0395] $NH_2$-[UBP]-[胞苷脱氨酶]-[NAP]-[napDNAbp]-COOH;

[0396] $NH_2$-[UBP]-[胞苷脱氨酶]-[napDNAbp]-[NAP]-COOH;

[0397] $NH_2$-[NAP]-[UBP]-[napDNAbp]-[胞苷脱氨酶]-COOH;

[0398] $NH_2$-[UBP]-[NAP]-[napDNAbp]-[胞苷脱氨酶]-COOH;

[0399] $NH_2$-[UBP]-[napDNAbp]-[NAP]-[胞苷脱氨酶]-COOH;

[0400] $NH_2$-[UBP]-[napDNAbp]-[胞苷脱氨酶]-[NAP]-COOH;

[0401] $NH_2$-[NAP]-[napDNAbp]-[UBP]-[胞苷脱氨酶]-COOH;

[0402] $NH_2$-[napDNAbp]-[NAP]-[UBP]-[胞苷脱氨酶]-COOH;

[0403] $NH_2$-[napDNAbp]-[UBP]-[NAP]-[胞苷脱氨酶]-COOH;

[0404] $NH_2$-[napDNAbp]-[UBP]-[胞苷脱氨酶]-[NAP]-COOH;

[0405] $NH_2$-[NAP]-[napDNAbp]-[胞苷脱氨酶]-[UBP]-COOH;

[0406] $NH_2$-[napDNAbp]-[NAP]-[胞苷脱氨酶]-[UBP]-COOH;

[0407] $NH_2$-[napDNAbp]-[胞苷脱氨酶]-[NAP]-[UBP]-COOH;或

[0408] $NH_2$-[napDNAbp]-[胞苷脱氨酶]-[UBP]-[NAP]-COOH。

[0409] 在一些实施方案中,包含胞苷脱氨酶,napDNAbp(例如,Cas9域),UBP和NAP的融合蛋白不包括接头序列。在一些实施方案中,在胞苷脱氨酶域与napDNAbp,NAP和/或UBP之间存在接头。在一些实施方案中,在napDNAbp和胞苷脱氨酶域,NAP和/或UBP之间存在接头。在一些实施方案中,在NAP和胞苷脱氨酶,napDNAbp和/或UBP之间存在接头。在一些实施方案中,在UBP和胞苷脱氨酶,napDNAbp和NAP之间存在接头。在一些实施方案中,以上通用结构中使用的“-”表示任选的接头的存在。在一些实施方案中,接头是本文提供的任何接头,例如在标题为“接头”的部分中。在一些实施方案中,接头包含1至200个氨基酸。在一些实施方案中,接头包含长度1至5、1至10、1至20、1至30、1至40、1至50、1至60、1至80、1至100、1至

150、1至200、5至10、5至20、5至30、5至40、5至60、5至80、5至100、5至150、5至200、10至20、10至30、10至40、10至50、10至60、10至80、10至100、10至150、10至200、20至30、20至40、20至50、20至60、20至80、20至100、20至150、20至200、30至40、30至50、30至60、30至80、30至100、30至150、30至200、40至50、40至60、40至80、40至100、40至150、40至200、50至6050至80、50至100、50至150、50至200、60至80、60至100、60至150、60至200、80至100、80至150、80至200、100至150、100至200、或150至200个氨基酸。在一些实施方案中，接头长度为4、16、32或104个氨基酸。在一些实施方案中，接头包含SGSETPGTSESATPES(SEQ ID NO:102)，SGGS(SEQ ID NO:103)，SGGSSGSETPGTSESATPESSGGS(SEQ ID NO:107)，SGGSSGGSSGSETPGTSESATPESSGGSSGGS(SEQ ID NO:108)，GGSGGSPGSPAGSPTSTEEGTSESATPESGPGTSTEPSEGSAPGSPAGSPTSTEEGTSTEPSEGSAPGTSTEPSEGSAPGTSESATPESGPGSEPATSGGSGGS(SEQ ID NO:109)，或SGGSGGSGGS(SEQ ID NO:120)的氨基酸序列。在一些实施方案中，接头包含氨基酸序列SGSETPGTSESATPES(SEQ ID NO:102)，其也可以称为XTEN接头。

[0410] 融合蛋白，其包含核酸酶可编程DNA结合蛋白(napDNAbp)和碱基切除酶(BEE)

[0411] 本公开内容的一些方面提供了融合蛋白，其包含核酸可编程DNA结合蛋白(napDNAbp)和碱基切除酶。在一些实施方案中，本文提供的任何融合蛋白是碱基编辑器。在一些实施方案中，碱基切除酶(BEE)是胞嘧啶，胸腺嘧啶，腺嘌呤，鸟嘌呤或尿嘧啶碱基切除酶。在一些实施方案中，碱基切除酶(BEE)是胞嘧啶碱基切除酶。在一些实施方案中，BEE是胸腺嘧啶碱基切除酶。在一些实施方案中，碱基切除酶包含与天然存在的BEE至少75％，80％，85％，90％，95％，96％，97％，98％，99％或99.5％相同的氨基酸序列。在一些实施方案中，碱基切除酶包含与SEQ ID NO:65-66中任一项至少75％，80％，85％，90％，95％，96％，97％，98％，99％或99.5％相同的氨基酸序列。在一些实施方案中，碱基切除酶包含SEQ ID NO:65-66中任一项的氨基酸序列。

[0412] 在一些实施方案中，napDNAbp是Cas9域，Cpf1域，CasX域，CasY域，C2c1域，C2c2域，aC2c3域或Argonaute域。在一些实施方案中，napDNAbp是本文提供的任何napDNAbp。在一些实施方案中，本文提供的任何融合蛋白的napDNAbp是Cas9域。Cas9域可以是本文提供的任何Cas9域或Cas9蛋白(例如，dCas9或nCas9)。在一些实施方案中，本文提供的任何Cas9域或Cas9蛋白(例如，dCas9或nCas9)可以与本文提供的任何胞苷脱氨酶融合。在一些实施方案中，融合蛋白包含以下结构：

[0413] $NH_2$-[BEE]-[napDNAbp]-COOH；或

[0414] $NH_2$-[napDNAbp]-[BEE]-COOH；

[0415] 在一些实施方案中，融合蛋白还包含核酸聚合酶(NAP)。在一些实施方案中，NAP是真核核酸聚合酶。在一些实施方案中，NAP是DNA聚合酶。在一些实施方案中，NAP具有跨损伤聚合酶活性。在一些实施方案中，NAP是跨损伤DNA聚合酶。在一些实施方案中，NAP是Rev7，Rev1复合物，聚合酶iota，聚合酶kappa或聚合酶eta。在一些实施方案中，NAP是真核聚合酶alpha，beta，gamma，delta，epsilon，gamma，eta，iota，kappa，lambda，mu，或nu。在一些实施方案中，NAP包含与核酸聚合酶(例如，跨损伤DNA聚合酶)至少75％，80％，85％，90％，95％，96％，97％，98％，99％或99.5％相同的氨基酸序列。在一些实施方案中，NAP包含与本文提供的任何核酸聚合酶至少75％，80％，85％，90％，95％，96％，97％，98％，99％或99.5％相同的氨基酸序列。例如，NAP可以包含与SEQ ID NO:54-64中任一项至少75％，80％，85％，

90％,95％,96％,97％,98％,99％或99.5％相同的氨基酸序列。在一些实施方案中,NAP包含SEQ ID NO:54-64中任一项的氨基酸序列。在一些实施方案中,融合蛋白包含以下结构:

[0416] $NH_2$-[BEE]-[napDNAbp]-[NAP]-COOH;

[0417] $NH_2$-[BEE]-[NAP]-[napDNAbp]-COOH;

[0418] $NH_2$-[NAP]-[BEE]-[napDNAbp]-COOH;

[0419] $NH_2$-[NAP]-[napDNAbp]-[BEE]-COOH;

[0420] $NH_2$-[napDNAbp]-[NAP]-[BEE]-COOH;或

[0421] $NH_2$-[napDNAbp]-[BEE]-[NAP]-COOH

[0422] 在一些实施方案中,包含napDNAbp(例如,Cas9域)和BEE的融合蛋白不包括接头序列。在一些实施方案中,包含napDNAbp(例如,Cas9域),BEE和NAP的融合蛋白不包括接头序列。在一些实施方案中,在napDNAbp和BEE之间存在接头。在一些实施方案中,在BEE和NAP和/或napDNAbp之间存在接头。在一些实施方案中,在NAP和BEE和/或napDNAbp之间存在接头。在一些实施方案中,在napDNAbp和BEE和/或NAP之间存在接头。在一些实施例中,以上通用结构中使用的“-”表示任选的接头的存在。在一些实施方案中,接头是本文提供的任何接头,例如在标题为“接头”的部分中。在一些实施方案中,接头包含1至200个氨基酸。在一些实施方案中,接头包含1至5、1至10、1至20、1至30、1至40、1至50、1至60、1至80、1至100、1至150、1至200、5至10、5至20、5至30、5至40、5至60、5至80、5至100、5至150、5至200、10至20、10至30、10至40、10至50、10至60、10至80、10至100、10至150、10至200、20至30、20至40、20至50、20至60、20至80、20至100、20至150、20至200、30至40、30至50、30至60、30至80、30至100、30至150、30至200、40至50、40至60、40至80、40至100、40至150、40至200、50至60 50至80、50至100、50至150、50至200、60至80、60至100、60至150、60至200、80至100、80至150、80至200、100至150、100至200、或150至200个氨基酸的长度。在一些实施方案中,接头长度为4、16、32或104个氨基酸。在一些实施方案中,接头包括SGSETPGTSESATPES(SEQ ID NO:102),SGGS(SEQ ID NO:103),SGGSSGSETPGTSESATPESSGGS(SEQ ID NO:107),SGGSSGGSSGSETPGTSESATPESSGGSSGGS(SEQ ID NO:108),GGSGGSPGSPAGSPTSTEEGTSESATPESGPGTSTEPSEGSAPGSPAGSPTSTEEGTSTEPSEGSAPGTSTEPSEGSAPGTSESATPESGPGSEPATSGGSGGS(SEQ ID NO:109),或SGGSGGSGGS(SEQ ID NO:120)的氨基酸序列。在一些实施方案中,接头包含氨基酸序列SGSETPGTSESATPES(SEQ ID NO:102),其也可以称为XTEN接头。

[0423] 包含核定位序列(NLS)的融合蛋白

[0424] 在一些实施方案中,本文提供的任何融合蛋白进一步包含一个或多个核靶向序列,例如核定位序列(NLS)。在一些实施方案中,NLS包含促进包含NLS的蛋白质输入细胞核中(例如,通过核转运)的氨基酸序列。在一些实施方案中,本文提供的任何融合蛋白进一步包含核定位序列(NLS)。在一些实施方案中,NLS与融合蛋白的N端融合。在一些实施方案中,NLS与融合蛋白的C端融合。在一些实施方案中,NLS与napDNAbp的N端融合。在一些实施方案中,NLS与napDNAbp的C端融合。在一些实施方案中,NLS与NAP的N端融合。在一些实施方案中,NLS与NAP的C端融合。在一些实施方案中,NLS与胞苷脱氨酶的N端融合。在一些实施方案中,NLS与胞苷脱氨酶的C端融合。在一些实施方案中,NLS与UBP的N端融合。在一些实施方案中,NLS与UBP的C端融合。在一些实施方案中,NLS与BEE的N端融合。在一些实施方案中,NLS与BEE的C端融合。在一些实施方案中,NLS经由一个或多个接头与融合蛋白融合。在一些实

施方案中,NLS与融合蛋白在没有接头的情况下融合。在一些实施方案中,NLS包含本文提供或引用的NLS序列的任一个的氨基酸序列。在一些实施方案中,NLS包含如SEQ ID NO:41或SEQ ID NO:42中所示的氨基酸序列。另外的核定位序列是本领域已知的并且对于技术人员是显而易见的。例如,NLS序列描述于Plank et al.,PCT/EP2000/011690中,其内容通过引用并入本文,用于其对示例性的核定位序列的公开。在一些实施方案中,NLS包含氨基酸序列PKKKRKV(SEQ ID NO:41),MDSLLMNRRKFLYQFKNVRWAKGRRETYLC(SEQ ID NO:42),KRTADGSEFESPKKKRKV(SEQ ID NO:43),KRGINDRNFWRGENGRKTR(SEQ ID NO:44),KKTGGPIYRRVDGKWRR(SEQ ID NO:45),RRELILYDKEEIRRIWR(SEQ ID NO:46),或AVSRKRKA(SEQ ID NO:47)。

[0425] 接头

[0426] 在某些实施方案中,接头可以用于连接本文所述的任何蛋白质或蛋白质域。接头可以简单地为共价键,或者它可以是长度为许多原子的聚合物接头。在某些实施方案中,接头是多肽或基于氨基酸。在其他实施方案中,接头不是肽样的。在某些实施方案中,接头是共价键(例如,碳-碳键、二硫键、碳-杂原子键等)。在某些实施方案中,接头是酰胺连接的碳-氮键。在某些实施方案中,接头是环状或非环状,取代或未取代的,分支或未分支的脂肪族或杂脂肪族接头。在某些实施方案中,接头是聚合物(例如,聚乙烯、聚乙二醇、聚酰胺、聚酯等)。在某些实施方案中,接头包含氨基烷酸的单体、二聚体或聚合物。在某些实施方案中,接头包含氨基烷酸(例如甘氨酸、乙酸、丙氨酸、β-丙氨酸、3-氨基丙酸、4-氨基丁酸、5-戊酸等)。在某些实施方案中,接头包含氨基己酸(Ahx)的单体、二聚体或聚合物。在某些实施方案中,接头基于碳环部分(例如,环戊烷、环己烷)。在其他实施方案中,接头包含聚乙二醇部分(PEG)。在其他实施方案中,接头包含氨基酸。在某些实施方案中,接头包含肽。在某些实施方案中,接头包含芳基或杂芳基部分。在某些实施方案中,接头基于苯环。接头可以包括官能化部分以促进亲核试剂(例如,硫醇、氨基)从肽附接至接头。任何亲电子试剂都可以用作接头的一部分。示例性亲电子试剂包括但不限于活化酯、活化酰胺、迈克尔受体、烷基卤化物、芳基卤化物、酰基卤化物和异硫氰酸酯。

[0427] 在一些实施方案中,接头是一个氨基酸或多个氨基酸(例如肽或蛋白质)。在一些实施方案中,接头是键(例如,共价键)、有机分子、基团、聚合物或化学部分。在一些实施方案中,接头的长度为5-100个氨基酸,例如长度为5、6、7、8、9、10、11、12、13、14、15、16、17、18、19、20、21、22、23、24、25、26、27、28、29、30、30-35、35-40、40-45、45-50、50-60、60-70、70-80、80-90、90-100、100-110、110-120、120-130、130-140、140-150或150-200个氨基酸。也考虑了更长或更短的接头。在一些实施方案中,接头包含氨基酸序列SGSETPGTSESATPES(SEQ ID NO:10),其也可以称为XTEN接头。在一些实施方案中,接头包含氨基酸序列SGGS(SEQ ID NO:103)。在一些实施方案中,接头包含$(SGGS)_n$(SEQ ID NO:103),$(GGGS)_n$(SEQ ID NO:104),$(GGGGS)_n$(SEQ ID NO:105),$(G)_n$(SEQ ID NO:121),$(EAAAK)_n$(SEQ ID NO:106),$(GGS)_n$(SEQ ID NO:122),SGSETPGTSESATPES(SEQ ID NO:102),SGGSGGSGGS(SEQ ID NO:120),或$(XP)_n$基序(SEQ ID NO:123)基序,或这些中任何的组合,其中n独立地是1和30之间的整数,并且其中X是任何氨基酸。在一些实施方案中,n是1、2、3、4、5、6、7、8、9、10、11、12、13、14或15。在一些实施方案中,接头包含SGSETPGTSESATPES(SEQ ID NO:102)和SGGS(SEQ ID NO:103)。在一些实施方案中,接头包含SGGSSGSETPGTSESATPESSGGS(SEQ ID NO:

107)。在一些实施方案中,接头包含SGGSSGGSSGSETPGTSESATPESSGGSSGGS(SEQ ID NO:108)。

[0428] 在一些实施方案中,接头包含GGSGGSPGSPAGSPTSTEEGTSESATPESGPGTSTEPSEGSAPGSPAGSPTSTEEGTSTEPSEGSAPGTSTEPSEGSAPGTSESATPESGPGSEPATSGGSGGS(SEQ ID NO:109)。在一些实施方案中,接头包含SGGSGGSGGS(SEQ ID NO:120)。

[0429] 具有引导核酸的核酸可编程DNA结合蛋白(napDNAbp)复合物

[0430] 本公开内容的一些方面提供了包含本文提供的任何融合蛋白,以及与融合蛋白的napDNAbp结合的引导核酸的复合物。本公开的一些方面提供了包含本文提供的任何融合蛋白,以及与融合蛋白的Cas9域(例如,dCas9、核酸酶活性Cas9或Cas9切口酶)结合的引导RNA的复合物。

[0431] 在一些实施方案中,引导核酸(例如引导RNA)为15-100个核苷酸长,并且包含与靶序列互补的至少10个连续核苷酸的序列。在一些实施方案中,引导RNA长度为15、16、17、18、19、20、21、22、23、24、25、26、27、28、29、30、31、32、33、34、35、36、37、38、39、40、41、42、43、44、45、46、47、48、49或50个核苷酸。在一些实施方案中,引导RNA包含与靶序列互补的15、16、17、18、19、20、21、22、23、24、25、26、27、28、29、30、31、32、33、34、35、36、37、38、39或40个连续核苷酸的序列。在一些实施方案中,靶序列是DNA序列。在一些实施方案中,靶序列是RNA序列。在一些实施方案中,靶序列是哺乳动物基因组中的序列。在一些实施方案中,靶序列是人基因组中的序列。在一些实施方案中,靶序列的3'末端紧邻规范PAM序列(NGG)。在一些实施方案中,引导核酸(例如,引导RNA)与疾病或病症相关的序列互补。在一些实施方案中,引导核酸(例如,引导RNA)与疾病或病症相关的序列互补,所述疾病或病症具有与本文中提供的任何疾病或病症相关的基因中的突变。在一些实施方案中,引导核酸(例如,引导RNA)与如本文中提供的疾病或病症相关的任何基因互补。

[0432] 使用融合蛋白的方法

[0433] 本公开内容的一些方面提供了使用本文中提供的任何融合蛋白(例如碱基编辑器)或包含引导核酸(例如gRNA)和本文提供的融合蛋白(例如碱基编辑器)的复合物的方法。例如,本公开内容的一些方面提供了包括使DNA或RNA分子与本文提供的任何融合蛋白或碱基编辑器接触,以及与至少一种引导核酸(例如,引导RNA)接触的方法,其中引导核酸(例如,引导RNA)约15-100个核苷酸长,并且包含与靶序列互补的至少10个连续核苷酸的序列。在一些实施方案中,靶序列的3'末端紧邻规范spCas9 PAM序列(NGG)。在一些实施方案中,靶序列的3'末端不与spCas9规范PAM序列(NGG)直接相邻。在一些实施方案中,靶序列的3'末端紧邻AGC、GAG、TTT、GTG或CAA序列。

[0434] 在一些实施方案中,靶DNA序列包含与疾病或病症相关的序列。在一些实施方案中,靶DNA序列包含与疾病或病症相关的点突变。在一些实施方案中,融合蛋白(例如包含napDNAbp、胞苷脱氨酶和尿嘧啶结合蛋白UBP)或复合物的活性导致点突变的校正。在一些实施方案中,靶DNA序列包含与疾病或病症相关的G至C或C至G点突变,并且其中突变体C碱基的脱氨基和/或切除导致与疾病或病症不相关的序列。在一些实施方案中,靶DNA序列编码蛋白质,并且点突变是密码子,并且与野生型密码子相比,导致由突变体密码子编码的氨基酸的变化。在一些实施方案中,突变体C的脱氨基导致突变体密码子编码的氨基酸的变化。在一些实施方案中,突变体C的脱氨基导致编码野生型氨基酸的密码子。在一些实施方

案中，接触是在受试者体内。在一些实施方案中，受试者患有或已经诊断出疾病或病症。在一些实施方案中，疾病或病症是22q13.3缺失综合征;2-甲基-3-羟基丁酸尿;3甲基巴豆酰-CoA羧化酶1缺陷;3-甲基巴豆酰CoA羧化酶2缺陷;3-甲基戊烯二酸尿2型;3-甲基戊烯二酸尿3型;3-甲基戊烯二酸尿V型;3-氧-5alpha-类固醇delta 4-脱氢酶缺陷;46,XY性逆转,1型;46,XY真两性畸形,SRY-相关;4-羟苯丙酮酸二加氧酶缺陷;异常脸型(Abnormal facial shape);异常糖基化(CDG IIa);软骨发生不全2型;全色盲2;全色盲5;全色盲6;全色盲7;获得性血红蛋白H病;尖头并指I型;肢端骨发育不全1伴有或不伴有激素抗性;肢端骨发育不全2,伴有或不伴有激素抗性;面骨发育不全,Cincinnati型;ACTH抗性;急性神经原戈谢病;Adams-Oliver综合征;Adams-Oliver综合征2;Adams-Oliver综合征4;Adams-Oliver综合征6;腺嘌呤磷酸核糖基转移酶缺陷;腺苷酸基琥珀酸裂解酶缺陷;青春期肾痨(Adolescent肾痨);肾上腺脑白质营养不良;成人交界型大疱性表皮松解;成人神经元蜡样质脂褐质沉积症;ADULT综合征;年龄相关黄斑变性14;年龄相关黄斑变性3;Aicardi Goutieres综合征5;Aicardi-goutieres综合征6;亚历山大病;alpha地中海贫血;Alpha-B晶体蛋白病变(crystallinopathy);奥尔波特综合征,常染色体隐性;奥尔波特综合征,X连锁隐性;儿童交替性偏瘫2;阿尔茨海默病;阿尔茨海默病,1型;阿尔茨海默病,3型;釉质形成不全,成熟不足型(Hypomaturation type),IIA3;釉质形成不全,1E型;Amish致死性小头畸形;AML-急性髓样白血病;淀粉样蛋白生成性转甲状腺素蛋白淀粉样变;肌萎缩侧索硬化16,青少年;肌萎缩侧索硬化6,常染色体隐性;肌萎缩侧索硬化1型;肌萎缩侧索硬化10型;肌萎缩侧索硬化2型;肌萎缩侧索硬化9型;Andersen Tawil综合征;贫血,红细胞生成障碍性先天性,IV型;贫血,非球形红细胞溶血性,由于G6PD缺陷;贫血,铁粒幼细胞,吡多辛难治性,常染色体隐性;Angelman综合征;血管病,遗传性,伴有肾病,动脉瘤和肌肉痉挛;无汗性外胚层发育不良伴有免疫缺陷;甲缺如;Antley-Bixler综合征伴有生殖器异常和类固醇生成障碍;Antley-Bixler综合征不伴有生殖器异常和类固醇生成障碍;再生障碍性贫血;载脂蛋白a-i缺陷;精氨酸酶缺乏症;心律失常性右心室心肌病;心律失常性右心室心肌病,11型;心律失常性右心室心肌病,9型;婴儿动脉石灰化(Arterial calcification of infancy);动脉扭曲综合征;先天性多数关节弯曲远端1型;关节炎肾功能不全胆汁淤积综合征(Arthrogryposis renal dysfunction cholestasis syndrome);关节挛缩,远端,5d型;Arts综合征;天冬氨酰基葡萄糖胺尿,finnish型;窒息性腔部营养不良2;共济失调伴有维生素E缺陷;共济失调-毛细血管扩张综合征;共济失调-毛细血管扩张样病症;骨发育不全症1型;心房颤动;心房颤动,家族性,10;房间隔缺损4;胎儿虹膜炎综合征;ATR-X综合征;非典型溶血尿毒症综合征1;听觉神经病常染色体隐性,1;耳髁综合征(Auriculocondylar syndrome)1;自身免疫性疾病,多系统,婴儿发作性;自身免疫性淋巴增殖性综合征,1A型;自身免疫性淋巴增殖性综合征,V型;常染色体显性遗传夜间发作性额叶癫痫;常染色体显性进行性外眼肌麻痹伴有线粒体DNA缺失2;常染色体显性进行性外眼肌麻痹伴有线粒体DNA缺失3;常染色体显性进行性外眼肌麻痹伴有线粒体DNA缺失4;常染色体隐性先天性鱼鳞病1;常染色体隐性先天性鱼鳞病5;常染色体隐性低磷维生素D难治性佝偻病(hypophosphatemic vitamin D refractory rickets);Axenfeld-rieger异常;Axenfeld-Rieger综合征1型;Axenfeld-Rieger综合征3型;Baraitser-Winter综合征1;Bardet-Biedl综合征;Bardet-Biedl合征10;Bardet-Biedl综合征12;Bardet-Biedl综合征2;Bardet-

Biedl综合征3;Bardet-Biedl综合征4;Bardet-Biedl综合征9;Bartter综合征出生前型2;Bartter综合征,4b型;基底神经节疾病,生物素响应;贝克肌营养不良;良性家族性新生儿癫痫1;性家族性新生儿-婴儿癫痫;良性复发性肝内胆汁淤积2;Bernard-Soulier综合征,B型;beta地中海贫血;Bietti结晶性角质视网膜营养不良;胆汁酸合成缺陷,先天性,2;生物素酰胺酶缺乏症;出血障碍,血小板型,19;血型--Lutheran抑制剂;布卢姆综合征;Bosley-Salih-Alorainy综合征;Boucher Neuhauser综合征;短指B2型;乳腺癌;乳腺-卵巢癌,家族性1;乳腺-卵巢癌,家族性2;支气管扩张;Brown-Vialetto-Van laere综合征;Brown-Vialetto-Van Laere综合征2;大疱性鱼鳞病样红斑;伯基特淋巴瘤;屈肢骨发育不全;Cap肌病2;碳水化合物缺陷糖蛋白综合征I型;碳水化合物缺陷糖蛋白综合征II型;结肠癌;胰腺癌;心律失常;心性脑肌病(Cardioencephalomyopathy),致死性婴儿,由于细胞色素C氧化酶缺陷3;心面皮肤综合征(Cardiofaciocutaneous syndrome);心面皮肤综合征2;心肌病;心肌病,限制性;卡尼复合征(Carney complex),1型;肉碱软脂酰转移酶I缺陷;白内障1;白内障,先天性,伴有感觉神经性聋,唐氏综合征样脸部外貌,身材矮小症,和智力低下;儿茶酚胺依赖性室性心动过速;中央轴空病;中枢性性早熟,;小脑性共济失调和低促性腺素性功能减退症;婴儿小脑性共济失调伴有进行性外眼肌麻痹;小脑性共济失调,耳聋和发作性睡病;脑淀粉样蛋白血管病,APP相关;脑常染色体显性动脉病伴有皮质下梗死和脑白质病;脑海绵状血管畸形1;大脑性瘫痪,痉挛性四肢瘫痪性,1;大脑-肋骨-下颌综合征;神经元蜡样脂褐质沉积症1;神经元蜡样脂褐质沉积症10;神经元蜡样脂褐质沉积症6;神经元蜡样脂褐质沉积症7;神经元蜡样脂褐质沉积症8;蜡样脂褐质沉积症,神经元,13;蜡样脂褐质沉积症,神经元,2;Ch\xc3\xa9diak-Higashi综合征;Char综合征;夏科-马里-图斯病;夏科-马里-图斯病1B型;夏科-马里-图斯病2B型;夏科-马里-图斯病type 2D型;夏科-马里-图斯病2I型;夏科-马里-图斯病2K型;夏科-马里-图斯病,轴突,伴有声带麻痹,常染色体隐性;夏科-马里-图斯病,脱髓鞘,1C型;夏科-马里-图斯病,显性中间E;夏科-马里-图斯病,2型;夏科-马里-图斯病,2A2型;夏科-马里-图斯病,4C型;夏科-马里-图斯病,4G型;夏科-马里-图斯病,IA型;夏科-马里-图斯病,IE型;夏科-马里-图斯病,IF型;夏科-马里-图斯病,X-连锁隐性,5型;CHARGE关联;儿童综合征;胆固烷醇贮积病;胆固醇单加氧酶(侧链切割)缺陷;点状软骨发育不良1,X连锁隐性;Chops综合征;染色体9q缺失综合征;慢性肉芽肿病,X-连锁;纤毛运动异常,原发性,14;纤毛运动异常,原发性,19;纤毛运动异常,原发性,3;纤毛运动异常,原发性,7;颅骨锁骨发育不良;科凯恩综合征A型;科-勒二氏综合征;科恩综合征;Cole疾病;结肠直肠癌,遗传性,非息肉病性,1型;组合性细胞和激素免疫缺陷伴有肉芽肿;组合性氧化磷酸化缺陷24;组合性氧化磷酸化缺陷9;常见可变免疫缺陷7;补体组分9缺陷;视锥-视杆营养不良10;视锥-视杆营养不良11;视锥-视杆营养不良3;视锥-视杆营养不良5;视锥-视杆营养不良6;先天性肾上腺发育不良,X-连锁;先天性无巨核细胞性血小板减少症(Congenital amegakaryocytic血小板减少);先天性无虹膜;先天性输精管双侧缺乏(Congenital bilateral absence of the vas deferens);先天性白内障,听力损失,和神经变性;先天性挛缩性蜘蛛指综合征;先天性叶酸吸收缺陷;先天性糖基化病症1K型;先天性糖基化病症1M型;先天性糖基化病症1t型;先天性糖基化病症1u型;先天性糖基化病症2C型;先天性全身性脂肪营养不良1型;先天性全身性脂肪营养不良2型;先天性心脏缺陷,多种类型,1,X连锁;先天性乳糖酶缺陷;先天性长QT综合征;先天性肌营养不良症-肌营养不

良蛋白聚糖病变(dystroglycanopathy)伴有脑和眼异常,A2型;先天性肌营养不良症-肌营养不良蛋白聚糖病变伴有脑和眼异常,A7型;先天性肌营养不良症-肌营养不良蛋白聚糖病变伴有智力低下,B1型;先天性肌营养不良症-肌营养不良蛋白聚糖病变伴有智力低下,B2型;先天性肌病伴有纤维类型不相称;先天性肌强直,常染色体显性形式;先天性肌强直,常染色体隐性形式;先天性稳定性夜盲症,常染色体显性3;先天性稳定性夜盲症,1A型;先天性稳定性夜盲症,1F型;粪卟啉病;角膜营养不良,Fuchs内皮,8;角膜上皮营养障碍;球形角膜脆性,蓝巩膜和关节活动过度;Cornelia de Lange综合征1;Cornelia de Lange综合征4;皮质发育异常,复杂,伴有其他脑畸形3;可的松还原酶缺陷1;Cowden综合征2;颅骨外胚层发育不良1;颅面耳聋手综合征(Craniofacial deafness hand syndrome);颅骨关节病(Cranioosteoarthropathy);颅缝早闭;颅缝早闭3;颅缝早闭和齿异常;肌酸缺陷,X-连锁;Crigler Najjar综合征,1型;Crouzon综合征;隐眼综合征;隐睾,单侧或双侧;Cushing指关节粘连;Beare和Stevenson的Cutis Gyrata综合征;半胱胺甲硫胺酸尿症;囊性纤维化病;胱氨酸病,眼肾病性;细胞色素-c氧化酶缺陷;Danon病;耳聋,常染色体显性12;耳聋,常染色体显性20;耳聋,常染色体隐性1A;耳聋,常染色体隐性63;耳聋,常染色体隐性8;耳聋,常染色体隐性9;乙酰基-CoA乙酰基转移酶缺陷;alpha-甘露糖苷酶缺陷;亚铁氧化酶缺陷;甘油激酶缺陷;胍基乙酸甲基转移酶缺陷;羟基甲基戊二酰-CoA裂合酶缺陷;碘化物过氧化物酶缺陷;丙二酰-CoA脱羧酶缺陷;UDP葡萄糖-己糖-1-磷酸尿苷基转化酶缺陷;延迟的言语和语言发育;delta地中海贫血;Dent病1;Desbuquois综合征;Desmosterolosis;DFNA 2非综合征听力损伤;糖尿病2型;糖尿病,胰岛素依赖性,20;Digitorenocerebral综合征;扩张型心肌病1FF;扩张型心肌病1G;扩张型心肌病1S;扩张型心肌病1X;扩张型心肌病3B;类固醇生成障碍,由于细胞色素p450氧化还原酶缺陷;远端遗传性运动神经元病2B型;双行睫-淋巴水肿综合征;Drash综合征;杜氏肌营养不良;先天性角化不良常染色体显性;先天性角化不良X-连锁;先天性角化不良,常染色体显性,2;先天性角化不良,常染色体隐性,5;张力失常1;张力失常27;张力失常5,多巴响应类型;张力失常,多巴响应类型,伴有或不伴有高苯丙氨酸血症,常染色体隐性;早期婴儿癫痫性脑病13;早期婴儿癫痫性脑病2;早期婴儿癫痫性脑病8;早期婴儿癫痫性脑病9;早期肌阵挛型脑病;外胚层发育不良症-并趾综合征1;先天性缺指,外胚层发育不良症,和唇裂/腭综合征3;埃勒斯-当洛斯综合征,经典类型;埃勒斯-当洛斯综合征,羟赖氨酸缺陷;埃勒斯-当洛斯综合征,肌挛缩类型;埃勒斯-当洛斯综合征,4型;Eichsfeld型先天性肌营养不良症;椭圆形红细胞性贫血3;子宫内膜癌;终板乙酰胆碱酯酶缺陷;扩张前庭水管综合征;肠激酶缺乏症;单纯性大疱性表皮松解症,Koebner型;癫痫,夜间额叶,3型;癫痫,进行性肌阵挛1A (Unverricht和Lundborg);癫痫,进行性肌阵挛2b;癫痫性脑病,早期婴儿,1;癫痫性脑病,早期婴儿,24;癫痫性脑病,早期婴儿,28;癫痫性脑病,早期婴儿,31;骺软骨发育不全,miura型;发作性共济失调(Episodic ataxia)1型;发作性共济失调,6型;发作性疼痛综合征,家族性,3;红细胞增多,家族性,2;红细胞增多,家族性,3;红角皮病,伴有共济失调;渗出性玻璃体视网膜病变1;渗出性玻璃体视网膜病变5;法布里病;法布里病,心脏变体;因子v和因子viii,组合缺陷,2;家族性淀粉样蛋白肾病伴有荨麻疹和耳聋;家族性乳腺癌;家族性寒冷性荨麻疹;家族性热性癫痫发作8;家族性偏瘫型偏头痛3型;家族性肥大性心肌病1;家族性肥大性心肌病10;家族性肥大性心肌病11;家族性肥大性心肌病20;家族性肥大性心肌病23;家族性肥大性心肌病4;家族性肥大性

心肌病6;家族性发育不全,肾小球囊肿肾;家族性婴儿肌无力;家族性幼年痛风;家族性地中海热;家族性血小板病症伴有相关髓样恶性;家族性孔洞脑;家族性迟发性皮肤卟啉病;家族性内脏淀粉样变性,Ostertag型;范科尼贫血,互补组C;范科尼贫血,互补组F;范科尼贫血,互补组G;范科尼贫血,互补组J;范科尼贫血,互补组T;法伯脂肪肉芽肿病;致死性血红蛋白定量性状基因座1;致死性血红蛋白定量性状基因座6;纤维软骨增生;局限型癫痫伴有言语障碍,伴有或不伴有智力低下;局灶性节段性肾小球硬化症6;中央凹发育不良和早老白内障综合征;额鼻发育异常1;额鼻发育异常2;额颞叶痴呆;果糖二磷酸酶缺陷;延胡索酸酶缺陷;半乳糖神经酰胺beta-半乳糖苷酶缺陷;胆囊疾病4;Gamstorp-Wohlfart综合征;神经节苷脂唾液酸酶缺陷;神经节苷脂贮积病GM1 3型;加德纳综合征;GATA-1相关血小板减少伴有异常红系造血(dyserythropoiesis);戈谢病;戈谢病3C型;戈谢病,围产期致死性;戈谢病,1型;全身性癫痫伴有发热性惊厥+,1型;全身性癫痫伴有发热性惊厥+,2型;全身性癫痫伴有发热性惊厥+,9型;Gerstmann-Straussler-Scheinker综合征;Glanzmann血小板机能不全;青光眼1,开角,F;青光眼,先天性;全局发育延迟;糖皮质激素缺陷4;戊二酸尿症,1型;糖原贮积病IIIa;糖原贮积病IV,先天性神经肌肉;糖原贮积病IXb;心脏糖原贮积病,致死性先天性;糖原贮积病,II型;糖原贮积病,IV型;糖原贮积病,V型;糖原贮积病,VI型;糖基磷脂酰肌醇缺陷;灰色血小板综合征;Griscelli综合征2型;生长和智力低下,下颌面骨发育不全,小头畸形和腭裂;生长激素不敏感伴有免疫缺陷;血色素沉积症1型;血色素沉积症3型;溶血性贫血,由于己糖激酶缺陷;溶血性贫血,非球形,由于葡萄糖磷酸异构酶缺陷;含铁血黄素沉着症,系统性,由于无铜蓝蛋白血症(aceruloplasminemia);Hennekam淋巴管扩张-淋巴水肿综合征;遗传性肠病性肢端皮炎;遗传性血管性水肿1型;遗传性乳腺和卵巢癌综合征;遗传性癌症素因性综合征;遗传性弥漫性胃癌;遗传性弥散性脑白质病伴有球形体;遗传性因子II缺陷疾病;遗传性因子IX缺陷疾病;遗传性因子VIII缺陷疾病;遗传性因子XI缺陷疾病;遗传性果糖尿症;遗传性平滑肌瘤病和肾细胞癌;遗传性淋巴水肿I型;遗传性神经痛性肌萎缩;遗传性非息肉病性结肠直肠癌5性;遗传性非息肉病性结肠直肠新生物;遗传性胰腺炎;遗传性副神经节瘤-嗜铬细胞瘤综合征;遗传性热变性异形红细胞增多症(pyropoikilocytosis);遗传性感觉神经病1D型;遗传性铁粒幼细胞性贫血;异位,内脏,X-连锁;异位;Hirschsprung病神经节成神经细胞瘤;组织细胞性髓性网状细胞增多症;前脑无裂畸形11;前脑无裂畸形2;前脑无裂畸形3;前脑无裂畸形4;同型半胱氨酸血症,由于MTHFR缺陷;高胱氨酸尿,由于CBS缺陷;Hurler综合征;Hurthle细胞甲状腺癌;Hutchinson-Gilford综合征;高钙尿症,儿童,自身限制;高胆固醇血症;过度惊骇(Hyperekplexia)3;遗传性过度惊骇;高铁蛋白血症白内障综合征;高脂蛋白血症,I型;高脂蛋白血症,ID型;高赖氨酸血症;高鸟氨酸血症-高氨血症-高瓜氨酸尿综合征;高前胰岛素血症;器官间距过远,重度,伴有面中部隆起(midface prominence),近视,智力低下,和骨脆性;肥大性心肌病;低钙血症,常染色体显性1;低钙血症,常染色体显性1,伴有巴特综合征;软骨发育不良;低色素性小红血球贫血伴有铁超负荷;低血糖症伴有肝中糖原合成酶缺陷;低促性腺素性功能减退症13伴有或不伴有嗅觉丧失;少汗性X连锁外胚层发育不良症;低钾性周期性麻痹1;低镁血症1,肠;低镁血症5,肾脏,伴有眼累及;低镁血症,癫痫,和智力低下;低髓鞘化脑白质营养不良症7;低髓鞘化脑白质营养不良症8,伴有或不伴有少牙畸形和/或低促性腺素性功能减退症;低蛋白血症,分解代谢过度;甲状腺功能减退症,先天

性,非甲状腺肿,1;甲状腺功能减退症,先天性,非甲状腺肿,5;甲状腺功能减退症,先天性,非甲状腺肿,6;稀毛症6;稀毛症-淋巴水肿-毛细管扩张综合征;I细胞疾病;寻常鳞癣;特发性基底神经节钙化5;免疫缺陷12;免疫缺陷23;免疫缺陷24;免疫缺陷30;免疫缺陷31a;免疫缺陷31C;免疫缺陷伴有超IgM 1型;包涵体肌病2;婴儿脑-视网膜变性;婴儿GM1神经节苷脂贮积病;婴儿低磷酸酯酶症;婴儿眼震,X连锁;胰岛素抗性糖尿病和黑棘皮症;智力失能;中间槭糖浆尿疾病(Intermediate maple syrup urine disease)2型;侵入性肺炎球菌病,复发性孤立性,2;虹膜-角膜-小梁发生不全;脑中铁积累;Jackson-Weiss综合征;雅-克二氏病;Joubert综合征23;青少年GM>1<神经节苷脂贮积病;幼年性息肉病综合征;歌舞伎面谱综合征(Kabuki make-up syndrome);Kallmann综合征3;Kallmann综合征4;Kallmann综合征5;Kallmann综合征6;圆锥角膜1;Kohlschutter综合征;库格尔贝格-韦兰德病(Kugelberg-Welander disease);拉福拉病(Lafora disease);Langer肢中部发育不良综合征;Laron型孤立性生长激素缺陷;拉尔森综合征,显性类型;Lchad缺陷伴有母亲急性妊娠脂肪肝;利伯先天性黑矇13;利伯先天性黑矇4;利伯先天性黑矇9;利氏病;LEOPARD综合征;LEOPARD综合征1;LEOPARD综合征2;矮妖精貌综合征;Leri-Weill软骨骨生成障碍;莱施-奈恩综合征;脑白质营养不良症,低髓鞘化,6;脑白质病伴有共济失调;脑白质病伴有脑干和脊髓累及和乳酸盐升高;脑白质病伴有消除的白质;莱迪希细胞无生殖力;Li-Fraumeni综合征1;肢肌营养不良;肢肌营养不良,1B型;肢肌营养不良,1C;肢肌营养不良,1E型;肢肌营养不良,2A型;肢肌营养不良,2B型;肢肌营养不良型,2E型;肢肌营养不良,2F型;肢肌营养不良,2L型;肢肌营养不良-肌营养不良蛋白聚糖病,C1型;肢肌营养不良-肌营养不良蛋白聚糖病,C14型;肢肌营养不良-肌营养不良蛋白聚糖病,C2型;肢肌营养不良-肌营养不良蛋白聚糖病,C7型;无脑回1;长QT综合征1;长QT综合征13;长QT综合征15;长QT综合征2;长QT综合征9;长QT综合征,LQT1亚型;长链3-羟基乙酰基-CoA脱氢酶缺陷;Lowe综合征,;促黄体激素抗性,女性;淋巴细胞增生综合征1;淋巴细胞增生综合征1,X连锁;Lynch综合征I;Lynch综合征II;巨血小板减少症,家族性,Bernard-Soulier类型;黄斑营养不良伴有中心锥累及;Majeed综合征;恶性食管肿瘤;恶性前列腺肿瘤;下颌末端成骨不全;槭糖尿病;槭糖尿病1A型;槭糖尿病2型;马方综合征;Marie Unna遗传性稀毛症1;年轻人成年型糖尿病,2型;年轻人成年型糖尿病,3型;中链酰基辅酶A脱氢酶缺陷;Meier-Gorlin综合征5;梅-弗综合征;MEN2表型:未分类;MEN2表型:未知;Menkes头发纽结综合征;绝经,天然,年龄(age at),定量性状基因座3;智力低下30,X连锁;智力低下和小头畸形伴有脑桥和脑发育不全;智力低下,常染色体显性13;智力低下,常染色体显性16;智力低下,常染色体显性29;智力低下,常染色体显性38;智力低下,常染色体显性7;智力低下,常染色体隐性34;智力低下,常染色体隐性49;智力低下,刻板型运动,癫痫,和/或脑畸形;智力低下,综合征性,Claes-Jensen型,X连锁;智力低下,X连锁,综合征性13;智力低下,X连锁,综合征性32;智力低下,X连锁,综合征性,雷蒙(Raymond)型;智力低下,X连锁,综合征性,wu型;智力低下-低张性面容综合征X连锁,1;分层蛋白缺陷先天性肌营养不良症;异染性脑白质营养不良;干骺端软骨发育不全,施米德(Schmid)型;甲基钴胺素缺陷,cblg型;甲基丙二酸尿症,mut(0)型;小头畸形和脉络膜视网膜病变,常染色体隐性,2;小头畸形伴有或不伴有脉络膜视网膜病变,淋巴水肿或智力低下;小红细胞贫血;小阴茎;小眼综合征性3;小眼综合征性5;小眼,孤立性3;小眼,孤立性6;小眼,孤立性,伴有眼缺损7;糖尿病微血管并发症7;轻度非-PKU高

苯丙氨酸血症;线粒体复合物I缺陷;线粒体复合物II缺陷;线粒体复合物III缺陷;线粒体DNA减液综合征13(脑肌病类型);线粒体DNA减液综合征2;线粒体DNA减液综合征9(脑肌病性伴有甲基丙二酸尿症);线粒体短链烯酰-CoA水合酶1缺陷;线粒体三功能性蛋白缺陷;Miyoshi肌营养不良症1;Miyoshi肌营养不良症3;Mohr-Tranebjaerg综合征;镶嵌非整倍杂色体综合征(Mosaic variegated aneuploidy syndrome);Mowat-Wilson综合征;粘脂质累积III Gamma;粘多糖贮积症VI型;粘多糖贮积症,MPS-II;粘多糖贮积症,MPS-III-B;粘多糖贮积症,MPS-I-S;粘多糖贮积症,MPS-IV-A;粘多糖贮积症,MPS-IV-B;Muenke综合征;穆利布瑞侏儒综合征;多发性先天性异常;多发性内分泌瘤病,1型;多发性内分泌瘤病,2型;多发性内分泌瘤病,2a型;多发性骨骺发育不良1;多发性骨骺发育不良5;多发性外生骨疣2型;多发性翼状胬肉综合征Escobar型;多发性硫酸酯酶缺陷;残毁性角皮病;肌无力,肢带,家族性;肌无力综合征,先天性,9,与乙酰胆碱受体缺陷肌相关;无力综合征,先天性,9,与乙酰胆碱受体缺陷肌相关;肌无力综合征,先天性,伴有突触前和突触后缺陷;肌无力综合征,先天性,伴有管状聚集体2;肌无力综合征,缓慢通道先天性;肌阵挛癫痫肌病感觉性共济失调;肌阵挛,家族性皮质;肌纤维肌病1;肌纤维颤搐1;肌病伴有体位性肌萎缩,X连锁;肌病,肌动蛋白,先天性,伴有过量的细肌丝;肌病,中央核性;肌病,远端,1;肌病,孤立性线粒体,常染色体线性;肌病,还原体,X连锁,早发性,重度;先天性肌强直;甲病,非综合征性先天性,8;小眼4;发作性睡病7;美洲土著肌病(Native American myopathy);Navajo神经肝病变(neurohepatopathy);线形体肌病3;新生儿张力减退;新生儿胰岛素依赖性糖尿病;由柠檬素缺陷引起的新生儿肝内胆汁淤积;卵巢新生物;肾结石/骨质疏松,低磷酸盐血性,2;肾痨16;肾痨18;肾病综合征,10型;Neu-Laxova综合征1;神经变性伴有脑铁积累5;垂体神经部尿崩症;Nicolaides-Baraitser综合征;尼曼皮克病C1型;尼曼皮克病,A型;尼曼皮克病,B型;尼曼皮克病,c1型,幼年型形式;Nonaka肌病;非酮性高甘氨酸血症;努南综合征1;努南综合征5;努南综合征7;努南综合征8;未提供;未规定;眼皮肤白化病3型;眼咽肌营养不良症;Opsismodysplasia;视神经萎缩9;视神经萎缩和白内障,常染色体显性;视神经发育不全和中枢神经系统异常;口-面-指综合征;鸟氨酸氨基转移酶缺陷;鸟氨酸氨甲酰转移酶缺陷;口面裂11;口面指综合征6;乳清酸尿症;成骨不全12型;成骨不全13型;成骨不全III型;成骨不全伴有正常巩膜,显性形式;成骨不全,隐性围产期致死性;骨硬化症常染色体显性1型;骨硬化症常染色体隐性7;耳-腭-指综合征,I型;厚皮性骨膜病综合征;Pallister-Hall综合征;Papillon-Lef\xc3\xa8vre综合征;副神经节瘤1;副神经节瘤4;副甲状腺癌;顶骨孔2;帕金森病1;帕金森病7;帕金森病9;发作性睡眠性血经蛋白尿1;部分次黄嘌呤-鸟嘌呤磷酸核糖转移酶缺陷;脱皮综合征(Peeling skin syndrome),肢端类型;Pelger-Hu\xc3\xabt异常;佩利措伊斯-梅茨巴赫病;彭德莱综合征;永久性新生儿糖尿病;过氧化物酶体生物发生病症(Peroxisome biogenesis disorder)6B;过氧化物酶体生物发生病症9B;波伊茨-耶格综合征;斐弗综合征;苯丙酮尿;嗜铬细胞瘤;磷酸甘油酸酯激酶1缺陷;磷酸核糖焦磷酸合成酶超活性;光敏感毛发低硫营养不良;Pierson综合征;色素性苍白球变性;Pitt-Hopkins综合征;Pitt-Hopkins样综合征2;垂体依赖型皮质醇增多症;垂体激素缺陷,组合性1;垂体激素缺陷,组合性4;垂体激素缺陷,组合性5;血小板型出血病症16;多可凝集性红细胞综合征;结节性多动脉炎;多囊肾病,婴儿型;葡聚糖体肌病(Polyglucosan body myopathy)2;多小脑回,双侧额顶骨;多神经病,听力损失,共济失调,

色素性视网膜炎,和白内障;脑桥小脑发育不全,1B型;脑桥小脑发育不全,1c型;脑桥小脑发育不全,9型;Poretti-boltshauser综合征;轴前多指症2;过早染色单体分离性状;卵巢功能早衰5;卵巢功能早衰7;卵巢功能早衰9;原发性常染色体隐性小头畸形1;原发性常染色体隐性小头畸形2;原发性常染色体隐性小头畸形5;原发性常染色体隐性小头畸形6;原发性纤毛运动异常;原发性扩张性心肌病;原发性家族性肥大心肌病;原发性高草酸尿,I型;原发性高草酸尿,III型;原发性局限性皮肤淀粉样变性病1;原发性开角型青光眼幼发型1;原发性肺动脉高压;原发性肺动脉高压4;Primrose综合征;进行性骨化性肌炎;进行性硬化性灰质萎缩;增殖性血管病和积水性无脑畸形-脑积水综合征;备解素缺陷,X连锁;丙酸血症;假性胡尔勒多种营养不良;假性醛固酮减少症1型常染色体显性;假性醛固酮减少症2B型;假性醛固酮减少症,2型;假性甲状旁腺功能减退症1A型;弹性假黄色瘤;弹性假黄色瘤样病症伴有多发性凝固因子缺陷;与遗传性出血性毛细血管扩张症相关的肺动脉高压;肺纤维化和/或骨髓衰竭,端粒相关,2;骨发育障碍矮小症;吡多辛依赖性癫痫;丙酮酸脱氢酶E1-alpha缺陷;桡骨发育不全-血小板减少综合征;Raine综合征;Ras病变(Rasopathy);隐性营养不良性大疱性表皮松解;赖芬斯坦综合征;肾性肉毒碱转运缺陷;肾细胞癌,乳头状,1;肾发育不良;肾少尿2;肾小管性酸中毒,远端,伴有溶血性贫血;视网膜锥体营养不良3A;色素性视网膜炎;色素性视网膜炎10;色素性视网膜炎11;色素性视网膜炎14;色素性视网膜炎2;色素性视网膜炎25;色素性视网膜炎33;色素性视网膜炎35;色素性视网膜炎4;色素性视网膜炎43;色素性视网膜炎50;色素性视网膜炎56;色素性视网膜炎73;色素性视网膜炎74;视网膜母细胞瘤;Rett病症;Rett综合征,先天性变型;Rett综合征,zappella变型;杆状瘤素因综合征2;肢根点状软骨发育不良1型;Rienhoff综合征;Roberts-SC短肢综合征;Robinow综合征;RRM2B-相关线粒体疾病;Rubinstein-Taybi综合征;Saethre-Chotzen综合征;肩腓肌病,X连锁显性;Schindler病,1型;Schindler病,3型;施奈德结晶状角膜营养不良;Seckel综合征1;癫痫;选择齿发育不全(Selective tooth agenesis)1;Senior-Loken综合征8;感觉性共济失调神经病,构音困难,和眼肌瘫痪;SeSAME综合征;由于ADA缺陷引起的重度组合免疫缺陷;重度组合免疫缺陷伴有小头畸形,生长延迟和对电离辐射的敏感性;重度先天性嗜中性白血球减少症;重度先天性嗜中性白血球减少症4,常染色体隐性;婴儿期重度肌阵挛癫痫;重度X连锁肌小管肌病;短QT综合征;短QT综合征2;身材矮小症伴有非特异性骨骼异常;身材矮小症,耳道闭锁,下颌骨发育不全,骨骼异常;身材矮小症,特发性,常染色体;身材矮小症,特发性,X连锁;短肋骨胸廓发育不良(Short-Rib Thoracic Dysplasia)13伴有或不伴有多指;短肋骨胸廓发育不良14伴有多指;短肋骨胸廓发育不良3伴有或不伴有多指;斯普林泽综合征;Shprintzen-Goldberg综合征;Shwachman综合征;唾液酸贮积病,重度婴儿型;唾液酸沉积症,II型;Sick窦综合征2,常染色体显性;铁粒幼细胞性贫血伴有B细胞免疫缺陷,周期热和发育延迟;谷固醇血症;Sj\xc3\xb6gren-Larsson综合征;Smith-Lemli-Opitz综合征;Sorsby眼底营养不良;Sotos综合征1;Sotos综合征2;痉挛性共济失调Charlevoix-Saguenay型;痉挛性截瘫11,常染色体隐性;痉挛性截瘫30,常染色体隐性;痉挛性截瘫4,常染色体显性;痉挛性截瘫54,常染色体隐性;痉挛性截瘫6;痉挛性截瘫7;痉挛性截瘫8;精子发生失败8;球形细胞增多症4型;鞘脂活化蛋白1缺陷;鞘磷脂/胆固醇脂沉积;脊髓性肌萎缩,下肢显性2,常染色体显性;脊髓性肌萎缩,II型;脊髓小脑性共济失调14;脊髓小脑性共济失调21;脊髓小脑性共济失调35;

脊髓小脑性共济失调38;Spino小脑性共济失调,常染色体隐性12;脊椎肋骨发育不全2;脊椎干骺端发育不全伴有关节松弛;脊椎干骺端发育不全,pakistani型;先天性椎体骨骺结构不良;脊椎干骺端结构不良伴有锥-杆营养不良;鳞状细胞头和颈癌;斯塔加特病1;斯塔加特病3;Steel综合征;Stickler综合征1型;僵硬性皮肤综合征;Sting相关血管病,婴儿发作性;亚急性神经原戈谢病;琥珀酰-CoA乙酰乙酸转移酶缺陷;超氧化物歧化酶,升高的胞内;主动脉瓣上狭窄(Supravalvar aortic stenosis);指关节粘连-短指综合征;并指型9;丹吉尔病;跗腕联合综合征(Tarsal carpal coalition syndrome);泰-萨克斯病;泰-萨克斯病,B1变型;T-细胞幼淋巴细胞白血病;Temple-Baraitser综合征;Temtamy轴前短指综合征;法乐四联症;胸主动脉动脉瘤和主动脉夹层;血小板减少2;血小板减少,X连锁;血小板减少,X连锁,间歇性;由于活化的蛋白C抗性引起的血栓形成倾向;血栓形成倾向,遗传性,由于蛋白C缺陷,常染色体显性;血栓形成倾向,遗传性,由于蛋白C缺陷,常染色体隐性;甲状腺癌,非髓性,4;甲状腺内分泌障碍1;甲状腺毒性周期性麻痹;提策(Tietz)综合征;齿发育不全,选择性,3;齿发育不全,选择性,X连锁,1;暂时性新生儿糖尿病1;暂时性新生儿糖尿病2;Treacher collins综合征2;毛鼻指发育异常(Trichorhinophalangeal dysplasia)I型;三甘油脂贮存病伴鳞癣;磷酸丙糖异构酶缺陷;拇指三节指骨;结节性硬化症1;结节性硬化症2;结节性硬化症综合征;酪氨酸酶阴性眼皮肤白化病;酪氨酸酶阳性眼皮肤白化病;酪氨酸血症2型;乌尔里希(Ullrich)先天性肌营养不良症;未分类;Unverricht-Lundborg综合征;Upshaw-Schulman综合征;尿苷5'单磷酸水解酶缺陷,溶血性贫血;乌斯赫尔(Usher)综合征,1D型;乌斯赫尔综合征,1F型;乌斯赫尔综合征,2A型;范德沃德综合征(Vander Woude syndrome);多样性卟啉病;Vater与大头和巨脑室的关联;室间隔缺损3;维生素D依赖性佝偻病,1型;维生素D依赖性佝偻病,2型;维生素k依赖性凝固因子,组合缺陷,1;卵黄状变性;Von Hippel-Lindau综合征;血管性血友病(von Willebrand disease),2b型;瓦尔登布尔(Waardenburg)综合征1型;瓦尔登布尔综合征2E型,不伴有神经病学累及;瓦尔登布尔综合征4A型;瓦尔登布尔综合征4B型;瓦尔登布尔综合征4C型;Walker-Warburg先天性肌营养不良症;Warburg微综合征3;Warts,低丙球蛋白血症,感染和先天性骨髓粒细胞缺乏症;韦尼印-霍夫曼病;维尔纳综合征(Werner syndrome);Wieacker综合征;Wiedemann-Steiner综合征;Winchester综合征;Wolfram综合征2;干瘪细胞增多症;着色性干皮病,D组;着色性干皮病,G组;X连锁丙种球蛋白缺乏血症;X连锁遗传性运动和感觉神经病;X连锁鳞癣伴有固醇硫酸酯酶缺陷;X连锁智力低下41;X连锁智力低下90;X连锁室周异位;济-拉二氏综合征;或济-拉二氏综合征2。在一些实施方案中,靶DNA序列包含与疾病或病症相关的序列。在一些实施方案中,靶DNA序列包含与疾病或病症相关的点突变。在一些实施方案中,与疾病或病症相关的点突变在与疾病或病症相关的基因中。在一些实施方案中,与疾病或病症相关的基因选自下组:AARS2、AASS、ABCA1、ABCA4、ABCB11、ABCB6、ABCC6、ABCC8、ABCD1、ABCG8、ABHD12、ABHD5、ACADM、ACAT1、ACE、ACO2、ACTA1、ACTB、ACTG1、ACTN2、ACVR1、ACVRL1、ADA、ADAMTS13、ADAR、ADGRG1、ADSL、AFF4、AGA、AGBL1、AGL、AGPAT2、AGRN、AGXT、AIPL1、AKR1D1、ALAD、ALAS2、ALDH3A2、ALDH7A1、ALDOB、ALG1、ALPL、ALS2、ALX3、ALX4、AMPD2、AMT、ANKS6、ANO5、APC、APOA1、APOE、APP、APRT、AQP2、AR、ARHGEF9、ARID2、ARL6、ARSA、ARSB、ARSE、ARX、ASAH1、ASB10、ASPM、ATF6、ATL1、ATM、ATP13A2、ATP1A3、ATP6V1B2、ATP7A、ATR、ATRX、AVP、B2M、B3GALT6、BAAT、BARD1、BBS10、BBS12、BBS2、BBS4、BBS9、BCKDHA、BCKDHB、

BCS1L、BEST1、BHLHA9、BICD2、BLM、BMP1、BMP4、BMPR2、BRAF、BRCA1、BRCA2、BRIP1、BTD、BTK、C10orf2、C1GALT1C1、C5orf42、C9、CA1、CACNA1S、CALM2、CANT1、CAPN3、CASK、CASQ2、CASR、CAV3、CBS、CCBE1、CCDC39、CD40LG、CDC6、CDC73、CDH1、CDH23、CDKL5、CDKN2A、CDON、CECR1、CENPJ、CEP120、CEP83、CFP、CFTR、CHAT、CHCHD10、CHD7、CHRNA1、CHRNB2、CHRNG、CHST14、CHSY1、CLCN1、CLCN2、CLCN5、CLCNKA、CLDN16、CLDN19、CLIC2、CLN6、CLN8、CNGA3、CNNM2、CNTNAP2、COA5、COL11A1、COL1A1、COL1A2、COL27A1、COL2A1、COL3A1、COL4A1、COL4A5、COL5A1、COL5A2、COL6A1、COL6A3、COL7A1、COLQ、COMP、CP、CPOX、CPT1A、CPT2、CR2、CRADD、CREBBP、CRH、CRX、CRYAB、CSF1R、CSTB、CTH、CTLA4、CTNS、CTPS1、CTSC、CTSD、CTSF、CTSK、CUL3、CXCR4、CYBB、CYP1B1、CYP27A1、CYP27B1、CYP4F22、CYP4V2、CYP7B1、DARS2、DBT、DCLRE1C、DCX、DDHD2、DES、DGUOK、DHCR24、DHCR7、DKC1、DLG3、DLL4、DMD、DMP1、DNAH11、DNAH5、DNAJB6、DNAJC19、DNM1、DNM2、DNMT1、DOCK6、DOK7、DOLK、DPAGT1、DPM2、DSC2、DSP、DYNC1H1、DYNC2H1、DYRK1A、DYSF、ECEL1、ECHS1、EDA、EDN3、EEF1A2、EFHC1、EFTUD2、EGLN1、EHMT1、EIF2B5、ELN、ELOVL4、ELOVL5、EMP2、ENPP1、EOGT、ERCC2、ERCC8、ESCO2、ETFDH、EXOSC3、EXOSC8、EXT2、EYA1、EYS、F12、F2、F5、F8、F9、FAM20C、FANCA、FANCF、FANCG、FAS、FBLN5、FBN1、FBN2、FBP1、FBXL4、FCGR3B、FGF8、FGFR1、FGFR2、FGFR3、FH、FHL1、FKTN、FLCN、FLG、FLNA、FLNB、FLT4、FLVCR2、FOXC1、FOXE1、FOXG1、FOXL2、FRAS1、FRMD7、FTL、FUS、G6PC3、G6PD、GAA、GABRA1、GABRG2、GAD1、GALC、GALNS、GALT、GAMT、GARS、GATA1、GATA6、GBA、GBA2、GBE1、GCDH、GCH1、GCK、GDAP1、GDI1、GFAP、GGCX、GHR、GJA8、GJB1、GJB2、GK、GLB1、GLI3、GLRA1、GMPPB、GNAI3、GNAS、GNAT1、GNE、GNPTAB、GNPTG、GPI、GPIHBP1、GPT2、GRIA3、GRIN2A、GRIN2B、GRIP1、GRN、GSC、GUCY2D、GYG1、GYS2、H6PD、HADHB、HBB、HBD、HBG1、HBG2、HCN1、HCN4、HESX1、HEXA、HFE、HFM1、HGSNAT、HINT1、HK1、HMGCL、HNF1A、HNF1B、HOGA1、HOXA1、HPD、HPGD、HPRT1、HR、HSD17B10、HSPB1、IDS、IDUA、IFT122、IFT80、IGHMBP2、IKBKG、IL11RA、IL12RB1、IMPDH1、IMPG2、INF2、ING1、INPPL1、INSL3、INSR、IRF6、IRX5、ISPD、ITGA2B、ITGB3、ITK、JAGN1、KCNA1、KCNH1、KCNH2、KCNJ1、KCNJ10、KCNJ11、KCNJ18、KCNJ2、KCNJ5、KCNK3、KCNQ1、KCNQ2、KCNQ4、KDM5C、KIAA0196、KIAA0586、KIF11、KIF1A、KIF2A、KISS1、KISS1R、KLF1、KMT2A、KMT2D、KRAS、KRIT1、KRT1、KRT5、KRT6A、LAMA1、LAMA2、LAMB2、LAMB3、LAMP2、LBR、LCT、LDLR、LIPA、LITAF、LMBR1、LMNA、LPIN2、LPL、LRIT3、LRP5、LRRC6、LRTOMT、LYST、LYZ、MAD1L1、MAF、MALT1、MAN2B1、MAPK1、MASTL、MATN3、MC2R、MCCC1、MCCC2、MCFD2、MCM8、MCOLN1、MCPH1、MECP2、MEF2C、MEFV、MEN1、MESP2、MET、MFN2、MFSD8、MGAT2、MITF、MKKS、MLH1、MLYCD、MMACHC、MMP14、MOG、MPL、MPV17、MPZ、MRE11A、MRPL3、MSH2、MSH6、MSR1、MSX1、MT-ATP6、MTHFR、MTM1、MT-ND1、MTR、MUSK、MUT、MYBPC3、MYC、MYH7、MYL2、MYL3、MYO1E、MYOC、NAGA、NAGLU、NARS2、NBEAL2、NBN、NDP、NDUFA1、NDUFA13、NDUFAF3、NDUFS8、NEFL、NEU1、NEXN、NFIX、NHEJ1、NHLRC1、NIPA1、NIPBL、NKX2-5、NLRP3、NMNAT1、NNT、NOBOX、NOG、NOL3、NOTCH3、NPC1、NPR2、NROB1、NR3C2、NR5A1、NRXN1、NSD1、NSDHL、NT5C3A、NYX、OAT、OCA2、OCRL、OFD1、OPA3、OPCML、OSMR、OTC、OTOF、OTX2、OXCT1、PAFAH1B1、PAH、PAK3、PALB2、PANK2、PAPSS2、PARK7、PAX2、PAX3、PAX6、PAX9、PCCA、PCCB、PCDH15、PCDH19、PCYT1A、PDE4D、PDE6A、PDE6B、PDE6C、PDE6H、PDGFB、PDHA1、PET100、PEX10、PEX7、PGK1、PGM1、PGM3、PHGDH、PHKB、PHOX2B、PIEZO1、PIGM、PITPNM3、PITX2、PKHD1、PKP2、PLA2G6、PLK4、PLOD1、PLP1、PMM2、PMP22、PMS2、PNPLA6、POLG、POLG2、POLR1A、POLR1D、POLR3A、POLR3B、POMT1、POMT2、POR、POU1F1、PPOX、PPT1、PRKACG、

PRKAG2、PRKAR1A、PRKCG、PRNP、PROC、PROK2、PROKR2、PRPF31、PRPS1、PRSS56、PSAP、PSEN1、PTEN、PTPN11、PURA、PVRL4、PYGL、PYGM、RAB18、RAB27A、RAB7A、RAD21、RAD51C、RAF1、RAG2、RAX、RAX2、RB1、RBM8A、RDH12、RET、RHO、RIT1、RNF216、ROGDI、RP2、RPGR、RPS6KA3、RRM2B、RSPO4、RUNX1、RUNX2、RYR1、RYR2、SACS、SAMHD1、SBDS、SCN11A、SCN1A、SCN2A、SCN5A、SCN8A、SCNN1B、SDHAF1、SDHB、SDHD、SEMA4A、SEPN1、SERPINF1、SERPING1、SETBP1、SGCB、SGCD、SH2D1A、SH3TC2、SHANK3、SHH、SHOX、SIGMAR1、SIX3、SKI、SLC11A2、SLC17A5、SLC19A3、SLC1A3、SLC22A5、SLC25A13、SLC25A15、SLC25A19、SLC25A22、SLC25A38、SLC25A4、SLC26A4、SLC2A10、SLC2A9、SLC33A1、SLC35C1、SLC39A4、SLC46A1、SLC4A1、SLC52A2、SLC52A3、SLC5A5、SLC6A5、SLC6A8、SLC9A3R1、SMAD2、SMAD4、SMARCA2、SMARCA4、SMN1、SMPD1、SNCA、SNRNP200、SNRPB、SOD1、SOD3、SOX9、SPAST、SPATA5、SPG11、SPG7、SPTB、SRD5A2、SRY、STAC3、STAR、STAT1、STAT3、STAT5B、STK11、STS、STX1B、STXBP1、SUCLG1、SUMF1、TARDBP、TAZ、TBC1D24、TBX1、TBX20、TCF12、TCF4、TECTA、TERC、TERT、TFAP2B、TFR2、TGFB3、TGFBI、TGFBR2、TGIF1、TGM1、TGM5、TGM6、THRA、THRB、TIMM8A、TK2、TMEM173、TMEM240、TMEM98、TMPRSS15、TMPRSS3、TMPRSS6、TNFRSF11A、TNNI3、TNNT1、TOR1A、TP53、TP63、TPI1、TPM1、TPM2、TPM3、TPO、TPP1、TRIM37、TRNT1、TRPM6、TRPS1、TSC1、TSC2、TSHR、TSPAN12、TTPA、TTR、TUBB4A、TULP1、TYMP、TYR、TYRP1、UBE2T、UBE3A、UBIAD1、UMOD、UMPS、UROD、USH2A、USP8、VDR、VHL、VPS13B、VPS33B、VWF、WAS、WDR19、WDR45、WDR62、WDR72、WFS1、WNK4、WNT5A、WRN、WT1、WWOX、ZBTB20、ZC4H2、ZDHHC9、ZEB2、ZFP57、ZIC3、或ZNF469。

[0435] 一些实施方案提供了用于使用本文提供的DNA编辑融合蛋白的方法。在一些实施方案中，融合蛋白用于通过使靶核碱基(例如C残基)脱氨基而将点突变引入到核酸中。在一些实施方案中，使用融合蛋白使靶C脱氨基为U，其然后被除去以创建先前由C残基占据的无碱基位点。在一些实施方案中，靶核碱基的脱氨基导致遗传缺陷的校正，例如在校正导致基因产物中功能丧失的点突变中。在一些实施方案中，本文提供的方法用于将失活性点突变引入到编码与疾病或病症相关的基因产物的基因或等位基因中。例如，在一些实施方案中，本文提供了采用DNA编辑融合蛋白将失活性点突变引入到癌基因中的方法(例如，在增殖性疾病的治疗中)。在一些实施方案中，失活性突变可以在编码序列中产生提前终止密码子，其导致截短的基因产物(例如缺乏全长蛋白质功能的截短的蛋白质)的表达。

[0436] 在一些实施方案中，本文中提供的方法的目的是经由基因组编辑恢复功能失调基因的功能。可以验证本文提供的核碱基编辑蛋白以用于体外基于基因编辑的人治疗学，例如通过校正人细胞培养物中的疾病相关的突变。本领域技术人员将理解，本文提供的核碱基编辑蛋白，例如包含核酸可编程DNA结合蛋白(例如，Cas9)和胞苷脱氨酶域和尿嘧啶结合蛋白的融合蛋白可以用于校正任何单点C至G或G至C突变。在第一种情况下，突变体C至U的脱氨基和随后U的切割校正突变，而在后一种情况下，C至U的脱氨基以及随后与突变体G碱基配对的U的切割及随后的一轮复制校正突变。

[0437] 疾病相关基因和等位基因中的点突变的成功校正为治疗学和基础研究中的应用开辟了基因校正的新策略。像所公开的包含核酸可编程DNA结合蛋白(napDNAbp)、胞苷脱氨酶和尿嘧啶结合蛋白的融合蛋白一样，位点特异性单碱基修饰系统也应用于“反向”基因治疗，其中某些基因功能得以有目的地抑制或消除。在这些情况下，导致蛋白质中突变性失活的位点特异性突变残基或抑制蛋白质的功能的突变可以用于在体外、离体或体内消除或抑

制蛋白质功能。

[0438] 本公开提供了用于治疗诊断患有与点突变相关或由点突变引起的疾病的受试者的方法,所述点突变可以通过本文提供的DNA编辑融合蛋白进行校正。例如,在一些实施方案中,提供的方法包括向患有此类疾病(例如,与如上所述的点突变相关的癌症)的受试者施用有效量的校正点突变(例如C至G或G至C点突变)或将失活性突变引入到疾病相关基因中的碱基编辑器融合蛋白。在一些实施方案中,疾病是增殖性疾病。在一些实施方案中,疾病是遗传疾病。在一些实施方案中,疾病是新生性疾病。在一些实施方案中,疾病是代谢性疾病。在一些实施方案中,疾病是溶酶体贮积病。可以通过校正点突变或将失活性突变引入到疾病相关基因中来治疗的其他疾病对于本领域技术人员来说是已知的,并且本公开内容在这方面不受限制。

[0439] 本公开内容提供了包含致病性G至C或C至G突变的基因的列表。可以使用本文提供的方法和组合物校正此类致病性的G至C或C至G突变,例如通过将C突变为G和/或将G突变为C,从而恢复基因功能。

[0440] 在一些实施方案中,融合蛋白识别规范的PAM,因此可以校正在侧翼序列中分别具有规范PAM(例如NGG)的致病性G至C或C至G的突变。例如,识别规范PAM的Cas9蛋白包含与如SEQ ID NO:6提供的化脓链球菌Cas9的氨基酸序列或与其包含SEQ ID NO:6的RuvC和HNH域的片段至少80%,85%,90%,95%,97%,98%或99%相同的氨基酸序列。

[0441] 对于本领域技术人员而言显而易见的是,为了将包含napDNAbp(例如Cas9域)的本文提供的任何融合蛋白靶向到靶位点,例如包含要编辑的点突变的位点,通常必需将融合蛋白与引导RNA(例如sgRNA)一起共表达。如本文其他地方更详细地解释的,引导RNA通常包含允许Cas9结合的tracrRNA框架和引导序列,其对Cas9:核酸编辑酶/域融合蛋白序列赋予特异性。在一些实施方案中,引导RNA包含结构5'-[指导序列]-guuuuagagcuagaaauagcaaguuaaaauaaaggcuaguccguuaucaacuugaaaaaguggcaccgagucggugcuuuuu-3'(SEQ ID NO:119),其中引导序列包含与靶序列互补的序列。在一些实施方案中,引导序列包含与靶核酸互补的核酸序列。引导序列通常为20个核苷酸长。基于本公开内容,用于将Cas9:核酸编辑酶/域融合蛋白靶向到特定基因组靶位点的合适的引导RNA的序列对于本领域技术人员将是显而易见的。此类合适的引导RNA序列通常包含与待编辑的靶核苷酸上游或下游50个核苷酸内的核酸序列互补的引导序列。

[0442] 碱基编辑器效率

[0443] 本公开的一些方面基于认识到本文提供的任何碱基编辑器能够修饰特定的核苷酸碱基而不产生显著比例的插入/缺失。如本文所用,"插入/缺失"是指核酸内的核苷酸碱基的插入或缺失。此类插入或缺失可以导致基因编码区内的移码突变。在一些实施方案中,期望产生有效修饰(例如突变或脱氨基)核酸内的特定核苷酸而不在核酸中产生大量插入或缺失(即插入/缺失)的碱基编辑器。在某些实施方案中,本文提供的任何碱基编辑器能够相对于插入/缺失产生更大比例的意图修饰(例如,点突变或脱氨基)。在一些实施方案中,本文提供的碱基编辑器能够生成大于1:1的意图点突变与插入/缺失的比率。在一些实施方案中,本文提供的碱基编辑器能够产生意图点突变与插入/缺失的比率,其为至少1.5:1、至少2:1、至少2.5:1、至少3:1、至少3.5:1、至少4:1、至少4.5:1、至少5:1、至少5.5:1、至少6:1、至少6.5:1、至少7:1、至少7.5:1、至少8:1、至少10:1、至少12:1、至少15:1、至少20:1、至

少25:1、至少30:1、至少40:1、至少50:1、至少100:1、至少200:1、至少300:1、至少400:1、至少500:1、至少600:1、至少700:1、至少800:1、至少900:1或至少1000:1或更多。可以使用任何合适的方法，例如以下实施例中使用的方法来测定意图的突变和插入/缺失的数目。在一些实施方案中，为了计算插入/缺失频率，扫描测序读段以与两个10-bp序列精确匹配，所述两个10-bp序列在可能发生插入/缺失的窗口的两侧侧翼。若未定位精确匹配，则从分析中排除读段。若该插入/缺失窗口的长度与参照序列精确匹配，则读段分类为不含有插入/缺失。若插入/缺失窗口比参照序列长或短两个或更多个碱基，则测序读段分别分类为插入或缺失。

[0444] 在一些实施方案中，本文提供的碱基编辑器能够限制核酸的区域中的插入/缺失的形成。在一些实施方案中，区域在由碱基编辑器靶向的核苷酸处或由碱基编辑器靶向的核苷酸的2、3、4、5、6、7、8、9或10个核苷酸内的区域。在一些实施方案中，本文提供的任何碱基编辑器能够将核酸的区域处的插入/缺失的形成限制至小于1％、小于1.5％、小于2％、小于2.5％、小于3％、小于3.5％、小于4％、小于4.5％、小于5％、小于6％、小于7％、小于8％、小于9％、小于10％、小于12％、小于15％或小于20％。在核酸区域处形成的缺失/缺失的数目可以取决于核酸(例如细胞的基因组内的核酸)暴露于碱基编辑器的时间的量。在一些实施方案中，在将核酸(例如，细胞的基因组内的核酸)暴露于碱基编辑器的至少1小时、至少2小时、至少6小时、至少12小时、至少24小时、至少36小时、至少48小时、至少3天、至少4天、至少5天、至少7天、至少10天或至少14天后测定插入/缺失的数目或比例。

[0445] 本公开内容的一些方面基于认识到本文提供的任何碱基编辑器能够有效地在核酸(例如受试者的基因组内的核酸)中产生意图的突变(如点突变)而不产生大量的非意图突变(如非意图的点突变)。在一些实施方案中，意图的突变是由与gRNA结合的特定碱基编辑器产生的突变，所述碱基编辑器特别设计为产生意图的突变。在一些实施方案中，意图的突变是与疾病或病症相关的突变。在一些实施方案中，意图的突变是与疾病或病症相关的胞嘧啶(C)至鸟嘌呤(G)点突变。在一些实施方案中，意图的突变是与疾病或病症相关的鸟嘌呤(G)至胞嘧啶(C)点突变。在一些实施方案中，意图的突变是基因的编码区内的胞嘧啶(C)至鸟嘌呤(G)点突变。在一些实施方案中，意图的突变是基因的编码区内的鸟嘌呤(G)至胞嘧啶(C)点突变。在一些实施方案中，意图的突变是产生终止密码子，例如基因的编码区内的提前终止密码子的点突变。在一些实施方案中，意图的突变是消除终止密码子的突变。在一些实施方案中，意图的突变是改变基因的剪接的突变。在一些实施方案中，意图的突变是改变基因的调控序列(例如，基因启动子或基因阻抑物)的突变。在一些实施方案中，本文提供的任何碱基编辑器能够产生大于1:1的意图突变与非意图突变(例如，意图的点突变:非意图的点突变)的比率。在一些实施方案中，本文提供的任何碱基编辑器能够产生意图突变与非意图突变的比率(例如，意图的点突变:非意图的点突变)，其为至少1.5:1、至少2:1、至少2.5:1、至少3:1、至少3.5:1、至少4:1、至少4.5:1、至少5:1、至少5.5:1、至少6:1、至少6.5:1、至少7:1、至少7.5:1、至少8:1、至少10:1、至少12:1、至少15:1、至少20:1、至少25:1、至少30:1、至少40:1、至少50:1、至少100:1、至少150:1、至少200:1、至少250:1、至少500:1或至少1000:1或更多。应当理解，本文“碱基编辑器效率”部分中描述的碱基编辑器的特性可以应用于任何融合蛋白或使用本文提供的融合蛋白的方法。

[0446] 用于编辑核酸的方法

[0447] 本公开的一些方面提供了用于编辑核酸的方法。在一些实施方案中，该方法是用于编辑核酸的核碱基(例如，双链DNA序列的碱基对)的方法。在一些实施方案中，方法包括以下步骤：a)使核酸(例如双链DNA序列)的靶区域与包含碱基编辑器(例如与胞苷脱氨酶或尿嘧啶结合蛋白融合的Cas9域)和引导核酸(例如gRNA)的复合物接触，其中所述靶区域包含靶定的核碱基对，b)诱导所述靶区域的链分离，c)将靶区域的单链中的所述靶核碱基对的第一核碱基转换成第二核碱基，以及d)切除第二核碱基，从而创建无碱基位点，并且e)用作为胞嘧啶(C)的第四核碱基替换与第一核碱基碱基互补的第三核碱基。在一些实施方案中，该方法导致核酸中小于20％的插入/缺失形成。应当理解的是，在一些实施方案中，省略步骤b。在一些实施方案中，第一核碱基是胞嘧啶(C)。在一些实施方案中，第二核碱基是脱氨基的胞嘧啶或尿嘧啶。在一些实施方案中，第三核碱基是鸟嘌呤(G)。在一些实施方案中，第四核碱基是胞嘧啶(C)。在一些实施方案中，将第五核碱基连接到步骤(d)中产生的无碱基位点中。在一些实施方案中，第五核碱基是鸟嘌呤(G)。在一些实施方案中，方法导致小于19％、18％、16％、14％、12％、10％、8％、6％、4％、2％、1％、0.5％、0.2％或小于0.1％插入/缺失形成。在一些实施方案中，编辑至少5％的意图碱基对。在一些实施方案中，编辑至少10％、15％、20％、25％、30％、35％、40％、45％或50％的意图碱基对。

[0448] 在一些实施方案中，靶核苷酸中意图产物与非意图产物的比率为至少2:1、5:1、10:1、20:1、30:1、40:1、50:1、60:1、70:1、80:1、90:1、100:1或200:1或更多。在一些实施方案中，意图点突变与插入/缺失形成的比率大于1:1、10:1、50:1、100:1、500:1、或1000:1或更多。在一些实施方案中，切割单链(切口链)与引导核酸杂交。在一些实施方案中，切割单链与包含第一核碱基的链相对。在一些实施方案中，碱基编辑器包含Cas9域。在一些实施方案中，碱基编辑器包含切口酶活性。在一些实施方案中，意图编辑的碱基对在PAM位点的上游。在一些实施方案中，意图编辑的碱基对在PAM位点的上游1、2、3、4、5、6、7、8、9、10、11、12、13、14、15、16、17、18、19或20个核苷酸。在一些实施方案中，意图编辑的碱基对在PAM位点的下游。在一些实施方案中，意图编辑的碱基对在PAM位点的下游1、2、3、4、5、6、7、8、9、10、11、12、13、14、15、16、17、18、19或20个核苷酸。在一些实施方案中，方法不需要规范(例如，NGG)PAM位点。在一些实施方案中，核碱基编辑器包含接头。在一些实施方案中，接头的长度为1-25个氨基酸。在一些实施方案中，接头的长度为5-20个氨基酸。在一些实施方案中，接头的长度为10、11、12、13、14、15、16、17、18、19或20个氨基酸。在一些实施方案中，靶区域包含靶窗口，其中靶窗口包含靶核碱基对。在一些实施方案中，靶窗口包含1-10个核苷酸。在一些实施方案中，靶窗口的长度为1-9、1-8、1-7、1-6、1-5、1-4、1-3、1-2或1个核苷酸。在一些实施方案中，靶窗口的长度为1、2、3、4、5、6、7、8、9、10、11、12、13、14、15、16、17、18、19或20个核苷酸。在一些实施方案中，意图编辑的碱基对在靶窗口内。在一些实施方案中，靶窗口包含意图编辑的碱基对。在一些实施方案中，使用本文提供的任何碱基编辑器进行方法。在一些实施方案中，靶窗口是脱氨基窗。

[0449] 在一些实施方案中，本公开提供了用于编辑核苷酸的方法。在一些实施方案中，本公开提供了用于编辑双链DNA序列的核碱基对的方法。在一些实施方案中，方法包括a)使双链DNA序列的靶区域与包含碱基编辑器和引导核酸(例如gRNA)的复合物接触，其中靶区域包含靶核碱基对，b)诱导所述靶区域的链分离，c)将靶区域的单链中的所述靶核碱基对的第一核碱基转换为第二核碱基，d)切除第二核碱基，从而创建无碱基位点，并且e)用作为胞

嘧啶(C)的第四核碱基替换与第一核碱基碱基互补的第三核碱基，由此产生意图编辑的碱基对，其中产生意图编辑的碱基对的效率是至少5％。应当理解的是，在一些实施方案中，省略步骤b。在一些实施方案中，编辑至少5％的意图碱基对。在一些实施方案中，编辑至少10％、15％、20％、25％、30％、35％、40％、45％或50％的意图碱基对。在一些实施方案中，方法引起小于19％、18％、16％、14％、12％、10％、8％、6％、4％、2％、1％、0.5％、0.2％或小于0.1％的插入/缺失形成。在一些实施方案中，靶核苷酸处的意图产物与非意图产物的比率为至少2:1、5:1、10:1、20:1、30:1、40:1、50:1、60:1、70:1、80:1、90:1、100:1或200:1或更多。在一些实施方案中，意图点突变与插入/缺失形成的比率大于1:1、10:1、50:1、100:1、500:1或1000:1或更多。在一些实施方案中，切割单链与引导核酸杂交。在一些实施方案中，核碱基编辑器包含切口酶活性。在一些实施方案中，意图编辑的碱基对在PAM位点的上游。在一些实施方案中，意图编辑的碱基对在PAM位点的上游1、2、3、4、5、6、7、8、9、10、11、12、13、14、15、16、17、18、19或20个核苷酸。在一些实施方案中，意图编辑的碱基对在PAM位点的下游。在一些实施方案中，意图编辑的碱基对在PAM位点的下游1、2、3、4、5、6、7、8、9、10、11、12、13、14、15、16、17、18、19或20个核苷酸。在一些实施方案中，该方法不需要规范(例如，NGG)PAM位点。在一些实施方案中，核碱基编辑器包含接头。在一些实施方案中，接头的长度为1-25个氨基酸。在一些实施方案中，接头的长度为5-20个氨基酸。在一些实施方案中，接头的长度为10、11、12、13、14、15、16、17、18、19或20个氨基酸。在一些实施方案中，靶区域包含靶窗口，其中靶窗口包含靶核碱基对。在一些实施方案中，靶窗口包含1-10个核苷酸。在一些实施方案中，靶窗口的长度为1-9、1-8、1-7、1-6、1-5、1-4、1-3、1-2或1个核苷酸。在一些实施方案中，靶窗口的长度为1、2、3、4、5、6、7、8、9、10、11、12、13、14、15、16、17、18、19或20个核苷酸。在一些实施方案中，意图编辑的碱基对发生在靶窗口内。在一些实施方案中，靶窗口包含意图编辑的碱基对。在一些实施方案中，核碱基编辑器是本文提供的碱基编辑器的任一个。

[0450] 药物组合物

[0451] 本公开的其他方面涉及药物组合物，其包含本文所述的碱基编辑器、融合蛋白或融合蛋白-gRNA复合物中的任一种。如本文所用，术语“药物组合物”是指配制用于药物用途的组合物。在一些实施方案中，药物组合物进一步包含药学上可接受的载体。在一些实施方案中，药物组合物包含另外的试剂(例如用于特异性递送，增加半衰期或其他治疗性化合物)。

[0452] 如本文所用，术语“药学上可接受的载体”是指药学上可接受的材料、组合物或媒介物，例如液体或固体填充剂、稀释剂、赋形剂、制造助剂(例如，润滑剂、滑石粉、硬脂酸镁、钙或锌或硬脂酸)或溶剂包封材料，涉及将化合物从身体的一个部位(例如，递送部位)运送或运输到另一个部位(例如，器官、组织或身体的一部分)。药学上可接受的载体是“可接受的”，意思是与制剂的其他成分相容并且对受试者的组织无害(例如，生理学相容的、无菌的、生理学的pH等)。可以充当药学上可接受的载体的材料的一些实例包括：(1)糖，例如乳糖、葡萄糖和蔗糖；(2)淀粉，如玉米淀粉和马铃薯淀粉；(3)纤维素及其衍生物，如羧甲基纤维素钠、甲基纤维素、乙基纤维素、微晶纤维素和醋酸纤维素；(4)粉末黄蓍胶；(5)麦芽；(6)明胶；(7)润滑剂，如硬脂酸镁、十二烷基硫酸钠和滑石粉；(8)赋形剂，如可可脂和栓剂蜡；(9)油，如花生油、棉籽油、红花油、芝麻油、橄榄油、玉米油和豆油；(10)二醇，如丙二醇；

(11)多元醇,如甘油、山梨糖醇、甘露醇和聚乙二醇(PEG);(12)酯类,如油酸乙酯和月桂酸乙酯;(13)琼脂;(14)缓冲剂,如氢氧化镁和氢氧化铝;(15)海藻酸;(16)无热原水;(17)等渗盐水;(18)林格氏液;(19)乙醇;(20)pH缓冲溶液;(21)聚酯,聚碳酸酯和/或聚酸酐;(22)增量剂(bulking agent),如多肽和氨基酸(23)血清成分,如血清白蛋白、HDL和LDL;(22)C2-C12醇,如乙醇;和(23)药物制剂中采用的其他无毒相容物质。润湿剂、着色剂、脱模剂、包衣剂、甜味剂、调味剂、芳香剂、防腐剂和抗氧化剂也可以存在于制剂中。诸如“赋形剂”、“载体”、“药学上可接受的载体”等术语在本文中可互换使用。

[0453] 在一些实施方案中,配制药物组合物用于递送至受试者,例如用于基因编辑。施用本文所述药物组合物的合适途径包括但不限于:局部、皮下、经皮、皮内、病灶内、关节内、腹膜内、膀胱内、经粘膜、牙龈、牙内、耳蜗内、经鼓室、器官内、硬膜外、鞘内、肌肉内、静脉内、血管内、骨内(intraosseus)、眼周、肿瘤内、脑内和脑室内施用。

[0454] 在一些实施方案中,将本文所述的药物组合物局部施用于患病部位(例如,肿瘤部位)。在一些实施方案中,本文所述的药物组合物通过注射、通过导管的方法、通过栓剂的方法或通过植入物的方法施用于受试者,所述植入物是多孔的、无孔的或凝胶状的材料,包括膜(例如硅橡胶膜(sialastic membrane))或纤维。

[0455] 在其他实施方案中,本文所述的药物组合物在控释系统中递送。在一个实施方案中,可以使用泵(参见例如,Langer,1990,Science 249:1527-1533;Sefton,1989,CRC Crit.Ref.Biomed.Eng.14:201;Buchwald et al.,1980,Surgery 88:507;Saudek et al.,1989,N.Engl.J.Med.321:574)。在另一个实施方案中,可以使用聚合物材料。(参见例如,Medical Applications of Controlled Release(Langer and Wise编,CRC Press,Boca Raton,Fla.,1974);Controlled Drug Bioavailability,Drug Product Design and Performance(Smolen and Ball编,Wiley,New York,1984);Ranger and Peppas,1983,Macromol.Sci.Rev.Macromol.Chem.23:61。还参见Levy et al.,1985,Science 228:190;During et al.,1989,Ann.Neurol.25:351;Howard et al.,1989,J.Neurosurg.71:105.)其他控释系统讨论于例如Langer,同上中。

[0456] 在一些实施方案中,药物组合物根据常规规程配制为适合于对受试者(例如人)进行静脉内或皮下施用的组合物。在一些实施方案中,用于通过注射施用的药物组合物是无菌等渗水性缓冲液中的溶液。必要时,药物还可以包括增溶剂和局部麻醉剂如利多卡因,以缓解注射部位的疼痛。通常,成分单独供应或以单位剂量形式混合在一起,例如,作为干燥的冻干粉末或无水浓缩物,在密封容器如安瓿或小药囊中,其表明活性剂的量。当药物通过输注施用时,可以用含有无菌药用级水或盐水的输液瓶分配。当药物组合物通过注射施用时,可以提供一安瓿的无菌注射用水或盐水,以便在施用前混合成分。

[0457] 用于全身施用的药物组合物可以是液体,例如无菌盐水、乳酸林格氏液或汉克氏液。此外,药物组合物可以是固体形式,并在使用前立即重新溶解或悬浮。还考虑了冻干形式。

[0458] 药物组合物可以包含在脂质颗粒或囊泡(例如脂质体或微晶)内,其也适用于肠胃外施用。颗粒可以是任何合适的结构的,例如单层或多层,只要其中含有组合物即可。化合物可以包埋在含有融合脂质二油酰磷脂酰乙醇胺(DOPE)、低水平(5-10mol%)的阳离子脂质的“稳定质粒-脂质颗粒”(SPLP)中,并通过聚乙二醇(PEG)涂层稳定(Zhang Y.P.et al.,

Gene Ther.1999,6:1438-47)。对于此类颗粒和囊泡,特别优选带正电荷的脂质如N-[1-(2,3-二油酰氧基)丙基]-N,N,N-三甲基-甲基硫酸铵或"DOTAP"。此类脂质颗粒的制备是众所周知的。参见例如,美国专利号4,880,635;4,906,477;4,911,928;4,917,951;4,920,016;和4,921,757;其每一个通过引用并入本文。

[0459] 例如,本文所述的药物组合物可以作为单位剂量施用或包装。当用于提及本公开的药物组合物时,术语"单位剂量"是指适合作为受试者的单一剂量的物理上离散的单位,每个单位含有预定量的活性物质,其经计算与所需稀释剂联合产生所需治疗效果;即载体(carrier)或媒介物(vehicle)。

[0460] 此外,药物组合物可以作为药物试剂盒提供,其包含(a)含有冻干形式的本发明的化合物(例如融合蛋白或碱基编辑器)的容器和(b)含有用于注射的药学上可接受的稀释剂(例如无菌水)的第二容器。药学上可接受的稀释剂可以用于重构或稀释本发明的冻干化合物。任选地与这种容器结合的可以是由管理药物或生物制品的制造、使用或销售的政府机构规定的形式的通知,该通知反映了制造、使用或销售机构对人类施用的批准。

[0461] 在另一方面,包括含有可用于治疗上述疾病的材料的制品。在一些实施方案中,制品包含容器和标签。合适的容器包括例如瓶子、小瓶、注射器和试管。容器可以由多种材料形成,例如玻璃或塑料。在一些实施方案中,容器容纳有效治疗本文所述疾病的组合物,并且可以具有无菌进入口。例如,容器可以是静脉内溶液袋或具有可由皮下注射针刺穿的塞子的小瓶。组合物中的活性剂是本发明的化合物。在一些实施方案中,容器上或与容器结合的标签表明组合物用于治疗所选择的疾病。制品可以进一步包含第二容器,其包含药学上可接受的缓冲液,例如磷酸盐缓冲盐水、林格氏溶液或右旋糖溶液。它可以进一步包括从商业和用户角度所需的其他材料,包括其他缓冲剂、稀释剂、过滤器、针头、注射器和具有使用说明的包装说明书。

[0462] 试剂盒、载体、细胞

[0463] 本公开内容的一些方面提供了包含核酸构建体的试剂盒,所述核酸构建体包含(a)编码如本文提供的任何融合蛋白的核苷酸序列;和(b)驱动(a)序列表达的异源启动子。在一些实施方案中,试剂盒进一步包含编码引导RNA主链的表达构建体,其中构建体包含克隆位点,其定位成允许将与靶序列相同或互补的核酸序列克隆到引导RNA主链中。

[0464] 本公开内容的一些方面提供了编码如本文提供的融合蛋白的napDNAbp(例如,Cas9蛋白)的多核苷酸。本公开内容的一些方面提供了包含此类多核苷酸的载体。在一些实施方案中,载体包含驱动多核苷酸表达的异源启动子。

[0465] 本公开内容的一些方面提供了细胞,其包含本文提供的任何融合蛋白,编码本文提供的任何融合蛋白的核酸分子,包含本文提供的任何融合蛋白和gRNA的复合物,和/或任何本文提供的载体。

[0466] 以上对报告物系统的示例性实施方案的描述仅出于说明目的而提供,并不意味着限制。本公开内容还包括其他报告物系统,例如上面详细描述的示例性系统的变型。

## 实施例

[0467] 经由无碱基位点生成和工程化特异性修复的胞嘧啶(C)至鸟嘌呤(G)的碱基编辑器

[0468] 在下面给出HEK2、RNF2和FANCF位点的测序数据。呈现的数据代表窗中编辑最多的C的碱基编辑值。对于HEK2，这是C6；对于RNF2，这是C6；对于FANCF，这是C6。碱基编辑之前和之后的三个不同位点的序列如下：HEK2:GAACACAAAGCATAGACTGC(SEQ ID NO:110)(测序读段CTTGTGTTTCGTATCTGACG(SEQ ID NO:111))；RNF2:GTCATCTTAGTCATTACCTG(SEQ ID NO:112)(测序读段CAGTAGAATCAGTAATGGAC(SEQ ID NO:113))；和FANCF:GGAATCCCTTCTGCAGCACC(SEQ ID NO:114)(测序读段相同)。对于HEK2和RNF2两者，对非靶链进行测序(此链含有与靶C互补的G)。对于FANCF，对靶链进行测序(此链含有靶C)。在图1和2中显示了C至T碱基编辑(例如，使用BE3，它是C至T碱基编辑器)和C至G碱基编辑的示意图。已知某些DNA聚合酶用G替换无碱基位点相对的碱基。实现C至G碱基编辑的一种策略是诱导无碱基位点的创建，然后募集或栓系此类聚合酶，以用C替换无碱基位点相对的G。若基于聚合酶的预先确定的碱基偏爱，可以将C和T切割并且用所有聚合酶修复，则这可以提供对所有编辑器的使用。

[0469] 下文汇总了不同融合构建体并且在表1中显示。UdgX是已知以最小的尿嘧啶切除活性紧密结合尿嘧啶的UDG同种型，UdgX*是UdgX的突变形式(Sang et al.NAR,2015)，其在Sang等人的体外测定法中观察到缺乏尿嘧啶切除活性。UdgX_On是UdgX的另一种突变形式(Sang et al.NAR,2015)，在Sang等人的相同体外测定法中观察到该突变形式具有增加的尿嘧啶切除活性。UDG是负责从DNA切除尿嘧啶以创建无碱基位点的酶。Rev7是已知掺入无碱基位点相对的C的Rev1/Rev3/Rev7复合物的组分。Rev1是上述复合物的酶促成分。聚合酶Alpha,Beta,Gamma,Delta,Epsilon,Gamma,Eta,Iota,Kappa,Lambda,Mu,和Nu是对于无碱基位点相对的碱基掺入具有不同偏爱的真核聚合酶。

[0470] 表1:构建体参考关键词

[0471]

| 构建体 | 定义 |
|---|---|
| BE3 | 发表的碱基编辑构建体 |
| BE3_UdgX | 用尿嘧啶结合蛋白,UdgX替换UGI |
| BE3_UdgX* | 用对尿嘧啶具有降低的结合亲和力的UdgX同种型替换UGI |
| BE3_REV7 | 用C整合跨损伤合成机器的组分替换UGI |
| BE2_UDG | 基于dCas9的构建体(无切口产生)，其中尿嘧啶脱糖基化酶(deglycosylase)替换UGI |
| BE3_UDG | 用尿嘧啶脱糖基化酶(BE3)替换UGI |
| BE2_UdgX_On | dCas9构建体，其中用具有增加尿嘧啶切除的激活突变的UdgX替换UGI |
| BE3_UdgX_On | 用具有增加尿嘧啶切割的激活突变的UdgX替换UGI |
| SMUG1 | 用SMUG1，一种ssDNA尿嘧啶脱糖基化酶替换UGI |

[0472] 实施例中使用的构建体：

[0473] BE3_全长-这是C至T碱基编辑器构建体,其包含胞苷脱氨酶、nCas9和尿嘧啶糖基化抑制剂(UGI)域。

[0474] MSSETGPVAVDPTLRRRIEPHEFEVFFDPRELRKETCLLYEINWGGRHSIWRHTSQNTNKHVEVNFIE
KFTTERYFCPNTRCSITWFLSWSPCGECSRAITEFLSRYPHVTLFIYIARLYHHADPRNRQGLRDLISSGVTIQIM
TEQESGYCWRNFVNYSPSNEAHWPRYPHLWVRLYVLELYCIILGLPPCLNILRRKQPQLTFFTIALQSCHYQRLPP
HILWATGLKSGSETPGTSESATPESDKKYSIGLAIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALL
FDSGETAEATRLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFFHRLEESFLVEEDKKHERHPIFGNIVDEVAY
HEKYPTIYHLRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFEENPINAS
GVDAKAILSARLSKSRRLENLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLA
QIGDQYADLFLAAKNLSDAILLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSK
NGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLRKQRTFDNGSIPHQIHLGELHAILRRQEDFYPF
LKDNREKIEKILTFRIPYYVGPLARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQSFIERMTNFDKNLPNEKVL
PKHSLLYEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVE
DRFNASLGTYHDLLKIIKDKDFLDNEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQLKRRRYTGWG
RLSRKLINGIRDKQSGKTILDFLKSDGFANRNFMQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKG
ILQTVKVVDELVKVMGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHPVENTQLQNEKLY
LYYLQNGRDMYVDQELDINRLSDYDVDHIVPQSFLKDDSIDNKVLTRSDKNRGKSDNVPSEEVVKKMKNYWRQLLN
AKLITQRKFDNLTKAERGGLSELDKAGFIKRQLVETRQITKHVAQILDSRMNTKYDENDKLIREVKVITLKSKLVS
DFRKDFQFYKVREINNYHHAHDAYLNAVVGTALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQEIGKATAKYFFYS
NIMNFFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEVQTGGFSKESILPKRNS
DKLIARKKDWDPKKYGGFDSPTVAYSVLVVAKVEKGKSKKLKSVKELLGITIMERSSFEKNPIDFLEAKGYKEVKK
DLIIKLPKYSLFELENGRKRMLASAGELQKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQHKHYLDE
IIEQISEFSKRVILADANLDKVLSAYNKHRDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVLD
ATLIHQSITGLYETRIDLSQLGGDSGGSTNLSDIIEKETGKQLVIQESILMLPEEVEEVIGNKPESDILVHTAYDE
STDENVMLLTSDAPEYKPWALVIQDSNGENKIKMLSGGSPKKKRKV(SEQ ID NO:115)

[0475] BE3_无UGI-此构建体是缺乏UGI域的上述BE3构建体。

[0476] MSSETGPVAVDPTLRRRIEPHEFEVFFDPRELRKETCLLYEINWGGRHSIWRHTSQNTNKHVEVNFIE
KFTTERYFCPNTRCSITWFLSWSPCGECSRAITEFLSRYPHVTLFIYIARLYHHADPRNRQGLRDLISSGVTIQIM
TEQESGYCWRNFVNYSPSNEAHWPRYPHLWVRLYVLELYCIILGLPPCLNILRRKQPQLTFFTIALQSCHYQRLPP
HILWATGLKSGSETPGTSESATPESDKKYSIGLAIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALL
FDSGETAEATRLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFFHRLEESFLVEEDKKHERHPIFGNIVDEVAY
HEKYPTIYHLRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFEENPINAS
GVDAKAILSARLSKSRRLENLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLA
QIGDQYADLFLAAKNLSDAILLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSK
NGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLRKQRTFDNGSIPHQIHLGELHAILRRQEDFYPF
LKDNREKIEKILTFRIPYYVGPLARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQSFIERMTNFDKNLPNEKVL
PKHSLLYEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVE
DRFNASLGTYHDLLKIIKDKDFLDNEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQLKRRRYTGWG
RLSRKLINGIRDKQSGKTILDFLKSDGFANRNFMQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKG

ILQTVKVVDELVKVMGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHPVENTQLQNEKLY
LYYLQNGRDMYVDQELDINRLSDYDVDHIVPQSFLKDDSIDNKVLTRSDKNRGKSDNVPSEEVVKKMKNYWRQLLN
AKLITQRKFDNLTKAERGGLSELDKAGFIKRQLVETRQITKHVAQILDSRMNTKYDENDKLIREVKVITLKSKLVS
DFRKDFQFYKVREINNYHHAHDAYLNAVVGTALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQEIGKATAKYFFYS
NIMNFFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEVQTGGFSKESILPKRNS
DKLIARKKDWDPKKYGGFDSPTVAYSVLVVAKVEKGKSKKLKSVKELLGITIMERSSFEKNPIDFLEAKGYKEVKK
DLIIKLPKYSLFELENGRKRMLASAGELQKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQHKHYLDE
IIEQISEFSKRVILADANLDKVLSAYNKHRDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVLD
ATLIHQSITGLYETRIDLSQLGGD(SEQ ID NO:116)

[0477] Cas9切口酶序列-用于BE3。

[0478] MDKKYSIGLAIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETAEATRLKRTA
RRRYTRRKNRICYLQEIFSNEMAKVDDSFFHRLEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVD
STDKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFEENPINASGVDAKAILSARLSKSR
RLENLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNL
SDAILLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEF
YKFIKPILEKMDGTEELLVKLNREDLLRKQRTFDNGSIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRI
PYYVGPLARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQSFIERMTNFDKNLPNEKVLPKHSLLYEYFTVYNEL
TKVKYVTEGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKI
IKDKDFLDNEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSG
KTILDFLKSDGFANRNFMQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKVVDELVKVMG
RHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQEL
DINRLSDYDVDHIVPQSFLKDDSIDNKVLTRSDKNRGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNLTKAE
RGGLSELDKAGFIKRQLVETRQITKHVAQILDSRMNTKYDENDKLIREVKVITLKSKLVSDFRKDFQFYKVREINN
YHHAHDAYLNAVVGTALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQEIGKATAKYFFYSNIMNFFKTEITLANGE
IRKRPLIETNGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEVQTGGFSKESILPKRNSDKLIARKKDWDPKKYG
GFDSPTVAYSVLVVAKVEKGKSKKLKSVKELLGITIMERSSFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFELEN
GRKRMLASAGELQKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQHKHYLDEIIEQISEFSKRVILAD
ANLDKVLSAYNKHRDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVLDATLIHQSITGLYETRI
DLSQLGGD(SEQ ID NO:21)

[0479] dCas9序列-用于BE2

[0480] MDKKYSIGLAIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETAEATRLKRTA
RRRYTRRKNRICYLQEIFSNEMAKVDDSFFHRLEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVD
STDKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFEENPINASGVDAKAILSARLSKSR
RLENLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNL
SDAILLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEF
YKFIKPILEKMDGTEELLVKLNREDLLRKQRTFDNGSIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRI
PYYVGPLARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQSFIERMTNFDKNLPNEKVLPKHSLLYEYFTVYNEL
TKVKYVTEGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKI
IKDKDFLDNEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSG

KTILDFLKSDGFANRNFMQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKVVDELVKVMG
RHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQEL
DINRLSDYDVDAIVPQSFLKDDSIDNKVLTRSDKNRGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNLTKAE
RGGLSELDKAGFIKRQLVETRQITKHVAQILDSRMNTKYDENDKLIREVKVITLKSKLVSDFRKDFQFYKVREINN
YHHAHDAYLNAVVGTALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQEIGKATAKYFFYSNIMNFFKTEITLANGE
IRKRPLIETNGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEVQTGGFSKESILPKRNSDKLIARKKDWDPKKYG
GFDSPTVAYSVLVVAKVEKGKSKKLKSVKELLGITIMERSSFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFELEN
GRKRMLASAGELQKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQHKHYLDEIIEQISEFSKRVILAD
ANLDKVLSAYNKHRDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVLDATLIHQSITGLYETRI
DLSQLGGD(SEQ ID NO:22)

[0481] BE3_用UDG,UdgX变体、聚合酶替换UGI-在下述构建体中,NLS通过加下划线鉴定,并且接头以斜体字鉴定。下述序列中指示的“[UGI]”鉴定插入UDG,UDG变体(例如UDG,UdgX*(R107S),和UdgX_On(H109S)),Rev7,和Smug1(而不是BE3的UGI)的位置。下述序列中指示的“[聚合酶]”鉴定插入聚合酶(例如Pol Beta,Pol Lambda,Pol Eta,Pol Mu,Pol Iota,Pol Kappa,Pol Alpha,Pol Delta,Pol Gamma和Pol Nu)和Rev1的位置。

[0482] MSSETGPVAVDPTLRRRIEPHEFEVFFDPRELRKETCLLYEINWGGRHSIWRHTSQNTN
KHVEVNFIEKFTTERYFCPNTRCSITWFLSWSPCGECSRAITEFLSRYPHVTLFIYIARLY
HHADPRNRQGLRDLISSGVTIQIMTEQESGYCWRNFVNYSPSNEAHWPRYPHLWVRL
YVLELYCIILGLPPCLNILRRKQPQLTFFTIALQSCHYQRLPPHILWATGLKSGSETPGTS
ESATPESDKKYSIGLAIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLF
DSGETAEATRLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFFHRLEESFLVEEDK
KHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKADLRLIYLALAHMIKFRGHFLI
EGDLNPDNSDVDKLFIQLVQTYNQLFEENPINASGVDAKAILSARLSKSRRLENLIAQL
PGEKKNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYA
DLFLAAKNLSDAILLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEK
YKEIFFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLRKQRTF
DNGSIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPYYVGPLARGNSRFAWM
TRKSEETITPWNFEEVVDKGASAQSFIERMTNFDKNLPNEKVLPKHSLLYEYFTVYNE
LTKVKYVTEGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEIS
GVEDRFNASLGTYHDLLKIIKDKDFLDNEENEDILEDIVLTLTLFEDREMIEERLKTYA
HLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFANRNFMQLI
HDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKVVDELVKVMGRH
KPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHPVENTQLQNEKLYL
YYLQNGRDMYVDQELDINRLSDYDVDHIVPQSFLKDDSIDNKVLTRSDKNRGKSDNV
PSEEVVKKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKAGFIKRQLVETRQIT
KHVAQILDSRMNTKYDENDKLIREVKVITLKSKLVSDFRKDFQFYKVREINNYHHAH
DAYLNAVVGTALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQEIGKATAKYFFYSNI
MNFFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEV
QTGGFSKESILPKRNSDKLQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVDHIVPQS
FLKDDSIDNKVLTRSDKNRGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNLTK
AERGGLSELDKAGFIKRQLVETRQITKHVAQILDSRMNTKYDENDKLIREVKVITLKSK
LVSDFRKDFQFYKVREINNYHHAHDAYLNAVVGTALIKKYPKLESEFVYGDYKVYDV
RKMIAKSEQEIGKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWDKGR
DFATVRKVLSMPQVNIVKKTEVQTGGFSKESILPKRNSDKLIARKKDWDPKKYGGFDS
PTVAYSVLVVAKVEKGKSKKLKSVKELLGITIMERSSFEKNPIDFLEAKGYKEVKKDLI
IKLPKYSLFELENGRKRMLASAGELQKGNELALPSKYVNFLYLASHYEKLKGSPEDNE
QKQLFVEQHKHYLDEIIEQISEFSKRVILADANLDKVLSAYNKHRDKPIREQAENIIHLF
TLTNLGAPAAFKYFDTTIDRKRYTSTKEVLDATLIHQSITGLYETRIDLSQLGGD*SGGS*(
SEQ ID NO: 117)[UGI]*SGGSGGSGGS*(SEQ ID NO: 120)[聚合酶]PKKKRKV (SEQ ID NO:
41)

[0483] N-端UDG(插入物UDG(Tyr147Ala)或UDG(Asn204Asp))+C端的Cas9切口酶和聚合酶-在下述构建体中,NLS序列通过加下划线鉴定,并且接头通过斜体字鉴定。下述序列中指示的“[UDG变体]”鉴定插入UDG Tyr147Ala和UDG Asn204Asp的位置。下述序列中指示的“[聚合酶]”鉴定插入聚合酶(例如Pol Beta,Pol Lambda,Pol Eta,Pol Mu,Pol Iota,Pol Kappa,Pol Alpha,Pol Delta,Pol Gamma和Pol Nu)和Rev1的位置。

[0484] [UDG变体]

[0485] SETPGTSESATPESDKKYSIGLAIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKN
LIGALLFDSGETAEATRLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFFHRLEESF
LVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKADLRLIYLALAHMIK
FRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFEENPINASGVDAKAILSARLSKSRRL
ENLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLA
QIGDQYADLFLAAKNLSDAILLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALV
RQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDL
LRKQRTFDNGSIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPYYVGPLARG
NSRFAWMTRKSEETITPWNFEEVVDKGASAQSFIERMTNFDKNLPNEKVLPKHSLLYE
YFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIE
CFDSVEISGVEDRFNASLGTYHDLLKIIKDKDFLDNEENEDILEDIVLTLTLFEDREMIEE
RLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFANR
NFMQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKVVDELVK
VMGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHPVENTQLQ
NEKLYLYYLQNGRDMYVDQELDINRLSDYDVDHIVPQSFLKDDSIDNKVLTRSDKNR
GKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKAGFIKRQL
VETRQITKHVAQILDSRMNTKYDENDKLIREVKVITLKSKLVSDFRKDFQFYKVREINN
YHHAHDAYLNAVVGTALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQEIGKATAKY
FFYSNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSMPQVNIV
KKTEVQTGGFSKESILPKRNSDKLIARKKDWDPKKYGGFDSPTVAYSVLVVAKVEKG
KSKKLKSVKELLGITIMERSSFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFELENGRKR
MLASAGELQKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQHKHYLDEII
EQISEFSKRVILADANLDKVLSAYNKHRDKPIREQAENIIHLFTLTNLGAPAAFKYFDTT
IDRKRYTSTKEVLDATLIHQSITGLYETRIDLSQLGGD*SGGS*(SEQ ID NO: 118)[聚合酶]PKKKRKV(SEQ ID NO: 41)

[0486] 实施例1:C至G方法1-增加无碱基位点形成

[0487] 若较有效地产生无碱基位点,则预期通过C至G碱基编辑途径的总通量将增加。在图3和4中显示了此方法中使用的碱基编辑器的示意图。使用UdgX(一种鉴定为以最小的尿嘧啶切除活性紧密结合尿嘧啶的UDG直向同源物)增加C至G编辑的量。不希望受任何特定理论的束缚,UdgX与U的近共价结合模拟促进跨损伤聚合酶型修复的损伤。此外,UdgX具有低水平的催化活性,其与紧密结合组合切割U并导致无碱基位点形成。无碱基位点形成允许脱靶产物,并且此损伤的优先产生导致更多产物。这通过不同的实验和碱基编辑器得到支持,它们在图5和6中显示。

[0488] 在图7至15中显示了使用七种碱基编辑器(BE3;BE3_UdgX;BE3_UdgX*;BE2_UdgX_On;BE3_UdgX_On;BE2_UDG;和BE3_UDG)在WT细胞中的HEK2、RNF2和FANCF位点处进行C至G碱基编辑的结果。这些图显示了在对来自标准C至T编辑的序列扰动具有高、中等和低容限的三个代表性位点处在最多编辑的位置(C6)处进行C至G编辑的结果。

[0489] 图16至24中显示了使用各种C至G碱基编辑器(BE3;BE3_UdgX;BE2_UNG;BE3_UNG;BE2UdgX_On;BE3UdgX_On;和SMUG1)在UDG-/-细胞中的HEK2、RNF2和FANCF位点处进行C至G碱基编辑的结果。

[0490] 图25至30中显示了使用各种C至G碱基编辑器(BE3;BE3_UdgX;BE2_UNG;BE3_UNG;BE2UdgX_On;BE3UdgX_On;和SMUG1)在REV1-/-细胞中的HEK2、RNF2和FANCF位点处进行C至G碱基编辑的结果。

[0491] 图31和32中汇总了使用各种C至G的碱基编辑器(BE3;BE3_UdgX;BE2_UNG;BE3_UNG;BE2UdgX_On;BE3UdgX_On;和SMUG1)在三种各自的细胞类型(WT、UDG-/-和REV1-/-细胞)中的HEK2、RNF2和FANCF位点处进行C至G碱基编辑的结果。

[0492] 实施例2:C至G的方法2-增加无碱基相对的C的掺入

[0493] 无碱基位点相对的C整合的偏爱增加应当导致总的C至G碱基编辑的增加。图33和34中显示了此方法的示意图以及此方法中使用的碱基编辑器。图35中显示了可用于C至G碱基编辑的此方法的各种聚合酶。简言之,无碱基位点产生导致C至非T产物形成。Rev1具有dC转移酶活性。消除此途径或改变无碱基损伤得到修复的方式应当导致新的碱基编辑器。若此途径仅仅造成此产物的形成,则Rev1-/-敲除细胞系应当缺乏C至G编辑。基于对无碱基位点相对的修复的聚合酶偏爱,各种聚合酶的融合物应当导致相对链的修复,从而导致C至G碱基的编辑增加。图36中显示了示例性碱基编辑器。

[0494] 在图37至39中显示了使用各种碱基编辑器(BE3;BE3_UdgX;BE2_UdgX_On;BE3_UdgX_On;BE2_UDG;和BE3_UDG)在WT细胞中的HEK2、RNF2和FANCF位点处进行C至G碱基编辑的结果。

[0495] 表2给出了通过人聚合酶η,ι,κ,和REV1进行的在无碱基位点和G相对的一碱基掺入的稳态动力学参数。参见Choi et al.J mol Bio.2010。

[0496] 表2:聚合酶η,ι,κ,和REV1的稳态动力学参数

[0497]

| 聚合酶 | 模板 | dNTP | $K_m$ (μM) | $k_{cat}$ ($s^{-1}$) | $k_{cat}/K_m$ ($mM^{-1}\,s^{-1}$) | dNTP选择性比率[a] | 相对效率[b] |
|---|---|---|---|---|---|---|---|
| η | AP位点 | A | 40 ± 6 | 0.12 ± 0.004 | 3.0 | 0.95 | 0.065 |
| | | T | 290 ± 50 | 0.92 ± 0.05 | 3.2 | 1 | 0.070 |
| | | G | 8.5 ± 1.0 | 0.005 ± 0.0001 | 0.59 | 0.19 | 0.013 |
| | | C | 210 ± 20 | 0.14 ± 0.01 | 0.67 | 0.21 | 0.015 |
| | G | C | 2.6 ± 0.1 | 0.12 ± 0.005 | 46 | | 1 |
| ι | AP位点 | A | 210 ± 40 | 0.54 ± 0.04 | 2.6 | 0.45 | 1.4 |
| | | T | 130 ± 20 | 0.74 ± 0.02 | 5.7 | 1 | 3.0 |
| | | G | 120 ± 10 | 0.47 ± 0.01 | 3.9 | 0.69 | 2.1 |
| | | C | 570 ± 140 | 0.77 ± 0.05 | 1.4 | 0.24 | 0.74 |
| | G | C | 300 ± 30 | 0.57 ± 0.02 | 1.9 | | 1 |
| κ | AP位点 | A | 1600 ± 200 | 0.077 ± 0.005 | 0.048 | 0.77 | 0.00065 |
| | | T | 2300 ± 700 | 0.017 ± 0.002 | 0.0074 | 0.12 | 0.00010 |
| | | G | 400 ± 70 | 0.0032 ± 0.0002 | 0.008 | 0.13 | 0.00011 |
| | | C | 780 ± 220 | 0.049 ± 0.005 | 0.063 | 1 | 0.00085 |
| | G | C | 3.8 ± 0.5 | 0.28 ± 0.01 | 74 | | 1 |
| REV1 | AP位点 | A | 140 ± 50 | 0.000025 ± 0.000002 | 0.00018 | 0.0031 | 0.00019 |
| | | T | 190 ± 30 | 0.000072 ± 0.000003 | 0.00038 | 0.0067 | 0.00040 |
| | | G | 190 ± 50 | 0.000031 ± 0.000003 | 0.00016 | 0.0029 | 0.00017 |
| | | C | 210 ± 30 | 0.012 ± 0.001 | 0.057 | 1 | 0.061 |
| | G | C | 12.8 ± 50 | 0.012 ± 0.0003 | 0.94 | | 1 |

[0498] [a]dNTP选择性比率,其通过用每次dNTP掺入的$k_{cat}/K_m$除以AP位点相对的dNTP掺入的最高$k_{cat}/K_m$计算。

[0499] [b]相对效率,其通过用AP位点相对的每次dNTP掺入的$k_{cat}/K_m$除以G相对的dCTP掺入的$k_{cat}/K_m$计算。

[0500] 表3中给出了通过人聚合酶α和δ/PCNA进行的在无碱基位点和G相对的一碱基掺入的稳态动力学参数。

[0501] 表3:聚合酶α和δ/PCNA的稳态动力学参数

[0502] 通过人聚合酶α和δ/PCNA进行的在AP位点和G相对的一碱基掺入的稳态动力学参数

[0503]

| 聚合酶 | 模板 | dNTP | $K_m$ (μM) | $k_{cat}$ ($s^{-1}$) | $k_{cat}/K_m$ ($mM^{-1}\ s^{-1}$) | dNTP选择性比率[a] | 相对效率[b] |
|---|---|---|---|---|---|---|---|
| α | AP位点 | A | 570 ± 100 | 0.0083 ± 0.0001 | 0.015 | 1 | 0.0010 |
| | | T | 250 ± 60 | 0.00046 ± 0.00003 | 0.0018 | 0.12 | 0.00012 |
| | | G | 550 ± 120 | 0.00024 ± 0.00002 | 0.0004 | 0.027 | 0.00003 |
| | | C | 980 ± 50 | 0.00047 ± 0.000001 | 0.0005 | 0.033 | 0.00003 |
| | G | C | 0.42 ± 0.09 | 0.0064 ± 0.0003 | 15 | | 1 |
| δ/PCNA | AP位点 | A | 25 ± 6 | 0.0067 ± 0.0004 | 0.27 | 1 | 0.012 |
| | | T | 62 ± 16 | 0.0060 ± 0.0004 | 0.097 | 0.36 | 0.0044 |
| | | G | 110 ± 20 | 0.010 ± 0.001 | 0.091 | 0.34 | 0.0041 |
| | | C | 880 ± 160 | 0.0069 ± 0.0006 | 0.0078 | 0.029 | 0.0004 |
| | G | C | 0.27 ± 0.05 | 0.0059 ± 0.0002 | 22 | | 1 |

[0504] [a]dNTP选择性比率,其通过用每次dNTP掺入的$k_{cat}/K_m$除以AP位点相对的dNTP掺入的最高$k_{cat}/K_m$计算。

[0505] [b]相对效率,其通过用AP位点相对的每次dNTP掺入的$k_{cat}/K_m$除以G相对的dCTP掺入的$k_{cat}/K_m$计算。

[0506] 表4-可以用于碱基编辑方法2的聚合酶

[0507]

| 聚合酶 | 大小(氨基酸) |
|---|---|
| 家族 **X** | |
| Beta | 335 |
| Lambda | 575 |
| Mu | 494 |
| 家族 **B** | |
| Alpha | 1462 |
| Delta | 1107 |
| Epsilon | 2286 |
| 家族 **Y** | |
| Eta | 713 |
| Iota | 740 |
| Kappa | 870 |
| Rev1 | 1251 |
| Zeta (Rev3/Rev7) | 3130 |

[0508] 实施例3:C至G的方法3-增加无碱基位点形成和C掺入两者

[0509] 在图40中显示了用于增加无碱基位点形成和C掺入两者以实现增加的C至G碱基编

辑的碱基编辑器的示意图。聚合酶栓系的构建体(特别是Pol Kappa)的添加增加C至G碱基编辑。图41中显示了对于Pol Iota栓系的构建体使用Pol Kappa在HEK2、RNF2和FANCF位点处进行碱基编辑的结果。图42到47中显示了在WT细胞中的HEK2、RNF2和FANCF位点处胞嘧啶残基处,使用其他聚合酶栓系的构建体进行碱基编辑的结果。UDG147是一种直接除去T并增加C至G碱基编辑的酶(图42到44),而UDG 204是一种直接除去C并增加C至G碱基编辑的酶(图45至47)。

[0510] 实施例4:C至G的方法4-消除增加C至G通量的旁路修复途径

[0511] 改善C至G编辑的一种方法是消除或下调旁路修复途径。作为一个实例,消除修复途径蛋白$MSH2^{-/-}$可以导致C至G碱基编辑的增加,显示在图48中。在图49至51中显示了使用各种碱基编辑器(BE3;BE3_UdgX;BE2_UdgX_On;BE3_UdgX_On;BE2_UDG;和BE3_UDG)在$MSH2^{-/-}$细胞中的HEK2、RNF2和FANCF位点处进行的C至G碱基编辑的结果。

[0512] 实施例5:C至G的方法5-反式表达组分

[0513] 用于鉴定一起发挥功能的碱基编辑器组分的一种方法是在细胞中反式地一起表达那些组分。一旦鉴定出诱导C至G突变的碱基编辑器成分(例如,聚合酶,尿嘧啶结合蛋白,碱基切除酶,胞苷脱氨酶和/或核酸可编程DNA结合蛋白),就可以将它们栓系以产生碱基编辑器。与APOBEC-Cas9切口酶融合的表达的UDG和UdgX变体以及同时反式过表达的TLS聚合酶导致RNF2位点处的C至G编辑。图52中显示了反式表达组分的示意图。

[0514] 在图53至55中显示了使用与各种聚合酶(Pol Kappa,Pol Eta,Pol Iota,REV1,Pol Beta和Pol Delta)一起反式表达的五种不同碱基编辑器(BE3;BE3_UdgX;BE2_UdgX_On;BE3_UdgX_On;BE2_UDG;和BE3_UDG)在HEK293细胞中的HEK2、RNF2和FANCF处进行碱基编辑的结果。

[0515] 参考文献

[0516] 1.Chan,K.,Resnick,M.A.,Gordenin,D.A.The choice of nucleotide inserted opposite abasic sites formed within chromosomal DNA reveals the polymerase activities participating in translesion DNA synthesis.DNA Repair 12,878-889(2013).

