



(12)发明专利

(10)授权公告号 CN 106460005 B

(45)授权公告日 2020.01.03

(21)申请号 201580024026.2

(22)申请日 2015.03.27

(65)同一申请的已公布的文献号
申请公布号 CN 106460005 A

(43)申请公布日 2017.02.22

(30)优先权数据
PCT/CA2014/050326 2014.03.28 CA
61/971,274 2014.03.27 US

(85)PCT国际申请进入国家阶段日
2016.11.07

(86)PCT国际申请的申请数据
PCT/CA2015/050240 2015.03.27

(87)PCT国际申请的公布数据
W02015/143567 EN 2015.10.01

(73)专利权人 莫迪卡戈公司
地址 加拿大魁北克省

(72)发明人 马克-安德烈·蒂奥斯特
皮埃尔-奥利维耶·拉瓦伊

(74)专利代理机构 北京英赛嘉华知识产权代理
有限责任公司 11204

代理人 王达佐 洪欣

(51)Int.Cl.
C12N 15/82(2006.01)

(56)对比文件
CN 103998601 A, 2014.08.20,
CN 103930435 A, 2014.07.16,
CN 101952446 A, 2011.01.19,
CN 103282501 A, 2013.09.04,
Selvaraju Kanagarajan等. Transient
Expression of Hemagglutinin Antigen from
Low Pathogenic Avian Influenza A (H7N7)
in Nicotiana benthamiana.《PLOS ONE》.2012,
第7卷(第3期), e33010, 第1-10页.

Francesca De Amicis等. Improvement of
the pBI121 plant expression vector by
leader replacement with a sequence
combining a poly(CAA) and a CT motif..
《Transgenic Res》.2007, 第16卷731-738.

审查员 彭郁葱

权利要求书2页 说明书35页

序列表48页 附图35页

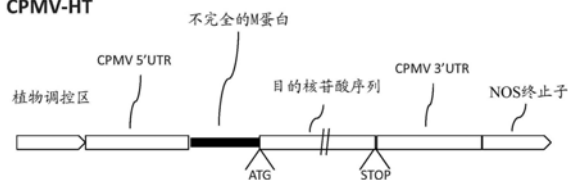
(54)发明名称

修饰的CPMV增强子元件

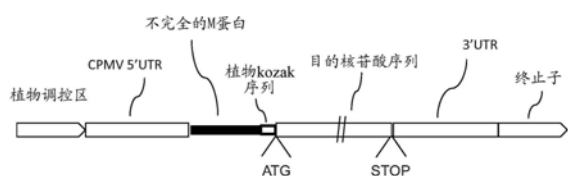
(57)摘要

提供了表达增强子,其依次包含CPMV 5'UTR核苷酸序列和填充片段,所述CPMV 5'UTR核苷酸序列包含SEQ ID NO:1的核苷酸1-160或者包含与SEQ ID NO:1具有约80%至100%的序列相似性的核苷酸序列;所述填充片段包含编码不完全的M蛋白的核苷酸序列以及在植物中有活性的一种或多种kozak序列。还描述了包含所述表达增强子的植物和植物物质,以及使用所述表达增强子的方法。

CPMV-HT



CPMV-HT+



1. 表达增强子,其依次包含CPMV 5' UTR和填充片段,所述CPMV 5' UTR的核苷酸序列由核苷酸1-160组成,其中核苷酸1-114和118-160即为SEQ ID NO:1的核苷酸1-114和118-160,核苷酸115-117是ATG或非-ATG密码子;所述填充片段包含编码不完全的M蛋白的核苷酸序列以及一种Kozak序列,所述不完全的M蛋白的核苷酸序列由核苷酸161-509或161-511组成,其中核苷酸161-163是ATG或非-ATG密码子,核苷酸164-509或164-511即为SEQ ID NO:1的核苷酸164-509或164-511,所述一种Kozak序列为核苷酸序列AGAAA,AGACA,AGGAA或AAGCA。

2. 如权利要求1所述的表达增强子,其中所述填充片段包含349至约500个核苷酸的长度或者其间的任意长度。

3. 如权利要求2所述的表达增强子,其中所述填充片段还包含多克隆位点。

4. 如权利要求1所述的表达增强子,其中所述编码不完全的M蛋白的核苷酸序列为349或351个核苷酸的长度。

5. 如权利要求1所述的表达增强子,其中所述CPMV 5' UTR的核苷酸115为腺嘌呤或鸟嘌呤。

6. 植物表达系统,其包含含有调控区的核酸序列,所述调控区与权利要求1所述的表达增强子以及目的核苷酸序列可操作地连接。

7. 如权利要求6所述的植物表达系统,其还包含豇豆花叶病毒组3' UTR。

8. 如权利要求7所述的植物表达系统,其还包含第二核酸序列,所述第二核酸序列编码沉默抑制子。

9. 如权利要求8所述的植物表达系统,其中所述沉默抑制子选自HcPro和p19。

10. 如权利要求6所述的植物表达系统,其中所述调控区选自:质体蓝素启动子、CaMV 35S启动子、2x CaMV35S启动子、CAS启动子、RbcS启动子、Ubi启动子或者肌动蛋白启动子。

11. 如权利要求6所述的植物表达系统,其中所述目的核苷酸序列编码病毒蛋白或抗体。

12. 如权利要求11所述的植物表达系统,其中所述病毒蛋白为选自下述的流感病毒血凝素:H1、H2、H3、H4、H5、H6、H7、H8、H9、H10、H11、H12、H13、H14、H15、H16以及B型流感病毒血凝素。

13. 如权利要求12所述的植物表达系统,其中编码病毒蛋白或抗体的核苷酸序列包含天然的信号肽序列或者非天然的信号肽。

14. 如权利要求13所述的植物表达系统,其中所述非天然的信号肽来自蛋白质二硫键异构酶(PDI)。

15. 在植物或者植物部分中产生目的蛋白的方法,所述方法包括:向所述植物或所述植物部分引入权利要求6所述的植物表达系统,以及在允许编码所述目的蛋白的核苷酸序列表达的条件下培育所述植物或所述植物部分。

16. 用于产生表达目的核苷酸序列的植物或植物部分的方法,其包括用权利要求6所述的植物表达系统瞬时转染或者稳定转化植物或植物部分。

17. 核酸,其包含与目的核苷酸序列可操作地连接的权利要求1所述的表达增强子。

18. 如权利要求17所述的核酸,其中所述目的核苷酸序列为选自下述的流感病毒血凝素(HA):B HA、C、H1、H2、H3、H4、H5、H6、H7、H8、H9、H10、H11、H12、H13、H14、H15以及H16。

19. 如权利要求18所述的核酸,其中所述HA为嵌合的HA,其中所述HA的天然跨膜结构域被异源跨膜结构域替代。

20. 如权利要求19所述的核酸,其中所述异源跨膜结构域获自H1California。

21. 如权利要求1所述的表达增强子,其中所述编码不完全的M蛋白的核苷酸序列由以下组成:SEQ ID NO:1-4中任一个的核苷酸161-509、SEQ ID NO:55的核苷酸161-511或SEQ ID NO:56的核苷酸161-511。

修饰的CPMV增强子元件

发明领域

[0001] 本发明涉及目的蛋白在植物中的表达。本发明还提供用于在植物中产生目的蛋白的方法和组合物。

[0002] 发明背景

[0003] 植物作为重组蛋白的产生系统具有巨大的潜力。在植物中产生外源蛋白的一种方法是产生稳定的转基因植物系。然而,该方法耗时且劳动密集。转基因植物的替代方案是使用基于植物病毒的表达载体。基于植物病毒的载体允许在植物中快速、高水平、瞬时表达蛋白。

[0004] 在植物中实现外源蛋白的高水平瞬时表达的一种方法涉及使用基于RNA植物病毒的载体,包括豇豆花叶病毒组,如豇豆花叶病毒(Cowpea mosaic virus) (CPMV;参见,例如W02007/135480;W02009/087391;US 2010/0287670,Sainsbury F.等,2008,Plant Physiology;148:121-1218;Sainsbury F.等,2008,Plant Biotechnology Journal;6:82-92;Sainsbury F.等,2009,Plant Biotechnology Journal;7:682-693;Sainsbury F.等,2009,Methods in Molecular Biology,Recombinant Proteins From Plants,vol.483:25-39)。

[0005] 豇豆花叶病毒组是具有二重基因组的RNA病毒。豇豆花叶病毒组RNA基因组的片段被称为RNA-1和RNA-2。RNA-1编码VPg、复制酶和蛋白酶蛋白。复制酶是病毒所需的,用于病毒基因组的复制。豇豆花叶病毒组豇豆花叶病毒(CPMV)的RNA-2产生105kDa或95kDa的多肽,其被加工为4个功能肽。

[0006] CPMV RNA-2的5'区包含位于115位、161位、512位和524位的起始密码子(AUG)。位于161位和512位的起始密码子处于同一三联体阅读框内。在位于161位起始密码子处的起始导致105K多蛋白的合成,而在位于512位起始密码子处的起始指导95K多蛋白的合成。在CPMV的512位起始密码子处的翻译起始比在161位处的起始更高效,导致产生比105K多蛋白更多的95K多蛋白。位于115位的起始密码子对于病毒复制不是必要的(Wellink等,1993Biochimie.75(8):741-7)。

[0007] RNA-2通过RNA-1编码的复制酶的高效复制需要维持CPMV RNA-2中161位和512位的起始位点之间的框架(Holness等,1989;Virology 172,311-320;van Bokhoven等,1993,Virology 195,377-386;Rohll等,1993 Virology 193,672-679;Wellink等,1993,Biochimie.75(8):741-7)。该需求影响可被插入复制形式的CPMV RNA-2表达载体中512位处起始密码子上游的序列的长度。此外,多接头的使用应当谨慎,因为它们可使开放阅读框(ORF)在这些起始位点之间移动。

[0008] CPMV被用作发展适于在植物中产生异源多肽的载体系统的基础(Liu等,2005Vaccine 23,1788-1792;Sainsbury等,2007Virus Expression Vectors (Hefferon, K.ed),pp.339-555)。这些系统基于RNA-2的修饰,但是不同之处在于使用全长形式还是或者缺失形式。通过与RNA-1共接种(co-inoculation)实现修饰的RNA-2的复制。外源蛋白与RNA-2来源的多蛋白的C端融合。N端多肽的释放由来自口蹄疫病毒的2A催化肽序列的作用

介导(Gopinath等,2000,Virology 267:159-173)。得到的RNA-2分子既能在植物内传播,又能在植物间传播。该策略已被用于在豇豆植物中表达多种重组蛋白,如乙型肝炎核心抗原(HBcAg)和小免疫蛋白(SIP)(Mechtcheriakova等,J.Virol.Methods 131,10-15;2006;Monger等,2006,植物Biotechnol.J.4,623-631;Alamillo等,2006,Biotechnol.J.1,1103-1111)。尽管成功,但是全长病毒载体的使用限制了插入序列的大小,并且在植物之间的移动引发了关于病毒的生物防护的担忧。

[0009] 为了解决生物防护和插入大小的问题,Canizares等人(2006PlantBiotechnol,J 4:183-193)用目的序列替代了RNA-2的大部分编码区,以产生无能形式的CPMV RNA-2(deIRNA-2)。将待表达的序列与RNA-2位于512位的AUG融合,紧接3'非翻译区(UTR)的上游,以产生模拟RNA-2的分子。在RNA-1和沉默抑制基因的存在下,当被引入植物时,此类构建体能够复制,并指导高水平的异源蛋白的合成(Sainsbury等,2008Plant Biotechnol J 6:82-92)。

[0010] CPMV RNA-2载体中位于161位的起始密码子的突变(U162C;HT)增加了由位于512位的起始密码子之后插入的序列所编码的蛋白的表达水平。这允许高水平的外源蛋白的产生,而无需病毒复制,并将其称为CPMV-HT系统(WO2009/087391;Sainsbury和Lomonosoff,2008,Plant Physiol.148,1212-1218)。在pEAQ表达质粒(Sainsbury等,2009,Plant Biotechnology Journal,7,pp 682-693;US 2010/0287670)中,待表达的序列位于5' UTR和3' UTR之间。pEAQ系列中的5' UTR携带U162C(HT)突变。

[0011] 发明概述

[0012] 本发明涉及目的蛋白在植物中的表达。本发明还提供用于在植物中产生目的蛋白的方法和组合物。

[0013] 如本文所述,提供表达增强子,其依次包含CPMV 5' UTR核苷酸序列和填充序列,所述CPMV 5' UTR核苷酸序列包含SEQ ID NO:1的核苷酸1-160或者包含与SEQ ID NO:1具有约80%至100%序列相似性的核苷酸序列,所述填充片段包含编码不完全M蛋白的核苷酸序列、在植物中有活性的一种或多种kozak序列,或者这两种均包含。填充片段可以包含10至约500个核苷酸的长度,或者其间的任意长度。填充片段的不完全的M蛋白可以包含约10至约351个核苷酸的长度,或者其间的任意长度。填充片段还可以包含多克隆位点。多克隆位点包含约0至约100个核苷酸的长度,或者其间的任意长度。

[0014] 还提供如上文所述的表达增强子,其中kozak序列选自如SEQ ID NO:5-17中所示的序列。

[0015] 还提供包含含有调控区的核酸序列的植物表达系统,所述核酸序列与表达增强子可操作地连接,所述表达增强子依次包含CPMV 5' UTR核苷酸序列和填充片段,所述CPMV 5' UTR核苷酸序列包含SEQ ID NO:1的核苷酸1-160或者包含与SEQ ID NO:1具有约80%至100%的序列相似性的核苷酸序列,所述填充片段包含编码不完全的M蛋白的核苷酸序列和一种或多种kozak序列,表达增强子与目的核苷酸序列可操作地连接。植物表达系统还可以包含豇豆花叶病毒组3' UTR。植物表达系统还可以包含编码沉默抑制因子的第二核酸序列,例如HcPro或p19。

[0016] 如上文所限定的植物表达系统的目的核苷酸序列可以编码病毒蛋白或抗体。例如,病毒蛋白可以选自:H1、H2、H3、H4、H5、H6、H7、H8、H9、H10、H11、H12、H13、H14、H15、H16、B

型流感病毒血凝素。编码病毒蛋白或抗体的核苷酸序列可以包含天然的信号肽序列或者非天然的信号肽,例如所述非天然的信号肽可以获自蛋白质二硫键异构酶(PDI)。

[0017] 如本文所述,提供在植物中或者在植物的一部分中产生目的蛋白的方法,其包括向植物或植物的一部分引入上文所限定的植物表达系统,以及在允许编码目的蛋白的核苷酸序列表达的条件下培养所述植物或植物部分。

[0018] 本发明还提供瞬时转染或者稳定转化有如上文所述的植物表达系统的植物或植物部分。

[0019] 优选地,基于植物的表达系统具有许多特性,如例如含有目的基因的便利克隆位点,可以以成本有效的方式容易地感染植物,可以引起接种植物的高效局部感染或整体感染。此外,感染应当提供有用蛋白材料的良好收率。

[0020] 该发明概述不一定描述本发明的全部特征。

[0021] 附图简述

[0022] 根据以下描述,参考附图,本发明的这些以及其它特征将变得更加显而易见,其中:

[0023] 图1A显示了与目的核苷酸序列融合的现有技术的增强子序列实例(CPMV HT)和如本文所述的增强子序列实例(CPMV HT+)的一般示意图。显示于该图中的元件并非都是增强子序列中所需的。在目的核苷酸序列的3'末端可以包含其它元件(未显示),所述元件包括编码豇豆花叶病毒组3'非翻译区(UTR)的序列、质体蓝素3'UTR的序列或者豇豆花叶病毒组3'UTR与质体蓝素3'UTR的组合的序列。图1B显示了包含如现有技术所述的增强子序列(CPMV HT)以及如本发明所述的增强子序列的构建体,所述增强子序列在其5'末端与植物调控区(在这些非限制性实例中为2X35S)可操作地连接,以及在其3'末端与填充片段、目的核苷酸序列、含有ATG起始位点的“GOI”可操作地连接。在这些实例中,CPMV HT的填充片段包含不完全的M蛋白和多克隆位点,以及在CPMV HT+的实例中,填充片段包含不完全的M蛋白、多克隆位点和植物kozak序列。

[0024] 图2显示了包含与目的核苷酸序列可操作地连接的CPMV-HT(现有技术)表达构建体,和基于CPMV HT+的表达构建体的植物中产生的蛋白的粗蛋白提取物中的相对血凝滴度(HMG)。显示了来自下述的HA的表达数据:具有PDI信号肽的H1A/California/07/2009(构建体编号484:CPMV HT;和构建体编号1805:CPMV HT+;参见实施例4);具有PDI信号肽的H3A/Victoria/361/2011(构建体编号1391:CPMV HT;和构建体编号1819:CPMV HT+;分别参见实施例1和2);具有缺失的蛋白水解环并具有PDI信号肽的B Brisbane/60/08(构建体编号1039:CPMV HT;和构建体编号1829:CPMV HT+;参见实施例8);具有缺失的蛋白水解环,跨膜结构域和胞质尾区被H1A/California/07/2009的跨膜结构域和胞质尾区替代,并具有PDI信号肽的B Brisbane/60/08+H1TM(构建体编号1067:CPMV HT;和构建体编号1875:CPMV HT+;参见实施例9);具有缺失的蛋白水解环并具有PDI信号肽的B Massachusetts/2/2012(构建体编号2072:CPMV HT;和构建体编号2052:CPMV HT+;参见实施例10);具有缺失的蛋白水解环,跨膜结构域和胞质尾区被H1A/California/07/2009的跨膜结构域和胞质尾区替代,并具有PDI信号肽的B Massachusetts/2/2012+H1Tm(构建体编号2074:CPMV HT;和构建体编号2062:CPMV HT+;参见实施例11);具有缺失的蛋白水解环并具有天然信号肽的B Wisconsin/1/2010(构建体编号1445:CPMV HT;和构建体编号1839:CPMV HT+;参见实施例

12);以及具有缺失的蛋白水解环,跨膜结构域和胞质尾区被H1A/California/07/2009的跨膜结构域和胞质尾区替代,并具有天然的信号肽的B Wisconsin/1/2010+H1Tm(构建体编号1454:CPMV HT;和构建体编号1860:CPMV HT+;参见实施例13)。

[0025] 图3显示了在包含CPMV-HT(现有技术)表达构建体和基于CPMV HT+的表达构建体的植物中产生的蛋白的粗蛋白提取物中的相对血凝滴度(HMG)。显示了下述的表达数据:具有PDI信号肽的来自流感病毒A/Indonesia/5/2005的H5(H5 Indo;构建体编号409:CPMV HT;和构建体编号2319:CPMV HT+;参见实施例5);具有苜蓿蛋白质二硫键异构酶的来自流感病毒A/Hangzhou/1/2013的H7(H7 Han;构建体编号2140:CPMV HT;和构建体编号2142:CPMV HT+;参见实施例6);与来自流感病毒A/Indonesia/5/2005的H5的跨膜结构域和胞质尾区(TMCT)融合,并具有苜蓿蛋白质二硫键异构酶的信号肽的来自流感病毒A/Hangzhou/1/2013的H7(H7Han+H5Tm;构建体编号2130:CPMV HT;构建体编号2146:CPMV HT+;参见实施例7)。

[0026] 图4A显示了测试的植物Kozak序列的变体的实例。显示了CPMV HT+、植物调控区、5' UTR、填充片段以及目的核苷酸序列(GOI)的部分序列的构建体。在该非限制性实例中,该构建体包含2X35S调控区、含有填充片段的CPMV HT+,所述填充片段包含不完全的M蛋白和多克隆位点(限制性位点)、植物kozak序列(也显示了目的核苷酸序列的5'末端:“ATG...GOI”;其中GOI为H3A/Victoria/361)。在所述序列下面也显示了植物kozak序列的变体(也参见图7)。各变体植物Kozak序列位于填充片段的3'末端与目的核苷酸序列(在这些非限制性实例中为H3A/Victoria/361)的5'-ATG位点之间。构建体的其它元件保持相同。图4B显示了含有CPMV HT+表达构建体和所显示的变体植物Kozak序列的植物中产生的目的核苷酸序列的HA滴度。

[0027] 图5显示了用于制备构建体编号1391的序列组件(A-2X35S CPMV-HT PDISP H3Victoria NOS;参见实施例1)。构建体编号1391掺入了现有技术CPMV-HT序列(在161位具有突变的起始密码子,与编码不完全的M蛋白的序列融合的CPMV 5' UTR,并且其在5' UTR与目的核苷酸序列之间不包含异源kozak序列(PDISP/H3 Victoria))。PDISP:蛋白质二硫键异构酶信号肽。NOS:胭脂碱合成酶终止子。图5A显示了引物序列IF-PDI.S1=3c(SEQ ID NO:18)。图5B显示了引物序列IF-H3V36111.s1-4r(SEQ ID NO:19)。图5C显示了PDISP/H3 Victoria的序列(SEQ ID NO:20)。图5D显示了构建体1191的示意图。图5E显示了构建体1191;从左至右t-DNA边界(加下划线),具有质体蓝素-P19-质体蓝素沉默抑制子表达盒的2X35S CPMV-HT NOS(SEQ ID NO:21)。图5F显示了从2X35S启动子至NOS终止子的表达盒编号1391。PDISP/H3 Victoria核苷酸序列加下划线;CPMV 5' UTR加粗;不完全的M蛋白为斜体(SEQ ID NO:22)。图5G显示了PDISP/H3 Victoria的氨基酸序列(SEQ ID NO:23)。图5H显示了构建体编号1391(参照构建体)的示意图。

[0028] 图6显示了用于制备构建体编号1819的序列组件(2X35S CPMV-HT+PDISP H3Victoria NOS;参见实施例1)。构建体编号1819掺入了CPMV-HT+序列(在161位具有突变的起始密码子,与编码不完全的M蛋白、多克隆位点的填充片段融合的CPMV 5' UTR,并且在所述多克隆位点与目的核苷酸序列之间包含植物kozak序列(PDISP/H3 Victoria))。PDISP:蛋白质二硫键异构酶信号肽。NOS:胭脂碱合成酶终止子。图6A显示了引物序列IF(SacII)-Kozac_PDI.c(SEQ ID NO:24)。图6B显示了构建体2181的示意图。图6C显示了构建

体2181的序列(从左至右t-DNA边界,加下划线;具有质体蓝素-P19-质体蓝素沉默抑制子表达盒的2X35S/CPMV-HT+/NOS;SEQ ID NO:25)。图6D显示了从2X35S启动子至NOS终止子的表达盒编号1819。PDISP/H3 Victoria核苷酸序列加下划线(SEQ ID NO:26)。图6E显示了构建体1819的示意图。

[0029] 图7显示了包含用于制备选择的基于“CPMV HT+”的构建体的植物kozak序列变型的序列(构建体编号1952至1959)。显示了2X35S/CPMV HT+/NOS表达系统中的SacII限制性位点与PDISP/H3 Victoria的ATG之间的序列变型,其包含植物kozak序列的变型(序列显示为来自构建体1819的对应序列的变型;参见实施例2)。变体植物kozak序列加下划线。PDISP:蛋白质二硫键异构酶信号肽。图7A显示了IF-HT1*-PDI.c的核苷酸序列(SEQ ID NO:27;用于制备构建体编号1952)。图7B显示了IF-HT2*-PDI.c的核苷酸序列(SEQ ID NO:28;用于制备构建体编号1953)。图7C显示了IF-HT3*-PDI.c的核苷酸序列(SEQ ID NO:29;用于制备构建体编号1954)。图7D显示了IF-HT4*-PDI.c的核苷酸序列(SEQ ID NO:30;用于制备构建体编号1955)。图7E显示了IF-HT5*-PDI.c的核苷酸序列(SEQ ID NO:31;用于制备构建体编号1956)。图7F显示了IF-HT6*-PDI.c的核苷酸序列(SEQ ID NO:32;用于制备构建体编号1957)。图7G显示了IF-HT7*-PDI.c的核苷酸序列(SEQ ID NO:33;用于制备构建体编号1958)。图7H显示了IF-HT8*-PDI.c的核苷酸序列(SEQ ID NO:34;用于制备构建体编号1959)。图7I显示了利用SEQ ID NO:27(图7A)的包含植物kozak序列(Kozak1)的构建体编号1952的示意图。除了各构建体(1953-1959)分别包含如图7B至7H(SEQ ID NO:28至34)中所示的修饰的植物Kozak序列(Kozak1)之外,构建体1953-1959包含与构建体1952相同的特征。

[0030] 图8显示了用于制备构建体编号484和1805的序列组件(分别为2X35S/CPMV HT PDISP/H1 California NOS和2X35S/CPMV HT+PDISP/H1 California NOS;参见实施例4)。构建体编号484掺入了现有技术的CPMV-HT序列(在161位处具有突变的起始密码子,与编码不完全的M蛋白的序列融合的CPMV 5' UTR),并且在5' UTR与目的核苷酸序列(PDISP/H1 California)之间不包含异源kozak序列。构建体编号1805包含含有160个核苷酸的CPMV 5' UTR,含有不完全的M蛋白、多克隆位点和植物kozak序列的填充片段,并且是基于CPMV HT+的构建体的实例。PDISP:蛋白质二硫键异构酶信号肽。NOS:胭脂碱合成酶终止子。图8A显示了PDISP/H1 California的核苷酸序列(SEQ ID NO:35)。图8B显示了PDISP/H1 California的氨基酸序列(SEQ ID NO:36)。图8C显示了构建体编号484(2X35S/CPMV HT;参照构建体)的示意图。图8D显示了构建体编号1805(2X35S/CPMV HT+)的示意图。

[0031] 图9显示了用于制备构建体编号409和2319的序列组件(分别为2X35S/CPMV HT PDISP/H5 Indonesia NOS;CPMV HT+PDISP/H5 Indonesia NOS;参见实施例5)。构建体编号409掺入了现有技术的CPMV-HT序列(在161位处具有突变的起始密码子,与编码的不完全的M蛋白序列融合的CPMV 5' UTR),并且在5' UTR与目的核苷酸序列(PDISP/H5 Indonesia)之间不包含异源kozak序列。构建体编号2319包含含有160个核苷酸的CPMV 5' UTR,含有不完全的M蛋白、多克隆位点和植物kozak序列的填充片段,并且是基于CPMV HT+的构建体的实例。NOS:胭脂碱合成酶终止子。图9A显示了PDISP/H5 Indonesia的核苷酸序列(SEQ ID NO:37)。图9B显示了PDISP/H5 Indonesia的氨基酸序列(SEQ ID NO:38)。图9C显示了构建体编号409(2X35S/CPMV HT,参照构建体)的示意图。图9D显示了构建体编号2319(2X35S/CPMV

HT+) 的示意图。

[0032] 图10显示了用于制备构建体编号2140和2142的序列组件(分别为2X35S/CPMV HT H7 Hangzhou NOS;CPMV HT+H7 Hangzhou NOS;参见实施例6)。构建体编号2140掺入了现有技术的CPMV-HT序列(在161位处具有突变的起始密码子,与编码不完全的M蛋白的序列融合的CPMV 5' UTR),并且在5' UTR与目的核苷酸序列(PDISP/H7 Hangzhou)之间不包含异源kozak序列。构建体编号2142包含含有160个核苷酸的CPMV 5' UTR,含有不完全的M蛋白、多克隆位点和植物kozak序列的填充片段,并且是基于CPMV HT+的构建体的实例。NOS:胭脂碱合成酶终止子。图10A显示了天然的H7 Hangzhou的核苷酸序列(SEQ ID NO:39)。图10B显示了天然的H7 Hangzhou的氨基酸序列(SEQ ID NO:40)。图10C显示了构建体编号2140(2X35S/CPMV HT;参照构建体)的示意图。图10D显示了构建体编号2142(2X35S/CPMV HT+)的示意图。

[0033] 图11显示了用于制备构建体编号2130和2146的序列组件(分别为2X35S/CPMV HT H7 Hangzhou+H5 Indonesia TMCT NOS;CPMV HT+H7 Hangzhou+H5 Indonesia TMCT NOS;参见实施例7)。构建体编号2130掺入了现有技术的CPMV-HT序列(在161位处具有突变的起始密码子,与编码不完全的M蛋白的序列融合的CPMV 5' UTR),并且在5' UTR与目的核苷酸序列(PDISP/H7 Hangzhou+H5 Indonesia TMCT)之间不包含异源kozak序列。构建体编号2146包含含有160个核苷酸的CPMV 5' UTR,含有不完全的M蛋白、多克隆位点和植物kozak序列的填充片段,并且是基于CPMV HT+的构建体的实例。NOS:胭脂碱合成酶终止子。图11A显示了H7 Hangzhou+H5 Indonesia TMCT的核苷酸序列(SEQ ID NO:41)。图11B显示了H7 Hangzhou+H5 Indonesia TMCT的氨基酸序列(SEQ ID NO:42)。图11C显示了构建体编号2130(2X35S/CPMV HT;参照构建体)的示意图。图11D显示了构建体编号2146(2X35S/CPMV HT+)的示意图。

[0034] 图12显示了用于制备构建体编号1039和1829的序列组件(分别为2X35S/CPMV HT PDISP/HA B Brisbane (PrL-) NOS和2X35S/CPMV HT+PDISP/HA B Brisbane (PrL-) NOS;参见实施例8)。构建体编号1039掺入了现有技术的CPMV-HT序列(在161位处具有突变的起始密码子,与编码不完全的M蛋白的序列融合的CPMV 5' UTR),并且在5' UTR与目的核苷酸序列(PDISP/HA B Brisbane (PrL-))之间不包含异源kozak序列。构建体编号1829包含含有160个核苷酸的CPMV 5' UTR,含有不完全的M蛋白、多克隆位点和植物kozak序列的填充片段,并且是基于CPMV HT+的构建体的实例。PDISP:蛋白质二硫键异构酶信号肽;NOS:胭脂碱合成酶终止子;PrL-:缺失的蛋白水解环。图12A显示了PDISP/HA B Brisbane (PrL-)的核苷酸序列(SEQ ID NO:43)。图12B显示了PDISP/HA B Brisbane (PrL-)的氨基酸序列(SEQ ID NO:44)。图12C显示了构建体编号1039(2X35S/CPMV HT;参照构建体)的示意图。图12D显示了构建体编号1829(2X35S/CPMV HT+)示意图。

[0035] 图13显示了用于制备构建体编号1067和1875的序列组件(分别为2X35S/CPMV HT PDISP/HA B Brisbane (PrL-)+H1 California TMCT NOS和2X35S/CPMV HT+PDISP/HA B Brisbane (PrL-)+H1 California TMCT NOS;参见实施例9)。构建体编号1067掺入了现有技术的CPMV-HT序列(在161位处具有突变的起始密码子,与编码不完全的M蛋白的序列融合的CPMV 5' UTR),并且在5' UTR与目的核苷酸序列(PDISP/HA B Brisbane (PrL-)+H1 California TMCT)之间不包含异源kozak序列。构建体编号1875包含含有160个核苷酸的

CPMV 5' UTR, 含有不完全的M蛋白、多克隆位点和植物kozak序列的填充片段, 并且是基于 CPMV HT+的构建体的实例。PDISP: 蛋白质二硫键异构酶信号肽; NOS: 胭脂碱合成酶终止子; PrL-: 缺失的蛋白水解环; TMCT: 跨膜结构域胞质尾区。图13A显示了PDISP/HA B Brisbane (PrL-)+H1 California TMCT的核苷酸序列 (SEQ ID NO:45)。图13B显示了PDISP/HA B Brisbane (PrL-)+H1 California TMCT的氨基酸序列 (SEQ ID NO:46)。图13C显示了构建体编号1067 (2X35S/CPMV HT; 参照构建体) 的示意图。图13D显示了构建体编号1875 (2X35S/CPMV160+) 的示意图。

[0036] 图14显示了用于制备构建体编号2072和2052的序列组件 (分别为2X35S/CPMV HT PDISP/HA B Massachusetts (PrL-) NOS和2X35S/CPMV HT+PDISP/HA B Massachusetts (PrL-) NOS; 参见实施例10)。构建体编号2072掺入了现有技术的CPMV-HT序列 (在161位处具有突变的起始密码子, 与编码不完全的M蛋白的序列融合的CPMV 5' UTR), 并且在5' UTR与目的核苷酸序列 (PDISP/HA B Massachusetts (PrL-)) 之间不包含异源kozak序列。构建体编号2052包含含有160个核苷酸的CPMV 5' UTR, 含有不完全的M蛋白、多克隆位点和植物kozak序列的填充片段, 并且是基于CPMV HT+的构建体的实例。PDISP: 蛋白质二硫键异构酶信号肽; NOS: 胭脂碱合成酶终止子; PrL-: 缺失的蛋白水解环。图14A显示了PDISP/HA B Massachusetts (PrL-) 的核苷酸序列 (SEQ ID NO:47)。图14B显示了PDISP/HA B Massachusetts (PrL-) 的氨基酸序列 (SEQ ID NO:48)。图14C显示了构建体编号2072 (2X35S/CPMV HT; 参照构建体) 的示意图。图14D显示了构建体编号2052 (2X35S/CPMV HT+) 的示意图。

[0037] 图15显示了用于制备构建体编号2074和2062的序列组件 (分别为2X35S/CPMV HT PDISP/HA B Massachusetts (PrL-)+H1 California TMCT NOS和2X35S/CPMV HT+PDISP/HA B Massachusetts (PrL-)+H1 California TMCT NOS; 参见实施例11)。构建体编号2074掺入了现有技术的CPMV-HT序列 (在161位处具有突变的起始密码子, 与编码不完全的M蛋白的序列融合的CPMV 5' UTR), 并且在5' UTR与目的核苷酸序列 (PDISP/HA B Massachusetts (PrL-)+H1 California TMCT) 之间不包含异源kozak序列。构建体编号2062包含含有160个核苷酸的CPMV 5' UTR, 含有不完全的M蛋白、多克隆位点和植物kozak序列的填充片段, 并且是基于CPMV HT+的构建体的实例。PDISP: 蛋白质二硫键异构酶信号肽; NOS: 胭脂碱合成酶终止子; PrL-: 缺失的蛋白水解环; TMCT: 跨膜结构域胞质尾区。图15A显示了PDISP/HA B Massachusetts (PrL-)+H1 California TMCT的核苷酸序列 (SEQ ID NO:49)。图15B显示了PDISP/HA B Massachusetts (PrL-)+H1 California TMCT的氨基酸序列 (SEQ ID NO:50)。图15C显示了构建体编号2074 (2X35S/CPMV HT; 参照构建体) 的示意图。图15D显示了构建体编号2062 (2X35S/CPMV HT+) 的示意图。

[0038] 图16显示了用于制备构建体编号1445和1839的序列组件 (分别为2X35S/CPMV HT HA B Wisconsin (PrL-) NOS和2X35S/CPMV HT+HA B Wisconsin (PrL-) NOS; 参见实施例12)。构建体编号1445掺入了现有技术的CPMV-HT序列 (在161位处具有突变的起始密码子, 与编码不完全的M蛋白的序列融合的CPMV 5' UTR), 并且在5' UTR与目的核苷酸序列 (HA B Wisconsin (PrL-)) 之间不包含异源kozak序列。构建体编号1839包含含有160个核苷酸的CPMV 5' UTR, 含有不完全的M蛋白、多克隆位点和植物kozak序列的填充片段, 并且是基于 CPMV HT+的构建体的实例。PrL-: 缺失的蛋白水解环; NOS: 胭脂碱合成酶终止子。图16A显示

了HA B Wisconsin(PrL-)的核苷酸序列(SEQ ID NO:51)。图16B显示了HA B Wisconsin(PrL-)的氨基酸序列(SEQ ID NO:52)。图16C显示了构建体编号1445(2X35S/CPMV HT;参照构建体)的示意图。图16D显示了构建体编号1839(2X35S/CPMV HT+)的示意图。

[0039] 图17显示了用于制备构建体编号1454和1860的序列组件(分别为2X35S/CPMV HT HA B Wisconsin(PrL-)+H1 California TMCT NOS和2X35S/CPMV HT+HA B Wisconsin(PrL-)+H1 California TMCT NOS;参见实施例13)。构建体编号1454掺入了现有技术的CPMV-HT序列(在161位处具有突变的起始密码子,与编码不完全的M蛋白的序列融合的CPMV 5' UTR),并且在5' UTR与目的核苷酸序列(HA B Wisconsin(PrL-)+H1 California TMCT)之间不包含异源kozak序列。构建体编号1860包含含有160个核苷酸的CPMV 5' UTR,含有不完全的M蛋白、多克隆位点和植物kozak序列的填充片段,并且是基于CPMV HT+的构建体的实例。NOS:胭脂碱合成酶终止子;PrL-:缺失的蛋白水解环;TMCT:跨膜结构域胞质尾区。图17A显示了HA B Wisconsin(PrL-)+H1 California TMCT的核苷酸序列(SEQ ID NO:53)。图17B显示了PDISP/HA B Wisconsin(PrL-)+H1 California TMCT的氨基酸序列(SEQ ID NO:54)。图17C显示了构建体编号1454(2X35S/CPMV HT;参照构建体)的示意图。图17D显示了构建体编号1893(2X35S/CPMV HT+)的示意图。

[0040] 详细描述

[0041] 本方面涉及目的蛋白在植物中的表达。本发明还提供用于在植物中产生目的蛋白的方法和组合物。

[0042] 在下述描述中,广泛使用了许多术语,提供以下定义以便于理解本发明的各个方面。在说明书中使用实例,包括术语的实例仅出于示例性目的,并且不意图限制本发明的实施方案的范围和意义。

[0043] 如本文所使用的,当与术语“包含”连同使用时,单词“一个/一种(a)”或“一个/一种(an)”的使用可以意为“一个/一种(one)”,但是其也符合“一个/一种或多个/多种(one or more)”、“至少一个/一种(at least one)”和“一个/一种或多于一个/一种(one or more than one)”的意思。术语“约”指给定值的大约 $\pm 10\%$ 的偏差。术语“多个/多种(plurality)”意为多于一个/一种,例如两个/两种或者更多个/更多种、三个/三种或者更多个/更多种、四个/四种或者更多个/更多种等。

[0044] 本发明提供核酸、表达增强子和表达系统,以在植物中产生目的蛋白,包括豇豆花叶病毒组5'非翻译区(UTR)以及修饰的、延长的或截短的填充序列。还提供包含目的序列的植物细胞、植物组织、全植株、接种物、核酸、构建体,以及在植物中表达目的蛋白的方法。

[0045] 本发明的表达增强子的实例包括但不限于如图1A中所示的CPMV HT+。提供了CPMV HT+的核苷酸序列的非限制性实例,参考SEQ ID NO:1、3和4。

[0046] 表达增强子CPMV HT+可以在增强子序列的5'末端与在植物中有活性的调控区可操作地连接,以及在表达增强子的3'末端与目的核苷酸序列可操作地连接(图1A),以驱使目的核苷酸序列在植物宿主内的表达。

[0047] 还提供使用CPMV HT+,在植物中产生一种或多种目的蛋白的表达系统。本文所述的表达系统包含含有CPMV HT+或者与CPMV HT+具有80%至100%或者其间任意量的序列相似性的序列的表达盒。包含CPMV HT+的表达盒还可以包含与表达增强子的5'末端可操作地连接并在植物中有活性的调控区。目的核苷酸序列可以与表达盒的3'末端可操作地连接,

以当被引入植物中时实现目的核苷酸序列在植物宿主中的表达。

[0048] 参考图1A和1B,显示了包含来自豇豆花叶病毒组基因组(豇豆花叶病毒组5' UTR)的5' UTR序列以及修饰的M蛋白的表达增强子。在图1A和1B中,显示了包含不完全的M蛋白的现有技术的CPMV HT增强子(Sainsbury和Lomonosoff,2008,Plant Physiology;148:1212-1218;WO 2009/087391;将它们通过引用并入本文),以及本发明的CPMV HT+。

[0049] 现有技术的CPMV HT表达增强子(Sainsbury和Lomonosoff,2008,Plant Physiology;148:1212-1218;WO 2009/087391;将它们通过引用并入本文)包含115位处具有修饰的ATG(ATG修饰为CGT)的CPMV 5' UTR的核苷酸1-160,其与包含SEQ ID NO:2的核苷酸161-509的不完全的M蛋白融合,所述不完全的M蛋白由位于162位的修饰的ATG(ATG修饰为ACG)和接头序列组成(参见SEQ ID NO:2)。目的核苷酸与接头序列的3'末端融合(参见图1A和1B)。现有技术的CPMV HT缺乏紧接目的核苷酸与表达增强子融合位置上游的植物Kozak序列。

[0050] 本发明的CPMV HT+可以包含115-117位处具有修饰的ATG的CPMV 5' UTR的核苷酸1-160(其中ATG修饰为GTG;位于115位的A修饰为G),其与修饰的填充序列融合,所述填充序列包含:修饰的ATG(位于161-163位的ATG修饰为ACG;其中位于162位的“T”修饰为“C”)、不完全的M蛋白、接头序列和植物kozak序列。CPMV HT+核苷酸序列的非限制性实例显示于SEQ ID NO:1和4中。目的核苷酸可以与kozak序列的3'末端融合,例如CPMV HT+的植物kozak序列(参见图1A和1B)。

[0051] 如SEQ ID NO:4所示,本发明的表达增强子的5' UTR也可以包含位于115位的“A”,并包含核苷酸115-117处的天然ATG序列,其与修饰的填充序列融合,所述填充序列包含:位于核苷酸161-163处修饰为ACG(其中位于162位的“T”修饰为“C”)、不完全的M蛋白、接头序列和植物kozak序列。包含位于115位天然的或野生型“A”的表达增强子可被称作CPMV HT+[WT115]。CPMV HT+[WT115]核苷酸序列的非限制性实例显示于SEQ ID NO:3中。目的核苷酸可以与CPMV HT+[WT115]的植物kozak序列的3'末端融合(参见图1A和1B)。CPMV HT+[WT115]可被认为是CPMV HT+的变体,因为CPMV HT+在115位包含核苷酸“G”(SEQ ID NO:1)。而且,CPMV HT+[511]也可被认为是CPMV HT+的变体,因为CPMV HT+[511]包含不完全M蛋白的核苷酸161-511(参见SEQ ID NO:56)。如此,术语“CPMV HT+”是包括诸如CPMV HT+[WT115]和CPMV HT+[511]的变体的通用术语。

[0052] CPMV HT+表达增强子的非限制性实例呈现于SEQ ID NO:1、3和4中,然而,应当理解,可以在不脱离本发明的情况下对CPMV HT+和CPMV HT+[WT115]的填充序列进行变型或修饰,条件是填充片段包含不完全的M蛋白,并且当与目的核苷酸序列融合时,在植物中有活性的kozak序列(例如植物kozak序列)位于目的核苷酸序列的5'末端。SEQ ID NO:56中提供了CPMV HT+表达增强子的另一实例(包含M蛋白-天然kozak序列)(CPMV HT+[511])。

[0053] 可以通过插入、截短或缺失来修饰填充序列,以使当与现有技术的增强子CPMV HT(Sainsbury和Lomonosoff,2008,Plant Physiology;148:1212-1218;WO 2009/087391;将它们通过引用并入本文)的最初的(未改变的)不完全的M蛋白序列相比,不完全的M蛋白序列更长、截短或变短。除了不完全的M蛋白之外,本发明的CPMV HT+表达增强子的填充序列还可以包含一个或多个限制性位点(多接头、多克隆位点、一个或多个克隆位点)、一种或多种kozak序列或M蛋白-天然kozak序列、一种或多种接头序列、一个或多个重组位点或者它

们的组合。例如(不应被认为是限制性的),填充序列可以依次包含不完全的M蛋白(SEQ ID NO:1或2的核苷酸161-509,或者SEQ ID NO:56的核苷酸161-511)、与在植物中有活性的kozak序列(例如植物kozak序列)融合的期望长度的多克隆位点。

[0054] 可以利用多种方法将目的核苷酸序列与本发明的CPMV HT+(或者CPMV HT+[WT115]或CPMV HT+[511])增强子序列融合(可操作地连接)。例如(不应被认为是限制性的):

[0055] 1) 目的核苷酸序列(例如编码目的蛋白的核酸)可以紧接在由不完全的M蛋白序列组成的填充片段(含有或不含多克隆位点)之后与CPMV HT+表达增强子融合。在此种情况下,目的核苷酸序列与不完全M蛋白序列(或多克隆位点(如果存在的话))的3'末端融合,并且所述目的核苷酸序列在其5'末端包含紧接目的核苷酸序列的ATG起始位点上游的植物kozak序列。

[0056] 2) 目的核苷酸序列可以紧接位于填充片段(包含不完全的M蛋白、任选的多克隆位点和植物kozak序列)的3'末端的植物kozak序列之后,与包含填充片段的表达增强子融合。在此种情况下,目的核苷酸序列将不包含相应的多克隆位点或植物kozak序列。

[0057] 3) 目的核苷酸序列可以利用多克隆位点与CPMV HT+表达增强子融合。在此种情况下,目的核苷酸序列可以在其5'末端包含相应的克隆位点,以允许紧接目的核苷酸序列的ATG起始位点上游与表达增强子和植物kozak序列融合。

[0058] 利用任何以上方法得到的整体结果是构建体或表达盒,其包含与5' UTR序列可操作地结合(可操作地连接)的植物调控区,所述5' UTR序列与包含同植物kozak序列融合的不完全M蛋白的填充片段的5'末端融合,所述植物kozak序列的3'末端与目的核苷酸序列的5'末端融合。构建体还可以包含含有豇豆花叶病毒组3'非翻译区(UTR),例如质体蓝素3' UTR或在植物中有活性的其它3' UTR的序列,以及终止子序列(例如NOS终止子)的序列,所述终止子序列与目的核苷酸序列的3'末端可操作地连接(参见图1A)。

[0059] 还提供植物表达系统,其包含含有调控区的第一核酸序列,所述第一核酸序列与本文所述的一种或多种表达增强子(例如CPMV HT+、CPMV HT+[WT115]、CPMV HT+[511])以及目的核苷酸序列可操作地连接。此外,描述了包含启动子(调控区)序列、含有豇豆花叶病毒组5' UTR和填充序列的表达增强子(例如CPMV HT+、CPMV HT+[WT115]或CPMV HT+[511])的核酸,所述填充序列具有与编码一种或多种目的蛋白的核酸序列融合的植物kozak序列。所述核酸还可以包含含有豇豆花叶病毒组3'非翻译区(UTR),例如质体蓝素3' UTR或在植物中有活性的其它3' UTR的序列以及终止子序列(例如NOS终止子),所述终止子序列与目的核苷酸序列的3'末端可操作地连接(参见图1A),以使目的核苷酸序列插入豇豆花叶病毒组3'非翻译区(UTR)、质体蓝素3' UTR或其它3' UTR序列的上游。

[0060] “可操作地连接”意为,具体序列直接或者间接相互作用,以实现预期功能,如介导或调控核酸序列的表达。可操作地连接的序列的相互作用可以,例如被与可操作地连接的序列相互作用的蛋白介导。

[0061] 本文所称的“表达增强子”、“增强子序列”或“增强子元件”包括源自SEQ ID NO:1的核苷酸1-160的序列,或者与SEQ ID NO:1的核苷酸1-160的序列共有约80%至100%或者其间任意量的序列相似性的序列。增强子序列能够增强与之连接的下游异源开放阅读框(ORF)的表达。

[0062] 术语“5' UTR”或“5' 非翻译区”或“5' 前导序列”指不翻译的mRNA区。5' UTR通常开始于转录开始位点,并结束于紧接编码区的翻译起始位点或起始密码子(在mRNA中通常为AUG,在DNA序列中为ATG)之前。可以通过下述修饰5' UTR:5' UTR的长度,或者使天然存在的起始密码子或翻译起始位点突变,以使所述密码子不再作为起始密码子发挥作用,并且翻译可以在替代的起始位点处起始。例如可以修饰天然CPMV 5' UTR的115位、161位、512位、524位或其组合处起始的ATG,以移除ATG序列,例如修饰为ACG、GTG或者CGT或者其它非ATG序列。

[0063] “不完全的M蛋白”意为包含SEQ ID NO:1的核苷酸161-509或者SEQ ID NO:56的核苷酸161-511的约10个至约351个核苷酸或者其间任意量的M蛋白编码序列。SEQ ID NO:1的核苷酸161-509或者SEQ ID NO:56的核苷酸161-511对应于存在于天然CPMV RNA 2基因组序列中的M蛋白的序列(参见本文的SEQ ID NO:55;也参见WO 2009/087391的表1;将它们通过引用并入本文),其中修饰(置换)位于161-163位的一个或多个核苷酸以移除“atg”开始位点。不完全的M蛋白序列的非限制性实例包括这样的核苷酸序列:其与SEQ ID NO:1-4的核苷酸161至509、SEQ ID NO:55的核苷酸161-511或者SEQ ID NO:56的核苷酸161-511所限定的序列具有约80%至100%的序列相似性或者其间任意量的序列相似性。可以修饰序列以移除起始密码子或起始位点,以使所述序列不包含“atg”密码子。此外,不完全的M蛋白的非限制性实例包括SEQ ID NO:1-4的核苷酸161至509、SEQ ID NO:55的核苷酸161-511或者SEQ ID NO:56的核苷酸161-511。如本文所述使用的包含对天然CPMV RNA 2序列(SEQ ID NO:55)改变的序列的非限制性实例包括CPMV HT+(SEQ ID NO:1)和CPMV HT+[511](SEQ ID NO:56)。

[0064] SEQ ID NO:55包含核苷酸1-514(天然CMPV RNA-2基因组片段的全长序列存在于WO 2009/087391的表1中;将其通过引用并入本文)的CPMV RNA 2基因组的天然(野生型)序列的片段。来自SEQ ID NO:55的核苷酸1-511的5' UTR序列包含在115位和161位起始的(斜体加粗)野生型“atg”序列以及核苷酸161-511的加下划线的不完全的M蛋白(为了上下文也显示了位于512位的“atg”;在CPMV RNA-2的天然5' 区的512位“atg”处的起始导致95K多蛋白的合成):

[0065] 1tattaaaatc ttaatagggtt ttgataaaag cgaacgtggg gaaacccgaa ccaaacccttc

[0066] 61ttctaaactc tctctcatct ctcttaaagc aaacttctct cttgtctttc ttgc~~atg~~agc

[0067] 121gatcttcaac gttgtcagat cgtgcttcgg caccagtaca~~atg~~ttttctt tcaactgaagc

[0068] 181gaaatcaaag atctctttgt ggacacgtag tgcggcgcca ttaaataacg tgtacttgct

[0069] 241ctattcttgt cgggtgtggtc ttgggaaaag aaagcttgct ggaggctgct gttcagcccc

[0070] 301atacattact tgttacgatt ctgctgactt tcggcggtg caatatctct acttctgctt

[0071] 361gacgaggtat tggtgcctgt acttctttct tcttcttctt gctgattggt tctataagaa

[0072] 421atctagtatt ttctttgaaa cagagttttc ccgtggtttt cgaacttgga gaaagattgt

[0073] 481taagcttctg tatattctgc ccaaatttga aatg...SEQ ID NO:55

[0074] CPMV RNA-2序列的核苷酸1-160的5' UTR(参见SEQ ID NO:1),起始于转录开始位点,止于框内第一起始密码子(位于161位),其作为用于产生由野生型豇豆花叶病毒基因组片段编码的两种共羧基端蛋白中较长的蛋白的起始位点(两种编码蛋白中较短的蛋白起始于155位)。此外,也可以突变、缺失或者以其它方式修饰位于(或者对应于)CPMV RNA-2基

基因组序列115位处的‘第三’起始位点。已经显示,当与不完全的M蛋白组合时,除了移除位于161位的AUG之外,移除AUG 115增强了目的核苷酸序列的表达(Sainsbury和Lomonosoff, 2008, Plant Physiology; 148:1212-1218; WO 2009/087391; 将它们通过引用并入本文)。

[0075] 表达增强子可以包含SEQ ID NO:1、3或4的核苷酸1-509或者SEQ ID NO:55的核苷酸1-511或者SEQ ID NO:56的核苷酸1-511,并包含与填充片段融合的CPMV 5' 非翻译区(UTR),所述填充片段包含不完全的M蛋白和位于填充片段3'末端的植物kozak序列以及任选地多克隆位点;或者表达增强子可以包含与SEQ ID NO:1的核苷酸1-160、SEQ ID NO:1的核苷酸161-509、SEQ ID NO:1的核苷酸1-509、SEQ ID NO:55的核苷酸1-511或SEQ ID NO:56的核苷酸1-511有80%至100%序列相似性或者其间任意量的序列相似性的序列,并包含含有不完全M蛋白的填充片段,并且在填充片段的3'末端还包含在植物中有活性的kozak序列(植物kozak序列或M蛋白-天然kozak序列)以及任选地多克隆位点,并且当与同现有技术“CPMV HT”增强子序列(SEQ ID NO:2; Sainsbury F., 和Lomonosoff G.P., 2008, Plant Physiol. 148:pp. 1212-1218; 将其通过引用并入本文)可操作地连接的编码异源开放阅读框的核苷酸序列的表达相比时,显示出增强与上述表达增强子可操作地连接的编码异源开放阅读框的同样的核苷酸序列表达的特性。

[0076] 填充序列可以包含不完全M蛋白的约50个至351个核苷酸,或者其间任何量的核苷酸,例如但不限于SEQ ID NO:1的核苷酸161-509或SEQ ID NO:55或56的核苷酸161-511、位于填充片段的3'末端的植物kozak序列或者M蛋白-天然kozak序列,以及位于编码不完全的M蛋白的序列与kozak序列(例如植物kozak序列)之间的长度为约0至100个核苷酸或者其间任意量的核苷酸的多克隆位点(MCS)。因此,填充序列的长度可以为与CPMV 5' UTR序列的3'末端融合的约50个至约500个核苷酸或者其间任何量,例如填充序列的长度可以为与CPMV 5' UTR序列的3'末端融合的50、60、70、80、90、100、120、140、160、180、200、220、240、260、280、300、320、340、360、380、400、420、440、460、480、500个核苷酸,或者其间任何量。

