



(12) 发明专利申请

(10) 申请公布号 CN 102812040 A

(43) 申请公布日 2012. 12. 05

(21) 申请号 201080061134. 4

R · A · 戴维斯 - 泰伯 E · 冯

(22) 申请日 2010. 10. 28

(74) 专利代理机构 中国专利代理 (香港) 有限公司 72001

(30) 优先权数据

61/256, 544 2009. 10. 30 US

代理人 李连涛 郭文洁

(85) PCT 申请进入国家阶段日

2012. 06. 29

(51) Int. Cl.

C07K 16/00 (2006. 01)

C12P 21/02 (2006. 01)

C12P 21/06 (2006. 01)

(86) PCT 申请的申请数据

PCT/US2010/054475 2010. 10. 28

(87) PCT 申请的公布数据

W02011/053699 EN 2011. 05. 05

(71) 申请人 雅培制药有限公司

地址 美国伊利诺伊州

(72) 发明人 G · R · 卡森 W · R · 焦恩

Y · Z · 库尼斯 W · F · 莱斯三世

权利要求书 3 页 说明书 58 页

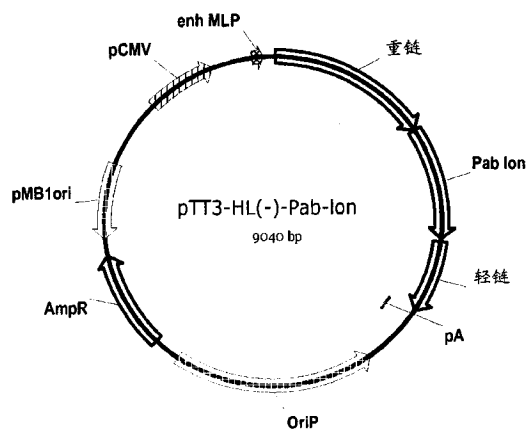
序列表 59 页 附图 14 页

(54) 发明名称

SORF 构建体和多基因表达

(57) 摘要

本发明的实施方案涉及用于表达多肽的载体构建体和方法,所述多肽包括多聚的产物,诸如治疗性抗体。特定构建体允许从单个开放读码框(sORF)制备表达产物。一个实施方案提供了用于产生一种或多种重组蛋白产物的分离的或纯化的表达载体,所述表达载体包含单个开放读码框插入物;所述插入物包含:编码信号肽的信号肽核酸序列;编码第一多肽的第一核酸序列;编码第一蛋白切割位点的第一插入核酸序列,其中所述第一蛋白切割位点由火球菌属的 lon 蛋白酶基因的内含肽区段、或火球菌属或甲烷球菌属的 k1bA 基因的内含肽区段、或源自它们的修饰的内含肽区段提供;和编码第二多肽的第二核酸序列。构建体和方法的某些实施方案采用:Pyrococcus abyssi、激烈火球菌或掘越氏火球菌 OT3 的 lon 蛋白酶基因的内含肽区段;或 Pyrococcus abyssi、激烈火球菌或詹氏甲烷球菌的 k1bA 基因的内含肽区段;或其它内含肽区段。



1. 用于产生一种或多种重组蛋白产物的分离的或纯化的表达载体,所述表达载体包含单个开放读码框插入物;所述插入物包含:

(a) 编码信号肽的信号肽核酸序列;

(b) 编码第一多肽的第一核酸序列;

(c) 编码第一蛋白切割位点的第一插入核酸序列,其中所述第一蛋白切割位点由火球菌属的 lon 蛋白酶基因的内含肽区段、或火球菌属或甲烷球菌属的 k1bA 基因的内含肽区段、或源自它们的修饰的内含肽区段提供;和

(d) 编码第二多肽的第二核酸序列;

其中所述编码所述第一蛋白切割位点的第一插入核酸序列可操作地位于所述第一核酸序列和所述第二核酸序列之间;

其中所述编码所述信号肽的信号肽核酸序列可操作地位于所述第一核酸序列之前;且

其中所述表达载体能够表达单个开放读码框多肽,所述多肽可在所述第一蛋白切割位点处被切割。

2. 根据权利要求 1 所述的表达载体,其中所述第一蛋白切割位点由下述内含肽区段提供:Pyrococcus abyssi、激烈火球菌或掘越氏火球菌 OT3 的 lon 蛋白酶基因的内含肽区段;或 Pyrococcus abyssi、激烈火球菌或詹氏甲烷球菌的 k1bA 基因的内含肽区段;或分别源自它们的修饰的内含肽区段。

3. 根据权利要求 1 所述的表达载体,其中所述内含肽区段或修饰的内含肽区段编码倒数第二位残基,所述残基是赖氨酸、丝氨酸或不是组氨酸。

4. 根据权利要求 1 所述的表达载体,其中所述内含肽区段或修饰的内含肽区段能够被切割,但是不能使所述第一多肽与所述第二多肽完全相连。

5. 根据权利要求 1 所述的表达载体,其中所述第一蛋白切割位点是由下述内含肽区段提供:包含选自 SEQ ID NO:1、3、4、6、7、55、35、37 和 39 的序列的内含肽区段,和源自它们的修饰的内含肽区段。

6. 根据权利要求 1-5 中任一项所述的表达载体,其中所述第一多肽和第二多肽能够多聚装配。

7. 根据权利要求 1-6 中任一项所述的表达载体,其中所述第一多肽和第二多肽中的至少一种能够胞外分泌。

8. 根据权利要求 1-7 中任一项所述的表达载体,其中所述第一多肽和第二多肽中的至少一种具有哺乳动物起源。

9. 根据权利要求 1-8 中任一项所述的表达载体,其中所述第一多肽包含免疫球蛋白重链或其功能片段,且所述第二多肽包含免疫球蛋白轻链或其功能片段,且所述第一多肽是在所述第二多肽的上游。

10. 根据权利要求 1-9 中任一项所述的表达载体,所述载体仅包含一个所述信号肽核酸序列。

11. 根据权利要求 1-10 中任一项所述的表达载体,所述表达载体另外包含编码第三多肽的第三核酸序列和编码第二蛋白切割位点的第二插入核酸序列;其中所述第二插入核酸序列和第三核酸序列以该次序可操作地位于所述第二核酸序列之后。

12. 根据权利要求 1-11 中任一项所述的表达载体,其中所述第一多肽和所述第二多肽

包含功能抗体或其它抗原识别分子；其中抗原特异性指向结合选自下述的抗原：TNF α （肿瘤坏死因子- α ）、促红细胞生成素受体、RSV、EL/选择蛋白、白介素-1、白介素-12、白介素-13、白介素-18、白介素-23、白介素-33、CD81、CD19、IGF1、IGF2、EGFR、CXCL-13、GLP-1R、前列腺素 E2 和淀粉样蛋白 β 。

13. 根据权利要求 1-12 中任一项所述的表达载体，其中所述第一多肽和第二多肽包含一对免疫球蛋白链，它们来自 D2E7、EL246、ABT-007、ABT-325 或 ABT-874 的抗体。

14. 根据权利要求 1-13 中任一项所述的表达载体，其中所述第一多肽和第二多肽各自独立地选自：免疫球蛋白重链或免疫球蛋白轻链区段，它们来自 D2E7、EL246、ABT-007、ABT-325、ABT-874 或其它抗体的类似区段。

15. 根据权利要求 1-14 中任一项所述的表达载体，其中所述载体另外包含所述插入物的启动子调控元件。

16. 根据权利要求 15 所述的表达载体，其中所述启动子调控元件是诱导型或组成型。

17. 根据权利要求 15 所述的表达载体，其中所述启动子调控元件是组织特异性的。

18. 根据权利要求 15 所述的表达载体，其中所述启动子包含腺病毒主要晚期启动子。

19. 一种宿主细胞，其包含根据权利要求 1-18 中任一项所述的载体。

20. 根据权利要求 19 所述的宿主细胞，其中所述宿主细胞是原核细胞。

21. 根据权利要求 20 所述的宿主细胞，其中所述宿主细胞是大肠杆菌。

22. 根据权利要求 19 所述的宿主细胞，其中所述宿主细胞是真核细胞。

23. 根据权利要求 22 所述的宿主细胞，其中所述真核细胞选自：原生生物细胞、动物细胞、植物细胞和真菌细胞。

24. 根据权利要求 23 所述的宿主细胞，其中所述真核细胞是选自下述的动物细胞：哺乳动物细胞、禽细胞和昆虫细胞。

25. 根据权利要求 24 所述的宿主细胞，其中所述宿主细胞是哺乳动物细胞系。

26. 根据权利要求 24 所述的宿主细胞，其中所述宿主细胞是 CHO 细胞或二氢叶酸还原酶-缺陷型 CHO 细胞。

27. 根据权利要求 24 所述的宿主细胞，其中所述宿主细胞是 COS 细胞或 HEK 细胞。

28. 根据权利要求 23 所述的宿主细胞，其中所述宿主细胞是酵母细胞。

29. 根据权利要求 28 所述的宿主细胞，其中所述酵母细胞是酿酒酵母。

30. 根据权利要求 24 所述的宿主细胞，其中所述宿主细胞是草地贪夜蛾 Sf9 昆虫细胞。

31. 一种用于生产重组多蛋白或多种蛋白的方法，所述方法包括：在足以允许载体蛋白表达的条件下，在培养基中培养根据权利要求 19 所述的宿主细胞。

32. 根据权利要求 31 所述的方法，所述方法另外包括：回收和 / 或纯化所述载体蛋白。

33. 根据权利要求 31-32 中任一项所述的方法，其中所述多种蛋白能够多聚装配。

34. 根据权利要求 31-33 中任一项所述的方法，其中所述重组多蛋白或多种蛋白是生物学上有功能的和 / 或治疗性的。

35. 一种用于生产重组产物的方法，其中所述产物是免疫球蛋白蛋白质或其功能片段、装配的抗体或其它抗原识别分子，所述方法包括：在足以生产所述重组产物的条件下，在培养基中培养根据权利要求 19 所述的宿主细胞。

36. 根据权利要求 31-35 中任一项所述的方法生产的蛋白。

37. 根据权利要求 31-36 中任一项所述的方法生产的多蛋白。

38. 根据权利要求 31-37 中任一项所述的方法生产的装配的免疫球蛋白、装配的其它抗原识别分子、或单个免疫球蛋白链或其功能片段。

39. 根据权利要求 38 所述的免疫球蛋白、其它抗原识别分子或单个免疫球蛋白链或其功能片段,其中存在实现或促进特定抗原与下述物质的结合的能力:肿瘤坏死因子- α 、促红细胞生成素受体、RSV、EL/ 选择蛋白、白介素-1、白介素-12、白介素-13、白介素-18、白介素-23、白介素-33、CD81、CD19、IGF1、IGF2、EGFR、前列腺素 E2、CXCL-13、GLP-1R 或淀粉样蛋白 β 。

40. 根据权利要求 39 所述的免疫球蛋白或其功能片段,其中所述免疫球蛋白是 D2E7 或 ABT-874,或者其中所述功能片段是它们各自的片段。

41. 一种药物组合物,其包含治疗有效量的根据权利要求 36-40 中任一项所述的蛋白和药学上可接受的载体。

42. 根据权利要求 1-18 中任一项所述的表达载体,其另外包含编码标签的核酸序列。

43. 根据权利要求 1-18 中任一项所述的表达载体,其中所述插入核酸序列另外编码标签。

44. 一种表达载体、具有所述载体的宿主细胞、载体表达产物、药物组合物和 / 或制备或使用前述任一种的方法,其中所述载体是根据权利要求 1-18 中任一项所述的载体,且另外包含编码轻链信号肽的区段。

45. 根据权利要求 44 所述的载体、宿主细胞、载体表达产物、药物组合物和 / 或制备或使用方法,其中所述编码的轻链信号肽是选自下述的 κ 轻链信号肽:A17、A18、A19、A26 和 H2G。

46. 根据权利要求 44 所述的载体、宿主细胞、载体表达产物、药物组合物和 / 或制备或使用方法,其中所述编码的轻链信号肽是 VKII κ 轻链信号肽 A18,SEQ ID NO:82(氨基酸序列 MRLPAQLLGLLMLWIPGSSA)。

SORF 构建体和多基因表达

[0001] 相关申请的交叉引用

[0002] 本申请要求 Gerald R. Carson 等人于 2009 年 10 月 30 日提交的美国临时专利申请系列号 61/256,544 的权益,该申请通过引用整体并入本文。

[0003] 关于联邦资助研究或开发的声明

[0004] 不适用

[0005] 对序列表、表格或计算机程序列表光盘附件的引用

[0006] 不适用(提供了序列表,但不是作为光盘附件)。

[0007] 背景

[0008] 在重组表达技术领域,希望的蛋白产物的高生产水平的实现和产生希望的纯度的产物的能力,代表正在发生的挑战。对于包括生物治疗剂(它们是抗体)在内的蛋白产物而言,这样的挑战是特别相关的,但是在该领域的进展还与其它生物制品有关。本发明的某些实施方案至少部分地解决了这些挑战的一个或多个方面。

发明内容

[0009] 下面的缩写是适用的:ORF,开放读码框;sORF,单个开放读码框;MW,分子量;HC 或 H,免疫球蛋白重链;LC 或 L,免疫球蛋白轻链;pab, *Pyrococcus abyssi*;pfu,激烈火球菌(*Pyrococcus furiosus*);pho,掘越氏火球菌(*Pyrococcus horikoshii*)OT3;aa 或 AA,氨基酸;SP,信号肽;LCSP,轻链信号肽;MTX,甲氨蝶呤。

[0010] 本发明的实施方案一般地涉及表达盒、载体构建体、重组宿主细胞以及重组多蛋白和前蛋白的重组表达和加工(包括翻译后加工)的方法。在一些实施方案中,一种或多种表达的产物是免疫球蛋白。

[0011] 在一些实施方案中,所述表达载体包含一个或多个内含肽区段。在一些实施方案中,所述内含肽区段源自生物体 *Pyrococcus abyssi*、激烈火球菌和掘越氏火球菌 OT3 的一种或多种 1on 内含肽。

[0012] 在一些实施方案中,关于某些元件的次序和存在与否,构造构建体的结构。在一个实施方案中,某些载体基因区段的次序是 HL,其中 H 和 L 分别指示免疫球蛋白重链和轻链。在另一个实施方案中,所述次序是 LH。在一个具体实施方案中,所述构建体具有标记为(-)的设计,其中减号指示,所述构建体具有在 ORF 开始处的一个信号肽和插入在内含肽的最后一个氨基酸和第二个蛋白亚基(例如,在内含肽之后的成熟抗体链)的第一个氨基酸之间的甲硫氨酸。在一个具体实施方案中,所述构建体具有标记为(+)的设计,其中加号指示,第一个信号肽在 ORF 开始处的存在和第二个信号肽在位于内含肽下游的第二个蛋白亚基开始处的存在。在一个具体实施方案中,所述构型是 HL(-)。

[0013] 在一些实施方案中,本发明提供了 sORF(单个开放读码框)构建体设计,当在来自在瞬时表达条件下的实验的培养物上清液中测量时,其能够生产大于 2、5、10、20、30、40 或 50 微克/ml 的分泌产物的蛋白表达水平。在一些实施方案中,本发明提供了 sORF 构建体,当在来自使用稳定的 CHO(中国仓鼠卵巢)细胞表达系统条件的实验的培养物上清液中测

量时,其能够生产大于 20 微克 /ml/ 天的蛋白表达水平。在一个实施方案中,所述表达水平 ($\mu\text{g/ml/天}$) 是在 1-24、大于 10 或大于 20 的范围内。在一个具体实施方案中,所述表达水平是 $24\mu\text{g/ml/天}$ 。在一些实施方案中,所述蛋白表达属于分泌型抗体,其已经自我装配成重链和轻链的多聚单元。在一些实施方案中,所述抗体属于 IgG 同种型。

[0014] 在一个实施方案中,本发明提供了用于产生一种或多种重组蛋白产物的分离的或纯化的表达载体,所述表达载体包含单个开放读码框插入物;所述插入物包含:

[0015] (a) 编码信号肽的信号肽核酸序列;

[0016] (b) 编码第一多肽的第一核酸序列;

[0017] (c) 编码第一蛋白切割位点的第一插入核酸序列,其中所述第一蛋白切割位点由火球菌属的 lon 蛋白酶基因的内含肽区段、或火球菌属或甲烷球菌属的 k1bA 基因的内含肽区段、或源自它们的修饰的内含肽区段提供;和

[0018] (d) 编码第二多肽的第二核酸序列;

[0019] 其中所述编码所述第一蛋白切割位点的第一插入核酸序列可操作地位于所述第一核酸序列和所述第二核酸序列之间;

[0020] 其中所述编码所述信号肽的信号肽核酸序列可操作地位于所述第一核酸序列之前;且

[0021] 其中所述表达载体能够表达单个开放读码框多肽,所述多肽可在所述第一蛋白切割位点处被切割。

[0022] 为了清楚,在包含不同的插入区段和方法的实施方案的背景下,编码蛋白切割位点的插入核酸序列可以是这样的:所述插入核酸序列至少编码第一蛋白切割位点。在规范的内含肽中,例如,切割反应通常以自动持续的且快速的方式进行。另一种解释部分地依赖于根本机理的理解。从观察外显肽组分的加工后角度看,可以理解,存在分别朝向内含肽区段的 N-端和 C-端的第一蛋白切割位点和第二蛋白切割位点。切割位点的命名无意必然地与切割反应可能发生的次序相对应,并认识到,可以认为切割反应是在一个切割反应位点处的单个且相对协调的事件,即使认识到给定机理中的动力学上不同的步骤。本说明书也提供了组合物和方法的实施方案,如本领域会理解的,它们具有包含一个或多个切割位点的插入区段。另外,根据加工机理的理解,包含 1 个切割位点或 2 个切割位点的区段可以各自允许插入区段的部分或完全切除。

[0023] 在一个实施方案中,插入核酸序列另外编码第二蛋白切割位点。

[0024] 在表达载体的一个实施方案中,所述第一蛋白切割位点由下述内含肽区段提供: *Pyrococcus abyssi*、激烈火球菌或掘越氏火球菌 OT3 的 lon 蛋白酶基因的内含肽区段;或 *Pyrococcus abyssi*、激烈火球菌或詹氏甲烷球菌 (*Methanococcus jannaschii*) 的 k1bA 基因的内含肽区段;或分别源自它们的修饰的内含肽区段。

[0025] 在一个实施方案中,所述内含肽区段或修饰的内含肽区段编码倒数第二位残基,所述残基是赖氨酸、丝氨酸或不是组氨酸。在一个实施方案中,所述内含肽区段或修饰的内含肽区段能够被切割,但是不能使所述第一多肽与所述第二多肽完全相连。

[0026] 在一个实施方案中,所述第一蛋白切割位点是由下述内含肽区段提供:包含选自 SEQ ID NO:1、3、4、6、7、55、35、37 和 39 的序列的内含肽区段,和源自它们的修饰的内含肽区段。

[0027] 在一个实施方案中,所述第一多肽和第二多肽能够多聚装配。在一个实施方案中,所述第一多肽和第二多肽中的至少一种能够胞外分泌。在一个实施方案中,所述第一多肽和第二多肽中的至少一种具有哺乳动物起源。在一个实施方案中,所述第一多肽包含免疫球蛋白重链或其功能片段,且所述第二多肽包含免疫球蛋白轻链或其功能片段,且所述第一多肽是在所述第二多肽的上游(5'侧)。

[0028] 在表达载体的一个实施方案中,所述载体仅包含一个信号肽核酸序列。

[0029] 在一个实施方案中,表达载体另外包含编码第三多肽的第三核酸序列和编码第二蛋白切割位点的第二插入核酸序列;其中所述第二插入核酸序列和第三核酸序列以该次序可操作地位于所述第二核酸序列之后。

[0030] 在表达载体的一个实施方案中,所述第一多肽和所述第二多肽包含功能抗体或其它抗原识别分子;其中抗原特异性指向结合选自下述的抗原:肿瘤坏死因子- α 、促红细胞生成素受体、RSV、EL/选择蛋白、白介素-1、白介素-12、白介素-13、白介素-17、白介素-18、白介素-23、白介素-33、CD81、CD19、IGF1、IGF2、EGFR、CXCL-13、GLP-1R、前列腺素E2和淀粉样蛋白 β 。

[0031] 在本发明的一个实施方案中,就表达载体而言,所述第一多肽和第二多肽包含一对免疫球蛋白链,它们来自D2E7、EL246、ABT-007、ABT-325或ABT-874的抗体。在一个实施方案中,所述第一多肽和第二多肽各自独立地选自:免疫球蛋白重链或免疫球蛋白轻链区段,它们来自D2E7、EL246、ABT-007、ABT-325、ABT-874或其它抗体的类似区段。

[0032] 在一个实施方案中,表达载体另外包含所述插入物的启动子调控元件。在一个实施方案中,所述启动子调控元件是诱导型或组成型。在一个实施方案中,所述启动子调控元件是组织特异性的。在一个实施方案中,所述启动子包含腺病毒主要晚期启动子。

[0033] 在一个实施方案中,本发明提供了一种宿主细胞,其包含本文所述的载体。在一个实施方案中,所述宿主细胞是原核细胞。在一个实施方案中,所述宿主细胞是大肠杆菌。在一个实施方案中,所述宿主细胞是真核细胞。在一个实施方案中,所述真核细胞选自:原生生物细胞、动物细胞、植物细胞和真菌细胞。在一个实施方案中,所述真核细胞是选自下述的动物细胞:哺乳动物细胞、禽细胞和昆虫细胞。在一个实施方案中,所述宿主细胞是哺乳动物细胞系。在一个实施方案中,所述宿主细胞是CHO细胞或二氢叶酸还原酶-缺陷型CHO细胞。在一个实施方案中,所述宿主细胞是HEK(人胚胎肾)细胞或非洲绿猴肾细胞,例如,COS细胞。在一个实施方案中,所述宿主细胞是酵母细胞。在一个实施方案中,所述酵母细胞是酿酒酵母。在一个实施方案中,所述宿主细胞是草地贪夜蛾(*Spodoptera frugiperda*) Sf9昆虫细胞。

[0034] 在一个实施方案中,本发明提供了一种用于生产重组多蛋白或多种蛋白的方法,所述方法包括:在足以允许载体蛋白表达的条件下,在培养基中培养宿主细胞。在一个实施方案中,所述方法另外包括:回收和/或纯化所述载体蛋白。在生产方法的一个实施方案中,所述多种蛋白能够多聚装配。在一个实施方案中,所述重组多蛋白或多种蛋白是生物学上有功能的和/或治疗性的。

[0035] 在一个实施方案中,本发明提供了一种用于生产重组产物的方法,其中所述产物是免疫球蛋白蛋白质或其功能片段、装配的抗体或其它抗原识别分子,所述方法包括:在足以生产所述重组产物的条件下,在培养基中培养宿主细胞。在一个实施方案中,本发明提供

了根据本文所述的方法生产的蛋白质或多蛋白。在实施方案中,本发明提供了根据本文的方法生产的装配的免疫球蛋白、装配的其它抗原识别分子、或单个免疫球蛋白链或其功能片段。在一个实施方案中,关于所述免疫球蛋白、其它抗原识别分子或单个免疫球蛋白链或其功能片段,存在实现或促进特定抗原(其中抗原可以是配体或反受体等)与下述物质的结合的能力:肿瘤坏死因子- α 、促红细胞生成素受体、RSV、EL/选择蛋白、白介素-1、白介素-12、白介素-13、白介素-17、白介素-18、白介素-23、白介素-33、CD81、CD19、IGF1、IGF2、EGFR、CXCL-13、GLP-1R、前列腺素 E2 或淀粉样蛋白 β 。在一个实施方案中,所述免疫球蛋白或其功能片段是免疫球蛋白 D2E7 或 ABT-874,或者所述功能片段是它们各自的片段。

[0036] 在一个实施方案中,本发明提供了一种药物组合物,其包含治疗有效量的蛋白和药学上可接受的载体。

[0037] 在一个实施方案中,本发明提供了如本文所述的表达载体,其另外包含编码标签的核酸序列。在载体构建体的一个实施方案中,所述插入核酸序列另外编码标签。

[0038] 在一个实施方案中,所述第一多肽和所述第二多肽包含功能抗体或其它抗原识别分子;其中抗原特异性指向结合选自下述的抗原:肿瘤坏死因子- α 、促红细胞生成素受体、RSV、EL/选择蛋白、白介素-1、白介素-12、白介素-13、白介素-17、白介素-18、白介素-23、白介素-33、CD81、CD19、IGF1、IGF2、EGFR、CXCL-13、GLP-1R、前列腺素 E2 和淀粉样蛋白 β 。在一个实施方案中,所述第一多肽和第二多肽包含一对免疫球蛋白链,它们来自 D2E7、EL246、ABT-007、ABT-325 或 ABT-874 的抗体。在一个实施方案中,所述第一多肽和第二多肽各自独立地选自:免疫球蛋白重链或免疫球蛋白轻链区段,它们来自 D2E7、EL246、ABT-007、ABT-325、ABT-874 或其它抗体的类似区段。

[0039] 在一个实施方案中,载体另外包含所述 sORF 插入物的启动子调控元件。在一个实施方案中,所述启动子调控元件是诱导型或组成型。在一个实施方案中,所述启动子调控元件是组织特异性的。在一个实施方案中,所述启动子包含腺病毒主要晚期启动子。

[0040] 在一个实施方案中,载体另外包含编码蛋白酶的核酸,所述蛋白酶能够切割所述第一蛋白切割位点。在一个实施方案中,所述编码蛋白酶的核酸可操作地位于所述 sORF 插入物内;所述表达载体另外包含编码第二切割位点的额外核酸,其位于所述编码蛋白酶的核酸与所述第一核酸和所述第二核酸中的至少一种之间。

[0041] 在一个实施方案中,本发明提供了一种宿主细胞,其包含本文所述的载体。在一个实施方案中,所述宿主细胞是原核细胞。在一个实施方案中,所述宿主细胞是大肠杆菌。在一个实施方案中,所述宿主细胞是真核细胞。在一个实施方案中,所述真核细胞选自:原生生物细胞、动物细胞、植物细胞和真菌细胞。在一个实施方案中,所述真核细胞是选自下述的动物细胞:哺乳动物细胞、禽细胞和昆虫细胞。在一个优选的实施方案中,所述宿主细胞是 CHO 细胞或二氢叶酸还原酶-缺陷型 CHO 细胞。在一个实施方案中,所述宿主细胞是 COS 细胞。在一个实施方案中,所述宿主细胞是酵母细胞。在一个实施方案中,所述酵母细胞是酿酒酵母。在一个实施方案中,所述宿主细胞是昆虫草地贪夜蛾 Sf9 细胞。在一个实施方案中,所述宿主细胞是人胚胎肾细胞。

[0042] 在一个实施方案中,本发明提供了一种用于生产重组多蛋白或多种蛋白的方法,所述方法包括:在足以允许载体蛋白表达的条件下,在培养基中培养宿主细胞。在一个实施方案中,所述方法另外包括:回收和/或纯化所述载体蛋白。在一个实施方案中,所述多种

蛋白能够多聚装配。在一个实施方案中,所述重组多蛋白或多种蛋白是生物学上有功能的和/或治疗性的。

[0043] 在一个实施方案中,本发明提供了一种用于生产免疫球蛋白蛋白质或其功能片段、装配的抗体或其它抗原识别分子的方法,所述方法包括:在足以生产免疫球蛋白蛋白质或其功能片段、装配的抗体或其它抗原识别分子条件下,在培养基中培养根据权利要求38所述的宿主细胞。

[0044] 在一个实施方案中,本发明提供了根据本文的方法生产的蛋白质或多蛋白。在一个实施方案中,本发明提供了根据本文的方法生产的装配的免疫球蛋白、装配的其它抗原识别分子、或单个免疫球蛋白链或其功能片段。在一个实施方案中,所述免疫球蛋白、其它抗原识别分子或单个免疫球蛋白链或其功能片段具有实现或促进特定抗原与下述物质的结合的能力:肿瘤坏死因子- α 、促红细胞生成素受体、白介素-18、EL/选择蛋白或白介素-12。在一个实施方案中,所述免疫球蛋白是D2E7,或者其中所述功能片段是D2E7的片段。

[0045] 在一个实施方案中,本发明提供了一种表达载体、具有所述载体的宿主细胞、载体表达产物、药物组合物和/或制备或使用前述任一种的方法,其中所述载体是根据权利要求1-9中任一项所述的载体,且另外包含编码轻链信号肽的区段。在一个实施方案中,所述编码的轻链信号肽是选自下述的 κ 轻链信号肽:A17、A18、A19、A26和H2G。在一个实施方案中,所述编码的轻链信号肽是VKII κ 轻链信号肽A18, SEQ ID NO:82(氨基酸序列MRLPAQLLGLLMLWIPGSSA)。

[0046] 在一个实施方案中,本发明的组合物是分离的或纯化的。

[0047] 在一个实施方案中,本发明的组合物是肽化合物。在一个实施方案中,本发明的组合物是核酸化合物。在一个实施方案中,本发明的肽化合物被装配在具有所述肽或至少一种其它肽的多聚复合物中。

[0048] 在一个实施方案中,本发明提供了一种药物制剂,其包含本发明的组合物。在一个实施方案中,本发明提供了一种合成本发明的组合物或其药物制剂的方法。在一个实施方案中,药物制剂包含本领域会理解的一种或多种赋形剂、载体和/或其它组分。在一个实施方案中,本发明的组合物的有效量可以是治疗有效量。

[0049] 在一个实施方案中,使用重组方法学或合成技术,制备本发明的肽组合物。在一个实施方案中,使用重组方法学或合成技术,制备本发明的核酸组合物。

[0050] 在一些实施方案中,本发明提供了用于药剂生产中的方法。

[0051] 在技术领域的背景下,结合附图,从下面的描述中会明白本发明的实施方案的其它方面、特征和优点。

[0052] 一般而言,在本文中使用的术语和短语具有它们的本领域公认的含义,这些含义可以通过参考本领域技术人员已知的标准教科书、期刊参考文献和背景来找到。在本文中提供的定义意图澄清它们在本发明的实施方案的背景下的具体用途。

[0053] 不希望受任何具体理论的约束,在本文中可以讨论与本发明有关的根本原理或机理的信念或理解。应当认识到,不论任何解释或假设的最终正确性如何,但是本发明的实施方案可以是可操作的和有用的。

附图说明

[0054] 图 1 解释了 sORF 表达构建体 pTT3pab lon HL(-) 的示意图。

[0055] 图 2 解释了在 D2E7 抗体的表达构建体中的 sORF 组分结构。

[0056] 图 3 解释了 sORF 表达产物的蛋白分析的 SDS-PAGE 结果。通过蛋白 A 亲和色谱法纯化分泌的 IgG 分子,并在非还原 (A) 和还原 (B) 条件下通过 SDS-PAGE 进行分离。泳道和样品从左向右为:(泳道 1) 分子量参照标志物;(2) 对照构建体产物;(3)Pab-lon mut A1;(4)Pab-lon mut A2;(5)pTT3 pfu lon YP,和 (6)pTT3 pfu lon MA。

[0057] 图 4 解释了其它 sORF 表达产物的蛋白分析的 SDS-PAGE 结果。通过蛋白 A 亲和色谱法纯化分泌的 IgG,并在非还原 (A) 和还原 (B) 条件下通过 SDS-PAGE 进行分离。泳道和样品从左向右为:(泳道 1) 分子量标志物;(2) 对照;(3)pTT3 pfu lon HL(-);和 (4)pTT3pfu lon MutA。

[0058] 图 5 解释了使用 k1bA 内含肽从 sORF 构建生产的分泌型抗体的分析。通过蛋白 A 亲和色谱法纯化从 Pab-k1bA HL(-) 和 Mja-k1bA HL(-) 构建体分泌的 IgG 产物,并在还原 (图 A、B 和 C) 和非还原 (图 D) 条件下通过 SDS-PAGE 进行分离。图 A 和 D 显示了染色的凝胶的图像;图 B 是使用针对人 IgG1 Fc 的抗体的免疫印迹;图 C 是使用针对人 κ 轻链的抗体的免疫印迹。泳道和样品从左向右为:(泳道 1) 对照;(2)Pab-k1bA HL(-);和 (3) Mja-k1bA HL(-)。所述对照是从 2 个分开的开放读码框的表达生成的相同抗体。

[0059] 图 6 解释了使用 Pab k1bA 内含肽 (其具有对在 N-端剪接点处的氨基酸残基的修饰) 的单个开放读码框构建体的表达结果。通过蛋白 A 亲和色谱法纯化分泌的 IgG 蛋白,并在非还原和还原条件下通过 SDS-PAGE 进行分离。泳道和样品从左向右为:(泳道 1) 分子量标志物;(2) 使用常规载体 (对照) 生产的相同抗体;(3)pTT3 Pab k1ba HL(-)wt;(4) pTT3 Pab k1ba HL(-)GC;和 (5)pTT3 Pab k1ba HL(-)KC。

[0060] 图 7 解释了 sORF 表达构建体 pA190-Pab-lon HL(-) 的示意图,所述构建体适合用作在 CHO 细胞系统中的稳定表达载体。

[0061] 图 8 解释了 sORF 构建体转染克隆 (sORF Pab lon 构建体) 的稳定表达系统达到培养物汇合的时间和频率的结果。

[0062] 图 9 解释了具有轻链接头突变的瞬时表达构建体的 sORF 组分结构示意图。一系列构建体被命名为“M1-X”(第 1 行)和 D2-X”(第 2 行),其中 X 指示任意氨基酸。

[0063] 图 10 解释了具有轻链接头突变 (基于 Met1 残基的变化) 的一系列 sORF 构建体的 IgG 分泌结果。

[0064] 图 11 解释了具有轻链接头突变 (基于 Asp2 残基的变化) 的一系列 sORF 构建体的 IgG 分泌结果。

[0065] 图 12 解释了蛋白产物的 SDS-PAGE 分析结果,所述蛋白产物来自具有轻链接头突变的 Met1 和 Asp2 系列构建体中的每一个的实施例。

[0066] 图 13 解释了能够表达 ABT-874 抗体的瞬时表达构建体的 sORF 组分结构示意图。

[0067] 图 14 解释了在用于导入插入物 (包括标签) 的溶剂可接近的环的优选位置 (虚线箭头指示区段 H 和 F 的接头附近,朝向 DOD 内切核酸酶结构域的末端) 的位置的背景下,内含肽的某些结构基序。

[0068] 图 15 解释了具有轻链信号肽 A18 的表达构建体的质粒图谱,所述构建体用于 HEK293 细胞的瞬时转染系统中。

[0069] 图 16 解释了抗体表达构建体的产物的 SDS-PAGE 分析的结果。

[0070] 图 17 解释了来自转染的细胞系(包括瞬时转染的 HEK293 细胞和稳定转染的 CHO 细胞)的抗体表达构建体的产物的蛋白印迹分析的结果。

[0071] 图 18 解释了具有轻链信号肽 A18 的表达构建体的质粒图谱,所述构建体用于 CHO 细胞的稳定转染系统中。

具体实施方式

[0072] 通过下述的非限制性实施例,可以进一步理解本发明。

[0073] 本领域的公开内容(包括根据 Carson 等人于 2007 年 3 月 22 日提交的 US 20070065912 的公开内容)会理解某些信息。

[0074] 本发明提供了系统,例如,构建体和方法,其用于表达化合物结构或生物活性蛋白诸如酶、激素(例如,胰岛素)、细胞因子、趋化因子、受体、抗体或其它分子。优选地,所述蛋白是免疫调节蛋白诸如白介素、全长免疫球蛋白、其片段、本领域理解的其它抗原识别分子或其它生物治疗分子。这样的系统的概述是在免疫球蛋白分子的具体背景下,其中重组生产是基于在单个启动子的转录控制下的重链和轻链编码序列的表达,其中单一翻译产物(多蛋白)向分离的重链和轻链的转化是由内含肽组分介导。

[0075] 在一个实施方案中,免疫球蛋白多蛋白分子的第一链或第二链可以是重链或轻链。编码重组免疫球蛋白区段的序列可以是全长编码序列或其片段。在一个具体的实施方案中,第二轻链编码序列必须是编码要在本发明的实践中加工的多蛋白的序列的一部分;即,总共存在 3 个区段,包括任意次序的 2 个轻链和 1 个重链。在具体实施方案中,用这些组分以下述次序构建成构建体:a) IgH-IgL;b) IgL-IgH;c) IgH-IgL-IgL;d) IgL-IgH-IgL;e) IgL-IgL-IgH;f) IgH-IgH-IgL;g) IgH-IgL-IgH;和/或 h) IgL-IgH-IgH。在一个实施方案中,所述连字符可以指示切割位点序列所在的位置。

[0076] 或者,免疫球蛋白重链和轻链编码序列与位于它们之间的内含肽编码序列在框架内融合,其中所述内含肽天然地能够或被修饰成缺少剪接活性,或者重链和轻链的末端被设计成,使得剪接优选地不会发生,或使得剪接以低效率发生,因而未剪接的抗体分子占优势。另外,修饰的内含肽可以进一步被修饰,进一步使得,不存在内切核酸酶区域(在内切核酸酶区域以前已经存在的情况下),条件是,保留位点特异性的蛋白水解性裂解活性,使得轻和重抗体多肽从初级翻译产物的插入内含肽部分中释放出。轻或重抗体多肽可以是 N- 外显肽,且任一个可以是 C- 外显肽。

[0077] 所述载体可以是能够表达全长多蛋白的任意重组载体,例如,腺伴随病毒(AAV)载体、慢病毒载体、逆转录病毒载体、复制活性的腺病毒载体、复制缺陷的腺病毒载体和无肠(gutless)腺病毒载体、疱疹病毒载体或非病毒载体(质粒)或本领域已知的任意其它载体,选择适合在其中表达免疫球蛋白或其它蛋白的宿主细胞的载体。杆状病毒载体可用于在昆虫细胞中表达基因。众多载体是本领域已知的,且许多是可商业得到的,或在本领域可以其它方式容易地得到。

[0078] 包括启动子在内的调节序列;宿主细胞

[0079] 用于重组免疫球蛋白或其它蛋白表达的载体可以包括本领域已知的许多启动子中的任一种,其中所述启动子是组成性的、可调节的或可诱导的、细胞类型特异性的、组织特异性的或物种特异性的。其它具体的实例包括,例如,四环素响应性的启动子(Gossen M, Bujard H, Proc Natl Acad Sci U S A. 1992, 15 ;89(12) :5547-51)。所述载体是适合要在其中表达嵌合基因的宿主细胞的复制子,且它合乎需要地还包括也在细菌细胞(有利地,大肠杆菌,即对于分子生物学操作而言方便的细胞)中有功能的复制子。

[0080] 用于基因表达的宿主细胞可以是,但不限于,动物细胞,特别是哺乳动物细胞,或它可以是微生物细胞(细菌、酵母、真菌,但是优选真核生物)或植物细胞。特别合适的宿主细胞包括:昆虫培养的细胞诸如草地贪夜蛾细胞、酵母细胞诸如酿酒酵母或巴氏毕赤酵母(*Pichia pastoris*)、真菌诸如里氏木霉(*Trichoderma reesei*)、曲霉属、短梗霉属(*Aureobasidium*)和青霉属(*Penicillium*)种以及哺乳动物细胞诸如CHO(中国仓鼠卵巢)、BHK(幼仓鼠肾)、COS、293、3T3(小鼠)、Vero(非洲绿猴)细胞,也可以使用不同的转基因动物系统,包括、但不限于、猪、小鼠、大鼠、绵羊、山羊、母牛。已知鸡系统用于在卵白中表达,转基因的绵羊、山羊和母牛系统用于在乳汁中表达,以及其它。杆状病毒(特别是AcNPV)载体可以用于本发明的单个ORF抗体表达和切割,例如在多角体蛋白启动子或其它强启动子的调节控制下在昆虫细胞系中表达sORF;这样的载体和细胞系是本领域众所周知的和可商业得到的。用于哺乳动物细胞中的启动子可以是组成型(疱疹病毒TK启动子,McKnight, Cell 31 :355, 1982;SV40早期启动子, Benoist 等人 Nature 290 :304, 1981; 鲁斯肉瘤病毒启动子, Gorman 等人 Proc. Natl. Acad. Sci. USA 79 :6777, 1982; 巨细胞病毒启动子, Foecking 等人 Gene 45 :101, 1980; 小鼠乳腺肿瘤病毒启动子,一般地参见 Etcheverry in Protein Engineering: Principles and Practice, Cleland 等人编,第162-181页, Wiley & Sons, 1996)或受调节的(例如金属硫蛋白启动子, Hamer 等人 J. Molec. Appl. Genet. 1 :273, 1982)。载体可以基于感染特定哺乳动物细胞的病毒,特别是逆转录病毒、牛痘和腺病毒,它们的衍生物是本领域已知的和可商业得到的。启动子包括、但不限于:巨细胞病毒启动子、腺病毒晚期启动子和牛痘 7.5K 启动子。酵母和真菌载体(参见,例如, Van den Handel, C. 等人(1991)见: Bennett, J. W. 和 Lasure, L. L. (编), More Gene Manipulations in Fungi, Academy Press, Inc., New York, 397-428)和启动子也是众所周知的和可广泛得到的。烯醇化酶是一种众所周知的组成型酵母启动子,醇脱氢酶是一种众所周知的受调节的启动子。

[0081] 具体的启动子、转录终止序列和其它任选的序列(诸如编码组织特异性序列的序列)的选择,在很大程度上由希望在其中进行表达的细胞的类型决定。所述细胞可以是细菌、酵母、真菌、哺乳动物、昆虫、鸡或其它动物细胞。

[0082] 信号序列

[0083] 要切割、蛋白水解地加工或自加工的蛋白的编码序列(其掺入在载体中)可以另外包含编码一个或多个信号序列的一个或多个序列。这些编码的信号序列可以与多蛋白内的一个或多个成熟的区段相结合。例如,编码免疫球蛋白重链前导序列的序列可以在重链的编码序列之前,与多蛋白编码序列的剩余部分可操作地连接且在框架内。类似地,轻链前导序列肽编码序列或其它前导序列肽编码序列可以与免疫球蛋白轻链编码序列之一或二者在框架内结合,其中前导序列-链与自加工位点(诸如2A)被邻近链隔开,或被编码蛋白

酶识别序列的序列隔开,以维持适当的读码框。

[0084] 免疫球蛋白重链和轻链的化学计量学

[0085] 在本文的许多实施方案中,免疫球蛋白 / 抗体轻链 (IgL) 和重链 (IgH) 在载体水平或在表达的细胞内水平以约 1 : 1 比例 (IgL : IgH) 存在于宿主细胞内。然而,在本文中和别处的重组方案已经依赖于重链和轻链的等摩尔表达 (参见,例如,美国专利公开 2005/0003482A1 或国际公开 W02004/113493),在其它实施方案中,本发明提供了方法和表达盒和载体,其具有 2 : 1 比例的轻链和重链编码序列,并在初级翻译产物是多蛋白时,与所述链的自加工或蛋白水解性加工共表达。在一些实施方案中,所述比例大于 1 : 1,诸如约 2 : 1 或大于 2 : 1。在一个具体实施方案中,以大于 1 : 1 (IgL : IgH) 的比例使用轻链编码序列。在一个具体的实施方案中, IgL : IgH 的比例是 2 : 1。因而,在一些实施方案中, sORF 抗体表达技术提供的优点包括:操纵重链和轻链的基因剂量比例的能力、用于内质网 (ER) 中的多亚基装配的重链和轻链多肽的接近和高效率蛋白分泌的潜力。