[0517] 2.Choi,J.Y.,Lim,S.,Kim,E.J.,Jo,A.,and Guengerich F.P.Translesion synthesis across abasic lesions by human B-family and Y-family DNA polymerases alpha,delta,eta,iota,kappa,and Rev1.Journal of Molecular Biology 404,34-44(2010).

[0518] 3.Dianov,G.L.and Hubsher U.Mammalian base excision repair:the forgotten archangel.Nucleic Acids Research,1-8(2013).

[0519] 4.Fortini,P.,Pasucci,B.,Sobol,R.W.,Wilson,S.H.,and Dogliotti,E.Different DNA polymerases are involved in the Short-and lon-patch base excision repair in mammalian cells.Biochemistry 37,3575-3580(1998).

[0520] 5.Jiricny,J.The multifaceted mismatch-repair system.Nature Rev.Molecular Cell Biology 7,335-346(2006).

[0521] 6.Katafuchi A.and Nohmi T.DNA polymerases involved in the incorporation of oxidized nucelotides into DNA:their efficiency and template

base preference.Mutation Research 703,24-31(2010).

[0522] 7.Kavli,B.,Slupphaug,G.,Mol,C.D.,Arvai,A.S.,Peterson,S.B.,Tainer,J.A.,and Krokan,E.H.Excision of cytosine and thymine from DNA by mutants of human uracil-DNA glycosylase.EMBO 15,3442-3447(1996).

[0523] 8.Krokan,H.E.and Bjoras,M.Base Excision Repair,Cold Spring Harbor Perspectives in Biology,1-22(2013).

[0524] 9.Kunkel,T.A.and Erie,D.A.Eukaryotic mismatch repair in relation to RNA replication.Annual Reviews Genetics 49,291-313(2015).

[0525] 10.Li,G.M.Mechanisms and functions of DNA mismatch repair.Cell Research 18,85-98(2008).

[0526] 11.Lin,W.,Xin,H.,Wu,X.,Yuan,F.,and Wang,Z.The human REV1 gene codes for a DNA template-dependent dCMP transferase.Nucleic Acids Research 27,4468-4475(1999).

[0527] 12.Mol,C.D.,Arvai,A.S.,Slupphaug,G.,Kavil,B.,Alseth,I.,Krokan,H.E.,and Tainer,J.A.Crystal structure and mutational analysis of human uracil-DNA glycosylase:structural basis for specificity and catalysis.Cell 80,869-878(1995).

[0528] 13.Prasad,R.,Poltoratsky,V.,Hou,E.W.,and Wilson,S.H.Rev1 is a base excision repair enzyme with 5'deoxyribose phosphate lyase activity.

[0529] Nucleic Acid Research,1-10(2016).

[0530] 14.Robertson,A.B.,Klungland,A.,Rognes,T.,and Leiros,I.Base excision repair:the long and the short of it.Cell Molecular Life Sciences 66,981-993(2009).

[0531] 15.Sale,J.E.,Lehmann,A.R.,and Woodgate,R.Y-Family DNA polymerases and their role in tolerance of cellular DNA damage.Nature Rev.Molecular Cell Biology 13,141-152(2012).

[0532] 16.Sang,P.B.,Srinath,T.,Patil,A.G.,Woo,E.J.,and Varshney,U.A unique uracil-DNA binding protein of the uracil DNA glycosylase superfamily.

[0533] Nucleic Acids Research,1-12(2015).

[0534] 17.Savva,R.,McAuley-Hecht,K.,Brown,T.,and Pearl,L.The structural basis of specific base-excision repair by uracil-DNA glycosylase.Nature 373,487-493(1995).

[0535] 18.Slupphaug,G.,Mol,C.D.,Kavli,B.,Arvai,A.S.,Krokan,H.E.,and Tainer,J.A.A nucleotide-flipping mechanism from the structure of human uracil-DNA glycosylase bound to DNA.Nature 384,87-92(1996).

[0536] 19.Weill,J.C.and Reynaud C.A.DNA polymerases in adaptive immunity.Nature Rev.Immunology 8,302-312(2008).

[0537] 20.Yasui,A.Alternative excision repair pathways.Cold Spring Harbor Perspectives in Biology,1-8(2013).

[0538] 实施例6:Cas9变体序列

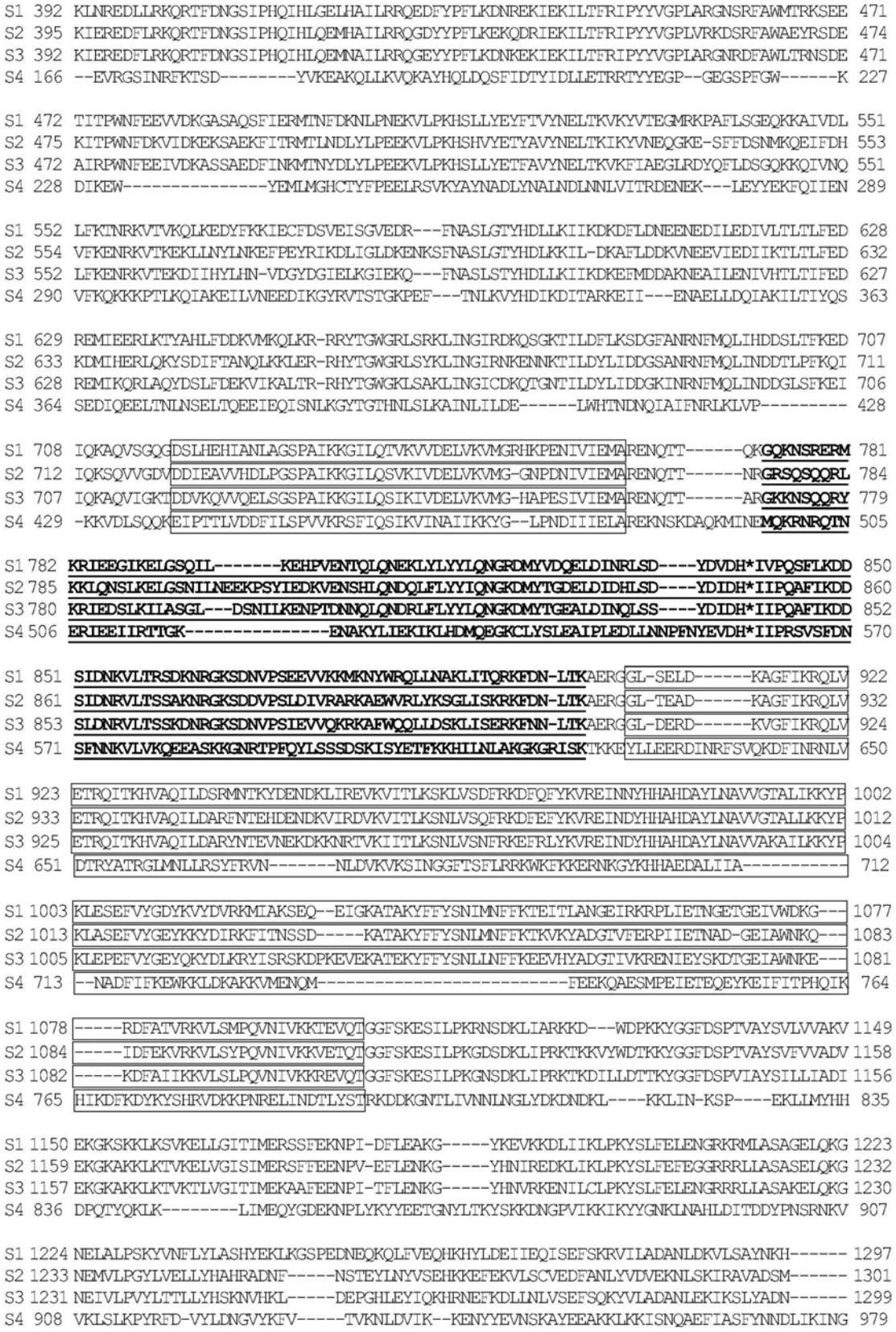
[0539] 本公开提供了Cas9变体,例如来自一种或多种生物体的Cas9蛋白,其可以包含一个或多个突变(例如,以产生dCas9或Cas9切口酶)。在一些实施方案中,可以突变Cas9蛋白的一个或多个氨基酸残基(下文通过星号标识)。在一些实施方案中,SEQ ID NO:6中提供的氨基酸序列的D10和/或H840残基,或本文中提供的任何Cas9,诸如SEQ ID NO:4-26中提供的任一种氨基酸序列中的相应的突变得以突变。在一些实施方案中,SEQ ID NO:6中提供的氨基酸序列的D10残基,或本文中提供的任何Cas9,诸如SEQ ID NO:4-26中提供的任一种氨基酸序列中的相应的突变得以突变为除D之外的任何氨基酸残基。在一些实施方案中,SEQ ID NO:6中提供的氨基酸序列的D10残基,或本文中提供的任何Cas9,诸如SEQ ID NO:4-26中提供的任一种氨基酸序列中的相应的突变得以突变为A。在一些实施方案中,SEQ ID NO:6中提供的氨基酸序列的H840残基,或本文中提供的任何Cas9,诸如SEQ ID NO:4-26中提供的任一种氨基酸序列中的相应的残基是H。在一些实施方案中,SEQ ID NO:6中提供的氨基酸序列的H840残基,或本文中提供的任何Cas9,诸如SEQ ID NO:4-26中提供的任一种氨基酸序列中的相应的突变得以突变为除H之外的任何氨基酸残基。在一些实施方案中,SEQ ID NO:6中提供的氨基酸序列的H840残基,或本文中提供的任何Cas9,诸如SEQ ID NO:4-26中提供的任一种氨基酸序列中的相应的突变得以突变为A。在一些实施方案中,SEQ ID NO:6中提供的氨基酸序列的D10残基,或本文中提供的任何Cas9,诸如SEQ ID NO:4-26中提供的任一种氨基酸序列中的相应的残基是D。

[0540] 对来自各种物种的Cas9序列进行比对以确定是否可以在其他Cas9蛋白中鉴定SEQ ID NO:6的D10和H840的相应的同源氨基酸残基,从而允许具有同源氨基酸残基的相应的突变的Cas9变体产生。使用NCBI基于约束的多重比对工具(NCBI Constraint-based Multiple Alignment Tool)(COBALT(可在st-va.ncbi.nlm.nih.gov/tools/cobalt获得)在以下参数的情况下进行比对。比对参数:空位罚分-11,-1;末端空位罚分-5,-1。CDD参数:使用RPS BLAST开启(on);Blast E值0.003;查找保守列和重新计算开启。查询聚类参数:使用查询聚类开启;词大小4;最大聚类距离0.8;字母常规。

[0541] 以下提供了四个Cas9序列的示例性比对。比对中的Cas9序列是:序列1(S1):SEQ ID NO:23|WP_010922251|gi 499224711|II型CRISPR RNA引导的内切核酸酶Cas9[酿脓链球菌(Streptococcus pyogenes)];序列2(S2):SEQ ID NO:24|WP_039695303|gi 746743737|II型CRISPR RNA引导的内切核酸酶Cas9[解没食子酸链球菌(Streptococcus gallolyticus)];序列3(S3):SEQ ID NO:25|WP_045635197|gi 782887988|II型CRISPR RNA引导的内切核酸酶Cas9[缓症链球菌(Streptococcus mitis)];序列4(S4):SEQ ID NO:26|5AXW_A|gi 924443546|金黄色葡萄球菌(Staphylococcus Aureus)Cas9。针对四个序列中的每一个鉴定HNH域(粗体和下划线)和RuvC域(加框)。S1中的氨基酸残基10和840以及比对序列中的同源氨基酸在相应的氨基酸残基后用星号标识。


S1 1 --MDKK-YSIGLD*IGINSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLI--GALLFDSG--ETAEATRLKRTARRRYT 73
S2 1 --MTKKNYSIGLD*IGINSVGWAVITDDYKVPAKKMKVLGNTDKKYIKKNLL--GALLFDSG--ETAEATRLKRTARRRYT 74
S3 1 --M-KKGYSIGLD*IGINSVGFAVITDDYKVPSKKMKVLGNTDKRFIKKNLI--GALLFDEG--TTAEARRLKRTARRRYT 73
S4 1 GSHMKRNYILGLD*IGITSVGYGII--DYET----------RDVIDAGVRLFKEANVENNEGRRSKRGARRLKR 61
S1 74 RRKNRICYLQEIFSNEMAKVDDSFFHRLEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKADLRL 153
S2 75 RRKNRLRYLQEIFANEIAKVDESFFQRLDESFLTDDDKTFDSHPIFGNKAEEDAYHQKFPTIYHLRKHLADSSEKADLRL 154
S3 74 RRKNRLRYLQEIFSEEMSKVDSSFFHRLDDSFLIPEDKRESKYPIFATLTEEKEYHKQFPTIYHLRKQLADSKEKTDLRL 153
S4 62 RRRHRIQRVKKLL----------FDYNLLTD----------HSELSGINPYEARVKGLSQKLSEEE 107
S1 154 IYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFEENPINASGVDAKAILSARLSKSRRLENLIAQLPGEK 233
S2 155 VYLALAHMIKFRGHFLIEGELNAENTDVQKIFADFVGVYNRTFDDSHLSEITVDVASILTEKISKSRRLENLIKYYPTEK 234
S3 154 IYLALAHMIKYRGHFLYEEAFDIKNNDIQKIFNEFISIYDNTFEGSSLSGQNAQVEAIFTDKISKSAKRERVLKLFPDEK 233
S4 108 FSAALLHLAKRRG----------VHNVNEVEEDT---------- 131
S1 234 KNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAILLSDILRVNTEIT 313
S2 235 KNTLFGNLIALALGLQPNFKTNFKLSEDAKLQFSKDTYEEDLEELLGKIGDDYADLFTSAKNLYDAILLSGILTVDDNST 314
S3 234 STGLFSEFLKLIVGNQADFKKHFDLEDKAPLQFSKDTYDEDLENLLGQIGDDFTDLFVSAKKLYDAILLSGILTVTDPST 313
S4 132 -----GNELS----------TKEQISRN---------- 144
S1 314 KAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIKPILEKM--DGTEELLV 391
S2 315 KAPLSASMIKRYVEHHEDLEKLKEFIKANKSELYHDIFKDKNKNGYAGYIENGVKQDEFYKYLKNILSKIKIDGSDYFLD 394
S3 314 KAPLSASMIERYENHQNDLAALKQFIKNNLPEKYDEVFSDQSKDGYAGYIDGKTTQETFYKYIKNLLSKF--EGTDYFLD 391
S4 145 ----SKALEEKYVAELQ----------LERLKKDG------ 165

[0542]



[0543]

```
S1 1298 RDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVLDATLIHQSIT--------GLYETRI----DLSQL 1365
S2 1302 DNFSIEEISNSFINLLTLTALGAPADFNFLGEKIPRKRYTSTKECLNATLIHQSIT--------GLYETRI----DLSKL 1369
S3 1300 EQADIEILANSFINLLTFTALGAPAAFKFFGKDIDRKRYTTVSEILNATLIHQSIT--------GLYETWI----DLSKL 1367
S4 980  ELYRVIGVNNDLLNRIEVNMIDITYR-EYLENMNDKRPPRIIKTIASKT---QSIKKYSTDILGNLYEVKSKKHPQIIKK 1055
```

[0544]

```
S1 1366 GGD 1368 (SEQ ID NO: 23)
S2 1370 GEE 1372 (SEQ ID NO: 24)
S3 1368 GED 1370 (SEQ ID NO: 25)
S4 1056 G-- 1056 (SEQ ID NO: 26)
```

[0545] 比对证明,可以通过使用本领域已知的比对程序和算法鉴定与参照序列或参照残基比对的氨基酸序列或残基从而在Cas9序列变体间鉴定与参照Cas9氨基酸序列或氨基酸残基同源的氨基酸序列和氨基酸残基,所述Cas9序列变体包括但不限于来自不同物种的Cas9序列。本公开提供了Cas9变体,其中如本文所述突变通过SEQ ID NO:23-26中的星号标识(例如,分别为S1、S2、S3和S4)的一个或多个氨基酸残基。SEQ ID NO:6的Cas9中的残基D10和H840(其对应于SEQ ID NO:23-26中通过星号标识的残基)在本文中称为“同源的”或“相应的”残基。此类同源的残基可以通过序列比对来鉴定,例如,如上所述,和通过鉴定与参照序列或残基比对的序列或残基来鉴定。类似地,对应于本文SEQ ID NO:6中鉴定的突变的Cas9序列中的突变,例如SEQ ID NO:6中的残基10和840的突变,在本文中称为“同源的”或“相应的”突变。例如,对应于SEQ ID NO:6或S1(SEQ ID NO:23)中的D10A突变,对于上述四个比对序列的突变是S2中的D11A、S3中的D10A和S4中的D13A;SEQ ID NO:6或S1(SEQ ID NO:23)中H840A的相应的突变是S2中的H850A、S3中的H842A和S4中的H560A。

[0546] 此外,已经使用上文概述的相同算法和比对参数来比对来自不同物种的几个Cas9序列。使用上文概述的相同算法和比对参数,将来自不同物种的几个Cas9序列(‘632公开文本的SEQ ID NO:11-260)进行比对,并在例如于2017年4月27日公布,标题为“Nucleobase editors and uses thereof”的专利公开文本号WO2017/070632(“632公开文本”)(其通过引用并入本文)中显示。可以使用该方法鉴定与其他Cas9蛋白的残基同源的氨基酸残基,所述方法可以用于将相应的突变掺入其他Cas9蛋白中。以与上文概述相同的方式鉴定与SEQ ID NO:6的残基10和840同源的氨基酸残基。在本文中提供了比对并且通过引用并入。针对四个序列(SEQ ID NO:23-26)中的每种,鉴定出HNH域(粗体和加下划线)和RuvC域(框示)。在比对中在SEQ ID NO:23中框示对应于SEQ ID NO:6中氨基酸残基10和840的单一残基,从而允许鉴定比对序列中的相应氨基酸残基。

[0547] 等同方案和范围,通过引用并入

[0548] 本领域技术人员将认识到或能够仅仅使用常规的实验确定本文所述的本发明具体实施方案的许多等同方案。本发明的范围不意图限于以上说明书,而是如所附权利要求中所述。

[0549] 在权利要求中,诸如“一种”、“一个”和“该”的冠词可以表示一个或超出一个,除非相反地指出或者从上下文中显而易见。若一个、超出一个或所有组成员在给定产物或过程中存在、使用或以其他方式相关,则认为在组中的一个或多个成员之间包括“或”的权利要求或说明书是满足的,除非另有说明或从上下文中显而易见。本发明包括实施方案,其中组的恰好一个成员在给定产物或过程中存在、使用或以其他方式相关。本发明还包括实施方案,其中超出一个或所有组成员在给定产物或过程中存在、使用或以其他方式相关。

[0550] 此外,应理解,本发明涵盖所有变型、组合和置换,其中来自一个或多个权利要求

或来自说明书的相关部分的一个或多个限制、元素、条款、描述性术语等被引入另一个权利要求中。例如，可以修改从属于另一个权利要求的任何权利要求以包括在从属于相同基本权利要求的任何其他权利要求中找到的一个或多个限制。此外，在权利要求叙述组合物的情况下，应当理解包括将组合物用于本文公开的任何目的的方法，并且包括根据本文公开的任何制备方法或本领域中已知的其他方法制备组合物的方法，除非另有说明或者除非本领域普通技术人员明白会出现矛盾或不一致。

[0551] 在将元素呈现为列表(例如，以马库什群组格式)的情况下，应当理解，还公开了元素的每个子群，并且可以从群组中移除任何元素。还应注意，术语“包含”旨在是开放的并且允许包含另外的元素或步骤。应当理解，通常，在本发明或本发明的方面称为包含特定元素、特征、步骤等的情况下，本发明或本发明的方面的某些实施方案由此类元素、特征、步骤等组成，或基本上由此类元素、特征、步骤等组成。出于简化的目的，那些实施方案未在本文中以同样的描述具体阐述。因此，对于包含一个或多个元素、特征、步骤等的本发明的每个实施方案，本发明还提供了由那些元素、特征、步骤等组成或基本上由那些元素、特征、步骤等组成的实施方案。

[0552] 在给出范围的情况下，端点包括在内。此外，应当理解，除非另有说明或从上下文和/或本领域普通技术人员的理解中明显看出，否则表示为范围的值可以假定在本发明的不同实施方案中的所述范围内的任何特定值，至该范围下限的单位的十分之一，除非上下文另有明确规定。还应当理解，除非另有说明或从上下文和/或本领域普通技术人员的理解中明显看出，否则表示为范围的值可以假定给定范围内的任何子范围，其中子范围的端点表示为与范围的下限的单位的十分之一相同的精度。

[0553] 此外，应当理解，本发明的任何具体实施方案可以明确地从任何一个或多个权利要求中排除。在给出范围的情况下，该范围内的任何值可以明确地从任何一个或多个权利要求中排除。本发明的组合物和/或方法的任何实施方案、元素、特征、应用或方面可以从任何一个或多个权利要求中排除。出于简洁的目的，本文未明确阐述其中排除一个或多个元素、特征、目的或方面的所有实施方案。

[0554] 本文提及的所有出版物、专利和序列数据库条目，包括上面列出的那些项，通过引用整体并入本文，如同每个单独的出版物或专利被具体和单独地指出通过引用并入。在冲突的情况下，以本申请(包括本文中的任何定义)为准。

序列表

<110> 哈佛大学的校长及成员们

<120> 胞嘧啶至鸟嘌呤碱基编辑器

<130> H0824.70253WO00

<140> 尚未分配

<141> 2018-03-09

<150> US 62/470,175

<151> 2017-03-10

<160> 123

<170> PatentIn version 3.5

<210> 1

<211> 4104

<212> DNA

<213> 酿脓链球菌

<400> 1

atggataaga aatactcaat aggcttagat atcggcacaa atagcgtcgg atgggcggtg 60
atcactgatg attataaggt tccgtctaaa aagttcaagg ttctgggaaa tacagaccgc 120
cacagtatca aaaaaaatct tataggggct cttttatttg gcagtggaga gacagcggaa 180
gcgactcgtc tcaaacggac agctcgtaga aggtatacac gtcggaagaa tcgtatttgt 240
tatctacagg agatttttc aaatgagatg gcgaaagtag atgatagttt ctttcatcga 300
cttgaagagt ctttttggt ggaagaagac aagaagcatg aacgtcatcc tatttttgga 360
aatatagtag atgaagttgc ttatcatgag aaatatccaa ctatctatca tctgcgaaaa 420
aaattggcag attctactga taaagcggat ttgcgcttaa tctatttggc cttagcgcat 480
atgattaagt ttcgtggtca ttttttgatt gagggagatt taaatcctga taatagtgat 540
gtggacaaac tatttatcca gttggtacaa atctacaatc aattatttga agaaaaccct 600
attaacgcaa gtagagtaga tgctaaagcg attctttctg cacgattgag taaatcaaga 660
cgattagaaa atctcattgc tcagctcccc ggtgagaaga gaaatggctt gtttgggaat 720
ctcattgctt tgtcattggg attgacccct aattttaaat caaattttga tttggcagaa 780
gatgctaaat tacagctttc aaaagatact tacgatgatg atttagataa tttattggcg 840
caaattggag atcaatatgc tgatttgttt ttggcagcta agaatttatc agatgctatt 900
ttactttcag atatcctaag agtaaatagt gaaataacta aggctcccct atcagcttca 960
atgattaagc gctacgatga acatcatcaa gacttgactc ttttaaaagc tttagttcga 1020
caacaacttc cagaaaagta taaagaaatc ttttttgatc aatcaaaaaa cggatatgca 1080
ggttatattg atgggggagc tagccaagaa gaattttata aatttatcaa accaatttta 1140
gaaaaaatgg atggtactga ggaattattg gtgaaactaa atcgtgaaga tttgctgcgc 1200
aagcaacgga cctttgacaa cggctctatt ccccatcaaa ttcacttggg tgagctgcat 1260
gctattttga gaagacaaga agacttttat ccatttttaa aagacaatcg tgagaagatt 1320
gaaaaaatct tgacttttcg aattccttat tatgttggtc cattggcgcg tggcaatagt 1380

cgttttgcat ggatgactcg gaagtctgaa gaaacaatta ccccatggaa ttttgaagaa 1440
gttgtcgata aaggtgcttc agctcaatca tttattgaac gcatgacaaa ctttgataaa 1500
aatcttccaa atgaaaaagt actaccaaaa catagtttgc tttatgagta ttttacggtt 1560
tataacgaat tgacaaaggt caaatatgtt actgagggaa tgcgaaaacc agcatttctt 1620
tcaggtgaac agaagaaagc cattgttgat ttactcttca aaacaaatcg aaaagtaacc 1680
gttaagcaat taaaagaaga ttatttcaaa aaaatagaat gttttgatag tgttgaaatt 1740
tcaggagttg aagatagatt taatgcttca ttaggcgcct accatgattt gctaaaaatt 1800
attaaagata aagatttttt ggataatgaa gaaaatgaag atatcttaga ggatattgtt 1860
ttaacattga ccttatttga agataggggg atgattgagg aaagacttaa aacatatgct 1920
cacctctttg atgataaggt gatgaaacag cttaaacgtc gccgttatac tggttgggga 1980
cgtttgtctc gaaaattgat taatggtatt agggataagc aatctggcaa aacaatatta 2040
gatttttga aatcagatgg ttttgccaat cgcaatttta tgcagctgat ccatgatgat 2100
agtttgacat ttaaagaaga tattcaaaaa gcacaggtgt ctggacaagg ccatagttta 2160
catgaacaga ttgctaactt agctggcagt cctgctatta aaaaaggtat tttacagact 2220
gtaaaaattg ttgatgaact ggtcaaagta atggggcata agccagaaaa tatcgttatt 2280
gaaatggcac gtgaaaatca gacaactcaa aagggccaga aaaattcgcg agagcgtatg 2340
aaacgaatcg aagaaggtat caaagaatta ggaagtcaga ttcttaaaga gcatcctgtt 2400
gaaaatactc aattgcaaaa tgaaaagctc tatctctatt atctacaaaa tggaagagac 2460
atgtatgtgg accaagaatt agatattaat cgtttaagtg attatgatgt cgatcacatt 2520
gttccacaaa gtttcattaa agacgattca atagacaata aggtactaac gcgttctgat 2580
aaaaatcgtg gtaaatcgga taacgttcca agtgaagaag tagtcaaaaa gatgaaaaac 2640
tattggagac aacttctaaa cgccaagtta atcactcaac gtaagtttga taatttaacg 2700
aaagctgaac gtggaggttt gagtgaactt gataaagctg gttttatcaa acgccaattg 2760
gttgaaactc gccaaatcac taagcatgtg gcacaaattt tggatagtcg catgaatact 2820
aaatacgatg aaaatgataa acttattcga gaggttaaag tgattacctt aaaatctaaa 2880
ttagtttctg acttccgaaa agatttccaa ttctataaag tacgtgagat taacaattac 2940
catcatgccc atgatgcgta tctaaatgcc gtcgttggaa ctgctttgat taagaaatat 3000
ccaaaacttg aatcggagtt tgtctatggt gattataaag tttatgatgt tcgtaaaatg 3060
attgctaagt ctgagcaaga aataggcaaa gcaaccgcaa aatatttctt ttactctaat 3120
atcatgaact tcttcaaaac agaaattaca cttgcaaatg gagagattcg caaacgccct 3180
ctaatcgaaa ctaatgggga aactggagaa attgtctggg ataaagggcg agattttgcc 3240
acagtgcgca aagtattgtc catgccccaa gtcaatattg tcaagaaaac agaagtacag 3300
acaggcggat tctccaagga gtcaatttta ccaaaaagaa attcggacaa gcttattgct 3360
cgtaaaaaag actgggatcc aaaaaaatat ggtggttttg atagtccaac ggtagcttat 3420
tcagtcctag tggttgctaa ggtggaaaaa gggaaatcga agaagttaaa atccgttaaa 3480
gagttactag ggatcacaat tatggaaaga agttcctttg aaaaaaatcc gattgacttt 3540
ttagaagcta aaggatataa ggaagttaaa aaagacttaa tcattaaact acctaaatat 3600
agtctttttg agttagaaaa cggtcgtaaa cggatgctgg ctagtgccgg agaattacaa 3660
aaaggaaatg agctggctct gccaagcaaa tatgtgaatt ttttatattt agctagtcat 3720

tatgaaaagt tgaagggtag tccagaagat aacgaacaaa aacaattgtt tgtggagcag 3780
cataagcatt atttagatga gattattgag caaatcagtg aattttctaa gcgtgttatt 3840
ttagcagatg ccaatttaga taaagttctt agtgcatata acaaacatag agacaaacca 3900
atacgtgaac aagcagaaaa tattattcat ttatttacgt tgacgaatct tggagctccc 3960
gctgctttta aatattttga tacaacaatt gatcgtaaac gatatacgtc tacaaaagaa 4020
gttttagatg ccactcttat ccatcaatcc atcactggtc tttatgaaac acgcattgat 4080
ttgagtcagc taggaggtga ctga 4104

<210> 2
<211> 4212
<212> DNA
<213> 人工序列
<220>
<223> 合成多核苷酸
<400> 2

atggataaaa agtattctat tggtttagac atcggcacta attccgttgg atgggctgtc 60
ataaccgatg aatacaaagt accttcaaag aaatttaagg tgttggggaa cacagaccgt 120
cattcgatta aaaagaatct tatcggtgcc ctcctattcg atagtggcga aacggcagag 180
gcgactcgcc tgaaacgaac cgctcggaga aggtatacac gtcgcaagaa ccgaatatgt 240
tacttacaag aaatttttag caatgagatg gccaaagttg acgattcttt ctttcaccgt 300
ttggaagagt ccttccttgt cgaagaggac aagaaacatg aacggcaccc catctttgga 360
aacatagtag atgaggtggc atatcatgaa aagtacccaa cgatttatca cctcagaaaa 420
aagctagttg actcaactga taaagcggac ctgaggttaa tctacttggc tcttgcccat 480
atgataaagt tccgtgggca ctttctcatt gagggtgatc taaatccgga caactcggat 540
gtcgacaaac tgttcatcca gttagtacaa acctataatc agttgtttga agagaaccct 600
ataaatgcaa gtggcgtgga tgcgaaggct attcttagcg cccgcctctc taaatcccga 660
cggctagaaa acctgatcgc acaattaccc ggagagaaga aaaatgggtt gttcggtaac 720
cttatagcgc tctcactagg cctgacacca aattttaagt cgaacttcga cttagctgaa 780
gatgccaaat tgcagcttag taaggacacg tacgatgacg atctcgacaa tctactggca 840
caaattggag atcagtatgc ggacttattt ttggctgcca aaaaccttag cgatgcaatc 900
ctcctatctg acatactgag agttaatact gagattacca aggcgccgtt atccgcttca 960
atgatcaaaa ggtacgatga acatcaccaa gacttgacac ttctcaaggc cctagtccgt 1020
cagcaactgc ctgagaaata taaggaaata ttctttgatc agtcgaaaaa cgggtacgca 1080
ggttatattg acggcggagc gagtcaagag gaattctaca agtttatcaa acccatatta 1140
gagaagatgg atgggacgga agagttgctt gtaaaactca atcgcgaaga tctactgcga 1200
aagcagcgga ctttcgacaa cggtagcatt ccacatcaaa tccacttagg cgaattgcat 1260
gctatactta gaaggcagga ggatttttat ccgttcctca aagacaatcg tgaaaagatt 1320
gagaaaatcc taacctttcg cataccttac tatgtgggac ccctggcccg agggaactct 1380
cggttcgcat ggatgacaag aaagtccgaa gaaacgatta ctccatggaa ttttgaggaa 1440
gttgtcgata aaggtgcgtc agctcaatcg ttcatcgaga ggatgaccaa ctttgacaag 1500

aatttaccga acgaaaaagt attgcctaag cacagtttac tttacgagta tttcacagtg 1560
tacaatgaac tcacgaaagt taagtatgtc actgagggca tgcgtaaacc cgcctttcta 1620
agcggagaac agaagaaagc aatagtagat ctgttattca agaccaaccg caaagtgaca 1680
gttaagcaat tgaaagagga ctactttaag aaaattgaat gcttcgattc tgtcgagatc 1740
tccggggtag aagatcgatt taatgcgtca cttggtacgt atcatgacct cctaaagata 1800
attaaagata aggacttcct ggataacgaa gagaatgaag atatcttaga agatatagtg 1860
ttgactctta ccctctttga agatcgggaa atgattgagg aaagactaaa aacatacgct 1920
cacctgttcg acgataaggt tatgaaacag ttaaagaggc gtcgctatac gggctgggga 1980
cgattgtcgc ggaaacttat caacgggata agagacaagc aaagtggtaa aactattctc 2040
gattttctaa agagcgacgg cttcgccaat aggaacttta tgcagctgat ccatgatgac 2100
tctttaacct tcaaagagga tatacaaaag gcacaggttt ccggacaagg ggactcattg 2160
cacgaacata ttgcgaatct tgctggttcg ccagccatca aaaagggcat actccagaca 2220
gtcaaagtag tggatgagct agttaaggtc atgggacgtc acaaaccgga aaacattgta 2280
atcgagatgg cacgcgaaaa tcaaacgact cagaaggggc aaaaaaacag tcgagagcgg 2340
atgaagagaa tagaagaggg tattaaagaa ctgggcagcc agatcttaaa ggagcatcct 2400
gtggaaaata cccaattgca gaacgagaaa ctttacctct attacctaca aaatggaagg 2460
gacatgtatg ttgatcagga actggacata aaccgtttat ctgattacga cgtcgatcac 2520
attgtacccc aatccttttt gaaggacgat tcaatcgaca ataaagtgct tacacgctcg 2580
gataagaacc gagggaaaag tgacaatgtt ccaagcgagg aagtcgtaaa gaaaatgaag 2640
aactattggc ggcagctcct aaatgcgaaa ctgataacgc aaagaaagtt cgataactta 2700
actaaagctg agaggggtgg cttgtctgaa cttgacaagg ccggatttat taaacgtcag 2760
ctcgtggaaa cccgccaaat cacaaagcat gttgcacaga tactagattc ccgaatgaat 2820
acgaaatacg acgagaacga taagctgatt cgggaagtca aagtaatcac tttaaagtca 2880
aaattggtgt cggacttcag aaaggatttt caattctata aagttaggga gataaataac 2940
taccaccatg cgcacgacgc ttatcttaat gccgtcgtag ggaccgcact cattaagaaa 3000
tacccgaagc tagaaagtga gtttgtgtat ggtgattaca aagtttatga cgtccgtaag 3060
atgatcgcga aaagcgaaca ggagataggc aaggctacag ccaaatactt cttttattct 3120
aacattatga atttctttaa gacggaaatc actctggcaa acggagagat acgcaaacga 3180
cctttaattg aaaccaatgg ggagacaggt gaaatcgtat gggataaggg ccgggacttc 3240
gcgacggtga gaaaagtttt gtccatgccc caagtcaaca tagtaaagaa aactgaggtg 3300
cagaccggag ggttttcaaa ggaatcgatt cttccaaaaa ggaatagtga taagctcatc 3360
gctcgtaaaa aggactggga cccgaaaaag tacggtggct tcgatagccc tacagttgcc 3420
tattctgtcc tagtagtggc aaaagttgag aagggaaaat ccaagaaact gaagtcagtc 3480
aaagaattat tggggataac gattatggag cgctcgtctt ttgaaaagaa ccccatcgac 3540
ttccttgagg cgaaaggtta caaggaagta aaaaaggatc tcataattaa actaccaaag 3600
tatagtctgt ttgagttaga aaatggccga aaacggatgt tggctagcgc cggagagctt 3660
caaaagggga acgaactcgc actaccgtct aaatacgtga atttcctgta tttagcgtcc 3720
cattacgaga agttgaaagg ttcacctgaa gataacgaac agaagcaact ttttgttgag 3780
cagcacaaac attatctcga cgaaatcata gagcaaattt cggaattcag taagagagtc 3840

atcctagctg atgccaatct ggacaaagta ttaagcgcat acaacaagca cagggataaa 3900
cccatacgtg agcaggcgga aaatattatc catttgttta ctcttaccaa cctcggcgct 3960
ccagccgcat tcaagtattt tgacacaacg atagatcgca aacgatacac ttctaccaag 4020
gaggtgctag acgcgacact gattcaccaa tccatcacgg gattatatga aactcggata 4080
gatttgtcac agcttggggg tgacggatcc cccaagaaga agaggaaagt ctcgagcgac 4140
tacaaagacc atgacggtga ttataaagat catgacatcg attacaagga tgacgatgac 4200
aaggctgcag ga 4212

<210> 3
<211> 4107
<212> DNA
<213> 酿脓链球菌
<400> 3

atggataaga aatactcaat aggcttagat atcggcacaa atagcgtcgg atgggcggtg 60
atcactgatg aatataaggt tccgtctaaa aagttcaagg ttctgggaaa tacagaccgc 120
cacagtatca aaaaaaatct tataggggct cttttatttg acagtggaga gacagcggaa 180
gcgactcgtc tcaaacggac agctcgtaga aggtatacac gtcggaagaa tcgtatttgt 240
tatctacagg agatttttc aaatgagatg gcgaaagtag atgatagttt ctttcatcga 300
cttgaagagt ctttttggt ggaagaagac aagaagcatg aacgtcatcc tatttttgga 360
aatatagtag atgaagttgc ttatcatgag aaatatccaa ctatctatca tctgcgaaaa 420
aaattggtag attctactga taaagcggat ttgcgcttaa tctatttggc cttagcgcat 480
atgattaagt ttcgtggtca ttttttgatt gagggagatt taaatcctga taatagtgat 540
gtggacaaac tatttatcca gttggtacaa acctacaatc aattatttga agaaaaccct 600
attaacgcaa gtggagtaga tgctaaagcg attctttctg cacgattgag taaatcaaga 660
cgattagaaa atctcattgc tcagctcccc ggtgagaaga aaaatggctt atttgggaat 720
ctcattgctt tgtcattggg tttgacccct aattttaaat caaattttga tttggcagaa 780
gatgctaaat tacagctttc aaaagatact tacgatgatg atttagataa tttattggcg 840
caaattggag atcaatatgc tgatttgttt ttggcagcta agaatttatc agatgctatt 900
ttactttcag atatcctaag agtaaatact gaaataacta aggctcccct atcagcttca 960
atgattaaac gctacgatga acatcatcaa gacttgactc ttttaaaagc tttagttcga 1020
caacaacttc cagaaaagta taaagaaatc ttttttgatc aatcaaaaaa cggatatgca 1080
ggttatattg atgggggagc tagccaagaa gaattttata aatttatcaa accaatttta 1140
gaaaaaatgg atggtactga ggaattattg gtgaaactaa atcgtgaaga tttgctgcgc 1200
aagcaacgga cctttgacaa cggctctatt ccccatcaaa ttcacttggg tgagctgcat 1260
gctattttga gaagacaaga agacttttat ccatttttaa aagacaatcg tgagaagatt 1320
gaaaaaatct tgacttttcg aattccttat tatgttggtc cattggcgcg tggcaatagt 1380
cgttttgcat ggatgactcg gaagtctgaa gaaacaatta ccccatggaa ttttgaagaa 1440
gttgtcgata aaggtgcttc agctcaatca tttattgaac gcatgacaaa ctttgataaa 1500
aatcttccaa atgaaaaagt actaccaaaa catagtttgc tttatgagta ttttacggtt 1560
tataacgaat tgacaaaggt caaatatgtt actgaaggaa tgcgaaaacc agcatttctt 1620

tcaggtgaac agaagaaagc cattgttgat ttactcttca aaacaaatcg aaaagtaacc 1680
gttaagcaat taaaagaaga ttatttcaaa aaaatagaat gttttgatag tgttgaaatt 1740
tcaggagttg aagatagatt taatgcttca ttaggtacct accatgattt gctaaaaatt 1800
attaaagata aagatttttt ggataatgaa gaaaatgaag atatcttaga ggatattgtt 1860
ttaacattga ccttatttga agatagggag atgattgagg aaagacttaa aacatatgct 1920
cacctctttg atgataaggt gatgaaacag cttaaacgtc gccgttatac tggttgggga 1980
cgtttgtctc gaaaattgat taatggtatt agggataagc aatctggcaa aacaatatta 2040
gatttttga aatcagatgg ttttgccaat cgcaatttta tgcagctgat ccatgatgat 2100
agtttgacat ttaaagaaga cattcaaaaa gcacaagtgt ctggacaagg cgatagttta 2160
catgaacata ttgcaaattt agctggtagc cctgctatta aaaaaggtat tttacagact 2220
gtaaaagttg ttgatgaatt ggtcaaagta atggggcggc ataagccaga aaatatcgtt 2280
attgaaatgg cacgtgaaaa tcagacaact caaaagggcc agaaaaattc gcgagagcgt 2340
atgaaacgaa tcgaagaagg tatcaaagaa ttaggaagtc agattcttaa agagcatcct 2400
gttgaaaata ctcaattgca aaatgaaaag ctctatctct attatctcca aaatggaaga 2460
gacatgtatg tggaccaaga attagatatt aatcgtttaa gtgattatga tgtcgatcac 2520
attgttccac aaagtttcct taaagacgat tcaatagaca ataaggtctt aacgcgttct 2580
gataaaaatc gtggtaaatc ggataacgtt ccaagtgaag aagtagtcaa aaagatgaaa 2640
aactattgga gacaacttct aaacgccaag ttaatcactc aacgtaagtt tgataattta 2700
acgaaagctg aacgtggagg tttgagtgaa cttgataaag ctggttttat caaacgccaa 2760
ttggttgaaa ctcgccaaat cactaagcat gtggcacaaa ttttggatag tcgcatgaat 2820
actaaatacg atgaaaatga taaacttatt cgagaggtta aagtgattac cttaaaatct 2880
aaattagttt ctgacttccg aaaagatttc caattctata aagtacgtga gattaacaat 2940
taccatcatg cccatgatgc gtatctaaat gccgtcgttg gaactgcttt gattaagaaa 3000
tatccaaaac ttgaatcgga gtttgtctat ggtgattata aagtttatga tgttcgtaaa 3060
atgattgcta agtctgagca agaaataggc aaagcaaccg caaaatattt cttttactct 3120
aatatcatga acttcttcaa aacagaaatt acacttgcaa atggagagat tcgcaaacgc 3180
cctctaatcg aaactaatgg ggaaactgga gaaattgtct gggataaagg gcgagatttt 3240
gccacagtgc gcaaagtatt gtccatgccc caagtcaata ttgtcaagaa aacagaagta 3300
cagacaggcg gattctccaa ggagtcaatt ttaccaaaaa gaaattcgga caagcttatt 3360
gctcgtaaaa aagactggga tccaaaaaaa tatggtggtt ttgatagtcc aacggtagct 3420
tattcagtcc tagtggttgc taaggtggaa aaagggaaat cgaagaagtt aaaatccgtt 3480
aaagagttac tagggatcac aattatggaa agaagttcct ttgaaaaaaa tccgattgac 3540
tttttagaag ctaaaggata taaggaagtt aaaaaagact taatcattaa actacctaaa 3600
tatagtcttt ttgagttaga aaacggtcgt aaacggatgc tggctagtgc cggagaatta 3660
caaaaaggaa atgagctggc tctgccaagc aaatatgtga atttttata tttagctagt 3720
cattatgaaa agttgaaggg tagtccagaa gataacgaac aaaaacaatt gtttgtggag 3780
cagcataagc attatttaga tgagattatt gagcaaatca gtgaattttc taagcgtgtt 3840
attttagcag atgccaattt agataaagtt cttagtgcat ataacaaaca tagagacaaa 3900
ccaatacgtg aacaagcaga aaatattatt catttattta cgttgacgaa tcttggagct 3960

cccgctgctt ttaaatattt tgatacaaca attgatcgta aacgatatac gtctacaaaa 4020
gaagttttag atgccactct tatccatcaa tccatcactg gtctttatga aacacgcatt 4080
gatttgagtc agctaggagg tgactga 4107

<210> 4
<211> 1367
<212> PRT
<213> 酿脓链球菌
<400> 4

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

245 250 255
Asp Leu Ala Glu Asp Ala Lys Leu Gln Leu Ser Lys Asp Thr Tyr Asp
260 265 270
Asp Asp Leu Asp Asn Leu Leu Ala Gln Ile Gly Asp Gln Tyr Ala Asp
275 280 285
Leu Phe Leu Ala Ala Lys Asn Leu Ser Asp Ala Ile Leu Leu Ser Asp
290 295 300
Ile Leu Arg Val Asn Ser Glu Ile Thr Lys Ala Pro Leu Ser Ala Ser
305 310 315 320
Met Ile Lys Arg Tyr Asp Glu His His Gln Asp Leu Thr Leu Leu Lys
325 330 335
Ala Leu Val Arg Gln Gln Leu Pro Glu Lys Tyr Lys Glu Ile Phe Phe
340 345 350
Asp Gln Ser Lys Asn Gly Tyr Ala Gly Tyr Ile Asp Gly Gly Ala Ser
355 360 365
Gln Glu Glu Phe Tyr Lys Phe Ile Lys Pro Ile Leu Glu Lys Met Asp
370 375 380
Gly Thr Glu Glu Leu Leu Val Lys Leu Asn Arg Glu Asp Leu Leu Arg
385 390 395 400
Lys Gln Arg Thr Phe Asp Asn Gly Ser Ile Pro His Gln Ile His Leu
405 410 415
Gly Glu Leu His Ala Ile Leu Arg Arg Gln Glu Asp Phe Tyr Pro Phe
420 425 430
Leu Lys Asp Asn Arg Glu Lys Ile Glu Lys Ile Leu Thr Phe Arg Ile
435 440 445
Pro Tyr Tyr Val Gly Pro Leu Ala Arg Gly Asn Ser Arg Phe Ala Trp
450 455 460
Met Thr Arg Lys Ser Glu Glu Thr Ile Thr Pro Trp Asn Phe Glu Glu
465 470 475 480
Val Val Asp Lys Gly Ala Ser Ala Gln Ser Phe Ile Glu Arg Met Thr
485 490 495
Asn Phe Asp Lys Asn Leu Pro Asn Glu Lys Val Leu Pro Lys His Ser
500 505 510
Leu Leu Tyr Glu Tyr Phe Thr Val Tyr Asn Glu Leu Thr Lys Val Lys
515 520 525
Tyr Val Thr Glu Gly Met Arg Lys Pro Ala Phe Leu Ser Gly Glu Gln
530 535 540
Lys Lys Ala Ile Val Asp Leu Leu Phe Lys Thr Asn Arg Lys Val Thr
545 550 555 560

Val Lys Gln Leu Lys Glu Asp Tyr Phe Lys Lys Ile Glu Cys Phe Asp
565 570 575
Ser Val Glu Ile Ser Gly Val Glu Asp Arg Phe Asn Ala Ser Leu Gly
580 585 590
Ala Tyr His Asp Leu Leu Lys Ile Ile Lys Asp Lys Asp Phe Leu Asp
595 600 605
Asn Glu Glu Asn Glu Asp Ile Leu Glu Asp Ile Val Leu Thr Leu Thr
610 615 620
Leu Phe Glu Asp Arg Gly Met Ile Glu Glu Arg Leu Lys Thr Tyr Ala
625 630 635 640
His Leu Phe Asp Asp Lys Val Met Lys Gln Leu Lys Arg Arg Arg Tyr
645 650 655
Thr Gly Trp Gly Arg Leu Ser Arg Lys Leu Ile Asn Gly Ile Arg Asp
660 665 670
Lys Gln Ser Gly Lys Thr Ile Leu Asp Phe Leu Lys Ser Asp Gly Phe
675 680 685
Ala Asn Arg Asn Phe Met Gln Leu Ile His Asp Asp Ser Leu Thr Phe
690 695 700
Lys Glu Asp Ile Gln Lys Ala Gln Val Ser Gly Gln Gly His Ser Leu
705 710 715 720
His Glu Gln Ile Ala Asn Leu Ala Gly Ser Pro Ala Ile Lys Lys Gly
725 730 735
Ile Leu Gln Thr Val Lys Ile Val Asp Glu Leu Val Lys Val Met Gly
740 745 750
His Lys Pro Glu Asn Ile Val Ile Glu Met Ala Arg Glu Asn Gln Thr
755 760 765
Thr Gln Lys Gly Gln Lys Asn Ser Arg Glu Arg Met Lys Arg Ile Glu
770 775 780
Glu Gly Ile Lys Glu Leu Gly Ser Gln Ile Leu Lys Glu His Pro Val
785 790 795 800
Glu Asn Thr Gln Leu Gln Asn Glu Lys Leu Tyr Leu Tyr Tyr Leu Gln
805 810 815
Asn Gly Arg Asp Met Tyr Val Asp Gln Glu Leu Asp Ile Asn Arg Leu
820 825 830
Ser Asp Tyr Asp Val Asp His Ile Val Pro Gln Ser Phe Ile Lys Asp
835 840 845
Asp Ser Ile Asp Asn Lys Val Leu Thr Arg Ser Asp Lys Asn Arg Gly
850 855 860
Lys Ser Asp Asn Val Pro Ser Glu Glu Val Val Lys Lys Met Lys Asn

865 870 875 880
Tyr Trp Arg Gln Leu Leu Asn Ala Lys Leu Ile Thr Gln Arg Lys Phe
885 890 895
Asp Asn Leu Thr Lys Ala Glu Arg Gly Gly Leu Ser Glu Leu Asp Lys
900 905 910
Ala Gly Phe Ile Lys Arg Gln Leu Val Glu Thr Arg Gln Ile Thr Lys
915 920 925
His Val Ala Gln Ile Leu Asp Ser Arg Met Asn Thr Lys Tyr Asp Glu
930 935 940
Asn Asp Lys Leu Ile Arg Glu Val Lys Val Ile Thr Leu Lys Ser Lys
945 950 955 960
Leu Val Ser Asp Phe Arg Lys Asp Phe Gln Phe Tyr Lys Val Arg Glu
965 970 975
Ile Asn Asn Tyr His His Ala His Asp Ala Tyr Leu Asn Ala Val Val
980 985 990
Gly Thr Ala Leu Ile Lys Lys Tyr Pro Lys Leu Glu Ser Glu Phe Val
995 1000 1005
Tyr Gly Asp Tyr Lys Val Tyr Asp Val Arg Lys Met Ile Ala Lys
1010 1015 1020
Ser Glu Gln Glu Ile Gly Lys Ala Thr Ala Lys Tyr Phe Phe Tyr
1025 1030 1035
Ser Asn Ile Met Asn Phe Phe Lys Thr Glu Ile Thr Leu Ala Asn
1040 1045 1050
Gly Glu Ile Arg Lys Arg Pro Leu Ile Glu Thr Asn Gly Glu Thr
1055 1060 1065
Gly Glu Ile Val Trp Asp Lys Gly Arg Asp Phe Ala Thr Val Arg
1070 1075 1080
Lys Val Leu Ser Met Pro Gln Val Asn Ile Val Lys Lys Thr Glu
1085 1090 1095
Val Gln Thr Gly Gly Phe Ser Lys Glu Ser Ile Leu Pro Lys Arg
1100 1105 1110
Asn Ser Asp Lys Leu Ile Ala Arg Lys Lys Asp Trp Asp Pro Lys
1115 1120 1125
Lys Tyr Gly Gly Phe Asp Ser Pro Thr Val Ala Tyr Ser Val Leu
1130 1135 1140
Val Val Ala Lys Val Glu Lys Gly Lys Ser Lys Lys Leu Lys Ser
1145 1150 1155
Val Lys Glu Leu Leu Gly Ile Thr Ile Met Glu Arg Ser Ser Phe
1160 1165 1170

Glu Lys Asn Pro Ile Asp Phe Leu Glu Ala Lys Gly Tyr Lys Glu
1175 1180 1185
Val Lys Lys Asp Leu Ile Ile Lys Leu Pro Lys Tyr Ser Leu Phe
1190 1195 1200
Glu Leu Glu Asn Gly Arg Lys Arg Met Leu Ala Ser Ala Gly Glu
1205 1210 1215
Leu Gln Lys Gly Asn Glu Leu Ala Leu Pro Ser Lys Tyr Val Asn
1220 1225 1230
Phe Leu Tyr Leu Ala Ser His Tyr Glu Lys Leu Lys Gly Ser Pro
1235 1240 1245
Glu Asp Asn Glu Gln Lys Gln Leu Phe Val Glu Gln His Lys His
1250 1255 1260
Tyr Leu Asp Glu Ile Ile Glu Gln Ile Ser Glu Phe Ser Lys Arg
1265 1270 1275
Val Ile Leu Ala Asp Ala Asn Leu Asp Lys Val Leu Ser Ala Tyr
1280 1285 1290
Asn Lys His Arg Asp Lys Pro Ile Arg Glu Gln Ala Glu Asn Ile
1295 1300 1305
Ile His Leu Phe Thr Leu Thr Asn Leu Gly Ala Pro Ala Ala Phe
1310 1315 1320
Lys Tyr Phe Asp Thr Thr Ile Asp Arg Lys Arg Tyr Thr Ser Thr
1325 1330 1335
Lys Glu Val Leu Asp Ala Thr Leu Ile His Gln Ser Ile Thr Gly
1340 1345 1350
Leu Tyr Glu Thr Arg Ile Asp Leu Ser Gln Leu Gly Gly Asp
1355 1360 1365

<210> 5
<211> 1368
<212> PRT
<213> 人工序列
<220>
<223> 合成多肽
<400> 5

Met Asp Lys Lys Tyr Ser Ile Gly Leu Ala Ile Gly Thr Asn Ser Val
1 5 10 15
Gly Trp Ala Val Ile Thr Asp Glu Tyr Lys Val Pro Ser Lys Lys Phe
20 25 30
Lys Val Leu Gly Asn Thr Asp Arg His Ser Ile Lys Lys Asn Leu Ile
35 40 45

Gly Ala Leu Leu Phe Asp Ser Gly Glu Thr Ala Glu Ala Thr Arg Leu
50 55 60
Lys Arg Thr Ala Arg Arg Arg Tyr Thr Arg Arg Lys Asn Arg Ile Cys
65 70 75 80
Tyr Leu Gln Glu Ile Phe Ser Asn Glu Met Ala Lys Val Asp Asp Ser
85 90 95
Phe Phe His Arg Leu Glu Glu Ser Phe Leu Val Glu Glu Asp Lys Lys
100 105 110
His Glu Arg His Pro Ile Phe Gly Asn Ile Val Asp Glu Val Ala Tyr
115 120 125
His Glu Lys Tyr Pro Thr Ile Tyr His Leu Arg Lys Lys Leu Val Asp
130 135 140
Ser Thr Asp Lys Ala Asp Leu Arg Leu Ile Tyr Leu Ala Leu Ala His
145 150 155 160
Met Ile Lys Phe Arg Gly His Phe Leu Ile Glu Gly Asp Leu Asn Pro
165 170 175
Asp Asn Ser Asp Val Asp Lys Leu Phe Ile Gln Leu Val Gln Thr Tyr
180 185 190
Asn Gln Leu Phe Glu Glu Asn Pro Ile Asn Ala Ser Gly Val Asp Ala
195 200 205
Lys Ala Ile Leu Ser Ala Arg Leu Ser Lys Ser Arg Arg Leu Glu Asn
210 215 220
Leu Ile Ala Gln Leu Pro Gly Glu Lys Lys Asn Gly Leu Phe Gly Asn
225 230 235 240
Leu Ile Ala Leu Ser Leu Gly Leu Thr Pro Asn Phe Lys Ser Asn Phe
245 250 255
Asp Leu Ala Glu Asp Ala Lys Leu Gln Leu Ser Lys Asp Thr Tyr Asp
260 265 270
Asp Asp Leu Asp Asn Leu Leu Ala Gln Ile Gly Asp Gln Tyr Ala Asp
275 280 285
Leu Phe Leu Ala Ala Lys Asn Leu Ser Asp Ala Ile Leu Leu Ser Asp
290 295 300
Ile Leu Arg Val Asn Thr Glu Ile Thr Lys Ala Pro Leu Ser Ala Ser
305 310 315 320
Met Ile Lys Arg Tyr Asp Glu His His Gln Asp Leu Thr Leu Leu Lys
325 330 335
Ala Leu Val Arg Gln Gln Leu Pro Glu Lys Tyr Lys Glu Ile Phe Phe
340 345 350
Asp Gln Ser Lys Asn Gly Tyr Ala Gly Tyr Ile Asp Gly Gly Ala Ser

355 360 365

Gln Glu Glu Phe Tyr Lys Phe Ile Lys Pro Ile Leu Glu Lys Met Asp
370 375 380

Gly Thr Glu Glu Leu Leu Val Lys Leu Asn Arg Glu Asp Leu Leu Arg
385 390 395 400

Lys Gln Arg Thr Phe Asp Asn Gly Ser Ile Pro His Gln Ile His Leu
405 410 415

Gly Glu Leu His Ala Ile Leu Arg Arg Gln Glu Asp Phe Tyr Pro Phe
420 425 430

Leu Lys Asp Asn Arg Glu Lys Ile Glu Lys Ile Leu Thr Phe Arg Ile
435 440 445

Pro Tyr Tyr Val Gly Pro Leu Ala Arg Gly Asn Ser Arg Phe Ala Trp
450 455 460

Met Thr Arg Lys Ser Glu Glu Thr Ile Thr Pro Trp Asn Phe Glu Glu
465 470 475 480

Val Val Asp Lys Gly Ala Ser Ala Gln Ser Phe Ile Glu Arg Met Thr
485 490 495

Asn Phe Asp Lys Asn Leu Pro Asn Glu Lys Val Leu Pro Lys His Ser
500 505 510

Leu Leu Tyr Glu Tyr Phe Thr Val Tyr Asn Glu Leu Thr Lys Val Lys
515 520 525

Tyr Val Thr Glu Gly Met Arg Lys Pro Ala Phe Leu Ser Gly Glu Gln
530 535 540

Lys Lys Ala Ile Val Asp Leu Leu Phe Lys Thr Asn Arg Lys Val Thr
545 550 555 560

Val Lys Gln Leu Lys Glu Asp Tyr Phe Lys Lys Ile Glu Cys Phe Asp
565 570 575

Ser Val Glu Ile Ser Gly Val Glu Asp Arg Phe Asn Ala Ser Leu Gly
580 585 590

Thr Tyr His Asp Leu Leu Lys Ile Ile Lys Asp Lys Asp Phe Leu Asp
595 600 605

Asn Glu Glu Asn Glu Asp Ile Leu Glu Asp Ile Val Leu Thr Leu Thr
610 615 620

Leu Phe Glu Asp Arg Glu Met Ile Glu Glu Arg Leu Lys Thr Tyr Ala
625 630 635 640

His Leu Phe Asp Asp Lys Val Met Lys Gln Leu Lys Arg Arg Arg Tyr
645 650 655

Thr Gly Trp Gly Arg Leu Ser Arg Lys Leu Ile Asn Gly Ile Arg Asp
660 665 670

Lys Gln Ser Gly Lys Thr Ile Leu Asp Phe Leu Lys Ser Asp Gly Phe
675 680 685
Ala Asn Arg Asn Phe Met Gln Leu Ile His Asp Asp Ser Leu Thr Phe
690 695 700
Lys Glu Asp Ile Gln Lys Ala Gln Val Ser Gly Gln Gly Asp Ser Leu
705 710 715 720
His Glu His Ile Ala Asn Leu Ala Gly Ser Pro Ala Ile Lys Lys Gly
725 730 735
Ile Leu Gln Thr Val Lys Val Val Asp Glu Leu Val Lys Val Met Gly
740 745 750
Arg His Lys Pro Glu Asn Ile Val Ile Glu Met Ala Arg Glu Asn Gln
755 760 765
Thr Thr Gln Lys Gly Gln Lys Asn Ser Arg Glu Arg Met Lys Arg Ile
770 775 780
Glu Glu Gly Ile Lys Glu Leu Gly Ser Gln Ile Leu Lys Glu His Pro
785 790 795 800
Val Glu Asn Thr Gln Leu Gln Asn Glu Lys Leu Tyr Leu Tyr Tyr Leu
805 810 815
Gln Asn Gly Arg Asp Met Tyr Val Asp Gln Glu Leu Asp Ile Asn Arg
820 825 830
Leu Ser Asp Tyr Asp Val Asp His Ile Val Pro Gln Ser Phe Leu Lys
835 840 845
Asp Asp Ser Ile Asp Asn Lys Val Leu Thr Arg Ser Asp Lys Asn Arg
850 855 860
Gly Lys Ser Asp Asn Val Pro Ser Glu Glu Val Val Lys Lys Met Lys
865 870 875 880
Asn Tyr Trp Arg Gln Leu Leu Asn Ala Lys Leu Ile Thr Gln Arg Lys
885 890 895
Phe Asp Asn Leu Thr Lys Ala Glu Arg Gly Gly Leu Ser Glu Leu Asp
900 905 910
Lys Ala Gly Phe Ile Lys Arg Gln Leu Val Glu Thr Arg Gln Ile Thr
915 920 925
Lys His Val Ala Gln Ile Leu Asp Ser Arg Met Asn Thr Lys Tyr Asp
930 935 940
Glu Asn Asp Lys Leu Ile Arg Glu Val Lys Val Ile Thr Leu Lys Ser
945 950 955 960
Lys Leu Val Ser Asp Phe Arg Lys Asp Phe Gln Phe Tyr Lys Val Arg
965 970 975
Glu Ile Asn Asn Tyr His His Ala His Asp Ala Tyr Leu Asn Ala Val

980 985 990

Val Gly Thr Ala Leu Ile Lys Lys Tyr Pro Lys Leu Glu Ser Glu Phe

995 1000 1005

Val Tyr Gly Asp Tyr Lys Val Tyr Asp Val Arg Lys Met Ile Ala

1010 1015 1020

Lys Ser Glu Gln Glu Ile Gly Lys Ala Thr Ala Lys Tyr Phe Phe

1025 1030 1035

Tyr Ser Asn Ile Met Asn Phe Phe Lys Thr Glu Ile Thr Leu Ala

1040 1045 1050

Asn Gly Glu Ile Arg Lys Arg Pro Leu Ile Glu Thr Asn Gly Glu

1055 1060 1065

Thr Gly Glu Ile Val Trp Asp Lys Gly Arg Asp Phe Ala Thr Val

1070 1075 1080

Arg Lys Val Leu Ser Met Pro Gln Val Asn Ile Val Lys Lys Thr

1085 1090 1095

Glu Val Gln Thr Gly Gly Phe Ser Lys Glu Ser Ile Leu Pro Lys

1100 1105 1110

Arg Asn Ser Asp Lys Leu Ile Ala Arg Lys Lys Asp Trp Asp Pro

1115 1120 1125

Lys Lys Tyr Gly Gly Phe Asp Ser Pro Thr Val Ala Tyr Ser Val

1130 1135 1140

Leu Val Val Ala Lys Val Glu Lys Gly Lys Ser Lys Lys Leu Lys

1145 1150 1155

Ser Val Lys Glu Leu Leu Gly Ile Thr Ile Met Glu Arg Ser Ser

1160 1165 1170

Phe Glu Lys Asn Pro Ile Asp Phe Leu Glu Ala Lys Gly Tyr Lys

1175 1180 1185

Glu Val Lys Lys Asp Leu Ile Ile Lys Leu Pro Lys Tyr Ser Leu

1190 1195 1200

Phe Glu Leu Glu Asn Gly Arg Lys Arg Met Leu Ala Ser Ala Gly

1205 1210 1215

Glu Leu Gln Lys Gly Asn Glu Leu Ala Leu Pro Ser Lys Tyr Val

1220 1225 1230

Asn Phe Leu Tyr Leu Ala Ser His Tyr Glu Lys Leu Lys Gly Ser

1235 1240 1245

Pro Glu Asp Asn Glu Gln Lys Gln Leu Phe Val Glu Gln His Lys

1250 1255 1260

His Tyr Leu Asp Glu Ile Ile Glu Gln Ile Ser Glu Phe Ser Lys

1265 1270 1275

Arg Val Ile Leu Ala Asp Ala Asn Leu Asp Lys Val Leu Ser Ala
1280 1285 1290
Tyr Asn Lys His Arg Asp Lys Pro Ile Arg Glu Gln Ala Glu Asn
1295 1300 1305
Ile Ile His Leu Phe Thr Leu Thr Asn Leu Gly Ala Pro Ala Ala
1310 1315 1320
Phe Lys Tyr Phe Asp Thr Thr Ile Asp Arg Lys Arg Tyr Thr Ser
1325 1330 1335
Thr Lys Glu Val Leu Asp Ala Thr Leu Ile His Gln Ser Ile Thr
1340 1345 1350
Gly Leu Tyr Glu Thr Arg Ile Asp Leu Ser Gln Leu Gly Gly Asp
1355 1360 1365

<210> 6
<211> 1368
<212> PRT
<213> 酿脓链球菌
<400> 6

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

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

485 490 495

Asn Phe Asp Lys Asn Leu Pro Asn Glu Lys Val Leu Pro Lys His Ser

500 505 510

Leu Leu Tyr Glu Tyr Phe Thr Val Tyr Asn Glu Leu Thr Lys Val Lys

515 520 525

Tyr Val Thr Glu Gly Met Arg Lys Pro Ala Phe Leu Ser Gly Glu Gln

530 535 540

Lys Lys Ala Ile Val Asp Leu Leu Phe Lys Thr Asn Arg Lys Val Thr

545 550 555 560

Val Lys Gln Leu Lys Glu Asp Tyr Phe Lys Lys Ile Glu Cys Phe Asp

565 570 575

Ser Val Glu Ile Ser Gly Val Glu Asp Arg Phe Asn Ala Ser Leu Gly

580 585 590

Thr Tyr His Asp Leu Leu Lys Ile Ile Lys Asp Lys Asp Phe Leu Asp

595 600 605

Asn Glu Glu Asn Glu Asp Ile Leu Glu Asp Ile Val Leu Thr Leu Thr

610 615 620

Leu Phe Glu Asp Arg Glu Met Ile Glu Glu Arg Leu Lys Thr Tyr Ala

625 630 635 640

His Leu Phe Asp Asp Lys Val Met Lys Gln Leu Lys Arg Arg Arg Tyr

645 650 655

Thr Gly Trp Gly Arg Leu Ser Arg Lys Leu Ile Asn Gly Ile Arg Asp

660 665 670

Lys Gln Ser Gly Lys Thr Ile Leu Asp Phe Leu Lys Ser Asp Gly Phe

675 680 685

Ala Asn Arg Asn Phe Met Gln Leu Ile His Asp Asp Ser Leu Thr Phe

690 695 700

Lys Glu Asp Ile Gln Lys Ala Gln Val Ser Gly Gln Gly Asp Ser Leu

705 710 715 720

His Glu His Ile Ala Asn Leu Ala Gly Ser Pro Ala Ile Lys Lys Gly

725 730 735

Ile Leu Gln Thr Val Lys Val Val Asp Glu Leu Val Lys Val Met Gly

740 745 750

Arg His Lys Pro Glu Asn Ile Val Ile Glu Met Ala Arg Glu Asn Gln

755 760 765

Thr Thr Gln Lys Gly Gln Lys Asn Ser Arg Glu Arg Met Lys Arg Ile

770 775 780

Glu Glu Gly Ile Lys Glu Leu Gly Ser Gln Ile Leu Lys Glu His Pro

785 790 795 800

Val Glu Asn Thr Gln Leu Gln Asn Glu Lys Leu Tyr Leu Tyr Tyr Leu
805 810 815
Gln Asn Gly Arg Asp Met Tyr Val Asp Gln Glu Leu Asp Ile Asn Arg
820 825 830
Leu Ser Asp Tyr Asp Val Asp His Ile Val Pro Gln Ser Phe Leu Lys
835 840 845
Asp Asp Ser Ile Asp Asn Lys Val Leu Thr Arg Ser Asp Lys Asn Arg
850 855 860
Gly Lys Ser Asp Asn Val Pro Ser Glu Glu Val Val Lys Lys Met Lys
865 870 875 880
Asn Tyr Trp Arg Gln Leu Leu Asn Ala Lys Leu Ile Thr Gln Arg Lys
885 890 895
Phe Asp Asn Leu Thr Lys Ala Glu Arg Gly Gly Leu Ser Glu Leu Asp
900 905 910
Lys Ala Gly Phe Ile Lys Arg Gln Leu Val Glu Thr Arg Gln Ile Thr
915 920 925
Lys His Val Ala Gln Ile Leu Asp Ser Arg Met Asn Thr Lys Tyr Asp
930 935 940
Glu Asn Asp Lys Leu Ile Arg Glu Val Lys Val Ile Thr Leu Lys Ser
945 950 955 960
Lys Leu Val Ser Asp Phe Arg Lys Asp Phe Gln Phe Tyr Lys Val Arg
965 970 975
Glu Ile Asn Asn Tyr His His Ala His Asp Ala Tyr Leu Asn Ala Val
980 985 990
Val Gly Thr Ala Leu Ile Lys Lys Tyr Pro Lys Leu Glu Ser Glu Phe
995 1000 1005
Val Tyr Gly Asp Tyr Lys Val Tyr Asp Val Arg Lys Met Ile Ala
1010 1015 1020
Lys Ser Glu Gln Glu Ile Gly Lys Ala Thr Ala Lys Tyr Phe Phe
1025 1030 1035
Tyr Ser Asn Ile Met Asn Phe Phe Lys Thr Glu Ile Thr Leu Ala
1040 1045 1050
Asn Gly Glu Ile Arg Lys Arg Pro Leu Ile Glu Thr Asn Gly Glu
1055 1060 1065
Thr Gly Glu Ile Val Trp Asp Lys Gly Arg Asp Phe Ala Thr Val
1070 1075 1080
Arg Lys Val Leu Ser Met Pro Gln Val Asn Ile Val Lys Lys Thr
1085 1090 1095
Glu Val Gln Thr Gly Gly Phe Ser Lys Glu Ser Ile Leu Pro Lys

1100 1105 1110

Arg Asn Ser Asp Lys Leu Ile Ala Arg Lys Lys Asp Trp Asp Pro

1115 1120 1125

Lys Lys Tyr Gly Gly Phe Asp Ser Pro Thr Val Ala Tyr Ser Val

1130 1135 1140

Leu Val Val Ala Lys Val Glu Lys Gly Lys Ser Lys Lys Leu Lys

1145 1150 1155

Ser Val Lys Glu Leu Leu Gly Ile Thr Ile Met Glu Arg Ser Ser

1160 1165 1170

Phe Glu Lys Asn Pro Ile Asp Phe Leu Glu Ala Lys Gly Tyr Lys

1175 1180 1185

Glu Val Lys Lys Asp Leu Ile Ile Lys Leu Pro Lys Tyr Ser Leu

1190 1195 1200

Phe Glu Leu Glu Asn Gly Arg Lys Arg Met Leu Ala Ser Ala Gly

1205 1210 1215

Glu Leu Gln Lys Gly Asn Glu Leu Ala Leu Pro Ser Lys Tyr Val

1220 1225 1230

Asn Phe Leu Tyr Leu Ala Ser His Tyr Glu Lys Leu Lys Gly Ser

1235 1240 1245

Pro Glu Asp Asn Glu Gln Lys Gln Leu Phe Val Glu Gln His Lys

1250 1255 1260

His Tyr Leu Asp Glu Ile Ile Glu Gln Ile Ser Glu Phe Ser Lys

1265 1270 1275

Arg Val Ile Leu Ala Asp Ala Asn Leu Asp Lys Val Leu Ser Ala

1280 1285 1290

Tyr Asn Lys His Arg Asp Lys Pro Ile Arg Glu Gln Ala Glu Asn

1295 1300 1305

Ile Ile His Leu Phe Thr Leu Thr Asn Leu Gly Ala Pro Ala Ala

1310 1315 1320

Phe Lys Tyr Phe Asp Thr Thr Ile Asp Arg Lys Arg Tyr Thr Ser

1325 1330 1335

Thr Lys Glu Val Leu Asp Ala Thr Leu Ile His Gln Ser Ile Thr

1340 1345 1350

Gly Leu Tyr Glu Thr Arg Ile Asp Leu Ser Gln Leu Gly Gly Asp

1355 1360 1365

<210> 7

<211> 1368

<212> PRT

<213> 人工序列

<220>

<223> 合成多肽

<400> 7

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

Leu Phe Leu Ala Ala Lys Asn Leu Ser Asp Ala Ile Leu Leu Ser Asp
290 295 300
Ile Leu Arg Val Asn Thr Glu Ile Thr Lys Ala Pro Leu Ser Ala Ser
305 310 315 320
Met Ile Lys Arg Tyr Asp Glu His His Gln Asp Leu Thr Leu Leu Lys
325 330 335
Ala Leu Val Arg Gln Gln Leu Pro Glu Lys Tyr Lys Glu Ile Phe Phe
340 345 350
Asp Gln Ser Lys Asn Gly Tyr Ala Gly Tyr Ile Asp Gly Gly Ala Ser
355 360 365
Gln Glu Glu Phe Tyr Lys Phe Ile Lys Pro Ile Leu Glu Lys Met Asp
370 375 380
Gly Thr Glu Glu Leu Leu Val Lys Leu Asn Arg Glu Asp Leu Leu Arg
385 390 395 400
Lys Gln Arg Thr Phe Asp Asn Gly Ser Ile Pro His Gln Ile His Leu
405 410 415
Gly Glu Leu His Ala Ile Leu Arg Arg Gln Glu Asp Phe Tyr Pro Phe
420 425 430
Leu Lys Asp Asn Arg Glu Lys Ile Glu Lys Ile Leu Thr Phe Arg Ile
435 440 445
Pro Tyr Tyr Val Gly Pro Leu Ala Arg Gly Asn Ser Arg Phe Ala Trp
450 455 460
Met Thr Arg Lys Ser Glu Glu Thr Ile Thr Pro Trp Asn Phe Glu Glu
465 470 475 480
Val Val Asp Lys Gly Ala Ser Ala Gln Ser Phe Ile Glu Arg Met Thr
485 490 495
Asn Phe Asp Lys Asn Leu Pro Asn Glu Lys Val Leu Pro Lys His Ser
500 505 510
Leu Leu Tyr Glu Tyr Phe Thr Val Tyr Asn Glu Leu Thr Lys Val Lys
515 520 525
Tyr Val Thr Glu Gly Met Arg Lys Pro Ala Phe Leu Ser Gly Glu Gln
530 535 540
Lys Lys Ala Ile Val Asp Leu Leu Phe Lys Thr Asn Arg Lys Val Thr
545 550 555 560
Val Lys Gln Leu Lys Glu Asp Tyr Phe Lys Lys Ile Glu Cys Phe Asp
565 570 575
Ser Val Glu Ile Ser Gly Val Glu Asp Arg Phe Asn Ala Ser Leu Gly
580 585 590
Thr Tyr His Asp Leu Leu Lys Ile Ile Lys Asp Lys Asp Phe Leu Asp

595 600 605

Asn Glu Glu Asn Glu Asp Ile Leu Glu Asp Ile Val Leu Thr Leu Thr

610 615 620

Leu Phe Glu Asp Arg Glu Met Ile Glu Glu Arg Leu Lys Thr Tyr Ala

625 630 635 640

His Leu Phe Asp Asp Lys Val Met Lys Gln Leu Lys Arg Arg Arg Tyr

645 650 655

Thr Gly Trp Gly Arg Leu Ser Arg Lys Leu Ile Asn Gly Ile Arg Asp

660 665 670

Lys Gln Ser Gly Lys Thr Ile Leu Asp Phe Leu Lys Ser Asp Gly Phe

675 680 685

Ala Asn Arg Asn Phe Met Gln Leu Ile His Asp Asp Ser Leu Thr Phe

690 695 700

Lys Glu Asp Ile Gln Lys Ala Gln Val Ser Gly Gln Gly Asp Ser Leu

705 710 715 720

His Glu His Ile Ala Asn Leu Ala Gly Ser Pro Ala Ile Lys Lys Gly

725 730 735

Ile Leu Gln Thr Val Lys Val Val Asp Glu Leu Val Lys Val Met Gly

740 745 750

Arg His Lys Pro Glu Asn Ile Val Ile Glu Met Ala Arg Glu Asn Gln

755 760 765

Thr Thr Gln Lys Gly Gln Lys Asn Ser Arg Glu Arg Met Lys Arg Ile

770 775 780

Glu Glu Gly Ile Lys Glu Leu Gly Ser Gln Ile Leu Lys Glu His Pro

785 790 795 800

Val Glu Asn Thr Gln Leu Gln Asn Glu Lys Leu Tyr Leu Tyr Tyr Leu

805 810 815

Gln Asn Gly Arg Asp Met Tyr Val Asp Gln Glu Leu Asp Ile Asn Arg

820 825 830

Leu Ser Asp Tyr Asp Val Asp Ala Ile Val Pro Gln Ser Phe Leu Lys

835 840 845

Asp Asp Ser Ile Asp Asn Lys Val Leu Thr Arg Ser Asp Lys Asn Arg

850 855 860

Gly Lys Ser Asp Asn Val Pro Ser Glu Glu Val Val Lys Lys Met Lys

865 870 875 880

Asn Tyr Trp Arg Gln Leu Leu Asn Ala Lys Leu Ile Thr Gln Arg Lys

885 890 895

Phe Asp Asn Leu Thr Lys Ala Glu Arg Gly Gly Leu Ser Glu Leu Asp

900 905 910

Lys Ala Gly Phe Ile Lys Arg Gln Leu Val Glu Thr Arg Gln Ile Thr
915 920 925
Lys His Val Ala Gln Ile Leu Asp Ser Arg Met Asn Thr Lys Tyr Asp
930 935 940
Glu Asn Asp Lys Leu Ile Arg Glu Val Lys Val Ile Thr Leu Lys Ser
945 950 955 960
Lys Leu Val Ser Asp Phe Arg Lys Asp Phe Gln Phe Tyr Lys Val Arg
965 970 975
Glu Ile Asn Asn Tyr His His Ala His Asp Ala Tyr Leu Asn Ala Val
980 985 990
Val Gly Thr Ala Leu Ile Lys Lys Tyr Pro Lys Leu Glu Ser Glu Phe
995 1000 1005
Val Tyr Gly Asp Tyr Lys Val Tyr Asp Val Arg Lys Met Ile Ala
1010 1015 1020
Lys Ser Glu Gln Glu Ile Gly Lys Ala Thr Ala Lys Tyr Phe Phe
1025 1030 1035
Tyr Ser Asn Ile Met Asn Phe Phe Lys Thr Glu Ile Thr Leu Ala
1040 1045 1050
Asn Gly Glu Ile Arg Lys Arg Pro Leu Ile Glu Thr Asn Gly Glu
1055 1060 1065
Thr Gly Glu Ile Val Trp Asp Lys Gly Arg Asp Phe Ala Thr Val
1070 1075 1080
Arg Lys Val Leu Ser Met Pro Gln Val Asn Ile Val Lys Lys Thr
1085 1090 1095
Glu Val Gln Thr Gly Gly Phe Ser Lys Glu Ser Ile Leu Pro Lys
1100 1105 1110
Arg Asn Ser Asp Lys Leu Ile Ala Arg Lys Lys Asp Trp Asp Pro
1115 1120 1125
Lys Lys Tyr Gly Gly Phe Asp Ser Pro Thr Val Ala Tyr Ser Val
1130 1135 1140
Leu Val Val Ala Lys Val Glu Lys Gly Lys Ser Lys Lys Leu Lys
1145 1150 1155
Ser Val Lys Glu Leu Leu Gly Ile Thr Ile Met Glu Arg Ser Ser
1160 1165 1170
Phe Glu Lys Asn Pro Ile Asp Phe Leu Glu Ala Lys Gly Tyr Lys
1175 1180 1185
Glu Val Lys Lys Asp Leu Ile Ile Lys Leu Pro Lys Tyr Ser Leu
1190 1195 1200
Phe Glu Leu Glu Asn Gly Arg Lys Arg Met Leu Ala Ser Ala Gly

1205 1210 1215
Glu Leu Gln Lys Gly Asn Glu Leu Ala Leu Pro Ser Lys Tyr Val
1220 1225 1230
Asn Phe Leu Tyr Leu Ala Ser His Tyr Glu Lys Leu Lys Gly Ser
1235 1240 1245
Pro Glu Asp Asn Glu Gln Lys Gln Leu Phe Val Glu Gln His Lys
1250 1255 1260
His Tyr Leu Asp Glu Ile Ile Glu Gln Ile Ser Glu Phe Ser Lys
1265 1270 1275
Arg Val Ile Leu Ala Asp Ala Asn Leu Asp Lys Val Leu Ser Ala
1280 1285 1290
Tyr Asn Lys His Arg Asp Lys Pro Ile Arg Glu Gln Ala Glu Asn
1295 1300 1305
Ile Ile His Leu Phe Thr Leu Thr Asn Leu Gly Ala Pro Ala Ala
1310 1315 1320
Phe Lys Tyr Phe Asp Thr Thr Ile Asp Arg Lys Arg Tyr Thr Ser
1325 1330 1335
Thr Lys Glu Val Leu Asp Ala Thr Leu Ile His Gln Ser Ile Thr
1340 1345 1350
Gly Leu Tyr Glu Thr Arg Ile Asp Leu Ser Gln Leu Gly Gly Asp
1355 1360 1365

<210> 8
<211> 1367
<212> PRT
<213> 人工序列
<220>
<223> 合成多肽
<400> 8

Asp Lys Lys Tyr Ser Ile Gly Leu Ala Ile Gly Thr Asn Ser Val Gly
1 5 10 15
Trp Ala Val Ile Thr Asp Glu Tyr Lys Val Pro Ser Lys Lys Phe Lys
20 25 30
Val Leu Gly Asn Thr Asp Arg His Ser Ile Lys Lys Asn Leu Ile Gly
35 40 45
Ala Leu Leu Phe Asp Ser Gly Glu Thr Ala Glu Ala Thr Arg Leu Lys
50 55 60
Arg Thr Ala Arg Arg Arg Tyr Thr Arg Arg Lys Asn Arg Ile Cys Tyr
65 70 75 80
Leu Gln Glu Ile Phe Ser Asn Glu Met Ala Lys Val Asp Asp Ser Phe

85 90 95

Phe His Arg Leu Glu Glu Ser Phe Leu Val Glu Glu Asp Lys Lys His
100 105 110

Glu Arg His Pro Ile Phe Gly Asn Ile Val Asp Glu Val Ala Tyr His
115 120 125

Glu Lys Tyr Pro Thr Ile Tyr His Leu Arg Lys Lys Leu Val Asp Ser
130 135 140

Thr Asp Lys Ala Asp Leu Arg Leu Ile Tyr Leu Ala Leu Ala His Met
145 150 155 160

Ile Lys Phe Arg Gly His Phe Leu Ile Glu Gly Asp Leu Asn Pro Asp
165 170 175

Asn Ser Asp Val Asp Lys Leu Phe Ile Gln Leu Val Gln Thr Tyr Asn
180 185 190

Gln Leu Phe Glu Glu Asn Pro Ile Asn Ala Ser Gly Val Asp Ala Lys
195 200 205

Ala Ile Leu Ser Ala Arg Leu Ser Lys Ser Arg Arg Leu Glu Asn Leu
210 215 220

Ile Ala Gln Leu Pro Gly Glu Lys Lys Asn Gly Leu Phe Gly Asn Leu
225 230 235 240

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

Leu Ala Glu Asp Ala Lys Leu Gln Leu Ser Lys Asp Thr Tyr Asp Asp
260 265 270

Asp Leu Asp Asn Leu Leu Ala Gln Ile Gly Asp Gln Tyr Ala Asp Leu
275 280 285

Phe Leu Ala Ala Lys Asn Leu Ser Asp Ala Ile Leu Leu Ser Asp Ile
290 295 300

Leu Arg Val Asn Thr Glu Ile Thr Lys Ala Pro Leu Ser Ala Ser Met
305 310 315 320

Ile Lys Arg Tyr Asp Glu His His Gln Asp Leu Thr Leu Leu Lys Ala
325 330 335

Leu Val Arg Gln Gln Leu Pro Glu Lys Tyr Lys Glu Ile Phe Phe Asp
340 345 350

Gln Ser Lys Asn Gly Tyr Ala Gly Tyr Ile Asp Gly Gly Ala Ser Gln
355 360 365

Glu Glu Phe Tyr Lys Phe Ile Lys Pro Ile Leu Glu Lys Met Asp Gly
370 375 380

Thr Glu Glu Leu Leu Val Lys Leu Asn Arg Glu Asp Leu Leu Arg Lys
385 390 395 400

Gln Arg Thr Phe Asp Asn Gly Ser Ile Pro His Gln Ile His Leu Gly
405 410 415
Glu Leu His Ala Ile Leu Arg Arg Gln Glu Asp Phe Tyr Pro Phe Leu
420 425 430
Lys Asp Asn Arg Glu Lys Ile Glu Lys Ile Leu Thr Phe Arg Ile Pro
435 440 445
Tyr Tyr Val Gly Pro Leu Ala Arg Gly Asn Ser Arg Phe Ala Trp Met
450 455 460
Thr Arg Lys Ser Glu Glu Thr Ile Thr Pro Trp Asn Phe Glu Glu Val
465 470 475 480
Val Asp Lys Gly Ala Ser Ala Gln Ser Phe Ile Glu Arg Met Thr Asn
485 490 495
Phe Asp Lys Asn Leu Pro Asn Glu Lys Val Leu Pro Lys His Ser Leu
500 505 510
Leu Tyr Glu Tyr Phe Thr Val Tyr Asn Glu Leu Thr Lys Val Lys Tyr
515 520 525
Val Thr Glu Gly Met Arg Lys Pro Ala Phe Leu Ser Gly Glu Gln Lys
530 535 540
Lys Ala Ile Val Asp Leu Leu Phe Lys Thr Asn Arg Lys Val Thr Val
545 550 555 560
Lys Gln Leu Lys Glu Asp Tyr Phe Lys Lys Ile Glu Cys Phe Asp Ser
565 570 575
Val Glu Ile Ser Gly Val Glu Asp Arg Phe Asn Ala Ser Leu Gly Thr
580 585 590
Tyr His Asp Leu Leu Lys Ile Ile Lys Asp Lys Asp Phe Leu Asp Asn
595 600 605
Glu Glu Asn Glu Asp Ile Leu Glu Asp Ile Val Leu Thr Leu Thr Leu
610 615 620
Phe Glu Asp Arg Glu Met Ile Glu Glu Arg Leu Lys Thr Tyr Ala His
625 630 635 640
Leu Phe Asp Asp Lys Val Met Lys Gln Leu Lys Arg Arg Arg Tyr Thr
645 650 655
Gly Trp Gly Arg Leu Ser Arg Lys Leu Ile Asn Gly Ile Arg Asp Lys
660 665 670
Gln Ser Gly Lys Thr Ile Leu Asp Phe Leu Lys Ser Asp Gly Phe Ala
675 680 685
Asn Arg Asn Phe Met Gln Leu Ile His Asp Asp Ser Leu Thr Phe Lys
690 695 700
Glu Asp Ile Gln Lys Ala Gln Val Ser Gly Gln Gly Asp Ser Leu His

705 710 715 720
Glu His Ile Ala Asn Leu Ala Gly Ser Pro Ala Ile Lys Lys Gly Ile
725 730 735
Leu Gln Thr Val Lys Val Val Asp Glu Leu Val Lys Val Met Gly Arg
740 745 750
His Lys Pro Glu Asn Ile Val Ile Glu Met Ala Arg Glu Asn Gln Thr
755 760 765
Thr Gln Lys Gly Gln Lys Asn Ser Arg Glu Arg Met Lys Arg Ile Glu
770 775 780
Glu Gly Ile Lys Glu Leu Gly Ser Gln Ile Leu Lys Glu His Pro Val
785 790 795 800
Glu Asn Thr Gln Leu Gln Asn Glu Lys Leu Tyr Leu Tyr Tyr Leu Gln
805 810 815
Asn Gly Arg Asp Met Tyr Val Asp Gln Glu Leu Asp Ile Asn Arg Leu
820 825 830
Ser Asp Tyr Asp Val Asp Ala Ile Val Pro Gln Ser Phe Leu Lys Asp
835 840 845
Asp Ser Ile Asp Asn Lys Val Leu Thr Arg Ser Asp Lys Asn Arg Gly
850 855 860
Lys Ser Asp Asn Val Pro Ser Glu Glu Val Val Lys Lys Met Lys Asn
865 870 875 880
Tyr Trp Arg Gln Leu Leu Asn Ala Lys Leu Ile Thr Gln Arg Lys Phe
885 890 895
Asp Asn Leu Thr Lys Ala Glu Arg Gly Gly Leu Ser Glu Leu Asp Lys
900 905 910
Ala Gly Phe Ile Lys Arg Gln Leu Val Glu Thr Arg Gln Ile Thr Lys
915 920 925
His Val Ala Gln Ile Leu Asp Ser Arg Met Asn Thr Lys Tyr Asp Glu
930 935 940
Asn Asp Lys Leu Ile Arg Glu Val Lys Val Ile Thr Leu Lys Ser Lys
945 950 955 960
Leu Val Ser Asp Phe Arg Lys Asp Phe Gln Phe Tyr Lys Val Arg Glu
965 970 975
Ile Asn Asn Tyr His His Ala His Asp Ala Tyr Leu Asn Ala Val Val
980 985 990
Gly Thr Ala Leu Ile Lys Lys Tyr Pro Lys Leu Glu Ser Glu Phe Val
995 1000 1005
Tyr Gly Asp Tyr Lys Val Tyr Asp Val Arg Lys Met Ile Ala Lys
1010 1015 1020

Ser Glu Gln Glu Ile Gly Lys Ala Thr Ala Lys Tyr Phe Phe Tyr
1025 1030 1035
Ser Asn Ile Met Asn Phe Phe Lys Thr Glu Ile Thr Leu Ala Asn
1040 1045 1050
Gly Glu Ile Arg Lys Arg Pro Leu Ile Glu Thr Asn Gly Glu Thr
1055 1060 1065
Gly Glu Ile Val Trp Asp Lys Gly Arg Asp Phe Ala Thr Val Arg
1070 1075 1080
Lys Val Leu Ser Met Pro Gln Val Asn Ile Val Lys Lys Thr Glu
1085 1090 1095
Val Gln Thr Gly Gly Phe Ser Lys Glu Ser Ile Leu Pro Lys Arg
1100 1105 1110
Asn Ser Asp Lys Leu Ile Ala Arg Lys Lys Asp Trp Asp Pro Lys
1115 1120 1125
Lys Tyr Gly Gly Phe Asp Ser Pro Thr Val Ala Tyr Ser Val Leu
1130 1135 1140
Val Val Ala Lys Val Glu Lys Gly Lys Ser Lys Lys Leu Lys Ser
1145 1150 1155
Val Lys Glu Leu Leu Gly Ile Thr Ile Met Glu Arg Ser Ser Phe
1160 1165 1170
Glu Lys Asn Pro Ile Asp Phe Leu Glu Ala Lys Gly Tyr Lys Glu
1175 1180 1185
Val Lys Lys Asp Leu Ile Ile Lys Leu Pro Lys Tyr Ser Leu Phe
1190 1195 1200
Glu Leu Glu Asn Gly Arg Lys Arg Met Leu Ala Ser Ala Gly Glu
1205 1210 1215
Leu Gln Lys Gly Asn Glu Leu Ala Leu Pro Ser Lys Tyr Val Asn
1220 1225 1230
Phe Leu Tyr Leu Ala Ser His Tyr Glu Lys Leu Lys Gly Ser Pro
1235 1240 1245
Glu Asp Asn Glu Gln Lys Gln Leu Phe Val Glu Gln His Lys His
1250 1255 1260
Tyr Leu Asp Glu Ile Ile Glu Gln Ile Ser Glu Phe Ser Lys Arg
1265 1270 1275
Val Ile Leu Ala Asp Ala Asn Leu Asp Lys Val Leu Ser Ala Tyr
1280 1285 1290
Asn Lys His Arg Asp Lys Pro Ile Arg Glu Gln Ala Glu Asn Ile
1295 1300 1305
Ile His Leu Phe Thr Leu Thr Asn Leu Gly Ala Pro Ala Ala Phe

1310 1315 1320
Lys Tyr Phe Asp Thr Thr Ile Asp Arg Lys Arg Tyr Thr Ser Thr
1325 1330 1335
Lys Glu Val Leu Asp Ala Thr Leu Ile His Gln Ser Ile Thr Gly
1340 1345 1350
Leu Tyr Glu Thr Arg Ile Asp Leu Ser Gln Leu Gly Gly Asp
1355 1360 1365

<210> 9
<211> 1368
<212> PRT
<213> 人工序列
<220>
<223> 合成多肽
<400> 9

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

195 200 205

Lys Ala Ile Leu Ser Ala Arg Leu Ser Lys Ser Arg Arg Leu Glu Asn

210 215 220

Leu Ile Ala Gln Leu Pro Gly Glu Lys Lys Asn Gly Leu Phe Gly Asn

225 230 235 240

Leu Ile Ala Leu Ser Leu Gly Leu Thr Pro Asn Phe Lys Ser Asn Phe

245 250 255

Asp Leu Ala Glu Asp Ala Lys Leu Gln Leu Ser Lys Asp Thr Tyr Asp

260 265 270

Asp Asp Leu Asp Asn Leu Leu Ala Gln Ile Gly Asp Gln Tyr Ala Asp

275 280 285

Leu Phe Leu Ala Ala Lys Asn Leu Ser Asp Ala Ile Leu Leu Ser Asp

290 295 300

Ile Leu Arg Val Asn Thr Glu Ile Thr Lys Ala Pro Leu Ser Ala Ser

305 310 315 320

Met Ile Lys Arg Tyr Asp Glu His His Gln Asp Leu Thr Leu Leu Lys

325 330 335

Ala Leu Val Arg Gln Gln Leu Pro Glu Lys Tyr Lys Glu Ile Phe Phe

340 345 350

Asp Gln Ser Lys Asn Gly Tyr Ala Gly Tyr Ile Asp Gly Gly Ala Ser

355 360 365

Gln Glu Glu Phe Tyr Lys Phe Ile Lys Pro Ile Leu Glu Lys Met Asp

370 375 380

Gly Thr Glu Glu Leu Leu Val Lys Leu Asn Arg Glu Asp Leu Leu Arg

385 390 395 400

Lys Gln Arg Thr Phe Asp Asn Gly Ser Ile Pro His Gln Ile His Leu

405 410 415

Gly Glu Leu His Ala Ile Leu Arg Arg Gln Glu Asp Phe Tyr Pro Phe

420 425 430

Leu Lys Asp Asn Arg Glu Lys Ile Glu Lys Ile Leu Thr Phe Arg Ile

435 440 445

Pro Tyr Tyr Val Gly Pro Leu Ala Arg Gly Asn Ser Arg Phe Ala Trp

450 455 460

Met Thr Arg Lys Ser Glu Glu Thr Ile Thr Pro Trp Asn Phe Glu Glu

465 470 475 480

Val Val Asp Lys Gly Ala Ser Ala Gln Ser Phe Ile Glu Arg Met Thr

485 490 495

Asn Phe Asp Lys Asn Leu Pro Asn Glu Lys Val Leu Pro Lys His Ser

500 505 510

Leu Leu Tyr Glu Tyr Phe Thr Val Tyr Asn Glu Leu Thr Lys Val Lys
515 520 525
Tyr Val Thr Glu Gly Met Arg Lys Pro Ala Phe Leu Ser Gly Glu Gln
530 535 540
Lys Lys Ala Ile Val Asp Leu Leu Phe Lys Thr Asn Arg Lys Val Thr
545 550 555 560
Val Lys Gln Leu Lys Glu Asp Tyr Phe Lys Lys Ile Glu Cys Phe Asp
565 570 575
Ser Val Glu Ile Ser Gly Val Glu Asp Arg Phe Asn Ala Ser Leu Gly
580 585 590
Thr Tyr His Asp Leu Leu Lys Ile Ile Lys Asp Lys Asp Phe Leu Asp
595 600 605
Asn Glu Glu Asn Glu Asp Ile Leu Glu Asp Ile Val Leu Thr Leu Thr
610 615 620
Leu Phe Glu Asp Arg Glu Met Ile Glu Glu Arg Leu Lys Thr Tyr Ala
625 630 635 640
His Leu Phe Asp Asp Lys Val Met Lys Gln Leu Lys Arg Arg Arg Tyr
645 650 655
Thr Gly Trp Gly Arg Leu Ser Arg Lys Leu Ile Asn Gly Ile Arg Asp
660 665 670
Lys Gln Ser Gly Lys Thr Ile Leu Asp Phe Leu Lys Ser Asp Gly Phe
675 680 685
Ala Asn Arg Asn Phe Met Gln Leu Ile His Asp Asp Ser Leu Thr Phe
690 695 700
Lys Glu Asp Ile Gln Lys Ala Gln Val Ser Gly Gln Gly Asp Ser Leu
705 710 715 720
His Glu His Ile Ala Asn Leu Ala Gly Ser Pro Ala Ile Lys Lys Gly
725 730 735
Ile Leu Gln Thr Val Lys Val Val Asp Glu Leu Val Lys Val Met Gly
740 745 750
Arg His Lys Pro Glu Asn Ile Val Ile Glu Met Ala Arg Glu Asn Gln
755 760 765
Thr Thr Gln Lys Gly Gln Lys Asn Ser Arg Glu Arg Met Lys Arg Ile
770 775 780
Glu Glu Gly Ile Lys Glu Leu Gly Ser Gln Ile Leu Lys Glu His Pro
785 790 795 800
Val Glu Asn Thr Gln Leu Gln Asn Glu Lys Leu Tyr Leu Tyr Tyr Leu
805 810 815
Gln Asn Gly Arg Asp Met Tyr Val Asp Gln Glu Leu Asp Ile Asn Arg