[0077] 表达增强子还可以包含一个或多个“限制性位点”或“限制性识别位点”、“多克隆位点”、“MCS”、“克隆位点”、“多接头序列”或“多接头”,以便于目的核苷酸插入植物表达系统。本领域众所周知,限制性位点是被限制酶识别的特异性序列基序。表达增强子可以包含位于不完全的M蛋白的下游(3')的一个或多个限制性位点或克隆位点。所述一个或多个限制性位点或克隆位点可以位于的上游一种或多种kozak序列的上游(5'),并位于不完全的M蛋白和kozak序列之间。多接头序列(多克隆位点)可以包含可用于向5' UTR的3'末端添加和移除核酸序列(包括编码目的蛋白的核苷酸序列)的任何核酸序列。多接头序列可以包含4至约100个核酸或者其间任何量的核酸。

[0078] 由SEQ ID NO:1、3、4或者56的序列所限定的CPMV HT+表达增强子(包括CPMV HT+[WT115]; CPMV HT+[511])是本发明的表达增强子的实例,其不应被认为是限制性的。本发明还包括这样的表达盒或载体,其包含如SEQ ID NO:1、3、4或56所限定的CPMV HT+、CPMV HT+[WT115]或CPMV HT+[511],并且还包含与表达增强子序列可操作地结合的植物调控区以及与表达增强子的3'末端融合的目的核苷酸序列(GOI; 具有3' UTR和终止子序列)。本发明还包括如SEQ ID NO:1、3、4或56所限定的CPMV HT+表达增强子的变体。

[0079] 如果增强子序列包含CPMV 5' UTR序列、不完全的M蛋白以及位于所述增强子3'末端的在植物中有活性的kozak序列(例如植物kozak序列),则所述增强子序列可被称为

“CPMV HT+增强子序列”，条件是当与同现有技术中的“CPMV HT”增强子序列 (SEQ ID NO:2; Sainsbury F.和Lomonosoff G.P., 2008, Plant Physiol. 148: pp. 1212-1218; 将其通过引用并入本文) 可操作地连接的编码异源开放阅读框的核苷酸序列的表达相比时，所述 CPMV HT+增强子序列显示出增强同所述表达增强子可操作地连接的编码异源开放阅读框的同样的核苷酸序列的表达的特性。

[0080] 可以存在于 CPMV HT+ 中的变化可以包括：

[0081] 1) 不完全的 M 蛋白的长度从约 50 至约 351 个核苷酸或者其间任何量的改变；

[0082] 2) 多克隆位点 (MCS) 的修饰，既包括包含的序列的修饰，又包括 MCS 的长度从约 0 至约 100 个核苷酸或者其间任何长度的改变；以及

[0083] 3) kozak 序列可以是在植物中有活性的任何 kozak 序列，并且包括植物 kozak 序列或 M 蛋白-天然 kozak 序列。

[0084] 例如，CPMV 5' UTR 包含 SEQ ID NO:1 或 3 的核苷酸 1 至 160 或者与 SEQ ID NO:1 或 3 的核苷酸 1-160 具有约 80% 至约 100% 或者其间任何量的序列相似性的序列。

[0085] SEQ ID NO:1 是包含 CPMV HT+ 的表达增强子的实例 (核苷酸 1-160, 5' UTR, 包含位于 115 位的修饰的 ATG (GTG) (小写粗体且斜体)；包含以下的填充片段：不完全的 M 蛋白 (加下划线)，核苷酸 161-509, 162 位处具有修饰的核苷酸 (ACG)；多克隆位点 (斜体)，核苷酸 510-528；以及植物 kozak 序列 (大写并加粗)，核苷酸 529-534)：

[0086] 1tattaaaatc ttaatagggtt ttgataaaag cgaacgtggg gaaacccgaa ccaaacccttc

[0087] 61ttctaaactc tctctcatct ctcttaaagc aaacttctct cttgtctttc ttgc**gtg**agc

[0088] 121gatcttcaac gttgtcagat cgtgcttcgg caccagtaca **acg**tttttctt tcaactgaagc

[0089] 181gaaatcaaag atctctttgt ggacacgtag tgccgcgcca ttaaataaacg tgtacttgct

[0090] 241ctattcttgt cgggtgtggtc ttgggaaaag aaagcttgct ggaggctgct gttcagcccc

[0091] 301atacattact tgttacgatt ctgctgactt tcggcgggtg caatatctct acttctgctt

[0092] 361gacgaggtat tggtgcctgt acttctttct tcttcttctt gctgattggt tctataagaa

[0093] 421atctagtatt ttctttgaaa cagagttttc ccgtgggttt cgaacttgga gaaagattgt

[0094] 481taagcttctg tatattctgc ccaaatttgt tcgggcccaa taccgcgg**AG AAAA**

[0095] (SEQ ID NO:1)

[0096] 构建体 1819、1805、2319、2142、2146、1829、1875、2052、2062、1839 和 1860 (分别参见实施例 1 和 4-13) 是包含含有 SEQ ID NO:1 的 CPMV HT+ 表达增强子的代表性构建体。使用这些构建体的结果可见于图 2 和 3。

[0097] SEQ ID NO:2 包含如现有技术中已知的“CPMV HT”表达增强子 (例如 Sainsbury 和 Lomonosoff 2008, Plant Physiol. 148: pp. 1212-1218 的图 1; 将其通过引用并入本文)。CPMV HT 包含 115 位处具有修饰的核苷酸 (cgt) 的来自 SEQ ID NO:2 的核苷酸 1-160 的 5' UTR 序列以及含有 162 位处具有修饰的核苷酸 (acg) 的不完全的 M 蛋白，并且缺乏植物 kozak 序列 (5' UTR: 核苷酸 1-160; 不完全的 M 蛋白 (加下划线)，核苷酸 161-509)。SEQ ID NO:2 还包含 Sainsbury 和 Lomonosoff 2008 中所述的现有技术的 CPMV HT 序列中不存在的多克隆位点 (斜体，核苷酸 510-517)：

[0098] 1tattaaaatc ttaatagggtt ttgataaaag cgaacgtggg gaaacccgaa ccaaacccttc

[0099] 61ttctaaactc tctctcatct ctcttaaagc aaacttctct cttgtctttc ttgcgtgagc

[0100] 12l gatcttcaac gttgtcagat cgtgcttcgg caccagtaca **acg** tttttctt tcactgaagc
[0101] 18l gaaatcaaag atctctttgt ggacacgtag tgcggcgccca ttaaataacg tgtacttgtc
[0102] 24l ctattcttgt cgggtgtggtc ttgggaaaag aaagcttgct ggaggctgct gttcagcccc
[0103] 30l atacattact tgttacgatt ctgctgactt tcggcgggtg caatatctct acttctgctt
[0104] 36l gacgaggtat tgttgccgtg acttctttct tcttcttctt gctgattggt tctataagaa
[0105] 42l atctagtatt ttctttgaaa cagagttttc ccgtggtttt cgaacttga gaaagattgt
[0106] 48l taagcttctg tatattctgc ccaaatttgt cgggccc SEQ ID NO:2

[0107] 包含CPMV HT的构建体在本文被用作参照构建体,以便可以将目的核苷酸序列的表达水平或者利用CPMV HT+、CPMV HT+[WT115]或CPMV HT+[511]的构建体产生的由所述目的核苷酸序列编码的产物进行比较。构建体1391、484、409、2140、2130、1039、1067、2072、2074、1445和1454(分别参见实施例1和4-13)是参照构建体CPMV HT的代表,并且包含SEQ ID NO:2。利用这些构建体的结果可见图2和3。

[0108] 如本文所述的CPMV HT+的植物kozak序列可以是任何植物kozak序列,包括但不限于SEQ ID NO:5-17的序列之一(也参见图4,具有植物kozak的CPMV HT+;图4的构建体包含SEQ ID NO:1的核苷酸1-528、植物kozak序列的几种实例之一,并包含植物调控区和目的核苷酸序列:GOI的转录起始位点ATG)。

[0109] 表达增强子还可以包含一种或多种“kozak共有序列”或“kozak序列”。kozak序列可以包括在植物中有活性的任何kozak序列,例如,植物kozak序列(例如CPMV HT+;CPMV HT WT115)或存在于天然M蛋白中的kozak序列(例如CPMV HT+511)。Kozak序列在翻译起始中发挥主要作用。可以通过确保以下来优化翻译速率:从转基因构建体除去任何mRNA不稳定的序列,以及根据需要使翻译开始位点或起始位点匹配植物的kozak共有序列(Gutierrez, R.A.等,1999,Trends Plant Sci.4,429-438;Kawaguchi,R.和Bailey-Serres,J.,2002,Curr.Opin.Plant Biol.5,460-465)。该基序中最高度保守的位点是ATG密码子上游的嘌呤(其最通常为A)三核苷酸,其指示翻译的起始(Kozak,M.,1987,J.Mol.Biol.20:947-950,通过引用并入本文)。kozak序列可以包含存在于不完全的M蛋白中的天然序列。例如,如SEQ ID NO:55或56所示的紧接不完全的M蛋白的核苷酸512上游的序列,例如SEQ ID NO:55的核苷酸508-511。

[0110] 植物kozak共有序列在本领域是已知的(参见例如Rangan等,Mol.Biotechnol.,2008,July 39(3),pp.207-213)。天然存在的和合成的Kozak序列均可用于表达增强子或者均可与如本文所述的目的核苷酸序列融合。

[0111] 植物kozak序列可以是任何已知的植物kozak序列(参见例如L.Rangan等,Mol.Biotechnol.2008),包括但不限于下述植物共有序列:

[0112] caA (A/C) a (SEQ ID NO:5;植物界)

[0113] aaA (A/C) a (SEQ ID NO:6;双子叶植物)

[0114] aa (A/G) (A/C) a (SEQ ID NO:7;拟南芥)

[0115] 植物kozak序列还可以选自(参见图4):

- AGAAA (SEQ ID NO: 8)
- AGACA (SEQ ID NO: 9)
- AGGAA (SEQ ID NO: 10)
- AAAAA (SEQ ID NO: 11)
- [0116] AAACA (SEQ ID NO: 12)
- AAGCA (SEQ ID NO: 13)
- AAGAA (SEQ ID NO: 14)
- AAAGAA (SEQ ID NO: 15)
- AAAGAA (SEQ ID NO: 16)
- [0117] (A/-)A(A/G)(A/G)(A/C)A. (SEQ ID NO:17;共有序列)
- [0118] SEQ ID NO:4中提供了具有植物kozak共有序列的CPMV HT+ (核苷酸1-160,5' UTR, 包含115位修饰的ATG (gTG) (小写加粗并斜体);包含以下的填充片段:不完全的M蛋白(加下划线)、162位处具有修饰的核苷酸(ACG)的核苷酸161-509;多克隆位点(斜体),核苷酸510-528;以及共有植物kozak序列(大写并加粗),核苷酸529-534)。
- [0119] 1tattaaaatc ttaatagggtt ttgataaaag cgaacgtggg gaaacccgaa ccaaacccttc
- [0120] 61ttctaaactc tctctcatct ctcttaaagc aaacttctct cttgtctttc ttgc~~**gtg**~~agc
- [0121] 12lgatcttcaac gttgtcagat cgtgcttcgg caccagtaca~~**acg**~~ttttctt tcaactgaagc
- [0122] 18lgaaatcaaag atctctttgt ggacacgtag tgcggcgcca ttaaataacg tgtacttgct
- [0123] 24lctattcttgt cgggtgtggtc ttgggaaaag aaagcttgct ggaggctgct gttcagcccc
- [0124] 30latacattact tgttacgatt ctgctgactt tcggcgggtg caatatctct acttctgctt
- [0125] 36lgacgaggtat tgttgccgtg acttctttct tcttcttctt gctgattggt tctataagaa
- [0126] 42latctagtatt ttctttgaaa cagagttttc ccgtgggttt cgaacttgga gaaagattgt
- [0127] 48ltaagcttctg tatattctgc ccaaatttgt tcgggcccac taccgcgg(A/-)A(A/G)
- [0128] (A/G)(A/C)A (SEQ ID NO:4)
- [0129] SEQ ID NO:56 (“CPMV HT+[511]”)包含核苷酸1-154的CPMV RNA2基因组的天然序列的一段。来自SEQ ID NO:56的核苷酸1-511的5' UTR序列包含115位修饰的“atg”序列(“g”代替“a”;斜体加粗)和162位修饰的“atg”序列(“c”代替“t”;斜体加粗)以及核苷酸161-511的不完全的M蛋白(加下划线)。CPMV HT+[511]包含天然的M蛋白kozak共有序列(核苷酸508-511;加粗):
- [0130] 1tattaaaatc ttaatagggtt ttgataaaag cgaacgtggg gaaacccgaa ccaaacccttc
- [0131] 61ttctaaactc tctctcatct ctcttaaagc aaacttctct cttgtctttc ttgc~~**gtg**~~agc
- [0132] 12lgatcttcaac gttgtcagat cgtgcttcgg caccagtaca~~**acg**~~ttttctt tcaactgaagc
- [0133] 18lgaaatcaaag atctctttgt ggacacgtag tgcggcgcca ttaaataacg tgtacttgct
- [0134] 24lctattcttgt cgggtgtggtc ttgggaaaag aaagcttgct ggaggctgct gttcagcccc
- [0135] 30latacattact tgttacgatt ctgctgactt tcggcgggtg caatatctct acttctgctt

[0136] 361gacgaggtat tggtgcctgt acttctttct tcttcttctt gctgattgggt tctataagaa
[0137] 421atctagtatt ttctttgaaa cagagttttc ccgtgggtttt cgaacttgga gaaagattgt
[0138] 481taagcttctg tatattctgc ccaaatttga a...SEQ ID NO:56

[0139] 构建体1952-1959是包含具有不同的植物kozak序列的SEQ ID NO:1的CPMV HT+表达增强子的实例(参见实施例3,以及图4和7)。

[0140] 通过SEQ ID NO:3的序列(CPMV HT+[WT115])提供CPMV HT+增强子序列的另一非限制性实例。包含CPMV HT+和与SEQ ID NO:3的表达增强子序列可操作结合的植物调控区以及在3'末端与目的核苷酸序列(GOI)融合的转录开始位点(ATG)的表达盒或载体也是本发明的一部分。

[0141] SEQ ID NO:3(CPMV HT+[WT115])包含:核苷酸1-160,5' UTR,115-117位处具有ATG(小写加粗);包含以下的填充片段:不完全的M蛋白(加下划线)、161-163位具有修饰的ATG(acg)(小写加粗并加下划线)的核苷酸161-509的填充片段;多克隆位点(斜体),核苷酸510-528;以及植物kozak序列(大写并加粗),核苷酸529-534。

[0142] 1tattaaaatc ttaataggtt ttgataaaag cgaacgtggg gaaacccgaa ccaaacccttc
[0143] 61ttctaaactc tctctcatct ctcttaaagc aaacttctct cttgtctttc ttgc**atg**agc
[0144] 121gatcttcaac gttgtcagat cgtgcttcgg caccagtaca**acg**ttttctt tcaactgaagc
[0145] 181gaaatcaaag atctctttgt ggacacgtag tgcggcgcca ttaaataacg tgtacttgct
[0146] 241ctattcttgt cgggtgtggtc ttgggaaaag aaagcttgct ggaggctgct gttcagcccc
[0147] 301atacattact tgttacgatt ctgctgactt tcggcggtg caatatctct acttctgctt
[0148] 361gacgaggtat tggtgcctgt acttctttct tcttcttctt gctgattgggt tctataagaa
[0149] 421atctagtatt ttctttgaaa cagagttttc ccgtgggtttt cgaacttgga gaaagattgt
[0150] 481taagcttctg tatattctgc ccaaatttgt tcgggccc aa taccgcgg**AG AAAA**
[0151] (SEQ ID NO:3)

[0152] SEQ ID NO:3的植物kozak序列可以是任何植物kozak序列,包括但不限于SEQ ID NO:5-17的序列之一(也参见图4,具有植物kozak的CPMV HT+;图4的构建体包含SEQ ID NO:1的核苷酸1-528、植物kozak序列的几种实例之一,并包含植物调控区和目的核苷酸序列:GOI的转录起始位点ATG)。

[0153] CPMV HT+表达增强子可以与SEQ ID NO:1、3或4的核苷酸1-160、SEQ ID NO:1、3或4的核苷酸161-509或者SEQ ID NO:1、3或4的核苷酸1-509所限定的序列具有至少100%、99%、98%、97%、96%、95%、90%、85%和80%的同一性。例如增强子序列可以与SEQ ID NO:1、3或4的核苷酸1-160、SEQ ID NO:1、3或4的核苷酸161-509或者SEQ ID NO:1、3或4的核苷酸1-509所限定的序列具有约80%至约100%或者其间任何量的同一性,或者与SEQ ID NO:1、3或4的核苷酸1-160、SEQ ID NO:1、3或4的核苷酸161-509或者SEQ ID NO:1、3或4的核苷酸1-509所限定的序列具有约90%至约100%或者其间任何量的同一性,或者与SEQ ID NO:1、3或4的核苷酸1-160、SEQ ID NO:1、3或4的核苷酸161-509或者SEQ ID NO:1、3或4的核苷酸1-509所限定的序列具有约95%至约100%或者其间任何量的同一性,其中当与如本文所述的植物调控区和植物kozak序列可操作地连接时,与利用同样的植物调控区与CPMV HT(SEQ ID NO:2)融合的目的核苷酸序列的表达水平相比,所述表达增强子增加了与表达

增强子可操作地连接的目的核苷酸序列的表达水平。

[0154] 当提及具体序列时,例如如威斯康星大学GCG软件程序所述的或者通过人工比对和目测,使用术语“相似性百分比”或“同一性百分比”(参见,例如Current Protocols in Molecular Biology,Ausubel等,eds.1995supplement)。用于将序列对齐进行比较的方法在本领域众所周知。可以如下所述进行比较序列的最佳对齐:利用例如Smith&Waterman的算法(1981,Adv.Appl.Math.2:482),通过Needleman&Wunsch的比对算法(1970,J.Mol.Biol.48:443),通过Pearson&Lipman的搜索相似性的方法(1988,Proc.Nat'l.Acad.Sci.USA 85:2444),通过这些算法的计算机化实施(例如:威斯康星遗传学软件包Genetics Computer Group(GCG)中的GAP、BESTFIT、FASTA和TFASTA,575Science Dr.,Madison,Wis.)。

[0155] 适合于确定序列序列同一性和序列相似性百分比的算法的实例为BLAST和BLAST 2.0算法,其分别描述于Altschul等人(1977,Nuc.Acids Res.25:3389-3402)和Altschul等人(1990,J.Mol.Biol.215:403-410)中。使用BLAST和BLAST 2.0,用本文所述的参数,确定本发明的核酸和蛋白的序列同一性百分比。例如BLASTN程序(对于核苷酸序列)可以使用字长(W)11、期望值(E)10、M=5、N=-4以及双链比对作为缺省值。对于氨基酸序列,BLASTP程序可以将下述用作缺省值:字长3和期望值(E)10,以及BLOSUM62得分矩阵(参见Henikoff&Henikoff,1989,Proc.Natl.Acad.Sci.USA 89:10915)对齐(B)50、期望值(E)10、M=5、N=-4以及双链比对。实施BLAST分析的软件可以通过美国国家生物技术信息中心(参见URL:ncbi.nlm.nih.gov/)公开获得。

[0156] 如图2-4所示,与使用同样的启动子、3' UTR和终止子序列且与现有技术表达增强子可操作地连接的目的核苷酸序列的表达相比,使用如本文所述的表达增强子通常导致同样的目的核苷酸序列的表达增加。例如,参考图2和图3,显示了在植物中产生的蛋白的表达的比较,所述植物包含可操作地连接下述的CPMV-HT(现有技术)表达构建体和基于CPMV HT+的表达构建体:

[0157] H1A/California/07/2009(“PDI-H1 Cal”或“H1A/California/07/2009”):基于CPMV HT+的构建体编号1805,基于CPMV HT的构建体编号484(参见实施例4);

[0158] H3A/Victoria/361/2011(“PDI-H3 Vic”或“H3A/Victoria/361/2011”):基于CPMV HT+的构建体编号1819;基于CPMV HT的构建体编号1391(分别参见实施例1和2);

[0159] 具有缺失的蛋白水解环且具有天然信号肽的B/Wisconsin/1/2010(“WtSp-B Wis-PrL”或“B/Wisconsin/1/2010”):基于CPMV HT+的构建体编号1839;基于CPMV HT的构建体编号1445(参见实施例12);

[0160] 具有缺失的蛋白水解环且具有PDI信号肽的B Brisbane/60/08(“B Brisbane/60/08”):基于CPMV HT+的构建体编号1829;基于CPMV HT的构建体编号1039(参见实施例8);

[0161] 具有与跨膜结构域和胞质尾区融合的缺失的蛋白水解环且具有PDI信号肽的B Brisbane/60/08+H1Tm(“B Brisbane/60/08+H1Tm”):基于CPMV HT+的构建体编号1875;基于CPMV HT的构建体1067(参见实施例9),

[0162] 具有缺失的蛋白水解环且具有PDI信号肽的B Massachusetts/2/20122012(“B Massachusetts/2/2012 2012”):基于CPMV HT+的构建体编号2052;基于CPMV HT的构建体编号2072(参见实施例10),

[0163] 具有与跨膜结构域和胞质尾区融合的缺失的蛋白水解环且具有PDI信号肽的B Massachusetts/2/2012+H1Tm(“B Massachusetts/2/2012+H1T m”):基于CPMV HT+的构建体编号2062;基于CPMV HT的构建体2074(参见实施例11),

[0164] 具有与跨膜结构域和胞质尾区融合的缺失的蛋白水解环且具有天然信号肽的B Wisconsin/1/2010+H1Tm(“B Wisconsin/1/2010+H1Tm”):基于CPMV HT+的构建体编号1860;基于CPMV HT的构建体1454(参见实施例13);

[0165] 具有PDI信号肽的来自流感病毒A/Indonesia/5/2005的H5(PDI H5Indo):基于CPMV HT+的构建体编号2319;基于CPMV HT的构建体编号409(参见实施例5);

[0166] 来自流感病毒A/Hangzhou/1/2013的H7,其中天然信号肽已被苜蓿蛋白质二硫键异构酶的信号肽替代(H7 Han):基于CPMV HT+的构建体编号2142;基于CPMV HT的构建体编号2140(参见实施例6);

[0167] 来自流感病毒A/Hangzhou/1/2013的H7,其与来自流感病毒A/Indonesia/5/2005的H5的跨膜结构域和胞质尾区(TMCT)融合,并具有苜蓿蛋白质二硫键异构酶的信号肽(H7 Han+H5Tm):基于CPMV HT+的构建体编号2146;基于CPMV HT的构建体编号2130(参见实施例7)。

[0168] 大体上,当与基于现有技术CPMV HT的构建体中的表达相比时,在基于CPMV HT+的构建体中表达增加(测定为血凝活性)。

[0169] 此外,一些目的核苷酸序列编码嵌合的或修饰的蛋白,例如其包含异源信号肽(例如PDI)、异源跨膜结构域胞质尾区序列(TDCT)和/或含有缺失的蛋白水解环(PrL-)的修饰序列,并且仍观察到活性的增加。

[0170] 如果用于以上基于CPMV HT+的构建体的植物kozak序列被其它植物kozak序列(例如SEQ ID NO:8-16中所限定的那些植物kozak序列之一)替代,仍可观察到使用基于CPMV HT+的构建体所观察到的表达的增加。例如,参考图4,显示了在植物中产生的蛋白的表达的比较,所述植物包含基于CPMV HT+的表达构建体,所述构建体可操作地连接各自与不同的植物kozak序列融合的目的核苷酸序列(H3A/Victoria/361)。在不同的情况下,基于CPMV HT+的构建体的表达(测定为血凝滴度)表明显著的表达水平,并且高于基于现有技术的CPMV HT的构建体。

[0171] 编码蛋白的目的核苷酸序列需要位于待表达的基因上游的“翻译起始位点”或“起始位点”或者“翻译开始位点”或“开始位点”或者“起始密码子”的存在。可以将此类起始位点提供为增强子序列的一部分或者作为编码目的蛋白的核苷酸序列的一部分。

[0172] “表达盒”指包含目的核酸的核苷酸序列,所述目的核酸在合适的启动子或用于在宿主细胞中转录目的核酸的其它调控元件的控制下,或者与合适的启动子或用于在宿主细胞中转录目的核酸的其它调控元件可操作地(或者有效地)连接。

[0173] “蛋白水解环”或“切割位点”意为参与前体HA0切割的蛋白水解位点的共有序列。本文使用的“共有的”或“共有序列(consensus sequence)”意为基于多序列(例如特定流感病毒HA0序列的亚型)比对的分析,包含相关序列的序列变异性的序列(氨基酸序列或者核苷酸序列)。流感病毒HA0切割位点的共有序列可以包含流感病毒A共有的血凝素氨基酸序列,包括例如共有的H1、共有的H3、共有的H5,或者流感病毒B共有的血凝素氨基酸序列,例如但不限于B Florida、B Malaysia、B Wisconsin以及B Massachusetts。蛋白水解环区的

序列的非限制性实例显示于美国临时申请第61/806,227号的图15和18B中(提交于2013年3月28日,将其通过引用并入本文;也参见Bianchi等,2005,Journal of Virology,79:7380-7388;通过引用并入本文)。

[0174] 蛋白水解环或切割位点中的残基可以是突变的,例如但不限于点突变、置换、插入或缺失。本文使用的术语“氨基酸突变”或“氨基酸修饰”意图涵盖氨基酸置换、缺失、插入和修饰。可以如美国临时申请第61/806,227号(提交于2013年3月28日,将其通过引用并入本文)所述对置换、缺失、插入和修饰进行任何组合,以得到最终的构建体,条件是最终的构建体具有期望的特征,例如蛋白水解环或切割位点减少或消除的蛋白酶切割。

[0175] 如本文所述,提供核酸构建体(表达系统),其包含与编码目的蛋白的目的核苷酸序列可操作地连接的表达增强子序列。还提供包含如本文所述的增强子序列的植物表达系统。还提供包含植物调控区的植物表达系统,所述调控区与增强子序列可操作结合,所述增强子序列与目的核苷酸序列可操作地连接,所述目的核苷酸序列编码目的蛋白。所述增强子序列可以选自SEQ ID NO:1或3,或者与SEQ ID NO:1、3或4的核苷酸1-160、SEQ ID NO:1、3或4的核苷酸161-509、或者SEQ ID NO:1、3或4的核苷酸1-509呈现100%、99%、98%、97%、96%、95%、90%、85%或80%同一性的核苷酸序列,其中当如本文所述与植物调控区和植物kozak序列可操作地连接时,与利用相同植物调控区的与CPMV HT(SEQ ID NO:2;如Sainsbury F.和Lomonosoff G.P.,2008,Plant Physiol.148:pp.1212-1218中所述包含不完全的M蛋白的现有技术的增强子序列;将其通过引用并入本文)融合的目的核苷酸序列的表达水平相比,所述表达增强子增加了与所述表达增强子可操作地连接的目的核苷酸序列的表达水平。

[0176] 可将本发明的增强子序列用于在宿主有机体(例如植物)中表达目的蛋白。在此种情况下,目的蛋白也可以与所讨论的宿主有机体异源,并且利用本领域已知的转化技术将其引入植物细胞。有机体中的异源基因可以替代内源的等同基因,即通常发挥相同或相似功能的基因,或者插入序列可以是所述内源基因或其它序列的外加序列。

[0177] 本发明还提供表达盒,其依次包含与如本文所述的表达增强子序列可操作地连接的启动子或植物调控区,所述表达增强子序列与目的核苷酸序列和3' UTR序列以及终止子序列融合。可将增强子序列限定为例如但不限于下述的任何一项:SEQ ID NO:1、3和4,或者与SEQ ID NO:1、3或4的核苷酸1-160、SEQ ID NO:1、3或4的核苷酸161-509或者SEQ ID NO:1、3或4的核苷酸1-509呈现100%、99%、98%、97%、96%、95%、90%、85%或80%同一性的核苷酸序列。表达增强子或目的核苷酸序列可以包含植物kozak序列。

[0178] 本领域技术人员将会理解,终止(终止子)序列可以是在植物宿主中有活性的任何序列,例如终止序列可以来源于二重RNA病毒(例如豇豆花叶病毒组)的RNA-2基因组片段,或者终止序列可以是NOS终止子。

[0179] 本发明的构建体还可以包含3' 非翻译区(UTR)。3' 非翻译区含有多腺苷酸化信号和能够影响mRNA加工或基因表达的任何其它调控信号。多腺苷酸化信号通常以实现向mRNA前体的3' 末端添加多聚腺苷酸轨迹为特征。多腺苷酸化信号通常由经典形式的5' AATAAA-3' 同源物的存在所识别,尽管变化并不罕见。合适的3' 区的非限制性实例是包含下述基因的多腺苷酸化信号的3' 转录的非翻译区:农杆菌(Agrobacterium)肿瘤诱导(Ti)质粒基因如胭脂碱合成酶(Nos基因),以及植物基因如大豆储藏蛋白基因、核酮酶-1,5-二磷酸羧化

酶基因的小亚基(ssRUBISCO;US4,962,028;将其通过引用并入本文)、用于调节质体蓝素的表达的启动子(Pwee和Gray 1993;将其通过引用并入本文)。终止(终止子)序列可由苜蓿质体蓝素基因的3' UTR获得。

[0180] “目的核苷酸(或核酸)序列”或者“目的编码区”意为待在宿主有机体(例如植物)内表达以产生目的蛋白的任何核苷酸序列或编码区(这些术语可互换使用)。此类目的核苷酸序列可以编码但不限于天然的或修饰的蛋白、工业酶或修饰的工业酶、农业蛋白(agricultural protein)或修饰的农业蛋白、辅助蛋白、蛋白补充物、药理学活性蛋白、营养物、附加值产品或者用于饲料、食物或者饲料和食物用途的它们的片段。

[0181] 目的蛋白可以包含天然的或非天然的信号肽;所述非天然的信号肽可以是植物源的或者可从动物或细菌多肽获得。天然的信号肽可能对应于待表达的目的蛋白的信号肽,此外,所述信号肽可以来源于除流感病毒之外的病毒的结构蛋白或血凝素。可以使用的信号肽的非限制性实例为苜蓿蛋白质二硫键异构酶的信号肽(PDI SP;登录号Z11499的核苷酸32-103)或者马铃薯贮藏蛋白信号肽(PatA SP;位于GenBank登录号A08215的核苷酸1738-1806)。该登录号的PatA SP的核苷酸序列为:ATGGCAACTACTAAAACCTTTTTTAATTTTATTTTATGATATTAGCAACTACTAGTTCAACATGTGCT(SEQ ID NO:57);

[0182] 马铃薯贮藏蛋白A信号肽的氨基酸序列为:

[0183] MATTKTFLILFFMILATTSSTCA(SEQ ID NO:58)

[0184] 目的核苷酸序列或目的编码区还可以包含编码药理学活性蛋白的核苷酸序列,例如生长因子、生长调节素、抗体、抗原以及可用于免疫或疫苗接种等的它们的片段或它们的衍生物。此类蛋白包括但不限于:是人类病原体、病毒蛋白的蛋白,例如但不限于来自呼吸道合胞体病毒(RSV)、轮状病毒、流感病毒、人类免疫缺陷病毒(HIV)、狂犬病毒、人乳头瘤病毒(HPV)、肠病毒71(EV71)中的一种或多种蛋白;或者白介素,例如IL-1至IL-24、IL-26和IL-27中的一种或多种;细胞因子,红细胞生成素(EPO)、胰岛素、G-CSF、GM-CSF、hPG-CSF、M-CSF或其组合;干扰素,例如干扰素- α 、干扰素- β 、干扰素- γ ;凝血因子,例如因子VIII、因子IX;或者tPA hGH;受体;受体激动剂;抗体例如但不限于利妥昔单抗;神经多肽;胰岛素;疫苗;生长因子例如但不限于表皮生长因子、角质细胞生长因子、转化生长因子;生长调节素;抗原;自身抗原;或者它们的片段或者它们的组合。

[0185] 目的蛋白还可以包含流感病毒血凝素(HA;参见W0 2009/009876,将其通过引用并入本文)。HA是同源三聚体膜I型糖蛋白,通常包含信号肽、HA1结构域和在C端含有跨膜锚定位点的HA2结构域以及小胞质尾区。编码HA的核苷酸序列众所周知并且是可获得的(参见,例如URL:biohealthbase.org/GSearch/home.do?decorator=Influenza的BioDefense and Public Health Database(Influenza Research Database;Squires等,2008Nucleic Acids Research 36:D497-D503);或者由美国国家生物技术信息中心维护的数据库(参见URL:ncbi.nlm.nih.gov),将它们通过引用并入本文)。

[0186] HA蛋白可以是A型流感病毒、B型流感病毒的HA蛋白,或者是选自以下的A型流感病毒HA的亚型:H1、H2、H3、H4、H5、H6、H7、H8、H9、H10、H11、H12、H13、H14、H15和H16。在本发明的一些方面,HA可以来自A型流感病毒,其选自H1、H2、H3、H5、H6、H7和H9。以上列出的HA的片段也可被认为是目的蛋白。此外,可以将来自以上列出的HA类型或亚型的结构域组合以产生嵌合的HA(参见例如W02009/076778,将其通过引用并入本文)。

[0187] 包含HA蛋白的亚型的实例包括:A/New Caledonia/20/99 (H1N1)、A/Indonesia/5/2006 (H5N1)、A/chicken/New York/1995、A/herring gull/DE/677/88 (H2N8)、A/Texas/32/2003、A/mallard/MN/33/00、A/duck/Shanghai/1/2000、A/northern pintail/TX/828189/02、A/Turkey/Ontario/6118/68 (H8N4)、A/shoveler/Iran/G54/03、A/chicken/Germany/N/1949 (H10N7)、A/duck/England/56 (H11N6)、A/duck/Alberta/60/76 (H12N5)、A/Gull/Maryland/704/77 (H13N6)、A/Mallard/Gurjev/263/82、A/duck/Australia/341/83 (H15N8)、A/black-headed gull/Sweden/5/99 (H16N3)、B/Lee/40、C/Johannesburg/66、A/PuertoRico/8/34 (H1N1)、A/Brisbane/59/2007 (H1N1)、A/Solomon Islands3/2006 (H1N1)、A/Brisbane 10/2007 (H3N2)、A/Wisconsin/67/2005 (H3N2)、B/Malaysia/2506/2004、B/Florida/4/2006、A/Singapore/1/57 (H2N2)、A/Anhui/1/2005 (H5N1)、A/Vietnam/1194/2004 (H5N1)、A/Teal/HongKong/W312/97 (H6N1)、A/Equine/Prague/56 (H7N7)、A/HongKong/1073/99 (H9N2))。

[0188] HA蛋白可以是H1、H2、H3、H5、H6、H7或H9亚型。例如,H1蛋白可以来自A/New Caledonia/20/99 (H1N1)、A/PuertoRico/8/34 (H1N1)、A/Brisbane/59/2007 (H1N1)、A/Solomon Islands 3/2006 (H1N1)、A/California/04/2009 (H1N1) 或A/California/07/2009 (H1N1) 株。H3蛋白也可以来自A/Brisbane 10/2007 (H3N2)、A/Wisconsin/67/2005 (H3N2)、A/Victoria/361/2011 (H3N2)、A/Texas/50/2012 (H3N2)、A/Hawaii/22/2012 (H3N2)、A/New York/39/2012 (H3N2) 或A/Perth/16/2009 (H3N2) 株。在本发明的另一方面,H2蛋白可以来自A/Singapore/1/57 (H2N2) 株。H5蛋白可以来自A/Anhui/1/2005 (H5N1)、A/Vietnam/1194/2004 (H5N1) 或者A/Indonesia/5/2005株。在本发明的一个方面,H6蛋白可以来自A/Teal/HongKong/W312/97 (H6N1) 株。H7蛋白可以来自A/Equine/Prague/56 (H7N7) 株,或者H7A/ Hangzhou/1/2013、A/Anhui/1/2013 (H7N9) 或者A/Shanghai/2/2013 (H7N9) 株。在本发明的一个方面,H9蛋白来自A/HongKong/1073/99 (H9N2) 株。在本发明的另一方面,HA蛋白可以来自可能是B型病毒的流感病毒,包括B/Malaysia/2506/2004、B/Florida/4/2006、B/Brisbane/60/08、B/Massachusetts/2/2012样病毒(Yamagata谱系) 或者B/Wisconsin/1/2010 (Yamagata谱系)。来自H1、H2、H3、H5、H6、H7、H9或者B亚型的HA蛋白的氨基酸序列的非限制性实例包括如WO 2009/009876、WO 2009/076778、WO 2010/003225(将它们通过引用并入本文) 中所述的序列。流感病毒HA蛋白可以是H5 Indonesia。

[0189] HA也可以是嵌合HA,其中HA的天然跨膜结构域被异源跨膜结构域替代。HA蛋白的跨膜结构域高度保守(参见,例如WO 2010/148511的图1C;将其通过引用并入本文)。异源跨膜结构域可以获自任何HA跨膜结构域,例如但不限于来自下述的跨膜结构域:H1 California、B/Florida/4/2006 (GenBank登录号ACA33493.1)、B/Malaysia/2506/2004 (GenBank登录号ABU99194.1)、H1/Bri (GenBank登录号ADE28750.1)、H1A/Solomon Islands/3/2006 (GenBank登录号ABU99109.1)、H1/NC (GenBank登录号AAP34324.1)、H2A/Singapore/1/1957 (GenBank登录号AAA64366.1)、H3A/Brisbane/10/2007 (GenBank登录号ACI26318.1)、H3A/Wisconsin/67/2005 (GenBank登录号AB037599.1)、H5A/Anhui/1/2005 (GenBank登录号ABD28180.1)、H5A/Vietnam/1194/2004 (GenBank登录号ACR48874.1)、H5-Indo (GenBank登录号ABW06108.1)。也可以通过以下共有氨基酸序列来限定跨膜结构域:

[0190] iLXiYstvAiSslXlXXmlagXsXwmcs (SEQ ID NO:59)

[0191] 包含具有异源跨膜结构域的嵌合HA的构建体的实例包括：构建体编号1875 (Brisbane/60/08+H1TM, 其中跨膜结构域和胞质尾区被H1A/California/07/2009替代；参见实施例9)、构建体编号2074 (B Massachusetts/2/2012+H1TM, 其中跨膜结构域和胞质尾区被H1A/California/07/2009的跨膜结构域和胞质尾区替代；参见实施例11) 和构建体编号1860 (B Wisconsin/1/2010+H1TM, 其中跨膜结构域和胞质尾区被H1A/California/07/2009的跨膜结构域和胞质尾区替代；参见实施例13)。这些嵌合HA的活性显示于图2中。

[0192] 本文限定的表达增强子也可用于驱使编码一种或多种目的蛋白的任何目的核苷酸序列的表达。蛋白的实例包括，例如但不限于工业酶，例如，纤维素酶、木聚糖酶、蛋白酶、过氧化物酶、枯草杆菌蛋白酶；蛋白补充物，营养物，附加值产品或者用于饲料、食物或者饲料和食物用途的它们的片段；药学活性蛋白，例如但不限于生长因子，生长调节素，抗体，抗原以及用于免疫或疫苗接种等的它们的片段或它们的衍生物。其它目的蛋白可以包括但不限于：白介素，例如IL-1至IL-24、IL-26和IL-27中的一种或多种；细胞因子，红细胞生成素(EPO)、胰岛素、G-CSF、GM-CSF、hPG-CSF、M-CSF或其组合；干扰素，例如干扰素- α 、干扰素- β 、干扰素- γ ；凝血因子，例如因子VIII、因子IX；或者tPA hGH；受体；受体激动剂；抗体；神经多肽；胰岛素；疫苗；生长因子，例如但不限于表皮生长因子、角质细胞生长因子、转化生长因子；生长调节素；抗原；自身抗原或其片段；抗体，单克隆抗体、嵌合单克隆抗体、单链单克隆抗体、病毒样颗粒(VLP)；或者它们的组合。

[0193] 如果目的蛋白是VLP，则VLP可以包含HA0前体形式，或者通过二硫桥保持在一起的HA1或HA2结构域形式。VLP可以具有约20nm至1 μ m或者其间任何量的平均大小，例如60、65、70、75、80、85、90、95、100、105、110、120、130、140、150 160、170、180、190或200nm，或者其间任何量，例如100nm，并且其可以包含脂膜。VLP可以是有包膜的或者无包膜的，例如病毒包膜蛋白、病毒结构蛋白、病毒衣壳蛋白或者病毒外壳蛋白。VLP还可以包含一种或多种脂质、磷脂、核酸、膜等。

[0194] HA可以包含天然的或非天然的信号肽；所述非天然的信号肽可以是植物源的。例如，信号肽可以是蛋白质二硫键异构酶信号肽(PDI)。天然的信号肽可以对应于表达的血凝素的信号肽，或者可以对应于第二血凝素。

[0195] 本发明还提供包含编码HA蛋白的序列的核酸分子。核酸分子还可以包含与编码HA蛋白的序列可操作地连接的一个或多个调控区。核酸分子可以包含编码H1、H2、H3、H4、H5、H6、H7、H8、H9、H10、H11、H12、H13、H14、H15、H16或者来自B型流感病毒的HA的序列。例如，由所述核酸分子编码的HA蛋白可以是H1、H2、H3、H5、H6、H7、H9亚型、来自B型的HA。由所述核酸编码的H1蛋白可以来自A/New Caledonia/20/99 (H1N1)、A/PuertoRico/8/34 (H1N1)、A/Brisbane/59/2007 (H1N1)、A/Solomon Islands 3/2006 (H1N1)、A/California/04/2009 (H1N1) 或者A/California/07/2009 (H1N1) 株。由所述核酸分子编码的H3蛋白可以来自A/Brisbane 10/2007 (H3N2)、A/Wisconsin/67/2005 (H3N2)、A/Victoria/361/2011 (H3N2)、A/Texas/50/2012 (H3N2)、A/Hawaii/22/2012 (H3N2)、A/New York/39/2012 (H3N2) 或者A/Perth/16/2009 (H3N2) 株。由所述核酸分子编码的H2蛋白可以来自A/Singapore/1/57 (H2N2) 株。由所述核酸分子编码的H5蛋白来自A/Anhui/1/2005 (H5N1)、A/Vietnam/1194/2004 (H5N1) 或者A/Indonesia/5/2005株。由所述核酸分子编码的H6蛋白可以来自A/Teal/HongKong/W312/97 (H6N1) 株。由所述核酸分子编码的H7蛋白可以来自A/Equine/

Prague/56 (H7N7) 株或者H7A/Hangzhou/1/2013、A/Anhui/1/2013 (H7N9) 或者A/Shanghai/2/2013 (H7N9) 株。此外,由所述核酸分子编码的H9蛋白可以来自A/HongKong/1073/99 (H9N2) 株。由所述核酸分子编码的HA蛋白可以来自B型流感病毒,包括B/Malaysia/2506/2004、B/Florida/4/2006、B/Brisbane/60/08、B/Massachusetts/2/2012样病毒 (Yamagata 谱系) 或者B/Wisconsin/1/2010 (Yamagata谱系)。来自H1、H2、H3、H5、H6、H7、H9或者B亚型的HA蛋白的氨基酸序列的非限制性实例包括如W0 2009/009876、W0 2009/076778、W0 2010/003225 (将它们通过引用并入本文) 中所述的序列。流感病毒HA蛋白可以是H5 Indonesia。

[0196] 表1:如本文所述制备的构建体的实例:

[0197] 基于CPMV-HT的构建体 (包含SEQ ID NO:2的构建体;现有技术)

[0198]

构建体#	SP ¹	目的序列	实施例
1391	PDI ²	H3 Victoria	1
484	PDI	H1 California	4
409	PDI	H5 Indonesia	5
2140	PDI	H7 Hangzhou	6
2130	PDI	H7 Hangzhou+H5 Indonesia TMCT ⁴	7
1039	PDI	B Brisbane (PrL-)	8
1067	PDI	B Brisbane (PrL-)+H1 California TMCT	9
2072	PDI	B Massachusetts (PrL-)	10
2074	PDI	B Massachusetts (PrL-)+H1 California TMCT	11
1445	WT ³	B Wisconsin (PrL-)	12
1454	WT	B Wisconsin (PrL-)+H1 California TMCT	13

[0199] 基于CPMV HT+的构建体 (包含SEQ ID NO:1的构建体)

[0200]

构建体#	SP	目的序列	实施例
1819	PDI	H3 Victoria	2
1805	PDI	H1 California	4
2319	WT	H5 Indonesia	5
2142	PDI	H7 Hangzhou	6
2146	PDI	H7 Hangzhou+ H5 Indonesia TMCT	7
1829	PDI	B Brisbane(PrL-)	8
1875	PDI	B Brisbane(PrL-)+H1 California TMCT	9
2052	PDI	B Massachusetts (PrL-)	10
2062	PDI	B Massachusetts (PrL-)+H1 California TMCT	11
1839	WT	B Wisconsin (PrL-)	12
1860	WT	B Wisconsin (PrL-)+H1 California TMCT	13

[0201] ¹:SP-信号肽

[0202] ²:PDI-苜蓿蛋白质二硫键异构酶

[0203] ³:WT-野生型或者天然的

[0204] ⁴:TMCT-跨膜结构域和胞质尾区

[0205] 如果目的核酸序列编码对植物直接或者间接有毒的产物,则可以通过在期望的组

织内或者在植物发育的期望阶段选择性表达目的核苷酸序列来降低此类毒性。

[0206] 可以在任何合适的植物宿主中表达目的编码区或者目的核苷酸序列,所述植物宿主转化有或者包含本发明的核苷酸序列或核酸分子或遗传构建体或者载体。合适的宿主的实例包括但不限于,拟南芥,农作物包括例如:双低油菜(*canola*)、芸薹属、玉米(*maize*)、烟草属(烟草)例如本氏烟草(*Nicotiana benthamiana*)、苜蓿、马铃薯、甘薯(*Ipomoea batatas*)、人参、豌豆、燕麦、大米、大豆、小麦、大麦、向日葵、棉花、玉米(*corn*)、黑麦(*Secale cereale*)、高粱(*Sorghum bicolor*、*Sorghum vulgare*)、红花(*Carthamus tinctorius*)。

[0207] 本文使用的术语“生物质”和“植物物质”指来源于植物的任何材料。生物质或植物物质可以包括整株植物或者植物部分,所述植物部分包括叶、根、茎、花、种子,所述生物质或植物物质还可以包括植物的任何组织、植物的任何细胞或者植物、植物部分、组织或细胞的任何部分。此外,生物质或植物物质可以包括细胞内植物组分、细胞外植物组分、植物的液体或固体提取物或者它们的组合。此外,生物质或植物物质可以包括植物,来自植物叶、茎、果实、根或其组合的植物细胞、组织、液体提取物或其组合。植物部分可以包含植物物质或生物质。

[0208] “调控区”、“调控元件”或“启动子”意为通常但不总是位于基因的蛋白编码区的上游的核酸的一部分,其由DNA或RNA或者DNA和RNA二者组成。当调控区具有活性并与目的基因可操作结合或者可操作地连接时,这可以导致目的基因的表达。调控元件可能能够介导器官特异性或者控制发育基因或时序基因的激活。“调控区”包括启动子元件、显示出基础启动子活性的核心启动子元件、可以响应外部刺激而被诱导的元件、介导启动子活性的元件如负调控元件或转录增强子。本文使用的“调控区”还包括转录之后有活性的元件,例如调节基因表达的调控元件,如翻译和转录增强子、翻译和转录抑制子、上游激活序列以及mRNA不稳定决定子。这些后面的元件中的一些可以位于近编码区。

[0209] 在本公开的背景,术语“调控元件”或“调控区”通常指通常但不总是位于结构基因的编码序列上游(5')的DNA序列,其通过提供RNA聚合酶和/或在特定位点开始转录所需的其它因子的识别来控制编码区的表达。然而,应当理解,位于内含子内的其它核苷酸序列或者所述序列的3'也可能有助于目的编码区的表达调控。提供RNA聚合酶或者其它转录因子的识别以确保在特定位点处起始的调控元件的实例是启动子元件。大部分但非全部的真核生物启动子元件含有TATA盒,所述TATA盒为由通常位于转录开始位点上游大约25个碱基对的腺苷和胸苷核苷酸碱基对组成的保守核酸序列。启动子元件可以包括负责转录起始的基础启动子元件以及修饰基因表达的其它调控元件(如以上所列出的)。

[0210] 存在数种类型的调控区,包括发育性调控的调控区、诱导型的调控区或组成型的调控区。在某些器官或器官的组织发育期间,在特定时间,发育性调控的调控区或者控制受其控制的基因的差异表达的调控区在所述器官或器官的组织内被激活。然而,一些发育性调控的调控区可能在特定发育阶段的某些器官或者组织中优先被激活,它们还可能以发育性调控的方式被激活或者也处于在植物内的其它器官或组织中的基础水平。组织特异性调控区例如种子特异性调控区的实例包括napin启动子和cruciferin启动子(Rask等,1998, *J. Plant Physiol.* 152:595-599; Bilodeau et al., 1994, *Plant Cell* 14:125-130)。叶特异性启动子的实例包括质体蓝素启动子(参见US 7,125,978,将其通过引用并入本文)。

[0211] 诱导型调控区是能够响应诱导物,直接或者间接激活一个或多个DNA序列或基因的转录的调控区。在诱导物不存在的情况下,所述DNA序列或基因将不转录。通常,与诱导型调控区特异性结合以激活转录的蛋白因子可以以无活性的形式存在,然后通过诱导物将其直接或者间接转化为活性形式。然而,蛋白因子也可以不存在。诱导物可以是诸如蛋白、代谢物、生长调节剂、除草剂或酚类化合物的化学物质,或者由热、冷或毒性元素直接施加的生理应激,或者通过病原体或疾病原如病毒的作用间接施加的生理应激。通过将诱导物外部应用于细胞或植物如通过喷射、浇水、加热或类似的方法,可以将含有诱导型调控区的植物细胞暴露于所述诱导物。诱导型调控元件可以来源于植物基因或非植物基因(例如Gatz, C和Lenk, I.R.P., 1998, Trends Plant Sci. 3, 352-358; 将其通过引用并入)。潜在的诱导型启动子的实例包括但不限于四环素可诱导的启动子(Gatz, C., 1997, Ann. Rev. Plant Physiol. Plant Mol. Biol. 48, 89-108; 将其通过引用并入), 类固醇可诱导的启动子(Aoyama, T. and Chua, N.H., 1997, Plant J. 2, 397-404; 将其通过引用并入)和乙醇可诱导的启动子(Salter, M.G. 等, 1998, Plant Journal 16, 127-132; Caddick, M.X. deng, 1998, Nature Biotech. 16, 177-180; 将其通过引用并入)、细胞分裂素可诱导的IB6和CKI1基因(Brandstatter, I. 和 Kieber, J.J., 1998, Plant Cell 10, 1009-1019; Kakimoto, T., 1996, Science 274, 982-985; 将其通过引用并入)以及植物生长素可诱导的元件DR5(Ulmasov, T. 等, 1997, Plant Cell 9, 1963-1971; 将其通过引用并入)。

[0212] 组成型调控区引导遍及植物不同部分的基因的表达,以及连续贯穿植物发育的基因的表达。已知的组成型调控元件的实例包括与CaMV 35S转录本相关的启动子(p35S; Odell et al., 1985, Nature, 313:810-812), 稻肌动蛋白1(Zhang等, 1991, Plant Cell, 3: 1155-1165)、肌动蛋白2(An等, 1996, Plant J., 10:107-121)或者tms 2(U.S. 5, 428, 147, 将它们通过引用并入本文), 以及磷酸丙糖异构酶1(Xu等, 1994, Plant Physiol. 106:459-467)基因, 玉米泛素1基因(Cornejo等, 1993, Plant Mol. Biol. 29:637-646), 拟南芥泛素1和6基因(Holtorf等, 1995, Plant Mol. Biol. 29:637-646), 烟草翻译起始因子4A基因(Mandel等, 1995 Plant Mol. Biol. 29:995-1004), 木薯脉花叶病毒(Cassava Vein Mosaic Virus)启动子pCAS(Verdaguer等, 1996), 核酮酶二磷酸羧化酶的小亚基的启动子pRbcS:(Outchkourov等, 2003), pUbi(对于单子叶植物和双子叶植物)。

[0213] 如本文所述,已经发现包含在叶的表达中具有证实效力的增强子序列的调控区在瞬时表达中有效。在不希望受理论的束缚的情况下,通过与核基质连接而与光合基因的上游调控元件的连接可以介导强的表达。例如从豌豆质体蓝素的翻译开始位点至-784(US 7, 125, 978, 将其通过引用并入本文)用于介导强报告基因的表达。

[0214] 本文使用的术语“组成型的”不一定指示,受组成型调控区控制的核苷酸序列在所有细胞类型中的表达均处于同一水平,而是所述序列在广泛的细胞类型中表达,尽管经常观察到丰度的变化。

[0215] 以上所述的表达构建体可以存在于载体中。载体可以包含允许表达盒转移并整合进有机体或宿主的基因组的边界序列。构建体可以是植物二元载体,例如基于pPZP的双元转化载体(Hajdukiewicz等1994)。其它示例构建体包括pBin19(参见Frisch, D.A., L.W.Harris-Haller等, 1995, Plant Molecular Biology 27:405-409)。

[0216] 如果需要,本发明的构建体可进一步操作为包含选择性标志物。然而,这可能不需

要。可用的选择性标志物包括提供针对化学物质的抗性的酶,所述化学物质如抗生素,例如庆大霉素、潮霉素、卡那霉素;除草剂,如草铵膦(phosphinothrycin)、草甘膦、氯磺隆等。类似地,可以使用提供可通过颜色变化鉴定化合物的产生的酶,如GUS(B-葡萄糖醛酸苷酶),或者通过发冷光鉴定的酶,如荧光素酶或GFP。

[0217] 载体还可以包含本文所述的表达增强子。表达增强子可以位于T-DNA上,所述T-DNA还包含基因沉默抑制子和NPTII。多接头也可以编码一组或两组6×组氨酸残基,以允许目的蛋白包含N-端或C-端His标签,从而有助于蛋白纯化。