[0086] 本发明另外提供了用载体转化或感染的宿主细胞或稳定的宿主细胞克隆,所述载体包含:编码免疫球蛋白 (即,抗体) 的重链和 1 个或至少 2 个轻链的序列;编码切割位点 (诸如自加工位点、蛋白酶识别位点或在它们之间的信号肽) 的序列;且可能另外包含一个或多个编码额外的蛋白水解性切割位点的序列。在本发明的范围内还包括这样的细胞或克隆在制备全长重组免疫球蛋白或其片段或包含多个亚基的其它生物活性蛋白 (例如,双链或多链分子,或在自然界中生成成为前蛋白并经切割或加工以释放出前体衍生的蛋白和活性部分的那些) 中的用途。非限制性实例包括:胰岛素、白介素 -18、白介素 -1、骨形态形成蛋白 4、骨形态形成蛋白 2、任意其它双链骨形态形成蛋白、神经生长因子、肾素、胰凝乳蛋白酶、转化生长因子 β 和白介素 1 β 。

[0087] 在一个有关的方面,本发明提供了一种重组免疫球蛋白分子或其片段或由这样的细胞或克隆生产的其它蛋白,其中所述免疫球蛋白包含源自自加工切割位点的氨基酸 (诸如内含肽或刺猬结构域),切割位点或信号肽切割和方法,用于生产它们的载体和宿主细胞。在一些实施方案中,本发明提供了含有本文所述的一种或多种构建体的宿主细胞。

[0088] 本发明提供了用于表达免疫球蛋白分子或其片段的单个载体构建体,以及将它们用于体外或体内用途的方法。所述载体具有自加工或其它蛋白酶识别序列,该序列在第一和第二免疫球蛋白编码序列之间以及在第二和第三免疫球蛋白编码序列之间,从而允许使用单个启动子和转录物来表达功能抗体分子。示例性的载体构建体包含编码在开放读码框之间的自加工切割位点的序列,且可能另外包含与自加工切割位点邻近的额外的蛋白水解性切割位点,用于在切割以后去除包含自加工切割位点的氨基酸。所述载体构建体可用于与全长生物活性的免疫球蛋白或其片段的增强的体外和体内生产有关的方法中。使用相同的策略,可以制备具有至少 2 个不同链的其它生物活性蛋白,尽管应当理解,可能不要求任一条链的编码序列相对于其它链的编码序列以大于 1 的比例存在。

[0089] 尽管在本文中例证了具体的组合物和方法,应当理解,许多替代性组合物和方法中的任一种是可适用的,且适合用于实践本发明。还应当理解,使用本领域的标准规程,可以评价本发明的多蛋白表达盒和载体、宿主细胞和方法。除非另有说明,本发明的实践将采用细胞生物学、分子生物学 (包括重组技术)、微生物学、生物化学和免疫学的常规技术,它们是在本领域技术人员的范围内。这样的技术在文献中得到了充分解释,所述

文献例如, Molecular Cloning :A Laboratory Manual, 第 2 版 (Sambrook 等人, 1989) ; Oligonucleotide Synthesis (M. J. Gait 编, 1984) ; Animal Cell Culture (R. I. Freshney 编, 1987) ; Methods in Enzymology (Academic Press, Inc.) ; Handbook of Experimental Immunology (D. M. Weir & C. C. Blackwell 编) ; Vectors for Mammalian Cells (J. M. Miller & M. P. Calos 编, 1987) ; Current Protocols in Molecular Biology (F. M. Ausubel 等人编, 1993) ; PCR: The Polymerase Chain Reaction, (Mullis 等人编, 1994) ; 和 Current Protocols in Immunology (J. E. Coligan 等人编, 1991), 它们中的每一篇特此通过引用并入本文中。

[0090] 除非另有说明, 在本文中使用的术语具有与本领域技术人员通常理解相同的含义, 本发明的实践将采用微生物学和重组 DNA 技术的常规技术, 它们是在本领域技术人员的知识内。

[0091] 在蛋白的背景下, 在本文中一般地使用的术语“修饰的”表示这样的区段: 其中在提及的分子中置换、删除或添加了至少一个氨基酸残基。类似地, 在核酸的背景下, 该术语表示这样的区段: 其中在提及的分子中置换、删除或添加了至少一个核酸亚基。

[0092] 本文使用的术语“内含肽”通常表示蛋白的内部区段, 该内部区段会促进它自身的去除, 并影响称作外显肽的侧接区段的连接。在多种类型的生物体中识别出了内含肽的许多实例, 在有些情况下, 它们具有共同的结构和 / 或功能特征。本发明广泛地能够采用内含肽及其变体, 只要认为存在, 并进一步被识别或发现。参见, 例如, Gogarten JP 等人, 2002, Annu Rev Microbiol. 2002 ; 56 : 263-87 ; Perler, F. B. (2002), InBase, the Intein Database. Nucleic Acids Res. 30, 383-384 (也通过 New England Biolabs, Inc., Ipswich, MA 的因特网网站 ; <http://www.neb.com/neb/inteins.html> ; Amitai G 等人, Mol Microbiol. 2003, 47 (1) : 61-73 ; Gorbalenya AE, Nucleic Acids Res. 1998 ; 26 (7) : 1741-1748. Non-canonical inteins)。在蛋白中, 含有内含肽的单元或内含肽剪接单元可以理解为包括侧接外显肽的部分, 其中结构方面可以促进切割、连接等反应。该术语也可以理解为提及含有“修饰的内含肽”组分的基于内含肽的系统的范畴。

[0093] 本文使用的术语“修饰的内含肽”可以表示合成的内含肽或天然的内含肽, 其中在内含肽剪接单元中置换、删除或添加了至少一个氨基酸残基, 使得切割的或切离的外显肽不与所述内含肽完全相连。

[0094] 本文使用的术语“载体”表示 DNA 或 RNA 分子, 诸如质粒、病毒或其它媒介物, 其含有一个或多个异源的或重组的 DNA 序列, 且被设计成用于在不同的宿主细胞之间转移。术语“表达载体”和“基因治疗载体”表示, 可以有效地在细胞中掺入和表达异源 DNA 片段的任意载体。克隆或表达载体可以包含额外的元件, 例如, 表达载体可以具有 2 个复制系统, 因而, 允许它维持在 2 种生物体中, 例如在人细胞中进行表达和在原核宿主中进行克隆和扩增。可以采用可有效地将核酸导入细胞中使得蛋白或多肽表达发生的任意合适的载体, 例如病毒载体或非病毒质粒载体。对于表达而言有效的任意细胞 (例如, 昆虫细胞和真核细胞诸如酵母或哺乳动物细胞) 可用于实践本发明。

[0095] 术语“异源 DNA”和“异源 RNA”表示这样的核苷酸: 其对于细胞而言不是内源的 (天然的), 或不是它们存在于其中的基因组或载体的一部分。通常, 通过转导、感染、转染、转化、电穿孔、基因枪转化等, 将异源 DNA 或 RNA 添加给细胞。这样的核苷酸通常包括至少

一个编码序列,但是所述编码序列不需要表达。术语“异源 DNA”可以表示“异源编码序列”或“转基因”。

[0096] 本文使用的术语“蛋白”和“多肽”可以可互换地使用,并通常表示使用本发明的含有自加工切割位点的载体表达的目标“蛋白”和“多肽”。这样的“蛋白”和“多肽”可以是可用于下文进一步描述的研究、诊断或治疗目的的任意蛋白或多肽。本文使用的多蛋白是这样的蛋白:其用于加工,以生成 2 种或更多种多肽产物。

[0097] 本文使用的术语“多聚体”表示由 2 个或更多个多肽链(有时称作“亚基”)组成的蛋白,所述多肽链装配形成功能蛋白。多聚体可以由 2 个(二聚体)、3 个(三聚体)、4 个(四聚体)或更多个(例如,五聚体,诸如此类)肽链组成。多聚体可能源自自装配,或可能需要诸如催化剂等组分来辅助装配。多聚体可以仅由相同的肽链组成(同型多聚体),或由 2 个或更多个不同的肽链组成(异型多聚体)。这样的多聚体可以具有结构或化学功能。许多多聚体是本领域已知的和使用的,包括、但不限于酶、激素、抗体、细胞因子、趋化因子和受体。这样,多聚体可以具有生物学(例如,药学)和工业(例如,生物加工/生物生产)用途。

[0098] 本文使用的术语“标签”表示肽,其可以掺入表达载体中,其可能起允许检测和/或纯化载体插入物的一种或多种表达产物的功能。这样的标签是本领域众所周知,且可以包括放射性标记的氨基酸或与生物素基部分的多肽的连接,所述多肽可以通过标记的抗生物素蛋白(例如,抗生物素链菌素,其含有可通过光学或比色测量方法检测到的荧光标志物或酶活性)来检测。亲和标签(诸如 FLAG、谷胱甘肽-S-转移酶、麦芽糖结合蛋白、纤维素-结合域、硫氧还蛋白、NusA、mistin、甲壳质-结合域、角质酶、AGT、GFP 和其它亲和标签)被广泛地用于例如蛋白表达和纯化系统中。多肽标签的其它非限制性实例包括、但不限于下述的:组氨酸标签、放射性同位素或放射性核素(例如, ^3H 、 ^{14}C 、 ^{35}S 、 ^{90}Y 、 ^{99}Tc 、 ^{111}In 、 ^{125}I 、 ^{131}I 、 ^{177}Lu 、 ^{166}Ho 或 ^{153}Sm);荧光标签(例如, FITC、罗丹明、镧系元素、磷光体)、酶标签(例如,辣根过氧化物酶、萤光素酶、碱性磷酸酶);化学发光的标签;生物素基基团;可被第二报道物识别的预定多肽表位(例如,亮氨酸拉链对序列、第二抗体的结合位点、金属结合域、表位标签);和磁性剂诸如钕螯合物。

[0099] 本文关于本发明的病毒基因治疗载体所用的术语“复制缺陷的”是指,所述病毒载体不能独立地进一步复制和包装它的基因组。例如,当用 rAAV 病毒粒子感染受试者的细胞时,异源基因在感染的细胞中表达,但是,由于感染的细胞缺少 AAV rep 和 cap 基因及附加功能基因的事实, rAAV 不能复制。

[0100] 本文使用的“逆转录病毒转移载体”是指这样的表达载体:其包含编码转基因的核苷酸序列,且另外包含包装载体必需的核苷酸序列。优选地,逆转录病毒转移载体还包含在细胞中表达转基因必需的序列。

[0101] 本文使用的“包装系统”是指一系列病毒构建体,其包含编码参与包装重组病毒的病毒蛋白质的基因。通常,包装系统的构建体最后掺入包装细胞中。

[0102] 本文使用的“第二代”慢病毒载体系统是指缺乏功能性附加基因的慢病毒包装系统,诸如其中附加基因 vif、vpr、vpu 和 nef 已经缺失或者灭活的慢病毒包装系统。参见,例如, Zufferey 等人 1997. Nat. Biotechnol. 15:871-875。

[0103] 本文使用的“第三代”慢病毒载体系统是指这样的慢病毒包装系统:其具有第二

代载体系统特征,并进一步缺乏功能性 tat 基因,诸如其中 tat 基因已经缺失或者灭活的慢病毒包装系统。通常,编码 rev 的基因提供在单独的表达构建体上。参见,例如,Du11 等人 1998. J. Virol. 72 :8463-8471。

[0104] 本文关于病毒或病毒载体使用的“假型 (pseudotyped)”是指,用异源的或者功能修饰的病毒包膜蛋白替换天然的病毒包膜蛋白。

[0105] 本文关于重组 DNA 构建体或载体使用的术语“可操作地连接的”是指,重组 DNA 构建体或载体的核苷酸组分通常彼此共价地连接。通常,“可操作地连接的”DNA 序列是邻接的,且在分泌型前导序列的情况下,是邻接的且在相同读码框中。但是,增强子不一定与被增量调节其表达的序列邻接。该术语与“可操作地放置的”相一致。

[0106] 增强子序列会影响启动子依赖性的基因表达,且可能位于天然基因的 5' 或 3' 区域。“增强子”是顺式作用元件,其刺激或抑制邻近基因的转录。抑制转录的增强子也称作“沉默子”。增强子可以在离编码序列和被转录区域的下游位置多达几千碱基对 (kb) 的距离处在任一个方向起作用 (即,可以与编码序列相关联)。另外,隔离物或染色质打开序列,诸如基质附着区域 (Chung, Cell, 1993, Aug 13 ;74 (3) :505-14, Frisch 等人, Genome Research, 2001, 12 :349-354, Kim 等人, J. Biotech107, 2004, 95-105), 可以用于增强稳定整合的基因盒的转录。

[0107] 本文使用的术语“基因”或“编码序列”是指,当与适当的调节序列可操作地连接时,在体外或在体内转录 (DNA) 和翻译 (mRNA) 成多肽的核酸序列。所述基因可以包括或不包括在编码区之前和之后的区域,例如 5' 不翻译序列 (5' UTR) 或“先导”序列和 3' UTR 序列或“尾巴”序列以及在单个编码区段 (外显子) 之间的插入序列 (内含子)。

[0108] “启动子”是指导 RNA 聚合酶的结合并从而促进 RNA 合成的 DNA 序列,即,足以指导转录的最小序列。启动子和对应的蛋白或多肽表达可以是细胞类型特异性的、组织特异性的或物种特异性的。在本发明的核酸构建体或载体中也包括增强子序列,其可以与或不与启动子序列邻接。

[0109] 在本文中广泛使用的“转录调节序列”或表达控制序列包括启动子序列和物理地有关的序列,其调控或调节有关的编码序列的转录,经常是响应于营养或环境信号。那些有关的序列可以决定组织或细胞特异性的表达、对环境信号的应答、增加或减少转录的蛋白的结合等。“可调节的启动子”是其活性受顺式或反式作用因子影响的任意启动子 (例如,被外部信号或试剂活化的可诱导的启动子)。

[0110] “组成型启动子”是在大多数情况下指导许多或所有组织 / 细胞类型中的 RNA 生产的任意启动子,例如,人 CMV 立即早期增强子 / 启动子区域,其促进克隆的 DNA 插入物在哺乳动物细胞中的组成性表达。

[0111] 术语“转录调节蛋白”、“转录调节因子”和“转录因子”在本文中可互换地使用,并表示这样的核蛋白:其结合 DNA 应答元件,并从而转录地调节有关的一个或多个基因的表达。转录调节蛋白通常直接地结合 DNA 应答元件,但是在有些情况下,与 DNA 的结合可以是间接的,这通过与其它蛋白的结合来实现,所述其它蛋白又结合或被结合在 DNA 应答元件上。

[0112] 本文使用的术语“免疫球蛋白”和“抗体”表示完整的分子及其片段,诸如 Fa、F(ab')₂ 和 Fv,它们能够结合目标抗原决定簇。这样的“免疫球蛋白”和“抗体”由 2 个相

同的分子量为约 23,000 道尔顿的多肽轻链和 2 个相同的分子量为 53,000-70,000 道尔顿的重链组成。所述 4 个链通过二硫键连接成“Y”构型。重链被分类为 γ (IgG)、 μ (IgM)、 α (IgA)、 δ (IgD) 或 ϵ (IgE), 且是免疫球蛋白的分类命名的基础, 这决定了给定抗体的效应子功能。轻链被分类为 κ 或 λ 。当在本文中提及“免疫球蛋白或其片段”时, 应该理解, 这样的“其片段”是免疫学上有功能的免疫球蛋白片段, 特别是以完整免疫球蛋白的至少 10% 的结合亲和力结合它的同源配体的片段。

[0113] 抗体的 Fab 片段是抗体分子的单价抗原结合片段。Fv 片段是遗传工程化的片段, 其含有表达为 2 个链的轻链可变区和重链可变区。

[0114] 术语“人源化的抗体”表示这样的抗体分子: 其中已经替换非抗原结合区中的一个或多个氨基酸, 以便更接近地模仿人抗体, 同时仍然保留所述抗体的原始结合活性。参见, 例如, 美国专利号 6,602,503。

[0115] 本文使用的术语“抗原决定簇”表示, 与特定抗体接触的分子片段 (即, 表位)。蛋白或肽或蛋白糖肽或糖蛋白的众多区域可以诱导抗体的生成, 所述抗体特异性地结合所述蛋白上的给定区域或三维结构。这些区域或结构称作抗原决定簇或表位。抗原决定簇可以与完整抗原 (即, 用于引起免疫应答的免疫原) 竞争与抗体的结合。

[0116] 当提及本发明的重组蛋白或多肽时, 术语“片段”是指这样的肽或多肽: 它的氨基酸序列与对应的全长蛋白或多肽的氨基酸序列的一部分 (但不是全部) 相同, 它保留对应的全长蛋白或多肽的至少一种功能或活性。所述片段优选地包括全长蛋白或多肽的至少 20-100 个连续氨基酸残基。

[0117] 本文使用的术语“施用”或“导入”是指, 通过本领域已知的任意途径, 将蛋白 (包括免疫球蛋白) 递送给有此需要的人或动物。药用载体和制剂或组合物也是本领域众所周知的。给药途径可以包括: 静脉内的、肌肉内的、真皮内的、皮下的、透皮的、粘膜的、瘤内的或粘膜的。或者, 这些术语可以表示, 将用于重组蛋白表达的载体递送给细胞或培养的细胞和 / 或受试者的细胞或器官。这样的施用或导入可以在体内、在体外或先体外后体内地发生。通过下述方式, 可以将用于重组蛋白或多肽表达的载体导入细胞中: 转染, 其通常是指, 通过物理方式 (例如, 磷酸钙转染、电穿孔、显微注射或脂转染) 将异源 DNA 插入细胞中; 感染, 其通常是指, 借助于传染性病原体 (即病毒) 的导入; 或转导, 其通常是指, 病毒对细胞的稳定感染, 或遗传物质通过病毒剂 (例如, 噬菌体) 从一种微生物向另一种微生物的转移。

[0118] “转化”通常用于表示, 包含异源 DNA 的细菌, 或表达癌基因且因此已经转变成连续生长模式的细胞, 例如, 肿瘤细胞。用于“转化”细胞的载体可以是质粒、病毒或其它媒介物。

[0119] 通常, 根据用于将异源 DNA (即, 载体) 施用、导入或插入细胞中的方式, 将细胞称作“转导的”、“感染的”、“转染的”或“转化的”细胞。术语“转导的”、“转染的”和“转化的”可以在本文中可互换地使用, 不论异源 DNA 的导入方法。

[0120] 本文使用的术语“稳定转化的”、“稳定转染的”和“转基因的”是指, 具有整合到基因组中的非天然的 (异源的) 核酸序列的细胞。通过由含有转染的 DNA (其通过整合进它们的基因组中或作为附加型元件而稳定地复制) 的子代细胞群体组成的细胞系或克隆的建立, 证实稳定的转染。在某些情况下, “转染”是不稳定的, 即它是瞬时的。在瞬时转染的

情况下,外源或异源 DNA 被表达,然而导入的序列没有整合到基因组中,或宿主细胞不能复制。

[0121] 本文使用的“先体外后体内施用”表示一种方法,其中从受试者取出原代细胞,将载体施用给所述细胞,以产生转导的、感染的或转染的重组细胞,并将所述重组细胞重新施用给相同或不同的受试者。

[0122] “多顺反子转录物”表示,含有超过一个蛋白编码区或顺反子的 mRNA 分子。包含 2 个编码区的 mRNA 被称作“双顺反子转录物”。“5′ - 近端”编码区或顺反子是这样的编码区:其翻译起始密码子(通常 AUG)与多顺反子 mRNA 分子的 5′ 末端最接近。“5′ - 远端”编码区或顺反子是这样的编码区:其翻译起始密码子(通常 AUG)不是与 mRNA 的 5′ 末端最接近的起始密码子。

[0123] 术语“5′ - 远端”和“下游”同义地用于表示,不与 mRNA 分子的 5′ 末端邻近的编码区。

[0124] 本文使用的“共转录”是指,2 个(或更多个)开放读码框或编码区或多核苷酸是在包含启动子的单个转录控制或调控元件的转录控制下。

[0125] 本文使用的术语“宿主细胞”表示,已经用载体转导、感染、转染或转化的细胞。所述载体可以是质粒、病毒颗粒、噬菌体等。诸如温度、pH 等培养条件是以前与选择用于表达的宿主细胞一起使用的那些,且是本领域技术人员显而易见的。应当理解,术语“宿主细胞”表示原始的转导的、感染的、转染的或转化的细胞及其后代。

[0126] 本文使用的术语“生物活性”和“生物学上有活性的”表示,在培养的细胞系中或在无细胞系统(诸如在 ELISA 平板中的配体-受体试验)中归因于特定蛋白的活性。“免疫球蛋白”、“抗体”或其片段的“生物活性”表示,结合抗原决定簇并从而促进免疫学功能的能力。激素或白介素的“生物活性”是本领域已知的。

[0127] 本文使用的术语“肿瘤”和“癌症”表示,表现出对正常生长和/或发育的控制的至少部分丧失的细胞。例如,肿瘤或癌细胞通常已经丧失接触抑制,且可能是侵袭性的,和/或具有转移的能力。

[0128] 抗体是作为重链和轻链的异源二聚体的免疫球蛋白蛋白质。一种典型的抗体是具有结合到一起的 2 个重链和 2 个轻链(或其功能片段)的多聚体。抗体可以具有其它的聚合结构级别,是二聚的、三聚的、四聚的、五聚的等,经常取决于同种型。已经证实,它们非常难以在哺乳动物培养表达系统中从单一载体或从 2 个载体以全长形式表达。几种方法目前被用于生产抗体:给动物体内免疫接种以生产“多克隆的”抗体,对 B- 细胞杂交瘤进行体外细胞培养以生产单克隆抗体(Kohler 等人 1988. Eur. J. Immunol. 6:511;Antibodies: A Laboratory Manual, Cold Spring Harbor Laboratory, 1988;通过引用并入本文),和重组 DNA 技术(例如描述在 Cabilly 等人,美国专利号 6331415,通过引用并入本文)。

[0129] 众所周知,免疫球蛋白多肽的基础分子结构包括 2 个相同的分子量为约 23,000 道尔顿的轻链和 2 个相同的分子量为 53,000-70,000 道尔顿的重链,其中所述 4 个链通过二硫键连接成“Y”构型。氨基酸序列在 Y 的顶部是 N- 末端,在每条链的底部是 C- 末端。在 N- 末端是可变区(长度为约 100 个氨基酸),其提供抗原结合的特异性。

[0130] 本发明涉及改进的用于生产所有类型的免疫球蛋白的方法,所述类型包括、但不限于:具有天然序列(即响应于抗原刺激而生成的序列)的全长抗体和抗体片段,在单个稳

定折叠的多肽链中组合了重链和轻链的抗原结合可变区的单链抗体；单价抗体（其包含与第二重链的 Fc 区结合的重链 / 轻链二聚体）；“Fab 片段”，其包括免疫球蛋白分子的整个“Y”区域，即，“Y”的分支，单独的轻链或重链或其部分（即，1 个重链和 1 个轻链的聚集体，通常称作 Fab'）；“杂种免疫球蛋白”，其具有对 2 种或更多种不同抗原的特异性（例如，细胞杂交瘤或双特异性的抗体，其例如描述在美国专利号 6,623,940 中）；“复合免疫球蛋白”，其中重链和轻链模仿来自不同的物种或特异性的那些；和“嵌合的抗体”，其中重链和轻链的每个氨基酸序列部分源自超过一个物种（即，可变区源自一个来源诸如鼠抗体，同时恒定区源自另一个来源诸如人抗体）。

[0131] 本发明的组合物和方法可以用于生产免疫球蛋白或其片段，其中重链或轻链是“哺乳动物的”、“嵌合的”或以增强它的效力的方式进行了修饰。修饰的抗体包括：保留未修饰形式的相同生物活性的氨基酸和核酸序列变体，和以使所述活性发生变化（即，增强补体结合、与膜的相互作用和其它效应子功能的恒定区中的变化，或增强抗原结合特征的可变区中的变化）的方式进行修饰的那些。本发明的组合物和方法可以另外包括催化免疫球蛋白或其片段。

[0132] 编码“变体”免疫球蛋白的多核苷酸序列可以编码相对于参照多肽序列而言改变了一个或多个氨基酸的“变体”免疫球蛋白氨基酸序列。下面的这种相同讨论适用于其它目标生物活性蛋白序列（和它们的编码序列）。所述变体多核苷酸序列可以编码含有“保守”置换的变体氨基酸序列，其中所述置换氨基酸具有与它替换的氨基酸类似的结构或化学性质。应当理解，可以制备目标蛋白的变体，其氨基酸序列与天然存在的序列的氨基酸序列基本上相同的（至少约 80-99% 同一性和在它们之间的任意整数），且其形成功能上等效的三维结构，并保留天然存在的蛋白的生物活性。在生物学领域众所周知，可以在蛋白序列中做出某些氨基酸置换，而不影响蛋白的功能。通常，可耐受保守氨基酸置换或类似氨基酸置换，而不影响蛋白功能。类似的氨基酸可以是大小和 / 或电荷性质类似的那些，例如，天冬氨酸和谷氨酸以及异亮氨酸和缬氨酸是两对类似的氨基酸。当天然的二级和三级结构形成不被破坏时（除了预见到以外），允许彼此的置换。在本领域已经用多种方式评估了氨基酸对之间的相似性。例如，Dayhoff 等人，Atlas of Protein Sequence and Structure, 1978. 第 5 卷，增刊 3，第 22 章，第 345-352 页（它通过引用并入本文），提供了可以用作氨基酸相似性的量度的氨基酸置换频率表。Dayhoff 等人的频率表是基于来自多种进化上不同来源的、具有相同功能的蛋白的氨基酸序列的对比。

[0133] 通过本领域众所周知的方法，可以容易地制备公开的核苷酸（和氨基酸）序列的置换突变、插入和缺失变体。这些变体可以以与具体例证的序列相同的方式使用，只要所述变体与本发明的具体例证的序列具有实质的序列同一性，并保留希望的功能性。

[0134] 本文使用的实质的序列同一性表示这样的同源性（或同一性）：其足以使变体多核苷酸或蛋白以与所述变体的来源多核苷酸或蛋白相同的能力发挥作用。优选地，该序列同一性是大于 70% 或 80%，更优选地，该同一性是大于 85%，或该同一性是大于 90%，和 / 或者，该同一性是大于 95% 以及在 70-100% 之间的所有整数。制备置换突变、插入和缺失突变是在本领域经过训练的人员的技能范围内，所述突变在功能上等同于所述序列的功能或被设计成改善所述序列的功能或以其它方式提供方法学优点。在任意天然存在的蛋白上可以读出的或在定量现有技术产品中可以读出的实施方案 / 变体无意落入要求保护的

发明范围内。本领域众所周知,可以截短和/或以其它方式突变本发明的多核苷酸序列,使得原始全长序列的某些得到的片段和/或突变体可以保留全长序列的希望的特征。适用于从更大的核酸分子产生片段的多种限制性酶是众所周知的。另外,众所周知,Bal31 外切核酸酶可以方便地用于 DNA 的时间控制的有限消化。参见,例如,Maniatis 等人 1982. *Molecular Cloning: A Laboratory Manual*, Cold Spring Harbor Laboratory, New York, 第 135-139 页,通过引用并入本文。也参见 Wei 等人 1983. *J. Biol. Chem.* 258 :13006-13512。通过使用 Bal31 外切核酸酶(通常称作“erase-a-base”规程),普通技术人员可以从主题核酸的任一端或两端去除核苷酸,以产生在功能上与主题核苷酸序列等效的宽谱片段。以此方式,本领域普通技术人员可以产生数百种具有受控的不同长度的片段,它们来自沿着原始编码序列的所有位置。普通技术人员可以常规地测试或筛选产生的片段的特征,并确定本文教导的片段的实用性。还众所周知,利用定位诱变,可以容易地生产全长序列的突变体序列或其片段。参见,例如, Larionov, O. A. 和 Nikiforov, V. G. 1982. *Genetika* 18 : 349-59 ; Shortle 等人 (1981) *Annu. Rev. Genet.* 15 :265-94 ;二者通过引用并入本文。技术人员可以常规地生产缺失、插入或置换型突变,并鉴定含有全长野生型序列的希望特征的那些得到的突变体或其片段,例如,保留激素、细胞因子、抗原结合或其它生物活性的那些。

[0135] 额外地或替代地,所述变体多核苷酸序列可以编码含有“非保守”置换的变体氨基酸序列,其中所述置换的氨基酸具有与它替换的氨基酸不相似的结构或化学性质。编码变体免疫球蛋白的多核苷酸也可以编码含有氨基酸插入或缺失或二者的变体氨基酸序列。此外,编码变体免疫球蛋白的多核苷酸可以编码与参照多核苷酸序列相同的多肽,但是由于遗传密码的简并性,所述多肽的多核苷酸序列相对于参照多核苷酸序列改变了一个或多个碱基。

[0136] 当提及本发明的重组免疫球蛋白时,术语“片段”是指这样的多肽:所述多肽的氨基酸序列与对应的全长免疫球蛋白蛋白质的氨基酸序列的一部分(但不是全部)相同,所述多肽要么基本上保留与对应的全长蛋白相同的生物学功能或活性,要么保留对应的全长蛋白的至少一种功能或活性。所述片段优选地包括全长免疫球蛋白的至少 20-100 个连续氨基酸残基,且优选地,保留结合与全长抗体相同的抗原的能力。

[0137] 本文使用的术语“序列同一性”是指,当使用序列比对程序进行比对时,2 个或更多个比对的序列之间的核酸或氨基酸序列同一性。术语“%同源性”在本文中与术语“%同一性”可互换地使用,并表示,当使用序列比对程序进行比对时,2 个或更多个比对的序列之间的核酸或氨基酸序列同一性水平。例如,本文使用的 80%同源性是指,通过本领域理解的确定的算法,测得与 80%序列同一性相同的物质,因此,给定序列的同系物在给定序列的长度上具有大于 80%序列同一性。

[0138] 通过例如下述算法,可以进行比较序列的最佳比对:Smith 和 Waterman. 1981. *Adv. Appl. Math.* 2 :482 的局部同源性算法, Needleman 和 Wunsch. 1970. *J. Mol. Biol.* 48 : 443 的同源性比对算法, Pearson 和 Lipman. 1988. *Proc. Natl. Acad. Sci. USA* 85 :2444 的相似性检索方法,这些算法的计算机化实现(在威斯康星遗传学软件包中的 GAP、BESTFIT、FASTA 和 TFASTA, Genetics Computer Group, Madison, Wis.), Altschul 等人 1990. *J. Mol. Biol.* 215 :403-410 的 BLAST 算法,利用从国家生物技术信息中心网站(参见 nml.nih.gov/)上可公开得到的软件,或通过目检(通常参见 Ausubel 等人,同下文)。为了本发明

的目的,最优选地通过 Smith 和 Waterman, 1981. Adv. Appl. Math. 2 :482 的局部同源性算法进行比较序列的最佳比对。也参见, Altschul 等人 1990 和 Altschul 等人 1997。

[0139] 在 2 个或更多个核酸或蛋白序列的背景下,术语“相同的”或“同一性百分比”表示,当对比和比对最大对应时,使用本文所述的序列对比算法之一(例如 Smith-Waterman 算法)、本领域已知的其它算法(例如, BLAST)或通过目检测得,2 个或更多个序列或子序列是相同的,或具有相同的氨基酸残基或核苷酸的指定百分比。

[0140] 根据本发明,也包括这样的序列变体:其编码自加工切割多肽,和其本身与天然或参照序列具有 80、85、88、89、90、91、92、93、94、95、96、97、98、99% (和在 80 至 100 之间的所有整数百分比值)或更多序列同一性的多肽。也包括所述多肽的氨基酸片段,其代表至少 5 个、至少 10 个或至少 15 个单元连续段;和根据所述的同一性条件与其同源的片段;以及代表至少 15 个、至少 30 个或至少 45 个单元连续段的核酸序列片段。在一个具体实施方案中,核酸序列或氨基酸序列与一并公开的各个序列具有 90%、91%、92%、93%、94%、95%、96%、97%、98%、99% 或 99.5% 同一性。

[0141] 如果两个序列在中等至高严谨性杂交和洗涤条件下彼此特异性地杂交,那么认为核酸序列与参照核酸序列“可选择性地杂交”。杂交条件是基于核酸结合复合物或探针的解链温度 (T_m)。例如,“最大严谨性”典型地发生在约 $T_m-5^{\circ}\text{C}$ (探针 T_m 以下 5°C);“高严谨性”则在 T_m 以下约 $5-10^{\circ}\text{C}$;“中等严谨性”在探针 T_m 以下约 $10-20^{\circ}\text{C}$;“低严谨性”在 T_m 以下约 $20-25^{\circ}\text{C}$ 。在功能上,最大严谨性条件可以用于鉴定与杂交探针具有严格同一性或近似严格同一性的序列;而高严谨性条件被用于鉴定与探针具有约 80% 或更高序列同一性的序列。

[0142] 中等和高严谨性杂交条件是本领域众所周知的(参见,例如, Sambrook 等人, 1989, 第 9 和 11 章,和 Ausubel, F.M. 等人, 1993)。高严谨性条件的一个实例包括:在约 42°C 、在 50% 甲酰胺、5X SSC、5X 登哈特溶液、0.5% SDS 和 $100\mu\text{g/ml}$ 变性载体 DNA 中进行杂交,然后在 2X SSC 和 0.5% SDS 中在室温洗涤两次,在 42°C 在 0.1X SSC 和 0.5% SDS 中洗涤另外两次。编码具有与天然存在的目标蛋白相同的生物活性的多肽、并在中等至高严谨性杂交条件下杂交的 2A 序列变体,被认为是在本发明的范围内。

[0143] 作为遗传密码的简并性的结果,可以生产许多编码序列,它们编码相同的多肽序列,包括诸如结构组分、自加工组分(例如内含肽)、调节组分(例如,信号肽酶切割序列)或其它组分。例如,三联体 CGT 编码氨基酸精氨酸。或者,精氨酸由三联体核苷酸序列 CGA、CGC、CGG、AGA 和 AGG 编码。因此,应当理解,编码区中的同义密码子的这种置换落入本发明所涵盖的序列变体内。

[0144] 另外应当理解,这样的序列变体可以在高严谨性条件下与亲本序列杂交或不杂交。例如,当序列变体包括由亲本核苷酸编码的每个氨基酸的不同密码子时,这是可能的。尽管如此,本发明具体地预见到和包括这样的变体。

[0145] 抗体作为治疗模态的潜力目前受到当前技术的生产能力和费用的限制。用于免疫球蛋白(或其它蛋白)生产的改进的病毒或非病毒单个表达载体会促进 2 个或更多个编码序列(即,来自单个载体的、具有双特异性或多特异性的免疫球蛋白或其它蛋白)的表达和递送。本发明解决了这些限制,并适用于本文进一步详述的任意免疫球蛋白(即抗体)或其片段或其它多组分蛋白或结合蛋白对,包括工程化的抗体诸如单链抗体、全长抗体或抗

体片段、双链激素、双链细胞因子、双链趋化因子、双链受体等。

[0146] 内含肽

[0147] 本文使用的内含肽是在表达的蛋白内的区段,其向初级表达产物的 N- 端方向结合 N- 外显肽,并向初级表达产物的 C- 端方向结合 C- 外显肽。天然存在的内含肽会介导内含肽的切除和 N- 和 C- 外显肽的重新连接(蛋白连接)。但是,在本发明的表达产物的背景下,内含肽或侧接外显肽氨基酸序列的基本序列使得,在没有外显肽连接存在下,或在减少的或最小量的外显肽连接下,发生蛋白骨架的切割,因而外显肽蛋白从初级翻译产物(多蛋白)释放出来,它们不会连接形成融合蛋白。初级表达产物(在任何蛋白水解性裂解之前,由 mRNA 合成的蛋白)的内含肽部分会介导在 N- 外显肽 / 内含肽和内含肽 / C- 外显肽接头处的蛋白水解性裂解。一般而言,天然存在的内含肽也介导 N- 外显肽和 C- 外显肽剪接到一起(通过肽键的形成而连接)。但是,在适用于表达 2 种多肽(具体地由抗体分子的重链和轻链来例证)的目标的本发明中,优选地,不会发生蛋白连接。这可以通过掺入内含肽来实现,所述内含肽天然地或由于突变不具有连接活性。或者,通过突变可以阻止剪接,所述突变会改变在剪接位点处或之后的氨基酸,以防止释放的蛋白的连接。参见 Xu 和 Perler, 1996, EMBO J. 15 :5146-5153。例如, Ser、Thr 或 Cys 通常存在于 C- 外显肽的起点处,且可以被改变以修饰或中断剪接。在一种具体的内含肽中,对剪接的影响会阻止或减少表达的蛋白的连接。

[0148] 内含肽是一类蛋白,它们的基因仅存在于其它蛋白的基因内。与称作外显肽的侧接宿主基因一起,内含肽被转录为单个 mRNA,并翻译为单个多肽。翻译后,内含肽启动自催化事件,以去除它们自身并用新的多肽键连接侧接的宿主蛋白区段。该反应仅由内含肽催化,不需要其它细胞蛋白、辅因子或 ATP。内含肽存在于多种单细胞生物体中,且它们具有不同的大小。许多内含肽含有内切核酸酶结构域,该结构域造成它们在基因组内的可动性。

[0149] 内含肽介导的反应已经被用于生物技术中,特别是体外场合,诸如用于纯化和用于蛋白芯片构建,和用于植物品种改良(Perler, F.B. 2005. IUBMB Life 57(7) :469-76)。已经将突变导入天然内含肽核苷酸序列中,据报道,这些突变体中的一些具有改变的性质(Xu 和 Perler, 1996. EMBO J. 15(9), 5146-5153)。除了内含肽以外,也已知细菌内含肽 - 样(BIL) 结构域和刺猬(Hog) 自加工结构域(Hog/ 内含肽(HINT) 超家族中的另外 2 个成员)会通过类似的机理催化翻译后自加工(Dassa 等人 2004. J. Biol. Chem. 279(31) : 32001-32007)。

[0150] 内含肽作为框架内插入存在于特定宿主蛋白中。在自剪接反应中,内含肽将它们自身切离前体蛋白,同时侧接区域(即外显肽)相连以恢复宿主基因功能。这些元件也含有内切核酸酶功能,该功能造成它们在基因组内的可动性。内含肽以一定的大小范围(134-1650 个氨基酸)存在,已经在真细菌、真核生物和古细菌的基因组中鉴定出它们。使用模型剪接 / 报告系统的实验已经证实,可以分离内切核酸酶、蛋白切割和蛋白剪接功能(Xu 和 Perler. 1996. EMBO J. 15 :5146-5153)。下述的实施例使用来自掘越氏火球菌 Pho Pol I、酿酒酵母 VMA 和集胞藻属的内含肽,以建立具有来自抗体重链和轻链的序列的融合蛋白。设计成删除内含肽的剪接能力的内含肽突变会产生这样的单个多肽:其发生自切割,以生成正确地编码的抗体重链和轻链。该策略可以类似地用于表达其它多链蛋白、激素或细胞因子,且它也可以适用于将前体蛋白(前蛋白)加工成它们的成熟的、生物活性的形

式。尽管在本文中具体地例证了掘越氏火球菌 Pho PoI I、酿酒酵母 VMA 和集胞藻属内含肽的使用,本领域已知的其它内含肽可以用于本发明的多蛋白表达载体和方法中。

[0151] 除了掘越氏火球菌 Pho PoI I、酿酒酵母 VMA 和集胞藻属内含肽以外,许多其它内含肽是本领域已知的(参见,例如,Perler, F.B. 2002, InBase, the Intein Database, Nucl. Acids Res. 30(1):383-384 和 the Intein Database and Registry, 可通过新英格兰 Biolabs 网站得到,例如,在 <http://tools.neb.com/inbase/>)。已经在宽范围的生物体(诸如酵母、分枝杆菌和极端嗜热古细菌)中鉴定出内含肽。某些内含肽具有内切核酸酶活性以及位点特异性的蛋白切割和剪接活性。内切核酸酶活性不是本发明的实践所必需的;可以删除内切核酸酶编码区,条件是,维持蛋白切割活性。

[0152] 已经非常详细地研究了蛋白剪接过程的机理(Chong 等人 1996. J. Biol. Chem. 271:22159-22168; Xu 和 Perler. 1996. EMBO J 15:5146-5153), 并已经在内含肽和外显肽剪接点处发现了保守氨基酸(Xu 等人 1994. EMBO J 13:5517-5522)。本文描述的某些构建体含有与所述第一编码序列的 3' - 端融合的内含肽序列,第二编码序列在内含肽的 C- 端处在框架内融合。合适的内含肽序列可以选自己知含有蛋白剪接元件的任意蛋白。在环球网上可以发现含有所有已知的内含肽的数据库(Perler, F.B. 1999. Nucl. Acids Res. 27:346-347)。内含肽编码序列在 3' 末端处与第二编码序列的 5' 末端融合(在框架内)。为了使该蛋白靶向特定细胞器,可以使适当的肽信号与蛋白的编码序列融合。

[0153] 在第二外显肽编码序列以后,内含肽编码序列 - 外显肽编码序列可以如在相同细胞中表达多种蛋白所希望地经常重复。对于含有多个内含肽的构建体,可以有用地使用来自不同来源的内含肽元件。在要表达的最后一个基因的序列之后,合乎需要地插入转录终止序列(并有利地包括聚腺苷酸化序列)。聚腺苷酸化序列和终止序列的次序可以是本领域所理解的。在一个实施方案中,聚腺苷酸化序列可以在终止序列前面。

[0154] 已经设计修饰的内含肽剪接单元,使得这样的修饰的目标内含肽可以催化外显肽从内含肽切除,但是不催化外显肽的连接(参见,例如,美国专利 7026526 和美国专利公开 20020129400)。火球菌属物种 GB-D DNA 聚合酶中的 C- 端外显肽接头的诱变会生成改变的剪接元件,其诱导外显肽和内含肽的切割,但是阻止外显肽的随后连接(Xu 和 Perler. 1996. EMBO J 15:5146-5153)。丝氨酸 538 向丙氨酸或甘氨酸(Ser 向 Ala 或 Gly)的突变会诱导切割,但是阻止连接。在这样的位置,Ser 向 Met 或 Ser 向 Thr 的突变也被用于实现多蛋白的表达,所述多蛋白被切割成单独的区段,且至少部分地不重新连接。在其它内含肽剪接单元中的等效残基的突变也可以阻止外显肽区段的连接,这是由于与内含肽相连的 C- 端外显肽处的氨基酸的相对保守。在低保守/同源性的情况下,例如,系统地改变 C- 外显肽的前几个(例如,约 5 个)残基和/或内含肽区段的最后几个残基,并筛选它们的支持切割、但是不支持给定外显肽区段(尤其是本文公开的和本领域理解的外显肽区段)的剪接的能力。存在不含有内切核酸酶结构域的内含肽;这些包括集胞藻属 dnaE 内含肽和蟾分枝杆菌 GyrA 蛋白(Magnasco 等人, Biochemistry, 2004, 43, 10265-10276; Telenti 等人 1997. J. Bacteriol. 179:6378-6382)。其它内含肽已经在自然界中发现,或已经通过从编码含有内切核酸酶的内含肽的序列去除编码内切核酸酶的结构域来人工地产生(Chong 等人 1997. J. Biol. Chem. 272:15587-15590)。在需要的情况下,最初选择内含肽,使得它由执行剪接功能所需的最小数目的氨基酸组成,诸如来自蟾分枝杆菌 GyrA 蛋白的内含肽

