



(10) 授权公告号 CN 111566216 B

(45) 授权公告日 2024. 07. 02

(21) 申请号 201880068443.0	(74) 专利代理机构 北京集佳知识产权代理有限公司 11227
(22) 申请日 2018.08.22	专利代理师 郑斌 刘振佳
(65) 同一申请的已公布的文献号 申请公布号 CN 111566216 A	(51) Int.Cl. C12N 15/63 (2006.01)
(43) 申请公布日 2020.08.21	(56) 对比文件 JP 2004081143 A, 2004.03.18 US 2007111258 A1, 2007.05.17 W0 2021195491 A2, 2021.09.30 Peschek J等.A conformational RNA zipper promotes intron ejection during non-conventional XBP1 mRNA splicing.EMBO Rep.2015,第16卷(第16期),摘要,第1689页左栏第2段,图1,图EV1 A.
(30) 优先权数据 1713545.0 2017.08.23 GB	审查员 李怡澜
(85) PCT国际申请进入国家阶段日 2020.04.20	权利要求书4页 说明书44页 序列表31页 附图10页
(86) PCT国际申请的申请数据 PCT/GB2018/052387 2018.08.22	
(87) PCT国际申请的公布数据 W02019/038544 EN 2019.02.28	
(73) 专利权人 阿斯比奥英国有限公司 地址 英国爱丁堡	
(72) 发明人 格拉哈姆·怀特西德	

(54) 发明名称
使用可调节内含子的表达控制

(57) 摘要
本发明涉及能够在真核细胞中调节基因表达并且对未折叠蛋白反应 (UPR) 具有响应性的调节性核酸序列的用途。公开了能够调节基因表达的可调节内含子和UPR诱导型启动子。还公开了包含这样的调节性核酸序列的重组表达构建体，由此可通过在包含所述构建体的真核细胞中引起未折叠蛋白反应 (UPR) 来诱导所编码的表达产物的表达；使用这样的构建体和相关载体、细胞等的方法。

1. 用于在细胞中产生表达产物的合成核酸表达构建体,所述合成核酸表达构建体包含与编码所述表达产物的核酸序列有效连接的启动子序列,编码所述表达产物的所述核酸序列包含编码可调节内含子的序列,所述可调节内含子是包含可切除序列的内含子,所述可切除序列能够通过细胞中的未折叠蛋白反应 (UPR) 体系从所述合成核酸表达构建体产生的转录物中剪接出去/剪接掉,从而产生编码功能性表达产物的转录物,

其中所述表达产物不是XBP1蛋白、Hac1蛋白或bZIP60蛋白,并且

其中所述可调节内含子包含序列CNG/CNG-Xn-CNG/CNG,其中Xn代表长度为n个碱基的序列,其中/代表切割位点,并且其中序列CNG-Xn-CNG从所述转录物中切除;或

其中所述可调节内含子包含序列CNG/CNG-Xn-CNG/CNG[CG],其中Xn代表长度为n个核苷酸的序列,其中/代表切割位点,使得在剪接时可切除序列CNG-Xn-CNG从所述转录物中切除,并且其中序列Xn的5'端的核苷酸为C或G。

2. 权利要求1所述的合成核酸表达构建体,其中所述可调节内含子能够被IRE1蛋白或者其同源物或直系同源物剪接掉。

3. 权利要求1或2所述的合成核酸表达构建体,其中剪接掉所述可调节内含子的所述可切除序列消除所述转录物中的提前终止密码子。

4. 权利要求1或2所述的合成核酸表达构建体,其中剪接掉所述可调节内含子的所述可切除序列导致所述转录物中位于所述可调节内含子下游的序列发生阅读框移位。

5. 权利要求1或2所述的合成核酸表达构建体,其中所述可切除序列的核苷酸长度不可被3整除。

6. 权利要求1所述的合成核酸表达构建体,其中Xn的长度为10至500个核苷酸。

7. 权利要求6所述的合成核酸表达构建体,其中Xn的长度为15至350个核苷酸。

8. 权利要求6所述的合成核酸表达构建体,其中Xn的长度为15至100个核苷酸。

9. 权利要求6所述的合成核酸表达构建体,其中Xn的长度为15至35个核苷酸。

10. 权利要求6所述的合成核酸表达构建体,其中Xn的长度为20至25个核苷酸。

11. 权利要求6至10中任一项所述的合成核酸表达构建体,其中Xn由序列CACUCAGACUACGUGCACCU (SEQ ID NO:1) 组成。

12. 权利要求11所述的合成核酸表达构建体,其中Xn包含以下序列之一或由其组成:

-CACUCAGACUACGUGCACCU (SEQ ID NO:1);

-CACUCAGACUACGUGCUCU (SEQ ID NO:2);

-CACUCAGACUACGUGCCCCU (SEQ ID NO:3);

-CACUCAGACUACGUGCGCCU (SEQ ID NO:4); 和

-CACUCAGACUAUGUGCACCU (SEQ ID NO:5)。

13. 权利要求1或2所述的合成核酸表达构建体,其中所述可调节内含子由序列CNG/CNGCACUCAGACUACGUGCACCUCNG/CNGC (SEQ ID NO:6) 组成。

14. 权利要求1或2所述的合成核酸表达构建体,其中所述可调节内含子由序列CAG/CAGCACUCAGACUACGUGCACCUCUG/CUGC (SEQ ID NO:7) 组成。

15. 权利要求1或2所述的合成核酸表达构建体,其中所述可调节内含子包含以下序列之一或由其组成:

-CNG/CAGCACUCAGACUACGUGCACCUCUG/CNG (SEQ ID NO:8);

- CNG/CAGCACUCAGACUACGUGCUCCUCUG/CNG (SEQ ID NO:9);
- CNG/CAGCACUCAGACUACGUGCCCCUCUG/CNG (SEQ ID NO:10);
- CNG/CAGCACUCAGACUACGUGCGCCUCUG/CNG (SEQ ID NO:11);和
- CNG/CAGCACUCAGACUAUGUGCACCUCUG/CNG (SEQ ID NO:12)。

16. 权利要求1或2所述的合成核酸表达构建体,其中所述可调节内含子包含以下序列之一或由其组成:

- CAG/CAGCACUCAGACUACGUGCACCUCUG/CUGC (SEQ ID NO:7);
- CAG/CAGCACUCAGACUACGUGCUCCUCUG/CUGC (SEQ ID NO:13);
- CAG/CAGCACUCAGACUACGUGCCCCUCUG/CUGC (SEQ ID NO:14);
- CAG/CAGCACUCAGACUACGUGCGCCUCUG/CUGC (SEQ ID NO:15);和
- CAG/CAGCACUCAGACUAUGUGCACCUCUG/CUGC (SEQ ID NO:16)。

17. 权利要求1或2所述的合成核酸表达构建体,其中所述可调节内含子包含序列CAG/CUGCAGCACUCAGACUACGUGCACCUCUG/CUG (SEQ ID NO:17) 或 CAG/CUGCAGCACUCAGACUACGUGCACCUCUG/CUGG (SEQ ID NO:27),其中/代表切割位点。

18. 权利要求1或2所述的合成核酸表达构建体,其包含与编码所述表达产物的核酸序列有效连接的诱导型启动子。

19. 权利要求18所述的合成核酸表达构建体,其中所述诱导型启动子是未折叠蛋白反应 (UPR) 诱导型启动子。

20. 权利要求1或2所述的合成核酸表达构建体,其中所述编码表达产物的核酸序列是转基因。

21. 权利要求1或2所述的合成核酸表达构建体,其中所述编码表达产物的核酸序列编码蛋白质。

22. 核酸,其包含编码表达产物的序列,所述编码表达产物的序列包含编码可调节内含子的序列,所述可调节内含子是包含可切除序列的内含子,所述可切除序列能够通过细胞中的未折叠蛋白反应 (UPR) 体系从合成表达构建体中产生的转录物中剪接掉,从而产生编码功能性表达产物的转录物,并且其中所述表达产物不是XBP1蛋白、Hac1蛋白、bZIP60蛋白,或其同源物,

其中所述可调节内含子包含序列CNG/CNG-Xn-CNG/CNG,其中Xn代表长度为n个碱基的序列,其中/代表切割位点,并且其中序列CNG-Xn-CNG从所述转录物中切除;或

其中所述可调节内含子包含序列CNG/CNG-Xn-CNG/CNG[CG],其中Xn代表长度为n个核苷酸的序列,其中/代表切割位点,使得在剪接时可切除序列CNG-Xn-CNG从所述转录物中切除,并且其中序列Xn的5'端的核苷酸为C或G。

23. 载体,其包含根据权利要求1至21中任一项所述的合成核酸表达构建体或根据权利要求22所述的核酸。

24. 药物组合物,其包含根据权利要求1至21中任一项所述的合成核酸表达构建体、根据权利要求22所述的核酸或根据权利要求23所述的载体。

25. 根据权利要求1至21中任一项所述的核酸表达构建体、根据权利要求22所述的核酸或根据权利要求23所述的载体用于制备药物组合物的用途。

26. 细胞,其包含根据权利要求1至21中任一项所述的核酸表达构建体、根据权利要求

22所述的核酸或根据权利要求23所述的载体。

27.根据权利要求26所述的细胞,其是真核细胞。

28.根据权利要求26或27所述的细胞,其中根据权利要求1至21中任一项所述的核酸表达构建体、根据权利要求22所述的核酸或根据权利要求23所述的载体编码对所述细胞具有毒性的表达产物。

29.用于产生表达产物的方法,所述方法包括:

a) 提供包含根据根据权利要求1至21中任一项所述的合成核酸表达构建体的真核细胞群;

b) 处理所述细胞群,以诱导未折叠蛋白反应,从而从所述可调节内含子中剪接掉所述可切除序列;

c) 在合适的条件下孵育所述细胞群以产生所述表达产物;以及

d) 从所述细胞群中分离所述表达产物。

30.根据权利要求29所述的方法,其包括在步骤b) 处理所述细胞群以诱导未折叠蛋白反应 (UPR) 之前,在适合于所述细胞生长的条件下孵育所述细胞群。

31.根据权利要求29所述的方法,其中步骤b) 包括向所述细胞群施加应激,所述应激适于诱导UPR。

32.根据权利要求29至31中任一项所述的方法,其中步骤b) 包括施用能够在所述细胞群中诱导UPR的化学剂。

33.根据权利要求32所述的方法,其中所述能够在所述细胞群中诱导UPR的化学剂包含毛喉素、二硫苏糖醇 (DTT)、衣霉素、毒胡萝卜素、饱和脂肪酸、能够下调硬脂酰辅酶A脱饱和酶活性的试剂中的一种或更多种。

34.根据权利要求29至31中任一项所述的方法,其中步骤b) 包括在所述真核细胞群中表达诱导物蛋白,从而在所述细胞群中诱导UPR响应。

35.根据权利要求33所述的方法,其中步骤b) 包括用能够在所述细胞群中表达诱导物蛋白的表达载体转染所述细胞群,或其中:

步骤b) 包括从先前引入到所述细胞群内的表达载体中诱导诱导物蛋白的表达。

36.根据权利要求35所述的方法,其中所述诱导物蛋白是异源蛋白。

37.根据权利要求29至31中任一项所述的方法,其中步骤b) 包括使所述细胞群暴露于低氧或碳水化合物剥夺。

38.根据权利要求29至31中任一项所述的方法,其中所述细胞群包含真菌细胞、动物细胞或植物细胞。

39.根据权利要求38所述的方法,其中所述细胞群包含哺乳动物细胞。

40.根据权利要求29至31中任一项所述的方法,其包括在步骤a) 之前将所述核酸表达构建体引入到所述细胞群内的步骤。

41.根据权利要求1至21中任一项所述的核酸表达构建体、根据权利要求22所述的核酸、根据权利要求23所述的载体、根据权利要求24所述的药物组合物或根据权利要求26至28所述的细胞,其用于制备药物。

42.根据权利要求1至21中任一项所述的核酸表达构建体、根据权利要求22所述的核酸、根据权利要求23所述的载体、根据权利要求24所述的药物组合物用于制备用于基因治

疗的药物的用途。

使用可调节内含子的表达控制

[0001] 本发明涉及能够在真核细胞中调节基因表达并且对未折叠蛋白反应(unfolded protein response,UPR)具有响应性的调节性核酸序列的用途。在一些方面中,本发明涉及可调节内含子用于控制基因表达的用途。在另一些方面中,其涉及UPR诱导型启动子。本发明还涉及包含这样的调节性核酸序列的重组表达构建体,以及涉及使用这种构建体和相关的载体、细胞等的方法,由此可通过在包含该构建体的真核细胞中引起未折叠蛋白反应(UPR)来诱导编码的表达产物的表达。

背景技术

[0002] 在许多情况下,可调节的基因表达是理想的,在这种情况下,控制表达产物的表达水平是有益的或必要的。例如,在工业生物技术的情况下,能够在发酵过程中的期望时间诱导表达产物(例如蛋白质)的产生是非常有利的。在另一个实例中,在基因治疗中,能够在治疗的期望时间和/或位置的位置诱导治疗性产物(例如治疗性蛋白质)的表达可以是理想的。

[0003] 诱导型启动子是本领域已知的,例如四环素(Tet)-on和-off诱导型表达体系(Gossen M & Bujard H.PNAS.1992 Jun 15;89(12):5547-51;Gossen M,Freundlieb S,Bender G,Müller G,Hillen W,& Bujard H.Science.1995 Jun 23;268(5218))。

[0004] 诱导型启动子通过在转录水平上调节表达,控制由相关表达体系产生的mRNA的量来发挥作用。虽然这可提供有用的基因表达水平,但仍需要用于控制表达的其他理想地改进的体系。

[0005] 特别地,需要用于毒性蛋白(即对产生它们的细胞具有毒性的蛋白质)的表达的改进的表达控制体系。就毒性蛋白而言,即使少量蛋白质的表达也常常会导致细胞死亡或非常差的生成物(production)。

[0006] 此外,在细胞治疗的基因的情况下,需要允许在期望的时间和/或位置诱导治疗性蛋白质或RNA或其他表达产物的表达的体系。

[0007] 本发明涉及在翻译时发挥作用的表达调节。特别地,其涉及由于真核细胞中的未折叠蛋白反应(UPR)而被剪接出的内含子用于控制表达的用途。

[0008] 未折叠蛋白反应(UPR)是内质网应激的细胞应对机制。UPR响应于内质网(endoplasmic reticulum,ER)腔中的未折叠或错误折叠的蛋白质积累而激活。UPR旨在通过多种机制恢复细胞的正常功能。如果在一定时间范围内未达到这些目标或延长了中断时间,则UPR的目标是凋亡。可通过提高蛋白质的合成和折叠(例如异源蛋白的产生)或通过其他细胞应激(例如,化学诱导的应激,例如阻断糖基化途径或二硫键形成)来触发UPR。UPR在所有真核生物中都高度保守。

[0009] 在哺乳动物细胞中,存在三种机制来感知ER应激和激活UPR:

[0010] 1) XBP-1 mRNA的IRE-1剪接

[0011] 通过ER跨膜蛋白IRE1经由分子伴侣BiP的解离检测由蛋白质折叠需求提高或通过ER过程的化学抑制引起的ER应激。BiP的这种解离通过允许蛋白质的低聚和磷酸化激活

IRE1,从而导致面向RNA酶结构域的ER腔暴露出来。随后,RNA酶结构域以不依赖于剪接体的方式催化从XBP-1 mRNA去除非规范内含子(non-canonical intron)。在非应激条件下,XBP1 mRNA是未剪接的(XBP1u),其在翻译时会形成261个氨基酸的ORF,这是一种非功能性蛋白质。然而,当检测到ER应激时,XBP1u被IRE1 RNA酶剪接形成XBP1s,其编码功能性的376个氨基酸的蛋白质。该功能性蛋白质XBP1是一种转录因子,其控制参与蛋白质稳态的数个基因的表达,例如分子伴侣、二硫键异构酶和参与磷脂生物合成的酶。其通过与特定序列、ER应激反应元件(ER stress response element,ERSE)或未折叠蛋白反应元件(unfolded protein response element,UPRE)结合并增强基因表达来调节这些过程。因此,基因表达的控制是通过IRE1从XBP1u去除mRNA内含子。非常相似的体系在非哺乳动物的真核细胞中运行。

[0012] 2)ATF6

[0013] 激活的转录因子(activated transcription factor,ATF6)ATF6是一种2型跨膜蛋白,在该蛋白的胞质区具有转录因子结构域。它被合成为无活性的前体,并通过与BiP/GRP78缔合而保留在ER中。响应于应激条件,ATF6解离并运输到高尔基体,在那里发生加工以释放转录因子结构域。然后,该结构域被转运到细胞核,在那里其可与ERSE和UPRE结合,以增强参与蛋白质稳态的基因的表达。

[0014] 3)双链RNA激活的蛋白激酶(PERK)

[0015] PERK是一种1型ER跨膜蛋白,其包含ER腔应激传感元件和胞质蛋白激酶结构域。PERK响应于内质网应激而被激活,并通过经由磷酸化使真核起始因子(eIF2a)失活,从而抑制了哺乳动物细胞的ER中的正常蛋白质翻译。

[0016] 机制1)和2)与本发明最相关。

发明内容

[0017] 根据本发明,提供了用于在细胞中产生表达产物的合成核酸表达构建体,该核酸表达构建体包含与编码表达产物的核酸序列有效连接的启动子序列,所述编码表达产物的核酸序列包含编码可调节内含子的序列,所述可调节内含子是包含可切除序列的内含子,所述可切除序列能够通过细胞中的未折叠蛋白反应(UPR)体系从合成表达构建体产生的转录物中剪接出,从而产生编码功能性表达产物的转录物。

[0018] 因此,本发明基于可调节内含子的用途,该内含子能够通过UPR机制从编码表达产物的核酸序列的RNA转录物中剪接出,以调节表达。UPR机制在真核生物中普遍存在,并且该机制和相关序列都非常保守。因此,可在所有真核细胞上实施本发明。可调节内含子优选能够被IRE1蛋白或者其同源物或直系同源物剪接出(IRE1的同源物或直系同源物存在于所有真核生物中,包括真菌、植物和哺乳动物)。

[0019] 启动子通常与编码表达产物的核酸序列异源。也就是说,未天然地发现启动子与编码表达产物的序列有效连接。例如,在天然存在的基因中通常不会发现启动子和编码表达产物的核酸序列在一起。在本发明的一些实施方案中,启动子是合成启动子,即不是天然存在的启动子。本领域中已知的适用于真核细胞中的组成型和非组成型启动子具有广泛的范围,作为非限制性实例,可提及CAG启动子、CMV启动子、SV40。

[0020] 在一些优选实施方案中,表达产物是蛋白质。在这样的实施方案中,由于可调节内

含子的存在,由编码表达产物的核酸序列产生的未剪接的转录物编码蛋白质的截短的或在其他情况下有缺陷的形式(version),但当该转录物通过细胞中的UPR机制处理时,内含子的可切除序列被剪接出,并可从转录物中产生功能性蛋白质。因此,从可调节内含子剪接出产生了编码功能性蛋白质表达产物的功能性mRNA。

[0021] 然而,在某些情况下,表达产物可以是除蛋白质以外的产物。合适地,表达产物可以是RNA分子,例如核酶、RNA适配体、siRNA、反义RNA或miRNA。在这样的实施方案中,产生非功能性形式的RNA分子作为未剪接的转录物,并且剪接出内含子的可切除部分使RNA分子转化为活性形式。

[0022] 在本发明的优选实施方案中,剪接出可调节内含子的可切除序列允许转录物从编码表达产物的核酸序列正确翻译,从而允许产生期望的表达产物(例如蛋白质)。在这样的实施方案中,内含子在来自编码表达产物的核酸序列的转录物中的存在导致无功能的蛋白质从转录物中翻译出。通常,这是由内含子插入编码的氨基酸的翻译的蛋白质中引起的,或更优选地,是由于在内含子的转录物3'中引入了终止密码子或移码(相对于编码功能性蛋白质的正常转录物)而导致的。由于许多原因,由未剪接的转录物编码的蛋白质可以是无功能的,例如:

[0023] -内含子可导致向内含子下游(即以3'方向)的移码。当内含子的可切除序列的长度不是三个核苷酸的倍数时,通常会导致这种情况。这样的移码通常导致终止密码子的引入,导致截短的蛋白。在其他情况下,它可简单地导致内含子下游的编码的氨基酸序列的完全改变。

[0024] -引入编码序列,这导致氨基酸存在于翻译的蛋白质中,这破坏了该蛋白质的功能。在这种情况下,内含子下游的氨基酸序列没有改变,但是由内含子编码的氨基酸序列以未剪接的形式存在于翻译的蛋白质中,这破坏了蛋白质的功能。

[0025] -在内含子序列中引入终止密码子。在这种情况下,内含子本身可包含终止密码子,这将导致翻译的提前终止和截短的蛋白的产生。

[0026] 在本发明的某些优选实施方案中,合成核酸表达构建体被配置为使得剪接出内含子的可切除序列消除了转录物中的提前终止密码子(premature stop codon)。在本文中,“提前”意指终止密码子,其出现在正常转录物(即编码功能性蛋白质(例如野生型mRNA)的转录物)的终止密码子的上游(即,沿5'方向)。

[0027] 优选地,可调节内含子被配置为使得剪接出内含子的可切除序列导致位于可调节内含子下游(即3')的转录物中的序列的阅读框的移位。

[0028] 在本发明的某些优选实施方案中,可调节内含子被配置为使得当发生剪接时,从转录物中切除核苷酸长度不是3的倍数的序列。换句话说,可切除序列的长度为n个核苷酸,其中n不能被3整除。

[0029] 在本发明的优选实施方案中,可调节内含子包含序列CNG/CNG-Xn-CNG/CNG,其中Xn代表长度为n个碱基的序列,其中/代表切割位点,并且其中序列CNG-Xn-CNG从转录物中切除。

[0030] 因此,换句话说,可调节内含子合适地包含中央序列(Xn),其侧翼为两个剪接位点靶序列,每个剪接位点靶序列具有序列CNG/CNG,其中/代表切割位点。

[0031] CNG/CNG是真核细胞中的UPR体系以高度保守的方式靶向的共有剪接位点序列。如

本领域已知的,当诱导UPR响应时,该剪接位点共有序列被IRE1蛋白(其同源物或直系同源物在所有真核生物包括真菌、植物和哺乳动物中均存在)靶向。在自然界中,具有该共有剪接位点靶序列的内含子见于编码在UPR中激活的转录因子的mRNA中,所述转录因子例如XBP1蛋白(在后生动物中)、Hac1蛋白(在酵母中)和bZIP60蛋白(在植物中)。IRE1或者其同源物或直系同源物的内切核糖核酸酶活性发挥作用以在由/指示的位置切割RNA转录物,以去除可切除的内含子序列,然后将切割的RNA通过RNA连接酶蛋白(酿酒酵母(*S.cerevisiae*)中的RNA连接酶Rlg1p,哺乳动物细胞中的RNA连接酶RtcB)连接在一起。对哺乳动物、酵母和植物的细胞中的UPR体系进行了广泛的研究,机理也相对较好地进行了表征(参见,例如,Yoshida等,Cell,Vol.107,881-891,December 28,2001;Lu,等Mol Cell.2014 September 4;55(5):758-770;Samali等,International Journal of Cell Biology,Vol.2010,Article ID 830307,11页doi:10.1155/2010/830307;Chakraborty等,Appl Biol Chem DOI,10.1007/s13765-016-0167-6,Online ISSN 2468-0842,Print ISSN 2468-0834;以及Nagashima等,Scientific Reports 1,Article number:29(2011),DOI:10.1038/srep00029)。因此,这里将不深入讨论UPR的机制。然而,重要的是应当注意,鉴于IRE1介导的剪接的高度保守,来源于一个物种的内含子可被进化高度多样化的另一个物种成功地剪接。例如,Yoshida等(同上)解释了哺乳动物细胞可成功地从酵母Hac1 mRNA(对应于哺乳动物细胞中XBP1的Hac1)中剪接出内含子。

[0032] 应该注意的是,可调节内含子的长度可有很大的变化。例如,尽管已经观察到20和23个核苷酸的变体,但是哺乳动物和植物中的XBP1内含子的长度通常为26个核苷酸。酵母的Hac1内含子要长得多,通常长为约252个核苷酸(尽管如上所述,但在哺乳动物细胞中这个长得多的Hac1内含子仍然可被剪接出)。

[0033] 因此,在本发明的多种实施方案中,可调节内含子或Xn的长度可为10至500个核苷酸,更优选长度可为15至350个核苷酸,更优选长度可为15至100个核苷酸,更优选长度可为15至35个核苷酸,更优选长度可为20至25个核苷酸。调节内含子的可切除序列因此合适地长度为16至506个核苷酸,更优选长度为21至356个核苷酸,更优选长度为21至106个核苷酸,更优选长度为21至41个核苷酸,以及更优选长度为26至31个核苷酸。如上所述,在本发明的一些实施方案中,优选地,选择长度Xn,使得可切除序列的长度不能被3整除。

[0034] 关于Xn的特定序列有很大的自由度。优选的序列在下面列出,但是当然可使用许多其他变体,只要可调节内含子保持功能性,即,其通过UPR体系以合适的水平从转录物中剪接出。当然,某些可能的序列可以不是最佳的或者其干扰剪接过程,例如由于形成不期望的二级结构,但是本领域技术人员可容易地评估任何给定序列的作用,以确定其是否对剪接有不良影响。

[0035] 对可调节内含子的功能性(即在诱导UPR时可调节内含子被成功从转录物中剪接出的能力)的评估,本领域技术人员可使用广泛范围的方法容易地进行评估,并且这些可适用于旨在其中使用构建体的特定表达体系。作为一个优选的实例,可使用以下实施例中描述的方法,例如实施例1。例如,可将任何待评估的候选可调节内含子的功能性替换为实施例1中所述的构建体(称为SYNP-XBP-01),以代替实施例1中使用的示例性内含子,并且可通过评估通过2mM DTT诱导UPR前后EGFP表达的水平来测量在诱导UPR时所述内含子被成功剪接出的能力,这与实施例1中进行的完全相同。功能性可调节内含子是这样的内含子,其

能够在诱导UPR后成功地剪接出,以导致功能性EGFP的表达。优选地,功能性内含子使得在用2mM DTT诱导UPR之后24小时表达提高至少5倍,更优选EGFP的表达提高至少10倍,更优选表达提高至少100倍,并且更优选地,表达至少提高1000倍。优选的是,在诱导UPR之前,EGFP的表达水平是最低的,并且优选地可忽略不计。最低表达可定义为例如小于如实施例1中使用的对照构建体(即,没有可调节内含子的构建体,其中编码EGFP的序列的表达由CMV-mp驱动)的表达水平的50%,优选小于20%,更优选小于10%,更优选小于5%,更优选小于1%。可忽略的表达水平是使用实施例1的方法基本上不可检测到的表达水平。

[0036] 然而,应理解,技术人员可容易地修改实施例1中采用的方法。例如,这可涉及使用不同的细胞类型,使用不同的表达产物,使用成功剪接的不同指示剂(例如,测量编码功能性表达产物的剪接的mRNA的水平,或使用不同的报道蛋白),以及使用不同的UPR诱导物。在这种改进的方法中,仍然存在这样的情况,即功能性可调节内含子是一种能够在UPR诱导之后被成功剪接出从而在UPR诱导之后表达功能性表达产物的内含子。优选地,这导致在UPR诱导之后24小时编码功能性表达产物的转录物的表达提高至少5倍,更优选提高10倍,更优选提高100倍,并且更优选提高至少1000倍。优选的是,在诱导UPR之前,编码功能性表达产物的转录物的表达水平是最低的或可忽略不计。

[0037] CNG/CNG[CG]是哺乳动物细胞的一个优选剪接位点共有靶序列,其中优选在指示的位置处存在C或G(尽管不是必需的)。通常在该位置存在C优于G。因此,在本发明的一些优选实施方案中,特别是当合成核酸表达构建体旨在用于哺乳动物细胞时,内含子在内含子的一端、另一端或两端(优选两端)包含序列CNG/CNG[CG]。

[0038] 因此,在本发明的一些优选实施方案中,可调节内含子包含序列CNG/CNG-Xn-CNG/CNG[CG],其中Xn代表长度为n个核苷酸的序列,其中/代表切割位点,使得可切除序列CNG-Xn-CNG在剪接后从转录物中切除,其中序列Xn的5'端的核苷酸为C或G。Xn的合适长度如上所示。

[0039] 在本发明的一些优选实施方案中,Xn包含序列CACUCAGACUACGUGCACCU (SEQ ID NO:1) 或与其至少60%相同,更优选与其至少70%相同,更优选与其至少80%相同,更优选与其至少90%相同,并且更优选与其至少95%、96%、97%、98%或99%相同的序列。

[0040] 在本发明的一些优选实施方案中,Xn由以下序列组成:序列CACUCAGACUACGUGCACCU (SEQ ID NO:1) 或与其至少60%相同,更优选与其至少70%相同,更优选与其至少80%相同,更优选与其至少90%相同,并且更优选与其至少95%、96%、97%、98%或99%相同的序列。

[0041] 序列CACUCAGACUACGUGCACCU (SEQ ID NO:1) 对应于如上所述的位于IRE1切割位点内部的哺乳动物XBP1内含子的区域。因此,这代表了本发明的一个优选实施方案,特别是当合成核酸表达构建体旨在用于哺乳动物细胞中时。然而,在一系列非哺乳动物XBP1内含子中也发现了与此高度相似的序列。

[0042] 在本发明的一些实施方案中,Xn包含以下序列之一或由其组成:

[0043] -CACUCAGACUACGUGCACCU (SEQ ID NO:1);

[0044] -CACUCAGACUACGUGCUCU (SEQ ID NO:2);

[0045] -CACUCAGACUACGUGCCCU (SEQ ID NO:3);

[0046] -CACUCAGACUACGUGCGCU (SEQ ID NO:4);和

[0047] -CACUCAGACUAUGUGCACCU (SEQ ID NO:5)。

[0048] 在本发明的其他实施方案中, X_n 包含以下序列或由其组成: 序列 ACGGGCAACUUUACACGACG (SEQ ID NO:49) 或与其至少60%相同, 更优选与其至少70%相同, 更优选与其至少80%相同, 更优选与其至少90%相同, 并且更优选与其至少95%、96%、97%、98%或99%相同的序列。

[0049] 在本发明的一个特别优选的实施方案中, 可调节内含子包含以下序列或由其组成: 序列 CNG/CNGCACUCAGACUACGUGCACCUCNG/CNGC (SEQ ID NO:6), 或与其至少60%相同, 更优选与其至少70%相同, 更优选与其至少80%相同, 更优选与其至少90%相同, 并且更优选与其至少95%、96%、97%、98%或99%相同的序列, 其中/代表切割位点。在根据上述序列同一性水平的变体序列中, 剪接位点靶序列优选保留为 CNG/CNGC, 并且序列变异发生在其他区域中。

[0050] 合适地, 可调节内含子包含以下序列或由其组成: 序列 CAG/CAGCACUCAGACUACGUGCACCUCUG/CUGC (SEQ ID NO:7), 或与其至少60%相同, 更优选与其至少70%相同, 更优选与其至少80%相同, 更优选与其至少90%相同, 并且更优选与其至少95%、96%、97%、98%或99%相同的序列, 其中/代表切割位点。在根据上述序列同一性水平的变体序列中, 剪接位点靶序列优选保留为 CAG/CUGC, 并且序列变异发生在其他区域中。

[0051] 在本发明的一些优选实施方案中, 可调节内含子包含以下序列之一或由其组成:

[0052] -CNG/CAGCACUCAGACUACGUGCACCUCUG/CNG (SEQ ID NO:8);

[0053] -CNG/CAGCACUCAGACUACGUGCUCCUCUG/CNG (SEQ ID NO:9);

[0054] -CNG/CAGCACUCAGACUACGUGCCCCUCUG/CNG (SEQ ID NO:10);

[0055] -CNG/CAGCACUCAGACUACGUGCGCCUCUG/CNG (SEQ ID NO:11); 和

[0056] -CNG/CAGCACUCAGACUAUGUGCACCUCUG/CNG (SEQ ID NO:12)。

[0057] 在本发明的其他优选实施方案中, 可调节内含子包含以下序列之一或由其组成:

[0058] -CAG/CAGCACUCAGACUACGUGCACCUCUG/CUGC (SEQ ID NO:7);

[0059] -CAG/CAGCACUCAGACUACGUGCUCCUCUG/CUGC (SEQ ID NO:13);

[0060] -CAG/CAGCACUCAGACUACGUGCCCCUCUG/CUGC (SEQ ID NO:14);

[0061] -CAG/CAGCACUCAGACUACGUGCGCCUCUG/CUGC (SEQ ID NO:15); 和

[0062] -CAG/CAGCACUCAGACUAUGUGCACCUCUG/CUGC (SEQ ID NO:16)。

[0063] 在本发明的另一个实施方案中, 可调节内含子包含序列 CAG/CUGCAGCACUCAGACUACGUGCACCUCUG/CAG (SEQ ID NO:17) 或 CAG/CUGCAGCACUCAGACUACGUGCACCUCUG/CUGG (SEQ ID NO:27), 其中/代表切割位点。该序列是由于在具有下划线的位置向哺乳动物 XBP1 内含子序列添加了三核苷酸 CUG。据信, 添加这种三核苷酸会使内含子的剪接稍微不优化, 以降低细胞中任何不期望的剪接 (并因此表达产物的背景表达)。

[0064] 因此, 在本发明的一些优选实施方案中, X_n 包含 CAGCACUCAGACUACGUGCACCU (SEQ ID NO:23) 或由其组成。

[0065] 在本发明的另一个实施方案中, 可调节内含子包含序列: CNG/CAGACGGGCAACUUUACACGACGUG/CNG (SEQ ID NO:50), 或与其至少60%相同, 更优选与其至少70%相同, 更优选与其至少80%相同, 更优选与其至少90%相同, 并且更优选与其至少

95%、96%、97%、98%或99%相同的序列,其中/代表切割位点。在根据上述序列同一性水平的变体序列中,剪接位点靶序列优选保留为CNG/CNG,并且序列变异发生在其他区域中。

[0066] 为避免疑问,应注意,在本发明的内含子序列中,优选将内含子各末端的剪接位点靶序列限制为CNG/CNG,更优选限制为CNG/CNG[CG]。因此,尽管这些序列在剪接位点靶序列中提供了一些变化,但是如果需要的话,位于两个剪接位点靶序列之间的中央序列应容纳另外的变化。

[0067] 在本发明的一些优选实施方案中,可调节内含子是XBP1内含子、Hac1内含子、bZIP60内含子,或其同源物。这意味着内含子可以是XBP1、Hac1或bZIP60内含子的野生型形式,或其天然存在的同源物。

[0068] 在本发明的一些实施方案中,可优选的是,转录物中的剪接位点靶序列(即包含序列CNG/CNG)侧翼为能够相互作用以形成茎环结构的序列。因此,剪接位点靶序列优选地侧翼为彼此互补的序列,使得它们将彼此杂交以形成茎环结构,其中剪接位点靶序列至少部分地,优选整个地位于在转录物中形成的茎环结构的环区域内。

[0069] 关于野生型XBP1、HAC1和bZIP60内含子,已经假设在mRNA转录物中的剪接位点处形成茎环结构。这涉及内含子和与剪接位点靶序列相邻的外显子中的序列之间的杂交,当以相反方向阅读时,所述序列在核苷酸序列中是互补的。本文中报道的实验表明,在预期不会形成这种茎环结构的情况下,将内含子插入到编码序列中时,剪接成功地进行。因此,形成茎环结构对于成功地剪接本发明的可调节内含子而言似乎不是必要的。然而,在某些情况下,可期望形成茎环结构,因为这可例如导致最佳的剪接活性。在一个替代方案中,在某些情况下可期望避免提供可被修饰以形成茎环结构的序列,因为这可导致内含子的不期望的主动剪接,从而潜在地导致表达漏失(expression leakage)。

[0070] 在本发明的某些实施方案中,在待形成茎环结构的情况下,由所述转录物形成的茎环结构优选包含长度为6至9个核苷酸的环和长度为3至10个核苷酸的茎。更优选地,茎环结构包含长度为7至8个核苷酸的环和长度为4至8个核苷酸的茎。

[0071] 在待形成茎环结构的本发明的某些实施方案中,内含子可合适地在剪接靶位点包含如下序列:

[0072] -Yn-CNG/CNG-A-Zn-

[0073] 其中A是长度为0至3个核苷酸(优选1或2个核苷酸)的序列,

[0074] 其中/代表切割位点,

[0075] 并且其中Yn和Zn表示当在相反方向上阅读时在核苷酸序列上是互补的序列,因此能够杂交形成茎环结构的茎。Yn和Zn的长度优选为3至10个核苷酸,更优选为4至8个核苷酸。

[0076] 在一些实施方案中,内含子可合适地在剪接靶位点包含如下序列:

[0077] -Zn-CNG/CNG[CG]-A-Yn-,其中组分具有与以上相同的含义。在这种情况下,A优选具有0、1或2个核苷酸的长度。

[0078] 显然,提供合适的互补序列(例如以上结构中的Yn和Zn)以提供茎结构可通过使内含子的序列进行改变以提供与相邻的编码(即外显子)序列中的相应序列互补的合适区域来实现。也可能或期望在某种程度上改变编码区的序列,例如通过利用遗传密码中的冗余来改变核酸序列而不影响编码的氨基酸序列;通常应避免表达产物的氨基酸序列的改变。

[0079] 在本发明的一些实施方案中,核酸表达构建体包含与编码表达产物的核酸序列有效连接的诱导型启动子,所述编码表达产物的核酸序列包含编码可调节内含子的序列。如上所提及的,诱导型启动子是本领域已知的。通过组合本发明的诱导型启动子与可调节内含子,可实现双重表达水平,即在转录水平和翻译水平二者上的控制。这可允许非常严密的表达控制,例如避免任何表达“漏失”。例如,在毒性蛋白的表达期间,或在所期望的时间诱导表达之前,必须将背景表达量保持在绝对最低水平的任何情况下,这一点都可以是重要的。

[0080] 在一个优选实施方案中,诱导型启动子是未折叠蛋白反应(UPR)诱导型启动子,即,其自身被UPR诱导的启动子。在这样的实施方案中,UPR的诱导既在驱动转录方面诱导表达,又由于可调节内含子的剪接而允许表达功能性表达产物。

[0081] 在本发明的一些实施方案中,UPR诱导型启动子合适地包含针对以下的至少一个结合位点:在UPR中驱动基因表达的ATF6、XBP1、bZIP60,或者同源或在其他方面等效的转录因子。

[0082] 合适地,UPR诱导型启动子包含至少一个以下序列中的一个或更多个拷贝:

[0083] -TGACGTG(ATF6共有序列),

[0084] -TGACGTGCT(以上的变体),

[0085] -TGACGTG[TG](称为UPRE位点),

[0086] -CCAAT-N9-CCACG(称为ERSE1位点)(SEQ ID NO:18),和

[0087] -ATTGG-N-CCACG(称为ERSE2位点)(SEQ ID NO:19)。

[0088] 这些位点由ATF6、XBP1和bZIP60结合。

[0089] 合适地,启动子包含序列TGACGTG(任选地作为TGACGTGCT或TGACGTG[TG]的一部分)的一个或更多个拷贝,优选地序列TGACGTG[TG](任选地作为TGACGTGCT或TGACGTG[TG]的一部分)的3个或更多个拷贝,并且优选地序列TGACGTG[TG](任选作为TGACGTGCT或TGACGTG[TG]的一部分)的5个或更多拷贝。

[0090] 一个示例性的UPR诱导型启动子序列包含以下序列(SEQ ID NO:20):TGACGTGCTTGACGTGCTTGACGTGCTTGACGTGCTTGACGTGCT。

[0091] 该序列包含UPRE位点的6个串联拷贝。

[0092] 另一个示例性的UPR诱导型启动子序列包含以下序列(SEQ ID NO:47):

TGACGTGCTGATGATGCGTAGCTAGTAGTTGACGTGCTGATGATGCGTAGCTAGTAGTTGACGTGCTG

[0093] ATGATGCGTAGCTAGTAGTTGACGTGCTGATGATGCGTAGCTAGTAGTTGACGTGCTGATGATGCGTAGCTAGTAGTTGACGTGCT
GCTAGTAGTTGACGTGCT

[0094] 该序列包含UPRE位点的6个拷贝,每个拷贝间隔20个核苷酸。

[0095] 合适地,UPR诱导型启动子包含所述针对以下的至少一个连接位点:ATF6、XBP1,或者驱动UPR与最小启动子序列(例如,CMV最小启动子)有效连接的同源或在其他方面等效的转录因子。其他合适的最小启动子是本领域已知的。

[0096] CMV最小启动子具有以下序列(SEQ ID NO:21):

AGGTCTATATAAGCAGAGCTCGTTTAACTGAACCGTCAGATCGCCTAGATACGCCATCCACGCTGTTTT

[0097] GACCTCCATAGAAGAT

[0098] 因此,示例性诱导型启动子包含含有根据SEQ ID NO:20的序列的核酸,该核酸位

于具有根据SEQ ID NO:21的序列的核酸的上游并与之有效连接。

[0099] 例如,诱导型启动子可合适地包含以下序列(SEQ ID NO:22):

TGACGTGCTTGACGTGCTTGACGTGCTTGACGTGCTTGACGTGCTTGGTACCGTCGACGA

[0100] TATCGGATCCAGGTCTATATAAGCAGAGCTCGTTTGTAGTGAACCGTCAGATCGCCTAGATACGCCATCC
ACGCTGTTTTGACCTCCATAGAAGATCGCCACC

[0101] 下文提供了可与可调节内含子结合使用的UPR诱导型启动子的更多细节。

[0102] 编码表达产物的核酸序列合适地是转基因。转基因通常编码基因表达产物,例如RNA或多肽(蛋白质)。转基因可以是全长cDNA或基因组DNA序列,或其具有至少一些生物学活性的任何片段、亚基或突变体。转基因可合适地是微小基因(minigene),即,缺少其部分、大部分或全部的天然内含子序列的基因序列。转基因任选地可包含常规内含子序列(即除了可调节内含子之外)。任选地,转基因可以是杂交核酸序列,即由同源和/或异源cDNA和/或基因组DNA片段构建的序列。“突变体”意指包含与野生型或天然存在的序列不同的一个或更多个核苷酸的核酸序列,即,突变的核酸序列包含一个或更多个核苷酸替换、缺失和/或插入。核苷酸的替换、缺失和/或插入可产生其氨基酸/核酸序列与野生型氨基酸/核酸序列不同的基因产物(即蛋白质或核酸)。在一些情况下,转基因还可包括编码前导肽的序列或信号序列,使得转基因产物将从细胞中分泌出。

[0103] 除了上面讨论的可调节内含子之外,常规的(即不可调节的)内含子也可用于编码表达产物的核酸序列中。术语“常规内含子”涵盖整个内含子的任何部分,该部分足够大以被核剪接装置识别和剪接。通常,短的、功能性的内含子序列是优选的,以使表达组件的尺寸保持尽可能小,这有利于表达组件的构建和操作。在一些实施方案中,内含子是从编码由编码表达产物的核酸序列编码的蛋白质的基因中获得的。常规内含子可位于编码表达的序列的5',编码表达的序列的3'或位于编码表达的序列内。因此,在一些实施方案中,编码表达产物的核酸序列还包含常规内含子。合适的内含子的一些非限制性实例是小鼠微小病毒(Minute Virus of Mice,MVM)内含子、 β -珠蛋白内含子(β alVS-II)、IX因子(FIX)内含子A、猿猴病毒40(SV40)小t内含子和 β -肌动蛋白内含子。如本领域公知的,内含子可在表达水平提高方面具有益处。

[0104] 在本发明的一些优选实施方案中,编码表达产物的核酸序列编码蛋白质。基本上可使用任何蛋白质,并且作为非限制性实例,该蛋白质可以是酶、抗体或抗体片段(例如单克隆抗体)、病毒蛋白(例如REP-CAP、REV、VSV-G或RD114)、治疗性蛋白质或毒性蛋白(例如胰天蛋白酶3、8或9)。

[0105] 在本发明的一些优选实施方案中,编码表达产物的核酸序列编码毒性蛋白。在这种情况下,“毒性蛋白”意指对使用中产生表达产物的细胞具有毒性的蛋白质。例如,毒性蛋白可以是以下之一:胰天蛋白酶3、胰天蛋白酶8、胰天蛋白酶9和毒性病毒蛋白(例如VSV-G或AAV REP蛋白)。

[0106] 在本发明的一些优选实施方案中,编码表达产物的核酸序列编码治疗性表达产物。治疗性表达产物可以是蛋白质,例如可分泌蛋白质,例如凝血因子(例如IX因子或VIII因子)、细胞因子、生长因子、抗体或纳米抗体、趋化因子、血浆因子、胰岛素、促红细胞生成素、脂蛋白脂肪酶或毒性蛋白。

[0107] 作为备选地,治疗性表达产物可以是RNA,例如siRNA或miRNA。多种治疗性siRNA已

经在本领域中描述,并且作为非限制性实例,该siRNA可用于治疗FTDP-17(额颞痴呆(frontotemporal dementia))、DYT1张力失常、生长激素缺乏症、阿尔茨海默病中的BACE1、白血病(例如靶向c-raf、bcl-2)、黑素瘤(例如靶向ATF2、BRAF)、前列腺癌(例如靶向P110B)和胰腺癌(例如靶向K-Ras)。siRNA治疗概述于“Therapeutic potentials of short interfering RNAs”,Appl Microbiol Biotechnol,DOI10.1007/s00253-017-8433-z中。类似地,对于miRNA,可根据本发明实施的多种miRNA治疗方法概述于“MicroRNA therapeutics:towards a new era for the management of cancer and other diseases”,Nature Reviews Drug Discovery;16,203-222(2017)中。

[0108] 合适地,核酸表达构建体包含提供或编码核糖体结合位点、起始密码子、终止密码子中的一个或多个,优选所有的序列,以及转录终止序列。

[0109] 在另一方面中,本发明提供了核酸,其包含编码表达产物的序列(例如基因),该编码表达产物的序列包含编码可调节内含子的序列,所述可调节内含子是包含可切除序列的内含子,所述可切除序列能够通过细胞中的未折叠蛋白反应(UPR)体系从合成表达构建体产生的转录物中剪接出,从而产生编码功能性表达产物的转录物,并且其中表达产物不是XBP1蛋白、Hac1蛋白、bZIP60蛋白,或其同源物。

[0110] 因此,编码表达产物的序列不是天然包含根据本发明的可调节内含子的基因。换句话说,可调节内含子与编码在其中发现其的表达产物的序列异源。

[0111] 可调节内含子的优选特征和编码表达产物的序列如上所述。

[0112] 可将这样的核酸插入到任何合适的表达构建体,例如使得与启动子有效连接的表达载体,和驱动核酸转录所需的任何其他元件中。功能性表达产物的表达将由可调节内含子控制。

[0113] 在另一方面中,本发明提供了载体,其包含如上所述的合成核酸表达构建体。

[0114] 术语“载体”在本领域中是公知的,并且如在本申请中使用的是指核酸分子,例如双链DNA,其可已经插入到根据本发明的核酸表达构建体中。载体适用于将插入的核酸分子转运到合适的宿主细胞中。载体通常包含允许转录插入核酸分子,并且优选地将转录物翻译成多肽的所有必需元件。载体通常包含所有必需的元件,使得一旦载体进入宿主细胞中,载体就可独立于宿主染色体DNA复制或与宿主染色体DNA一致复制;可产生载体及其插入的核酸分子的数个拷贝。本发明的载体可以是游离型载体(即,其不整合到宿主细胞的基因组中),或者可以是整合到宿主细胞的基因组中的载体。该定义包括非病毒和病毒载体二者。非病毒载体包括但不限于质粒载体(例如pMA-RQ、pUC载体、bluescript载体(pBS)和pBR322,或其不含细菌序列(微环)的衍生物)、基于转座子的载体(例如PiggyBac(PB)载体或Sleeping Beauty(SB)载体)等。更大的载体(例如人工染色体(细菌(BAC)、酵母(YAC)或人(HAC)))可用于容纳更大的插入片段。病毒载体来源于病毒,包括但不限于以下病毒的载体:逆转录病毒、慢病毒、腺相关病毒、腺病毒、疱疹病毒、肝炎病毒等。通常但不是必要的,病毒载体具有复制缺陷,因为它们已经失去了在给定细胞中繁殖的能力,因为复制必需的病毒基因已经从病毒载体中消除。然而,一些病毒载体也可适合于在给定的细胞(例如如,癌细胞)中特异性复制,并且通常用于触发细胞(癌细胞)特异性分解(溶瘤分解)。病毒体是同时包含病毒和非病毒成分的载体的一个非限制性实例,特别是它们将脂质体与灭活的HIV或流感病毒组合(Yamada等,2003)。另一个实例包括与阳离子脂质混合的病毒载体。

[0115] 在一些优选实施方案中,载体是病毒载体,例如逆转录病毒载体、慢病毒载体、腺病毒载体或腺相关病毒(AAV)载体,更优选AAV载体。为了克服AAV转导中的限制步骤之一(即单链至双链AAV转化),AAV载体优选以自互补双链AAV载体(scAAV)使用(McCarty,2001,2003;Nathwani等,2002,2006,2011;Wu等,2008),尽管本文也涵盖了单链AAV载体(ssAAV)的用途。

[0116] 在一些优选实施方案中,载体是质粒。这样的质粒可包括多种其他功能性核酸序列,例如一种或更多种选择性标志物、一个或更多个的复制起点、多克隆位点等。

[0117] 在本发明的一些优选实施方案中,载体是用于在真核细胞中表达的表达载体。真核表达载体的一些实例包括但不限于可从Stratagene获得的pW-LNE0、pSV2CAT、pOG44、pXT1和pSG;可从Amersham Pharmacia Biotech获得的pSVK3、pBPV、pMSG和pSVL;和可从Clontech获得的pCMVDsRed2-express、pIRES2-DsRed2、pDsRed2-Mito、pCMV-EGFP。许多其他载体是公知的且可商购。对于哺乳动物细胞腺病毒载体,pSV和pCMV系列载体是特别公知的非限制性实例。有许多公知的酵母表达载体,包括但不限于酵母整合型质粒(yeast integrative plasmid,YIp)和酵母复制型质粒(yeast replicative plasmid,YRp)。对于植物,农杆菌属(agrobacterium)的Ti质粒是一个示例性表达载体,并且植物病毒也提供合适的表达载体,例如烟草花叶病毒(tobacco mosaic virus,TMV)、马铃薯X病毒和豇豆花叶病毒。

[0118] 在一些优选实施方案中,载体是基因治疗载体。多种基因治疗载体是本领域已知的,并且可提及AAV载体、腺病毒载体、逆转录病毒载体和慢病毒载体。当载体是基因治疗载体时,编码表达产物的核酸序列合适地编码治疗性蛋白质。治疗性蛋白质可以是可分泌的蛋白质。可分泌的蛋白质,特别是可分泌的治疗性蛋白质的非限制性实例包括凝血因子(例如VII I因子或IX因子)、胰岛素、促红细胞生成素、脂蛋白脂肪酶、抗体或纳米抗体、生长因子、细胞因子、趋化因子、血浆因子、毒性蛋白等。

[0119] 本发明的核酸表达构建体和载体可与可药用赋形剂(即一种或更多种可药用载体物质和/或添加剂,例如缓冲液、载体、赋形剂、稳定剂等)一起配制成药物组合物。药物组合物可以药盒的形式提供。本文中所用的术语“可药用”与本领域一致,并且意指与药物组合物的其他成分相容并且对其接受者无害。

[0120] 因此,本发明的另一方面提供了药物组合物,其包含本文中所述的核酸表达构建体或载体。

[0121] 在本发明的另一方面中,提供了根据本发明多个方面的核酸表达构建体和载体用于制备药物组合物的用途。

[0122] 根据本发明的另一方面,提供了细胞,其包含根据本发明的合成核酸表达构建体或载体。

[0123] 优选地,细胞是真核细胞。真核细胞可合适地是真菌细胞(例如酵母细胞)、动物(后生动物)细胞(例如哺乳动物细胞)或植物细胞。

[0124] 在本发明的一些实施方案中,细胞可以是原核细胞;尽管原核细胞不具有UPR,但是原核细胞仍可用于产生合成核酸表达构建体或者用于处理合成核酸表达构建体的其他步骤中。

[0125] 在本发明的一些优选实施方案中,细胞是离体的,例如在细胞培养中。在本发明的

另一些实施方案中,细胞可以是组织或多细胞生物体的一部分。

[0126] 在一些优选实施方案中,表达产物对其中存在构建体或载体的细胞具有毒性。在一个这样的实施方案中,细胞是用于细胞治疗的细胞(例如治疗性免疫细胞、例如治疗性T细胞),其包含根据本发明的载体的合成核酸表达构建体,其中表达产物为对细胞具有毒性。在这样的一个实施方案中,对UPR的诱导可被用于诱导细胞死亡,即作为生死开关(kill switch)。合适的毒性表达产物包括胱天蛋白酶,例如胱天蛋白酶3、胱天蛋白酶8或胱天蛋白酶9。

[0127] 合成核酸表达构建体可插入到细胞的基因组中,或者可存在于游离型载体中。

[0128] 在另一方面中,本发明提供了产生表达产物的方法,所述方法包括:

[0129] a) 提供包含根据本发明的合成核酸表达构建体的真核细胞群;

[0130] b) 处理所述细胞群,以诱导未折叠蛋白反应,从而从可调节内含子中剪接出可切除序列;

[0131] c) 在合适的条件下孵育所述细胞群以产生表达产物;以及

[0132] d) 从所述细胞群中分离表达产物。

[0133] 该方法是适合的细胞培养方法。因此,可在对于所用细胞类型而言适合的细胞培养条件下提供细胞。合适的细胞培养条件是技术人员公知的。

[0134] 合成核酸表达构建体可存在于基因组中或可以是游离的。

[0135] 显然,本发明允许将表达产物(或表达产物的活性形式)的产生延迟直至细胞培养过程中的所需时间。例如,这可允许细胞群扩增,直到达到所需细胞数量或浓度,或达到所需生长期为止。

[0136] 例如,在毒性蛋白的情况下,可避免功能性(即毒性)表达产物的产生,直到细胞培养体系处于所需阶段为止。一旦毒性蛋白被表达,细胞当然将受到不良影响或被杀伤。

[0137] 该方法合适地包括在步骤b)处理所述细胞群,以诱导未折叠蛋白反应(UPR)之前,在适合于所述细胞生长的条件下孵育所述细胞群。

[0138] 通常,步骤b)包括向所述细胞施加应激(stress),所述应激适合于诱导UPR。有广泛范围的应激可用于诱导UPR,并且在文献中对此进行了广泛的描述。

[0139] 在一些优选实施方案中,诱导UPR的步骤合适地包括施用能够在所述细胞中诱导UPR的化学剂(即,诱导UPR的化学剂)。

[0140] 本领域技术人员可容易地评估任何特定应激(例如化学剂)诱导UPR的能力。例如,本领域技术人员可在实施例1的方法中合适地施加应激(例如通过施用候选化学剂)来代替DTT。试剂诱导UPR的能力将通过施加应激(例如化学剂)对功能性表达产物(即EGFP,在实施例1的情况下)的表达的影响来鉴定。当然,如果需要的话,可修改实施例1的方法,例如使用不同的细胞类型或不同的构建体。具体地,实施例4证明了如何测试多种候选化学剂诱导UPR的能力。评估应激(例如化学剂)诱导UPR的能力的多种其他方法对于本领域技术人员将是明显的。

[0141] 可诱导ER应激并诱导UPR的许多化学剂是本领域已知的。能够诱导IRE1途径的诱导UPR的化学剂适用于本发明,因为它们将导致可调节内含子的剪接。

[0142] 可用于诱导UPR的一些示例性诱导UPR的化学剂包含:

[0143] • 二硫苏糖醇(DTT) - 该试剂可降低蛋白质的二硫桥。变性的蛋白质在ER内部积

累。

- [0144] • 衣霉素-该试剂抑制N-连接的糖基化。
- [0145] • 布雷菲德菌素A(brefeldin A)-通常用作未折叠蛋白反应的诱导物。
- [0146] • 毒胡萝卜素(thapsigargin)-该试剂由于抑制Sarco/内质网Ca²⁺-ATP酶(Sarco/Endoplasmic Reticulum Ca²⁺-ATPase, SERCA)而导致ER Ca²⁺耗竭。
- [0147] • 2-脱氧葡萄糖
- [0148] • A23187(CAS编号52665-69-7)
- [0149] • 硼替佐米(bortezomib)(Velcade)
- [0150] • 槲皮素(queracetin)
- [0151] • 破坏细胞中的脂质平衡使得诱导UPR的试剂,例如饱和脂肪酸(例如棕榈酸)-即那些诱导脂质诱导的ER应激反应/UPR的试剂。

[0152] 可以合适的浓度施用这种诱导UPR的化学剂,并且本领域技术人员可通过常规实验和文献查阅容易地确定这种浓度。多种药剂的合适施用浓度如下:DTT 2mM;衣霉素2.5至5μg/ml;布雷菲德菌素A 0.5μg/ml;毒胡萝卜素0.1至1μM;2-脱氧葡萄糖4mM;A23187(CAS编号52665-69-7)0.5μM;硼替佐米(Velcade)5至30nM;和棕榈酸(或其他脂肪酸)100μM。这些浓度是指试剂在细胞所暴露的培养基中的浓度。

[0153] 用于本发明的另一种诱导UPR的化学剂是毛喉素(forskolin)。毛喉素(coleonol)是由印度锦紫苏植物(毛喉鞘蕊花(Plectranthus barbatus))产生的一种半日花烷型二萜(labdane diterpene)。毛喉素的其他名称包括pashanabhedi、印度锦紫苏、makandi、HL-362和NKH477。

[0154] 二硫苏糖醇(DTT)、衣霉素和毒胡萝卜素在文献中广泛用于诱导UPR,因此在本发明的一些实施方案中代表优选的诱导UPR的化学剂。毛喉素是另一种优选的诱导UPR的化学剂,特别是但不仅限于其在体内使用的安全性。

[0155] 在本发明的一个特别优选的实施方案中,施用能够破坏所述细胞中脂质平衡的诱导UPR的化学剂以诱导UPR。在文献中已经广泛报道了脂质和脂质代谢在诱导UPR中的作用,该现象被称为“脂质诱导的ER应激反应/UPR”。参见,例如Iwao和Shidoj, PLOS ONE | DOI: 10.1371/journal.pone.0132761 July 17, 2015; Robblee等“Saturated Fatty Acids Engage an IRE1α-Dependent Pathway to Activate the NLRP3 Inflammasome in Myeloid Cells”-Cell Reports 14, 2611-2623, March 22, 2016; Ariyama等, “Decrease in Membrane Phospholipid Unsaturation Induces Unfolded Protein Response”-THE JOURNAL OF BIOLOGICAL CHEMISTRY VOL. 285, NO. 29, pp. 22027-22035, July 16, 2010; Basseri和Austin“Endoplasmic Reticulum Stress and Lipid Metabolism: Mechanisms and Therapeutic Potential”-Biochemistry Research International Volume 2012, Article ID 841362, 13页, doi:10.1155/2012/841362; Kitai等“Membrane lipid saturation activates IRE1α without inducing clustering”-Genes to Cells(2013) 18, 798-809。因此,合适地,诱导UPR的化学剂能够以UPR被诱导的这样的方式改变细胞的脂质平衡。存在广泛范围可实现此目的的试剂。例如,已经表明,以提高脂质饱和水平(从而降低去饱和脂质水平)的这样的方式破坏细胞中的脂质平衡导致UPR的诱导。因此,合适地,诱导UPR的化学剂能够以提高脂质饱和水平的这样的方式改变细胞的脂质平衡。

[0156] 优选地,诱导UPR的化学剂能够改变细胞膜中饱和脂肪酸与不饱和脂肪酸的比,从而提高饱和脂肪酸的比例。这可通过数种方式来实现,例如将饱和脂肪酸引入至细胞或抑制将饱和脂肪酸转化为不饱和脂肪酸的酶的活性。

[0157] 因此,在一个特别优选的实施方案中,诱导UPR的化学剂包含饱和脂肪酸,合适地是中链或长链饱和脂肪酸。在本发明的某些实施方案中,脂肪酸的脂族链长为6至26个碳,更优选9至22个碳,更优选12至20个碳,更优选14至20个碳。

[0158] 合适地,所述诱导UPR的化学剂包含选自以下的至少一种脂肪酸:己酸、庚酸、辛酸、壬酸、癸酸、十一烷酸、月桂酸、十三烷酸、肉豆蔻酸、十五烷酸、棕榈酸、十七烷酸、硬脂酸、十九烷酸、二十烷酸、二十一烷酸、二十二烷酸、二十三烷酸、二十四烷酸、二十五烷酸和二十六烷酸。更优选地,诱导UPR的化学剂包含选自以下的至少一种脂肪酸:壬酸、癸酸、十一烷酸、月桂酸、十三烷酸、肉豆蔻酸、十五烷酸、棕榈酸、十七烷酸、硬脂酸、十九烷酸、二十烷酸、二十一烷酸和二十二烷酸。更优选地,诱导UPR的化学剂包含选自以下的至少一种脂肪酸:月桂酸、十三烷酸、肉豆蔻酸、十五烷酸、棕榈酸、十七烷酸、硬脂酸、十九烷酸和二十烷酸。更优选地,诱导UPR的化学剂包含选自以下的至少一种脂肪酸:肉豆蔻酸、十五烷基酸、棕榈酸、十七烷酸、硬脂酸、十九烷酸和二十烷酸。

[0159] 在一个特别优选的实施方案中,诱导UPR的化学剂包含棕榈酸或硬脂酸。

[0160] 应当注意的是,在提到脂肪酸的情况下,可以任何合适的形式提供它们,例如作为盐(棕榈酸盐、硬脂酸盐等)。

[0161] 在另一个优选实施方案中,诱导UPR的化学剂包含能够在细胞中下调硬脂酰辅酶A脱饱和酶活性的试剂。硬脂酰辅酶A脱饱和酶是一种ER标记酶(resident enzyme),其可在饱和脂肪酸中引入双键。例如,诱导UPR的化学剂合适地包括硬脂酰辅酶A脱饱和酶的抑制剂。MF-43是硬脂酰辅酶A脱饱和酶的合适抑制剂的一个实例。作为备选地,诱导UPR的化学剂可能下调硬脂酰辅酶A脱饱和酶的表达,例如通过敲低硬脂酰辅酶A脱饱和酶表达(例如通过RNA干扰),或敲除硬脂酰辅酶A脱饱和酶基因。也可靶向参与脂肪酸去饱和的其他基因。

[0162] 其他示例性的诱导UPR的化学剂包含香叶基香叶酸(geranylgeranoic acid, GGA)、2,3-二氢GGA、9-顺式视黄酸和全反式视黄酸(参见-Chieko Iwao和Yoshihiro Shidoj, PLOS ONE | DOI:10.1371/journal.pone.0132761 July 17, 2015)。

[0163] 在本发明的一些实施方案中可使用数种诱导UPR的化学剂的组合。例如,饱和脂肪酸(例如棕榈酸)可与能够下调硬脂酰辅酶A脱饱和酶活性的试剂(例如MF-43)组合使用。

[0164] 在本发明的另一个实施方案中,在步骤b)中诱导UPR合适地包括在所述真核细胞群中表达诱导物蛋白,从而在细胞群中诱导UPR响应。这样的蛋白被称为“诱导物蛋白”,因为当其在细胞群中表达时其通常通过产生ER应激发挥作用以诱导UPR响应。诱导物蛋白通常是与根据本发明第一方面的合成核酸表达构建体所编码的蛋白质不同的蛋白质。诱导物蛋白合适地是异源蛋白,但是在一些实施方案中,它可以是过表达的同源蛋白。重要的是诱导物蛋白在细胞中的表达诱导了UPR,这导致了可调节内含子的剪接。有许多方法可在细胞群中实现诱导物蛋白的表达。在一些合适的实施方案中,用适于在细胞中表达诱导物蛋白的表达载体转染细胞群。在一个实施方案中,可用病毒感染细胞,这导致了引起ER应激和UPR诱导的病毒蛋白表达。也可使用编码病毒或非病毒蛋白的重组病毒载体,下面描述一个

实例,其中在细胞中表达AAV以诱导UPR。这形成了本发明的一个优选实施方案。作为备选地,可将基本上任何其他形式的表达载体(例如质粒)引入到细胞内以表达异源蛋白。用合适的表达载体转染细胞的合适方法是本领域公知的。诱导物蛋白的性质通常不受特别关注,尽管通常优选该蛋白是无毒的。相反地,异源蛋白产生在细胞中产生的ER应激是重要的。

[0165] 在一些实施方案中,步骤b)包括用能够在所述细胞中表达诱导物蛋白,优选异源蛋白的表达载体转染所述细胞群。作为备选地,步骤b)可包括从先前(例如在步骤a)之前)引入到细胞中的表达载体诱导所述诱导物蛋白的表达。

[0166] 诱导物蛋白的表达可在组成型或非组成型启动子的控制下。一个示例性的非组成型启动子是诱导型启动子(在这种情况下,诱导型启动子将不是UPR诱导型启动子)。

[0167] 诱导未折叠蛋白反应的其他方法包括使细胞暴露于缺氧或碳水化合物(例如葡萄糖)剥夺。

[0168] 如上所述,鉴于IRE1介导的内含子剪接在整个真核生物中的普遍存在,该方法可用任何类型的真核细胞进行。因此,该方法可例如在真菌细胞(例如酵母细胞)、动物(后生动物)细胞(例如哺乳动物细胞)和植物细胞中进行。

[0169] 在某些优选实施方案中,真核细胞群是动物(后生动物)细胞群。合适地,动物细胞可以是来自无脊椎动物或脊椎动物的细胞。

[0170] 在一些优选实施方案中,真核细胞群是哺乳动物细胞群。可使用广泛范围的哺乳动物细胞,包括但不限于中国仓鼠卵巢(CHO)细胞、人胚肾(HEK)细胞(例如HEK-293)、人视网膜细胞、人羊水细胞、小鼠骨髓瘤淋巴母细胞样细胞。在这样的实施方案中,可调节内含子是这样的内含子可以是优选的,所述内含子在内含子的一个、另一个或两个(优选两个)末端包含序列CNG/CNG[CG],即哺乳动物剪接靶共有序列。合适地,例如,可调节内含子具有序列

CNG/CNGCACUCAGACUACGUGCACCUCNG/CNGC (SEQ ID NO: 6),

[0171] **CAG/CAGCACUCAGACUACGUGCACCUCUG/CUGC (SEQ ID NO: 7) 或**

CAG/CUGCAGCACUCAGACUACGUGCACCUCUG/CUGC (SEQ ID NO: 17) 。

[0172] 在另一些实施方案中,真核细胞群是昆虫细胞群。在方法中使用的合适的昆虫细胞包括杆状病毒感染的细胞和未感染的细胞,例如来自以下昆虫物种的细胞:草地贪夜蛾(*Spodoptera frugiperda*) (例如Sf9或Sf21)、粉纹夜蛾(*Trichoplusia ni*) (例如Hi-5)、黑腹果蝇(*Drosophila melanogaster*) (例如Schneider 2细胞和Schneider 3细胞)。

[0173] 在本发明的另一些实施方案中,真核细胞群合适地是真菌细胞群,优选酵母细胞。用于本方法的合适的真菌细胞包括但不限于啤酒酵母(*Saccharomyces cerevisiae*)、毕赤酵母(*Pichia pastoris*)、曲霉属(*Aspergillus* spp.)、木霉属(*Trichoderma* spp.)和嗜热毁丝霉(*Myceliophthora thermophila*)。

[0174] 在本发明的另一些实施方案中,真核细胞群合适地是植物细胞群或植物原生质体群。

[0175] 在本发明的另一些实施方案中,真核细胞群合适地是原生动物细胞群,例如蜥蜴利什曼原虫(*Leishmania tarentolae*)。

[0176] 步骤d),即从所述细胞群中分离表达产物,可使用本领域公知的常规技术进行。当

然,这些技术将根据表达产物的性质而变化。

[0177] 该方法可合适地包括将核酸表达构建体引入到细胞内的步骤。有许多公知的转染真核细胞的方法,并且技术人员可容易地为任何细胞类型选择合适的方法。当然可在任何合适的载体中提供核酸表达构建体。

[0178] 在另一方面中,本发明提供了根据本发明多个方面的核酸表达构建体、载体、细胞或药物组合物,其用于治疗方法或治疗。

[0179] 如本文中所用,术语“治疗”是指治疗性处理和预防性或预防性措施二者。有益或期望的临床结果包括但不限于预防不期望的临床状态或疾病、降低疾病的发生率、减轻与疾病有关的症状、减轻疾病的程度、使疾病状态稳定(即不恶化)、使疾病进展延缓或减慢、改善或减轻疾病状态、可检测或不可检测的缓解(无论是部分还是全部),或其组合。与未接受治疗的预期生存期相比,“治疗”还意味着生存期延长。

[0180] 如本文中所用,术语“治疗性处理”或“治疗”等是指其目的是:使对象的身体或其部分(element)从不期望的生理变化或病症达到期望状态,例如不那么严重或不那么不愉快的状态(例如,改善或减轻),或者恢复到其正常的健康状态(例如,恢复对象的健康、身体完整性和身体健康);使其保持所述不期望的生理变化或病症(例如,稳定或不恶化);或与所述不期望的生理变化或病症相比阻止或减慢其进展至更严重或更差的状态。

[0181] 如本文中所用,术语“预防”、“预防性治疗”或“预防性治疗”等包括预防疾病或病症的发作,包括在患所述疾病或病症之前降低与其相关的疾病或病症的严重程度或症状。这种在患病之前的预防或降低是指将本文中所述的核酸表达构建体、载体或药物组合物施用至在施用时尚未患有明显疾病或病症的症状的患者。“预防”还包括例如在一段时间的改善之后预防疾病或病症的再现或复发预防。在一些实施方案中,本文中所述的核酸表达构建体、载体或药物组合物可用于基因治疗。

[0182] 本发明还提供了本文中所述的核酸表达构建体、载体或药物组合物用于制备用于基因治疗的药物的用途。

[0183] 本文中公开了用于在需要所述基因治疗的对象中进行基因治疗的方法,其包括:

[0184] -向所述对象中引入包含含有本发明的核酸表达构建体的药物组合物的基因治疗载体,所述核酸表达构建体包含编码治疗性表达产物的序列,使得所述基因治疗载体将所述核酸表达构建体递送至所述对象的靶细胞;以及

[0185] -在对象的靶细胞中表达治疗有效量的功能性治疗性表达产物。

[0186] 显然,只有当表达产物在UPR具有活性的细胞中表达时,才发生治疗有效量的功能性治疗性表达产物的表达。在许多情况下,UPR将在处于应激下的细胞(例如癌细胞或被病原体(例如病毒)感染的细胞)中具有活性。因此,本发明的一个优点是仅在UPR具有活性的细胞中发生功能性治疗产物(例如其可以是毒性蛋白或其他细胞毒性剂)的表达。这对于减轻或避免功能性治疗性表达产物的不期望的脱靶表达可以是有用的。

[0187] 因此,在本发明的涉及治疗的多个方面中,优选地,待治疗的病症是癌症或感染(例如,病毒感染)。

[0188] 作为替代地,可使用合适的UPR诱导剂在细胞中诱导UPR。上文讨论了多种UPR诱导剂,并且合适的UPR诱导剂包括诱导UPR的可药用试剂(例如,硼替佐米(Bortezomib)、毛喉

素)和干扰细胞中的脂质平衡以诱导UPR的试剂(例如,以上讨论的饱和脂肪酸)。可将UPR诱导剂直接递送至靶位点(例如通过注射)或全身性给予。

[0189] 治疗性表达产物可以是多肽/蛋白质,例如可分泌蛋白质或肽,例如如凝血因子(例如VII I因子或IX因子)、胰岛素、促红细胞生成素、脂蛋白脂肪酶、抗体或纳米抗体、生长因子、细胞因子、趋化因子、血浆因子、毒性蛋白等。作为备选地,治疗性表达产物可以是RNA,例如siRNA或miRNA。

[0190] 在一些优选实施方案中,治疗性表达产物对于已经引入基因治疗载体的细胞可以是具有毒性的。在这样的实施方案中,UPR的诱导可用于诱导细胞中的毒性,例如导致细胞死亡。因此,UPR的诱导和治疗性表达产物的随后表达可用于诱导所述细胞中的生死开关。合适的毒性表达产物,例如蛋白质,在本领域中是公知的,并且可提及例如胰天蛋白酶3、胰天蛋白酶8和胰天蛋白酶9。

[0191] 上面讨论了用于本发明该方面的合适的基因治疗载体。

[0192] 基因治疗方案已在本领域中广泛描述。这些包括但不限于:肌肉注射合适的载体;在包括肌肉的多种组织中进行流体动力学基因递送;间质内注射;在气道中滴注;应用于内皮、肝内实质;静脉内或动脉内施用。已经开发出多种装置来增强DNA对靶细胞的可利用性。一种简单的方法是使靶细胞与含有DNA的导管或可植入材料物理接触。另一种方法是利用无针喷射注射装置,该装置在高压下将一柱液体直接投射到靶组织中。这些递送范例也可用于递送载体。靶向基因递送的另一种方法是使用分子偶联物,该偶联物由蛋白质或合成配体组成,核酸结合剂或DNA结合剂已附着在该蛋白质或合成配体上,用于将核酸特异性地靶向细胞(Cristiano等,1993)。

[0193] 术语“对象”和“患者”在本文可互换使用,并且是指动物,优选脊椎动物,更优选哺乳动物,并且特别包括人患者和非人哺乳动物。“哺乳动物”对象包括但不限于人。优选的患者或对象是人对象。