[0218] 转录后基因沉默(PTGS)可能涉及限制植物中转基因的表达,并且来自马铃薯病毒Y(HcPro)的沉默抑制子的共表达可用于消解转基因mRNA的特异性降解(Brigneti等,1998,EMBO J.17,6739-6746,将其通过引用并入本文)。可选的沉默抑制子在本领域众所周知,并且可如本文所述使用(Chiba等,2006,Virology 346:7-14;将其通过引用并入本文),例如但不限于:TEV-p1/HC-Pro(烟草蚀刻病毒-p1/HC-Pro)、BYV-p21、番茄丛矮病毒的p19(TBSV p19;p19的结构描述于WO 2010/0003225中,将其通过引用并入本文)、番茄皱缩病毒的衣壳蛋白(TCV-CP)、黄瓜花叶病毒的2b(CMV-2b)、马铃薯病毒X的p25(PVX-p25)、马铃薯病毒M的p11(PVM-p11)、马铃薯病毒S的p11(PVS-p11)、蓝莓枯黄病毒的p16(BScV-p16)、柑橘衰退病毒的p23(CTV-p23)、葡萄卷叶伴随病毒2号的p24(GLRaV-2p24)、葡萄病毒A的p10(GVA-p10)、葡萄病毒B的p14(GVB-p14)、独活潜隐病毒的p10(HLV-p10)或者大蒜普通潜隐病毒的p16(GCLV-p16)。

[0219] 因此,可将一种或多种沉默抑制子与基于豇豆花叶病毒组的表达盒、来源于双生病毒的扩增元件以及编码目的蛋白的核酸序列共表达,以进一步确保植物内高水平的蛋白产生,所述沉默抑制子例如但不限于:HcPro、TEV-p1/HC-Pro、BYV-p21、TBSV p19、TCV-CP、CMV-2b、PVX-p25、rgscam、来自FHV的B2蛋白、CPMV的小外壳蛋白和来自TCV的外壳蛋白、PVM-p11、PVS-p11、BScV-p16、CTV-p23、GLRaV-2p24、GBV-p14、HLV-p10、GCLV-p16或者GVA-p10。

[0220] 可以利用Ti质粒、Ri质粒、植物病毒载体、直接DNA转化、微注射、电穿孔等将本发明的构建体引入植入细胞。关于此类技术的综述,参见例如Weissbach和Weissbach, *Methods for Plant Molecular Biology*, Academy Press, New York VIII, pp.421-463 (1988); Geierman和Corey, *Plant Molecular Biology*, 2d Ed. (1988); 以及Miki和Iyer, *Fundamentals of Gene Transfer in Plants*. In *Plant Metabolism*, 2d Ed. DT. Dennis, DH Turpin, DD Lefebvre, DB Layzell (eds), Addison Wesley, Langmans Ltd. London, pp.561-579 (1997)。其它方法包括直接DNA摄取、使用脂质体、电穿孔(例如使用原生质体)、微注射、微弹轰击法(microprojectile)或晶须(whisker)以及真空渗入。参见,例如, Bilang等(1991, *Gene* 100:247-250), Scheid等(1991, *Mol. Gen. Genet.* 228:104-112), Guerche等(1987, *Plant Science* 52:111-116), Neuhauser等(1987, *Theor. Appl. Genet.* 75:30-36), Klein等(1987, *Nature* 327:70-73); Freeman等(1984, *Plant Cell Physiol.* 29:1353), Howell等(1980, *Science* 208:1265), Horsch等(1985, *Science* 227:1229-1231), DeBlock等(1989, *Plant Physiology* 91:694-701), *Methods for Plant Molecular Biology* (Weissbach和Weissbach, eds., Academic Press Inc., 1988), *Methods in Plant Molecular Biology* (Schuler和Zielinski, eds., Academic Press Inc., 1989), WO 92/

09696, WO 94/00583, EP 331083, EP 175966, Liu和Lomonossoff (2002, J Virol Meth, 105: 343-348), EP290395; WO 8706614; 美国专利号4,945,050、5,036,006和5,100,792; 提交于1995年5月10日的美国专利申请连续号08/438,666和提交于1992年9月25日的美国专利申请连续号07/951,715(在此将它们通过引用全部并入)。

[0221] 可将瞬时表达方法用于表达本发明的构建体(参见D' Aoust等, 2009, Methods in molecular biology, Vol 483, 41-50页; Liu和Lomonossoff, 2002, Journal of Virological Methods, 105:343-348; 将其通过引用并入本文)。可选地, 可以使用如Kapila等人(1997, Plant Sci. 122, 101-108; 将其通过引用并入本文)或者WO 00/063400、WO 00/037663(将它们通过引用并入本文)所述的基于真空的瞬时表达方法。这些方法可以包括, 例如但不限于农杆菌接种(Agro-inoculation)或农杆菌渗入(Agro-infiltration)、注射渗入的方法, 但是如以上所示, 也可以使用其它瞬时方法。随着农杆菌接种、农杆菌渗入或者注射渗入, 包含期望核酸的农杆菌(Agrobacteria)的混合物进入组织的细胞间隙, 例如叶、植物的地上部分(包括茎、叶和花)、植物的其它部分(茎、根、花)或者全植株。穿过表皮之后, 农杆菌感染细胞, 并将t-DNA拷贝转移至所述细胞内。t-DNA以附加体的形式转录, 进行mRNA翻译, 导致感染的细胞中目的蛋白的产生, 然而t-DNA在细胞核内的传代是短暂的。

[0222] 本发明的另外考虑部分是含有本发明的基因构建体的转基因植物、植物细胞或种子, 它们可用作适于本文所述的瞬时蛋白表达的平台植物。由植物细胞再生为全植株的方法也是本领域已知的(例如参见Guerineau和Mullineaux(1993, Plant transformation and expression vectors. In: Plant Molecular Biology Labfax (Croy RRD ed) Oxford, BIOS Scientific Publishers, pp 121-148)。通常, 在合适的培养基中培养转化的植物细胞, 如果选择性标志物被用于帮助鉴定转化的植物细胞, 所述培养基可以含有诸如抗生素的选择剂。一旦愈伤组织形成, 可以通过根据已知的方法使用合适的植物激素促进芽的形成, 并且将所述芽转移至生根培养基以用于植物再生。然后可以由种子或者利用无性繁殖技术将所述植物用于建立重复子代。也可以不利用组织培养而产生转基因植物。本领域已经建立用于稳定转化以及这些有机体再生的方法, 并且其是本领域技术人员已知的。在Vasil等人(Cell Culture and Somatic Cell Genetics of Plants, Vol I, II and III, Laboratory Procedures and Their Applications, Academic Press, 1984)以及Weissbach和Weissbach(Methods for Plant Molecular Biology, Academic Press, 1989)中评述了可用的技术。获得转化的和再生的植物的方法不是本发明的关键。

[0223] 如果将用两种或更多种核酸构建体转化或者共转化植物、植物部分或者植物细胞, 则在单次转染事件中将所述核酸构建体引入农杆菌, 将核酸合并, 并且如所述转染所述细菌细胞。可选地, 可以连续引入所述构建体。在这种情况下, 如所述将第一构建体引入农杆菌, 使细胞在选择性条件下生长(例如在抗生素存在的情况下), 其中只有单转化的细菌可以生长。在该第一选择步骤之后, 如所述将第二核酸构建体引入农杆菌, 并且使细胞在双选择性条件下生长, 其中只有双重转化的细菌可以生长。然后如本文所述可将双重转化的细菌用于转化植物、植物部分或者植物细胞, 或者可使其进行再次转化步骤, 以容纳第三核酸构建体。

[0224] 可选地, 如果将用两种或更多种核酸构建体转化或者共转化植物、植物部分或植物细胞, 可通过用植物、植物部分或植物细胞共渗入农杆菌细胞的混合物而将核酸构建体

引入植物,各农杆菌细胞可以包含一种或多种待引入所述植物的构建体。为了改变在构建体内的目的核苷酸序列在植物、植物部分或植物细胞内的相对表达水平,在渗入步骤期间,包含期望构建体的各农杆菌群的浓度可以不同。

[0225] 本公开还提供包含如本文所限定的表达系统的转基因植物,其中当与缺乏如本文所述的表达系统中的一种或多种组件的其它相似的表达系统(例如CPMV HT (SEQ ID NO: 2))相比时,盒中的异源目的核酸的表达水平升高。

[0226] 本公开还包括产生目的蛋白的方法,其包括下述步骤:提供表达如本文所述的表达系统的植物、植物部分,至少收获已表达目的蛋白的组织,以及任选地,从所述组织中分离目的蛋白。

[0227] 因此在各方面,并且不受限制,本发明提供:

[0228] -表达增强子,其包含选自SEQ ID NO:1或3的豇豆花叶病毒组5' UTR或者与SEQ ID NO:1或3中任一项所示的序列显示出100%、99%、98%、97%、96%、95%、90%、85%或80%同一性的核苷酸序列,其中当如本文所述与植物调控区和在植物中有活性的kozak序列(例如植物kozak序列)可操作地连接时,与利用同样的植物调控区的与CPMV HT (SEQ ID NO:2; 如Sainsbury F.和Lomonosoff G.P.,2008,Plant Physiol.148:pp.1212-1218中所述包含不完全的M蛋白的现有技术的增强子序列;将其通过引用并入本文)融合的目的核苷酸序列的表达水平相比,所述表达增强子增加了与所述表达增强子可操作地连接的所述目的核苷酸序列的表达水平。

[0229] -一种或多种表达系统,其包含如以上所限定的基于豇豆花叶病毒组的表达增强子或表达盒、启动子(调控区),任选地多接头、kozak序列、编码目的蛋白的核酸以及终止子。

[0230] -利用如本文所述的一种或多种表达系统或载体,在宿主有机体(如植物)中表达目的蛋白的方法。

[0231] -由本发明的一种或多种表达系统或载体表达目的蛋白的宿主细胞和有机体,以及产生所述宿主和有机体的方法。

[0232] 表2:序列列表

[0233]

SEQ ID NO	描述	SEQ ID NO	描述
1	CPMV HT+	31	引物 IF-HT5*-PDI.c
2	CPMV HT	32	引物 IF-HT6*-PDI.c
3	CPMV HT+ WT115	33	IF-HT7*-PDI.c
4	CPMV HT+ (植物 kozak 共有序列)	34	IF-HT8*-PDI.c
5	共有的植物界 kozak 序列	35	PDISP/H1 California 的核苷酸序列
6	共有的双子叶植物 kozak 序列	36	PDISP/H1 California 的氨基酸序列
7	共有的拟南芥 kozak 序列	37	PDISP H5 Indonesia 的核苷酸序列
8	kozak 序列 AGAAA	38	PDISP H5 Indonesia 的氨基酸序列
9	kozak 序列 AGACA	39	PDISP/H7 Hangzhou 的核苷酸序列
10	kozak 序列 AGGAA	40	PDISP/H7 Hangzhou 的氨基酸序列
11	kozak 序列 AAAAA	41	PDISP/H7 Hangzhou+H5 Indonesia TMCT 的核苷酸序列
12	kozak 序列 AAACA	42	PDISP/H7 Hangzhou+H5 Indonesia TMCT 的氨基酸序列
13	kozak 序列 AAGCA	43	PDISP/HA B Brisbane (PrL-) 的核苷酸序列
14	kozak 序列 AAGAA	44	PDISP/HA B Brisbane (PrL-) 的氨基酸序列
15	kozak 序列 AAAGAA	45	PDISP/HA B Brisbane (PrL-)+H1 California TMCT 的核苷酸序列
16	kozak 序列 AAAAGAA	46	PDISP/HA B Brisbane (PrL-)+H1 California TMCT

[0234]

			的氨基酸序列
17	共有的 kozak 序列 (A/-)A(A/G)(A/G)(A/C)A	47	PDISP/HA B Massachusetts (PrL-)的核苷酸序列
18	引物 IF-PDI.S1+3c	48	PDISP/HA B Massachusetts (PrL-)的氨基酸序列
19	引物 IF-H3V36111.s1-4r	49	PDISP/HA B Massachusetts (PrL-)+H1 California TMCT 的核苷酸序列
20	PDISP/H3 Victoria 的核苷酸序 列	50	PDISP/HA B Massachusetts (PrL-)+H1 California TMCT 的氨基酸序列
21	构建体 1191 的核苷酸序列	51	HA B Wisconsin (PrL-)的核 苷酸序列
22	表达盒 1391 的核苷酸序列	52	HA B Wisconsin (PrL-)的氨 基酸序列
23	PDISP/H3 的氨基酸序列	53	HA B Wisconsin (PrL-)+H1 California TMCT 的核苷酸 序列
24	引物 IF (SacII)-Kozac_PDI.c	54	HA B Wisconsin (PrL-)+H1 California TMC 的氨基酸序 列
25	构建体 2181 的核苷酸序列	55	天然 CPMV RNA 2 的基因组 片段, 核苷酸 1-514, (WO2009/087391, 表 1)
26	表达盒 1819 的核苷酸序列	56	CPMV RNA 2 基因组, 核苷 酸 1-514, 115 位和 162 位具 有修饰
27	引物 IF-HT1*-PDI.c	57	马铃薯贮藏蛋白信号肽(核 苷酸序列)
28	引物 IF-HT2*-PDI.c	58	马铃薯贮藏蛋白信号肽(氨 基酸序列)
29	引物 IF-HT3*-PDI.c	59	HA 跨膜结构域共有氨基酸 序列
30	引物 IF-HT4*-PDI.c		

[0235] 实施例1:2X35S/CPMV-HT/PDISP/H3 Victoria/NOS(构建体编号1391)

[0236] 利用以下基于PCR的方法,将编码天然信号肽已被苜蓿蛋白质二硫键异构酶的信号肽替代的来自流感病毒A/Victoria/361/2011的H3的序列(PDISP/H3 Victoria)克隆进2X35S-CPMV-HT-NOS表达体。使用引物IF-PDI.S1+3c(图5A,SEQ ID NO:18)和IF-H3V36111.s1-4r(图5B,SEQ ID NO:19),使用PDISP/H3 Victoria序列(图5C,SEQ ID NO:20)作为模板扩增含有PDISP/H3 Victoria编码序列的片段。使用In-Fusion克隆系统(Clontech,Mountain View,CA)将PCR产物克隆进2X35S/CPMV-HT/NOS表达系统。用SacII和StuI限制酶消化构建体编号1191(图5D和5E,SEQ ID NO:21),并将线性化的质粒用于In-

Fusion装配反应。构建体编号1191是旨在将目的基因“In Fusion”克隆进基于CPMV-HT的表达盒的受体质粒。并入基因构建体以在苜蓿质体蓝素基因启动子和终止子下共表达TBSV P19沉默抑制子。骨架为pCAMBIA双元质粒,并且由左至右的t-DNA边界序列显示于图5E (SEQ ID NO:21)中。得到的构建体给予编号1391 (图5F, SEQ ID NO:22)。与PDISP融合的来自流感病毒A/Victoria/361/2011的成熟H3的氨基酸序列显示于图5G (SEQ ID NO:23)中。质粒1391的示意图显示于图5H中。

[0237] 实施例2:2X35S/CPMV-HT+/PDISP/H3 Victoria/NOS (构建体编号1819)

[0238] 利用以下基于PCR的方法,将编码天然信号肽已被苜蓿蛋白质二硫键异构酶的信号肽替代的来自流感病毒A/Victoria/361/2011的H3的序列(PDISP/H3 Victoria)克隆进2X35S-CPMV-HT+/NOS表达体。使用引物IF (SacII)-Kozac_PDI.c (图6A, SEQ ID NO:24)和IF-H3V36111.s1-4r (图5B, SEQ ID NO:19),使用PDISP/H3 Victoria序列 (图8C, SEQ ID NO:20)作为模板,扩增PDISP/H3 Victoria编码序列的片段。使用In-Fusion克隆系统 (Clontech, Mountain View, CA),将PCR产物克隆进2X35S/CPMV-HT+/NOS表达系统。用SacII和StuI限制酶消化构建体编号2181 (图6B),并将线性化的质粒用于In-Fusion装配反应。构建体编号2181是旨在将目的基因“In Fusion”克隆进基于CPMV-HT+的表达盒的受体质粒。同时并入基因构建体,以在苜蓿质体蓝素基因启动子和终止子下共表达TBSV P19沉默抑制子。骨架为pCAMBIA双元质粒,并且由左至右的t-DNA边界显示于图6C (SEQ ID NO:25)中。得到的构建体给予编号1819 (图6D, SEQ ID NO:26)。与PDISP融合的来自流感病毒的成熟H3的氨基酸序列显示于图5G (SEQ ID NO:23)中。质粒1819的示意图显示于图6E中。

[0239] 实施例3 2X35S/CPMV HT+/NOS表达系统中的SacII限制性位点与PDISP/H3 Victoria的ATG之间的序列变化 (构建体编号1952至1959)

[0240] 利用与构建体编号1819 (参见实施例2)相同的基于PCR的方法,用修饰的正向引物并保持所有的其它步骤相同,创建八个包含2X35S/CPMV HT+/NOS表达系统中的SacII限制性位点与PDISP/H3Victoria的ATG之间序列变化的构建体。使用图7A-7H中列出的引物扩增变体HT1*至HT8*,以分别创建构建体编号1952至1959:

[0241] IF-HT1*(-Mprot)-PDI.c (图7A, SEQ ID NO:27),

[0242] IF-HT2*(-Mprot)-PDI.c (图7B, SEQ ID NO:28),

[0243] IF-HT3*(-Mprot)-PDI.c (图7C, SEQ ID NO:29)

[0244] IF-HT4*(-Mprot)-PDI.c (图7D, SEQ ID NO:30)

[0245] IF-HT5*(-Mprot)-PDI.c (图7E, SEQ ID NO:31)

[0246] IF-HT6*(-Mprot)-PDI.c (图7F, SEQ ID NO:32)

[0247] IF-HT7*(-Mprot)-PDI.c (图7G, SEQ ID NO:33) 以及

[0248] IF-HT8*(-Mprot)-PDI.c (图7H, SEQ ID NO:34)。

[0249] 质粒1952的示意图显示于图7I中。使用类似的特性来制备构建体1953-1959。

[0250] 实施例4:PDISP/H1 California的2X35S/CPMV HT (构建体编号484) 和HT+ (构建体编号1805)

[0251] 利用与构建体1391和1819相同的基于PCR的方法 (参见实施例1和2),但是使用为PDISP/H1 California特别设计的修饰的PCR引物,将编码序列分别克隆进原始的HT和修饰的HT+中,所述编码序列对应于天然信号肽已被苜蓿蛋白质二硫键异构酶的信号肽替代的

来自流感病毒A/California/7/2009 (PDISP/H1 California) 的H1 (图8A, SEQ ID NO:35)。与PDISP融合的来自流感病毒A/California/7/2009的成熟H1的氨基酸序列显示于图8B (SEQ ID NO:36) 中。质粒484和1805的示意图显示于图8C和图8D中。

[0252] 实施例5:H5 Indonesia的2X35S/CPMV HT (构建体编号409) 和HT+ (构建体编号2319)

[0253] 利用与构建体1391和1819相同的基于PCR的方法 (参见实施例1和2), 但是使用为PDI/H5 Indonesia特别设计的修饰的PCR引物, 将编码序列分别克隆进原始的HT和修饰的HT+中, 所述编码序列对应于天然信号肽已被苜蓿蛋白质二硫键异构酶的信号肽替代的来自流感病毒A/Indonesia/5/2005的H5 (图9A, SEQ ID NO:37)。与PDISP融合的来自流感病毒A/Indonesia/5/2005的成熟H5的氨基酸序列显示于图9B (SEQ ID NO:38) 中。质粒409和2319的示意图显示于图9C和图9D中。

[0254] 实施例6:PDISP-H7 Hangzhou的2X35S/CPMV HT (构建体编号2140) 和HT+ (构建体编号2142)

[0255] 利用与构建体1391和1819相同的基于PCR的方法 (参见实施例1和2), 但是使用为PDISP/H7 Hangzhou特别设计的修饰的PCR引物, 将编码序列分别克隆进原始的HT和修饰的HT+中, 所述编码序列对应于天然信号肽已被苜蓿蛋白质二硫键异构酶的信号肽替代的来自流感病毒A/Hangzhou/1/2013的H7 (PDISP/H7 Hangzhou) (图10A, SEQ ID NO:39)。与PDISP融合的来自流感病毒A/Hangzhou/1/2013的成熟H7的氨基酸序列显示于图10B (SEQ ID NO:40) 中。质粒2140、2142和2168的示意图显示于图10C和图10D中。

[0256] 实施例7:PDISP/H7 Hangzhou+H5 Indonesia TMCT的2X35S/CPMV HT (构建体编号2130) 和HT+ (构建体编号2146)

[0257] 利用与构建体1391和1819相同的基于PCR的方法 (参见实施例1和2), 但是使用为PDISP/H7 Hangzhou+H5 Indonesia TMCT特别设计的修饰的PCR引物, 将嵌合体血凝素编码序列分别克隆进原始的HT和修饰的HT+中, 所述嵌合体血凝素编码序列对应于与来自流感病毒A/Indonesia/5/2005的H5的跨膜结构域和胞质尾区 (TMCT) 融合并具有苜蓿蛋白质二硫键异构酶的信号肽的来自流感病毒A/Hangzhou/1/2013的H7的胞外域 (PDISP/H7 Hangzhou+H5 Indonesia TMCT) (图11A, SEQ ID NO:41)。与PDISP融合的H7 Hangzhou+H5 Indonesia TMCT的氨基酸序列显示于图11B (SEQ ID NO:42) 中。质粒2130和2146的示意图显示于图11C和11D中。

[0258] 实施例8:PDISP/HA B Brisbane (PrL-) 的2X35S/CPMV HT (构建体编号1039) 和HT+ (构建体编号1829)

[0259] 利用与构建体1391和1819相同的基于PCR的方法 (参见实施例1和2), 但是使用为PDISP/HA B Brisbane (PrL-) 特别设计的修饰的PCR引物, 将编码序列分别克隆进原始的HT和修饰的HT+中, 所述编码序列对应于天然信号肽已被苜蓿蛋白质二硫键异构酶的信号肽替代的具有缺失的蛋白水解环 (PrL-) 的来自流感病毒B/Brisbane/60/2008的HA (PDISP/HA B Brisbane (PrL-; 图12A, SEQ ID NO:43)。与PDISP融合的成熟HA B Brisbane (PrL-) 的氨基酸序列显示于图12B (SEQ ID NO:44) 中。质粒1039和1829的示意图显示于图12C和12D中。

[0260] 实施例9:PDISP/HA B Brisbane (PrL-) +H1 California TMCT的2X35S/CPMV HT (构建体编号1067) 和HT+ (构建体编号1875)

[0261] 利用与构建体1391和1819相同的基于PCR的方法(参见实施例1和2),但是使用为PDISP/HA B Brisbane (PrL-)+H1 California TMCT特别设计的修饰的PCR引物,将嵌合体血凝素编码序列分别克隆进原始的HT和修饰的HT+中,所述嵌合体血凝素编码序列对应于具有缺失的蛋白水解环(PrL-)的来自流感病毒B/Brisbane/60/08的HA的胞外域,其与来自流感病毒A/California/7/2009的H1的跨膜结构域和胞质尾区(TMCT)融合,并且具有苜蓿蛋白质二硫键异构酶的信号肽(PDISP/HA B Brisbane (PrL-)+H1 California TMCT)(图13A,SEQ ID NO:45)。与PDISP融合的成熟HA B Brisbane (PrL-)+H1 California TMCT的氨基酸序列显示于图13B(SEQ ID NO:46)中。质粒1067和1875的示意图显示于图13C和13D中。

[0262] 实施例10:PDISP/HA B Massachusetts (PrL-)的2X35S/CPMV HT(构建体编号2072)和HT+(构建体编号2052)

[0263] 利用与构建体1391和1819相同的基于PCR的方法(参见实施例1和2),但是使用为PDISP/HA B Massachusetts (PrL-)特别设计的修饰的PCR引物,将编码序列分别克隆进原始的HT和修饰的HT+中,所述编码序列对应于天然信号肽已被苜蓿蛋白质二硫键异构酶的信号肽替代的具有缺失的蛋白水解环(PrL-)的来自流感病毒B/Massachusetts/2/2012的HA(PDISP/HA B Massachusetts (PrL-)) (图14A,SEQ ID NO:47)。与PDISP融合的成熟HA B Massachusetts (PrL-)的氨基酸序列显示于图14B(SEQ ID NO:48)中。质粒2072和2052的示意图显示于图14C和14D中。

[0264] 实施例11:PDISP/HA B Massachusetts (PrL-)+H1 California TMCT的2X35S/CPMV HT(构建体编号2074)和HT+(构建体编号2062)

[0265] 利用与构建体1391和1819相同的基于PCR的方法(参见实施例1和2),但是使用为PDISP/HA B Massachusetts (PrL-)+H1 California TMCT特别设计的修饰的PCR引物,将嵌合体血凝素编码序列分别克隆进原始的HT和修饰的HT+中,所述嵌合体血凝素编码序列对应于具有缺失的蛋白水解环(PrL-)的来自流感病毒B/Massachusetts/2/2012的HA的胞外域,其与来自流感病毒A/California/7/2009的H1的跨膜结构域和胞质尾区(TMCT)融合,并且具有苜蓿蛋白质二硫键异构酶的信号肽(PDISP/HA B Massachusetts (PrL-)+H1 California TMCT)(图15A,SEQ ID NO:49)。与PDISP融合的成熟HA B Massachusetts (PrL-)+H1 California TMCT的氨基酸序列显示于图15B(SEQ ID NO:50)中。质粒2074和2062的示意图显示于图15C和15D中。

[0266] 实施例12:HA B Wisconsin (PrL-)的2X35S/CPMV HT(构建体编号1445)和HT+(构建体编号1839)

[0267] 利用与构建体1391和1819相同的基于PCR的方法(参见实施例1和2),但是使用为HA B Wisconsin (PrL-)特别设计的修饰的PCR引物,将编码序列分别克隆进原始的HT和修饰的HT+中,所述编码序列对应于具有自身的天然信号肽并具有缺失的蛋白水解环(PrL-)的来自流感病毒B/Wisconsin/1/2010的HA(HA B Wisconsin (PrL-)) (图16A,SEQ ID NO:51)。具有自身的天然信号肽的HA B Wisconsin (PrL-)的氨基酸序列显示于图16B(SEQ ID NO:52)中。质粒1445和1839的示意图显示于图16C和16D中。

[0268] 实施例13:HA B Wisconsin (PrL-)+H1 California TMCT的2X35S/CPMVHT(构建体编号1454)和HT+(构建体编号1860)

[0269] 利用与构建体1391和1819相同的基于PCR的方法(参见实施例1和2),但是使用为

HA B Wisconsin(PrL-)+H1 California TMCT特别设计的修饰的PCR引物,将嵌合体血凝素编码序列分别克隆进原始的HT和修饰的HT+中,所述嵌合体血凝素编码序列对应于具有缺失的蛋白水解环(PrL-)的来自流感病毒B/Wisconsin/2/2012的HA的胞外域,其与来自流感病毒A/California/7/2009的H1的跨膜结构域和胞质尾区(TMCT)融合,并且具有HA B Wisconsin的天然信号肽(HA B Wisconsin(PrL-)+H1California TMCT)(图17A,SEQ ID NO:53)。HA B Wisconsin(PrL-)+H1California TMCT的氨基酸序列显示于图17B(SEQ ID NO:54)中。质粒1454和1860的示意图显示于图17C和17D中。

[0270] 实施例14:农杆菌转染

[0271] 利用D' Aoust等2008(Plant Biotechnology Journal 6:930-940)所述的方法,通过电穿孔,使农杆菌菌株AGL1转染有DNA构建体。使转染的农杆菌在pH5.6的补充有10mM 2-(N-吗啉)乙磺酸(MES)、20μM乙酰丁香酮、50μg/ml卡那霉素以及25μg/ml的羧苄青霉素的YEB培养基中生长至0.6至1.6的OD₆₀₀。在使用之前将农杆菌悬液离心,并在渗入介质中重悬(10mM MgCl₂和10mM MES,pH 5.6)。

[0272] 制备植物生物物质、接种物和农杆菌渗入物

[0273] 本文使用的术语“生物物质”和“植物物质”意为反映来源于植物的任何物质。生物物质或植物物质可以包含整个植株、组织、细胞或其任何部分。此外,生物物质或植物物质可以包含细胞内植物组分、细胞外植物组分、植物的液体或固体提取物或者它们的组合。此外,生物物质或植物物质可以包含植物,植物细胞,组织,来自植物叶、茎、果实、根或其组合的液体提取物或者它们的组合。植物部分可以包含植物物质或生物物质。

[0274] 在铺满市售的泥炭苔底层的平地上由种子长出本氏烟草植株。允许植物在16/8光周期和25℃白天/20℃夜晚的温度方案下的温室中生长。播种后三周,挑选单个小植株,移植进盆中,并留在温室中在同样的环境条件下生长另外三周。

[0275] 使转染有各构建体的农杆菌在pH5.6的补充有10mM 2-(N-吗啉)乙磺酸(MES)、20μM乙酰丁香酮、50μg/ml卡那霉素和25μg/ml羧苄青霉素的YEB培养基中生长,直至它们达到0.6至1.6的OD₆₀₀。使用之前将农杆菌悬液离心,并在渗入介质(10mM MgCl₂和10mM MES,pH 5.6)中重悬,以及在4℃保存过夜。在渗入当天,以2.5倍培养体积稀释培养物批次,并且允许在使用前升温。将本氏烟草的全植株倒置在密封的不锈钢槽中的细菌悬液中,在20-40托的真空下持续2-min。将植物放回温室中,持续2-6天培育期,直至收获。

[0276] 叶收获和全蛋白提取

[0277] 培育之后,收获植物的地上部分,在-80℃冷冻并破碎成片。通过在3体积冷的50mM Tris pH 8.0、0.15M NaCl、0.1%Triton X-100和1mM苯甲基磺酰氟中均化(Polytron)各冷冻-破碎的植物材料样本,提取总的可溶蛋白。均化之后,将浆料在10,000g,4℃下离心10min,并保留这些澄清的粗提取物(上清液)用于分析。

[0278] 实施例15:血凝分析

[0279] 通过Bradford分析(Bio-Rad,Hercules,CA),使用牛血清白蛋白作为参考标准,测定澄清的粗提取物中的总蛋白含量。

[0280] 血凝分析基于Nayak和Reichl(2004)所述的方法。简言之,在含有100μL PBS的V形底96孔微量滴定板中对测试样本(100μL)进行连续二倍稀释,每孔剩余100μL稀释样本。向各孔添加一百微升0.25%火鸡红细胞悬液(Bio Link Inc.,Syracuse,NY;对于所有B株、

H1、H5和H7)或者0.5%的豚鼠红细胞悬液(对于H3),并将板在室温孵育2h。将显示完全血凝的最高稀释度的倒数记作HA活性。

[0281] 在此将所有引用通过引用并入。

[0282] 针对一个或多个实施方案,描述了本发明。然而,对本领域技术人员而言显而易见的是,在不脱离权利要求所限定的本发明的范围的情况下,可以进行多种改变和修改。

- [0001] 序列表
- [0002] <110> 莫迪卡戈公司
- [0003] <120> 修饰的CPMV增强子元件
- [0004] <130> V86817W0
- [0005] <150> US 61/971,274
- [0006] <151> 2014-03-27
- [0007] <150> PCT/CA04/050326
- [0008] <151> 2014-03-28
- [0009] <160> 59
- [0010] <170> PatentIn version 3.5
- [0011] <210> 1
- [0012] <211> 534
- [0013] <212> DNA
- [0014] <213> 人工序列
- [0015] <220>
- [0016] <223> 包含CPMV HT+的表达增强子
- [0017] <400> 1
- [0018] tattaaaatc ttaatagggt ttgataaaag cgaacgtggg gaaacccgaa ccaaacccttc 60
- [0019] ttctaaactc tctctcatct ctcttaaagc aaacttctct cttgtctttc ttgcgtgagc 120
- [0020] gatcttcaac gttgtcagat cgtgcttcgg caccagtaca acgttttctt tcaactgaagc 180
- [0021] gaaatcaaag atctctttgt ggacacgtag tgcggcgcca ttaaataacg tgtacttgct 240
- [0022] ctattcttgt cgggtgtggtc ttgggaaaag aaagcttgct ggaggctgct gttcagcccc 300
- [0023] atacattact tgttacgatt ctgctgactt tcggcggtg caatatctct acttctgctt 360
- [0024] gacgaggtat tgttgcctgt acttctttct tcttcttctt gctgattggt tctataagaa 420
- [0025] atctagtatt ttctttgaaa cagagttttc ccgtggtttt cgaacttgga gaaagattgt 480
- [0026] taagcttctg tatattctgc ccaaatttgt tcgggcccaa taccgcggag aaaa 534
- [0027] <210> 2
- [0028] <211> 517
- [0029] <212> DNA
- [0030] <213> 人工序列
- [0031] <220>
- [0032] <223> 表达增强子CPMV HT
- [0033] <400> 2
- [0034] tattaaaatc ttaatagggt ttgataaaag cgaacgtggg gaaacccgaa ccaaacccttc 60
- [0035] ttctaaactc tctctcatct ctcttaaagc aaacttctct cttgtctttc ttgcgtgagc 120
- [0036] gatcttcaac gttgtcagat cgtgcttcgg caccagtaca acgttttctt tcaactgaagc 180
- [0037] gaaatcaaag atctctttgt ggacacgtag tgcggcgcca ttaaataacg tgtacttgct 240
- [0038] ctattcttgt cgggtgtggtc ttgggaaaag aaagcttgct ggaggctgct gttcagcccc 300

[0039] atacattact tgttacgatt ctgctgactt tcggcgggtg caatatctct acttctgctt 360
 [0040] gacgaggtat tgttgcctgt acttctttct tcttcttctt gctgattggg tctataagaa 420
 [0041] atctagtatt ttctttgaaa cagagttttc ccgtgggttt cgaacttgga gaaagattgt 480
 [0042] taagcttctg tatattctgc ccaaatttgt cgggccc 517
 [0043] <210> 3
 [0044] <211> 534
 [0045] <212> DNA
 [0046] <213> 人工序列
 [0047] <220>
 [0048] <223> CPMV HT+[WT115]
 [0049] <400> 3
 [0050] tattaaaatc ttaatagggt ttgataaaag cgaacgtggg gaaacccgaa ccaaaccctt 60
 [0051] ttctaaactc tctctcatct ctcttaaagc aaacttctct cttgtctttc ttgcatgagc 120
 [0052] gatcttcaac gttgtcagat cgtgcttcgg caccagtaca acgttttctt tcaactgaagc 180
 [0053] gaaatcaaag atctctttgt ggacacgtag tgcggcgcca ttaaataacg tgtacttgct 240
 [0054] ctattcttgt cgggtgtggtc ttgggaaaag aaagcttgct ggaggctgct gttcagcccc 300
 [0055] atacattact tgttacgatt ctgctgactt tcggcgggtg caatatctct acttctgctt 360
 [0056] gacgaggtat tgttgcctgt acttctttct tcttcttctt gctgattggg tctataagaa 420
 [0057] atctagtatt ttctttgaaa cagagttttc ccgtgggttt cgaacttgga gaaagattgt 480
 [0058] taagcttctg tatattctgc ccaaatttgt tcgggcccac taccgcggag aaaa 534
 [0059] <210> 4
 [0060] <211> 534
 [0061] <212> DNA
 [0062] <213> 人工序列
 [0063] <220>
 [0064] <223> CPMV HT+ (植物kozak共有序列)
 [0065] <220>
 [0066] <221> misc_feature
 [0067] <222> (529) .. (529)
 [0068] <220>
 [0069] <221> misc_feature
 [0070] <222> (529) .. (529)
 [0071] <223> a可能存在或者不存在
 [0072] <400> 4
 [0073] tattaaaatc ttaatagggt ttgataaaag cgaacgtggg gaaacccgaa ccaaaccctt 60
 [0074] ttctaaactc tctctcatct ctcttaaagc aaacttctct cttgtctttc ttgctgagc 120
 [0075] gatcttcaac gttgtcagat cgtgcttcgg caccagtaca acgttttctt tcaactgaagc 180
 [0076] gaaatcaaag atctctttgt ggacacgtag tgcggcgcca ttaaataacg tgtacttgct 240
 [0077] ctattcttgt cgggtgtggtc ttgggaaaag aaagcttgct ggaggctgct gttcagcccc 300

[0078] atacattact tgttacgatt ctgctgactt tcggcgggtg caatatctct acttctgctt 360
[0079] gacgaggtat tgttgcctgt acttctttct tcttcttctt gctgattggg tctataagaa 420
[0080] atctagtatt ttctttgaaa cagagttttc ccgtggtttt cgaacttgga gaaagattgt 480
[0081] taagcttctg tatattctgc ccaaatttgt tcgggcccac taccgcggna rrna 534
[0082] <210> 5
[0083] <400> 5
[0084] 000
[0085] <210> 6
[0086] <400> 6
[0087] 000
[0088] <210> 7
[0089] <400> 7
[0090] 000
[0091] <210> 8
[0092] <400> 8
[0093] 000
[0094] <210> 9
[0095] <400> 9
[0096] 000
[0097] <210> 10
[0098] <400> 10
[0099] 000
[0100] <210> 11
[0101] <400> 11
[0102] 000
[0103] <210> 12
[0104] <400> 12
[0105] 000
[0106] <210> 13
[0107] <400> 13
[0108] 000
[0109] <210> 14
[0110] <400> 14
[0111] 000
[0112] <210> 15
[0113] <400> 15
[0114] 000
[0115] <210> 16
[0116] <400> 16

[0117]	000
[0118]	<210> 17
[0119]	<400> 17
[0120]	000
[0121]	<210> 18
[0122]	<211> 48
[0123]	<212> DNA
[0124]	<213> 人工序列
[0125]	<220>
[0126]	<223> IF-PDI.S1+3c
[0127]	<400> 18
[0128]	aaatttgtcg ggcccatggc gaaaaacgtt gcgattttcg gcttattg 48
[0129]	<210> 19
[0130]	<211> 51
[0131]	<212> DNA
[0132]	<213> 人工序列
[0133]	<220>
[0134]	<223> IF-H3V36111.s1-4r
[0135]	<400> 19
[0136]	actaaagaaa ataggccttc aaatgcaa atgtgcaccta atgttgcct t 51
[0137]	<210> 20
[0138]	<211> 1725
[0139]	<212> DNA
[0140]	<213> 人工序列
[0141]	<220>
[0142]	<223> PDISP/H3 Victoria的核苷酸序列
[0143]	<400> 20
[0144]	atggcgaaaa acgttgcgat tttcggctta ttgttttctc ttcttgtgtt ggttccttct 60
[0145]	cagatcttcg cccaaaaact tcctggaaat gacaacagca cggcaacgct gtgccttggg 120
[0146]	caccatgcag taccaaacgg aacgatagtg aaaacaatca cgaatgacca aattgaagtt 180
[0147]	actaatgcta ctgagctggc tcagaattcc tcaatagggt aaatatgcga cagtcctcat 240
[0148]	cagatccttg atggagaaaa ctgcacacta atagatgctc tattgggaga ccctcagtgt 300
[0149]	gatggcttcc aaaataagaa atgggacctt tttgttgaac gaagcaaagc ctacagcaac 360
[0150]	tgttaccett atgatgtgcc ggattatgcc tccttaggt cactagtgc ctcatecggc 420
[0151]	acactggagt ttaacaatga aagcttcaat tggactggag tcaactcaaaa cggaacaagt 480
[0152]	tctgcttgea taaggagatc taataatagt ttcttttagta gattaaattg gttgaccac 540
[0153]	ttaaacttca aatacccagc attgaacgtg actatgcaa acaatgaaca atttgacaaa 600
[0154]	ttgtacattt ggggggttca ccacccgggt acggacaagg accaaatctt cctgtatgct 660
[0155]	caatcatcag gaagaatcac agtatctacc aaaagaagcc aacaagctgt aatcccgaat 720

[0156]	atcggatcta gacccagaat aaggaatatc cctagcagaa taagcatcta ttggacaata	780
[0157]	gtaaaaccgg gagacatact tttgattaac agcacaggga atctaattgc tcctaggggt	840
[0158]	tacttcaaaa tacgaagtgg gaaaagctca ataatgagat cagatgcacc cattggcaaa	900
[0159]	tgcaattctg aatgcatcac tccaaatgga agcattccca atgacaaacc attccaaaat	960
[0160]	gtaaacagga tcacatacgg ggcctgtccc agatatgtta agcaaagcac tctgaaattg	1020
[0161]	gcaacaggaa tgcgaaatgt accagagaaa caaactagag gcatatttgg cgcaatagcg	1080
[0162]	ggtttcatag aaaatggttg ggagggaatg gtggatggtt ggtacggttt caggcatcaa	1140
[0163]	aattctgagg gaagaggaca agcagcagat ctcaaaagca ctcaagcagc aatcgatcaa	1200
[0164]	atcaatggga agctgaatcg attgatcggg aaaaccaacg agaaattcca tcagattgaa	1260
[0165]	aaagaattct cagaagtcca agggagaatt caggacctg agaaatatgt tgaggacact	1320
[0166]	aaaatagatc tctggtcata caacgcggag cttcttgttg ccctggagaa ccaacataca	1380
[0167]	attgatctaa ctgactcaga aatgaacaaa ctgtttgaaa aaacaaagaa gcaactaagg	1440
[0168]	gaaaatgctg aggatatggg caatggttgt ttcaaaatat accacaaatg tgacaatgcc	1500
[0169]	tgcataggat caatcagaaa tggaacttat gaccacgatg tatacagaga tgaagcatta	1560
[0170]	aacaaccggt tccagatcaa gggagttgag ctgaagtcag ggtacaaaga ttggatccta	1620
[0171]	tggatttcct ttgccatata atgttttttg ctttgtgttg ctttgttggg gttcatcatg	1680
[0172]	tgggcctgcc aaaagggcaa cattaggtgc aacatttgca tttga	1725
[0173]	<210> 21	
[0174]	<211> 4903	
[0175]	<212> DNA	
[0176]	<213> 人工序列	
[0177]	<220>	
[0178]	<223> 构建体1191的核苷酸序列	
[0179]	<400> 21	
[0180]	tggcaggata tatttgtgtg taaacaaatt gacgcttaga caacttaata acacattgcg	60
[0181]	gacgttttta atgtactgaa ttaacgccga atcccgggct ggtatattta tatgttgtca	120
[0182]	aataactcaa aaaccataaa agttttaagtt agcaagtgtg tacattttta cttgaacaaa	180
[0183]	aatattcacc tactactggt ataaatcatt attaaacatt agagtaaaga aatatggatg	240
[0184]	ataagaacaa gagtagtgat attttgacaa caattttgtt gcaacatttg agaaaaat	300
[0185]	gttggttctc cttttcattg gtcaaaaaca atagagagag aaaaaggaag agggagaata	360
[0186]	aaaacataat gtgagtatga gagagaaagt tgtacaaaag ttgtacaaa atagttgtac	420
[0187]	aaatatcatt gaggaatttg acaaaagcta cacaaataag ggttaattgc tgtaaataaa	480
[0188]	taaggatgac gcattagaga gatgtaccat tagagaattt ttggcaagtc attaaaaaga	540
[0189]	agaataaat tatttttaaa attaaaagtt gagtcatttg attaaacatg tgattattta	600
[0190]	atgaattgat gaaagagttg gattaaagtt gtattagtaa ttagaatttg gtgtcaaatt	660
[0191]	taatttgaca tttgatcttt tcctatatat tgccccatag agtcagttaa ctcatTTTTA	720
[0192]	tatttcatag atcaaataag agaaataacg gtatattaat ccctccaaaa aaaaaaacg	780
[0193]	gtatatttac taaaaaatct aagccacgta ggaggataac aggatccccg taggaggata	840
[0194]	acatccaatc caaccaatca caacaatcct gatgagataa cccactttaa gcccacgcat	900

[0195]	ctgtggcaca tctacattat ctaaadcaca cattcttcca cacatctgag ccacacaaaa	960
[0196]	accaatccac atctttatca cccattctat aaaaaatcac actttgtgag tctacacttt	1020
[0197]	gattcccttc aaacacatac aaagagaaga gactaattaa ttaattaatc atcttgagag	1080
[0198]	aaaatggaac gagctataca aggaaacgac gctaggggaa aagctaacag tgaacgttgg	1140
[0199]	gatggaggat caggaggtag cacttctccc ttcaaacttc ctgacgaaag tccgagttgg	1200
[0200]	actgagtggc ggctacataa cgatgagacg aattcgaatc aagataatcc ccttggtttc	1260
[0201]	aaggaaagct ggggtttcgg gaaagttgta ttttaagagat atctcagata cgacaggacg	1320
[0202]	gaagcttcac tgcacagagt ccttgatctt tggacgggag attcggttaa ctatgcagca	1380
[0203]	tctcgatttt tcggtttcga ccagatcgga tgtacctata gtattcgggt tcgaggagtt	1440
[0204]	agtatcccg tttctggagg gtgcggaact cttcagcatc tctgtgagat ggcaattcgg	1500
[0205]	tctaagcaag aactgctaca gcttgcccca atcgaagtgg aaagtaatgt atcaagagga	1560
[0206]	tgccctgaag gtactcaaac cttcgaaaaa gaaagcgagt aagttaaaat gcttcttcgt	1620
[0207]	ctcctattta taatatggtt tggtattgtt aattttgttc ttgtagaaga gcttaattaa	1680
[0208]	tcgttgttgt tatgaaatac tatttgtatg agatgaactg gtgtaatgta attcatttac	1740
[0209]	ataagtggag tcagaatcag aatgtttcct ccataactaa ctagacatga agacctgccg	1800
[0210]	cgtacaattg tcttatattt gaacaactaa aattgaacat cttttgccac aactttataa	1860
[0211]	gtggttaata tagctcaaat atatggtcaa gttcaataga ttaataatgg aaatatcagt	1920
[0212]	tatcgaaatt cattaacaat caacttaacg ttattaacta ctaattttat atcatccct	1980
[0213]	ttgataaatg atagtacacc aattaggaag gagcatgctc gcctaggaga ttgtcgtttc	2040
[0214]	ccgccttcag tttgcaagct gctctagccg tgtagccaat acgcaaaccg cctctccccg	2100
[0215]	cgcggttgga attactagcg cgtgtcgaca agcttgcatg ccggtcaaca tgggtggagca	2160
[0216]	cgacacactt gtctactcca aaaatatcaa agatacagtc tcagaagacc aaagggcaat	2220
[0217]	tgagactttt caacaaaggg taatatccgg aaacctcctc ggattccatt gccagctat	2280
[0218]	ctgtcacttt attgtgaaga tagtggaaaa ggaaggtggc tcctacaaat gccatcattg	2340
[0219]	cgataaagga aaggccatcg ttgaagatgc ctctgccgac agtggtccca aagatggacc	2400
[0220]	cccaccacg aggagcatcg tggaaaaaga agacgttcca accacgtctt caaagcaagt	2460
[0221]	ggattgatgt gataacatgg tggagcacga cacactgtc tactccaaaa atatcaaaga	2520
[0222]	tacagtctca gaagaccaaa gggcaattga gacttttcaa caaagggtaa tatccgaaa	2580
[0223]	cctcctcgga ttccattgcc cagctatctg tcactttatt gtgaagatag tggaaaagga	2640
[0224]	aggtggctcc tacaatgcc atcattgcga taaaggaaag gccatcgttg aagatgcctc	2700
[0225]	tgccgacagt ggtcccaaaag atggaccccc acccagagg agcatcgttg aaaaagaaga	2760
[0226]	cgttccaacc acgtcttcaa agcaagtgga ttgatgtgat atctccactg acgtaaggga	2820
[0227]	tgacgcacaa tcccactatc cttcgcaaga ccttctctt atataaggaa gttcatttca	2880
[0228]	tttgagagag tattaaaatc ttaatagggt ttgataaaag cgaacgtggg gaaacccgaa	2940
[0229]	ccaaaccttc ttctaaactc tctctcatct ctcttaaagc aaacttctct cttgtctttc	3000
[0230]	ttgcgtgagc gatcttcaac gttgtcagat cgtgttcgg caccagtaca acgttttctt	3060
[0231]	tactgaagc gaaatcaaag atctctttgt ggacacgtag tgcggcgcca ttaaataacg	3120
[0232]	tgtacttgtc ctattcttgt cgggtgtggtc ttgggaaaag aaagcttgct ggaggctgct	3180
[0233]	gttcagcccc atacattact tgttacgatt ctgctgactt tcggcgggtg caatatctct	3240

[0234]	acttctgctt gacgaggtat tgttgccctgt acttctttct tcttcttctt gctgattggt	3300
[0235]	tctataagaa atctagtatt ttctttgaaa cagagtttcc ccgtggtttt cgaacttgga	3360
[0236]	gaaagattgt taagcttctg tatattctgc ccaaatttgt cgggcccgcg gatggcgaaa	3420
[0237]	aacgttgcca ttttcggctt attgttttct cttcttgtgt tggttccttc tcagatcttc	3480
[0238]	gcctgcaggc tcctcagcca aaacgacacc cccatctgtc tatccactgg cccctggatc	3540
[0239]	tgctgcccac actaactcca tggtagacct gggatgcctg gtcaagggtt atttccctga	3600
[0240]	gccagtgaca gtgacctgga actctggatc cctgtccagc ggtgtgcaca ccttcccagc	3660
[0241]	tgtcctgcag tctgacctct acactctgag cagctcagtg actgtccctt ccagcacctg	3720
[0242]	gcccagcgag accgtcacct gcaacgttgc ccacccggcc agcagcacca aggtggacaa	3780
[0243]	gaaaattgtg cccagggatt gtggttgtaa gccttgcata tgtacagtcc cagaagtatc	3840
[0244]	atctgtcttc atcttcccc caaagcccaa ggatgtgtc accattactc tgactcctaa	3900
[0245]	ggtcacgtgt gttgtggtag acatcagcaa ggatgatccc gaggtccagt tcagctggtt	3960
[0246]	tgtagatgat gtggagggtc acacagctca gacgcaacc cgggaggagc agttcaacag	4020
[0247]	cactttccgc tcagtcaagt aacttcccat catgcaccag gactggctca atggcaagga	4080
[0248]	gcgatcgctc accatcacca tcaccatcac catcaccatt aaaggcctat tttctttagt	4140
[0249]	ttgaatttac tgttattcgg tgtgcatttc tatgtttggt gagcggtttt ctgtgctcag	4200
[0250]	agtgtgttta ttttatgtaa ttttaatttct ttgtgagctc ctgttttagca ggtcgtccct	4260
[0251]	tcagcaagga cacaaaaaga ttttaatttt attaaaaaaa aaaaaaaaaa agaccgggaa	4320
[0252]	ttcgatatca agcttatcga cctgcagatc gttcaaacat ttggcaataa agtttcttaa	4380
[0253]	gattgaatcc tgttgccggt cttgcgatga ttatcatata atttctgttg aattacgtta	4440
[0254]	agcatgtaat aattaacatg taatgcatga cgttatttat gagatgggtt tttatgatta	4500
[0255]	gagtcccga attatacatt taatacgcca tagaaaacaa aatatagcgc gcaaactagg	4560
[0256]	ataaattatc gcgcgcggtg tcacttatgt tactagatct ctagagtctc aagcttggcg	4620
[0257]	cgcccacgtg actagtggca ctggccgtcg ttttacaacg tcgtgactgg gaaaaccctg	4680
[0258]	gcgttaccca acttaatcgc cttgcagcac atccccctt cgccagctgg cgtaatagcg	4740
[0259]	aagaggcccg caccgatcgc ccttcccaac agttgcgcag cctgaatggc gaatgctaga	4800
[0260]	gcagcttgag cttggatcag attgtcgttt cccgccttca gtttaacta tcagtgtttg	4860
[0261]	acaggatata ttggcgggta aacctaagag aaaagagcgt tta	4903
[0262]	<210> 22	
[0263]	<211> 3465	
[0264]	<212> DNA	
[0265]	<213> 人工序列	
[0266]	<220>	
[0267]	<223> 表达盒1391的核苷酸序列	
[0268]	<400> 22	
[0269]	gtcaacatgg tggagcacga cacacttgct tactccaaaa atatcaaaga tacagtctca	60
[0270]	gaagacccaa gggcaattga gacttttcaa caaagggtaa tatccggaaa cctcctcgga	120
[0271]	ttccattgcc cagctatctg tcactttatt gtgaagatag tggaaaagga aggtggctcc	180
[0272]	tacaaatgcc atcattgcga taaaggaaag gccatcgttg aagatgcctc tgccgacagt	240

