



(12) 发明专利

(10) 授权公告号 CN 101065491 B

(45) 授权公告日 2012.01.11

(21) 申请号 200580040821.7

(22) 申请日 2005.09.28

(30) 优先权数据

60/613,719 2004.09.29 US

(85) PCT申请进入国家阶段日

2007.05.29

(86) PCT申请的申请数据

PCT/IL2005/001045 2005.09.28

(87) PCT申请的公布数据

W02006/035442 EN 2006.04.06

(73) 专利权人 胶原植物有限公司

地址 以色列谢莫纳村

(72) 发明人 O·肖塞约夫 H·斯泰恩

(74) 专利代理机构 中国专利代理(香港)有限公

司 72001

代理人 权陆军 黄可峻

(51) Int. Cl.

C12N 15/82(2006.01)

A01H 5/00(2006.01)

(56) 对比文件

WO 0134647 A, 2001.05.17, 全文.

WO 0129242 A, 2001.04.26, 说明书第 85 页
第 28 行至第 86 页第 17 行、附图 18.

VUORI K. CHARACTERIZATION OF THE
HUMAN PRILYL 4-HYDROXYLASE TETRAMER
AND ITSMULTIFUNCTIONAL PROTEIN
DISULFIDE-ISOMERASESUBUNIT SYNTHESIZED
IN A BACULOCIRUSEXPRESSON SYSTEM.
PROCEEDINGS OF THE NATIONAL ACADEMY OF
SCIENCES OF THE UNITED STATES OF AMERIC89
16. 1992, 89(16), 7467-7470.

审查员 李振鹏

权利要求书 1 页 说明书 20 页
序列表 57 页 附图 10 页

(54) 发明名称

生产胶原的植物及其生成和使用方法

(57) 摘要

提供了在植物中生产胶原的方法和生产胶原的植物。该方法通过以能够使胶原 α 链积聚在缺乏内源性 P4H 活性的亚细胞区室中的方式,在植物中表达至少一种类型的胶原 α 链来实现,从而在植物中生产胶原。

1. 在植物或被分离的植物细胞中生产胶原的方法,其包括使至少一种类型的胶原 α 链和哺乳动物脯氨酰-4-羟化酶 (P4H) 靶向所述植物或所述被分离的植物细胞的液泡,从而允许所述至少一种类型的所述胶原 α 链由所述哺乳动物 P4H 羟基化,且不由内源性 P4H 羟基化,从而在所述植物或被分离的植物细胞中生产胶原。

2. 权利要求 1 的方法,其进一步包括在所述缺乏内源性 P4H 活性的亚细胞区室中表达外源性赖氨酰羟化酶 (LH3)。

3. 权利要求 1 的方法,其中所述至少一种类型的所述胶原 α 链包括用于靶向液泡的信号肽。

4. 权利要求 1 的方法,其中所述至少一种类型的所述胶原 α 链缺乏 ER 靶向序列或保留序列。

5. 权利要求 1 的方法,其中所述至少一种类型的所述胶原 α 链表达于包含 DNA 的植物细胞器中。

6. 权利要求 1 的方法,其中所述外源性哺乳动物 P4H 包括用于靶向液泡的信号肽。

7. 权利要求 1 的方法,其中所述外源性哺乳动物 P4H 缺乏 ER 靶向序列或保留序列。

8. 权利要求 1 的方法,其中所述外源性哺乳动物 P4H 表达于包含 DNA 的植物细胞器中。

9. 权利要求 1 的方法,其中所述至少一种类型的所述胶原 α 链是 $\alpha 1$ 链。

10. 权利要求 1 的方法,其中所述至少一种类型的所述胶原 α 链是 $\alpha 2$ 链。

11. 权利要求 1 的方法,其中所述至少一种类型的所述胶原 α 链包括 C-末端和 / 或 N-末端前肽。

12. 权利要求 1 的方法,其中所述植物选自烟草、玉米、苜蓿、稻、马铃薯、大豆、番茄、小麦、大麦、低芥酸菜子、胡萝卜和棉花。

13. 权利要求 1 的方法,其中所述至少一种类型的所述胶原 α 链或所述外源性 P4H 只表达于所述植物的部分中。

14. 权利要求 13 的方法,其中所述植物的所述部分是叶、种子、根、块茎或茎。

15. 权利要求 1 的方法,其中所述外源性哺乳动物 P4H 能够使所述至少一种类型的所述胶原 α 链的 Gly-X-Y 三联体的 Y 位置特异性羟基化。

16. 权利要求 14 的方法,其中所述外源性哺乳动物 P4H 是人 P4H。

17. 权利要求 1 的方法,其中对所述植物实施应激条件。

18. 权利要求 17 的方法,其中所述应激条件选自干旱、盐度、损伤、寒冷和喷射应激诱导化合物。

19. 核酸构建体,其包括置于在植物细胞中起作用的启动子的转录控制下、编码人 P4H 的多核苷酸,所述人 P4H 包括用于靶向液泡的信号肽。

20. 权利要求 19 的核酸构建体,其中所述启动子选自 CaMV 35S 启动子、遍在蛋白启动子、rbcS 启动子和 SVBV 启动子。

生产胶原的植物及其生成和使用方法

[0001] 发明领域和背景

[0002] 本发明涉及生产胶原的植物及其生成和使用方法。更具体而言,本发明涉及用于生成能够生产高水平羟基化胶原链的植物的新方法,所述胶原链能够形成天然的三股螺旋 I 型胶原纤维。

[0003] 胶原是负责脊椎动物和许多其他多细胞生物的结构完整性的主要结构蛋白。I 型胶原代表原型的纤维胶原,并且是大多数组织中的主要胶原类型。

[0004] I 型胶原是骨和腱中占优势的胶原组分,并且大量发现于皮肤、主动脉和肺中。I 型胶原纤维提供极大的抗张强度和有限的伸长性。I 型胶原的最丰富的分子形式是由 2 种不同的 α 链 [$\alpha 1(I)$]₂ 和 $\alpha 2(I)$ 组成的异源三聚体 (Inkinen, 2003)。所有纤维胶原分子都包含由重复的 Gly-X-Y 三联体构建的 3 种多肽链,其中 X 和 Y 可以是任何氨基酸,但通常是亚氨基酸脯氨酸和羟脯氨酸。

[0005] 形成原纤维的胶原作为包含球状 N- 和 C- 末端延伸前肽的前体前胶原而合成。前胶原的生物合成是牵涉多种不同的翻译后修饰的复杂过程,所述翻译后修饰包括脯氨酸和赖氨酸羟基化、N- 连接的和 O- 连接的糖基化以及链内和链间二硫键的形成。执行这些修饰的酶以协同的方式起作用以确保正确排列和热稳定的三股螺旋分子的折叠和装配。

[0006] 每种前胶原分子由 3 种多肽链组成成分在粗面内质网内装配。因为多肽链共翻译地穿过内质网膜移位 (translocate), 所以脯氨酸和赖氨酸残基的羟基化发生在 Gly-X-Y 重复区域内。一旦多肽链完全移位到内质网腔内, C- 前肽就折叠。3 种前- α 链随后经由其 C- 前肽结合以形成三聚体分子,从而允许 Gly-X-Y 重复区域在其 C- 末端形成成核点,确保链的正确排列。Gly-X-Y 区域随后在 C- 至 -N 方向折叠以形成三股螺旋。

[0007] 多肽链修饰和三股螺旋形成之间的时间关系是关键性的,因为脯氨酸残基的羟基化是确保三股螺旋在体温下的稳定性必需的,一旦形成,三股螺旋就不再充当羟基化酶的底物。C- 前肽 (以及较少程度上的 N- 前肽) 在前胶原通过细胞的过程中保持前胶原可溶 (Bulleid 等人, 2000)。在前胶原分子分泌进入细胞外基质后或在这个过程中,通过前胶原 N- 和 C- 蛋白酶去除前肽,从而触发胶原分子自发地自身装配为原纤维 (Hulmes, 2002)。通过前胶原 N- 和 C- 蛋白酶去除前肽使前胶原的溶解性降低了 > 10000 倍,并且是发动胶原自身装配成纤维所必需和充分的。对这个装配过程关键的是在三股螺旋结构域末端被称为端肽的短的非三股螺旋肽,它确保原纤维结构内的胶原分子的正确记录,并降低自身装配的临界浓度 (Bulleid 等人, 2000)。在自然中,胶原的三股螺旋结构的稳定性需要由脯氨酰-4- 羟化酶 (P4H) 对脯氨酸的羟基化,以在胶原链内形成羟脯氨酸残基。

[0008] 表达胶原链的植物是本领域已知的,参见例如,美国专利号 6, 617, 431 和 (Merle 等人, 2002, Ruggiero 等人, 2000)。尽管植物能够合成包含羟脯氨酸的蛋白质,但与哺乳动物 P4H 相比,负责在植物细胞中合成羟脯氨酸的脯氨酰羟化酶显示出相对宽松的底物序列特异性,并因此只在 Gly-X-Y 三联体的 Y 位置包含羟脯氨酸的胶原生产需要植物共表达胶原和 P4H 基因 (Olsen 等人, 2003)。

[0009] 生产依赖于植物中天然存在的羟基化机制的人胶原的尝试得到了在脯氨酸羟基

化方面弱的胶原 (Merle 等人, 2002)。这种胶原在低于 30°C 的温度下溶解或失去其三股螺旋结构。胶原和脯氨酰-羟化酶的共表达产生稳定羟基化的胶原, 它在生物学上与体温时的应用相关 (Merle 等人, 2002)。

[0010] 赖氨酰羟化酶 (LH, EC 1.14.11.4)、半乳糖基转移酶 (EC 2.4.1.50) 和葡糖基转移酶 (EC 2.4.1.66) 是在胶原翻译后修饰中牵涉的酶。它们序贯地将特定位置中的赖氨酰残基修饰为羟赖氨酰、半乳糖基羟赖氨酰和葡糖基半乳糖基羟赖氨酰残基。这些结构是胶原特有的并且是其功能活性必需的 (Wang 等人, 2002)。单独的人的酶, 赖氨酰羟化酶 3 (LH3) 可以催化羟赖氨酸连接的碳水化合物形成中的所有 3 个连续步骤 (Wang 等人, 2002)。

[0011] 烟草中表达的人胶原的羟赖氨酸形成少于 2% 的在牛胶原中发现的羟赖氨酸 (0.04% 残基 / 1.88% 残基)。这提示植物内源性的赖氨酰羟化酶不能充分地使胶原中的赖氨酸羟基化。

[0012] 在将本发明变为实践的同时, 本发明人揭示了胶原链的有效羟基化依赖于胶原链连同能够正确修饰这种多肽的酶的隔离。

[0013] 发明概述

[0014] 根据本发明的一个方面, 提供了在植物或被分离的植物细胞中生产胶原的方法, 其包括在植物或被分离的植物细胞中, 以能够使至少一种类型的胶原 α 链和外源性 P4H 积聚在缺乏内源性 P4H 活性的亚细胞区室中的方式, 表达至少一种类型的胶原 α 链和外源性 P4H, 从而在植物中生产胶原。

[0015] 根据本发明的一个另外的方面, 提供了

[0016] 根据下文描述的本发明的优选实施方案中进一步的特征, 该方法进一步包括在缺乏内源性 P4H 活性的亚细胞区室中表达外源性 LH3。

[0017] 根据所述优选实施方案中再进一步的特征, 至少一种类型的胶原 α 链包括用于靶向质外体或液泡的信号肽。

[0018] 根据所述优选实施方案中再进一步的特征, 至少一种类型的胶原 α 链缺乏 ER 靶向序列或保留序列。

[0019] 根据所述优选实施方案中再进一步的特征, 至少一种类型的胶原 α 链表达于包含 DNA 的植物细胞器中。

[0020] 根据所述优选实施方案中再进一步的特征, 外源性 P4H 包括用于靶向质外体或液泡的信号肽。

[0021] 根据所述优选实施方案中再进一步的特征, 外源性 P4H 缺乏 ER 靶向序列或保留序列。

[0022] 根据所述优选实施方案中再进一步的特征, 外源性 P4H 表达于包含 DNA 的植物细胞器中。

[0023] 根据所述优选实施方案中再进一步的特征, 至少一种类型的胶原 α 链是 $\alpha 1$ 链。

[0024] 根据所述优选实施方案中再进一步的特征, 至少一种类型的胶原 α 链是 $\alpha 2$ 链。

[0025] 根据所述优选实施方案中再进一步的特征, 至少一种类型的胶原 α 链包括 C- 末端和 / 或 N- 末端前肽。

[0026] 根据所述优选实施方案中再进一步的特征, 植物选自烟草、玉米、苜蓿、稻、马铃薯、大豆、番茄、小麦、大麦、低芥酸菜子 (Canola) 和棉花。

[0027] 根据所述优选实施方案中再进一步的特征,至少一种类型的胶原 α 链或外源性 P4H 只在植物的部分中表达。

[0028] 根据所述优选实施方案中再进一步的特征,植物的部分是叶、种子、根、块茎或茎。

[0029] 根据所述优选实施方案中再进一步的特征,外源性 P4H 能够使至少一种类型的胶原 α 链的 Gly-X-Y 三联体的 Y 位置特异性羟基化。

[0030] 根据所述优选实施方案中再进一步的特征,外源性 P4H 是人 P4H。

[0031] 根据所述优选实施方案中再进一步的特征,对植物实施应激条件。

[0032] 根据所述优选实施方案中再进一步的特征,应激条件选自干旱、盐度、损伤、寒冷和喷射应激诱导化合物。

[0033] 根据本发明的另一方面,提供了基因修饰植物或被分离的植物细胞,其能够积聚具有羟基化模式的胶原 α 链,所述羟基化模式与当胶原 α 链表达于人细胞中时产生的模式相同。

[0034] 根据本发明的再一方面,提供了基因修饰植物或被分离的植物细胞,其能够在缺乏内源性 P4H 活性的亚细胞区室中积聚胶原 α 链。

[0035] 根据所述优选实施方案中再进一步的特征,基因修饰植物进一步包括外源性 P4H。

[0036] 根据所述优选实施方案中再进一步的特征,至少一种类型的胶原 α 链包括用于靶向质外体或液泡的信号肽。

[0037] 根据所述优选实施方案中再进一步的特征,至少一种类型的胶原 α 链缺乏 ER 靶向序列或保留序列。

[0038] 根据所述优选实施方案中再进一步的特征,至少一种类型的胶原 α 链表达于包含 DNA 的植物细胞器中。

[0039] 根据所述优选实施方案中再进一步的特征,外源性 P4H 包括用于靶向质外体或液泡的信号肽。

[0040] 根据所述优选实施方案中再进一步的特征,外源性 P4H 缺乏 ER 靶向序列或保留序列。

[0041] 根据所述优选实施方案中再进一步的特征,外源性 P4H 表达于包含 DNA 的植物细胞器中。

[0042] 根据所述优选实施方案中再进一步的特征,胶原 α 链是 $\alpha 1$ 链。

[0043] 根据所述优选实施方案中再进一步的特征,胶原 α 链是 $\alpha 2$ 链。

[0044] 根据所述优选实施方案中再进一步的特征,胶原 α 链包括 C-末端和 / 或 N-末端前肽。

[0045] 根据本发明的再一方面,提供了植物系统,其包括能够积聚胶原 $\alpha 1$ 链的第一种基因修饰植物,和能够积聚胶原 $\alpha 2$ 链的第二种基因修饰植物。

[0046] 根据本发明的再一方面,提供了植物系统,其包括能够积聚胶原 $\alpha 1$ 链和胶原 $\alpha 2$ 链的第一种基因修饰植物,以及能够积聚 P4H 的第二种基因修饰植物。

[0047] 根据所述优选实施方案中再进一步的特征,第一种基因修饰植物和第二种基因修饰植物中的至少一种进一步包括外源性 P4H。

[0048] 根据本发明的再一方面,提供了生产纤维胶原的方法,其包括:(a) 在第一种植物中表达胶原 $\alpha 1$ 链;(b) 在第二种植物中表达胶原 $\alpha 2$ 链;其中在第一种植物和第二种植物

中的表达被这样设置,从而使得胶原 $\alpha 1$ 链和胶原 $\alpha 2$ 链各自能够在缺乏内源性 P4H 活性的亚细胞区室中积聚;和 (c) 使第一种植物与第二种植物杂交,并选择表达胶原 $\alpha 1$ 链和胶原 $\alpha 2$ 链的后代,从而生产纤维胶原。

[0049] 根据所述优选实施方案中再进一步的特征,该方法进一步包括在第一种植物和第二种植物的每一种中表达外源性 P4H。

[0050] 根据所述优选实施方案中再进一步的特征,胶原 $\alpha 1$ 链和胶原 $\alpha 2$ 链中的每一种包括用于靶向质外体或液泡的信号肽。

[0051] 根据所述优选实施方案中再进一步的特征,胶原 $\alpha 1$ 链和胶原 $\alpha 2$ 链中的每一种缺乏 ER 靶向序列或保留序列。

[0052] 根据所述优选实施方案中再进一步的特征,步骤 (a) 和 (b) 经由在包含 DNA 的植物细胞器中的表达来实现。

[0053] 根据所述优选实施方案中再进一步的特征,外源性 P4H 包括用于靶向质外体或液泡的信号肽。

[0054] 根据所述优选实施方案中再进一步的特征,外源性 P4H 缺乏 ER 靶向序列或保留序列。

[0055] 根据所述优选实施方案中再进一步的特征,外源性 P4H 表达于包含 DNA 的植物细胞器中。

[0056] 根据所述优选实施方案中再进一步的特征,胶原 $\alpha 1$ 链和胶原 $\alpha 2$ 链中的每一种包括 C-末端和 / 或 N-末端前肽。

[0057] 根据所述优选实施方案中再进一步的特征,外源性 P4H 能够使至少一种类型的胶原 α 链的 Gly-X-Y 三联体的 Y 位置特异性羟基化。

[0058] 根据所述优选实施方案中再进一步的特征,外源性 P4H 是人 P4H。

[0059] 根据所述优选实施方案中再进一步的特征,对第一种植物和第二种植物实施应激条件。

[0060] 根据所述优选实施方案中再进一步的特征,应激条件选自干旱、盐度、损伤、重金属毒性和寒冷应激。

[0061] 根据本发明的再一方面,提供了生产纤维胶原的方法,其包括:(a) 在第一种植物中表达胶原 $\alpha 1$ 链和胶原 $\alpha 2$ 链,其中在第一种植物中的表达被这样设置,从而使得胶原 $\alpha 1$ 链和胶原 $\alpha 2$ 链各自能够在缺乏内源性 P4H 活性的亚细胞区室中积聚;(b) 在第二种植物中表达外源性 P4H,其能够在缺乏内源性 P4H 活性的亚细胞区室中积聚;和 (c) 使第一种植物与第二种植物杂交,并选择表达胶原 $\alpha 1$ 链、胶原 $\alpha 2$ 链和 P4H 的后代,从而生产纤维胶原。

[0062] 根据本发明的再一方面,提供了核酸构建体,其包括置于在植物细胞中起作用的启动子的转录控制下、编码人 P4H 的多核苷酸。

[0063] 根据所述优选实施方案中再进一步的特征,启动子选自 CaMV 35S 启动子、遍在蛋白启动子、rbcS 启动子和 SVBV 启动子。

[0064] 根据本发明的再一方面,提供了基因修饰植物或被分离的植物细胞,其能够表达胶原 $\alpha 1$ 链、胶原 $\alpha 2$ 链、P4H、LH3 和蛋白酶 C 和 / 或蛋白酶 N。

[0065] 根据所述优选实施方案中再进一步的特征,胶原 $\alpha 1$ 链和胶原 $\alpha 2$ 链各自能够在

缺乏内源性植物 P4H 活性的亚细胞区室中积聚。

[0066] 根据本发明的再一方面,提供了基因修饰植物或被分离的植物细胞,其能够积聚具有温度稳定性特征的胶原,所述温度稳定性特征与哺乳动物胶原的特征相同。

[0067] 根据所述优选实施方案中再进一步的特征,胶原是 I 型胶原。

[0068] 根据所述优选实施方案中再进一步的特征,哺乳动物胶原是人胶原。

[0069] 根据本发明的再一方面,提供了对于在植物中的表达进行最优化的胶原编码序列。

[0070] 根据所述优选实施方案中再进一步的特征,胶原编码序列为由 SEQ ID NO :1 所阐明的。

[0071] 本发明通过提供能够表达正确羟基化的胶原链成功地解决了目前已知构造的缺点,所述胶原链能够装配为具有与人胶原类似特性的胶原。

[0072] 除非另外定义,本文所使用的所有技术和科学术语具有与本发明所属领域普通技术人员通常理解相同的含义。尽管与本文所述那些相似或等同的方法和材料可以用于本发明的实践或测试中,但下文还是描述了适当的方法和材料。在冲突的情况下,以本专利说明书包括定义为准。此外,材料、方法和实施例仅是举例说明的,并不预期是限制性的。

[0073] 附图简述

[0074] 本发明在此处仅通过例子参考附图进行描述。现在特别详细地参考附图,要强调的是所显示的细节通过例子来进行并且仅用于本发明优选实施方案的举例说明性讨论的目的,并且为了提供被认为是本发明的原理和概念方面最有用和易于理解的描述而呈现。在这点上,不试图显示比基本理解本发明所需细节更多的本发明的结构细节,附图采用的描述使得本发明的几种形式如何能够在实践中体现对于本领域技术人员是显而易见的。

[0075] 在附图中:

[0076] 图 1a-d 举例说明了用于转化测试植物的各种表达盒和载体的构建。作为本研究的部分合成的所有编码序列对于烟草中的表达是最优化的。

[0077] 图 2 举例说明了各种共转化方法。每种表达盒由编码序列的短名表示。编码序列在表 1 中详细说明。每个共转化通过 2 个 pBINPLUS 二元载体来进行。每个长方形表示携带 1、2 或 3 个表达盒的单个 pBINPLUS 载体。启动子和终止子在实施例 1 中详细说明。

[0078] 图 3 是转化体的多重 PCR 筛选,显示出对于胶原 $\alpha 1$ 链 (324bp 片段) 或胶原 $\alpha 2$ 链 (537bp 片段) 或二者阳性的植物。

[0079] 图 4 是由共转化 2、3 和 4 产生的转基因植物的蛋白质印迹分析。

[0080] 从烟草共转化体 #2、#3 和 #4 中提取总可溶性蛋白,并用抗 - 胶原 I 抗体 (来自 Chemicon Inc. 的 #AB745) 进行检测。大小标记是来自 Fermentas Inc. 的 #SM0671。W. T. 是野生型烟草。阳性胶原条带在对于 I 型胶原 $\alpha 1$ 或 $\alpha 2$ 或二者为 PCR 阳性的植物中可见。来自人胎盘的 500ng I 型胶原 (来自 Chemicon Inc. 的 #CC050, 通过胃蛋白酶消化从人胎盘中提取) 的阳性对照条带代表来自转基因植物的样品中的约 0.3% 总可溶性蛋白 (约 150 μ g)。如通过 I 型胶原抗体的抗羧基末端前肽 (来自 Chemicon Inc. 的 #MAB1913) 所检测的,在人胶原样品中约 140kDa 的较大条带是含有其 C- 前肽的前胶原。在人胶原样品中约 120kDa 的较小条带是不含前肽的胶原。由于其罕见的组成,富含脯氨酸的蛋白质 (包括胶原) 作为具有高于预期的分子量的条带在聚丙烯酰胺凝胶上一致迁移。因此,分子量

约 95kDa、不含前肽的胶原链作为约 120kDa 的条带迁移。

[0081] 图 5 是由共转化 #8 (携带与胶原链翻译地融合的质外体 (appoplast) 信号) 产生的转基因植物的蛋白质印迹分析。从转基因烟草叶中提取总可溶性蛋白,并用抗-胶原 I 抗体 (来自 Chemicon Inc. 的 #AB745) 进行检测。阳性胶原 $\alpha 2$ 条带在植物 8-141 中可见。来自人胎盘的 I 型胶原 (来自 Chemicon Inc. 的 #CC050) 充当对照。

[0082] 图 6a-b 举例说明根据热处理和胰蛋白酶或胃蛋白酶消化证实的胶原三股螺旋装配和热稳定性,在图 6a 中 - 对来自烟草 2-9 (只表达胶原 $\alpha 1$ 且不表达 P4H) 和 3-5 (表达胶原 $\alpha 1+2$ 以及人 P4H α 和 β 亚基) 的总可溶性蛋白实施热处理 (38°C 或 43°C 15 分钟),随后为胰蛋白酶消化 (室温 20 分钟),并用抗-胶原 I 抗体在蛋白质印迹方法中进行检测。阳性对照是 500ng 人胶原 I 样品 + 野生型烟草的总可溶性蛋白。在图 6b 中 - 从转基因烟草 13-6 (表达胶原 I $\alpha 1$ 和 $\alpha 2$ 链 - 由箭标指出,人 P4H α 和 β 亚基以及人 LH3) 提取总可溶性蛋白,并实施热处理 (33°C、38°C 或 42°C 20 分钟),立即在冰上冷却以防止三股螺旋重新装配,并于室温 (约 22°C) 与胃蛋白酶一起温育 30 分钟,随后在标准蛋白质印迹方法中用抗-胶原 I 抗体 (来自 Chemicon Inc. 的 #AB745) 进行检测。阳性对照是 ~ 50ng 人胶原 I 样品 (来自 Chemicon Inc. 的 #CC050,通过胃蛋白酶消化从人胎盘中提取),将其加到从野生型烟草中提取的总可溶性蛋白中。

[0083] 图 7 举例说明了对野生型烟草进行的 RNA 印迹分析。印迹用烟草 P4HcDNA 进行探测。

[0084] 图 8 是由共转化 2、3 和 13 产生的转基因植物的蛋白质印迹分析。从烟草共转化体提取总可溶性蛋白,并用抗人 P4H α 和 β 以及抗胶原 I 抗体进行检测。

[0085] 优选实施方案的描述

[0086] 本发明是关于表达并积聚胶原的植物,所述植物可用于生产显示哺乳动物胶原特征的胶原和胶原纤维。

[0087] 参考附图和附加的描述可以更好地理解本发明的原理和操作。

[0088] 在详细解释本发明的至少一个实施方案之前,应当理解本发明在其应用方面不局限于下文描述中阐明或由实施例示例的细节。本发明能够是其他实施方案、或以各种方式加以实践或执行。同样,应当理解本文所采用的措辞和术语是用于描述的目的且不应被视为限制。

[0089] 生产胶原的植物是本领域已知的。尽管此类植物可以用于生产胶原链以及胶原,但此类链是不正确地羟基化的,并因此无论是在植物中还是不在植物中,其自身的装配导致内在不稳定的胶原。

[0090] 在将本发明变为实践的同时,本发明人设计了植物表达方法,它确保胶原链的正确羟基化并因此使得能够在植物中生产胶原,所述胶原紧密地模仿人 I 型胶原的特征 (例如,温度稳定性)。

[0091] 因此,根据本发明的一个方面,提供了基因修饰植物,它能够表达至少一种类型的胶原 α 链,并将其积聚在缺乏内源性 P4H 活性的亚细胞区室中。

[0092] 如本文所使用的,短语“基因修饰植物”指任何低等 (例如藓类) 或高等 (维管) 植物或组织或其被分离的细胞 (例如细胞悬浮液),它用外源性多核苷酸序列稳定或瞬时转化。植物的例子包括烟草、玉米、苜蓿、稻、马铃薯、大豆、番茄、小麦、大麦、低芥酸菜子、棉

花、胡萝卜以及低等植物例如藓类。

[0093] 如本文所使用的,短语“胶原链”指胶原亚基,例如胶原纤维的 $\alpha 1$ 或 2 链,优选 I 型纤维。如本文所使用的,短语“胶原”指被装配的胶原三聚体,在 I 型胶原的情况下它包括 2 个 $\alpha 1$ 链和 1 个 $\alpha 2$ 链。胶原纤维是缺乏末端前肽 C 和 N 的胶原。

[0094] 如本文所使用的,短语“缺乏内源性 P4H 活性的亚细胞区室”指细胞的任何被区室化的区域,它不包括植物 P4H 或具有植物样 P4H 活性的酶。此类亚细胞区室的例子包括液泡、质外体和细胞质,以及细胞器例如叶绿体、线粒体等。

[0095] 任何类型的胶原链都可以由本发明的基因修饰植物表达。例子包括形成原纤维的胶原(I、II、III、V 和 XI 型)、形成网络的胶原(IV、VIII 和 X 型)、与原纤维表面结合的胶原(IX、XII 和 XIV 型)、作为跨膜蛋白存在的胶原(XIII 和 XVII 型)、或形成 11nm 周期性珠状丝的胶原(VI 型)。关于进一步的描述请参见 Hulmes, 2002。

[0096] 优选地,表达的胶原链是 I 型胶原 $\alpha 1$ 和 2 链。表达的胶原 α 链可以由来源于任何哺乳动物的任何多核苷酸序列编码。优选地,编码胶原 α 链的序列是人的并且由 SEQ ID NO :1 和 4 阐明。

[0097] 一般地,在植物中表达的 α 胶原链可以包括或不包括其末端前肽(即前肽 C 和前肽 N)。

[0098] Ruggiero 等人(2000)指出前胶原经由植物蛋白水解活性的处理不同于在人中的正常处理,并且前肽 C 通过植物蛋白水解活性去除,尽管切割位点不明。前肽 C 的切割可以在三聚体装配之前(3 种 C-前肽的结合是发动三聚体装配所必需的)发生在前胶原肽上。

[0099] 经由植物蛋白水解活性的 N-前肽切割发生在成熟植物中而不发生在小植物中。此类切割去除来自 N 端肽的 2 个氨基酸(17 个中的 2 个)。

[0100] C-前肽(以及较少程度上的 N-前肽)在前胶原通过动物细胞的过程中保持前胶原可溶(Bulleid 等人,2000),并且预期在植物细胞中有类似作用。在前胶原分子分泌进入细胞外基质后或在这个过程中,通过前胶原 N-和 C-蛋白酶去除前肽,从而触发胶原分子自发地自身装配为原纤维(Hulmes, 2002)。通过前胶原 N-和 C-蛋白酶去除前肽使前胶原的溶解性降低了 > 10000 倍,并且是发动胶原自身装配成纤维所必需和充分的。对这个装配过程关键的是在三股螺旋结构域末端被称为端肽的短的非三股螺旋肽,它确保原纤维结构内的胶原分子的正确记录,并降低自身装配的临界浓度(Bulleid 等人,2000)。现有技术描述了在胶原生产过程中利用胃蛋白酶切割前肽(Bulleid 等人 2000)。然而胃蛋白酶破坏端肽,并且因此,胃蛋白酶提取的胶原不能形成有序的原纤维结构(Bulleid 等人 2000)。