(Telenti 等人 1997. 同上)。在一个替代实施方案中,选择没有内切核酸酶活性的内含肽,诸如来自蟾分枝杆菌 GyrA 蛋白的内含肽,或已经被修饰以去除内切核酸酶结构域的酿酒酵母 VMA 内含肽 (Chong 等人 1997. 同上)。

[0155] 内含肽剪接单元的其它修饰可以允许改变切割反应的反应速率,允许通过简单地修饰剪接单元的基因序列来控制蛋白剂量。

[0156] 在一个实施方案中,将 C- 端外显肽的第一个残基工程化成含有甘氨酸或丙氨酸,即经证实会阻止使用火球菌属物种 GB-D DNA 聚合酶的外显肽连接的修饰 (Xu 和 Perler. 1996. EMBO J 15 :5146-5153)。在该实施方案中,优选的 C- 端外显肽蛋白天然地含有甘氨酸或丙氨酸残基,该残基在天然氨基酸序列中在 N- 端甲硫氨酸之后。外显肽的甘氨酸或丙氨酸与内含肽的 C- 端的融合,会在多蛋白的加工以后提供天然氨基酸序列。在另一个实施方案中,通过改变天然序列,或通过将额外的氨基酸残基添加在天然序列的 N- 端上,将人工的甘氨酸或丙氨酸放置在 C- 端外显肽中。在该实施方案中,在多蛋白加工以后,蛋白的天然氨基酸序列将被改变一个氨基酸。在其它实施方案中,在本发明中有用的其它修饰描述在 US 7026526 中。

[0157] 在一个实施方案中,内含肽是根据美国专利 7,026,526。在一个具体实施方案中,内含肽是火球菌属物种 GB-D DNA 聚合酶内含肽。在一个实施方案中,C- 端外显肽丝氨酸向丙氨酸或甘氨酸的突变会形成修饰的内含肽剪接元件,其能够促进多蛋白的切除,但是不促进外显肽单位的连接。在一个实施方案中,内含肽是美国专利 7,026,526 的蟾分枝杆菌 GyrA 最小内含肽。在一个实施方案中,C- 端外显肽苏氨酸向丙氨酸或甘氨酸的突变会形成修饰的内含肽剪接元件,其促进多蛋白的切除,但是不连接外显肽单元。

[0158] 应当理解,对于本文所述的某些内含肽,构建体和方案的实施方案可以产生分泌型蛋白产物 (特别是对于包括抗体在内的多聚蛋白而言) 的提高了的表达。

[0159] 信号肽和信号肽酶

[0160] 已经知道信号假设 (其中蛋白在它们的氨基酸序列内含有关于靶向膜的蛋白的信息) 超过 30 年。Milstein 和同事发现,当将内质网囊泡 (微粒体) 加入翻译系统中时,来自骨髓瘤细胞的 IgG 的轻链以更高分子量形式合成,并转化成它的成熟形式,提出的模型是基于这些结果,其中微粒体含有蛋白酶,所述蛋白酶通过去除氨基- 端延伸肽将前体蛋白形式转化成成熟形式。不久将信号假设扩展至在蛋白内包括独特的靶向序列,所述蛋白定位在不同的细胞内膜 (诸如线粒体和叶绿体) 上。以后发现,这些独特的靶向序列被特定信号肽酶 (SPase) 从输出的蛋白切掉。

[0161] 存在至少 3 种涉入细菌中的信号肽切割的独特 SPase。SPase I 可以加工非脂蛋白底物,所述底物由 SecYEG 途径或双精氨酸易位 (Tat) 途径输出。由 Sec 途径输出的脂蛋白被 SPase II 切割。SPase IV 切割 IV 型前菌毛素和前菌毛素- 样蛋白 (它们是 II 型分泌器官的组分)。

[0162] 在真核生物中,靶向内质网 (ER) 膜的蛋白由信号肽介导,所述信号肽共翻译地或翻译后地将蛋白靶向 Sec61 易位机构。ER 信号肽具有与它们的细菌对应物的那些类似的特征。ER 信号肽在被信号肽酶复合物 (SPC) 输出到内质网腔中以后从输出的蛋白上切掉。

[0163] 将蛋白分类至真核细胞内的不同位置的信号肽必须是不同的,因为这些细胞含有许多不同的膜状的和水性的隔室。靶向内质网的蛋白经常含有可切割的信号序列。令人惊

讶的是,许多人工肽可以起易位信号的作用。认为最重要的关键特征是高于特定阈值的疏水性。ER 信号肽具有比细菌信号肽更高的亮氨酸残基含量。信号识别颗粒 (SRP) 在它们从核糖体出现以后结合可切割的信号肽。新生蛋白向 ER 膜的靶向需要 SRP。在蛋白易位至内质网腔以后,输出的蛋白被 SPC 加工。另一个实施方案利用在真核细胞中天然地存在的信号 (先导) 肽加工酶。

[0164] 大多数已知的 ER 信号肽是 N- 端可切割的或内部不可切割的。最近,发现许多病毒多蛋白 (诸如在丙型肝炎病毒、汉坦病毒、黄病毒、风疹病毒和 C 型流感病毒中发现的那些) 含有最可能被 ER SPC 切割的内部信号肽。这些对多蛋白的成熟的研究表明,SPC 不仅可以切割位于氨基端的信号肽,而且在内部信号肽之后切割。预测的信号肽酶底物特异性元件的诱变因而可以阻断病毒传染性。

[0165] 早老素 - 型天冬氨酸蛋白酶信号肽肽酶 (SPP) 会切割在它们的跨膜区域内的信号肽。SPP 是在人类中产生信号肽 - 衍生的 HLA-E 表位所必需的。

[0166] 信号肽酶是本领域众所周知的。参见,例如,Paetzel M. 2002. Chem. Rev. 102(12) : 4549 ; Pekosz A. 1998. Proc. Natl. Acad. Sci. USA. 95 : 13233-13238 ; Marius K. 2002. Molecular Cell 10 : 735-744 ; Okamoto K. 2004. J. Virol. 78 : 6370-6380, Vol. 78 ; Martoglio B. 2003. Human Molecular Genetics 12 : R201-R206 ; 和 Xia W. 2003. J. Cell Sci. 116 : 2839-2844。

[0167] 本发明的实施方案利用内部可切割的信号肽,用于表达在单一转录物中的多肽。单个转录的多肽随后被 SPC 切割,剩下单独的单个肽,或单个肽被装配成蛋白。本发明的方法适用于在单个转录的多肽中表达免疫球蛋白重链和轻链,随后切割,然后装配成成熟的免疫球蛋白。该技术适用于多肽细胞因子、生长因子或多种其它蛋白,例如,在单个转录的多肽中的 IL-12p40 和 IL-12p35,然后装配成在单个转录的多肽中的 IL-12 或 IL-12p40 和 IL-23p 19,然后装配成 IL-23。

[0168] 信号肽酶方案适用于哺乳动物表达载体,所述载体导致来自前体或多蛋白的功能抗体或其它加工产物的表达。在抗体的情况下,它从载体生成,作为含有重链和轻链的多蛋白,其中在重链和轻链之间的插入序列是内部可切割的信号肽。该内部可切割的信号肽可以被定位于内质网的蛋白酶 (主要是信号肽酶)、早老素或早老素 - 样蛋白酶切割,导致重链和轻链折叠和装配,以产生功能分子,并合乎需要地分泌它。除了源自丙型肝炎病毒的内部可切割的信号肽以外,可以被定位于内质网的蛋白酶切割的其它内部可切割的序列可以替代它们。类似地,本发明的实践不一定限于其中信号肽酶引起切割的宿主细胞,但是它也包括蛋白酶,包括、但不限于 : 早老素、早老素 - 样蛋白酶和用于加工多肽的其它蛋白酶。那些蛋白酶已经综述在引用的文章以及其它文章中。

[0169] 另外,本发明不限于免疫球蛋白重链和轻链的表达,而且也包括其它多肽和多蛋白,它们在单一转录物中表达,随后进行内部信号肽切割,以释放出每种单个肽或蛋白。这些蛋白可以在成熟的产物中装配到一起或不装配到一起。

[0170] 在本发明的范围内也包括表达构建体,其中单个多肽以交替的次序存在,即,“肽 1- 内部可切割的信号肽 - 肽 2”或“肽 2- 内部可切割的信号肽 - 肽 1”。本发明另外包括通过内部可切割的信号肽相连的超过 2 种肽的表达,诸如“肽 1- 内部可切割的信号肽 - 肽 2- 内部可切割的信号肽 - 肽 3”,诸如此类。

[0171] 另外,本发明适用于 I 型和 II 型跨膜蛋白的表达,并适用于与表达构建体相关的其它蛋白酶切割位点。

[0172] 本发明可以进一步利用内部可切割的信号肽,用于一种或多种多肽的成熟,所述多肽是在单一转录物内编码的多蛋白内。所述单个转录的多肽然后被 SPC 切割,剩下单独的单个肽,或单个肽装配成蛋白。本发明的实施方案包括用于在单个转录的多肽中表达免疫球蛋白重链和轻链、并最终装配成成熟的免疫球蛋白的组合物和方法。本发明适用于表达多肽细胞因子、生长因子或多种其它蛋白,例如,在单个转录的多肽中表达 IL-12p40 和 IL-12p35,然后装配成在单个转录的多肽中的 IL-12 或 IL-12p40 和 IL-23p19,然后装配成 IL-23。

[0173] 信号肽的修饰

[0174] 在 sORF 构建体的实施方案中,采用修饰的信号肽。例如,在重链-int-轻链的构建体中,当通过定位诱变降低轻链信号肽序列的疏水性时,抗体分泌水平增加至约 10 倍。可以如 Carson 等人于 2007 年 3 月 22 日提交的 US 20070065912 所述,采用信号肽。

[0175] 标签

[0176] 本发明的 sORF 构建体设计的实施方案包括:使用含有标签(优选内部标签)的经修饰的内含肽。多种标签是本领域已知的。本发明的标签包括、但不限于:荧光标签和化学发光的标签。使用这样的构建体,使用在单个细胞中的荧光检测,可以监测表达的多蛋白的量。另外,使用荧光活化细胞分选(FACS),根据蛋白表达水平,可以分选这些细胞。这样的标签的使用在稳定细胞系产生中是特别有用的,因为这允许通过 FACS 分析来选择高产细胞或细胞系。如在本发明中所教导的,在它们从侧接抗体重链和轻链自动切割下以后,已经在细胞裂解物中观察到全长内含肽。这会提供用于检测荧光标记的内含肽的基础和它们在稳定的细胞系产生中的应用的基础。标签也可以用于蛋白的纯化中。

[0177] 小型内含肽(Mini-Intein)

[0178] 因为存在于许多内含肽(包括掘越氏火球菌 PoII 内含肽和酿酒酵母 VMA 内含肽)中的内切核酸酶区域对于基因表达系统而言不是特别有利的,可以任选地删除内切核酸酶结构域,并用小连接物替代,以建立“小型内含肽”。这些工程化的小型内含肽也可用于描述的构建体设计中,且它们具有下述优点:内含肽编码区明显更小,因而允许更大的序列编码目标多肽和/或更容易地操纵重组 DNA 分子。

[0179] 在一些实施方案中,采用具有全人特征的抗体或其类似物是有利的。这些试剂会避免由源自非人物种的抗体或类似物诱发的不希望免疫应答。为了解决对源自自加工肽的氨基酸残基的可能的宿主免疫应答,可以将蛋白水解性切割位点的编码序列插入(使用本领域已知的标准方法学)在所述第一蛋白的编码序列和所述自加工肽的编码序列之间,从而从表达的多肽(即抗体)去除自加工肽序列。该发现特别适用于在体内使用的治疗性或诊断性抗体。

[0180] 基因递送和包括病毒载体在内的载体

[0181] 本发明预见到众多载体中的任一种用于将构建体导入细胞中的应用,所述构建体包含两种或更多种多肽或蛋白质的编码序列和自加工切割序列。基因表达载体的众多实例是本领域已知的,且可以具有病毒或非病毒起源。可用于实践本发明的非病毒基因递送方法包括、但不限于:质粒、脂质体、核酸/脂质体复合物、阳离子脂质等。

[0182] 病毒和其它载体可有效地转导细胞,并将它们自身的 DNA 导入宿主细胞中。在产生重组病毒载体中,用编码目标蛋白质或多肽的可表达序列替换非必需基因。示例性的载体包括、但不限于病毒和非病毒载体,如逆转录病毒载体(包括慢病毒载体)、腺病毒(Ad)载体(包括它们的复制活性的、复制缺陷的和无肠的形式)、腺伴随病毒(AAV)载体、猴病毒 40(SV-40)载体、牛乳头状瘤病毒载体、Epstein-Barr 载体、疱疹载体、牛痘载体、莫洛尼鼠白血病病毒载体、Harvey 鼠肉瘤病毒载体、鼠乳腺肿瘤病毒载体、鲁斯肉瘤病毒载体和非病毒质粒。杆状病毒载体是众所周知的,且适用于在昆虫细胞中表达。众多适用于在哺乳动物或其它真核细胞中表达的载体是本领域众所周知的,且许多是可商业得到的。商业来源包括、但不限于:Stratagene, La Jolla, CA; Invitrogen, Carlsbad, CA; Promega, Madison, WI 和 Sigma-Aldrich, St. Louis, MO。许多载体序列可从 GenBank 得到,并且关于载体的其它信息可在因特网上通过 Riken BioSource Center 得到。

[0183] 在一个实施方案中,载体通常包含复制起点,并且载体可以另外包含或者不包含“标志物”或“选择标记”功能,通过它们可以鉴别和选择所述载体。尽管可以使用任何选择标记,用于重组载体中的选择标记通常是本领域已知的,适当的选择标记的选择取决于宿主细胞。编码赋予抗生素或其它毒素抗性的蛋白质的选择标记基因的实例包括、但不限于:氨苄西林、甲氨蝶呤、四环素、新霉素(Southern 等人 1982. J Mol Appl Genet. 1:327-41)、霉酚酸(Mulligan 等人 1980. Science 209:1422-7)、嘌呤霉素、zeomycin、潮霉素(Sugden 等人 1985. Mol Cell Biol. 5:410-3)、二氢叶酸还原酶、谷氨酰胺合成酶和 G418。本领域技术人员会理解,表达载体通常包括适合被表达的编码序列的复制起点、与编码序列或要表达的序列可操作地连接的启动子以及核糖体结合位点、RNA 剪接位点、聚腺苷酸化位点和转录终止子序列。

[0184] 将载体或其它 DNA 序列称作“重组的”,仅仅承认在自然状态下分离的或者发现的通常未可操作地连接的 DNA 序列的可操作连接。当表达和/或控制序列调节核酸序列的转录和翻译时(适当时),该调节(表达和/或控制)序列与核酸编码序列可操作地连接。因此,表达和/或控制序列可包括启动子、增强子、转录终止子、编码序列的 5' 起始密码子(即 ATG)、内含子的剪接信号和终止密码子。

[0185] 已知腺病毒基因治疗载体表现出强烈的瞬时表达、优良的效价以及在体内转导分裂的和未分裂的细胞的能力(Hitt 等人 2000. Adv in Virus Res 55:479-505)。重组 Ad 载体可以包含:使得载体掺入复制缺陷型 Ad 病毒粒子中的包装位点;两种或更多种目标多肽或蛋白质的编码序列,例如,目标免疫球蛋白的重链和轻链;以及编码单独的自加工切割位点或者与额外的蛋白水解性切割位点相组合的自加工切割位点的序列。对于掺入感染性病毒粒子中而言必需的或辅助的其它元件包括:5' 和 3' Ad ITR、E2 基因、E4 基因的部分及任选的 E3 基因。

[0186] 通过本领域已知的标准技术,使用 Ad 包装细胞和包装技术,生产包囊重组 Ad 载体的复制缺陷型 Ad 病毒粒子。这些方法的实例可参见,例如,美国专利号 5,872,005。两种或更多种目标多肽或蛋白质的编码序列通常插入在腺病毒的病毒基因组的缺失的 E3 区域中。用于实践本发明的优选的腺病毒载体不表达一种或多种野生型 Ad 基因产物,例如, E1a、E1b、E2、E3 和 E4。优选的实施方案是这样的病毒粒子:其通常与包装细胞系一起使用,所述包装细胞系补充 E1、E2A、E4 及任选的 E3 基因区域的功能。参见,例如美国专利号

5, 872, 005、5, 994, 106、6, 133, 028 和 6, 127, 175。

[0187] 除了另外指出之外,本文使用的“腺病毒”和“腺病毒颗粒”表示病毒自身或其衍生物,并且覆盖所有血清型和亚型以及天然存在的和重组的形式。这种腺病毒可以是野生型,或者以本领域已知的或者本文公开的不同方式修饰。这种修饰包括:对包装在颗粒中的腺病毒基因组的修饰,以便产生感染性的病毒。这种修饰包括本领域已知的缺失,如缺失 E1a、E1b、E2a、E2b、E3 或 E4 编码区中的一个或多个。示例性的包装和生产细胞源自 293、A549 或者 HeLa 细胞。使用本领域已知的标准技术,纯化和配制腺病毒载体。

[0188] 腺伴随病毒 (AAV) 是一种依赖于辅助病毒的人细小病毒,其能通过染色体整合而潜伏地感染细胞。因为它的整合进染色体中的能力和它的非病原性质,AAV 作为人基因治疗载体具有重要潜力。为了用于实践本发明,可以使用本领域技术人员已知的标准方法产生 rAAV 病毒粒子,并构建成使得它们包括控制序列(包括转录起始和终止序列)及目标编码序列(作为在转录方向可操作地连接的组分)。更具体地,本发明的重组 AAV 载体包含:能使载体整合入复制缺陷 AAV 病毒粒子中的包装位点;两种或更多种目标多肽或蛋白质的编码序列,例如,目标免疫球蛋白的重链和轻链;编码单独的自加工切割位点或者与一个或多个额外蛋白水解性切割位点相组合的自加工切割位点的序列。可以构建用于实践本发明的 AAV 载体,使得它们还包括控制序列(包括转录起始和终止序列)作为在转录方向可操作地连接的组分。这些组分在 5' 和 3' 末端上侧接功能性的 AAV ITR 序列。“功能性的 AAV ITR 序列”是指,ITR 序列具有预期的挽救、复制和包装 AAV 病毒粒子的功能。

[0189] 重组 AAV 载体的特征还在于,它们能够指导选择的重组多肽或蛋白产物在靶细胞中的表达和生产。因此,重组载体至少包含包囊化必需的所有 AAV 序列及重组 AAV (rAAV) 病毒粒子的感染必需的物理结构。因此,用于表达载体中的 AAV ITR 不需要具有野生型核苷酸序列(例如在 Kotin. 1994. Hum. Gene Ther. 5:793-801 中所述),且可以通过核苷酸的插入、缺失或置换而改变,或者 AAV ITR 可以源自几种 AAV 血清型中的任一种。通常,AAV 载体可以是源自本领域已知的腺伴随病毒血清型的任意载体。

[0190] 通常,将 AAV 表达载体导入生产细胞中,随后导入 AAV 辅助构建体中,其中所述辅助构建体包括这样的 AAV 编码区:其能在生产细胞中表达,并且补充在 AAV 载体中缺少的 AAV 辅助功能。所述辅助构建体可以设计为减量调节大 Rep 蛋白 (Rep78 和 Rep68) 的表达,典型地通过将在 p5 之后的起始密码子从 ATG 突变为 ACG,如在通过引用并入本文的美国专利号 6, 548, 286 中所述。随后在生产细胞中导入辅助病毒和 / 或另外的载体,其中所述辅助病毒和 / 或另外的载体提供能支持有效的 rAAV 病毒生产的附属功能。然后培养生产细胞以生产 rAAV。使用标准方法进行这些步骤。通过本领域已知的标准技术,使用 AAV 包装细胞和包装技术,制备包囊化本发明的重组 AAV 载体的复制缺陷型 AAV 病毒粒子。这些方法的实例可以参见,例如,美国专利号 5, 436, 146、5, 753, 500、6, 040, 183、6, 093, 570 和 6, 548, 286, 它们通过引用整体并入本文。用于包装的其它组合物和方法描述在 Wang 等人(美国专利公开 2002/0168342, 也通过引用整体并入本文)中,并包括属于本领域技术人员的那些技术。

[0191] 在实践本发明中,用于生产 rAAV 或其它载体表达载体病毒粒子的宿主细胞包括哺乳动物细胞、昆虫细胞、微生物和酵母。宿主细胞也可以是包装细胞,其中 AAV (或其它) rep 和 cap 基因稳定地维持在宿主细胞或生产细胞中,其中 AAV 载体基因组被稳定地维持

和包装。示例性的包装和生产细胞源自 293、A549 或者 HeLa 细胞。使用本领域已知的标准技术,纯化和配制 AAV 载体。其它合适的宿主细胞(取决于载体)包括:中国仓鼠卵巢(CHO)细胞、CHO 二氢叶酸还原酶缺陷型变体诸如 CHO DX B11 或 CHO DG44 细胞(参见,例如,Urlaub 和 Chasin. 1980. Proc. Natl. Acad. Sci. 77 :4216-4220)、PerC. 6 细胞(Jones 等人 2003. Biotechnol. Prog. 19 :163-168)或 Sp/20 小鼠骨髓瘤细胞(Coney 等人 1994. Cancer Res. 54 :2448-2455)。

[0192] 逆转录病毒载体

[0193] 逆转录病毒载体可以用于基因递送(Miller. 1992. Nature 357 :455-460)。逆转录病毒载体及更具体的慢病毒载体可用于实施本发明。因此,本文所用的术语“逆转录病毒”或者“逆转录病毒载体”是指分别包括“慢病毒”和“慢病毒载体”。已经对逆转录病毒载体进行测试,并发现是将目标基因稳定地导入广范围的靶细胞的基因组中的合适的递送媒介物。逆转录病毒载体将未重排的单拷贝转基因递送进细胞中的能力,使得逆转录病毒载体非常适合将基因转移进细胞中。另外,通过逆转录病毒包膜糖蛋白与宿主细胞上的特定细胞表面受体的结合,逆转录病毒进入宿主细胞中。结果,假型逆转录病毒载体也可以用于实践本发明,在所述假型逆转录病毒载体中,编码的天然包膜蛋白被异源包膜蛋白替代,所述异源包膜蛋白具有与天然包膜蛋白不同的细胞特异性(例如与天然包膜蛋白相比,结合不同的细胞表面受体)。指导编码一种或多种靶蛋白编码序列的逆转录病毒载体向特定靶细胞递送的能力,是实践本发明所需的。

[0194] 本发明提供了逆转录病毒载体,其包括例如:包含一个或多个转基因序列的逆转录病毒转移载体,及包含一个或多个包装元件的逆转录病毒包装载体。具体地,本发明提供了编码异源的或者功能修饰的包膜蛋白的假型逆转录病毒载体,其用于生产假型逆转录病毒。

[0195] 本发明的逆转录病毒载体的核心序列可容易地源自众多的逆转录病毒,包括例如 B、C 和 D 型逆转录病毒以及泡沫病毒和慢病毒(参见 RNA Tumor Viruses,第 2 版, Cold Spring Harbor Laboratory, 1985)。适用于本发明的组合物和方法中的逆转录病毒的实例包括、但不限于:慢病毒。适用于本发明的组合物和方法中的其它逆转录病毒包括、但不限于:禽类白血病病毒、牛白血病病毒、鼠白血病病毒、貂致细胞灶病毒、鼠肉瘤病毒、网状内皮组织增殖病毒和鲁斯肉瘤病毒。特别优选的鼠白血病病毒包括 4070A 和 1504A(Hartley 和 Rowe. 1976. J. Virol. 19 :19-25)、Abelson(ATCC 号 VR-999)、Friend(ATCC 号 VR-245)、Graffi、Gross(ATCC 号 VR-590)、Kirsteni Harvey 肉瘤病毒和 Rauscher(ATCC 号 VR-998)和莫洛尼鼠白血病病毒(ATCC 号 VR-190)。这种逆转录病毒可从诸如美国典型培养物保藏中心(ATCC ;Manassas, VA)的保藏物或收集库容易地得到,或者使用常用技术从已知来源分离。其它来源是可商业获得的。

[0196] 在一个实施方案中,本发明的逆转录病毒载体序列可以源自慢病毒。优选的慢病毒是人免疫缺陷病毒,例如 1 型或者 2 型(即 HIV-1 或者 HIV-2,其中 HIV-1 以前被称作淋巴结病相关病毒 3(HTLV-III)及获得性免疫缺陷综合征(AIDS)相关病毒(ARV)),或者与 HIV-1 或 HIV-2 相关的、已经被修饰并与 AIDS 或 AIDS 样疾病相关的另一种病毒。其它慢病毒包括绵羊脱髓鞘性脑白质炎/梅迪病毒、猫免疫缺陷病毒(FIV)、牛慢病毒、猿猴免疫缺陷病毒(SIV)、马感染性贫血病毒(EIAV)及山羊关节炎-脑炎病毒(CAEV)。

[0197] 合适的逆转录病毒属和株是本领域众所周知的（参见，例如，Fields Virology，第 3 版，B.N.Fields 等人编，1996.Lippincott-Raven Publishers，参见例如，第 58 章，Retroviridae: The Viruses and Their Replication, Classification, 第 1768-1771 页，包括其中的表 1，通过引用并入本文）。用于制备生产逆转录病毒的生产细胞和生产细胞系的逆转录病毒包装系统以及制备这样的包装系统的方法，也是本领域已知的。

[0198] 典型的包装系统包含至少两个包装载体：第一个包装载体，其包含第一个核苷酸序列，该序列包含 gag、pol 或 gag 和 pol 基因；和第二个包装载体，其包含第二个核苷酸序列，该序列包含异源的或者功能修饰的包膜基因。逆转录病毒元件可以源自慢病毒，如 HIV。所述载体可以缺少功能性 tat 基因和 / 或功能性附加基因 (vif、vpr、vpu、vpx、nef)。所述系统可以另外包含第三个包装载体，其具有包含 rev 基因的核苷酸序列。所述包装系统可以以含有所述第一个、第二个及任选的第三个核苷酸序列的包装细胞的形式提供。

[0199] 在一些实施方案中，可适用于多种表达系统，特别是具有真核细胞和有利的哺乳动物细胞的那些。在天然蛋白被糖基化的情况下，优选的实施方案可以包含为表达的蛋白提供天然 - 样糖基化的表达系统。

[0200] 慢病毒具有一些共有的结构性病毒粒子蛋白，包括：包膜糖蛋白 SU(gp 120) 和 TM(gp41)，它们由 env 基因编码；CA(p24)、MA(p17) 和 NC(p7-11)，它们由 gag 基因编码；及由 pol 基因编码的 RT、PR 和 IN。HIV-1 和 HIV-2 含有参与调节病毒 RNA 的合成和加工及其它复制功能的辅助蛋白质和其它蛋白质。由 vif、vpr、vpu/vpx 和 nef 基因编码的辅助蛋白质可以在重组系统中省略（或者灭活）。另外，tat 和 rev 可例如通过突变或者缺失而省略或者灭活。

[0201] 第一代慢病毒载体包装系统为 gag/pol 和 env 提供了单独的包装构建体，并且出于安全原因，通常采用异源的或者功能修饰的包膜蛋白。在第二代慢病毒载体系统中，附加基因 vif、vpr、vpu 和 nef 被缺失或者灭活。第三代慢病毒载体系统是其中 tat 基因已经缺失或者以其它方式灭活（例如通过突变）的那些。

[0202] 补偿通常由 tat 提供的转录调节，可以通过使用强组成型启动子而提供，如人巨细胞病毒立即早期 (HCAAV-IE) 增强子 / 启动子。基于组成型启动子活性的强度、对靶组织的特异性（例如，肝特异性的启动子）或者与希望的表达控制有关的其它因素，可以选择其它启动子 / 增强子，这些为本领域所已知。例如，在一些实施方案中，采用诱导型启动子如 tet 来实现受控表达是合乎需要的。编码 rev 的基因可以提供在单独的表达构建体上，所以典型的第三代慢病毒载体系统将包含四个质粒：gagpol、rev、包膜和转移载体各一个。无论采用哪一代包装系统，gag 和 pol 均可以提供在单个构建体上或者分开的构建体上。

[0203] 通常，包装载体被包含在包装细胞中，并通过转染、转导或者感染而导入细胞中。进行转染、转导或者感染的方法为本领域技术人员所熟知。本发明的逆转录病毒转移载体可以通过转染、转导或者感染而导入包装细胞系中，以产生生产细胞或者细胞系。本发明的包装载体可以通过标准方法导入人细胞或者细胞系中，所述标准方法包括例如：磷酸钙转染、脂转染或电穿孔。在一些实施方案中，包装载体与显性选择标记（诸如 neo、二氢叶酸还原酶 (DHFR)、谷氨酰胺合成酶或 ADA）一起导入细胞中，随后在合适药物存在下选择，并分离克隆。选择标记基因可以与由包装载体编码的基因物理连接。

[0204] 已知其中包装功能设计为由合适的包装细胞表达的的稳定细胞系。例如，参见美国

专利号 5,686,279 ;和 Ory 等人 1996. Proc. Natl. Acad. Sci. 93 :11400-11406, 它们描述了包装细胞。关于稳定的细胞系生产的其它描述,可以参见 Dull 等人 1998. J. Virol. 72(11) : 8463-8471 ;和 Zufferey 等人 1998. J. Virol. 72 :9873-9880。

[0205] Zufferey 等人 1997. Nat. Biotechnol. 15 :871-75 教导了一种慢病毒包装质粒,其中包括 HIV-1 包膜基因的 pol 的 3' 序列缺失。该构建体含有 tat 和 rev 序列,并且 3' LTR 被聚腺苷酸序列替代。5' LTR 和 psi 序列被另一种启动子(诸如诱导型启动子)替换。例如,可以使用 CMV 启动子或其衍生物。

[0206] 包装载体可以含有另外的包装功能改变,以增强慢病毒蛋白表达及增加安全性。例如,可以除去在 gag 上游的所有 HIV 序列。也可以除去包膜的下游序列。另外,可以采用修饰载体以增强 RNA 剪接和翻译的步骤。

[0207] 任选地,使用条件包装系统,诸如 Dull 等人 1998(同上)描述的包装系统。也优选地使用自身灭活载体(SIN),其通过缺失 HIV-1 长末端重复序列(LTR)而提高载体的生物安全性,例如 Zufferey 等人 1998. J. Virol. 72 :9873-9880 所述。也可以使用诱导型载体,诸如通过四环素可诱导的 LTR。

[0208] 启动子

[0209] 在一些实施方案中,本发明的载体通常包括异源控制序列,其包括、但不限于:组成型启动子,诸如巨细胞病毒(CMV)立即早期启动子、RSV LTR、MOMLV LTR 和 PGK 启动子;组织或细胞类型特异性的启动子,包括 mTTR、TK、HBV、hAAT、可调节的或可诱导的启动子、增强子等。

[0210] 某些有用的启动子包括:LSP 启动子(III 等人 1997. Blood Coagul. Fibrinolysis 8S2 :23-30)、EF1- α 启动子(Kim 等人 1990. Gene 91(2) :217-23)和 Guo 等人 1996. Gene Ther. 3(9) :802-10)。最优的启动子包括延伸因子 1- α (EF1a) 启动子、磷酸甘油酸激酶-1 (PGK) 启动子、巨细胞病毒立即早期基因(CMV)启动子、嵌合的肝特异性的启动子(LSP)、巨细胞病毒增强子/鸡 β -肌动蛋白(CAG)启动子、四环素应答启动子(TRE)、转甲状腺素蛋白启动子(TTR)、猴病毒 40(SV40)启动子和 CK6 启动子。可用于实践本发明的一种有利的启动子是腺病毒主要晚期启动子(Berkner 和 Sharp. 1985. Nucl. Acids Res. 13 : 841-857)。有关启动子的结构和功能信息是本领域已知的。相关序列可从公共数据库容易地得到,并掺入载体中,用于实践本发明的方面。

[0211] 在本发明的实践中一种特别优选的启动子是腺病毒主要晚期启动子。表达盒可以包含:在 5' 至 3' 方向,腺病毒主要晚期启动子、与目标蛋白或目标蛋白链的第一个编码序列可操作地连接的三联体前导序列、编码自加工序列或蛋白酶切割序列的序列、目标蛋白或蛋白链的第二个编码序列、和任选的编码自加工序列或蛋白酶切割序列的序列、随后的目标蛋白或蛋白链的第三个编码序列。所有这些编码序列共价地连接在同一个读码框中,使得在多蛋白编码序列内不终止翻译。在蛋白合成过程中或在多肽合成结束以后,自加工或蛋白水解性加工会将多蛋白切割成适当的蛋白链或蛋白。在免疫球蛋白合成的情况下,轻链的编码序列在多蛋白编码序列内出现 2 次。有利地,前导序列编码区可以与蛋白或蛋白链序列相结合;信号肽酶的加工可以具有下述额外益处:在加工位点下游的蛋白的 N-端处,去除某种残余的氨基酸残基。免疫球蛋白重链的组分是:Met, 蛋白起始甲硫氨酸;HC, 重链;LC, 轻链;SPPC, 自加工或蛋白酶切割位点。免疫球蛋白合成的表达构建体可

以包括下述的:Met-蛋白酶-SPPC-HC前导序列-HC-SPPC-LC前导序列-LC-SPPC-LC前导序列-LC;Met-蛋白酶-SPPC-LC前导序列-LC-SPPC-LC前导序列-LC-SPPC-HC前导序列-HC;Met-蛋白酶-SPPC-LC前导序列-LC-SPPC-HC前导序列-HC-SPPC-LC前导序列-LC;HC前导序列-HC-SPPC-LC前导序列-LC-SPPC-LC前导序列-LC;LC前导序列-LC-SPPC-HC前导序列-HC-SPPC-LC前导序列-LC;LC前导序列-LC-SPPC-LC前导序列-LC-SPPC-HC前导序列-HC;Met-蛋白酶-SPPC-HC前导序列-HC-SPPC-LC前导序列-LC。

[0212] 包括抗体在内的生物治疗分子

[0213] 在本发明的范围内,除了别的以外,特定表达的抗体(免疫球蛋白)可以包括特异性地结合下述因子的那些:肿瘤坏死因子(与HUMIRA/D2E7相对应和/或源自它的工程化的抗体;所述HUMIRA/D2E7是百慕大哈密尔顿的Abbott Biotechnology Ltd.的阿达木单抗的商标);白介素-12(源自ABT-874的工程化的抗体);白介素-18(源自ABT-325的工程化的抗体);重组促红细胞生成素受体(源自ABT-007的工程化的抗体);或E/L选择蛋白(源自EL246-GG的工程化的抗体)。工程化的多蛋白的编码序列和氨基酸序列公开在本文中,或是本领域可得到的。适用于本发明的其它抗体包括,例如,类克(英夫利昔单抗);Rituxan/美罗华(利妥昔单抗);赫赛汀(曲妥单抗);阿瓦斯丁(贝伐单抗);Synagis(帕利珠单抗);爱必妥(西妥昔单抗);Reopro(阿昔单抗);Orthoclone OKT3(莫罗单抗-CD3);赛尼哌(达珠单抗);舒莱(巴利昔单抗);麦罗塔(吉姆单抗);Campath(阿仑珠单抗);泽娃灵(替伊莫单抗);Xolair(奥马珠单抗);Bexxar(托西莫单抗);和Raptiva(依法珠单抗);其中在商标-品牌名称之后通常在括号中注明各自的通用名。其它合适的蛋白包括,例如,下述的一种或多种:红细胞生成素 α 、红细胞生成素 β 、依那西普、达依泊汀 α 、非格司亭、干扰素 β 1a、干扰素 β 1b、干扰素 α -2b、甘精胰岛素、生长激素、特立帕肽、促滤泡素 α 、脱氧核糖核酸酶(dornase)、因子VIII、因子VII、因子IX、伊米苷酶、奈西立肽、来格司亭和Von Willebrand因子;其中一个或多个通用名可以各自对应产物的一个或多个商标-品牌名称。其它抗体和蛋白适用于本发明,如本领域所理解的。

[0214] 本发明也预见到两种或更多种目标多肽或蛋白或前蛋白的编码序列的受控表达。基因调节系统可用于调节特定一个或多个基因的表达。在一个示例性的方案中,基因调节系统或开关包括嵌合的转录因子,其具有配体结合结构域、转录激活结构域和DNA结合结构域。所述结构域可从基本上任何来源得到,并且可以以许多方式中的任一种相组合,以获得新的蛋白质。可调节的基因系统还包括DNA应答元件,其与嵌合的转录因子相互作用。该转录调控元件位于要调节的基因邻近。

[0215] 可用于实践本发明的示例性的转录调节系统包括,例如:果蝇蜕皮激素系统(Yao等人1996.Proc.Natl.Acad.Sci.93:3346);蚕蜕皮激素系统(Suhr等人1998.Proc.Natl.Acad.Sci.95:7999);GeneSwitch(Valentis, The Woodlands, TX的商标);合成的孕酮受体系统,其采用RU486作为诱导物(Osterwalder等人2001.Proc.Natl.Acad.Sci.USA 98(22):12596-601);Tet和RevTet系统(四环素调节的表达系统,BD Biosciences Clontech, Mountain View, CA的商标),其采用小分子如四环素(Tc)或类似物例如多西环素来调节(启动或关闭)靶标的转录(Knott等人2002.Biotechnologys 32(4):796,798,800);ARIAD调节技术(Ariad, Cambridge, MA),其基于使用小分子使两个细胞内分子结合在一起,每个分子均与转录活化剂或DNA结合蛋白相连。当这些组分集合在一起时,会激活

目标基因的转录。Ariad 具有基于同源二聚化的系统和基于异源二聚化的系统 (Rivera 等人 1996. *Nature Med.* 2(9):1028-1032; Ye 等人 2000. *Science* 283:88-91)。

[0216] 本发明的表达载体构建体的实施方案 (其包含编码自加工或蛋白酶切割的重组多肽形式的抗体或其片段或其它异源蛋白或前蛋白的核酸序列) 可以在体外、先体外后体内或在体内导入细胞中, 用于向细胞 (例如, 体细胞) 递送外源的治疗性基因或转基因, 或用于由载体转导的细胞生产重组多肽。

[0217] 宿主细胞和载体的递送

[0218] 使用本领域已知的标准方法, 可以将本发明的载体构建体在体外或者先体外后体内导入合适的细胞中。这种技术包括, 例如: 使用磷酸钙的转染、显微注射进培养的细胞中 (Capecchi. 1980. *Cell* 22:479-488)、电穿孔 (Shigekawa 等人 1988. *Biotechnology* 6:742-751)、脂质体介导的基因转移 (Mannino 等人 1988. *Biotechnology* 6:682-690)、脂质介导的转导 (Feigner 等人 1987. *Proc. Natl. Acad. Sci. USA* 84:7413-7417) 及使用高速微射弹进行的核酸递送 (Klein 等人 1987. *Nature* 327:70-73)。

[0219] 对于体外或者先体外后体内表达, 可以采用有效地表达功能性蛋白产物的任何细胞。本领域已知许多用于蛋白质表达的细胞和细胞系的实例。例如, 原核细胞和昆虫细胞可用于表达。另外, 可以使用真核微生物, 如酵母。重组蛋白质在原核、昆虫和酵母系统中的表达为本领域所通常已知的, 并且可以适用于使用本发明的组合物和方法的抗体或其它蛋白表达。

[0220] 可用于表达的细胞的实例另外包括: 哺乳动物细胞诸如成纤维细胞、来自非人哺乳动物的细胞诸如羊、猪、鼠和牛细胞、昆虫细胞等。哺乳动物细胞的具体实例包括、但不限于: COS 细胞、VERO 细胞、HeLa 细胞、中国仓鼠卵巢 (CHO) 细胞、CHO DX B11 细胞、CHO DG44 细胞、PerC. 6 细胞、Sp2/0 细胞、293 细胞、NSO 细胞、3T3 成纤维细胞、W138 细胞、BHK 细胞、HEPG2 细胞和 MDCK 细胞。

[0221] 在经改良的常规营养培养基中培养宿主细胞, 所述改良是为了适合诱导启动子、选择转化体或者扩增编码希望的序列的基因。可以在多种培养基中培养哺乳动物宿主细胞。可商业得到的培养基诸如 Ham' s F10 (Sigma)、最低必需培养基 (MEM) (Sigma)、RPMI 1640 (Sigma)、最低基础培养基 (MEM) Alpha 培养基和 Dulbecco 氏改良的伊格尔培养基 (DMEM) (Sigma) 通常适用于培养宿主细胞。给定的培养基通常根据需要补加激素和 / 或其它生长因子 (如胰岛素、转铁蛋白或表皮生长因子)、盐 (如氯化钠、钙、镁和磷酸盐)、缓冲液 (如 HEPES)、核苷 (如腺苷和胸苷)、抗生素、痕量元素和葡萄糖或者等效的能量源。也可以包含适当浓度的任何其它必需补充物, 这些为本领域技术人员所已知。对于特定细胞系而言合适的培养条件 (如温度、pH 等) 通常为本领域所已知, 为多种细胞系培养建议的培养条件参见例如 ATCC 目录 (可在因特网上在 "atcc.org/SearchCatalogs/AllCollections.cfm" 得到), 或按照商业供应商的指示。

[0222] 所述表达载体可以通过不同途径在体内施用 (例如真皮内、静脉内、肿瘤内、脑内、门静脉内、腹膜内、肌肉内、膀胱内等), 以递送经由自加工切割序列连接的多个基因, 从而在动物模型或人受试者中表达两种或更多种蛋白或多肽。根据施用途径, 治疗性蛋白质局部地 (在脑或膀胱中) 或者全身地 (其它施用途径) 发挥其作用。在开放读框的 5' 侧的组织特异性启动子的使用, 会导致由该整个开放读码框编码的蛋白质或多肽的组织特异

性表达。

[0223] 将携带转基因的重组表达载体在体外、先体外后体内或者在体内导入靶细胞中的多种方法,在先前已经描述并且为本领域所熟知。本发明提供了治疗方法、疫苗和癌症疗法,其中用含有 2 种或更多种目标蛋白或多肽的编码序列的重组载体感染靶细胞,并在靶细胞中表达所述蛋白或多肽。