820 825 830

Leu Ser Asp Tyr Asp Val Asp Ala Ile Val Pro Gln Ser Phe Leu Lys

835 840 845

Asp Asp Ser Ile Asp Asn Lys Val Leu Thr Arg Ser Asp Lys Asn Arg

850 855 860

Gly Lys Ser Asp Asn Val Pro Ser Glu Glu Val Val Lys Lys Met Lys

865 870 875 880

Asn Tyr Trp Arg Gln Leu Leu Asn Ala Lys Leu Ile Thr Gln Arg Lys

885 890 895

Phe Asp Asn Leu Thr Lys Ala Glu Arg Gly Gly Leu Ser Glu Leu Asp

900 905 910

Lys Ala Gly Phe Ile Lys Arg Gln Leu Val Glu Thr Arg Gln Ile Thr

915 920 925

Lys His Val Ala Gln Ile Leu Asp Ser Arg Met Asn Thr Lys Tyr Asp

930 935 940

Glu Asn Asp Lys Leu Ile Arg Glu Val Lys Val Ile Thr Leu Lys Ser

945 950 955 960

Lys Leu Val Ser Asp Phe Arg Lys Asp Phe Gln Phe Tyr Lys Val Arg

965 970 975

Glu Ile Asn Asn Tyr His His Ala His Asp Ala Tyr Leu Asn Ala Val

980 985 990

Val Gly Thr Ala Leu Ile Lys Lys Tyr Pro Lys Leu Glu Ser Glu Phe

995 1000 1005

Val Tyr Gly Asp Tyr Lys Val Tyr Asp Val Arg Lys Met Ile Ala

1010 1015 1020

Lys Ser Glu Gln Glu Ile Gly Lys Ala Thr Ala Lys Tyr Phe Phe

1025 1030 1035

Tyr Ser Asn Ile Met Asn Phe Phe Lys Thr Glu Ile Thr Leu Ala

1040 1045 1050

Asn Gly Glu Ile Arg Lys Arg Pro Leu Ile Glu Thr Asn Gly Glu

1055 1060 1065

Thr Gly Glu Ile Val Trp Asp Lys Gly Arg Asp Phe Ala Thr Val

1070 1075 1080

Arg Lys Val Leu Ser Met Pro Gln Val Asn Ile Val Lys Lys Thr

1085 1090 1095

Glu Val Gln Thr Gly Gly Phe Ser Lys Glu Ser Ile Leu Pro Lys

1100 1105 1110

Arg Asn Ser Asp Lys Leu Ile Ala Arg Lys Lys Asp Trp Asp Pro

1115 1120 1125

Lys Lys Tyr Gly Gly Phe Asp Ser Pro Thr Val Ala Tyr Ser Val
1130 1135 1140
Leu Val Val Ala Lys Val Glu Lys Gly Lys Ser Lys Lys Leu Lys
1145 1150 1155
Ser Val Lys Glu Leu Leu Gly Ile Thr Ile Met Glu Arg Ser Ser
1160 1165 1170
Phe Glu Lys Asn Pro Ile Asp Phe Leu Glu Ala Lys Gly Tyr Lys
1175 1180 1185
Glu Val Lys Lys Asp Leu Ile Ile Lys Leu Pro Lys Tyr Ser Leu
1190 1195 1200
Phe Glu Leu Glu Asn Gly Arg Lys Arg Met Leu Ala Ser Ala Gly
1205 1210 1215
Glu Leu Gln Lys Gly Asn Glu Leu Ala Leu Pro Ser Lys Tyr Val
1220 1225 1230
Asn Phe Leu Tyr Leu Ala Ser His Tyr Glu Lys Leu Lys Gly Ser
1235 1240 1245
Pro Glu Asp Asn Glu Gln Lys Gln Leu Phe Val Glu Gln His Lys
1250 1255 1260
His Tyr Leu Asp Glu Ile Ile Glu Gln Ile Ser Glu Phe Ser Lys
1265 1270 1275
Arg Val Ile Leu Ala Asp Ala Asn Leu Asp Lys Val Leu Ser Ala
1280 1285 1290
Tyr Asn Lys His Arg Asp Lys Pro Ile Arg Glu Gln Ala Glu Asn
1295 1300 1305
Ile Ile His Leu Phe Thr Leu Thr Asn Leu Gly Ala Pro Ala Ala
1310 1315 1320
Phe Lys Tyr Phe Asp Thr Thr Ile Asp Arg Lys Arg Tyr Thr Ser
1325 1330 1335
Thr Lys Glu Val Leu Asp Ala Thr Leu Ile His Gln Ser Ile Thr
1340 1345 1350
Gly Leu Tyr Glu Thr Arg Ile Asp Leu Ser Gln Leu Gly Gly Asp
1355 1360 1365

<210> 10
<211> 1367
<212> PRT
<213> 人工序列
<220>
<223> 合成多肽
<400> 10

Asp Lys Lys Tyr Ser Ile Gly Leu Ala Ile Gly Thr Asn Ser Val Gly
1 5 10 15
Trp Ala Val Ile Thr Asp Glu Tyr Lys Val Pro Ser Lys Lys Phe Lys
20 25 30
Val Leu Gly Asn Thr Asp Arg His Ser Ile Lys Lys Asn Leu Ile Gly
35 40 45
Ala Leu Leu Phe Asp Ser Gly Glu Thr Ala Glu Ala Thr Arg Leu Lys
50 55 60
Arg Thr Ala Arg Arg Arg Tyr Thr Arg Arg Lys Asn Arg Ile Cys Tyr
65 70 75 80
Leu Gln Glu Ile Phe Ser Asn Glu Met Ala Lys Val Asp Asp Ser Phe
85 90 95
Phe His Arg Leu Glu Glu Ser Phe Leu Val Glu Glu Asp Lys Lys His
100 105 110
Glu Arg His Pro Ile Phe Gly Asn Ile Val Asp Glu Val Ala Tyr His
115 120 125
Glu Lys Tyr Pro Thr Ile Tyr His Leu Arg Lys Lys Leu Val Asp Ser
130 135 140
Thr Asp Lys Ala Asp Leu Arg Leu Ile Tyr Leu Ala Leu Ala His Met
145 150 155 160
Ile Lys Phe Arg Gly His Phe Leu Ile Glu Gly Asp Leu Asn Pro Asp
165 170 175
Asn Ser Asp Val Asp Lys Leu Phe Ile Gln Leu Val Gln Thr Tyr Asn
180 185 190
Gln Leu Phe Glu Glu Asn Pro Ile Asn Ala Ser Gly Val Asp Ala Lys
195 200 205
Ala Ile Leu Ser Ala Arg Leu Ser Lys Ser Arg Arg Leu Glu Asn Leu
210 215 220
Ile Ala Gln Leu Pro Gly Glu Lys Lys Asn Gly Leu Phe Gly Asn Leu
225 230 235 240
Ile Ala Leu Ser Leu Gly Leu Thr Pro Asn Phe Lys Ser Asn Phe Asp
245 250 255
Leu Ala Glu Asp Ala Lys Leu Gln Leu Ser Lys Asp Thr Tyr Asp Asp
260 265 270
Asp Leu Asp Asn Leu Leu Ala Gln Ile Gly Asp Gln Tyr Ala Asp Leu
275 280 285
Phe Leu Ala Ala Lys Asn Leu Ser Asp Ala Ile Leu Leu Ser Asp Ile
290 295 300
Leu Arg Val Asn Thr Glu Ile Thr Lys Ala Pro Leu Ser Ala Ser Met

305 310 315 320
Ile Lys Arg Tyr Asp Glu His His Gln Asp Leu Thr Leu Leu Lys Ala
325 330 335
Leu Val Arg Gln Gln Leu Pro Glu Lys Tyr Lys Glu Ile Phe Phe Asp
340 345 350
Gln Ser Lys Asn Gly Tyr Ala Gly Tyr Ile Asp Gly Gly Ala Ser Gln
355 360 365
Glu Glu Phe Tyr Lys Phe Ile Lys Pro Ile Leu Glu Lys Met Asp Gly
370 375 380
Thr Glu Glu Leu Leu Val Lys Leu Asn Arg Glu Asp Leu Leu Arg Lys
385 390 395 400
Gln Arg Thr Phe Asp Asn Gly Ser Ile Pro His Gln Ile His Leu Gly
405 410 415
Glu Leu His Ala Ile Leu Arg Arg Gln Glu Asp Phe Tyr Pro Phe Leu
420 425 430
Lys Asp Asn Arg Glu Lys Ile Glu Lys Ile Leu Thr Phe Arg Ile Pro
435 440 445
Tyr Tyr Val Gly Pro Leu Ala Arg Gly Asn Ser Arg Phe Ala Trp Met
450 455 460
Thr Arg Lys Ser Glu Glu Thr Ile Thr Pro Trp Asn Phe Glu Glu Val
465 470 475 480
Val Asp Lys Gly Ala Ser Ala Gln Ser Phe Ile Glu Arg Met Thr Asn
485 490 495
Phe Asp Lys Asn Leu Pro Asn Glu Lys Val Leu Pro Lys His Ser Leu
500 505 510
Leu Tyr Glu Tyr Phe Thr Val Tyr Asn Glu Leu Thr Lys Val Lys Tyr
515 520 525
Val Thr Glu Gly Met Arg Lys Pro Ala Phe Leu Ser Gly Glu Gln Lys
530 535 540
Lys Ala Ile Val Asp Leu Leu Phe Lys Thr Asn Arg Lys Val Thr Val
545 550 555 560
Lys Gln Leu Lys Glu Asp Tyr Phe Lys Lys Ile Glu Cys Phe Asp Ser
565 570 575
Val Glu Ile Ser Gly Val Glu Asp Arg Phe Asn Ala Ser Leu Gly Thr
580 585 590
Tyr His Asp Leu Leu Lys Ile Ile Lys Asp Lys Asp Phe Leu Asp Asn
595 600 605
Glu Glu Asn Glu Asp Ile Leu Glu Asp Ile Val Leu Thr Leu Thr Leu
610 615 620

Phe Glu Asp Arg Glu Met Ile Glu Glu Arg Leu Lys Thr Tyr Ala His
625 630 635 640
Leu Phe Asp Asp Lys Val Met Lys Gln Leu Lys Arg Arg Arg Tyr Thr
645 650 655
Gly Trp Gly Arg Leu Ser Arg Lys Leu Ile Asn Gly Ile Arg Asp Lys
660 665 670
Gln Ser Gly Lys Thr Ile Leu Asp Phe Leu Lys Ser Asp Gly Phe Ala
675 680 685
Asn Arg Asn Phe Met Gln Leu Ile His Asp Asp Ser Leu Thr Phe Lys
690 695 700
Glu Asp Ile Gln Lys Ala Gln Val Ser Gly Gln Gly Asp Ser Leu His
705 710 715 720
Glu His Ile Ala Asn Leu Ala Gly Ser Pro Ala Ile Lys Lys Gly Ile
725 730 735
Leu Gln Thr Val Lys Val Val Asp Glu Leu Val Lys Val Met Gly Arg
740 745 750
His Lys Pro Glu Asn Ile Val Ile Glu Met Ala Arg Glu Asn Gln Thr
755 760 765
Thr Gln Lys Gly Gln Lys Asn Ser Arg Glu Arg Met Lys Arg Ile Glu
770 775 780
Glu Gly Ile Lys Glu Leu Gly Ser Gln Ile Leu Lys Glu His Pro Val
785 790 795 800
Glu Asn Thr Gln Leu Gln Asn Glu Lys Leu Tyr Leu Tyr Tyr Leu Gln
805 810 815
Asn Gly Arg Asp Met Tyr Val Asp Gln Glu Leu Asp Ile Asn Arg Leu
820 825 830
Ser Asp Tyr Asp Val Asp His Ile Val Pro Gln Ser Phe Leu Lys Asp
835 840 845
Asp Ser Ile Asp Asn Lys Val Leu Thr Arg Ser Asp Lys Asn Arg Gly
850 855 860
Lys Ser Asp Asn Val Pro Ser Glu Glu Val Val Lys Lys Met Lys Asn
865 870 875 880
Tyr Trp Arg Gln Leu Leu Asn Ala Lys Leu Ile Thr Gln Arg Lys Phe
885 890 895
Asp Asn Leu Thr Lys Ala Glu Arg Gly Gly Leu Ser Glu Leu Asp Lys
900 905 910
Ala Gly Phe Ile Lys Arg Gln Leu Val Glu Thr Arg Gln Ile Thr Lys
915 920 925
His Val Ala Gln Ile Leu Asp Ser Arg Met Asn Thr Lys Tyr Asp Glu

930 935 940

Asn Asp Lys Leu Ile Arg Glu Val Lys Val Ile Thr Leu Lys Ser Lys
945 950 955 960

Leu Val Ser Asp Phe Arg Lys Asp Phe Gln Phe Tyr Lys Val Arg Glu
965 970 975

Ile Asn Asn Tyr His His Ala His Asp Ala Tyr Leu Asn Ala Val Val
980 985 990

Gly Thr Ala Leu Ile Lys Lys Tyr Pro Lys Leu Glu Ser Glu Phe Val
995 1000 1005

Tyr Gly Asp Tyr Lys Val Tyr Asp Val Arg Lys Met Ile Ala Lys
1010 1015 1020

Ser Glu Gln Glu Ile Gly Lys Ala Thr Ala Lys Tyr Phe Phe Tyr
1025 1030 1035

Ser Asn Ile Met Asn Phe Phe Lys Thr Glu Ile Thr Leu Ala Asn
1040 1045 1050

Gly Glu Ile Arg Lys Arg Pro Leu Ile Glu Thr Asn Gly Glu Thr
1055 1060 1065

Gly Glu Ile Val Trp Asp Lys Gly Arg Asp Phe Ala Thr Val Arg
1070 1075 1080

Lys Val Leu Ser Met Pro Gln Val Asn Ile Val Lys Lys Thr Glu
1085 1090 1095

Val Gln Thr Gly Gly Phe Ser Lys Glu Ser Ile Leu Pro Lys Arg
1100 1105 1110

Asn Ser Asp Lys Leu Ile Ala Arg Lys Lys Asp Trp Asp Pro Lys
1115 1120 1125

Lys Tyr Gly Gly Phe Asp Ser Pro Thr Val Ala Tyr Ser Val Leu
1130 1135 1140

Val Val Ala Lys Val Glu Lys Gly Lys Ser Lys Lys Leu Lys Ser
1145 1150 1155

Val Lys Glu Leu Leu Gly Ile Thr Ile Met Glu Arg Ser Ser Phe
1160 1165 1170

Glu Lys Asn Pro Ile Asp Phe Leu Glu Ala Lys Gly Tyr Lys Glu
1175 1180 1185

Val Lys Lys Asp Leu Ile Ile Lys Leu Pro Lys Tyr Ser Leu Phe
1190 1195 1200

Glu Leu Glu Asn Gly Arg Lys Arg Met Leu Ala Ser Ala Gly Glu
1205 1210 1215

Leu Gln Lys Gly Asn Glu Leu Ala Leu Pro Ser Lys Tyr Val Asn
1220 1225 1230

Phe Leu Tyr Leu Ala Ser His Tyr Glu Lys Leu Lys Gly Ser Pro
1235 1240 1245
Glu Asp Asn Glu Gln Lys Gln Leu Phe Val Glu Gln His Lys His
1250 1255 1260
Tyr Leu Asp Glu Ile Ile Glu Gln Ile Ser Glu Phe Ser Lys Arg
1265 1270 1275
Val Ile Leu Ala Asp Ala Asn Leu Asp Lys Val Leu Ser Ala Tyr
1280 1285 1290
Asn Lys His Arg Asp Lys Pro Ile Arg Glu Gln Ala Glu Asn Ile
1295 1300 1305
Ile His Leu Phe Thr Leu Thr Asn Leu Gly Ala Pro Ala Ala Phe
1310 1315 1320
Lys Tyr Phe Asp Thr Thr Ile Asp Arg Lys Arg Tyr Thr Ser Thr
1325 1330 1335
Lys Glu Val Leu Asp Ala Thr Leu Ile His Gln Ser Ile Thr Gly
1340 1345 1350
Leu Tyr Glu Thr Arg Ile Asp Leu Ser Gln Leu Gly Gly Asp
1355 1360 1365

<210> 11
<211> 1367
<212> PRT
<213> 人工序列
<220>
<223> 合成多肽
<400> 11

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

Glu Arg His Pro Ile Phe Gly Asn Ile Val Asp Glu Val Ala Tyr His
115 120 125
Glu Lys Tyr Pro Thr Ile Tyr His Leu Arg Lys Lys Leu Val Asp Ser
130 135 140
Thr Asp Lys Ala Asp Leu Arg Leu Ile Tyr Leu Ala Leu Ala His Met
145 150 155 160
Ile Lys Phe Arg Gly His Phe Leu Ile Glu Gly Asp Leu Asn Pro Asp
165 170 175
Asn Ser Asp Val Asp Lys Leu Phe Ile Gln Leu Val Gln Thr Tyr Asn
180 185 190
Gln Leu Phe Glu Glu Asn Pro Ile Asn Ala Ser Gly Val Asp Ala Lys
195 200 205
Ala Ile Leu Ser Ala Arg Leu Ser Lys Ser Arg Arg Leu Glu Asn Leu
210 215 220
Ile Ala Gln Leu Pro Gly Glu Lys Lys Asn Gly Leu Phe Gly Asn Leu
225 230 235 240
Ile Ala Leu Ser Leu Gly Leu Thr Pro Asn Phe Lys Ser Asn Phe Asp
245 250 255
Leu Ala Glu Asp Ala Lys Leu Gln Leu Ser Lys Asp Thr Tyr Asp Asp
260 265 270
Asp Leu Asp Asn Leu Leu Ala Gln Ile Gly Asp Gln Tyr Ala Asp Leu
275 280 285
Phe Leu Ala Ala Lys Asn Leu Ser Asp Ala Ile Leu Leu Ser Asp Ile
290 295 300
Leu Arg Val Asn Thr Glu Ile Thr Lys Ala Pro Leu Ser Ala Ser Met
305 310 315 320
Ile Lys Arg Tyr Asp Glu His His Gln Asp Leu Thr Leu Leu Lys Ala
325 330 335
Leu Val Arg Gln Gln Leu Pro Glu Lys Tyr Lys Glu Ile Phe Phe Asp
340 345 350
Gln Ser Lys Asn Gly Tyr Ala Gly Tyr Ile Asp Gly Gly Ala Ser Gln
355 360 365
Glu Glu Phe Tyr Lys Phe Ile Lys Pro Ile Leu Glu Lys Met Asp Gly
370 375 380
Thr Glu Glu Leu Leu Val Lys Leu Asn Arg Glu Asp Leu Leu Arg Lys
385 390 395 400
Gln Arg Thr Phe Asp Asn Gly Ser Ile Pro His Gln Ile His Leu Gly
405 410 415
Glu Leu His Ala Ile Leu Arg Arg Gln Glu Asp Phe Tyr Pro Phe Leu

420 425 430

Lys Asp Asn Arg Glu Lys Ile Glu Lys Ile Leu Thr Phe Arg Ile Pro

435 440 445

Tyr Tyr Val Gly Pro Leu Ala Arg Gly Asn Ser Arg Phe Ala Trp Met

450 455 460

Thr Arg Lys Ser Glu Glu Thr Ile Thr Pro Trp Asn Phe Glu Glu Val

465 470 475 480

Val Asp Lys Gly Ala Ser Ala Gln Ser Phe Ile Glu Arg Met Thr Asn

485 490 495

Phe Asp Lys Asn Leu Pro Asn Glu Lys Val Leu Pro Lys His Ser Leu

500 505 510

Leu Tyr Glu Tyr Phe Thr Val Tyr Asn Glu Leu Thr Lys Val Lys Tyr

515 520 525

Val Thr Glu Gly Met Arg Lys Pro Ala Phe Leu Ser Gly Glu Gln Lys

530 535 540

Lys Ala Ile Val Asp Leu Leu Phe Lys Thr Asn Arg Lys Val Thr Val

545 550 555 560

Lys Gln Leu Lys Glu Asp Tyr Phe Lys Lys Ile Glu Cys Phe Asp Ser

565 570 575

Val Glu Ile Ser Gly Val Glu Asp Arg Phe Asn Ala Ser Leu Gly Thr

580 585 590

Tyr His Asp Leu Leu Lys Ile Ile Lys Asp Lys Asp Phe Leu Asp Asn

595 600 605

Glu Glu Asn Glu Asp Ile Leu Glu Asp Ile Val Leu Thr Leu Thr Leu

610 615 620

Phe Glu Asp Arg Glu Met Ile Glu Glu Arg Leu Lys Thr Tyr Ala His

625 630 635 640

Leu Phe Asp Asp Lys Val Met Lys Gln Leu Lys Arg Arg Arg Tyr Thr

645 650 655

Gly Trp Gly Arg Leu Ser Arg Lys Leu Ile Asn Gly Ile Arg Asp Lys

660 665 670

Gln Ser Gly Lys Thr Ile Leu Asp Phe Leu Lys Ser Asp Gly Phe Ala

675 680 685

Asn Arg Asn Phe Met Gln Leu Ile His Asp Asp Ser Leu Thr Phe Lys

690 695 700

Glu Asp Ile Gln Lys Ala Gln Val Ser Gly Gln Gly Asp Ser Leu His

705 710 715 720

Glu His Ile Ala Asn Leu Ala Gly Ser Pro Ala Ile Lys Lys Gly Ile

725 730 735

Leu Gln Thr Val Lys Val Val Asp Glu Leu Val Lys Val Met Gly Arg
740 745 750
His Lys Pro Glu Asn Ile Val Ile Glu Met Ala Arg Glu Asn Gln Thr
755 760 765
Thr Gln Lys Gly Gln Lys Asn Ser Arg Glu Arg Met Lys Arg Ile Glu
770 775 780
Glu Gly Ile Lys Glu Leu Gly Ser Gln Ile Leu Lys Glu His Pro Val
785 790 795 800
Glu Asn Thr Gln Leu Gln Asn Glu Lys Leu Tyr Leu Tyr Tyr Leu Gln
805 810 815
Asn Gly Arg Asp Met Tyr Val Asp Gln Glu Leu Asp Ile Asn Arg Leu
820 825 830
Ser Asp Tyr Asp Val Asp His Ile Val Pro Gln Ser Phe Leu Lys Asp
835 840 845
Asp Ser Ile Asp Asn Lys Val Leu Thr Arg Ser Asp Lys Asn Arg Gly
850 855 860
Lys Ser Asp Asn Val Pro Ser Glu Glu Val Val Lys Lys Met Lys Asn
865 870 875 880
Tyr Trp Arg Gln Leu Leu Asn Ala Lys Leu Ile Thr Gln Arg Lys Phe
885 890 895
Asp Asn Leu Thr Lys Ala Glu Arg Gly Gly Leu Ser Glu Leu Asp Lys
900 905 910
Ala Gly Phe Ile Lys Arg Gln Leu Val Glu Thr Arg Gln Ile Thr Lys
915 920 925
His Val Ala Gln Ile Leu Asp Ser Arg Met Asn Thr Lys Tyr Asp Glu
930 935 940
Asn Asp Lys Leu Ile Arg Glu Val Lys Val Ile Thr Leu Lys Ser Lys
945 950 955 960
Leu Val Ser Asp Phe Arg Lys Asp Phe Gln Phe Tyr Lys Val Arg Glu
965 970 975
Ile Asn Asn Tyr His His Ala His Asp Ala Tyr Leu Asn Ala Val Val
980 985 990
Gly Thr Ala Leu Ile Lys Lys Tyr Pro Lys Leu Glu Ser Glu Phe Val
995 1000 1005
Tyr Gly Asp Tyr Lys Val Tyr Asp Val Arg Lys Met Ile Ala Lys
1010 1015 1020
Ser Glu Gln Glu Ile Gly Lys Ala Thr Ala Lys Tyr Phe Phe Tyr
1025 1030 1035
Ser Asn Ile Met Asn Phe Phe Lys Thr Glu Ile Thr Leu Ala Asn

1040 1045 1050

Gly Glu Ile Arg Lys Arg Pro Leu Ile Glu Thr Asn Gly Glu Thr
1055 1060 1065

Gly Glu Ile Val Trp Asp Lys Gly Arg Asp Phe Ala Thr Val Arg
1070 1075 1080

Lys Val Leu Ser Met Pro Gln Val Asn Ile Val Lys Lys Thr Glu
1085 1090 1095

Val Gln Thr Gly Gly Phe Ser Lys Glu Ser Ile Leu Pro Lys Arg
1100 1105 1110

Asn Ser Asp Lys Leu Ile Ala Arg Lys Lys Asp Trp Asp Pro Lys
1115 1120 1125

Lys Tyr Gly Gly Phe Asp Ser Pro Thr Val Ala Tyr Ser Val Leu
1130 1135 1140

Val Val Ala Lys Val Glu Lys Gly Lys Ser Lys Lys Leu Lys Ser
1145 1150 1155

Val Lys Glu Leu Leu Gly Ile Thr Ile Met Glu Arg Ser Ser Phe
1160 1165 1170

Glu Lys Asn Pro Ile Asp Phe Leu Glu Ala Lys Gly Tyr Lys Glu
1175 1180 1185

Val Lys Lys Asp Leu Ile Ile Lys Leu Pro Lys Tyr Ser Leu Phe
1190 1195 1200

Glu Leu Glu Asn Gly Arg Lys Arg Met Leu Ala Ser Ala Gly Glu
1205 1210 1215

Leu Gln Lys Gly Asn Glu Leu Ala Leu Pro Ser Lys Tyr Val Asn
1220 1225 1230

Phe Leu Tyr Leu Ala Ser His Tyr Glu Lys Leu Lys Gly Ser Pro
1235 1240 1245

Glu Asp Asn Glu Gln Lys Gln Leu Phe Val Glu Gln His Lys His
1250 1255 1260

Tyr Leu Asp Glu Ile Ile Glu Gln Ile Ser Glu Phe Ser Lys Arg
1265 1270 1275

Val Ile Leu Ala Asp Ala Asn Leu Asp Lys Val Leu Ser Ala Tyr
1280 1285 1290

Asn Lys His Arg Asp Lys Pro Ile Arg Glu Gln Ala Glu Asn Ile
1295 1300 1305

Ile His Leu Phe Thr Leu Thr Asn Leu Gly Ala Pro Ala Ala Phe
1310 1315 1320

Lys Tyr Phe Asp Thr Thr Ile Asp Arg Lys Arg Tyr Thr Ser Thr
1325 1330 1335

Lys Glu Val Leu Asp Ala Thr Leu Ile His Gln Ser Ile Thr Gly
1340 1345 1350
Leu Tyr Glu Thr Arg Ile Asp Leu Ser Gln Leu Gly Gly Asp
1355 1360 1365

<210> 12
<211> 1052
<212> PRT
<213> 人工序列
<220>
<223> 合成多肽
<400> 12

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

Ile Lys Glu Trp Tyr Glu Met Leu Met Gly His Cys Thr Tyr Phe Pro
225 230 235 240
Glu Glu Leu Arg Ser Val Lys Tyr Ala Tyr Asn Ala Asp Leu Tyr Asn
245 250 255
Ala Leu Asn Asp Leu Asn Asn Leu Val Ile Thr Arg Asp Glu Asn Glu
260 265 270
Lys Leu Glu Tyr Tyr Glu Lys Phe Gln Ile Ile Glu Asn Val Phe Lys
275 280 285
Gln Lys Lys Lys Pro Thr Leu Lys Gln Ile Ala Lys Glu Ile Leu Val
290 295 300
Asn Glu Glu Asp Ile Lys Gly Tyr Arg Val Thr Ser Thr Gly Lys Pro
305 310 315 320
Glu Phe Thr Asn Leu Lys Val Tyr His Asp Ile Lys Asp Ile Thr Ala
325 330 335
Arg Lys Glu Ile Ile Glu Asn Ala Glu Leu Leu Asp Gln Ile Ala Lys
340 345 350
Ile Leu Thr Ile Tyr Gln Ser Ser Glu Asp Ile Gln Glu Glu Leu Thr
355 360 365
Asn Leu Asn Ser Glu Leu Thr Gln Glu Glu Ile Glu Gln Ile Ser Asn
370 375 380
Leu Lys Gly Tyr Thr Gly Thr His Asn Leu Ser Leu Lys Ala Ile Asn
385 390 395 400
Leu Ile Leu Asp Glu Leu Trp His Thr Asn Asp Asn Gln Ile Ala Ile
405 410 415
Phe Asn Arg Leu Lys Leu Val Pro Lys Lys Val Asp Leu Ser Gln Gln
420 425 430
Lys Glu Ile Pro Thr Thr Leu Val Asp Asp Phe Ile Leu Ser Pro Val
435 440 445
Val Lys Arg Ser Phe Ile Gln Ser Ile Lys Val Ile Asn Ala Ile Ile
450 455 460
Lys Lys Tyr Gly Leu Pro Asn Asp Ile Ile Ile Glu Leu Ala Arg Glu
465 470 475 480
Lys Asn Ser Lys Asp Ala Gln Lys Met Ile Asn Glu Met Gln Lys Arg
485 490 495
Asn Arg Gln Thr Asn Glu Arg Ile Glu Glu Ile Ile Arg Thr Thr Gly
500 505 510
Lys Glu Asn Ala Lys Tyr Leu Ile Glu Lys Ile Lys Leu His Asp Met
515 520 525
Gln Glu Gly Lys Cys Leu Tyr Ser Leu Glu Ala Ile Pro Leu Glu Asp

530 535 540

Leu Leu Asn Asn Pro Phe Asn Tyr Glu Val Asp His Ile Ile Pro Arg
545 550 555 560

Ser Val Ser Phe Asp Asn Ser Phe Asn Asn Lys Val Leu Val Lys Gln
565 570 575

Glu Glu Asn Ser Lys Lys Gly Asn Arg Thr Pro Phe Gln Tyr Leu Ser
580 585 590

Ser Ser Asp Ser Lys Ile Ser Tyr Glu Thr Phe Lys Lys His Ile Leu
595 600 605

Asn Leu Ala Lys Gly Lys Gly Arg Ile Ser Lys Thr Lys Lys Glu Tyr
610 615 620

Leu Leu Glu Glu Arg Asp Ile Asn Arg Phe Ser Val Gln Lys Asp Phe
625 630 635 640

Ile Asn Arg Asn Leu Val Asp Thr Arg Tyr Ala Thr Arg Gly Leu Met
645 650 655

Asn Leu Leu Arg Ser Tyr Phe Arg Val Asn Asn Leu Asp Val Lys Val
660 665 670

Lys Ser Ile Asn Gly Gly Phe Thr Ser Phe Leu Arg Arg Lys Trp Lys
675 680 685

Phe Lys Lys Glu Arg Asn Lys Gly Tyr Lys His His Ala Glu Asp Ala
690 695 700

Leu Ile Ile Ala Asn Ala Asp Phe Ile Phe Lys Glu Trp Lys Lys Leu
705 710 715 720

Asp Lys Ala Lys Lys Val Met Glu Asn Gln Met Phe Glu Glu Lys Gln
725 730 735

Ala Glu Ser Met Pro Glu Ile Glu Thr Glu Gln Glu Tyr Lys Glu Ile
740 745 750

Phe Ile Thr Pro His Gln Ile Lys His Ile Lys Asp Phe Lys Asp Tyr
755 760 765

Lys Tyr Ser His Arg Val Asp Lys Lys Pro Asn Arg Glu Leu Ile Asn
770 775 780

Asp Thr Leu Tyr Ser Thr Arg Lys Asp Asp Lys Gly Asn Thr Leu Ile
785 790 795 800

Val Asn Asn Leu Asn Gly Leu Tyr Asp Lys Asp Asn Asp Lys Leu Lys
805 810 815

Lys Leu Ile Asn Lys Ser Pro Glu Lys Leu Leu Met Tyr His His Asp
820 825 830

Pro Gln Thr Tyr Gln Lys Leu Lys Leu Ile Met Glu Gln Tyr Gly Asp
835 840 845

Glu Lys Asn Pro Leu Tyr Lys Tyr Tyr Glu Glu Thr Gly Asn Tyr Leu
850 855 860
Thr Lys Tyr Ser Lys Lys Asp Asn Gly Pro Val Ile Lys Lys Ile Lys
865 870 875 880
Tyr Tyr Gly Asn Lys Leu Asn Ala His Leu Asp Ile Thr Asp Asp Tyr
885 890 895
Pro Asn Ser Arg Asn Lys Val Val Lys Leu Ser Leu Lys Pro Tyr Arg
900 905 910
Phe Asp Val Tyr Leu Asp Asn Gly Val Tyr Lys Phe Val Thr Val Lys
915 920 925
Asn Leu Asp Val Ile Lys Lys Glu Asn Tyr Tyr Glu Val Asn Ser Lys
930 935 940
Cys Tyr Glu Glu Ala Lys Lys Leu Lys Lys Ile Ser Asn Gln Ala Glu
945 950 955 960
Phe Ile Ala Ser Phe Tyr Asn Asn Asp Leu Ile Lys Ile Asn Gly Glu
965 970 975
Leu Tyr Arg Val Ile Gly Val Asn Asn Asp Leu Leu Asn Arg Ile Glu
980 985 990
Val Asn Met Ile Asp Ile Thr Tyr Arg Glu Tyr Leu Glu Asn Met Asn
995 1000 1005
Asp Lys Arg Pro Pro Arg Ile Ile Lys Thr Ile Ala Ser Lys Thr
1010 1015 1020
Gln Ser Ile Lys Lys Tyr Ser Thr Asp Ile Leu Gly Asn Leu Tyr
1025 1030 1035
Glu Val Lys Ser Lys Lys His Pro Gln Ile Ile Lys Lys Gly
1040 1045 1050

<210> 13
<211> 1052
<212> PRT
<213> 人工序列
<220>
<223> 合成多肽
<400> 13

Lys Arg Asn Tyr Ile Leu Gly Leu Asp Ile Gly Ile Thr Ser Val Gly
1 5 10 15
Tyr Gly Ile Ile Asp Tyr Glu Thr Arg Asp Val Ile Asp Ala Gly Val
20 25 30
Arg Leu Phe Lys Glu Ala Asn Val Glu Asn Asn Glu Gly Arg Arg Ser
35 40 45

Lys Arg Gly Ala Arg Arg Leu Lys Arg Arg Arg Arg His Arg Ile Gln
50 55 60
Arg Val Lys Lys Leu Leu Phe Asp Tyr Asn Leu Leu Thr Asp His Ser
65 70 75 80
Glu Leu Ser Gly Ile Asn Pro Tyr Glu Ala Arg Val Lys Gly Leu Ser
85 90 95
Gln Lys Leu Ser Glu Glu Glu Phe Ser Ala Ala Leu Leu His Leu Ala
100 105 110
Lys Arg Arg Gly Val His Asn Val Asn Glu Val Glu Glu Asp Thr Gly
115 120 125
Asn Glu Leu Ser Thr Lys Glu Gln Ile Ser Arg Asn Ser Lys Ala Leu
130 135 140
Glu Glu Lys Tyr Val Ala Glu Leu Gln Leu Glu Arg Leu Lys Lys Asp
145 150 155 160
Gly Glu Val Arg Gly Ser Ile Asn Arg Phe Lys Thr Ser Asp Tyr Val
165 170 175
Lys Glu Ala Lys Gln Leu Leu Lys Val Gln Lys Ala Tyr His Gln Leu
180 185 190
Asp Gln Ser Phe Ile Asp Thr Tyr Ile Asp Leu Leu Glu Thr Arg Arg
195 200 205
Thr Tyr Tyr Glu Gly Pro Gly Glu Gly Ser Pro Phe Gly Trp Lys Asp
210 215 220
Ile Lys Glu Trp Tyr Glu Met Leu Met Gly His Cys Thr Tyr Phe Pro
225 230 235 240
Glu Glu Leu Arg Ser Val Lys Tyr Ala Tyr Asn Ala Asp Leu Tyr Asn
245 250 255
Ala Leu Asn Asp Leu Asn Asn Leu Val Ile Thr Arg Asp Glu Asn Glu
260 265 270
Lys Leu Glu Tyr Tyr Glu Lys Phe Gln Ile Ile Glu Asn Val Phe Lys
275 280 285
Gln Lys Lys Lys Pro Thr Leu Lys Gln Ile Ala Lys Glu Ile Leu Val
290 295 300
Asn Glu Glu Asp Ile Lys Gly Tyr Arg Val Thr Ser Thr Gly Lys Pro
305 310 315 320
Glu Phe Thr Asn Leu Lys Val Tyr His Asp Ile Lys Asp Ile Thr Ala
325 330 335
Arg Lys Glu Ile Ile Glu Asn Ala Glu Leu Leu Asp Gln Ile Ala Lys
340 345 350
Ile Leu Thr Ile Tyr Gln Ser Ser Glu Asp Ile Gln Glu Glu Leu Thr

355 360 365

Asn Leu Asn Ser Glu Leu Thr Gln Glu Glu Ile Glu Gln Ile Ser Asn

370 375 380

Leu Lys Gly Tyr Thr Gly Thr His Asn Leu Ser Leu Lys Ala Ile Asn

385 390 395 400

Leu Ile Leu Asp Glu Leu Trp His Thr Asn Asp Asn Gln Ile Ala Ile

405 410 415

Phe Asn Arg Leu Lys Leu Val Pro Lys Lys Val Asp Leu Ser Gln Gln

420 425 430

Lys Glu Ile Pro Thr Thr Leu Val Asp Asp Phe Ile Leu Ser Pro Val

435 440 445

Val Lys Arg Ser Phe Ile Gln Ser Ile Lys Val Ile Asn Ala Ile Ile

450 455 460

Lys Lys Tyr Gly Leu Pro Asn Asp Ile Ile Ile Glu Leu Ala Arg Glu

465 470 475 480

Lys Asn Ser Lys Asp Ala Gln Lys Met Ile Asn Glu Met Gln Lys Arg

485 490 495

Asn Arg Gln Thr Asn Glu Arg Ile Glu Glu Ile Ile Arg Thr Thr Gly

500 505 510

Lys Glu Asn Ala Lys Tyr Leu Ile Glu Lys Ile Lys Leu His Asp Met

515 520 525

Gln Glu Gly Lys Cys Leu Tyr Ser Leu Glu Ala Ile Pro Leu Glu Asp

530 535 540

Leu Leu Asn Asn Pro Phe Asn Tyr Glu Val Asp His Ile Ile Pro Arg

545 550 555 560

Ser Val Ser Phe Asp Asn Ser Phe Asn Asn Lys Val Leu Val Lys Gln

565 570 575

Glu Glu Ala Ser Lys Lys Gly Asn Arg Thr Pro Phe Gln Tyr Leu Ser

580 585 590

Ser Ser Asp Ser Lys Ile Ser Tyr Glu Thr Phe Lys Lys His Ile Leu

595 600 605

Asn Leu Ala Lys Gly Lys Gly Arg Ile Ser Lys Thr Lys Lys Glu Tyr

610 615 620

Leu Leu Glu Glu Arg Asp Ile Asn Arg Phe Ser Val Gln Lys Asp Phe

625 630 635 640

Ile Asn Arg Asn Leu Val Asp Thr Arg Tyr Ala Thr Arg Gly Leu Met

645 650 655

Asn Leu Leu Arg Ser Tyr Phe Arg Val Asn Asn Leu Asp Val Lys Val

660 665 670

Lys Ser Ile Asn Gly Gly Phe Thr Ser Phe Leu Arg Arg Lys Trp Lys
675 680 685
Phe Lys Lys Glu Arg Asn Lys Gly Tyr Lys His His Ala Glu Asp Ala
690 695 700
Leu Ile Ile Ala Asn Ala Asp Phe Ile Phe Lys Glu Trp Lys Lys Leu
705 710 715 720
Asp Lys Ala Lys Lys Val Met Glu Asn Gln Met Phe Glu Glu Lys Gln
725 730 735
Ala Glu Ser Met Pro Glu Ile Glu Thr Glu Gln Glu Tyr Lys Glu Ile
740 745 750
Phe Ile Thr Pro His Gln Ile Lys His Ile Lys Asp Phe Lys Asp Tyr
755 760 765
Lys Tyr Ser His Arg Val Asp Lys Lys Pro Asn Arg Glu Leu Ile Asn
770 775 780
Asp Thr Leu Tyr Ser Thr Arg Lys Asp Asp Lys Gly Asn Thr Leu Ile
785 790 795 800
Val Asn Asn Leu Asn Gly Leu Tyr Asp Lys Asp Asn Asp Lys Leu Lys
805 810 815
Lys Leu Ile Asn Lys Ser Pro Glu Lys Leu Leu Met Tyr His His Asp
820 825 830
Pro Gln Thr Tyr Gln Lys Leu Lys Leu Ile Met Glu Gln Tyr Gly Asp
835 840 845
Glu Lys Asn Pro Leu Tyr Lys Tyr Tyr Glu Glu Thr Gly Asn Tyr Leu
850 855 860
Thr Lys Tyr Ser Lys Lys Asp Asn Gly Pro Val Ile Lys Lys Ile Lys
865 870 875 880
Tyr Tyr Gly Asn Lys Leu Asn Ala His Leu Asp Ile Thr Asp Asp Tyr
885 890 895
Pro Asn Ser Arg Asn Lys Val Val Lys Leu Ser Leu Lys Pro Tyr Arg
900 905 910
Phe Asp Val Tyr Leu Asp Asn Gly Val Tyr Lys Phe Val Thr Val Lys
915 920 925
Asn Leu Asp Val Ile Lys Lys Glu Asn Tyr Tyr Glu Val Asn Ser Lys
930 935 940
Cys Tyr Glu Glu Ala Lys Lys Leu Lys Lys Ile Ser Asn Gln Ala Glu
945 950 955 960
Phe Ile Ala Ser Phe Tyr Asn Asn Asp Leu Ile Lys Ile Asn Gly Glu
965 970 975
Leu Tyr Arg Val Ile Gly Val Asn Asn Asp Leu Leu Asn Arg Ile Glu

980 985 990
Val Asn Met Ile Asp Ile Thr Tyr Arg Glu Tyr Leu Glu Asn Met Asn
995 1000 1005
Asp Lys Arg Pro Pro Arg Ile Ile Lys Thr Ile Ala Ser Lys Thr
1010 1015 1020
Gln Ser Ile Lys Lys Tyr Ser Thr Asp Ile Leu Gly Asn Leu Tyr
1025 1030 1035
Glu Val Lys Ser Lys Lys His Pro Gln Ile Ile Lys Lys Gly
1040 1045 1050

<210> 14
<211> 1052
<212> PRT
<213> 人工序列
<220>
<223> 合成多肽
<400> 14

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

180 185 190

Asp Gln Ser Phe Ile Asp Thr Tyr Ile Asp Leu Leu Glu Thr Arg Arg

195 200 205

Thr Tyr Tyr Glu Gly Pro Gly Glu Gly Ser Pro Phe Gly Trp Lys Asp

210 215 220

Ile Lys Glu Trp Tyr Glu Met Leu Met Gly His Cys Thr Tyr Phe Pro

225 230 235 240

Glu Glu Leu Arg Ser Val Lys Tyr Ala Tyr Asn Ala Asp Leu Tyr Asn

245 250 255

Ala Leu Asn Asp Leu Asn Asn Leu Val Ile Thr Arg Asp Glu Asn Glu

260 265 270

Lys Leu Glu Tyr Tyr Glu Lys Phe Gln Ile Ile Glu Asn Val Phe Lys

275 280 285

Gln Lys Lys Lys Pro Thr Leu Lys Gln Ile Ala Lys Glu Ile Leu Val

290 295 300

Asn Glu Glu Asp Ile Lys Gly Tyr Arg Val Thr Ser Thr Gly Lys Pro

305 310 315 320

Glu Phe Thr Asn Leu Lys Val Tyr His Asp Ile Lys Asp Ile Thr Ala

325 330 335

Arg Lys Glu Ile Ile Glu Asn Ala Glu Leu Leu Asp Gln Ile Ala Lys

340 345 350

Ile Leu Thr Ile Tyr Gln Ser Ser Glu Asp Ile Gln Glu Glu Leu Thr

355 360 365

Asn Leu Asn Ser Glu Leu Thr Gln Glu Glu Ile Glu Gln Ile Ser Asn

370 375 380

Leu Lys Gly Tyr Thr Gly Thr His Asn Leu Ser Leu Lys Ala Ile Asn

385 390 395 400

Leu Ile Leu Asp Glu Leu Trp His Thr Asn Asp Asn Gln Ile Ala Ile

405 410 415

Phe Asn Arg Leu Lys Leu Val Pro Lys Lys Val Asp Leu Ser Gln Gln

420 425 430

Lys Glu Ile Pro Thr Thr Leu Val Asp Asp Phe Ile Leu Ser Pro Val

435 440 445

Val Lys Arg Ser Phe Ile Gln Ser Ile Lys Val Ile Asn Ala Ile Ile

450 455 460

Lys Lys Tyr Gly Leu Pro Asn Asp Ile Ile Ile Glu Leu Ala Arg Glu

465 470 475 480

Lys Asn Ser Lys Asp Ala Gln Lys Met Ile Asn Glu Met Gln Lys Arg

485 490 495

Asn Arg Gln Thr Asn Glu Arg Ile Glu Glu Ile Ile Arg Thr Thr Gly
500 505 510
Lys Glu Asn Ala Lys Tyr Leu Ile Glu Lys Ile Lys Leu His Asp Met
515 520 525
Gln Glu Gly Lys Cys Leu Tyr Ser Leu Glu Ala Ile Pro Leu Glu Asp
530 535 540
Leu Leu Asn Asn Pro Phe Asn Tyr Glu Val Asp His Ile Ile Pro Arg
545 550 555 560
Ser Val Ser Phe Asp Asn Ser Phe Asn Asn Lys Val Leu Val Lys Gln
565 570 575
Glu Glu Ala Ser Lys Lys Gly Asn Arg Thr Pro Phe Gln Tyr Leu Ser
580 585 590
Ser Ser Asp Ser Lys Ile Ser Tyr Glu Thr Phe Lys Lys His Ile Leu
595 600 605
Asn Leu Ala Lys Gly Lys Gly Arg Ile Ser Lys Thr Lys Lys Glu Tyr
610 615 620
Leu Leu Glu Glu Arg Asp Ile Asn Arg Phe Ser Val Gln Lys Asp Phe
625 630 635 640
Ile Asn Arg Asn Leu Val Asp Thr Arg Tyr Ala Thr Arg Gly Leu Met
645 650 655
Asn Leu Leu Arg Ser Tyr Phe Arg Val Asn Asn Leu Asp Val Lys Val
660 665 670
Lys Ser Ile Asn Gly Gly Phe Thr Ser Phe Leu Arg Arg Lys Trp Lys
675 680 685
Phe Lys Lys Glu Arg Asn Lys Gly Tyr Lys His His Ala Glu Asp Ala
690 695 700
Leu Ile Ile Ala Asn Ala Asp Phe Ile Phe Lys Glu Trp Lys Lys Leu
705 710 715 720
Asp Lys Ala Lys Lys Val Met Glu Asn Gln Met Phe Glu Glu Lys Gln
725 730 735
Ala Glu Ser Met Pro Glu Ile Glu Thr Glu Gln Glu Tyr Lys Glu Ile
740 745 750
Phe Ile Thr Pro His Gln Ile Lys His Ile Lys Asp Phe Lys Asp Tyr
755 760 765
Lys Tyr Ser His Arg Val Asp Lys Lys Pro Asn Arg Lys Leu Ile Asn
770 775 780
Asp Thr Leu Tyr Ser Thr Arg Lys Asp Asp Lys Gly Asn Thr Leu Ile
785 790 795 800
Val Asn Asn Leu Asn Gly Leu Tyr Asp Lys Asp Asn Asp Lys Leu Lys

805 810 815
Lys Leu Ile Asn Lys Ser Pro Glu Lys Leu Leu Met Tyr His His Asp
820 825 830
Pro Gln Thr Tyr Gln Lys Leu Lys Leu Ile Met Glu Gln Tyr Gly Asp
835 840 845
Glu Lys Asn Pro Leu Tyr Lys Tyr Tyr Glu Glu Thr Gly Asn Tyr Leu
850 855 860
Thr Lys Tyr Ser Lys Lys Asp Asn Gly Pro Val Ile Lys Lys Ile Lys
865 870 875 880
Tyr Tyr Gly Asn Lys Leu Asn Ala His Leu Asp Ile Thr Asp Asp Tyr
885 890 895
Pro Asn Ser Arg Asn Lys Val Val Lys Leu Ser Leu Lys Pro Tyr Arg
900 905 910
Phe Asp Val Tyr Leu Asp Asn Gly Val Tyr Lys Phe Val Thr Val Lys
915 920 925
Asn Leu Asp Val Ile Lys Lys Glu Asn Tyr Tyr Glu Val Asn Ser Lys
930 935 940
Cys Tyr Glu Glu Ala Lys Lys Leu Lys Lys Ile Ser Asn Gln Ala Glu
945 950 955 960
Phe Ile Ala Ser Phe Tyr Lys Asn Asp Leu Ile Lys Ile Asn Gly Glu
965 970 975
Leu Tyr Arg Val Ile Gly Val Asn Asn Asp Leu Leu Asn Arg Ile Glu
980 985 990
Val Asn Met Ile Asp Ile Thr Tyr Arg Glu Tyr Leu Glu Asn Met Asn
995 1000 1005
Asp Lys Arg Pro Pro His Ile Ile Lys Thr Ile Ala Ser Lys Thr
1010 1015 1020
Gln Ser Ile Lys Lys Tyr Ser Thr Asp Ile Leu Gly Asn Leu Tyr
1025 1030 1035
Glu Val Lys Ser Lys Lys His Pro Gln Ile Ile Lys Lys Gly
1040 1045 1050

<210> 15
<211> 1367
<212> PRT
<213> 人工序列
<220>
<223> 合成多肽
<400> 15
Asp Lys Lys Tyr Ser Ile Gly Leu Asp Ile Gly Thr Asn Ser Val Gly

1 5 10 15
Trp Ala Val Ile Thr Asp Glu Tyr Lys Val Pro Ser Lys Lys Phe Lys
20 25 30
Val Leu Gly Asn Thr Asp Arg His Ser Ile Lys Lys Asn Leu Ile Gly
35 40 45
Ala Leu Leu Phe Asp Ser Gly Glu Thr Ala Glu Ala Thr Arg Leu Lys
50 55 60
Arg Thr Ala Arg Arg Arg Tyr Thr Arg Arg Lys Asn Arg Ile Cys Tyr
65 70 75 80
Leu Gln Glu Ile Phe Ser Asn Glu Met Ala Lys Val Asp Asp Ser Phe
85 90 95
Phe His Arg Leu Glu Glu Ser Phe Leu Val Glu Glu Asp Lys Lys His
100 105 110
Glu Arg His Pro Ile Phe Gly Asn Ile Val Asp Glu Val Ala Tyr His
115 120 125
Glu Lys Tyr Pro Thr Ile Tyr His Leu Arg Lys Lys Leu Val Asp Ser
130 135 140
Thr Asp Lys Ala Asp Leu Arg Leu Ile Tyr Leu Ala Leu Ala His Met
145 150 155 160
Ile Lys Phe Arg Gly His Phe Leu Ile Glu Gly Asp Leu Asn Pro Asp
165 170 175
Asn Ser Asp Val Asp Lys Leu Phe Ile Gln Leu Val Gln Thr Tyr Asn
180 185 190
Gln Leu Phe Glu Glu Asn Pro Ile Asn Ala Ser Gly Val Asp Ala Lys
195 200 205
Ala Ile Leu Ser Ala Arg Leu Ser Lys Ser Arg Arg Leu Glu Asn Leu
210 215 220
Ile Ala Gln Leu Pro Gly Glu Lys Lys Asn Gly Leu Phe Gly Asn Leu
225 230 235 240
Ile Ala Leu Ser Leu Gly Leu Thr Pro Asn Phe Lys Ser Asn Phe Asp
245 250 255
Leu Ala Glu Asp Ala Lys Leu Gln Leu Ser Lys Asp Thr Tyr Asp Asp
260 265 270
Asp Leu Asp Asn Leu Leu Ala Gln Ile Gly Asp Gln Tyr Ala Asp Leu
275 280 285
Phe Leu Ala Ala Lys Asn Leu Ser Asp Ala Ile Leu Leu Ser Asp Ile
290 295 300
Leu Arg Val Asn Thr Glu Ile Thr Lys Ala Pro Leu Ser Ala Ser Met
305 310 315 320

Ile Lys Arg Tyr Asp Glu His His Gln Asp Leu Thr Leu Leu Lys Ala
325 330 335
Leu Val Arg Gln Gln Leu Pro Glu Lys Tyr Lys Glu Ile Phe Phe Asp
340 345 350
Gln Ser Lys Asn Gly Tyr Ala Gly Tyr Ile Asp Gly Gly Ala Ser Gln
355 360 365
Glu Glu Phe Tyr Lys Phe Ile Lys Pro Ile Leu Glu Lys Met Asp Gly
370 375 380
Thr Glu Glu Leu Leu Val Lys Leu Asn Arg Glu Asp Leu Leu Arg Lys
385 390 395 400
Gln Arg Thr Phe Asp Asn Gly Ser Ile Pro His Gln Ile His Leu Gly
405 410 415
Glu Leu His Ala Ile Leu Arg Arg Gln Glu Asp Phe Tyr Pro Phe Leu
420 425 430
Lys Asp Asn Arg Glu Lys Ile Glu Lys Ile Leu Thr Phe Arg Ile Pro
435 440 445
Tyr Tyr Val Gly Pro Leu Ala Arg Gly Asn Ser Arg Phe Ala Trp Met
450 455 460
Thr Arg Lys Ser Glu Glu Thr Ile Thr Pro Trp Asn Phe Glu Glu Val
465 470 475 480
Val Asp Lys Gly Ala Ser Ala Gln Ser Phe Ile Glu Arg Met Thr Asn
485 490 495
Phe Asp Lys Asn Leu Pro Asn Glu Lys Val Leu Pro Lys His Ser Leu
500 505 510
Leu Tyr Glu Tyr Phe Thr Val Tyr Asn Glu Leu Thr Lys Val Lys Tyr
515 520 525
Val Thr Glu Gly Met Arg Lys Pro Ala Phe Leu Ser Gly Glu Gln Lys
530 535 540
Lys Ala Ile Val Asp Leu Leu Phe Lys Thr Asn Arg Lys Val Thr Val
545 550 555 560
Lys Gln Leu Lys Glu Asp Tyr Phe Lys Lys Ile Glu Cys Phe Asp Ser
565 570 575
Val Glu Ile Ser Gly Val Glu Asp Arg Phe Asn Ala Ser Leu Gly Thr
580 585 590
Tyr His Asp Leu Leu Lys Ile Ile Lys Asp Lys Asp Phe Leu Asp Asn
595 600 605
Glu Glu Asn Glu Asp Ile Leu Glu Asp Ile Val Leu Thr Leu Thr Leu
610 615 620
Phe Glu Asp Arg Glu Met Ile Glu Glu Arg Leu Lys Thr Tyr Ala His

625 630 635 640
Leu Phe Asp Asp Lys Val Met Lys Gln Leu Lys Arg Arg Arg Tyr Thr
645 650 655
Gly Trp Gly Arg Leu Ser Arg Lys Leu Ile Asn Gly Ile Arg Asp Lys
660 665 670
Gln Ser Gly Lys Thr Ile Leu Asp Phe Leu Lys Ser Asp Gly Phe Ala
675 680 685
Asn Arg Asn Phe Met Gln Leu Ile His Asp Asp Ser Leu Thr Phe Lys
690 695 700
Glu Asp Ile Gln Lys Ala Gln Val Ser Gly Gln Gly Asp Ser Leu His
705 710 715 720
Glu His Ile Ala Asn Leu Ala Gly Ser Pro Ala Ile Lys Lys Gly Ile
725 730 735
Leu Gln Thr Val Lys Val Val Asp Glu Leu Val Lys Val Met Gly Arg
740 745 750
His Lys Pro Glu Asn Ile Val Ile Glu Met Ala Arg Glu Asn Gln Thr
755 760 765
Thr Gln Lys Gly Gln Lys Asn Ser Arg Glu Arg Met Lys Arg Ile Glu
770 775 780
Glu Gly Ile Lys Glu Leu Gly Ser Gln Ile Leu Lys Glu His Pro Val
785 790 795 800
Glu Asn Thr Gln Leu Gln Asn Glu Lys Leu Tyr Leu Tyr Tyr Leu Gln
805 810 815
Asn Gly Arg Asp Met Tyr Val Asp Gln Glu Leu Asp Ile Asn Arg Leu
820 825 830
Ser Asp Tyr Asp Val Asp His Ile Val Pro Gln Ser Phe Leu Lys Asp
835 840 845
Asp Ser Ile Asp Asn Lys Val Leu Thr Arg Ser Asp Lys Asn Arg Gly
850 855 860
Lys Ser Asp Asn Val Pro Ser Glu Glu Val Val Lys Lys Met Lys Asn
865 870 875 880
Tyr Trp Arg Gln Leu Leu Asn Ala Lys Leu Ile Thr Gln Arg Lys Phe
885 890 895
Asp Asn Leu Thr Lys Ala Glu Arg Gly Gly Leu Ser Glu Leu Asp Lys
900 905 910
Ala Gly Phe Ile Lys Arg Gln Leu Val Glu Thr Arg Gln Ile Thr Lys
915 920 925
His Val Ala Gln Ile Leu Asp Ser Arg Met Asn Thr Lys Tyr Asp Glu
930 935 940

Asn Asp Lys Leu Ile Arg Glu Val Lys Val Ile Thr Leu Lys Ser Lys
945 950 955 960
Leu Val Ser Asp Phe Arg Lys Asp Phe Gln Phe Tyr Lys Val Arg Glu
965 970 975
Ile Asn Asn Tyr His His Ala His Asp Ala Tyr Leu Asn Ala Val Val
980 985 990
Gly Thr Ala Leu Ile Lys Lys Tyr Pro Lys Leu Glu Ser Glu Phe Val
995 1000 1005
Tyr Gly Asp Tyr Lys Val Tyr Asp Val Arg Lys Met Ile Ala Lys
1010 1015 1020
Ser Glu Gln Glu Ile Gly Lys Ala Thr Ala Lys Tyr Phe Phe Tyr
1025 1030 1035
Ser Asn Ile Met Asn Phe Phe Lys Thr Glu Ile Thr Leu Ala Asn
1040 1045 1050
Gly Glu Ile Arg Lys Arg Pro Leu Ile Glu Thr Asn Gly Glu Thr
1055 1060 1065
Gly Glu Ile Val Trp Asp Lys Gly Arg Asp Phe Ala Thr Val Arg
1070 1075 1080
Lys Val Leu Ser Met Pro Gln Val Asn Ile Val Lys Lys Thr Glu
1085 1090 1095
Val Gln Thr Gly Gly Phe Ser Lys Glu Ser Ile Leu Pro Lys Arg
1100 1105 1110
Asn Ser Asp Lys Leu Ile Ala Arg Lys Lys Asp Trp Asp Pro Lys
1115 1120 1125
Lys Tyr Gly Gly Phe Asp Ser Pro Thr Val Ala Tyr Ser Val Leu
1130 1135 1140
Val Val Ala Lys Val Glu Lys Gly Lys Ser Lys Lys Leu Lys Ser
1145 1150 1155
Val Lys Glu Leu Leu Gly Ile Thr Ile Met Glu Arg Ser Ser Phe
1160 1165 1170
Glu Lys Asn Pro Ile Asp Phe Leu Glu Ala Lys Gly Tyr Lys Glu
1175 1180 1185
Val Lys Lys Asp Leu Ile Ile Lys Leu Pro Lys Tyr Ser Leu Phe
1190 1195 1200
Glu Leu Glu Asn Gly Arg Lys Arg Met Leu Ala Ser Ala Gly Glu
1205 1210 1215
Leu Gln Lys Gly Asn Glu Leu Ala Leu Pro Ser Lys Tyr Val Asn
1220 1225 1230
Phe Leu Tyr Leu Ala Ser His Tyr Glu Lys Leu Lys Gly Ser Pro

1235 1240 1245
Glu Asp Asn Glu Gln Lys Gln Leu Phe Val Glu Gln His Lys His
1250 1255 1260
Tyr Leu Asp Glu Ile Ile Glu Gln Ile Ser Glu Phe Ser Lys Arg
1265 1270 1275
Val Ile Leu Ala Asp Ala Asn Leu Asp Lys Val Leu Ser Ala Tyr
1280 1285 1290
Asn Lys His Arg Asp Lys Pro Ile Arg Glu Gln Ala Glu Asn Ile
1295 1300 1305
Ile His Leu Phe Thr Leu Thr Asn Leu Gly Ala Pro Ala Ala Phe
1310 1315 1320
Lys Tyr Phe Asp Thr Thr Ile Asp Arg Lys Arg Tyr Thr Ser Thr
1325 1330 1335
Lys Glu Val Leu Asp Ala Thr Leu Ile His Gln Ser Ile Thr Gly
1340 1345 1350
Leu Tyr Glu Thr Arg Ile Asp Leu Ser Gln Leu Gly Gly Asp
1355 1360 1365

<210> 16
<211> 1367
<212> PRT
<213> 人工序列
<220>
<223> 合成多肽
<400> 16

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

115 120 125

Glu Lys Tyr Pro Thr Ile Tyr His Leu Arg Lys Lys Leu Val Asp Ser
130 135 140

Thr Asp Lys Ala Asp Leu Arg Leu Ile Tyr Leu Ala Leu Ala His Met
145 150 155 160

Ile Lys Phe Arg Gly His Phe Leu Ile Glu Gly Asp Leu Asn Pro Asp
165 170 175

Asn Ser Asp Val Asp Lys Leu Phe Ile Gln Leu Val Gln Thr Tyr Asn
180 185 190

Gln Leu Phe Glu Glu Asn Pro Ile Asn Ala Ser Gly Val Asp Ala Lys
195 200 205

Ala Ile Leu Ser Ala Arg Leu Ser Lys Ser Arg Arg Leu Glu Asn Leu
210 215 220

Ile Ala Gln Leu Pro Gly Glu Lys Lys Asn Gly Leu Phe Gly Asn Leu
225 230 235 240

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

Leu Ala Glu Asp Ala Lys Leu Gln Leu Ser Lys Asp Thr Tyr Asp Asp
260 265 270

Asp Leu Asp Asn Leu Leu Ala Gln Ile Gly Asp Gln Tyr Ala Asp Leu
275 280 285

Phe Leu Ala Ala Lys Asn Leu Ser Asp Ala Ile Leu Leu Ser Asp Ile
290 295 300

Leu Arg Val Asn Thr Glu Ile Thr Lys Ala Pro Leu Ser Ala Ser Met
305 310 315 320

Ile Lys Arg Tyr Asp Glu His His Gln Asp Leu Thr Leu Leu Lys Ala
325 330 335

Leu Val Arg Gln Gln Leu Pro Glu Lys Tyr Lys Glu Ile Phe Phe Asp
340 345 350

Gln Ser Lys Asn Gly Tyr Ala Gly Tyr Ile Asp Gly Gly Ala Ser Gln
355 360 365

Glu Glu Phe Tyr Lys Phe Ile Lys Pro Ile Leu Glu Lys Met Asp Gly
370 375 380

Thr Glu Glu Leu Leu Val Lys Leu Asn Arg Glu Asp Leu Leu Arg Lys
385 390 395 400

Gln Arg Thr Phe Asp Asn Gly Ser Ile Pro His Gln Ile His Leu Gly
405 410 415

Glu Leu His Ala Ile Leu Arg Arg Gln Glu Asp Phe Tyr Pro Phe Leu
420 425 430

Lys Asp Asn Arg Glu Lys Ile Glu Lys Ile Leu Thr Phe Arg Ile Pro
435 440 445
Tyr Tyr Val Gly Pro Leu Ala Arg Gly Asn Ser Arg Phe Ala Trp Met
450 455 460
Thr Arg Lys Ser Glu Glu Thr Ile Thr Pro Trp Asn Phe Glu Glu Val
465 470 475 480
Val Asp Lys Gly Ala Ser Ala Gln Ser Phe Ile Glu Arg Met Thr Asn
485 490 495
Phe Asp Lys Asn Leu Pro Asn Glu Lys Val Leu Pro Lys His Ser Leu
500 505 510
Leu Tyr Glu Tyr Phe Thr Val Tyr Asn Glu Leu Thr Lys Val Lys Tyr
515 520 525
Val Thr Glu Gly Met Arg Lys Pro Ala Phe Leu Ser Gly Glu Gln Lys
530 535 540
Lys Ala Ile Val Asp Leu Leu Phe Lys Thr Asn Arg Lys Val Thr Val
545 550 555 560
Lys Gln Leu Lys Glu Asp Tyr Phe Lys Lys Ile Glu Cys Phe Asp Ser
565 570 575
Val Glu Ile Ser Gly Val Glu Asp Arg Phe Asn Ala Ser Leu Gly Thr
580 585 590
Tyr His Asp Leu Leu Lys Ile Ile Lys Asp Lys Asp Phe Leu Asp Asn
595 600 605
Glu Glu Asn Glu Asp Ile Leu Glu Asp Ile Val Leu Thr Leu Thr Leu
610 615 620
Phe Glu Asp Arg Glu Met Ile Glu Glu Arg Leu Lys Thr Tyr Ala His
625 630 635 640
Leu Phe Asp Asp Lys Val Met Lys Gln Leu Lys Arg Arg Arg Tyr Thr
645 650 655
Gly Trp Gly Arg Leu Ser Arg Lys Leu Ile Asn Gly Ile Arg Asp Lys
660 665 670
Gln Ser Gly Lys Thr Ile Leu Asp Phe Leu Lys Ser Asp Gly Phe Ala
675 680 685
Asn Arg Asn Phe Met Gln Leu Ile His Asp Asp Ser Leu Thr Phe Lys
690 695 700
Glu Asp Ile Gln Lys Ala Gln Val Ser Gly Gln Gly Asp Ser Leu His
705 710 715 720
Glu His Ile Ala Asn Leu Ala Gly Ser Pro Ala Ile Lys Lys Gly Ile
725 730 735
Leu Gln Thr Val Lys Val Val Asp Glu Leu Val Lys Val Met Gly Arg

740 745 750

His Lys Pro Glu Asn Ile Val Ile Glu Met Ala Arg Glu Asn Gln Thr

755 760 765

Thr Gln Lys Gly Gln Lys Asn Ser Arg Glu Arg Met Lys Arg Ile Glu

770 775 780

Glu Gly Ile Lys Glu Leu Gly Ser Gln Ile Leu Lys Glu His Pro Val

785 790 795 800

Glu Asn Thr Gln Leu Gln Asn Glu Lys Leu Tyr Leu Tyr Tyr Leu Gln

805 810 815

Asn Gly Arg Asp Met Tyr Val Asp Gln Glu Leu Asp Ile Asn Arg Leu

820 825 830

Ser Asp Tyr Asp Val Asp His Ile Val Pro Gln Ser Phe Leu Lys Asp

835 840 845

Asp Ser Ile Asp Asn Lys Val Leu Thr Arg Ser Asp Lys Asn Arg Gly

850 855 860

Lys Ser Asp Asn Val Pro Ser Glu Glu Val Val Lys Lys Met Lys Asn

865 870 875 880

Tyr Trp Arg Gln Leu Leu Asn Ala Lys Leu Ile Thr Gln Arg Lys Phe

885 890 895

Asp Asn Leu Thr Lys Ala Glu Arg Gly Gly Leu Ser Glu Leu Asp Lys

900 905 910

Ala Gly Phe Ile Lys Arg Gln Leu Val Glu Thr Arg Gln Ile Thr Lys

915 920 925

His Val Ala Gln Ile Leu Asp Ser Arg Met Asn Thr Lys Tyr Asp Glu

930 935 940

Asn Asp Lys Leu Ile Arg Glu Val Lys Val Ile Thr Leu Lys Ser Lys

945 950 955 960

Leu Val Ser Asp Phe Arg Lys Asp Phe Gln Phe Tyr Lys Val Arg Glu

965 970 975

Ile Asn Asn Tyr His His Ala His Asp Ala Tyr Leu Asn Ala Val Val

980 985 990

Gly Thr Ala Leu Ile Lys Lys Tyr Pro Lys Leu Glu Ser Glu Phe Val

995 1000 1005

Tyr Gly Asp Tyr Lys Val Tyr Asp Val Arg Lys Met Ile Ala Lys

1010 1015 1020

Ser Glu Gln Glu Ile Gly Lys Ala Thr Ala Lys Tyr Phe Phe Tyr

1025 1030 1035

Ser Asn Ile Met Asn Phe Phe Lys Thr Glu Ile Thr Leu Ala Asn

1040 1045 1050

Gly Glu Ile Arg Lys Arg Pro Leu Ile Glu Thr Asn Gly Glu Thr
1055 1060 1065
Gly Glu Ile Val Trp Asp Lys Gly Arg Asp Phe Ala Thr Val Arg
1070 1075 1080
Lys Val Leu Ser Met Pro Gln Val Asn Ile Val Lys Lys Thr Glu
1085 1090 1095
Val Gln Thr Gly Gly Phe Ser Lys Glu Ser Ile Leu Pro Lys Arg
1100 1105 1110
Asn Ser Asp Lys Leu Ile Ala Arg Lys Lys Asp Trp Asp Pro Lys
1115 1120 1125
Lys Tyr Gly Gly Phe Asp Ser Pro Thr Val Ala Tyr Ser Val Leu
1130 1135 1140
Val Val Ala Lys Val Glu Lys Gly Lys Ser Lys Lys Leu Lys Ser
1145 1150 1155
Val Lys Glu Leu Leu Gly Ile Thr Ile Met Glu Arg Ser Ser Phe
1160 1165 1170
Glu Lys Asn Pro Ile Asp Phe Leu Glu Ala Lys Gly Tyr Lys Glu
1175 1180 1185
Val Lys Lys Asp Leu Ile Ile Lys Leu Pro Lys Tyr Ser Leu Phe
1190 1195 1200
Glu Leu Glu Asn Gly Arg Lys Arg Met Leu Ala Ser Ala Gly Glu
1205 1210 1215
Leu Gln Lys Gly Asn Glu Leu Ala Leu Pro Ser Lys Tyr Val Asn
1220 1225 1230
Phe Leu Tyr Leu Ala Ser His Tyr Glu Lys Leu Lys Gly Ser Pro
1235 1240 1245
Glu Asp Asn Glu Gln Lys Gln Leu Phe Val Glu Gln His Lys His
1250 1255 1260
Tyr Leu Asp Glu Ile Ile Glu Gln Ile Ser Glu Phe Ser Lys Arg
1265 1270 1275
Val Ile Leu Ala Asp Ala Asn Leu Asp Lys Val Leu Ser Ala Tyr
1280 1285 1290
Asn Lys His Arg Asp Lys Pro Ile Arg Glu Gln Ala Glu Asn Ile
1295 1300 1305
Ile His Leu Phe Thr Leu Thr Asn Leu Gly Ala Pro Ala Ala Phe
1310 1315 1320
Lys Tyr Phe Asp Thr Thr Ile Asp Arg Lys Arg Tyr Thr Ser Thr
1325 1330 1335
Lys Glu Val Leu Asp Ala Thr Leu Ile His Gln Ser Ile Thr Gly

1340 1345 1350

Leu Tyr Glu Thr Arg Ile Asp Leu Ser Gln Leu Gly Gly Asp

1355 1360 1365

<210> 17

<211> 1367

<212> PRT

<213> 人工序列

<220>

<223> 合成多肽

<400> 17

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

225 230 235 240
Ile Ala Leu Ser Leu Gly Leu Thr Pro Asn Phe Lys Ser Asn Phe Asp
245 250 255
Leu Ala Glu Asp Ala Lys Leu Gln Leu Ser Lys Asp Thr Tyr Asp Asp
260 265 270
Asp Leu Asp Asn Leu Leu Ala Gln Ile Gly Asp Gln Tyr Ala Asp Leu
275 280 285
Phe Leu Ala Ala Lys Asn Leu Ser Asp Ala Ile Leu Leu Ser Asp Ile
290 295 300
Leu Arg Val Asn Thr Glu Ile Thr Lys Ala Pro Leu Ser Ala Ser Met
305 310 315 320
Ile Lys Arg Tyr Asp Glu His His Gln Asp Leu Thr Leu Leu Lys Ala
325 330 335
Leu Val Arg Gln Gln Leu Pro Glu Lys Tyr Lys Glu Ile Phe Phe Asp
340 345 350
Gln Ser Lys Asn Gly Tyr Ala Gly Tyr Ile Asp Gly Gly Ala Ser Gln
355 360 365
Glu Glu Phe Tyr Lys Phe Ile Lys Pro Ile Leu Glu Lys Met Asp Gly
370 375 380
Thr Glu Glu Leu Leu Val Lys Leu Asn Arg Glu Asp Leu Leu Arg Lys
385 390 395 400
Gln Arg Thr Phe Asp Asn Gly Ser Ile Pro His Gln Ile His Leu Gly
405 410 415
Glu Leu His Ala Ile Leu Arg Arg Gln Glu Asp Phe Tyr Pro Phe Leu
420 425 430
Lys Asp Asn Arg Glu Lys Ile Glu Lys Ile Leu Thr Phe Arg Ile Pro
435 440 445
Tyr Tyr Val Gly Pro Leu Ala Arg Gly Asn Ser Arg Phe Ala Trp Met
450 455 460
Thr Arg Lys Ser Glu Glu Thr Ile Thr Pro Trp Asn Phe Glu Glu Val
465 470 475 480
Val Asp Lys Gly Ala Ser Ala Gln Ser Phe Ile Glu Arg Met Thr Asn
485 490 495
Phe Asp Lys Asn Leu Pro Asn Glu Lys Val Leu Pro Lys His Ser Leu
500 505 510
Leu Tyr Glu Tyr Phe Thr Val Tyr Asn Glu Leu Thr Lys Val Lys Tyr
515 520 525
Val Thr Glu Gly Met Arg Lys Pro Ala Phe Leu Ser Gly Glu Gln Lys
530 535 540

Lys Ala Ile Val Asp Leu Leu Phe Lys Thr Asn Arg Lys Val Thr Val
545 550 555 560
Lys Gln Leu Lys Glu Asp Tyr Phe Lys Lys Ile Glu Cys Phe Asp Ser
565 570 575
Val Glu Ile Ser Gly Val Glu Asp Arg Phe Asn Ala Ser Leu Gly Thr
580 585 590
Tyr His Asp Leu Leu Lys Ile Ile Lys Asp Lys Asp Phe Leu Asp Asn
595 600 605
Glu Glu Asn Glu Asp Ile Leu Glu Asp Ile Val Leu Thr Leu Thr Leu
610 615 620
Phe Glu Asp Arg Glu Met Ile Glu Glu Arg Leu Lys Thr Tyr Ala His
625 630 635 640
Leu Phe Asp Asp Lys Val Met Lys Gln Leu Lys Arg Arg Arg Tyr Thr
645 650 655
Gly Trp Gly Arg Leu Ser Arg Lys Leu Ile Asn Gly Ile Arg Asp Lys
660 665 670
Gln Ser Gly Lys Thr Ile Leu Asp Phe Leu Lys Ser Asp Gly Phe Ala
675 680 685
Asn Arg Asn Phe Met Gln Leu Ile His Asp Asp Ser Leu Thr Phe Lys
690 695 700
Glu Asp Ile Gln Lys Ala Gln Val Ser Gly Gln Gly Asp Ser Leu His
705 710 715 720
Glu His Ile Ala Asn Leu Ala Gly Ser Pro Ala Ile Lys Lys Gly Ile
725 730 735
Leu Gln Thr Val Lys Val Val Asp Glu Leu Val Lys Val Met Gly Arg
740 745 750
His Lys Pro Glu Asn Ile Val Ile Glu Met Ala Arg Glu Asn Gln Thr
755 760 765
Thr Gln Lys Gly Gln Lys Asn Ser Arg Glu Arg Met Lys Arg Ile Glu
770 775 780
Glu Gly Ile Lys Glu Leu Gly Ser Gln Ile Leu Lys Glu His Pro Val
785 790 795 800
Glu Asn Thr Gln Leu Gln Asn Glu Lys Leu Tyr Leu Tyr Tyr Leu Gln
805 810 815
Asn Gly Arg Asp Met Tyr Val Asp Gln Glu Leu Asp Ile Asn Arg Leu
820 825 830
Ser Asp Tyr Asp Val Asp His Ile Val Pro Gln Ser Phe Leu Lys Asp
835 840 845
Asp Ser Ile Asp Asn Lys Val Leu Thr Arg Ser Asp Lys Asn Arg Gly

850 855 860

Lys Ser Asp Asn Val Pro Ser Glu Glu Val Val Lys Lys Met Lys Asn
865 870 875 880

Tyr Trp Arg Gln Leu Leu Asn Ala Lys Leu Ile Thr Gln Arg Lys Phe
885 890 895

Asp Asn Leu Thr Lys Ala Glu Arg Gly Gly Leu Ser Glu Leu Asp Lys
900 905 910

Ala Gly Phe Ile Lys Arg Gln Leu Val Glu Thr Arg Gln Ile Thr Lys
915 920 925

His Val Ala Gln Ile Leu Asp Ser Arg Met Asn Thr Lys Tyr Asp Glu
930 935 940

Asn Asp Lys Leu Ile Arg Glu Val Lys Val Ile Thr Leu Lys Ser Lys
945 950 955 960

Leu Val Ser Asp Phe Arg Lys Asp Phe Gln Phe Tyr Lys Val Arg Glu
965 970 975

Ile Asn Asn Tyr His His Ala His Asp Ala Tyr Leu Asn Ala Val Val
980 985 990

Gly Thr Ala Leu Ile Lys Lys Tyr Pro Lys Leu Glu Ser Glu Phe Val
995 1000 1005

Tyr Gly Asp Tyr Lys Val Tyr Asp Val Arg Lys Met Ile Ala Lys
1010 1015 1020

Ser Glu Gln Glu Ile Gly Lys Ala Thr Ala Lys Tyr Phe Phe Tyr
1025 1030 1035

Ser Asn Ile Met Asn Phe Phe Lys Thr Glu Ile Thr Leu Ala Asn
1040 1045 1050

Gly Glu Ile Arg Lys Arg Pro Leu Ile Glu Thr Asn Gly Glu Thr
1055 1060 1065

Gly Glu Ile Val Trp Asp Lys Gly Arg Asp Phe Ala Thr Val Arg
1070 1075 1080

Lys Val Leu Ser Met Pro Gln Val Asn Ile Val Lys Lys Thr Glu
1085 1090 1095

Val Gln Thr Gly Gly Phe Ser Lys Glu Ser Ile Leu Pro Lys Arg
1100 1105 1110

Asn Ser Asp Lys Leu Ile Ala Arg Lys Lys Asp Trp Asp Pro Lys
1115 1120 1125

Lys Tyr Gly Gly Phe Glu Ser Pro Thr Val Ala Tyr Ser Val Leu
1130 1135 1140

Val Val Ala Lys Val Glu Lys Gly Lys Ser Lys Lys Leu Lys Ser
1145 1150 1155

Val Lys Glu Leu Leu Gly Ile Thr Ile Met Glu Arg Ser Ser Phe
1160 1165 1170
Glu Lys Asn Pro Ile Asp Phe Leu Glu Ala Lys Gly Tyr Lys Glu
1175 1180 1185
Val Lys Lys Asp Leu Ile Ile Lys Leu Pro Lys Tyr Ser Leu Phe
1190 1195 1200
Glu Leu Glu Asn Gly Arg Lys Arg Met Leu Ala Ser Ala Gly Glu
1205 1210 1215
Leu Gln Lys Gly Asn Glu Leu Ala Leu Pro Ser Lys Tyr Val Asn
1220 1225 1230
Phe Leu Tyr Leu Ala Ser His Tyr Glu Lys Leu Lys Gly Ser Pro
1235 1240 1245
Glu Asp Asn Glu Gln Lys Gln Leu Phe Val Glu Gln His Lys His
1250 1255 1260
Tyr Leu Asp Glu Ile Ile Glu Gln Ile Ser Glu Phe Ser Lys Arg
1265 1270 1275
Val Ile Leu Ala Asp Ala Asn Leu Asp Lys Val Leu Ser Ala Tyr
1280 1285 1290
Asn Lys His Arg Asp Lys Pro Ile Arg Glu Gln Ala Glu Asn Ile
1295 1300 1305
Ile His Leu Phe Thr Leu Thr Asn Leu Gly Ala Pro Ala Ala Phe
1310 1315 1320
Lys Tyr Phe Asp Thr Thr Ile Asp Arg Lys Gln Tyr Arg Ser Thr
1325 1330 1335
Lys Glu Val Leu Asp Ala Thr Leu Ile His Gln Ser Ile Thr Gly
1340 1345 1350
Leu Tyr Glu Thr Arg Ile Asp Leu Ser Gln Leu Gly Gly Asp
1355 1360 1365

<210> 18
<211> 1367
<212> PRT
<213> 人工序列
<220>
<223> 合成多肽
<400> 18

Asp Lys Lys Tyr Ser Ile Gly Leu Ala Ile Gly Thr Asn Ser Val Gly
1 5 10 15
Trp Ala Val Ile Thr Asp Glu Tyr Lys Val Pro Ser Lys Lys Phe Lys
20 25 30

Val Leu Gly Asn Thr Asp Arg His Ser Ile Lys Lys Asn Leu Ile Gly
35 40 45
Ala Leu Leu Phe Asp Ser Gly Glu Thr Ala Glu Ala Thr Arg Leu Lys
50 55 60
Arg Thr Ala Arg Arg Arg Tyr Thr Arg Arg Lys Asn Arg Ile Cys Tyr
65 70 75 80
Leu Gln Glu Ile Phe Ser Asn Glu Met Ala Lys Val Asp Asp Ser Phe
85 90 95
Phe His Arg Leu Glu Glu Ser Phe Leu Val Glu Glu Asp Lys Lys His
100 105 110
Glu Arg His Pro Ile Phe Gly Asn Ile Val Asp Glu Val Ala Tyr His
115 120 125
Glu Lys Tyr Pro Thr Ile Tyr His Leu Arg Lys Lys Leu Val Asp Ser
130 135 140
Thr Asp Lys Ala Asp Leu Arg Leu Ile Tyr Leu Ala Leu Ala His Met
145 150 155 160
Ile Lys Phe Arg Gly His Phe Leu Ile Glu Gly Asp Leu Asn Pro Asp
165 170 175
Asn Ser Asp Val Asp Lys Leu Phe Ile Gln Leu Val Gln Thr Tyr Asn
180 185 190
Gln Leu Phe Glu Glu Asn Pro Ile Asn Ala Ser Gly Val Asp Ala Lys
195 200 205
Ala Ile Leu Ser Ala Arg Leu Ser Lys Ser Arg Arg Leu Glu Asn Leu
210 215 220
Ile Ala Gln Leu Pro Gly Glu Lys Lys Asn Gly Leu Phe Gly Asn Leu
225 230 235 240
Ile Ala Leu Ser Leu Gly Leu Thr Pro Asn Phe Lys Ser Asn Phe Asp
245 250 255
Leu Ala Glu Asp Ala Lys Leu Gln Leu Ser Lys Asp Thr Tyr Asp Asp
260 265 270
Asp Leu Asp Asn Leu Leu Ala Gln Ile Gly Asp Gln Tyr Ala Asp Leu
275 280 285
Phe Leu Ala Ala Lys Asn Leu Ser Asp Ala Ile Leu Leu Ser Asp Ile
290 295 300
Leu Arg Val Asn Thr Glu Ile Thr Lys Ala Pro Leu Ser Ala Ser Met
305 310 315 320
Ile Lys Arg Tyr Asp Glu His His Gln Asp Leu Thr Leu Leu Lys Ala
325 330 335
Leu Val Arg Gln Gln Leu Pro Glu Lys Tyr Lys Glu Ile Phe Phe Asp

340 345 350

Gln Ser Lys Asn Gly Tyr Ala Gly Tyr Ile Asp Gly Gly Ala Ser Gln

355 360 365

Glu Glu Phe Tyr Lys Phe Ile Lys Pro Ile Leu Glu Lys Met Asp Gly

370 375 380

Thr Glu Glu Leu Leu Val Lys Leu Asn Arg Glu Asp Leu Leu Arg Lys

385 390 395 400

Gln Arg Thr Phe Asp Asn Gly Ser Ile Pro His Gln Ile His Leu Gly

405 410 415

Glu Leu His Ala Ile Leu Arg Arg Gln Glu Asp Phe Tyr Pro Phe Leu

420 425 430

Lys Asp Asn Arg Glu Lys Ile Glu Lys Ile Leu Thr Phe Arg Ile Pro

435 440 445

Tyr Tyr Val Gly Pro Leu Ala Arg Gly Asn Ser Arg Phe Ala Trp Met

450 455 460

Thr Arg Lys Ser Glu Glu Thr Ile Thr Pro Trp Asn Phe Glu Glu Val

465 470 475 480

Val Asp Lys Gly Ala Ser Ala Gln Ser Phe Ile Glu Arg Met Thr Asn

485 490 495

Phe Asp Lys Asn Leu Pro Asn Glu Lys Val Leu Pro Lys His Ser Leu

500 505 510

Leu Tyr Glu Tyr Phe Thr Val Tyr Asn Glu Leu Thr Lys Val Lys Tyr

515 520 525

Val Thr Glu Gly Met Arg Lys Pro Ala Phe Leu Ser Gly Glu Gln Lys

530 535 540

Lys Ala Ile Val Asp Leu Leu Phe Lys Thr Asn Arg Lys Val Thr Val

545 550 555 560

Lys Gln Leu Lys Glu Asp Tyr Phe Lys Lys Ile Glu Cys Phe Asp Ser

565 570 575

Val Glu Ile Ser Gly Val Glu Asp Arg Phe Asn Ala Ser Leu Gly Thr

580 585 590

Tyr His Asp Leu Leu Lys Ile Ile Lys Asp Lys Asp Phe Leu Asp Asn

595 600 605

Glu Glu Asn Glu Asp Ile Leu Glu Asp Ile Val Leu Thr Leu Thr Leu

610 615 620

Phe Glu Asp Arg Glu Met Ile Glu Glu Arg Leu Lys Thr Tyr Ala His

625 630 635 640

Leu Phe Asp Asp Lys Val Met Lys Gln Leu Lys Arg Arg Arg Tyr Thr

645 650 655

Gly Trp Gly Arg Leu Ser Arg Lys Leu Ile Asn Gly Ile Arg Asp Lys
660 665 670
Gln Ser Gly Lys Thr Ile Leu Asp Phe Leu Lys Ser Asp Gly Phe Ala
675 680 685
Asn Arg Asn Phe Met Gln Leu Ile His Asp Asp Ser Leu Thr Phe Lys
690 695 700
Glu Asp Ile Gln Lys Ala Gln Val Ser Gly Gln Gly Asp Ser Leu His
705 710 715 720
Glu His Ile Ala Asn Leu Ala Gly Ser Pro Ala Ile Lys Lys Gly Ile
725 730 735
Leu Gln Thr Val Lys Val Val Asp Glu Leu Val Lys Val Met Gly Arg
740 745 750
His Lys Pro Glu Asn Ile Val Ile Glu Met Ala Arg Glu Asn Gln Thr
755 760 765
Thr Gln Lys Gly Gln Lys Asn Ser Arg Glu Arg Met Lys Arg Ile Glu
770 775 780
Glu Gly Ile Lys Glu Leu Gly Ser Gln Ile Leu Lys Glu His Pro Val
785 790 795 800
Glu Asn Thr Gln Leu Gln Asn Glu Lys Leu Tyr Leu Tyr Tyr Leu Gln
805 810 815
Asn Gly Arg Asp Met Tyr Val Asp Gln Glu Leu Asp Ile Asn Arg Leu
820 825 830
Ser Asp Tyr Asp Val Asp His Ile Val Pro Gln Ser Phe Leu Lys Asp
835 840 845
Asp Ser Ile Asp Asn Lys Val Leu Thr Arg Ser Asp Lys Asn Arg Gly
850 855 860
Lys Ser Asp Asn Val Pro Ser Glu Glu Val Val Lys Lys Met Lys Asn
865 870 875 880
Tyr Trp Arg Gln Leu Leu Asn Ala Lys Leu Ile Thr Gln Arg Lys Phe
885 890 895
Asp Asn Leu Thr Lys Ala Glu Arg Gly Gly Leu Ser Glu Leu Asp Lys
900 905 910
Ala Gly Phe Ile Lys Arg Gln Leu Val Glu Thr Arg Gln Ile Thr Lys
915 920 925
His Val Ala Gln Ile Leu Asp Ser Arg Met Asn Thr Lys Tyr Asp Glu
930 935 940
Asn Asp Lys Leu Ile Arg Glu Val Lys Val Ile Thr Leu Lys Ser Lys
945 950 955 960
Leu Val Ser Asp Phe Arg Lys Asp Phe Gln Phe Tyr Lys Val Arg Glu

965 970 975

Ile Asn Asn Tyr His His Ala His Asp Ala Tyr Leu Asn Ala Val Val

980 985 990

Gly Thr Ala Leu Ile Lys Lys Tyr Pro Lys Leu Glu Ser Glu Phe Val

995 1000 1005

Tyr Gly Asp Tyr Lys Val Tyr Asp Val Arg Lys Met Ile Ala Lys

1010 1015 1020

Ser Glu Gln Glu Ile Gly Lys Ala Thr Ala Lys Tyr Phe Phe Tyr

1025 1030 1035

Ser Asn Ile Met Asn Phe Phe Lys Thr Glu Ile Thr Leu Ala Asn

1040 1045 1050

Gly Glu Ile Arg Lys Arg Pro Leu Ile Glu Thr Asn Gly Glu Thr

1055 1060 1065

Gly Glu Ile Val Trp Asp Lys Gly Arg Asp Phe Ala Thr Val Arg

1070 1075 1080

Lys Val Leu Ser Met Pro Gln Val Asn Ile Val Lys Lys Thr Glu

1085 1090 1095

Val Gln Thr Gly Gly Phe Ser Lys Glu Ser Ile Leu Pro Lys Arg

1100 1105 1110

Asn Ser Asp Lys Leu Ile Ala Arg Lys Lys Asp Trp Asp Pro Lys

1115 1120 1125

Lys Tyr Gly Gly Phe Val Ser Pro Thr Val Ala Tyr Ser Val Leu

1130 1135 1140

Val Val Ala Lys Val Glu Lys Gly Lys Ser Lys Lys Leu Lys Ser

1145 1150 1155

Val Lys Glu Leu Leu Gly Ile Thr Ile Met Glu Arg Ser Ser Phe

1160 1165 1170

Glu Lys Asn Pro Ile Asp Phe Leu Glu Ala Lys Gly Tyr Lys Glu

1175 1180 1185

Val Lys Lys Asp Leu Ile Ile Lys Leu Pro Lys Tyr Ser Leu Phe

1190 1195 1200

Glu Leu Glu Asn Gly Arg Lys Arg Met Leu Ala Ser Ala Gly Glu

1205 1210 1215

Leu Gln Lys Gly Asn Glu Leu Ala Leu Pro Ser Lys Tyr Val Asn

1220 1225 1230

Phe Leu Tyr Leu Ala Ser His Tyr Glu Lys Leu Lys Gly Ser Pro

1235 1240 1245

Glu Asp Asn Glu Gln Lys Gln Leu Phe Val Glu Gln His Lys His

1250 1255 1260

Tyr Leu Asp Glu Ile Ile Glu Gln Ile Ser Glu Phe Ser Lys Arg
1265 1270 1275
Val Ile Leu Ala Asp Ala Asn Leu Asp Lys Val Leu Ser Ala Tyr
1280 1285 1290
Asn Lys His Arg Asp Lys Pro Ile Arg Glu Gln Ala Glu Asn Ile
1295 1300 1305
Ile His Leu Phe Thr Leu Thr Asn Leu Gly Ala Pro Ala Ala Phe
1310 1315 1320
Lys Tyr Phe Asp Thr Thr Ile Asp Arg Lys Gln Tyr Arg Ser Thr
1325 1330 1335
Lys Glu Val Leu Asp Ala Thr Leu Ile His Gln Ser Ile Thr Gly
1340 1345 1350
Leu Tyr Glu Thr Arg Ile Asp Leu Ser Gln Leu Gly Gly Asp
1355 1360 1365

<210> 19
<211> 1367
<212> PRT
<213> 人工序列
<220>
<223> 合成多肽
<400> 19

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

Thr Asp Lys Ala Asp Leu Arg Leu Ile Tyr Leu Ala Leu Ala His Met
145 150 155 160
Ile Lys Phe Arg Gly His Phe Leu Ile Glu Gly Asp Leu Asn Pro Asp
165 170 175
Asn Ser Asp Val Asp Lys Leu Phe Ile Gln Leu Val Gln Thr Tyr Asn
180 185 190
Gln Leu Phe Glu Glu Asn Pro Ile Asn Ala Ser Gly Val Asp Ala Lys
195 200 205
Ala Ile Leu Ser Ala Arg Leu Ser Lys Ser Arg Arg Leu Glu Asn Leu
210 215 220
Ile Ala Gln Leu Pro Gly Glu Lys Lys Asn Gly Leu Phe Gly Asn Leu
225 230 235 240
Ile Ala Leu Ser Leu Gly Leu Thr Pro Asn Phe Lys Ser Asn Phe Asp
245 250 255
Leu Ala Glu Asp Ala Lys Leu Gln Leu Ser Lys Asp Thr Tyr Asp Asp
260 265 270
Asp Leu Asp Asn Leu Leu Ala Gln Ile Gly Asp Gln Tyr Ala Asp Leu
275 280 285
Phe Leu Ala Ala Lys Asn Leu Ser Asp Ala Ile Leu Leu Ser Asp Ile
290 295 300
Leu Arg Val Asn Thr Glu Ile Thr Lys Ala Pro Leu Ser Ala Ser Met
305 310 315 320
Ile Lys Arg Tyr Asp Glu His His Gln Asp Leu Thr Leu Leu Lys Ala
325 330 335
Leu Val Arg Gln Gln Leu Pro Glu Lys Tyr Lys Glu Ile Phe Phe Asp
340 345 350
Gln Ser Lys Asn Gly Tyr Ala Gly Tyr Ile Asp Gly Gly Ala Ser Gln
355 360 365
Glu Glu Phe Tyr Lys Phe Ile Lys Pro Ile Leu Glu Lys Met Asp Gly
370 375 380
Thr Glu Glu Leu Leu Val Lys Leu Asn Arg Glu Asp Leu Leu Arg Lys
385 390 395 400
Gln Arg Thr Phe Asp Asn Gly Ser Ile Pro His Gln Ile His Leu Gly
405 410 415
Glu Leu His Ala Ile Leu Arg Arg Gln Glu Asp Phe Tyr Pro Phe Leu
420 425 430
Lys Asp Asn Arg Glu Lys Ile Glu Lys Ile Leu Thr Phe Arg Ile Pro
435 440 445
Tyr Tyr Val Gly Pro Leu Ala Arg Gly Asn Ser Arg Phe Ala Trp Met

450 455 460
Thr Arg Lys Ser Glu Glu Thr Ile Thr Pro Trp Asn Phe Glu Glu Val
465 470 475 480
Val Asp Lys Gly Ala Ser Ala Gln Ser Phe Ile Glu Arg Met Thr Asn
485 490 495
Phe Asp Lys Asn Leu Pro Asn Glu Lys Val Leu Pro Lys His Ser Leu
500 505 510
Leu Tyr Glu Tyr Phe Thr Val Tyr Asn Glu Leu Thr Lys Val Lys Tyr
515 520 525
Val Thr Glu Gly Met Arg Lys Pro Ala Phe Leu Ser Gly Glu Gln Lys
530 535 540
Lys Ala Ile Val Asp Leu Leu Phe Lys Thr Asn Arg Lys Val Thr Val
545 550 555 560
Lys Gln Leu Lys Glu Asp Tyr Phe Lys Lys Ile Glu Cys Phe Asp Ser
565 570 575
Val Glu Ile Ser Gly Val Glu Asp Arg Phe Asn Ala Ser Leu Gly Thr
580 585 590
Tyr His Asp Leu Leu Lys Ile Ile Lys Asp Lys Asp Phe Leu Asp Asn
595 600 605
Glu Glu Asn Glu Asp Ile Leu Glu Asp Ile Val Leu Thr Leu Thr Leu
610 615 620
Phe Glu Asp Arg Glu Met Ile Glu Glu Arg Leu Lys Thr Tyr Ala His
625 630 635 640
Leu Phe Asp Asp Lys Val Met Lys Gln Leu Lys Arg Arg Arg Tyr Thr
645 650 655
Gly Trp Gly Arg Leu Ser Arg Lys Leu Ile Asn Gly Ile Arg Asp Lys
660 665 670
Gln Ser Gly Lys Thr Ile Leu Asp Phe Leu Lys Ser Asp Gly Phe Ala
675 680 685
Asn Arg Asn Phe Met Gln Leu Ile His Asp Asp Ser Leu Thr Phe Lys
690 695 700
Glu Asp Ile Gln Lys Ala Gln Val Ser Gly Gln Gly Asp Ser Leu His
705 710 715 720
Glu His Ile Ala Asn Leu Ala Gly Ser Pro Ala Ile Lys Lys Gly Ile
725 730 735
Leu Gln Thr Val Lys Val Val Asp Glu Leu Val Lys Val Met Gly Arg
740 745 750
His Lys Pro Glu Asn Ile Val Ile Glu Met Ala Arg Glu Asn Gln Thr
755 760 765

Thr Gln Lys Gly Gln Lys Asn Ser Arg Glu Arg Met Lys Arg Ile Glu
770 775 780
Glu Gly Ile Lys Glu Leu Gly Ser Gln Ile Leu Lys Glu His Pro Val
785 790 795 800
Glu Asn Thr Gln Leu Gln Asn Glu Lys Leu Tyr Leu Tyr Tyr Leu Gln
805 810 815
Asn Gly Arg Asp Met Tyr Val Asp Gln Glu Leu Asp Ile Asn Arg Leu
820 825 830
Ser Asp Tyr Asp Val Asp His Ile Val Pro Gln Ser Phe Leu Lys Asp
835 840 845
Asp Ser Ile Asp Asn Lys Val Leu Thr Arg Ser Asp Lys Asn Arg Gly
850 855 860
Lys Ser Asp Asn Val Pro Ser Glu Glu Val Val Lys Lys Met Lys Asn
865 870 875 880
Tyr Trp Arg Gln Leu Leu Asn Ala Lys Leu Ile Thr Gln Arg Lys Phe
885 890 895
Asp Asn Leu Thr Lys Ala Glu Arg Gly Gly Leu Ser Glu Leu Asp Lys
900 905 910
Ala Gly Phe Ile Lys Arg Gln Leu Val Glu Thr Arg Gln Ile Thr Lys
915 920 925
His Val Ala Gln Ile Leu Asp Ser Arg Met Asn Thr Lys Tyr Asp Glu
930 935 940
Asn Asp Lys Leu Ile Arg Glu Val Lys Val Ile Thr Leu Lys Ser Lys
945 950 955 960
Leu Val Ser Asp Phe Arg Lys Asp Phe Gln Phe Tyr Lys Val Arg Glu
965 970 975
Ile Asn Asn Tyr His His Ala His Asp Ala Tyr Leu Asn Ala Val Val
980 985 990
Gly Thr Ala Leu Ile Lys Lys Tyr Pro Lys Leu Glu Ser Glu Phe Val
995 1000 1005
Tyr Gly Asp Tyr Lys Val Tyr Asp Val Arg Lys Met Ile Ala Lys
1010 1015 1020
Ser Glu Gln Glu Ile Gly Lys Ala Thr Ala Lys Tyr Phe Phe Tyr
1025 1030 1035
Ser Asn Ile Met Asn Phe Phe Lys Thr Glu Ile Thr Leu Ala Asn
1040 1045 1050
Gly Glu Ile Arg Lys Arg Pro Leu Ile Glu Thr Asn Gly Glu Thr
1055 1060 1065
Gly Glu Ile Val Trp Asp Lys Gly Arg Asp Phe Ala Thr Val Arg

1070 1075 1080
Lys Val Leu Ser Met Pro Gln Val Asn Ile Val Lys Lys Thr Glu
1085 1090 1095
Val Gln Thr Gly Gly Phe Ser Lys Glu Ser Ile Leu Pro Lys Arg
1100 1105 1110
Asn Ser Asp Lys Leu Ile Ala Arg Lys Lys Asp Trp Asp Pro Lys
1115 1120 1125
Lys Tyr Gly Gly Phe Val Ser Pro Thr Val Ala Tyr Ser Val Leu
1130 1135 1140
Val Val Ala Lys Val Glu Lys Gly Lys Ser Lys Lys Leu Lys Ser
1145 1150 1155
Val Lys Glu Leu Leu Gly Ile Thr Ile Met Glu Arg Ser Ser Phe
1160 1165 1170
Glu Lys Asn Pro Ile Asp Phe Leu Glu Ala Lys Gly Tyr Lys Glu
1175 1180 1185
Val Lys Lys Asp Leu Ile Ile Lys Leu Pro Lys Tyr Ser Leu Phe
1190 1195 1200
Glu Leu Glu Asn Gly Arg Lys Arg Met Leu Ala Ser Ala Arg Glu
1205 1210 1215
Leu Gln Lys Gly Asn Glu Leu Ala Leu Pro Ser Lys Tyr Val Asn
1220 1225 1230
Phe Leu Tyr Leu Ala Ser His Tyr Glu Lys Leu Lys Gly Ser Pro
1235 1240 1245
Glu Asp Asn Glu Gln Lys Gln Leu Phe Val Glu Gln His Lys His
1250 1255 1260
Tyr Leu Asp Glu Ile Ile Glu Gln Ile Ser Glu Phe Ser Lys Arg
1265 1270 1275
Val Ile Leu Ala Asp Ala Asn Leu Asp Lys Val Leu Ser Ala Tyr
1280 1285 1290
Asn Lys His Arg Asp Lys Pro Ile Arg Glu Gln Ala Glu Asn Ile
1295 1300 1305
Ile His Leu Phe Thr Leu Thr Asn Leu Gly Ala Pro Ala Ala Phe
1310 1315 1320
Lys Tyr Phe Asp Thr Thr Ile Asp Arg Lys Glu Tyr Arg Ser Thr
1325 1330 1335
Lys Glu Val Leu Asp Ala Thr Leu Ile His Gln Ser Ile Thr Gly
1340 1345 1350
Leu Tyr Glu Thr Arg Ile Asp Leu Ser Gln Leu Gly Gly Asp
1355 1360 1365

<210> 20
<211> 1367
<212> PRT
<213> 人工序列
<220>
<223> 合成多肽
<400> 20

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

Leu Ala Glu Asp Ala Lys Leu Gln Leu Ser Lys Asp Thr Tyr Asp Asp
260 265 270
Asp Leu Asp Asn Leu Leu Ala Gln Ile Gly Asp Gln Tyr Ala Asp Leu
275 280 285
Phe Leu Ala Ala Lys Asn Leu Ser Asp Ala Ile Leu Leu Ser Asp Ile
290 295 300
Leu Arg Val Asn Thr Glu Ile Thr Lys Ala Pro Leu Ser Ala Ser Met
305 310 315 320
Ile Lys Arg Tyr Asp Glu His His Gln Asp Leu Thr Leu Leu Lys Ala
325 330 335
Leu Val Arg Gln Gln Leu Pro Glu Lys Tyr Lys Glu Ile Phe Phe Asp
340 345 350
Gln Ser Lys Asn Gly Tyr Ala Gly Tyr Ile Asp Gly Gly Ala Ser Gln
355 360 365
Glu Glu Phe Tyr Lys Phe Ile Lys Pro Ile Leu Glu Lys Met Asp Gly
370 375 380
Thr Glu Glu Leu Leu Val Lys Leu Asn Arg Glu Asp Leu Leu Arg Lys
385 390 395 400
Gln Arg Thr Phe Asp Asn Gly Ser Ile Pro His Gln Ile His Leu Gly
405 410 415
Glu Leu His Ala Ile Leu Arg Arg Gln Glu Asp Phe Tyr Pro Phe Leu
420 425 430
Lys Asp Asn Arg Glu Lys Ile Glu Lys Ile Leu Thr Phe Arg Ile Pro
435 440 445
Tyr Tyr Val Gly Pro Leu Ala Arg Gly Asn Ser Arg Phe Ala Trp Met
450 455 460
Thr Arg Lys Ser Glu Glu Thr Ile Thr Pro Trp Asn Phe Glu Glu Val
465 470 475 480
Val Asp Lys Gly Ala Ser Ala Gln Ser Phe Ile Glu Arg Met Thr Ala
485 490 495
Phe Asp Lys Asn Leu Pro Asn Glu Lys Val Leu Pro Lys His Ser Leu
500 505 510
Leu Tyr Glu Tyr Phe Thr Val Tyr Asn Glu Leu Thr Lys Val Lys Tyr
515 520 525
Val Thr Glu Gly Met Arg Lys Pro Ala Phe Leu Ser Gly Glu Gln Lys
530 535 540
Lys Ala Ile Val Asp Leu Leu Phe Lys Thr Asn Arg Lys Val Thr Val
545 550 555 560
Lys Gln Leu Lys Glu Asp Tyr Phe Lys Lys Ile Glu Cys Phe Asp Ser

565 570 575
Val Glu Ile Ser Gly Val Glu Asp Arg Phe Asn Ala Ser Leu Gly Thr
580 585 590
Tyr His Asp Leu Leu Lys Ile Ile Lys Asp Lys Asp Phe Leu Asp Asn
595 600 605
Glu Glu Asn Glu Asp Ile Leu Glu Asp Ile Val Leu Thr Leu Thr Leu
610 615 620
Phe Glu Asp Arg Glu Met Ile Glu Glu Arg Leu Lys Thr Tyr Ala His
625 630 635 640
Leu Phe Asp Asp Lys Val Met Lys Gln Leu Lys Arg Arg Arg Tyr Thr
645 650 655
Gly Trp Gly Ala Leu Ser Arg Lys Leu Ile Asn Gly Ile Arg Asp Lys
660 665 670
Gln Ser Gly Lys Thr Ile Leu Asp Phe Leu Lys Ser Asp Gly Phe Ala
675 680 685
Asn Arg Asn Phe Met Ala Leu Ile His Asp Asp Ser Leu Thr Phe Lys
690 695 700
Glu Asp Ile Gln Lys Ala Gln Val Ser Gly Gln Gly Asp Ser Leu His
705 710 715 720
Glu His Ile Ala Asn Leu Ala Gly Ser Pro Ala Ile Lys Lys Gly Ile
725 730 735
Leu Gln Thr Val Lys Val Val Asp Glu Leu Val Lys Val Met Gly Arg
740 745 750
His Lys Pro Glu Asn Ile Val Ile Glu Met Ala Arg Glu Asn Gln Thr
755 760 765
Thr Gln Lys Gly Gln Lys Asn Ser Arg Glu Arg Met Lys Arg Ile Glu
770 775 780
Glu Gly Ile Lys Glu Leu Gly Ser Gln Ile Leu Lys Glu His Pro Val
785 790 795 800
Glu Asn Thr Gln Leu Gln Asn Glu Lys Leu Tyr Leu Tyr Tyr Leu Gln
805 810 815
Asn Gly Arg Asp Met Tyr Val Asp Gln Glu Leu Asp Ile Asn Arg Leu
820 825 830
Ser Asp Tyr Asp Val Asp His Ile Val Pro Gln Ser Phe Leu Lys Asp
835 840 845
Asp Ser Ile Asp Asn Lys Val Leu Thr Arg Ser Asp Lys Asn Arg Gly
850 855 860
Lys Ser Asp Asn Val Pro Ser Glu Glu Val Val Lys Lys Met Lys Asn
865 870 875 880