[0194] 如本文中所用,“治疗量”或“治疗有效量”是指有效治疗对象的疾病或病症,即获得所需的局部或全身作用的表达产物的量。因此,该术语是指引起研究人员、兽医、医生或其他临床医师寻求的在组织、系统、动物或人中的生物学或医学反应的表达产物的量。这样的量通常取决于基因产物和疾病的严重程度,但是可由技术人员确定,可能地通过常规实验确定。

[0195] 表达产物(例如蛋白质)的表达水平可通过多种常规手段来测量,例如通过基于抗体的测定,例如通过Western印迹或ELISA分析,例如以评价是否达到了表达产物的治疗性表达。表达产物的表达也可在检测基因产物的酶活性或生物学活性的生物测定中测量。

[0196] 在另一方面中,本发明提供了合成UPR诱导型启动子,其包含合成UPR响应性顺式调节元件。合适地,UPR响应性顺式调节元件包含针对以下的至少一个结合位点:作为UPR的一部分驱动基因表达的ATF6、XBP1或bZIP60,或者同源或在其他方面等效的转录因子。

[0197] 这样的启动子可用于在UPR诱导之后选择性驱动所期望的表达产物在真核细胞中的表达。尽管就严密控制基因表达而言,将这种启动子与上文讨论的可调节内含子组合使用可以是有利的,但在其他情况下,可在没有可调节内含子的情况下使用UPR诱导型启动子。

[0198] UPR响应性顺式调节元件是这样的序列,其包含针对一个或更多个转录因子的功

能性转录因子结合位点(transcription factor binding site,TFBS),所述转录因子作为UPR的一部分驱动基因表达。如上所述,这些包括但不限于ATF6、XBP1和bZIP60。包含针对ATF6的TFBS的UPR响应性顺式调节元件在本发明中特别令人关注。优选地,UPR响应性顺式调节元件是UPR特异性的,即,其仅在UPR期间增强表达。例如,通常优选地,对于不参与UPR的转录因子,它不包含任何TFBS。

[0199] 合成UPR诱导型启动子通常包含至少一个与最小启动子或近端启动子有效连接的合成UPR响应性顺式调节元件。当顺式调节元件与近端启动子有效连接时,近端启动子本身应为UPR诱导型启动子,当未诱导UPR时,该启动子不会驱动真核细胞中有效连接的基因的转录。在没有其他调节元件的情况下,最小启动子通常无法驱动表达。用于本发明的合适的最小启动子的一些实例包括但不限于CMV最小(SEQ ID NO:21)启动子和MinTk最小启动子。其他合适的最小启动子是本领域已知的。

[0200] 合适地,UPR诱导型启动子包含UPR响应性顺式调节元件,其包含以下转录因子靶序列中的至少一种的一个或更多个拷贝:

[0201] -TGACGTG(ATF6转录因子结合位点共有序列),

[0202] -TGACGTGCT(以上的变体)

[0203] -TGACGTG[TG](称为UPRE位点),

[0204] -CCAAT-N9-CCACG(称为ERSE1位点)(SEQ ID NO:18),和

[0205] -ATTGG-N-CCACG(称为ERSE2位点)(SEQ ID NO:19)。

[0206] 合适地,合成UPR响应性顺式调节元件包含以上列出的至少一个转录因子靶序列的两个或更多个拷贝,优选三个或更多个拷贝,合适地五个或更多个拷贝。作为替选地或另外地,合成UPR响应性顺式调节元件合适地包含以上列出的至少两个转录因子靶序列的一个或更多个拷贝。

[0207] 转录因子靶序列可以是彼此直接相邻的(串联重复序列)或可例如通过间隔序列或另外的功能性序列(例如另外的转录因子靶序列)间隔开。通常,间隔序列(如果存在的话)的长度为5至50个核苷酸,但在某些情况下可更长或更短。例如,间隔序列合适地长度为2至50个核苷酸,合适地长度为4至30个核苷酸,或合适地长度为5至20个核苷酸。其可优选长度为5个核苷酸的倍数,因为这提供了整数个半圈的DNA双螺旋(完整一圈对应于染色质中的约10个核苷酸)。间隔序列的长度为10个核苷酸的倍数可以是更优选的,因为这提供了整数个完整圈的DNA双螺旋。间隔序列可具有基本上任何序列,只要它不阻止UPR响应性顺式调节元件如所期望的发挥作用(例如,其包含沉默子序列,阻止所期望的转录因子的结合等)。每个转录因子靶序列之间的间隔序列可以是相同的或者它们可以是不同的。

[0208] 在一个优选实施方案中,UPR响应性顺式调节元件包含转录因子靶序列TGACGTG(即ATF6共有序列)的一个或更多个拷贝,优选转录因子靶序列TGACGTG的3个或更多个拷贝,优选转录因子靶序列TGACGTG的5个或更多个拷贝,例如转录因子靶序列TGACGTG的6个或更多个拷贝。如上所提及的,这些转录因子靶序列可串联重复或可彼此间隔开。通常,优选存在于UPR响应性顺式调节元件中的至少两个,并且优选全部的转录因子靶序列彼此间隔,例如通过如上所讨论的间隔序列。

[0209] 合适地,UPR响应性顺式调节元件包含转录因子靶序列TGACGTGCT的一个或更多个拷贝,优选转录因子靶序列TGACGTGCT的3个或更多个拷贝,优选转录因子靶序列TGACGTGCT

的5个或更多个拷贝,例如转录因子靶序列TGACGTGCT的6个或更多个拷贝。如上所提及的,它们可串联重复,或者可彼此间隔开。通常,优选存在于UPR响应性顺式调节元件中的至少两个,并且优选全部的转录因子靶序列彼此间隔,例如通过如上所讨论的间隔序列。当以UPR响应顺式调节元件中的多拷贝数使用时,无论是作为串联重复序列还是包括间隔序列,转录因子靶序列TGACGTGCT都被发现是特别有效的。

[0210] 在本发明的一些实施方案中,UPR响应性顺式调节元件包含序列

[0211] TGACGTG-S-TGACGTG-S-TGACGTG-S-TGACGTG-S-TGACGTG-S-TGACGTG (SEQ ID NO: 54)

[0212] 其中S代表如上定义的任意的间隔序列。优选地,如上所定义的间隔序列存在于至少两个并且优选地所有转录因子靶序列(TGACGTG)之间。

[0213] 在本发明的一些实施方案中,UPR响应性顺式调节元件包含序列

[0214] TGACGTGCTTGACGTGCTTGACGTGCTTGACGTGCTTGACGTGCTTGACGIGCT (SEQ ID NO: 20)

[0215] 在本发明的另一些实施方案中,UPR响应性顺式调节元件包含序列

[0216] TGACGTGCT-S-TGACGTGCT-S-TGACGTGCT-S-TGACGTGCT-S-TGACGTGCT-S-TGACGTGCT-S-TGACGTGCT (SEO ID NO: 55)

[0217] 其中S代表如上定义的任意的间隔序列。优选地,如上所定义的间隔序列存在于至少两个,优选所有的转录因子靶序列之间。

[0218] 在本发明的另一些实施方案中,UPR响应性顺式调节元件包含序列

[0219] TGACGTGCTGATGATGCGTAGCTAGTAGTTGACGTGCTGATGATGCGTAGCTAGTAGTTGACGTGCTG
ATGATGCGTAGCTAGTAGTTGACGTGCTGATGATGCGTAGCTAGTAGTTGACGTGCTGATGATGCGTA
GCTAGTAGTTGACGTGCT (SEQ ID NO: 56)

[0220] 或与其至少50%相同,更优选与其至少70%相同,更优选与其至少80%相同,更优选与其至少85%、90%、95%、98%或99%相同的序列。高度优选仅在不是转录因子靶序列(即具有序列TGACGTGCT的那些)的序列中发生序列变异。通常优选仅在间隔序列(即具有序列GATGATGCGTAGCTAGTAGT (SEQ ID NO:61)的那些)中发生序列变异。

[0221] 在本发明的一些实施方案中,UPR诱导型启动子包含以下序列

[0222] TGACGTGCTTGACGTGCTTGACGTGCTTGACGTGCTTGACGTGCTTGACGTGCTGGTACCGTCGACGA
TATCGGATCCAGGTCTATATAAGCAGAGCTCGTTTAGTGAACCGTCAGATCGCCTAGATACGCCATCC
ACGCTGTTTTGACCTCCATAGAAGATCGCCACC (SEQ ID NO: 22)

[0223] 或与其至少70%相同,更优选与其至少80%相同,更优选与其至少90%相同,更优选与其至少95%、96%、97%、98%或99%相同的序列。该UPR诱导型启动子包含与CMV-MP最小启动子有效连接的SEQ ID NO:20的UPR响应性顺式调节元件。高度优选仅在不是转录因子靶序列(即具有序列TGACGTGCT的那些)的序列中发生序列变异,也不在CMV-MP序列中发生序列变异。

[0224] 在本发明的其它实施方案中UPR诱导型启动子包含以下序列:

[0225] TGACGTGCTGATGATGCGTAGCTAGTAGTTGACGTGCTGATGATGCGTAGCTAGTAGTTGACGTGCTG
ATGATGCGTAGCTAGTAGTTGACGTGCTGATGATGCGTAGCTAGTAGTTGACGTGCTGATGATGCGTA
GCTAGTAGTTGACGTGCTGATGATGCGTAGCTAGTAGTGAGTACCGTACCGTACCGTACCGTACCGT
ATATCGGATCCAGGTCTATATAAGCAGAGCTCGTTAGTGAAACCGTCAGAT (SEQ ID NO: 57)

[0226] 或与其至少50%相同,更优选与其至少70%相同,更优选与其至少80%相同,更优选与其至少85%、90%、995%、98%或99%相同的序列。该UPR诱导型启动子包含与CMV-MP最小启动子有效连接的SEQ ID NO:56的UPR响应性顺式调节元件。高度优选仅在不是转录因子靶序列(即具有序列TGACGTGCT的那些)的序列中发生序列变异,也不在CMV-MP序列中发生序列变异。通常优选仅在间隔序列(即具有序列GATGATGCGTAGCTAGTAGT (SEQ ID NO: 61)的那些)中发生序列变异。

[0227] 在本发明的另一些实施方案中,UPR诱导型启动子包含以下序列:

TGACGTGCTGATGATGCGTAGCTAGTAGTTGACGTGCTGATGATGCGTAGCTAGTAGTTGACGTGCTG
ATGATGCGTAGCTAGTAGTTGACGTGCTGATGATGCGTAGCTAGTAGTTGACGTGCTGATGATGCGTA
[0228] GCTAGTAGTTGACGTGCTGATGATGCGTAGCTAGTAGTGCAGTTAGCGTAGCTGAGGTACCGTCGACG
ATATCGGATCCTTCGCATATTAAGGTGACGCGTGTGGCCTCGAACACCGAG (SEQ ID NO: 58)

[0229] 或与其至少50%相同,更优选与其至少70%相同,更优选与其至少80%相同,更优选与其至少85%、90%、995%、98%或99%相同的序列。该UPR诱导型启动子包含与MinTK最小启动子有效连接的SEQ ID NO:56的UPR响应顺式调节元件。高度优选仅在不是转录因子靶序列(即具有序列TGACGTGCT的那些)的序列中发生序列变异,也不在MinTK序列中发生序列变异。通常优选仅在间隔序列(即具有序列GATGATGCGTAGCTAGTAGT (SEQ ID NO:61)的那些)中发生序列变异。

[0230] 在不存在UPR的情况下,当UPR诱导型启动子存在于真核细胞中时,其优选不驱动有效连接的基因的转录。当UPR在真核细胞中发生时,UPR诱导型启动子在存在于真核细胞中时驱动有效连接的基因的转录。技术人员可使用广泛范围的方法容易地评估在UPR的诱导下UPR诱导型启动子选择性地驱动转录的能力的评估,并且这些方法可针对其中旨在使用构建体的特定表达体系进行定制。作为一个优选的实例,可使用以下实施例(例如实施例8)中描述的方法。例如,可将任何待评估的候选UPR诱导型启动子替换为实施例8中所述的构建体,以代替实施例8中使用的示例性的含有ATF6的UPR诱导型启动子,并且可通过评估2mM DTT诱导UPR前后萤光素酶表达的水平,来测量所述候选UPR诱导型启动子在诱导UPR时选择性驱动转录的能力,这与实施例8中进行的完全相同。UPR诱导型启动子是能够在UPR的诱导之后成功诱导有效连接的基因(在实施例8的情况下,萤光素酶基因)的转录显著提高,从而导致该基因的表达。优选地,UPR诱导型启动子在用2mM DTT诱导UPR之后24小时使表达提高至少5倍,更优选使表达提高至少10倍,更优选使表达提高至少100倍,更优选基因(例如萤光素酶)的表达提高至少1000倍。优选的是,在诱导UPR之前,基因(例如萤光素酶)的表达水平是最低的,并且优选地可忽略不计。最低表达可定义为例如等于或小于实施例8中所用的阴性对照构建体(即,没有的构建体,其中编码萤光素酶的序列的表达仅由CMV-MP驱动)的表达水平,优选小于阴性对照构建体表达水平的50%,优选小于20%,更优选小于10%,更优选小于5%,更优选小于1%。可忽略不计的表达水平是,例如,使用实施例8的方法基本上不可检测的表达水平。

[0231] 本发明还提供了表达构建体或载体,其包含与编码表达产物的核酸序列有效连接的如上所述的合成UPR诱导型启动子。表达构建体或载体可以是如上文针对本发明的其他方面所讨论的任何表达构建体或载体。表达产物可以是如上文针对本发明的其他方面所讨论的任何表达产物(例如,编码蛋白质)。

[0232] 在一些优选实施方案中,表达产物不是报道蛋白,即它不编码常规用作表达水平的指标的蛋白质。许多报道基因是本领域已知的,特别地包含荧光蛋白、发光蛋白和显色蛋白。因此,在一些优选实施方案中,表达产物不是荧光蛋白或发光蛋白,例如它不是萤光素酶。如上所述,优选的表达产物包括治疗性蛋白质和毒性蛋白。

[0233] 在另一方面中,本发明提供了用于产生表达产物的方法,所述方法包括:

[0234] a) 提供包含合成核酸表达构建体的真核细胞群,所述合成核酸表达构建体包含与编码根据本发明的表达产物的核酸序列有效连接的UPR诱导型启动子;

[0235] b) 处理所述细胞群以诱导未折叠蛋白反应,从而诱导从UPR诱导型启动子的转录;

[0236] c) 在合适的条件下孵育所述细胞群以产生表达产物;以及

[0237] d) 从所述细胞群中分离表达产物。

[0238] 上文针对本发明的其他方面讨论了用于产生表达产物的方法的其他任选和优选特征,并且这些特征在细节上作必要修改后也适用于本发明。优选表达产物是治疗性蛋白质或毒性蛋白。优选表达产物不是报道蛋白

[0239] 因此,本发明的另一方面提供了药物组合物,其包含核酸表达构建体或载体,所述核酸表达构建体或载体包含与编码根据本发明的表达产物的核酸序列有效连接的UPR诱导型启动子。上文针对本发明的其他方面讨论了药物组合物的其他任选和优选特征,这些特征在细节上作必要修改后也适用于本发明。

[0240] 在本发明的另一方面中,提供了包含与编码根据本发明的表达产物的核酸序列有效连接的UPR诱导型启动子的核酸表达构建体和载体用于制备药物组合物的用途。

[0241] 根据本发明的另一方面,提供了细胞,其包含含有根据本发明的UPR诱导型启动子的合成核酸表达构建体或载体。上面针对本发明的其他方面讨论了这种细胞的其他任选和优选特征,这些特征在细节上作必要修改后也适用于本发明。

[0242] 在另一方面中,本发明提供了包含根据本发明的UPR诱导型启动子的核酸表达构建体、载体、细胞或药物组合物,其用于治疗或治疗。上面针对本发明的其他方面讨论了这种方法的其他任选和优选特征,这些特征在细节上作必要修改后也适用于本发明。

[0243] 为了促进对本发明的理解,下面定义了许多术语。本文所定义的术语具有与和本发明有关的领域中的普通技术人员通常理解的含义。本文中的术语用于描述本发明的特定实施方案,但是它们的用途并不限制本发明,除非权利要求书中阐明。

[0244] 本文包括对本发明背景的讨论,以解释本发明的上下文。这不应被视为承认所提及的任何材料自任何权利要求的优先权日起已在任何国家发布、已知或作为共同的一般知识的一部分。

[0245] 在整个本公开内容中,通过标识引文引用了多种出版物、专利和公开的专利说明书。本说明书中引用的所有文件均通过引用整体并入本文。特别地,本文中具体提及的这样的文件的教导或部分通过引用并入。

[0246] 除非另有说明,否则本发明的实施将采用细胞生物学、细胞培养、分子生物学、转基因生物学、微生物学、重组DNA和免疫学的常规技术,其在本领域技术范围内。文献中对这种技术进行了充分的解释。参见,例如,Current Protocols in Molecular Biology (Ausubel,2000,Wiley and son Inc,Library of Congress,USA);Molecular Cloning:A Laboratory Manual,Third Edition,(Sambrook等,2001,Cold Spring Harbor,New York:

Cold Spring Harbor Laboratory Press); Oligonucleotide Synthesis (M.J.Gait ed., 1984); U.S.Pat.No.4,683,195; Nucleic Acid Hybridization (Harries和Higgins eds.1984); Transcription and Translation (Hames和Higgins eds.1984); Culture of Animal Cells (Freshney, Alan R.Liss, Inc., 1987); Immobilized Cells and Enzymes (IRL Press, 1986); Perbal, A Practical Guide to Molecular Cloning (1984); the series, Methods in Enzymology (Abelson和Simon, eds.-in-chief, Academic Press, Inc., New York), specifically, Vols.154和155 (Wu等eds.) 和Vol.185, "Gene Expression Technology" (Goeddel, ed.); Gene Transfer Vectors For Mammalian Cells (Miller和Calos eds., 1987, Cold Spring Harbor Laboratory); Immunochemical Methods in Cell and Molecular Biology (Mayer和Walker, eds., Academic Press, London, 1987); Handbook of Experimental Immunology, Vols.I-IV (Weir和Blackwell, eds., 1986); 和 Manipulating the Mouse Embryo, (Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y., 1986)。

[0247] 没有数量词修饰的名词用于指代“一个/种或更多个/种”。

[0248] 如本文中所用, 术语“包含”或其变体与“包括”及其变体或“含有”及其变体同义, 并且是包括性的或开放式的, 并且不排除其他的, 非叙述性的特征、要素或方法步骤。

[0249] 通过端点对数值范围的叙述包括相应范围内包括的所有数字和分数, 以及所叙述的端点。

[0250] 本文中所用的术语“核酸”通常是指基本上由核苷酸组成的任何长度的低聚物或聚合物 (优选线性聚合物)。核苷酸单元通常包括杂环碱基、糖基团和至少一个 (例如一个、两个或三个) 磷酸基团, 包括经修饰或取代的磷酸基团。杂环碱基可尤其包含嘌呤和嘧啶碱基, 例如腺嘌呤 (A)、鸟嘌呤 (G)、胞嘧啶 (C)、胸腺嘧啶 (T) 和尿嘧啶 (U), 它们广泛存在于天然存在的核酸、其他天然存在的碱 (例如黄嘌呤, 肌苷、次黄嘌呤) 以及化学或生物化学修饰的 (例如甲基化的)、非天然或衍生化的碱中。糖基团可尤其包括戊糖 (戊呋喃糖) 基团, 例如优选在天然存在的核酸中常见的核糖和/或2-脱氧核糖, 或阿拉伯糖、2-脱氧阿拉伯糖、蔗糖或己糖糖基, 以及经修饰或取代的糖基团。如本文中预期的, 核酸可包括天然存在的核苷酸、经修饰的核苷酸或其混合物。经修饰的核苷酸可包括经修饰的杂环碱基、经修饰的糖部分、经修饰的磷酸基团或其组合。可引入对磷酸基团或糖的修饰以提高稳定性、对酶促降解的抵抗力或另一些有用的性质。术语“核酸”还优选涵盖DNA、RNA和DNA/RNA杂交分子, 具体包括hnRNA、前mRNA、mRNA、cDNA、基因组DNA、扩增产物、寡核苷酸, 以及合成的 (例如化学合成的) DNA、RNA或DNA/RNA杂交体。核酸可以是天然存在的, 例如存在于自然中或从自然分离; 或可以是非天然存在的, 例如重组的, 即通过重组DNA技术产生的, 和/或部分或全部通过化学或生物化学合成的。“核酸”可以是双链的、部分双链的或单链的。在单链的情况下, 核酸可以有义链或反义链。另外, 核酸可以是环形或线性的。

[0251] 如本文中所用, “核酸表达构建体”是指包含在一个或更多个所期望的细胞类型、组织或器官中指导表达的一种或更多种转录控制元件 (例如但不限于启动子、增强子和/或调节元件、多腺苷酸化序列和内含子) 的核酸分子。本发明的核酸表达构建体是合成核酸分子。

[0252] 在本申请中, “合成的”意指自然界中不存在的核酸分子。本发明的合成核酸表达

构建体通常通过重组技术人工产生。这样的合成核酸可包含天然存在的序列(例如启动子、增强子、内含子和其他这样的调节序列),但是它们在非天然存在的背景下存在。例如,合成基因(或基因的一部分)通常包含一个或更多个本质上不连续的核酸序列(嵌合序列),和/或可包括替换、插入和缺失及其组合。

[0253] 如本文中所用,术语“有效连接”或等同表达是指多种核酸元件相对于彼此的排列,使得所述元件在功能上联系并且能够以预期的方式彼此相互作用。这样的元件可包括但不限于启动子、增强子和/或调节元件、多腺苷酸化序列、一个或更多个内含子和/或外显子,以及待表达的目的基因的编码序列。核酸序列元件在正确定向或有效连接时一起发挥作用以调节彼此的活性,并最终可影响表达产物的表达水平。调节意指提高、降低或维持特定元件的活性水平。每个元件相对于其他元件的位置可用每个元件的5'末端和3'末端来表示,并且任何特定元件之间的距离可由所述元件之间的中间核苷酸或碱基对的数目来参考。如本领域技术人员所理解的,有效连接意味着功能活性,并且不一定与天然的位置联系有关。实际上,当用于核酸表达组件中时,顺式调节元件通常将位于紧接着启动子的上游(尽管通常是这种情况,但绝对不应将其解释为对核酸表达组件内位置的限制或排除),但体内情况并不一定如此。例如,天然存在于基因下游的调节元件序列(所述基因的转录受到所述序列的影响),当位于启动子的上游时能够以相同的方式发挥作用。因此,根据一个具体实施方案,调节元件的调节或增强作用是不依赖于位置的。

[0254] “共有序列”-共有序列的含义在本领域中是公知的。在本申请中,除非上下文另有指示,否则以下符号用于共有序列。考虑以下示例性DNA序列:

[0255] A[CT]N{A}YR

[0256] A意指始终在该位置发现A;[CT]代表该位置的C或T;N代表该位置的任何碱基;{A}意指在该位置发现除A以外的任何碱基。Y代表任何嘧啶,R代表任何嘌呤。

[0257] 除非上下文另外指出,否则本申请中的“可调节内含子”是指存在于RNA分子(通常是转录物)中的核酸序列,该序列包含侧翼为核糖核酸酶(通常是IRE1或同源物或直系同源物)的靶位点的可切除序列,其中由于未折叠蛋白反应,可通过所述核糖核酸酶的作用从RNA分子上切除可切除序列。在本领域的某些情况下,术语内含子仅用于指代从RNA分子切除(即剪接出)的序列,但是在本发明的可调节内含子的情况下不太合适。

[0258] 术语“同一性”和“相同”等是指两个聚合物分子之间,例如两个核酸分子(例如两个DNA分子)之间的序列相似性。序列比对和序列同一性的确定可例如使用由Altschul等1990(J Mol Biol 215:403-10)最初描述的局部序列排比检索基本工具(Basic Local Alignment Search Tool,BLAST)来完成,例如Tatusova和Madden 1999(FEMS Microbiol Lett 174:247-250)所描述的“Blast 2序列”算法。

[0259] 用于比较的序列比对方法是本领域公知的。多种程序和比对算法在例如以下中描述:Smith和Waterman(1981)Adv. Appl. Math. 2:482;Needleman和Wunsch(1970)J. Mol. Biol. 48:443;Pearson和Lipman(1988)Proc. Natl. Acad. Sci. U.S.A. 85:2444;Higgins和Sharp(1988)Gene 73:237-44;Higgins和Sharp(1989)CABIOS 5:151-3;Corpet等(1988)Nucleic Acids Res. 16:10881-90;Huang等(1992)Comp. Appl. Biosci. 8:155-65;Pearson等(1994)Methods Mol. Biol. 24:307-31;Tatiana等(1999)FEMS Microbiol. Lett. 174:247-50。序列比对方法和同源性计算的详细考虑可见于例如

Altschul等(1990)J.Mol.Biol.215:403-10中。

[0260] 美国国家生物技术信息中心(National Center for Biotechnology Information,NCBI)局部序列排比检索基本工具(BLAST™;Altschul等(1990))可从数种来源(包括美国国家生物技术信息中心(Bethesda,MD))以及在互联网上获得,与数个序列分析程序结合使用。互联网上BLAST™的“帮助”部分提供了如何使用此程序确定序列同一性的描述。为了比较核酸序列,可使用默认参数使用BLAST™(Blastn)程序的“Blast 2序列”功能。当通过这种方法评估时,与参考序列具有更大相似性的核酸序列将显示出提高的同一性百分比。通常,序列同一性百分比是在序列的整个长度上计算的。

[0261] 例如,通过Needleman-Wunsch算法合适地找到具有以下得分参数的全局最优比对:匹配得分:+2,不匹配得分:-3;空位罚分:空位开放5,空位延伸2。通过比对碱基的数量与比对总长度的比,可合适地计算出所得的最佳全局比对的同一性百分比,其中比对长度包括匹配和不匹配二者,乘以100。

[0262] 如本文中所用,“细胞培养物”是指可处于未分化或分化状态的细胞增殖团。

[0263] 如本文中所用,“顺式调节元件”或“CRE”是技术人员已知的术语;它涉及调节邻近基因转录(即顺式)的非编码DNA区域。CRE通常通过与转录因子结合来调节基因转录。CRE可以是例如增强子、启动子、绝缘子或沉默。在当前情况下,UPR诱导型CRE通常是增强子元件,其与转录因子结合,所述转录因子充当UPR的一部分来诱导转录。在本上下文中,当提供CRE作为启动子和编码表达产物的基因的一部分时,优选的是,UPR诱导型CRE位于距转录起始位点(transcription start site,TSS)1500个核苷酸或更少的核苷酸处,更优选距TSS 1000个核苷酸或更少的核苷酸处,更优选地距TSS 500个核苷酸或更少的核苷酸处,并且合适地距TSS 250个、200个、150个或100个核苷酸或更少的核苷酸处。

[0264] 如本文中所用,“互补的”或“互补性”是指两个核酸序列的沃森-克里克碱基配对。例如,对于序列5'-AGT-3',其与互补序列3'-TCA-5'结合。两个核酸序列之间的互补性可以是“部分的”,其中只有一些碱基与其互补序列结合,或者当序列中的每个碱基与其互补碱基结合时,它可以是完整的。核酸链之间的互补程度对核酸链之间的杂交效率和强度具有重要影响。

[0265] 如本文中所用,“常规内含子”是指通常存在于前mRNA中的内含子,通常是剪接体内含子,其不编码蛋白质合成信息,并且在mRNA翻译之前从mRNA分子中除去。该术语用于将这样的内含子与本发明的非规范调节内含子区分开。

[0266] 本申请中的“转染”广泛地是指将核酸有意引入到细胞内的任何过程,并且涵盖了病毒载体和非病毒载体的引入,并且包括转化、转导和类似术语。实例包括但不限于:用病毒载体转染;用质粒载体转化;电穿孔(Fromm等(1986)Nature 319:791-3);脂质体转染(Feigner等(1987)Proc.Natl.Acad.Sci.USA 84:7413-7);显微注射(Mueller等(1978)Cell 15:579-85);农杆菌介导的转移(Fraley等(1983)Proc.Natl.Acad.Sci.USA80:4803-7);直接DNA摄取;颈须介导的转化(whiskers-mediated transformation);和基因枪法(microprojectile bombardment)(Klein等(1987)Nature 327:70)。

[0267] 如本文中所用,短语“转基因”是指外源核酸序列。在一个实例中,转基因是基因序列,编码工业上或药学上有用的化合物的基因或编码期望性状的基因。在另一个实例中,转基因是反义核酸序列,其中反义核酸序列的表达抑制靶核酸序列的表达。

[0268] 如本文中所用,短语“启动子”是指通常位于待转录核酸序列上游的DNA区域,其是转录发生所需的。启动子允许对在其控制下的序列的转录进行合适的激活或抑制。启动子通常包含被转录因子识别和结合的特定序列,例如增强子序列。转录因子与启动子DNA序列结合,导致RNA聚合酶的募集,RNA聚合酶是一种从基因编码区合成RNA的酶。许多启动子是本领域已知的。

[0269] 如本文中所用,“最小启动子”是指短的DNA片段,其自身是无活性的或大部分是无活性的,但是当与其他转录调节元件组合时可介导强转录。最小启动子序列可来源于多种不同的来源,包括原核生物基因和真核生物基因。最小启动子的实例是多巴胺 β -羟化酶基因最小启动子和巨细胞病毒(cytomegalovirus,CMV)即早期基因最小启动子(CMV-MP)以及疱疹胸苷激酶最小启动子(MinTK)。

[0270] “RNA转录物”或“转录物”是指由RNA聚合酶催化的DNA序列转录产生的产物。当RNA转录物是DNA序列的典型的完美互补拷贝时,它被称为初级转录物,或者可以是来源于初级转录物的转录后加工的RNA序列,并且被称为成熟RNA。

[0271] “信使RNA”或“(mRNA)”是指没有内含子并且可被细胞翻译成蛋白质的转录RNA的加工形式。

[0272] 现在将参照附图通过非限制性实例描述本发明的实施方案。

附图说明

[0273] 图1示出了在通过2mM DTT诱导前后,来自SYNP-XBP-01转染的细胞的GFP荧光的比。

[0274] 图2在图a)中示出了诱导前后GFP荧光阳性对照的比。在图b)中,示出了在具有和不具有2mM DTT的情况下的SYNP-XBP-01的荧光显微术。

[0275] 图3示出了在通过2mM DTT诱导之后两种质粒中GFP的表达。数字是与CMV-IE的比。

[0276] 图4示出了在添加2mM DTT前后来自SYNP-ATF6-01的表达的原始数据。

[0277] 图5示出了在添加2mM DTT前后来自SYNP-ATF6-02的表达的原始数据。

[0278] 图6示出了在用2mM DTT处理前后用SYNP-ATF6-01或SYNP-ATF6-02转染的HEK293F细胞的荧光显微术。

[0279] 图7示出了用wtAAV诱导前后来自HEK293F细胞的EGFP荧光。Zs Green是重组AAV,可产生GFP的变体,其用作对照,pre-AS-Green是无转染(即空白)。

[0280] 图8示出了在wtAAV产生前后来自HEK293F细胞的EGFP表达。

[0281] 图9示出了以相对于CMV-IE的比测量的EGFP表达。使用的构建体为SYNP-ATF6-02;其包含内含子和ATF6诱导型启动子。使用Maxcyte试剂对HEK293F细胞进行转染。转染之后24小时,添加诱导物,并在接下来的24小时追踪GFP表达。诱导物浓度:DTT 2mM,棕榈酸0.5mM,MF-43 1 μ M。

[0282] 图10示出了pMA-RQ的质粒图。

[0283] 图11示出了在CMV-IE启动子控制下,与SEAP的表达相比,在CHO和HEK293细胞中用DTT诱导之后的SEAP的表达图(示出的结果为HEK293中)。(X轴示出了诱导之后的小时数)

[0284] 图12示出了在CHO (A) 和HEK293 (B) 细胞中用DTT或棕榈酸诱导之后萤光素酶的表达图(X轴单位为小时)。为了比较,示出了在CMV-IE启动子和CMV最小启动子(CMV-MP)的控

制下荧光素酶的表达。

[0285] 图13示出了SYNP-CAS9_INT的质粒图。

[0286] 图14A示出了在CMV-IE启动子控制下,与阴性对照和CASP9相比,用DTT或毛喉素诱导胱天蛋白酶9(CASP9)之后HEK293细胞生存力的图(X轴显示诱导之后的小时数)。图14B示出了在表达构建体不含可调节内含子的细胞(左图)和在表达构建体含有可调节内含子的细胞(右图)中与SASP9融合的荧光标志物Spark的显微图。

[0287] 图15示出了在CMV-IE启动子控制下,与阴性对照(HEK293细胞加DTT)和CASP9相比,用DTT诱导胱天蛋白酶9之后HEK293细胞生存力的图。

[0288] 图16A至D示出了将内含子用于以下情况时形成的预测二级结构:

[0289] A. SEAP (CAG/CAGACGGGCAACTTTACACGACGCTG/CAG)

[0290] B. EGFP (CAG/CTGGAGCACTCAGACTACGTGCACCTCTG/CTG)

[0291] C. CASP9 (CAG/CAGACGGGCAACTTTACACGACGCTG/CTG)

[0292] D. 荧光素酶 (CCG/CAGACGGGCAACTTTACACGACGCTG/CAG)

[0293] 图17示出了当含有5x ATF6结合位点的UPR响应顺式调节元件与两个不同的最小启动子组合时,用DDT诱导之后荧光素酶的表达图(X轴单位是诱导之后的小时数)。对照是没有任何其他调节序列的CMV-IE启动子和CMV最小启动子。

具体实施方式

[0294] 实施例1:使用XBP1内含子作为蛋白质表达的调节机制。

[0295] XBP1的mRNA中存在一个26bp的非常规内含子(可调节),该内含子在剪接之前编码无意义的蛋白质。一旦通过IRE1蛋白检测到内质网(ER)应激,就会通过独特的剪接位点去除该非常规内含子,随后将mRNA重新连接以形成可翻译产生XBP1蛋白的mRNA转录物。ER应激可以多种方式诱导,例如通过化学添加或通过表达异源蛋白。

[0296] 发明人已经认识到,通过在目的基因的序列中包含可调节内含子、包含合适的剪接位点,该过程可适于在mRNA/翻译水平上调节蛋白质表达。在合适的启动子的控制下,产生了mRNA,但是不能对其进行加工以形成目的基因的功能性表达产物。然而,一旦诱导了UPR,例如通过施用DTT或表达另一种异源蛋白,mRNA被加工,剪接出可调节内含子,并可将mRNA翻译成功能性蛋白质。

[0297] 通过使用这种机制,可调整蛋白质(例如毒性蛋白)的产生以适合制造过程,因为在产生蛋白质之前需要进行诱导。因此,例如,可延迟蛋白质的翻译,直到达到发酵中的所需阶段。该方法也可用于非蛋白质表达产物,例如功能性RNA(例如siRNA或miRNA)。

[0298] 剪接位点和内含子

[0299] IRE1的共有剪接识别位点序列:

[0300] CNG/CNG

[0301] 该共有序列在整个真核生物中是保守的。考虑到可使用CNG/CNG共有序列,共有序列CNG/CNG[CG],优选CNG/CNGC通常见于哺乳动物细胞中。

[0302] WT哺乳动物XBP1剪接识别位点序列:

[0303] 5' 位点:CCG/CAGC

[0304] 3' 位点:CUG/CAGC

[0305] 来自哺乳动物的WT内含子序列:

[0306] 切除的内含子序列 (SEQ ID NO:23):

[0307] **cagcacucagacuacgugcaccucug**

[0308] WT哺乳动物内含子,其包含侧翼为剪接识别位点序列的切除序列:

[0309] **CCG/cagcacucagacuacgugcaccucug/CAGC (SEQ ID NO: 24)**

[0310] 包含可调节内含子的EGFP序列的构建

[0311] 选择用于插入内含子的区域是由于与剪接识别位点的序列相似性。在3' 识别位点只需单个沉默突变,即CTCG到CTGG,就可插入内含子。

[0312] 具有用于内含子插入的带下划线区域的EGFP基因序列 (SEQ ID NO:25):

ATGGTGAGCAAGGGCGAGGAGCTCTTCACCGGGTGGTGCCCATCCTGGTGGAGCTGGACG
GCGACGTAAACGGCCACAAGTTCAGCGTGTCCGGCGAGGGCGAGGGCGATGCCACCTACGG
CAAGCTGACCCTGAAGTTCATCTGCACCAACGGCAAGCTGCCCGTGGCCCTGGCCCACCCTC
GTGACCACCTGACCTACGGCGTGCAGTGTCTCAGCCGCTACCCCGACCATGAAGCAGC
ACGACTTCTTCAAGTCCGCCATGCCCGAAGGCTACGTCCAGGAGCGCACCATCTTCTTCAA
GGACGACGGCAACTACAAGACCCGCGCCGAGGTGAAGTTGAGGGCGACACCTGGTGAAC
CGCATCGAGCTGAAGGCCATCGACTTCAAGGAGGACGGCAACATCCTGGGGCCACAAGCTGG
AGTACAACACTACAACAGCCACAACGTCATATATCATGGCCGACAAGCAGAAGAACGGCATCAA
GGTGAACCTTCAAGATCCGCCACAACATCGAGGACGGCAGCGTGCAG/CTCGCCGACCACTA
CCAGCAGAACACCCCATCGGCGACGGCCCGTGCTGCTGCCCGACAACCACTACCTGAGC
ACCCAGTCCGCCCTGAGCAAAGACCCCAACGAGAAGCGCGATCACATGGTCCTGCTGGAST
TCTGTACCGCCGCCGGGATCACTCTCGGCATGGACGAGCTGTACAAG