[0273]	ggtcccaaag atggaccccc acccacgagg agcatcgtgg aaaaagaaga cgttccaacc	300
[0274]	acgtcttcaa agcaagtgga ttgatgtgat aacatgggtg agcacgacac acttgtctac	360
[0275]	tccaaaaata tcaaagatac agtctcagaa gaccaaaggg caattgagac ttttcaacaa	420
[0276]	agggtaatat ccggaaacct cctcggattc cattgcccag ctatctgtca ctttattgtg	480
[0277]	aagatagtgg aaaaggaagg tggctcctac aaatgccatc attgcgataa aggaaaggcc	540
[0278]	atcgttgaag atgcctctgc cgacagtggg cccaaagatg gacccccacc cacgaggagc	600
[0279]	atcgtggaaa aagaagacgt tccaaccacg tcttcaaagc aagtggattg atgtgatatc	660
[0280]	tccactgacg taagggatga cgcacaatcc cactatcctt cgcaagaccc ttcctctata	720
[0281]	taaggaagtt catttcattt ggagaggtat taaaatctta ataggttttg ataaaagcga	780
[0282]	acgtggggaa acccgaacca aaccttcttc taaactctct ctcatctctc ttaaagcaaa	840
[0283]	cttctctctt gtctttcttg cgtgagcgat cttcaacgtt gtcagatcgt gcttcggcac	900
[0284]	cagtacaacg ttttctttca ctgaagegaa atcaaagatc tctttgtgga cacgtagtgc	960
[0285]	ggcgccatta aataacgtgt acttgtctca ttcttgcgg tgtggtcttg ggaaaagaaa	1020
[0286]	gcttgctgga ggctgctgtt cagccccata cttacttgt tacgattctg ctgactttcg	1080
[0287]	gcgggtgcaa tatctctact tctgcttgac gaggtattgt tgctgtact tctttcttct	1140
[0288]	tcttcttgct gattggttct ataagaaatc tagtattttc tttgaaacag agttttcccg	1200
[0289]	tggttttcga acttgagaa agattgttaa gcttctgtat attctgcca aatttgtcgg	1260
[0290]	gcccattggcg aaaaacgttg cgattttcgg cttattgttt tctcttcttg tgttggttcc	1320
[0291]	ttctcagatc ttcgccccaa aacttcctgg aaatgacaac agcacggcaa cgctgtgcct	1380
[0292]	tgggcacat gcagtaccaa acggaacgat agtgaaaaca atcacgaatg accaaattga	1440
[0293]	agttactaat gctactgagc tggttcagaa ttctcaata ggtgaaatat gcgacagtcc	1500
[0294]	tcatcagatc cttgatggag aaaactgcac actaatagat gctctattgg gagaccctca	1560
[0295]	gtgtgatggc ttccaaaata agaaatggga ctttttgtt gaacgaagca aagcctacag	1620
[0296]	caactgttac cttatgatg tgccggatta tgctccctt aggtcactag ttgcctcatc	1680
[0297]	cggcacactg gagtttaaca atgaaagctt caattggact ggagtcactc aaaacggaac	1740
[0298]	aagttctgct tgcataagga gatctaataa tagtttcttt agtagattaa attggttgac	1800
[0299]	ccacttaaac ttcaaatacc cagcattgaa cgtgactatg ccaaacaatg aacaatttga	1860
[0300]	caaattgtac atttgggggg ttaccaccc ggttacggac aaggaccaa tcttctgtta	1920
[0301]	tgctcaatca tcaggaagaa tcacagtatc taccaaaaga agccaacaag ctgtaatccc	1980
[0302]	gaatatcgga tctagacca gaataaggaa tatccctagc agaataagca tctattggac	2040
[0303]	aatagtaaaa ccgggagaca tacttttgat taacagcaca gggaatctaa ttgctcctag	2100
[0304]	gggttacttc aaaatacgaa gtgggaaaag ctcaataatg agatcagatg caccatttgg	2160
[0305]	caaatgcaat tctgaatgca tactccaaa tggaagcatt cccaatgaca aaccattcca	2220
[0306]	aaatgtaaac aggatcacat acggggcctg tcccagatat gttaagcaaa gactctgaa	2280
[0307]	attggcaaca ggaatgcgaa atgtaccaga gaaacaaact agaggcatat ttggcgcaat	2340
[0308]	agcgggtttc atagaaaatg gttgggaggg aatggtggat ggttggtacg gtttcaggca	2400
[0309]	tcaaaattct gaggaagag gacaagcagc agatctcaaa agcactcaag cagcaatcga	2460
[0310]	tcaaatcaat gggaagctga atcgattgat cgggaaaacc aacgagaaat tccatcagat	2520
[0311]	tgaaaaagaa ttctcagaag tcgaaggag aattcaggac cttgagaaat atgttgagga	2580

[0312]	cactaaaata gatctctggt catacaacgc ggagcttctt gttgccctgg agaaccaaca	2640
[0313]	tacaattgat ctaactgact cagaaatgaa caaactgttt gaaaaaacaagaagcaact	2700
[0314]	aagggaataat gctgaggata tgggcaatgg ttgtttcaaa atataccaca aatgtgacaa	2760
[0315]	tgctgcata ggatcaatca gaaatggaac ttatgaccac gatgtataca gagatgaagc	2820
[0316]	attaaacaac cggttccaga tcaaggaggt tgagctgaag tcagggtaca aagattggat	2880
[0317]	cctatggatt tcctttgcca tatcatgttt tttgctttgt gttgctttgt tggggttcat	2940
[0318]	catgtggggc tgccaaaagg gcaacattag gtgcaacatt tgcatttgaa ggcctatatt	3000
[0319]	ctttagtttg aattttactgt tttcggtgt gcatttctat gtttggtgag cggttttctg	3060
[0320]	tgctcagagt gtgtttatatt tatgtaattt aattttcttg tgagctcctg tttagcaggt	3120
[0321]	cgtcccttca gcaaggacac aaaaagattt taattttatt aaaaaaaaaa aaaaaaaga	3180
[0322]	ccgggaattc gatatcaagc ttatcgacct gcagatcggt caaacatttg gcaataaagt	3240
[0323]	ttcttaagat tgaatcctgt tgccggtctt gcgatgatta tcatataatt tctgttgaat	3300
[0324]	tacgttaagc atgtaataat taacatgtaa tgcatgacgt tatttatgag atgggttttt	3360
[0325]	atgattagag tcccgaatt atacatttaa tacgcgatag aaaacaaaat atagcgcgca	3420
[0326]	aactaggata aattatcgcg cgcggtgtca tctatgttac tagat	3465
[0327]	<210> 23	
[0328]	<211> 574	
[0329]	<212> PRT	
[0330]	<213> 人工序列	
[0331]	<220>	
[0332]	<223> PDISP/H3 Victoria的氨基酸序列	
[0333]	<400> 23	
[0334]	Met Ala Lys Asn Val Ala Ile Phe Gly Leu Leu Phe Ser Leu Leu Val	
[0335]	1 5 10 15	
[0336]	Leu Val Pro Ser Gln Ile Phe Ala Gln Lys Leu Pro Gly Asn Asp Asn	
[0337]	20 25 30	
[0338]	Ser Thr Ala Thr Leu Cys Leu Gly His His Ala Val Pro Asn Gly Thr	
[0339]	35 40 45	
[0340]	Ile Val Lys Thr Ile Thr Asn Asp Gln Ile Glu Val Thr Asn Ala Thr	
[0341]	50 55 60	
[0342]	Glu Leu Val Gln Asn Ser Ile Gly Glu Ile Cys Asp Ser Pro His	
[0343]	65 70 75 80	
[0344]	Gln Ile Leu Asp Gly Glu Asn Cys Thr Leu Ile Asp Ala Leu Leu Gly	
[0345]	85 90 95	
[0346]	Asp Pro Gln Cys Asp Gly Phe Gln Asn Lys Lys Trp Asp Leu Phe Val	
[0347]	100 105 110	
[0348]	Glu Arg Ser Lys Ala Tyr Ser Asn Cys Tyr Pro Tyr Asp Val Pro Asp	
[0349]	115 120 125	
[0350]	Tyr Ala Ser Leu Arg Ser Leu Val Ala Ser Ser Gly Thr Leu Glu Phe	

[0351]	130	135	140
[0352]	Asn Asn Glu Ser Phe Asn Trp Thr Gly Val Thr Gln Asn Gly Thr Ser		
[0353]	145	150	155
[0354]	Ser Ala Cys Ile Arg Arg Ser Asn Asn Ser Phe Phe Ser Arg Leu Asn		
[0355]	165	170	175
[0356]	Trp Leu Thr His Leu Asn Phe Lys Tyr Pro Ala Leu Asn Val Thr Met		
[0357]	180	185	190
[0358]	Pro Asn Asn Glu Gln Phe Asp Lys Leu Tyr Ile Trp Gly Val His His		
[0359]	195	200	205
[0360]	Pro Gly Thr Asp Lys Asp Gln Ile Phe Leu Tyr Ala Gln Ser Ser Gly		
[0361]	210	215	220
[0362]	Arg Ile Thr Val Ser Thr Lys Arg Ser Gln Gln Ala Val Ile Pro Asn		
[0363]	225	230	235
[0364]	Ile Gly Ser Arg Pro Arg Ile Arg Asn Ile Pro Ser Arg Ile Ser Ile		
[0365]	245	250	255
[0366]	Tyr Trp Thr Ile Val Lys Pro Gly Asp Ile Leu Leu Ile Asn Ser Thr		
[0367]	260	265	270
[0368]	Gly Asn Leu Ile Ala Pro Arg Gly Tyr Phe Lys Ile Arg Ser Gly Lys		
[0369]	275	280	285
[0370]	Ser Ser Ile Met Arg Ser Asp Ala Pro Ile Gly Lys Cys Asn Ser Glu		
[0371]	290	295	300
[0372]	Cys Ile Thr Pro Asn Gly Ser Ile Pro Asn Asp Lys Pro Phe Gln Asn		
[0373]	305	310	315
[0374]	Val Asn Arg Ile Thr Tyr Gly Ala Cys Pro Arg Tyr Val Lys Gln Ser		
[0375]	325	330	335
[0376]	Thr Leu Lys Leu Ala Thr Gly Met Arg Asn Val Pro Glu Lys Gln Thr		
[0377]	340	345	350
[0378]	Arg Gly Ile Phe Gly Ala Ile Ala Gly Phe Ile Glu Asn Gly Trp Glu		
[0379]	355	360	365
[0380]	Gly Met Val Asp Gly Trp Tyr Gly Phe Arg His Gln Asn Ser Glu Gly		
[0381]	370	375	380
[0382]	Arg Gly Gln Ala Ala Asp Leu Lys Ser Thr Gln Ala Ala Ile Asp Gln		
[0383]	385	390	395
[0384]	Ile Asn Gly Lys Leu Asn Arg Leu Ile Gly Lys Thr Asn Glu Lys Phe		
[0385]	405	410	415
[0386]	His Gln Ile Glu Lys Glu Phe Ser Glu Val Glu Gly Arg Ile Gln Asp		
[0387]	420	425	430
[0388]	Leu Glu Lys Tyr Val Glu Asp Thr Lys Ile Asp Leu Trp Ser Tyr Asn		
[0389]	435	440	445

[0390]	Ala Glu Leu Leu Val Ala Leu Glu Asn Gln His Thr Ile Asp Leu Thr
[0391]	450 455 460
[0392]	Asp Ser Glu Met Asn Lys Leu Phe Glu Lys Thr Lys Lys Gln Leu Arg
[0393]	465 470 475 480
[0394]	Glu Asn Ala Glu Asp Met Gly Asn Gly Cys Phe Lys Ile Tyr His Lys
[0395]	485 490 495
[0396]	Cys Asp Asn Ala Cys Ile Gly Ser Ile Arg Asn Gly Thr Tyr Asp His
[0397]	500 505 510
[0398]	Asp Val Tyr Arg Asp Glu Ala Leu Asn Asn Arg Phe Gln Ile Lys Gly
[0399]	515 520 525
[0400]	Val Glu Leu Lys Ser Gly Tyr Lys Asp Trp Ile Leu Trp Ile Ser Phe
[0401]	530 535 540
[0402]	Ala Ile Ser Cys Phe Leu Leu Cys Val Ala Leu Leu Gly Phe Ile Met
[0403]	545 550 555 560
[0404]	Trp Ala Cys Gln Lys Gly Asn Ile Arg Cys Asn Ile Cys Ile
[0405]	565 570
[0406]	<210> 24
[0407]	<211> 48
[0408]	<212> DNA
[0409]	<213> 人工序列
[0410]	<220>
[0411]	<223> IF (SacII) -Kozac_PDI.c
[0412]	<400> 24
[0413]	gtcgggccca ataccgcgga gaaaatggcg aaaaacgttg cgatttttc 48
[0414]	<210> 25
[0415]	<211> 4914
[0416]	<212> DNA
[0417]	<213> 人工序列
[0418]	<220>
[0419]	<223> 构建体2181的核苷酸序列
[0420]	<400> 25
[0421]	tggcaggata tatttggtg taaacaaatt gacgcttaga caacttaata acacattgcg 60
[0422]	gacgttttta atgtactgaa ttaacgccga atcccgggct ggtatatatta tatgttgtca 120
[0423]	aataactcaa aaaccataaa agtttaagtt agcaagtgtg tacattttta cttgaacaaa 180
[0424]	aatattcacc tactactggt ataatcatt attaaacatt agagtaaaga aatatggatg 240
[0425]	ataagaacaa gagtagtgat attttgacaa caattttgtt gcaacatttg agaaaatttt 300
[0426]	gttgttctct cttttcattg gtcaaaaaca atagagagag aaaaaggaag agggagaata 360
[0427]	aaaacataat gtgagtatga gagagaaagt tgtacaaaag ttgtacacaa atagttgtac 420
[0428]	aaatatcatt gaggaatttg acaaaagcta cacaaataag ggtaattgc tgtaaataaa 480

[0429]	taaggatgac gcattagaga gatgtacat tagagaattt ttggcaagtc attaaaaaga	540
[0430]	aagaataaat tattttttaa attaaaagtt gagtcatttg attaaacatg tgattattta	600
[0431]	atgaattgat gaaagagttg gattaaagtt gtattagtaa ttagaatttg gtgtcaaatt	660
[0432]	taatttgaca ttgatcttt tcctatatat tgcccatag agtcagttaa ctcatTTTTA	720
[0433]	tatttcatag atcaaataag agaaataacg gtatattaat ccctccaaaa aaaaaaacg	780
[0434]	gtatatttac taaaaaatct aagccacgta ggaggataac aggatccccg taggaggata	840
[0435]	acatccaatc caaccaatca caacaatcct gatgagataa cccactttta gcccacgcat	900
[0436]	ctgtggcaca tctacattat ctaaatacaca cattcttcca cacatctgag ccacacaaaa	960
[0437]	accaatccac atctttatca cccattctat aaaaaatcac actttgtgag tctacacttt	1020
[0438]	gattcccttc aaacacatac aaagagaaga gactaattaa ttaattaaatc atcttgagag	1080
[0439]	aaaatggaac gagctataca aggaaacgac gctagggaaac aagctaacag tgaacgttgg	1140
[0440]	gatggaggat caggaggtag cacttctccc ttcaaacttc ctgacgaaag tccgagttgg	1200
[0441]	actgagtggc ggctacataa cgatgagacg aattcgaatc aagataatcc ccttggtttc	1260
[0442]	aaggaaagct ggggtttcgg gaaagttgta tttaagagat atctcagata cgacaggacg	1320
[0443]	gaagcttcac tgcacagagt ccttggatct tggacgggag attcggttaa ctatgcagca	1380
[0444]	tctcgatttt tcggtttcga ccagatcgga tgtacctata gtattcggtt tcgaggagtt	1440
[0445]	agtatcccg tttctggagg gtcgcgaact cttcagcatc tctgtgagat ggcaattcgg	1500
[0446]	tctaagcaag aactgctaca gcttgcccca atcgaagtgg aaagtaatgt atcaagagga	1560
[0447]	tgccctgaag gtactcaaac cttcgaaaaa gaaagcgagt aagttaaaat gcttcttcgt	1620
[0448]	ctcctattta taatatggtt tgttattgtt aattttgttc ttgtagaaga gcttaattaa	1680
[0449]	tcgttgttgt tatgaaatac tatttgtatg agatgaactg gtgtaatgta attcatttac	1740
[0450]	ataagtggag tcagaatcag aatgtttcct ccataactaa ctagacatga agacctgccg	1800
[0451]	cgtacaattg tcttatattt gaacaactaa aattgaacat cttttgccac aactttataa	1860
[0452]	gtggttaata tagctcaaat atatggtcaa gttcaataga ttaataatgg aaatatcagt	1920
[0453]	tatcgaaatt cattaacaat caacttaacg ttattaacta ctaattttat atcatcccct	1980
[0454]	ttgataaatg atagtacacc aattaggaag gagcatgctc gcctaggaga ttgtcgtttc	2040
[0455]	ccgccttcag tttgcaagct gctctagccg tgtagccaat acgcaaaccg cctctccccg	2100
[0456]	cgcggttgga attactagcg cgtgtcgaca agcttgcatt ccggtcaaca tgggtggagca	2160
[0457]	cgacacactt gtctactcca aaaatatcaa agatacagtc tcagaagacc aaagggaat	2220
[0458]	tgagactttt caacaaaggg taatatccgg aaacctctc ggattccatt gccagctat	2280
[0459]	ctgtcacttt attgtgaaga tagtggaaaa ggaaggtggc tcctacaaat gccatcattg	2340
[0460]	cgataaagga aaggccatcg ttgaagatgc ctctgccgac agtggtccca aagatggacc	2400
[0461]	cccaccacg aggagcatcg tggaaaaaga agacgttcca accacgtctt caaagcaagt	2460
[0462]	ggattgatgt gataacatgg tggagcacga cacacttgtc tactccaaaa atatcaaaga	2520
[0463]	tacagtctca gaagaccaa gggcaattga gacttttcaa caaagggtaa tatccgaaa	2580
[0464]	cctctcgga ttccattgcc cagctatctg tcactttatt gtgaagatag tggaaaagga	2640
[0465]	aggtggctcc tacaatatgcc atcattgcga taaaggaaag gccatcggtg aagatgcctc	2700
[0466]	tgccgacagt ggtcccaaag atggaccccc acccacgagg agcatcggtg aaaaagaaga	2760
[0467]	cgttccaacc acgtcttcaa agcaagtgga ttgatgtgat atctccactg acgtaaggga	2820

[0468]	tgacgcacaa tcccactatc cttcgcaaga cccttcctct atataaggaa gttcatttca	2880
[0469]	tttggagagg tattaaaatc ttaatagggt ttgataaaag cgaacgtggg gaaacccgaa	2940
[0470]	ccaaaccttc ttctaaactc tctctcatct ctcttaaagc aaacttctct cttgtctttc	3000
[0471]	ttgcgtgagc gatcttcaac gttgtcagat cgtgcttcgg caccagtaca acgttttctt	3060
[0472]	tcactgaagc gaaatcaaag atctctttgt ggacacgtag tgcggcgcca ttaaataacg	3120
[0473]	tgtacttgtc ctattcttgt cgggtgtggtc ttgggaaaag aaagcttgct ggaggctgct	3180
[0474]	gttcagcccc atacattact tgttacgatt ctgctgactt tcggcgggtg caatatctct	3240
[0475]	acttctgctt gacgaggtat tgttgccctgt acttctttct tcttcttctt gctgattggt	3300
[0476]	tctataagaa atctagtatt ttctttgaaa cagagtttcc cgtgggtttt cgaacttgga	3360
[0477]	gaaagattgt taagcttctg tatattctgc ccaaatttgt cgggcccatt accgcggaga	3420
[0478]	aatggcgaa aaacgttgcg attttcggct tattgttttc tcttcttggt ttggttccct	3480
[0479]	ctcagatctt cgcgacgtca ctctcagcc aaaacgacac ccccatctgt ctatccactg	3540
[0480]	gcccctggat ctgctgcccc aactaactcc atggtgacct tgggatgcct ggtcaagggc	3600
[0481]	tatttccctg agccagtgc agtgacctgg aactctggat ccctgtccag cgggtgtgcac	3660
[0482]	accttcccag ctgtcctgca gtctgacctc tacactctga gcagctcagt gactgtcccc	3720
[0483]	tccagcacct ggcccagcga gaccgtcacc tgcaacgttg cccacccggc cagcagcacc	3780
[0484]	aaggtggaca agaaaattgt gccagggat tgtggttgta agccttgcat atgtacagtc	3840
[0485]	ccagaagtat catctgtctt catcttcccc ccaaagccca aggatgtgct caccattact	3900
[0486]	ctgactccta aggtcacgtg tgttggtgta gacatcagca aggatgatcc cgaggtccag	3960
[0487]	ttcagctggt ttgtagatga tgtggagggt cacacagctc agacgcaacc ccgggaggag	4020
[0488]	cagttcaaca gcactttccg ctgagtcagt gaacttccca tcatgcacca ggactggctc	4080
[0489]	aatggcaagg agcgatcgct caccatcacc atcaccatca ccatcaccat taaaggccta	4140
[0490]	ttttctttag ttggaattta ctgttatctg gtgtgcattt ctatgtttgg tgagcggttt	4200
[0491]	tctgtgctca gagtgtgttt attttatgta atttaatttc tttgtgagct cctgttttagc	4260
[0492]	aggtcgtccc ttcagcaagg acacaaaaag attttaattt tattaaaaaa aaaaaaaaaa	4320
[0493]	aagaccggga attcgatata aagcttatcg acctgcagat cgttcaaaca tttggcaata	4380
[0494]	aagtttctta agattgaatc ctgttgccgg tcttgcatg attatcatat aatttctggt	4440
[0495]	gaattacgtt aagcatgtaa taattaacat gtaatgcatg acgttattta tgagatgggt	4500
[0496]	ttttatgatt agagtcccg c aattatacat ttaatacgcg atagaaaaca aaatatagcg	4560
[0497]	cgcaaactag gataaattat cgcgcgcggg gtcactatg ttactagatc tctagagtct	4620
[0498]	caagcttggc gcgcccacgt gactagtggc actggccgtc gttttacaac gtcgtgactg	4680
[0499]	ggaaaacctt ggcgttacct aacttaatcg ccttgacga catccccctt tcgccagctg	4740
[0500]	gcgtaatagc gaagaggccc gcaccgatcg cccttcccaa cagttgcgca gcctgaatgg	4800
[0501]	cgaatgctag agcagcttga gcttgatca gattgtcgtt tcccgccttc agttttaaact	4860
[0502]	atcagtgttt gacaggatat attggcgggt aaacctaaga gaaaagagcg tttta	4914
[0503]	<210> 26	
[0504]	<211> 3480	
[0505]	<212> DNA	
[0506]	<213> 人工序列	

[0507]	<220>	
[0508]	<223>	表达盒1819的核苷酸序列
[0509]	<400>	26
[0510]	gtcaacatgg tggagcacga cacacttgtc tactccaaaa atatcaaaga tacagtctca	60
[0511]	gaagaccaa gggcaattga gacttttcaa caaagggtaa tatccggaaa cctcctcgga	120
[0512]	ttccattgcc cagctatctg tcactttatt gtgaagatag tggaaaagga aggtggctcc	180
[0513]	tacaaatgcc atcattgcga taaaggaaag gccatcgttg aagatgcctc tgccgacagt	240
[0514]	ggtcccaaag atggaccccc acccacgagg agcatcgtgg aaaaagaaga cgttccaacc	300
[0515]	acgtcttcaa agcaagtgga ttgatgtgat aacatgggtg agcacgacac acttgtctac	360
[0516]	tccaaaaata tcaaagatac agtctcagaa gaccaaaggg caattgagac ttttcaacaa	420
[0517]	agggtaatat ccggaacct cctcggattc cattgccag ctatctgtca ctttattgtg	480
[0518]	aagatagtgg aaaaggaagg tggtcctac aaatgccatc attgcgataa aggaaaggcc	540
[0519]	atcgttgaag atgcctctgc cgacagtggc ccaaagatg gacccccacc cagaggagc	600
[0520]	atcgtggaaa aagaagacgt tccaaccacg tcttcaaagc aagtggattg atgtgatatc	660
[0521]	tccactgacg taagggatga cgcacaatcc cactatcctt cgcaagacc ttcctctata	720
[0522]	taaggaagtt catctcattt ggagaggtat taaaatctta ataggttttg ataaaagcga	780
[0523]	acgtggggaa acccgaacca aaccttcttc taaactctct ctcatctctc ttaaagcaaa	840
[0524]	cttctctctt gtctttcttg cgtgagcgat cttcaacgtt gtcagatcgt gcttcggcac	900
[0525]	cagtacaacg ttttctttca ctgaagcgaa atcaaagatc tctttgtgga cacgtagtgc	960
[0526]	ggcgccatta aataacgtgt acttgtccta ttcttgtcgg tgtggtcttg ggaaaagaaa	1020
[0527]	gcttgctgga ggctgctgtt cagccccata cattacttgt tacgattctg ctgactttcg	1080
[0528]	gcgggtgcaa tatctctact tctgcttgac gaggtattgt tgccgtgact tctttcttct	1140
[0529]	tcttcttgct gattggttct ataagaaatc tagtattttc tttgaaacag agttttcccg	1200
[0530]	tggttttcga acttgagaa agattgttaa gcttctgtat attctgcca aatttgctcg	1260
[0531]	gccaataacc gcggagaaaa tggcgaaaaa cgttgcgatt ttcggcttat tgttttctct	1320
[0532]	tcttgtgttg gttccttctc agatcttcgc caaaaactt cctggaaatg acaacagcac	1380
[0533]	ggcaacgctg tgccttgggc accatgcagt accaaacgga acgatagtga aaacaatcac	1440
[0534]	gaatgaccaa attgaagtta ctaatgctac tgagctggtt cagaattcct caataggtga	1500
[0535]	aatatgcgac agtcctcatc agatccttga tggagaaaac tgcacactaa tagatgctct	1560
[0536]	attgggagac cctcagtgtg atggcttcca aaataagaaa tgggaccttt ttgttgaacg	1620
[0537]	aagcaaagcc tacagcaact gttaccctta tgatgtgccg gattatgcct cccttaggtc	1680
[0538]	actagttgcc tcattccgga cactggagtt taacaatgaa agcttcaatt ggactggagt	1740
[0539]	cactcaaac ggaacaagtt ctgcttgcac aaggagatct aataatagtt tcttttagtag	1800
[0540]	attaaattgg ttgaccact taaacttcaa ataccagca ttgaacgtga ctatgccaaa	1860
[0541]	caatgaacaa ttgacaaat tgtacatttg gggggttcac caccgggta cggacaagga	1920
[0542]	ccaaatcttc ctgtatgctc aatcatcagg agaatacaca gtatctacca aaagaagcca	1980
[0543]	acaagctgta atcccgaata tcgcatctag acccagaata aggaatatcc ctagcagaat	2040
[0544]	aagcatctat tggacaatag taaaaccggg agacatactt ttgattaaca gcacaggga	2100
[0545]	tctaattgct cctaggggtt acttcaaaat acgaagtggg aaaagctcaa taatgagatc	2160

[0546]	agatgcaccc attggcaaat gcaattctga atgcatcact ccaaatggaa gcattcccaa	2220
[0547]	tgacaaacca ttccaaaatg taaacaggat cacatacggg gcctgtccca gatatgttaa	2280
[0548]	gcaaagcact ctgaaattgg caacaggaat gcgaaatgta ccagagaaac aaactagagg	2340
[0549]	catatttggc gcaatagcgg gtttcataga aaatggttgg gagggaatgg tggatggttg	2400
[0550]	gtacggtttc aggcatcaaa attctgaggg aagaggacaa gcagcagatc tcaaaagcac	2460
[0551]	tcaagcagca atcgatcaaa tcaatgggaa gctgaatcga ttgatcgga aaaccaacga	2520
[0552]	gaaattccat cagattgaaa aagaattctc agaagtcgaa gggagaattc aggacctga	2580
[0553]	gaaatatgtt gaggacacta aaatagatct ctggtcatac aacgcggagc ttcttggtgc	2640
[0554]	cctggagaac caacatacaa ttgatctaac tgactcagaa atgaacaaac tgtttgaaaa	2700
[0555]	aacaaagaag caactaaggg aaaatgctga ggatatgggc aatggttggt tcaaaatata	2760
[0556]	ccacaaatgt gacaatgcct gcataggatc aatcagaaat ggaacttatg accacgatgt	2820
[0557]	atacagagat gaagcattaa acaaccggtt ccagatcaag ggagttgagc tgaagtcagg	2880
[0558]	gtacaaagat tggatcctat ggatttctt tgccatatca tgtttttgc tttgtgtgc	2940
[0559]	tttgttgggg ttcatcatgt gggcctgcc aagggaac attaggtgca acatttgc	3000
[0560]	ttgaaggcct attttcttta gtttgaattt actgttattc ggtgtgcatt tctatgttg	3060
[0561]	gtgagcgtt ttctgtgctc agagtgtgtt tattttatgt aatttaattt ctttgtgagc	3120
[0562]	tcctgtttag caggtcgtcc cttcagcaag gacacaaaa gattttaatt ttattaaaa	3180
[0563]	aaaaaaaa aaagaccggg aattcgatat caagcttattc gacctgcaga tcgttcaa	3240
[0564]	atttggaat aaagtttctt aagattgaat cctgttgccg gtcttgcgat gattatcata	3300
[0565]	taatttctgt tgaattacgt taagcatgta ataattaaca tgtaatgcat gacgttattt	3360
[0566]	atgagatggg tttttatgat tagagtcccg caattataca tttaatacgc gatagaaa	3420
[0567]	aaaatatagc gcgcaacta ggataaatta tcgcgcgcgg tgtcatctat gttactagat	3480
[0568]	<210>	27
[0569]	<211>	52
[0570]	<212>	DNA
[0571]	<213>	人工序列
[0572]	<220>	
[0573]	<223>	引物IF-HT1*-PDI.c
[0574]	<400>	27
[0575]	gtcgggcca ataccgcgga gacaatggcg aaaaacgttg cgattttcgg ct	52
[0576]	<210>	28
[0577]	<211>	52
[0578]	<212>	DNA
[0579]	<213>	人工序列
[0580]	<220>	
[0581]	<223>	引物IF-HT2*-PDI.c
[0582]	<400>	28
[0583]	gtcgggcca ataccgcgga ggaaatggcg aaaaacgttg cgattttcgg ct	52
[0584]	<210>	29

[0585] <211> 52
[0586] <212> DNA
[0587] <213> 人工序列
[0588] <220>
[0589] <223> 引物IF-HT3*-PDI.c
[0590] <400> 29
[0591] gtcgggcccc ataccgcgga aaaaatggcg aaaaacgttg cgatttttcgg ct 52
[0592] <210> 30
[0593] <211> 52
[0594] <212> DNA
[0595] <213> 人工序列
[0596] <220>
[0597] <223> 引物IF-HT4*-PDI.c
[0598] <400> 30
[0599] gtcgggcccc ataccgcgga aacaatggcg aaaaacgttg cgatttttcgg ct 52
[0600] <210> 31
[0601] <211> 52
[0602] <212> DNA
[0603] <213> 人工序列
[0604] <220>
[0605] <223> 引物IF-HT5*-PDI.c
[0606] <400> 31
[0607] gtcgggcccc ataccgcgga agcaatggcg aaaaacgttg cgatttttcgg ct 52
[0608] <210> 32
[0609] <211> 52
[0610] <212> DNA
[0611] <213> 人工序列
[0612] <220>
[0613] <223> 引物IF-HT6*-PDI.c
[0614] <400> 32
[0615] gtcgggcccc ataccgcgga agaaatggcg aaaaacgttg cgatttttcgg ct 52
[0616] <210> 33
[0617] <211> 53
[0618] <212> DNA
[0619] <213> 人工序列
[0620] <220>
[0621] <223> IF-HT7*-PDI.c
[0622] <400> 33
[0623] gtcgggcccc ataccgcgga aagaaatggc gaaaaacgtt gcgatttttcg gct 53

[0624]	<210> 34	
[0625]	<211> 54	
[0626]	<212> DNA	
[0627]	<213> 人工序列	
[0628]	<220>	
[0629]	<223> IF-HT8*-PDI.c	
[0630]	<400> 34	
[0631]	gtcgggcccc ataccgcgga aaagaaatgg cgaaaaacgt tgcgattttc ggct	54
[0632]	<210> 35	
[0633]	<211> 1722	
[0634]	<212> DNA	
[0635]	<213> 人工序列	
[0636]	<220>	
[0637]	<223> PDISP/H1 California的核苷酸序列	
[0638]	<400> 35	
[0639]	atggcgaaaa acgttgcgat tttcggctta ttgttttctc ttcttgtgtt ggttccttct	60
[0640]	cagatcttcg ctgacacatt atgtataggt tatcatgcga acaattcaac agacactgta	120
[0641]	gacacagtac tagaaaagaa tgtaacagta acacactctg ttaaccttct agaagacaag	180
[0642]	cataacggga aactatgcaa actaagaggg gtagcccat tgcatttggg taaatgtaac	240
[0643]	attgctggct ggatcctggg aaatccagag tgtgaatcac tctccacagc aagctcatgg	300
[0644]	tcctacattg tggaaacacc tagttcagac aatggaacgt gttaccagg agatttcac	360
[0645]	gattatgagg agctaagaga gcaattgagc tcagtgtcat catttgaaag gtttgagata	420
[0646]	ttccccaaga caagttcatg gcccaatcat gactcgaaca aaggtgtaac ggcagcatgt	480
[0647]	cctcatgctg gagcaaaaag cttctacaaa aatttaatat ggctagttaa aaaaggaaat	540
[0648]	tcatacccaa agctcagcaa atcctacatt aatgataaag ggaaagaagt cctcgtgcta	600
[0649]	tggggcattc accatccatc tactagtgtc gaccaacaaa gtctctatca gaatgcagat	660
[0650]	gcatatgttt ttgtggggtc atcaagatac agcaagaagt tcaagccgga aatagcaata	720
[0651]	agacccaaag tgagggatca agaagggaga atgaactatt actggacact agtagagccg	780
[0652]	ggagacaaaa taacattcga agcaactgga aatctagtgg taccgagata tgcattcgca	840
[0653]	atggaaagaa atgctggatc tgggtattatc atttcagata caccagtcca cgattgcaat	900
[0654]	acaacttgtc aaacacccaa ggggtgctata aacaccagcc tcccatTTca gaatatacat	960
[0655]	ccgatcacia ttggaaaatg tccaaaatat gtaaaaagca caaaattgag actggccaca	1020
[0656]	ggattgagga atatcccgtc tattcaatct agaggactat ttggggccat tgccggtttc	1080
[0657]	attgaagggg ggtggacagg gatggtagat ggatggtacg gttatcacca tcaaaatgag	1140
[0658]	caggggtcag gatatgcagc cgacctgaag agcacacaga atgccattga cgagattact	1200
[0659]	aacaaagtaa attctgttat tgaaaagatg aatacacagt tcacagcagt aggtaaagag	1260
[0660]	ttcaaccacc tggaaaaaag aatagagaat ttaaataaaa aagttgatga tggtttctctg	1320
[0661]	gacatttggc cttacaatgc cgaactgttg gttctattgg aaaatgaaag aactttggac	1380
[0662]	taccacgatt caaatgtgaa gaacttatat gaaaaggtaa gaagccagct aaaaaacaat	1440

[0663] gccaaagaaa ttggaaacgg ctgctttgaa ttttaccaca aatgcgataa cacgtgcatg 1500
 [0664] gaaagtgtca aaaatgggac ttatgactac caaaataact cagaggaagc aaaattaaac 1560
 [0665] agagaagaaa tagatgggggt aaagctggaa tcaacaagga ttaccagat tttggcgatc 1620
 [0666] tattcaactg tcgccagtgc atttggtactg gtagtctccc tgggggcaat cagtttctgg 1680
 [0667] atgtgctcta atgggtctct acagtgtaga atatgtattt aa 1722
 [0668] <210> 36
 [0669] <211> 573
 [0670] <212> PRT
 [0671] <213> 人工序列
 [0672] <220>
 [0673] <223> PDISP/H1 California的氨基酸序列
 [0674] <400> 36
 [0675] Met Ala Lys Asn Val Ala Ile Phe Gly Leu Leu Phe Ser Leu Leu Val
 [0676] 1 5 10 15
 [0677] Leu Val Pro Ser Gln Ile Phe Ala Asp Thr Leu Cys Ile Gly Tyr His
 [0678] 20 25 30
 [0679] Ala Asn Asn Ser Thr Asp Thr Val Asp Thr Val Leu Glu Lys Asn Val
 [0680] 35 40 45
 [0681] Thr Val Thr His Ser Val Asn Leu Leu Glu Asp Lys His Asn Gly Lys
 [0682] 50 55 60
 [0683] Leu Cys Lys Leu Arg Gly Val Ala Pro Leu His Leu Gly Lys Cys Asn
 [0684] 65 70 75 80
 [0685] Ile Ala Gly Trp Ile Leu Gly Asn Pro Glu Cys Glu Ser Leu Ser Thr
 [0686] 85 90 95
 [0687] Ala Ser Ser Trp Ser Tyr Ile Val Glu Thr Pro Ser Ser Asp Asn Gly
 [0688] 100 105 110
 [0689] Thr Cys Tyr Pro Gly Asp Phe Ile Asp Tyr Glu Glu Leu Arg Glu Gln
 [0690] 115 120 125
 [0691] Leu Ser Ser Val Ser Ser Phe Glu Arg Phe Glu Ile Phe Pro Lys Thr
 [0692] 130 135 140
 [0693] Ser Ser Trp Pro Asn His Asp Ser Asn Lys Gly Val Thr Ala Ala Cys
 [0694] 145 150 155 160
 [0695] Pro His Ala Gly Ala Lys Ser Phe Tyr Lys Asn Leu Ile Trp Leu Val
 [0696] 165 170 175
 [0697] Lys Lys Gly Asn Ser Tyr Pro Lys Leu Ser Lys Ser Tyr Ile Asn Asp
 [0698] 180 185 190
 [0699] Lys Gly Lys Glu Val Leu Val Leu Trp Gly Ile His His Pro Ser Thr
 [0700] 195 200 205
 [0701] Ser Ala Asp Gln Gln Ser Leu Tyr Gln Asn Ala Asp Ala Tyr Val Phe

[0702]	210	215	220
[0703]	Val Gly Ser Ser Arg Tyr Ser Lys Lys Phe Lys Pro Glu Ile Ala Ile		
[0704]	225	230	235
[0705]	Arg Pro Lys Val Arg Asp Gln Glu Gly Arg Met Asn Tyr Tyr Trp Thr		240
[0706]		245	250
[0707]	Leu Val Glu Pro Gly Asp Lys Ile Thr Phe Glu Ala Thr Gly Asn Leu		255
[0708]		260	265
[0709]	Val Val Pro Arg Tyr Ala Phe Ala Met Glu Arg Asn Ala Gly Ser Gly		270
[0710]		275	280
[0711]	Ile Ile Ile Ser Asp Thr Pro Val His Asp Cys Asn Thr Thr Cys Gln		285
[0712]		290	295
[0713]	Thr Pro Lys Gly Ala Ile Asn Thr Ser Leu Pro Phe Gln Asn Ile His		300
[0714]	305	310	315
[0715]	Pro Ile Thr Ile Gly Lys Cys Pro Lys Tyr Val Lys Ser Thr Lys Leu		320
[0716]		325	330
[0717]	Arg Leu Ala Thr Gly Leu Arg Asn Ile Pro Ser Ile Gln Ser Arg Gly		335
[0718]		340	345
[0719]	Leu Phe Gly Ala Ile Ala Gly Phe Ile Glu Gly Gly Trp Thr Gly Met		350
[0720]		355	360
[0721]	Val Asp Gly Trp Tyr Gly Tyr His His Gln Asn Glu Gln Gly Ser Gly		365
[0722]		370	375
[0723]	Tyr Ala Ala Asp Leu Lys Ser Thr Gln Asn Ala Ile Asp Glu Ile Thr		380
[0724]	385	390	395
[0725]	Asn Lys Val Asn Ser Val Ile Glu Lys Met Asn Thr Gln Phe Thr Ala		400
[0726]		405	410
[0727]	Val Gly Lys Glu Phe Asn His Leu Glu Lys Arg Ile Glu Asn Leu Asn		415
[0728]		420	425
[0729]	Lys Lys Val Asp Asp Gly Phe Leu Asp Ile Trp Thr Tyr Asn Ala Glu		430
[0730]		435	440
[0731]	Leu Leu Val Leu Leu Glu Asn Glu Arg Thr Leu Asp Tyr His Asp Ser		445
[0732]		450	455
[0733]	Asn Val Lys Asn Leu Tyr Glu Lys Val Arg Ser Gln Leu Lys Asn Asn		460
[0734]	465	470	475
[0735]	Ala Lys Glu Ile Gly Asn Gly Cys Phe Glu Phe Tyr His Lys Cys Asp		480
[0736]		485	490
[0737]	Asn Thr Cys Met Glu Ser Val Lys Asn Gly Thr Tyr Asp Tyr Pro Lys		495
[0738]		500	505
[0739]	Tyr Ser Glu Glu Ala Lys Leu Asn Arg Glu Glu Ile Asp Gly Val Lys		510
[0740]		515	520
			525

[0741]	Leu Glu Ser Thr Arg Ile Tyr Gln Ile Leu Ala Ile Tyr Ser Thr Val	
[0742]	530	540
[0743]	Ala Ser Ser Leu Val Leu Val Val Ser Leu Gly Ala Ile Ser Phe Trp	
[0744]	545	555
[0745]	Met Cys Ser Asn Gly Ser Leu Gln Cys Arg Ile Cys Ile	560
[0746]	565	570
[0747]	<210> 37	
[0748]	<211> 1728	
[0749]	<212> DNA	
[0750]	<213> 人工序列	
[0751]	<220>	
[0752]	<223> PDISP/H5 Indonesia的核苷酸序列	
[0753]	<400> 37	
[0754]	atggcgaaaa acgttgcatg tttcggttta ttgttttctc ttcttgtgtt ggttccttct	60
[0755]	cagatcttcg ccgatcagat ttgcattggt taccatgcaa acaattcaac agagcaggtt	120
[0756]	gacacaatca tggaaaagaa cgttactgtt acacatgccc aagacatact ggaaaagaca	180
[0757]	cacaacggga agctctgcga tctagatgga gtgaagcctc taattttaag agattgtagt	240
[0758]	gtagctggat ggctcctcgg gaacccaatg tgtgacgaat tcatcaatgt accggaatgg	300
[0759]	tcttacatag tggagaaggc caatccaacc aatgacctct gttaccagg gagtttcaac	360
[0760]	gactatgaag aactgaaaca cctattgagc agaataaacc attttgagaa aattcaaate	420
[0761]	atcccaaaaa gttcttggtc cgatcatgaa gcctcatcag gagttagctc agcatgtcca	480
[0762]	tacctgggaa gtccctcctt ttttagaaat gtggtatggc ttatcaaaaa gaacagtaca	540
[0763]	taccaacaa taaagaaaag ctacaataat accaaccaag aggatctttt ggtactgtgg	600
[0764]	ggaattcacc atcctaata tgcggcagag cagacaaggc tatatcaaaa cccaaccacc	660
[0765]	tatatctcca ttgggacatc aacactaaac cagagattgg taccaaaaat agctactaga	720
[0766]	tccaaagtaa acgggcaaag tggaaggatg gagttcttct ggacaatttt aaacctaate	780
[0767]	gatgcaatca acttcgagag taatggaaat ttcatgtctc cagaatatgc atacaaaatt	840
[0768]	gtcaagaaag gggactcagc aattatgaaa agtgaattgg aatatggtaa ctgcaacacc	900
[0769]	aagtgtcaaa ctccaatggg ggcgataaac tctagtatgc cattccacaa catacacct	960
[0770]	ctcaccatcg gggaatgccc caaatatgtg aatcaaaaca gattagtcct tgcaacaggg	1020
[0771]	ctcagaaata gccctcaaag agagagcaga agaaaaaga gaggactatt tggagctata	1080
[0772]	gcaggtttta tagaggagg atggcaggga atggtagatg gttggtatgg gtaccaccat	1140
[0773]	agcaatgagc aggggagtg gtacgtgca gacaaagaat ccactcaaaa ggcaatagat	1200
[0774]	ggagtacca ataaggtcaa ctcaatcatt gacaaaatga aactcagtt tgaggccgtt	1260
[0775]	ggaagggaat ttaataactt agaaaggaga atagagaatt taaacaagaa gatggaagac	1320
[0776]	gggtttctag atgtctggac ttataatgcc gaacttctgg ttctcatgga aatgagaga	1380
[0777]	actctagact ttcatgactc aaatgttaag aacctctacg acaaggtecg actacagctt	1440
[0778]	agggataatg caaaggagct gggtaacggt tgtttcgagt tctatcaca atgtgataat	1500
[0779]	gaatgtatgg aaagtataag aaacggaacg tacaactatc cgcagtattc agaagaagca	1560

[0780] agattaaaaa gagaggaaat aagtggggta aaattggaat caataggaac ttaccaaata 1620
 [0781] ctgtcaattt attcaacagt ggcgagttcc ctagcactgg caatcatgat ggctggtcta 1680
 [0782] tctttatgga tgtgtccaa tggatcgta caatgcagaa ttgcatt 1728
 [0783] <210> 38
 [0784] <211> 576
 [0785] <212> PRT
 [0786] <213> 人工序列
 [0787] <220>
 [0788] <223> PDISP/H5 Indonesia的氨基酸序列
 [0789] <400> 38
 [0790] Met Ala Lys Asn Val Ala Ile Phe Gly Leu Leu Phe Ser Leu Leu Val
 [0791] 1 5 10 15
 [0792] Leu Val Pro Ser Gln Ile Phe Ala Asp Gln Ile Cys Ile Gly Tyr His
 [0793] 20 25 30
 [0794] Ala Asn Asn Ser Thr Glu Gln Val Asp Thr Ile Met Glu Lys Asn Val
 [0795] 35 40 45
 [0796] Thr Val Thr His Ala Gln Asp Ile Leu Glu Lys Thr His Asn Gly Lys
 [0797] 50 55 60
 [0798] Leu Cys Asp Leu Asp Gly Val Lys Pro Leu Ile Leu Arg Asp Cys Ser
 [0799] 65 70 75 80
 [0800] Val Ala Gly Trp Leu Leu Gly Asn Pro Met Cys Asp Glu Phe Ile Asn
 [0801] 85 90 95
 [0802] Val Pro Glu Trp Ser Tyr Ile Val Glu Lys Ala Asn Pro Thr Asn Asp
 [0803] 100 105 110
 [0804] Leu Cys Tyr Pro Gly Ser Phe Asn Asp Tyr Glu Glu Leu Lys His Leu
 [0805] 115 120 125
 [0806] Leu Ser Arg Ile Asn His Phe Glu Lys Ile Gln Ile Ile Pro Lys Ser
 [0807] 130 135 140
 [0808] Ser Trp Ser Asp His Glu Ala Ser Ser Gly Val Ser Ser Ala Cys Pro
 [0809] 145 150 155 160
 [0810] Tyr Leu Gly Ser Pro Ser Phe Phe Arg Asn Val Val Trp Leu Ile Lys
 [0811] 165 170 175
 [0812] Lys Asn Ser Thr Tyr Pro Thr Ile Lys Lys Ser Tyr Asn Asn Thr Asn
 [0813] 180 185 190
 [0814] Gln Glu Asp Leu Leu Val Leu Trp Gly Ile His His Pro Asn Asp Ala
 [0815] 195 200 205
 [0816] Ala Glu Gln Thr Arg Leu Tyr Gln Asn Pro Thr Thr Tyr Ile Ser Ile
 [0817] 210 215 220
 [0818] Gly Thr Ser Thr Leu Asn Gln Arg Leu Val Pro Lys Ile Ala Thr Arg

[0819]	225	230	235	240
[0820]	Ser Lys Val Asn Gly Gln Ser Gly Arg Met Glu Phe Phe Trp Thr Ile			
[0821]		245	250	255
[0822]	Leu Lys Pro Asn Asp Ala Ile Asn Phe Glu Ser Asn Gly Asn Phe Ile			
[0823]		260	265	270
[0824]	Ala Pro Glu Tyr Ala Tyr Lys Ile Val Lys Lys Gly Asp Ser Ala Ile			
[0825]		275	280	285
[0826]	Met Lys Ser Glu Leu Glu Tyr Gly Asn Cys Asn Thr Lys Cys Gln Thr			
[0827]		290	295	300
[0828]	Pro Met Gly Ala Ile Asn Ser Ser Met Pro Phe His Asn Ile His Pro			
[0829]	305	310	315	320
[0830]	Leu Thr Ile Gly Glu Cys Pro Lys Tyr Val Lys Ser Asn Arg Leu Val			
[0831]		325	330	335
[0832]	Leu Ala Thr Gly Leu Arg Asn Ser Pro Gln Arg Glu Ser Arg Arg Lys			
[0833]		340	345	350
[0834]	Lys Arg Gly Leu Phe Gly Ala Ile Ala Gly Phe Ile Glu Gly Gly Trp			
[0835]		355	360	365
[0836]	Gln Gly Met Val Asp Gly Trp Tyr Gly Tyr His His Ser Asn Glu Gln			
[0837]		370	375	380
[0838]	Gly Ser Gly Tyr Ala Ala Asp Lys Glu Ser Thr Gln Lys Ala Ile Asp			
[0839]	385	390	395	400
[0840]	Gly Val Thr Asn Lys Val Asn Ser Ile Ile Asp Lys Met Asn Thr Gln			
[0841]		405	410	415
[0842]	Phe Glu Ala Val Gly Arg Glu Phe Asn Asn Leu Glu Arg Arg Ile Glu			
[0843]		420	425	430
[0844]	Asn Leu Asn Lys Lys Met Glu Asp Gly Phe Leu Asp Val Trp Thr Tyr			
[0845]		435	440	445
[0846]	Asn Ala Glu Leu Leu Val Leu Met Glu Asn Glu Arg Thr Leu Asp Phe			
[0847]		450	455	460
[0848]	His Asp Ser Asn Val Lys Asn Leu Tyr Asp Lys Val Arg Leu Gln Leu			
[0849]	465	470	475	480
[0850]	Arg Asp Asn Ala Lys Glu Leu Gly Asn Gly Cys Phe Glu Phe Tyr His			
[0851]		485	490	495
[0852]	Lys Cys Asp Asn Glu Cys Met Glu Ser Ile Arg Asn Gly Thr Tyr Asn			
[0853]		500	505	510
[0854]	Tyr Pro Gln Tyr Ser Glu Glu Ala Arg Leu Lys Arg Glu Glu Ile Ser			
[0855]		515	520	525
[0856]	Gly Val Lys Leu Glu Ser Ile Gly Thr Tyr Gln Ile Leu Ser Ile Tyr			
[0857]		530	535	540

[0858]	Ser Thr Val Ala Ser Ser Leu Ala Leu Ala Ile Met Met Ala Gly Leu	
[0859]	545	550 555 560
[0860]	Ser Leu Trp Met Cys Ser Asn Gly Ser Leu Gln Cys Arg Ile Cys Ile	
[0861]	565	570 575
[0862]	<210> 39	
[0863]	<211> 1701	
[0864]	<212> DNA	
[0865]	<213> 人工序列	
[0866]	<220>	
[0867]	<223> PDISP/H7 Hangzhou的核苷酸序列	
[0868]	<400> 39	
[0869]	atggcgaaaa acgttgcgat tttcggttta ttgttttctc ttcttgtgtt ggttccttct	60
[0870]	cagatcttcg cggacaaaat ctgcctcgga catcatgccg tgtcaaacgg aaccaaagta	120
[0871]	aacacattaa ctgaaagagg agtggaagtc gtcaatgcaa ctgaaacagt ggaacgaaca	180
[0872]	aacatcccca ggatctgctc aaaagggaaa aggacagttg acctcggtca atgttgactc	240
[0873]	ctggggacaa tcaactggacc acctcaatgt gaccaattcc tagaattttc agccgattta	300
[0874]	attattgaga ggcgagaagg aagtgatgtc tgttatcctg ggaaattcgt gaatgaagaa	360
[0875]	gctctgaggc aaattctcag agaatcaggc ggaattgaca aggaagcaat gggattcaca	420
[0876]	tacagtggaa taagaactaa tggagcaacc agtgcattga ggagatcagg atcttcattc	480
[0877]	tatgcagaaa tgaaatggct cctgtcaaac acagataatg ctgcattccc gcagatgact	540
[0878]	aagtcatata aaaatacaag aaaaagccca gctctaatag tatgggggat ccatcattcc	600
[0879]	gtatcaactg cagagcaaac caagctatat gggagtggaa acaaactggg gacagttggg	660
[0880]	agttctaatt atcaacaatc ttttgtaccg agtccaggag cgagaccaca agttaatggg	720
[0881]	atatctggaa gaattgactt tcattggcta atgctaaatc ccaatgatac agtcactttc	780
[0882]	agtttcaatg gggctttcat agctccagac cgtgcaagct tcctgagagg aaaatctatg	840
[0883]	ggaatccaga gtggagtaca ggttgatgcc aattgtgaag gggactgcta tcatagtga	900
[0884]	gggacaataa taagtaactt gccatttcag aacatagata gcagggcagt tggaaaatgt	960
[0885]	ccgagatatg ttaagcaaag gagtctgctg ctagcaacag ggaatgaaga tgttcctgag	1020
[0886]	attccaaagg gaagaggcct atttggtgct atagcgggtt tcattgaaaa tggatgggaa	1080
[0887]	ggcctaattg atggttggtg tggtttcaga caccagaatg cacagggaga gggaactgct	1140
[0888]	gcagattaca aaagcactca atcggaatt gatcaataa caggaaaatt aaaccggctt	1200
[0889]	atagaaaaaa ccaaccaaca atttgagttg atcgacaatg aattcaatga ggtagagaag	1260
[0890]	caaatcggtg atgtgataaa ttggaccaga gattctataa cagaagtgtg gtcatacaat	1320
[0891]	gctgaactct tggtagcaat ggagaaccag catacaattg atctggctga ttcagaaatg	1380
[0892]	gacaaactgt acgaacgagt gaaaagacag ctgagagaga atgctgaaga agatggcact	1440
[0893]	ggttgctttg aaatatattca caagtgtgat gatgactgta tggccagtat tagaaataac	1500
[0894]	acctatgatc acagcaaata cagggaagag gcaatgcaaa atagaataca gattgacca	1560
[0895]	gtcaaactaa gcagcggtc caaagatgtg atactttggt ttagcttcgg ggcattcatgt	1620
[0896]	ttcatacttc tagccattgt aatgggcctt gtcttcatat gtgtaaagaa tggaaacatg	1680