Tyr Trp Arg Gln Leu Leu Asn Ala Lys Leu Ile Thr Gln Arg Lys Phe
885 890 895
Asp Asn Leu Thr Lys Ala Glu Arg Gly Gly Leu Ser Glu Leu Asp Lys
900 905 910
Ala Gly Phe Ile Lys Arg Gln Leu Val Glu Thr Arg Ala Ile Thr Lys
915 920 925
His Val Ala Gln Ile Leu Asp Ser Arg Met Asn Thr Lys Tyr Asp Glu
930 935 940
Asn Asp Lys Leu Ile Arg Glu Val Lys Val Ile Thr Leu Lys Ser Lys
945 950 955 960
Leu Val Ser Asp Phe Arg Lys Asp Phe Gln Phe Tyr Lys Val Arg Glu
965 970 975
Ile Asn Asn Tyr His His Ala His Asp Ala Tyr Leu Asn Ala Val Val
980 985 990
Gly Thr Ala Leu Ile Lys Lys Tyr Pro Lys Leu Glu Ser Glu Phe Val
995 1000 1005
Tyr Gly Asp Tyr Lys Val Tyr Asp Val Arg Lys Met Ile Ala Lys
1010 1015 1020
Ser Glu Gln Glu Ile Gly Lys Ala Thr Ala Lys Tyr Phe Phe Tyr
1025 1030 1035
Ser Asn Ile Met Asn Phe Phe Lys Thr Glu Ile Thr Leu Ala Asn
1040 1045 1050
Gly Glu Ile Arg Lys Arg Pro Leu Ile Glu Thr Asn Gly Glu Thr
1055 1060 1065
Gly Glu Ile Val Trp Asp Lys Gly Arg Asp Phe Ala Thr Val Arg
1070 1075 1080
Lys Val Leu Ser Met Pro Gln Val Asn Ile Val Lys Lys Thr Glu
1085 1090 1095
Val Gln Thr Gly Gly Phe Ser Lys Glu Ser Ile Leu Pro Lys Arg
1100 1105 1110
Asn Ser Asp Lys Leu Ile Ala Arg Lys Lys Asp Trp Asp Pro Lys
1115 1120 1125
Lys Tyr Gly Gly Phe Asp Ser Pro Thr Val Ala Tyr Ser Val Leu
1130 1135 1140
Val Val Ala Lys Val Glu Lys Gly Lys Ser Lys Lys Leu Lys Ser
1145 1150 1155
Val Lys Glu Leu Leu Gly Ile Thr Ile Met Glu Arg Ser Ser Phe
1160 1165 1170
Glu Lys Asn Pro Ile Asp Phe Leu Glu Ala Lys Gly Tyr Lys Glu

1175 1180 1185
Val Lys Lys Asp Leu Ile Ile Lys Leu Pro Lys Tyr Ser Leu Phe
1190 1195 1200
Glu Leu Glu Asn Gly Arg Lys Arg Met Leu Ala Ser Ala Gly Glu
1205 1210 1215
Leu Gln Lys Gly Asn Glu Leu Ala Leu Pro Ser Lys Tyr Val Asn
1220 1225 1230
Phe Leu Tyr Leu Ala Ser His Tyr Glu Lys Leu Lys Gly Ser Pro
1235 1240 1245
Glu Asp Asn Glu Gln Lys Gln Leu Phe Val Glu Gln His Lys His
1250 1255 1260
Tyr Leu Asp Glu Ile Ile Glu Gln Ile Ser Glu Phe Ser Lys Arg
1265 1270 1275
Val Ile Leu Ala Asp Ala Asn Leu Asp Lys Val Leu Ser Ala Tyr
1280 1285 1290
Asn Lys His Arg Asp Lys Pro Ile Arg Glu Gln Ala Glu Asn Ile
1295 1300 1305
Ile His Leu Phe Thr Leu Thr Asn Leu Gly Ala Pro Ala Ala Phe
1310 1315 1320
Lys Tyr Phe Asp Thr Thr Ile Asp Arg Lys Arg Tyr Thr Ser Thr
1325 1330 1335
Lys Glu Val Leu Asp Ala Thr Leu Ile His Gln Ser Ile Thr Gly
1340 1345 1350
Leu Tyr Glu Thr Arg Ile Asp Leu Ser Gln Leu Gly Gly Asp
1355 1360 1365

<210> 21
<211> 1368
<212> PRT
<213> 人工序列
<220>
<223> 合成多肽
<400> 21

Met Asp Lys Lys Tyr Ser Ile Gly Leu Ala Ile Gly Thr Asn Ser Val
1 5 10 15
Gly Trp Ala Val Ile Thr Asp Glu Tyr Lys Val Pro Ser Lys Lys Phe
20 25 30
Lys Val Leu Gly Asn Thr Asp Arg His Ser Ile Lys Lys Asn Leu Ile
35 40 45
Gly Ala Leu Leu Phe Asp Ser Gly Glu Thr Ala Glu Ala Thr Arg Leu

50 55 60

Lys Arg Thr Ala Arg Arg Arg Tyr Thr Arg Arg Lys Asn Arg Ile Cys
65 70 75 80

Tyr Leu Gln Glu Ile Phe Ser Asn Glu Met Ala Lys Val Asp Asp Ser
85 90 95

Phe Phe His Arg Leu Glu Glu Ser Phe Leu Val Glu Glu Asp Lys Lys
100 105 110

His Glu Arg His Pro Ile Phe Gly Asn Ile Val Asp Glu Val Ala Tyr
115 120 125

His Glu Lys Tyr Pro Thr Ile Tyr His Leu Arg Lys Lys Leu Val Asp
130 135 140

Ser Thr Asp Lys Ala Asp Leu Arg Leu Ile Tyr Leu Ala Leu Ala His
145 150 155 160

Met Ile Lys Phe Arg Gly His Phe Leu Ile Glu Gly Asp Leu Asn Pro
165 170 175

Asp Asn Ser Asp Val Asp Lys Leu Phe Ile Gln Leu Val Gln Thr Tyr
180 185 190

Asn Gln Leu Phe Glu Glu Asn Pro Ile Asn Ala Ser Gly Val Asp Ala
195 200 205

Lys Ala Ile Leu Ser Ala Arg Leu Ser Lys Ser Arg Arg Leu Glu Asn
210 215 220

Leu Ile Ala Gln Leu Pro Gly Glu Lys Lys Asn Gly Leu Phe Gly Asn
225 230 235 240

Leu Ile Ala Leu Ser Leu Gly Leu Thr Pro Asn Phe Lys Ser Asn Phe
245 250 255

Asp Leu Ala Glu Asp Ala Lys Leu Gln Leu Ser Lys Asp Thr Tyr Asp
260 265 270

Asp Asp Leu Asp Asn Leu Leu Ala Gln Ile Gly Asp Gln Tyr Ala Asp
275 280 285

Leu Phe Leu Ala Ala Lys Asn Leu Ser Asp Ala Ile Leu Leu Ser Asp
290 295 300

Ile Leu Arg Val Asn Thr Glu Ile Thr Lys Ala Pro Leu Ser Ala Ser
305 310 315 320

Met Ile Lys Arg Tyr Asp Glu His His Gln Asp Leu Thr Leu Leu Lys
325 330 335

Ala Leu Val Arg Gln Gln Leu Pro Glu Lys Tyr Lys Glu Ile Phe Phe
340 345 350

Asp Gln Ser Lys Asn Gly Tyr Ala Gly Tyr Ile Asp Gly Gly Ala Ser
355 360 365

Gln Glu Glu Phe Tyr Lys Phe Ile Lys Pro Ile Leu Glu Lys Met Asp
370 375 380
Gly Thr Glu Glu Leu Leu Val Lys Leu Asn Arg Glu Asp Leu Leu Arg
385 390 395 400
Lys Gln Arg Thr Phe Asp Asn Gly Ser Ile Pro His Gln Ile His Leu
405 410 415
Gly Glu Leu His Ala Ile Leu Arg Arg Gln Glu Asp Phe Tyr Pro Phe
420 425 430
Leu Lys Asp Asn Arg Glu Lys Ile Glu Lys Ile Leu Thr Phe Arg Ile
435 440 445
Pro Tyr Tyr Val Gly Pro Leu Ala Arg Gly Asn Ser Arg Phe Ala Trp
450 455 460
Met Thr Arg Lys Ser Glu Glu Thr Ile Thr Pro Trp Asn Phe Glu Glu
465 470 475 480
Val Val Asp Lys Gly Ala Ser Ala Gln Ser Phe Ile Glu Arg Met Thr
485 490 495
Asn Phe Asp Lys Asn Leu Pro Asn Glu Lys Val Leu Pro Lys His Ser
500 505 510
Leu Leu Tyr Glu Tyr Phe Thr Val Tyr Asn Glu Leu Thr Lys Val Lys
515 520 525
Tyr Val Thr Glu Gly Met Arg Lys Pro Ala Phe Leu Ser Gly Glu Gln
530 535 540
Lys Lys Ala Ile Val Asp Leu Leu Phe Lys Thr Asn Arg Lys Val Thr
545 550 555 560
Val Lys Gln Leu Lys Glu Asp Tyr Phe Lys Lys Ile Glu Cys Phe Asp
565 570 575
Ser Val Glu Ile Ser Gly Val Glu Asp Arg Phe Asn Ala Ser Leu Gly
580 585 590
Thr Tyr His Asp Leu Leu Lys Ile Ile Lys Asp Lys Asp Phe Leu Asp
595 600 605
Asn Glu Glu Asn Glu Asp Ile Leu Glu Asp Ile Val Leu Thr Leu Thr
610 615 620
Leu Phe Glu Asp Arg Glu Met Ile Glu Glu Arg Leu Lys Thr Tyr Ala
625 630 635 640
His Leu Phe Asp Asp Lys Val Met Lys Gln Leu Lys Arg Arg Arg Tyr
645 650 655
Thr Gly Trp Gly Arg Leu Ser Arg Lys Leu Ile Asn Gly Ile Arg Asp
660 665 670
Lys Gln Ser Gly Lys Thr Ile Leu Asp Phe Leu Lys Ser Asp Gly Phe

675 680 685

Ala Asn Arg Asn Phe Met Gln Leu Ile His Asp Asp Ser Leu Thr Phe

690 695 700

Lys Glu Asp Ile Gln Lys Ala Gln Val Ser Gly Gln Gly Asp Ser Leu

705 710 715 720

His Glu His Ile Ala Asn Leu Ala Gly Ser Pro Ala Ile Lys Lys Gly

725 730 735

Ile Leu Gln Thr Val Lys Val Val Asp Glu Leu Val Lys Val Met Gly

740 745 750

Arg His Lys Pro Glu Asn Ile Val Ile Glu Met Ala Arg Glu Asn Gln

755 760 765

Thr Thr Gln Lys Gly Gln Lys Asn Ser Arg Glu Arg Met Lys Arg Ile

770 775 780

Glu Glu Gly Ile Lys Glu Leu Gly Ser Gln Ile Leu Lys Glu His Pro

785 790 795 800

Val Glu Asn Thr Gln Leu Gln Asn Glu Lys Leu Tyr Leu Tyr Tyr Leu

805 810 815

Gln Asn Gly Arg Asp Met Tyr Val Asp Gln Glu Leu Asp Ile Asn Arg

820 825 830

Leu Ser Asp Tyr Asp Val Asp His Ile Val Pro Gln Ser Phe Leu Lys

835 840 845

Asp Asp Ser Ile Asp Asn Lys Val Leu Thr Arg Ser Asp Lys Asn Arg

850 855 860

Gly Lys Ser Asp Asn Val Pro Ser Glu Glu Val Val Lys Lys Met Lys

865 870 875 880

Asn Tyr Trp Arg Gln Leu Leu Asn Ala Lys Leu Ile Thr Gln Arg Lys

885 890 895

Phe Asp Asn Leu Thr Lys Ala Glu Arg Gly Gly Leu Ser Glu Leu Asp

900 905 910

Lys Ala Gly Phe Ile Lys Arg Gln Leu Val Glu Thr Arg Gln Ile Thr

915 920 925

Lys His Val Ala Gln Ile Leu Asp Ser Arg Met Asn Thr Lys Tyr Asp

930 935 940

Glu Asn Asp Lys Leu Ile Arg Glu Val Lys Val Ile Thr Leu Lys Ser

945 950 955 960

Lys Leu Val Ser Asp Phe Arg Lys Asp Phe Gln Phe Tyr Lys Val Arg

965 970 975

Glu Ile Asn Asn Tyr His His Ala His Asp Ala Tyr Leu Asn Ala Val

980 985 990

Val Gly Thr Ala Leu Ile Lys Lys Tyr Pro Lys Leu Glu Ser Glu Phe
995 1000 1005
Val Tyr Gly Asp Tyr Lys Val Tyr Asp Val Arg Lys Met Ile Ala
1010 1015 1020
Lys Ser Glu Gln Glu Ile Gly Lys Ala Thr Ala Lys Tyr Phe Phe
1025 1030 1035
Tyr Ser Asn Ile Met Asn Phe Phe Lys Thr Glu Ile Thr Leu Ala
1040 1045 1050
Asn Gly Glu Ile Arg Lys Arg Pro Leu Ile Glu Thr Asn Gly Glu
1055 1060 1065
Thr Gly Glu Ile Val Trp Asp Lys Gly Arg Asp Phe Ala Thr Val
1070 1075 1080
Arg Lys Val Leu Ser Met Pro Gln Val Asn Ile Val Lys Lys Thr
1085 1090 1095
Glu Val Gln Thr Gly Gly Phe Ser Lys Glu Ser Ile Leu Pro Lys
1100 1105 1110
Arg Asn Ser Asp Lys Leu Ile Ala Arg Lys Lys Asp Trp Asp Pro
1115 1120 1125
Lys Lys Tyr Gly Gly Phe Asp Ser Pro Thr Val Ala Tyr Ser Val
1130 1135 1140
Leu Val Val Ala Lys Val Glu Lys Gly Lys Ser Lys Lys Leu Lys
1145 1150 1155
Ser Val Lys Glu Leu Leu Gly Ile Thr Ile Met Glu Arg Ser Ser
1160 1165 1170
Phe Glu Lys Asn Pro Ile Asp Phe Leu Glu Ala Lys Gly Tyr Lys
1175 1180 1185
Glu Val Lys Lys Asp Leu Ile Ile Lys Leu Pro Lys Tyr Ser Leu
1190 1195 1200
Phe Glu Leu Glu Asn Gly Arg Lys Arg Met Leu Ala Ser Ala Gly
1205 1210 1215
Glu Leu Gln Lys Gly Asn Glu Leu Ala Leu Pro Ser Lys Tyr Val
1220 1225 1230
Asn Phe Leu Tyr Leu Ala Ser His Tyr Glu Lys Leu Lys Gly Ser
1235 1240 1245
Pro Glu Asp Asn Glu Gln Lys Gln Leu Phe Val Glu Gln His Lys
1250 1255 1260
His Tyr Leu Asp Glu Ile Ile Glu Gln Ile Ser Glu Phe Ser Lys
1265 1270 1275
Arg Val Ile Leu Ala Asp Ala Asn Leu Asp Lys Val Leu Ser Ala

1280 1285 1290
Tyr Asn Lys His Arg Asp Lys Pro Ile Arg Glu Gln Ala Glu Asn
1295 1300 1305
Ile Ile His Leu Phe Thr Leu Thr Asn Leu Gly Ala Pro Ala Ala
1310 1315 1320
Phe Lys Tyr Phe Asp Thr Thr Ile Asp Arg Lys Arg Tyr Thr Ser
1325 1330 1335
Thr Lys Glu Val Leu Asp Ala Thr Leu Ile His Gln Ser Ile Thr
1340 1345 1350
Gly Leu Tyr Glu Thr Arg Ile Asp Leu Ser Gln Leu Gly Gly Asp
1355 1360 1365

<210> 22
<211> 1368
<212> PRT
<213> 人工序列
<220>
<223> 合成多肽
<400> 22

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

165 170 175

Asp Asn Ser Asp Val Asp Lys Leu Phe Ile Gln Leu Val Gln Thr Tyr

180 185 190

Asn Gln Leu Phe Glu Glu Asn Pro Ile Asn Ala Ser Gly Val Asp Ala

195 200 205

Lys Ala Ile Leu Ser Ala Arg Leu Ser Lys Ser Arg Arg Leu Glu Asn

210 215 220

Leu Ile Ala Gln Leu Pro Gly Glu Lys Lys Asn Gly Leu Phe Gly Asn

225 230 235 240

Leu Ile Ala Leu Ser Leu Gly Leu Thr Pro Asn Phe Lys Ser Asn Phe

245 250 255

Asp Leu Ala Glu Asp Ala Lys Leu Gln Leu Ser Lys Asp Thr Tyr Asp

260 265 270

Asp Asp Leu Asp Asn Leu Leu Ala Gln Ile Gly Asp Gln Tyr Ala Asp

275 280 285

Leu Phe Leu Ala Ala Lys Asn Leu Ser Asp Ala Ile Leu Leu Ser Asp

290 295 300

Ile Leu Arg Val Asn Thr Glu Ile Thr Lys Ala Pro Leu Ser Ala Ser

305 310 315 320

Met Ile Lys Arg Tyr Asp Glu His His Gln Asp Leu Thr Leu Leu Lys

325 330 335

Ala Leu Val Arg Gln Gln Leu Pro Glu Lys Tyr Lys Glu Ile Phe Phe

340 345 350

Asp Gln Ser Lys Asn Gly Tyr Ala Gly Tyr Ile Asp Gly Gly Ala Ser

355 360 365

Gln Glu Glu Phe Tyr Lys Phe Ile Lys Pro Ile Leu Glu Lys Met Asp

370 375 380

Gly Thr Glu Glu Leu Leu Val Lys Leu Asn Arg Glu Asp Leu Leu Arg

385 390 395 400

Lys Gln Arg Thr Phe Asp Asn Gly Ser Ile Pro His Gln Ile His Leu

405 410 415

Gly Glu Leu His Ala Ile Leu Arg Arg Gln Glu Asp Phe Tyr Pro Phe

420 425 430

Leu Lys Asp Asn Arg Glu Lys Ile Glu Lys Ile Leu Thr Phe Arg Ile

435 440 445

Pro Tyr Tyr Val Gly Pro Leu Ala Arg Gly Asn Ser Arg Phe Ala Trp

450 455 460

Met Thr Arg Lys Ser Glu Glu Thr Ile Thr Pro Trp Asn Phe Glu Glu

465 470 475 480

Val Val Asp Lys Gly Ala Ser Ala Gln Ser Phe Ile Glu Arg Met Thr
485 490 495
Asn Phe Asp Lys Asn Leu Pro Asn Glu Lys Val Leu Pro Lys His Ser
500 505 510
Leu Leu Tyr Glu Tyr Phe Thr Val Tyr Asn Glu Leu Thr Lys Val Lys
515 520 525
Tyr Val Thr Glu Gly Met Arg Lys Pro Ala Phe Leu Ser Gly Glu Gln
530 535 540
Lys Lys Ala Ile Val Asp Leu Leu Phe Lys Thr Asn Arg Lys Val Thr
545 550 555 560
Val Lys Gln Leu Lys Glu Asp Tyr Phe Lys Lys Ile Glu Cys Phe Asp
565 570 575
Ser Val Glu Ile Ser Gly Val Glu Asp Arg Phe Asn Ala Ser Leu Gly
580 585 590
Thr Tyr His Asp Leu Leu Lys Ile Ile Lys Asp Lys Asp Phe Leu Asp
595 600 605
Asn Glu Glu Asn Glu Asp Ile Leu Glu Asp Ile Val Leu Thr Leu Thr
610 615 620
Leu Phe Glu Asp Arg Glu Met Ile Glu Glu Arg Leu Lys Thr Tyr Ala
625 630 635 640
His Leu Phe Asp Asp Lys Val Met Lys Gln Leu Lys Arg Arg Arg Tyr
645 650 655
Thr Gly Trp Gly Arg Leu Ser Arg Lys Leu Ile Asn Gly Ile Arg Asp
660 665 670
Lys Gln Ser Gly Lys Thr Ile Leu Asp Phe Leu Lys Ser Asp Gly Phe
675 680 685
Ala Asn Arg Asn Phe Met Gln Leu Ile His Asp Asp Ser Leu Thr Phe
690 695 700
Lys Glu Asp Ile Gln Lys Ala Gln Val Ser Gly Gln Gly Asp Ser Leu
705 710 715 720
His Glu His Ile Ala Asn Leu Ala Gly Ser Pro Ala Ile Lys Lys Gly
725 730 735
Ile Leu Gln Thr Val Lys Val Val Asp Glu Leu Val Lys Val Met Gly
740 745 750
Arg His Lys Pro Glu Asn Ile Val Ile Glu Met Ala Arg Glu Asn Gln
755 760 765
Thr Thr Gln Lys Gly Gln Lys Asn Ser Arg Glu Arg Met Lys Arg Ile
770 775 780
Glu Glu Gly Ile Lys Glu Leu Gly Ser Gln Ile Leu Lys Glu His Pro

785 790 795 800
Val Glu Asn Thr Gln Leu Gln Asn Glu Lys Leu Tyr Leu Tyr Tyr Leu
805 810 815
Gln Asn Gly Arg Asp Met Tyr Val Asp Gln Glu Leu Asp Ile Asn Arg
820 825 830
Leu Ser Asp Tyr Asp Val Asp Ala Ile Val Pro Gln Ser Phe Leu Lys
835 840 845
Asp Asp Ser Ile Asp Asn Lys Val Leu Thr Arg Ser Asp Lys Asn Arg
850 855 860
Gly Lys Ser Asp Asn Val Pro Ser Glu Glu Val Val Lys Lys Met Lys
865 870 875 880
Asn Tyr Trp Arg Gln Leu Leu Asn Ala Lys Leu Ile Thr Gln Arg Lys
885 890 895
Phe Asp Asn Leu Thr Lys Ala Glu Arg Gly Gly Leu Ser Glu Leu Asp
900 905 910
Lys Ala Gly Phe Ile Lys Arg Gln Leu Val Glu Thr Arg Gln Ile Thr
915 920 925
Lys His Val Ala Gln Ile Leu Asp Ser Arg Met Asn Thr Lys Tyr Asp
930 935 940
Glu Asn Asp Lys Leu Ile Arg Glu Val Lys Val Ile Thr Leu Lys Ser
945 950 955 960
Lys Leu Val Ser Asp Phe Arg Lys Asp Phe Gln Phe Tyr Lys Val Arg
965 970 975
Glu Ile Asn Asn Tyr His His Ala His Asp Ala Tyr Leu Asn Ala Val
980 985 990
Val Gly Thr Ala Leu Ile Lys Lys Tyr Pro Lys Leu Glu Ser Glu Phe
995 1000 1005
Val Tyr Gly Asp Tyr Lys Val Tyr Asp Val Arg Lys Met Ile Ala
1010 1015 1020
Lys Ser Glu Gln Glu Ile Gly Lys Ala Thr Ala Lys Tyr Phe Phe
1025 1030 1035
Tyr Ser Asn Ile Met Asn Phe Phe Lys Thr Glu Ile Thr Leu Ala
1040 1045 1050
Asn Gly Glu Ile Arg Lys Arg Pro Leu Ile Glu Thr Asn Gly Glu
1055 1060 1065
Thr Gly Glu Ile Val Trp Asp Lys Gly Arg Asp Phe Ala Thr Val
1070 1075 1080
Arg Lys Val Leu Ser Met Pro Gln Val Asn Ile Val Lys Lys Thr
1085 1090 1095

Glu Val Gln Thr Gly Gly Phe Ser Lys Glu Ser Ile Leu Pro Lys
1100 1105 1110
Arg Asn Ser Asp Lys Leu Ile Ala Arg Lys Lys Asp Trp Asp Pro
1115 1120 1125
Lys Lys Tyr Gly Gly Phe Asp Ser Pro Thr Val Ala Tyr Ser Val
1130 1135 1140
Leu Val Val Ala Lys Val Glu Lys Gly Lys Ser Lys Lys Leu Lys
1145 1150 1155
Ser Val Lys Glu Leu Leu Gly Ile Thr Ile Met Glu Arg Ser Ser
1160 1165 1170
Phe Glu Lys Asn Pro Ile Asp Phe Leu Glu Ala Lys Gly Tyr Lys
1175 1180 1185
Glu Val Lys Lys Asp Leu Ile Ile Lys Leu Pro Lys Tyr Ser Leu
1190 1195 1200
Phe Glu Leu Glu Asn Gly Arg Lys Arg Met Leu Ala Ser Ala Gly
1205 1210 1215
Glu Leu Gln Lys Gly Asn Glu Leu Ala Leu Pro Ser Lys Tyr Val
1220 1225 1230
Asn Phe Leu Tyr Leu Ala Ser His Tyr Glu Lys Leu Lys Gly Ser
1235 1240 1245
Pro Glu Asp Asn Glu Gln Lys Gln Leu Phe Val Glu Gln His Lys
1250 1255 1260
His Tyr Leu Asp Glu Ile Ile Glu Gln Ile Ser Glu Phe Ser Lys
1265 1270 1275
Arg Val Ile Leu Ala Asp Ala Asn Leu Asp Lys Val Leu Ser Ala
1280 1285 1290
Tyr Asn Lys His Arg Asp Lys Pro Ile Arg Glu Gln Ala Glu Asn
1295 1300 1305
Ile Ile His Leu Phe Thr Leu Thr Asn Leu Gly Ala Pro Ala Ala
1310 1315 1320
Phe Lys Tyr Phe Asp Thr Thr Ile Asp Arg Lys Arg Tyr Thr Ser
1325 1330 1335
Thr Lys Glu Val Leu Asp Ala Thr Leu Ile His Gln Ser Ile Thr
1340 1345 1350
Gly Leu Tyr Glu Thr Arg Ile Asp Leu Ser Gln Leu Gly Gly Asp
1355 1360 1365

<210> 23

<211> 1368

<212> PRT

<213> 酿脓链球菌

<400> 23

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

290 295 300

Ile Leu Arg Val Asn Thr Glu Ile Thr Lys Ala Pro Leu Ser Ala Ser
305 310 315 320

Met Ile Lys Arg Tyr Asp Glu His His Gln Asp Leu Thr Leu Leu Lys
325 330 335

Ala Leu Val Arg Gln Gln Leu Pro Glu Lys Tyr Lys Glu Ile Phe Phe
340 345 350

Asp Gln Ser Lys Asn Gly Tyr Ala Gly Tyr Ile Asp Gly Gly Ala Ser
355 360 365

Gln Glu Glu Phe Tyr Lys Phe Ile Lys Pro Ile Leu Glu Lys Met Asp
370 375 380

Gly Thr Glu Glu Leu Leu Val Lys Leu Asn Arg Glu Asp Leu Leu Arg
385 390 395 400

Lys Gln Arg Thr Phe Asp Asn Gly Ser Ile Pro His Gln Ile His Leu
405 410 415

Gly Glu Leu His Ala Ile Leu Arg Arg Gln Glu Asp Phe Tyr Pro Phe
420 425 430

Leu Lys Asp Asn Arg Glu Lys Ile Glu Lys Ile Leu Thr Phe Arg Ile
435 440 445

Pro Tyr Tyr Val Gly Pro Leu Ala Arg Gly Asn Ser Arg Phe Ala Trp
450 455 460

Met Thr Arg Lys Ser Glu Glu Thr Ile Thr Pro Trp Asn Phe Glu Glu
465 470 475 480

Val Val Asp Lys Gly Ala Ser Ala Gln Ser Phe Ile Glu Arg Met Thr
485 490 495

Asn Phe Asp Lys Asn Leu Pro Asn Glu Lys Val Leu Pro Lys His Ser
500 505 510

Leu Leu Tyr Glu Tyr Phe Thr Val Tyr Asn Glu Leu Thr Lys Val Lys
515 520 525

Tyr Val Thr Glu Gly Met Arg Lys Pro Ala Phe Leu Ser Gly Glu Gln
530 535 540

Lys Lys Ala Ile Val Asp Leu Leu Phe Lys Thr Asn Arg Lys Val Thr
545 550 555 560

Val Lys Gln Leu Lys Glu Asp Tyr Phe Lys Lys Ile Glu Cys Phe Asp
565 570 575

Ser Val Glu Ile Ser Gly Val Glu Asp Arg Phe Asn Ala Ser Leu Gly
580 585 590

Thr Tyr His Asp Leu Leu Lys Ile Ile Lys Asp Lys Asp Phe Leu Asp
595 600 605

Asn Glu Glu Asn Glu Asp Ile Leu Glu Asp Ile Val Leu Thr Leu Thr
610 615 620
Leu Phe Glu Asp Arg Glu Met Ile Glu Glu Arg Leu Lys Thr Tyr Ala
625 630 635 640
His Leu Phe Asp Asp Lys Val Met Lys Gln Leu Lys Arg Arg Arg Tyr
645 650 655
Thr Gly Trp Gly Arg Leu Ser Arg Lys Leu Ile Asn Gly Ile Arg Asp
660 665 670
Lys Gln Ser Gly Lys Thr Ile Leu Asp Phe Leu Lys Ser Asp Gly Phe
675 680 685
Ala Asn Arg Asn Phe Met Gln Leu Ile His Asp Asp Ser Leu Thr Phe
690 695 700
Lys Glu Asp Ile Gln Lys Ala Gln Val Ser Gly Gln Gly Asp Ser Leu
705 710 715 720
His Glu His Ile Ala Asn Leu Ala Gly Ser Pro Ala Ile Lys Lys Gly
725 730 735
Ile Leu Gln Thr Val Lys Val Val Asp Glu Leu Val Lys Val Met Gly
740 745 750
Arg His Lys Pro Glu Asn Ile Val Ile Glu Met Ala Arg Glu Asn Gln
755 760 765
Thr Thr Gln Lys Gly Gln Lys Asn Ser Arg Glu Arg Met Lys Arg Ile
770 775 780
Glu Glu Gly Ile Lys Glu Leu Gly Ser Gln Ile Leu Lys Glu His Pro
785 790 795 800
Val Glu Asn Thr Gln Leu Gln Asn Glu Lys Leu Tyr Leu Tyr Tyr Leu
805 810 815
Gln Asn Gly Arg Asp Met Tyr Val Asp Gln Glu Leu Asp Ile Asn Arg
820 825 830
Leu Ser Asp Tyr Asp Val Asp His Ile Val Pro Gln Ser Phe Leu Lys
835 840 845
Asp Asp Ser Ile Asp Asn Lys Val Leu Thr Arg Ser Asp Lys Asn Arg
850 855 860
Gly Lys Ser Asp Asn Val Pro Ser Glu Glu Val Val Lys Lys Met Lys
865 870 875 880
Asn Tyr Trp Arg Gln Leu Leu Asn Ala Lys Leu Ile Thr Gln Arg Lys
885 890 895
Phe Asp Asn Leu Thr Lys Ala Glu Arg Gly Gly Leu Ser Glu Leu Asp
900 905 910
Lys Ala Gly Phe Ile Lys Arg Gln Leu Val Glu Thr Arg Gln Ile Thr

915 920 925

Lys His Val Ala Gln Ile Leu Asp Ser Arg Met Asn Thr Lys Tyr Asp

930 935 940

Glu Asn Asp Lys Leu Ile Arg Glu Val Lys Val Ile Thr Leu Lys Ser

945 950 955 960

Lys Leu Val Ser Asp Phe Arg Lys Asp Phe Gln Phe Tyr Lys Val Arg

965 970 975

Glu Ile Asn Asn Tyr His His Ala His Asp Ala Tyr Leu Asn Ala Val

980 985 990

Val Gly Thr Ala Leu Ile Lys Lys Tyr Pro Lys Leu Glu Ser Glu Phe

995 1000 1005

Val Tyr Gly Asp Tyr Lys Val Tyr Asp Val Arg Lys Met Ile Ala

1010 1015 1020

Lys Ser Glu Gln Glu Ile Gly Lys Ala Thr Ala Lys Tyr Phe Phe

1025 1030 1035

Tyr Ser Asn Ile Met Asn Phe Phe Lys Thr Glu Ile Thr Leu Ala

1040 1045 1050

Asn Gly Glu Ile Arg Lys Arg Pro Leu Ile Glu Thr Asn Gly Glu

1055 1060 1065

Thr Gly Glu Ile Val Trp Asp Lys Gly Arg Asp Phe Ala Thr Val

1070 1075 1080

Arg Lys Val Leu Ser Met Pro Gln Val Asn Ile Val Lys Lys Thr

1085 1090 1095

Glu Val Gln Thr Gly Gly Phe Ser Lys Glu Ser Ile Leu Pro Lys

1100 1105 1110

Arg Asn Ser Asp Lys Leu Ile Ala Arg Lys Lys Asp Trp Asp Pro

1115 1120 1125

Lys Lys Tyr Gly Gly Phe Asp Ser Pro Thr Val Ala Tyr Ser Val

1130 1135 1140

Leu Val Val Ala Lys Val Glu Lys Gly Lys Ser Lys Lys Leu Lys

1145 1150 1155

Ser Val Lys Glu Leu Leu Gly Ile Thr Ile Met Glu Arg Ser Ser

1160 1165 1170

Phe Glu Lys Asn Pro Ile Asp Phe Leu Glu Ala Lys Gly Tyr Lys

1175 1180 1185

Glu Val Lys Lys Asp Leu Ile Ile Lys Leu Pro Lys Tyr Ser Leu

1190 1195 1200

Phe Glu Leu Glu Asn Gly Arg Lys Arg Met Leu Ala Ser Ala Gly

1205 1210 1215

Glu Leu Gln Lys Gly Asn Glu Leu Ala Leu Pro Ser Lys Tyr Val
1220 1225 1230
Asn Phe Leu Tyr Leu Ala Ser His Tyr Glu Lys Leu Lys Gly Ser
1235 1240 1245
Pro Glu Asp Asn Glu Gln Lys Gln Leu Phe Val Glu Gln His Lys
1250 1255 1260
His Tyr Leu Asp Glu Ile Ile Glu Gln Ile Ser Glu Phe Ser Lys
1265 1270 1275
Arg Val Ile Leu Ala Asp Ala Asn Leu Asp Lys Val Leu Ser Ala
1280 1285 1290
Tyr Asn Lys His Arg Asp Lys Pro Ile Arg Glu Gln Ala Glu Asn
1295 1300 1305
Ile Ile His Leu Phe Thr Leu Thr Asn Leu Gly Ala Pro Ala Ala
1310 1315 1320
Phe Lys Tyr Phe Asp Thr Thr Ile Asp Arg Lys Arg Tyr Thr Ser
1325 1330 1335
Thr Lys Glu Val Leu Asp Ala Thr Leu Ile His Gln Ser Ile Thr
1340 1345 1350
Gly Leu Tyr Glu Thr Arg Ile Asp Leu Ser Gln Leu Gly Gly Asp
1355 1360 1365

<210> 24

<211> 1372

<212> PRT

<213> 解没食子酸链球菌

<400> 24

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

Thr Phe Asp Ser His Pro Ile Phe Gly Asn Lys Ala Glu Glu Asp Ala
115 120 125
Tyr His Gln Lys Phe Pro Thr Ile Tyr His Leu Arg Lys His Leu Ala
130 135 140
Asp Ser Ser Glu Lys Ala Asp Leu Arg Leu Val Tyr Leu Ala Leu Ala
145 150 155 160
His Met Ile Lys Phe Arg Gly His Phe Leu Ile Glu Gly Glu Leu Asn
165 170 175
Ala Glu Asn Thr Asp Val Gln Lys Ile Phe Ala Asp Phe Val Gly Val
180 185 190
Tyr Asn Arg Thr Phe Asp Asp Ser His Leu Ser Glu Ile Thr Val Asp
195 200 205
Val Ala Ser Ile Leu Thr Glu Lys Ile Ser Lys Ser Arg Arg Leu Glu
210 215 220
Asn Leu Ile Lys Tyr Tyr Pro Thr Glu Lys Lys Asn Thr Leu Phe Gly
225 230 235 240
Asn Leu Ile Ala Leu Ala Leu Gly Leu Gln Pro Asn Phe Lys Thr Asn
245 250 255
Phe Lys Leu Ser Glu Asp Ala Lys Leu Gln Phe Ser Lys Asp Thr Tyr
260 265 270
Glu Glu Asp Leu Glu Glu Leu Leu Gly Lys Ile Gly Asp Asp Tyr Ala
275 280 285
Asp Leu Phe Thr Ser Ala Lys Asn Leu Tyr Asp Ala Ile Leu Leu Ser
290 295 300
Gly Ile Leu Thr Val Asp Asp Asn Ser Thr Lys Ala Pro Leu Ser Ala
305 310 315 320
Ser Met Ile Lys Arg Tyr Val Glu His His Glu Asp Leu Glu Lys Leu
325 330 335
Lys Glu Phe Ile Lys Ala Asn Lys Ser Glu Leu Tyr His Asp Ile Phe
340 345 350
Lys Asp Lys Asn Lys Asn Gly Tyr Ala Gly Tyr Ile Glu Asn Gly Val
355 360 365
Lys Gln Asp Glu Phe Tyr Lys Tyr Leu Lys Asn Ile Leu Ser Lys Ile
370 375 380
Lys Ile Asp Gly Ser Asp Tyr Phe Leu Asp Lys Ile Glu Arg Glu Asp
385 390 395 400
Phe Leu Arg Lys Gln Arg Thr Phe Asp Asn Gly Ser Ile Pro His Gln
405 410 415
Ile His Leu Gln Glu Met His Ala Ile Leu Arg Arg Gln Gly Asp Tyr

420 425 430

Tyr Pro Phe Leu Lys Glu Lys Gln Asp Arg Ile Glu Lys Ile Leu Thr
435 440 445

Phe Arg Ile Pro Tyr Tyr Val Gly Pro Leu Val Arg Lys Asp Ser Arg
450 455 460

Phe Ala Trp Ala Glu Tyr Arg Ser Asp Glu Lys Ile Thr Pro Trp Asn
465 470 475 480

Phe Asp Lys Val Ile Asp Lys Glu Lys Ser Ala Glu Lys Phe Ile Thr
485 490 495

Arg Met Thr Leu Asn Asp Leu Tyr Leu Pro Glu Glu Lys Val Leu Pro
500 505 510

Lys His Ser His Val Tyr Glu Thr Tyr Ala Val Tyr Asn Glu Leu Thr
515 520 525

Lys Ile Lys Tyr Val Asn Glu Gln Gly Lys Glu Ser Phe Phe Asp Ser
530 535 540

Asn Met Lys Gln Glu Ile Phe Asp His Val Phe Lys Glu Asn Arg Lys
545 550 555 560

Val Thr Lys Glu Lys Leu Leu Asn Tyr Leu Asn Lys Glu Phe Pro Glu
565 570 575

Tyr Arg Ile Lys Asp Leu Ile Gly Leu Asp Lys Glu Asn Lys Ser Phe
580 585 590

Asn Ala Ser Leu Gly Thr Tyr His Asp Leu Lys Lys Ile Leu Asp Lys
595 600 605

Ala Phe Leu Asp Asp Lys Val Asn Glu Glu Val Ile Glu Asp Ile Ile
610 615 620

Lys Thr Leu Thr Leu Phe Glu Asp Lys Asp Met Ile His Glu Arg Leu
625 630 635 640

Gln Lys Tyr Ser Asp Ile Phe Thr Ala Asn Gln Leu Lys Lys Leu Glu
645 650 655

Arg Arg His Tyr Thr Gly Trp Gly Arg Leu Ser Tyr Lys Leu Ile Asn
660 665 670

Gly Ile Arg Asn Lys Glu Asn Asn Lys Thr Ile Leu Asp Tyr Leu Ile
675 680 685

Asp Asp Gly Ser Ala Asn Arg Asn Phe Met Gln Leu Ile Asn Asp Asp
690 695 700

Thr Leu Pro Phe Lys Gln Ile Ile Gln Lys Ser Gln Val Val Gly Asp
705 710 715 720

Val Asp Asp Ile Glu Ala Val Val His Asp Leu Pro Gly Ser Pro Ala
725 730 735

Ile Lys Lys Gly Ile Leu Gln Ser Val Lys Ile Val Asp Glu Leu Val
740 745 750
Lys Val Met Gly Gly Asn Pro Asp Asn Ile Val Ile Glu Met Ala Arg
755 760 765
Glu Asn Gln Thr Thr Asn Arg Gly Arg Ser Gln Ser Gln Gln Arg Leu
770 775 780
Lys Lys Leu Gln Asn Ser Leu Lys Glu Leu Gly Ser Asn Ile Leu Asn
785 790 795 800
Glu Glu Lys Pro Ser Tyr Ile Glu Asp Lys Val Glu Asn Ser His Leu
805 810 815
Gln Asn Asp Gln Leu Phe Leu Tyr Tyr Ile Gln Asn Gly Lys Asp Met
820 825 830
Tyr Thr Gly Asp Glu Leu Asp Ile Asp His Leu Ser Asp Tyr Asp Ile
835 840 845
Asp His Ile Ile Pro Gln Ala Phe Ile Lys Asp Asp Ser Ile Asp Asn
850 855 860
Arg Val Leu Thr Ser Ser Ala Lys Asn Arg Gly Lys Ser Asp Asp Val
865 870 875 880
Pro Ser Leu Asp Ile Val Arg Ala Arg Lys Ala Glu Trp Val Arg Leu
885 890 895
Tyr Lys Ser Gly Leu Ile Ser Lys Arg Lys Phe Asp Asn Leu Thr Lys
900 905 910
Ala Glu Arg Gly Gly Leu Thr Glu Ala Asp Lys Ala Gly Phe Ile Lys
915 920 925
Arg Gln Leu Val Glu Thr Arg Gln Ile Thr Lys His Val Ala Gln Ile
930 935 940
Leu Asp Ala Arg Phe Asn Thr Glu His Asp Glu Asn Asp Lys Val Ile
945 950 955 960
Arg Asp Val Lys Val Ile Thr Leu Lys Ser Asn Leu Val Ser Gln Phe
965 970 975
Arg Lys Asp Phe Glu Phe Tyr Lys Val Arg Glu Ile Asn Asp Tyr His
980 985 990
His Ala His Asp Ala Tyr Leu Asn Ala Val Val Gly Thr Ala Leu Leu
995 1000 1005
Lys Lys Tyr Pro Lys Leu Ala Ser Glu Phe Val Tyr Gly Glu Tyr
1010 1015 1020
Lys Lys Tyr Asp Ile Arg Lys Phe Ile Thr Asn Ser Ser Asp Lys
1025 1030 1035
Ala Thr Ala Lys Tyr Phe Phe Tyr Ser Asn Leu Met Asn Phe Phe

1040 1045 1050

Lys Thr Lys Val Lys Tyr Ala Asp Gly Thr Val Phe Glu Arg Pro
1055 1060 1065

Ile Ile Glu Thr Asn Ala Asp Gly Glu Ile Ala Trp Asn Lys Gln
1070 1075 1080

Ile Asp Phe Glu Lys Val Arg Lys Val Leu Ser Tyr Pro Gln Val
1085 1090 1095

Asn Ile Val Lys Lys Val Glu Thr Gln Thr Gly Gly Phe Ser Lys
1100 1105 1110

Glu Ser Ile Leu Pro Lys Gly Asp Ser Asp Lys Leu Ile Pro Arg
1115 1120 1125

Lys Thr Lys Lys Val Tyr Trp Asp Thr Lys Lys Tyr Gly Gly Phe
1130 1135 1140

Asp Ser Pro Thr Val Ala Tyr Ser Val Phe Val Val Ala Asp Val
1145 1150 1155

Glu Lys Gly Lys Ala Lys Lys Leu Lys Thr Val Lys Glu Leu Val
1160 1165 1170

Gly Ile Ser Ile Met Glu Arg Ser Phe Phe Glu Glu Asn Pro Val
1175 1180 1185

Glu Phe Leu Glu Asn Lys Gly Tyr His Asn Ile Arg Glu Asp Lys
1190 1195 1200

Leu Ile Lys Leu Pro Lys Tyr Ser Leu Phe Glu Phe Glu Gly Gly
1205 1210 1215

Arg Arg Arg Leu Leu Ala Ser Ala Ser Glu Leu Gln Lys Gly Asn
1220 1225 1230

Glu Met Val Leu Pro Gly Tyr Leu Val Glu Leu Leu Tyr His Ala
1235 1240 1245

His Arg Ala Asp Asn Phe Asn Ser Thr Glu Tyr Leu Asn Tyr Val
1250 1255 1260

Ser Glu His Lys Lys Glu Phe Glu Lys Val Leu Ser Cys Val Glu
1265 1270 1275

Asp Phe Ala Asn Leu Tyr Val Asp Val Glu Lys Asn Leu Ser Lys
1280 1285 1290

Ile Arg Ala Val Ala Asp Ser Met Asp Asn Phe Ser Ile Glu Glu
1295 1300 1305

Ile Ser Asn Ser Phe Ile Asn Leu Leu Thr Leu Thr Ala Leu Gly
1310 1315 1320

Ala Pro Ala Asp Phe Asn Phe Leu Gly Glu Lys Ile Pro Arg Lys
1325 1330 1335

Arg Tyr Thr Ser Thr Lys Glu Cys Leu Asn Ala Thr Leu Ile His
1340 1345 1350
Gln Ser Ile Thr Gly Leu Tyr Glu Thr Arg Ile Asp Leu Ser Lys
1355 1360 1365
Leu Gly Glu Glu
1370

<210> 25
<211> 1370
<212> PRT
<213> 格氏链球菌(Streptococcus gordonii)
<400> 25

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

Val Leu Lys Leu Phe Pro Asp Glu Lys Ser Thr Gly Leu Phe Ser Glu
225 230 235 240
Phe Leu Lys Leu Ile Val Gly Asn Gln Ala Asp Phe Lys Lys His Phe
245 250 255
Asp Leu Glu Asp Lys Ala Pro Leu Gln Phe Ser Lys Asp Thr Tyr Asp
260 265 270
Glu Asp Leu Glu Asn Leu Leu Gly Gln Ile Gly Asp Asp Phe Thr Asp
275 280 285
Leu Phe Val Ser Ala Lys Lys Leu Tyr Asp Ala Ile Leu Leu Ser Gly
290 295 300
Ile Leu Thr Val Thr Asp Pro Ser Thr Lys Ala Pro Leu Ser Ala Ser
305 310 315 320
Met Ile Glu Arg Tyr Glu Asn His Gln Asn Asp Leu Ala Ala Leu Lys
325 330 335
Gln Phe Ile Lys Asn Asn Leu Pro Glu Lys Tyr Asp Glu Val Phe Ser
340 345 350
Asp Gln Ser Lys Asp Gly Tyr Ala Gly Tyr Ile Asp Gly Lys Thr Thr
355 360 365
Gln Glu Thr Phe Tyr Lys Tyr Ile Lys Asn Leu Leu Ser Lys Phe Glu
370 375 380
Gly Thr Asp Tyr Phe Leu Asp Lys Ile Glu Arg Glu Asp Phe Leu Arg
385 390 395 400
Lys Gln Arg Thr Phe Asp Asn Gly Ser Ile Pro His Gln Ile His Leu
405 410 415
Gln Glu Met Asn Ala Ile Leu Arg Arg Gln Gly Glu Tyr Tyr Pro Phe
420 425 430
Leu Lys Asp Asn Lys Glu Lys Ile Glu Lys Ile Leu Thr Phe Arg Ile
435 440 445
Pro Tyr Tyr Val Gly Pro Leu Ala Arg Gly Asn Arg Asp Phe Ala Trp
450 455 460
Leu Thr Arg Asn Ser Asp Glu Ala Ile Arg Pro Trp Asn Phe Glu Glu
465 470 475 480
Ile Val Asp Lys Ala Ser Ser Ala Glu Asp Phe Ile Asn Lys Met Thr
485 490 495
Asn Tyr Asp Leu Tyr Leu Pro Glu Glu Lys Val Leu Pro Lys His Ser
500 505 510
Leu Leu Tyr Glu Thr Phe Ala Val Tyr Asn Glu Leu Thr Lys Val Lys
515 520 525
Phe Ile Ala Glu Gly Leu Arg Asp Tyr Gln Phe Leu Asp Ser Gly Gln

530 535 540

Lys Lys Gln Ile Val Asn Gln Leu Phe Lys Glu Asn Arg Lys Val Thr
545 550 555 560

Glu Lys Asp Ile Ile His Tyr Leu His Asn Val Asp Gly Tyr Asp Gly
565 570 575

Ile Glu Leu Lys Gly Ile Glu Lys Gln Phe Asn Ala Ser Leu Ser Thr
580 585 590

Tyr His Asp Leu Leu Lys Ile Ile Lys Asp Lys Glu Phe Met Asp Asp
595 600 605

Ala Lys Asn Glu Ala Ile Leu Glu Asn Ile Val His Thr Leu Thr Ile
610 615 620

Phe Glu Asp Arg Glu Met Ile Lys Gln Arg Leu Ala Gln Tyr Asp Ser
625 630 635 640

Leu Phe Asp Glu Lys Val Ile Lys Ala Leu Thr Arg Arg His Tyr Thr
645 650 655

Gly Trp Gly Lys Leu Ser Ala Lys Leu Ile Asn Gly Ile Cys Asp Lys
660 665 670

Gln Thr Gly Asn Thr Ile Leu Asp Tyr Leu Ile Asp Asp Gly Lys Ile
675 680 685

Asn Arg Asn Phe Met Gln Leu Ile Asn Asp Asp Gly Leu Ser Phe Lys
690 695 700

Glu Ile Ile Gln Lys Ala Gln Val Ile Gly Lys Thr Asp Asp Val Lys
705 710 715 720

Gln Val Val Gln Glu Leu Ser Gly Ser Pro Ala Ile Lys Lys Gly Ile
725 730 735

Leu Gln Ser Ile Lys Ile Val Asp Glu Leu Val Lys Val Met Gly His
740 745 750

Ala Pro Glu Ser Ile Val Ile Glu Met Ala Arg Glu Asn Gln Thr Thr
755 760 765

Ala Arg Gly Lys Lys Asn Ser Gln Gln Arg Tyr Lys Arg Ile Glu Asp
770 775 780

Ser Leu Lys Ile Leu Ala Ser Gly Leu Asp Ser Asn Ile Leu Lys Glu
785 790 795 800

Asn Pro Thr Asp Asn Asn Gln Leu Gln Asn Asp Arg Leu Phe Leu Tyr
805 810 815

Tyr Leu Gln Asn Gly Lys Asp Met Tyr Thr Gly Glu Ala Leu Asp Ile
820 825 830

Asn Gln Leu Ser Ser Tyr Asp Ile Asp His Ile Ile Pro Gln Ala Phe
835 840 845

Ile Lys Asp Asp Ser Leu Asp Asn Arg Val Leu Thr Ser Ser Lys Asp
850 855 860
Asn Arg Gly Lys Ser Asp Asn Val Pro Ser Ile Glu Val Val Gln Lys
865 870 875 880
Arg Lys Ala Phe Trp Gln Gln Leu Leu Asp Ser Lys Leu Ile Ser Glu
885 890 895
Arg Lys Phe Asn Asn Leu Thr Lys Ala Glu Arg Gly Gly Leu Asp Glu
900 905 910
Arg Asp Lys Val Gly Phe Ile Lys Arg Gln Leu Val Glu Thr Arg Gln
915 920 925
Ile Thr Lys His Val Ala Gln Ile Leu Asp Ala Arg Tyr Asn Thr Glu
930 935 940
Val Asn Glu Lys Asp Lys Lys Asn Arg Thr Val Lys Ile Ile Thr Leu
945 950 955 960
Lys Ser Asn Leu Val Ser Asn Phe Arg Lys Glu Phe Arg Leu Tyr Lys
965 970 975
Val Arg Glu Ile Asn Asp Tyr His His Ala His Asp Ala Tyr Leu Asn
980 985 990
Ala Val Val Ala Lys Ala Ile Leu Lys Lys Tyr Pro Lys Leu Glu Pro
995 1000 1005
Glu Phe Val Tyr Gly Glu Tyr Gln Lys Tyr Asp Leu Lys Arg Tyr
1010 1015 1020
Ile Ser Arg Ser Lys Asp Pro Lys Glu Val Glu Lys Ala Thr Glu
1025 1030 1035
Lys Tyr Phe Phe Tyr Ser Asn Leu Leu Asn Phe Phe Lys Glu Glu
1040 1045 1050
Val His Tyr Ala Asp Gly Thr Ile Val Lys Arg Glu Asn Ile Glu
1055 1060 1065
Tyr Ser Lys Asp Thr Gly Glu Ile Ala Trp Asn Lys Glu Lys Asp
1070 1075 1080
Phe Ala Ile Ile Lys Lys Val Leu Ser Leu Pro Gln Val Asn Ile
1085 1090 1095
Val Lys Lys Arg Glu Val Gln Thr Gly Gly Phe Ser Lys Glu Ser
1100 1105 1110
Ile Leu Pro Lys Gly Asn Ser Asp Lys Leu Ile Pro Arg Lys Thr
1115 1120 1125
Lys Asp Ile Leu Leu Asp Thr Thr Lys Tyr Gly Gly Phe Asp Ser
1130 1135 1140
Pro Val Ile Ala Tyr Ser Ile Leu Leu Ile Ala Asp Ile Glu Lys

1145 1150 1155
Gly Lys Ala Lys Lys Leu Lys Thr Val Lys Thr Leu Val Gly Ile
1160 1165 1170
Thr Ile Met Glu Lys Ala Ala Phe Glu Glu Asn Pro Ile Thr Phe
1175 1180 1185
Leu Glu Asn Lys Gly Tyr His Asn Val Arg Lys Glu Asn Ile Leu
1190 1195 1200
Cys Leu Pro Lys Tyr Ser Leu Phe Glu Leu Glu Asn Gly Arg Arg
1205 1210 1215
Arg Leu Leu Ala Ser Ala Lys Glu Leu Gln Lys Gly Asn Glu Ile
1220 1225 1230
Val Leu Pro Val Tyr Leu Thr Thr Leu Leu Tyr His Ser Lys Asn
1235 1240 1245
Val His Lys Leu Asp Glu Pro Gly His Leu Glu Tyr Ile Gln Lys
1250 1255 1260
His Arg Asn Glu Phe Lys Asp Leu Leu Asn Leu Val Ser Glu Phe
1265 1270 1275
Ser Gln Lys Tyr Val Leu Ala Asp Ala Asn Leu Glu Lys Ile Lys
1280 1285 1290
Ser Leu Tyr Ala Asp Asn Glu Gln Ala Asp Ile Glu Ile Leu Ala
1295 1300 1305
Asn Ser Phe Ile Asn Leu Leu Thr Phe Thr Ala Leu Gly Ala Pro
1310 1315 1320
Ala Ala Phe Lys Phe Phe Gly Lys Asp Ile Asp Arg Lys Arg Tyr
1325 1330 1335
Thr Thr Val Ser Glu Ile Leu Asn Ala Thr Leu Ile His Gln Ser
1340 1345 1350
Ile Thr Gly Leu Tyr Glu Thr Trp Ile Asp Leu Ser Lys Leu Gly
1355 1360 1365
Glu Asp
1370

<210> 26
<211> 1056
<212> PRT
<213> 金黄色葡萄球菌
<400> 26

Gly Ser His Met Lys Arg Asn Tyr Ile Leu Gly Leu Asp Ile Gly Ile
1 5 10 15
Thr Ser Val Gly Tyr Gly Ile Ile Asp Tyr Glu Thr Arg Asp Val Ile

20 25 30

Asp Ala Gly Val Arg Leu Phe Lys Glu Ala Asn Val Glu Asn Asn Glu

35 40 45

Gly Arg Arg Ser Lys Arg Gly Ala Arg Arg Leu Lys Arg Arg Arg Arg

50 55 60

His Arg Ile Gln Arg Val Lys Lys Leu Leu Phe Asp Tyr Asn Leu Leu

65 70 75 80

Thr Asp His Ser Glu Leu Ser Gly Ile Asn Pro Tyr Glu Ala Arg Val

85 90 95

Lys Gly Leu Ser Gln Lys Leu Ser Glu Glu Glu Phe Ser Ala Ala Leu

100 105 110

Leu His Leu Ala Lys Arg Arg Gly Val His Asn Val Asn Glu Val Glu

115 120 125

Glu Asp Thr Gly Asn Glu Leu Ser Thr Lys Glu Gln Ile Ser Arg Asn

130 135 140

Ser Lys Ala Leu Glu Glu Lys Tyr Val Ala Glu Leu Gln Leu Glu Arg

145 150 155 160

Leu Lys Lys Asp Gly Glu Val Arg Gly Ser Ile Asn Arg Phe Lys Thr

165 170 175

Ser Asp Tyr Val Lys Glu Ala Lys Gln Leu Leu Lys Val Gln Lys Ala

180 185 190

Tyr His Gln Leu Asp Gln Ser Phe Ile Asp Thr Tyr Ile Asp Leu Leu

195 200 205

Glu Thr Arg Arg Thr Tyr Tyr Glu Gly Pro Gly Glu Gly Ser Pro Phe

210 215 220

Gly Trp Lys Asp Ile Lys Glu Trp Tyr Glu Met Leu Met Gly His Cys

225 230 235 240

Thr Tyr Phe Pro Glu Glu Leu Arg Ser Val Lys Tyr Ala Tyr Asn Ala

245 250 255

Asp Leu Tyr Asn Ala Leu Asn Asp Leu Asn Asn Leu Val Ile Thr Arg

260 265 270

Asp Glu Asn Glu Lys Leu Glu Tyr Tyr Glu Lys Phe Gln Ile Ile Glu

275 280 285

Asn Val Phe Lys Gln Lys Lys Lys Pro Thr Leu Lys Gln Ile Ala Lys

290 295 300

Glu Ile Leu Val Asn Glu Glu Asp Ile Lys Gly Tyr Arg Val Thr Ser

305 310 315 320

Thr Gly Lys Pro Glu Phe Thr Asn Leu Lys Val Tyr His Asp Ile Lys

325 330 335

Asp Ile Thr Ala Arg Lys Glu Ile Ile Glu Asn Ala Glu Leu Leu Asp
340 345 350
Gln Ile Ala Lys Ile Leu Thr Ile Tyr Gln Ser Ser Glu Asp Ile Gln
355 360 365
Glu Glu Leu Thr Asn Leu Asn Ser Glu Leu Thr Gln Glu Glu Ile Glu
370 375 380
Gln Ile Ser Asn Leu Lys Gly Tyr Thr Gly Thr His Asn Leu Ser Leu
385 390 395 400
Lys Ala Ile Asn Leu Ile Leu Asp Glu Leu Trp His Thr Asn Asp Asn
405 410 415
Gln Ile Ala Ile Phe Asn Arg Leu Lys Leu Val Pro Lys Lys Val Asp
420 425 430
Leu Ser Gln Gln Lys Glu Ile Pro Thr Thr Leu Val Asp Asp Phe Ile
435 440 445
Leu Ser Pro Val Val Lys Arg Ser Phe Ile Gln Ser Ile Lys Val Ile
450 455 460
Asn Ala Ile Ile Lys Lys Tyr Gly Leu Pro Asn Asp Ile Ile Ile Glu
465 470 475 480
Leu Ala Arg Glu Lys Asn Ser Lys Asp Ala Gln Lys Met Ile Asn Glu
485 490 495
Met Gln Lys Arg Asn Arg Gln Thr Asn Glu Arg Ile Glu Glu Ile Ile
500 505 510
Arg Thr Thr Gly Lys Glu Asn Ala Lys Tyr Leu Ile Glu Lys Ile Lys
515 520 525
Leu His Asp Met Gln Glu Gly Lys Cys Leu Tyr Ser Leu Glu Ala Ile
530 535 540
Pro Leu Glu Asp Leu Leu Asn Asn Pro Phe Asn Tyr Glu Val Asp His
545 550 555 560
Ile Ile Pro Arg Ser Val Ser Phe Asp Asn Ser Phe Asn Asn Lys Val
565 570 575
Leu Val Lys Gln Glu Glu Ala Ser Lys Lys Gly Asn Arg Thr Pro Phe
580 585 590
Gln Tyr Leu Ser Ser Ser Asp Ser Lys Ile Ser Tyr Glu Thr Phe Lys
595 600 605
Lys His Ile Leu Asn Leu Ala Lys Gly Lys Gly Arg Ile Ser Lys Thr
610 615 620
Lys Lys Glu Tyr Leu Leu Glu Glu Arg Asp Ile Asn Arg Phe Ser Val
625 630 635 640
Gln Lys Asp Phe Ile Asn Arg Asn Leu Val Asp Thr Arg Tyr Ala Thr

645 650 655

Arg Gly Leu Met Asn Leu Leu Arg Ser Tyr Phe Arg Val Asn Asn Leu
660 665 670

Asp Val Lys Val Lys Ser Ile Asn Gly Gly Phe Thr Ser Phe Leu Arg
675 680 685

Arg Lys Trp Lys Phe Lys Lys Glu Arg Asn Lys Gly Tyr Lys His His
690 695 700

Ala Glu Asp Ala Leu Ile Ile Ala Asn Ala Asp Phe Ile Phe Lys Glu
705 710 715 720

Trp Lys Lys Leu Asp Lys Ala Lys Lys Val Met Glu Asn Gln Met Phe
725 730 735

Glu Glu Lys Gln Ala Glu Ser Met Pro Glu Ile Glu Thr Glu Gln Glu
740 745 750

Tyr Lys Glu Ile Phe Ile Thr Pro His Gln Ile Lys His Ile Lys Asp
755 760 765

Phe Lys Asp Tyr Lys Tyr Ser His Arg Val Asp Lys Lys Pro Asn Arg
770 775 780

Glu Leu Ile Asn Asp Thr Leu Tyr Ser Thr Arg Lys Asp Asp Lys Gly
785 790 795 800

Asn Thr Leu Ile Val Asn Asn Leu Asn Gly Leu Tyr Asp Lys Asp Asn
805 810 815

Asp Lys Leu Lys Lys Leu Ile Asn Lys Ser Pro Glu Lys Leu Leu Met
820 825 830

Tyr His His Asp Pro Gln Thr Tyr Gln Lys Leu Lys Leu Ile Met Glu
835 840 845

Gln Tyr Gly Asp Glu Lys Asn Pro Leu Tyr Lys Tyr Tyr Glu Glu Thr
850 855 860

Gly Asn Tyr Leu Thr Lys Tyr Ser Lys Lys Asp Asn Gly Pro Val Ile
865 870 875 880

Lys Lys Ile Lys Tyr Tyr Gly Asn Lys Leu Asn Ala His Leu Asp Ile
885 890 895

Thr Asp Asp Tyr Pro Asn Ser Arg Asn Lys Val Val Lys Leu Ser Leu
900 905 910

Lys Pro Tyr Arg Phe Asp Val Tyr Leu Asp Asn Gly Val Tyr Lys Phe
915 920 925

Val Thr Val Lys Asn Leu Asp Val Ile Lys Lys Glu Asn Tyr Tyr Glu
930 935 940

Val Asn Ser Lys Ala Tyr Glu Glu Ala Lys Lys Leu Lys Lys Ile Ser
945 950 955 960

Asn Gln Ala Glu Phe Ile Ala Ser Phe Tyr Asn Asn Asp Leu Ile Lys
965 970 975
Ile Asn Gly Glu Leu Tyr Arg Val Ile Gly Val Asn Asn Asp Leu Leu
980 985 990
Asn Arg Ile Glu Val Asn Met Ile Asp Ile Thr Tyr Arg Glu Tyr Leu
995 1000 1005
Glu Asn Met Asn Asp Lys Arg Pro Pro Arg Ile Ile Lys Thr Ile
1010 1015 1020
Ala Ser Lys Thr Gln Ser Ile Lys Lys Tyr Ser Thr Asp Ile Leu
1025 1030 1035
Gly Asn Leu Tyr Glu Val Lys Ser Lys Lys His Pro Gln Ile Ile
1040 1045 1050
Lys Lys Gly
1055

<210> 27
<211> 345
<212> PRT
<213> 冰岛硫化叶菌
<400> 27

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

Ala Gly Trp Val Leu Thr Arg Leu Gly Lys Ala Lys Val Ser Glu Gly
165 170 175
Asp Tyr Val Gly Val Asn Val Phe Thr Pro Thr Arg Gly Ile Leu Tyr
180 185 190
Ser Leu Ile Gln Asn Val Asn Gly Ile Val Pro Gly Ile Lys Pro Glu
195 200 205
Thr Ala Phe Gly Leu Trp Ile Ala Arg Lys Val Val Ser Ser Val Thr
210 215 220
Asn Pro Asn Val Ser Val Val Arg Ile Tyr Thr Ile Ser Asp Ala Val
225 230 235 240
Gly Gln Asn Pro Thr Thr Ile Asn Gly Gly Phe Ser Ile Asp Leu Thr
245 250 255
Lys Leu Leu Glu Lys Arg Tyr Leu Leu Ser Glu Arg Leu Glu Ala Ile
260 265 270
Ala Arg Asn Ala Leu Ser Ile Ser Ser Asn Met Arg Glu Arg Tyr Ile
275 280 285
Val Leu Ala Asn Tyr Ile Tyr Glu Tyr Leu Thr Gly Ser Lys Arg Leu
290 295 300
Glu Asp Leu Leu Tyr Phe Ala Asn Arg Asp Leu Ile Met Asn Leu Asn
305 310 315 320
Ser Asp Asp Gly Lys Val Arg Asp Leu Lys Leu Ile Ser Ala Tyr Val
325 330 335
Asn Gly Glu Leu Ile Arg Gly Glu Gly
340 345

<210> 28
<211> 345
<212> PRT
<213> 冰岛硫化叶菌
<400> 28

Met Glu Val Pro Leu Tyr Asn Ile Phe Gly Asp Asn Tyr Ile Ile Gln
1 5 10 15
Val Ala Thr Glu Ala Glu Asn Ser Thr Ile Tyr Asn Asn Lys Val Glu
20 25 30
Ile Asp Asp Glu Glu Leu Arg Asn Val Leu Asn Leu Ala Tyr Lys Ile
35 40 45
Ala Lys Asn Asn Glu Asp Ala Ala Ala Glu Arg Arg Gly Lys Ala Lys
50 55 60
Lys Lys Lys Gly Glu Glu Gly Glu Thr Thr Thr Ser Asn Ile Ile Leu
65 70 75 80

Pro Leu Ser Gly Asn Asp Lys Asn Pro Trp Thr Glu Thr Leu Lys Cys
85 90 95
Tyr Asn Phe Pro Thr Thr Val Ala Leu Ser Glu Val Phe Lys Asn Phe
100 105 110
Ser Gln Val Lys Glu Cys Glu Glu Val Ser Ala Pro Ser Phe Val Lys
115 120 125
Pro Glu Phe Tyr Lys Phe Gly Arg Ser Pro Gly Met Val Glu Arg Thr
130 135 140
Arg Arg Val Lys Leu Glu Val Glu Pro His Tyr Leu Ile Met Ala Ala
145 150 155 160
Ala Gly Trp Val Leu Thr Arg Leu Gly Lys Ala Lys Val Ser Glu Gly
165 170 175
Asp Tyr Val Gly Val Asn Val Phe Thr Pro Thr Arg Gly Ile Leu Tyr
180 185 190
Ser Leu Ile Gln Asn Val Asn Gly Ile Val Pro Gly Ile Lys Pro Glu
195 200 205
Thr Ala Phe Gly Leu Trp Ile Ala Arg Lys Val Val Ser Ser Val Thr
210 215 220
Asn Pro Asn Val Ser Val Val Ser Ile Tyr Thr Ile Ser Asp Ala Val
225 230 235 240
Gly Gln Asn Pro Thr Thr Ile Asn Gly Gly Phe Ser Ile Asp Leu Thr
245 250 255
Lys Leu Leu Glu Lys Arg Asp Leu Leu Ser Glu Arg Leu Glu Ala Ile
260 265 270
Ala Arg Asn Ala Leu Ser Ile Ser Ser Asn Met Arg Glu Arg Tyr Ile
275 280 285
Val Leu Ala Asn Tyr Ile Tyr Glu Tyr Leu Thr Gly Ser Lys Arg Leu
290 295 300
Glu Asp Leu Leu Tyr Phe Ala Asn Arg Asp Leu Ile Met Asn Leu Asn
305 310 315 320
Ser Asp Asp Gly Lys Val Arg Asp Leu Lys Leu Ile Ser Ala Tyr Val
325 330 335
Asn Gly Glu Leu Ile Arg Gly Glu Gly
340 345

<210> 29
<211> 1210
<212> PRT
<213> 未培养的Parcubacteria组细菌
<400> 29

Met Ser Lys Arg His Pro Arg Ile Ser Gly Val Lys Gly Tyr Arg Leu
1 5 10 15
His Ala Gln Arg Leu Glu Tyr Thr Gly Lys Ser Gly Ala Met Arg Thr
20 25 30
Ile Lys Tyr Pro Leu Tyr Ser Ser Pro Ser Gly Gly Arg Thr Val Pro
35 40 45
Arg Glu Ile Val Ser Ala Ile Asn Asp Asp Tyr Val Gly Leu Tyr Gly
50 55 60
Leu Ser Asn Phe Asp Asp Leu Tyr Asn Ala Glu Lys Arg Asn Glu Glu
65 70 75 80
Lys Val Tyr Ser Val Leu Asp Phe Trp Tyr Asp Cys Val Gln Tyr Gly
85 90 95
Ala Val Phe Ser Tyr Thr Ala Pro Gly Leu Leu Lys Asn Val Ala Glu
100 105 110
Val Arg Gly Gly Ser Tyr Glu Leu Thr Lys Thr Leu Lys Gly Ser His
115 120 125
Leu Tyr Asp Glu Leu Gln Ile Asp Lys Val Ile Lys Phe Leu Asn Lys
130 135 140
Lys Glu Ile Ser Arg Ala Asn Gly Ser Leu Asp Lys Leu Lys Lys Asp
145 150 155 160
Ile Ile Asp Cys Phe Lys Ala Glu Tyr Arg Glu Arg His Lys Asp Gln
165 170 175
Cys Asn Lys Leu Ala Asp Asp Ile Lys Asn Ala Lys Lys Asp Ala Gly
180 185 190
Ala Ser Leu Gly Glu Arg Gln Lys Lys Leu Phe Arg Asp Phe Phe Gly
195 200 205
Ile Ser Glu Gln Ser Glu Asn Asp Lys Pro Ser Phe Thr Asn Pro Leu
210 215 220
Asn Leu Thr Cys Cys Leu Leu Pro Phe Asp Thr Val Asn Asn Asn Arg
225 230 235 240
Asn Arg Gly Glu Val Leu Phe Asn Lys Leu Lys Glu Tyr Ala Gln Lys
245 250 255
Leu Asp Lys Asn Glu Gly Ser Leu Glu Met Trp Glu Tyr Ile Gly Ile
260 265 270
Gly Asn Ser Gly Thr Ala Phe Ser Asn Phe Leu Gly Glu Gly Phe Leu
275 280 285
Gly Arg Leu Arg Glu Asn Lys Ile Thr Glu Leu Lys Lys Ala Met Met
290 295 300
Asp Ile Thr Asp Ala Trp Arg Gly Gln Glu Gln Glu Glu Glu Leu Glu

305 310 315 320
Lys Arg Leu Arg Ile Leu Ala Ala Leu Thr Ile Lys Leu Arg Glu Pro
325 330 335
Lys Phe Asp Asn His Trp Gly Gly Tyr Arg Ser Asp Ile Asn Gly Lys
340 345 350
Leu Ser Ser Trp Leu Gln Asn Tyr Ile Asn Gln Thr Val Lys Ile Lys
355 360 365
Glu Asp Leu Lys Gly His Lys Lys Asp Leu Lys Lys Ala Lys Glu Met
370 375 380
Ile Asn Arg Phe Gly Glu Ser Asp Thr Lys Glu Glu Ala Val Val Ser
385 390 395 400
Ser Leu Leu Glu Ser Ile Glu Lys Ile Val Pro Asp Asp Ser Ala Asp
405 410 415
Asp Glu Lys Pro Asp Ile Pro Ala Ile Ala Ile Tyr Arg Arg Phe Leu
420 425 430
Ser Asp Gly Arg Leu Thr Leu Asn Arg Phe Val Gln Arg Glu Asp Val
435 440 445
Gln Glu Ala Leu Ile Lys Glu Arg Leu Glu Ala Glu Lys Lys Lys Lys
450 455 460
Pro Lys Lys Arg Lys Lys Lys Ser Asp Ala Glu Asp Glu Lys Glu Thr
465 470 475 480
Ile Asp Phe Lys Glu Leu Phe Pro His Leu Ala Lys Pro Leu Lys Leu
485 490 495
Val Pro Asn Phe Tyr Gly Asp Ser Lys Arg Glu Leu Tyr Lys Lys Tyr
500 505 510
Lys Asn Ala Ala Ile Tyr Thr Asp Ala Leu Trp Lys Ala Val Glu Lys
515 520 525
Ile Tyr Lys Ser Ala Phe Ser Ser Ser Leu Lys Asn Ser Phe Phe Asp
530 535 540
Thr Asp Phe Asp Lys Asp Phe Phe Ile Lys Arg Leu Gln Lys Ile Phe
545 550 555 560
Ser Val Tyr Arg Arg Phe Asn Thr Asp Lys Trp Lys Pro Ile Val Lys
565 570 575
Asn Ser Phe Ala Pro Tyr Cys Asp Ile Val Ser Leu Ala Glu Asn Glu
580 585 590
Val Leu Tyr Lys Pro Lys Gln Ser Arg Ser Arg Lys Ser Ala Ala Ile
595 600 605
Asp Lys Asn Arg Val Arg Leu Pro Ser Thr Glu Asn Ile Ala Lys Ala
610 615 620

Gly Ile Ala Leu Ala Arg Glu Leu Ser Val Ala Gly Phe Asp Trp Lys
625 630 635 640
Asp Leu Leu Lys Lys Glu Glu His Glu Glu Tyr Ile Asp Leu Ile Glu
645 650 655
Leu His Lys Thr Ala Leu Ala Leu Leu Leu Ala Val Thr Glu Thr Gln
660 665 670
Leu Asp Ile Ser Ala Leu Asp Phe Val Glu Asn Gly Thr Val Lys Asp
675 680 685
Phe Met Lys Thr Arg Asp Gly Asn Leu Val Leu Glu Gly Arg Phe Leu
690 695 700
Glu Met Phe Ser Gln Ser Ile Val Phe Ser Glu Leu Arg Gly Leu Ala
705 710 715 720
Gly Leu Met Ser Arg Lys Glu Phe Ile Thr Arg Ser Ala Ile Gln Thr
725 730 735
Met Asn Gly Lys Gln Ala Glu Leu Leu Tyr Ile Pro His Glu Phe Gln
740 745 750
Ser Ala Lys Ile Thr Thr Pro Lys Glu Met Ser Arg Ala Phe Leu Asp
755 760 765
Leu Ala Pro Ala Glu Phe Ala Thr Ser Leu Glu Pro Glu Ser Leu Ser
770 775 780
Glu Lys Ser Leu Leu Lys Leu Lys Gln Met Arg Tyr Tyr Pro His Tyr
785 790 795 800
Phe Gly Tyr Glu Leu Thr Arg Thr Gly Gln Gly Ile Asp Gly Gly Val
805 810 815
Ala Glu Asn Ala Leu Arg Leu Glu Lys Ser Pro Val Lys Lys Arg Glu
820 825 830
Ile Lys Cys Lys Gln Tyr Lys Thr Leu Gly Arg Gly Gln Asn Lys Ile
835 840 845
Val Leu Tyr Val Arg Ser Ser Tyr Tyr Gln Thr Gln Phe Leu Glu Trp
850 855 860
Phe Leu His Arg Pro Lys Asn Val Gln Thr Asp Val Ala Val Ser Gly
865 870 875 880
Ser Phe Leu Ile Asp Glu Lys Lys Val Lys Thr Arg Trp Asn Tyr Asp
885 890 895
Ala Leu Thr Val Ala Leu Glu Pro Val Ser Gly Ser Glu Arg Val Phe
900 905 910
Val Ser Gln Pro Phe Thr Ile Phe Pro Glu Lys Ser Ala Glu Glu Glu
915 920 925
Gly Gln Arg Tyr Leu Gly Ile Asp Ile Gly Glu Tyr Gly Ile Ala Tyr

930 935 940

Thr Ala Leu Glu Ile Thr Gly Asp Ser Ala Lys Ile Leu Asp Gln Asn
945 950 955 960

Phe Ile Ser Asp Pro Gln Leu Lys Thr Leu Arg Glu Glu Val Lys Gly
965 970 975

Leu Lys Leu Asp Gln Arg Arg Gly Thr Phe Ala Met Pro Ser Thr Lys
980 985 990

Ile Ala Arg Ile Arg Glu Ser Leu Val His Ser Leu Arg Asn Arg Ile
995 1000 1005

His His Leu Ala Leu Lys His Lys Ala Lys Ile Val Tyr Glu Leu
1010 1015 1020

Glu Val Ser Arg Phe Glu Glu Gly Lys Gln Lys Ile Lys Lys Val
1025 1030 1035

Tyr Ala Thr Leu Lys Lys Ala Asp Val Tyr Ser Glu Ile Asp Ala
1040 1045 1050

Asp Lys Asn Leu Gln Thr Thr Val Trp Gly Lys Leu Ala Val Ala
1055 1060 1065

Ser Glu Ile Ser Ala Ser Tyr Thr Ser Gln Phe Cys Gly Ala Cys
1070 1075 1080

Lys Lys Leu Trp Arg Ala Glu Met Gln Val Asp Glu Thr Ile Thr
1085 1090 1095

Thr Gln Glu Leu Ile Gly Thr Val Arg Val Ile Lys Gly Gly Thr
1100 1105 1110

Leu Ile Asp Ala Ile Lys Asp Phe Met Arg Pro Pro Ile Phe Asp
1115 1120 1125

Glu Asn Asp Thr Pro Phe Pro Lys Tyr Arg Asp Phe Cys Asp Lys
1130 1135 1140

His His Ile Ser Lys Lys Met Arg Gly Asn Ser Cys Leu Phe Ile
1145 1150 1155

Cys Pro Phe Cys Arg Ala Asn Ala Asp Ala Asp Ile Gln Ala Ser
1160 1165 1170

Gln Thr Ile Ala Leu Leu Arg Tyr Val Lys Glu Glu Lys Lys Val
1175 1180 1185

Glu Asp Tyr Phe Glu Arg Phe Arg Lys Leu Lys Asn Ile Lys Val
1190 1195 1200

Leu Gly Gln Met Lys Lys Ile
1205 1210

<210> 30

<211> 1300

<212> PRT

<213> 新凶手弗朗西斯菌

<400> 30

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

Phe Val Asn Gly Glu Asn Thr Lys Arg Lys Gly Ile Asn Glu Tyr Ile
290 295 300
Asn Leu Tyr Ser Gln Gln Ile Asn Asp Lys Thr Leu Lys Lys Tyr Lys
305 310 315 320
Met Ser Val Leu Phe Lys Gln Ile Leu Ser Asp Thr Glu Ser Lys Ser
325 330 335
Phe Val Ile Asp Lys Leu Glu Asp Asp Ser Asp Val Val Thr Thr Met
340 345 350
Gln Ser Phe Tyr Glu Gln Ile Ala Ala Phe Lys Thr Val Glu Glu Lys
355 360 365
Ser Ile Lys Glu Thr Leu Ser Leu Leu Phe Asp Asp Leu Lys Ala Gln
370 375 380
Lys Leu Asp Leu Ser Lys Ile Tyr Phe Lys Asn Asp Lys Ser Leu Thr
385 390 395 400
Asp Leu Ser Gln Gln Val Phe Asp Asp Tyr Ser Val Ile Gly Thr Ala
405 410 415
Val Leu Glu Tyr Ile Thr Gln Gln Ile Ala Pro Lys Asn Leu Asp Asn
420 425 430
Pro Ser Lys Lys Glu Gln Glu Leu Ile Ala Lys Lys Thr Glu Lys Ala
435 440 445
Lys Tyr Leu Ser Leu Glu Thr Ile Lys Leu Ala Leu Glu Glu Phe Asn
450 455 460
Lys His Arg Asp Ile Asp Lys Gln Cys Arg Phe Glu Glu Ile Leu Ala
465 470 475 480
Asn Phe Ala Ala Ile Pro Met Ile Phe Asp Glu Ile Ala Gln Asn Lys
485 490 495
Asp Asn Leu Ala Gln Ile Ser Ile Lys Tyr Gln Asn Gln Gly Lys Lys
500 505 510
Asp Leu Leu Gln Ala Ser Ala Glu Asp Asp Val Lys Ala Ile Lys Asp
515 520 525
Leu Leu Asp Gln Thr Asn Asn Leu Leu His Lys Leu Lys Ile Phe His
530 535 540
Ile Ser Gln Ser Glu Asp Lys Ala Asn Ile Leu Asp Lys Asp Glu His
545 550 555 560
Phe Tyr Leu Val Phe Glu Glu Cys Tyr Phe Glu Leu Ala Asn Ile Val
565 570 575
Pro Leu Tyr Asn Lys Ile Arg Asn Tyr Ile Thr Gln Lys Pro Tyr Ser
580 585 590
Asp Glu Lys Phe Lys Leu Asn Phe Glu Asn Ser Thr Leu Ala Asn Gly

595 600 605

Trp Asp Lys Asn Lys Glu Pro Asp Asn Thr Ala Ile Leu Phe Ile Lys

610 615 620

Asp Asp Lys Tyr Tyr Leu Gly Val Met Asn Lys Lys Asn Asn Lys Ile

625 630 635 640

Phe Asp Asp Lys Ala Ile Lys Glu Asn Lys Gly Glu Gly Tyr Lys Lys

645 650 655

Ile Val Tyr Lys Leu Leu Pro Gly Ala Asn Lys Met Leu Pro Lys Val

660 665 670

Phe Phe Ser Ala Lys Ser Ile Lys Phe Tyr Asn Pro Ser Glu Asp Ile

675 680 685

Leu Arg Ile Arg Asn His Ser Thr His Thr Lys Asn Gly Ser Pro Gln

690 695 700

Lys Gly Tyr Glu Lys Phe Glu Phe Asn Ile Glu Asp Cys Arg Lys Phe

705 710 715 720

Ile Asp Phe Tyr Lys Gln Ser Ile Ser Lys His Pro Glu Trp Lys Asp

725 730 735

Phe Gly Phe Arg Phe Ser Asp Thr Gln Arg Tyr Asn Ser Ile Asp Glu

740 745 750

Phe Tyr Arg Glu Val Glu Asn Gln Gly Tyr Lys Leu Thr Phe Glu Asn

755 760 765

Ile Ser Glu Ser Tyr Ile Asp Ser Val Val Asn Gln Gly Lys Leu Tyr

770 775 780

Leu Phe Gln Ile Tyr Asn Lys Asp Phe Ser Ala Tyr Ser Lys Gly Arg

785 790 795 800

Pro Asn Leu His Thr Leu Tyr Trp Lys Ala Leu Phe Asp Glu Arg Asn

805 810 815

Leu Gln Asp Val Val Tyr Lys Leu Asn Gly Glu Ala Glu Leu Phe Tyr

820 825 830

Arg Lys Gln Ser Ile Pro Lys Lys Ile Thr His Pro Ala Lys Glu Ala

835 840 845

Ile Ala Asn Lys Asn Lys Asp Asn Pro Lys Lys Glu Ser Val Phe Glu

850 855 860

Tyr Asp Leu Ile Lys Asp Lys Arg Phe Thr Glu Asp Lys Phe Phe Phe

865 870 875 880

His Cys Pro Ile Thr Ile Asn Phe Lys Ser Ser Gly Ala Asn Lys Phe

885 890 895

Asn Asp Glu Ile Asn Leu Leu Leu Lys Glu Lys Ala Asn Asp Val His

900 905 910

Ile Leu Ser Ile Asp Arg Gly Glu Arg His Leu Ala Tyr Tyr Thr Leu
915 920 925
Val Asp Gly Lys Gly Asn Ile Ile Lys Gln Asp Thr Phe Asn Ile Ile
930 935 940
Gly Asn Asp Arg Met Lys Thr Asn Tyr His Asp Lys Leu Ala Ala Ile
945 950 955 960
Glu Lys Asp Arg Asp Ser Ala Arg Lys Asp Trp Lys Lys Ile Asn Asn
965 970 975
Ile Lys Glu Met Lys Glu Gly Tyr Leu Ser Gln Val Val His Glu Ile
980 985 990
Ala Lys Leu Val Ile Glu Tyr Asn Ala Ile Val Val Phe Glu Asp Leu
995 1000 1005
Asn Phe Gly Phe Lys Arg Gly Arg Phe Lys Val Glu Lys Gln Val
1010 1015 1020
Tyr Gln Lys Leu Glu Lys Met Leu Ile Glu Lys Leu Asn Tyr Leu
1025 1030 1035
Val Phe Lys Asp Asn Glu Phe Asp Lys Thr Gly Gly Val Leu Arg
1040 1045 1050
Ala Tyr Gln Leu Thr Ala Pro Phe Glu Thr Phe Lys Lys Met Gly
1055 1060 1065
Lys Gln Thr Gly Ile Ile Tyr Tyr Val Pro Ala Gly Phe Thr Ser
1070 1075 1080
Lys Ile Cys Pro Val Thr Gly Phe Val Asn Gln Leu Tyr Pro Lys
1085 1090 1095
Tyr Glu Ser Val Ser Lys Ser Gln Glu Phe Phe Ser Lys Phe Asp
1100 1105 1110
Lys Ile Cys Tyr Asn Leu Asp Lys Gly Tyr Phe Glu Phe Ser Phe
1115 1120 1125
Asp Tyr Lys Asn Phe Gly Asp Lys Ala Ala Lys Gly Lys Trp Thr
1130 1135 1140
Ile Ala Ser Phe Gly Ser Arg Leu Ile Asn Phe Arg Asn Ser Asp
1145 1150 1155
Lys Asn His Asn Trp Asp Thr Arg Glu Val Tyr Pro Thr Lys Glu
1160 1165 1170
Leu Glu Lys Leu Leu Lys Asp Tyr Ser Ile Glu Tyr Gly His Gly
1175 1180 1185
Glu Cys Ile Lys Ala Ala Ile Cys Gly Glu Ser Asp Lys Lys Phe
1190 1195 1200
Phe Ala Lys Leu Thr Ser Val Leu Asn Thr Ile Leu Gln Met Arg

1205 1210 1215
Asn Ser Lys Thr Gly Thr Glu Leu Asp Tyr Leu Ile Ser Pro Val
1220 1225 1230
Ala Asp Val Asn Gly Asn Phe Phe Asp Ser Arg Gln Ala Pro Lys
1235 1240 1245
Asn Met Pro Gln Asp Ala Asp Ala Asn Gly Ala Tyr His Ile Gly
1250 1255 1260
Leu Lys Gly Leu Met Leu Leu Gly Arg Ile Lys Asn Asn Gln Glu
1265 1270 1275
Gly Lys Lys Leu Asn Leu Val Ile Lys Asn Glu Glu Tyr Phe Glu
1280 1285 1290
Phe Val Gln Asn Arg Asn Asn
1295 1300

<210> 31
<211> 1300
<212> PRT
<213> 人工序列
<220>
<223> 合成多肽
<400> 31

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

145 150 155 160
Asp Ile Asp Glu Ala Leu Glu Ile Ile Lys Ser Phe Lys Gly Trp Thr
165 170 175
Thr Tyr Phe Lys Gly Phe His Glu Asn Arg Lys Asn Val Tyr Ser Ser
180 185 190
Asn Asp Ile Pro Thr Ser Ile Ile Tyr Arg Ile Val Asp Asp Asn Leu
195 200 205
Pro Lys Phe Leu Glu Asn Lys Ala Lys Tyr Glu Ser Leu Lys Asp Lys
210 215 220
Ala Pro Glu Ala Ile Asn Tyr Glu Gln Ile Lys Lys Asp Leu Ala Glu
225 230 235 240
Glu Leu Thr Phe Asp Ile Asp Tyr Lys Thr Ser Glu Val Asn Gln Arg
245 250 255
Val Phe Ser Leu Asp Glu Val Phe Glu Ile Ala Asn Phe Asn Asn Tyr
260 265 270
Leu Asn Gln Ser Gly Ile Thr Lys Phe Asn Thr Ile Ile Gly Gly Lys
275 280 285
Phe Val Asn Gly Glu Asn Thr Lys Arg Lys Gly Ile Asn Glu Tyr Ile
290 295 300
Asn Leu Tyr Ser Gln Gln Ile Asn Asp Lys Thr Leu Lys Lys Tyr Lys
305 310 315 320
Met Ser Val Leu Phe Lys Gln Ile Leu Ser Asp Thr Glu Ser Lys Ser
325 330 335
Phe Val Ile Asp Lys Leu Glu Asp Asp Ser Asp Val Val Thr Thr Met
340 345 350
Gln Ser Phe Tyr Glu Gln Ile Ala Ala Phe Lys Thr Val Glu Glu Lys
355 360 365
Ser Ile Lys Glu Thr Leu Ser Leu Leu Phe Asp Asp Leu Lys Ala Gln
370 375 380
Lys Leu Asp Leu Ser Lys Ile Tyr Phe Lys Asn Asp Lys Ser Leu Thr
385 390 395 400
Asp Leu Ser Gln Gln Val Phe Asp Asp Tyr Ser Val Ile Gly Thr Ala
405 410 415
Val Leu Glu Tyr Ile Thr Gln Gln Ile Ala Pro Lys Asn Leu Asp Asn
420 425 430
Pro Ser Lys Lys Glu Gln Glu Leu Ile Ala Lys Lys Thr Glu Lys Ala
435 440 445
Lys Tyr Leu Ser Leu Glu Thr Ile Lys Leu Ala Leu Glu Glu Phe Asn
450 455 460

Lys His Arg Asp Ile Asp Lys Gln Cys Arg Phe Glu Glu Ile Leu Ala
465 470 475 480
Asn Phe Ala Ala Ile Pro Met Ile Phe Asp Glu Ile Ala Gln Asn Lys
485 490 495
Asp Asn Leu Ala Gln Ile Ser Ile Lys Tyr Gln Asn Gln Gly Lys Lys
500 505 510
Asp Leu Leu Gln Ala Ser Ala Glu Asp Asp Val Lys Ala Ile Lys Asp
515 520 525
Leu Leu Asp Gln Thr Asn Asn Leu Leu His Lys Leu Lys Ile Phe His
530 535 540
Ile Ser Gln Ser Glu Asp Lys Ala Asn Ile Leu Asp Lys Asp Glu His
545 550 555 560
Phe Tyr Leu Val Phe Glu Glu Cys Tyr Phe Glu Leu Ala Asn Ile Val
565 570 575
Pro Leu Tyr Asn Lys Ile Arg Asn Tyr Ile Thr Gln Lys Pro Tyr Ser
580 585 590
Asp Glu Lys Phe Lys Leu Asn Phe Glu Asn Ser Thr Leu Ala Asn Gly
595 600 605
Trp Asp Lys Asn Lys Glu Pro Asp Asn Thr Ala Ile Leu Phe Ile Lys
610 615 620
Asp Asp Lys Tyr Tyr Leu Gly Val Met Asn Lys Lys Asn Asn Lys Ile
625 630 635 640
Phe Asp Asp Lys Ala Ile Lys Glu Asn Lys Gly Glu Gly Tyr Lys Lys
645 650 655
Ile Val Tyr Lys Leu Leu Pro Gly Ala Asn Lys Met Leu Pro Lys Val
660 665 670
Phe Phe Ser Ala Lys Ser Ile Lys Phe Tyr Asn Pro Ser Glu Asp Ile
675 680 685
Leu Arg Ile Arg Asn His Ser Thr His Thr Lys Asn Gly Ser Pro Gln
690 695 700
Lys Gly Tyr Glu Lys Phe Glu Phe Asn Ile Glu Asp Cys Arg Lys Phe
705 710 715 720
Ile Asp Phe Tyr Lys Gln Ser Ile Ser Lys His Pro Glu Trp Lys Asp
725 730 735
Phe Gly Phe Arg Phe Ser Asp Thr Gln Arg Tyr Asn Ser Ile Asp Glu
740 745 750
Phe Tyr Arg Glu Val Glu Asn Gln Gly Tyr Lys Leu Thr Phe Glu Asn
755 760 765
Ile Ser Glu Ser Tyr Ile Asp Ser Val Val Asn Gln Gly Lys Leu Tyr

770 775 780

Leu Phe Gln Ile Tyr Asn Lys Asp Phe Ser Ala Tyr Ser Lys Gly Arg
785 790 795 800

Pro Asn Leu His Thr Leu Tyr Trp Lys Ala Leu Phe Asp Glu Arg Asn
805 810 815

Leu Gln Asp Val Val Tyr Lys Leu Asn Gly Glu Ala Glu Leu Phe Tyr
820 825 830

Arg Lys Gln Ser Ile Pro Lys Lys Ile Thr His Pro Ala Lys Glu Ala
835 840 845

Ile Ala Asn Lys Asn Lys Asp Asn Pro Lys Lys Glu Ser Val Phe Glu
850 855 860

Tyr Asp Leu Ile Lys Asp Lys Arg Phe Thr Glu Asp Lys Phe Phe Phe
865 870 875 880

His Cys Pro Ile Thr Ile Asn Phe Lys Ser Ser Gly Ala Asn Lys Phe
885 890 895

Asn Asp Glu Ile Asn Leu Leu Leu Lys Glu Lys Ala Asn Asp Val His
900 905 910

Ile Leu Ser Ile Ala Arg Gly Glu Arg His Leu Ala Tyr Tyr Thr Leu
915 920 925

Val Asp Gly Lys Gly Asn Ile Ile Lys Gln Asp Thr Phe Asn Ile Ile
930 935 940

Gly Asn Asp Arg Met Lys Thr Asn Tyr His Asp Lys Leu Ala Ala Ile
945 950 955 960

Glu Lys Asp Arg Asp Ser Ala Arg Lys Asp Trp Lys Lys Ile Asn Asn
965 970 975

Ile Lys Glu Met Lys Glu Gly Tyr Leu Ser Gln Val Val His Glu Ile
980 985 990

Ala Lys Leu Val Ile Glu Tyr Asn Ala Ile Val Val Phe Glu Asp Leu
995 1000 1005

Asn Phe Gly Phe Lys Arg Gly Arg Phe Lys Val Glu Lys Gln Val
1010 1015 1020

Tyr Gln Lys Leu Glu Lys Met Leu Ile Glu Lys Leu Asn Tyr Leu
1025 1030 1035

Val Phe Lys Asp Asn Glu Phe Asp Lys Thr Gly Gly Val Leu Arg
1040 1045 1050

Ala Tyr Gln Leu Thr Ala Pro Phe Glu Thr Phe Lys Lys Met Gly
1055 1060 1065

Lys Gln Thr Gly Ile Ile Tyr Tyr Val Pro Ala Gly Phe Thr Ser
1070 1075 1080

Lys Ile Cys Pro Val Thr Gly Phe Val Asn Gln Leu Tyr Pro Lys
1085 1090 1095
Tyr Glu Ser Val Ser Lys Ser Gln Glu Phe Phe Ser Lys Phe Asp
1100 1105 1110
Lys Ile Cys Tyr Asn Leu Asp Lys Gly Tyr Phe Glu Phe Ser Phe
1115 1120 1125
Asp Tyr Lys Asn Phe Gly Asp Lys Ala Ala Lys Gly Lys Trp Thr
1130 1135 1140
Ile Ala Ser Phe Gly Ser Arg Leu Ile Asn Phe Arg Asn Ser Asp
1145 1150 1155
Lys Asn His Asn Trp Asp Thr Arg Glu Val Tyr Pro Thr Lys Glu
1160 1165 1170
Leu Glu Lys Leu Leu Lys Asp Tyr Ser Ile Glu Tyr Gly His Gly
1175 1180 1185
Glu Cys Ile Lys Ala Ala Ile Cys Gly Glu Ser Asp Lys Lys Phe
1190 1195 1200
Phe Ala Lys Leu Thr Ser Val Leu Asn Thr Ile Leu Gln Met Arg
1205 1210 1215
Asn Ser Lys Thr Gly Thr Glu Leu Asp Tyr Leu Ile Ser Pro Val
1220 1225 1230
Ala Asp Val Asn Gly Asn Phe Phe Asp Ser Arg Gln Ala Pro Lys
1235 1240 1245
Asn Met Pro Gln Asp Ala Asp Ala Asn Gly Ala Tyr His Ile Gly
1250 1255 1260
Leu Lys Gly Leu Met Leu Leu Gly Arg Ile Lys Asn Asn Gln Glu
1265 1270 1275
Gly Lys Lys Leu Asn Leu Val Ile Lys Asn Glu Glu Tyr Phe Glu
1280 1285 1290
Phe Val Gln Asn Arg Asn Asn
1295 1300

<210> 32
<211> 1300
<212> PRT
<213> 人工序列
<220>
<223> 合成多肽
<400> 32

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

Leu Arg Phe Glu Leu Ile Pro Gln Gly Lys Thr Leu Glu Asn Ile Lys
20 25 30
Ala Arg Gly Leu Ile Leu Asp Asp Glu Lys Arg Ala Lys Asp Tyr Lys
35 40 45
Lys Ala Lys Gln Ile Ile Asp Lys Tyr His Gln Phe Phe Ile Glu Glu
50 55 60
Ile Leu Ser Ser Val Cys Ile Ser Glu Asp Leu Leu Gln Asn Tyr Ser
65 70 75 80
Asp Val Tyr Phe Lys Leu Lys Lys Ser Asp Asp Asp Asn Leu Gln Lys
85 90 95
Asp Phe Lys Ser Ala Lys Asp Thr Ile Lys Lys Gln Ile Ser Glu Tyr
100 105 110
Ile Lys Asp Ser Glu Lys Phe Lys Asn Leu Phe Asn Gln Asn Leu Ile
115 120 125
Asp Ala Lys Lys Gly Gln Glu Ser Asp Leu Ile Leu Trp Leu Lys Gln
130 135 140
Ser Lys Asp Asn Gly Ile Glu Leu Phe Lys Ala Asn Ser Asp Ile Thr
145 150 155 160
Asp Ile Asp Glu Ala Leu Glu Ile Ile Lys Ser Phe Lys Gly Trp Thr
165 170 175
Thr Tyr Phe Lys Gly Phe His Glu Asn Arg Lys Asn Val Tyr Ser Ser
180 185 190
Asn Asp Ile Pro Thr Ser Ile Ile Tyr Arg Ile Val Asp Asp Asn Leu
195 200 205
Pro Lys Phe Leu Glu Asn Lys Ala Lys Tyr Glu Ser Leu Lys Asp Lys
210 215 220
Ala Pro Glu Ala Ile Asn Tyr Glu Gln Ile Lys Lys Asp Leu Ala Glu
225 230 235 240
Glu Leu Thr Phe Asp Ile Asp Tyr Lys Thr Ser Glu Val Asn Gln Arg
245 250 255
Val Phe Ser Leu Asp Glu Val Phe Glu Ile Ala Asn Phe Asn Asn Tyr
260 265 270
Leu Asn Gln Ser Gly Ile Thr Lys Phe Asn Thr Ile Ile Gly Gly Lys
275 280 285
Phe Val Asn Gly Glu Asn Thr Lys Arg Lys Gly Ile Asn Glu Tyr Ile
290 295 300
Asn Leu Tyr Ser Gln Gln Ile Asn Asp Lys Thr Leu Lys Lys Tyr Lys
305 310 315 320
Met Ser Val Leu Phe Lys Gln Ile Leu Ser Asp Thr Glu Ser Lys Ser

325 330 335
Phe Val Ile Asp Lys Leu Glu Asp Asp Ser Asp Val Val Thr Thr Met
340 345 350
Gln Ser Phe Tyr Glu Gln Ile Ala Ala Phe Lys Thr Val Glu Glu Lys
355 360 365
Ser Ile Lys Glu Thr Leu Ser Leu Leu Phe Asp Asp Leu Lys Ala Gln
370 375 380
Lys Leu Asp Leu Ser Lys Ile Tyr Phe Lys Asn Asp Lys Ser Leu Thr
385 390 395 400
Asp Leu Ser Gln Gln Val Phe Asp Asp Tyr Ser Val Ile Gly Thr Ala
405 410 415
Val Leu Glu Tyr Ile Thr Gln Gln Ile Ala Pro Lys Asn Leu Asp Asn
420 425 430
Pro Ser Lys Lys Glu Gln Glu Leu Ile Ala Lys Lys Thr Glu Lys Ala
435 440 445
Lys Tyr Leu Ser Leu Glu Thr Ile Lys Leu Ala Leu Glu Glu Phe Asn
450 455 460
Lys His Arg Asp Ile Asp Lys Gln Cys Arg Phe Glu Glu Ile Leu Ala
465 470 475 480
Asn Phe Ala Ala Ile Pro Met Ile Phe Asp Glu Ile Ala Gln Asn Lys
485 490 495
Asp Asn Leu Ala Gln Ile Ser Ile Lys Tyr Gln Asn Gln Gly Lys Lys
500 505 510
Asp Leu Leu Gln Ala Ser Ala Glu Asp Asp Val Lys Ala Ile Lys Asp
515 520 525
Leu Leu Asp Gln Thr Asn Asn Leu Leu His Lys Leu Lys Ile Phe His
530 535 540
Ile Ser Gln Ser Glu Asp Lys Ala Asn Ile Leu Asp Lys Asp Glu His
545 550 555 560
Phe Tyr Leu Val Phe Glu Glu Cys Tyr Phe Glu Leu Ala Asn Ile Val
565 570 575
Pro Leu Tyr Asn Lys Ile Arg Asn Tyr Ile Thr Gln Lys Pro Tyr Ser
580 585 590
Asp Glu Lys Phe Lys Leu Asn Phe Glu Asn Ser Thr Leu Ala Asn Gly
595 600 605
Trp Asp Lys Asn Lys Glu Pro Asp Asn Thr Ala Ile Leu Phe Ile Lys
610 615 620
Asp Asp Lys Tyr Tyr Leu Gly Val Met Asn Lys Lys Asn Asn Lys Ile
625 630 635 640

Phe Asp Asp Lys Ala Ile Lys Glu Asn Lys Gly Glu Gly Tyr Lys Lys
645 650 655
Ile Val Tyr Lys Leu Leu Pro Gly Ala Asn Lys Met Leu Pro Lys Val
660 665 670
Phe Phe Ser Ala Lys Ser Ile Lys Phe Tyr Asn Pro Ser Glu Asp Ile
675 680 685
Leu Arg Ile Arg Asn His Ser Thr His Thr Lys Asn Gly Ser Pro Gln
690 695 700
Lys Gly Tyr Glu Lys Phe Glu Phe Asn Ile Glu Asp Cys Arg Lys Phe
705 710 715 720
Ile Asp Phe Tyr Lys Gln Ser Ile Ser Lys His Pro Glu Trp Lys Asp
725 730 735
Phe Gly Phe Arg Phe Ser Asp Thr Gln Arg Tyr Asn Ser Ile Asp Glu
740 745 750
Phe Tyr Arg Glu Val Glu Asn Gln Gly Tyr Lys Leu Thr Phe Glu Asn
755 760 765
Ile Ser Glu Ser Tyr Ile Asp Ser Val Val Asn Gln Gly Lys Leu Tyr
770 775 780
Leu Phe Gln Ile Tyr Asn Lys Asp Phe Ser Ala Tyr Ser Lys Gly Arg
785 790 795 800
Pro Asn Leu His Thr Leu Tyr Trp Lys Ala Leu Phe Asp Glu Arg Asn
805 810 815
Leu Gln Asp Val Val Tyr Lys Leu Asn Gly Glu Ala Glu Leu Phe Tyr
820 825 830
Arg Lys Gln Ser Ile Pro Lys Lys Ile Thr His Pro Ala Lys Glu Ala
835 840 845
Ile Ala Asn Lys Asn Lys Asp Asn Pro Lys Lys Glu Ser Val Phe Glu
850 855 860
Tyr Asp Leu Ile Lys Asp Lys Arg Phe Thr Glu Asp Lys Phe Phe Phe
865 870 875 880
His Cys Pro Ile Thr Ile Asn Phe Lys Ser Ser Gly Ala Asn Lys Phe
885 890 895
Asn Asp Glu Ile Asn Leu Leu Leu Lys Glu Lys Ala Asn Asp Val His
900 905 910
Ile Leu Ser Ile Asp Arg Gly Glu Arg His Leu Ala Tyr Tyr Thr Leu
915 920 925
Val Asp Gly Lys Gly Asn Ile Ile Lys Gln Asp Thr Phe Asn Ile Ile
930 935 940
Gly Asn Asp Arg Met Lys Thr Asn Tyr His Asp Lys Leu Ala Ala Ile

945 950 955 960
Glu Lys Asp Arg Asp Ser Ala Arg Lys Asp Trp Lys Lys Ile Asn Asn
965 970 975
Ile Lys Glu Met Lys Glu Gly Tyr Leu Ser Gln Val Val His Glu Ile
980 985 990
Ala Lys Leu Val Ile Glu Tyr Asn Ala Ile Val Val Phe Ala Asp Leu
995 1000 1005
Asn Phe Gly Phe Lys Arg Gly Arg Phe Lys Val Glu Lys Gln Val
1010 1015 1020
Tyr Gln Lys Leu Glu Lys Met Leu Ile Glu Lys Leu Asn Tyr Leu
1025 1030 1035
Val Phe Lys Asp Asn Glu Phe Asp Lys Thr Gly Gly Val Leu Arg
1040 1045 1050
Ala Tyr Gln Leu Thr Ala Pro Phe Glu Thr Phe Lys Lys Met Gly
1055 1060 1065
Lys Gln Thr Gly Ile Ile Tyr Tyr Val Pro Ala Gly Phe Thr Ser
1070 1075 1080
Lys Ile Cys Pro Val Thr Gly Phe Val Asn Gln Leu Tyr Pro Lys
1085 1090 1095
Tyr Glu Ser Val Ser Lys Ser Gln Glu Phe Phe Ser Lys Phe Asp
1100 1105 1110
Lys Ile Cys Tyr Asn Leu Asp Lys Gly Tyr Phe Glu Phe Ser Phe
1115 1120 1125
Asp Tyr Lys Asn Phe Gly Asp Lys Ala Ala Lys Gly Lys Trp Thr
1130 1135 1140
Ile Ala Ser Phe Gly Ser Arg Leu Ile Asn Phe Arg Asn Ser Asp
1145 1150 1155
Lys Asn His Asn Trp Asp Thr Arg Glu Val Tyr Pro Thr Lys Glu
1160 1165 1170
Leu Glu Lys Leu Leu Lys Asp Tyr Ser Ile Glu Tyr Gly His Gly
1175 1180 1185
Glu Cys Ile Lys Ala Ala Ile Cys Gly Glu Ser Asp Lys Lys Phe
1190 1195 1200
Phe Ala Lys Leu Thr Ser Val Leu Asn Thr Ile Leu Gln Met Arg
1205 1210 1215
Asn Ser Lys Thr Gly Thr Glu Leu Asp Tyr Leu Ile Ser Pro Val
1220 1225 1230
Ala Asp Val Asn Gly Asn Phe Phe Asp Ser Arg Gln Ala Pro Lys
1235 1240 1245

Asn Met Pro Gln Asp Ala Asp Ala Asn Gly Ala Tyr His Ile Gly
1250 1255 1260
Leu Lys Gly Leu Met Leu Leu Gly Arg Ile Lys Asn Asn Gln Glu
1265 1270 1275
Gly Lys Lys Leu Asn Leu Val Ile Lys Asn Glu Glu Tyr Phe Glu
1280 1285 1290
Phe Val Gln Asn Arg Asn Asn
1295 1300

<210> 33
<211> 1300
<212> PRT
<213> 人工序列
<220>
<223> 合成多肽
<400> 33

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

Asn Asp Ile Pro Thr Ser Ile Ile Tyr Arg Ile Val Asp Asp Asn Leu
195 200 205
Pro Lys Phe Leu Glu Asn Lys Ala Lys Tyr Glu Ser Leu Lys Asp Lys
210 215 220
Ala Pro Glu Ala Ile Asn Tyr Glu Gln Ile Lys Lys Asp Leu Ala Glu
225 230 235 240
Glu Leu Thr Phe Asp Ile Asp Tyr Lys Thr Ser Glu Val Asn Gln Arg
245 250 255
Val Phe Ser Leu Asp Glu Val Phe Glu Ile Ala Asn Phe Asn Asn Tyr
260 265 270
Leu Asn Gln Ser Gly Ile Thr Lys Phe Asn Thr Ile Ile Gly Gly Lys
275 280 285
Phe Val Asn Gly Glu Asn Thr Lys Arg Lys Gly Ile Asn Glu Tyr Ile
290 295 300
Asn Leu Tyr Ser Gln Gln Ile Asn Asp Lys Thr Leu Lys Lys Tyr Lys
305 310 315 320
Met Ser Val Leu Phe Lys Gln Ile Leu Ser Asp Thr Glu Ser Lys Ser
325 330 335
Phe Val Ile Asp Lys Leu Glu Asp Asp Ser Asp Val Val Thr Thr Met
340 345 350
Gln Ser Phe Tyr Glu Gln Ile Ala Ala Phe Lys Thr Val Glu Glu Lys
355 360 365
Ser Ile Lys Glu Thr Leu Ser Leu Leu Phe Asp Asp Leu Lys Ala Gln
370 375 380
Lys Leu Asp Leu Ser Lys Ile Tyr Phe Lys Asn Asp Lys Ser Leu Thr
385 390 395 400
Asp Leu Ser Gln Gln Val Phe Asp Asp Tyr Ser Val Ile Gly Thr Ala
405 410 415
Val Leu Glu Tyr Ile Thr Gln Gln Ile Ala Pro Lys Asn Leu Asp Asn
420 425 430
Pro Ser Lys Lys Glu Gln Glu Leu Ile Ala Lys Lys Thr Glu Lys Ala
435 440 445
Lys Tyr Leu Ser Leu Glu Thr Ile Lys Leu Ala Leu Glu Glu Phe Asn
450 455 460
Lys His Arg Asp Ile Asp Lys Gln Cys Arg Phe Glu Glu Ile Leu Ala
465 470 475 480
Asn Phe Ala Ala Ile Pro Met Ile Phe Asp Glu Ile Ala Gln Asn Lys
485 490 495
Asp Asn Leu Ala Gln Ile Ser Ile Lys Tyr Gln Asn Gln Gly Lys Lys