[0314] EGFP蛋白序列 (SEQ ID NO:26):

MVSKGEELFT GVPILVELD GDVNGHKFSV SGEGEDATY GKLTCLKFICT TGKLPVPWPT
LVTTLTLYGVQ CFSRYPDHMK QHDFFKSAMP EGYVQERTIF FKDDGNYKTR AEVKFEGDTL
VNRIELKGID FKEDGNILGH KLEYNYNSHN VYIMADKQKN GIKVNFKIRH NIEDGSVQLA
DHYQQNTPIG DGPVLLPDNH YLSTQSALSK DPNEKRDHNV LLEFVTAAGI TLGMDELYK

[0316] 本研究中使用的序列如下所示。GFP带有下划线,内含子和剪接位点以粗体显示。在本实施例中,向内含子的5' 端添加了另外的3bp的序列,以产生预期具有稍微次优剪接的内含子。目的是将背景表达水平,即尚未诱导UPR时的水平保持在最低。已知悬液适应的细胞通常比贴壁细胞处于稍高的应激状态,因此,这种轻微修饰在有望避免背景表达中具有益处。当然可使用野生型内含子,并且在某些情况下可以是优选的。本实施例中使用的内含子序列如下(另外的3bp序列带有下划线):

[0317] **CAG/cugcagcacucagacuacgugcaccucug/CUGG (SEQ ID NO: 27)**

[0318] 具有可调节XBP1内含子,CMV最小启动子(CMV-MP)和SV40 polyA尾的EGFP序列,XBP1内含子序列为粗体,EGFP和内含子编码序列带有下划线(该构建体称为SYNP-XBP-01)(SEQ ID NO:28):

- GGTACCGTCGACGATATCGGATCCAGGTCTATATAAGCAGAGCTCGTTTAGTGAACCGTCAGATCGCC
TAGATACGCCATCCACGCTGTTTTGACCTCCATAGAAGATCGCCACCATGGTGAGCAAGGGCGAGGAG
CTGTTACACGGGGTGGTGCCCATCCTGGTCGAGCTGGACGGCGACGTAACCGGCCACAAGTTCAGCGT
GTCCGGCGAGGGCGAGGGCGATGCCACCTACGGCAAGCTGACCCCTGAAGTTCATCTGCACCACCGGCA
AGCTGCCCCGTGCCCTGGCCACCCCTCGTGACCACCCCTGACCTACGGCGTGCACTGCTTCAGCCGCTAC
CCCGACCACATGAAGCAGCAGACTTCTTCAAGTCCGCCATGCCCGAAGGCTACGTCCAGGAGCGCAC
CATCTTCTTCAAGGACGACGGCAACTACAAGACCCGCGCCGAGGTGAAGTTCGAGGGCGACACCCTGG
TGAACCGCATCGAGCTGAAGGGCATCGACTTCAAGGAGGACGGCAACATCCTGGGGCACAAAGCTGGAG
[0319] TACAACTACAACAGCCACAACGTCTATATCATGGCCGACAAGCAGAAGAACGGCATCAAGGTGAACCT
CAAGATCCGCCACAACATCGAGGACGGCAGCGTGAG/ctgcagcactcagactacgtgcacctctg/
CTGGCCGACCACTACCAGCAGAACACCCCCATCGGCGACGGCCCCGTGCTGCTGCCCGACAACCACTA
CCTGAGCACCCAGTCCGCCCTGAGCAAAGACCCCCAACGAGAAGCGCGATCACATGGTCTGCTGGAGT
TCGTGACCGCCGCCGGGATCACTCTCGGCATGGACGAGCTGTACAAGTGAGACGATCCTTATCGGAIT
TTACCACATTTGTAGAGGTTTTACTTGCTTTAAAAAACCTCCACATCTCCCCCTGAACCTGAAACAT
AAAATGAATGCAATTGTTGTTGTTAACTTGTTTATTGCAGCTIATAATGGTTACAAATAAAGCAATAG
CATCACAAATTTACAAATAAAGCATTTTTTTTCACTGCATTCTAGTTGTGGTTTGTCCAAACTCATCA
ATGTATCTTATCATGTCTGCTCGAAGGCGCGCCCTGGGCTCATGGCCCTCCGCTCACTGCCC
[0320] 可看出,具有如上所述插入的可调节XBP1内含子的EGFP编码截短的蛋白(SEQ ID
NO:29、51、52、53-SEQ ID NO:51、52、53是指在初始终止密码子之后散落(falling)的零碎
序列,考虑到这些不会翻译):
MVSKGEELFTGVVPIIIVELDGDVNGHKFSVSGEGEGDATYGKLTCLKFICTTGKLPVPWFTLVTTLTYG
VQCFSTRYPDHMKQHDFFKSAMPEGYVQERTIFFKDDGNYKTRAEVKFEGLTLVNRIELKGIDFKEDGN
[0321] ILGHKLEYNYNSHNVYIMADKQKNGIKVNFKIRHNIEDGSVQLEHSDYVHLCWPTTTSRTPPSATAPC
CCPTTTTStopAPSPPStopAKTPTRSALTWSCWSSSStopPPPGSLSAWTSCT
[0322] 本实施例中使用的DNA构建体:
[0323] -CMV-MP-GFP:控制EGFP的表达的CMV-MP。
[0324] -SYNP-XBP-01:控制具有XBP1内含子的EGFP的表达的CMV-MP。
[0325] 如上所述,这些构建体是相同的,只是在EGFP编码序列中存在内含子,如上所述。
[0326] 所有这些构建体均由Geneart合成,并提供在质粒pMA-RQ中(质粒图参见图10)。将
含有这些构建体的质粒pMA-RQ直接转染到相关细胞中。
[0327] 本实施例中使用的细胞系:
[0328] -Freestyle HEK293F,Invitrogen目录号:R790-07。
[0329] -Freestyle CHO-S,Invitrogen目录号:R80007。
[0330] HEK293F和CHO-S细胞的生长:
[0331] -按照制造商的说明维持细胞。
[0332] -使用MAX试剂(Invitrogen目录号:16447100)转染细胞。
[0333] 如下针对24孔板修改了用于这些细胞的转染的标准方案:
[0334] -使40ml细胞在250ml排气式锥形烧瓶(Sigma-Aldrich CLS431144)中于37℃,8%
CO₂和100rpm搅拌下生长。按照制造商的说明接种细胞。
[0335] -在转染之前1天,使用血细胞计数器对细胞进行计数,并将其分成500,000个细

胞/ml。

[0336] -在转染当天,将细胞以100万个细胞/ml接种到24孔板中的500 μ l适当培养基中。

[0337] -然后将0.625 μ g DNA/孔添加至10 μ l OptiMem培养基(Thermofisher;11058021),并在室温下孵育5分钟。

[0338] -同时通过添加OptiMem将0.625 μ l的Max试剂补足至10 μ l,并在室温下孵育5分钟。

[0339] -孵育后,将两种混合物添加至同一管中,并在室温下孵育25至30分钟。

[0340] -然后将DNA/Max试剂混合物(20 μ l/孔)直接添加至细胞,并如前所述孵育细胞。

[0341] -然后在24小时之后测量细胞的GFP荧光。

[0342] -通过添加2mM DTT并在1小时之后监测GFP荧光来测量从GFP构建体中的内含子剪接。DTT是一种强还原剂,其通过破坏ER内二硫键的形成来诱导ER应激。

[0343] GFP荧光的测量:

[0344] -通过用无菌水稀释成五分之一,从萤光素酶测定体系(Promega;E1500)准备细胞裂解缓冲液。

[0345] -将细胞以900xg沉淀5分钟,然后以100 μ l/1 $\times$ 10⁶个细胞重悬于来自萤光素酶试剂盒的裂解缓冲液中。

[0346] -然后将其在室温下孵育10至15分钟以进行裂解。

[0347] -以900xg持续5分钟收获细胞核和碎片。

[0348] -收集上清液,将最大100 μ l/样品收集至96孔黑色板。

[0349] -将上清液用PBS稀释成二分之一(这是为了从裂解缓冲液中稀释影响GFP信号的巯基乙醇;在存在还原剂的情况下GFP信号降低)。

[0350] -使样品在室温下孵育5分钟。

[0351] -使用读板器在激发和发射分别设置在485nm和520nm处的情况下测量GFP荧光。

[0352] -GFP荧光也直接在荧光显微镜下可视化。

[0353] 结果

[0354] SYNP-XBP-01实验

[0355] HEK293F实验:

[0356] -如上所述建立24孔转染。每个条件一式两份/一式三份进行。

[0357] -转染之后24小时,将细胞用2mM DTT处理或通过用等体积的水的模拟处理来处理。

[0358] -处理之后1小时,如上所述测量细胞GFP荧光。

[0359] -在诱导前后计算GFP荧光与对照质粒的比。图1示出了3次独立实验的累积结果。

[0360] 图1示出了在通过2mM DTT诱导前后,来自SYNP-XBP-01转染的细胞的GFP荧光的比。结果清楚地表明,在诱导ER应激后,GFP的产生被诱导比阳性对照高4至10倍,所产生的活性比阳性对照高2.5至7倍。另外,背景显著低于对照(虚线),清楚地表明EGFP的表达受到可调节内含子的控制。这表明内含子在位于异源基因(即XBP1以外的基因)中时具有功能,并有助于诱导型基因表达和更高的表达水平。

[0361] CHO-S实验:

[0362] 使用CHO-S细胞的实验以与针对HEK293F细胞所述的相同的方式进行。实验结果可见于图2a和图2b中。

[0363] 图2在图a)中示出了诱导前后GFP荧光阳性对照的比。在图b)中,示出了在具有和不具有2mM DTT的情况下的SYNP-XBP-01的荧光显微术。这些数据表明,在CHO-S细胞中,诱导之前根本没有背景,因此添加2mM DTT后GFP荧光被诱导>1000倍。CHO-S细胞中的控制可比HEK293F细胞中的控制更严密。

[0364] 实施例2:在可调节的XBP1内含子存在下,ATF6响应元件用于增强基因表达和提供精确的基因控制的用途

[0365] ATF6是一种被ER应激激活的转录因子。一旦被激活,该转录因子便与ERSE或UPRE结合,并激活蛋白质稳态的重要组成部分的基因转录(参见Yoshida等,Cell,Vol.107,881-891,December 28,2001)。

[0366] 在这项研究中,发明人研究了在存在和不存在XBP1内含子的情况下ATF6结合位点哺乳动物UPRE的结合和基因表达增强。该UPRE具有TGACGTG的共有序列,并且除ATF6外,还被XBP-1结合,因此基于ER应激产生了潜在强大的基因表达的反馈环(feedback loop)。内含子的添加还允许研究在转录和翻译水平上的控制是否比各自单独赋予更好的诱导性。

[0367] 为进行这项研究,准备了两种构建体:

[0368] 1) 位于CMV最小启动子(CMV-MP)和EGFP上游的6x ATF6元件(具有序列TGACGTGCT);该构建体被命名为SYNP-ATF6-01(参见下文,SEQ ID NO:30)。

[0369] 2) 位于CMV-MP和EGFP上游具有XBP1内含子插入的6x ATF6元件;该构建体被命名为SYNP-ATF6-02(参见下文,SEQ ID NO:31)。

[0370] 如实施例1中所述将内含子插入到EGFP中。

[0371] HEK293F细胞用于这项研究。如实施例1所述进行所有生长条件、转染和分析。

[0372] 本研究中使用的质粒:

[0373] -含有CMV-IE-GFP构建体的pMA-RQ。

[0374] -含有CMV-MP-GFP构建体的pMA-RQ。

[0375] -含有SYNP-ATF6-01构建体的pMA-RQ。

[0376] -含有SYNP-ATF6-02构建体的pMA-RQ。

[0377] SYNP-ATF6-01序列,6x ATF6位点带有下列划线(SEQ ID NO:30):

[0378] TGACGTGCTTGACGTGCTTGACGTGCTTGACGTGCTTGACGTGCTTGACGTGCTGGTACCGTCGACGA
TATCGGATCCAGGTCTATATAAGCAGAGCTCGTTTAGTGAACCGTCAGATCGCCTAGATACGCCATCC
ACGCTGTTTTGACCTCCATAGAAGATCGCCACCATGGTGAGCAAGGGCGAGGAGCTGTTCACCGGGGT
GGTGCCCATCCTGGTCGAGCTGGACGCGCAGCTAAACGGCCACAAGTTCAGCGTGTCCGGCGAGGGCG
AGGGCGATGCCACCTACGGCAAGCTGACCCTGAAGTTCATCTGCACCACCGCAAGCTGCCCGTGCCC
TGGCCACCCCTCGTGACCACCCCTGACCTACGGCGTGAGTGCTTCAGCCGCTACCCCGACCACATGAA
GCAGCAGCACTTCTTCAAGTCCGCCAIGCCGAAGGCTACGTCCAGGAGCGCACCATCTTCTTCAAGG
ACGACGGCAACTACAAGACCCGCGCCGAGGTGAAGTTCGAGGGCGACACCCTGGTGAACCGCATCGAG
CTGAAGGGCATCGACTTCAAGGAGGACGGCAACATCCTGGGGCACAAGCTGGAGTACAACCTACAACAG
CCACAACGTCTATATCATGCGCGACAAGCAGAAGAACGGCATCAAGGTGAACTTCAAGATCCGCCACA
ACATCGAGGACGGCAGCGTGCAGCTCGCCGACCACTACCAGCAGAACACCCCCATCGGCGACGGCCCC
GTGCTGCTGCCCCGACAACCACTACCTGAGCACCCAGTCCGCCCTGAGCAAAGACCCCAACGAGAAGCG
CGATCACATGGTCCTGCTGGAGTTCGTGACCGCCGCCGGGATCACTCTCGGCATGGACGAGCTGTACA
AGTGA

[0379] SYNP-ATF6-02的序列,6x ATF6位点带有下划线,切除的内含子序列以小写和粗体显示(SEQ ID NO:31):

TGACGTGCTTGACGTGCTTGACGTGCTTGACGTGCTTGACGTGCTGGTACCGTCGACGATATCG
GATCCAGGTCTATATAAGCAGAGCTCGTTTAGTGAACCGTCAGATCGCCTAGATACGCCATCCACGCTGTTTTG
ACCTCCATAGAAGATCGCCACCATGGTGAGCAAGGGCGAGGAGCTGTTCACCGGGGTGGTGCCCATCCTGGT
CGAGCTGGACGGCGACGTAAACGGCCACAAGTTCAGCGTGTCGGCGAGGGCGAGGGCGATGCCACCTACG
GCAAGCTGACCCTGAAGTTCATCTGCACCACCGGCAAGCTGCCCGTGGCCCTGGCCACCCCTCGTGACCACCT
[0380] **GACCTACGGCGTGCA**GTGCTTCAGCCGCTACCCCGACCACATGAAGCAGCAGCACTTCTTCAAGTCCGCCATG
CCCGAAGGCTACGTCCAGGAGCGACCATCTTCTTCAAGGACGACGGCAACTACAAGACCCGCGCCGAGGTG
AAGTTCGAGGGCGACACCTGGTGAACCGCATCGAGCTGAAGGGCATCGACTTCAAGGAGGACGGCAACAT
CCTGGGGCACAAGCTGGAGTACAACTACAACAGCCACAACGTCTATATCATGGCCGACAAGCAGAAGAACGG
CATCAAGGTGAACTTCAAGATCCGCCACAACATCGAGGACGGCAGCGTGCA**gtgagcactcagactacgtgcacct**
ctgCTGGCCGACCACTACCAGCAGAACACCCCATCGGCGACGGCCCCGTGCTGCTGCCGACAACCACTACCT
GAGACCCAGTCCGCCCTGAGCAAAGACCCCAACGAGAAGCGCGATCACATGGTCCTGCTGGAGTTCGTGA

[0381] 结果

[0382] 如上所述建立24孔转染。每个条件一式两份/一式三份进行。转染之后24小时,将细胞用2mM DTT处理以诱导UPR,或者用等体积的水进行模拟处理。细胞处理之后1、3、5和24小时,如先前所述测量GFP荧光。

[0383] 在诱导前后计算GFP荧光与对照质粒的比。图3示出了3次独立实验的累积结果。

[0384] 图3示出了在通过2mM DTT诱导之后,来自两个质粒的GFP的表达。数字是与CMV-IE的比。这表明6x ATF6元件可被ER应激高度诱导,并可将基因表达增强至CMV-IE水平的3倍。此外,向EGFP添加XBP1内含子使表达提高至CMV-IE水平的4倍。当分析原始背景数据时,发现SYNP-ATF6-01具有与CMV-MP相当的背景活性,向EGFP添加内含子使GFP荧光的背景水平降低到几乎为零(图4和5)。通过荧光显微术证实了这一点(图6)。这表明本文提供的蛋白质表达的双重控制方法,即通过可调节的启动子的转录和通过可调节内含子的翻译,可提供严密的调节的控制和高表达水平。

[0385] 实施例3:在wtAAV产生过程中来自SYNP-ATF6-01和SYNP-ATF6-02的EGFP的诱导型表达

[0386] 该实施例的目的是确定是否可通过产生异源蛋白来激活ER应激反应,作为对化学ER应激的替代(或补充),如前所述。进行了实验,由此在AAV产生过程中测量了来自SYNP-ATF6-01和SYNP-ATF6-02的GFP表达。

[0387] HEK293F细胞用于这项研究。如实施例1所述进行所有生长条件、转染和分析。

[0388] 在本实验中,如前所述在24孔板中进行转染。将ATF6质粒(pMA-RQ载体中的SYNP-ATF6-01和SYNP-ATF6-02构建体)转染到HEK293F细胞中(使用MAX试剂,如上所述),并在24小时之后测量GFP。在GFP测量之后,将相同细胞用于wtAAV产生的质粒pGRG25AAV2(由Adrien Savvy提供)和来自Takara/Clontech的pHelper质粒以1:1的比转染。然后在1、3、5和24小时之后测量EGFP荧光(图7和8-并未示出所有数据)。两种质粒都是wtAAV产生所必需的,其中pGRG25AAV2提供wtAAV2基因组,pHelper质粒提供病毒复制所需的E2、VA和E4辅助功能。

[0389] wtAAV2病毒基因组是本领域公知的(参见Srivastava等“Nucleotide sequence and organization of the adeno-associated virus 2 genome”, J Virol.1983 Feb;45 (2):555-564),并且AAV的表达体系在本领域中也是公知的。在当前情况下,将wtAAV2病毒基因组插入到质粒pGRG25中,这描述于McKenzie和Craig,“Fast,easy and efficient: site-specific insertion of transgenes into Enterobacterial chromosomes using Tn7 without need for selection of the insertion event”;BMC Microbiology 2006, 6:39中。pHelper质粒可获自Takara/Clontech(AAVpro体系的目录号6234)。合适的用于诱导AAV表达的AAV2表达体系广泛地可商购,例如从Takara (Clontech)作为“AAVpro无辅助体系(AAV2)”-参见http://www.clontech.com/US/Products/Viral_Transduction/AAV_Vector_Systems/Helper_Efree_Expression_System。

[0390] 质粒pAAV-CMV-ZsGreen(目录号6231,来自Takara/Clontech AAV载体体系)用作对照,以确认在细胞中实现了AAV表达。将pAAV-CMV-ZsGreen和pHelper质粒以1:1的比转染到单独的HeK293F细胞群中。这证实了成功的AAV表达。Zs green是eGFP的变体,因此同样地如前所述,对这些细胞进行了GFP测量。

[0391] 这些数据表明,通过产生wtAAV,两种质粒都可被诱导至CMV-IE的水平。此外,可观察到ATF6-01的背景水平与先前实验相似,而ATF6-02的背景水平为零,只有wtAAV的合成才能诱导EGFP的产生。这支持了我们的初步发现,即完全控制表达需要两个ER应激元件。

[0392] 实施例4:用多种诱导物诱导UPR

[0393] 进行实验以评估数种候选试剂诱导UPR的能力。使用基本上如上所述的技术评估这些候选诱导物诱导UPR的能力。受测的其他候选物为0.5mM棕榈酸、1μM MF-43(2-甲基-5-(6-(4-(2-(三氟甲基)苯氧基)哌啶-1-基)哒嗪-3-基)-1,3,4-噻二唑),其是硬脂酰辅酶A脱饱和酶的抑制剂,以及二者的组合(使用各自的相同的浓度)。

[0394] 在这些实验中,如前所述在24孔板中进行转染。将SYNP-ATF6-02质粒(即pMA-RQ载体中的SYNP-ATF6-02)转染到HEK293F细胞中,并在24小时之后测量GFP。然后使用上述诱导物刺激UPR的诱导,并在诱导之后0、1、3、5和24小时时取样并测量。如前所述测量GFP表达。结果是3次独立实验的平均值,误差棒代表标准偏差。

[0395] 实验步骤简述:

[0396] -使40ml细胞在250ml排气式锥形烧瓶(Sigma-Aldrich CLS431144)中于37℃,8% CO₂和100rpm搅拌下生长。按照制造商的说明接种细胞。

[0397] -在转染之前1天,使用血细胞计数器对细胞进行计数,并将其分成500,000个细胞/ml。

[0398] -在转染当天,将细胞以100万个细胞/ml接种到24孔板中的500μl适当培养基中。

[0399] -然后将0.625μg DNA/孔添加至10μl OptiMem培养基(Thermofisher;11058021),并在室温下孵育5分钟。

[0400] -同时通过添加OptiMem将0.625μl的Max试剂补足至10μl,并在室温下孵育5分钟。

[0401] -孵育后,将两种混合物添加至同一管中,并在室温下孵育25至30分钟。

[0402] -然后将DNA/Max试剂混合物(20μl/孔)直接添加至细胞,并如前所述孵育细胞。

[0403] -然后在24小时之后测量细胞的GFP荧光。

[0404] -在添加2mM DTT、0.5mM棕榈酸、1μM MF-43、0.5mM棕榈酸和1μM MF-43的组合之后

测量从GFP构建体中的内含子剪接,以及在1、3、5和24小时之后通过监测GFP荧光进行AAV分析。

[0405] 针对UPR诱导的AAV产生:如前所述,将ATF6质粒(即pMA-RQ载体中的SYNP-ATF6-02)转染到HEK293细胞中。24小时之后,测量GFP。在GFP测量之后,如上所述,将相同细胞用于wtAAV产生的质粒pGRG25AAV2和pHelper质粒以1:1的比转染到细胞中。然后在1、3、5和24小时之后测量EGFP荧光。

[0406] GFP荧光的测量

[0407] -通过用无菌水稀释成五分之一,从萤光素酶测定体系(Promega;E1500)准备细胞裂解缓冲液。

[0408] -将细胞以900xg沉淀5分钟,然后以100 μ l/1x10⁶个细胞重悬于来自萤光素酶试剂盒的裂解缓冲液中。

[0409] -然后将其在室温下孵育10至15分钟以进行裂解。

[0410] -以900xg持续5分钟收获细胞核和碎片。

[0411] -收集上清液,将最大100 μ l/样品收集至96孔黑色板。

[0412] -将上清液用PBS稀释成二分之一(这是为了从裂解缓冲液中稀释影响GFP信号的巯基乙醇;在存在还原剂的情况下GFP信号降低)。

[0413] -使样品在室温下孵育5分钟。

[0414] -使用读板器在激发和发射分别设置在485nm和520nm处的情况下测量GFP荧光。

[0415] -GFP荧光也直接在荧光显微镜下可视化。

[0416] 这些实验的结果显示在图9中,从中可看出,所有试剂均成功诱导了UPR,如添加试剂后功能性EGFP的表达所表明的。每种试剂的诱导强度不同。DTT是UPR的强大诱导物。影响细胞中脂质平衡的棕榈酸和MF-43均能有效地单独诱导UPR,当联合使用时,诱导水平更高。WtAAV产生是有效的诱导物,但其作用不如其他诱导物强。

[0417] 来自本实验的令人感兴趣且潜在有用的观察结果是,可使用不同的诱导物以不同水平诱导UPR,从而允许控制剪接水平和功能性表达产物(在这种情况下为EGFP)的该表达。作为替代地或另外地,多种试剂的不同剂量水平可用于调节功能性表达产物的表达水平。

[0418] 该实施例中列出的程序可用于评估任何试剂诱导UPR的能力。

[0419] 实施例5-可调节内含子用于调节分泌碱性磷酸酶(SEAP)表达的用途

[0420] 分泌的碱性磷酸酶是生物加工工业中用作标志物的标准蛋白。当蛋白质通过细胞的所有蛋白质质量控制(转录、翻译、翻译后修饰然后分泌)检查点时,它是分泌型蛋白的理想标志物。

[0421] 包含可调节内含子的SEAP序列的构建。

[0422] 在GeneART通过化学合成来合成构建体SEAP-ATF6-001

[0423] 使用的UPR诱导型顺式调节元件(增强子区域)为6x ATF6 (TGACGTGCT),每个ATF6间隔20bp,与CMV-MP偶联。这是上面使用的串联重复序列6x ATF6启动子的修饰,其序列在下面的序列中带有下划线。

[0424] 在SEAP编码序列的1314位的2个CAG密码子之间设计了内含子。编码插入的内含子的切除区域的DNA如下:

[0425] **CAGACGGGCAACTTTACACGACGCTG (SEQ ID NO: 32)**

[0426] 包含剪接位点,这将导致以下序列:

[0427] **CAG/CAGACGGGCAACTTTACACGACGCTG/CAG. (SEQ ID NO: 33)**

[0428] 该序列没有导致文献中针对XBP1野生型内含子描述的内含子二级结构(关于这一点的讨论参见下文实施例6)。

[0429] 包含启动子和可调节内含子的SEAP表达构建体序列如下所示(切除的内含子序列以粗体显示,并且6x ATF6增强子区域带有下列线):

TGACGTGCTGATGATGCGTAGCTAGTAGTTGACGTGCTGATGATGCGTAGCTAGTAGTTGACGTGCTG
ATGATGCGTAGCTAGTAGTTGACGTGCTGATGATGCGTAGCTAGTAGTTGACGTGCTGATGATGCGTA
GCTAGTAGTTGACGTGCTGATGATGCGTAGCTAGTAGTGCAGTTAGCGTAGCTGAGGTACCGTCGACG
ATATCGGATCCAGGTCTATATAAGCAGAGCTCGTTTAGTGAACCGTCAGATCGCCTAGATACGCCATC
CACGCTGTTTTGACCTCCATAGAAGATCGCCACCATGCTGCTGCTGCTGCTGCTGCTGGGCCTGAGGC
TACAGCTCTCCCTGGGCATCATCCAGTTGAGGAGGAGAACCCGGAATTCTGGAACCGCGAGGCAGCC
GAGGCCCTGGGTGCCGCCAAGAAGCTGCAGCCTGCACAGACAGCCGCCAAGAACCTCATCATCTTCCT
GGGCGATGGGATGGGGGTGTCTACGGTGACAGCTGCCAGGATCCTAAAAGGGCAGAAGAAGGACAAAC
TGGGGCCTGAGATACCCCTGGCTATGGACCGCTTCCCATATGTGGCTCTGTCCAAGACATACAATGTA
GACAAACATGTGCCAGACAGTGGAGCCACAGCCACGGCCTACCTGTGCGGGGTCAAGGGCAACTTCCA
GACCATTGGCTTGAGTGCAGCCGCCCGCTTTAACCAGTGCAACACGACACGCGGCAACGAGGTCATCT
CCGTGATGAATCGGGCCAAGAAAGCAGGGAAGTCAGTGGGAGTGGTAACCACCACACGAGTGCAGCAC
GCCTCGCCAGCCGGCACCTACGCCCACACGGTGAACCGCAACTGGTACTCGGACGCCGACGTGCCTGC
CTCGGCCCCGCCAGGAGGGGTGCCAGGACATCGCTACGCAGCTCATCTCCAACATGGACATTGACGTGA

[0430] TCCTAGGTGGAGGCCGAAAGTACATGTTTCGCATGGGAACCCAGACCCTGAGTACCCAGATGACTAC
AGCCAAGGTGGGACCAGGCTGGACGGGAAGAATCTGGTGCAGGAATGGCTGGCGAAGCGCCAGGGTGC
CCGGTATGTGTGGAACCGCACTGAGCTCATGCAGGCTTCCCTGGACCCGTCTGTGACCCATCTCATGG
GCCTCTTTGAGCCTGGAGACATGAAATACGAGATCCACCGAGACTCCACACTGGACCCCTCCCTGATG
GAGATGACAGAGGCTGCCCTGCGCCTGCTGAGCAGGAACCCCCGCGGCTTCTTCCTCTTCGTGGAGGG
TGGTGCATCGACCAGGTCATCATGAAAGCAGGGCTTACCGGGCACTGACTGAAACGATCATGTTTCG
ACGACGCCATTGAGAGGGCGGGCCAGCTCACCAGCGAGGAGGACACGCTGAGCCTCGTCACTGCCGAC
CACTCCCACGTCTTCTCCTTCGGAGGCTACCCCTGCGAGGGAGCTCCATCTTCGGGCTGGCCCTGG
CAAGGCCCGGGACAGGAAGGCCTACACGGTCTCTTATACGGAAACGGTCCAGGCTATGTGCTCAAGG
ACGGCGCCCCGGCCGATGTTACCGAGAGCGAGAGCGGGAGCCCCGAGTATCGGCAG**CAGACGGGCAAC**
TTTACACGACGCTGCAGTCAGCAGTGCCCTGGACGAAGAGACACACGCAGCGAGGACGTGGCGGTG
TTCGCGCGCGGGCCCGCAGGCGCACCTGGTTACGCGCTGCAGGAGCAGACCTTCATAGCGCACGTCTAT
GGCCTTCGCCGCTGCTGGAGCCCTACACCGCTGCGACCTGGCGCCCCCGCGGCACCACCGACG
CCGCGACCCGGGTTACTCTAGAGTCGGGGCGGCCGCGCTTCGAGCAGACATGA SEQ ID NO:

34)

[0431] 包含内含子的编码序列的翻译导致以下截短的蛋白序列:

[0432] MLLLLLLLGLRLQLSLGIIPVEEENPDFWNREAAEALGAACKLQPAQTAAKNLII FLGDGMGVSTVTA
ARILKGQKKDKLGPEIPLAMDRFPYVALSKTYNVDKHVPDSGATATAYLCGVKGNFQTIGLSAAARFN
QCNTTRGNEVISVMNRAKKAGKSVGVVTTTRVQHASPAGTYAHTVNRNWYSADVPASARQEGCQDIA
TQLISNMDIDVILGGGRKYMFRMGTPDPEYPDDYSQGGTRLGKNLVQEWLAKRQGARYVWNRTELMQ
ASLDPSVTHLMGLFEPGDMKYEIHRDSTLDPSLMEMTEAALRLLSRNPRGFFLFVEGGRIDHGHESR
AYRALTETIMFDDAIERAGQLTSEEDTSLSLVTADHSHVFSFGGYPLRGSSIFGLAPGKARDRKAYTVL
LYGNPGYVVLKDGARPDVTESESGSPEYRQQTGNFTRRCSQQCPWTKRHTQARTWRCSRAARRRTWFT
ACRSRPS (SEQ ID NO: 35)

[0433] 去除内含子允许完全翻译SEAP蛋白:

[0434] MLLLLLLLGLRLQLSLGIIPVEEENPDFWNREAAEALGAACKLQPAQTAAKNLII FLGDGMGVSTVTA
ARILKGQKKDKLGPEIPLAMDRFPYVALSKTYNVDKHVPDSGATATAYLCGVKGNFQTIGLSAAARFN
QCNTTRGNEVISVMNRAKKAGKSVGVVTTTRVQHASPAGTYAHTVNRNWYSADVPASARQEGCQDIA
TQLISNMDIDVILGGGRKYMFRMGTPDPEYPDDYSQGGTRLGKNLVQEWLAKRQGARYVWNRTELMQ
ASLDPSVTHLMGLFEPGDMKYEIHRDSTLDPSLMEMTEAALRLLSRNPRGFFLFVEGGRIDHGHESR
AYRALTETIMFDDAIERAGQLTSEEDTSLSLVTADHSHVFSFGGYPLRGSSIFGLAPGKARDRKAYTVL
LYGNPGYVVLKDGARPDVTESESGSPEYRQQSAVPLDEETHAGEDVAVFARGPQAHLVHGVQEQT FIA
HVMAFAACLEPYTACDLAPPAGTTDAHPGYSRVGAAGRFEQT (SEQ ID NO: 36)

[0435] 基本按照上述方法所述进行实验,即,用上述构建体(pMA-RQ)转染HEK或CHO-s细胞并孵育24小时。在此时间之后,以2mM终浓度添加活化化合物DTT,然后在诱导之后3、5和24小时测量SEAP活性。

[0436] 更详细地:

[0437] 24孔板的转染方案如实施例1中所述。此后:

[0438] -24小时之后测量上清液的SEAP活性。

[0439] -在添加2mM DTT后,通过监测3、5和24小时之后的SEAP活性来测量来自SEAP构建体的内含子剪接。

[0440] -按照制造商的说明,Roche,SEAP报道基因测定试剂盒,购自Sigma-Aldrich,目录号11779842001,来测量SEAP活性。

[0441] 如图11所示,在诱导之前没有SEAP表达。然而,在添加DTT后,SEAP活性迅速提高,实际上3小时之后,SEAP活性的水平类似于CMV-IE(一种强大的组成型启动子)。这种表达速度太快,以至于无法通过单独的转录来解释,这表明mRNA是由漏失的启动子产生的,并且几乎可通过去除内含子而立即翻译。实际上,在24小时之后,两种细胞类型中的表达水平都比CMV-IE高约2倍。

[0442] 这些结果表明内含子适合与分泌型蛋白一起使用。此外,它暗示了关于由内含子形成的二级结构和关于侧翼为剪接位点靶序列的中央内含子序列(即上述序列称为Xn的序列)具有相当大的灵活性,其中从mRNA成功剪接出内含子的重要因素是剪接位点靶序列。本实验中使用的可调节内含子序列(CAGACGGGCAACUUUACACGACGCUG (SEQ ID NO:37))与野生型XBP1内含子序列(CAGCACUCAGACUACGUGACCUCUG (SEQ ID NO:23))完全不同,但仍实现了有效的剪接。此外,如下面更详细地讨论的,预期本实施例中使用的内含子不会形成类似于由野生型XBP1内含子形成的二级结构。

[0443] 实施例6-使用萤火虫萤光素酶基因的可调节内含子用于调节蛋白质表达的用途。

[0444] 在现有技术中已经断言,由XBP1内含子形成的二级结构对于剪接是必不可少的。然而,上述所有实验都使用了预期不会形成这样的二级结构或类似于野生型XBP1内含子的结构的内含子结构,但剪接体系的性能非常强大。

[0445] 因此,本实验的一个方面是确定专门设计为具有现有技术中所述的二级结构的内含子的作用。选择萤光素酶,一种胞内蛋白作为报道蛋白,由于易于测定表达水平,因此也是有利的。

[0446] 包含可调节内含子的萤光素酶序列的构建。

[0447] 在GeneART上通过化学合成来合成表达构建体。如上所述,增强子区域是与CMV-MP偶联的间隔为20bp的6x ATF6(TGACGTGCT)。

[0448] 将内含子插入到位于萤光素酶编码序列的1447位的CCG和CAG密码子之间。

[0449] 编码插入的内含子序列的切除区域的DNA如下:

[0450] **CAGACGGGCAACTTTACACGACGCTG (SEQ ID NO: 32)**

[0451] 包含剪接位点,这导致以下序列:

[0452] **CCG/CAGACGGGCAACTTTACACGACGCTG/CAG (SEQ ID NO: 38)**

[0453] 如下所述,当将这种可调节内含子序列插入到萤光素酶基因中时,预计将提供相同的二级结构,正如针对在XBP1基因中的天然位置中的XBP1野生型内含子所描述的(使用位于<http://rna.tbi.univie.ac.at/cgi-bin/RNAWebSuite/RNAfold.cgi>的RNA折叠网络服务器计算)。萤光素酶中的由内含子形成的预测二级结构如图16D所示。为了进行比较,图16还示出了当内含子插入到其他实验中所使用的相关基因中时形成的预测二级结构:

[0454] -图16B-如在EGFP中使用的内含子(即如在实施例1至4中使用的,由CAG/CTGCAGCACTCAGACTACGTGCACCTCTG/CTG,SEQ ID NO:48编码);

[0455] -图16A-如在SEAP中使用的内含子(即如在实施例5中使用的,由CAG/CAGACGGGCAACTTTACACGACGCTG/CAG,SEQ ID NO:33编码);

[0456] -图16C-如在CASP9中使用的内含子(即如在实施例7中使用的,由CAG/CAGACGGGCAACTTTACACGACGCTG/CTG,SEQ ID NO:43编码),参见图16C