[0101] 形成人 P4H β 亚基的蛋白质二硫键异构酶(PDI)显示在三聚体装配之前与 C-前肽结合,因此在链的装配过程中还充当分子伴侣(Ruggiero 等人,2000)。在不同植物中表达的人前胶原 I N-蛋白酶和前胶原 C-蛋白酶的使用可以产生更类似于天然的人胶原的胶原,并且可以形成有序的原纤维结构。

[0102] 在其中 N 或 C 前肽或二者包括在表达的胶原链中的情况下,本发明的基因修饰植物还可表达分别的蛋白酶(即, C 或 N 或二者)。编码此类蛋白酶的多核苷酸序列由 SEQ ID NO :18(蛋白酶 C)和 20(蛋白酶 N)示例。此类蛋白酶可以被这样表达,从而使得它们在与胶原链相同的亚细胞区室中积聚。

[0103] 表达的胶原链在缺乏内源性 P4H 活性的亚细胞区室中的积聚可以经由几种方法

中的任何一种来实现。

[0104] 例如,表达的胶原链可以包括用于将表达的蛋白靶向亚细胞区室例如质外体或细胞器(例如叶绿体)的信号序列。适当的信号序列的例子包括叶绿体转运肽(包括在 Swiss-Prot 的登录号 P07689 中,氨基酸 1-57)和线粒体转运肽(包括在 Swiss-Prot 的登录号 P46643 中,氨基酸 1-28)。下文实施例部分提供了适当的信号序列的另外例子以及关于在植物细胞的胶原链表达中采用此类信号序列的指导方针。

[0105] 可替代地,当在植物中表达时,胶原链的序列可以以改变胶原细胞定位的方式进行修饰。

[0106] 如上文所提及的,植物的 ER 包括不能使胶原链正确羟基化的 P4H。胶原 α 链天然地包括 ER 靶向序列,它指导表达的胶原进入 ER,在其中它被翻译后修饰(包括不正确的羟基化)。因此,ER 靶向序列的去除将导致缺乏翻译后修饰(包括任何羟基化)的胶原链的细胞质积聚。

[0107] 下文实施例部分的实施例 1 描述了缺乏 ER 序列的胶原序列的生产。

[0108] 再可替代地,胶原链可以被表达并积聚在包含 DNA 的细胞器例如叶绿体或线粒体中。叶绿体表达的进一步描述在下文中提供。

[0109] 如上文所提及的, α 链的羟基化是稳定的 I 型胶原装配所必需的。因为通过本发明基因修饰植物表达的 α 链积聚在缺乏内源性 P4H 活性的区室内,所以必须从植物、植物组织或细胞中分离此类链并在体外进行羟基化。此类羟基化可以通过 Turpeenniemi-Hujanen 和 Myllyla (Concomitant hydroxylation of proline and lysine residues in collagen using purified enzymes in vitro. *Biochim Biophys Acta*. 1984 Jul 16 ;800 (1) :59-65) 所述方法实现。

[0110] 尽管此类体外羟基化可以产生正确羟基化的胶原链,但它的实现是困难且代价高的。

[0111] 为了克服体外羟基化的局限,本发明的基因修饰植物优选还共表达 P4H,它能使胶原 α 链正确羟基化[即只在 Gly-X-Y 三联体的脯氨酸(Y)位置上羟基化]。P4H 是由 2 个亚基 α 和 β 组成的酶。二者都是形成活性酶所必需的,而 β 亚基还具有伴侣功能。

[0112] 通过本发明的基因修饰植物表达的 P4H 优选是人 P4H,它由例如 SEQ ID NO:12 和 14 编码。此外,还可使用显示增强的底物特异性的 P4H 突变体或 P4H 同源物。

[0113] 适当的 P4H 同源物由 NCBI 编号 NP_179363 鉴定的鼠耳芥属 (*Arabidopsis*) 氧化还原酶示例。由本发明人进行的这种蛋白质序列和人 P4H α 亚基的逐对比对揭示了任何已知的植物 P4H 同源物的功能结构域之间的最高同源性。

[0114] 因为 P4H 需要与表达的胶原链共积聚,所以其编码序列优选相应地进行修饰(加入信号序列、可能阻止 ER 靶向的缺失,等等)。

[0115] 在哺乳动物细胞中,胶原还通过赖氨酰羟化酶、半乳糖基转移酶和葡糖基转移酶进行修饰。这些酶序贯地将特定位置的赖氨酰残基修饰为羟赖氨酰、半乳糖基羟赖氨酰和葡糖基半乳糖基羟赖氨酰残基。单独的人的酶,赖氨酰羟化酶 3 (LH3) 可以催化羟赖氨酸连接的碳水化合物形成中的所有 3 个连续步骤。

[0116] 因此,本发明的基因修饰植物还优选表达哺乳动物 LH3。编码 LH3 的序列例如 SEQ ID NO:22 所阐明的那些可以用于此类用途。

[0117] 上文描述的胶原链和修饰酶可以由稳定整合或瞬时表达的核酸构建体进行表达,所述核酸构建体包括编码 α 链和 / 或修饰酶 (例如 P4H 和 LH3)、置于植物功能启动子的转录控制下的多核苷酸序列。此类核酸构建体 (它在本文中也被称为表达构建体) 可以被设置为在遍及完整植物、限定的植物组织或限定的植物细胞、或在植物的限定发育阶段进行表达。此类构建体还可包括选择标记 (例如抗生素抗性)、增强子元件和用于细菌复制的复制起点。

[0118] 应当理解包括 2 种可表达的插入片段 (例如 2 个 α 链类型、或 α 链和 P4H) 的构建体优选包括适合于每种插入片段的单独启动子,或可替代地,此类构建体可以表达包括来自单个启动子的 2 个插入序列的单个转录物嵌合体。在这种情况下,嵌合转录物包括在 2 个插入序列之间的 IRES 序列,从而使得下游插入片段可以从那里被翻译。

[0119] 本发明的构建体可以采用众多植物功能表达启动子和增强子,它们可以是组织特异性、发育特异性、组成型或诱导型的,下文提供了一些例子。

[0120] 如本说明书此处和随后的权利要求部分所使用的,短语“植物启动子”或“启动子”包括可以指导植物细胞 (包括包含 DNA 的细胞器) 中的基因表达的启动子。此类启动子可以来源于植物、细菌、病毒、真菌或动物来源。此类启动子可以是组成型的,即能够指导在多个植物组织中的高水平基因表达,组织特异性的,即能够指导在一种或多种特定植物组织中的基因表达,诱导型的,即能够在刺激下指导基因表达,或嵌合的,即由至少 2 个不同的启动子部分形成。

[0121] 因此,所使用的植物启动子可以是组成型启动子、组织特异性启动子、诱导型启动子或嵌合启动子。

[0122] 组成型植物启动子的例子包括,但不限于, CaMV35S 和 CaMV19S 启动子、FMV34S 启动子、甘蔗杆状病毒 (sugarcane bacilliformbadnavirus) 启动子、CsVMV 启动子、鼠耳芥 ACT2/ACT8 肌动蛋白启动子、鼠耳芥遍在蛋白 UBQ1 启动子、大麦叶硫素 BTH6 启动子和稻肌动蛋白启动子。

[0123] 组织特异性启动子的例子包括,但不限于,豆菜豆蛋白贮存蛋白启动子、DLEC 启动子、PHS 启动子、玉米醇溶蛋白贮存蛋白启动子、来自大豆的羽扇豆球蛋白 γ 启动子、AT2S1 基因启动子、来自鼠耳芥的 ACT11 肌动蛋白启动子、来自欧洲油菜 (*Brassica napus*) 的 napA 启动子和马铃薯 patatin 基因启动子。

[0124] 诱导型启动子是由特异性的刺激例如应激条件 (包括例如光、温度、化学制剂、干旱、高盐度、渗透压休克)、氧化剂条件或在致病性的情况下诱导的启动子,并且包括,但不限于,来源于豌豆 rbcS 基因的光诱导型启动子,来自苜蓿 rbcS 基因的启动子,在干旱时活化的启动子 DRE、MYC 和 MYB; 在高盐度和渗透压应激下活化的启动子 INT、INPS、prxEa、Ha hsp17.7G4 和 RD21,以及在致病性应激下活化的启动子 hsr203J 和 str246C。

[0125] 优选地,本发明使用的启动子是强组成型启动子,从而使得构建体插入片段的超表达在植物转化后实现。

[0126] 应当理解,本发明所使用的任何构建体类型都可以利用每种构建体类型中相同或不同的选择标记共转化到相同的植物中。可替代地,第一种构建体类型可以导入第一种植物中,而第二种构建体类型可以导入第二种等基因植物中,之后由此产生的转基因植物可以进行杂交并选择双重转化体的后代。此类后代的进一步自交可以用于产生 2 种构建体的

纯合系。

[0127] 存在将核酸构建体导入单子叶植物和双子叶植物的各种方法 (Potrykus, I., *Annu. Rev. Plant. Physiol., Plant. Mol. Biol.* (1991)42 :205-225 ;Shimamoto 等人, *Nature*(1989)338 :274-276)。此类方法依赖于核酸构建体或其部分稳定整合到植物基因组中,或在其中这些序列不由植物后代遗传的情况下依赖于核酸构建体的瞬时表达。

[0128] 此外,存在可以将核酸构建体直接导入包含 DNA 的细胞器例如叶绿体的 DNA 中的几种方法。

[0129] 存在实现外源序列例如包括在本发明的核酸构建体内的那些稳定地基因组整合到植物基因组中的 2 种基本方法:

[0130] (i) 土壤杆菌属 (*Agrobacterium*)-介导的基因转移 :Klee 等人 (1987)*Annu. Rev. Plant Physiol.* 38 :467-486 ;Klee 和 Rogers, *Cell Culture and Somatic Cell Genetics of Plants*, 第 6 卷, *Molecular Biology of Plant Nuclear Genes*, eds. Schell, J., 和 Vasil, L. K., Academic Publishers, San Diego, Calif. (1989) 第 2-25 页 ;Gatenby, *Plant Biotechnology*, eds. Kung, S. 和 Arntzen, C. J., Butterworth Publishers, Boston, Mass. (1989) 第 93-112 页。

[0131] (ii) DNA 直接摄取 :Paszkowski 等人, *Cell Culture and Somatic Cell Genetics of Plants*, 第 6 卷, *Molecular Biology of Plant Nuclear Genes* eds. Schell, J., 和 Vasil, L. K., Academic Publishers, San Diego, Calif. (1989) 第 52-68 页 ;包括用于将 DNA 直接摄入原生质体中的方法, Toriyama, K. 等人 (1988)*Bio/Technology* 6 :1072-1074。通过植物细胞短暂电休克诱导的 DNA 摄取 :Zhang 等人 *Plant Cell Rep.* (1988)7 :379-384。Fromm 等人 *Nature* (1986)319 :791-793。通过粒子轰击, Klein 等人 *Bio/Technology* (1988)6 :559-563 ;McCabe 等人 *Bio/Technology* (1988)6 :923-926 ;Sanford, *Physiol. Plant.* (1990)79 :206-209 ;通过使用微量移液管系统, Neuhaus 等人, *Theor. Appl. Genet.* (1987)75 :30-36 ;Neuhaus 和 Spangenberg, *Physiol. Plant.* (1990)79 :213-217 ;或通过将 DNA 与发芽的花粉直接温育, DeWet 等人 *Experimental Manipulation of Ovule Tissue*, eds. Chapman, G. P. 和 Mantell, S. H. 和 Daniels, W. Longman, London, (1985) 第 197-209 页 ;和 Ohta, *Proc. Natl. Acad. Sci. USA* (1986)83 :715-719, 将 DNA 注射到植物细胞或组织中。

[0132] 土壤杆菌属系统包括质粒载体的使用,所述质粒载体包含整合到植物基因组 DNA 中的限定的 DNA 片段。植物组织的接种方法依赖于植物物种和土壤杆菌属递送系统而变。广泛使用的方法是叶盘法,它可以由任何组织外植体来进行,所述外植体提供用于发动整个植物分化的良好来源。Horsch 等人 *Plant Molecular Biology Manual A5*, Kluwer Academic Publishers, Dordrecht (1988) 第 1-9 页。补充方法采用与真空渗透组合的土壤杆菌属递送系统。土壤杆菌属系统在转基因双子叶植物的制备中尤其可行。

[0133] 存在直接将 DNA 转移到植物细胞中的各种方法。在电穿孔中,使原生质体短暂地暴露在强电场中。在显微注射中,利用极小的微量移液管将 DNA 机械地直接注射到细胞中。在微粒轰击中, DNA 吸附在微粒例如硫酸镁晶体、钨颗粒或金颗粒上,并且微粒物理加速进入细胞或植物组织中。

[0134] 转化后进行植物繁殖。最常见的植物繁殖法是通过种子。然而通过种子繁殖的

再生具有由于杂合性而在农作物中缺少一致性的缺陷,因为根据孟德尔定律决定的遗传变异,种子通过植物产生。基本上,每个种子在遗传上不同并且每个将伴随其自身的特殊性状生长。因此,优选被转化的植物被这样生产,从而使得再生的植物具有与亲本转基因植物相同的性状和特征。因此,优选被转化的植物通过微繁殖(micropropagation)进行再生,所述微繁殖提供被转化植物的快速、一致的生殖。

[0135] 可以用于瞬时表达包括在本发明的核酸构建体内的被分离核酸的瞬时表达法包括,但不限于,如上所述的显微注射和轰击,但在有利于瞬时表达的条件下,和病毒介导的表达,其中包括核酸构建体的被包装或未包装的重组病毒载体用于感染植物组织或细胞,从而使得其中确立的繁殖的重组病毒表达非病毒的核酸序列。

[0136] 已显示对于植物宿主转化有用的病毒包括 CaMV、TMV 和 BV。利用植物病毒的植物转化描述于美国专利号 4,855,237(BGV)、EP-A67,553(TMV)、日本公开申请号 63-14693(TMV)、EPA 194,809(BV)、EPA 278,667(BV); 和 Gluzman, Y. 等人, *Communications in Molecular Biology: Viral Vectors*, Cold Spring Harbor Laboratory, New York, pp. 172-189(1988) 中。用于在多种宿主包括植物中表达外来 DNA 的假病毒颗粒描述于 WO 87/06261 中。

[0137] 用于在植物中导入和表达非病毒性外源核酸序列的植物 RNA 病毒的构建由上述参考文献以及 Dawson, W. O. 等人, *Virology* (1989) 172 :285-292 ;Takamatsu 等人 *EMBOJ.* (1987) 6 :307-311 ;French 等人 *Science* (1986) 231 :1294-1297 ;和 Takamatsu 等人 *FEBS Letters* (1990) 269 :73-76 证实。

[0138] 当病毒是 DNA 病毒时,可以对病毒自身进行构建。可替代地,病毒可以首先克隆到细菌质粒中以用于容易地构建含有外来 DNA 的所需病毒载体。病毒随后可以从质粒中切除。如果病毒是 DNA 病毒,则可以将细菌的复制起点附着在病毒 DNA 上,所述病毒 DNA 随后通过细菌进行复制。这种 DNA 的转录和翻译将产生外壳蛋白,它将使病毒 DNA 壳体化。如果病毒是 RNA 病毒,则该病毒通常作为 cDNA 被克隆并插入质粒中。质粒随后用于制备所有构建体。RNA 病毒随后通过转录质粒的病毒序列并翻译病毒基因以产生使病毒 RNA 壳体化的外壳蛋白来产生。

[0139] 用于在植物中导入和表达非病毒性外源核酸序列的植物 RNA 病毒的构建由上述参考文献以及美国专利号 5,316,931 证实,所述外源核酸序列例如本发明构建体中包括的那些。

[0140] 在一个实施方案中,提供了植物病毒性核酸,其中天然的外壳蛋白编码序列已从病毒核酸中缺失,已插入非天然的植物病毒外壳蛋白的编码序列和非天然的启动子,优选非天然的外壳蛋白编码序列的亚基因组启动子,其能够在植物宿主中表达,包装重组的植物病毒核酸,并确保通过重组的植物病毒核酸全身地感染宿主。可替代地,外壳蛋白基因可以通过在其中插入非天然的核酸序列来灭活,从而产生蛋白质。重组的植物病毒核酸可以包含一个或多个另外的非天然的亚基因组启动子。每个非天然的亚基因组启动子都能够在植物宿主中转录或表达邻近基因或核酸序列,并且不能互相重组以及与天然的亚基因组启动子重组。如果包括超过一个的核酸序列,则插入的非天然(外来)核酸序列可以邻近天然的植物病毒亚基因组启动子、或邻近天然的和非天然的植物病毒亚基因组启动子。非天然的核酸序列在亚基因组启动子控制下在宿主植物中被转录或表达,以生产所需产品。

[0141] 在第二个实施方案中,提供的重组的植物病毒核酸与第一个实施方案中的一样,只是除了将天然的外壳蛋白编码序列置于邻近一个非天然的外壳蛋白亚基因组启动子而不是非天然的外壳蛋白编码序列。

[0142] 在第三个实施方案中,提供重组的植物病毒核酸,其中天然外壳蛋白的基因邻近其亚基因组启动子,且一个或多个非天然亚基因组启动子已插入到病毒核酸中。插入的非天然的亚基因组启动子能够在植物宿主中转录或表达邻近基因,并且不能互相重组及与天然的亚基因组启动子重组。插入的非天然的核酸序列可以邻近非天然的亚基因组植物病毒启动子,从而使得所述序列在亚基因组启动子控制下在宿主植物中被转录或表达,以生产所需产品。

[0143] 在第四个实施方案中,提供的重组的植物病毒核酸与第三个实施方案中的一样,只是除了天然的外壳蛋白编码序列由非天然的外壳蛋白编码序列替代。

[0144] 病毒载体通过由重组的植物病毒核酸编码的外壳蛋白进行壳体化,以生产重组的植物病毒。重组的植物病毒核酸或重组的植物病毒用于感染适当的宿主植物。重组的植物病毒核酸能够在宿主中复制、在宿主中全身地扩散、并且在宿主中转录或表达外来基因(被分离的核酸),以生产所需蛋白。

[0145] 用于将外源核酸序列导入叶绿体基因组的技术是已知的。这种技术包括下列过程。首先,植物细胞被化学处理以便将每个细胞的叶绿体数目减少至约一个。随后,外源核酸经由粒子轰击导入细胞中以用于将至少一种外源核酸分子导入叶绿体中。选择外源核酸从而使得它能够经由同源重组整合到叶绿体基因组中,所述同源重组易于通过叶绿体固有的酶来实现。为此,外源核酸除了目的基因之外还包括至少一个核酸段,它来源于叶绿体基因组。此外,外源核酸包括选择标记,它通过序贯选择过程以用于确定所有或基本上所有的叶绿体基因组拷贝在此类选择后将包括外源核酸。关于这种技术的进一步的细节在美国专利号 4,945,050 和 5,693,507 中找到,所述专利引入本文作为参考。因此多肽可以通过叶绿体的蛋白质表达系统来生产并且变得整合到叶绿体内膜中。

[0146] 上述转化方法可以用于在任何物种的植物、或植物组织或来源于那里的被分离的植物细胞中生产胶原链和 / 或修饰酶以及被装配的胶原(含或不含前肽)。

[0147] 优选的植物是能够积聚大量本文描述的胶原链、胶原和 / 或加工酶的那些。还可根据其对应激条件的抗性和其中表达的组分或被装配的胶原可以被提取的容易性来选择此类植物。优选的植物的例子包括烟草、玉米、苜蓿、稻、马铃薯、大豆、番茄、小麦、大麦、低芥酸菜子和棉花。

[0148] 胶原纤维广泛地用于食物和化妆品工业中。因此,尽管由植物表达的胶原纤维组分(α 链)和修饰酶在胶原的工业合成中找到功用,但由于其简单性和成本效率,优选在植物中生产完整的胶原。

[0149] 几种方法可以用于在植物中生成 I 型胶原。例如可以从表达胶原 $\alpha 1$ 和 P4H(和任选的 LH3)的植物中分离胶原 $\alpha 1$ 链并且将其与胶原 $\alpha 2$ 链混合,所述胶原 $\alpha 2$ 链从表达胶原 $\alpha 2$ 和 P4H(和任选的 LH3 以及蛋白酶 C 和 / 或 N)的植物中分离。因为胶原 $\alpha 1$ 链自动地自身装配为三股螺旋,所以在混合之前可能必须使此类同源三聚体变性,且与胶原 $\alpha 2$ 链一起复性。

[0150] 优选地,表达胶原 $\alpha 1$ 和 P4H(和任选的 LH3 以及蛋白酶 C 和 / 或 N)的第一种植

物可以与表达胶原 $\alpha 2$ 的第二种（且优选等基因的）植物杂交，或可替代地，表达 2 种 α 链的第一种植物可以与表达 P4H 和任选的 LH3 以及蛋白酶 C 和 / 或 N 的第二种植物杂交。

[0151] 应当指出尽管上述植物育种方法使用 2 种单独转化的植物，但使用 3 种或更多种单独转化的植物的方法也可采用，所述植物各自表达 1 种或 2 种组分。

[0152] 本领域的普通技术人员知道各种植物育种技术并且因此本文不提供此类技术的进一步描述。

[0153] 尽管植物育种方法是优选的，但应当指出表达胶原 $\alpha 1$ 和 2、P4H 和 LH3（和任选的蛋白酶 C 和 / 或 N）的单个植物可以经由数个转化事件产生，每个所述转化事件被设计用于将再一个可表达的组分导入细胞中。在这种情况下，每个转化事件的稳定性可以利用特异性的选择标记进行验证。

[0154] 在任何情况下，转化和植物育种方法都可以用于生成表达任何数目的组分的任何植物。目前优选的是表达胶原 $\alpha 1$ 和 2 链、P4H、LH3 和至少一种蛋白酶（例如蛋白酶 C 和 / 或 N）的植物。如下文实施例部分将进一步描述的，此类植物积聚在最高达 42°C 的温度下显示出稳定性的胶原。

[0155] 由育种产生的后代，或可替代地，多重转化的植物可以通过使用核酸或蛋白质探针（例如抗体）证实外源 mRNA 和 / 或多肽的存在来选择。后面的方法是优选的，因为它能够定位所表达的多肽组分（通过例如，探测分级分离的植物提取物），并因此同样证实了用于正确加工和装配的潜力。适当的探针的例子在下文实施例部分提供。

[0156] 一旦表达胶原的后代被鉴别，此类植物就进一步在使胶原链以及修饰酶最大化表达的条件下进一步培育。

[0157] 因为游离的脯氨酸积聚可以促进不同的富含脯氨酸的蛋白质的过度生产，所述蛋白质包括由本发明的基因修饰植物表达的胶原链，所以优选的培育条件是增加被培育植物中游离脯氨酸积聚的那些。

[0158] 游离脯氨酸响应于广泛的环境应激而在多种植物中积聚，所述环境应激包括剥夺、盐化、低温、高温、病原体感染、重金属毒性、无氧生活、养分缺乏、大气污染和紫外线照射（Hare 和 Cress, 1997）。

[0159] 游离脯氨酸还可响应于经由化合物例如 ABA，或应激诱导化合物例如铜盐、百草枯（paraquat）、水杨酸等处理植物或土壤而积聚。

[0160] 因此，表达胶原的后代可以在不同的应激条件（例如 50mM- 最高达 250mM 的不同的 NaCl 浓度）下生长。为了进一步增强胶原的生产，将检查各种应激条件对胶原表达的作用，并关于植物生活力、生物量和胶原积聚进行最优化。

[0161] 植物组织 / 细胞优选在成熟期进行收获，并且利用众所周知的现有技术的提取方法分离胶原纤维，下文描述了一种此类方法。

[0162] 转基因植物的叶子在液氮下被磨成粉末，并且于 4°C 在包含 0.2M NaCl 的 0.5M 乙酸中提取匀浆物 60 小时。通过离心去除不溶的材料。包含重组胶原的上清液在 0.4M 和 0.7M NaCl 下进行盐分级分离。包含重组的异源三聚体胶原的 0.7M NaCl 沉淀物溶解于 0.1M 乙酸中并针对其进行透析，并储存于 -20°C（根据 Ruggiero 等人，2000）。

[0163] 在检查下列实施例后本发明的其他目的、优点和新型特征对于本领域的普通技术人员而言将是显而易见的，所述实施例不预期是限制性的。另外，如上文所描绘的和下文权

利要求部分所请求保护的发明的各种实施方案和方面中的每一个在下列实施例中都能找到实验支持。

[0164] 实施例

[0165] 现在参考下列实施例,所述实施例与上述说明书一起以非限制性的方式举例说明本发明。

[0166] 一般地,本文所使用的术语和本发明中使用的实验程序包括分子、生物化学、微生物学和重组DNA技术。此类技术在文献中得到充分说明。参见,例如“Molecular Cloning :A laboratory Manual”Sambrook等人(1989);“Current Protocols in Molecular Biology”第I-III卷 Ausubel, R. M., ed. (1994); Ausubel等人,“Current Protocols in Molecular Biology”, John Wiley and Sons, Baltimore, Maryland(1989); Perbal, “A Practical Guide to Molecular Cloning”, John Wiley & Sons, New York(1988); Watson等人,“Recombinant DNA”, Scientific American Books, New York; Birren等人(eds), “Genome Analysis: A Laboratory Manual Series”, 第1-4卷, Cold Spring Harbor Laboratory Press, New York(1998); 美国专利号 4,666,828; 4,683,202; 4,801,531; 5,192,659 和 5,272,057 中所阐明的方法; “Cell Biology :A Laboratory Handbook”, 第I-III卷 Cellis, J. E., ed. (1994); “Current Protocols in Immunology”第I-III卷 Coligan J. E., ed. (1994); Stites等人(eds), “Basic and Clinical Immunology”(第8版), Appleton & Lange, Norwalk, CT(1994); Mishell 和 Shiigi(eds), “Selected Methods in Cellular Immunology”, W. H. Freeman and Co., New York(1980); 可用的免疫测定广泛描述于专利和科学文献中,参见,例如,美国专利号 3,791,932; 3,839,153; 3,850,752; 3,850,578; 3,853,987; 3,867,517; 3,879,262; 3,901,654; 3,935,074; 3,984,533; 3,996,345; 4,034,074; 4,098,876; 4,879,219; 5,011,771 和 5,281,521; “Oligonucleotide Synthesis” Gait, M. J., ed. (1984); “Nucleic Acid Hybridization” Hames, B. D., 和 Higgins S. J., eds. (1985); “Transcription and Translation” Hames, B. D., 和 Higgins S. J., Eds. (1984); “Animal Cell Culture” Freshney, R. I., ed. (1986); “Immobilized Cells and Enzymes” IRL Press, (1986); “A Practical Guide to Molecular Cloning” Perbal, B., (1984) 和 “Methods in Enzymology” 第1-317卷, Academic Press; “PCR Protocols :A Guide To Methods And Applications”, Academic Press, San Diego, CA(1990); Marshak等人, “Strategies for Protein Purification and Characterization-A Laboratory Course Manual” CSHL Press(1996); 所有这些都如本文完全阐明的那样引入本文作为参考。这个文件自始至终提供了其他的一般参考文献。其中的程序被认为是本领域众所周知的并且为了读者的方便而提供。其中包含的所有信息都引入本文作为参考。

[0167] 实施例 1

[0168] 构建体和转化方案

[0169] 图 1a-d 中举例说明了在这项工作中使用的表达盒和载体的构建。这项工作中的所有编码序列都对于在烟草中的表达进行最优化,并与所需侧翼区(SEQ ID NO :1、4、7、12、14、16、18、20、22)一起进行化学合成。图 1a- 编码与液泡信号或质外体信号(由 SEQ IDNO : 7 编码)融合的、或不含信号的 Co11 和 Co12 的合成基因(SEQ ID1、4)在由菊花 rbcS1 启

动子和 5' UTR(SEQ ID NO:10) 以及菊花 rbcS13' UTR 和终止子 (SEQ ID NO:11) 组成的表达盒中进行克隆。将完整的表达盒克隆到 pBINPLUS 植物转化载体的多克隆位点中 (van Engelen 等人,1995, Transgenic Res 4:288-290)。图 1b- 编码与液泡信号或质外体信号 (由 SEQ ID NO:7 编码) 融合的、或不含信号的 P4H β - 人、P4H α - 人和 P4H- 植物 (SEQ ID NO:12、14 和 16) 的合成基因在由载体 pJD330 携带的 CaMV 35S 启动子和 TMV ω 序列和土壤杆菌属胭脂碱合成酶 (NOS) 终止子组成的表达盒中进行克隆 (Galili 等人,1987, Nucleic Acids Res 15:3257-3273)。将完整的表达盒克隆到携带 Co11 或 Co12 表达盒的 pBINPLUS 载体的多克隆位点中。图 1c- 编码与液泡信号或质外体信号 (由 SEQ ID NO:7 编码) 融合的蛋白酶 C 和蛋白酶 N (SEQ ID NO:18、20) 的合成基因在由菊花 rbcS1 启动子和 5' UTR (SEQ ID NO:10) 以及菊花 rbcS13' UTR 和终止子 (SEQ ID NO:11) 组成的表达盒中进行克隆。将完整的表达盒克隆到 pBINPLUS 植物转化载体的多克隆位点中。图 1d- 编码与液泡信号或质外体信号 (由 SEQ ID NO:7 编码) 融合的、或不含信号的、具有侧翼的草莓镶脉病毒 (SVBV) 启动子 (NCBI 编号 AF331666REGION:623..950 版本 AF331666.1GI:13345788) 并由土壤杆菌属章鱼碱 (octopin) 合酶 (OCS) 终止子 (NCBI 编号 Z37515REGION:1344..1538 版本 Z37515.1GI:886843) 终止的 LH3 的合成基因 (SEQ ID NO:22) 在携带 Co11 和 P4H β 表达盒的 pBINPLUS 载体的多克隆位点中进行克隆。

[0170] 在图 2 中举例说明了利用图 1 中描述的表达盒进入宿主植物的共转化方案。每个表达盒插入片段由编码序列的短名表示。编码序列和相关的 SEQ ID NO 描述于表 1 中。每次共转化由 2 个 pBINPLUS 二元载体来进行。每个长方形表示携带 1、2 或 3 个表达盒的单个 pBINPLUS 载体。启动子和终止子在图 1 中详细说明。

[0171] 实施例 2

[0172] 植物胶原表达

[0173] 下表 1 中列出的编码蛋白质的合成的多核苷酸序列对于在烟草植物中的表达进行设计和最优化。

[0174] 表 1- 表达的蛋白质的列表

[0175]