500 505 510
Asp Leu Leu Gln Ala Ser Ala Glu Asp Asp Val Lys Ala Ile Lys Asp
515 520 525
Leu Leu Asp Gln Thr Asn Asn Leu Leu His Lys Leu Lys Ile Phe His
530 535 540
Ile Ser Gln Ser Glu Asp Lys Ala Asn Ile Leu Asp Lys Asp Glu His
545 550 555 560
Phe Tyr Leu Val Phe Glu Glu Cys Tyr Phe Glu Leu Ala Asn Ile Val
565 570 575
Pro Leu Tyr Asn Lys Ile Arg Asn Tyr Ile Thr Gln Lys Pro Tyr Ser
580 585 590
Asp Glu Lys Phe Lys Leu Asn Phe Glu Asn Ser Thr Leu Ala Asn Gly
595 600 605
Trp Asp Lys Asn Lys Glu Pro Asp Asn Thr Ala Ile Leu Phe Ile Lys
610 615 620
Asp Asp Lys Tyr Tyr Leu Gly Val Met Asn Lys Lys Asn Asn Lys Ile
625 630 635 640
Phe Asp Asp Lys Ala Ile Lys Glu Asn Lys Gly Glu Gly Tyr Lys Lys
645 650 655
Ile Val Tyr Lys Leu Leu Pro Gly Ala Asn Lys Met Leu Pro Lys Val
660 665 670
Phe Phe Ser Ala Lys Ser Ile Lys Phe Tyr Asn Pro Ser Glu Asp Ile
675 680 685
Leu Arg Ile Arg Asn His Ser Thr His Thr Lys Asn Gly Ser Pro Gln
690 695 700
Lys Gly Tyr Glu Lys Phe Glu Phe Asn Ile Glu Asp Cys Arg Lys Phe
705 710 715 720
Ile Asp Phe Tyr Lys Gln Ser Ile Ser Lys His Pro Glu Trp Lys Asp
725 730 735
Phe Gly Phe Arg Phe Ser Asp Thr Gln Arg Tyr Asn Ser Ile Asp Glu
740 745 750
Phe Tyr Arg Glu Val Glu Asn Gln Gly Tyr Lys Leu Thr Phe Glu Asn
755 760 765
Ile Ser Glu Ser Tyr Ile Asp Ser Val Val Asn Gln Gly Lys Leu Tyr
770 775 780
Leu Phe Gln Ile Tyr Asn Lys Asp Phe Ser Ala Tyr Ser Lys Gly Arg
785 790 795 800
Pro Asn Leu His Thr Leu Tyr Trp Lys Ala Leu Phe Asp Glu Arg Asn
805 810 815

Leu Gln Asp Val Val Tyr Lys Leu Asn Gly Glu Ala Glu Leu Phe Tyr
820 825 830
Arg Lys Gln Ser Ile Pro Lys Lys Ile Thr His Pro Ala Lys Glu Ala
835 840 845
Ile Ala Asn Lys Asn Lys Asp Asn Pro Lys Lys Glu Ser Val Phe Glu
850 855 860
Tyr Asp Leu Ile Lys Asp Lys Arg Phe Thr Glu Asp Lys Phe Phe Phe
865 870 875 880
His Cys Pro Ile Thr Ile Asn Phe Lys Ser Ser Gly Ala Asn Lys Phe
885 890 895
Asn Asp Glu Ile Asn Leu Leu Leu Lys Glu Lys Ala Asn Asp Val His
900 905 910
Ile Leu Ser Ile Asp Arg Gly Glu Arg His Leu Ala Tyr Tyr Thr Leu
915 920 925
Val Asp Gly Lys Gly Asn Ile Ile Lys Gln Asp Thr Phe Asn Ile Ile
930 935 940
Gly Asn Asp Arg Met Lys Thr Asn Tyr His Asp Lys Leu Ala Ala Ile
945 950 955 960
Glu Lys Asp Arg Asp Ser Ala Arg Lys Asp Trp Lys Lys Ile Asn Asn
965 970 975
Ile Lys Glu Met Lys Glu Gly Tyr Leu Ser Gln Val Val His Glu Ile
980 985 990
Ala Lys Leu Val Ile Glu Tyr Asn Ala Ile Val Val Phe Glu Asp Leu
995 1000 1005
Asn Phe Gly Phe Lys Arg Gly Arg Phe Lys Val Glu Lys Gln Val
1010 1015 1020
Tyr Gln Lys Leu Glu Lys Met Leu Ile Glu Lys Leu Asn Tyr Leu
1025 1030 1035
Val Phe Lys Asp Asn Glu Phe Asp Lys Thr Gly Gly Val Leu Arg
1040 1045 1050
Ala Tyr Gln Leu Thr Ala Pro Phe Glu Thr Phe Lys Lys Met Gly
1055 1060 1065
Lys Gln Thr Gly Ile Ile Tyr Tyr Val Pro Ala Gly Phe Thr Ser
1070 1075 1080
Lys Ile Cys Pro Val Thr Gly Phe Val Asn Gln Leu Tyr Pro Lys
1085 1090 1095
Tyr Glu Ser Val Ser Lys Ser Gln Glu Phe Phe Ser Lys Phe Asp
1100 1105 1110
Lys Ile Cys Tyr Asn Leu Asp Lys Gly Tyr Phe Glu Phe Ser Phe

1115 1120 1125
Asp Tyr Lys Asn Phe Gly Asp Lys Ala Ala Lys Gly Lys Trp Thr
1130 1135 1140
Ile Ala Ser Phe Gly Ser Arg Leu Ile Asn Phe Arg Asn Ser Asp
1145 1150 1155
Lys Asn His Asn Trp Asp Thr Arg Glu Val Tyr Pro Thr Lys Glu
1160 1165 1170
Leu Glu Lys Leu Leu Lys Asp Tyr Ser Ile Glu Tyr Gly His Gly
1175 1180 1185
Glu Cys Ile Lys Ala Ala Ile Cys Gly Glu Ser Asp Lys Lys Phe
1190 1195 1200
Phe Ala Lys Leu Thr Ser Val Leu Asn Thr Ile Leu Gln Met Arg
1205 1210 1215
Asn Ser Lys Thr Gly Thr Glu Leu Asp Tyr Leu Ile Ser Pro Val
1220 1225 1230
Ala Asp Val Asn Gly Asn Phe Phe Asp Ser Arg Gln Ala Pro Lys
1235 1240 1245
Asn Met Pro Gln Asp Ala Ala Ala Asn Gly Ala Tyr His Ile Gly
1250 1255 1260
Leu Lys Gly Leu Met Leu Leu Gly Arg Ile Lys Asn Asn Gln Glu
1265 1270 1275
Gly Lys Lys Leu Asn Leu Val Ile Lys Asn Glu Glu Tyr Phe Glu
1280 1285 1290
Phe Val Gln Asn Arg Asn Asn
1295 1300

<210> 34
<211> 1300
<212> PRT
<213> 人工序列
<220>
<223> 合成多肽
<400> 34

Met Ser Ile Tyr Gln Glu Phe Val Asn Lys Tyr Ser Leu Ser Lys Thr
1 5 10 15
Leu Arg Phe Glu Leu Ile Pro Gln Gly Lys Thr Leu Glu Asn Ile Lys
20 25 30
Ala Arg Gly Leu Ile Leu Asp Asp Glu Lys Arg Ala Lys Asp Tyr Lys
35 40 45
Lys Ala Lys Gln Ile Ile Asp Lys Tyr His Gln Phe Phe Ile Glu Glu

50 55 60

Ile Leu Ser Ser Val Cys Ile Ser Glu Asp Leu Leu Gln Asn Tyr Ser
65 70 75 80

Asp Val Tyr Phe Lys Leu Lys Lys Ser Asp Asp Asp Asn Leu Gln Lys
85 90 95

Asp Phe Lys Ser Ala Lys Asp Thr Ile Lys Lys Gln Ile Ser Glu Tyr
100 105 110

Ile Lys Asp Ser Glu Lys Phe Lys Asn Leu Phe Asn Gln Asn Leu Ile
115 120 125

Asp Ala Lys Lys Gly Gln Glu Ser Asp Leu Ile Leu Trp Leu Lys Gln
130 135 140

Ser Lys Asp Asn Gly Ile Glu Leu Phe Lys Ala Asn Ser Asp Ile Thr
145 150 155 160

Asp Ile Asp Glu Ala Leu Glu Ile Ile Lys Ser Phe Lys Gly Trp Thr
165 170 175

Thr Tyr Phe Lys Gly Phe His Glu Asn Arg Lys Asn Val Tyr Ser Ser
180 185 190

Asn Asp Ile Pro Thr Ser Ile Ile Tyr Arg Ile Val Asp Asp Asn Leu
195 200 205

Pro Lys Phe Leu Glu Asn Lys Ala Lys Tyr Glu Ser Leu Lys Asp Lys
210 215 220

Ala Pro Glu Ala Ile Asn Tyr Glu Gln Ile Lys Lys Asp Leu Ala Glu
225 230 235 240

Glu Leu Thr Phe Asp Ile Asp Tyr Lys Thr Ser Glu Val Asn Gln Arg
245 250 255

Val Phe Ser Leu Asp Glu Val Phe Glu Ile Ala Asn Phe Asn Asn Tyr
260 265 270

Leu Asn Gln Ser Gly Ile Thr Lys Phe Asn Thr Ile Ile Gly Gly Lys
275 280 285

Phe Val Asn Gly Glu Asn Thr Lys Arg Lys Gly Ile Asn Glu Tyr Ile
290 295 300

Asn Leu Tyr Ser Gln Gln Ile Asn Asp Lys Thr Leu Lys Lys Tyr Lys
305 310 315 320

Met Ser Val Leu Phe Lys Gln Ile Leu Ser Asp Thr Glu Ser Lys Ser
325 330 335

Phe Val Ile Asp Lys Leu Glu Asp Asp Ser Asp Val Val Thr Thr Met
340 345 350

Gln Ser Phe Tyr Glu Gln Ile Ala Ala Phe Lys Thr Val Glu Glu Lys
355 360 365

Ser Ile Lys Glu Thr Leu Ser Leu Leu Phe Asp Asp Leu Lys Ala Gln
370 375 380
Lys Leu Asp Leu Ser Lys Ile Tyr Phe Lys Asn Asp Lys Ser Leu Thr
385 390 395 400
Asp Leu Ser Gln Gln Val Phe Asp Asp Tyr Ser Val Ile Gly Thr Ala
405 410 415
Val Leu Glu Tyr Ile Thr Gln Gln Ile Ala Pro Lys Asn Leu Asp Asn
420 425 430
Pro Ser Lys Lys Glu Gln Glu Leu Ile Ala Lys Lys Thr Glu Lys Ala
435 440 445
Lys Tyr Leu Ser Leu Glu Thr Ile Lys Leu Ala Leu Glu Glu Phe Asn
450 455 460
Lys His Arg Asp Ile Asp Lys Gln Cys Arg Phe Glu Glu Ile Leu Ala
465 470 475 480
Asn Phe Ala Ala Ile Pro Met Ile Phe Asp Glu Ile Ala Gln Asn Lys
485 490 495
Asp Asn Leu Ala Gln Ile Ser Ile Lys Tyr Gln Asn Gln Gly Lys Lys
500 505 510
Asp Leu Leu Gln Ala Ser Ala Glu Asp Asp Val Lys Ala Ile Lys Asp
515 520 525
Leu Leu Asp Gln Thr Asn Asn Leu Leu His Lys Leu Lys Ile Phe His
530 535 540
Ile Ser Gln Ser Glu Asp Lys Ala Asn Ile Leu Asp Lys Asp Glu His
545 550 555 560
Phe Tyr Leu Val Phe Glu Glu Cys Tyr Phe Glu Leu Ala Asn Ile Val
565 570 575
Pro Leu Tyr Asn Lys Ile Arg Asn Tyr Ile Thr Gln Lys Pro Tyr Ser
580 585 590
Asp Glu Lys Phe Lys Leu Asn Phe Glu Asn Ser Thr Leu Ala Asn Gly
595 600 605
Trp Asp Lys Asn Lys Glu Pro Asp Asn Thr Ala Ile Leu Phe Ile Lys
610 615 620
Asp Asp Lys Tyr Tyr Leu Gly Val Met Asn Lys Lys Asn Asn Lys Ile
625 630 635 640
Phe Asp Asp Lys Ala Ile Lys Glu Asn Lys Gly Glu Gly Tyr Lys Lys
645 650 655
Ile Val Tyr Lys Leu Leu Pro Gly Ala Asn Lys Met Leu Pro Lys Val
660 665 670
Phe Phe Ser Ala Lys Ser Ile Lys Phe Tyr Asn Pro Ser Glu Asp Ile

675 680 685

Leu Arg Ile Arg Asn His Ser Thr His Thr Lys Asn Gly Ser Pro Gln

690 695 700

Lys Gly Tyr Glu Lys Phe Glu Phe Asn Ile Glu Asp Cys Arg Lys Phe

705 710 715 720

Ile Asp Phe Tyr Lys Gln Ser Ile Ser Lys His Pro Glu Trp Lys Asp

725 730 735

Phe Gly Phe Arg Phe Ser Asp Thr Gln Arg Tyr Asn Ser Ile Asp Glu

740 745 750

Phe Tyr Arg Glu Val Glu Asn Gln Gly Tyr Lys Leu Thr Phe Glu Asn

755 760 765

Ile Ser Glu Ser Tyr Ile Asp Ser Val Val Asn Gln Gly Lys Leu Tyr

770 775 780

Leu Phe Gln Ile Tyr Asn Lys Asp Phe Ser Ala Tyr Ser Lys Gly Arg

785 790 795 800

Pro Asn Leu His Thr Leu Tyr Trp Lys Ala Leu Phe Asp Glu Arg Asn

805 810 815

Leu Gln Asp Val Val Tyr Lys Leu Asn Gly Glu Ala Glu Leu Phe Tyr

820 825 830

Arg Lys Gln Ser Ile Pro Lys Lys Ile Thr His Pro Ala Lys Glu Ala

835 840 845

Ile Ala Asn Lys Asn Lys Asp Asn Pro Lys Lys Glu Ser Val Phe Glu

850 855 860

Tyr Asp Leu Ile Lys Asp Lys Arg Phe Thr Glu Asp Lys Phe Phe Phe

865 870 875 880

His Cys Pro Ile Thr Ile Asn Phe Lys Ser Ser Gly Ala Asn Lys Phe

885 890 895

Asn Asp Glu Ile Asn Leu Leu Leu Lys Glu Lys Ala Asn Asp Val His

900 905 910

Ile Leu Ser Ile Ala Arg Gly Glu Arg His Leu Ala Tyr Tyr Thr Leu

915 920 925

Val Asp Gly Lys Gly Asn Ile Ile Lys Gln Asp Thr Phe Asn Ile Ile

930 935 940

Gly Asn Asp Arg Met Lys Thr Asn Tyr His Asp Lys Leu Ala Ala Ile

945 950 955 960

Glu Lys Asp Arg Asp Ser Ala Arg Lys Asp Trp Lys Lys Ile Asn Asn

965 970 975

Ile Lys Glu Met Lys Glu Gly Tyr Leu Ser Gln Val Val His Glu Ile

980 985 990

Ala Lys Leu Val Ile Glu Tyr Asn Ala Ile Val Val Phe Ala Asp Leu
995 1000 1005
Asn Phe Gly Phe Lys Arg Gly Arg Phe Lys Val Glu Lys Gln Val
1010 1015 1020
Tyr Gln Lys Leu Glu Lys Met Leu Ile Glu Lys Leu Asn Tyr Leu
1025 1030 1035
Val Phe Lys Asp Asn Glu Phe Asp Lys Thr Gly Gly Val Leu Arg
1040 1045 1050
Ala Tyr Gln Leu Thr Ala Pro Phe Glu Thr Phe Lys Lys Met Gly
1055 1060 1065
Lys Gln Thr Gly Ile Ile Tyr Tyr Val Pro Ala Gly Phe Thr Ser
1070 1075 1080
Lys Ile Cys Pro Val Thr Gly Phe Val Asn Gln Leu Tyr Pro Lys
1085 1090 1095
Tyr Glu Ser Val Ser Lys Ser Gln Glu Phe Phe Ser Lys Phe Asp
1100 1105 1110
Lys Ile Cys Tyr Asn Leu Asp Lys Gly Tyr Phe Glu Phe Ser Phe
1115 1120 1125
Asp Tyr Lys Asn Phe Gly Asp Lys Ala Ala Lys Gly Lys Trp Thr
1130 1135 1140
Ile Ala Ser Phe Gly Ser Arg Leu Ile Asn Phe Arg Asn Ser Asp
1145 1150 1155
Lys Asn His Asn Trp Asp Thr Arg Glu Val Tyr Pro Thr Lys Glu
1160 1165 1170
Leu Glu Lys Leu Leu Lys Asp Tyr Ser Ile Glu Tyr Gly His Gly
1175 1180 1185
Glu Cys Ile Lys Ala Ala Ile Cys Gly Glu Ser Asp Lys Lys Phe
1190 1195 1200
Phe Ala Lys Leu Thr Ser Val Leu Asn Thr Ile Leu Gln Met Arg
1205 1210 1215
Asn Ser Lys Thr Gly Thr Glu Leu Asp Tyr Leu Ile Ser Pro Val
1220 1225 1230
Ala Asp Val Asn Gly Asn Phe Phe Asp Ser Arg Gln Ala Pro Lys
1235 1240 1245
Asn Met Pro Gln Asp Ala Asp Ala Asn Gly Ala Tyr His Ile Gly
1250 1255 1260
Leu Lys Gly Leu Met Leu Leu Gly Arg Ile Lys Asn Asn Gln Glu
1265 1270 1275
Gly Lys Lys Leu Asn Leu Val Ile Lys Asn Glu Glu Tyr Phe Glu

1280 1285 1290

Phe Val Gln Asn Arg Asn Asn

1295 1300

<210> 35

<211> 1300

<212> PRT

<213> 人工序列

<220>

<223> 合成多肽

<400> 35

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

225 230 235 240
Glu Leu Thr Phe Asp Ile Asp Tyr Lys Thr Ser Glu Val Asn Gln Arg
245 250 255
Val Phe Ser Leu Asp Glu Val Phe Glu Ile Ala Asn Phe Asn Asn Tyr
260 265 270
Leu Asn Gln Ser Gly Ile Thr Lys Phe Asn Thr Ile Ile Gly Gly Lys
275 280 285
Phe Val Asn Gly Glu Asn Thr Lys Arg Lys Gly Ile Asn Glu Tyr Ile
290 295 300
Asn Leu Tyr Ser Gln Gln Ile Asn Asp Lys Thr Leu Lys Lys Tyr Lys
305 310 315 320
Met Ser Val Leu Phe Lys Gln Ile Leu Ser Asp Thr Glu Ser Lys Ser
325 330 335
Phe Val Ile Asp Lys Leu Glu Asp Asp Ser Asp Val Val Thr Thr Met
340 345 350
Gln Ser Phe Tyr Glu Gln Ile Ala Ala Phe Lys Thr Val Glu Glu Lys
355 360 365
Ser Ile Lys Glu Thr Leu Ser Leu Leu Phe Asp Asp Leu Lys Ala Gln
370 375 380
Lys Leu Asp Leu Ser Lys Ile Tyr Phe Lys Asn Asp Lys Ser Leu Thr
385 390 395 400
Asp Leu Ser Gln Gln Val Phe Asp Asp Tyr Ser Val Ile Gly Thr Ala
405 410 415
Val Leu Glu Tyr Ile Thr Gln Gln Ile Ala Pro Lys Asn Leu Asp Asn
420 425 430
Pro Ser Lys Lys Glu Gln Glu Leu Ile Ala Lys Lys Thr Glu Lys Ala
435 440 445
Lys Tyr Leu Ser Leu Glu Thr Ile Lys Leu Ala Leu Glu Glu Phe Asn
450 455 460
Lys His Arg Asp Ile Asp Lys Gln Cys Arg Phe Glu Glu Ile Leu Ala
465 470 475 480
Asn Phe Ala Ala Ile Pro Met Ile Phe Asp Glu Ile Ala Gln Asn Lys
485 490 495
Asp Asn Leu Ala Gln Ile Ser Ile Lys Tyr Gln Asn Gln Gly Lys Lys
500 505 510
Asp Leu Leu Gln Ala Ser Ala Glu Asp Asp Val Lys Ala Ile Lys Asp
515 520 525
Leu Leu Asp Gln Thr Asn Asn Leu Leu His Lys Leu Lys Ile Phe His
530 535 540

Ile Ser Gln Ser Glu Asp Lys Ala Asn Ile Leu Asp Lys Asp Glu His
545 550 555 560
Phe Tyr Leu Val Phe Glu Glu Cys Tyr Phe Glu Leu Ala Asn Ile Val
565 570 575
Pro Leu Tyr Asn Lys Ile Arg Asn Tyr Ile Thr Gln Lys Pro Tyr Ser
580 585 590
Asp Glu Lys Phe Lys Leu Asn Phe Glu Asn Ser Thr Leu Ala Asn Gly
595 600 605
Trp Asp Lys Asn Lys Glu Pro Asp Asn Thr Ala Ile Leu Phe Ile Lys
610 615 620
Asp Asp Lys Tyr Tyr Leu Gly Val Met Asn Lys Lys Asn Asn Lys Ile
625 630 635 640
Phe Asp Asp Lys Ala Ile Lys Glu Asn Lys Gly Glu Gly Tyr Lys Lys
645 650 655
Ile Val Tyr Lys Leu Leu Pro Gly Ala Asn Lys Met Leu Pro Lys Val
660 665 670
Phe Phe Ser Ala Lys Ser Ile Lys Phe Tyr Asn Pro Ser Glu Asp Ile
675 680 685
Leu Arg Ile Arg Asn His Ser Thr His Thr Lys Asn Gly Ser Pro Gln
690 695 700
Lys Gly Tyr Glu Lys Phe Glu Phe Asn Ile Glu Asp Cys Arg Lys Phe
705 710 715 720
Ile Asp Phe Tyr Lys Gln Ser Ile Ser Lys His Pro Glu Trp Lys Asp
725 730 735
Phe Gly Phe Arg Phe Ser Asp Thr Gln Arg Tyr Asn Ser Ile Asp Glu
740 745 750
Phe Tyr Arg Glu Val Glu Asn Gln Gly Tyr Lys Leu Thr Phe Glu Asn
755 760 765
Ile Ser Glu Ser Tyr Ile Asp Ser Val Val Asn Gln Gly Lys Leu Tyr
770 775 780
Leu Phe Gln Ile Tyr Asn Lys Asp Phe Ser Ala Tyr Ser Lys Gly Arg
785 790 795 800
Pro Asn Leu His Thr Leu Tyr Trp Lys Ala Leu Phe Asp Glu Arg Asn
805 810 815
Leu Gln Asp Val Val Tyr Lys Leu Asn Gly Glu Ala Glu Leu Phe Tyr
820 825 830
Arg Lys Gln Ser Ile Pro Lys Lys Ile Thr His Pro Ala Lys Glu Ala
835 840 845
Ile Ala Asn Lys Asn Lys Asp Asn Pro Lys Lys Glu Ser Val Phe Glu

850 855 860

Tyr Asp Leu Ile Lys Asp Lys Arg Phe Thr Glu Asp Lys Phe Phe Phe
865 870 875 880

His Cys Pro Ile Thr Ile Asn Phe Lys Ser Ser Gly Ala Asn Lys Phe
885 890 895

Asn Asp Glu Ile Asn Leu Leu Leu Lys Glu Lys Ala Asn Asp Val His
900 905 910

Ile Leu Ser Ile Ala Arg Gly Glu Arg His Leu Ala Tyr Tyr Thr Leu
915 920 925

Val Asp Gly Lys Gly Asn Ile Ile Lys Gln Asp Thr Phe Asn Ile Ile
930 935 940

Gly Asn Asp Arg Met Lys Thr Asn Tyr His Asp Lys Leu Ala Ala Ile
945 950 955 960

Glu Lys Asp Arg Asp Ser Ala Arg Lys Asp Trp Lys Lys Ile Asn Asn
965 970 975

Ile Lys Glu Met Lys Glu Gly Tyr Leu Ser Gln Val Val His Glu Ile
980 985 990

Ala Lys Leu Val Ile Glu Tyr Asn Ala Ile Val Val Phe Glu Asp Leu
995 1000 1005

Asn Phe Gly Phe Lys Arg Gly Arg Phe Lys Val Glu Lys Gln Val
1010 1015 1020

Tyr Gln Lys Leu Glu Lys Met Leu Ile Glu Lys Leu Asn Tyr Leu
1025 1030 1035

Val Phe Lys Asp Asn Glu Phe Asp Lys Thr Gly Gly Val Leu Arg
1040 1045 1050

Ala Tyr Gln Leu Thr Ala Pro Phe Glu Thr Phe Lys Lys Met Gly
1055 1060 1065

Lys Gln Thr Gly Ile Ile Tyr Tyr Val Pro Ala Gly Phe Thr Ser
1070 1075 1080

Lys Ile Cys Pro Val Thr Gly Phe Val Asn Gln Leu Tyr Pro Lys
1085 1090 1095

Tyr Glu Ser Val Ser Lys Ser Gln Glu Phe Phe Ser Lys Phe Asp
1100 1105 1110

Lys Ile Cys Tyr Asn Leu Asp Lys Gly Tyr Phe Glu Phe Ser Phe
1115 1120 1125

Asp Tyr Lys Asn Phe Gly Asp Lys Ala Ala Lys Gly Lys Trp Thr
1130 1135 1140

Ile Ala Ser Phe Gly Ser Arg Leu Ile Asn Phe Arg Asn Ser Asp
1145 1150 1155

Lys Asn His Asn Trp Asp Thr Arg Glu Val Tyr Pro Thr Lys Glu
1160 1165 1170
Leu Glu Lys Leu Leu Lys Asp Tyr Ser Ile Glu Tyr Gly His Gly
1175 1180 1185
Glu Cys Ile Lys Ala Ala Ile Cys Gly Glu Ser Asp Lys Lys Phe
1190 1195 1200
Phe Ala Lys Leu Thr Ser Val Leu Asn Thr Ile Leu Gln Met Arg
1205 1210 1215
Asn Ser Lys Thr Gly Thr Glu Leu Asp Tyr Leu Ile Ser Pro Val
1220 1225 1230
Ala Asp Val Asn Gly Asn Phe Phe Asp Ser Arg Gln Ala Pro Lys
1235 1240 1245
Asn Met Pro Gln Asp Ala Ala Ala Asn Gly Ala Tyr His Ile Gly
1250 1255 1260
Leu Lys Gly Leu Met Leu Leu Gly Arg Ile Lys Asn Asn Gln Glu
1265 1270 1275
Gly Lys Lys Leu Asn Leu Val Ile Lys Asn Glu Glu Tyr Phe Glu
1280 1285 1290
Phe Val Gln Asn Arg Asn Asn
1295 1300

<210> 36
<211> 1300
<212> PRT
<213> 人工序列
<220>
<223> 合成多肽
<400> 36

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

Asp Phe Lys Ser Ala Lys Asp Thr Ile Lys Lys Gln Ile Ser Glu Tyr
100 105 110
Ile Lys Asp Ser Glu Lys Phe Lys Asn Leu Phe Asn Gln Asn Leu Ile
115 120 125
Asp Ala Lys Lys Gly Gln Glu Ser Asp Leu Ile Leu Trp Leu Lys Gln
130 135 140
Ser Lys Asp Asn Gly Ile Glu Leu Phe Lys Ala Asn Ser Asp Ile Thr
145 150 155 160
Asp Ile Asp Glu Ala Leu Glu Ile Ile Lys Ser Phe Lys Gly Trp Thr
165 170 175
Thr Tyr Phe Lys Gly Phe His Glu Asn Arg Lys Asn Val Tyr Ser Ser
180 185 190
Asn Asp Ile Pro Thr Ser Ile Ile Tyr Arg Ile Val Asp Asp Asn Leu
195 200 205
Pro Lys Phe Leu Glu Asn Lys Ala Lys Tyr Glu Ser Leu Lys Asp Lys
210 215 220
Ala Pro Glu Ala Ile Asn Tyr Glu Gln Ile Lys Lys Asp Leu Ala Glu
225 230 235 240
Glu Leu Thr Phe Asp Ile Asp Tyr Lys Thr Ser Glu Val Asn Gln Arg
245 250 255
Val Phe Ser Leu Asp Glu Val Phe Glu Ile Ala Asn Phe Asn Asn Tyr
260 265 270
Leu Asn Gln Ser Gly Ile Thr Lys Phe Asn Thr Ile Ile Gly Gly Lys
275 280 285
Phe Val Asn Gly Glu Asn Thr Lys Arg Lys Gly Ile Asn Glu Tyr Ile
290 295 300
Asn Leu Tyr Ser Gln Gln Ile Asn Asp Lys Thr Leu Lys Lys Tyr Lys
305 310 315 320
Met Ser Val Leu Phe Lys Gln Ile Leu Ser Asp Thr Glu Ser Lys Ser
325 330 335
Phe Val Ile Asp Lys Leu Glu Asp Asp Ser Asp Val Val Thr Thr Met
340 345 350
Gln Ser Phe Tyr Glu Gln Ile Ala Ala Phe Lys Thr Val Glu Glu Lys
355 360 365
Ser Ile Lys Glu Thr Leu Ser Leu Leu Phe Asp Asp Leu Lys Ala Gln
370 375 380
Lys Leu Asp Leu Ser Lys Ile Tyr Phe Lys Asn Asp Lys Ser Leu Thr
385 390 395 400
Asp Leu Ser Gln Gln Val Phe Asp Asp Tyr Ser Val Ile Gly Thr Ala

405 410 415

Val Leu Glu Tyr Ile Thr Gln Gln Ile Ala Pro Lys Asn Leu Asp Asn

420 425 430

Pro Ser Lys Lys Glu Gln Glu Leu Ile Ala Lys Lys Thr Glu Lys Ala

435 440 445

Lys Tyr Leu Ser Leu Glu Thr Ile Lys Leu Ala Leu Glu Glu Phe Asn

450 455 460

Lys His Arg Asp Ile Asp Lys Gln Cys Arg Phe Glu Glu Ile Leu Ala

465 470 475 480

Asn Phe Ala Ala Ile Pro Met Ile Phe Asp Glu Ile Ala Gln Asn Lys

485 490 495

Asp Asn Leu Ala Gln Ile Ser Ile Lys Tyr Gln Asn Gln Gly Lys Lys

500 505 510

Asp Leu Leu Gln Ala Ser Ala Glu Asp Asp Val Lys Ala Ile Lys Asp

515 520 525

Leu Leu Asp Gln Thr Asn Asn Leu Leu His Lys Leu Lys Ile Phe His

530 535 540

Ile Ser Gln Ser Glu Asp Lys Ala Asn Ile Leu Asp Lys Asp Glu His

545 550 555 560

Phe Tyr Leu Val Phe Glu Glu Cys Tyr Phe Glu Leu Ala Asn Ile Val

565 570 575

Pro Leu Tyr Asn Lys Ile Arg Asn Tyr Ile Thr Gln Lys Pro Tyr Ser

580 585 590

Asp Glu Lys Phe Lys Leu Asn Phe Glu Asn Ser Thr Leu Ala Asn Gly

595 600 605

Trp Asp Lys Asn Lys Glu Pro Asp Asn Thr Ala Ile Leu Phe Ile Lys

610 615 620

Asp Asp Lys Tyr Tyr Leu Gly Val Met Asn Lys Lys Asn Asn Lys Ile

625 630 635 640

Phe Asp Asp Lys Ala Ile Lys Glu Asn Lys Gly Glu Gly Tyr Lys Lys

645 650 655

Ile Val Tyr Lys Leu Leu Pro Gly Ala Asn Lys Met Leu Pro Lys Val

660 665 670

Phe Phe Ser Ala Lys Ser Ile Lys Phe Tyr Asn Pro Ser Glu Asp Ile

675 680 685

Leu Arg Ile Arg Asn His Ser Thr His Thr Lys Asn Gly Ser Pro Gln

690 695 700

Lys Gly Tyr Glu Lys Phe Glu Phe Asn Ile Glu Asp Cys Arg Lys Phe

705 710 715 720

Ile Asp Phe Tyr Lys Gln Ser Ile Ser Lys His Pro Glu Trp Lys Asp
725 730 735
Phe Gly Phe Arg Phe Ser Asp Thr Gln Arg Tyr Asn Ser Ile Asp Glu
740 745 750
Phe Tyr Arg Glu Val Glu Asn Gln Gly Tyr Lys Leu Thr Phe Glu Asn
755 760 765
Ile Ser Glu Ser Tyr Ile Asp Ser Val Val Asn Gln Gly Lys Leu Tyr
770 775 780
Leu Phe Gln Ile Tyr Asn Lys Asp Phe Ser Ala Tyr Ser Lys Gly Arg
785 790 795 800
Pro Asn Leu His Thr Leu Tyr Trp Lys Ala Leu Phe Asp Glu Arg Asn
805 810 815
Leu Gln Asp Val Val Tyr Lys Leu Asn Gly Glu Ala Glu Leu Phe Tyr
820 825 830
Arg Lys Gln Ser Ile Pro Lys Lys Ile Thr His Pro Ala Lys Glu Ala
835 840 845
Ile Ala Asn Lys Asn Lys Asp Asn Pro Lys Lys Glu Ser Val Phe Glu
850 855 860
Tyr Asp Leu Ile Lys Asp Lys Arg Phe Thr Glu Asp Lys Phe Phe Phe
865 870 875 880
His Cys Pro Ile Thr Ile Asn Phe Lys Ser Ser Gly Ala Asn Lys Phe
885 890 895
Asn Asp Glu Ile Asn Leu Leu Leu Lys Glu Lys Ala Asn Asp Val His
900 905 910
Ile Leu Ser Ile Asp Arg Gly Glu Arg His Leu Ala Tyr Tyr Thr Leu
915 920 925
Val Asp Gly Lys Gly Asn Ile Ile Lys Gln Asp Thr Phe Asn Ile Ile
930 935 940
Gly Asn Asp Arg Met Lys Thr Asn Tyr His Asp Lys Leu Ala Ala Ile
945 950 955 960
Glu Lys Asp Arg Asp Ser Ala Arg Lys Asp Trp Lys Lys Ile Asn Asn
965 970 975
Ile Lys Glu Met Lys Glu Gly Tyr Leu Ser Gln Val Val His Glu Ile
980 985 990
Ala Lys Leu Val Ile Glu Tyr Asn Ala Ile Val Val Phe Ala Asp Leu
995 1000 1005
Asn Phe Gly Phe Lys Arg Gly Arg Phe Lys Val Glu Lys Gln Val
1010 1015 1020
Tyr Gln Lys Leu Glu Lys Met Leu Ile Glu Lys Leu Asn Tyr Leu

1025 1030 1035

Val Phe Lys Asp Asn Glu Phe Asp Lys Thr Gly Gly Val Leu Arg

1040 1045 1050

Ala Tyr Gln Leu Thr Ala Pro Phe Glu Thr Phe Lys Lys Met Gly

1055 1060 1065

Lys Gln Thr Gly Ile Ile Tyr Tyr Val Pro Ala Gly Phe Thr Ser

1070 1075 1080

Lys Ile Cys Pro Val Thr Gly Phe Val Asn Gln Leu Tyr Pro Lys

1085 1090 1095

Tyr Glu Ser Val Ser Lys Ser Gln Glu Phe Phe Ser Lys Phe Asp

1100 1105 1110

Lys Ile Cys Tyr Asn Leu Asp Lys Gly Tyr Phe Glu Phe Ser Phe

1115 1120 1125

Asp Tyr Lys Asn Phe Gly Asp Lys Ala Ala Lys Gly Lys Trp Thr

1130 1135 1140

Ile Ala Ser Phe Gly Ser Arg Leu Ile Asn Phe Arg Asn Ser Asp

1145 1150 1155

Lys Asn His Asn Trp Asp Thr Arg Glu Val Tyr Pro Thr Lys Glu

1160 1165 1170

Leu Glu Lys Leu Leu Lys Asp Tyr Ser Ile Glu Tyr Gly His Gly

1175 1180 1185

Glu Cys Ile Lys Ala Ala Ile Cys Gly Glu Ser Asp Lys Lys Phe

1190 1195 1200

Phe Ala Lys Leu Thr Ser Val Leu Asn Thr Ile Leu Gln Met Arg

1205 1210 1215

Asn Ser Lys Thr Gly Thr Glu Leu Asp Tyr Leu Ile Ser Pro Val

1220 1225 1230

Ala Asp Val Asn Gly Asn Phe Phe Asp Ser Arg Gln Ala Pro Lys

1235 1240 1245

Asn Met Pro Gln Asp Ala Ala Ala Asn Gly Ala Tyr His Ile Gly

1250 1255 1260

Leu Lys Gly Leu Met Leu Leu Gly Arg Ile Lys Asn Asn Gln Glu

1265 1270 1275

Gly Lys Lys Leu Asn Leu Val Ile Lys Asn Glu Glu Tyr Phe Glu

1280 1285 1290

Phe Val Gln Asn Arg Asn Asn

1295 1300

<210> 37

<211> 1300

<212> PRT
<213> 人工序列
<220>
<223> 合成多肽
<400> 37

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

Leu Asn Gln Ser Gly Ile Thr Lys Phe Asn Thr Ile Ile Gly Gly Lys
275 280 285
Phe Val Asn Gly Glu Asn Thr Lys Arg Lys Gly Ile Asn Glu Tyr Ile
290 295 300
Asn Leu Tyr Ser Gln Gln Ile Asn Asp Lys Thr Leu Lys Lys Tyr Lys
305 310 315 320
Met Ser Val Leu Phe Lys Gln Ile Leu Ser Asp Thr Glu Ser Lys Ser
325 330 335
Phe Val Ile Asp Lys Leu Glu Asp Asp Ser Asp Val Val Thr Thr Met
340 345 350
Gln Ser Phe Tyr Glu Gln Ile Ala Ala Phe Lys Thr Val Glu Glu Lys
355 360 365
Ser Ile Lys Glu Thr Leu Ser Leu Leu Phe Asp Asp Leu Lys Ala Gln
370 375 380
Lys Leu Asp Leu Ser Lys Ile Tyr Phe Lys Asn Asp Lys Ser Leu Thr
385 390 395 400
Asp Leu Ser Gln Gln Val Phe Asp Asp Tyr Ser Val Ile Gly Thr Ala
405 410 415
Val Leu Glu Tyr Ile Thr Gln Gln Ile Ala Pro Lys Asn Leu Asp Asn
420 425 430
Pro Ser Lys Lys Glu Gln Glu Leu Ile Ala Lys Lys Thr Glu Lys Ala
435 440 445
Lys Tyr Leu Ser Leu Glu Thr Ile Lys Leu Ala Leu Glu Glu Phe Asn
450 455 460
Lys His Arg Asp Ile Asp Lys Gln Cys Arg Phe Glu Glu Ile Leu Ala
465 470 475 480
Asn Phe Ala Ala Ile Pro Met Ile Phe Asp Glu Ile Ala Gln Asn Lys
485 490 495
Asp Asn Leu Ala Gln Ile Ser Ile Lys Tyr Gln Asn Gln Gly Lys Lys
500 505 510
Asp Leu Leu Gln Ala Ser Ala Glu Asp Asp Val Lys Ala Ile Lys Asp
515 520 525
Leu Leu Asp Gln Thr Asn Asn Leu Leu His Lys Leu Lys Ile Phe His
530 535 540
Ile Ser Gln Ser Glu Asp Lys Ala Asn Ile Leu Asp Lys Asp Glu His
545 550 555 560
Phe Tyr Leu Val Phe Glu Glu Cys Tyr Phe Glu Leu Ala Asn Ile Val
565 570 575
Pro Leu Tyr Asn Lys Ile Arg Asn Tyr Ile Thr Gln Lys Pro Tyr Ser

580 585 590

Asp Glu Lys Phe Lys Leu Asn Phe Glu Asn Ser Thr Leu Ala Asn Gly

595 600 605

Trp Asp Lys Asn Lys Glu Pro Asp Asn Thr Ala Ile Leu Phe Ile Lys

610 615 620

Asp Asp Lys Tyr Tyr Leu Gly Val Met Asn Lys Lys Asn Asn Lys Ile

625 630 635 640

Phe Asp Asp Lys Ala Ile Lys Glu Asn Lys Gly Glu Gly Tyr Lys Lys

645 650 655

Ile Val Tyr Lys Leu Leu Pro Gly Ala Asn Lys Met Leu Pro Lys Val

660 665 670

Phe Phe Ser Ala Lys Ser Ile Lys Phe Tyr Asn Pro Ser Glu Asp Ile

675 680 685

Leu Arg Ile Arg Asn His Ser Thr His Thr Lys Asn Gly Ser Pro Gln

690 695 700

Lys Gly Tyr Glu Lys Phe Glu Phe Asn Ile Glu Asp Cys Arg Lys Phe

705 710 715 720

Ile Asp Phe Tyr Lys Gln Ser Ile Ser Lys His Pro Glu Trp Lys Asp

725 730 735

Phe Gly Phe Arg Phe Ser Asp Thr Gln Arg Tyr Asn Ser Ile Asp Glu

740 745 750

Phe Tyr Arg Glu Val Glu Asn Gln Gly Tyr Lys Leu Thr Phe Glu Asn

755 760 765

Ile Ser Glu Ser Tyr Ile Asp Ser Val Val Asn Gln Gly Lys Leu Tyr

770 775 780

Leu Phe Gln Ile Tyr Asn Lys Asp Phe Ser Ala Tyr Ser Lys Gly Arg

785 790 795 800

Pro Asn Leu His Thr Leu Tyr Trp Lys Ala Leu Phe Asp Glu Arg Asn

805 810 815

Leu Gln Asp Val Val Tyr Lys Leu Asn Gly Glu Ala Glu Leu Phe Tyr

820 825 830

Arg Lys Gln Ser Ile Pro Lys Lys Ile Thr His Pro Ala Lys Glu Ala

835 840 845

Ile Ala Asn Lys Asn Lys Asp Asn Pro Lys Lys Glu Ser Val Phe Glu

850 855 860

Tyr Asp Leu Ile Lys Asp Lys Arg Phe Thr Glu Asp Lys Phe Phe Phe

865 870 875 880

His Cys Pro Ile Thr Ile Asn Phe Lys Ser Ser Gly Ala Asn Lys Phe

885 890 895

Asn Asp Glu Ile Asn Leu Leu Leu Lys Glu Lys Ala Asn Asp Val His
900 905 910
Ile Leu Ser Ile Ala Arg Gly Glu Arg His Leu Ala Tyr Tyr Thr Leu
915 920 925
Val Asp Gly Lys Gly Asn Ile Ile Lys Gln Asp Thr Phe Asn Ile Ile
930 935 940
Gly Asn Asp Arg Met Lys Thr Asn Tyr His Asp Lys Leu Ala Ala Ile
945 950 955 960
Glu Lys Asp Arg Asp Ser Ala Arg Lys Asp Trp Lys Lys Ile Asn Asn
965 970 975
Ile Lys Glu Met Lys Glu Gly Tyr Leu Ser Gln Val Val His Glu Ile
980 985 990
Ala Lys Leu Val Ile Glu Tyr Asn Ala Ile Val Val Phe Ala Asp Leu
995 1000 1005
Asn Phe Gly Phe Lys Arg Gly Arg Phe Lys Val Glu Lys Gln Val
1010 1015 1020
Tyr Gln Lys Leu Glu Lys Met Leu Ile Glu Lys Leu Asn Tyr Leu
1025 1030 1035
Val Phe Lys Asp Asn Glu Phe Asp Lys Thr Gly Gly Val Leu Arg
1040 1045 1050
Ala Tyr Gln Leu Thr Ala Pro Phe Glu Thr Phe Lys Lys Met Gly
1055 1060 1065
Lys Gln Thr Gly Ile Ile Tyr Tyr Val Pro Ala Gly Phe Thr Ser
1070 1075 1080
Lys Ile Cys Pro Val Thr Gly Phe Val Asn Gln Leu Tyr Pro Lys
1085 1090 1095
Tyr Glu Ser Val Ser Lys Ser Gln Glu Phe Phe Ser Lys Phe Asp
1100 1105 1110
Lys Ile Cys Tyr Asn Leu Asp Lys Gly Tyr Phe Glu Phe Ser Phe
1115 1120 1125
Asp Tyr Lys Asn Phe Gly Asp Lys Ala Ala Lys Gly Lys Trp Thr
1130 1135 1140
Ile Ala Ser Phe Gly Ser Arg Leu Ile Asn Phe Arg Asn Ser Asp
1145 1150 1155
Lys Asn His Asn Trp Asp Thr Arg Glu Val Tyr Pro Thr Lys Glu
1160 1165 1170
Leu Glu Lys Leu Leu Lys Asp Tyr Ser Ile Glu Tyr Gly His Gly
1175 1180 1185
Glu Cys Ile Lys Ala Ala Ile Cys Gly Glu Ser Asp Lys Lys Phe

1190 1195 1200
Phe Ala Lys Leu Thr Ser Val Leu Asn Thr Ile Leu Gln Met Arg
1205 1210 1215
Asn Ser Lys Thr Gly Thr Glu Leu Asp Tyr Leu Ile Ser Pro Val
1220 1225 1230
Ala Asp Val Asn Gly Asn Phe Phe Asp Ser Arg Gln Ala Pro Lys
1235 1240 1245
Asn Met Pro Gln Asp Ala Ala Ala Asn Gly Ala Tyr His Ile Gly
1250 1255 1260
Leu Lys Gly Leu Met Leu Leu Gly Arg Ile Lys Asn Asn Gln Glu
1265 1270 1275
Gly Lys Lys Leu Asn Leu Val Ile Lys Asn Glu Glu Tyr Phe Glu
1280 1285 1290
Phe Val Gln Asn Arg Asn Asn
1295 1300

<210> 38
<211> 887
<212> PRT
<213> 格氏嗜盐碱杆菌
<400> 38

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

145 150 155 160
Trp Thr Leu His Thr Tyr Thr Leu Thr Ala Thr Asp Gly Ala Lys Thr
165 170 175
Asp Thr Glu Tyr Ala Arg Arg Thr Leu Ala Tyr Thr Val Arg Gln Glu
180 185 190
Leu Tyr Thr Asp His Asp Ala Ala Pro Val Ala Thr Asp Gly Leu Met
195 200 205
Leu Leu Thr Pro Glu Pro Leu Gly Glu Thr Pro Leu Asp Leu Asp Cys
210 215 220
Gly Val Arg Val Glu Ala Asp Glu Thr Arg Thr Leu Asp Tyr Thr Thr
225 230 235 240
Ala Lys Asp Arg Leu Leu Ala Arg Glu Leu Val Glu Glu Gly Leu Lys
245 250 255
Arg Ser Leu Trp Asp Asp Tyr Leu Val Arg Gly Ile Asp Glu Val Leu
260 265 270
Ser Lys Glu Pro Val Leu Thr Cys Asp Glu Phe Asp Leu His Glu Arg
275 280 285
Tyr Asp Leu Ser Val Glu Val Gly His Ser Gly Arg Ala Tyr Leu His
290 295 300
Ile Asn Phe Arg His Arg Phe Val Pro Lys Leu Thr Leu Ala Asp Ile
305 310 315 320
Asp Asp Asp Asn Ile Tyr Pro Gly Leu Arg Val Lys Thr Thr Tyr Arg
325 330 335
Pro Arg Arg Gly His Ile Val Trp Gly Leu Arg Asp Glu Cys Ala Thr
340 345 350
Asp Ser Leu Asn Thr Leu Gly Asn Gln Ser Val Val Ala Tyr His Arg
355 360 365
Asn Asn Gln Thr Pro Ile Asn Thr Asp Leu Leu Asp Ala Ile Glu Ala
370 375 380
Ala Asp Arg Arg Val Val Glu Thr Arg Arg Gln Gly His Gly Asp Asp
385 390 395 400
Ala Val Ser Phe Pro Gln Glu Leu Leu Ala Val Glu Pro Asn Thr His
405 410 415
Gln Ile Lys Gln Phe Ala Ser Asp Gly Phe His Gln Gln Ala Arg Ser
420 425 430
Lys Thr Arg Leu Ser Ala Ser Arg Cys Ser Glu Lys Ala Gln Ala Phe
435 440 445
Ala Glu Arg Leu Asp Pro Val Arg Leu Asn Gly Ser Thr Val Glu Phe
450 455 460

Ser Ser Glu Phe Phe Thr Gly Asn Asn Glu Gln Gln Leu Arg Leu Leu
465 470 475 480
Tyr Glu Asn Gly Glu Ser Val Leu Thr Phe Arg Asp Gly Ala Arg Gly
485 490 495
Ala His Pro Asp Glu Thr Phe Ser Lys Gly Ile Val Asn Pro Pro Glu
500 505 510
Ser Phe Glu Val Ala Val Val Leu Pro Glu Gln Gln Ala Asp Thr Cys
515 520 525
Lys Ala Gln Trp Asp Thr Met Ala Asp Leu Leu Asn Gln Ala Gly Ala
530 535 540
Pro Pro Thr Arg Ser Glu Thr Val Gln Tyr Asp Ala Phe Ser Ser Pro
545 550 555 560
Glu Ser Ile Ser Leu Asn Val Ala Gly Ala Ile Asp Pro Ser Glu Val
565 570 575
Asp Ala Ala Phe Val Val Leu Pro Pro Asp Gln Glu Gly Phe Ala Asp
580 585 590
Leu Ala Ser Pro Thr Glu Thr Tyr Asp Glu Leu Lys Lys Ala Leu Ala
595 600 605
Asn Met Gly Ile Tyr Ser Gln Met Ala Tyr Phe Asp Arg Phe Arg Asp
610 615 620
Ala Lys Ile Phe Tyr Thr Arg Asn Val Ala Leu Gly Leu Leu Ala Ala
625 630 635 640
Ala Gly Gly Val Ala Phe Thr Thr Glu His Ala Met Pro Gly Asp Ala
645 650 655
Asp Met Phe Ile Gly Ile Asp Val Ser Arg Ser Tyr Pro Glu Asp Gly
660 665 670
Ala Ser Gly Gln Ile Asn Ile Ala Ala Thr Ala Thr Ala Val Tyr Lys
675 680 685
Asp Gly Thr Ile Leu Gly His Ser Ser Thr Arg Pro Gln Leu Gly Glu
690 695 700
Lys Leu Gln Ser Thr Asp Val Arg Asp Ile Met Lys Asn Ala Ile Leu
705 710 715 720
Gly Tyr Gln Gln Val Thr Gly Glu Ser Pro Thr His Ile Val Ile His
725 730 735
Arg Asp Gly Phe Met Asn Glu Asp Leu Asp Pro Ala Thr Glu Phe Leu
740 745 750
Asn Glu Gln Gly Val Glu Tyr Asp Ile Val Glu Ile Arg Lys Gln Pro
755 760 765
Gln Thr Arg Leu Leu Ala Val Ser Asp Val Gln Tyr Asp Thr Pro Val

770 775 780
Lys Ser Ile Ala Ala Ile Asn Gln Asn Glu Pro Arg Ala Thr Val Ala
785 790 795 800
Thr Phe Gly Ala Pro Glu Tyr Leu Ala Thr Arg Asp Gly Gly Gly Leu
805 810 815
Pro Arg Pro Ile Gln Ile Glu Arg Val Ala Gly Glu Thr Asp Ile Glu
820 825 830
Thr Leu Thr Arg Gln Val Tyr Leu Leu Ser Gln Ser His Ile Gln Val
835 840 845
His Asn Ser Thr Ala Arg Leu Pro Ile Thr Thr Ala Tyr Ala Asp Gln
850 855 860
Ala Ser Thr His Ala Thr Lys Gly Tyr Leu Val Gln Thr Gly Ala Phe
865 870 875 880
Glu Ser Asn Val Gly Phe Leu
885

<210> 39
<211> 1129
<212> PRT
<213> 酸土脂环酸芽孢杆菌
<400> 39
Met Ala Val Lys Ser Ile Lys Val Lys Leu Arg Leu Asp Asp Met Pro
1 5 10 15
Glu Ile Arg Ala Gly Leu Trp Lys Leu His Lys Glu Val Asn Ala Gly
20 25 30
Val Arg Tyr Tyr Thr Glu Trp Leu Ser Leu Leu Arg Gln Glu Asn Leu
35 40 45
Tyr Arg Arg Ser Pro Asn Gly Asp Gly Glu Gln Glu Cys Asp Lys Thr
50 55 60
Ala Glu Glu Cys Lys Ala Glu Leu Leu Glu Arg Leu Arg Ala Arg Gln
65 70 75 80
Val Glu Asn Gly His Arg Gly Pro Ala Gly Ser Asp Asp Glu Leu Leu
85 90 95
Gln Leu Ala Arg Gln Leu Tyr Glu Leu Leu Val Pro Gln Ala Ile Gly
100 105 110
Ala Lys Gly Asp Ala Gln Gln Ile Ala Arg Lys Phe Leu Ser Pro Leu
115 120 125
Ala Asp Lys Asp Ala Val Gly Gly Leu Gly Ile Ala Lys Ala Gly Asn
130 135 140
Lys Pro Arg Trp Val Arg Met Arg Glu Ala Gly Glu Pro Gly Trp Glu

145 150 155 160
Glu Glu Lys Glu Lys Ala Glu Thr Arg Lys Ser Ala Asp Arg Thr Ala
165 170 175
Asp Val Leu Arg Ala Leu Ala Asp Phe Gly Leu Lys Pro Leu Met Arg
180 185 190
Val Tyr Thr Asp Ser Glu Met Ser Ser Val Glu Trp Lys Pro Leu Arg
195 200 205
Lys Gly Gln Ala Val Arg Thr Trp Asp Arg Asp Met Phe Gln Gln Ala
210 215 220
Ile Glu Arg Met Met Ser Trp Glu Ser Trp Asn Gln Arg Val Gly Gln
225 230 235 240
Glu Tyr Ala Lys Leu Val Glu Gln Lys Asn Arg Phe Glu Gln Lys Asn
245 250 255
Phe Val Gly Gln Glu His Leu Val His Leu Val Asn Gln Leu Gln Gln
260 265 270
Asp Met Lys Glu Ala Ser Pro Gly Leu Glu Ser Lys Glu Gln Thr Ala
275 280 285
His Tyr Val Thr Gly Arg Ala Leu Arg Gly Ser Asp Lys Val Phe Glu
290 295 300
Lys Trp Gly Lys Leu Ala Pro Asp Ala Pro Phe Asp Leu Tyr Asp Ala
305 310 315 320
Glu Ile Lys Asn Val Gln Arg Arg Asn Thr Arg Arg Phe Gly Ser His
325 330 335
Asp Leu Phe Ala Lys Leu Ala Glu Pro Glu Tyr Gln Ala Leu Trp Arg
340 345 350
Glu Asp Ala Ser Phe Leu Thr Arg Tyr Ala Val Tyr Asn Ser Ile Leu
355 360 365
Arg Lys Leu Asn His Ala Lys Met Phe Ala Thr Phe Thr Leu Pro Asp
370 375 380
Ala Thr Ala His Pro Ile Trp Thr Arg Phe Asp Lys Leu Gly Gly Asn
385 390 395 400
Leu His Gln Tyr Thr Phe Leu Phe Asn Glu Phe Gly Glu Arg Arg His
405 410 415
Ala Ile Arg Phe His Lys Leu Leu Lys Val Glu Asn Gly Val Ala Arg
420 425 430
Glu Val Asp Asp Val Thr Val Pro Ile Ser Met Ser Glu Gln Leu Asp
435 440 445
Asn Leu Leu Pro Arg Asp Pro Asn Glu Pro Ile Ala Leu Tyr Phe Arg
450 455 460

Asp Tyr Gly Ala Glu Gln His Phe Thr Gly Glu Phe Gly Gly Ala Lys
465 470 475 480
Ile Gln Cys Arg Arg Asp Gln Leu Ala His Met His Arg Arg Arg Gly
485 490 495
Ala Arg Asp Val Tyr Leu Asn Val Ser Val Arg Val Gln Ser Gln Ser
500 505 510
Glu Ala Arg Gly Glu Arg Arg Pro Pro Tyr Ala Ala Val Phe Arg Leu
515 520 525
Val Gly Asp Asn His Arg Ala Phe Val His Phe Asp Lys Leu Ser Asp
530 535 540
Tyr Leu Ala Glu His Pro Asp Asp Gly Lys Leu Gly Ser Glu Gly Leu
545 550 555 560
Leu Ser Gly Leu Arg Val Met Ser Val Asp Leu Gly Leu Arg Thr Ser
565 570 575
Ala Ser Ile Ser Val Phe Arg Val Ala Arg Lys Asp Glu Leu Lys Pro
580 585 590
Asn Ser Lys Gly Arg Val Pro Phe Phe Phe Pro Ile Lys Gly Asn Asp
595 600 605
Asn Leu Val Ala Val His Glu Arg Ser Gln Leu Leu Lys Leu Pro Gly
610 615 620
Glu Thr Glu Ser Lys Asp Leu Arg Ala Ile Arg Glu Glu Arg Gln Arg
625 630 635 640
Thr Leu Arg Gln Leu Arg Thr Gln Leu Ala Tyr Leu Arg Leu Leu Val
645 650 655
Arg Cys Gly Ser Glu Asp Val Gly Arg Arg Glu Arg Ser Trp Ala Lys
660 665 670
Leu Ile Glu Gln Pro Val Asp Ala Ala Asn His Met Thr Pro Asp Trp
675 680 685
Arg Glu Ala Phe Glu Asn Glu Leu Gln Lys Leu Lys Ser Leu His Gly
690 695 700
Ile Cys Ser Asp Lys Glu Trp Met Asp Ala Val Tyr Glu Ser Val Arg
705 710 715 720
Arg Val Trp Arg His Met Gly Lys Gln Val Arg Asp Trp Arg Lys Asp
725 730 735
Val Arg Ser Gly Glu Arg Pro Lys Ile Arg Gly Tyr Ala Lys Asp Val
740 745 750
Val Gly Gly Asn Ser Ile Glu Gln Ile Glu Tyr Leu Glu Arg Gln Tyr
755 760 765
Lys Phe Leu Lys Ser Trp Ser Phe Phe Gly Lys Val Ser Gly Gln Val

770 775 780

Ile Arg Ala Glu Lys Gly Ser Arg Phe Ala Ile Thr Leu Arg Glu His
785 790 795 800

Ile Asp His Ala Lys Glu Asp Arg Leu Lys Lys Leu Ala Asp Arg Ile
805 810 815

Ile Met Glu Ala Leu Gly Tyr Val Tyr Ala Leu Asp Glu Arg Gly Lys
820 825 830

Gly Lys Trp Val Ala Lys Tyr Pro Pro Cys Gln Leu Ile Leu Leu Glu
835 840 845

Glu Leu Ser Glu Tyr Gln Phe Asn Asn Asp Arg Pro Pro Ser Glu Asn
850 855 860

Asn Gln Leu Met Gln Trp Ser His Arg Gly Val Phe Gln Glu Leu Ile
865 870 875 880

Asn Gln Ala Gln Val His Asp Leu Leu Val Gly Thr Met Tyr Ala Ala
885 890 895

Phe Ser Ser Arg Phe Asp Ala Arg Thr Gly Ala Pro Gly Ile Arg Cys
900 905 910

Arg Arg Val Pro Ala Arg Cys Thr Gln Glu His Asn Pro Glu Pro Phe
915 920 925

Pro Trp Trp Leu Asn Lys Phe Val Val Glu His Thr Leu Asp Ala Cys
930 935 940

Pro Leu Arg Ala Asp Asp Leu Ile Pro Thr Gly Glu Gly Glu Ile Phe
945 950 955 960

Val Ser Pro Phe Ser Ala Glu Glu Gly Asp Phe His Gln Ile His Ala
965 970 975

Asp Leu Asn Ala Ala Gln Asn Leu Gln Gln Arg Leu Trp Ser Asp Phe
980 985 990

Asp Ile Ser Gln Ile Arg Leu Arg Cys Asp Trp Gly Glu Val Asp Gly
995 1000 1005

Glu Leu Val Leu Ile Pro Arg Leu Thr Gly Lys Arg Thr Ala Asp
1010 1015 1020

Ser Tyr Ser Asn Lys Val Phe Tyr Thr Asn Thr Gly Val Thr Tyr
1025 1030 1035

Tyr Glu Arg Glu Arg Gly Lys Lys Arg Arg Lys Val Phe Ala Gln
1040 1045 1050

Glu Lys Leu Ser Glu Glu Glu Ala Glu Leu Leu Val Glu Ala Asp
1055 1060 1065

Glu Ala Arg Glu Lys Ser Val Val Leu Met Arg Asp Pro Ser Gly
1070 1075 1080

Ile Ile Asn Arg Gly Asn Trp Thr Arg Gln Lys Glu Phe Trp Ser
1085 1090 1095
Met Val Asn Gln Arg Ile Glu Gly Tyr Leu Val Lys Gln Ile Arg
1100 1105 1110
Ser Arg Val Pro Leu Gln Asp Ser Ala Cys Glu Asn Thr Gly Asp
1115 1120 1125
Ile

<210> 40
<211> 1389
<212> PRT
<213> Leptotrichia shahii
<400> 40

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

210 215 220

Leu Arg Glu Lys Leu Leu Lys Asp Asp Lys Ile Asp Val Ile Leu Thr
225 230 235 240
Asn Phe Met Glu Ile Arg Glu Lys Ile Lys Ser Asn Leu Glu Ile Leu
245 250 255
Gly Phe Val Lys Phe Tyr Leu Asn Val Gly Gly Asp Lys Lys Lys Ser
260 265 270
Lys Asn Lys Lys Met Leu Val Glu Lys Ile Leu Asn Ile Asn Val Asp
275 280 285
Leu Thr Val Glu Asp Ile Ala Asp Phe Val Ile Lys Glu Leu Glu Phe
290 295 300
Trp Asn Ile Thr Lys Arg Ile Glu Lys Val Lys Lys Val Asn Asn Glu
305 310 315 320
Phe Leu Glu Lys Arg Arg Asn Arg Thr Tyr Ile Lys Ser Tyr Val Leu
325 330 335
Leu Asp Lys His Glu Lys Phe Lys Ile Glu Arg Glu Asn Lys Lys Asp
340 345 350
Lys Ile Val Lys Phe Phe Val Glu Asn Ile Lys Asn Asn Ser Ile Lys
355 360 365
Glu Lys Ile Glu Lys Ile Leu Ala Glu Phe Lys Ile Asp Glu Leu Ile
370 375 380
Lys Lys Leu Glu Lys Glu Leu Lys Lys Gly Asn Cys Asp Thr Glu Ile
385 390 395 400
Phe Gly Ile Phe Lys Lys His Tyr Lys Val Asn Phe Asp Ser Lys Lys
405 410 415
Phe Ser Lys Lys Ser Asp Glu Glu Lys Glu Leu Tyr Lys Ile Ile Tyr
420 425 430
Arg Tyr Leu Lys Gly Arg Ile Glu Lys Ile Leu Val Asn Glu Gln Lys
435 440 445
Val Arg Leu Lys Lys Met Glu Lys Ile Glu Ile Glu Lys Ile Leu Asn
450 455 460
Glu Ser Ile Leu Ser Glu Lys Ile Leu Lys Arg Val Lys Gln Tyr Thr
465 470 475 480
Leu Glu His Ile Met Tyr Leu Gly Lys Leu Arg His Asn Asp Ile Asp
485 490 495
Met Thr Thr Val Asn Thr Asp Asp Phe Ser Arg Leu His Ala Lys Glu
500 505 510
Glu Leu Asp Leu Glu Leu Ile Thr Phe Phe Ala Ser Thr Asn Met Glu
515 520 525

Leu Asn Lys Ile Phe Ser Arg Glu Asn Ile Asn Asn Asp Glu Asn Ile
530 535 540
Asp Phe Phe Gly Gly Asp Arg Glu Lys Asn Tyr Val Leu Asp Lys Lys
545 550 555 560
Ile Leu Asn Ser Lys Ile Lys Ile Ile Arg Asp Leu Asp Phe Ile Asp
565 570 575
Asn Lys Asn Asn Ile Thr Asn Asn Phe Ile Arg Lys Phe Thr Lys Ile
580 585 590
Gly Thr Asn Glu Arg Asn Arg Ile Leu His Ala Ile Ser Lys Glu Arg
595 600 605
Asp Leu Gln Gly Thr Gln Asp Asp Tyr Asn Lys Val Ile Asn Ile Ile
610 615 620
Gln Asn Leu Lys Ile Ser Asp Glu Glu Val Ser Lys Ala Leu Asn Leu
625 630 635 640
Asp Val Val Phe Lys Asp Lys Lys Asn Ile Ile Thr Lys Ile Asn Asp
645 650 655
Ile Lys Ile Ser Glu Glu Asn Asn Asn Asp Ile Lys Tyr Leu Pro Ser
660 665 670
Phe Ser Lys Val Leu Pro Glu Ile Leu Asn Leu Tyr Arg Asn Asn Pro
675 680 685
Lys Asn Glu Pro Phe Asp Thr Ile Glu Thr Glu Lys Ile Val Leu Asn
690 695 700
Ala Leu Ile Tyr Val Asn Lys Glu Leu Tyr Lys Lys Leu Ile Leu Glu
705 710 715 720
Asp Asp Leu Glu Glu Asn Glu Ser Lys Asn Ile Phe Leu Gln Glu Leu
725 730 735
Lys Lys Thr Leu Gly Asn Ile Asp Glu Ile Asp Glu Asn Ile Ile Glu
740 745 750
Asn Tyr Tyr Lys Asn Ala Gln Ile Ser Ala Ser Lys Gly Asn Asn Lys
755 760 765
Ala Ile Lys Lys Tyr Gln Lys Lys Val Ile Glu Cys Tyr Ile Gly Tyr
770 775 780
Leu Arg Lys Asn Tyr Glu Glu Leu Phe Asp Phe Ser Asp Phe Lys Met
785 790 795 800
Asn Ile Gln Glu Ile Lys Lys Gln Ile Lys Asp Ile Asn Asp Asn Lys
805 810 815
Thr Tyr Glu Arg Ile Thr Val Lys Thr Ser Asp Lys Thr Ile Val Ile
820 825 830
Asn Asp Asp Phe Glu Tyr Ile Ile Ser Ile Phe Ala Leu Leu Asn Ser

835 840 845

Asn Ala Val Ile Asn Lys Ile Arg Asn Arg Phe Phe Ala Thr Ser Val

850 855 860

Trp Leu Asn Thr Ser Glu Tyr Gln Asn Ile Ile Asp Ile Leu Asp Glu

865 870 875 880

Ile Met Gln Leu Asn Thr Leu Arg Asn Glu Cys Ile Thr Glu Asn Trp

885 890 895

Asn Leu Asn Leu Glu Glu Phe Ile Gln Lys Met Lys Glu Ile Glu Lys

900 905 910

Asp Phe Asp Asp Phe Lys Ile Gln Thr Lys Lys Glu Ile Phe Asn Asn

915 920 925

Tyr Tyr Glu Asp Ile Lys Asn Asn Ile Leu Thr Glu Phe Lys Asp Asp

930 935 940

Ile Asn Gly Cys Asp Val Leu Glu Lys Lys Leu Glu Lys Ile Val Ile

945 950 955 960

Phe Asp Asp Glu Thr Lys Phe Glu Ile Asp Lys Lys Ser Asn Ile Leu

965 970 975

Gln Asp Glu Gln Arg Lys Leu Ser Asn Ile Asn Lys Lys Asp Leu Lys

980 985 990

Lys Lys Val Asp Gln Tyr Ile Lys Asp Lys Asp Gln Glu Ile Lys Ser

995 1000 1005

Lys Ile Leu Cys Arg Ile Ile Phe Asn Ser Asp Phe Leu Lys Lys

1010 1015 1020

Tyr Lys Lys Glu Ile Asp Asn Leu Ile Glu Asp Met Glu Ser Glu

1025 1030 1035

Asn Glu Asn Lys Phe Gln Glu Ile Tyr Tyr Pro Lys Glu Arg Lys

1040 1045 1050

Asn Glu Leu Tyr Ile Tyr Lys Lys Asn Leu Phe Leu Asn Ile Gly

1055 1060 1065

Asn Pro Asn Phe Asp Lys Ile Tyr Gly Leu Ile Ser Asn Asp Ile

1070 1075 1080

Lys Met Ala Asp Ala Lys Phe Leu Phe Asn Ile Asp Gly Lys Asn

1085 1090 1095

Ile Arg Lys Asn Lys Ile Ser Glu Ile Asp Ala Ile Leu Lys Asn

1100 1105 1110

Leu Asn Asp Lys Leu Asn Gly Tyr Ser Lys Glu Tyr Lys Glu Lys

1115 1120 1125

Tyr Ile Lys Lys Leu Lys Glu Asn Asp Asp Phe Phe Ala Lys Asn

1130 1135 1140

Ile Gln Asn Lys Asn Tyr Lys Ser Phe Glu Lys Asp Tyr Asn Arg
1145 1150 1155
Val Ser Glu Tyr Lys Lys Ile Arg Asp Leu Val Glu Phe Asn Tyr
1160 1165 1170
Leu Asn Lys Ile Glu Ser Tyr Leu Ile Asp Ile Asn Trp Lys Leu
1175 1180 1185
Ala Ile Gln Met Ala Arg Phe Glu Arg Asp Met His Tyr Ile Val
1190 1195 1200
Asn Gly Leu Arg Glu Leu Gly Ile Ile Lys Leu Ser Gly Tyr Asn
1205 1210 1215
Thr Gly Ile Ser Arg Ala Tyr Pro Lys Arg Asn Gly Ser Asp Gly
1220 1225 1230
Phe Tyr Thr Thr Thr Ala Tyr Tyr Lys Phe Phe Asp Glu Glu Ser
1235 1240 1245
Tyr Lys Lys Phe Glu Lys Ile Cys Tyr Gly Phe Gly Ile Asp Leu
1250 1255 1260
Ser Glu Asn Ser Glu Ile Asn Lys Pro Glu Asn Glu Ser Ile Arg
1265 1270 1275
Asn Tyr Ile Ser His Phe Tyr Ile Val Arg Asn Pro Phe Ala Asp
1280 1285 1290
Tyr Ser Ile Ala Glu Gln Ile Asp Arg Val Ser Asn Leu Leu Ser
1295 1300 1305
Tyr Ser Thr Arg Tyr Asn Asn Ser Thr Tyr Ala Ser Val Phe Glu
1310 1315 1320
Val Phe Lys Lys Asp Val Asn Leu Asp Tyr Asp Glu Leu Lys Lys
1325 1330 1335
Lys Phe Lys Leu Ile Gly Asn Asn Asp Ile Leu Glu Arg Leu Met
1340 1345 1350
Lys Pro Lys Lys Val Ser Val Leu Glu Leu Glu Ser Tyr Asn Ser
1355 1360 1365
Asp Tyr Ile Lys Asn Leu Ile Ile Glu Leu Leu Thr Lys Ile Glu
1370 1375 1380
Asn Thr Asn Asp Thr Leu
1385

<210> 41
<211> 7
<212> PRT
<213> 人工序列
<220>

<223> 合成多肽
<400> 41
Pro Lys Lys Lys Arg Lys Val
1 5
<210> 42
<211> 30
<212> PRT
<213> 人工序列
<220>
<223> 合成多肽
<400> 42
Met Asp Ser Leu Leu Met Asn Arg Arg Lys Phe Leu Tyr Gln Phe Lys
1 5 10 15
Asn Val Arg Trp Ala Lys Gly Arg Arg Glu Thr Tyr Leu Cys
20 25 30
<210> 43
<211> 18
<212> PRT
<213> 人工序列
<220>
<223> 合成多肽
<400> 43
Lys Arg Thr Ala Asp Gly Ser Glu Phe Glu Ser Pro Lys Lys Lys Arg
1 5 10 15
Lys Val
<210> 44
<211> 19
<212> PRT
<213> 人工序列
<220>
<223> 合成多肽
<400> 44
Lys Arg Gly Ile Asn Asp Arg Asn Phe Trp Arg Gly Glu Asn Gly Arg
1 5 10 15
Lys Thr Arg
<210> 45
<211> 17
<212> PRT
<213> 人工序列

<220>
<223> 合成多肽
<400> 45
Lys Lys Thr Gly Gly Pro Ile Tyr Arg Arg Val Asp Gly Lys Trp Arg
1 5 10 15
Arg
<210> 46
<211> 17
<212> PRT
<213> 人工序列
<220>
<223> 合成多肽
<400> 46
Arg Arg Glu Leu Ile Leu Tyr Asp Lys Glu Glu Ile Arg Arg Ile Trp
1 5 10 15
Arg
<210> 47
<211> 8
<212> PRT
<213> 人工序列
<220>
<223> 合成多肽
<400> 47
Ala Val Ser Arg Lys Arg Lys Ala
1 5
<210> 48
<211> 313
<212> PRT
<213> 人工序列
<220>
<223> 合成多肽
<400> 48
Met Ile Gly Gln Lys Thr Leu Tyr Ser Phe Phe Ser Pro Ser Pro Ala
1 5 10 15
Arg Lys Arg His Ala Pro Ser Pro Glu Pro Ala Val Gln Gly Thr Gly
20 25 30
Val Ala Gly Val Pro Glu Glu Ser Gly Asp Ala Ala Ala Ile Pro Ala
35 40 45
Lys Lys Ala Pro Ala Gly Gln Glu Glu Pro Gly Thr Pro Pro Ser Ser

50 55 60

Pro Leu Ser Ala Glu Gln Leu Asp Arg Ile Gln Arg Asn Lys Ala Ala
65 70 75 80
Ala Leu Leu Arg Leu Ala Ala Arg Asn Val Pro Val Gly Phe Gly Glu
85 90 95
Ser Trp Lys Lys His Leu Ser Gly Glu Phe Gly Lys Pro Tyr Phe Ile
100 105 110
Lys Leu Met Gly Phe Val Ala Glu Glu Arg Lys His Tyr Thr Val Tyr
115 120 125
Pro Pro Pro His Gln Val Phe Thr Trp Thr Gln Met Cys Asp Ile Lys
130 135 140
Asp Val Lys Val Val Ile Leu Gly Gln Asp Pro Tyr His Gly Pro Asn
145 150 155 160
Gln Ala His Gly Leu Cys Phe Ser Val Gln Arg Pro Val Pro Pro Pro
165 170 175
Pro Ser Leu Glu Asn Ile Tyr Lys Glu Leu Ser Thr Asp Ile Glu Asp
180 185 190
Phe Val His Pro Gly His Gly Asp Leu Ser Gly Trp Ala Lys Gln Gly
195 200 205
Val Leu Leu Leu Asn Ala Val Leu Thr Val Arg Ala His Gln Ala Asn
210 215 220
Ser His Lys Glu Arg Gly Trp Glu Gln Phe Thr Asp Ala Val Val Ser
225 230 235 240
Trp Leu Asn Gln Asn Ser Asn Gly Leu Val Phe Leu Leu Trp Gly Ser
245 250 255
Tyr Ala Gln Lys Lys Gly Ser Ala Ile Asp Arg Lys Arg His His Val
260 265 270
Leu Gln Thr Ala His Pro Ser Pro Leu Ser Val Tyr Arg Gly Phe Phe
275 280 285
Gly Cys Arg His Phe Ser Lys Thr Asn Glu Leu Leu Gln Lys Ser Gly
290 295 300
Lys Lys Pro Ile Asp Trp Lys Glu Leu
305 310

<210> 49
<211> 209
<212> PRT
<213> 人工序列
<220>
<223> 合成多肽

<400> 49

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

<210> 50

<211> 209

<212> PRT

<213> 人工序列

<220>

<223> 合成多肽

<400> 50

Met Ala Gly Ala Gln Asp Phe Val Pro His Thr Ala Asp Leu Ala Glu
1 5 10 15
Leu Ala Ala Ala Ala Gly Glu Cys Arg Gly Cys Gly Leu Tyr Arg Asp
20 25 30

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

<210> 51
<211> 209
<212> PRT
<213> 人工序列
<220>
<223> 合成多肽
<400> 51

Met Ala Gly Ala Gln Asp Phe Val Pro His Thr Ala Asp Leu Ala Glu
1 5 10 15
Leu Ala Ala Ala Ala Gly Glu Cys Arg Gly Cys Gly Leu Tyr Arg Asp
20 25 30
Ala Thr Gln Ala Val Phe Gly Ala Gly Gly Arg Ser Ala Arg Ile Met
35 40 45
Met Ile Gly Glu Gln Pro Gly Asp Lys Glu Asp Leu Ala Gly Leu Pro
50 55 60
Phe Val Gly Pro Ala Gly Arg Leu Leu Asp Arg Ala Leu Glu Ala Ala

65 70 75 80
Asp Ile Asp Arg Asp Ala Leu Tyr Val Thr Asn Ala Val Lys His Phe
85 90 95
Lys Phe Thr Arg Ala Ala Gly Gly Lys Arg Arg Ile Ser Lys Thr Pro
100 105 110
Ser Arg Thr Glu Val Val Ala Cys Arg Pro Trp Leu Ile Ala Glu Met
115 120 125
Thr Ser Val Glu Pro Asp Val Val Val Leu Leu Gly Ala Thr Ala Ala
130 135 140
Lys Ala Leu Leu Gly Asn Asp Phe Arg Val Thr Gln His Arg Gly Glu
145 150 155 160
Val Leu His Val Asp Asp Val Pro Gly Asp Pro Ala Leu Val Ala Thr
165 170 175
Val His Pro Ser Ser Leu Leu Arg Gly Pro Lys Glu Glu Arg Glu Ser
180 185 190
Ala Phe Ala Gly Leu Val Asp Asp Leu Arg Val Ala Ala Asp Val Arg
195 200 205
Pro

<210> 52
<211> 211
<212> PRT
<213> 人工序列
<220>
<223> 合成多肽
<400> 52
Met Thr Thr Leu Thr Arg Gln Asp Leu Asn Phe Gly Gln Val Val Ala
1 5 10 15
Asp Val Leu Cys Glu Phe Leu Glu Val Ala Val His Leu Ile Leu Tyr
20 25 30
Val Arg Glu Val Tyr Pro Val Gly Ile Phe Gln Lys Arg Lys Lys Tyr
35 40 45
Asn Val Pro Val Gln Met Ser Cys His Pro Glu Leu Asn Gln Tyr Ile
50 55 60
Gln Asp Thr Leu His Cys Val Lys Pro Leu Leu Glu Lys Asn Asp Val
65 70 75 80
Glu Lys Val Val Val Val Ile Leu Asp Lys Glu His Arg Pro Val Glu
85 90 95
Lys Phe Val Phe Glu Ile Thr Gln Pro Pro Leu Leu Ser Ile Ser Ser
100 105 110

Asp Ser Leu Leu Ser His Val Glu Gln Leu Leu Arg Ala Phe Ile Leu
115 120 125
Lys Ile Ser Val Cys Asp Ala Val Leu Asp His Asn Pro Pro Gly Cys
130 135 140
Thr Phe Thr Val Leu Val His Thr Arg Glu Ala Ala Thr Arg Asn Met
145 150 155 160
Glu Lys Ile Gln Val Ile Lys Asp Phe Pro Trp Ile Leu Ala Asp Glu
165 170 175
Gln Asp Val His Met His Asp Pro Arg Leu Ile Pro Leu Lys Thr Met
180 185 190
Thr Ser Asp Ile Leu Lys Met Gln Leu Tyr Val Glu Glu Arg Ala His
195 200 205
Lys Gly Ser
210

<210> 53
<211> 270
<212> PRT
<213> 人工序列
<220>
<223> 合成多肽
<400> 53

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

Phe Arg Asn Leu Cys Gly Gln Pro Glu Val Phe Phe His His Cys Phe
145 150 155 160
Val His Asn Leu Cys Pro Leu Leu Phe Leu Ala Pro Ser Gly Arg Asn
165 170 175
Leu Thr Pro Ala Glu Leu Pro Ala Lys Gln Arg Glu Gln Leu Leu Gly
180 185 190
Ile Cys Asp Ala Ala Leu Cys Arg Gln Val Gln Leu Leu Gly Val Arg
195 200 205
Leu Val Val Gly Val Gly Arg Leu Ala Glu Gln Arg Ala Arg Arg Ala
210 215 220
Leu Ala Gly Leu Met Pro Glu Val Gln Val Glu Gly Leu Leu His Pro
225 230 235 240
Ser Pro Arg Asn Pro Gln Ala Asn Lys Gly Trp Glu Ala Val Ala Lys
245 250 255
Glu Arg Leu Asn Glu Leu Gly Leu Leu Pro Leu Leu Leu Lys
260 265 270

<210> 54
<211> 335
<212> PRT
<213> 人工序列
<220>
<223> 合成多肽
<400> 54

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

Glu Asp Lys Leu Asn His His Gln Arg Ile Gly Leu Lys Tyr Phe Gly
130 135 140
Asp Phe Glu Lys Arg Ile Pro Arg Glu Glu Met Leu Gln Met Gln Asp
145 150 155 160
Ile Val Leu Asn Glu Val Lys Lys Val Asp Ser Glu Tyr Ile Ala Thr
165 170 175
Val Cys Gly Ser Phe Arg Arg Gly Ala Glu Ser Ser Gly Asp Met Asp
180 185 190
Val Leu Leu Thr His Pro Ser Phe Thr Ser Glu Ser Thr Lys Gln Pro
195 200 205
Lys Leu Leu His Gln Val Val Glu Gln Leu Gln Lys Val His Phe Ile
210 215 220
Thr Asp Thr Leu Ser Lys Gly Glu Thr Lys Phe Met Gly Val Cys Gln
225 230 235 240
Leu Pro Ser Lys Asn Asp Glu Lys Glu Tyr Pro His Arg Arg Ile Asp
245 250 255
Ile Arg Leu Ile Pro Lys Asp Gln Tyr Tyr Cys Gly Val Leu Tyr Phe
260 265 270
Thr Gly Ser Asp Ile Phe Asn Lys Asn Met Arg Ala His Ala Leu Glu
275 280 285
Lys Gly Phe Thr Ile Asn Glu Tyr Thr Ile Arg Pro Leu Gly Val Thr
290 295 300
Gly Val Ala Gly Glu Pro Leu Pro Val Asp Ser Glu Lys Asp Ile Phe
305 310 315 320
Asp Tyr Ile Gln Trp Lys Tyr Arg Glu Pro Lys Asp Arg Ser Glu
325 330 335

<210> 55
<211> 575
<212> PRT
<213> 人工序列
<220>
<223> 合成多肽
<400> 55

Met Asp Pro Arg Gly Ile Leu Lys Ala Phe Pro Lys Arg Gln Lys Ile
1 5 10 15
His Ala Asp Ala Ser Ser Lys Val Leu Ala Lys Ile Pro Arg Arg Glu
20 25 30
Glu Gly Glu Glu Ala Glu Glu Trp Leu Ser Ser Leu Arg Ala His Val
35 40 45

Val Arg Thr Gly Ile Gly Arg Ala Arg Ala Glu Leu Phe Glu Lys Gln
50 55 60
Ile Val Gln His Gly Gly Gln Leu Cys Pro Ala Gln Gly Pro Gly Val
65 70 75 80
Thr His Ile Val Val Asp Glu Gly Met Asp Tyr Glu Arg Ala Leu Arg
85 90 95
Leu Leu Arg Leu Pro Gln Leu Pro Pro Gly Ala Gln Leu Val Lys Ser
100 105 110
Ala Trp Leu Ser Leu Cys Leu Gln Glu Arg Arg Leu Val Asp Val Ala
115 120 125
Gly Phe Ser Ile Phe Ile Pro Ser Arg Tyr Leu Asp His Pro Gln Pro
130 135 140
Ser Lys Ala Glu Gln Asp Ala Ser Ile Pro Pro Gly Thr His Glu Ala
145 150 155 160
Leu Leu Gln Thr Ala Leu Ser Pro Pro Pro Pro Pro Thr Arg Pro Val
165 170 175
Ser Pro Pro Gln Lys Ala Lys Glu Ala Pro Asn Thr Gln Ala Gln Pro
180 185 190
Ile Ser Asp Asp Glu Ala Ser Asp Gly Glu Glu Thr Gln Val Ser Ala
195 200 205
Ala Asp Leu Glu Ala Leu Ile Ser Gly His Tyr Pro Thr Ser Leu Glu
210 215 220
Gly Asp Cys Glu Pro Ser Pro Ala Pro Ala Val Leu Asp Lys Trp Val
225 230 235 240
Cys Ala Gln Pro Ser Ser Gln Lys Ala Thr Asn His Asn Leu His Ile
245 250 255
Thr Glu Lys Leu Glu Val Leu Ala Lys Ala Tyr Ser Val Gln Gly Asp
260 265 270
Lys Trp Arg Ala Leu Gly Tyr Ala Lys Ala Ile Asn Ala Leu Lys Ser
275 280 285
Phe His Lys Pro Val Thr Ser Tyr Gln Glu Ala Cys Ser Ile Pro Gly
290 295 300
Ile Gly Lys Arg Met Ala Glu Lys Ile Ile Glu Ile Leu Glu Ser Gly
305 310 315 320
His Leu Arg Lys Leu Asp His Ile Ser Glu Ser Val Pro Val Leu Glu
325 330 335
Leu Phe Ser Asn Ile Trp Gly Ala Gly Thr Lys Thr Ala Gln Met Trp
340 345 350
Tyr Gln Gln Gly Phe Arg Ser Leu Glu Asp Ile Arg Ser Gln Ala Ser

355 360 365

Leu Thr Thr Gln Gln Ala Ile Gly Leu Lys His Tyr Ser Asp Phe Leu
370 375 380

Glu Arg Met Pro Arg Glu Glu Ala Thr Glu Ile Glu Gln Thr Val Gln
385 390 395 400

Lys Ala Ala Gln Ala Phe Asn Ser Gly Leu Leu Cys Val Ala Cys Gly
405 410 415

Ser Tyr Arg Arg Gly Lys Ala Thr Cys Gly Asp Val Asp Val Leu Ile
420 425 430

Thr His Pro Asp Gly Arg Ser His Arg Gly Ile Phe Ser Arg Leu Leu
435 440 445

Asp Ser Leu Arg Gln Glu Gly Phe Leu Thr Asp Asp Leu Val Ser Gln
450 455 460

Glu Glu Asn Gly Gln Gln Gln Lys Tyr Leu Gly Val Cys Arg Leu Pro
465 470 475 480

Gly Pro Gly Arg Arg His Arg Arg Leu Asp Ile Ile Val Val Pro Tyr
485 490 495

Ser Glu Phe Ala Cys Ala Leu Leu Tyr Phe Thr Gly Ser Ala His Phe
500 505 510

Asn Arg Ser Met Arg Ala Leu Ala Lys Thr Lys Gly Met Ser Leu Ser
515 520 525

Glu His Ala Leu Ser Thr Ala Val Val Arg Asn Thr His Gly Cys Lys
530 535 540

Val Gly Pro Gly Arg Val Leu Pro Thr Pro Thr Glu Lys Asp Val Phe
545 550 555 560

Arg Leu Leu Gly Leu Pro Tyr Arg Glu Pro Ala Glu Arg Asp Trp
565 570 575

<210> 56
<211> 713
<212> PRT
<213> 人工序列
<220>
<223> 合成多肽
<400> 56

Met Ala Thr Gly Gln Asp Arg Val Val Ala Leu Val Asp Met Asp Cys
1 5 10 15

Phe Phe Val Gln Val Glu Gln Arg Gln Asn Pro His Leu Arg Asn Lys
20 25 30

Pro Cys Ala Val Val Gln Tyr Lys Ser Trp Lys Gly Gly Gly Ile Ile

35 40 45

Ala Val Ser Tyr Glu Ala Arg Ala Phe Gly Val Thr Arg Ser Met Trp
50 55 60

Ala Asp Asp Ala Lys Lys Leu Cys Pro Asp Leu Leu Leu Ala Gln Val
65 70 75 80

Arg Glu Ser Arg Gly Lys Ala Asn Leu Thr Lys Tyr Arg Glu Ala Ser
85 90 95

Val Glu Val Met Glu Ile Met Ser Arg Phe Ala Val Ile Glu Arg Ala
100 105 110

Ser Ile Asp Glu Ala Tyr Val Asp Leu Thr Ser Ala Val Gln Glu Arg
115 120 125

Leu Gln Lys Leu Gln Gly Gln Pro Ile Ser Ala Asp Leu Leu Pro Ser
130 135 140

Thr Tyr Ile Glu Gly Leu Pro Gln Gly Pro Thr Thr Ala Glu Glu Thr
145 150 155 160

Val Gln Lys Glu Gly Met Arg Lys Gln Gly Leu Phe Gln Trp Leu Asp
165 170 175

Ser Leu Gln Ile Asp Asn Leu Thr Ser Pro Asp Leu Gln Leu Thr Val
180 185 190

Gly Ala Val Ile Val Glu Glu Met Arg Ala Ala Ile Glu Arg Glu Thr
195 200 205

Gly Phe Gln Cys Ser Ala Gly Ile Ser His Asn Lys Val Leu Ala Lys
210 215 220

Leu Ala Cys Gly Leu Asn Lys Pro Asn Arg Gln Thr Leu Val Ser His
225 230 235 240

Gly Ser Val Pro Gln Leu Phe Ser Gln Met Pro Ile Arg Lys Ile Arg
245 250 255

Ser Leu Gly Gly Lys Leu Gly Ala Ser Val Ile Glu Ile Leu Gly Ile
260 265 270

Glu Tyr Met Gly Glu Leu Thr Gln Phe Thr Glu Ser Gln Leu Gln Ser
275 280 285

His Phe Gly Glu Lys Asn Gly Ser Trp Leu Tyr Ala Met Cys Arg Gly
290 295 300

Ile Glu His Asp Pro Val Lys Pro Arg Gln Leu Pro Lys Thr Ile Gly
305 310 315 320

Cys Ser Lys Asn Phe Pro Gly Lys Thr Ala Leu Ala Thr Arg Glu Gln
325 330 335

Val Gln Trp Trp Leu Leu Gln Leu Ala Gln Glu Leu Glu Glu Arg Leu
340 345 350

Thr Lys Asp Arg Asn Asp Asn Asp Arg Val Ala Thr Gln Leu Val Val
355 360 365
Ser Ile Arg Val Gln Gly Asp Lys Arg Leu Ser Ser Leu Arg Arg Cys
370 375 380
Cys Ala Leu Thr Arg Tyr Asp Ala His Lys Met Ser His Asp Ala Phe
385 390 395 400
Thr Val Ile Lys Asn Cys Asn Thr Ser Gly Ile Gln Thr Glu Trp Ser
405 410 415
Pro Pro Leu Thr Met Leu Phe Leu Cys Ala Thr Lys Phe Ser Ala Ser
420 425 430
Ala Pro Ser Ser Ser Thr Asp Ile Thr Ser Phe Leu Ser Ser Asp Pro
435 440 445
Ser Ser Leu Pro Lys Val Pro Val Thr Ser Ser Glu Ala Lys Thr Gln
450 455 460
Gly Ser Gly Pro Ala Val Thr Ala Thr Lys Lys Ala Thr Thr Ser Leu
465 470 475 480
Glu Ser Phe Phe Gln Lys Ala Ala Glu Arg Gln Lys Val Lys Glu Ala
485 490 495
Ser Leu Ser Ser Leu Thr Ala Pro Thr Gln Ala Pro Met Ser Asn Ser
500 505 510
Pro Ser Lys Pro Ser Leu Pro Phe Gln Thr Ser Gln Ser Thr Gly Thr
515 520 525
Glu Pro Phe Phe Lys Gln Lys Ser Leu Leu Leu Lys Gln Lys Gln Leu
530 535 540
Asn Asn Ser Ser Val Ser Ser Pro Gln Gln Asn Pro Trp Ser Asn Cys
545 550 555 560
Lys Ala Leu Pro Asn Ser Leu Pro Thr Glu Tyr Pro Gly Cys Val Pro
565 570 575
Val Cys Glu Gly Val Ser Lys Leu Glu Glu Ser Ser Lys Ala Thr Pro
580 585 590
Ala Glu Met Asp Leu Ala His Asn Ser Gln Ser Met His Ala Ser Ser
595 600 605
Ala Ser Lys Ser Val Leu Glu Val Thr Gln Lys Ala Thr Pro Asn Pro
610 615 620
Ser Leu Leu Ala Ala Glu Asp Gln Val Pro Cys Glu Lys Cys Gly Ser
625 630 635 640
Leu Val Pro Val Trp Asp Met Pro Glu His Met Asp Tyr His Phe Ala
645 650 655
Leu Glu Leu Gln Lys Ser Phe Leu Gln Pro His Ser Ser Asn Pro Gln

660 665 670
Val Val Ser Ala Val Ser His Gln Gly Lys Arg Asn Pro Lys Ser Pro
675 680 685
Leu Ala Cys Thr Asn Lys Arg Pro Arg Pro Glu Gly Met Gln Thr Leu
690 695 700
Glu Ser Phe Phe Lys Pro Leu Thr His
705 710

<210> 57
<211> 494
<212> PRT
<213> 人工序列
<220>
<223> 合成多肽
<400> 57

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

195 200 205

Ser Ser Arg Val Val Gln Glu Leu Leu Glu His Gly Val Cys Glu Glu
210 215 220

Val Glu Arg Val Arg Arg Ser Glu Arg Tyr Gln Thr Met Lys Leu Phe
225 230 235 240

Thr Gln Ile Phe Gly Val Gly Val Lys Thr Ala Asp Arg Trp Tyr Arg
245 250 255

Glu Gly Leu Arg Thr Leu Asp Asp Leu Arg Glu Gln Pro Gln Lys Leu
260 265 270

Thr Gln Gln Gln Lys Ala Gly Leu Gln His His Gln Asp Leu Ser Thr
275 280 285

Pro Val Leu Arg Ser Asp Val Asp Ala Leu Gln Gln Val Val Glu Glu
290 295 300

Ala Val Gly Gln Ala Leu Pro Gly Ala Thr Val Thr Leu Thr Gly Gly
305 310 315 320

Phe Arg Arg Gly Lys Leu Gln Gly His Asp Val Asp Phe Leu Ile Thr
325 330 335

His Pro Lys Glu Gly Gln Glu Ala Gly Leu Leu Pro Arg Val Met Cys
340 345 350

Arg Leu Gln Asp Gln Gly Leu Ile Leu Tyr His Gln His Gln His Ser
355 360 365

Cys Cys Glu Ser Pro Thr Arg Leu Ala Gln Gln Ser His Met Asp Ala
370 375 380

Phe Glu Arg Ser Phe Cys Ile Phe Arg Leu Pro Gln Pro Pro Gly Ala
385 390 395 400

Ala Val Gly Gly Ser Thr Arg Pro Cys Pro Ser Trp Lys Ala Val Arg
405 410 415

Val Asp Leu Val Val Ala Pro Val Ser Gln Phe Pro Phe Ala Leu Leu
420 425 430

Gly Trp Thr Gly Ser Lys Leu Phe Gln Arg Glu Leu Arg Arg Phe Ser
435 440 445

Arg Lys Glu Lys Gly Leu Trp Leu Asn Ser His Gly Leu Phe Asp Pro
450 455 460

Glu Gln Lys Thr Phe Phe Gln Ala Ala Ser Glu Glu Asp Ile Phe Arg
465 470 475 480

His Leu Gly Leu Glu Tyr Leu Pro Pro Glu Gln Arg Asn Ala
485 490

<210> 58

<211> 740

<212> PRT
<213> 人工序列
<220>
<223> 合成多肽
<400> 58

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

Lys Cys Leu Glu Ala Leu Gly Ile Asn Ser Val Arg Asp Leu Gln Thr
275 280 285
Phe Ser Pro Lys Ile Leu Glu Lys Glu Leu Gly Ile Ser Val Ala Gln
290 295 300
Arg Ile Gln Lys Leu Ser Phe Gly Glu Asp Asn Ser Pro Val Ile Leu
305 310 315 320
Ser Gly Pro Pro Gln Ser Phe Ser Glu Glu Asp Ser Phe Lys Lys Cys
325 330 335
Ser Ser Glu Val Glu Ala Lys Asn Lys Ile Glu Glu Leu Leu Ala Ser
340 345 350
Leu Leu Asn Arg Val Cys Gln Asp Gly Arg Lys Pro His Thr Val Arg
355 360 365
Leu Ile Ile Arg Arg Tyr Ser Ser Glu Lys His Tyr Gly Arg Glu Ser
370 375 380
Arg Gln Cys Pro Ile Pro Ser His Val Ile Gln Lys Leu Gly Thr Gly
385 390 395 400
Asn Tyr Asp Val Met Thr Pro Met Val Asp Ile Leu Met Lys Leu Phe
405 410 415
Arg Asn Met Val Asn Val Lys Met Pro Phe His Leu Thr Leu Leu Ser
420 425 430
Val Cys Phe Cys Asn Leu Lys Ala Leu Asn Thr Ala Lys Lys Gly Leu
435 440 445
Ile Asp Tyr Tyr Leu Met Pro Ser Leu Ser Thr Thr Ser Arg Ser Gly
450 455 460
Lys His Ser Phe Lys Met Lys Asp Thr His Met Glu Asp Phe Pro Lys
465 470 475 480
Asp Lys Glu Thr Asn Arg Asp Phe Leu Pro Ser Gly Arg Ile Glu Ser
485 490 495
Thr Arg Thr Arg Glu Ser Pro Leu Asp Thr Thr Asn Phe Ser Lys Glu
500 505 510
Lys Asp Ile Asn Glu Phe Pro Leu Cys Ser Leu Pro Glu Gly Val Asp
515 520 525
Gln Glu Val Phe Lys Gln Leu Pro Val Asp Ile Gln Glu Glu Ile Leu
530 535 540
Ser Gly Lys Ser Arg Glu Lys Phe Gln Gly Lys Gly Ser Val Ser Cys
545 550 555 560
Pro Leu His Ala Ser Arg Gly Val Leu Ser Phe Phe Ser Lys Lys Gln
565 570 575
Met Gln Asp Ile Pro Ile Asn Pro Arg Asp His Leu Ser Ser Ser Lys

580 585 590
Gln Val Ser Ser Val Ser Pro Cys Glu Pro Gly Thr Ser Gly Phe Asn
595 600 605
Ser Ser Ser Ser Ser Tyr Met Ser Ser Gln Lys Asp Tyr Ser Tyr Tyr
610 615 620
Leu Asp Asn Arg Leu Lys Asp Glu Arg Ile Ser Gln Gly Pro Lys Glu
625 630 635 640
Pro Gln Gly Phe His Phe Thr Asn Ser Asn Pro Ala Val Ser Ala Phe
645 650 655
His Ser Phe Pro Asn Leu Gln Ser Glu Gln Leu Phe Ser Arg Asn His
660 665 670
Thr Thr Asp Ser His Lys Gln Thr Val Ala Thr Asp Ser His Glu Gly
675 680 685
Leu Thr Glu Asn Arg Glu Pro Asp Ser Val Asp Glu Lys Ile Thr Phe
690 695 700
Pro Ser Asp Ile Asp Pro Gln Val Phe Tyr Glu Leu Pro Glu Ala Val
705 710 715 720
Gln Lys Glu Leu Leu Ala Glu Trp Lys Arg Ala Gly Ser Asp Phe His
725 730 735
Ile Gly His Lys
740

<210> 59
<211> 870
<212> PRT
<213> 人工序列
<220>
<223> 合成多肽
<400> 59
Met Asp Ser Thr Lys Glu Lys Cys Asp Ser Tyr Lys Asp Asp Leu Leu
1 5 10 15
Leu Arg Met Gly Leu Asn Asp Asn Lys Ala Gly Met Glu Gly Leu Asp
20 25 30
Lys Glu Lys Ile Asn Lys Ile Ile Met Glu Ala Thr Lys Gly Ser Arg
35 40 45
Phe Tyr Gly Asn Glu Leu Lys Lys Glu Lys Gln Val Asn Gln Arg Ile
50 55 60
Glu Asn Met Met Gln Gln Lys Ala Gln Ile Thr Ser Gln Gln Leu Arg
65 70 75 80
Lys Ala Gln Leu Gln Val Asp Arg Phe Ala Met Glu Leu Glu Gln Ser

85 90 95

Arg Asn Leu Ser Asn Thr Ile Val His Ile Asp Met Asp Ala Phe Tyr
100 105 110

Ala Ala Val Glu Met Arg Asp Asn Pro Glu Leu Lys Asp Lys Pro Ile
115 120 125

Ala Val Gly Ser Met Ser Met Leu Ser Thr Ser Asn Tyr His Ala Arg
130 135 140

Arg Phe Gly Val Arg Ala Ala Met Pro Gly Phe Ile Ala Lys Arg Leu
145 150 155 160

Cys Pro Gln Leu Ile Ile Val Pro Pro Asn Phe Asp Lys Tyr Arg Ala
165 170 175

Val Ser Lys Glu Val Lys Glu Ile Leu Ala Asp Tyr Asp Pro Asn Phe
180 185 190

Met Ala Met Ser Leu Asp Glu Ala Tyr Leu Asn Ile Thr Lys His Leu
195 200 205

Glu Glu Arg Gln Asn Trp Pro Glu Asp Lys Arg Arg Tyr Phe Ile Lys
210 215 220

Met Gly Ser Ser Val Glu Asn Asp Asn Pro Gly Lys Glu Val Asn Lys
225 230 235 240

Leu Ser Glu His Glu Arg Ser Ile Ser Pro Leu Leu Phe Glu Glu Ser
245 250 255

Pro Ser Asp Val Gln Pro Pro Gly Asp Pro Phe Gln Val Asn Phe Glu
260 265 270

Glu Gln Asn Asn Pro Gln Ile Leu Gln Asn Ser Val Val Phe Gly Thr
275 280 285

Ser Ala Gln Glu Val Val Lys Glu Ile Arg Phe Arg Ile Glu Gln Lys
290 295 300

Thr Thr Leu Thr Ala Ser Ala Gly Ile Ala Pro Asn Thr Met Leu Ala
305 310 315 320

Lys Val Cys Ser Asp Lys Asn Lys Pro Asn Gly Gln Tyr Gln Ile Leu
325 330 335

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

Lys Val Ser Gly Ile Gly Lys Val Thr Glu Lys Met Leu Lys Ala Leu
355 360 365

Gly Ile Ile Thr Cys Thr Glu Leu Tyr Gln Gln Arg Ala Leu Leu Ser
370 375 380

Leu Leu Phe Ser Glu Thr Ser Trp His Tyr Phe Leu His Ile Ser Leu
385 390 395 400

Gly Leu Gly Ser Thr His Leu Thr Arg Asp Gly Glu Arg Lys Ser Met
405 410 415
Ser Val Glu Arg Thr Phe Ser Glu Ile Asn Lys Ala Glu Glu Gln Tyr
420 425 430
Ser Leu Cys Gln Glu Leu Cys Ser Glu Leu Ala Gln Asp Leu Gln Lys
435 440 445
Glu Arg Leu Lys Gly Arg Thr Val Thr Ile Lys Leu Lys Asn Val Asn
450 455 460
Phe Glu Val Lys Thr Arg Ala Ser Thr Val Ser Ser Val Val Ser Thr
465 470 475 480
Ala Glu Glu Ile Phe Ala Ile Ala Lys Glu Leu Leu Lys Thr Glu Ile
485 490 495
Asp Ala Asp Phe Pro His Pro Leu Arg Leu Arg Leu Met Gly Val Arg
500 505 510
Ile Ser Ser Phe Pro Asn Glu Glu Asp Arg Lys His Gln Gln Arg Ser
515 520 525
Ile Ile Gly Phe Leu Gln Ala Gly Asn Gln Ala Leu Ser Ala Thr Glu
530 535 540
Cys Thr Leu Glu Lys Thr Asp Lys Asp Lys Phe Val Lys Pro Leu Glu
545 550 555 560
Met Ser His Lys Lys Ser Phe Phe Asp Lys Lys Arg Ser Glu Arg Lys
565 570 575
Trp Ser His Gln Asp Thr Phe Lys Cys Glu Ala Val Asn Lys Gln Ser
580 585 590
Phe Gln Thr Ser Gln Pro Phe Gln Val Leu Lys Lys Lys Met Asn Glu
595 600 605
Asn Leu Glu Ile Ser Glu Asn Ser Asp Asp Cys Gln Ile Leu Thr Cys
610 615 620
Pro Val Cys Phe Arg Ala Gln Gly Cys Ile Ser Leu Glu Ala Leu Asn
625 630 635 640
Lys His Val Asp Glu Cys Leu Asp Gly Pro Ser Ile Ser Glu Asn Phe
645 650 655
Lys Met Phe Ser Cys Ser His Val Ser Ala Thr Lys Val Asn Lys Lys
660 665 670
Glu Asn Val Pro Ala Ser Ser Leu Cys Glu Lys Gln Asp Tyr Glu Ala
675 680 685
His Pro Lys Ile Lys Glu Ile Ser Ser Val Asp Cys Ile Ala Leu Val
690 695 700
Asp Thr Ile Asp Asn Ser Ser Lys Ala Glu Ser Ile Asp Ala Leu Ser

705 710 715 720
Asn Lys His Ser Lys Glu Glu Cys Ser Ser Leu Pro Ser Lys Ser Phe
725 730 735
Asn Ile Glu His Cys His Gln Asn Ser Ser Ser Thr Val Ser Leu Glu
740 745 750
Asn Glu Asp Val Gly Ser Phe Arg Gln Glu Tyr Arg Gln Pro Tyr Leu
755 760 765
Cys Glu Val Lys Thr Gly Gln Ala Leu Val Cys Pro Val Cys Asn Val
770 775 780
Glu Gln Lys Thr Ser Asp Leu Thr Leu Phe Asn Val His Val Asp Val
785 790 795 800
Cys Leu Asn Lys Ser Phe Ile Gln Glu Leu Arg Lys Asp Lys Phe Asn
805 810 815
Pro Val Asn Gln Pro Lys Glu Ser Ser Arg Ser Thr Gly Ser Ser Ser
820 825 830
Gly Val Gln Lys Ala Val Thr Arg Thr Lys Arg Pro Gly Leu Met Thr
835 840 845
Lys Tyr Ser Thr Ser Lys Lys Ile Lys Pro Asn Asn Pro Lys His Thr
850 855 860
Leu Asp Ile Phe Phe Lys
865 870

<210> 60
<211> 1468
<212> PRT
<213> 人工序列
<220>
<223> 合成多肽
<400> 60

Met Ala Pro Val His Gly Asp Asp Cys Glu Ile Gly Ala Ser Ala Leu
1 5 10 15
Ser Asp Ser Gly Ser Phe Val Ser Ser Arg Ala Arg Arg Glu Lys Lys
20 25 30
Ser Lys Lys Gly Arg Gln Glu Ala Leu Glu Arg Leu Lys Lys Ala Lys
35 40 45
Ala Gly Glu Lys Tyr Lys Tyr Glu Val Glu Asp Phe Thr Gly Val Tyr
50 55 60
Glu Glu Val Asp Glu Glu Gln Tyr Ser Lys Leu Val Gln Ala Arg Gln
65 70 75 80
Asp Asp Asp Trp Ile Val Asp Asp Asp Gly Ile Gly Tyr Val Glu Asp

85 90 95

Gly Arg Glu Ile Phe Asp Asp Asp Leu Glu Asp Asp Ala Leu Asp Ala
100 105 110

Asp Glu Lys Gly Lys Asp Gly Lys Ala Arg Asn Lys Asp Lys Arg Asn
115 120 125

Val Lys Lys Leu Ala Val Thr Lys Pro Asn Asn Ile Lys Ser Met Phe
130 135 140

Ile Ala Cys Ala Gly Lys Lys Thr Ala Asp Lys Ala Val Asp Leu Ser
145 150 155 160

Lys Asp Gly Leu Leu Gly Asp Ile Leu Gln Asp Leu Asn Thr Glu Thr
165 170 175

Pro Gln Ile Thr Pro Pro Pro Val Met Ile Leu Lys Lys Lys Arg Ser
180 185 190

Ile Gly Ala Ser Pro Asn Pro Phe Ser Val His Thr Ala Thr Ala Val
195 200 205

Pro Ser Gly Lys Ile Ala Ser Pro Val Ser Arg Lys Glu Pro Pro Leu
210 215 220

Thr Pro Val Pro Leu Lys Arg Ala Glu Phe Ala Gly Asp Asp Val Gln
225 230 235 240

Val Glu Ser Thr Glu Glu Glu Gln Glu Ser Gly Ala Met Glu Phe Glu
245 250 255

Asp Gly Asp Phe Asp Glu Pro Met Glu Val Glu Glu Val Asp Leu Glu
260 265 270

Pro Met Ala Ala Lys Ala Trp Asp Lys Glu Ser Glu Pro Ala Glu Glu
275 280 285

Val Lys Gln Glu Ala Asp Ser Gly Lys Gly Thr Val Ser Tyr Leu Gly
290 295 300

Ser Phe Leu Pro Asp Val Ser Cys Trp Asp Ile Asp Gln Glu Gly Asp
305 310 315 320

Ser Ser Phe Ser Val Gln Glu Val Gln Val Asp Ser Ser His Leu Pro
325 330 335

Leu Val Lys Gly Ala Asp Glu Glu Gln Val Phe His Phe Tyr Trp Leu
340 345 350

Asp Ala Tyr Glu Asp Gln Tyr Asn Gln Pro Gly Val Val Phe Leu Phe
355 360 365

Gly Lys Val Trp Ile Glu Ser Ala Glu Thr His Val Ser Cys Cys Val
370 375 380

Met Val Lys Asn Ile Glu Arg Thr Leu Tyr Phe Leu Pro Arg Glu Met
385 390 395 400

Lys Ile Asp Leu Asn Thr Gly Lys Glu Thr Gly Thr Pro Ile Ser Met
405 410 415
Lys Asp Val Tyr Glu Glu Phe Asp Glu Lys Ile Ala Thr Lys Tyr Lys
420 425 430
Ile Met Lys Phe Lys Ser Lys Pro Val Glu Lys Asn Tyr Ala Phe Glu
435 440 445
Ile Pro Asp Val Pro Glu Lys Ser Glu Tyr Leu Glu Val Lys Tyr Ser
450 455 460
Ala Glu Met Pro Gln Leu Pro Gln Asp Leu Lys Gly Glu Thr Phe Ser
465 470 475 480
His Val Phe Gly Thr Asn Thr Ser Ser Leu Glu Leu Phe Leu Met Asn
485 490 495
Arg Lys Ile Lys Gly Pro Cys Trp Leu Glu Val Lys Ser Pro Gln Leu
500 505 510
Leu Asn Gln Pro Val Ser Trp Cys Lys Val Glu Ala Met Ala Leu Lys
515 520 525
Pro Asp Leu Val Asn Val Ile Lys Asp Val Ser Pro Pro Pro Leu Val
530 535 540
Val Met Ala Phe Ser Met Lys Thr Met Gln Asn Ala Lys Asn His Gln
545 550 555 560
Asn Glu Ile Ile Ala Met Ala Ala Leu Val His His Ser Phe Ala Leu
565 570 575
Asp Lys Ala Ala Pro Lys Pro Pro Phe Gln Ser His Phe Cys Val Val
580 585 590
Ser Lys Pro Lys Asp Cys Ile Phe Pro Tyr Ala Phe Lys Glu Val Ile
595 600 605
Glu Lys Lys Asn Val Lys Val Glu Val Ala Ala Thr Glu Arg Thr Leu
610 615 620
Leu Gly Phe Phe Leu Ala Lys Val His Lys Ile Asp Pro Asp Ile Ile
625 630 635 640
Val Gly His Asn Ile Tyr Gly Phe Glu Leu Glu Val Leu Leu Gln Arg
645 650 655
Ile Asn Val Cys Lys Ala Pro His Trp Ser Lys Ile Gly Arg Leu Lys
660 665 670
Arg Ser Asn Met Pro Lys Leu Gly Gly Arg Ser Gly Phe Gly Glu Arg
675 680 685
Asn Ala Thr Cys Gly Arg Met Ile Cys Asp Val Glu Ile Ser Ala Lys
690 695 700
Glu Leu Ile Arg Cys Lys Ser Tyr His Leu Ser Glu Leu Val Gln Gln