[0457] 包含启动子和可调节内含子的萤光素酶表达构建体序列如下所示(内含子序列以粗体显示,并且6x ATF6增强子区域带有下列线):

[0458] 包含内含子的编码序列的翻译产生以下截短的蛋白:

**MEDAKNIKKGPAPFYPLEDGTAGEQLHKAMKRYALVPGTIAFTDAHIEVDITYAEYFEMSVR
LAEAMKRYGLNTNHRIVVCSNSLQFFMPVLGALFIGVAVAPANDIYNERELLNSMGISQPT
VVFVSKKGLQKILNVQKKLPPIQKIIIMDSKTDYQGFQSMYTFVTSHLPPGFNEYDFVPESFD
RDKTIALIMNSSGSTGLPKGVALPHRTACVRFSHARDPIFGNQIIPDTAILSVPFHGFGMF
TTLGYLICGFRVVLMYRFEEELFLRSLQDYKIQSALLVPTLFSFFAKSTLIDKYDLSNLHEIAS
GGAPLSKEVGEAVAKRFHLPGRQGYGLTETTSAILITPEGDDKPGAVGKVVPFFFEAKVVDL
DTGKTLGVNQRGELCVRGPMIMSGYVNNPEATNALIDKDGWLHSGDIAYWDEDEHFFIVD
RLKSLIKYKGYQVAPAELESILLQHPNIFDAGVAGLPDDDAGELPAADGQLYTTLTQTSCWNT
VKP (SEQ ID NO: 39)**

[0460] 去除内含子允许翻译完整的萤光素酶蛋白:

[0461] MEDAKNIKKGPAPFYPLEDGTAGEQLHKAMKRYALVPGTIAFTDAHIEVDITYAEYFEMSVR
LAEAMKRYGLNTNHRIVVCSSENSLQFFMPVLGALFIGVAVAPANDIYNERELLNSMGISQPT
VVFVSKKGLQKILNVQKKLPPIQKIIIMDSKTDYQGFQSMYTFVTSHLPPGFNEYDFVPESFD
RDKTIALIMNSSGSTGLPKGVALPHRTACVRFSHARDPIFGNQIIPDTAILSVVPPFHHGFGMF
TTLGYLICGFRVVLMYRFEEELFLRSLQDYKIQSALLVPTLFSFFAKSTLIDKYDLSNLHEIAS
GGAPLSKEVGEAVAKRFHLPGRQGYGLTETTSAILITPEGDDKPGAVGKVVPFFFEAKVVDL
DTGKTLGVNQRGELCVRGPMIMSGYVNNPEATNALIDKDGWLHSGDIAYWDEDEHFFIVD
RLKSLIKYKGYQVAPAELESILLQHPNIFDAGVAGLPDDDAGELPAADWVLEHGKTMTEKEIV
DYVASQVTTAKLRGGVVFVDEVPKGLTGKLDARKIREILIKAKKGGKIAV (SEQ ID NO: 40)

[0462] 基本上如实施例5所述进行实验,即用上述构建体转染HEK或CHO-s细胞并孵育24小时。在此时间后,添加活化化合物DTT或棕榈酸。然后在诱导之后3、5和24小时测量萤光素酶活性。萤光素酶测量如下:

[0463] -通过用无菌水稀释成五分之一,从萤光素酶测定体系(Promega;E1500)准备细胞裂解缓冲液。

[0464] -将细胞以900xg沉淀5分钟,然后以100 μ l/1x10⁶个细胞重悬于来自萤光素酶试剂盒的裂解缓冲液中。

[0465] -然后将其在室温下孵育10至15分钟以发生裂解。

[0466] -以900xg持续5分钟收获细胞核和碎片。

[0467] -收集上清液,将10 μ l/样品添加96孔白板中。

[0468] -使用用自动进样器的读板仪测量生物发光,并注入50 μ l底物。按照制造商的说明(Promega;E1500)准备底物。

[0469] 如图12所示,在诱导之前没有萤光素酶表达,然而添加任何一种诱导物后,活性都会迅速提高,实际上3小时之后,萤光素酶的活性是CMV-IE(一种强大的组成型启动子)的2至2.5倍。再次,这种表达速度太快以至于不能由单独的转录来解释,并且表明一些mRNA已经由启动子产生并且在诱导之后去除内含子后几乎立即被翻译。实际上,在24小时之后,两种细胞类型中的表达水平平均比CMV-IE高约3至4倍。

[0470] 这些结果在某种程度上暗示了“完美的”内含子二级结构在某些情况下可提供更好的诱导水平或更高的剪接效率。然而,结构对于体系的整体功能显然并不重要。此外,SEAP和萤光素酶是完全不同的蛋白质,不可进行直接比较。SEAP是分泌型的,因此经历了进一步修饰和表达的另外的瓶颈。

[0471] 实施例7-使用可调节内含子控制胱天蛋白酶9基因表达

[0472] 进行本实验以证实来自可调节内含子的表达控制是适当严密的,以控制毒性表达产物的表达,所述毒性表达产物对表达它们的细胞具有致命性。可调节内含子被工程化为引起蛋白质胱天蛋白酶9的凋亡。该蛋白质在HEK细胞中的过表达导致迅速死亡。因此,来自未经修饰的序列的该蛋白质的任何表达都会导致培养物中死细胞的显著提高。这使我们能够确定内含子所提供的控制的严密性。胱天蛋白酶9也与GFP-Spark融合,如果内含子存在于转录物中,这将是不可见的。

[0473] 包含可调节内含子的萤光素酶序列的构建。

[0474] 为了表达,将内含子与CMV-IE组成型启动子偶联。使用BsaI限制性位点和2个互补的寡核苷酸将内含子构建体克隆到质粒载体SYNP-CASPSp-001中。用BsaI(NEB R0535S)消化质粒,并通过加热至98 $^{\circ}$ C持续5分钟使寡核苷酸退火,以使二级结构解链,然后在55 $^{\circ}$ C下

孵育20分钟。这使得寡核苷酸形成双链DNA。该DNA的5'和3'端均带有单链突出端(overhang),可与经消化的SYNP-CASPSp-001连接。然后将双链DNA连接到质粒中,随后转化到one shot top ten化学感受态细胞(Thermofisher,C404003)中。对分离的DNA进行测序以确认内含子的存在。完整的质粒称为SYNP-CASP9-INT,质粒图谱如图13所示。

[0475] 添加至载体中的内含子寡核苷酸具有以下序列:

[0476] INTC9FP:ACGTCAGACGGGCAACTTTACACGACGCTG (SEQ ID NO:41)

[0477] INTC9RP:ACGTCAGCGTCGTGTAAAGTTGCCCCGTCTG (SEQ ID NO:42)

[0478] 因此,在添加的CCAG和胱天蛋白酶9编码序列的1087位处的CTG密码子之间对可调节内含子进行工程化。

[0479] 编码插入的内含子序列的切除区域的DNA如下:

[0480] **CAGACGGGCAACTTTACACGACGCTG (SEQ ID NO: 32)**

[0481] 包含剪接位点,这将导致以下序列:

[0482] **CAG/CAGACGGGCAACTTTACACGACGCTG/CTG (SEQ ID NO: 43)**

[0483] 该序列不会导致针对XBP1野生型内含子描述的“完美的”内含子二级结构。

[0484] 包含CASP9-内含子构建体序列的SYNP-CASP9-INT载体如下(内含子序列以粗体显示):

[0485] GACGGATCGGGAGATCTCCCGATCCCCTATGGTGCACCTCTCAGTACAATCTGCTCTGATGCCGCATAG
TTAAGCCAGTATCTGCTCCCTGCTTGTGTGTTGGAGGTCGCTGAGTAGTGCGCGAGCAAAATTTAAGC
TACAACAAGGCAAGGCTTGACCGACAATTGCATGAAGAATCTGCTTAGGGTTAGGCGTTTTGCGCTGC
TTCGCGAGTACATTTATATTGGCTCATGTCCAATATGACCGCCATGTTGACATTGATTATTGACTAGT
TATTAATAGTAATCAATTACGGGGTCATTAGTTTCATAGCCCATATATGGAGTTCGCGGTTACATAACT
TACGGTAAATGGCCCCGCTGGCTGACCGCCCAACGACCCCCGCCATTGACGTCAATAATGACGTATG
TTCCCATAGTAACGCCAATAGGGACTTTCCATTGACGTCAATGGGTGGAGTATTTACGGTAAACTGCC
CACTTGGCAGTACATCAAGTGATCATATGCCAAGTCCGCCCCCTATTGACGTCAATGACGGTAAATG
GCCCCGCTGGCATTATGCCCAGTACATGACCTTACGGGACTTTTCCTACTTGGCAGTACATCTACGTAT
TAGTCATCGCTATTACCATGGTGATGCGGTTTTGGCAGTACACCAATGGGCGTGGATAGCGGTTTTGAC
TCACGGGGATTTCCAAGTCTCCACCCATTGACGTCAATGGGAGTTTGTTTTGGCACCAAAATCAACG
GGACTTTCCAAAATGTCGTAATAACCCCGCCCCGTTGACGCAAATGGGCGGTAGGCGGTGACGGTGGG
AGGTCTATATAAGCAGAGCTCGTTTTAGTGAACCGTCAGATCCTCACTCTCTTCCGCATCGCTGTCTGC
GAGGGCCAGCTGTTGGGCTCGCGGTTGAGGACAAACTCTTCGCGGTCTTTCCAGTACTCTTGATCGG
AAACCCGTCGGCCTCCGAACGGTACTCCGCCACCGAGGGACCTGAGCGAGTCCGCATCGACCGGATCG
GAAAACCTCTCGAGAAAGGCGTCTAACCAGTCACAGTCGCAAGGTAGGCTGAGCACCGTGGCGGGCGG
CAGCGGGTGGCGGTGCGGGTTGTTCTGGCGGAGGTGCTGCTGATGATGTAATTAAAGTAGGCGGTCT
TGAGACGGCGGATGGTCGAGGTGAGGTGTGGGTTTAGTGAACCGTCAGATCCTCACTCTCTTCCGCAT
CGCTGTCTGCGAGGGCCAGCTGTCAGGCTTGAGATCCAGCTGTTGGGGTGAGTACTCCCTCTCAAAG
CGGGCATTACTTCTGCGCTAAGATTGTCAGTTTCCAAAACGAGGAGGATTTGATATTACCTGGCCC
GATCTGGCCATACACTTGAGTGACAATGACATCCACTTTGCCTTTCTCTCCACAGGTGTCCACTCCCA
GGTCCAAGTTTAAACTTTAATACGACTCACTATAGGGGCCGCCACCAAGCTTGGTACATGGACGAAGC
GGATCGGCGGCTCCTGCGGCGGTGCCGGCTGCGGCTGGTGGAAGAGCTGCAGGTGGACCAGCTCTGGG
ACGCCCTGCTGAGCCGCGAGCTGTTACAGCCCCATATGATCGAGGACATCCAGCGGGCAGGCTCTGGA

[0486]

TCTCGGCGGGATCAGGCCAGGCAGCTGATCATAGATCTGGAGACTCGAGGGAGTCAGGCTCTTCCTTT
GTTTCATCTCCTGCTTAGAGGACACAGGCCAGGACATGCTGGCTTCGTTTCTGCGAACTAACAGGCAAG
CAGCAAAGTTGTGGAAGCCAAACCCTAGAAAACCTTACCCCAGTGGTGCTCAGACCAGAGATTGCGAAA
CCAGAGGTTCTCAGACCGGAAACACCCAGACCAGTGGACATTGGTTCTGGAGGATTCGGTGATGTCGG
TGCTCTTGAGAGTTTGAGGGGAAATGCAGATTGCTTACATCCTGAGCATGGAGCCCTGTGGCCACT
GCCTCATTATCAACAATGTGAACCTCTGCCGTGAGTCCGGGCTCCGCACCCGCACTGGCTCCAACATC
GACTGTGAGAAGTTGCGGCGTCGCTTCTCCTCGCTGCATTTTCATGGTGGAGGTGAAGGGCGACCTGAC
TGCCAAGAAAATGGTGCTGGCTTTGCTGGAGCTGGCGCAGCAGGACCACGGTGCTCTGCACTGCTGCG
TGGTGGTCATTCTCTCTCACGGCTGTCAGGCCAGCCACCTGCAGTTCACAGGGGCTGTCTACGGCACA
GATGGATGCCCTGTGTCGGTCGAGAAGATTGTGAACATCTTCAATGGGACCAGCTGCCCCAGCCTGGG
AGGGAAGCCCAAGCTCTTTTTCATCCAGGCCCTGTGGTGGGGAGCAGAAAGACCATGGGTTTGAGGTGG
CCTCCACTTCCCCTGAAGACGAGTCCCCTGGCAGTAACCCCGAGCCAGATGCCACCCCGTTCCAGGAA
GGTTTGAGGACCTTCGACCAGCTGGACGCCATATCTAGTTTGGCCACACCCAGTGACATCTTTGTGTC
CTACTCTACTTTCCCAGGTTTTGTTTCTGGAGGGACCCCAAGAGTGGCTCCTGGTCAGACGGGGCAAC
TTTACACGACGCTGGAGACCCTGGACGACATCTTGAGCAGTGGGCTCACTCTGAAGACCTGCAGTCC
CTCCTGCTTAGGGTCGCTAATGCTGTTTCGGTGAAAGGGATTTATAACAGATGCCTGGTTGCTTTAA
TTTCCTCCGAAAAAACTTTTCTTTAAACATCAGGGGGTGGAGGCTCTGTGAGCAAGGGCGAGGAGC
TGTTACACGGGGTGGTGCCATCCTGGTCGAGCTGGACGGCGACGTAACGGCCACAAGTTTCAGCGTG
TCCGGCGAGGGCGAGGGCGATGCCACCTACGGCAAGCTGACCCCTGAAGTTTCATCTGCACCACCGGCAA
GCTGCCCGTGCCCTGGCCCCACCTCGTGACCACCTGACCTACGGCGTGCACTGCTTCAGCCGCTACC
CCGACCACATGAAGAAGCAGACTTCTCAAGTCCGCCATGCCGAAGGCTACGTCCAGGAGCGCACC
ATCTTCTTCAAGGACGACGGCAACTACAAGACCCGCGCCGAGGTGAAGTTCGAGGGCGACACCCTGGT
GAACCGCATCGAGCTGAAGGGCATCGACTTCAAGGAGGACGGCAACATCCTGGGGCACAAGCTGGAGT
ACAACCTACAACAGCCACAACGCTCTATATCATGGCCGACAAGCAGAAGAAGCGCATCAAGGCTAACTTC
AAGGTTGCGCCACAACATCGAGGACGGCAGCGTGCACTCGCCGACCACTACCAGCAGAACACCCCCAT
CGGCGACGGCCCCGTGCTGCTGCCCCGACAACCACTACCTGAGCACCCAGTCCGCCCTGAGCAAAGACC
CCAACGAGAAGCGCGATCACATGGTCCTGCTGGAGTTCGTGACCGCCGCGGGGATCACTCTCGGCATG
GACGAGCTGTACAAGTAACTCGAGTCTAGAGCGGCCCGCGAATTCGGGCCCCGTTTAAACCCGCTGAT
CAGCCTCGACTGTGCCTTCTAGTTGCCAGCCATCTGTTGTTTGGCCCTCCCCCGTGCTTCTTTGACC
CTGGAAGGTGCCACTCCCCTGTCCTTTCTTAATAAAATGAGGAAATTGCATCGCATTTGTCTGAGTAG
GTGTCAATTCTATTCTGGGGGGTGGGGTGGGGCAGGACAGCAAGGGGGAGGATTGGGAAGACAATAGCA
GGCATGCTGGGGATGCGGTGGGCTCTATGGCTTCTGAGGCGGAAGAACCAGCTGGGGCTCTAGGGGG
TATCCCCACGCGCCCTGTAGCGGCGCATTAAGCGCGGGGGTGTGGTGGTTACGCGCAGCGTGACCGC
TACACTTGCCAGCGCCCTAGCGCCCGCTCCTTTGCTTCTTCCCTTCTTCTCGCCACGTTTCGAG
GCTTTCCCCGTCAAGCTCTAAATCGGGGGCTCCCTTTAGGGTTCCGATTTAGTGCTTTACGGCACCTC
GACCCCAAAAACTTGATTAGGGTGATGGTTACGTAAGTGGGCCATCGCCCTGATAGACGGTTTTTTCG
CCCTTTGACGTTGGAGTCCACGTTCTTTAATAGTGGACTCTTGTTCAAACTGGAACAACACTCAACC
CTATCTCGGTCTATTCTTTGATTTATAAGGGATTTTGCCGATTTCCGGCTATTGGTTAAAAAATGAG
CTGATTTAACAAAAATTTAACGCGAATTAPTCTCTGGAATGIGTGTCAGTTAGGGTGTGGAAAGTCC
CCAGGCTCCCCAGCAGGCAGGAAGTATGCAAGCAIGCATCTCAATTAGTCAGCAACCAGGTGTGAAA
GTCCCCAGGCTCCCCAGCAGGCAGGAAGTATGCAAGCAIGCATCTCAATTAGTCAGCAACCATAGTCC
CGCCCCCTAACTCCGCCCCATCCCGCCCCCTAACTCCGCCCCAGTTCGCCCCATCTCCGCCCCATGGCTGA
CTAATTTTTTTTATTTATGCAGAGGCCGAGGCCGCTCTGCCTCTGAGCTATTCCAGAAGTAGTGAGG
AGGCTTTTTTTGGAGGCCTAGGCTTTTGCAAAAAGCTCTCGGGAGCTTGTATATCCATTTTCGGATCTG
ATCAGCACGTGATGAAAAAGCCTGAACTCACCGCGACGCTCTGICGAGAAGTTTCTGATCGAAAAGTTC
GACAGCGTCTCCGACCTGATGCAGCTCTCGGAGGGCGAAGATCTCGTGCTTTACGCTTCGATGTAGG
AGGGCGTGGATATGTCCTGCGGGTAAATAGCTGCGCCGATGGTTTCTACAAAGATCGTTATGTTTATC
GGCACTTTGCATCGGCCGCGCTCCCGATTCGCGAAGTGCTTGACATTGGGGAATTACGCGAGAGCCTG

[0487]

ACCTATTGCATCTCCCGCCGTGCACAGGGTGTACAGTTGCAAGACCTGCCTGAAACCGAACTGCCCGC
TGTTCTGCAGCCGGTTCGCGGAGGCCATGGATGCGATCGCTGCGGCCGATCTTAGCCAGACGAGCGGGT
TCGGCCCATTCGGACCGCAAGGAATCGGTCAATACACTACATGGCGTGATTTTCATATGCGCGATTGCT
GATCCCCATGTGTATCACTGGCAAACGTGTATGGACGACACCCTCAGTGCCTCCGTCGCGCAGGCTCT
CGATGAGCTGATGCTTTGGGCCGAGGACTGCCCGAAGTCCGSCACCTCGTGACGCGGATTTTCGGCT
CCAACAATGTCTTGACGGACAATGGCCGCA¹AACAGCGGTCAATTGACTGGAGCGAGGCGATGTTCCGG
GATTCCCAATACGAGGTCCGCAACATCTTCTTCTGGAGGCCGIGGTTGGCTTGATGGAGCAGCAGAC
GCGCTACTTCGAGCGGAGGCATCCGGAGCTTGCAAGATCGCCCGGGCTCCGGGCGTATATGCTCCGCA
TTGGTCTTGACCAACTCTATCAGAGCTTGGTTGACGGCAATTTTCGATGATGCAGCTTGGCGCAGGGT
CGATGCGACGCAATCGTCCGATCCGGAGCCGGGACTGTGGGCGTACACAAATCGCCCGCAGAAGCGC
GGCCGTCTGGACCGATGGCTGTGTAGAAGTACTCGCCGATAGTGGAAACCGACGCCCCAGCACTCGTC
CGAGGGCAAAGGAATAGCACGTGCTACGAGATTTTCGAT²CCACCGCCGCTTCTATGAAAGGTTGGGC
TTCGGAATCGTTTTCCGGGACGCTGGCTGGATGAICCTCCAGCGCGGGGATCTCATGCTGGAGTTCIT
CGCCCACCCCAACTTGTTTATTGCAGCTTATAATGGTTACAAATAAAGCAATAGCATCACAAATTTCA
CAAATAAAGCATTTTTTCACTGCATTCTAGTTGIGGTTTGTCCAACTCATCAATGTATCTTATCAT
GTCTGTATACCGTCGACCTCTAGCTAGAGCTTGGCGTAATCAIGGTCATAGCTGTTTCTGTGTGAA
TTGTTATCCGCTCACAAATCCACACAACATACGAGCCGGAAGCATAAAGTGTAAGCCTGGGGTGCCT
AATGAGTGAGCTAACTCACATTAATTGCGTTGCGCTCACTGCCCGCTTTCCAGTCGGGAAACCTGTG
TGCCAGCTGCATTAATGAATCGGCCAACGCGCGGGGAGAGGCGGTTTTCGATATTGGGCGCTCTTCCGC
TTCCTCGCTCACTGACTCGCTGCGCTCGGTCGTTTCGGCTGCGGCGAGCGGTATCAGCTCACTCAAAGG
CGGTAATACGGTTATCCACAGAATCAGGGGATAACGCAGGAAAGAATGTGAGCAAAGGCCAGCAA
AAGGCCAGGAACCGTAAAAAGGCCGCGTTGCTGGCGTT³TTCCATAGGCTCCGCCCCCTGACGAGCA
TCACAAAAATCGACGCTCAAGTCAGAGGTGGCGAAACCCGACAGGACTATAAAGATACCAGGCGTTTC
CCCCGGAAGCTCCCTCGTGCGCTCTCCTGTTCCGACCTGCCGCTTACCGGATACCTGTCCGCCTTT
CTCCCTTCGGGAAGCGTGGCGCTTCTCATAGCTCACGCTGTAGGTATCTCAGTTCGGTGTAGGTCGT
TCGCTCCAAGCTGGGCTGTGTGCACGAACCCCCCGTTACGCCCGACCGCTGCGCCTTATCCGGTAACT
ATCGTCTTGAGTCCAACCCGTAAGACACGACTTATCGCCACTGGCAGCAGCCACTGGTAAACAGGATT
AGCAGAGCGAGGTATGTAGGCGGTGCTACAGAGTTCCTGAAGTGGTGGCCTAACTACGGCTACACTAG
AAGAACAGTATTTGGTATCTGCGCTCTGCTGAAGCCAGTTACCTTCGGAAAAAGAGTTGGTAGCTCTT
GATCCGGCAAACAAACCACCGCTGGTAGCGGTGGTTTTTTGTTTGCAAGCAGCAGATTACGCGCAGA
AAAAAAGGATCTCAAGAAGATCCTTTGATCTTTTCTACGGGGTCTGAGCGCGGAACCCCTATTTGTTT
ATTTTTCTAAATACATTCAAATATGTATCCGCTCATGAATTAATTCTTAGAAAACTCATCGAGCATC
AAATGAACTGCAATTTATTCATATCAGGATTATCAATACCAIATTTTTGAAAAAGCCGTTTCTGTAA
TGAAGGAGAAAACTCACCGAGGCAGTTCCATAGGATGGCAAGATCCTGGTATCGGTCTGCGATTCCGA
CTCGTCCAACATCAATACAACCTATTAATTTCCCCTCGTCAAAAATAAGGTTATCAAGTGAGAAATCA
CCATGAGTGACGACTGAATCCGGTGAGAAATGGCAAAGTTTAIGCATTTCTTCCAGACTTGTTCAAC
AGGCCAGCCATTACGCTCGTCATCAAATCACTCGCATCAACCAAACCGTTATTCATTCTGTGATTGCG
CCTGAGCGAGACGAAATACGCGATCGCTGTAAAAGGACAATTACAAACASGAATCGAATGCAACCGG
CGCAGGAACACTGCCAGCGCATCAACAATATTTTACCTGAATCAGGATAITCTTCTAATACCTGGAA
TGCTGTTTTCCAGGGATCGCAGTGGTGAGTAACCATGCATCATCAGGAGTACGGATAAAATGCTTGA
TGGTCGGAAGAGGCATAAATTCGTCAGCCAGTTIAGTCTGACCATCTCAICTGTAAACATCATTTGGCA
ACGCTACCTTTGCCATGTTTCAGAAACAAC⁴CTGCGCATCGGGCTTCCCATACAATCGATAGATTGT
CGCACCTGATTGCCGACATTATCGCGAGCCCATITATACCCATATAAATCAGCATCCATGTTGGAAT
TTAATCGCGGCTAGAGCAAGACGTTTCCGTTGAATATGGCTCATAACACCCCTTGATTACTGTIT
ATGTAAGCAGACAGTTTTATTGTTTCATGACCAAAATCCCTAACGTGAGTTTTCGTTCCACTGAGCGT
CAGACCCCGTAGAAATCCGCGCACATTTCCCGAAAAGTGCCACCTGACGTC (SEQ ID NO: 44)

[0488]

包含内含子的编码序列导致以下截短的蛋白：

[0489] MDEADRRLLRRCLRLVEELQVDQLWDALLSRELFRPHMIEDIQRAGSGSRRDQARQLIIDLETRGSQ
ALPLFISCLEDTGQDMLASFLRTNRQAAKLSKPTLENLTPVVLRPEIRKPEVLRPETPRPVDIGSGGF
GDVGALESRLGNADLAY ILSMEPCGHCLI INNVNFCRESGLRTRTGSNIDCEKLRRRFSSLHFMVEVK
GDLTAKKMVLALLELAQQDHGALDCCVVVILSHGCQASHLQFFGAVYGTGCPVSVEKIVNIFNGTSC
PSLGGKPKLFFIQACGGEQKDHGFEVASTSPEDESPGSNPEPDATPFQEGRLTFDQLDAISSLPTPSD
IFVSYSTFPGFVSWRDPKSGSWAGDPGRHL (SEQ ID NO: 45)

[0490] 去除内含子允许胱天蛋白酶9蛋白完全翻译:

MDEADRRLLRRCLRLVEELQVDQLWDALLSRELFRPHMIEDIQRAGSGSRRDQARQLIIDLETRGSQ
ALPLFISCLEDTGQDMLASFLRTNRQAAKLSKPTLENLTPVVLRPEIRKPEVLRPETPRPVDIGSGGF
GDVGALESRLGNADLAY ILSMEPCGHCLI INNVNFCRESGLRTRTGSNIDCEKLRRRFSSLHFMVEVK
[0491] GDLTAKKMVLALLELAQQDHGALDCCVVVILSHGCQASHLQFFGAVYGTGCPVSVEKIVNIFNGTSC
PSLGGKPKLFFIQACGGEQKDHGFEVASTSPEDESPGSNPEPDATPFQEGRLTFDQLDAISSLPTPSD
IFVSYSTFPGFVSWRDPKSGSWADGCLYTTLLETLDLFEQWAHSEDLQSLLLRVANAVSVKGIYKQ
MPGCFNFLRKKLFFKTSGGGGS (SEQ ID NO: 46)

[0492] 如前所述进行转染。添加2mM的DTT,添加10 μ M的毛喉素。进行后者以确定毛喉素是否可诱导内含子的去除。毛喉素具有潜在的基因治疗应用等。内含子去除的诱导被测量为细胞死亡的函数,较高的细胞死亡指示了内含子去除。

[0493] 如方法中所述进行实验,即用上述构建体转染HEK或CHO-s细胞并孵育24小时。在这段时间之后,添加活化化合物DTT或毛喉素,然后在诱导之后3、5和24小时测量%细胞死亡。

[0494] 使用Countess II细胞计数器(Thermofisher),将内含子的诱导测量为细胞死亡的函数。用台盼蓝将细胞染色以确定细胞生存力。将10 μ l细胞悬液添加至血细胞计数器并在countess上进行分析。

[0495] 如图14A所示,在诱导之前没有细胞死亡,然而在添加任何一种诱导物之后,死细胞的%迅速提高,实际上在DTT处理3小时之后,死细胞的数量接近于不含内含子的构建体的对照。实际上,在DTT处理之后5小时,死细胞的%等于阳性对照,而毛喉素处理仅在24小时之后达到该水平,这表明它是弱的剪接诱导物。没有添加诱导物的细胞几乎没有或没有细胞死亡,这表明直到添加诱导物胱天蛋白酶9才表达。这表明,由于CMV-IE启动子是组成型的,因此不断产生转录物,因此来自其的对基因表达的控制非常严密(实际上明显是绝对的),并且是在mRNA水平上。图14B还示出了来自内含子的对基因表达的控制非常严密。这是因为胱天蛋白酶9与称为SPARK的GFP变体融合,因此胱天蛋白酶9的表达将导致GFP变体的表达。该图示出了直到添加诱导物并且细胞死亡开始,GFP才表达。

[0496] 在另一实验中,使用与上述相同的构建体(SYNP-CASP9-INT),并测量细胞死亡速率。从该图可看出,约50%的细胞在诱导的1小时内死亡,这表明反应迅速且受到严密控制。本实验的结果如图15所示。

[0497] 实施例8-使用不同的最小启动子的UPR诱导型启动子

[0498] 本实验的目的是用不同的最小启动子测试上述实施例中使用的含ATF6的UPR响应性顺式调节元件(在本文中也称为增强子),并进一步评估其控制的诱导性和严密性。为此,将含有ATF6的UPR响应性顺式调节元件与CMV-MP和MinTK(疱疹胸苷激酶最小启动子)最小启动子有效连接。

[0499] 如上所述,通过GeneART通过化学合成来合成构建体。

[0500] 如上所述,增强子序列包含序列TGACGTGCT(其包含ATF6共有序列TGACGTG)的6个

重复序列,其间隔20bp的间隔序列。增强子序列与MinTK (在ATF06-MP001构建体中) 或CMV-MP (在ATF06-MP002构建体中) 偶联。

[0501] 如前所述进行HEK293-F细胞的转染。以2mM的浓度添加DTT。如前所述测量萤光素酶活性。结果如图17所示-DTT是在时间=0小时时添加的,如X轴所示(单位为小时)。

[0502] 在添加DTT之前,两种构建体均显示出良好的诱导性和可忽略的表达。与包含MinTK的启动子相比,包含CMV-MP的启动子观察到稍高的表达。然而,两种启动子构建体在通过DTT诱导UPR后均显示高表达水平,明显高于由组成型CMV-IE启动子提供的表达水平。

[0503] ATF06-MP-001 (带下划线的ATF06和MinTK启动子序列)

[0504]

GAAGTTCCTATTCCGAAGTTCCTATTCTCTAGAAAGTATAGGAACTTCTGACGTGCTGATGATGCGTA
GCTAGTAGTTGACGTGCTGATGATGCGTAGCTAGTAGTTGACGTGCTGATGATGCGTAGCTAGTAGTT
GACGTGCTGATGATGCGTAGCTAGTAGTTGACGTGCTGATGATGCGTAGCTAGTAGTTGACGTGCTGA
TGATGCGTAGCTAGTAGTGCAGTTAGCGTAGCTGAGGTACCGTCGACGATATCGGATCCTTCGCATAT
TAAGGTGACGCGTGTGGCCTCGAACACCGAGCGACCCCTGCAGCGACCCGCTTAAATGGAAGATGCCAA
AAACATTAAGAAGGGCCCAGCGCCATTCTACCCACTCGAAGACGGGACCGCCGGCGAGCAGCTGCACA
AAGCCATGAAGCGCTACGCCCTGGTGCCCGGCACCATCGCCTTTACCGACGCACATATCGAGGTGGAC
ATTACCTACGCCGAGTACTTCGAGATGAGCGTTCCGGCTGGCAGAAGCTATGAAGCGCTATGGGCTGAA
TACAAACCATCGGATCGTGGTGTGCAGCGAGAATAGCTTGCAGTTCTTCATGCCCGTGTGGGTGCCC
TGTTTCATCGGTGTGGCTGTGGCCCCAGCTAACGACATCTACAACGAGCGCGAGCTGCTGAACAGCATG
GGCATCAGCCAGCCCACCGTCGTATTCGTGAGCAAGAAAGGGCTGCAAAGATCCTCAACGTGCAAAA
GAAGCTACCGATCATACAAAAGATCATCATCATGGATAGCAAGACCGACTACCAGGGCTTCCAAAGCA
TGTAACACCTTCGTGACTTCCCATTGTCACCCGGCTTCAACGAGTACGACTTCGTGCCCGAGAGCTTC
GACCGGGACAAAACCATCGCCCTGATCATGAACAGTAGTGGCAGTACCGGATTGCCCAAGGGCGTAGC
CCTACCGCACCGCACCGCTTGTGTCCGATTTCAGTCATGCCCGCGACCCCATCTTCGGCAACCAGATCA
TCCCCGACACCGCTATCCTCAGCGTGGTGCCATTTACCACGGCTTCGGCATGTTACCACGCTGGGC
TACTTGATCTGCGGCTTTCGGCTCGTGCTCATGTACCGCTTCGAGGAGGAGCTATTCTTCGCGAGCTT
GCAAGACTATAAGATTCAATCTGCCCTGCTGGTGCCACACTATTTAGCTTCTTCGCTAAGAGCACTC
TCATCGACAAGTACGACCTAAGCAACTTGCACGAGATCGCCAGCGGGCGGGCGCCGCTCAGCAAGGAG
GTAGGTGAGGCGGTGGCCAAACGCTTCCACCTACCAGGCATCCGCCAGGGCTACGGCCTGACAGAAAC
AACCAGCGCCATTCTGATCACCCCCGAAGGGGACGACAAGCCTGGCGCAGTAGGCAAGGTGGTGCCCT
TCTTCGAGGCTAAGGTGGTGGACTTGGACACCGGTAAGACACTGGGTGTGAACCAGCGCGGCGAGCTG
TGCGTCCGTGGCCCCATGATCATGAGCGGTACGTTAACAACCCGAGGCTACAAACGCTCTCATCGA
CAAGGACGGCTGGCTGCACAGCGGCGACATCGCCTACTGGGACGAGGACGAGCACTTCTTCATCGTGG
ACCGGCTGAAGAGCCTGATCAAATACAAGGGCTACCAGGTAGCCCCAGCCGAAGTGGAGAGCATCCTG
CTGCAACACCCCAACATCTTCGACGCGGGGTGCGCGGCTGCCCGACGACGATGCCGCGGAGCTGCC
CGCCGAGTCGTCTGTGGAACACGGTAAAACCATGACCGAGAAGGAGATCGTGGACTATGTGGCCA
GCCAGGTTACAACCGCCAAGAAGCTGCGCGGTGGTGTGTGTTCGTGGACGAGGTGCCTAAAGGACTG
ACCGGCAAGTTGGACGCCCCGAAGATCCGCGAGATTCTCATTAAGGCCAAGAAGGGCGGCAAGATCGC
CGTGTAATGAAAGCTTGGTCTCTACGAGTAATAGACGCCAGTTGAATTCCTTCGAGCAGACATGATA
AGATACATTGATGAGTTTGGACAAACCACAACCTAGAATGCAGTGAAAAAATGCTTTATTTGTGAAAT
TTGTGATGCTATTGCTTTATTTGTAACCATTATAAGCTGCAATAAACAAGTTAACAACAACAATTGCA

TTCATTTTATGTTTCAGGTTTCAGGGGGAGGTGTGGGAGGTTTTTTAAAGCAAGTAAAACCTCTACAA
TGTGGTAAAATCGATAAGGATCCGT (SEQ ID NO: 59)

[0505]

ATF06-MP-002 (带下划线的ATF06和CMV-MP启动子序列)