名称:	SwissProt 编号	氨基酸	剪接同工型	缺失	名称	包括在 SEQ ID NO 中	由 SEQ ID NO 编码
胶原 α 1(I) 链 [前体]	p02452	1442	一种形式	ER 信号	Co11	3	1
胶原 α 2(I) 链 [前体]	p08123 在 p08123 中进行 2 种改变: D549A 和 N249I	1342	一种形式	ER 信号	Co12	6	4
脯氨酰 4-羟化酶 β 亚基	p07237	487	一种形式	ER 信号, KDEL	P4H β 人	13	12

脯氨酰 4-羟化酶 $\alpha 1$ 亚基	p13674	517	P13674-1	ER 信号	P4H α 人	15	14
脯氨酰 4-羟化酶植物	没有记入 SwissProt。NCBI 编号 :gi : 15227885	252	一种形式	线粒体信号 预期为 :氨基酸 1-39	P4H 植物	17	16
前胶原 C-蛋白酶	p13497	866	P13497-1 BMP1-3	ER 信号, 前肽	蛋白酶 C	19	18
前胶原 I N-蛋白酶	o95450	958	O95450-1 LpNPI	ER 信号, 前肽	蛋白酶 N	21	20
赖氨酰羟化酶 3	o60568	714	一种形式	ER 信号	LH3	23	22

[0176] 信号肽

[0177] (i) 用于巯基蛋白酶 aleurain 前体 (NCBI 编号 P05167GI :113603) 的大麦基因的液泡信号序列

[0178] MAHARVLLLALAVLATAAVAVASSSSSFADSNPIRPVTDRAASTLA (SEQ IDNO :24)。

[0179] (ii) 鼠耳芥 (*Arabidopsis thaliana*) 内切 -1,4- β -葡聚糖酶的质外体信号 (Cell, NCBI 编号 CAA67156.1GI :2440033) ;SEQ ID NO.9, 由 SEQ ID NO.7 编码。

[0180] 质粒的构建

[0181] 如实施例 1 所教导的构建植物表达载体, 每种被构建的表达载体的组成经由限制酶切分析和测序进行确认。

[0182] 构建包括下列表达盒的表达载体 :

[0183] 1. 胶原 $\alpha 1$

[0184] 2. 胶原 $\alpha 1$ + 人 P4H β 亚基

[0185] 3. 胶原 $\alpha 1$ + 人 P4H β 亚基 + 人 LH3

[0186] 4. 胶原 $\alpha 2$

[0187] 5. 胶原 $\alpha 2$ + 人 P4H α 亚基

[0188] 6. 胶原 $\alpha 2$ + 鼠耳芥 P4H

[0189] 7. 人 P4H β 亚基 + 人 LH3

[0190] 8. 人 P4H α 亚基

[0191] 上述编码序列中的每一种都与液泡转运肽或质外体转运肽翻译地融合, 或在其中预期细胞质积聚的情况下, 缺乏任何转运肽序列。

[0192] 植物转化和 PCR 筛选

[0193] 烟草植物 (烟草 (*Nicotiana tabacum*), Samsun NN) 根据图 2 中教导的转化方案由上述表达载体进行转化。

[0194] 所得的转基因植物经由多重 PCR 利用 4 种引物进行筛选, 所述引物被设计为能够扩增胶原 $\alpha 1$ 的 324bp 片段和胶原 $\alpha 2$ 的 537bp 片段 (表 2)。图 3 举例说明了多重 PCR 筛选的结果。

[0195] 表 2- 用于扩增胶原 $\alpha 1$ 的 324bp 片段和胶原 $\alpha 2$ 的 537bp 片段的多重 PCR 的引物列表

[0196]

Co11 正向引物 (24-聚体):	5' ATCACCAGGAGAACAGGGACCATC 3'	SEQ ID 25
Co11 反向引物 (29-聚体):	5' TCCACTTCCAAATCTCTATCCCTAACAAAC 3'	SEQ ID 26
Co12 正向引物 (23-聚体):	5' AGGCATTAGAGGCGATAAGGGAG 3'	SEQ ID 27
Co12 反向引物 (27-聚体):	5' TCAATCCAATAATAGCCACTTGACCAC 3'	SEQ ID 28

[0197] 实施例 3

[0198] 在转基因烟草植物中的人胶原的检测

[0199] 通过在 0.5ml 50mM Tris-HCl pH = 7.5 中与“完整的”蛋白酶抑制剂混合物 (来自 Roche Diagnostics GmbH 的产品 #1836145, 1 片剂 /50ml 缓冲液) 一起磨碎 500mg 叶子, 从烟草转化体 2、3 和 4 中提取总可溶性蛋白。粗提取物与 250 μ l 包含 10% β -巯基乙醇和 8% SDS 的 4 \times 样品应用缓冲液混合, 将样品煮沸 7 分钟并在 13000rpm 下离心 8 分钟。将 20 μ l 上清液加载到 10% 聚丙烯酰胺凝胶中并在标准蛋白质印迹程序中用抗-胶原 I (变性的) 抗体 (来自 Chemicon Inc. 的 #AB745) 进行测试 (图 4)。W. T. 是野生型烟草。在对于 I 型胶原 $\alpha 1$ 或 $\alpha 2$ 或二者 PCR 阳性的植物中可见阳性胶原条带。来自人胎盘的 500ng I 型胶原 (来自 Chemicon Inc. 的 #CC050) 的阳性对照条带代表来自转基因植物的样品中的约 0.3% 总可溶性蛋白 (约 150 μ g)。

[0200] 当胶原靶向液泡时, 检测到表达预期分子量的、最高达 ~1% 总可溶性蛋白的胶原的植物 (图 4)。成功地实现了全长胶原亚细胞靶向质外体 (图 5)。在细胞质中表达胶原 (即没有靶向肽) 的植物没有将胶原积聚至可检测的水平, 从而显示在植物中的胶原亚细胞靶向作用对于成功是关键。

[0201] 此外, 与显示对缺少 N-前肽的胶原实施重要的蛋白质水解作用的 Ruggiero 等人 2000 和 Merle 等人 2002 研究相反, 使用本方法, 含有 C-前肽和 N-前肽的全长胶原蛋白质以高水平积聚在亚细胞区室中。

[0202] 现在的数据还清晰地显示各自表达不同胶原链类型的 2 种植物杂交是有利的, 因为它使得能够选择表达最佳水平的每种链类型的植物和后续的植物杂交, 以获得所需生产胶原的植物。

[0203] 由本发明的植物生产的胶原包括天然的前肽, 并因此预期形成比通过蛋白质水解纯化的人对照大的蛋白质。计算出的不含羟基化或糖基化的胶原 $\alpha 1$ 和 $\alpha 2$ 链分子量如

下:含有前肽的 Co11-136kDa, 不含前肽的 Co11-95kDa、含有前肽的 Co12-127kDa、不含前肽的 Co12-92kDa。

[0204] 如图 4 中可见的, 转化体 3-5 和 3-49 中的 Co11 条带看起来比其他植物中的 Co11 条带大。这提示胶原链中通过由 α 和 β 亚基组成的人脯氨酸 4-羟化酶全酶进行脯氨酸羟基化, 所述 α 和 β 亚基在这些植物中共表达并靶向与人胶原链相同的亚细胞区室(例如液泡)。

[0205] 实施例 4

[0206] 在转基因植物中的胶原三股螺旋装配和热稳定性

[0207] 通过转基因植物的总粗蛋白提取物的热变性、随后为胰蛋白酶或胃蛋白酶消化, 来测试在转基因植物中的胶原三股螺旋的装配和螺旋的热稳定性(图 6a-b)。

[0208] 在第一个实验中, 来自烟草 2-9(只表达胶原 $\alpha 1$, 且不表达 P4H) 和 3-5(表达胶原 $\alpha 1+2$ 和 P4H) 的总可溶性蛋白通过下列操作提取: 在 0.5ml 50 mM Tris-HCl pH = 7.5 中磨碎 500mg 叶子, 在 13000rpm 下离心 10 分钟, 并收集上清液。对 50 μ l 上清液实施热处理(于 33°C 或 43°C 15 分钟), 并随后立即置于冰上。通过向每种样品加入 6 μ l 溶于 50mM Tris-HCl pH = 7.5 的 1mg/ml 胰蛋白酶开始胰蛋白酶消化。样品在室温(约 22°C)下温育 20 分钟。通过加入 20 μ l 包含 10% β 巯基乙醇和 8% SDS 的 4 \times 样品应用缓冲液来终止消化, 将样品煮沸 7 分钟并在 13000rpm 下离心 7 分钟。将 50 μ l 上清液加载到 10% 聚丙烯酰胺凝胶上并利用标准蛋白质印迹程序经由抗-胶原 I 抗体(来自 Chemicon Inc. 的 #AB745)进行测试。阳性对照是 ~500ng 人胶原 I 样品(来自 Chemicon Inc. 的 #CC050, 通过胃蛋白酶消化从人胎盘中提取), 将其加入从野生型烟草中提取的 50 μ l 总可溶性蛋白中。

[0209] 如图 6a 中所示, 在植物 #3-5 中形成的胶原三股螺旋以及对照人胶原在 33°C 对变性有抗性。相反, 由植物 #2-9 形成的胶原在 33°C 变性。这种热稳定性方面的差异提示在转化体 #3-5 中的成功的三股螺旋装配和翻译后的脯氨酸羟基化, 所述转化体 #3-5 表达胶原 $\alpha 1$ 和胶原 $\alpha 2$ 以及 P4H β 和 α 亚基。

[0210] 转化体 #2-9 中的 2 个条带可以表示二聚体或三聚体, 它们在与 SDS 和巯基乙醇一起煮沸 7 分钟后是稳定的。类似的条带在人胶原(上面的图)和转化体 #3-5 中可见。可能的解释是不同三股螺旋中的 2 种肽之间的共价键(交联), 所述共价键在 2 个赖氨酸通过赖氨酸氧化酶氧化脱氨后形成。

[0211] 在第二个实验中, 来自转基因烟草 13-6(表达胶原 I $\alpha 1$ 和 $\alpha 2$ 链 - 由箭标指出, 人 P4H α 和 β 亚基以及人 LH3) 的总可溶性蛋白通过下列操作提取: 在 0.5ml 100 mM Tris-HCl pH = 7.5 和 300mM NaCl 中磨碎 500mg 叶子, 在 10000rpm 下离心 7 分钟, 并收集上清液。对 50 μ l 上清液实施热处理(于 33°C、38°C 或 42°C 20 分钟), 并随后立即置于冰上。通过向每种样品加入溶于 10mM 乙酸的 4.5 μ l 0.1M HCl 和 4 μ l 2.5mg/ml 胃蛋白酶开始胃蛋白酶消化。样品在室温(约 22°C)下温育 30 分钟。通过加入 5 μ l 无缓冲的 1M Tris 来终止消化。每种样品与 22 μ l 包含 10% β 巯基乙醇和 8% SDS 的 4 \times 样品应用缓冲液混合, 煮沸 7 分钟并在 13000rpm 下离心 7 分钟。将 40 μ l 上清液加载到 10% 聚丙烯酰胺凝胶上并在标准蛋白质印迹程序中用抗-胶原 I 抗体(来自 Chemicon Inc. 的 #AB745)进行测试。阳性对照是加入从野生型烟草中提取的总可溶性蛋白中的 ~50ng 人胶原 I(来

自 Chemicon Inc. 的 #CC050, 通过胃蛋白酶消化从人胎盘中提取)。

[0212] 如图 6b 中举例说明的, 在植物 #13-6 中形成的胶原三股螺旋在 42°C 对变性有抗性。前肽的切割首先在 33°C 可见, 并且当温度升至 38°C 并再升至 42°C 时效力逐渐增加。被切割的胶原三股螺旋结构域在凝胶上显示出的迁移与胃蛋白酶处理的人胶原的迁移相似。在这个实验中使用的人胶原通过胃蛋白酶的蛋白水解作用从人胎盘中提取出来, 并且因此缺少前肽和某些端肽。

[0213] 实施例 5

[0214] 植物 P4H 表达

[0215] 天然的植物 P4H 的诱导

[0216] 烟草 P4H cDNA 被克隆并用作探针以测定将诱导内源性 P4H 表达的条件和处理。RNA 印迹分析 (图 7) 清晰地显示, P4H 以相对高的水平表达于苗端并以低水平表达于叶中。P4H 水平在磨损处理 (在下面的图中的“创伤”) 后 4 小时在叶中被显著诱导。利用其他应激条件达到类似结果 (未显示)。

[0217] 转基因烟草植物中入 P4H α 和 β 亚基以及胶原 $\alpha 1$ 和 $\alpha 2$ 链的检测

[0218] 使用抗人 P4H α 亚基抗体 (来自 ICN Biomedicals Inc. 的 #63-163)、抗人 P4H β 亚基抗体 (来自 Chemicon Inc. 的 #MAB2701) 和抗胶原 I 抗体 (来自 Chemicon Inc. 的 #AB745) 实现转基因烟草植物中人 P4H α 和 β 亚基以及 I 型胶原 $\alpha 1$ 和 $\alpha 2$ 链的检测。用这些抗体探测的蛋白质印迹结果显示于图 8 中。

[0219] 在植物 13-6 (还转化有人 LH3) 中确认 P4H α 、P4H β 和胶原 I $\alpha 1$ 和 $\alpha 2$ 条带的表达。计算出的包括液泡信号肽的 P4H α 和 β 的分子量分别为 65.5kDa 和 53.4kDa。计算出的含有前肽、不含羟基化或糖基化的胶原 $\alpha 1$ 和 $\alpha 2$ 链的分子量分别为 136kDa 和 127kDa。

[0220] 应当理解本发明的某些特征还可与单个实施方案组合提供, 所述特征为了清晰起见在分开的实施方案的背景中进行描述。相反地, 本发明的各种特征还可分开或以任何适当的亚组合提供, 所述特征为了简短起见在单个实施方案的背景中进行描述。

[0221] 尽管本发明已结合其具体实施方案进行了描述, 但明显的是, 多种替代方案、修饰和改变对于本领域技术人员将是显而易见的。因此, 它预期包含落在附加权利要求的精神和广泛范围内的所有此类替代方案、修饰和改变。本说明书中提到的所有公开出版物、专利和专利申请以及 GenBank 编号都整体引入本说明书作为参考, 其程度相当于每个单独的公开出版物、专利或专利申请或 GenBank 编号被具体地和单独地说明引入本文作为参考。此外, 本申请中的任何参考文献的引用或鉴定不应被解释为承认此类参考文献可作为本发明的现有技术。

[0222] 参考文献

[0223] (其他参考文献在文件中引用)

[0224] 1. Bulleid NJ, John DC, Kadler KE. Recombinant expression systems for the production of collagen. *Biochem Soc Trans.* 2000 ;28(4) :350-3. Review. PMID : 10961917 [PubMed-indexed for MEDLINE]

[0225] 2. Hare PD, Cress WA. Metabolic implications of stress-induced proline accumulation in plants. *Plant Growth Regulation* 1997 ;21 :79-102.

[0226] 3. Hieta R, Myllyharju J. Cloning and characterization of a low

molecularweight prolyl 4-hydroxylase from Arabidopsis thaliana. Effective hydroxylation of proline-rich, collagen-like, and hypoxia-inducible transcription factor alpha-like peptides. J Biol Chem. 2002 Jun 28; 277(26): 23965-71. Epub 2002 Apr 25. PMID: 11976332 [PubMed-indexed for MEDLINE]

[0227] 4. Hulmes DJ. Building collagen molecules, fibrils, and suprafibrillar structures. J Struct Biol. 2002 Jan-Feb; 137(1-2): 2-10. Review. PMID: 12064927 [PubMed-indexed for MEDLINE]

[0228] 5. Inkinen K. Connective tissue formation in wound healing. An experimental study. Academic Dissertation, September 2003. University of Helsinki, Faculty of Science, Department of Biosciences, Division of Biochemistry (ISBN 952-10-1313-3) <http://ethesis.helsinki.fi/iulkaikut/mat/bioti/vk/inkinen/>

[0229] 6. Merle C, Perret S, Lacour T, Jonval V, Hudaverdian S, Garrone R, Ruggiero F, Theisen M. Hydroxylated human homotrimeric collagen I in Agrobacterium tumefaciens-mediated transient expression and in transgenic tobacco plant. FEBS Lett. 2002 Mar 27; 515(1-3): 114-8. PMID: 11943205 [PubMed-indexed for MEDLINE]

[0230] 7. Olsen D, Yang C, Bodo M, Chang R, Leigh S, Baez J, Carmichael D, Perala M, Hamalainen ER, Jarvinen M, Polarek J. Recombinant collagen and gelatin for drug delivery. Adv Drug Deliv Rev. 2003 Nov 28; 55(12): 1547-67. PMID: 14623401 [PubMed-in process]

[0231] 8. Ruggiero F, Exposito JY, Boumat P, Gruber V, Perret S, Comte J, Olganier B, Garrone R, Theisen M.

[0232] Triple helix assembly and processing of human collagen produced in transgenic tobacco plants. FEBS Lett. 2000 Mar 3; 469(1): 132-6. PMID: 10708770 [PubMed-indexed for MEDLINE]

[0233] 9. Tanaka M, Sato K, Uchida T. Plant prolyl hydroxylase recognizes poly(L-proline)II helix. J Biol Chem. 1981 Nov 25; 256(22): 11397-400. PMID: 6271746 [PubMed-indexed for MEDLINE]

[0234] 10. Wang C, Luosujarvi H, Heikkinen J, Risteli M, Uitto L, Myllyla R. The third activity for lysyl hydroxylase 3: galactosylation of hydroxylysyl residues in collagen *in vitro*. Matrix Biol. 2002 Nov; 21(7): 559-66. PMID: 12475640 [PubMed-indexed for MEDLINE]

序列表

<110>CollPlant Ltd.

<120> 生产胶原的植物及其生成和使用方法

<130>CPCH0761218P

<160>28

<170>PatentIn version 3.3

<210>1

<211>4662

<212>DNA

<213> 人工序列

<220>

<223> 含有与人胶原 $\alpha 1(I)$ 链融合的大麦巯基蛋白酶 aleurain 前体基因的维管信号序列

编码区和侧翼区的合成序列

<400>1

```

gcgatgcatg taatgtcatg agccacatga tccaatggcc acaggaacgt aagaatgtag      60
atagatttga ttttgtccgt tagatagcaa acaacattat aaaagggtgtg tatcaatacg      120
aactaattca ctcatgggat tcatagaagt ccattcctcc taagtatcta aaccatggct      180
cacgctcgtg ttctctcct cgcctcgcct gttttggcaa cagctgctgt ggcgtgtggt      240
tctagttctt cttttgctga ttcaaacct attagacctg ttactgatag agcagcttcc      300
actttggctc aattgcaaga ggaggggcag gttgagggcc aagatgagga tatecctcca      360
attacatgcg tgcaaaatgg cttgcggttac cacgataggg atgtgtggaa acctgaacct      420
tgtcgtatct gtgtgtgtga taacggcaag gtgctctgcg atgatgttat ctgcgatgag      480
acaaaaaatt gccctggcgc tgaagttcct gagggcgagt gttgccctgt gtgccctgat      540
ggttccgagt ccccaactga tcaggaaact actggcgtgg agggcccaaa aggagatact      600
ggtccacgtg gtccctagggg tccagcaggt cctccaggta gagatggtat tccaggccag      660
cctggattgc caggaccacc aggccaccct ggcccaccag gacctcctgg tcttgggtgga      720
aatttcgctc cacaactctc ttatggctat gatgagaagt caacagggtg tatttccggt      780
ccaggctcta tgggaccatc cggaccaaga ggtctcccag gtcctccagg tgcctctgga      840
cctcaaggct ttcaaggacc tccaggcgaa ccaggagAAC caggcgttc tggaccaatg      900
ggcccaaggg gaccacctgg cccaccagga aaaaatggcg atgatggcga agctggaaag      960

```

cctggtcgtc	ctggagagag	aggteectct	ggccccacagg	gtgcaagagg	cttgccagga	1020
actgctggct	tgcctggaat	gaagggacat	aggggcttct	ccggcctcga	tggcgctaag	1080
ggtgatgctg	gccctgctgg	accaaagggc	gagccagggt	cccctggaga	aaacggtgct	1140
cctggacaaa	tgggtectcg	tggacttcca	ggagaaaggg	gtcgtccagg	cgctccagga	1200
ccagcaggtg	ctaggggaaa	cgatggtgca	acaggcgctg	ctggccctcc	tggcccaact	1260
ggtcctgctg	gccctccagg	attcccaggc	gcagttggag	ctaaaggaga	agcaggacca	1320
cagggcccta	ggggttctga	aggacctcag	ggtgttagag	gtgaaccagg	tcctccaggc	1380
ccagctggag	cagctggtec	agcaggaaat	ccaggtgctg	atggtcaacc	tggagctaag	1440
ggcgctaata	gcgaccagg	tatcgaggc	gcaccagggt	ttcctggcgc	tagaggccca	1500
agtggtcctc	aaggaccagg	tggaccacca	ggtccaaaag	gcaattctgg	cgaacctggc	1560
gctccagggt	ctaaaggaga	tactggtgct	aaaggcgaac	caggacctgt	tgggttctag	1620
ggtcctcctg	gtcctgctgg	agaagaagga	aaaagagggt	ctcgtggaga	accaggacca	1680
actggacttc	ctggacctcc	tggatgaacgt	ggcggacctg	gctcaagggg	tttccctgga	1740
gctgatggag	tggcaggtec	aaaaggccct	gctggagaga	gaggttcacc	aggtccagct	1800
ggtcctaagg	gctcccctgg	tgaagcaggt	agaccaggcg	aagcaggatt	gccaggcgca	1860
aagggattga	caggctctcc	tggtagtctc	ggcccagatg	gaaaaacagg	cccaccaggt	1920
ccagcaggac	aagatggacg	tccaggccca	ccaggtcttc	ctggagcaag	gggacaagct	1980
ggcgttatgg	gttttccagg	acctaaaggt	gctgctggag	agccaggaaa	ggcaggtgaa	2040
agaggagttc	ctggtcacc	aggagcagtg	ggtcctgctg	gcaaagatgg	tgaagctgga	2100
gcacagggcc	ctccaggccc	tgctggccca	gctggcgaac	gtggagaaca	aggcccagct	2160
ggtagtccag	gatttcaagg	attgcctggc	cctgctggcc	ctccaggaga	agcaggaaaa	2220
cctggagaac	aaggagttec	tggtgatttg	ggagcacctg	gaccttcagg	agcacgtggt	2280
gaaagaggct	tccctggcga	gaggggtggt	caaggtccac	caggtccagc	aggacctaga	2340
ggtgctaata	gcgctectgg	caacgatgga	gcaaaaggtg	atgctggtgc	tcctggcgca	2400
cctggaagtc	agggtgctcc	tggattgcaa	ggaatgcctg	gagagagggg	tgctgctggc	2460
ttgccaggcc	caaagggcga	taggggtgat	gctggaccaa	aaggtgctga	tggatcccca	2520
ggaaaagatg	gagttcgtgg	tcttactggc	ccaatcggac	ctccaggccc	tgctggcgct	2580
ccaggtgata	agggcgaaag	tggcccaagt	ggacctgctg	gacctactgg	tgctagaggt	2640
gcacctggtg	ataggggtga	acctggacca	cctggtccag	ctggttttgc	tggctctcct	2700
ggagctgatg	gacaacctgg	cgcaaagggg	gaaccagggt	atgctggcgc	aaagggagat	2760
gctggtccac	ctggacctgc	tggtcagca	ggccccctg	ggccaatcgg	taatgttgga	2820
gcaccagggt	ctaagggagc	taggggttcc	gctggtccac	ctggagcaac	aggatttcca	2880
ggcgctgctg	gtagagttag	cccaccaggc	ccatccggaa	acgcaggccc	tcctggtcct	2940
ccaggtcctg	ctggcaagga	gggtggcaaa	ggaccaaggg	gcgaaactgg	ccctgctggt	3000
agacctggcg	aagttggccc	tcctggacca	ccaggtccag	caggagaaaa	aggttcccca	3060
ggagctgatg	gcccagctgg	tgctccagga	actccaggcc	ctcaaggtat	tgctggacag	3120
agaggcgttg	tgggactccc	tggatcaaagg	ggagagagag	gatttccagg	cttgccagga	3180
cctagtggag	aacctggaaa	acaaggccca	tcaggcgcta	gtggagagcg	tggacctcct	3240
ggccctatgg	gacctcctgg	attggctggc	ccacctggcg	aatcaggctg	tgaagcgca	3300

ccaggcgcag aaggatcacc tggaagagat ggatcccctg gtgctaaagg cgatcgtgga	3360
gaaactggtc cagcaggccc accaggcgca ccagggtcac ctggcgctcc aggacctgtg	3420
ggaccagctg gaaaatccgg agataggggc gagacaggcc cagcaggacc agctggacct	3480
gttggccctg ctggcgctcg tggaccagca ggacctcaag gaccaagggg agataagga	3540
gaaacaggcg aacaaggcga taggggcatt aagggtcata ggggttttag tggcctccag	3600
ggtcctcctg gcccacctgg atcaccagga gaacaggac catctggtgc ttccggccca	3660
gctggtccaa gaggacctcc aggatcagct ggtgcacctg gaaaagatgg tcttaacggt	3720
ctcccaggac caatcgccc tccaggacct agaggaagaa caggagatgc tggccctggt	3780
ggccctccag gacctcctgg tccaccagggt ccacctggtc ctccatcagc tggattcgat	3840
ttttcatttc ttccacagcc accacaagag aaagctcacg atggcggcag atattaccgt	3900
gctgatgatg ctaacgttgt tagggataga gatttgggaag tggatacaac tttgaaatcc	3960
ctctcccagc aaattgaaaa cattagatct ccagaagggt cacgtaaaaa cccagctaga	4020
acatgtcgtg atttgaaaat gtgtcactcc gattggaaaa gtggtgaata ctggattgat	4080
ccaaatcagg gctgtaatct cgatgctatc aaagttttct gtaacatgga aacaggcgaa	4140
acatgcgttt atcctactca accttccgtg gctcagaaaa attggtacat ctcaaaaaat	4200
cctaaagata agaggcacgt ttggttcggt gaaagtatga ctgatggatt tcaatttgag	4260
tacggcggtc aaggtagtga tccagctgat gtggctattc aactcacatt tttgctctt	4320
atgtccacag aggcacaca aaacatcact taccactgca aaaacagtgt ggcttatatg	4380
gatcaacaaa caggaaacct taagaaggct cttcttttga agggctcaaa cgagattgag	4440
attagagcag agggcaactc aaggtttact tattcagtta ctgttgatgg ctgcacttca	4500
catactggcg cttggggtaa aacagttatc gagtataaga ctacaaaaac atcaagactc	4560
ccaatcattg atgttgctcc tctcgatggt ggcgctcctg atcaagagtt cggttttgat	4620
gtgggcccag tttgtttcct ctaatgagct cgcggccgca tc	4662

<210>2

<211>4662

<212>DNA

<213> 人工序列

<220>

<223> 与人胶原 $\alpha 1(I)$ 链融合的大麦巯基蛋白酶 aleurain 前体基因的维管信号序列和侧

翼区的合成序列

<220>

<221>CDS

<222>(175)..(4644)

<400>2

gcgatgcatg taatgtcatg agccacatga tccaatggcc acaggaacgt aagaatgtag	60
atagatttga tttgtccgt tagatagcaa acaacattat aaaaggtgtg tatcaatagc	120
aactaattca ctcatggat tcatagaagt ccattcctcc taagtatcta aacc atg	177
	Met
	1
gct cac gct cgt gtt ctc ctc ctc gct ctc gct gtt ttg gca aca gct	225
Ala His Ala Arg Val Leu Leu Leu Ala Leu Ala Val Leu Ala Thr Ala	
5 10 15	
gct gtg gct gtg gct tct agt tct tct ttt gct gat tca aac cct att	273
Ala Val Ala Val Ala Ser Ser Ser Ser Phe Ala Asp Ser Asn Pro Ile	
20 25 30	
aga cct gtt act gat aga gca gct tcc act ttg gct caa ttg caa gag	321
Arg Pro Val Thr Asp Arg Ala Ala Ser Thr Leu Ala Gln Leu Gln Glu	
35 40 45	
gag ggc cag gtt gag ggc caa gat gag gat atc cct cca att aca tgc	369
Glu Gly Gln Val Glu Gly Gln Asp Glu Asp Ile Pro Pro Ile Thr Cys	
50 55 60 65	
gtg caa aat ggc ttg cgt tac cac gat agg gat gtg tgg aaa cct gaa	417
Val Gln Asn Gly Leu Arg Tyr His Asp Arg Asp Val Trp Lys Pro Glu	
70 75 80	
cct tgt cgt atc tgt gtg tgt gat aac ggc aag gtg ctc tgc gat gat	465
Pro Cys Arg Ile Cys Val Cys Asp Asn Gly Lys Val Leu Cys Asp Asp	
85 90 95	
gtt atc tgc gat gag aca aaa aat tgc cct ggc gct gaa gtt cct gag	513
Val Ile Cys Asp Glu Thr Lys Asn Cys Pro Gly Ala Glu Val Pro Glu	
100 105 110	
ggc gag tgt tgc cct gtg tgc cct gat ggt tcc gag tcc cca act gat	561
Gly Glu Cys Cys Pro Val Cys Pro Asp Gly Ser Glu Ser Pro Thr Asp	
115 120 125	
cag gaa act act ggc gtg gag ggc cca aaa gga gat act ggt cca cgt	609
Gln Glu Thr Thr Gly Val Glu Gly Pro Lys Gly Asp Thr Gly Pro Arg	
130 135 140 145	
ggt cct agg ggt cca gca ggt cct cca ggt aga gat ggt att cca ggc	657
Gly Pro Arg Gly Pro Ala Gly Pro Pro Gly Arg Asp Gly Ile Pro Gly	
150 155 160	
cag cct gga ttg cca gga cca cca ggc cca cct ggc cca cca gga cct	705
Gln Pro Gly Leu Pro Gly Pro Pro Gly Pro Pro Gly Pro Pro Gly Pro	
165 170 175	
cct ggt ctt ggt gga aat ttc gct cca caa ctc tct tat ggc tat gat	753