[0897] cgggtgcacta tttgtatata a 1701
 [0898] <210> 40
 [0899] <211> 566
 [0900] <212> PRT
 [0901] <213> 人工序列
 [0902] <220>
 [0903] <223> PDISP/H7 Hangzhou的氨基酸序列
 [0904] <400> 40
 [0905] Met Ala Lys Asn Val Ala Ile Phe Gly Leu Leu Phe Ser Leu Leu Val
 [0906] 1 5 10 15
 [0907] Leu Val Pro Ser Gln Ile Phe Ala Asp Lys Ile Cys Leu Gly His His
 [0908] 20 25 30
 [0909] Ala Val Ser Asn Gly Thr Lys Val Asn Thr Leu Thr Glu Arg Gly Val
 [0910] 35 40 45
 [0911] Glu Val Val Asn Ala Thr Glu Thr Val Glu Arg Thr Asn Ile Pro Arg
 [0912] 50 55 60
 [0913] Ile Cys Ser Lys Gly Lys Arg Thr Val Asp Leu Gly Gln Cys Gly Leu
 [0914] 65 70 75 80
 [0915] Leu Gly Thr Ile Thr Gly Pro Pro Gln Cys Asp Gln Phe Leu Glu Phe
 [0916] 85 90 95
 [0917] Ser Ala Asp Leu Ile Ile Glu Arg Arg Glu Gly Ser Asp Val Cys Tyr
 [0918] 100 105 110
 [0919] Pro Gly Lys Phe Val Asn Glu Glu Ala Leu Arg Gln Ile Leu Arg Glu
 [0920] 115 120 125
 [0921] Ser Gly Gly Ile Asp Lys Glu Ala Met Gly Phe Thr Tyr Ser Gly Ile
 [0922] 130 135 140
 [0923] Arg Thr Asn Gly Ala Thr Ser Ala Cys Arg Arg Ser Gly Ser Ser Phe
 [0924] 145 150 155 160
 [0925] Tyr Ala Glu Met Lys Trp Leu Leu Ser Asn Thr Asp Asn Ala Ala Phe
 [0926] 165 170 175
 [0927] Pro Gln Met Thr Lys Ser Tyr Lys Asn Thr Arg Lys Ser Pro Ala Leu
 [0928] 180 185 190
 [0929] Ile Val Trp Gly Ile His His Ser Val Ser Thr Ala Glu Gln Thr Lys
 [0930] 195 200 205
 [0931] Leu Tyr Gly Ser Gly Asn Lys Leu Val Thr Val Gly Ser Ser Asn Tyr
 [0932] 210 215 220
 [0933] Gln Gln Ser Phe Val Pro Ser Pro Gly Ala Arg Pro Gln Val Asn Gly
 [0934] 225 230 235 240
 [0935] Ile Ser Gly Arg Ile Asp Phe His Trp Leu Met Leu Asn Pro Asn Asp

[0936]		245		250		255
[0937]	Thr Val Thr Phe Ser Phe Asn Gly Ala Phe Ile Ala Pro Asp Arg Ala					
[0938]		260		265		270
[0939]	Ser Phe Leu Arg Gly Lys Ser Met Gly Ile Gln Ser Gly Val Gln Val					
[0940]		275		280		285
[0941]	Asp Ala Asn Cys Glu Gly Asp Cys Tyr His Ser Gly Gly Thr Ile Ile					
[0942]		290		295		300
[0943]	Ser Asn Leu Pro Phe Gln Asn Ile Asp Ser Arg Ala Val Gly Lys Cys					
[0944]	305		310		315	320
[0945]	Pro Arg Tyr Val Lys Gln Arg Ser Leu Leu Leu Ala Thr Gly Met Lys					
[0946]		325		330		335
[0947]	Asn Val Pro Glu Ile Pro Lys Gly Arg Gly Leu Phe Gly Ala Ile Ala					
[0948]		340		345		350
[0949]	Gly Phe Ile Glu Asn Gly Trp Glu Gly Leu Ile Asp Gly Trp Tyr Gly					
[0950]		355		360		365
[0951]	Phe Arg His Gln Asn Ala Gln Gly Glu Gly Thr Ala Ala Asp Tyr Lys					
[0952]		370		375		380
[0953]	Ser Thr Gln Ser Ala Ile Asp Gln Ile Thr Gly Lys Leu Asn Arg Leu					
[0954]	385		390		395	400
[0955]	Ile Glu Lys Thr Asn Gln Gln Phe Glu Leu Ile Asp Asn Glu Phe Asn					
[0956]		405		410		415
[0957]	Glu Val Glu Lys Gln Ile Gly Asn Val Ile Asn Trp Thr Arg Asp Ser					
[0958]		420		425		430
[0959]	Ile Thr Glu Val Trp Ser Tyr Asn Ala Glu Leu Leu Val Ala Met Glu					
[0960]		435		440		445
[0961]	Asn Gln His Thr Ile Asp Leu Ala Asp Ser Glu Met Asp Lys Leu Tyr					
[0962]		450		455		460
[0963]	Glu Arg Val Lys Arg Gln Leu Arg Glu Asn Ala Glu Glu Asp Gly Thr					
[0964]	465		470		475	480
[0965]	Gly Cys Phe Glu Ile Phe His Lys Cys Asp Asp Asp Cys Met Ala Ser					
[0966]		485		490		495
[0967]	Ile Arg Asn Asn Thr Tyr Asp His Ser Lys Tyr Arg Glu Glu Ala Met					
[0968]		500		505		510
[0969]	Gln Asn Arg Ile Gln Ile Asp Pro Val Lys Leu Ser Ser Gly Tyr Lys					
[0970]		515		520		525
[0971]	Asp Val Ile Leu Trp Phe Ser Phe Gly Ala Ser Cys Phe Ile Leu Leu					
[0972]		530		535		540
[0973]	Ala Ile Val Met Gly Leu Val Phe Ile Cys Val Lys Asn Gly Asn Met					
[0974]	545		550		555	560

[0975]	Arg Cys Thr Ile Cys Ile	
[0976]	565	
[0977]	<210> 41	
[0978]	<211> 1698	
[0979]	<212> DNA	
[0980]	<213> 人工序列	
[0981]	<220>	
[0982]	<223> PDISP/H7 Hangzhou+H5 Indonesia TMCT的核苷酸序列	
[0983]	<400> 41	
[0984]	atggcgaaaa acgttgcgat tttcggccta ttgttttctc ttcttgtgtt gggttccttct	60
[0985]	cagatcttcg cggacaaaat ctgcctcgga catcatgccg tgtcaaacgg aaccaaagta	120
[0986]	aacacattaa ctgaaagagg agtggaagtc gtcaatgcaa ctgaaacagt ggaacgaaca	180
[0987]	aacatcccca ggatctgctc aaaagggaaa aggacagttg acctcggtca atgttgactc	240
[0988]	ctggggacaa tcaactggacc acctcaatgt gaccaattcc tagaattttc agccgattta	300
[0989]	attattgaga ggcgagaagg aagtgatgtc tgttatcctg ggaaattcgt gaatgaagaa	360
[0990]	gctctgaggc aaattctcag agaatcaggc ggaattgaca aggaagcaat gggattcaca	420
[0991]	tacagtggaa taagaactaa tggagcaacc agtgcattga ggagatcagg atcttcattc	480
[0992]	tatgcagaaa tgaaatggct cctgtcaaac acagataatg ctgcattccc gcagatgact	540
[0993]	aagtcatata aaaatacaag aaaaagccca gctctaatag tatgggggat ccatcattcc	600
[0994]	gtatcaactg cagagcaaac caagctatat gggagtggaa acaaactggg gacagttggg	660
[0995]	agttctaatt atcaacaatc ttttgtaccg agtccaggag cgagaccaca agttaatggg	720
[0996]	atatctggaa gaattgactt tcattggcta atgctaaatc ccaatgatac agtcactttc	780
[0997]	agtttcaatg gggctttcat agctccagac cgtgcaagct tcctgagagg aaaatctatg	840
[0998]	ggaatccaga gtggagtaca ggttgatgcc aattgtgaag gggactgcta tcatagtgga	900
[0999]	gggacaataa taagtaactt gccatttcag aacatagata gcagggcagt tggaaaatgt	960
[1000]	ccgagatatg ttaagcaaag gagtctgctg ctagcaacag ggatgaagaa tgttcctgag	1020
[1001]	attccaaagg gaagaggcct atttggtgct atagcgggtt tcattgaaaa tggatgggaa	1080
[1002]	ggcctaattg atggttggtg tggtttcaga caccagaatg cacagggaga gggaactgct	1140
[1003]	gcagattaca aaagcactca atcggcaatt gatcaaataa caggaaaatt aaaccggctt	1200
[1004]	atagaaaaaa ccaaccaaca atttgagttg atcgacaatg aattcaatga ggtagagaag	1260
[1005]	caaatcggtg atgtgataaa ttggaccaga gattctataa cagaagtgtg gtcatacaat	1320
[1006]	gctgaactct tggtagcaat ggagaaccag catacaattg atctggctga ttcagaaatg	1380
[1007]	gacaaactgt acgaacgagt gaaaagacag ctgagagaga atgctgaaga agatggcact	1440
[1008]	ggttgctttg aaatatattc caagtgtgat gatgactgta tggccagtat tagaaataac	1500
[1009]	acctatgatc acagcaaata cagggaagag gcaatgcaaa atagaataca gattgacca	1560
[1010]	gtcaaactaa gcagcggcta ccaaatactg tcaatttatt caacagtggc gagttcccta	1620
[1011]	gcactggcaa tcatgatggc tggctctatc ttatggatgt gctccaatgg atcgttacaa	1680
[1012]	tgcagaattt gcatttaa	1698
[1013]	<210> 42	

[1014]	<211>	565
[1015]	<212>	PRT
[1016]	<213>	人工序列
[1017]	<220>	
[1018]	<223>	PDISP/H7 Hangzhou+H5 Indonesia TMCT的氨基酸序列
[1019]	<400>	42
[1020]	Met Ala Lys Asn Val Ala Ile Phe Gly Leu Leu Phe Ser Leu Leu Val	
[1021]	1 5 10 15	
[1022]	Leu Val Pro Ser Gln Ile Phe Ala Asp Lys Ile Cys Leu Gly His His	
[1023]	20 25 30	
[1024]	Ala Val Ser Asn Gly Thr Lys Val Asn Thr Leu Thr Glu Arg Gly Val	
[1025]	35 40 45	
[1026]	Glu Val Val Asn Ala Thr Glu Thr Val Glu Arg Thr Asn Ile Pro Arg	
[1027]	50 55 60	
[1028]	Ile Cys Ser Lys Gly Lys Arg Thr Val Asp Leu Gly Gln Cys Gly Leu	
[1029]	65 70 75 80	
[1030]	Leu Gly Thr Ile Thr Gly Pro Pro Gln Cys Asp Gln Phe Leu Glu Phe	
[1031]	85 90 95	
[1032]	Ser Ala Asp Leu Ile Ile Glu Arg Arg Glu Gly Ser Asp Val Cys Tyr	
[1033]	100 105 110	
[1034]	Pro Gly Lys Phe Val Asn Glu Glu Ala Leu Arg Gln Ile Leu Arg Glu	
[1035]	115 120 125	
[1036]	Ser Gly Gly Ile Asp Lys Glu Ala Met Gly Phe Thr Tyr Ser Gly Ile	
[1037]	130 135 140	
[1038]	Arg Thr Asn Gly Ala Thr Ser Ala Cys Arg Arg Ser Gly Ser Ser Phe	
[1039]	145 150 155 160	
[1040]	Tyr Ala Glu Met Lys Trp Leu Leu Ser Asn Thr Asp Asn Ala Ala Phe	
[1041]	165 170 175	
[1042]	Pro Gln Met Thr Lys Ser Tyr Lys Asn Thr Arg Lys Ser Pro Ala Leu	
[1043]	180 185 190	
[1044]	Ile Val Trp Gly Ile His His Ser Val Ser Thr Ala Glu Gln Thr Lys	
[1045]	195 200 205	
[1046]	Leu Tyr Gly Ser Gly Asn Lys Leu Val Thr Val Gly Ser Ser Asn Tyr	
[1047]	210 215 220	
[1048]	Gln Gln Ser Phe Val Pro Ser Pro Gly Ala Arg Pro Gln Val Asn Gly	
[1049]	225 230 235 240	
[1050]	Ile Ser Gly Arg Ile Asp Phe His Trp Leu Met Leu Asn Pro Asn Asp	
[1051]	245 250 255	
[1052]	Thr Val Thr Phe Ser Phe Asn Gly Ala Phe Ile Ala Pro Asp Arg Ala	

[1053]	260	265	270
[1054]	Ser Phe Leu Arg Gly Lys Ser Met Gly Ile Gln Ser Gly Val Gln Val		
[1055]	275	280	285
[1056]	Asp Ala Asn Cys Glu Gly Asp Cys Tyr His Ser Gly Gly Thr Ile Ile		
[1057]	290	295	300
[1058]	Ser Asn Leu Pro Phe Gln Asn Ile Asp Ser Arg Ala Val Gly Lys Cys		
[1059]	305	310	315
[1060]	Pro Arg Tyr Val Lys Gln Arg Ser Leu Leu Leu Ala Thr Gly Met Lys		
[1061]	325	330	335
[1062]	Asn Val Pro Glu Ile Pro Lys Gly Arg Gly Leu Phe Gly Ala Ile Ala		
[1063]	340	345	350
[1064]	Gly Phe Ile Glu Asn Gly Trp Glu Gly Leu Ile Asp Gly Trp Tyr Gly		
[1065]	355	360	365
[1066]	Phe Arg His Gln Asn Ala Gln Gly Glu Gly Thr Ala Ala Asp Tyr Lys		
[1067]	370	375	380
[1068]	Ser Thr Gln Ser Ala Ile Asp Gln Ile Thr Gly Lys Leu Asn Arg Leu		
[1069]	385	390	395
[1070]	Ile Glu Lys Thr Asn Gln Gln Phe Glu Leu Ile Asp Asn Glu Phe Asn		
[1071]	405	410	415
[1072]	Glu Val Glu Lys Gln Ile Gly Asn Val Ile Asn Trp Thr Arg Asp Ser		
[1073]	420	425	430
[1074]	Ile Thr Glu Val Trp Ser Tyr Asn Ala Glu Leu Leu Val Ala Met Glu		
[1075]	435	440	445
[1076]	Asn Gln His Thr Ile Asp Leu Ala Asp Ser Glu Met Asp Lys Leu Tyr		
[1077]	450	455	460
[1078]	Glu Arg Val Lys Arg Gln Leu Arg Glu Asn Ala Glu Glu Asp Gly Thr		
[1079]	465	470	475
[1080]	Gly Cys Phe Glu Ile Phe His Lys Cys Asp Asp Asp Cys Met Ala Ser		
[1081]	485	490	495
[1082]	Ile Arg Asn Asn Thr Tyr Asp His Ser Lys Tyr Arg Glu Glu Ala Met		
[1083]	500	505	510
[1084]	Gln Asn Arg Ile Gln Ile Asp Pro Val Lys Leu Ser Ser Gly Tyr Gln		
[1085]	515	520	525
[1086]	Ile Leu Ser Ile Tyr Ser Thr Val Ala Ser Ser Leu Ala Leu Ala Ile		
[1087]	530	535	540
[1088]	Met Met Ala Gly Leu Ser Leu Trp Met Cys Ser Asn Gly Ser Leu Gln		
[1089]	545	550	555
[1090]	Cys Arg Ile Cys Ile		
[1091]	565		

[1092]	<210> 43	
[1093]	<211> 1734	
[1094]	<212> DNA	
[1095]	<213> 人工序列	
[1096]	<220>	
[1097]	<223> PDISP/HA B Brisbane (PrL-)的核苷酸序列	
[1098]	<400> 43	
[1099]	atggcgaaaa acgttgcgat tttcggctta ttgttttctc ttcttgtgtt ggttccttct	60
[1100]	cagatcttcg ccgatcgaat ctgcactgga ataacatcgt caaactcacc acatgtcgtc	120
[1101]	aaaactgcta ctcaaggga ggtcaatgtg actggtgtaa taccactgac aacaacaccc	180
[1102]	accaaattctc attttgcaaa tctcaaagga acagaaacca gggggaaact atgccccaaa	240
[1103]	tgctcaact gcacagatct ggacgtagcc ttgggcagac caaatgcac ggggaaaata	300
[1104]	ccctcggcaa gagtttcaat actccatgaa gtcagacctg ttacatctgg gtgcttttct	360
[1105]	ataatgcacg acagaacaaa aattagacag ctgcctaacc ttctccgagg atacgaacat	420
[1106]	atcaggttat caaccataa cgttatcaat gcagaaaatg caccaggagg accctacaaa	480
[1107]	attggaacct cagggtcttg ccctaacatt accaatggaa acggattttt cgcaacaatg	540
[1108]	gcttgggccg tccccaaaaa cgacaaaaac aaaacagcaa caaatccatt aacaatagaa	600
[1109]	gtaccataca tttgtacaga aggagaagac caaattaccg tttgggggtt ccactctgac	660
[1110]	aacgagaccc aaatggcaaa gctctatggg gactcaaagc cccagaagtt cacctcatct	720
[1111]	gccaacggag tgaccacaca ttacgtttca cagattgggtg gcttcccaaa tcaaacagaa	780
[1112]	gacggaggac taccacaaag tggtagaatt gttgttgatt acatgggtgca aaaatctggg	840
[1113]	aaaacaggaa caattaccta tcaaaggggt attttattgc ctcaaaaggt gtggtgcgca	900
[1114]	agtggcagga gcaaggtaat aaaaggatcc ttgcctttaa ttggagaagc agattgcctc	960
[1115]	cacgaaaaat acggttgatt aaacaaaagc aagccttact acacagggga acatgcaaag	1020
[1116]	gccataggaa attgcccatt atgggtgaaa acacccttga agctggccaa tggaacaaa	1080
[1117]	tatagacctc ctggttgagg atgggaagga atgattgcag gttggcacgg atacacatcc	1140
[1118]	catggggcac atggagtagc ggtggcagca gaccttaaga gcactcaaga ggccataaac	1200
[1119]	aagataacaa aaaatctcaa ctcttttagt gagctggaag taaagaatct tcaaagacta	1260
[1120]	agcggtgcca tggatgaact ccacaacgaa atactagaac tagatgagaa agtggatgat	1320
[1121]	ctcagagctg atacaataag ctcaaaaata gaactcgag tcctgctttc caatgaagga	1380
[1122]	ataataaaca gtgaagatga acatctcttg gcgcttgaaa gaaagctgaa gaaaatgctg	1440
[1123]	ggccccctctg ctgtagagat agggaatgga tgctttgaaa ccaaacacaa gtgcaaccag	1500
[1124]	acctgtctcg acagaatagc tgctggtacc tttgatgcag gagaattttc tctccccacc	1560
[1125]	tttgattcac tgaatattac tgctgcatct ttaaatgacg atggattgga taatcatact	1620
[1126]	atactgcttt actactcaac tgctgcctcc agtttggtg taacactgat gatagctatc	1680
[1127]	tttgtgtttt atatggtctc cagagacaat gtttcttget ccatctgtct ataa	1734
[1128]	<210> 44	
[1129]	<211> 577	
[1130]	<212> PRT	

[1131]	<213> 人工序列															
[1132]	<220>															
[1133]	<223> PDISP/HA B Brisbane (PrL-)的氨基酸序列															
[1134]	<400> 44															
[1135]	Met	Ala	Lys	Asn	Val	Ala	Ile	Phe	Gly	Leu	Leu	Phe	Ser	Leu	Leu	Val
[1136]	1			5					10					15		
[1137]	Leu	Val	Pro	Ser	Gln	Ile	Phe	Ala	Asp	Arg	Ile	Cys	Thr	Gly	Ile	Thr
[1138]				20					25					30		
[1139]	Ser	Ser	Asn	Ser	Pro	His	Val	Val	Lys	Thr	Ala	Thr	Gln	Gly	Glu	Val
[1140]				35					40					45		
[1141]	Asn	Val	Thr	Gly	Val	Ile	Pro	Leu	Thr	Thr	Thr	Pro	Thr	Lys	Ser	His
[1142]		50						55					60			
[1143]	Phe	Ala	Asn	Leu	Lys	Gly	Thr	Glu	Thr	Arg	Gly	Lys	Leu	Cys	Pro	Lys
[1144]	65					70					75				80	
[1145]	Cys	Leu	Asn	Cys	Thr	Asp	Leu	Asp	Val	Ala	Leu	Gly	Arg	Pro	Lys	Cys
[1146]					85					90					95	
[1147]	Thr	Gly	Lys	Ile	Pro	Ser	Ala	Arg	Val	Ser	Ile	Leu	His	Glu	Val	Arg
[1148]				100					105					110		
[1149]	Pro	Val	Thr	Ser	Gly	Cys	Phe	Pro	Ile	Met	His	Asp	Arg	Thr	Lys	Ile
[1150]			115					120					125			
[1151]	Arg	Gln	Leu	Pro	Asn	Leu	Leu	Arg	Gly	Tyr	Glu	His	Ile	Arg	Leu	Ser
[1152]		130					135						140			
[1153]	Thr	His	Asn	Val	Ile	Asn	Ala	Glu	Asn	Ala	Pro	Gly	Gly	Pro	Tyr	Lys
[1154]	145					150					155				160	
[1155]	Ile	Gly	Thr	Ser	Gly	Ser	Cys	Pro	Asn	Ile	Thr	Asn	Gly	Asn	Gly	Phe
[1156]					165					170					175	
[1157]	Phe	Ala	Thr	Met	Ala	Trp	Ala	Val	Pro	Lys	Asn	Asp	Lys	Asn	Lys	Thr
[1158]				180					185					190		
[1159]	Ala	Thr	Asn	Pro	Leu	Thr	Ile	Glu	Val	Pro	Tyr	Ile	Cys	Thr	Glu	Gly
[1160]			195					200					205			
[1161]	Glu	Asp	Gln	Ile	Thr	Val	Trp	Gly	Phe	His	Ser	Asp	Asn	Glu	Thr	Gln
[1162]		210						215					220			
[1163]	Met	Ala	Lys	Leu	Tyr	Gly	Asp	Ser	Lys	Pro	Gln	Lys	Phe	Thr	Ser	Ser
[1164]	225					230					235				240	
[1165]	Ala	Asn	Gly	Val	Thr	Thr	His	Tyr	Val	Ser	Gln	Ile	Gly	Gly	Phe	Pro
[1166]					245					250					255	
[1167]	Asn	Gln	Thr	Glu	Asp	Gly	Gly	Leu	Pro	Gln	Ser	Gly	Arg	Ile	Val	Val
[1168]				260						265				270		
[1169]	Asp	Tyr	Met	Val	Gln	Lys	Ser	Gly	Lys	Thr	Gly	Thr	Ile	Thr	Tyr	Gln

[1170]	275	280	285
[1171]	Arg Gly Ile Leu Leu Pro Gln Lys Val Trp Cys Ala Ser Gly Arg Ser		
[1172]	290	295	300
[1173]	Lys Val Ile Lys Gly Ser Leu Pro Leu Ile Gly Glu Ala Asp Cys Leu		
[1174]	305	310	315
[1175]	His Glu Lys Tyr Gly Gly Leu Asn Lys Ser Lys Pro Tyr Tyr Thr Gly		
[1176]	325	330	335
[1177]	Glu His Ala Lys Ala Ile Gly Asn Cys Pro Ile Trp Val Lys Thr Pro		
[1178]	340	345	350
[1179]	Leu Lys Leu Ala Asn Gly Thr Lys Tyr Arg Pro Pro Gly Gly Gly Trp		
[1180]	355	360	365
[1181]	Glu Gly Met Ile Ala Gly Trp His Gly Tyr Thr Ser His Gly Ala His		
[1182]	370	375	380
[1183]	Gly Val Ala Val Ala Ala Asp Leu Lys Ser Thr Gln Glu Ala Ile Asn		
[1184]	385	390	395
[1185]	Lys Ile Thr Lys Asn Leu Asn Ser Leu Ser Glu Leu Glu Val Lys Asn		
[1186]	405	410	415
[1187]	Leu Gln Arg Leu Ser Gly Ala Met Asp Glu Leu His Asn Glu Ile Leu		
[1188]	420	425	430
[1189]	Glu Leu Asp Glu Lys Val Asp Asp Leu Arg Ala Asp Thr Ile Ser Ser		
[1190]	435	440	445
[1191]	Gln Ile Glu Leu Ala Val Leu Leu Ser Asn Glu Gly Ile Ile Asn Ser		
[1192]	450	455	460
[1193]	Glu Asp Glu His Leu Leu Ala Leu Glu Arg Lys Leu Lys Lys Met Leu		
[1194]	465	470	475
[1195]	Gly Pro Ser Ala Val Glu Ile Gly Asn Gly Cys Phe Glu Thr Lys His		
[1196]	485	490	495
[1197]	Lys Cys Asn Gln Thr Cys Leu Asp Arg Ile Ala Ala Gly Thr Phe Asp		
[1198]	500	505	510
[1199]	Ala Gly Glu Phe Ser Leu Pro Thr Phe Asp Ser Leu Asn Ile Thr Ala		
[1200]	515	520	525
[1201]	Ala Ser Leu Asn Asp Asp Gly Leu Asp Asn His Thr Ile Leu Leu Tyr		
[1202]	530	535	540
[1203]	Tyr Ser Thr Ala Ala Ser Ser Leu Ala Val Thr Leu Met Ile Ala Ile		
[1204]	545	550	555
[1205]	Phe Val Val Tyr Met Val Ser Arg Asp Asn Val Ser Cys Ser Ile Cys		
[1206]	565	570	575
[1207]	Leu		
[1208]	<210> 45		

[1209]	<211> 1734	
[1210]	<212> DNA	
[1211]	<213> 人工序列	
[1212]	<220>	
[1213]	<223> PDISP/HA B Brisbane (PrL-)+H1 California	
[1214]	TMCT的核苷酸序列	
[1215]	<400> 45	
[1216]	atggcgaaaa acgttgcgat tttcggctta ttgttttctc ttcttgtgtt ggttccttct	60
[1217]	cagatcttcg ccgatcgaat ctgcactgga ataacatcgt caaactcacc acatgtcgtc	120
[1218]	aaaactgcta ctcaaggga ggtcaatgtg actggtgtaa taccactgac aacaacaccc	180
[1219]	accaaattctc attttgcaaa tctcaaagga acagaaacca gggggaaact atgccccaaa	240
[1220]	tgcttcaact gcacagatct ggacgtagcc ttgggcagac caaatgcac ggggaaaata	300
[1221]	ccctcggcaa gagtttcaat actccatgaa gtcagacctg ttacatctgg gtgcttttct	360
[1222]	ataatgcacg acagaacaaa aattagacag ctgcctaacc ttctccgagg atacgaacat	420
[1223]	atcaggttat caaccataa cgttatcaat gcagaaaatg caccaggagg accctacaaa	480
[1224]	attggaacct caggtcttg ccctaacatt accaatggaa acggattttt cgcaacaatg	540
[1225]	gcttgggccg tccccaaaaa cgacaaaaac aaaacagcaa caaatccatt aacaatagaa	600
[1226]	gtaccataca tttgtacaga aggagaagac caaattaccg tttgggggtt ccactctgac	660
[1227]	aacgagaccc aaatggcaaa gctctatggg gactcaaagc ccagaagtt cacctcatct	720
[1228]	gccaacggag tgaccacaca ttacgtttca cagattgggtg gcttcccaaa tcaaacagaa	780
[1229]	gacggaggac taccacaaag tggtagaatt gttgttgatt acatggtgca aaaatctggg	840
[1230]	aaaacaggaa caattaccta tcaaaggggt attttattgc ctcaaaagggt gtggtgcgca	900
[1231]	agtggcagga gcaaggtaat aaaaggatcc ttgcctttaa ttggagaagc agattgcctc	960
[1232]	cacgaaaaat acggttgatt aaacaaaagc aagccttact acacagggga acatgcaaag	1020
[1233]	gccataggaa attgcccatt atgggtgaaa acacccttga agctggccaa tggaacaaa	1080
[1234]	tatagacctc ctggttgagg atgggaagga atgattgcag gttggcacgg atacacatcc	1140
[1235]	catggggcac atggagtagc ggtggcagca gaccttaaga gcactcaaga ggccataaac	1200
[1236]	aagataacaa aaaatctcaa ctcttttagt gagctggaag taaagaatct tcaaagacta	1260
[1237]	agcggtgcca tggatgaact ccacaacgaa atactagaac tagatgagaa agtggatgat	1320
[1238]	ctcagagctg atacaataag ctcaaaaata gaactcgag tcctgctttc caatgaagga	1380
[1239]	ataataaaca gtgaagatga acatctcttg gcgcttgaaa gaaagctgaa gaaaatgctg	1440
[1240]	ggccccctctg ctgtagagat agggaatgga tgctttgaaa ccaaacacaa gtgcaaccag	1500
[1241]	acctgtctcg acagaatagc tgctggtacc tttgatgcag gagaattttc tctccccacc	1560
[1242]	tttgattcac tgaatattac tgctgcatct ttaaattgacg atggattgga taattaccag	1620
[1243]	attttggcga tctattcaac tgctgccagt tcattggtac tggtagtctc cctgggggca	1680
[1244]	atcagtttct ggatgtgctc taatgggtct ctacagtgtg gaatatgtat ttaa	1734
[1245]	<210> 46	
[1246]	<211> 577	
[1247]	<212> PRT	

[1248]	<213> 人工序列															
[1249]	<220>															
[1250]	<223> PDISP/HA B Brisbane (PrL-)+H1 California															
[1251]	TMCT的氨基酸序列															
[1252]	<400> 46															
[1253]	Met	Ala	Lys	Asn	Val	Ala	Ile	Phe	Gly	Leu	Leu	Phe	Ser	Leu	Leu	Val
[1254]	1				5					10					15	
[1255]	Leu	Val	Pro	Ser	Gln	Ile	Phe	Ala	Asp	Arg	Ile	Cys	Thr	Gly	Ile	Thr
[1256]					20					25					30	
[1257]	Ser	Ser	Asn	Ser	Pro	His	Val	Val	Lys	Thr	Ala	Thr	Gln	Gly	Glu	Val
[1258]					35					40					45	
[1259]	Asn	Val	Thr	Gly	Val	Ile	Pro	Leu	Thr	Thr	Thr	Pro	Thr	Lys	Ser	His
[1260]					50					55					60	
[1261]	Phe	Ala	Asn	Leu	Lys	Gly	Thr	Glu	Thr	Arg	Gly	Lys	Leu	Cys	Pro	Lys
[1262]	65					70					75					80
[1263]	Cys	Leu	Asn	Cys	Thr	Asp	Leu	Asp	Val	Ala	Leu	Gly	Arg	Pro	Lys	Cys
[1264]					85					90						95
[1265]	Thr	Gly	Lys	Ile	Pro	Ser	Ala	Arg	Val	Ser	Ile	Leu	His	Glu	Val	Arg
[1266]					100					105					110	
[1267]	Pro	Val	Thr	Ser	Gly	Cys	Phe	Pro	Ile	Met	His	Asp	Arg	Thr	Lys	Ile
[1268]					115					120					125	
[1269]	Arg	Gln	Leu	Pro	Asn	Leu	Leu	Arg	Gly	Tyr	Glu	His	Ile	Arg	Leu	Ser
[1270]					130					135					140	
[1271]	Thr	His	Asn	Val	Ile	Asn	Ala	Glu	Asn	Ala	Pro	Gly	Gly	Pro	Tyr	Lys
[1272]	145					150					155				160	
[1273]	Ile	Gly	Thr	Ser	Gly	Ser	Cys	Pro	Asn	Ile	Thr	Asn	Gly	Asn	Gly	Phe
[1274]					165						170				175	
[1275]	Phe	Ala	Thr	Met	Ala	Trp	Ala	Val	Pro	Lys	Asn	Asp	Lys	Asn	Lys	Thr
[1276]					180					185					190	
[1277]	Ala	Thr	Asn	Pro	Leu	Thr	Ile	Glu	Val	Pro	Tyr	Ile	Cys	Thr	Glu	Gly
[1278]					195					200					205	
[1279]	Glu	Asp	Gln	Ile	Thr	Val	Trp	Gly	Phe	His	Ser	Asp	Asn	Glu	Thr	Gln
[1280]					210					215					220	
[1281]	Met	Ala	Lys	Leu	Tyr	Gly	Asp	Ser	Lys	Pro	Gln	Lys	Phe	Thr	Ser	Ser
[1282]	225					230					235				240	
[1283]	Ala	Asn	Gly	Val	Thr	Thr	His	Tyr	Val	Ser	Gln	Ile	Gly	Gly	Phe	Pro
[1284]					245					250					255	
[1285]	Asn	Gln	Thr	Glu	Asp	Gly	Gly	Leu	Pro	Gln	Ser	Gly	Arg	Ile	Val	Val
[1286]					260					265					270	

[1287]	Asp	Tyr	Met	Val	Gln	Lys	Ser	Gly	Lys	Thr	Gly	Thr	Ile	Thr	Tyr	Gln
[1288]			275					280					285			
[1289]	Arg	Gly	Ile	Leu	Leu	Pro	Gln	Lys	Val	Trp	Cys	Ala	Ser	Gly	Arg	Ser
[1290]			290					295					300			
[1291]	Lys	Val	Ile	Lys	Gly	Ser	Leu	Pro	Leu	Ile	Gly	Glu	Ala	Asp	Cys	Leu
[1292]	305						310				315					320
[1293]	His	Glu	Lys	Tyr	Gly	Gly	Leu	Asn	Lys	Ser	Lys	Pro	Tyr	Tyr	Thr	Gly
[1294]					325					330						335
[1295]	Glu	His	Ala	Lys	Ala	Ile	Gly	Asn	Cys	Pro	Ile	Trp	Val	Lys	Thr	Pro
[1296]				340					345					350		
[1297]	Leu	Lys	Leu	Ala	Asn	Gly	Thr	Lys	Tyr	Arg	Pro	Pro	Gly	Gly	Gly	Trp
[1298]			355					360					365			
[1299]	Glu	Gly	Met	Ile	Ala	Gly	Trp	His	Gly	Tyr	Thr	Ser	His	Gly	Ala	His
[1300]	370						375					380				
[1301]	Gly	Val	Ala	Val	Ala	Ala	Asp	Leu	Lys	Ser	Thr	Gln	Glu	Ala	Ile	Asn
[1302]	385					390					395					400
[1303]	Lys	Ile	Thr	Lys	Asn	Leu	Asn	Ser	Leu	Ser	Glu	Leu	Glu	Val	Lys	Asn
[1304]					405					410						415
[1305]	Leu	Gln	Arg	Leu	Ser	Gly	Ala	Met	Asp	Glu	Leu	His	Asn	Glu	Ile	Leu
[1306]				420					425					430		
[1307]	Glu	Leu	Asp	Glu	Lys	Val	Asp	Asp	Leu	Arg	Ala	Asp	Thr	Ile	Ser	Ser
[1308]			435					440					445			
[1309]	Gln	Ile	Glu	Leu	Ala	Val	Leu	Leu	Ser	Asn	Glu	Gly	Ile	Ile	Asn	Ser
[1310]	450						455					460				
[1311]	Glu	Asp	Glu	His	Leu	Leu	Ala	Leu	Glu	Arg	Lys	Leu	Lys	Lys	Met	Leu
[1312]	465					470					475					480
[1313]	Gly	Pro	Ser	Ala	Val	Glu	Ile	Gly	Asn	Gly	Cys	Phe	Glu	Thr	Lys	His
[1314]					485					490					495	
[1315]	Lys	Cys	Asn	Gln	Thr	Cys	Leu	Asp	Arg	Ile	Ala	Ala	Gly	Thr	Phe	Asp
[1316]				500					505				510			
[1317]	Ala	Gly	Glu	Phe	Ser	Leu	Pro	Thr	Phe	Asp	Ser	Leu	Asn	Ile	Thr	Ala
[1318]			515						520				525			
[1319]	Ala	Ser	Leu	Asn	Asp	Asp	Gly	Leu	Asp	Asn	Tyr	Gln	Ile	Leu	Ala	Ile
[1320]	530						535					540				
[1321]	Tyr	Ser	Thr	Val	Ala	Ser	Ser	Leu	Val	Leu	Val	Val	Ser	Leu	Gly	Ala
[1322]	545					550					555					560
[1323]	Ile	Ser	Phe	Trp	Met	Cys	Ser	Asn	Gly	Ser	Leu	Gln	Cys	Arg	Ile	Cys
[1324]					565					570					575	
[1325]	Ile															

[1326]	<210> 47	
[1327]	<211> 1731	
[1328]	<212> DNA	
[1329]	<213> 人工序列	
[1330]	<220>	
[1331]	<223> PDISP/HA B Massachussetts (PrL-)的核苷酸序列	
[1332]	<400> 47	
[1333]	atggcgaaaa acgttgcgat tttcggctta ttgttttctc ttcttgtgtt ggttccttct	60
[1334]	cagatcttcg ccgatcgaat ctgcactggg ataacatctt caaactcacc tcatgtggtc	120
[1335]	aaaacagcta ctcaaggga ggtcaatgtg actggtgtga taccactaac aacaacacca	180
[1336]	acaaaatctt attttgcaaa tctcaaagga acaagacca gagggaaact atgccagac	240
[1337]	tgtctcaact gtacagatct ggatgtggcc ctgggcaggc caatgtgtgt gggaactaca	300
[1338]	ccttctgcga aagcttcaat acttcacgaa gtcagacctg ttacatccgg gtgtcttcct	360
[1339]	ataatgcacg acagaacaaa aatcaggcaa ctagccaatc ttctcagagg atatgaaaat	420
[1340]	atcaggttat caacccaaaa cgttatcgat gcagaaaagg caccaggagg accctacaga	480
[1341]	cttgaacct caggatcttg ccctaacgct accagtaaaa gcggattttt cgcaacaatg	540
[1342]	gcttgggctg tcccaaagga caacaacaaa aatgcaacga acccattaac agtagaagta	600
[1343]	ccatacattt gtgcagaagg ggaagaccaa attactgttt gggggttcca ttcagataac	660
[1344]	aaaacccaaa tgaagaacct ctatggagac tcaaatcctc aaaagttcac ctcatctgct	720
[1345]	aatggagtaa ccacacatta tgtttctcag attggcggct tcccagatca aacagaagac	780
[1346]	ggaggactac cacaaagcgg cagaattgtc gttgattaca tgatgcaaaa acctgggaaa	840
[1347]	acaggaacaa ttgtctatca aagaggtgtt ttgttgccctc aaaaggtgtg gtgcgcgagt	900
[1348]	ggcaggagca aagtaataaa agggtccttg cttttaattg gtgaagcaga ttgccttcat	960
[1349]	gaaaaatacg gtggattaaa caaaagcaag cttactaca caggagaaca tgcaaaagcc	1020
[1350]	ataggaaatt gcccaatatg ggtgaaaaca ctttgaagc ttgccaatgg aaccaaatat	1080
[1351]	agacctctg gtggaggatg ggaaggaatg attgcaggtt ggcacggata cacatctcac	1140
[1352]	ggagcacatg gagtggcagt tgctgcagac ctaagagca cacaagaagc tataaacaag	1200
[1353]	ataacaaaa atctcaactc tttgagttag ctagaagtaa agaattctca aaggctaagt	1260
[1354]	ggtgccatgg atgaactcca caacgaaata ctgcagctgg atgagaaagt ggatgacctc	1320
[1355]	agagctgaca ctataagttc acaaatagaa cttgcagtct tgctttccaa cgaaggaata	1380
[1356]	ataaacagtg aagacgagca tctattggca cttgagagaa aactaaagaa aatgctgggt	1440
[1357]	ccctctgctg tagacatagg aaatggatgc ttcgaaacca aacacaaatg caaccagacc	1500
[1358]	tgcttagaca ggatagctgc tggcaccttt aatgcaggag agttttctct cccactttt	1560
[1359]	gattcattga acattactgc tgcattctta aatgatgatg gattggataa ccatactata	1620
[1360]	ctgctctatt actcaactgc tgcttctagt ttggctgtaa cattgatget agctattttt	1680
[1361]	attgtttata tggctccag agacaacgtt tcatgtcca tctgtctata a	1731
[1362]	<210> 48	
[1363]	<211> 576	
[1364]	<212> PRT	

[1365]	<213> 人工序列																		
[1366]	<220>																		
[1367]	<223> PDISP/HA B Massachussetts (PrL-)的氨基酸序列																		
[1368]	<400> 48																		
[1369]	Met	Ala	Lys	Asn	Val	Ala	Ile	Phe	Gly	Leu	Leu	Phe	Ser	Leu	Leu	Val			
[1370]	1				5					10					15				
[1371]	Leu	Val	Pro	Ser	Gln	Ile	Phe	Ala	Asp	Arg	Ile	Cys	Thr	Gly	Ile	Thr			
[1372]					20					25					30				
[1373]	Ser	Ser	Asn	Ser	Pro	His	Val	Val	Lys	Thr	Ala	Thr	Gln	Gly	Glu	Val			
[1374]					35					40					45				
[1375]	Asn	Val	Thr	Gly	Val	Ile	Pro	Leu	Thr	Thr	Thr	Pro	Thr	Lys	Ser	Tyr			
[1376]		50						55					60						
[1377]	Phe	Ala	Asn	Leu	Lys	Gly	Thr	Lys	Thr	Arg	Gly	Lys	Leu	Cys	Pro	Asp			
[1378]	65					70					75					80			
[1379]	Cys	Leu	Asn	Cys	Thr	Asp	Leu	Asp	Val	Ala	Leu	Gly	Arg	Pro	Met	Cys			
[1380]					85					90					95				
[1381]	Val	Gly	Thr	Thr	Pro	Ser	Ala	Lys	Ala	Ser	Ile	Leu	His	Glu	Val	Arg			
[1382]					100					105					110				
[1383]	Pro	Val	Thr	Ser	Gly	Cys	Phe	Pro	Ile	Met	His	Asp	Arg	Thr	Lys	Ile			
[1384]					115					120					125				
[1385]	Arg	Gln	Leu	Ala	Asn	Leu	Leu	Arg	Gly	Tyr	Glu	Asn	Ile	Arg	Leu	Ser			
[1386]		130						135					140						
[1387]	Thr	Gln	Asn	Val	Ile	Asp	Ala	Glu	Lys	Ala	Pro	Gly	Gly	Pro	Tyr	Arg			
[1388]	145					150					155					160			
[1389]	Leu	Gly	Thr	Ser	Gly	Ser	Cys	Pro	Asn	Ala	Thr	Ser	Lys	Ser	Gly	Phe			
[1390]					165					170					175				
[1391]	Phe	Ala	Thr	Met	Ala	Trp	Ala	Val	Pro	Lys	Asp	Asn	Asn	Lys	Asn	Ala			
[1392]					180					185					190				
[1393]	Thr	Asn	Pro	Leu	Thr	Val	Glu	Val	Pro	Tyr	Ile	Cys	Ala	Glu	Gly	Glu			
[1394]					195					200					205				
[1395]	Asp	Gln	Ile	Thr	Val	Trp	Gly	Phe	His	Ser	Asp	Asn	Lys	Thr	Gln	Met			
[1396]		210						215					220						
[1397]	Lys	Asn	Leu	Tyr	Gly	Asp	Ser	Asn	Pro	Gln	Lys	Phe	Thr	Ser	Ser	Ala			
[1398]	225					230					235					240			
[1399]	Asn	Gly	Val	Thr	Thr	His	Tyr	Val	Ser	Gln	Ile	Gly	Gly	Phe	Pro	Asp			
[1400]					245					250					255				
[1401]	Gln	Thr	Glu	Asp	Gly	Gly	Leu	Pro	Gln	Ser	Gly	Arg	Ile	Val	Val	Asp			
[1402]					260					265					270				
[1403]	Tyr	Met	Met	Gln	Lys	Pro	Gly	Lys	Thr	Gly	Thr	Ile	Val	Tyr	Gln	Arg			

[1404]	275	280	285
[1405]	Gly Val Leu Leu Pro Gln Lys Val Trp Cys Ala Ser Gly Arg Ser Lys		
[1406]	290	295	300
[1407]	Val Ile Lys Gly Ser Leu Pro Leu Ile Gly Glu Ala Asp Cys Leu His		
[1408]	305	310	315
[1409]	Glu Lys Tyr Gly Gly Leu Asn Lys Ser Lys Pro Tyr Tyr Thr Gly Glu		
[1410]	325	330	335
[1411]	His Ala Lys Ala Ile Gly Asn Cys Pro Ile Trp Val Lys Thr Pro Leu		
[1412]	340	345	350
[1413]	Lys Leu Ala Asn Gly Thr Lys Tyr Arg Pro Pro Gly Gly Gly Trp Glu		
[1414]	355	360	365
[1415]	Gly Met Ile Ala Gly Trp His Gly Tyr Thr Ser His Gly Ala His Gly		
[1416]	370	375	380
[1417]	Val Ala Val Ala Ala Asp Leu Lys Ser Thr Gln Glu Ala Ile Asn Lys		
[1418]	385	390	395
[1419]	Ile Thr Lys Asn Leu Asn Ser Leu Ser Glu Leu Glu Val Lys Asn Leu		
[1420]	405	410	415
[1421]	Gln Arg Leu Ser Gly Ala Met Asp Glu Leu His Asn Glu Ile Leu Glu		
[1422]	420	425	430
[1423]	Leu Asp Glu Lys Val Asp Asp Leu Arg Ala Asp Thr Ile Ser Ser Gln		
[1424]	435	440	445
[1425]	Ile Glu Leu Ala Val Leu Leu Ser Asn Glu Gly Ile Ile Asn Ser Glu		
[1426]	450	455	460
[1427]	Asp Glu His Leu Leu Ala Leu Glu Arg Lys Leu Lys Lys Met Leu Gly		
[1428]	465	470	475
[1429]	Pro Ser Ala Val Asp Ile Gly Asn Gly Cys Phe Glu Thr Lys His Lys		
[1430]	485	490	495
[1431]	Cys Asn Gln Thr Cys Leu Asp Arg Ile Ala Ala Gly Thr Phe Asn Ala		
[1432]	500	505	510
[1433]	Gly Glu Phe Ser Leu Pro Thr Phe Asp Ser Leu Asn Ile Thr Ala Ala		
[1434]	515	520	525
[1435]	Ser Leu Asn Asp Asp Gly Leu Asp Asn His Thr Ile Leu Leu Tyr Tyr		
[1436]	530	535	540
[1437]	Ser Thr Ala Ala Ser Ser Leu Ala Val Thr Leu Met Leu Ala Ile Phe		
[1438]	545	550	555
[1439]	Ile Val Tyr Met Val Ser Arg Asp Asn Val Ser Cys Ser Ile Cys Leu		
[1440]	565	570	575
[1441]	<210> 49		
[1442]	<211> 1731		

[1443]	<212> DNA	
[1444]	<213> 人工序列	
[1445]	<220>	
[1446]	<223> PDISP/HA B Massachussetts (PrL-)+H1	
[1447]	California TMCT的核苷酸序列	
[1448]	<400> 49	
[1449]	atggcgaaaa acgttgcgat tttcggctta ttgttttctc ttcttgtgtt ggttccttct	60
[1450]	cagatcttcg ccgatcgaat ctgcactggg ataacatctt caaactcacc tcattgtggtc	120
[1451]	aaaacagcta ctcaagggga ggtcaatgtg actggtgtga taccactaac aacaacacca	180
[1452]	acaaaatctt attttgcaaa tctcaaagga acaaagacca gagggaaact atgcccagac	240
[1453]	tgtctcaact gtacagatct ggatgtggcc ctgggcaggc caatgtgtgt gggaactaca	300
[1454]	ccttctgcga aagcttcaat acttcacgaa gtcagacctg ttacatccgg gtgcttcct	360
[1455]	ataatgcacg acagaacaaa aatcaggcaa ctagccaatc ttctcagagg atatgaaaat	420
[1456]	atcaggttat caacccaaaa cgttatcgat gcagaaaagg caccaggagg accctacaga	480
[1457]	cttggaaact caggatcttg ccctaacgct accagtaaaa gcggattttt cgcaacaatg	540
[1458]	gcttgggctg tcccaaagga caacaacaaa aatgcaacga acccattaac agtagaagta	600
[1459]	ccatacatth gtgcagaagg ggaagaccaa attactgttt gggggttcca ttcagataac	660
[1460]	aaaacccaaa tgaagaacct ctatggagac tcaaatctc aaaagttcac ctcatctgct	720
[1461]	aatggagtaa ccacacatta tgtttctcag attggcggct tcccagatca aacagaagac	780
[1462]	ggaggactac cacaagcgg cagaattgtc gttgattaca tgatgcaaaa acctgggaaa	840
[1463]	acaggaacaa ttgtctatca aagaggtgtt ttgttgccctc aaaaggtgtg gtgcgcgagt	900
[1464]	ggcaggagca aagtaataaa agggctcctg cctttaattg gtgaagcaga ttgccttcat	960
[1465]	gaaaaatagc gtggattaaa caaaagcaag cttactaca caggagaaca tgcaaaagcc	1020
[1466]	ataggaaatt gcccaatatg ggtgaaaaca ctttgaagc ttgccaatgg aaccaaata	1080
[1467]	agacctctg gtggaggatg ggaaggaatg attgcaggtt ggcacggata cacatctcac	1140
[1468]	ggagcacatg gattggcagt tgctgcagac ctaagagca cacaagaagc tataaacaag	1200
[1469]	ataacaaaa atctcaactc tttgagttag ctagaagtaa agaattctca aaggctaagt	1260
[1470]	ggtgccatgg atgaactcca caacgaata ctgcagctgg atgagaaagt ggatgacctc	1320
[1471]	agagctgaca ctataagttc acaaatagaa cttgcagtct tgctttccaa cgaaggaata	1380
[1472]	ataaacagtg aagacgagca tctattggca cttgagagaa aactaaagaa aatgctgggt	1440
[1473]	ccctctgctg tagacatagg aatggatgc ttcgaaacca aacacaaatg caaccagacc	1500
[1474]	tgcttagaca ggatagctgc tggcaccttt aatgcaggag agttttctct cccactttt	1560
[1475]	gattcattga acattactgc tgcattctta aatgatgatg gattggataa ctaccagatt	1620
[1476]	ttggcgatct attcaactgt cgccagttca ttggtactgg tagtctccct gggggcaatc	1680
[1477]	agtttctgga tgtgtcttaa tgggtctcta cagtgtagaa tatgtattta a	1731
[1478]	<210> 50	
[1479]	<211> 576	
[1480]	<212> PRT	
[1481]	<213> 人工序列	

[1521]	275	280	285
[1522]	Gly Val Leu Leu Pro Gln Lys Val Trp Cys Ala Ser Gly Arg Ser Lys		
[1523]	290	295	300
[1524]	Val Ile Lys Gly Ser Leu Pro Leu Ile Gly Glu Ala Asp Cys Leu His		
[1525]	305	310	315
[1526]	Glu Lys Tyr Gly Gly Leu Asn Lys Ser Lys Pro Tyr Tyr Thr Gly Glu		
[1527]	325	330	335
[1528]	His Ala Lys Ala Ile Gly Asn Cys Pro Ile Trp Val Lys Thr Pro Leu		
[1529]	340	345	350
[1530]	Lys Leu Ala Asn Gly Thr Lys Tyr Arg Pro Pro Gly Gly Gly Trp Glu		
[1531]	355	360	365
[1532]	Gly Met Ile Ala Gly Trp His Gly Tyr Thr Ser His Gly Ala His Gly		
[1533]	370	375	380
[1534]	Val Ala Val Ala Ala Asp Leu Lys Ser Thr Gln Glu Ala Ile Asn Lys		
[1535]	385	390	395
[1536]	Ile Thr Lys Asn Leu Asn Ser Leu Ser Glu Leu Glu Val Lys Asn Leu		
[1537]	405	410	415
[1538]	Gln Arg Leu Ser Gly Ala Met Asp Glu Leu His Asn Glu Ile Leu Glu		
[1539]	420	425	430
[1540]	Leu Asp Glu Lys Val Asp Asp Leu Arg Ala Asp Thr Ile Ser Ser Gln		
[1541]	435	440	445
[1542]	Ile Glu Leu Ala Val Leu Leu Ser Asn Glu Gly Ile Ile Asn Ser Glu		
[1543]	450	455	460
[1544]	Asp Glu His Leu Leu Ala Leu Glu Arg Lys Leu Lys Lys Met Leu Gly		
[1545]	465	470	475
[1546]	Pro Ser Ala Val Asp Ile Gly Asn Gly Cys Phe Glu Thr Lys His Lys		
[1547]	485	490	495
[1548]	Cys Asn Gln Thr Cys Leu Asp Arg Ile Ala Ala Gly Thr Phe Asn Ala		
[1549]	500	505	510
[1550]	Gly Glu Phe Ser Leu Pro Thr Phe Asp Ser Leu Asn Ile Thr Ala Ala		
[1551]	515	520	525
[1552]	Ser Leu Asn Asp Asp Gly Leu Asp Asn Tyr Gln Ile Leu Ala Ile Tyr		
[1553]	530	535	540
[1554]	Ser Thr Val Ala Ser Ser Leu Val Leu Val Val Ser Leu Gly Ala Ile		
[1555]	545	550	555
[1556]	Ser Phe Trp Met Cys Ser Asn Gly Ser Leu Gln Cys Arg Ile Cys Ile		
[1557]	565	570	575
[1558]	<210> 51		
[1559]	<211> 1704		