705 710 715 720
Ile Leu Lys Thr Glu Arg Val Val Ile Pro Met Glu Asn Ile Gln Asn
725 730 735
Met Tyr Ser Glu Ser Ser Gln Leu Leu Tyr Leu Leu Glu His Thr Trp
740 745 750
Lys Asp Ala Lys Phe Ile Leu Gln Ile Met Cys Glu Leu Asn Val Leu
755 760 765
Pro Leu Ala Leu Gln Ile Thr Asn Ile Ala Gly Asn Ile Met Ser Arg
770 775 780
Thr Leu Met Gly Gly Arg Ser Glu Arg Asn Glu Phe Leu Leu Leu His
785 790 795 800
Ala Phe Tyr Glu Asn Asn Tyr Ile Val Pro Asp Lys Gln Ile Phe Arg
805 810 815
Lys Pro Gln Gln Lys Leu Gly Asp Glu Asp Glu Glu Ile Asp Gly Asp
820 825 830
Thr Asn Lys Tyr Lys Lys Gly Arg Lys Lys Ala Ala Tyr Ala Gly Gly
835 840 845
Leu Val Leu Asp Pro Lys Val Gly Phe Tyr Asp Lys Phe Ile Leu Leu
850 855 860
Leu Asp Phe Asn Ser Leu Tyr Pro Ser Ile Ile Gln Glu Phe Asn Ile
865 870 875 880
Cys Phe Thr Thr Val Gln Arg Val Ala Ser Glu Ala Gln Lys Val Thr
885 890 895
Glu Asp Gly Glu Gln Glu Gln Ile Pro Glu Leu Pro Asp Pro Ser Leu
900 905 910
Glu Met Gly Ile Leu Pro Arg Glu Ile Arg Lys Leu Val Glu Arg Arg
915 920 925
Lys Gln Val Lys Gln Leu Met Lys Gln Gln Asp Leu Asn Pro Asp Leu
930 935 940
Ile Leu Gln Tyr Asp Ile Arg Gln Lys Ala Leu Lys Leu Thr Ala Asn
945 950 955 960
Ser Met Tyr Gly Cys Leu Gly Phe Ser Tyr Ser Arg Phe Tyr Ala Lys
965 970 975
Pro Leu Ala Ala Leu Val Thr Tyr Lys Gly Arg Glu Ile Leu Met His
980 985 990
Thr Lys Glu Met Val Gln Lys Met Asn Leu Glu Val Ile Tyr Gly Asp
995 1000 1005
Thr Asp Ser Ile Met Ile Asn Thr Asn Ser Thr Asn Leu Glu Glu
1010 1015 1020

Val Phe Lys Leu Gly Asn Lys Val Lys Ser Glu Val Asn Lys Leu
1025 1030 1035
Tyr Lys Leu Leu Glu Ile Asp Ile Asp Gly Val Phe Lys Ser Leu
1040 1045 1050
Leu Leu Leu Lys Lys Lys Lys Tyr Ala Ala Leu Val Val Glu Pro
1055 1060 1065
Thr Ser Asp Gly Asn Tyr Val Thr Lys Gln Glu Leu Lys Gly Leu
1070 1075 1080
Asp Ile Val Arg Arg Asp Trp Cys Asp Leu Ala Lys Asp Thr Gly
1085 1090 1095
Asn Phe Val Ile Gly Gln Ile Leu Ser Asp Gln Ser Arg Asp Thr
1100 1105 1110
Ile Val Glu Asn Ile Gln Lys Arg Leu Ile Glu Ile Gly Glu Asn
1115 1120 1125
Val Leu Asn Gly Ser Val Pro Val Ser Gln Phe Glu Ile Asn Lys
1130 1135 1140
Ala Leu Thr Lys Asp Pro Gln Asp Tyr Pro Asp Lys Lys Ser Leu
1145 1150 1155
Pro His Val His Val Ala Leu Trp Ile Asn Ser Gln Gly Gly Arg
1160 1165 1170
Lys Val Lys Ala Gly Asp Thr Val Ser Tyr Val Ile Cys Gln Asp
1175 1180 1185
Gly Ser Asn Leu Thr Ala Ser Gln Arg Ala Tyr Ala Pro Glu Gln
1190 1195 1200
Leu Gln Lys Gln Asp Asn Leu Thr Ile Asp Thr Gln Tyr Tyr Leu
1205 1210 1215
Ala Gln Gln Ile His Pro Val Val Ala Arg Ile Cys Glu Pro Ile
1220 1225 1230
Asp Gly Ile Asp Ala Val Leu Ile Ala Thr Trp Leu Gly Leu Asp
1235 1240 1245
Pro Thr Gln Phe Arg Val His His Tyr His Lys Asp Glu Glu Asn
1250 1255 1260
Asp Ala Leu Leu Gly Gly Pro Ala Gln Leu Thr Asp Glu Glu Lys
1265 1270 1275
Tyr Arg Asp Cys Glu Arg Phe Lys Cys Pro Cys Pro Thr Cys Gly
1280 1285 1290
Thr Glu Asn Ile Tyr Asp Asn Val Phe Asp Gly Ser Gly Thr Asp
1295 1300 1305
Met Glu Pro Ser Leu Tyr Arg Cys Ser Asn Ile Asp Cys Lys Ala

1310 1315 1320
Ser Pro Leu Thr Phe Thr Val Gln Leu Ser Asn Lys Leu Ile Met
1325 1330 1335
Asp Ile Arg Arg Phe Ile Lys Lys Tyr Tyr Asp Gly Trp Leu Ile
1340 1345 1350
Cys Glu Glu Pro Thr Cys Arg Asn Arg Thr Arg His Leu Pro Leu
1355 1360 1365
Gln Phe Ser Arg Thr Gly Pro Leu Cys Pro Ala Cys Met Lys Ala
1370 1375 1380
Thr Leu Gln Pro Glu Tyr Ser Asp Lys Ser Leu Tyr Thr Gln Leu
1385 1390 1395
Cys Phe Tyr Arg Tyr Ile Phe Asp Ala Glu Cys Ala Leu Glu Lys
1400 1405 1410
Leu Thr Thr Asp His Glu Lys Asp Lys Leu Lys Lys Gln Phe Phe
1415 1420 1425
Thr Pro Lys Val Leu Gln Asp Tyr Arg Lys Leu Lys Asn Thr Ala
1430 1435 1440
Glu Gln Phe Leu Ser Arg Ser Gly Tyr Ser Glu Val Asn Leu Ser
1445 1450 1455
Lys Leu Phe Ala Gly Cys Ala Val Lys Ser
1460 1465

<210> 61
<211> 1107
<212> PRT
<213> 人工序列
<220>
<223> 合成多肽
<400> 61

Met Asp Gly Lys Arg Arg Pro Gly Pro Gly Pro Gly Val Pro Pro Lys
1 5 10 15
Arg Ala Arg Gly Gly Leu Trp Asp Asp Asp Asp Ala Pro Arg Pro Ser
20 25 30
Gln Phe Glu Glu Asp Leu Ala Leu Met Glu Glu Met Glu Ala Glu His
35 40 45
Arg Leu Gln Glu Gln Glu Glu Glu Glu Leu Gln Ser Val Leu Glu Gly
50 55 60
Val Ala Asp Gly Gln Val Pro Pro Ser Ala Ile Asp Pro Arg Trp Leu
65 70 75 80
Arg Pro Thr Pro Pro Ala Leu Asp Pro Gln Thr Glu Pro Leu Ile Phe

85 90 95

Gln Gln Leu Glu Ile Asp His Tyr Val Gly Pro Ala Gln Pro Val Pro
100 105 110

Gly Gly Pro Pro Pro Ser His Gly Ser Val Pro Val Leu Arg Ala Phe
115 120 125

Gly Val Thr Asp Glu Gly Phe Ser Val Cys Cys His Ile His Gly Phe
130 135 140

Ala Pro Tyr Phe Tyr Thr Pro Ala Pro Pro Gly Phe Gly Pro Glu His
145 150 155 160

Met Gly Asp Leu Gln Arg Glu Leu Asn Leu Ala Ile Ser Arg Asp Ser
165 170 175

Arg Gly Gly Arg Glu Leu Thr Gly Pro Ala Val Leu Ala Val Glu Leu
180 185 190

Cys Ser Arg Glu Ser Met Phe Gly Tyr His Gly His Gly Pro Ser Pro
195 200 205

Phe Leu Arg Ile Thr Val Ala Leu Pro Arg Leu Val Ala Pro Ala Arg
210 215 220

Arg Leu Leu Glu Gln Gly Ile Arg Val Ala Gly Leu Gly Thr Pro Ser
225 230 235 240

Phe Ala Pro Tyr Glu Ala Asn Val Asp Phe Glu Ile Arg Phe Met Val
245 250 255

Asp Thr Asp Ile Val Gly Cys Asn Trp Leu Glu Leu Pro Ala Gly Lys
260 265 270

Tyr Ala Leu Arg Leu Lys Glu Lys Ala Thr Gln Cys Gln Leu Glu Ala
275 280 285

Asp Val Leu Trp Ser Asp Val Val Ser His Pro Pro Glu Gly Pro Trp
290 295 300

Gln Arg Ile Ala Pro Leu Arg Val Leu Ser Phe Asp Ile Glu Cys Ala
305 310 315 320

Gly Arg Lys Gly Ile Phe Pro Glu Pro Glu Arg Asp Pro Val Ile Gln
325 330 335

Ile Cys Ser Leu Gly Leu Arg Trp Gly Glu Pro Glu Pro Phe Leu Arg
340 345 350

Leu Ala Leu Thr Leu Arg Pro Cys Ala Pro Ile Leu Gly Ala Lys Val
355 360 365

Gln Ser Tyr Glu Lys Glu Glu Asp Leu Leu Gln Ala Trp Ser Thr Phe
370 375 380

Ile Arg Ile Met Asp Pro Asp Val Ile Thr Gly Tyr Asn Ile Gln Asn
385 390 395 400

Phe Asp Leu Pro Tyr Leu Ile Ser Arg Ala Gln Thr Leu Lys Val Gln
405 410 415
Thr Phe Pro Phe Leu Gly Arg Val Ala Gly Leu Cys Ser Asn Ile Arg
420 425 430
Asp Ser Ser Phe Gln Ser Lys Gln Thr Gly Arg Arg Asp Thr Lys Val
435 440 445
Val Ser Met Val Gly Arg Val Gln Met Asp Met Leu Gln Val Leu Leu
450 455 460
Arg Glu Tyr Lys Leu Arg Ser Tyr Thr Leu Asn Ala Val Ser Phe His
465 470 475 480
Phe Leu Gly Glu Gln Lys Glu Asp Val Gln His Ser Ile Ile Thr Asp
485 490 495
Leu Gln Asn Gly Asn Asp Gln Thr Arg Arg Arg Leu Ala Val Tyr Cys
500 505 510
Leu Lys Asp Ala Tyr Leu Pro Leu Arg Leu Leu Glu Arg Leu Met Val
515 520 525
Leu Val Asn Ala Val Glu Met Ala Arg Val Thr Gly Val Pro Leu Ser
530 535 540
Tyr Leu Leu Ser Arg Gly Gln Gln Val Lys Val Val Ser Gln Leu Leu
545 550 555 560
Arg Gln Ala Met His Glu Gly Leu Leu Met Pro Val Val Lys Ser Glu
565 570 575
Gly Gly Glu Asp Tyr Thr Gly Ala Thr Val Ile Glu Pro Leu Lys Gly
580 585 590
Tyr Tyr Asp Val Pro Ile Ala Thr Leu Asp Phe Ser Ser Leu Tyr Pro
595 600 605
Ser Ile Met Met Ala His Asn Leu Cys Tyr Thr Thr Leu Leu Arg Pro
610 615 620
Gly Thr Ala Gln Lys Leu Gly Leu Thr Glu Asp Gln Phe Ile Arg Thr
625 630 635 640
Pro Thr Gly Asp Glu Phe Val Lys Thr Ser Val Arg Lys Gly Leu Leu
645 650 655
Pro Gln Ile Leu Glu Asn Leu Leu Ser Ala Arg Lys Arg Ala Lys Ala
660 665 670
Glu Leu Ala Lys Glu Thr Asp Pro Leu Arg Arg Gln Val Leu Asp Gly
675 680 685
Arg Gln Leu Ala Leu Lys Val Ser Ala Asn Ser Val Tyr Gly Phe Thr
690 695 700
Gly Ala Gln Val Gly Lys Leu Pro Cys Leu Glu Ile Ser Gln Ser Val

705 710 715 720
Thr Gly Phe Gly Arg Gln Met Ile Glu Lys Thr Lys Gln Leu Val Glu
725 730 735
Ser Lys Tyr Thr Val Glu Asn Gly Tyr Ser Thr Ser Ala Lys Val Val
740 745 750
Tyr Gly Asp Thr Asp Ser Val Met Cys Arg Phe Gly Val Ser Ser Val
755 760 765
Ala Glu Ala Met Ala Leu Gly Arg Glu Ala Ala Asp Trp Val Ser Gly
770 775 780
His Phe Pro Ser Pro Ile Arg Leu Glu Phe Glu Lys Val Tyr Phe Pro
785 790 795 800
Tyr Leu Leu Ile Ser Lys Lys Arg Tyr Ala Gly Leu Leu Phe Ser Ser
805 810 815
Arg Pro Asp Ala His Asp Arg Met Asp Cys Lys Gly Leu Glu Ala Val
820 825 830
Arg Arg Asp Asn Cys Pro Leu Val Ala Asn Leu Val Thr Ala Ser Leu
835 840 845
Arg Arg Leu Leu Ile Asp Arg Asp Pro Glu Gly Ala Val Ala His Ala
850 855 860
Gln Asp Val Ile Ser Asp Leu Leu Cys Asn Arg Ile Asp Ile Ser Gln
865 870 875 880
Leu Val Ile Thr Lys Glu Leu Thr Arg Ala Ala Ser Asp Tyr Ala Gly
885 890 895
Lys Gln Ala His Val Glu Leu Ala Glu Arg Met Arg Lys Arg Asp Pro
900 905 910
Gly Ser Ala Pro Ser Leu Gly Asp Arg Val Pro Tyr Val Ile Ile Ser
915 920 925
Ala Ala Lys Gly Val Ala Ala Tyr Met Lys Ser Glu Asp Pro Leu Phe
930 935 940
Val Leu Glu His Ser Leu Pro Ile Asp Thr Gln Tyr Tyr Leu Glu Gln
945 950 955 960
Gln Leu Ala Lys Pro Leu Leu Arg Ile Phe Glu Pro Ile Leu Gly Glu
965 970 975
Gly Arg Ala Glu Ala Val Leu Leu Arg Gly Asp His Thr Arg Cys Lys
980 985 990
Thr Val Leu Thr Gly Lys Val Gly Gly Leu Leu Ala Phe Ala Lys Arg
995 1000 1005
Arg Asn Cys Cys Ile Gly Cys Arg Thr Val Leu Ser His Gln Gly
1010 1015 1020

Ala Val Cys Glu Phe Cys Gln Pro Arg Glu Ser Glu Leu Tyr Gln
1025 1030 1035
Lys Glu Val Ser His Leu Asn Ala Leu Glu Glu Arg Phe Ser Arg
1040 1045 1050
Leu Trp Thr Gln Cys Gln Arg Cys Gln Gly Ser Leu His Glu Asp
1055 1060 1065
Val Ile Cys Thr Ser Arg Asp Cys Pro Ile Phe Tyr Met Arg Lys
1070 1075 1080
Lys Val Arg Lys Asp Leu Glu Asp Gln Glu Gln Leu Leu Arg Arg
1085 1090 1095
Phe Gly Pro Pro Gly Pro Glu Ala Trp
1100 1105

<210> 62
<211> 1239
<212> PRT
<213> 人工序列
<220>
<223> 合成多肽
<400> 62

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

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

465 470 475 480
Glu Asp Pro Trp Leu Trp Asp Leu Glu Trp Asp Leu Gln Glu Phe Lys
485 490 495
Gln Lys Lys Ala Lys Lys Val Lys Lys Glu Pro Ala Thr Ala Ser Lys
500 505 510
Leu Pro Ile Glu Gly Ala Gly Ala Pro Gly Asp Pro Met Asp Gln Glu
515 520 525
Asp Leu Gly Pro Cys Ser Glu Glu Glu Glu Phe Gln Gln Asp Val Met
530 535 540
Ala Arg Ala Cys Leu Gln Lys Leu Lys Gly Thr Thr Glu Leu Leu Pro
545 550 555 560
Lys Arg Pro Gln His Leu Pro Gly His Pro Gly Trp Tyr Arg Lys Leu
565 570 575
Cys Pro Arg Leu Asp Asp Pro Ala Trp Thr Pro Gly Pro Ser Leu Leu
580 585 590
Ser Leu Gln Met Arg Val Thr Pro Lys Leu Met Ala Leu Thr Trp Asp
595 600 605
Gly Phe Pro Leu His Tyr Ser Glu Arg His Gly Trp Gly Tyr Leu Val
610 615 620
Pro Gly Arg Arg Asp Asn Leu Ala Lys Leu Pro Thr Gly Thr Thr Leu
625 630 635 640
Glu Ser Ala Gly Val Val Cys Pro Tyr Arg Ala Ile Glu Ser Leu Tyr
645 650 655
Arg Lys His Cys Leu Glu Gln Gly Lys Gln Gln Leu Met Pro Gln Glu
660 665 670
Ala Gly Leu Ala Glu Glu Phe Leu Leu Thr Asp Asn Ser Ala Ile Trp
675 680 685
Gln Thr Val Glu Glu Leu Asp Tyr Leu Glu Val Glu Ala Glu Ala Lys
690 695 700
Met Glu Asn Leu Arg Ala Ala Val Pro Gly Gln Pro Leu Ala Leu Thr
705 710 715 720
Ala Arg Gly Gly Pro Lys Asp Thr Gln Pro Ser Tyr His His Gly Asn
725 730 735
Gly Pro Tyr Asn Asp Val Asp Ile Pro Gly Cys Trp Phe Phe Lys Leu
740 745 750
Pro His Lys Asp Gly Asn Ser Cys Asn Val Gly Ser Pro Phe Ala Lys
755 760 765
Asp Phe Leu Pro Lys Met Glu Asp Gly Thr Leu Gln Ala Gly Pro Gly
770 775 780

Gly Ala Ser Gly Pro Arg Ala Leu Glu Ile Asn Lys Met Ile Ser Phe
785 790 795 800
Trp Arg Asn Ala His Lys Arg Ile Ser Ser Gln Met Val Val Trp Leu
805 810 815
Pro Arg Ser Ala Leu Pro Arg Ala Val Ile Arg His Pro Asp Tyr Asp
820 825 830
Glu Glu Gly Leu Tyr Gly Ala Ile Leu Pro Gln Val Val Thr Ala Gly
835 840 845
Thr Ile Thr Arg Arg Ala Val Glu Pro Thr Trp Leu Thr Ala Ser Asn
850 855 860
Ala Arg Pro Asp Arg Val Gly Ser Glu Leu Lys Ala Met Val Gln Ala
865 870 875 880
Pro Pro Gly Tyr Thr Leu Val Gly Ala Asp Val Asp Ser Gln Glu Leu
885 890 895
Trp Ile Ala Ala Val Leu Gly Asp Ala His Phe Ala Gly Met His Gly
900 905 910
Cys Thr Ala Phe Gly Trp Met Thr Leu Gln Gly Arg Lys Ser Arg Gly
915 920 925
Thr Asp Leu His Ser Lys Thr Ala Thr Thr Val Gly Ile Ser Arg Glu
930 935 940
His Ala Lys Ile Phe Asn Tyr Gly Arg Ile Tyr Gly Ala Gly Gln Pro
945 950 955 960
Phe Ala Glu Arg Leu Leu Met Gln Phe Asn His Arg Leu Thr Gln Gln
965 970 975
Glu Ala Ala Glu Lys Ala Gln Gln Met Tyr Ala Ala Thr Lys Gly Leu
980 985 990
Arg Trp Tyr Arg Leu Ser Asp Glu Gly Glu Trp Leu Val Arg Glu Leu
995 1000 1005
Asn Leu Pro Val Asp Arg Thr Glu Gly Gly Trp Ile Ser Leu Gln
1010 1015 1020
Asp Leu Arg Lys Val Gln Arg Glu Thr Ala Arg Lys Ser Gln Trp
1025 1030 1035
Lys Lys Trp Glu Val Val Ala Glu Arg Ala Trp Lys Gly Gly Thr
1040 1045 1050
Glu Ser Glu Met Phe Asn Lys Leu Glu Ser Ile Ala Thr Ser Asp
1055 1060 1065
Ile Pro Arg Thr Pro Val Leu Gly Cys Cys Ile Ser Arg Ala Leu
1070 1075 1080
Glu Pro Ser Ala Val Gln Glu Glu Phe Met Thr Ser Arg Val Asn

1085 1090 1095
Trp Val Val Gln Ser Ser Ala Val Asp Tyr Leu His Leu Met Leu
1100 1105 1110
Val Ala Met Lys Trp Leu Phe Glu Glu Phe Ala Ile Asp Gly Arg
1115 1120 1125
Phe Cys Ile Ser Ile His Asp Glu Val Arg Tyr Leu Val Arg Glu
1130 1135 1140
Glu Asp Arg Tyr Arg Ala Ala Leu Ala Leu Gln Ile Thr Asn Leu
1145 1150 1155
Leu Thr Arg Cys Met Phe Ala Tyr Lys Leu Gly Leu Asn Asp Leu
1160 1165 1170
Pro Gln Ser Val Ala Phe Phe Ser Ala Val Asp Ile Asp Arg Cys
1175 1180 1185
Leu Arg Lys Glu Val Thr Met Asp Cys Lys Thr Pro Ser Asn Pro
1190 1195 1200
Thr Gly Met Glu Arg Arg Tyr Gly Ile Pro Gln Gly Glu Ala Leu
1205 1210 1215
Asp Ile Tyr Gln Ile Ile Glu Leu Thr Lys Gly Ser Leu Glu Lys
1220 1225 1230
Arg Ser Gln Pro Gly Pro
1235

<210> 63
<211> 900
<212> PRT
<213> 人工序列
<220>
<223> 合成多肽
<400> 63

Met Glu Asn Tyr Glu Ala Leu Val Gly Phe Asp Leu Cys Asn Thr Pro
1 5 10 15
Leu Ser Ser Val Ala Gln Lys Ile Met Ser Ala Met His Ser Gly Asp
20 25 30
Leu Val Asp Ser Lys Thr Trp Gly Lys Ser Thr Glu Thr Met Glu Val
35 40 45
Ile Asn Lys Ser Ser Val Lys Tyr Ser Val Gln Leu Glu Asp Arg Lys
50 55 60
Thr Gln Ser Pro Glu Lys Lys Asp Leu Lys Ser Leu Arg Ser Gln Thr
65 70 75 80
Ser Arg Gly Ser Ala Lys Leu Ser Pro Gln Ser Phe Ser Val Arg Leu

85 90 95

Thr Asp Gln Leu Ser Ala Asp Gln Lys Gln Lys Ser Ile Ser Ser Leu
100 105 110

Thr Leu Ser Ser Cys Leu Ile Pro Gln Tyr Asn Gln Glu Ala Ser Val
115 120 125

Leu Gln Lys Lys Gly His Lys Arg Lys His Phe Leu Met Glu Asn Ile
130 135 140

Asn Asn Glu Asn Lys Gly Ser Ile Asn Leu Lys Arg Lys His Ile Thr
145 150 155 160

Tyr Asn Asn Leu Ser Glu Lys Thr Ser Lys Gln Met Ala Leu Glu Glu
165 170 175

Asp Thr Asp Asp Ala Glu Gly Tyr Leu Asn Ser Gly Asn Ser Gly Ala
180 185 190

Leu Lys Lys His Phe Cys Asp Ile Arg His Leu Asp Asp Trp Ala Lys
195 200 205

Ser Gln Leu Ile Glu Met Leu Lys Gln Ala Ala Ala Leu Val Ile Thr
210 215 220

Val Met Tyr Thr Asp Gly Ser Thr Gln Leu Gly Ala Asp Gln Thr Pro
225 230 235 240

Val Ser Ser Val Arg Gly Ile Val Val Leu Val Lys Arg Gln Ala Glu
245 250 255

Gly Gly His Gly Cys Pro Asp Ala Pro Ala Cys Gly Pro Val Leu Glu
260 265 270

Gly Phe Val Ser Asp Asp Pro Cys Ile Tyr Ile Gln Ile Glu His Ser
275 280 285

Ala Ile Trp Asp Gln Glu Gln Glu Ala His Gln Gln Phe Ala Arg Asn
290 295 300

Val Leu Phe Gln Thr Met Lys Cys Lys Cys Pro Val Ile Cys Phe Asn
305 310 315 320

Ala Lys Asp Phe Val Arg Ile Val Leu Gln Phe Phe Gly Asn Asp Gly
325 330 335

Ser Trp Lys His Val Ala Asp Phe Ile Gly Leu Asp Pro Arg Ile Ala
340 345 350

Ala Trp Leu Ile Asp Pro Ser Asp Ala Thr Pro Ser Phe Glu Asp Leu
355 360 365

Val Glu Lys Tyr Cys Glu Lys Ser Ile Thr Val Lys Val Asn Ser Thr
370 375 380

Tyr Gly Asn Ser Ser Arg Asn Ile Val Asn Gln Asn Val Arg Glu Asn
385 390 395 400

Leu Lys Thr Leu Tyr Arg Leu Thr Met Asp Leu Cys Ser Lys Leu Lys
405 410 415
Asp Tyr Gly Leu Trp Gln Leu Phe Arg Thr Leu Glu Leu Pro Leu Ile
420 425 430
Pro Ile Leu Ala Val Met Glu Ser His Ala Ile Gln Val Asn Lys Glu
435 440 445
Glu Met Glu Lys Thr Ser Ala Leu Leu Gly Ala Arg Leu Lys Glu Leu
450 455 460
Glu Gln Glu Ala His Phe Val Ala Gly Glu Arg Phe Leu Ile Thr Ser
465 470 475 480
Asn Asn Gln Leu Arg Glu Ile Leu Phe Gly Lys Leu Lys Leu His Leu
485 490 495
Leu Ser Gln Arg Asn Ser Leu Pro Arg Thr Gly Leu Gln Lys Tyr Pro
500 505 510
Ser Thr Ser Glu Ala Val Leu Asn Ala Leu Arg Asp Leu His Pro Leu
515 520 525
Pro Lys Ile Ile Leu Glu Tyr Arg Gln Val His Lys Ile Lys Ser Thr
530 535 540
Phe Val Asp Gly Leu Leu Ala Cys Met Lys Lys Gly Ser Ile Ser Ser
545 550 555 560
Thr Trp Asn Gln Thr Gly Thr Val Thr Gly Arg Leu Ser Ala Lys His
565 570 575
Pro Asn Ile Gln Gly Ile Ser Lys His Pro Ile Gln Ile Thr Thr Pro
580 585 590
Lys Asn Phe Lys Gly Lys Glu Asp Lys Ile Leu Thr Ile Ser Pro Arg
595 600 605
Ala Met Phe Val Ser Ser Lys Gly His Thr Phe Leu Ala Ala Asp Phe
610 615 620
Ser Gln Ile Glu Leu Arg Ile Leu Thr His Leu Ser Gly Asp Pro Glu
625 630 635 640
Leu Leu Lys Leu Phe Gln Glu Ser Glu Arg Asp Asp Val Phe Ser Thr
645 650 655
Leu Thr Ser Gln Trp Lys Asp Val Pro Val Glu Gln Val Thr His Ala
660 665 670
Asp Arg Glu Gln Thr Lys Lys Val Val Tyr Ala Val Val Tyr Gly Ala
675 680 685
Gly Lys Glu Arg Leu Ala Ala Cys Leu Gly Val Pro Ile Gln Glu Ala
690 695 700
Ala Gln Phe Leu Glu Ser Phe Leu Gln Lys Tyr Lys Lys Ile Lys Asp

705 710 715 720
Phe Ala Arg Ala Ala Ile Ala Gln Cys His Gln Thr Gly Cys Val Val
725 730 735
Ser Ile Met Gly Arg Arg Arg Pro Leu Pro Arg Ile His Ala His Asp
740 745 750
Gln Gln Leu Arg Ala Gln Ala Glu Arg Gln Ala Val Asn Phe Val Val
755 760 765
Gln Gly Ser Ala Ala Asp Leu Cys Lys Leu Ala Met Ile His Val Phe
770 775 780
Thr Ala Val Ala Ala Ser His Thr Leu Thr Ala Arg Leu Val Ala Gln
785 790 795 800
Ile His Asp Glu Leu Leu Phe Glu Val Glu Asp Pro Gln Ile Pro Glu
805 810 815
Cys Ala Ala Leu Val Arg Arg Thr Met Glu Ser Leu Glu Gln Val Gln
820 825 830
Ala Leu Glu Leu Gln Leu Gln Val Pro Leu Lys Val Ser Leu Ser Ala
835 840 845
Gly Arg Ser Trp Gly His Leu Val Pro Leu Gln Glu Ala Trp Gly Pro
850 855 860
Pro Pro Gly Pro Cys Arg Thr Glu Ser Pro Ser Asn Ser Leu Ala Ala
865 870 875 880
Pro Gly Ser Pro Ala Ser Thr Gln Pro Pro Pro Leu His Phe Ser Pro
885 890 895
Ser Phe Cys Leu
900

<210> 64
<211> 1251
<212> PRT
<213> 人工序列
<220>
<223> 合成多肽
<400> 64
Met Arg Arg Gly Gly Trp Arg Lys Arg Ala Glu Asn Asp Gly Trp Glu
1 5 10 15
Thr Trp Gly Gly Tyr Met Ala Ala Lys Val Gln Lys Leu Glu Glu Gln
20 25 30
Phe Arg Ser Asp Ala Ala Met Gln Lys Asp Gly Thr Ser Ser Thr Ile
35 40 45
Phe Ser Gly Val Ala Ile Tyr Val Asn Gly Tyr Thr Asp Pro Ser Ala

50 55 60

Glu Glu Leu Arg Lys Leu Met Met Leu His Gly Gly Gln Tyr His Val
65 70 75 80

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

Asn Ala Lys Ile Lys Glu Leu Lys Gly Glu Lys Val Ile Arg Pro Glu
100 105 110

Trp Ile Val Glu Ser Ile Lys Ala Gly Arg Leu Leu Ser Tyr Ile Pro
115 120 125

Tyr Gln Leu Tyr Thr Lys Gln Ser Ser Val Gln Lys Gly Leu Ser Phe
130 135 140

Asn Pro Val Cys Arg Pro Glu Asp Pro Leu Pro Gly Pro Ser Asn Ile
145 150 155 160

Ala Lys Gln Leu Asn Asn Arg Val Asn His Ile Val Lys Lys Ile Glu
165 170 175

Thr Glu Asn Glu Val Lys Val Asn Gly Met Asn Ser Trp Asn Glu Glu
180 185 190

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

Gly Arg Lys Gln Asn Gly Ile Pro His Pro Arg Gly Ser Thr Ala Ile
210 215 220

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

Cys Leu Val Pro Met Val Asn Ser Val Ala Ser Arg Leu Ser Pro Ala
245 250 255

Phe Ser Gln Glu Glu Asp Lys Ala Glu Lys Ser Ser Thr Asp Phe Arg
260 265 270

Asp Cys Thr Leu Gln Gln Leu Gln Gln Ser Thr Arg Asn Thr Asp Ala
275 280 285

Leu Arg Asn Pro His Arg Thr Asn Ser Phe Ser Leu Ser Pro Leu His
290 295 300

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

Ser Ser Thr Lys Ser Thr Ser Ser Val Ser Thr Phe Ser Lys Ala Ala
325 330 335

Pro Ser Val Pro Ser Lys Pro Ser Asp Cys Asn Phe Ile Ser Asn Phe
340 345 350

Tyr Ser His Ser Arg Leu His His Ile Ser Met Trp Lys Cys Glu Leu
355 360 365

Thr Glu Phe Val Asn Thr Leu Gln Arg Gln Ser Asn Gly Ile Phe Pro
370 375 380
Gly Arg Glu Lys Leu Lys Lys Met Lys Thr Gly Arg Ser Ala Leu Val
385 390 395 400
Val Thr Asp Thr Gly Asp Met Ser Val Leu Asn Ser Pro Arg His Gln
405 410 415
Ser Cys Ile Met His Val Asp Met Asp Cys Phe Phe Val Ser Val Gly
420 425 430
Ile Arg Asn Arg Pro Asp Leu Lys Gly Lys Pro Val Ala Val Thr Ser
435 440 445
Asn Arg Gly Thr Gly Arg Ala Pro Leu Arg Pro Gly Ala Asn Pro Gln
450 455 460
Leu Glu Trp Gln Tyr Tyr Gln Asn Lys Ile Leu Lys Gly Lys Ala Ala
465 470 475 480
Asp Ile Pro Asp Ser Ser Leu Trp Glu Asn Pro Asp Ser Ala Gln Ala
485 490 495
Asn Gly Ile Asp Ser Val Leu Ser Arg Ala Glu Ile Ala Ser Cys Ser
500 505 510
Tyr Glu Ala Arg Gln Leu Gly Ile Lys Asn Gly Met Phe Phe Gly His
515 520 525
Ala Lys Gln Leu Cys Pro Asn Leu Gln Ala Val Pro Tyr Asp Phe His
530 535 540
Ala Tyr Lys Glu Val Ala Gln Thr Leu Tyr Glu Thr Leu Ala Ser Tyr
545 550 555 560
Thr His Asn Ile Glu Ala Val Ser Cys Asp Glu Ala Leu Val Asp Ile
565 570 575
Thr Glu Ile Leu Ala Glu Thr Lys Leu Thr Pro Asp Glu Phe Ala Asn
580 585 590
Ala Val Arg Met Glu Ile Lys Asp Gln Thr Lys Cys Ala Ala Ser Val
595 600 605
Gly Ile Gly Ser Asn Ile Leu Leu Ala Arg Met Ala Thr Arg Lys Ala
610 615 620
Lys Pro Asp Gly Gln Tyr His Leu Lys Pro Glu Glu Val Asp Asp Phe
625 630 635 640
Ile Arg Gly Gln Leu Val Thr Asn Leu Pro Gly Val Gly His Ser Met
645 650 655
Glu Ser Lys Leu Ala Ser Leu Gly Ile Lys Thr Cys Gly Asp Leu Gln
660 665 670
Tyr Met Thr Met Ala Lys Leu Gln Lys Glu Phe Gly Pro Lys Thr Gly

675 680 685

Gln Met Leu Tyr Arg Phe Cys Arg Gly Leu Asp Asp Arg Pro Val Arg

690 695 700

Thr Glu Lys Glu Arg Lys Ser Val Ser Ala Glu Ile Asn Tyr Gly Ile

705 710 715 720

Arg Phe Thr Gln Pro Lys Glu Ala Glu Ala Phe Leu Leu Ser Leu Ser

725 730 735

Glu Glu Ile Gln Arg Arg Leu Glu Ala Thr Gly Met Lys Gly Lys Arg

740 745 750

Leu Thr Leu Lys Ile Met Val Arg Lys Pro Gly Ala Pro Val Glu Thr

755 760 765

Ala Lys Phe Gly Gly His Gly Ile Cys Asp Asn Ile Ala Arg Thr Val

770 775 780

Thr Leu Asp Gln Ala Thr Asp Asn Ala Lys Ile Ile Gly Lys Ala Met

785 790 795 800

Leu Asn Met Phe His Thr Met Lys Leu Asn Ile Ser Asp Met Arg Gly

805 810 815

Val Gly Ile His Val Asn Gln Leu Val Pro Thr Asn Leu Asn Pro Ser

820 825 830

Thr Cys Pro Ser Arg Pro Ser Val Gln Ser Ser His Phe Pro Ser Gly

835 840 845

Ser Tyr Ser Val Arg Asp Val Phe Gln Val Gln Lys Ala Lys Lys Ser

850 855 860

Thr Glu Glu Glu His Lys Glu Val Phe Arg Ala Ala Val Asp Leu Glu

865 870 875 880

Ile Ser Ser Ala Ser Arg Thr Cys Thr Phe Leu Pro Pro Phe Pro Ala

885 890 895

His Leu Pro Thr Ser Pro Asp Thr Asn Lys Ala Glu Ser Ser Gly Lys

900 905 910

Trp Asn Gly Leu His Thr Pro Val Ser Val Gln Ser Arg Leu Asn Leu

915 920 925

Ser Ile Glu Val Pro Ser Pro Ser Gln Leu Asp Gln Ser Val Leu Glu

930 935 940

Ala Leu Pro Pro Asp Leu Arg Glu Gln Val Glu Gln Val Cys Ala Val

945 950 955 960

Gln Gln Ala Glu Ser His Gly Asp Lys Lys Lys Glu Pro Val Asn Gly

965 970 975

Cys Asn Thr Gly Ile Leu Pro Gln Pro Val Gly Thr Val Leu Leu Gln

980 985 990

Ile Pro Glu Pro Gln Glu Ser Asn Ser Asp Ala Gly Ile Asn Leu Ile
995 1000 1005
Ala Leu Pro Ala Phe Ser Gln Val Asp Pro Glu Val Phe Ala Ala
1010 1015 1020
Leu Pro Ala Glu Leu Gln Arg Glu Leu Lys Ala Ala Tyr Asp Gln
1025 1030 1035
Arg Gln Arg Gln Gly Glu Asn Ser Thr His Gln Gln Ser Ala Ser
1040 1045 1050
Ala Ser Val Pro Lys Asn Pro Leu Leu His Leu Lys Ala Ala Val
1055 1060 1065
Lys Glu Lys Lys Arg Asn Lys Lys Lys Lys Thr Ile Gly Ser Pro
1070 1075 1080
Lys Arg Ile Gln Ser Pro Leu Asn Asn Lys Leu Leu Asn Ser Pro
1085 1090 1095
Ala Lys Thr Leu Pro Gly Ala Cys Gly Ser Pro Gln Lys Leu Ile
1100 1105 1110
Asp Gly Phe Leu Lys His Glu Gly Pro Pro Ala Glu Lys Pro Leu
1115 1120 1125
Glu Glu Leu Ser Ala Ser Thr Ser Gly Val Pro Gly Leu Ser Ser
1130 1135 1140
Leu Gln Ser Asp Pro Ala Gly Cys Val Arg Pro Pro Ala Pro Asn
1145 1150 1155
Leu Ala Gly Ala Val Glu Phe Asn Asp Val Lys Thr Leu Leu Arg
1160 1165 1170
Glu Trp Ile Thr Thr Ile Ser Asp Pro Met Glu Glu Asp Ile Leu
1175 1180 1185
Gln Val Val Lys Tyr Cys Thr Asp Leu Ile Glu Glu Lys Asp Leu
1190 1195 1200
Glu Lys Leu Asp Leu Val Ile Lys Tyr Met Lys Arg Leu Met Gln
1205 1210 1215
Gln Ser Val Glu Ser Val Trp Asn Met Ala Phe Asp Phe Ile Leu
1220 1225 1230
Asp Asn Val Gln Val Val Leu Gln Gln Thr Tyr Gly Ser Thr Leu
1235 1240 1245
Lys Val Thr
1250

<210> 65
<211> 313
<212> PRT

<213> 人工序列

<220>

<223> 合成多肽

<400> 65

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

275 280 285

Gly Cys Arg His Phe Ser Lys Thr Asn Glu Leu Leu Gln Lys Ser Gly
290 295 300

Lys Lys Pro Ile Asp Trp Lys Glu Leu
305 310

<210> 66

<211> 313

<212> PRT

<213> 人工序列

<220>

<223> 合成多肽

<400> 66

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

Arg Lys Arg His Ala Pro Ser Pro Glu Pro Ala Val Gln Gly Thr Gly
20 25 30

Val Ala Gly Val Pro Glu Glu Ser Gly Asp Ala Ala Ala Ile Pro Ala
35 40 45

Lys Lys Ala Pro Ala Gly Gln Glu Glu Pro Gly Thr Pro Pro Ser Ser
50 55 60

Pro Leu Ser Ala Glu Gln Leu Asp Arg Ile Gln Arg Asn Lys Ala Ala
65 70 75 80

Ala Leu Leu Arg Leu Ala Ala Arg Asn Val Pro Val Gly Phe Gly Glu
85 90 95

Ser Trp Lys Lys His Leu Ser Gly Glu Phe Gly Lys Pro Tyr Phe Ile
100 105 110

Lys Leu Met Gly Phe Val Ala Glu Glu Arg Lys His Tyr Thr Val Tyr
115 120 125

Pro Pro Pro His Gln Val Phe Thr Trp Thr Gln Met Cys Asp Ile Lys
130 135 140

Asp Val Lys Val Val Ile Leu Gly Gln Asp Pro Tyr His Gly Pro Asn
145 150 155 160

Gln Ala His Gly Leu Cys Phe Ser Val Gln Arg Pro Val Pro Pro Pro
165 170 175

Pro Ser Leu Glu Asn Ile Tyr Lys Glu Leu Ser Thr Asp Ile Glu Asp
180 185 190

Phe Val His Pro Gly His Gly Asp Leu Ser Gly Trp Ala Lys Gln Gly
195 200 205

Val Leu Leu Leu Asp Ala Val Leu Thr Val Arg Ala His Gln Ala Asn

210 215 220
Ser His Lys Glu Arg Gly Trp Glu Gln Phe Thr Asp Ala Val Val Ser
225 230 235 240
Trp Leu Asn Gln Asn Ser Asn Gly Leu Val Phe Leu Leu Trp Gly Ser
245 250 255
Tyr Ala Gln Lys Lys Gly Ser Ala Ile Asp Arg Lys Arg His His Val
260 265 270
Leu Gln Thr Ala His Pro Ser Pro Leu Ser Val Tyr Arg Gly Phe Phe
275 280 285
Gly Cys Arg His Phe Ser Lys Thr Asn Glu Leu Leu Gln Lys Ser Gly
290 295 300
Lys Lys Pro Ile Asp Trp Lys Glu Leu
305 310

<210> 67
<211> 198
<212> PRT
<213> 智人
<400> 67

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

165 170 175
Arg Arg Ile Leu Leu Pro Leu Tyr Glu Val Asp Asp Leu Arg Asp Ala
180 185 190
Phe Arg Thr Leu Gly Leu
195
<210> 68
<211> 198
<212> PRT
<213> 小家鼠
<400> 68
Met Asp Ser Leu Leu Met Lys Gln Lys Lys Phe Leu Tyr His Phe Lys
1 5 10 15
Asn Val Arg Trp Ala Lys Gly Arg His Glu Thr Tyr Leu Cys Tyr Val
20 25 30
Val Lys Arg Arg Asp Ser Ala Thr Ser Cys Ser Leu Asp Phe Gly His
35 40 45
Leu Arg Asn Lys Ser Gly Cys His Val Glu Leu Leu Phe Leu Arg Tyr
50 55 60
Ile Ser Asp Trp Asp Leu Asp Pro Gly Arg Cys Tyr Arg Val Thr Trp
65 70 75 80
Phe Thr Ser Trp Ser Pro Cys Tyr Asp Cys Ala Arg His Val Ala Glu
85 90 95
Phe Leu Arg Trp Asn Pro Asn Leu Ser Leu Arg Ile Phe Thr Ala Arg
100 105 110
Leu Tyr Phe Cys Glu Asp Arg Lys Ala Glu Pro Glu Gly Leu Arg Arg
115 120 125
Leu His Arg Ala Gly Val Gln Ile Gly Ile Met Thr Phe Lys Asp Tyr
130 135 140
Phe Tyr Cys Trp Asn Thr Phe Val Glu Asn Arg Glu Arg Thr Phe Lys
145 150 155 160
Ala Trp Glu Gly Leu His Glu Asn Ser Val Arg Leu Thr Arg Gln Leu
165 170 175
Arg Arg Ile Leu Leu Pro Leu Tyr Glu Val Asp Asp Leu Arg Asp Ala
180 185 190
Phe Arg Met Leu Gly Phe
195
<210> 69
<211> 197
<212> PRT

<213> 家犬(Canis lupus familiaris)

<400> 69

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

<210> 70

<211> 198

<212> PRT

<213> 牛(Bos taurus)

<400> 70

Met Asp Ser Leu Leu Lys Lys Gln Arg Gln Phe Leu Tyr Gln Phe Lys
1 5 10 15
Asn Val Arg Trp Ala Lys Gly Arg His Glu Thr Tyr Leu Cys Tyr Val
20 25 30
Val Lys Arg Arg Asp Ser Pro Thr Ser Phe Ser Leu Asp Phe Gly His
35 40 45

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

<210> 71
<211> 238
<212> PRT
<213> 褐家鼠(Rattus norvegicus)
<400> 71

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

Leu Phe Leu Arg Tyr Ile Ser Asp Trp Asp Leu Asp Pro Gly Arg Cys
115 120 125
Tyr Arg Val Thr Trp Phe Thr Ser Trp Ser Pro Cys Tyr Asp Cys Ala
130 135 140
Arg His Val Ala Asp Phe Leu Arg Gly Asn Pro Asn Leu Ser Leu Arg
145 150 155 160
Ile Phe Thr Ala Arg Leu Thr Gly Trp Gly Ala Leu Pro Ala Gly Leu
165 170 175
Met Ser Pro Ala Arg Pro Ser Asp Tyr Phe Tyr Cys Trp Asn Thr Phe
180 185 190
Val Glu Asn His Glu Arg Thr Phe Lys Ala Trp Glu Gly Leu His Glu
195 200 205
Asn Ser Val Arg Leu Ser Arg Arg Leu Arg Arg Ile Leu Leu Pro Leu
210 215 220
Tyr Glu Val Asp Asp Leu Arg Asp Ala Phe Arg Thr Leu Gly
225 230 235

<210> 72
<211> 429
<212> PRT
<213> 小家鼠
<400> 72

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

Arg Leu Val Gln Glu Gly Ala Gln Val Ala Ala Met Asp Leu Tyr Glu
145 150 155 160
Phe Lys Lys Cys Trp Lys Lys Phe Val Asp Asn Gly Gly Arg Arg Phe
165 170 175
Arg Pro Trp Lys Arg Leu Leu Thr Asn Phe Arg Tyr Gln Asp Ser Lys
180 185 190
Leu Gln Glu Ile Leu Arg Pro Cys Tyr Ile Pro Val Pro Ser Ser Ser
195 200 205
Ser Ser Thr Leu Ser Asn Ile Cys Leu Thr Lys Gly Leu Pro Glu Thr
210 215 220
Arg Phe Cys Val Glu Gly Arg Arg Met Asp Pro Leu Ser Glu Glu Glu
225 230 235 240
Phe Tyr Ser Gln Phe Tyr Asn Gln Arg Val Lys His Leu Cys Tyr Tyr
245 250 255
His Arg Met Lys Pro Tyr Leu Cys Tyr Gln Leu Glu Gln Phe Asn Gly
260 265 270
Gln Ala Pro Leu Lys Gly Cys Leu Leu Ser Glu Lys Gly Lys Gln His
275 280 285
Ala Glu Ile Leu Phe Leu Asp Lys Ile Arg Ser Met Glu Leu Ser Gln
290 295 300
Val Thr Ile Thr Cys Tyr Leu Thr Trp Ser Pro Cys Pro Asn Cys Ala
305 310 315 320
Trp Gln Leu Ala Ala Phe Lys Arg Asp Arg Pro Asp Leu Ile Leu His
325 330 335
Ile Tyr Thr Ser Arg Leu Tyr Phe His Trp Lys Arg Pro Phe Gln Lys
340 345 350
Gly Leu Cys Ser Leu Trp Gln Ser Gly Ile Leu Val Asp Val Met Asp
355 360 365
Leu Pro Gln Phe Thr Asp Cys Trp Thr Asn Phe Val Asn Pro Lys Arg
370 375 380
Pro Phe Trp Pro Trp Lys Gly Leu Glu Ile Ile Ser Arg Arg Thr Gln
385 390 395 400
Arg Arg Leu Arg Arg Ile Lys Glu Ser Trp Gly Leu Gln Asp Leu Val
405 410 415
Asn Asp Phe Gly Asn Leu Gln Leu Gly Pro Pro Met Ser
420 425

<210> 73
<211> 429
<212> PRT

<213> 褐家鼠

<400> 73

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

290 295 300

Val Ile Ile Thr Cys Tyr Leu Thr Trp Ser Pro Cys Pro Asn Cys Ala
305 310 315 320

Trp Gln Leu Ala Ala Phe Lys Arg Asp Arg Pro Asp Leu Ile Leu His
325 330 335

Ile Tyr Thr Ser Arg Leu Tyr Phe His Trp Lys Arg Pro Phe Gln Lys
340 345 350

Gly Leu Cys Ser Leu Trp Gln Ser Gly Ile Leu Val Asp Val Met Asp
355 360 365

Leu Pro Gln Phe Thr Asp Cys Trp Thr Asn Phe Val Asn Pro Lys Arg
370 375 380

Pro Phe Trp Pro Trp Lys Gly Leu Glu Ile Ile Ser Arg Arg Thr Gln
385 390 395 400

Arg Arg Leu His Arg Ile Lys Glu Ser Trp Gly Leu Gln Asp Leu Val
405 410 415

Asn Asp Phe Gly Asn Leu Gln Leu Gly Pro Pro Met Ser
420 425

<210> 74

<211> 370

<212> PRT

<213> 恒河猴(Macaca mulatta)

<400> 74

Met Val Glu Pro Met Asp Pro Arg Thr Phe Val Ser Asn Phe Asn Asn
1 5 10 15

Arg Pro Ile Leu Ser Gly Leu Asn Thr Val Trp Leu Cys Cys Glu Val
20 25 30

Lys Thr Lys Asp Pro Ser Gly Pro Pro Leu Asp Ala Lys Ile Phe Gln
35 40 45

Gly Lys Val Tyr Ser Lys Ala Lys Tyr His Pro Glu Met Arg Phe Leu
50 55 60

Arg Trp Phe His Lys Trp Arg Gln Leu His His Asp Gln Glu Tyr Lys
65 70 75 80

Val Thr Trp Tyr Val Ser Trp Ser Pro Cys Thr Arg Cys Ala Asn Ser
85 90 95

Val Ala Thr Phe Leu Ala Lys Asp Pro Lys Val Thr Leu Thr Ile Phe
100 105 110

Val Ala Arg Leu Tyr Tyr Phe Trp Lys Pro Asp Tyr Gln Gln Ala Leu
115 120 125

Arg Ile Leu Cys Gln Lys Arg Gly Gly Pro His Ala Thr Met Lys Ile

130 135 140

Met Asn Tyr Asn Glu Phe Gln Asp Cys Trp Asn Lys Phe Val Asp Gly
145 150 155 160
Arg Gly Lys Pro Phe Lys Pro Arg Asn Asn Leu Pro Lys His Tyr Thr
165 170 175
Leu Leu Gln Ala Thr Leu Gly Glu Leu Leu Arg His Leu Met Asp Pro
180 185 190
Gly Thr Phe Thr Ser Asn Phe Asn Asn Lys Pro Trp Val Ser Gly Gln
195 200 205
His Glu Thr Tyr Leu Cys Tyr Lys Val Glu Arg Leu His Asn Asp Thr
210 215 220
Trp Val Pro Leu Asn Gln His Arg Gly Phe Leu Arg Asn Gln Ala Pro
225 230 235 240
Asn Ile His Gly Phe Pro Lys Gly Arg His Ala Glu Leu Cys Phe Leu
245 250 255
Asp Leu Ile Pro Phe Trp Lys Leu Asp Gly Gln Gln Tyr Arg Val Thr
260 265 270
Cys Phe Thr Ser Trp Ser Pro Cys Phe Ser Cys Ala Gln Glu Met Ala
275 280 285
Lys Phe Ile Ser Asn Asn Glu His Val Ser Leu Cys Ile Phe Ala Ala
290 295 300
Arg Ile Tyr Asp Asp Gln Gly Arg Tyr Gln Glu Gly Leu Arg Ala Leu
305 310 315 320
His Arg Asp Gly Ala Lys Ile Ala Met Met Asn Tyr Ser Glu Phe Glu
325 330 335
Tyr Cys Trp Asp Thr Phe Val Asp Arg Gln Gly Arg Pro Phe Gln Pro
340 345 350
Trp Asp Gly Leu Asp Glu His Ser Gln Ala Leu Ser Gly Arg Leu Arg
355 360 365
Ala Ile
370

<210> 75
<211> 384
<212> PRT
<213> 黑猩猩属(Pan troglodytes)
<400> 75

Met Lys Pro His Phe Arg Asn Pro Val Glu Arg Met Tyr Gln Asp Thr
1 5 10 15
Phe Ser Asp Asn Phe Tyr Asn Arg Pro Ile Leu Ser His Arg Asn Thr

20 25 30

Val Trp Leu Cys Tyr Glu Val Lys Thr Lys Gly Pro Ser Arg Pro Pro
35 40 45

Leu Asp Ala Lys Ile Phe Arg Gly Gln Val Tyr Ser Lys Leu Lys Tyr
50 55 60

His Pro Glu Met Arg Phe Phe His Trp Phe Ser Lys Trp Arg Lys Leu
65 70 75 80

His Arg Asp Gln Glu Tyr Glu Val Thr Trp Tyr Ile Ser Trp Ser Pro
85 90 95

Cys Thr Lys Cys Thr Arg Asp Val Ala Thr Phe Leu Ala Glu Asp Pro
100 105 110

Lys Val Thr Leu Thr Ile Phe Val Ala Arg Leu Tyr Tyr Phe Trp Asp
115 120 125

Pro Asp Tyr Gln Glu Ala Leu Arg Ser Leu Cys Gln Lys Arg Asp Gly
130 135 140

Pro Arg Ala Thr Met Lys Ile Met Asn Tyr Asp Glu Phe Gln His Cys
145 150 155 160

Trp Ser Lys Phe Val Tyr Ser Gln Arg Glu Leu Phe Glu Pro Trp Asn
165 170 175

Asn Leu Pro Lys Tyr Tyr Ile Leu Leu His Ile Met Leu Gly Glu Ile
180 185 190

Leu Arg His Ser Met Asp Pro Pro Thr Phe Thr Ser Asn Phe Asn Asn
195 200 205

Glu Leu Trp Val Arg Gly Arg His Glu Thr Tyr Leu Cys Tyr Glu Val
210 215 220

Glu Arg Leu His Asn Asp Thr Trp Val Leu Leu Asn Gln Arg Arg Gly
225 230 235 240

Phe Leu Cys Asn Gln Ala Pro His Lys His Gly Phe Leu Glu Gly Arg
245 250 255

His Ala Glu Leu Cys Phe Leu Asp Val Ile Pro Phe Trp Lys Leu Asp
260 265 270

Leu His Gln Asp Tyr Arg Val Thr Cys Phe Thr Ser Trp Ser Pro Cys
275 280 285

Phe Ser Cys Ala Gln Glu Met Ala Lys Phe Ile Ser Asn Asn Lys His
290 295 300

Val Ser Leu Cys Ile Phe Ala Ala Arg Ile Tyr Asp Asp Gln Gly Arg
305 310 315 320

Cys Gln Glu Gly Leu Arg Thr Leu Ala Lys Ala Gly Ala Lys Ile Ser
325 330 335

Ile Met Thr Tyr Ser Glu Phe Lys His Cys Trp Asp Thr Phe Val Asp
340 345 350
His Gln Gly Cys Pro Phe Gln Pro Trp Asp Gly Leu Glu Glu His Ser
355 360 365
Gln Ala Leu Ser Gly Arg Leu Arg Ala Ile Leu Gln Asn Gln Gly Asn
370 375 380

<210> 76
<211> 377
<212> PRT
<213> 绿猴(Chlorocebus sabaeus)
<400> 76

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

Glu Arg Ser His Asn Asp Thr Trp Val Leu Leu Asn Gln His Arg Gly
225 230 235 240
Phe Leu Arg Asn Gln Ala Pro Asp Arg His Gly Phe Pro Lys Gly Arg
245 250 255
His Ala Glu Leu Cys Phe Leu Asp Leu Ile Pro Phe Trp Lys Leu Asp
260 265 270
Asp Gln Gln Tyr Arg Val Thr Cys Phe Thr Ser Trp Ser Pro Cys Phe
275 280 285
Ser Cys Ala Gln Lys Met Ala Lys Phe Ile Ser Asn Asn Lys His Val
290 295 300
Ser Leu Cys Ile Phe Ala Ala Arg Ile Tyr Asp Asp Gln Gly Arg Cys
305 310 315 320
Gln Glu Gly Leu Arg Thr Leu His Arg Asp Gly Ala Lys Ile Ala Val
325 330 335
Met Asn Tyr Ser Glu Phe Glu Tyr Cys Trp Asp Thr Phe Val Asp Arg
340 345 350
Gln Gly Arg Pro Phe Gln Pro Trp Asp Gly Leu Asp Glu His Ser Gln
355 360 365
Ala Leu Ser Gly Arg Leu Arg Ala Ile
370 375

<210> 77
<211> 384
<212> PRT
<213> 智人
<400> 77

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

Lys Val Thr Leu Thr Ile Phe Val Ala Arg Leu Tyr Tyr Phe Trp Asp
115 120 125
Pro Asp Tyr Gln Glu Ala Leu Arg Ser Leu Cys Gln Lys Arg Asp Gly
130 135 140
Pro Arg Ala Thr Met Lys Ile Met Asn Tyr Asp Glu Phe Gln His Cys
145 150 155 160
Trp Ser Lys Phe Val Tyr Ser Gln Arg Glu Leu Phe Glu Pro Trp Asn
165 170 175
Asn Leu Pro Lys Tyr Tyr Ile Leu Leu His Ile Met Leu Gly Glu Ile
180 185 190
Leu Arg His Ser Met Asp Pro Pro Thr Phe Thr Phe Asn Phe Asn Asn
195 200 205
Glu Pro Trp Val Arg Gly Arg His Glu Thr Tyr Leu Cys Tyr Glu Val
210 215 220
Glu Arg Met His Asn Asp Thr Trp Val Leu Leu Asn Gln Arg Arg Gly
225 230 235 240
Phe Leu Cys Asn Gln Ala Pro His Lys His Gly Phe Leu Glu Gly Arg
245 250 255
His Ala Glu Leu Cys Phe Leu Asp Val Ile Pro Phe Trp Lys Leu Asp
260 265 270
Leu Asp Gln Asp Tyr Arg Val Thr Cys Phe Thr Ser Trp Ser Pro Cys
275 280 285
Phe Ser Cys Ala Gln Glu Met Ala Lys Phe Ile Ser Lys Asn Lys His
290 295 300
Val Ser Leu Cys Ile Phe Thr Ala Arg Ile Tyr Asp Asp Gln Gly Arg
305 310 315 320
Cys Gln Glu Gly Leu Arg Thr Leu Ala Glu Ala Gly Ala Lys Ile Ser
325 330 335
Ile Met Thr Tyr Ser Glu Phe Lys His Cys Trp Asp Thr Phe Val Asp
340 345 350
His Gln Gly Cys Pro Phe Gln Pro Trp Asp Gly Leu Asp Glu His Ser
355 360 365
Gln Asp Leu Ser Gly Arg Leu Arg Ala Ile Leu Gln Asn Gln Glu Asn
370 375 380

<210> 78
<211> 373
<212> PRT
<213> 智人
<400> 78

Met Lys Pro His Phe Arg Asn Thr Val Glu Arg Met Tyr Arg Asp Thr
1 5 10 15
Phe Ser Tyr Asn Phe Tyr Asn Arg Pro Ile Leu Ser Arg Arg Asn Thr
20 25 30
Val Trp Leu Cys Tyr Glu Val Lys Thr Lys Gly Pro Ser Arg Pro Arg
35 40 45
Leu Asp Ala Lys Ile Phe Arg Gly Gln Val Tyr Ser Gln Pro Glu His
50 55 60
His Ala Glu Met Cys Phe Leu Ser Trp Phe Cys Gly Asn Gln Leu Pro
65 70 75 80
Ala Tyr Lys Cys Phe Gln Ile Thr Trp Phe Val Ser Trp Thr Pro Cys
85 90 95
Pro Asp Cys Val Ala Lys Leu Ala Glu Phe Leu Ala Glu His Pro Asn
100 105 110
Val Thr Leu Thr Ile Ser Ala Ala Arg Leu Tyr Tyr Tyr Trp Glu Arg
115 120 125
Asp Tyr Arg Arg Ala Leu Cys Arg Leu Ser Gln Ala Gly Ala Arg Val
130 135 140
Lys Ile Met Asp Asp Glu Glu Phe Ala Tyr Cys Trp Glu Asn Phe Val
145 150 155 160
Tyr Ser Glu Gly Gln Pro Phe Met Pro Trp Tyr Lys Phe Asp Asp Asn
165 170 175
Tyr Ala Phe Leu His Arg Thr Leu Lys Glu Ile Leu Arg Asn Pro Met
180 185 190
Glu Ala Met Tyr Pro His Ile Phe Tyr Phe His Phe Lys Asn Leu Arg
195 200 205
Lys Ala Tyr Gly Arg Asn Glu Ser Trp Leu Cys Phe Thr Met Glu Val
210 215 220
Val Lys His His Ser Pro Val Ser Trp Lys Arg Gly Val Phe Arg Asn
225 230 235 240
Gln Val Asp Pro Glu Thr His Cys His Ala Glu Arg Cys Phe Leu Ser
245 250 255
Trp Phe Cys Asp Asp Ile Leu Ser Pro Asn Thr Asn Tyr Glu Val Thr
260 265 270
Trp Tyr Thr Ser Trp Ser Pro Cys Pro Glu Cys Ala Gly Glu Val Ala
275 280 285
Glu Phe Leu Ala Arg His Ser Asn Val Asn Leu Thr Ile Phe Thr Ala
290 295 300
Arg Leu Tyr Tyr Phe Trp Asp Thr Asp Tyr Gln Glu Gly Leu Arg Ser

305 310 315 320
Leu Ser Gln Glu Gly Ala Ser Val Glu Ile Met Gly Tyr Lys Asp Phe
325 330 335
Lys Tyr Cys Trp Glu Asn Phe Val Tyr Asn Asp Asp Glu Pro Phe Lys
340 345 350
Pro Trp Lys Gly Leu Lys Tyr Asn Phe Leu Phe Leu Asp Ser Lys Leu
355 360 365
Gln Glu Ile Leu Glu
370
<210> 79
<211> 382
<212> PRT
<213> 智人
<400> 79
Met Asn Pro Gln Ile Arg Asn Pro Met Glu Arg Met Tyr Arg Asp Thr
1 5 10 15
Phe Tyr Asp Asn Phe Glu Asn Glu Pro Ile Leu Tyr Gly Arg Ser Tyr
20 25 30
Thr Trp Leu Cys Tyr Glu Val Lys Ile Lys Arg Gly Arg Ser Asn Leu
35 40 45
Leu Trp Asp Thr Gly Val Phe Arg Gly Gln Val Tyr Phe Lys Pro Gln
50 55 60
Tyr His Ala Glu Met Cys Phe Leu Ser Trp Phe Cys Gly Asn Gln Leu
65 70 75 80
Pro Ala Tyr Lys Cys Phe Gln Ile Thr Trp Phe Val Ser Trp Thr Pro
85 90 95
Cys Pro Asp Cys Val Ala Lys Leu Ala Glu Phe Leu Ser Glu His Pro
100 105 110
Asn Val Thr Leu Thr Ile Ser Ala Ala Arg Leu Tyr Tyr Tyr Trp Glu
115 120 125
Arg Asp Tyr Arg Arg Ala Leu Cys Arg Leu Ser Gln Ala Gly Ala Arg
130 135 140
Val Thr Ile Met Asp Tyr Glu Glu Phe Ala Tyr Cys Trp Glu Asn Phe
145 150 155 160
Val Tyr Asn Glu Gly Gln Gln Phe Met Pro Trp Tyr Lys Phe Asp Glu
165 170 175
Asn Tyr Ala Phe Leu His Arg Thr Leu Lys Glu Ile Leu Arg Tyr Leu
180 185 190
Met Asp Pro Asp Thr Phe Thr Phe Asn Phe Asn Asn Asp Pro Leu Val

195 200 205

Leu Arg Arg Arg Gln Thr Tyr Leu Cys Tyr Glu Val Glu Arg Leu Asp
210 215 220

Asn Gly Thr Trp Val Leu Met Asp Gln His Met Gly Phe Leu Cys Asn
225 230 235 240

Glu Ala Lys Asn Leu Leu Cys Gly Phe Tyr Gly Arg His Ala Glu Leu
245 250 255

Arg Phe Leu Asp Leu Val Pro Ser Leu Gln Leu Asp Pro Ala Gln Ile
260 265 270

Tyr Arg Val Thr Trp Phe Ile Ser Trp Ser Pro Cys Phe Ser Trp Gly
275 280 285

Cys Ala Gly Glu Val Arg Ala Phe Leu Gln Glu Asn Thr His Val Arg
290 295 300

Leu Arg Ile Phe Ala Ala Arg Ile Tyr Asp Tyr Asp Pro Leu Tyr Lys
305 310 315 320

Glu Ala Leu Gln Met Leu Arg Asp Ala Gly Ala Gln Val Ser Ile Met
325 330 335

Thr Tyr Asp Glu Phe Glu Tyr Cys Trp Asp Thr Phe Val Tyr Arg Gln
340 345 350

Gly Cys Pro Phe Gln Pro Trp Asp Gly Leu Glu Glu His Ser Gln Ala
355 360 365

Leu Ser Gly Arg Leu Arg Ala Ile Leu Gln Asn Gln Gly Asn
370 375 380

<210> 80

<211> 396

<212> PRT

<213> 褐家鼠

<400> 80

Met Gln Pro Gln Gly Leu Gly Pro Asn Ala Gly Met Gly Pro Val Cys
1 5 10 15

Leu Gly Cys Ser His Arg Arg Pro Tyr Ser Pro Ile Arg Asn Pro Leu
20 25 30

Lys Lys Leu Tyr Gln Gln Thr Phe Tyr Phe His Phe Lys Asn Val Arg
35 40 45

Tyr Ala Trp Gly Arg Lys Asn Asn Phe Leu Cys Tyr Glu Val Asn Gly
50 55 60

Met Asp Cys Ala Leu Pro Val Pro Leu Arg Gln Gly Val Phe Arg Lys
65 70 75 80

Gln Gly His Ile His Ala Glu Leu Cys Phe Ile Tyr Trp Phe His Asp

85 90 95

Lys Val Leu Arg Val Leu Ser Pro Met Glu Glu Phe Lys Val Thr Trp
100 105 110

Tyr Met Ser Trp Ser Pro Cys Ser Lys Cys Ala Glu Gln Val Ala Arg
115 120 125

Phe Leu Ala Ala His Arg Asn Leu Ser Leu Ala Ile Phe Ser Ser Arg
130 135 140

Leu Tyr Tyr Tyr Leu Arg Asn Pro Asn Tyr Gln Gln Lys Leu Cys Arg
145 150 155 160

Leu Ile Gln Glu Gly Val His Val Ala Ala Met Asp Leu Pro Glu Phe
165 170 175

Lys Lys Cys Trp Asn Lys Phe Val Asp Asn Asp Gly Gln Pro Phe Arg
180 185 190

Pro Trp Met Arg Leu Arg Ile Asn Phe Ser Phe Tyr Asp Cys Lys Leu
195 200 205

Gln Glu Ile Phe Ser Arg Met Asn Leu Leu Arg Glu Asp Val Phe Tyr
210 215 220

Leu Gln Phe Asn Asn Ser His Arg Val Lys Pro Val Gln Asn Arg Tyr
225 230 235 240

Tyr Arg Arg Lys Ser Tyr Leu Cys Tyr Gln Leu Glu Arg Ala Asn Gly
245 250 255

Gln Glu Pro Leu Lys Gly Tyr Leu Leu Tyr Lys Lys Gly Glu Gln His
260 265 270

Val Glu Ile Leu Phe Leu Glu Lys Met Arg Ser Met Glu Leu Ser Gln
275 280 285

Val Arg Ile Thr Cys Tyr Leu Thr Trp Ser Pro Cys Pro Asn Cys Ala
290 295 300

Arg Gln Leu Ala Ala Phe Lys Lys Asp His Pro Asp Leu Ile Leu Arg
305 310 315 320

Ile Tyr Thr Ser Arg Leu Tyr Phe Tyr Trp Arg Lys Lys Phe Gln Lys
325 330 335

Gly Leu Cys Thr Leu Trp Arg Ser Gly Ile His Val Asp Val Met Asp
340 345 350

Leu Pro Gln Phe Ala Asp Cys Trp Thr Asn Phe Val Asn Pro Gln Arg
355 360 365

Pro Phe Arg Pro Trp Asn Glu Leu Glu Lys Asn Ser Trp Arg Ile Gln
370 375 380

Arg Arg Leu Arg Arg Ile Lys Glu Ser Trp Gly Leu
385 390 395

<210> 81
<211> 227
<212> PRT
<213> 牛
<400> 81

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

<210> 82
<211> 490
<212> PRT
<213> 黑猩猩属

<400> 82

Met Asn Pro Gln Ile Arg Asn Pro Met Glu Trp Met Tyr Gln Arg Thr
1 5 10 15
Phe Tyr Tyr Asn Phe Glu Asn Glu Pro Ile Leu Tyr Gly Arg Ser Tyr
20 25 30
Thr Trp Leu Cys Tyr Glu Val Lys Ile Arg Arg Gly His Ser Asn Leu
35 40 45
Leu Trp Asp Thr Gly Val Phe Arg Gly Gln Met Tyr Ser Gln Pro Glu
50 55 60
His His Ala Glu Met Cys Phe Leu Ser Trp Phe Cys Gly Asn Gln Leu
65 70 75 80
Ser Ala Tyr Lys Cys Phe Gln Ile Thr Trp Phe Val Ser Trp Thr Pro
85 90 95
Cys Pro Asp Cys Val Ala Lys Leu Ala Lys Phe Leu Ala Glu His Pro
100 105 110
Asn Val Thr Leu Thr Ile Ser Ala Ala Arg Leu Tyr Tyr Tyr Trp Glu
115 120 125
Arg Asp Tyr Arg Arg Ala Leu Cys Arg Leu Ser Gln Ala Gly Ala Arg
130 135 140
Val Lys Ile Met Asp Asp Glu Glu Phe Ala Tyr Cys Trp Glu Asn Phe
145 150 155 160
Val Tyr Asn Glu Gly Gln Pro Phe Met Pro Trp Tyr Lys Phe Asp Asp
165 170 175
Asn Tyr Ala Phe Leu His Arg Thr Leu Lys Glu Ile Ile Arg His Leu
180 185 190
Met Asp Pro Asp Thr Phe Thr Phe Asn Phe Asn Asn Asp Pro Leu Val
195 200 205
Leu Arg Arg His Gln Thr Tyr Leu Cys Tyr Glu Val Glu Arg Leu Asp
210 215 220
Asn Gly Thr Trp Val Leu Met Asp Gln His Met Gly Phe Leu Cys Asn
225 230 235 240
Glu Ala Lys Asn Leu Leu Cys Gly Phe Tyr Gly Arg His Ala Glu Leu
245 250 255
Arg Phe Leu Asp Leu Val Pro Ser Leu Gln Leu Asp Pro Ala Gln Ile
260 265 270
Tyr Arg Val Thr Trp Phe Ile Ser Trp Ser Pro Cys Phe Ser Trp Gly
275 280 285
Cys Ala Gly Gln Val Arg Ala Phe Leu Gln Glu Asn Thr His Val Arg
290 295 300

Leu Arg Ile Phe Ala Ala Arg Ile Tyr Asp Tyr Asp Pro Leu Tyr Lys
305 310 315 320
Glu Ala Leu Gln Met Leu Arg Asp Ala Gly Ala Gln Val Ser Ile Met
325 330 335
Thr Tyr Asp Glu Phe Glu Tyr Cys Trp Asp Thr Phe Val Tyr Arg Gln
340 345 350
Gly Cys Pro Phe Gln Pro Trp Asp Gly Leu Glu Glu His Ser Gln Ala
355 360 365
Leu Ser Gly Arg Leu Arg Ala Ile Leu Gln Val Arg Ala Ser Ser Leu
370 375 380
Cys Met Val Pro His Arg Pro Pro Pro Pro Pro Gln Ser Pro Gly Pro
385 390 395 400
Cys Leu Pro Leu Cys Ser Glu Pro Pro Leu Gly Ser Leu Leu Pro Thr
405 410 415
Gly Arg Pro Ala Pro Ser Leu Pro Phe Leu Leu Thr Ala Ser Phe Ser
420 425 430
Phe Pro Pro Pro Ala Ser Leu Pro Pro Leu Pro Ser Leu Ser Leu Ser
435 440 445
Pro Gly His Leu Pro Val Pro Ser Phe His Ser Leu Thr Ser Cys Ser
450 455 460
Ile Gln Pro Pro Cys Ser Ser Arg Ile Arg Glu Thr Glu Gly Trp Ala
465 470 475 480
Ser Val Ser Lys Glu Gly Arg Asp Leu Gly
485 490

<210> 83
<211> 190
<212> PRT
<213> 智人
<400> 83

Met Asn Pro Gln Ile Arg Asn Pro Met Lys Ala Met Tyr Pro Gly Thr
1 5 10 15
Phe Tyr Phe Gln Phe Lys Asn Leu Trp Glu Ala Asn Asp Arg Asn Glu
20 25 30
Thr Trp Leu Cys Phe Thr Val Glu Gly Ile Lys Arg Arg Ser Val Val
35 40 45
Ser Trp Lys Thr Gly Val Phe Arg Asn Gln Val Asp Ser Glu Thr His
50 55 60
Cys His Ala Glu Arg Cys Phe Leu Ser Trp Phe Cys Asp Asp Ile Leu
65 70 75 80

Ser Pro Asn Thr Lys Tyr Gln Val Thr Trp Tyr Thr Ser Trp Ser Pro
85 90 95
Cys Pro Asp Cys Ala Gly Glu Val Ala Glu Phe Leu Ala Arg His Ser
100 105 110
Asn Val Asn Leu Thr Ile Phe Thr Ala Arg Leu Tyr Tyr Phe Gln Tyr
115 120 125
Pro Cys Tyr Gln Glu Gly Leu Arg Ser Leu Ser Gln Glu Gly Val Ala
130 135 140
Val Glu Ile Met Asp Tyr Glu Asp Phe Lys Tyr Cys Trp Glu Asn Phe
145 150 155 160
Val Tyr Asn Asp Asn Glu Pro Phe Lys Pro Trp Lys Gly Leu Lys Thr
165 170 175
Asn Phe Arg Leu Leu Lys Arg Arg Leu Arg Glu Ser Leu Gln
180 185 190

<210> 84
<211> 190
<212> PRT
<213> 大猩猩
<400> 84

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

Val Tyr Asn Asp Asp Glu Pro Phe Lys Pro Trp Lys Gly Leu Lys Tyr
165 170 175
Asn Phe Arg Phe Leu Lys Arg Arg Leu Gln Glu Ile Leu Glu
180 185 190

<210> 85
<211> 199
<212> PRT
<213> 智人
<400> 85

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

<210> 86
<211> 202
<212> PRT
<213> 恒河猴

<400> 86

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

<210> 87

<211> 185

<212> PRT

<213> 牛

<400> 87

Met Asp Glu Tyr Thr Phe Thr Glu Asn Phe Asn Asn Gln Gly Trp Pro
1 5 10 15
Ser Lys Thr Tyr Leu Cys Tyr Glu Met Glu Arg Leu Asp Gly Asp Ala
20 25 30
Thr Ile Pro Leu Asp Glu Tyr Lys Gly Phe Val Arg Asn Lys Gly Leu
35 40 45
Asp Gln Pro Glu Lys Pro Cys His Ala Glu Leu Tyr Phe Leu Gly Lys

50 55 60

Ile His Ser Trp Asn Leu Asp Arg Asn Gln His Tyr Arg Leu Thr Cys
65 70 75 80

Phe Ile Ser Trp Ser Pro Cys Tyr Asp Cys Ala Gln Lys Leu Thr Thr
85 90 95

Phe Leu Lys Glu Asn His His Ile Ser Leu His Ile Leu Ala Ser Arg
100 105 110

Ile Tyr Thr His Asn Arg Phe Gly Cys His Gln Ser Gly Leu Cys Glu
115 120 125

Leu Gln Ala Ala Gly Ala Arg Ile Thr Ile Met Thr Phe Glu Asp Phe
130 135 140

Lys His Cys Trp Glu Thr Phe Val Asp His Lys Gly Lys Pro Phe Gln
145 150 155 160

Pro Trp Glu Gly Leu Asn Val Lys Ser Gln Ala Leu Cys Thr Glu Leu
165 170 175

Gln Ala Ile Leu Lys Thr Gln Gln Asn
180 185

<210> 88
<211> 200
<212> PRT
<213> 智人
<400> 88

Met Ala Leu Leu Thr Ala Glu Thr Phe Arg Leu Gln Phe Asn Asn Lys
1 5 10 15

Arg Arg Leu Arg Arg Pro Tyr Tyr Pro Arg Lys Ala Leu Leu Cys Tyr
20 25 30

Gln Leu Thr Pro Gln Asn Gly Ser Thr Pro Thr Arg Gly Tyr Phe Glu
35 40 45

Asn Lys Lys Lys Cys His Ala Glu Ile Cys Phe Ile Asn Glu Ile Lys
50 55 60

Ser Met Gly Leu Asp Glu Thr Gln Cys Tyr Gln Val Thr Cys Tyr Leu
65 70 75 80

Thr Trp Ser Pro Cys Ser Ser Cys Ala Trp Glu Leu Val Asp Phe Ile
85 90 95

Lys Ala His Asp His Leu Asn Leu Gly Ile Phe Ala Ser Arg Leu Tyr
100 105 110

Tyr His Trp Cys Lys Pro Gln Gln Lys Gly Leu Arg Leu Leu Cys Gly
115 120 125

Ser Gln Val Pro Val Glu Val Met Gly Phe Pro Lys Phe Ala Asp Cys

130 135 140

Trp Glu Asn Phe Val Asp His Glu Lys Pro Leu Ser Phe Asn Pro Tyr
145 150 155 160
Lys Met Leu Glu Glu Leu Asp Lys Asn Ser Arg Ala Ile Lys Arg Arg
165 170 175
Leu Glu Arg Ile Lys Ile Pro Gly Val Arg Ala Gln Gly Arg Tyr Met
180 185 190
Asp Ile Leu Cys Asp Ala Glu Val
195 200

<210> 89
<211> 210
<212> PRT
<213> 恒河猴
<400> 89

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

195 200 205

Ser Arg
210

<210> 90
<211> 386
<212> PRT
<213> 智人
<400> 90

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

245 250 255
Pro Glu Thr His Cys His Ala Glu Arg Cys Phe Leu Ser Trp Phe Cys
260 265 270
Asp Asp Ile Leu Ser Pro Asn Thr Asn Tyr Glu Val Thr Trp Tyr Thr
275 280 285
Ser Trp Ser Pro Cys Pro Glu Cys Ala Gly Glu Val Ala Glu Phe Leu
290 295 300
Ala Arg His Ser Asn Val Asn Leu Thr Ile Phe Thr Ala Arg Leu Cys
305 310 315 320
Tyr Phe Trp Asp Thr Asp Tyr Gln Glu Gly Leu Cys Ser Leu Ser Gln
325 330 335
Glu Gly Ala Ser Val Lys Ile Met Gly Tyr Lys Asp Phe Val Ser Cys
340 345 350
Trp Lys Asn Phe Val Tyr Ser Asp Asp Glu Pro Phe Lys Pro Trp Lys
355 360 365
Gly Leu Gln Thr Asn Phe Arg Leu Leu Lys Arg Arg Leu Arg Glu Ile
370 375 380
Leu Gln
385

<210> 91
<211> 236
<212> PRT
<213> 智人
<400> 91

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

115 120 125

Gln Gly Leu Arg Asp Leu Val Asn Ser Gly Val Thr Ile Gln Ile Met
130 135 140

Arg Ala Ser Glu Tyr Tyr His Cys Trp Arg Asn Phe Val Asn Tyr Pro
145 150 155 160

Pro Gly Asp Glu Ala His Trp Pro Gln Tyr Pro Pro Leu Trp Met Met
165 170 175

Leu Tyr Ala Leu Glu Leu His Cys Ile Ile Leu Ser Leu Pro Pro Cys
180 185 190

Leu Lys Ile Ser Arg Arg Trp Gln Asn His Leu Thr Phe Phe Arg Leu
195 200 205

His Leu Gln Asn Cys His Tyr Gln Thr Ile Pro Pro His Ile Leu Leu
210 215 220

Ala Thr Gly Leu Ile His Pro Ser Val Ala Trp Arg
225 230 235

<210> 92
<211> 229
<212> PRT
<213> 小家鼠
<400> 92

Met Ser Ser Glu Thr Gly Pro Val Ala Val Asp Pro Thr Leu Arg Arg
1 5 10 15

Arg Ile Glu Pro His Glu Phe Glu Val Phe Phe Asp Pro Arg Glu Leu
20 25 30

Arg Lys Glu Thr Cys Leu Leu Tyr Glu Ile Asn Trp Gly Gly Arg His
35 40 45

Ser Val Trp Arg His Thr Ser Gln Asn Thr Ser Asn His Val Glu Val
50 55 60

Asn Phe Leu Glu Lys Phe Thr Thr Glu Arg Tyr Phe Arg Pro Asn Thr
65 70 75 80

Arg Cys Ser Ile Thr Trp Phe Leu Ser Trp Ser Pro Cys Gly Glu Cys
85 90 95

Ser Arg Ala Ile Thr Glu Phe Leu Ser Arg His Pro Tyr Val Thr Leu
100 105 110

Phe Ile Tyr Ile Ala Arg Leu Tyr His His Thr Asp Gln Arg Asn Arg
115 120 125

Gln Gly Leu Arg Asp Leu Ile Ser Ser Gly Val Thr Ile Gln Ile Met
130 135 140

Thr Glu Gln Glu Tyr Cys Tyr Cys Trp Arg Asn Phe Val Asn Tyr Pro

145 150 155 160
Pro Ser Asn Glu Ala Tyr Trp Pro Arg Tyr Pro His Leu Trp Val Lys
165 170 175
Leu Tyr Val Leu Glu Leu Tyr Cys Ile Ile Leu Gly Leu Pro Pro Cys
180 185 190
Leu Lys Ile Leu Arg Arg Lys Gln Pro Gln Leu Thr Phe Phe Thr Ile
195 200 205
Thr Leu Gln Thr Cys His Tyr Gln Arg Ile Pro Pro His Leu Leu Trp
210 215 220
Ala Thr Gly Leu Lys
225

<210> 93
<211> 229
<212> PRT
<213> 褐家鼠
<400> 93

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

180 185 190

Leu Asn Ile Leu Arg Arg Lys Gln Pro Gln Leu Thr Phe Phe Thr Ile
195 200 205

Ala Leu Gln Ser Cys His Tyr Gln Arg Leu Pro Pro His Ile Leu Trp
210 215 220

Ala Thr Gly Leu Lys
225

<210> 94
<211> 224
<212> PRT
<213> 智人
<400> 94

Met Ala Gln Lys Glu Glu Ala Ala Val Ala Thr Glu Ala Ala Ser Gln
1 5 10 15

Asn Gly Glu Asp Leu Glu Asn Leu Asp Asp Pro Glu Lys Leu Lys Glu
20 25 30

Leu Ile Glu Leu Pro Pro Phe Glu Ile Val Thr Gly Glu Arg Leu Pro
35 40 45

Ala Asn Phe Phe Lys Phe Gln Phe Arg Asn Val Glu Tyr Ser Ser Gly
50 55 60

Arg Asn Lys Thr Phe Leu Cys Tyr Val Val Glu Ala Gln Gly Lys Gly
65 70 75 80

Gly Gln Val Gln Ala Ser Arg Gly Tyr Leu Glu Asp Glu His Ala Ala
85 90 95

Ala His Ala Glu Glu Ala Phe Phe Asn Thr Ile Leu Pro Ala Phe Asp
100 105 110

Pro Ala Leu Arg Tyr Asn Val Thr Trp Tyr Val Ser Ser Ser Pro Cys
115 120 125

Ala Ala Cys Ala Asp Arg Ile Ile Lys Thr Leu Ser Lys Thr Lys Asn
130 135 140

Leu Arg Leu Leu Ile Leu Val Gly Arg Leu Phe Met Trp Glu Glu Pro
145 150 155 160

Glu Ile Gln Ala Ala Leu Lys Lys Leu Lys Glu Ala Gly Cys Lys Leu
165 170 175

Arg Ile Met Lys Pro Gln Asp Phe Glu Tyr Val Trp Gln Asn Phe Val
180 185 190

Glu Gln Glu Glu Gly Glu Ser Lys Ala Phe Gln Pro Trp Glu Asp Ile
195 200 205

Gln Glu Asn Phe Leu Tyr Tyr Glu Glu Lys Leu Ala Asp Ile Leu Lys

210 215 220

<210> 95
<211> 224
<212> PRT
<213> 小家鼠
<400> 95

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

<210> 96
<211> 224
<212> PRT
<213> 褐家鼠
<400> 96

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

<210> 97
<211> 224
<212> PRT
<213> 牛
<400> 97

Met Ala Gln Lys Glu Glu Ala Ala Ala Ala Ala Glu Pro Ala Ser Gln
1 5 10 15
Asn Gly Glu Glu Val Glu Asn Leu Glu Asp Pro Glu Lys Leu Lys Glu
20 25 30
Leu Ile Glu Leu Pro Pro Phe Glu Ile Val Thr Gly Glu Arg Leu Pro
35 40 45

Ala His Tyr Phe Lys Phe Gln Phe Arg Asn Val Glu Tyr Ser Ser Gly
50 55 60
Arg Asn Lys Thr Phe Leu Cys Tyr Val Val Glu Ala Gln Ser Lys Gly
65 70 75 80
Gly Gln Val Gln Ala Ser Arg Gly Tyr Leu Glu Asp Glu His Ala Thr
85 90 95
Asn His Ala Glu Glu Ala Phe Phe Asn Ser Ile Met Pro Thr Phe Asp
100 105 110
Pro Ala Leu Arg Tyr Met Val Thr Trp Tyr Val Ser Ser Ser Pro Cys
115 120 125
Ala Ala Cys Ala Asp Arg Ile Val Lys Thr Leu Asn Lys Thr Lys Asn
130 135 140
Leu Arg Leu Leu Ile Leu Val Gly Arg Leu Phe Met Trp Glu Glu Pro
145 150 155 160
Glu Ile Gln Ala Ala Leu Arg Lys Leu Lys Glu Ala Gly Cys Arg Leu
165 170 175
Arg Ile Met Lys Pro Gln Asp Phe Glu Tyr Ile Trp Gln Asn Phe Val
180 185 190
Glu Gln Glu Glu Gly Glu Ser Lys Ala Phe Glu Pro Trp Glu Asp Ile
195 200 205
Gln Glu Asn Phe Leu Tyr Tyr Glu Glu Lys Leu Ala Asp Ile Leu Lys
210 215 220

<210> 98
<211> 208
<212> PRT
<213> 海七鳃鳗
<400> 98

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

Cys Ala Asp Cys Ala Glu Lys Ile Leu Glu Trp Tyr Asn Gln Glu Leu
100 105 110
Arg Gly Asn Gly His Thr Leu Lys Ile Trp Ala Cys Lys Leu Tyr Tyr
115 120 125
Glu Lys Asn Ala Arg Asn Gln Ile Gly Leu Trp Asn Leu Arg Asp Asn
130 135 140
Gly Val Gly Leu Asn Val Met Val Ser Glu His Tyr Gln Cys Cys Arg
145 150 155 160
Lys Ile Phe Ile Gln Ser Ser His Asn Gln Leu Asn Glu Asn Arg Trp
165 170 175
Leu Glu Lys Thr Leu Lys Arg Ala Glu Lys Arg Arg Ser Glu Leu Ser
180 185 190
Ile Met Ile Gln Val Lys Ile Leu His Thr Thr Lys Ser Pro Ala Val
195 200 205

<210> 99
<211> 384
<212> PRT
<213> 智人
<400> 99

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

Trp Ser Lys Phe Val Tyr Ser Gln Arg Glu Leu Phe Glu Pro Trp Asn
165 170 175
Asn Leu Pro Lys Tyr Tyr Ile Leu Leu His Ile Met Leu Gly Glu Ile
180 185 190
Leu Arg His Ser Met Asp Pro Pro Thr Phe Thr Phe Asn Phe Asn Asn
195 200 205
Glu Pro Trp Val Arg Gly Arg His Glu Thr Tyr Leu Cys Tyr Glu Val
210 215 220
Glu Arg Met His Asn Asp Thr Trp Val Leu Leu Asn Gln Arg Arg Gly
225 230 235 240
Phe Leu Cys Asn Gln Ala Pro His Lys His Gly Phe Leu Glu Gly Arg
245 250 255
His Ala Glu Leu Cys Phe Leu Asp Val Ile Pro Phe Trp Lys Leu Asp
260 265 270
Leu Asp Gln Asp Tyr Arg Val Thr Cys Phe Thr Ser Trp Ser Pro Cys
275 280 285
Phe Ser Cys Ala Gln Glu Met Ala Lys Phe Ile Ser Lys Asn Lys His
290 295 300
Val Ser Leu Cys Ile Phe Thr Ala Arg Ile Tyr Arg Arg Gln Gly Arg
305 310 315 320
Cys Gln Glu Gly Leu Arg Thr Leu Ala Glu Ala Gly Ala Lys Ile Ser
325 330 335
Ile Met Thr Tyr Ser Glu Phe Lys His Cys Trp Asp Thr Phe Val Asp
340 345 350
His Gln Gly Cys Pro Phe Gln Pro Trp Asp Gly Leu Asp Glu His Ser
355 360 365
Gln Asp Leu Ser Gly Arg Leu Arg Ala Ile Leu Gln Asn Gln Glu Asn
370 375 380

<210> 100
<211> 184
<212> PRT
<213> 智人
<400> 100

Met Asp Pro Pro Thr Phe Thr Phe Asn Phe Asn Asn Glu Pro Trp Val
1 5 10 15
Arg Gly Arg His Glu Thr Tyr Leu Cys Tyr Glu Val Glu Arg Met His
20 25 30
Asn Asp Thr Trp Val Leu Leu Asn Gln Arg Arg Gly Phe Leu Cys Asn
35 40 45

Gln Ala Pro His Lys His Gly Phe Leu Glu Gly Arg His Ala Glu Leu
50 55 60
Cys Phe Leu Asp Val Ile Pro Phe Trp Lys Leu Asp Leu Asp Gln Asp
65 70 75 80
Tyr Arg Val Thr Cys Phe Thr Ser Trp Ser Pro Cys Phe Ser Cys Ala
85 90 95
Gln Glu Met Ala Lys Phe Ile Ser Lys Asn Lys His Val Ser Leu Cys
100 105 110
Ile Phe Thr Ala Arg Ile Tyr Asp Asp Gln Gly Arg Cys Gln Glu Gly
115 120 125
Leu Arg Thr Leu Ala Glu Ala Gly Ala Lys Ile Ser Ile Met Thr Tyr
130 135 140
Ser Glu Phe Lys His Cys Trp Asp Thr Phe Val Asp His Gln Gly Cys
145 150 155 160
Pro Phe Gln Pro Trp Asp Gly Leu Asp Glu His Ser Gln Asp Leu Ser
165 170 175
Gly Arg Leu Arg Ala Ile Leu Gln
180

<210> 101
<211> 184
<212> PRT
<213> 智人
<400> 101

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

Leu Arg Thr Leu Ala Glu Ala Gly Ala Lys Ile Ser Ile Met Thr Tyr
130 135 140
Ser Glu Phe Lys His Cys Trp Asp Thr Phe Val Asp His Gln Gly Cys
145 150 155 160
Pro Phe Gln Pro Trp Asp Gly Leu Asp Glu His Ser Gln Asp Leu Ser
165 170 175
Gly Arg Leu Arg Ala Ile Leu Gln
180

<210> 102
<211> 16
<212> PRT
<213> 人工序列
<220>
<223> 合成多肽
<400> 102
Ser Gly Ser Glu Thr Pro Gly Thr Ser Glu Ser Ala Thr Pro Glu Ser
1 5 10 15

<210> 103
<211> 120
<212> PRT
<213> 人工序列
<220>
<223> 合成多肽
<220>
<221> MISC_FEATURE
<222> (5)..(120)
<223> 可以不存在
<400> 103
Ser Gly Gly Ser Ser Gly Gly Ser Ser Gly Gly Ser Ser Gly Gly Ser
1 5 10 15
Ser Gly Gly Ser Ser Gly Gly Ser Ser Gly Gly Ser Ser Gly Gly Ser
20 25 30
Ser Gly Gly Ser Ser Gly Gly Ser Ser Gly Gly Ser Ser Gly Gly Ser
35 40 45
Ser Gly Gly Ser Ser Gly Gly Ser Ser Gly Gly Ser Ser Gly Gly Ser
50 55 60
Ser Gly Gly Ser Ser Gly Gly Ser Ser Gly Gly Ser Ser Gly Gly Ser
65 70 75 80
Ser Gly Gly Ser Ser Gly Gly Ser Ser Gly Gly Ser Ser Gly Gly Ser

85 90 95
Ser Gly Gly Ser Ser Gly Gly Ser Ser Gly Gly Ser Ser Gly Gly Ser
100 105 110
Ser Gly Gly Ser Ser Gly Gly Ser
115 120
<210> 104
<211> 120
<212> PRT
<213> 人工序列
<220>
<223> 合成多肽
<220>
<221> MISC_FEATURE
<222> (5)..(120)
<223> 可以不存在
<400> 104
Gly Gly Gly Ser Gly Gly Gly Ser Gly Gly Gly Ser Gly Gly Gly Ser
1 5 10 15
Gly Gly Gly Ser Gly Gly Gly Ser Gly Gly Gly Ser Gly Gly Gly Ser
20 25 30
Gly Gly Gly Ser Gly Gly Gly Ser Gly Gly Gly Ser Gly Gly Gly Ser
35 40 45
Gly Gly Gly Ser Gly Gly Gly Ser Gly Gly Gly Ser Gly Gly Gly Ser
50 55 60
Gly Gly Gly Ser Gly Gly Gly Ser Gly Gly Gly Ser Gly Gly Gly Ser
65 70 75 80
Gly Gly Gly Ser Gly Gly Gly Ser Gly Gly Gly Ser Gly Gly Gly Ser
85 90 95
Gly Gly Gly Ser Gly Gly Gly Ser Gly Gly Gly Ser Gly Gly Gly Ser
100 105 110
Gly Gly Gly Ser Gly Gly Gly Ser
115 120
<210> 105
<211> 150
<212> PRT
<213> 人工序列
<220>
<223> 合成多肽
<220>

<221> MISC_FEATURE
<222> (6)..(150)
<223> 可以不存在
<400> 105

Gly Gly Gly Gly Ser Gly Gly Gly Gly Ser Gly Gly Gly Gly Ser Gly
1 5 10 15
Gly Gly Gly Ser Gly Gly Gly Gly Ser Gly Gly Gly Gly Ser Gly Gly
20 25 30
Gly Gly Ser Gly Gly Gly Gly Ser Gly Gly Gly Gly Ser Gly Gly Gly
35 40 45
Gly Ser Gly Gly Gly Gly Ser Gly Gly Gly Gly Ser Gly Gly Gly Gly
50 55 60
Ser Gly Gly Gly Gly Ser Gly Gly Gly Gly Ser Gly Gly Gly Gly Ser
65 70 75 80
Gly Gly Gly Gly Ser Gly Gly Gly Gly Ser Gly Gly Gly Gly Ser Gly
85 90 95
Gly Gly Gly Ser Gly Gly Gly Gly Ser Gly Gly Gly Gly Ser Gly Gly
100 105 110
Gly Gly Ser Gly Gly Gly Gly Ser Gly Gly Gly Gly Ser Gly Gly Gly
115 120 125
Gly Ser Gly Gly Gly Gly Ser Gly Gly Gly Gly Ser Gly Gly Gly Gly
130 135 140
Ser Gly Gly Gly Gly Ser
145 150

<210> 106
<211> 150
<212> PRT
<213> 人工序列
<220>
<223> 合成多肽
<220>
<221> MISC_FEATURE
<222> (6)..(150)
<223> 可以不存在
<400> 106

Glu Ala Ala Ala Lys Glu Ala Ala Ala Lys Glu Ala Ala Ala Lys Glu
1 5 10 15
Ala Ala Ala Lys Glu Ala Ala Ala Lys Glu Ala Ala Ala Lys Glu Ala
20 25 30

Ala Ala Lys Glu Ala Ala Ala Lys Glu Ala Ala Ala Lys Glu Ala Ala
35 40 45
Ala Lys Glu Ala Ala Ala Lys Glu Ala Ala Ala Lys Glu Ala Ala Ala
50 55 60
Lys Glu Ala Ala Ala Lys Glu Ala Ala Ala Lys Glu Ala Ala Ala Lys
65 70 75 80
Glu Ala Ala Ala Lys Glu Ala Ala Ala Lys Glu Ala Ala Ala Lys Glu
85 90 95
Ala Ala Ala Lys Glu Ala Ala Ala Lys Glu Ala Ala Ala Lys Glu Ala
100 105 110
Ala Ala Lys Glu Ala Ala Ala Lys Glu Ala Ala Ala Lys Glu Ala Ala
115 120 125
Ala Lys Glu Ala Ala Ala Lys Glu Ala Ala Ala Lys Glu Ala Ala Ala
130 135 140
Lys Glu Ala Ala Ala Lys
145 150

<210> 107
<211> 24
<212> PRT
<213> 人工序列
<220>
<223> 合成多肽
<400> 107
Ser Gly Gly Ser Ser Gly Ser Glu Thr Pro Gly Thr Ser Glu Ser Ala
1 5 10 15
Thr Pro Glu Ser Ser Gly Gly Ser
20

<210> 108
<211> 32
<212> PRT
<213> 人工序列
<220>
<223> 合成多肽
<400> 108
Ser Gly Gly Ser Ser Gly Gly Ser Ser Gly Ser Glu Thr Pro Gly Thr
1 5 10 15
Ser Glu Ser Ala Thr Pro Glu Ser Ser Gly Gly Ser Ser Gly Gly Ser
20 25 30

<210> 109

<211> 104
<212> PRT
<213> 人工序列
<220>
<223> 合成多肽
<400> 109

Gly Gly Ser Gly Gly Ser Pro Gly Ser Pro Ala Gly Ser Pro Thr Ser
1 5 10 15
Thr Glu Glu Gly Thr Ser Glu Ser Ala Thr Pro Glu Ser Gly Pro Gly
20 25 30
Thr Ser Thr Glu Pro Ser Glu Gly Ser Ala Pro Gly Ser Pro Ala Gly
35 40 45
Ser Pro Thr Ser Thr Glu Glu Gly Thr Ser Thr Glu Pro Ser Glu Gly
50 55 60
Ser Ala Pro Gly Thr Ser Thr Glu Pro Ser Glu Gly Ser Ala Pro Gly
65 70 75 80
Thr Ser Glu Ser Ala Thr Pro Glu Ser Gly Pro Gly Ser Glu Pro Ala
85 90 95
Thr Ser Gly Gly Ser Gly Gly Ser
100

<210> 110
<211> 20
<212> DNA
<213> 人工序列
<220>
<223> 合成多核苷酸
<400> 110

gaacacaaag catagactgc 20

<210> 111
<211> 20
<212> DNA
<213> 人工序列
<220>
<223> 合成多核苷酸
<400> 111

cttgtgtttc gtatctgacg 20

<210> 112
<211> 20
<212> DNA

<213> 人工序列
<220>
<223> 合成多核苷酸
<400> 112
gtcatcttag tcattacctg 20
<210> 113
<211> 20
<212> DNA
<213> 人工序列
<220>
<223> 合成多核苷酸
<400> 113
cagtagaatc agtaatggac 20
<210> 114
<211> 20
<212> DNA
<213> 人工序列
<220>
<223> 合成多核苷酸
<400> 114
ggaatccctt ctgcagcacc 20
<210> 115
<211> 1710
<212> PRT
<213> 人工序列
<220>
<223> 合成多肽
<400> 115

| | | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Met | Ser | Ser | Glu | Thr | Gly | Pro | Val | Ala | Val | Asp | Pro | Thr | Leu | Arg | Arg |
| 1 | | | | 5 | | | | 10 | | | | | 15 | | |
| Arg | Ile | Glu | Pro | His | Glu | Phe | Glu | Val | Phe | Phe | Asp | Pro | Arg | Glu | Leu |
| | | | 20 | | | | | 25 | | | | 30 | | | |
| Arg | Lys | Glu | Thr | Cys | Leu | Leu | Tyr | Glu | Ile | Asn | Trp | Gly | Gly | Arg | His |
| | | 35 | | | | | 40 | | | | | 45 | | | |
| Ser | Ile | Trp | Arg | His | Thr | Ser | Gln | Asn | Thr | Asn | Lys | His | Val | Glu | Val |
| | 50 | | | | | 55 | | | | | 60 | | | | |
| Asn | Phe | Ile | Glu | Lys | Phe | Thr | Thr | Glu | Arg | Tyr | Phe | Cys | Pro | Asn | Thr |
| 65 | | | | | 70 | | | | | 75 | | | | | 80 |
| Arg | Cys | Ser | Ile | Thr | Trp | Phe | Leu | Ser | Trp | Ser | Pro | Cys | Gly | Glu | Cys |

85 90 95
Ser Arg Ala Ile Thr Glu Phe Leu Ser Arg Tyr Pro His Val Thr Leu
100 105 110
Phe Ile Tyr Ile Ala Arg Leu Tyr His His Ala Asp Pro Arg Asn Arg
115 120 125
Gln Gly Leu Arg Asp Leu Ile Ser Ser Gly Val Thr Ile Gln Ile Met
130 135 140
Thr Glu Gln Glu Ser Gly Tyr Cys Trp Arg Asn Phe Val Asn Tyr Ser
145 150 155 160
Pro Ser Asn Glu Ala His Trp Pro Arg Tyr Pro His Leu Trp Val Arg
165 170 175
Leu Tyr Val Leu Glu Leu Tyr Cys Ile Ile Leu Gly Leu Pro Pro Cys
180 185 190
Leu Asn Ile Leu Arg Arg Lys Gln Pro Gln Leu Thr Phe Phe Thr Ile
195 200 205
Ala Leu Gln Ser Cys His Tyr Gln Arg Leu Pro Pro His Ile Leu Trp
210 215 220
Ala Thr Gly Leu Lys Ser Gly Ser Glu Thr Pro Gly Thr Ser Glu Ser
225 230 235 240
Ala Thr Pro Glu Ser Asp Lys Lys Tyr Ser Ile Gly Leu Ala Ile Gly
245 250 255
Thr Asn Ser Val Gly Trp Ala Val Ile Thr Asp Glu Tyr Lys Val Pro
260 265 270
Ser Lys Lys Phe Lys Val Leu Gly Asn Thr Asp Arg His Ser Ile Lys
275 280 285
Lys Asn Leu Ile Gly Ala Leu Leu Phe Asp Ser Gly Glu Thr Ala Glu
290 295 300
Ala Thr Arg Leu Lys Arg Thr Ala Arg Arg Arg Tyr Thr Arg Arg Lys
305 310 315 320
Asn Arg Ile Cys Tyr Leu Gln Glu Ile Phe Ser Asn Glu Met Ala Lys
325 330 335
Val Asp Asp Ser Phe Phe His Arg Leu Glu Glu Ser Phe Leu Val Glu
340 345 350
Glu Asp Lys Lys His Glu Arg His Pro Ile Phe Gly Asn Ile Val Asp
355 360 365
Glu Val Ala Tyr His Glu Lys Tyr Pro Thr Ile Tyr His Leu Arg Lys
370 375 380
Lys Leu Val Asp Ser Thr Asp Lys Ala Asp Leu Arg Leu Ile Tyr Leu
385 390 395 400

Ala Leu Ala His Met Ile Lys Phe Arg Gly His Phe Leu Ile Glu Gly
405 410 415
Asp Leu Asn Pro Asp Asn Ser Asp Val Asp Lys Leu Phe Ile Gln Leu
420 425 430
Val Gln Thr Tyr Asn Gln Leu Phe Glu Glu Asn Pro Ile Asn Ala Ser
435 440 445
Gly Val Asp Ala Lys Ala Ile Leu Ser Ala Arg Leu Ser Lys Ser Arg
450 455 460
Arg Leu Glu Asn Leu Ile Ala Gln Leu Pro Gly Glu Lys Lys Asn Gly
465 470 475 480
Leu Phe Gly Asn Leu Ile Ala Leu Ser Leu Gly Leu Thr Pro Asn Phe
485 490 495
Lys Ser Asn Phe Asp Leu Ala Glu Asp Ala Lys Leu Gln Leu Ser Lys
500 505 510
Asp Thr Tyr Asp Asp Asp Leu Asp Asn Leu Leu Ala Gln Ile Gly Asp
515 520 525
Gln Tyr Ala Asp Leu Phe Leu Ala Ala Lys Asn Leu Ser Asp Ala Ile
530 535 540
Leu Leu Ser Asp Ile Leu Arg Val Asn Thr Glu Ile Thr Lys Ala Pro
545 550 555 560
Leu Ser Ala Ser Met Ile Lys Arg Tyr Asp Glu His His Gln Asp Leu
565 570 575
Thr Leu Leu Lys Ala Leu Val Arg Gln Gln Leu Pro Glu Lys Tyr Lys
580 585 590
Glu Ile Phe Phe Asp Gln Ser Lys Asn Gly Tyr Ala Gly Tyr Ile Asp
595 600 605
Gly Gly Ala Ser Gln Glu Glu Phe Tyr Lys Phe Ile Lys Pro Ile Leu
610 615 620
Glu Lys Met Asp Gly Thr Glu Glu Leu Leu Val Lys Leu Asn Arg Glu
625 630 635 640
Asp Leu Leu Arg Lys Gln Arg Thr Phe Asp Asn Gly Ser Ile Pro His
645 650 655
Gln Ile His Leu Gly Glu Leu His Ala Ile Leu Arg Arg Gln Glu Asp
660 665 670
Phe Tyr Pro Phe Leu Lys Asp Asn Arg Glu Lys Ile Glu Lys Ile Leu
675 680 685
Thr Phe Arg Ile Pro Tyr Tyr Val Gly Pro Leu Ala Arg Gly Asn Ser
690 695 700
Arg Phe Ala Trp Met Thr Arg Lys Ser Glu Glu Thr Ile Thr Pro Trp

705 710 715 720
Asn Phe Glu Glu Val Val Asp Lys Gly Ala Ser Ala Gln Ser Phe Ile
725 730 735
Glu Arg Met Thr Asn Phe Asp Lys Asn Leu Pro Asn Glu Lys Val Leu
740 745 750
Pro Lys His Ser Leu Leu Tyr Glu Tyr Phe Thr Val Tyr Asn Glu Leu
755 760 765
Thr Lys Val Lys Tyr Val Thr Glu Gly Met Arg Lys Pro Ala Phe Leu
770 775 780
Ser Gly Glu Gln Lys Lys Ala Ile Val Asp Leu Leu Phe Lys Thr Asn
785 790 795 800
Arg Lys Val Thr Val Lys Gln Leu Lys Glu Asp Tyr Phe Lys Lys Ile
805 810 815
Glu Cys Phe Asp Ser Val Glu Ile Ser Gly Val Glu Asp Arg Phe Asn
820 825 830
Ala Ser Leu Gly Thr Tyr His Asp Leu Leu Lys Ile Ile Lys Asp Lys
835 840 845
Asp Phe Leu Asp Asn Glu Glu Asn Glu Asp Ile Leu Glu Asp Ile Val
850 855 860
Leu Thr Leu Thr Leu Phe Glu Asp Arg Glu Met Ile Glu Glu Arg Leu
865 870 875 880
Lys Thr Tyr Ala His Leu Phe Asp Asp Lys Val Met Lys Gln Leu Lys
885 890 895
Arg Arg Arg Tyr Thr Gly Trp Gly Arg Leu Ser Arg Lys Leu Ile Asn
900 905 910
Gly Ile Arg Asp Lys Gln Ser Gly Lys Thr Ile Leu Asp Phe Leu Lys
915 920 925
Ser Asp Gly Phe Ala Asn Arg Asn Phe Met Gln Leu Ile His Asp Asp
930 935 940
Ser Leu Thr Phe Lys Glu Asp Ile Gln Lys Ala Gln Val Ser Gly Gln
945 950 955 960
Gly Asp Ser Leu His Glu His Ile Ala Asn Leu Ala Gly Ser Pro Ala
965 970 975
Ile Lys Lys Gly Ile Leu Gln Thr Val Lys Val Val Asp Glu Leu Val
980 985 990
Lys Val Met Gly Arg His Lys Pro Glu Asn Ile Val Ile Glu Met Ala
995 1000 1005
Arg Glu Asn Gln Thr Thr Gln Lys Gly Gln Lys Asn Ser Arg Glu
1010 1015 1020