Pro Gly Leu Gly Gly Asn Phe Ala Pro Gln Leu Ser Tyr Gly Tyr Asp	
180	185
gag aag tca aca ggt ggt att tcc gtt cca ggt cct atg gga cca tcc	801
Glu Lys Ser Thr Gly Gly Ile Ser Val Pro Gly Pro Met Gly Pro Ser	
195	200
gga cca aga ggt ctc cca ggt cct cca ggt gct cct gga cct caa ggc	849
Gly Pro Arg Gly Leu Pro Gly Pro Pro Gly Ala Pro Gly Pro Gln Gly	
210	215
ttt caa gga cct cca ggc gaa cca gga gaa cca ggc gct tct gga cca	897
Phe Gln Gly Pro Pro Gly Glu Pro Gly Glu Pro Gly Ala Ser Gly Pro	
230	235
atg ggc cca agg gga cca cct ggc cca cca gga aaa aat ggc gat gat	945
Met Gly Pro Arg Gly Pro Pro Gly Pro Pro Gly Lys Asn Gly Asp Asp	
245	250
ggc gaa gct gga aag cct ggt cgt cct gga gag aga ggt cct cct ggc	993
Gly Glu Ala Gly Lys Pro Gly Arg Pro Gly Glu Arg Gly Pro Pro Gly	
260	265
cca cag ggt gca aga ggc ttg cca gga act gct ggc ttg cct gga atg	1041
Pro Gln Gly Ala Arg Gly Leu Pro Gly Thr Ala Gly Leu Pro Gly Met	
275	280
aag gga cat agg ggc ttc tcc ggc ctc gat ggc gct aag ggt gat gct	1089
Lys Gly His Arg Gly Phe Ser Gly Leu Asp Gly Ala Lys Gly Asp Ala	
290	295
ggc cct gct gga cca aag ggc gag cca ggt tcc cct gga gaa aac ggt	1137
Gly Pro Ala Gly Pro Lys Gly Glu Pro Gly Ser Pro Gly Glu Asn Gly	
310	315
gct cct gga caa atg ggt cct cgt gga ctt cca gga gaa agg ggt cgt	1185
Ala Pro Gly Gln Met Gly Pro Arg Gly Leu Pro Gly Glu Arg Gly Arg	
325	330
cca ggc gct cca gga cca gca ggt gct agg gga aac gat ggt gca aca	1233
Pro Gly Ala Pro Gly Pro Ala Gly Ala Arg Gly Asn Asp Gly Ala Thr	
340	345
ggc gct gct ggc cct cct ggc cca act ggt cct gct ggc cct cca gga	1281
Gly Ala Ala Gly Pro Pro Gly Pro Thr Gly Pro Ala Gly Pro Pro Gly	
355	360
ttc cca ggc gca gtt gga gct aaa gga gaa gca gga cca cag ggc cct	1329
Phe Pro Gly Ala Val Gly Ala Lys Gly Glu Ala Gly Pro Gln Gly Pro	
370	375
agg ggt tct gaa gga cct cag ggt gtt aga ggt gaa cca ggt cct cca	1377

Arg Gly Ser Glu Gly Pro Gln Gly Val Arg Gly Glu Pro Gly Pro Pro	
390	395 400
ggc cca gct gga gca gct ggt cca gca gga aat cca ggt gct gat ggt	1425
Gly Pro Ala Gly Ala Ala Gly Pro Ala Gly Asn Pro Gly Ala Asp Gly	
405	410 415
caa cct gga gct aag ggc gct aat ggc gca cca ggt atc gca ggc gca	1473
Gln Pro Gly Ala Lys Gly Ala Asn Gly Ala Pro Gly Ile Ala Gly Ala	
420	425 430
cca ggt ttt cct ggc gct aga ggc cca agt ggt cct caa gga cca ggt	1521
Pro Gly Phe Pro Gly Ala Arg Gly Pro Ser Gly Pro Gln Gly Pro Gly	
435	440 445
gga cca cca ggt cca aaa ggc aat tct ggc gaa cct ggc gct cca ggt	1569
Gly Pro Pro Gly Pro Lys Gly Asn Ser Gly Glu Pro Gly Ala Pro Gly	
450	455 460 465
tct aaa gga gat act ggt gct aaa ggc gaa cca gga cct gtt ggt gtt	1617
Ser Lys Gly Asp Thr Gly Ala Lys Gly Glu Pro Gly Pro Val Gly Val	
470	475 480
cag ggt cct cct ggt cct gct gga gaa gaa gga aaa aga ggt gct cgt	1665
Gln Gly Pro Pro Gly Pro Ala Gly Glu Glu Gly Lys Arg Gly Ala Arg	
485	490 495
gga gaa cca gga cca act gga ctt cct gga cct cct ggt gaa cgt ggc	1713
Gly Glu Pro Gly Pro Thr Gly Leu Pro Gly Pro Pro Gly Glu Arg Gly	
500	505 510
gga cct ggc tca agg ggt ttc cct gga gct gat gga gtg gca ggt cca	1761
Gly Pro Gly Ser Arg Gly Phe Pro Gly Ala Asp Gly Val Ala Gly Pro	
515	520 525
aaa ggc cct gct gga gag aga ggt tca cca ggt cca gct ggt cct aag	1809
Lys Gly Pro Ala Gly Glu Arg Gly Ser Pro Gly Pro Ala Gly Pro Lys	
530	535 540 545
ggc tcc cct ggt gaa gca ggt aga cca ggc gaa gca gga ttg cca ggc	1857
Gly Ser Pro Gly Glu Ala Gly Arg Pro Gly Glu Ala Gly Leu Pro Gly	
550	555 560
gca aag gga ttg aca ggc tct cct ggt agt cct ggc cca gat gga aaa	1905
Ala Lys Gly Leu Thr Gly Ser Pro Gly Ser Pro Gly Pro Asp Gly Lys	
565	570 575
aca ggc cca cca ggt cca gca gga caa gat gga cgt cca ggc cca cca	1953
Thr Gly Pro Pro Gly Pro Ala Gly Gln Asp Gly Arg Pro Gly Pro Pro	
580	585 590
ggt cct cct gga gca agg gga caa gct ggc gtt atg ggt ttt cca gga	2001

Gly Pro Pro Gly Ala Arg Gly Gln Ala Gly Val Met Gly Phe Pro Gly	
595	600
cct aaa ggt gct gct gga gag cca gga aag gca ggt gaa aga gga gtt	2049
Pro Lys Gly Ala Ala Gly Glu Pro Gly Lys Ala Gly Glu Arg Gly Val	
610	615
cct ggt cca cca gga gca gtg ggt cct gct ggc aaa gat ggt gaa gct	2097
Pro Gly Pro Pro Gly Ala Val Gly Pro Ala Gly Lys Asp Gly Glu Ala	
630	635
gga gca cag ggc cct cca ggc cct gct ggc cca gct ggc gaa cgt gga	2145
Gly Ala Gln Gly Pro Pro Gly Pro Ala Gly Pro Ala Gly Glu Arg Gly	
645	650
gaa caa ggc cca gct ggt agt cca gga ttt caa gga ttg cct ggc cct	2193
Glu Gln Gly Pro Ala Gly Ser Pro Gly Phe Gln Gly Leu Pro Gly Pro	
660	665
gct ggc cct cca gga gaa gca gga aaa cct gga gaa caa gga gtt cct	2241
Ala Gly Pro Pro Gly Glu Ala Gly Lys Pro Gly Glu Gln Gly Val Pro	
675	680
ggt gat ttg gga gca cct gga cct tca gga gca cgt ggt gaa aga ggc	2289
Gly Asp Leu Gly Ala Pro Gly Pro Ser Gly Ala Arg Gly Glu Arg Gly	
690	695
ttc cct ggc gag agg ggt gtt caa ggt cca cca ggt cca gca gga cct	2337
Phe Pro Gly Glu Arg Gly Val Gln Gly Pro Pro Gly Pro Ala Gly Pro	
710	715
aga ggt gct aat ggc gct cct ggc aac gat gga gca aaa ggt gat gct	2385
Arg Gly Ala Asn Gly Ala Pro Gly Asn Asp Gly Ala Lys Gly Asp Ala	
725	730
ggt gct cct ggc gca cct gga agt cag ggt gct cct gga ttg caa gga	2433
Gly Ala Pro Gly Ala Pro Gly Ser Gln Gly Ala Pro Gly Leu Gln Gly	
740	745
atg cct gga gag agg ggt gct gct ggc ttg cca ggc cca aag ggc gat	2481
Met Pro Gly Glu Arg Gly Ala Ala Gly Leu Pro Gly Pro Lys Gly Asp	
755	760
agg ggt gat gct gga cca aaa ggt gct gat gga tcc cca gga aaa gat	2529
Arg Gly Asp Ala Gly Pro Lys Gly Ala Asp Gly Ser Pro Gly Lys Asp	
770	775
gga gtt cgt ggt ctt act ggc cca atc gga cct cca ggc cct gct ggc	2577
Gly Val Arg Gly Leu Thr Gly Pro Ile Gly Pro Pro Gly Pro Ala Gly	
790	795
gct cca ggt gat aag ggc gaa agt ggc cca agt gga cct gct gga cct	2625

Ala Pro Gly Asp Lys Gly Glu Ser Gly Pro Ser Gly Pro Ala Gly Pro	
805	810
act ggt gct aga ggt gca cct ggt gat agg ggt gaa cct gga cca cct	2673
Thr Gly Ala Arg Gly Ala Pro Gly Asp Arg Gly Glu Pro Gly Pro Pro	
820	825
ggt cca gct ggt ttt gct ggt cct cct gga gct gat gga caa cct ggc	2721
Gly Pro Ala Gly Phe Ala Gly Pro Pro Gly Ala Asp Gly Gln Pro Gly	
835	840
gca aag ggt gaa cca ggt gat gct ggc gca aag gga gat gct ggt cca	2769
Ala Lys Gly Glu Pro Gly Asp Ala Gly Ala Lys Gly Asp Ala Gly Pro	
850	855
cct gga cct gct ggt cca gca ggc ccc cct ggg cca atc ggt aat gtt	2817
Pro Gly Pro Ala Gly Pro Ala Gly Pro Pro Gly Pro Ile Gly Asn Val	
870	875
gga gca cca ggt gct aag gga gct agg ggt tcc gct ggt cca cct gga	2865
Gly Ala Pro Gly Ala Lys Gly Ala Arg Gly Ser Ala Gly Pro Pro Gly	
885	890
gca aca gga ttt cca ggc gct gct ggt aga gtt ggc cca cca ggc cca	2913
Ala Thr Gly Phe Pro Gly Ala Ala Gly Arg Val Gly Pro Pro Gly Pro	
900	905
tcc gga aac gca ggc cct cct ggt cct cca ggt cct gct ggc aag gag	2961
Ser Gly Asn Ala Gly Pro Pro Gly Pro Pro Gly Pro Ala Gly Lys Glu	
915	920
ggt ggc aaa gga cca agg ggc gaa act ggc cct gct ggt aga cct ggc	3009
Gly Gly Lys Gly Pro Arg Gly Glu Thr Gly Pro Ala Gly Arg Pro Gly	
930	935
gaa gtt ggc cct cct gga cca cca ggt cca gca gga gaa aaa ggt tcc	3057
Glu Val Gly Pro Pro Gly Pro Pro Gly Pro Ala Gly Glu Lys Gly Ser	
950	955
cca gga gct gat ggc cca gct ggt gct cca gga act cca ggc cct caa	3105
Pro Gly Ala Asp Gly Pro Ala Gly Ala Pro Gly Thr Pro Gly Pro Gln	
965	970
ggt att gct gga cag aga ggc gtt gtg gga ctc cct ggt caa agg gga	3153
Gly Ile Ala Gly Gln Arg Gly Val Val Gly Leu Pro Gly Gln Arg Gly	
980	985
gag aga gga ttt cca ggc ttg cca gga cct agt gga gaa cct gga aaa	3201
Glu Arg Gly Phe Pro Gly Leu Pro Gly Pro Ser Gly Glu Pro Gly Lys	
995	1000
caa ggc cca tca ggc gct agt gga gag cgt gga cct cct ggc cct	3246

Gln Gly Pro Ser Gly Ala Ser Gly Glu Arg Gly Pro Pro Gly Pro		
1010	1015	1020
atg gga cct cct gga ttg gct ggc cca cct ggc gaa tca ggt cgt		3291
Met Gly Pro Pro Gly Leu Ala Gly Pro Pro Gly Glu Ser Gly Arg		
1025	1030	1035
gaa ggc gca cca ggc gca gaa gga tca cct gga aga gat gga tcc		3336
Glu Gly Ala Pro Gly Ala Glu Gly Ser Pro Gly Arg Asp Gly Ser		
1040	1045	1050
cct ggt gct aaa ggc gat cgt gga gaa act ggt cca gca ggc cca		3381
Pro Gly Ala Lys Gly Asp Arg Gly Glu Thr Gly Pro Ala Gly Pro		
1055	1060	1065
cca ggc gca cca ggt gca cct ggc gct cca gga cct gtg gga cca		3426
Pro Gly Ala Pro Gly Ala Pro Gly Ala Pro Gly Pro Val Gly Pro		
1070	1075	1080
gct gga aaa tcc gga gat agg ggc gag aca ggc cca gca gga cca		3471
Ala Gly Lys Ser Gly Asp Arg Gly Glu Thr Gly Pro Ala Gly Pro		
1085	1090	1095
gct gga cct gtt ggc cct gct ggc gct cgt gga cca gca gga cct		3516
Ala Gly Pro Val Gly Pro Ala Gly Ala Arg Gly Pro Ala Gly Pro		
1100	1105	1110
caa gga cca agg gga gat aag gga gaa aca ggc gaa caa ggc gat		3561
Gln Gly Pro Arg Gly Asp Lys Gly Glu Thr Gly Glu Gln Gly Asp		
1115	1120	1125
agg ggc att aag ggt cat agg ggt ttt agt ggc ctc cag ggt cct		3606
Arg Gly Ile Lys Gly His Arg Gly Phe Ser Gly Leu Gln Gly Pro		
1130	1135	1140
cct ggc cca cct gga tca cca gga gaa cag gga cca tct ggt gct		3651
Pro Gly Pro Pro Gly Ser Pro Gly Glu Gln Gly Pro Ser Gly Ala		
1145	1150	1155
tcc ggc cca gct ggt cca aga gga cct cca gga tca gct ggt gca		3696
Ser Gly Pro Ala Gly Pro Arg Gly Pro Pro Gly Ser Ala Gly Ala		
1160	1165	1170
cct gga aaa gat ggt ctt aac ggt ctc cca gga cca atc ggc cct		3741
Pro Gly Lys Asp Gly Leu Asn Gly Leu Pro Gly Pro Ile Gly Pro		
1175	1180	1185
cca gga cct aga gga aga aca gga gat gct ggc cct gtt ggc cct		3786
Pro Gly Pro Arg Gly Arg Thr Gly Asp Ala Gly Pro Val Gly Pro		
1190	1195	1200
cca gga cct cct ggt cca cca ggt cca cct ggt cct cca tca gct		3831

Pro Gly Pro Pro Gly Pro Pro Gly Pro Pro Gly Pro Pro Ser Ala 1205	1210	1215	
gga ttc gat ttt tca ttt ctt cca cag cca cca caa gag aaa gct Gly Phe Asp Phe Ser Phe Leu Pro Gln Pro Pro Gln Glu Lys Ala 1220	1225	1230	3876
cac gat ggc ggc aga tat tac cgt gct gat gat gct aac gtt gtt His Asp Gly Gly Arg Tyr Tyr Arg Ala Asp Asp Ala Asn Val Val 1235	1240	1245	3921
agg gat aga gat ttg gaa gtg gat aca act ttg aaa tcc ctc tcc Arg Asp Arg Asp Leu Glu Val Asp Thr Thr Leu Lys Ser Leu Ser 1250	1255	1260	3966
cag caa att gaa aac att aga tct cca gaa ggt tca cgt aaa aac Gln Gln Ile Glu Asn Ile Arg Ser Pro Glu Gly Ser Arg Lys Asn 1265	1270	1275	4011
cca gct aga aca tgt cgt gat ttg aaa atg tgt cac tcc gat tgg Pro Ala Arg Thr Cys Arg Asp Leu Lys Met Cys His Ser Asp Trp 1280	1285	1290	4056
aaa agt ggt gaa tac tgg att gat cca aat cag ggc tgt aat ctc Lys Ser Gly Glu Tyr Trp Ile Asp Pro Asn Gln Gly Cys Asn Leu 1295	1300	1305	4101
gat gct atc aaa gtt ttc tgt aac atg gaa aca ggc gaa aca tgc Asp Ala Ile Lys Val Phe Cys Asn Met Glu Thr Gly Glu Thr Cys 1310	1315	1320	4146
gtt tat cct act caa cct tcc gtg gct cag aaa aat tgg tac atc Val Tyr Pro Thr Gln Pro Ser Val Ala Gln Lys Asn Trp Tyr Ile 1325	1330	1335	4191
tca aaa aat cct aaa gat aag agg cac gtt tgg ttc ggt gaa agt Ser Lys Asn Pro Lys Asp Lys Arg His Val Trp Phe Gly Glu Ser 1340	1345	1350	4236
atg act gat gga ttt caa ttt gag tac ggc ggt caa ggt agt gat Met Thr Asp Gly Phe Gln Phe Glu Tyr Gly Gly Gln Gly Ser Asp 1355	1360	1365	4281
cca gct gat gtg gct att caa ctc aca ttt ttg cgt ctt atg tcc Pro Ala Asp Val Ala Ile Gln Leu Thr Phe Leu Arg Leu Met Ser 1370	1375	1380	4326
aca gag gca tca caa aac atc act tac cac tgc aaa aac agt gtg Thr Glu Ala Ser Gln Asn Ile Thr Tyr His Cys Lys Asn Ser Val 1385	1390	1395	4371
gct tat atg gat caa caa aca gga aac ctt aag aag gct ctt ctt			4416

Ala Tyr Met Asp Gln Gln Thr Gly Asn Leu Lys Lys Ala Leu Leu
 1400 1405 1410
 ttg aag ggc tca aac gag att gag att aga gca gag ggc aac tca 4461
 Leu Lys Gly Ser Asn Glu Ile Glu Ile Arg Ala Glu Gly Asn Ser
 1415 1420 1425
 agg ttt act tat tca gtt act gtt gat ggc tgc act tca cat act 4506
 Arg Phe Thr Tyr Ser Val Thr Val Asp Gly Cys Thr Ser His Thr
 1430 1435 1440
 ggc gct tgg ggt aaa aca gtt atc gag tat aag act aca aaa aca 4551
 Gly Ala Trp Gly Lys Thr Val Ile Glu Tyr Lys Thr Thr Lys Thr
 1445 1450 1455
 tca aga ctc cca atc att gat gtt gct cct ctc gat gtt ggc gct 4596
 Ser Arg Leu Pro Ile Ile Asp Val Ala Pro Leu Asp Val Gly Ala
 1460 1465 1470
 cct gat caa gag ttc ggt ttt gat gtg ggc cca gtt tgt ttc ctc 4641
 Pro Asp Gln Glu Phe Gly Phe Asp Val Gly Pro Val Cys Phe Leu
 1475 1480 1485
 taa tgagctcgcg gccgcac 4662

<210>3

<211>1489

<212>PRT

<213>人工序列

<220>

<223> 与人胶原 $\alpha 1(I)$ 链融合的大麦巯基蛋白酶 aleurain 前体基因的维管信号序列
 和侧

翼区的合成序列

<400>3

Met Ala His Ala Arg Val Leu Leu Leu Ala Leu Ala Val Leu Ala Thr
 1 5 10 15
 Ala Ala Val Ala Val Ala Ser Ser Ser Ser Phe Ala Asp Ser Asn Pro
 20 25 30
 Ile Arg Pro Val Thr Asp Arg Ala Ala Ser Thr Leu Ala Gln Leu Gln
 35 40 45
 Glu Glu Gly Gln Val Glu Gly Gln Asp Glu Asp Ile Pro Pro Ile Thr
 50 55 60

370	375	380
Pro Arg Gly Ser Glu Gly Pro Gln Gly Val Arg Gly Glu Pro Gly Pro		
385	390	395
Pro Gly Pro Ala Gly Ala Ala Gly Pro Ala Gly Asn Pro Gly Ala Asp		
	405	410
Gly Gln Pro Gly Ala Lys Gly Ala Asn Gly Ala Pro Gly Ile Ala Gly		
	420	425
Ala Pro Gly Phe Pro Gly Ala Arg Gly Pro Ser Gly Pro Gln Gly Pro		
	435	440
Gly Gly Pro Pro Gly Pro Lys Gly Asn Ser Gly Glu Pro Gly Ala Pro		
	450	455
Gly Ser Lys Gly Asp Thr Gly Ala Lys Gly Glu Pro Gly Pro Val Gly		
465	470	475
Val Gln Gly Pro Pro Gly Pro Ala Gly Glu Glu Gly Lys Arg Gly Ala		
	485	490
Arg Gly Glu Pro Gly Pro Thr Gly Leu Pro Gly Pro Pro Gly Glu Arg		
	500	505
Gly Gly Pro Gly Ser Arg Gly Phe Pro Gly Ala Asp Gly Val Ala Gly		
	515	520
Pro Lys Gly Pro Ala Gly Glu Arg Gly Ser Pro Gly Pro Ala Gly Pro		
	530	535
Lys Gly Ser Pro Gly Glu Ala Gly Arg Pro Gly Glu Ala Gly Leu Pro		
545	550	555
Gly Ala Lys Gly Leu Thr Gly Ser Pro Gly Ser Pro Gly Pro Asp Gly		
	565	570
Lys Thr Gly Pro Pro Gly Pro Ala Gly Gln Asp Gly Arg Pro Gly Pro		
	580	585
Pro Gly Pro Pro Gly Ala Arg Gly Gln Ala Gly Val Met Gly Phe Pro		
	595	600
Gly Pro Lys Gly Ala Ala Gly Glu Pro Gly Lys Ala Gly Glu Arg Gly		
	610	615
Val Pro Gly Pro Pro Gly Ala Val Gly Pro Ala Gly Lys Asp Gly Glu		
625	630	635
Ala Gly Ala Gln Gly Pro Pro Gly Pro Ala Gly Pro Ala Gly Glu Arg		
	645	650
Gly Glu Gln Gly Pro Ala Gly Ser Pro Gly Phe Gln Gly Leu Pro Gly		
	660	665
Pro Ala Gly Pro Pro Gly Glu Ala Gly Lys Pro Gly Glu Gln Gly Val		
	675	680
		685

Pro Gly Asp Leu Gly Ala Pro Gly Pro Ser Gly Ala Arg Gly Glu Arg
 690 695 700
 Gly Phe Pro Gly Glu Arg Gly Val Gln Gly Pro Pro Gly Pro Ala Gly
 705 710 715 720
 Pro Arg Gly Ala Asn Gly Ala Pro Gly Asn Asp Gly Ala Lys Gly Asp
 725 730 735
 Ala Gly Ala Pro Gly Ala Pro Gly Ser Gln Gly Ala Pro Gly Leu Gln
 740 745 750
 Gly Met Pro Gly Glu Arg Gly Ala Ala Gly Leu Pro Gly Pro Lys Gly
 755 760 765
 Asp Arg Gly Asp Ala Gly Pro Lys Gly Ala Asp Gly Ser Pro Gly Lys
 770 775 780
 Asp Gly Val Arg Gly Leu Thr Gly Pro Ile Gly Pro Pro Gly Pro Ala
 785 790 795 800
 Gly Ala Pro Gly Asp Lys Gly Glu Ser Gly Pro Ser Gly Pro Ala Gly
 805 810 815
 Pro Thr Gly Ala Arg Gly Ala Pro Gly Asp Arg Gly Glu Pro Gly Pro
 820 825 830
 Pro Gly Pro Ala Gly Phe Ala Gly Pro Pro Gly Ala Asp Gly Gln Pro
 835 840 845
 Gly Ala Lys Gly Glu Pro Gly Asp Ala Gly Ala Lys Gly Asp Ala Gly
 850 855 860
 Pro Pro Gly Pro Ala Gly Pro Ala Gly Pro Pro Gly Pro Ile Gly Asn
 865 870 875 880
 Val Gly Ala Pro Gly Ala Lys Gly Ala Arg Gly Ser Ala Gly Pro Pro
 885 890 895
 Gly Ala Thr Gly Phe Pro Gly Ala Ala Gly Arg Val Gly Pro Pro Gly
 900 905 910
 Pro Ser Gly Asn Ala Gly Pro Pro Gly Pro Pro Gly Pro Ala Gly Lys
 915 920 925
 Glu Gly Gly Lys Gly Pro Arg Gly Glu Thr Gly Pro Ala Gly Arg Pro
 930 935 940
 Gly Glu Val Gly Pro Pro Gly Pro Pro Gly Pro Ala Gly Glu Lys Gly
 945 950 955 960
 Ser Pro Gly Ala Asp Gly Pro Ala Gly Ala Pro Gly Thr Pro Gly Pro
 965 970 975
 Gln Gly Ile Ala Gly Gln Arg Gly Val Val Gly Leu Pro Gly Gln Arg
 980 985 990
 Gly Glu Arg Gly Phe Pro Gly Leu Pro Gly Pro Ser Gly Glu Pro Gly

995	1000	1005
Lys Gln Gly Pro Ser Gly	Ala Ser Gly Glu Arg	Gly Pro Pro Gly
1010	1015	1020
Pro Met Gly Pro Pro Gly	Leu Ala Gly Pro Pro	Gly Glu Ser Gly
1025	1030	1035
Arg Glu Gly Ala Pro Gly	Ala Glu Gly Ser Pro	Gly Arg Asp Gly
1040	1045	1050
Ser Pro Gly Ala Lys Gly	Asp Arg Gly Glu Thr	Gly Pro Ala Gly
1055	1060	1065
Pro Pro Gly Ala Pro Gly	Ala Pro Gly Ala Pro	Gly Pro Val Gly
1070	1075	1080
Pro Ala Gly Lys Ser Gly	Asp Arg Gly Glu Thr	Gly Pro Ala Gly
1085	1090	1095
Pro Ala Gly Pro Val Gly	Pro Ala Gly Ala Arg	Gly Pro Ala Gly
1100	1105	1110
Pro Gln Gly Pro Arg Gly	Asp Lys Gly Glu Thr	Gly Glu Gln Gly
1115	1120	1125
Asp Arg Gly Ile Lys Gly	His Arg Gly Phe Ser	Gly Leu Gln Gly
1130	1135	1140
Pro Pro Gly Pro Pro Gly	Ser Pro Gly Glu Gln	Gly Pro Ser Gly
1145	1150	1155
Ala Ser Gly Pro Ala Gly	Pro Arg Gly Pro Pro	Gly Ser Ala Gly
1160	1165	1170
Ala Pro Gly Lys Asp Gly	Leu Asn Gly Leu Pro	Gly Pro Ile Gly
1175	1180	1185
Pro Pro Gly Pro Arg Gly	Arg Thr Gly Asp Ala	Gly Pro Val Gly
1190	1195	1200
Pro Pro Gly Pro Pro Gly	Pro Pro Gly Pro Pro	Gly Pro Pro Ser
1205	1210	1215
Ala Gly Phe Asp Phe Ser	Phe Leu Pro Gln Pro	Pro Gln Glu Lys
1220	1225	1230
Ala His Asp Gly Gly Arg	Tyr Tyr Arg Ala Asp	Asp Ala Asn Val
1235	1240	1245
Val Arg Asp Arg Asp Leu	Glu Val Asp Thr Thr	Leu Lys Ser Leu
1250	1255	1260
Ser Gln Gln Ile Glu Asn	Ile Arg Ser Pro Glu	Gly Ser Arg Lys
1265	1270	1275
Asn Pro Ala Arg Thr Cys	Arg Asp Leu Lys Met	Cys His Ser Asp
1280	1285	1290

Trp Lys Ser Gly Glu Tyr Trp Ile Asp Pro Asn Gln Gly Cys Asn
 1295 1300 1305
 Leu Asp Ala Ile Lys Val Phe Cys Asn Met Glu Thr Gly Glu Thr
 1310 1315 1320
 Cys Val Tyr Pro Thr Gln Pro Ser Val Ala Gln Lys Asn Trp Tyr
 1325 1330 1335
 Ile Ser Lys Asn Pro Lys Asp Lys Arg His Val Trp Phe Gly Glu
 1340 1345 1350
 Ser Met Thr Asp Gly Phe Gln Phe Glu Tyr Gly Gly Gln Gly Ser
 1355 1360 1365
 Asp Pro Ala Asp Val Ala Ile Gln Leu Thr Phe Leu Arg Leu Met
 1370 1375 1380
 Ser Thr Glu Ala Ser Gln Asn Ile Thr Tyr His Cys Lys Asn Ser
 1385 1390 1395
 Val Ala Tyr Met Asp Gln Gln Thr Gly Asn Leu Lys Lys Ala Leu
 1400 1405 1410
 Leu Leu Lys Gly Ser Asn Glu Ile Glu Ile Arg Ala Glu Gly Asn
 1415 1420 1425
 Ser Arg Phe Thr Tyr Ser Val Thr Val Asp Gly Cys Thr Ser His
 1430 1435 1440
 Thr Gly Ala Trp Gly Lys Thr Val Ile Glu Tyr Lys Thr Thr Lys
 1445 1450 1455
 Thr Ser Arg Leu Pro Ile Ile Asp Val Ala Pro Leu Asp Val Gly
 1460 1465 1470
 Ala Pro Asp Gln Glu Phe Gly Phe Asp Val Gly Pro Val Cys Phe
 1475 1480 1485