[0224] 例如,本发明的重组载体的体内递送可以靶向多种器官类型,包括、但不限于:脑、肝、血管、肌肉、心、肺和皮肤。

[0225] 在先体外后体内基因转移的情况下,使用本发明的重组载体和本领域熟知的方法,从宿主取出靶细胞,并在实验室中进行遗传修饰。

[0226] 使用常规施用模式,包括、但不限于上述模式,可以施用本发明的重组载体。本发明的重组载体可以在多种制剂中,所述制剂包括、但不限于:液体溶液和悬浮液、微泡、脂质体和可注射的或可输注的溶液。优选的形式依赖于施用模式和治疗用途。

[0227] 在一些实施方案中,表达载体构建体在免疫球蛋白或其它生物活性蛋白体内生产中的优点包括:在患者中长期施用单个载体和持续的抗体表达;具有完整生物活性的抗体或其片段(或其它生物活性蛋白)的体内表达;和在入细胞中产生的抗体的天然的翻译后修饰。理想地,表达的蛋白与天然存在的蛋白是相同的或充分相同的,使得在将表达的蛋白施用于多个场合或在需要所述蛋白的患者中持续表达的情况下,减少或不触发免疫应答。

[0228] 本发明的重组载体构建体的实施方案可以另外用于在体外生产重组抗体和其它生物活性蛋白,它们用于治疗中或用于研究中。重组蛋白的生产方法是本领域众所周知的,且可以用于表达重组抗体,其中使用本文所述的含有自加工切割位点或其它蛋白酶切割位点的载体构建体。

[0229] 在一个方面,本发明提供了用于生产重组免疫球蛋白或其片段的方法,其中将表达载体(诸如上述的表达载体)导入细胞中,以得到转染的细胞,其中所述载体在 5' 至 3' 方向包含:与免疫球蛋白重链和 2 个轻链或其片段的编码序列可操作地连接的启动子,在每个所述链之间的自加工序列。应当理解,免疫球蛋白重链的编码序列或免疫球蛋白轻链的编码序列可以相对于给定的载体构建体中的自加工序列在 5' 侧(即,前面)。

[0230] 在抗体的构建体的一个实施方案中,编码抗体或免疫球蛋白或其片段的第一链或第二链的序列包括:源自 IgG、IgM、IgD、IgE 或 IgA 的重链或其片段。概括地讲,编码抗体或免疫球蛋白或其片段的链的序列也包括来自 IgG、IgM、IgD、IgE 或 IgA 的轻链或其片段。本发明的实施方案涉及与整个抗体分子的蛋白以及它们的修饰的或衍生的形式相对应的基因,所述形式包括,例如,其它抗原识别分子片段如 Fab、单链 Fv(scFv) 和 F(ab')₂。所述抗体和片段可以是动物衍生的、人-小鼠嵌合的、人源化的、通过去免疫化(Deimmunisation)[™](Biovation Ltd) 改变的、为了改变对 Fc 受体的亲和力而改变的、或全人的。配体结合分子的实施方案可以是亲和力成熟的,如本领域理解的。在优选的实施方案中,抗体或其它重组蛋白在施用它的人或动物中不会引起或最小地引起免疫应答。

[0231] 所述抗体可以是双特异性的,且包括、但不限于:双抗体,细胞杂交瘤,小型抗体,ScBs 抗体和把手在孔中(knobs-into-holes)抗体。

[0232] 以本领域众所周知的不同方式(Harlow 等人 1988. 抗体, A Laboratory Manual, Cold Spring Harbor Laboratory),可以实现抗体本身的生产和回收。根据本领域众所周

知的方法,收集和 / 或纯化和 / 或使用其它目标蛋白。

[0233] 在实践本发明的实施方案中,通过在适合宿主细胞生长和编码序列表达的培养条件下培养修饰的重组宿主细胞,可以使用重组 DNA 技术生产抗体或其变体(类似物)。为了监测表达的成功,使用标准技术诸如 ELISA、RIA 等,可以监测与抗原有关的抗体水平。使用本领域已知的标准技术,从培养物上清液中回收抗体。当然,通过标准的纯化技术,包括、但不限于经由蛋白 A、蛋白 G 或蛋白 L 柱的亲色谱法,或关于特定抗原,或甚至关于希望具有特异性的抗原的特定表位,可以容易地制备这些抗体的纯化形式。使用常规色谱法,诸如离子交换、疏水相互作用、亲和力或尺寸排阻柱,也可以如本领域理解地纯化抗体。也参见 Rinderknecht 等人于 1997 年 6 月 24 日提交的 US 5,641,870,“Low pH hydrophobic interaction chromatography for antibody purification”和 Shukla 等人于 2008 年 9 月 23 日提交的 US 7,427,659,“Process for purifying proteins in a hydrophobic interaction chromatography flow-through fraction”,它们公开了纯化技术。纯化技术可以以不同的组合或与其它技术(诸如硫酸铵沉淀和尺寸限制的膜过滤)结合进行。在将表达系统设计成包括信号肽的情况下,得到的抗体会分泌进培养基或上清液中;但是,细胞内生产也是可能的。可以回收细胞内的内容物,并进行纯化。

[0234] 可以选择细胞培养条件,所述条件会促进希望的多蛋白加工水平,例如,最完全加工。这样的加工可以发生在细胞内,但是也可以发生在细胞外、在细胞培养过程中或以后。

[0235] 以前已经描述了来自用人 Ig 基因座工程化的小鼠的抗原特异性的全人单克隆抗体的生产和选择(Jakobovits 等人 1998. *Advanced Drug Delivery Reviews* 31:33-42; Mendez 等人 1997. *Nature Genetics* 15:146-156; Jakobovits 等人 1995. *Curr Opin Biotechnol* 6:561-566; Green 等人 1994. *Nature Genetics* Vol. 7:13-21)。

[0236] 已经在转基因山羊的乳汁中实现治疗性单克隆抗体的高水平表达,并已经证实,抗原结合水平与使用常规细胞培养技术生产的单克隆抗体的水平等同。该方法是基于人治疗蛋白在转基因动物的乳汁中的形成,所述转基因动物携带允许它们在它们的乳汁中表达人治疗蛋白的遗传信息。这些重组蛋白一旦生成,使用标准的技术可以从乳汁中有效地纯化它们。参见例如,Pollock 等人 1999. *J. Immunol. Meth.* 231:147-157 和 Young 等人 1998. *Res Immunol.* 149(6):609-610。来自转基因动物的动物乳汁、卵白、血液、尿、精浆和蚕虫茧已经表现出作为在工业规模生产重组蛋白的来源的潜力(Houdebine L M. 2002. *Curr Opin Biotechnol* 13:625-629; Little 等人 2000. *Immunol Today*, 21(8):364-70; 和 Gura T. 2002. *Nature*, 417:584-5860)。本发明预见到转基因动物表达系统用于表达目标重组抗体或其变体(类似物)或其它蛋白的用途,其中使用本发明的编码自加工切割位点的载体和 / 或蛋白酶识别位点载体。

[0237] 还已经成功地证实重组蛋白在植物中的生产,所述植物包括、但不限于:通过土壤杆菌感染、基因枪转化、原生质体转化等转化的马铃薯、番茄、烟草、稻和其它植物。已经证实重组人 GM-CSF 在转基因烟草植物的种子中的表达和包括单链抗体在内的抗体在植物中的表达。参见,例如,Streafield 和 Howard. 2003. *Int. J. Parasitol.* 33:479-93; Schillberg 等人 2003. *Cell Mol Life Sci.* 60:433A5; Pogue 等人 2002. *Annu. Rev. Phytopathol.* 40:45-74; 和 McCormick 等人 2003. *J Immunological Methods*, 278:95-104。本发明预见到转基因的植物表达系统用于表达目标重组免疫球蛋白或其片段或其它蛋白

的用途,其中使用本发明的编码蛋白酶切割位点或自加工切割位点的载体。

[0238] 与昆虫细胞相结合的杆状病毒载体表达系统也正在成为重组蛋白生产的可行平台的基础。已经报道,杆状病毒载体表达系统会提供相对于哺乳动物细胞培养的优点,诸如容易培养和更高的表达水平。参见,例如,Ghosh 等人 2002. *Mol Ther.* 6 :5-11,和 Ikonomou 等人 2003. *Appl Microbiol Biotechnol.* 62 :1-20。本发明另外预见到杆状病毒载体表达系统用于表达重组免疫球蛋白或其片段的用途,其中使用本发明的编码自加工切割位点的载体。杆状病毒载体和合适的宿主细胞是本领域众所周知的和可商业得到的。

[0239] 基于酵母的系统也可以用于表达目标重组免疫球蛋白或其片段或其它蛋白,包括双杂种或三杂种系统,其中使用本发明的编码自加工切割位点的载体。参见,例如,美国专利号 5,643,745,它通过引用并入本文。

[0240] 应当理解,本发明的表达盒和载体和重组宿主细胞(其包含单独的自加工肽的编码序列,或与蛋白水解性切割位点的额外编码序列相组合的自加工肽的编码序列)可以用于在任意蛋白表达系统中表达双杂种和三杂种系统的重组免疫球蛋白或其片段、前蛋白、生物活性蛋白和蛋白组分,它们中的许多是本领域已知的,它们的实例描述在本文中。本领域技术人员可以容易地改变本发明的载体、宿主细胞和方法的实施方案,用于任意蛋白表达系统中。

[0241] 实施例 1. Lon 蛋白酶内含肽和表达构建体

[0242] 在新英格兰 Biolabs (NEB, Ipswich, Massachusetts, USA) 内含肽数据库 (InBase, The Intein Database and Registry; 在 <http://www.neb.com/neb/inteins.html>) 中报道了 3 种 ATP 依赖性的 Lon 蛋白酶内含肽。参见 Perler, F. B. (2002). InBase, the Intein Database. *Nucleic Acids Res.* 30,383-384。这些内含肽来自生物体 *Pyrococcus abyssi* (Pab Lon 内含肽)、激烈火球菌 (Pfu Lon 内含肽) 和掘越氏火球菌 OT3 (Pho Lon 内含肽)。这些 Lon 内含肽具有提出的内切核酸酶结构域、赖氨酸替代组氨酸作为内含肽的倒数第二位残基、和不同的长度(分别是 333、401 和 474 个氨基酸)。在 NEB 数据库中,所有 3 种 lon 内含肽被指示为理论内含肽,根据数据库指示,列出的贡献者没有指示,已经证实给定的内含肽项目的剪接产物的存在。应当指出,已经在实验中发现 Pab Lon 内含肽的内切核酸酶结构域不具有活性 (Saves I, Morlot C, Thion L, Rolland JL, Dietrich J, Masson JM. Investigating the endonuclease activity of four *Pyrococcus abyssi* inteins. *Nucleic Acids Res.* 2002 Oct 1 ;30(19) :4158-65)。

[0243] 我们已经发现,被包含在 ATP 依赖性的蛋白酶 lon 家族的基因中的内含肽可以非常有效地介导在不同单个开放读码框构建体设计中的抗体重链和轻链的切割。一提供了与这些内含肽有关的序列信息。

[0244] Lon 内含肽序列和载体构建体设计

[0245] 表 1 提供了 Pab Lon 内含肽(在 NCBI/ 蛋白中的登记号 CAB50486. 1)、PAB1313 Pab Lon 内含肽(包括 -1 和 +1 外显肽残基)的蛋白序列信息 (SEQ ID NO :1)。

[0246] 表 1. Pab Lon 内含肽氨基酸序列 (SEQ ID NO :1)。

[0247] QCFSGEETVIRENGEVKVLRLKDFVEKALEKPSGEGLDGDVKVYHDFRNENVEVLTKDGFTKLLYANK

[0248] RIGKQKLRRVNVLEKDYWFALTPDHKVYTTDGLKEAGEITEKDELISVPITVFDCEDEDLKKIGLLPLTS

[0249] DDERLRKIATLMGILFNGGSIDGLGVLTLKSERSVIEKFVITLKELFGKFYEYIIEENTILKTRDPRI

- [0250] IKFLVGLGAPIEGKDLKMPWWVKLKPSLFLAFLEGFRAHIVEQLVDDPNKNLPFFQELSWYLGLFGIKAD
- [0251] IKVEEVGDKHKIIFDAGRLDVKQFIETWEDVEVTYNLTTEKGNLLANGLFVKNS
- [0252] 表 2 描述了已经进行密码子选择优化的 Pab Lon 内含肽的核苷酸序列。
- [0253] 表 2. Pab Lon 内含肽核苷酸序列 (SEQ ID NO :2)。
- [0254] tgcttcagcggcgaggaaaccgtggtgatccgggagaacggcgaggtgaaggtgctgcggct
- [0255] gaaggacttcgtggagaaggccctggaaaagccctccggcgagggcctggacggcgacgtga
- [0256] aagtgggtgtaccacgacttcggaacgagaacgtggaggtgctgaccaaggacggcttcacc
- [0257] aagctgctgtacgccaacaagcggatcggcaagcagaaactgcggcggggtggtgaacctgga
- [0258] aaaggactactggttcgccctgacccccgaccacaaggtgtacaccaccgacggcctgaaag
- [0259] aggccggcgagatcaccgagaaggacgagctgatcagcgtgcccataccctgttcgactgc
- [0260] gaggacgaggacctgaagaagatcggcctgctgccccctgaccagcgacgacgagcggctgcg
- [0261] gaagatcgccaccctgatgggcatcctgttcaacggcggcagcatcgatgagggcctgggcg
- [0262] tgctgacctgaagagcgagcggagcgtgatcgagaagtctcgatcacctgaaagagctg
- [0263] ttccgcaagtctcgagtacgagatcatcaaagaggaaaacaccatcctgaaaacccgggaccc
- [0264] ccggatcatcaagtcttctggtgggcctgggagcccccatcgagggaaggtatctgaagatgc
- [0265] cttggtgggtgaagctgaagcccagcctgttcctggccttcctggaaggcttcggggccac
- [0266] atcgtggagcagctggctgacgacccccacaagaatctgcccttctttcaggaaactgagctg
- [0267] gtatctgggcctgttcggcatcaaggccgacatcaaggtggaggaagtgggcgacaagcaca
- [0268] agatcatcttcgacgcggcgaggctggacgtggacaagcagttcatcgagacctgggaggat
- [0269] gtggaggtgacctacaacctgaccacagagaagggaatctgctggccaacggcctgttcgt
- [0270] gaagaac
- [0271] 表 3 描述了由 SEQ ID NO :2 编码的蛋白序列 SEQ ID NO :3。
- [0272] 表 3. Pab Lon 内含肽氨基酸序列 (SEQ ID NO :3)。
- [0273] CFSGEETVVIRENGEVKVLRLKDFVEKALEKPSGEGLDGDVKVYHDFRNENVEVLTKDGF
- [0274] KLLYANKRIGKQKLRRVVNLEKDYWFALTPDHKVYTTDGLKEAGEITEKDELSVPITVFD
- [0275] EDEDLKKIGLLPLTSDDERLRKIATLMGILFNNGSIDGLGLVTLKSERSVIEKFVITLKE
- [0276] FGKFEYEIIKEENTILKTRDPRIKFLVGLGAPIEGKDLKMPWWVKLKPSLFLAFLEGFRAH
- [0277] IVEQLVDDPNKNLPFFQELSWYLGLFGIKADIKVEEVGDKHKIIFDAGRLDVKQFIETWED
- [0278] VEVTYNLTTEKGNLLANGLFVK
- [0279] 表 4 提供了 Pfu Lon 内含肽 (在 NCBI/ 蛋白中的登记号 AAL80591. 1)、PF0467 (包
- 括 -1 和 +1 外显肽残基) 的蛋白序列信息 (SEQ ID NO :4)。
- [0280] 表 4. Pfu Lon 内含肽氨基酸序列 (SEQ ID NO :4)。
- [0281] QCFSGEEVILIEKDGEKKVFKLREFVDGLLKEASGEGMDGSIRVVYKDLQGENIKILTKDGLVKLLYV
- [0282] NRREGKQKLKIVNLEKDYWLALTPEHKVYTIKGLKEAGEITKDDEIIRVPLTILDGFDVAEKSIRE
- [0283] LERLSLLPLNSEDRLKIAIGIMGALFGSGGIDENLNTLSFVSSEKKTIEQFVKALSELFGEFDYKIE
- [0284] EKENSIIFRTCDKRIVTFFATLGAPVGDKSKVKLKPWWVKLKPSLFLAFMDGLYSSNRNDKEILEIT
- [0285] QLTDNVETTFEEISWYLSFFGIKAEAEDEEKDKYRARLTLSSSIDNMLNFIEFIPISFSPAKREKFF
- [0286] KEIEKYLEYSIPEKTEDLKKRVKRVKKGERRNFLESWEEVEVTYNVTETGNLLANGLFVKNS
- [0287] 表 5 提供了天然 Pfu Lon 内含肽的核苷酸序列信息 (SEQ ID NO :5)。

[0288] 表 5. Pfu Lon 内含肽核苷酸序列 (SEQ ID NO :5)。

[0289] tgttttagcggtgaagaagttatcttaattgaaaaggacggagagaaaaaagtcttcaaact
[0290] tagggagttcgttgacgggtctccttaaggaggcgtctggagaagggatggacggaagtatta
[0291] gagtagttttataaagatcttcaaggggaaaacataaaaatactcacaaaagacggacttgta
[0292] aagctcctttatgtcaatagaagagaagggaagcaaaagcttagaaaaatagtaaactcttga
[0293] aaaggattatttggttgcatataacacctgaacataaagtgtacacaataaagggccttaag
[0294] aagctggagagataactaaagatgatgagataataagagtgcctctcacaaattcttgacggc
[0295] tttgacgtagccgagaagagtataagagaggaacttgaaaggcttagcctacttccactaaa
[0296] tagtgaagacagtagactagaaaagatagcaggaatcatgggcgcactctttggtagtggag
[0297] gtatcgatgagaatctcaatacccttagctttgtttctagcgagaagaaaacaattgaacag
[0298] tttgttaaagcactcagcgagctcttcggggaatttgactataaaattgaagaaaaagaaaa
[0299] cagcattattttcagaacatgtgataaaagaatagtgccttctttgctacacttggtgcac
[0300] cagttggagacaaaagcaaaagttaagcttaagcttccatggtgggtcaagcttaagccgtca
[0301] cttttcctcgccttcatggatggtctctacagtagcaataggaatgacaaagaaatcctcga
[0302] aataactcaacttactgacaacgtcgaaacgttcttcgaggaaatatcttggtatctgagct
[0303] tcttttgaattaaggcagaagctgaagaggatgaagaaaaagataaaatacagggctagactt
[0304] acgctatcctcatcaatagacaacatgcttaatttcattgagttcattccaataagcttttc
[0305] tccagcaaagagagagaaaaattctttaaggaaattgaaaaatatctggaatatagcattccccg
[0306] aaaagactgaggatcttaagaaacgagttagagagtttaagaaggagagagaaggaatttc
[0307] ctcgaaagctgggaggaagttgaagttacttacaacgtaactacagagacaggaaatctact
[0308] tgctaacggtctatattgttaagaac

[0309] 表 6 描述了由 SEQ ID NO :5 编码的蛋白序列 SEQ ID NO :6。

[0310] 表 6. Pfu Lon 内含肽氨基酸序列。

[0311] CFSGEEVILIEKDGEKKVFKLREFVDGLLKEASGEGMDGSIRVVYKDLQGENIKILTKDGLV
[0312] KLLYVNRREGKQKLRLKIVNLEKDYWLALTPEHKVYTIKGLKEAGEITKDDEIRVPLTILDG
[0313] FDVAEKSIREELERLSLLPLNSEDRLKIAIGIMGALFGSGGIDENLNTLSFVSSEKKTIEQ
[0314] FVKALSELFGEFDYKIEEKENSIIFRTCDKRIVTFFATLGAPVGDKSKVKLKLPPWWKLKPS
[0315] LFLAFMDGLYSSNRNDKEILEITQLTDNVETFFEEISWYLSFFGIKAEAEDEEKDKYRRL
[0316] TLSSSIDNMLNFIEFIPISFSPAKREKFFKEIEKYLEYSIPEKTEDLKKRVKRVKKGERRNF
[0317] LESWEEVEVTYNVTTETGNLLANGLFVKN

[0318] 使用 PCR 技术,克隆 Pfu Lon 内含肽。通过根据哺乳动物密码子选择的设计,合成 Pab Lon 内含肽核苷酸序列。从 Inbase 得到 *Pyrococcus abyssi* lon 蛋白酶内含肽的蛋白序列,所述 Inbase 是由马萨诸塞州 Ipswich 的新英格兰 Biolabs 资助的公众监护的内含肽数据库 (<http://www.neb.com/neb/inteins.html>)。得到的蛋白序列被列出为 EMBL 登录号 CAB50486.1,gi5459000;但是,使用在网站上列出的蛋白序列。在表 7 中指示了 Pab-lon 内含肽蛋白序列。

[0319] 表 7. Pab-lon 内含肽氨基酸序列 (SEQ ID NO :7)。

[0320] CFSGEETVIRENGEVKVLRLKDFVEKALEKPSGEGLDGDVKVYHDFRNENVEVLTKDGFTKLLYANK
[0321] RIGKQKLRRVNVNLEKDYWFALTPDHKVYTTDGLKEAGEITEKDELISVPITVFDCEDEDLKKIGLLPLTS

[0322] DDERLRKIATLMGILFNNGGSIDEGLVLTLSERSVIEKFVITLKELFGKFEYEI I KEENTILKTRDPRI

[0323] IKFLVGLGAPIEGKDLKMPWWVKLKPSLFLAFLEGFRAHIVEQLVDDPNKNLPFFQELSWYLGFLGIKAD

[0324] IKVEEVGDKHKI I FDAGRDLVDKQFIETWEDVEVTYNLTTEKGNLLANGLFVKN

[0325] 使用专有的方法,通过 GeneArt (GeneArt AG, Regensburg, 德国),将该 Pab-lon 蛋白序列回译成为哺乳动物表达优化的 DNA 序列。在表 8 中指出了得到的 Pab-lon 内含肽 DNA 序列。应当理解, DNA 构建体可以任选地提供额外的侧接接头,这取决于特定克隆和表达载体的选择和使用常规技术的对应分子生物学方案。将 DNA 序列合成 (GeneArt 合成编号 0611467) 为 999bp 片段,并在 GeneArt 载体质粒 pGA4 中递送。参见在 <http://partsregistry.org> 上的生物学标准部分 (包括 Part :BBa_J70003) 的登记。对接收的 DNA 材料重新测序,并确定与设计的序列相对应。该 DNA 材料直接地用作以后的含有 Pab-lon 内含肽的质粒构建体的来源 / 模板 ;所述质粒不进行重新繁殖。

[0326] 表 8. Pab-lon 内含肽核酸序列 (SEQ ID NO :8)。

[0327] TGCTTCAGCGGCGAGGAAACCGTGATCCGGGAGAACGGCGAGGTGAAGGTGCTGCGGCT

[0328] GAAGGACTTCGTGGAGAAGGCCCTGGAAAAGCCCTCCGGCGAGGGCCTGGACGGCGACGTGA

[0329] AAGTGGTGTACCACGACTTCCGGAACGAGAACGTGGAGGTGCTGACCAAGGACGGCTTCACC

[0330] AAGCTGCTGTACGCCAACAAAGCGGATCGGCAAGCAGAAACTGCGGCGGGTGGTGAACCTGGA

[0331] AAAGGACTACTGGTTCGCCCTGACCCCCGACCACAAGGTGTACACCACCGACGGCCTGAAAAG

[0332] AGGCCGGCGAGATCACCGAGAAGGACGAGCTGATCAGCGTGCCCATCACCGTGTTCGACTGC

[0333] GAGGACGAGGACCTGAAGAAGATCGGCCTGTGCCCCTGACCAGCGACGACGAGCGGCTGCG

[0334] GAAGATCGCCACCCTGATGGGCATCCTGTTCAACGGCGGCAGCATCGATGAGGGCCTGGGCG

[0335] TGCTGACCCTGAAGAGCGAGCGGAGCGTGATCGAGAAGTTCGTGATCACCTGAAAGAGCTG

[0336] TTCGGCAAGTTCGAGTACGAGATCATCAAAGAGGAAAAACACCATCCTGAAAACCCGGGACCC

[0337] CCGGATCATCAAGTTTCTGGTGGGCCTGGGAGCCCCCATCGAGGGCAAGGATCTGAAGATGC

[0338] CTTGGTGGGTGAAGCTGAAGCCCAGCCTGTTCTTGGCCTTCCTGGAAGGCTTCCGGGCCCCAC

[0339] ATCGTGGAGCAGCTGGTTCGACGACCCCCAACAAAGAATCTGCCCTTCTTTCAGGAACTGAGCTG

[0340] GTATCTGGGCCTGTTCCGGCATCAAGGCCGACATCAAGGTGGAGGAAGTGGGCGACAAGCACA

[0341] AGATCATCTTCGACGCCGGCAGGCTGGACGTGGACAAGCAGTTCATCGAGACCTGGGAGGAT

[0342] GTGGAGGTGACCTACAACCTGACCACAGAGAAGGGCAATCTGCTGGCCAACGGCCTGTTCGT

[0343] GAAGAAC

[0344] 构建下述的哺乳动物表达载体 :pTT3-pfu lon HL(+)、pTT3-pfu lon HL(-)、pTT3-pfu lon LH(+)、pTT3-pfu lon LH(-)、pTT3-pfu lon LKH(+)、pTT3-pfu lon LKH(-)、pTT3-pab lon HL(+)、pTT3 pab lon HL(-)、pTT3-pab lon LH(+)、pTT3-pab lon LH(-)、pTT3-pab lon LKH(+)、pTT3-pab lon LKH(-)。在这里, H 和 L 组分代表命名为 D2E7 的抗体的免疫球蛋白重链和轻链。关于 pTT3pab lon HL(-) 构建体的图示,参见图 1。图 2 解释了能够表达 D2E7 抗体的这些瞬时表达载体的 sORF 组分的结构的方面。

[0345] 尽管 pTT3 载体代表一个具体实施方案,其它实施方案可以包括关于分离的核酸的方面,所述分离的核酸编码一种或多种本文公开的蛋白。另一个实施方案提供了一种包含分离的核酸序列的载体,其中所述载体选自 :pcDNA ;pTT (Durocher 等人, Nucleic Acids Research 2002, Vol 30, No. 2 :E9) ;pTT3 ;pEFBOS (Mizushima, S. 和 Nagata, S., 1990,

Nucleic Acids Research Vol 18, No. 17 :5322) ;pBV ;pJV ;和 pBJ。如上面所指出的,在 pTT3 载体骨架上制作不同的构建体。该载体具有 EBV 复制起点,这允许它在悬浮培养的经转染的 293E 细胞 (其表达爱泼斯坦-巴尔病毒核抗原 1) 中进行附加型扩增。参见 Durocher 等人,其描述了载体 pTT。与 pTT 相比,pTT3 具有额外的多克隆位点,如在 Ghayur, Tariq 等人于 2005 年 7 月 7 日提交的美国专利申请公开 20050147610 所指示。

[0346] 每个基于 pTT3 的载体具有一个受 CMV 启动子调节的 ORF。在 ORF 中,内含肽序列插入在抗体重链和轻链 (分别是 HC 和 LC,或简称 H 和 L) 之间的框架内,其次序为 HC- 内含肽 -LC 或 LC- 内含肽 -HC。具有“HL”命名的构建体具有抗体 HC 编码序列,其后面是内含肽,然后是 LC 编码序列;具有“LH”命名的构建体具有 LC 编码序列,其后面是内含肽,然后是 HC 编码序列。具有“LKH”命名的构建体具有插入在 LC 和内含肽之间的赖氨酸 (K)。具有“(-)”命名的构建体具有在 ORF 开始处的一个信号肽和插入在内含肽的最后一个氨基酸与跟随内含肽的成熟抗体重链或轻链的第一个氨基酸之间的甲硫氨酸。具有“(+)”命名的构建体具有在 ORF 开始处的一个信号肽和在内含肽的下游抗体亚基的开始处的第二个信号肽。

[0347] 通过瞬时转染,将构建体导入 293E 细胞中。简而言之,使用编码 ORF 构建体的 pTT3 载体和聚乙烯亚胺 (PEI),制备复合物。使用 PEI-DNA 复合物转染 HEK293E 细胞;参见 Durocher 等人,2002,Nucl. Acids Res. 30 :E9。在转染后 4-7 天,收集细胞和培养物上清液用于分析。

[0348] 通过构建体进行蛋白表达。在多个瞬时表达实验中,在转染后第 7 天或第 8 天,收集培养物上清液样品。通过 ELISA 评估样品,且所述样品含有如下所示的、来自 IgG 测量的分泌的抗体的水平或范围。

[0349] 表 9. Lon 内含肽免疫球蛋白 sORF 构建体和抗体生产。

[0350]

构建体	IgG (分泌的), μg/ml
sORF 构建体	
pTT3-pfu lon HL (+)	1.4 – 2.1
pTT3-pfu lon HL (-)	31-40
pTT3-pfu lon LH (+)	< 0.1
pTT3-pfu lon LH (-)	1.6
pTT3-pfu lon LKH (+)	< 0.1
pTT3-pfu lon LKH (-)	10
pTT3-pab lon HL (+)	1.3
pTT3- pab lon HL (-)	41-68
pTT3- pab lon LH (+)	< 0.1
pTT3- pab lon LH (-)	0.5
pTT3- pab lon LKH (+)	< 0.1
pTT3- pab lon LKH (-)	0.9
其它构建体	
对照载体	10- 60

[0351] 在这些实验中,包括常规的双载体系统 (其表达与上述的构建体系列相同的抗体,并使用相同的调控元件) 作为对照 (参见表,末行)。因而,对照载体表达 D2E7 抗体,其使用导入来自 2 个分开的 pTT3 载体所负载的 2 个分开的 ORF 的抗体重链和轻链的常规方

案。由该对照载体系统生产的抗体分泌水平的范围是 10–60 $\mu\text{g/ml}$ ，如该表所示。

[0352] 与使用常规对照载体生产的水平相比，由使用 Lon 内含肽的几种 sORF 构建体设计生产的 IgG 分泌水平是在相同的范围内或更高。这些水平显著高于使用“2A”技术生产的那些水平，后者据报道在哺乳动物细胞中是 1.6 $\mu\text{g/ml}$ (Fang 等人, 2005, Nature Biotechnology 23 :584–590)。尽管 Pab Lon 和 Pfu Lon 内含肽可以用于构建体设计中以产生希望的抗体生产水平，在描述的 pTT3 构建体中的 Pab Lon 内含肽允许更高水平的抗体分泌。这些数据也提示，当使用 HL 构建体设计时，抗体分泌水平通常比使用 LH 构建体设计时更大。通过组合免疫球蛋白链的次序特征和信号序列的方面，HL(-) 构建体能够产生在研究的那些中最高水平的分泌型抗体产物。

[0353] 表达产物的进一步表征

[0354] 关于表达的方面（包括表达产物分析），进一步表征在表 9 中列出的某些 sORF 构建体。这些构建体包括 4 个实施例，它们生产相对更高水平的分泌的抗体：pTT3 pfu lon HL(-)、pTT3 pfu lon HL(+)、pTT3 pfu lon LKH(-) 和 pTT3 pab lon HL(-)。通过蛋白 A 亲和色谱法，纯化从这些构建体生产的分泌型抗体，并在还原和非还原 SDS-PAGE 凝胶上进行分析，并测定它们的 HL 和 LC 的 N- 端氨基酸序列。

[0355] 使用 pTT3 pfu lon HL(-) 生产的样品含有与抗体 HC、抗体 LC 和完全装配的抗体相对应的凝胶迁移带（在非还原凝胶上），它们的迁移与使用常规载体（诸如上述的对照 D2E7 载体）的传统方法所生产的抗体不可区分。在还原凝胶上，除了与抗体 HC 和 LC 相对应的带以外，还出现了 2 个更高分子量 (MW) 带，它们对应着未加工的三联体蛋白 (HC- 内含肽-LC) 和部分地加工的 HC- 内含肽融合体。该评估是基于蛋白印迹分析和质谱法分析。这 2 个带的丰度似乎取决于培养条件，并且可以通过修改培养条件来减小。使用根据本文提供的其它描述的和 / 或本领域理解的方法，可以方便地从完全加工的抗体药用物质中去除这些更高分子量产物。

[0356] 使用 pTT3 pfu lon HL(+) 生产的样品含有与抗体 HC、抗体 LC 和完整抗体相对应的带（在非还原凝胶上），它们的迁移与传统方法所生产的抗体不可区分。另外，存在一个与三联体多蛋白相对应的更大的分子量带。使用 pTT3 pfu lon LKH(-) 生产的样品也含有与 HC、LC 和完整抗体相对应的带（在非还原凝胶上），它们的迁移与使用常规载体生产的抗体不可区分。在还原凝胶上，除了与 HC 和 LC 相对应的带以外，还存在 2 个更高分子量带。它们中的第一个对应着三联体多蛋白，如上面关于其它载体设计所述；第二个带对应着 LC- 内含肽融合产物，其源自在该接头处的不完全切割。关于产物的相对丰度，似乎存在与切割的 LC 相同量的 LC- 内含肽融合体。

[0357] 使用 pTT3 pab lon HL(-) 生产的样品含有与 HC、LC 和完整抗体相对应的带（在非还原凝胶上），它们的迁移与通过传统方法生产的抗体不可区分。在还原凝胶上，除了与 HC 和 LC 相对应的带以外，还存在 1 个主要的更高分子量带，它似乎对应着未加工的三联体蛋白（基于蛋白印迹分析）。与使用 pTT3 pfu lon HL(-) 生产的样品相比，在更高分子量带中存在相对更小量的该三联体。该结果提示，尽管在我们的载体设计中 Pfu lon 内含肽和 Pab lon 内含肽是同源的且在功能上是类似的，Pab lon 介导的 N- 端切割比由 Pfu lon 介导的切割更完全，因为在从 Pab lon 构建体表达以后，没有观察到 HC- 内含肽融合体。但是，应当指出，两种构建体产生完整装配的抗体产物。

[0358] 在一方面,相对于其它构建体产品,诸如完全加工的和完全自我装配的抗体产物,某些蛋白产物(如未加工的和部分加工的蛋白)可以视作污染物产物。另一方面,这样的某些蛋白产物可能是有用的,例如作为用于进一步加工反应和/或定向装配的材料,这可以产生完整抗体产物。如果这些蛋白产物视作污染物副产物,那么如指出的,存在这样的选择和方案:其便利去除,并从而富集或纯化希望的组分,诸如完整抗体。

[0359] 除了来自不同构建体表达系统的培养物上清液的胞外样品以外,还得到细胞内样品,并使用具有对 HC 和 LC 的特异性的检测抗体,通过蛋白印迹分析法进行分析。如关于在培养的上清液样品中的那些物质所述,观察类似的蛋白物质。

[0360] 测定了不同构建体的重链和轻链产物的 N-端氨基酸序列(参见下表)。结果表明,内含肽-介导的蛋白切割精确地发生在 2 个剪接点处,即在 HC 和内含肽组分的接头处和在 LC 和内含肽组分的接头处。

[0361] 表 10. 来自 sORF 构建体的表达产物的主要物质的重链和轻链 N-端氨基酸序列。

构建体	HC, N-端AA序列	SEQ ID NO:	LC, N-端AA序列	SEQ ID NO:
对照、成熟的HC或LC	EVQLVESGGG	9	DIQMTQSPSS	11
sORF 构建体				
pTT3 pfu lon HL (+)	EVQLVESGGG	9	DIQMTQSPSS	11
pTT3 pfu lon HL (-)	EVQLVESGGG	9	MDIQMTQSPS	12
pTT3 pfu lon LKH (-)	MEVQLVESGG	10	DIQMTQSPS	11
pTT3 pab lon HL (-)	EVQLVESGGG	9	MDIQMTQSPS	12

[0363] 来自 Lon 内含肽构建体的 IgG1 抗体的功能性质

[0364] 还通过抗原特异性的 ELISA,分析了来自表 9 的 sORF 构建体设计的分泌的 D2E7 抗体产物。分析结果证实,所述构建体抗体产物结合人 TNF α ,即 D2E7 抗体的配体。因而,所述基于内含肽的构建体和表达系统能够表达和产生 sORF 产物,其产生完全自我装配的多聚抗体,所述抗体是有功能的和抗原特异性的。

[0365] 通过蛋白 A 亲和纯化、随后的 SEC、尺寸排阻色谱法,纯化使用 pTT3 pfu lon HL(-) 构建体生产的抗体。使用 BiaCore™ 系统,通过表面等离子体共振技术,分析纯化的抗体。sORF 构建体产品的表征包括它与有关的配体 TNF α 的结合的方面。在表 11 中,指出了来自 BiaCore 分析的值的結果,关于结合速率常数 (k_a ,单位为 1/Ms)、解离速率常数 (k_d ,单位为 1/s) 和平衡解离常数 (K_D ,单位为 M) 的动力学参数。认为解离常数值 (K_D) 与使用常规载体(使用 2 个独特的免疫球蛋白链 ORF)生产的阿达木单抗(D2E7)抗体的对应值类似。

[0366] 表 11. 从 sORF 构建体生产的抗体的动力学参数。

[0367]

构建体	k_a (1/Ms)	k_d (1/s)	K_D (M)
pTT3 pfu lon HL(-)	1.51E+06	1.10E-04	7.29E-11

[0368] 实施例 2. 生产具有轻链序列变化的抗体的 sORF 构建体

[0369] 产生了几种 sORF 构建体,其轻链序列具有相对于 D2E7 的免疫球蛋白轻链的变化。将这些 sORF 构建体工程化成具有也象在 D2E7 中一样的重链,并因而能够生产 IgG1 抗体物质。使用构建体 pTT3 pfu lon HL(-) 和 pTT3 pab lon HL(-) 作为骨架,我们产生和测试

了这样的构建体：其具有在 C-端剪接点（即，在内含肽和下游免疫球蛋白轻链组分之间的接头）处的序列变化（参见表 12）。通过蛋白 A 亲和纯化，纯化某些构建体的分泌的免疫球蛋白，并在还原和非还原 SDS-PAGE 凝胶上分析。还使用针对 HC 和 LC 的抗体，通过蛋白印迹分析来分析细胞内样品。参见，例如，图 3 和图 4。

[0370] 图 3 解释了 sORF 表达产物的 SDS-PAGE 凝胶蛋白分析的结果。通过蛋白 A 亲和色谱法纯化分泌的 IgG 分子，并在非还原 (A) 和还原条件 (B) 下在 SDS-PAGE 凝胶上进行分离。泳道和样品从左向右为：(泳道 1) 分子量参照标志物；(2) 对照构建体产物，即来自非-sORF 表达系统的 D2E7 抗体；(3) Pab-lon mut A1；(4) Pab-lon mut A2；(5) pTT3 pfu lon YP, 和 (6) pTT3 pfu lon MA。

[0371] 图 4 解释了其它 sORF 表达产物的 SDS-PAGE 凝胶蛋白分析的结果。通过蛋白 A 亲和色谱法纯化分泌的 IgG，并在非还原 (A) 和还原 (B) 条件下通过 SDS-PAGE 凝胶进行分离。泳道和样品从左向右为：(泳道 1) 分子量标志物；(2) 对照 D2E7 产物；(3) pTT3 pfu lon HL(-)；和 (4) pTT3 pfu lon MutA。

[0372] 表 12. sORF 构建体在内含肽-LC 的 C-端剪接接头处的氨基酸序列

[0373]

构建体	内含肽	SEQ ID NO:	C-端接头	SEQ ID NO:	成熟的LC	SEQ ID NO:
pTT3 pfu lon HL (-)	ANGLFVKN	13	<u>M</u>		DIQMTQS	17
pTT3 pfu lon MutA	ANGLFVKN	13	<u>MRAKR</u>	14	DIQMTQS	17
pTT3 pfu lon MutB	ANGLFVKN	13			DIQMTQS	17
pTT3 pfu lon YP	ANGLFVKN	13	<u>YP</u>		DIQMTQS	17
pTT3 pfu lon RP	ANGLFVKN	13	<u>RP</u>		DIQMTQS	17
pTT3 pfu lon VP	ANGLFVKN	13	<u>VP</u>		DIQMTQS	17
pTT3 pfu lon QP	ANGLFVKN	13	<u>QP</u>		DIQMTQS	17
pTT3 pfu lon AP	ANGLFVKN	13	<u>AP</u>		DIQMTQS	17
pTT3 pfu lon HA	ANGLFVKN	13	<u>HA</u>		DIQMTQS	17
pTT3 pfu lon YA	ANGLFVKN	13	<u>YA</u>		DIQMTQS	17
pTT3 pfu lon MP	ANGLFVKN	13	<u>MP</u>		DIQMTQS	17
pTT3 pfu lon MA	ANGLFVKN	13	<u>MA</u>		DIQMTQS	17
pTT3 pab lon MutA1	ANGLFVKN	13	<u>HA</u> RGVFRR	15	DIQMTQS	17
pTT3 pab lon MutA2	ANGLFVKN	13	<u>MD</u> RGVFRR	16	DIQMTQS	17
pTT3 pab lon AIQ	ANGLFVKN	13	--		AIQMTQS	18
pTT3 pab lon NIQ	ANGLFVKN	13	--		NIQMTQS	19
pTT3 pab lon NFQ	ANGLFVKN	13	--		NFQMTQS	20

[0374] 由表 12 的构建体生成的免疫球蛋白分泌水平如表 13 所示。测定了在成熟轻链的 N-端处的氨基酸,并显示了表征部分序列的结果。

[0375] 表 13. 来自 sORF 构建体的抗体的轻链的 IgG 水平和 N-端蛋白序列。

[0376]

构建体	IgG (ug/ml)	LC的N-端氨基酸序列	SEQ ID NO:
pTT3 pfu lon HL (-)		<u>M</u> DIQMTQS	21
pTT3 pfu lon MutA	17	<u>MRAKR</u> DIQMTQS	22
pTT3 pfu lon MutB	6	DIQMTQS	17
pTT3 pfu lon YP	32	<u>YP</u> DIQMTQS	23
pTT3 pfu lon RP	22	<u>RP</u> DIQMTQS	24
pTT3 pfu lon VP	20	<u>VP</u> DIQMTQS	25
pTT3 pfu lon QP	13	<u>QP</u> DIQMTQS	26
pTT3 pfu lon AP	21	<u>AP</u> DIQMTQS	27
pTT3 pfu lon HA	18	<u>HA</u> DIQMTQS	28
pTT3 pfu lon YA	15	<u>YA</u> DIQMTQS	29
pTT3 pfu lon MP	29	<u>MP</u> DIQMTQS	30
pTT3 pfu lon MA	33	<u>MA</u> DIQMTQS	31
pTT3 pab lon MutA1	16	<u>HA</u> RGVFRR DIQMTQS	32
pTT3 pab lon MutA2	11	<u>MD</u> RGVFRR DIQMTQS	33
pTT3 pab lon AIQ	24	AIQMTQS	18
pTT3 pab lon NIQ	20	NIQMTQS	19
pTT3 pab lon NFQ	18	NFQMTQS	20