Arg Met Lys Arg Ile Glu Glu Gly Ile Lys Glu Leu Gly Ser Gln
1025 1030 1035
Ile Leu Lys Glu His Pro Val Glu Asn Thr Gln Leu Gln Asn Glu
1040 1045 1050
Lys Leu Tyr Leu Tyr Tyr Leu Gln Asn Gly Arg Asp Met Tyr Val
1055 1060 1065
Asp Gln Glu Leu Asp Ile Asn Arg Leu Ser Asp Tyr Asp Val Asp
1070 1075 1080
His Ile Val Pro Gln Ser Phe Leu Lys Asp Asp Ser Ile Asp Asn
1085 1090 1095
Lys Val Leu Thr Arg Ser Asp Lys Asn Arg Gly Lys Ser Asp Asn
1100 1105 1110
Val Pro Ser Glu Glu Val Val Lys Lys Met Lys Asn Tyr Trp Arg
1115 1120 1125
Gln Leu Leu Asn Ala Lys Leu Ile Thr Gln Arg Lys Phe Asp Asn
1130 1135 1140
Leu Thr Lys Ala Glu Arg Gly Gly Leu Ser Glu Leu Asp Lys Ala
1145 1150 1155
Gly Phe Ile Lys Arg Gln Leu Val Glu Thr Arg Gln Ile Thr Lys
1160 1165 1170
His Val Ala Gln Ile Leu Asp Ser Arg Met Asn Thr Lys Tyr Asp
1175 1180 1185
Glu Asn Asp Lys Leu Ile Arg Glu Val Lys Val Ile Thr Leu Lys
1190 1195 1200
Ser Lys Leu Val Ser Asp Phe Arg Lys Asp Phe Gln Phe Tyr Lys
1205 1210 1215
Val Arg Glu Ile Asn Asn Tyr His His Ala His Asp Ala Tyr Leu
1220 1225 1230
Asn Ala Val Val Gly Thr Ala Leu Ile Lys Lys Tyr Pro Lys Leu
1235 1240 1245
Glu Ser Glu Phe Val Tyr Gly Asp Tyr Lys Val Tyr Asp Val Arg
1250 1255 1260
Lys Met Ile Ala Lys Ser Glu Gln Glu Ile Gly Lys Ala Thr Ala
1265 1270 1275
Lys Tyr Phe Phe Tyr Ser Asn Ile Met Asn Phe Phe Lys Thr Glu
1280 1285 1290
Ile Thr Leu Ala Asn Gly Glu Ile Arg Lys Arg Pro Leu Ile Glu
1295 1300 1305
Thr Asn Gly Glu Thr Gly Glu Ile Val Trp Asp Lys Gly Arg Asp

1310 1315 1320

Phe Ala Thr Val Arg Lys Val Leu Ser Met Pro Gln Val Asn Ile

1325 1330 1335

Val Lys Lys Thr Glu Val Gln Thr Gly Gly Phe Ser Lys Glu Ser

1340 1345 1350

Ile Leu Pro Lys Arg Asn Ser Asp Lys Leu Ile Ala Arg Lys Lys

1355 1360 1365

Asp Trp Asp Pro Lys Lys Tyr Gly Gly Phe Asp Ser Pro Thr Val

1370 1375 1380

Ala Tyr Ser Val Leu Val Val Ala Lys Val Glu Lys Gly Lys Ser

1385 1390 1395

Lys Lys Leu Lys Ser Val Lys Glu Leu Leu Gly Ile Thr Ile Met

1400 1405 1410

Glu Arg Ser Ser Phe Glu Lys Asn Pro Ile Asp Phe Leu Glu Ala

1415 1420 1425

Lys Gly Tyr Lys Glu Val Lys Lys Asp Leu Ile Ile Lys Leu Pro

1430 1435 1440

Lys Tyr Ser Leu Phe Glu Leu Glu Asn Gly Arg Lys Arg Met Leu

1445 1450 1455

Ala Ser Ala Gly Glu Leu Gln Lys Gly Asn Glu Leu Ala Leu Pro

1460 1465 1470

Ser Lys Tyr Val Asn Phe Leu Tyr Leu Ala Ser His Tyr Glu Lys

1475 1480 1485

Leu Lys Gly Ser Pro Glu Asp Asn Glu Gln Lys Gln Leu Phe Val

1490 1495 1500

Glu Gln His Lys His Tyr Leu Asp Glu Ile Ile Glu Gln Ile Ser

1505 1510 1515

Glu Phe Ser Lys Arg Val Ile Leu Ala Asp Ala Asn Leu Asp Lys

1520 1525 1530

Val Leu Ser Ala Tyr Asn Lys His Arg Asp Lys Pro Ile Arg Glu

1535 1540 1545

Gln Ala Glu Asn Ile Ile His Leu Phe Thr Leu Thr Asn Leu Gly

1550 1555 1560

Ala Pro Ala Ala Phe Lys Tyr Phe Asp Thr Thr Ile Asp Arg Lys

1565 1570 1575

Arg Tyr Thr Ser Thr Lys Glu Val Leu Asp Ala Thr Leu Ile His

1580 1585 1590

Gln Ser Ile Thr Gly Leu Tyr Glu Thr Arg Ile Asp Leu Ser Gln

1595 1600 1605

Leu Gly Gly Asp Ser Gly Gly Ser Thr Asn Leu Ser Asp Ile Ile
1610 1615 1620
Glu Lys Glu Thr Gly Lys Gln Leu Val Ile Gln Glu Ser Ile Leu
1625 1630 1635
Met Leu Pro Glu Glu Val Glu Glu Val Ile Gly Asn Lys Pro Glu
1640 1645 1650
Ser Asp Ile Leu Val His Thr Ala Tyr Asp Glu Ser Thr Asp Glu
1655 1660 1665
Asn Val Met Leu Leu Thr Ser Asp Ala Pro Glu Tyr Lys Pro Trp
1670 1675 1680
Ala Leu Val Ile Gln Asp Ser Asn Gly Glu Asn Lys Ile Lys Met
1685 1690 1695
Leu Ser Gly Gly Ser Pro Lys Lys Lys Arg Lys Val
1700 1705 1710

<210> 116
<211> 1612
<212> PRT
<213> 人工序列
<220>
<223> 合成多肽
<400> 116

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

Thr Glu Gln Glu Ser Gly Tyr Cys Trp Arg Asn Phe Val Asn Tyr Ser
145 150 155 160
Pro Ser Asn Glu Ala His Trp Pro Arg Tyr Pro His Leu Trp Val Arg
165 170 175
Leu Tyr Val Leu Glu Leu Tyr Cys Ile Ile Leu Gly Leu Pro Pro Cys
180 185 190
Leu Asn Ile Leu Arg Arg Lys Gln Pro Gln Leu Thr Phe Phe Thr Ile
195 200 205
Ala Leu Gln Ser Cys His Tyr Gln Arg Leu Pro Pro His Ile Leu Trp
210 215 220
Ala Thr Gly Leu Lys Ser Gly Ser Glu Thr Pro Gly Thr Ser Glu Ser
225 230 235 240
Ala Thr Pro Glu Ser Asp Lys Lys Tyr Ser Ile Gly Leu Ala Ile Gly
245 250 255
Thr Asn Ser Val Gly Trp Ala Val Ile Thr Asp Glu Tyr Lys Val Pro
260 265 270
Ser Lys Lys Phe Lys Val Leu Gly Asn Thr Asp Arg His Ser Ile Lys
275 280 285
Lys Asn Leu Ile Gly Ala Leu Leu Phe Asp Ser Gly Glu Thr Ala Glu
290 295 300
Ala Thr Arg Leu Lys Arg Thr Ala Arg Arg Arg Tyr Thr Arg Arg Lys
305 310 315 320
Asn Arg Ile Cys Tyr Leu Gln Glu Ile Phe Ser Asn Glu Met Ala Lys
325 330 335
Val Asp Asp Ser Phe Phe His Arg Leu Glu Glu Ser Phe Leu Val Glu
340 345 350
Glu Asp Lys Lys His Glu Arg His Pro Ile Phe Gly Asn Ile Val Asp
355 360 365
Glu Val Ala Tyr His Glu Lys Tyr Pro Thr Ile Tyr His Leu Arg Lys
370 375 380
Lys Leu Val Asp Ser Thr Asp Lys Ala Asp Leu Arg Leu Ile Tyr Leu
385 390 395 400
Ala Leu Ala His Met Ile Lys Phe Arg Gly His Phe Leu Ile Glu Gly
405 410 415
Asp Leu Asn Pro Asp Asn Ser Asp Val Asp Lys Leu Phe Ile Gln Leu
420 425 430
Val Gln Thr Tyr Asn Gln Leu Phe Glu Glu Asn Pro Ile Asn Ala Ser
435 440 445
Gly Val Asp Ala Lys Ala Ile Leu Ser Ala Arg Leu Ser Lys Ser Arg

450 455 460
Arg Leu Glu Asn Leu Ile Ala Gln Leu Pro Gly Glu Lys Lys Asn Gly
465 470 475 480
Leu Phe Gly Asn Leu Ile Ala Leu Ser Leu Gly Leu Thr Pro Asn Phe
485 490 495
Lys Ser Asn Phe Asp Leu Ala Glu Asp Ala Lys Leu Gln Leu Ser Lys
500 505 510
Asp Thr Tyr Asp Asp Asp Leu Asp Asn Leu Leu Ala Gln Ile Gly Asp
515 520 525
Gln Tyr Ala Asp Leu Phe Leu Ala Ala Lys Asn Leu Ser Asp Ala Ile
530 535 540
Leu Leu Ser Asp Ile Leu Arg Val Asn Thr Glu Ile Thr Lys Ala Pro
545 550 555 560
Leu Ser Ala Ser Met Ile Lys Arg Tyr Asp Glu His His Gln Asp Leu
565 570 575
Thr Leu Leu Lys Ala Leu Val Arg Gln Gln Leu Pro Glu Lys Tyr Lys
580 585 590
Glu Ile Phe Phe Asp Gln Ser Lys Asn Gly Tyr Ala Gly Tyr Ile Asp
595 600 605
Gly Gly Ala Ser Gln Glu Glu Phe Tyr Lys Phe Ile Lys Pro Ile Leu
610 615 620
Glu Lys Met Asp Gly Thr Glu Glu Leu Leu Val Lys Leu Asn Arg Glu
625 630 635 640
Asp Leu Leu Arg Lys Gln Arg Thr Phe Asp Asn Gly Ser Ile Pro His
645 650 655
Gln Ile His Leu Gly Glu Leu His Ala Ile Leu Arg Arg Gln Glu Asp
660 665 670
Phe Tyr Pro Phe Leu Lys Asp Asn Arg Glu Lys Ile Glu Lys Ile Leu
675 680 685
Thr Phe Arg Ile Pro Tyr Tyr Val Gly Pro Leu Ala Arg Gly Asn Ser
690 695 700
Arg Phe Ala Trp Met Thr Arg Lys Ser Glu Glu Thr Ile Thr Pro Trp
705 710 715 720
Asn Phe Glu Glu Val Val Asp Lys Gly Ala Ser Ala Gln Ser Phe Ile
725 730 735
Glu Arg Met Thr Asn Phe Asp Lys Asn Leu Pro Asn Glu Lys Val Leu
740 745 750
Pro Lys His Ser Leu Leu Tyr Glu Tyr Phe Thr Val Tyr Asn Glu Leu
755 760 765

Thr Lys Val Lys Tyr Val Thr Glu Gly Met Arg Lys Pro Ala Phe Leu
770 775 780
Ser Gly Glu Gln Lys Lys Ala Ile Val Asp Leu Leu Phe Lys Thr Asn
785 790 795 800
Arg Lys Val Thr Val Lys Gln Leu Lys Glu Asp Tyr Phe Lys Lys Ile
805 810 815
Glu Cys Phe Asp Ser Val Glu Ile Ser Gly Val Glu Asp Arg Phe Asn
820 825 830
Ala Ser Leu Gly Thr Tyr His Asp Leu Leu Lys Ile Ile Lys Asp Lys
835 840 845
Asp Phe Leu Asp Asn Glu Glu Asn Glu Asp Ile Leu Glu Asp Ile Val
850 855 860
Leu Thr Leu Thr Leu Phe Glu Asp Arg Glu Met Ile Glu Glu Arg Leu
865 870 875 880
Lys Thr Tyr Ala His Leu Phe Asp Asp Lys Val Met Lys Gln Leu Lys
885 890 895
Arg Arg Arg Tyr Thr Gly Trp Gly Arg Leu Ser Arg Lys Leu Ile Asn
900 905 910
Gly Ile Arg Asp Lys Gln Ser Gly Lys Thr Ile Leu Asp Phe Leu Lys
915 920 925
Ser Asp Gly Phe Ala Asn Arg Asn Phe Met Gln Leu Ile His Asp Asp
930 935 940
Ser Leu Thr Phe Lys Glu Asp Ile Gln Lys Ala Gln Val Ser Gly Gln
945 950 955 960
Gly Asp Ser Leu His Glu His Ile Ala Asn Leu Ala Gly Ser Pro Ala
965 970 975
Ile Lys Lys Gly Ile Leu Gln Thr Val Lys Val Val Asp Glu Leu Val
980 985 990
Lys Val Met Gly Arg His Lys Pro Glu Asn Ile Val Ile Glu Met Ala
995 1000 1005
Arg Glu Asn Gln Thr Thr Gln Lys Gly Gln Lys Asn Ser Arg Glu
1010 1015 1020
Arg Met Lys Arg Ile Glu Glu Gly Ile Lys Glu Leu Gly Ser Gln
1025 1030 1035
Ile Leu Lys Glu His Pro Val Glu Asn Thr Gln Leu Gln Asn Glu
1040 1045 1050
Lys Leu Tyr Leu Tyr Tyr Leu Gln Asn Gly Arg Asp Met Tyr Val
1055 1060 1065
Asp Gln Glu Leu Asp Ile Asn Arg Leu Ser Asp Tyr Asp Val Asp

1070 1075 1080

His Ile Val Pro Gln Ser Phe Leu Lys Asp Asp Ser Ile Asp Asn

1085 1090 1095

Lys Val Leu Thr Arg Ser Asp Lys Asn Arg Gly Lys Ser Asp Asn

1100 1105 1110

Val Pro Ser Glu Glu Val Val Lys Lys Met Lys Asn Tyr Trp Arg

1115 1120 1125

Gln Leu Leu Asn Ala Lys Leu Ile Thr Gln Arg Lys Phe Asp Asn

1130 1135 1140

Leu Thr Lys Ala Glu Arg Gly Gly Leu Ser Glu Leu Asp Lys Ala

1145 1150 1155

Gly Phe Ile Lys Arg Gln Leu Val Glu Thr Arg Gln Ile Thr Lys

1160 1165 1170

His Val Ala Gln Ile Leu Asp Ser Arg Met Asn Thr Lys Tyr Asp

1175 1180 1185

Glu Asn Asp Lys Leu Ile Arg Glu Val Lys Val Ile Thr Leu Lys

1190 1195 1200

Ser Lys Leu Val Ser Asp Phe Arg Lys Asp Phe Gln Phe Tyr Lys

1205 1210 1215

Val Arg Glu Ile Asn Asn Tyr His His Ala His Asp Ala Tyr Leu

1220 1225 1230

Asn Ala Val Val Gly Thr Ala Leu Ile Lys Lys Tyr Pro Lys Leu

1235 1240 1245

Glu Ser Glu Phe Val Tyr Gly Asp Tyr Lys Val Tyr Asp Val Arg

1250 1255 1260

Lys Met Ile Ala Lys Ser Glu Gln Glu Ile Gly Lys Ala Thr Ala

1265 1270 1275

Lys Tyr Phe Phe Tyr Ser Asn Ile Met Asn Phe Phe Lys Thr Glu

1280 1285 1290

Ile Thr Leu Ala Asn Gly Glu Ile Arg Lys Arg Pro Leu Ile Glu

1295 1300 1305

Thr Asn Gly Glu Thr Gly Glu Ile Val Trp Asp Lys Gly Arg Asp

1310 1315 1320

Phe Ala Thr Val Arg Lys Val Leu Ser Met Pro Gln Val Asn Ile

1325 1330 1335

Val Lys Lys Thr Glu Val Gln Thr Gly Gly Phe Ser Lys Glu Ser

1340 1345 1350

Ile Leu Pro Lys Arg Asn Ser Asp Lys Leu Ile Ala Arg Lys Lys

1355 1360 1365

Asp Trp Asp Pro Lys Lys Tyr Gly Gly Phe Asp Ser Pro Thr Val
1370 1375 1380
Ala Tyr Ser Val Leu Val Val Ala Lys Val Glu Lys Gly Lys Ser
1385 1390 1395
Lys Lys Leu Lys Ser Val Lys Glu Leu Leu Gly Ile Thr Ile Met
1400 1405 1410
Glu Arg Ser Ser Phe Glu Lys Asn Pro Ile Asp Phe Leu Glu Ala
1415 1420 1425
Lys Gly Tyr Lys Glu Val Lys Lys Asp Leu Ile Ile Lys Leu Pro
1430 1435 1440
Lys Tyr Ser Leu Phe Glu Leu Glu Asn Gly Arg Lys Arg Met Leu
1445 1450 1455
Ala Ser Ala Gly Glu Leu Gln Lys Gly Asn Glu Leu Ala Leu Pro
1460 1465 1470
Ser Lys Tyr Val Asn Phe Leu Tyr Leu Ala Ser His Tyr Glu Lys
1475 1480 1485
Leu Lys Gly Ser Pro Glu Asp Asn Glu Gln Lys Gln Leu Phe Val
1490 1495 1500
Glu Gln His Lys His Tyr Leu Asp Glu Ile Ile Glu Gln Ile Ser
1505 1510 1515
Glu Phe Ser Lys Arg Val Ile Leu Ala Asp Ala Asn Leu Asp Lys
1520 1525 1530
Val Leu Ser Ala Tyr Asn Lys His Arg Asp Lys Pro Ile Arg Glu
1535 1540 1545
Gln Ala Glu Asn Ile Ile His Leu Phe Thr Leu Thr Asn Leu Gly
1550 1555 1560
Ala Pro Ala Ala Phe Lys Tyr Phe Asp Thr Thr Ile Asp Arg Lys
1565 1570 1575
Arg Tyr Thr Ser Thr Lys Glu Val Leu Asp Ala Thr Leu Ile His
1580 1585 1590
Gln Ser Ile Thr Gly Leu Tyr Glu Thr Arg Ile Asp Leu Ser Gln
1595 1600 1605
Leu Gly Gly Asp
1610

<210> 117
<211> 1929
<212> PRT
<213> 人工序列
<220>

<223> 合成多肽

<400> 117

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

290 295 300

Ala Thr Arg Leu Lys Arg Thr Ala Arg Arg Arg Tyr Thr Arg Arg Lys
305 310 315 320
Asn Arg Ile Cys Tyr Leu Gln Glu Ile Phe Ser Asn Glu Met Ala Lys
325 330 335
Val Asp Asp Ser Phe Phe His Arg Leu Glu Glu Ser Phe Leu Val Glu
340 345 350
Glu Asp Lys Lys His Glu Arg His Pro Ile Phe Gly Asn Ile Val Asp
355 360 365
Glu Val Ala Tyr His Glu Lys Tyr Pro Thr Ile Tyr His Leu Arg Lys
370 375 380
Lys Leu Val Asp Ser Thr Asp Lys Ala Asp Leu Arg Leu Ile Tyr Leu
385 390 395 400
Ala Leu Ala His Met Ile Lys Phe Arg Gly His Phe Leu Ile Glu Gly
405 410 415
Asp Leu Asn Pro Asp Asn Ser Asp Val Asp Lys Leu Phe Ile Gln Leu
420 425 430
Val Gln Thr Tyr Asn Gln Leu Phe Glu Glu Asn Pro Ile Asn Ala Ser
435 440 445
Gly Val Asp Ala Lys Ala Ile Leu Ser Ala Arg Leu Ser Lys Ser Arg
450 455 460
Arg Leu Glu Asn Leu Ile Ala Gln Leu Pro Gly Glu Lys Lys Asn Gly
465 470 475 480
Leu Phe Gly Asn Leu Ile Ala Leu Ser Leu Gly Leu Thr Pro Asn Phe
485 490 495
Lys Ser Asn Phe Asp Leu Ala Glu Asp Ala Lys Leu Gln Leu Ser Lys
500 505 510
Asp Thr Tyr Asp Asp Asp Leu Asp Asn Leu Leu Ala Gln Ile Gly Asp
515 520 525
Gln Tyr Ala Asp Leu Phe Leu Ala Ala Lys Asn Leu Ser Asp Ala Ile
530 535 540
Leu Leu Ser Asp Ile Leu Arg Val Asn Thr Glu Ile Thr Lys Ala Pro
545 550 555 560
Leu Ser Ala Ser Met Ile Lys Arg Tyr Asp Glu His His Gln Asp Leu
565 570 575
Thr Leu Leu Lys Ala Leu Val Arg Gln Gln Leu Pro Glu Lys Tyr Lys
580 585 590
Glu Ile Phe Phe Asp Gln Ser Lys Asn Gly Tyr Ala Gly Tyr Ile Asp
595 600 605

Gly Gly Ala Ser Gln Glu Glu Phe Tyr Lys Phe Ile Lys Pro Ile Leu
610 615 620
Glu Lys Met Asp Gly Thr Glu Glu Leu Leu Val Lys Leu Asn Arg Glu
625 630 635 640
Asp Leu Leu Arg Lys Gln Arg Thr Phe Asp Asn Gly Ser Ile Pro His
645 650 655
Gln Ile His Leu Gly Glu Leu His Ala Ile Leu Arg Arg Gln Glu Asp
660 665 670
Phe Tyr Pro Phe Leu Lys Asp Asn Arg Glu Lys Ile Glu Lys Ile Leu
675 680 685
Thr Phe Arg Ile Pro Tyr Tyr Val Gly Pro Leu Ala Arg Gly Asn Ser
690 695 700
Arg Phe Ala Trp Met Thr Arg Lys Ser Glu Glu Thr Ile Thr Pro Trp
705 710 715 720
Asn Phe Glu Glu Val Val Asp Lys Gly Ala Ser Ala Gln Ser Phe Ile
725 730 735
Glu Arg Met Thr Asn Phe Asp Lys Asn Leu Pro Asn Glu Lys Val Leu
740 745 750
Pro Lys His Ser Leu Leu Tyr Glu Tyr Phe Thr Val Tyr Asn Glu Leu
755 760 765
Thr Lys Val Lys Tyr Val Thr Glu Gly Met Arg Lys Pro Ala Phe Leu
770 775 780
Ser Gly Glu Gln Lys Lys Ala Ile Val Asp Leu Leu Phe Lys Thr Asn
785 790 795 800
Arg Lys Val Thr Val Lys Gln Leu Lys Glu Asp Tyr Phe Lys Lys Ile
805 810 815
Glu Cys Phe Asp Ser Val Glu Ile Ser Gly Val Glu Asp Arg Phe Asn
820 825 830
Ala Ser Leu Gly Thr Tyr His Asp Leu Leu Lys Ile Ile Lys Asp Lys
835 840 845
Asp Phe Leu Asp Asn Glu Glu Asn Glu Asp Ile Leu Glu Asp Ile Val
850 855 860
Leu Thr Leu Thr Leu Phe Glu Asp Arg Glu Met Ile Glu Glu Arg Leu
865 870 875 880
Lys Thr Tyr Ala His Leu Phe Asp Asp Lys Val Met Lys Gln Leu Lys
885 890 895
Arg Arg Arg Tyr Thr Gly Trp Gly Arg Leu Ser Arg Lys Leu Ile Asn
900 905 910
Gly Ile Arg Asp Lys Gln Ser Gly Lys Thr Ile Leu Asp Phe Leu Lys

915 920 925

Ser Asp Gly Phe Ala Asn Arg Asn Phe Met Gln Leu Ile His Asp Asp

930 935 940

Ser Leu Thr Phe Lys Glu Asp Ile Gln Lys Ala Gln Val Ser Gly Gln

945 950 955 960

Gly Asp Ser Leu His Glu His Ile Ala Asn Leu Ala Gly Ser Pro Ala

965 970 975

Ile Lys Lys Gly Ile Leu Gln Thr Val Lys Val Val Asp Glu Leu Val

980 985 990

Lys Val Met Gly Arg His Lys Pro Glu Asn Ile Val Ile Glu Met Ala

995 1000 1005

Arg Glu Asn Gln Thr Thr Gln Lys Gly Gln Lys Asn Ser Arg Glu

1010 1015 1020

Arg Met Lys Arg Ile Glu Glu Gly Ile Lys Glu Leu Gly Ser Gln

1025 1030 1035

Ile Leu Lys Glu His Pro Val Glu Asn Thr Gln Leu Gln Asn Glu

1040 1045 1050

Lys Leu Tyr Leu Tyr Tyr Leu Gln Asn Gly Arg Asp Met Tyr Val

1055 1060 1065

Asp Gln Glu Leu Asp Ile Asn Arg Leu Ser Asp Tyr Asp Val Asp

1070 1075 1080

His Ile Val Pro Gln Ser Phe Leu Lys Asp Asp Ser Ile Asp Asn

1085 1090 1095

Lys Val Leu Thr Arg Ser Asp Lys Asn Arg Gly Lys Ser Asp Asn

1100 1105 1110

Val Pro Ser Glu Glu Val Val Lys Lys Met Lys Asn Tyr Trp Arg

1115 1120 1125

Gln Leu Leu Asn Ala Lys Leu Ile Thr Gln Arg Lys Phe Asp Asn

1130 1135 1140

Leu Thr Lys Ala Glu Arg Gly Gly Leu Ser Glu Leu Asp Lys Ala

1145 1150 1155

Gly Phe Ile Lys Arg Gln Leu Val Glu Thr Arg Gln Ile Thr Lys

1160 1165 1170

His Val Ala Gln Ile Leu Asp Ser Arg Met Asn Thr Lys Tyr Asp

1175 1180 1185

Glu Asn Asp Lys Leu Ile Arg Glu Val Lys Val Ile Thr Leu Lys

1190 1195 1200

Ser Lys Leu Val Ser Asp Phe Arg Lys Asp Phe Gln Phe Tyr Lys

1205 1210 1215

Val Arg Glu Ile Asn Asn Tyr His His Ala His Asp Ala Tyr Leu
1220 1225 1230
Asn Ala Val Val Gly Thr Ala Leu Ile Lys Lys Tyr Pro Lys Leu
1235 1240 1245
Glu Ser Glu Phe Val Tyr Gly Asp Tyr Lys Val Tyr Asp Val Arg
1250 1255 1260
Lys Met Ile Ala Lys Ser Glu Gln Glu Ile Gly Lys Ala Thr Ala
1265 1270 1275
Lys Tyr Phe Phe Tyr Ser Asn Ile Met Asn Phe Phe Lys Thr Glu
1280 1285 1290
Ile Thr Leu Ala Asn Gly Glu Ile Arg Lys Arg Pro Leu Ile Glu
1295 1300 1305
Thr Asn Gly Glu Thr Gly Glu Ile Val Trp Asp Lys Gly Arg Asp
1310 1315 1320
Phe Ala Thr Val Arg Lys Val Leu Ser Met Pro Gln Val Asn Ile
1325 1330 1335
Val Lys Lys Thr Glu Val Gln Thr Gly Gly Phe Ser Lys Glu Ser
1340 1345 1350
Ile Leu Pro Lys Arg Asn Ser Asp Lys Leu Gln Asn Glu Lys Leu
1355 1360 1365
Tyr Leu Tyr Tyr Leu Gln Asn Gly Arg Asp Met Tyr Val Asp Gln
1370 1375 1380
Glu Leu Asp Ile Asn Arg Leu Ser Asp Tyr Asp Val Asp His Ile
1385 1390 1395
Val Pro Gln Ser Phe Leu Lys Asp Asp Ser Ile Asp Asn Lys Val
1400 1405 1410
Leu Thr Arg Ser Asp Lys Asn Arg Gly Lys Ser Asp Asn Val Pro
1415 1420 1425
Ser Glu Glu Val Val Lys Lys Met Lys Asn Tyr Trp Arg Gln Leu
1430 1435 1440
Leu Asn Ala Lys Leu Ile Thr Gln Arg Lys Phe Asp Asn Leu Thr
1445 1450 1455
Lys Ala Glu Arg Gly Gly Leu Ser Glu Leu Asp Lys Ala Gly Phe
1460 1465 1470
Ile Lys Arg Gln Leu Val Glu Thr Arg Gln Ile Thr Lys His Val
1475 1480 1485
Ala Gln Ile Leu Asp Ser Arg Met Asn Thr Lys Tyr Asp Glu Asn
1490 1495 1500
Asp Lys Leu Ile Arg Glu Val Lys Val Ile Thr Leu Lys Ser Lys

1505 1510 1515

Leu Val Ser Asp Phe Arg Lys Asp Phe Gln Phe Tyr Lys Val Arg
1520 1525 1530

Glu Ile Asn Asn Tyr His His Ala His Asp Ala Tyr Leu Asn Ala
1535 1540 1545

Val Val Gly Thr Ala Leu Ile Lys Lys Tyr Pro Lys Leu Glu Ser
1550 1555 1560

Glu Phe Val Tyr Gly Asp Tyr Lys Val Tyr Asp Val Arg Lys Met
1565 1570 1575

Ile Ala Lys Ser Glu Gln Glu Ile Gly Lys Ala Thr Ala Lys Tyr
1580 1585 1590

Phe Phe Tyr Ser Asn Ile Met Asn Phe Phe Lys Thr Glu Ile Thr
1595 1600 1605

Leu Ala Asn Gly Glu Ile Arg Lys Arg Pro Leu Ile Glu Thr Asn
1610 1615 1620

Gly Glu Thr Gly Glu Ile Val Trp Asp Lys Gly Arg Asp Phe Ala
1625 1630 1635

Thr Val Arg Lys Val Leu Ser Met Pro Gln Val Asn Ile Val Lys
1640 1645 1650

Lys Thr Glu Val Gln Thr Gly Gly Phe Ser Lys Glu Ser Ile Leu
1655 1660 1665

Pro Lys Arg Asn Ser Asp Lys Leu Ile Ala Arg Lys Lys Asp Trp
1670 1675 1680

Asp Pro Lys Lys Tyr Gly Gly Phe Asp Ser Pro Thr Val Ala Tyr
1685 1690 1695

Ser Val Leu Val Val Ala Lys Val Glu Lys Gly Lys Ser Lys Lys
1700 1705 1710

Leu Lys Ser Val Lys Glu Leu Leu Gly Ile Thr Ile Met Glu Arg
1715 1720 1725

Ser Ser Phe Glu Lys Asn Pro Ile Asp Phe Leu Glu Ala Lys Gly
1730 1735 1740

Tyr Lys Glu Val Lys Lys Asp Leu Ile Ile Lys Leu Pro Lys Tyr
1745 1750 1755

Ser Leu Phe Glu Leu Glu Asn Gly Arg Lys Arg Met Leu Ala Ser
1760 1765 1770

Ala Gly Glu Leu Gln Lys Gly Asn Glu Leu Ala Leu Pro Ser Lys
1775 1780 1785

Tyr Val Asn Phe Leu Tyr Leu Ala Ser His Tyr Glu Lys Leu Lys
1790 1795 1800

Gly Ser Pro Glu Asp Asn Glu Gln Lys Gln Leu Phe Val Glu Gln
1805 1810 1815
His Lys His Tyr Leu Asp Glu Ile Ile Glu Gln Ile Ser Glu Phe
1820 1825 1830
Ser Lys Arg Val Ile Leu Ala Asp Ala Asn Leu Asp Lys Val Leu
1835 1840 1845
Ser Ala Tyr Asn Lys His Arg Asp Lys Pro Ile Arg Glu Gln Ala
1850 1855 1860
Glu Asn Ile Ile His Leu Phe Thr Leu Thr Asn Leu Gly Ala Pro
1865 1870 1875
Ala Ala Phe Lys Tyr Phe Asp Thr Thr Ile Asp Arg Lys Arg Tyr
1880 1885 1890
Thr Ser Thr Lys Glu Val Leu Asp Ala Thr Leu Ile His Gln Ser
1895 1900 1905
Ile Thr Gly Leu Tyr Glu Thr Arg Ile Asp Leu Ser Gln Leu Gly
1910 1915 1920
Gly Asp Ser Gly Gly Ser
1925

<210> 118
<211> 1385
<212> PRT
<213> 人工序列
<220>
<223> 合成多肽
<400> 118

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

Arg Leu Glu Glu Ser Phe Leu Val Glu Glu Asp Lys Lys His Glu Arg
115 120 125
His Pro Ile Phe Gly Asn Ile Val Asp Glu Val Ala Tyr His Glu Lys
130 135 140
Tyr Pro Thr Ile Tyr His Leu Arg Lys Lys Leu Val Asp Ser Thr Asp
145 150 155 160
Lys Ala Asp Leu Arg Leu Ile Tyr Leu Ala Leu Ala His Met Ile Lys
165 170 175
Phe Arg Gly His Phe Leu Ile Glu Gly Asp Leu Asn Pro Asp Asn Ser
180 185 190
Asp Val Asp Lys Leu Phe Ile Gln Leu Val Gln Thr Tyr Asn Gln Leu
195 200 205
Phe Glu Glu Asn Pro Ile Asn Ala Ser Gly Val Asp Ala Lys Ala Ile
210 215 220
Leu Ser Ala Arg Leu Ser Lys Ser Arg Arg Leu Glu Asn Leu Ile Ala
225 230 235 240
Gln Leu Pro Gly Glu Lys Lys Asn Gly Leu Phe Gly Asn Leu Ile Ala
245 250 255
Leu Ser Leu Gly Leu Thr Pro Asn Phe Lys Ser Asn Phe Asp Leu Ala
260 265 270
Glu Asp Ala Lys Leu Gln Leu Ser Lys Asp Thr Tyr Asp Asp Asp Leu
275 280 285
Asp Asn Leu Leu Ala Gln Ile Gly Asp Gln Tyr Ala Asp Leu Phe Leu
290 295 300
Ala Ala Lys Asn Leu Ser Asp Ala Ile Leu Leu Ser Asp Ile Leu Arg
305 310 315 320
Val Asn Thr Glu Ile Thr Lys Ala Pro Leu Ser Ala Ser Met Ile Lys
325 330 335
Arg Tyr Asp Glu His His Gln Asp Leu Thr Leu Leu Lys Ala Leu Val
340 345 350
Arg Gln Gln Leu Pro Glu Lys Tyr Lys Glu Ile Phe Phe Asp Gln Ser
355 360 365
Lys Asn Gly Tyr Ala Gly Tyr Ile Asp Gly Gly Ala Ser Gln Glu Glu
370 375 380
Phe Tyr Lys Phe Ile Lys Pro Ile Leu Glu Lys Met Asp Gly Thr Glu
385 390 395 400
Glu Leu Leu Val Lys Leu Asn Arg Glu Asp Leu Leu Arg Lys Gln Arg
405 410 415
Thr Phe Asp Asn Gly Ser Ile Pro His Gln Ile His Leu Gly Glu Leu

420 425 430

His Ala Ile Leu Arg Arg Gln Glu Asp Phe Tyr Pro Phe Leu Lys Asp
435 440 445

Asn Arg Glu Lys Ile Glu Lys Ile Leu Thr Phe Arg Ile Pro Tyr Tyr
450 455 460

Val Gly Pro Leu Ala Arg Gly Asn Ser Arg Phe Ala Trp Met Thr Arg
465 470 475 480

Lys Ser Glu Glu Thr Ile Thr Pro Trp Asn Phe Glu Glu Val Val Asp
485 490 495

Lys Gly Ala Ser Ala Gln Ser Phe Ile Glu Arg Met Thr Asn Phe Asp
500 505 510

Lys Asn Leu Pro Asn Glu Lys Val Leu Pro Lys His Ser Leu Leu Tyr
515 520 525

Glu Tyr Phe Thr Val Tyr Asn Glu Leu Thr Lys Val Lys Tyr Val Thr
530 535 540

Glu Gly Met Arg Lys Pro Ala Phe Leu Ser Gly Glu Gln Lys Lys Ala
545 550 555 560

Ile Val Asp Leu Leu Phe Lys Thr Asn Arg Lys Val Thr Val Lys Gln
565 570 575

Leu Lys Glu Asp Tyr Phe Lys Lys Ile Glu Cys Phe Asp Ser Val Glu
580 585 590

Ile Ser Gly Val Glu Asp Arg Phe Asn Ala Ser Leu Gly Thr Tyr His
595 600 605

Asp Leu Leu Lys Ile Ile Lys Asp Lys Asp Phe Leu Asp Asn Glu Glu
610 615 620

Asn Glu Asp Ile Leu Glu Asp Ile Val Leu Thr Leu Thr Leu Phe Glu
625 630 635 640

Asp Arg Glu Met Ile Glu Glu Arg Leu Lys Thr Tyr Ala His Leu Phe
645 650 655

Asp Asp Lys Val Met Lys Gln Leu Lys Arg Arg Arg Tyr Thr Gly Trp
660 665 670

Gly Arg Leu Ser Arg Lys Leu Ile Asn Gly Ile Arg Asp Lys Gln Ser
675 680 685

Gly Lys Thr Ile Leu Asp Phe Leu Lys Ser Asp Gly Phe Ala Asn Arg
690 695 700

Asn Phe Met Gln Leu Ile His Asp Asp Ser Leu Thr Phe Lys Glu Asp
705 710 715 720

Ile Gln Lys Ala Gln Val Ser Gly Gln Gly Asp Ser Leu His Glu His
725 730 735

Ile Ala Asn Leu Ala Gly Ser Pro Ala Ile Lys Lys Gly Ile Leu Gln
740 745 750
Thr Val Lys Val Val Asp Glu Leu Val Lys Val Met Gly Arg His Lys
755 760 765
Pro Glu Asn Ile Val Ile Glu Met Ala Arg Glu Asn Gln Thr Thr Gln
770 775 780
Lys Gly Gln Lys Asn Ser Arg Glu Arg Met Lys Arg Ile Glu Glu Gly
785 790 795 800
Ile Lys Glu Leu Gly Ser Gln Ile Leu Lys Glu His Pro Val Glu Asn
805 810 815
Thr Gln Leu Gln Asn Glu Lys Leu Tyr Leu Tyr Tyr Leu Gln Asn Gly
820 825 830
Arg Asp Met Tyr Val Asp Gln Glu Leu Asp Ile Asn Arg Leu Ser Asp
835 840 845
Tyr Asp Val Asp His Ile Val Pro Gln Ser Phe Leu Lys Asp Asp Ser
850 855 860
Ile Asp Asn Lys Val Leu Thr Arg Ser Asp Lys Asn Arg Gly Lys Ser
865 870 875 880
Asp Asn Val Pro Ser Glu Glu Val Val Lys Lys Met Lys Asn Tyr Trp
885 890 895
Arg Gln Leu Leu Asn Ala Lys Leu Ile Thr Gln Arg Lys Phe Asp Asn
900 905 910
Leu Thr Lys Ala Glu Arg Gly Gly Leu Ser Glu Leu Asp Lys Ala Gly
915 920 925
Phe Ile Lys Arg Gln Leu Val Glu Thr Arg Gln Ile Thr Lys His Val
930 935 940
Ala Gln Ile Leu Asp Ser Arg Met Asn Thr Lys Tyr Asp Glu Asn Asp
945 950 955 960
Lys Leu Ile Arg Glu Val Lys Val Ile Thr Leu Lys Ser Lys Leu Val
965 970 975
Ser Asp Phe Arg Lys Asp Phe Gln Phe Tyr Lys Val Arg Glu Ile Asn
980 985 990
Asn Tyr His His Ala His Asp Ala Tyr Leu Asn Ala Val Val Gly Thr
995 1000 1005
Ala Leu Ile Lys Lys Tyr Pro Lys Leu Glu Ser Glu Phe Val Tyr
1010 1015 1020
Gly Asp Tyr Lys Val Tyr Asp Val Arg Lys Met Ile Ala Lys Ser
1025 1030 1035
Glu Gln Glu Ile Gly Lys Ala Thr Ala Lys Tyr Phe Phe Tyr Ser

1040 1045 1050

Asn Ile Met Asn Phe Phe Lys Thr Glu Ile Thr Leu Ala Asn Gly
1055 1060 1065

Glu Ile Arg Lys Arg Pro Leu Ile Glu Thr Asn Gly Glu Thr Gly
1070 1075 1080

Glu Ile Val Trp Asp Lys Gly Arg Asp Phe Ala Thr Val Arg Lys
1085 1090 1095

Val Leu Ser Met Pro Gln Val Asn Ile Val Lys Lys Thr Glu Val
1100 1105 1110

Gln Thr Gly Gly Phe Ser Lys Glu Ser Ile Leu Pro Lys Arg Asn
1115 1120 1125

Ser Asp Lys Leu Ile Ala Arg Lys Lys Asp Trp Asp Pro Lys Lys
1130 1135 1140

Tyr Gly Gly Phe Asp Ser Pro Thr Val Ala Tyr Ser Val Leu Val
1145 1150 1155

Val Ala Lys Val Glu Lys Gly Lys Ser Lys Lys Leu Lys Ser Val
1160 1165 1170

Lys Glu Leu Leu Gly Ile Thr Ile Met Glu Arg Ser Ser Phe Glu
1175 1180 1185

Lys Asn Pro Ile Asp Phe Leu Glu Ala Lys Gly Tyr Lys Glu Val
1190 1195 1200

Lys Lys Asp Leu Ile Ile Lys Leu Pro Lys Tyr Ser Leu Phe Glu
1205 1210 1215

Leu Glu Asn Gly Arg Lys Arg Met Leu Ala Ser Ala Gly Glu Leu
1220 1225 1230

Gln Lys Gly Asn Glu Leu Ala Leu Pro Ser Lys Tyr Val Asn Phe
1235 1240 1245

Leu Tyr Leu Ala Ser His Tyr Glu Lys Leu Lys Gly Ser Pro Glu
1250 1255 1260

Asp Asn Glu Gln Lys Gln Leu Phe Val Glu Gln His Lys His Tyr
1265 1270 1275

Leu Asp Glu Ile Ile Glu Gln Ile Ser Glu Phe Ser Lys Arg Val
1280 1285 1290

Ile Leu Ala Asp Ala Asn Leu Asp Lys Val Leu Ser Ala Tyr Asn
1295 1300 1305

Lys His Arg Asp Lys Pro Ile Arg Glu Gln Ala Glu Asn Ile Ile
1310 1315 1320

His Leu Phe Thr Leu Thr Asn Leu Gly Ala Pro Ala Ala Phe Lys
1325 1330 1335

Tyr Phe Asp Thr Thr Ile Asp Arg Lys Arg Tyr Thr Ser Thr Lys
1340 1345 1350
Glu Val Leu Asp Ala Thr Leu Ile His Gln Ser Ile Thr Gly Leu
1355 1360 1365
Tyr Glu Thr Arg Ile Asp Leu Ser Gln Leu Gly Gly Asp Ser Gly
1370 1375 1380
Gly Ser
1385

<210> 119
<211> 82
<212> RNA
<213> 人工序列
<220>
<223> 合成多核苷酸
<400> 119
guuuuagagc uagaaauagc aaguuaaaau aaaggcuagu ccguuaucaa cuugaaaaag 60
uggcaccgag ucggugcuuu uu 82

<210> 120
<211> 10
<212> PRT
<213> 人工序列
<220>
<223> 合成多肽
<400> 120
Ser Gly Gly Ser Gly Gly Ser Gly Gly Ser
1 5 10

<210> 121
<211> 30
<212> PRT
<213> 人工序列
<220>
<223> 合成多肽
<220>
<221> MISC_FEATURE
<222> (2)..(30)
<223> 可以不存在
<400> 121
Gly Gly Gly Gly Gly Gly Gly Gly Gly Gly Gly Gly Gly Gly Gly
1 5 10 15

Gly Gly Gly Gly Gly Gly Gly Gly Gly Gly Gly Gly Gly Gly
20 25 30

<210> 122
<211> 90
<212> PRT
<213> 人工序列
<220>
<223> 合成多肽
<220>
<221> MISC_FEATURE
<222> (4)..(90)
<223> 可以不存在
<400> 122

Gly Gly Ser Gly Gly Ser Gly Gly Ser Gly Gly Ser Gly Gly Ser Gly
1 5 10 15
Gly Ser Gly Gly Ser Gly Gly Ser Gly Gly Ser Gly Gly Ser Gly Gly
20 25 30
Ser Gly Gly Ser Gly Gly Ser Gly Gly Ser Gly Gly Ser Gly Gly Ser
35 40 45
Gly Gly Ser Gly Gly Ser Gly Gly Ser Gly Gly Ser Gly Gly Ser Gly
50 55 60
Gly Ser Gly Gly Ser Gly Gly Ser Gly Gly Ser Gly Gly Ser Gly Gly
65 70 75 80
Ser Gly Gly Ser Gly Gly Ser Gly Gly Ser
85 90

<210> 123
<211> 60
<212> PRT
<213> 人工序列
<220>
<223> 合成多肽
<220>
<221> misc_feature
<222> (1)..(1)
<223> Xaa可以是任何天然存在的氨基酸
<220>
<221> MISC_FEATURE
<222> (3)..(60)
<223> 可以不存在

<400> 123

Xaa Pro Xaa Pro Xaa Pro Xaa Pro Xaa Pro Xaa Pro Xaa Pro Xaa Pro
1 5 10 15
Xaa Pro Xaa Pro Xaa Pro Xaa Pro Xaa Pro Xaa Pro Xaa Pro Xaa Pro
20 25 30
Xaa Pro Xaa Pro Xaa Pro Xaa Pro Xaa Pro Xaa Pro Xaa Pro Xaa Pro
35 40 45
Xaa Pro Xaa Pro Xaa Pro Xaa Pro Xaa Pro Xaa Pro
50 55 60

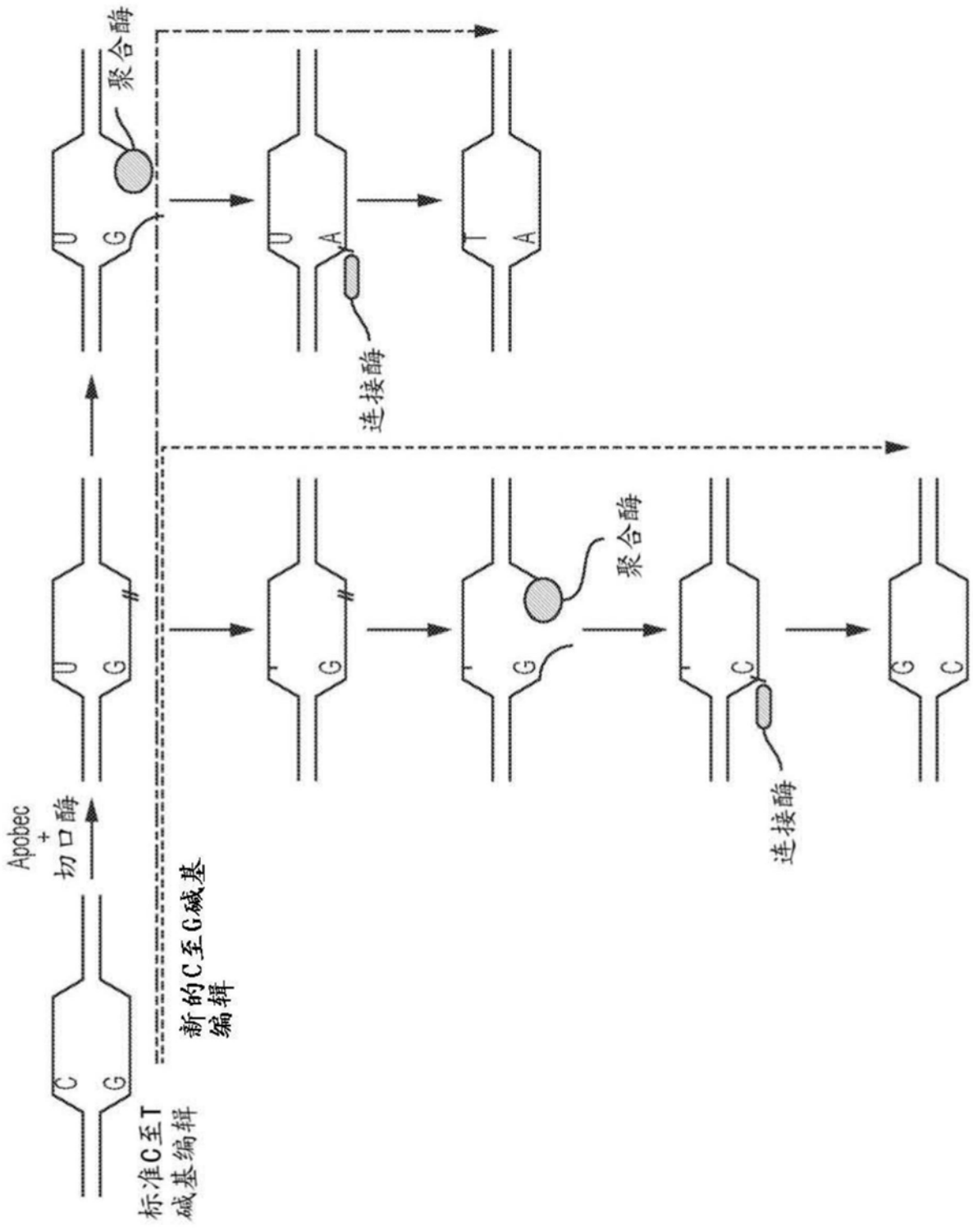



图1

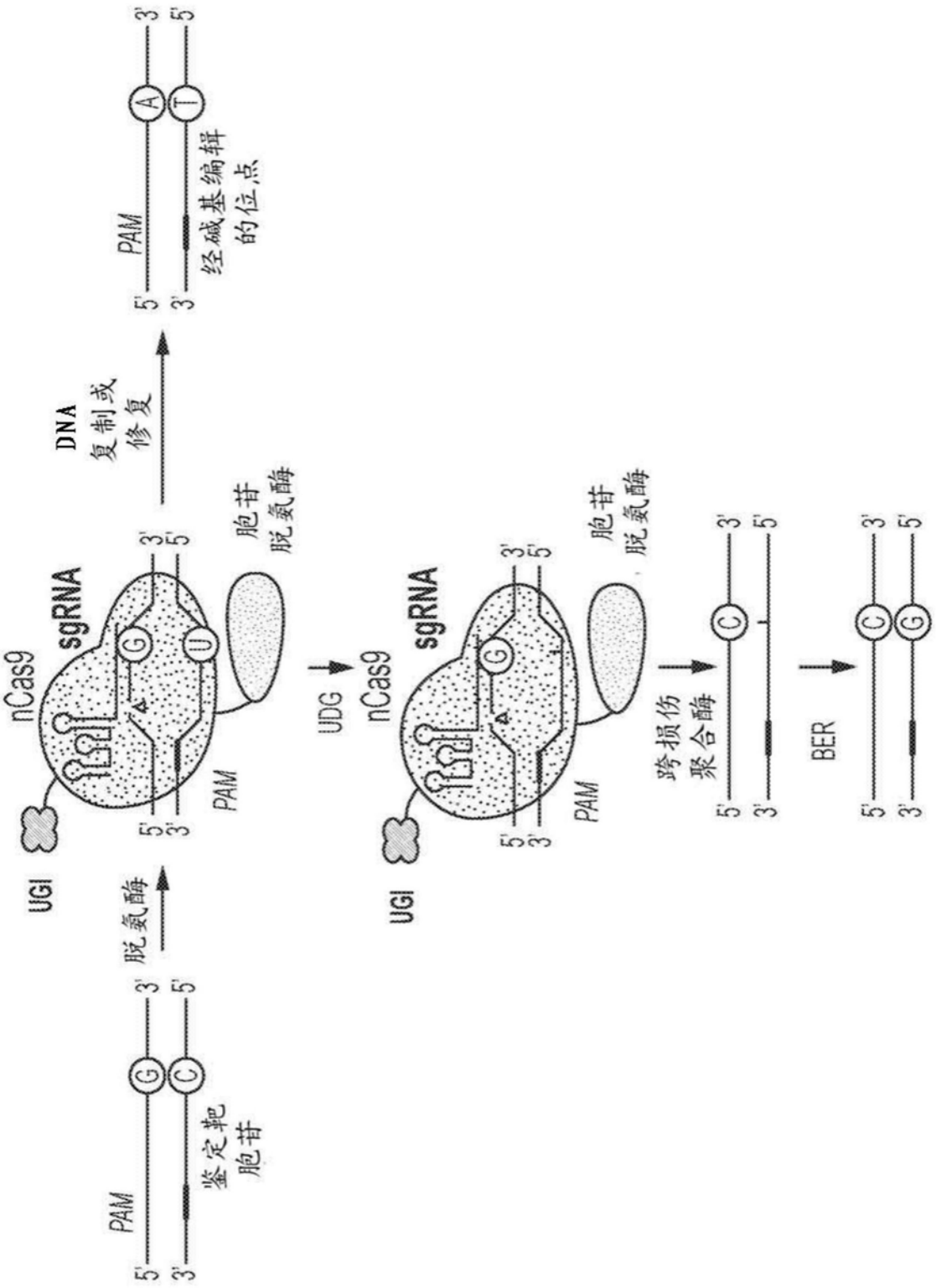

PAM
鉴定靶
胞苷
脱氨酶
UGI
nCas9
sgRNA
胞苷
脱氨酶
DNA
复制或
修复
经碱基编辑
的位点
UDG
跨损伤
聚合酶
BER

图2

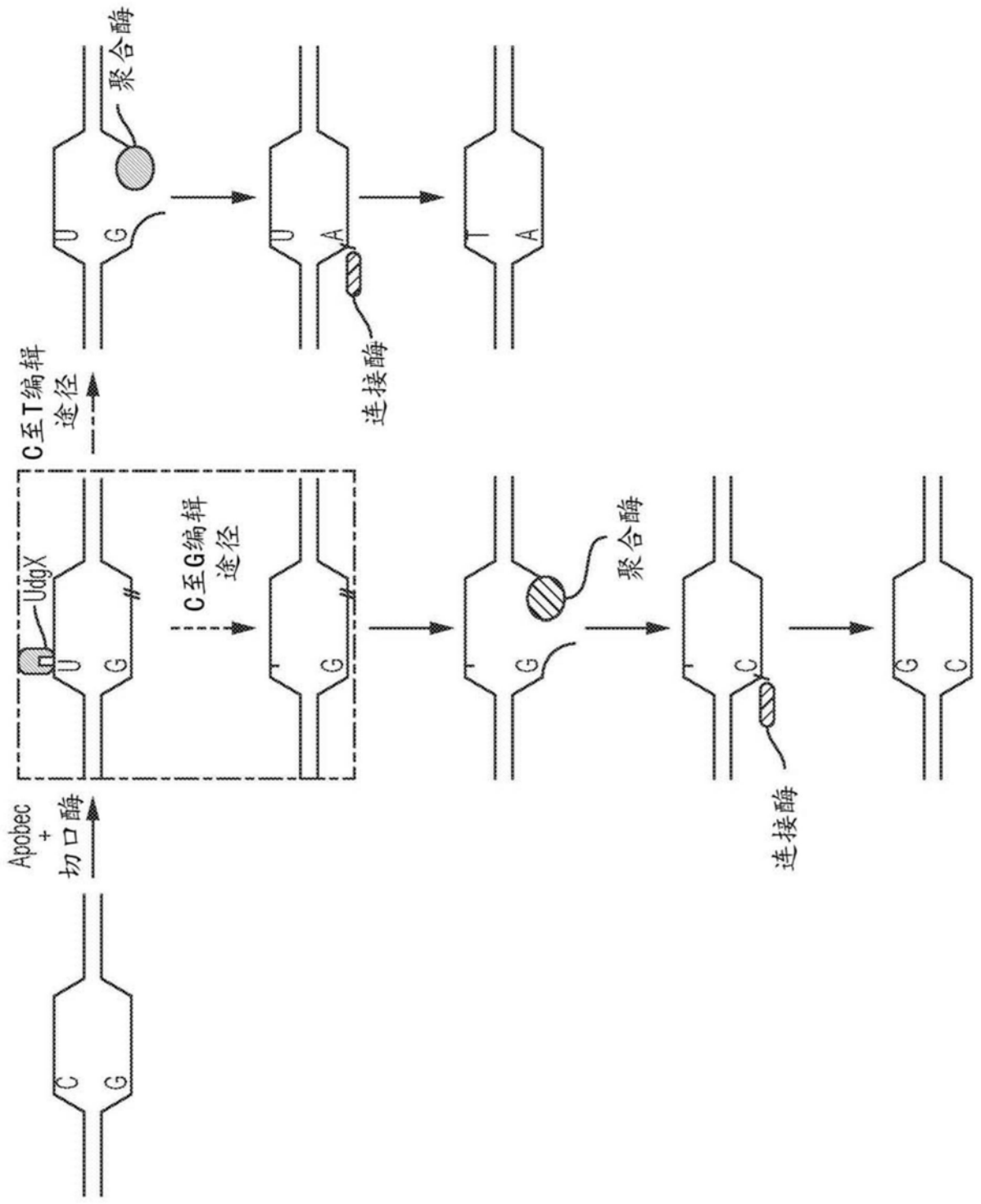



图3

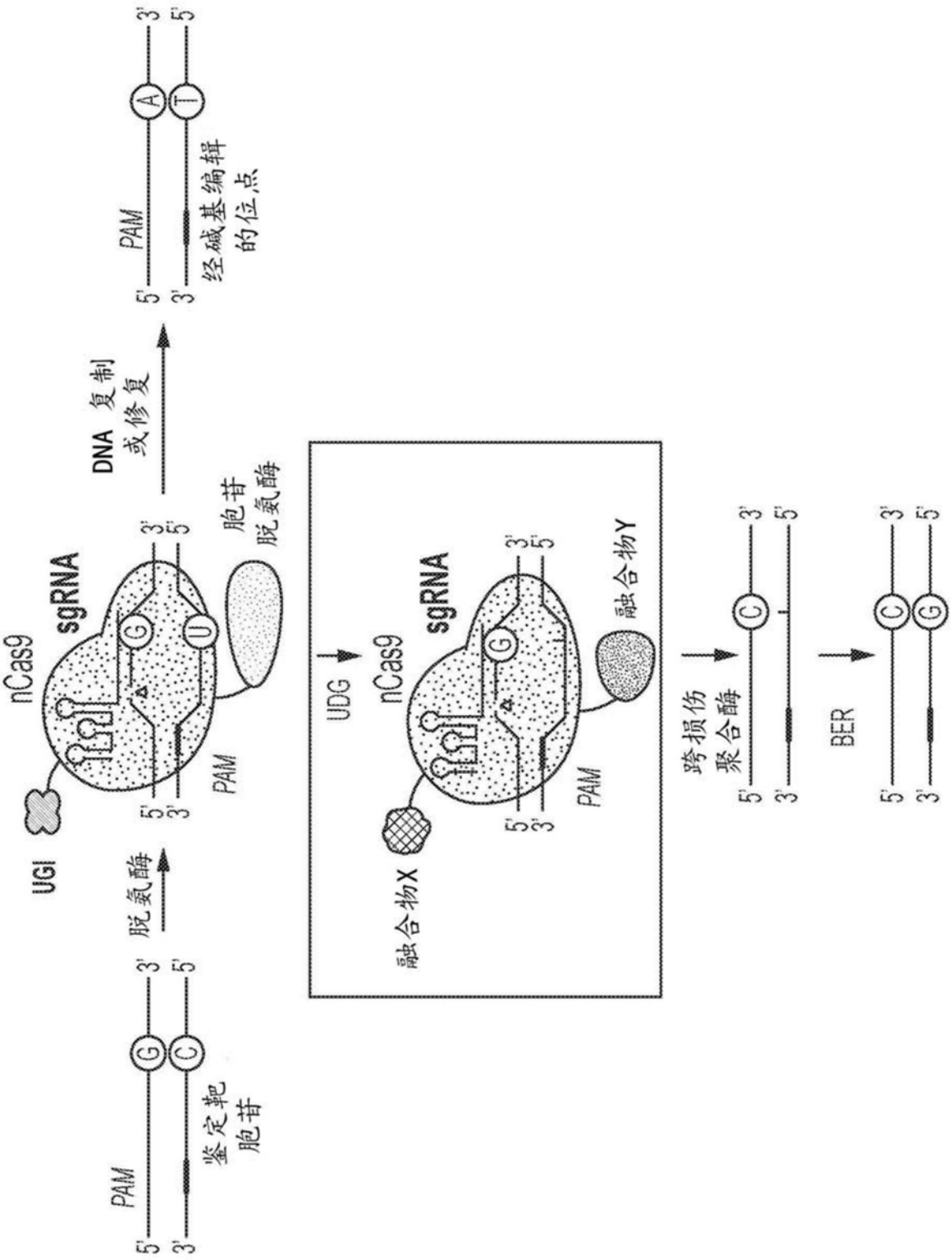

PAM
鉴定靶
胞苷
脱氨酶
UGI
nCas9
sgRNA
胞苷
脱氨酶
DNA 复制
或修复
经碱基编辑
的位点
UDG
融合物X
融合物Y
跨损伤
聚合酶
BER

图4

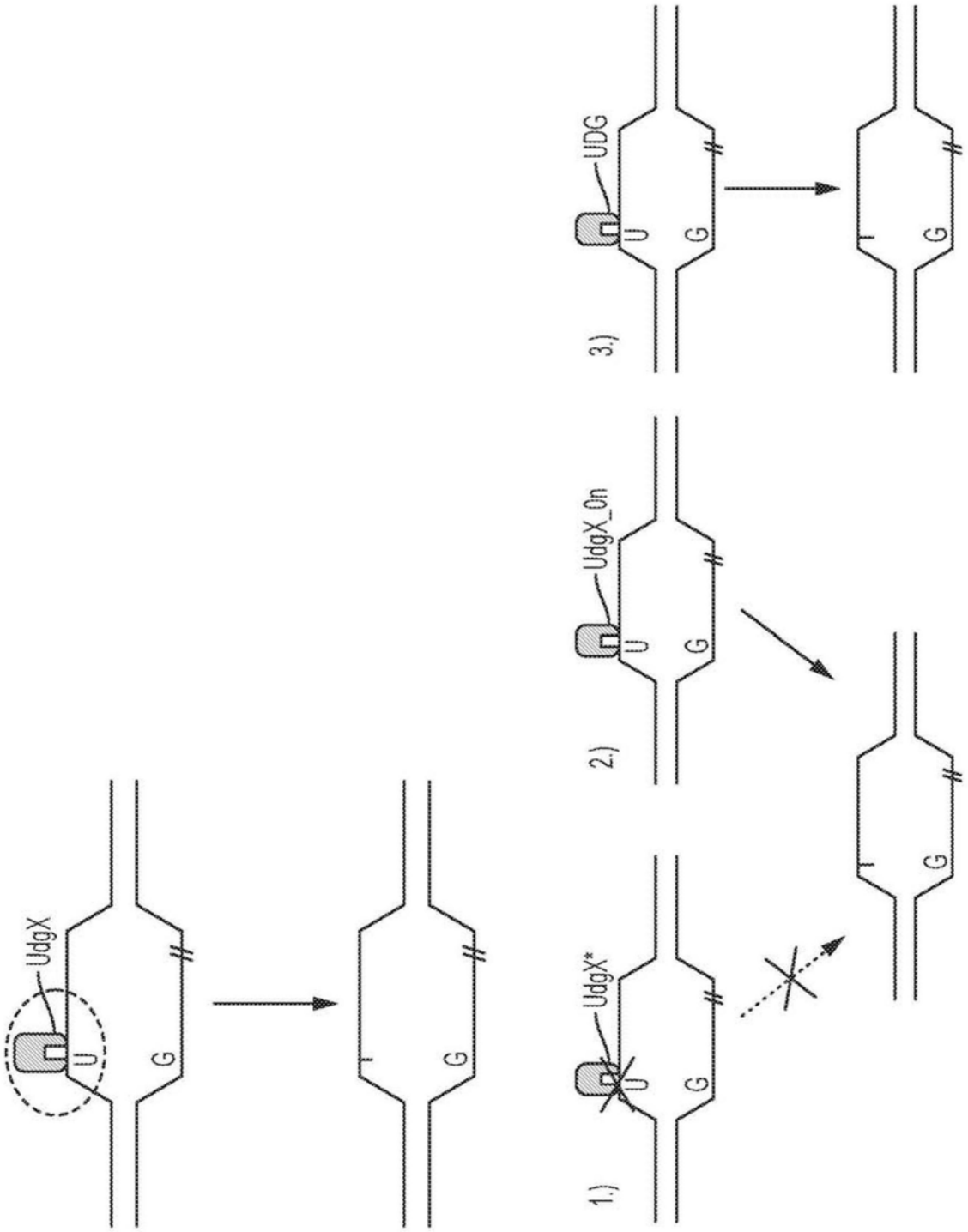



图5

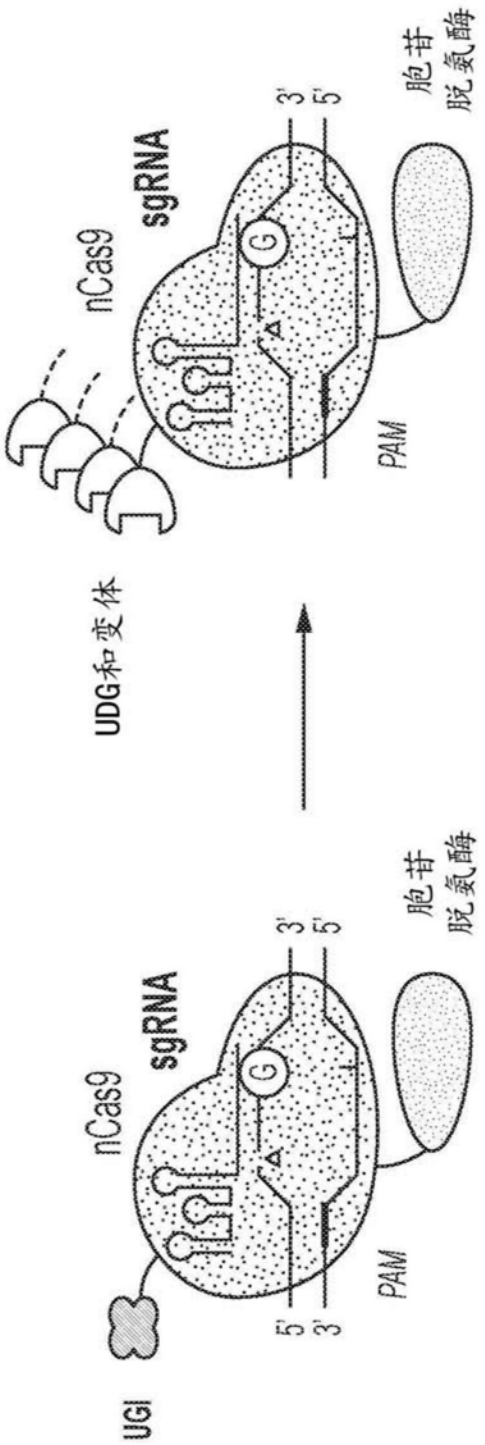
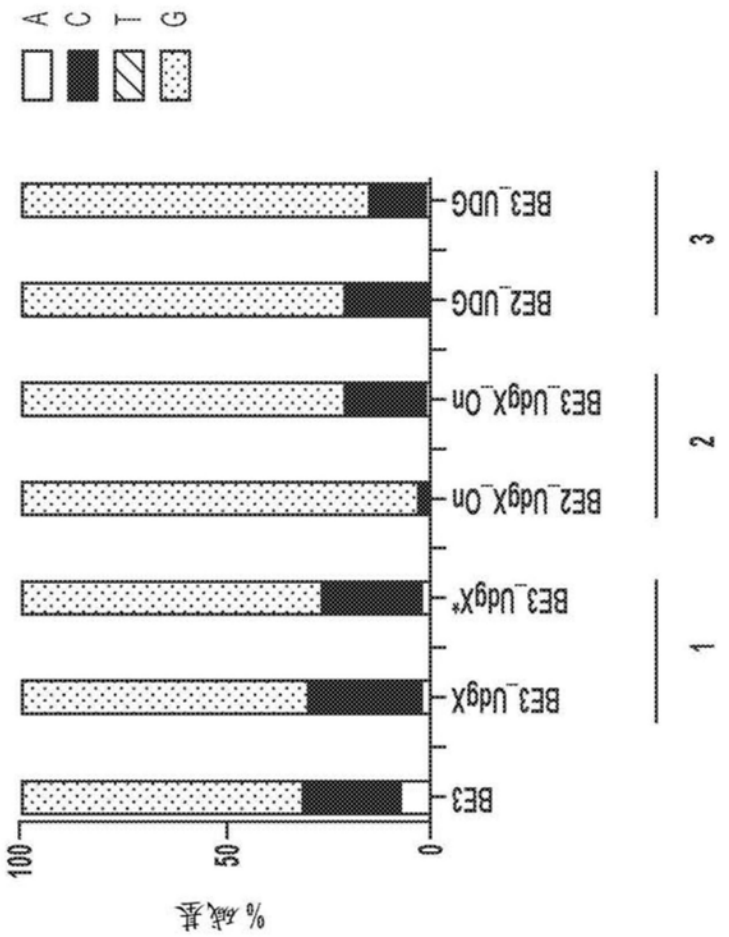



图6




| | A | C | T | G |
|---|---|---|---|---|
| WT Hap1 | | | | |
| BE3 | 7.029168 | 23.98752 | 0.311287 | 68.67203 |
| BE3_UdgX | 2.429073 | 26.86903 | 0.154274 | 70.54762 |
| BE3_UdgX* | 2.293083 | 24.10467 | 0.133893 | 73.46835 |
| BE2_UdgX_On | 0.597649 | 2.235434 | 0.088334 | 97.07858 |
| BE3_UdgX_On | 1.761788 | 19.16713 | 0.155196 | 78.91588 |
| BE2_UDG | 0.479271 | 1.871172 | 0.026613 | 97.62294 |
| BE3_UDG | 1.666401 | 13.09618 | 0.14136 | 85.09606 |

图7

HEK2

WT Hap1

| | A | C | T | G |
|---|---|---|---|---|
| BE3 | 7.029168 | 23.98752 | 0.311287 | 68.67203 |
| BE3 | 5.872296 | 27.17249 | 0.27076 | 66.68446 |
| BE3_UdgX | 2.429073 | 26.86903 | 0.154274 | 70.54762 |
| BE3_UdgX | 2.908524 | 26.48944 | 0.438746 | 70.1633 |
| BE3_UdgX* | 2.293083 | 24.10467 | 0.133893 | 73.46835 |
| BE3_REV7 | 2.010573 | 12.94713 | 0.013356 | 85.02894 |
| BE2_UDG | 0.479271 | 1.871172 | 0.026613 | 97.62294 |
| BE3_UDG | 1.666401 | 13.09618 | 0.14136 | 85.09606 |
| BE2_UdgX_On | 0.597649 | 2.235434 | 0.088334 | 97.07858 |
| BE3_UdgX_On | 1.761788 | 19.16713 | 0.155196 | 78.91588 |
| SMUG1 | 5.542508 | 10.38046 | 0.234852 | 83.84218 |

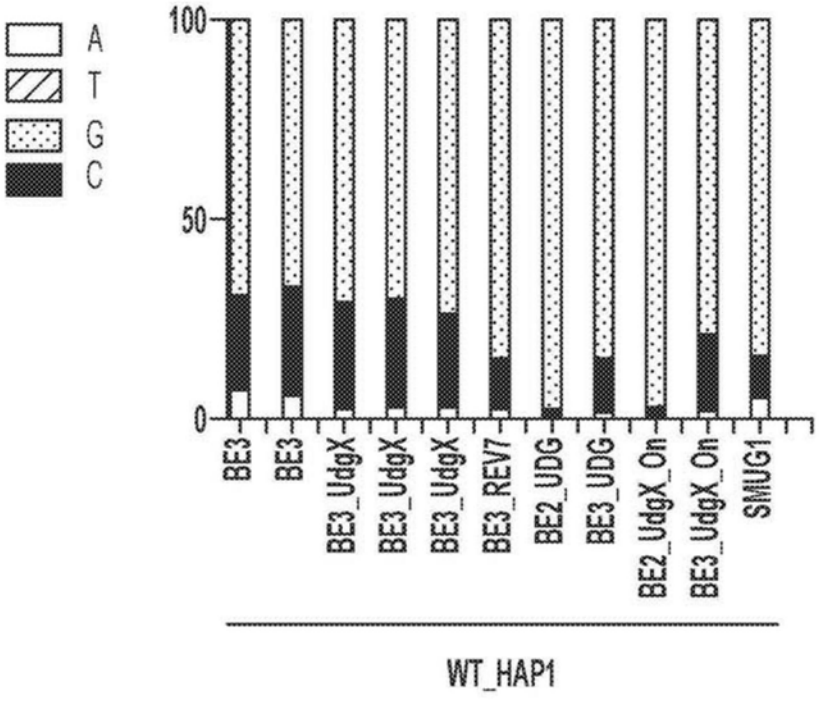
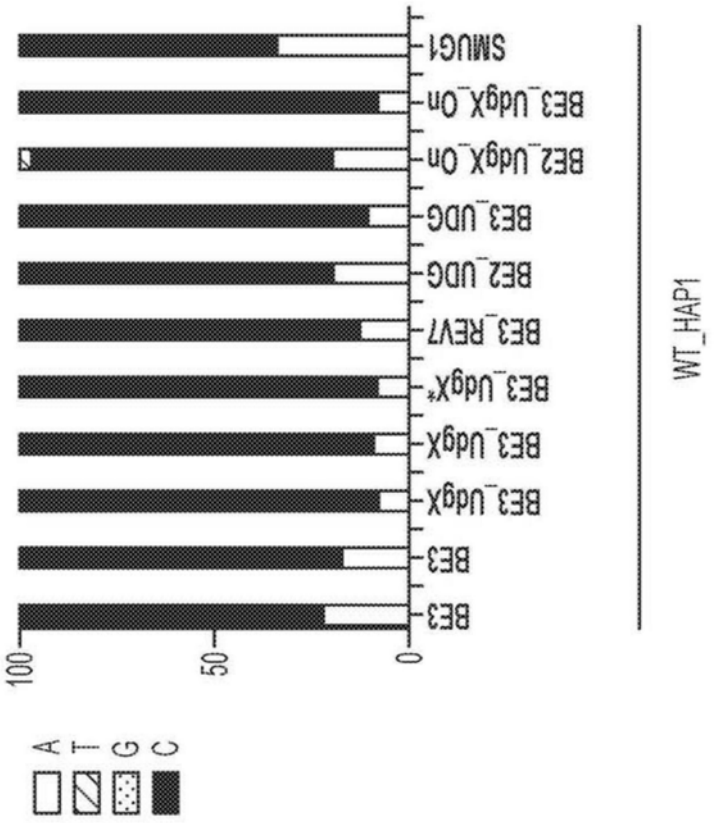



图8




| HEK2 | | | | | |
|---|---|---|---|---|---|
| WT Hap1 | 总编辑物 | A | C | T | |
| BE3 | 31.3279752 | 22.43735 | 76.56901 | 0.99364 | |
| BE3 | 33.315546 | 17.62629 | 81.56099 | 0.812714 | |
| BE3_UdgX | 29.4523771 | 8.24746 | 91.22873 | 0.523809 | |
| BE3_UdgX | 29.8367097 | 9.748139 | 88.78137 | 1.47049 | |
| BE3_UdgX* | 26.5316462 | 8.642822 | 90.85252 | 0.504655 | |
| BE3_REV7 | 14.97105859 | 13.42973 | 86.48106 | 0.089209 | |
| BE2_UDG | 2.37705617 | 20.16239 | 78.71804 | 1.119573 | |
| BE3_UDG | 14.9039407 | 11.18094 | 87.87059 | 0.948472 | |
| BE2_UdgX_On | 2.921416827 | 20.45751 | 76.51882 | 3.023668 | |
| BE3_UdgX_On | 21.08411823 | 8.355998 | 90.90792 | 0.73608 | |
| SMUG1 | 16.15782057 | 34.30233 | 64.24419 | 1.453488 | |

图9

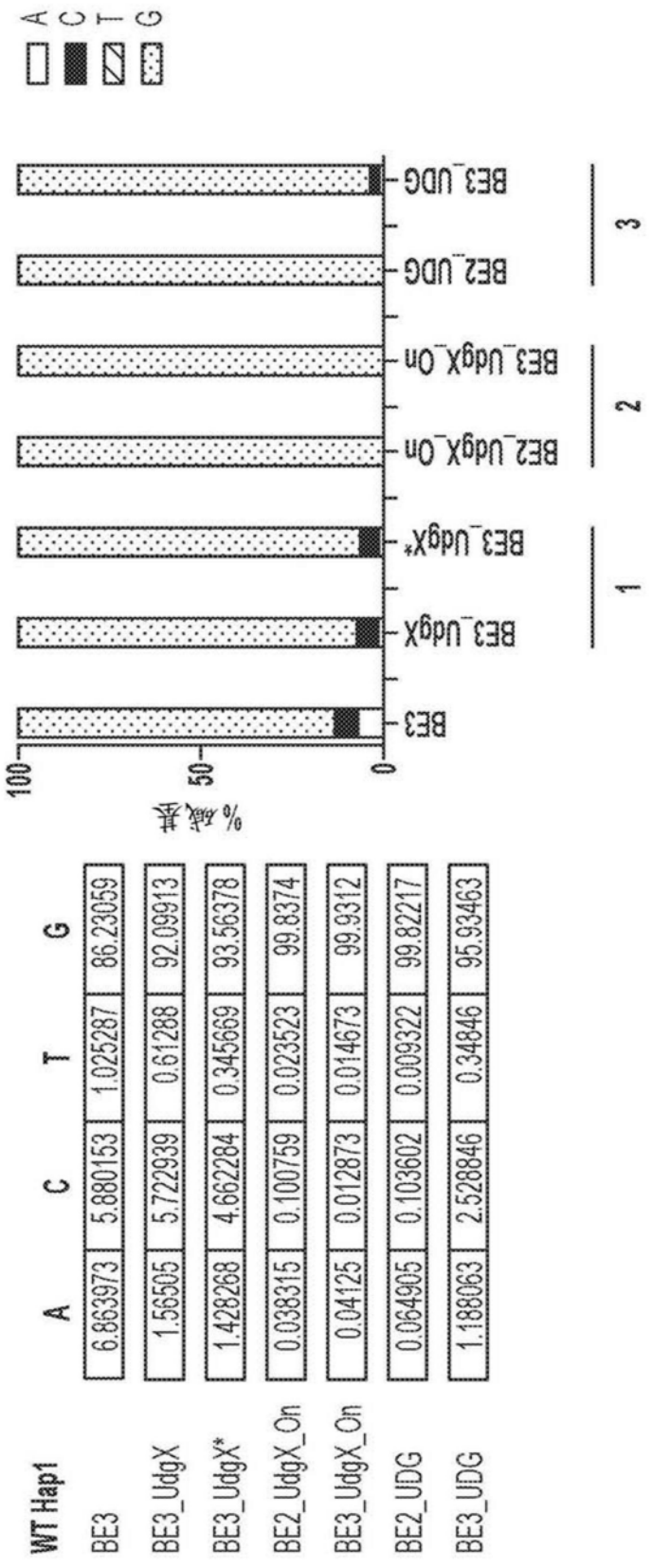

A
C
T
G
BE3
BE3_UdgX
BE3_UdgX*
BE2_UdgX_On
BE3_UdgX_On
BE2_UDG
BE3_UDG
1
2
3
100
50
0
% 碱基
WT Hap1
A C T G
6.863973 5.880153 1.025287 86.23059
1.56505 5.722939 0.61288 92.09913
1.428268 4.662284 0.345669 93.56378
0.038315 0.100759 0.023523 99.8374
0.04125 0.012873 0.014673 99.9312
0.064905 0.103602 0.009322 99.82217
1.188063 2.528846 0.34846 95.93463

图10

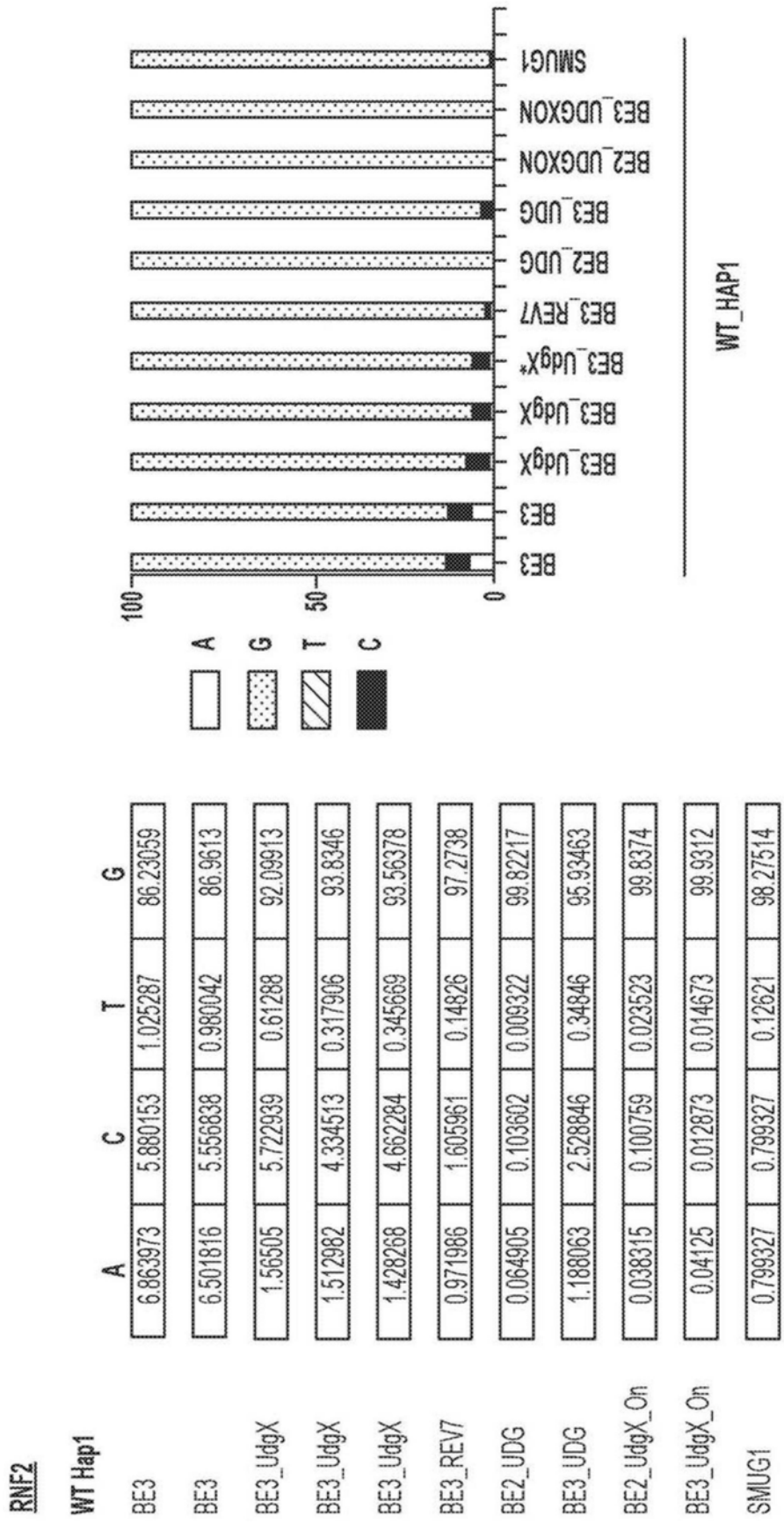

RNF2
WT Hap1
BE3
BE3
BE3_UdgX
BE3_UdgX
BE3_UdgX
BE3_REV7
BE2_UDG
BE3_UDG
BE2_UdgX_On
BE3_UdgX_On
SMUG1
A C T G
6.863973 5.880153 1.025287 86.23059
6.501816 5.556838 0.980042 86.9613
1.56505 5.722939 0.61288 92.09913
1.512982 4.334513 0.317906 93.8346
1.428268 4.662284 0.345669 93.56378
0.971986 1.605961 0.14826 97.2738
0.064905 0.103602 0.009322 99.82217
1.188063 2.528846 0.34846 95.93463
0.038315 0.100759 0.023523 99.8374
0.04125 0.012873 0.014673 99.9312
0.799327 0.799327 0.12621 98.27514
100
50
0
A
G
T
C
BE3
BE3
BE3_UdgX
BE3_UdgX
BE3_UdgX*
BE3_REV7
BE2_UDG
BE3_UDG
BE2_UDGXON
BE3_UDGXON
SMUG1
WT_HAP1

图11

RNF2

WT Hap1

| | 总编辑物 | A | C | T |
|---|---|---|---|---|
| BE3 | 13.769413 | 49.84942 | 42.70446 | 7.44612 |
| BE3 | 13.038696 | 49.86554 | 42.61805 | 7.516411 |
| BE3_UdgX | 7.9008694 | 19.80858 | 72.43429 | 7.757126 |
| BE3_UdgX | 6.1654013 | 24.53988 | 70.30383 | 5.156295 |
| BE3_UdgX | 6.4362206 | 22.1911 | 72.43823 | 5.370677 |
| BE3_REV7 | 2.7262062 | 35.65341 | 58.90827 | 5.438312 |
| BE2_UDG | 0.177828896 | 36.49852 | 58.25915 | 5.242335 |
| BE3_UDG | 4.0653692 | 29.22399 | 62.20458 | 8.571428 |
| BE2_UdgX_On | 0.162597289 | 23.5645 | 61.96868 | 14.46682 |
| BE3_UdgX_On | 0.068796831 | 59.95976 | 18.71227 | 21.32797 |
| SMUG1 | 1.724863273 | 46.34146 | 46.34146 | 7.317073 |

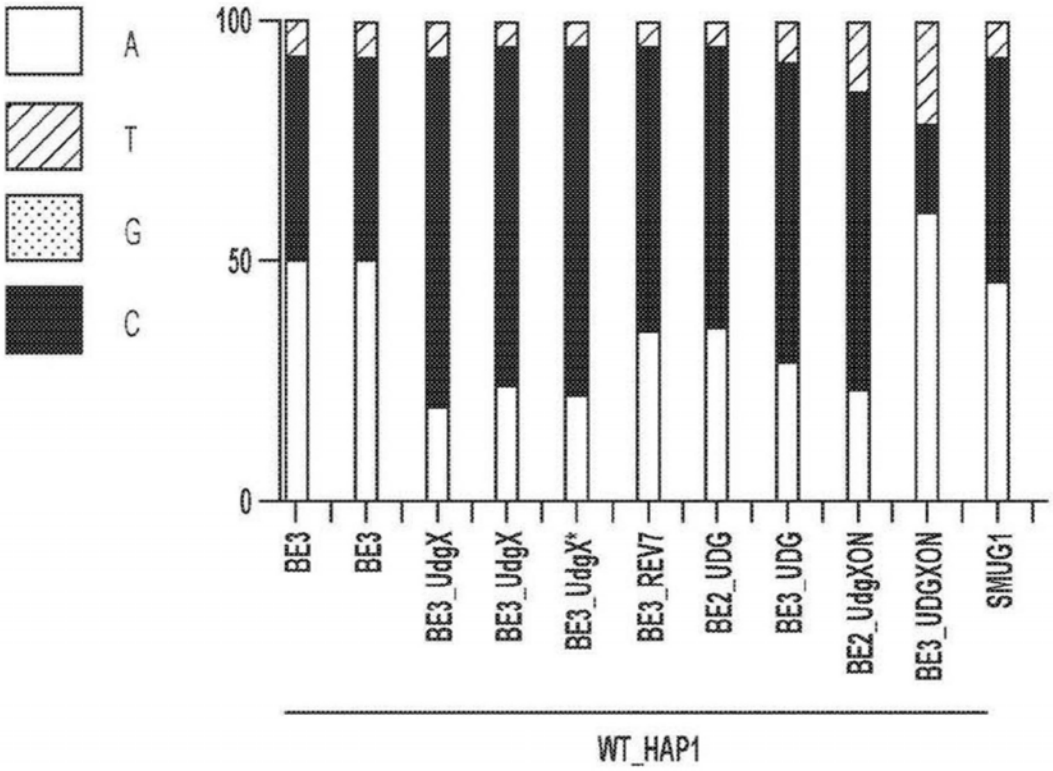
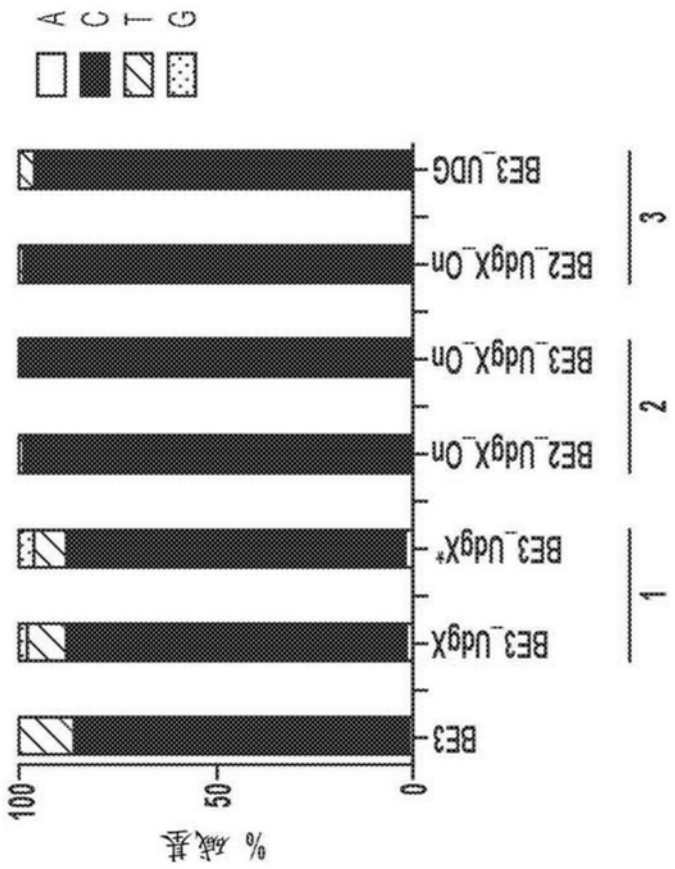



图12




| WT Hap1 | A | C | T | G |
|---|---|---|---|---|
| BE3 | 0.957861 | 85.13409 | 13.29985 | 0.608199 |
| BE3_UdgX | 1.778677 | 86.27834 | 9.689997 | 2.252991 |
| BE3_Udgx* | 1.61547 | 86.04843 | 9.358341 | 2.97776 |
| BE2_UdgX_On | 0.183828 | 99.0089 | 0.540035 | 0.267237 |
| BE3_UdgX_On | 0.10635 | 99.66768 | 0.215432 | 0.010537 |
| BE2_UDG | 0.081416 | 98.62175 | 1.196048 | 0.100783 |
| BE3_UDG | 0.446354 | 95.72025 | 3.352711 | 0.480688 |

图13

FANCF

WT Hap1

| | A | C | T | G |
|---|---|---|---|---|
| BE3 | 0.957861 | 85.13409 | 13.29985 | 0.608199 |
| BE3 | 1.318066 | 78.56728 | 19.15355 | 0.961108 |
| BE3_UdgX | 1.778677 | 86.27834 | 9.689997 | 2.252991 |
| BE3_UdgX | 1.764677 | 87.32044 | 8.332098 | 2.582781 |
| BE3_UdgX | 1.61547 | 86.04843 | 9.689997 | 2.97776 |
| BE3_REV7 | 0.514632 | 93.70418 | 5.31733 | 0.463864 |
| BE2_UDG | 0.081416 | 98.62175 | 1.196048 | 0.100783 |
| BE3_UDG | 0.446354 | 95.72025 | 3.352711 | 0.480688 |
| BE2_UdgXON | 0.183828 | 99.0089 | 0.540035 | 0.267237 |
| BE3_UDGXON | 0.10635 | 99.66768 | 0.215432 | 0.010537 |
| SMUG1 | 0.330974 | 97.17443 | 2.400242 | 0.094351 |

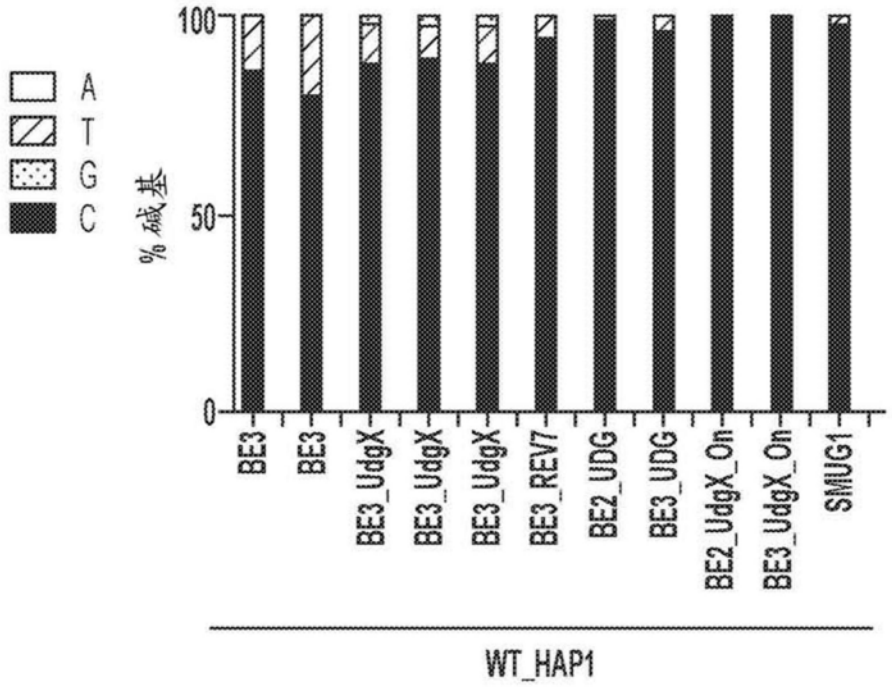



图14

FANCF

WT Hap1

| | 总编辑物 | A | T | G |
|---|---|---|---|---|
| BE3 | 14.8659096 | 6.443336 | 89.46543 | 4.091234 |
| BE3 | 21.4327257 | 6.149792 | 89.36591 | 4.4843 |
| BE3_UdgX | 13.721665 | 12.96254 | 70.61823 | 16.41922 |
| BE3_UdgX | 12.679556 | 13.9175 | 65.71285 | 20.36965 |
| BE3_UdgX | 13.951571 | 11.57913 | 67.07733 | 21.34355 |
| BE3_REV7 | 6.2958255 | 8.174178 | 84.45803 | 7.367795 |
| BE2_UDG | 1.37824733 | 5.907229 | 86.78036 | 7.31241 |
| BE2_UDG | 4.279753 | 10.42942 | 78.33889 | 11.23169 |
| BE2_UdgXON | 0.991099728 | 18.54785 | 54.48845 | 26.9637 |
| BE3_UDGXON | 0.332319917 | 32.00235 | 64.82678 | 3.170875 |
| SMUG1 | 2.825567162 | 11.71353 | 84.94728 | 3.339192 |

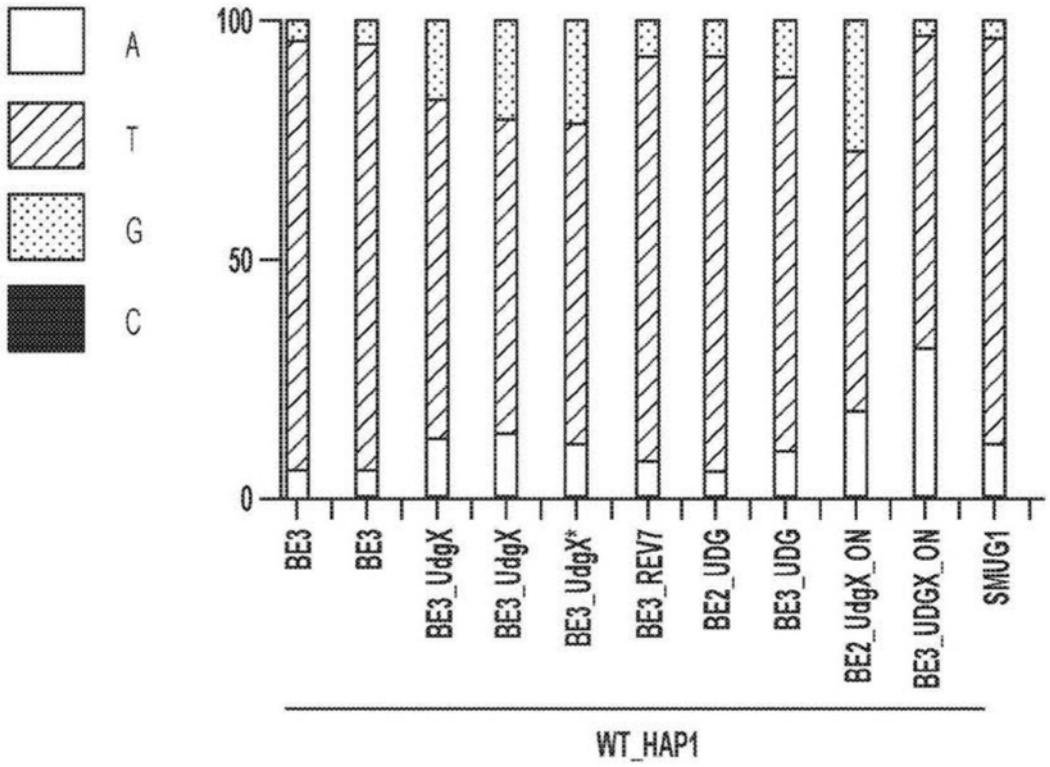
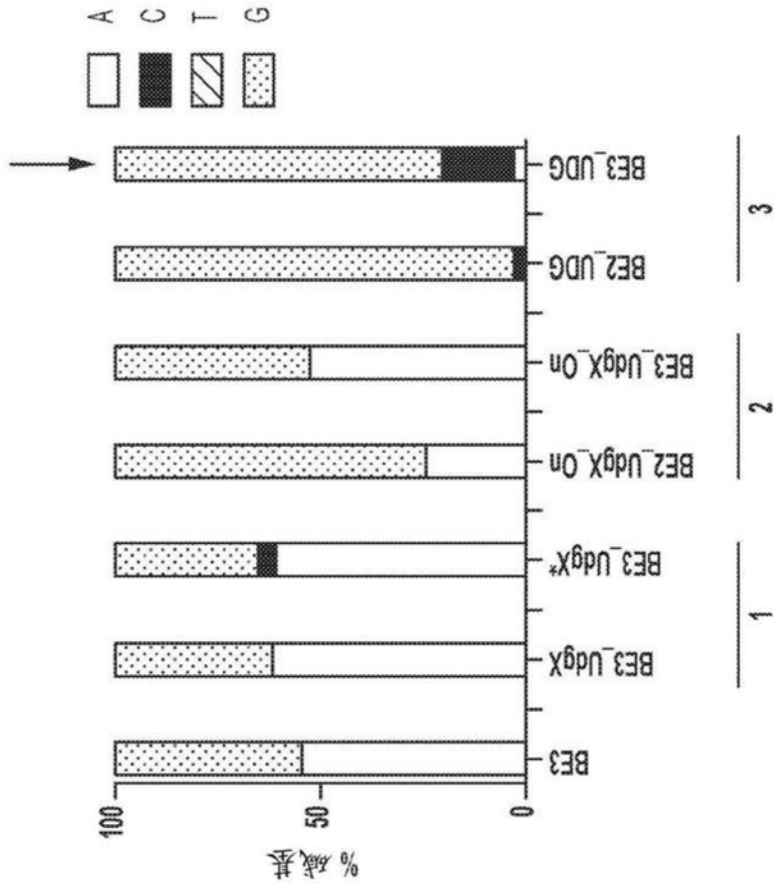



图15




| UDG-/-HAP1 | A | C | T | G |
|---|---|---|---|---|
| BE3 | 54.48563 | 1.099159 | 0.008312 | 44.4069 |
| BE3_UdgX | 61.34017 | 0.562122 | 0.045397 | 38.05231 |
| BE3_UdgX* | 60.98913 | 4.55437 | 0.003002 | 34.4535 |
| BE3_UdgX_On | 24.22648 | 0.224724 | 0.048873 | 75.49992 |
| BE3_UdgX_On | 52.57995 | 0.73715 | 0.056083 | 46.62682 |
| BE2_UDG | 1.126263 | 1.972256 | 0.025385 | 96.8761 |
| BE3_UDG | 3.034692 | 17.10462 | 0.179678 | 79.68102 |

图16

| UDG-/- HAP1 | A | C | T | G |
|---|---|---|---|---|
| BE3 | 54.48563 | 1.099159 | 0.008312 | 44.4069 |
| BE3 | 54.88124 | 0.280479 | 0.037441 | 44.80084 |
| BE3_UdgX | 61.34017 | 0.562122 | 0.045397 | 38.05231 |
| BE3_UDGX | 71.34399 | 1.200882 | 0.015775 | 27.43935 |
| BE3_UDGX | 60.98913 | 4.55437 | 0.003002 | 34.4535 |
| BE3_REV7 | 52.13184 | 1.191155 | 0.028973 | 46.64804 |
| BE2_UDG | 1.126263 | 1.972256 | 0.025385 | 96.8761 |
| BE3_UDG | 3.034692 | 17.10462 | 0.179678 | 79.68102 |
| BE2_UdgX_On | 24.22648 | 0.224724 | 0.048873 | 75.49992 |
| BE3_UdgX_On | 52.57995 | 0.73715 | 0.056083 | 46.62682 |
| SMUG1 | 16.2789 | 0.700013 | 0.052671 | 82.96842 |

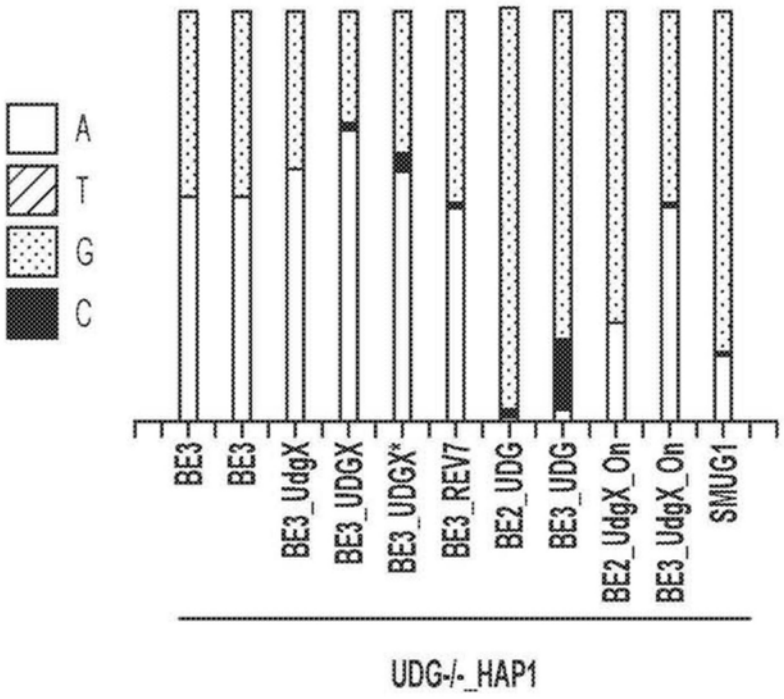



图17

| UDG-/-HAP1 | 总编辑物 | A | C | T |
|---|---|---|---|---|
| BE3 | 55.59310115 | 98.0079 | 1.97715 | 0.014952 |
| BE3 | 55.19916029 | 99.42405 | 0.508121 | 0.06783 |
| BE3_UdgX | 61.94768881 | 99.0193 | 0.907414 | 0.073283 |
| BE3_UDGX | 72.56064713 | 98.32325 | 1.655005 | 0.021741 |
| BE3_UDGX | 65.54650222 | 93.04712 | 6.948304 | 0.00458 |
| BE3_REV7 | 53.35196778 | 97.71306 | 2.232636 | 0.054305 |
| BE2_UDG | 3.12390396 | 36.05306 | 63.13434 | 0.812604 |
| BE3_UDG | 20.3189896 | 14.93525 | 84.18047 | 0.884284 |
| BE2_UdgX_On | 24.50008069 | 98.88328 | 0.917237 | 0.19948 |
| BE3_UdgX_On | 53.37318037 | 98.5138 | 1.381124 | 0.105076 |
| SMUG1 | 17.03158156 | 95.58066 | 4.110088 | 0.309256 |

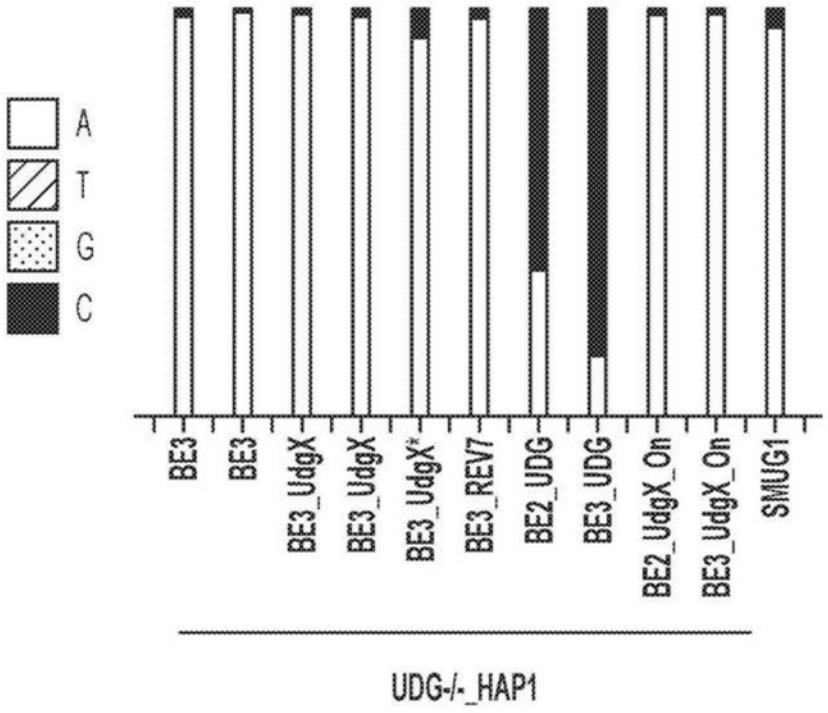
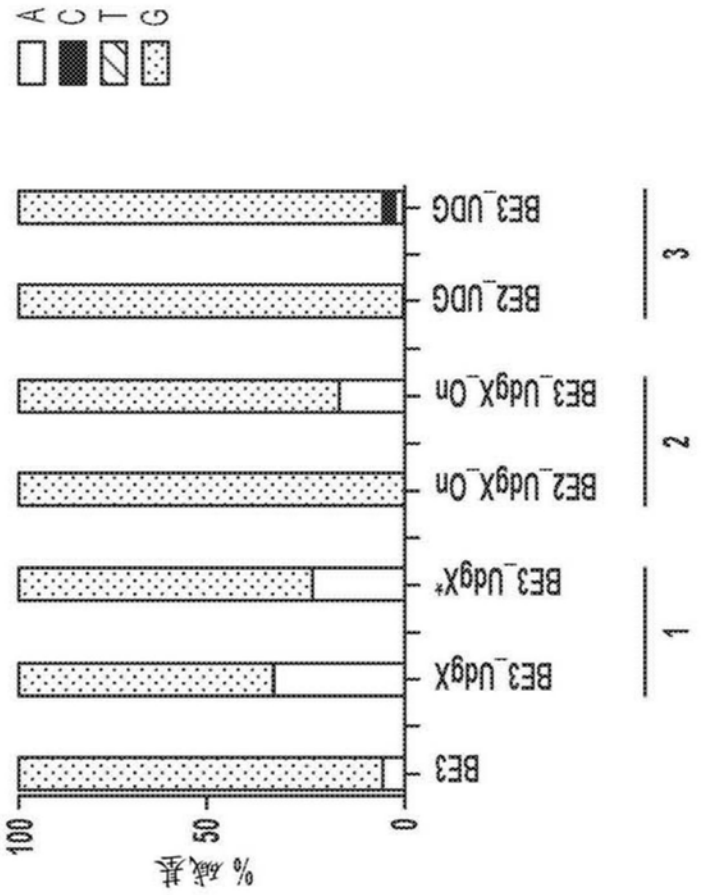



图18




| UDG-/- HAP1 | A | C | T | G |
|---|---|---|---|---|
| BE3 | 6.074682 | 0.14494 | 0.016881 | 93.7635 |
| BE3_UdgX | 33.90168 | 0.033551 | 0.054252 | 66.01052 |
| BE3_UdgX* | 23.21214 | 0.55525 | 0.050559 | 76.18205 |
| BE2_UdgX_On | 0.510098 | 0.010289 | 0.013813 | 99.4658 |
| BE3_UdgX_On | 16.38801 | 0.073378 | 0.015595 | 83.52302 |
| BE2_UDG | 0.181713 | 0.175531 | 0.013027 | 99.62973 |
| BE3_UDG | 1.925522 | 3.84065 | 0.390487 | 93.84334 |

图19

| UDG-/- HAP1 | A | C | T | G |
|---|---|---|---|---|
| BE3 | 6.074682 | 0.14494 | 0.016881 | 93.7635 |
| BE3 | 17.20021 | 0.014772 | 0.073862 | 82.71116 |
| BE3_UdgX | 33.90168 | 0.033551 | 0.054252 | 66.01052 |
| BE3_UDGX | 29.92735 | 0.262275 | 0.056048 | 69.75433 |
| BE3_UDGX | 23.21214 | 0.55525 | 0.050559 | 76.18205 |
| BE3_REV7 | 16.81834 | 0.510068 | 0.045985 | 82.62561 |
| BE2_UDG | 0.181713 | 0.175531 | 0.013027 | 99.62973 |
| BE3_UDG | 1.925522 | 3.84065 | 0.390487 | 93.84334 |
| BE2_UdgX_On | 0.510098 | 0.010289 | 0.013813 | 99.4658 |
| BE3_UdgX_On | 16.38801 | 0.073378 | 0.015595 | 83.52302 |
| SMUG1 | 2.500624 | 0.048749 | 0.021911 | 97.42872 |