Leu

<210>4

<211>4362

<212>DNA

<213> 人工序列

<220>

<223> 含有与人胶原 $\alpha 2(I)$ 链融合的大麦巯基蛋白酶 aleurain 前体基因的维管信号序列

编码区和侧翼区的合成序列

<400>4

gcgatgcatg taatgtcatg agccacatga tccaatggcc acaggaacgt aagaatgtag	60
atagatttga ttttgtccgt tagatagcaa acaacattat aaaaggtgtg tatcaatagc	120
aactaattca ctcatggat tcatagaagt ccattcctcc taagtatcta aaccatggct	180
cacgctcgtg ttctctect egctctcgt gttttggcaa cagctgctgt ggctgtggct	240
tcaagttcta gttttgctga ttccaacca attcgtccag ttactgatag agcagcttcc	300
actttggctc aattgcttca agaagaaact gtgaggaagg gccctgctgg cgataggggc	360
cctaggggcg aaaggggtcc accaggacct ccaggcaggg atggcgaaga tggccaact	420
ggccctcctg gacctcctgg ccctccaggg ccacccggct tgggcgaaa cttcgcagct	480
caatacgatg gcaaggggtg tggctttggc cctggctcta tgggcttgat gggacctaga	540
ggcccacctg gtgctgctgg tgctcctgga ccacagggtt ttcagggacc agctggcgag	600
ccaggagagc caggccaaac aggaccagct ggtgcaaggg gacctgctgg acctcctgga	660
aaagctgggtg aagatgggtca cccaggcaaa ccaggacgtc ctggcgaaag aggtgttgtt	720
ggaccacaag gcgctagggg atttccaggt acacctggat tgccaggttt taagggcatt	780
cgtaggtcata acggcctcga tggattgaag ggacagcctg gcgcacctgg cgtaagggt	840
gaacctggag caccaggtga aaacgggtact cctggccaga ctggtgcaag aggactccca	900
ggtgaaaggg gtagagttgg tgctcctgga cctgctggag ctaggggtag tgatggtagt	960
gttggtcctg tgggacctgc tggccaatc ggttccgctg gcccacctgg attcccaggc	1020
gctccaggac ctaaaggaga aatcggtgct gtgggtaacg caggtcctac tggctccagca	1080
ggtcctcgtg gagaagtggg attgccagga ctttctggtc cagtgggccc tccaggcaac	1140
cctggagcta acggcttgac aggagctaaa ggccgagcag gactccctgg agtggctggc	1200
gcaccaggat tgcctggtec aaggggtatc ccaggccctg ttggcgcagc tggagctact	1260
ggtgcacgtg gacttgttgg cgaaccaggc cctgctggat caaaaggcga gtctggaaat	1320
aagggagaac ctggttctgc tggacctcaa ggtcctcctg gaccttctgg agaagaagga	1380
aaaaggggac caaatggcga ggctggatca gcaggctccac caggaccacc tggacttctg	1440
ggatcccctg gtagtagagg acttccaggc gctgatggta gagcaggcgt tatgggacca	1500
ccaggaagta gaggagcatc cggctccagca ggagttaggg gtcctaacgg agatgctggt	1560
agaccaggtg aaccaggtct tatgggceca aggggcctcc caggtagtcc aggaaatata	1620
ggccctgctg gaaaagaagg ccctgttggc cttccaggta ttgatggacg tcctggccct	1680
attggcccag caggtgcaag aggagaacct ggcaatattg gatttccagg accaaagggt	1740
ccaacaggcg atcctggaaa aaatggagat aagggctcatg ctggattggc aggcgcaagg	1800
ggcgctcctg gtccagatgg aaacaacggc gcacagggtc cacctggccc tcagggtgtt	1860
caaggcgga aaggcgaaca aggcccagct ggaccaccag gctttcaagg cttgccagga	1920
ccaagtggtc cagcaggtga agttggcaag ccaggcgagc gtggacttca tggcgagttt	1980
ggactccctg gaccagcagg accaaggggt gaaagaggcc ctccctggaga gagtggcgt	2040
gctggaccaa caggcccaat cggtagtaga ggtcctagt gacctccagg ccagatgga	2100
aataagggtg aaccaggagt tgtgggcgt gttggaacag ctggctcttc aggaccatca	2160
ggactcccag gcgagagagg cgctgctggc attcctggag gaaaagggtga aaaagcgaa	2220
cctggcctcc gtggcgaaat cggaaatcct ggacgtgatg gtgctcgtgg tgcacacggc	2280
gctgtgggcg ctccaggccc tgctgggtgct actggtgata gaggagaggc tggcgcagct	2340

ggcccagcag	gtcctgctgg	cccaaggggt	agtcctggtg	aaagaggcga	agttggacct	2400
gctggcccta	acggctttgc	tggccctgct	ggagcagcag	gtcaacctgg	cgctaaaggt	2460
gaaaggggcg	gaaagggccc	aaaaggtgaa	aatggcgttg	tgggaccaac	tgggccagtg	2520
ggcgcagctg	gacctgctgg	tccaaatgga	ccaccaggac	cagcaggtag	tagaggagat	2580
ggtggacctc	caggaatgac	aggttttcca	ggtgctgctg	gtagaacagg	acctcctggt	2640
cctagtggta	tttctggtec	accaggacca	ccaggctctg	ctggaaaaga	aggattgagg	2700
ggtccacgtg	gtgatcaagg	accagtgggc	agaactggtg	aagttggcgc	agtgggacca	2760
cctggttttg	ctggagaaaa	gggcccttct	ggagaggcag	gaacagctgg	tcctcctggt	2820
acacctggac	ctcaaggact	tttgggtgca	cctggtattc	tcggattgcc	aggaagtagg	2880
ggcgaacgtg	gacttcctgg	cgtggcagga	gcagttggag	aacctggccc	tctcggaatc	2940
gcaggcccac	caggcgcaag	aggaccacca	ggagctgttg	gatcaccagg	cgtgaatggt	3000
gcacctggcg	aggctggteg	tgatggaaac	ccaggaaatg	atggcccacc	aggaagagat	3060
ggtcaacctg	gacacaaagg	cgagaggggc	taccaggaa	atattggccc	agttggtgct	3120
gctggcgcac	caggcccaca	cggtccagtt	ggaccagcag	gaaaacacgg	taatcgtggc	3180
gaaacaggcc	cttcaggccc	agtgggacct	gctggtgctg	ttggcccagg	aggaccatct	3240
ggacctcaag	gcattagagg	cgataagggg	gagcctggcg	aaaaaggacc	tagaggcttg	3300
cctggtttta	aaggacacaa	cggtctccaa	ggacttccag	gtatcgtctg	tcatcatgga	3360
gatcagggtg	ctcctggatc	agtgggtcca	gcaggctcta	gaggcccagc	aggcccttcc	3420
ggtccagcag	gaaaggatgg	acgtactggc	caccctggaa	ctgtgggccc	tgctggaatt	3480
agaggtcctc	aaggteatca	gggccctgct	ggccctccag	gtccaccagg	tcctccaggc	3540
ccaccaggag	tttcaggtgg	tggttacgat	tttggttacg	atggtgattt	ttaccgtgct	3600
gatcaacctc	gaagtgctcc	ttctctccgt	cctaaagatt	atgaagtga	tgctactttg	3660
aaatacctta	acaaccagat	tgagactctt	ctcacacctg	agggatcaag	aaagaatcca	3720
gcacgtacat	gccgtgatct	cagacttagt	caccagagat	ggtcaagtgg	ctattattgg	3780
attgatccta	atcagggttg	tacaatggag	gctatcaaag	tttactgtga	ttttccaact	3840
ggagagacat	gtattagggc	acaacctgag	aacattccag	ctaaaaattg	gtatcgttcc	3900
tctaaagata	agaaacatgt	ttggctcgga	gagactatta	acgctggttc	tcagttcgag	3960
tataatgttg	agggcgttac	ttctaaagag	atggcaactc	agctcgcttt	tatgagattg	4020
ctcgctaact	acgcatccca	aaacatcact	tatcactgca	aaaattccat	tgcatatatg	4080
gatgaggaga	caggaaatth	gaagaaagca	gttattctcc	aaggtagtaa	cgatggtgag	4140
cttgtggctg	agggaaatag	tagattcact	tacacagttt	tgggtgatgg	atgctcaaag	4200
aaaactaatg	agtggggcaa	gacaatcatt	gagtacaaga	caaataagcc	ttctaggtct	4260
ccatttctcg	atattgcacc	tcttgatata	ggaggagctg	atcacagatt	ttttgttgat	4320
atcggacctg	tttgttttaa	gtaatgagct	cgcgcccgca	tc		4362

<210>5

<211>4362

<212>DNA

<213>人工序列

<220>

<223> 与人胶原 $\alpha 2(I)$ 链融合的大麦巯基蛋白酶 aleurain 前体基因的维管信号序列和侧

翼区的合成序列

<220>

<221>CDS

<222>(175)..(4344)

<400>5

```

gcgatgcatg taatgtcatg agccacatga tccaatggcc acaggaacgt aagaatgtag 60
atagatttga ttttgtccgt tagatagcaa acaacattat aaaagggtgtg tatcaatagc 120
aactaattca ctcatgggat tcatagaagt ccattcctcc taagtatcta aacc atg 177
Met
1
gct cac gct cgt gtt ctc ctc ctc gct ctc gct gtt ttg gca aca gct 225
Ala His Ala Arg Val Leu Leu Leu Ala Leu Ala Val Leu Ala Thr Ala
5 10 15
gct gtg gct gtg gct tca agt tct agt ttt gct gat tcc aac cca att 273
Ala Val Ala Val Ala Ser Ser Ser Ser Phe Ala Asp Ser Asn Pro Ile
20 25 30
cgt cca gtt act gat aga gca gct tcc act ttg gct caa ttg ctt caa 321
Arg Pro Val Thr Asp Arg Ala Ala Ser Thr Leu Ala Gln Leu Leu Gln
35 40 45
gaa gaa act gtg agg aag ggc cct gct ggc gat agg ggc cct agg ggc 369
Glu Glu Thr Val Arg Lys Gly Pro Ala Gly Asp Arg Gly Pro Arg Gly
50 55 60 65
gaa agg ggt cca cca gga cct cca ggc agg gat ggc gaa gat ggt cca 417
Glu Arg Gly Pro Pro Gly Pro Pro Gly Arg Asp Gly Glu Asp Gly Pro
70 75 80
act ggc cct cct gga cct cct ggc cct cca ggg cca ccc ggc ttg ggc 465
Thr Gly Pro Pro Gly Pro Pro Gly Pro Pro Gly Pro Pro Gly Leu Gly
85 90 95
gga aac ttc gca gct caa tac gat ggc aag ggt gtt ggt ctt ggt cct 513
Gly Asn Phe Ala Ala Gln Tyr Asp Gly Lys Gly Val Gly Leu Gly Pro
100 105 110
ggt cct atg ggc ttg atg gga cct aga ggc cca cct ggt gct gct ggt 561
Gly Pro Met Gly Leu Met Gly Pro Arg Gly Pro Pro Gly Ala Ala Gly

```

115	120	125	
gct cct gga cca cag ggt ttt cag gga cca gct ggc gag cca gga gag			609
Ala Pro Gly Pro Gln Gly Phe Gln Gly Pro Ala Gly Glu Pro Gly Glu			
130	135	140	145
cca ggc caa aca gga cca gct ggt gca agg gga cct gct gga cct cct			657
Pro Gly Gln Thr Gly Pro Ala Gly Ala Arg Gly Pro Ala Gly Pro Pro			
	150	155	160
gga aaa gct ggt gaa gat ggt cac cca ggc aaa cca gga cgt cct ggc			705
Gly Lys Ala Gly Glu Asp Gly His Pro Gly Lys Pro Gly Arg Pro Gly			
	165	170	175
gaa aga ggt gtt gtt gga cca caa ggc gct agg gga ttt cca ggt aca			753
Glu Arg Gly Val Val Gly Pro Gln Gly Ala Arg Gly Phe Pro Gly Thr			
	180	185	190
cct gga ttg cca ggt ttt aag ggc att cgt ggt cat aac ggc ctc gat			801
Pro Gly Leu Pro Gly Phe Lys Gly Ile Arg Gly His Asn Gly Leu Asp			
	195	200	205
gga ttg aag gga cag cct ggc gca cct ggc gtt aag ggt gaa cct gga			849
Gly Leu Lys Gly Gln Pro Gly Ala Pro Gly Val Lys Gly Glu Pro Gly			
210	215	220	225
gca cca ggt gaa aac ggt act cct ggc cag act ggt gca aga gga ctc			897
Ala Pro Gly Glu Asn Gly Thr Pro Gly Gln Thr Gly Ala Arg Gly Leu			
	230	235	240
cca ggt gaa agg ggt aga gtt ggt gct cct gga cct gct gga gct agg			945
Pro Gly Glu Arg Gly Arg Val Gly Ala Pro Gly Pro Ala Gly Ala Arg			
	245	250	255
ggt agt gat ggt agt gtt ggt cct gtg ggc cct gct ggt cca atc ggt			993
Gly Ser Asp Gly Ser Val Gly Pro Val Gly Pro Ala Gly Pro Ile Gly			
	260	265	270
tcc gct ggc cca cct gga ttc cca ggc gct cca gga cct aaa gga gaa			1041
Ser Ala Gly Pro Pro Gly Phe Pro Gly Ala Pro Gly Pro Lys Gly Glu			
	275	280	285
atc ggt gct gtg ggt aac gca ggt cct act ggt cca gca ggt cct cgt			1089
Ile Gly Ala Val Gly Asn Ala Gly Pro Thr Gly Pro Ala Gly Pro Arg			
290	295	300	305
gga gaa gtg gga ttg cca gga ctt tct ggt cca gtg ggc cct cca ggc			1137
Gly Glu Val Gly Leu Pro Gly Leu Ser Gly Pro Val Gly Pro Pro Gly			
	310	315	320
aac cct gga gct aac ggc ttg aca gga gct aaa ggc gca gca gga ctc			1185
Asn Pro Gly Ala Asn Gly Leu Thr Gly Ala Lys Gly Ala Ala Gly Leu			

325	330	335	
cct gga gtg gct ggc gca cca gga ttg cct ggt cca agg ggt atc cca			1233
Pro Gly Val Ala Gly Ala Pro Gly Leu Pro Gly Pro Arg Gly Ile Pro			
340	345	350	
ggc cct gtt ggc gca gct gga gct act ggt gca cgt gga ctt gtt ggc			1281
Gly Pro Val Gly Ala Ala Gly Ala Thr Gly Ala Arg Gly Leu Val Gly			
355	360	365	
gaa cca ggc cct gct gga tca aaa ggc gag tct gga aat aag gga gaa			1329
Glu Pro Gly Pro Ala Gly Ser Lys Gly Glu Ser Gly Asn Lys Gly Glu			
370	375	380	385
cct ggt tct gct gga cct caa ggt cct cct gga cct tct gga gaa gaa			1377
Pro Gly Ser Ala Gly Pro Gln Gly Pro Pro Gly Pro Ser Gly Glu Glu			
390	395	400	
gga aaa agg gga cca aat ggc gag gct gga tca gca ggt cca cca gga			1425
Gly Lys Arg Gly Pro Asn Gly Glu Ala Gly Ser Ala Gly Pro Pro Gly			
405	410	415	
cca cct gga ctt cgt gga tcc cct ggt agt aga gga ctt cca ggc gct			1473
Pro Pro Gly Leu Arg Gly Ser Pro Gly Ser Arg Gly Leu Pro Gly Ala			
420	425	430	
gat ggt aga gca ggc gtt atg gga cca cca gga agt aga gga gca tcc			1521
Asp Gly Arg Ala Gly Val Met Gly Pro Pro Gly Ser Arg Gly Ala Ser			
435	440	445	
ggt cca gca gga gtt agg ggt cct aac gga gat gct ggt aga cca ggt			1569
Gly Pro Ala Gly Val Arg Gly Pro Asn Gly Asp Ala Gly Arg Pro Gly			
450	455	460	465
gaa cca ggt ctt atg ggc cca agg ggc ctc cca ggt agt cca gga aat			1617
Glu Pro Gly Leu Met Gly Pro Arg Gly Leu Pro Gly Ser Pro Gly Asn			
470	475	480	
atc ggc cct gct gga aaa gaa ggc cct gtt gga ctt cca ggt att gat			1665
Ile Gly Pro Ala Gly Lys Glu Gly Pro Val Gly Leu Pro Gly Ile Asp			
485	490	495	
gga cgt cct ggc cct att ggc cca gca ggt gca aga gga gaa cct ggc			1713
Gly Arg Pro Gly Pro Ile Gly Pro Ala Gly Ala Arg Gly Glu Pro Gly			
500	505	510	
aat att gga ttt cca gga cca aag ggt cca aca ggc gat cct gga aaa			1761
Asn Ile Gly Phe Pro Gly Pro Lys Gly Pro Thr Gly Asp Pro Gly Lys			
515	520	525	
aat gga gat aag ggt cat gct gga ttg gca ggc gca agg ggc gct cct			1809
Asn Gly Asp Lys Gly His Ala Gly Leu Ala Gly Ala Arg Gly Ala Pro			

530	535	540	545	
ggt cca gat gga aac aac ggc gca cag ggt cca cct ggc cct cag ggt				1857
Gly Pro Asp Gly Asn Asn Gly Ala Gln Gly Pro Pro Gly Pro Gln Gly				
	550	555	560	
gtt caa ggc gga aaa ggc gaa caa ggc cca gct gga cca cca ggc ttt				1905
Val Gln Gly Gly Lys Gly Glu Gln Gly Pro Ala Gly Pro Pro Gly Phe				
	565	570	575	
caa ggc ttg cca gga cca agt ggt cca gca ggt gaa gtt ggc aag cca				1953
Gln Gly Leu Pro Gly Pro Ser Gly Pro Ala Gly Glu Val Gly Lys Pro				
	580	585	590	
ggc gag cgt gga ctt cat ggc gag ttt gga ctc cct gga cca gca gga				2001
Gly Glu Arg Gly Leu His Gly Glu Phe Gly Leu Pro Gly Pro Ala Gly				
	595	600	605	
cca agg ggt gaa aga ggc cct cct gga gag agt ggc gct gct gga cca				2049
Pro Arg Gly Glu Arg Gly Pro Pro Gly Glu Ser Gly Ala Ala Gly Pro				
610	615	620	625	
aca ggc cca atc ggt agt aga ggt cct agt gga cct cca ggc cca gat				2097
Thr Gly Pro Ile Gly Ser Arg Gly Pro Ser Gly Pro Pro Gly Pro Asp				
	630	635	640	
gga aat aag ggt gaa cca gga gtt gtg ggc gct gtt gga aca gct ggt				2145
Gly Asn Lys Gly Glu Pro Gly Val Val Gly Ala Val Gly Thr Ala Gly				
	645	650	655	
cct tca gga cca tca gga ctc cca ggc gag aga ggc gct gct ggc att				2193
Pro Ser Gly Pro Ser Gly Leu Pro Gly Glu Arg Gly Ala Ala Gly Ile				
	660	665	670	
cct gga gga aaa ggt gaa aaa ggc gaa cct ggc ctc cgt ggc gaa atc				2241
Pro Gly Gly Lys Gly Glu Lys Gly Glu Pro Gly Leu Arg Gly Glu Ile				
	675	680	685	
gga aat cct gga cgt gat ggt gct cgt ggt gca cac ggc gct gtg ggc				2289
Gly Asn Pro Gly Arg Asp Gly Ala Arg Gly Ala His Gly Ala Val Gly				
690	695	700	705	
gct cca ggc cct gct ggt gct act ggt gat aga gga gag gct ggc gca				2337
Ala Pro Gly Pro Ala Gly Ala Thr Gly Asp Arg Gly Glu Ala Gly Ala				
	710	715	720	
gct ggc cca gca ggt cct gct ggc cca agg ggt agt cct ggt gaa aga				2385
Ala Gly Pro Ala Gly Pro Ala Gly Pro Arg Gly Ser Pro Gly Glu Arg				
	725	730	735	
ggc gaa gtt gga cct gct ggc cct aac ggc ttt gct ggc cct gct gga				2433
Gly Glu Val Gly Pro Ala Gly Pro Asn Gly Phe Ala Gly Pro Ala Gly				

740	745	750	
gca gca ggt caa cct ggc gct aaa ggt gaa agg ggc gga aag ggc cca			2481
Ala Ala Gly Gln Pro Gly Ala Lys Gly Glu Arg Gly Gly Lys Gly Pro			
755	760	765	
aaa ggt gaa aat ggc gtt gtg gga cca act ggt cca gtg ggc gca gct			2529
Lys Gly Glu Asn Gly Val Val Gly Pro Thr Gly Pro Val Gly Ala Ala			
770	775	780	785
gga cct gct ggt cca aat gga cca cca gga cca gca ggt agt aga gga			2577
Gly Pro Ala Gly Pro Asn Gly Pro Pro Gly Pro Ala Gly Ser Arg Gly			
790	795	800	
gat ggt gga cct cca gga atg aca ggt ttt cca ggt gct gct ggt aga			2625
Asp Gly Gly Pro Pro Gly Met Thr Gly Phe Pro Gly Ala Ala Gly Arg			
805	810	815	
aca gga cct cct ggt cct agt ggt att tct ggt cca cca gga cca cca			2673
Thr Gly Pro Pro Gly Pro Ser Gly Ile Ser Gly Pro Pro Gly Pro Pro			
820	825	830	
ggt cct gct gga aaa gaa gga ttg agg ggt cca cgt ggt gat caa gga			2721
Gly Pro Ala Gly Lys Glu Gly Leu Arg Gly Pro Arg Gly Asp Gln Gly			
835	840	845	
cca gtg ggc aga act ggt gaa gtt ggc gca gtg gga cca cct ggt ttt			2769
Pro Val Gly Arg Thr Gly Glu Val Gly Ala Val Gly Pro Pro Gly Phe			
850	855	860	865
gct gga gaa aag ggc cct tct gga gag gca gga aca gct ggt cct cct			2817
Ala Gly Glu Lys Gly Pro Ser Gly Glu Ala Gly Thr Ala Gly Pro Pro			
870	875	880	
ggt aca cct gga cct caa gga ctt ttg ggt gca cct ggt att ctc gga			2865
Gly Thr Pro Gly Pro Gln Gly Leu Leu Gly Ala Pro Gly Ile Leu Gly			
885	890	895	
ttg cca gga agt agg ggc gaa cgt gga ctt cct ggc gtg gca gga gca			2913
Leu Pro Gly Ser Arg Gly Glu Arg Gly Leu Pro Gly Val Ala Gly Ala			
900	905	910	
gtt gga gaa cct ggc cct ctc gga atc gca ggc cca cca ggc gca aga			2961
Val Gly Glu Pro Gly Pro Leu Gly Ile Ala Gly Pro Pro Gly Ala Arg			
915	920	925	
gga cca cca gga gct gtt gga tca cca ggc gtg aat ggt gca cct ggc			3009
Gly Pro Pro Gly Ala Val Gly Ser Pro Gly Val Asn Gly Ala Pro Gly			
930	935	940	945
gag gct ggt cgt gat gga aac cca gga aat gat ggc cca cca gga aga			3057
Glu Ala Gly Arg Asp Gly Asn Pro Gly Asn Asp Gly Pro Pro Gly Arg			

	950	955	960	
	gat ggt caa cct gga cac aaa ggc gag agg ggc tac cca gga aat att			3105
	Asp Gly Gln Pro Gly His Lys Gly Glu Arg Gly Tyr Pro Gly Asn Ile			
	965	970	975	
	ggc cca gtt ggt gct gct ggc gca cca ggc cca cac ggt cca gtt gga			3153
	Gly Pro Val Gly Ala Ala Gly Ala Pro Gly Pro His Gly Pro Val Gly			
	980	985	990	
	cca gca gga aaa cac ggt aat cgt ggc gaa aca ggc cct tca ggc cca			3201
	Pro Ala Gly Lys His Gly Asn Arg Gly Glu Thr Gly Pro Ser Gly Pro			
	995	1000	1005	
	gtg gga cct gct ggt gct gtt ggc cca aga gga cca tct gga cct			3246
	Val Gly Pro Ala Gly Ala Val Gly Pro Arg Gly Pro Ser Gly Pro			
	1010	1015	1020	
	caa ggc att aga ggc gat aag gga gag cct ggc gaa aaa gga cct			3291
	Gln Gly Ile Arg Gly Asp Lys Gly Glu Pro Gly Glu Lys Gly Pro			
	1025	1030	1035	
	aga ggc ttg cct ggt ttt aaa gga cac aac ggt ctc caa gga ctt			3336
	Arg Gly Leu Pro Gly Phe Lys Gly His Asn Gly Leu Gln Gly Leu			
	1040	1045	1050	
	cca ggt atc gct ggt cat cat gga gat cag ggt gct cct gga tca			3381
	Pro Gly Ile Ala Gly His His Gly Asp Gln Gly Ala Pro Gly Ser			
	1055	1060	1065	
	gtg ggt cca gca ggt cct aga ggc cca gca ggc cct tcc ggt cca			3426
	Val Gly Pro Ala Gly Pro Arg Gly Pro Ala Gly Pro Ser Gly Pro			
	1070	1075	1080	
	gca gga aag gat gga cgt act ggc cac cct gga act gtg ggc cct			3471
	Ala Gly Lys Asp Gly Arg Thr Gly His Pro Gly Thr Val Gly Pro			
	1085	1090	1095	
	gct gga att aga ggt cct caa ggt cat cag ggc cct gct ggc cct			3516
	Ala Gly Ile Arg Gly Pro Gln Gly His Gln Gly Pro Ala Gly Pro			
	1100	1105	1110	
	cca ggt cca cca ggt cct cca ggc cca cca gga gtt tca ggt ggt			3561
	Pro Gly Pro Pro Gly Pro Pro Gly Pro Pro Gly Val Ser Gly Gly			
	1115	1120	1125	
	ggt tac gat ttt ggt tac gat ggt gat ttt tac cgt gct gat caa			3606
	Gly Tyr Asp Phe Gly Tyr Asp Gly Asp Phe Tyr Arg Ala Asp Gln			
	1130	1135	1140	
	cct aga agt gct cct tct ctc cgt cct aaa gat tat gaa gtt gat			3651
	Pro Arg Ser Ala Pro Ser Leu Arg Pro Lys Asp Tyr Glu Val Asp			

1145	1150	1155	
gct act ttg aaa tca ctt aac aac cag att gag act ctt ctc aca			3696
Ala Thr Leu Lys Ser Leu Asn Asn Gln Ile Glu Thr Leu Leu Thr			
1160	1165	1170	
cct gag gga tca aga aag aat cca gca cgt aca tgc cgt gat ctc			3741
Pro Glu Gly Ser Arg Lys Asn Pro Ala Arg Thr Cys Arg Asp Leu			
1175	1180	1185	
aga ctt agt cac cca gag tgg tca agt ggc tat tat tgg att gat			3786
Arg Leu Ser His Pro Glu Trp Ser Ser Gly Tyr Tyr Trp Ile Asp			
1190	1195	1200	
cct aat cag ggt tgt aca atg gag gct atc aaa gtt tac tgt gat			3831
Pro Asn Gln Gly Cys Thr Met Glu Ala Ile Lys Val Tyr Cys Asp			
1205	1210	1215	
ttt cca act gga gag aca tgt att agg gca caa cct gag aac att			3876
Phe Pro Thr Gly Glu Thr Cys Ile Arg Ala Gln Pro Glu Asn Ile			
1220	1225	1230	
cca gct aaa aat tgg tat cgt tcc tct aaa gat aag aaa cat gtt			3921
Pro Ala Lys Asn Trp Tyr Arg Ser Ser Lys Asp Lys Lys His Val			
1235	1240	1245	
tgg ctc gga gag act att aac gct ggt tct cag ttc gag tat aat			3966
Trp Leu Gly Glu Thr Ile Asn Ala Gly Ser Gln Phe Glu Tyr Asn			
1250	1255	1260	
gtt gag ggc gtt act tct aaa gag atg gca act cag ctc gct ttt			4011
Val Glu Gly Val Thr Ser Lys Glu Met Ala Thr Gln Leu Ala Phe			
1265	1270	1275	
atg aga ttg ctc gct aac tac gca tcc caa aac atc act tat cac			4056
Met Arg Leu Leu Ala Asn Tyr Ala Ser Gln Asn Ile Thr Tyr His			
1280	1285	1290	
tgc aaa aat tcc att gca tat atg gat gag gag aca gga aat ttg			4101
Cys Lys Asn Ser Ile Ala Tyr Met Asp Glu Glu Thr Gly Asn Leu			
1295	1300	1305	
aag aaa gca gtt att ctc caa ggt agt aac gat gtt gag ctt gtg			4146
Lys Lys Ala Val Ile Leu Gln Gly Ser Asn Asp Val Glu Leu Val			
1310	1315	1320	
gct gag gga aat agt aga ttc act tac aca gtt ttg gtg gat gga			4191
Ala Glu Gly Asn Ser Arg Phe Thr Tyr Thr Val Leu Val Asp Gly			
1325	1330	1335	
tgc tca aag aaa act aat gag tgg ggc aag aca atc att gag tac			4236
Cys Ser Lys Lys Thr Asn Glu Trp Gly Lys Thr Ile Ile Glu Tyr			

1340 1345 1350
aag aca aat aag cct tct agg etc cca ttt etc gat att gca cct 4281
Lys Thr Asn Lys Pro Ser Arg Leu Pro Phe Leu Asp Ile Ala Pro
1355 1360 1365
ctt gat atc gga gga gct gat cac gag ttt ttt gtt gat atc gga 4326
Leu Asp Ile Gly Gly Ala Asp His Glu Phe Phe Val Asp Ile Gly
1370 1375 1380
cct gtt tgt ttt aag taa tgagctcgcg gccgcatc 4362
Pro Val Cys Phe Lys
1385

<210>6

<211>1389

<212>PRT

<213>人工序列

<220>

<223> 含有与人胶原 $\alpha 2(I)$ 链融合的大麦巯基蛋白酶 aleurain 前体基因的维管信号
序列

编码区和侧翼区的合成序列

<400>6

Met Ala His Ala Arg Val Leu Leu Leu Ala Leu Ala Val Leu Ala Thr
1 5 10 15
Ala Ala Val Ala Val Ala Ser Ser Ser Ser Phe Ala Asp Ser Asn Pro
 20 25 30
Ile Arg Pro Val Thr Asp Arg Ala Ala Ser Thr Leu Ala Gln Leu Leu
 35 40 45
Gln Glu Glu Thr Val Arg Lys Gly Pro Ala Gly Asp Arg Gly Pro Arg
 50 55 60
Gly Glu Arg Gly Pro Pro Gly Pro Pro Gly Arg Asp Gly Glu Asp Gly
65 70 75 80
Pro Thr Gly Pro Pro Gly Pro Pro Gly Pro Pro Gly Pro Pro Gly Leu
 85 90 95
Gly Gly Asn Phe Ala Ala Gln Tyr Asp Gly Lys Gly Val Gly Leu Gly
 100 105 110
Pro Gly Pro Met Gly Leu Met Gly Pro Arg Gly Pro Pro Gly Ala Ala
 115 120 125