[0377] 结果

[0378] 对于这些构建体,在 +1 位置(紧随在内含肽之后)处使用不同的 AA 残基,似乎是分泌的抗体的产生因素。在该位置处使用氨基酸残基 H、Y、R、V、Q、A、N 和 M,会产生相对更高水平的抗体表达。轻链的 N-端氨基酸(表 13)的分析提示,在内含肽的 C-末端处的完全的且精确的切割。与由构建体 pTT3 pfu lon HL(-) 和 pTT3 pab lon HL(-) 生成的抗体类似地,加工的 HC 和 LC 以及装配的完整抗体占分泌的蛋白物质的大部分。但是,当氨基酸

天冬氨酸 (Asp ;D) 直接地用在内含肽之后 (象在构建体 pTT3 pfu lon MutB 中一样) 时, 几乎没有抗体分泌。分析了由该构建体生成的细胞内蛋白。经测定, 当 D 是在内含肽之后的第一个氨基酸时, 在 C- 端剪接点处几乎没有切割, 几乎不产生抗体 LC, 并产生相对大量的内含肽-LC 融合蛋白。

[0379] κ 同种型可变区 (V_{κ}) 的成熟种系轻链的氨基酸序列通常以 D、E、N、A 或 V 开始; λ 同种型 (V_{λ}) 的成熟种系轻链的氨基酸序列通常以 Q、S、L 或 N 开始。根据我们的结果, 使用 Pab lon 或 Pfu lon 内含肽的 sORF 载体用于生产抗体的实施方案包括: 具有以任意氨基酸开始的 LC 的那些。在优选的实施方案中, 为了实现更高的总完全加工效率的目的, LC 以除了 D 或 E 以外的氨基酸开始, 尽管这样的氨基酸可以充当有效选择。

[0380] 我们发现, 在内含肽和位于内含肽下游的成熟抗体 LC 之间的区域似乎会促进在 N- 端剪接点处的切割效率。例如, 我们对比了构建体 pTT3 pfu lon HL(-) 和 pTT3 pfu lon MutA 的产品。尽管当使用 pTT3 pfu lon HL(-) 时在 N- 端剪接点处的切割是不完全的, 产生一些部分地加工的 HC- 内含肽融合蛋白, 当替代性地使用构建体 pTT3 pfu lon MutA 时, 该蛋白物质的量显著减少 (参见图 2)。因而, 尽管不同的构建体可以用于产生希望的产物, 某些构建体可能具有特定属性, 诸如产生相对更高产量的特定目标产物 (例如, 完全加工的和自我装配的多聚的分泌型抗体) 的能力。

[0381] 实施例 3. sORF 构建体的内含肽组分的其它选择

[0382] 来自詹氏甲烷球菌和 Pyrococcus abyssi 的 k1bA 基因的内含肽

[0383] 我们研究了 sORF 构建体的其它内含肽选择, 包括诸如来自甲烷球菌属和火球菌属物种的 k1bA 基因的内含肽。我们发现, 在 k1bA 基因中包含的内含肽也是用于介导蛋白表达和加工的有效选择, 包括在不同的单个开放读码框构建体设计中切割抗体重链和轻链的背景下。

[0384] 具体地, 我们检查了来自詹氏甲烷球菌 (Mja k1bA 内含肽)、Pyrococcus abyssi (Pab k1bA 内含肽) 和激烈火球菌 (Pfu k1bA 内含肽) 的 k1bA 内含肽。来自前两种生物体的内含肽是小型内含肽, 它们缺少内切核酸酶结构域, 而 Pfu k1bA 是全尺寸的内含肽。天然内含肽蛋白区段的序列长度分别是 168、333 和 522 个氨基酸。在下表中提供了与这些内含肽相结合的序列信息。因而为表达系统开发了内含肽、修饰的内含肽和构建体。

[0385] K1bA 内含肽序列和载体构建体设计

[0386] 修饰 Mja k1bA 的核苷酸序列, 以允许哺乳动物密码子选择的相对优化。表 14 提供了已经如此修饰的詹氏甲烷球菌的 Mja k1bA 内含肽基因的核酸序列信息 (SEQ ID NO :34)。表 15 提供了 Mja K1bA 内含肽区段的蛋白序列信息。也参见在 NCBI/ 蛋白 (MJ0781) 中的登记号 Q58191。表 16 提供了 Pab k1bA 内含肽基因的核酸序列信息, 所述基因相对于天然序列在密码子选择方面进行了修饰。对于天然序列, 关于 Pab K1bA 内含肽的 NEB Inbase 信息, 参见登记号 [B75050, 在 NCBI/ 蛋白 PAB 1457 中]。根据该来源, 指示的蛋白氨基酸序列被指示为包括 -1 和 +1 外显肽残基, 它们分别似乎是 G 和 C。表 17 提供了 Pab K1bA 内含肽蛋白区段的氨基酸序列信息。表 18 提供了 Pfu k1bA 内含肽基因的核酸序列信息 (天然)。表 19 提供了 Pfu K1bA 内含肽蛋白区段的氨基酸序列信息; 也参见 NCBI 中的登记号 AE010211。

[0387] 表 14. Mja k1bA 内含肽基因核苷酸序列 (SEQ ID NO :34), 修饰的密码子选择。

- [0388] Gctctggcctacgacgagcccatctacctgagcgacggcaacatcatcaacatcggcgagtt
[0389] cgtggacaagttcttcaagaagtacaagaacagcatcaagaaaggagacaacggcttcggct
[0390] ggatcgacatcggcaacgagaacatctacatcaagagcttcaacaagctgtccctgatcatc
[0391] gaggacaagcggatcctgagagtggtggcggaagaagtacagcggcaagctgatcaagatcac
[0392] caccaagaaccggcgaggagatcacctgacccacgaccaccccggtgtacatcagcaagaccg
[0393] gcgaggtgctggaaatcaacgccgagatggtgaaagtgggcgactacatctatatccccaag
[0394] aacaacaccatcaacctggacgaggtgatcaaggtggagaccgtggactacaacggccacat
[0395] ctacgacctgacctggaggacaaccacacctacatcgccggcaagaacgagggttcgccg
[0396] tgagcaac
[0397] 表 15. Mja K1bA 内含肽蛋白氨基酸序列 (SEQ ID NO :35)。
[0398] ALAYDEPIYLSGNIINIGEFVDKFFKKYKNSIKKEDNGFGWIDIGNENIYIKSFNKLII
[0399] EDKRILRVWRKKYSGKLIKITTKNRREITLTHDHPVYISKTEVLEINAEMVKVGDYIYIPK
[0400] NNTINLDEVIKVTVDYNGHIYDLTVEDNHTYIAGKNEGFAVSN
[0401] 表 16. Pab klba 内含肽基因核苷酸序列 (SEQ ID NO :36), 修饰的密码子选择。
[0402] Gctctgtactacttcagcgagatccagctgcccacggcaaagagttcatcggcaaactggt
[0403] ggacgagctgttcgagaagtaccacgacaagatcggcaagtacaaggacatggaatacgtgg
[0404] agctgaacgaagaggacaccttcgaggtgatcagcatcgccccgacctgagcgccaggcgg
[0405] cacaaggtgaccacgtgtggcgggcggaaggtgaaagacggcgagaagctggtgaagatccg
[0406] gaccgccagcggcaaagaactggtgctgaccaggaccaccccggtgttcgtgctgctgggcc
[0407] gggacgtggccagacgggacgccggcaacgtgaaagtgggcgacgagatcgccgtgctgaac
[0408] accaggcccgacttcagcgtgctgtccccccctgccatgcccagctgctgtccgagccctt
[0409] caactacgagctgtccagcatcggcgacgtggcctgggacgaggtggtggaggtggacgaga
[0410] tcgacgccaagggcctgggcgtggagtacctgtacgacctgacctggacatcaaccacaac
[0411] tacgtggccaacggcatcgtggtgtccaac
[0412] 表 17. Pab K1ba 内含肽蛋白氨基酸序列 (SEQ ID NO :37)。
[0413] ALYYFSEIQLPNGKEFIGKLVDLFEKYHDKIGKYKDMEYVELNEEDTFEVISIGPDLSARR
[0414] HKVTHVWRRKVKDGEKLVKIRTASGKELVLTQDHPVFVLLGRDVARRDAGNVKVGDEIAVLN
[0415] TRPDFSVLSPPAMPELLSEPFNYELSSIGDVAWDEVVEVDEIDAKGLGVEYLYDLTVDINHN
[0416] YVANGIVVSN
[0417] 表 18. Pfu klba 内含肽基因核苷酸序列 (SEQ ID NO :38), 天然的。
[0418] gcactttacgattttctctgtcatccaactatctaattggtagattttgtacttataggagattt
[0419] agtcgaggaattattcaagaagtatgccgagaaaaattaaaacatacaaaagaccttgagtaca
[0420] tagagcttaacgaggaagaccgttttgaagttgttagtgtagtccagatttgaaggctaatt
[0421] aaacatgttgtctcaagagtttggagaagaaaggtcagagagggggaaaaagctaatacgcatt
[0422] aaagacgagaactggcaacgaataatcctcactagaaatcatccgctattttgccttctcca
[0423] atggagacgtagtcagaaaaggccgagaagctcaaagttggggatagagttgcagtgatg
[0424] atgagacctccttcacctcctcaactaaagctgtagttgacctgcaatttacgtgaaaaat
[0425] aagtgattactaccttgttccgaacggaaaaggtatgataaaagttcctaacgatggtatc
[0426] ctccagaaaaggcccaatatcttcttttcagtaaatcatatcctgtaaaaattagtcagagaa

[0427] gttgatgagaagttatcctatctcgtctggagttatactcggatgatgggtatatatcatcgaa
[0428] tggatactacatctcagctacatttgacgacgaagcttacatggatgcctttgtctctgtag
[0429] tctcggactttatccctaactatgtccccagttataaggaagaacggagattacacaattgta
[0430] actgtttggctcgaagatTTTTGTGAAATGCTCTCAAGGATATTTGGAATACCAAGGGGCGAG
[0431] AAAATCTATGTGGGATATTCAGACGCTAGTACTTTCAAATGACGATCTTATGAGATACTTCA
[0432] TAGCTGGACTTTTCGACGCTGATGGGTACGTAGATGAAAATGGGCCCTCCATAGTCTTAGTA
[0433] ACAAAGAGTGAAACCGTGGCAAGGAAGATTTGGTACGTTCTTCAGAGGTTGGGGATCATAAG
[0434] TACAGTTTCCCGTGTAAGAGCAGAGGGTTTAAAGAAGGCGAGCTGTTCAAGGGTAATTATTA
[0435] GTGGTGTGAAGATCTTGCTAAATTTGCAAAATTCATACCCCTACGTCCTCAAGAAAGAGG
[0436] GCCAACTTATGGAGATATTAAGGACTAAGAAGCCATATCGGGGAAGAAGAAGCTTACCGCGT
[0437] GCCGATATCCAGTGATATGATAGCTCCTCTCCGTCAAATGTTGGGATTAAGTGTTCAGAGC
[0438] TGTCTAAGTTAGCGTCTTATTATGCAGGGGAAAAAGTTTCTGAAAGCCTAATTAGGCATATA
[0439] GAAAAGGGAAGGGTCAAAGAGATAAGACGCTCTACGCTCAAGGGGATTGCCCTTGCTCTCCA
[0440] GCAGATAGCTAAAGATGTGGGTAAACGAAGAAGCTTGGGTGAGAGCCAAGAGGCTTCAATTGA
[0441] TAGCTGAGGGAGATGTTTACTGGGATGAAGTCGTAAGTGTGAGGAAGTTGATCCGAAGGAG
[0442] CTTGGCATTGAGTACGTCTATGACCTCACGGTTGAGGACGACCACAATTATGTGGCAAATGG
[0443] CATACTAGTCTCAAAC

[0444] 表 19. Pfu K1ba 内含肽蛋白氨基酸序列 (SEQ ID NO :39)。

[0445] ALYDFSVIQLSNGRFVLIGDLVEELFKKYAEKIKTYKDLEY IELNEEDRFEVVSVPDLKAN
[0446] KHVVSrvWRRKVREGEKLIRIKTRTGNEIILTRNHPLFAFSNGDVVRKEAEKLVGDRVAVM
[0447] MRPPSPPTKAVVDPAIYVKISDYLLVPNGKGMKVPNDGIPPEKAQYLLSVNSYPVKLVRE
[0448] VDEKLSYLAGVILGDGYISSNGYYISATFDDEAYMDAFVSVVSDFIPNYVPSIRKNGDYTIV
[0449] TVGSKIFAEMLSRIFGIPRGRKSMWDIPDVVLSNDDLRYFIAGLFDADGYVDENGPSIVLV
[0450] TKSETVARKIWIYVLQRLGIISTVSRVKS RGFKEGELFRV IISGVEDLAKFAKFIPLRHSRKR
[0451] AKLMEILRTKKPYRGRRTYRVP IISDMIAPLRQMLGLTVAELSKLASYYAGEKVS ESLIRHI
[0452] EKGRVKEIRRSTLKGIALALQQIAKDVGN EEA WVR AKRLQLIAEGDVY WDEVVSVEE VDPKE
[0453] LGIEYVYDLTVEDDHNYVANGILVSN

[0454] 我们使用采用哺乳动物密码子选择的序列,合成了 Mja k1ba 内含肽和 Pab k1ba 内含肽的核苷酸序列。构建下述哺乳动物表达载体 :pTT3-Pab k1ba HL(-) ;pTT3-Pab k1ba HL(+) ;pTT3-Pab k1ba LH(-) ;pTT3-Mja k1ba HL(-) ;pTT3-Mja k1ba HL(+) ;pTT3-Mja k1ba LH(-)。如在本文别处所述,在 PTT3 载体骨架上制备这些构建体。Pfu k1ba 内含肽核苷酸序列是天然序列,也构建了 pTT3-Pfu-k1ba-HL(+)

[0455] 如关于使用 Lon 蛋白酶内含肽的构建体所指示的,所有具有“HL”命名的构建体都具有抗体免疫球蛋白重链 (HC) 编码序列,其后面是内含肽区段,再后面是轻链 (LC) 编码序列。同样地,所有具有“LH”命名的构建体都具有抗体 LC 编码序列,其后面是内含肽区段,再后面是 HC 编码序列。具有“(-)”命名的构建体具有在 ORF 开始处的一个信号肽和插入在内含肽区段的最后一个氨基酸和下游外显肽区段(例如,在内含肽之后的成熟抗体重链或轻链)的第一个氨基酸之间的甲硫氨酸。具有“(+)”命名的构建体具有在 ORF 开始处的一个信号肽和在内含肽的下游抗体亚基开始处的第二个信号肽。

[0456] 通过瞬时转染技术,将具有不同的 K1bA 内含肽区段和构型的构建体导入 293E 细胞中。在转染后 7-8 天时,通过使用 ELISA 测量 IgG 水平,分析培养物上清液中分泌的抗体。这些结果参见表 20,对于每种构建体表达系统,值的单位是:微克/ml 样品。

[0457] 表 20. K1bA 内含肽 sORF 构建体和分泌的抗体生产。

[0458]

构建体	IgG (μ g/ml)
pTT3-Pab k1bA HL(-)	19
pTT3-Pab k1bA HL(+),	6
pTT3-Pab k1bA LH(-)	0.4
pTT3-Mja k1bA HL(-)	13
pTT3-Mja k1bA HL(+),	4
pTT3-Mja k1bA LH(-)	< 0.1
PTT3-Pfu-k1bA HL(+)	< 0.1

[0459] 我们纯化和分析了由构建体 pTT3-Pab k1bA HL(-) 和 pTT3-Mja k1bA HL(-) 表达的分泌型抗体产物。通过蛋白 A 亲和色谱法纯化所述抗体产物,并通过在还原和非还原条件下的 SDS-PAGE 电泳技术来表征。参见图 5。在非还原条件下,来自这 2 种载体的培养物上清液样品主要作为单个带迁移,但是,具有比对照抗体明显更大的分子量。在还原条件下,我们发现来自这 2 种载体的培养物上清液样品含有在大小上与抗体 LC 和 HC- 内含肽融合组分相对应的可检测带。使用针对 IgG1 Fc 或 κ 轻链的任一种抗体的对应的免疫印迹与这些带的表征相一致。这些结果提示,在 C- 端剪接点处存在相对有效的切割,但是在 N- 端剪接点处总体上存在不太有效的切割或甚至几乎没有切割。但是,即使对于其中实现了小于完全切割效率的构建体和表达系统,免疫球蛋白重链和轻链亚基能够装配和变成分泌为完整的 IgG 抗体分子。

[0460] K1b 内含肽在 N- 端内含肽剪接接头处的修饰

[0461] 考虑到上面关于在 N- 端剪接点处的切割所述的结果,我们进行了其它工作,以提供增强的切割效率。我们注意到,这 2 种内含肽 (Pab k1bA 和 Mja k1bA) 中的每一种的第一个氨基酸是丙氨酸 (Ala ;A),而不是半胱氨酸 (Cys ;C)。我们理解为,半胱氨酸是能够在其它内含肽系统中起亲核基团功能的残基。我们测试了与在免疫球蛋白 HC 区段的末端处导入一个氨基酸甘氨酸 (它对于内含肽的上游 k1bA 外显肽而言是天然的) 相组合地在该位置重新导入亲核的氨基酸半胱氨酸的影响。参见表 21,该表提供了这些额外的构建体的蛋白区段的序列信息。该表提供了在天然 Pab k1ba 内含肽、Pab k1ba HL(-) 构建体 (在该背景下,它称作 WT) 和 3 个具有在 N- 端剪接点处的突变的构建体 (Pab k1ba HL(-)GC ;Pab k1ba HL(-)GA ;和 Pab k1ba HL(-)KC) 的 2 个剪接点处的氨基酸残基。星号 (*) 指示已经在突变型构建体中导入变体氨基酸残基的位置。在这些构建体中,Pab-k1bA HL(-)GC 表现

出表达和加工蛋白的能力,其具有在N-端内含肽接头处的有效切割。也参见图6,该图解释了来自某些构建体的IgG蛋白的表达和SDS-PAGE分析的结果。

[0462] 表 21. 具有在N-端内含肽接头处的修饰的Pab-klbA构建体的区段的蛋白序列
[0463]

构建体	在C-端处的 外显肽/ HC	SEQ ID NO:	在N-端处的 内含肽	SEQ ID NO:	在C-端处的 内含肽	在N-端处的 外显肽/ LC	SEQ ID NO:
天然的Pab klbA	* GHD G	40	* A LYY	42	VSN	CMGT	44
Pab-klbA HL(-) WT	SPG K	41	A LYY	42	VSN	MDIQ	45
Pab-klbA HL(-) GC	SP G		C LYY	43	VSN	MDIQ	45
Pab-klbA HL(-) GA	SP G		A LYY	42	VSN	MDIQ	45
Pab-klbA HL(-) KC	SPG K	41	C LYY	43	VSN	MDIQ	45

[0464] * 星号指示已经导入变化的位置。

[0465] 表 22. 在Pab-klbA构建体中的蛋白序列区段的其它信息。

[0466]

构建体	HC	SEQ ID NO:	N-端接头	内含肽	SEQ ID NO:
pTT3- Pab klbA HL (-)	LSLSPGK	46	M	ALYYFSEIQ	48
pTT3- Pab klbA GC	LSLSPG	47	C	ALYYFSEIQ	48
pTT3- Pab klbA GA,	LSLSPG	47	-	ALYYFSEIQ	48
pTT3- Pab klbA KC	LSLSPGK	46	C	ALYYFSEIQ	48
pTT3- Pab klbA KA	LSLSPGK	46	-	ALYYFSEIQ	48

[0467] 用于制备载体的材料和方法:Pab-klbA HL(-) 变体 GA、GC 和 KC 的构建

[0468] 制备了某些载体构建体的几种变体形式。通过PCR,构建了Pab-klbA HL(-) 突变体GA、GC和KC。使用正向引物HC-F和反向引物Hint-R来PCR扩增重链的3'末端,以产生PCR产物#1。正向引物GA-F、GC-F和KC-F含有希望的突变以及与重链的3'末端互补的序列。使用引物GA-F、GC-F或KC-F和反向引物内含肽-R-2来扩增Pab-klbA内含肽的5'末端,以生成PCR产物#2。然后使用Qiagen凝胶提取试剂盒,纯化PCR产物#1和#2。使用外部引物HC-F和内含肽-R-2,将纯化的PCR产物对合到一起并扩增,以产生PCR产物#3。使用限制性酶SacII和RssII消化载体Pab-klbA HL(-),然后使用Qiagen凝胶提取试剂盒进行纯化。通过在最高效率DH5 α 细胞(Invitrogen)中的同源重组,将PCR产物#3亚克隆进Pab-klbA-HL(-)(用SacII和RssII切割)中。使用集落PCR,测定携带正确突变的转化体,随后测序。使用Qiagen Maxi试剂盒,扩增和纯化来自正确克隆的DNA。在下表中指示了引物序列。

[0469] 表 23. 突变体GA、GC、KC的引物序列。

[0470]

引物命名	核酸序列	SEQ ID NO:
GA-F	GCCTCTCCCTGTCTCCGGGTGCTCTGTACTACTTCAGCGAGATC	49
GC-F	GCCTCTCCCTGTCTCCGGGTGTCTGTACTACTTCAGCGAGATC	50
KC-F	TCTCCCTGTCTCCGGGTAAATGTCTGTACTACTTCAGCGAGATC	51
HC-F	CGGCGTGGAGGTGCATAATG	52
HINT-R	ACCCGGAGACAGGGAGAG	53
INTEIN-R-2	GGGTCAGCACCAAGTTCTTTG	54

[0471] 实施例 4. 稳定的载体和表达 sORF 构建体的细胞系的制备

[0472] 使用 sORF 构建体的实施方案, 可以开发出稳定表达载体和表达这种载体的细胞系。作为一个实例, 将含有 sORF (其具有 Pab lon 内含肽) 的稳定表达载体稳定地转染进 CHO (中国仓鼠卵巢) 细胞系中。我们使用包括 CMV 增强子、腺病毒主要晚期启动子、SV40 聚腺苷酸序列、胃泌素转录终止子、由 SV40 启动子驱动的 DHFR 编码序列和与在 pTT3 pab lon HL(-) (其用于瞬时表达系统中) 中的 ORF 相同的 ORF 在内的元件, 设计和制备了稳定的 sORF 表达载体。从而制备出 sORF 构建体 pA190-Pab-lon HL(-), 并能够用作稳定表达载体; 参见图 7。类似地制备其它构建体。

[0473] 使用磷酸钙转染技术, 将 pA190 构建体导入 CHO 细胞 (命名为 CHO B3.2) 中, 将所述细胞以 200 个细胞/孔的密度铺板在 48 个 96-孔平板内的选择培养基 (含有 MEM 和 5% FBS) 中。监测转染平板的细胞/菌落生长和 IgG 分泌。

[0474] 选择来自 sORF 稳定表达载体 pA190-Pab-lon HL(-) 的 30 个克隆和来自对照稳定表达载体 pA190 转染反应物的 32 个克隆的样品, 并培养。在该阶段, 不经扩增地评估选定的克隆的 IgG 分泌水平, 并在用 20nM 甲氨蝶呤 (MTX) 扩增以后也评估所述水平。对于稳定表达系统, sORF 载体产生显著频率的生长阳性孔 (共 4608 个孔中的 2304 个阳性孔), 其具有可察觉数目 (443/2304) 和比例 (19%) 的 IgG 分泌阳性的样品。在 12-孔平板中, 在 0nM MTX 的条件下, 约 29 个选定的克隆的 IgG 分泌水平范围是约 0.3 至约 2.5 微克/ml 培养物上清液。在含有 20nM MTX 的条件下, 约 24 个选定的克隆的 IgG 分泌水平范围是约 0.1 至约 6 微克/ml, 约半数挑选的克隆表现出大于 2 μ g/ml 的分泌水平。还注意到, sORF 构建体克隆在贴壁培养容器中表现出相对快速的汇合, 例如, 在含有 20nM MTX 的贴壁培养中, sORF 克隆比常规载体克隆更快速地生长且更早地达到汇合。在 20nM MTX 中在第 1 代时, 28% 的来自常规载体的克隆在 6 天内 (在 4 天、5 天或 6 天) 达到汇合; 而 77% 的来自 sORF pab lon 载体的克隆在 6 天内达到汇合 (参见图 8)。这些数据提示, 在稳定表达系统的开发 (包括 CHO 细胞系开发) 中, 使用 sORF 表达载体与常规载体相比的显著优点。在使用 100nM MTX 进行直接扩增的条件下, sORF 克隆也表现出有利的抗体分泌水平。在一个实验中, 16 个暴露于 100nM MTX 的克隆产生平均值为 6 μ g/ml 的 IgG 分泌水平, 前 5 个克隆的平均值为 12 μ g/ml, 最高的克隆具有 24 μ g/ml 的生产水平。下表显示了使用 MTX 扩增的结果, 在每个扩增步骤中具有最高表达水平。对于具有 pA190-Pab-lon-HL(-) 构建体的不同克隆, 值的单位是微克/ml IgG。

[0475] 表 24. 具有 MTX 扩增的 IgG 表达水平。

[0476]

克隆#	IgG, ug/ml		
	0 nM MTX	20 nM MTX	100 nM MTX (直接从0 nM)
1	0.94	0.64	
2	0.64	0.39	
3	0.75		
4	0.40	0.57	
5	0.83	3.90	
6	0.32	2.19	5.48
7	1.53	1.29	3.91
8	0.46	0.16	4.02
9	1.11	1.14	
10	0.88	2.17	
11	2.50	2.23	
12	1.30	1.60	6.23
13	0.84	2.41	
14	1.25	1.93	
15	0.86	2.68	3.78
16	0.50	0.30	3.51
17	1.05	1.13	1.42
18	2.23	2.45	
19	0.88	2.01	7.80
20	1.52	2.46	3.69
21	2.57	5.62	7.95
22	1.59	4.20	24.0
23	0.94	3.47	14.78
24	0.76	0.68	
25	1.22	4.35	
26	3.67	2.62	7.56
27	1.29	4.26	6.96
28	1.29	3.14	9.71
29	0.96		14.77
30	0.43	2.76	

[0477] 用于稳定的表达系统的材料和方法

[0478] 使用磷酸钙共沉淀方法,用表达载体转染中国仓鼠卵巢细胞,所述细胞培养在补充了 H/T 和 10% 透析的 FBS 的 Alpha MEM 中。参见 Kingston, R. E. 等人 (1993), 第 16.23 单元: Amplification Using CHOCell Expression Vectors, Current Protocols in Molecular Biology (Ausubel, F. M., Brent, R., Moore, D. M., Kingston, R. E., Seidman, J. G., Smith, J. A., 和 Struhl, K. 编; Wiley Interscience, New York), 2:16.23.1。次日,使用胰蛋白酶/EDTA 在室温转移细胞,并重新悬浮于补充了 5% 透析的 FBS 的 Alpha MEM (α -MEM+5% dFBS) 中,所述 Alpha MEM 是表达来自表达载体的 DHFR 的转染细胞的选择性生长培养基。使用对人 IgG γ 链特异性的 ELISA,筛选度过选择的培养物上清液。在含有 MTX 的 α -MEM+5% dFBS 中,培养产生最高 ELISA 信号的细胞系。MTX 是 DHFR 的抑制剂,其选择由于载体的扩增而生成更高水平的酶的细胞。在不同浓度的 MTX 中培养细胞系,并监测抗体的表达。

[0479] 在一些实施方案中,本发明的组合物和方法可以采用 pA205 载体构建体或其衍生物,例如在 Gion 等人于 2008 年 10 月 2 日提交的 US20080241883 中所述。

[0480] 因此,使用不同的 sORF 设计和构建体,制备了稳定的细胞表达系统。具体地,sORF 系统非常适用于与 CHO 平台相整合,用于表达生物治疗剂诸如抗体分子。

[0481] 实施例 5. sORF 构建体中内含肽 C- 端剪接接头的方面的表征

[0482] 我们在 sORF 构建体的背景下研究了内含肽 C- 端剪接点的方面。我们制备了约 40 个新构建体,以部分地表征与内含肽的下游第一个氨基酸以及会影响在 C- 端剪接点处的切割效率的剪接点长度有关的方面。这些其它的构建体具有轻链接头突变的变化。作为概述,我们聚焦于在轻链的 N- 端末端处或附近的残基;用所有 20 种可能的天然氨基酸残基替换在位置 1 处的甲硫氨酸 (Met1) 和在位置 2 处的天冬氨酸 (Asp2) 中的每一个。参见图 9。这 2 个系列的轻链接头突变构建体使用 D2E7 抗体编码序列和处于 HC- 内含肽 -LC 构型的 Pab lon 内含肽区段。

[0483] 将所述构建体转染进 293 细胞中。使用大多数 Met 置换构建体的瞬时表达产生高 IgG 滴度,这些构建体中的许多产生比对照构建体 (在该位置具有 Met) 更高的表达水平。参见图 10。Asp 置换构建体通常产生更低水平的抗体表达;参见图 11。

[0484] 研究了多蛋白加工的效率。参见,例如,图 12。基于 Met 的变化的构建体文库生成来自多蛋白的 HC 和 LC 的有效加工,这类似于以前描述的 Pab lon HL(-) 构建体。基于 Asp 的变化的构建体文库似乎具有相对受损的 C- 端加工,几乎不产生 LC,并产生显著量的内含肽 -LC 融合蛋白质。该不完全切割的结果被解释为,独立于在剪接点处的氨基酸的性质,因此似乎与一个氨基酸单位的总长度差异有关。但是,应当指出,即使相对无效的构建体仍然可以生成一些 IgG 产物。

[0485] 在一个实验中,进一步分析了来自甲硫氨酸 - 变化文库的 20 个构建体中的 10 个的 IgG 抗体产物。使用蛋白 A 亲和色谱法分批纯化了样品,并通过质谱法 (包括分子量的评价) 分析了轻链组分。质谱法的结果指示,来自某些构建体 (Pab lon M1 A、Pab lon M1 D、Pab lon M1 E、Pab lon M1 F、Pab lon M1 G、Pab lon M1 H、Pab lon M1 I、Pab lon M1 K、Pab lon M1 L 和 Pab lon M1 C) 的抗体轻链如在构建体设计中所工程化的那样形成,通过内含肽介导的反应发生精确切割。在其中用 Cys 替代 Met 的构建体生成更少加工的 HC 和 LC 组分,且含有一种作为剪接产物的蛋白质,这提示,在抗体轻链的 +1 位置处的 Cys 可以在一定程度上支持蛋白剪接。在生成的所有抗体轻链中,LC 的第一个氨基酸的存在证实,使用这些载体生成的抗体产物将在 LCN- 端区域是同源的,并且 LC 对内源的氨基肽酶活性的加工是敏感的。

[0486] 实施例 6. 使用 sORF 构建体表达抗体 ABT-847

[0487] 我们开发了 sORF 构建体,它们适用于表达包含人 λ 轻链的抗体。我们使用了 ABT-847,即具有对抗原白介素 -12 的特异性的全人抗体。该抗体具有人 IgG 1 同种型的重链和 λ 同种型的轻链。关于结合人 IL-12 的人抗体和生产方法,参见 Salfeld 等人于 2005 年 7 月 5 日提交的美国专利 6,914,128。

[0488] 使用同源重组,制备了能够表达 ABT-874 的 5 种 sORF 构建体。图 13 解释了这些构建体的 sORF 组分的结构。所述构建体中的 3 种具有 HC- 内含肽 -LC 构型,2 种构建体具有 LC- 内含肽 -HC 构型。通过瞬时转染,将这些载体导入 HEK293 细胞中,并通过 IgG ELISA,评估它们的抗体表达水平。所述 3 种 HC- 内含肽 -LC 构建体在培养物上清液样品中产生与具有类似构型 (HC- 内含肽 -LC)、但是具有 D2E7 抗体区段的构建体类似的抗体滴度。在

两种情况下,ABT-874 和 D2E7sORF 表达系统采用 Pab Lon HL(-) 方面。具有处于 LC- 内含肽 -HC 构型的 ABT-874 抗体区段的 2 种构建体产生更低水平的 IgG 滴度。

[0489] 通过蛋白 A 亲和色谱法,分批纯化在上清液中生成的 IgG 样品,并通过 SDS-PAGE 电泳进行分析。这些分析揭示,从所有 3 种 HC- 内含肽 -LC 构建体中的多蛋白完全加工出 LC 组分。

[0490] 还使用质谱法表征了 IgG 样品。该分析证实,根据所述构建体设计,从以适当氨基酸开始的 3 种 HC- 内含肽 -LC 构建体生成 LC。该结果与在所述构型中使用的内含肽介导的精确切割相一致。从不具有额外甲硫氨酸残基的 2 个构建体生成的 LC 组分与在来自 ABT-874 抗体的对照样品的材料中相同。因而,尽管可以象关于 D2E7 抗体那样进行修饰,考虑到在来自所述构建体的表达产物中实现的希望的 N- 端 LC 氨基酸序列,这样的修饰是任选的。还使用质谱法分析来评价 HC 的分子量,其表现出根据构建体设计预见到的分子量。

[0491] 实施例 7. 具有基于内含肽的纯化标签的 sORF 构建体

[0492] 在本发明的一些实施方案中,将构建体设计成具有插入物。在具有插入物的构建体的一些实施方案中,插入物能够提供可检测的信号,或可用于提供结合或识别元件。在本发明的一些实施方案中,将构建体设计成促进某些构建体相关的表达产物与一种或多种这样的产物的分离。例如,制备载体设计,以允许完全加工的、装配的多聚抗体产物从组分混合物中的纯化,所述混合物包括来自内含肽剪接反应的部分地加工的蛋白。在表达 HL 构建体的背景下,H- 内含肽 -L 的结构可能导致在 H- 内含肽或内含肽 -L 接头中的一个或两个处的不完全切割反应,因而产生 H- 内含肽、内含肽 -L 或三联体 H- 内含肽 -L 的蛋白副产物,这不同于会产生 H、内含肽和 L 组分的完全有效切割的实现。但是,即使在后一种情况下,去除内含肽组分可能是有用的。因此,形成了下述策略:给内含肽装配标签、优选内部标签,从而实现内含肽切割和 / 或连接反应的至少部分效率。

[0493] 如本文别处所述,在来自 sORF 构建体的培养物上清液样品中,我们已经观察到几种部分地加工的中间体,它们含有与免疫球蛋白重链和 / 或轻链相连的内含肽蛋白。我们已经设计了 sORF 构建体,其中 Pfu lon 和 Pab lon 内含肽含有内部多组氨酸标签 (IHT)。这会提供允许快速地且有效地分离未加工的污染物的组合物和方法。

[0494] 我们已经发现,通过插入肽或大蛋白,可以修饰内含肽。优选地,插入溶剂可接近的环中。通过与结构建模相结合地分析几种内含肽的序列比对,我们在 *Pyrococcus abyssi* (PAB) LON 内含肽和激烈火球菌 (PFU) LON 内含肽内鉴别出溶剂可接近的环。该环位于内切核酸酶 (H) 结构域基序和内含肽的 F/G 块之间 (参见图 14)。图 14 解释了在溶剂可接近的环的优选位置 (虚线箭头) 定位在 H 和 F/G 块之间 (用于导入插入物,包括标签) 的背景下,内含肽的某些结构基序。也参见在因特网网站 http://tools.neb.com/inbase/motifs_endo.php 上可得到的信息 (InBase, The Intein Database:DOD Homing Endonuclease Motifs;InBase Reference:Perler, F.B.,2002, InBase, the Intein Database,Nucleic Acids Res. 30,383-384) (解释包括基序在内的某些内含肽结构特征的示意图来源)。我们已经确定,在指示的结构域之间的区域允许许多可能的插入位点,它们可以用于该溶剂可接近的环内。根据上述来源,在图 14 中,如下指出了保守残基的某些特征:带框的氨基酸,在标准的剪接反应中的亲核基团;大写字母,在标准的内含肽中的保守氨基酸;小写字母,通过改进的机理可以剪接的多态内含肽的氨基酸。

[0495] 使用定位诱变,我们在该环中插入多组氨酸亲和标签,并测试了这些内含肽的发挥功能的能力。IHT 不会破坏 PAB-LON 或 PFU-LON 内含肽的自加工能力,且所述多组氨酸标签可以用于纯化蛋白,因此证实该环是溶剂可接近的。另外,PFU-RIR1-1 结构的晶体结构的检查提示,任意蛋白可以插入在该区域,只要它不会实质上或完全地破坏氨基端和羧基端自催化反应。例如,具有紧密靠近的氨基和羧基端残基的多肽标签或蛋白不会实质上破坏内含肽的自催化活性。在优选的实施方案中,提供了具有插入物(诸如标签)的构建体,其中所述构建体能够表现出一种或多种希望的内含肽活性(例如,切割和/或连接)。因此,可以修饰包括内含肽的该溶剂可接近的环的构建体组分,以建立许多不同的功能分子。

[0496] 对于 Pfu lon 内含肽,插入标签序列。优选的插入位置是这样的位置:其中插入位点的上游氨基酸序列是 IEFIP(AA 323-327),且插入位点的下游氨基酸序列是 ISFSP(AA 328-332)。对于 Pab lon 内含肽,插入标签序列。优选的插入位置是这样的位置:其中插入位点的上游氨基酸序列是 IIFDA(AA 291-295),且插入位点的下游氨基酸序列是 GRLDV(AA 296-300)。在每个 Pfu lon 和 Pab lon 内含肽构建体中,插入的标签序列包括下述的: HHHHHH(SEQ ID NO:56), HHHHHHHHHH(SEQ ID NO:57),和 HHHHHH(SEQ ID NO:58)。如由后一个标签序列(其中 H=组氨酸和 Q=谷氨酰胺)所证实的,插入物可以是多组氨酸标签以外的插入物。如本领域理解的,可以使用其它插入序列。

[0497] IHT-修饰的内含肽 sORF 构建体会产生与没有 IHT 修饰的那些类似的抗体分泌水平。该结果提示,IHT 修饰不会破坏内含肽在 sORF 产物的自加工中起作用的能力,IHT 也不会阻止正确加工的抗体分泌进培养基中。

[0498] 使用固定化的金属亲和色谱法,我们证实,IHT 可以从蛋白 A 纯化的抗体制品中快速地且有效地去除含有污染物的内含肽。这些 IHT 构建体已经使我们能够分离正确地加工的抗体,并代表用于生产 sORF-衍生的生物制品(包括治疗性抗体)的补充方法。

[0499] 使用内部地 His 标记的 sORF 构建体,使用镍柱色谱法技术,以流过模式从含有内含肽的蛋白物质中有效地分离出 D2E7 抗体分子。类似地,使用 Q-柱,也从含有内含肽的蛋白物质中有效地分离出使用 Pab lon HL(-)构建体生产的 D2E7 抗体。还通过 SEC 分级分离分析了下述样品:通过蛋白 A 技术纯化的、从 Pab lon HL(-)生产的 IgG 样品,和通过 proA 和 Ni-树脂纯化的、从 Pab lon HL(-)/10His 构建体生产的 IgG 样品。纯化后的样品表现出提高的单体 IgG 物质的纯度。尺寸排阻色谱法(SEC)进一步去除残余的微量污染物。通过 BiaCore,分析纯化的 IgG 样品对特定抗原 TNFa 的结合亲和力。这些样品的亲和力与使用常规载体生产的 D2E7 抗体不可区分。

[0500] 实施例 8. Pho lon 内含肽

[0501] 掘越氏火球菌 OT3 的 Pho lon 内含肽的氨基酸序列如下所示。也参见 NCBI/蛋白(PH0452)中的登记号 BAA29538.1,根据 Inbase,即 NEB 内含肽数据库。

[0502] 表 25. Pho lon 氨基酸序列(SEQ ID NO:55)。

[0503] QCFSGEEVI IVEKGKDRKVVKLREFVEDALKEPSGEGMDGDIKVITYKDLRGEDVRILTKDGFVKLLYVNK

[0504] REGKQKLRKIVNLKDYLAVTPDHKVFTSEGLKEAGEITEKDEIIRVPLVILDGPKIASTYGEDGKFDD

[0505] YIRWKKYYEKTGNGYKRAAKELNIKESTLRWWTQGAKPNSLKMIEELEKLNLLPLTSEDSRLEKVAIILG

[0506] ALFSDGNIDRNFTLSFISSEKAIERFVETLKELFGEFNIEIRDNHESLGKSILFRTWDRRIIRFFVAL

[0507] GAPVGNKTKVKLELPWWIKLKPSLFLAFMDGLYSGDGSVPRFARYEEGIKFNFTFEIAQLTDDVEKKLPF

[0508] FEEIAWYLSFFGIKAKVRVDKTGDKYKVRLIFSQSIDNVLFLEFIPISLSPAKREKFLREVESYLAAMP

[0509] ESSLAGRIEELREHFNRIKKGERRSF IETWEVVNVTYNVTTETGNLLANGLFVKNS

[0510] 实施例 9. 载体和轻链信号肽

[0511] 在用于表达蛋白的单个开放读码框载体的背景下,已经取得进一步的进展。在一些实施方案中,所述载体采用用于在哺乳动物细胞中表达的免疫球蛋白的蛋白质。在这样的载体中,设计和制备了包括轻链信号肽的轻链组分的构型。

[0512] 在单个开放读码框构建体的一些实施方案中,采用来自天然人抗体轻链来源的信号肽。例如,在 V BASE 中报道了人轻链信号肽,所述 V BASE 包括来自诸如 Genbank 和 EMBL 数据文库等来源的关于人种系可变区序列的信息的数据库。所述 V BASE 数据库是英国剑桥的 MRC Centre for Protein Engineering 的分部(可在因特网地址 <http://vbase.mrc-cpe.cam.ac.uk/> 得到)。也参见 Giudicelli V 等人, Nucleic Acids Research, 2006, 第 34 卷, 数据库号 D781-D784 ; Retter I 等人, Nucleic Acids Res. 2005 年 1 月 1 日 ; 33 (数据库号) : D671-D674。在具体实施方案中,提供了多种载体设计,它们可以用于生产具有不同氨基酸(包括天然氨基酸)的 IgG1 抗体。

[0513] 提供了某些人轻链信号肽,它们通常在 19-23 个氨基酸的长度范围内,并具有在下表中指出的序列(Hu Vk, 人 κ 可变区 ; LCSP, 轻链信号肽)。也提供了变体肽,它们与天然肽(包括人起源的天然肽)相比具有氨基酸序列和 / 或长度的变化。对于给定的氨基酸序列,可以为了相对于一个或多个其它序列进行对比的目的,指示间隙。在一个实施方案中,提供了给定的轻链信号肽或编码它的核酸序列。在一个实施方案中,将氨基酸或核酸序列提供为序列或其区段的合成构建体,诸如在表达载体、合成的(诸如化学合成的)分子或重组表达产物中。