[1560]	<212> DNA	
[1561]	<213> 人工序列	
[1562]	<220>	
[1563]	<223> HA B Wisconsin (PrL-) 的核苷酸序列	
[1564]	<400> 51	
[1565]	atgaaggcaa taattgtact actcatggta gtaacatcca atgcagatcg aatctgcact	60
[1566]	gggataacat cttcaaactc acctcatgtg gtcaaaacag ctactcaagg ggaggtcaat	120
[1567]	gtgactggcg tgataccact gacaacaaca ccaacaaaat cttattttgc aaatctcaaa	180
[1568]	ggaacaagga ccagagggaa actatgcccg gactgtctca actgtacaga tctggatgtg	240
[1569]	gccttgggca ggccaatgtg tgtggggacc acaccttctg ctaaagcttc aatactccac	300
[1570]	gaggtcagac ctgttacatc cgggtgcttt cctataatgc acgacagaac aaaaatcagg	360
[1571]	caactacca atcttctcag aggatatgaa aatatcaggt tatcaacca aaacgttata	420
[1572]	gatgcagaaa aagcaccagg aggaccctac agacttgga cctcaggatc ttgcctaac	480
[1573]	gctaccagta aaatcggatt ttttgcaaca atggcttggg ctgtccaaa ggacaactac	540
[1574]	aaaaatgcaa cgaaccact aacagtagaa gtaccataca tttgtacaga aggggaagac	600
[1575]	caaattactg tttgggggtt ccattcagat aacaaaacc aaatgaagag cctctatgga	660
[1576]	gactcaaate ctcaaaagt caccctatct gctaattggag taaccacaca ttatgtttct	720
[1577]	cagattggcg acttcccaga tcaaacagaa gacggaggac taccacaaag cggcagaatt	780
[1578]	gttgttgatt acatgatgca aaaacctggg aaaacaggaa caattgtcta tcaaagaggt	840
[1579]	gttttgttgc ctcaaaaggt gtggtgcgcg agtggcagga gcaaagtaat aaaagggtca	900
[1580]	ttgcctttaa ttggtgaagc agattgcctt catgaaaaat acggtggatt aaacaaaagc	960
[1581]	aagccttact acacaggaga acatgcaaaa gccataggaa attgccaat atgggtaaaa	1020
[1582]	acaccttga agcttgccaa tggaaccaa tatagacctc ctggtggagg atgggaagga	1080
[1583]	atgattgcag gttggcacgg atacacatct cacggagcac atggagtggc agtggcggca	1140
[1584]	gaccttaaga gtacacaaga agctataaat aagataacaa aaaatctcaa ttctttgagt	1200
[1585]	gagctagaag taaagaacct tcaaagacta agtggtgcca tggatgaact ccacaacgaa	1260
[1586]	atactcgagc tggatgagaa agtggatgat ctcagagctg acactataag ctacaaata	1320
[1587]	gaacttgcag tcttgctttc caacgaagga ataataaaca gtgaagacga gcatctattg	1380
[1588]	gcacttgaga gaaaactaaa gaaaatgctg ggtccctctg ctgtagacat aggaaacgga	1440
[1589]	tgcttcgaaa ccaaacacaa atgcaaccag acctgcttag acaggatagc tgctggcacc	1500
[1590]	tttaatgcag gagaattttc tctccccact tttgattcat tgaacattac tgctgcatct	1560
[1591]	ttaaatgatg atggattgga taaccatact atactgtct attactcaac tgctgcttct	1620
[1592]	agtttggctg taacattaat gctagctatt tttattgttt atatggtctc cagagacaac	1680
[1593]	gtttcatget ccatctgtct ataa	1704
[1594]	<210> 52	
[1595]	<211> 567	
[1596]	<212> PRT	
[1597]	<213> 人工序列	
[1598]	<220>	

[1599] <223> HA B Wisconsin (PrL-) 的氨基酸序列

[1600] <400> 52

[1601] Met Lys Ala Ile Ile Val Leu Leu Met Val Val Thr Ser Asn Ala Asp

[1602] 1 5 10 15

[1603] Arg Ile Cys Thr Gly Ile Thr Ser Ser Asn Ser Pro His Val Val Lys

[1604] 20 25 30

[1605] Thr Ala Thr Gln Gly Glu Val Asn Val Thr Gly Val Ile Pro Leu Thr

[1606] 35 40 45

[1607] Thr Thr Pro Thr Lys Ser Tyr Phe Ala Asn Leu Lys Gly Thr Arg Thr

[1608] 50 55 60

[1609] Arg Gly Lys Leu Cys Pro Asp Cys Leu Asn Cys Thr Asp Leu Asp Val

[1610] 65 70 75 80

[1611] Ala Leu Gly Arg Pro Met Cys Val Gly Thr Thr Pro Ser Ala Lys Ala

[1612] 85 90 95

[1613] Ser Ile Leu His Glu Val Arg Pro Val Thr Ser Gly Cys Phe Pro Ile

[1614] 100 105 110

[1615] Met His Asp Arg Thr Lys Ile Arg Gln Leu Pro Asn Leu Leu Arg Gly

[1616] 115 120 125

[1617] Tyr Glu Asn Ile Arg Leu Ser Thr Gln Asn Val Ile Asp Ala Glu Lys

[1618] 130 135 140

[1619] Ala Pro Gly Gly Pro Tyr Arg Leu Gly Thr Ser Gly Ser Cys Pro Asn

[1620] 145 150 155 160

[1621] Ala Thr Ser Lys Ile Gly Phe Phe Ala Thr Met Ala Trp Ala Val Pro

[1622] 165 170 175

[1623] Lys Asp Asn Tyr Lys Asn Ala Thr Asn Pro Leu Thr Val Glu Val Pro

[1624] 180 185 190

[1625] Tyr Ile Cys Thr Glu Gly Glu Asp Gln Ile Thr Val Trp Gly Phe His

[1626] 195 200 205

[1627] Ser Asp Asn Lys Thr Gln Met Lys Ser Leu Tyr Gly Asp Ser Asn Pro

[1628] 210 215 220

[1629] Gln Lys Phe Thr Ser Ser Ala Asn Gly Val Thr Thr His Tyr Val Ser

[1630] 225 230 235 240

[1631] Gln Ile Gly Asp Phe Pro Asp Gln Thr Glu Asp Gly Gly Leu Pro Gln

[1632] 245 250 255

[1633] Ser Gly Arg Ile Val Val Asp Tyr Met Met Gln Lys Pro Gly Lys Thr

[1634] 260 265 270

[1635] Gly Thr Ile Val Tyr Gln Arg Gly Val Leu Leu Pro Gln Lys Val Trp

[1636] 275 280 285

[1637] Cys Ala Ser Gly Arg Ser Lys Val Ile Lys Gly Ser Leu Pro Leu Ile

[1638]	290	295	300
[1639]	Gly Glu Ala Asp Cys Leu His Glu Lys Tyr Gly Gly Leu Asn Lys Ser		
[1640]	305	310	315
[1641]	Lys Pro Tyr Tyr Thr Gly Glu His Ala Lys Ala Ile Gly Asn Cys Pro		
[1642]	325	330	335
[1643]	Ile Trp Val Lys Thr Pro Leu Lys Leu Ala Asn Gly Thr Lys Tyr Arg		
[1644]	340	345	350
[1645]	Pro Pro Gly Gly Gly Trp Glu Gly Met Ile Ala Gly Trp His Gly Tyr		
[1646]	355	360	365
[1647]	Thr Ser His Gly Ala His Gly Val Ala Val Ala Ala Asp Leu Lys Ser		
[1648]	370	375	380
[1649]	Thr Gln Glu Ala Ile Asn Lys Ile Thr Lys Asn Leu Asn Ser Leu Ser		
[1650]	385	390	395
[1651]	Glu Leu Glu Val Lys Asn Leu Gln Arg Leu Ser Gly Ala Met Asp Glu		
[1652]	405	410	415
[1653]	Leu His Asn Glu Ile Leu Glu Leu Asp Glu Lys Val Asp Asp Leu Arg		
[1654]	420	425	430
[1655]	Ala Asp Thr Ile Ser Ser Gln Ile Glu Leu Ala Val Leu Leu Ser Asn		
[1656]	435	440	445
[1657]	Glu Gly Ile Ile Asn Ser Glu Asp Glu His Leu Leu Ala Leu Glu Arg		
[1658]	450	455	460
[1659]	Lys Leu Lys Lys Met Leu Gly Pro Ser Ala Val Asp Ile Gly Asn Gly		
[1660]	465	470	475
[1661]	Cys Phe Glu Thr Lys His Lys Cys Asn Gln Thr Cys Leu Asp Arg Ile		
[1662]	485	490	495
[1663]	Ala Ala Gly Thr Phe Asn Ala Gly Glu Phe Ser Leu Pro Thr Phe Asp		
[1664]	500	505	510
[1665]	Ser Leu Asn Ile Thr Ala Ala Ser Leu Asn Asp Asp Gly Leu Asp Asn		
[1666]	515	520	525
[1667]	His Thr Ile Leu Leu Tyr Tyr Ser Thr Ala Ala Ser Ser Leu Ala Val		
[1668]	530	535	540
[1669]	Thr Leu Met Leu Ala Ile Phe Ile Val Tyr Met Val Ser Arg Asp Asn		
[1670]	545	550	555
[1671]	Val Ser Cys Ser Ile Cys Leu		
[1672]	565		
[1673]	<210> 53		
[1674]	<211> 1704		
[1675]	<212> DNA		
[1676]	<213> 人工序列		

- [1677] <220>
- [1678] <223> HA B Wisconsin (PrL-)+H1 California TMCT的核苷酸序列
- [1679] <400> 53
- [1680] atgaaggcaa taattgtact actcatggta gtaacatcca atgcagatcg aatctgcact 60
- [1681] gggataacat cttcaaactc acctcatgtg gtcaaaacag ctactcaagg ggaggtcaat 120
- [1682] gtgactggcg tgataccact gacaacaaca ccaacaaaat cttattttgc aaatctcaaa 180
- [1683] ggaacaagga ccagagggaa actatgcccc gactgtctca actgtacaga tctggatgtg 240
- [1684] gccttgggca ggccaatgtg tgtggggacc acaccttctg ctaaagcttc aatactccac 300
- [1685] gaggtcagac ctgttacatc cgggtgcttt cctataatgc acgacagAAC aaaaatcagg 360
- [1686] caactaccca atcttctcag aggatatgaa aatatcaggT tatcaaccca aaacgttatc 420
- [1687] gatgcagaaa aagcaccagg aggaccctac agacttggaa cctcaggatc ttgcctaac 480
- [1688] gctaccagta aaatcggatt ttttgcAaca atggcttggg ctgtcccaaa ggacaactac 540
- [1689] aaaaatgcaa cgaaccact aacagtagaa gtaccataca tttgtacaga aggggaagac 600
- [1690] caaattactg tttgggggtt ccattcagat acaaaaacc aaatgaagag cctctatgga 660
- [1691] gactcaaate ctcaaaagtT cacctcatct gctaattggag taaccacaca ttatgtttct 720
- [1692] cagattggcg acttcccaga tcaaacagaa gacggaggac taccacaaag cggcagaatt 780
- [1693] gttgttgatt acatgatgca aaaacctggg aaacaggaa caattgtcta tcaaagaggt 840
- [1694] gttttgttgc ctcaaaaggt gtggtgcgcg agtggcagga gcaaagtaat aaaagggtca 900
- [1695] ttgcctttaa ttggtgaagc agattgcctt catgaaaaat acggtggatt aaacaaaagc 960
- [1696] aagccttact acacaggaga acatgcaaaa gccataggaa attgccaat atgggtaaaa 1020
- [1697] acaccttga agcttgccaa tggaaccaa tatagacctc ctggtggagg atgggaagga 1080
- [1698] atgattgcag gttggcacgg atacacatct cacggagcac atggagtggc agtggcggca 1140
- [1699] gaccttaaga gtacacaaga agctataaat aagataacaa aaaatctcaa ttctttgagt 1200
- [1700] gagctagaag taaagaacct tcaaagacta agtggtgcca tggatgaact ccacaacgaa 1260
- [1701] atactcgagc tggatgagaa agtggatgat ctcagagctg acactataag ctcaaaata 1320
- [1702] gaacttgcag tcttgctttc caacgaagga ataataaaca gtgaagacga gcatctattg 1380
- [1703] gcacttgaga gaaaactaaa gaaaatgctg gtcctctctg ctgtagacat aggaaacgga 1440
- [1704] tgcttcgaaa ccaaacacaa atgcaaccag acctgcttag acaggatagc tgctggcacc 1500
- [1705] tttaatgcag gagaattttc tctccccact tttgattcat tgaacattac tgctgcatct 1560
- [1706] ttaaattgat atggattgga taactaccag attttggcga tctattcaac tgctgccagt 1620
- [1707] tcattggtac tggtagtctc cctgggggca atcagtttct ggatgtgctc taatgggtct 1680
- [1708] ctacagtgta gaatatgtat ttaa 1704
- [1709] <210> 54
- [1710] <211> 567
- [1711] <212> PRT
- [1712] <213> 人工序列
- [1713] <220>
- [1714] <223> HA B Wisconsin (PrL-)+H1 California TMC的氨基酸序列
- [1715] <400> 54

[1716]	Met	Lys	Ala	Ile	Ile	Val	Leu	Leu	Met	Val	Val	Thr	Ser	Asn	Ala	Asp
[1717]	1				5					10					15	
[1718]	Arg	Ile	Cys	Thr	Gly	Ile	Thr	Ser	Ser	Asn	Ser	Pro	His	Val	Val	Lys
[1719]				20					25					30		
[1720]	Thr	Ala	Thr	Gln	Gly	Glu	Val	Asn	Val	Thr	Gly	Val	Ile	Pro	Leu	Thr
[1721]				35					40					45		
[1722]	Thr	Thr	Pro	Thr	Lys	Ser	Tyr	Phe	Ala	Asn	Leu	Lys	Gly	Thr	Arg	Thr
[1723]		50					55					60				
[1724]	Arg	Gly	Lys	Leu	Cys	Pro	Asp	Cys	Leu	Asn	Cys	Thr	Asp	Leu	Asp	Val
[1725]	65					70					75				80	
[1726]	Ala	Leu	Gly	Arg	Pro	Met	Cys	Val	Gly	Thr	Thr	Pro	Ser	Ala	Lys	Ala
[1727]					85					90					95	
[1728]	Ser	Ile	Leu	His	Glu	Val	Arg	Pro	Val	Thr	Ser	Gly	Cys	Phe	Pro	Ile
[1729]				100					105					110		
[1730]	Met	His	Asp	Arg	Thr	Lys	Ile	Arg	Gln	Leu	Pro	Asn	Leu	Leu	Arg	Gly
[1731]				115					120					125		
[1732]	Tyr	Glu	Asn	Ile	Arg	Leu	Ser	Thr	Gln	Asn	Val	Ile	Asp	Ala	Glu	Lys
[1733]		130						135					140			
[1734]	Ala	Pro	Gly	Gly	Pro	Tyr	Arg	Leu	Gly	Thr	Ser	Gly	Ser	Cys	Pro	Asn
[1735]	145					150					155				160	
[1736]	Ala	Thr	Ser	Lys	Ile	Gly	Phe	Phe	Ala	Thr	Met	Ala	Trp	Ala	Val	Pro
[1737]				165					170					175		
[1738]	Lys	Asp	Asn	Tyr	Lys	Asn	Ala	Thr	Asn	Pro	Leu	Thr	Val	Glu	Val	Pro
[1739]				180					185					190		
[1740]	Tyr	Ile	Cys	Thr	Glu	Gly	Glu	Asp	Gln	Ile	Thr	Val	Trp	Gly	Phe	His
[1741]		195						200					205			
[1742]	Ser	Asp	Asn	Lys	Thr	Gln	Met	Lys	Ser	Leu	Tyr	Gly	Asp	Ser	Asn	Pro
[1743]		210					215					220				
[1744]	Gln	Lys	Phe	Thr	Ser	Ser	Ala	Asn	Gly	Val	Thr	Thr	His	Tyr	Val	Ser
[1745]	225					230					235				240	
[1746]	Gln	Ile	Gly	Asp	Phe	Pro	Asp	Gln	Thr	Glu	Asp	Gly	Gly	Leu	Pro	Gln
[1747]				245					250					255		
[1748]	Ser	Gly	Arg	Ile	Val	Val	Asp	Tyr	Met	Met	Gln	Lys	Pro	Gly	Lys	Thr
[1749]				260				265					270			
[1750]	Gly	Thr	Ile	Val	Tyr	Gln	Arg	Gly	Val	Leu	Leu	Pro	Gln	Lys	Val	Trp
[1751]		275					280						285			
[1752]	Cys	Ala	Ser	Gly	Arg	Ser	Lys	Val	Ile	Lys	Gly	Ser	Leu	Pro	Leu	Ile
[1753]		290					295					300				
[1754]	Gly	Glu	Ala	Asp	Cys	Leu	His	Glu	Lys	Tyr	Gly	Gly	Leu	Asn	Lys	Ser

[1755]	305	310	315	320
[1756]	Lys Pro Tyr Tyr Thr Gly Glu His Ala Lys Ala Ile Gly Asn Cys Pro			
[1757]		325	330	335
[1758]	Ile Trp Val Lys Thr Pro Leu Lys Leu Ala Asn Gly Thr Lys Tyr Arg			
[1759]		340	345	350
[1760]	Pro Pro Gly Gly Gly Trp Glu Gly Met Ile Ala Gly Trp His Gly Tyr			
[1761]		355	360	365
[1762]	Thr Ser His Gly Ala His Gly Val Ala Val Ala Ala Asp Leu Lys Ser			
[1763]		370	375	380
[1764]	Thr Gln Glu Ala Ile Asn Lys Ile Thr Lys Asn Leu Asn Ser Leu Ser			
[1765]	385	390	395	400
[1766]	Glu Leu Glu Val Lys Asn Leu Gln Arg Leu Ser Gly Ala Met Asp Glu			
[1767]		405	410	415
[1768]	Leu His Asn Glu Ile Leu Glu Leu Asp Glu Lys Val Asp Asp Leu Arg			
[1769]		420	425	430
[1770]	Ala Asp Thr Ile Ser Ser Gln Ile Glu Leu Ala Val Leu Leu Ser Asn			
[1771]		435	440	445
[1772]	Glu Gly Ile Ile Asn Ser Glu Asp Glu His Leu Leu Ala Leu Glu Arg			
[1773]		450	455	460
[1774]	Lys Leu Lys Lys Met Leu Gly Pro Ser Ala Val Asp Ile Gly Asn Gly			
[1775]	465	470	475	480
[1776]	Cys Phe Glu Thr Lys His Lys Cys Asn Gln Thr Cys Leu Asp Arg Ile			
[1777]		485	490	495
[1778]	Ala Ala Gly Thr Phe Asn Ala Gly Glu Phe Ser Leu Pro Thr Phe Asp			
[1779]		500	505	510
[1780]	Ser Leu Asn Ile Thr Ala Ala Ser Leu Asn Asp Asp Gly Leu Asp Asn			
[1781]		515	520	525
[1782]	Tyr Gln Ile Leu Ala Ile Tyr Ser Thr Val Ala Ser Ser Leu Val Leu			
[1783]		530	535	540
[1784]	Val Val Ser Leu Gly Ala Ile Ser Phe Trp Met Cys Ser Asn Gly Ser			
[1785]	545	550	555	560
[1786]	Leu Gln Cys Arg Ile Cys Ile			
[1787]		565		
[1788]	<210> 55			
[1789]	<211> 514			
[1790]	<212> DNA			
[1791]	<213> 人工序列			
[1792]	<220>			
[1793]	<223> 天然CPMV RNA 2 基因组的片段,核苷酸1-514			

- [1794] <400> 55
- [1795] tattaaaatc ttaatagggt ttgataaaag cgaacgtggg gaaacccgaa ccaaacccttc 60
- [1796] ttctaaactc tctctcatct ctcttaaagc aaacttctct cttgtctttc ttgcatgagc 120
- [1797] gatcttcaac gttgtcagat cgtgcttcgg caccagtaca atgttttctt tctactgaagc 180
- [1798] gaaatcaaag atctctttgt ggacacgtag tgcggcgcca ttaaataacg tgtacttgct 240
- [1799] ctattcttgt cgggtgtggtc ttgggaaaag aaagcttgct ggaggctgct gttcagcccc 300
- [1800] atacattact tgttacgatt ctgctgactt tcggcggtg caatatctct acttctgctt 360
- [1801] gacgaggtat tgttgcctgt acttctttct tcttcttctt gctgattggt tctataagaa 420
- [1802] atctagtatt ttctttgaaa cagagttttc cgtggtttt cgaacttgga gaaagattgt 480
- [1803] taagcttctg tatattctgc ccaaatttga aatg 514
- [1804] <210> 56
- [1805] <211> 511
- [1806] <212> DNA
- [1807] <213> 人工序列
- [1808] <220>
- [1809] <223> CPMV RNA2基因组,核苷酸1-514,115位和162位处具有修饰
- [1810] <400> 56
- [1811] tattaaaatc ttaatagggt ttgataaaag cgaacgtggg gaaacccgaa ccaaacccttc 60
- [1812] ttctaaactc tctctcatct ctcttaaagc aaacttctct cttgtctttc ttgcgtgagc 120
- [1813] gatcttcaac gttgtcagat cgtgcttcgg caccagtaca acgttttctt tctactgaagc 180
- [1814] gaaatcaaag atctctttgt ggacacgtag tgcggcgcca ttaaataacg tgtacttgct 240
- [1815] ctattcttgt cgggtgtggtc ttgggaaaag aaagcttgct ggaggctgct gttcagcccc 300
- [1816] atacattact tgttacgatt ctgctgactt tcggcggtg caatatctct acttctgctt 360
- [1817] gacgaggtat tgttgcctgt acttctttct tcttcttctt gctgattggt tctataagaa 420
- [1818] atctagtatt ttctttgaaa cagagttttc cgtggtttt cgaacttgga gaaagattgt 480
- [1819] taagcttctg tatattctgc ccaaatttga a 511
- [1820] <210> 57
- [1821] <211> 69
- [1822] <212> DNA
- [1823] <213> 人工序列
- [1824] <220>
- [1825] <223> 马铃薯贮藏蛋白信号肽(核苷酸序列)
- [1826] <400> 57
- [1827] atggcaacta ctaaaacttt ttttaatttta ttttttatga tattagcaac tactagttca 60
- [1828] acatgtgct 69
- [1829] <210> 58
- [1830] <211> 23
- [1831] <212> PRT
- [1832] <213> 人工序列

[1833]	<220>
[1834]	<223> 马铃薯贮藏蛋白信号肽(氨基酸序列)
[1835]	<400> 58
[1836]	Met Ala Thr Thr Lys Thr Phe Leu Ile Leu Phe Phe Met Ile Leu Ala
[1837]	1 5 10 15
[1838]	Thr Thr Ser Ser Thr Cys Ala
[1839]	20
[1840]	<210> 59
[1841]	<211> 29
[1842]	<212> PRT
[1843]	<213> 人工序列
[1844]	<220>
[1845]	<223> HA跨膜结构域共有氨基酸序列
[1846]	<220>
[1847]	<221> misc_feature
[1848]	<222> (3) .. (3)
[1849]	<223> Xaa可以是任何天然存在的氨基酸
[1850]	<220>
[1851]	<221> misc_feature
[1852]	<222> (15) .. (15)
[1853]	<223> Xaa可以是任何天然存在的氨基酸
[1854]	<220>
[1855]	<221> misc_feature
[1856]	<222> (17) .. (18)
[1857]	<223> Xaa可以是任何天然存在的氨基酸
[1858]	<220>
[1859]	<221> misc_feature
[1860]	<222> (23) .. (23)
[1861]	<223> Xaa可以是任何天然存在的氨基酸
[1862]	<220>
[1863]	<221> misc_feature
[1864]	<222> (25) .. (25)
[1865]	<223> Xaa可以是任何天然存在的氨基酸
[1866]	<400> 59
[1867]	Ile Leu Xaa Ile Tyr Tyr Ser Thr Val Ala Ile Ser Ser Leu Xaa Leu
[1868]	1 5 10 15
[1869]	Xaa Xaa Met Leu Ala Gly Xaa Ser Xaa Trp Met Cys Ser
[1870]	20 25

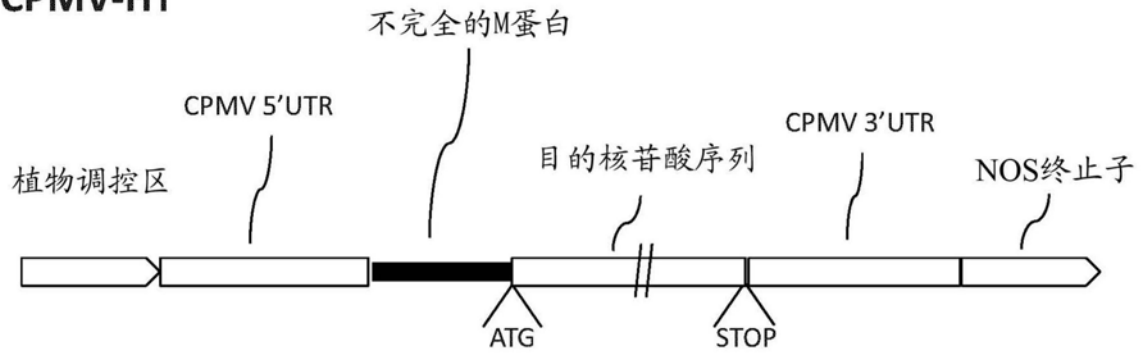
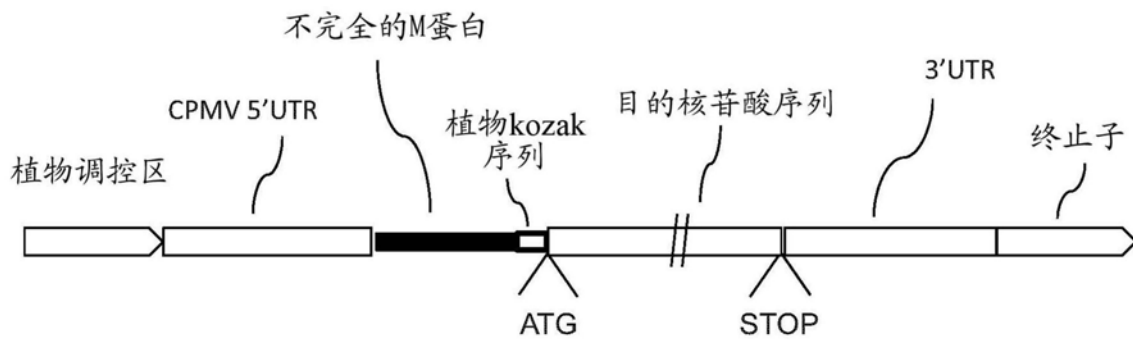
CPMV-HT**CPMV-HT+**

图1a

- CPMV HT

-

2X35S...TTCATTTGGAGAGGTATTAAAA...(CPMV HT 5'UTR)...ACCAGTACA(ACG)TTTT...M蛋白
(不完全的)...AAATTTGTCGGGCCCATG...(GOI)
- CPMV HT+

-

2X35S...TTCATTTGGAGAGGTATTAAAA...(CPMV HT 5'UTR)...ACCAGTACA(ACG)TTTT...M蛋白
(不完全的)...AAATTTGTCGGGCCCAATACCGCGGAGAAAAATG...(GOI)

图1b

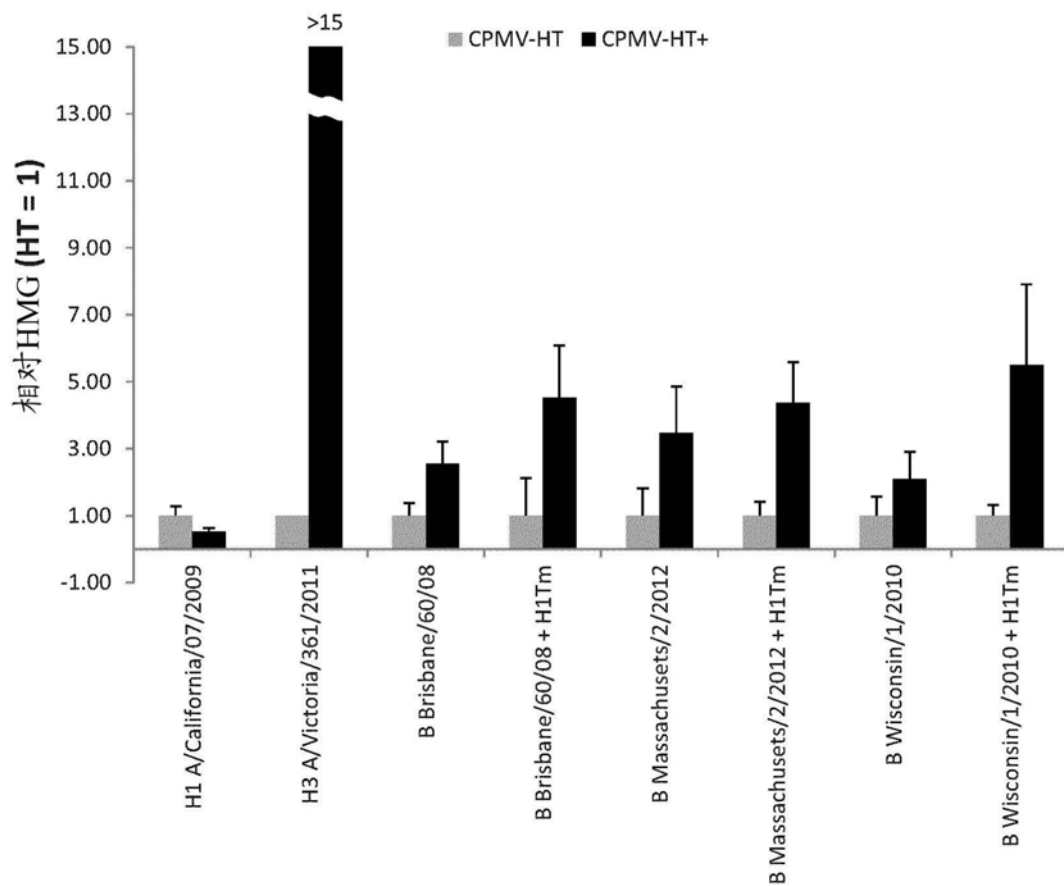


图2

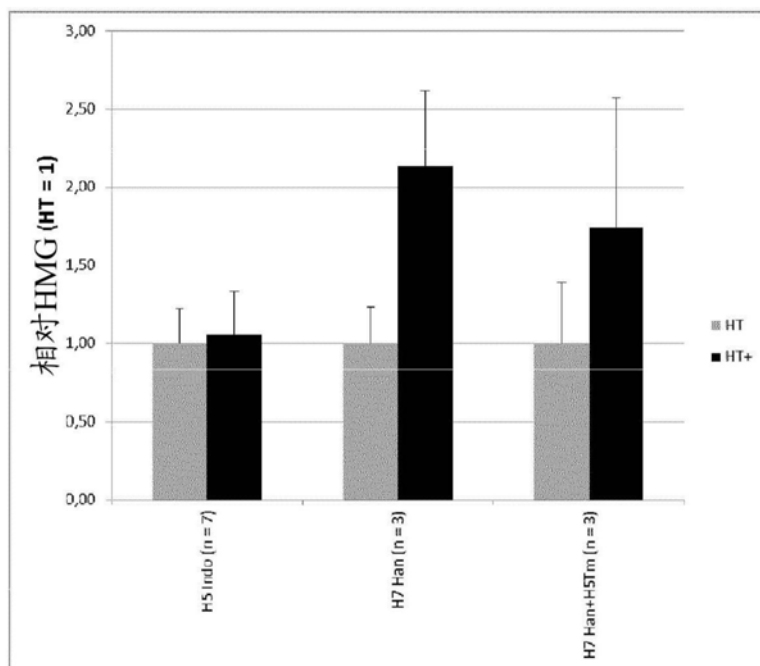


图3

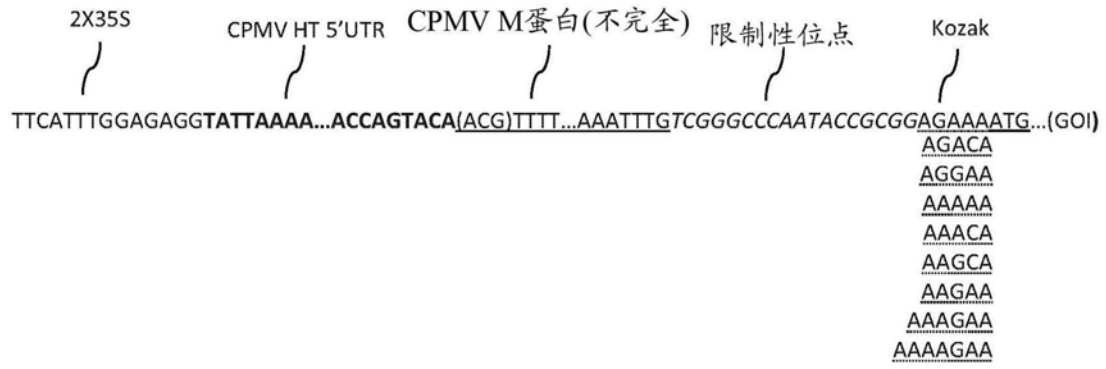


图4a

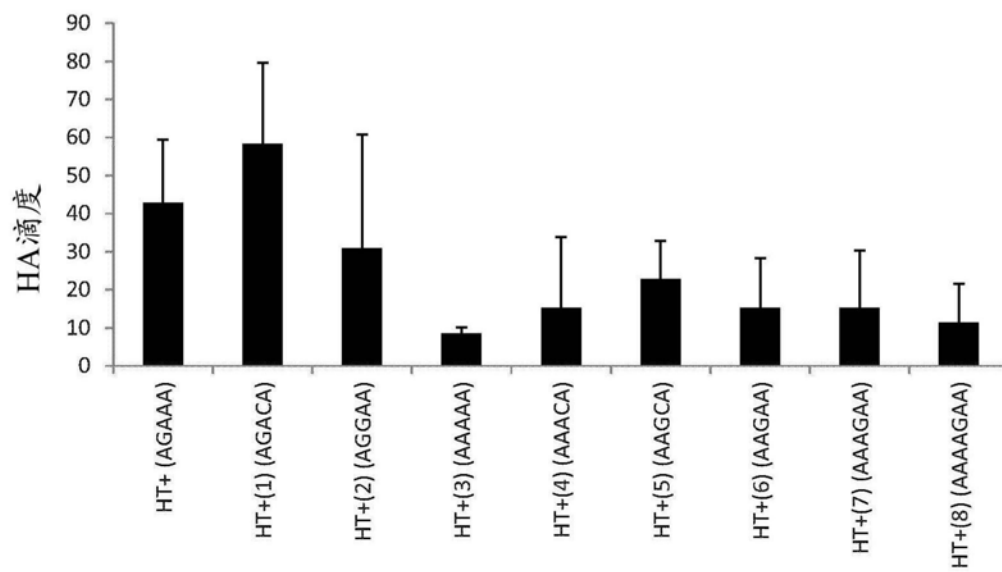


图4b

图 5A IF-PDI.S1+3c (SEQ ID NO:18)

AAATTTGTCGGGCCCCATGGCGAAAAACGTTGCGATTTTCGGCTTATTG

图 5B IF-H3V36111.s1-4r (SEQ ID NO:19)

ACTAAAGAAAAATAGGCCCTCAAATGCAAATGTTGCACCTAATGTTGCCCTT

图 5C PDISP/H3 Victoria的核苷酸序列 (SEQ ID NO: 20)

ATGGCGAAAAACGTTGCGATTTTCGGCTTATTGTTTTCTTCTTGTGTTGGTTCCCTTCTCAGATCTTCGCCCCAAAAAC
TTCTTGAAATGACAAACAGCACGGCAACGCTGTGCCCTGGGCACCATGCAGTACCAACGGAACGATAGTGAAAACAAT
CACGAATGACCAAATTGAAGTTACTAATGCTACTGAGCTGGTTTCAAGATTCCTCAATAGGTGAAATATGCGACAGTCCCT
CATCAGATCCCTTGATGGAGAAAACCTGCACACTAATAGATGCTCTATTGGGAGACCCCTCAGTGTGATGGCTTCCAAAATA
AGAAATGGGACCTTTTTGTTGAACGAAGCAAAGCCTACAGCAACTGTTACCCTTATGATGTGCCGGATTATGCCTCCCT
TAGGTCACTAGTTGCCTCATCCGGCACACTGGAGTTTAAACAATGAAAGCTTCAATTGGACTGGAGTCACTCAAAACGGA
ACAAGTTCTGCTTGCATAAGGAGATCTAATAATAGTTTCTTTAGTAGATTAAATTGGTTGACCCACTTAAACTTCAAT
ACCCAGCATTGAACGTGACTATGCCAAACAATGAACAATTTGACAAATTGTACATTTGGGGGGTTCACCACCCGGGTAC
GGACAAGGACCAAATCTTCTGTATGCTCAATCATCAGGAAGAATCACAGTATCTACCAAAAGAGCCACAAGCTGTA
ATCCCGAATATCGGATCTAGACCCAGAATAAGGAATATCCCTAGCAGAATAAGCATCTATTGGACAATAGTAAACCGG
GAGACATACTTTTGATTAAACAGCACAGGGAATCTAATTGCTCCTAGGGGTTACTTCAAAATACGAAGTGGGAAAAGCTC
AATAATGAGATCAGATGCACCCATTGGCAAATGCAATTTCTGAATGCATCACTCCAAATGGAAGCATTCCTCAATGACAAA
CCATTCCAAATGTAACAGGATCACATACGGGGCCTGTCCCAGATATGTTAAGCAAAGCACTCTGAAATTGGCAACAG
GAATGCGAAATGTACCAGAGAAACAACTAGAGGCATATTTGGCGCAATAGCGGGTTTCATAGAAAATGGTTGGGAGGG
AATGGTGGATGGTTGGTACGGTTTCAGGCATCAAAATTTCTGAGGGAAGAGGACAAGCAGCAGATCTCAAAAGCACTCAA
GCAGCAATCGATCAAATCAATGGGAAGCTGAATCGATTGATCGGGAAAACCAACGAGAAATTCATCAGATTGAAAAAG
AATTCTCAGAAGTCGAAGGGAGAATTCAGGACCTTGAGAAATATGTTGAGGACACTAAAATAGATCTCTGGTCATACAA
CGCGGAGCTTCTTGTGGCCCTGGAGAACCAACATACAATTGATCTAACTGACTCAGAAATGAACAACTGTTTGAAAAA
ACAAAGAAGCAACTAAGGGAAAATGCTGAGGATATGGGCAATGGTTGTTTCAAAATATACCACAAATGTGACAATGCCT
GCATAGGATCAATCAGAAATGGAACCTTATGACCACGATGTATACAGAGATGAAGCATTAACAACCGGTTCCAGATCAA
GGGAGTTGAGCTGAAGTCAGGGTACAAGATTGGATCCTATGGATTTCCTTTGCCATATCATGTTTTTTGCTTTGTGTT
GCTTTGTTGGGGTTCATCATGTGGGCCTGCCAAAAGGGCAACATTAGGTGCAACATTTCATTGA

图 5D 构建体1191的示意图。图示中对用于质粒线性化的SacII和StuI限制酶位点进行了注释。

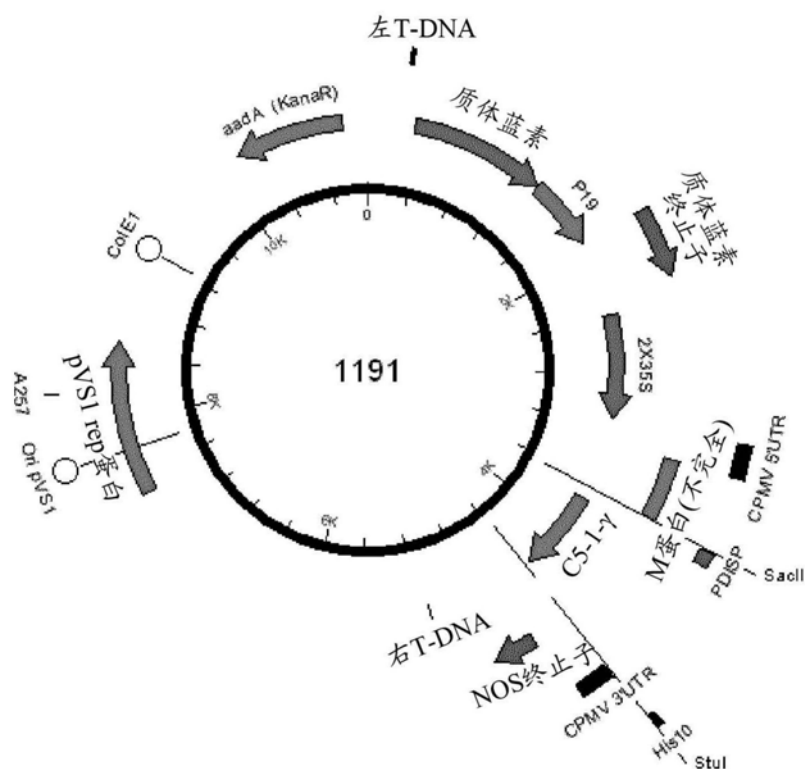


图5E 构建体1191, 从左至右t-DNA边界(加下划线)。具有质体蓝素-P19-质体蓝素沉默抑制子表达盒的2X35S/CPMV-HT/ NOS (SEQ ID NO:21)

TGGCAGGATATATTGTGGTGTAACAAATTGACGCTTAGACAACCTAATAACACATTGCGGACGTTTTTAAATGTACTGA
ATTACGCCGAATCCCGGGCTGGTATATTTATATGTTGTGCAATAACTCAAAAACCATAAAAGTTAAGTTAGCAAGTG
TGTACATTTTTACTTGAACAAAAATATTCACCTACTACTGTTATAAATCATTATTAAACATTAGAGTAAAGAAATATGG
ATGATAAGAACAAGAGTAGTGATATTTTGACAACAATTTTGTGCAACATTTGAGAAAATTTTGTGTTCTCTCTTTTC
ATTGGTCAAAAACAATAGAGAGAGAAAAAGGAAGAGGGAGAATAAAAACATAATGTGAGTATGAGAGAGAAAGTTGTAC
AAAAGTTGTACAAAATAGTTGTACAAATATCATTGAGGAATTTGACAAAAGCTACACAAATAAGGGTTAATTGCTGTGA
AATAAATAAGGATGACGCATTAGAGAGATGTACCATTAGAGAATTTTGGCAAGTCATTAAAAAGAAAGAATAAATTAT
TTTTAAATTTAAAGTTGAGTCATTTGATTAAACATGTGATTATTTAATGAATTGATGAAAGAGTTGGATTAAAGTTGT
ATTAGTAATTAGAATTTGGTGTCAAATTTAATTTGACATTTGATCTTTTCCATATATATTGCCCATAGAGTCAGTTAAC
TCATTTTATATTTTCATAGATCAAATAAGAGAAATAACGGTATATTAATCCCTCCAAAAAAGGATATATTAC
TAAAAAATCTAAGCCACGTAGGAGGATAACAGGATCCCGGTAGGAGGATAACATCCAATCCAACCAATCACAACAATCC
TGATGAGATAACCCACTTTAAGCCACGCATCTGTGGCACATCTACATTATCTAAATCACACATTCTTCCACACATCTG
AGCCACACAAAAACCAATCCACATCTTTATCACCATTCTATAAAAAATCACACTTTGTGAGTCTACACTTTGATTCCC
TTCAAACACATACAAAGAGAAGAGACTAATTAATTAATTAATCATCTTGAGAGAAAATGGAACGAGCTATACAAGGAAA
CGACGCTAGGGAACAAGCTAACAGTGAACGTTGGGATGGAGGATCAGGAGGTACCACTTCTCCCTTCAAACCTCTGAC
GAAAGTCCGAGTTGGACTGAGTGGCGGCTACATAACGATGAGACGAATTCGAATCAAGATAATCCCTTGGTTTCAAGG
AAAGCTGGGGTTTCGGGAAAGTTGTATTTAAGAGATATCTCAGATACGACAGGACGGAAGCTTCACTGCACAGAGTCCT
TGGATCTTGGACGGGAGATTCCGTTAACTATGCAGCATCTCGATTTTTCGGTTTCGACCAGATCGGATGTACCTATAGT
ATTCGGTTTCGAGGAGTTAGTATCACCCTTTCTGGAGGGTCGCAACTCTTCAGCATCTCTGTGAGATGGCAATTCGGT
CTAAGCAAGAACTGCTACAGCTTGCCCAATCGAAGTGGAAAGTAATGTATCAAGAGGATGCCCTGAAGGTACTCAAAC
CTTCGAAAAAGAAAGCGAGTAAGTTAAATGCTTCTTCGTCTCCTATTTATAATATGGTTTGTATTGTAAATTTTGT
CTTGTAGAAGAGCTTAATTAATCGTTGTTGTTATGAAATACTATTTGTATGAGATGAACTGGTGTAATGTAATTCATTT
ACATAAGTGGAGTCAGAATCAGAATGTTTCTCCATAACTAACTAGACATGAAGACCTGCCGCTACAATGTCTTTATA
TTTGAACAATAAAATTGAACATCTTTTGCCACAACCTTTATAAGTGGTTAATATAGCTCAAATATATGGTCAAGTTCAA
TAGATTAATAATGGAATATCAGTTATCGAAATTCATTAACAATCAACTAACGTTATTAACACTAATTTTATATCAT
CCCTTTTGATAAATGATAGTACACCAATTAGGAAGGAGCATGCTCGCTAGGAGATTGTCGTTTCCCGCCTTCAGTTTG
CAAGCTGCTCTAGCCGTGTAGCCAATACGCAACCGCCTCTCCCCGCGCTTGGGAATTACTAGCGCGTGTGACAAAGC
TTGCATGCCGCTCAACATGGTGGAGCACGACACACTTGTCTACTCCAAAAATATCAAAGATACAGTCTCAGAAGACCAA
AGGGCAATTGAGACTTTTCAACAAAGGTAATATCCGGAACCTCTCGGATTCATTGCCAGCTATCTGTCACTTTA
TTGTGAAGATAGTGGAAAAGGAAGGTGGCTCCTACAAATGCCATCATTGCGATAAAGGAAAGGCCATCGTTGAAGATGC
CTCTGCCGACAGTGGTCCCAAAGATGGACCCCCACCCACGAGGAGCATCGTGGAAAAAGAACGTTCCAAACCAGCTCT
TCAAAGCAAGTGGATTGATGTGATAACATGGTGGAGCACGACACACTTGTCTACTCCAAAAATATCAAAGATACAGTCT
CAGAAGACCAAAGGGCAATTGAGACTTTTCAACAAAGGTAATATCCGGAACCTCTCGGATTCATTGCCAGCTAT
CTGTCACTTTATTGTGAAGATAGTGGAAAAGGAAGGTGGCTCCTACAAATGCCATCATTGCGATAAAGGAAAGGCCATC
GTTGAAGATGCCTCTGCCGACAGTGGTCCCAAAGATGGACCCCCACCCACGAGGAGCATCGTGGAAAAAGAACGTTT
CAACCACGCTCTCAAAGCAAGTGGATTGATGTGATATCTCCACTGACGTAAGGGATGACGCACAATCCCCTATCTCTC
GCAAGACCTTCTCTATATAAGGAAGTTCAATTTCAATTTGGAGAGGTATTAATACTTAATAGGTTTGTATAAAGCGA
ACGTGGGGAACCCGAACCAACCTTCTCTAACTCTCTCATCTCTCTTAAAGCAAACTCTCTCTGTCTTTCTT
GCGTGAGCGATCTTCAACGTTGTCAGATCGTGCTTCGGCACCAAGTACAACGTTTCTTTCACTGAAGCGAAATCAAAGA
TCTCTTTGTGGACACGTAGTGCGGCGCCATTAATAACGTGTACTTGTCTATTCTTGTGCGTGTGGTCTTGGGAAAAAG
AAAGCTTGTGGAGGCTGCTGTTTCAGCCCCATACATTACTTGTACGATTCTGCTGACTTTCGGCGGGTCAATATCTC
TACTTCTGCTTGACGAGGTATTGTTGCCTGTACTTCTTTCTTCTTCTTCTTCTGCTGATTGGTTCTATAAGAAATCTAGTA
TTTTCTTTGAAACAGAGTTTTCCCGTGGTTTTTCGAACTTGAGAAAGATTGTTAAGCTTCTGTATATTTGCCCAAATT
TGTCGGGCCCCGCGGATGGCGAAAAACGTTGCGATTTTCGGCTTATGTTTTCTCTTCTGTGTTGGTCTCTCTCAGAT
CTTCGCTTGACAGGCTCCTCAGCCAAAACGACACCCCATCTGTCTATCCACTGGCCCTGGATCTGCTGCCAAACTAA
CTCCATGGTGACCTGGGATGGCTGGTCAAGGGCTATTTCCCTGAGCCAGTGACAGTGACCTGGAACTCTGGATCCCTG
TCCAGCGGTGTGCACACCTTCCAGCTGTCTGAGCTGACCTTACACTCTGAGCAGCTCAGTGACTGTCCCTCCA
GCACCTGGCCCGAGACCGTCACCTGCAACGTTGCCACCCGGCCAGCAGCACCAAGGTGGACAAGAAAATTTGTGCC
CAGGGATTGTGGTTGTAAGCCTTGCATATGTACAGTCCCAGAAGTATCATCTGTCTTCTCTCTTCCCCCAAAGCCCAAG

图 5E续

GATGTGCTCACCATTACTCTGACTCCTAAGGTCACGTGTGTTGTGGTAGACATCAGCAAGGATGATCCCGAGGTCCAGT
TCAGCTGGTTTGTAGATGATGTGGAGGTGCACACAGCTCAGACGCAACCCCGGAGGAGCAGTTCAACAGCACTTTCCG
CTCAGTCAGTGAACCTCCCATCATGCACCAGGACTGGCTCAATGGCAAGGAGCGATCGCTCACCATCACCATCACCATC
ACCATCACCATTAAAGGCCTATTTCTTTAGTTTGAATTTACTGTTATTCGGTGTGCATTTCTATGTTTGGTGAGCGGT
TTTCTGTGCTCAGAGTGTGTTTATTTTATGTAATTTAATTTCTTTGTGAGCTCCTGTTTAGCAGGTCGTCCTTCAGCA
AGGACACAAAAAGATTTTAAATTTTATTAATAAAAAAAAAAAAAAAAAAGACCGGAATTCGATATCAAGCTTATCGACCTGC
AGATCGTTCAAACATTTGGCAATAAAGTTTCTTAAGATTGAATCCTGTTGCCGCTTGCATGATTATCATATAATTT
CTGTTGAATTACGTTAAGCATGTAATAATTAACATGTAATGCATGACGTTATTTATGAGATGGGTTTTTATGATTAGAG
TCCCGCAATTATACATTTAATACGCGATAGAAAACAAATATAGCGCGCAACTAGGATAAATTATCGCGCGCGGTGTC
ATCTATGTTACTAGATCTCTAGAGTCTCAAGCTTGGCGCGCCACGTGACTAGTGGCACTGGCCGTCGTTTACAACGT
CGTGACTGGGAAAACCTTGGCGTTACCCAACCTTAATCGCCTTGCAGCACATCCCCCTTTCGCCAGCTGGCGTAATAGCG
AAGAGGCCCGCACCGATCGCCCTTCCCAACAGTTGCGCAGCCTGAATGGCGAATGCTAGAGCAGCTTGAGCTTGATCA
GATTGTCGTTTCCGCGCTTCAGTTTAACTATCAGTGTGTTGACAGGATATATGGCGGGTAAACCTAAGAGAAAAGAGC
GTTTA

图 5F 表达盒编号1391, 从2X35S启动子至NOS终止子。PDISP/H3 Victoria核苷酸序列如下划线。(SEQ ID NO: 22)

GTCAACATGGTGGAGCACGACACACTTGTCTACTCCAAAAATATCAAAGATACAGTCTCAGAAGACCAAAGGGCAATTG
AGACTTTTCAACAAAGGGTAATATCCGAAACCTCCTCGGATTCCATTGCCAGCTATCTGTCACTTTATTGTGAAGAT
AGTGAAAAGGAAGGTGGCTCCTACAAATGCCATCATTGCGATAAAGGAAAGGCCATCGTTGAAGATGCCTCTGCCGAC
AGTGGTCCCAAAGATGGACCCCAACCCACGAGGAGCATCGTGAAAAAGAGACGTTCCAACACGCTCTCAAAGCAAG
TGGATTGATGTGATAACATGGTGGAGCACGACACACTTGTCTACTCCAAAAATATCAAAGATACAGTCTCAGAAGACCA
AAGGGCAATTGAGACTTTTCAACAAAGGGTAATATCCGAAACCTCCTCGGATTCCATTGCCAGCTATCTGTCACTTT
ATTGTGAAGATAGTGGAAAAGGAAGGTGGCTCCTACAAATGCCATCATTGCGATAAAGGAAAGGCCATCGTTGAAGATG
CCTCTGCCGACAGTGGTCCCAAAGATGGACCCCAACCCACGAGGAGCATCGTGAAAAAGAGACGTTCCAACACGCTC
TTCAAAGCAAGTGGATTGATGTGATATCTCCACTGACGTAAGGGATGACGCACAATCCCCTATCCTTCGCAAGACCCCT
TCCTCTATATAAGGAAGTTTCAATTTATTGGAGAGGTATTAATACTTAATAGGTTTTGATAAAGCGAACGTGGGGAA
ACCCGAACCAAACTTCTTCTAACTCTCTCTCATCTCTTAAAGCAAACTTCTCTTGTCTTTCTTGCCTGAGCGA
TCTTCAACGTTGTGATCGTCTTCCGACACAGTACAACGTTTTCTTCTACTGAAGCGAAATCAAAGATCTCTTTGTG
GACACGTAGTGGCGGCCATTAAATAACGTGTACTTGTCTTATCTTGTGCGTGTGGTCTTGGGAAAAGAAAGCTTGCT
GGAGGCTGCTGTTACGCCCCATACATTACTTGTACGATTCTGCTGACTTTCGGCGGGTGCAATATCTCTACTTCTGCT
TGACGAGGTATGTTGCCTGTACTTCTTCTTCTTCTTCTGCTGATTGGTTCTATAAGAAATCTAGTATTTTCTTTGA
AACAGAGTTTTCCCGTGGTTTTCGAACTTGGAGAAAGATTGTTAAGCTTCTGTATATCTGCCCAAATTTGTGGGGCC
ATGGCGAAAACGTTGCGATTTTCGGCTTATTGTTTTCTTCTTGTGTTGGTTCTTCTCAGATCTTCGCCCAAAAC
TTCTTGGAAATGACAAACAGCACGGCAACGCTGTGCCCTGGGCACCATGCAGTACCAACGGAACGATAGTGAAACAAT
CACGAATGACCAAAATGAAGTTACTAATGCTACTGAGCTGGTTTCAAGATTCCTCAATAGGTGAAATATGCGACAGTCTT
CATCAGATCCTTGATGGAGAAAACCTGCACACTAATAGATGCTCTATTGGGAGACCTCAGTGTGATGGCTTCCAAAATA
AGAAATGGGACCTTTTTGTTGAACGAAGCAAAGCCTACAGCAACTGTTACCTTATGATGTGCCGATTATGCCTCCCT
TAGGTCACTAGTTGCCTCATCCGGCACACTGGAGTTTAACAATGAAAGCTTCAATTGGACTGGAGTCACTCAAAACGGA
ACAAGTTCTGCTTGCATAAGGAGATCTAATAATAGTTTCTTTAGTAGATTAAATGGTTGACCCACTTAACTTCAAAAT
ACCCAGCATTGAACGTGACTATGCCAAACAATGAACAATTTGACAAATTTGACATTTGGGGGGTTTACCACCCGGGTAC
GGACAAGGACCAAAATCTTCTGTATGCTCAATCATCAGGAAGAATCACAGTATCTACCAAAAGAGCCAACAAGCTGTA
ATCCGAATATCGGATCTAGACCCAGAATAAGGAATATCCCTAGCAGAATAAGCATCTATTGGACAATAGTAAACCGG
GAGACATACTTTGATTAAACAGCACAGGAATCTAATGCTCTAGGGGTTACTTCAAAATACGAAGTGGGAAAGCTC
AATAATGAGATCAGATGCACCCATTGGCAAAATGCAATTTGTAATGCATCACTCCAAATGGAAGCATTCCCAATGACAAA
CCATTCCAAATGTAAACAGGATCACATACGGGGCCTGTCCAGATATGTTAAGCAAAGCACTCTGAAATTTGGCAACAG
GAATGCGAAATGTACCAGAGAAACAACTAGAGGCATATTTGGCGCAATAGCGGGTTTCATAGAAAATGGTTGGGAGGG
AATGGTGGATGGTTGGTACGGTTTCAAGGCATCAAAATTTGAGGGAAGAGGACAAGCAGCAGATCTCAAAGCACTCAA
GCAGCAATCGATCAAATCAATGGGAAGCTGAATCGATTGATCGGGGAAAACCAACGAGAAATTCATCAGATTGAAAAAG
AATTCTCAGAAGTCGAAGGGAGAATTGAGGACCTTGAGAAATATGTTGAGGACACTAAAATAGATCTCTGGTCATACAA
CGCGGAGCTTCTGTTGCCCTGGAGAACCAACATACAATTGATCTAATGACTCAGAAATGAACAACTGTTTGAAAAA