Gly Ala Pro Gly Pro Gln Gly Phe Gln Gly Pro Ala Gly Glu Pro Gly
 130 135 140
 Glu Pro Gly Gln Thr Gly Pro Ala Gly Ala Arg Gly Pro Ala Gly Pro
 145 150 155 160
 Pro Gly Lys Ala Gly Glu Asp Gly His Pro Gly Lys Pro Gly Arg Pro
 165 170 175
 Gly Glu Arg Gly Val Val Gly Pro Gln Gly Ala Arg Gly Phe Pro Gly
 180 185 190
 Thr Pro Gly Leu Pro Gly Phe Lys Gly Ile Arg Gly His Asn Gly Leu
 195 200 205
 Asp Gly Leu Lys Gly Gln Pro Gly Ala Pro Gly Val Lys Gly Glu Pro
 210 215 220
 Gly Ala Pro Gly Glu Asn Gly Thr Pro Gly Gln Thr Gly Ala Arg Gly
 225 230 235 240
 Leu Pro Gly Glu Arg Gly Arg Val Gly Ala Pro Gly Pro Ala Gly Ala
 245 250 255
 Arg Gly Ser Asp Gly Ser Val Gly Pro Val Gly Pro Ala Gly Pro Ile
 260 265 270
 Gly Ser Ala Gly Pro Pro Gly Phe Pro Gly Ala Pro Gly Pro Lys Gly
 275 280 285
 Glu Ile Gly Ala Val Gly Asn Ala Gly Pro Thr Gly Pro Ala Gly Pro
 290 295 300
 Arg Gly Glu Val Gly Leu Pro Gly Leu Ser Gly Pro Val Gly Pro Pro
 305 310 315 320
 Gly Asn Pro Gly Ala Asn Gly Leu Thr Gly Ala Lys Gly Ala Ala Gly
 325 330 335
 Leu Pro Gly Val Ala Gly Ala Pro Gly Leu Pro Gly Pro Arg Gly Ile
 340 345 350
 Pro Gly Pro Val Gly Ala Ala Gly Ala Thr Gly Ala Arg Gly Leu Val
 355 360 365
 Gly Glu Pro Gly Pro Ala Gly Ser Lys Gly Glu Ser Gly Asn Lys Gly
 370 375 380
 Glu Pro Gly Ser Ala Gly Pro Gln Gly Pro Pro Gly Pro Ser Gly Glu
 385 390 395 400
 Glu Gly Lys Arg Gly Pro Asn Gly Glu Ala Gly Ser Ala Gly Pro Pro
 405 410 415
 Gly Pro Pro Gly Leu Arg Gly Ser Pro Gly Ser Arg Gly Leu Pro Gly
 420 425 430
 Ala Asp Gly Arg Ala Gly Val Met Gly Pro Pro Gly Ser Arg Gly Ala

435	440	445
Ser Gly Pro Ala Gly Val Arg Gly Pro Asn Gly Asp Ala Gly Arg Pro		
450	455	460
Gly Glu Pro Gly Leu Met Gly Pro Arg Gly Leu Pro Gly Ser Pro Gly		
465	470	475
Asn Ile Gly Pro Ala Gly Lys Glu Gly Pro Val Gly Leu Pro Gly Ile		
485	490	495
Asp Gly Arg Pro Gly Pro Ile Gly Pro Ala Gly Ala Arg Gly Glu Pro		
500	505	510
Gly Asn Ile Gly Phe Pro Gly Pro Lys Gly Pro Thr Gly Asp Pro Gly		
515	520	525
Lys Asn Gly Asp Lys Gly His Ala Gly Leu Ala Gly Ala Arg Gly Ala		
530	535	540
Pro Gly Pro Asp Gly Asn Asn Gly Ala Gln Gly Pro Pro Gly Pro Gln		
545	550	555
Gly Val Gln Gly Gly Lys Gly Glu Gln Gly Pro Ala Gly Pro Pro Gly		
565	570	575
Phe Gln Gly Leu Pro Gly Pro Ser Gly Pro Ala Gly Glu Val Gly Lys		
580	585	590
Pro Gly Glu Arg Gly Leu His Gly Glu Phe Gly Leu Pro Gly Pro Ala		
595	600	605
Gly Pro Arg Gly Glu Arg Gly Pro Pro Gly Glu Ser Gly Ala Ala Gly		
610	615	620
Pro Thr Gly Pro Ile Gly Ser Arg Gly Pro Ser Gly Pro Pro Gly Pro		
625	630	635
Asp Gly Asn Lys Gly Glu Pro Gly Val Val Gly Ala Val Gly Thr Ala		
645	650	655
Gly Pro Ser Gly Pro Ser Gly Leu Pro Gly Glu Arg Gly Ala Ala Gly		
660	665	670
Ile Pro Gly Gly Lys Gly Glu Lys Gly Glu Pro Gly Leu Arg Gly Glu		
675	680	685
Ile Gly Asn Pro Gly Arg Asp Gly Ala Arg Gly Ala His Gly Ala Val		
690	695	700
Gly Ala Pro Gly Pro Ala Gly Ala Thr Gly Asp Arg Gly Glu Ala Gly		
705	710	715
Ala Ala Gly Pro Ala Gly Pro Ala Gly Pro Arg Gly Ser Pro Gly Glu		
725	730	735
Arg Gly Glu Val Gly Pro Ala Gly Pro Asn Gly Phe Ala Gly Pro Ala		
740	745	750

Gly Ala Ala Gly Gln Pro Gly Ala Lys Gly Glu Arg Gly Gly Lys Gly
 755 760 765
 Pro Lys Gly Glu Asn Gly Val Val Gly Pro Thr Gly Pro Val Gly Ala
 770 775 780
 Ala Gly Pro Ala Gly Pro Asn Gly Pro Pro Gly Pro Ala Gly Ser Arg
 785 790 795 800
 Gly Asp Gly Gly Pro Pro Gly Met Thr Gly Phe Pro Gly Ala Ala Gly
 805 810 815
 Arg Thr Gly Pro Pro Gly Pro Ser Gly Ile Ser Gly Pro Pro Gly Pro
 820 825 830
 Pro Gly Pro Ala Gly Lys Glu Gly Leu Arg Gly Pro Arg Gly Asp Gln
 835 840 845
 Gly Pro Val Gly Arg Thr Gly Glu Val Gly Ala Val Gly Pro Pro Gly
 850 855 860
 Phe Ala Gly Glu Lys Gly Pro Ser Gly Glu Ala Gly Thr Ala Gly Pro
 865 870 875 880
 Pro Gly Thr Pro Gly Pro Gln Gly Leu Leu Gly Ala Pro Gly Ile Leu
 885 890 895
 Gly Leu Pro Gly Ser Arg Gly Glu Arg Gly Leu Pro Gly Val Ala Gly
 900 905 910
 Ala Val Gly Glu Pro Gly Pro Leu Gly Ile Ala Gly Pro Pro Gly Ala
 915 920 925
 Arg Gly Pro Pro Gly Ala Val Gly Ser Pro Gly Val Asn Gly Ala Pro
 930 935 940
 Gly Glu Ala Gly Arg Asp Gly Asn Pro Gly Asn Asp Gly Pro Pro Gly
 945 950 955 960
 Arg Asp Gly Gln Pro Gly His Lys Gly Glu Arg Gly Tyr Pro Gly Asn
 965 970 975
 Ile Gly Pro Val Gly Ala Ala Gly Ala Pro Gly Pro His Gly Pro Val
 980 985 990
 Gly Pro Ala Gly Lys His Gly Asn Arg Gly Glu Thr Gly Pro Ser Gly
 995 1000 1005
 Pro Val Gly Pro Ala Gly Ala Val Gly Pro Arg Gly Pro Ser Gly
 1010 1015 1020
 Pro Gln Gly Ile Arg Gly Asp Lys Gly Glu Pro Gly Glu Lys Gly
 1025 1030 1035
 Pro Arg Gly Leu Pro Gly Phe Lys Gly His Asn Gly Leu Gln Gly
 1040 1045 1050
 Leu Pro Gly Ile Ala Gly His His Gly Asp Gln Gly Ala Pro Gly

1055	1060	1065
Ser Val Gly Pro Ala Gly	Pro Arg Gly Pro Ala Gly	Pro Ser Gly
1070	1075	1080
Pro Ala Gly Lys Asp Gly	Arg Thr Gly His Pro Gly	Thr Val Gly
1085	1090	1095
Pro Ala Gly Ile Arg Gly	Pro Gln Gly His Gln Gly	Pro Ala Gly
1100	1105	1110
Pro Pro Gly Pro Pro Gly	Pro Pro Gly Pro Pro Gly	Val Ser Gly
1115	1120	1125
Gly Gly Tyr Asp Phe Gly	Tyr Asp Gly Asp Phe Tyr	Arg Ala Asp
1130	1135	1140
Gln Pro Arg Ser Ala Pro	Ser Leu Arg Pro Lys Asp	Tyr Glu Val
1145	1150	1155
Asp Ala Thr Leu Lys Ser	Leu Asn Asn Gln Ile Glu	Thr Leu Leu
1160	1165	1170
Thr Pro Glu Gly Ser Arg	Lys Asn Pro Ala Arg Thr	Cys Arg Asp
1175	1180	1185
Leu Arg Leu Ser His Pro	Glu Trp Ser Ser Gly Tyr	Tyr Trp Ile
1190	1195	1200
Asp Pro Asn Gln Gly Cys	Thr Met Glu Ala Ile Lys	Val Tyr Cys
1205	1210	1215
Asp Phe Pro Thr Gly Glu	Thr Cys Ile Arg Ala Gln	Pro Glu Asn
1220	1225	1230
Ile Pro Ala Lys Asn Trp	Tyr Arg Ser Ser Lys Asp	Lys Lys His
1235	1240	1245
Val Trp Leu Gly Glu Thr	Ile Asn Ala Gly Ser Gln	Phe Glu Tyr
1250	1255	1260
Asn Val Glu Gly Val Thr	Ser Lys Glu Met Ala Thr	Gln Leu Ala
1265	1270	1275
Phe Met Arg Leu Leu Ala	Asn Tyr Ala Ser Gln Asn	Ile Thr Tyr
1280	1285	1290
His Cys Lys Asn Ser Ile	Ala Tyr Met Asp Glu Glu	Thr Gly Asn
1295	1300	1305
Leu Lys Lys Ala Val Ile	Leu Gln Gly Ser Asn Asp	Val Glu Leu
1310	1315	1320
Val Ala Glu Gly Asn Ser	Arg Phe Thr Tyr Thr Val	Leu Val Asp
1325	1330	1335
Gly Cys Ser Lys Lys Thr	Asn Glu Trp Gly Lys Thr	Ile Ile Glu
1340	1345	1350

Tyr Lys Thr Asn Lys Pro Ser Arg Leu Pro Phe Leu Asp Ile Ala
 1355 1360 1365
 Pro Leu Asp Ile Gly Gly Ala Asp His Glu Phe Phe Val Asp Ile
 1370 1375 1380
 Gly Pro Val Cys Phe Lys
 1385

<210>7

<211>127

<212>DNA

<213> 人工序列

<220>

<223> 含有拟南芥内切 -1,4- β -葡聚糖酶的质外体信号编码区和侧翼区的合成序列

<400>7

gccatggcta ggaagtcttt gattttccca gtgattcttc ttgctgtgct tcttttctct 60
 ccacctatatt actctgctgg acacgattat agggatgctc ttaggaagtc atctatggct 120
 caattgc 127

<210>8

<211>127

<212>DNA

<213> 人工序列

<220>

<223> 拟南芥内切 -1,4- β -葡聚糖酶的质外体信号和侧翼区的合成序列

<220>

<221>CDS

<222>(10)..(120)

<400>8

gccatggct agg aag tct ttg att ttc cca gtg att ctt ctt gct gtg ctt 51
 Arg Lys Ser Leu Ile Phe Pro Val Ile Leu Leu Ala Val Leu
 1 5 10
 ctt ttc tct cca cct att tac tct gct gga cac gat tat agg gat get 99
 Leu Phe Ser Pro Pro Ile Tyr Ser Ala Gly His Asp Tyr Arg Asp Ala
 15 20 25 30
 ctt agg aag tca tct atg gct caattgc 127

Leu Arg Lys Ser Ser Met Ala

35

<210>9

<211>37

<212>PRT

<213> 人工序列

<220>

<223> 拟南芥内切 -1,4- β -葡聚糖酶的质外体信号和侧翼区的合成序列

<400>9

Arg Lys Ser Leu Ile Phe Pro Val Ile Leu Leu Ala Val Leu Leu Phe

1 5 10 15

Ser Pro Pro Ile Tyr Ser Ala Gly His Asp Tyr Arg Asp Ala Leu Arg

20 25 30

Lys Ser Ser Met Ala

35

<210>10

<211>1037

<212>DNA

<213> 人工序列

<220>

<223> 菊花 rbcS1 启动子和 5' UTR

<400>10

aaatggcgcg ccaagcttag acaaacaccc cttgttatac aaagaatttc gctttacaaa 60

atcaaattcg agaaaataat atatgcacta aataagatca ttcggatcca atctaacc aa 120

ttacgatacg ctttgggtac acttgatttt tgtttcagta gttacatata tcttgtttta 180

tatgctatct ttaaggatct tcaactcaag actattttgtt gatgttcttg atggggctcg 240

gaagatttga tatgatacac tctaactctt aggagatacc agccaggatt atattcagta 300

agacaatcaa attttacgtg ttcaaactcg ttatcttttc atttaatgga tgagccagaa 360

tctctataga atgattgcaa tcgagaatat gttcggccga tatccctttg ttggettcaa 420

tattctacat atcacacaag aatcgaccgt attgtaccct ctttccataa aggaacacac 480

agtatgcaga tgcttttttc ccacatgcag taacataggt attcaaaaat ggctaaaaga 540

agtttgataa caaattgaca actatttcca tttctgttat ataaatttca caacacacaa 600

aagcccgtaa tcaagagtct gcccatgtac gaaataactt ctattatttg gtattgggccc 660

taagcccagc tcagagtacg tgggggtacc acatatagga aggtaacaaa atactgcaag	720
atagccccat aacgtaccag cctctcetta ccacgaagag ataagatata agaccaccc	780
tgccacgtgt cacatcgtca tgggtggttaa tgataaggga ttacatcctt ctatgtttgt	840
ggacatgatg catgtaatgt catgagccac atgatccaat ggccacagga acgtaagaat	900
gtagatagat ttgattttgt ccgttagata gcaaacaaca ttataaaagg tgtgtatcaa	960
tacgaactaa ttcactcatt ggattcatag aagtccattc ctcttaagta tctaacata	1020
tgcaattgtc gactaaa	1037

<210>11

<211>975

<212>DNA

<213> 人工序列

<220>

<223> 菊花 rbcS13' UTR 和终止子

<400>11

aaaaggatcc gcggccecat aagttttact atttaccaag acttttgaat attaaccctc	60
ttgtaacgag teggttaaat ttgattgttt agggttttgt attatttttt tttggctttt	120
taattcatca ctttaattcc ctaattgtct gttcatttcg ttgtttgttt ccggatcgat	180
aatgaaatgt aagagatata atatataaat aataaattgt cgtttcataat ttgcaatcct	240
tttttacaaa cttttaatta attgtatgta tgacattttc ttcttggtat attagggggga	300
aataatgtta aataaaaagta caaaataaac tacagtacat cgtactgaat aaattaccta	360
gccaaaaagt acacctttcc atatacttcc tacatgaagg cattttcaac attttcaaat	420
aaggaatgct acaaccgcat aataacatcc acaaattttt ttataaaata acatgtcaga	480
cagtgattga aagattttat tatagtttcg ttatcttctt ttctcattaa gcgaatcact	540
acctaacacg tcattttgtg aaatattttt tgaatgtttt tatatagttg tagcattcct	600
cttttcaaat tagggtttgt ttgagatagc atttcagccg gttcatacaa cttaaaagca	660
tactctaata ctggaaaaaa gactaaaaaa tcttghtaagt tagcgcagaa tattgaccca	720
aattatatac acacatgacc ccatatagag actaattaca cttttaacca ctaataatta	780
ttactgtatt ataacatcta ctaattaaac ttgtgagttt ttgctagaat tattatcata	840
tatactaaaa ggcaggaacg caaacattgc cccggtactg tagcaactac ggtagacgca	900
ttaattgtct atagtggacg cattaattaa ccaaaaccgc ctctttcccc ttcttcttga	960
agcttgagct ctttt	975

<210>12

<211>1633

<212>DNA

<213> 人工序列

<220>

<223> 含有与人脯氨酰 4-羟化酶 β 亚基融合的大麦巯基蛋白酶 aleurain 前体基因的维管

信号序列编码区和侧翼区的合成序列

<400>12

```

ctcgagtaaa ccatggctca tgctagggtt ttgcttttgg ctcttgctgt tcttgctact 60
gctgctgttg ctgtggcttc ttcttcatct ttcgctgatt ctaacccaat taggccagtg 120
actgatagag ctgcttctac tcttgctcaa ttggctgcaca tggatgctcc agaagaggag 180
gatcacgttc ttgtgcttag gaagtctaac ttcgctgaag ctcttgctgc tcacaagtac 240
cttcttgtgg agttttatgc tccttggtgc ggacattgca aagctcttgc tccagagtat 300
gctaaggctg ctggaaagt gaaggctgag ggatctgaaa ttaggcttgc taaagtggat 360
gctactgagg agtctgatct tgctcaacag tacggagtta ggggataccc aactattaag 420
ttcttcagga acggagatac tgcttctcca aaggagtata ctgctggaag ggaggctgat 480
gatattgtga actggcttaa gaagagaact ggaccagctg ctactactct tccagatgga 540
gctgctgctg aatctcttgt ggagtcactt gaggtggcag tgattggatt cttcaaggat 600
gtggagtctg attctgctaa gcagttcctt caagctgctg aggctattga tgatattcca 660
ttcggaatta cttctaactc tgatgtgttc tctaagtacc agcttgataa ggatggagtg 720
gtgcttttca agaaattcga tgaggggaagg aacaatttcg agggagaggt gacaaaggag 780
aaccttcttg atttcattaa gcacaaccag cttccacttg tgattgagtt cactgagcag 840
actgctccaa agattttcgg aggagagatt aagactcaca ttcttctttt cttccaaag 900
tctgtgtctg attacgatgg aaagttgtct aacttcaaga ctgctgctga gtctttcaag 960
ggaaagattc ttttcatatt cattgattct gatcacactg ataaccagag gattcttgag 1020
ttcttcggac ttaagaagga agagtgecca gctgttaggc ttattactct tgaggaggag 1080
atgactaagt acaagccaga gtctgaagaa cttactgctg agaggattac tgagttctgc 1140
cacagattcc ttgagggaaa gattaagcca caccttatgt ctcaagagct tccagaggat 1200
tgggataagc agccagttaa ggtgttgggt ggtaaaaact tcgaggatgt ggctttcgat 1260
gagaagaaga acgtgttcgt ggagttctac gcaccttggg gtggctactg taagcagctt 1320
gctccaattt gggataagt gggagagact tacaaggatc acgagaacat tgtgattgct 1380
aagatggatt ctactgctaa cgaggtggag gctgttaagg ttcactcttt cccaactttg 1440
aagttcttcc cagcttctgc tgataggact gtgattgatt acaacggaga aaggactctt 1500
gatggattca agaagttcct tgagtctgga ggacaagatg gagctggaga tgatgatgat 1560
cttgaggatt tggaagaagc tgaggagcca gatatggagg aggatgatga tcagaagget 1620
gtgtgatgag etc 1633

```

<210>13

<211>537

<212>PRT

<213> 人工序列

<220>

<223> 含有与人脯氨酰 4-羟化酶 β 亚基融合的大麦巯基蛋白酶 aleurain 前体基因的
维管

信号序列和侧翼区的合成序列

<400>13

```

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

```

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

<210>14

<211>1723

<212>DNA

<213> 人工序列

<220>

<223> 含有与人脯氨酰 4-羟化酶 α -1 亚基融合的大麦巯基蛋白酶 aleurain 前体基因的维

管信号序列编码区和侧翼区的合成序列

<400>14

```

ctcgagtaaa ccatggctca tgctagggtt ttgcttttgg ctcttgctgt tcttgctact 60
gctgctgttg ctgtggcttc ttcttcatct ttcgctgatt ctaacccaat taggccagtg 120
actgatagag ctgcttctac tcttgctcaa ttggctcgaca tgcacccagg attcttact 180
tctattggac agatgactga tcttattcac actgagaagg atcttgtgac ttctcttaag 240
gattacatta aggctgagga ggataagttg gagcagatta agaagtgggc tgagaagttg 300
gataggctta cttctactgc taaaaagat ccagagggat tcgttggcca tccagtgaac 360
gctttcaagt tgatgaagag gcttaacact gagtggagtg agcttgagaa ccttgtgctt 420
aaggatatgt ctgatggatt catttctaac ctactattc agaggcagta ctcccaaat 480
gatgaggatc aagtgggagc tgctaaggct cttcttaggc ttcaggatac ttacaacctt 540
gatactgata caatttctaa gggaaacctt ccaggagtta agcacaagtc tttccttact 600
gctgaggatt gcttcgagct tggaaaggtt gcatacactg aggctgatta ctaccacact 660
gagctttgga tggaacaagc tcttaggcaa cttgatgagg gagagatttc tactattgat 720
aaggtgtcag tgcttgatta ctttcttac gctgtgtacc agcagggtga tcttgataag 780
gctcttttgc ttactaagaa gttgcttgag cttgatccag aacatcagag ggctaacgga 840
aaccttaagt acttcgagta cattatggct aagggaaaagg atgtgaacaa gtctgcttct 900
gatgatcagt ctgatcaaaa gactactcca aagaagaagg gagtggctgt tgattatctt 960
cctgagagggc agaagtatga gatgttgtgt aggggagagg gtattaagat gactccaagg 1020
aggcagaaga agttgttctg caggtatcac gatggaaaca ggaacccaaa gttcattctt 1080
gctccagcta agcaagaaga tgagtgggat aagccaagga ttattaggtt ccacgatatt 1140
atctctgatg ctgagattga gattgtgaag gatcttgcta agccaagact taggagggtt 1200
actatttcta accctattac tggatgctt gagactgtgc actacaggat ttctaagtct 1260
gcttgctttt ctggatacga gaaccagtg gtgtctagga ttaacatgag gattcaggat 1320
cttactggac ttgatgtgtc tactgctgag gagcttcaag ttgctaacta cggagtggga 1380
ggacaatatg agccacactt cgatttcgct aggaaggatg agccagatgc ttttaaggag 1440
cttggaaactg gaaacaggat tgctacttgg cttttctaca tgtctgatgt ttctgctgga 1500
ggagctactg ttttccaga agtgggagct tctgtttggc caaagaaggg aactgctgtg 1560
ttctggtaca accttttgc ttctggagag ggagattact ctactaggca tgetgcttgc 1620
ccagttcttg ttggaaacaa gtgggtgtca aacaagtggc ttcattgagag gggacaagag 1680

```

tttagaaggc catgcactct ttctgagctt gagtgatgag etc

1723

<210>15

<211>567

<212>PRT

<213>人工序列

<220>

<223>含有与人脯氨酰 4-羟化酶 α -1 亚基融合的大麦巯基蛋白酶 aleurain 前体基因的维

管信号序列和侧翼区的合成序列

<400>15

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

Thr Glu Ala Asp Tyr Tyr His Thr Glu Leu Trp Met Glu Gln Ala Leu
 210 215 220
 Arg Gln Leu Asp Glu Gly Glu Ile Ser Thr Ile Asp Lys Val Ser Val
 225 230 235 240
 Leu Asp Tyr Leu Ser Tyr Ala Val Tyr Gln Gln Gly Asp Leu Asp Lys
 245 250 255
 Ala Leu Leu Leu Thr Lys Lys Leu Leu Glu Leu Asp Pro Glu His Gln
 260 265 270
 Arg Ala Asn Gly Asn Leu Lys Tyr Phe Glu Tyr Ile Met Ala Lys Glu
 275 280 285
 Lys Asp Val Asn Lys Ser Ala Ser Asp Asp Gln Ser Asp Gln Lys Thr
 290 295 300
 Thr Pro Lys Lys Lys Gly Val Ala Val Asp Tyr Leu Pro Glu Arg Gln
 305 310 315 320
 Lys Tyr Glu Met Leu Cys Arg Gly Glu Gly Ile Lys Met Thr Pro Arg
 325 330 335
 Arg Gln Lys Lys Leu Phe Cys Arg Tyr His Asp Gly Asn Arg Asn Pro
 340 345 350
 Lys Phe Ile Leu Ala Pro Ala Lys Gln Glu Asp Glu Trp Asp Lys Pro
 355 360 365
 Arg Ile Ile Arg Phe His Asp Ile Ile Ser Asp Ala Glu Ile Glu Ile
 370 375 380
 Val Lys Asp Leu Ala Lys Pro Arg Leu Arg Arg Ala Thr Ile Ser Asn
 385 390 395 400
 Pro Ile Thr Gly Asp Leu Glu Thr Val His Tyr Arg Ile Ser Lys Ser
 405 410 415
 Ala Trp Leu Ser Gly Tyr Glu Asn Pro Val Val Ser Arg Ile Asn Met
 420 425 430
 Arg Ile Gln Asp Leu Thr Gly Leu Asp Val Ser Thr Ala Glu Glu Leu
 435 440 445
 Gln Val Ala Asn Tyr Gly Val Gly Gly Gln Tyr Glu Pro His Phe Asp
 450 455 460
 Phe Ala Arg Lys Asp Glu Pro Asp Ala Phe Lys Glu Leu Gly Thr Gly
 465 470 475 480
 Asn Arg Ile Ala Thr Trp Leu Phe Tyr Met Ser Asp Val Ser Ala Gly
 485 490 495
 Gly Ala Thr Val Phe Pro Glu Val Gly Ala Ser Val Trp Pro Lys Lys
 500 505 510
 Gly Thr Ala Val Phe Trp Tyr Asn Leu Phe Ala Ser Gly Glu Gly Asp

515	520	525
Tyr Ser Thr Arg His Ala Ala Cys Pro Val Leu Val Gly Asn Lys Trp		
530	535	540
Val Ser Asn Lys Trp Leu His Glu Arg Gly Gln Glu Phe Arg Arg Pro		
545	550	555
Cys Thr Leu Ser Glu Leu Glu		
	565	

<210>16

<211>928

<212>DNA

<213>人工序列

<220>

<223> 含有与植物脯氨酰 4- 羟化酶植物融合的大麦巯基蛋白酶 aleurain 前体基因的
维管
信号序列编码区和侧翼区的合成序列

<400>16

```

ctcgagtaaa ccatggctca tgctagggtt ttgcttttgg ctcttgctgt tcttgctact 60
gctgctgttg ctgtggcttc ttcttcatct ttcgctgatt ctaacccaat taggccagt 120
actgatagag ctgcttctac tcttgctcaa ttggctgaca tgcttggtat tctttctctt 180
ccaaacgcta acaggaactc ttctaagact aacgatctta ctaacattgt gaggaagtct 240
gagacttctt ctggagatga ggagggaaat ggagaaagat ggggtggaagt gatttcttgg 300
gagccaaggg ctgtttgtta ccacaacttc cttactaatg aggagtgcga gcaccttatt 360
tctcttgcta agccatctat ggtgaagtct actgtggtgg atgagaaaac tggaggatct 420
aaggattcaa gagtgaggac ttcatctggt actttcctta ggaggggaca tgatgaagtt 480
gtggaagtta ttgagaagag gatttctgat ttcactttca ttccagtgga gaacggagaa 540
ggacttcaag ttcttcacta ccaagtggga caaaagtacg agccacacta cgattacttc 600
cttgatgagt tcaacactaa gaacggagga cagaggattg ctactgtgct tatgtacett 660
tctgatgtgg atgatggagg agagactggt tttccagctg ctaggggaaa catttctgct 720
gttccttggg ggaacgagct ttctaagtgt ggaaaggagg gactttctgt gcttccaaag 780
aaaagggatg ctcttctttt ctggaacatg aggccagatg cttctcttga tccatcttct 840
cttcatggag gatgcccagt tgtaaggga aacaagtggc catctactaa gtggttccac 900
gtgcacgagt tcaagtgta atgagctc 928

```

<210>17

<211>302

<212>PRT

<213> 人工序列

<220>

<223> 含有与植物脯氨酰 4- 羟化酶植物融合的大麦巯基蛋白酶 aleurain 前体基因的
维管

信号序列和侧翼区的合成序列

<400>17

```

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

```

Asn	Glu	Leu	Ser	Lys	Cys	Gly	Lys	Glu	Gly	Leu	Ser	Val	Leu	Pro	Lys
				245					250					255	
Lys	Arg	Asp	Ala	Leu	Leu	Phe	Trp	Asn	Met	Arg	Pro	Asp	Ala	Ser	Leu
			260					265					270		
Asp	Pro	Ser	Ser	Leu	His	Gly	Gly	Cys	Pro	Val	Val	Lys	Gly	Asn	Lys
			275					280					285		
Trp	Ser	Ser	Thr	Lys	Trp	Phe	His	Val	His	Glu	Phe	Lys	Val		
			290				295					300			

<210>18

<211>2689

<212>DNA

<213>人工序列

<220>

<223>含有人前胶原 C- 蛋白酶编码区和侧翼区的合成序列

<400>18

```

agatctatcg atgcatgcca tggtagcgcg ccatggctca attggctgca acatcaaggc 60
ctgaaagagt ttggccagat ggtgttattc ctttcgttat tggtagaac tttactggat 120
ctcagagagc agtttttaga caagctatga gacattggga aaagcacact tgtgtgacat 180
tccttgaaag gactgatgaa gattcttata ttgtgttcac ataccgcca tgtggatgct 240
gctcatatgt tggtagaagg ggaggaggtc cacaagcaat ttctattgga aaaaactgcg 300
ataagtccgg aattgtggtg catgaattgg gacatgttgt tggtttctgg cacgaacaca 360
caaggccaga tagggatagg cacgtgtcta ttgtgaggga aaacattcag ccaggccaag 420
agtacaatth tcttaagatg gaacctcaag aggtggaatc tctcggagag acttacgact 480
tcgactccat catgcactac gcaaggaata ctttcagcag gggcatcttc ttggatacca 540
ttgtgcctaa gtacgaggtg aacggcgta agccacctat tggtagaac actaggctct 600
ctaagggtga tattgcacag gctaggaagc tctacaaatg tccagcatgc ggagaaactc 660
ttcaggattc cactggcaac ttctcatctc cagagtacc aaacggatac tctgtctata 720
tgcactgtgt ttggaggatc tcagtgactc ctggagagaa gatcatcctc aacttcaact 780
ccctcgatct ctatcgttct aggcctctgtt ggtacgacta tgtggaagtg agagatggt 840
tctggagaaa ggctccactt agaggaaggt tctgcggatc taaacttct gagccaatcg 900
tgtctactga ttccagattg tgggtggagt tcaggctctc ttctaattgg gttggcaagg 960
gcttttttgc tgtgtacgag gctatttgtg gcggcgacgt gaaaaaggac tacggacata 1020
ttcaaagtcc aaattacca gatgattacc gtccttcaaa agtgtgtatt tggaggattc 1080
aagtgagtga gggtttccat gttggattga cattccaatc ttctgaaatt gagagacag 1140
attcatgcgc atacgattat ttggaagtga gagatggaca ctctgaatct tctacactta 1200
ttggaaggta ctgcggttat gagaaacctg atgatattaa gtctacttct agtaggttgt 1260