CGGACGTAATACGACTCACTATAGGGCGAATTGGCGGAAGGCCGTCAAGGCCGCATGCATAA7AAAAT
ATCTTTATTTTCATTACATCTGTGTGTTGGTTTTTGTGTGTGACGTGCTGATGATGCGTAGCTAGTA
GTTGACGTGCTGATGATGCGTAGCTAGTAGTTGACGTGCTGAIGATGCGTAGCTAGTAGTTGACGTGC
TGATGATGCGTAGCTAGTAGTTGACGTGCTGATGATGCGTAGCTAGTAGTTGACGTGCTGATGATGCG
TAGCTAGTAGTGACGTTAGCGTAGCTGAGGTACCGTCGACGATATCGGATCCAGGTCTATATAAGCAG
AGCTCGTTTAGTGAACCGTCAGATCGCCTAGATACGCCATCCACGCTGTTTTGACCTCCATAGAAGAT
CGCCACCATGGAAGATGCCAAAAACATTAAAGAGGGCCAGCGCCATTCTACCCACTCGAAGACGGGA
CCGCCGGCGAGCAGCTGCACAAAGCCATGAAGCGCTACGCCCTGGTGCCCGGCACCATCGCCTTTACC
GACGCACATATCGAGGTGGACATTACCTACGCCGAGTACTTCGAGATGAGCGTTCGGCTGGCAGAAGC
TATGAAGCGCTATGGGCTGAATACAAACCATCGGATCGTGGTGTGCAGCGAGAATAGCTTGCAGTTCT
TCATGCCCCGTGTTGGGTGCCCTGTTTCATCGGTGTGGCTGTGGCCCCAGCTAACGACATCTACAACGAG
CGCGAGCTGCTGAACAGCATGGGCATCAGCCAGCCACCGTCGTATTTCGTGAGCAAGAAAGGGCTGCA
AAAGATCCTCAACGTGCAAAAGAAGCTACCGATCATACAAAGATCATCATCATGGATAGCAAGACCG
ACTACCAGGGCTTCCAAAGCATGTACACCTTCGTGACTTCCCATTTCGCCACCCGGCTTCAACGAGTAC
GACTTCGTGCCCCGAGAGCTTCGACCGGGACAAAACCATCGCCCTGATCATGAACAGTAGTGGCAGTAC
CGGATTGCCCAAGGGCGTAGCCCTACCGCACCGCACCGCTTGTGTCCGATTTCAGTCATGCCCCGACCC
CCATCTTCGGCAACCAGATCATCCCCGACACCGCTATCCTCAGCGTGGTGCCATTTACACACGGCTTC
GGCATGTTACACACGCTGGGCTACTTGATCTGCGGCTTTCGGGTCGTGCTCATGTACCGCTTCGAGGA
GGAGCTATTCTTGCGCAGCTTGCAAGACTATAAGATTCAATCTGCCCTGCTGGTGCCACACTATTTA
GCTTCTTCGCTAAGAGCACTCTCATCGACAAGTACGACCTAAGCAACTTGACAGAGATCGCCAGCGGC
GGGGCGCCGCTCAGCAAGGAGGTAGGTGAGGCGGTGGCCAAACGCTTCCACCTACCAGGCATCCGCCA
GGGCTACGGCCTGACAGAAACAACCAGCGCCATTCTGATCACCCCCGAAGGGGACGACAAGCCTGGCG
CAGTAGGCAAGGTGGTGCCCTTCTTCGAGGCTAAGGTGGTGGACTTGGACACCGGTAAGACACTGGGT
GTGAACCAGCGCGGCGAGCTGTGCGTCCGTGGCCCCATGATCATGAGCGGCTACGTTAACAACCCCCGA
GGCTACAAACGCTCTCATCGACAAGGACGGCTGGCTGCACAGCGGCACATCGCCTACTGGGACGAGG
ACGAGCACTTCTTCATCGTGGACCGGTGAAGAGCCTGATCAAATACAAGGGCTACCAGGTAGCCCCA
GCCGAACCTGGAGAGCATCCTGCTGCAACACCCCAACATCTTCGACGCCGGGGTCGCCGGCCTGCCCGA
CGACGATGCCGGCGAGCTGCCCCGCCGAGTCGTCTGCTGGAACACGGTAAAACCATGACCGAGAAGG
AGATCGTGGACTATGTGGCCAGCCAGGTTACAACCGCCAAGAAGCTGCGCGGTGGTGTGTGTTCTGTG
GACGAGGTGCCTAAAGGACTGACCGGCAAGTTGGACGCCCGCAAGATCCGCGAGATTCTCATTAAGGC
CAAGAAGGGCGCAAGATCGCCGTGTAATGAAAGCTTGGTCTCTACGAGTAATAGACGCCAGTTGAA
TTCCTTCGAGCAGACATGATAAGATACATTGATGAGTTTGGACAAACCACAACCTAGAATGCAGTGAAA
AAAATGCTTTATTTGTGAAATTTGTGATGCTATTGCTTTATTIGTAACCATTATAAGCTGCAATAAAC

AAGTTAACAACAACAATTGCATTCATTTTATGTTTCAGGTTTCAGGGGGAGGTGTGGGAGGTTTTTTAA
AGCAAGTAAACCTCTACAAATGTGGTAAAATCGATAAGGATCCGTCTGGGCCTCATGGGCCTTCCGC
TCACTGCCCCGCTTTCAGTCGGGAAACCTGTGCTGCCAGCTGCAT (SEQ ID NO: 60)

[0507] 尽管上面详细讨论了本发明的多种实施方案的制造和使用,但是应当理解,本发明提供了许多可应用的发明构思,这些构思可在广泛多种特定的上下文中体现。本文讨论的特定实施方案仅是制造和使用本发明的特定方式的举例说明,并不限制本发明的范围。

	<110> Synpromics Limited	
	<120> 使用可调节内含子的表达控制	
	<130> P241689W0	
	<150> GB1713545.0	
	<151> 2017-08-23	
	<160> 61	
	<170> PatentIn version 3.5	
	<210> 1	
	<211> 20	
	<212> RNA	
	<213> 未知	
	<220>	
	<223> IRE 1切割位点内的哺乳动物XBP1内含子的序列	
	<400> 1	
	cacucagacu acgugcaccu	20
	<210> 2	
	<211> 20	
	<212> RNA	
	<213> 未知	
	<220>	
	<223> IRE 1切割位点内的哺乳动物XBP1内含子的变体	
	<400> 2	
	cacucagacu acgugcuccu	20
[0001]	<210> 3	
	<211> 20	
	<212> RNA	
	<213> 未知	
	<220>	
	<223> IRE 1切割位点内的哺乳动物XBP1内含子的变体	
	<400> 3	
	cacucagacu acgugccuccu	20
	<210> 4	
	<211> 20	
	<212> RNA	
	<213> 未知	
	<220>	
	<223> IRE 1切割位点内的哺乳动物XBP1内含子的变体	
	<400> 4	
	cacucagacu acgugcgccu	20
	<210> 5	
	<211> 20	
	<212> RNA	
	<213> 未知	
	<220>	
	<223> IRE 1切割位点内的哺乳动物XBP1内含子的变体	
	<400> 5	
	cacucagacu augugcaccu	20
	<210> 6	

	<211> 33	
	<212> RNA	
	<213> 人工序列	
	<220>	
	<223> 可调节内含子	
	<220>	
	<221> misc_feature	
	<222> (2)..(2)	
	<223> n是a、c、g或u	
	<220>	
	<221> misc_feature	
	<222> (5)..(5)	
	<223> n是a、c、g或u	
	<220>	
	<221> misc_feature	
	<222> (28)..(28)	
	<223> n是a、c、g或u	
	<220>	
	<221> misc_feature	
	<222> (31)..(31)	
	<223> n是a、c、g或u	
	<400> 6	
	cngcngcacu cagacuacgu gcaccucngc ngc	33
	<210> 7	
	<211> 33	
	<212> RNA	
	<213> 未知	
[0002]	<220>	
	<223> 哺乳动物XBP1内含子序列	
	<400> 7	
	cagcagcacu cagacuacgu gcaccucugc ugc	33
	<210> 8	
	<211> 32	
	<212> RNA	
	<213> 人工序列	
	<220>	
	<223> 可调节内含子	
	<220>	
	<221> misc_feature	
	<222> (2)..(2)	
	<223> n是a、c、g或u	
	<220>	
	<221> misc_feature	
	<222> (31)..(31)	
	<223> n是a、c、g或u	
	<400> 8	
	cngcagcacu cagacuacgu gcaccucugc ng	32
	<210> 9	
	<211> 32	
	<212> RNA	
	<213> 人工序列	
	<220>	
	<223> 可调节内含子	

	<220>		
	<221>	misc_feature	
	<222>	(2)..(2)	
	<223>	n是a、c、g或u	
	<220>		
	<221>	misc_feature	
	<222>	(31)..(31)	
	<223>	n是a、c、g或u	
	<400>	9	
	cngcagcacu	cagacuacgu	gcuccucugc ng 32
	<210>	10	
	<211>	32	
	<212>	RNA	
	<213>	人工序列	
	<220>		
	<223>	可调节内含子	
	<220>		
	<221>	misc_feature	
	<222>	(2)..(2)	
	<223>	n是a、c、g或u	
	<220>		
	<221>	misc_feature	
	<222>	(31)..(31)	
	<223>	n是a、c、g或u	
	<400>	10	
	cngcagcacu	cagacuacgu	gccccucugc ng 32
[0003]	<210>	11	
	<211>	32	
	<212>	RNA	
	<213>	人工序列	
	<220>		
	<223>	可调节内含子	
	<220>		
	<221>	misc_feature	
	<222>	(2)..(2)	
	<223>	n是a、c、g或u	
	<220>		
	<221>	misc_feature	
	<222>	(31)..(31)	
	<223>	n是a、c、g或u	
	<400>	11	
	cngcagcacu	cagacuacgu	gcgccucugc ng 32
	<210>	12	
	<211>	32	
	<212>	RNA	
	<213>	人工序列	
	<220>		
	<223>	可调节内含子	
	<220>		
	<221>	misc_feature	
	<222>	(2)..(2)	
	<223>	n是a、c、g或u	
	<220>		
	<221>	misc_feature	

	<222> (31)..(31)	
	<223> n是a、c、g或u	
	<400> 12	
	cngcagcacu cagacuaugu gcaccucugc ng	32
	<210> 13	
	<211> 33	
	<212> RNA	
	<213> 人工序列	
	<220>	
	<223> 可调节内含子	
	<400> 13	
	cagcagcacu cagacuacgu gcuccucugc ugc	33
	<210> 14	
	<211> 33	
	<212> RNA	
	<213> 人工序列	
	<220>	
	<223> 可调节内含子	
	<400> 14	
	cagcagcacu cagacuacgu gccccucugc ugc	33
	<210> 15	
	<211> 33	
	<212> RNA	
	<213> 人工序列	
[0004]	<220>	
	<223> 可调节内含子	
	<400> 15	
	cagcagcacu cagacuacgu gcgccucugc ugc	33
	<210> 16	
	<211> 33	
	<212> RNA	
	<213> 人工序列	
	<220>	
	<223> 可调节内含子	
	<400> 16	
	cagcagcacu cagacuaugu gcaccucugc ugc	33
	<210> 17	
	<211> 35	
	<212> RNA	
	<213> 人工序列	
	<220>	
	<223> 具有添加的三核苷酸CUG的哺乳动物XBP1内含子序列	
	<400> 17	
	cagcugcagc acucagacua cgugcaccuc ugacag	35
	<210> 18	
	<211> 19	
	<212> DNA	
	<213> 未知	
	<220>	
	<223> ERSE1位点	

[0005]	<220>			
	<221>	misc_feature		
	<222>	(6)..(14)		
	<223>	n是a、c、g或t		
	<400>	18		
		ccaatnnnnn nnnnccacg	19	
	<210>	19		
	<211>	11		
	<212>	DNA		
	<213>	未知		
	<220>			
	<223>	ERSE2位点		
	<220>			
	<221>	misc_feature		
	<222>	(6)..(6)		
	<223>	n是a、c、g或t		
	<400>	19		
		attggnccac g	11	
	<210>	20		
	<211>	54		
	<212>	DNA		
	<213>	未知		
	<220>			
	<223>	UPR诱导型启动子序列		
	<400>	20		
		tgacgtgctt gacgtgcttg acgtgcttga cgtgcttgac gtgcttgacg tgct	54	
	<210>	21		
	<211>	84		
	<212>	DNA		
	<213>	未知		
	<220>			
	<223>	CMV最小启动子		
	<400>	21		
		aggtctatat aagcagagct cgtttagtga accgtcagat cgcctagata cgccatccac	60	
		gctgttttga cctccataga agat	84	
	<210>	22		
	<211>	169		
	<212>	DNA		
	<213>	未知		
	<220>			
	<223>	诱导型启动子		
	<400>	22		
		tgacgtgctt gacgtgcttg acgtgcttga cgtgcttgac gtgcttgacg tgctgtgacc	60	
		gtcgacgata tcggatccag gtctatataa gcagagctcg tttagtgaac cgtcagatcg	120	
		cctagatacg ccatccacgc tgttttgacc tccatagaag atcgccacc	169	
	<210>	23		
	<211>	26		
	<212>	RNA		
	<213>	未知		
	<220>			
	<223>	哺乳动物WT内含子序列		

	<400> 23 cagcacucag acuacgugca ccucug	26
	<210> 24 <211> 33 <212> RNA <213> 未知	
	<220> <223> 侧翼为剪接识别位点序列的哺乳动物WT内含子序列	
	<400> 24 ccgcagcacu cagacuacgu gcaccucugc agc	33
	<210> 25 <211> 717 <212> DNA <213> 人工序列	
	<220> <223> 具有内含子插入的EGFP基因	
[0006]	<400> 25 atggtgagca agggcgagga gctgttcacc ggggtggtgc ccatcctggt cgagctggac	60
	ggcgacgtaa acggccacaa gttcagcgtg tccggcgagg gcgagggcga tgccacctac	120
	ggcaagctga ccctgaagtt catctgcacc accggcaagc tgcccgtgcc ctggcccacc	180
	ctcgtgacca ccctgacctg cggcgtgcag tgcttcagcc gctaccccgga ccacatgaag	240
	cagcacgact tcttcaagtc cgccatgccc gaaggctacg tccaggagcg caccatcttc	300
	ttcaaggacg acggcaacta caagaccgcg gccgaggtga agttcgaggg cgacaccctg	360
	gtgaaccgca tcgagctgaa gggcatcgac ttcaaggagg acggcaacat cctggggcac	420
	aagctggagt acaactacaa cagccacaac gtctatatca tggccgacaa gcagaagaac	480
	ggcatcaagg tgaacttcaa gatccgccac aacatcgagg acggcagcgt gcagctcgcc	540
	gaccactacc agcagaacac ccccatcggc gacggccccc tgctgctgcc cgacaaccac	600
	tacctgagca cccagtccgc cctgagcaaa gaccccaacg agaagcgaga tcacatggtc	660
	ctgctggagt tcgtgaccgc cgccgggata actctcggca tggacgagct gtacaag	717
	<210> 26 <211> 239 <212> PRT <213> 未知	
	<220> <223> EGFP 蛋白质序列	
	<400> 26 Met Val Ser Lys Gly Glu Glu Leu Phe Thr Gly Val Val Pro Ile Leu 1 5 10 15	
	Val Glu Leu Asp Gly Asp Val Asn Gly His Lys Phe Ser Val Ser Gly 20 25 30	
	Glu Gly Glu Gly Asp Ala Thr Tyr Gly Lys Leu Thr Leu Lys Phe Ile 35 40 45	
	Cys Thr Thr Gly Lys Leu Pro Val Pro Trp Pro Thr Leu Val Thr Thr 50 55 60	

	Leu Thr Tyr Gly Val Gln Cys Phe Ser Arg Tyr Pro Asp His Met Lys 65 70 75 80	
	Gln His Asp Phe Phe Lys Ser Ala Met Pro Glu Gly Tyr Val Gln Glu 85 90 95	
	Arg Thr Ile Phe Phe Lys Asp Asp Gly Asn Tyr Lys Thr Arg Ala Glu 100 105 110	
	Val Lys Phe Glu Gly Asp Thr Leu Val Asn Arg Ile Glu Leu Lys Gly 115 120 125	
	Ile Asp Phe Lys Glu Asp Gly Asn Ile Leu Gly His Lys Leu Glu Tyr 130 135 140	
	Asn Tyr Asn Ser His Asn Val Tyr Ile Met Ala Asp Lys Gln Lys Asn 145 150 155 160	
	Gly Ile Lys Val Asn Phe Lys Ile Arg His Asn Ile Glu Asp Gly Ser 165 170 175	
	Val Gln Leu Ala Asp His Tyr Gln Gln Asn Thr Pro Ile Gly Asp Gly 180 185 190	
	Pro Val Leu Leu Pro Asp Asn His Tyr Leu Ser Thr Gln Ser Ala Leu 195 200 205	
[0007]	Ser Lys Asp Pro Asn Glu Lys Arg Asp His Met Val Leu Leu Glu Phe 210 215 220	
	Val Thr Ala Ala Gly Ile Thr Leu Gly Met Asp Glu Leu Tyr Lys 225 230 235	
	<210> 27 <211> 36 <212> RNA <213> 人工序列	
	<220> <223> 具有添加的三核苷酸CUG的哺乳动物XBP1内含子序列	
	<400> 27 cagcugcagc acucagacua cgugcaccuc ugcugg	36
	<210> 28 <211> 1150 <212> DNA <213> 人工序列	
	<220> <223> 具有可调节的XBP1内含子、CMV-MP和SV40 polyA尾的EGFP序列	
	<400> 28 ggtaccgtcg acgatatcgg atccaggtct atataagcag agctcgttta gtgaaccgtc agatcgccata gatacgccat ccacgtgtt ttgacctcca tagaagatcg ccaccatggt gagcaagggc gaggagctgt tcaccggggt ggtgcccatc ctggtcgagc tggacggcga cgtaaacggc cacaagttca gcgtgtccgg cgaggcgag ggcgatgcc cctacggcaa gctgaccctg aagttcatct gcaccaccgg caagctgccc gtgccctggc ccaccctcgt	60 120 180 240 300

```

gaccacccctg acctacggcg tgcagtgtt cagccgtac cccgaccaca tgaagcagca 360
cgacttcttc aagtccgccg tgcccgaagg ctacgtccag gagcgcacca tcttcttcaa 420
ggacgaagcg aactacaaga cccgcgccga ggtgaagttc gagggcgaca ccctggtgaa 480
ccgcacgcag ctgaagggca tcgacttcaa ggaggacggc aacatcctgg ggcacaagct 540
ggagtacaac tacaacagcc acaacgtcta tatcatggcc gacaagcaga agaacggcat 600
caaggtgaac ttcaagatcc gccacaacat cgaggacggc agcgtgcagc tgcagcactc 660
agactacgtg cacctctgct ggccgaccac taccagcaga acacccccat cggcgcggcg 720
cccgctgtgc tgcccgacaa cactacctg agcaccagc cggccctgag caaagacccc 780
aacgagaagc gcatcacat ggtcctgtg gagttcgtga ccgccgccg gatcactctc 840
ggcatggacg agctgtacaa gtgagacgat cttatcgga tttaccaca tttgtagagg 900
ttttacttgc tttaaaaaac ctccacatc tcccctgaa cctgaaacat aaaatgaatg 960
caattgttgt tgtaacttg tttatgacg cttataatgg ttacaaataa agcaatagca 1020
tcacaaatit cacaaataaa gcattttttt cactgcattc tagttgtggt ttgtccaaac 1080
tcatcaatgt atcttatcat gtctgtcga aggcgcgcc tgggcctcat gggccttcg 1140
ctcactgccc 1150

```

<210> 29
 <211> 211
 <212> PRT
 <213> 人工序列

[0008] <220>
 <223> 截短的EGFP蛋白 - 开始至第一终止密码子

<400> 29

Met Val Ser Lys Gly Glu Glu Leu Phe Thr Gly Val Val Pro Ile Leu
 1 5 10 15

Val Glu Leu Asp Gly Asp Val Asn Gly His Lys Phe Ser Val Ser Gly
 20 25 30

Glu Gly Glu Gly Asp Ala Thr Tyr Gly Lys Leu Thr Leu Lys Phe Ile
 35 40 45

Cys Thr Thr Gly Lys Leu Pro Val Pro Trp Pro Thr Leu Val Thr Thr
 50 55 60

Leu Thr Tyr Gly Val Gln Cys Phe Ser Arg Tyr Pro Asp His Met Lys
 65 70 75 80

Gln His Asp Phe Phe Lys Ser Ala Met Pro Glu Gly Tyr Val Gln Glu
 85 90 95

Arg Thr Ile Phe Phe Lys Asp Asp Gly Asn Tyr Lys Thr Arg Ala Glu
 100 105 110

Val Lys Phe Glu Gly Asp Thr Leu Val Asn Arg Ile Glu Leu Lys Gly
 115 120 125

Ile Asp Phe Lys Glu Asp Gly Asn Ile Leu Gly His Lys Leu Glu Tyr
 130 135 140

Asn Tyr Asn Ser His Asn Val Tyr Ile Met Ala Asp Lys Gln Lys Asn
145 150 155 160

Gly Ile Lys Val Asn Phe Lys Ile Arg His Asn Ile Glu Asp Gly Ser
165 170 175

Val Gln Leu Glu His Ser Asp Tyr Val His Leu Cys Trp Pro Thr Thr
180 185 190

Thr Ser Arg Thr Pro Pro Ser Ala Thr Ala Pro Cys Cys Cys Pro Thr
195 200 205

Thr Thr Thr
210

<210> 30
<211> 889
<212> DNA
<213> 人工序列

<220>
<223> SYNP-ATF6-02构建体

[0009]

<400> 30
tgacgtgctt gacgtgcttg acgtgcttga cgtgcttgac gtgcttgacg tgctggtacc 60
gtcgacgata tcggatccag gtctatataa gcagagctcg tttagtgaac cgtcagatcg 120
cctagatacg ccatccacgc tgttttgacc tccatagaag atcgccacca tggtagagcaa 180
gggcgaggag ctgttcacgc ggggtggtgcc catcctggtc gagctggacg gcgacgtaaa 240
cgccacaag ttcagcgtgt ccggcgaggc cgaggcgat gccacctacg gcaagctgac 300
cctgaagtgc atctgcacca ccggcaagct gcccggtccc tggccacccc tcgtgaccac 360
cctgacctac ggctgtcagt gcttcagccg ctaccccgac cacatgaagc agcacgactt 420
cttcaagtcc gccatgcccg aaggctacgt ccaggagcgc accatcttct tcaaggacga 480
cggcaactac aagaccgcg ccgaggtgaa gttcgagggc gacacctgg tgaaccgcat 540
cgagctgaag ggcatcgact tcaaggagga cggcaacatc ctggggcaca agctggagta 600
caactacaac agccacaacg tctatatcat ggccgacaag cagaagaacg gcatcaaggt 660
gaacttcaag atccgccaca acatcgagga cggcagcgtg cagctcgccg accactacca 720
gcagaacacc cccatcgccg acggccccgt gctgctgccc gacaaccact acctgagcac 780
ccagtccgcc ctgagcaaaag accccaacga gaagcgcgat cacatggtcc tgctggagtt 840
cgtgaccgcc gccgggatca ctctcggcat ggacgagctg tacaagtga 889

<210> 31
<211> 845
<212> DNA
<213> 人工序列

<220>
<223> 具有切除的内含子序列的SYNP-ATF6-02

<400> 31
tgacgtgctt gacgtgcttg acgtgcttga cgtgcttgac gtgcttgacg tgctggtacc 60
gtcgacgata tcggatccag gtctatataa gcagagctcg tttagtgaac cgtcagatcg 120
cctagatacg ccatccacgc tgttttgacc tccatagaag atcgccacca tggtagagcaa 180

```

gggcgaggag ctgttcaccg ggggtgggtgcc catcctggtc gagctggacg gcgacgtaaa      240
cggccacaag ttcagcgtgt ccggcgaggg cgaggcgat gccacctacg gcaagctgac      300
cctgaagttc atctgcacca ccggcaagct gcccggtgcc tggcccaccc tcgtgaccac      360
cctgacctac ggctgagct gcttcagccg ctaccccgac cacatgaagc agcacgactt      420
cttcaagtcc gccatgcccg aaggctacgt ccaggagcgc accatcttct tcaaggacga      480
cggcaactac aagaccccg ccgaggtgaa gtctgagggc gacaccctgg tgaaccgcat      540
cgagctgaag ggcctgact tcaaggagga cgcaacatc ctggggcaca agctggagta      600
caactacaac agccacaacg tctatatcat ggccgacaag cagaagaacg gcatcaaggt      660
gaacttcaag atccgccaca acatcgagga cggcagcgtg cagctggccg accactacca      720
gcagaacacc cccatcggcg acggccccgt gctgctgccc gacaaccact acctgagcac      780
ccagtccgcc ctgagcaaag accccaacga gaagcgcat cacatgttcc tgctggagtt      840
cgtga      845

```

<210> 32
 <211> 26
 <212> DNA
 <213> 人工序列

<220>
 <223> 编码内含子切除区域的DNA

<400> 32
 cagacgggca actttacacg acgctg 26

[0010]

<210> 33
 <400> 33
 000

<210> 34
 <211> 1892
 <212> DNA
 <213> 人工序列

<220>
 <223> SEAP表达构建体

```

<400> 34
tgacgtgctg atgatgcgta gctagtagtt gacgtgctga tgatgcgtag ctagtagttg      60
acgtgctgat gatgcgtagc tagtagttga cgtgctgatg atgcgtagct agtagttgac      120
gtgctgatga tgcgtagcta gtagttgacg tgcctgatgat gcgtagctag tagtgacgtt      180
agcgtagctg aggtaccgtc gacgataatc gatccaggtc tatataagca gagctcgttt      240
agtgaaccgt cagatgcctt agatacgcca tccacgctgt tttgacctcc atagaagatc      300
gccaccatgc tgcctgctgt gctgctgctg ggccctgagc tacagctctc cctgggcacg      360
atcccagttg aggaggagaa cccggacttc tggaaccgag aggcagccga ggccctgggt      420
gccgccaaga agctgcagcc tgcacagaca gccccaaga acctcatcat ctctctgggc      480
gatgggatgg ggggtgtctac ggtgacagct gccaggatcc taaaagggca gaagaaggac      540
aaactggggc ctgagatacc cctggctatg gaccgcttcc catatgtggc tctgtccaag      600
acatacaatg tagacaaaca tgtgccagac agtggagcca cagccacggc ctacctgtgc      660
ggggtcaagg gcaacttcca gaccattggc ttgagtcgag ccgcccgtt taaccagtgc      720

```

aacacgacac gcggaacga ggtcatctcc gtgatgaatc gggccaagaa agcagggaag	780
tcagtgggag tggtaacac cacacgagtg cagcacgcct cgccagccgg cacctacgcc	840
cacacggtga accgcaactg gtactcggac gccgacgtgc ctgcctcggc ccgccaggag	900
gggtgccagg acatcgctac gcagctcacc tccaacatgg acattgacgt gatcctaggt	960
ggaggccgaa agtacatgtt tcgcatggga accccagacc ctgagtacc agatgactac	1020
agccaaggtg ggaccaggct ggacgggaag aatctgggtc aggaatggct ggcgaagcgc	1080
cagggtgccc ggtatgtgtg gaaccgcaat gagctcatgc aggcttccct ggaccgctct	1140
gtgaccatc tcattggcct ctttgagcct ggagacatga aatacgagat ccaccgagac	1200
tccacactgg accctccct gatggagatg acagaggctg ccctgcgcct gctgagcagg	1260
aacccccgag gcttcttct cttcgtggag ggtgggtcga tcgaccacgg tcacatgaa	1320
agcagggtt accgggcaat gactgaaacg atcatgttcg acgacgcat tgagaggcg	1380
ggccagctca ccagcagga ggacacgtg agcctcgtca ctgccagca cttccacgtc	1440
ttctccttcg gaggtaccc cctgcgagg agctccatct tcgggctggc ccctggcaag	1500
gccccggaca ggaaggccta cacggctcct ctatacgaa acggtccagg ctatgtgctc	1560
aaggacggcg cccggccgga tgttaccgag agcgagagcg ggagccccga gtatcggcag	1620
cagacgggca actttacac acgtgcagt cagcagtgcc cctggacgaa gagacacacg	1680
caggcagga cgtggcggtg ttccgcgcg gcccgagcg gcacctggtt cagggcgtgc	1740
aggagcagac cttcatagcg cagtcattg ccttcgccgc ctgcctggag ccctacaccg	1800
cctgcgacct ggcccccc gccggcaca ccgacgccg gcacccgggt tactctagag	1860
[0011] tcggggcggc cgccgcttc gagcagacat ga	1892

<210> 35

<211> 483

<212> PRT

<213> 人工序列

<220>

<223> 截短的SEAP蛋白

<400> 35

Met Leu Leu Leu Leu Leu Leu Gly Leu Arg Leu Gln Leu Ser Leu
1 5 10 15

Gly Ile Ile Pro Val Glu Glu Glu Asn Pro Asp Phe Trp Asn Arg Glu
20 25 30

Ala Ala Glu Ala Leu Gly Ala Ala Lys Lys Leu Gln Pro Ala Gln Thr
35 40 45

Ala Ala Lys Asn Leu Ile Ile Phe Leu Gly Asp Gly Met Gly Val Ser
50 55 60

Thr Val Thr Ala Ala Arg Ile Leu Lys Gly Gln Lys Lys Asp Lys Leu
65 70 75 80

Gly Pro Glu Ile Pro Leu Ala Met Asp Arg Phe Pro Tyr Val Ala Leu
85 90 95

Ser Lys Thr Tyr Asn Val Asp Lys His Val Pro Asp Ser Gly Ala Thr

100	105	110
Ala Thr Ala Tyr Leu Cys Gly Val Lys Gly Asn Phe Gln Thr Ile Gly 115 120 125		
Leu Ser Ala Ala Ala Arg Phe Asn Gln Cys Asn Thr Thr Arg Gly Asn 130 135 140		
Glu Val Ile Ser Val Met Asn Arg Ala Lys Lys Ala Gly Lys Ser Val 145 150 155 160		
Gly Val Val Thr Thr Thr Arg Val Gln His Ala Ser Pro Ala Gly Thr 165 170 175		
Tyr Ala His Thr Val Asn Arg Asn Trp Tyr Ser Asp Ala Asp Val Pro 180 185 190		
Ala Ser Ala Arg Gln Glu Gly Cys Gln Asp Ile Ala Thr Gln Leu Ile 195 200 205		
Ser Asn Met Asp Ile Asp Val Ile Leu Gly Gly Gly Arg Lys Tyr Met 210 215 220		
Phe Arg Met Gly Thr Pro Asp Pro Glu Tyr Pro Asp Asp Tyr Ser Gln 225 230 235 240		
Gly Gly Thr Arg Leu Asp Gly Lys Asn Leu Val Gln Glu Trp Leu Ala 245 250 255		
[0012] Lys Arg Gln Gly Ala Arg Tyr Val Trp Asn Arg Thr Glu Leu Met Gln 260 265 270		
Ala Ser Leu Asp Pro Ser Val Thr His Leu Met Gly Leu Phe Glu Pro 275 280 285		
Gly Asp Met Lys Tyr Glu Ile His Arg Asp Ser Thr Leu Asp Pro Ser 290 295 300		
Leu Met Glu Met Thr Glu Ala Ala Leu Arg Leu Leu Ser Arg Asn Pro 305 310 315 320		
Arg Gly Phe Phe Leu Phe Val Glu Gly Gly Arg Ile Asp His Gly His 325 330 335		
His Glu Ser Arg Ala Tyr Arg Ala Leu Thr Glu Thr Ile Met Phe Asp 340 345 350		
Asp Ala Ile Glu Arg Ala Gly Gln Leu Thr Ser Glu Glu Asp Thr Leu 355 360 365		
Ser Leu Val Thr Ala Asp His Ser His Val Phe Ser Phe Gly Gly Tyr 370 375 380		
Pro Leu Arg Gly Ser Ser Ile Phe Gly Leu Ala Pro Gly Lys Ala Arg 385 390 395 400		
Asp Arg Lys Ala Tyr Thr Val Leu Leu Tyr Gly Asn Gly Pro Gly Tyr		

	405	410	415
	Val Leu Lys Asp Gly Ala Arg Pro Asp Val Thr Glu Ser Glu Ser Gly 420 425 430		
	Ser Pro Glu Tyr Arg Gln Gln Thr Gly Asn Phe Thr Arg Arg Cys Ser 435 440 445		
	Gln Gln Cys Pro Trp Thr Lys Arg His Thr Gln Ala Arg Thr Trp Arg 450 455 460		
	Cys Ser Arg Ala Ala Arg Arg Arg Thr Trp Phe Thr Ala Cys Arg Ser 465 470 475 480		
	Arg Pro Ser		
<210>	36		
<211>	519		
<212>	PRT		
<213>	未知		
<220>			
<223>	全长SEAP蛋白		
<400>	36		
	Met Leu Leu Leu Leu Leu Leu Leu Gly Leu Arg Leu Gln Leu Ser Leu 1 5 10 15		
[0013]	Gly Ile Ile Pro Val Glu Glu Glu Asn Pro Asp Phe Trp Asn Arg Glu 20 25 30		
	Ala Ala Glu Ala Leu Gly Ala Ala Lys Lys Leu Gln Pro Ala Gln Thr 35 40 45		
	Ala Ala Lys Asn Leu Ile Ile Phe Leu Gly Asp Gly Met Gly Val Ser 50 55 60		
	Thr Val Thr Ala Ala Arg Ile Leu Lys Gly Gln Lys Lys Asp Lys Leu 65 70 75 80		
	Gly Pro Glu Ile Pro Leu Ala Met Asp Arg Phe Pro Tyr Val Ala Leu 85 90 95		
	Ser Lys Thr Tyr Asn Val Asp Lys His Val Pro Asp Ser Gly Ala Thr 100 105 110		
	Ala Thr Ala Tyr Leu Cys Gly Val Lys Gly Asn Phe Gln Thr Ile Gly 115 120 125		
	Leu Ser Ala Ala Ala Arg Phe Asn Gln Cys Asn Thr Thr Arg Gly Asn 130 135 140		
	Glu Val Ile Ser Val Met Asn Arg Ala Lys Lys Ala Gly Lys Ser Val 145 150 155 160		
	Gly Val Val Thr Thr Thr Arg Val Gln His Ala Ser Pro Ala Gly Thr 165 170 175		

	Tyr	Ala	His	Thr	Val	Asn	Arg	Asn	Trp	Tyr	Ser	Asp	Ala	Asp	Val	Pro
				180					185						190	
	Ala	Ser	Ala	Arg	Gln	Glu	Gly	Cys	Gln	Asp	Ile	Ala	Thr	Gln	Leu	Ile
			195					200					205			
	Ser	Asn	Met	Asp	Ile	Asp	Val	Ile	Leu	Gly	Gly	Gly	Arg	Lys	Tyr	Met
		210					215					220				
	Phe	Arg	Met	Gly	Thr	Pro	Asp	Pro	Glu	Tyr	Pro	Asp	Asp	Tyr	Ser	Gln
	225					230					235					240
	Gly	Gly	Thr	Arg	Leu	Asp	Gly	Lys	Asn	Leu	Val	Gln	Glu	Trp	Leu	Ala
					245					250					255	
	Lys	Arg	Gln	Gly	Ala	Arg	Tyr	Val	Trp	Asn	Arg	Thr	Glu	Leu	Met	Gln
				260					265					270		
	Ala	Ser	Leu	Asp	Pro	Ser	Val	Thr	His	Leu	Met	Gly	Leu	Phe	Glu	Pro
			275					280					285			
	Gly	Asp	Met	Lys	Tyr	Glu	Ile	His	Arg	Asp	Ser	Thr	Leu	Asp	Pro	Ser
		290					295					300				
	Leu	Met	Glu	Met	Thr	Glu	Ala	Ala	Leu	Arg	Leu	Leu	Ser	Arg	Asn	Pro
	305					310					315					320
[0014]	Arg	Gly	Phe	Phe	Leu	Phe	Val	Glu	Gly	Gly	Arg	Ile	Asp	His	Gly	His
					325					330					335	
	His	Glu	Ser	Arg	Ala	Tyr	Arg	Ala	Leu	Thr	Glu	Thr	Ile	Met	Phe	Asp
				340					345					350		
	Asp	Ala	Ile	Glu	Arg	Ala	Gly	Gln	Leu	Thr	Ser	Glu	Glu	Asp	Thr	Leu
			355					360					365			
	Ser	Leu	Val	Thr	Ala	Asp	His	Ser	His	Val	Phe	Ser	Phe	Gly	Gly	Tyr
		370					375					380				
	Pro	Leu	Arg	Gly	Ser	Ser	Ile	Phe	Gly	Leu	Ala	Pro	Gly	Lys	Ala	Arg
	385					390					395					400
	Asp	Arg	Lys	Ala	Tyr	Thr	Val	Leu	Leu	Tyr	Gly	Asn	Gly	Pro	Gly	Tyr
					405					410					415	
	Val	Leu	Lys	Asp	Gly	Ala	Arg	Pro	Asp	Val	Thr	Glu	Ser	Glu	Ser	Gly
				420					425					430		
	Ser	Pro	Glu	Tyr	Arg	Gln	Gln	Ser	Ala	Val	Pro	Leu	Asp	Glu	Glu	Thr
			435					440					445			
	His	Ala	Gly	Glu	Asp	Val	Ala	Val	Phe	Ala	Arg	Gly	Pro	Gln	Ala	His
		450					455					460				
	Leu	Val	His	Gly	Val	Gln	Glu	Gln	Thr	Phe	Ile	Ala	His	Val	Met	Ala
	465					470					475					480

Phe Ala Ala Cys Leu Glu Pro Tyr Thr Ala Cys Asp Leu Ala Pro Pro
485 490 495

Ala Gly Thr Thr Asp Ala Ala His Pro Gly Tyr Ser Arg Val Gly Ala
500 505 510

Ala Gly Arg Phe Glu Gln Thr
515

<210> 37
<211> 26
<212> RNA
<213> 人工序列

<220>
<223> 可调节内含子

<400> 37
cagacgggca acuuuacacg acgcug 26

<210> 38
<211> 32
<212> DNA
<213> 人工序列

<220>
<223> 插入萤光素酶编码序列中的内含子

<400> 38
ccgcagacgg gcaactttac acgacgtgc ag 32

[0015]