图 5F 续

ACAAAGAAGCAACTAAGGGAAAATGCTGAGGATATGGGCAATGGTTGTTTCAAAATATACCACAAATGTGACAATGCCT
GCATAGGATCAATCAGAAATGGAACCTTATGACCACGATGTATACAGAGATGAAGCATTAAACAACCGGTTCCAGATCAA
GGGAGTTGAGCTGAAGTCAGGGTACAAAGATTGGATCCTATGGATTTCCTTTGCCATATCATGTTTTTGGCTTTGTGT
GCTTTGTTGGGGTTCATCATGTGGGCTGCCAAAAGGGCAACATTAGGTGCAACATTGCATTGGAAGGCCTATTTTCT
TTAGTTTGAATTTACTGTTATTCGGTGTGCATTTCTATGTTTGGTGAGCGGTTTTCTGTGCTCAGAGTGTGTTTATTTT
ATGTAATTTAATTTCTTTGTGAGCTCCTGTTTAGCAGGTCGTCCTTCAGCAAGGACACAAAAGATTTTAATTTTATT
AAAAAAAAAAAAAAAAAGACCGGAATTCGATATCAAGCTTATCGACCTGCAGATCGTTCAACATTTGGCAATAAAG
TTTCTTAAGATTGAATCCTGTTGCCGCTCTGCGATGATTATCATATAATTTCTGTTGAATTACGTTAAGCATGTAATA
ATTAACATGTAATGCATGACGTTATTTATGAGATGGGTTTTATGATTAGAGTCCCGCAATTATACATTTAATACGCGA
TAGAAAACAAAATATAGCGCGCAAACTAGGATAAATTATCGCGCGGGTGTCTATCTATGTTACTAGAT

图5G PDISP/H3 Victoria的氨基酸序列(SEQ ID NO:23)

MAKNVAIFGLLFSLVLVPSQIFAQKLPGNDNSTATLCLGHAVPNGTIVKTIITNDQIEVTNATELVQNSSIGEICDSP
HQILDGENCTLIDALLGDPQCDGFQNKWDLFVERSKAYSNCPYPDYPDYLRLSLVASSGTLEFNNESEFNWTGVTQNG
TSSACIRRSNNSFFSRLNWLHLNFKYPALNVTMPNNEQFDKLYIWGVHHPGTDKDQIFLYAQSSGRITVSTKRSQAV
IPNIGSRPRI RNIPSRISYWTIVKPGDILLINSTGNLIAPRGYFKIRSGKSSIMRSDAPIGKCNSCITPNGSIPNDK
PFQNVNRITYGACPRYVKQSTLKLATGMNRNVEKQTRGIFGAIAGFIENGWEGMVDGWYGFRRHQNSEGRQAADLKSTQ
AAIDQINGKLNRLIGKTNEKFHQIEKEFSEVEGRIQDLEKYVEDTKIDLWSYNAELLVALENQHTIDLTDSEMNKLFK
TKKQLRENAEDMGNGCFKIYHKCDNACIGSIRNGTYDHDVYRDEALNNRFQIKGVELKSGYKDWILWISFAISCFLLCV
ALLGFIMWACQKGNIRCNICI*

图5H 构建体编号1391的示意图

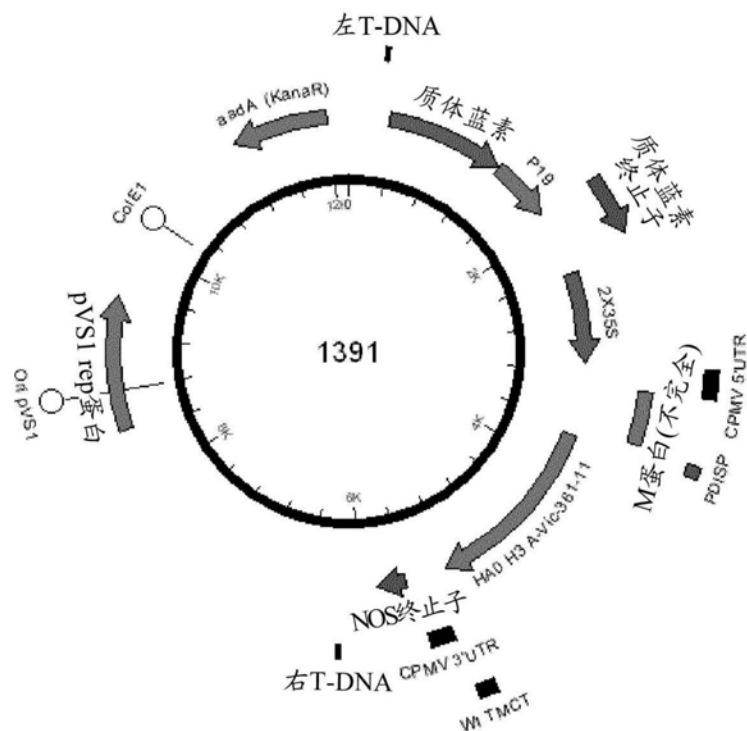


图5:A-2X35S/CPMV-HT/PDISP/H3 Victoria/NOS (构建体编号1391)

图 6C 续

AAAAGTTGTACCAAAATAGTTGTACAAATATCATTGAGGAATTTGACAAAAGCTACACAAATAAGGGTTAATTGCTGTA
AATAAATAAGGATGAGCATTAGAGAGATGTACATTAGAGAATTTTGGCAAGTCATTAAAAAGAAAGATAAATTTAT
TTTTAAATTTAAAGTTGAGTCATTTGATTAAACATGTGATTATTTAATGAATTGATGAAAGAGTTGGATTAAAGTTGT
ATTAGTAATTAGAATTTGGTGTCAAATTTAATTTGACATTTGATCTTTTCCATATATATTGCCCATAGAGTCAGTTAAC
TCATTTTATATTTTCATAGATCAAATAAGAGAAATAACGGTATATTAATCCCTCCAAAAAAGGATATATTAC
TAAAAATCTAAGCCACGTAGGAGGATAACAGGATCCCGTAGGAGGATAACATCCAATCCAACCAATCACAACAATCC
TGATGAGATAACCCACTTTAAGCCACGCATCTGTGGCAGATCTACATTATCTAAATCACACATCTTCCACACATCTG
AGCCACACAAAAACCAATCCACATCTTTATCACCATTCTATAAAAAATCACACTTTGTGAGTCTACACTTTGATTCCC
TTCAAACACATACAAAGAGAAGAGACTAATTAATTAATTAATCATCTTGAGAGAAAATGGAACGAGCTATACAAGGAAA
CGACGCTAGGGAACAAGCTAACAGTGAACGTTGGGATGGAGGATCAGGAGGTACCCTTCTCCCTTCAAACCTCCTGAC
GAAAGTCCGAGTTGGACTGAGTGGCGGTACATAACGATGAGACGAATTCGAATCAAGATAATCCCTTGGTTTCAAGG
AAAGCTGGGGTTTCGGGAAAGTTGTATTTAAGAGATATCTCAGATACGACAGGACGGAAGCTTCACTGCACAGAGTCCT
TGGATCTTGGACGGGAGATTCCGTTAACTATGCAGCATCTCGATTTTTCGGTTTCGACCAGATCGGATGTACCTATAGT
ATTCGGTTTCGAGGAGTTAGTATCACCCTTCTGGAGGGTCGCGAATCTTCAAGCATCTCTGTGAGATGGCAATTCGGT
CTAAGCAAGAACTGCTACAGCTTGCCCCAATCGAAGTGGAAGTAATGTATCAAGAGGATGCCCTGAAGGTACTCAAAC
CTTCGAAAAAGAAAGCGAGTAAGTTAAATGCTTCTTCGTCCTCTTCTTATAATATGTTTGTATTGTTAATTTGTGT
CTTGTAGAAGAGCTTAATTAATCGTTGTGTGTATGAAATCTATTTGTATGAGATGAACTGGTGTAAATGTAATTCATTT
ACATAAGTGGAGTCAGAATCAGAATGTTTCTCCATACTAACTAGACATGAAGACCTGCCGCTACAATTGTCTTATA
TTTGAACAATAAAATGAACATCTTTTGCACAACCTTTATAAGTGGTAAATATAGCTCAAATATATGGTCAAGTTCAA
TAGATTAATAATGGAAATATCAGTTATCGAAATTCATTAACAATCAACTTAACGTTATTAATACTAATTTTATATCAT
CCCCTTTGATAAATGATAGTACACCAATTAGGAAGGAGCATGCTCGCTAGGAGATTGTCGTTTCCCGCTTCAGTTTG
CAAGCTGCTCTAGCCGTGTAGCCAATACGCAAAACCGCTCTCCCCGCGCTTGGGAATTACTAGCGCGTGTGACAAAGC
TTGCATGCCGTTCAACATGGTGGAGCAGACACACTTGTCTACTCCAAAAATATCAAGATACAGTCTCAGAAGACCAA
AGGGCAATTGAGACTTTTCAACAAAGGGTAATATCCGGAACCTCCTCGGATTCATTGCCCAGCTATCTGTCACTTTA
TTGTGAAGATAGTGGAAAAGGAAGGTGGCTCCTACAAATGCCATTCGCGATAAAGGAAAGGCCATCGTTGAAGATGC
CTCTGCCGAGAGTGGTCCCAAGATGGACCCCAACACGAGGAGCATCGTGAAAAAGAGAGCTTCCAACCAAGCTCT
TCAAGCAAGTGGATTGATGTGATAACATGGTGGAGCAGACACACTTGTCTACTCCAAAAATATCAAGATACAGTCT
CAGAAGACCAAGGGCAATTGAGACTTTTCAACAAAGGGTAATATCCGGAACCTCCTCGGATTCATTGCCCAGCTAT
CTGTCACTTTATTGTGAAGATAGTGGAAAAGGAAGGTGGCTCCTACAAATGCCATCATTGCGATAAAGGAAAGGCCATC
GTTGAAGATGCTCTGCCGACAGTGGTCCCAAGATGGACCCCAACACGAGGAGCATCGTGAAAAAGAGAGCTTC
CAACCACGTCTTCAAAGCAAGTGGATTGATGTGATATCTCCACTGACGTAAGGGATGACGCACAATCCCACTATCCTTC
GCAAGACCTTCTCTATATAAGGAAGTTCAATTTCAATTTGGAGAGGTATTAATACTTAATAGGTTTGTATAAAGCGA
ACGTGGGGAACCCGAACCAACCTTCTTCAACTCTCTCTCATCTCTCTTAAAGCAAACTTCTCTTGTCTTTCTT
GCGTGAGCGATCTTCAACGTTGTCAGATCGTGCTTCGGCACCAGTACAACGTTTCTTTCACTGAAGCGAAATCAAAGA
TCTCTTTGTGGACACGTAGTGGCGGCCATTAAATAACGTGTACTTGTCTTCTTCTTGTCTGCTGGTCTTGGGAAAAAG
AAAGCTTGTGGAGGCTGCTGTTTCAGCCCCATACATTACTTGTACGATTCTGCTGACTTTTCGGCGGGTGAATATCTC
TACTTCTGCTTGACGAGGTATTGTTGCCGTGACTTCTTTCTTCTTCTTCTTGTGCTGATTGGTTCTATAAGAAATCTAGTA
TTTTCTTTGAAACAGAGTTTCCCGTGGTTTTCGAACCTGGGAGAAAGATTGTTAAGCTTCTGTATATCTTGCCCAAAAT
TGTCGGGCCCCAATACCGCGAGAAAATGGCGAAAACGTTGCGATTTTCGGCTTATTGTTTTCTTCTTGTGTTGGTT
CCTTCTCAGATCTTCGCGACGTCACTCCTCAGCCAAAACGACACCCCATCTGTCTATCCACTGGCCCCGATCTGCT
GCCCCAACTAACTCCATGGTGACCCTGGGATGCCTGGTCAAGGGCTATTTCCCTGAGCCAGTGACAGTGACCTGGAAC
CTGGATCCCTGTCCAGCGGTGTGCACACCTTCCAGCTGTCTGCAGTCTGACCTCTACACTCTGAGCAGCTCAGTGAC
TGTCCTTCCAGCACCTGGCCAGCGAGACCGTCACCTGCAACGTTGCCACCCGGCCAGCAGCACCAAGGTGGACAAG
AAAATGTGCGCAGGATTGTGGTTGTAAGCCTTGCATATGTACAGTCCAGAAAGTATCATCTGCTTCTTCACTTCCCCC
CAAAGCCCAAGGATGTGCTCACCATTACTCTGACTCCTAAGGTACGCTGTGTGTGGTAGACATCAGCAAGGATGATCC
CGAGGTCCAGTTCAGCTGGTTTGTAGATGATGTGGAGGTGCACACAGCTCAGACGCAACCCCGGAGGAGCAGTTCAAC
AGCACTTTCCGCTCAGTCAGTGAATTTCCATCATGCACCAGGACTGGCTCAATGGCAAGGAGCGATCGCTCACCATCA
CCATCACCATCACCATCACCATTAAAGGCTATTTTCTTTAGTTTGAATTTACTGTTATTTCGGTGTGCAATTTCTATGTT
TGGTGAGCGGTTTTCTGTGCTCAGAGTGTGTTTTTATGTAATTTAATTTCTTTGTGAGCTCCTGTTTAGCAGGTGCG
TCCCTTCAGCAAGGACACAAAAAGATTTTAATTTTATTAAAAAAAGACCGGAATTCGATATCAAGCT
TATCGACCTGCAGATCGTTCAAACATTTGGCAATAAGTTTCTTAAGATTGAATCCTGTTGCCGGTCTTGCGATGATTA
TCATATAATTTCTGTGAATTACGTTAAGCATGTAATAATTAACATGTAATGCATGACGTTATTTATGAGATGGGTTTT
TATGATTAGAGTCCCGCAATTATACATTTAATACGCGATAGAAAACAAATATAGCGCGCAAACTAGGATAAATATATCG
CGCGGGTGTGTCATCTGTGTTACTAGATCTCAGAGTCTCAAGCTTGGCGCGCCACGTGACTAGTGGCACTGGCGTCTG
TTTTACAACGTCGTGACTGGGAAAACCTGGCGTTACCCAACCTAATCGCTTGCAGCACATCCCCCTTTCGCCAGCTG

图 6C续

CGCTAATAGCGAAGAGGCCCGCACCGATCGCCCTTCCCAACAGTTGCGCAGCCTGAATGGCGAATGCTAGAGCAGCTTG
AGCTTGGATCAGATTGTCGTTTCCCGCCTTCAGTTTAACTATCAGTGTTTGACAGGATATATTGGCGGGTAAACCTAA
GAGAAAAGAGCGTTTA

图 6D 表达盒编号1819, 从2X35S启动子至NOS终止子。PDISP/H3 Victoria核糖
酸序列加有下划线。(SEQ ID NO: 26)

GTCAACATGGTGGAGCAGCACACTTGTCTACTCCAAAAATATCAAAGATACAGTCTCAGAAGACCAAAGGGCAATTG
AGACTTTTCAACAAAGGGTAATATCCGGAAACCTCCTCGGATTCCATTGCCAGCTATCTGTCACTTTATTGTGAAGAT
AGTGGAAAAGGAAGGTGGCTCCTACAAATGCCATCATTGCGATAAAGGAAAGGCCATCGTTGAAGATGCCTCTGCCGAC
AGTGGTCCCAAGATGGACCCCAACCCACGAGGAGCATCGTGAAAAAGAAGACGTTCAACACGCTCTCAAAGCAAG
TGGATTGATGTGATAACATGGTGGAGCAGCACACTTGTCTACTCCAAAAATATCAAAGATACAGTCTCAGAAGACCA
AAGGCAATTGAGACTTTTCAACAAAGGGTAATATCCGGAAACCTCCTCGGATTCCATTGCCAGCTATCTGTCACTTT
ATTGTGAAGATAGTGGAAAAGGAAGGTGGCTCCTACAAATGCCATCATTGCGATAAAGGAAAGGCCATCGTTGAAGATG
CCTCTGCCGACAGTGGTCCCAAGATGGACCCCAACCCACGAGGAGCATCGTGAAAAAGAAGACGTTCAACACGCTC
TTCAAAGCAAGTGGATTGATGTGATATCTCCACTGACGTAAGGGATGACGCACAATCCACTATCCTTCGCAAGACCCCT
TCTCTATATAAGGAAGTTCATTTCAATTGGAGAGGTATTAATACTTAATAGGTTTGATAAAAGCGAACGTGGGGAA
ACCCGAACCAACCTTCTTCTAACTCTCTCTCATCTCTCTTAAAGCAAACTTCTCTCTTGTCTTTCTTGCCTGAGCGA
TCTTCAACGTTGTGAGATCGTGCTTCGGCACCAAGTACAACGTTTTCTTTCTACTGAAGCGAAATCAAAGATCTCTTTGTG
GACACGTAGTGC GGCGCCATTAAATAACGTGTACTTGTCTATTCTTGTGCGTGTGGTCTTGGGAAAAGAAAGCTTGCT
GGAGGCTGCTGTTTCAGCCCCATACATTACTTGTACGATTCTGCTGACTTTTCGGCGGGTGCAATATCTCTACTTCTGCT
TGACGAGGTATTGTTGCCTGTACTTCTTCTTCTTCTTCTTCTGCTGATTGGTCTATAAGAAATCTAGTATTTTCTTTGA
AACAGAGTTTTTCCCGTGGTTTTCGAACTTGGAGAAAGATTGTTAAGCTTCTGTATATTCTGCCCAAATTTGTGCGGGCC
AATACCGCGGAGAAAATGGCGAAAAACGTTGCGATTTTCGGCTTATTGTTTTCTTCTTCTTGTGTTGGTTCCCTTCTCAGA
TCTTCGCCCCAAAACCTTCCTGGAAATGACAACAGCACGGCAACGCTGTGCCTTGGGCACCATGCAGTACCAACCGGAAC
GATAGTGAAAACAATCAGCAATGACCAAAATTGAAGTTACTAATGCTACTGAGCTGGTTCAGAATTCTCTCAATAGGTGAA
ATATGCGACAGTCCTCATCAGATCCTTGATGGAGAAAACGACACTAATAGATGCTCTATTGGGAGACCCCTCAGTGTG
ATGGCTTCCAAAATAAGAAATGGGACCTTTTTGTTGAACGAAGCAAGCCTACAGCAACTGTTACCCTTATGATGTGCC
GGATTATGCCCTCCCTTAGGTCACTAGTTGCCTCATCCGGCACACTGGAGTTAACAATGAAAGCTTCAATTGGACTGGA
GTCACCTCAAAACGGAACAAGTCTGCTTGCATAAGGAGATCTAATAATAGTTTCTTTAGTAGATTAAATTTGGTTGACCC
ACTTAACTTCAAATACCCAGCATTGAACGTGACTATGCCAAACAATGAACAATTTGACAAATTTGTACATTTGGGGGGT
TCACCACCCGGGTACGGACAAGGACCAATCTTCTGTATGCTCAATCATCAGGAAGAATCAGAGTATCTACCAAAAGA
AGCCAACAAGCTGTAATCCCGAATATCGGATCTAGACCCAGAATAAGGAATATCCCTAGCAGAATAAGCATCTATTGGA
CAATAGTAAAACCGGGAGACATACTTTTGATTAACAGCACAGGGAATCTAATTGCTCCTAGGGGTACTTCAAAATACG
AAGTGGGAAAAGCTCAATAATGAGATCAGATGCACCCATTGGCAAATGCAATCTGAATGCATCACTCCAAATGGAAGC
ATTTCCCAATGACAAACCATTCCAAAATGTAAACAGGATCACATACGGGGCCTGTCCAGATATGTTAAGCAAAGCACTC
TGAAATTTGGCAACAGGAATGCGAAATGTACCAGAGAAACAACTAGAGGCATATTTGGCGCAATAGCGGGTTTCATAGA
AAATGGTTGGGAGGGAATGGTGGATGGTTGGTACGGTTTCAGGCATCAAAATCTGAGGGAAGAGGACAAAGCAGCAGAT
CTCAAAAGCACTCAAGCAGCAATCGATCAAAATCAATGGGAAGCTGAATCGATTGATCGGGAAAACCAACGAGAAATTC
ATCAGATTGAAAAAGAAATTCAGAAAGTGAAGGAGAAATTCAGGACCTTGAGAAATATGTTGAGGACACTAAAATAGA
TCTCTGGTCAACAACGCGGAGCTTCTTGTGCGCTGGAGAACCAACATACAATTGATCTAACTGACTCAGAAATGAAC
AACTGTTTGAAAAACAAAGAAAGCAACTAAGGAAAAATGCTGAGGATATGGGCAATGGTTGTTTCAAAATATACCACA
AATGTGACAATGCCTGCATAGGATCAATCAGAAATGGAACCTATGACCACGATGTATACAGAGATGAAGCATTAACAA
CCGGTTCAGATCAAGGGAGTTGAGCTGAAGTCAGGGTACAAAGATTGGATCCTATGGATTTTCTTTGCCATATCATGT
TTTTTGCTTTGTGTTGCTTTGTGTTGGGTTTCATCATGTGGGCTGCCAAAAGGGCAACATTAGTGCAACATTTGCATT
GAAGCCTATTTCTTTAGTTTGAATTTACTGTTATTCGGTGTGCATTTCTATGTTTGGTGAGCGGTTTCTGTGCTCA
GAGTGTGTTATTTTATGTAATTTAATTTCTTTGTGAGCTCCTGTTTAGCAGGTGTCCTTTCAGCAAGGACACAAAAA
GATTTTAAATTTTATTAATAAAAAAAAAAAAAAAAAAGACCGGGAATTCGATATCAAGCTTATCGACCTGCAGATCGTTCAAA
CATTTGGCAATAAAGTTTCTTAAGATTGAATCCTGTTGCCGGTCTGCGATGATTATCATATAATTTCTGTTGAATTAC
GTTAAGCATGTAATAATTAACATGTAATGCATGACGTTATTTATGAGATGGGTTTTTATGATTAGAGTCCCGCAATTAT
ACATTTAATACGCGATAGAAAACAAATATAGCGCGCAAACTAGGATAAATTATCGCGCGGGTGTCTATGTTACT
AGAT

图6E 构建体编号1819的示意图

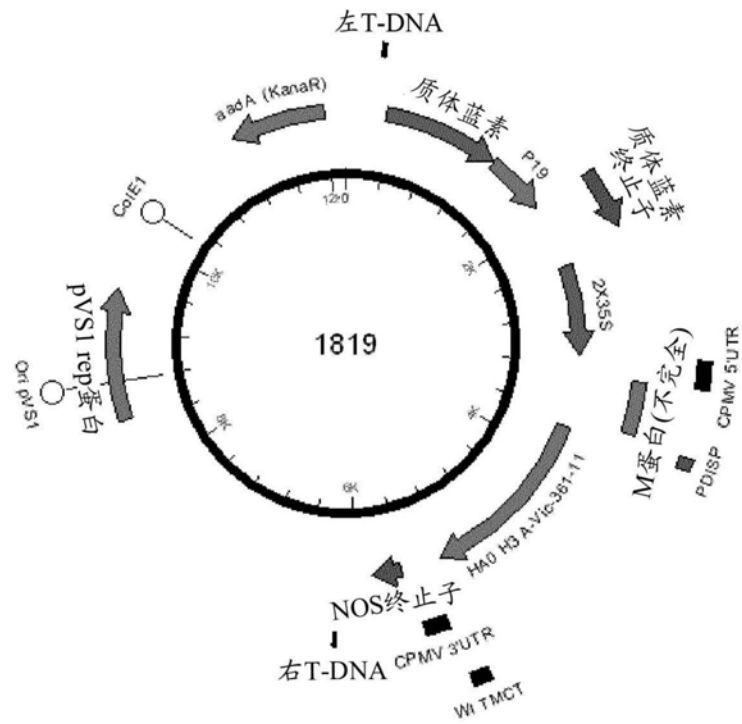


图6:B-2X35S/CPMV-HT+/PDISP/H3 Victoria/NOS (构建体编号1819)

图 7A (SEQ ID NO:27) IF-HT1*-PDI.c (来自原始1819构建体的修饰序列加下划线)

GTCGGGGCCCAATACCGCGGAGACAATGGCGAAAAACGTTGCGATTTTCGGCT

图 7B (SEQ ID NO:28) IF-HT2*-PDI.c (来自原始1819构建体的修饰序列加下划线)

GTCGGGGCCCAATACCGCGGAGGAAATGGCGAAAAACGTTGCGATTTTCGGCT

图 7C (SEQ ID NO:29) IF-HT3*-PDI.c (来自原始1819构建体的修饰序列加下划线)

GTCGGGGCCCAATACCGCGGAAAAAATGGCGAAAAACGTTGCGATTTTCGGCT

图 7D (SEQ ID NO:30) IF-HT4*-PDI.c (来自原始1819构建体的修饰序列加下划线)

GTCGGGGCCCAATACCGCGGAAACAATGGCGAAAAACGTTGCGATTTTCGGCT

图 7E (SEQ ID NO:31) IF-HT5*-PDI.c (来自原始1819构建体的修饰序列加下划线)

GTCGGGGCCCAATACCGCGGAAGCAATGGCGAAAAACGTTGCGATTTTCGGCT

图 7F (SEQ ID NO:32) IF-HT6*-PDI.c (来自原始1819构建体的修饰序列加下划线)

GTCGGGGCCCAATACCGCGGAAGAAATGGCGAAAAACGTTGCGATTTTCGGCT

图 7G (SEQ ID NO:33) IF-HT7*-PDI.c (来自原始1819构建体的修饰序列加下划线)

GTCGGGGCCCAATACCGCGGAAGAAATGGCGAAAAACGTTGCGATTTTCGGCT

图 7H (SEQ ID NO:34) IF-HT8*-PDI.c (来自原始1819构建体的修饰序列加下划线)

GTCGGGGCCCAATACCGCGGAAAAGAAATGGCGAAAAACGTTGCGATTTTCGGCT

图 7I: 构建体编号1952的示意图。使用类似的特征来制备构建体1953-1959

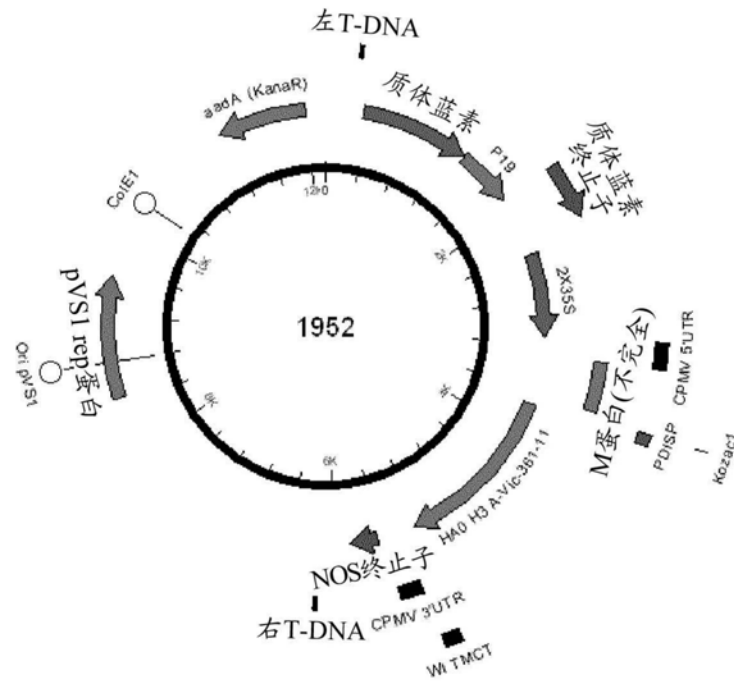


图7:2X35S/CPMV HT+/NOS表达系统中的SacII限制性位点与PDISP/H3 Victoria的ATG之间的序列的E-变型

图 8A (SEQ ID NO: 35)
PDISP/H1 California的核苷酸序列

ATGGCGAAAAACGTTGCGATTTTCGGCTTATTGTTTTCTTCTTGTTGGTTCCTTCTCAGATCTTCGCTGACACATTATGTATAG
GTTATCATGCGAACAAATCAACAGACACTGTAGACACAGTACTAGAAAAGAATGTAACAGTAACACACTCTGTTAACCTTCTAGAA
GACAAGCATAACGGGAACTATGCAAACTAAGAGGGGTAGCCCCATTGCATTTGGGTAAATGTAACATTGCTGGCTGGATCCTGG
GAAATCCAGAGTGTGAATCACTCTCCACAGCAAGCTCATGGTCTACATTGTGGAAACACCTAGTTCAGACAATGGAACGTGTTAC
CCAGGAGATTTATCGATTATGAGGAGCTAAGAGAGCAATTGAGCTCAGTGTATCATTTGAAAGGTTTGAGATATCCCCAAGA
CAAGTTCATGGCCCAATCATGACTCGAACAAAGGTGTACGGCAGCATGCTCTCATGCTGGAGCAAAAAGCTTACAAAAATTT
AATATGGCTAGTTAAAAAGGAAATTCATACCCAAAGCTCAGCAAATCCTACATTAATGATAAAGGGAAAGAAGTCTCGTGCTA
TGGGGCATTACCATCCATCTACTAGTGTGACCAACAAAGTCTCTATCAGAATGCAGATGCATATGTTTTGTGGGGTCATCAAG
ATACAGCAAGAAGTTCAAGCCGGAATAGCAATAAGACCCAAAGTGAGGGATCAAGAAGGGAGAATGAATATTACTGGACACT
AGTAGAGCCGGGAGACAAAATAACATTGCAAGCACTGGAATCTAGTGGTACCGAGATATGCATTGCAATGGAAAGAAATGC
TGGATCTGGTATTATCATTTAGATACACAGTCCACGATTGCAATACAACCTGTCAAACACCAAGGGTGTATAAACACACAGCC
TCCCATTTCAGAAATATACATCCGATCACAATTGGAAAAATGTCCAAAATATGTAAAAAGCACAAAATTGAGACTGGCCACAGGATTG
AGGAATATCCCGTCTATTCAATCTAGAGGACTATTTGGGGCCATTGCCGGTTTCATTGAAGGGGGGTGGACAGGGATGGTAGAT
GGATGGTACGGTTATCACCATCAAAATGAGCAGGGGTGAGGATATGCAGCCGACCTGAAGAGCACACAGAATGCCATTGACGAG
ATTACTAACAAAGTAAATCTGTTATTGAAAAGATGAATACACAGTTTCACAGCAGTAGGTAAAGAGTTCAACCACCTGGAAAAA
GAATAGAGAATTTAAATAAAAAAGTTGATGATGGTTTCTGGACATTTGGACTTACAATGCCGAACGTGTTGGTTCTATTGAAAAA
GAAAGAATTTGGACTACACGATTCAAATGTGAAGAACTTATATGAAAAGGTAAGAAGCCAGCTAAAAACAATGCCAAGGAA
ATTGGAACGGCTGCTTTGAATTTACCACAAATGCGATAACACGTGCATGGAAAGTGTCAAAAATGGGACTTATGACTACCCAA
AATACTCAGAGGAAGCAAAATTAACAGAGAAGAAATAGATGGGGTAAAGCTGGAATCAACAAGGATTACCAGATTTGGCGA
TCTATTCACTGTCGCCAGTTCAATGGTACTGGTAGTCTCCTGGGGGAATCAGTTTCTGGATGTGCTCTAATGGGTCTCTACAGT
GTAGAATATGTATTAA

图 8B (SEQ ID NO :36)
PDISP/H1 California的氨基酸序列

MAKNVAIFGLLFSLLVLPVSQIFADTLCIGYHANNSTDVDTVLEKNVTVTHSVNLLDKHNGKLCRLRGVAPLHLGKCNIAGWILGNPE
CESLSTASSWSYIVETPSSDNGTCYPGDFIDYEELREQLSSVSSEFERFEIPKTSWPNHDSNKGVTAAACPHAGAKSFYKNLIWLKKGNS
YPKLSKSYINDKGKEVLVLWGIHHPSTSADQQSLYQNADAYVFGSSRYSKFKPEIAIRPKVRDQEGRMNYYWTLVEPGDKITFEATGN
LVVPRYAFAMERNAGSGIIISDTPVHDCNTTCQTPKGAINSLPFQNIHPITIGKCPKYVKSTKLRLATGLRNIPISQSRGLFGAIAAGFIEGG
WTGMVDGWYGYHHQNEQGGGYAADLKSTQNAIDEITNKVNSVIEKMNTQFTAVGKEFNHLEKRIENLNKKVDDGFLDIWTYNAELL
VLENERLTDYHDSNVKNLYEKVRSQKNNAKEIGNGCFEYHKCDNTCMESVKNGTYDYPKYSEEAKLNREEIDGVKLESTRIYQILAIYS
TVASSLVVSLGAISFWMCSNGLQCRICI*

图8C 构建体编号484 (2X35S/CPMV HT)的示意图

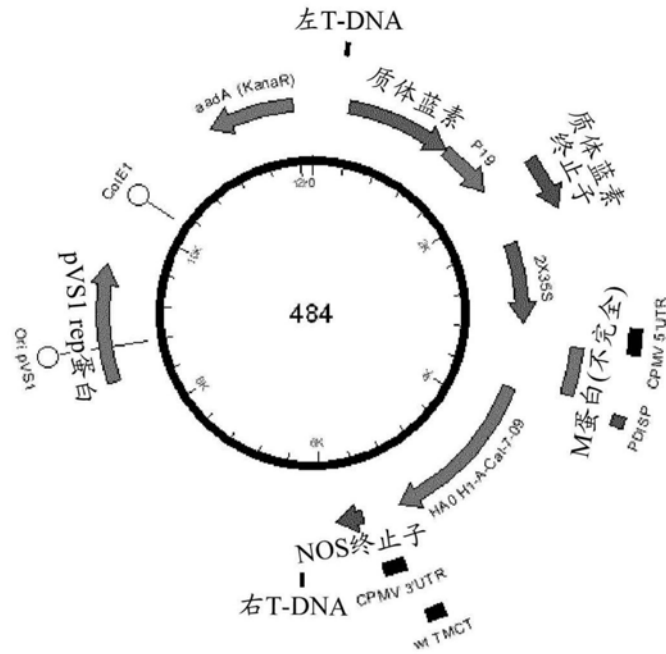


图8D 构建体编号1805 (2X35S/CPMV HT+)的示意图

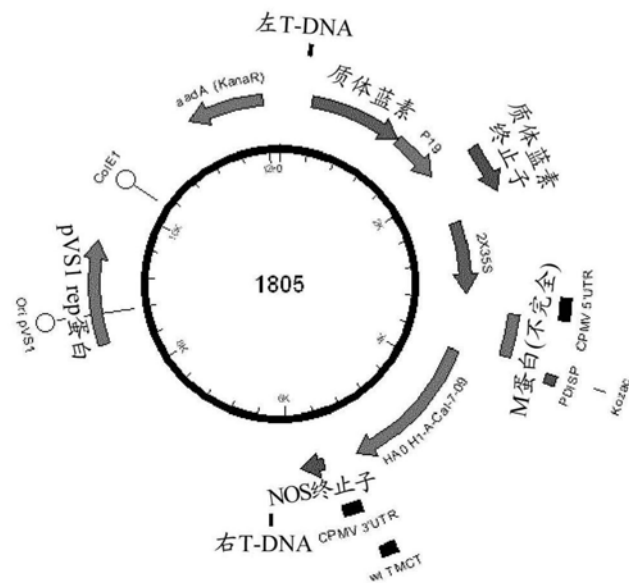


图8:PDISP/H1 California的G-2X35S/CPMV HT (构建体编号484) 和HT+ (构建体编号1805)

图9A (SEQ ID NO: 37)

PDISP/H5 Indonesia的核苷酸序列

ATGGCGAAAAACGTTGCGATTTTCGGCTTATTGTTTTCTTCTTGTTGGTTCCTTCTCAGATCTTCGCCGATCAGATTGCAATTG
GTTACCATGCAAAACAATTCAACAGAGCAGGTTGACACAATCATGGAAGAAGACGTTACTGTTACACATGCCAAGACATACTGGA
AAAGACACACAACGGGAAGCTCTGCGATCTAGATGGAGTGAAGCCTCTAATTTAAGAGATTGTAGTGTAGCTGGATGGCTCCTC
GGGAACCAATGTGTGACGAATTCATCAATGTACCGGAATGGTCTTACATAGTGGAGAAGGCCAATCCAACCAATGACCTCTGTT
ACCCAGGGAGTTTCAACGACTATGAAGAACTGAAACACCTATTGAGCAGAATAAACCATTTTGAGAAAATCAAATCATCCCCAAA
AGTTCTTGGTCCGATCATGAAGCCTCATCAGGAGTTAGCTCAGCATGTCCATACCTGGGAAGTCCCTCTTTTGTAGAAATGTGGT
ATGGCTTATCAAAAAGAAGTACATACCCAACAATAAAGAAAAGCTACAATAATACCAACCAAGAGGATCTTTGGTACTGTGG
GGAATTCACCATCCTAATGATGCGGCAGAGCAGACAAGGCTATATCAAAACCAACCACTATATTTCCATTGGGACATCAACACT
AAACCAGAGATTGGTACCAAAAATAGCTACTAGATCCAAAGTAAACGGGCAAGTGGGAAGGATGGAGTTCCTTCTGGACAATTTTA
AAACCTAATGATGCAATCAACTTCGAGAGTAATGGAATTTTCAATGCTCCAGAATATGCATACAAAATTGTCAAGAAAGGGGACTC
AGCAATTATGAAAAGTGAATTGGAATATGGTAAGTGAACACCAAGTGTCAAACTCCAATGGGGGCGATAAACTCTAGTATGCCA
TTCCACAACATACACCTCTCACCATCGGGGAATGCCCCAAATATGTGAAATCAAACAGATTAGTCTTGAACAGGGGCTCAGAAA
TAGCCCTCAAAGAGAGAGCAGAAGAAAAAGAGAGGACTATTTGGAGCTATAGCAGGTTTATAGAGGGAGGATGGCAGGGAA
TGGTAGATGGTTGGTATGGGTACCACCATAGCAATGAGCAGGGGAGTGGGTACGCTGCAGACAAAGAATCCACTCAAAAGGCAA
TAGATGGAGTCACCAATAAGGTCAACTCAATCATTGACAAAATGAACACTCAGTTTGAGGCCGTTGGAAGGGAATTTAATAACTT
AGAAAGGAGAATAGAGAATTTAAACAAGAAGTGAAGACGGGTTTCTAGATGTCTGGACTTATAATGCCGAACCTCTGGTTCTC
ATGGAATGAGAGAATCTAGACTTTTATGACTCAATGTTAAGAACCTCTACGACAAGGTCGACTACAGCTTAGGGATAATG
CAAAGGAGCTGGGTAACGGTGTTCGAGTTCTATCACAATGTGATAATGAATGTATGGAAAGTATAAGAAACGGAACGTACAA
CTATCCGAGTATTACAGAAGAAGCAAGATTAAGAGAGGAAATAAGTGGGGTAAATGGAATCAATAGGAACCTACCAAT
ACTGTCAATTTATTCAACAGTGGCGAGTTCCTAGCACTGGCAATCATGATGGCTGGTCTATCTTTATGGATGTGTCCAATGGATC
GTTACAATGCAGAATTTGCATT

图9B (SEQ ID NO: 38)

PDISP/H5 Indonesia的氨基酸序列

MAKNVAIFGLLFSLLVLPVSQIFADQICIGYHANNSTEQVDTIMEKNVTVTHAQDILEKTHNGKLCDLGDKVLPILRDCSVAGWLLGNP
MCDEFINVPESYIVEKANPTNDLCYPGSFNDYEELKHLISRINHFEKIQIPKSSWSDHEASSGVSSACPYLGSFRRNVVWLKKNSTY
PTIKKSYNNTNQEDLLVLWGIHPNDAAEQTRLYQNPTTYISIGTSLNQRLVPKIATRSKVNGQSGRMEFFWTILKPNDAINFESNGNF
IAPEYAYKIVKKGDSAIMKSELEYGNCNTKCQTPMGAINSSMPFHNIHPLTIGECPKYVKSRLVATGLRNSPQRESRRKRGLFGAIAAG
FIEGGWQGMVDGWYGYHHSNEQSGYAADKESTQKAIDGVTNKVNSIIDKMNTQFEAVGREFNLERRIENLNKKMEDGFLDVWT
YNAELLVLMENERTLDFHDSNVKNLYDKVRLQLRDNAKELNGCFEYHKKCDNECMESIRNGTYNYPQYSEEARLKREEISGVKLESIGT
YQILSIYSTVASSLALAIMMAGLSLWMCNSGLQCRICI

图 9C 构建体编号489 (2X35S/CPMV HT)的示意图

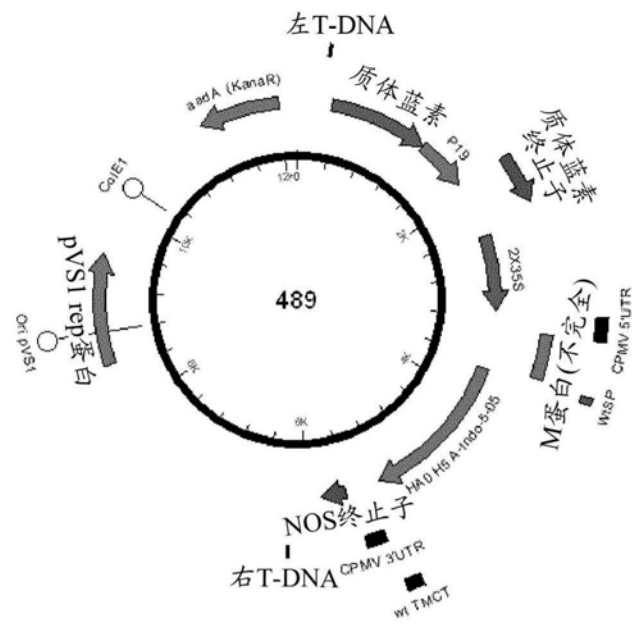


图 9D 构建体编号1889 (2X35S/CPMV HT+)的示意图

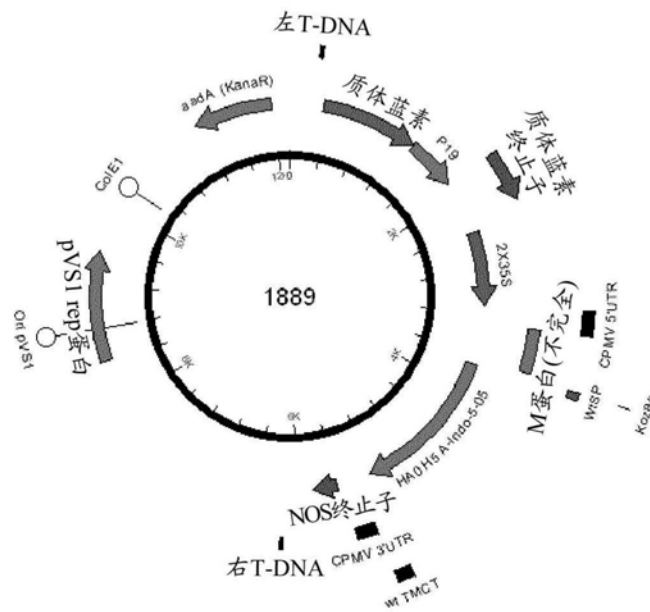


图9:H5 Indonesia的2X35S/CPMV HT (构建体编号409) 和HT+ (构建体编号2319)

图 10A (SEQ ID NO : 39)

PDISP/H7 Hangzhou的核苷酸序列

```
ATGGCGAAAAACGTTGCGATTTTCGGCTTATTGTTTTCTTCTTGTGTTGGTTCCTTCTCAGATCTTCGCGGACAAAATCTGCCTCG
GACATCATGCCGTGTCAAACGGAACCAAAGTAAACACATTAAGTAAAGAGGAGTGGAAAGTCGTCATGCAACTGAAACAGTGG
AACGAACAAACATCCCCAGGATCTGCTCAAAAGGGAAAAGGACAGTTGACCTCGGTCAATGTGGACTCTGGGGACAATCACTG
GACCACTCAATGTGACCAATTCCTAGAATTTTCAGCCGATTAAATTATTGAGAGGCGAGAAGGAAGTGATGTCTGTTATCCTGGG
AAATTCGTGAATGAAGAAGCTCTGAGGCAAATTCTCAGAGAATCAGGCGGAATTGACAAGGAAGCAATGGGATTCACATACAGT
GGAATAAGAACTAATGGAGCAACCAAGTGCATGTAGGAGATCAGGATCTTCATTCTATGCAGAAATGAAATGGCTCCTGTCAAACA
CAGATAATGTGCATTCCCGCAGATGACTAAGTCATATAAAATACAAGAAAAAGCCAGCTCTAATAGTATGGGGGATCCATCA
TTCCGTATCAACTGCAGAGCAAACCAAGCTATATGGGAGTGGAAACAACTGGTGACAGTTGGGAGTTCTAATTATCAACAATCT
TTTGTACCGAGTCCAGGAGCGAGACCACAAGTTAATGGTATATCTGGAAGAATTGACTTTCATTGGCTAATGCTAAATCCCAATGA
TACAGTCACTTTCAGTTTCAATGGGGCTTTCATAGCTCCAGACCGTGCAAGCTTCTGAGAGGAAAATCTATGGGAATCCAGAGTG
GAGTACAGGTTGATGCCAATTGTGAAGGGGACTGCTATCATAGTGGAGGGACAATAATAAGTAACTTGCCATTTCAGAACATAGA
TAGCAGGGCAGTTGGAAAATGTCCGAGATATGTTAAGCAAAGGAGTCTGCTGCTAGCAACAGGGATGAAGAATGTTCTGAGAT
TCCAAAGGGAAGAGGCCTATTTGGTGCTATAGCGGGTTTCATTGAAAATGGATGGGAAGGCCTAATTGATGGTTGGTATGGTTTC
AGACACCAGAATGCACAGGGAGAGGGAAGTGTGCAGATTACAAAAGCACTCAATCGGCAATTGATCAAATAACAGGAAAATTA
AACCGGCTTATAGAAAAACCAACCAACAATTTGAGTTGATCGACAATGAATTCATGAGGTAGAGAAGCAAATCGGTAATGTGA
TAAATTGGACCAGAGATTCTATAACAGAAGTGTGGTCATACATGCTGAACCTTTGGTAGCAATGGAGAACCAGCATACAATTGA
TCTGGCTGATTACAGAAATGGACAACTGTACGAACGAGTGAAAAGACAGCTGAGAGAGAATGCTGAAGAAGATGGCACTGGTTG
CTTTGAAATATTTACAAAGTGTGATGATGACTGTATGGCCAGTATTAGAAATAACACCTATGATCACAGCAAATACAGGGAAGAG
GCAATGCAAAATAGAATACAGATTGACCCAGTCAAACCTAAGCAGCGGCTACAAAGATGTGATACTTTGGTTTAGCTTCGGGGCAT
CATGTTTCATACTTCTAGCCATTGTAATGGGCCTTGTCTTCATATGTGTAAAGAATGGAACATGCGGTGCACTATTTGTATATAA
```

图 10B (SEQ ID NO :40)

PDISP/H7 Hangzhou的氨基酸序列

```
MAKNVAIFGLLFSLLVLPVSIQIFADKICLGHHAVSNGTKVNTLTERGVEVVNATETVERTNIPRICSKGKRTVDLQQCGLLGITGPPQCD
QFLEFSADLIIRREGSDVCYPGKFVNEEALRQILRESGGIDKEAMGFTYSGIRTNGATSACRRSGSSFYAEMKWLLSNTDNAAFPMQTK
SYKNTRKSPALIVWGIHHSVSTAEQTKLYGSGNKLVTGSSNYQQSFVPSPGARPQVNGISGRIDFWLMLNPNDTVTFNFNGAFIAPD
RASFLRGKSMGIQSGVQVDANCEGDCYHSGGTIISNLPFQNIIDSRVAGKCPRYVKQRSLLATGMKNVPEIPKGRGLFGAIAIGFIENGW
EGLIDGWYGRFRHQNAQEGTAADYKSTQSAIDQITGKLNRLIEKTNQQFELIDNEFNEVEKQIGNVINWTRDSITEVWSYNAELLVAME
NQHTIDLADSEMDKLYERVKRQLRENAEEDGTGCFEIFHKDDDCMASIRNNTYDHSKYREEAMQNRIQIDPVKLSSGYKDVILWFSFG
ASCFILLAIVMGLVFICVKNGNMRCTICI*
```

图 10C 构建体编号2140 (2X35S/CPMV HT)的示意图

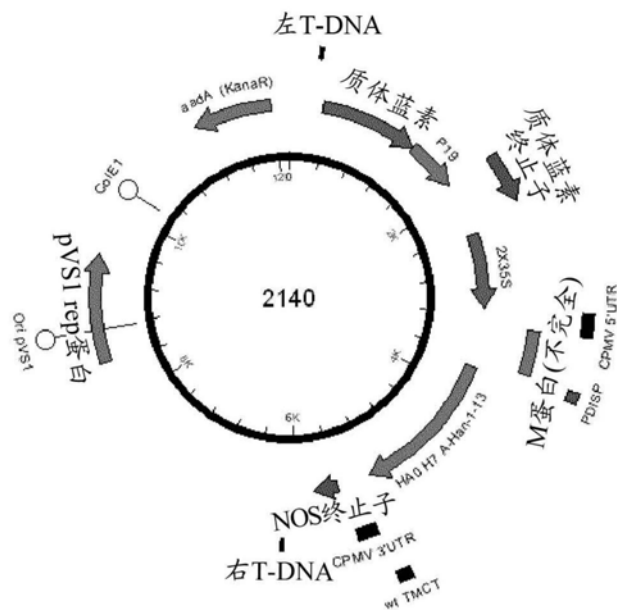


图 10D 构建体编号2142 (2X35S/CPMV HT+)的示意图

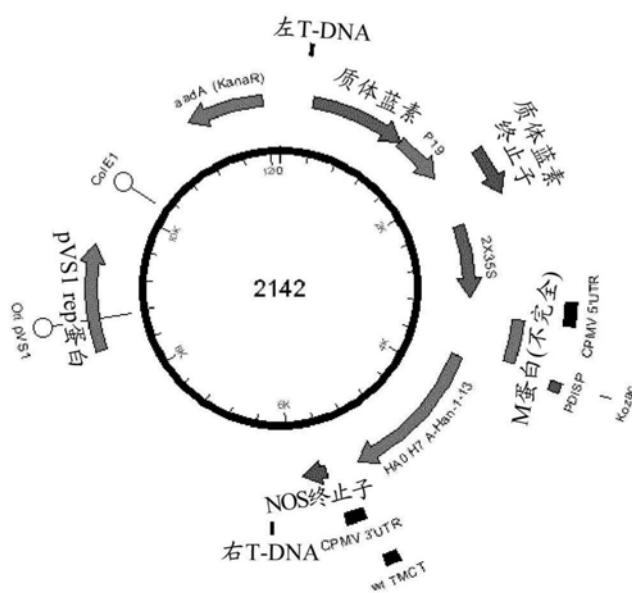


图10:PDISP/H7 Hangzhou的I-2X35S/CPMV HT H7 (构建体编号2140) 和HT+ (构建体编号2142)

图 11A (SEQ ID NO : 41)