```


ggcttaaatt	tgtgtcagat	ggttctatta	acaaggctgg	tttcgcagtg	aacttcttca	1320
aggaagtgga	tgaatgctca	agacctaa	gaggaggatg	tgagcaaaga	tgcttaaca	1380
ctttgggaag	ttacaagtgt	tcttgcgac	ctggatacga	gttggctcct	gataagagaa	1440
gatgcgaagc	tgcttgcggt	ggttttttga	caaaattgaa	cggatctatt	acttctcctg	1500
gatggccaaa	agagtacca	cctaataaga	attgcatttg	gcagcttggt	gcacctactc	1560
agtaccgtat	ttcattgcaa	ttcgattttt	tcgagactga	gggtaatgat	gtgtgcaagt	1620
acgatttcgt	ggaagtgaga	tcaggtctta	ctgctgatag	taaattgcac	ggaaagttct	1680
gcggatctga	aaaaccagaa	gtgattacat	cacagtacaa	caatatgagg	gtggagtcca	1740
aatctgataa	tactgtttct	aaaaaagggt	ttaaggcaca	tttcttttct	gataaggacg	1800
agtgtcttaa	agataatggt	ggttgccagc	aggattgcgt	gaacacattc	ggttcatatg	1860
agtccaatg	ccgtagtgga	tttgttcttc	acgataacaa	acatgattgc	aaagaggcag	1920
gttgcgatca	caaggtgaca	tctacttcag	gtactatcac	atctccaaac	tggcctgata	1980
agtatccttc	aaaaaaagaa	tgtacatggg	caatttcttc	tacaccaggt	catagggtta	2040
agttgacatt	catggagatg	gatattgaga	gtcaaccaga	gtgcgcttat	gatcatcttg	2100
aggtgttcga	tggaagggat	gctaaggctc	ctgttcttgg	tagattctgt	gtagtaaaa	2160
agccagaacc	agtgtcttga	acaggatcta	ggatgttctc	tagattctac	tctgataact	2220
cagttcagag	gaaaggattc	caagctagtc	acgcaactga	atgcggtgga	caagttagag	2280
cagatgttaa	gactaaggat	ctttactcac	acgcacagtt	cggagataac	aactaccctg	2340
gaggagtcca	ttgcgagtgg	gttattgtgg	ctgaagaggg	atacggagtt	gagcttgttt	2400
tccagacatt	cgaggtggag	gaggaaactg	attgcggtta	cgattatatg	gaactttttg	2460
atggatacga	tagtactgct	ccaagacttg	gaaggtattg	tggtagtggg	ccaccagaag	2520
aggtgtactc	agctggagat	agtgttcttg	ttaagttcca	cagtgatgat	acaattacta	2580
agaagggatt	ccatcttaga	tataactcaa	ctaagttcca	ggatactctt	cattctagga	2640
agtaatgagc	tcgcggccgc	atccaagctt	ctgcagacgc	gtcgacgctc		2689

<210>19

<211>870

<212>PRT

<213>人工序列

<220>

<223>含有人前胶原 C- 蛋白酶和侧翼区的合成序列

<400>19

Met	Ala	Gln	Leu	Ala	Ala	Thr	Ser	Arg	Pro	Glu	Arg	Val	Trp	Pro	Asp
1				5					10					15	
Gly	Val	Ile	Pro	Phe	Val	Ile	Gly	Gly	Asn	Phe	Thr	Gly	Ser	Gln	Arg
			20					25						30	

Ala Val Phe Arg Gln Ala Met Arg His Trp Glu Lys His Thr Cys Val
35 40 45
Thr Phe Leu Glu Arg Thr Asp Glu Asp Ser Tyr Ile Val Phe Thr Tyr
50 55 60
Arg Pro Cys Gly Cys Cys Ser Tyr Val Gly Arg Arg Gly Gly Gly Pro
65 70 75 80
Gln Ala Ile Ser Ile Gly Lys Asn Cys Asp Lys Phe Gly Ile Val Val
85 90 95
His Glu Leu Gly His Val Val Gly Phe Trp His Glu His Thr Arg Pro
100 105 110
Asp Arg Asp Arg His Val Ser Ile Val Arg Glu Asn Ile Gln Pro Gly
115 120 125
Gln Glu Tyr Asn Phe Leu Lys Met Glu Pro Gln Glu Val Glu Ser Leu
130 135 140
Gly Glu Thr Tyr Asp Phe Asp Ser Ile Met His Tyr Ala Arg Asn Thr
145 150 155 160
Phe Ser Arg Gly Ile Phe Leu Asp Thr Ile Val Pro Lys Tyr Glu Val
165 170 175
Asn Gly Val Lys Pro Pro Ile Gly Gln Arg Thr Arg Leu Ser Lys Gly
180 185 190
Asp Ile Ala Gln Ala Arg Lys Leu Tyr Lys Cys Pro Ala Cys Gly Glu
195 200 205
Thr Leu Gln Asp Ser Thr Gly Asn Phe Ser Ser Pro Glu Tyr Pro Asn
210 215 220
Gly Tyr Ser Ala His Met His Cys Val Trp Arg Ile Ser Val Thr Pro
225 230 235 240
Gly Glu Lys Ile Ile Leu Asn Phe Thr Ser Leu Asp Leu Tyr Arg Ser
245 250 255
Arg Leu Cys Trp Tyr Asp Tyr Val Glu Val Arg Asp Gly Phe Trp Arg
260 265 270
Lys Ala Pro Leu Arg Gly Arg Phe Cys Gly Ser Lys Leu Pro Glu Pro
275 280 285
Ile Val Ser Thr Asp Ser Arg Leu Trp Val Glu Phe Arg Ser Ser Ser
290 295 300
Asn Trp Val Gly Lys Gly Phe Phe Ala Val Tyr Glu Ala Ile Cys Gly
305 310 315 320
Gly Asp Val Lys Lys Asp Tyr Gly His Ile Gln Ser Pro Asn Tyr Pro
325 330 335
Asp Asp Tyr Arg Pro Ser Lys Val Cys Ile Trp Arg Ile Gln Val Ser

	340		345		350
Glu Gly Phe His Val Gly Leu Thr Phe Gln Ser Phe Glu Ile Glu Arg					
	355		360		365
His Asp Ser Cys Ala Tyr Asp Tyr Leu Glu Val Arg Asp Gly His Ser					
	370		375		380
Glu Ser Ser Thr Leu Ile Gly Arg Tyr Cys Gly Tyr Glu Lys Pro Asp					
385		390		395	400
Asp Ile Lys Ser Thr Ser Ser Arg Leu Trp Leu Lys Phe Val Ser Asp					
	405		410		415
Gly Ser Ile Asn Lys Ala Gly Phe Ala Val Asn Phe Phe Lys Glu Val					
	420		425		430
Asp Glu Cys Ser Arg Pro Asn Arg Gly Gly Cys Glu Gln Arg Cys Leu					
	435		440		445
Asn Thr Leu Gly Ser Tyr Lys Cys Ser Cys Asp Pro Gly Tyr Glu Leu					
	450		455		460
Ala Pro Asp Lys Arg Arg Cys Glu Ala Ala Cys Gly Gly Phe Leu Thr					
465		470		475	480
Lys Leu Asn Gly Ser Ile Thr Ser Pro Gly Trp Pro Lys Glu Tyr Pro					
	485		490		495
Pro Asn Lys Asn Cys Ile Trp Gln Leu Val Ala Pro Thr Gln Tyr Arg					
	500		505		510
Ile Ser Leu Gln Phe Asp Phe Phe Glu Thr Glu Gly Asn Asp Val Cys					
	515		520		525
Lys Tyr Asp Phe Val Glu Val Arg Ser Gly Leu Thr Ala Asp Ser Lys					
	530		535		540
Leu His Gly Lys Phe Cys Gly Ser Glu Lys Pro Glu Val Ile Thr Ser					
545		550		555	560
Gln Tyr Asn Asn Met Arg Val Glu Phe Lys Ser Asp Asn Thr Val Ser					
	565		570		575
Lys Lys Gly Phe Lys Ala His Phe Phe Ser Asp Lys Asp Glu Cys Ser					
	580		585		590
Lys Asp Asn Gly Gly Cys Gln Gln Asp Cys Val Asn Thr Phe Gly Ser					
	595		600		605
Tyr Glu Cys Gln Cys Arg Ser Gly Phe Val Leu His Asp Asn Lys His					
	610		615		620
Asp Cys Lys Glu Ala Gly Cys Asp His Lys Val Thr Ser Thr Ser Gly					
625		630		635	640
Thr Ile Thr Ser Pro Asn Trp Pro Asp Lys Tyr Pro Ser Lys Lys Glu					
	645		650		655

Cys Thr Trp Ala Ile Ser Ser Thr Pro Gly His Arg Val Lys Leu Thr
 660 665 670
 Phe Met Glu Met Asp Ile Glu Ser Gln Pro Glu Cys Ala Tyr Asp His
 675 680 685
 Leu Glu Val Phe Asp Gly Arg Asp Ala Lys Ala Pro Val Leu Gly Arg
 690 695 700
 Phe Cys Gly Ser Lys Lys Pro Glu Pro Val Leu Ala Thr Gly Ser Arg
 705 710 715 720
 Met Phe Leu Arg Phe Tyr Ser Asp Asn Ser Val Gln Arg Lys Gly Phe
 725 730 735
 Gln Ala Ser His Ala Thr Glu Cys Gly Gly Gln Val Arg Ala Asp Val
 740 745 750
 Lys Thr Lys Asp Leu Tyr Ser His Ala Gln Phe Gly Asp Asn Asn Tyr
 755 760 765
 Pro Gly Gly Val Asp Cys Glu Trp Val Ile Val Ala Glu Glu Gly Tyr
 770 775 780
 Gly Val Glu Leu Val Phe Gln Thr Phe Glu Val Glu Glu Glu Thr Asp
 785 790 795 800
 Cys Gly Tyr Asp Tyr Met Glu Leu Phe Asp Gly Tyr Asp Ser Thr Ala
 805 810 815
 Pro Arg Leu Gly Arg Tyr Cys Gly Ser Gly Pro Pro Glu Glu Val Tyr
 820 825 830
 Ser Ala Gly Asp Ser Val Leu Val Lys Phe His Ser Asp Asp Thr Ile
 835 840 845
 Thr Lys Lys Gly Phe His Leu Arg Tyr Thr Ser Thr Lys Phe Gln Asp
 850 855 860
 Thr Leu His Ser Arg Lys
 865 870

<210>20

<211>2912

<212>DNA

<213> 人工序列

<220>

<223> 含有人前胶原 I N- 蛋白酶编码区和侧翼区的合成序列

<400>20

gcgccatggc tcaattgagg agaagggcta ggagacacgc agctgatgat gattacaaca 60

ttgaagtttt	gcttgggtgtt	gatgatagtg	tggtgcaatt	ccacggaaaa	gagcatgttc	120
agaaatatct	tttgacactt	atgaatattg	tgaacgaaat	ctacatgat	gagtctttgg	180
gagcacacat	taacgtgggt	cttgtgagga	ttattcttct	ttcatacgg	aaatctatgt	240
cacttattga	gattggaaac	ccttctcagt	ctcttgagaa	tgtgtgcaga	tgggcatacc	300
ttcaacagaa	gcctgatact	ggacacgatg	agtatcacga	tcacgctatt	ttccttacia	360
ggcaggattt	cggccaagt	ggaatgcaag	gatatgctcc	tgttactgg	atgtgccacc	420
ctgttaggtc	ttgtacactt	aaccacgagg	atggtttttc	atctgctttc	gtgggtggctc	480
atgagacagg	tcatgttttg	ggaatggaac	atgatggaca	gggtaataga	tgtggagatg	540
aagtgagact	tggttcaatt	atggctcctc	ttgttcaagc	tgcttttcat	aggttccact	600
ggagtaggtg	ttcacagcaa	gagttgagta	gataccttca	ttcttacgat	tgcttgcttg	660
atgatccatt	tgctcatgat	tggccagctt	tgcccaact	tcctggattg	cactactcta	720
tgaacgagca	gtgcagattt	gatttcggtc	ttggttacat	gatgtgcaca	gctttcagga	780
ctttcgatcc	atgcaaacag	ttgtgggtgtt	cacaccaga	taaccaatat	ttctgtaaaa	840
caaaaaaagg	tccaccactt	gatggtacta	tgtgcgcacc	tggaaagcac	tgcttcaagg	900
gacactgcat	ttggcttact	cctgatattc	ttaaaaggga	tggatcatgg	ggagcttgg	960
ctccattcgg	aagttgctca	agaacttgcg	gaacaggtgt	taagtttaga	actaggcagt	1020
gcgataatcc	acaccctgct	aatgggtggt	gaacttgctc	tggacttgct	tacgattttc	1080
agttgtgttc	taggcaagat	tgccctgata	gtcttgctga	ttttagagaa	gagcaatgta	1140
gacagtggga	tctttacttt	gagcacggcg	acgctcagca	ccactggctt	ccacacgagc	1200
atagagatgc	aaaagaaagg	tgtcaccttt	attgcgagag	tagagagact	ggagaggtgg	1260
tgtcaatgaa	gagaatggtg	cacgatggta	caaggtgttc	ttataaggat	gcattctctt	1320
tgtgtgtgag	gggagattgc	aggaaagtgg	gttgtgatgg	agtgattgga	tctagtaagc	1380
aagaagataa	gtgcggagtg	tgcggaggag	ataactctca	ttgcaagggt	gtgaaaggaa	1440
cttttacaag	atcaccaaaa	aaacacgggt	acattaagat	gttcgaaatt	cctgctggag	1500
caaggcattt	gcttattcag	gaagtggatg	caacatctca	ccacttggca	gtgaaaaacc	1560
ttgagactgg	aaaattcatt	ttgaacgagg	agaacgatgt	tgatgcatct	agtaagactt	1620
tcattgcaat	gggtgttgaa	tgggagtata	gggatgagga	tggaagggaa	acacttcaaa	1680
caatgggtcc	tcttcatgga	acaattactg	tgttggatg	tccagtggga	gatacaaggg	1740
tgtcattgac	atacaagtat	atgattcacg	aggatagtct	taacgttgat	gataacaacg	1800
ttttggaaga	agattctgtg	gtttacgagt	gggctcttaa	gaaatggta	ccttgcctca	1860
agccatgtgg	tggaggaagt	cagttcacta	agtatggttg	taggaggagg	cttgatcata	1920
agatggttca	taggggattt	tgcgcagcac	ttagtaagcc	aaaggcaatt	aggagggtt	1980
gtaacctca	agaatgctca	caaccagttt	gggtgacagg	agagtgggag	ccatgttcc	2040
aaacatgcgg	aagaactgga	atgcaagtta	gatcagttag	atgcattcaa	cctcttcatg	2100
ataacactac	aagaagtgtg	cacgcaaac	actgtaacga	tgctaggcca	gagagtagaa	2160
gagcttgctc	tagggaactt	tgccctggta	gatggagggc	aggaccttgg	agtcagtget	2220
ctgtgacatg	tggaaacgg	actcaggaaa	gacctgttcc	atgtagaact	gctgatgata	2280
gtttcggaat	ttgtcaggag	gaaaggccag	aaacagctag	gactttaga	cttgacett	2340
gtcctaggaa	tatttctgat	cctagtaaaa	aatcatacgt	ggtgcaatgg	ttgagtaggc	2400

cagatccaga ttcaccaatt aggaagattt cttcaaaagg aactgccag ggtgataaga 2460
 gtatcttctg cagaatggaa gttcttagta ggtactgttc tattccaggt tataacaaac 2520
 tttcttgtaa gagttgcaac ttgtataaca atcttactaa cgtggagggt agaattgaac 2580
 ctccaccagg aaagcacaac gatattgatg tgtttatgcc tactcttcct gtgccaacag 2640
 ttgcaatgga agttagacct tctccateta ctccacttga ggtgccactt aatgcatcaa 2700
 gtactaacgc tactgaggat cacccagaga ctaacgcagt tgatgagcct tataagattc 2760
 acggacttga ggatgagggt cagccaccaa accttattcc taggaggcca agtccttacg 2820
 aaaaaactag aatcagagg attcaggagc ttattgatga gatgaggaaa aaggagatgc 2880
 ttggaaagtt ctaatgagct cgcggccgca tc 2912

<210>21

<211>962

<212>PRT

<213>人工序列

<220>

<223>含有人前胶原 I N-蛋白酶和侧翼区的合成序列

<400>21

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

145	150	155	160
Thr Gly His Val Leu Gly Met Glu His Asp Gly Gln Gly Asn Arg Cys			
	165	170	175
Gly Asp Glu Val Arg Leu Gly Ser Ile Met Ala Pro Leu Val Gln Ala			
	180	185	190
Ala Phe His Arg Phe His Trp Ser Arg Cys Ser Gln Gln Glu Leu Ser			
	195	200	205
Arg Tyr Leu His Ser Tyr Asp Cys Leu Leu Asp Asp Pro Phe Ala His			
	210	215	220
Asp Trp Pro Ala Leu Pro Gln Leu Pro Gly Leu His Tyr Ser Met Asn			
	225	230	235
Glu Gln Cys Arg Phe Asp Phe Gly Leu Gly Tyr Met Met Cys Thr Ala			
	245	250	255
Phe Arg Thr Phe Asp Pro Cys Lys Gln Leu Trp Cys Ser His Pro Asp			
	260	265	270
Asn Pro Tyr Phe Cys Lys Thr Lys Lys Gly Pro Pro Leu Asp Gly Thr			
	275	280	285
Met Cys Ala Pro Gly Lys His Cys Phe Lys Gly His Cys Ile Trp Leu			
	290	295	300
Thr Pro Asp Ile Leu Lys Arg Asp Gly Ser Trp Gly Ala Trp Ser Pro			
	305	310	315
Phe Gly Ser Cys Ser Arg Thr Cys Gly Thr Gly Val Lys Phe Arg Thr			
	325	330	335
Arg Gln Cys Asp Asn Pro His Pro Ala Asn Gly Gly Arg Thr Cys Ser			
	340	345	350
Gly Leu Ala Tyr Asp Phe Gln Leu Cys Ser Arg Gln Asp Cys Pro Asp			
	355	360	365
Ser Leu Ala Asp Phe Arg Glu Glu Gln Cys Arg Gln Trp Asp Leu Tyr			
	370	375	380
Phe Glu His Gly Asp Ala Gln His His Trp Leu Pro His Glu His Arg			
	385	390	395
Asp Ala Lys Glu Arg Cys His Leu Tyr Cys Glu Ser Arg Glu Thr Gly			
	405	410	415
Glu Val Val Ser Met Lys Arg Met Val His Asp Gly Thr Arg Cys Ser			
	420	425	430
Tyr Lys Asp Ala Phe Ser Leu Cys Val Arg Gly Asp Cys Arg Lys Val			
	435	440	445
Gly Cys Asp Gly Val Ile Gly Ser Ser Lys Gln Glu Asp Lys Cys Gly			
	450	455	460

Val Cys Gly Gly Asp Asn Ser His Cys Lys Val Val Lys Gly Thr Phe
 465 470 475 480
 Thr Arg Ser Pro Lys Lys His Gly Tyr Ile Lys Met Phe Glu Ile Pro
 485 490 495
 Ala Gly Ala Arg His Leu Leu Ile Gln Glu Val Asp Ala Thr Ser His
 500 505 510
 His Leu Ala Val Lys Asn Leu Glu Thr Gly Lys Phe Ile Leu Asn Glu
 515 520 525
 Glu Asn Asp Val Asp Ala Ser Ser Lys Thr Phe Ile Ala Met Gly Val
 530 535 540
 Glu Trp Glu Tyr Arg Asp Glu Asp Gly Arg Glu Thr Leu Gln Thr Met
 545 550 555 560
 Gly Pro Leu His Gly Thr Ile Thr Val Leu Val Ile Pro Val Gly Asp
 565 570 575
 Thr Arg Val Ser Leu Thr Tyr Lys Tyr Met Ile His Glu Asp Ser Leu
 580 585 590
 Asn Val Asp Asp Asn Asn Val Leu Glu Glu Asp Ser Val Val Tyr Glu
 595 600 605
 Trp Ala Leu Lys Lys Trp Ser Pro Cys Ser Lys Pro Cys Gly Gly Gly
 610 615 620
 Ser Gln Phe Thr Lys Tyr Gly Cys Arg Arg Arg Leu Asp His Lys Met
 625 630 635 640
 Val His Arg Gly Phe Cys Ala Ala Leu Ser Lys Pro Lys Ala Ile Arg
 645 650 655
 Arg Ala Cys Asn Pro Gln Glu Cys Ser Gln Pro Val Trp Val Thr Gly
 660 665 670
 Glu Trp Glu Pro Cys Ser Gln Thr Cys Gly Arg Thr Gly Met Gln Val
 675 680 685
 Arg Ser Val Arg Cys Ile Gln Pro Leu His Asp Asn Thr Thr Arg Ser
 690 695 700
 Val His Ala Lys His Cys Asn Asp Ala Arg Pro Glu Ser Arg Arg Ala
 705 710 715 720
 Cys Ser Arg Glu Leu Cys Pro Gly Arg Trp Arg Ala Gly Pro Trp Ser
 725 730 735
 Gln Cys Ser Val Thr Cys Gly Asn Gly Thr Gln Glu Arg Pro Val Pro
 740 745 750
 Cys Arg Thr Ala Asp Asp Ser Phe Gly Ile Cys Gln Glu Glu Arg Pro
 755 760 765
 Glu Thr Ala Arg Thr Cys Arg Leu Gly Pro Cys Pro Arg Asn Ile Ser

770	775	780
Asp Pro Ser Lys Lys Ser Tyr Val Val Gln Trp Leu Ser Arg Pro Asp		
785	790	795
Pro Asp Ser Pro Ile Arg Lys Ile Ser Ser Lys Gly His Cys Gln Gly		
	805	810
Asp Lys Ser Ile Phe Cys Arg Met Glu Val Leu Ser Arg Tyr Cys Ser		
	820	825
Ile Pro Gly Tyr Asn Lys Leu Ser Cys Lys Ser Cys Asn Leu Tyr Asn		
	835	840
Asn Leu Thr Asn Val Glu Gly Arg Ile Glu Pro Pro Pro Gly Lys His		
850	855	860
Asn Asp Ile Asp Val Phe Met Pro Thr Leu Pro Val Pro Thr Val Ala		
865	870	875
Met Glu Val Arg Pro Ser Pro Ser Thr Pro Leu Glu Val Pro Leu Asn		
	885	890
Ala Ser Ser Thr Asn Ala Thr Glu Asp His Pro Glu Thr Asn Ala Val		
	900	905
Asp Glu Pro Tyr Lys Ile His Gly Leu Glu Asp Glu Val Gln Pro Pro		
915	920	925
Asn Leu Ile Pro Arg Arg Pro Ser Pro Tyr Glu Lys Thr Arg Asn Gln		
930	935	940
Arg Ile Gln Glu Leu Ile Asp Glu Met Arg Lys Lys Glu Met Leu Gly		
945	950	955
Lys Phe		960

<210>22

<211>2888

<212>DNA

<213> 人工序列

<220>

<223> 含有与人赖氨酰羟化酶3融合的大麦巯基蛋白酶 aleurain 前体基因的维管信号序列

编码区和侧翼区的合成序列

<400>22

gcgaattcgc tagctatcac tgaagagaca gcaagacaat ggtgtctcga tgcaccagaa	60
ccacatcttt gcagcagatg tgaagcagcc agagtgggtcc acaagacgca ctcagaaaag	120
gcatcttcta ccgacacaga aaaagacaac cacagctcat catccaacat gtagactgtc	180

gttatgcgtc	ggctgaagat	aagactgacc	ccaggccagc	actaaagaag	aaataatgca	240
agtggtccta	gctccacttt	agctttaata	attatgtttc	attattattc	tctgcttttg	300
ctctctatat	aaagagcttg	tattttcatt	tgaaggcaga	ggcgaacaca	cacacagaac	360
ctccctgctt	acaaaccaga	tcttaacca	tggetcacgc	tagggttttg	cttcttgctc	420
ttgctgttct	tgctactgct	gctgttgctg	tggettcttc	aagtcttttc	gctgattcta	480
accaatttag	gccagtgact	gatagagctg	cttctactct	tgctcaattg	agatctatgt	540
ctgatagacc	aagggaagg	gatccagtta	atccagagaa	gttgcttggtg	attactgtgg	600
ctactgctga	gactgaagga	taccttagat	tccttaggag	tgctgagttc	ttcaactaca	660
ctgtgaggac	tcttgactt	ggagaagaat	ggaggggagg	agatgttgct	agaactgttg	720
gaggaggaca	gaaagtgaga	tggttaaga	aagagatgga	gaagtacgct	gatagggagg	780
atatgattat	tatgttcgtg	gattcttacg	atgtgattct	tgctggatct	ccaactgagc	840
ttttgaagaa	attcgttcag	tctggatcta	ggcttctttt	ctctgctgag	tctttttggt	900
ggccagaatg	gggacttgct	gagcaatata	cagaagtggg	aactggaaag	agattcctta	960
actctggagg	attcattgga	ttcgctacta	ctattcacca	gattgtgagg	cagtggaaagt	1020
acaaggatga	cgatgatgat	cagcttttct	acactaggct	ttaccttgat	ccaggactta	1080
gggagaagtt	gtctcttaac	cttgatcaca	agtctaggat	ttcccagaac	cttaacggtg	1140
ctcttgatga	ggttgtgctt	aagttcgata	ggaacagagt	gaggattagg	aacgtggctt	1200
acgatactct	tcctattgtg	gtgcatggaa	acggaccaac	aaaactccag	cttaactacc	1260
ttggaaacta	cgttccaaac	ggatggactc	cagaaggagg	atgtggattc	tgcaatcagg	1320
ataggagaac	tcttcagga	ggacaaccac	caccaagagt	ttccttgct	gtgttcgttg	1380
aacagccaac	tccattcctt	ccaagattcc	ttcagaggct	tcttcttttg	gattaccac	1440
cagataggtt	gacacttttc	cttcacaaca	acgaggtttt	ccacgagcca	cacattgctg	1500
attcttggcc	acagcttcag	gatcatttct	ctgctgtgaa	gttggttggg	ccagaagaag	1560
ctctttctcc	aggagaagct	agggatatgg	ctatggattt	gtgcaggcag	gatccagagt	1620
gcgagtctta	cttctctctt	gatgctgatg	ctgtgcttac	taaccttcag	actcttagga	1680
ttcttattga	ggagaacagg	aaagtgattg	ctccaatgct	ttctaggcac	ggaaagtgtg	1740
ggtctaattt	ctggggtgct	cttctctctg	atgagtacta	cgctagatca	gaggactacg	1800
tggagcttgt	tcagagaaag	agagtgggag	tttggaacgt	tccttatatt	tctcaggctt	1860
acgtgattag	gggagatact	cttaggatgg	agcttcacac	gagggatggt	ttctctggat	1920
ctgatactga	tccagatatg	gctttctgca	agtctttcag	ggataagggg	attttcttc	1980
acctttctaa	ccagcatgag	ttcggaagat	tgcttgctac	ttcaagatac	gatactgagc	2040
accttcatcc	tgatctttgg	cagattttcg	ataaccagct	ggattggaag	gagcagtaca	2100
ttcacgagaa	ctactctagg	gctcttgaag	gagaaggaat	tgtggagcaa	ccatgcccag	2160
atgtttactg	gttcccactt	cttctgagc	aatgtgcca	tgagcttgtt	gctgagatgg	2220
agcattacgg	acaatggagt	ggaggtagac	atgaggattc	taggcttgct	ggaggatacg	2280
agaacgttcc	aactgtggat	attccatga	agcaagtggg	atacgaggat	caatggettc	2340
agcttcttag	gacttatgtg	ggaccaatga	ctgagtctct	tttcccagga	taccacacta	2400
aggctagggc	tgttatgaac	ttcgttgtga	ggtatcgtcc	agatgagcaa	ccatctctta	2460
ggccacacca	cgattcttct	actttcactc	ttaactgtgc	tcttaaccac	aagggacttg	2520

attatgaggg aggaggatgc cgtttcetta gatacgattg cgtgatttct tcaccaagaa 2580
 agggatgggc tcttcttcat ccaggaagge ttactcatta ccacgagga cttccaacta 2640
 cttggggaac tagatatatt atgggtgtctt tcgtggatcc atgactgctt taatgagata 2700
 tgcgagacgc ctatgategc atgatatttg ctttcaattc tgttgtgcac gttgtaaaaa 2760
 acctgagcat gtgtagctca gatcettacc gccggtttcg gttcattcta atgaatatat 2820
 caccggttac tategtatth ttatgaataa tattctccgt tcaatttact gattgtccag 2880
 aattcgcg 2888

<210>23

<211>764

<212>PRT

<213>人工序列

<220>

<223>含有与人赖氨酰羟化酶3融合的大麦巯基蛋白酶 aleurain 前体基因的维管信号序列

和侧翼区的合成序列

<400>23

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

145	150	155	160
Arg Leu Leu Phe Ser Ala Glu Ser Phe Cys Trp Pro Glu Trp Gly Leu			
	165	170	175
Ala Glu Gln Tyr Pro Glu Val Gly Thr Gly Lys Arg Phe Leu Asn Ser			
	180	185	190
Gly Gly Phe Ile Gly Phe Ala Thr Thr Ile His Gln Ile Val Arg Gln			
	195	200	205
Trp Lys Tyr Lys Asp Asp Asp Asp Asp Gln Leu Phe Tyr Thr Arg Leu			
210	215	220	
Tyr Leu Asp Pro Gly Leu Arg Glu Lys Leu Ser Leu Asn Leu Asp His			
225	230	235	240
Lys Ser Arg Ile Phe Gln Asn Leu Asn Gly Ala Leu Asp Glu Val Val			
	245	250	255
Leu Lys Phe Asp Arg Asn Arg Val Arg Ile Arg Asn Val Ala Tyr Asp			
	260	265	270
Thr Leu Pro Ile Val Val His Gly Asn Gly Pro Thr Lys Leu Gln Leu			
	275	280	285
Asn Tyr Leu Gly Asn Tyr Val Pro Asn Gly Trp Thr Pro Glu Gly Gly			
290	295	300	
Cys Gly Phe Cys Asn Gln Asp Arg Arg Thr Leu Pro Gly Gly Gln Pro			
305	310	315	320
Pro Pro Arg Val Phe Leu Ala Val Phe Val Glu Gln Pro Thr Pro Phe			
	325	330	335
Leu Pro Arg Phe Leu Gln Arg Leu Leu Leu Leu Asp Tyr Pro Pro Asp			
	340	345	350
Arg Val Thr Leu Phe Leu His Asn Asn Glu Val Phe His Glu Pro His			
	355	360	365
Ile Ala Asp Ser Trp Pro Gln Leu Gln Asp His Phe Ser Ala Val Lys			
370	375	380	
Leu Val Gly Pro Glu Glu Ala Leu Ser Pro Gly Glu Ala Arg Asp Met			
385	390	395	400
Ala Met Asp Leu Cys Arg Gln Asp Pro Glu Cys Glu Phe Tyr Phe Ser			
	405	410	415
Leu Asp Ala Asp Ala Val Leu Thr Asn Leu Gln Thr Leu Arg Ile Leu			
	420	425	430
Ile Glu Glu Asn Arg Lys Val Ile Ala Pro Met Leu Ser Arg His Gly			
	435	440	445
Lys Leu Trp Ser Asn Phe Trp Gly Ala Leu Ser Pro Asp Glu Tyr Tyr			
450	455	460	