<210> 39
<211> 501
<212> PRT
<213> 人工序列

<220>
<223> 截短的萤光素酶蛋白

<400> 39

Met Glu Asp Ala Lys Asn Ile Lys Lys Gly Pro Ala Pro Phe Tyr Pro
1 5 10 15

Leu Glu Asp Gly Thr Ala Gly Glu Gln Leu His Lys Ala Met Lys Arg
20 25 30

Tyr Ala Leu Val Pro Gly Thr Ile Ala Phe Thr Asp Ala His Ile Glu
35 40 45

Val Asp Ile Thr Tyr Ala Glu Tyr Phe Glu Met Ser Val Arg Leu Ala
50 55 60

Glu Ala Met Lys Arg Tyr Gly Leu Asn Thr Asn His Arg Ile Val Val
65 70 75 80

Cys Ser Glu Asn Ser Leu Gln Phe Phe Met Pro Val Leu Gly Ala Leu
85 90 95

Phe Ile Gly Val Ala Val Ala Pro Ala Asn Asp Ile Tyr Asn Glu Arg
100 105 110

Glu Leu Leu Asn Ser Met Gly Ile Ser Gln Pro Thr Val Val Phe Val

115	120	125
Ser Lys Lys Gly Leu Gln Lys Ile Leu Asn Val Gln Lys Lys Leu Pro 130 135 140		
Ile Ile Gln Lys Ile Ile Ile Met Asp Ser Lys Thr Asp Tyr Gln Gly 145 150 155 160		
Phe Gln Ser Met Tyr Thr Phe Val Thr Ser His Leu Pro Pro Gly Phe 165 170 175		
Asn Glu Tyr Asp Phe Val Pro Glu Ser Phe Asp Arg Asp Lys Thr Ile 180 185 190		
Ala Leu Ile Met Asn Ser Ser Gly Ser Thr Gly Leu Pro Lys Gly Val 195 200 205		
Ala Leu Pro His Arg Thr Ala Cys Val Arg Phe Ser His Ala Arg Asp 210 215 220		
Pro Ile Phe Gly Asn Gln Ile Ile Pro Asp Thr Ala Ile Leu Ser Val 225 230 235 240		
Val Pro Phe His His Gly Phe Gly Met Phe Thr Thr Leu Gly Tyr Leu 245 250 255		
Ile Cys Gly Phe Arg Val Val Leu Met Tyr Arg Phe Glu Glu Glu Leu 260 265 270		
[0016] Phe Leu Arg Ser Leu Gln Asp Tyr Lys Ile Gln Ser Ala Leu Leu Val 275 280 285		
Pro Thr Leu Phe Ser Phe Phe Ala Lys Ser Thr Leu Ile Asp Lys Tyr 290 295 300		
Asp Leu Ser Asn Leu His Glu Ile Ala Ser Gly Gly Ala Pro Leu Ser 305 310 315 320		
Lys Glu Val Gly Glu Ala Val Ala Lys Arg Phe His Leu Pro Gly Ile 325 330 335		
Arg Gln Gly Tyr Gly Leu Thr Glu Thr Thr Ser Ala Ile Leu Ile Thr 340 345 350		
Pro Glu Gly Asp Asp Lys Pro Gly Ala Val Gly Lys Val Val Pro Phe 355 360 365		
Phe Glu Ala Lys Val Val Asp Leu Asp Thr Gly Lys Thr Leu Gly Val 370 375 380		
Asn Gln Arg Gly Glu Leu Cys Val Arg Gly Pro Met Ile Met Ser Gly 385 390 395 400		
Tyr Val Asn Asn Pro Glu Ala Thr Asn Ala Leu Ile Asp Lys Asp Gly 405 410 415		
Trp Leu His Ser Gly Asp Ile Ala Tyr Trp Asp Glu Asp Glu His Phe		

	420	425	430
	Phe Ile Val Asp Arg Leu Lys Ser Leu Ile Lys Tyr Lys Gly Tyr Gln 435 440 445		
	Val Ala Pro Ala Glu Leu Glu Ser Ile Leu Leu Gln His Pro Asn Ile 450 455 460		
	Phe Asp Ala Gly Val Ala Gly Leu Pro Asp Asp Ala Gly Glu Leu 465 470 475 480		
	Pro Ala Ala Asp Gly Gln Leu Tyr Thr Thr Leu Gln Thr Ser Cys Trp 485 490 495		
	Asn Thr Val Lys Pro 500		
	<210> 40 <211> 550 <212> PRT <213> 未知		
	<220> <223> 全长萤光素酶蛋白		
	<400> 40		
	Met Glu Asp Ala Lys Asn Ile Lys Lys Gly Pro Ala Pro Phe Tyr Pro 1 5 10 15		
[0017]	Leu Glu Asp Gly Thr Ala Gly Glu Gln Leu His Lys Ala Met Lys Arg 20 25 30		
	Tyr Ala Leu Val Pro Gly Thr Ile Ala Phe Thr Asp Ala His Ile Glu 35 40 45		
	Val Asp Ile Thr Tyr Ala Glu Tyr Phe Glu Met Ser Val Arg Leu Ala 50 55 60		
	Glu Ala Met Lys Arg Tyr Gly Leu Asn Thr Asn His Arg Ile Val Val 65 70 75 80		
	Cys Ser Glu Asn Ser Leu Gln Phe Phe Met Pro Val Leu Gly Ala Leu 85 90 95		
	Phe Ile Gly Val Ala Val Ala Pro Ala Asn Asp Ile Tyr Asn Glu Arg 100 105 110		
	Glu Leu Leu Asn Ser Met Gly Ile Ser Gln Pro Thr Val Val Phe Val 115 120 125		
	Ser Lys Lys Gly Leu Gln Lys Ile Leu Asn Val Gln Lys Lys Leu Pro 130 135 140		
	Ile Ile Gln Lys Ile Ile Ile Met Asp Ser Lys Thr Asp Tyr Gln Gly 145 150 155 160		
	Phe Gln Ser Met Tyr Thr Phe Val Thr Ser His Leu Pro Pro Gly Phe 165 170 175		

	Asn	Glu	Tyr	Asp	Phe	Val	Pro	Glu	Ser	Phe	Asp	Arg	Asp	Lys	Thr	Ile	
				180					185					190			
	Ala	Leu	Ile	Met	Asn	Ser	Ser	Gly	Ser	Thr	Gly	Leu	Pro	Lys	Gly	Val	
			195					200					205				
	Ala	Leu	Pro	His	Arg	Thr	Ala	Cys	Val	Arg	Phe	Ser	His	Ala	Arg	Asp	
		210					215					220					
	Pro	Ile	Phe	Gly	Asn	Gln	Ile	Ile	Pro	Asp	Thr	Ala	Ile	Leu	Ser	Val	
	225					230					235					240	
	Val	Pro	Phe	His	His	Gly	Phe	Gly	Met	Phe	Thr	Thr	Leu	Gly	Tyr	Leu	
					245					250					255		
	Ile	Cys	Gly	Phe	Arg	Val	Val	Leu	Met	Tyr	Arg	Phe	Glu	Glu	Glu	Leu	
				260					265					270			
	Phe	Leu	Arg	Ser	Leu	Gln	Asp	Tyr	Lys	Ile	Gln	Ser	Ala	Leu	Leu	Val	
			275					280					285				
	Pro	Thr	Leu	Phe	Ser	Phe	Phe	Ala	Lys	Ser	Thr	Leu	Ile	Asp	Lys	Tyr	
		290					295					300					
	Asp	Leu	Ser	Asn	Leu	His	Glu	Ile	Ala	Ser	Gly	Gly	Ala	Pro	Leu	Ser	
	305					310					315					320	
[0018]	Lys	Glu	Val	Gly	Glu	Ala	Val	Ala	Lys	Arg	Phe	His	Leu	Pro	Gly	Ile	
					325					330					335		
	Arg	Gln	Gly	Tyr	Gly	Leu	Thr	Glu	Thr	Thr	Ser	Ala	Ile	Leu	Ile	Thr	
				340					345					350			
	Pro	Glu	Gly	Asp	Asp	Lys	Pro	Gly	Ala	Val	Gly	Lys	Val	Val	Pro	Phe	
			355					360					365				
	Phe	Glu	Ala	Lys	Val	Val	Asp	Leu	Asp	Thr	Gly	Lys	Thr	Leu	Gly	Val	
		370					375					380					
	Asn	Gln	Arg	Gly	Glu	Leu	Cys	Val	Arg	Gly	Pro	Met	Ile	Met	Ser	Gly	
	385					390					395					400	
	Tyr	Val	Asn	Asn	Pro	Glu	Ala	Thr	Asn	Ala	Leu	Ile	Asp	Lys	Asp	Gly	
					405					410					415		
	Trp	Leu	His	Ser	Gly	Asp	Ile	Ala	Tyr	Trp	Asp	Glu	Asp	Glu	His	Phe	
				420					425					430			
	Phe	Ile	Val	Asp	Arg	Leu	Lys	Ser	Leu	Ile	Lys	Tyr	Lys	Gly	Tyr	Gln	
			435					440					445				
	Val	Ala	Pro	Ala	Glu	Leu	Glu	Ser	Ile	Leu	Leu	Gln	His	Pro	Asn	Ile	
		450					455					460					
	Phe	Asp	Ala	Gly	Val	Ala	Gly	Leu	Pro	Asp	Asp	Asp	Ala	Gly	Glu	Leu	
	465					470					475					480	

	Pro Ala Ala Asp Val Val Leu Glu His Gly Lys Thr Met Thr Glu Lys 485 490 495	
	Glu Ile Val Asp Tyr Val Ala Ser Gln Val Thr Thr Ala Lys Lys Leu 500 505 510	
	Arg Gly Gly Val Val Phe Val Asp Glu Val Pro Lys Gly Leu Thr Gly 515 520 525	
	Lys Leu Asp Ala Arg Lys Ile Arg Glu Ile Leu Ile Lys Ala Lys Lys 530 535 540	
	Gly Gly Lys Ile Ala Val 545 550	
	<210> 41 <211> 30 <212> DNA <213> 人工序列	
	<220> <223> INTC9FP内含子	
	<400> 41 acgtcagacg ggcaacttta cacgacgtg	30
	<210> 42 <211> 30 <212> DNA <213> 人工序列	
[0019]	<220> <223> INTC9RP内含子序列	
	<400> 42 acgtcagcgt cgtgttaaagt tgcccgtctg	30
	<210> 43 <211> 32 <212> DNA <213> 人工序列	
	<220> <223> 插入胰天蛋白酶9编码序列中的可调节内含子	
	<400> 43 cagcagacgg gcaactttac acgacgtctg tg	32
	<210> 44 <211> 8008 <212> DNA <213> 人工序列	
	<220> <223> 具有内含子构建体的SYNP-CASP9-INT载体	
	<400> 44 gacggatcgg gagatctccc gatcccctat ggtgcactct cagtacaatc tgctctgatg	60
	ccgcatagtt aagccagtat ctgctccctg cttgtgtgtt ggaggtcgct gagtagtgcg	120
	cgagcaaaat ttaagctaca acaaggcaag gcttgaccga caattgcatg aagaatctgc	180
	ttagggttag gcgttttgcg ctgcttcgcg agtacattta tattggctca tgtccaatat	240
	gaccgccatg ttgacattga ttattgacta gttattaata gtaatcaatt acgggggtcat	300

[0020]

tagttcatag cccatataatg gagttccgcg ttacataact tacggtaaact ggcccgcctg	360
gctgaccgcc caacgacccc cgcccattga cgtcaataat gacgtatgtt cccatagtaa	420
cgccaatagg gactttccat tgacgtcaat ggggtggagta ttacggtaa actgcccact	480
tggcagtaca tcaagtgtat catatgccaa gtccgcccc tattgacgtc aatgacggta	540
aatggcccgc ctggcattat gcccagtaca tgaccttacg ggactttcct acttggcagt	600
acatctacgt attagtcatc gctattacca tggatgatcg gttttggcag tacaccaatg	660
ggcgtggata gcggtttgac tcacggggat ttccaagtct ccacccatt gacgtcaatg	720
ggagtttgtt ttggcaccaa aatcaacggg actttccaaa atgtcgtaat aacccgccc	780
cgttgacgca aatgggcggg aggcgtgtac ggtgggaggt ctatataagc agagctcgtt	840
tagtgaaccg tcagatcctc actctcttcc gcatcgctgt ctgcgagggc cagctgttgg	900
gctcgcgggtt gaggacaaac tcttcgcggt ctttcagta ctcttggatc ggaaacccgt	960
cggcctccga acggtactcc gccaccgagg gacctgagcg agtccgcatc gaccggatcg	1020
gaaaacctct cgagaaaggc gtctaaccag tcacagtcgc aaggtaggct gagcaccgtg	1080
gcgggcggca gcgggtggcg gtcggggttg ttcttggcgg aggtgctgct gatgatgtaa	1140
ttaaagtagg cggctcttag acggcggatg gtcgaggtga ggtgtgggtt tagtgaaccg	1200
tcagatcctc actctcttcc gcatcgctgt ctgcgagggc cagctgtcag gcttgagatc	1260
cagctgttgg ggtgagtact ccctctcaaa agcgggcatt acttctgcgc taagattgtc	1320
agtttccaaa aacgaggagg atttgataatt cacctggccc gatctggcca tacacttgag	1380
tgacaatgac atccactttg cttttctctc cacaggtgtc cactcccagg tccaagttaa	1440
aactttaata cgactcacta taggggccgc caccaagctt ggtacatgga cgaagcggat	1500
cggcggctcc tgcggcgggtg ccggctgcgg ctggtggaag agctgcaggt ggaccagctc	1560
tgggacgccc tgcagagccg cgagctgttc agggcccata tgatcgagga catccagcgg	1620
gcaggctctg gatctcggcg ggatcaggcc aggcagctga tcatagatct ggagactcga	1680
gggagtcagg ctcttctttt gttcatctcc tgcttagagg acacaggcca ggacatgctg	1740
gttctgtttc tgcgaactaa caggcaagca gcaaagttgt cgaagccaac ctagaaaac	1800
cttaccacag tgggtgctcag accagagatt cgcaaccag aggttctcag accggaaca	1860
cccagaccag tggacattgg ttctggagga ttcggtgatg tcggtgctct tgagagtttg	1920
aggggaaatg cagatttggc ttacatcctg agcatggagc cctgtggcca ctgcctcatt	1980
atcaacaatg tgaacttctg ccgtgagtcg gggctccgca cccgactgg ctccaacatc	2040
gactgtgaga agttgcggcg tcgttcttcc tcgtgcatt tcatggtgga ggtgaagggc	2100
gacctgactg ccaagaaaat ggtgctggct ttgctggagc tggcgagca ggaccacggt	2160
gctctggact gctgcgtggt ggtcattctc tctcaggct gtcaggccag ccacctgcag	2220
ttcccagggg ctgtctacgg cacagatgga tgccctgtgt cggtcgagaa gattgtgaac	2280
atcttcaatg ggaccagctg cccagcctg ggaggggaagc ccaagctctt ttcatccag	2340
gcctgtgggtg gggagcagaa agaccatggg tttaggtgg cctccacttc cctgaagac	2400
gagtcacctg gcagtaacc cgagccagat gccacccgt tccaggaagg tttaggacc	2460
ttcgaccagc tggacgcat atctagtttg cccacacca gtgacatctt tgtgtcctac	2520
tctactttcc caggttttgt ttctggagg gacccaaga gggctcctg gtcagacggg	2580

[0021]

caactttaca cgacgttgga gaccttgga gacatctttg agcagtgggc tcactctgaa	2640
gacctgcagt cctcctgct tagggtcgct aatgctgttt cggtgaaagg gatttataaa	2700
cagatgcctg gttgctttaa ttctctccga aaaaaacttt tctttaaaac atcaggggggt	2760
ggaggctctg tgagcaaggc cgaggagctg ttcaccgggg tggtgcccat cctggtcgag	2820
ctggacggcg acgtaaacgg ccacaagttc agcgtgtccg gcgagggcga gggcgatgcc	2880
acctacggca agctgacct gaagttcatc tgcaccaccg gcaagctgcc cgtgccctgg	2940
cccacctcg tgaccacct gacctacggc gtgcagtgtc tcagccgcta ccccgaccac	3000
atgaagaagc acgacttctt caagtccgcc atgcccgaag gctacgtcca ggagcgcacc	3060
atcttcttca aggacgacgg caactacaag acccgcgccg aggtgaagtt cgagggcgac	3120
accttggtag accgcatcga gctgaaggc atcgacttca aggaggacgg caacatcctg	3180
gggcacaagc tggagtacaa ctacaacagc cacaacgtct atatcatggc cgacaagcag	3240
aagaacggca tcaaggctaa cttcaagggt cgccacaaca tcgaggacgg cagcgtgcag	3300
ctcgccgacc actaccagca gaacaccccc atcgcgacg gccccgtgct gctgcccgac	3360
aaccactacc tgagcaccca gtccgccctg agcaaagacc ccaacgagaa gcgcgatcac	3420
atggtcctgc tggagttcgt gaccgccgc gggatcactc tcggcatgga cgagctgtac	3480
aagtaaaact gagtctagag cggcgccga attcggggcc gtttaaaccc gctgatcagc	3540
ctcgactgtg ccttctagtt gccagccatc tgttgtttgc ccctccccg tgccttcctt	3600
gaccttgaa ggtgccactc ccactgtcct ttcctaataa aatgaggaaa ttgcatcgca	3660
ttgtctgagt aggtgtcatt ctattctggg ggggtgggtg gggcaggaca gcaaggggga	3720
ggatttgga gacaatagca ggcatgctgg ggatcggtg ggccttatgg cttctgaggc	3780
ggaaagaacc agctggggct ctagggggta tccccacgcg cctgttagcg gcgcattaag	3840
cgcggcgggt gtggtggta cgcgcagct gaccgtaca cttgccagcg ccctagcgcc	3900
cgtctcttcc gctttctcc cttctttct cgcacgttc gcaggcttcc cccgtcaagc	3960
tctaaatcgg gggctccctt tagggttccg atttagtct ttacggcacc tcgaccccaa	4020
aaaacttgat tagggtgatg gtacacgtag tggccatcg cctgataga cggtttttcg	4080
ccctttgacg ttggagtcca cgttctttaa tagtggactc ttgttccaaa ctggaacaac	4140
actcaacct atctcggtct attcttttga ttataaggg attttgccga tttcggccta	4200
tgggttaaaa aatgagctga ttaacaaaa atttaacgcg aattaattct gtggaatgtg	4260
tgtcagttag ggtgtgaaa gtccccagc tccccagcag gcagaagtat gcaaagcatg	4320
catctcaatt agtcagcaac cagggttgga aagtccccag gctccccagc aggcagaagt	4380
atgcaaagca tgcatctcaa ttagtcagca accatagtcc cggccctaac tccgccatc	4440
ccgcccttaa ctccgccctg ttccgcccat tctccgccc atggtgact aattttttt	4500
atttatgcag aggccgaggc cgcctctgcc tctgagctat tccagaagta gtgaggaggc	4560
ttttttggag gcctaggctt ttgcaaaaag ctctcgggag cttgtatata cattttcgga	4620
tctgatcagc acgtgatgaa aaagcctgaa ctcaccgcga cgtctgtcga gaagtttctg	4680
atcgaaaagt tcgacagcgt ctccgacctg atgcagctct cggagggcga agaattctgt	4740
gctttcagct tcgatgtagg agggcgtgga tatgtcctgc gggtaaatag ctgcgccgat	4800
ggtttctaca aagatcgtta tgtttatcgg cactttgcat cggccgcgt cccgattccg	4860

[0022]

gaagtgccttg acattgggga attcagcgag agcctgacct attgcatctc ccgccgtgca	4920
cagggtgtca cgttgcaaga cctgccctgaa accgaactgc ccgctgttct gcagccggtc	4980
gcggaggcca tggatgcgat cgctgcggcc gatcttagcc agacgagcgg gttcggccca	5040
ttcggaccgc aaggaatcgg tcaatacact acatggcgtg atttcatatg cgcgattgct	5100
gatccccatg tgtatcactg gcaaaactgtg atggacgaca ccgtcagtgc gtccgtcgcg	5160
caggctctcg atgagctgat gctttgggcc gaggactgcc ccgaagtccg gcacctcgtg	5220
cacgcggatt tcggctcaa caatgtcctg acggacaatg gccgcataac agcggtcatt	5280
gactggagcg aggcgatgtt cggggattcc caatacgagg tcgccaacat cttcttctgg	5340
aggccgtggt tggcttgtat ggagcagcag acgcgtact tcgagcggag gcatccggag	5400
cttgcaggat cgcgcgggct ccgggcgtat atgctccgca ttggtcttga ccaactctat	5460
cagagcttgg ttgacggcaa tttcgtgat gcagcttggg cgcagggtcg atgcgacgca	5520
atcgtccgat ccggagccgg gactgtcggg cgtacacaaa tcgcccgag aagcgcggcc	5580
gtctggaccg atggctgtgt agaagtact gccgatagt gaaaccgacg cccagcact	5640
cgcccgagg caaaggaata gcacgtgcta cgagatttcg attccaccgc cgccttctat	5700
gaaaggttgg gcttcggaat cgttttccgg gacgctggct ggatgatcct ccagcgcggg	5760
gatctcatgc tggagttctt cgccccccc aacttgttta ttgcagctta taatggttac	5820
aaataagca atagcatcac aaatttcaca aataaagcat ttttttact gcattctagt	5880
tgtggtttgt ccaaactcat caatgtatct tatcatgtct gtataccgtc gacctctagc	5940
tagagcttgg cgtaatcatg gtcatactg tttcctgtgt gaaattgtta tccgtcaca	6000
attccacaca acatacgagc cggaagcata aagtgtaaag cctgggggtgc ctaatgagtg	6060
agctaactca cattaattgc gttgcgtca ctgcccgctt tccagtcggg aaacctgtcg	6120
tgccagctgc attaatgaat cggccaacgc gcggggagag gcggtttgcg tattgggcgc	6180
tcttcgcgtt cctcgtcac tgaactcgtg cgctcggctg ttcggctgcg gcgagcggta	6240
tcagctcact caaaggcggg aatacggta tccacagaat caggggataa cgcaggaaag	6300
aacatgtgag caaaaggcca gcaaaaggcc aggaaccgta aaaaggccgc gttgctggcg	6360
tttttcata ggctccgcc ccctgacgag catcacaaaa atcgacgctc aagtcagagg	6420
tggcgaacc cgacaggact ataaagatac caggcgtttc cccttgaag ctccctcgtg	6480
cgctctcctg ttccgacct gccgcttacc ggatacctgt ccgctttct cccttcggga	6540
agcgtggcgc tttctcatag ctacgcgtg aggtatctca gttcgggtga ggtcgttcgc	6600
tccaagctgg gctgtgtgca cgaaccccc gttcagccc accgctgcgc cttatccggt	6660
aactatcgtc ttgagtccaa cccggttaaga cacgacttat cgcactggc agcagccact	6720
ggtaacagga ttagcagagc gaggtatgta ggccgtgcta cagagttctt gaagtgttg	6780
cctaactacg gctacactag aagaacagta ttggtatct gcgctctgct gaagccagtt	6840
accttcgaa aaagagtgg tagctcttga tccggcaaac aaaccaccgc tggtagcgg	6900
ggtttttttg ttgcaagca gcagattacg cgcagaaaaa aaggatctca agaagatcct	6960
ttgatcttt ctacggggtc tgagcgcgga acccctatgt gtttatttt ctaatacat	7020
tcaaatatgt atccgctcat gaattaattc ttagaaaaac tcatacgagca tcaaatgaaa	7080
cigcaattta ttcatatcag gattatcaat accatatgtt tgaanaagcc gtttctgtaa	7140

```

tgaaggagaa aactcacga ggcagttcca taggatggca agatcctggt atcgggtctgc 7200
gattccgact cgtccaacat caatacaacc tattaatttc ccctcgtcaa aaataaggtt 7260
atcaagttag aaatcaccat gagtgacgac tgaatccggt gagaatggca aaagtttatg 7320
catttctttc cagacttggt caacaggcca gccattacgc tcgtcatcaa aatcactcgc 7380
atcaacaaaa ccgttattca ttctgtattg cgctgagcgc agacgaaata cgcgatcgtc 7440
gttaaaagga caattacaaa caggaatcga atgcaaccgg cgcaggaaca ctgccagcgc 7500
atcaacaata ttttcacctg aatcaggata ttcttctaata acctggaatg ctgtttttccc 7560
agggatcgca gtggtgagta accatgcacg atcaggagta cggataaaat gcttgatggt 7620
cggaagaggc ataaattccg tcagccagtt tagtctgacc atctcatctg taacatcatt 7680
ggcaacgcta cctttgccat gtttcagaaa caactctggc gcatcgggct tcccatacaa 7740
tcgatagatt gtcgcacctg attgcccgcg attatcgca gccatttat acccatataa 7800
atcagcatcc atgttggaat ttaatcgagg cctagagcaa gacgtttccc gttgaatatg 7860
gtcataaaca ccccttgatg tactgtttat gtaagcagac agttttattg ttcattacca 7920
aaatccctta acgtgagttt tcgttccact gagcgtcaga ccccgtagaa atccgcgcac 7980
atttccccga aaagtgccac ctgacgtc 8008

```

```

<210> 45
<211> 371
<212> PRT
<213> 人工序列

```

[0023]

```

<220>
<223> 截短的半胱天冬酶9

```

```

<400> 45

```

```

Met Asp Glu Ala Asp Arg Arg Leu Leu Arg Arg Cys Arg Leu Arg Leu
1          5          10          15

```

```

Val Glu Glu Leu Gln Val Asp Gln Leu Trp Asp Ala Leu Leu Ser Arg
20          25          30

```

```

Glu Leu Phe Arg Pro His Met Ile Glu Asp Ile Gln Arg Ala Gly Ser
35          40          45

```

```

Gly Ser Arg Arg Asp Gln Ala Arg Gln Leu Ile Ile Asp Leu Glu Thr
50          55          60

```

```

Arg Gly Ser Gln Ala Leu Pro Leu Phe Ile Ser Cys Leu Glu Asp Thr
65          70          75          80

```

```

Gly Gln Asp Met Leu Ala Ser Phe Leu Arg Thr Asn Arg Gln Ala Ala
85          90          95

```

```

Lys Leu Ser Lys Pro Thr Leu Glu Asn Leu Thr Pro Val Val Leu Arg
100         105         110

```

```

Pro Glu Ile Arg Lys Pro Glu Val Leu Arg Pro Glu Thr Pro Arg Pro
115         120         125

```

```

Val Asp Ile Gly Ser Gly Gly Phe Gly Asp Val Gly Ala Leu Glu Ser
130         135         140

```

	Leu	Arg	Gly	Asn	Ala	Asp	Leu	Ala	Tyr	Ile	Leu	Ser	Met	Glu	Pro	Cys	
	145					150					155					160	
	Gly	His	Cys	Leu	Ile	Ile	Asn	Asn	Val	Asn	Phe	Cys	Arg	Glu	Ser	Gly	
				165						170						175	
	Leu	Arg	Thr	Arg	Thr	Gly	Ser	Asn	Ile	Asp	Cys	Glu	Lys	Leu	Arg	Arg	
				180					185						190		
	Arg	Phe	Ser	Ser	Leu	His	Phe	Met	Val	Glu	Val	Lys	Gly	Asp	Leu	Thr	
			195					200					205				
	Ala	Lys	Lys	Met	Val	Leu	Ala	Leu	Leu	Glu	Leu	Ala	Gln	Gln	Asp	His	
	210						215					220					
	Gly	Ala	Leu	Asp	Cys	Cys	Val	Val	Val	Ile	Leu	Ser	His	Gly	Cys	Gln	
	225				230						235					240	
	Ala	Ser	His	Leu	Gln	Phe	Pro	Gly	Ala	Val	Tyr	Gly	Thr	Asp	Gly	Cys	
				245					250						255		
	Pro	Val	Ser	Val	Glu	Lys	Ile	Val	Asn	Ile	Phe	Asn	Gly	Thr	Ser	Cys	
				260					265						270		
	Pro	Ser	Leu	Gly	Gly	Lys	Pro	Lys	Leu	Phe	Phe	Ile	Gln	Ala	Cys	Gly	
			275					280					285				
[0024]	Gly	Glu	Gln	Lys	Asp	His	Gly	Phe	Glu	Val	Ala	Ser	Thr	Ser	Pro	Glu	
	290						295					300					
	Asp	Glu	Ser	Pro	Gly	Ser	Asn	Pro	Glu	Pro	Asp	Ala	Thr	Pro	Phe	Gln	
	305					310					315					320	
	Glu	Gly	Leu	Arg	Thr	Phe	Asp	Gln	Leu	Asp	Ala	Ile	Ser	Ser	Leu	Pro	
					325					330					335		
	Thr	Pro	Ser	Asp	Ile	Phe	Val	Ser	Tyr	Ser	Thr	Phe	Pro	Gly	Phe	Val	
				340					345					350			
	Ser	Trp	Arg	Asp	Pro	Lys	Ser	Gly	Ser	Trp	Ser	Ala	Gly	Asp	Pro	Gly	
			355					360					365				
	Arg	His	Leu														
			370														
	<210>	46															
	<211>	430															
	<212>	PRT															
	<213>	人工序列															
	<220>																
	<223>	全长半胱天冬酶9															
	<400>	46															
	Met	Asp	Glu	Ala	Asp	Arg	Arg	Leu	Leu	Arg	Arg	Cys	Arg	Leu	Arg	Leu	
	1				5					10					15		
	Val	Glu	Glu	Leu	Gln	Val	Asp	Gln	Leu	Trp	Asp	Ala	Leu	Leu	Ser	Arg	

	20	25	30
	Glu Leu Phe Arg Pro His Met 35	Ile Glu Asp Ile Gln Arg Ala Gly Ser 40	45
	Gly Ser Arg Arg Asp Gln Ala Arg Gln Leu Ile 50	Ile Asp Leu Glu Thr 55	60
	Arg Gly Ser Gln Ala Leu Pro Leu Phe Ile 65	Ser Cys Leu Glu Asp Thr 70	75 80
	Gly Gln Asp Met Leu Ala Ser Phe Leu Arg Thr Asn Arg Gln Ala Ala 85	90	95
	Lys Leu Ser Lys Pro Thr Leu Glu Asn Leu Thr Pro Val Val Leu Arg 100	105	110
	Pro Glu Ile Arg Lys Pro Glu Val Leu Arg Pro Glu Thr Pro Arg Pro 115	120	125
	Val Asp Ile Gly Ser Gly Gly Phe Gly Asp Val Gly Ala Leu Glu Ser 130	135	140
	Leu Arg Gly Asn Ala Asp Leu Ala Tyr Ile Leu Ser Met Glu Pro Cys 145	150	155 160
	Gly His Cys Leu Ile Ile Asn Asn Val Asn Phe Cys Arg Glu Ser Gly 165	170	175
[0025]	Leu Arg Thr Arg Thr Gly Ser Asn Ile Asp Cys Glu Lys Leu Arg Arg 180	185	190
	Arg Phe Ser Ser Leu His Phe Met Val Glu Val Lys Gly Asp Leu Thr 195	200	205
	Ala Lys Lys Met Val Leu Ala Leu Leu Glu Leu Ala Gln Gln Asp His 210	215	220
	Gly Ala Leu Asp Cys Cys Val Val Val Ile Leu Ser His Gly Cys Gln 225	230	235 240
	Ala Ser His Leu Gln Phe Pro Gly Ala Val Tyr Gly Thr Asp Gly Cys 245	250	255
	Pro Val Ser Val Glu Lys Ile Val Asn Ile Phe Asn Gly Thr Ser Cys 260	265	270
	Pro Ser Leu Gly Gly Lys Pro Lys Leu Phe Phe Ile Gln Ala Cys Gly 275	280	285
	Gly Glu Gln Lys Asp His Gly Phe Glu Val Ala Ser Thr Ser Pro Glu 290	295	300
	Asp Glu Ser Pro Gly Ser Asn Pro Glu Pro Asp Ala Thr Pro Phe Gln 305	310	315 320
	Glu Gly Leu Arg Thr Phe Asp Gln Leu Asp Ala Ile Ser Ser Leu Pro		

	325	330	335
	Thr Pro Ser Asp Ile Phe Val Ser Tyr Ser Thr Phe Pro Gly Phe Val 340	345	350
	Ser Trp Arg Asp Pro Lys Ser Gly Ser Trp Ser Ala Asp Gly Gln Leu 355	360	365
	Tyr Thr Thr Leu Leu Glu Thr Leu Asp Asp Ile Phe Glu Gln Trp Ala 370	375	380
	His Ser Glu Asp Leu Gln Ser Leu Leu Leu Arg Val Ala Asn Ala Val 385	390	395
	Ser Val Lys Gly Ile Tyr Lys Gln Met Pro Gly Cys Phe Asn Phe Leu 405	410	415
	Arg Lys Lys Leu Phe Phe Lys Thr Ser Gly Gly Gly Gly Ser 420	425	430
	<210> 47 <211> 154 <212> DNA <213> 人工序列		
	<220> <223> UPR诱导型启动子		
[0026]	<400> 47 tgacgtgctg atgatgcgta gctagtagtt gacgtgctga tgatgcgtag ctagtagttg acgtgctgat gatgcgtagc tagtagttga cgtgctgatg atgcgtagct agtagttgac gtgctgatga tgcgtagcta gtagttgacg tgct	60 120 154	
	<210> 48 <211> 35 <212> DNA <213> 人工序列		
	<220> <223> 可调节内含子		
	<400> 48 cagctgcagc actcagacta cgtgcacctc tgctg	35	
	<210> 49 <211> 20 <212> RNA <213> 人工序列		
	<220> <223> intron sequence inside cleavage sites		
	<400> 49 acgggcaacu uuacacgacg	20	
	<210> 50 <211> 32 <212> RNA <213> 人工序列		
	<220> <223> 可调节内含子序列		

<220>
 <221> misc_feature
 <222> (2)..(2)
 <223> n是a、c、g或u

 <220>
 <221> misc_feature
 <222> (31)..(31)
 <223> n是a、c、g或u

 <400> 50
 cngcagacgg gcaacuuuac acgacgcugc ng 32

 <210> 51
 <211> 5
 <212> PRT
 <213> 人工序列

 <220>
 <223> 截短的EGFP蛋白 - 第一终止密码子至第二终止密码子

 <400> 51
 Ala Pro Ser Pro Pro
 1 5

 <210> 52
 <211> 16
 <212> PRT
 <213> 人工序列

 <220>
 <223> 截短的EGFP蛋白 - 从第二终止密码子至第三终止密码子

 <400> 52
 [0027] Ala Lys Thr Pro Thr Arg Ser Ala Ile Thr Trp Ser Cys Trp Ser Ser
 1 5 10 15

 <210> 53
 <211> 13
 <212> PRT
 <213> 人工序列

 <220>
 <223> 从第三终止密码子至蛋白质末端的截短的EGFP蛋白

 <400> 53
 Pro Pro Pro Gly Ser Leu Ser Ala Trp Thr Ser Cys Thr
 1 5 10

 <210> 54
 <211> 47
 <212> DNA
 <213> 人工序列

 <220>
 <223> UPR响应性CRE

 <220>
 <221> misc_feature
 <222> (8)..(8)
 <223> 任选的间隔序列

 <220>
 <221> misc_feature
 <222> (16)..(16)
 <223> 任选的间隔序列

 <220>

	<221> misc_feature	
	<222> (24)..(24)	
	<223> 任选的间隔序列	
	<220>	
	<221> misc_feature	
	<222> (32)..(32)	
	<223> 任选的间隔序列	
	<220>	
	<221> misc_feature	
	<222> (40)..(40)	
	<223> 任选的间隔序列	
	<400> 54	
	tgacgtgntg acgtgntgac gtgntgacgt gntgacgtgn tgacgtg	47
	<210> 55	
	<211> 59	
	<212> DNA	
	<213> 人工序列	
	<220>	
	<223> UPR响应性CRE	
	<220>	
	<221> misc_feature	
	<222> (10)..(10)	
	<223> 任选的间隔序列	
	<220>	
	<221> misc_feature	
	<222> (20)..(20)	
	<223> 任选的间隔序列	
[0028]	<220>	
	<221> misc_feature	
	<222> (30)..(30)	
	<223> 任选的间隔序列	
	<220>	
	<221> misc_feature	
	<222> (40)..(40)	
	<223> 任选的间隔序列	
	<220>	
	<221> misc_feature	
	<222> (50)..(50)	
	<223> 任选的间隔序列	
	<400> 55	
	tgacgtgctn tgacgtgctn tgacgtgctn tgacgtgctn tgacgtgctn tgacgtgct	59
	<210> 56	
	<211> 154	
	<212> DNA	
	<213> 人工序列	
	<220>	
	<223> UPR响应性CRE	
	<400> 56	
	tgacgtgctg atgatgcgta gctagtagtt gacgtgctga tgatgcgtag ctagtagttg	60
	acgtgctgat gatgcgtagc tagtagttga cgtgctgatg atgcgtagct agtagttgac	120
	gtgctgatga tgcgtagcta gtagttgacg tgct	154
	<210> 57	
	<211> 255	
	<212> DNA	
	<213> 人工序列	