PDISP/H7 Hangzhou+H5 Indonesia TMCT的核苷酸序列

ATGGCGAAAAACGTTGCGATTTTCGGCTTATTGTTTCTCTTCTTGTTGGTTCCTTCTCAGATCTTCGCGGACAAAATCTGCCTCG
GACATCATGCCGTGTCAAACGGAACCAAGTAAACACATTAAGTAAAGAGGAGTGGAAGTCGTCATGCAACTGAAACAGTGG
AACGAACAAACATCCCCAGGATCTGCTCAAAGGGAAAAGGACAGTTGACCTCGGTCAATGTGGACTCCTGGGGACAATCACTG
GACCACCTCAATGTGACCAATTCCTAGAAATTTTCAGCCGATTTAATTATTGAGAGGCGAGAAGGAAGTGATGTCTGTTATCCTGGG
AAATTCGTGAATGAAGAAGCTCTGAGGCAAATTCAGAGAATCAGGCGGAATTGACAAGGAAGCAATGGGATTCACATACAGT
GGAATAAGAACTAATGGAGCAACCAAGTGCATGTAGGAGATCAGGATCTTCATTCTATGCAGAAATGAAATGGCTCCTGTCAAACA
CAGATAATGTGCATTCCCGCAGATGACTAAGTCATATAAAAATACAAGAAAAAGCCAGCTCTAATAGTATGGGGGATCCATCA
TTCCGTATCAACTGCAGAGCAAACCAAGCTATATGGGAGTGGAACAAACTGGTGACAGTTGGGAGTTCTAATTATCAACAATCT
TTTGTACCGAGTCCAGGAGCGAGACCACAAGTTAATGGTATATCTGGAAGAATTGACTTTCATTGGCTAATGCTAAATCCCAATGA
TACAGTCACTTTCAAGTTTCAATGGGGCTTTCATAGCTCCAGACCGTGCAAGCTTCTGAGAGGAAAATCTATGGGAATCCAGAGTG
GAGTACAGGTTGATGCCAATTGTGAAGGGGACTGTATCATAGTGAGGGACAATAAAGTAACTTGCCATTTCCAGAACATAGA
TAGCAGGGCAGTTGGAAAATGTCCGAGATATGTTAAGCAAAGGAGTCTGCTGCTAGCAACAGGGATGAAGAATGTTCTGAGAT
TCCAAAGGGAAGAGGCCTATTTGGTGCTATAGCGGGTTTCATTGAAAATGGATGGGAAGGCCTAATTGATGGTTGGTATGGTTTC
AGACACCAGAATGCACAGGGAGAGGGAACTGCTGCAGATTACAAAAGCACTCAATCGGCAATTGATCAAATAACAGGAAAATTA
AACCGGCTTATAGAAAAACCAACCAACAATTTGAGTTGATCGACAATGAATTCATGAGGTAGAGAAGCAAATCGGTAATGTGA
TAAATTGGACCAGAGATTCTATAACAGAAGTGTGGTCATACATGCTGAACCTTTGGTAGCAATGGAGAACCAGCATACAATTGA
TCTGGCTGATTACAGAAATGGACAACTGTACGAACGAGTGAAAAGACAGCTGAGAGAGAATGCTGAAGAAGATGGCACTGGTTG
CTTTGAAATATTTACAAAGTGTGATGATGACTGTATGGCCAGTATTAGAAATAACACCTATGATCACAGCAAATACAGGGAAGAG
GCAATGCAAAATAGAAATACAGATTGACCCAGTCAAACCTAAGCAGCGGCTACCAAATACTGTCAATTTATTCAACAGTGGCGAGTTC
CCTAGCACTGGCAATCATGATGGCTGGTCTATCTTTATGGATGTGCTCCAATGGATCGTTACAATGCAGAATTTGCATTTAA

图 11B (SEQ ID NO : 42)

PDISP/H7 Hangzhou+H5 Indonesia TMCT的氨基酸序列

MAKNVAIFGLLFSLLVLPVSQIFADKICLGHAVSNGTKVNTLTERGVEVVNATETVERTNIPRICSKGKRTVDLGQCGLLGTITGPPQCD
QFLEFSADLIERREGSDVCYPGKFVNEEALRQILRESGGIDKEAMGFTYSGIRTNGATSACRRSGSSFYAEMKWLLSNTDAAFPQMTK
SYKNTRKSPALIVWGIHHSVSTAEQTKLYGSGNKLVTGSSNYQQSFVSPGARPQVNGISGRIDFWLMLNPNDTVTFSENGAFIAPD
RASFLRGKSMGIQSGVQVDANCEGDCYHSGGTIISNLPFQNIIDSRVVGKCPRYVKQRSLLATGMKNVPEIPKGRGLFGAIAIGFIENGW
EGLIDGWYGRFHQNAQGEFTAADYKSTQSAIDQITGKLNRLIEKTNQQFELIDNEFNEVEKQIGNVINWTRDSITEVWSYNAELLVAME
NQHTIDLADSEMDKLYERVKRQLRENAEEDGTGCFEIFHKDDDCMASIRNNTYDHSKYREEAMQNRQIDPVLKSSGYQLSIYSTVAS
SLALAIMMAGLSLWMCNSGSLQCRICI*

图 11C 构建体编号2130 (2X35S/CPMV HT)的示意图

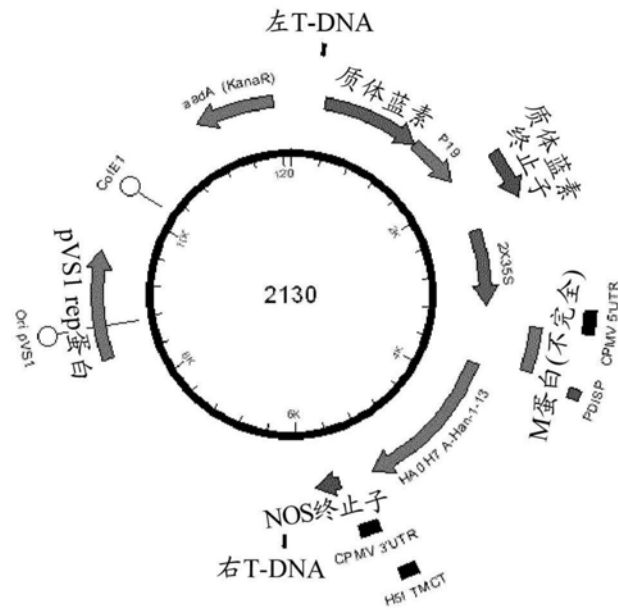


图 11D 构建体编号2146 (2X35S/CPMV HT+)的示意图

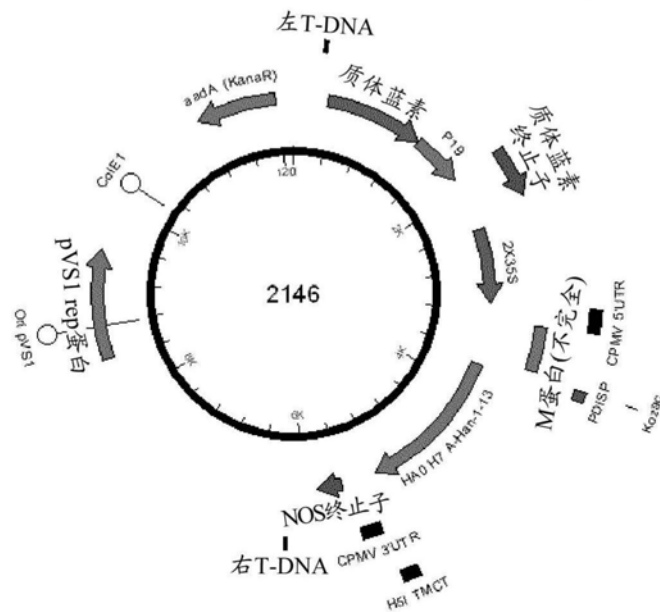


图11:PDISP/H7 Hangzhou+H5 Indonesia TMCT的2X35S/CPMV HT (构建体编号2130) 和HT+ (构建体编号2146)

图 12A (SEQ ID NO : 43)

PDISP/HA B Brisbane (PrL-)的核苷酸序列

ATGGCGAAAAACGTTGCGATTTTCGGCTTATTGTTTTCTTCTTGTTGGTTCCTTCTCAGATCTTCGCCGATCGAATCTGCACTG
GAATAACATCGTCAAACCTACCACATGTCGTCAAACCTGCTACTCAAGGGGAGGTCAATGTGACTGGTGAATACCACTGACAAC
AACACCCACCAAATCTCATTTTGCAAATCTCAAAGGAACAGAAACCAGGGGAAACTATGCCCAAATGCCTCAACTGCACAGAT
CTGGACGTAGCCTTGGGCAGACCAAATGCACGGGGAAAAATACCTCGGCAAGAGTTCAATACTCCATGAAGTCAGACCTGTTA
CATCTGGGTGCTTCTATAATGCACGACAGAACAAAAATTAGACAGCTGCCTAACCTTCTCCGAGGATACGAACATATCAGGTTA
TCAACCCATAACGTTATCAATGCAGAAAATGCACCAGGAGGACCTACAAAATTGGAACCTCAGGGTCTTGCCCTAACATTACCAA
TGGAACGGATTTTTCGAACAATGGCTTGGGCCGTCCAAAAAACGACAAAAACAAACAGCAACAAATCCATTAACAATAGAA
GTACCATACATTTGTACAGAAGGAGAAGACCAATTACCGTTTGGGGTTCCTACTCTGACAACGAGACCCAAATGGCAAAGCTCT
ATGGGGACTCAAAGCCCCAGAAAGTTCACCTCATCTGCCAAGGAGTGACCACACATTACGTTTCACAGATTGGTGGCTTCCCAAAT
CAAACAGAAGACGAGGACTACCACAAAGTGGTAGAATTGTTGTTGATTACATGGTGCAAAAATCTGGGAAAAACAGGAACAATT
ACCTATCAAAGGGGTATTTTATTGCCTCAAAGGTGTGGTGCGCAAGTGGCAGGAGCAAGGTAATAAAAGGATCCTTGCCTTTAA
TTGGAGAAGCAGATTGCCTCCACGAAAAATACGGTGGATTAAACAAAAGCAAGCCTTACTACACAGGGGAACATGCAAAGGCCA
TAGGAAATTGCCAATATGGGTGAAACACCCCTGAAGCTGGCCAATGGAACCAAATATAGACCTCCTGGTGGAGGATGGGAAG
GAATGATTGCAGTTGGCAGGATACACATCCCATGGGGCACATGGAGTAGCGGTGGCAGCAGACCTTAAGAGCACTCAAGAGG
CCATAACAAGATACAAAAAATCTCAACTCTTTGAGTGAGCTGGAAGTAAAGAATCTTCAAAGACTAAGCGGTGCCATGGATGA
ACTCCACAACGAAATACTAGAACTAGATGAGAAAGTGGATGATCTCAGAGCTGATACAATAAGCTCACAAATAGAACTCGCAGTC
CTGCTTCCAATGAAGGAATAATAAACAGTGAAGATGAACATCTTGGCGCTTGAAAGAAAGCTGAAGAAAATGCTGGGCCCT
CTGCTGTAGAGATAGGGAATGGATGCTTTGAAACCAACACAAGTGCAACCAGACCTGCTCGACAGAATAGCTGCTGGTACCTT
TGATGCAGGAGAATTTCTCTCCACCTTTGATTCACTGAATTAATGCTGCATCTTTAAATGACGATGGATTGGATAATCATA
TATACTGCTTACTACTCAACTGCTGCCCTCCAGTTTGGCTGTAACACTGATGATAGCTATCTTTGTTGTTTATATGGTCTCCAGAGAC
AATGTTTCTGCTCCATCTGTCTATAA

图 12B (SEQ ID NO :44)

PDISP/HA B Brisbane (PrL-)的氨基酸序列

MAKNVAIFGLLFSLLVLVPSQIFADRICTGITSSNSPHVVKATQGEVNVTVIPLTTPTKSHFANLKGTETRGKLCPKCLNCTDLVALG
RPKCTGKIP SARVSILHEVRPVTSGCFPIIMHRTKIRQLPNLLRGYEHIRLSTHNVINAENAPGGPYKIGTSGSCPNTNNGNFFATMAW
AVPKNDKNKTATNPLTIEVPYICTEGEDQITVWGFHSDNETQMAKLYGDSKPQKFTSSANGVTTHYVSQIGGFNPQTEDGGLPQSGRI
VVDYMVQKSGKTGTITYQRGILLPQKVWCASGRSKVIKGSPLIGEADCLHEKYGGLNKSPPYTGHAHAIGNCPIWVKTPCLKLANGTK
YRPPGGGWEGMIAGWGHYTSAGAHGVAADLKSTQEAINKITKNLSLSELEVKNLQRLSGAMDELHNEILELDEKVDLDRADTISS
QIELAVLLSNEGIINSEDEHLLALERKLKMLGPSAVEIGNGCFETKHKNQCTLDRIAAGTFDAGEFSLPTFDSLNTAASLNDGDLNHT
ILLYYSTAASSLAVTLMIAIFVVMVSRDNVSCSICL*

图 12C 构建体编号1039 (2X35S/CPMV HT)的示意图

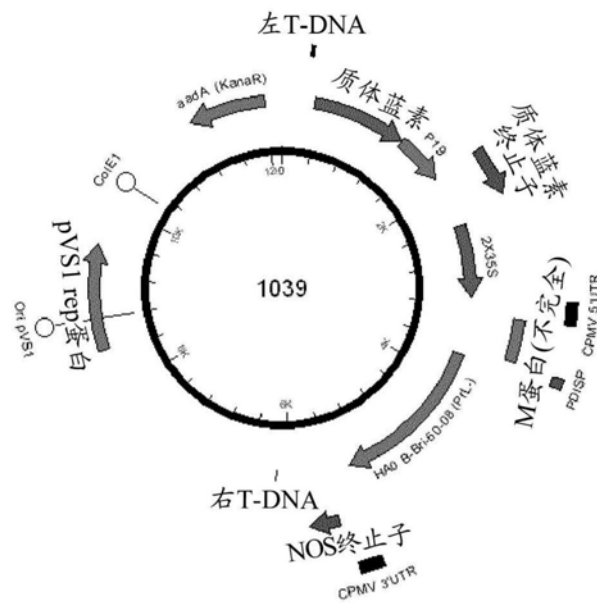


图 12D 构建体编号1829 (2X35S/CPMV HT+)示意图

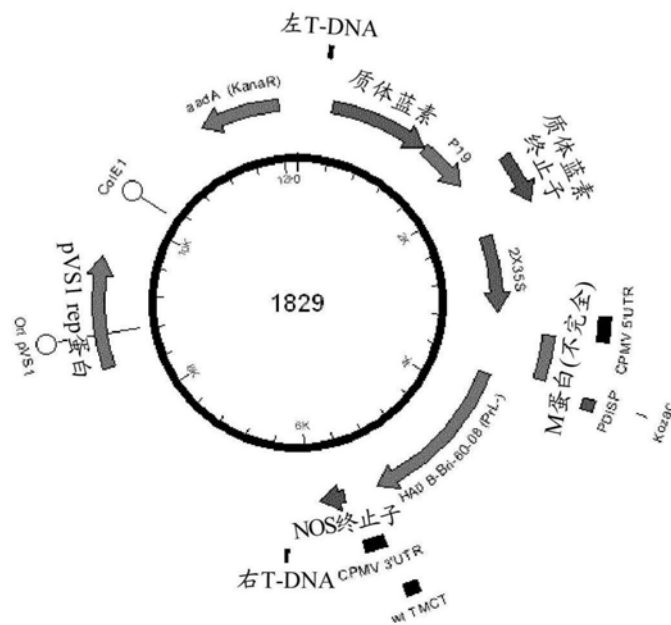


图12:PDISP/HA B Brisbane (PrL-) 的2X35S/CPMV HT (构建体编号1039) 和K-2X35S/CPMV H T+ (构建体编号1829)

图 13A (SEQ ID NO : 45)

PDISP/HA B Brisbane (PrL-)+H1 California TMCT的核苷酸序列

```
ATGGCGAAAAACGTTGCGATTTTCGGCTTATTGTTTTCTTCTTGTGTTGGTTCCTTCTCAGATCTTCGCCGATCGAATCTGCACTG
GAATAACATCGTCAAACCTACCACATGTCGTCAAACTGCTACTCAAGGGGAGGTCAATGTGACTGGTGAATACCACTGACAAC
AACACCCACCAAATCTCATTTTGCAAATCTCAAAGGAACAGAAACCAGGGGGAAACTATGCCCAAAATGCCTCAACTGCACAGAT
CTGGACGTAGCCTTGGGCAGACCAAAATGCACGGGGAAAAATACCTCGGCAAGAGTTTCAATACTCCATGAAGTCAGACCTGTTA
CATCTGGGTGCTTCTCTATAATGCACGACAGAACAAAAATTAGACAGCTGCCTAACCTTCTCCGAGGATACGAACATATCAGGTTA
TCAACCCATAACGTTATCAATGCAGAAAATGCACCAGGAGGACCCTACAAAATTGGAACCTCAGGGTCTTGCCCTAACATTACCAA
TGGAACGGATTTCGCAACAATGGCTTGGGCCGTCCAAAAAACGACAAAAACAAACAGCAACAAATCCATTAACAATAGAA
GTACCATACATTTGTACAGAAGGAGAAGACCAATTACCGTTTGGGGGTTCCACTCTGACAACGAGACCCAAATGGCAAAGCTCT
ATGGGGACTCAAAGCCCCAGAAAGTTCACCTCATCTGCCAACGGAGTGACCACACATTACGTTTCACAGATTGGTGGCTTCCCAAAT
CAAACAGAAGACGGAGGACTACCACAAAGTGGTAGAATTGTTGTTGATTACATGGTGCAAAAATCTGGGAAAAACAGGAACAATT
ACCTATCAAAGGGGTATTTTATTCCTCAAAGGTGTGGTGCGCAAGTGCGAGGAGCAAGGTAATAAAAGGATCCTTGCTTTAA
TTGGAGAAGCAGATTGCCTCCACGAAAAATACGGTGGATTAAACAAAAGCAAGCCTTACTACACAGGGGAACATGCAAAGGCCA
TAGGAAATTGCCAATATGGGTGAAACACCCCTGAAGCTGGCCAATGGAACCAATATAGACCTCCTGGTGGAGGATGGGAAG
GAATGATTGCAGGTTGGCACGGATACACATCCCATGGGGCACATGGAGTAGCGGTGGCAGCAGACCTTAAGAGCACTCAAGAGG
CCATAAACAGATAACAAAAATCTCAACTCTTTGAGTGAGCTGGAAGTAAAGAATCTTCAAAGACTAAGCGGTGCCATGGATGA
ACTCCACAACGAAATACTAGAACTAGATGAGAAAGTGGATGATCTCAGAGCTGATACAATAAGCTCACAAATAGAACTCGCAGTC
CTGCTTCCAATGAAGGAATAATAAACAGTGAAGATGAACATCTCTTGGCGCTTGAAAGAAAGCTGAAGAAAATGCTGGGCCCT
CTGCTGTAGAGATAGGAATGGATGCTTTGAAACCAACACAAAGTGCAACCAGACCTGTCTCGACAGAATAGCTGCTGGTACCTT
TGATGCAAGGAGAATTTCTCTCCACCTTTGATTACTGAATTAATGCTGTCATCTTAAATGACGATGGATTGGATAATTACCA
GATTTTGGCGATCTATTCAACTGTCGCCAGTTCATTGGTACTGGTAGTCTCCCTGGGGGCAATCAGTTTCTGGATGTGCTCTAATG
GGTCTCTACAGTGTAGAATATGTATTAA
```

图 13B (SEQ ID NO : 46)

PDISP/HA B Brisbane (PrL-)+H1 California TMCT的氨基酸序列

```
MAKNVAIFGLLFSLVLVPSQIFADRICTGITSSNSPHVVKATQGEVNVTVIPLTTPTKSHFANLKGTETRGKLCPKCLNCTDL DVALG
RPKCTGKIPSARVSVILHEVRPVTS GCFPIIMH DRTKIRQLPNLLRGYEHIRLSTHN VINAENAPGGPYKIGTS GSCP NITNGNGFFATMAW
AVPKNDKNKTATNPLTIEVPYICTEGEDQITVWGFHSDNETQMAKLYGDSKPQKFTSSANGVTTHYVSQIGGFNPQTEDGGLPQSGRI
VVDYMVQKSGKTGTITYQRGILLPQKVWCASGRSKVIKGSPLIGEADCLHEKYGGLNLSKPYTGEHAKAIGNCPIWVKTPKLKLANGTK
YRPPGGGWEGMIAGWHGYTSHGAHGVAVAADLKSTQEAINKITKNLSLSELEVKNLQRLSGAMDELHNEILELDEKVDLDRADTISS
QJELAVLLSNEGIINSEDEHLLALERKLLKMLGPSAVEIGNGCFETKHKNQTCLDRIAAGTFDAGEFSLPTFDSL NITAASLND DGLDNYQ
ILAIYSTVASSLVLVSLGAISFWMC SNGSLQCRICI*
```

图 13C 构建体编号1067 (2X35S/CPMV HT)的示意图

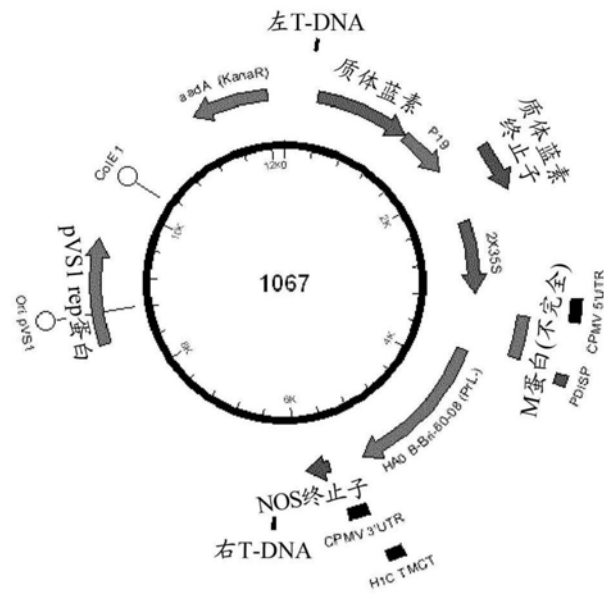


图 13D 构建体编号1875 (2X35S/CPMV HT+)的示意图

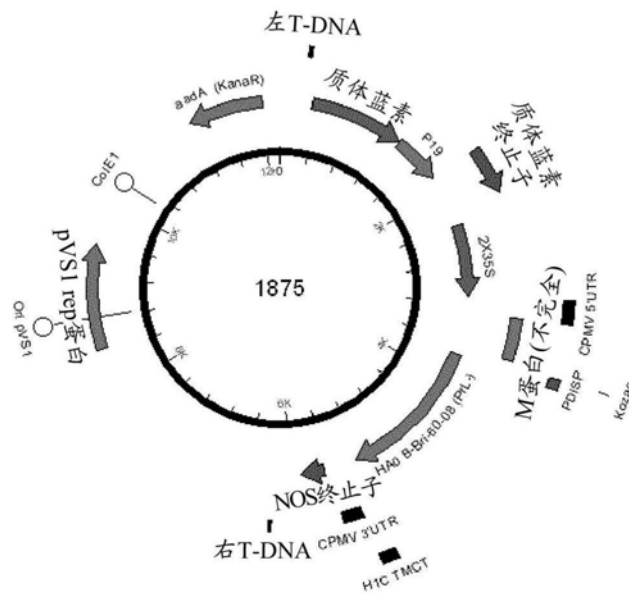


图13:PPDISP/HA B Brisbane (PrL-) +H1 California TMCT的L-2X35S/CPMV HT (构建体编号1067) 和HT+ (构建体编号1875)

图 14A (SEQ ID NO : 47)

PDISP/HA B Massachussetts (PrL-)的核苷酸序列

```
ATGGCGAAAAACGTTGCGATTTTCGGCTTATTGTTTTCTTCTTGTTGGTTCCTTCTCAGATCTTCGCCGATCGAATCTGCACTG
GGATAACATCTTCAAACCTCATGTGGTCAAAACAGCTACTCAAGGGGAGGTCAATGTGACTGGTGTGATACCACTAACAAAC
AACACCAACAAAATCTTATTTTGCAATCTCAAAGGAACAAAGACCAGAGGGAACTATGCCAGACTGTCTCAACTGTACAGATC
TGGATGTGGCCTGGGCAGGCCAATGTGTGTGGGAACACACCTTCTGCGAAAGCTTCAATACTTCACGAAGTCAGACCTGTTAC
ATCCGGGTGCTTCCCTATAATGCACGACAGAACAAAATCAGGCAACTAGCCAATCTTCTCAGAGGATATGAAAATATCAGGTTAT
CAACCCAAAACGTTATCGATGCAGAAAAGGCACCAGGAGGACCCTACAGACTTGGAACCTCAGGATCTTGCCTAACGCTACCAG
TAAAAGCGGATTTTTCGCAACAATGGCTTGGGCTGTCCCAAAGGACAACAACAAAATGCAACGAACCCATTAAACAGTAGAAGTA
CCATACATTTGTGCAGAAGGGGAAGACCAATTACTGTTGGGGGTTCCATTAGATAACAAAACCCAAATGAAGAACCTCTATG
GAGACTCAAATCCTCAAAAGTTCACCTCATCTGCTAATGGAGTAACACACATTATGTTTCTCAGATTGGCGGCTTCCAGATCAA
CAGAAGACGGAGGACTACCACAAAGCGGCAGAAATTGCGTTGATTACATGATGCAAAAACCTGGGAAAACAGGAACAATTGTCT
ATCAAAGAGGTGTTTTGTTGCCTCAAAAGGTGTGGTGCAGAGTGGCAGGAGCAAAGTAATAAAAGGGTCTTGCCTTTAATTGG
TGAAGCAGATTGCCTTCATGAAAAATACGGTGGATTAACAAAAGCAAGCCTTACTACACAGGAGAACATGCAAAAGCCATAGG
AAATTGCCCAATATGGGTGAAAACACCTTTGAAGCTTGCCCAATGGAACCAATATAGACCTCCTGGTGGAGGATGGGAAGGAAT
GATTGCAAGTTGGCACGGATACACATCTCACGGAGCACATGGAGTGGCAGTTGCTGCAGACCTTAAGAGCACACAAGAAGCTAT
AAACAAGATAACAAAAATCTCAACTCTTTGAGTGAGCTAGAAGTAAAGAATCTTCAAAGGCTAAGTGGTGCCATGGATGAACTC
CACACGAAATACTCGAGCTGGATGAGAAAGTGGATGACCTCAGAGCTGACACTATAAGTTCACAAATAGAACTTGCACTCTTGC
TTTCCAACGAAGGAATAATAACAGTGAAGACGAGCATCTATTGGCACTTGAGAGAAAATAAGAAAATGCTGGGTCCCTCTGC
TGTAAGACATAGGAAATGGATGCTTCGAAACCAACACAAATGCAACGAGACCTGCTTAGACAGGATAGCTGCTGGCACCTTAAAT
GCAGGAGAGTTTTCTCTCCCACTTTTGATTGATTGAACATTACTGCTGCATCTTTAAATGATGATGGATTGGATAACCATATA
CTGCTCTATTACTCAACTGCTGCTTAGTTGGCTGTAAACATTGATGCTAGCTATTTTATTGTTTATATGGTCTCCAGAGACAACG
TTTCATGCTCCATCTGTCTATAA
```

图 14B (SEQ ID NO :48)

PDISP/HA B Massachussetts (PrL-)的氨基酸序列

```
MAKNVAIFGLLFSLVLVPSQIFADRICTGITSSNSPHVVKATQGEVNVTVGIPLTTPTKSYFANLKGTKTRGKLCPCDCLNCTDLDVALG
RPMCVGTTPSAKASILHEVRPVTSGCFPIIMHDRTKIRQLANLLRGYENIRLSTQNVIDAEKAPGGPYRLGTSGSCP NATSKSGFFATMA
WAVPKDNNKNATNPLTVEVPYICAEGEDQITVWGFHSDNKTQMKNLYGDSNPQKFTSSANGVTTHYVSQIGGFPDQTEDGGLPQSG
RIVVDYMMQKPGKGTIVYQRGVLLPQKVWCASGRSKVIKGSPLIGEADCLHEKYGLNKSHPYYTGEHAKAIGNCPIWVKTPKLKLAN
GTKYRPPGGGWEGMIAGWHGYTSHGAHGVAVAADLKSTQEAINKITKNLSLSELEVKNLQRLSGAMDELHNEILELDEKVDLDRAD
TISSQIELAVLLSNEGIINSEDEHLLALERLKKMLGPSAVDIGNGCFETKHKCNQCLDRIAAGTFNAGEFSLPTFDSL NITAASLNDDGLD
NHTILLYSTAASSLAVTLM LAIFIVYMVSRDNVSCICL*
```

图 14C 构建体编号2072 (2X35S/CPMV HT)的示意图

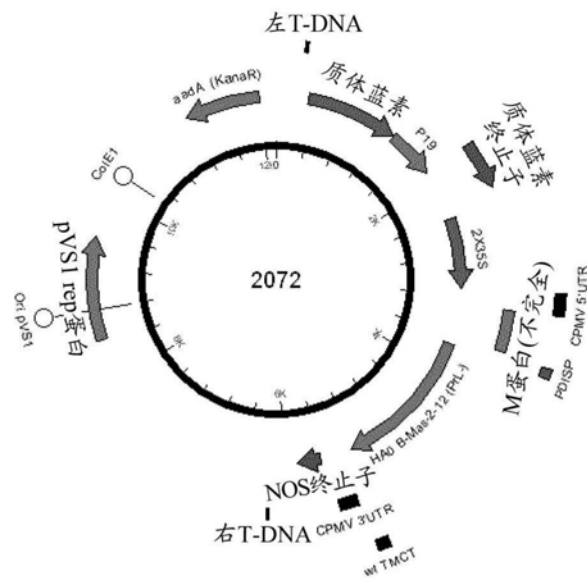


图 14D 构建体编号2052 (2X35S/CPMV HT+)的示意图

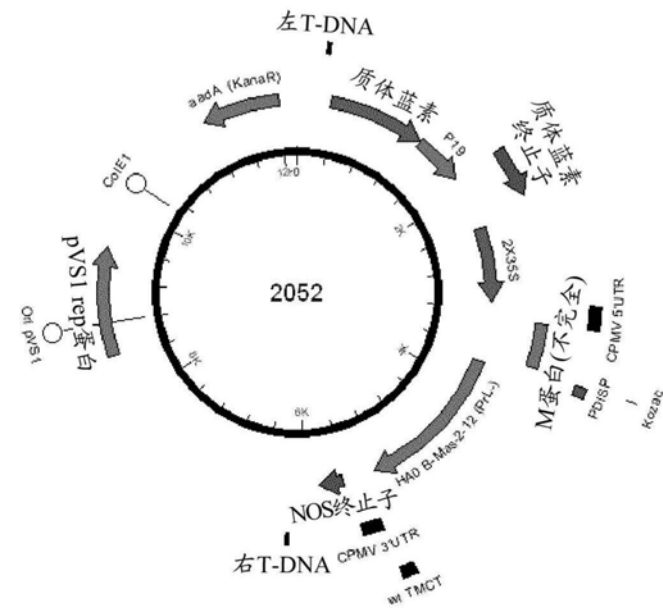


图14:PDISP/HA B Massachussetts (PrL-)的M-2X35S/CPMV HT (构建体编号2072) 和HT+ (构建体编号2052)

图 15A (SEQ ID NO : 49)

PDISP/HA B Massachussetts (PrL-)+H1 California TMCT的核苷酸序列

```
ATGGCGAAAAACGTTGCGATTTTCGGCTTATTGTTTTCTTCTTGTTGGTTCCTTCTCAGATCTTCGCCGATCGAATCTGCACTG
GGATAACATCTTCAAACCTCACCTCATGTGGTCAAAACAGCTACTCAAGGGGAGGTCAATGTGACTGGTGTGATACCACTAACAAC
AACACCAACAAAATCTTATTTTGCAAATCTCAAAGGAACAAAGACCAGAGGGAACTATGCCCAGAGTGTCTCAACTGTACAGATC
TGGATGTGGCCCTGGGCAGGCCAATGTGTGTGGGAACACACCTTCTGCGAAAGCTTCAATACTTCACGAAGTCAGACCTGTTAC
ATCCGGGTGCTTCCCTATAATGCACGACAGAACAAAAATCAGGCAACTAGCCAATCTTCTCAGAGGATATGAAAATATCAGGTTAT
CAACCCAAAACGTTATCGATGCAGAAAAGGCACCAGGAGGACCCTACAGACTTGGAACCTCAGGATCTTGCCCTAACGCTACCAG
TAAAAGCGGATTTTTCGCAACAATGGCTTGGGCTGTCCAAAGGACAACAACAAAATGCAACGAACCCATTAAACAGTAGAAGTA
CCATACATTTGTGCAGAAGGGGAAGACCAATTACTGTTTGGGGGTTCCATTGAGATAACAAAACCCAAATGAAGAACCTCTATG
GAGACTCAAATCCTCAAAAGTTACCTCATCTGCTAATGGAGTAACCACACATTATGTTTCTCAGATTGGCGGCTTCCAGATCAAA
CAGAAGACGGAGGACTACCACAAAGCGGCAGAAATTGTCGTTGATTACATGATGCAAAAACCTGGGAAAACAGGAACAATTGTCT
ATCAAAGAGGTGTTTTGTTGCCTCAAAGGTGTGGTGCGCGAGTGGCAGGAGCAAAGTAATAAAAGGGTCCTTGCCTTTAATTGG
TGAAGCAGATTGCCTTCATGAAAAATACGGTGGATTAACAAAAGCAAGCCTTACTACACAGGAGAACATGCAAAAGCCATAGG
AAATTGCCCAATATGGGTGAAAACACCTTTGAAGCTTGCCATGGAACCAAATATAGACCTCCTGGTGGAGGATGGGAAGGAAT
GATTGCAAGTTGGCAGCGATACACATCTCACGGAGCACATGGAGTGGCAGTTGCTGCAGACCTTAAGAGCACACAAGAAGCTAT
AAACAAGATAACAAAAATCTCAACTCTTTGAGTGAGCTAGAAGTAAAGAATCTTCAAAGGCTAAGTGGTGCCATGGATGAACTC
CACACGAAATACTCGAGCTGGATGAGAAAGTGGATGACCTCAGAGCTGACACTATAAGTTCACAAATAGAAGTTGCAGTCTTGC
TTTCCAACGAAGGAATAATAACAGTGAAGACGAGCATCTATTGGCACTTGAGAGAAAACCTAAAGAAAATGCTGGGTCCTCTGC
TGTAGACATAGGAAATGGATGCTTCGAAACCAACACAAATGCAACCAGACCTGCTTAGACAGGATAGCTGCTGGCACCTTTAAT
GCAGGAGAGTTTTCTCTCCCACTTTTGATTGATTGAACATTACTGCTGCATCTTTAAATGATGATGGATTGGATAACTACCAGATT
TTGGCGATCTATTCAACTGTCGCCAGTTCATTGGTACTGGTAGTCTCCCTGGGGGCAATCAGTTTCTGGATGTGCTAATGGGTCT
CTACAGTGTAGAATATGTATTAA
```

图 15B (SEQ ID NO :50)

PDISP/HA B Massachussetts (PrL-)+H1 California TMCT的氨基酸序列

```
MAKNVAIFGLLFSLLVLPVPSQIFADRICTGITSSNSPHVVKATQGEVNVTVGVIPLTTTPTKSYFANLKGTKTRGKLCPCDCLNCTDLVALG
RPMCVGTTPSAKASILHEVRPVTS GCFPIMHDR TKIRQLANLLRGYENIRLSTQNVIDAEKAPGGPYRLGTS GSCPNATSKSGFFATMA
WAVPKDNNKNATNPLTVEVPYICAEGEDQITVWGFHSDNKTQMKNLYGDSNPQKFTSSANGVTTHYVSQIGGFPDQTEDGGLPQSG
RIVVDYMMQKPGKTGTIVYQRGVLLPQKVWCASGRSKVIKGSPLIGEADCLHEKYGGLNKSKPYTGEHAKAIGNCPIWVKTPCLKLAN
GTKYRPPGGGWEGMIAGWHGYTSHGAHGVAVAADLKSTQEAINKITKNLNSLSELEVKNLQRLSGAMDELHNEILEDEKVDDLRAD
TISSQIELAVLLSNEGIINSEDEHLLALERKLLKMLGPSAVDIGNGCFETKHKCNQTCLDRIAAGTFNAGEFSLPTFDSL NITAASLNDDGLD
NYQILAIYSTVASSLVVSLGAISFWMCNGLQCRICI*
```

图 15C 构建体编号2074 (2X35S/CPMV HT)的示意图

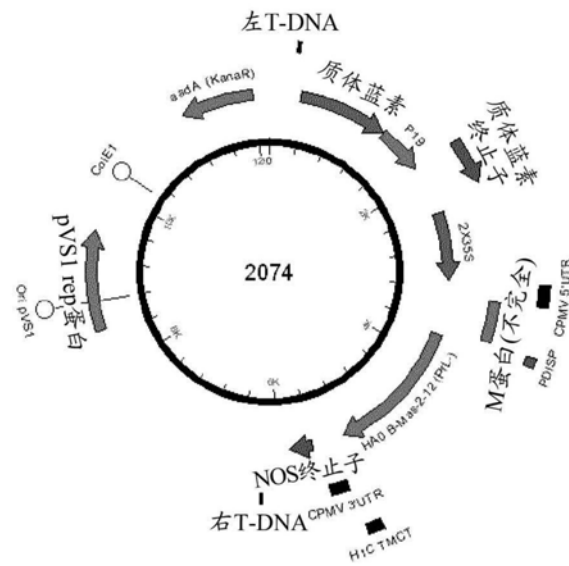


图 15D 构建体编号2062 (2X35S/CPMV HT+)的示意图

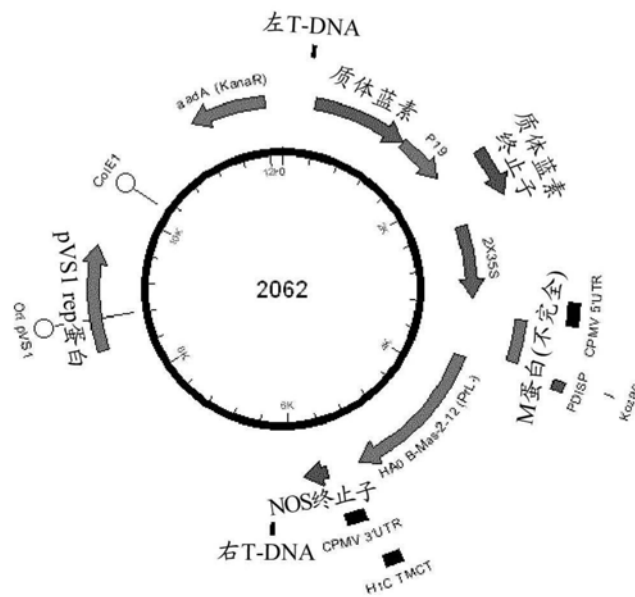


图15:PDISP/HA B Massachusetts (PrL-) +H1 California TMCT的N-2X35S/CPMV HT (构建体编号2074) 和HT+ (构建体编号2062)

图 16A (SEQ ID NO :51)
HA B Wisconsin (PrL-)的核苷酸序列

```
ATGAAGCAATAATTGTACTACTCATGGTAGTAACATCCAATGCAGATCGAATCTGCACTGGGATAACATCTTCAAACCTCACCTCA
TGTGGTCAAAACAGCTACTCAAGGGGAGGTCAATGTGACTGGCGTGATACCACTGACAACAACCAACAAAATCTTATTTTGCA
AATCTCAAAGGAACAAGGACCAGAGGGAACTATGCCCGGACTGTCTCAACTGTACAGATCTGGATGTGGCCTTGGGCAGGCCA
ATGTGTGTGGGGACCACACCTTCTGCTAAAGCTTCAATACTCCACGAGGTCAGACCTGTTACATCCGGGTGCTTCTCTATAATGCA
CGACAGAACAAAATCAGGCAACTACCAATCTTCTCAGAGGATATGAAAATATCAGGTTATCAACCCAAAACGTTATCGATGCA
GAAAAAGCACCAGGAGGACCCTACAGACTTGGAACCTCAGGATCTTGCCCTAACGCTACCAGTAAATCGGATTTTTTGCAACAA
TGGCTTGGGCTGTCCAAAGGACAACACAAAATGCAACGAACCCACTAACAGTAGAAGTACCATACATTTGTACAGAAGGGGA
AGACCAAATTACTGTTTGGGGTTCATTTCAGATAACAAAACCAATGAAGAGCCTCTATGGAGACTCAATCCTCAAAAGTTCA
CCTCATCTGCTAATGGAGTAACCACACATTATGTTTCTCAGATTGGCGACTTCCAGATCAACAGAGAGCGGAGGACTACCACAA
AGCGGCAGAAATTGTTGTTGATTACATGATGCAAAAACCTGGGAAAACAGGAACAATTGTCTATCAAAGAGGTGTTTTGTTGCCTC
AAAAGGTGTGGTGCGCGAGTGGCAGGAGCAAAGTAATAAAAGGGTCATTGCCTTTAATTGGTGAAGCAGATTGCCTTCATGAAA
AATACGGTGGATTAACAAAAGCAAGCCTTACTACACAGGAGAACATGCAAAAGCCATAGGAAATTGCCCAATATGGGTAAAAA
CACCTTTGAAGCTTGCCAATGGAACCAATATAGACCTCTGTTGGAGGATGGGAAGGAATGATTGCAGGTTGGCACGGATACA
CATCTCACGGAGCATGGAGTGGCAGTGGCGGCAGACCTTAAGAGTACACAAGAAGGTATAAATAAGATAACAAAAAATCTCA
ATTCTTTGAGTGAGCTAGAAGTAAAGAACCTTCAAAGACTAAGTGGTGCCATGGATGAACTCCACAACGAAATACTCGAGCTGGA
TGAGAAAGTGGATGATCTCAGAGCTGACACTATAAGCTCACAAATAGAAGTTCAGTCTTGCTTCCAACGAAGGAATAATAAAC
AGTGAAGACGAGCATCTATTGGCACTTGAGAGAAAACCTAAAGAAAATGCTGGGTCCCTCTGCTGTAGACATAGGAAACGGATGC
TTCGAAACCAACACAAATGCAACCAGACCTGCTTAGACAGGATAGCTGCTGGCACCTTAATGCAGGAGAAATTTCTCTCCCAAC
TTTTGATTCATTGAACATTACTGCTGCATCTTTAAATGATGATGGATTGGATAACCATACTATACTGCTCTATTACTCAACTGCTGCT
TCTAGTTTGGCTGAACATTAATGCTAGCTATTTTATTGTTTATATGGTCTCCAGAGACAACGTTTCATGCTCCATCTGCTATAA
```

图 16B (SEQ ID NO :52)
HA B Wisconsin (PrL-)的氨基酸序列

```
MKAIIVLLMVVTSNADRICTGITSSNSPHVVKATQGEVNVTVIPLTTPTKSYFANLKGTRTRGKLCPCDCLNCTDLDVALGRPMC VGT
TPSAKASILHEVRPVTS GCFPIIMH DRTKIRQLPNLLRGYENIRLSTQNVDAEKAPGGPYRLGTSGSCP NATSKIGFFATMAWAVPKDNY
KNATNPLTVEVPYICTEGEDQITVWGFHSDNKTQMKSLYGDSNPQKFTSSANGVTTHYVSQIGDFPDQTEDGGLPQSGRIVVDYMM
QKPGKTGTIVYQRGVLLPQKVWCASGRSKVIKSLPLIGEADCLHEKYGGLNKSPPYTGEHAKAIGNCPIWVKTPKLANGTKYRPPGG
GWEGMIAGWHGYTSHGAHGVAVAADLKSTQEAINKITKNLSLSELEVKNLQRLSGAMDELHNEILELDEKVDDL RADTISSQIELAVL
LSNEGIINSEDEHLLALERKLLKMLGPSAVDIGNGCFETKHKCNQTCLDRIAAGTFNAGEFSLPTFDSL NITAASLND DGLDNHTILLYST
AASSLAVTLM LAIFIVYMVSRDNVSCICL*
```

图 16C 构建体编号1445 (2X35S/CPMV HT)的示意图

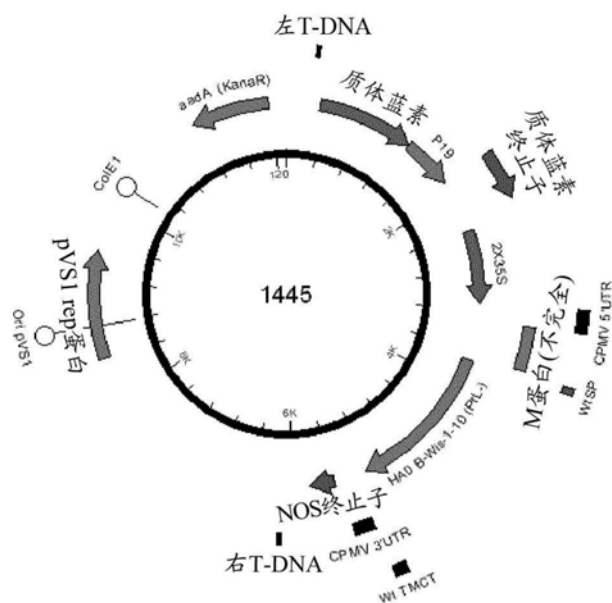


图 16D 构建体编号1839 (2X35S/CPMV HT+)的示意图

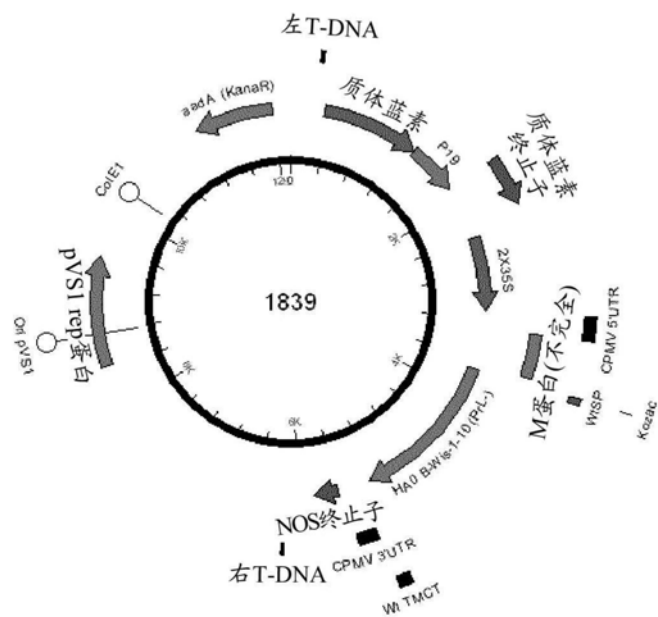


图16:HA B Wisconsin (PrL-) 的0-2X35S/CPMV HT (构建体编号1445)、HT+ (构建体编号1839)

图 17A (SEQ ID NO :53)

HA B Wisconsin (PrL-)+H1 California TMCT的核苷酸序列

```
ATGAAGGCAATAATTGTACTACTCATGGTAGTAACATCCAATGCAGATCGAATCTGCACTGGGATAACATCTTCAAACCTCACCTCA
TGTGGTCAAAACAGCTACTCAAGGGGAGGTCAATGTGACTGGCGTGATACCACTGACAACAACCAACAAAATCTTATTTTGCA
AATCTCAAAGGAACAAGGACCAGAGGGAACTATGCCCGGACTGTCTCAACTGTACAGATCTGGATGTGGCCTTGGGCAGGCCA
ATGTGTGTGGGGACCACACCTTCTGCTAAAGCTTCAATACTCCACGAGGTGAGACCTGTACATCCGGGTGCTTTCCTATAATGCA
CGACAGAACAAAATCAGGCAACTACCCAATCTTCTCAGAGGATATGAAAATATCAGGTTATCAACCCAAAACGTTATCGATGCA
GAAAAGCACCAGGAGGACCCTACAGACTTGGAACCTCAGGATCTTGCCCTAACGCTACCAGTAAAATCGGATTTTTTGCAACAA
TGGCTTGGGCTGTCCCAAAGGACAACACAAAATGCAACGAACCCACTAACAGTAGAAGTACCATACATTTGTACAGAAGGGGA
AGACCAAATTACTGTTTGGGGGTTCCATTAGATAACAAAACCCAAATGAAGAGCCTCTATGGAGACTCAAACTCTCAAAAGTTCA
CCTCATCTGCTAATGGAGTAACACACATTATGTTTCTCAGATTGGCGACTTCCAGATCAAACAGAAGACGGAGGACTACCACAA
AGCGGCAGAAATTGTTGTTGATTACATGATGCAAAAACCTGGGAAAACAGGAACAATTGTCTATCAAAGAGGTGTTTGTTCCTC
AAAAGGTGTGGTGCAGAGTGGCAGGAGCAAAGTAATAAAAGGGTCATTGCCTTTAATTGGTGAAGCAGATTGCCTTCATGAAA
AATACGGTGGATTAAACAAAAGCAAGCCTTACTACACAGGAGAACATGCAAAAGCCATAGGAAATTGCCCAATATGGGTAAAAA
CACCTTTGAAGCTTGCCAATGGAACCAATATAGACCTCTGGTGGAGGATGGGAAGGAATGATTGCAGGTTGGCACGGATACA
CATCTCACGGAGCACATGGAGTGGCAGTGGCGGCAGACCTTAAGAGTACACAAGAAGCTATAAATAAGATAACAAAAAATCTCA
ATTCTTTGAGTGAGCTAGAAGTAAAGAACCTTCAAAGACTAAGTGGTGCCATGGATGAACTCCACAACGAAATACTCGAGCTGGA
TGAGAAAGTGGATGATCTCAGAGCTGACACTATAAGCTCACAAATAGAAGTTCAGTCTTGCTTCCAACGAAGGAATAATAAAC
AGTGAAGACGAGCATCTATTGGCACTTGAGAGAAAATAAGAAAATGCTGGGTCCCTCTGCTGTAGACATAGGAAACGGATGC
TTCGAAACCAACACAAATGCAACCAGACCTGCTTAGACAGGATAGCTGCTGGCACCTTTAATGCAGGAGAAATTTCTCTCCCCAC
TTTTGATTATTGAACATTACTGCTCATCTTAAATGATGATGGATTGGATACTACCAGATTTTGGCGATCTATTCAACTGTCGC
CAGTTCATTGGTACTGGTAGTCTCCCTGGGGCAATCAGTTTCTGGATGTGCTCTAATGGGTCTCTACAGTGTAGAATATGTATT
AA
```

图 17B (SEQ ID NO : 54)

HA B Wisconsin (PrL-)+H1 California TMCT的氨基酸序列

```
MKAIIVLLMVVTSNADRICTGITSNSPHVVKATQGEVNVTVGIPLTTPTKSYFANLKGTRTRGKLCPDCLNCTDLDVALGRPMC VGT
TPSAKASILHEVRPVTS GCFPIMHDRTKIRQLPNLLRGYENIRLSTQNVIDAEKAPGGPYRLGTS GSCP NATSKIGFFATMAWAVPKDNY
KNATNPLTVEVPYICTEGEDQITVWGFHSDNKTQMKSLYGDSNPQKFTSSANGVTTHYVSQIGDFPDQTEDGGLPQSGRIVVDYMM
QKPGKTGTIVYQRGVLLPQKVWCASGRSKVIKGSPLIGEADCLHEKYGGLNKSPPYTG EHA KAIGNCPIWVKTPKLANGTKYRPPGG
GWEGMIAGWHGYTSHGAHGVAVAADLKSTQEAINKITKNLNSLELVKNLQRLSGAMDELHNEILELDEKVDDL RADTISSQIELAVL
LSNEGIINSEDEHLLALERK LKMLGPSAVDIGNGCFETKHKNQTC LDRIAAGTFNAGEFSLPTFDSL NITAASLNDDGLDNYQLAIYST
VASSLVLVVSLGAISFWMCSNGLQCRICI*
```

图 17C 构建体编号1454 (2X35S/CPMV HT)的示意图

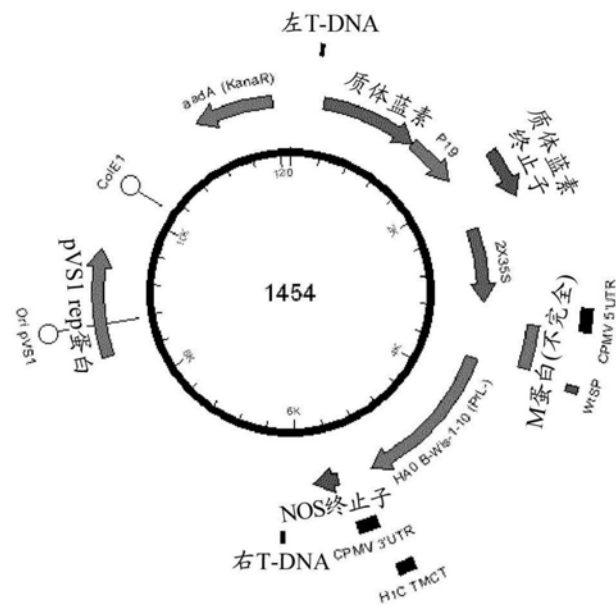


图 17D 构建体编号1860 (2X35S/CPMV HT+)的示意图

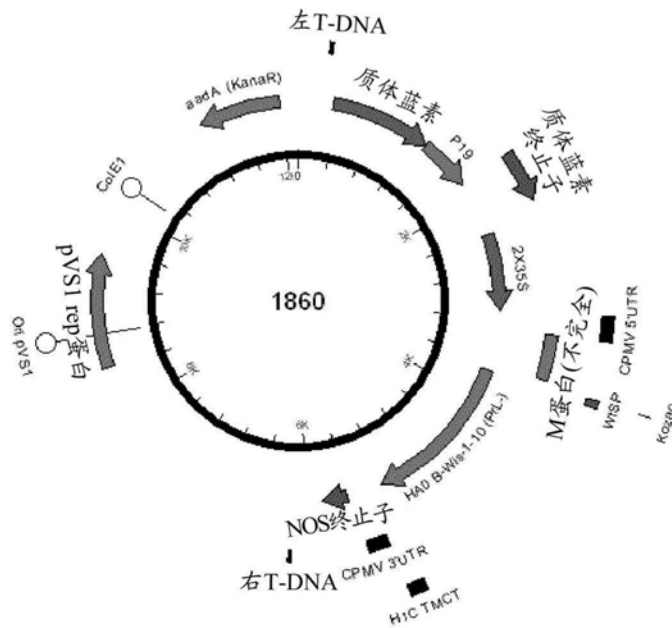


图17:HA B Wisconsin(PrL-) +H1 California TMCT的P-2X35S/CPMV HT (构建体编号1454) 和HT+ (构建体编号1860)