Ala Arg Ser Glu Asp Tyr Val Glu Leu Val Gln Arg Lys Arg Val Gly
465 470 475 480
Val Trp Asn Val Pro Tyr Ile Ser Gln Ala Tyr Val Ile Arg Gly Asp
485 490 495
Thr Leu Arg Met Glu Leu Pro Gln Arg Asp Val Phe Ser Gly Ser Asp
500 505 510
Thr Asp Pro Asp Met Ala Phe Cys Lys Ser Phe Arg Asp Lys Gly Ile
515 520 525
Phe Leu His Leu Ser Asn Gln His Glu Phe Gly Arg Leu Leu Ala Thr
530 535 540
Ser Arg Tyr Asp Thr Glu His Leu His Pro Asp Leu Trp Gln Ile Phe
545 550 555 560
Asp Asn Pro Val Asp Trp Lys Glu Gln Tyr Ile His Glu Asn Tyr Ser
565 570 575
Arg Ala Leu Glu Gly Glu Gly Ile Val Glu Gln Pro Cys Pro Asp Val
580 585 590
Tyr Trp Phe Pro Leu Leu Ser Glu Gln Met Cys Asp Glu Leu Val Ala
595 600 605
Glu Met Glu His Tyr Gly Gln Trp Ser Gly Gly Arg His Glu Asp Ser
610 615 620
Arg Leu Ala Gly Gly Tyr Glu Asn Val Pro Thr Val Asp Ile His Met
625 630 635 640
Lys Gln Val Gly Tyr Glu Asp Gln Trp Leu Gln Leu Leu Arg Thr Tyr
645 650 655
Val Gly Pro Met Thr Glu Ser Leu Phe Pro Gly Tyr His Thr Lys Ala
660 665 670
Arg Ala Val Met Asn Phe Val Val Arg Tyr Arg Pro Asp Glu Gln Pro
675 680 685
Ser Leu Arg Pro His His Asp Ser Ser Thr Phe Thr Leu Asn Val Ala
690 695 700
Leu Asn His Lys Gly Leu Asp Tyr Glu Gly Gly Gly Cys Arg Phe Leu
705 710 715 720
Arg Tyr Asp Cys Val Ile Ser Ser Pro Arg Lys Gly Trp Ala Leu Leu
725 730 735
His Pro Gly Arg Leu Thr His Tyr His Glu Gly Leu Pro Thr Thr Trp
740 745 750
Gly Thr Arg Tyr Ile Met Val Ser Phe Val Asp Pro
755 760

<210>24

<211>45

<212>PRT

<213> 人工序列

<220>

<223> 大麦巯基蛋白酶 aleurain 前体基因的液泡信号序列

<400>24

```

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

```

<210>25

<211>24

<212>DNA

<213> 人工序列

<220>

<223> 单链 DNA 寡核苷酸

<400>25

atcaccagga gaacaggac catc

24

<210>26

<211>29

<212>DNA

<213> 人工序列

<220>

<223> 单链 DNA 寡核苷酸

<400>26

tccacttcca aatctctatc cctaacaac

29

<210>27

<211>23

<212>DNA

<213> 人工序列

<220>

<223> 单链 DNA 寡核苷酸

<400>27

aggcattaga ggcgataagg gag

23

<210>28

<211>27

<212>DNA

<213> 人工序列

<220>

<223> 单链 DNA 寡核苷酸

<400>28

tcaatccaat aatagccact tgaccac

27

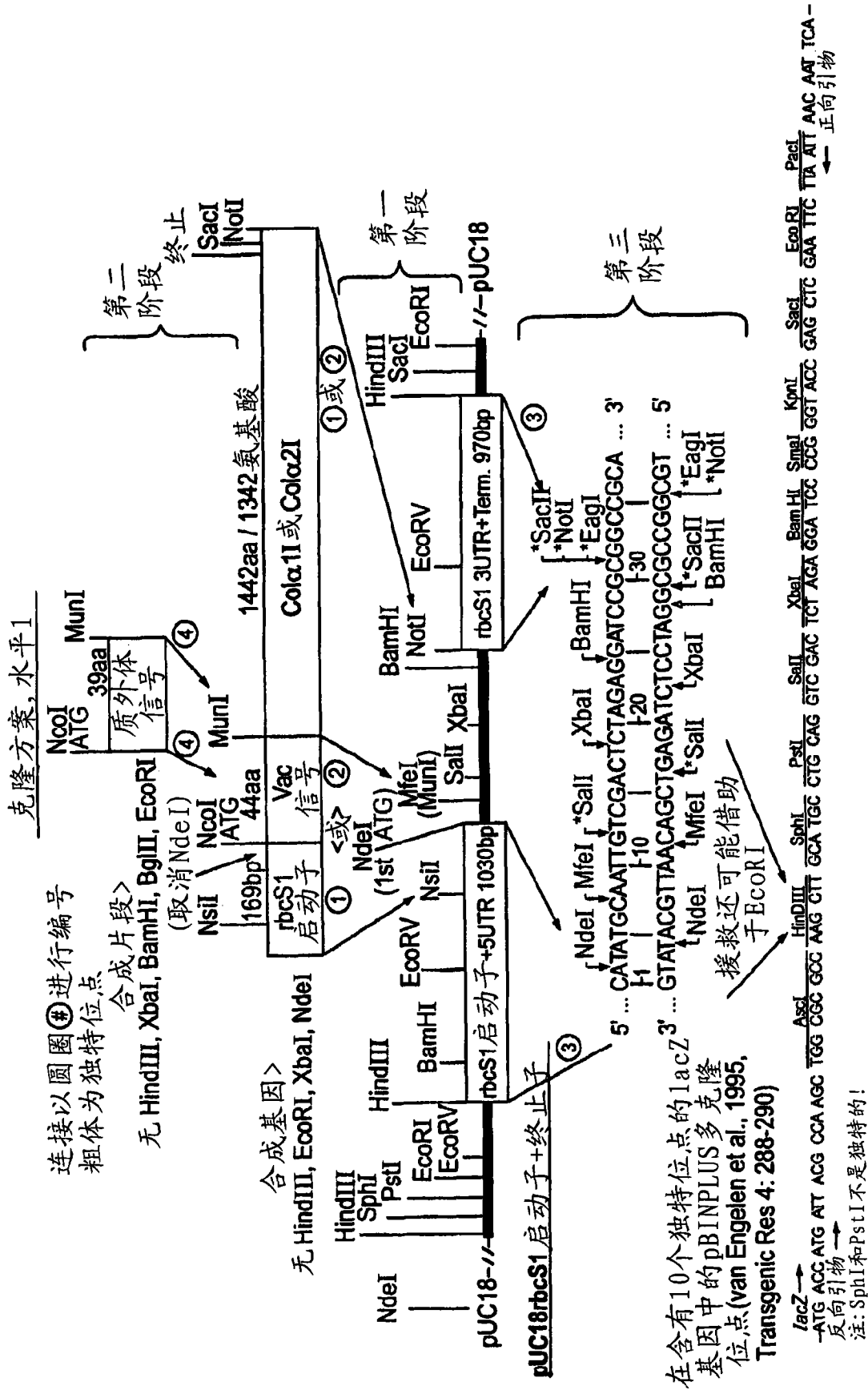


图 1a

克隆方案, 水平3

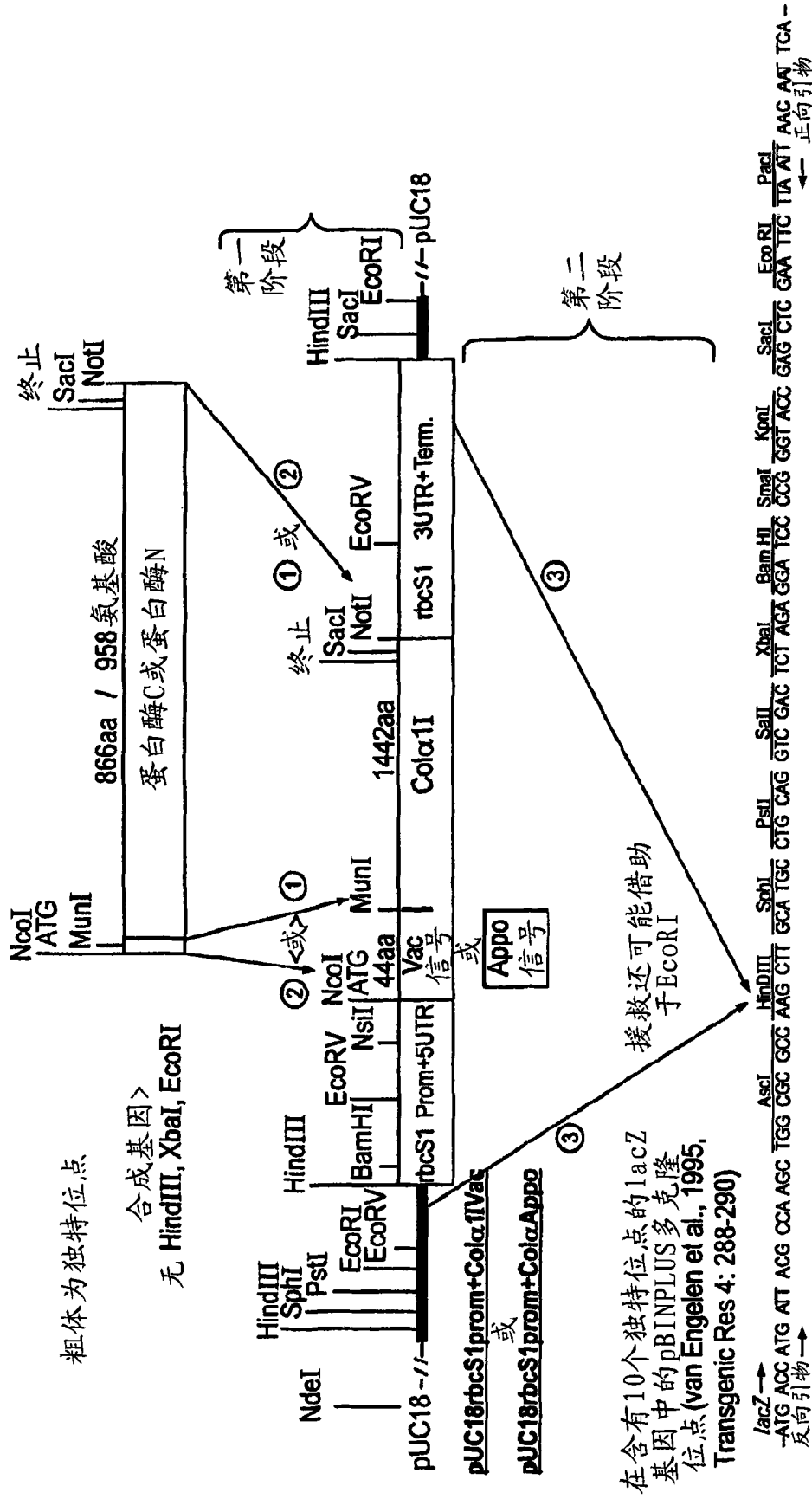


图 1c

克隆方案, 水平4

粗体为独特位点

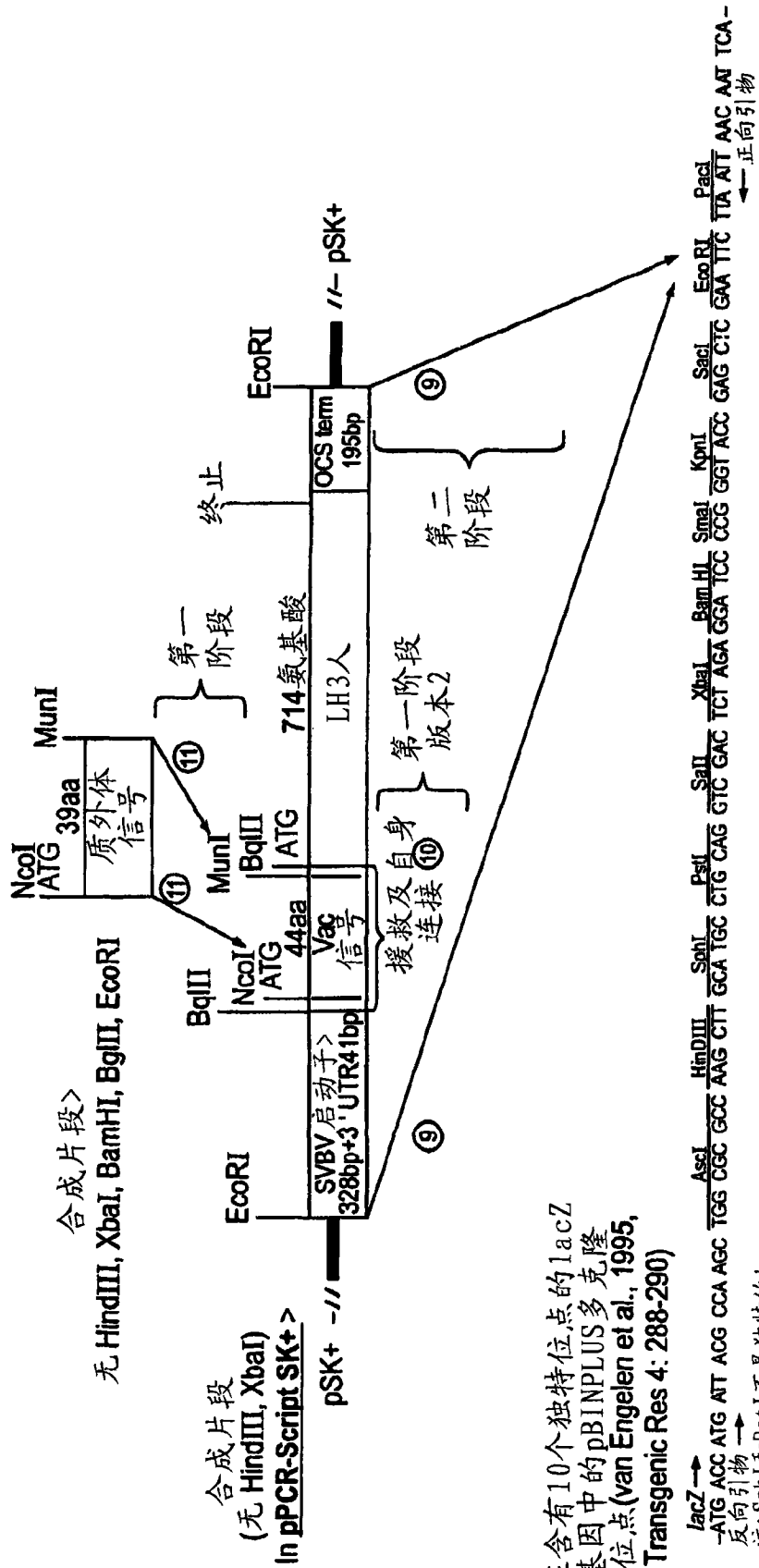


图 1d

液泡	细胞质	质外体
<p>2</p> <p>Cole a1(I) P4Hb Cole a2(I)</p>	<p>1</p> <p>Cole a1(I) Cole a2(I)</p>	<p>7</p> <p>Cole a1(I) Cole a2(I)</p>
<p>3</p> <p>Cole a1(I) P4Hb Cole a2(I) P4Ha</p>	<p>5</p> <p>Cole a1(I) P4Hb Cole a2(I) P4Ha</p>	<p>8</p> <p>Cole a1(I) P4Hb Cole a2(I) P4Ha</p>
<p>4</p> <p>Cole a1(I) P4Hb Cole a2(I) P4Hb植物</p>	<p>6</p> <p>Cole a1(I) P4Hb Cole a2(I) P4Hb植物</p>	<p>9</p> <p>Cole a1(I) P4Hb Cole a2(I) P4Hb植物</p>
<p>10</p> <p>蛋白酶C 蛋白酶N</p>	<p>11</p> <p>蛋白酶C 蛋白酶N</p>	<p>12</p> <p>蛋白酶C 蛋白酶N</p>
<p>13</p> <p>Cole a1(I) P4Hb LH3 Cole a2(I) P4Ha</p>	<p>14</p> <p>Cole a1(I) P4Hb LH3 Cole a2(I) P4Ha</p>	<p>15</p> <p>Cole a1(I) P4Hb LH3 Cole a2(I) P4Ha</p>
<p>20</p> <p>P4Hb LH3 P4Ha</p>		<p>21</p> <p>P4Hb LH3 P4Ha</p>

图 2

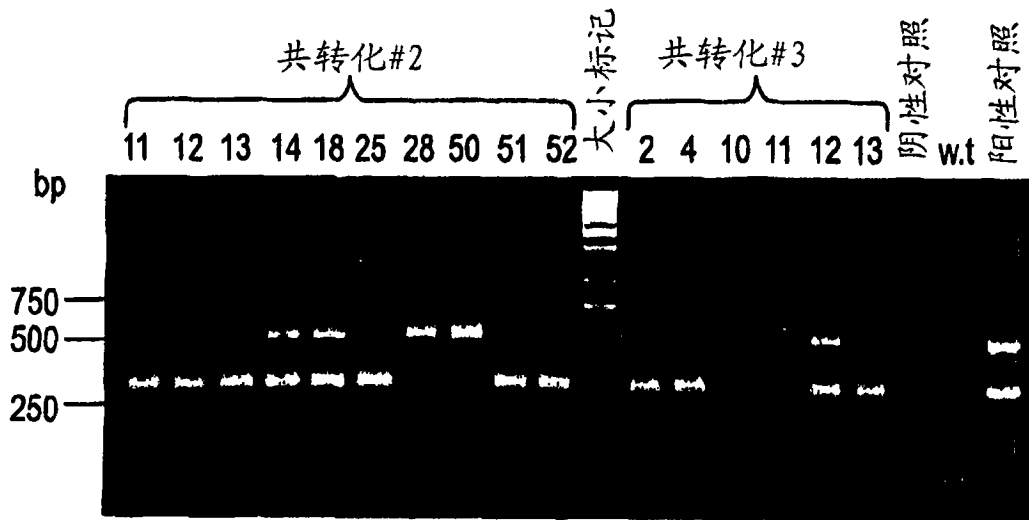


图 3

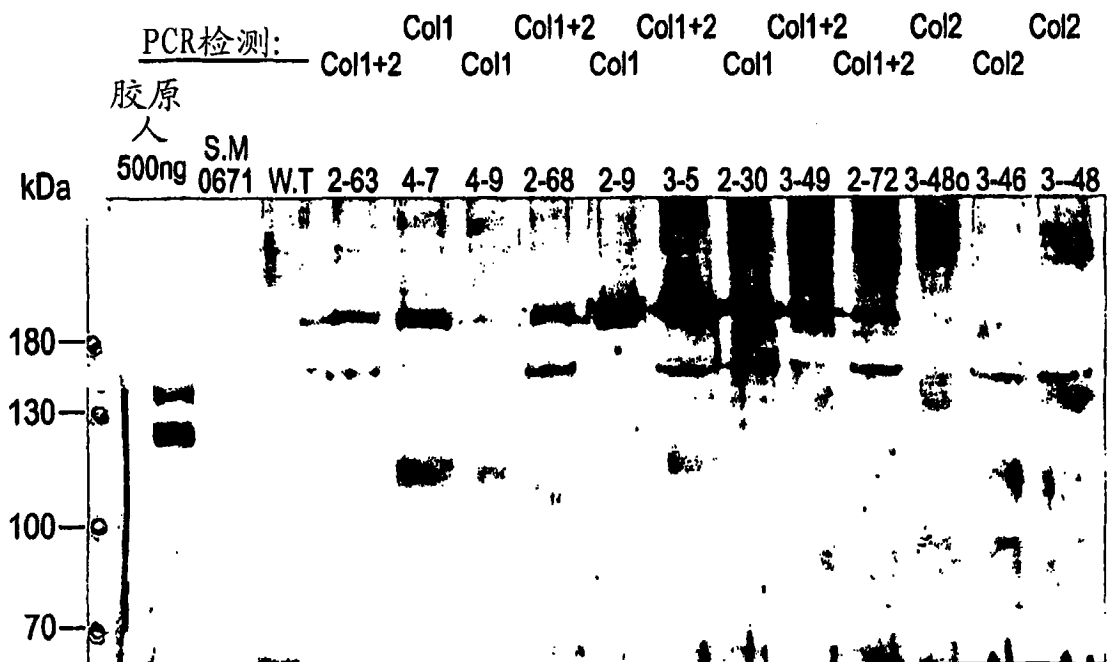


图 4

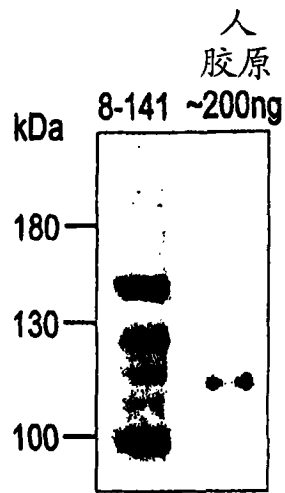


图 5

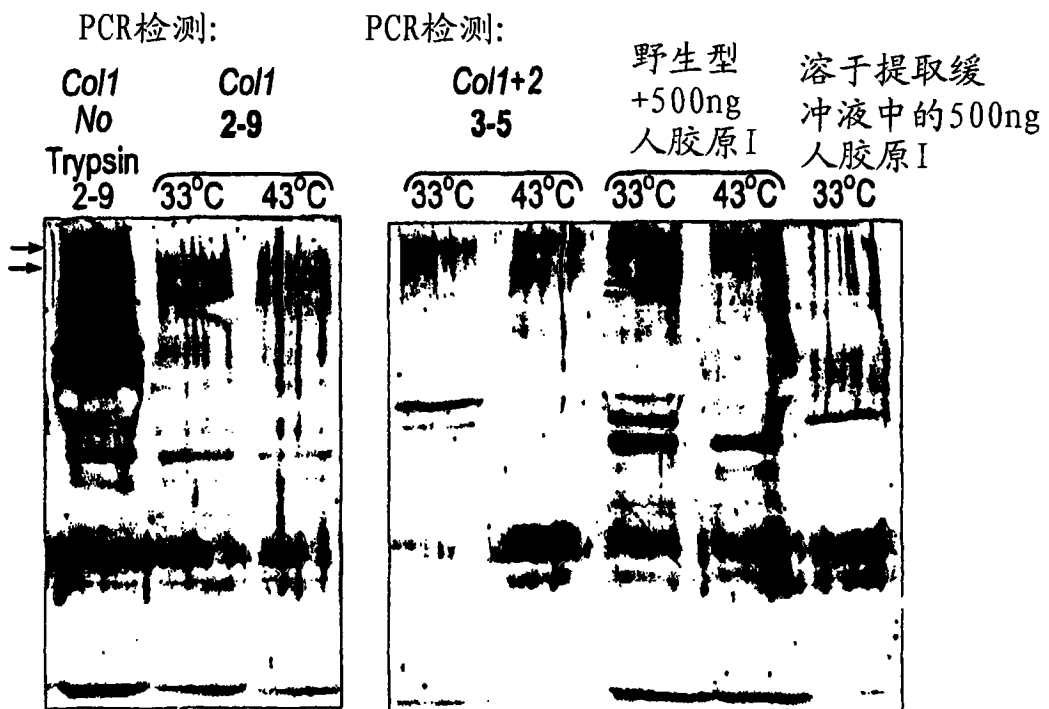


图 6a

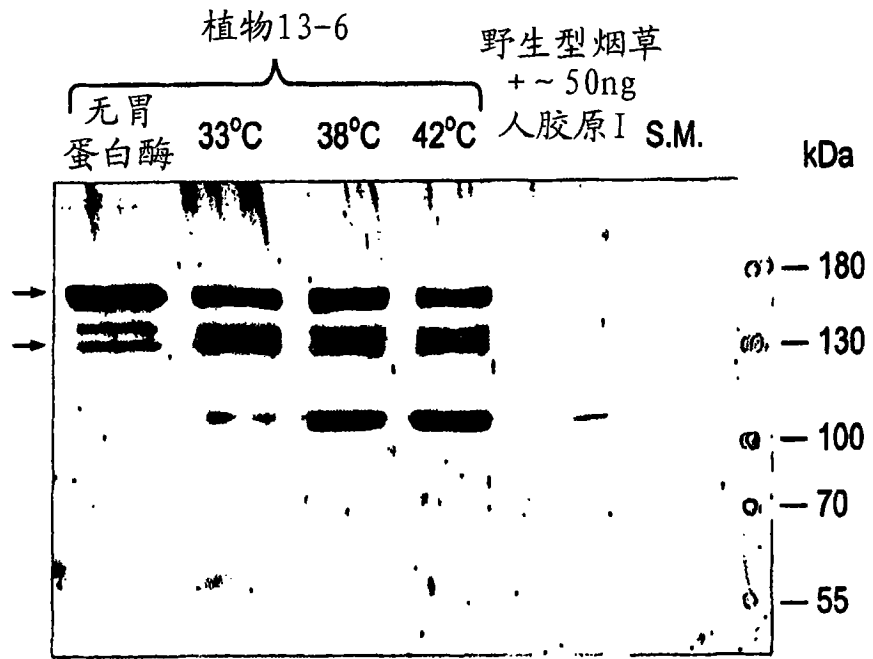


图 6b

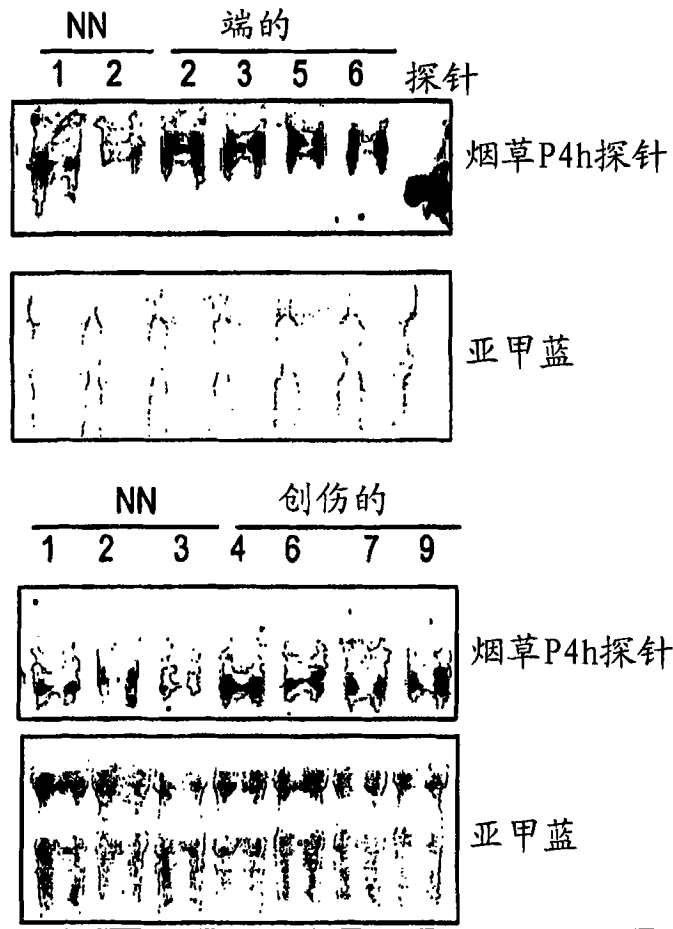


图 7

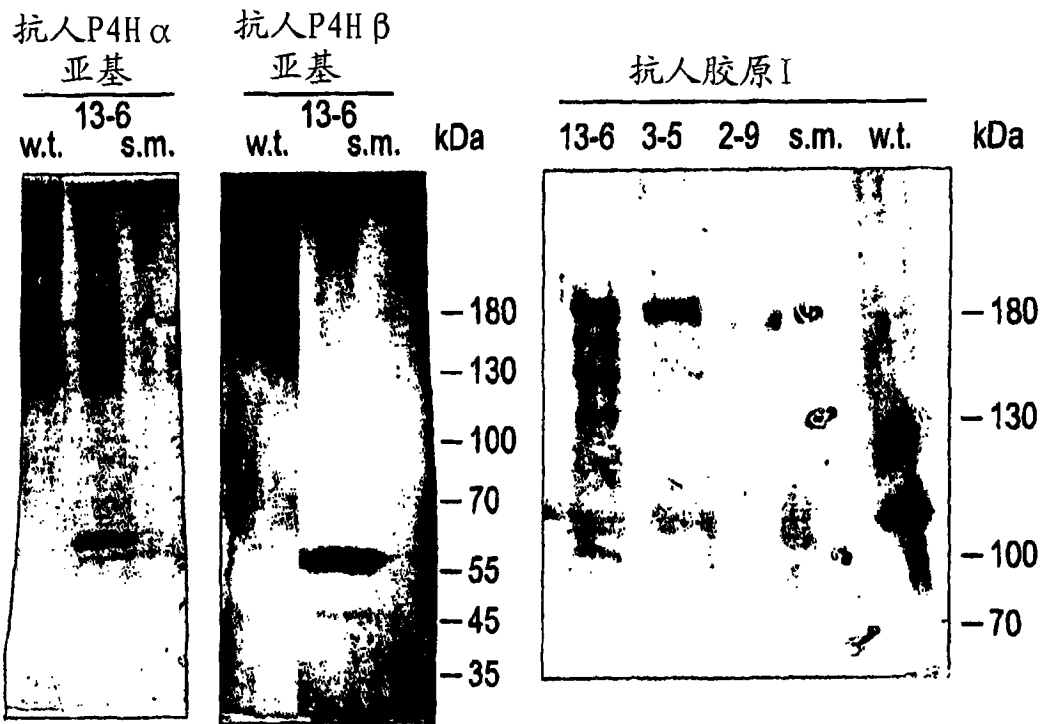


图 8