[0514] 表 26. VK 前导序列, 部分 1

人Vk	LCSP 项目	氨基酸序列	SEQ ID NO:
-	-	-20 -10 -1	-
-	-		-
VKI	O12	MDMRVPAQLLGILLLLWLRGARC	59
	O2	MDMRVPAQLLGILLLLWLRGARC	60
	O18	MDMRVPAQLLGILLQLWLSGARC	61
	O8	MDMRVPAQLLGILLLLWLSGARC	62
	A20	MDMRVPAQLLGILLLLWLPDTRC	63
	A30	MDMRVPAQLLGILLLLWFPGARC	64
	L14	MDMRVPAQLLGILLLLWFPGARC	65
	L1	MDMRVLAQLLGILLLLCFPGARC	66
	L15	MDMRVLAQLLGILLLLCFPGARC	67
	L4	MDMRVPAQLLGILLLLWLPGARC	68
	L18	MDMRVPAQLLGILLLLWLPGARC	69
	L5	MDMRVPAQLLGILLLLWFPGSRC	70
	L19	MDMRVPAQLLGILLLLWFPGSRC	71
	L8	MDMRVPAQLLGILLLLWLPGARC	72
	L23	MDMRVPAQRLGILLLLWFPGARC	73
	L9	MRVPAQLLGILLLLWLPGARC	74
	L24	MDMRVPAQLLGILLLLWLPGARC	75
	L11	MDMRVPAQLLGILLLLWLPGARC	76
	L12	MDMRVPAQLLGILLLLWLPGAKC	77
VKII	O11	MRLPAQLLGILLMLWVPGSSE	78
	O1	MRLPAQLLGILLMLWVPGSSE	79
	A17	MRLPAQLLGILLMLWVPGSSG	80
	A1	MRLPAQLLGILLMLWVPGSSG	81
	A18	MRLPAQLLGILLMLWIPGSSA	82
	A2	MRLPAQLLGILLMLWIPGSSA	83
	A19	MRLPAQLLGILLMLWVSGSSG	84
	A3	MRLPAQLLGILLMLWVSGSSG	85
	A23	MRLLAQLLGILLMLWVPGSSG	86
VKIII	A27	METPAQLLFLLLLWLPD TTG	87
	A11	METPAQLLFLLLLWLPD TTG	88
	L2	MEAPAQLLFLLLLWLPD TTG	89
	L16	MEAPAQLLFLLLLWLPD TTG	90
	L6	MEAPAQLLFLLLLWLPD TTG	91
	L20	MEAPAQLLFLLLLWLTDTT G	92
	L25	MEPWKPQHSFFFLLLLLWLPD TTG	93
VKIV	B3	MVLQTQVFISLLLWISGAYG	94
VKV	B2	MGSQVHLLSFLLLWISDTRA	95
VKVI	A26	MLPSQLIGFLLLWVPASRG	96
	A10	MLPSQLIGFLLLWVPASRG	97
	A14	MVSPLQFLRLLLLWVPASRG	98

[0515] 表 27. VK 前导序列, 部分 2

人Vk	LCSP 项目	氨基酸序列	SEQ ID NO:
-	-	-20 -10 -1	-
-	-		-
VKI	L5	MDMRVPAQLLG L L L L W F P G S R C	70
突变体	2G	MDMRVPAQLLG L L L L W F P G S G G	99
	3G	MDMRVPAQLLG L L L L W F P G S G G G	100
	4G	MDMRVPAQLLG L L L L W F P G S G G G G	101
	5G	MDMRVPAQLLG L L L L W F P G S G G G G G	102
	1R	M R M R V P A Q L L G L L L L W F P G S R C	103
	1R2G	M R M R V P A Q L L G L L L L W F P G S G G	104
	2R	M R R M R V P A Q L L G L L L L W F P G S R C	105
	2R2G	M R R M R V P A Q L L G L L L L W F P G S G G	106
	3R2G	M R R R M R V P A Q L L G L L L L W F P G S G G	107
	H2G	MDMRVPAQLLG D E W F P G S G G	108
	-	-	-

[0517] [0518] 用某些 V_K 信号肽替换 L5, 后者是命名为 E7 的免疫球蛋白轻链 (与具有对肿瘤坏死因子 α 的抗原特异性的抗体分子 D2E7 的轻链相对应) 的信号肽。另外, 构建了 L5 的某些突变体的突变文库, 并用突变体 L5 肽替换天然的 L5 肽。使用 *Pyrococcus abyssi* lon 内含肽, 以 HL 方向构建哺乳动物表达载体。制备了下述载体: pTT3-A14-E7-PablonHL,

[0519] pTT3-A17-E7-PablonHL, pTT3-A18-E7-PablonHL, pTT3-A19-E7-PablonHL, pTT3-A23-E7-PablonHL, pTT3-A26-E7-PablonHL, pTT3-A27-E7-PablonHL, pTT3-B2-E7-PablonHL, pTT3-B3-E7-PablonHL, pTT3-L2-E7-PablonHL, pTT3-L20-E7-PablonHL, pTT3-L25-E7-PablonHL, pTT3-mut-1R-E7-PablonHL, pTT3-mut-1R2G-E7-PablonHL, pTT3-mut-2R-E7-PablonHL, pTT3-mut-2G-E7-PablonHL, pTT3-mut-2R2G-E7-PablonHL, pTT3-mut-3G-E7-PablonHL, pTT3-mut-3R2G-E7-PablonHL, pTT3-mut-4G-E7-PablonHL, pTT3-mut-H+2G-E7-PablonHL。在 PTT3 载体骨架上制备这些构建体。该载体具有 EBV 复制起点, 所述 EBV 复制起点允许它在悬浮培养的被感染的 293E 细胞 (表达爱泼斯坦-巴尔病毒核抗原 1 的细胞) 中的附加型扩增。每种载体具有一个由 CMV 启动子驱动的 ORF。在 ORF 中, 所述内含肽序列以 HC-内含肽-LC 的次序插入在抗体重链和轻链之间的框架内。pTT3-A18-E7-PablonHL 的构建体结构的示意图如图 15 所示。

[0520] 通过瞬时转染, 将所述构建体导入 293E 细胞中, 并进行多个瞬时表达实验。在给定的实验中, 在转染后第 8 天, 从转染的细胞的培养物上清液中收集样品, 并分析。通过 IgG ELISA, 评估样品含有的分泌的抗体的水平, 它们的数据显示在下表中, 单位为微克抗体/毫升样品。天然对照是使用 L5LCSP 序列的载体。作为另一种对照 (在表中未显示), 在这些实验中包括常规的双载体系统, 其表达相同的抗体并使用相同的调控元件; 从该对照载体系统生成的抗体分泌水平的范围是 80-206 μg/ml。由几种 sORF 构建体设计 (它们使用这些轻链信号肽) 生成的 IgG 分泌水平与使用常规载体生成的范围量级相当。这些表达水平显著高于使用天然 L5E7 信号肽的水平 (2.0 μg/ml, 使用 Pablon HL(+) 构建体)。这些抗体分泌水平也显著高于使用“2A”技术生成的水平, 后者据报道在哺乳动物细胞中是 1.6 μg/ml (Fang 等人, 2005, Nature Biotechnology 23:584-590)。

[0521] 表 28. 来自 LCSP 构建体的抗体水平。

[0522]

项目	LCSP载体组分	IgG, μg/ml
1	天然对照 (L5)	2.15
2	A14	7.25
3	A17	56.85
4	A18	41.9
5	A19	15.7
6	A23	3.25
7	A26	27.5
8	A27	4.6
9	B2	9.1
10	B3	1.7
11	L2	6.65
12	L20	1.9
13	L25	0.15
14	2G	9.25
15	3G	1.9
16	4G	3.05
17	5G	4.25
18	1R	0.5
19	1R2G	4.25
20	2R	0.1
21	2R2G	1.55
22	3R2G	1.2
23	H2G	99.3

[0523] 在测量其抗体产物水平的指示的构建体中,选择生产最高水平的分泌型抗体的5种构建体的产物,用于进一步分析。所述产物与下述5种构建体相对应:pTT3-A17-E7-PablonHL、pTT3-A18-E7-PablonHL、pTT3-A19-E7-PablonHL、pTT3-A26-E7-PablonHL和pTT3-mutH+2G-E7-PablonHL。通过蛋白A亲和色谱法,纯化从这些构建体生成的分泌型抗体,并在还原SDS-PAGE凝胶上分析,并测定它们的HC和LC的N-端氨基酸序列。使用pTT3-A18-E7-PablonHL生成的样品含有与抗体重链和轻链相对应的蛋白迁移带,它们的迁移与通过传统方法生产的类似抗体不可区分。在还原凝胶上,除了与抗体HC和LC相对应的带以外,还存在2个更高分子量的带,它们似乎对应着未加工的三联体蛋白(HC-内含肽-LC)。如本文所述和根据常规技术,可以方便地去除这样的构建体有关的污染物(作为未加工的或部分地加工的蛋白)。参见图16,它描绘了SDS-PAGE分析结果的一个实施例。通过蛋白A亲和色谱法纯化分泌的IgG抗体,并使用在还原条件下在凝胶中的SDS-PAGE技术进行分离。泳道中的样品从左向右为:泳道(1)SeeBlue Plus2蛋白标准品(Invitrogen)蛋白分子量标志物;(2)对照,用传统的非-sORF表达载体系统生产的D2E7抗体;(3)Pab-lon HL(-);(4)pTT3-A18-E7-PablonHL。

[0524] 还使用针对HC和LC的抗体,通过蛋白印迹分析来分析表达构建体的细胞内产物样品。象在培养的上清液中一样,观察到类似的蛋白物质。通过质谱法分析,确定重链和轻链的N-端氨基酸序列是天然的。分析证实,A18信号肽切割已经精确地发生在轻链的表达所需的正确点处。另外,使用阳离子交换色谱法(CIEX)证实了与传统地表达的D2E7的相似性,所述CIEX基于净表面电荷来分离蛋白,并能够检测D2E7变体和杂质。因此,A18信号肽可以用于抗体表达的sORF载体中,且能够有效地表达完全加工的和装配的抗体产物。

[0525] 除了上述的瞬时表达系统以外,还制备了稳定的细胞系用于表达抗体产物。使用具有A18轻链信号肽组分的载体,用sORF表达构建体制备稳定的CHO细胞系。使用

重组技术,将 sORF 构建体 A18-E7-PablonHL 克隆进质粒 pA190 中。参见图 18,它提供了 pBJ-A18-LC-Pablon-HL(也称作 pA190-A18-E7-PablonHL) 的构建体结构的示意图。通过磷酸钙方法,将该构建体转染进 CHO B3.2 细胞中,并铺板在 48 个 96-孔平板中,在含有 5% FBS 的最低基础培养基 (MEM) Alpha 培养基中。筛选转染样品,并用最大 100nM MTX 进行扩增方法。表征了培养物中的构建体的表达结果。在 100nM MTX,具有构建体 pA190-A18-E7-PablonHL 的细胞在培养物上清液样品中表达在 1.1-16.9 $\mu\text{g/ml}$ 范围内的抗体。通过有限稀释,亚克隆 4 个最高表达克隆。测试亚克隆,并发现以 2.9-31.8 $\mu\text{g/ml}$ 的量表达抗体。4 个具有最高表达水平的亚克隆的细胞系适合悬浮培养,并生成 31-44 $\mu\text{g/ml}$ 的平均量,这从取自第 4 天的培养物的样品测得。

[0526] 评估了所述构建体的产生成熟的轻链产物的能力。参见图 17,它提供了蛋白印迹实验结果。表征了细胞内抗体产物的样品。在还原条件下,根据在凝胶中的 SDS-PAGE,分离整个细胞裂解物,并转移至硝酸纤维素膜,用封闭溶液(在 TTBS 中的脱脂奶粉,所述 TTBS 是含有吐温 20 的三羟甲基氨基甲烷缓冲盐水)封闭,用辣根过氧化物酶缀合的针对重链或轻链的抗体温育,并使用 ECL(增强的化学发光)试剂显影。在印迹泳道中根据构建体命名的样品从左向右是:在左侧未编号的泳道,分子量标志物;(泳道 1)对照,来自 CHO 细胞的 D2E7;(2)对照,在瞬时转染 HEK293 细胞中的 pTT3-A18-E7-PablonHL;(3-11)来自 pA190-A18-E7-PablonHL 的不同克隆,分别对应着克隆编号 1、3、7、9、12、14、18、15 和 13。用字母“a”和“b”标记的箭头指示下述表达产物:(a)上面的带,具有信号肽的轻链,和(b)下面的带,成熟的轻链。在成熟的轻链产物中,信号肽已经被切掉,产生具有比前体更低分子量的产物。结果表明,具有 A18 轻链信号肽组分的构建体能够表达和生产完全成熟的轻链产物,其与 D2E7 抗体相当。

[0527] 关于通过引用并入的声明和变化

[0528] 任意序列信息视作本说明书的一部分。

[0529] 在本申请中提及的所有参考文献(例如专利文件,包括授权的专利或等同物、专利申请公开文本、未公开的专利申请,和非专利文献文件或其它来源材料),均通过引用整体并入本文中,如同单个地通过引用并入。

[0530] 在引用的参考文献和本申请的公开内容出现任何不一致的情况下,以本文的公开内容为准。在本文中提供的一些参考文献通过引用并入,以提供关于原料来源、其它原料、其它试剂、其它合成方法、其它分析方法、其它生物材料、其它细胞和本发明的其它用途的信息,例如,细节。

[0531] 在本文中提及的所有专利和出版物指示本发明所属领域的技术人员的技术水平。在本文中引用的参考文献可以指示截止至它们的公开日或提交日的技术水平,如果需要的话,该信息意图用于本文中,以排除在现有技术中的特定实施方案。例如,当在本文中要求保护物质的组合物时,应当理解,在本文要求保护的物质组合物中无意包括在申请人的发明之前在本领域中已知的和可得到的化合物(包括在本文引用的参考文献中提供了其实现公开内容的化合物)。

[0532] 本申请的任何一个或多个附件通过引用并入,作为说明书和/或附图的一部分。

[0533] 当要求保护化合物、构建体或组合物时,应当理解,无意包括本领域已知的化合物、构建体和组合物(包括在本文公开的参考文献中教导的那些)。当在本文中使用马库什

组或其它分组时,所述组的所有单个成员和从所述组可能产生的所有组合和子组合意图单个地予以阐述,并被包括在公开内容中。

[0534] 在本文使用术语“包括”、“包含”、“包含有”或“含有”的情况下,它们应当解释为说明所提及的所述特征、整数、步骤或组分的存在,但是并不排除一个或多个其它特征、整数、步骤、组分或其组的存在或加入。因而,本文使用的“包含”与“包括”、“含有”、“具有”或“特征在于”同义,且是包括在内的或开放式的,不排除其它未列举的要素或方法步骤。本文使用的“由……组成”排除在权利要求描述中未指出的任意要素、步骤或成分等。本文使用的“基本上由……组成”不排除不会实质上影响权利要求的基础的和新颖的特征(例如,与活性成分有关)的材料或步骤。在本文的每种情况下,任一个术语“包含”、“基本上由……组成”和“由……组成”可以用其它2个术语中的至少任一个替换,从而公开不一定同延的(coextensive)单独的实施方案和/或范围。在没有在本文中具体公开的任何要素或限制存在下,可以适当地实践在本文中例证地描述的本发明的实施方案。当在本文中公开范围(例如,温度范围、时间范围、组合物或浓度范围或其它值范围等)时,意图在公开内容中包括所有中间范围和子范围以及在给定范围内包括的所有单个值。本发明不受公开的实施方案(包括在附图中显示的或在说明书中例证的任一个)的限制,所述实施方案作为实施例或例证给出,而不是进行限制。应该理解,可以从本文的权利要求中排除在本文的说明书中包括的范围或子范围中的任意子范围或单个值。

[0535] 已经参照不同的具体的和/或优选的实施方案和技术对本发明进行了描述。但是,应当理解,可以做出许多改变和改进,同时保留在本发明的精神和范围内。本领域的普通技术人员可以理解,除了在本文中具体描述的那些以外的组合物、方法、装置、装置元件、材料、规程和技术,也可用于实施在本文中广泛地公开的发明,无需求助于过度的实验;这可以扩展至,例如,除了具体例证的那些以外的原料、生物材料、试剂、合成的方法、纯化方法、分析方法、试验方法和生物方法。本发明意图囊括本文所述的前述物质(例如,组合物、方法、装置、装置元件、材料、规程和技术等)的所有本领域已知的功能等价物。已经采用的术语和表述用作描述而不是限制的术语,在这样的术语和表述的应用中,无意排除所示和所述的特征或其部分的任意等效物,但是应当认识到,不同的改进在要求保护的本发明的范围内是可能的。因而,应当理解,尽管实施方案已经具体地公开了本发明,本领域技术人员可以实施优选的实施方案和本文公开的概念的任选的特征、修改和变化,并且这样的修改和变化被视作在所附权利要求书限定的本发明的范围内。

[0536] 参考文献

[0537] 本申请通过引用具体地整体并入下述每一篇:

[0538] Gerald R. Carson 等人于2009年10月30日提交的美国临时专利申请系列号61/256,544;Gerald R. Carson 等人于2010年6月24日提交的美国专利申请系列号12/822,598;Gerald R. Carson 等人于2006年7月21日提交的美国专利申请系列号11/459,098(于2007年3月22日公开为US 20070065912);Gerald R. Carson 等人于2005年7月21日提交的美国临时专利申请系列号60/701,855;和Gerald R. Carson 等人于2006年7月21日提交的PCT国际申请号PCT/US06/28691(于2007年2月1日公开为WO/2007/014162)。

[0539] 美国专利文件:Jacobs, Jr. 等人于1999年11月9日提交的5,981,182;Lorens

等人于 2008 年 3 月 27 日提交的 7105341、7,378,248 ;Belfort 等人于 2005 年 8 月 23 日提交的 6,933,362。Salfeld 等人的于 2000 年 7 月 18 日授权的 US 6,090,382,它是关于结合人 TNF α 的人抗体。Salfeld 等人的于 2005 年 7 月 5 日授权的 US 6,914,128,它是关于结合人 IL-12 的人抗体及其生产方法。Salfeld 等人的于 2001 年 7 月 10 日授权的 US 6,258,562,它是关于结合人 TNF α 的人抗体。

[0540] 美国专利文件 :Jin, Cheng He 等人的 20030036643A1,于 2003 年 2 月 20 日公开 ;Kinsella,Todd M. 的 20050158820A1,于 2005 年 7 月 21 日公开,它是关于环肽的体内生产。Ghayur,Tariq 等人的美国专利申请公开 20050147610,于 2005 年 7 月 7 日公开。Jutila 的 US5,756,095,于 1998 年 5 月 26 日公开,它是关于具有在 E- 选择蛋白和 L- 选择蛋白上的共有表位的特异性的抗体。

[0541] Hoffman、Rebecca S. 等人的 US 20070081996,于 2007 年 4 月 12 日公开,它是关于使用 TNF α 抗体来治疗抑郁症的方法。

[0542] Chen L,Benner J,Perler FB.,Protein splicing in the absence of an intein penultimate histidine. J Biol Chem. 2000 年 7 月 7 日 ;275(27) :20431-5。

[0543] Cohen GN 等 人, An integrated analysis of the genome of the hyperthermophilic archaeon *Pyrococcus abyssi*. Mol Microbiol. 2003 年 3 月 ;47(6) :1495-512。

[0544] Durocher Y, Perret S, Kamen A., Nucleic Acids Res. 2002 Jan 15 ;30(2) :E9, High-level and high-throughput recombinant protein production by transient transfection of suspension-growing human 293-EBNA1 cells。

[0545] Fukui T, Eguchi T, Atomi H, Imanaka T. A membrane-bound archaeal Lon protease displays ATP-independent proteolytic activity towards unfolded proteins and ATP-dependent activity for folded proteins. J Bacteriol. 2002 年 7 月 ;184(13) :3689-98. PMID :12057965。

[0546] Gandor C, 等人, 1995 FEBS Letters 377 :290-294。

[0547] Goddard MR, Burt A. Recurrent invasion and extinction of a selfish gene. Proc Natl Acad Sci U S A. 1999 年 11 月 23 日 ;96(24) :13880-5。

[0548] Gogarten JP, Hilario E. Inteins, introns, and homing endonucleases : recent revelations about the life cycle of parasitic genetic elements. BMC Evol Biol. 2006 年 11 月 13 日 ;6 :94. PMID :17101053。

[0549] Gogarten JP, Senejani AG, Zhaxybayeva O, Olendzenski L, Hilario E. Inteins : structure, function, and evolution. Annu Rev Microbiol. 2002 ;56 :263-87。

[0550] Wood David W 等人的国际公开号 W0/2005/086654, 其国际申请号为 PCT/US2005/005763, 公开日为 2005 年 9 月 22 日。

[0551] Kimball AB, 等 人, Arch Dermatol. 2008 年 2 月 ;144(2) :200-7, Safety and efficacy of ABT-874, a fully human interleukin 12/23 monoclonal antibody, in the treatment of moderate to severe chronic plaque psoriasis : results of a randomized, placebo-controlled, phase 2 trial。

[0552] Liao YD, Jeng JC, Wang CF, Wang SC, Chang ST. Removal of N-terminal

methionine from recombinant proteins by engineered E.coli methionine aminopeptidase. *Protein Sci.* 2004 年 7 月 ;13(7) :1802-10。

[0553] Lecompte, O. ;Ripp, R. ;Puzos-Barbe, V. ;Duprat, S. ;Heilig, R. ;Dietrich, J. ;Thierry, J. C. ;Poch, O. (2001) Genome evolution at the genus level :comparison of three complete genomes of hyperthermophilic archaea. *Genome Res.* 11(6) :981-93.. PubMed ID :11381026。

[0554] Mills Kenneth V., Jennifer S.Manning, Alicia M.Garcia, 和 Lisa A.Wuerdeman, Protein Splicing of a *Pyrococcus abyssi* Intein with a C-terminal Glutamine, *The Journal of Biological Chemistry*, Vol. 279, No. 20, 2004 年 5 月 14 日出版, 第 20685-20691 页。

[0555] Mills Kenneth V., Deirdre M.Dorval, 和 Katherine T.Lewandowski, Kinetic Analysis of the Individual Steps of Protein Splicing for the *Pyrococcus abyssi* PolIII Intein, *The Journal of Biological Chemistry*, Vol. 280, No. 4, 2005 年 1 月 28 日出版, 第 2714-2720 页。

[0556] Powell KT 和 Weaver JC, 1990 *Bio/Technology* 8 :333-337。

[0557] Saves I, Morlot C, Thion L, Rolland JL, Diétrich J, Masson JM, *Nucleic Acids Res.* 2002 Oct 1 ;30(19) :4158-65. Investigating the endonuclease activity of four *Pyrococcus abyssi* inteins。

[0558] Senejani AG, Hilario E, Gogarten JP. The intein of the *Thermoplasma* A-ATPase A subunit :structure, evolution and expression in E.coli. *BMC Biochem.* 2001 ;2 :13. PMID :11722801。

[0559] Southworth MW, Benner J, Perler FB, *EMBO J.* 2000 ;19(18) :5019-26. An alternative protein splicing mechanism for inteins lacking an N-terminal nucleophile。

[0560] Xie, J. ;Juang, J. F. ;Shi, X. F. ;Liu, C. Q. (2001) Analysis of the characteristic sequence of intein and revision of its motifs. *Chinese Sci Bull* 46 :758-761。

[0561] Mannon PJ 等 人, 2004, *N Engl J Med.* 2004 ;351(20) :2069-79, Anti-interleukin-12 antibody for active Crohn' s disease。

[0562] Xu 和 Perler, 1996. *EMBO J.* 15(9), 5146-5153。

[0563] Wu C, 等人, *Nat Biotechnol.* 2007 Nov ;25(11) :1290-7. Simultaneous targeting of multiple disease mediators by a dual-variable-domain immunoglobulin。

[0564] *Molecular Cloning :A Laboratory Manual*, 第 2 版 (Sambrook 等 人, 1989) ; *Oligonucleotide Synthesis* (M. J. Gait 编, 1984) ; *Animal Cell Culture* (R. I. Freshney 编, 1987) ; *Methods in Enzymology* (Academic Press, Inc.) ; *Handbook of Experimental Immunology* (D. M. Weir & C. C. Blackwell 编) ; *Gene Transfer Vectors for Mammalian Cells* (J. M. Miller & M. P. Calos 编, 1987) ; *Current Protocols in Molecular Biology* (F. M. Ausubel 等人编, 1993) ; *PCR :The Polymerase Chain Reaction*, (Mullis 等人编, 1994) ; and *Current Protocols in Immunology* (J. E. Coligan 等人编, 1991)。

[0001]

序列表

<110> APPLICANT: Abbott Laboratories
AGENT/REPRESENTATIVE: Greenlee Sullivan P.C.
CARSON, Gerald R.

<120> SORF 构建体和多基因表达

<130> 54-08 WO

<140> PCT/US _____ (待指定)
<141> 2010-10-28

<150> US 61/256,544
<151> 2009-10-30

<160> 108

<170> PatentIn version 3.5

<210> 1
<211> 335
<212> PRT
<213> Pyrococcus abyssi

<400> 1

Gln Cys Phe Ser Gly Glu Glu Thr Val Val Ile Arg Glu Asn Gly Glu
1 5 10 15

Val Lys Val Leu Arg Leu Lys Asp Phe Val Glu Lys Ala Leu Glu Lys
20 25 30

Pro Ser Gly Glu Gly Leu Asp Gly Asp Val Lys Val Val Tyr His Asp
35 40 45

Phe Arg Asn Glu Asn Val Glu Val Leu Thr Lys Asp Gly Phe Thr Lys
50 55 60

Leu Leu Tyr Ala Asn Lys Arg Ile Gly Lys Gln Lys Leu Arg Arg Val

[0002]

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

[0003]

245	250	255
Asp Pro Asn Lys Asn Leu Pro Phe Phe Gln Glu Leu Ser Trp Tyr Leu		
260	265	270
Gly Leu Phe Gly Ile Lys Ala Asp Ile Lys Val Glu Glu Val Gly Asp		
275	280	285
Lys His Lys Ile Ile Phe Asp Ala Gly Arg Leu Asp Val Asp Lys Gln		
290	295	300
Phe Ile Glu Thr Trp Glu Asp Val Glu Val Thr Tyr Asn Leu Thr Thr		
305	310	315
		320
Glu Lys Gly Asn Leu Leu Ala Asn Gly Leu Phe Val Lys Asn Ser		
325	330	335

<210> 2

<211> 999

<212> DNA

<213> Pyrococcus abyssi

<400> 2

tgcttcagcg gcgaggaaac cgtggtgata cgggagaacg gcgaggtgaa ggtgctgcgg	60
ctgaaggact tcgtggagaa ggccctggaa aagccctccg gcgagggcct ggacggcgac	120
gtgaaagtgg tgtaccacga cttccggaac gagaacgtgg aggtgctgac caaggacggc	180
ttcaccaagc tgctgtacgc caacaagcgg atcggcaagc agaaactgcg gcgggtggtg	240
aacctggaaa aggactactg gttcgccctg acccccgacc acaaggtgta caccaccgac	300
ggcctgaaag aggccggcga gatcaccgag aaggacgagc tgatcagcgt gcccatcacc	360
gtgttcgact gcgaggacga ggacctgaag aagatcggcc tgctgcccct gaccagcgac	420
gacgagcggc tgcggaagat cgccaccctg atgggcatcc tgttcaacgg cggcagcatc	480

[0004]

```

gatgagggcc tgggcgtgct gaccctgaag agcgagcgga gcgtgatcga gaagttcgtg      540
atcacccctga aagagctgtt cggcaagttc gagtacgaga tcatcaaaga ggaaaacacc      600
atcctgaaaa cccgggaccc ccggatcatc aagtttctgg tgggcctggg agcccccatc      660
gagggcaagg atctgaagat gccttggtgg gtgaagctga agcccagcct gttcctggcc      720
ttcctggaag gcttccgggc ccacatcgtg gagcagctgg tcgacgaccc caacaagaat      780
ctgcccttct ttcaggaact gagctggtat ctgggcctgt tcggcatcaa ggccgacatc      840
aaggtggagg aagtgggcga caagcacaag atcatcttcg acgccggcag gctggacgtg      900
gacaagcagt tcatcgagac ctgggaggat gtggaggatga cctacaacct gaccacagag      960
aagggcaatc tgctggccaa cggcctgttc gtgaagaac                                999

```

<210> 3

<211> 333

<212> PRT

<213> *Pyrococcus abyssi*

<400> 3

```

Cys Phe Ser Gly Glu Glu Thr Val Val Ile Arg Glu Asn Gly Glu Val
1           5           10          15

```

```

Lys Val Leu Arg Leu Lys Asp Phe Val Glu Lys Ala Leu Glu Lys Pro
          20           25           30

```

```

Ser Gly Glu Gly Leu Asp Gly Asp Val Lys Val Val Tyr His Asp Phe
          35           40           45

```

```

Arg Asn Glu Asn Val Glu Val Leu Thr Lys Asp Gly Phe Thr Lys Leu
          50           55           60

```

```

Leu Tyr Ala Asn Lys Arg Ile Gly Lys Gln Lys Leu Arg Arg Val Val

```

[0005]

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

[0006]

245	250	255
Pro Asn Lys Asn Leu Pro Phe Phe Gln Glu Leu Ser Trp Tyr Leu Gly		
260	265	270
Leu Phe Gly Ile Lys Ala Asp Ile Lys Val Glu Glu Val Gly Asp Lys		
275	280	285
His Lys Ile Ile Phe Asp Ala Gly Arg Leu Asp Val Asp Lys Gln Phe		
290	295	300
Ile Glu Thr Trp Glu Asp Val Glu Val Thr Tyr Asn Leu Thr Thr Glu		
305	310	315 320
Lys Gly Asn Leu Leu Ala Asn Gly Leu Phe Val Lys Asn		
325	330	
<210> 4		
<211> 403		
<212> PRT		
<213> 激烈火球菌		
<400> 4		
Gln Cys Phe Ser Gly Glu Glu Val Ile Leu Ile Glu Lys Asp Gly Glu		
1	5	10 15
Lys Lys Val Phe Lys Leu Arg Glu Phe Val Asp Gly Leu Leu Lys Glu		
20	25	30
Ala Ser Gly Glu Gly Met Asp Gly Ser Ile Arg Val Val Tyr Lys Asp		
35	40	45
Leu Gln Gly Glu Asn Ile Lys Ile Leu Thr Lys Asp Gly Leu Val Lys		
50	55	60

[0007]

Leu Leu Tyr Val Asn Arg Arg Glu Gly Lys Gln Lys Leu Arg Lys Ile
65 70 75 80

Val Asn Leu Glu Lys Asp Tyr Trp Leu Ala Leu Thr Pro Glu His Lys
85 90 95

Val Tyr Thr Ile Lys Gly Leu Lys Glu Ala Gly Glu Ile Thr Lys Asp
100 105 110

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

Ala Glu Lys Ser Ile Arg Glu Glu Leu Glu Arg Leu Ser Leu Leu Pro
130 135 140

Leu Asn Ser Glu Asp Ser Arg Leu Glu Lys Ile Ala Gly Ile Met Gly
145 150 155 160

Ala Leu Phe Gly Ser Gly Gly Ile Asp Glu Asn Leu Asn Thr Leu Ser
165 170 175

Phe Val Ser Ser Glu Lys Lys Thr Ile Glu Gln Phe Val Lys Ala Leu
180 185 190

Ser Glu Leu Phe Gly Glu Phe Asp Tyr Lys Ile Glu Glu Lys Glu Asn
195 200 205

Ser Ile Ile Phe Arg Thr Cys Asp Lys Arg Ile Val Thr Phe Phe Ala
210 215 220

Thr Leu Gly Ala Pro Val Gly Asp Lys Ser Lys Val Lys Leu Lys Leu
225 230 235 240

[0008]

Pro Trp Trp Val Lys Leu Lys Pro Ser Leu Phe Leu Ala Phe Met Asp
245 250 255

Gly Leu Tyr Ser Ser Asn Arg Asn Asp Lys Glu Ile Leu Glu Ile Thr
260 265 270

Gln Leu Thr Asp Asn Val Glu Thr Phe Phe Glu Glu Ile Ser Trp Tyr
275 280 285

Leu Ser Phe Phe Gly Ile Lys Ala Glu Ala Glu Glu Asp Glu Glu Lys
290 295 300

Asp Lys Tyr Arg Ala Arg Leu Thr Leu Ser Ser Ser Ile Asp Asn Met
305 310 315 320

Leu Asn Phe Ile Glu Phe Ile Pro Ile Ser Phe Ser Pro Ala Lys Arg
325 330 335

Glu Lys Phe Phe Lys Glu Ile Glu Lys Tyr Leu Glu Tyr Ser Ile Pro
340 345 350

Glu Lys Thr Glu Asp Leu Lys Lys Arg Val Lys Arg Val Lys Lys Gly
355 360 365

Glu Arg Arg Asn Phe Leu Glu Ser Trp Glu Glu Val Glu Val Thr Tyr
370 375 380

Asn Val Thr Thr Glu Thr Gly Asn Leu Leu Ala Asn Gly Leu Phe Val
385 390 395 400

Lys Asn Ser

[0009]

<210>	5	
<211>	1203	
<212>	DNA	
<213>	激烈火球菌	
<400>	5	
tgttttagcg gtgaagaagt tatcttaatt gaaaaggacg gagagaaaaa agtcttcaaa	60	
cttagggagt tcgttgacgg tctccttaag gaggcgtctg gagaagggat ggacggaagt	120	
attagagtag ttataaaga tcttcaaggg gaaaacataa aaatactcac aaaagacgga	180	
cttgtaaagc tcctttatgt caatagaaga gaagggaagc aaaagcttag aaaaatagta	240	
aatcttgaag aggattattg gcttgcatga acacctgaac ataaagtgt cacaataaag	300	
ggccttaaag aagctggaga gataactaaa gatgatgaga taataagagt gcctctcaca	360	
attcttgacg gctttgacgt agccgagaag agtataagag aggaacttga aaggcttagc	420	
ctacttcac taaatagtga agacagtaga ctagaaaaga tagcaggaat catgggcgca	480	
ctctttggta gtggaggtat cgatgagaat ctcaataccc ttagctttgt ttctagcgag	540	
aagaaaacaa ttgaacagtt tgttaaagca ctcagcgagc tcttcgggga atttgactat	600	
aaaattgaag aaaaagaaaa cagcattatt ttcagaacat gtgataaaag aatagtgaac	660	
ttctttgcta cacttggtgc accagttgga gacaaaagca aagttaagct taagcttcca	720	
tggtgggtca agcttaagcc gtcacttttc ctcgccttca tggatggtct ctacagtagc	780	
aataggaatg acaaagaaat cctcgaaata actcaactta ctgacaacgt cgaaacgttc	840	
ttcgaggaaa tatcttggtg tctgagcttc ttggaatta aggcagaagc tgaagaggat	900	
gaagaaaaag ataaatacag ggctagactt acgctatcct catcaataga caacatgctt	960	
aatttcattg agttcattcc aataagcttt tctccagcaa agagagaaaa attctttaag	1020	
gaaattgaaa aatatctgga atatagcatt cccgaaaaga ctgaggatct taagaaacga	1080	

[0010]

gttaagagag ttaagaaggg agagagaagg aatttcctcg aaagctggga ggaagttgaa 1140
gttacttaca acgtaactac agagacagga aatctacttg ctaacggtct atttgtaag 1200
aac 1203

<210> 6

<211> 401

<212> PRT

<213> 激烈火球菌

<400> 6

Cys Phe Ser Gly Glu Glu Val Ile Leu Ile Glu Lys Asp Gly Glu Lys
1 5 10 15

Lys Val Phe Lys Leu Arg Glu Phe Val Asp Gly Leu Leu Lys Glu Ala
20 25 30

Ser Gly Glu Gly Met Asp Gly Ser Ile Arg Val Val Tyr Lys Asp Leu
35 40 45

Gln Gly Glu Asn Ile Lys Ile Leu Thr Lys Asp Gly Leu Val Lys Leu
50 55 60

Leu Tyr Val Asn Arg Arg Glu Gly Lys Gln Lys Leu Arg Lys Ile Val
65 70 75 80

Asn Leu Glu Lys Asp Tyr Trp Leu Ala Leu Thr Pro Glu His Lys Val
85 90 95

Tyr Thr Ile Lys Gly Leu Lys Glu Ala Gly Glu Ile Thr Lys Asp Asp
100 105 110

Glu Ile Ile Arg Val Pro Leu Thr Ile Leu Asp Gly Phe Asp Val Ala
115 120 125

[0011]

Glu Lys Ser Ile Arg Glu Glu Leu Glu Arg Leu Ser Leu Leu Pro Leu
130 135 140

Asn Ser Glu Asp Ser Arg Leu Glu Lys Ile Ala Gly Ile Met Gly Ala
145 150 155 160

Leu Phe Gly Ser Gly Gly Ile Asp Glu Asn Leu Asn Thr Leu Ser Phe
165 170 175

Val Ser Ser Glu Lys Lys Thr Ile Glu Gln Phe Val Lys Ala Leu Ser
180 185 190

Glu Leu Phe Gly Glu Phe Asp Tyr Lys Ile Glu Glu Lys Glu Asn Ser
195 200 205

Ile Ile Phe Arg Thr Cys Asp Lys Arg Ile Val Thr Phe Phe Ala Thr
210 215 220

Leu Gly Ala Pro Val Gly Asp Lys Ser Lys Val Lys Leu Lys Leu Pro
225 230 235 240

Trp Trp Val Lys Leu Lys Pro Ser Leu Phe Leu Ala Phe Met Asp Gly
245 250 255

Leu Tyr Ser Ser Asn Arg Asn Asp Lys Glu Ile Leu Glu Ile Thr Gln
260 265 270

Leu Thr Asp Asn Val Glu Thr Phe Phe Glu Glu Ile Ser Trp Tyr Leu
275 280 285

Ser Phe Phe Gly Ile Lys Ala Glu Ala Glu Glu Asp Glu Glu Lys Asp
290 295 300

[0012]

Lys Tyr Arg Ala Arg Leu Thr Leu Ser Ser Ser Ile Asp Asn Met Leu
305 310 315 320

Asn Phe Ile Glu Phe Ile Pro Ile Ser Phe Ser Pro Ala Lys Arg Glu
325 330 335

Lys Phe Phe Lys Glu Ile Glu Lys Tyr Leu Glu Tyr Ser Ile Pro Glu
340 345 350

Lys Thr Glu Asp Leu Lys Lys Arg Val Lys Arg Val Lys Lys Gly Glu
355 360 365

Arg Arg Asn Phe Leu Glu Ser Trp Glu Glu Val Glu Val Thr Tyr Asn
370 375 380

Val Thr Thr Glu Thr Gly Asn Leu Leu Ala Asn Gly Leu Phe Val Lys
385 390 395 400

Asn

<210> 7
<211> 333
<212> PRT
<213> Pyrococcus abyssi

<400> 7

Cys Phe Ser Gly Glu Glu Thr Val Val Ile Arg Glu Asn Gly Glu Val
1 5 10 15

Lys Val Leu Arg Leu Lys Asp Phe Val Glu Lys Ala Leu Glu Lys Pro
20 25 30

[0013]

Ser Gly Glu Gly Leu Asp Gly Asp Val Lys Val Val Tyr His Asp Phe
35 40 45

Arg Asn Glu Asn Val Glu Val Leu Thr Lys Asp Gly Phe Thr Lys Leu
50 55 60

Leu Tyr Ala Asn Lys Arg Ile Gly Lys Gln Lys Leu Arg Arg Val Val
65 70 75 80

Asn Leu Glu Lys Asp Tyr Trp Phe Ala Leu Thr Pro Asp His Lys Val
85 90 95

Tyr Thr Thr Asp Gly Leu Lys Glu Ala Gly Glu Ile Thr Glu Lys Asp
100 105 110

Glu Leu Ile Ser Val Pro Ile Thr Val Phe Asp Cys Glu Asp Glu Asp
115 120 125

Leu Lys Lys Ile Gly Leu Leu Pro Leu Thr Ser Asp Asp Glu Arg Leu
130 135 140

Arg Lys Ile Ala Thr Leu Met Gly Ile Leu Phe Asn Gly Gly Ser Ile
145 150 155 160

Asp Glu Gly Leu Gly Val Leu Thr Leu Lys Ser Glu Arg Ser Val Ile
165 170 175

Glu Lys Phe Val Ile Thr Leu Lys Glu Leu Phe Gly Lys Phe Glu Tyr
180 185 190

Glu Ile Ile Lys Glu Glu Asn Thr Ile Leu Lys Thr Arg Asp Pro Arg
195 200 205

[0014]

Ile Ile Lys Phe Leu Val Gly Leu Gly Ala Pro Ile Glu Gly Lys Asp
210 215 220

Leu Lys Met Pro Trp Trp Val Lys Leu Lys Pro Ser Leu Phe Leu Ala
225 230 235 240

Phe Leu Glu Gly Phe Arg Ala His Ile Val Glu Gln Leu Val Asp Asp
245 250 255

Pro Asn Lys Asn Leu Pro Phe Phe Gln Glu Leu Ser Trp Tyr Leu Gly
260 265 270

Leu Phe Gly Ile Lys Ala Asp Ile Lys Val Glu Glu Val Gly Asp Lys
275 280 285

His Lys Ile Ile Phe Asp Ala Gly Arg Leu Asp Val Asp Lys Gln Phe
290 295 300

Ile Glu Thr Trp Glu Asp Val Glu Val Thr Tyr Asn Leu Thr Thr Glu
305 310 315 320

Lys Gly Asn Leu Leu Ala Asn Gly Leu Phe Val Lys Asn
325 330

<210> 8

<211> 999

<212> DNA

<213> *Pyrococcus abyssi*

<400> 8

tgcttcagcg gcgaggaaac cgtggtgatc cgggagaacg gcgaggtgaa ggtgctgcgg 60

ctgaaggact tcgtggagaa ggccctggaa aagccctccg gcgagggcct ggacggcgac 120

gtgaaagtgg tgtaccacga cttccggaac gagaacgtgg aggtgctgac caaggacggc 180

[0015]

ttcaccaagc tgctgtacgc caacaagcgg atcggcaagc agaaactgcg gcgggtggtg	240
aacctggaaa aggactactg gttcgccctg acccccgacc acaagggtgta caccaccgac	300
ggcctgaaag aggccggcga gatcaccgag aaggacgagc tgatcagcgt gcccatcacc	360
gtgttcgact gcgaggacga ggacctgaag aagatcggcc tgctgcccct gaccagcgac	420
gacgagcggc tgcggaagat cgccaccctg atgggcatcc tgttcaacgg cggcagcatc	480
gatgagggcc tgggcgtgct gaccctgaag agcagcgga gcgtgatcga gaagttcgtg	540
atcacctga aagagctgtt cggcaagttc gagtacgaga tcatcaaaga ggaaaacacc	600
atcctgaaaa cccgggaccc ccggatcacc aagtttctgg tgggcctggg agccccatc	660
gagggcaagg atctgaagat gccttggtgg gtgaagctga agcccagcct gttcctggcc	720
ttcctggaag gcttcgggc ccacatcgtg gagcagctgg tcgacgaccc caacaagaat	780
ctgcccttct ttcaggaact gagctggtat ctgggcctgt tcggcatcaa ggccgacatc	840
aagggtggagg aagtgggcga caagcacaag atcatcttcg acgccggcag gctggacgtg	900
gacaagcagt tcatcgagac ctgggaggat gtggaggtga cctacaacct gaccacagag	960
aagggcaatc tgctggccaa cggcctgttc gtgaagaac	999