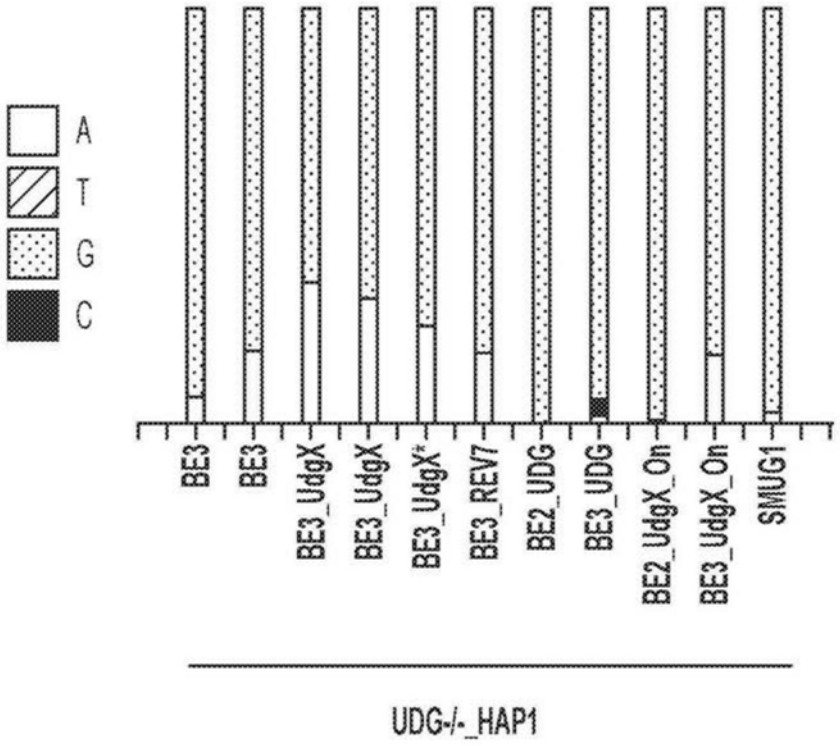
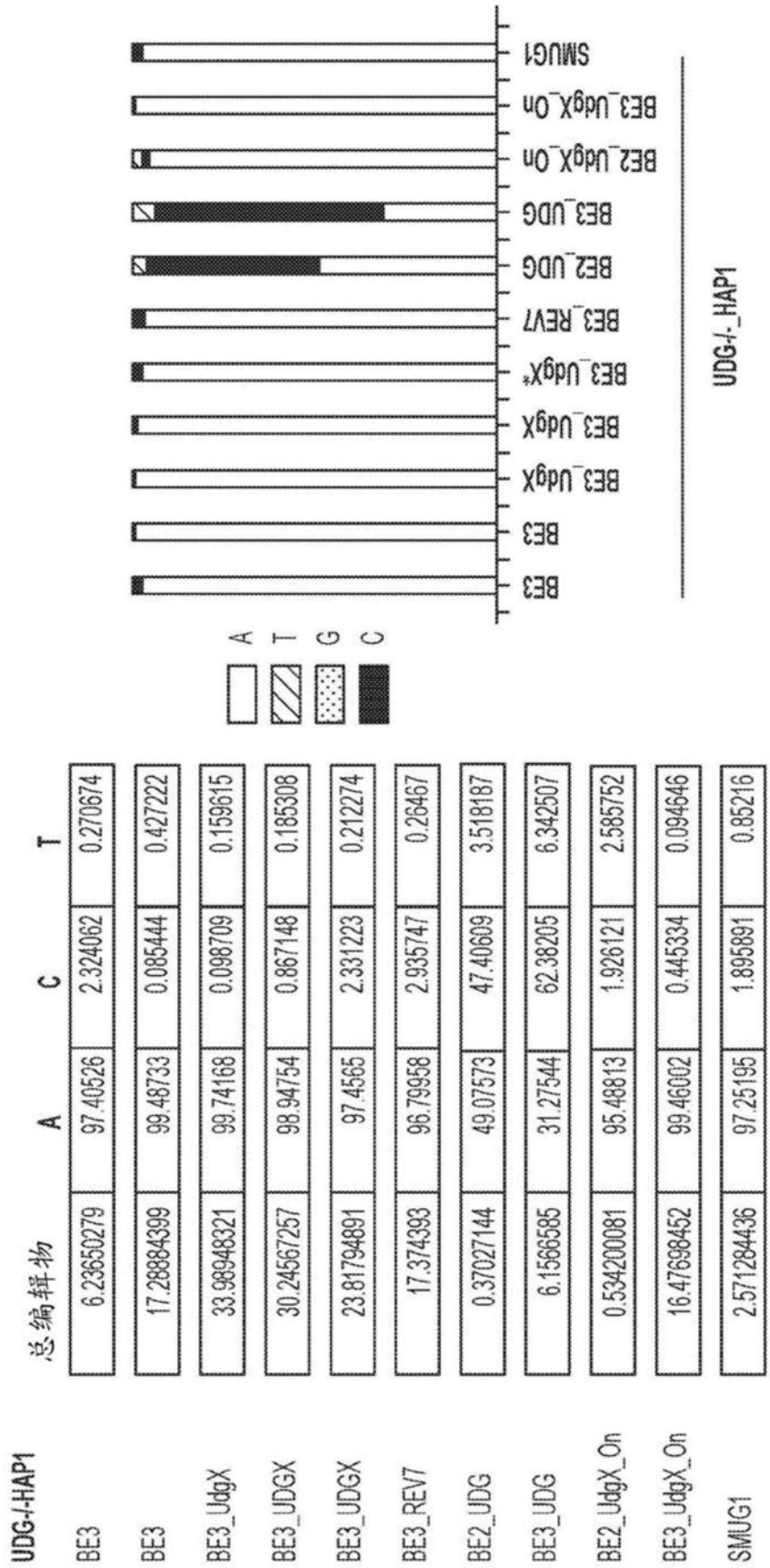
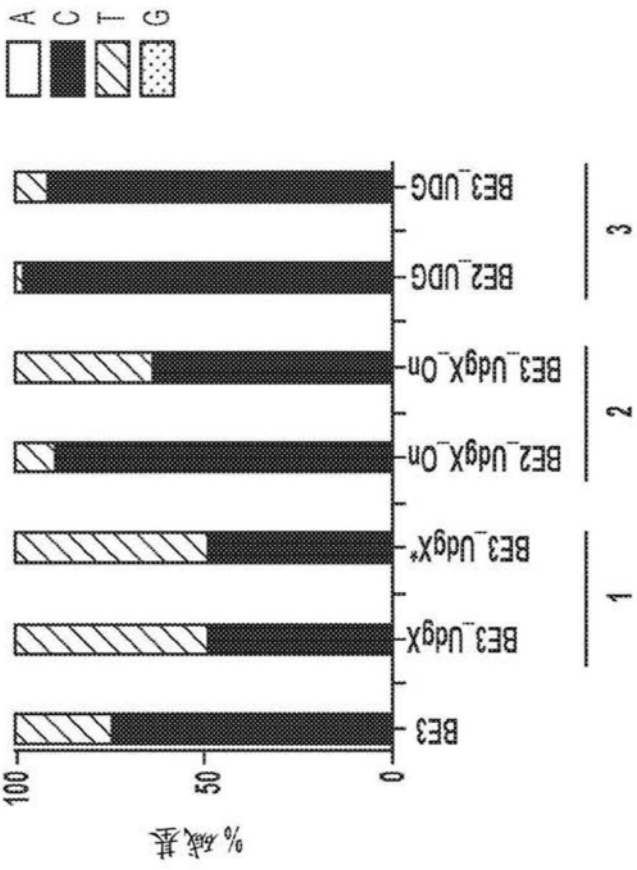



图20




| UDG-/-HAP1 | 总编辑物 | A | C | T |
|---|---|---|---|---|
| BE3 | 6.23650279 | 97.40526 | 2.324062 | 0.270674 |
| BE3 | 17.28884399 | 99.48733 | 0.085444 | 0.427222 |
| BE3_UdgX | 33.98948321 | 99.74168 | 0.098709 | 0.159615 |
| BE3_UDGX | 30.24567257 | 98.94754 | 0.867148 | 0.185308 |
| BE3_UDGX | 23.81794891 | 97.4565 | 2.331223 | 0.212274 |
| BE3_REV7 | 17.374393 | 96.79958 | 2.935747 | 0.26467 |
| BE2_UDG | 0.37027144 | 49.07573 | 47.40609 | 3.518187 |
| BE3_UDG | 6.1566585 | 31.27544 | 62.38205 | 6.342507 |
| BE2_UdgX_On | 0.534200081 | 95.48813 | 1.926121 | 2.585752 |
| BE3_UdgX_On | 16.47698452 | 99.46002 | 0.445334 | 0.094646 |
| SMUG1 | 2.571284436 | 97.25195 | 1.895891 | 0.85216 |

图21




| UDG-/- HAP1 | A | C | T | G |
|---|---|---|---|---|
| BE3 | 0.081524 | 74.54174 | 25.2736 | 0.10314 |
| BE3_UdgX | 0.007288 | 49.28903 | 50.48537 | 0.218305 |
| BE3_UdgX* | 0.280479 | 48.59955 | 50.86376 | 0.256216 |
| BE2_UdgX_On | 0.068791 | 89.49933 | 10.39908 | 0.032799 |
| BE3_UdgX_On | 0.083133 | 63.67724 | 36.17167 | 0.0728 |
| BE2_UDG | 0.106566 | 98.30479 | 1.362403 | 0.226247 |
| BE3_UDG | 0.717993 | 91.0367 | 7.638342 | 0.606967 |

图22

| UDG-/- HAP1 | A | C | T | G |
|---|---|---|---|---|
| BE3 | 0.081524 | 74.54174 | 25.2736 | 0.10314 |
| BE3 | 0.010804 | 45.91037 | 53.89299 | 0.185833 |
| BE3_UdgX | 0.007288 | 49.28903 | 50.48537 | 0.218305 |
| BE3_UDGX | 0.09751 | 48.31471 | 51.38799 | 0.199794 |
| BE3_UDGX | 0.280479 | 48.59955 | 50.86376 | 0.256216 |
| BE3_REV7 | 0.937192 | 66.39211 | 30.96824 | 1.702469 |
| BE2_UDG | 0.106566 | 98.30479 | 1.362403 | 0.226247 |
| BE3_UDG | 0.717993 | 91.0367 | 7.638342 | 0.606967 |
| BE2_UdgXON | 0.068791 | 89.49933 | 10.39908 | 0.032799 |
| BE3_UDGXON | 0.083133 | 63.6724 | 36.17167 | 0.0728 |
| SMUG1 | 0.08642 | 85.30366 | 14.44968 | 0.160237 |

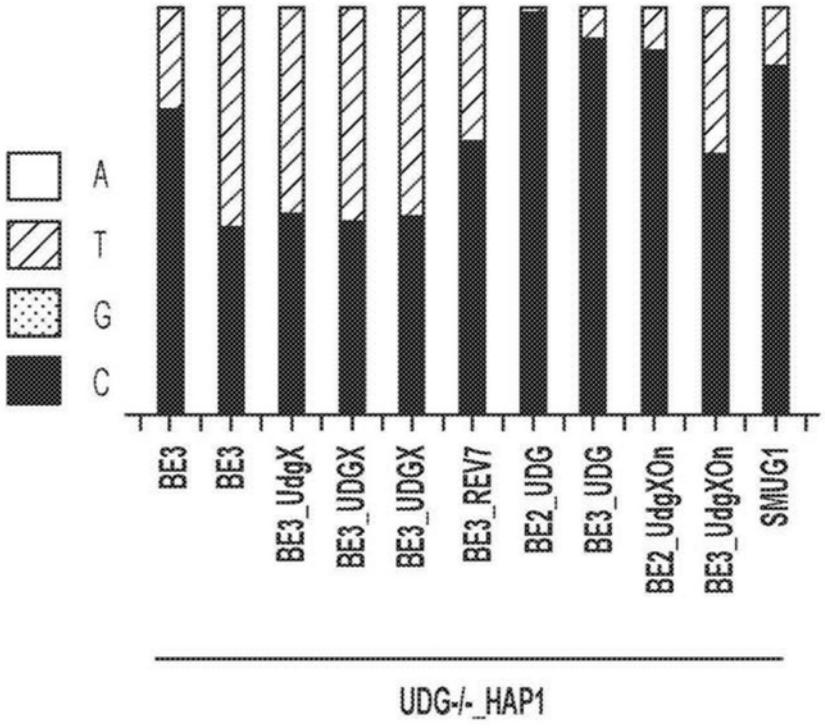



图23

| UDG-/- HAP1 | 总编辑物 | A | T | G |
|---|---|---|---|---|
| BE3 | 25.45826365 | 0.320225 | 99.27464 | 0.405133 |
| BE3 | 54.08962767 | 0.019975 | 99.63646 | 0.343566 |
| BE3_UdgX | 50.7109634 | 0.014372 | 99.55514 | 0.430489 |
| BE3_UDGX | 51.68529451 | 0.188662 | 99.42478 | 0.386559 |
| BE3_UDGX | 51.4004551 | 0.545674 | 98.95586 | 0.49847 |
| BE3_REV7 | 33.6079013 | 2.788607 | 92.14571 | 5.065681 |
| BE2_UDG | 1.6952162 | 6.286266 | 80.36751 | 13.34623 |
| BE3_UDG | 8.9633013 | 8.010358 | 85.21795 | 6.771688 |
| BE2_UdgXON | 10.50066614 | 0.655112 | 99.03253 | 0.312353 |
| BE3_UDGXON | 36.32759937 | 0.228842 | 99.57076 | 0.200398 |
| SMUG1 | 14.69633751 | 0.588037 | 98.32164 | 1.090319 |

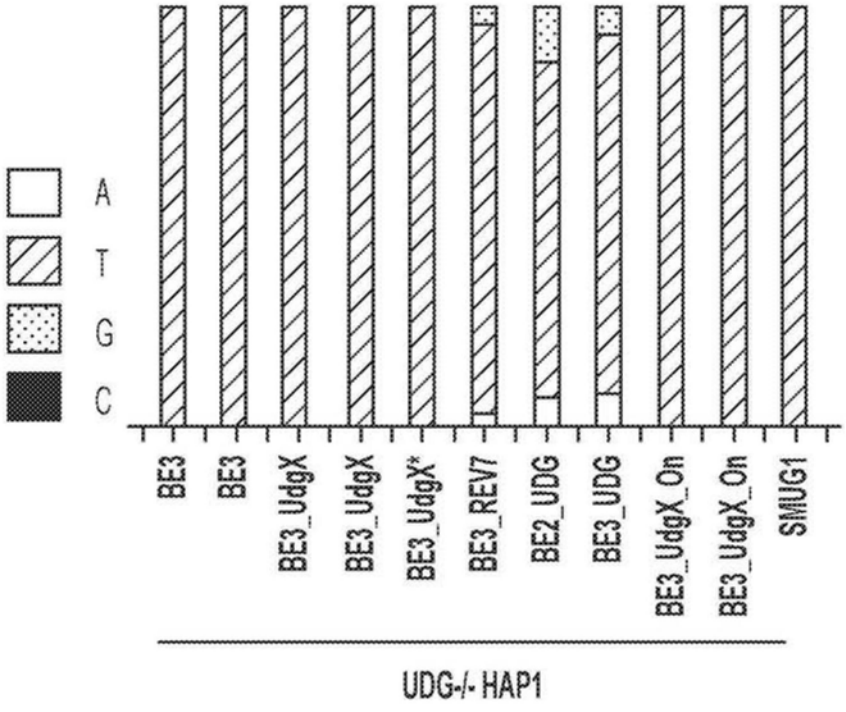



图24

REV1-/-_HAP1

| | A | C | T | G |
|---|---|---|---|---|
| BE3 | 27.90154 | 38.82184 | 0.21661 | 33.06002 |
| BE3_UdgX | 5.394911 | 62.86946 | 0.220717 | 31.51492 |
| BE2_UDG | 0.583164 | 3.171128 | 0.018236 | 96.22747 |
| BE3_UDG | 3.612539 | 33.5042 | 0.285178 | 62.59809 |
| BE2_UdgX_On | 0.787453 | 2.146845 | 0.072979 | 96.99272 |
| BE3_UdgX_On | 2.845842 | 27.20856 | 0.126609 | 69.81899 |
| SMUG1 | 2.136539 | 12.36191 | 0.104799 | 85.39675 |

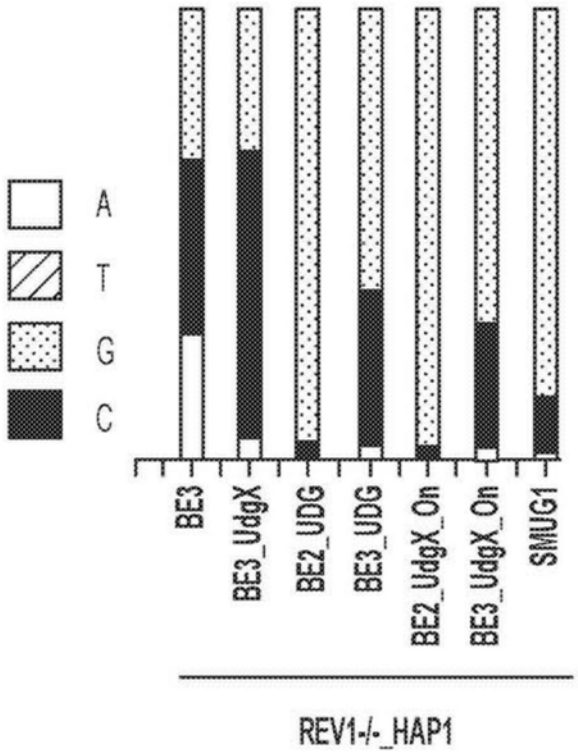



图25

| REV1-/-HAP1 | 总编辑物 | A | C | T |
|---|---|---|---|---|
| BE3 | 66.9399898 | 41.68142 | 57.99499 | 0.323588 |
| BE3_UdgX | 68.4850877 | 7.877497 | 91.80022 | 0.322284 |
| BE2_UDG | 3.77252838 | 15.45818 | 84.05843 | 0.483394 |
| BE3_UDG | 37.4019174 | 9.6587 | 89.57883 | 0.76247 |
| BE2_UdgX_On | 3.007276279 | 26.18491 | 71.38834 | 2.426748 |
| BE3_UdgX_On | 30.18101476 | 9.429247 | 90.151126 | 0.419498 |
| SMUG1 | 14.60324913 | 14.63057 | 84.65179 | 0.717639 |

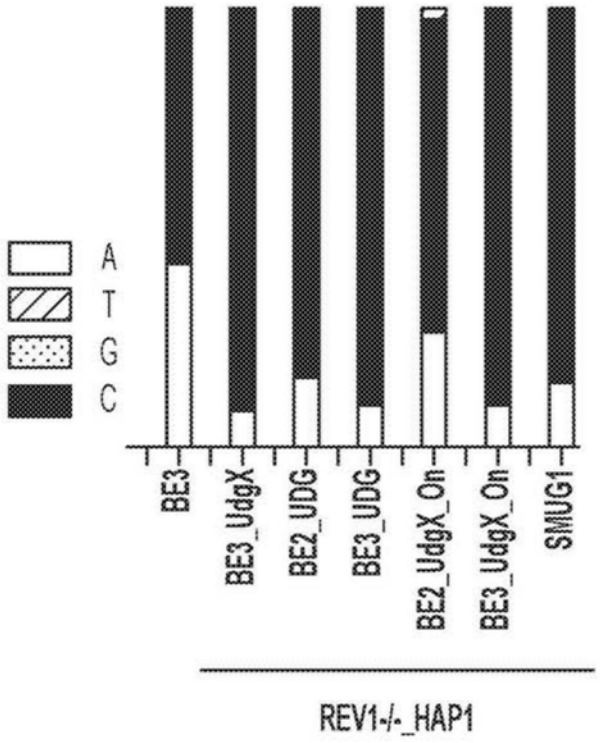
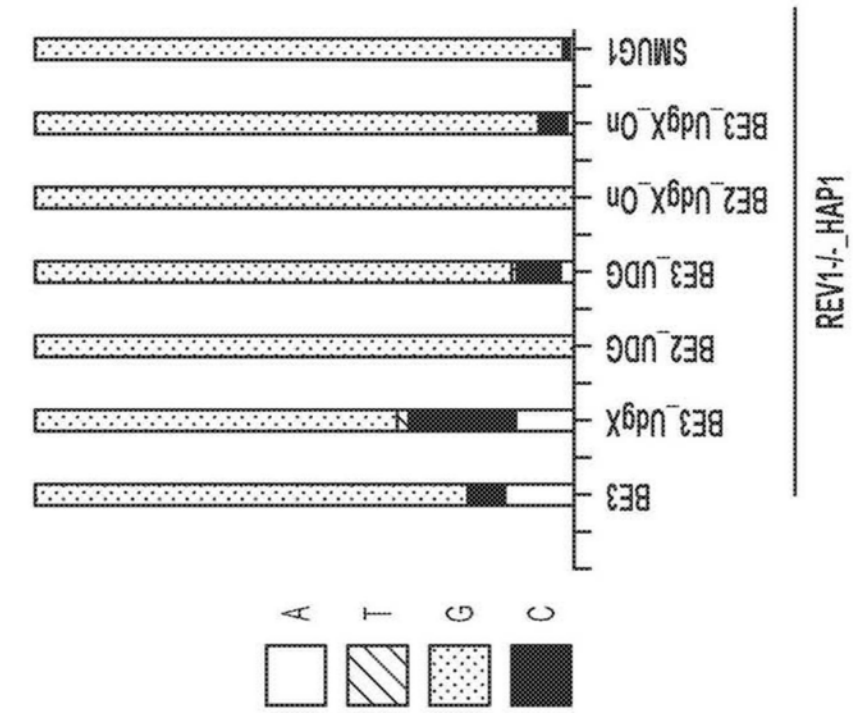
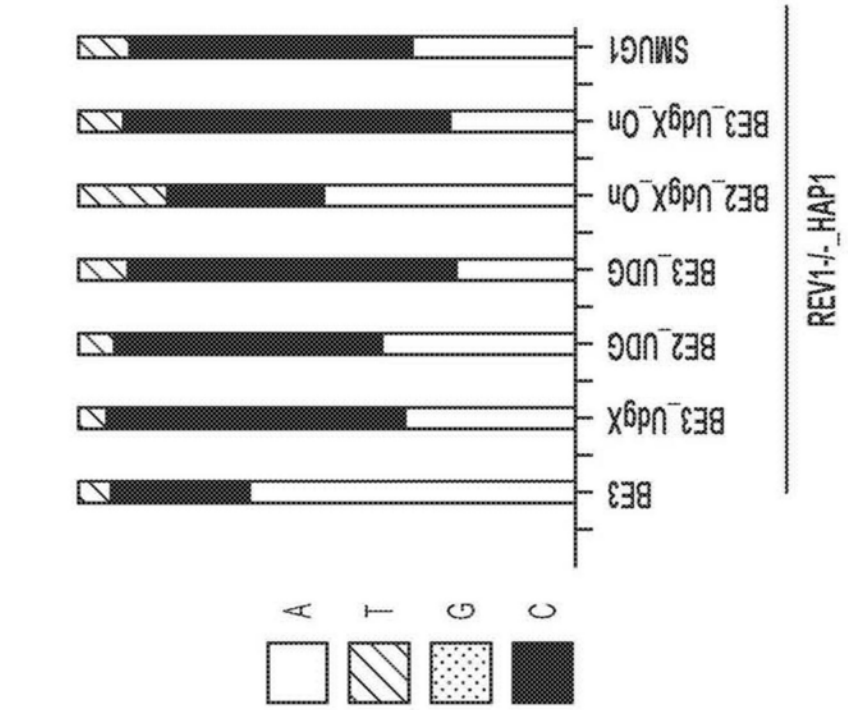



图26




| REV1-/- HAP1 | A | C | T | G |
|---|---|---|---|---|
| BE3 | 13.37062 | 5.67339 | 1.300502 | 79.6555 |
| BE3_UdgX | 11.30944 | 19.73378 | 1.890332 | 67.06644 |
| BE2_UDG | 0.085476 | 0.118742 | 0.015709 | 99.78008 |
| BE3_UDG | 2.875416 | 7.888452 | 1.180925 | 88.05521 |
| BE2_UdgX_On | 0.066387 | 0.041177 | 0.023698 | 99.86874 |
| BE3_UdgX_On | 1.706863 | 4.37727 | 0.625263 | 93.2906 |
| SMUG1 | 0.741226 | 1.273957 | 0.234637 | 97.75018 |

图27




| REV1-/- HAP1 | 总编辑物 | A | C | T |
|---|---|---|---|---|
| BE3 | 20.344512 | 65.7210161 | 27.8865868 | 6.39239712 |
| BE3_UdgX | 32.933552 | 34.3401769 | 59.9199868 | 5.73983638 |
| BE2_UDG | 0.21992751 | 38.865538 | 53.9916084 | 7.14285357 |
| BE3_UDG | 11.944793 | 24.0725478 | 66.0409268 | 9.88652545 |
| BE2_UdgX_On | 0.131260725 | 50.5761844 | 31.3700384 | 18.0537772 |
| BE3_UdgX_On | 6.709395472 | 25.4398874 | 65.2408983 | 9.31921428 |
| SMUG1 | 2.24982028 | 32.946027 | 56.6248126 | 10.4291604 |

图28

REV1-/-_HAP1

| | A | C | T | G |
|---|---|---|---|---|
| BE3 | 1.850796 | 59.33426 | 38.02208 | 0.792859 |
| BE3_UdgX | 3.486185 | 65.86237 | 22.91265 | 7.738802 |
| BE2_UDG | 0.222439 | 98.32175 | 1.1797 | 0.276117 |
| BE3_UDG | 0.840939 | 91.43678 | 4.79438 | 2.927905 |
| BE2_UdgXON | 0.208095 | 98.36564 | 1.119386 | 0.306881 |
| BE3_UDGXON | 2.108873 | 82.15436 | 12.85805 | 2.878713 |
| SMUG1 | 0.428121 | 95.46036 | 3.901712 | 0.20981 |

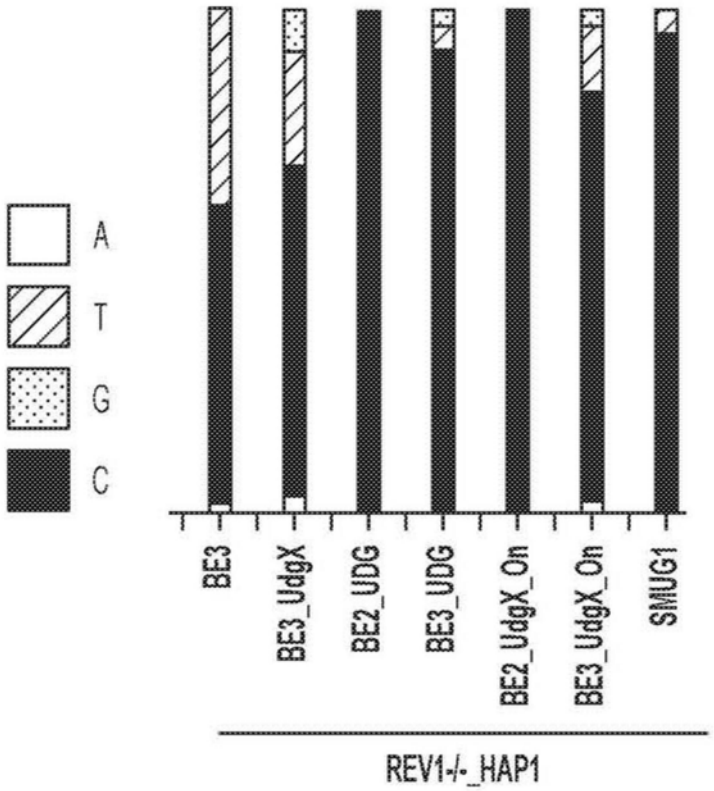



图29

| REV1 -/- HAP1 | 总编辑物 | A | T | G |
|---|---|---|---|---|
| BE3 | 40.665735 | 4.55124197 | 93.4990601 | 1.94969795 |
| BE3_UdgX | 34.137637 | 10.212145 | 67.1184417 | 22.6694132 |
| BE2_UDG | 1.6782564 | 13.2541786 | 70.2931924 | 16.4526291 |
| BE3_UDG | 8.5632244 | 9.82035926 | 55.9880225 | 34.1916183 |
| BE2_UdgXON | 1.634361336 | 12.732475 | 68.490701 | 18.776824 |
| BE3_UDGXON | 17.84563906 | 11.8173022 | 72.051513 | 16.1311847 |
| SMUG1 | 4.539642813 | 9.43071161 | 85.9475655 | 4.62172285 |

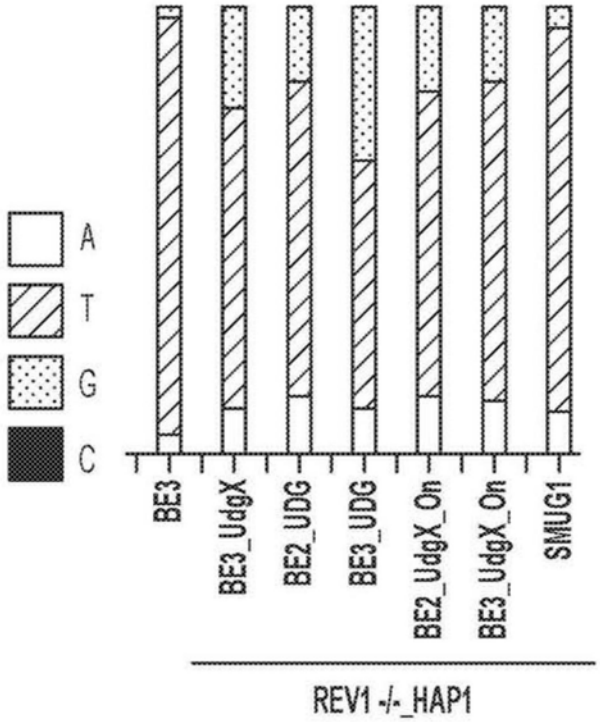



图30

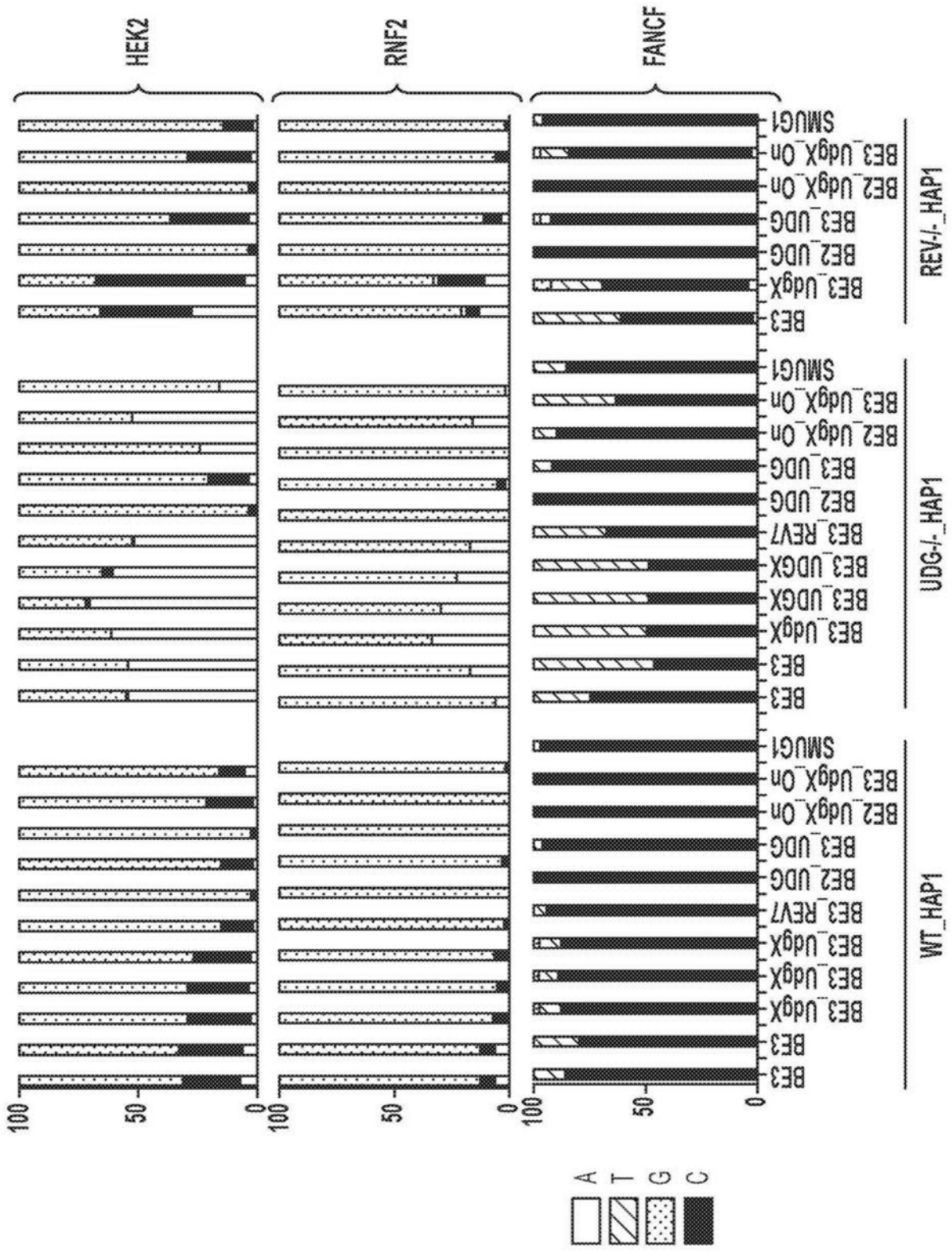

HEK2
RNF2
FANCF
SMUG1
BE3_UdgX_On
BE2_UdgX_On
BE3_UDG
BE2_UDG
BE3_UdgX
BE3
REV-/-_HAP1
SMUG1
BE3_UdgX_On
BE2_UdgX_On
BE3_UDG
BE2_UDG
BE3_REV7
BE3_UDGX
BE3_UDGX
BE3_UdgX
BE3
BE3
UDG-/-_HAP1
SMUG1
BE3_UdgX_On
BE2_UdgX_On
BE3_UDG
BE2_UDG
BE3_REV7
BE3_UdgX
BE3_UdgX
BE3_UdgX
BE3
BE3
WT_HAP1
100
50
0
A
T
G
C

图31

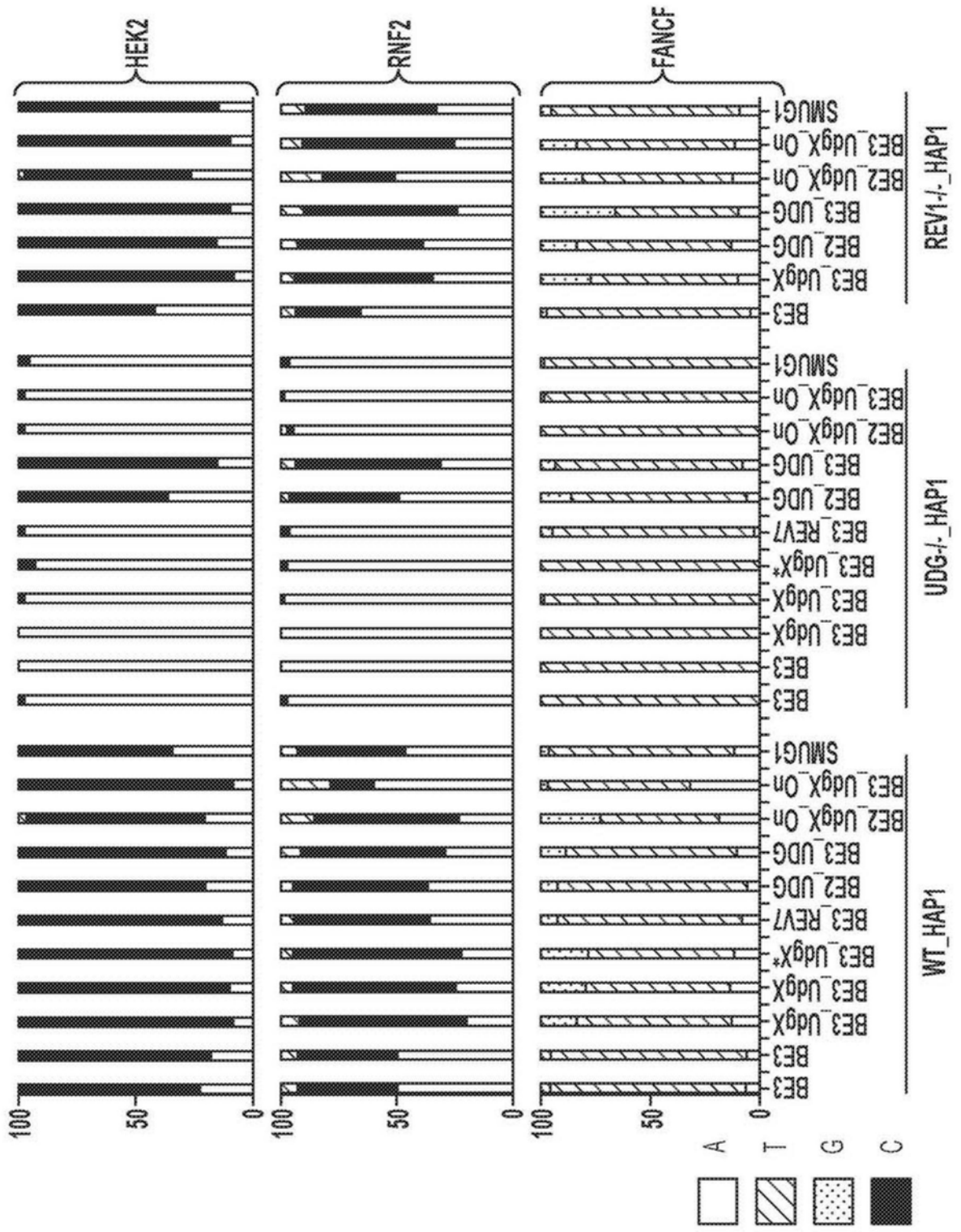

HEK2
RNF2
FANCF
100
50
0
BE3
BE3
BE3_UdgX
BE3_UdgX
BE3_UdgX*
BE3_REV7
BE2_UDG
BE3_UDG
BE2_UdgX_On
BE3_UdgX_On
SMUG1
WT_HAP1
BE3
BE3
BE3_UdgX
BE3_UdgX
BE3_UdgX*
BE3_REV7
BE2_UDG
BE3_UDG
BE2_UdgX_On
BE3_UdgX_On
SMUG1
UDG-/-_HAP1
BE3
BE3_UdgX
BE3_UdgX
BE3_UdgX*
BE3_REV7
BE2_UDG
BE3_UDG
BE2_UdgX_On
BE3_UdgX_On
SMUG1
REV1-/-_HAP1
A
T
G
C

图32

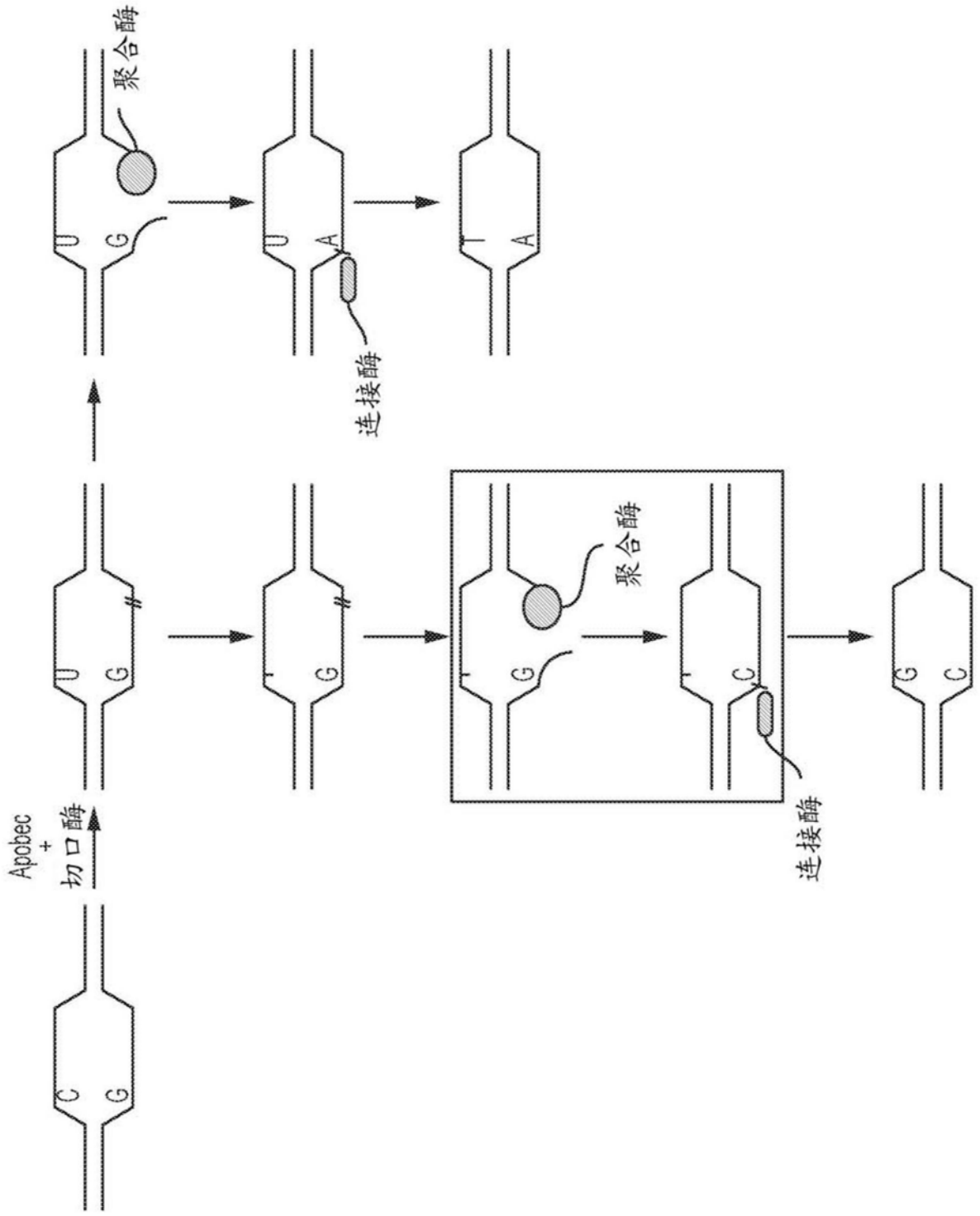



图33

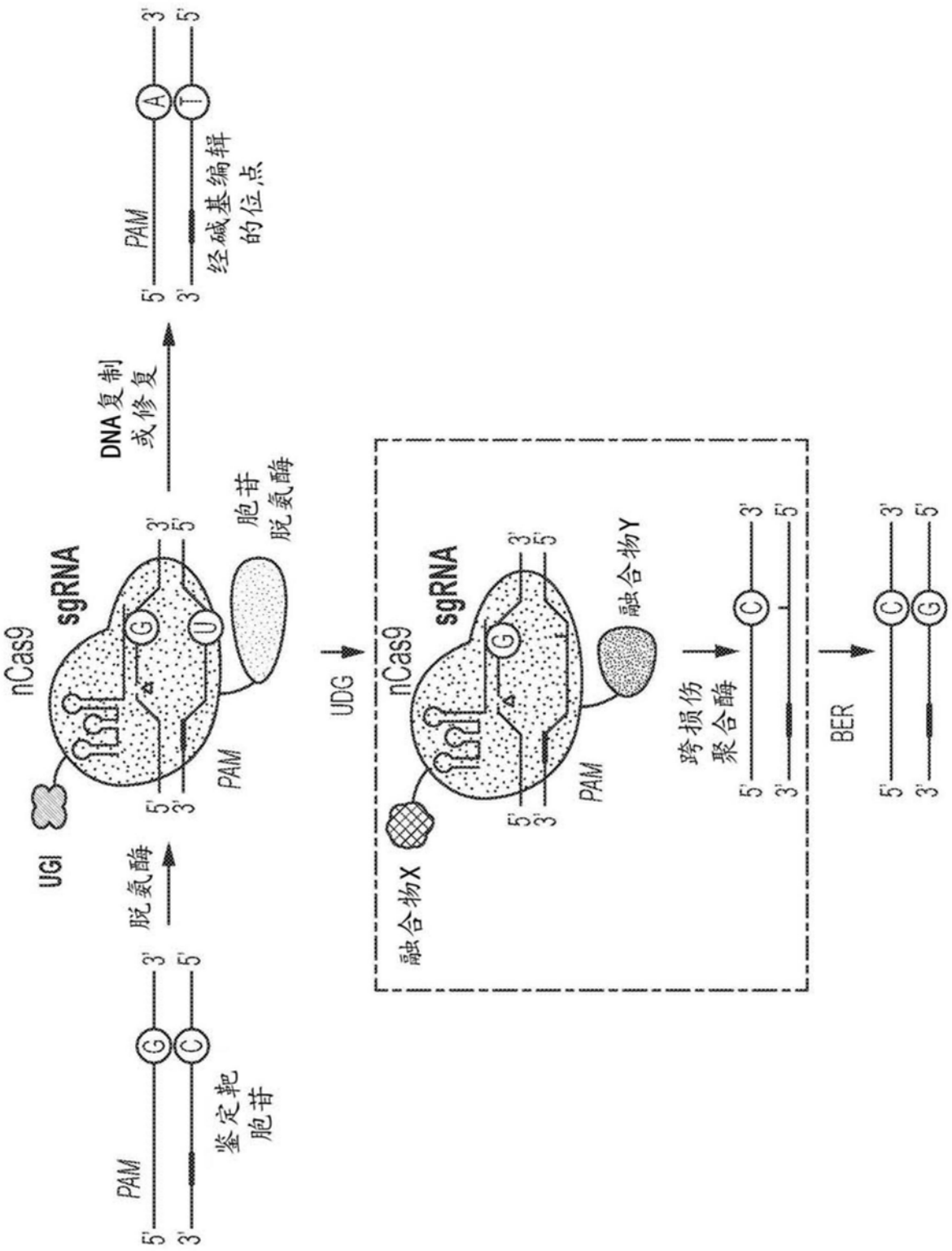

PAM
5'
3'
G
C
鉴定靶
胞苷
脱氨酶
UGI
nCas9
sgRNA
G
U
PAM
胞苷
脱氨酶
DNA复制
或修复
PAM
A
T
经碱基编辑
的位点
UDG
融合物X
融合物Y
跨损伤
聚合酶
BER

图34

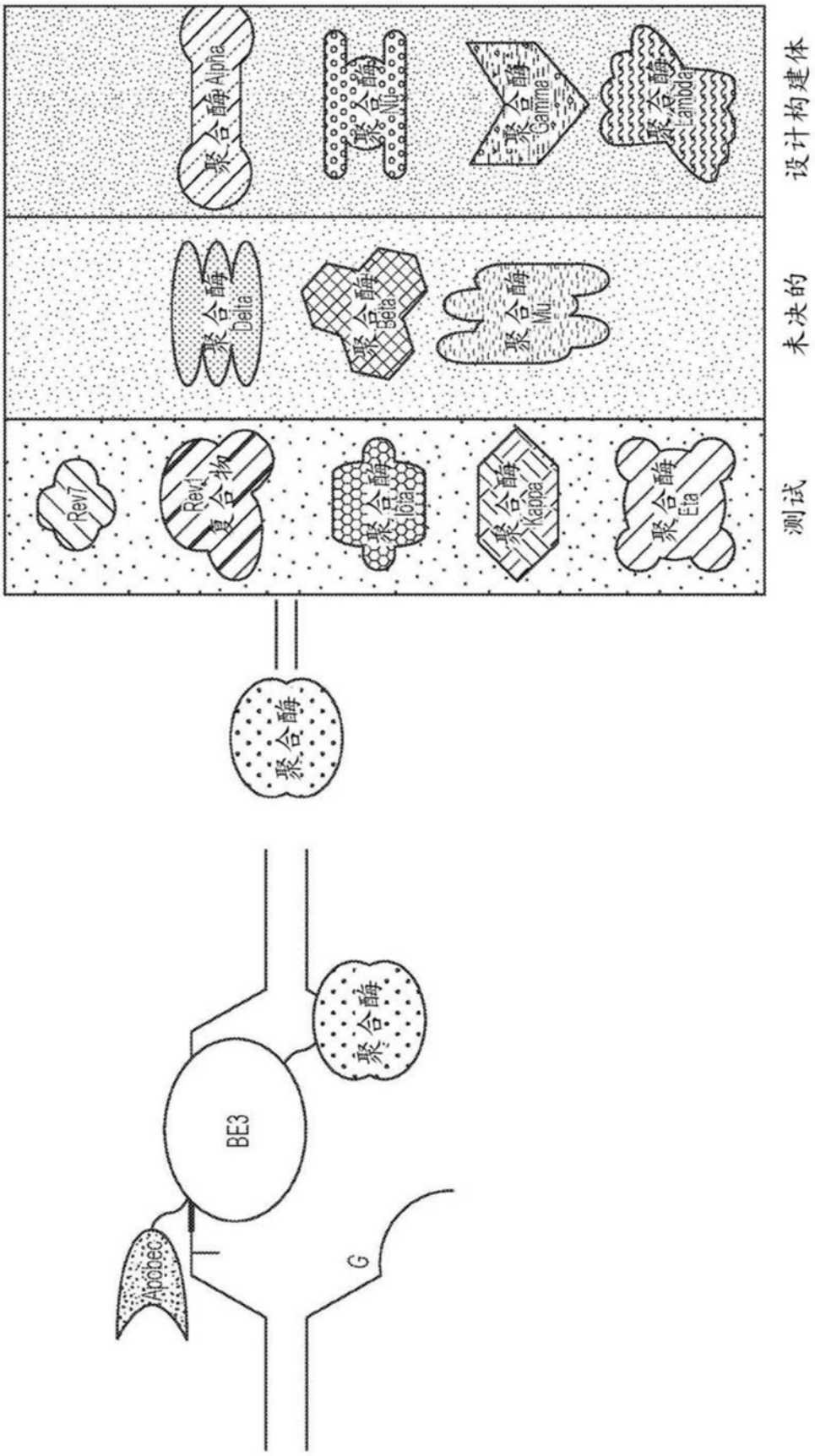

设计构建体
未决的
测试
聚合酶 Alpha
聚合酶 Nu
聚合酶 Gamma
聚合酶 Lamda
聚合酶 Delta
聚合酶 Beta
聚合酶 Mu
Rev7
Rev1 聚合物
聚合酶 Iota
聚合酶 Kappa
聚合酶 Eta
聚合酶
BE3
聚合酶
Apobec
T
G

图35

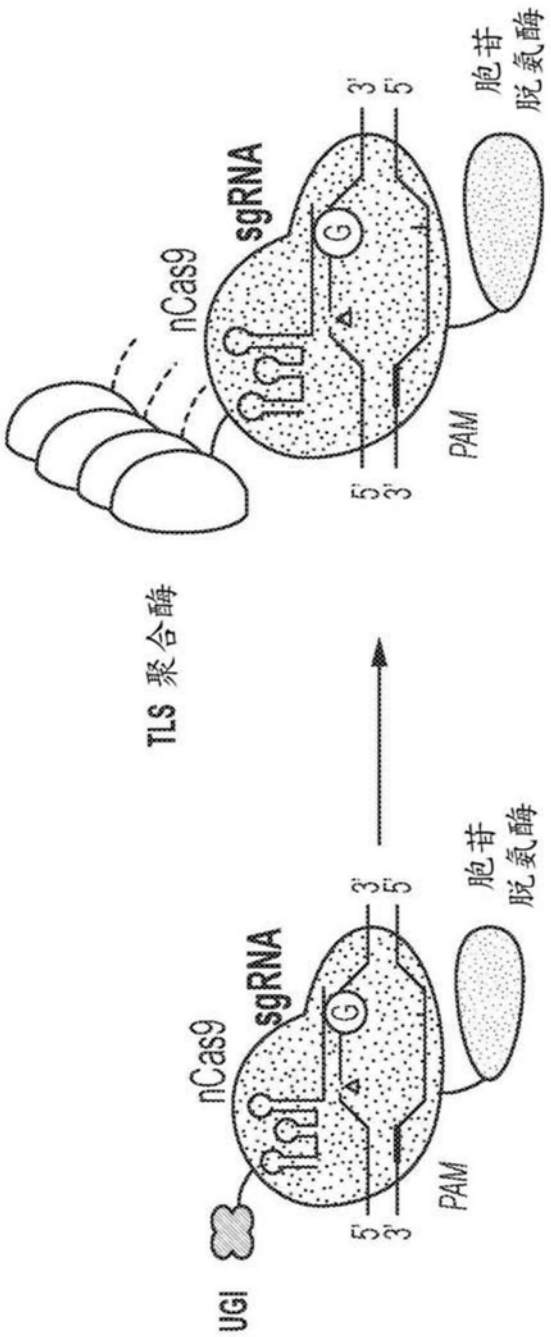



图36

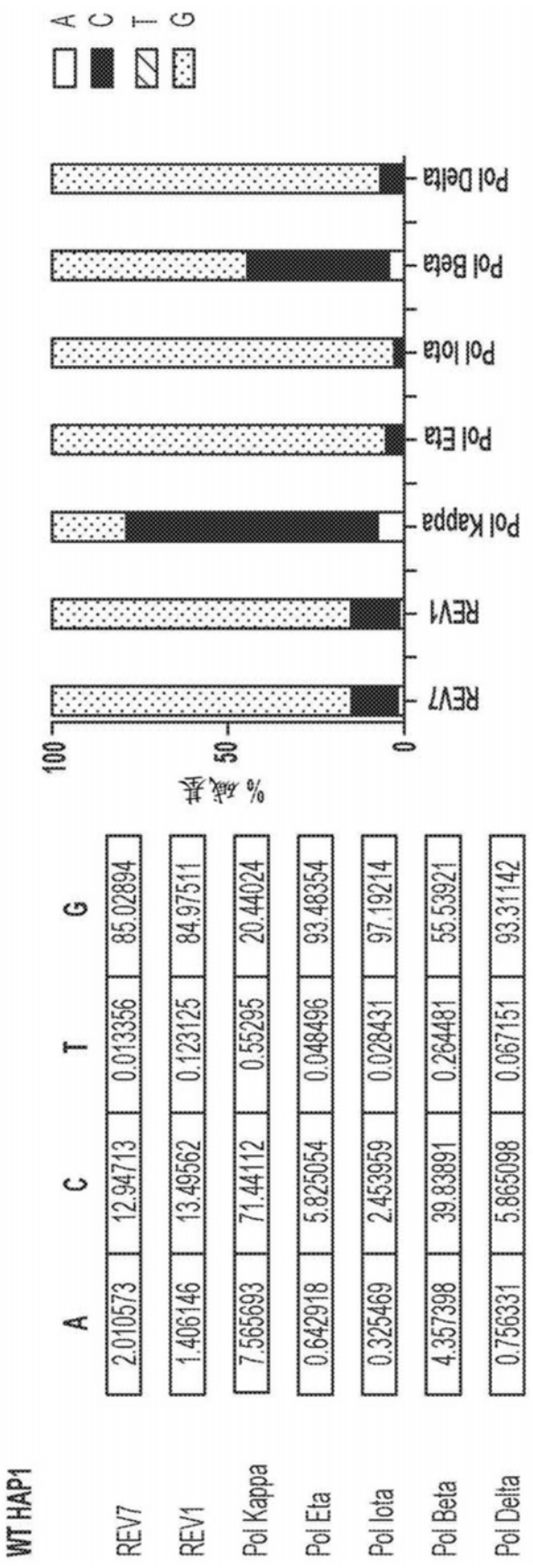

A
C
T
G
% 碱基
100
50
0
REV7
REV1
Pol Kappa
Pol Eta
Pol Iota
Pol Beta
Pol Delta
WT HAP1
A C T G
REV7 2.010573 12.94713 0.013356 85.02894
REV1 1.406146 13.49562 0.123125 84.97511
Pol Kappa 7.565693 71.44112 0.55295 20.44024
Pol Eta 0.642918 5.825054 0.048496 93.48354
Pol Iota 0.325469 2.453959 0.028431 97.19214
Pol Beta 4.357398 39.83891 0.264481 55.53921
Pol Delta 0.756331 5.865098 0.067151 93.31142

图37

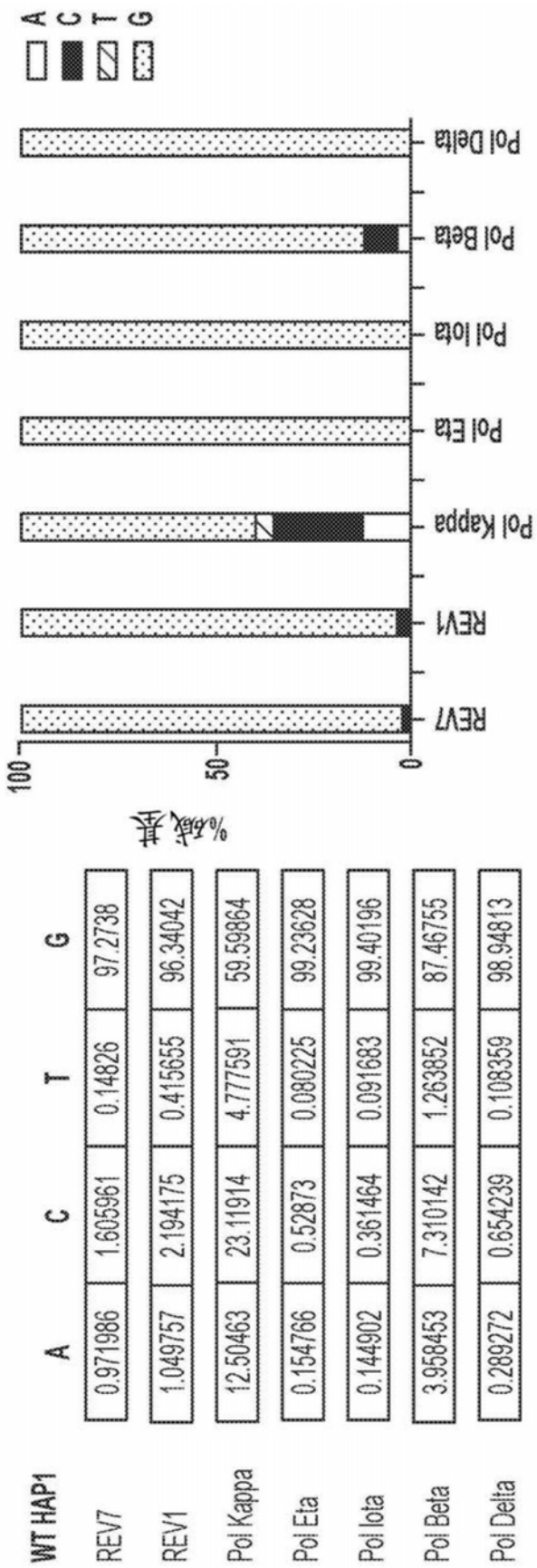
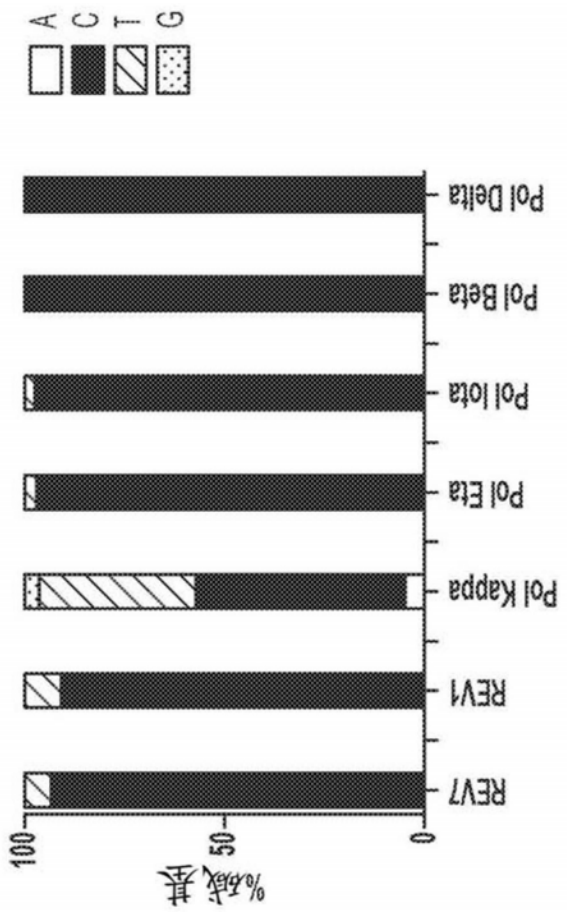

A
C
T
G
Pol Delta
Pol Beta
Pol Iota
Pol Eta
Pol Kappa
REV1
REV7
100
50
0
%碱基
WT HAP1
REV7
REV1
Pol Kappa
Pol Eta
Pol Iota
Pol Beta
Pol Delta
A
C
T
G
0.971986 1.605961 0.14826 97.2738
1.049757 2.194175 0.415655 96.34042
12.50463 23.11914 4.777591 59.59864
0.154766 0.52873 0.080225 99.23628
0.144902 0.361464 0.091683 99.40196
3.958453 7.310142 1.263852 87.46755
0.289272 0.654239 0.108359 98.94813

图38




| WT HAP1 | A | C | T | G |
|---|---|---|---|---|
| REV7 | 0.514632 | 93.70418 | 5.31733 | 0.463664 |
| REV1 | 0.914949 | 90.84767 | 7.636108 | 0.601268 |
| Pol Kappa | 4.345517 | 53.269 | 39.77689 | 2.608598 |
| Pol Eta | 0.266049 | 97.62926 | 1.852404 | 0.252288 |
| Pol Iota | 0.178977 | 98.23811 | 1.49925 | 0.083658 |
| Pol Beta | 0.001701 | 99.94274 | 0.052159 | 0.003402 |
| Pol Delta | 0.000882 | 99.98706 | 0.010879 | 0.001176 |

图39

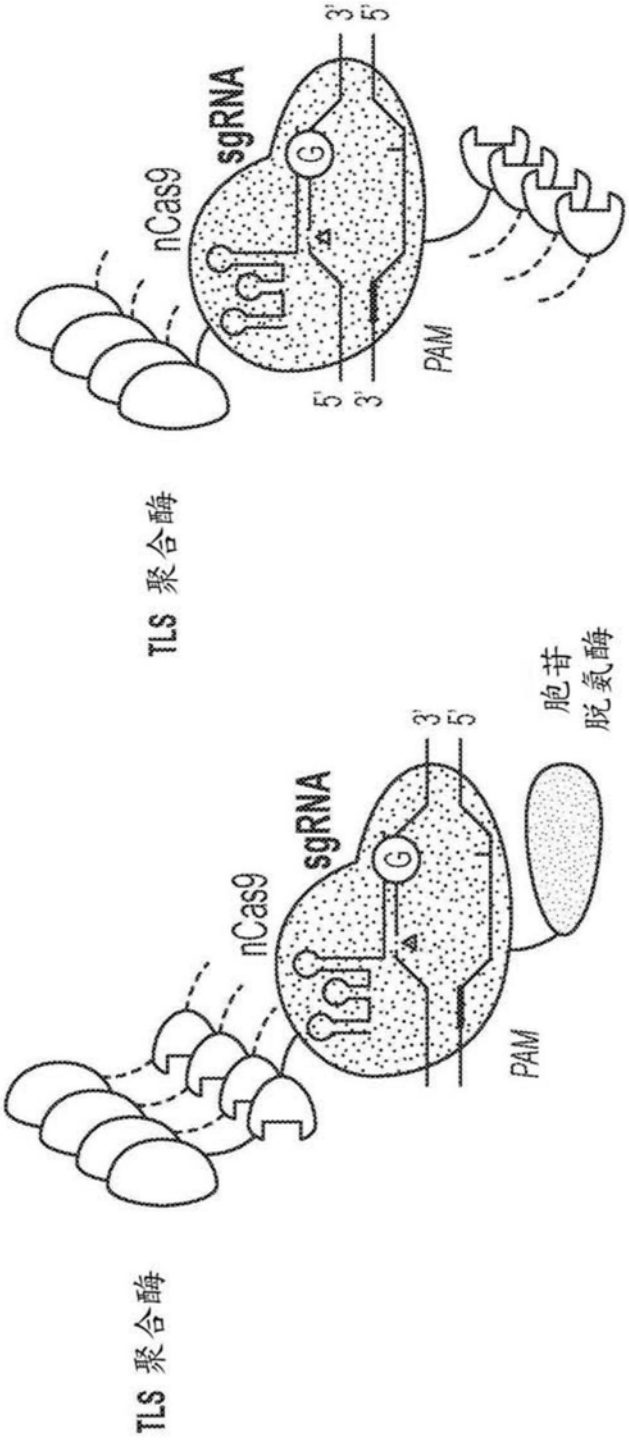

3'
5'
sgRNA
G
nCas9
TLS
5'
3'
PAM
TLS 聚合酶
胞苷
脱氨酶
3'
5'
sgRNA
G
nCas9
TLS
PAM
TLS 聚合酶

图40

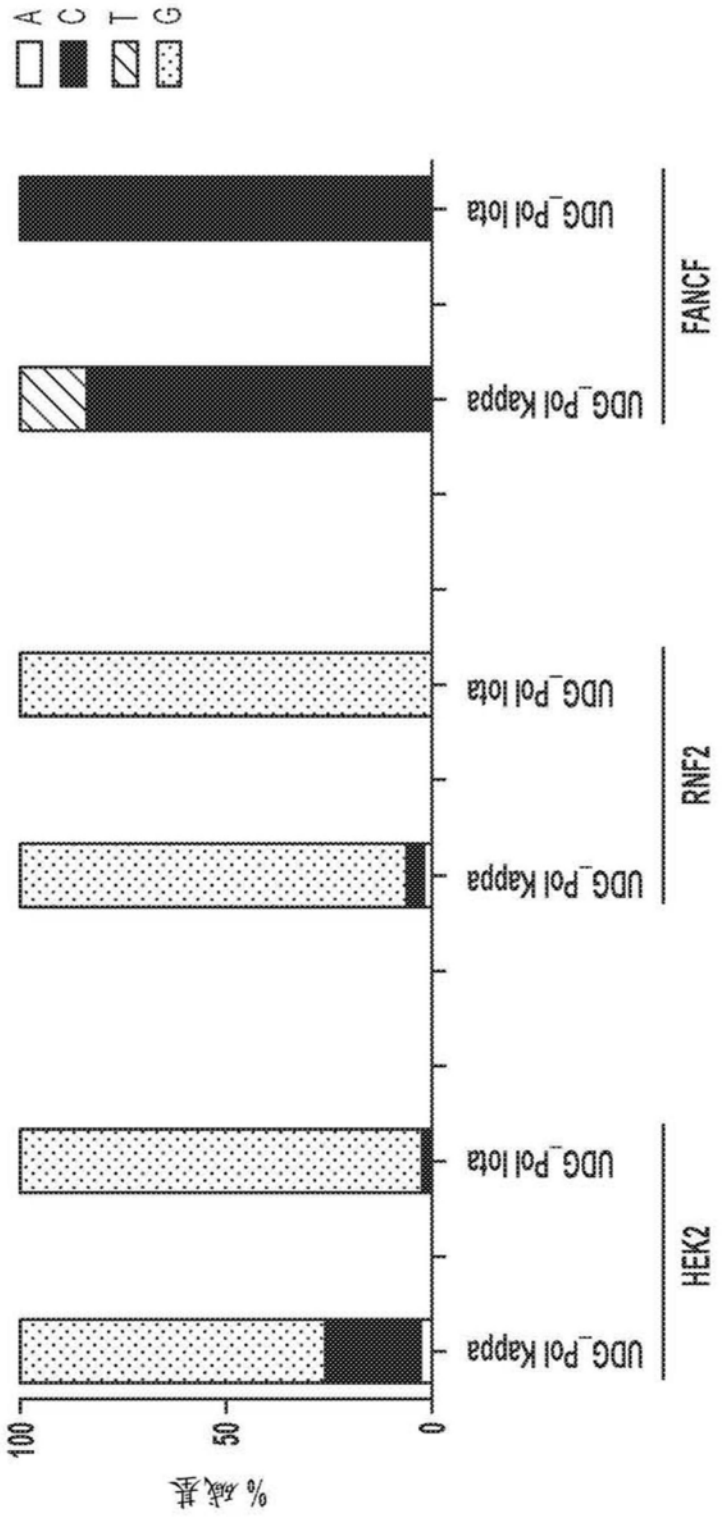

A
C
T
G
UDG_Pol iota
UDG_Pol kappa
FANCF
UDG_Pol iota
UDG_Pol kappa
RNF2
UDG_Pol iota
UDG_Pol kappa
HEK2
100
50
0
% 碱基

图41

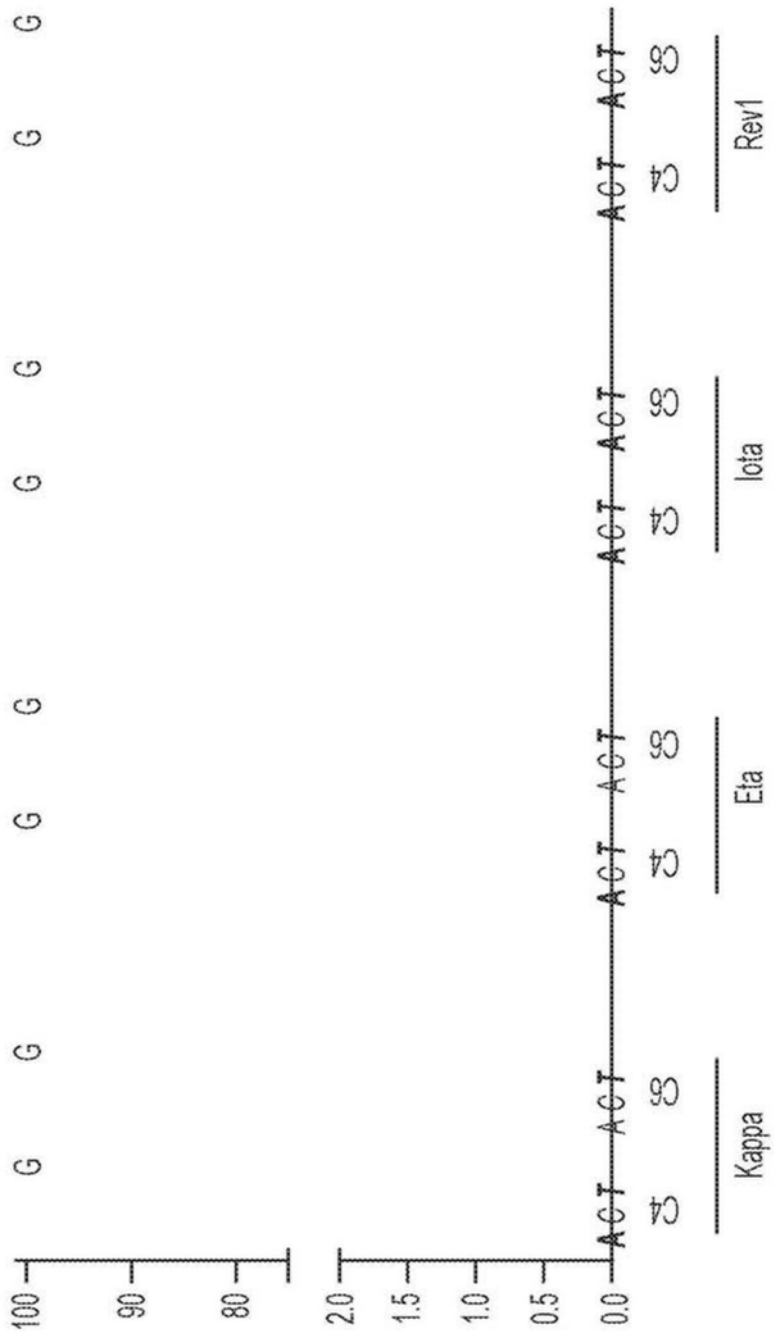

G
G
G
G
G
G
G
G
G
G
100
90
80
2.0
1.5
1.0
0.5
0.0
ACT
ACT
ACT
ACT
ACT
ACT
ACT
ACT
ACT
ACT
C4
C6
C4
C6
C4
C6
C4
C6
Kappa
Eta
Iota
Rev1

图42

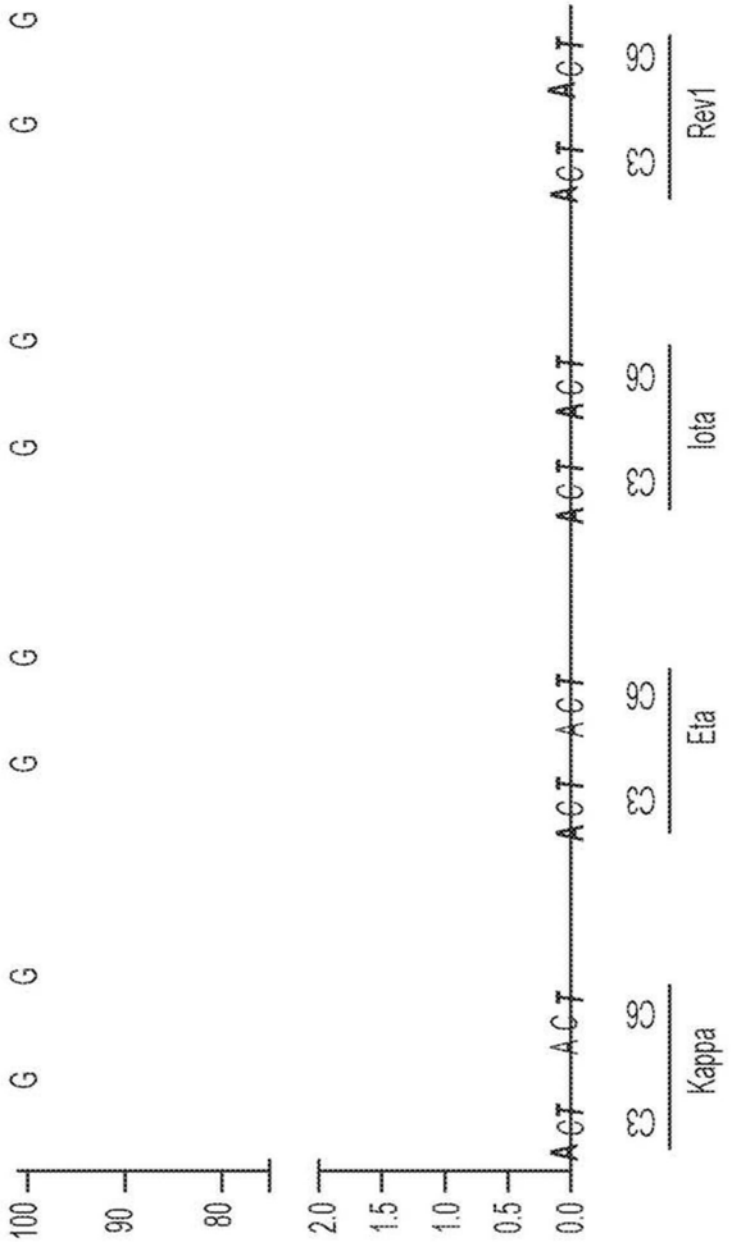



图43

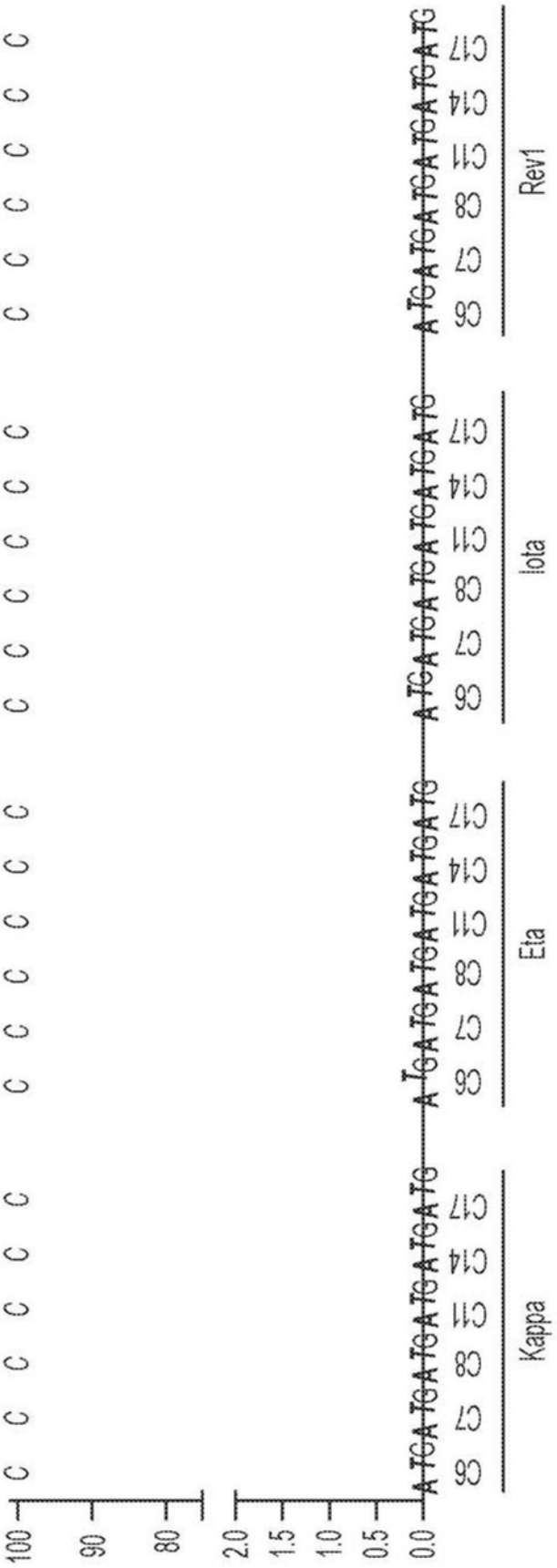



图44

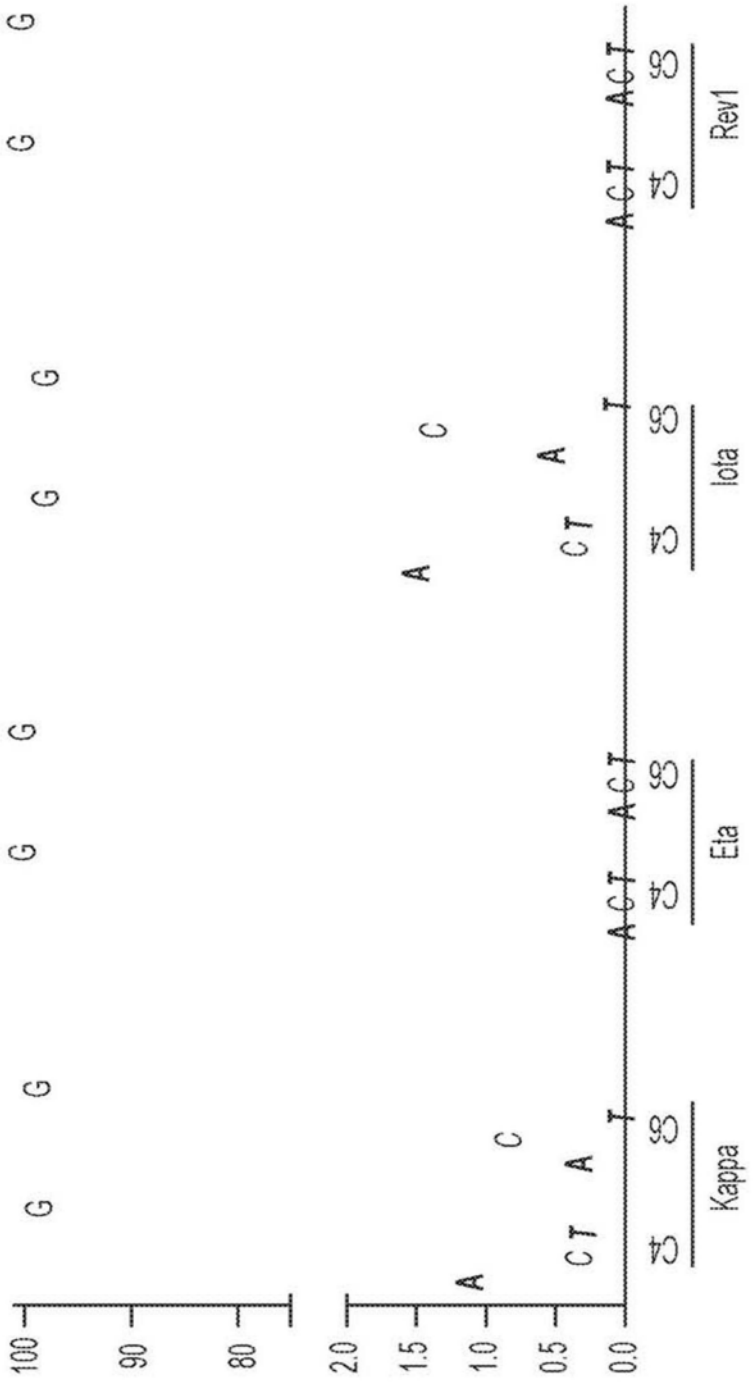



图45

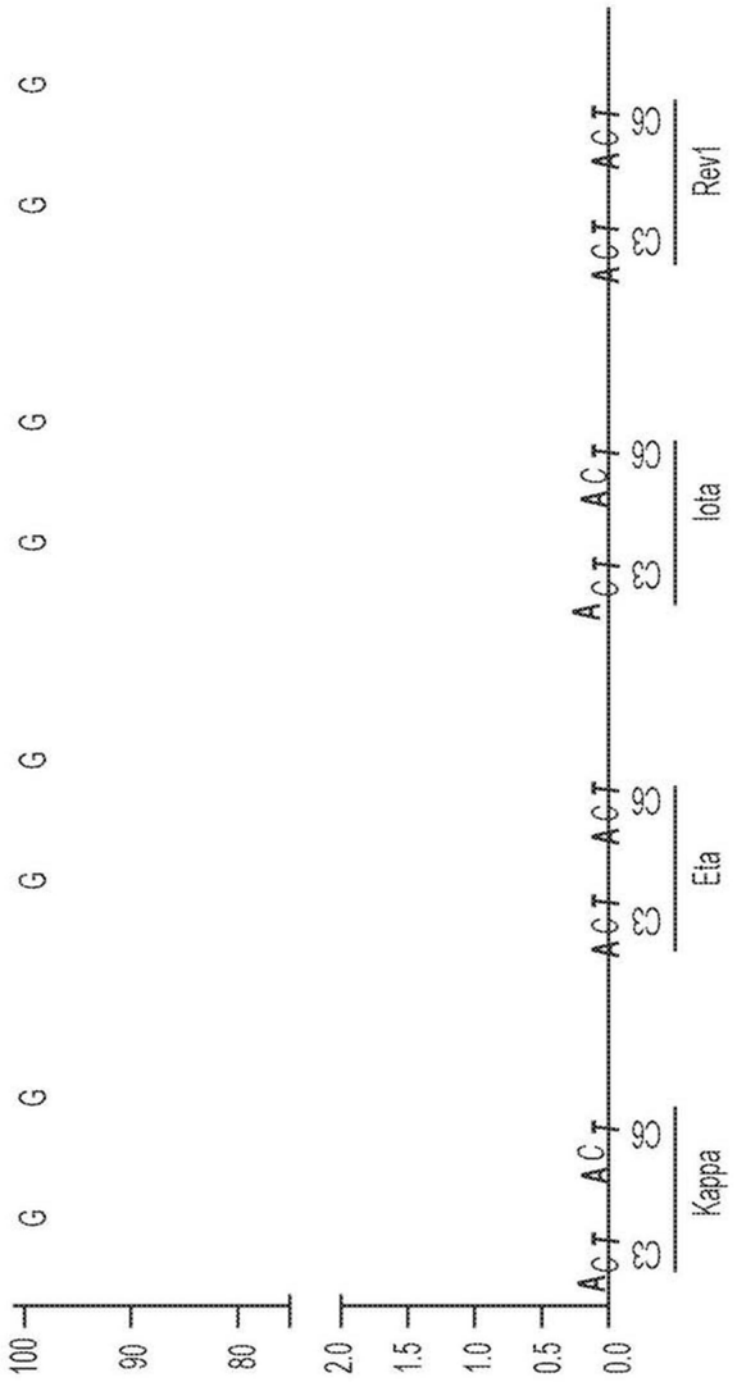



图46

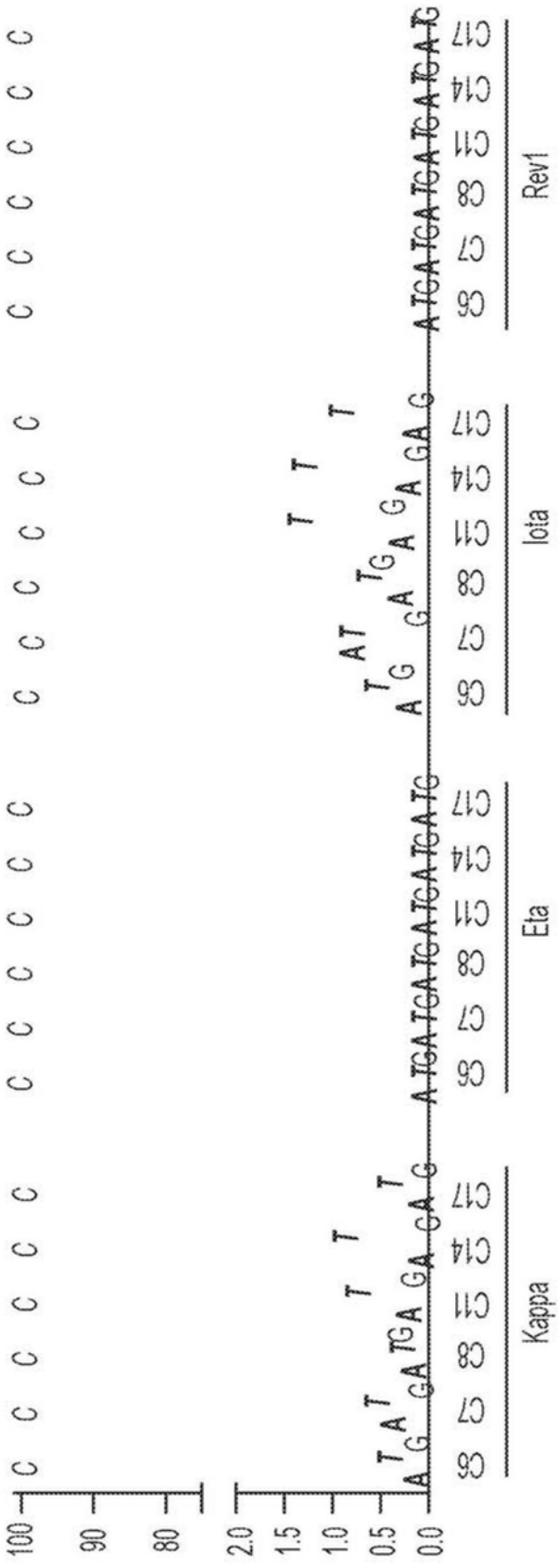



图47

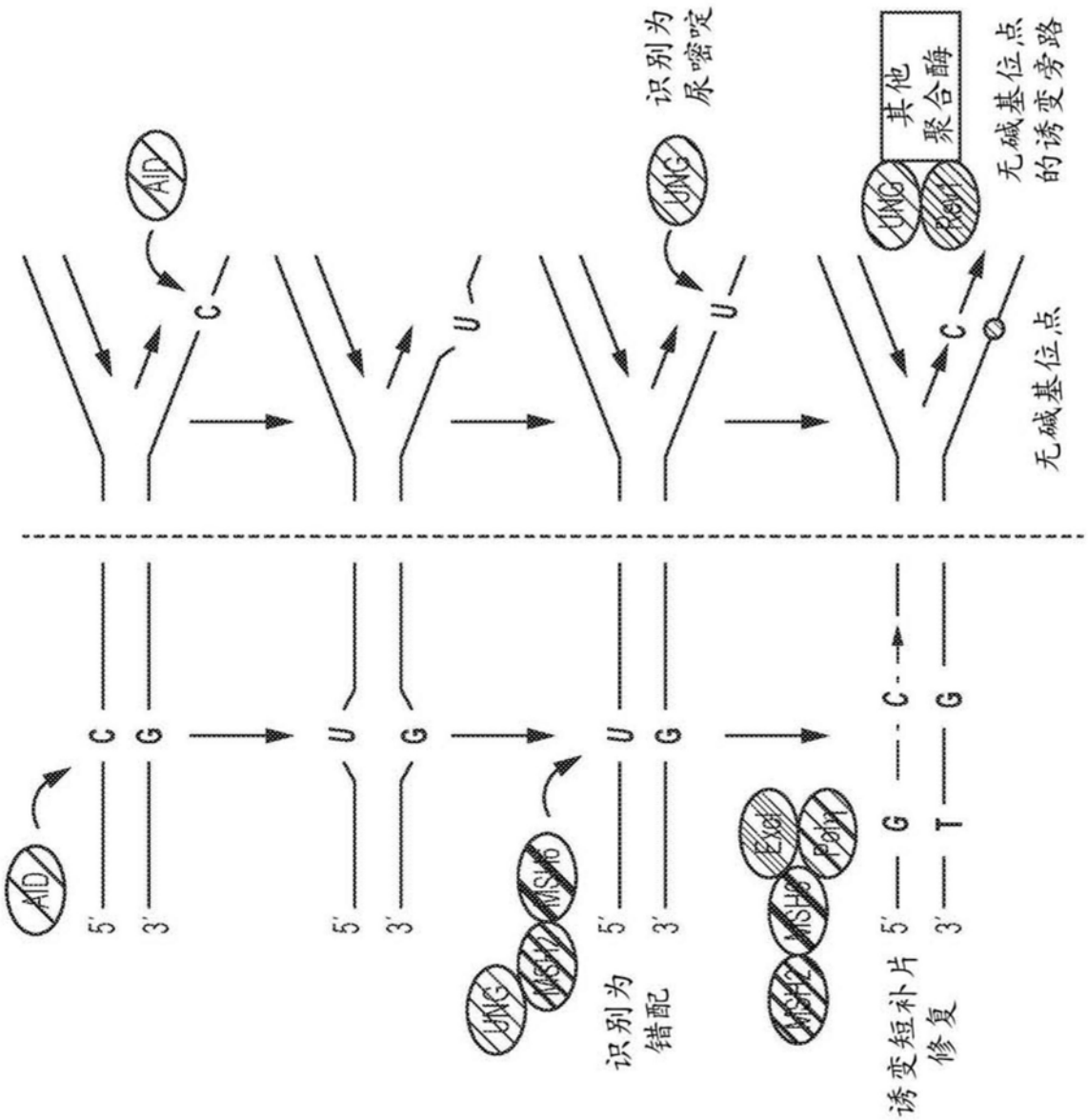
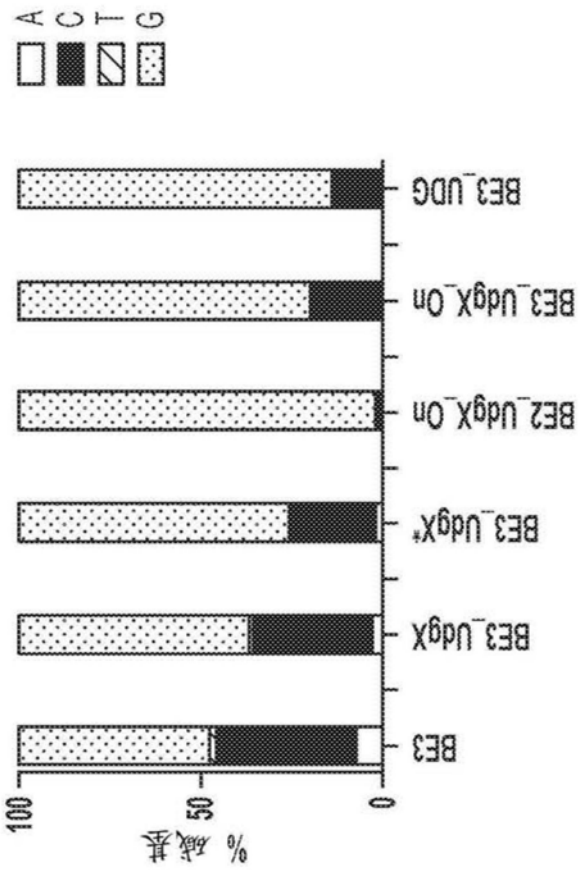
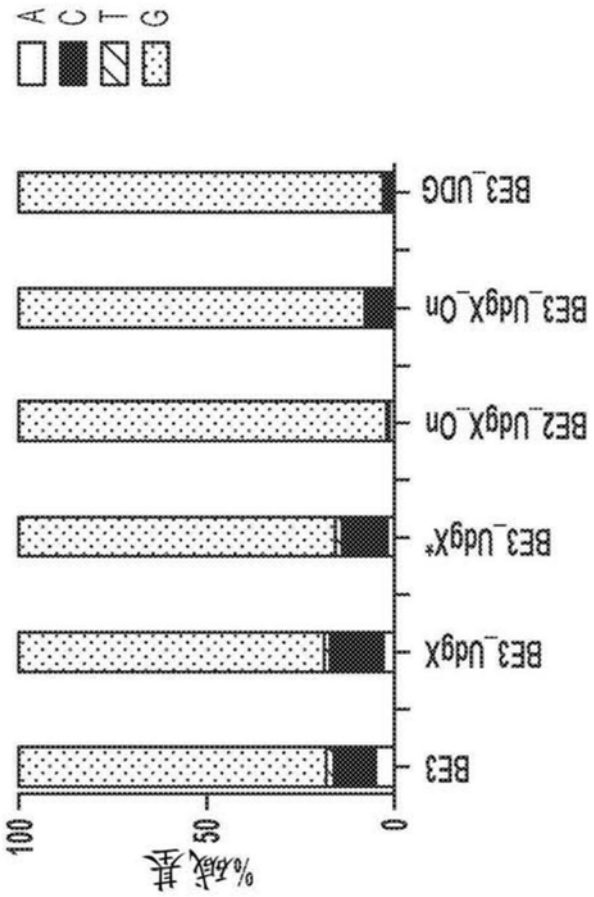

AID
C
G
5′
3′
U
识别为
错配
UNG
MSH2
MSH6
Exo1
Pol η
识别为
尿嘧啶
其他
聚合酶
Rev1
无碱基位点
无碱基位点
的诱变旁路
诱变短补片
修复
T

图48




| MSH2-/- HAP1 | A | C | T | G |
|---|---|---|---|---|
| BE3 | 7.531235 | 38.16771 | 1.534232 | 52.76683 |
| BE3_UdgX | 2.716642 | 32.76543 | 0.906873 | 63.61105 |
| BE3_UDGX* | 2.13603 | 23.11333 | 0.617898 | 74.13274 |
| BE2_UdgX_On | 0.20333 | 2.012962 | 0.080061 | 97.70365 |
| BE3_UdgX_On | 1.533642 | 18.05043 | 0.530436 | 79.8855 |
| BE3_UDG | 1.36582 | 12.56294 | 0.433865 | 85.63737 |

图49




| MSH2-/- HAP1 | A | C | T | G |
|---|---|---|---|---|
| BE3 | 8.072351 | 9.490587 | 1.359173 | 81.07789 |
| BE3_UdgX | 2.928139 | 15.37726 | 1.290227 | 80.40437 |
| BE3_UDGX* | 1.72001 | 13.87825 | 0.795514 | 83.60623 |
| BE2_UdgX_On | 1.994799 | 0.789305 | 0.110912 | 97.10498 |
| BE3_UdgX_On | 1.403443 | 7.205661 | 0.596091 | 90.7948 |
| BE3_UDG | 0.983205 | 3.27285 | 0.368807 | 95.37514 |

图50

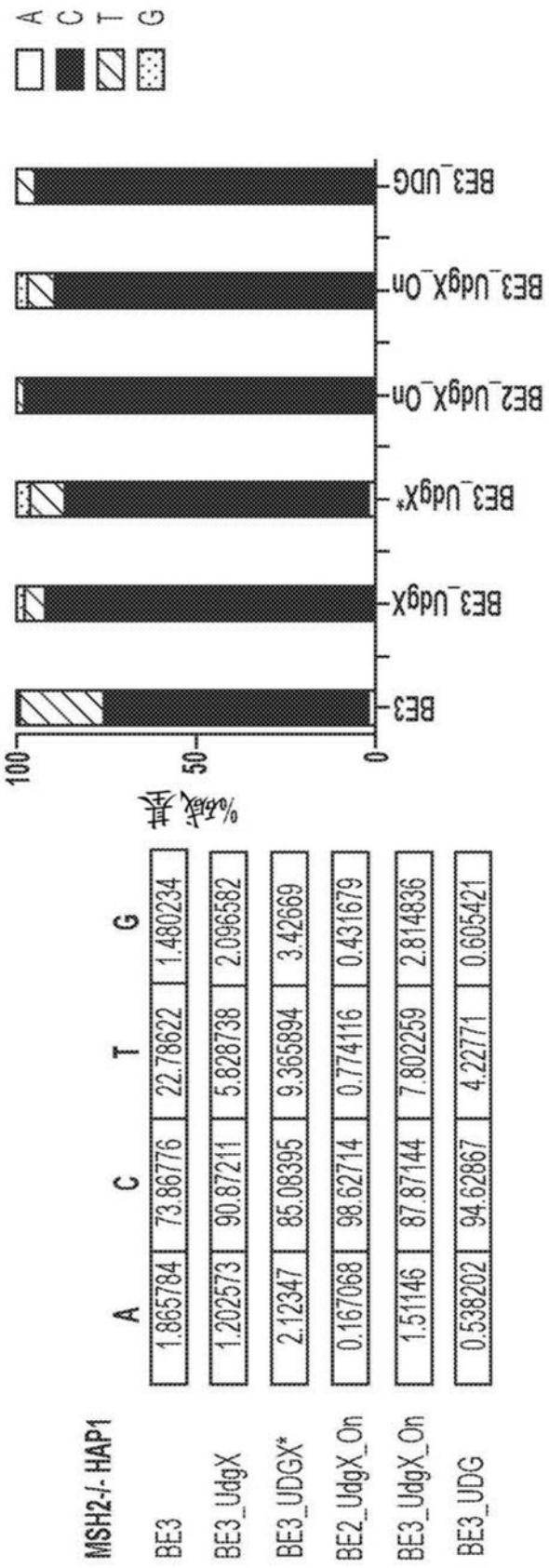

A C T G
BE3
BE3_UdgX
BE3_UdgX*
BE2_UdgX_On
BE3_UdgX_On
BE3_UDG
0
50
100
%碱基
MSH2-/- HAP1
A C T G
1.865784 73.86776 22.78622 1.480234
1.202573 90.87211 5.828738 2.096582
2.12347 85.08395 9.365894 3.42669
0.167068 98.62714 0.774116 0.431679
1.51146 87.87144 7.802259 2.814836
0.538202 94.62867 4.22771 0.605421

图51

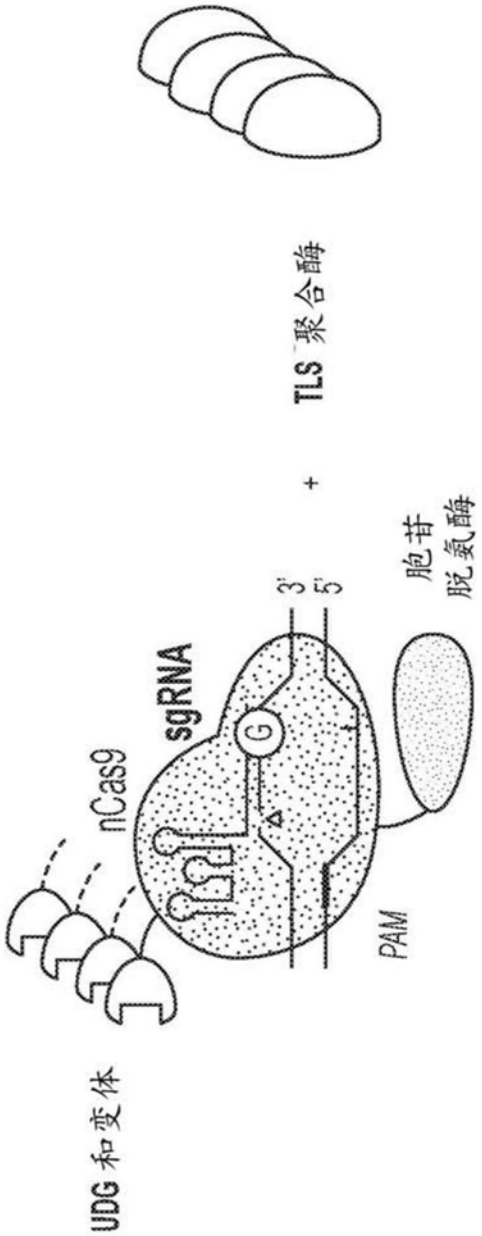



图52

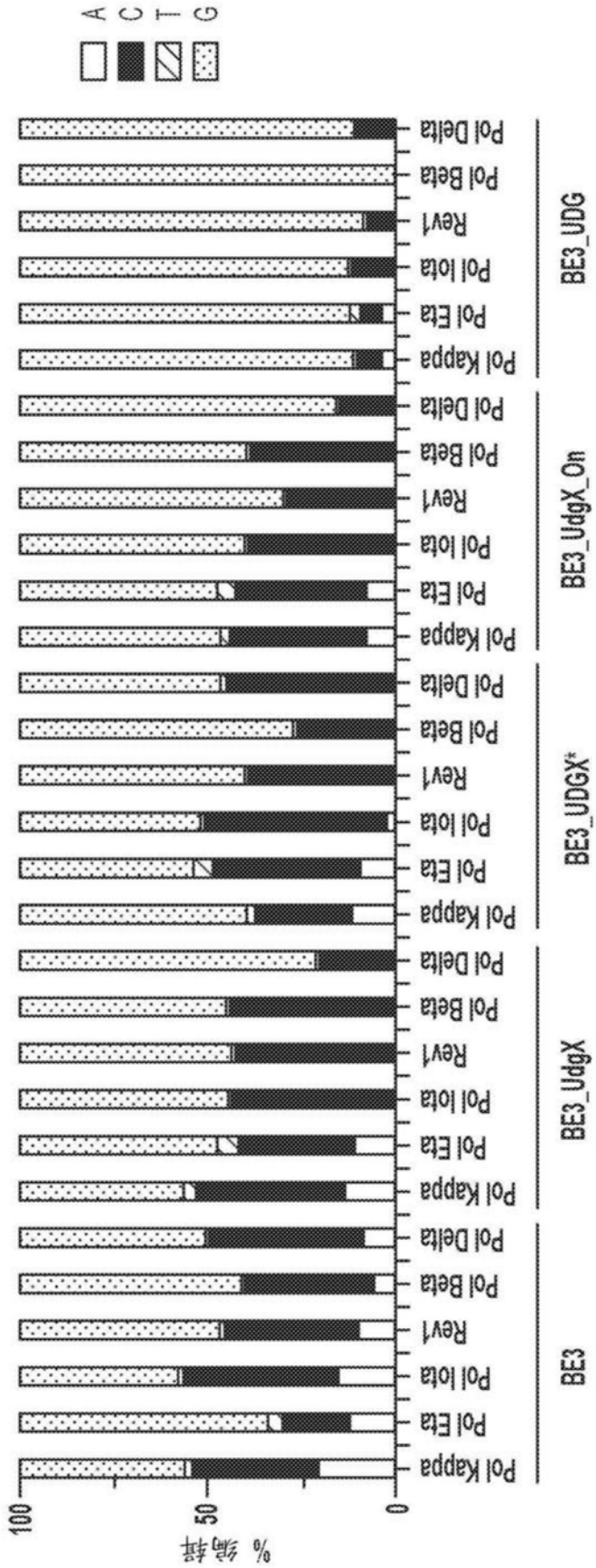

A
C
T
G
Pol Delta
Pol Beta
Rev1
Pol Iota
Pol Eta
Pol Kappa
BE3_UDG
BE3_UdgX_On
BE3_UDGX*
BE3_UdgX
BE3
100
50
0
% 编辑

图53

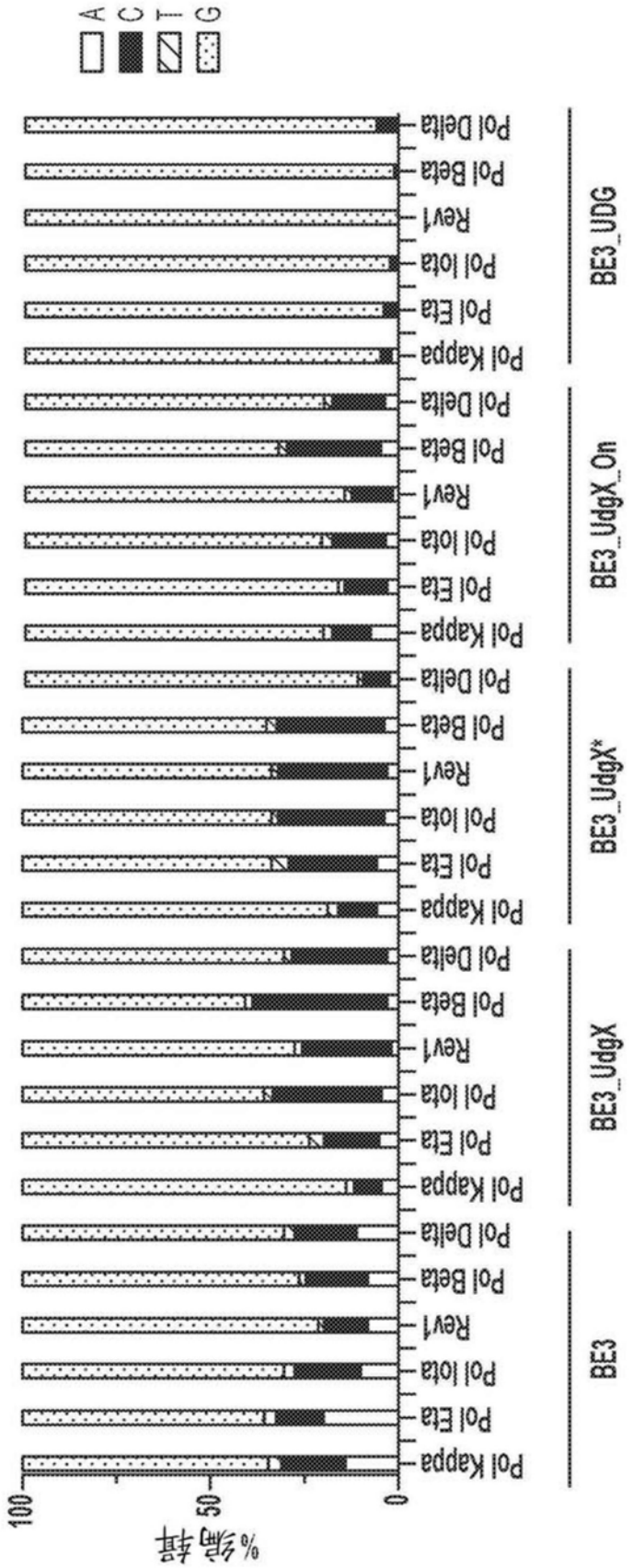

A
C
T
G
Pol Delta
Pol Beta
Rev1
Pol Iota
Pol Eta
Pol Kappa
BE3_UDG
BE3_UdgX_On
BE3_UdgX*
BE3_UdgX
BE3
100
50
0
% 编辑

图54

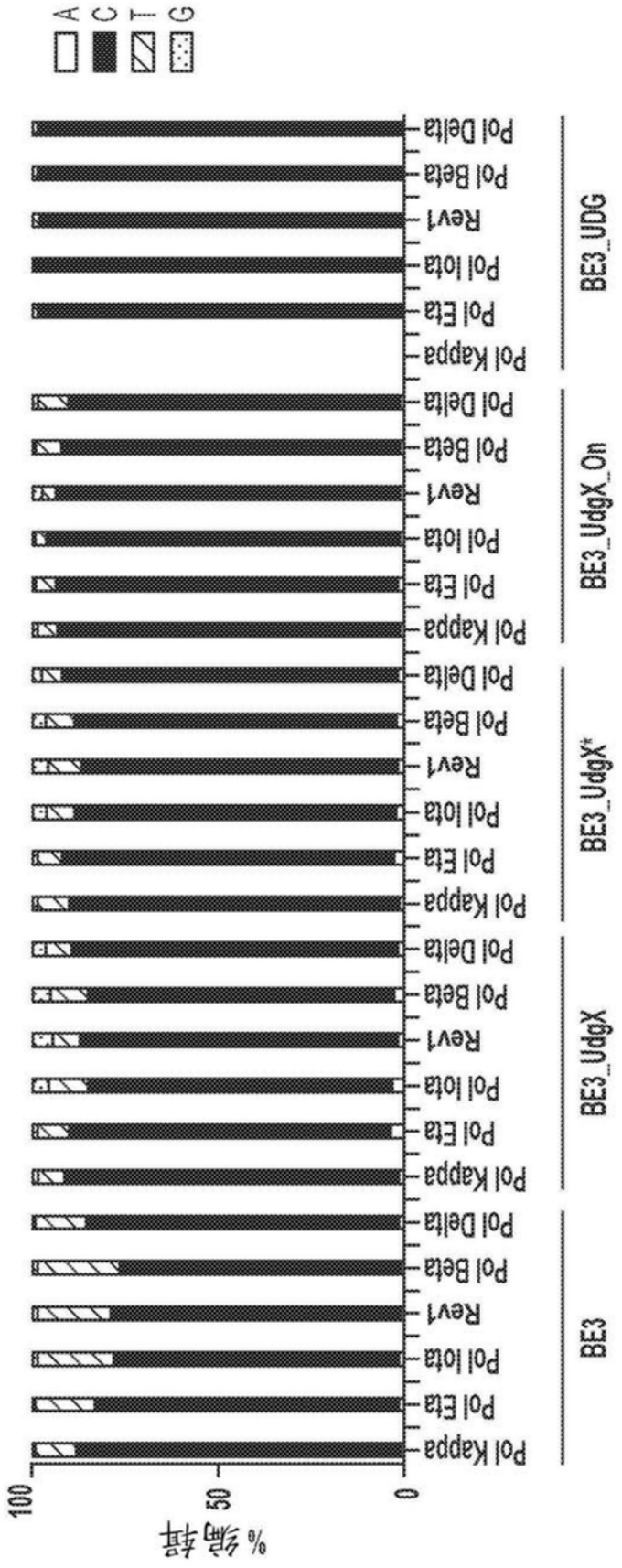

A
C
T
G
Pol Delta
Pol Beta
Rev1
Pol Iota
Pol Eta
Pol Kappa
BE3_UDG
BE3_UdgX_On
BE3_UdgX*
BE3_UdgX
BE3
100
50
0
% 编辑

图55