	<220>		
	<223>	UPR诱导型启动子	
	<400>	57	
	tgacgtgctg atgatgcgta gctagtagtt gacgtgctga tgatgcgtag ctagtagttg	60	
	acgtgctgat gatgcgtagc tagtagttga cgtgctgatg atgcgtagct agtagttgac	120	
	gtgctgatga tgcgtagcta gtagttgacg tgctgatgat gcgtagctag tagtgcagtt	180	
	agcgtagctg aggtaccgtc gacgatatcg gatccaggtc tatataagca gagctcgttt	240	
	agtgaaccgt cagat	255	
	<210>	58	
	<211>	255	
	<212>	DNA	
	<213>	人工序列	
	<220>		
	<223>	UPR诱导型启动子	
	<400>	58	
	tgacgtgctg atgatgcgta gctagtagtt gacgtgctga tgatgcgtag ctagtagttg	60	
	acgtgctgat gatgcgtagc tagtagttga cgtgctgatg atgcgtagct agtagttgac	120	
	gtgctgatga tgcgtagcta gtagttgacg tgctgatgat gcgtagctag tagtgcagtt	180	
	agcgtagctg aggtaccgtc gacgatatcg gaccccttcgc atattaaggt gacgcgtgtg	240	
	gcctcgaaca ccgag	255	
	<210>	59	
	<211>	2269	
[0029]	<212>	DNA	
	<213>	人工序列	
	<220>		
	<223>	构建体 ATF06-MP-001	
	<400>	59	
	gaagttccta ttccgaagtt cctattctct agaaagtata ggaacttctg acgtgctgat	60	
	gatgcgtagc tagtagttga cgtgctgatg atgcgtagct agtagttgac gtgctgatga	120	
	tgcgtagcta gtagttgacg tgctgatgat gcgtagctag tagttgacgt gctgatgatg	180	
	cgtagctagt agttgacgtg ctgatgatgc gtagctagta gtgcagttag cgtagctgag	240	
	gtaccgtcga cgatatcgga tccttcgcat attaagggtga cgcgtgtggc ctcgaacacc	300	
	gagcgaccct gcagcgaccc gcttaaatgg aagatgccaa aaacattaag aagggccccag	360	
	cgccattcta cccactcgaa gacgggaccg ccggcgagca gctgcacaaa gccatgaagc	420	
	gctacgccct ggtgcccggc accatcgctt ttaccgacgc acatatcgag gtggacatta	480	
	cctacgccga gtacttcgag atgagcggtc ggctggcaga agctatgaag cgctatgggc	540	
	tgaatacaaa ccacgggacg gtggtgtgca gcgagaatag cttgcagttc ttcatgcccg	600	
	tgttgggtgc cctgttcacg ggtgtggctg tggccccagc taacgacatc tacaacgagc	660	
	gcgagctgct gaacagcatg ggcatcagcc agcccaccgt cgtattcgtg agcaagaaag	720	
	ggctgcaaaa gatcctcaac gtgcaaaaga agctaccgat catacaaaag atcatcatca	780	
	tggatagcaa gaccgactac cagggcttcc aaagcatgta caccttcgtg acttccatt	840	
	tgccaccggg cttcaacgag tacgacttcg tgcccagagag cttcgaccgg gacaaaacca	900	
	tcgccctgat catgaacagt agtggcagta ccgattgcc caaggcgta gccctaccgc	960	

[0030]

```

accgcaccgc ttgtgtccga ttcagtcgat cccgcgaccc catcttcggc aaccagatca 1020
tccccgacac cgctatcctc agcgtgggtc catttcacca cggcttcggc atgttcacca 1080
cgctggggcta cttgatctgc ggctttcggg tcgtgtcat gtaccgttc gaggaggagc 1140
tattcttgcg cagcttgcaa gactataaga ttcaatctgc cctgctgggtg cccacactat 1200
ttagcttctt cgtaagagc actctcatcg acaagtacga cctaagcaac ttgcacgaga 1260
tcgccagcgg cggggcgccg ctacgaagg aggtagggtga ggccgtggcc aaacgcttcc 1320
acctaccagg catccgccag ggctacggcc tgacagaaac aaccagcgcc attctgatca 1380
ccccgaagg ggacgacaag cctggcgag taggcaaggt ggtgcccttc ttcgaggcta 1440
aggtgggtgga ctggacacc ggtaagacac tgggtgtgaa ccagcgcggc gagctgtgcg 1500
tccgtggccc catgatcatg agcggctacg ttaacaaccc cgaggctaca aacgtctca 1560
tcgacaagga cggtggctg cacagcggcg acatcgcta ctgggacgag gacgagcact 1620
tcttcacgt ggaccggctg aagagcctga tcaatacaa gggctaccag gtagccccag 1680
ccgaactgga gagcatcctg ctgcaacacc ccaacatctt cgacgcggg gtcgccggcc 1740
tgcccgaca cgatgccgc gagctgccc cgcagtcgt cgtgctggaa cacggtaaaa 1800
ccatgaccga gaaggagatc gtggactatg tggccagcca ggttacaacc gccaagaagc 1860
tgcgcggtgg tgttgtgtc gtggacgagg tgcctaaagg actgaccggc aagttggacg 1920
cccgaagat ccgcgagatt ctcatgaagg ccaagaaggg cgcaagatc gccgtgtaat 1980
gaaagcttgg tctctacgag taatagacgc ccagttgaat tccttcgagc agacatgata 2040
agatacattg atgagtttgg acaaacaca actagaatgc agtgaaaaaa atgctttatt 2100
tgtgaaattt gtgatgctat tgctttattt gtaaccatta taagctgcaa taaacaagtt 2160
aacaacaaca attgcattca ttttatgttt caggttcagg gggagggtgtg ggagggtttt 2220
taaagcaagt aaaacctcta caaatgttgt aaaatcgata aggatccgt 2269

```

<210> 60
 <211> 2425
 <212> DNA
 <213> 人工序列

<220>
 <223> 构建体ATF06-MP-002

```

<400> 60
gcgacgtaat acgactcact atagggcgaa ttggcggaag gccgtcaagg ccgcatgcat 60
aataaaatat ctttattttc attacatctg tgtgttggtt ttttgttgt gacgtgctga 120
tgatgcgtag ctagtagttg acgtgctgat gatgcgtagc tagtagttga cgtgctgatg 180
atgcgtagct agtagttgac gtgctgatga tgcgtagcta gtagttgacg tgctgatgat 240
gcgtagctag tagttgacgt gctgatgatg cgtagctagt agtgcagtta gcgtagctga 300
ggtaccgtcg acgatatcgg atccaggctt atataagcag agctcgttta gtgaaccgtc 360
agatcgccca gatacgccat ccacgtgtt ttgacctca tagaagatcg ccaccatgga 420
agatgcaaaa aacattaaga agggcccagc gccattctac ccactcgaag acgggaccgc 480
cggcgagcag ctgcacaaag ccatgaagcg ctacgccctg gtgcccggca ccacgcctt 540
taccgacgca catatcgagg tggacattac ctacgccgag tacttcgaga tgagcggttcg 600
gctggcgaaa gctatgaagc gctatgggct gaatacaaac catcggatcg tgggtgtcag 660

```

	cgagaatagc ttgcagttct tcatgcccgt gttgggtgcc ctgttcatcg gtgtggctgt	720
	ggccccagct aacgacatct acaacgagcg cgagctgctg aacagcatgg gcatcagcca	780
	gccaccgctc gtattcgtga gcaagaaagg gctgcaaaag atcctcaacg tgcaaaagaa	840
	gctaccgata atacaaaaga tcatcatcat ggatagcaag accgactacc agggcttcca	900
	aagcatgtac accttcgtga cttcccatct gccacccggc ttcaacgagt acgacttcgt	960
	gcccagagac ttcgaccggg acaaaacat cgccctgata atgaacagta gtggcagtac	1020
	cggattgccc aaggcgtag ccctaccgca ccgacccgct tgtgtccgat tcatcctgc	1080
	ccgcgacccc atcttcggca accagatcat ccccgacacc gctatcctca gcgtggtgcc	1140
	atttcaccac ggcttcggca tgttaccac gctgggttac ttgatctgcg gctttcgggt	1200
	cgtgctcatg taccgcttcg aggaggagct attcttgcgc agcttgcaag actataagat	1260
	tcaatctgcc ctgctggtgc ccacactatt tagcttcttc gctaagagca ctctcatcga	1320
	caagtacgac ctaagcaact tgcacgagat cgccagcggc gggcgccgc tcagcaagga	1380
	ggtaggtagg gccgtggcca aacgcttcca cctaccaggc atccgccagg gctacggcct	1440
	gacagaaaca accagcgcca ttctgatcac ccccgaaagg gacgacaagc ctggcgcagt	1500
	aggcaagggt gtgcccttct tcgaggctaa ggtgggtggc ttggacaccg gtaagacact	1560
	gggtgtgaac cagcgccggc agctgtgcgt ccgtggcccc atgatcatga gcggctacgt	1620
	taacaacccc gaggtacaa acgctctcat cgacaaggac ggctggctgc acagcggcga	1680
[0031]	catcgcttac tgggacgagg acgagcactt cttcatcgtg gaccggctga agagcctgat	1740
	caataacaag ggctaccagg tagccccagc cgaactggag agcatcctgc tgcaacaccc	1800
	caacatcttc gacgccgggg tcgccgcct gcccgacgac gatgccggcg agctgccccg	1860
	cgcagtcgtc gtgctggaac acggtaaaac catgaccgag aaggagatcg tggactatgt	1920
	ggccagccag gttacaaccg ccaagaagct gcgcgggtgt gttgtgttcg tggacgaggt	1980
	gcctaaagga ctgaccggca agttggacgc ccgcaagatc cgcgagattc tcattaaggc	2040
	caagaagggc ggcaagatcg ccgtgtaatg aaagcttggt ctctacgagt aatagacgcc	2100
	cagttgaatt ccttcgagca gacatgataa gatacattga tgagtttgga caaaccacaa	2160
	ctagaatgca gtgaaaaaaa tgctttatct gtgaaatttg tgatgctatt gctttatttg	2220
	taaccattat aagctgcaat aaacaagtta acaacaaca ttgcattcat tttatgtttc	2280
	aggttcaggg ggaggtgtgg gaggtttttt aaagcaagta aaacctctac aaatgtggta	2340
	aaatcgataa ggatccgtct gggcctcatg ggccttccgc tcaactgccc ctttccagtc	2400
	gggaaacctg tcgtgccagc tgcatt	2425
	<210> 61	
	<211> 20	
	<212> DNA	
	<213> 人工序列	
	<220>	
	<223> 间隔序列	
	<400> 61	
	gatgatgcgt agctagtagt	20

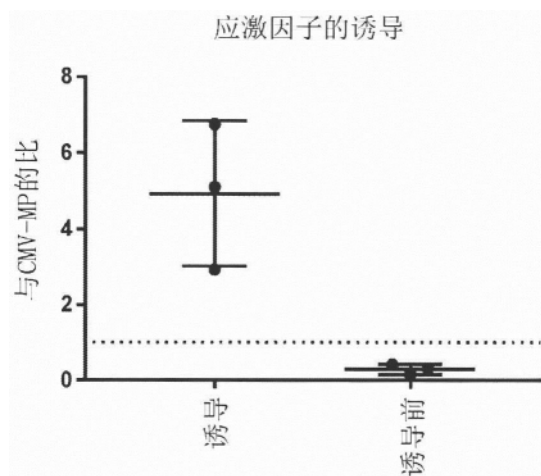


图1

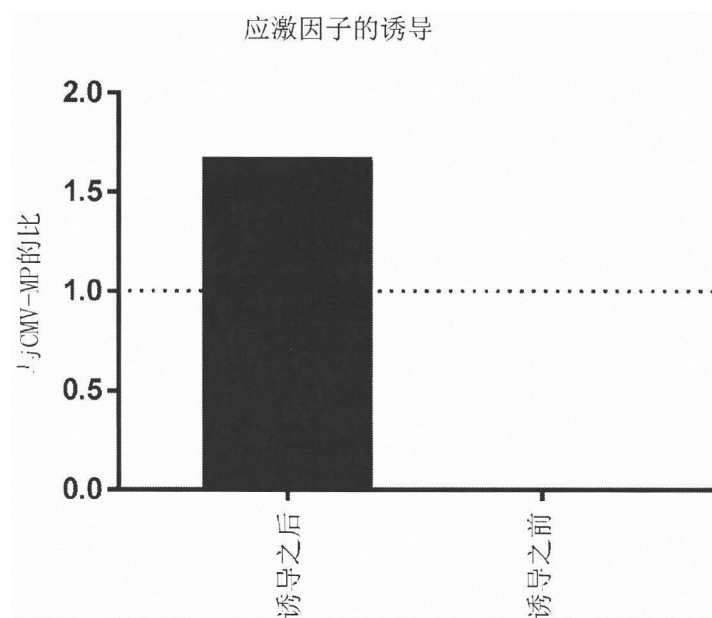


图2a

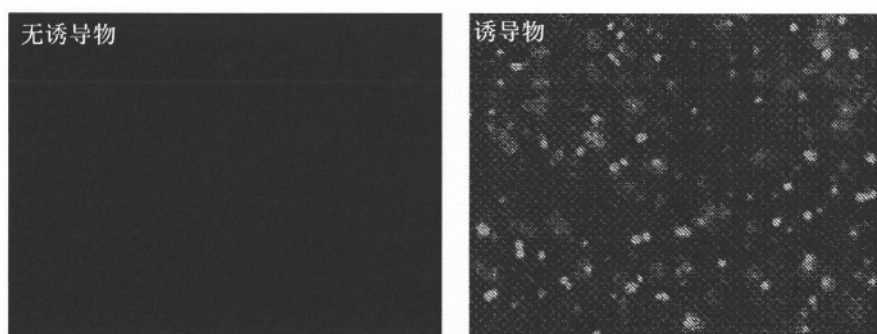


图2b

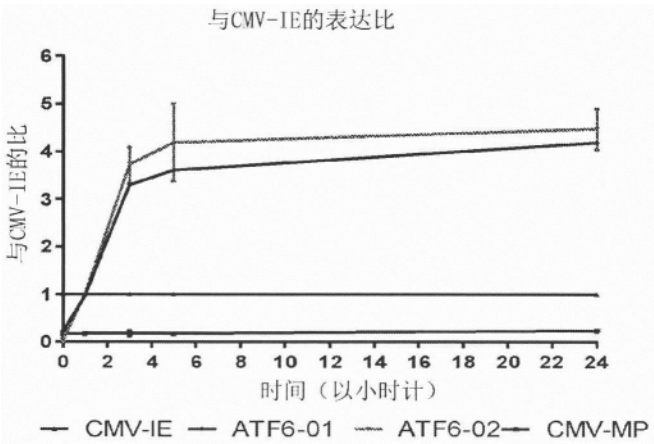


图3

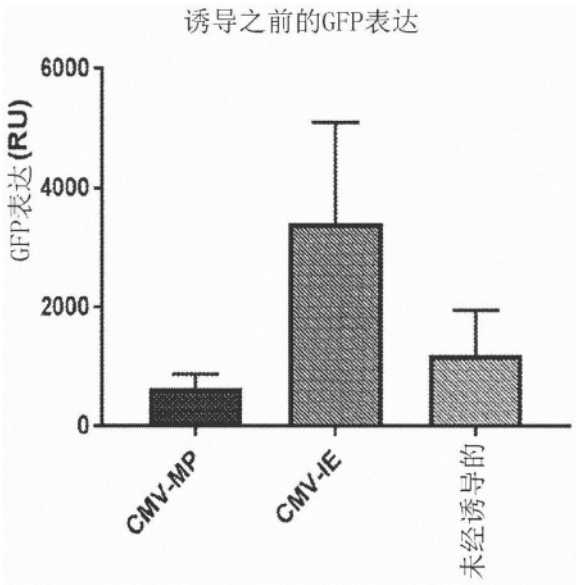


图4

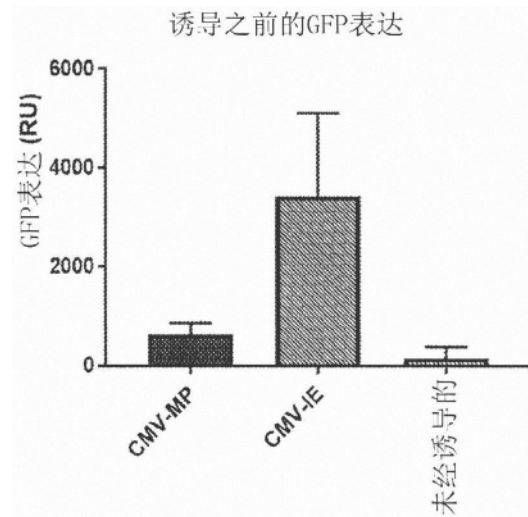


图5

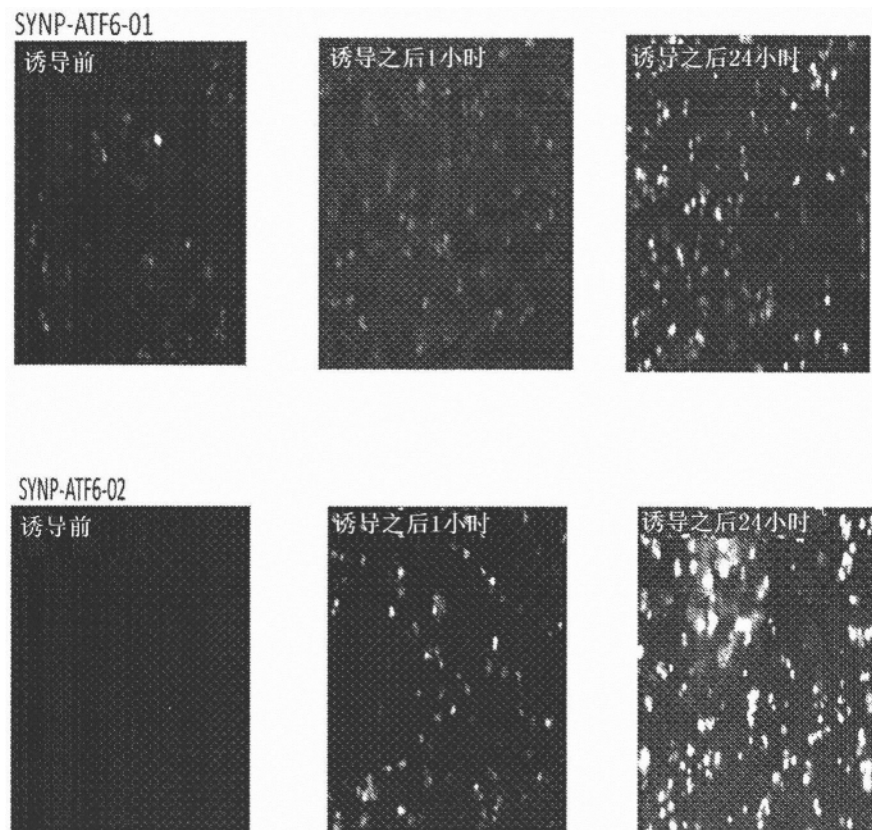


图6

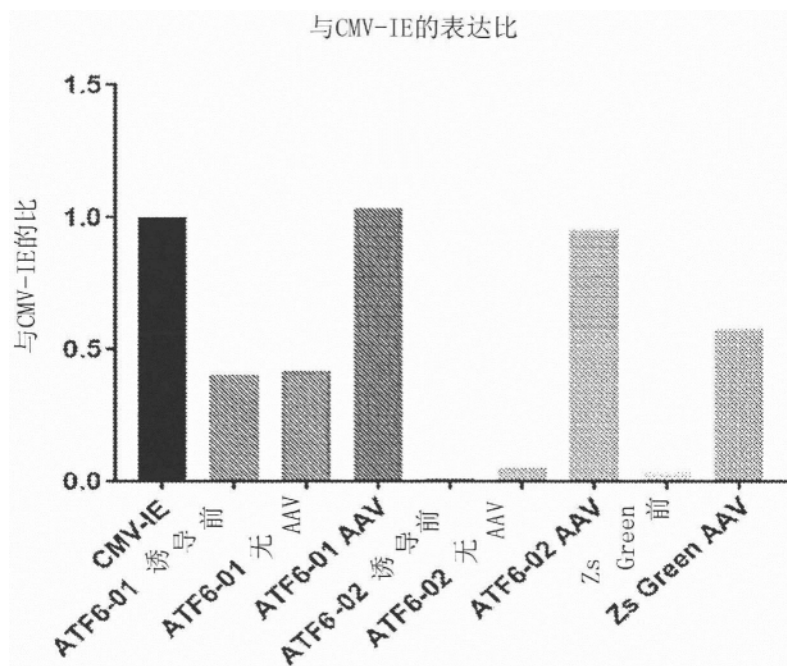


图7

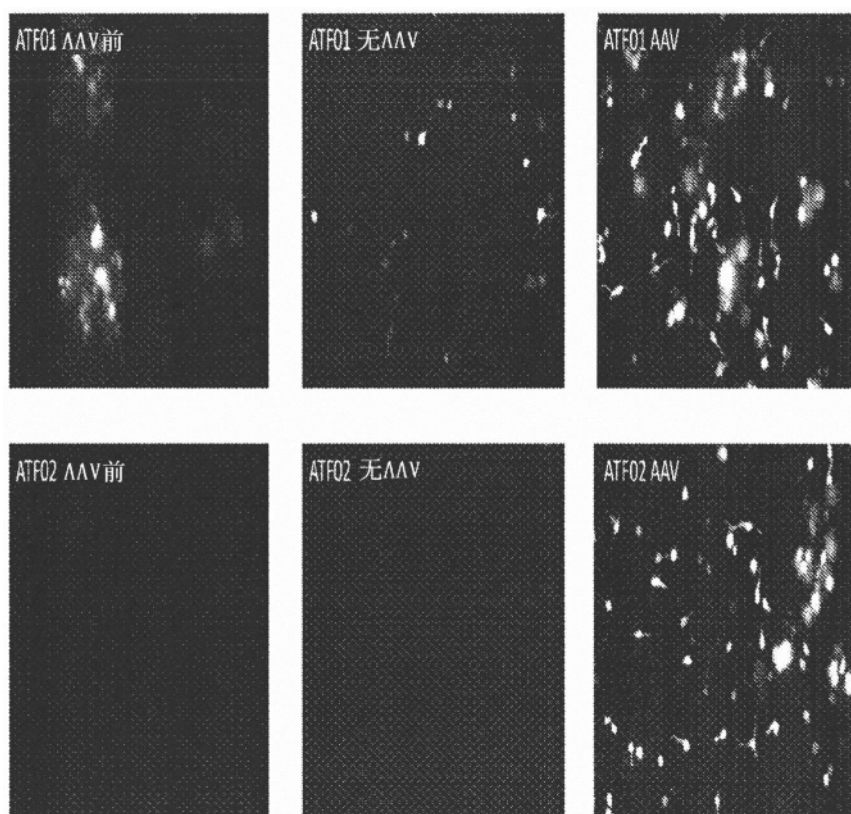


图8

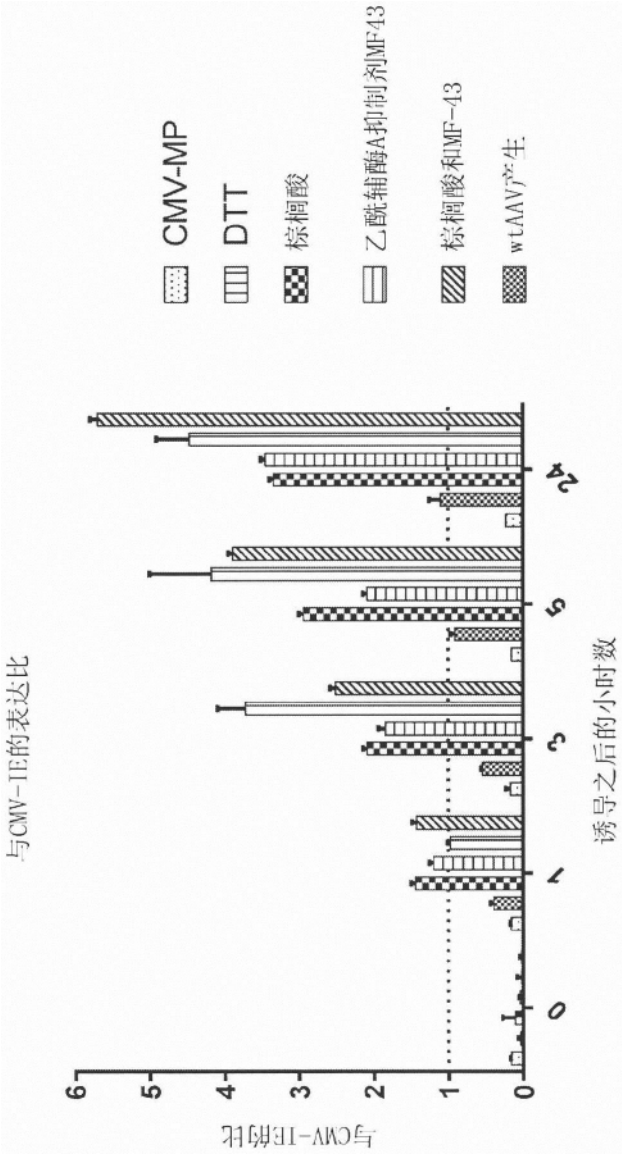


图9

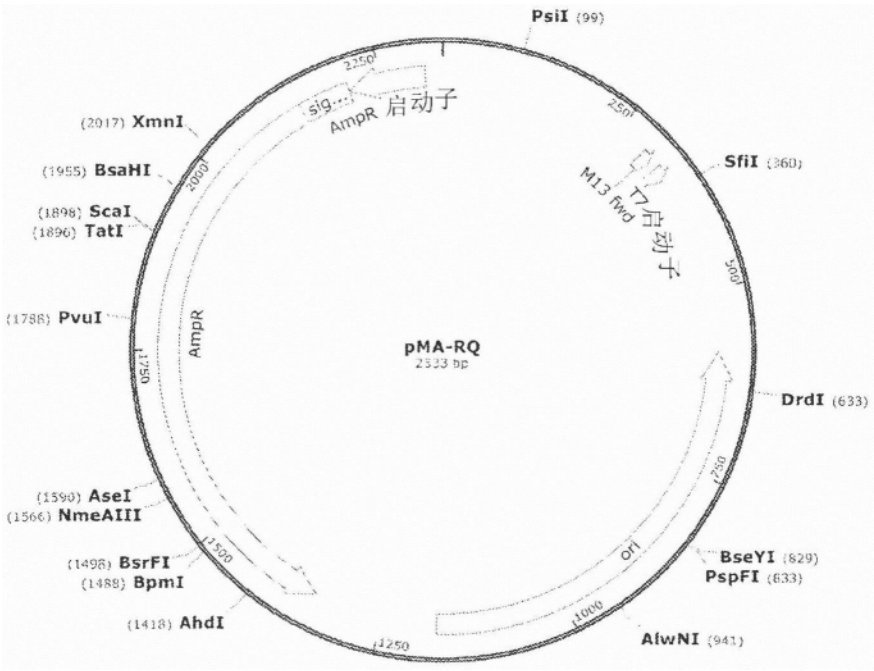


图10

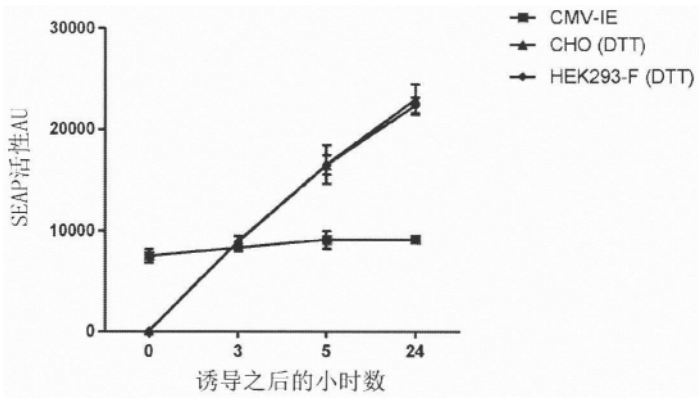


图11

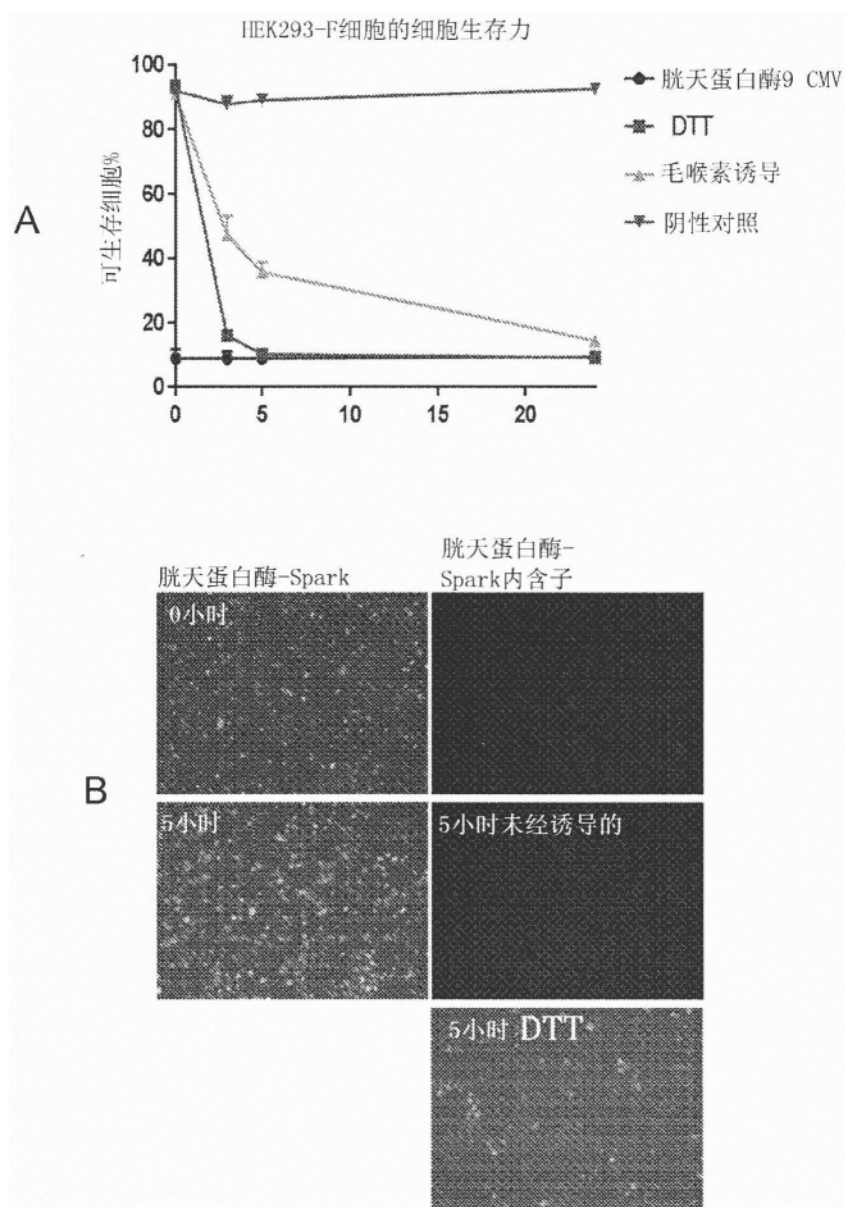


图14

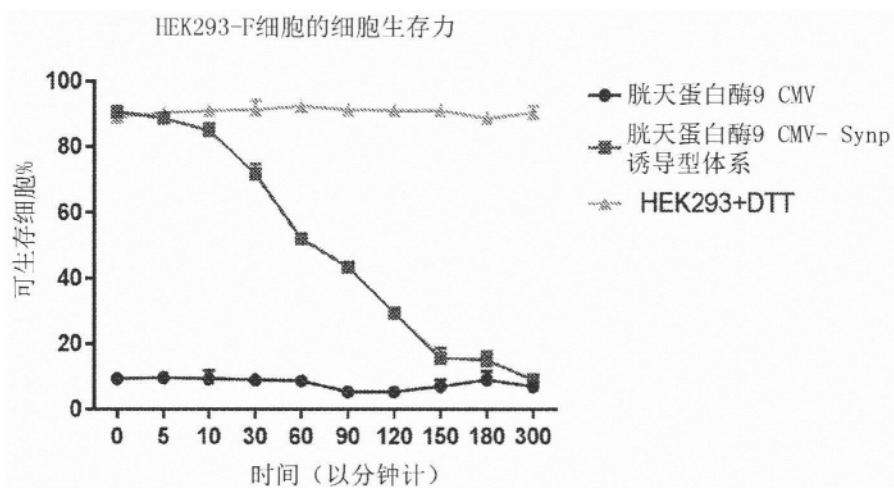


图15

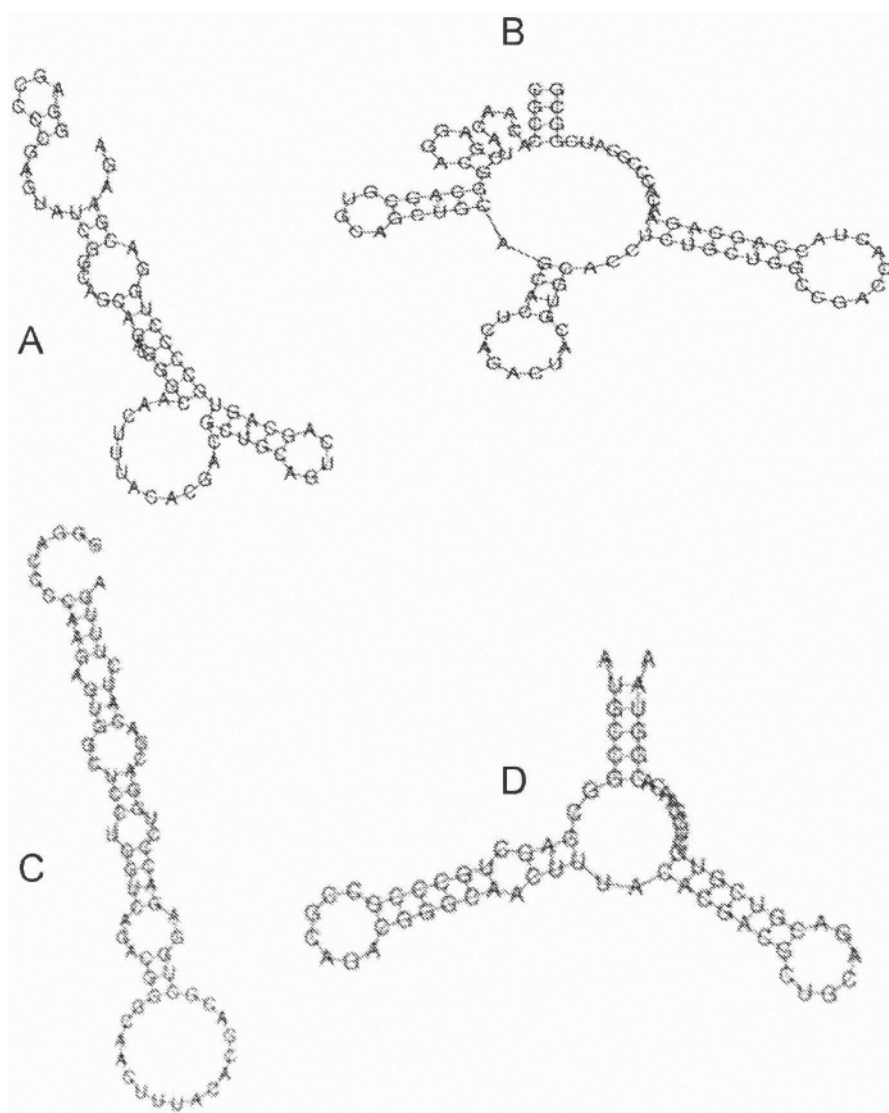


图16

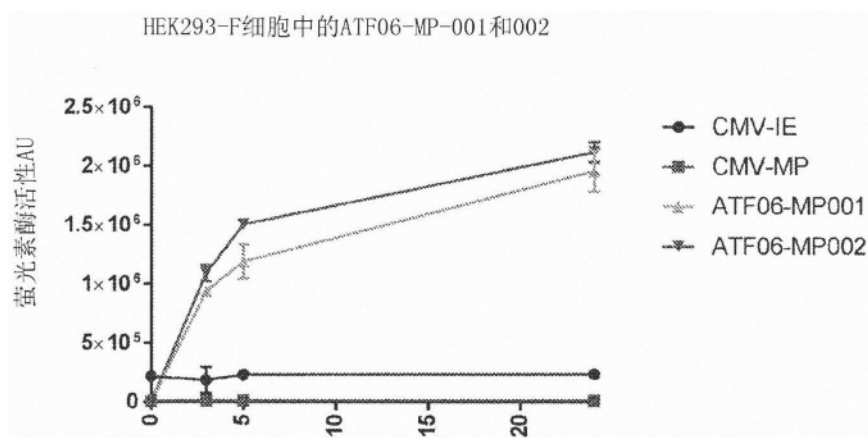


图17