<210> 9

<211> 10

<212> PRT

<213> 人工序列

<220>

<223> 合成构建体

<400> 9

Glu Val Gln Leu Val Glu Ser Gly Gly Gly

1

5

10

[0016]

<210> 10
<211> 10
<212> PRT
<213> 人工序列

<220>
<223> 合成构建体

<400> 10

Met Glu Val Gln Leu Val Glu Ser Gly Gly
1 5 10

<210> 11
<211> 10
<212> PRT
<213> 人工序列

<220>
<223> 合成构建体

<400> 11

Asp Ile Gln Met Thr Gln Ser Pro Ser Ser
1 5 10

<210> 12
<211> 10
<212> PRT
<213> 人工序列

<220>
<223> 合成构建体

<400> 12

Met Asp Ile Gln Met Thr Gln Ser Pro Ser
1 5 10

<210> 13
<211> 8

[0017]

<212> PRT

<213> 人工序列

<220>

<223> 合成构建体

<400> 13

Ala Asn Gly Leu Phe Val Lys Asn

1

5

<210> 14

<211> 5

<212> PRT

<213> 人工序列

<220>

<223> 合成构建体

<400> 14

Met Arg Ala Lys Arg

1

5

<210> 15

<211> 8

<212> PRT

<213> 人工序列

<220>

<223> 合成构建体

<400> 15

His Ala Arg Gly Val Phe Arg Arg

1

5

<210> 16

<211> 8

<212> PRT

<213> 人工序列

[0018]

<220>

<223> 合成构建体

<400> 16

Met Asp Arg Gly Val Phe Arg Arg
1 5

<210> 17

<211> 7

<212> PRT

<213> 人工序列

<220>

<223> 合成构建体

<400> 17

Asp Ile Gln Met Thr Gln Ser
1 5

<210> 18

<211> 7

<212> PRT

<213> 人工序列

<220>

<223> 合成构建体

<400> 18

Ala Ile Gln Met Thr Gln Ser
1 5

<210> 19

<211> 7

<212> PRT

<213> 人工序列

<220>

[0019]

<223> 合成构建体

<400> 19

Asn Ile Gln Met Thr Gln Ser
1 5

<210> 20

<211> 7

<212> PRT

<213> 人工序列

<220>

<223> 合成构建体

<400> 20

Asn Phe Gln Met Thr Gln Ser
1 5

<210> 21

<211> 8

<212> PRT

<213> 人工序列

<220>

<223> 合成构建体

<400> 21

Met Asp Ile Gln Met Thr Gln Ser
1 5

<210> 22

<211> 12

<212> PRT

<213> 人工序列

<220>

<223> 合成构建体

[0020]

<400> 22

Met Arg Ala Lys Arg Asp Ile Gln Met Thr Gln Ser
1 5 10

<210> 23

<211> 9

<212> PRT

<213> 人工序列

<220>

<223> 合成构建体

<400> 23

Tyr Pro Asp Ile Gln Met Thr Gln Ser
1 5

<210> 24

<211> 9

<212> PRT

<213> 人工序列

<220>

<223> 合成构建体

<400> 24

Arg Pro Asp Ile Gln Met Thr Gln Ser
1 5

<210> 25

<211> 9

<212> PRT

<213> 人工序列

<220>

<223> 合成构建体

<400> 25

[0021]

Val Pro Asp Ile Gln Met Thr Gln Ser
1 5

<210> 26
<211> 9
<212> PRT
<213> 人工序列

<220>
<223> 合成构建体

<400> 26

Gln Pro Asp Ile Gln Met Thr Gln Ser
1 5

<210> 27
<211> 9
<212> PRT
<213> 人工序列

<220>
<223> 合成构建体

<400> 27

Ala Pro Asp Ile Gln Met Thr Gln Ser
1 5

<210> 28
<211> 9
<212> PRT
<213> 人工序列

<220>
<223> 合成构建体

<400> 28

His Ala Asp Ile Gln Met Thr Gln Ser
1 5

[0022]

<210> 29
<211> 9
<212> PRT
<213> 人工序列

<220>
<223> 合成构建体

<400> 29

Tyr Ala Asp Ile Gln Met Thr Gln Ser
1 5

<210> 30
<211> 9
<212> PRT
<213> 人工序列

<220>
<223> 合成构建体

<400> 30

Met Pro Asp Ile Gln Met Thr Gln Ser
1 5

<210> 31
<211> 9
<212> PRT
<213> 人工序列

<220>
<223> 合成构建体

<400> 31

Met Ala Asp Ile Gln Met Thr Gln Ser
1 5

[0023]

<210> 32

<211> 15

<212> PRT

<213> 人工序列

<220>

<223> 合成构建体

<400> 32

His	Ala	Arg	Gly	Val	Phe	Arg	Arg	Asp	Ile	Gln	Met	Thr	Gln	Ser
1				5					10					15

<210> 33

<211> 15

<212> PRT

<213> 人工序列

<220>

<223> 合成构建体

<400> 33

Met	Asp	Arg	Gly	Val	Phe	Arg	Arg	Asp	Ile	Gln	Met	Thr	Gln	Ser
1				5					10					15

<210> 34

<211> 504

<212> DNA

<213> 詹氏甲烷球菌

<400> 34

gctctggcct acgacgagcc catctacctg agcgacggca acatcatcaa catcggcgag 60

ttcgtggaca agttcttcaa gaagtacaag aacagcatca agaaagagga caacggcttc 120

ggctggatcg acatcggaac cgagaacatc tacatcaaga gcttcaacaa gctgtccctg 180

atcatcgagg acaagcggat cctgagagtg tggcggaaga agtacagcgg caagctgac 240

aagatcacca ccaagaaccg gcgggagatc accctgaccc acgaccaccc cgtgtacatc 300

[0024]

agcaagaccg gcgaggtgct ggaaatcaac gccgagatgg tgaaagtggg cgactacatc 360
tatatcccca agaacaacac catcaacctg gacgaggtga tcaaggtgga gaccgtggac 420
tacaacggcc acatctacga cctgaccgtg gaggacaacc acacctacat cgccggcaag 480
aacgagggct tcgccgtgag caac 504

<210> 35

<211> 168

<212> PRT

<213> 詹氏甲烷球菌

<400> 35

Ala Leu Ala Tyr Asp Glu Pro Ile Tyr Leu Ser Asp Gly Asn Ile Ile
1 5 10 15

Asn Ile Gly Glu Phe Val Asp Lys Phe Phe Lys Lys Tyr Lys Asn Ser
20 25 30

Ile Lys Lys Glu Asp Asn Gly Phe Gly Trp Ile Asp Ile Gly Asn Glu
35 40 45

Asn Ile Tyr Ile Lys Ser Phe Asn Lys Leu Ser Leu Ile Ile Glu Asp
50 55 60

Lys Arg Ile Leu Arg Val Trp Arg Lys Lys Tyr Ser Gly Lys Leu Ile
65 70 75 80

Lys Ile Thr Thr Lys Asn Arg Arg Glu Ile Thr Leu Thr His Asp His
85 90 95

Pro Val Tyr Ile Ser Lys Thr Gly Glu Val Leu Glu Ile Asn Ala Glu
100 105 110

[0025]

Met Val Lys Val Gly Asp Tyr Ile Tyr Ile Pro Lys Asn Asn Thr Ile
 115 120 125

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

Ile Tyr Asp Leu Thr Val Glu Asp Asn His Thr Tyr Ile Ala Gly Lys
 145 150 155 160

Asn Glu Gly Phe Ala Val Ser Asn
 165

<210> 36
 <211> 588
 <212> DNA
 <213> Pyrococcus abyssi

<400> 36
 gctctgtact acttcagcga gatccagctg cccaacggca aagagttcat cggcaaactg 60
 gtggacgagc tgttcgagaa gtaccacgac aagatcggca agtacaagga catggaatac 120
 gtggagctga acgaagagga caccttcgag gtgatcagca tcggccccga cctgagcgcc 180
 aggcggcaca aggtgaccca cgtgtggcgg cggaagggtga aagacggcga gaagctgggtg 240
 aagatccgga ccgccagcgg caaagaactg gtgctgaccc aggaccaccc cgtgttcgtg 300
 ctgctgggcc gggacgtggc cagacgggac gccggcaacg tgaaagtggg cgacgagatc 360
 gccgtgctga acaccaggcc cgacttcagc gtgctgtccc cccctgccat gcccgagctg 420
 ctgtccgagc ccttcaacta cgagctgtcc agcatcggcg acgtggcctg ggacgaggtg 480
 gtggaggtgg acgagatcga cgccaagggc ctgggcgtgg agtacctgta cgacctgacc 540
 gtggacatca accacaacta cgtggccaac ggcatcgtgg tgtccaac 588

<210> 37

[0026]

<400> 37

000

<210> 38

<211> 1566

<212> DNA

<213> 激烈火球菌

<400> 38

gcacttttacg atttctctgt catccaacta tctaattggtg gatttgtact tataggagat	60
ttagtcgagg aattattcaa gaagtatgcc gagaaaatta aaacatacaa agaccttgag	120
tacatagagc ttaacgagga agaccgtttt gaagttgta gtgttagtcc agatttgaag	180
gctaataaac atgttgtctc aagagtttgg agaagaaagg tcagagaggg ggaaaagcta	240
atacgcataa agacgagaac tggcaacgaa ataatcctca ctagaaatca tccgctatit	300
gccttctcca atggagacgt agtcagaaaa gaggccgaga agctcaaagt tggggataga	360
gttgcagtga tgatgagacc tccttcacct cctcaaacta aagctglagt tgaccctgca	420
atttacgtga aaataagtga ttactacctt gtccgaacg gaaaaggat gataaaagtt	480
cctaacgatg gtattcctcc agaaaaggcc caatatcttc tttcagtaaa ttcatactct	540
gtaaaattag tcagagaagt tgatgagaag ttatcctatc tcgctggagt tatactcgg	600
gatgggtata tatcatcgaa tggatactac atctcagcta catttgacga cgaagcttac	660
atggatgcct ttgtctctgt agtctcgac tttatcccta actatgtccc cagtataagg	720
aagaacggag attacacaat tgtaactgtt ggctcgaaga ttttgctga aatgctctca	780
aggatatattg gaataccaag gggcagaaaa tctatgtggg atattccaga cgtagtactt	840
tcaaatgacg atcttatgag atacttcata gctggacttt tcgacgctga tgggtacgta	900
gatgaaaatg ggccctccat agtcctagta acaaagagt aaaccgtggc aaggaagatt	960
tggtagcttc ttcagaggtt ggggatcata agtacagttt cccgtgtaaa gagcagaggg	1020

[0027]

tttaaagaag gcgagctggt cagggttaatt attagtgggtg ttgaagatct tgctaaattt	1080
gcaaaattca taccctacg tcactcaaga aagaggcca aacttatgga gatattaagg	1140
actaagaagc catatcgggg aagaagaact taccgcgtgc cgatatccag tgatatgata	1200
gctcctctcc gtcaaatgtt gggattaact gttgcagagc tgtctaagtt agcgtcttat	1260
tatgcagggg aaaaagtctc tgaaagccta attaggcata tagaaaaggg aagggtcaaa	1320
gagataagac gctctacgct caaggggatt gcccttgctc tccagcagat agctaaagat	1380
gtgggtaacg aagaagcttg ggtgagagcc aagaggcttc aattgatagc tgaggagat	1440
gtttactggg atgaagtcgt aagtgttgag gaagttgatc cgaaggagct tggcattgag	1500
tacgtctatg acctcacggt tgaggacgac cacaattatg tggcaaatgg catactagtc	1560
tcaaac	1566

<210> 39

<211> 522

<212> PRT

<213> 激烈火球菌

<400> 39

Ala	Leu	Tyr	Asp	Phe	Ser	Val	Ile	Gln	Leu	Ser	Asn	Gly	Arg	Phe	Val
1				5					10					15	

Leu	Ile	Gly	Asp	Leu	Val	Glu	Glu	Leu	Phe	Lys	Lys	Tyr	Ala	Glu	Lys
		20						25						30	

Ile	Lys	Thr	Tyr	Lys	Asp	Leu	Glu	Tyr	Ile	Glu	Leu	Asn	Glu	Glu	Asp
		35						40						45	

Arg	Phe	Glu	Val	Val	Ser	Val	Ser	Pro	Asp	Leu	Lys	Ala	Asn	Lys	His
		50					55						60		

[0028]

Val Val Ser Arg Val Trp Arg Arg Lys Val Arg Glu Gly Glu Lys Leu
65 70 75 80

Ile Arg Ile Lys Thr Arg Thr Gly Asn Glu Ile Ile Leu Thr Arg Asn
85 90 95

His Pro Leu Phe Ala Phe Ser Asn Gly Asp Val Val Arg Lys Glu Ala
100 105 110

Glu Lys Leu Lys Val Gly Asp Arg Val Ala Val Met Met Arg Pro Pro
115 120 125

Ser Pro Pro Gln Thr Lys Ala Val Val Asp Pro Ala Ile Tyr Val Lys
130 135 140

Ile Ser Asp Tyr Tyr Leu Val Pro Asn Gly Lys Gly Met Ile Lys Val
145 150 155 160

Pro Asn Asp Gly Ile Pro Pro Glu Lys Ala Gln Tyr Leu Leu Ser Val
165 170 175

Asn Ser Tyr Pro Val Lys Leu Val Arg Glu Val Asp Glu Lys Leu Ser
180 185 190

Tyr Leu Ala Gly Val Ile Leu Gly Asp Gly Tyr Ile Ser Ser Asn Gly
195 200 205

Tyr Tyr Ile Ser Ala Thr Phe Asp Asp Glu Ala Tyr Met Asp Ala Phe
210 215 220

Val Ser Val Val Ser Asp Phe Ile Pro Asn Tyr Val Pro Ser Ile Arg
225 230 235 240

[0029]

Lys Asn Gly Asp Tyr Thr Ile Val Thr Val Gly Ser Lys Ile Phe Ala
245 250 255

Glu Met Leu Ser Arg Ile Phe Gly Ile Pro Arg Gly Arg Lys Ser Met
260 265 270

Trp Asp Ile Pro Asp Val Val Leu Ser Asn Asp Asp Leu Met Arg Tyr
275 280 285

Phe Ile Ala Gly Leu Phe Asp Ala Asp Gly Tyr Val Asp Glu Asn Gly
290 295 300

Pro Ser Ile Val Leu Val Thr Lys Ser Glu Thr Val Ala Arg Lys Ile
305 310 315 320

Trp Tyr Val Leu Gln Arg Leu Gly Ile Ile Ser Thr Val Ser Arg Val
325 330 335

Lys Ser Arg Gly Phe Lys Glu Gly Glu Leu Phe Arg Val Ile Ile Ser
340 345 350

Gly Val Glu Asp Leu Ala Lys Phe Ala Lys Phe Ile Pro Leu Arg His
355 360 365

Ser Arg Lys Arg Ala Lys Leu Met Glu Ile Leu Arg Thr Lys Lys Pro
370 375 380

Tyr Arg Gly Arg Arg Thr Tyr Arg Val Pro Ile Ser Ser Asp Met Ile
385 390 395 400

Ala Pro Leu Arg Gln Met Leu Gly Leu Thr Val Ala Glu Leu Ser Lys
405 410 415

[0030]

Leu Ala Ser Tyr Tyr Ala Gly Glu Lys Val Ser Glu Ser Leu Ile Arg
420 425 430

His Ile Glu Lys Gly Arg Val Lys Glu Ile Arg Arg Ser Thr Leu Lys
435 440 445

Gly Ile Ala Leu Ala Leu Gln Gln Ile Ala Lys Asp Val Gly Asn Glu
450 455 460

Glu Ala Trp Val Arg Ala Lys Arg Leu Gln Leu Ile Ala Glu Gly Asp
465 470 475 480

Val Tyr Trp Asp Glu Val Val Ser Val Glu Glu Val Asp Pro Lys Glu
485 490 495

Leu Gly Ile Glu Tyr Val Tyr Asp Leu Thr Val Glu Asp Asp His Asn
500 505 510

Tyr Val Ala Asn Gly Ile Leu Val Ser Asn
515 520

<210> 40

<211> 4

<212> PRT

<213> 人工序列

<220>

<223> 合成构建体

<400> 40

Gly His Asp Gly

1

<210> 41

<211> 4

[0031]

<212> PRT

<213> 人工序列

<220>

<223> 合成构建体

<400> 41

Ser Pro Gly Lys

1

<210> 42

<211> 4

<212> PRT

<213> 人工序列

<220>

<223> 合成构建体

<400> 42

Ala Leu Tyr Tyr

1

<210> 43

<211> 4

<212> PRT

<213> 人工序列

<220>

<223> 合成构建体

<400> 43

Cys Leu Tyr Tyr

1

<210> 44

<211> 4

<212> PRT

<213> 人工序列

[0032]

<220>

<223> 合成构建体

<400> 44

Cys Met Gly Thr

1

<210> 45

<211> 4

<212> PRT

<213> 人工序列

<220>

<223> 合成构建体

<400> 45

Met Asp Ile Gln

1

<210> 46

<211> 7

<212> PRT

<213> 人工序列

<220>

<223> 合成构建体

<400> 46

Leu Ser Leu Ser Pro Gly Lys

1

5

<210> 47

<211> 6

<212> PRT

<213> 人工序列

<220>

[0033]

<223> 合成构建体

<400> 47

Leu Ser Leu Ser Pro Gly
1 5

<210> 48

<211> 9

<212> PRT

<213> 人工序列

<220>

<223> 合成构建体

<400> 48

Ala Leu Tyr Tyr Phe Ser Glu Ile Gln
1 5

<210> 49

<211> 44

<212> DNA

<213> 人工序列

<220>

<223> 合成构建体

<400> 49

gcctctccct gtctccgggt gctctgtact acttcagcga gatc

44

<210> 50

<211> 44

<212> DNA

<213> 人工序列

<220>

<223> 合成构建体

<400> 50

gcctctccct gtctccgggt tgtctgtact acttcagcga gatc

44

[0034]

<210> 51

<211> 44

<212> DNA

<213> 人工序列

<220>

<223> 合成构建体

<400> 51

tctccctgtc tccgggtaaa tgtctgtact acttcagcga gatc

44

<210> 52

<211> 20

<212> DNA

<213> 人工序列

<220>

<223> 合成构建体

<400> 52

cggcgtggag gtgcataatg

20

<210> 53

<211> 18

<212> DNA

<213> 人工序列

<220>

<223> 合成构建体

<400> 53

acccggagac agggagag

18

<210> 54

<211> 20

<212> DNA

<213> 人工序列

<220>

[0035]

<223> 合成构建体

<400> 54

gggtcagcac cagttctttg

20

<210> 55

<211> 476

<212> PRT

<213> 掘越氏火球菌

<400> 55

Gln Cys Phe Ser Gly Glu Glu Val Ile Ile Val Glu Lys Gly Lys Asp
1 5 10 15

Arg Lys Val Val Lys Leu Arg Glu Phe Val Glu Asp Ala Leu Lys Glu
20 25 30

Pro Ser Gly Glu Gly Met Asp Gly Asp Ile Lys Val Thr Tyr Lys Asp
35 40 45

Leu Arg Gly Glu Asp Val Arg Ile Leu Thr Lys Asp Gly Phe Val Lys
50 55 60

Leu Leu Tyr Val Asn Lys Arg Glu Gly Lys Gln Lys Leu Arg Lys Ile
65 70 75 80

Val Asn Leu Asp Lys Asp Tyr Trp Leu Ala Val Thr Pro Asp His Lys
85 90 95

Val Phe Thr Ser Glu Gly Leu Lys Glu Ala Gly Glu Ile Thr Glu Lys
100 105 110

Asp Glu Ile Ile Arg Val Pro Leu Val Ile Leu Asp Gly Pro Lys Ile
115 120 125

[0036]

Ala Ser Thr Tyr Gly Glu Asp Gly Lys Phe Asp Asp Tyr Ile Arg Trp
130 135 140

Lys Lys Tyr Tyr Glu Lys Thr Gly Asn Gly Tyr Lys Arg Ala Ala Lys
145 150 155 160

Glu Leu Asn Ile Lys Glu Ser Thr Leu Arg Trp Trp Thr Gln Gly Ala
165 170 175

Lys Pro Asn Ser Leu Lys Met Ile Glu Glu Leu Glu Lys Leu Asn Leu
180 185 190

Leu Pro Leu Thr Ser Glu Asp Ser Arg Leu Glu Lys Val Ala Ile Ile
195 200 205

Leu Gly Ala Leu Phe Ser Asp Gly Asn Ile Asp Arg Asn Phe Asn Thr
210 215 220

Leu Ser Phe Ile Ser Ser Glu Arg Lys Ala Ile Glu Arg Phe Val Glu
225 230 235 240

Thr Leu Lys Glu Leu Phe Gly Glu Phe Asn Tyr Glu Ile Arg Asp Asn
245 250 255

His Glu Ser Leu Gly Lys Ser Ile Leu Phe Arg Thr Trp Asp Arg Arg
260 265 270

Ile Ile Arg Phe Phe Val Ala Leu Gly Ala Pro Val Gly Asn Lys Thr
275 280 285

Lys Val Lys Leu Glu Leu Pro Trp Trp Ile Lys Leu Lys Pro Ser Leu
290 295 300

[0037]

Phe Leu Ala Phe Met Asp Gly Leu Tyr Ser Gly Asp Gly Ser Val Pro
305 310 315 320

Arg Phe Ala Arg Tyr Glu Glu Gly Ile Lys Phe Asn Gly Thr Phe Glu
325 330 335

Ile Ala Gln Leu Thr Asp Asp Val Glu Lys Lys Leu Pro Phe Phe Glu
340 345 350

Glu Ile Ala Trp Tyr Leu Ser Phe Phe Gly Ile Lys Ala Lys Val Arg
355 360 365

Val Asp Lys Thr Gly Asp Lys Tyr Lys Val Arg Leu Ile Phe Ser Gln
370 375 380

Ser Ile Asp Asn Val Leu Asn Phe Leu Glu Phe Ile Pro Ile Ser Leu
385 390 395 400

Ser Pro Ala Lys Arg Glu Lys Phe Leu Arg Glu Val Glu Ser Tyr Leu
405 410 415

Ala Ala Val Pro Glu Ser Ser Leu Ala Gly Arg Ile Glu Glu Leu Arg
420 425 430

Glu His Phe Asn Arg Ile Lys Lys Gly Glu Arg Arg Ser Phe Ile Glu
435 440 445

Thr Trp Glu Val Val Asn Val Thr Tyr Asn Val Thr Thr Glu Thr Gly
450 455 460

Asn Leu Leu Ala Asn Gly Leu Phe Val Lys Asn Ser
465 470 475

[0038]

<210> 56
<211> 6
<212> PRT
<213> 人工序列

<220>
<223> 合成构建体

<400> 56

His His His His His His
1 5

<210> 57
<211> 10
<212> PRT
<213> 人工序列

<220>
<223> 合成构建体

<400> 57

His His His His His His His His His His
1 5 10

<210> 58
<211> 6
<212> PRT
<213> 人工序列

<220>
<223> 合成构建体

<400> 58

His Gln His Gln His Gln
1 5

<210> 59

[0039]

<211> 22
<212> PRT
<213> 人工序列

<220>
<223> 合成构建体

<400> 59

Met	Asp	Met	Arg	Val	Pro	Ala	Gln	Leu	Leu	Gly	Leu	Leu	Leu	Leu	Trp
1				5				10					15		

Leu	Arg	Gly	Ala	Arg	Cys
					20

<210> 60
<211> 22
<212> PRT
<213> 人工序列

<220>
<223> 合成构建体

<400> 60

Met	Asp	Met	Arg	Val	Pro	Ala	Gln	Leu	Leu	Gly	Leu	Leu	Leu	Leu	Trp
1				5				10					15		

Leu	Arg	Gly	Ala	Arg	Cys
					20

<210> 61
<211> 22
<212> PRT
<213> 人工序列

<220>
<223> 合成构建体

<400> 61

[0040]

Met Asp Met Arg Val Pro Ala Gln Leu Leu Gly Leu Leu Gln Leu Trp
1 5 10 15

Leu Ser Gly Ala Arg Cys
20

<210> 62
<211> 22
<212> PRT
<213> 人工序列

<220>
<223> 合成构建体

<400> 62

Met Asp Met Arg Val Pro Ala Gln Leu Leu Gly Leu Leu Leu Leu Trp
1 5 10 15

Leu Ser Gly Ala Arg Cys
20

<210> 63
<211> 22
<212> PRT
<213> 人工序列

<220>
<223> 合成构建体

<400> 63

Met Asp Met Arg Val Pro Ala Gln Leu Leu Gly Leu Leu Leu Leu Trp
1 5 10 15

Leu Pro Asp Thr Arg Cys
20

[0041]

<210> 64
<211> 22
<212> PRT
<213> 人工序列

<220>
<223> 合成构建体

<400> 64

Met Asp Met Arg Val Pro Ala Gln Leu Leu Gly Leu Leu Leu Leu Trp
1 5 10 15

Phe Pro Gly Ala Arg Cys
20

<210> 65
<211> 22
<212> PRT
<213> 人工序列

<220>
<223> 合成构建体

<400> 65

Met Asp Met Arg Val Pro Ala Gln Leu Leu Gly Leu Leu Leu Leu Trp
1 5 10 15

Phe Pro Gly Ala Arg Cys
20

<210> 66
<211> 22
<212> PRT
<213> 人工序列

<220>
<223> 合成构建体

[0042]

<400> 66

Met Asp Met Arg Val Leu Ala Gln Leu Leu Gly Leu Leu Leu Cys
1 5 10 15

Phe Pro Gly Ala Arg Cys
20

<210> 67

<211> 22

<212> PRT

<213> 人工序列

<220>

<223> 合成构建体

<400> 67

Met Asp Met Arg Val Leu Ala Gln Leu Leu Gly Leu Leu Leu Cys
1 5 10 15

Phe Pro Gly Ala Arg Cys
20

<210> 68

<211> 22

<212> PRT

<213> 人工序列

<220>

<223> 合成构建体

<400> 68

Met Asp Met Arg Val Pro Ala Gln Leu Leu Gly Leu Leu Leu Trp
1 5 10 15

Leu Pro Gly Ala Arg Cys

[0043]

20

<210> 69
<211> 22
<212> PRT
<213> 人工序列

<220>
<223> 合成构建体

<400> 69

Met	Asp	Met	Arg	Val	Pro	Ala	Gln	Leu	Leu	Gly	Leu	Leu	Leu	Leu	Trp
1				5				10						15	

Leu Pro Gly Ala Arg Cys
20

<210> 70
<211> 22
<212> PRT
<213> 人工序列

<220>
<223> 合成构建体

<400> 70

Met	Asp	Met	Arg	Val	Pro	Ala	Gln	Leu	Leu	Gly	Leu	Leu	Leu	Leu	Trp
1				5				10						15	

Phe Pro Gly Ser Arg Cys
20

<210> 71
<211> 22
<212> PRT
<213> 人工序列

[0044]

<220>

<223> 合成构建体

<400> 71

Met	Asp	Met	Arg	Val	Pro	Ala	Gln	Leu	Leu	Gly	Leu	Leu	Leu	Trp
1				5				10					15	

Phe	Pro	Gly	Ser	Arg	Cys
					20

<210> 72

<211> 22

<212> PRT

<213> 人工序列

<220>

<223> 合成构建体

<400> 72

Met	Asp	Met	Arg	Val	Pro	Ala	Gln	Leu	Leu	Gly	Leu	Leu	Leu	Trp
1				5				10					15	

Leu	Pro	Gly	Ala	Arg	Cys
					20

<210> 73

<211> 22

<212> PRT

<213> 人工序列

<220>

<223> 合成构建体

<400> 73

Met	Asp	Met	Arg	Val	Pro	Ala	Gln	Arg	Leu	Gly	Leu	Leu	Leu	Trp
1				5				10					15	

[0045]

Phe Pro Gly Ala Arg Cys
20

<210> 74
<211> 20
<212> PRT
<213> 人工序列

<220>
<223> 合成构建体

<400> 74

Met Arg Val Pro Ala Gln Leu Leu Gly Leu Leu Leu Leu Trp Leu Pro
1 5 10 15

Gly Ala Arg Cys
20

<210> 75
<211> 22
<212> PRT
<213> 人工序列

<220>
<223> 合成构建体

<400> 75

Met Asp Met Arg Val Pro Ala Gln Leu Leu Gly Leu Leu Leu Leu Trp
1 5 10 15

Leu Pro Gly Ala Arg Cys
20

<210> 76
<211> 22
<212> PRT

[0046]

<213> 人工序列

<220>

<223> 合成构建体

<400> 76

Met Asp Met Arg Val Pro Ala Gln Leu Leu Gly Leu Leu Leu Leu Trp
1 5 10 15

Leu Pro Gly Ala Arg Cys
20

<210> 77

<211> 22

<212> PRT

<213> 人工序列

<220>

<223> 合成构建体

<400> 77

Met Asp Met Arg Val Pro Ala Gln Leu Leu Gly Leu Leu Leu Leu Trp
1 5 10 15

Leu Pro Gly Ala Lys Cys
20

<210> 78

<211> 20

<212> PRT

<213> 人工序列

<220>

<223> 合成构建体

<400> 78

Met Arg Leu Pro Ala Gln Leu Leu Gly Leu Leu Met Leu Trp Val Pro

[0047]

1	5	10	15
---	---	----	----

Gly Ser Ser Glu
20

<210> 79
<211> 20
<212> PRT
<213> 人工序列

<220>
<223> 合成构建体

<400> 79

Met Arg Leu Pro Ala Gln Leu Leu Gly Leu Leu Met Leu Trp Val Pro			
1	5	10	15

Gly Ser Ser Glu
20

<210> 80
<211> 20
<212> PRT
<213> 人工序列

<220>
<223> 合成构建体

<400> 80

Met Arg Leu Pro Ala Gln Leu Leu Gly Leu Leu Met Leu Trp Val Pro			
1	5	10	15

Gly Ser Ser Gly
20

<210> 81

[0048]

<211> 20
<212> PRT
<213> 人工序列

<220>
<223> 合成构建体

<400> 81

Met	Arg	Leu	Pro	Ala	Gln	Leu	Leu	Gly	Leu	Leu	Met	Leu	Trp	Val	Pro
1				5				10					15		

Gly Ser Ser Gly
20

<210> 82
<211> 20
<212> PRT
<213> 人工序列

<220>
<223> 合成构建体

<400> 82

Met	Arg	Leu	Pro	Ala	Gln	Leu	Leu	Gly	Leu	Leu	Met	Leu	Trp	Ile	Pro
1				5				10					15		

Gly Ser Ser Ala
20

<210> 83
<211> 20
<212> PRT
<213> 人工序列

<220>
<223> 合成构建体

<400> 83

[0049]

Met Arg Leu Pro Ala Gln Leu Leu Gly Leu Leu Met Leu Trp Ile Pro
1 5 10 15

Gly Ser Ser Ala
20

<210> 84
<211> 20
<212> PRT
<213> 人工序列

<220>
<223> 合成构建体

<400> 84

Met Arg Leu Pro Ala Gln Leu Leu Gly Leu Leu Met Leu Trp Val Ser
1 5 10 15

Gly Ser Ser Gly
20

<210> 85
<211> 20
<212> PRT
<213> 人工序列

<220>
<223> 合成构建体

<400> 85

Met Arg Leu Pro Ala Gln Leu Leu Gly Leu Leu Met Leu Trp Val Ser
1 5 10 15

Gly Ser Ser Gly
20

[0050]

<210> 86
<211> 20
<212> PRT
<213> 人工序列

<220>
<223> 合成构建体

<400> 86

Met	Arg	Leu	Leu	Ala	Gln	Leu	Leu	Gly	Leu	Leu	Met	Leu	Trp	Val	Pro
1				5				10					15		

Gly Ser Ser Gly
20

<210> 87
<211> 20
<212> PRT
<213> 人工序列

<220>
<223> 合成构建体

<400> 87

Met	Glu	Thr	Pro	Ala	Gln	Leu	Leu	Phe	Leu	Leu	Leu	Leu	Trp	Leu	Pro
1				5				10					15		

Asp Thr Thr Gly
20

<210> 88
<211> 20
<212> PRT
<213> 人工序列

<220>
<223> 合成构建体

[0051]

<400> 88

Met Glu Thr Pro Ala Gln Leu Leu Phe Leu Leu Leu Trp Leu Pro
1 5 10 15

Asp Thr Thr Gly
20

<210> 89

<211> 20

<212> PRT

<213> 人工序列

<220>

<223> 合成构建体

<400> 89

Met Glu Ala Pro Ala Gln Leu Leu Phe Leu Leu Leu Trp Leu Pro
1 5 10 15

Asp Thr Thr Gly
20

<210> 90

<211> 20

<212> PRT

<213> 人工序列

<220>

<223> 合成构建体

<400> 90

Met Glu Ala Pro Ala Gln Leu Leu Phe Leu Leu Leu Trp Leu Pro
1 5 10 15

Asp Thr Thr Gly

[0052]

20

<210> 91
<211> 20
<212> PRT
<213> 人工序列

<220>
<223> 合成构建体

<400> 91

Met	Glu	Ala	Pro	Ala	Gln	Leu	Leu	Phe	Leu	Leu	Leu	Leu	Trp	Leu	Pro
1				5				10						15	

Asp Thr Thr Gly
20

<210> 92
<211> 20
<212> PRT
<213> 人工序列

<220>
<223> 合成构建体

<400> 92

Met	Glu	Ala	Pro	Ala	Gln	Leu	Leu	Phe	Leu	Leu	Leu	Leu	Trp	Leu	Thr
1				5				10						15	

Asp Thr Thr Gly
20

<210> 93
<211> 23
<212> PRT
<213> 人工序列

[0053]

<220>

<223> 合成构建体

<400> 93

Met Glu Pro Trp Lys Pro Gln His Ser Phe Phe Phe Leu Leu Leu Leu
1 5 10 15

Trp Leu Pro Asp Thr Thr Gly
20

<210> 94

<211> 20

<212> PRT

<213> 人工序列

<220>

<223> 合成构建体

<400> 94

Met Val Leu Gln Thr Gln Val Phe Ile Ser Leu Leu Leu Trp Ile Ser
1 5 10 15

Gly Ala Tyr Gly
20

<210> 95

<211> 20

<212> PRT

<213> 人工序列

<220>

<223> 合成构建体

<400> 95

Met Gly Ser Gln Val His Leu Leu Ser Phe Leu Leu Leu Trp Ile Ser
1 5 10 15

[0054]

Asp Thr Arg Ala
20

<210> 96
<211> 19
<212> PRT
<213> 人工序列

<220>
<223> 合成构建体

<400> 96

Met	Leu	Pro	Ser	Gln	Leu	Ile	Gly	Phe	Leu	Leu	Leu	Trp	Val	Pro	Ala
1				5				10						15	

Ser Arg Gly

<210> 97
<211> 19
<212> PRT
<213> 人工序列

<220>
<223> 合成构建体

<400> 97

Met	Leu	Pro	Ser	Gln	Leu	Ile	Gly	Phe	Leu	Leu	Leu	Trp	Val	Pro	Ala
1				5				10						15	

Ser Arg Gly

<210> 98
<211> 20
<212> PRT

[0055]

<213> 人工序列

<220>

<223> 合成构建体

<400> 98

Met Val Ser Pro Leu Gln Phe Leu Arg Leu Leu Leu Leu Trp Val Pro
1 5 10 15

Ala Ser Arg Gly
20

<210> 99

<211> 22

<212> PRT

<213> 人工序列

<220>

<223> 合成构建体

<400> 99

Met Asp Met Arg Val Pro Ala Gln Leu Leu Gly Leu Leu Leu Leu Trp
1 5 10 15

Phe Pro Gly Ser Gly Gly
20

<210> 100

<211> 23

<212> PRT

<213> 人工序列

<220>

<223> 合成构建体

<400> 100

Met Asp Met Arg Val Pro Ala Gln Leu Leu Gly Leu Leu Leu Leu Trp

[0056]

1 5 10 15

Phe Pro Gly Ser Gly Gly Gly
20

<210> 101
<211> 24
<212> PRT
<213> 人工序列

<220>
<223> 合成构建体

<400> 101

Met Asp Met Arg Val Pro Ala Gln Leu Leu Gly Leu Leu Leu Leu Trp
1 5 10 15

Phe Pro Gly Ser Gly Gly Gly Gly
20

<210> 102
<211> 25
<212> PRT
<213> 人工序列

<220>
<223> 合成构建体

<400> 102

Met Asp Met Arg Val Pro Ala Gln Leu Leu Gly Leu Leu Leu Leu Trp
1 5 10 15

Phe Pro Gly Ser Gly Gly Gly Gly Gly
20 25

<210> 103

[0057]

<211> 22
<212> PRT
<213> 人工序列

<220>
<223> 合成构建体

<400> 103

Met	Arg	Met	Arg	Val	Pro	Ala	Gln	Leu	Leu	Gly	Leu	Leu	Leu	Leu	Trp
1				5				10					15		

Phe Pro Gly Ser Arg Cys
20

<210> 104
<211> 22
<212> PRT
<213> 人工序列

<220>
<223> 合成构建体

<400> 104

Met	Arg	Met	Arg	Val	Pro	Ala	Gln	Leu	Leu	Gly	Leu	Leu	Leu	Leu	Trp
1				5				10					15		

Phe Pro Gly Ser Gly Gly
20

<210> 105
<211> 23
<212> PRT
<213> 人工序列

<220>
<223> 合成构建体

<400> 105

[0058]

Met Arg Arg Met Arg Val Pro Ala Gln Leu Leu Gly Leu Leu Leu Leu
1 5 10 15

Trp Phe Pro Gly Ser Arg Cys
20

<210> 106
<211> 23
<212> PRT
<213> 人工序列

<220>
<223> 合成构建体

<400> 106

Met Arg Arg Met Arg Val Pro Ala Gln Leu Leu Gly Leu Leu Leu Leu
1 5 10 15

Trp Phe Pro Gly Ser Gly Gly
20

<210> 107
<211> 24
<212> PRT
<213> 人工序列

<220>
<223> 合成构建体

<400> 107

Met Arg Arg Arg Met Arg Val Pro Ala Gln Leu Leu Gly Leu Leu Leu
1 5 10 15

Leu Trp Phe Pro Gly Ser Gly Gly
20

[0059]

<210> 108

<211> 20

<212> PRT

<213> 人工序列

<220>

<223> 合成构建体

<400> 108

Met	Asp	Met	Arg	Val	Pro	Ala	Gln	Leu	Leu	Gly	Asp	Glu	Trp	Phe	Pro
1				5				10					15		

Gly	Ser	Gly	Gly												
				20											

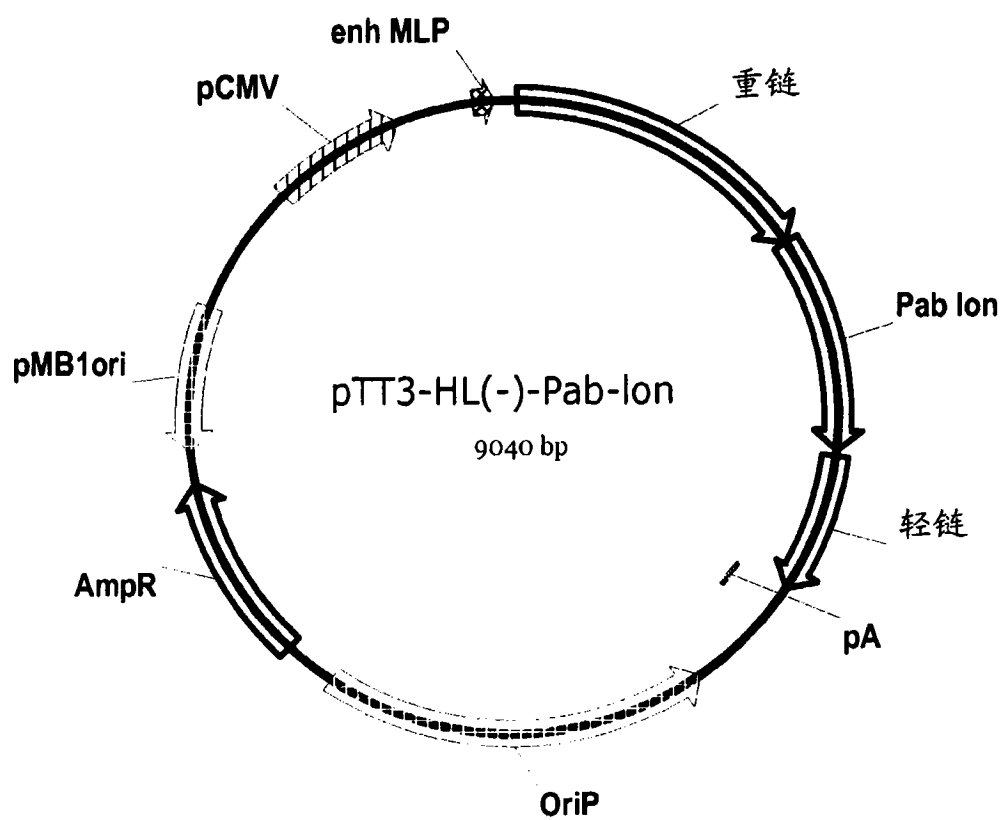


图 1

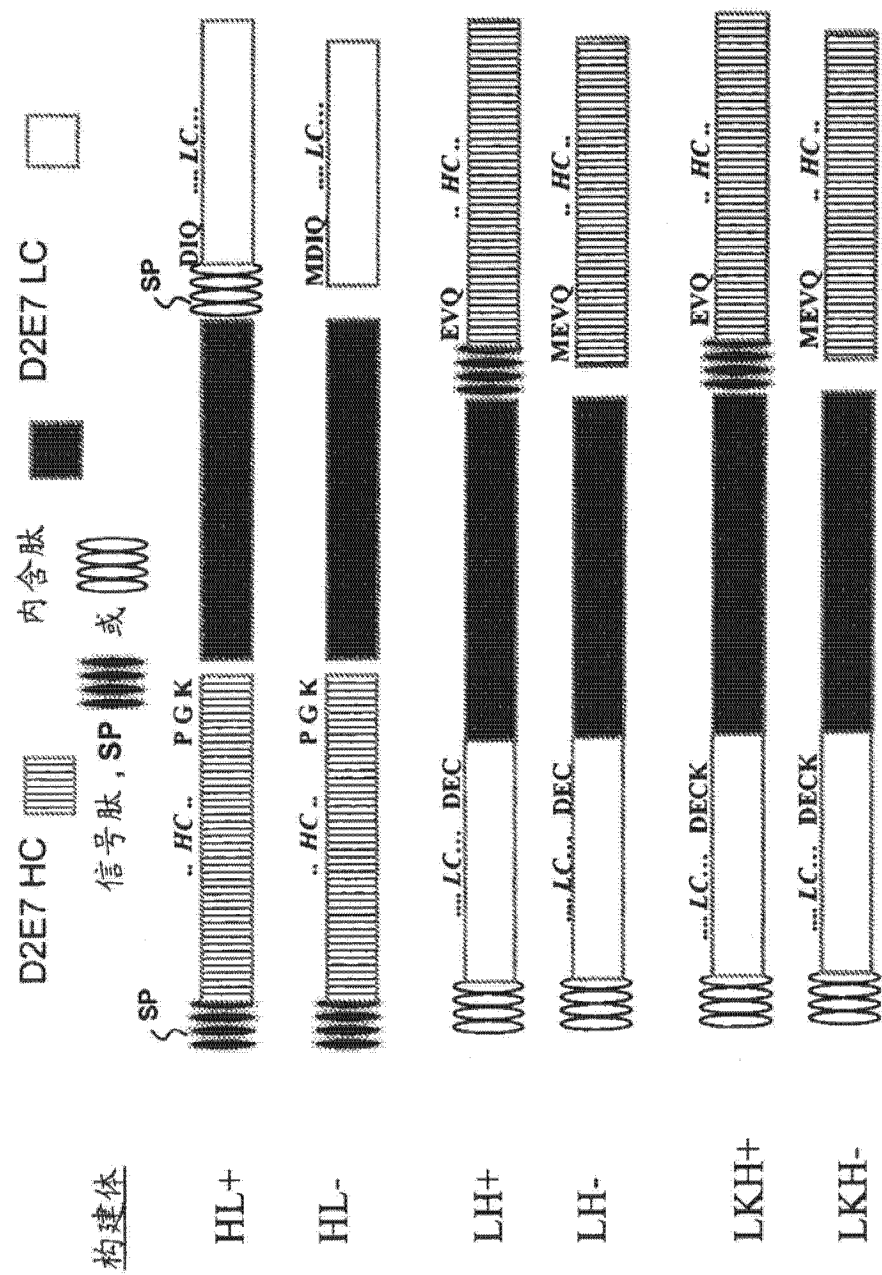


图 2

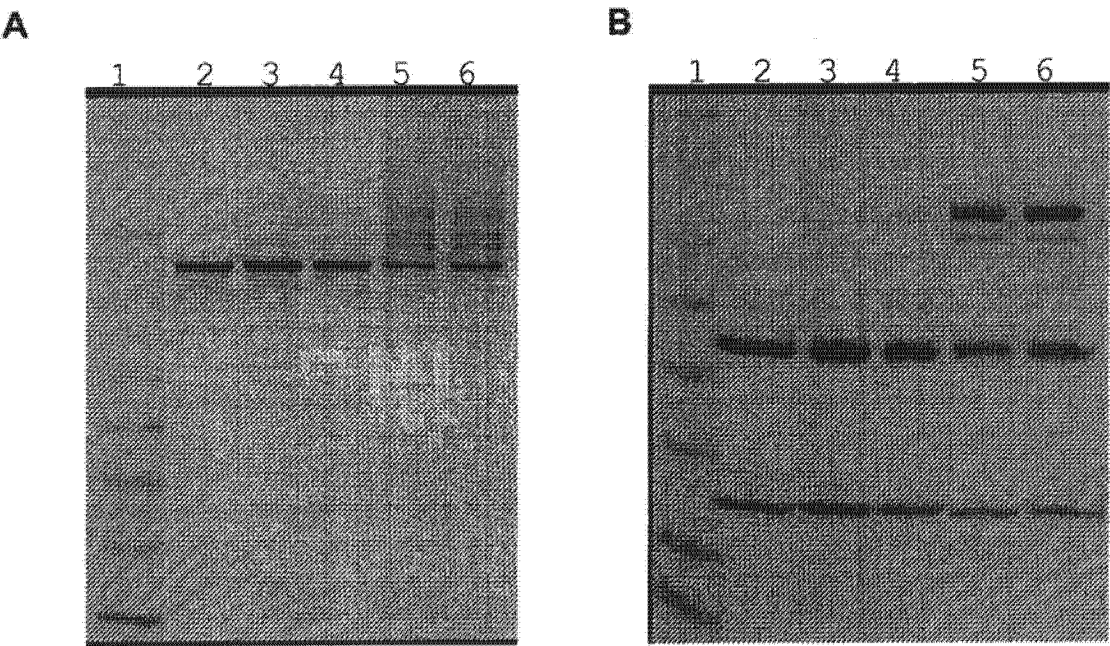


图 3

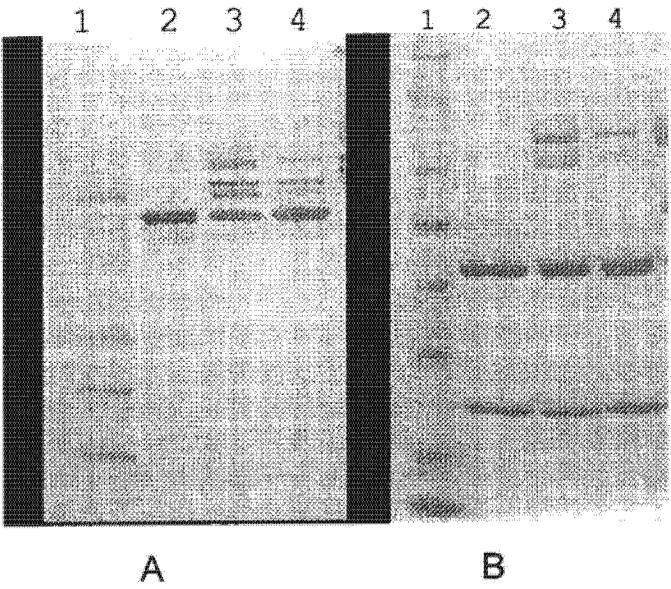


图 4

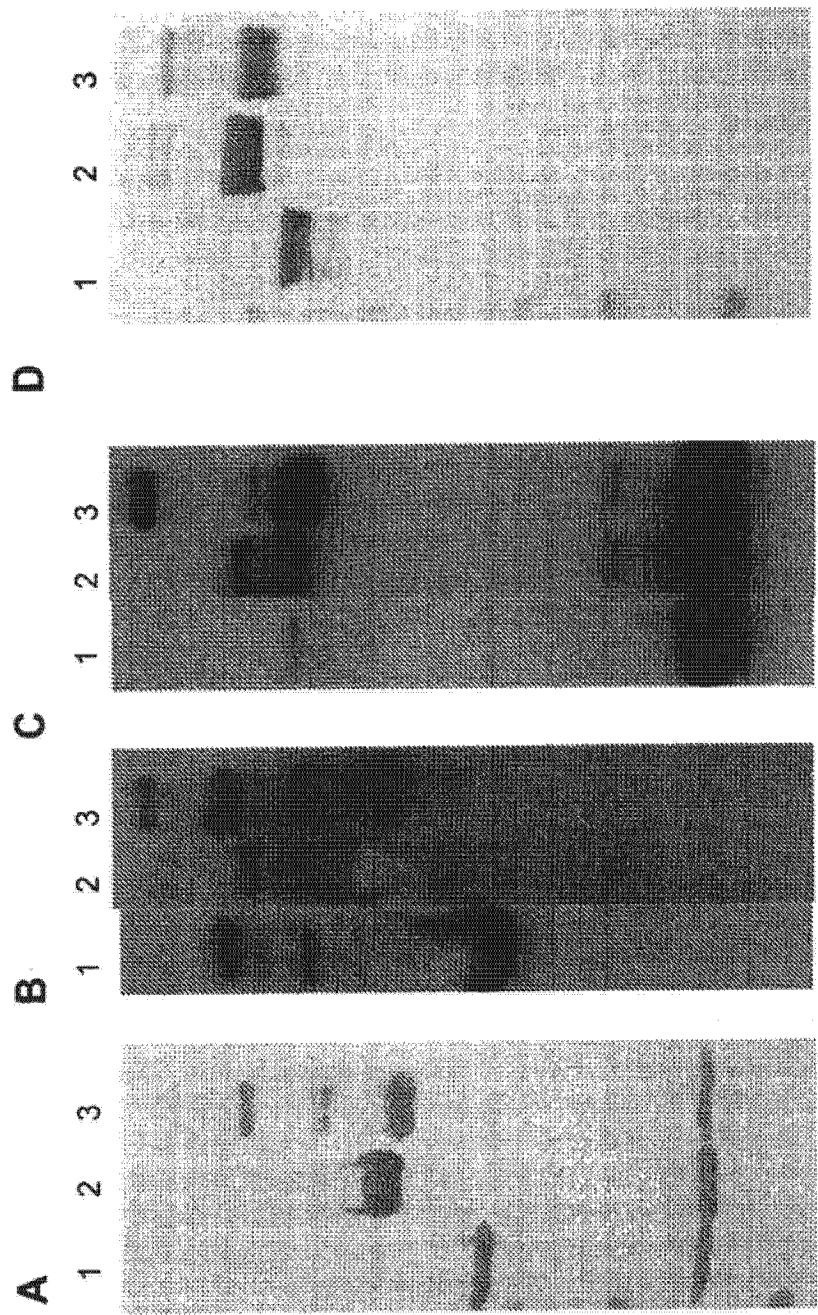


图 5

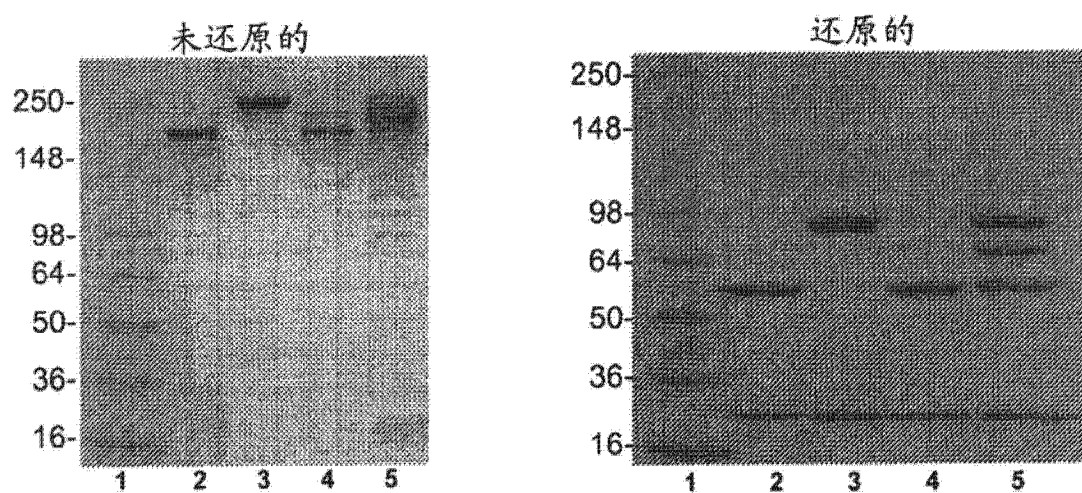


图 6

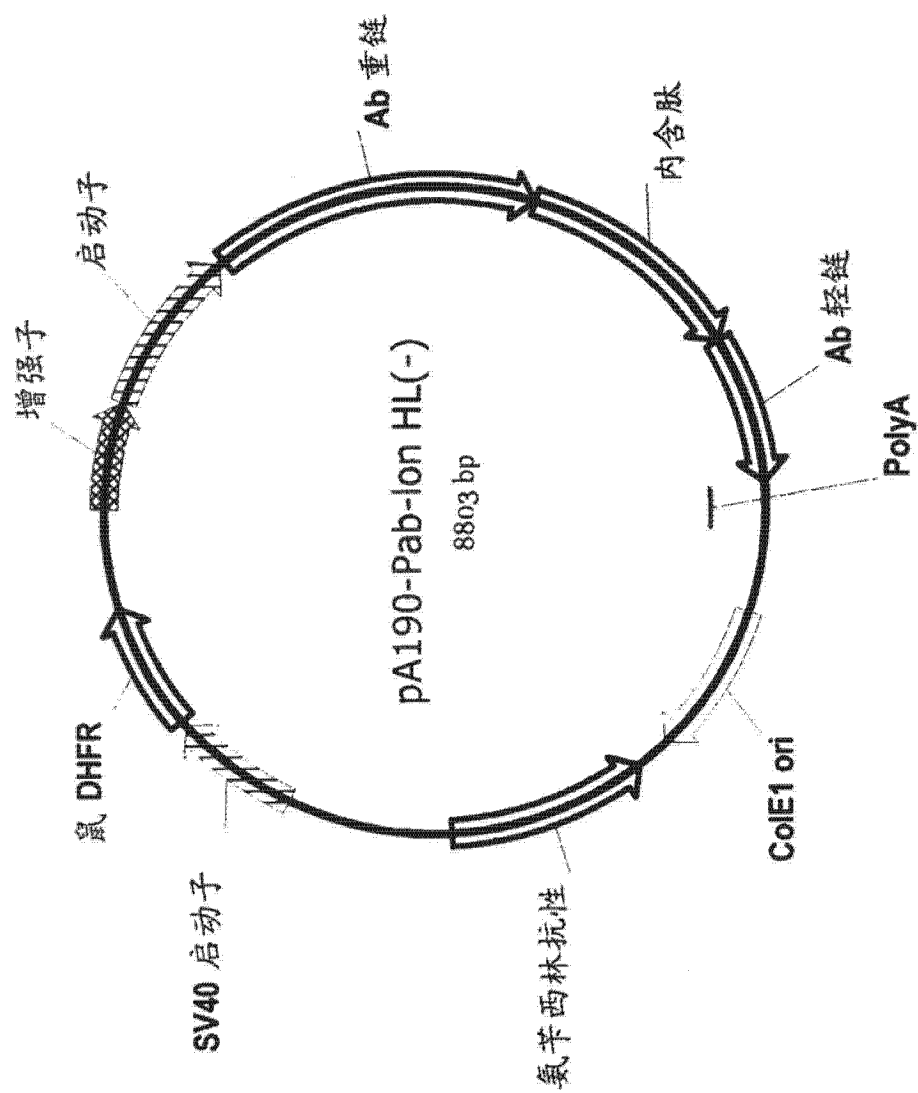


图 7

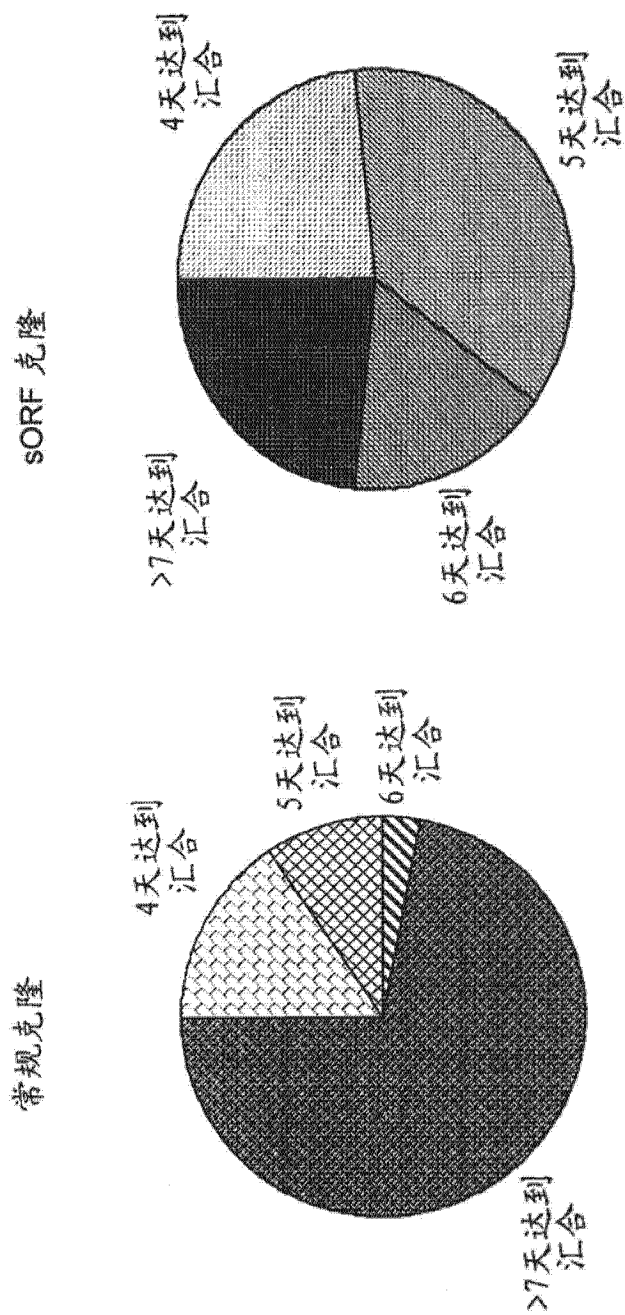


图 8

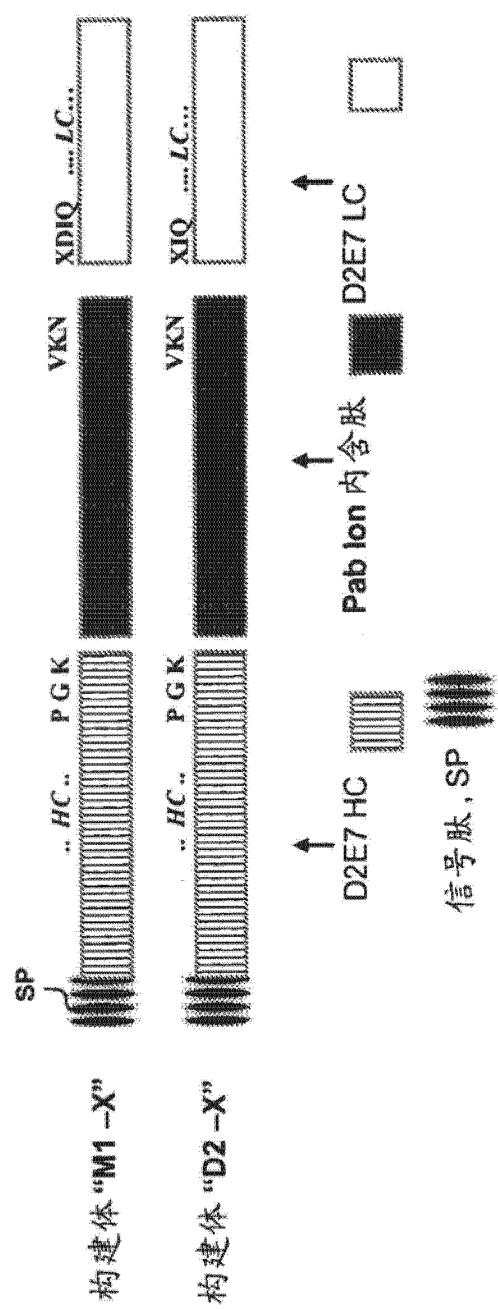


图 9

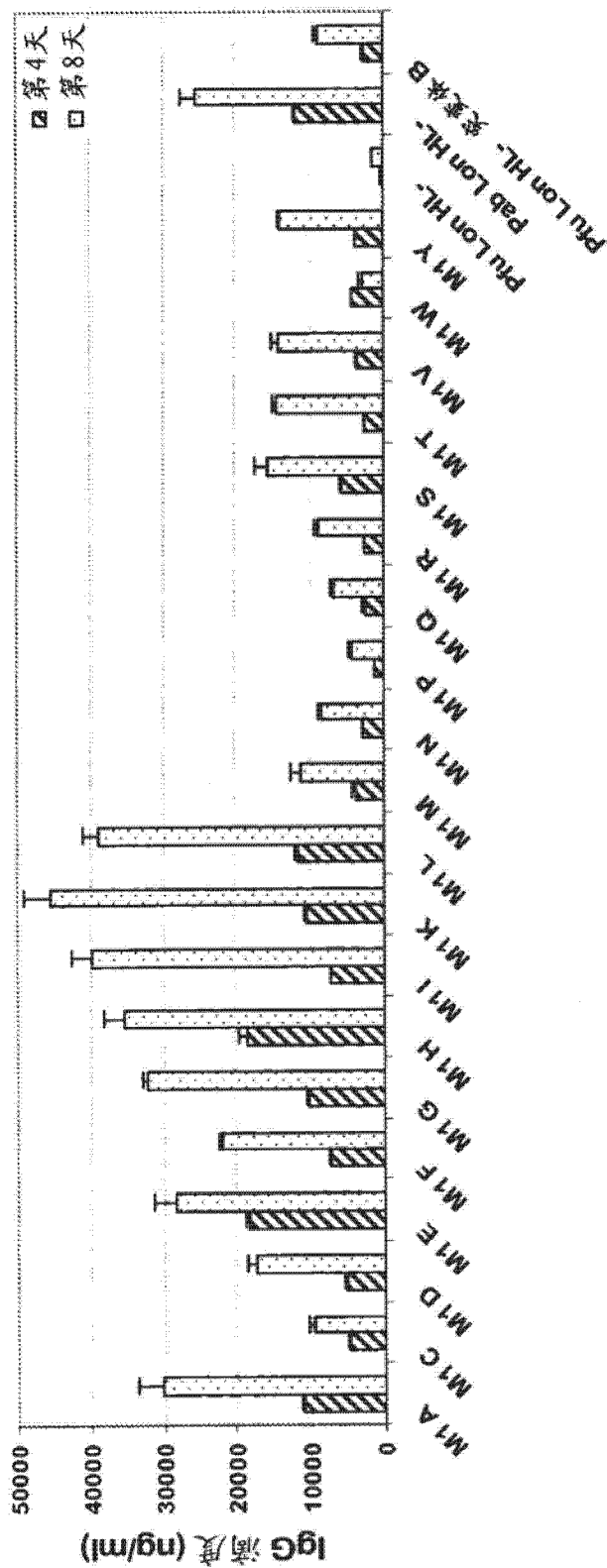


图 10

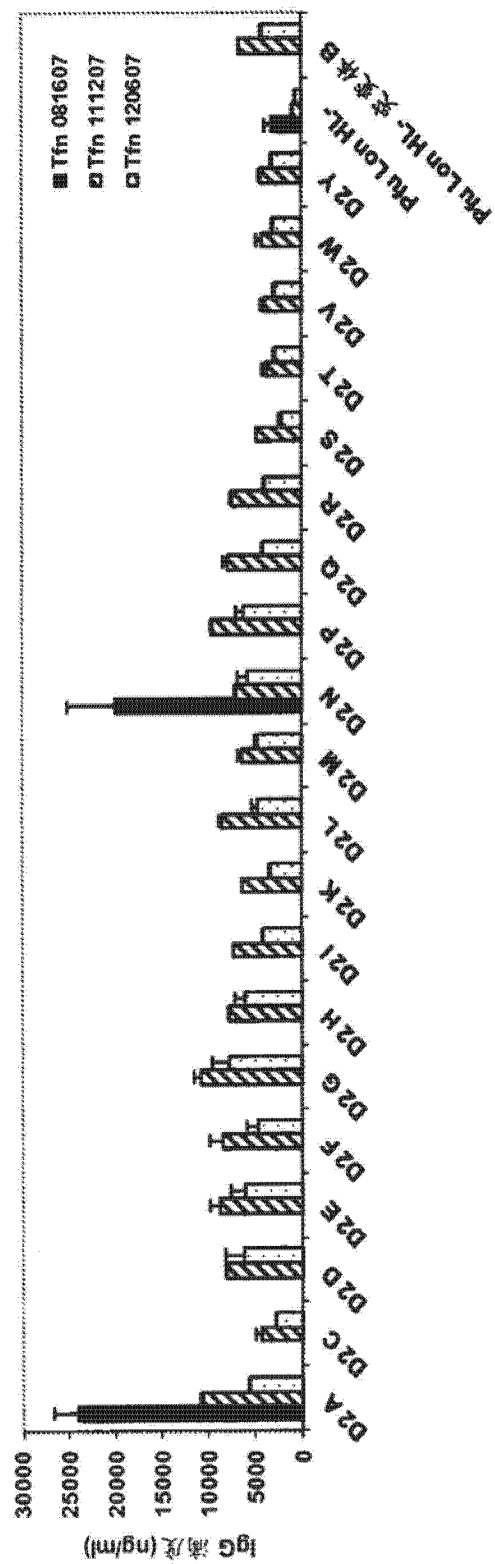


图 11

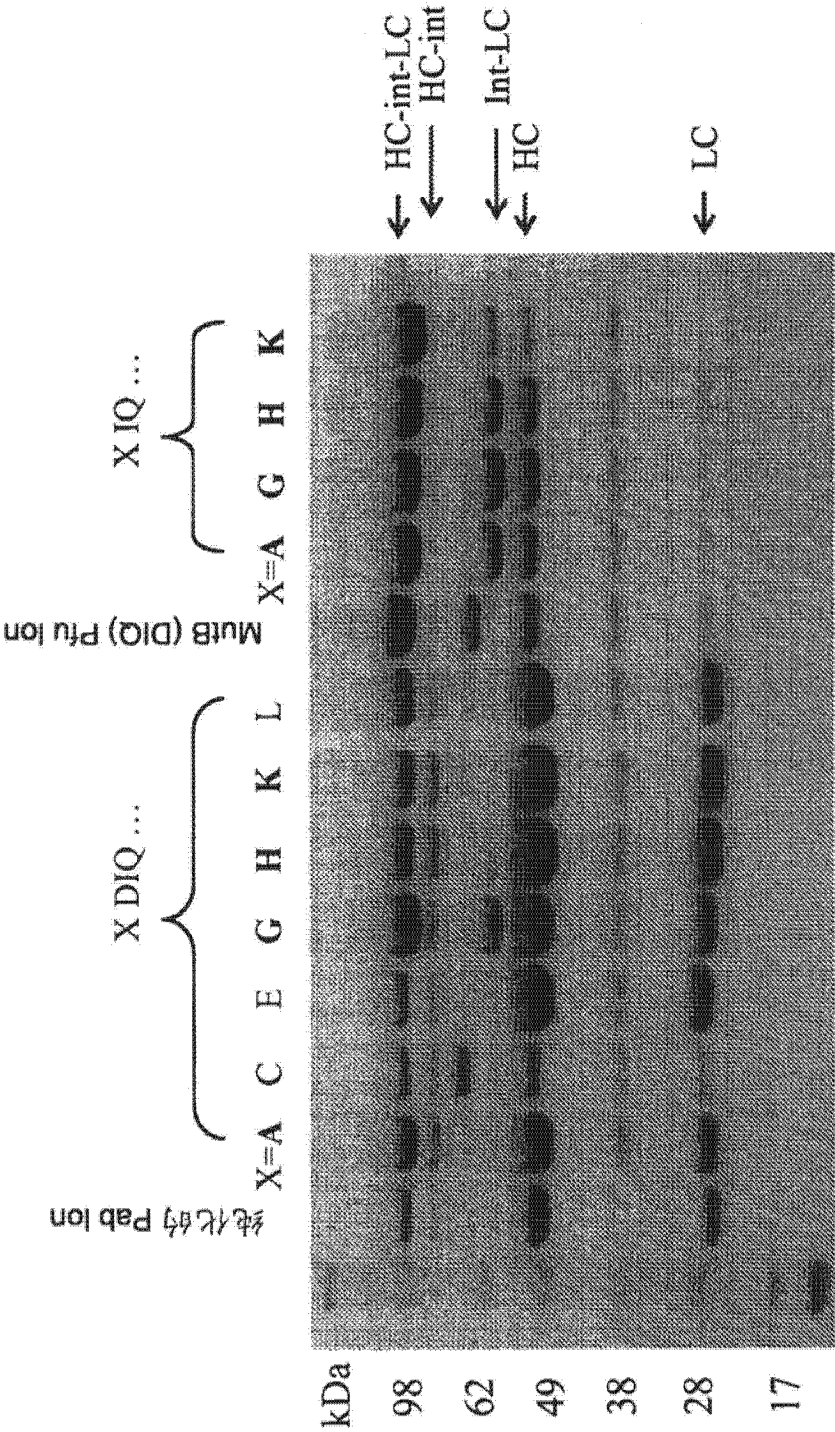


图 12

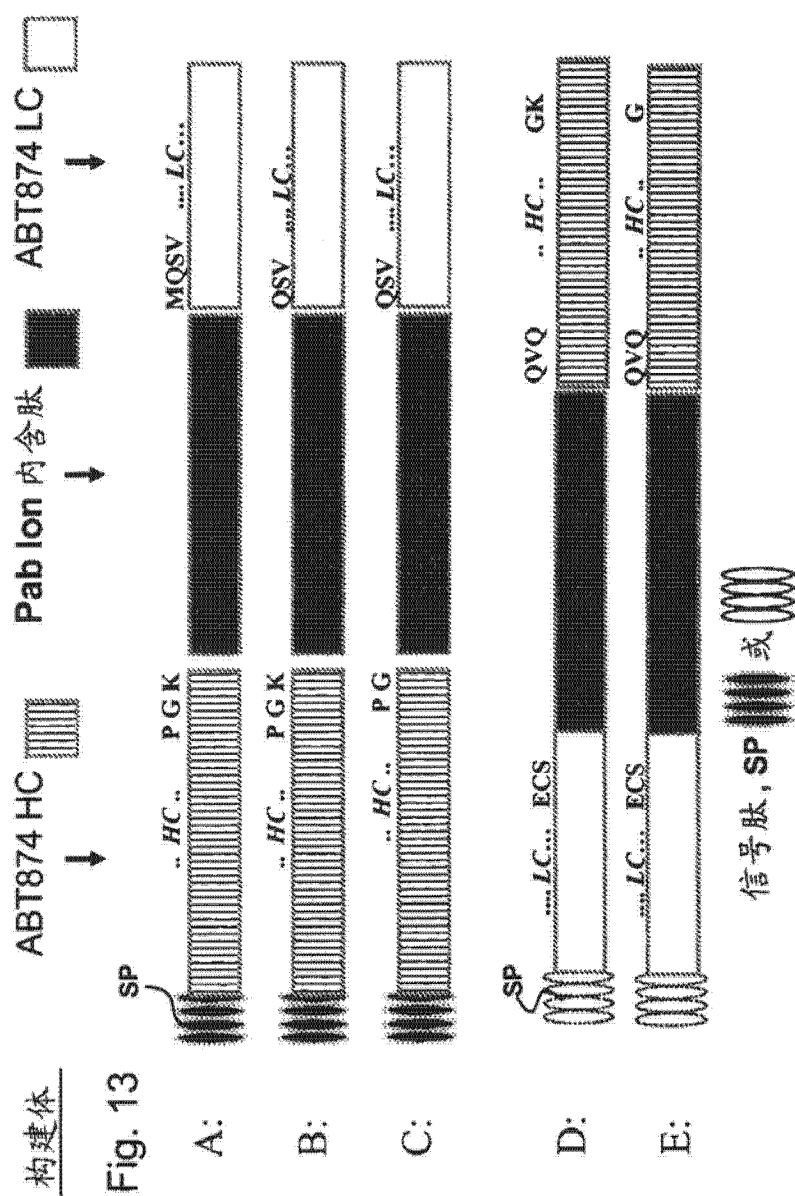


图 13

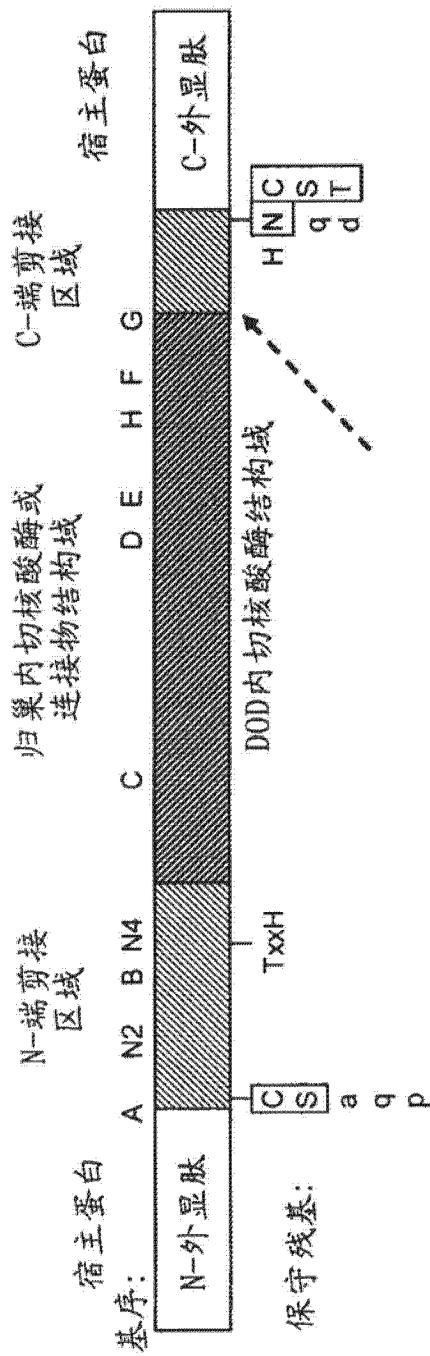


图 14

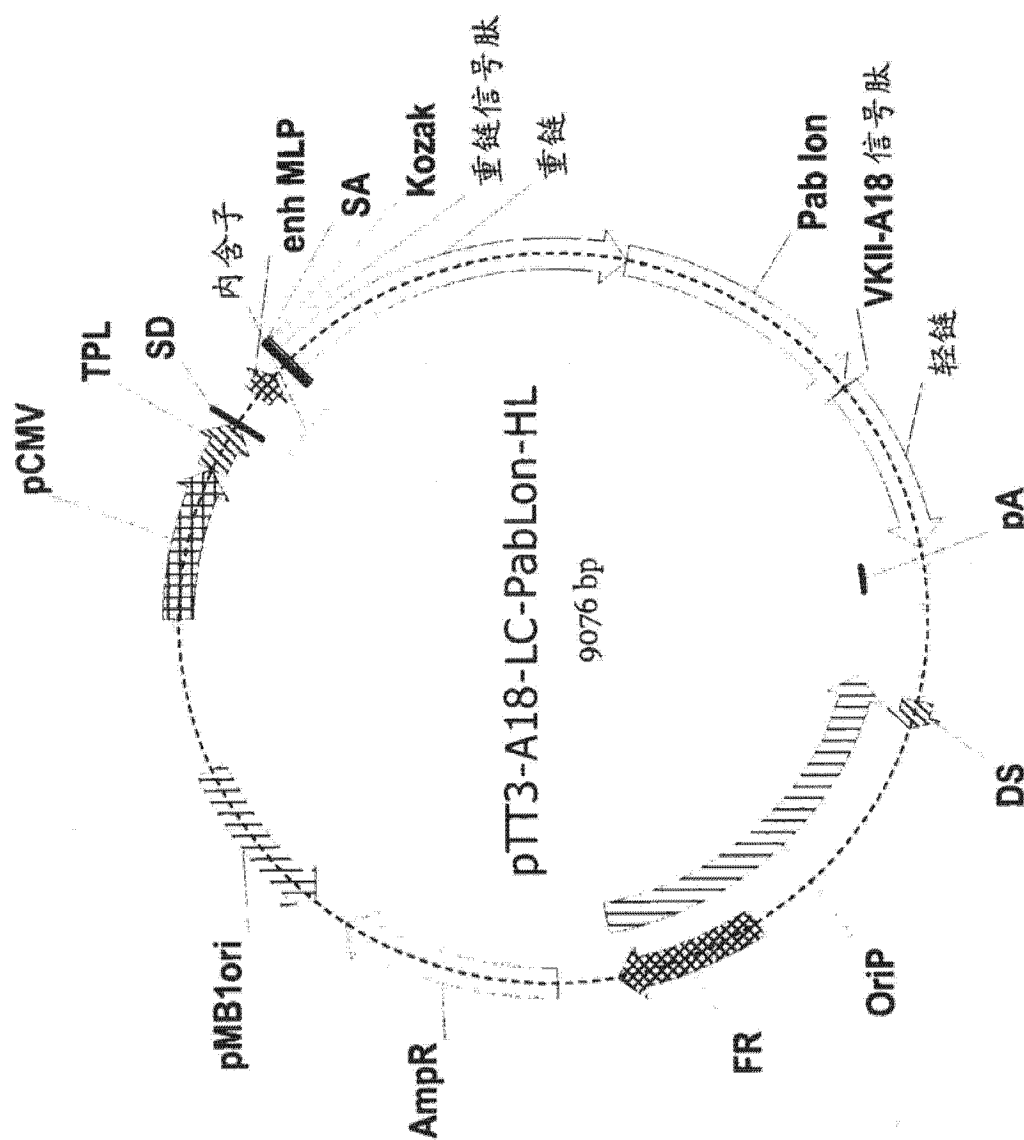


图 15

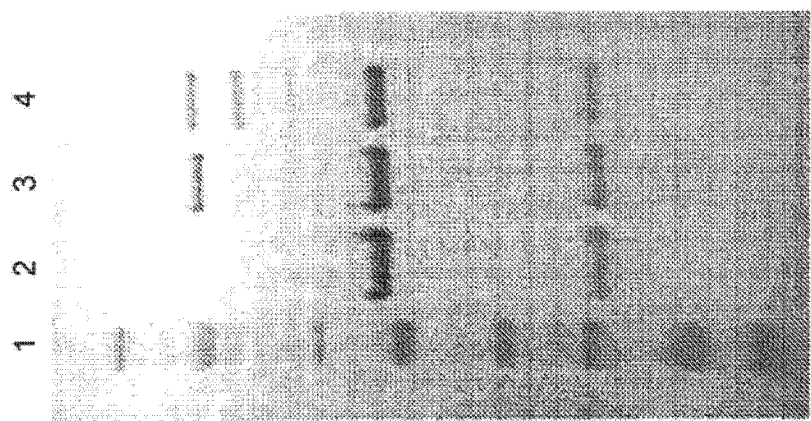


图 16

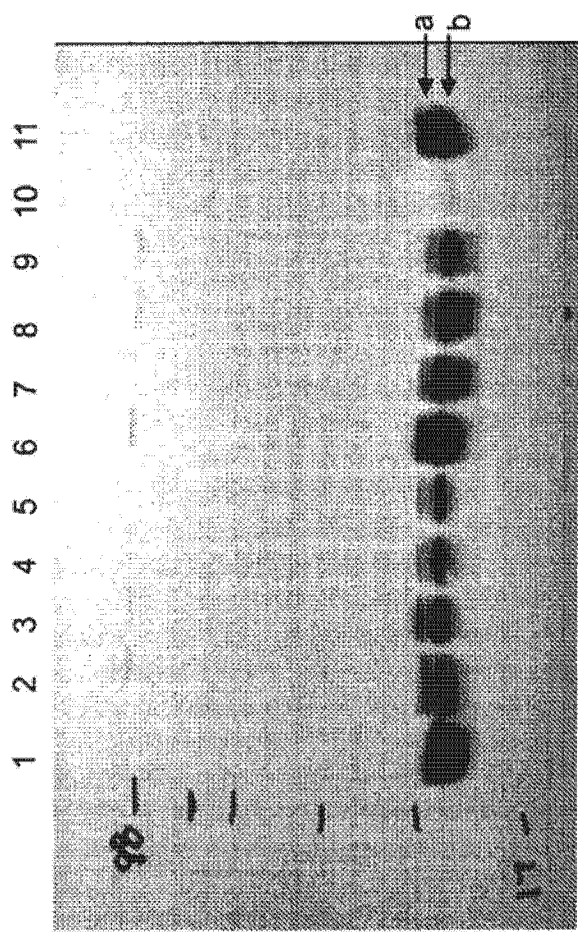


图 17

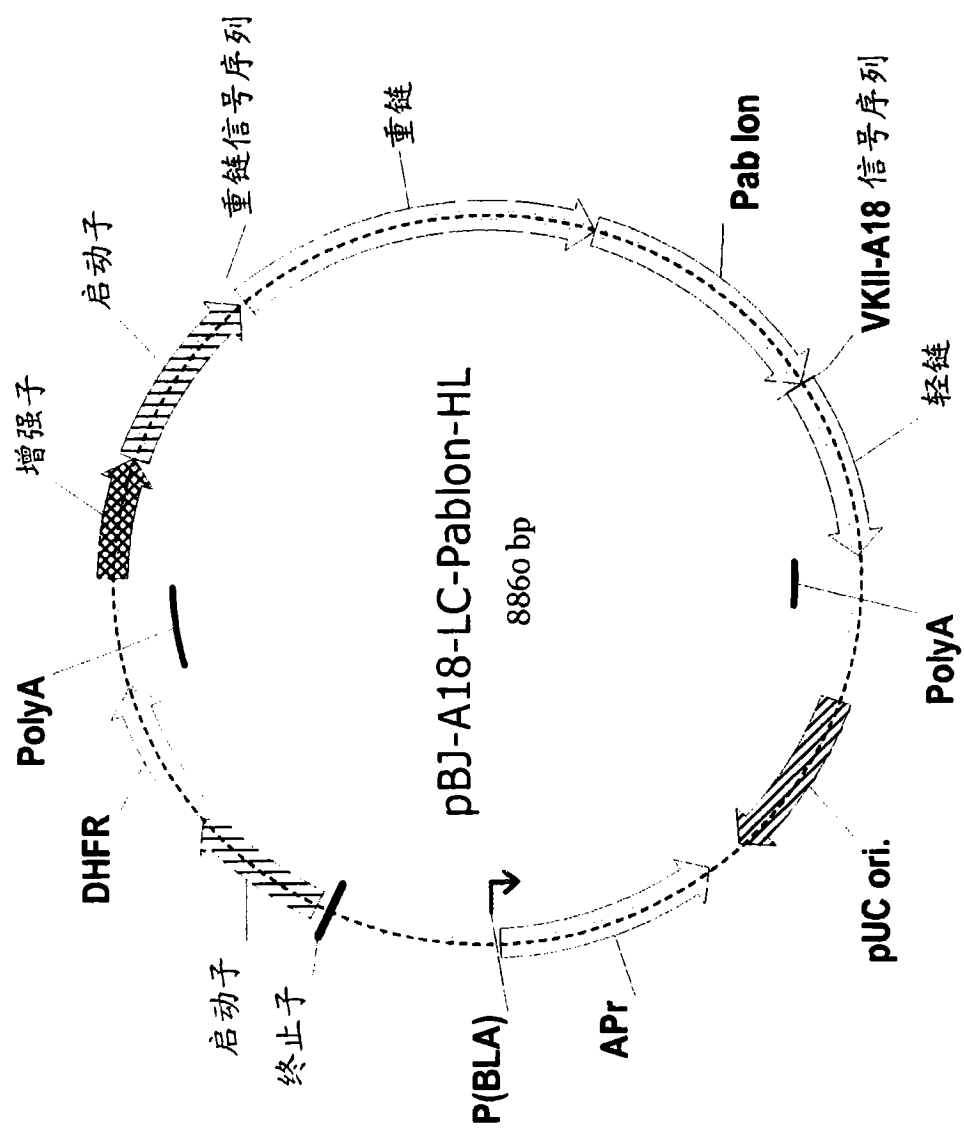


图 18